

THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

ADVANCING PULMONARY HEALTH: IMAGE PROCESSING AND MACHINE LEARNING IN FIBRE-BUNDLE ENDO-MICROSCOPY SYSTEMS

TAREK HALOUBI



Doctor of Philosophy

School of Engineering

The University of Edinburgh

2025

Tarek Haloubi: *Advancing Pulmonary Health: Image Processing and Machine Learning in Fibre-Bundle Endo-Microscopy Systems*, Doctor of Philosophy, ©2025

DECLARATION

I hereby declare that this thesis and the work reported herein was composed and originated entirely by myself, under the supervision of Prof. James R. Hopgood in the School of Engineering at The University of Edinburgh. Other people who have contributed to this work are acknowledged and referenced appropriately within.

Edinburgh, 2025

Tarek Haloubi

ABSTRACT

This thesis presents a comprehensive investigation into advanced image processing and machine learning techniques aimed at enhancing the utility of real-time fluorescence lifetime imaging (FLIm) for *in vivo* pulmonary optical endomicroscopy (OEM). Addressing the inherent challenges of clinical FLIm applications—such as motion artefacts, uninformative frames, and registration difficulties—this work introduces significant advancements across several key areas.

A novel temporal reliability and accuracy via correlation enhanced registration (TRACER) pipeline is developed to integrate a sequence of pre-processing steps, including uninformative frame removal and motion characterisation via dense optical flow (OF), alongside a tracking-based normalised cross correlation (NCC) image registration method. TRACER yields marked improvements in registration performance, achieving a 20 % to 30 % enhancement across quality of alignment (QA), structural similarity index measure (SSIM), and normalised root mean squared error (NRMSE) metrics, while operating an order of magnitude faster than the next best registration method. This improved registration efficiency is critical for the real-time clinical applicability of FLIm.

A multi-objective image registration framework is also introduced, leveraging both intensity and lifetime information inherent to FLIm sequences to address complex spatio-temporal dynamics. A novel multi-objective optimisation (MOO) strategy, dynamic multi-objective search (DYNMOS), is proposed, which effectively balances intensity and lifetime contrast for improved registration robustness, particularly in scenarios involving significant displacements or repetitive structures. The use of Pareto fronts further clarifies the trade-offs between competing registration objectives.

Building on these image processing advances, the thesis explores the impact of pipeline robustness on machine learning (ML) applications for OEM-FLIm data. It addresses uninformative frame detection and automated analysis of biologically relevant probes. A transfer learning (TL) strategy employing a pre-trained VGG16-based convolutional neural network (CNN) is proposed to accurately identify uninformative frames, enhancing data quality and processing efficiency. Additionally, the effects of different pipeline elements are examined in the context of a U-Net-based deep learning framework for the automatic detection of neutrophil activation probe (NAP) signals within FLIm images. These results underscore the importance of robust processing in enabling clinically meaningful, automated disease characterisation at the alveolar level.

In conclusion, this thesis delivers a systematic suite of image processing and ML methodologies that significantly enhance the reliability, accuracy, and automation of FLIm in pulmonary imaging. The combined developments in registration (TRACER, DYNMOS) and ML-driven analysis of frame quality and NAP activity pave the way for clinically translatable applications of fibre-bundle OEM systems. Future directions should focus on refining these techniques, integrating them into real-time workflows, and broadening their applicability to a wider array of pulmonary diseases and imaging probes. The findings presented here hold strong potential for improving diagnostic precision and advancing our understanding of lung pathology at the microscopic scale.

LAY SUMMARY

Lung diseases remain a major cause of illness worldwide, yet many begin in the smallest structures of the lung, called alveoli. These tiny air sacs are responsible for transferring oxygen into the blood. Understanding how diseases affect alveoli at a very early stage could improve diagnosis and treatment, but seeing them clearly inside a living, breathing patient is extremely difficult.

This PhD thesis focuses on a key challenge in lung imaging: how to obtain clear, stable microscopic images from deep inside the lung while the patient continues to breathe. Conventional medical imaging methods, such as X-rays and CT scans, cannot show individual alveoli in detail. A newer technique, known as optical endomicroscopy with fluorescence lifetime imaging, uses a very thin fibre-optic probe to look directly at lung tissue from inside the airways. While this technology can capture highly detailed images, it is very sensitive to movement. Normal breathing could distort the images, making them hard to interpret.

To address this problem, the thesis introduces a new software system called TRACER. This system acts like specialised video stabilisation software for lung microscopy. It automatically removes poor-quality images, tracks how lung tissue moves over time, and digitally aligns images to produce a clearer and more stable view. As a result, image quality is substantially improved, and the processing is fast enough to support future real-time use in clinical settings. These performance improvements are reported outcomes of the work and are based on experimental evaluation within the thesis. The research also develops a method for combining two types of information collected at the same time: how bright the tissue appears and how long fluorescent signals last. These two measurements provide complementary structural and chemical information about the lung. By carefully balancing both types of data during image alignment, the method helps preserve important biological detail even in complex tissue regions.

Finally, the thesis explores how artificial intelligence can be used to analyse these improved images automatically. Machine learning models are trained to quickly identify usable images and to detect biological signals linked to inflammation. This reduces the need for time-consuming manual review and improves the reliability of automated analysis when compared with unstabilised data.

In summary, this thesis develops new software tools that make it possible to stabilise, enhance, and automatically analyse microscopic images of the living lung. These advances support more detailed and reliable investigation of lung disease and represent a step towards bringing advanced optical imaging technologies into routine clinical practice.

ACKNOWLEDGMENTS

First and foremost, I wish to extend my deepest gratitude to my primary supervisor, Professor James R. Hopgood, for his unwavering guidance and steadfast support throughout my doctoral journey. His thought-provoking discussions, analytical rigour, and ability to steer my research in the right direction were indispensable in shaping the progress of this thesis. I am equally grateful to my secondary supervisor, Professor Kev Dhaliwal, whose inspiring "bigger-picture" perspective, enthusiastic mentorship, and entrepreneurial spirit continually encouraged me to push boundaries. His words of encouragement proved invaluable in sustaining my motivation and confidence.

I also owe tremendous thanks to my industrial supervisors, whose guidance and expertise were instrumental. At the National Physical Laboratory (NPL), Spencer A. Thomas provided invaluable technical insights and analytical perspectives, ensuring I stayed on track at every milestone and teaching me the importance of clear, summative discussions. Meanwhile, Catherine Hines at GSK offered leadership by example, pairing honesty with a refreshing outlook that helped me develop a strategic approach not only to my doctoral work but also to my career trajectory beyond the PhD.

I am further indebted to my mentors at NPL, Leah Chapman and Linden Fradet, for their exceptional professionalism, nurturing demeanour, and readiness to offer support. Working with them across multiple roles—be it as a Communication Ambassador, Communication Stream Leader, or University Liaison Management Support Officer—expanded my creativity, refined my strategic thinking, and underscored the value of fostering collaboration between academia and industry.

Finally, I wish to acknowledge the boundless love and encouragement I have received from my family and friends throughout this journey. Their support has been the bedrock that allowed me to persevere during the more challenging moments. In particular, I am profoundly grateful to Ieva, my devoted partner in research, whose companionship, understanding, and parallel dedication to completing her own PhD have been a constant source of inspiration and solace. I look forward to the day when we can both celebrate our graduations together.

CONTENTS

1 Introduction 1
1.1 Motivation and Aims of the Thesis . . . 1
1.2 Main Contributions . . . 4
1.3 Project outputs and recognitions . . . 6
1.4 Overview of the Thesis . . . 7
2 Background and Related Work 9
2.1 Introduction . . . 9
2.1.1 Chapter Overview . . . 11
2.2 Medical and Engineering Background . . . 11
2.2.1 Pulmonary Imaging . . . 11
2.2.2 Fluorescence Microscopy Imaging . . . 13
2.2.3 Fluorescence Lifetime Microscopy Image Representations 18
2.2.4 Fluorescence Lifetime Microscopy Image Characterisation 22
2.3 Optical Endomicroscopy Using Fluorescence Lifetime Imaging . . 27
2.3.1 The KronoScan System . . . 29
2.3.2 Biological Setup . . . 31
2.3.3 Pulmonary Motion and Respiratory Mechanics . . . 33
2.3.4 Implications for Medical Imaging . . . 33
2.4 Motion Artefacts and Their Compensation . . . 34
2.4.1 Understanding and Addressing Uninformative Frames . . 36
2.4.2 Medical Image Registration and Alignment Techniques . . 37
2.4.3 Limitations and Challenges in Image Registration . . . 40
2.5 Conclusion . . . 42
3 Temporal Reliability and Accuracy via Correlation Enhanced Registration 43
3.1 Introduction . . . 43
3.1.1 Motivation . . . 43
3.1.2 Problem Statement . . . 44
3.1.3 Chapter overview . . . 45
3.2 Processing Pipeline Development . . . 45
3.2.1 Simulation and Real Data . . . 46
3.2.2 Image Sequence Pre-processing . . . 49
3.2.3 Image-to-image characterisation . . . 50
3.2.4 Image Registration . . . 53
3.3 Addressing Image Registration Challenges . . . 55
3.3.1 Feature Channels in the Context of FLIm Sequences . . . 58
3.3.2 Channel and Spatial Reliability Peak Tracking Registration 58

3.4 Results . . . 61
3.4.1 Quantitative Analysis . . . 62
3.4.2 Statistical Assessment of Registration Performance . . . 70
3.4.3 Computational Runtime . . . 71
3.4.4 Qualitative Example . . . 72
3.4.5 Application Example to Segmentation and Detection . . . 72
3.5 Conclusion . . . 74
4 A Multi-Objective Approach to Image Registration 79
4.1 Introduction . . . 79
4.1.1 Motivation . . . 80
4.1.2 Problem Statement . . . 80
4.1.3 Chapter Overview . . . 82
4.2 Imaging Data Representation . . . 83
4.3 Algorithmic Development . . . 83
4.3.1 Formulating the Multi-Objective Function . . . 85
4.3.2 Multi-Objective Optimisation Strategies . . . 85
4.3.3 Overview of Optimisation Methods . . . 90
4.4 Analysis of Multi-Objective Optimisation . . . 95
4.4.1 Tukey's HSD Test Overview . . . 95
4.4.2 Weight Assignment Metric . . . 96
4.4.3 Quality Alignment Metric . . . 96
4.4.4 Comparative Analysis . . . 96
4.5 Analysis of the Normalised Cross-Correlation Space . . . 97
4.5.1 Intensity Dominance . . . 97
4.5.2 Lifetime Selection . . . 99
4.6 Pareto Optimisation Analysis . . . 100
4.6.1 Notation and Setup . . . 102
4.6.2 Interpretation in FLIm Context . . . 103
4.6.3 Pareto Front Analysis . . . 103
4.7 Dynamic Multi-Objective Image Registration . . . 104
4.8 Results . . . 108
4.9 Conclusion . . . 111
5 Machine Learning Applications in FLIm-OEM Data Processing 113
5.1 Introduction . . . 113
5.1.1 Motivation . . . 113
5.1.2 Problem Statement . . . 114
5.1.3 Chapter Overview . . . 115
5.2 Machine Learning for OEM-FLIm Processing . . . 116
5.2.1 From Artificial Intelligence to Transfer Learning . . . 116
5.2.2 Model-Driven vs. Data-Driven Approaches . . . 118

5.2.3 A New Approach for Uninformative Frame Detection . . . 120
5.2.4 Results . . . 124
5.3 Automatic Detection of Neutrophil Activation Probe . . . 128
5.3.1 U-Net-Based Framework . . . 129
5.3.2 Results . . . 133
5.4 Importance of a Robust Image Processing Pipeline . . . 135
5.4.1 Image Reconstruction . . . 136
5.4.2 Image Registration . . . 139
5.5 Conclusion . . . 141
6 Conclusions and Recommendations 143
6.1 Summary . . . 144
6.2 Thesis Outlook . . . 146
6.2.1 Limitations . . . 146
6.2.2 Future Work . . . 148
6.2.3 Recommendations . . . 150
6.2.4 Concluding Remarks . . . 151

LIST OF FIGURES

Figure 1.1 Visual overview of the OEM-FLIm system imaging the alveolar space . . . 2
Figure 1.2 Representative FLIm data example showing intensity, lifetime, and intensity-weighted lifetime images . . . 3
Figure 1.3 Examples of motion artefacts in OEM-FLIm images . . . 4
Figure 2.1 Hierarchical structure of the lungs showing alveolar anatomy and gas exchange . . . 10
Figure 2.2 Representative examples of conventional pulmonary imaging modalities . . . 13
Figure 2.3 Pulmonary fluorescence images showing intensity, lifetime, and intensity-weighted visualisations . . . 17
Figure 2.4 Comparing the signal-to-noise ratio (SNR), texture homogeneity and correlation between intensity and lifetime data. . . . 28
Figure 2.5 Schematic of the KronoScan OEM-FLIm system used in this study . . . 30
Figure 3.1 Overview of the TRACER pipeline for FLIm image processing . . . 45
Figure 3.2 Illustration of simulation sequences and variability scenarios used in the study . . . 47
Figure 3.3 Illustration of FLIm data sequences across different variability scenarios . . . 48
Figure 3.4 Sequential QA metric trends across four FLIm imaging scenarios . . . 50
Figure 3.5 Optical flow-based characterisation of motion direction and angular variance in simulated sequences . . . 51
Figure 3.6 Impact of image complexity on NCC peak detection and the role of channel and spatial reliability tracker (CSRT) tracking . . . 57
Figure 3.7 Quantitative evaluation of pre-processing effects on simulation and FLIm data . . . 69
Figure 3.8 Average computation time per frame for benchmarked registration methods . . . 72
Figure 3.9 Qualitative comparison of registration methods based on alignment accuracy . . . 73

Figure 3.10 Comparison of NAP region detection using different image registration methods 75
Figure 4.1 [Registration performance comparison across different FLIm image representations 81
Figure 4.2 Comparison of lifetime image representations with low and high temporal resolution 84
Figure 4.3 Figure showing a schematic comparison of the TRACER image registration and the multi-objective approach. . . 86
Figure 4.4 Performance comparison of optimisation methods for multi-objective registration. 93
Figure 4.5 Performance comparison of MOO methods across QA, runtime, and iteration count 94
Figure 4.6 Comparison of NCC maps for intensity and lifetime FLIm images 98
Figure 4.7 NCC space comparison highlighting the reliability of intensity data for registration 99
Figure 4.8 Distribution of optimal registration weights from simplicial homology global optimisation (SHGO) optimisation . . . 100
Figure 4.9 Example of lifetime modality dominance in NCC-based registration 101
Figure 4.10 Pareto fronts illustrating trade-offs between intensity and lifetime registration objectives 105
Figure 4.11 Aggregated Pareto front analysis highlighting intensity dominance across a large number of sequences 106
Figure 4.12 Performance comparison of registration strategies including DYNMOS and SHGO-optimised MOO 109
Figure 4.13 Computation time and iteration comparison between DYNMOS and SHGO methods 110
Figure 5.1 Example of NAP imaging using fluorescence intensity imaging (FII) and FLIm modalities with region highlighting 114
Figure 5.2 Examples of informative and uninformative FLIm frames from KronoScan data 117
Figure 5.3 Transfer learning TL framework using VGG16 for FLIm uninformative frame detection 122
Figure 5.4 Training and validation performance curves for the TL model 126
Figure 5.5 Sample outputs from the transfer learning model on FLIm images 127
Figure 5.6 Confusion matrix obtained by applying the TL model to the entire test set. 128

Figure 5.7 Schematic of the proposed U-Net framework for NAP detection, including input FLIm data, network architecture, and resulting binary segmentation mask 131
Figure 5.8 Training and validation performance of the U-Net model over 150 epochs, showing combined loss (left) and IoU score (right), indicating effective convergence and improved NAP segmentation capability 133
Figure 5.9 Qualitative comparison of NAP detection methods using overlay masks 134
Figure 5.10 Qualitative comparison of various image types and representations for FLIm 137

LIST OF TABLES

Table 3.1 Summary of motion vectors' angular variance across different imaging sequences. 54
Table 3.2 Impact of step 1 on registration methods and TRACER, using simulation image sequences. Simulation sequences 1 to 3 (Part 1 of 3) 63
Table 3.3 Impact of step 1 on registration methods and TRACER, using simulation image sequences. Simulation sequences 3 to 6 (Part 2 of 3) 64
Table 3.4 Impact of step 1 on registration methods and TRACER, using simulation image sequences. Simulation sequence 7 (Part 3 of 3) 64
Table 3.5 Impact of step 1 on registration methods and TRACER, using FLIm image sequences. FLIm sequences 1 to 3 (Part 1 of 2) 65
Table 3.6 Impact of step 1 on registration methods and TRACER, using FLIm image sequences. FLIm sequences 4 to 6 (Part 2 of 2) 66
Table 3.7 Impact of step 2 on registration methods and TRACER, using simulation image sequences. 67
Table 3.8 Impact of step 2 on registration methods and TRACER, using FLIm image sequences. 68
Table 3.9 Paired t-test results comparing registration methods before and after applying Step 1 and 2. Significant differences ($p < 0.05$) indicated with *. 70
Table 3.10 Comparing the detection of NAP in fused images from various registration techniques with the proposed TRACER method. 75
Table 3.11 Statistical comparison between TRACER and other registration methods for Precision, Recall, and F1-Score. . . 76
Table 4.1 Tukey honestly significant difference (HSD) Test Results for Weight Assignment Metric 96
Table 4.2 Tukey HSD Test Results for QA Metric 96
Table 5.1 A table comparing the performance of VGG16, InceptionV3, and ResNet50 in the proposed transfer learning model for uninformative frame detection. 124

Table 5.2 Comparing the TL-based approach to the one reported in [1]. 128
Table 5.3 Varying the filtering method and data representation affects U-Net performance for automatic NAP detection. 134
Table 5.4 Varying the filtering method and data representation affects U-Net performance for automatic NAP detection. 139
Table 5.5 A table to summarise different approaches to the proposed transfer learning model for uninformative frame detection. 140

ACRONYMS

ADMNet	Attention-guided densely multi-scale network
AI	artificial intelligence
AUC	area under the curve
BCE	Binary cross entropy
BFGS	Broyden-Fletcher-Goldfarb-Shanno
CBCT	cone-beam CT
CEUS	contrast enhanced ultrasound
CI	confidence interval
CNN	convolutional neural network
COPD	chronic obstructive pulmonary disease
CSRT	channel and spatial reliability tracker
CT	computed tomography
CXR	chest X-ray
DCF	discriminative correlation filter
DEAP	distributed evolutionary algorithms in Python
DL	deep learning
DYNMOS	dynamic multi-objective search
ECC	enhanced correlation coefficient
EVLV	*Ex vivo* lung ventilation
FAD	flavin adenine dinucleotide
FII	fluorescence intensity imaging
FLIm	fluorescence lifetime imaging
FLIM	fluorescence lifetime microscopy
FoV	field of view

fps	frames per second
FRET	Förster resonance energy transfer
GLCM	grey-level co-occurrence matrix
GMM	gaussian mixture modelling
HSD	honestly significant difference
IoU	intersection over union
L-BFGS-B	limited-memory Broyden-Fletcher-Goldfarb-Shanno with bounds
MC	Monte Carlo
MMI	Mattes mutual information
ML	machine learning
MOO	multi-objective optimisation
MRI	magnetic resonance imaging
NADPH	nicotinamide adenine dinucleotide phosphate
NAP	neutrophil activation probe
NCC	normalised cross correlation
NRMSE	normalised root mean squared error
NSGA-II	non-dominated sorting genetic algorithm - II
OEM	optical endomicroscopy
OF	optical flow
PET	positron emission tomography
QA	quality of alignment
RLD	rapid lifetime determination
RoC	rate of change
ROC	receiver operating characteristic
RoI	region of interest
SHGO	simplicial homology global optimisation
SNR	signal-to-noise ratio

SPAD	single-photon avalanche diode
SSIM	structural similarity index measure
TCSPC	time-correlated single photon counting
TL	transfer learning
TNC	truncated Newton constrained
TRACER	temporal reliability and accuracy via correlation enhanced registration
XAI	explainable artificial intelligence

1 INTRODUCTION

1.1 MOTIVATION AND AIMS OF THE THESIS

The lungs are among the most vital organs in the human body, enabling oxygen to reach every cell while removing carbon dioxide—a waste product of cellular activity. At the heart of this exchange are tiny air sacs known as alveoli, numbering over 400 million in a healthy adult [2]. These microscopic structures create a massive internal surface area, roughly the size of a tennis court, allowing for efficient gas exchange and playing a key role in immune defence [2, 3]. Disruptions in alveolar function, whether due to inflammation, fibrosis, infection, or cancer, can lead to severe respiratory conditions. Therefore, accurately visualising and assessing alveolar health is crucial for both diagnosis and treatment of pulmonary disease [3].

Traditional imaging techniques such as chest X-ray (CXR), computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) scans have long served as the backbone of respiratory diagnostics [4, 5, 6]. These methods offer valuable anatomical and functional perspectives but typically lack the spatial resolution to capture microscopic events within the alveoli [7, 8]. Moreover, many rely on ionising radiation or are limited by motion artefacts from breathing. Bronchoscopy allows direct visualisation of the airways, yet still falls short of capturing the biochemical and metabolic dynamics occurring at the cellular level in alveolar tissue [9]. In response, there is growing interest in optical imaging technologies capable of operating at microscopic resolution *in vivo* [7].

Among these, fibre-based optical endomicroscopy (OEM) integrated with fluorescence lifetime imaging (FLIm). OEM-FLIm stands out as a promising approach [10, 7, 8, 11]. This technology delivers pulses of light through a miniature flexible fibre to excite fluorescent molecules within lung tissue and detect their emitted signals [12, 13]. A simplified visual overview of the OEM-FLIm system and its role in imaging the alveolar space is shown in Fig. 1.1, highlighting the fibre-based access route and the endo-microscopic imaging system workflow. While intensity-based imaging provides structural detail, the fluorescence lifetime dimension—a measure of how long the signal persists after excitation—adds a powerful layer of functional insight [14, 15]. This metric is less affected by

signal brightness or probe concentration, and can reveal subtle changes in tissue metabolism, inflammation, or cellular stress [16, 17].

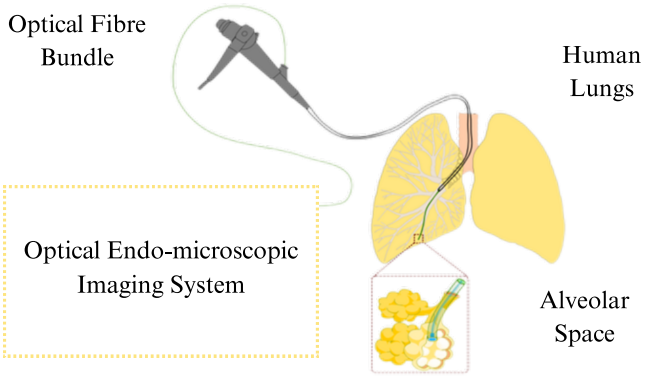
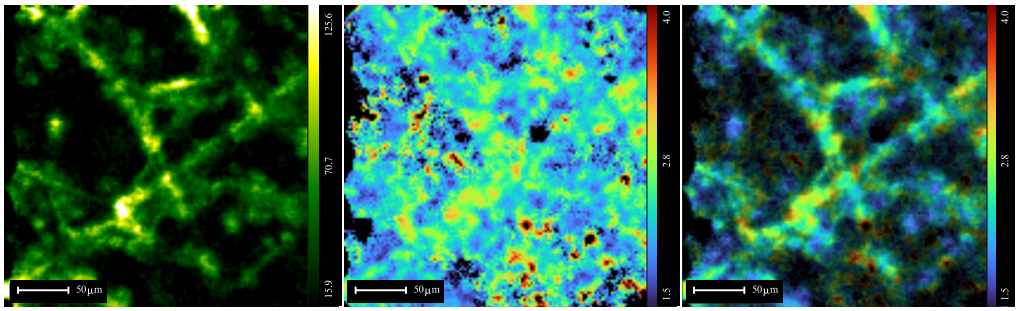


Figure 1.1: A conceptual illustration of the OEM-FLIm imaging process. An optical fibre bundle is guided into the human lung via a bronchoscope and is connected to an optical endo-microscopic imaging system. The distal end of the fibre reaches the alveolar space, where it delivers excitation light and captures the resulting fluorescence emission. This configuration enables minimally invasive imaging of structural and functional characteristics deep within the lungs.

The dual-contrast capability of OEM-FLIm, combining both intensity and lifetime information, enables real-time and functionally informative imaging from deep within the lungs. This opens avenues for earlier disease detection, intraoperative guidance, and more personalised therapeutic strategies [18, 19, 8]. In particular, the ability to visualise alveolar regions at a microscopic level offers significant potential for pulmonary diagnostics, especially in contexts where conventional imaging lacks sensitivity to early or subtle changes [17, 20]. A representative example of the FLIm data used in this work—including intensity, lifetime, and intensity-weighted lifetime images—is shown in Fig. 1.2, demonstrating the complementary contrast mechanisms captured by this imaging technology.

Nonetheless, the clinical deployment of FLIM faces several challenges. Motion artefacts from respiration and probe manipulation introduce spatial inconsistencies, geometric distortion, and signal blurring [7, 21]. Fig. 1.3 presents examples of motion-related artefacts that commonly affect OEM-FLIm image sequences. Moreover, lifetime signals often exhibit low signal-to-noise ratio (SNR), limiting the robustness of quantitative analyses [22, 17]. Therefore, the analysis of the multi-dimensional data acquired via FLIm, combining both structural and functional information, requires advanced image processing techniques. Conventional tools developed for intensity-based imaging are insufficient to extract the deeper



(a) Intensity Image (b) Lifetime Image (c) Intensity-weighted Lifetime Image

Figure 1.2: Representative pulmonary fluorescence lifetime microscopy (FLIM) images of the same alveolar region: **(a)** fluorescence intensity highlighting alveolar walls and capillaries, **(b)** fluorescence lifetime map showing pixel-wise decay times and associated biochemical contrast, and **(c)** intensity-weighted lifetime image combining structural and functional information. Together, these representations illustrate the dual structural and biochemical contrast provided by FLIm at the alveolar scale.

functional signals embedded in fluorescence lifetime. There is thus an increasing need for integrated pipelines capable of registration, denoising, and automated interpretation that fully leverage both imaging dimensions [23, 20, 24].

To address these challenges, this thesis aims to advance OEM-FLIm through the development and validation of novel image processing methodologies. The key objectives are:

- **Establish a foundational understanding of OEM-FLIm in pulmonary imaging:** Review relevant literature and clinical applications to highlight the imaging modality's potential and limitations at the alveolar level. This includes a detailed overview of the biological and engineering context of OEM-FLIm.
- **Develop and implement the a robust image processing pipeline:** Address motion artefacts and signal degradation by removing uninformative frames, estimating motion between subsequent frames, and applying image alignment to generate temporally stable and spatially consistent FLIm sequences.
- **Enhance registration performance through multi-objective optimisation:** Integrate both intensity and lifetime channels, treating image alignment as a multi-objective problem to better accommodate the dynamic, multimodal nature of FLIm data [25, 26].
- **Investigate machine learning (ML)-based methods:** Explore the impact of the image processing workflow on deep learning (DL) architecture

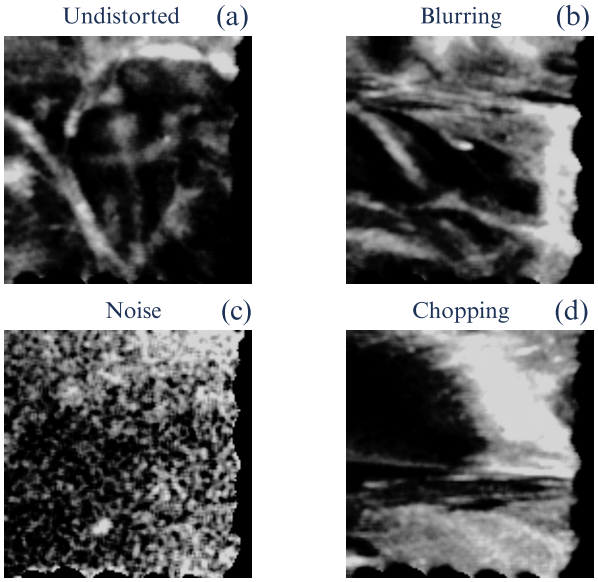


Figure 1.3: Representative examples of common motion artefacts observed in OEM-FLIm intensity images. (a) Undistorted frame showing clear alveolar structures with minimal artefacts. (b) Blurring due to rapid probe movement during acquisition, reducing contrast and structural fidelity. (c) Noise artefact arising from poor fibre contact or insufficient fluorescence signal. (d) Chopping artefact resulting from discontinuous scanning or abrupt fibre displacement. These artefacts pose significant challenges for accurate frame interpretation and downstream image analysis.

for identifying and removing uninformative frames and detecting specific biomarkers such as the neutrophil activation probe (NAP), ultimately supporting automated clinical decision-making [17, 27].

- **Validate the methods on both simulated and real data:** Evaluate the proposed pipelines in terms of quantitative performance, computational efficiency, and clinical relevance, with an emphasis on reproducibility and translational potential.

By pursuing these objectives, this thesis intends to establish a more robust and reliable image processing framework for OEM-FLIm, enhancing its applicability for both research and clinical use in pulmonary health.

1.2 MAIN CONTRIBUTIONS

This thesis delivers several key novel contributions that significantly advance the field of FLIm in pulmonary fibre-based OEM. Departing from traditional

approaches, this research presents targeted methodologies for addressing critical challenges in FLIm data processing, as evidenced by the findings in the subsequent chapters:

- **A Novel temporal reliability and accuracy via correlation enhanced registration (TRACER) Pipeline:** Chapter 3 introduces a new image processing pipeline, termed TRACER, specifically designed to tackle the pervasive issue of motion artefacts in *in vivo* pulmonary OEM-FLIm. The novelty of TRACER lies in its comprehensive approach, which integrates uninformative frame removal, motion characterisation using dense optical flow (OF), and a tracking-based normalised cross correlation (NCC) image registration method. Empirical evaluation demonstrates that TRACER achieves substantial improvements in image alignment—ranging from 20 % to 30 % across various quantitative metrics, compared to conventional registration techniques. Furthermore, the tracking-based NCC image registration approach exhibits superior computational efficiency, operating at an order of magnitude faster than the next best method, making it well-suited for real-time applications. This development provides a robust foundation for accurate temporal analysis of dynamic alveolar processes.

- **A Multi-Objective Image Registration Framework Leveraging Intensity and Lifetime Information:** Chapter 4 presents a significant advancement in FLIm image registration by formulating it as a multi-objective optimisation problem. This novel approach moves beyond traditional intensity-based registration by simultaneously considering both intensity and fluorescence lifetime data to achieve more accurate alignment in the complex spatio-temporal dynamics inherent in FLIm sequences. The introduction of dynamic multi-objective search (DYNMOS) as the image registration method enables an efficient exploration of the trade-offs between registration accuracy based on intensity and lifetime information. Notably, this multi-objective strategy achieves comparable registration performance to state-of-the-art intensity-based methods while offering significantly reduced computational overhead (up to 2.46× faster). This contribution highlights the synergistic benefits of leveraging the multi-dimensional nature of FLIm data for enhanced image registration.

- **The Successful Application of Machine Learning for Automated FLIm Data Enhancement and Analysis:** Chapter 5 explores the impact of TRACER and the potential of ML techniques to address key limitations in OEM-FLIm data processing. A key novelty is the development of a transfer learning (TL)-based approach for the automated identification and

removal of uninformative frames from FLIm sequences. This method demonstrates high accuracy in distinguishing informative from uninformative data, thereby improving overall data quality and the efficiency of downstream processing. Furthermore, this chapter investigates the effects of different elements of the image processing workflow on a U-Net-based deep learning framework for the automatic detection of NAP within FLIm images. This offers important insights towards automated disease characterisation at the alveolar level, showcasing the transformative potential of robust image processing pipelines and ML in extracting clinically relevant information from complex FLIm datasets.

Collectively, these main contributions provide a suite of novel and rigorously validated image processing and ML methodologies that address critical challenges in pulmonary OEM-FLIm. The advancements presented in this thesis enhance the temporal reliability and spatial accuracy of FLIm data and pave the way for more automated and informative analyses, ultimately bolstering the potential of OEM-FLIm as a powerful tool for advancing pulmonary health research and clinical applications.

1.3 PROJECT OUTPUTS AND RECOGNITIONS

The following contributions and awards are as a result of the research conducted for this thesis, which has not been submitted for any other degree or professional qualification.

Journal

- Gouzou, D. Taimori, A. Haloubi, T. Finlayson, N. Wang, Q. Hopgood, J. Vallejo, M. (2024) "Applications of Machine Learning in Time-domain Fluorescence Lifetime Imaging: a Review." In IOP Methods and Applications in Fluorescence.
- Haloubi, T., Thomas, S., Hines, C, Dhaliwal, K., Hopgood, J. (2025). "Motion Compensation in Pulmonary Fluorescence Lifetime Imaging: An Image Processing Pipeline for Artefact Reduction and Clinical Precision." In IEEE Open Journal of Engineering in Medicine and Biology (OJEMB).

Conferences

- Haloubi, T., Thomas, S., Hines, C, Dhaliwal, K., Hopgood, J. (2024). "Navigating Noise and Texture: Motion Compensation Methodology for Fluorescence Lifetime Imaging in Pulmonary Research." In: Proceedings of the 46th IEEE Engineering in Medicine and Biology Conference (EMBC), Orlando, USA

Awards

- **Best Oral Presentation.** Awarded to the best presentation given by a PhD student at the 6th Postgraduate Institute for Measurement Science (PGI) conference 2021.
- **Best poster.** Awarded to the best poster at the University of Edinburgh's School of Engineering Conference 2023.
- **Peter Grant award for best student-led conference paper.** Awarded to the best conference paper from the Institute for Imaging, Data, and Communications (IDCOM) 2024.

1.4 OVERVIEW OF THE THESIS

The remainder of this thesis is structured as follows:

CHAPTER 2 Provides a comprehensive Background and Related Work, establishing the necessary context for understanding the challenges and the novelty of the proposed solutions. This chapter meticulously outlines the principles of pulmonary imaging, fluorescence lifetime microscopy, and the specific issues associated with their application in fibre-optic endomicroscopy. This detailed foundation is crucial as it highlights the current state of the art and pinpoints the specific gaps in knowledge and technological limitations that this thesis directly addresses, ensuring that the subsequent chapters are understood within the broader scientific landscape and the clinical need for improved alveolar-level imaging.

CHAPTER 3 Introduces TRACER, a novel image processing pipeline designed to mitigate motion artefacts in *in vivo* pulmonary OEM-FLIm data. TRACER integrates uninformative frame removal, motion characterisation using dense OF, and tracking-based NCC for enhanced image registration. The development and validation of TRACER are a significant contribution, as it provides a robust method for achieving temporally consistent and accurately aligned FLIm image sequences. This is fundamental for reliable quantitative analysis and the extraction of meaningful biological information from dynamic alveolar processes.

CHAPTER 4 Focuses on the multi-objective approach to image registration, extending beyond traditional intensity-based methods by integrating both intensity and fluorescence lifetime information. By formulating registration as a multi-objective optimisation problem and introducing the DYNMOS approach, this chapter demonstrates alignment improvements and reduced computational

overhead. This plays an important role in robustly processing FLIM data as it leverages the inherent multi-dimensionality of FLIm data, unlocking a richer source of information for accurate image registration, particularly in scenarios with complex motion and multi-modal data. The achievement of comparable performance to state-of-the-art methods with significantly reduced computational cost enhances the feasibility of advanced FLIm analysis in resource-constrained settings.

CHAPTER 5 Explores the importance of TRACER in ML applications for OEM-FLIm data processing, investigating DL techniques for automated uninformative frame detection and NAP identification. The successful application of TL for frame selection and a U-Net-based framework for NAP detection demonstrates the potential for automated data enhancement and disease characterisation. This exploration is crucial for advancing the clinical utility of OEM-FLIm by enabling more efficient data processing and providing tools for automated, objective analysis of disease markers, ultimately streamlining the pathway from image acquisition to clinically relevant insights.

CHAPTER 6 Concludes the thesis by summarising the main contributions, discussing the broader implications of the research, and outlining potential avenues for future work. This concluding chapter serves as a critical synthesis of the findings, reinforcing the impact of the developed methodologies and charting a course for future research to further enhance the capabilities and clinical translation of OEM-FLIm for advancing pulmonary health. By reflecting on the achieved advancements and identifying remaining challenges, this chapter underscores the significance of the thesis within the ongoing efforts to improve the diagnosis and treatment of pulmonary diseases.

2

BACKGROUND AND RELATED WORK

2.1 INTRODUCTION

This chapter provides the wider clinical and engineering context for the thesis and situates OEM-FLIm within the evolving field of pulmonary imaging. It begins with an overview of lung anatomy and physiology, highlighting how advanced imaging modalities are crucial for studying healthy respiratory function and the pathological processes that disrupt it. The discussion then underscores the importance of FLIm in capturing subtle metabolic and structural details at the alveolar level—an insight that conventional imaging techniques often struggle to achieve due to limitations in resolution or concerns over ionising radiation. Finally, the chapter introduces a major practical challenge in pulmonary imaging: motion artefacts driven by continuous respiratory mechanics. It outlines current strategies for motion compensation and clarifies why robust solutions to this issue are vital for optimising FLIm data processing for enhancing downstream image analysis.

The lungs serve as the principal organs of respiration, delivering oxygen to the bloodstream while removing carbon dioxide, a by-product of cellular metabolism [2, 28, 29]. Despite being relatively compact within the thoracic cavity, they contain over 400 million alveoli—microscopic air sacs that maximise the internal surface area for gas exchange [2]. These alveoli also host immune cells known as alveolar macrophages, which patrol the respiratory surfaces and neutralise pathogens or foreign particles [3, 30]. Beyond gas exchange, the lungs maintain a critical defence against inhaled threats, supporting systemic oxygenation and immune homeostasis [2, 3, 30]. Fig. 2.1 illustrates the hierarchical structure of the lungs, zooming in on the alveolar space and depicting the exchange of oxygen and carbon dioxide between the alveoli and surrounding blood vessels.

Clinically, pulmonary imaging has typically relied on modalities like CXR, CT, MRI, and PET, each of which can provide valuable anatomical and functional information [4, 5, 31]. These techniques have been vital for diagnosing, staging, and monitoring respiratory diseases and guiding therapeutic interventions. However, despite continued advances—such as dose reductions in CT [32], improved functional sequences in MRI [33], and novel hybrid methods [34]—many existing modalities remain constrained by radiation exposure limits, incomplete spatial resolution, or reduced specificity at the alveolar scale [7, 23]. Optical-based meth-

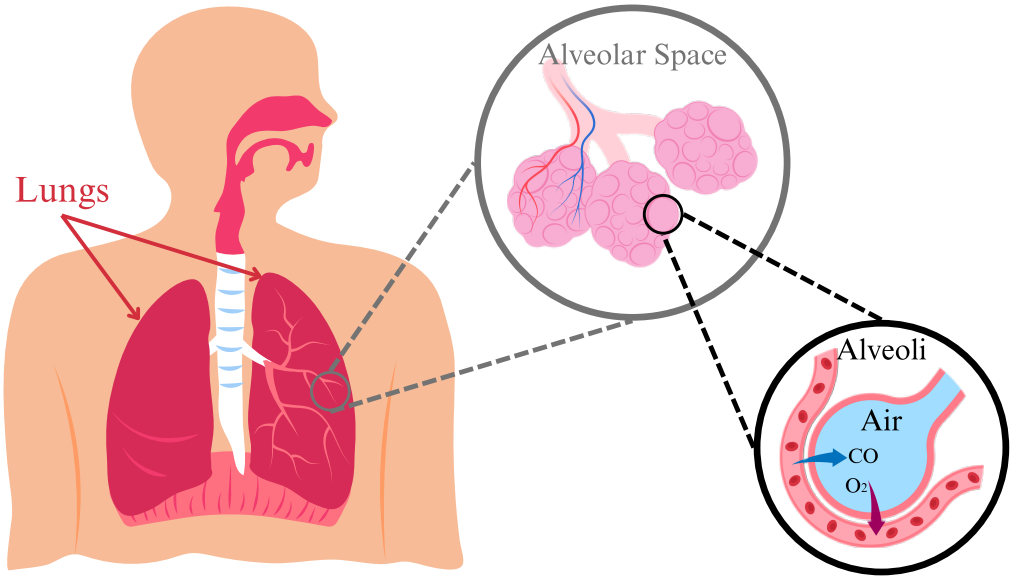


Figure 2.1: Illustration of the lung's hierarchical organisation, from the macroscopic lung (left) to a zoomed-in view of the alveolar space (centre), and finally a close-up inset (right) showing alveoli surrounded by a blood vessel. Arrows indicate the transfer of oxygen (O_2) from alveolar air into the bloodstream and the removal of carbon dioxide (CO_2) from the blood back into the alveolar space for expiration.

ods, including emerging approaches like FLIM, are now being actively considered to address these shortcomings [35, 18]. In particular, OEM-FLIm offers metabolic and molecular insights not readily available from purely anatomical scans [8, 1]. By measuring fluorescence decay times, FLIm can discriminate between various biochemical states of tissues, potentially revealing disease processes at an earlier stage and with greater cellular detail [16, 23, 15]. Moreover, the modality could specifically characterise immune responses by probing the metabolic states of neutrophils or other inflammatory cells [36, 17].

Yet, the very dynamics that make the lung an efficient respiratory organ also complicate imaging [1, 21]. Continuous motion from breathing can distort or blur optical measurements, introducing artefacts that obscure critical structural or functional information. Successful FLIM acquisition, therefore, requires strategies to mitigate the effects of respiratory motion [7]. Whether through gating, synchronisation with the respiratory cycle, or sophisticated computational corrections, motion compensation is essential to preserve image fidelity and enable precise data interpretation [37, 21, 1, 38].

2.1.1 *Chapter Overview*

In the remainder of this chapter, the optical system "KronoScan" and biological setup employed in this work are described, followed by an examination of the pulmonary motion phenomena and motion compensation approaches prevalent in the literature. The chapter concludes by outlining how these insights pave the way for the subsequent development of a novel FLIm-based image processing pipeline. This pipeline also considers the dual challenges of intensity and lifetime data analysis from FLIm and the integration of ML techniques to enhance the accuracy and automation of clinical or research workflows. In doing so, it lays the foundation for the thesis's central objective: to harness the promising potential of OEM-FLIm for comprehensive pulmonary imaging while minimising the confounding effects of respiratory motion.

2.2 MEDICAL AND ENGINEERING BACKGROUND

Building on the introductory overview of pulmonary physiology and the motivations for exploring optical imaging in the lung, this section delves into the medical and engineering principles that underpin pulmonary medical imaging research, with emphasis on the application of OEM-FLIm. The lungs, with their intricate alveolar architecture and dynamic breathing mechanics, present unique challenges to image acquisition and interpretation.

To address these demands, a review of established and emerging pulmonary imaging modalities is presented, highlighting their respective advantages and limitations in characterising alveolar structures and identifying pathological changes. Next, the fundamental concepts of fluorescence microscopy imaging are considered, tracing its evolution from basic intensity-based techniques to modern, high-resolution approaches capable of revealing subtle biochemical and molecular details. Finally, FLIm characterisation is introduced, demonstrating how lifetime measurements add a new dimension of functional and metabolic insight to conventional fluorescent images. This integrated background sets the stage for understanding why FLIM, coupled with careful engineering and computational strategies, holds promise for advancing both diagnostic and research applications in real-time pulmonary medicine.

2.2.1 *Pulmonary Imaging*

Clinically, widely used imaging modalities guide diagnosis and treatment of pulmonary disorders [2, 39]. The most commonly employed imaging modalities for pulmonary assessment include:

1. **CXR:** Commonly employed for initial lung assessments due to its availability and low cost. It can detect infections, pneumothorax, fluid accumulation, and chronic disease progression [4, 31].

2. **CT:** Offers cross-sectional imaging of lung parenchyma, allowing for detection of interstitial lung disease, emphysema, and tumour nodules [40]. Advances in ultra-low-dose CT protocols have improved patient safety while maintaining diagnostic performance [41].

3. **MRI:** Though not as commonly used as CT for lung imaging, MRI has gained traction due to its capability to visualise soft tissues and assess lung ventilation and perfusion without ionising radiation [42].

4. **PET:** It is usually used in combination with CT (PET-CT), is essential for oncological imaging, providing metabolic information that aids in tumour staging and treatment response evaluation [6].

5. **Bronchoscopy:** Primarily a diagnostic and therapeutic procedure that enables direct visualisation of the airways, sample collection, and removal of obstructions [9, 43].

Examples of various pulmonary imaging modalities—along with representative clinical findings and artefacts—are illustrated in Fig. 2.2. Despite the utility of these established methods, research continues to push for more sophisticated techniques to enhance the detection and characterisation of complex pulmonary diseases. They are considered conventional and invaluable for clinical diagnostics. However, they suffer from inherent limitations:

- Spatial and Temporal Resolution: CXR and CT lack the temporal resolution required to capture dynamic biological processes at the cellular level [7].

- Lack of Functional Information: Conventional techniques primarily provide structural information, with limited insight into tissue metabolism [16].

- Radiation Exposure: X-ray-based modalities expose patients to ionising radiation, raising concerns for repeated imaging [48].

- Motion Artefacts: Respiratory and cardiac motion can introduce distortions in CT and MRI images, reducing diagnostic accuracy [37, 45, 38].

Given these challenges with conventional methods, optical approaches represent promising alternatives. Specifically, advanced imaging techniques such as OEM have emerged, offering real-time, microscopic imaging of lung tissue. When integrated with OEM, FLIm further enhances contrast by mapping fluorescence decay times, thereby providing insights into the tissue's biochemical properties and

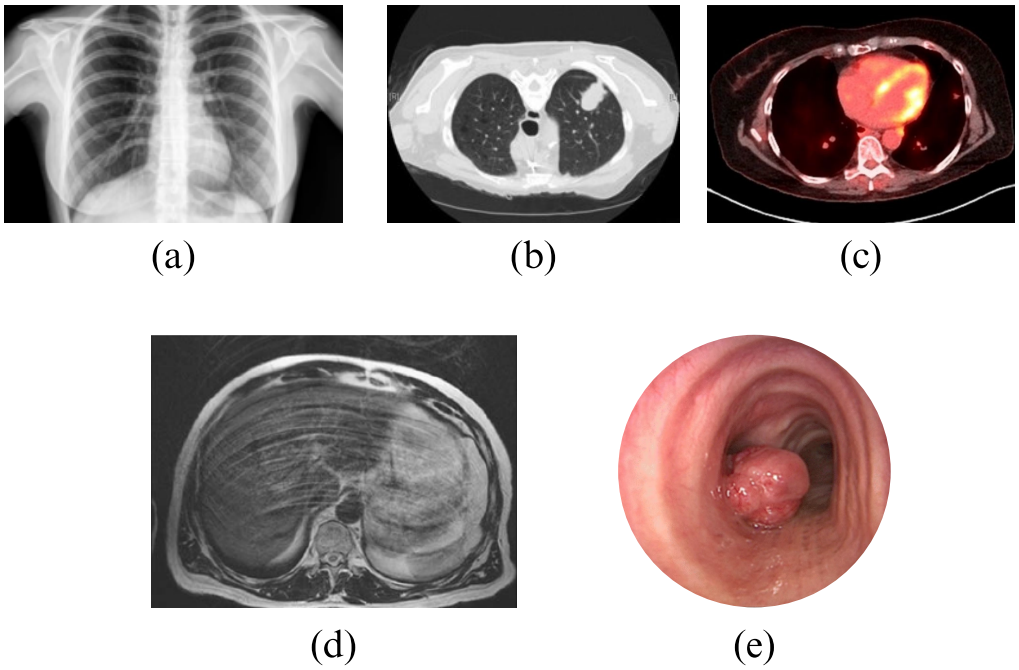


Figure 2.2: Examples of common pulmonary imaging modalities showcasing different clinical findings: **(a)** CXR image of healthy lungs [44], **(b)** CT scan revealing a tumour (visible as a white region) in the upper right portion of the lung [45], **(c)** combined PET/CT with yellow/orange areas indicative of sarcoidosis [46], **(d)** MRI example exhibiting motion artefacts due to respiration [47], and **(e)** bronchoscopy image capturing a tumour obstructing the left mainstream bronchus [9].

allowing for clear differentiation between healthy and pathological states [14, 8, 49]. Before delving into the operational details of OEM-FLIm, it is instructive to review the fundamental principles of fluorescence microscopy imaging.

2.2.2 *Fluorescence Microscopy Imaging*

The field of pulmonary imaging has evolved significantly over the past decades [5, 7], driven by the need for improved diagnostic precision and a deeper understanding of lung physiology and pathology. Given the intricate nature of the alveolar space, advanced imaging modalities play a pivotal role in capturing the structural and functional characteristics of lung tissue at microscopic levels.

Fluorescence, a light emission process, occurs due to photon energy absorption from an external source, inducing energy transitions from the ground to excited states. This phenomenon manifests in emission and excitation spectra distinguished by parameters such as fluorescence intensity and lifetime—the latter corresponding to the average interval between excitation and the return to the

ground state [14]. Fluorescence microscopy is widely regarded as an indispensable tool in biological research, enabling both structural and functional molecular imaging of biological specimens [50]. Although recent low-cost, high-sensitivity OEM systems [51] have further refined fluorescence imaging capabilities—particularly in visualising multiple colour signatures—these approaches typically rely on intensity-based signals alone. Such methods can yield rich morphological detail but may not fully capture the biochemical or metabolic nuances within tissues.

In contrast, FLIM not only records intensity but also characterises the decay profiles of fluorophores in a sample [52, 14]. By extending conventional fluorescence microscopy to incorporate lifetime measurements, researchers gain deeper insight into the local molecular environment and tissue metabolism. Understanding the alveolar space is crucial in pulmonary research [2, 7], and FLIm thus emerges as a powerful tool for exploring alveolar-level processes. Indeed, its capacity to map metabolic states and monitor disease progression at the microscopic scale affords a more comprehensive understanding of how pathologies alter lung function and how potential therapeutic interventions may be optimised [53, 54, 55].

With the relevance of fluorescence imaging in pulmonary research established, the discussion now shifts to two complementary approaches for detecting and analysing fluorescence signals: intensity imaging and lifetime imaging. The former focuses on the amplitude of emitted light, while the latter captures the temporal decay properties of fluorescence. Each approach presents distinct advantages and challenges, as outlined below.

2.2.2.1 *Intensity Imaging*

Fluorescence intensity imaging (FII) captures the amplitude of emitted fluorescence, providing a relatively straightforward means to visualise biological structures in situ. Because it measures the brightness of fluorescence signals at a given wavelength, FII often requires less complex hardware and fewer computational resources compared to lifetime-based methods, making it widely accessible in both research and clinical settings [50].

Despite these advantages, FII is inherently sensitive to factors such as fluorophore concentration, photobleaching, and tissue scattering. Varying levels of endogenous or exogenous fluorescent molecules can obscure the true biological signal, complicating quantitative comparisons across samples or over time [7, 23, 35]. Nonetheless, the technique is routinely employed to examine cell morphology, track molecular uptake or expression, and detect structural landmarks within tissues. Fig 2.3 illustrates representative examples of intensity images obtained from alveolar tissue, highlighting distinct features such as alveolar walls, capillaries, and different cells of interest. In particular, the distribution and concentration of endogenous fluorophores—typically absorbing around 485–520 nm—can delineate

cellular membranes and interstitial matrices, thereby providing valuable insight into lung microanatomy in both healthy and pathological states [30].

2.2.2.2 *Lifetime Imaging*

While the acronyms FLIM and FLIm are often used interchangeably in general discourse, a technical distinction is drawn in this thesis between Fluorescence Lifetime Microscopy (FLIM) and Fluorescence Lifetime Imaging (FLIm). FLIM strictly refers to the microscopic technique, whilst FLIm refers to the broader imaging modality applied. This distinction aligns with recent conventions in interventional biophotonics to differentiate the clinical data modality from traditional microscopy [10].

FLIm [14] is an emerging modality that measures the time a fluorophore remains in its excited state before returning to the ground state. It is less affected by variations in fluorophore concentration and photobleaching, providing quantitative information on molecular interactions and local environments. FLIm can also be used in Förster resonance energy transfer (FRET) studies to investigate protein–protein interactions at the nanometre scale [19, 16, 56, 57]. However, FLIm often requires more complex instrumentation and data analysis than intensity-based methods [58, 59].

Moreover, FLIM enhances traditional fluorescence microscopy by measuring the temporal decay of fluorescence emission, known as the fluorescence lifetime, rather than relying solely on the intensity of emitted light. Fluorescence lifetime is exquisitely sensitive to local environmental factors such as pH, ion concentration, refractive index, and the binding state of fluorescent molecules [60, 61]. Consequently, FLIm can offer unique insights into biochemical processes or tissue composition that may not be discernible from intensity images alone, as demonstrated in Fig 2.3.

In FLIM, a short pulse of light excites fluorophores in the sample. Instead of only recording the brightness of the emitted light, as in FII, FLIm captures how rapidly this light decays to baseline. The measured lifetime (τ) reflects the probabilistic time a fluorophore remains in its excited state, typically in the range of nanoseconds. There are two main methods to record these lifetimes:

- **Time-domain FLIm:** A pulsed excitation source is used, and the fluorescence decay is monitored in real-time. While it can approach the ideal shot noise limit, the approach often involves large datasets and longer integration times to build sufficient photon statistics [62]. Consequently, time-domain FLIm can be computationally intensive and may limit real-time operation.
- **Frequency-domain FLIm:** A sinusoidally modulated excitation light and phase-resolved detection scheme can estimate the fluorescence lifetime

based on amplitude and phase shifts. This approach, too, can reach near shot noise limited performance, but practical implementations must account for trade-offs in acquisition speed and hardware complexity [63].

Both methods hinge on detecting enough photons to reliably quantify the decay profile—any reduction in the "photon economy" can diminish the SNR and thus the precision of lifetime estimates [63, 62], nonetheless, with careful instrumentation design and calibration, FLIm can match or surpass the SNR of intensity-based imaging in some scenarios [64, 10, 65].

The ability of FLIm to reveal local molecular environments makes it widely appealing in biomedical research and diagnostics. FLIm has been used for:

- **Ion Sensing and Metabolic Studies:** pH, calcium, and other ion concentrations can be inferred via lifetime shifts in suitably engineered fluorescent probes [61].
- **Tissue Characterisation:** Auto-fluorescence lifetimes can distinguish healthy and diseased tissues, aiding in the real-time assessment of tumour margins or fibrotic lesions [55].
- **Multidimensional Investigations:** Emerging trends incorporate multispectral or multiparametric FLIm to disentangle overlapping fluorophore signatures and track multiple biochemical processes simultaneously [66].

Moreover, recent clinical applications leverage auto-fluorescence FLIm for endoscopic procedures, where the temporal contrast can guide surgeons to identify pathological changes *in vivo* [18, 35]. Such real-time capabilities hold promise for optical biopsies, reducing the need for invasive tissue sampling [7, 55].

At the alveolar level, FLIm enables the characterisation of lung tissue based on its intrinsic fluorescence properties, offering insights into:

- Tissue Metabolism: FLIm can distinguish between different metabolic states based on fluorescence lifetime signatures of nicotinamide adenine dinucleotide phosphate (NADPH) and flavin adenine dinucleotide (FAD) [56].
- Collagen and Elastin Distribution: The extracellular matrix components of lung tissue exhibit distinct fluorescence lifetimes, aiding in the differentiation of fibrotic and healthy regions [8, 7].
- Inflammatory and Neoplastic Changes: Changes in fluorescence lifetime can indicate inflammatory processes or early tumour development, providing potential applications in early disease detection [17, 16].

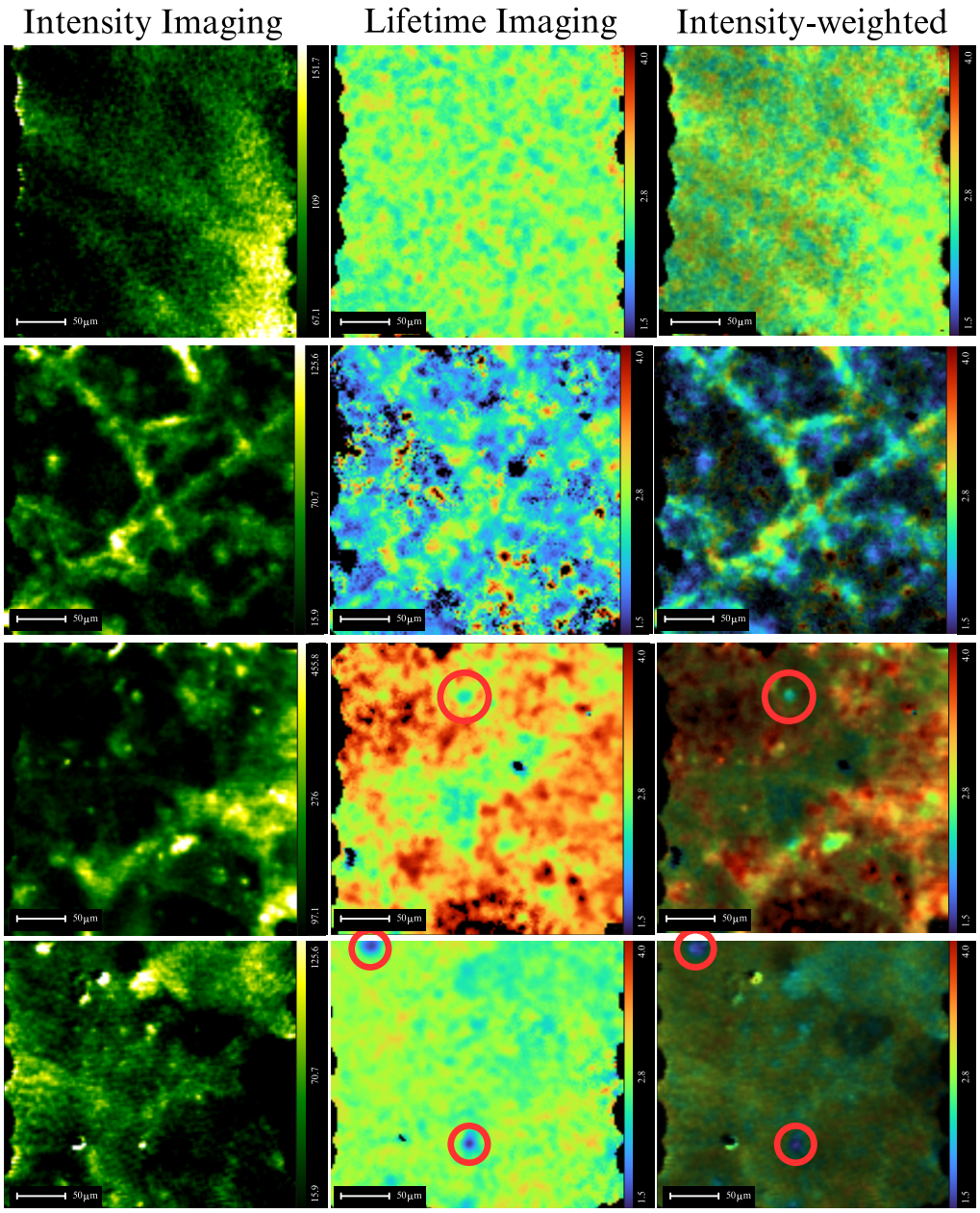


Figure 2.3: Representative pulmonary fluorescence images acquired using FLIM. Each row shows three type of images: **(left)** fluorescence intensity (dark green to light yellow, low to high photon counts), **(centre)** fluorescence lifetime (jet scale, 1.5 ns in blue to 4 ns in red), and **(right)** intensity-weighted lifetime (intensity-weighted overlay). Intensity highlights alveolar structure, whereas lifetime provides biochemical contrast. Regions circled in red illustrate areas of reduced lifetime that are not evident from intensity alone and are consistent with increased metabolic or NAP-related activity [17].

Despite its advantages, FLIm is more computationally demanding than conventional fluorescence imaging [65, 10]. Building sufficient photon statistics to estimate lifetimes accurately can necessitate longer acquisition times [63], especially in the time domain FLIm [62]. This requirement often conflicts with the desire for real-time imaging, particularly in clinical settings where rapid decisions may be needed [7]. Additionally:

- **Data Processing and Analysis:** Fluorescence decay data typically require complex, model-based inversions or multi-exponential fitting strategies. High-frame-rate acquisitions generate large volumes of data, placing demands on both hardware and software [66, 10].
- **Photon Economy and SNR:** Achieving near shot noise limited performance calls for meticulous control over excitation intensity, detector sensitivity, and timing electronics [63, 10].
- **Real-Time Capability:** Many algorithms in use today were not conceived with real-time performance in mind, limiting their clinical applicability and necessitating further research into fast, robust lifetime extraction methods [7, 10].

Nonetheless, FLIM stands at the forefront of functional microscopy and endoscopic imaging, offering a wealth of biochemical and biophysical information that can complement traditional intensity-based methods [55]. By overcoming current limitations—particularly in data processing speed and motion compensation, FLIM is poised to advance *in vivo*, high-resolution imaging and enable new discoveries in research and clinical medicine.

In summary, FLIm and FII each offer unique perspectives on lung tissue at the microscopic level. Advances in these techniques have the potential to improve our understanding of pulmonary structure and function, driving the development of more effective diagnostic and therapeutic approaches.

2.2.3 *Fluorescence Lifetime Microscopy Image Representations*

Mathematically, in time-domain FLIM, each pixel (x, y) contains a time-resolved histogram of photon arrival times following pulsed excitation [67]. These histograms are discretised into B time bins of equal width Δt, and the number of photons detected in the b-th bin is denoted $N_b(x, y)$.

As seen in Fig. 2.3, there are three key image representations that can be derived from this temporal data, each encoding different information:

2.2.3.1 *Fluorescence Intensity Image Representation*

An intensity image generally refers to any modality in which the image is directly formed by integrating photon counts recorded by the imaging sensors. This includes modalities such as standard microscopy, X-ray imaging, and CT, where the image intensity reflects the total photons detected at each pixel or detector element.

In fluorescence imaging specifically, the intensity image simply reflects the total fluorescence counts per pixel, obtained by summing photon counts over all time bins. Thus, the fluorescence intensity image $\mathbf{I}(x, y)$ is calculated as:

$$\mathbf{I}(x, y) = \sum_{b=1}^{B} N_b(x, y), \tag{2.1}$$

This corresponds to integrating the decay signal over the entire capture window and is akin to a conventional fluorescence image, dominated by structural and anatomical features. Under a single-exponential decay model, we can write:

$$I(t) = \mathbf{I}_0(x, y)\, e^{-t/\tau(x,y)}, \tag{2.2}$$

In practice, the measured signal deviates from this idealised form due to the presence of background photon counts and the stochastic nature of photon detection. Background contributions may arise from detector dark counts, ambient light, or autofluorescence, and add an approximately constant offset to the measured decay. In addition, photon arrival follows Poisson statistics, such that the variance of the detected counts in each time bin is proportional to the mean count level. These effects do not alter the expected form of the decay model but impact the noise characteristics and precision of any derived quantities [22, 13].

Defining bin edges t_b to t_{b+1}, the theoretical bin counts are:

$$\begin{aligned} N_b(x, y) &\propto \int_{t_b}^{t_{b+1}} \mathbf{I}_0(x, y)\, e^{-t/\tau(x,y)}\, dt \\ &= \mathbf{I}_0(x, y)\, \tau(x, y) \left[e^{-t_b/\tau(x,y)} - e^{-t_{b+1}/\tau(x,y)} \right], \end{aligned} \tag{2.3}$$

Summing across bins gives:

$$\mathbf{I}(x, y) \propto \mathbf{I}_0(x, y)\, \tau(x, y) \left[1 - e^{-T/\tau(x,y)} \right], \tag{2.4}$$

where $T = B\Delta t$ is the total integration time. If the window spans multiple lifetimes (i.e., $e^{-T/\tau(x,y)} \approx 0$), this simplifies to:

$$\mathbf{I}(x, y) \approx \mathbf{I}_0(x, y)\, \tau(x, y), \tag{2.5}$$

showing that high intensity may stem from either a large initial amplitude or a longer lifetime. However, in practice $\mathbf{I}(x, y)$ is just a raw count, and lifetime contrast is usually extracted separately, as described next [68, 67].

2.2.3.2 *Fluorescence Lifetime Image Representation*

The most rigorous method involves fitting the measured decay to an exponential model (or multi-exponential for complex decays). Under the single-exponential assumption, one performs a least-squares fit or maximum likelihood estimation of the model:

$$N_b \approx A \left(e^{-t_b/\tau} - e^{-t_{b+1}/\tau} \right), \tag{2.6}$$

with amplitude $A = I_0\tau$. This fitting is typically conducted pixel-by-pixel, although it can be computationally intensive for large images and high temporal resolution. Specialised software (e.g., `FLIMfit` [69]) commonly implements these per-pixel or global fitting algorithms [68].

In this work, we utilise the rapid lifetime determination (RLD) method [70], suitable for systems with few time bins and rapid acquisition. Unlike full time-correlated single photon counting (TCSPC) approaches that retain complete photon arrival timing information, RLD relies on gated integration and therefore represents a hardware-efficient, non-iterative approximation to lifetime estimation. Under the assumption of negligible background counts, the precision of lifetime estimates obtained using the RLD approach is ultimately limited by photon statistics, with a best-attainable standard deviation given by $\sigma_\tau \approx \tau/\sqrt{N}$, where N denotes the average number of detected signal photons [71]. This highlights that, when carefully designed, gated methods can approach the fundamental precision limits dictated by photon counting, while enabling significantly higher frame rates and compatibility with solid-state sensor architectures.

The lifetime image, $\mathbf{L}(x, y)$, mapping the decay constant at each pixel, is analytically derived from two gated integrals (early and late), each integrating photon counts over equal-width intervals:

$$\mathbf{L}(x, y) = \frac{T}{\ln\left(\frac{W_1(x,y)}{W_2(x,y)}\right)}, \tag{2.7}$$

where $W_1(x, y)$ and $W_2(x, y)$ represent the total photon counts in the early and late gates, respectively, and T is the duration of each gate. This method circumvents iterative fitting and is robust for high-speed imaging, although precision may diminish at low photon counts or when decay is multi-exponential [72, 68].

2.2.3.3 *Intensity-weighted Lifetime Image Representation*

Intensity-weighted lifetime images combine fluorescence intensity and lifetime information, providing enhanced interpretability of FLIM data [73]. The intensity-weighted lifetime image $\mathbf{I}_\tau(x, y)$ is computed as the scalar product of the total intensity and the calculated lifetime, using Eq. 2.7. This can be written as:

$$\mathbf{I}_\tau(x, y) = I_{total}(x, y) \cdot \tau(x, y), \tag{2.8}$$

where $I_{total}(x, y)$ is the sum of $W_1(x, y)$ and $W_2(x, y)$.

This representation serves two distinct purposes in the pipeline. Firstly, the product $\mathbf{I}_\tau$ produces a scalar field where the lifetime signal is effectively confidence-weighted by intensity. This suppresses noise in low-photon regions while emphasising structural features, providing a robust continuous variable for downstream image processing tasks such as registration. Secondly, for clinical display and qualitative assessment, this data is generated as a standard RGB image. In this visualisation, the lifetime value τ determines the colour (mapped via a colour map), while the intensity I_{total} modulates the brightness. This allows simultaneous observation of tissue structure and functional lifetime contrast.

2.2.3.4 *Statistical Representation of Fluorescence Imaging Modalities*

Understanding the statistical foundations underlying each fluorescence imaging modality is essential for accurate interpretation and robust data processing. As seen in Fig. 2.3, each representation has distinct clinical and analytical utility.

Statistically, the intensity image $\mathbf{I}(x, y)$ at each pixel is described by the integrated number of photons emitted by fluorophores within the pixel area over the integration period. The intensity image offers robust structural features, useful for navigation or registration [74, 68]. In contrast, the lifetime image $\mathbf{L}(x, y)$ provides a statistical representation of the expectation (mean value) of the fluorescence decay time distribution, explicitly weighted by the photon count detected at each pixel. Unlike intensity imaging, the lifetime image offers biochemical specificity, revealing metabolic or environmental contrasts [74, 19, 14].

The intensity-weighted lifetime image is derived by explicitly weighting the fluorescence lifetime values at each pixel by their corresponding intensities. This combined statistical metric, denoted as $\mathbf{I}_\tau(x, y)$, thus integrates both photon-count distributions and decay kinetics, effectively enhancing the visualisation of areas characterised by both significant intensity and distinct lifetime values [75]. This combined approach balances structural clarity from intensity imaging with the molecular specificity inherent in lifetime imaging, presenting a richer and more informative spatial representation, as exemplified in Fig. 2.3.

Although fluorescence lifetime is fundamentally a temporal decay phenomenon, many practical FLIM systems, including the one used in this research, routinely transform the raw photon histograms into 2D lifetime "maps" or intensity-weighted images. This transformation simplifies spatial alignment and enables the use of mature image-processing algorithms. The choice is thus a deliberate engineering decision, motivated by real-time constraints and limited access to data with higher temporal resolution in many clinical imaging settings.

2.2.4 *Fluorescence Lifetime Microscopy Image Characterisation*

FLIM inherently differs from conventional medical imaging in both the qualitative nature of its signal and the quantitative metrics it provides [55]. In traditional intensity-based imaging modalities, image contrast is primarily driven by the measured brightness of the signal at each pixel, which depends on factors such as fluorophore concentration, illumination intensity, and detector sensitivity [50]. Consequently, intensity images can be straightforward to acquire and interpret; however, they are prone to confounding effects such as uneven sample illumination, fluorophore bleaching, and variations in tissue optical properties.

By contrast, fluorescence lifetime imaging focuses on the temporal decay of fluorescence following excitation, providing a quantitative metric—namely, the fluorescence lifetime—that is relatively independent of the absolute fluorophore concentration. This temporal dimension can reveal subtle changes in the local molecular environment, including variations in pH, oxygen concentration, or protein binding states [19]. As a result, lifetime imaging carries additional functional and biochemical insights that are often unobservable in purely intensity-based measurements [66]. However, FLIm data tend to be acquired at lower spatial resolution and SNR compared with many standard anatomical imaging modalities, complicating both data acquisition and subsequent analysis [61].

These qualitative and quantitative differences have a considerable impact on the design of image processing pipelines for FLIm. Novel segmentation strategies [76], advanced de-noising algorithms [77], and robust feature extraction tools must account for the fact that lifetime maps are not merely another intensity channel, but rather reflect unique molecular decay kinetics [78]. Additionally, the relatively lower spatial resolution of FLIm requires careful attention to sampling strategies, detector performance, and image reconstruction algorithms to maximise information content and maintain clinical utility. Key Properties Defining Imaging Quality and Information Content:

- **Intensity Signal Dependence:** Traditional intensity-based fluorescence relies on excitation power, fluorophore concentration, and optical path

variations, making it susceptible to non-uniform illumination and tissue heterogeneities.

- **Lifetime Independence from Concentration:** Fluorescence lifetime offers a concentration-independent contrast mechanism, sensitive to the biochemical and biophysical environment, thus enabling the identification of subtle molecular changes.
- **Resolution and Signal-to-Noise Ratio:** FLIm often has lower spatial resolution than conventional high-resolution imaging (*e.g.,* MRI or CT) due to acquisition constraints, yet provides additional functional information that can guide diagnostic and therapeutic decisions.
- **Sensitivity to Environmental Factors:** Lifetime measurements can be heavily influenced by pH, temperature, and other local factors, demanding careful calibration and corrections in image processing to extract meaningful biological or clinical insights.
- **Temporal Decay Modelling:** The core of lifetime-based methods lies in accurately capturing and fitting decay curves, requiring robust computational models and appropriate noise-handling strategies within the image processing pipeline.

Understanding these qualitative and quantitative distinctions between intensity and lifetime imaging is therefore essential. It ensures that image acquisition protocols and subsequent processing methodologies are optimally tailored to leverage the unique information content FLIm can provide. This thesis addresses these challenges by proposing a novel FLIm-oriented image processing pipeline, paving the way for more insightful and reliable analyses of fluorescence lifetime data in clinical and research settings.

2.2.4.1 *Image Feature Metrics*

Texture analysis plays a pivotal role in comparing the qualitative and quantitative characteristics of different imaging modalities, including fluorescence lifetime imaging. One widely adopted method for quantifying texture features is the grey-level co-occurrence matrix (GLCM) [1], which captures the spatial relationship between pairs of grey-level intensities in an image. Formally, for a given offset $\mathbf{d}$ (often specified by a distance and orientation) and grey-levels i and j, the GLCM can be expressed as:

$$\text{GLCM}(i,j;\mathbf{d}) = \sum_{x=1}^{N_x}\sum_{y=1}^{N_y}\begin{cases}1 & \text{if } \mathbf{I}(x,y)=i \text{ and } \mathbf{I}(x+d_x,y+d_y)=j,\\ 0 & \text{otherwise,}\end{cases} \tag{2.9}$$

where (x, y) spans the valid spatial coordinates of the image $\mathbf{I}$, and (d_x, d_y) represents the chosen offset.

Once the GLCM is computed, it is often converted into a normalised probability matrix:

$$\Pr(i,j) = \frac{\text{GLCM}(i,j;\mathbf{d})}{\sum_{i,j} \text{GLCM}(i,j;\mathbf{d})}, \tag{2.10}$$

such that $\sum_{i,j} \Pr(i,j) = 1$. From $\Pr(i,j)$, several key texture features can be extracted to characterise an image quantitatively:

$$\text{Texture Correlation} = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)\ \Pr(i,j)}{\sigma_i \sigma_j}, \tag{2.11}$$

where μ_i, μ_j are the means and σ_i, σ_j are the standard deviations of the row and column marginals of $\Pr(i,j)$. Correlation measures the linear dependency of grey-level values in the GLCM; higher values indicate that similar or related intensity values co-occur more frequently at the specified offset.

$$\text{Texture Homogeneity} = \sum_{i,j} \frac{\Pr(i,j)}{1+(i-j)^2}. \tag{2.12}$$

Homogeneity favours uniform regions where $i \approx j$; larger values imply more uniform textures and reduced intensity variation.

In addition to these GLCM-derived features, other texture-based methods (*e.g.,* Local Binary Patterns, Wavelet Transforms) can be employed to capture higher-level textural attributes. However, GLCM remains a popular starting point due to its simplicity, interpretability, and the direct relation of its features to perceptual image characteristics such as smoothness, coarseness, and regularity.

2.2.4.2 *Entropy and Information Content*

Entropy is a fundamental concept for quantifying the richness and complexity of an image. While the GLCM approach calculates a second-order statistical measure of entropy (based on joint intensity distributions), one can also consider the first-order Shannon entropy of an image's intensity or lifetime histogram:

$$H = -\sum_{g=0}^{G-1} \Pr(g)\ \log(\Pr(g)), \tag{2.13}$$

where $\Pr(g)$ is the normalised histogram probability of grey-level g, and G is the total number of intensity (or lifetime) levels. High Shannon entropy in an image typically indicates a broader range of captured intensities or lifetime values, suggesting greater information content. In the context of FLIm, increased

entropy can arise from heterogeneous tissue properties or complex molecular environments, thereby providing richer diagnostic or research insights.

Furthermore, the concept of information content extends to multi-dimensional representations (e.g. intensity-plus-lifetime or multi-spectral data), in which joint entropy or mutual information can be used to evaluate the synergy or redundancy between different contrast mechanisms. Such measures are particularly relevant when fusing data from multiple imaging modalities or when seeking to integrate lifetime information with morphological or functional metrics for comprehensive image analysis.

In summary, the careful selection and computation of texture metrics and entropy-based measures allows for a deeper understanding of image complexity and feature richness. These metrics are central to robustly comparing fluorescence lifetime images to conventional imaging modalities and optimising processing pipelines that maximise diagnostic and research potential.

2.2.4.3 *Signal-to-Noise Ratio*

In microscopic imaging (and indeed many other imaging modalities) [50, 63], a dedicated "background" or "calibration" image is routinely acquired to capture noise inherent in the imaging system—such as detector read noise, thermal noise, and sensor-specific fixed pattern noise. By measuring this background separately (*i.e.,* without sample signal or with a uniform reference input), one can isolate these unwanted contributions rather than conflating them with the actual sample's signal.

Mathematically, the SNR is defined as the ratio of the desired signal to the underlying noise. Let $\mathbf{I}_{\text{raw}}(x, y)$ represent the raw image (containing both sample and system contributions) and $\mathbf{I}_{\text{bg}}(x, y)$ denote the background image (acquired under identical conditions but without the sample signal). If we let $\overline{\mathbf{I}_{\text{raw}}}$ be the mean pixel value in the region of interest (RoI) of the raw image and $\overline{\mathbf{I}_{\text{bg}}}$ be the mean pixel value in the corresponding RoI of the background image, then the effective signal (after background subtraction) can be approximated as:

$$\mathbf{S} = \overline{\mathbf{I}_{\text{raw}}} - \overline{\mathbf{I}_{\text{bg}}}, \tag{2.14}$$

Similarly, let σ_{bg} denote the standard deviation of the pixel values in the background image over the same RoI. By assuming that the principal noise source in the system is captured by σ_{bg}, one can estimate the SNR as:

$$\text{SNR} = \frac{\mathbf{S}}{\sigma_{\text{bg}}} = \frac{\overline{\mathbf{I}_{\text{raw}}} - \overline{\mathbf{I}_{\text{bg}}}}{\sigma_{\text{bg}}}. \tag{2.15}$$

This formulation ensures that the standard deviation reflects the variability introduced by the camera or the system itself, rather than variability due to biological or structural features in the sample [79, 80, 81]. Directly computing the standard deviation from the raw sample image alone would conflate system noise with genuine fluctuations in the sample, resulting in an overestimation of noise for regions containing the signal.

In practice, the background image is often subtracted from the raw image before calculating the mean sample signal, and the standard deviation from the background image is used to approximate the true noise floor [82]. This procedure yields a more accurate characterisation of the SNR, thereby improving subsequent image processing steps—such as thresholding or quantitative analysis—by providing a clearer distinction between true sample signal and inherent system variability. For lifetime images, image quality improves with increasing photon counts and with the provision a useful lower bound on achievable performance in the absence of dominant background or instrumental noise sources [71].

2.2.4.4 *A Multi-Metric Perspective*

When comparing two imaging modalities such as intensity-based fluorescence and FLIm, relying on a single feature metric (*e.g.,* SNR) can yield an incomplete or potentially misleading assessment of image quality and diagnostic utility. Each metric—be it entropy, SNR, texture correlation, or texture homogeneity—captures distinct aspects of image structure and content. Consequently, it is essential to evaluate multiple metrics in parallel in order to obtain a robust understanding of each modality's strengths and limitations.

Different metrics offer complementary insights. For instance, high entropy indicates complexity or diversity in image intensities, but without knowledge of the SNR, it is difficult to determine whether that complexity arises from meaningful signal variations or from random noise. Texture correlation assesses the spatial relationships between neighbouring pixels and can reveal structured patterns that may not be apparent from entropy alone. Meanwhile, texture homogeneity provides an indication of uniformity, highlighting whether distinct regions within the image exhibit markedly different properties. In practice, each modality or application domain—be it medical diagnosis or materials science—may place different emphasis on these metrics, depending on whether the goal is to detect subtle anomalies, characterize repetitive patterns, or identify broad structural changes.

A telling example arises when FII modality demonstrates simultaneously higher entropy, higher SNR, higher texture correlation, but lower texture homogeneity relative to FLIm, as shown in Fig. 2.4. Under these conditions, the higher entropy suggests a wealth of detailed information or complexity in the acquired images, and

the elevated SNR ensures that this complexity is predominantly due to genuine signal rather than noise artefacts. High texture correlation further indicates strong spatial relationships or repetitive patterns—useful for highlighting structural consistencies. At the same time, lower texture homogeneity signifies that the signal distribution varies considerably across different regions, which can be advantageous if the goal is to capture nuanced transitions or heterogeneities within the sample or tissue.

Overall, this multi-metric perspective highlights why clinicians and researchers should not confine themselves to a single scalar metric when comparing intensity vs. lifetime imaging. By evaluating metrics like entropy, SNR, texture correlation, and homogeneity collectively, one can more accurately interpret the strengths and limitations of each modality, ensuring that imaging pipelines and subsequent analyses are tailored to harness the most clinically or scientifically relevant information.

2.3 OPTICAL ENDOMICROSCOPY USING FLUORESCENCE LIFETIME IMAGING

In OEM-FLIm, the dynamic nature of the lung presents a unique challenge. Building on the insights gained from the comparative analysis of intensity and lifetime imaging, this section explores how fibre-based endomicroscopy can leverage both modalities to provide a multifaceted view of alveolar structures [83, 19]. As discussed previously, intensity measurements offer rapid and straightforward imaging of tissue features, while lifetime contrast reveals crucial metabolic and biochemical details that might otherwise remain undetected [83]. By incorporating both forms of data, FLIm opens new possibilities for pulmonary imaging, particularly in contexts where a deeper understanding of tissue function and pathology is required [7].

Early efforts in fluorescence lifetime measurement, such as those by Weber [84], paved the way for quantitative phase and modulation techniques, laying the groundwork for more advanced time-domain and frequency-domain approaches. Building upon this foundation, Lakowicz *et al.* contributed significantly since around 1992, developing methodologies to enhance fluorescence lifetime accuracy and sensitivity, eventually guiding the field toward improved instrumentation and data analysis [83, 14]. These pioneering studies set the stage for novel systems that optimised photon efficiency—an essential factor in obtaining high SNR in FLIm.

A prominent example of such a system is described by Colyer *et al.* [52], where a TCSPC framework was integrated to achieve precise lifetime measurements with minimal photon loss. Colyer *et al.* work exemplifies how hardware and computa-

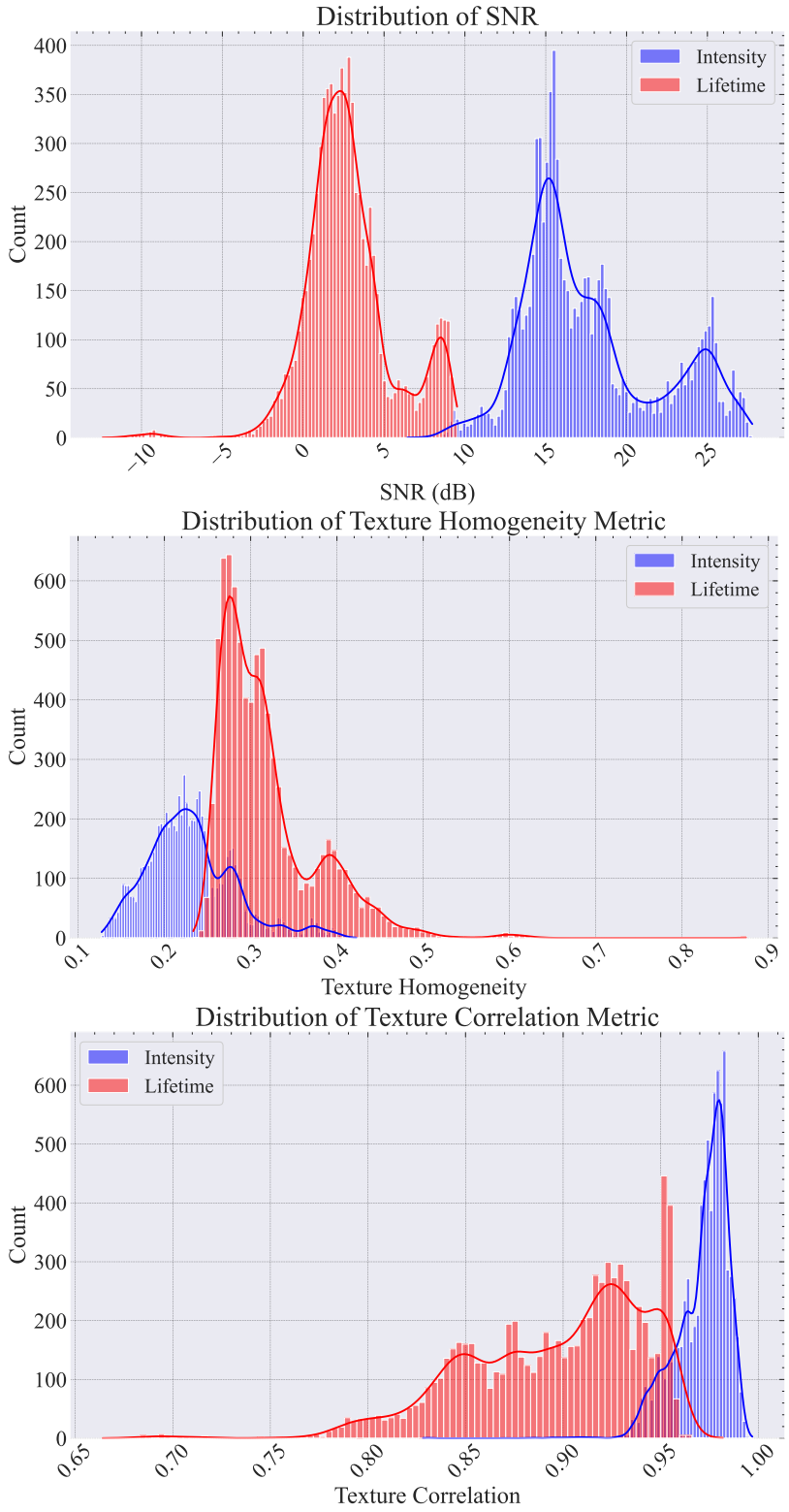


Figure 2.4: Comparing the SNR, texture homogeneity and correlation between intensity and lifetime data.

tional improvements have culminated in FLIm systems that can simultaneously attain high temporal resolution and robust photon economy. Recent advancements continue along this trajectory, as highlighted in [55], underscoring ongoing innovations in instrumentation and analysis techniques that further expand the utility and accuracy of fluorescence lifetime imaging.

2.3.1 *The KronoScan System*

The KronoScan system is another example of a clinic-ready OEM-FLIm platform, building on earlier innovations while incorporating on-chip histogramming for real-time [85], *in situ* imaging. Engineered for use with the Panoptes fibre [74]—a fully biocompatible, multi-functional endoscopic fibre bundle capable of both imaging and fluid delivery [74]—KronoScan provides simultaneous fluorescence intensity and lifetime data at video-rate acquisition speeds [12, 13, 67]. By addressing both structural and biochemical information within the alveoli, this integrated approach lays the groundwork for more accurate disease characterisation and real-time clinical decision-making in pulmonary imaging.

The system is built using the TCSPC technology [12], allowing for intensity and lifetime imaging at 5–8 frames per second (fps) [13]. It incorporates an advanced confocal laser scanning microscope to enhance optical resolution and imaging precision. Specifically, KronoScan operates as a confocal line-scan imaging system, in which a line of illumination is scanned across the distal end of the fibre bundle and detected synchronously by a corresponding line of sensor elements. This line-scanned configuration enables efficient spatial sampling while maintaining optical sectioning, distinguishing the system from point-scanned or wide-field FLIM implementations.

At the detector level, KronoScan employs a single-photon avalanche diode (SPAD)-based line-scan sensor, in which individual SPAD elements register single-photon arrival events with picosecond-scale temporal resolution. Each SPAD pixel operates in Geiger mode, producing a digital pulse for every detected photon, enabling direct compatibility with TCSPC-based timing electronics. The Panoptes fibre, with an outer diameter of 1.6 mm, houses up to an 8100-core imaging bundle alongside two capillary channels, facilitating fluid delivery and extraction during imaging. An overview of the KronoScan OEM-FLIm system employed in this thesis is shown in Fig. 2.5. The illustration highlights how the Panoptes fibre bundle, threaded through a bronchoscope's working channel, facilitates real-time fluorescence lifetime imaging of alveolar regions deep within the lung.

KronoScan captures simultaneous fluorescence intensity and lifetime video sequences at a spatial resolution of 128×128 pixels, corresponding to a 350×350

μm field of view (FoV). The system operates within two spectral emission bands: 498–565 nm (green) and 620–764 nm (red). For each spatial location, the raw data accessible from the acquisition system consist of early- and late-gated photon count images, derived from the on-chip temporal histogramming of detected photon events. These gated intensity images have dimensions consistent with the reconstructed 128 × 128 image frame.

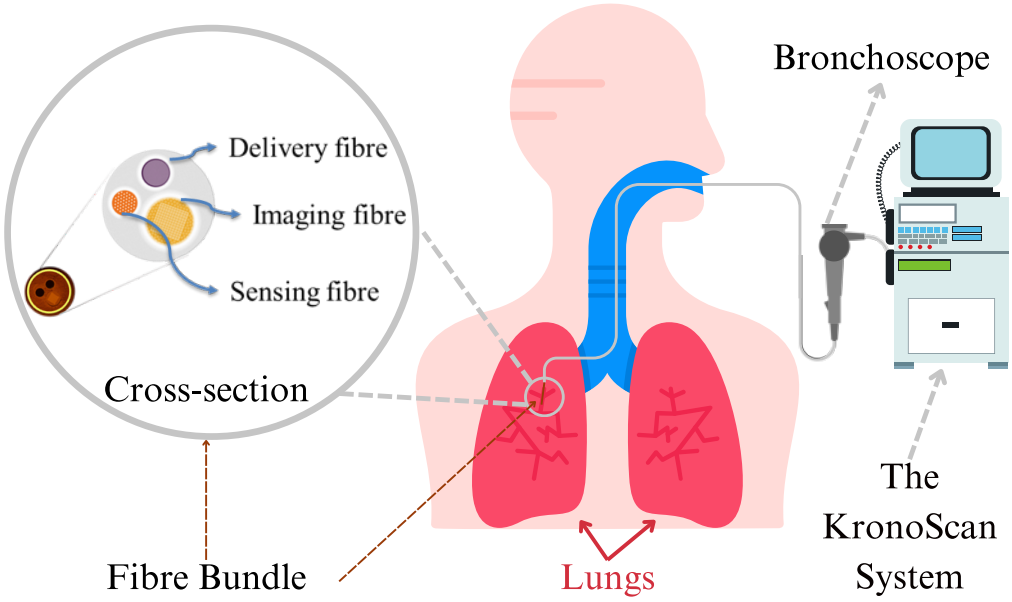


Figure 2.5: Schematic of the KronoScan OEM-FLIm system used for in vivo pulmonary imaging. The left side shows the KronoScan device connected to a standard bronchoscope, while the middle depicts the bronchoscope's working channel guiding the fibre bundle (Panoptes) from the patient's mouth into the lung. An inset on the right provides a cross-section of the fibre bundle. This configuration enables simultaneous fluorescence lifetime data acquisition within the alveolar space, supporting the algorithmic development and testing detailed later in this thesis.

Detected photon events from the SPAD array are accumulated on-chip into coarse temporal histograms, reducing data bandwidth and enabling video-rate operation. These histograms represent the photon arrival distribution following each excitation pulse and form the raw input to the lifetime estimation stage. Fluorescence lifetime decays are processed using a RLD method, which assumes a single-exponential decay model, enabling high-speed real-time processing.

Fluorescence lifetime values are not directly measured by the sensor, but are estimated from the early- and late-gated intensity readout using the RLD formulation described in Section 2.2.3.2. The resulting per-pixel lifetime and integrated intensity images constitute the primary image representations used

as inputs for the downstream image processing and analysis pipelines described later in this thesis.

Designed for clinical and biomedical research applications, KronoScan is capable of detecting labelled cells in human lungs [17, 74, 86], as well as imaging a variety of biological specimens [8]. Its dual-colour excitation capability, combined with an achromatic optical system spanning 400–900 nm, ensures robust fluorescence signal detection across a broad spectral range. The system can distinguish metabolic states of germ cells in live tissue using FLIM, providing insights into cellular activity. Additionally, it supports the detection of NAP signals [17], enabling the visualisation of activated neutrophils in pulmonary imaging applications. In these endoscopic applications, the FLIm modality provides contrast based on fluorescence decay dynamics rather than intensity alone, allowing functionally distinct tissue regions and labelled cellular populations to be differentiated under variable illumination and motion conditions. This is particularly beneficial in bronchoscopic imaging, where lifetime-based contrast improves robustness to intensity fluctuations caused by tissue heterogeneity, fibre coupling variability, and respiratory motion. Moreover, KronoScan is designed to integrate seamlessly with existing bronchoscopic and thoracic imaging platforms, broadening its clinical utility for real-time pulmonary imaging and fluorescence-guided diagnostics.

To evaluate and refine KronoScan's performance under controlled yet physiologically relevant conditions, the *Ex vivo* lung ventilation (EVLV) protocol is employed. This approach allows for the study of pulmonary immunology and imaging validation in an *ex vivo* environment that closely approximates *in vivo* conditions. The subsequent section details this biological setup, outlining how human lungs, preserved for research, are ventilated and maintained to enable systematic investigations of OEM-FLIm capabilities.

2.3.2 *Biological Setup*

Although the KronoScan system is clinic-ready, the experimental setup takes advantage of EVLV, a technique that enables the study of human lungs in a controlled environment outside the body. This method preserves the physiological characteristics of the lungs, closely mimicking *in vivo* conditions, and provides a platform for detailed investigations using OEM-FLIm. By maintaining functional lung tissue, EVLV facilitates the study of pulmonary immunology, imaging technology validation, and therapeutic development [8].

The lungs used in EVLV experiments are obtained from deceased donors when deemed unsuitable for transplantation. These lungs are transported to the laboratory within 48 hours of explantation, ensuring that tissue viability is preserved for experimental use. Upon arrival, the lungs are prepared and

ventilated using a ventilator, following the insertion of an endotracheal tube to establish a controlled airway. Ventilation parameters are carefully regulated to maintain physiological conditions. The system operates in volume-controlled mode, using air as the ventilatory medium. A tidal volume of 7 mL/kg of the donor's body mass is applied at a respiratory rate of 12 breaths per minute.

The EVLV platform serves as a robust experimental model for *in situ* immunolabelling and advanced imaging. The Panoptes multifunctional fibre, used in conjunction with the KronoScan optical endomicroscopy system, mentioned earlier in Section 2.3.1, enables the precise delivery of fluorescently labelled antibodies [17, 49] into the alveolar space. This is achieved through the fibre's dedicated delivery channel, allowing targeted labelling of specific immune cell populations. Following the administration of these molecular probes, fluorescence intensity and lifetime imaging are performed to visualise immune activity and cellular interactions. The Panoptes fibre is navigated through the distal airways using autofluorescence intensity imaging, ensuring accurate positioning within the lung tissue.

One of the key advantages of EVLV is its ability to maintain a physiologically relevant environment, allowing for more accurate assessments of lung function and immune cell behaviour. The technique is inherently compatible with OEM, supporting both intensity and lifetime imaging within intact lung tissue. Moreover, EVLV enables *in situ*, no-wash immunolabelling, permitting high-resolution investigations of immune cell interactions without requiring disruptive sample preparation.

In summary, EVLV represents a pivotal technique for maintaining viable human lungs *ex vivo*, facilitating advanced pulmonary imaging and immunological studies. When combined with the KronoScan system and Panoptes fibre, it provides a powerful tool for advancing our understanding of lung immunity, disease progression, and therapeutic interventions.

However, this *ex vivo* lung preparation is still subject to the inherent, cyclical movements that characterise normal respiration. The continual expansion and contraction of alveolar structures can introduce motion artefacts that obscure key details, diminish image quality, and complicate data interpretation [1, 7]. To address these challenges effectively, it is useful to revisit the fundamentals of pulmonary motion and respiratory mechanics, particularly how inhalation, exhalation, and other physiological processes govern lung function at the alveolar level.

2.3.3 *Pulmonary Motion and Respiratory Mechanics*

Pulmonary motion is essential to the primary function of the lungs: sustaining gas exchange by continuously cycling air in and out of the alveolar spaces [2]. During inhalation, negative intrapleural pressure expands the alveoli, whereas exhalation prompts recoil of the elastin-rich alveolar walls, driving air out of the lungs [2, 29]. Surfactant, a complex of lipids and proteins secreted by type II alveolar cells, reduces surface tension within each alveolus, preventing collapse (atelectasis) and allowing alveoli of varying sizes to remain open throughout the breathing cycle [2, 29].

These respiratory mechanics—spanning lung volumes and capacities, tissue compliance and elasticity, airway resistance, and the matching of ventilation to perfusion—govern how effectively oxygen and carbon dioxide are exchanged in the pulmonary capillaries [28, 87, 88]. Notably, diseases such as pulmonary fibrosis or chronic obstructive pulmonary disease (COPD) alter key parameters (*e.g.,* tissue stiffness or airway calibre), modifying local deformation and airflow patterns in ways that can complicate imaging or obscure structural details. OEM imaging of alveolar spaces provides a high-resolution view of alveolar dynamics in real-time. When the imaging probe is gently placed against the alveolar wall, the local motion arises from the cyclic expansion/contraction of the alveoli (at roughly 12 breaths per minute in humans) and from any physiological response triggered by the probe's contact.

In KronoScan, the OEM-FLIm system described earlier in Section 2.3.1, frames are acquired at a rate of about 5 Hz, sufficient to capture rapid phases of alveolar expansion and recoil throughout each respiratory cycle. Nonetheless, continuous lung movement remains a significant source of potential artefacts in high-resolution datasets. Cyclic alveolar motion can blur or misalign frames, especially if the fibre bundle's contact with the tissue momentarily changes. Thus, any meaningful analysis of OEM-FLIm data demands strategies to mitigate the impact of this inherent motion.

2.3.4 *Implications for Medical Imaging*

Breathing-induced motion poses substantial challenges for all high-resolution imaging modalities, including CT, MRI, and the aforementioned OEM-FLIm. Although techniques such as respiratory gating [89] and breath-hold protocols [90] can reduce lung motion artefacts, they may be impractical for patients who cannot tolerate breath holds or for scenarios requiring uninterrupted, real-time data capture. At the alveolar scale, even minute shifts in surfactant distribution

or alveolar inflation can degrade image clarity by causing misalignments between consecutively acquired frames [91].

These motion-related distortions have wide-ranging ramifications. First, they can obscure critical anatomical and functional signals, making it more difficult to detect subtle or early-stage pathological changes [21]. Second, they complicate quantitative analyses such as fluorescence lifetime measurements, which rely on stable, artefact-free frames to distinguish molecular variations [19]. Third, they interfere with downstream image processing [1], from automatic segmentation of alveolar walls to data-driven classification or machine-learning algorithms that expect consistent, comparable inputs across frames or time points [10, 92].

In response to these issues, robust motion compensation is a prerequisite for the accurate assessment of lung structure and function in pulmonary imaging. Strategies include:

- **Gating and Triggering:** Synchronising acquisition with specific phases of the respiratory cycle to minimise misalignment [93].
- **Rapid Acquisition:** Shortening image capture times to "freeze" lung motion, as in ultra-fast MRI [94] or advanced CT scans [95].
- **Algorithmic Registration:** Employing image registration to correct for frame-to-frame shifts and deformations, creating consistent datasets ready for quantitative or interpretive analysis [21, 19].

Incorporating these techniques can significantly improve the reliability and diagnostic precision of OEM-FLIm. By mitigating or correcting motion artefacts, researchers and clinicians can exploit the rich biochemical information embedded in lifetime signals—often elusive in more conventional imaging—thereby enhancing our ability to study alveolar-level processes, track disease progression, and monitor therapeutic interventions in real time.

2.4 MOTION ARTEFACTS AND THEIR COMPENSATION

Medical imaging is inherently susceptible to a range of artefacts that can compromise both the quality and interpretability of the resulting data. These artefacts may arise from diverse sources, such as hardware limitations, algorithmic assumptions, and inconsistencies during image acquisition [21]. Among the most prevalent and impactful are motion artefacts, which commonly manifest as blurring, ghosting, or spurious image shifts when the patient or organ of interest moves during scanning [38, 1]. For pulmonary imaging in particular, the cyclical nature of respiration introduces continuous motion that may distort or misrepresent key anatomical structures, thereby affecting diagnostic and therapeutic

decision-making. Other notable artefacts encountered in medical imaging include truncation artefacts in PET/CT scans, noise artefacts in OEM due to inadequate fibre-tissue contact, and geometric distortions that may arise during data fusion or image registration [96, 97]. In the context of pulmonary OEM-FLIm, successful image processing does not necessarily imply anatomically exact reconstruction of tissue geometry. Rather, the primary objective is to achieve sufficient temporal stability and feature consistency to support reliable human interpretation and downstream automated analysis. Small residual misalignments or intensity variations may be acceptable provided that diagnostically relevant patterns, contrast, and lifetime trends are preserved across frames.

Efforts to minimise or correct motion-induced artefacts generally fall under the umbrella of motion compensation. These approaches use image registration, gating, or model-based corrections to adjust the acquired data and enhance visual clarity [21]. Image registration, for example, aligns images or frames into a common spatial coordinate system to account for rotations, translations, and deformations induced by patient movement. Such registrations are integral to tasks like motion tracking, dose accumulation, and multi-modal image fusion. Robust methods often employ a combination of objective functions—e.g. image gradient, entropy-based terms, and smoothing penalties—to converge on a motion solution that best preserves anatomical detail while ensuring continuity of structures [38].

Advanced, application-specific techniques have also been proposed. In two-photon imaging of awake animals, for instance, real-time motion tracking is achieved via the Lucas–Kanade framework, alleviating both fast and slow displacements without the need for an external reference signal or distortion-free template [98]. Similar principles guide deformable image registration for thoracic CT images, where the lung's complex mechanics and variability in tissue density mandate flexible, non-rigid transformations [38]. In OEM-FLIm, one effective strategy for dealing with uninformative frames involves detecting and removing them based on textural cues such as GLCM measures, thereby preventing motion artefacts or pure-noise sequences from contaminating subsequent analyses [1].

Taken together, these strategies emphasise that effective motion compensation is paramount for accurate localisation, structural interpretation, and quantitative assessment in modern medical imaging. From a clinical standpoint, especially in pulmonary imaging, mitigating respiratory artefacts not only improves diagnostic confidence but also offers valuable physiological and functional information. It enables applications ranging from precise lesion tracking and image-guided interventions to a deeper, more reliable characterisation of dynamic processes in the lung, thereby reinforcing the importance of integrating robust motion compensation techniques into advanced imaging pipelines.

2.4.1 *Understanding and Addressing Uninformative Frames*

Uninformative frames occur in OEM, and in OEM-FLIm data specifically, when the imaging system fails to capture meaningful structural or functional details, typically due to poor contact with the tissue or excessive motion at the time of acquisition [1]. Such frames can be broadly classified into pure-noise frames, arising from inadequate probe–tissue contact, and motion-corrupted frames, in which abrupt or substantial movement distorts anatomical features. Notably, these corrupted frames can constitute a sizeable fraction of an OEM dataset, substantially increasing computational overhead for downstream analysis and diluting quantitative outcomes.

In an influential study, Perperidis *et al.* [1] observed that uninformative frames could exceed 25 % of OEM datasets. They employed a Gaussian mixture model classification scheme, leveraging texture descriptors extracted via the GLCM to detect and remove noise and motion-artefacted frames. The method reported 93 % sensitivity and specificity, markedly reducing manual post-analysis effort. However, this approach primarily addressed frame exclusion rather than spatial realignment—a distinction that becomes crucial when aiming to enhance real-time FLIm quality through motion compensation.

The significance of managing motion-induced artefacts extends well beyond OEM, pervading all major imaging modalities. In PET/CT, for instance, respiratory motion can misalign CT attenuation maps and PET emission data, leading to inaccurate tracer uptake values or false positives [37, 99]. Similarly, even sub-millimetre movements in cone-beam CT (CBCT) can degrade structural measurements, necessitating robust compensation [45]. In advanced microscopy such as two-photon imaging, uncorrected motion can distort fluorescence signals and inflate the risk of false positives [98]. Across these domains, mitigating motion artefacts is paramount for precise diagnosis, treatment planning, and reliable data interpretation.

In the context of OEM-FLIm, and in the data collected using the Kronoscan (Section 2.3.1), where each frame contains both intensity and lifetime information, motion disturbances pose a particular challenge for accurate metabolic and structural assessments. Fig. 1.3, in Chapter 1, Section 1.1, highlights common imaging defects in our this system, including blurred or “chopped” frames (Figs. 1.3.b, 1.3.d) and severely degraded images devoid of any discernible structure (Fig. 1.3.c). These impair the detection of clinically relevant biomarkers and hamper downstream tasks such as fusion, classification, or quantitative analysis of regions of interest [7].

A fundamental strategy for alleviating such artefacts is image registration, which aligns images or frames into a consistent spatial reference [100, 101]. By

estimating transformations (*e.g.,* translations, rotations, non-rigid deformations) that map one frame to another, registration-based approaches can detect irretrievably corrupted frames and correct motion distortions in partially salvageable frames. This process underpins automated motion compensation in OEM-FLIm, enabling seamless mosaicking [21], expanded fields of view, and enhanced signal quality through frame averaging [19]. Ultimately, robust registration techniques help ensure that higher-level analyses—whether segmentation, classification, or functional quantification—are conducted on artefact-free data, thereby increasing confidence in research findings and clinical decisions.

2.4.2 *Medical Image Registration and Alignment Techniques*

Image registration is a cornerstone of modern medical imaging [102], offering a systematic means to align or fuse images acquired at different time points, from various viewpoints, or through complementary modalities [97]. By transforming images into a single coordinate system, registration mitigates motion artefacts and enables comprehensive analyses across multiple datasets [100, 103]. In pulmonary imaging, for instance, it can track breathing-induced lung displacements; in two-photon microscopy of awake animals [96], it corrects both rapid and gradual shifts; and in clinical endomicroscopy, it supports the creation of larger mosaics from overlapping fields of view [21]. This section outlines the fundamental concepts of medical image registration and underscores its value in diverse diagnostic and therapeutic settings.

A typical registration task involves two images in a sequence: a *reference image* $\mathbf{I}_{[K]}(\boldsymbol{\chi})$ and a *moving image* $\mathbf{I}_{[K+1]}(\boldsymbol{u})$. Each image $\mathbf{I}_{[K]}$ in the sequence is indexed by $K = \{1, \ldots, n\}$. Here, $\boldsymbol{\chi} = (x, y)$ denotes the spatial coordinates in the reference image and $\boldsymbol{u} = (u, v)$ the corresponding coordinates in the moving image. The objective is to determine a transformation $\boldsymbol{T}$ that maps $\boldsymbol{u}$ in the moving image to $\boldsymbol{\chi}$ in the reference image, thereby aligning anatomical or functional structures across the two datasets. Depending on the nature and extent of deformation, $\boldsymbol{T}$ may be *rigid*, involving only rotations and translations, or *non-rigid*, accommodating shape changes and elasticity [104, 105].

Rigid registration is often adequate in scenarios where structures are solid and minimally deformable, such as bones in orthopaedic imaging [38]. However, dynamic organs like the brain or lungs can undergo significant deformation, requiring more complex *non-rigid* or *deformable* registration strategies. These may employ splines or finite-element models, optimised via similarity metrics (e.g. mutual information, cross-correlation) or feature-based descriptors (*e.g.,* landmark correspondences) [106, 107].

By substantially reducing artefacts and enhancing image clarity, registration bolsters both routine clinical practice and advanced research applications. In radiotherapy planning, for example, precise alignment of serial imaging studies ensures accurate dose delivery to tumours while sparing normal tissues, even when respiration shifts anatomical landmarks [106]. In drug-target interaction studies, consistent alignment of time-series data highlights how biological responses evolve following therapeutic interventions [19]. Equally important, the technique allows for low-power imaging of live tissues followed by registration-based averaging to improve SNR, facilitating the detection of subtle events—such as single-action-potential signals—without introducing false positives [96].

Taken together, image registration and alignment techniques form the operational backbone of robust motion compensation pipelines in medical imaging. Whether correcting diaphragmatic motion in thoracic PET/CT, compensating for sub-millimetre patient movements in CBCT scans, or synchronising *in vivo* fluorescence microscopy frames, rigorous registration practices optimise data quality and analytic power—thereby strengthening diagnostic confidence, improving research outcomes, and enabling more refined, patient-specific interventions.

Broadly, registration algorithms fall into two main categories: *feature-based* and *similarity-based* methods. Feature-based approaches identify and match key points or landmarks (*e.g.,* edges, corners, anatomical fiducials) to compute the alignment transformation. By contrast, similarity-based algorithms measure the resemblance (or dissimilarity) between intensity values in overlapping regions, iteratively adjusting the transformation to minimise intensity differences or maximise mutual information—an approach especially valuable for multi-modal data where pixel intensities exhibit complex relationships.

2.4.2.1 *Feature-Based Registration*

Feature-based registration extracts and matches distinctive geometric elements (*e.g.,* points, lines, or regions) across two images [26]. Algorithms such as scale-invariant feature transform [108], speeded up robust features [109], and oriented FAST and rotated BRIEF [110] are commonly used to identify these features and establish correspondences between the *reference* image $\mathbf{I}_{[K]}$ and the *moving* image $\mathbf{I}_{[K+1]}$. Once features are matched, an appropriate transformation $\boldsymbol{T}$ is estimated to align the two images. However, when distinct features are scarce or hard to detect, estimating $\boldsymbol{T}$ becomes challenging. This issue is prevalent in medical imaging [105, 101], especially in FLIm, where dynamic imaging sequences and subtle tissue contrast can hamper robust feature extraction.

Mathematically, let F_r and F_m denote the sets of features extracted from $\mathbf{I}_{[K]}$ (reference) and $\mathbf{I}_{[K+1]}$ (moving), respectively. We seek a transformation $\boldsymbol{T}$ that aligns these features by minimising a cost function $C(\boldsymbol{T})$:

$$C(\boldsymbol{T}) = \sum_{\substack{f_r \in F_r \\ f_m \in F_m}} w(f_r, f_m) \cdot d(\boldsymbol{T}(f_m), f_r), \tag{2.16}$$

where

- $d(\cdot, \cdot)$ is a distance metric (*e.g.,* Euclidean [111], Mahalanobis [112]) measuring the discrepancy between features.
- $w(f_r, f_m)$ represents an optional weight reflecting the confidence in matching f_m to f_r.
- $\boldsymbol{T}(f_m)$ is the location of feature f_m under the transformation $\boldsymbol{T}$.

The transformation $\boldsymbol{T}$ can be rigid (allowing only rotation and translation), affine (including shear and scale), or nonlinear (*e.g.,* spline-based). Solving the registration task thus involves finding:

$$\boldsymbol{T}^* = \arg\min_{\boldsymbol{T}} C(\boldsymbol{T}), \tag{2.17}$$

often achieved via iterative optimisation (*e.g.,* gradient descent, evolutionary algorithms, or other methods) [105].

Feature-based approaches excel when distinctive landmarks are plentiful and reliably extracted, but they may fail under low-contrast conditions or when features are absent or heavily distorted. In many clinical scenarios, tissue deformability and a paucity of clear fiducials complicate correspondence matching, underscoring the importance of alternative strategies such as similarity-based registration.

2.4.2.2 *Similarity-Based Registration*

Often referred to as *area-based* or *intensity-based* registration, *similarity-based* methods directly compare pixel intensities (or voxel intensities in 3D) across image regions to measure alignment [26]. By bypassing feature detection, similarity-based algorithms can be advantageous in settings where clear landmarks are scarce or inconsistent.

Mathematically, in similarity-based registration, we define a similarity measure $\mathcal{S}$ between the reference image $\mathbf{I}_{[K]}(\boldsymbol{\chi})$ and the transformed moving image $[\mathbf{I}_{[K+1]} \circ \boldsymbol{T}](\boldsymbol{\chi})$. The goal is to find the transformation $\boldsymbol{T}$ that maximises $\mathcal{S}$:

$$\boldsymbol{T}^* = \arg\max_{\boldsymbol{T}} \mathcal{S}\Big(\mathbf{I}_{[K]}, \mathbf{I}_{[K+1]} \circ \boldsymbol{T}\Big). \tag{2.18}$$

Common choices for $\mathcal{S}$ include:

- **enhanced correlation coefficient (ECC)** [113], which extends the standard cross-correlation by weighting pixels according to their deviation from mean intensity, increasing robustness to illumination changes.
- **Mattes mutual information (MMI)** [114], which quantifies the statistical dependence between intensity distributions in the two images.
- **Optical Flow Approaches** [115], which compute a per-pixel velocity field to align the moving image with the reference image.

When $\mathcal{S}$ is maximised, it indicates that $[\mathbf{I}_{[K+1]} \circ \boldsymbol{T}]$ most closely resembles $\mathbf{I}_{[K]}$, implying successful alignment.

In many cases, a gradient-based or iterative search algorithm is employed to optimise $\mathcal{S}$ [105, 102]. Alternatively, simpler approaches use a metric such as NCC to estimate an offset between two images without requiring iterative solutions [100, 101]. Although similarity-based registration has advanced significantly in modalities like X-ray, CT, or MRI [101, 100, 116], it remains relatively unexplored for FLIM.

2.4.3 *Limitations and Challenges in Image Registration*

While image registration is a powerful tool for aligning and fusing medical images, its application is far from straightforward. In the context of pulmonary OEM-FLIm, the scarcity of established solutions reflects the data- and application-specific nature of most existing registration methods. To leverage the potential of the KronoScan OEM-FLIm system (detailed in Section 2.3.1), it is essential to survey and evaluate both feature-based and similarity-based approaches, identifying strategies capable of offering robust performance under the unique constraints of endomicroscopic fluorescence lifetime data. Yet, in doing so, numerous challenges can limit the accuracy, reliability, and clinical utility of image registration techniques. Traditional methods employing NCC, mean squared error minimisation, and optical flow have not fully addressed the substantial spatial displacements or highly dynamic sequences that characterise real-time pulmonary FLIM. Consequently, alignment uncertainties persist, and practical validation in FLIm imaging remains necessary. Although recent efforts focus on computational efficiency and generalisability [117, 118], a robust, systematic approach tailored to the complex data attributes of FLIm remains an ongoing research challenge.

Medical image registration inherently depends on the quality of the data being aligned [102, 116, 73]. Low contrast, heavy noise, and significant motion artefacts—commonly encountered in dynamic lung imaging—can complicate the

search for accurate correspondences between frames. As discussed in Section 2.4.1, OEM data may contain uninformative frames caused by poor fibre–tissue contact or operator-induced movements, further hampering alignment. These issues become especially acute in multi-modal or multi-spectral settings, where conventional similarity measures (*e.g.,* mean squared error) may fail to capture meaningful correspondences between images with fundamentally different intensity distributions [100].

Because pulmonary structures are highly dynamic and continuously subjected to respiratory and cardiac motion, registration complexity is amplified. Efficiently optimising large-scale transformations in pulmonary image registration can be computationally demanding, often necessitating iterative processes with intensive parameter tuning. Moreover, the risk of converging to local rather than global optima can extend run times and undermine reliability. In clinical settings, where real-time or near real-time analyses may be desired, this computational overhead can impede the practical adoption of complex registration solutions. Additionally, many methods lack robust performance under challenging conditions such as extensive organ displacement, low signal-to-noise ratios, or image corruption. In these scenarios, knowledge of the underlying transformation model—or the capability to accurately characterise inter-frame motion—can be advantageous.

Evaluating registration accuracy poses another challenge. Ground-truth deformations are rarely available in *in vivo* pulmonary imaging, making it difficult to establish objective benchmarks. Moreover, manual annotation of anatomical landmarks is laborious, operator-dependent, and susceptible to inter-observer variability. Although various metrics—such as target registration error or cross-correlation—exist, no universal standard reliably captures all relevant facets of performance across diverse protocols and clinical contexts.

2.4.3.1 *Towards a Holistic Image Processing Pipeline*

Given these constraints, image registration should not be viewed in isolation but rather integrated into a broader, more comprehensive pipeline that encompasses:

1. **Data Acquisition Protocols:** Optimising scan parameters, refining sample preparation, and employing gating techniques can help mitigate motion artefacts before registration even begins.

2. **Pre-processing Steps:** Filtering noise, detecting and removing uninformative frames, and applying contrast enhancement strategies can substantially improve registration outcomes.

3. **Algorithmic Adaptation:** Tailoring registration algorithms to the specific properties of OEM-FLIm data is critical to achieving robust alignments.

This may include multi-scale methods, custom similarity metrics, or multi-objective strategies.

4. **Post-registration Analysis:** Downstream tasks such as mosaic generation, 3D reconstruction, and quantitative biomarker extraction all profit from precise alignment while also providing feedback on registration quality. Tight integration among these stages can reveal and correct residual misalignments early on.

In summary, while image registration holds immense promise for correcting motion artefacts, tracking structural changes, and enhancing multi-modal data fusion in pulmonary OEM-FLIm, a piecemeal application of individual routines can falter when confronted with real-world clinical and research demands. Addressing the broad range of limitations—from data quality and computational overhead to organ motion and validation challenges—facilitates the development of a holistic pipeline, unlocking the full potential of OEM-FLIm and establishing a strong foundation for this thesis's overarching objectives.

2.5 CONCLUSION

This chapter has outlined the evolution of pulmonary imaging, highlighting the advantages and limitations of both conventional and advanced imaging modalities. Fluorescence Lifetime Imaging FLIm demonstrates significant promise for enhancing lung diagnostics by offering functional insights at the alveolar level. However, its clinical applicability is curtailed by motion artefacts, necessitating the development of real-time motion compensation strategies.

A thorough understanding of pulmonary motion and respiratory mechanics is thus indispensable for refining imaging techniques and accurately interpreting lung images. By accounting for the lungs' dynamic behaviour and mitigating motion effects, imaging specialists can achieve clearer, more informative views of lung structure and function—ultimately improving patient outcomes in pulmonary medicine. Although direct comparisons with existing literature are challenging due to the innovative nature of these methods, a robust grounding in fundamental principles allows informed predictions about the range of motion expected when using a miniature fibre bundle to image alveolar spaces. Additional empirical research will further elucidate the dynamics observed in such advanced imaging techniques. With these challenges in mind, the development of reliable motion compensation algorithms is central to ensuring that FLIm data retain their integrity and diagnostic value in clinical environments. The subsequent chapters of this thesis explore strategies for mitigating these artefacts through a robust image processing pipeline, enabling accurate and stable FLIm data processing for pulmonary research and clinical applications.

3

TEMPORAL RELIABILITY AND ACCURACY VIA CORRELATION ENHANCED REGISTRATION

3.1 INTRODUCTION

The preceding chapters established the significant potential of OEM-FLIm for *in vivo* pulmonary diagnostics, while also highlighting a critical obstacle: motion artefacts inherent to real-time imaging within the lung. To unlock the full clinical and research utility of this technology, these artefacts must be addressed through a robust and computationally efficient processing framework. This chapter introduces TRACER, a bespoke image processing pipeline developed specifically to overcome these challenges.

The TRACER pipeline is engineered to deliver temporally stable and accurately aligned FLIm sequences by integrating a series of targeted processing steps. It begins with the systematic removal of uninformative frames, followed by a motion characterisation stage that employs dense OF to understand inter-frame dynamics. The core of the pipeline is a novel tracking-based registration method that leverages the channel and spatial reliability tracker (CSRT) algorithm to guide a NCC alignment process.

This integrated approach yields substantial gains in registration performance, enhancing key quality metrics by 20 % to 30 % compared to baseline methods. Critically, the TRACER registration method is not only more precise than current state-of-the-art techniques but is also significantly more efficient, operating at a speed approximately an order of magnitude faster than the next best-performing alternative. By effectively managing the complex motion patterns found in pulmonary FLIm data, TRACER provides the reliability and real-time capability essential for advancing clinical research applications.

3.1.1 *Motivation*

While image registration is a mature field in modalities like MRI and CT [100], techniques specifically tailored for FLIm are notably underdeveloped. Standard methods often rely on NCC as a similarity metric or cost function for alignment [100, 105]. Other applications have used NCC for object tracking, sometimes in conjunction with fiducial markers or mean-squared error minimisation [119, 120].

In the context of endoscopic FLIm, image registration serves a dual purpose: first, to stabilise video sequences for reliable human interpretation during real-time visualisation; and second, to ensure spatial consistency across frames for downstream computational processing, including frame averaging, feature extraction, and machine learning-based analysis.

However, these established approaches have proven insufficient for the unique demands of real-time pulmonary FLIm. They struggle to handle the intricate structural details, low signal-to-noise ratio, and significant spatial displacements that characterise these image sequences, leaving alignment uncertainties as a critical unresolved issue.

Without robust registration, apparent temporal variations in fluorescence intensity or lifetime can arise purely from motion artefacts, confounding both qualitative clinical assessment and quantitative algorithmic analysis. Although research has aimed to improve the computational speed and generalisability of registration algorithms [117, 118], rigorous validation within the specific context of FLIm imaging has been lacking.

3.1.2 *Problem Statement*

The inherent properties of FLIm data—characterised by lower spatial resolution and a reduced SNR [65, 121]—necessitate the development of sophisticated and highly specialised analysis methods. While ML approaches have begun to enhance FLIm's applicability [10], existing work, including the notable contributions by Perperidis *et al.* [1] and Sparks *et al.* [19], has yet to deliver a complete solution for real-time motion artefact correction.

This work addresses that gap by introducing the TRACER pipeline (Fig. 3.1), which provides an end-to-end framework for motion compensation explicitly designed for *in-vivo* pulmonary imaging. To our knowledge, no previously reported pipeline integrates these components to achieve a comparable improvement in both processing speed and final image quality. These enhancements are vital for preserving the temporal coherence of FLIm sequences and ensuring that downstream analyses, particularly those informing critical clinical decisions, are based on the most reliable data possible. Our experiments confirm that TRACER substantially improves the quality of fused images, leading to more dependable object detection and a significant step forward in the clinical and research utility of real-time OEM-FLIm.

3.1.3 *Chapter overview*

The structure of this chapter is as follows: Section 3.2 details the data used and the image processing techniques implemented for motion compensation in the TRACER pipeline . Section 3.3 focuses on the challenges with registering FLIm data. Section 3.4 presents the evaluation of pipeline performance through both qualitative and quantitative analyses, and discusses the findings, highlighting key advancements and limitations. Finally, Section 3.5 summarises the contributions and suggests directions for future research.

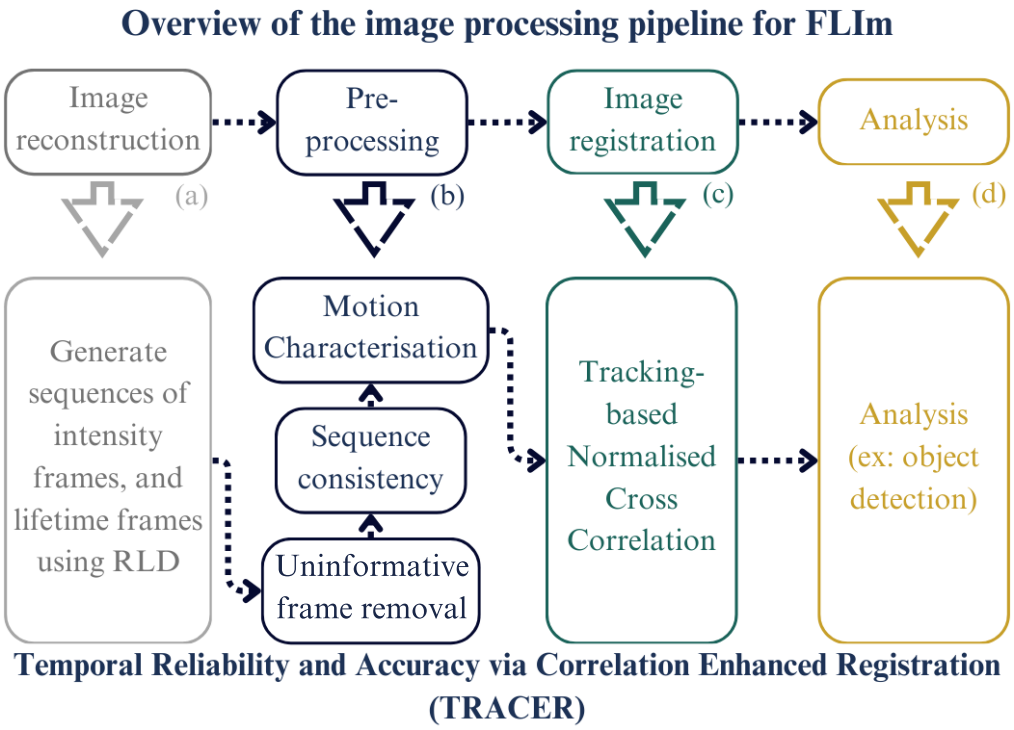


Figure 3.1: Overview of the TRACER image processing pipeline for FLIm, with emphasis on the proposed TRACER method. (a) Reconstruction of intensity and lifetime frames. (b) Three pre-processing steps ensure consistency and utility of the frames, along with motion model characterisation for the optimal choice of (c), where tracking-based NCC is selected. Finally, (d) FLIm analysis is performed, such as object detection, which is detailed later in the paper.

3.2 PROCESSING PIPELINE DEVELOPMENT

An overview of the complete TRACER pipeline is provided in Fig. 3.1. The process begins with the reconstruction of FLIm data into intensity and lifetime frames (Fig. 3.1.a). This is followed by a sequence of pre-processing steps designed to enhance data quality and consistency (Fig. 3.1.b). A motion model is then characterised to inform the optimal selection of an image registration

method (Fig. 3.1.c), leading to the final stage of downstream analysis, such as object tracking or quantitative image fusion (Fig. 3.1.d). To fully appreciate the development of these algorithmic components and their sequential impact, we first describe the imaging data used for validation.

3.2.1 *Simulation and Real Data*

To evaluate the efficacy, suitability, and limitations of the proposed FLIm image processing pipeline, this study utilised a combination of six real FLIm image sequences and seven simulated sequences of varying complexity. The real-world FLIm data were acquired with a custom-built system capable of capturing fluorescence intensity and lifetime images at up to 8 fps [13]. Each frame possesses a resolution of 128×128 pixels, with lifetime measurements extracted from the two-time-bin intensity data via the RLD method. These lifetime values were subsequently used to generate intensity-weighted lifetime image sequences, a technique known to enhance the interpretability of features [73].

The FLIm sequences were carefully selected to to evaluate the ability of the registration framework to stabilise sequences for both visual inspection and consistent downstream processing, and to represent four common scenarios encountered during acquisition, as illustrated in Fig. 3.3. These scenarios range from a stable, homogeneous sequence with a single field of view (Scenario 1), to sequences containing occasional corrupted frames (Scenario 2), and more complex sequences featuring transitions between two distinct scenes (Scenario 3) or multiple scene changes due to respiratory motion or operator movement (Scenario 4).

For the simulated data, shown in Fig. 3.2, a brick wall texture from the Brodatz database [123] was used as a base to analyse performance against repeating structural patterns, edges, and texture. This, along with four high-resolution fluorescence images, was used to generate seven simulated sequences to specifically study image-to-image motion and assess registration performance. To better reflect the characteristics of real FLIm acquisition data, controlled noise was added to the simulated sequences, including mild spatial blurring, salt-and-pepper noise, and global intensity reduction to emulate variations in signal quality and acquisition conditions. These simulations were designed to mimic Scenarios 2 and 3, focusing on the pipeline's ability to exclude uninformative frames and detect scene changes. Sequences 1 through 5 represent Scenario 2, containing some uninformative frames and small, variable displacements in both the x and y directions. Sequences 6 and 7 represent Scenario 3, containing corrupted frames alongside significant scene changes and larger simulated displacements.

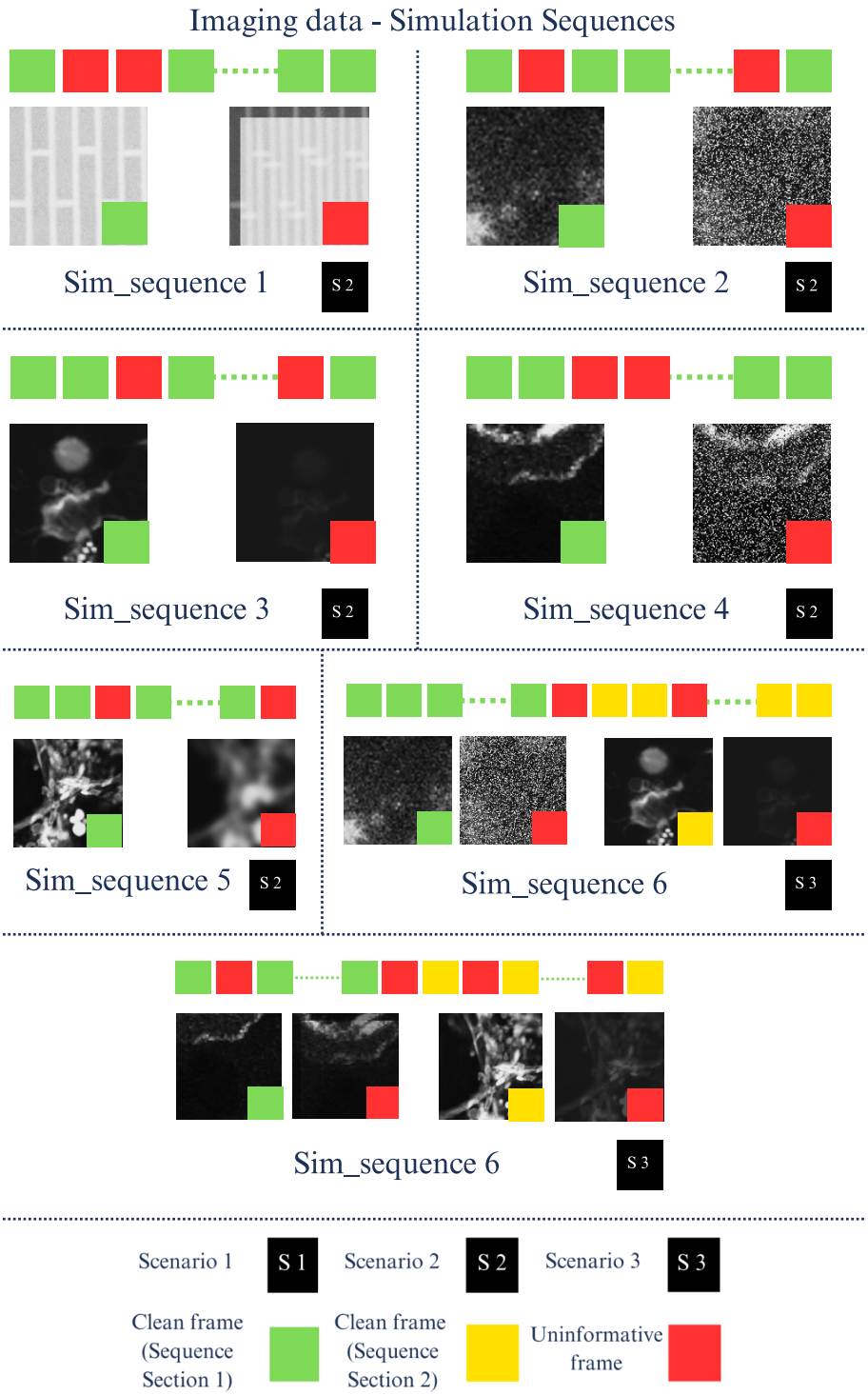


Figure 3.2: Overview of the simulated and FLIm sequences used in this study. The colour-coded squares above the image examples indicate sequence variability, corresponding to the scenarios described in 3.2.1. Scenario 1 (S1) sequences are the least variable, consisting of clean frames (green squares). Scenario 2 (S2) sequences show moderate variability, with occasional uninformative frames (red squares). Scenario 3 (S3) sequences exhibit increased variability, featuring both uninformative frames and transitions between two distinct scenes (yellow squares for the second scene).

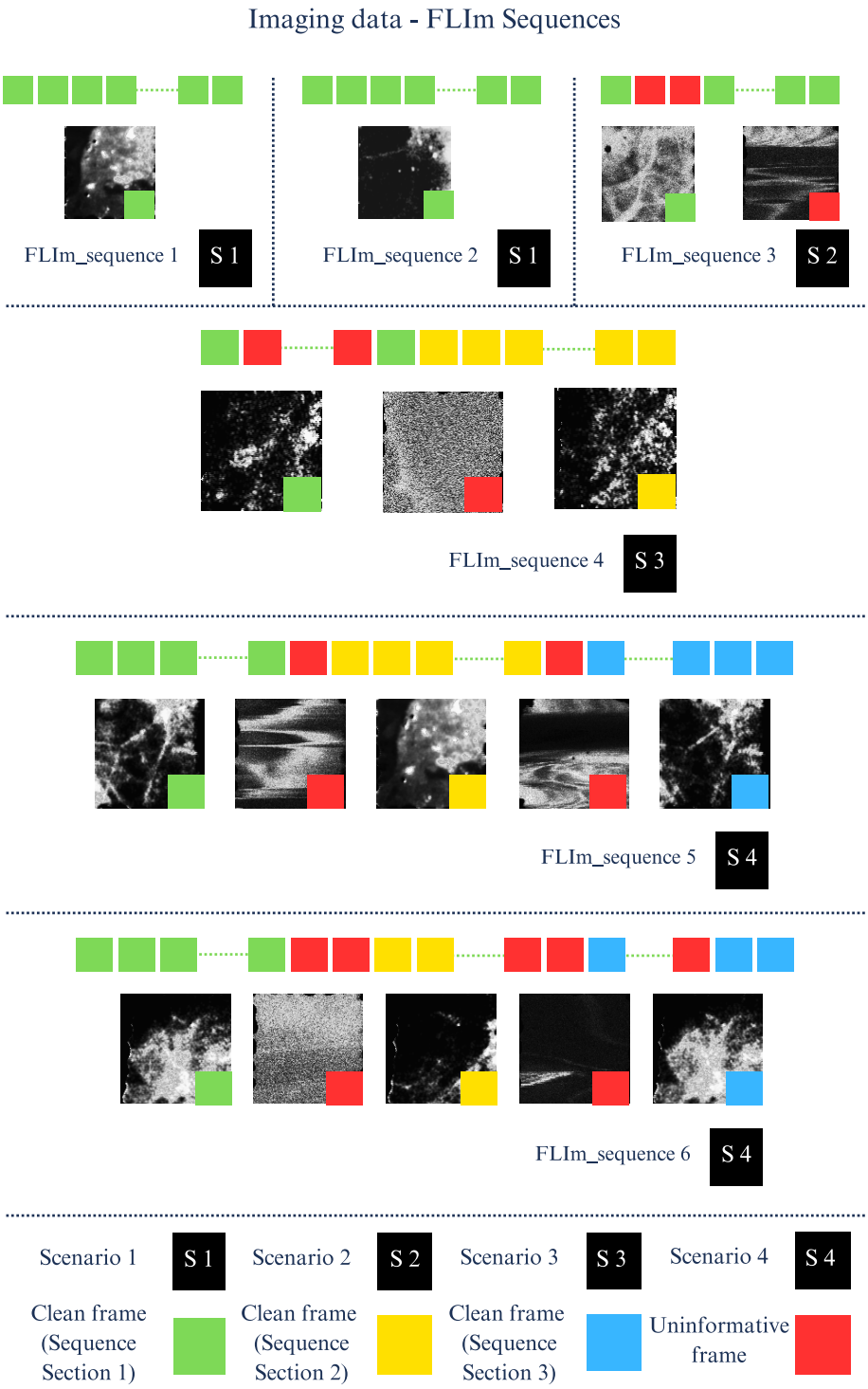


Figure 3.3: Overview of the simulated and FLIm sequences used in this study. All image examples are shown in grey-scale to facilitate the overlay of coloured bounding boxes used for indication. The colour-coded squares above the image examples indicate sequence variability, corresponding to the scenarios described in 3.2.1. Scenario 1 (S1) sequences are the least variable, consisting of clean frames (green squares). Scenario 2 (S2) sequences show moderate variability, with occasional uninformative frames (red squares). Scenario 3 (S3) sequences exhibit increased variability, featuring both uninformative frames and transitions between two distinct scenes (yellow squares for the second scene). Scenario 4 (S4) sequences are the most variable, involving multiple scene changes (blue squares for the third scene) due to operator movement of the fibre bundle.

3.2.2 *Image Sequence Pre-processing*

The pre-processing stage of the TRACER pipeline is composed of two critical steps: the removal of uninformative frames and the enforcement of scene consistency within the sequence. Firstly, corrupted frames are detected and discarded using the feature engineering approach described by Perperidis *et al.* [1], which was previously discussed in Section 2.4.1.

Secondly, to ensure that the remaining frames belong to a single, stable scene, we analyse the consistency across the sequence of N intensity-weighted lifetime images, where each image is a matrix $\mathbf{I}_{\tau\ [n]}$ for $n = \{1, \ldots, N\}$. For this, we employ the quality of alignment (QA) metric, which evaluates the alignment quality between a given image $\mathbf{I}_{\tau\ [n]}$ and the subsequent image $\mathbf{I}_{\tau\ [n+1]}$ in the sequence [19]:

$$QA_{[K+1,K]} = \mathrm{NCC}(\mathbf{I}_{\tau\ [n]}, \mathbf{I}_{\tau\ [n+1]})\,, \tag{3.1}$$

here, QA represents NCC, determined by adapting the NCC function, which can be written as:

$$NCC(\mathbf{A}, \mathbf{B}) = \frac{\sum_{(x,y)\in\Omega} \left[(\mathbf{A}(x,y) - \mu_{\mathbf{A}})\,(\mathbf{B}(x,y) - \mu_{\mathbf{B}})\right]}{\sigma_{\mathbf{A}} \cdot \sigma_{\mathbf{B}}}\,, \tag{3.2}$$

where $\mathbf{A}$ and $\mathbf{B}$ correspond to the images $\mathbf{I}_{\tau\ [n]}$ and $\mathbf{I}_{\tau\ [n+1]}$, $\mathbf{A}(x,y)$ is the scalar value at pixel coordinates (x,y), Ω is the set of all pixel coordinates, and $\mu_{\cdot}$ and $\sigma_{\cdot}$ are the mean and standard deviation of the respective images. When multi-channel image representations (*e.g.,* RGB or RGBA encodings) are used elsewhere in the pipeline, the NCC operation is applied to scalar-valued feature channels rather than directly to vector-valued colour images. In practice, this corresponds to computing NCC on a single derived two-dimensional field (such as intensity, lifetime, or an intensity-weighted lifetime map), or equivalently aggregating per-channel NCC terms where required. The formulation in Eq. (3.2) therefore remains valid and consistent, as the correlation is always performed on scalar image representations, with RGB or RGBA formats serving only as organisational or visual encodings outside the core registration computation.

Insight into scene consistency is gained by monitoring the rate of change (RoC) of the QA metric between consecutive images. The RoC for an image at position n in the sequence is defined as:

$$\mathrm{RoC}_{[n]} = \frac{|\mathrm{QA}_{[n+1,n]} - \mathrm{QA}_{[n,n-1]}|}{\mathrm{QA}_{[n]}} \times 100\,\%\,. \tag{3.3}$$

A scene change is flagged by applying a threshold to this RoC. The sequence is partitioned at frame n if the absolute RoC value exceeds a predefined threshold, $\text{thresh}_{\text{RoC}}$:

$$|\text{RoC}_{[n]}| \geq \text{thresh}_{\text{RoC}}\,\%\,. \tag{3.4}$$

Based on empirical analysis of the RoC across the sequences (Fig. 3.3), a threshold value of $\text{thresh}_{\text{RoC}} \approx 30\,\%$ or greater was found to reliably correspond to a transition between different scenes, typically caused by significant operator movement of the imaging probe. This consistency check, illustrated in Fig. 3.4, ensures that any given sequence submitted for registration depicts a stable scene, thereby enhancing the reliability of all subsequent processing steps.

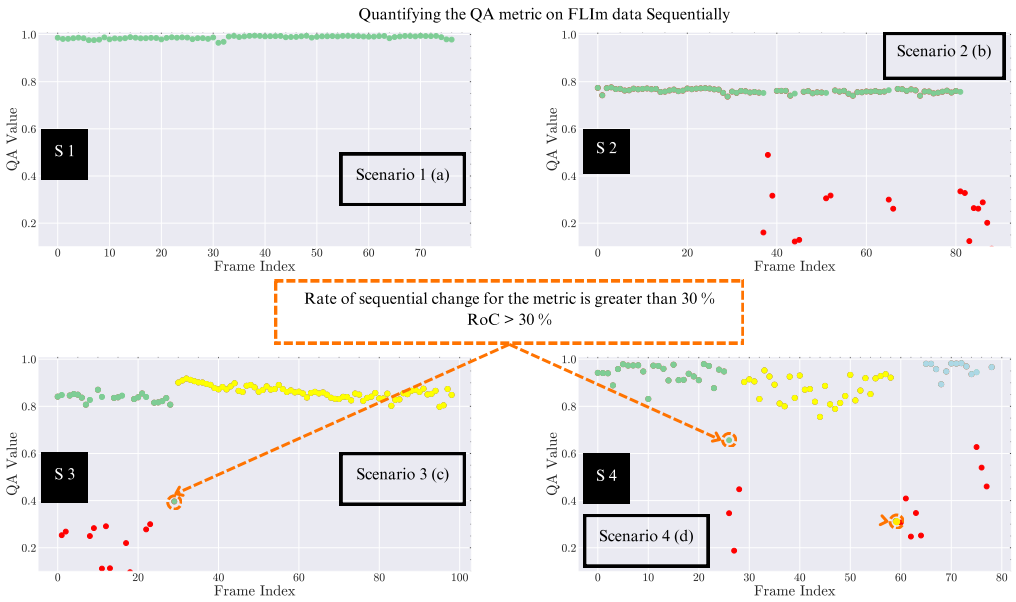


Figure 3.4: Sequential quantification of the QA metric for the FLIm image sequences from the four scenarios described in Section 3.2.1. (a) represents a homogeneous sequence as in scenario 1, where each point corresponds to the QA between consecutive frames. In scenario 2, with occasional uninformative frames, the plot appears as in (b). Red points indicate uninformative frames, which, when removed, allow for clearer detection of significant changes (RoC $\geq$ 30 %) in scenarios 3 and 4, shown in (c) and (d), respectively. This highlights transitions between distinct scenes or operator-induced movement.

3.2.3 *Image-to-image characterisation*

The next step in the TRACER pipeline involves characterising the motion between consecutive frames. Based on initial observations of the acquired FLIm data, it was assumed that rigid, translation-based registration would be sufficient.

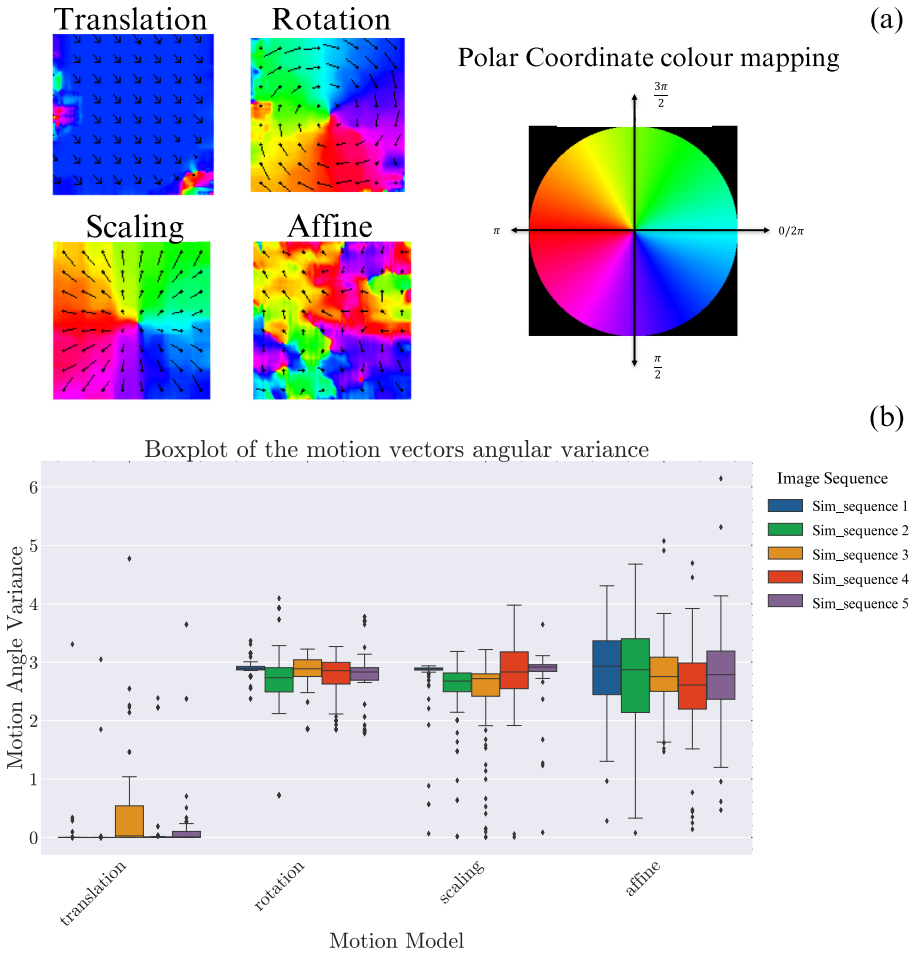


Figure 3.5: Evaluation of image-to-image motion characterisation using optical flow on simulated sequences. (a) Polar coordinate colour maps representing motion direction and magnitude: The angle of each motion vector is colour-coded according to the circular hue spectrum (0° mapped to red, 90° green, 180° cyan, and 270° purple), illustrating directionality, while magnitude (speed) of motion vectors is encoded by the length of the arrow representing motion vectors. (b) Angular variance of motion vectors across five simulated sequences, where translational motion shows the lowest angular variance, confirming consistent directional characteristics.

To validate this assumption quantitatively, we employed the robust two-frame motion estimation algorithm presented by Farnebäck [124]. This optical flow method characterises pixel displacement between images, which is crucial for understanding the underlying motion model prior to applying registration.

The algorithm functions by approximating the neighbourhood of each pixel with a quadratic polynomial via a polynomial expansion transform. Image motion is then efficiently characterised by observing how this polynomial representation transforms under various motion models. The displacement field between two images is estimated from the coefficients of these polynomial expansions. The process specifically involves:

1. Approximating each image neighbourhood with quadratic polynomials using the polynomial expansion transform.

2. Estimating the displacement field by analysing how the polynomial coefficients change due to motion.

3. Refining this estimate through iterative and multi-scale strategies to achieve a robust characterisation of motion.

The mathematical formulation for the polynomial expansion of a pixel's neighbourhood is given by:

$$\begin{aligned} f(x,y) \approx f(x_0,y_0) + \begin{bmatrix} x-x_0 & y-y_0 \end{bmatrix} \begin{bmatrix} f_x \\ f_y \end{bmatrix} \\ + \frac{1}{2} \begin{bmatrix} x-x_0 & y-y_0 \end{bmatrix} \begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix} \begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix}, \end{aligned} \tag{3.5}$$

where $f(x,y)$ represents the image intensity function $\mathbf{I}_{\tau\ [n]}(x,y)$, and (x_0,y_0) is the centre of the local neighbourhood. The terms

$$f_x = \left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0)}, \quad f_y = \left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0)},$$

$$f_{xx} = \left.\frac{\partial^2 f}{\partial x^2}\right|_{(x_0,y_0)}, \quad f_{xy} = \left.\frac{\partial^2 f}{\partial x \partial y}\right|_{(x_0,y_0)}, \quad f_{yy} = \left.\frac{\partial^2 f}{\partial y^2}\right|_{(x_0,y_0)},$$

denote the partial derivatives of f evaluated at (x_0,y_0), capturing the local image structure and variations resulting from motion.

From the polynomial expansion in Eq. 3.5, motion-induced changes in the image intensity between consecutive frames can be modelled by estimating the displacement vector $[x-x_0,\, y-y_0]$ for each pixel. By calculating the first-order

derivatives f_x and f_y, the intensity gradient along the x and y directions can be determined, forming the basis for computing the displacement field. The resulting motion vectors, derived from this displacement field, quantify the local movement of pixels and can be collectively represented as the vector field $\mathbf{V}$, where each vector $\mathbf{V}_i = (v_{x_i}, v_{y_i})$ encodes the pixel displacement in the x and y directions.

To facilitate interpretation of the vector field $\mathbf{V}$, polar coordinate colour maps are employed [125]. Fig 3.5 (a) illustrates these maps, wherein the angle θ_i of each vector $\mathbf{V}_i$ is mapped to hue on the colour wheel (indicating direction of motion), while the magnitude $\|\mathbf{V}_i\|$ is represented by the length of the corresponding motion vector arrow.

For a pure translational motion, all vectors in $\mathbf{V}$ should point in the same direction. To quantify the coherence of the motion model, we calculate the angular variance of the motion vectors, σ_θ^2:

$$\sigma_\theta^2 = \frac{1}{Z}\sum_{z=1}^{Z}(\theta_z - \bar{\theta})^2\,, \tag{3.6}$$

where θ_z is the angle (in degrees) of the z^{th} motion vector relative to a reference axis, $\bar{\theta}$ is the mean angle of all vectors in $\mathbf{V}$, and Z is the total number of motion vectors.

This equation directly relates to the vector field $\mathbf{V}$ by capturing the angular distribution of all vectors in the field. A lower σ_θ^2 indicates that the motion vectors are closely aligned, suggesting that the motion is well approximated by a translational model. Conversely, a higher σ_θ^2 signifies more diverse motion directions, indicative of non-translational or complex motion patterns.

The utility of σ_θ^2 as a metric is evident in Fig. 3.5(b), where angular variance values are shown for five simulated sequences subjected to different motion models, including translation, rotation, scaling, and affine transformations. As demonstrated, sequences with translational motion exhibit the lowest σ_θ^2, reflecting their uniform directional characteristics.

Moreover, the metric is applied to assess motion consistency in the six FLIm image sequences analysed in this study. Table 3.1 shows that the lower σ_θ^2 values align with the observations made during data acquisition, confirming the predominance of a translational motion model in these sequences.

3.2.4 *Image Registration*

This work focuses on similarity-based registration, an approach that omits an explicit feature detection step. Instead, it directly compares image regions by optimising a transformation function that maximises a chosen similarity metric between a reference and a moving image [26, 105]. The goal is to find the optimal

Table 3.1: Summary of motion vectors' angular variance across different imaging sequences.

Image sequences	**Mean Motion vectors σ_θ^2**
Simulation sequences	for translation* mostly $< 1.2°$
FLIm sequence 1	$0.18° \pm 0.13°$
FLIm sequence 2	$0.36° \pm 0.06°$
FLIm sequence 3	$0.29° \pm 0.05°$
FLIm sequence 4	$0.62° \pm 0.09°$
FLIm sequence 5	$0.41° \pm 0.13°$
FLIm sequence 6	$0.97° \pm 0.12°$

* See Fig. 3.5(b) for a visual representation of σ_θ^2 for simulated sequences.

transformation, T^*, that best aligns the moving image, $\mathbf{I}_{\tau\ [n+1]}$, with the reference image, $\mathbf{I}_{\tau\ [n]}$. At that point, Eq. (3.2) is used to include the spatial transformation operator T as $\mathrm{NCC}(\mathbf{I}_{\tau\ [n]}(\boldsymbol{\chi}), T(\mathbf{I}_{\tau\ [n+1]}))$, where, T is the transformation function mapping $\boldsymbol{\chi}$ to u, aligning $\mathbf{I}_{\tau\ [n+1]}$ with $\mathbf{I}_{\tau\ [n]}$.

$\mathbf{I}_\tau reg$ is defined as the composition, represented by $\circ$, of the optimal transformation function T^* with the moving image $\mathbf{I}_{\tau\ [n+1]}$, $\mathbf{I}_\tau reg = \mathbf{I}_{\tau\ [n+1]} \circ T^*$. Typically, T^* can be found by iteratively optimising a similarity function such as the NCC [105, 25] and can be written as:

$$T^* = \arg\max_T \mathrm{NCC}(\mathbf{I}_{\tau\ [n]}, T(\mathbf{I}_{\tau\ [n+1]})) \tag{3.7}$$

This can be solved using an optimisation algorithm, such as the Powell method [105], which is a widely used optimisation method that does not require the computation of gradients, which is advantageous when dealing with complex or non-differentiable objective functions, such as those often encountered in image registration. The image registration represented by Eq. 3.7 is referred to as NCC-General registration. Similarly, the ECC is a well-established, state-of-the-art, metric that can also be used. It is calculated as the correlation coefficient between the input and reference images, enhanced by a weighting function that prioritises pixels closer to the mean intensity, thereby increasing robustness to lighting changes [113]. Moreover, MMI is another approach that analyses the statistical dependence between the two images by comparing their joint pixel intensity histogram to their individual histograms and calculating a score based on the entropy of these histograms [114]. When well-aligned, both MMI and ECC scores are maximised, indicating greater similarity and shared information between the images. Furthermore, the study in [115] used OF to determine the optimal transformation for image alignment. The algorithm here approximates the pixel

displacement from the reference to the moving image, which is then used to compute a velocity field describing the apparent pixel motion. This field warps the moving image to align it with the reference image.

However, as noted in [100, 101], estimating the type of motion between images can enable direct image registration methods, eliminating the need for iterative solutions to determine the transformation function. That is, T in Eq. (3.7) can be restricted to linear translation parameters when considering a Translation motion model. We can also directly use the NCC offset between consecutive image pairs, performing a NCC-Translate image registration. Similarly, the ECC can also be restricted to Translation only.

Nonetheless, our empirical analysis revealed that the unique properties of FLIm, combined with significant displacements, present notable challenges for image registration. This is particularly true for optimisation processes, even when restricted to Translation. Moreover, methods involving the NCC become problematic when dealing with multiple peaks in the NCC space, complicating accurate offset computation.

3.3 ADDRESSING IMAGE REGISTRATION CHALLENGES

This section investigates the challenges of registering the image sequences described in Section 3.2.1 after we pre-process using the two steps mentioned in Section 3.2.2, and characterise the appropriate motion model.

3.3.0.1 *Multiple Peaks in Normalised Cross-Correlation*

Considering the direct method to achieve image registration and to understand the issue of multiple peaks in the NCC space, we first define the NCC 2D map $\mathbf{C}_{map}$. Utilising the Fourier domain variant of the NCC function provides a computationally efficient method for calculating and analysing the NCC $2D$ arrays, as described in [126]. Using this approach, Eq. (3.2) can be rewritten to compute $\mathbf{C}_{map}$ as follows:

$$\mathbf{C}_{map}(\mathbf{I}_{\tau\ [n]}, \mathbf{I}_{\tau\ [n+1]}) = \frac{\mathcal{F}^{-1}\left[\mathcal{F}(\mathbf{I}^{H}_{\tau\ [n+1]})^{*} \times \mathcal{F}(\mathbf{I}_{\tau\ [n]})\right]}{\sigma_{\mathbf{I}_{\tau\ [n]}} \cdot \sigma_{\mathbf{I}_{\tau\ [n+1]}}}, \tag{3.8}$$

where $\sigma_{\mathbf{I}_\tau\ [n]}$ and $\sigma_{\mathbf{I}_\tau\ [n+1]}$ are the standard deviation of the reference and moving images, respectively:

$$\begin{aligned}\sigma_{\mathbf{I}_\tau\ [n]} &= \sqrt{\sum_{(x,y)\in\Omega}\left(\mathbf{I}_{\tau\ [n]}(x,y) - \mu_{\mathbf{I}_\tau\ [n]}\right)^2},\\ \sigma_{\mathbf{I}_\tau\ [n+1]} &= \sqrt{\sum_{(x,y)\in\Omega}\left(\mathbf{I}_{\tau\ [n+1]}(x,y) - \mu_{\mathbf{I}_\tau\ [n+1]}\right)^2}\end{aligned} \tag{3.9}$$

$\mathcal{F}$ represents the Fast Fourier Transform, and $\mathcal{F}^{-1}$ is its inverse. $\mathbf{I}^H_{\tau\ [n+1]}$ denotes the horizontally flipped version of $\mathbf{I}_{\tau\ [n+1]}$, and $*$ indicates the complex conjugate, following the definition of cross-correlation rather than convolution [127].

Upon analysing the resulting $\mathbf{C}_{map}$ from FLIm sequences, a noisy NCC landscape with multiple peaks (Fig. 3.6.b) emerges due to edges and repetitive structures dominating the correlation space. This effect is evident in simulation sequences 1, 5, and 7, as these sequences produced multiple peaks when computing NCC between reference and moving images. Such complexities, primarily arising from repetitive patterns and large displacements, contribute significantly to registration challenges.

Fig. 3.6 demonstrates the impact of image complexity on the NCC space and highlights the necessity of a robust tracking mechanism to accurately identify the correct correlation peak. In simpler cases, such as sequence 1 (Fig. 3.6(a)), the $\mathbf{C}_{map}$ contains a single dominant peak, allowing straightforward identification of the correct translation between the reference and moving images. However, for more complex image sequences, as seen in Fig. 3.6(b) and Fig. 3.6(c), multiple peaks emerge in the $\mathbf{C}_{map}$, often due to repetitive structural elements or significant displacements, which complicates the accurate localisation of the peak representing the true transformation. Therefore, without a robust tracking mechanism, the identified peak (red square) may inaccurately reflect the motion, leading to errors in image registration.

To mitigate this, we propose integrating the CSRT [128] with the NCC for the image registration task. The CSRT algorithm enhances the traditional discriminative correlation filter (DCF) [129] by incorporating spatial and channel reliability mechanisms, enabling robust tracking even in complex scenarios with multiple similar features. Here, the CSRT is used to maintain consistent tracking of the correct peak (black square), even in the presence of multiple peaks, as shown in Fig. 3.6(d). This approach ensures the reliable computation of translation parameters, enhancing the robustness of the image registration process, especially in challenging conditions where the NCC landscape is noisy and contains several competing peaks.

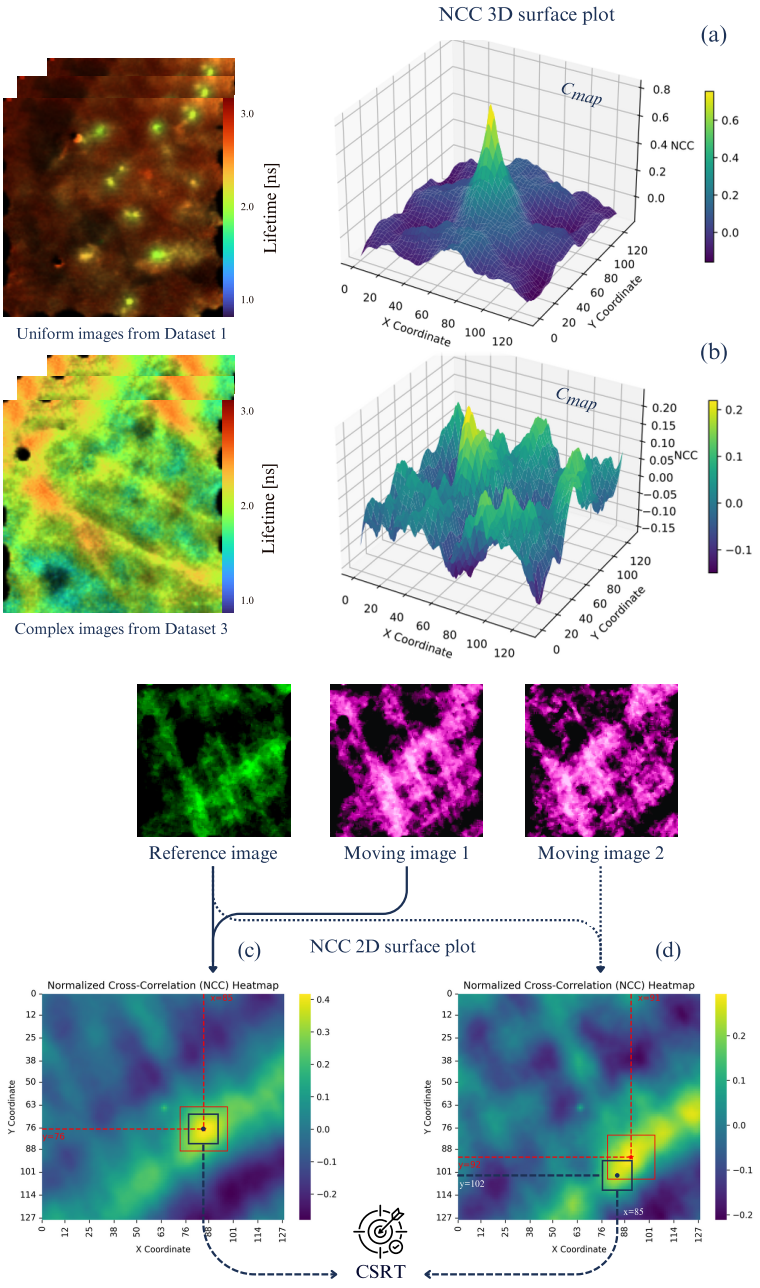


Figure 3.6: This figure illustrates the effect of image complexity on NCC and the resulting $\mathbf{C}_{map}$, along with the role of CSRT in addressing issues caused by multiple peaks. (a) shows a $\mathbf{C}_{map}$ with a single dominant peak, due to the low complexity in FLIm image sequence 1. In contrast, (b) and (c) display $\mathbf{C}_{map}$s from FLIm image sequences 2 and 3, where increased complexity results in multiple peaks. (c) also shows the 2D $\mathbf{C}_{map}$ used for tracking. When a single dominant peak exists, the identified peak (red square) coincides with the peak tracked by CSRT (black square). However, (d) illustrates the difficulty of identifying the correct peak in a $\mathbf{C}_{map}$ with multiple peaks, highlighting how CSRT ensures accurate tracking and localisation of the correct peak (black square) for precise computation of translation parameters.

3.3.1 *Feature Channels in the Context of FLIm Sequences*

In this approach, the feature channels are defined within the NCC correlation space rather than the image intensity or colour space. Whereas natural image processing operates in a spatial domain in which each pixel is associated with colour or intensity values (*e.g.,* RGB), the registration framework used here operates on correlation maps, in which each spatial location represents the strength of similarity between image patches.

Accordingly, the feature channels do not encode colour or semantic image content, but instead encode different measures of correlation reliability at each spatial location. These channels are extracted from the correlation map $\mathbf{C}_{map}$ and are used to characterise the local structure of the correlation surface, including the sharpness, consistency, and spatial stability of correlation peaks.

This reformulation replaces the conventional notion of "space + colour value" with "space + correlation value", allowing motion estimation to be driven by the structure of the similarity landscape rather than by appearance-based features. This distinction is particularly important for FLIm sequences, where intensity and lifetime values may vary independently of motion due to noise, biochemical heterogeneity, or acquisition variability.

3.3.2 *Channel and Spatial Reliability Peak Tracking Registration*

In practice, the correct correspondence between consecutive frames is strongly constrained by temporal and spatial proximity, as frame-to-frame motion in endoscopic FLIm sequences is typically limited. Empirically, this results in the true alignment peak appearing consistently near the previous frame's solution in the NCC map.

For image registration, the dominant peak in the NCC map $\mathbf{C}_{map}$, computed using Eq. 3.8, is tracked across frames to maintain alignment accuracy. Rather than selecting the global maximum independently at each frame, the tracking strategy prioritises continuity by following the correlation peak that remains spatially consistent with the previously identified alignment. Initially, $\mathbf{C}_{map}[0]$ is used to determine the location of the true dominant peak (i_0, j_0).

This approach ensures that the registration remains locked to the physically meaningful motion trajectory, even in cases where an alternative peak may temporarily exceed the tracked peak in magnitude and become the global maximum at later frames. Subsequently, the NCC is computed between a selected reference frame $\mathbf{I}_{\tau\ [n]}$ (typically the first frame in the sequence) and each moving image $\mathbf{I}_{\tau\ [n+1]}$ in a sequence of $N-1$ images, generating a series of NCC maps $\mathbf{C}_{map}[k]$, where $k \in \{1, 2, \ldots, N-1\}$. The objective is to track the peak representing

the correct alignment across these maps, despite the presence of multiple peaks caused by repetitive patterns or large displacements.

The CSRT learns a set of channel-specific DCFs to localise the dominant peak across frames. Let $c \in \{1, \ldots, C\}$ index the feature channels extracted from the correlation map $\mathbf{C}_{map}$, where C is the total number of channels. At the initial frame $\mathbf{C}_{map}[0]$, we extract feature channels $\{\xi_c\}_{c=1}^{C}$, while at subsequent frames $\mathbf{C}_{map}[k]$, we extract $\{\mathbf{x}_c\}_{c=1}^{C}$.

The CSRT formulates this tracking problem by learning a DCF for each channel, minimising the following objective:

$$\arg\min_{\{\mathrm{DCF}_c\}} \left\| \mathbf{S} \odot \left(\rho - \sum_{c=1}^{C} \xi_c * \mathrm{DCF}_c \right) \right\|^2 + \lambda \sum_{c=1}^{C} R_c \left\| \mathrm{DCF}_c \right\|^2, \tag{3.10}$$

where:

- ξ_c: the c-th feature channel extracted from $\mathbf{C}_{map}[0]$,
- DCF_c: the filter for the c-th channel,
- ρ: the desired response, typically a Gaussian centred at the target location,
- $\mathbf{S}$: the spatial reliability map, indicating the reliability of each region,
- R_c: the reliability map for the c-th channel,
- λ: a regularisation parameter,
- $\odot$: element-wise multiplication,
- $*$: convolution.

The spatial reliability map $\mathbf{S}$ focuses on the region around the dominant peak in $\mathbf{C}_{map}[0]$, and for subsequent frames $\mathbf{C}_{map}[k]$ is defined as:

$$\mathbf{S}(i, j) = \exp\left(-\frac{(i - i_0)^2 + (j - j_0)^2}{2\sigma_s^2} \right), \tag{3.11}$$

where (i_0, j_0) is the location of the dominant peak in $\mathbf{C}_{map}[0]$, and σ_s controls the spatial decay.

The channel reliability R_c determines how each channel contributes to the final response:

$$R_c = \exp\left(-\frac{\|\mathbf{d}_c\|^2}{2\sigma_c^2} \right), \tag{3.12}$$

where $\mathbf{d}_c$ represents the difference between the feature descriptors of the target in the current and previous $\mathbf{C}_{map}$ for the c-th channel, and σ_c controls how quickly the channel reliability decays. Solving Eq. (3.10) yields filters $\{\mathrm{DCF}_c\}$

that emphasise reliable channels and spatial regions, aiding the tracker in following the dominant peak across frames, even when multiple peaks exist.

Once the tracker is updated, the dominant peak for $\mathbf{C}_{map}[k]$ is obtained by:

$$(i_{\text{peak}}, j_{\text{peak}}) = \arg\max_{(i,j)} \left(\mathbf{S} \odot \sum_{c=1}^{C} \mathbf{x}_c * \text{DCF}_c \right)(i,j). \tag{3.13}$$

Here, $\mathbf{x}_c$ denotes the c-th feature channel extracted from $\mathbf{C}_{map}[k]$. Finally, the transformation T is determined using the offset

$$(\Delta i_k, \Delta j_k) = (i_{\text{peak}} - i_0, j_{\text{peak}} - j_0), \tag{3.14}$$

and the registered image is obtained as:

$$\mathbf{I}_\tau \text{reg}_{\ [n]} = \mathbf{I}_{\tau\ [n+1]} \circ T_{(\Delta i_k, \Delta j_k)}. \tag{3.15}$$

Our evaluation suggests that registration errors predominantly arise from the misidentification of the true optimal NCC peak, caused by the presence of competing secondary peaks with similar prominence, e.g. Fig. 3.6(b). Similarly, using the ECC method on complex image sequences confirmed convergence errors. Upon examining the 3D NCC map, multiple peaks were identified, complicating the registration process. By tracking the dominant peak using CSRT in the approach summarised in Algorithm 1, we achieve robust and efficient alignment without the need for iterative optimisation of similarity metrics, which can be particularly challenging for FLIm data due to image complexities.

Algorithm 1 CSRT for NCC-Based Image Registration

Require:
- A reference frame $\mathbf{I}_{\tau\ [n]}$.
- A sequence of $N-1$ moving frames $\{\mathbf{I}_{\tau\ [n+1]}, \mathbf{I}_{\tau\ [n+2]}, \ldots, \mathbf{I}_{\tau\ [n+N-1]}\}$.
- Regularisation parameter λ.
- Parameters σ_s, σ_c for spatial and channel reliability.

Ensure:
- Registered frames $\{\mathbf{I}_\tau\text{reg}_{\ [1]}, \mathbf{I}_\tau\text{reg}_{\ [2]}, \ldots, \mathbf{I}_\tau\text{reg}_{\ [N-1]}\}$.
- Estimated transformations $T_{(\Delta i_k, \Delta j_k)}$ for each frame $k \in \{1, N-1\}$.

1: **Compute** $\mathbf{C}_{map}[0]$ for the reference frame $\mathbf{I}_{\tau\ [n]}$ using Eq. (3.8).
2: **Identify** the location of true peak (i_0, j_0) in $\mathbf{C}_{map}[0]$.
3: **Construct feature channels** $\{\boldsymbol{\xi}_c\}_{c=1}^{C}$ from $\mathbf{C}_{map}[0]$, where C is the total number of channels.
4: **Define the spatial reliability map S** around the peak (i_0, j_0) using Eq. (3.11).
5: **Initialise channel reliability** R_c for each channel c as in Eq. (3.12) (often beginning with uniform reliability or basic channel difference measures).
6: **Learn the set of DCF filters** $\{\text{DCF}_c\}_{c=1}^{C}$ by minimising Eq. (3.10)
7: **for** $k = 1$ **to** $(N-1)$ **do**
8: **Compute** $\mathbf{C}_{map}[n]$ between $\mathbf{I}_{\tau\ [n]}$ and $\mathbf{I}_{\tau\ [n+1]}$ (Eq. 3.8).
9: **Extract feature channels** $\{\mathbf{x}_c\}_{c=1}^{C}$ from $\mathbf{C}_{map}[n+1]$.
10: **Update channel reliability** R_c using Eq. (3.12).
11: **Refine the DCF filters** if necessary by re-solving Eq. (3.10).
12: **Locate the new dominant peak** $(i_{\text{peak}}, j_{\text{peak}})$ using Eq. 3.13
13: **Compute the transformation offset** as per Eq. 3.14.
14: **Register the frame** $\mathbf{I}_{\tau\ [n]}$ by warping with the estimated transformation as per Eq. 3.15.
15: **end for**
16: **Output:** all registered images $\{\mathbf{I}_\tau\text{reg}_{\ [1]}, \ldots, \mathbf{I}_\tau\text{reg}_{\ [N-1]}\}$ and transformations $T_{(\Delta i_K, \Delta j_K)}$.

3.4 RESULTS

Recent studies, such as [25, 100], have highlighted the need to tailor these methods to the specific data and applications to ensure suitability. The TRACER pipeline was designed with this consideration in mind: by carefully characterising the motion model between images before the registration stage, we can select the most appropriate approach. Therefore, offering a comprehensive yet straightforward

solution for processing FLIm images to fuse temporal sequences, enabling the tracking and quantification of objects within the image scenes.

Our findings suggest that rigid transformation is effective for aligning image pairs within the TRACER pipeline. To demonstrate this, we compared the proposed registration approach in TRACER against the methods described in Section 3.2.4: NCC-Translate, NCC-General, ECC, MMI, and OF. An ablation study was conducted, evaluating registration performance with and without each step of the TRACER pipeline to assess the methods and the impact of the pre-processing steps. The evaluation was based on three key metrics: QA, structural similarity index measure (SSIM) [130], and normalised root mean squared error (NRMSE), and was conducted on all the image sequences described in Section 3.2.1.

3.4.1 *Quantitative Analysis*

The TRACER pipeline comprises two primary pre-processing steps, as detailed in Section 3.2.2. Step 1 involves the removal of uninformative frames—those that do not contribute to downstream analysis—while Step 2 ensures sequence consistency by detecting and addressing scene changes throughout the sequence. An ablation study was conducted, evaluating registration performance with and without each step of the TRACER pipeline to assess the methods and the impact of the pre-processing steps.

3.4.1.1 *Step 1 Ablation*

Tables 3.2, 3.3, 3.4 present the quantitative image registration results for the QA, SSIM, and NRMSE metrics on the simulated sequences before and after applying Step 1. As shown by the Tables 3.2, 3.3, 3.4, TRACER marginally outperforms other methods, with improvements of 6.5 % in QA, 7.6 % in SSIM, and 1.5 % in NRMSE. However, none of the registration methods perform as well without applying Step 1. After removing uninformative frames, QA improves by 4.3 %, SSIM by 5.6 %, and NRMSE by 27.6 % on average across all registration methods. Notably, TRACER is the only method capable of perfectly compensating for the small translational shifts observed across the initial five temporal sequences. In contrast, the ECC method succeeds in achieving this only within the context of simulation sequence 3. The errors observed in the NCC-Translate, NCC, and ECC methods are primarily due to incorrect localisation of the optimal solution. Additionally, localised registration errors in MMI and OF result in feature deformations within the images, negatively affecting metric evaluations.

When considering FLIm data, Tables 3.5, 3.6 show that for image sequences 1 and 2, Step 1 has no effect on registration performance. This is expected, as these

Table 3.2: Impact of step 1 on registration methods and TRACER, using simulation image sequences. Simulation sequences 1 to 3 (Part 1 of 3)

Data	Method	Without Step 1			With Step 1		
		QA ↑	SSIM ↑	NRMSE ↓	QA ↑	SSIM ↑	NRMSE ↓
Sim-seq 1	Unregistered	0.79 ± 0.22	0.21 ± 0.22	0.16 ± 0.21	0.86 ± 0.07	0.21 ± 0.23	0.10 ± 0.06
	NCC-Translate	0.55 ± 0.38	0.75 ± 0.24	0.21 ± 0.21	0.52 ± 0.37	0.82 ± 0.16	0.19 ± 0.17
	NCC-General	0.68 ± 0.16	0.41 ± 0.28	0.13 ± 0.17	0.86 ± 0.08	0.36 ± 0.26	0.09 ± 0.06
	ECC	**0.80** ± 0.19	0.74 ± 0.25	0.14 ± 0.23	0.86 ± 0.12	0.83 ± 0.15	0.04 ± 0.04
	MMI	0.60 ± 0.18	0.24 ± 0.24	0.21 ± 0.12	0.64 ± 0.12	0.22 ± 0.22	0.21 ± 0.13
	OF	0.77 ± 0.16	0.68 ± 0.23	0.13 ± 0.21	0.82 ± 0.11	0.75 ± 0.17	0.05 ± 0.04
	TRACER	**0.80** ± 0.19	**0.91** ± 0.19	**0.10** ± 0.18	**1.00** ± 0.00	**1.00** ± 0.00	**0.00** ± 0.00
Sim-seq 2	Unregistered	0.73 ± 0.19	0.43 ± 0.13	0.07 ± 0.08	0.76 ± 0.15	0.46 ± 0.09	0.06 ± 0.03
	NCC-Translate	0.61 ± 0.35	0.72 ± 0.26	0.12 ± 0.08	0.55 ± 0.34	0.78 ± 0.20	0.08 ± 0.07
	NCC-General	0.67 ± 0.12	0.71 ± 0.19	0.11 ± 0.04	0.92 ± 0.05	0.71 ± 0.20	0.03 ± 0.02
	ECC	**0.79** ± 0.25	**0.75** ± 0.28	**0.05** ± 0.07	0.84 ± 0.22	0.82 ± 0.24	0.03 ± 0.03
	MMI	0.64 ± 0.17	0.46 ± 0.11	0.12 ± 0.07	0.68 ± 0.14	0.46 ± 0.08	0.06 ± 0.04
	OF	0.69 ± 0.20	0.52 ± 0.21	0.06 ± 0.08	0.71 ± 0.17	0.56 ± 0.19	0.05 ± 0.03
	TRACER	0.75 ± 0.17	0.73 ± 0.20	0.06 ± 0.07	**1.00** ± 0.00	**1.00** ± 0.00	**0.00** ± 0.00
Sim-seq 3	Unregistered	0.84 ± 0.18	0.61 ± 0.17	0.09 ± 0.09	0.88 ± 0.12	0.65 ± 0.07	0.08 ± 0.04
	NCC-Translate	0.73 ± 0.23	0.74 ± 0.23	0.06 ± 0.07	0.74 ± 0.21	0.79 ± 0.19	0.06 ± 0.05
	NCC-General	0.77 ± 0.14	0.73 ± 0.17	0.06 ± 0.06	0.95 ± 0.05	0.75 ± 0.13	0.06 ± 0.04
	ECC	0.81 ± 0.16	0.79 ± 0.23	**0.04** ± 0.07	**1.00** ± 0.00	**1.00** ± 0.00	**0.00** ± 0.01
	MMI	0.82 ± 0.16	0.63 ± 0.17	0.09 ± 0.07	0.86 ± 0.11	0.66 ± 0.11	0.08 ± 0.05
	OF	0.88 ± 0.15	0.67 ± 0.20	0.07 ± 0.06	0.92 ± 0.05	0.72 ± 0.13	0.06 ± 0.04
	TRACER	**0.92** ± 0.17	**0.89** ± 0.21	0.05 ± 0.08	**1.00** ± 0.00	**1.00** ± 0.00	**0.00** ± 0.00

Table 3.3: Impact of step 1 on registration methods and TRACER, using simulation image sequences. Simulation sequences 3 to 6 (Part 2 of 3)

Data	Method	Without Step 1			With Step 1		
		QA ↑	SSIM ↑	NRMSE ↓	QA ↑	SSIM ↑	NRMSE ↓
Sim-seq 4	Unregistered	0.64 ± 0.21	0.50 ± 0.15	0.08 ± 0.09	0.67 ± 0.18	0.53 ± 0.10	0.06 ± 0.04
	NCC-Translate	0.80 ± 0.20	0.73 ± 0.27	0.05 ± 0.07	0.83 ± 0.16	0.79 ± 0.20	0.04 ± 0.03
	NCC-General	0.88 ± 0.15	0.85 ± 0.18	0.06 ± 0.05	0.97 ± 0.05	0.89 ± 0.14	0.02 ± 0.02
	ECC	0.72 ± 0.26	0.71 ± 0.27	0.06 ± 0.07	0.77 ± 0.24	0.78 ± 0.23	0.04 ± 0.04
	MMI	0.64 ± 0.21	0.47 ± 0.12	0.09 ± 0.10	0.66 ± 0.19	0.49 ± 0.06	0.07 ± 0.03
	OF	0.70 ± 0.18	0.57 ± 0.20	0.06 ± 0.06	0.73 ± 0.15	0.62 ± 0.16	0.05 ± 0.03
	TRACER	**0.90** ± 0.22	**0.90** ± 0.21	**0.04** ± 0.08	**1.00** ± 0.00	**1.00** ± 0.00	**0.00** ± 0.00
Sim-seq 5	Unregistered	0.86 ± 0.11	0.25 ± 0.14	0.22 ± 0.13	0.89 ± 0.08	0.27 ± 0.14	0.22 ± 0.13
	NCC-Translate	0.80 ± 0.20	0.74 ± 0.25	0.14 ± 0.13	0.81 ± 0.18	0.80 ± 0.19	0.13 ± 0.12
	NCC-General	0.82 ± 0.07	0.77 ± 0.24	0.09 ± 0.12	0.99 ± 0.01	0.84 ± 0.19	0.06 ± 0.08
	ECC	0.85 ± 0.12	0.81 ± 0.25	0.09 ± 0.12	0.98 ± 0.05	0.95 ± 0.17	0.05 ± 0.09
	MMI	0.85 ± 0.11	0.31 ± 0.19	0.20 ± 0.14	0.86 ± 0.09	0.31 ± 0.16	0.20 ± 0.14
	OF	0.78 ± 0.17	0.56 ± 0.25	0.13 ± 0.12	0.79 ± 0.16	0.61 ± 0.24	0.11 ± 0.09
	TRACER	**0.91** ± 0.10	**0.87** ± 0.19	**0.06** ± 0.10	**1.00** ± 0.00	**1.00** ± 0.00	**0.00** ± 0.00
Sim-seq 6	Unregistered	0.51 ± 0.23	0.42 ± 0.12	0.09 ± 0.10	0.54 ± 0.22	0.44 ± 0.02	0.08 ± 0.05
	NCC-Translate	0.52 ± 0.28	0.57 ± 0.29	0.10 ± 0.09	0.54 ± 0.32	0.56 ± 0.32	0.10 ± 0.08
	NCC-General	0.53 ± 0.19	0.58 ± 0.18	0.09 ± 0.06	0.56 ± 0.14	0.57 ± 0.17	0.18 ± 0.04
	ECC	0.76 ± 0.26	**0.72** ± 0.30	**0.06** ± 0.08	**0.77** ± 0.21	0.81 ± 0.24	**0.03** ± 0.03
	MMI	0.49 ± 0.20	0.44 ± 0.12	0.09 ± 0.09	0.49 ± 0.16	0.47 ± 0.06	0.08 ± 0.05
	OF	0.49 ± 0.21	0.46 ± 0.15	0.08 ± 0.06	0.52 ± 0.20	0.48 ± 0.09	0.07 ± 0.05
	TRACER	**0.78** ± 0.23	0.71 ± 0.20	**0.06** ± 0.08	0.72 ± 0.29	**0.82** ± 0.18	0.04 ± 0.04

Table 3.4: Impact of step 1 on registration methods and TRACER, using simulation image sequences. Simulation sequence 7 (Part 3 of 3)

Data	Method	Without Step 1			With Step 1		
		QA ↑	SSIM ↑	NRMSE ↓	QA ↑	SSIM ↑	NRMSE ↓
Sim-seq 7	Unregistered	0.50 ± 0.22	0.39 ± 0.16	0.18 ± 0.17	0.52 ± 0.21	0.41 ± 0.15	0.18 ± 0.17
	NCC-Translate	0.60 ± 0.27	0.59 ± 0.31	0.13 ± 0.13	0.63 ± 0.29	0.55 ± 0.34	0.13 ± 0.13
	NCC-General	0.62 ± 0.25	0.61 ± 0.18	0.12 ± 0.12	0.65 ± 0.25	0.59 ± 0.17	0.13 ± 0.12
	ECC	0.60 ± 0.27	0.67 ± 0.24	0.11 ± 0.12	0.64 ± 0.27	0.72 ± 0.23	0.11 ± 0.12
	MMI	0.48 ± 0.20	0.38 ± 0.13	0.18 ± 0.16	0.50 ± 0.20	0.40 ± 0.11	0.17 ± 0.16
	OF	0.54 ± 0.22	0.53 ± 0.16	**0.08** ± 0.07	0.56 ± 0.21	0.56 ± 0.14	**0.07** ± 0.05
	TRACER	**0.72** ± 0.34	**0.71** ± 0.25	0.15 ± 0.15	**0.71** ± 0.35	**0.76** ± 0.25	0.15 ± 0.15

sequences fall under scenario 1, where no uninformative frames are present. For the remaining sequences (3 to 6), OF generally performs the best, especially in sequences with multiple scenes. This is because, unlike the simulated sequences where scenes are entirely different, the FLIm images from different scenes are more similar in appearance. In such cases, deformative methods like OF are better suited to handle those subtle changes.

However, this observation holds before applying step 2 and primarily for QA. When looking at the SSIM metric, which is more sensitive to structural deformations [130, 131], OF falls behind other methods. The deformative nature of OF can introduce distortions that negatively impact the structural similarity between images, as reflected in the SSIM scores.

Table 3.5: Impact of step 1 on registration methods and TRACER, using FLIm image sequences. FLIm sequences 1 to 3 (Part 1 of 2)

Data	Method	Without Step 1			With Step 1		
		QA ↑	SSIM ↑	NRMSE ↓	QA ↑	SSIM ↑	NRMSE ↓
FLIm-seq 1	Unregistered	0.90 ± 0.04	0.64 ± 0.13	0.14 ± 0.10	0.90 ± 0.04	0.64 ± 0.13	0.14 ± 0.10
	NCC-Translate	0.91 ± 0.06	0.71 ± 0.17	0.13 ± 0.11	0.91 ± 0.06	0.71 ± 0.17	0.13 ± 0.11
	NCC-General	0.91 ± 0.03	0.60 ± 0.12	0.14 ± 0.09	0.91 ± 0.03	0.60 ± 0.12	0.14 ± 0.09
	ECC	0.92 ± 0.03	0.69 ± 0.10	0.13 ± 0.08	0.92 ± 0.03	0.69 ± 0.10	0.13 ± 0.08
	MMI	0.92 ± 0.03	0.69 ± 0.10	0.13 ± 0.09	0.92 ± 0.03	0.69 ± 0.10	0.13 ± 0.09
	OF	0.96 ± 0.02	0.76 ± 0.09	0.08 ± 0.06	0.96 ± 0.02	0.76 ± 0.09	0.08 ± 0.06
	TRACER	**0.98** ± 0.01	**0.91** ± 0.03	**0.06** ± 0.04	**0.98** ± 0.01	**0.91** ± 0.03	**0.06** ± 0.04
FLIm-seq 2	Unregistered	0.92 ± 0.04	0.73 ± 0.06	0.09 ± 0.06	0.92 ± 0.04	0.73 ± 0.06	0.09 ± 0.06
	NCC-Translate	0.93 ± 0.05	0.78 ± 0.11	0.08 ± 0.07	0.93 ± 0.05	0.78 ± 0.11	0.08 ± 0.07
	NCC-General	0.92 ± 0.04	0.76 ± 0.10	0.10 ± 0.08	0.92 ± 0.04	0.76 ± 0.10	0.10 ± 0.08
	ECC	0.92 ± 0.04	0.77 ± 0.06	0.08 ± 0.05	0.92 ± 0.04	0.77 ± 0.06	0.08 ± 0.05
	MMI	0.91 ± 0.04	0.73 ± 0.07	0.09 ± 0.06	0.91 ± 0.04	0.73 ± 0.07	0.09 ± 0.06
	OF	**0.98** ± 0.01	0.82 ± 0.04	0.05 ± 0.03	**0.98** ± 0.01	0.82 ± 0.04	0.05 ± 0.03
	TRACER	**0.98** ± 0.01	**0.94** ± 0.02	**0.04** ± 0.03	**0.98** ± 0.01	**0.94** ± 0.02	**0.04** ± 0.03
FLIm-seq 3	Unregistered	0.68 ± 0.18	0.31 ± 0.11	0.19 ± 0.20	0.75 ± 0.01	0.35 ± 0.02	0.15 ± 0.04
	NCC-Translate	0.74 ± 0.20	0.45 ± 0.25	0.18 ± 0.22	0.81 ± 0.09	0.50 ± 0.22	0.13 ± 0.09
	NCC-General	0.77 ± 0.05	0.38 ± 0.16	0.16 ± 0.10	0.77 ± 0.01	0.33 ± 0.07	0.14 ± 0.06
	ECC	0.76 ± 0.12	0.36 ± 0.08	0.15 ± 0.15	0.78 ± 0.02	0.38 ± 0.03	0.13 ± 0.05
	MMI	0.71 ± 0.19	0.29 ± 0.11	0.18 ± 0.20	0.78 ± 0.02	0.34 ± 0.06	0.14 ± 0.06
	OF	**0.79** ± 0.17	0.47 ± 0.11	**0.15** ± 0.17	0.85 ± 0.01	0.51 ± 0.02	0.11 ± 0.03
	TRACER	0.78 ± 0.18	**0.53** ± 0.25	0.15 ± 0.18	**0.92** ± 0.00	**0.83** ± 0.01	**0.09** ± 0.02

3.4.1.2 *Step 2 Ablation*

Table 3.7 compares the performance of different image registration methods across the QA, SSIM, and NRMSE metrics before and after applying Step 2 for sequences that require this step, such as simulation sequences 6 and 7. As shown in Table 3.7, registration performance struggles with these sequences compared to those benefiting only from Step 1 (simulation sequences 1 to 5). However,

Table 3.6: Impact of step 1 on registration methods and TRACER, using FLIm image sequences. FLIm sequences 4 to 6 (Part 2 of 2)

Data	Method	Without Step 1			With Step 1		
		QA ↑	SSIM ↑	NRMSE ↓	QA ↑	SSIM ↑	NRMSE ↓
FLIm-seq 4	Unregistered	0.55 ± 0.19	0.38 ± 0.15	0.23 ± 0.19	0.58 ± 0.16	0.40 ± 0.11	0.21 ± 0.14
	NCC-Translate	0.66 ± 0.21	0.52 ± 0.23	0.20 ± 0.16	0.67 ± 0.21	0.54 ± 0.23	0.19 ± 0.16
	NCC-General	0.71 ± 0.07	0.53 ± 0.17	0.20 ± 0.11	0.72 ± 0.07	0.51 ± 0.16	0.19 ± 0.10
	ECC	0.58 ± 0.16	0.39 ± 0.12	0.21 ± 0.15	0.59 ± 0.16	0.39 ± 0.11	0.21 ± 0.13
	MMI	0.55 ± 0.19	0.31 ± 0.14	0.23 ± 0.18	0.58 ± 0.16	0.35 ± 0.12	0.22 ± 0.14
	OF	**0.82** ± 0.12	0.54 ± 0.11	**0.13** ± 0.11	**0.85** ± 0.04	0.56 ± 0.08	**0.12** ± 0.06
	TRACER	0.69 ± 0.21	**0.56** ± 0.21	0.19 ± 0.16	0.79 ± 0.03	**0.57** ± 0.05	0.13 ± 0.05
FLIm-seq 5	Unregistered	0.62 ± 0.23	0.33 ± 0.22	0.25 ± 0.21	0.63 ± 0.22	0.33 ± 0.22	0.24 ± 0.21
	NCC-Translate	0.77 ± 0.14	0.49 ± 0.23	0.18 ± 0.14	0.77 ± 0.14	0.50 ± 0.23	0.18 ± 0.14
	NCC-General	0.78 ± 0.09	0.55 ± 0.22	0.17 ± 0.11	0.87 ± 0.09	0.55 ± 0.22	0.17 ± 0.11
	ECC	0.71 ± 0.14	0.43 ± 0.14	0.22 ± 0.18	0.72 ± 0.14	0.43 ± 0.14	0.22 ± 0.18
	MMI	0.64 ± 0.23	0.33 ± 0.20	0.24 ± 0.20	0.65 ± 0.23	0.33 ± 0.20	0.23 ± 0.20
	OF	**0.86** ± 0.11	0.53 ± 0.17	**0.13** ± 0.11	**0.87** ± 0.10	0.53 ± 0.16	0.13 ± 0.11
	TRACER	0.78 ± 0.09	**0.55** ± 0.14	0.15 ± 0.13	0.83 ± 0.02	**0.61** ± 0.06	**0.13** ± 0.05
FLIm-seq 6	Unregistered	0.62 ± 0.21	0.33 ± 0.16	0.33 ± 0.26	0.65 ± 0.21	0.36 ± 0.13	0.32 ± 0.26
	NCC-Translate	0.73 ± 0.23	0.53 ± 0.24	0.28 ± 0.26	0.77 ± 0.23	0.59 ± 0.22	0.26 ± 0.26
	NCC-General	0.72 ± 0.05	0.62 ± 0.19	0.18 ± 0.13	0.78 ± 0.04	0.62 ± 0.19	0.17 ± 0.10
	ECC	0.73 ± 0.17	0.47 ± 0.09	0.27 ± 0.21	0.73 ± 0.17	0.48 ± 0.08	0.27 ± 0.21
	MMI	0.67 ± 0.18	0.39 ± 0.13	0.29 ± 0.21	0.70 ± 0.16	0.41 ± 0.10	0.28 ± 0.21
	OF	**0.89** ± 0.09	0.56 ± 0.13	0.17 ± 0.16	**0.92** ± 0.05	0.60 ± 0.06	0.15 ± 0.12
	TRACER	0.77 ± 0.11	**0.68** ± 0.17	**0.15** ± 0.18	0.79 ± 0.02	**0.69** ± 0.05	**0.13** ± 0.10

after applying Step 2, registration performance improves significantly across all methods—by 15.6 % in QA, 30.5 % in SSIM, and 19.4 % in NRMSE. TRACER outperforms other methods, leading by 3.0 % in QA, 10.5 % in SSIM, and 2.5 % in NRMSE over the next best registration approach.

Table 3.7: Impact of step 2 on registration methods and TRACER, using simulation image sequences.

Data	Method	After step 1, without step 2			After step 1 and 2		
		QA ↑	SSIM ↑	NRMSE ↓	QA ↑	SSIM ↑	NRMSE ↓
Sim-seq 6	Unregistered	0.54 ± 0.22	0.44 ± 0.02	0.08 ± 0.05	0.72 ± 0.18	0.52 ± 0.05	0.08 ± 0.04
	NCC-Translate	0.54 ± 0.32	0.56 ± 0.32	0.10 ± 0.08	0.78 ± 0.29	0.78 ± 0.29	0.06 ± 0.08
	NCC-General	0.61 ± 0.14	0.68 ± 0.17	0.07 ± 0.04	0.86 ± 0.10	0.74 ± 0.17	0.05 ± 0.04
	ECC	**0.83** ± 0.21	0.81 ± 0.24	0.03 ± **0.03**	0.89 ± 0.16	0.88 ± 0.17	0.03 ± 0.04
	MMI	0.49 ± 0.16	0.47 ± 0.06	0.08 ± 0.05	0.66 ± 0.16	0.54 ± 0.08	0.08 ± 0.06
	OF	0.52 ± 0.20	0.48 ± 0.09	0.07 ± 0.05	0.70 ± 0.18	0.60 ± 0.12	0.06 ± 0.04
	TRACER	0.72 ± 0.29	**0.82** ± 0.18	0.04 ± 0.04	**0.98** ± 0.05	**0.98** ± 0.04	**0.01** ± 0.02
Sim-seq 7	Unregistered	0.52 ± 0.21	0.41 ± 0.15	0.18 ± 0.17	0.70 ± 0.18	0.39 ± 0.14	0.16 ± 0.10
	NCC-Translate	0.63 ± 0.29	0.55 ± 0.34	0.13 ± 0.13	0.84 ± 0.22	0.78 ± 0.28	0.09 ± 0.11
	NCC-General	0.63 ± 0.25	0.65 ± 0.17	0.15 ± 0.12	0.94 ± 0.10	0.86 ± 0.20	0.07 ± 0.08
	ECC	0.64 ± 0.27	0.72 ± 0.23	0.11 ± 0.12	0.88 ± 0.16	0.86 ± 0.22	0.06 ± 0.08
	MMI	0.50 ± 0.20	0.40 ± 0.11	0.17 ± 0.16	0.67 ± 0.18	0.40 ± 0.13	0.16 ± 0.11
	OF	0.56 ± 0.21	0.56 ± 0.14	0.07 ± 0.05	0.67 ± 0.19	0.58 ± 0.19	0.10 ± 0.08
	TRACER	**0.65** ± 0.35	**0.76** ± 0.25	**0.12** ± 0.15	**0.98** ± 0.06	**0.97** ± 0.05	**0.03** ± 0.04

3.4.1.3 *Ablation Summary*

As shown in Fig. 3.7, for simulation sequences, before Step 1 (Unprocessed), no registration method achieves optimal performance without Step 1. However, after applying step 1 and removing uninformative frames, QA improves by 4.3 %, SSIM by 5.6 %, and NRMSE by 2.6 % on average across all registration methods. Nonetheless, TRACER trails NCC-General and ECC due to tracking errors introduced by new scenes, underscoring the necessity of Step 2. For FLIm data, OF generally performs best, particularly in sequences containing multiple scenes. This is because, unlike simulated sequences—where scene differences are pronounced and localised registration errors in MMI and OF lead to feature deformations negatively impacting metrics—FLIm images from different scenes exhibit greater similarity in appearance. Consequently, deformative methods like OF are more adept at addressing these subtle variations (Tables 3.2, 3.3, and 3.4 for simulation data, and Tables 3.5 and 3.6 for FLIm data).

Notably, after applying Step 2, registration performance improves significantly across all methods, with increases of 15.6 % in QA, 30.5 % in SSIM, and 19.4 % in NRMSE (Table 3.7 and Table 3.8). Overall, the TRACER pipeline enhances registration performance by up to 50 % on average across all registration methods

Table 3.8: Impact of step 2 on registration methods and TRACER, using FLIm image sequences.

Data	Method	After step 1, without step 2			After step 1 and 2		
		QA ↑	SSIM ↑	NRMSE ↓	QA ↑	SSIM ↑	NRMSE ↓
FLIm-seq 4	Unregistered	0.58 ± 0.16	0.40 ± 0.11	0.21 ± 0.14	0.60 ± 0.11	0.40 ± 0.09	0.20 ± 0.11
	NCC-Translate	0.67 ± 0.21	0.54 ± 0.23	0.19 ± 0.16	0.76 ± 0.16	0.66 ± 0.17	0.14 ± 0.12
	NCC-General	0.72 ± 0.07	0.51 ± 0.16	0.19 ± 0.10	0.76 ± 0.07	0.50 ± 0.16	0.18 ± 0.10
	ECC	0.59 ± 0.16	0.39 ± 0.11	0.21 ± 0.13	0.64 ± 0.11	0.37 ± 0.11	0.19 ± 0.10
	MMI	0.58 ± 0.16	0.35 ± 0.12	0.22 ± 0.14	0.60 ± 0.11	0.34 ± 0.11	0.20 ± 0.11
	OF	**0.85** ± 0.04	0.56 ± 0.08	**0.12** ± 0.06	0.86 ± 0.03	0.56 ± 0.07	0.12 ± 0.06
	TRACER	0.79 ± 0.03	**0.57** ± 0.05	0.13 ± 0.05	**0.93** ± 0.02	**0.82** ± 0.04	**0.08** ± 0.04
FLIm-seq 5	Unregistered	0.63 ± 0.22	0.38 ± 0.22	0.24 ± 0.21	0.94 ± 0.02	0.75 ± 0.04	0.08 ± 0.05
	NCC-Translate	0.77 ± 0.14	0.50 ± 0.23	0.18 ± 0.14	0.97 ± 0.02	0.89 ± 0.07	0.06 ± 0.05
	NCC-General	0.77 ± 0.09	0.58 ± 0.22	0.17 ± 0.11	0.94 ± 0.01	0.71 ± 0.07	0.10 ± 0.06
	ECC	0.72 ± 0.14	0.43 ± 0.14	0.22 ± 0.18	0.95 ± 0.01	0.78 ± 0.04	0.08 ± 0.04
	MMI	0.65 ± 0.23	0.37 ± 0.20	0.23 ± 0.20	0.95 ± 0.01	0.73 ± 0.06	0.09 ± 0.05
	OF	**0.87** ± 0.10	0.53 ± 0.16	**0.13** ± 0.11	0.97 ± 0.01	0.85 ± 0.03	0.05 ± 0.03
	TRACER	0.83 ± 0.02	**0.61** ± 0.06	0.17 ± 0.05	**0.98** ± 0.01	**0.89** ± 0.01	**0.05** ± 0.02
FLIm-seq 6	Unregistered	0.65 ± 0.21	0.36 ± 0.13	0.32 ± 0.26	0.78 ± 0.09	0.47 ± 0.10	0.20 ± 0.12
	NCC-Translate	0.77 ± 0.23	0.59 ± 0.22	0.26 ± 0.26	0.89 ± 0.06	0.79 ± 0.08	0.15 ± 0.09
	NCC-General	0.79 ± 0.04	0.68 ± 0.19	0.15 ± 0.10	0.91 ± 0.04	0.65 ± 0.11	0.14 ± 0.09
	ECC	0.73 ± 0.17	0.48 ± 0.08	0.27 ± 0.21	0.83 ± 0.08	0.54 ± 0.08	0.17 ± 0.11
	MMI	0.70 ± 0.16	0.41 ± 0.10	0.28 ± 0.21	0.80 ± 0.08	0.46 ± 0.08	0.19 ± 0.12
	OF	**0.92** ± 0.05	0.60 ± 0.06	0.15 ± 0.12	0.94 ± 0.01	0.68 ± 0.05	**0.09** ± 0.05
	TRACER	0.79 ± 0.02	**0.69** ± 0.05	**0.13** ± 0.10	**0.95** ± 0.02	**0.85** ± 0.04	**0.09** ± 0.05

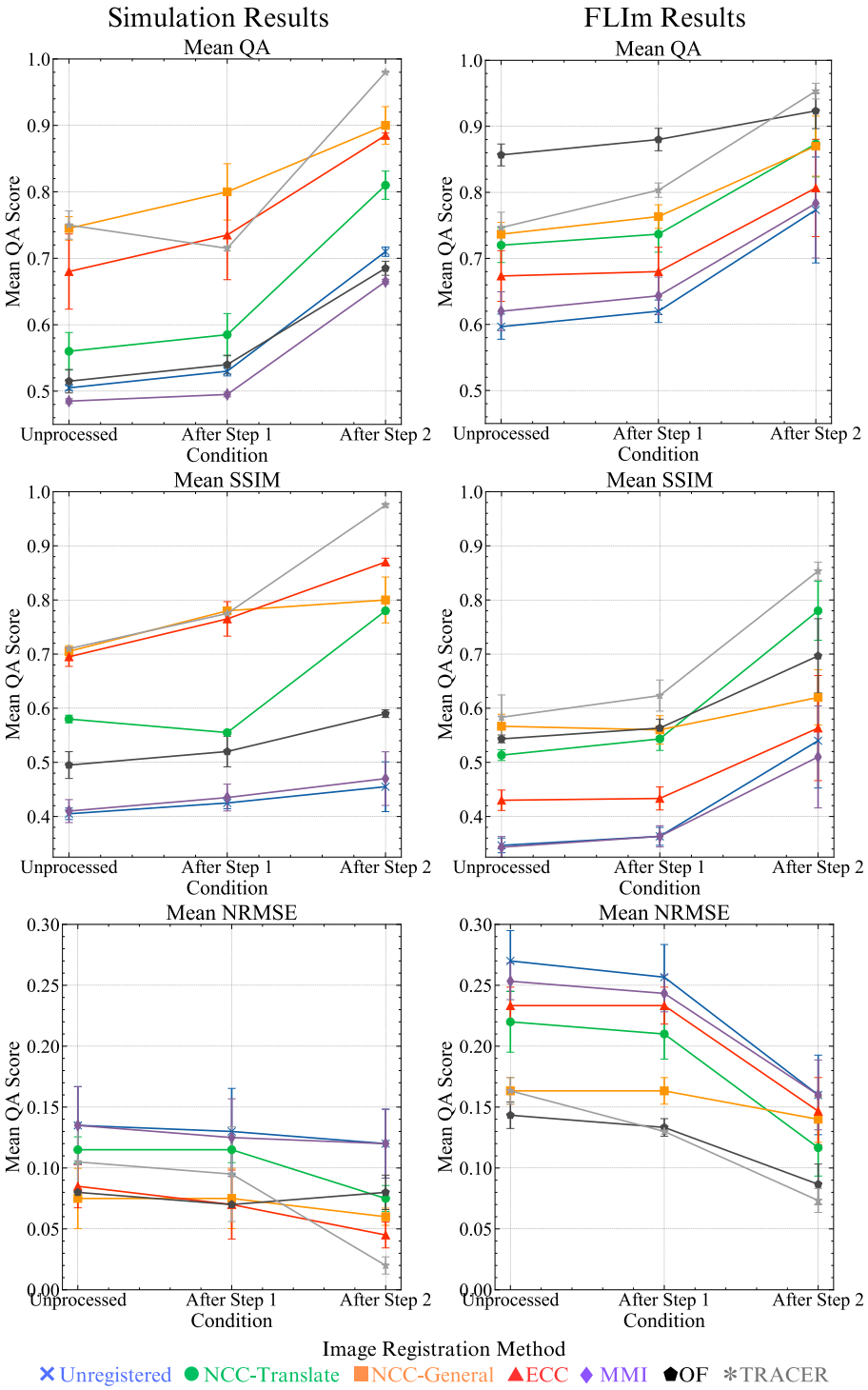


Figure 3.7: This figure illustrates the effect of the two primary pre-processing steps (1 and 2) described in Section 3.2.2 on the QA, SSIM, and NRMSE metrics. The error bars on the left represent the progressive improvements in the average metric values for the simulation sequences following each step, while those on the right display the corresponding results for the FLIm datasets.

and metrics. The proposed registration approach consistently outperforms state-of-the-art methods, achieving average improvements of 5 % in QA, 8 % in SSIM, and 3 % in NRMSE when both Steps are applied.

3.4.2 *Statistical Assessment of Registration Performance*

Table 3.9 summarises the statistical analysis of the mean improvements for QA, SSIM, and NRMSE across FLIm sequences 4, 5, and 6, before and after implementing Steps 1 and 2 of the TRACER pipeline. The analysis was performed with paired t-tests using the SciPy stats library [132] to determine whether the observed performance gains were statistically significant at the $p < 0.05$ level.

From Table 3.9, most registration methods exhibit clear and statistically significant improvements in all three metrics (QA, SSIM, NRMSE) once the TRACER pipeline's pre-processing is applied. For instance, the Unregistered case improves from a mean QA of 0.513 to 0.781, with $p = 0.012$, indicating that the simple act of removing uninformative frames and ensuring scene consistency yields a notable benefit before any actual image registration is performed. Similar significant gains are observed for NCC-General ($p = 0.000$), ECC ($p = 0.015$), MMI ($p = 0.012$), and OF ($p = 0.005$), confirming that the pipeline contributes substantially to improving alignment quality and structural fidelity.

Table 3.9: Paired t-test results comparing registration methods before and after applying Step 1 and 2. Significant differences ($p < 0.05$) indicated with *.

Method	Before using the proposed pipeline			After Steps 1 & 2			Paired t-test results		
	QA ↑	SSIM ↑	NRMSE ↓	QA ↑	SSIM ↑	NRMSE ↓	QA (p)	SSIM (p)	NRMSE (p)
Unregistered	0.513	0.374	0.173	0.781	0.539	0.135	0.012*	0.012*	0.012*
NCC-Direct	0.701	0.536	0.204	0.792	0.761	0.113	0.411	0.411	0.411
NCC-General	0.704	0.516	0.243	0.916	0.639	0.166	0.000*	0.000*	0.000*
ECC	0.571	0.410	0.217	0.780	0.548	0.198	0.015*	0.015*	0.015*
MMI	0.547	0.207	0.342	0.803	0.511	0.280	0.012*	0.012*	0.012*
OF	0.839	0.566	0.178	0.943	0.678	0.059	0.005*	0.005*	0.005*
TRACER	0.668	0.506	0.157	0.940	0.850	0.092	0.001*	0.001*	0.001*

A notable exception is NCC-Direct, whose increases (e.g. from QA=0.701 to 0.792) did not meet the threshold for statistical significance ($p = 0.411$). Closer inspection suggests that although some numerical gains are observed, they do not consistently outweigh the within-sequence variation when assessed by a paired t-test. This highlights that while the pipeline can facilitate better registration for most methods, its overall impact may vary when a method's core function has limited robustness against multiple peaks or intensity variations in certain FLIm frames.

Among all methods, our proposed TRACER approach shows particularly strong statistical significance ($p = 0.001$), with QA rising from 0.668 to 0.940 and

SSIM from 0.506 to 0.850. This confirms that TRACER's dedicated peak-tracking strategy, coupled with the prior removal of uninformative frames and splitting of inconsistent scene segments, yields the largest and most consistent benefits for rigid registration in FLIm data. Consequently, these results illustrate that carefully designed pre-processing steps can substantially aid most registration algorithms, and the final TRACER registration stage provides an additional, statistically validated improvement beyond those of standard methods.

It is important to note, the performance metrics reported in this chapter should be interpreted as indicators of practical robustness rather than absolute physical accuracy. The aim is to reduce motion-induced variability to a level at which temporal averaging, feature extraction, and subsequent learning-based analysis can operate reliably.

3.4.3 *Computational Runtime*

While both OF and MMI methods successfully achieved temporal registration, OF proved to be computationally intensive and lacked precision in cases where the deformative nature of the method was not needed. Moreover, the MMI suffered from sensitivity to initial conditions and susceptibility to local optima due to its reliance on gradient descent for maximising mutual information between image pairs [133].

Focusing the motion model on translational movements, primarily driven by natural respiratory motions, NCC and ECC demonstrated comparable or, in some cases, more accurate results. By using the NCC-Translate method to compute registration parameters, we were able to mitigate the computational complexity commonly associated with iterative optimisation in image registration, as seen with NCC.

The TRACER registration approach, on the other hand, maintains high accuracy and offers a computationally efficient solution, achieving about an order of magnitude faster performance than the next most accurate registration method. The computational runtime for registering each pair of images was measured using an Intel Core i7-6700 CPU with 8 GB of RAM, see Fig 3.8 for details.

Balancing runtime efficiency and accuracy is particularly important in real-time medical imaging, not just to improve speed but also to lower hardware costs, making devices more accessible. Additionally, reducing computational demands contributes to minimising environmental impact, a growing priority for future technologies.

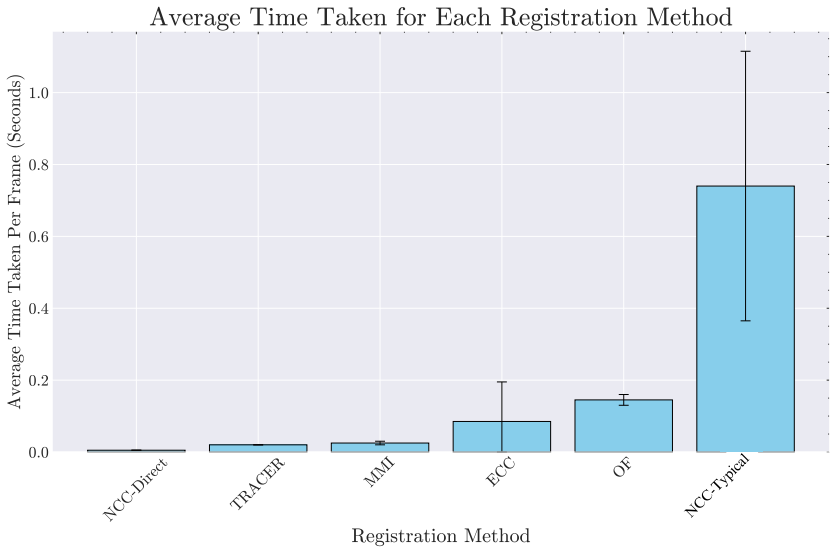


Figure 3.8: Comparison of the average time required to register a pair of images using the six registration methods benchmarked in this study, applied across all simulation and FLIm imaging datasets.

3.4.4 *Qualitative Example*

Figure 3.9 illustrates the fusing of a pair of reference and moving images from FLIm image sequence 3, shown in green and magenta, respectively. A perfect alignment results in a grayscale image, indicating precise alignment, whereas misalignment is characterised by green and magenta discrepancies, as demonstrated in Fig. 3.9.a. Additionally, Fig. 3.9 presents registration results fused with the reference image, scored by the QA. The results were obtained using the proposed TRACER registration approach alongside the other methods mentioned earlier. Notably, the TRACER method exhibits superior registration accuracy in Fig. 3.9.f, outperforming the other methods based on the quality of the fused image.

3.4.5 *Application Example to Segmentation and Detection*

To further verify the efficacy of the image registration task, this segmentation and detection experiment aimed to evaluate the performance of the detection of NAP signals [17] on the resulting fused images. A clinical expert provided a ground truth binary mask, shown in Fig. 3.10.b, highlighting areas of positive NAP signal throughout the image sequence for benchmarking purposes.

The detection methodology entails the application of a predefined threshold to segregate positive NAP signals within the FLIm imaging, as delineated by criteria established in prior studies [8, 17]. Positive regions manifest as distinct blobs against the FLIm image contrast, exemplified in Fig. 3.10.a. This method resulted in the generation of binary masks from fused images derived through the various

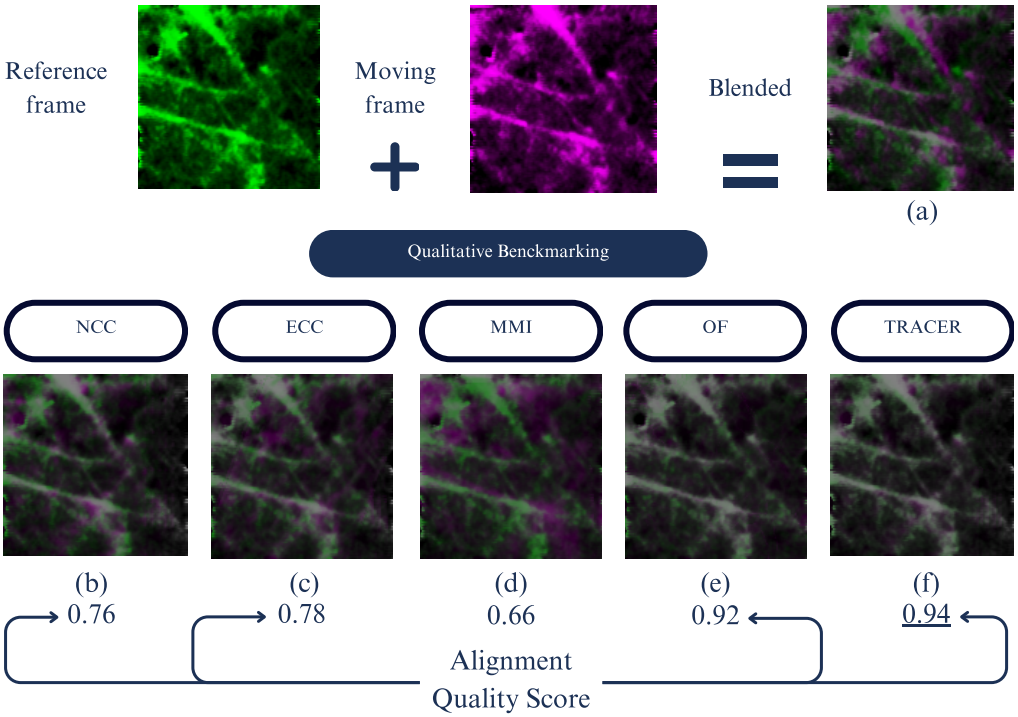


Figure 3.9: Evaluation of registration outcomes. The reference and moving images are represented in green and magenta, respectively. Perfect alignment results in a grayscale image, as opposed to the misalignment depicted in (a). Sub-chapter3 (b) to (f) showcase the most successful alignments from the benchmarking experiment: (b) and (c) display results using NCC (same result as NCC-Translate) and ECC, respectively; (d) and (e) present outcomes via MMI and OF, respectively; and (f) illustrates the alignment achieved with TRACER.

registration techniques, as shown in Fig. 3.10.c. The efficacy of each registration method was quantitatively assessed, and results are shown in Table 3.10 using precision, recall, and F1-score (also known as the Dice coefficient or DICE) measures, widely utilised metrics for assessing the accuracy of object detection algorithms [134].

Precision represents the ratio of the true positive predictions to the total positive predictions and is defined as:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN} \tag{3.16}$$

where TP denotes the true positives, FP represents the false positives, and FN denotes the false negatives. Recall is the ratio of the true positive predictions to the total number of relevant object predictions. The true positives (TP), false positives (FP), and false negatives (FN) are derived from comparing the ground truth binary mask with the binary mask generated by the detection method. This comparison is executed as follows:

- TP are pixels where both the predicted mask and the ground truth mask agree on a positive NAP signal, represented as pixels with a value of 1 in both masks.
- FP are pixels identified as positive signals in the predicted mask but are not present in the ground truth mask.
- FN represents the missed positive signals, where the predicted mask shows no signal despite the presence of a signal in the ground truth mask.

Table 3.11 summarises the statistical comparison of the proposed TRACER method against other registration techniques, specifically focusing on Precision, Recall, and the F1-score metrics. Precision alone, while informative, primarily reflects false positives (FP) and does not account for false negatives (FN). Thus, Recall, which accounts for FN, and the F1-Score, the harmonic mean of Precision and Recall, provide a more comprehensive assessment of detection performance:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.17}$$

This metric effectively summarises the overall accuracy, balancing both false positives and false negatives. Statistical analysis (Table 3.11) demonstrates that the proposed TRACER method significantly improves detection performance ($p < 0.05$) compared to NCC-Translate, MMI, and OF. However, no statistically significant differences were found between TRACER and the Unregistered, NCC-General, or ECC methods for Precision.

Table 3.10 revealed that TRACER stands out with increases in precision by up to 4 % and recall by up to 27 %. Additionally, the higher F1 score indicates a better balance between identifying relevant RoI and minimising erroneous detection, with up to a 19 % improvement seen with the image example, fused after applying the TRACER registration method. Notably, TRACER achieves consistently higher Recall and F1-Score, underscoring its superior ability to minimise false negatives and improve overall detection robustness. These findings strongly support the adoption of the proposed registration method for clinical applications requiring reliable and accurate image fusion.

3.5 CONCLUSION

This chapter presented temporal reliability and accuracy via correlation enhanced registration (TRACER), a comprehensive image processing pipeline developed to resolve the critical challenge of motion artefacts in real-time optical endomicroscopy (OEM)-fluorescence lifetime imaging (FLIm). By integrating a multi-stage process—which includes the removal of uninformative frames, mo-

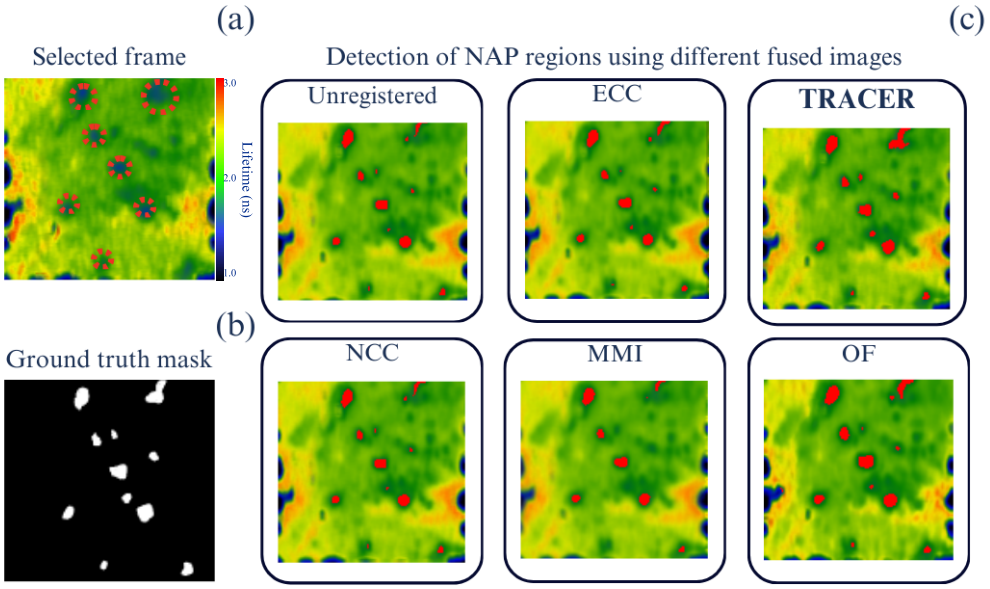


Figure 3.10: Evaluation of NAP blob detection, contrasting the efficacy of the proposed TRACER with outcomes from alternative image registration techniques. (a) presents a sample image from the sequence with doted red circles highlighting the targeted NAP RoIs. (b) displays the ground truth binary mask. (c) depicts the superimposed masks on the fused images, highlighting the detected regions in red.

Table 3.10: Comparing the detection of NAP in fused images from various registration techniques with the proposed TRACER method.

Method	Evaluation Metric		
	Precision ↑	Recall ↑	F1 score ↑
Unregistered	0.91 ± 0.07	0.63 ± 0.06	0.75 ± 0.08
NCC-Translate	0.92 ± 0.08	0.61 ± 0.05	0.72 ± 0.07
NCC-General	0.92 ± 0.06	0.61 ± 0.06	0.73 ± 0.08
ECC	0.92 ± 0.07	0.60 ± 0.09	0.73 ± 0.04
MMI	0.89 ± 0.08	0.50 ± 0.05	0.64 ± 0.07
OF	0.79 ± 0.09	0.68 ± 0.08	0.73 ± 0.04
TRACER	**0.93 ± 0.04**	**0.88 ± 0.04**	**0.90 ± 0.08**

Table 3.11: Statistical comparison between TRACER and other registration methods for Precision, Recall, and F1-Score.

Method	Precision		Recall		F1-Score	
	p-value	Significance	*p*-value	Significance	*p*-value	Significance
Unregistered	0.547	False	0.000	True	0.000	True
NCC-Translate	0.004	True	0.000	True	0.000	True
NCC-General	0.248	False	0.000	True	0.000	True
ECC	0.276	False	0.000	True	0.000	True
MMI	0.030	True	0.000	True	0.000	True
OF	0.000	True	0.000	True	0.000	True

tion characterisation via dense optical flow (OF), and a novel tracking-based normalised cross correlation (NCC) registration guided by channel and spatial reliability tracker (CSRT)—the TRACER pipeline provides an accurate and computationally efficient framework for processing raw FLIm data. This approach enables precise image alignment and fusion, even within dynamic sequences containing large displacements and multiple distinct scenes.

The empirical evaluation demonstrated that this pipeline yields substantial enhancements, with improvements of 20% to 30% observed across quality of alignment (QA), structural similarity index measure (SSIM), and normalised root mean squared error (NRMSE) metrics. The proposed registration method consistently surpassed existing state-of-the-art techniques, delivering average performance gains of 5 % in QA, 8 % in SSIM, and 3 % in NRMSE. Beyond its accuracy, the method is also computationally lean, operating an order of magnitude faster than the next best-performing approach. This combination of precision and speed establishes TRACER as a highly valuable tool for real-time FLIm analysis.

The robustness and real-time applicability of the pipeline were validated using both simulated and real FLIm datasets, highlighting its capacity to handle complex motion patterns efficiently and its potential for broader use in clinical and research settings. Looking ahead, future work will aim to build upon this foundation by expanding the validation dataset and exploring the potential of non-rigid registration techniques to manage more complex, non-linear motion. Further refinements to image fusion methods and optimisation of the pipeline's parameters will also be pursued.

In summary, the TRACER pipeline represents a significant step forward for the field, effectively addressing the unique registration challenges posed by FLIm data. By enhancing the quality and temporal reliability of real-time FLIm imaging,

this work makes a meaningful contribution to advancing pulmonary research and diagnostics.

4

A MULTI-OBJECTIVE APPROACH TO IMAGE REGISTRATION

4.1 INTRODUCTION

As discussed in Chapter 2, the FLIm data obtained using the custom OEM system, Kronoscan, provides both intensity and lifetime imaging data. Intensity and lifetime images capture different physical properties of the sample, they represent signal amplitude (Important spatial information), while lifetime images represent fluorescence decay times (Important biomedical insight). Each offering distinct insights into tissue characterisation [83]. While intensity images display higher signal-to-noise ratio (SNR), facilitating clearer anatomical structure visualisation, lifetime images encode valuable molecular information, albeit with inherently lower SNR due to fluorescence variability [27]. The difference in signal quality is discussed in more detail in Section 2.2.4.3, with illustrative examples provided in Fig. 2.4, which show that lifetime-based features exhibit greater variability than intensity features under comparable acquisition conditions. This complementary nature underpins the potential for combining both modalities to improve the robustness of image-processing workflows.

Traditionally, intensity-weighted lifetime images have been used to represent FLIm data, enhancing feature stability by leveraging intensity-derived contrast. As shown in Chapter 2, Section 2.2.4.4, Fig 2.4, the intensity images achieve superior SNR, higher texture correlation, and lower texture homogeneity compared to standard lifetime images, which often appear feature-sparse. These characteristics render intensity-weighted images more suitable for extracting key features, such as edges and textures, critical for subsequent analysis tasks [13, 73].

It has proven instrumental in various applications, including drug discovery [19], cancer detection [27], and the development of intelligent FLIm systems using machine learning [20, 10]. Furthermore, the TRACER pipeline, introduced in Chapter 3, utilised this representation to enhance registration accuracy and computational efficiency in real-time pulmonary FLIm.

It is also important to distinguish between the requirements for human versus machine consumption of fluorescence lifetime microscopy (FLIM) data. For human observers, effective visualisation prioritises perceptual clarity, intuitive colour encoding, and temporal stability, enabling rapid qualitative interpretation and clinical usability. In contrast, machine learning models are insensitive

to perceptual aesthetics and instead prioritise statistical consistency, feature separability, and robustness to noise and intensity variation across frames and datasets [130]. These requirements are not always aligned. Representations that are visually intuitive for human interpretation may not be optimal for automated analysis, while feature-efficient numerical representations may be difficult to interpret visually.

In summary, and based on the findings in Chapter 2, Section 2.2.3, the intensity-weighted lifetime representation can provide the following characteristics:

- **Highly Informative**: Due to the combination of high entropy and high SNR, capturing complex details clearly.
- **Pattern-Rich**: With high texture correlation indicating consistent structural patterns.
- **Variably Textured**: The low texture homogeneity means textures vary throughout the image, providing a comprehensive view of different regions.

4.1.1 *Motivation*

While intensity-weighted lifetime images have demonstrated efficacy, pre-generating these representations for registration can result in a loss of modality-specific information due to blending or merging of features [104, 25]. In contrast, by registering the original images, the integrity of the data from each modality is maintained, allowing the registration algorithm to leverage all available information for accurate alignment. For example, intensity images capture richer structural details, which can provide a robust foundation for registering both intensity and lifetime frames. This stabilises sequences by enhancing the localisation of corresponding regions [102, 105]. As illustrated in Fig. 4.1, registration metrics such as QA, SSIM, and NRMSE exhibit improved performance when using intensity images. Although intensity-weighted lifetime images surpass standard lifetime representations, they still underperform compared to intensity-based registration.

These findings underscore the need for a multi-objective registration strategy that integrates intensity and lifetime data. By preserving modality-specific information, this approach could improve registration accuracy while maintaining computational efficiency—an essential factor for real-time clinical applications.

4.1.2 *Problem Statement*

Although TRACER effectively mitigates motion artefacts and enhances registration performance, achieving consistent alignment remains challenging in scenarios with significant structural variability or scene transitions. The reliance

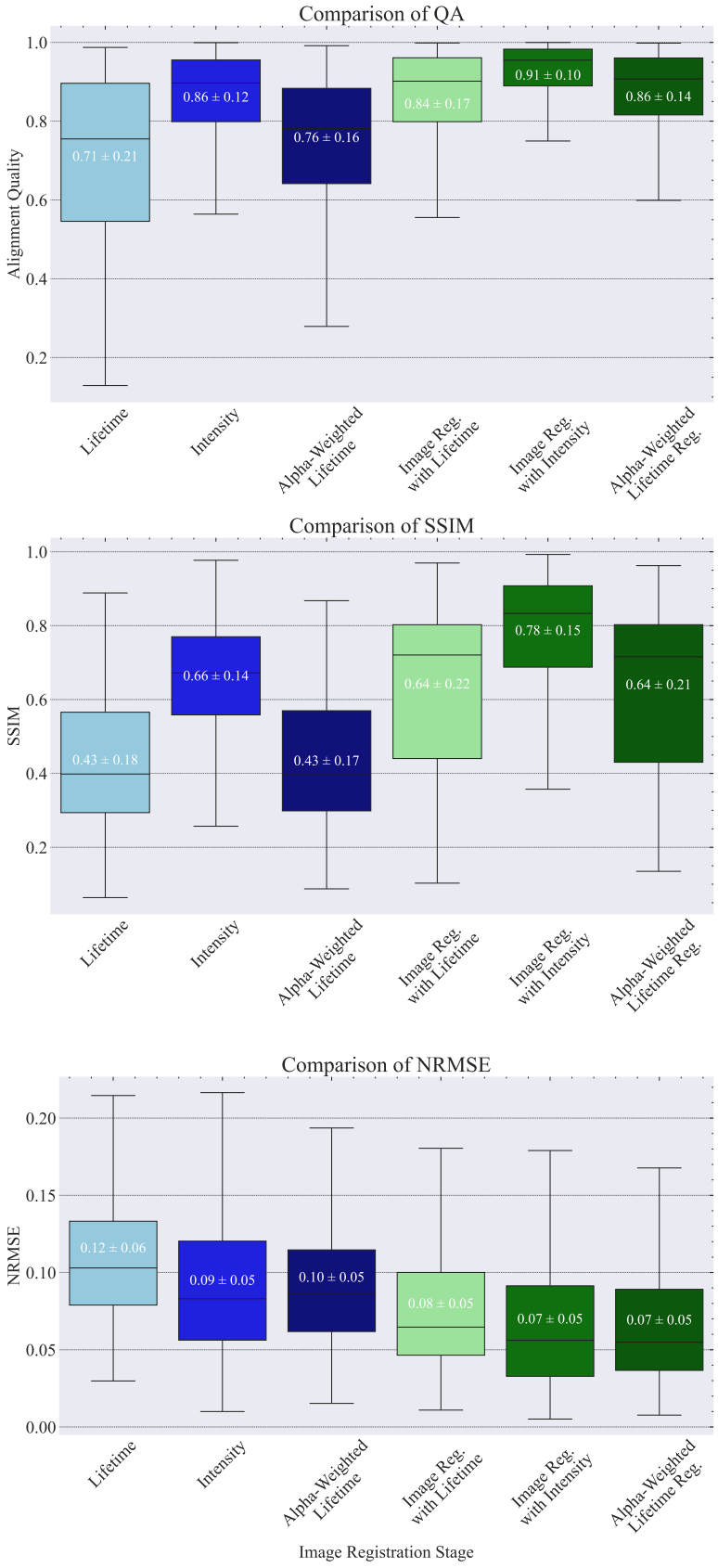


Figure 4.1: Comparison of image registration performance using intensity images, lifetime images, and intensity-weighted lifetime image representations. Metrics including QA, SSIM, and NRMSE highlight the superior performance achieved with intensity-based registration.

on intensity-weighted lifetime images highlights the need for approaches that directly integrate and exploit the complementary strengths of intensity and lifetime modalities.

This chapter addresses the following research questions:

- Can a multi-objective approach combining intensity and lifetime data enhance the accuracy and robustness of FLIm image registration?
- How can this approach be optimised to balance registration performance with the computational constraints of real-time applications?

4.1.3 *Chapter Overview*

This chapter introduces a multi-objective image registration framework designed to improve the alignment of FLIm imaging data by integrating both intensity and lifetime modalities. While intensity images (The fluorescence intensity imaging (FII) imaging data from the OEM-FLIm system) offer superior structural detail and SNR, lifetime images encode critical biochemical information. Traditional approaches, such as intensity-weighted lifetime representations, provide a compromise between these modalities but can obscure modality-specific details. To address this, a multi-objective strategy is proposed throughout Section 4.3, leveraging separate intensity and lifetime images to enhance registration accuracy and robustness.

Section 4.3.2 explores three multi-objective optimisation strategies: weighted sum-based registration, simultaneous multi-objective optimisation, and a multiplicative power-weighted formulation. These are evaluated using global and local optimisation techniques, with simplicial homology global optimisation (SHGO) demonstrating superior performance in identifying optimal weighting factors. A detailed analysis of the NCC space in Section 4.5 reveals a consistent dominance of intensity images in guiding alignment due to their higher SNR and structural consistency. However, in specific cases where lifetime images exhibit stronger feature correlation, they can contribute meaningfully to the registration process.

To further refine this process, a Pareto optimisation analysis in Section 4.6 is conducted using non-dominated sorting genetic algorithm - II (NSGA-II), providing insight into the trade-offs between the two modalities. Based on these findings, a novel dynamic multi-objective search (DYNMOS) algorithm is introduced in Section 4.7, which adaptively selects the most informative modality for registration, achieving comparable accuracy to SHGO while reducing computational overhead by 2.46×.

Section 4.8 demonstrates that integrating intensity and lifetime data through multi-objective registration leads to improved alignment performance while

maintaining computational efficiency. The findings provide a foundation for future adaptive FLIm registration methods with real-time clinical applications.

4.2 IMAGING DATA REPRESENTATION

As outlined in Chapter 2, Section 2.2.3, the various FLIM outputs, namely: intensity, lifetime, and intensity-weighted lifetime, each encode different aspects of the measured signal, from anatomical structure to functional contrast. In this work, the image-based representations $\mathbf{I}(x, y)$ and $\mathbf{L}(x, y)$ are favoured not only for their interpretability, but also for their compatibility with standard image processing techniques.

While fluorescence lifetime fundamentally arises from temporal photon decay, it is often expressed as a 2D image due to both hardware constraints and computational practicality. For downstream visualisation and machine-learning tasks, this two-dimensional intensity-weighted lifetime map may be rendered using colour encodings (*e.g.,* RGB) or combined with an explicit alpha channel (RGBA) to improve perceptual clarity and information salience. However, these representations are purely organisational constructs for display or network input, while the underlying data remain fundamentally two-dimensional and scalar-valued for computational analysis. Systems like the one described in Chapter 2, Section 2.3.1, which operate with a low temporal resolution (*e.g.,* two time gates), preclude robust analysis at the raw decay histogram level. Consequently, lifetime maps are precomputed as part of acquisition and treated as standalone images.

Figure 4.2 visually compares the lifetime image representations resulting from low temporal resolution (2-bin rapid lifetime determination (RLD)) and higher temporal resolution (16-bin decay histogram). While 16-bin systems capture finer decay kinetics, the image-based output is ultimately what is used for spatial analysis and clinical interpretation. Therefore, this work adopts a practical approach that treats $\mathbf{L}(x, y)$ and $\mathbf{I}(x, y)$ as aligned 2D images—despite their origins from time-resolved signals, in order to enable efficient registration, motion compensation, and fusion.

4.3 ALGORITHMIC DEVELOPMENT

The proposed multi-objective registration framework extends the TRACER pipeline by integrating an adaptive strategy that jointly processes intensity and lifetime data. The framework is designed to enhance registration accuracy and robustness through the following key steps:

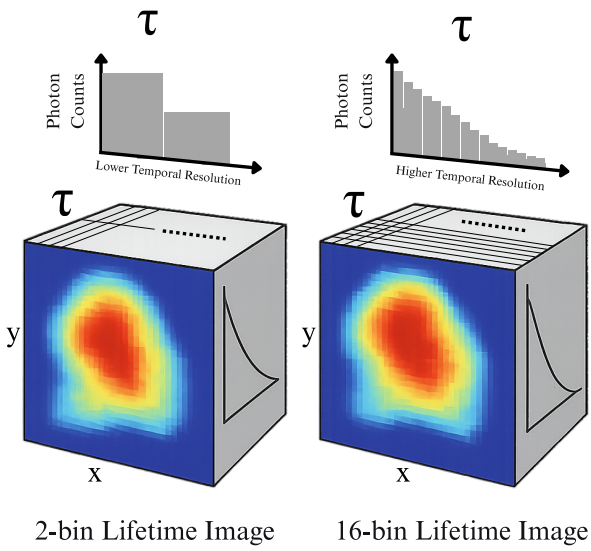


Figure 4.2: Comparison between 2-bin and 16-bin lifetime image representations. Both modalities generate spatial maps of the lifetime parameter $\mathbf{L}(x, y)$, but differ significantly in temporal resolution. Higher-bin systems preserve more of the decay dynamics, while low-bin systems (as used in this work) necessitate image-based lifetime mapping for real-time processing.

1. **Pre-processing**: Ensuring frame quality by eliminating uninformative frames and characterising motion patterns using the methods described in Chapter 3.

2. **Hybrid similarity metric**: Developing a composite similarity metric that combines structural details from intensity images with functional information from lifetime data to improve alignment accuracy.

3. **Optimisation**: Designing an efficient optimisation strategy to meet the computational requirements of real-time applications.

By leveraging the complementary strengths of intensity and lifetime modalities, this framework ensures that critical features from both domains contribute effectively to the registration process. The approach is expected to enhance robustness, particularly in challenging scenarios characterised by significant motion or low SNR.

The optimisation process is framed as a multi-objective problem, where both spatial and temporal alignments are simultaneously optimised, as illustrated in Fig. 4.3. Multi-objective optimisation is a robust approach for addressing complex problems involving multiple, often conflicting, objectives. It also offers improved convergence properties by enabling the algorithm to navigate the similarity landscape more effectively using information from both modalities [135, 136, 137].

By exploring trade-offs between competing objectives, it provides a framework for balanced and informed decision-making [135].

This method aims to align images obtained from the intensity and lifetime imaging modalities to integrate complementary information for enhanced analysis. The approach builds on similarity-based registration by incorporating multiple similarity metrics tailored to each modality [138]. The objective is to determine a transformation function that maximises a combined similarity measure between the reference and moving images from the two modalities.

Multi-objective similarity-based registration has been widely studied, with various similarity metrics and weighting strategies proposed to accommodate differences in image characteristics across modalities [102, 139]. Weighted combinations of similarity measures provide flexibility, allowing certain modalities to be prioritised based on their relevance to the application.

While numerous algorithms achieve multi-objective image registration, mutual information remains the most commonly used similarity metric for intensity-based multimodal image registration [133, 104, 140], with performance relying on the intensity correspondence between images [141]. In this framework, a similarity function is defined that integrates intensity and lifetime data. Specifically, NCC, traditionally used for mono-modal image registration, is extended to handle multiple modalities by combining intensity and lifetime data into a composite metric [135, 140].

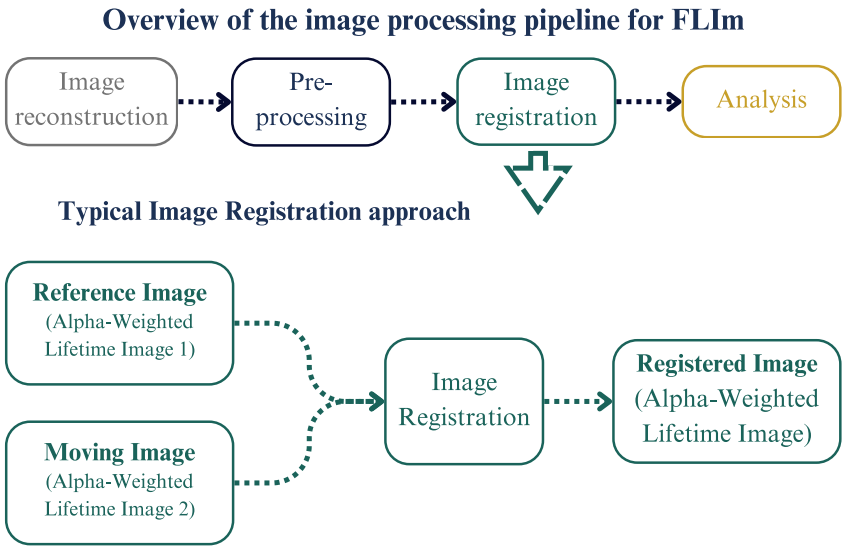
4.3.1 *Formulating the Multi-Objective Function*

Building on the concepts outlined in Chapter 3, particularly Sections 3.2.2 and 3.2.4, where the reference image was defined as $\mathbf{I}_{[n]}$ and the moving image as $\mathbf{I}_{[n+1]}$ within a sequence of n images ($n = \{1, \ldots, N\}$), this section extends the framework to incorporate additional modalities for multi-objective optimisation.

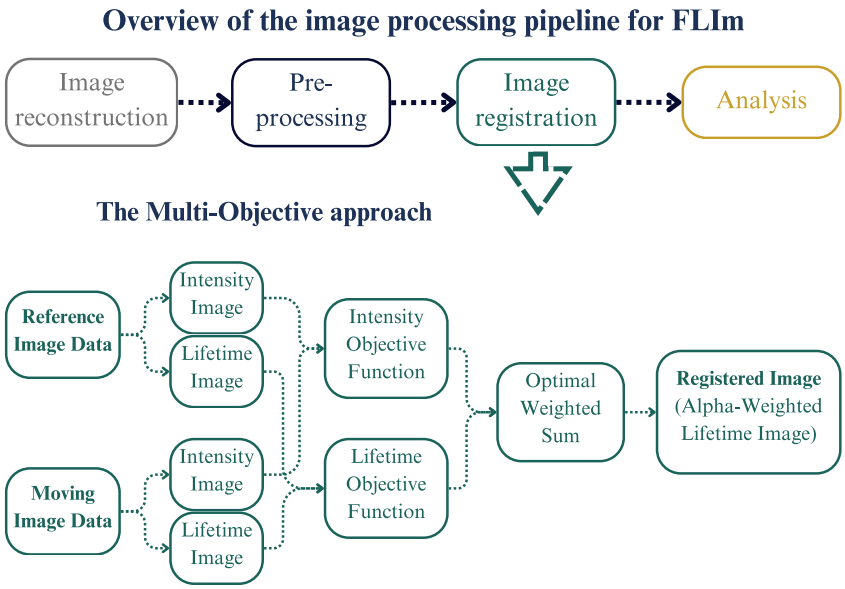
As depicted in Fig. 4.3(b), the proposed approach utilises two reference images: one from the intensity modality, denoted as $\mathbf{I}^{\text{ref}}$, and one from the lifetime modality, denoted as $\mathbf{L}^{\text{ref}}$. Similarly, the corresponding moving images are $\mathbf{I}^{\text{mov}}$ and $\mathbf{L}^{\text{mov}}$. This dual-modality configuration enables the formulation of a comprehensive objective function that integrates complementary information from both modalities, thereby enhancing registration accuracy and robustness.

4.3.2 *Multi-Objective Optimisation Strategies*

This section extends the initial investigation into image registration for FLIm by evaluating three strategies for optimising the joint or multi-objective function. The



(a) The typical image registration approach in FLIm



(b) Schematics of the proposed multi-objective approach to image registration

Figure 4.3: Figure showing a schematic comparison of the TRACER image registration and the multi-objective approach.

goal is to improve robustness in multi-modal registration, particularly addressing the challenges associated with intensity and lifetime image alignment.

4.3.2.1 *Weighted Sum-Based Optimisation*

This strategy is commonly used in multi-objective optimisation to simplify a problem by combining multiple objectives into a single scalar function, thereby enabling the use of conventional optimisation techniques. It serves as a baseline method due to its computational simplicity and flexibility in balancing competing modalities by adjusting the weighting factor. Recent studies have shown its utility in multi-modal medical image registration, where a trade-off between different image features, such as intensity and structural similarity, is crucial [142, 135]. Specifically, for two modalities—intensity and lifetime—the combined similarity measure can be formulated in a simple strategy, where the goal is to optimise the weight ω to find the registration parameters (restricted to translation for simplicity):

$$Maximise\, F(\omega) = \omega \cdot \mathrm{NCC}_{\mathbf{I}} + (1-\omega) \cdot \mathrm{NCC}_{\mathbf{L}}, \tag{4.1}$$

where:

- ω is the weighting factor ($\omega \in \mathbb{R} \cap [0,1]$) that balances the contribution of each modality.
- $\mathrm{NCC}_{\mathbf{I}}$ is the NCC between the intensity reference and moving images.
- $\mathrm{NCC}_{\mathbf{L}}$ is the NCC between the lifetime reference and moving images.

Each NCC term incorporates the spatial transformation T to align the moving image with the reference image, similar to Eq. 2.18 in Chapter 2, Section 2.4.2.2. The NCC for each modality is defined similarly to Eq. 3.2, with the respective images:

$$\mathrm{NCC}_{\mathbf{I}} = \frac{\sum\limits_{\boldsymbol{\chi}} \left(\mathbf{I}^{\mathrm{ref}}(\boldsymbol{\chi}) - \mu_r^{\mathbf{I}}\right)\left(\mathbf{I}^{\mathrm{mov}}(T(\boldsymbol{\chi})) - \mu_m^{\mathbf{I}}\right)}{\sqrt{\sum\limits_{\boldsymbol{\chi}} \left(\mathbf{I}^{\mathrm{ref}}(\boldsymbol{\chi}) - \mu_r^{\mathbf{I}}\right)^2} \sqrt{\sum\limits_{\boldsymbol{\chi}} \left(\mathbf{I}^{\mathrm{mov}}(T(\boldsymbol{\chi})) - \mu_m^{\mathbf{I}}\right)^2}} \tag{4.2}$$

Here, $\mathbf{I}^{\mathrm{ref}}$ and $\mathbf{I}^{\mathrm{mov}}$ are the intensity reference and moving images. $\mathrm{NCC}_{\mathbf{L}}$ is computed similarly with $\mathbf{L}^{\mathrm{ref}}$ and $\mathbf{L}^{\mathrm{mov}}$, the lifetime reference and moving images. The means μ are computed for each modality:

$$\mu_r^{\mathbf{I}} = \frac{1}{N}\sum_{\boldsymbol{\chi}} \mathbf{I}^{\mathrm{ref}}(\boldsymbol{\chi}), \quad \mu_m^{\mathbf{I}} = \frac{1}{N}\sum_{\boldsymbol{\chi}} \mathbf{I}^{\mathrm{mov}}(T(\boldsymbol{\chi})), \tag{4.3}$$

$$\mu_r^{\mathbf{L}} = \frac{1}{N}\sum_{\boldsymbol{\chi}} \mathbf{L}^{\mathrm{ref}}(\boldsymbol{\chi}), \quad \mu_m^{\mathbf{L}} = \frac{1}{N}\sum_{\boldsymbol{\chi}} \mathbf{L}^{\mathrm{mov}}(T(\boldsymbol{\chi})). \tag{4.4}$$

The optimisation problem aims to find the optimal translative transformation T^* using the optimal weighting factor w^* that maximise the combined similarity measure $F(\omega)$:

$$\{T^*\} = \arg\max_{T} F(T, \omega^*), \tag{4.5}$$

subject to $0 \leq \omega \leq 1$.

By optimising the weighting factor, the registration process accounts for the relative importance of each modality, potentially leading to a more accurate alignment.

We define the registered images for each modality as:

$$\mathbf{I}^{\text{reg}} = T^* \circ \mathbf{I}^{\text{mov}}, \quad \mathbf{L}^{\text{reg}} = T^* \circ \mathbf{L}^{\text{mov}}. \tag{4.6}$$

The final registered intensity-weighted lifetime image, $\mathbf{I}_\tau^{\text{reg}}$, is constructed by combining the registered lifetime image $\mathbf{L}^{\text{reg}}$ with the registered intensity image $\mathbf{I}^{\text{reg}}$. Specifically, it emphasises regions of higher fluorescence intensity, enhancing the visual and interpretative quality of lifetime information. The registered intensity-weighted lifetime image is defined as:

$$\mathbf{I}_\tau^{\text{reg}}(x, y) = \frac{\mathbf{I}^{\text{reg}}(x, y)\, \mathbf{L}^{\text{reg}}(x, y)}{\max\left(\mathbf{I}^{\text{reg}}\right)}, \tag{4.7}$$

which effectively normalises intensity to serve as an alpha weighting, highlighting pixels with higher photon counts. In this context, the term "alpha weighting" is used in a pragmatic sense to describe intensity-driven modulation of lifetime contrast, rather than implying a true four-channel (RGBA) physical measurement.

In this multi-objective strategy, the choice of the weighting factor ω plays a crucial role. It determines the influence of each modality in the registration process. In some cases, one modality may provide more reliable information for alignment, and adjusting ω accordingly can improve the registration outcome.

4.3.2.2 *Multi-Objective Simultaneous Optimisation*

Unlike the weighted sum strategy, simultaneous optimisation of weights and registration parameters allows for greater adaptability to varying image characteristics. This strategy acknowledges that the optimal weight may not be static but should be determined dynamically during the optimisation process. Multivariate optimisation techniques have been widely adopted in registration tasks involving diverse imaging modalities, with reported improvements in robustness and alignment accuracy [100]. This multi-objective function can be formulated by re-writing the ω scalarised equation (Eq. 4.1, considering both the weighting

factor ω and θ as the registration parameter variable, which in this case is a column vector (or array) of translation parameters. This can be done by either:

$$Maximise\, F(\omega, \theta) = \omega \cdot \mathrm{NCC}_{\mathbf{I}}(\theta) + (1 - \omega) \cdot \mathrm{NCC}_{\mathbf{L}}(\theta), \tag{4.8}$$

where θ is the registration parameter, ω, $\mathrm{NCC}_{\mathbf{I}}$, and $\mathrm{NCC}_{\mathbf{L}}$ are the same as in Eq. 4.1.

For this strategy, Eq. 4.5, can be re-written as:

$$\{\theta^*, \omega^*\} = \arg\max_{\theta,\omega} F(\theta, \omega), \tag{4.9}$$

4.3.2.3 *Multiplicative Power-Weighted Function*

The multiplicative formulation is inspired by geometric mean-based strategies often employed in cases where interaction effects between objectives are non-linear. By raising each modality to a power, the strategy allows for smooth interpolation between the dominance of one modality over the other, while preserving multiplicative coupling. This method is particularly relevant for FLIm data, where intensity and lifetime represent complementary but distinct sources of information. Multiplicative strategies have been explored in advanced image fusion and registration frameworks [117].

In this third strategy, instead of using a linear combination, the objective function is expressed as the product of the two modalities, each raised to a power corresponding to a weight parameter. This allowed for a non-linear interaction between intensity and lifetime data. This can be expressed as:

$$Maximise\, G(\omega, \theta) = (\mathrm{NCC}_{\mathbf{I}}(\theta))^{\omega} \times (\mathrm{NCC}_{\mathbf{L}}(\theta))^{(1-\omega)}, \tag{4.10}$$

This can be rewritten using logarithmic properties:

$$Maximise\, L(\omega, \theta) = \omega \cdot \log(\mathrm{NCC}_{\mathbf{I}}(\theta)) \times (1 - \omega) \cdot \log(\mathrm{NCC}_{\mathbf{L}}(\theta)), \tag{4.11}$$

Since logarithm is a monotonically increasing function, maximising $G(\omega, \theta)$ is directly equivalent to maximising $L(\omega, \theta)$.

The optimisation problem in the three strategies outlined above can be challenging due to the increased dimensionality and potential for multiple local maxima. Advanced optimisation techniques and careful initialisation will be required to solve it effectively.

4.3.3 *Overview of Optimisation Methods*

In the context of multi-objective image registration, as formulated in Eq. 4.5, the objective is to determine the optimal transformation T^* and weighting factor w^* that maximise the combined similarity measures for the three strategies: $F(\omega)$, $F(\omega, \theta)$, and $G(\omega, \theta)$. This optimisation problem is inherently complex due to the non-linear and non-convex nature of the similarity landscape, which may present multiple local optima, especially in the presence of significant motion or scene transitions.

Optimisation methods can broadly be classified into those dealing with univariate and multivariate functions. Univariate functions, involving a single decision variable, tend to exhibit simpler similarity landscapes, often characterised by a single prominent peak. As a result, straightforward algorithms such as the bi-linear method [140] can be applied with relatively low computational overhead. These methods are effective when optimising individual parameters, such as translation offsets in scenarios with minimal complexity.

Conversely, multivariate functions, which involve multiple interdependent variables, often exhibit more complex landscapes with multiple peaks. Such functions require more sophisticated optimisation algorithms capable of navigating high-dimensional, multi-modal spaces. Gradient-based approaches like Broyden-Fletcher-Goldfarb-Shanno (BFGS) or its limited-memory Broyden-Fletcher-Goldfarb-Shanno with bounds (L-BFGS-B) [143] are commonly employed for smooth, differentiable objective functions. However, in cases where the similarity landscape is discontinuous or highly irregular, evolutionary algorithms such as differential evolution [136] offer a more robust alternative by exploring the search space more globally.

A critical consideration in multi-objective registration is the distinction between uni-modal and multi-modal function spaces. Uni-modal landscapes, typically encountered in simpler registration strategies, contain a single dominant peak corresponding to the optimal solution. In contrast, multi-modal landscapes, which arise in more complex cases with significant displacements or repetitive patterns, exhibit multiple competing peaks. These require optimisation methods that can avoid local optima and reliably identify the global maximum. Hybrid approaches, combining gradient-based and evolutionary methods, are often necessary to balance computational efficiency with robustness in these challenging conditions [135, 140, 144, 138]

By carefully selecting appropriate optimisation strategies based on the complexity of the similarity landscape and the underlying data characteristics, reliable solutions can be achieved for both uni-modal and multi-modal registration problems. Below, we provide an overview of several optimisation techniques, discussing

their principles and applicability to the multi-objective image registration problem in FLIm.

4.3.3.1 *Bi-Linear*

Bi-linear interpolation is a common technique in image processing, including image registration. Bi-linear interpolation can be used to estimate the value of a function at a point that is not on the grid of sample points. This is necessary for evaluating the cost function during image registration, as the transformation may map pixels or voxels from one image to locations that do not correspond to pixel or voxel centres in the other image [137, 140]. Effective to find the maxima of $F(\omega)$.

4.3.3.2 *Differential Evolution*

Differential Evolution is a stochastic, population-based optimisation algorithm. It is well-suited for global optimisation problems, especially those that are non-differentiable, non-linear, and multi-objective. Principally, the method aims to maintain a population of candidate solutions. It utilises mutation, crossover, and selection operations to evolve the population. It is also capable of escaping local optima to find the global optimum [136].

Differential Evolution can effectively search the space of ω to find the global maximum of $F(\omega, \theta)$, and $G(\omega, \theta)$. Its ability to handle multi-objective functions makes it a strong candidate for optimizing the weighting factor when the similarity landscape is complex.

4.3.3.3 *Limited-memory Broyden-Fletcher-Goldfarb-Shanno with Bounds*

L-BFGS-B is an optimisation algorithm that handles large-scale problems with bound constraints [143]. It uses an approximation to the BFGS algorithm, incorporating bound constraints on variables. It requires gradient information (or numerical approximation).

It is primarily suited for smooth, differentiable functions and may not perform well if $F(\theta, \omega)$ is non-smooth or has multiple local maxima. Additionally, computing accurate gradients may be challenging if the similarity measure is not differentiable with respect to ω or θ.

4.3.3.4 *Truncated Newton Constrained*

Truncated Newton constrained (TNC) is an algorithm for large-scale optimisation with bound constraints, using a truncated Newton approach [145]. It employs Newton's method with constraints, which is suitable for smooth, large-scale problems.

Similar to L-BFGS-B, TNC it requires gradient information. However, it could aslo be used for optimising ω, or ω and θ within bounds. However, the same limitations regarding gradient requirements and susceptibility to local maxima apply. It may not be the most effective choice for a potentially non-smooth, multi-objective similarity function.

4.3.3.5 *Simplicial Homology Global Optimisation*

SHGO is a deterministic global optimisation algorithm that systematically explores the search space using simplicial subdivisions and homology theory [146]. The method divides the search space into simplices (triangles in higher dimensions). It uses sampling and partitioning strategies to cover the search space. This guarantees convergence to the global optimum under certain conditions.

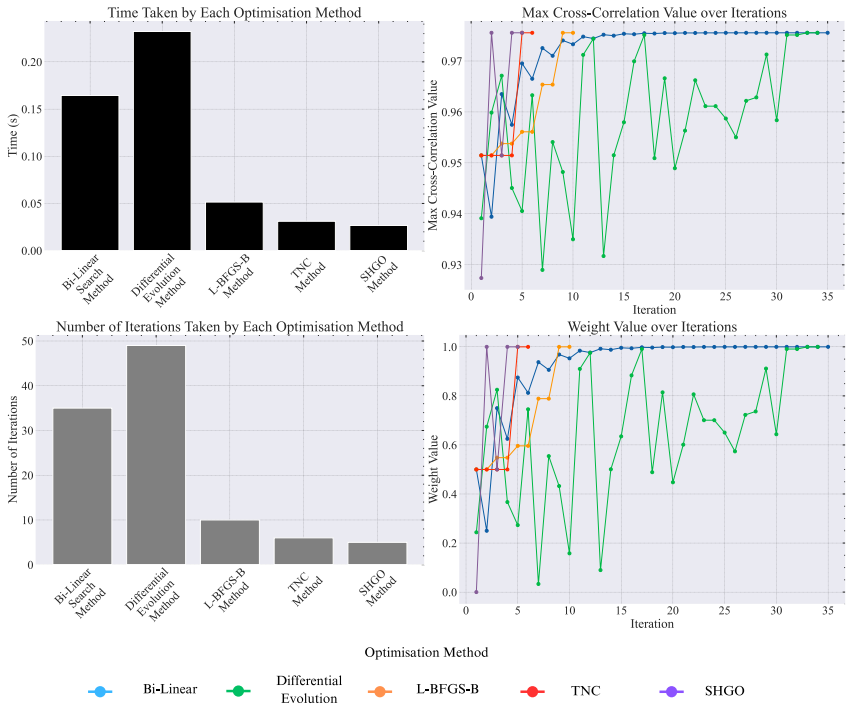
For the optimisation of ω or ω and θ, SHGO can efficiently navigate the search space to locate the global maximum of $F(\theta, \omega)$. Its deterministic nature ensures consistent results across runs, which is advantageous for reproducibility in image registration tasks.

4.3.3.6 *Comparison and Applicability*

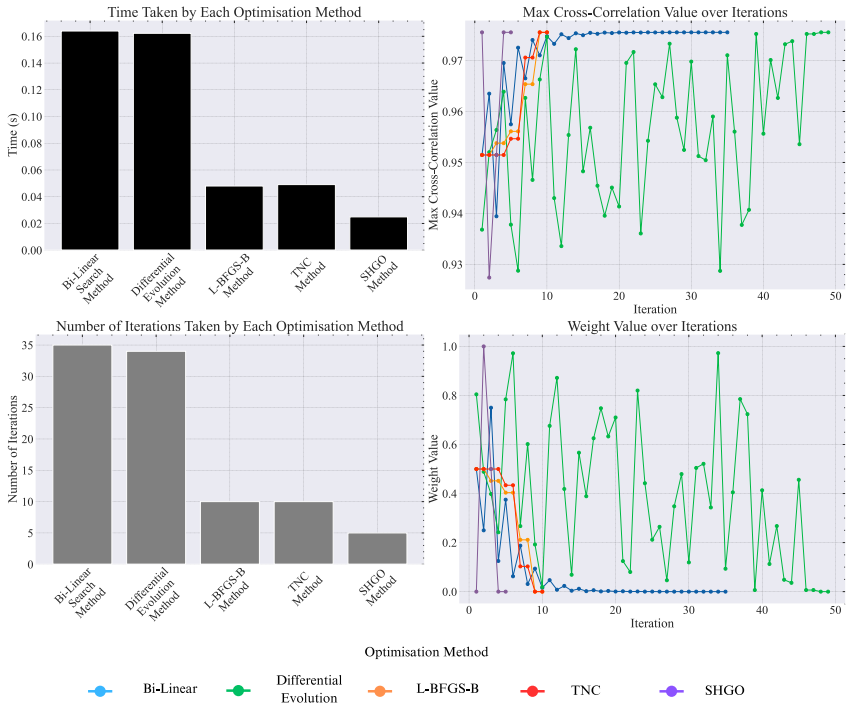
Fig. 4.4 presents the multi-objective results for a pair of intensity and lifetime images, comparing the performance of the multi-objective optimisation (MOO) methods discussed earlier: bi-linear search, differential evolution, L-BFGS-B, TNC, and SHGO. For intensity image registration, the SHGO method consistently identifies the optimal ω in as few as 4 iterations and approximately 4.5 ms. Notably, the selected ω for intensity-based MOO often defaults to 1, regardless of the optimisation method, as shown in Fig. 4.4(a). Conversely, Fig. 4.4(b) highlights the performance of these methods for lifetime image registration, where ω is frequently set to 0, favouring the intensity modality to align lifetime images.

An empirical evaluation on a larger sample of images, including multiple FLIm sequences, was conducted to further compare the optimisation methods detailed in Section 4.3.3. The results, depicted in Fig. 4.5, indicate that SHGO achieves efficient convergence with fewer iterations and higher similarity values. Its deterministic nature and global search capability make it particularly effective for identifying the optimal ω^* in this context. While differential evolution also locates the global maximum, it incurs significantly higher computational costs due to its stochastic nature. Bi-linear search, while prone to local maxima due to its unimodality assumption [137], and gradient-based methods (L-BFGS-B, TNC)—which are sensitive to unreliable gradient information and local maxima risks—achieve comparable image registration QA to other methods but lack the speed and efficiency of SHGO.

In summary:



(a) multi-objective optimisation comparison, registering intensity images.



(b) multi-objective optimisation comparison, registering lifetime images.

Figure 4.4: Performance comparison of optimisation methods for multi-objective registration. (Top left) Bar plot showing the time taken by each optimisation method, with TNC and SHGO being the fastest. (Bottom left) Bar plot comparing the number of iterations required by each method, displaying a trend similar to the time comparison. (Top right) Scatter plot illustrating the maximum QA value achieved at each iteration across all optimisation methods. (Bottom right) Scatter plot showing the ω value at each iteration for all methods, highlighting variations in convergence behaviour.

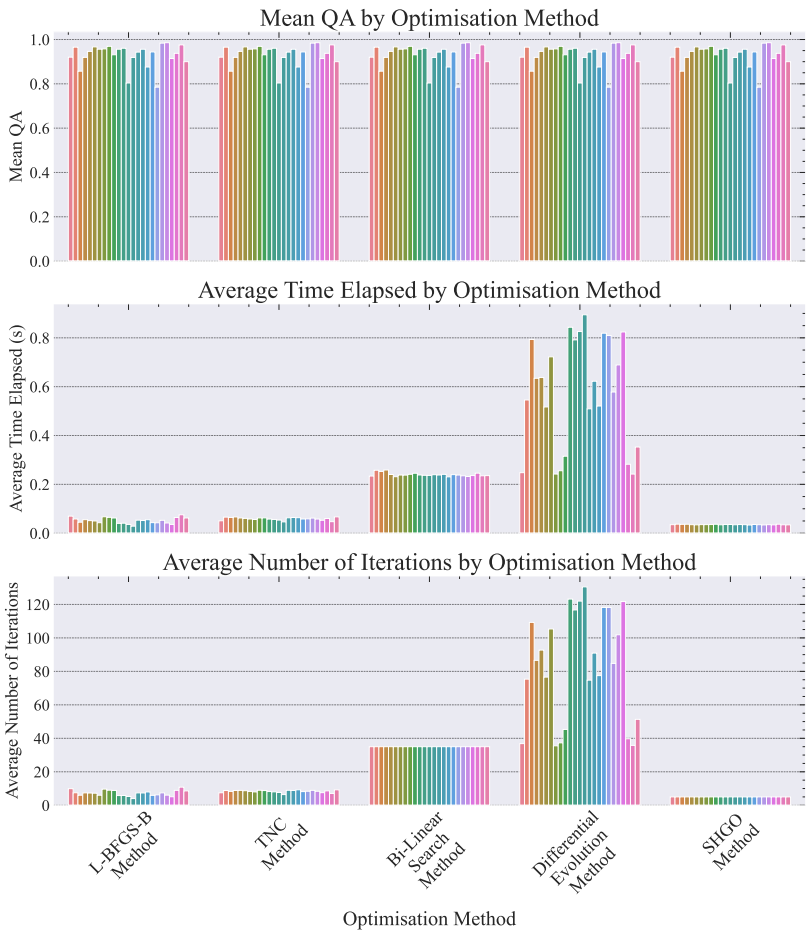


Figure 4.5: Comparison of the performance of different MOO methods outlined in Section 4.3.3. Each coloured bar represents a distinct dataset (image sequence) comprising at least 80 images. The **top** plot shows the maximum QA achieved by each method for each image sequence. The **middle** plot compares the average time taken per image sequence by each method. The **bottom** plot highlights the average number of iterations required by each method across the image sequences.

- **Global vs. Local Optimisation:** Global optimisers such as Differential Evolution and SHGO reliably find the global maximum for multi-objective functions, whereas local optimisers (bi-linear search, L-BFGS-B, TNC) may converge faster but are prone to local maxima.
- **Deterministic vs. Stochastic Approaches:** SHGO offers consistent, deterministic results, whereas Differential Evolution's stochastic approach introduces variability across runs.
- **Derivative Requirements:** Gradient-based methods (L-BFGS-B, TNC) require reliable gradient information, which may be unavailable or inconsistent for $F(\theta, \omega)$. Derivative-free methods, such as Differential Evolution and SHGO, are better suited for this application.

- **Computational Efficiency:** Computational cost is critical when the objective function involves resource-intensive operations, such as image transformations and similarity calculations. SHGO demonstrates superior efficiency, balancing thorough search with reduced computational overhead.

4.4 ANALYSIS OF MULTI-OBJECTIVE OPTIMISATION

A comparative analysis was conducted on three multi-objective image registration strategies, evaluated using two key metrics: the weight assigned to the intensity modality during registration (ω) and the QA metric. Statistical differences between the strategies were assessed using Tukey's honestly significant difference (HSD) test [147], which can be readily implemented via the "statsmodels" Python library [148].

4.4.1 *Tukey's HSD Test Overview*

Tukey's HSD test is a statistical method used to determine whether the means of multiple groups differ significantly. It is particularly effective for pairwise comparisons after an analysis of variance indicates that at least one group mean is significantly different. The test controls the error rate, minimising the probability of incorrectly identifying significant differences across multiple comparisons.

For each pair of groups, Tukey's HSD test evaluates:

- The difference between group means (Mean Diff).
- The adjusted p-value, representing the probability of observing the difference under the null hypothesis of no difference.
- The 95% confidence interval (CI) for the mean difference.
- Whether to reject the null hypothesis based on the adjusted p-value and CI.

In this analysis, Tukey's HSD test was applied to compare the performance of three registration strategies, providing insights into both ω assignment and QA metrics. The strategies were defined as follows:

- **Strategy 1**: Weighted sum-based approach (Section 4.3.2.1).
- **Strategy 2**: Simultaneous approach (Section 4.3.2.2).
- **Strategy 3**: Multiplicative approach (Section 4.3.2.3).

4.4.2 *Weight Assignment Metric*

The first metric analysed is the weight assigned to the intensity modality. Table 4.1 summarises the results of Tukey's HSD test for this metric.

Table 4.1: Tukey HSD Test Results for Weight Assignment Metric

Group 1	Group 2	Mean Diff	Adjusted p-value	Lower CI	Upper CI	Reject Null Hypothesis
Strategy 1	Strategy 2	0.1741	<0.001	0.1496	0.1987	Yes
Strategy 1	Strategy 3	0.1741	<0.001	0.1496	0.1987	Yes
Strategy 2	Strategy 3	0.0000	1.000	-0.0245	0.0245	No

The results indicate significant differences between Strategy 1 and both Strategies 2 and 3. However, no significant difference is observed between Strategies 2 and 3. These findings suggest that Strategy 1 systematically assigns weights differently, placing greater emphasis on the intensity modality. This distinction does not necessarily imply superior performance, as further discussed in the QA analysis.

4.4.3 *Quality Alignment Metric*

The QA metric evaluates registration quality by measuring the NCC between the registered and reference images. Table 4.2 presents the Tukey HSD results for this metric.

Table 4.2: Tukey HSD Test Results for QA Metric

Group 1	Group 2	Mean Diff	Adjusted p-value	Lower CI	Upper CI	Reject Null Hypothesis
Strategy 1	Strategy 2	0.0063	0.0067	0.0015	0.0112	Yes
Strategy 1	Strategy 3	0.0021	0.5654	-0.0028	0.0070	No
Strategy 2	Strategy 3	-0.0042	0.1066	-0.0091	0.0007	No

The results for the QA metric reveal a statistically significant difference between Strategy 1 and Strategy 2, with Strategy 1 exhibiting marginally better registration quality. However, no significant differences are observed in the comparisons between Strategy 1 and Strategy 3 or between Strategy 2 and Strategy 3. These findings indicate that the overall performance of the three strategies in terms of registration quality is broadly similar.

4.4.4 *Comparative Analysis*

The comparative analysis of weight (ω) assignment and QA metrics highlights the following key insights:

- Strategy 1 assigns weights to the intensity modality differently from Strategies 2 and 3, with statistically significant differences. However, this distinction does not appear to result in superior registration quality.
- The QA metric indicates broadly similar performance among the three strategies, with only a marginally significant improvement observed for Strategy 1 over Strategy 2.
- The absence of significant differences between Strategies 1 and 3, as well as between Strategies 2 and 3, suggests that all three strategies are functionally equivalent in achieving effective registration.

From a practical standpoint, any of the three strategies could be utilised for registration tasks, given their comparable performance in terms of QA. The choice of strategy may instead be guided by other considerations, such as computational efficiency, robustness under specific imaging conditions, or ease of implementation.

4.5 ANALYSIS OF THE NORMALISED CROSS-CORRELATION SPACE

The multi-objective image registration problem exhibits the following key characteristics:

- **Objective Function:** While the combined similarity measure $f(\omega)$ is univariate with respect to ω, it remains multi-objective due to the complex interplay between intensity and lifetime images.
- **Constraints:** The weighting factor ω is constrained within the range [0, 1].
- **Computational Cost:** Evaluating $f(\omega, \theta)$ involves computationally intensive operations, including image transformations and similarity calculations.

An in-depth analysis of the NCC space reveals key insights into the behaviour of the optimisation process. The analysis indicates a consistent tendency to favour the intensity modality over the lifetime modality in most cases, with occasional instances of the opposite. This trend persists regardless of the chosen MOO strategy.

4.5.1 *Intensity Dominance*

As outlined in Section 4.1, intensity images exhibit superior SNR, higher texture correlation, and lower texture homogeneity compared to standard lifetime images. These characteristics directly influence the NCC function derived from intensity images [133]. The consistency and reliability of intensity values enhance the

effectiveness of similarity measures like NCC, providing a robust foundation for image registration [100, 149].

This dominance is often reflected in the MOO, where higher ω values are assigned to $\mathrm{NCC}_{\mathbf{I}}$. Upon analysing the NCC space, a recurring pattern emerges, as depicted in Fig. 4.6. In this pattern, the peak of $\mathrm{NCC}_{\mathbf{I}}$ is more prominent and exhibits higher values compared to $\mathrm{NCC}_{\mathbf{L}}$. This observation highlights the inherent strength of intensity images in guiding the registration process.

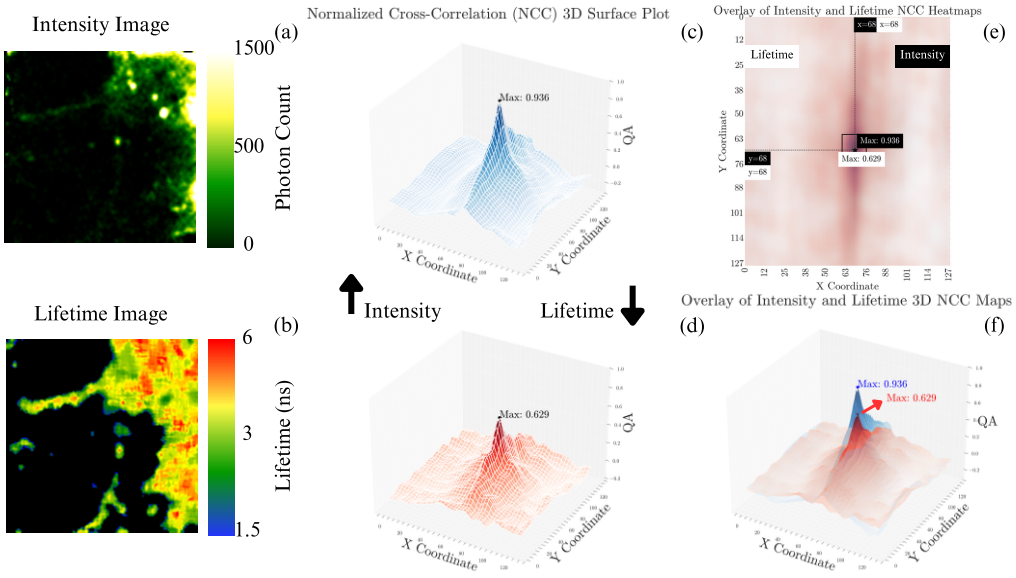


Figure 4.6: Comparison of the NCC spaces between intensity and lifetime imaging modalities. (a) An example intensity image, displaying features that result in a more prominent peak in the NCC space shown in (c), compared to the lifetime image in (b). Although the lifetime image exhibits a peak at the same location as the intensity image, as indicated by the heatmap overlay in (e), the peak is less prominent and lower in value. (f) Overlays the two 3D NCC maps, emphasising the dominance of the intensity NCC space.

Similarity measures like NCC are most effective when applied to images with prominent features and low noise levels [133, 117]. Given the reliability of intensity data, the MOO often prioritises intensity alignment to achieve superior overall image registration. Using intensity-derived transformations for aligning lifetime images reduces misalignments and improves registration outcomes for lifetime data.

As illustrated in Fig. 4.7, the localisation of the NCC peak for lifetime images is less accurate than that for intensity images. The lifetime NCC map not only exhibits a lower peak energy but also results in a computed translation that deviates from the true alignment achieved by intensity maps. This discrepancy is further emphasised in the heatmap overlay, where the lifetime-derived alignment fails to match the intensity-derived alignment.

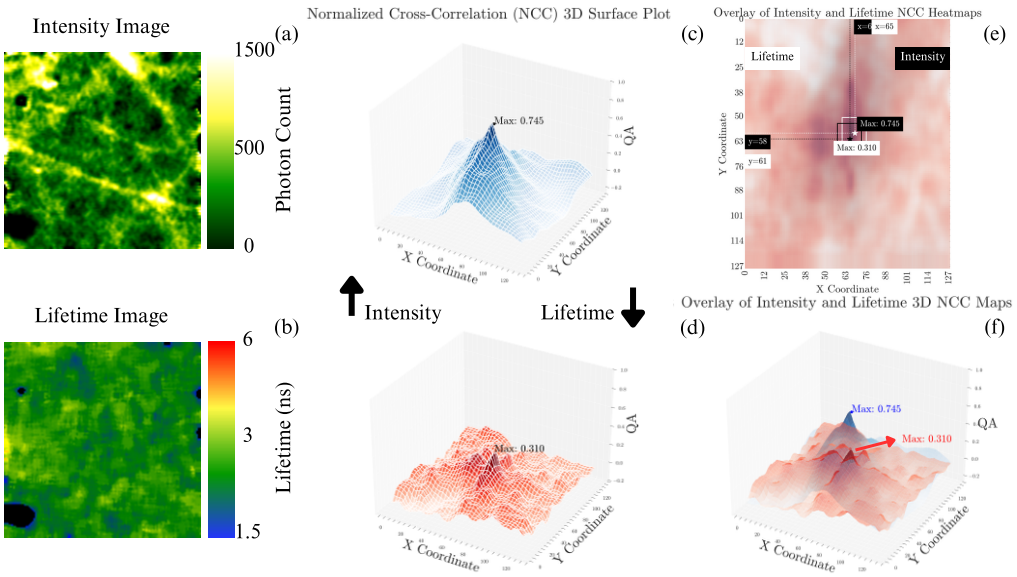


Figure 4.7: Reliability of intensity data demonstrated through NCC space analysis. (a) An example intensity image, producing an accurate and prominent peak in the NCC space (c), compared to the lifetime image (b), which exhibits a peak at a different location. This misalignment results in an inaccurate translation for lifetime registration. (e) The heatmap overlay shows the mismatch between intensity- and lifetime-derived peaks, with the lifetime NCC peak being less prominent and lower in value. (f) A 3D overlay of the NCC spaces further highlights the dominance and reliability of intensity data for registration.

4.5.2 *Lifetime Selection*

Analysing the distribution of the selected ω values in optimising the different strategies, as outlined in Section 4.3.2, and using SHGO—the best-performing optimisation method—reveals a strong preference for intensity images as the primary modality for registration. This preference arises from the superior reliability and clarity of intensity images. Fig. 4.8 highlights the dominance of the intensity modality over the lifetime modality. However, for image pairs where the optimisation assigned a weight of 1 to the lifetime modality, specific patterns emerge, as shown in Fig. 4.9:

- The dominant peak energy in the lifetime NCC space exceeds that of the intensity NCC space.
- Lifetime pairs achieving higher weighting than intensity exhibit higher SNR and feature correlation within the lifetime modality.
- Corresponding intensity pairs display a more uncertain NCC space, characterised by multiple peaks and greater displacements between frames.

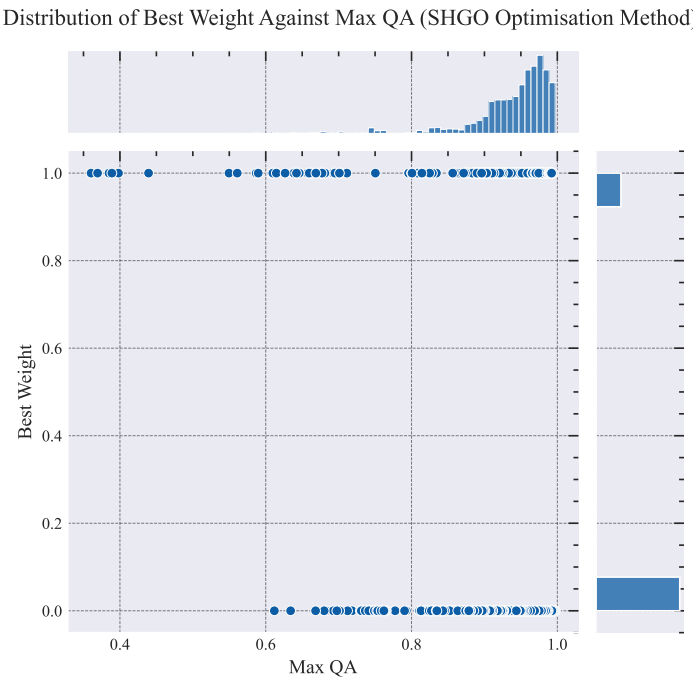


Figure 4.8: Distribution of optimal weight (ω) values selected during optimisation using SHGO. The plot demonstrates the predominance of intensity-dominant registrations, with only occasional cases favouring the lifetime modality.

This increased uncertainty in the intensity NCC space likely contributes to the selection of the lifetime modality for such pairs. Specifically, in these cases, the reduced peak energy and increased ambiguity in intensity NCC maps make lifetime NCC a more reliable choice. Feature metrics, as discussed in Section 4.1—namely SNR, texture homogeneity, and texture correlation—further support this finding. Lifetime images assigned a weight (ω) of 1 consistently exhibit the highest SNR, lowest texture homogeneity, and highest texture correlation. These characteristics occasionally allow the lifetime modality to dominate the registration process.

4.6 PARETO OPTIMISATION ANALYSIS

To further explore the relationship between the two modalities, a Pareto optimisation approach is implemented using the NSGA-II algorithm. In image registration, particularly for multi-modal imaging (*e.g.,* magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET)), different objective functions might capture different aspects of alignment quality—such as intensity-based similarity, structural similarity, or landmark alignment. NSGA-II helps discover trade-offs across these objectives simultaneously.

> A solution A is said to dominate solution B if A is at least as good in all objectives and strictly better in at least one.

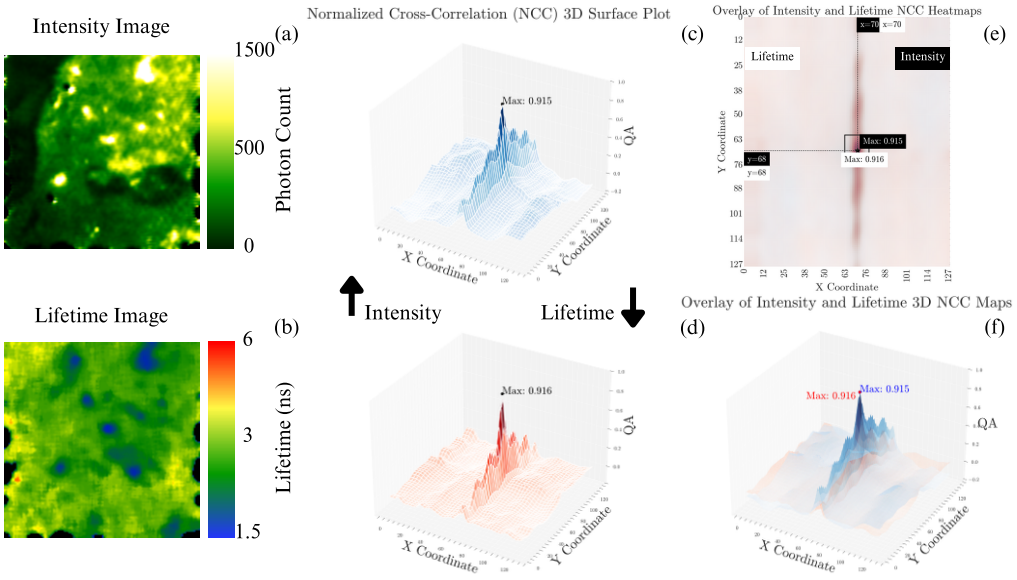


Figure 4.9: Analysis of the NCC space in cases where the lifetime modality is assigned a weight (ω) of 1. (a) An example lifetime image producing a slightly more prominent peak in the NCC space shown in (c) compared to the corresponding intensity image in (b). While the intensity NCC map (c) exhibits a peak, it is less dominant and does not reflect the true translation for optimal registration. (e) The heatmap overlay highlights the differences in peak localisation between lifetime and intensity modalities, with the lifetime NCC peak being more pronounced in this case. (f) The 3D overlay of NCC spaces further emphasises the occasional dominance of the lifetime modality.

Essentially, NSGA-II performs a sort of the population into non-dominated fronts, where the first front contains solutions not dominated by any member of the population, the second front includes solutions only dominated by members of the first front, and so on.

However, when the two objective functions are not of a competing nature, as in the case where both represent the NCC between images from two different modalities, the Pareto analysis is less about discovering trade-offs and more about: Confirming Alignment Robustness Across Modalities, where a small or single Pareto front suggests that the alignment algorithm is consistent and effective; and Evaluating Diversity Across Image Pairs, where a lack of a Pareto front (only one point) could signal that the modalities are strongly correlated, making alignment straightforward. Conversely, a larger Pareto front suggests variability that may need further investigation to understand its source (e.g., noise, differences in modality characteristics).

4.6.1 *Notation and Setup*

1. **Domain:** Let $\Omega \subset \mathbb{R}^d$ be the spatial domain of interest. Depending on the application, d can be 2 or 3.
2. **Intensity Image:** Denote by $\mathbf{I} : \Omega \to \mathbb{R}$ the intensity image (modality 1).
3. **Lifetime Image:** Denote by $\mathbf{L} : \Omega \to \mathbb{R}$ the lifetime image (modality 2).
4. **Transformation:** Let
$$T(\mathbf{x}; \theta)$$
be a spatial transformation that maps a point $\mathbf{x} \in \chi$ to another point in χ, parameterised by $\theta \in \theta \subseteq \mathbb{R}^n$. Typical choices for T include rigid, affine, or non-rigid models, depending on the application.

We aim to simultaneously align the intensity and lifetime images, leading to two objectives. Define:

$$\begin{aligned} f_1(\theta) &= \mathrm{NCC}_{\mathbf{I}}(\mathbf{I}, \mathbf{I} \circ T(\cdot; \theta)), \\ f_2(\theta) &= \mathrm{NCC}_{\mathbf{L}}(\mathbf{L}, \mathbf{L} \circ T(\cdot; \theta)), \end{aligned} \tag{4.12}$$

where, $\mathrm{NCC}_{\mathbf{I}}$, and $\mathrm{NCC}_{\mathbf{L}}$ measure the NCC similarity in the intensity and lifetime domains, as in Eq. 3.2.

Rather than combining f_1 and f_2, computed using Eq. 4.12, into a single cost function, we retain them as separate objectives. We seek a *Pareto-optimal* set of solutions θ^* by solving:

$$\min_{\theta \in \Theta} \Big(f_1(\theta),\ f_2(\theta) \Big). \tag{4.13}$$

Equivalently:

$$\theta^* \ = \ \arg\min_{\theta \in \Theta} \begin{bmatrix} f_1(\theta) \\ f_2(\theta) \end{bmatrix}, \tag{4.14}$$

where the outcome is a set of *non-dominated* (Pareto) solutions, each balancing the two objectives differently. Once the non-dominated fronts are established, NSGA-II encourages diversity within each front using a crowding distance measure, which estimates how "crowded" a solution's neighbourhood is. This fosters a spread of solutions along the Pareto front, preventing convergence to a narrow region of the solution space.

Often, we desire additional constraints to ensure physiologically or physically plausible deformations. For instance:

$$\text{Minimise}\big(f_1(\theta),\ f_2(\theta)\big) \tag{4.15}$$

Subject to :

$$R(\theta) \leq \kappa, \tag{4.16}$$

where $R(\theta)$ is a regularisation term (e.g., bending energy or gradient norm of the deformation field) and κ is a bound controlling its size.

4.6.2 *Interpretation in FLIm Context*

In FLIm, the **intensity dimension** captures the amplitude or photon count, while the **lifetime dimension** captures decay characteristics of the fluorescence signal. Aligning both simultaneously means we want:

- Good structural or spatial match based on intensity.
- Consistent alignment of lifetime signals, which can be highly sensitive to biochemical or molecular changes.

A multi-objective formulation naturally suits this setting, since different scientific or clinical goals may prioritise one metric over another. Bringing together Eq. 4.12, and Eq. 4.14, a concise notation for the FLIm registration problem (two objectives) is:

$$\theta^* = \arg\min_{\theta \in \Theta} \begin{bmatrix} D_{\mathbf{I}}(I,\, \mathbf{L} \circ T(\cdot;\theta)) \\ D_{\mathbf{L}}(\mathbf{L},\, \mathbf{L} \circ T(\cdot;\theta)) \end{bmatrix}, \tag{4.17}$$

potentially with a regularisation constraint to limit deformation complexity. The solution θ^* will typically be a *set* of optimal transformations forming the *Pareto front*, each representing a different compromise between intensity-based and lifetime-based alignment.

This can be implemented in Python using the distributed evolutionary algorithms in Python (DEAP) library [150]. DEAP is well-documented, modular, and designed to enable quick prototyping, customisation, and scaling of evolutionary algorithms. By separating solutions into non-dominated fronts and maintaining a diverse population, NSGA-II guides the search toward a Pareto front, offering a spectrum of trade-offs among registration objectives. This will help to gain deeper insights into how different alignment metrics interact, paving the way for better-informed decisions on which registration strategy best meets clinical or research needs.

4.6.3 *Pareto Front Analysis*

The outcome of implementing the NSGA-II using DEAP consistently indicated a strong correlation between the objectives—improving the alignment of one modality often improved the other. This is expected when two objective functions

are not competing, as in the case where both represent the NCC. When the objective values $(f_1(\theta), f_2(\theta))$ for the population solutions are plotted, as in Fig. 4.10, most non-Pareto solutions fall along a line trending toward the maximum NCC for both the intensity and lifetime modalities. Different image pairs might exhibit different relationships between NCC1 and NCC2: with highly correlated modalities, a small Pareto front indicates that alignment improves both NCC values simultaneously. For loosely correlated modalities, a more diverse Pareto front emerges, where aligning one modality sacrifices alignment for the other, leading to a larger set of trade-offs.

As shown in Fig. 4.10, among the sample of six image pairs, only pair 3 exhibits a small front with two solutions, suggesting a slight trade-off in alignment quality for this pair.

Investigating the NCC space, as detailed in Section 4.5, reveals that the maximum value of the computed NCC in $f_2(\theta)$ does not significantly exceed those of the other pairs. In contrast, $f_1(\theta)$ more often exhibits a markedly higher NCC, suggesting that $f_1(\theta)$ disproportionately affects the scalarised objective. This results in a higher weighting for the corresponding modality, thereby skewing the optimisation process to favour $f_1(\theta)$ over $f_2(\theta)$, resulting in a single Pareto solution.

Overlaying multiple Pareto fronts for different image pairs can reveal the dominance of one modality over the other, particularly when a small Pareto front or clustering along one axis is observed. As shown in Fig. 4.11, this approach highlights the dominance of intensity over lifetime when the slope of the clustering is steep or heavily skewed toward one axis.

This paired investigation provides deeper insights into the dominance of modalities and the behaviour of MOO methods like scalarisation in such strategies. By combining the dominance trend observed in Pareto-based solutions with the analytical findings about the scalarisation of the NCC space, a more comprehensive picture can be created of how and why one modality influences the optimisation process more than the other.

4.7 DYNAMIC MULTI-OBJECTIVE IMAGE REGISTRATION

The consistent dominance of the intensity modality observed in previous analyses suggests an opportunity to optimise computational resources by focusing on the dominant modality for alignment. To leverage this insight, a simple yet effective modality selection algorithm, DYNMOS, is proposed. DYNMOS selects the modality with the highest confidence in its NCC peak for registration, bypassing the need for complex multi-objective optimisation. By prioritising the most

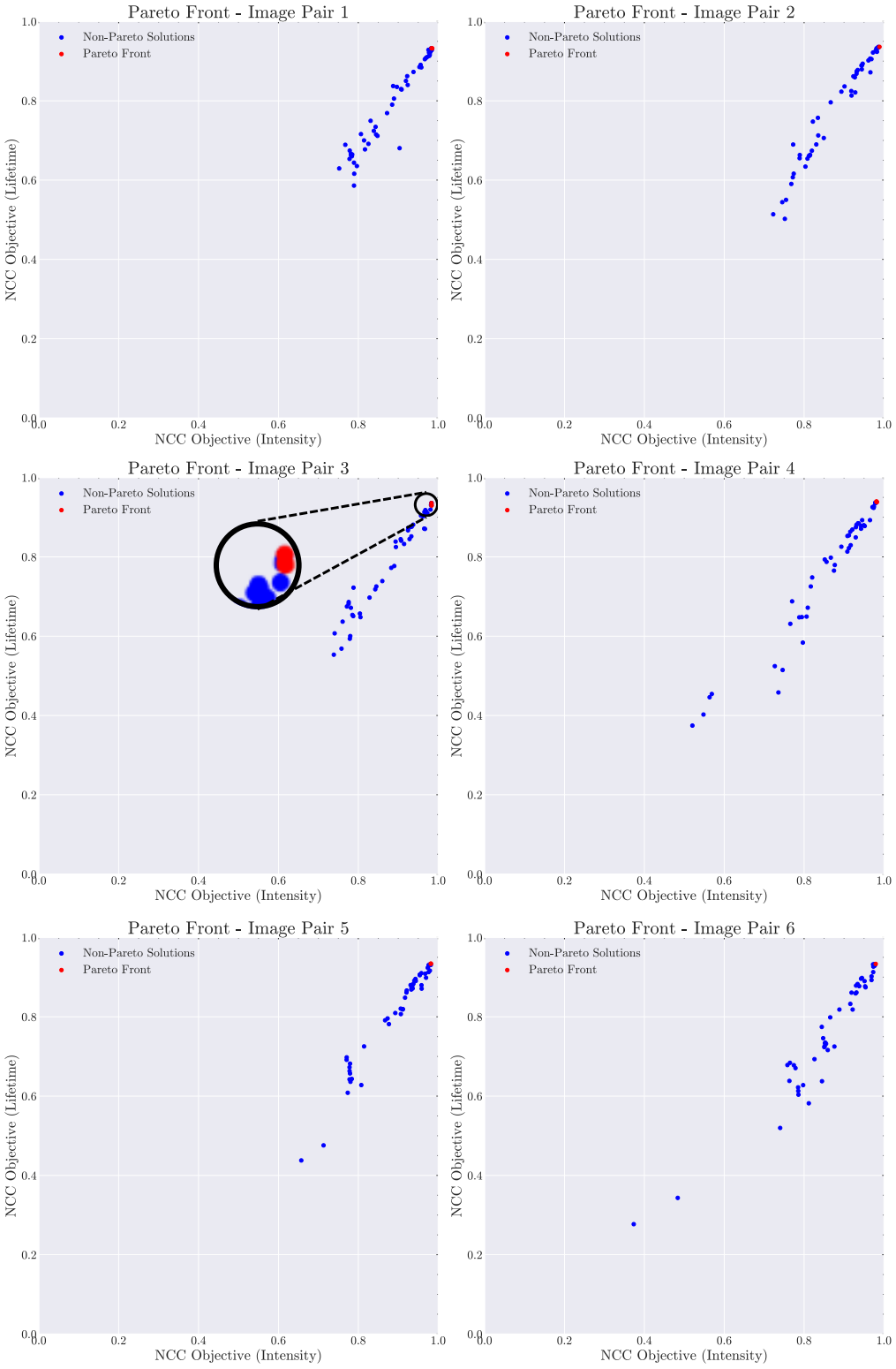


Figure 4.10: Pareto optimisation results for six image pairs showing the distribution of non-dominated solutions based on two objective functions, $f_1(\theta)$ (intensity) and $f_2(\theta)$ (lifetime). Each curve represents the Pareto front for a specific pair. Pair 3 has a minimal front (two solutions), highlighting a minor trade-off, while other pairs exhibit single solutions concentrated along a trend line, indicating correlated objectives and an effective solution for both modalities.

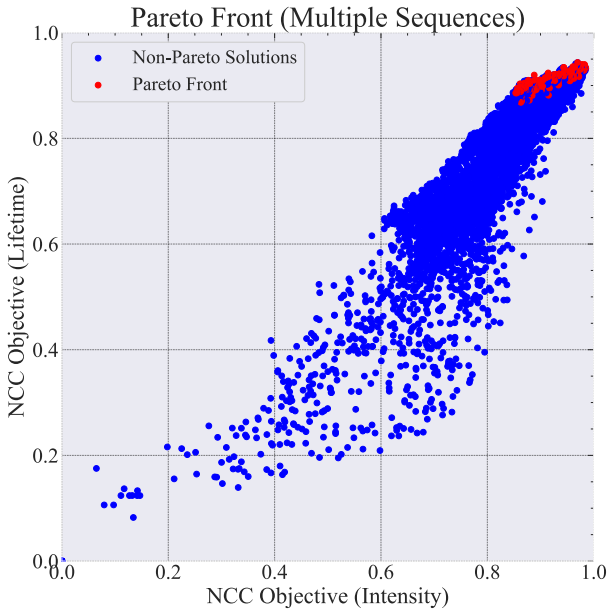


Figure 4.11: Comparison of Pareto fronts across modalities. The slope and clustering patterns of these fronts illustrate the dominance of the intensity objective function $f_1(\theta)$ over $f_2(\theta)$. This offers insights into the alignment behaviour across multiple image pairs. Steeper slopes indicate stronger dominance of intensity over lifetime.

reliable modality, this approach reduces computational overhead and enhances robustness, making it particularly suitable for real-time applications.

The DYNMOS method quantitatively measure and compare the confidence and peak energy between two NCC maps from two modalities can involve the following steps:

1. **Peak Detection**
   - Identify the primary peak location in each NCC map.
   - Measure the peak value, which represents the maximum correlation between the two images for each modality.
2. **Peak Confidence**
   - Compute the ratio of the primary peak value to the average of all other values in the NCC map [117, 113]:

$$C = \frac{p}{\mu_{np}}, \tag{4.18}$$

where, p is the primary peak value, $p = max(C_{map})$,and μ_{np} is the mean of non-peak Values [117, 113]. This provides a measure of how

dominant the primary peak is compared to other potential alignment positions. μ_{np} can be written as:

$$\mu_{np} = \frac{1}{|C_{map}| - 1} \sum_{x \in C_{map}, x \neq p} x, \tag{4.19}$$

3. **Peak Energy**

   - Define the peak energy as the sum of the values in a local neighbourhood (e.g., a 5x5 window) around the primary peak.

$$E = \sum_{(i,j) \in \text{window}(p,w)} C_{map(i,j)}, \tag{4.20}$$

   This metric captures the overall strength and spread of the primary peak.

4. **Comparison**

   - Compute the relative difference in confidence and peak energy between the two modalities:

$$\Delta C = |C_1 - C_2|, \quad \Delta E = |E_1 - E_2| \tag{4.21}$$

This approach, which can be seen summarised in Algorithm 2, ensures that comparing the prominence of both peaks (confidence) and the intensity of the correlation region (peak energy) is done in a systematic manner.

Algorithm 2 Modality Selection for Robust FLIm Multi-Objective Image Registration

Require: Two image modalities: intensity $\mathbf{I}$ and lifetime $\mathbf{L}$, window size w

Ensure: Selected modality for registration based on peak energy and confidence

Compute $\mathbf{C}_{map}$ for both modalities:
$\mathbf{C}_{map(intensity)}$ and $\mathbf{C}_{map(lifetime)}$, using Eq. 3.8
Identify primary peaks in $\mathbf{C}_{map}$:
$p_I = \max(\mathbf{C}_{map(intensity)}), \quad p_L = \max(\mathbf{C}_{map(lifetime)})$
Compute mean of non-peak values for both modalities:
μ_{npI} and μ_{npL}, using Eq. 4.19
Calculate confidence ratios:
C_I and C_L, using Eq. 4.18
Compute peak energy in a $w \times w$ window around the primary peaks:
E_I and E_L, using Eq. 4.20
Compare metrics:
if $(C_I > C_L)$ **and** $(E_I > E_L)$ **then**
    Select intensity modality for registration
    return "Intensity"
else if $(C_L > C_I)$ **and** $(E_L > E_I)$ **then**
    Select lifetime modality for registration
    return "Lifetime"
else
    Select modality with higher combined score: $S_I = C_I + E_I, \quad S_L = C_L + E_L$
    if $S_I > S_L$ **then**
        return "Intensity"
    else
        return "Lifetime"
    end if
end if

4.8 RESULTS

The evaluation of the proposed methods is summarised in Fig. 4.12, which compares image registration performance using QA, SSIM, and NRMSE metrics. For the multi-objective image registration task, the SHGO method demonstrated superior optimisation performance across the different strategies compared to traditional intensity-weighted image registration.

When registering intensity-weighted images directly, the results showed a significant improvement when leveraging the MOO framework. Specifically, MOO-derived intensity-weighted images achieved approximately a 4.65% increase in

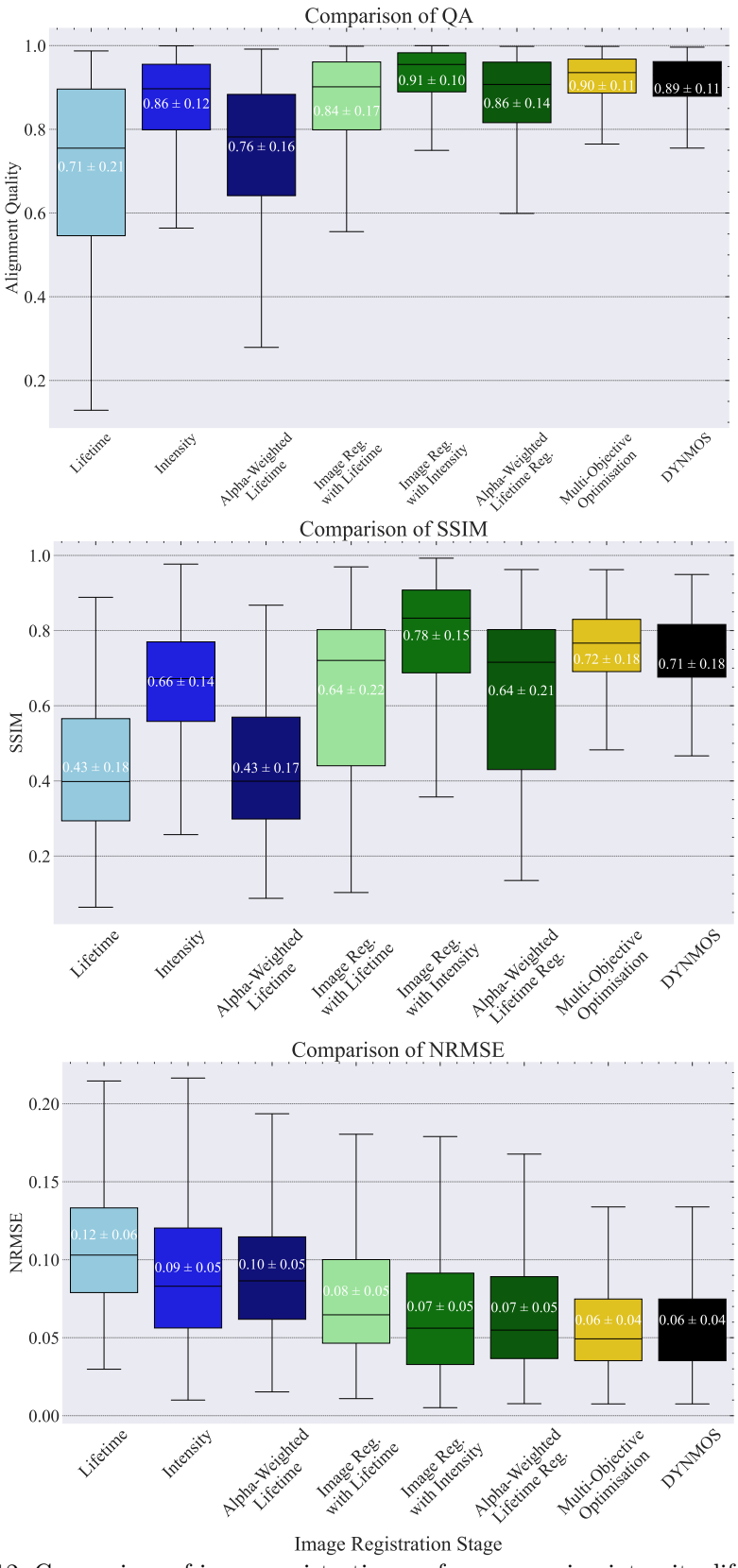


Figure 4.12: Comparison of image registration performance using intensity, lifetime, and intensity-weighted lifetime image representations. The figure highlights the improvements in QA, SSIM, and NRMSE achieved through the MOO framework and demonstrates the near-equivalent performance of the proposed DYNMOS method compared to the best MOO strategy optimised with the SHGO method.

QA, a 12.5% improvement in SSIM, and a 16.67% reduction in NRMSE compared to the conventional approach.

The proposed DYNMOS method achieves comparable performance to the best MOO results obtained using the SHGO method, as evidenced in Fig. 4.12. This near-equivalent performance underscores DYNMOS as a viable alternative for dynamic multi-objective FLIm registration.

Additionally, DYNMOS significantly outperforms SHGO in terms of computational efficiency. With its simple and direct approach, DYNMOS achieves dynamic multi-objective registration in a single iteration, offering a 2.46× faster computational performance compared to SHGO. This improvement is illustrated in Fig. 4.13, evaluated on the same sample of sequences as Fig. 4.5.

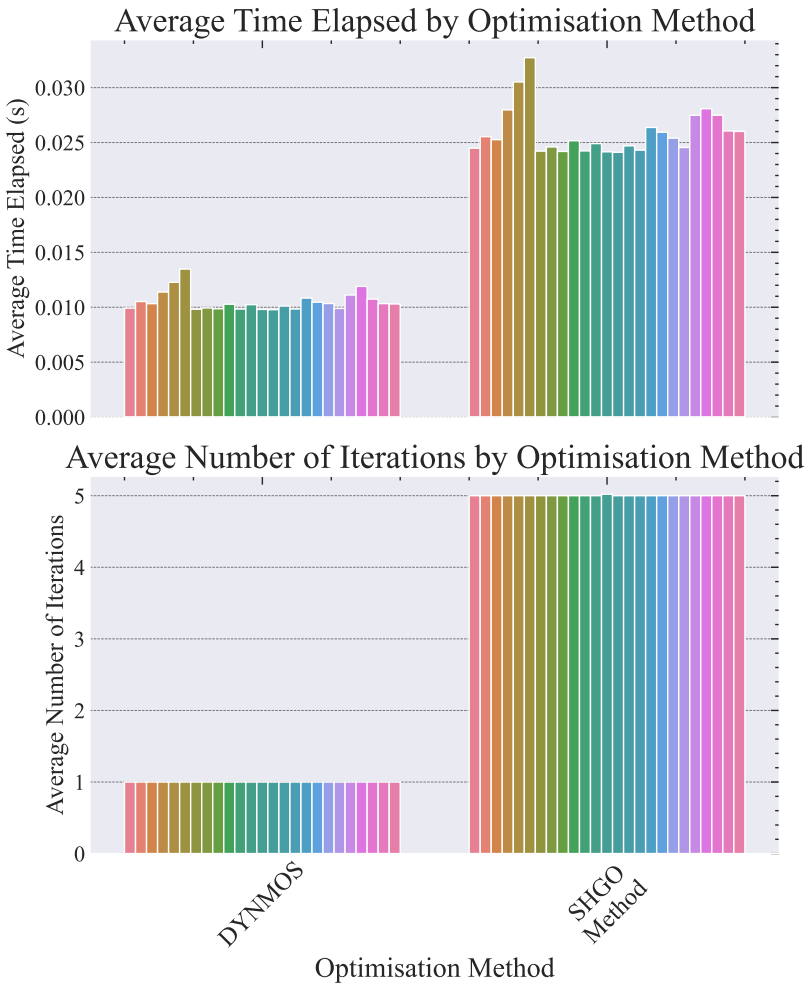


Figure 4.13: Bar plot comparing the average computation time for FLIm multi-objective image registration between the fastest MOO approach using SHGO and the proposed DYNMOS method. Each bar has a distinct colour representing an image sequence containing at least 80 images. Unlike SHGO, DYNMOS achieves registration in a single iteration across all datasets, resulting in a computational speed-up of approximately 2.46×.

4.9 CONCLUSION

This chapter explored various strategies for multi-objective image registration, aiming to integrate intensity and lifetime data to achieve accurate and robust alignment. The approaches outlined in Section 4.3.2 provided a comprehensive framework to evaluate the relative contributions and trade-offs of the two modalities, while highlighting their influence on registration accuracy.

The first strategy utilised a weighted sum formulation, where an optimal weight was determined through iterative exploration. Although straightforward, this approach relied on fixed weighting, limiting its adaptability to variations across image pairs. While effective in specific cases, its fixed nature often resulted in suboptimal performance.

The second strategy framed the problem as a multi-variate optimisation task, jointly optimising the weighting factor and registration parameters. This dynamic approach theoretically offered greater flexibility, adapting to differences in the image data. However, the results consistently revealed a dominance of the intensity modality, which contributed more significantly to alignment accuracy due to its richer structural features and higher correlation strength.

The third strategy adopted a non-linear approach, expressing the objective function as a product of the two modalities, each modulated by a weighting factor. This allowed for non-linear interactions between intensity and lifetime data. Despite this, the optimisation process similarly prioritised intensity, reaffirming its superior reliability and feature quality.

The analysis of these strategies using different optimisation methods demonstrated that the choice of optimisation method plays a critical role in achieving effective multi-objective registration. Global optimisation techniques, particularly SHGO, proved highly effective, efficiently navigating the complex search space to identify the optimal weighting factor ω. By leveraging both exploration and exploitation, SHGO consistently achieved better results compared to other methods.

To better understand the behaviour of the MOO and the dominance of intensity over lifetime, the chapter combined NCC space investigations with Pareto optimisation using NSGA-II. This pairing provided valuable insights into the scenarios where the weighting balance might shift in favour of the lifetime modality. It highlighted the conditions under which the lifetime NCC could dominate, including higher SNR, lower texture homogeneity, and greater feature correlation. These findings offer a deeper understanding of how the choice of weighting is influenced by specific imaging characteristics.

Additionally, the chapter proposed DYNMOS, a dynamic and computationally efficient modality selection algorithm. By prioritising the modality with the

highest confidence in its NCC peak, DYNMOS achieved comparable registration performance to the best MOO strategies while significantly reducing computational overhead. This approach demonstrated a 2.46× speed-up over SHGO, making it a practical solution for real-time applications.

In conclusion, the integration of intensity and lifetime data in image registration can be effectively addressed through multi-objective frameworks. The proposed strategies and methods, combined with insights gained from NCC space and Pareto optimisation, ensure that the final registered images leverage the complementary strengths of both modalities. This work lays the foundation for future advancements in dynamic and adaptive registration frameworks tailored to diverse imaging conditions in FLIm.

5

MACHINE LEARNING APPLICATIONS IN FLIM-OEM DATA PROCESSING

5.1 INTRODUCTION

The integration of machine learning (ML) into biomedical imaging has greatly improved data analysis [103], offering advanced solutions to persistent challenges. This chapter presents ML-driven methodologies aimed at enhancing the reliability of imaging sequences and analyses in OEM-FLIm systems. In particular, it investigates how the TRACER pipeline (Chapter 3) influences two key tasks: automatic neutrophil activation probe (NAP) detection via a U-Net-based deep learning (DL) framework and uninformative frame detection through a DL-transfer learning (TL) model.

Despite OEM-FLIm's potential in real-time *in-vivo* diagnostics, artefacts and motion-induced inconsistencies often generate uninformative frames. Traditional approaches relying on handcrafted features and statistical models lack generality under complex clinical conditions. By contrast, the DL-based solution proposed here leverages TL for uninformative frame detection, thus reducing manual intervention and improving overall system robustness.

Accurate NAP identification in FLIm data is critical for assessing neutrophil activity—an important biomarker of inflammation and immune response [36, 17, 8]. Conventional image analysis methods struggle to detect these signals when background intensity and motion artefacts vary. To address this, we explore the impact of the TRACER pipeline on a U-Net-based DL strategy for NAP detection, demonstrating high accuracy and computational efficiency.

5.1.1 *Motivation*

Findings from Chapter 3 reveal that removing uninformative frames before image registration (Section 3.4.1.1) improves QA by 4.3 %, SSIM by 5.6 %, and reduces NRMSE by 27.6 %. Moreover, Section 3.4.5 shows how these gains enhance the detection of positive NAP signals in FLIm data, further validating the proposed framework.

Although intensity imaging dominated in Chapter 4, lifetime imaging can sometimes provide more robust structural and textural cues. This observation motivates a hybrid approach incorporating both intensity and lifetime data,

prompting a reconsideration of earlier feature-engineering techniques [1] and encouraging new strategies for integrated analysis.

Additionally, as noted in Chapter 2, bacterial imaging methods have seen limited application to NAP detection in FLIm. While object detection frameworks exist, their effectiveness in this context remains unverified, creating an opportunity to explore their viability for NAP identification. Fig. 5.1 illustrates examples of NAP regions, highlighting the differences between intensity and lifetime imaging.

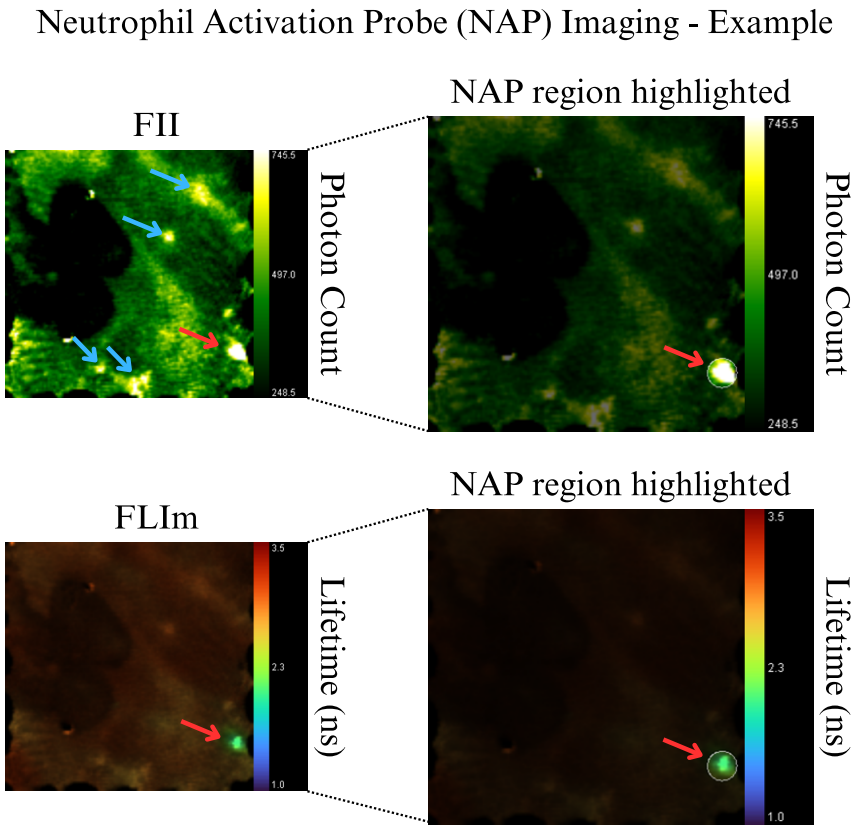


Figure 5.1: Example of a FII (top row) and FLIm (bottom row) from a dataset illustrating NAP imaging. Multiple region of interest (RoI)s (blue arrows) appear in the FII image, while one region (red arrow) exhibits a lower lifetime in the FLIm image, suggesting activation of the NAP probe. The zoomed-in views on the right highlight the importance of combining both modalities: FII alone may not fully capture bioenvironmental changes, whereas FLIm provides additional contrast through lifetime differences that help confirm probe activation.

5.1.2 *Problem Statement*

This chapter critically evaluates the advantages and limitations of ML-based approaches for uninformative frame detection and NAP identification, providing

insights into their empirical performance on real OEM-FLIm datasets. Based on these findings, recommendations are proposed for integrating these models into clinical workflows to improve diagnostic precision and streamline data processing. The chapter aims to address the following research questions:

- Can a DL approach leveraging TL improve the accuracy and generalisability of uninformative FLIm frame detection?
- Can a U-Net-based DL model be effectively applied to automate NAP region detection in FLIm imaging,
- and how does image registration or TRACER impact the automatic detection of NAP regions in FLIm?

5.1.3 *Chapter Overview*

This chapter explores the integration of ML techniques into OEM-FLIm data processing. Conventional FLIm analysis relies on predefined feature extraction and statistical modelling, which can be limited by noise, motion artefacts, and variability in fluorescence signals. ML presents an opportunity to enhance feature extraction, segmentation, and classification by learning data-driven representations that adapt to complex imaging conditions.

The chapter begins with Section 5.2 an overview of ML methodologies relevant to OEM-FLIm and other medical imaging modalities, focusing on DL approaches such as convolutional neural network (CNN) and TL. A comparative analysis is conducted in Section 5.2.2 between data-driven machine learning techniques and traditional model-based approaches, highlighting their respective advantages and limitations in handling FLIm datasets.

Section 5.2.3 demonstrates the application of DL for uninformative frame removal, a critical step in improving FLIm data quality. A TL framework is developed to automate this task, leveraging pre-trained networks to enhance generalisation while minimising computational overhead. Additionally, Section 5.3 explores advanced strategies for the automatic detection of NAP, before focusing on the importance of a robust image processing pipeline in Section 5.4 to further refine the application of ML in classification tasks relevant to OEM-FLIm. Finally, the chapter concludes with a discussion of potential future research directions aimed at refining these techniques and extending their applicability to OEM-FLIm imaging contexts.

5.2 MACHINE LEARNING FOR OEM-FLIM PROCESSING

Uninformative frames, defined in Chapter 2, Section 2.4.1, pose a major challenge in OEM-FLIm, where poor probe–tissue contact or significant motion at acquisition can produce frames devoid of meaningful structural or functional content [1]. These corrupted frames often increase computational overhead and confound quantitative analyses. Prior work by Perperidis *et al.* [1] showed that uninformative frames could account for over 25 % of OEM datasets, and proposed a gaussian mixture modelling (GMM) with grey-level co-occurrence matrix (GLCM)-based texture descriptors to exclude these frames. Although this strategy achieved high sensitivity and specificity, it focused on OEM-FII or intensity-based imaging, and the extension of the method to OEM-FLIm remains unexplored.

As discussed in Chapter 2, motion-induced artefacts broadly affect imaging modalities ranging from PET/CT to advanced microscopy [37, 99, 45, 98]. For OEM-FLIm, discarding uninformative frames can enhance registration accuracy (Chapter 3) and improve the detection of clinically relevant signals such as NAP activation. Fig. 5.2 offers a snapshot of these uninformative frames in FLIm, underscoring the need for automated solutions that can handle large datasets with minimal manual oversight.

To address these issues, this chapter explores ML-based approaches for automating uninformative frame detection and improving the overall TRACER pipeline. Moreover, ML can also bolster downstream tasks such as NAP detection by leveraging robust, data-driven models that reduce reliance on handcrafted features. From a clinical and analytical perspective, the success of the preprocessing and registration stages is ultimately judged by their impact on downstream tasks, such as frame selection, segmentation, and feature detection, rather than by registration accuracy alone. Therefore, the robustness of the image processing pipeline will play a crucial role in maximising the reliability and performance of ML applications in OEM-FLIm.

5.2.1 *From Artificial Intelligence to Transfer Learning*

Today, artificial intelligence (AI) is a rapidly growing domain aimed at creating systems capable of tasks typically associated with human intelligence. ML, a subset of AI, enables machines to learn patterns from data and make accurate predictions without explicit programming. Arthur Samuel first coined the term in 1959 [151], and Mitchell later formalised ML as:

> A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E [152].

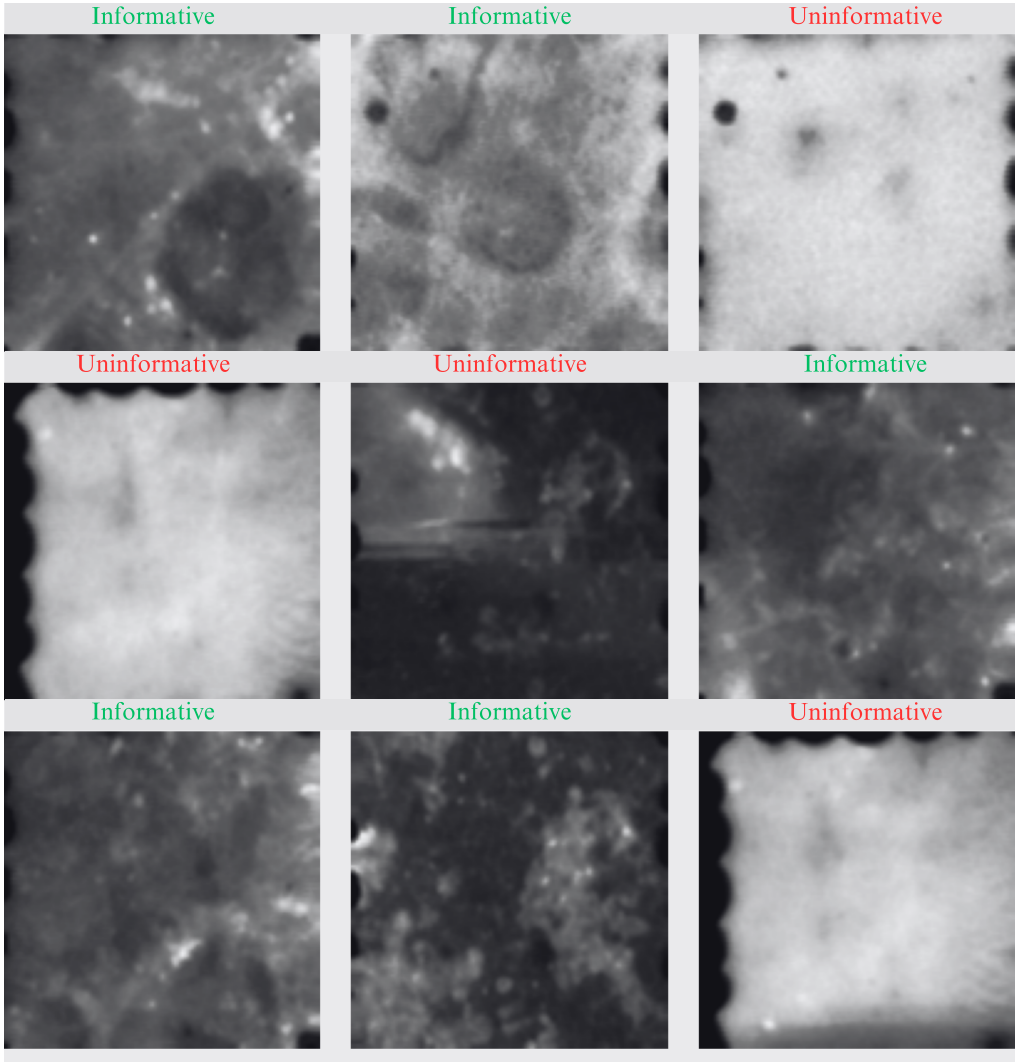


Figure 5.2: Example of "informative" and "uninformative" frames in the OEM-FLIm imaging data obtained using the KronoScan system (Chapter 2, Section 2.5).

In essence, ML leverages examples (training data) and quantitative performance metrics to iteratively refine its predictive capabilities. This data-driven perspective aligns well with the real-time and high-volume nature of FLIm acquisition, where rapid analysis can be critical.

ML algorithms can be broadly categorised into supervised and unsupervised learning, depending on whether labels are provided during training. In supervised learning, each input example is associated with a label or target; for instance, in a spam filter, the training samples are emails (inputs) labelled as either spam or not spam (outputs). By contrast, in unsupervised learning, the algorithm receives only input data and attempts to discover hidden patterns or structures. Applications include clustering (*e.g.,* K-means), dimensionality reduction, or density estimation [153, 103]. Additionally, semi-supervised learning combines aspects of the above

two paradigms, while reinforcement learning employs rewards or penalties to guide the training process.

Traditional ML approaches include regression (linear or logistic), random forests, K-means, or naive Bayes, whereas DL (based on neural networks) is the most prominent family of ML methods today [153]. In recent years, ML has been applied to numerous FLIm tasks [10], including lifetime estimation and tissue classification. For example, Wang *et al.* [154] employed a deep network to generate virtual hematoxylin and eosin staining directly from label-free autofluorescence lifetime images, while Chen *et al.* [155] developed a generative adversarial network-based framework that efficiently reconstructs high-quality FLIm outputs under low photon counts. However, none of these studies addressed the crucial problem of identifying uninformative frames or detecting NAP regions.

Within this broader ML landscape, TL is particularly appealing for biomedical imaging [153, 103], including OEM-FLIm. Instead of training ML models from scratch, TL methods transfer knowledge from models pre-trained on large, diverse datasets (often natural images) to a target domain of interest. This approach is especially valuable in OEM-FLIm, where annotated data can be expensive or hard to obtain and domain gaps may be substantial. Although recent domain adaptation studies in medical imaging [156] reveal promising directions, the feasibility and efficacy of such strategies in tackling uninformative FLIm frames and NAP detection require further investigation. Drawing on the foundational ideas in [1], the subsequent sections describe how TL can be integrated into our TRACER pipeline, paving the way for robust, data-driven solutions in OEM-FLIm image processing.

5.2.2 *Model-Driven vs. Data-Driven Approaches*

The task of detecting uninformative frames in FLIm data can be approached through two fundamentally different paradigms: model-driven methods, which rely on predefined mathematical formulations [157, 158], and data-driven techniques, which infer patterns from large datasets [159, 160]. Both approaches have their advantages and limitations, and the choice between them often depends on the specific application constraints, such as data availability, interpretability, and computational efficiency [161, 162].

Model-driven methods are grounded in theoretical formulations that describe system behaviour through structured mathematical models [157, 162]. These models are typically based on well-established physical or biological principles, offering interpretable parameters that correspond to real-world phenomena. Examples include regression models, differential equations, and compartmental models, all of which encode prior knowledge about the system — E.g.,: Chen et.

al [158] presented Learned Experts' Assessment-Based Reconstruction Network (LEARN), a novel DL framework for sparse-data CT reconstruction. By unrolling iterative reconstruction into a trainable CNN, it achieves state-of-the-art artefact suppression and computational efficiency — A key strength of model-driven approaches lies in their ability to perform well with relatively small datasets, provided that the underlying assumptions hold [157, 161, 158]. However, these assumptions can also be a major limitation—if the system deviates significantly from the modelled behaviour or exhibits high non-linearity, the model's predictive power may be compromised. Furthermore, traditional modelling frameworks often struggle with high-dimensional or highly variable data, requiring extensive manual adjustments or simplifications [161].

In contrast, data-driven methods, particularly machine learning-based approaches, forgo explicit assumptions about the system and instead learn patterns directly from the data [163, 160]. This adaptability makes them particularly suited for complex, high-dimensional problems where predefined models would be insufficiently expressive. DL architectures, such as CNN, have demonstrated remarkable success in identifying intricate patterns in imaging data, often surpassing handcrafted feature extraction techniques. E.g., Gong et. al [163] presented a DL-based PET denoising framework that outperforms traditional methods by leveraging perceptual loss and fine-tuning with real data. The combination of residual CNN, pretraining on simulated data, and domain adaptation through fine-tuning makes this approach highly effective for reducing noise while maintaining diagnostic quality. However, this flexibility comes at the cost of interpretability; while model-driven approaches offer direct physical meaning to parameters, data-driven models operate as "black boxes," making it difficult to trace how decisions are made [161]. Additionally, the success of ML solutions is heavily dependent on the availability of large, high-quality labelled datasets. Without sufficient data, these models risk overfitting and failing to generalise to unseen cases [161, 7].

Performance considerations also differentiate the two approaches. Model-driven techniques can achieve high accuracy when the underlying assumptions closely reflect reality, while data-driven models excel in scenarios where system behaviour is difficult to formalise [161, 162]. Furthermore, computational efficiency varies significantly between these methods. Traditional modelling approaches are often computationally lightweight and can run on standard hardware, whereas DL models, particularly those involving large neural networks, require significant computational resources and specialised hardware such as GPUs [153].

The choice between model-driven and data-driven approaches ultimately depends on the nature of the problem at hand. In well-characterised systems where theoretical insights provide a reliable basis for modelling, traditional approaches can offer efficiency and interoperability. Conversely, in scenarios involving com-

plex, non-linear, or poorly understood phenomena, data-driven methods often excel by discovering rich patterns without explicit assumptions regarding the underlying physics or biology. A hybrid strategy, leveraging both domain knowledge and ML, may offer a robust compromise, combining the interpretability of model-driven methods with the flexibility and predictive power of data-driven techniques [162].

In the context of pulmonary OEM, Perperidis et al. [1] employed a GLCM-based texture analysis and a GMM clustering scheme to identify informative versus uninformative frames. Their method extracted principal components from handcrafted features and used GMM thresholding to classify frames into clusters, assuming a consistent intensity distribution across different acquisitions. However, adapting this approach to diverse FLIm datasets is challenging, as the original feature distribution may not hold and thus risks misclassifying frames when imaging conditions vary. By contrast, ML, particularly DL, can accommodate substantial data heterogeneity and non-linear relationships, albeit at the cost of requiring large and well-curated training sets. Deep networks (*e.g.*, CNNs) introduce thousands of parameters capable of representing highly complex data patterns, potentially surpassing handcrafted feature extraction in performance.

Ultimately, the decision on which approach to adopt is governed by several considerations:

- **Nature of the Problem**: Traditional models may suffice for well-understood systems, whereas DL-based methods can better capture complexities in poorly characterised domains.
- **Data Availability**: ML thrives on large, high-quality datasets; conventional models can often deliver useful insights with smaller datasets.
- **Interpretability**: Model-driven techniques tend to yield interpretable parameters, while data-driven approaches may offer superior predictive accuracy at the expense of explainability.
- **Computational Resources**: Deep ML models often demand significant computing power, whereas most traditional algorithms can be executed on standard hardware.

5.2.3 *A New Approach for Uninformative Frame Detection*

Despite the strengths and weaknesses of each paradigm, this thesis proposes a TL framework to automate uninformative frame detection in FLIm. Unlike purely model-driven methods, TL obviates the need for large-scale domain-specific datasets by transferring pre-trained features from well-established architectures

(*e.g.,* VGG16 [164], InceptionV3 [165], and ResNet50 [166]) to the FLIm domain. As outlined in Fig. 5.3, this strategy not only alleviates the burden of collecting vast amounts of labelled data, but also benefits from the powerful representational capacity of deep networks, thereby supporting robust binary frame classification under varying imaging conditions.

For the task of categorising frames into "informative" and "uninformative," three well-established convolutional neural network architectures, namely, VGG16, InceptionV3 , and ResNet50 were evaluated via transfer learning. All three models incorporate pre-trained weights from ImageNet [167], providing a rich set of learned features that substantially reduce the amount of data and training time required. InceptionV3 incorporates innovative factorised convolutions and parallel Inception modules, whereas ResNet50 leverages skip (residual) connections to facilitate training of deeper networks. Both designs demonstrate high accuracy and strong performance in a range of computer vision tasks, particularly when sufficient training data and computational resources are available. By contrast, the VGG16 architecture is comparatively simpler, characterised by a straightforward stack of convolutional layers with small kernel sizes (3×3) and occasional max pooling [164]. This simplicity reduces hyperparameter complexity and improves interpretability, making it easier to adapt and fine-tune.

The training dataset used for this investigation comprises 135 OEM-FLIm imaging sequences recorded with the KronoScan system (Chapter 2, Section 2.3.1), totalling 12 780 individual frames. Of these frames, 400 were annotated by clinical experts via a Python-based classification tool, which prompts the user to label a frame as "uninformative" (1) or "informative" (0) according to the criteria detailed in Chapter 2, Section 2.4.1. This results in a label file mapping each frame's ID to its corresponding class. Fig. 5.2 illustrates representative frames judged by expert annotators to be "uninformative".

Leveraging these labelled data, the learned weights from each base network (VGG16, InceptionV3, or ResNet50) are transferred into the proposed TL model, depicted at the bottom of Fig. 5.3. Batches of FLIm frames are fed into the "frozen" feature-extraction layers of the pre-trained network, ensuring stable representations and minimising overfitting when the available data are limited [168, 169]. The resulting feature maps are then passed to the trainable classification head, comprising a Global Average Pooling 2D layer to reduce spatial dimensionality, followed by a dense layer with $ReLU$ activation and dropout for regularisation, and concluding with a sigmoid output for binary classification. This setup is initially trained for 50 epochs, after which the last four layers of the base network are unfrozen to allow domain-specific fine-tuning. This second phase, also lasting 50 epochs, further adapts the model's higher-level feature

representations to FLIm data variability, while limiting trainable parameters to mitigate the risk of catastrophic forgetting or overfitting.

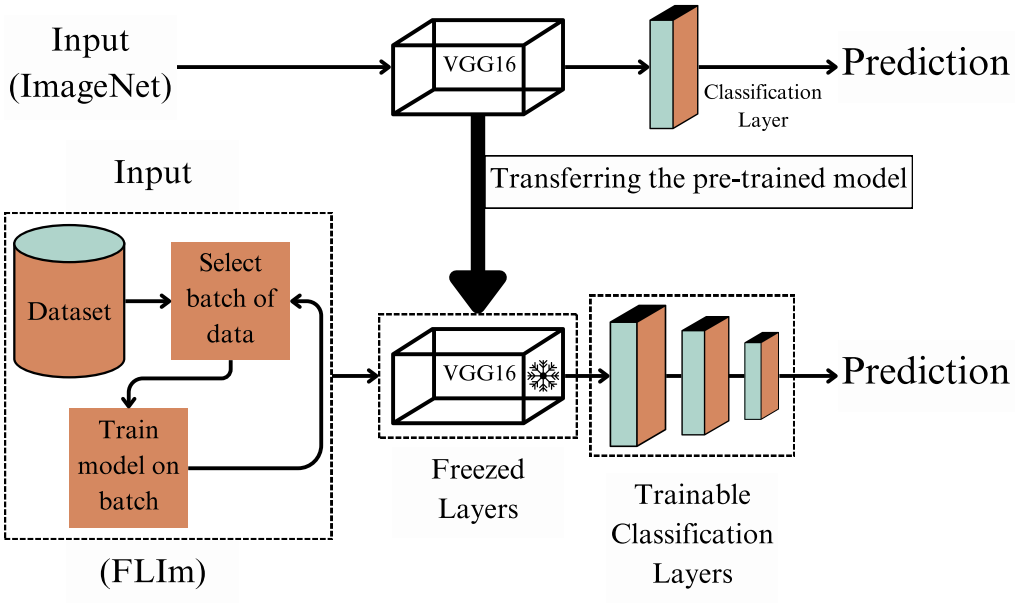


Figure 5.3: Illustration of the TL process using VGG16 pre-trained on ImageNet (top). The learned weights from VGG16 are transferred to the proposed TL model (bottom), which ingests batches of FLIm data for more efficient training. The bottom section shows the frozen VGG16 feature-extraction layers and the trainable classification layers, consisting of a Global Average Pooling 2D layer, a dense layer, ReLU activation with dropout, and a final sigmoid layer for binary classification. This design enables effective identification of uninformative frames in large FLIm datasets.

5.2.3.1 *Mathematical Formulation of the Binary Classification Task*

In the TL approach for uninformative frame detection, we define a training set

$$\mathcal{D} = \{(\mathbf{I}_i, y_i)\}_{i=1}^{N}, \tag{5.1}$$

where each image (frame) $\mathbf{I}_i$ has an associated label $y_i \in \{0, 1\}$. We use $y = 0$ to indicate an "informative" frame and $y = 1$ for an "uninformative" frame. In typical CNN-based pipelines, $\mathbf{I}_i$ is a three-dimensional tensor of size (H, W, C).

When transferring from a pre-trained network (e.g. VGG16, InceptionV3, or ResNet50), we adapt its top layers to produce a scalar probability $\hat{y}_i$. Formally, let

$$f(\mathbf{I}; \boldsymbol{\theta}) = \hat{y} \in [0, 1] \tag{5.2}$$

be the adapted network's output, where $\boldsymbol{\theta}$ encompasses the parameters of both the frozen and fine-tuned layers. Typically, $\hat{y}_i$ is computed as:

$$\hat{y}_i = \sigma(z_i), \tag{5.3}$$

with z_i the final logit (dense layer output) and $\sigma(\cdot)$ the sigmoid activation:

$$\sigma(z) = \frac{1}{1+e^{-z}}. \tag{5.4}$$

If $\hat{y}_i \geq 0.5$, we classify the frame as uninformative ($y = 1$); otherwise, it is deemed informative ($y = 0$).

To train this TL model, we minimise the Binary cross entropy (BCE) loss, quantifying how well the predicted probabilities match the ground truth labels. When the predictions are correct, the BCE is low, and when they are wrong, the BCE is high:

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{N}\sum_{i=1}^{N}\Big(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\Big). \tag{5.5}$$

This objective measures the discrepancy between predicted probabilities $\hat{y}_i$ and true labels y_i. The training procedure (*e.g.,* using Adam [170] or another gradient-based optimiser) iteratively updates $\boldsymbol{\theta}$ to minimise $\mathcal{L}$.

Once trained, the model's predictions $\hat{y}_i$ are evaluated against ground truth y_i. Common metrics include:

- **Accuracy (ACC):** The fraction of frames correctly classified as either informative or uninformative:

$$\text{ACC} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}(\tilde{y}_i = y_i), \tag{5.6}$$

  where $\tilde{y}_i \in \{0,1\}$ is the binary prediction after thresholding $\hat{y}_i$, and $\mathbb{I}(\cdot)$ is the indicator function.

- **Precision:** The proportion of predicted "uninformative" frames ($\tilde{y} = 1$) that are correct:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}. \tag{5.7}$$

- **Recall (Sensitivity):** The proportion of ground-truth "uninformative" frames that are correctly identified:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}. \tag{5.8}$$

- **Area Under the ROC Curve:** The receiver operating characteristic (ROC) curve plots the true positive rate (Recall) against the false positive rate at varying thresholds of $\hat{y}_i$. The area under the curve (AUC) summarises the overall ability of the classifier to distinguish between the two classes across all possible decision thresholds.

By transferring knowledge from a large-scale, pre-trained feature extractor and applying a sigmoid output layer, the network learns a probability function $\hat{y}_i = f(\mathbf{I}_i; \boldsymbol{\theta})$ for the "uninformative" label. Through iterative training to minimise BCE, the TL approach refines $\boldsymbol{\theta}$ for FLIm frame classification, optimising performance in terms of accuracy, precision, recall, and AUC.

5.2.4 *Results*

Table 5.1 summarises the performance of the three transfer learning architectures, mentioned in Section 5.2.3 for uninformative frame detection in FLIm. Among VGG16, InceptionV3, and ResNet50, VGG16 yields the highest accuracy, precision, and recall, as well as the lowest BCE loss. These metrics are reported along with an epistemic (model) uncertainty estimate using Monte Carlo (MC) dropout [171], which indicates that VGG16 achieves not only strong mean performance but also low variance compared with the deeper networks.

Table 5.1: A table comparing the performance of VGG16, InceptionV3, and ResNet50 in the proposed transfer learning model for uninformative frame detection.

Metric*	VGG16	InceptionV3	ResNet50
Accuracy ↑	**0.988 ± 0.001**	0.981 ± 0.001	0.923 ± 0.123
Precision ↑	**0.976 ± 0.001**	0.961 ± 0.001	0.809 ± 0.159
Recall ↑	**0.974 ± 0.005**	0.943 ± 0.001	0.826 ± 0.126
BCE Loss ↓	**0.047 ± 0.001**	0.062 ± 0.002	0.158 ± 0.052

* An epistemic uncertainty (model uncertainty) aware assessment, done using an MC dropout strategy [171].

Several factors may explain VGG16's superior results. First, its relatively shallow depth and uniform layer design reduce hyperparameter complexity, allowing the model to adapt quickly to domain-specific patterns even with a moderate dataset. Second, the standardised convolutional blocks help limit

overfitting in scenarios of limited inter-class variation, as is often the case in binary frame classification. In contrast, while deeper architectures such as InceptionV3 or ResNet50 offer potential advantages in large-scale and high-variance tasks, they can demand more data and computational resources to converge effectively, which may be a limiting factor in OEM-FLIm pipelines.

A key strength of employing MC Dropout is the ability to quantify prediction uncertainty in addition to point estimates. In Table 5.1, the standard deviations convey how stable each architecture is under random dropout sampling. VGG16's relatively low variance across accuracy, precision, recall, and loss suggests that it not only achieves high performance on average but also maintains consistency across multiple inference passes. In contrast, deeper models exhibit a higher variance, implying greater sensitivity to input perturbations, network initialisations, or dropout masks. From an operational standpoint, this indicates that clinicians or researchers can have added confidence when relying on VGG16 for automated frame filtering.

Figure 5.4 shows the training and validation curves for accuracy, loss, precision, and recall over multiple epochs, highlighting the validation metrics often exceeding those in training. While this behaviour is less common, it has been reported in the presence of regularisation and stochastic training effects, where mechanisms such as dropout and data augmentation introduce noise or constraints during training that are not active during evaluation, thereby reducing apparent training performance relative to validation performance [172]. This effect may be further amplified by the relatively small validation set used in this experiment, resulting in increased metric variance. Notably, both the training and validation losses converge to low values for the VGG16 model, suggesting stable optimisation rather than evidence of superior generalisation, and motivating a cautious interpretation of the observed metric differences.

Figure 5.5 provides sample outcomes of the final TL model on the same FLIm data shown in Fig. 5.2. Each sub-image is labelled with the ground truth (0 for "informative" or 1 for "uninformative") alongside the predicted class, illustrating that the model correctly classifies most frames in practice. These qualitative results corroborate the quantitative findings in Table 5.1, showcasing the robustness of DL methods in identifying corrupted or low-value FLIm frames.

It is instructive to recall the improvement in registration accuracy (Chapter 3, Section 3.4.1.1) realised by removing uninformative frames prior to image registration, leading to a 4.3 % gain in QA, a 5.6 % boost in SSIM, and a 27.6 % reduction in NRMSE across all tested methods. This underscores the practical value of reliable uninformative frame detection in the broader TRACER pipeline.

Figure 5.6 presents the confusion matrix for the TL model on the entire test set. Notably, the true positives (310) and true negatives (28) far exceed

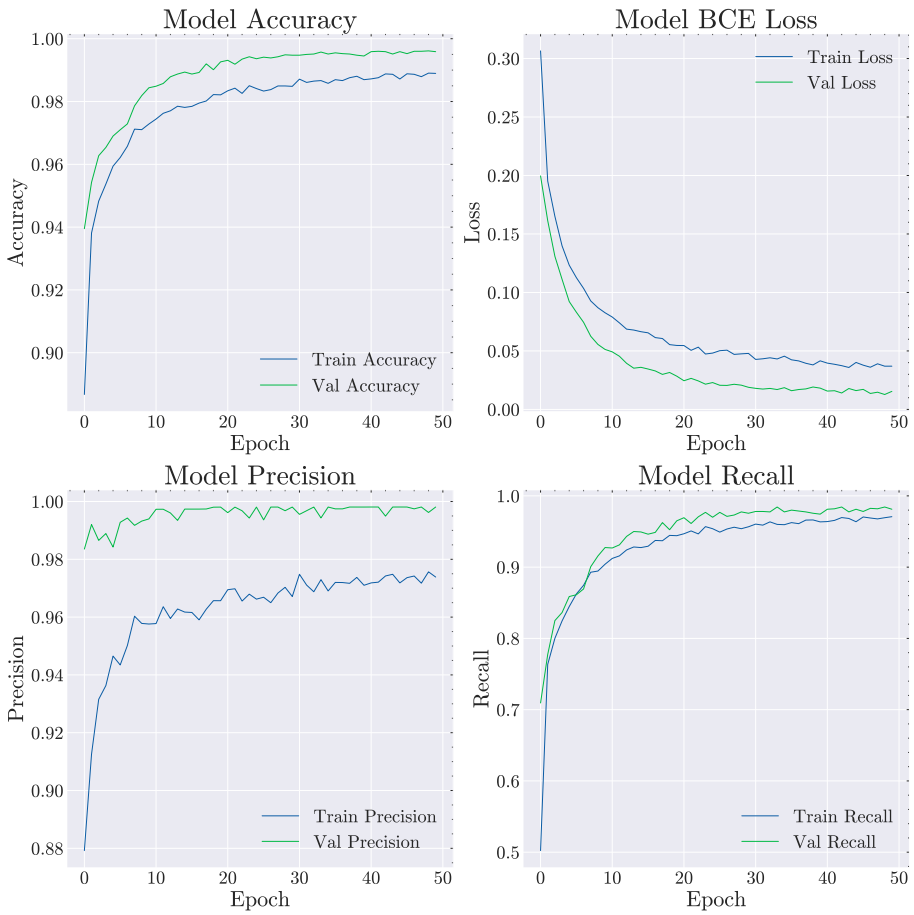


Figure 5.4: Training and validation curves for accuracy, loss, precision, and recall during fine-tuning of the TL model over multiple epochs. Notably, the validation metrics often exceed the training metrics, a behaviour that can arise when regularisation strategies such as dropout or data augmentation are applied during training but not during evaluation [172]. Given the limited validation sample size, these differences are interpreted cautiously and do not imply superior generalisation performance.

the misclassifications (7 false positives, 11 false negatives), indicating that the model is highly effective in distinguishing "informative" from "uninformative" frames. Such strong sensitivity and specificity are particularly relevant when poor-quality data can significantly degrade downstream analyses, as exemplified by the registration improvements highlighted above.

Table 5.2 compares these results to the feature-engineering approach in [1]. While both methods yield high sensitivity and specificity, the TL-based approach achieves a 2.85 % gain in sensitivity and a 5.28 % gain in specificity. Beyond numerical improvements, adopting a data-driven paradigm removes the reliance on handcrafted features and Gaussian mixture clustering, thus offering greater adaptability to evolving FLIm acquisition settings. In practical terms, a higher

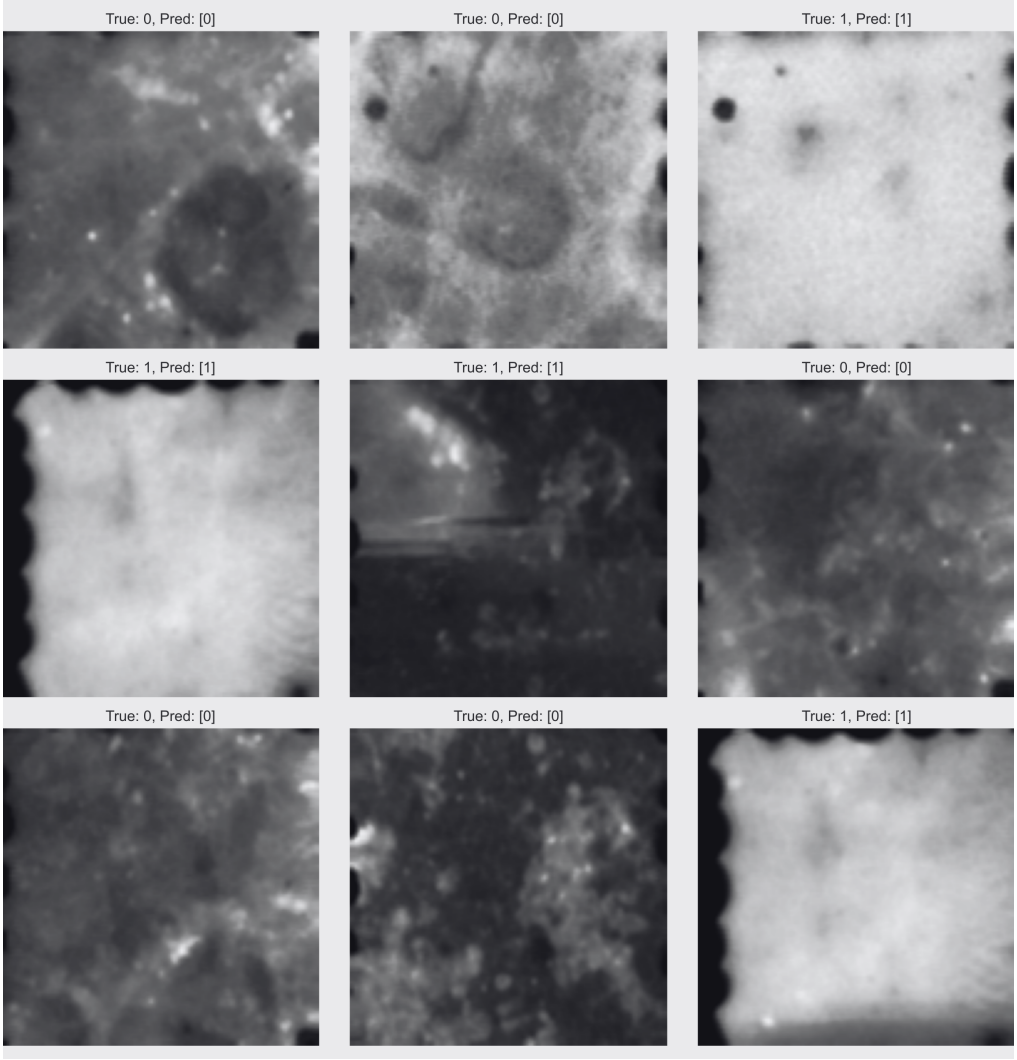


Figure 5.5: Sample outputs from the TL model evaluated on the same FLIm images shown in Fig. 5.2. Each sub-image is annotated with the true class (*0* for "informative" or *1* for "uninformative") and the model's predicted label. Accurate classifications (true label matches predicted label) highlight the effectiveness of the TL approach in identifying uninformative frames.

specificity means fewer "good" frames are erroneously discarded, while a stronger sensitivity ensures that most genuinely unusable frames are filtered out.

Overall, these findings demonstrate that even with a constrained dataset, TL can significantly boost classification effectiveness over purely handcrafted or conventional model-driven methods. In particular, VGG16 offers a compelling balance of interpretability, speed of training, and predictive accuracy, making it the most suitable architecture in this domain. This improvement in uninformative frame detection strengthens the reliability of the broader data-cleaning process in the FLIm imaging pipeline, facilitating downstream tasks such as image registration and NAP detection.

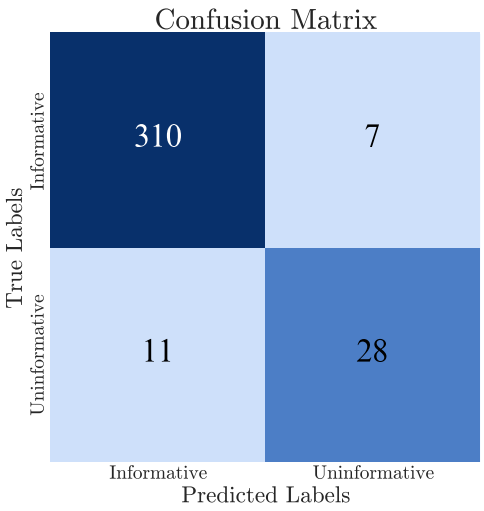


Figure 5.6: Confusion matrix obtained by applying the TL model to the entire test set. The high true positive (310) and true negative (28) rates indicate the effectiveness of the proposed approach, with only a small fraction of misclassifications (7 false positives, 11 false negatives). These results confirm the model's reliability in distinguishing informative from uninformative FLIm frames.

Table 5.2: Comparing the TL-based approach to the one reported in [1].

Metric	Feature-Engineering [1]	TL-based approach
Sensitivity	93 %	**95.65 %**
Specificity	92.6 %	**97.49 %**

5.3 AUTOMATIC DETECTION OF NEUTROPHIL ACTIVATION PROBE

As stated in Section 5.1, the principal objective of this chapter is to demonstrate the impact of the overall image processing pipeline, rather than to propose a novel or superior framework for detecting NAP signals in FLIm. Nonetheless, the choice of a U-Net architecture reflects its extensive popularity in biomedical image segmentation and its widely reported advantages, including efficient end-to-end training and strong performance in pixel-level classification tasks [173, 174]. By adopting such a mainstream model, we provide an illustrative validation of how reliable preprocessing and filtering contribute to consistent NAP detection in FLIm data.

NAP regions exhibit distinct patterns in both FII and FLIm. Empirical observations suggest that NAP signals have higher photon counts (above 500) in intensity (FII) images, whereas in FLIm, the corresponding lifetime values typically lie between 2.0 ns and 2.5 ns, which is considerably lower than surrounding tissue. This dual signature has been exploited in the built-in NAP detection

algorithm of the KronoScan system, whereby candidate pixels are deemed "informative" if they satisfy a maximum-lifetime threshold (2.5 ns) and percentile criteria for both lifetime and intensity. Specifically, pixels exceeding 20 % of the frame's green-lifetime distribution or below 98 % of its green-intensity range are discarded; the surviving pixels are grouped via a flood-fill step into clusters. Clusters containing fewer than five pixels (or failing an optional mode-based filtering process) are subsequently removed to reduce false positives from random noise.

Although this built-in method can capture gross features of NAP signals, it often produces clusters diverging significantly from expert-annotated ground truth. In particular, threshold-based or mode-based statistical filtering may fail to capture subtle morphological boundaries when the underlying distributions vary across frames or subjects. Inconsistent illumination, motion-induced artefacts, and inter-subject variability in lifetime values further complicate a purely heuristic approach.

To overcome these limitations, a U-Net-based DL approach is employed, which offers a more robust, data-driven mechanism to learn segmentation directly from examples of NAP annotations. While many advanced segmentation models exist, U-Net remains a standard reference in biomedical imaging research [175], known for its encoder-decoder structure with skip connections that preserve fine-grained spatial information. This architecture requires only a moderately sized annotated dataset to train effectively, aligning with our aims of both evaluating the image processing pipeline and illustrating how DL can outperform purely statistical methods in NAP detection.

5.3.1 *U-Net-Based Framework*

Unlike the uninformative frame detection task, which leveraged a larger dataset and a TL approach, here the dataset for NAP detection is notably smaller. Specifically, we use three OEM-FLIm sequences recorded with the KronoScan system (Chapter 2, Section 2.3.1), yielding 210 frames. Each frame was annotated using KronoScan's built-in tool, which allows clinical experts to draw circles around NAP regions of interest. These annotations were then converted into binary masks, where each pixel is classified as either "NAP" (1) or "non-NAP" (0). Because the available annotations are limited in number, transfer learning becomes less applicable, and an end-to-end segmentation architecture is employed instead.

U-Net [173] is a popular CNN architecture designed for biomedical image segmentation, consisting of a symmetrical encoder-decoder layout with skip connections between corresponding layers. In the encoder (contracting) path, multiple

3×3 convolution layers and 2×2 max-pooling operations progressively extract higher-level features at reduced spatial resolutions. The decoder (expanding) path employs a series of deconvolutions (often 2×2), merging high-level features with shallower feature maps through skip connections. These skip connections help preserve fine-grained spatial information, crucial for identifying subtle boundaries or smaller target structures such as NAP signals.

U-Net's encoder-decoder paradigm is well-suited to segmenting structures where both global context (*e.g.,* background variation in intensity or lifetime) and local details (*e.g.,* pixel-level NAP features) are important. In contrast to purely classification-based DL models, U-Net outputs a pixel-wise mask, allowing us to directly visualise and quantify NAP regions. Despite its depth, U-Net's inherent design—particularly the skip connections—enables effective training on relatively small datasets, which aligns with the practical constraints of FLIm-based annotation. Consequently, U-Net often exhibits high performance without requiring prohibitively large amounts of labelled data.

Figure 5.7 illustrates the proposed U-Net framework for NAP detection. The network input is a FLIm frame of size $128 \times 128 \times 3$, where the three channels may represent different spectral or lifetime-encoded information. After passing through the contracting path, feature maps are progressively up-sampled and merged with corresponding skip-level features from the encoder. This yields a final segmentation mask in the output layer, where each pixel is classified as either "NAP" or "non-NAP".

For diagnostic or reporting purposes, the predicted mask can be overlaid on the original FLIm frame, thereby highlighting the NAP regions. In addition, global measures—such as the fraction of NAP pixels relative to the entire image—can be computed to provide quantitative indicators of neutrophil activation. As will be shown in subsequent sections, this U-Net-based segmentation approach not only streamlines the detection of NAP signals but also provides a flexible platform for exploring how preprocessing choices (*e.g.,*, filtering or frame selection) affect the accuracy and robustness of NAP localisation.

5.3.1.1 *Technical Overview of the U-Net Architecture*

As shown in Fig. 5.7 and discussed in Section 5.3.1, the U-Net architecture comprises an encoder (contracting path) and a decoder (expanding path) joined by skip connections. The encoder reduces spatial dimensions via repeated convolution, ReLU activation, and max-pooling, whereas the decoder gradually restores the original image size through upsampling and transposed convolutions (also called deconvolutions).

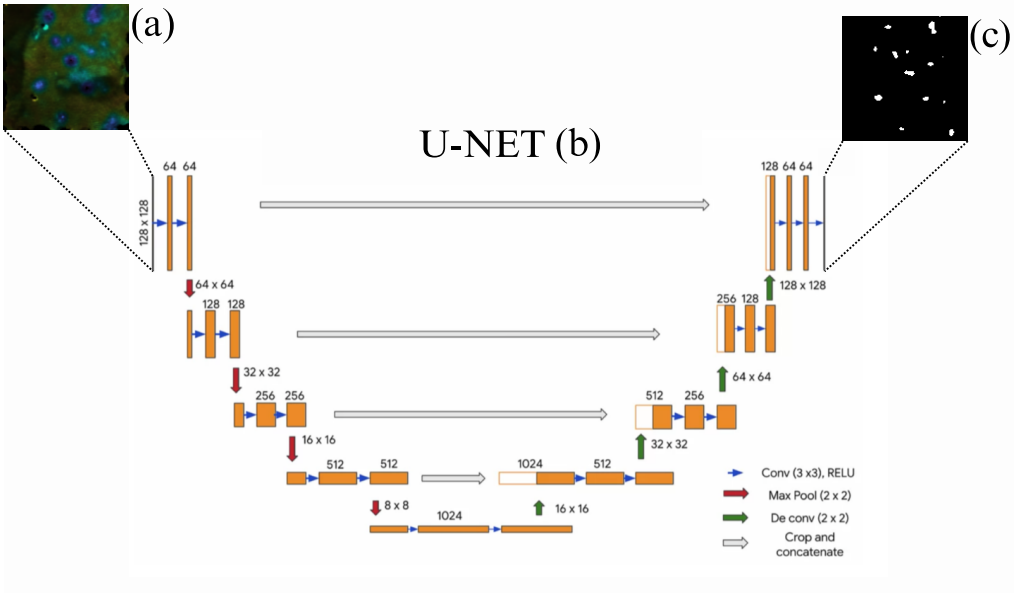


Figure 5.7: Schematic of the proposed U-Net framework for NAP detection. (a) Input FLIm data (128×128×3). (b) The U-Net model, featuring repeated 3×3 convolutions with *ReLU* activation, 2×2 max pooling operations, 2×2 deconvolutions, and skip connections to preserve spatial information. (c) Output binary mask where bright regions correspond to NAP signals.

Each layer in the encoder applies a convolution followed by a non-linear activation function. Given an input image $\mathbf{I}$ and a filter $\mathbf{F}$, the discrete 2D convolution evaluated at position (x, y) is:

$$[\mathbf{I} * \mathbf{F}](x, y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \mathbf{I}(i, j)\, \mathbf{F}(x - i,\, y - j), \tag{5.9}$$

where convolutional filters capture local spatial features. The Rectified Linear Unit (`ReLU`) function,

$$\mathrm{ReLU}(z) = \max(0, z), \tag{5.10}$$

introduces non-linearity and mitigates vanishing gradients. After these operations, a 2×2 max-pooling step reduces the feature map resolution. For example,

$$\mathrm{MaxPool}[\mathbf{M}]_{i,j} = \max\{\mathbf{M}_{2i,\,2j},\, \mathbf{M}_{2i+1,\,2j},\, \mathbf{M}_{2i,\,2j+1},\, \mathbf{M}_{2i+1,\,2j+1}\}. \tag{5.11}$$

Pooling compresses spatial information while retaining the most salient activations, enabling deeper layers to capture more global context.

The decoder (expanding path) inverts the encoding process by successively restoring the spatial dimensions. Transposed convolution (also referred to as deconvolution) is commonly used for upsampling, defined as:

$$[\mathbf{I} \circledast \mathbf{F}](x, y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \mathbf{I}(x - i,\, y - j)\, \mathbf{F}(i, j). \tag{5.12}$$

At each decoding level, features are concatenated with their corresponding encoder outputs (via skip connections) to recover fine-grained spatial detail lost through pooling. This mechanism is crucial for tasks like NAP detection, where small or irregularly shaped regions must be accurately segmented.

To produce the binary segmentation mask, U-Net ends with a 1×1 convolution per pixel, mapping the aggregated features to a single channel. A sigmoid activation function converts these values to probabilities:

$$\text{Sigmoid}(z) = \frac{1}{1 + e^{-z}}, \tag{5.13}$$

where $\text{Sigmoid}(z) \in [0, 1]$ represents the likelihood of a pixel belonging to the NAP class.

In line with Section 5.2.3.1, BCE remains a standard objective for pixel-wise classification:

$$\text{BCE}(y, p) = -\Big[y \log(p) + (1 - y) \log(1 - p)\Big], \tag{5.14}$$

where $y \in \{0, 1\}$ is the ground-truth pixel label and $p \in [0, 1]$ the predicted probability. Additionally, to address class imbalance or emphasise precise boundary localisation, the Dice Coefficient may be used:

$$\text{Dice}(A, B) = \frac{2\,|A \cap B|}{|A| + |B|}, \tag{5.15}$$

with Dice Loss defined as $1 - \text{Dice}(A, B)$. Combining BCE and Dice Loss offers a practical balance between local pixel-level accuracy and global overlap consistency:

$$\text{CombinedLoss} = 0.5 \times \text{BCE} + 0.5 \times \text{DiceLoss}. \tag{5.16}$$

This compound objective helps the network learn not only to label individual pixels correctly but also to maximise the overall match between predicted and true NAP regions.

5.3.2 *Results*

Figure 5.8 presents the training (yellow) and validation (red) metrics for the U-Net model across 150 epochs, showing the combined loss on the left and the intersection over union (IoU) score on the right. The consistent decrease in the combined loss, coupled with the steady rise in IoU, indicates that the model effectively learns to segment NAP pixels from the rest of the image. Although the dataset is relatively small, the encoder-decoder structure with skip connections appears sufficient to capture the essential features required for NAP detection.

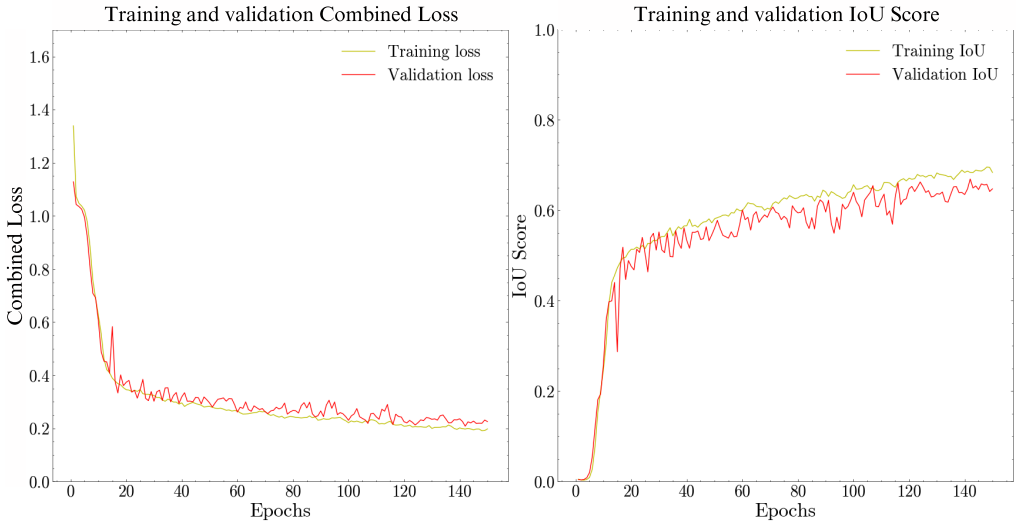


Figure 5.8: Training (yellow) and validation (red) metrics for the U-NET model over 150 epochs, highlighting: (left) the combined loss, and (right) IoU score. The U-NET shows progressive improvements in segmenting NAP regions, as indicated by rising IoU curves and decreasing combined loss.

As illustrated in Fig. 5.9, the proposed U-Net segmentation masks more closely match the annotated ground truth compared with either the built-in KronoScan software or a simple thresholding approach based on intensity (photon count $>$ 500) and lifetime ($<$ 2.5 ns). Table 5.3 quantitatively underscores this improvement: the built-in software yields an IoU of just 0.065, while thresholding reaches 0.356. U-Net, by contrast, achieves an IoU of 0.729, indicating a more coherent detection of NAP regions.

Although these results exceed those from heuristic-based methods, they represent only a moderate or "marginal" gain in certain real-world scenarios. The principal value of the U-Net approach here is to show how reliable preprocessing and segmentation can feed into broader FLIm analyses, rather than to introduce an entirely novel NAP detection paradigm. Indeed, as part of the subsequent

Table 5.3: Varying the filtering method and data representation affects U-Net performance for automatic NAP detection.

Detection Method	IoU Score
KronoScan Built-in Software	0.065
Image Thresholding	0.356
U-NET	**0.729**

ablation studies (Sections 5.4), we demonstrate how changes to filtering, data representation, and image registration can further influence U-Net segmentation performance.

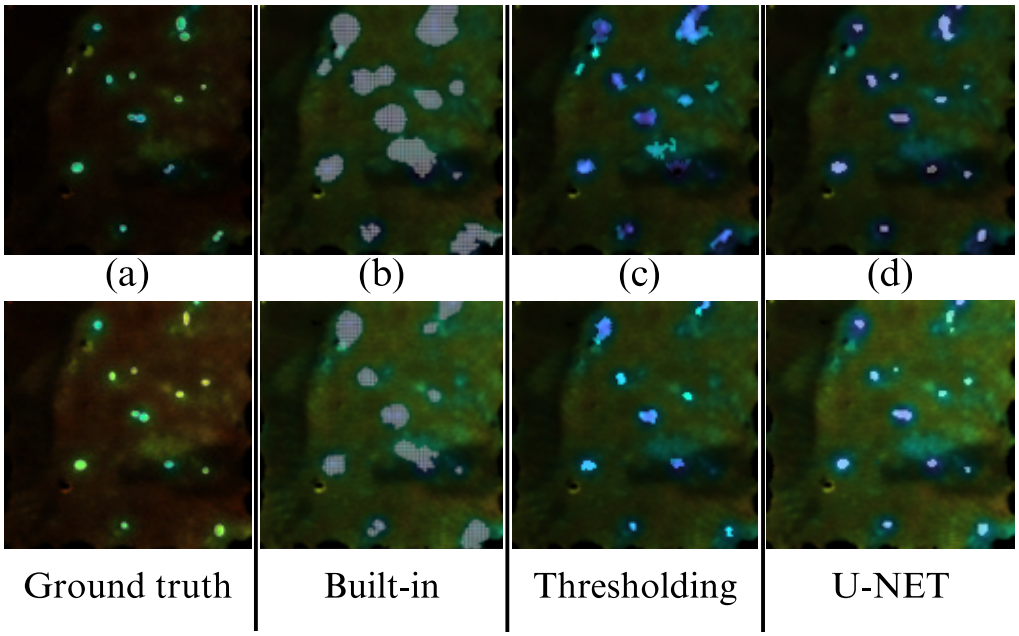


Figure 5.9: Qualitative examples comparing NAP detection to annotated ground truth (a). The built-in KronoScan software applies a statistical detection approach (b), while the thresholding method (c) labels pixels with an intensity above 500 photons and a lifetime below 2.5 ns. The proposed U-NET (d) yields more accurate and coherent segmentations, matching the ground truth annotations more closely than the other methods. Overlaid masks highlight regions where NAP signals are detected.

In summary, these experiments highlight how the U-Net architecture, even trained on a limited dataset, can robustly detect NAP signals in FLIm images. This offers a solid baseline for future optimisation, particularly where improved pipeline steps or larger labelled datasets might drive segmentation performance closer to a clinically actionable standard.

5.4 IMPORTANCE OF A ROBUST IMAGE PROCESSING PIPELINE

This section examines the influence of various processing elements in the TRACER pipeline (described in Chapter 3 and Section 5.2.3) on downstream analyses of FLIm data. Specifically, we focus on detecting NAP regions, a critical step that often dictates the reliability of subsequent quantitative measures.

This section demonstrates the significance of a robust image processing pipeline through various ablation studies. It highlights not only alternative methods for key tasks (such as using the U-Net model to detect NAP regions in FLIm imaging data and applying a DL-TL approach to detect and remove uninformative frames) but also the importance of standardisation and holistic performance evaluation. The image reconstruction process is also considered, including comparisons of filtering methods, data representation, and the implications of removing uninformative frames and registering the remaining data before training the U-Net.

From a clinical perspective, the target performance of such a pipeline is not defined by pixel-level precision or anatomically exact boundary delineation, but rather by the reliable and repeatable identification of clinically relevant tissue regions. In the context of endoscopic FLIm, moderate spatial imprecision can often be tolerated provided that the algorithm consistently distinguishes between tissue types or functional states that are meaningful for diagnosis, guidance, or longitudinal comparison. As such, the performance metrics considered in this chapter should be interpreted primarily as indicators of robust tissue differentiation and detection stability, rather than absolute segmentation accuracy.

By demonstrating U-Net's performance under the proposed pipeline, we provide a concrete example of how an improved, standardised FLIm reconstruction—combined with carefully curated training data—can yield consistent segmentations of NAP regions. In the following subsections, the reader is taken through considerations at different stages of implementing a ML model to automate a downstream task such as the automatic detection of NAP in OEM-FLIm data. It begins by discussing the image reconstruction process that involves the processing of FLIm and examine the effect of tweaking the process on the implementation of U-NET. It then reiterate the importance of dealing with uninformative frames comparing the performance of the proposed TL approach to the one developed by [1], and finally discuss the implications of the pipeline robustness and the value of registering the data and account for spatial correspondence between the images prior to training the U-NET for detecting NAP and what that means to the performance.

5.4.1 *Image Reconstruction*

Alpha-weighted lifetime images, along with the subtle differences between the two modalities discussed in Chapter 2, Section 2.2.2.1 and Section 2.2.2.2, can influence downstream tasks. This underscores the need to standardise the reconstruction of FLIm imaging data. The automated detection of NAP using the U-Net model (Section 5.3) relies on combining information from both intensity and lifetime images. By emphasising regions of higher intensity, alpha-weighted lifetime images can highlight lower-lifetime areas, making them especially valuable for NAP detection.

It is also crucial to test the reconstruction approach on a large dataset, such as the one used in this thesis, which contains over 100 image sequences and more than 10 000 images. Such a comprehensive dataset enhances trust in the generalisation of the method. Several studies [27, 73, 27] have applied Gaussian filtering (window sizes of 3–7) to reduce noise in lifetime images. However, experiments show that median filtering, combined with higher-dimensional data representation (RGB), can further enrich features and improve NAP detection.

5.4.1.1 *Image Filtering*

Gaussian filtering is a low-pass filter that smooths variations by blending each pixel with its neighbours, weighted by a Gaussian function [153]. It reduces high-frequency noise and preserves low-frequency structure but can blur edges when there is a large intensity difference. Inactive cores (zero-valued pixels) near boundaries can produce visible edge effects when smoothed by Gaussian filtering, especially if the field of view (FoV) does not span the full spatial resolution. These artefacts can degrade image quality and influence downstream tasks.

Median filtering, a common nonlinear approach, selects the median value within a local kernel and replaces the centre pixel with this value. This method suppresses noise (such as salt-and-pepper) more effectively than Gaussian filtering and preserves edges by avoiding the averaging effect. Zero-valued or inactive pixels at boundaries are handled better, reducing artificial smoothness and preserving edge detail [153].

Median Filtered Image: Median filtering manages outliers and inactive cores more effectively, producing images with sharper edges and fewer boundary artefacts.

Gaussian Filtered Image: Gaussian filtering reduces noise but can blur boundaries near inactive cores, generating unnatural gradients that affect edge clarity.

Choosing a filtering approach that aligns with the data characteristics can improve image quality and performance in downstream analyses, as illustrated in Fig. 5.10, distinct filtering methods and representations can lead to notable

qualitative differences in FLIm images. Comparing the raw intensity image with subsequent processing steps clarifies how median filtering and alpha weighting can visually emphasise NAP regions.

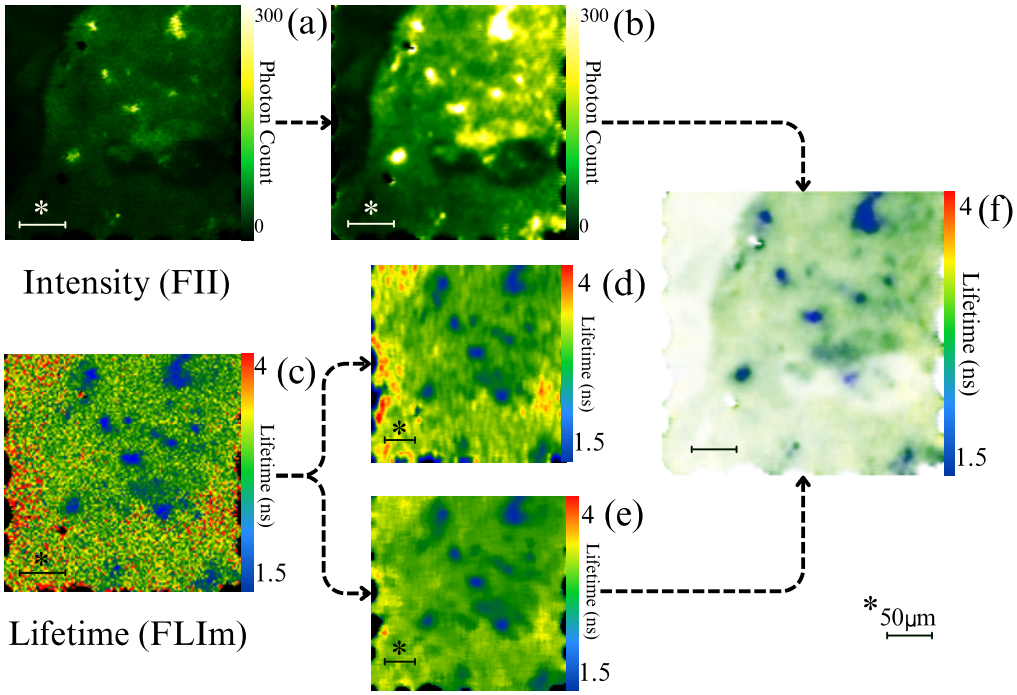


Figure 5.10: Qualitative comparison of various image types and representations for FLIm: (a) Raw intensity image, displayed in a dark-green-to-light-yellow colour scale. The colour bar associated with the intensity images represents detected photon counts. (b) Processed intensity image following background removal and lightfield normalisation. (c) Raw lifetime image with a jet colour scale. The colour bar associated with the lifetime images represents fluorescence lifetime in nanoseconds (ns). From (c), an arrow splits into: (d) Gaussian-filtered image and (e) Median-filtered image (which shows better edge preservation and fewer artefacts). Finally, an arrow from (e) leads to (f) Alpha-weighted lifetime image, which highlights NAP (blue regions) more clearly than the other representations. All images include a spatial scale bar corresponding to 50 μm.

5.4.1.2 *Data Representation*

When applying the same considerations about RGB versus grayscale representation to FLIm, additional nuances arise. FLIm measures the decay characteristics of fluorescent signals in biological tissues or fluorophores, yielding biochemical and physiological insights useful for distinguishing healthy from diseased tissue [53, 14].

By default, many FLIm datasets appear as single-channel (grayscale) intensity images or parametric maps, *i.e.,* lifetime values. While grayscale or simple pseudocolour maps can help observers detect regions with different lifetimes, richer multi-channel or "fake-colour" encodings can better preserve information. In this work, such encodings are used pragmatically to organise multiple scalar image fields (*e.g.,* intensity, lifetime, or intensity-weighted lifetime) into formats compatible with standard image-processing pipelines and learning-based models, rather than to imply additional physical measurement dimensions. For instance, distinct colour channels for different lifetime ranges, intensity levels, or even spectrally resolved signals can produce a representation analogous to RGB images [176], helping to avoid the information collapse associated with monochrome images.

From a ML perspective, colour or multi-channel FLIm data can facilitate learning more discriminative features [177]. Neural networks, especially convolutional architectures, benefit from channel-specific details that highlight aspects of tissue composition, metabolic states, or fluorophore distributions [14]. This advantage is not unique to RGB encodings; alternative multi-channel representations, such as intensity combined with RLD-derived, can provide comparable information content when appropriately structured. This not only supports classification or segmentation tasks but can also reduce noise resulting from collapsing everything into a single channel.

Artificial or "fake" colour maps have been widely used for qualitative assessments of FLIm data, especially in clinical contexts where distinct colours correspond to characteristic lifetime variations (*e.g.,* short-lifetime fluorophores mapped to blue, longer lifetimes mapped to red). Beyond qualitative interpretation, these colour maps can enrich the feature space for machine learning. They enable data augmentation (*e.g.,* adjusting contrast or channel weighting) and improve generalisation by emphasising subtle lifetime differences that a grayscale image might conceal [178]. In this sense, colour encodings act as a representational convenience that exploits the inductive biases of image-based learning architectures trained on natural RGB imagery, rather than as a strict requirement for downstream analysis.

In deciding whether to use grayscale or multi-channel FLIm representations, the key factor is whether additional lifetime or spectral channels yield diagnostic or predictive advantages. If the task focuses on large-scale spatial patterns, a single channel may be sufficient and computationally lighter. Conversely, if subtle lifetime variations are central to class discrimination—such as differentiating diseased from non-diseased tissue—multi-channel or fake-colour FLIm images can deliver superior machine learning performance. The choice of representation is therefore task-dependent, with RGB-style encodings adopted here to balance

information salience, computational convenience, and compatibility with established image-processing and machine-learning workflows. As with RGB data, architectural considerations must accommodate the increased dimensionality, but this is often a modest trade-off for improvements in diagnostic accuracy and robustness [177].

Table 5.4 summarises how different filtering methods and data representations affect U-Net performance IoU for NAP detection, with an uncertainty assessment derived via a bootstrapping strategy [179]. Moving from a simple grayscale representation (IoU $= 0.510 \pm 0.121$) to Gaussian-filtered RGB (IoU $= 0.554 \pm 0.098$) yields a moderate improvement, indicating the value of exploiting multiple colour channels to capture additional structural cues. Applying alpha-weighted Gaussian filtering (IoU $= 0.636 \pm 0.103$) further underscores the advantage of emphasising higher-intensity regions, which often correlate with relevant NAP features.

Table 5.4: Varying the filtering method and data representation affects U-Net performance for automatic NAP detection.

Image Type	IoU Score*
Gray-scale	0.510 ± 0.121
Gaussian-filtered RGB	0.554 ± 0.098
Alpha-weighted Gaussian-filtered RGB	0.636 ± 0.103
Median-filtered RGB	0.688 ± 0.075
Alpha-weighted median-filtered RGB	$\mathbf{0.729 \pm 0.091}$

* An aleatoric uncertainty (data uncertainty) aware assessment, done using bootstrapping strategy [179].

Median filtering delivers a more substantial gain (IoU $= 0.688 \pm 0.075$), consistent with its known edge-preserving and noise-reducing capabilities. The highest performance (IoU $= 0.729 \pm 0.091$) emerges from the combination of alpha-weighting and median filtering. These results highlight the importance of both filter selection and data representation in bolstering automated NAP detection. In particular, adopting multi-channel RGB and alpha-weighted images allows the model to leverage rich texture and intensity variations, while robust filtering methods can help preserve essential boundaries and minimise artefacts.

5.4.2 *Image Registration*

As discussed in Chapter 3, Section 3.4.5, aligning frames spatially before segmentation can substantially improve the accuracy of NAP detection. Previously, thresholding-based approaches on FLIm data showed that TRACER registration not only increased precision by up to 4 % and recall by up to 27 %, but also

yielded an up to 19 % higher F1 score. These results emphasise how correct spatial correspondence between frames can strengthen subsequent segmentation steps by reducing inconsistencies in pixel alignment.

Table 5.5 extends this evaluation to the U-Net model introduced in Section 5.3. The table compares performance metrics (IoU, Dice, and Combined Loss) for U-Net trained on unregistered *vs.* registered image sequences, with aleatoric (data) uncertainty estimates derived via bootstrapping [179]. On unregistered frames, the U-Net achieves an IoU of 0.729 ± 0.091, whereas the same network trained on registered data reaches 0.786 ± 0.016. Similar gains are seen for the Dice coefficient and in lowering the combined loss. These improvements likely stem from the more coherent spatial alignment of features, enabling the model to learn consistent NAP appearances across frames.

From a learning perspective, these improvements are consistent with the role of image registration as a preprocessing step that enforces spatial correspondence and reduces non-informative variability in the training data. As discussed by Chen et al. [180], registration can improve the effectiveness of downstream machine learning by ensuring that homologous anatomical or structural features are consistently aligned across samples, thereby stabilising feature representations learned by the model.

Table 5.5: A table to summarise different approaches to the proposed transfer learning model for uninformative frame detection.

Metric*	Unregistered Images	Registered Images**	Performance Gain
IoU ↑	0.729 ±0.091	**0.786 ±0.016**	7.82%
Dice ↑	0.814 ±0.012	**0.857 ±0.002**	5.29%
Combined Loss ↓	0.196 ±0.019	**0.152 ±0.004**	22.09%

* An aleatoric uncertainty (data uncertainty) aware assessment, done using a bootstrapping strategy [179].

** The images are registered using the TRACER registration (Chapter 3, Section 3.3.2).

In the context of frame-based FLIm sequences, registration reduces the risk that the network is repeatedly exposed to nominally similar structures appearing at different spatial locations, which could otherwise behave as a form of label noise or implicit data distortion during training. While this process may also reduce the apparent diversity of spatial configurations seen during optimisation, the results here suggest that the benefit of improved pixel-wise correspondence outweighs any reduction in effective sample variability for the NAP detection task considered. Consequently, these findings reinforce the conclusion from Chapter 3 that spatial alignment in FLIm imaging is not merely an isolated visual stabilisation step, but a foundational component that can positively influence downstream ML tasks, including NAP detection and broader automated analysis pipelines.

5.5 CONCLUSION

This chapter presented an integrated exploration of ML techniques for two key tasks in FLIm-based imaging pipelines: (1) uninformative frame detection, and (2) the automated detection of NAP. The opening sections demonstrated the efficacy of a TL approach in filtering out poor-quality frames by leveraging knowledge from pre-trained CNN architectures, namely VGG16, InceptionV3, and ResNet50. Empirically, VGG16 surpassed the performance of the other models, exhibiting relatively stable and reliable outcomes. Moreover, the TL approach achieved high sensitivity and specificity on a relatively modest dataset, outperforming the feature-engineering method presented in [1] by yielding a 2.85 % gain in sensitivity and a 5.28 % gain in specificity. Beyond these numerical improvements, adopting a data-driven paradigm obviates the need for handcrafted features and Gaussian mixture clustering, thereby offering greater adaptability to evolving FLIm acquisition settings. Crucially, the removal of uninformative frames was shown to provide a tangible increase in downstream imaging quality, notably improving registration accuracy and enhancing subsequent analysis metrics.

Next, we examined the detection of NAP signals—biologically significant markers of neutrophil activation [17]—using a U-Net-based segmentation framework [173]. While threshold-based or purely statistical methods (including the built-in KronoScan algorithm) capture some coarse features, the data-driven U-Net approach delivered improved pixel-level precision, producing more coherent NAP masks that closely align with expert annotations. Notably, both the filtering mechanisms (*e.g.,* alpha-weighted median filtering) and data representations (*e.g.,* RGB channels) proved instrumental in enhancing U-Net performance. Furthermore, the temporal alignment offered by the TRACER registration demonstrated that image registration can mitigate distortions or inter-frame motion displacement that might otherwise introduce noise or mismatches in the training data. This alignment ensures that NAP regions correspond well between consecutive frames—or across different imaging sessions—thereby boosting the performance of the same U-Net model trained on unregistered data by 7.28 % in IoU, 5.29 % in Dice, and 22.09 % in the combined loss, underscoring the importance of careful data preprocessing.

A central message of this chapter has been the interplay between classical image processing steps, such as filtering and image registration, and DL models. Empirical results demonstrated that even minor modifications in data representation or spatial alignment can have a significant impact on ML outcomes. In turn, stronger ML segmentation or classification pipelines further enhance the reliability of the TRACER framework. Taken holistically, each component of the imaging pipeline—frame filtering, image registration, data augmentation,

and deep segmentation—contributes to a more robust and accurate end-to-end solution.

Regarding future prospects, the findings offer several avenues for refinement. Incorporating domain-adaptation or semi-supervised techniques may mitigate the constraints imposed by small labelled datasets, while advanced architectures (*e.g.,* Vision Transformers or hybrid 3D models) may capture additional contextual or temporal cues. For instance, integrating multi-scale attention mechanisms and temporal consistency constraints could further refine feature extraction and enhance segmentation performance. Moreover, exploring cross-modality transfer learning may improve model robustness and facilitate better generalisation across heterogeneous clinical datasets. These directions will be explored in Chapter 6, which will provide a broader synthesis of the thesis contributions and outline a roadmap for continued research in OEM-FLIm real-time pulmonary imaging.

6

CONCLUSIONS AND RECOMMENDATIONS

Fibre-based optical endomicroscopy (OEM) imaging has emerged as a powerful medical imaging platform, particularly with the advent of miniaturised, flexible fibre-optic endoscopes. Fluorescence lifetime imaging (FLIm) has gained traction in clinical research, offering real-time diagnostics and drug discovery applications, with notable potential for investigating the lung's alveolar spaces; regions traditionally challenging to image using standard modalities. In this context, fluorescence lifetimes, derived from temporal decay profiles using methods such as rapid lifetime determination (RLD) or other advanced and machine learning (ML)-based approaches, provide crucial molecular insights alongside intensity-based imaging.

Despite its promise, fibre-based OEM-FLIm faces significant challenges, particularly motion artefacts induced by lung mechanics, probe movement, and structural distortions that degrade image quality and hinder quantitative analysis. This thesis addresses these limitations by developing a robust image processing pipeline tailored for OEM-FLIm, enhancing image fidelity through motion artefact correction, advanced registration techniques, and ML-based frame selection.

The key contributions of this thesis include:

- Establishing a foundational understanding of OEM-FLIm in pulmonary imaging, highlighting its clinical significance and inherent challenges.
- Developing the temporal reliability and accuracy via correlation enhanced registration (TRACER) pipeline to mitigate motion artefacts, noise, and signal degradation in real-time *in vivo* FLIm imaging.
- Enhancing image registration accuracy by leveraging both intensity and lifetime data to adapt to the dynamic nature of FLIm sequences.
- Investigating ML-based techniques to refine frame selection, improve disease characterisation, while highlighting the importance of the overall robustness of FLIm image processing.
- Validating the proposed methods on both simulated and real datasets, demonstrating improvements in data quality, computational efficiency, and clinical relevance.

Through these advancements, this work reinforces the viability of OEM-FLIm as a research and a promising diagnostic tool in pulmonary imaging. By addressing

both hardware and computational challenges, this thesis lays a stronger foundation for real-time lung imaging, ultimately aiming to improve diagnostic capabilities and expand the clinical utility of OEM-FLIm.

6.1 SUMMARY

Building on the motivation and aims set forth in Chapter 1, this thesis has focused on advancing fibre-based OEM-FLIm for pulmonary imaging by addressing motion artefacts, noise, and issues relating to image alignment and data quality. The overarching goal is to enable reliable imaging of the lung's alveolar regions in real-time, enhancing both clinical diagnostics potential and research into pulmonary diseases. The following chapters have contributed to a robust, holistic image processing pipeline, culminating in improved accuracy, speed, and interpretability of FLIm data. A concise summary of these chapters and their major findings is presented below.

Chapter 2 outlined the evolution of pulmonary imaging and the specific role FLIm can play by offering functional insights at the alveolar level. While FLIm holds promise for early diagnosis and personalised treatment, it is highly susceptible to motion artefacts, which can obscure or distort the signal. A review of respiratory mechanics showed that the lungs' continuous expansion and contraction present a challenging imaging environment, especially when using a miniature fibre bundle. Overcoming these motion effects is therefore central to maintaining data integrity and unlocking FLIm's full potential in clinical applications. Although direct comparisons with existing literature remain limited by the novelty of the approach, the fundamental investigation into pulmonary motion laid the groundwork for the subsequent development of real-time motion compensation strategies.

In Chapter 3, the thesis introduced the TRACER pipeline, a novel image processing pipeline specifically targeting the unique challenges of real-time OEM-FLIm. By combining strategic pre-processing, dense optical flow (OF) for motion characterisation, and channel and spatial reliability tracker (CSRT) for tracking-based normalised cross correlation (NCC) registration, TRACER demonstrated substantial gains in data quality. Specifically, TRACER improved quality of alignment (QA), structural similarity index measure (SSIM), and normalised root mean squared error (NRMSE) by 20 %–30 % over unprocessed imaging data, while outperforming state-of-the-art registration methods with a margin of up to 8 % in the aforementioned metrics. Crucially, it achieved these improvements at an order-of-magnitude reduction in imager registration computational time compared to the next best-performing approach, highlighting its suitability for real-time applications. Experiments on both simulated and real FLIm datasets

confirmed TRACER's robustness and adaptability to complex image texture and scenarios exhibiting significant displacement, reinforcing its potential for clinical use. Future directions include expanding the dataset to more varied lung imaging conditions and further refining non-rigid registration techniques.

Throughout Chapter 4, the work then delved deeper into integrating intensity and lifetime information through multi-objective image registration strategies. Three main approaches were assessed:

1. a weighted-sum formulation,
2. a multi-variate optimisation that jointly learns the weighting factor and registration parameters,
3. a non-linear combination of both modalities.

Across these strategies, intensity consistently emerged as the dominant modality, largely due to its richer structural features and stronger correlation in pulmonary images. Nevertheless, the analysis of the NCC space and Pareto fronts highlighted scenarios (*e.g.,* higher signal-to-noise ratio (SNR), less texture homogeneity) where lifetime data could carry comparable weight. The chapter also introduced dynamic multi-objective search (DYNMOS), a dynamic modality selection algorithm that determines the optimal modality to prioritise, offering comparable registration performance to the more computationally heavier multi-objective optimisation methods, at a 2.46× speed-up. This research underscores the complementary nature of intensity and lifetime data and shows how adaptive strategies can harness both to enhance registration accuracy under diverse imaging conditions.

Finally, in Chapter 5, the thesis investigated ML approaches for two critical tasks: uninformative frame detection and the automated detection of neutrophil activation probe (NAP), a biologically important marker for the detection of lung inflammation. In the first case, transfer learning (TL) methods using VGG16, InceptionV3, and ResNet50 were applied to classify and remove low-quality frames. VGG16 achieved the most robust performance, surpassing a traditional feature-engineering baseline by 2.85 % in sensitivity and 5.28 % in specificity, while eliminating the need for handcrafted features. This improvement in frame filtering directly bolstered subsequent registration and analysis steps.

For NAP detection, a U-Net-based segmentation framework produced superior pixel-level masks compared to threshold-based techniques, especially when combined with alpha-weighted median filtering and carefully designed data representations. Critically, the TRACER pipeline's registration step helped align regions between frames, improving U-Net performance by up to 7.28 % in intersection over union (IoU) and 5.29 % in Dice. These results emphasise the synergy between classical image processing (*e.g.,* alignment, filtering) and ML-based methods,

illustrating how small but targeted improvements in data quality can lead to significant gains in segmentation or classification accuracy.

Taken together, these findings demonstrate that a comprehensive, multi-component approach is essential for advancing OEM-FLIm in pulmonary imaging. By addressing motion artefacts at their source, enhancing image registration with dual-modal data, and integrating ML for improved frame selection and feature detection, the thesis achieves a level of accuracy, speed, and reliability not previously reported in the literature. This integrated pipeline—spanning hardware constraints, algorithmic adaptations, and data-driven insights—lays a solid foundation for future innovations in FLIm.

The chapters of this thesis collectively underscore the importance of a holistic methodology, wherein carefully tailored image processing and machine learning techniques collaborate to overcome the inherent challenges of fibre-based pulmonary imaging. With FLIm and OEM systems continuing to evolve, the methods presented here pave the way for real-time, high-fidelity lung imaging that can enhance clinical diagnostics, support longitudinal studies, and deepen our understanding of pulmonary disease mechanisms.

6.2 THESIS OUTLOOK

This thesis has presented a multi-stage pipeline specifically tailored to address the unique challenges of pulmonary OEM-FLIm. From data pre-processing to motion-model characterisation and registration, the proposed approach offers a coherent framework for handling both the molecular-level content of FLIm data and the inherent complexities of lung imaging (including breathing motion and variable probe handling). While promising, this pipeline remains a work in progress. This section critically evaluates its limitations, proposes avenues for future enhancement, and offers recommendations for researchers and clinicians seeking to bring FLIm closer to mainstream clinical practice.

6.2.1 *Limitations*

6.2.1.1 *Motion Modelling and Registration*

Despite successful application to several FLIm sequences, the current approach is predominantly designed around rigid transformations, limiting its ability to account for significant rotations, scaling, or affine distortions. Additionally, non-rigid deformations, common in complex respiratory and probe-handling scenarios, are only partly addressed. A key reason for this is the relatively small number

of available FLIm datasets, constraining the potential for robust training and testing of more advanced registration algorithms.

6.2.1.2 *Uninformative Frame Detection*

Although progress was made in filtering out frames that contain minimal or no clinical relevance, the current approach remains vulnerable to ambiguous cases or marginally informative frames. As a result, the pipeline might either discard frames that could harbour subtle yet important features or retain those lacking meaningful clinical signal. A more sophisticated framework is therefore required: one that could detect low-information frames and also recognises specific anatomical regions, thus refining what content is deemed "useful" or "uninformative."

6.2.1.3 *Data Requirements and Computational Costs*

In Chapter 5, several sections of this thesis drew on machine learning and deep learning approaches, which often demand significant quantities of annotated data. Although transfer learning strategies alleviate some of these requirements, the reality is that highly specialised FLIm data remain scarce. This scarcity restricts model generalisability and requires computationally intensive training cycles, typically on high-performance GPU hardware.

In this context, preprocessing steps such as frame-to-frame registration introduce an additional trade-off. While improved spatial alignment can stabilise feature learning and reduce non-informative variability, it may also reduce the effective diversity of spatial configurations presented during training. If not interpreted carefully, repeated exposure to highly similar registered frames could increase the risk of overfitting or give the appearance of improved performance due to reduced sample variability rather than genuinely enhanced generalisation.

Accordingly, the performance gains reported in this chapter should be interpreted as reflecting improved learning stability under constrained data conditions, rather than as a guarantee of superior generalisation to unseen acquisition scenarios. This consideration is particularly relevant for small or highly correlated FLIm datasets, where registration and data augmentation strategies must be balanced against the risk of implicit data redundancy.

6.2.1.4 *Interpretability and Clinical Trust*

Related to the point above is the problem of interpretability. Regulators and clinicians often request transparent decision pathways, which remain challenging in "black-box" deep learning (DL) architectures. Techniques from the field of

explainable artificial intelligence (XAI) [181] can aid in bridging this gap, but broader adoption is still lacking in preclinical and clinical FLIm contexts.

6.2.1.5 *Fusion Strategy and Loss of Detail*

The current averaging-based fusion strategy is effective in reducing noise but can lead to the partial loss of important, high-frequency image details. When objects of interest are small or represent transient features, averaging may dilute or obscure them [86].

6.2.2 *Future Work*

6.2.2.1 *Advanced Registration Methods and Expanded Datasets*

The scope of the motion analysis presented in this thesis was constrained by the availability of representative in vivo datasets acquired under controlled conditions. As such, the quantitative evaluation focused on translational motion, which was observed to dominate the majority of frame-to-frame displacements in the available data and provided an interpretable model for systematic analysis.

Extending beyond rigid transformations to incorporate non-rigid or deformable registration techniques is a logical progression. For instance, adaptive frameworks that automatically classify motion complexity (*e.g.,* low vs. high displacement) and select the most suitable registration method can substantially improve accuracy. Parallels from disparate fields highlight powerful strategies. In remote sensing, for example, the multi-temporal method of Yang *et al.* [182] illustrates how robust feature descriptors are essential for registering data with significant appearance variations—a challenge analogous to handling physiological changes or photobleaching artefacts in longitudinal FLIm studies. Their use of a multi-scale convolutional neural network (CNN) to generate biochemically-aware features from combined intensity and lifetime maps, alongside a "dynamic inlier selection" strategy that progressively refines alignment, presents a clear roadmap for improving feature-based robustness.

Complementing this feature-centric view, recent work in medical imaging on contrast enhanced ultrasound (CEUS)-based lesion tracking by Li *et al.* [93] demonstrates the power of explicitly decoupling motion estimation from the matching task. Their Attention-guided densely multi-scale network (ADMNet) architecture first employs a Deformable Motion Compensation module to predict and counteract non-rigid respiratory motion between frames. This generates a motion-compensated template, which is then passed to a separate Attention-based Matching module that performs robust localisation. This two-stage approach, which isolates and solves for large-scale physical deformation before attempting

fine-grained matching, is highly relevant for FLIm. Adopting a similar strategy could involve a preliminary network that corrects for patient-induced motion, providing a stably aligned series of images upon which a more sophisticated, feature-based registration (such as one inspired by Yang *et al.*) could be performed to ensure precise biochemical alignment. Translating these multi-stage, hybrid techniques into FLIm may therefore substantially improve registration robustness under dynamic conditions.

While this thesis has primarily operated in the image domain, registering $\mathbf{I}(x, y)$ and $\mathbf{L}(x, y)$, due to the limited temporal resolution available in 2-bin FLIm systems, future datasets with richer temporal sampling (*e.g.,* 16-bin or full time-correlated single photon counting (TCSPC) generated histograms) open the door to an alternative approach. Rather than relying solely on derived image maps, high-resolution decay profiles could be registered directly by aligning the full temporal signatures. This would enable new forms of multi-objective optimisation that treat each decay curve as a multi-dimensional signal, potentially improving robustness to photonic noise and enhancing biochemical fidelity in spatial alignment. The literature in [68, 67, 22] suggests that higher-bin systems yield finer kinetic detail, which could be explicitly incorporated into the registration framework rather than abstracted into 2D maps. Exploring such strategies—especially under conditions where the lifetime signal becomes increasingly informative—presents an important future direction for advancing motion compensation in FLIm.

6.2.2.2 *Enhancing Uninformative Frame Detection and Region Recognition*

Future efforts should target hybrid algorithms that combine classical image processing with DL, potentially incorporating CNN-based region-of-interest detection or vision transformers [183], to distinguish true uninformative frames from those containing subtle clinical cues. This would entail training on larger, more heterogeneous datasets to capture the wide variance in lung anatomy and motion artefacts.

6.2.2.3 *Refined Fusion Techniques*

While simple averaging reduces noise, it may compromise important morphological details. Future work could explore domain-specific approaches such as multi-resolution fusion, attention-based fusion, or approaches inspired by universal fluorescence microscopy image restoration (UniFMIR) [184], thus striking a balance between denoising and preserving clinically relevant structures in FLIm data.

6.2.2.4 *Machine Learning for Detection and Beyond*

Further exploration of architectures beyond U-Net, such as Vision Transformers [183] or hybrid 3D CNN [185], could enhance the automatic detection of NAP (as introduced in Chapter 5). Domain-adaptive or semi-supervised approaches have particular promise in scenarios where labelled data are sparse. The integration of multi-scale attention mechanisms and temporal constraints may also bolster performance, thereby realising the full potential of a robust pipeline for motion compensation and region-of-interest segmentation.

6.2.2.5 *Robustness, Reproducibility, and Clinical Validation*

Echoing the insights of Roberts *et al.* [186] regarding pitfalls in ML-based COVID-19 diagnostics, future research should emphasise multi-site validation, external testing cohorts, and cross-validation protocols. Bias in dataset composition or inconsistencies in acquisition protocols can severely undermine reproducibility and hinder clinical translation. By adopting rigorous statistical reporting, standardised metrics, and transparency in experimental design, FLIm studies can avoid many of these pitfalls.

6.2.3 *Recommendations*

6.2.3.1 *Establish Clear Protocols and Guidelines*

- **Data Acquisition:** Define strict protocols for image acquisition, incorporating consistent device settings, patient preparation, and scanning conditions to minimise inter-study variability.
- **Image Formatting:** Encourage a standard format (analogous to DICOM in radiology [187]), improving compatibility across different software and facilitating large-scale data exchange.

6.2.3.2 *Validation and Benchmarking*

- **Ground Truth Data:** Invest in multi-institutional efforts to create robust ground truth data, fostering reliable training and evaluation.
- **Performance Metrics:** Develop task-specific, standardised metrics for motion correction, segmentation, and classification. A multi-metric approach (including structural similarity, accuracy, interpretability, and uncertainty) is advisable for a comprehensive evaluation of algorithmic performance.

6.2.3.3 *Promoting Interpretability and Regulatory Readiness*

Implement XAI modules and transparent model architectures to address clinical and regulatory concerns. This may include saliency maps [188] or feature-importance graphs that elucidate the decision processes within deep networks, thereby fostering trust among practitioners and stakeholders.

6.2.3.4 *Building Collaborative Ecosystems*

Given that FLIm remains a niche but increasingly vital field, interdisciplinary and cross-institutional collaborations can significantly accelerate technical and clinical advancements. Partnerships with industry may yield more standardised hardware solutions, whereas academic and clinical networks can contribute diverse datasets, bridging the gap towards full clinical translation.

6.2.4 *Concluding Remarks*

The results achieved throughout this thesis underscore the feasibility and potential of a system-wide pipeline for pulmonary OEM-FLIm, even under challenging motion conditions. Nonetheless, to fully integrate such pipelines into routine clinical practice, it is imperative to address the limitations outlined above. Building larger, higher-quality datasets, refining motion compensation and fusion strategies, exploring new machine learning architectures, and establishing rigorous validation protocols all represent essential steps forward.

In drawing upon methodological parallels from remote sensing [182], contrast-enhanced ultrasound [93], universal fluorescence microscopy [184], and lessons from large-scale clinical studies [186], the future of FLIm research appears well-positioned to deliver more robust, interpretable, and clinically relevant imaging solutions. By combining these technological advances with carefully considered clinical and regulatory pathways, there is a clear roadmap for translation into routine pulmonary diagnostics and beyond.

REFERENCES

[1] A. Perperidis, A. Akram, Y. Altmann, P. McCool, J. Westerfeld, D. Wilson, K. Dhaliwal, and S. McLaughlin, "Automated detection of uninformative frames in pulmonary optical endomicroscopy," *IEEE Transactions on Biomedical Engineering*, vol. 64, 1 2017. [Online]. Available: https://doi.org/10.1109/tbme.2016.2538084

[2] J. B. West, A. M. Luks, and L. W. . Wilkins., *West's respiratory physiology : the essentials*, 11th ed. Philadelphia: Wolters Kluwer, 2021.

[3] L. P. Nicod, "Lung defences: an overview," *European Respiratory Review*, vol. 14, pp. 45–50, 12 2005.

[4] D. Bendayan, Y. Barziv, and M. Kramer, "Pulmonary calcifications: a review," *Respiratory Medicine*, vol. 94, pp. 190–193, 3 2000.

[5] P. M. Scherer and D. L. Chen, "Imaging pulmonary inflammation," *Journal of Nuclear Medicine*, vol. 57, no. 11, pp. 1764–1770, 2016. [Online]. Available: https://jnm.snmjournals.org/content/57/11/1764

[6] W. De Wever, S. Stroobants, J. Coolen, and J. Verschakelen, "Integrated PET/CT in the staging of nonsmall cell lung cancer: technical aspects and clinical integration," *European Respiratory Journal*, vol. 33, no. 1, pp. 201–212, 2008.

[7] A. Perperidis, K. Dhaliwal, S. McLaughlin, and T. Vercauteren, "Image computing for fibre-bundle endomicroscopy: A review," *Medical Image Analysis*, vol. 62, 5 2020. [Online]. Available: https://doi.org/10.1016/j.media.2019.101620

[8] D. C. Humphries, R. A. O'Connor, H. L. Stewart, T. M. Quinn, E. E. Gaughan, B. Mills, G. O. Williams, J. M. Stone, K. Finlayson, M. Chabaud-Riou, F. Boudet, K. Dhaliwal, and V. Pavot, "Specific in situ immuno-imaging of pulmonary-resident memory lymphocytes in human lungs," *Frontiers in Immunology*, vol. 14, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1100161

[9] C. Manley, "Bronchoscopy: Past, present, and future," *Current Pulmonology Reports*, vol. 8, no. 4, pp. 198–204, Dec 2019. [Online]. Available: https://doi.org/10.1007/s13665-019-00243-w

[10] D. Gouzou, A. Taimori, T. Haloubi, N. Finlayson, Q. Wang, J. R. Hopgood, and M. Vallejo, "Applications of machine learning in time-domain fluorescence lifetime imaging: a review," *Methods and Applications in Fluorescence*, 2024. [Online]. Available: http://iopscience.iop.org/article/10.1088/2050-6120/ad12f7

[11] T. Haloubi, S. A. Thomas, C. Hines, K. Dhaliwal, and J. R. Hopgood, "Motion compensation in pulmonary fluorescence lifetime imaging: An image processing pipeline for artefact reduction and clinical precision," *IEEE Open Journal of Engineering in Medicine and Biology*, pp. 1–11, 2025.

[12] A. T. Erdogan, R. Walker, N. Finlayson, N. Krstajic, G. Williams, J. Girkin, and R. Henderson, "A CMOS SPAD line sensor with per-pixel histogramming TDC for time-resolved multispectral imaging," *IEEE Journal of Solid-State Circuits*, vol. 54, 6 2019.

[13] E. Williams, G. O. S. Williams, A. T. Erdogan, N. Finlayson, R. K. Henderson, and J. M. Girkin, "High speed spectral fluorescence lifetime imaging for life science applications," *High-Speed Biomedical Imaging and Spectroscopy IV*, 3 2019. [Online]. Available: https://doi.org/10.1117/12.2509466

[14] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*. Boston, MA: Springer US, 2006. [Online]. Available: http://link.springer.com/10.1007/978-0-387-46312-4

[15] L. Marcu, "Fluorescence lifetime techniques in medical applications," *Annals of Biomedical Engineering*, vol. 40, no. 2, pp. 304–331, Feb 2012, epub 2012 Jan 25. Research supported by NIH grants R21 RR 025818, R01 HL067377, R42 CA117286. [Online]. Available: https://doi.org/10.1007/s10439-011-0495-y

[16] A. Rudkouskaya, J. T. Smith, M. Ochoa, X. Intes, and M. Barroso, "Quantification of anti-HER2 drug uptake into human breast cancer cells using fluorescence lifetime FRET imaging," in *Quantification of anti-HER2 drug uptake into human breast cancer cells using fluorescence lifetime FRET imaging*, C. L. Evans and K. F. Chan, Eds. SPIE, 2 2020.

[17] T. H. Craven, T. Walton, A. R. Akram, E. Scholefield, N. McDonald, A. D. Marshall, D. C. Humphries, B. Mills, T. A. Campbell, A. Bruce, J. Mair, J. W. Dear, D. E. Newby, A. T. Hill, T. S. Walsh, C. Haslett, and K. Dhaliwal, "Activated neutrophil fluorescent imaging technique for human lungs," *Scientific Reports*, vol. 11, 12 2021. [Online]. Available: https://doi.org/10.1038/s41598-020-80083-w

[18] D. Gorpas, J. Phipps, J. Bec, D. Ma, S. Dochow, D. Yankelevich, J. Sorger, J. Popp, A. Bewley, R. Gandour-Edwards, L. Marcu, and D. G. Farwell, "Autofluorescence lifetime augmented reality as a means for real-time robotic surgery guidance in human patients," *Scientific Reports*, vol. 9, p. 1187, 12 2019.

[19] H. Sparks, H. Kondo, S. Hooper, I. Munro, G. Kennedy, C. Dunsby, P. French, and E. Sahai, "Heterogeneity in tumor chromatin-doxorubicin binding revealed by in vivo fluorescence lifetime imaging confocal endomicroscopy," *Nature Communications*, vol. 9, 12 2018.

[20] V. Mannam, Y. Zhang, X. Yuan, C. Ravasio, and S. S. Howard, "Machine learning for faster and smarter fluorescence lifetime imaging microscopy," *Journal of Physics: Photonics*, vol. 2, 9 2020.

[21] A. D. Thrapp and M. R. Hughes, "Automatic motion compensation for structured illumination endomicroscopy using a flexible fiber bundle," *Journal of Biomedical Optics*, vol. 25, 2 2020.

[22] V. Zickus, M.-L. Wu, K. Morimoto, V. Kapitany, A. Fatima, A. Turpin, R. Insall, J. Whitelaw, L. Machesky, C. Bruschini, D. Faccio, and E. Charbon, "Fluorescence lifetime imaging with a megapixel SPAD camera and neural network lifetime estimation," *Scientific Reports*, vol. 10, 12 2020.

[23] J. Unger, C. Hebisch, J. E. Phipps, J. L. Lagarto, H. Kim, M. A. Darrow, R. J. Bold, and L. Marcu, "Real-time diagnosis and visualization of tumor margins in excised breast specimens using fluorescence lifetime imaging and machine learning," *Biomedical Optics Express*, vol. 11, 3 2020.

[24] S. Guo, A. Silge, H. Bae, T. Tolstik, T. Meyer, G. Matziolis, M. Schmitt, J. Popp, and T. Bocklitz, "Flim data analysis based on laguerre polynomial decomposition and machine-learning," *Journal of Biomedical Optics*, vol. 26, 1 2021.

[25] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "Deep learning in medical image registration: a review," *Physics in Medicine & Biology*, vol. 65, p. 20TR01, 10 2020.

[26] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885603001379

[27] Q. Wang, J. R. Hopgood, N. Finlayson, G. O. S. Williams, S. Fernandes, E. Williams, A. Akram, K. Dhaliwal, and M. Vallejo, "Deep learning in ex-vivo lung cancer discrimination using fluorescence lifetime endomicroscopic

images," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 7 2020. [Online]. Available: https://doi.org/10.1109/embc44109.2020.9175598

[28] A. B. Lumb and C. R. Thomas, *Nunn and Lumb's applied respiratory physiology*, ninth edition. ed. Edinburgh?: Elsevier, 2021.

[29] C. Novak, M. N. Ballinger, and S. Ghadiali, "Mechanobiology of pulmonary diseases: A review of engineering tools to understand lung mechanotransduction," *Journal of Biomechanical Engineering*, vol. 143, no. 11, p. 110801, Nov 2021.

[30] C. T. Robb, K. H. Regan, D. A. Dorward, and A. G. Rossi, "Key mechanisms governing resolution of lung inflammation," *Seminars in Immunopathology*, vol. 38, 7 2016.

[31] S. M. Lyon and M. D. Rossman, "Pulmonary tuberculosis," *Microbiology Spectrum*, vol. 5, 2 2017.

[32] M. Cristofaro, E. B. Rizzi, P. Piselli, E. Pianura, A. Petrone, N. Fusco, F. D. Stefano, and V. Schinina', "Image quality and radiation dose reduction in chest CT in pulmonary infection," *La radiologia medica*, vol. 125, pp. 451–460, 5 2020.

[33] S. M. Triphan, G. Bauman, P. Konietzke, M. Konietzke, M. O. Wielpütz, and I. W. for Pulmonary Functional Imaging (IWPFI), "Magnetic resonance imaging of lung perfusion," *Journal of Magnetic Resonance Imaging*, vol. 59, no. 3, pp. 784–796, 2024.

[34] V. Dilsizian and Y. Chandrashekhar, "Hybrid imaging for better diagnosis and treatment strategies: Hope, hype, and reality," *JACC Cardiovasc Imaging*, vol. 17, no. 4, pp. 467–469, Apr 2024, pMID: 38569799.

[35] M. Marsden, T. Fukazawa, Y.-C. Deng, B. W. Weyers, J. Bec, D. G. Farwell, and L. Marcu, "Flimbrush: dynamic visualization of intraoperative free-hand fiber-based fluorescence lifetime imaging," *Biomedical Optics Express*, vol. 11, 9 2020.

[36] P. M. Potey, A. G. Rossi, C. D. Lucas, and D. A. Dorward, "Neutrophils in the initiation and resolution of acute pulmonary inflammation: understanding biological function and therapeutic potential," *The Journal of Pathology*, vol. 247, 4 2019.

[37] W. Sureshbabu and O. Mawlawi, "PET/CT imaging artifacts," *Journal of Nuclear Medicine Technology*, vol. 33, no. 3, pp. 156–161, 2005. [Online]. Available: https://tech.snmjournals.org/content/33/3/156

[38] A. Sisniega, J. W. Stayman, Q. Cao, J. Yorkston, J. H. Siewerdsen, and W. Zbijewski, "Image-based motion compensation for high-resolution extremities cone-beam CT," in *Image-based motion compensation for high-resolution extremities cone-beam CT*, D. Kontos, T. G. Flohr, and J. Y. Lo, Eds., 3 2016.

[39] A. Mocumbi, M. Humbert, A. Saxena, Z.-C. Jing, K. Sliwa, F. Thienemann, S. L. Archer, and S. Stewart, "Pulmonary hypertension," *Nature Reviews Disease Primers*, vol. 10, p. 1, 1 2024.

[40] H. Kuehl, P. Veit, S. J. Rosenbaum, A. Bockisch, and G. Antoch, "Can PET/CT replace separate diagnostic CT for cancer imaging? optimizing CT protocols for imaging cancers of the chest and abdomen," *Journal of Nuclear Medicine*, vol. 48, no. 1 suppl, pp. 45S–57S, 2007. [Online]. Available: https://jnm.snmjournals.org/content/48/1_suppl/45S

[41] M. Tækker, B. Kristjánsdóttir, O. Graumann, C. B. Laursen, and P. I. Pietersen, "Diagnostic accuracy of low-dose and ultra-low-dose CT in detection of chest pathology: a systematic review," *Clinical Imaging*, vol. 74, pp. 139–148, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S089970712030557X

[42] S. R. Hopkins, "Ventilation/perfusion relationships and gas exchange: Measurement approaches," *Comprehensive Physiology*, vol. 10, no. 3, pp. 1155–1205, Jul 2020.

[43] G. J. Criner, R. Eberhardt, S. Fernandez-Bussy, D. Gompelmann, F. Maldonado, N. Patel, P. L. Shah, D.-J. Slebos, A. Valipour, M. M. Wahidi, M. Weir, and F. J. Herth, "Interventional bronchoscopy," *American Journal of Respiratory and Critical Care Medicine*, vol. 202, pp. 29–50, 7 2020.

[44] R. Smithuis, "The radiology assistant : chest X-Ray - lung disease." [Online]. Available: https://radiologyassistant.nl/chest/chest-x-ray/lung-disease

[45] S. Capostagno, A. Sisniega, J. W. Stayman, T. Ehtiati, C. R. Weiss, and J. H. Siewerdsen, "Deformable motion compensation for interventional cone-beam CT," *Physics in Medicine & Biology*, vol. 66, 3 2021.

[46] K. E. Schlarbaum, "PET/CT imaging in lung cancer," *Journal of Nuclear Medicine Technology*, vol. 52, no. 2, pp. 91–101, 2024. [Online]. Available: https://tech.snmjournals.org/content/52/2/91

[47] A. Stadler, W. Schima, A. Ba-Ssalamah, J. Kettenbach, and E. Eisenhuber, "Artifacts in body MR imaging: Their appearance and how to eliminate

them," *European Radiology*, vol. 17, no. 5, pp. 1242–1255, May 2007. [Online]. Available: https://doi.org/10.1007/s00330-006-0470-4

[48] X. Ou, X. Chen, X. Xu, L. Xie, X. Chen, Z. Hong, H. Bai, X. Liu, Q. Chen, L. Li *et al.*, "Recent development in X-Ray imaging technology: Future and challenges," *Research*, 2021.

[49] A. R. Akram, S. V. Chankeshwara, E. Scholefield, T. Aslam, N. McDonald, A. Megia-Fernandez, A. Marshall, B. Mills, N. Avlonitis, T. H. Craven, A. M. Smyth, D. S. Collie, C. Gray, N. Hirani, A. T. Hill, J. R. Govan, T. Walsh, C. Haslett, M. Bradley, and K. Dhaliwal, "In situ identification of gram-negative bacteria in human lungs using a topical fluorescent peptide targeting lipid A," *Science Translational Medicine*, vol. 10, 10 2018.

[50] J. W. Lichtman and J.-A. Conchello, "Fluorescence microscopy," *Nature Methods*, vol. 2, no. 12, pp. 910–919, 2005. [Online]. Available: https://doi.org/10.1038/nmeth817

[51] N. Krstajić, B. Mills, I. Murray, and A. Marshall, "Low-cost high sensitivity pulsed endomicroscopy to visualize tricolor optical signatures," *Journal of Biomedical Optics*, vol. 23, 7 2018.

[52] R. A. Colyer, C. Lee, and E. Gratton, "A novel fluorescence lifetime imaging system that optimizes photon efficiency," *Microscopy Research and Technique*, vol. 71, pp. 201–213, 3 2008.

[53] W. Becker, *Advanced Time-Correlated Single Photon Counting Techniques*. Springer Berlin Heidelberg, 2005, vol. 81. [Online]. Available: http://link.springer.com/10.1007/3-540-28882-1

[54] R. Datta, T. M. Heaster, J. T. Sharick, A. A. Gillette, and M. C. Skala, "Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications," *Journal of biomedical optics*, vol. 25, p. 1, 5 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32406215/

[55] R. Datta, A. Gillette, M. Stefely, and M. C. Skala, "Recent innovations in fluorescence lifetime imaging microscopy for biology and medicine," *Journal of Biomedical Optics*, vol. 26, no. 7, p. 070603, 2021. [Online]. Available: https://doi.org/10.1117/1.JBO.26.7.070603

[56] A. Rudkouskaya, D. E. Faulkner, N. Sinsuebphon, X. Intes, and M. Barroso, "Macroscopic fluorescence lifetime-based förster resonance energy transfer imaging for quantitative ligand–receptor binding," in *Biomaterials for Cancer Therapeutics (Second Edition)*, second edition ed.,

ser. Woodhead Publishing Series in Biomaterials. Woodhead Publishing, 2020, pp. 331–363. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780081029831000120

[57] R. I. Dmitriev, X. Intes, and M. M. Barroso, "Luminescence lifetime imaging of three-dimensional biological objects," *Journal of Cell Science*, vol. 134, 5 2021.

[58] A. B. Matheson, C. Hopkinson, M. G. Tanner, and R. K. Henderson, "Fluorescence lifetime imaging with distance and ranging using a miniaturised SPAD system," *Scientific Reports*, vol. 14, p. 13285, 6 2024.

[59] C. Poudel, I. Mela, and C. F. Kaminski, "High-throughput, multiparametric, and correlative fluorescence lifetime imaging," *Methods and Applications in Fluorescence*, vol. 8, p. 024005, 4 2020.

[60] Y. J. Won, W.-T. Han, and D. Y. Kim, "Precision and accuracy of the analog mean-delay method for high-speed fluorescence lifetime measurement," *J. Opt. Soc. Am. A*, vol. 28, no. 10, pp. 2026–2032, Oct 2011. [Online]. Available: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-28-10-2026

[61] M. A. K. Sagar, K. P. Cheng, J. N. Ouellette, J. C. Williams, J. J. Watters, and K. W. Eliceiri, "Machine learning methods for fluorescence lifetime imaging (FLIM) based label-free detection of microglia," *Frontiers in Neuroscience*, vol. 14, 9 2020.

[62] J. Philip and K. Carlsson, "Theoretical investigation of the signal-to-noise ratio in fluorescence lifetime imaging," *J. Opt. Soc. Am. A*, vol. 20, no. 2, pp. 368–379, Feb 2003. [Online]. Available: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-20-2-368

[63] Y. Zhang, A. A. Khan, G. D. Vigil, and S. S. Howard, "Investigation of signal-to-noise ratio in frequency-domain multiphoton fluorescence lifetime imaging microscopy," *J. Opt. Soc. Am. A*, vol. 33, no. 7, pp. B1–B11, Jul 2016. [Online]. Available: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-33-7-B1

[64] C. Chang, D. Sud, and M. Mycek, "Fluorescence lifetime imaging microscopy," in *Digital Microscopy, 3rd Edition*, ser. Methods in Cell Biology. Academic Press, 2007, vol. 81, pp. 495–524. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0091679X06810241

[65] X. Liu, D. Lin, W. Becker, J. Niu, B. Yu, L. Liu, and J. Qu, "Fast fluorescence lifetime imaging techniques: A review on challenge and development," *Journal of Innovative Optical Health Sciences*, vol. 12, 9 2019.

[66] J. T. Smith, R. Yao, N. Sinsuebphon, A. Rudkouskaya, N. Un, J. Mazurkiewicz, M. Barroso, P. Yan, and X. Intes, "Fast fit-free analysis of fluorescence lifetime imaging via deep learning," *Proceedings of the National Academy of Sciences*, vol. 116, 11 2019.

[67] G. O. S. Williams, E. Williams, N. Finlayson, A. T. Erdogan, Q. Wang, S. Fernandes, A. R. Akram, K. Dhaliwal, R. K. Henderson, J. M. Girkin, and M. Bradley, "Full spectrum fluorescence lifetime imaging with 0.5nm spectral and 50ps temporal resolution," *Nature Communications*, vol. 12, p. 6616, 11 2021.

[68] A. J. Walsh, J. T. Sharick, M. C. Skala, and H. T. Beier, "Temporal binning of time-correlated single photon counting data improves exponential decay fits and imaging speed," *Biomedical Optics Express*, vol. 7, p. 1385, 4 2016.

[69] S. C. Warren, A. Margineanu, D. Alibhai, D. J. Kelly, C. Talbot, Y. Alexandrov, I. Munro, M. Katan, C. Dunsby, and P. M. W. French, "Rapid global fitting of large fluorescence lifetime imaging microscopy datasets," *PLoS ONE*, vol. 8, p. e70687, 8 2013.

[70] X. Liu, D. Lin, W. Becker, J. Niu, B. Yu, L. Liu, and J. Qu, "Fast fluorescence lifetime imaging techniques: A review on challenge and development," *Journal of Innovative Optical Health Sciences*, vol. 12, no. 05, p. 1930003, 2019. [Online]. Available: https://doi.org/10.1142/S1793545819300039

[71] D. D.-U. Li, S. Ameer-Beg, J. Arlt, D. Tyndall, R. Walker, D. R. Matthews, V. Visitkul, J. Richardson, and R. K. Henderson, "Time-domain fluorescence lifetime imaging techniques suitable for solid-state imaging sensor arrays," pp. 5650–5669, 2012. [Online]. Available: https://www.mdpi.com/1424-8220/12/5/5650

[72] A. Taimori, D. Humphries, G. Williams, K. Dhaliwal, N. Finlayson, and J. Hopgood, "Fast and robust single-exponential decay recovery from noisy fluorescence lifetime imaging," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 12, pp. 3703–3716, 2022.

[73] Q. Wang, S. Fernandes, G. O. S. Williams, N. Finlayson, A. R. Akram, K. Dhaliwal, J. R. Hopgood, and M. Vallejo, "Deep learning-assisted co-registration of full-spectral autofluorescence lifetime microscopic images with H&E-stained histology images," *Communications Biology*, vol. 5, p. 1119, 10 2022.

[74] H. E. Parker, J. M. Stone, A. D. L. Marshall, T. R. Choudhary, R. R. Thomson, K. Dhaliwal, and M. G. Tanner, "Fibre-based

spectral ratio endomicroscopy for contrast enhancement of bacterial imaging and pulmonary autofluorescence," *Biomed. Opt. Express*, vol. 10, no. 4, pp. 1856–1869, Apr 2019. [Online]. Available: https://opg.optica.org/boe/abstract.cfm?URI=boe-10-4-1856

[75] M. Wang, F. Tang, X. Pan, L. Yao, X. Wang, Y. Jing, J. Ma, G. Wang, and L. Mi, "Rapid diagnosis and intraoperative margin assessment of human lung cancer with fluorescence lifetime imaging microscopy," *BBA Clinical*, vol. 8, pp. 7–13, 12 2017.

[76] G. Bianchetti, F. Ciccarone, M. R. Ciriolo, M. D. Spirito, G. Pani, and G. Maulucci, "Label-free metabolic clustering through unsupervised pixel classification of multiparametric fluorescent images," *Analytica Chimica Acta*, vol. 1148, 3 2021.

[77] V. Mannam, Y. Zhang, X. Yuan, and S. Howard, "Deep learning-based super-resolution fluorescence microscopy on small datasets," in *Deep learning-based super-resolution fluorescence microscopy on small datasets*, I. Gregor, R. Erdmann, and F. Koberling, Eds. SPIE, 3 2021.

[78] J. T. Smith, M. Ochoa, and X. Intes, "UNMIX-ME: spectral and lifetime fluorescence unmixing via deep learning," *Biomedical Optics Express*, vol. 11, 7 2020.

[79] J. B. Pawley, Ed., *Handbook Of Biological Confocal Microscopy*. Springer US, 2006.

[80] D. B. Murphy and M. W. Davidson, *Fundamentals of Light Microscopy and Electronic Imaging*. Wiley, 9 2012.

[81] S. K. Maji and H. Yahia, "Image denoising in fluorescence microscopy using feature based gradient reconstruction," *Journal of Medical Imaging*, vol. 10, 12 2023.

[82] J. Jonkman, C. M. Brown, G. D. Wright, K. I. Anderson, and A. J. North, "Tutorial: guidance for quantitative confocal microscopy," *Nature Protocols*, vol. 15, pp. 1585–1611, 5 2020.

[83] J. R. Lakowicz, H. Szmacinski, K. Nowaczyk, K. W. Berndt, and M. Johnson, "Fluorescence lifetime imaging," *Analytical Biochemistry*, vol. 202, pp. 316–330, 5 1992.

[84] G. Weber, "Resolution of the fluorescence lifetimes in a heterogeneous system by phase and modulation measurements," *The Journal of Physical Chemistry*, vol. 85, no. 8, pp. 949–953, Apr 1981. [Online]. Available: https://doi.org/10.1021/j150608a006

[85] A. T. Erdogan, N. Finlayson, G. S. Williams, E. Williams, and R. K. Henderson, "Video rate spectral fluorescence lifetime imaging with a 512 x 16 SPAD line sensor," in *Video rate spectral fluorescence lifetime imaging with a 512 x 16 SPAD line sensor*, K. Goda and K. K. Tsia, Eds. SPIE, 3 2019.

[86] M. Demirel, B. Mills, E. Gaughan, K. Dhaliwal, and J. R. Hopgood, "Bayesian statistical analysis for bacterial detection in pulmonary endomicroscopic fluorescence lifetime imaging," *IEEE Transactions on Image Processing*, vol. 33, pp. 1241–1256, 2024.

[87] J. C. Hogg, F. Chu, S. Utokaparch, R. Woods, W. M. Elliott, L. Buzatu, R. M. Cherniack, R. M. Rogers, F. C. Sciurba, H. O. Coxson, and P. D. Paré, "The nature of small-airway obstruction in chronic obstructive pulmonary disease," *New England Journal of Medicine*, vol. 350, pp. 2645–2653, 6 2004.

[88] J. V. Brugniaux, G. B. Coombs, O. F. Barak, Z. Dujic, M. S. Sekhon, and P. N. Ainslie, "Highs and lows of hyperoxia: physiological, performance, and clinical aspects," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 315, pp. R1–R27, 7 2018.

[89] D. A. Low, M. Nystrom, E. Kalinin, P. Parikh, J. F. Dempsey, J. D. Bradley, S. Mutic, S. H. Wahab, T. Islam, G. Christensen, D. G. Politte, and B. R. Whiting, "A method for the reconstruction of four-dimensional synchronized CT scans acquired during free breathing," *Medical Physics*, vol. 30, pp. 1254–1263, 6 2003.

[90] J. Biederer, S. Mirsadraee, M. Beer, F. Molinari, C. Hintze, G. Bauman, M. Both, E. J. R. V. Beek, J. Wild, and M. Puderbach, "MRI of the lung (3/3)—current applications and future perspectives," *Insights into Imaging*, vol. 3, pp. 373–386, 1 2012.

[91] L. Lonzetti, M. Zanon, G. S. Pacini, S. Altmayer, D. M. de Oliveira, A. S. Rubin, F. F. Gazzoni, M. C. Barros, and B. Hochhegger, "Magnetic resonance imaging of interstitial lung diseases: A state-of-the-art review," *Respiratory Medicine*, vol. 155, pp. 79–85, 8 2019.

[92] J. Fang, A. Swain, R. Unni, and Y. Zheng, "Decoding optical data with machine learning," *Laser & Photonics Reviews*, vol. 15, 2 2021.

[93] M.-D. Li, H.-T. Hu, S.-M. Ruan, M.-Q. Cheng, L.-D. Chen, Z.-R. Huang, W. Li, P. Lin, H. Yang, M. Kuang, M.-D. Lu, Q.-H. Huang, and W. Wang, "ADMNet: Adaptive-weighting dual mapping for online tracking with

respiratory motion estimation in contrast-enhanced ultrasound," *IEEE Transactions on Image Processing*, vol. 33, pp. 58–68, 2024. [Online]. Available: https://doi.org/10.1109/TIP.2023.3333195

[94] G. Bauman, O. Pusterla, and O. Bieri, "Ultra-fast steady-state free precession pulse sequence for fourier decomposition pulmonary MRI," *Magnetic Resonance in Medicine*, vol. 75, no. 4, pp. 1647–1653, 2016. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.25697

[95] F. Rengier, C. Melzig, T. Derlin, A. M. Marra, and J. Vogel-Claussen, "Advanced imaging in pulmonary hypertension: emerging techniques and applications," *The International Journal of Cardiovascular Imaging*, vol. 35, no. 8, pp. 1407–1420, Aug 2019. [Online]. Available: https://doi.org/10.1007/s10554-018-1448-4

[96] N. S. Alexander, G. Palczewska, P. Stremplewski, M. Wojtkowski, T. S. Kern, and K. Palczewski, "Image registration and averaging of low laser power two-photon fluorescence images of mouse retina," *Biomed Opt Express*, vol. 7, no. 7, pp. 2671–2691, Jun 20 2016. [Online]. Available: https://doi.org/10.1364/BOE.7.002671

[97] H. Kaur, D. Koundal, and V. Kadyan, "Image fusion techniques: A survey," *Archives of Computational Methods in Engineering*, vol. 28, no. 7, pp. 4425–4447, 2021. [Online]. Available: https://doi.org/10.1007/s11831-021-09540-7

[98] D. S. Greenberg and J. N. Kerr, "Automated correction of fast motion artifacts for two-photon imaging of awake animals," *Journal of Neuroscience Methods*, vol. 176, no. 1, pp. 1–15, Jan 15 2009. [Online]. Available: https://doi.org/10.1016/j.jneumeth.2008.08.020

[99] S. Neelakantan, T. Mukherjee, K. J. Myers, R. Rizi, and R. Avazmohammadi, "Physics-informed motion registration of lung parenchyma across static CT images," *ArXiv [Preprint]*, p. arXiv:2407.03457v1, Jul 3 2024. [Online]. Available: https://arxiv.org/abs/2407.03457v1

[100] S. Abbasi, M. Tavakoli, H. R. Boveiri, M. A. M. Shirazi, R. Khayami, H. Khorasani, R. Javidan, and A. Mehdizadeh, "Medical image registration using unsupervised deep neural network: A scoping literature review," *Biomedical Signal Processing and Control*, vol. 73, p. 103444, 3 2022.

[101] A. I. Abidi and S. K. Singh, *Deformable Registration Techniques for Thoracic CT Images: An Insight into Medical Image Registration*. Singapore: Springer Singapore Pte. Limited, 2020.

[102] F. P. Oliveira and J. M. R. Tavares, "Medical image registration: a review," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 17, pp. 73–93, 1 2014.

[103] X. Chen, X. Wang, K. Zhang, K.-M. Fung, T. C. Thai, K. Moore, R. S. Mannel, H. Liu, B. Zheng, and Y. Qiu, "Recent advances and clinical applications of deep learning in medical image analysis," *Medical Image Analysis*, vol. 79, p. 102444, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841522000913

[104] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Transactions on Medical Imaging*, vol. 32, pp. 1153–1190, 7 2013.

[105] G. Song, J. Han, Y. Zhao, Z. Wang, and H. Du, "A review on medical image registration as an optimization problem," *Current Medical Imaging Reviews*, vol. 13, 7 2017.

[106] J. Unger, T. Sun, Y.-L. Chen, J. E. Phipps, and R. J. Bold, "Method for accurate registration of tissue autofluorescence imaging data with corresponding histology: a means for enhanced tumor margin assessment," *Journal of Biomedical Optics*, vol. 23, p. 1, 1 2018.

[107] M. Wang and P. Li, "A review of deformation models in medical image registration," *Journal of Medical and Biological Engineering*, vol. 39, pp. 1–17, 2 2019.

[108] T. Nguyen, E.-A. Park, J. Han, D.-C. Park, and S.-Y. Min, "Object detection using scale invariant feature transform," in *Genetic and Evolutionary Computing*, J.-S. Pan, P. Krömer, and V. Snášel, Eds. Cham: Springer International Publishing, 2014, pp. 65–72.

[109] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Computer Vision – ECCV 2006*, A. Leonardis, H. Bischof, and A. Pinz, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 404–417.

[110] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *2011 International Conference on Computer Vision*, 2011, pp. 2564–2571.

[111] R. Fabbri, L. D. F. Costa, J. C. Torelli, and O. M. Bruno, "2D euclidean distance transform algorithms: A comparative survey," *ACM Comput. Surv.*, vol. 40, no. 1, Feb. 2008. [Online]. Available: https://doi.org/10.1145/1322432.1322434

[112] J. Zhang, K. Zhang, W. Niu, and J. Huang, "SAR image automatic registration based on PCA-SIFT and mahalanobis distance," in *Proceedings of the International Conference on Information Engineering and Applications (IEA) 2012*, Z. Zhong, Ed. London: Springer London, 2013, pp. 445–452.

[113] G. Evangelidis and E. Psarakis, "Parametric image alignment using enhanced correlation coefficient maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, 10 2008. [Online]. Available: https://doi.org/10.1109/TPAMI.2008.113

[114] D. Mattes, D. Haynor, H. Vesselle, T. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," *IEEE Transactions on Medical Imaging*, vol. 22, no. 1, pp. 120–128, Jan 2003.

[115] Y. Douini, J. Riffi, M. A. Mahraz, and H. Tairi, "Solving sub-pixel image registration problems using phase correlation and lucas-kanade optical flow method," in *2017 Intelligent Systems and Computer Vision (ISCV)*, April 2017, pp. 1–5.

[116] J. Zou, B. Gao, Y. Song, and J. Qin, "A review of deep learning-based deformable medical image registration," *Frontiers in Oncology*, vol. 12, 12 2022.

[117] A. Masullo and R. Theunissen, "On dealing with multiple correlation peaks in PIV," *Experiments in Fluids*, vol. 59, p. 89, 5 2018.

[118] D. Caulfield and K. Dawson-Howe, "Track validation using gradient-based normalised cross-correlation," in *Procedings of the British Machine Vision Conference 2011*. British Machine Vision Association, 2011, pp. 70.1–70.11.

[119] Z. Liu, C. Huang, and J. Luo, "A systematic investigation of lateral estimation using various interpolation approaches in conventional ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 8, pp. 1149–1160, 2017.

[120] P. Sebastian and Y. V. Voon, "Tracking using normalized cross correlation and color space," in *2007 International Conference on Intelligent and Advanced Systems*, Nov 2007, pp. 770–774.

[121] S. Guo, A. Silge, H. Bae, T. Tolstik, T. Meyer, G. Matziolis, M. Schmitt, J. Popp, and T. Bocklitz, "Flim data analysis based on laguerre polynomial decomposition and machine-learning," *Journal of Biomedical Optics*, vol. 26, no. 2, p. 022909, 2021. [Online]. Available: https://doi.org/10.1117/1.JBO.26.2.022909

[122] T. Haloubi, S. A. Thomas, C. Hines, K. Dhaliwal, and J. R. Hopgood, "Navigating noise and texture: Motion compensation methodology for fluorescence lifetime imaging in pulmonary research," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024, pp. 1–5.

[123] U. I. Database, "Broadtz brick wall texture dataset," https://sipi.usc.edu/database/database.php?volume=textures, 2023, [Online; accessed 17-December-2023].

[124] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Image Analysis*, J. Bigun and T. Gustavsson, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 363–370.

[125] Y. Adato, T. Zickler, and O. Ben-Shahar, "A polar representation of motion and implications for optical flow," in *CVPR 2011*, June 2011, pp. 1145–1152.

[126] J. TANG and Q. LI, "Fast template matching algorithm," *Journal of Computer Applications*, vol. 30, pp. 1559–1561, 6 2010.

[127] J.-C. Yoo and T. H. Han, "Fast normalized cross-correlation," *Circuits, Systems and Signal Processing*, vol. 28, pp. 819–843, 12 2009.

[128] A. Lukežič, T. Vojíř, L. Čehovin Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter tracker with channel and spatial reliability," *International Journal of Computer Vision*, vol. 126, pp. 671–688, 7 2018.

[129] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.

[130] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 4 2004.

[131] J. C. Mier, E. Huang, H. Talebi, F. Yang, and P. Milanfar, "Deep perceptual image quality assessment for compression," in *2021 IEEE International Conference on Image Processing (ICIP)*, Sep. 2021, pp. 1484–1488.

[132] SciPy, "ttest_rel — scipy v1.15.2 manual." [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_rel.html

[133] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1153–1190, July 2013.

[134] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.", 2022.

[135] N. Gunantara, "A review of multi-objective optimization: Methods and its applications," *Cogent Engineering*, vol. 5, no. 1, p. 1502242, 2018. [Online]. Available: https://doi.org/10.1080/23311916.2018.1502242

[136] K. V. Price, R. M. Storn, and J. A. Lampinen, "The differential evolution algorithm," *Differential evolution: a practical approach to global optimization*, pp. 37–134, 2005.

[137] P. Pardalos and V. Yatsenko, *Optimization and Control of Bilinear Systems: Theory, Algorithms, and Applications*, ser. Springer Optimization and Its Applications. Springer US, 2010. [Online]. Available: https://books.google.co.uk/books?id=w9aIeKYVxsIC

[138] M. P. Heinrich, M. Jenkinson, B. W. Papiez, S. M. Brady, and J. A. Schnabel, "Towards realtime multimodal fusion for image-guided interventions using self-similarities," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8149 LNCS, pp. 187–194, 2013.

[139] C. Wachinger and N. Navab, "Entropy and laplacian images: Structural representations for multi-modal registration," *Medical Image Analysis*, vol. 16, pp. 1–17, 1 2012.

[140] H. O. Velesaca, G. Bastidas, M. Rouhani, and A. D. Sappa, "Multimodal image registration techniques: a comprehensive survey," *Multimedia Tools and Applications*, vol. 83, pp. 63 919–63 947, 1 2024.

[141] J. P. Pluim, J. B. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE transactions on medical imaging*, vol. 22, pp. 986–1004, 8 2003. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/12906253/

[142] M. Gong, S. Zhao, L. Jiao, D. Tian, and S. Wang, "A novel coarse-to-fine scheme for automatic image registration based on sift and mutual information," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 7, pp. 4328–4338, 2014.

[143] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Transactions on Mathematical Software*, vol. 23, pp. 550–560, 12 1997. [Online]. Available: https://dl.acm.org/doi/10.1145/279232.279236

[144] D. Gorpas, H. Fatakdawala, J. Bec, D. Ma, D. R. Yankelevich, J. Qi, and L. Marcu, "Fluorescence lifetime imaging and intravascular ultrasound: Co-registration study using ex vivo human coronaries," *IEEE Transactions on Medical Imaging*, vol. 34, pp. 156–166, 1 2015.

[145] G. Di Pillo, S. Lucidi, and L. Palagi, "A truncated Newton method for constrained optimization," in *Nonlinear Optimization and Related Topics*, G. Di Pillo and F. Giannessi, Eds. Boston, MA: Springer US, 2000, pp. 79–103. [Online]. Available: https://doi.org/10.1007/978-1-4757-3226-9_5

[146] S. C. Endres, C. Sandrock, and W. W. Focke, "A simplicial homology algorithm for lipschitz optimisation," *Journal of Global Optimization*, vol. 72, pp. 181–217, 10 2018. [Online]. Available: https://link.springer.com/article/10.1007/s10898-018-0645-y

[147] H. Abdi and L. J. Williams, "Tukey's honestly significant difference (HSD) test," *Encyclopedia of research design*, 2010.

[148] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python," in *9th Python in Science Conference*, 2010.

[149] Y. Ye, B. Zhu, T. Tang, C. Yang, Q. Xu, and G. Zhang, "A robust multimodal remote sensing image registration method and system using steerable filters with first- and second-order gradients," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 188, pp. 331–350, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271622001083

[150] F.-A. Fortin, F.-M. De Rainville, M.-A. G. Gardner, M. Parizeau, and C. Gagné, "DEAP: evolutionary algorithms made easy," *J. Mach. Learn. Res.*, vol. 13, no. 1, p. 2171–2175, Jul. 2012.

[151] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210–229, 1959.

[152] T. M. Mitchell, "Machine learning and data mining," *Communications of the ACM*, vol. 42, no. 11, pp. 30–36, 1999.

[153] A. Gogoi and N. Mazumder, Eds., *Biomedical Imaging*. Springer Nature Singapore, 2024.

[154] Q. Wang, A. R. Akram, D. A. Dorward, S. Talas, B. Monks, C. Thum, J. R. Hopgood, M. Javidi, and M. Vallejo, "Deep learning-based virtual H&E staining from label-free autofluorescence lifetime images," *npj Imaging*, vol. 2, p. 17, 6 2024.

[155] Y.-I. Chen, Y.-J. Chang, S.-C. Liao, T. D. Nguyen, J. Yang, Y.-A. Kuo, S. Hong, Y.-L. Liu, H. G. Rylander, S. R. Santacruz, T. E. Yankeelov, and H.-C. Yeh, "Generative adversarial network enables rapid and robust fluorescence lifetime image analysis in live cells," *Communications Biology*, vol. 5, p. 18, 12 2022.

[156] H. Guan and M. Liu, "Domain adaptation for medical image analysis: A survey," *IEEE Transactions on Biomedical Engineering*, vol. 69, pp. 1173–1185, 3 2022.

[157] W. Xia, H. Shan, G. Wang, and Y. Zhang, "Physics-/model-based and data-driven methods for low-dose computed tomography: A survey," *IEEE Signal Processing Magazine*, vol. 40, pp. 89–100, 3 2023.

[158] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang, "LEARN: Learned experts' assessment-based reconstruction network for sparse-data CT," *IEEE Transactions on Medical Imaging*, vol. 37, pp. 1333–1347, 6 2018.

[159] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image reconstruction is a new frontier of machine learning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1289–1296, 2018.

[160] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.

[161] J. N. Hussain, "High dimensional data challenges in estimating multiple linear regression," *Journal of Physics: Conference Series*, vol. 1591, p. 012035, 7 2020.

[162] A. Maier, H. Köstler, M. Heisig, P. Krauss, and S. H. Yang, "Known operator learning and hybrid machine learning in medical imaging—a review of the past, the present, and the future," *Progress in Biomedical Engineering*, vol. 4, p. 022002, 4 2022.

[163] K. Gong, J. Guan, C.-C. Liu, and J. Qi, "PET image denoising using a deep neural network through fine tuning," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, pp. 153–161, 3 2019.

[164] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[165] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions,"

in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[166] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[167] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *CVPR09*, 2009.

[168] C. Parmar, J. D. Barry, A. Hosny, J. Quackenbush, and H. J. Aerts, "Data analysis strategies in medical imaging," *Clinical Cancer Research*, vol. 24, pp. 3492–3499, 8 2018.

[169] M. Gaillochet, C. Desrosiers, and H. Lombaert, "Active learning for medical image segmentation with stochastic batches," *Medical Image Analysis*, vol. 90, p. 102958, 12 2023.

[170] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Proceedings of the 3rd International Conference on Learning Representations*, vol. abs/1412.6980, 2014.

[171] A. Jungo and M. Reyes, *Assessing Reliability and Challenges of Uncertainty Estimations for Medical Image Segmentation*. Springer, Cham, 2019, pp. 48–56.

[172] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[173] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Springer, 2015, pp. 234–241.

[174] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.

[175] T. Falk, D. Mai, R. Bensch, O. Cicek, A. Abdulkadir, Y. Marrakchi, A. Bohm, J. Deubner, Z. Jäckel, K. Seiwald, A. Dovzhenko, O. Tietz, C. D. Bosco, S. Walsh, D. Saltukoglu, T. L. Tay, M. Prinz, K. Palme, M. Simons, I. Diester, T. Brox, and O. Ronneberger, "U-Net: deep learning for cell counting, detection, and morphometry," *Nature Methods*, vol. 16, no. 1, pp. 67–70, Jan 2019, erratum in: Nat Methods. 2019 Apr;16(4):351. doi: 10.1038/s41592-019-0356-4. [Online]. Available: https://doi.org/10.1038/s41592-018-0261-2

[176] R. Gonzalez and R. Woods, *Digital Image Processing*. Pearson, 2018.

[177] C. Pham and B. A. Plummer, "Enhancing feature diversity boosts channel-adaptive vision transformers," *Computer Vision and Pattern Recognition*, 5 2024. [Online]. Available: https://arxiv.org/abs/2405.16419v2

[178] M. Gu, C. J. R. Sheppard, and X. Gan, "Image formation in a fiber-optical confocal scanning microscope," *J. Opt. Soc. Am. A*, vol. 8, no. 11, pp. 1755–1761, Nov 1991. [Online]. Available: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-8-11-1755

[179] F. C. Ghesu, B. Georgescu, A. Mansoor, Y. Yoo, E. Gibson, R. Vishwanath, A. Balachandran, J. M. Balter, Y. Cao, R. Singh, S. R. Digumarthy, M. K. Kalra, S. Grbic, and D. Comaniciu, "Quantifying and leveraging predictive uncertainty for medical image assessment," *Medical Image Analysis*, vol. 68, p. 101855, 2 2021.

[180] M. Chen, N. J. Tustison, R. Jena *et al.*, "Image registration: Fundamentals and recent advances based on deep learning," in *Machine Learning for Brain Disorders*, O. Colliot, Ed. New York, NY, USA: Humana, 2023, ch. 14. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK597490/

[181] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, and R. Ranjan, "Explainable ai (xai): Core ideas, techniques, and solutions," *ACM Comput. Surv.*, vol. 55, no. 9, Jan. 2023. [Online]. Available: https://doi.org/10.1145/3561048

[182] Z. Yang, T. Dan, and Y. Yang, "Multi-temporal remote sensing image registration using deep convolutional features," *IEEE Access*, vol. 6, pp. 38 544–38 555, 2018.

[183] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *CoRR*, vol. abs/2010.11929, 2020. [Online]. Available: https://arxiv.org/abs/2010.11929

[184] C. Ma, W. Tan, R. He, and B. Yan, "Pretraining a foundation model for generalizable fluorescence microscopy-based image restoration," *Nature Methods*, vol. 21, pp. 1558–1567, 8 2024.

[185] M. Tan, F. Wu, B. Yang, J. Ma, D. Kong, Z. Chen, and D. Long, "Pulmonary nodule detection using hybrid two-stage 3D CNNs," *Medical Physics*, vol. 47, pp. 3376–3388, 8 2020.

[186] M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A. I. Aviles-Rivero, C. Etmann, C. McCague, L. Beer, J. R. Weir-McCall, Z. Teng, E. Gkrania-Klotsas, J. H. F. Rudd, E. Sala, and C.-B. Schönlieb, "Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans," *Nature Machine Intelligence*, vol. 3, 3 2021.

[187] A. Murphy, R. Chieng, and C. Moore, "Digital imaging and communications in medicine (DICOM)," 5 2017.

[188] N. Arun, N. Gaw, P. Singh, K. Chang, M. Aggarwal, B. Chen, K. Hoebel, S. Gupta, J. Patel, M. Gidwani, J. Adebayo, M. D. Li, and J. Kalpathy-Cramer, "Assessing the trustworthiness of saliency maps for localizing abnormalities in medical imaging," *Radiology: Artificial Intelligence*, vol. 3, 11 2021.