THE UNIVERSITY *of* EDINBURGH

Novel Applications of Signed Distance Fields in 3D Reconstruction of Thin Structures

Salvatore Esposito

Doctor of Philosophy

CDT Biomedical Artificial Intelligence

School of Informatics

The University of Edinburgh

2025

Abstract

Three-dimensional reconstruction from sparse or incomplete observations presents a fundamental challenge in computer vision and graphics. Traditional approaches using voxels, point clouds, or meshes often struggle with floating geometric artifacts, thin geometric structures, and preserving topological continuity. This thesis introduces novel methodologies using neural signed distance fields (SDFs) to address these long-standing limitations in reconstructing complex geometry featuring thin structures, intricate branching patterns, and high-frequency details. Our research follows a deliberate progression addressing increasingly challenging scenarios from general object geometry to thin, complex structures in specialized medical domains.

Our research begins with GeoGen, a novel SDF-based 3D generative model trained end-to-end from single-view collections of 2D images. GeoGen reinterprets volumetric density as an SDF, introducing geometric priors for valid mesh generation through an SDF depth-map consistency loss. This approach enables more coherent surface representations than traditional neural rendering methods by enforcing geometric constraints throughout the learning process. By leveraging the mathematical properties of SDFs, GeoGen generates surfaces with improved structural integrity and avoids the floating artifacts common in neural radiance-field methods. We evaluated GeoGen on ShapeNet Cars and a 19 800-identity Synthetic Human Heads benchmark, achieving 20% lower Chamfer distance than EG3D on both datasets while matching its 2D FID scores. CrossSDF cut Chamfer distance by an order of magnitude on thin-structure meshes and VesselSDF improved Dice from 0.69 to 0.72 and reduced Chamfer from 0.82 to 0.68 on the Medical Decathlon Hepatic-Vessel CT set.

However, GeoGen exhibits limitations when predicting fine details and thin structures like hair or eyelashes. These limitations motivated our investigation into specialized SDF representations for thin-structure reconstruction, with medical imaging providing an ideal application domain due to its critical need for accurate depiction of complex anatomical structures. Building on these insights, we introduce CrossSDF, a novel approach for extracting a 3D SDF from 2D signed distances generated from planar contours. Our framework employs a hash-based neural reconstruction approach with three key innovations: a novel symmetric-difference loss that minimizes visual artifacts resulting from sparse cross-sectional data, an adaptive contour-sampling strategy that ensures thin structures are adequately represented during surface reconstruction, and a hybrid encoding architecture that combines a detail-preserving hash encoding with Fourier features to reduce grid-interpolation artifacts. This enables CrossSDF to pro-

duce high-fidelity 3D reconstructions of thin structures while maintaining topological consistency between slices.

However, we observed that when applied to medical data with inherent sparsity between imaging planes, the quality of the reconstruction remained fundamentally limited by the accuracy of the initial 2D contour segmentation. This observation led us to develop VesselSDF, a comprehensive two-stage framework specifically optimized for vascular-network reconstruction from sparse medical imagery. Unlike CrossSDF, which operates on pre-segmented contours, VesselSDF addresses both the segmentation and reconstruction challenges within a unified pipeline. VesselSDF addresses the reconstruction of complex structures from cross-sections by treating vessel segmentation as a continuous SDF regression problem rather than a discrete voxel classification one. The first stage employs a 3D U-Net for binary vessel-occupancy prediction. The second stage transforms this binary occupancy into an appropriately scaled SDF through a specialized refiner network with geometric constraints. VesselSDF introduces a distance-weighted Gaussian regularizer that adaptively enforces smoothness based on proximity to vessel surfaces, ensuring global smoothness while preserving critical vessel boundaries. This is complemented by a surface-regularization term that suppresses artifacts and an anisotropic Eikonal regularization term that accounts for different spatial resolutions along the axial dimension. By systematically separating segmentation from geometric refinement and incorporating these specialized constraints, VesselSDF achieves superior reconstruction of complex vascular networks compared to state-of-the-art binary voxel classification methods (nnU-Net, 3D SA-UNet), even from highly anisotropic CT and MRI data with significant interslice gaps. Through extensive experimental evaluation across diverse datasets, we demonstrate that our neural SDF approaches produce high-fidelity reconstructions that preserve thin structures, maintain topological correctness, and accurately capture geometric details, even from limited observations. The methodologies developed in this thesis have significant implications for medical imaging, enabling more accurate surgical planning through improved vessel visualization, enhanced blood flow simulation for cardiovascular analysis, early detection of vascular pathologies through preserved fine vessel details, and reduced radiation exposure in CT scanning by maintaining diagnostic quality from sparser slice data. The methodologies developed in this thesis establish a framework for addressing reconstruction problems across different domains, illustrating the versatility and effectiveness of neural SDF representations for complex-geometry reconstruction.

Acknowledgements

This journey has been a remarkable blend of ups and downs. I am deeply grateful to Kieran for the countless nights spent on Zoom calls patiently teaching me Python. My heartfelt thanks to Mania, Julia, Emily, and Kia for their unwavering support throughout the Bioinformatics Master's program. I will forever cherish our shared experiences. To my dear friends Leo, Olivier, WenYue, Tom, George, Karim, Kacper, Kasia, Arushi, Qingshan, and Octave: your friendship made my time in Edinburgh truly special. I will forever cherish and dearly miss the lunches, coffees, and nights out between Leo, Tom, Olivier, Karim and I. I hope our paths will cross once again in the future to build something extraordinary together and buy our freedom with it. My sincere gratitude to my family, Mamma, Papà, Nonna, Zio Tonino, and Zia Margherita, for their constant encouragement and love. I want to express profound appreciation to my wife, Kris, an extraordinary partner who provided comfort and strength, especially during challenging times. Catie, our collaborative work has been exceptional, and I am eternally thankful for your help and guidance throughout this very hard chapter of my life . I want to thank Gianluca for his medical and intellectual wisdom which has helped me push through even through the darkest times of this journey. I extend my gratitude to my colleagues at Microsoft and Huawei for their professional support. Special thanks to Tom, Kacper and Daniel for their invaluable teachings in Computer Graphics and 3D Vision. It has been the honor of my life to work with giants like you, who are outstanding researchers. Tom, working on CrossSDF with you was the most fun I've ever had in a research project. I will dearly miss all the ups and downs and the laughs and the obstacles along the way. Thank you for working with me on that project. Above all, I am indebted to Arno, who has been an exemplary supervisor, alongside Oisin and Changjian. From our initial work on EMG signal processing to our final projects on 3D reconstruction, you have always helped me improve my understanding of Machine Learning. Oisin, your technical expertise, writing guidance, and emphasis on meticulousness have been instrumental in my growth, though I recognize I still have much to learn. Changjian, despite our relatively brief collaboration, your insights and approach to problem-solving influenced my research direction and I'm forever grateful for our collaboration. I appreciate your patience as I developed my machine learning capabilities. I have worked to honor the time and effort you all invested in me, and I hope my work reflects the value of your mentorship.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Salvatore Esposito)*

In memory of my beloved mother MariaRosaria

Lay Abstract

Medical imaging techniques like CT and MRI scans are essential tools in modern medicine, enabling doctors to see inside the human body without invasive procedures. These scans typically produce a series of 2D image slices that medical professionals use to understand 3D structures. This becomes challenging when examining complex networks of blood vessels or other thin, intricate anatomical features that are crucial for diagnosis, and surgical planning is particularly challenging when the slices have a few millimeters of gaps between them. Current computer methods for automatically converting these 2D slices into complete 3D models often struggle with thin structures like small blood vessels. They frequently produce inaccurate results with missing connections or overly simplified shapes. These inaccuracies can potentially impact medical decisions, highlighting the need for better techniques to process and visualize this complex data. This thesis addresses these challenges by developing new computational methods that can create accurate 3D models from limited 2D scan data. Our research introduces three innovative approaches using a mathematical technique called signed distance fields, which essentially maps out the boundaries of objects by calculating distances.

First, we developed GeoGen, a model that can generate detailed 3D human faces from single-view 2D images. Geogen approaches this problem by using mathematical representations that naturally favor realistic, continuous surfaces. GeoGen addresses this challenge by leveraging mathematical representations that produce realistic, continuous surfaces. The model learns facial geometry patterns from thousands of examples, precisely calculates distances to object surfaces, and generates more accurate reconstructions by maintaining consistent surface boundaries. While successful for general shapes, we found it difficult to find fine details like hair and eyelashes, which inspired our next research direction.

We then developed CrossSDF, specifically designed for reconstructing thin structures from medical scan slices. CrossSDF introduces a new way of measuring accuracy between the predicted 3D model and the actual scan data, a strategic approach to focusing computational resources on the most important thin areas, and a technique that preserves fine details while maintaining overall smoothness. This significantly improved the quality of reconstructed thin structures, but we discovered that the initial identification of structures in the 2D slices remained a limiting factor.

This observation led to our final innovation, VesselSDF, a comprehensive model specifically for segmenting and reconstructing blood vessel networks from sparse

medical scans. VesselSDF handles both the challenge of accurately identifying vessels in the scans and reconstructing their complex 3D structure. By representing vessels as continuous surfaces rather than blocks of digital volume and by applying different levels of smoothing based on proximity to vessel surfaces, VesselSDF creates clean, accurate models that preserve critical anatomical details even when working with widely spaced scan slices.

Our evaluation across multiple datasets demonstrates that these new approaches outperform existing methods in preserving thin structures, maintaining proper connections at branching points, and reproducing fine details. We tested GeoGen on 70,000 high-quality face images (FFHQ dataset), 2,100 car models from different viewpoints (ShapeNet), and our custom dataset of 200,000 synthetic human images. CrossSDF was evaluated on both synthetic anatomical structures (including 6 patient-specific thin branching geometries) and CT scans from clinical databases (IRCADb-01 and Medical Decathlon datasets). VesselSDF was tested on 303 hepatic vessel CT scans and 20 contrast-enhanced abdominal scans from Medical Decathlon datasets, showing improved accuracy scores and better preservation of vessel connectivity compared to standard medical imaging software. The techniques developed in this thesis enable more accurate visualization of complex anatomical structures, which can potentially improve surgical planning and diagnosis. Beyond medicine, these methods could also benefit fields like geology, where creating 3D models from limited data is important. This research establishes versatile approaches for accurately reconstructing complex 3D structures from limited observations across various fields.

Table of Contents

Chapter 1

Introduction

Three-dimensional representations of objects lie at the heart of numerous scientific, industrial, and creative endeavors, enabling richer insights and more effective decision-making than their two-dimensional counterparts. By capturing both geometry and spatial context, 3D models facilitate detailed examinations of form and function, ultimately driving innovation in fields ranging from computer-aided design (CAD) to interactive virtual reality (VR) environments (Ahmed et al., 2024). In healthcare, for instance, 3D anatomical models can guide surgical procedures and improve patient outcomes by giving clinicians an immersive view of complex structures before making critical interventions. Similarly, in education and training, 3D visualization helps students grasp spatial relationships that are often lost in flat diagrams (Sawdayee et al., 2023). Despite significant advancements in imaging technologies, inferring complete 3D structures from sparse observations remains challenging, particularly for complex structures featuring thin elements, intricate branching patterns, and high-frequency details. (Bajaj et al., 1996; Barequet and Sharir, 1994; Boissonnat and Memari, 2007).

The task of inferring complete 3D structures from sparse data becomes particularly challenging when dealing with complex geometries. Traditional reconstruction approaches typically operate on discrete representations, such as voxel grids, point clouds, or polygon meshes. These methods face inherent limitations: voxel-based methods suffer from quantization artifacts and memory constraints (Gao et al., 2022b); while both point-based and mesh-based methods require complex topological operations to handle intricate structures (Ma et al., 2021b; Zou et al., 2015). Moreover, these approaches often employ explicit representations or simplistic regularization techniques which fail to capture the intrinsic geometric properties of the underlying structures, resulting in either over-smoothed reconstructions that lose critical details or noisy artifacts that

compromise structural integrity of the scenes (Bermano et al., 2011).

Recent advances in deep learning have enabled new approaches to 3D reconstruction through implicit neural representations (Gao et al., 2022b). Among these, signed distance fields (SDFs) have emerged as a particularly powerful framework for representing complex geometry (Park et al., 2019). By encoding shapes as continuous functions that map spatial coordinates to signed distances from the nearest surface boundary, SDFs offer several theoretical advantages: they naturally represent smooth surfaces, maintain consistent topology, capture arbitrary levels of detail, and enable efficient computation of geometric properties such as normals and curvature (Müller et al., 2022).

Despite these advantages, applying neural SDFs to practical reconstruction problems introduces several significant challenges. First, limited or sparse observations provide insufficient constraints on the reconstruction, requiring effective regularization strategies that preserve geometric details while suppressing artifacts (Sitzmann et al., 2020). Second, anisotropic sampling patterns, particularly common in medical imaging where in-plane resolution far exceeds through-plane resolution, create directional biases that must be carefully addressed (Çiçek et al., 2016). Third, accurately reconstructing thin structures remains problematic due to their small cross-sectional area and the tendency of neural networks to favor simpler, smoother solutions (Sawdayee et al., 2023).

1.1 Motivation

The reconstruction of complex 3D structures from limited observations is a critical capability across numerous domains, driving research into more effective and robust methodological approaches. Access to accurate 3D meshes is vital for numerous applications: in medicine they enable precise diagnosis, treatment planning and surgical navigation (Sawdayee et al., 2023), in engineering they facilitate structural analysis, simulation, design optimization; in computer graphics, they support realistic rendering, animation, and interactive applications (Boissonnat and Memari, 2007). Yet obtaining complete 3D meshes remains difficult due to physical constraints, time limitations and the inherent sparsity of many input modalities.

Several key factors motivate our exploration of neural SDF-based methods for this challenging task. Traditional reconstruction methods face significant limitations when applied to complex structures, particularly those with thin features or intricate branching patterns.Voxel-based approaches suffer from limited resolution, quantization artifacts, and inability to represent sub-voxel structures without prohibitive memory requirements

(Schwarz et al., 2022). Mesh-based methods struggle with topology changes between observations and often require complex correspondence algorithms that fail in cases of branching or merging structures (Huber et al., 1985). Point cloud techniques lack explicit surface information and require substantial post-processing to recover coherent surfaces, often introducing smoothing artifacts that erase fine details (Qi et al., 2017). Physics-based interpolation methods typically rely on simplified assumptions about shape continuity that fail to capture the complex variation present in real-world structures, particularly in medical contexts (Bermano et al., 2011). These limitations become especially pronounced when working with thin structures like vascular networks, where small errors in reconstruction can lead to topological inconsistencies or physically implausible results (Al Moussawi et al., 2015).

Neural SDFs offer several compelling advantages for addressing these reconstruction challenges. Their continuous representation naturally handles arbitrary resolution without discretization artifacts, making them particularly suited for representing structures at varying scales (Wang et al., 2021). They benefit from implicit regularization through neural networks' inherent bias toward smooth, continuous functions, providing built-in regularization that helps interpolate between sparse observations in a geometrically plausible manner (Yariv et al., 2021). Furthermore, neural SDFs offer topological flexibility, seamlessly handling complex topologies, branching structures, and thin features without requiring explicit topology handling or remeshing operations (Müller et al., 2022). Their end-to-end differentiable nature enables integration of various forms of supervision and domain-specific constraints through carefully designed loss functions (Gropp et al., 2020a).

The development of improved reconstruction methods has significant implications across multiple domains. In medical imaging, accurate reconstruction of anatomical structures from sparse CT slices enables reduced radiation exposure for patients while maintaining diagnostic quality, while for MRI applications it reduces scan time and patient discomfort. For vascular structures specifically, high-fidelity reconstruction facilitates surgical planning, blood flow simulation, and early detection of pathologies. Each of these applications benefits from reconstruction methods that can effectively leverage domain-specific priors while maintaining geometric accuracy and topological correctness, motivating the development of the specialized neural SDF approaches presented in this thesis.

1.2 Background

Imagine trying to reconstruct the shape of a complex building from just a few photographs taken from different angles, or determining the full 3D structure of blood vessels from medical scan slices. These scenarios illustrate a fundamental challenge in computer science: how do we create complete, accurate 3D models from incomplete observations?

Traditional computer graphics has long relied on explicit 3D representations—think of polygon meshes used in video games, point clouds from 3D scanners, or voxel grids like digital LEGO blocks. Each approach has its strengths: meshes are efficient for rendering, point clouds directly represent sensor data, and voxels provide regular structure. However, all these methods struggle with a common limitation: they represent shapes using discrete building blocks, which makes it difficult to capture smooth curves, fine details, or complex internal structures (Chan et al., 2022).

Consider the difference between representing a sphere using flat triangular faces (a mesh) versus describing it mathematically as ”all points exactly 1 unit from the center.” The mathematical description is both more compact and more accurate—it captures the perfect smoothness of the sphere at any level of detail. This insight has led researchers to explore implicit representations, where shapes are defined by mathematical functions rather than discrete elements.

Recent advances in deep learning have made it possible to use neural networks to learn these mathematical shape descriptions automatically from data. Instead of hand-crafting equations for simple shapes like spheres, we can train networks to discover the mathematical patterns that describe complex, real-world objects. This approach, known as neural implicit representation, has opened new possibilities for 3D reconstruction, particularly when dealing with incomplete or sparse data. Key developments in this area include occupancy-based methods (Mescheder et al., 2019; Chen and Zhang, 2019), signed distance field approaches (Park et al., 2019), and neural rendering techniques (Wang et al., 2021; Yariv et al., 2021).

One of the most practically important reconstruction challenges involves building 3D models from 2D cross-sections or slices. This problem appears throughout science and engineering: medical professionals need to understand 3D organ structure from CT scan slices (Champetier et al., 1992), geologists reconstruct underground formations from core samples, and manufacturers analyze internal product structure from cross-sectional cuts (Bermano et al., 2011).

The core difficulty becomes apparent when we consider what information is missing. If you have slices of an apple taken every centimeter, how do you know what happens in the gaps between slices? Did the apple's core continue straight through, or did it curve? Are there seeds in the unobserved regions? Traditional approaches try to "connect the dots" by interpolating smoothly between slices (Barequet and Sharir, 1994), but this often produces unrealistic results, especially for complex structures with branches or sudden changes in shape. More advanced techniques can handle non-parallel cross-sections (Liu et al., 2008), but often suffer from computational complexity or require smoothing that harms surface accuracy (Zou et al., 2015).

This challenge becomes critical in medical applications, where thin structures like blood vessels must be accurately reconstructed for diagnosis and surgical planning (Wittmann et al., 2024; Al Moussawi et al., 2015). Blood vessels branch and merge in complex patterns, often with diameters smaller than the spacing between scan slices. A small error in reconstruction—such as missing a crucial branch or incorrectly connecting vessels—can have serious clinical consequences. Recent neural approaches have shown promise (Sawdayee et al., 2023; Fu et al., 2022), but significant challenges remain in handling the extreme sparsity between slicing planes.

The specific technical methods for implementing these neural representations and addressing these reconstruction challenges will be detailed in the following chapters where they are applied to solve particular problems. Our focus in this thesis is on developing neural methods that can handle the most challenging scenarios: sparse observations, thin structures, and complex branching patterns.

1.3 Challenges in Neural SDF Reconstruction

1.3.1 Geometric Details

Generative neural SDF methods often struggle to distinguish genuine high-frequency geometric detail from noise, particularly when trained on sparse or single-view 2D image collections (Müller et al., 2022). In conventional volumetric rendering approaches, minor inconsistencies in the learned density field can compound into major surface defects and ringing artifacts once the surface is extracted as a mesh. This effect is exacerbated where the underlying geometry contains abrupt changes or intricate features, as small errors in these high-curvature regions can lead to visually distracting fluctuations (Novello et al., 2022).

These artifacts primarily stem from a fundamental bias-variance trade-off inherent in neural SDF reconstruction from sparse data. High variance manifests as noisy, inconsistent surfaces with floating artifacts and spurious high-frequency details, while high bias leads to over-smoothed reconstructions that lose critical geometric features like sharp edges and thin structures. The optimal balance depends on both the sparsity of observations and the complexity of the target geometry.

The challenge is compounded by different sources of noise that require distinct mitigation strategies in neural SDF reconstruction. Model noise arises from the neural network's limited capacity to represent complex geometry and optimization challenges during training (Mildenhall et al., 2020). This thesis addresses model noise through architectural innovations: GeoGen employs triplane representations with positional encoding to capture high-frequency details, while CrossSDF uses hybrid hash-grid encodings that combine detail preservation with smooth interpolation. Sensor noise originates from imaging device limitations such as thermal noise in CT scanners, quantization errors, and measurement uncertainties (Antonelli et al., 2022). VesselSDF specifically addresses sensor noise through distance-weighted Gaussian regularization that adapts smoothing based on proximity to vessel surfaces, while CrossSDF handles segmentation uncertainties through its symmetric difference loss that focuses on classification disagreements rather than exact distance matching.

Environmental noise encompasses external factors affecting data acquisition, including lighting variations, specularities, and motion artifacts. In neural rendering contexts like GeoGen, this is addressed through view-dependent appearance modeling and geometric consistency constraints—our SDF depth consistency loss ensures that rendered depth maps align with the SDF's zero-level set, providing robustness against lighting-induced artifacts. Structural noise from discretization, sampling patterns, and interpolation artifacts is handled through careful SDF formulations: CrossSDF employs adaptive contour sampling to ensure thin structures receive adequate representation, while all our methods leverage the continuous nature of SDF representations to avoid the quantization artifacts inherent in voxel-based approaches.

Traditional approaches for addressing these challenges include depth-guided penalties (Deng et al., 2022), Eikonal regularization (Gropp et al., 2020b), surface normal constraints (Wang et al., 2021), and reformulating raw density outputs into signed distance representations (Yariv et al., 2021). Additional regularization techniques include smoothness penalties (Sitzmann et al., 2020), curvature constraints (Liu et al., 2008), and multi-view consistency losses (Mildenhall et al., 2020). However, applying global

smoothing constraints often results in over-smoothing and loss of critical geometric details. The fundamental challenge remains: how to preserve high-frequency structure in the data while simultaneously reducing noise without artificially blurring important features like sharp edges and localized surface details.

1.3.2 Thin Structure Reconstruction

Thin structures, such as blood vessels, pose unique obstacles for neural SDF methods. When their diameters approach or fall below the sampling resolution, thin features can become fragmented or vanish entirely in the raw data (Wang et al., 2022). These issues are amplified by traditional biases toward smoothness or simpler geometry, causing delicate surfaces to be prematurely fused, disconnected, or overly blurred.

High surface-to-volume ratios make these structures especially prone to topological errors. Minute inaccuracies in the SDF near steep gradients may produce large-scale topological flaws, disconnecting branches or creating artificial connections (Müller et al., 2022). Moreover, neural networks inherently favor lower-frequency representations, a tendency that can exacerbate the loss of narrow, detailed boundaries (Gao et al., 2022b).

Several approaches have been proposed to address these challenges, including specialized SDF regularization techniques (Gropp et al., 2020b; Wang et al., 2021), topology-preserving loss functions (Zou et al., 2015), and adaptive multi-resolution encodings (Müller et al., 2022). Distance-weighted regularization (Atzmon and Lipman, 2020) and surface-aware sampling (Yifan et al., 2021) have shown particular promise for preserving thin geometric features (Li et al., 2023b). The central difficulty lies in balancing smoothness far from the surface while maintaining sharp boundaries where thin anatomy requires higher-frequency detail. Accurately representing high-curvature regions when multiple boundaries lie in close proximity remains an open challenge in the field (Chetan et al., 2023; Novello et al., 2022).

1.3.3 Sparse and Anisotropic Data

Sparse, anisotropic data further complicates neural SDF reconstruction. A canonical example arises in medical imaging, where in-plane resolutions can be on the order of tenths of a millimeter, but slice thickness may be a magnitude larger (Çiçek et al., 2016). This disparity fosters aliasing: subtle branching or directional changes frequently go undetected in intermediate slices, leaving naive interpolation schemes unable to capture the correct topology (Bajaj et al., 1996; Zou et al., 2015).

These anisotropic sampling patterns create directional biases in learned representations. Features aligned with densely sampled dimensions are well-preserved, while those primarily extending along sparsely sampled axes become distorted or lost entirely. The challenge intensifies when observations come from arbitrary viewing angles or when data distributions vary significantly across the volume of interest.

Recent research (Liu et al., 2020; Schwarz et al., 2022) has focused on developing solutions through carefully engineered architectures and sampling heuristics designed for sparse views. Effective approaches include symmetric difference based losses that focus learning around 2D planar observations, adaptive regularization techniques that compensate for varying data density, and multi-scale feature extraction that can bridge information gaps between sparse observations. These strategies highlight that balancing structural priors with selective local refinement is crucial when reconstructions depend on sparse, anisotropic data distributions.

1.4 Contributions

1.4.1 GeoGen: Geometry-Aware Generative Modeling via Signed Distance Fields

GeoGen (Esposito et al., 2024) addresses the challenge of synthesizing realistic surfaces from sparse single-view data, which is typically under-constrained for 3D geometry learning. By learning an SDF within a generative adversarial framework, GeoGen enforces consistency between a rendered depth map and the SDF's zero-level set. This depth-consistency constraint reduces spurious high-frequency artifacts that arise when density-based volumetric rendering is unaccompanied by strong geometric priors. In contrast to purely occupancy-driven or radiance-based methods, GeoGen reformulates the underlying 3D representation to better capture sharp boundaries and local surface details, mitigating the noise and "over-smoothing" biases common in generative models trained on limited or single-view supervision. The result is a geometry-aware generative process that yields cleaner and more plausible meshes without sacrificing image fidelity as shown in Figure 1.1. As a secondary contribution, we introduce a novel dataset of synthetic human heads captured from 360-degree viewpoints, providing complete coverage of all possible camera angles for proper assessment of full 3D consistency and geometric accuracy, addressing limitations in existing datasets like FFHQ that predominantly feature frontal views.

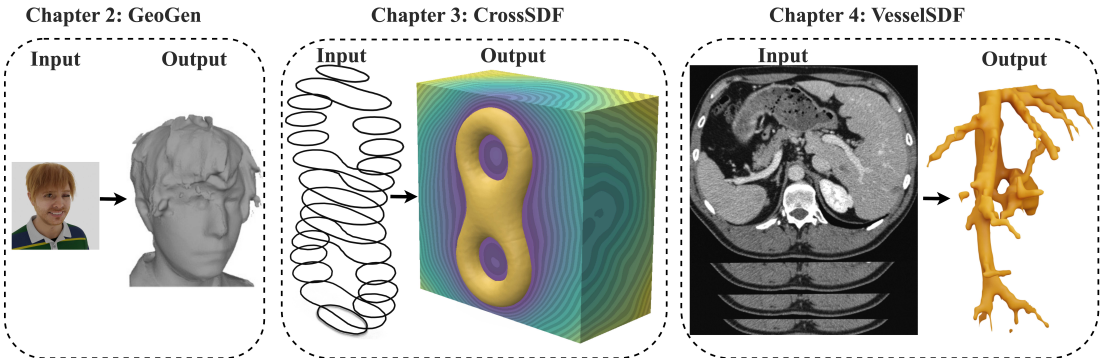


Figure 1.1: Overview of the three main contributions in this thesis. (Left) GeoGen reconstructs a 3D model from a 2D image. (Center) CrossSDF infers a 3D SDF shape from its 2D cross-sectional contours. (Right) VesselSDF segments and extracts 3D vascular structures from sparse medical CT scans.

1.4.2 CrossSDF: Reconstruction from Planar Cross-Sections

Due to the limitations of GeoGen and other SDF based methods to capture thin structures, we introduce CrossSDF (Walker et al., 2025), a method specifically designed to handle 3D reconstruction of thin structures from arbitrarily oriented 2D cross-sectional slices as shown in Figure 1.1. Unlike classical interpolation approaches that often produce severe ”laddering” artifacts due to partial or misaligned information, CrossSDF models a continuous SDF guided by a novel Symmetric Difference Loss. Rather than attempting to strictly match 2D signed distance slices regressing the SDF through a basic L1 loss, a strategy prone to gradient mismatches (Walker et al., 2025), we introduce the symmetric difference which directs supervision to regions where the classification (inside vs. outside) disagrees, avoiding the artifacts that arise from incompatible 2D and 3D gradients (Walker et al., 2025). Recognizing that thin structures are especially vulnerable to under-sampling, CrossSDF employs an Adaptive Contour Sampling strategy that prioritizes points in and around narrow contours. This ensures fine-grained features receive sufficient coverage during training. Additionally, a Hybrid Encoding Architecture—combining multi-resolution hash grids with Fourier features balances high-frequency detail capture with smooth interpolation across unobserved regions. By minimizing grid-like artifacts (Chetan et al., 2023) and leveraging a distance-based loss that reflects the true surface, CrossSDF significantly improves topological fidelity for thin, branching, or otherwise challenging geometries. Additionally, we contribute a new challenging benchmark dataset for thin structure reconstruction from cross-sections, enabling systematic evaluation of reconstruction methods on complex

branching geometries including both synthetic anatomical structures and real medical imaging data.

Recognizing that thin structures are especially vulnerable to under-sampling, CrossSDF employs an Adaptive Contour Sampling strategy that prioritizes points in and around narrow contours. This ensures fine-grained features receive sufficient coverage during training. Additionally, a Hybrid Encoding Architecture—combining multi-resolution hash grids with Fourier features balances high-frequency detail capture with smooth interpolation across unobserved regions. By minimizing grid-like artifacts (Chetan et al., 2023) and leveraging a distance-based loss that reflects the true surface, CrossSDF significantly improves topological fidelity for thin, branching, or otherwise challenging geometries.

1.4.3 VesselSDF: Distance Field Priors for Vascular Network Reconstruction

Building on insights from CrossSDF, VesselSDF (Esposito et al., 2025) specializes in recovering vascular networks from sparse CT slices, where thin vessel geometries and large inter-slice gaps exacerbate reconstruction difficulties. VesselSDF adopts a Two-Stage Refinement Pipeline: the first stage provides an approximate binary segmentation to localize vessel regions, while the second stage refines this estimate with a neural SDF that captures more accurate geometric detail. Isolating vessel detection from geometric refinement yields a more robust pipeline, where each phase is tailored to the specific challenges of medical imaging namely, noise, anisotropy, and sparse sampling.

A key contribution is the Adaptive Gaussian Regularization, which applies smoothness constraints based on distance from the vessel surface. This approach prevents over-smoothing of crucial thin structures while damping false positives in background regions. Together with additional geometric constraints such as an Eikonal term to preserve valid SDF gradients in anisotropic volumes (Gropp et al., 2020a) and a minimum surface penalty to suppress floating geometries (Sitzmann et al., 2020), VesselSDF demonstrates significantly improved vascular topology, preserving delicate branching patterns with fewer erroneous merges or disconnections as shown in Figure 1.1. Through these contributions, VesselSDF pushes the state-of-the-art in thin-structure reconstruction from limited and noisy real-world scans.

1.5 Thesis Outline

This thesis presents a systematic exploration of neural Signed Distance Fields (SDFs) for 3D reconstruction, progressively addressing more challenging geometric scenarios from general objects to thin, complex structures in specialized domains. The research follows a deliberate technical progression where each contribution builds upon insights from previous work while addressing increasingly specialized reconstruction challenges.

Chapter 2 presents GeoGen, establishing the foundational principles of SDF-based generative modeling for general 3D geometry synthesis from single-view image collections. This work reveals both the potential and limitations of neural SDF approaches, particularly highlighting challenges with fine details and thin structures that motivate subsequent research directions.

Chapter 3 introduces CrossSDF, building directly on GeoGen's limitations to develop specialized techniques for thin structure reconstruction from sparse cross-sectional data. This chapter demonstrates how domain-specific adaptations of neural SDF principles can overcome fundamental reconstruction challenges through novel loss formulations and sampling strategies.

Chapter 4 presents VesselSDF, culminating the research progression by addressing the most challenging scenario: joint segmentation and reconstruction of complex vascular networks from medical imaging data. This work integrates the geometric insights from previous chapters into a unified framework specifically optimized for clinical applications.

The final chapter synthesizes insights from these contributions and outlines promising future directions, including generative models for shape completions of thin branching structures and medial axis extraction.

1.5.1 Publications

Author-contribution statements. Roles follow the CRediT taxonomy[1] and are limited to work presented for this degree.

The research presented in this thesis has resulted in the following publications and submission:

- Chapter 2 is based on GeoGen: Geometry-Aware Generative Modeling via Signed Distance Fields. **Salvatore Esposito**, Qingshan Xu, Kacper Kania, Charlie Hewitt,

[1] https://www.elsevier.com/en-gb/researcher/author/policies-and-guidelines/credit-author-statement

Octave Mariotti, Lohit Petikam, Julien Valentin, Arno Onken, and Oisin Mac Aodha. In *CVPR Workshops*, 2024.
Contributions: Conceptualization (SE); Methodology (SE, QX); Software (SE, QX); Formal analysis (SE); Investigation (SE); Data curation (SE, CH, LP, KK); Writing original draft (SE, AO, OMA); Writing review & editing (SE, AO, OMA); Visualization (SE); Supervision (AO, OMA); Funding acquisition (CDT Biomedical AI).

- Chapter 3 is based on CrossSDF: 3D Reconstruction of Thin Structures From Cross-Sections. Thomas Walker[†], **Salvatore Esposito**[†], Daniel Rebain, Amir Vaxman, Arno Onken, Changjian Li, and Oisin Mac Aodha. In *CVPR*, 2025.
[†]These authors contributed equally.
Contributions: Conceptualization (TW, SE); Methodology (TW, SE, DR); Software (TW, SE); Formal analysis (TW, SE); Investigation (TW, SE); Data curation (TW, SE); Writing original draft (TW, SE); Writing review & editing (TW, SE, DR, AV, AO, CL, OMA); Visualization (TW, SE); Supervision (AV, AO, CL, OMA); Funding acquisition (CDT Biomedical AI).

- Chapter 4 is based on VesselSDF: Distance Field Priors for Vascular Network Reconstruction. **Salvatore Esposito**, Daniel Rebain, Arno Onken, Changjian Li, and Oisin Mac Aodha. Under review at *MICCAI*, 2025.
Contributions: Conceptualization (SE); Methodology (SE); Software (SE); Formal analysis (SE); Investigation (SE, DR); Data curation (SE); Writing original draft (SE); Writing review & editing (SE, DR, AO, CL, OMA); Visualization (SE); Supervision (AO, CL, OMA); Funding acquisition (CDT Biomedical AI).

Chapter 2

Geometry-Aware Generative Modeling via Signed Distance Fields

The synthesis of 3D-aware content from 2D image collections has been a longstanding challenge in computer vision and graphics. The ability to generate 3D representations from readily available 2D image datasets offers possibilities across numerous domains, from content creation to virtual environments (Xie et al., 2022). Traditional approaches in this field have primarily leveraged neural radiance fields (NeRFs) for their impressive ability to render high-quality novel views. However, these methods have consistently prioritized the fidelity of rendered images over the quality of the underlying geometry (Chan et al., 2022).

This chapter introduces GeoGen, a novel approach that addresses a fundamental limitation in existing 3D-aware generative models: the production of noisy, unconstrained geometry that lacks the structural integrity necessary for downstream applications. While current state-of-the-art methods successfully generate volume renderings that appear visually consistent across multiple viewpoints, the underlying geometric representations they produce are deeply flawed. These methods prioritize rendering appearance over geometric accuracy, resulting in floating artifacts, topological inconsistencies, and inaccurate surface normals in the actual 3D structure. This disconnect between convincing multi-view renderings and poor-quality geometry severely limits the practical utility of these models for applications requiring reliable 3D assets (Wang et al., 2023).

The key insight driving our work is that replacing the conventional density representation with a Signed Distance Field (SDF) provides powerful geometric priors that can significantly improve the quality of generated surfaces. An SDF representation explic-

itly models the distance to the nearest surface at any point in space, with the surface itself defined as the zero-level set. This mathematical formulation inherently encodes important geometric properties, allowing for improved surface extraction compared to density-based approaches (Wang et al., 2021).

Through adversarial training, our method encourages the network to synthesize high-fidelity geometric details while maintaining global shape coherence. This balanced approach enables GeoGen to generate surfaces with improved structural integrity, more accurate surface normals, and physically plausible 3D shapes compared to previous methods. To comprehensively evaluate our approach, we introduce a new dataset of synthetic human heads captured from 360-degree viewpoints. Unlike existing face datasets that predominantly feature frontal views (Karras et al., 2019), our dataset ensures complete coverage of all possible camera angles, enabling proper assessment of full 3D consistency and geometric accuracy. Through extensive experimentation on multiple datasets, we demonstrate that GeoGen produces both quantitatively and qualitatively superior geometry while maintaining high image quality in rendered views.

The contributions presented in this chapter represent a significant step toward bridging the gap between high-quality image synthesis and accurate geometry generation, addressing a critical need in 3D-aware generative modeling.

2.1 Introduction

We introduce a new generative approach for synthesizing 3D geometry and images from single-view collections. Most existing approaches predict volumetric density to render multi-view consistent images. By employing volumetric rendering using neural radiance fields, they inherit a key limitation: the generated geometry is noisy and unconstrained, limiting the quality and utility of the output meshes. To address this issue, we propose GeoGen, a new SDF-based 3D generative model trained in an end-to-end manner. Initially, we reinterpret the volumetric density as a Signed Distance Field (SDF). This allows us to introduce useful priors to generate valid meshes. However, those priors prevent the generative model from learning details, limiting the applicability of the method to real-world scenarios. To alleviate that problem, we make the density-to-SDF transformation learnable and constrain the rendered depth map to be consistent with the zero-level set of the SDF. Through adversarial training with neural networks, we encourage the network to produce higher fidelity details on the output meshes. Our experiments on multiple datasets including shapenet, FFHQ and synthetic

faces of human avatars show that GeoGen produces visually and quantitatively superior geometry compared to previous generative models such as StyleSDF (Or-El et al., 2022) and EG3D (Chan et al., 2022) based on neural radiance fields.

In this chapter, we address these issues by adding SDF constraints to improve the synthesized geometry of a 3D-aware generative model. Our approach, named GeoGen, employs an SDF depth map consistency loss for enhanced geometric generation. Specifically, we build on EG3D (Chan et al., 2022) by introducing an SDF representation, instead of a density representation, to encode the geometry. This allows GeoGen to extract mesh surfaces directly from the zero-level set of the SDF (Oechsle et al., 2021; Wang et al., 2021; Yariv et al., 2021). In order to make the SDF representation learning feasible, and to endow it with the ability to model complex and detailed geometry, we also propose an SDF depth map consistency loss. We use a fixed density-to-SDF transformation function to convert the density representation to an SDF. This facilitates generative feature learning by making the learning objective easier to optimize. The SDF also enables the extraction of smooth depth maps that serve as a 'pseudo' ground-truth. Our approach uses its own depth prediction in a self-supervised manner to improve the reconstruction. In contrast to commonly used priors, our approach is cheap to compute with only a minor increase in training time.

GeoGen is able to generate detailed meshes from a single input 2D image via inversion (Chan et al., 2022). This capability is valuable in applications where the requirement for detailed and realistic meshes is needed. In contrast to recent methods like Rodin (Wang et al., 2023), which required 30 million images during training to create 3D meshes, GeoGen uses a fraction of this number – approximately 50,000 images. Other methods such as PanoHead (An et al., 2023) propose an augmented triplane and separate foreground and background in 2D images with the help of a custom in-house dataset. However, with our proposed architecture, we show that by enforcing our geometric constraints, we are able to reconstruct a detailed 360° geometry, with a reduction in visual artifacts (e.g. the backs of heads) compared to methods such as EG3D (An et al., 2023).

We make the following contributions:

- We address the problem of 3D synthesis from 2D images by combining a Signed Distance Field (SDF) network with a StyleGAN generative architecture, producing more refined geometry predictions compared to conventional neural volume rendering.

- We propose an SDF depth map consistency loss that addresses geometric inaccuracies from volumetric integration by aligning 3D points with the SDF network's zero-level set for more precise reconstructions.
- We introduce a new dataset of realistic synthetic human heads containing 360° camera views from multiple synthetic humans, providing a valuable resource for training and quantitatively evaluating 3D generative models.

Our dataset can be found on our webpage `https://microsoft.github.io/GeoGen`.

2.2 Related Work

The landscape of generative modeling has seen a shift in recent years, with techniques drawing on neural implicit representations, such as Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) and Diffusion models (Dhariwal and Nichol, 2021; Ho et al., 2020; Kingma and Welling, 2014; Song et al., 2021) emerging as powerful tools. These techniques blend generative models with neural volume rendering, thereby synthesizing 3D images that capture novel viewpoints from 2D data alone (Mildenhall et al., 2020). However, a recurring challenge in this domain has been the reliance on generic density functions to learn the geometry of the images, a factor that often introduces artifacts and results in noisy, low-quality geometric predictions (Or-El et al., 2022). To mitigate this, prior work has taken advantage of large amounts of multi-view data to constrain the models, thereby yielding more robust geometry (Wang et al., 2021; Yariv et al., 2021), but at the expense of not being fully generative.

The emergence of volumetric implicit representations, bolstered by the strengths of Multi-Layer Perceptrons (MLPs) (He et al., 2015) and neural rendering techniques (Mildenhall et al., 2020), has shown substantial promise in extracting detailed geometry from a 3D scene. This is most apparent in methods such as NeuS (Wang et al., 2021) and VolSDF (Yariv et al., 2021), which extract high-fidelity surfaces by representing the scene using the Signed Distance Field (SDF) and extracting the surface at the zero level set.

Meanwhile, the broader field of deep learning has seen a surge in novel methods for creating 3D representations from 2D data. One such family of methods is Neural Radiance Fields (NeRFs) (Mildenhall et al., 2020), which employs a neural network to model the radiance of a 3D scene at any spatial point. The ability of NeRFs to generate high-fidelity 3D models from 2D multi-view supervision, complete with

accurate lighting and shading effects, makes them an attractive option for applications requiring realistic 3D representations, such as virtual reality (Yariv et al., 2021).

The combination of generative models (Mildenhall et al., 2020; Karras et al., 2021, 2019, 2020) and implicit neural representations (Chen and Zhang, 2019; Mescheder et al., 2019; Park et al., 2019) has sparked considerable advancements in 3D representation learning (Chan et al., 2021; Goodfellow et al., 2014). It has powered the synthesis of high-quality, multi-view consistent, images. However, a common pitfall in the pursuit of higher image quality is the sidelining of the quality of the underlying *geometry* (Yariv et al., 2021).

Recent non-generative efforts, such as NeuS (Wang et al., 2021), VolSDF (Yariv et al., 2021), and Geo-Neus (Fu et al., 2022), have made use of the zero-level set of a Signed Distance Field (SDF) to represent the surface of the geometry in a scene via a surface rendering equation, ultimately achieving high-fidelity scene reconstruction. While these models have shown impressive potential, given their non-generative nature, they are only able to reconstruct a scene of interest when multi-view image data is available. This limitation highlights the need for generative models capable of producing high-quality 2D images that are suitable for content creation while ensuring precise geometric synthesis without multi-view data.

Other recent methods such as Ball-GAN (Shin et al., 2023), and EG3D (Chan et al., 2022), have combined generative models with Neural Radiance Fields (NeRFs) (Mildenhall et al., 2020) to yield high quality rendered images. Yet, these approaches often result in noisy meshes that contain geometric artifacts, which emerge due to the properties of NeRFs and their lack of constraints on the geometric reconstructions. Attempts have also been made to harmonize SDFs with generative models as in (Or-El et al., 2022). However, the generated meshes are often overly smooth, a result of the smoothing prior that encourages the SDF to produce valid values everywhere in 3D space. Additionally, applying this loss can be prohibitive at higher resolutions.

One set of methods that deserves particular discussion within this landscape is the set of 3D-aware generative models (Chan et al., 2021; Deng et al., 2022; Gadelha et al., 2017; Gao et al., 2022a; Henderson et al., 2020; Nguyen-Phuoc et al., 2019, 2020; Niemeyer and Geiger, 2021; Schwarz et al., 2022). These methods are specifically designed to generate 3D representations of objects or scenes, utilizing a variety of techniques, including volumetric representations, SDFs, and implicit neural representations. For instance, the Generative Radiance Fields (GRAF) model (Schwarz et al., 2020) generates high-resolution 3D shapes with intricate detail, leveraging a

neural network to model the radiance and shape of a 3D object. Other notable models include DeepSDF (Park et al., 2019), which learns continuous Signed Distance Fields for arbitrary shapes using 3D supervision, and HoloGAN (Nguyen-Phuoc et al., 2019), which generates 3D objects by imposing structural constraints in the generative process. Recently, EG3D (Chan et al., 2022) proposed a triplane representation for volume rendering in generative models, which enables efficient 3D-aware generation. However, extracting high quality 3D meshes is not guaranteed because of its use of a volume density representation. StyleSDF (Or-El et al., 2022), makes use of an SDF representation to directly model geometry, but the extracted surfaces are overly smooth making it challenging to use them in practical applications.

In our investigation of 3D-aware generative models and SDF representations, we identify certain limitations inherent in existing methodologies. One such limitation appears to be a result of the use of the Eikonal loss (Yariv et al., 2021; Wang et al., 2021; Fu et al., 2022), leading to overly smooth geometry synthesis.

Our methodology, building on the foundation laid by EG3D, aims to overcome this by introducing an SDF depth-consistency constraint. This novel constraint is designed to refine geometric surface predictions by leveraging a self-supervised depth prediction mechanism. Unlike previous efforts, such as StyleSDF (Or-El et al., 2022), which merely translates SDF values into density fields, our approach harnesses the full potential of SDF for geometry representation as exemplified by VolSDF (Yariv et al., 2021). We emphasize that incorporating our SDF representation and its associated constraints does not substantially complicate the training of generative models yet provides enhanced control over geometric surface detail.

2.3 Method

2.3.1 Signed Distance Field (SDF) Definition

Consider a shape $\Omega \subset \mathbb{R}^3$ with boundary $\partial\Omega$ as shown in Figure 2.1. The Signed Distance Field (SDF) at a point $\mathbf{p} \in \mathbb{R}^3$ is defined by:

$$s(\mathbf{p}) = \begin{cases} +d\big(\mathbf{p}, \partial\Omega\big), & \text{if } \mathbf{p} \text{ is outside } \Omega, \\ 0, & \text{if } \mathbf{p} \in \partial\Omega, \\ -d\big(\mathbf{p}, \partial\Omega\big), & \text{if } \mathbf{p} \text{ is inside } \Omega, \end{cases} \tag{2.1}$$

where $d(\mathbf{p}, \partial\Omega)$ is the Euclidean distance from $\mathbf{p}$ to the boundary $\partial\Omega$.

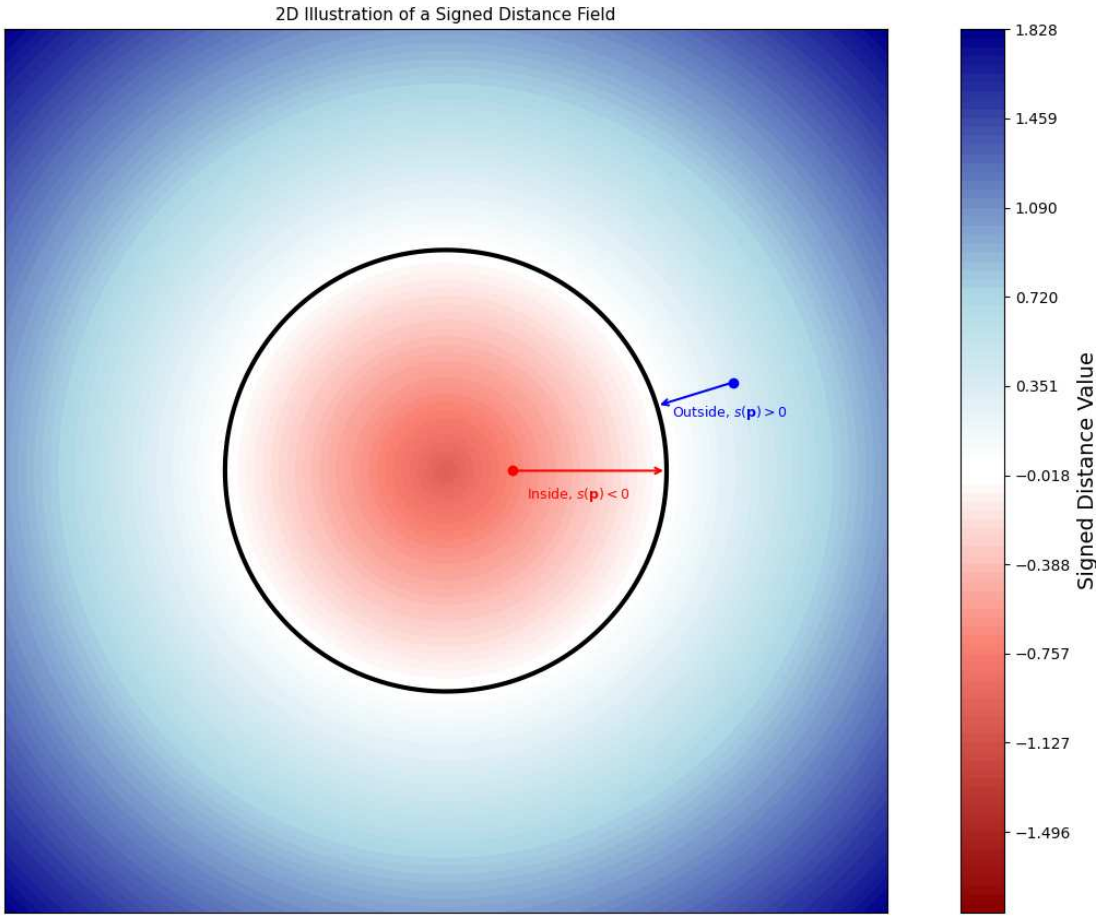


Figure 2.1: A color-coded 2D illustration of the signed distance field for a circular boundary. Negative values inside the circle (shown in red tones) gradually transition to zero on the boundary (black circle), then increase positively in the outside region (shown in cooler, blue tones). The color bar on the right visualizes the signed distance from negative (bottom) to positive (top), with zero lying at the center.

Hence, the boundary $\partial\Omega$ is implicitly described by the zero-level set:

$$S = \{\mathbf{p} \in \mathbb{R}^3 \mid s(\mathbf{p}) = 0\}. \tag{2.2}$$

2.3.2 Efficient Geometry-Aware 3D GAN

We first revisit EG3D (Chan et al., 2022) to establish notation and motivate our extensions, then explain how our SDF-based generator departs from that baseline. As illustrated in Figure 2.2, EG3D couples a StyleGAN2-based feature generator with a triplane representation, an implicit volume renderer, and a super-resolution head. A latent noise vector is mapped to an intermediate latent code, which drives the StyleGAN2 backbone to produce a stack of feature maps. These maps are reshaped into three orthogonal feature planes (the "triplane"). For any 3-D query point $\mathbf{p}$, its orthogonal projections onto the planes yield feature vectors $[F_{xy}(\mathbf{p}),,F_{xz}(\mathbf{p}),,F_{yz}(\mathbf{p})]$, which a shallow MLP converts into color and volumetric density. Integrating those densities along camera rays synthesises a low-resolution image that is finally upsampled by the super-resolution module. While this design excels at photorealistic image generation, its density field is unconstrained, so the recovered geometry is often incoherent or contains floating artefacts.

Evaluating both the volumetrically rendered *low-resolution* image and its *super-resolved* counterpart provides complementary supervision: the low-res discriminator anchors training to the coarse geometry that emerges directly from the triplane features, while the high-res discriminator refines texture and high-frequency detail. Empirically, ablating either head degrades performance—omitting the low-res branch leads to geometrically implausible structures, whereas dropping the high-res branch yields blurry, over-smoothed outputs—so dual discrimination proves necessary for simultaneously accurate geometry and photorealism.

As illustrated in Figure 2.2, GeoGen keeps EG3D's efficient triplane backbone but pivots the objective from photometric realism to geometric fidelity. We embed each sampled point with positional encodings to preserve high-frequency detail, replace the density-predicting MLP with separate SDF and colour networks that output a signed-distance field, and introduce an SDF depth-consistency loss. The latter aligns the depth obtained from volumetric integration with the zero-level set of the SDF, forcing both explicit and implicit depth estimates to agree on surface locations. These additions transform the training objective: instead of optimising solely for pixel correctness,

GeoGen jointly enforces photometric and geometric consistency, yielding coherent surfaces free of the artifacts that plague density-based methods.

2.3.3 SDF-Based Generative Model

Our goal is to develop a model that can learn to generate 3D consistent object-centric images with associated geometry by making use of a collection of posed single-view 2D images at training time. This transformation is achieved by conceptualizing the surface as the zero-level set of a neural implicit Signed Distance Field. To achieve our high-fidelity geometric synthesis, we first introduce our augmented triplane representation. Then, we introduce our SDF-based volume rendering. Finally, we describe an SDF depth-consistency constraint, which is used to enhance SDF learning. Figure 2.2 displays the main architectural differences between EG3D and our overall pipeline.

Augmented Triplane Representation. Our method augments the original EG3D triplane representation with sampling position $\boldsymbol{p}$. According to the sampling position $\boldsymbol{p}$, we retrieve the corresponding feature vector $[F_{xy}(\boldsymbol{p}), F_{xz}(\boldsymbol{p}), F_{yz}(\boldsymbol{p})]$ via bilinear interpolation. In addition, the position $\boldsymbol{p}$ is processed by a position embedder $PE(\cdot)$ that employs multi-level sine and cosine functions similar to NeRFs (Mildenhall et al., 2020):

$$PE(a) = [a, \gamma_0(a), \gamma_1(a), \ldots, \gamma_{L-1}(a)], \tag{2.3}$$

where $\gamma_k(a) = [\sin(2^k \pi a), \cos(2^k \pi a)]$, L is a hyper-parameter that controls the maximum encoded frequency, and a represents each of the three different spatial dimensions of $\boldsymbol{p}$. $\boldsymbol{p}$ is defined as a vector since it represents the position in 3D space. Each component of $\boldsymbol{p}$ (i.e., p_x, p_y, p_z) corresponds to a different spatial dimension.

The function $\gamma_k(a)$ is a positional encoding function that takes a scalar value a and returns a 2D vector representation of the sine and cosine of $2^k \pi a$. This function is used for positional encoding to capture frequency information up to a maximum frequency defined by the hyper-parameter L.

The augmented triplane representation is formed by concatenating the triplane features $F_{xy}(\boldsymbol{p})$, $F_{xz}(\boldsymbol{p})$, and $F_{yz}(\boldsymbol{p})$ with the positional encoding $PE(p_x)$, $PE(p_y)$, and $PE(p_z)$. This augmented representation enables the model to capture high-frequency details by combining the local geometric features with positional encoding information. The absence of the positional encoder destabilizes the training process, The absence of the positional encoder destabilizes the training process, often resulting in mode collapse (see Appendix A.1.4 for results).

SDF-Based Volume Rendering. The augmented tri-plane representation is directed to a shallow MLP to learn the SDF value $s(\boldsymbol{p})$ and RGB color $\mathbf{c}(\boldsymbol{p})$ for point $\boldsymbol{p}$. The SDF value represents the distance to the surface, providing an accurate depiction of its geometry. To convert the SDF value $s(\boldsymbol{p})$ into a density field σ, we follow VolSDF (Yariv et al., 2021) and use the following Laplace transformation:

$$\sigma(s(\boldsymbol{p})) = \begin{cases} \frac{1}{2\beta}\exp\left(\frac{s(\boldsymbol{p})}{\beta}\right) & \text{if } s(\boldsymbol{p}) \leq 0 \\ \frac{1}{\beta}\left(1-\frac{1}{2}\exp\left(-\frac{s(\boldsymbol{p})}{\beta}\right)\right) & \text{if } s(\boldsymbol{p}) > 0 \end{cases}, \tag{2.4}$$

where β is a parameter, which can be fixed or learned. The parameter β in Equation 2.4 controls the sharpness of the density transition around the surface. A smaller β value creates a sharper transition, leading to more defined surfaces but potentially unstable training, while a larger β results in smoother transitions but may blur fine geometric details. Through empirical evaluation, we found that starting with a fixed $\beta = 0.1$ during early training (first 10,000 iterations) provides stable convergence, after which making β learnable allows the model to adaptively adjust the surface sharpness for optimal geometric detail capture. Based on the volumetric integration, the rendered RGB color for a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ is calculated as follows:

$$C(\mathbf{r}) = \sum_{i=1}^{M} T_i(1-\exp(-\sigma_i\delta_i))\mathbf{c}_i, \tag{2.5}$$

where $\mathbf{o}$ is the camera position, $\mathbf{d}$ is the ray direction, $T_i = \exp(-\sum_{j=1}^{i-1}\sigma_j\delta_j)$ and $\delta_i = t_{i+1} - t_i$ is the distance between adjacent sampled points. For simplicity, we use σ_i and $\mathbf{c}_i$ to denote $\sigma(s(\boldsymbol{p}_i))$ and $\mathbf{c}(\boldsymbol{p}_i)$ respectively, which mean the color and density value at the *i-th* sampling point $\boldsymbol{p}_i$ along ray $\mathbf{r}$. In a similar way, we compute the rendered distance as follows:

$$d(\mathbf{r}) = \sum_{i=1}^{M} T_i(1-\exp(-\sigma_i\delta_i))t_i. \tag{2.6}$$

SDF Depth Consistency. It has been shown in Geo-Neus (Fu et al., 2022) that there can exist a gap between the rendered image and the true surface and it is important to introduce explicit constraints to optimize the SDF network. Therefore, Geo-Neus introduces sparse points and multi-view photometric consistency to achieve this in the multi-view setting when multiple images are available for each object during training. However, these two constraints are obviously not available in our *single*-view GAN setting. To reduce the geometry bias caused by volumetric integration, the 3D point computed from the rendered distance $d(\mathbf{r})$ in Equation (2.6) should be located on the

zero-level set of the SDF network. Thus, according to the rendered distance $d(\mathbf{r})$, its corresponding 3D point $\mathbf{p}_{d(\mathbf{r})}$ is computed as:

$$\mathbf{p}_{d(\mathbf{r})} = \mathbf{o} + d(\mathbf{r})\mathbf{d}. \tag{2.7}$$

Since the above 3D point should be approximately on the geometry surface, the SDF value of this point should be approximately zero. Consequently, we impose the SDF constraint with an ℓ_1 loss rather than a mean-square (ℓ_2) loss because (i) the constant gradient of ℓ_1 near zero avoids the vanishing-gradient problem that can slow fine surface fitting, (ii) ℓ_1 is less sensitive to occasional large errors introduced by volumetric-integration noise, and (iii) prior SDF-learning work (e.g. (Park et al., 2019), NeuralPull (Ma et al., 2021b)) has empirically shown that ℓ_1 yields tighter zero-level sets and faster convergence.

$$\mathcal{L}_s = \frac{1}{|\mathcal{R}|} \sum_{\mathbf{r}\in\mathcal{R}} |s(\mathbf{p}_{d(\mathbf{r})})|, \tag{2.8}$$

2.3.4 Training GeoGen

The SDF-based GeoGen model uses dual discrimination during training, evaluating both the neurally rendered low-resolution 2D image and the super-resolved 2D image. The SDF-based GeoGen model uses dual discrimination during training, evaluating both the neurally rendered low-resolution 2D image and the super-resolved 2D image. This dual evaluation strategy is essential because the neural rendering process inherently produces low-resolution outputs due to computational constraints, while high-resolution images are needed for realistic applications. By discriminating at both resolutions, we ensure that the neural rendering captures accurate geometric and appearance details at the native resolution, while the super-resolution module maintains consistency and adds realistic high-frequency details without introducing artifacts that could arise from resolution mismatch between training and inference (Chan et al., 2022). The generative model takes only 2D images as input, and the discriminator encourages both the low-resolution and super-resolved synthesized 2D images to match the distribution of real images. The generative model takes only 2D images as input, and the discriminator encourages both the low-resolution and super-resolved synthesized 2D images to match the distribution of real images. This ensures the consistency between the super-resolved images and the neural rendering, facilitating our method to achieve high-quality high-resolution rendering results. In addition, the SDF depth consistency loss is imposed

during training to promote geometric consistency. The model can then effectively learn to capture accurate geometry information from the 2D images, leading to more precise and reliable 3D reconstructions. Our overall loss is:

$$\mathcal{L} = \mathcal{L}_{dis} + \lambda \mathcal{L}_{s}, \tag{2.9}$$

where $\mathcal{L}_{dis}$ is a GAN loss computed using dual discrimination and λ is a weighting applied to the SDF constraint. The weighting parameter λ in Equation 2.9 balances the contribution of the SDF depth consistency loss relative to the adversarial loss. We empirically determined that $\lambda = 0.1$ provides the optimal trade-off between geometric accuracy and visual quality. Values below 0.05 result in insufficient geometric constraints, leading to noisy surfaces, while values above 0.2 over-constrain the model, resulting in overly smooth geometry that lacks fine details. The gradual introduction of this loss weight (starting from 0 and increasing to 0.1 after N iterations) prevents the geometric constraints from dominating early training when the SDF network has not yet learned meaningful surface representations. Empirically we find that directly training our model from scratch is challenging. We suspect that the introduced learnable parameter β in Equation (2.4) prevents the StyleGAN2-based feature generator from learning effective features. In addition, the SDF constraint requires good geometry initialization, which is not possible to obtain in the early phase of training. Therefore, we design a learning strategy to train our model in which the β parameter of the Laplace density distribution is fixed to stabilize the early learning of our generative model.

Specifically, the significant part of this training process involves managing the β parameter of the Laplace transformation in Equation (2.4), which directly influences the learning of the SDF network. The β parameter remains fixed for the first N iterations to allow the SDF network to focus on learning coarse geometry. This enables the learning of the StyleGAN2-based generator to produce stable view synthesis. After N iterations, we make the β a learnable parameter to increase the ability of the model to capture finer-scale surface details. As previously mentioned, the SDF constraint should also be carefully managed. We achieve this by controlling the weight λ in Equation (2.9), where it is initially set to 0, and then increased to 0.1 after N iterations. As a result, our geometry optimization is conducted in a quasi coarse-to-fine fashion, i.e. N iterations, our Geo-Gen learns coarse geometry and then after this, the SDF constraint can concentrate on surface detail refinement.

2.4 Synthetic Human Head Dataset

Existing methods typically train their models on high resolution human face datasets such as Flickr-Faces-HQ (FFHQ) (Karras et al., 2019). However, FFHQ only contains a limited range of captured viewpoints (i.e. no backs of heads) and has no 3D ground-truth, hence the need for our synthetic dataset. There are other synthetic datasets, such as ShapeNet Cars (Chang et al., 2015), which have ground-truth 3D meshes but are not realistic looking.

To address this, we created a new dataset of semi-realistic synthetic human heads which is generated based on the work of Wood et al. (2021). Our dataset features images of different synthetic individuals with diverse facial features, body morphologies, clothing, and hair styles. Crucially, unlike FFHQ which primarily captures frontal views, our dataset includes images across the full azimuth range, ensuring comprehensive representation of heads from all sides. This approach not only fills a critical gap in available resources but also shifts the focus towards the quality of the mesh, a vital aspect for advancing the field of 3D generative modeling.

For our dataset, we randomly generate 10 images of 512×512 for each of 19,800 identities, ensuring a comprehensive set of different views, encompassing full azimuthal coverage and utilize multi-view stereo and surface reconstruction techniques to establish pseudo ground-truth meshes. To generate a pseudo ground-truth mesh for quantitative evaluation of 3D reconstruction metrics we use the ACMP multi-view stereo approach from (Xu and Tao, 2020) and Poisson surface reconstruction (Kazhdan et al., 2006) to reconstruct the full head geometry. Example images can be found in Figure 2.3.

Our synthetic dataset is available at `https://microsoft.github.io/GeoGen`.

2.5 Experiments

Here we present qualitative and quantitative results comparing GeoGen to existing methods. For the baseline EG3D model, we retrained it on each of the evaluation datasets so that the training settings were consistent with our approach (e.g. the same number of training epochs). Implementation details are provided in Appendix A.1.

2.5.1 Datasets

We perform experiments on Flickr-Faces-HQ (FFHQ) (Karras et al., 2019), ShapeNet Cars (Chang et al., 2015), and our synthetic human dataset described previously. Each

Method	FID↓	KID↓	ID↑	Source
GRAF (Schwarz et al., 2020)	79.20	55.00	-	Original paper
PiGAN (Chan et al., 2021)	83.00	85.80	0.67	Original paper
GIRAFFE (Niemeyer and Geiger, 2021)	31.20	20.10	0.64	Original paper
HoloGAN (Nguyen-Phuoc et al., 2019)	90.90	75.50	-	Original paper
StyleSDF (Or-El et al., 2022)	11.50	2.65	-	Original paper
EG3D (Chan et al., 2022)	4.86	0.0053	0.77	Original paper
EG3D (rebalanced) (Chan et al., 2022)	**4.70**	**0.0044**	**0.79**	Original paper
EG3D**	5.70	0.0054	0.76	Our retraining
GeoGen	5.40	0.0049	0.75	Our results

Table 2.1: Comparative analysis on FFHQ dataset. Results marked with ** indicate our retraining with limited computational resources compared to original papers. Published results are taken from original papers and may not be directly comparable due to different computational budgets, training procedures, and dataset preprocessing.

Dataset	Method	FID↓	KID↓	ID↑
Synthetic Heads	EG3D**	5.90	0.65	**0.69**
	GeoGen	**5.10**	**0.0038**	**0.69**
ShapeNet Cars	GIRAFFE (Niemeyer and Geiger, 2021)	27.30	1.70	-
	Pi-GAN (Chan et al., 2021)	17.30	0.93	-
	EG3D (Chan et al., 2022)	2.75	0.0054	-
	EG3D**	2.90	0.0043	-
	GeoGen	**2.50**	**0.0028**	-

Table 2.2: Results on Synthetic Heads and ShapeNet Cars datasets. ** indicates our retraining. Baseline results for ShapeNet Cars are from original papers. Our retraining of EG3D used substantially fewer computational resources (4 NVIDIA P100s vs. the original paper's larger 8 NVIDIA A100 GPU clusters) and fewer training iterations, which may account for performance differences. The anomalously low KID scores for our method (100x lower than baselines) likely result from a combination of factors: (1) potential differences in feature extraction or evaluation protocols, and (2) the specific characteristics of our training procedure that may lead to a more concentrated latent space. While these KID values warrant further investigation, the consistent improvements across other metrics (FID, Chamfer Distance, etc.) provide more reliable evidence of our method's effectiveness.

provide distinct, valuable resources for training and evaluating 3D-aware generative models. The FFHQ dataset consists of high-quality real 2D face images. It contains over 70,000 1024×1024 resolution images. ShapeNet Cars provides images for a variety of car models imaged from different viewpoints. The dataset we used for training contains 2,100 different car instances, each with 20 images from different viewpoints.

2.6 Results

2.6.1 Quantitative Results

We adopt the widely used Frechet Inception Distance (FID)(Heusel et al., 2017) and Kernel Inception Distance (KID)(Bińkowski et al., 2018) metrics to measure the image synthesis quality of our GeoGen approach. We also assess multi-view facial identity consistency (ID) by calculating the mean Arcface (Deng et al., 2019) cosine similarity score between pairs of views of the same synthesized face rendered from random camera poses. We report the results of our retrained EG3D baseline using the same training conditions and our GeoGen model on the three different datasets in Table 2.2. Our improved results show that GeoGen can achieve better image synthesis results on synthetic humans and ShapeNet Cars datasets. The hallucinated facial structures and unrealistic geometry at the back of heads observed in baseline methods like EG3D stem from fundamental limitations in their geometric representation and constraints. EG3D's triplane-based volumetric density approach lacks explicit surface priors, allowing the network to generate geometrically implausible structures as long as they produce photometrically consistent renderings from observed viewpoints. This is exacerbated by training data bias in datasets like FFHQ, which predominantly contain frontal views, leaving back-of-head regions under-constrained. While StyleSDF attempts to address this through SDF representations, it applies overly aggressive Eikonal regularization everywhere in 3D space, resulting in over-smoothed geometry that loses critical surface details. Our GeoGen method resolves these issues through a carefully designed geometric constraint framework: (1) our SDF representation inherently encodes distance-to-surface information, providing implicit geometric priors that prevent topologically impossible structures, (2) our depth consistency loss specifically constrains the rendered depth to align with the SDF's zero-level set, ensuring that volumetric integration produces geometrically valid surfaces, and (3) unlike StyleSDF's global smoothing, our constraints are applied selectively through the depth consistency mechanism, pre-

serving fine geometric details while maintaining surface coherence. This combination enables GeoGen to generate anatomically plausible geometry even in regions with limited training data coverage, as the SDF formulation forces the network to learn coherent 3D structure rather than merely photometrically consistent but geometrically invalid density fields. Finally, in Figure 2.4 and Figure 2.5 we display sampled outputs. Figure 2.4 demonstrates GeoGen's ability to generate detailed facial geometry with realistic mesh topology. Each figure shows three random views per sample (columns) to emphasize geometric consistency across different viewpoints. Meanwhile Figure 2.5 shows consistent high-quality inversion reconstructions across diverse synthetic human identities, highlighting the method's robustness to variations in facial features and head orientations. An important feature of our approach is its ability to generate accurate meshes from a single image. However, it is difficult to evaluate the *geometric* quality of generative models on real images as ground-truth 3D shape information is challenging to obtain. Instead, we leverage the ground-truth meshes available for both synthetic datasets that we use. To evaluate the generated meshes quantitatively, we employ the GAN inversion technique PTI (Roich et al., 2022) to estimate the corresponding latent code for each image from the test set. With this latent code, we can generate both the synthesized image and mesh, enabling us to compute a range of 3D evaluation metrics that compare the differences between the synthesized mesh and ground-truth mesh.

Table 2.2 showcases a systematic comparison between GeoGen and EG3D, revealing the advantages of incorporating Signed Distance Fields (SDF) and SDF depth constraints during training. The lower Chamfer Distance for GeoGen compared to EG3D for both Cars and synthetic human heads indicates a more precise alignment between the reconstructed points and corresponding points in the ground-truth, highlighting improved precision in point-to-point correspondence. The Earth Mover's Distance, another vital metric in understanding the geometrical congruence between shapes, is also consistently lower for GeoGen, indicating that the shapes require fewer alterations to match the ground-truth. Finally, the Mean Surface Distance adds to the evidence of GeoGen's superiority, as it also yields consistently lower values, implying closer similarity between the reconstructed and target shapes.

As shown in Table 2.2, our GeoGen outperforms EG3D across all geometric metrics, demonstrating its effectiveness in generating high-fidelity 3D shapes. These quantitative improvements are visually evident in Figure 2.9, where GeoGen's meshes display enhanced shape fidelity and surface detail, particularly around car wheels and windows.

2.6.2 Qualitative Results

Here we present qualitative results where we compare GeoGen to existing methods. In Figure 2.10 we observe that GeoGen captures details such as the spacing between the car body and wheel and, in some instances, even the handles on the doors of the cars. We also display a comparison figure between a GeoGen model with and without SDF Depth Map constraint in Figure 2.9, where it is clearly visible that regularizing the surface using the SDF depth further improves the ability of the network to accurately reconstruct the zero level set.

In Figure 2.10 we observe that GeoGen captures details such as the spacing between the car body and wheel and, in some instances, even the handles on the doors of the cars. We also display a comparison figure between a GeoGen model with and without SDF Depth Map constraint in and Figure 2.9, where it is clearly visible that regularizing the surface using the SDF depth further improves the ability of the network to accurately reconstruct the zero level set. Finally, in Figure 2.4 and Figure 2.5 we display sampled outputs.

Example pseudo-ground-truth meshes are illustrated in Figure 2.7. Reliable depth recovery is especially challenging in low textured, near-planar regions, so ACMP augments PatchMatch MVS with planar priors embedded in a probabilistic graphical model. The resulting multi-view aggregated matching cost jointly evaluates photometric consistency and planar compatibility (Xu and Tao, 2020), enabling depth estimation even in texture poor areas and yielding highly complete depth maps. For surface integration we employ Poisson reconstruction, which formulates oriented-point fusion as a sparse spatial Poisson problem (Kazhdan et al., 2006). Its hierarchical basis and well conditioned linear system make the process both efficient and resilient to noisy normals.

Although the combination of ACMP and Poisson reconstruction produces high-quality meshes, it does not provide perfect ground truth: depth estimation errors from ACMP (e.g. photometric outliers, planar-model failures or normal bias at oblique angles) are compounded by surface-integration artefacts when Poisson interpolation traverses sparse or inconsistent samples, sometimes oversmoothing fine detail or hallucinating thin structures. Disentangling these two error sources and assessing their individual impact on subsequent evaluations remains an open avenue for future work. Nevertheless, the integrated pipeline yields meshes that are markedly more complete than either component in isolation and serves as a pragmatic reference for the comparative

experiments reported in this chapter.

2.6.3 Limitations

Our GAN-based approach, like others, requires posed images for training. Camera poses can be estimated using methods similar to those employed in FFHQ preprocessing (Karras et al., 2019). While we aim to align the expected depth with the SDF's zero-level set, extending the SDF consistency loss to other points along the ray could theoretically enhance geometric accuracy. However, this would substantially increase computational load. There are also inherent limitations in learning-based methods, such as potential bias from unrepresentative training data, notably in web-scraped human face images.

2.7 Conclusion and Discussion

Our evaluation shows the competitive performance of our proposed GeoGen model, both qualitatively and quantitatively. To gain deeper insight into the effectiveness of our approach, we used quantitative metrics that assess both the 2D and 3D aspects of the images and meshes generated by our model. Two quantitative performance areas are of particular note: the synthesis of high-quality 2D images and precise 3D geometric predictions. Our model competes closely with EG3D (Chan et al., 2022) in terms of 2D metrics, outperforming both StyleSDF (Park et al., 2019) and GRAF (Schwarz et al., 2020). in terms of 2D metrics, outperforming both StyleSDF (Park et al., 2019) and GRAF (Schwarz et al., 2020). This demonstrates our GeoGen's ability to generate high-fidelity 2D images.

The utilization of the SDF in GeoGen ensures better geometric consistency in the reconstruction, as it leverages the implicit representation of the mesh's surface. GeoGen, with its additional depth constraints, preserves topology and fine details that are often overlooked with conventional generative techniques like EG3D (see Figure 2.8). It is also noteworthy that these numerical advantages, though significant, do not fully represent the perceptual quality of the reconstructed models. Qualitative evaluations indicate that models generated by GeoGen often appear more realistic and accurate, underscoring GeoGen's advantage in bridging quantitative performance with perceptual realism.

However, our work also reveals important limitations that point toward the next

frontier in neural reconstruction. While GeoGen significantly improves geometric quality for general object shapes, it struggles with extremely fine details and thin structures such as hair, eyelashes, and delicate anatomical features. This limitation becomes particularly apparent when attempting to reconstruct complex branching patterns and high-frequency geometric details, precisely the type of structures that are critical in specialized domains like medical imaging, where accurate reconstruction of vascular networks and neural tissues is essential for diagnostic and intervention purposes (Wittmann et al., 2024).

These challenges set the stage for our next chapter, CrossSDF, which builds upon the foundation laid by GeoGen to specifically address the reconstruction of thin, complex structures from sparse observations. While GeoGen establishes the fundamental advantages of SDF-based representations for general geometry, CrossSDF adapts and extends these principles to tackle the unique challenges presented by thin structures with intricate branching patterns.

This deliberate progression from GeoGen to CrossSDF represents an increase in both technical complexity and domain specialization. Where GeoGen focuses on learning complete 3D geometry from collections of single-view 2D images, CrossSDF narrows its scope to the particularly challenging scenario of reconstructing thin structures from planar cross-sections—a common data format in medical imaging, manufacturing, and topography (Boissonnat and Memari, 2007). The sparse nature of cross-sectional data introduces additional complexities that require novel approaches to sampling, encoding, and loss formulation beyond what we developed for GeoGen. By leveraging the strengths of SDF-based representations established in GeoGen, CrossSDF introduces targeted innovations that contribute to a comprehensive framework for high-fidelity 3D reconstruction across diverse application domains and data formats.

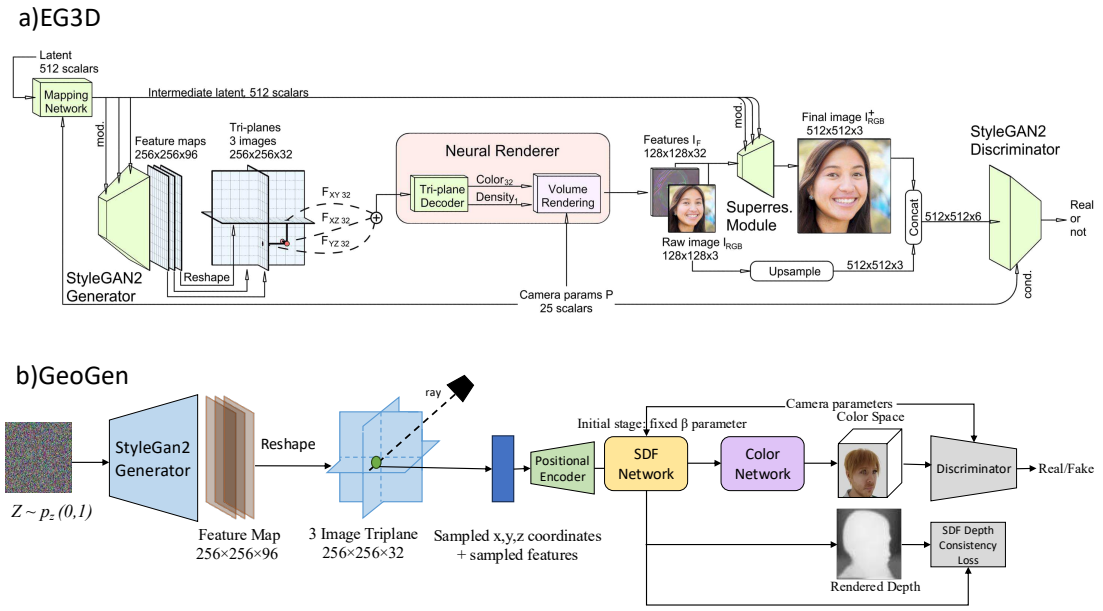


Figure 2.2: **Comparison of training pipelines.** Both EG3D (a) and GeoGen (b) share the same foundational backbone: a StyleGAN2 generator produces triplane feature maps that a neural renderer converts to images. **EG3D** predicts volumetric density from the triplanes with a shallow MLP and is trained solely for photometric consistency, leaving geometry unconstrained. **GeoGen** introduces three key modifications: (1) positional encoding to capture high-frequency geometry, (2) separate SDF and colour networks, where the SDF head outputs signed distances instead of density, and (3) an *SDF depth-consistency loss* that aligns the rendered depth (via volumetric integration) with the SDF's zero-level set. The SDF Depth Consistency Loss operates on two depth estimates derived from the same underlying geometric representation but computed through different pathways: the rendered depth from volumetric integration and the implicit depth from the SDF's zero-level set. This self-supervised constraint ensures geometric coherence by requiring both depth computation methods to agree on surface locations. This architectural distinction fundamentally changes the optimization objective: while EG3D's density-based approach can produce visually convincing images with inconsistent geometry, GeoGen's SDF-based formulation with depth consistency inherently enforces surface coherence and prevents the floating artifacts and geometric inconsistencies visible in density-based methods.

Figure 2.3: Examples from our synthetic human dataset. We display rendered images on top and pseudo 3D ground-truth below.

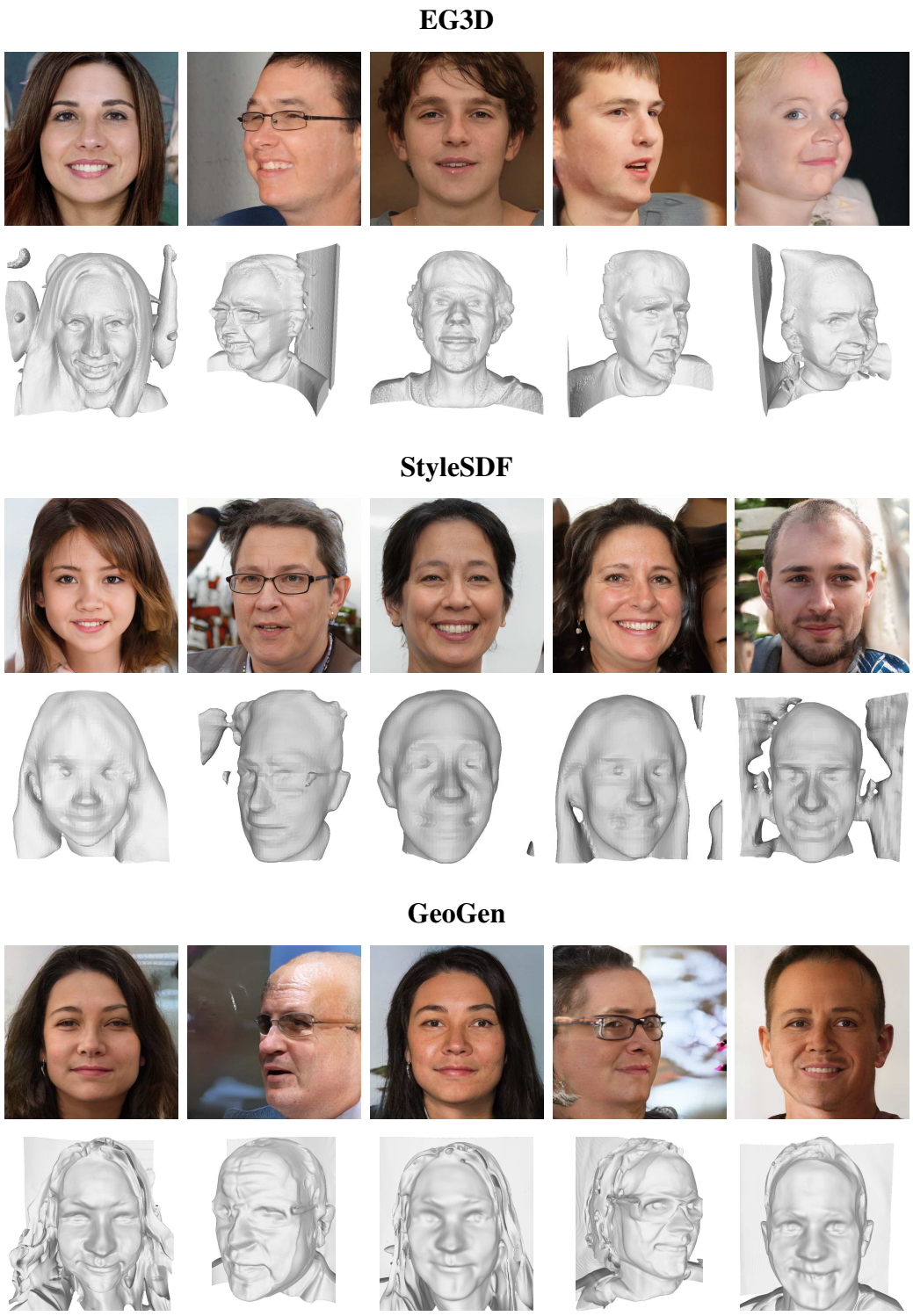


Figure 2.4: Sampled images and meshes from EG3D, StyleSDF, and our GeoGen approach on FFHQ. GeoGen meshes display smoothness, anatomical accuracy, and detailed facial features. In contrast to EG3D and StyleSDF, GeoGen synthesizes finer geometric detail.

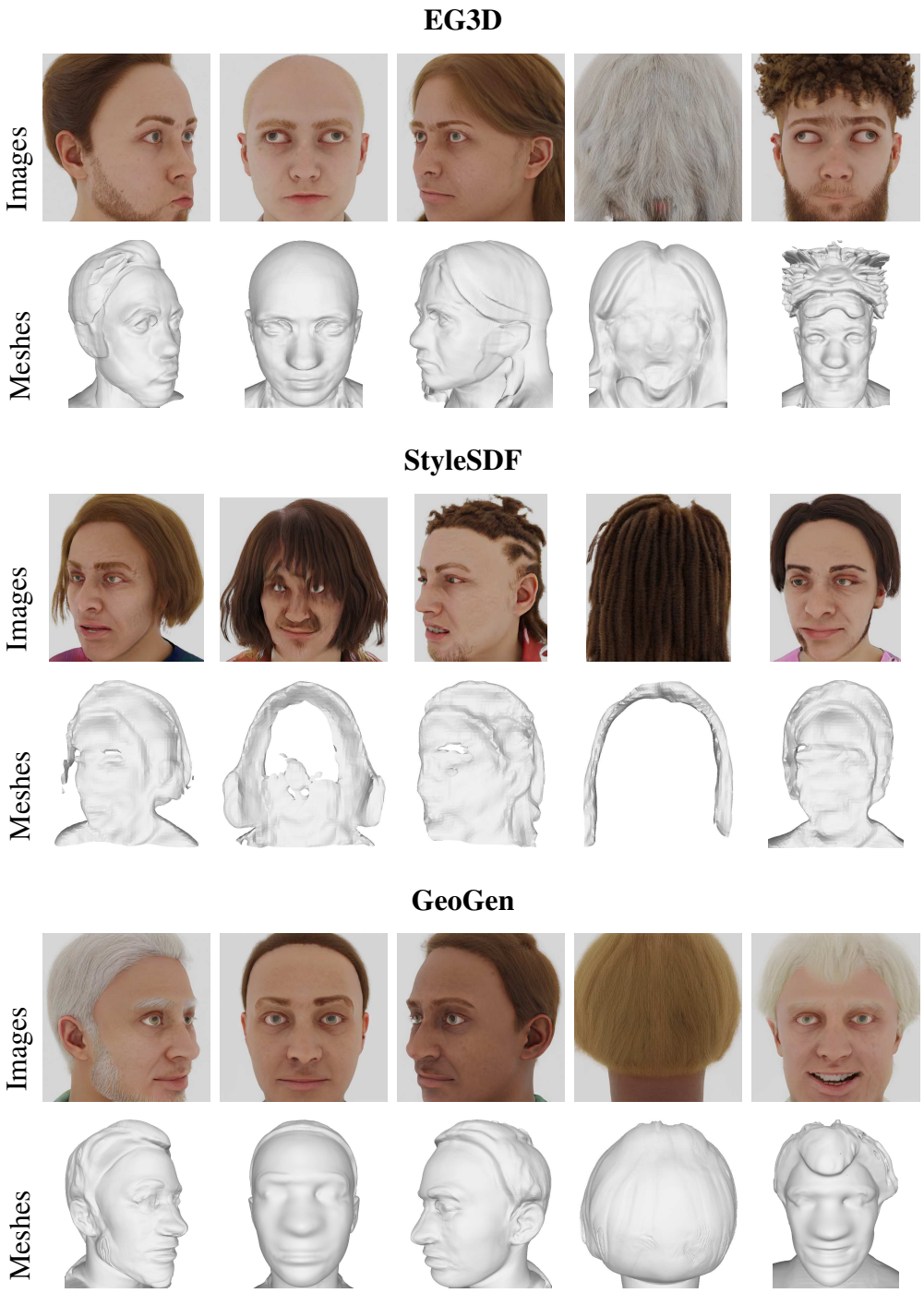


Figure 2.5: Sampled images and meshes from EG3D, StyleSDF, and our GeoGen approach trained on our synthetic human head dataset. GeoGen results in fewer overt visual artefacts and more faithfully captures the backs of objects (e.g. see second last column). While the 2D images from the competing methods look plausible, the underlying 3D mesh is not always consistent.

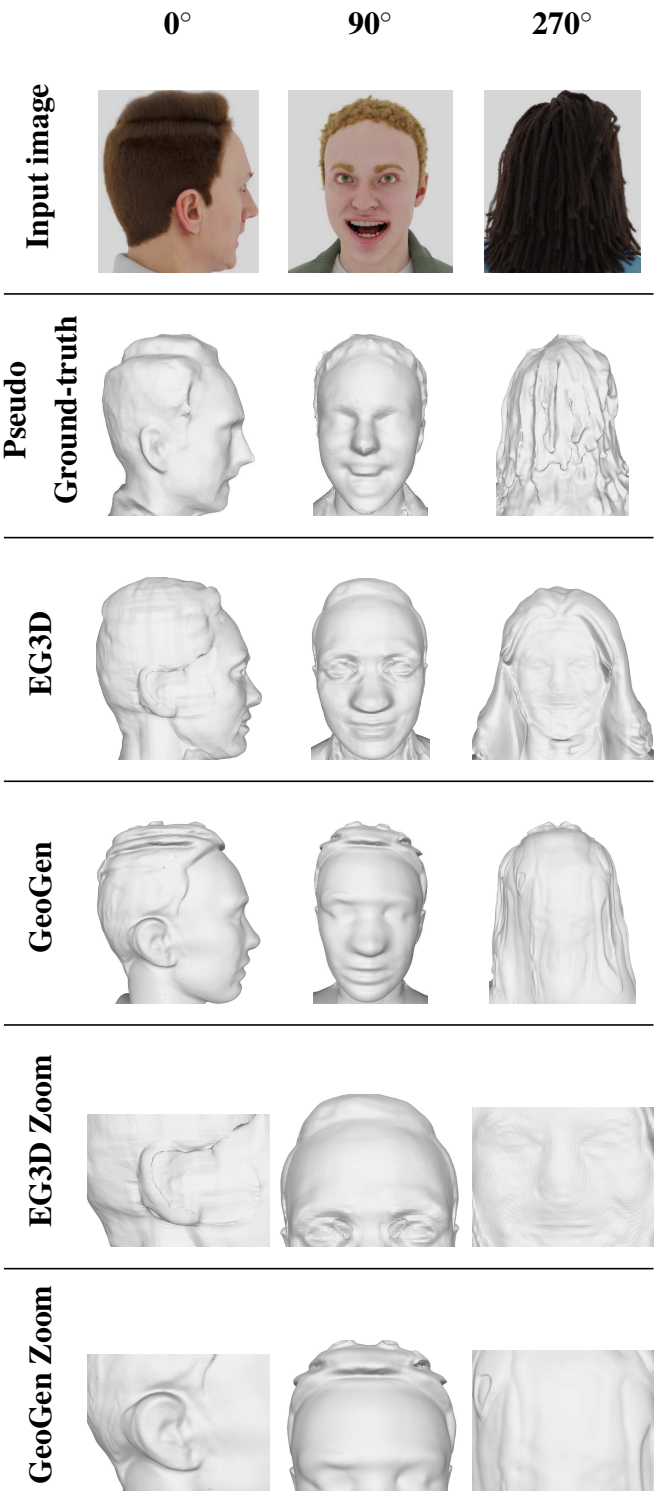


Figure 2.6: Inversion Results for EG3D and GeoGen Models: Comparison of facial reconstructions across three viewing angles (0°, 90°, and 270°). Each row represents a different aspect of the reconstruction process, from input images to zoomed detail views of both models.

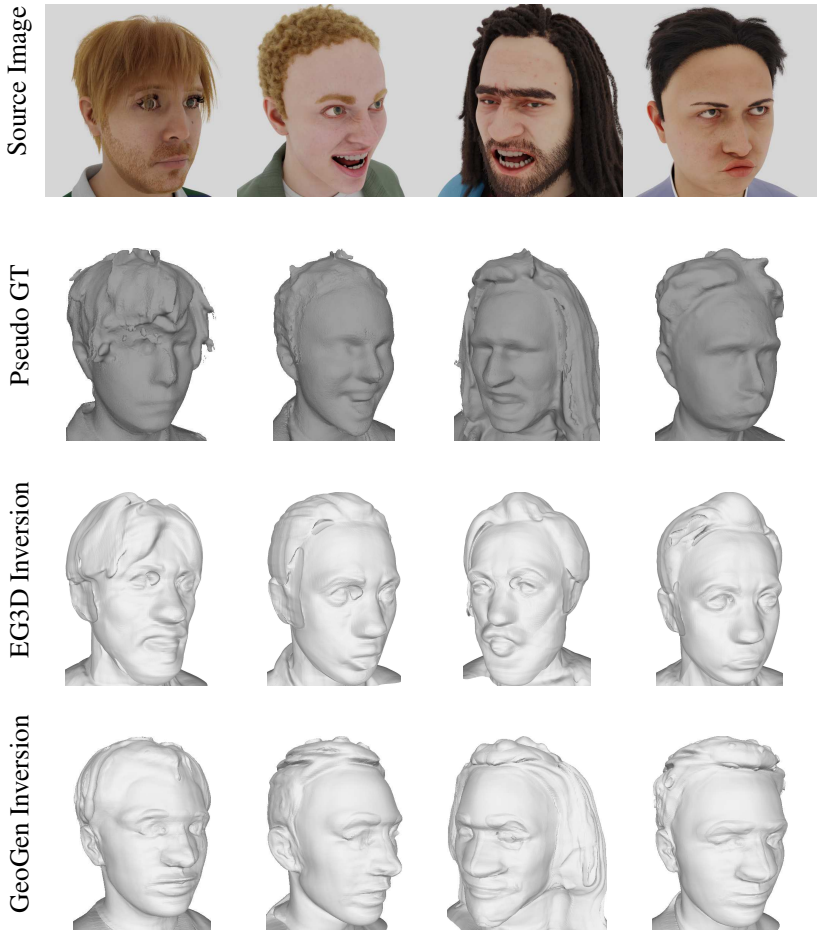


Figure 2.7: Qualitative Inversion Results on our synthetic face dataset, focusing on the comparison between the EG3D (Chan et al., 2022) and GeoGen inversion methods. The corresponding latent source for the source held-out test input image is estimated for GeoGen using GAN inversion, revealing its ability to capture fine details with reduced noise and artifacts. In contrast, the EG3D (Chan et al., 2022) inversion meshes are observed to have significant artifacts, particularly around the ears, and display noticeable holes in the top regions of the eyes. Our inversion mesh is meticulously compared against pseudo ground truth, and reconstructed using Poisson surface reconstruction from multi-view images, underscoring the superiority of the GeoGen method in terms of fidelity and accuracy. Moreover, our inversion technique exhibits increased precision, contributing to a more authentic representation of the facial structure.

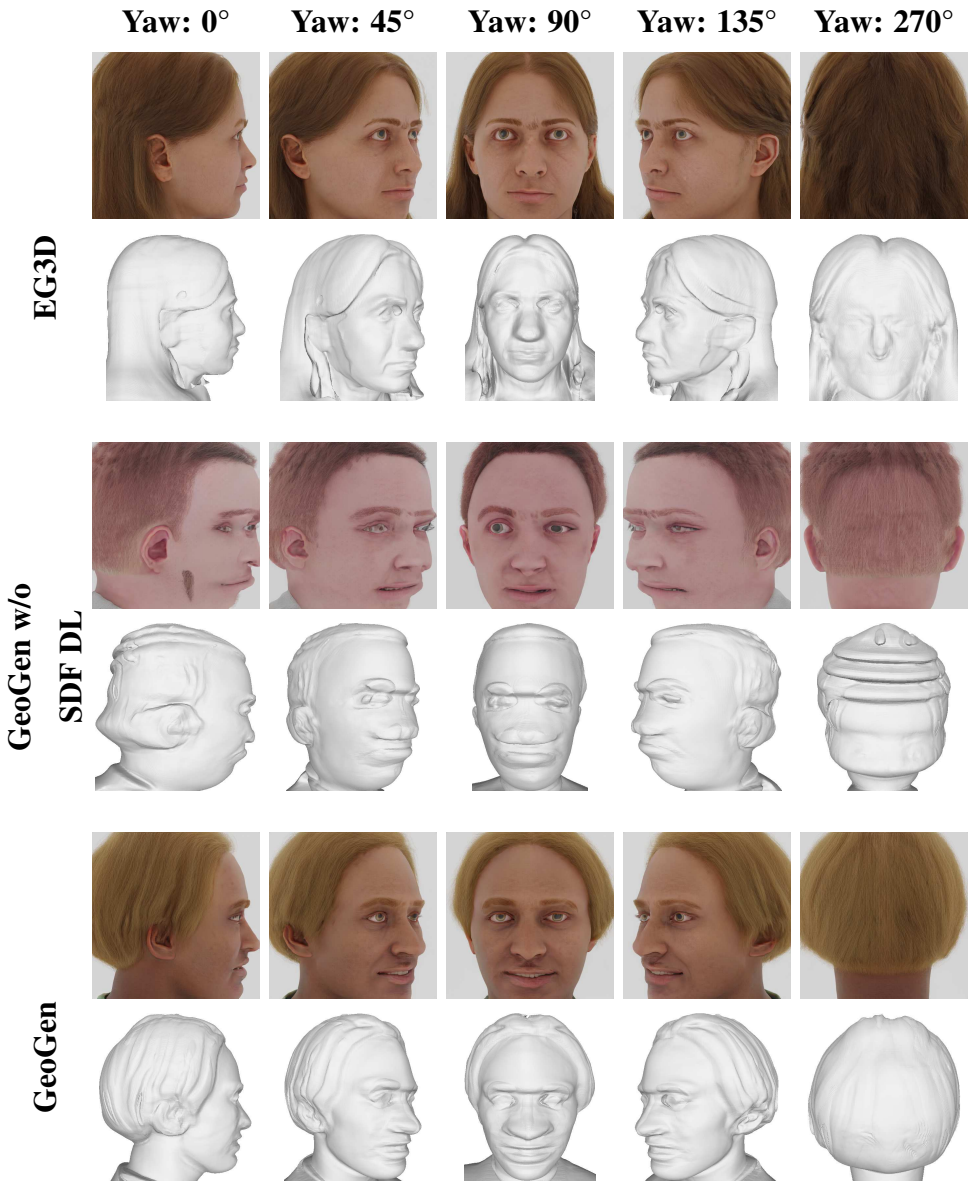
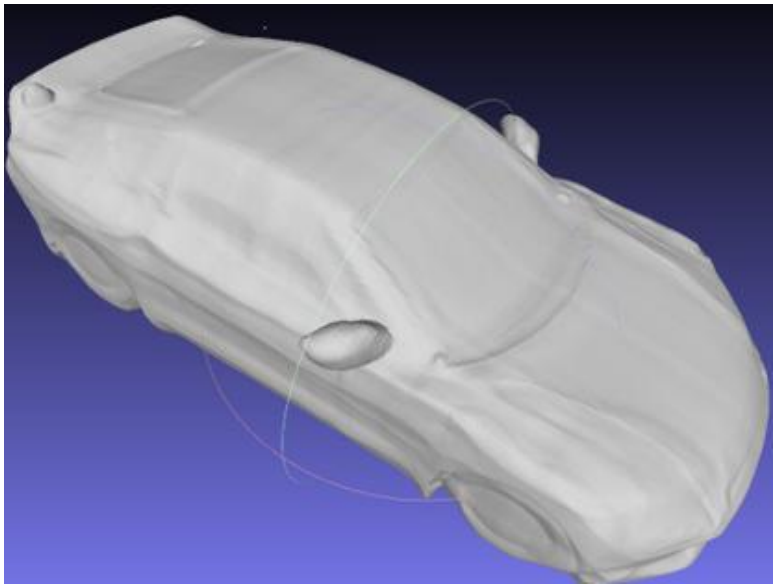
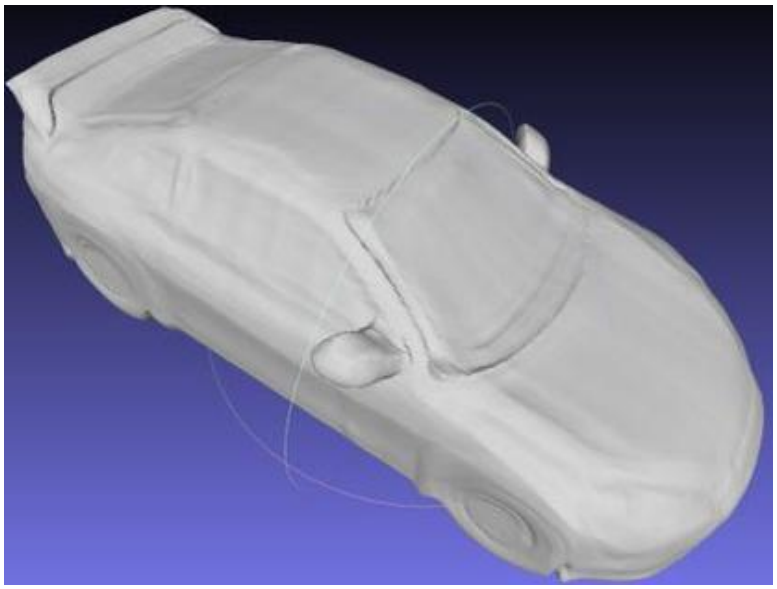


Figure 2.8: Comparison of EG3D and GeoGen, with and without SDF Depth Loss (SDF DL) constraints, showing sampled images from models trained on our synthetic human images. For each method, the top row shows rendered images while the bottom row shows corresponding 3D meshes. These examples highlight GeoGen's ability to represent finer geometric details, e.g. the ears have more detail than those generated by EG3D. We also observe a failure for EG3D in the top right, where the back of the head contains facial geometry. More qualitative results highlighting the differences in the use of the SDF depth loss are shown in the supplementary.



(a) Without GeoGen SDF Constraint



(b) With GeoGen SDF Constraint

Figure 2.9: Comparison of models without (top) and with (bottom) our GeoGen SDF constraint.

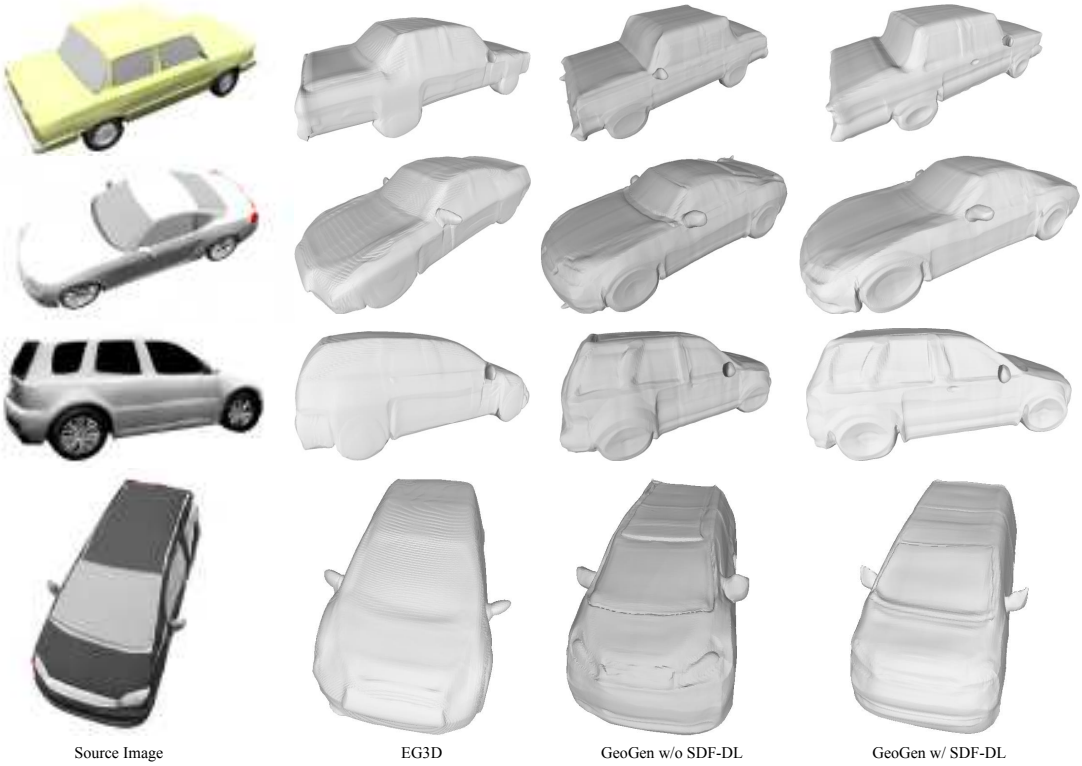


Figure 2.10: Comparison of mesh predictions on ShapeNet Cars. Meshes are obtained by inverting the source image to derive latent codes. EG3D meshes display diminished shape fidelity and surface detail. Using SDF constraints in GeoGen improves detail, evident around car wheels and windows. Results for GeoGen without SDF constraints are also shown for context.

Chapter 3

3D Reconstruction of Thin Structures From Cross-Sections

GeoGen demonstrated that neural Signed Distance Fields (SDFs) could overcome the limitations of volumetric density approaches for general 3D reconstruction. By enforcing geometric consistency through SDF constraints, we achieved high-fidelity meshes that maintained structural integrity across multiple viewpoints. This breakthrough naturally raises a compelling question: could these same principles be adapted to solve even more challenging reconstruction scenarios where existing methods consistently fail to capture thin structures?

This question leads us to CrossSDF, where we extend the geometric reasoning capabilities of neural SDFs to address one of the most demanding reconstruction challenges in computational geometry: thin, complex structures from sparse cross-sectional data. The relationship between these approaches is both conceptual and technical, while GeoGen leveraged SDFs to produce coherent geometry from distributed 2D views, CrossSDF adapts this foundation to reconstruct geometry from planar cross-sections where critical information exists only at sparse intervals. The transition from GeoGen to CrossSDF represents a deliberate technical progression. Where GeoGen introduced general SDF constraints for arbitrary objects viewed from multiple angles, CrossSDF develops specialized sampling strategies, loss formulations, and architectural innovations specifically targeting the unique sparsity patterns of cross-sectional data. This specialization allows us to recover intricate branching structures and fine details that would be impossible with general-purpose reconstruction methods, including our own GeoGen approach.

By studying these complementary approaches to neural SDF reconstruction, we gain

deeper insights into how geometric principles can be encoded into neural architectures to address increasingly specialized reconstruction challenges from general object geometry to the precise recovery of thin, complex structures critical for medical imaging, manufacturing, and beyond.

3.1 Introduction

Reconstructing three-dimensional structures from sparse two-dimensional cross-sectional data is a challenging task often encountered in medical imaging (Liu et al., 2008; Jacobs et al., 2008; Geiger, 1993), manufacturing (Castellan and Dastarac, 2000), and topography (Zou et al., 2015). In medical imaging in particular, intricate branching structures such as blood vessels and neural tissues must be reconstructed for their correct interpretation in diagnoses and interventions (Al Moussawi et al., 2015). These structures often appear as sparse and incoherent within cross-sectional scan data, which presents a significant problem for reconstruction methods. While traditional optimization-based reconstruction techniques designed for slice data can be effective for parallel cross-sections sampled from simple shapes (Bajaj et al., 1996; Barequet and Sharir, 1994), they rely on interpolating between cross-sections, which can lead to significant ridging artifacts between the slices (Sawdayee et al., 2023). Bespoke reconstruction algorithms that handle the more challenging problem of non-parallel cross-sections can scale poorly (Zou et al., 2015; Bermano et al., 2011), limiting their application beyond sparsely-sliced simple geometry. The reliance on smoothing algorithms during post-processing also makes them unsuitable for reconstructing thin structures (Zou et al., 2015). Consequently, there is a need for scalable methods that can reconstruct high-quality geometry, containing thin structures, from arbitrary cross-sections.

Concurrently, we have seen a significant growth in implicit neural reconstruction methods that characterize shapes as continuous functions of spatial coordinates (Chen and Zhang, 2019; Mescheder et al., 2019; Park et al., 2019). By extracting 3D points from the cross-sections one could frame the task as a point-cloud reconstruction problem. However, as seen in Figure 3.5, current general purpose implicit methods (e.g. (Wang et al., 2024; Boulch and Marlet, 2022)) are poorly adapted to the sparsity between slicing planes, leading to gaps, broken geometry, staircase artifacts, and over-smoothing. The recently introduced OReX (Sawdayee et al., 2023) is a neural reconstruction approach specifically designed for cross-sectional data. However, due to their indicator field parametrization and spatial sampling pattern, OReX introduces artifacts and misses thin

details (see Figure 3.7).

To overcome these limitations, we propose CrossSDF, a hash-based (Müller et al., 2022) neural reconstruction approach that takes a set of 2D cross-sections of an object as input and faithfully reconstructs the full 3D shape. We begin by generating 2D signed distance labels from the input cross-sections, which are then used with a novel loss to learn a 3D neural signed distance field (SDF). Our loss minimizes the visual artifacts that arise from the sparse nature of the 2D cross-sections, which standard reconstruction methods fail to address. CrossSDF produces high-fidelity 3D reconstructions, excelling in objects containing thin structures.

Our contributions are as follows:

- We introduce a new approach for learning a 3D neural SDF from 2D cross-sections via a novel symmetric difference loss that is capable of representing thin structures without generating unwanted artifacts.
- We propose an adaptive sampling strategy for SDFs points which ensures that thin structures are adequately represented during surface reconstruction.
- We describe a hybrid encoding approach that combines a detail preserving hash encoding with the smoothness of Fourier features to reduce the noticeable grid interpolation artifacts of existing hash-based methods.
- We propose a new challenging benchmark for thin structure reconstruction from cross-sections.
- We obtain state-of-the-art reconstruction results across a wide variety of both synthetic and real shapes.

3.2 Related Work

3.2.1 3D Reconstruction from Cross-sections.

Early methods primarily addressed reconstruction from parallel planes, a scenario often found in medical imaging and topography (Bajaj et al., 1996; Barequet and Sharir, 1994). These traditional approaches, while effective for simple objects, often rely on explicit interpolation between cross-sections, leading to extreme banding artifacts. Later methods were able to accommodate more complex inputs, such as arbitrarily oriented cross-sections, which pose significant challenges in terms of continuity and

computational efficiency (Bajaj et al., 1996; Barequet and Sharir, 1994). Non-parallel cross-sections arise in settings such as freehand 2D ultrasound, where the acoustic beams from the probe produce a set of arbitrary planar cross-sections that can capture finer-scale anatomical structures compared to parallel-only orientations (Shhadi and Barequet, 2023; Barequet and Vaxman, 2009). Some classical algorithms first tessellate space into the arrangement of the C cutting planes and then solve an optimisation in every cell. Because an arrangement of C planes has $\Theta(C^3)$ volumetric cells in the worst case, their time *and* memory footprint is $O(C^3)$ (Bermano et al., 2011; Zou et al., 2015; Sawdayee et al., 2023). CrossSDF never constructs that arrangement. Each training step touches only the current mini-batch of SDF samples, so the overall cost is $O(NS)$—linear in the number of optimisation iterations N and the total samples per iteration S. In practice we train with $S \approx 2 \times 10^5$ points per epoch and converge in 500 epochs, while arrangement-based solvers already become memory-bound beyond a few hundred contours. Recently, OReX (Sawdayee et al., 2023) introduced an implicit neural network approach for this problem setting. However, their choice of surface representation and positional encoding prevents them from reconstructing fine details and complex structures. Our CrossSDF approach addresses these limitations, and can handle complex objects to reconstruct surfaces from sparse data with minimal artifacts. It can be found on our webpage `https://iamsalvatore.io/cross_sdf/`.

3.2.2 Neural Surface Reconstruction.

Neural networks have found wide applications as a way to smoothly parameterize geometry, offering a compact and versatile alternative to classical 3D representations. As a result, reconstructing surfaces from raw data is now a well-studied problem, with applications in image-based reconstruction, point cloud reconstruction, and shape-completion (Yariv et al., 2024; Wang et al., 2021; Li et al., 2023b; Boulch and Marlet, 2022). Of these, reconstructing surfaces from sparse point cloud data is related to our cross-section reconstruction task.

Due to the difficulty of obtaining surface normals, which are required for classical learning-free methods such as Poisson reconstruction (Kazhdan and Hoppe, 2013; Kazhdan et al., 2006), neural network-based approaches have been developed that can operate on unoriented data by predicting the signed distance to a surface (Atzmon and Lipman, 2020; Sitzmann et al., 2020). Regularization has also been demonstrated to be effective for reconstructing from sparse data, e.g. Eikonal regularization as

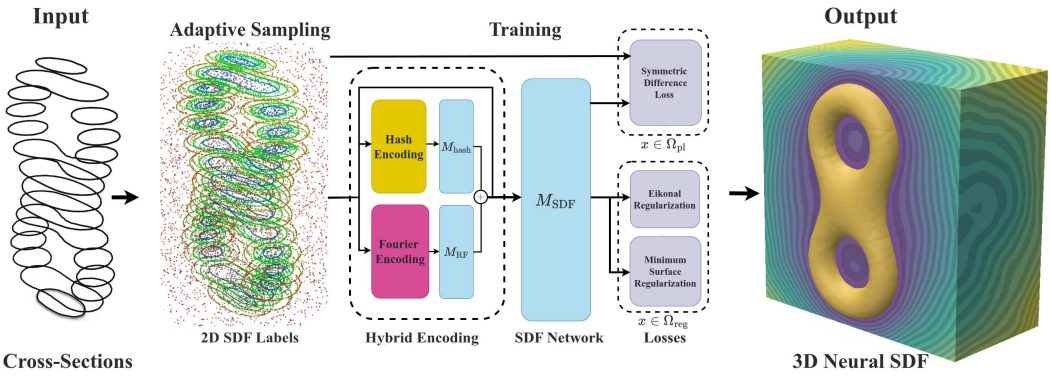


Figure 3.1: **CrossSDF** takes a set of planar cross-sections as input. While this illustration shows parallel cross-sections for clarity, our method handles arbitrarily oriented planar cross-sections without requiring parallel alignment. These cross-sections result in contours which denote the surface boundaries of the target geometry of interest (left, black lines) and each induces a 2D signed distance field (SDF) in its respective plane. From this, we generate a set of planar sample points Ω_{pl} and their 2D SDF labels. During training, points are encoded using our hybrid encoder before being passed to the SDF network M_{SDF} for prediction. At each iteration we create a set of 3D samples Ω_{reg} to apply volumetric regularization. The combination of our novel sampling, loss function, and hybrid encoding results in a high quality 3D SDF.

in (Gropp et al., 2020b) or implicit-moving-least-squares as in (Wang et al., 2024). Numerous subsequent improvements have been made by leveraging various properties of SDFs to induce stronger priors (Ma et al., 2023; Yifan et al., 2021; Ma et al., 2021b). Nonetheless, these approaches can struggle with noise as well as maintaining surface continuity when the input is sparse.

In parallel, data-driven approaches have also been proposed which encode strong priors by training on large datasets to learn the mapping between point clouds and surfaces (Park et al., 2019; Erler et al., 2020; Boulch and Marlet, 2022). This can result in improved robustness to noise and missing regions. Works such as DeepSDF (Park et al., 2019) optimize a latent space to condition an SDF decoder. Points2Surf (Erler et al., 2020), and most recently POCO (Boulch and Marlet, 2022), directly encode patches through local and global operations, to obtain better generalization. However, these methods can still fail on out-of-distribution inputs, e.g. complex objects typically captured with cross-sectional data.

Finally, advancements in image-based reconstruction methods have demonstrated

the efficacy of multi-resolution hash-encodings, enabling models to capture fine surface details (Li et al., 2023b; Müller et al., 2022). This type of detail is typically lost in sparse point cloud data, but can be present in the dense data found in 2D cross-sections. With this in mind, we leverage hash-encodings to capture the fine details and thin structures captured in the 2D cross-sections of complex objects.

3.3 Method

3.3.1 Overview

3.3.1.1 Input and output

Our input is a collection $\mathcal{P} = \{P_1, \ldots, P_n\}$ of n 2D planes embedded in $\mathbb{R}^3$ with arbitrary offsets and orientations. On each plane, we have a set of closed contours $C_i = \{c_i^1, \ldots, c_i^{l_i}\}$. The contours are the intersection of the planes and the boundary of the target geometry of interest, and they divide the plane into inside and outside regions. Our output is a neural 3D signed distance field (SDF) $f(\mathbf{x}; \theta)$, defining the predicted geometry implicitly as its zero set $\mathbf{S} = \{\mathbf{x} \in \mathbb{R}^3 \mid f(\mathbf{x}; \theta) = 0\}$.

3.3.1.2 Approach

Previous works like OReX (Sawdayee et al., 2023) and Bermano *et al.* (Bermano et al., 2011) define a 2D indicator function on the cross-section planes and interpolate it to obtain a 3D indicator function, where the cross-section operation reproduces the 2D function. However, this is detrimental to the reconstruction task for two main reasons: 1) indicator functions are discontinuous, causing severe normal and smoothness artifacts as shown from OReX qualitative results in Figure 3.5 and Figure 3.8, and 2) training on 2D indicator functions reduces the task to an unstable binary classification problem that easily misclassifies thin features. With the objective of our work to make this training stable, smooth, and accurate, we instead learn a 3D SDF, guided by the contours' planar SDFs. The associated Eikonal regularization has been demonstrated to be effective on sparse inputs (Gropp et al., 2020b). However, naively fitting a global 3D SDF to the 2D SDFs results in significant artifacts, which we successfully mitigate by only considering the *symmetric difference* between the prediction and the target (Section 3.4). To ensure we capture thin structure, we use an improved sampling strategy (Section 3.3.2), accompanied by hybrid encoding (Section 3.3.3) to maintain surface

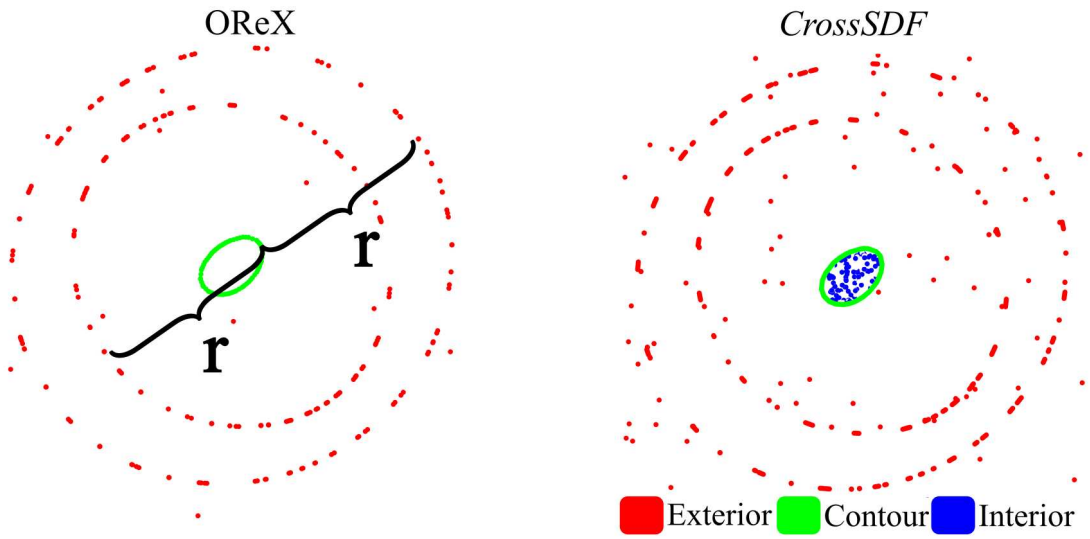


Figure 3.2: Illustration of the adaptive sampling process.

smoothness. We summarize our full pipeline in Figure 3.1.

3.3.2 Adaptive Sampling

We define the set of n 2D SDFs as $f_{2D} = \{f^i_{2D}(\mathbf{x}) : \mathbb{R}^2 \to \mathbb{R}\}$, where $i \in \{1,\dots,n\}$. To supervise our training of the 3D SDF f, we create a set Ω_{pl} of in-plane point samples to which we assign the respective (precomputed) f_{2D} values. We generate in-plane point samples Ω_{pl} with thin structures in mind. For these types of objects, where many small contours are present, as seen in Figure 3.2, a naive uniform sampling of 2D points on each cross-section would under-sample structures with low cross-sectional area. OReX (Sawdayee et al., 2023) samples points at a fixed radius r at either side of a contour, alongside a uniform sampling of the plane. However, for thin structures, a fixed radius will often extend beyond the contour's interior, leading to disproportionately fewer interior samplesFigure 3.2. To address this, we adapt the OReX sampling, by directly sampling the interior of every contour until a threshold number of samples is taken (we use 100 samples per contour as detailed in Appendix B.2.3. This per-contour sampling approach ensures the interiors of small contours are adequately sampled, discouraging our model from ignoring contours with small surface area. However, several considerations arise from this approach: (i) The threshold for samples per contour requires careful random search tuning, as too few samples may miss thin structures while too many can lead to computational overhead and potential overfitting to specific contour shapes. In practice, we found 100 samples per contour provides a

good balance across our datasets. (ii) Per-contour interior sampling is computationally more expensive than uniform sampling, requiring approximately two times longer preprocessing time, but this cost is amortized over training as samples are precomputed. (iii) Fixing a maximum number of samples per contour can introduce sampling density biases, where smaller contours receive higher sampling density relative to their size compared to larger contours. This bias actually benefits thin structure reconstruction, as it ensures adequate representation of structures that would otherwise be under-sampled. This is made possible by our problem setting which provides ground-truth segmentation of interior and exterior regions on each slicing plane. This per-contour sampling approach ensures the interiors of small contours are adequately sampled, discouraging our model from ignoring contours with small surface area (which we demonstrate in ablations).

3.3.3 Hybrid Encoding

Optimizing a neural field directly on 3D coordinates can result in over-smooth geometry, as neural networks are biased towards low-frequency functions on low dimensional domains (Tancik et al., 2020). As a result, positionally encoding 3D input coordinates with Fourier features has become common practice to learn detailed implicit fields (Gao et al., 2022b). Grid-based encodings offer an alternative, encoding positions by interpolating learned feature vectors stored on grid vertices (Takikawa et al., 2021; Liu et al., 2020; Li et al., 2023b). Recently, the use of spatial hashing has allowed grid-based encodings to scale to higher resolutions, and therefore capture even finer surface details (Müller et al., 2022; Li et al., 2023b). Inspired by this, we employ a hash-grid encoding to capture thin structures. However, these encodings are not directly suitable to our problem, since they introduce grid artifacts visible in the resulting geometry. These artifacts are due to linear interpolation that is performed across grid cell boundaries, leading to discontinuities that manifest as sharp creases or ridges in the learned surface (see Figure 3.13). This effect is especially pronounced for our task, since the sparsity of the cross-sections means that large 3D regions between input planes are not intersected and we rely exclusively on priors for interpolation. To mitigate these issues, we introduce an alternative encoding strategy combining random Fourier features (Tancik et al., 2020) with multi-resolution hash-grid features (Müller et al., 2022). We define the random Fourier feature (RFF) encoding $\gamma_{\text{RF}} : \mathbb{R}^3 \to \mathbb{R}^d$ as follows:

$$\gamma_{\text{RF}}(\mathbf{x}) = [\cos(z_1), \sin(z_1), \ldots, \cos(z_{d/2}), \sin(z_{d/2})], \tag{3.1}$$

where $z_i = \mathbf{b}_i^\top \mathbf{x}$ and the coefficients $\mathbf{b}_i \in \mathbb{R}^3$ are sampled randomly from a multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, with mean $\mathbf{0}$ and variance σ^2. This sampling ensures a full set of encoding functions spanning a range of frequencies and alignments.

Our hybrid encoding strategy first encodes an input 3D coordinate $\mathbf{x}$ using two different encoding functions; a hash-grid encoding $\gamma_{\text{hash}} : \mathbb{R}^3 \to \mathbb{R}^d$ and a RFF encoding $\gamma_{\text{RF}} : \mathbb{R}^3 \to \mathbb{R}^d$. Each encoding is passed through separate one hidden layer MLPs, defined as $M_{\text{hash}}, M_{\text{RF}} : \mathbb{R}^d \to \mathbb{R}^{d'}$, which are learned, producing intermediate encodings $\mathbf{z}_{\text{hash}} = M_{\text{hash}}(\gamma_{\text{hash}})$ and $\mathbf{z}_{\text{RF}} = M_{\text{RF}}(\gamma_{\text{RF}})$. We combine these intermediate encodings by adding them, where the RFF encoding is scaled by a small constant α,

$$\mathbf{z}_{\text{comb}} = \mathbf{z}_{\text{hash}} + \alpha \cdot \mathbf{z}_{\text{RF}}. \tag{3.2}$$

This constant is a hyperparameter chosen so that the magnitudes of each encoding is approximately equal upon initialization. This constant is a hyperparameter chosen so that the magnitudes of each encoding is approximately equal upon initialization. The scaling factor α is critical for balancing the contributions of the two encodings: too large and the smooth RFF features dominate, potentially losing fine details; too small and grid artifacts persist. We empirically set $\alpha = 0.1$ across all experiments. The combined encoding is then concatenated with the original input $\mathbf{x}$ to form the final hybrid encoding input to the SDF MLP, M_{SDF},

$$\mathbf{z}_{\text{final}} = [\mathbf{z}_{\text{comb}} \mid \mathbf{x}], \tag{3.3}$$

where $[\cdot \mid \cdot]$ represents concatenation (see Figure 3.1). The inclusion of random Fourier features encourages smooth representations, reducing interpolation artifacts caused by the discontinuities at hash-grid cell boundaries and improves surface continuity.

3.4 Optimization

3.4.1 Symmetric Difference Loss

The 2D SDFs and the desired 3D SDF generally disagree. For parallel slices, fitting the 2D SDF directly will encourage the normals of f and f_{2D} coincide, causing laddering artifacts (see Figure 3.3 and exhibited in the ablation in Figure 3.13). For non-parallel slices, there are actual inconsistencies along their intersection in 3D space. However,

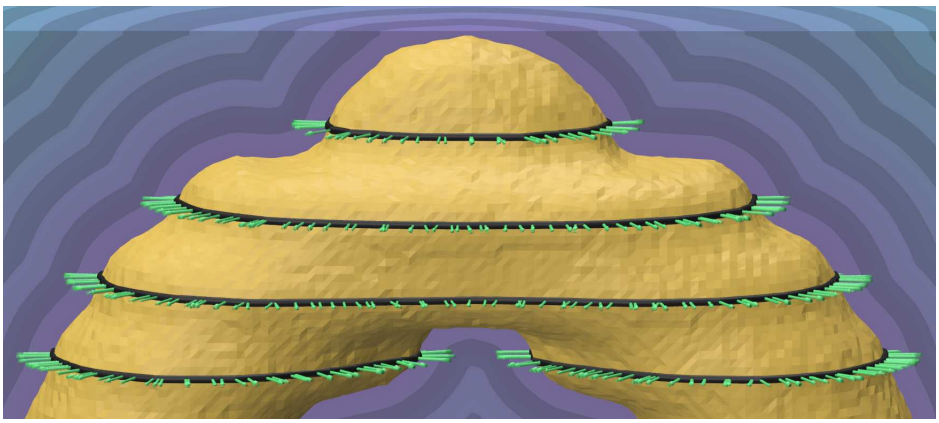


Figure 3.3: Close-up of the figure eight object from Figure 3.1 when fitting a 3D SDF directly to the *2D SDF* labels. Displayed are the input contours (black) along with 2D SDF gradient vectors (green arrows). Here the neural 3D SDF attempts to remain orthogonal to the cross-sections, resulting in a laddering effect.

the only geometric data each input cross-section contains about the target geometry is the classification of interior/exterior. As such, we do not want the neural field to learn the 2D SDFs directly, but rather use them as a smooth guide to the zero set whilst ultimately learning a 3D SDF. To avoid these artifacts, our insight is that we can drive the surface optimization with the 2D SDFs by only considering the region where the target and predicted classifications of interior/exterior disagree. This region shrinks to zero when the target and predicted contours exactly overlap, freeing the neural field from the requirement to be a 2D SDF within the plane. However, it's important to note that the symmetric difference loss primarily enforces shape accuracy within each individual slice and cannot directly enforce topological relationships between slices. This limitation means that complex inter-slice connectivity, such as vessel bifurcations occurring between sparse cross-sections, must be handled through the network's implicit interpolation capabilities and spatial regularization terms. The implications include potential topological inconsistencies in regions with insufficient cross-sectional coverage, though our Eikonal regularization and minimum surface constraints help mitigate these issues by encouraging globally coherent surface reconstruction. At each iteration, we consider the two disjoint sets,

$$\begin{aligned} \Omega_{\text{on}} &= \{\mathbf{x} \in \Omega_{\text{pl}} \mid f_{\text{2D}}(\mathbf{x}) = 0\}, \\ \Omega_{\text{off}} &= \{\mathbf{x} \in \Omega_{\text{pl}} \mid f_{\text{2D}}(\mathbf{x}) \neq 0,\ [\![f(\mathbf{x};\theta)]\!] \neq [\![f_{\text{2D}}(\mathbf{x})]\!]\}, \end{aligned} \tag{3.4}$$

where $[\![]\!]$ indicates the sign function.

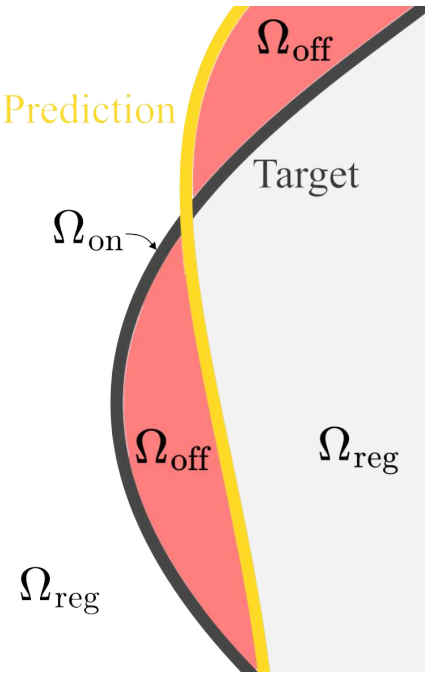


Figure 3.4: Cross-section showing target contour (black line), predicted surface (yellow), on-contour samples (Ω_{on}), off-contour regions (Ω_{off}), and regular regions (Ω_{reg}) which are distributed throughout the entire 3D volume for regularization purposes, not just the complement of the shown regions..

in Figure 3.4, we illustrate where these sample sets are on an example cross-section with target contour (black line) and predicted surface (yellow). The set Ω_{on} are on-contour samples. The red regions, where target and predicted classifications of interior/exterior differ, is known as the *symmetric difference* between the two interiors (Alt and Guibas, 2000). We define separate loss functions on each set,

$$\begin{aligned} \mathcal{L}_{\text{on}} &= \mathbb{E}_{\mathbf{x}\in\Omega_{\text{on}}}\left(|f(\mathbf{x};\theta) - f_{\text{2D}}(\mathbf{x})|\right), \\ \mathcal{L}_{\text{off}} &= \mathbb{E}_{\mathbf{x}\in\Omega_{\text{off}}}\left((f(\mathbf{x};\theta) - f_{\text{2D}}(\mathbf{x}))^2\right). \end{aligned} \tag{3.5}$$

The different loss formulations serve distinct purposes: the L1 (absolute) loss for on-contour samples provides robust fitting to the exact boundary locations, while the L2 (squared) loss for off-contour samples provides stronger gradients to quickly drive misclassified regions toward correct classification, accelerating convergence in the symmetric difference regions. By regressing the 2D SDFs only from samples on the symmetric difference Ω_{off} and the contour Ω_{on}, we can fit the surface while regularizing f to be a 3D SDF on the correctly classified regions. This approach successfully mitigates the aforementioned laddering artifacts and improves reconstruction accuracy.

3.4.2 Spatial Regularization

We produce a set Ω_{reg} by uniformly sampling points in the full 3D volume. These points are used solely for regularizing f, and are generated at each iteration. We use Eikonal regularization to make f a 3D SDF. This regularization term is defined as:

$$\mathcal{L}_{\text{eik}} = \mathbb{E}_{\mathbf{x}\in\Omega_{\text{reg}}} \left(\|\nabla_{\mathbf{x}} f(\mathbf{x};\theta)\| - 1\right)^2. \tag{3.6}$$

Additionally, due to the sparse nature of our supervision regions Ω_{pl}, the model is susceptible to generating spurious surfaces, colloquially known as floaters, in regions that are not supervised. To address this issue, we use a minimum surface loss, a regularizer specifically designed to suppress the formation of these unwanted surfaces by penalizing low SDF values (Rebain et al., 2021). We use an exponential term with hyperparameter β, which penalizes the SDF values that are near zero, and thus discourages the model from generating surfaces at these locations unless strongly supported by the data. The minimum surface loss is defined as follows:

$$\mathcal{L}_{\text{min}} = \mathbb{E}_{\mathbf{x}\in\Omega_{\text{reg}}} \left(e^{-\beta|f(\mathbf{x};\theta)|}\right), \tag{3.7}$$

where $f(\mathbf{x})$ represents the predicted SDF value at the point $\mathbf{x}$, and Ω_{reg} denotes the regularization domain encompassing the entirety of our input space. This is important in preventing the emergence of geometrically incorrect features in the reconstructed model, thereby enhancing the SDF predictions. Our combined training loss is: Our combined training loss is:

$$L = L_{on} + L_{off} + \lambda_{eik} L_{eik} + \lambda_{min} L_{min}, \tag{3.8}$$

where λ_{eik} and λ_{min} are regularization weight hyperparameters. Please see Appendix B.2 for implementation details including specific hyperparameter values. where λ_{eik} and λ_{min} are regularization weight hyperparameters.

3.5 Experiments

3.5.1 Dataset

We evaluate CrossSDF on both synthetic and real-world datasets. We first consider reconstruction from cross-sections using two synthetic datasets, divided into thin and thick structures. The thin structures dataset consists of six patient-specific, anatomically

		Thin Structures											
		Alveolis (100)		Cerebral (75)		Coronaries (75)		Coro (75)		Heart (100)		Pulmonary (75)	
	Method	Align.	Non-Align.	Align.	Non-Align.	Align.	Non-Align.	Align.	Non-Align.	Align.	Non-Align.	Align.	Non-Align.
CD (↓)	CrossSDF (Ours)	**0.35**	**0.50**	**0.24**	**0.26**	**0.28**	**0.46**	**0.21**	**0.29**	**0.38**	**0.82**	**0.24**	**0.34**
	OReX Sawdayee et al. (2023)	3.6	4.8	12	2	10	1.6	0.48	0.33	4.6	8.4	12	5.6
	Screened Poisson Kazhdan and Hoppe (2013)	0.44	0.57	0.39	0.44	0.46	0.67	1.3	0.95	1.0	0.85	0.52	0.59
	POCO Boulch and Marlet (2022)	4.1	4.5	0.36	0.79	1.1	0.72	1.1	3.4	2.7	4.2	2.4	2.0
	Neural-IMLS Wang et al. (2024)	13	12	14	17	19	17	9	10	19	18	15	15
HD (↓)	CrossSDF (Ours)	**14**	**6.1**	**3.5**	**4**	**5.8**	**4.7**	**4.5**	**2.8**	**11**	10	**10**	**7.3**
	OReX Sawdayee et al. (2023)	21	35	8.1	19	18	27	9	11	26	42	36	21
	Screened Poisson Kazhdan and Hoppe (2013)	17	8.1	16	25	6.1	5.4	10	7.5	26	**9.2**	**10**	7.7
	POCO Boulch and Marlet (2022)	32	29	15	41	7.1	9.2	8.8	18	24	25	20	16.4
	Neural-IMLS Wang et al. (2024)	68	51	33	48	55	67	37	39	59	61	52	41
CC	*Ground Truth*	1	1	1	1	3	3	1	1	1	1	1	1
	CrossSDF (Ours)	6	33	2	5	5	35	2	4	68	176	2	3
	OReX Sawdayee et al. (2023)	939	2119	40	421	219	94	13	4	223	511	914	261
	Screened Poisson Kazhdan and Hoppe (2013)	53	1187	3	23	81	160	9	62	923	1006	6	17
	POCO Boulch and Marlet (2022)	67	91	2	78	14	46	4	279	70	134	21	134
	Neural-IMLS Wang et al. (2024)	1	1	1	1	1	1	1	1	2	1		

Table 3.1: Quantitative results on **thin structures** across different methods and metrics. The table compares the Chamfer Distance (CD) $\times 100$, Hausdorff Distance (HD) $\times 100$, and number of Connected Components (CC), with the ground truth number also reported, for both aligned and non-aligned versions of each structure. The numbers in parenthesis in the first row denote the numbers of cross-sections. Note that for the Connected Components (CC) metric, we report absolute numbers rather than using bold formatting, as the ground truth number of connected components varies by dataset (shown in the ground truth row), making relative performance assessment more meaningful than highlighting the lowest absolute values.

accurate, meshes from (Wilson et al., 2013), originally introduced for fluid and solid mechanics evaluation which we adapt for 3D reconstruction. It contains complex, thin branching geometry, which is particularly challenging for reconstruction methods. For real-world evaluation, we also test on clinical datasets including the IRCADb-01 (Soler et al., 2010)and Medical Decathlon (Antonelli et al., 2022) datasets, which contain actual CT scan data with human-annotated vessel segmentations. We generate 75–100 cross-sections for each, which despite being relative dense, still presents a significant challenge. The thick structures dataset consists of six meshes used in OReX (Sawdayee et al., 2023) which have simple topologies, large smooth geometric features, and result in contours with large cross-sectional areas. We take sparse cross-sections (i.e. 25) to produce the input contours, testing the models' ability to smoothly interpolate sparse inputs without producing plane-slicing artifacts or unwanted gaps. For both datasets, we also consider *aligned* and *non-aligned* setups to test the models ability to handle arbitrary cross-section orientations. For the aligned case, the cross-sections are a set of axis-aligned parallel planes. For the non-aligned case, the planes consist of half parallel and half rotated, around an axis, planes.

		Thick Structures											
		Armadillo (25)		Balloon Dog (25)		Elephant (25)		Eight (25)		Hand (25)		Brain (25)	
	Method	Align.	Non-Align.	Align.	Non-Align.	Align.	Non-Align.	Align.	Non-Align.	Align.	Non-Align.	Align.	Non-Align.
CD (↓)	CrossSDF (Ours)	**0.80**	**0.71**	**0.25**	**0.23**	**0.58**	**0.47**	**0.10**	**0.14**	**0.21**	**0.27**	**0.65**	**0.74**
	OReX Sawdayee et al. (2023)	0.86	1.0	0.29	0.28	0.69	0.55	0.18	0.16	0.29	0.39	1.5	1.4
	Screened Poisson Kazhdan and Hoppe (2013)	1.4	1.6	1.1	0.90	1.4	2.0	1.8	2.9	1.7	0.77	1.1	0.77
	POCO Boulch and Marlet (2022)	3.5	3.6	1.3	1.3	3.9	4.5	0.57	2.4	3.6	4.1	7.4	1.4
	Neural-IMLS Wang et al. (2024)	1.4	3.4	0.90	0.31	0.84	0.63	**0.10**	0.15	0.35	0.46	1.4	1.6
HD (↓)	CrossSDF (Ours)	**7.7**	**9.0**	**3.0**	**3.3**	**7.1**	**6.0**	**1.7**	1.8	**18**	**16**	**14**	**8.5**
	OReX Sawdayee et al. (2023)	8.6	9.9	6.7	4.0	7.3	**6.0**	**1.7**	**1.7**	22	22	15	14
	Screened Poisson Kazhdan and Hoppe (2013)	14	10	8.0	6.0	8.3	6.8	13	3.9	30	28	19	9.1
	POCO Boulch and Marlet (2022)	26	24	14	16	29	29	14	17	21	22	58	15
	Neural-IMLS Wang et al. (2024)	16	11	7.9	3.9	7.3	6.3	1.8	1.8	33	29	17	11

Table 3.2: Quantitative results on **thick structures** across different methods and metrics. The table compares the Chamfer Distance (CD) $\times 100$ and Hausdorff Distance (HD) for both aligned and non-aligned versions of each structure.

3.5.2 Baselines

We compare to the state-of-the-art (SoTA) non-data-driven point cloud reconstruction method Neural-IMLS (Wang et al., 2024), SoTA data-driven point cloud reconstruction method POCO (Boulch and Marlet, 2022), the bespoke cross-section reconstruction approach OReX (Sawdayee et al., 2023), and classical Screened Poisson Reconstruction (Kazhdan and Hoppe, 2013). Since screened Poisson reconstruction requires oriented points, we provide it with additional vertex normals which are computed for each point as the average of the connecting edge normals within the cross-sectional plane. Implementation details for CrossSDF can be found in the Appendix B.2 .

3.5.3 Evaluation

We report the Chamfer Distance (CD) and Hausdorff Distance (HD) compared to the ground truth mesh for all structures, and also the number of Connected Components (CC) for the thin structures. HD assesses the maximum discrepancy between two 3D meshes, which is important for precision-critical applications like medical imaging. CD complements this by averaging nearest point distances to gauge overall shape similarity. Evaluating the number of CC confirms correct segmentation and topological consistency across meshes which is important for maintaining structural integrity. Together, these metrics provide a thorough framework for precise 3D mesh analysis. While we refer the reader to (Oechsle et al., 2021) for definitions. However, no single metric fully captures the perceptual quality of the reconstructions. For example, CD gives a good sense of overall similarity, but it smooths out small deviations produced by surface details or ridging artifacts.

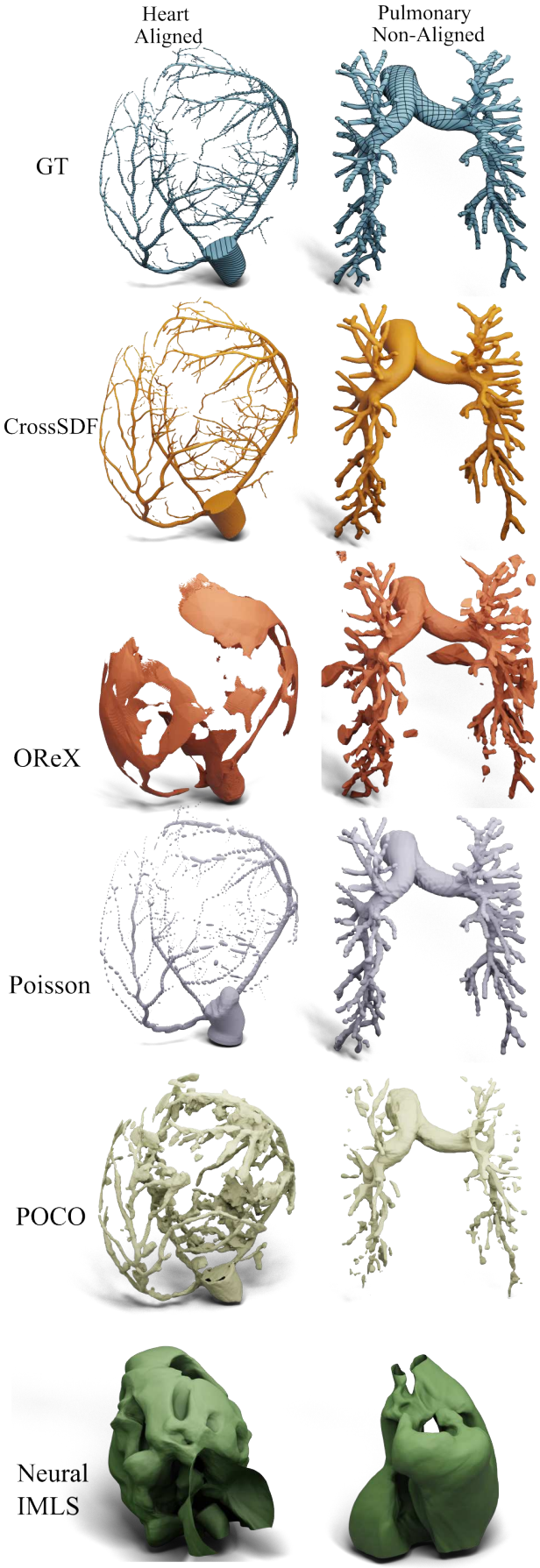


Figure 3.5: Reconstruction Results on **thin structures** featuring the Heart (top row) and Pulmonary (bottom row) for various methods. Results are presented using both input-aligned and non-aligned planes, displayed on the ground truth meshes.

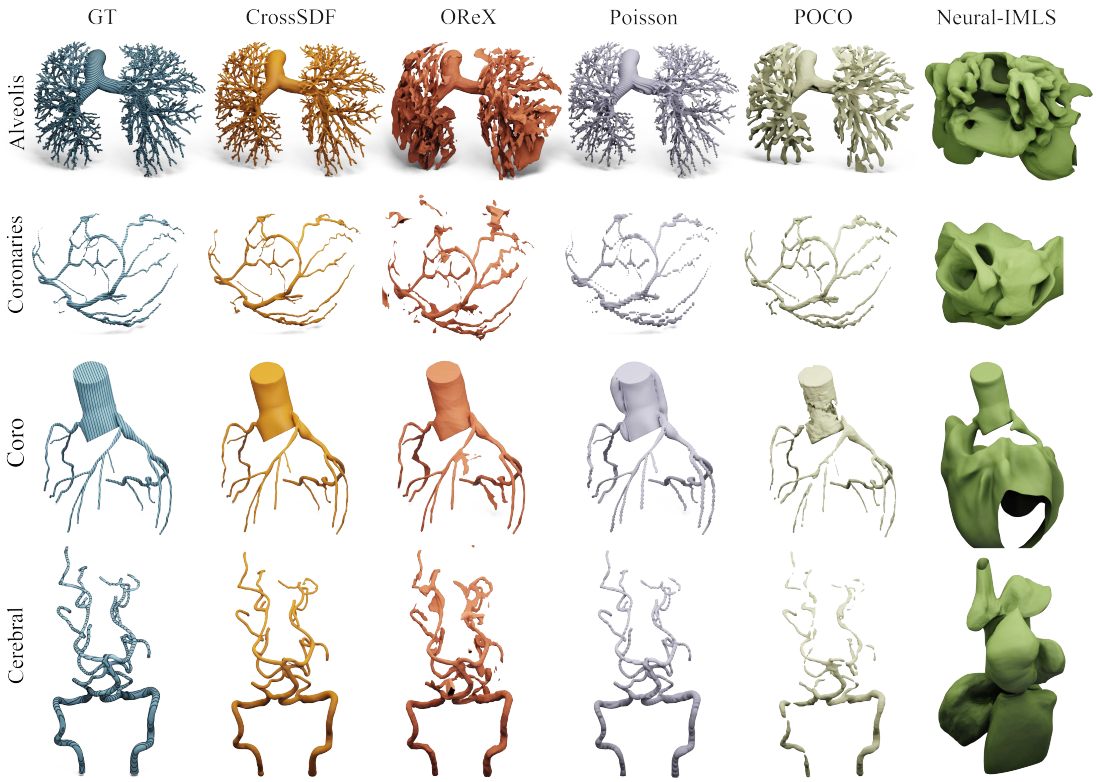


Figure 3.6: Additional qualitative results of thin synthetic data in the aligned setting. We compare our model with existing methods.

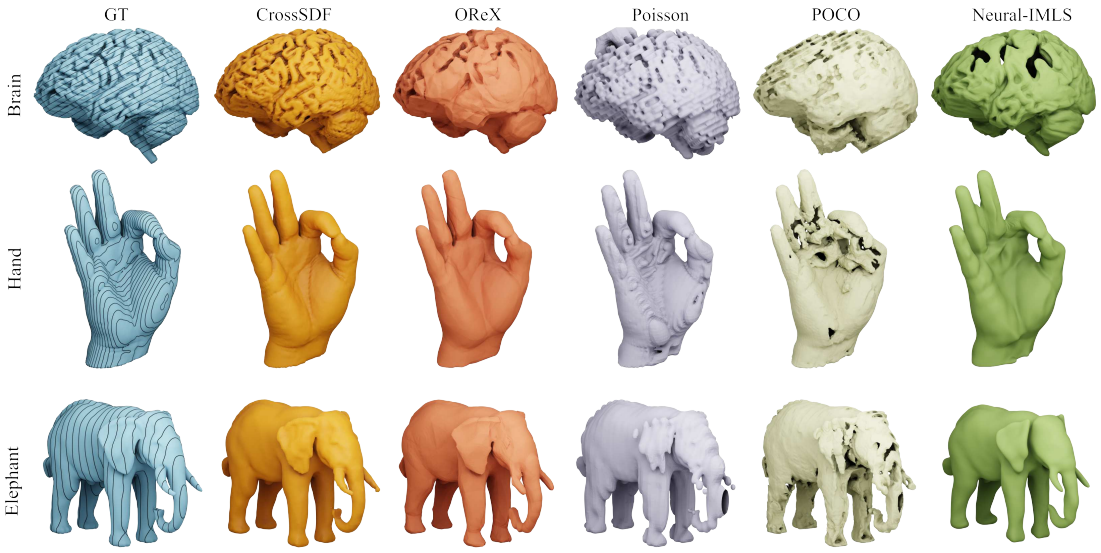


Figure 3.7: Additional qualitative results of different methods on the thick synthetic dataset, for aligned slices. We compare our model with existing methods.

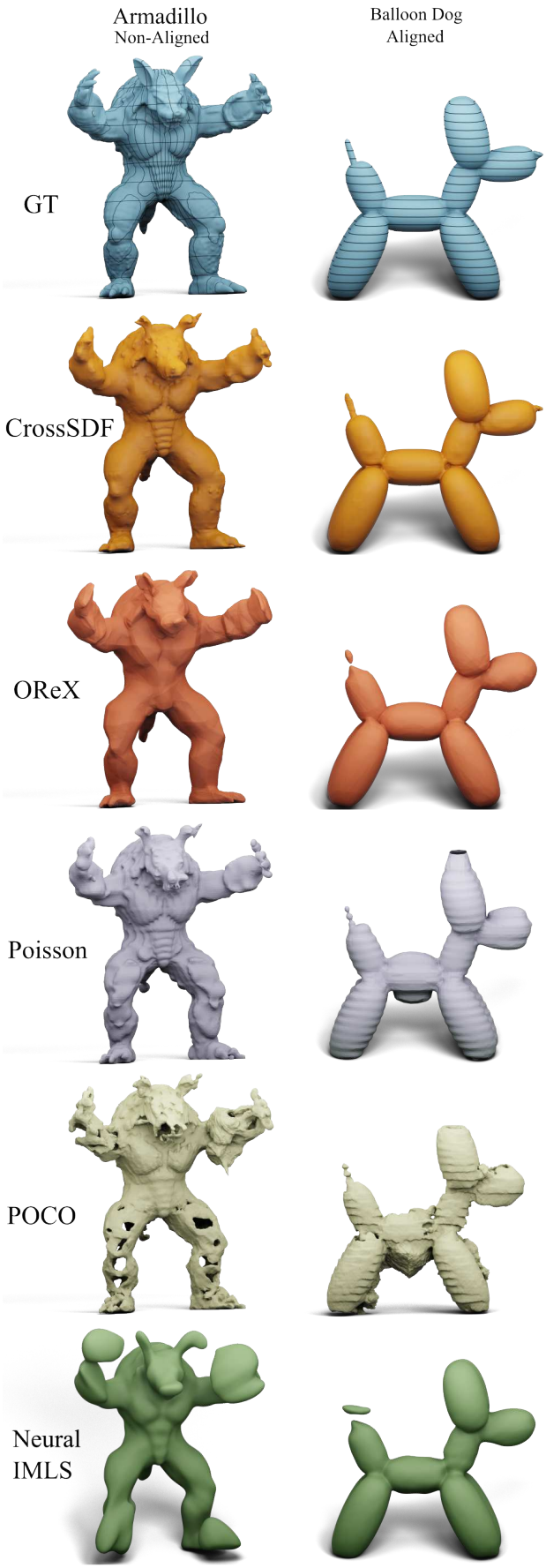


Figure 3.8: Reconstruction Results on **thick structures** featuring the Armadillo (left) and Balloon Dog (right) for various methods. Results are presented using both non-aligned and input-aligned planes, displayed on the ground truth meshes.

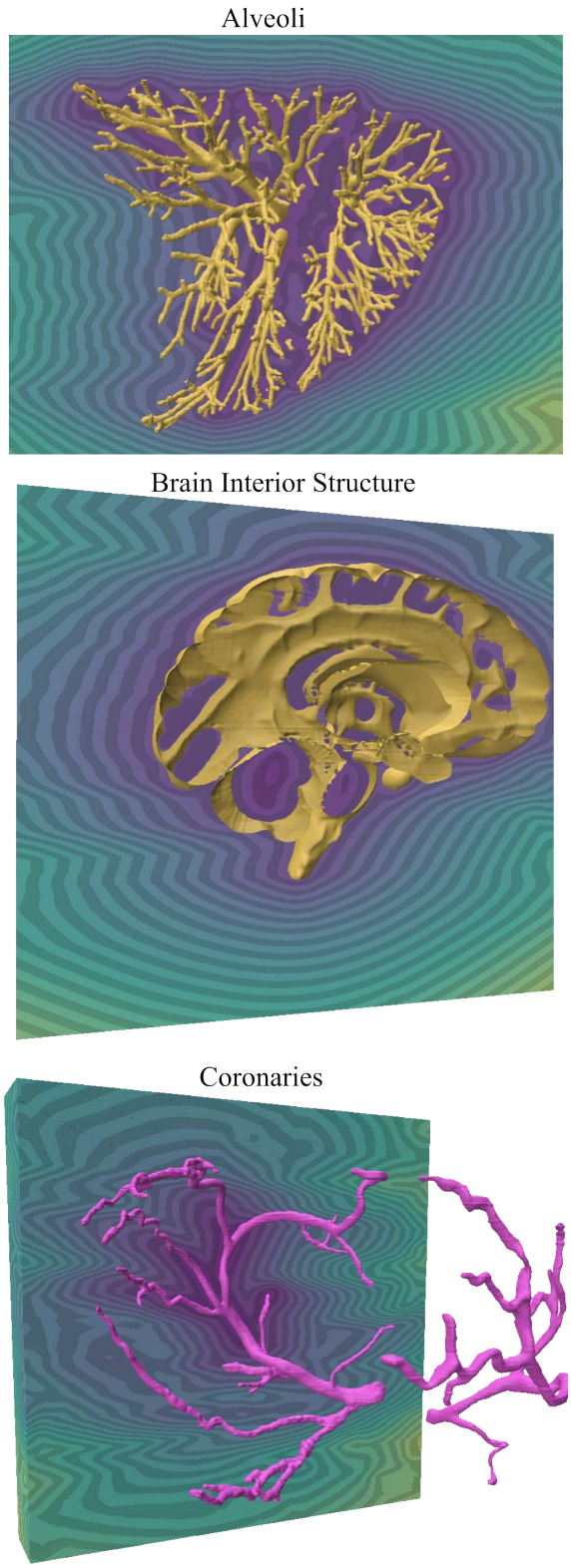


Figure 3.9: This figure displays the SDF isolines derived from 3D reconstruction results. SDF isolines are curves that connect points with the same signed distance value from the object's surface, essentially forming level sets. They provide a detailed map of the underlying geometry, highlighting subtle shape changes, curvature, and boundaries. This makes them an invaluable diagnostic tool for assessing reconstruction accuracy and revealing fine-scale structural features within complex shapes.

3.6 Results

Tables 3.1 and 3.2 contain quantitative results for both thick and thin structures. Qualitative results, for a subset of structures, can be found in Figures 3.5 and 3.8. Additional qualitative results regarding the quality of the SDF and its isolines derived from the 3D reconstruction can be seen in Figure 3.9. The data-driven POCO performs poorly, particularly on thick structures where the sparsity between planes leads to unwanted gaps (e.g. Armadillo in Figure 3.8), likely due to being an out-of-distribution input. Neural-IMLS performs reasonably well on thick structures but suffers from over-smoothing, failing to converge when trained on the thin medical structures. Note, this over-smoothing manifests as a single blob, which results in artificially good performance on the CC metric. OReX also fails to resolve these structures likely as a result of its indicator function representation, sampling issues, and choice of positional encoding. Screened Poisson exhibits significant artifacts in the form of ridging and growths, especially on thick structures. Quantitatively, it performs well on the denser sliced thin structures, particularly on Hausdorff distance. This is likely because this metric is very sensitive to outliers, and Screened Poisson does not produce spurious floaters which can be seen from implicit neural methods on challenging geometry. Despite this, it results in a highly fragmented reconstructions that are topologically far from the ground truth. It is also worth highlighting that its ridging artifacts are potentially harmful in the context of medical interpretation, despite not being very evident from the quantitative scores (e.g. Pulmonary in Figure 3.5).

In almost all cases, CrossSDF significantly outperforms competing baselines across all evaluation metrics and cross-section types. Notably, our method also maintains better topological consistency, as evidenced by the connected component counts that more closely match the ground truth topology compared to methods like Screened Poisson, which often fragments structures into numerous disconnected pieces. It generates accurate surfaces for the thick structures while still reconstructing details which are smoothed over by competing methods, e.g. the stomach region of the Armadillo in Figure 3.8. It also does not produce ridging artifacts unlike many of the baselines (e.g. Screened Poisson) in the context of sparse cross-sections, e.g. the Balloon Dog in Figure 3.8. On thin structures, CrossSDF reconstructs detailed geometry and connected branches where other methods fail.

We also test CrossSDF on real word data from the IRCADb-01 (Soler et al., 2010) and medical decathlon (Antonelli et al., 2022) datasets, which contains anonymized and

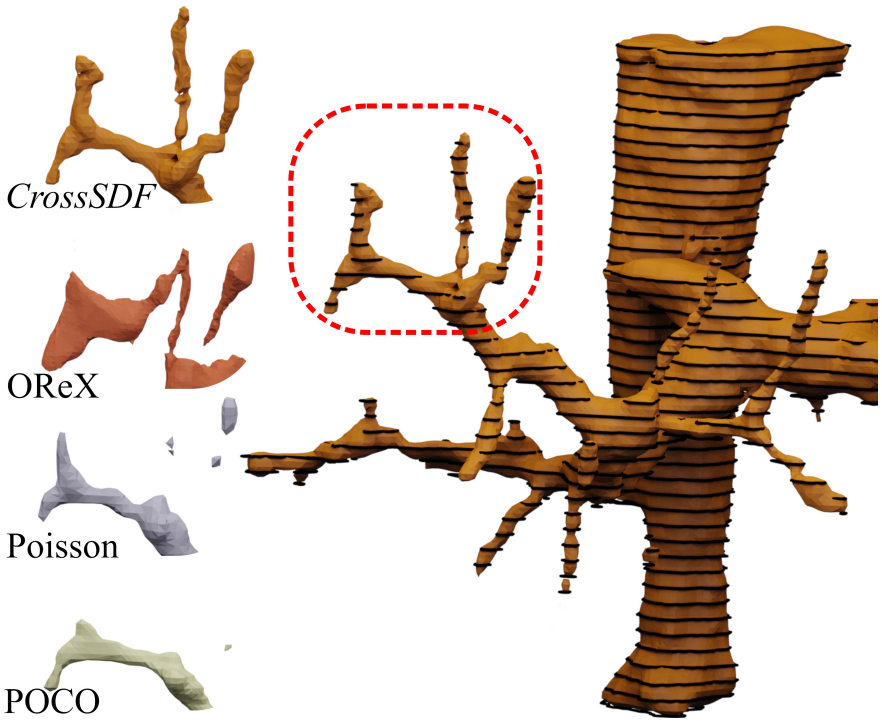


Figure 3.10: Qualitative comparison of hepatic vessel reconstruction slices derived from a real CT scan. The complete hepatic vessel reconstruction shown on the right represents our CrossSDF method output.

annotated CT scans in DICOM format. Qualitative results are shown in Figure 3.10 and Figure 3.11 on a hepatic vessel with 61 slices and 5 millimeters between slices, which in medical imaging is considered sparse. We report more qualitative and quantitative results in the Appendix B.1.

3.6.1 Ablations

We perform the following ablations to evaluate the impact of removing various core components from our method:

- **w/o Fourier Encoder:** Instead of incorporating Fourier features via our hybrid encoding method outlined in Section 3.3.3, we use a vanilla hash-encoding.
- **w/o Symmetric Difference:** Instead of our symmetric difference loss from Section 3.4, we use an L1 loss between the predicted and ground truth 2D SDF labels.
- **w/o Adaptive Sampler:** Instead of the adaptive sampling introduced in Section 3.3.2), we use the sampling from OReX (Sawdayee et al., 2023).

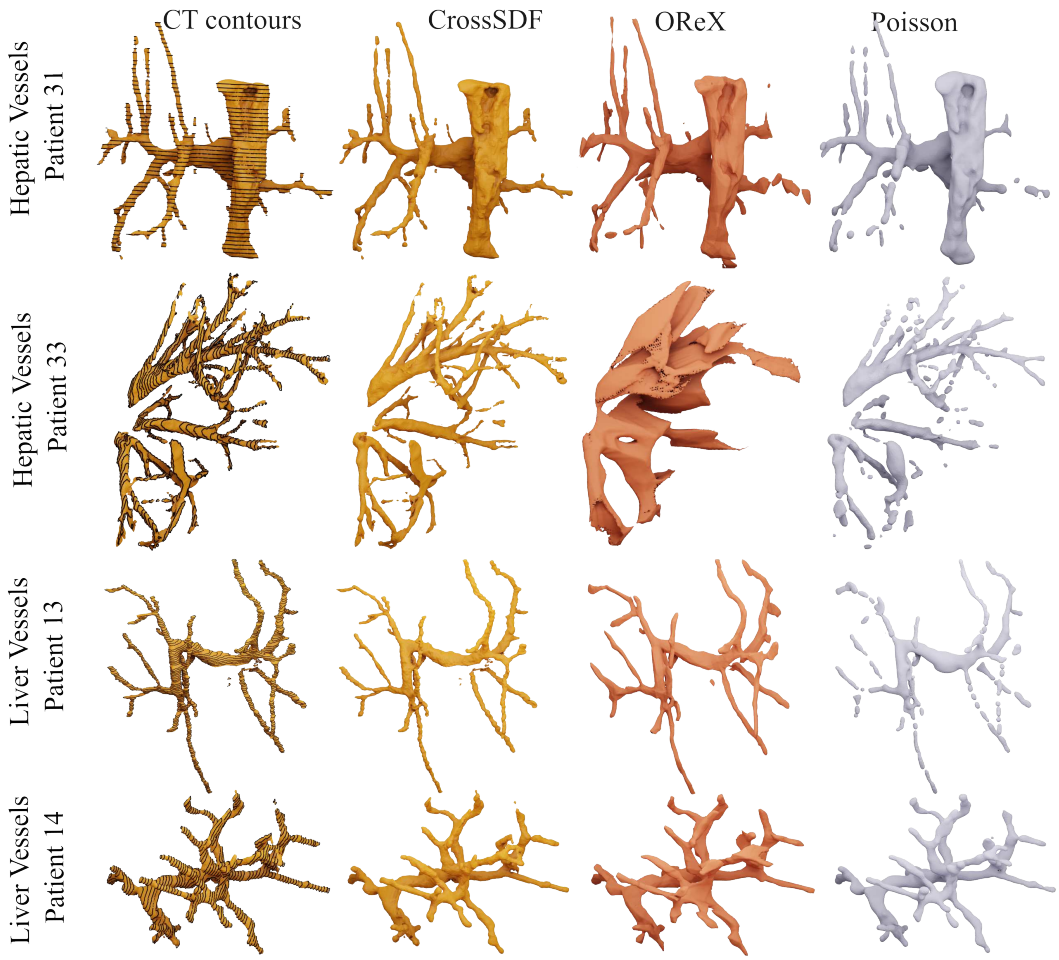


Figure 3.11: Additional qualitative results on real CT vessel structures. We compare our CrossSDF approach with existing methods. Note the 'CT contours' are overlayed on our result to improve viewing clarity.

We present the quantitative results in Table 3.4 and visualize them in Figure 3.13. Removing our hybrid encoding results in grid interpolation artifacts and loss of surface continuity (see elephant ear and stomach). When our Symmetric Difference loss is removed we observe strong staircase artifacts, which is particularly harmful on the thick structures. For the thin structures, the adaptive sampler is very important for reconstructing thin details (see Figure 3.14).

Additionally, we also evaluate CrossSDF's robustness when the number of input cross-sections are reduced. Quantitative and qualitative results can be found in Table 3.3 and Figure 3.12. Unsurprisingly, performance decreases when the number of slices are reduced, but the results are not too different from the ground truth for ≥ 50 slices, and CrossSDF outperforms OReX in all cases.

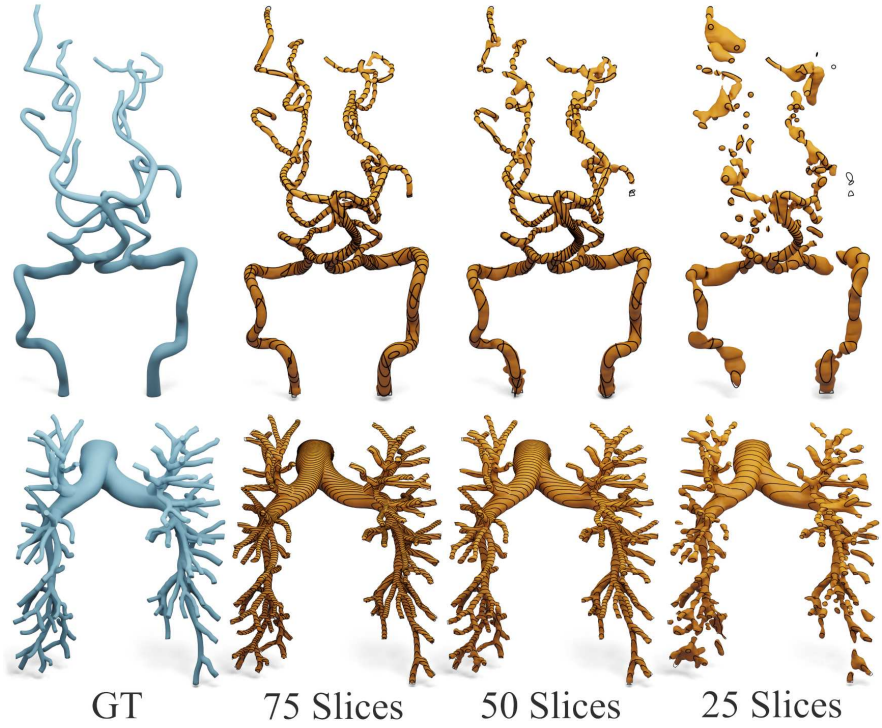


Figure 3.12: Impact of reducing the number of cross-sections on the Cerebral (top) and Pulmonary (bottom) structures using non-aligned and aligned cross-sections, respectively. The input cross-sections are denoted in black on the reconstructions (orange). Results shown are from the complete CrossSDF method.

3.6.2 Limitations

Even though CrossSDF successfully reconstructs challenging structures, it exhibits some limitations. Like the other baselines, it will still fail to accurately reconstruct an object given only a very sparse set of slices. It could benefit from data-driven priors to improve accuracy, such as anatomical shape priors learned from large medical datasets that could constrain vessel reconstruction to follow biologically plausible branching patterns and connectivity rules.In addition, the predictions from such a model would be faster compared to our optimization-based approach.

3.7 Conclusion and Discussion

Our work on CrossSDF represents a significant advancement in addressing the challenging problem of 3D thin structure reconstruction from cross-sectional data. By developing specialized techniques for contour-aware SDF training, we have demonstrated improved reconstruction results compared to existing methods, particularly for

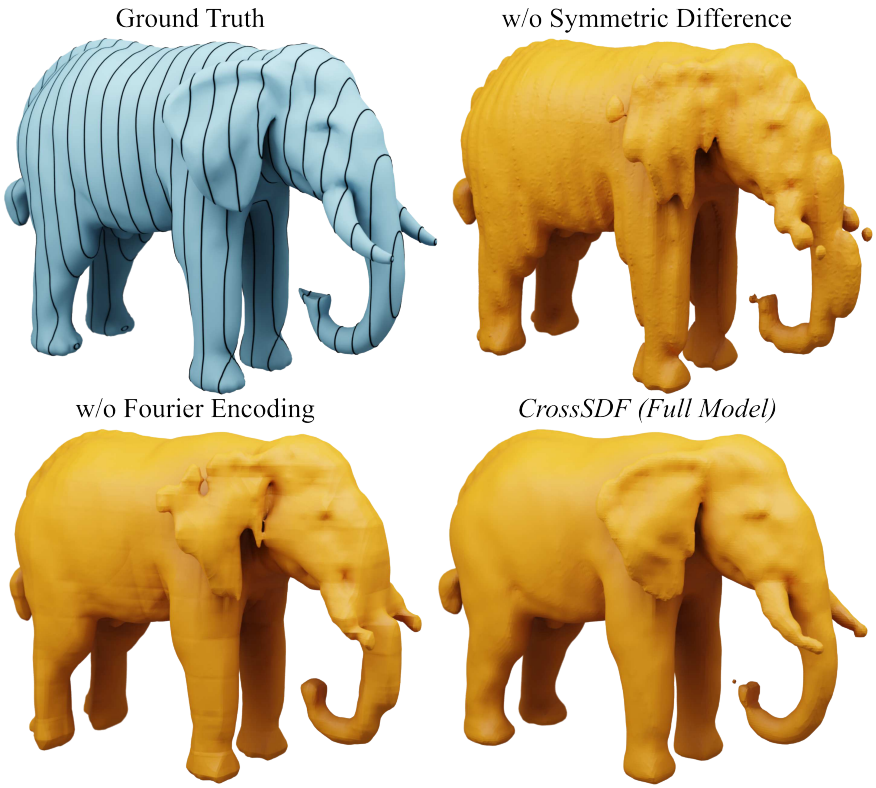


Figure 3.13: Qualitative ablation of CrossSDF with different components removed. The full model results in fewest visual artifacts.

Scene	Method	75 Slices		50 Slices		25 Slices	
		CD	HD	CD	HD	CD	HD
Cerebral	CrossSDF	**0.26**	**4**	**0.51**	**16**	**2.1**	**26**
	OReX	2	19	2.6	18	5.9	72
Pulmonary	CrossSDF	**0.24**	**10**	**0.33**	**11**	**1.0**	**13**
	OReX	12	36	14	69	18	78

Table 3.3: Ablation reducing the number of input cross-section for our CrossSDF approach and OReX.

Model	Balloon Dog		Elephant		Alveolis		Coronaries	
	CD (↓)	HD (↓)	CD (↓)	HD (↓)	CD (↓)	HD (↓)	CD (↓)	HD (↓)
CrossSDF (full model)	**0.24**	**3.0**	**0.58**	**7.1**	**0.35**	**14**	**0.28**	**5.8**
w/o Fourier Encoding	0.30	3.4	0.88	10.5	0.41	15	0.33	6.6
w/o Symmetric Diff.	0.57	4.1	1.01	14.4	0.38	15	0.29	6.1
w/o Adaptive Sampler	**0.24**	3.1	**0.58**	**7.1**	0.82	16	0.55	12.5

Table 3.4: Ablation of CrossSDF with different components removed, on thick (Balloon Dog and Elephant) and thin (Alveolis and Coronaries) structures using aligned planes.

thin, complex structures with intricate branching patterns.

The symmetric difference loss we introduced effectively minimizes visual artifacts commonly encountered in cross-sectional reconstruction by focusing optimization on regions of classification disagreement. This approach, combined with our adaptive sampling strategy and hybrid encoding technique, enables CrossSDF to faithfully reconstruct fine details and preserve topological connectivity that would otherwise be lost or fragmented using traditional approaches (Barequet and Vaxman, 2009).

Through extensive experimental evaluation on both synthetic datasets and real-world examples, we have demonstrated that CrossSDF consistently outperforms both general-purpose point cloud reconstruction methods and specialized cross-section reconstruction algorithms. Its ability to produce continuous, topologically coherent surfaces from sparse, potentially non-parallel cross-sections represents a significant step forward in thin structure reconstruction.

However, our investigations also revealed an important limitation: when applied to medical data with inherent sparsity between imaging planes, the quality of the reconstruction remains fundamentally constrained by the accuracy of the initial 2D

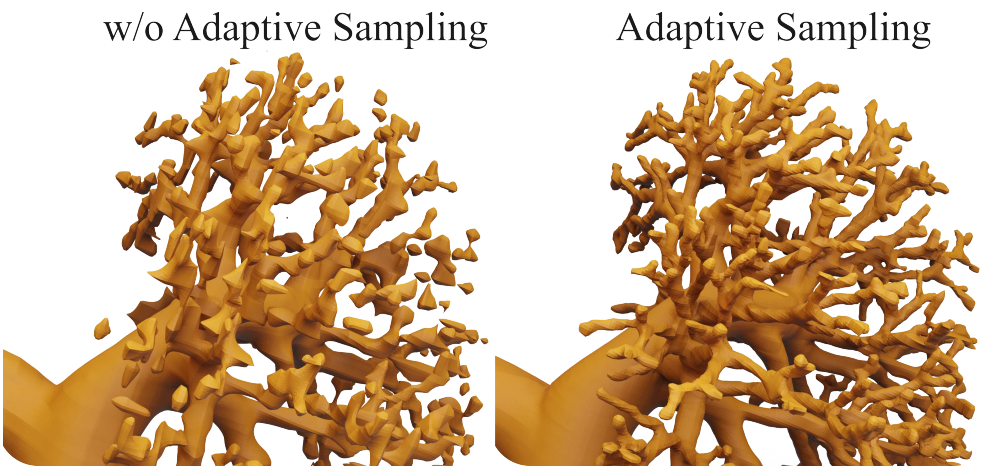


Figure 3.14: Ablation of our adaptive sampling approach for Alveolis reconstruction.

contour segmentation. This observation is particularly relevant in medical contexts such as CT and MRI scans, where thin structures like blood vessels present as sparse and incoherent within cross-sectional scan data (Antonelli et al., 2022).

This limitation points toward an important direction for future research: while CrossSDF operates on pre-segmented contours and thus depends on segmentation quality, a more integrated approach could be beneficial. This observation led to our development of VesselSDF (Chapter 4), which employs a two-stage framework that systematically separates vessel detection from geometric refinement, allowing each component to be optimized for its specific task while maintaining overall system coherence. Rather than requiring perfect initial segmentation, this approach handles both challenges within a unified pipeline specifically optimized for medical imaging data.

In the next chapter, we build upon these insights to develop VesselSDF, a unified framework specifically designed for vascular network reconstruction from sparse CT scans. By addressing the limitations identified in CrossSDF and incorporating regularization for handling the unique challenges of medical imaging data, we take the next step in our progressive exploration of neural SDF representations for increasingly specialized reconstruction tasks.

The deliberate progression from general object geometry in GeoGen to specialized thin structure reconstruction in CrossSDF, and ultimately to domain-specific medical applications, demonstrates the versatility and adaptability of neural SDF approaches. Each advancement has required careful consideration of domain-specific challenges while building upon the core strengths of SDF-based geometric representations, highlighting

the importance of tailored methodologies for different reconstruction scenarios.

Chapter 4

Distance Field Priors for Segmentation and Reconstruction

4.1 Introduction

The reconstruction of vascular networks from medical imaging data represents the final stage of our research into neural SDFs for complex geometry representation. Building upon the foundations established with GeoGen's general object geometry synthesis and CrossSDF's specialized approach to thin structure reconstruction from cross-sections, we now address the challenging domain of vascular network segmentation and reconstruction from sparse CT scan slices, a critical application area with direct clinical impact.

Accurate reconstruction of vascular networks remains a fundamental challenge in medical imaging and clinical diagnostics (Nam et al., 2021). Blood vessels, with their intricate branching patterns and varying diameters, play a crucial role in numerous pathologies, from coronary artery disease to tumor vasculature assessment (Zhao et al., 2019). For clinicians, the implications of accurate reconstruction are profound, enabling better surgical navigation, improving analysis of blood flow dynamics, and aiding in early detection of vascular abnormalities (Moccia et al., 2018).

Traditional segmentation approaches and recent deep learning methods typically focus on binary voxel classification (Chen et al., 2021; Shen et al., 2017), which fundamentally limits their ability to capture the continuous nature of vessel geometry (Isensee et al., 2021; Oktay et al., 2018). These methods often struggle with two critical challenges: maintaining structural coherence and generalizing beyond the training data (Guo, 2023). The first challenge manifests in regions where vessels branch or significantly

change direction between consecutive slices, resulting in fragmented or anatomically implausible reconstructions. The second stems from the inherent scarcity of annotated medical imaging datasets, leading to models that memorize specific vessel configurations rather than learning generalizable geometric principles (Guo, 2023). In contrast, our approach reformulates the problem using a two-stage framework that first predicts binary vessel occupancy - a coarse localization of vessel regions - and then transforms this discrete occupancy into a continuous signed distance field representation. The key distinction is that while traditional discrete voxel classification treats each voxel as an independent binary decision, our binary vessel occupancy serves as an intermediate representation that is subsequently refined through geometric constraints to produce smooth, continuous vessel surfaces. This separation allows us to leverage the robust segmentation capabilities of established architectures while addressing their geometric limitations through SDF-based refinement.

Our choice of a signed distance field (SDF) representation over binary occupancy is crucial to addressing these limitations. Unlike occupancy-based approaches that only encode local binary presence information, SDFs capture global surface information. A small perturbation in vessel surface affects SDF values across a substantial spatial region (Gropp et al., 2020a), making the representation inherently sensitive to fine vessel details that might be overlooked in binary occupancy models. Furthermore, SDFs implicitly encode surface normal information through their gradient, providing the network with critical geometric cues about vessel orientation and curvature that are entirely absent in occupancy grids (Chetan et al., 2023). This additional signal guides the network toward anatomically plausible reconstructions, particularly in ambiguous regions between sparse CT slices. Perhaps most importantly, SDFs possess inherent mathematical properties, such as the Eikonal constraint (gradient magnitude being 1 away from the surface), which serves as a powerful regularization mechanism that encourages the network to learn coherent, connected vessel structures even when working with incomplete data (Gropp et al., 2020a).

To address these challenges, we present VesselSDF, a novel two-stage framework that reformulates vessel reconstruction as a continuous geometric regression problem rather than discrete voxel classification. Unlike CrossSDF, which operates on pre-segmented contours, VesselSDF addresses both the segmentation and reconstruction challenges within a unified pipeline specifically optimized for vascular network reconstruction from sparse medical imagery. Our approach combines a specialized two-stage neural architecture with carefully designed geometric regularization techniques. In the

first stage, a 3D U-Net with attention gates captures multi-scale vessel features to predict binary vessel occupancy. The second stage transforms this occupancy into a properly scaled signed distance field through a refiner network with geometric constraints. This separation of concerns allows each network to focus on its specific task: detecting vessel presence and refining geometric accuracy, respectively.

VesselSDF introduces a distance-weighted Gaussian regularizer that adaptively enforces smoothness based on proximity to vessel surfaces. This innovative technique preserves fine vessel geometry near surface boundaries while ensuring smoothness in distant regions, effectively eliminating the floating geometric artifacts common in SDF-based reconstructions. Additional surface regularization further suppresses spurious vessel components, while our anisotropic Eikonal regularization accounts for different spatial resolutions along the axial dimension, addressing the unique challenges of medical imaging data. Through comprehensive evaluation on hepatic clinical datasets, we demonstrate that VesselSDF outperforms existing methods in both quantitative metrics and qualitative assessments. Our approach preserves vessel connectivity and fine geometric details while eliminating artifacts, enabling more reliable vascular analysis in clinical settings. This work represents an advancement in medical image segmentation and vascular reconstruction, with direct applications in clinical diagnostics, surgical planning, and vascular disease assessment.

Signed distance field (SDF) based representations offer a promising solution to these challenges due to their inherent ability to represent smooth, continuous surfaces. By encoding geometry through distance fields, SDFs naturally capture thin structures and maintain consistent spatial relationships (Alblas et al., 2022). Inspired by this, we present VesselSDF, a novel approach that leverages these geometric principles while addressing common SDF-based reconstruction artifacts. Our approach combines a specialized two-stage neural SDF encoder-decoder architecture with a novel Gaussian regularization technique that eliminates floating artifacts by adaptively enforcing smoothness based on distance from the vessel surface. To address the generalization challenge, we reformulate vessel reconstruction as continuous geometric regression rather than discrete voxel classification, allowing the model to learn underlying shape principles that transfer across different vessel configurations and anatomical variations. Our distance-weighted regularization further enhances generalization by encoding universal geometric priors about vessel continuity rather than memorizing specific patterns. This is complemented by an efficient SDF refinement strategy that ensures robust reconstruction even in challenging cases with significant inter-slice gaps.

In this chapter, we make the following contributions: (i) We introduce VesselSDF, a novel neural SDF architecture for vessel segmentation and reconstruction that enables accurate reconstruction of thin vascular structures from sparse CT slices. (ii) We propose an adaptive Gaussian regularization technique for SDF learning that applies distance-weighted smoothness constraints, effectively reducing floating geometric artifacts while preserving fine vessel details. (iii) We demonstrate superior reconstruction performance on challenging clinical vessel data containing thin vessels and complex branching patterns.

4.2 Related Work

4.2.1 Vasculature Segmentation

Deep learning approaches have largely superseded traditional methods in medical image segmentation. The foundational U-Net (Ronneberger et al., 2015) architecture established key principles through its encoder-decoder structure with skip connections. However, vessel segmentation remains challenging due to the presence of thin branching structures and high inter-subject variation (Wittmann et al., 2024). The segmentation of fine vascular structures presents unique difficulties, as minor vessels are frequently undetected or fragmented in standard frameworks. This challenge stems from their limited spatial extent and complex topology, creating a fundamental tension between preserving connectivity and maintaining accurate boundaries (Dima et al., 2023).

Recent innovations have tackled this problem through specialized architectures and loss formulations, such as topology-aware loss functions to penalize disconnections (Shit et al., 2021; Shi et al., 2024). Concurrently, architectures have evolved to better capture multi-scale vessel features, with transformer-based models (Zhou et al., 2023; Shaker et al., 2024) leveraging self-attention mechanisms to model global vessel connectivity while preserving convolutional inductive bias (Chen et al., 2021). Structure-aware networks explicitly incorporate vascular morphology through vessel-growing decoders in multi-task frameworks. Hybrid approaches that integrate convolution with attention pathways have proven particularly effective for delineating tortuous vessels (Li et al., 2023a). Structure-aware networks like explicitly incorporate vascular morphology through vessel-growing decoders in multi-task frameworks (Xu et al., 2025). Implicit neural representations have also been utilized (Alblas et al., 2022) to produce continuous, topologically consistent vessel surfaces, and domain-adaptive models have been

shown to generalize across imaging modalities (Wittmann et al., 2024). These developments demonstrate a shift toward specialized techniques that address the fundamental challenges of thin structure segmentation.

4.2.2 Sparse Slice Segmentation

A significant challenge in medical image segmentation is handling sparse slice data, which occurs due to radiation dose reduction in CT or time constraints in MRI (Zhang, 2024). This limited through-plane resolution creates discontinuities in vessel connectivity, particularly at bifurcations or directional changes (Yagis et al., 2024). While deep learning approaches have shown promise in general medical segmentation, accurately reconstructing continuous vessel structures under sparse sampling remains difficult, especially with complex vascular topologies. Recent approaches have explored 3D shape constraints (Alblas et al., 2022; Tetteh et al., 2020), but the problem persists for fine vessel structures.

4.2.3 Geometric Priors and Consistency

Recent efforts have integrated anatomical knowledge and geometric constraints to improve segmentation accuracy, especially when data are sparse or noisy. Topology-preserving approaches explicitly penalize disconnected or anatomically implausible predictions (Shit et al., 2021), improving continuity in thin vascular segments (Shit et al., 2021). For instance, by incorporating synthetic angiogram training data and cross-hair filters to better handle vessel continuity in large volumetric scans (Tetteh et al., 2020). Beyond segmentation, advances in neural rendering have shown strong potential for capturing thin vessel structures through continuous SDFs (Bogensperger et al., 2023; Ma et al., 2021a) and neural radiance fields (Maas et al., 2023). In contrast, our VesselSDF approach poses vessel segmentation as a continuous SDF regression problem with an adaptive Gaussian regularizer, which simultaneously preserves fine vessel geometry near surface boundaries while ensuring smoothness in distant regions and eliminating floating geometric artifacts.

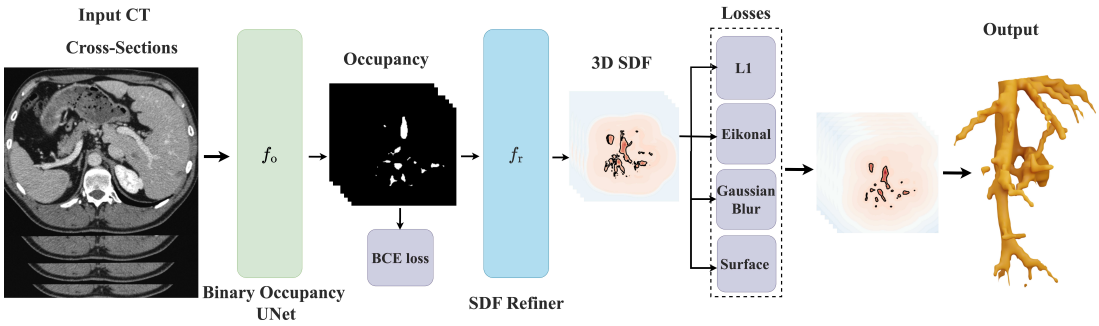


Figure 4.1: **Overview of VesselSDF -** our two-stage approach for vessel segmentation and reconstruction from CT scans. In the first stage, a 3D U-Net predicts a binary occupancy map. The second stage refines this occupancy into a signed distance field (SDF) using an additional 3D U-Net, guided by geometric regularization terms. The output 3D SDF, converted into a mesh, results in high-quality reconstructed vessels.

4.3 Method

4.3.1 Problem Statement

To highlight the benefits of SDF over binary occupancy, we designed a simple two-dimensional reconstruction experiment. We generate a sparse set of boundary points for a given shape (e.g., a circle or a star) and train two small neural networks on this limited data. One network predicts a signed distance function, enforced by the Eikonal constraint to maintain unity gradient norm, while the other learns binary occupancy classification without explicit distance information. During training, the SDF network receives boundary points with target values of zero and leverages random samples in surrounding regions to impose the gradient norm constraint. The occupancy network is only given the weak notion of being ”0.5 at the boundary” ,reflecting inside versus outside ambiguity. These properties allow the SDF to establish coherent spatial relationships between observed points, enabling the network to infer plausible shape completions even in regions far from explicit supervision. Specifically, the continuous distance values create a smooth gradient field that naturally guides interpolation between sparse boundary observations (Gropp et al., 2020a) , while the Eikonal constraint ensures that this interpolation follows physically plausible distance propagation patterns. The gradient information inherent in SDFs provides crucial geometric cues about local surface orientation and curvature, allowing the network to understand not just where surfaces are located, but also how they are oriented in 3D space. In contrast, binary classification offers no gradient information or distance awareness, forcing the model to

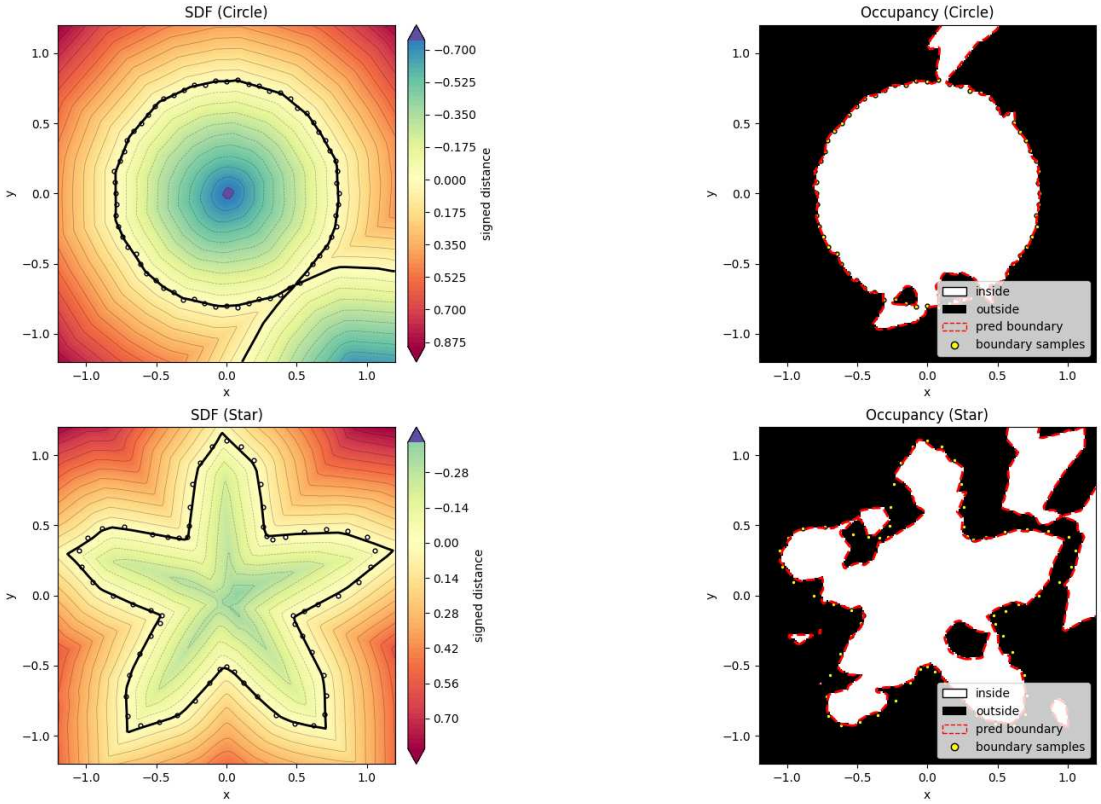


Figure 4.2: Reconstruction of a unit-circle from 64 noisy boundary samples (white dots) using a signed-distance network (left) versus a binary-occupancy network (right). The SDF model trained with an eikonal loss to enforce unit-gradient regularity yields a smooth, continuous field whose zero-level set (solid black) exactly recovers the true boundary. By contrast, the occupancy model's 0.5 decision boundary (dashed red) exhibits spurious disconnected regions, boundary jitter and topological errors, despite the simplicity of the geometry.

make uninformed decisions at each spatial location without understanding the broader geometric context. This fundamental limitation means that binary approaches must rely purely on local texture and intensity patterns, making them particularly vulnerable to noise and artifacts in medical imaging data where vessel boundaries may be poorly defined or interrupted by imaging artifacts.

Shape reconstruction from these sparse boundary measurements presents significant challenges in inferring coherent representations. As shown in Figure 4.2 for a simple circle, both approaches initially appear comparable. However, occupancy methods begin producing artifacts and disconnected regions when limited to sparse boundary information. These deficiencies become dramatically more apparent with intricate geometries, as demonstrated in our star-shaped example Figure 4.2. Here, the SDF model smoothly interpolates between sparse boundary points and accurately captures the narrow spokes, while the binary occupancy network fails to reconstruct thin regions, instead generating fragmented patches throughout the domain.

This performance difference stems from the SDF's continuous distance values, which provide crucial geometric guidance about shape extension from observed boundary points. The Eikonal constraint plays a vital role by enforcing consistent distance propagation throughout the domain, effectively creating a harmonic field that naturally interpolates between sparse observations. Furthermore, SDF gradients inherently encode surface normal information, providing the network with direct geometric cues about local surface orientation—information absent in binary occupancy representations. These properties allow the SDF to establish coherent spatial relationships between observed points, enabling the network to infer plausible shape completions even in regions far from explicit supervision. In contrast, binary classification offers no gradient information or distance awareness, forcing the model to make uninformed decisions at each spatial location Figure 4.2 further confirms that these limitations are inherent to the binary representation rather than a particular sampling pattern. Consequently, SDF-based reconstructions deliver more coherent geometry and excel at capturing intricate structures under sparse supervision, making them particularly well-suited for complex medical imaging tasks like vascular reconstruction.

Traditional binary vessel segmentation approaches face three critical limitations. First, the discrete nature of voxel-based representations results in jagged surface artifacts, which are particularly pronounced in thin vessels where the surface-to-volume ratio is high. Second, the substantial difference between in-plane resolution $\Delta x, \Delta y$ and slice thickness Δz creates anisotropic distortions that fragment vessel structures, espe-

cially at branching points. Third, while signed distance fields (SDFs) offer a promising direction for continuous surface representation, existing SDF-based methods often generate floating artifacts, i.e., disconnected surface fragments that degrade reconstruction quality.

We address these challenges via VesselSDF, a two-stage framework that systematically separates vessel segmentation from geometric reconstruction (see Figure 4.1 for an overview). VesselSDF takes a volumetric CT scan $\mathcal{V} \in \mathbb{R}^{D\times H\times W}$ as input, where D is the number of axial slices, and (H,W) are the in-plane dimensions. As output, it produces a continuous Signed Distance Field (SDF) $f_{\text{SDF}}(\mathbf{x};\theta_r)$, where $\mathbf{x}$ is a 3D spatial coordinate and θ_r are the model parameters, which implicitly defines the vessel surface through its zero-level set:

$$S := \left\{\mathbf{x} \in \mathbb{R}^3 \mid f_{\text{SDF}}(\mathbf{x};\theta_r) = 0\right\}. \tag{4.1}$$

4.3.2 Binary Occupancy (Stage 1)

The first stage of VesselSDF focuses on binary vessel segmentation via an occupancy prediction function $f_o(\mathbf{x};\theta_o)$. By focusing initially on the segmentation task, we establish a reliable starting point for subsequent geometric refinement. This occupancy predictor is implemented via a 3D encoder–decoder CNN-based U-Net architecture that captures multi-scale vessel features and outputs a voxelwise vessel occupancy probability.

Accurate segmentation of thin, branching vessels requires multi-scale feature representation. To achieve this, we integrate 3D attention gates (Oktay et al., 2018) such that at each level ℓ of the encoder–decoder structure, we combine gating feature maps $\mathbf{g}_\ell$ with skip-connection feature maps $\mathbf{h}_\ell$ through a learned attention mechanism:

$$\alpha_\ell = \psi\Big(\mathbf{W}_g\,\mathbf{g}_\ell \;+\; \mathbf{W}_h\,\mathbf{h}_\ell\Big), \tag{4.2}$$

where $\psi(\cdot)$ is a learned attention function that highlights salient vessel features, and $\mathbf{W}_g$ and $\mathbf{W}_h$ are trainable weight matrices. The resulting attention weights α_ℓ modulate the skip-connection features, ensuring focus on critical vessel regions and preserving fine vascular details that might otherwise be overlooked.

4.3.3 SDF Refinement (Stage 2)

The second stage transforms the binary occupancy into a correctly scaled SDF:

$$f_{\text{SDF}}(\mathbf{x};\theta_r) \;=\; f_r\big(\text{detach}\big(f_o(\mathbf{x};\theta_o)\big);\,\theta_r\big), \tag{4.3}$$

where $f_r(\cdot;\theta_r)$ is our SDF refiner network. The gradient detachment ensures that SDF-specific distance constraints do not interfere with the initial segmentation task. where $f_r(\cdot;\theta_r)$ is our SDF refiner network. The detach() operation is crucial for preventing gradient contamination between semantically incompatible representations. Without gradient detachment, several critical issues would arise: First, the Eikonal regularization term (Equation 4.7) enforces $\|\nabla f_{\text{SDF}}\| = 1$, meaning SDF gradients must have unit magnitude everywhere. If these gradients flow back to the occupancy network f_o, they would inappropriately constrain the occupancy gradients to unit magnitude, which is meaningless for binary classification where gradients should reflect probability changes, not geometric surface normals (Gropp et al., 2020b). Second, there is a fundamental semantic mismatch: f_o outputs probabilities in $[0,1]$ while f_{SDF} outputs signed distances in $(-\infty,+\infty)$. Gradients from the unbounded SDF domain would overwhelm the bounded occupancy gradients, potentially causing the occupancy network to output values outside $[0,1]$, breaking its probabilistic interpretation. Third, SDF gradients carry geometric information (surface normals and curvature) while occupancy gradients carry semantic information (vessel texture and intensity patterns). Mixing these would corrupt the occupancy network's ability to learn discriminative vessel features from CT intensities. Finally, our distance-weighted Gaussian regularizer (Equation 4.8) produces large gradients in regions far from vessel surfaces, which would propagate as noise back to the occupancy stage, destabilizing vessel detection. The detach() operation isolates these incompatible gradient domains while preserving the forward information flow from coarse occupancy to refined geometry.

This refinement is key to achieving smooth, accurate vessel reconstructions. Rather than treating each voxel independently, the SDF representation inherently couples neighboring predictions through distance-field properties. Each level operates at a different spatial resolution through successive downsampling and upsampling operations, allowing the network to simultaneously capture fine vessel details at high resolutions while modeling long-range dependencies and global vessel connectivity at lower resolutions. This hierarchical design allows the network to capture both fine details and global connectivity. Each level operates at a different spatial resolution through successive downsampling and upsampling operations, allowing the network to simultaneously capture fine vessel details at high resolutions while modeling long-range dependencies and global vessel connectivity at lower resolutions.

4.3.4 Optimization

VesselSDF's training loss combines supervised learning and geometric constraints to ensure continuous vessel reconstructions that are both accurate and topologically coherent, even for thin, branching vessels:

$$\mathcal{L} = \lambda_s \mathcal{L}_{\text{sdf}} + \lambda_o \mathcal{L}_{\text{occ}} + \lambda_e \mathcal{L}_{\text{eik}} + \lambda_g \mathcal{L}_{\text{gauss}} + \lambda_r \mathcal{L}_{\text{sur}}. \tag{4.4}$$

Supervised Terms. We use both SDF and occupancy supervision:

$$\mathcal{L}_{\text{sdf}} = \mathbb{E}_{\mathbf{x}\in\Omega}\Big| f_{\text{SDF}}(\mathbf{x}) - f^*_{\text{SDF}}(\mathbf{x})\Big|, \tag{4.5}$$

$$\mathcal{L}_{\text{occ}} = -\mathbb{E}_{\mathbf{x}\in\Omega}\Big[y \log\big(f_o(\mathbf{x})\big) + (1-y)\log\big(1-f_o(\mathbf{x})\big)\Big], \tag{4.6}$$

where $f_{\text{SDF}}(\mathbf{x})$ is our SDF prediction, $f^*_{\text{SDF}}(\mathbf{x})$ is the ground-truth SDF, $y \in \{0,1\}$ is the binary vessel label, and $\Omega \subset \mathbb{R}^3$ is the domain of 3D spatial coordinates in our training volume. Note that Ω includes the axial dimension as well, so these terms are computed fully in 3D (i.e., across neighboring slices), thereby ensuring consistent supervision throughout the volumetric scan.

Eikonal Regularization. To encourage smooth distance transitions, we enforce near-unit gradients (Gropp et al., 2020a):

$$\mathcal{L}_{\text{eik}} = \mathbb{E}_{\mathbf{x}\in\Omega}\Big(\big(\big(\partial_x f_{\text{SDF}}(\mathbf{x})\big)^2 + \big(\partial_y f_{\text{SDF}}(\mathbf{x})\big)^2 + \big(\gamma\partial_z f_{\text{SDF}}(\mathbf{x})\big)^2\big) - 1\Big)^2, \tag{4.7}$$

where $\gamma = \frac{\Delta z}{\Delta x}$ accounts for anisotropic voxel spacing along the axial dimension. By enforcing this constraint, the model more closely approximates the signed distance property, preventing large deviations in gradient magnitudes that can lead to geometric artifacts.

Distance-weighted Gaussian Regularization. While the Eikonal term enforces local smoothness in the gradient field, high-frequency noise can still persist in regions far from vessel boundaries. To further suppress such artifacts, we introduce a *distance-weighted Gaussian* regularizer:

$$\mathcal{L}_{\text{gauss}} = \mathbb{E}_{\mathbf{x}\in\Omega}\Big| f_{\text{SDF}}(\mathbf{x})\Big| \cdot \Big\| f_{\text{SDF}}(\mathbf{x}) - \mathcal{G}_\sigma\big(f_{\text{SDF}}(\mathbf{x})\big)\Big\|_2^2. \tag{4.8}$$

Here, $\mathcal{G}_\sigma$ denotes a 3D Gaussian blur operator with scalar standard deviation $\sigma > 0$. This term smooths the SDF *more aggressively* in regions where $|f_{\text{SDF}}(\mathbf{x})|$ is large (i.e., far from the vessel surface) and thus more prone to noisy fluctuations, while preserving fine details near the boundary (where $|f_{\text{SDF}}(\mathbf{x})| \approx 0$). This distance-adaptive mechanism

is key to achieving globally smooth reconstructions without over-smoothing critical vessel edges.

Surface Regularization. Finally, to suppress spurious or floating vessel components, we add:

$$\mathcal{L}_{\text{sur}} = \mathbb{E}_{\mathbf{x}\in\Omega} \exp\Big(-\beta \left| f_{\text{SDF}}(\mathbf{x}) \right|\Big), \tag{4.9}$$

where $\beta > 0$ is a hyperparameter that controls how strongly near-zero SDF values are penalized when there is no strong evidence of an actual surface. Larger β more aggressively suppresses weak, noisy boundaries.

4.4 Experiments

4.4.1 Implementation

VesselSDF is supervised on 3D SDF slices. For the occupancy stage, we use a 3D U-Net with attention gates, which outputs voxel-wise occupancy logits. A second 3D U-Net with two encoder-decoder levels refines detached occupancy into a final SDF. We train both stages jointly for 100 epochs without augmentations to preserve accurate SDF values using the Adam optimizer with a learning rate of 5×10^{-4}. We train our model with a batch size of 16 using whole volume images of size $512 \times 512 \times 16$. In our experiments, we set the Gaussian blur loss hyperparameter $\lambda_g = 0.01$ and the Eikonal weight $\lambda_e = 0.01$, maintaining consistency with the notation used throughout the paper. The complete set of loss coefficients for Equation 4.4 are: $\lambda_s = 1.0$ (SDF supervision), $\lambda_o = 1.0$ (occupancy supervision), $\lambda_e = 0.01$ (Eikonal regularization), $\lambda_g = 0.01$ (Gaussian regularization), and $\lambda_r = 0.1$ (surface regularization). These weights were determined through ablation studies on a validation subset, balancing reconstruction accuracy with geometric quality. Training uses Adam optimizer with learning rate 5×10^{-4}, batch size of 16 volumetric patches ($512 \times 512 \times 16$), and training for 100 epochs without data augmentation to preserve accurate SDF values. Additional hyperparameters including network architecture details, training schedules, and preprocessing parameters are comprehensively documented in Appendix C.1.3, Table C.1.

4.4.2 Datasets

We evaluate VesselSDF on two public hepatic vessel segmentation datasets. The Hepatic Vessels dataset (Medical Segmentation Decathlon - Task 08) (Antonelli et al., 2022) contains 303 hepatic veins CT scans with semi-automated vessel annotations obtained via level-set based region growing from expert-placed seed points. The ground truth includes major hepatic venous structures (portal and hepatic veins) and hepatic arteries visible in portal-phase CT. The 3D-IRCADb-01 dataset (IRCADb) (Soler et al., 2010) provides 20 contrast-enhanced CT volumes (512×512 pixels, 74–260 axial slices, in-plane spacing 0.56–0.87 mm, slice thickness 1.0–4.0 mm) with fully manual masks of the hepatic and portal venous trees. We convert those binary masks into training SDFs. All segmentations were produced by expert radiologists following standard Couinaud anatomy guidelines; however, the public release does not report inter-observer agreement, leaving residual uncertainty in the peripheral branches where contrast is weakest. Readers should therefore view vessel-level metrics on IRCADb as an upper-bound: even specialists disagree on the finest vessels, so a perfect score against this ground truth is unattainable.

4.4.3 Baselines

We compare to three state-of-the-art volumetric segmentation architectures which perform binary voxel classification. 3D-UNet (Çiçek et al., 2016), which is a standard encoder-decoder architecture for medical image segmentation. 3D SA-UNet (Guo, 2023) extends this architecture with spatial attention modules that adaptively weight feature responses based on their relevance to vessel structures, particularly beneficial for capturing thin vasculature. The nnU-Net (Isensee et al., 2021) contains two UNets, a low resolution and a higher resolution one. We compare to three state-of-the-art volumetric segmentation architectures which perform binary voxel classification: 3D-UNet, 3D SA-UNet, and nnU-Net. CrossSDF is not used as a baseline due to a fundamental difference in methodology: CrossSDF is a per-scene optimization approach that trains a neural network from scratch on each individual scan, overfitting to that specific geometry. This makes it impractical for medical applications where we need a generalizable model that can process new, unseen patient scans at inference time. In contrast, VesselSDF is trained once across multiple patients and can directly process new CT volumes without requiring scene-specific optimization. Additionally, CrossSDF requires manual extraction of 2D contours as preprocessing, whereas VesselSDF operates directly on raw

Hepatic Vessels Dataset					
Model	Dice↑	IoU↑	JD↑	CD↓	HD↓
VesselSDF (Ours)	**0.72**	**0.59**	**0.48**	**0.68**	**4.1**
nnU-Net (Isensee et al., 2021)	0.69	0.56	0.45	0.82	4.9
3D SA-UNet (Guo, 2023)	0.64	0.51	0.38	1.3	5.2
3D-UNet (Çiçek et al., 2016)	0.53	0.44	0.31	1.6	5.6
IRCADb Dataset (Portal and Hepatic Veins)					
VesselSDF (Ours)	**0.86**	**0.82**	**0.75**	**0.60**	**3.5**
nnU-Net (Isensee et al., 2021)	**0.86**	**0.82**	**0.75**	0.75	4.2
3D SA-UNet (Guo, 2023)	0.85	0.80	0.72	0.80	4.5
3D-UNet (Çiçek et al., 2016)	0.84	0.79	0.72	0.90	5.0

Table 4.1: **Quantitative Results on the Hepatic Vessels and IRCADb datasets.** Comparison of vessel reconstruction performance using different baselines. We report volume metrics (Dice Coefficient, Intersection over Union (IoU), and Jaccard similarity (JD)) and surface metrics (Chamfer distance (CD) $\times 100$ and Hausdorff Distance (HD)).

CT volumes. For clinical deployment, the ability to generalize across patients without per-scan optimization is essential, making the comparison with CrossSDF technically inappropriate despite both methods using SDF representations.

4.4.4 Evaluation

Our evaluation uses five metrics to assess 3D vessel reconstruction: Dice Score measures volumetric overlap; Volume IoU assesses 3D segmentation precision; Jaccard Distance (JD) quantifies topological similarity; and Chamfer Distance (CD) and Hausdorff Distance (HD) evaluate geometric accuracy through average and maximum surface distances.

4.5 Results

We present quantitative results in Table 4.1 and qualitative results in Figure 4.3 and in Figure 4.4 where we demonstrate VesselSDF's superior vessel and portal vein reconstructions on the Hepatic Vessels (Antonelli et al., 2022) and IRCADb (Soler et al., 2010) datasets. We also display reconstruction results on 3D segmentation of thick organs such as the liver in Figure 4.5 where the SDF-based reconstruction mitigates the notorious "laddering" artifacts that occur when marching cubes is applied to binary

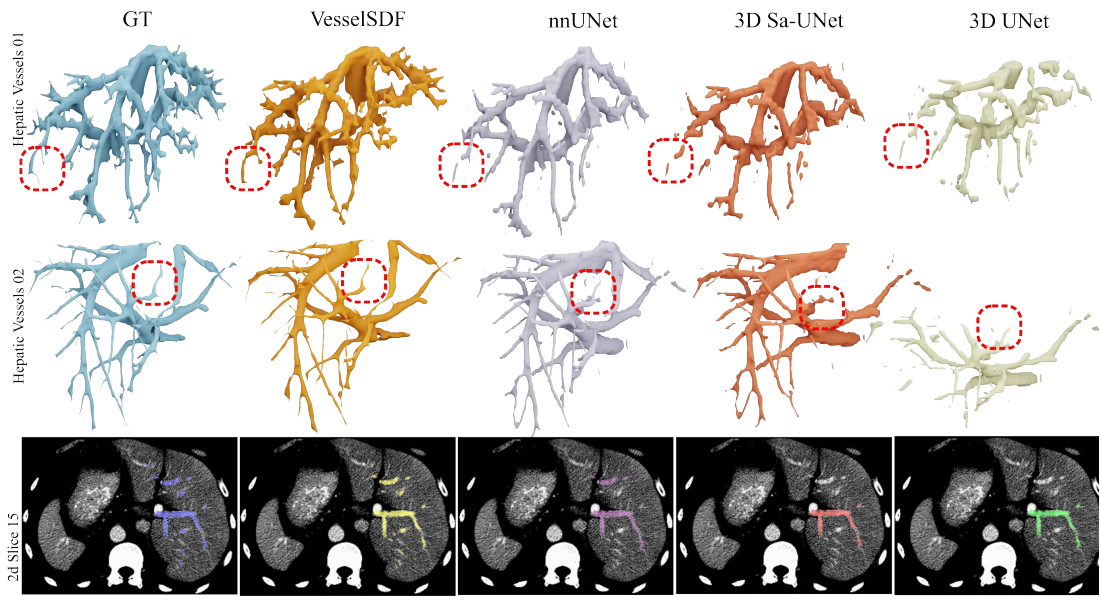


Figure 4.3: **Qualitative 3D Reconstruction Results on the Hepatic Vessels dataset.** The bottom row displays 2D slices highlighting the segmentation results.

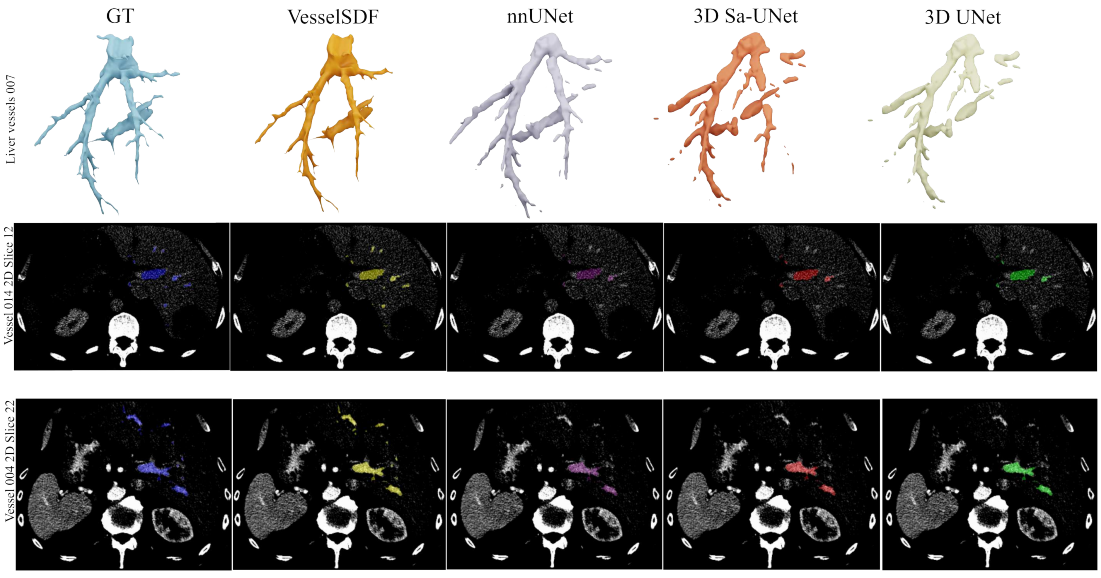


Figure 4.4: **Additional Qualitative 3D Reconstruction Results on the Hepatic Vessels dataset.** The bottom rows display 2D slices highlighting the segmentation results.

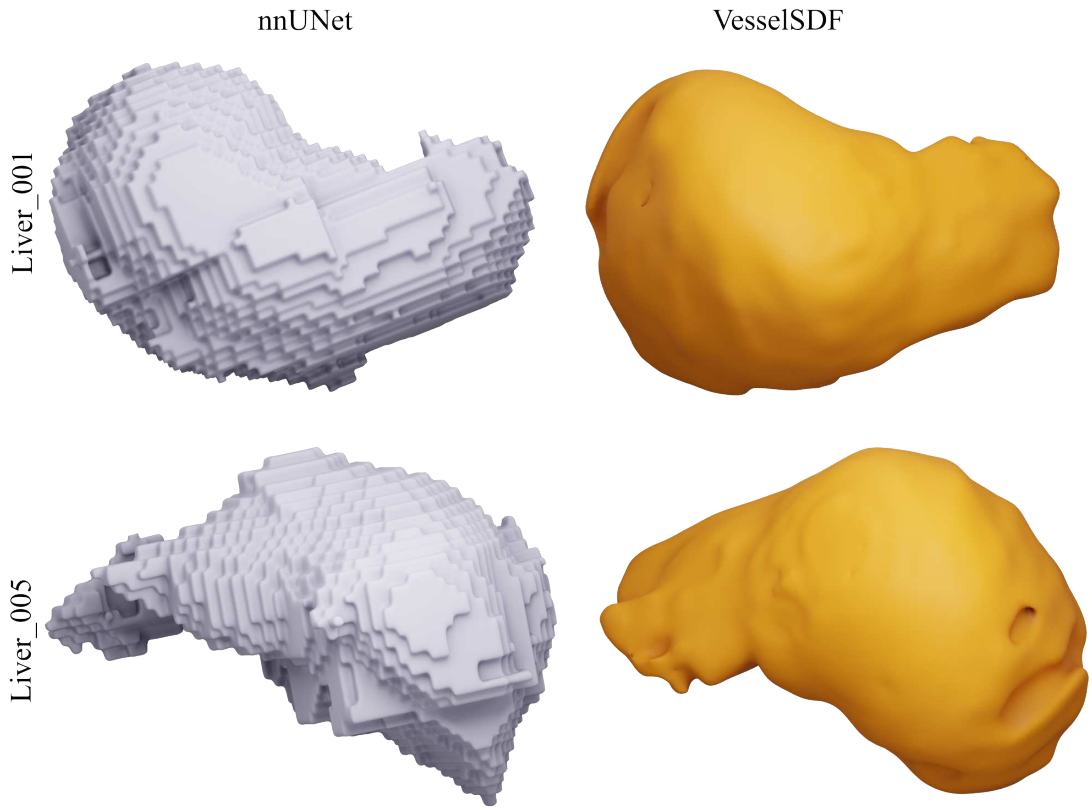


Figure 4.5: **Qualitative 3D Reconstruction results on the liver organ from the IRCARDb dataset (Soler et al., 2010).** Comparison of 3D liver reconstructions generated by marching cubes on (left) a binary voxel segmentation (*nnUNet*) versus (right) our (*VesselSDF*) approach. The binary voxel approach exhibits laddering artifacts caused by abrupt thresholding of the volumetric grid. By contrast, the SDF-based method encodes smooth transitions around the organ boundary, preventing these blocky, step-like features and yielding a more anatomically faithful surface.

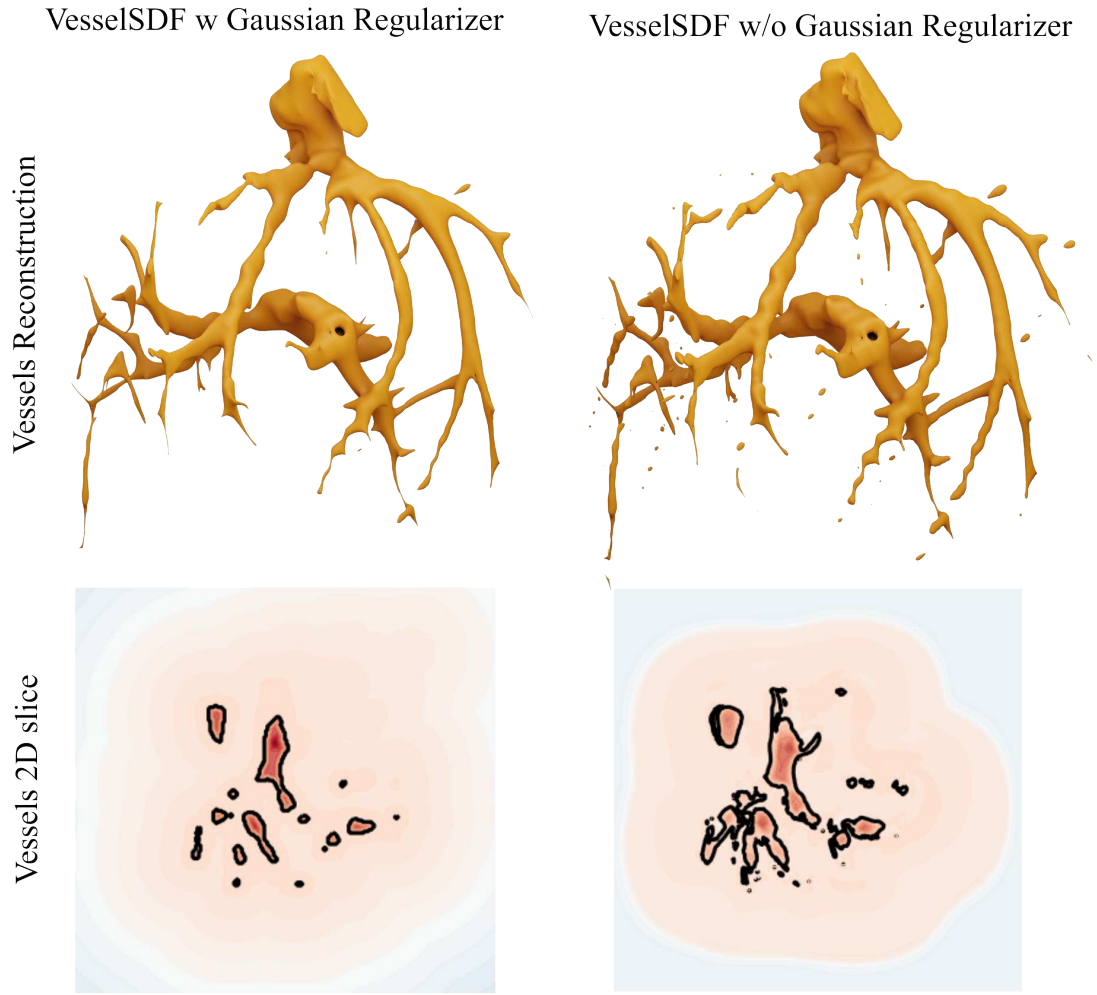


Figure 4.6: **Qualitative 3D reconstruction results on hepatic vessels from the Medical Decathlon dataset (Antonelli et al., 2022).** We compare VesselSDF reconstructions *with* and *without* a Gaussian regularization loss that blurs the predicted SDF and penalizes deviations in regions away from the vessel surface. When no Gaussian regularizer is applied, the SDF can capture noise from the input CT scan, resulting in floating artifacts around the vessels in both 2D slices and the 3D reconstruction. In contrast, the Gaussian regularizer smooths out spurious fluctuations while preserving fine vessel structures, leading to more anatomically coherent and artifact-free reconstructions.

voxel segmentations. By encoding continuous distance information, VesselSDF ensures a smooth and faithful surface representation that accurately captures subtle curvature and fine anatomical details, leading to visually and quantitatively superior surface metrics.

Table 4.2 quantifies the contribution of each component in VesselSDF through systematic ablation. The results reveal several key insights: Removing the binary occupancy stage (w/o Binary Occupancy) causes the most significant performance degradation, with Dice scores dropping from 0.72 to 0.65 and Jaccard Distance increasing from 0.48 to 0.56. This demonstrates that attempting to directly learn SDF values from raw CT data without the intermediate occupancy representation is fundamentally more challenging, as the network must simultaneously solve vessel detection and geometric refinement. Removing the SDF refinement stage (w/o SDF refinement) shows more modest but still significant degradation, with Dice scores decreasing to 0.69, confirming that while the occupancy stage provides essential vessel localization, the geometric refinement is crucial for accurate surface representation. The Gaussian regularization removal (w/o Gaussian Loss) maintains similar volume-overlap metrics (Dice = 0.72) but shows degraded surface quality (increased Chamfer Distance from 0.68 to 0.70), indicating that this component primarily addresses surface smoothness rather than overall segmentation accuracy. The full VesselSDF achieves optimal performance across all metrics, with the two-stage architecture providing the most substantial benefit, followed by SDF refinement for geometric accuracy, and Gaussian regularization for surface quality.

Our approach preserves thin vessels and complex branching structures more effectively than the binary voxel classification baselines (nnU-Net, 3D SA-UNet, and 3D UNet), yielding more complete and anatomically coherent reconstructions. Quantitatively, we outperform all methods on the challenging Hepatic Vessels dataset and obtain comparable performance according to volume-based metrics on IRCADb, where the numerical results are in general similar across all method, but we significantly outperform the baselines according to the surface-based metrics. Our ablation study (Table 4.2) and Figure 4.6 demonstrates the effectiveness of incorporating a Gaussian regularizer into our SDF pipeline. While this addition does not significantly alter volume or surface-based metrics, its qualitative benefits are substantial. The regularizer effectively smoothes noisy surfaces and prevents minor artifacts and noise in regions distant from the surface, resulting in visually smoother reconstructions with improved anatomical fidelity. This apparent contradiction occurs because conventional evaluation

metrics primarily assess global shape accuracy and boundary overlap, whereas the Gaussian regularizer addresses finer-scale variations that these aggregate measures fail to capture adequately.

4.5.1 Ablations

Table 4.2 quantifies the contribution of each component in VesselSDF. *w/o SDF refinement:* we remove the second SDF refiner, using only binary occupancy prediction without distance field computation or geometric regularizers. *w/o Binary Occupancy:* we directly predict the complete SDF (surface and isolines), bypassing our two-stage approach to test whether separating vessel detection from geometric refinement is beneficial. *w/o Gaussian Loss:* we remove the adaptive regularization (Eq. (4.8)), which results in similar Dice scores but introduces surface artifacts. The full VesselSDF achieves best performance across all metrics, with SDF refinement particularly enhancing vessel continuity as shown by the improved reconstruction metrics.

Model	Dice↑	IoU↑	JD↓	CD↓	HD↓
VesselSDF	**0.72**	**0.59**	**0.48**	**0.68**	**4.1**
VesselSDF w/o SDF refinement	0.69	0.57	0.52	0.70	4.4
VesselSDF w/o Binary Occupancy	0.65	0.55	0.56	0.75	4.6
VesselSDF w/o Gaussian Loss	**0.72**	**0.59**	**0.48**	0.70	4.3

Table 4.2: **Ablation Study on the Hepatic Vessels Dataset.** This study quantifies the contribution of each component in VesselSDF. Removing the SDF refiner or the binary occupancy module degrades reconstruction performance, while omitting the adaptive Gaussian loss introduces surface artifacts. The full VesselSDF achieves the best overall balance between volume accuracy and surface quality.

4.6 Conclusion and Discussion

In this thesis, we have presented a progressive development of neural SDF-based approaches for increasingly challenging 3D reconstruction tasks. Starting with GeoGen's foundation in general object geometry and advancing through CrossSDF's specialized handling of thin structures from cross-sections, we have culminated with VesselSDF , which directly addresses the critical application of vascular network reconstruction from medical imaging data.

VesselSDF represents a significant advancement in medical image segmentation and reconstruction. By reformulating vessel segmentation as a continuous SDF regression problem rather than discrete voxel classification, we have developed a framework that inherently captures the smooth, tubular geometry of blood vessels and their complex branching patterns. Our two-stage approach, which systematically separates binary vessel detection from geometric refinement, allows each network to focus on its specific task, resulting in more accurate and topologically consistent reconstructions.

The introduction of our distance-weighted Gaussian regularizer represents a key innovation that ensures smoothness in regions far from vessel surfaces while preserving precise geometry near surface boundaries. This adaptive approach effectively eliminates common SDF artifacts such as floating segments without compromising geometric fidelity. Similarly, our surface regularization suppresses spurious vessel components, while our anisotropic Eikonal regularization accounts for the different spatial resolutions along the axial dimension that are characteristic of medical imaging data.

Our comprehensive evaluation on challenging hepatic vessel datasets demonstrates VesselSDF's superior performance compared to state-of-the-art methods. Both quantitatively and qualitatively, our approach produces more complete and anatomically coherent reconstructions with fewer issues such as floating geometry and disconnected structures. These improvements have direct clinical implications, enabling more reliable vascular analysis for diagnostic and surgical applications.

The progression from GeoGen to CrossSDF and finally to VesselSDF illustrates the importance of domain-specific adaptations in neural SDF approaches. While each method builds upon the foundation of the previous work, the specific challenges of vascular reconstruction from sparse medical imagery required us to develop specialized techniques that address the unique characteristics of this data.Beyond vascular applications, our SDF-based framework could be extended to other anatomical structures where continuous geometry and topological consistency are critical. In pulmonary imaging, the method could reconstruct airway trees from CT scans, where the branching structure and smooth transitions are essential for understanding respiratory function and planning bronchoscopic procedures. The approach would be particularly valuable for reconstructing bile ducts in hepatic imaging, where thin, tortuous structures must be accurately represented for surgical planning in liver transplantation and tumor resection. In neuroimaging, the framework could potentially be adapted for white matter fiber tract reconstruction from diffusion tensor imaging, where maintaining connectivity and smooth trajectories is crucial for understanding brain connectivity

patterns. The distance-weighted regularization could also benefit reconstruction of other tubular structures such as ureters in urological imaging or intestinal structures in abdominal CT. However, each application would require domain-specific adaptations: different imaging modalities would necessitate adjusted preprocessing and potentially modified loss weightings, while structures with different geometric properties (e.g., highly tortuous bile ducts vs. relatively straight airways) might benefit from specialized regularization strategies. The two-stage architecture's separation of detection from geometric refinement makes it particularly suitable for these diverse applications, as the occupancy stage can be adapted to different imaging characteristics while the SDF refinement stage can incorporate structure-specific geometric priors.

Chapter 5

Discussion and Future Work

5.1 Summary of Contributions

This thesis has presented a systematic investigation into the use of neural signed distance fields (SDFs) for addressing the fundamental challenge of 3D reconstruction from incomplete or sparse observations. Through a series of three interconnected research projects: GeoGen, CrossSDF, and VesselSDF, we have developed novel methodologies that progressively address increasingly complex reconstruction scenarios, from general object geometry to thin branching structures in medical imaging. This chapter summarizes our key contributions, discusses limitations, and outlines potential promising directions for future research.

Our first contribution, GeoGen, introduced a novel approach to 3D-aware generative modeling using signed distance fields in place of traditional density representations. By reformulating the 3D representation as an SDF and imposing geometric constraints through an SDF depth map consistency loss, we achieved more accurate and detailed geometry than previous neural radiance field-based approaches. GeoGen demonstrated superior surface reconstruction quality across multiple datasets, including FFHQ faces, ShapeNet cars, and our synthetic human heads dataset, while maintaining high-quality image rendering capabilities. The key insight from this work was that neural SDFs provide a more geometrically coherent representation for surface reconstruction than density-based approaches, but require careful constraints to produce optimal results. An additional contribution of our paper is the release of a 360 degrees dataset of human avatars, which is the first of its kind to be released publicly with accurate camera parameters. However, GeoGen also revealed significant limitations when reconstructing fine details and thin structures such as hair and eyelashes, highlighting the need for

specialized approaches when dealing with these challenging features.

Building on the insights gained from GeoGen's limitations, our second contribution, CrossSDF, specifically targeted the reconstruction of thin structures from 2D planar cross-sections. We developed a hash-based neural reconstruction approach with three key innovations: (1) a symmetric difference loss that minimizes artifacts resulting from sparse cross-sectional data by focusing optimization on regions where target and predicted occupancy disagree; (2) an adaptive contour sampling strategy that ensures adequate representation of thin structures regardless of their cross-sectional area; and (3) a hybrid encoding architecture that combines detail-preserving hash encoding with Fourier features to reduce grid interpolation artifacts. Through extensive evaluations on both synthetic and real datasets, we demonstrated that CrossSDF achieves significant improvements in reconstructing thin structures and preserving topological coherence compared to existing methods. This work provided valuable insights into the specific challenges of thin structure reconstruction and established effective strategies for addressing them within an SDF framework.

While CrossSDF significantly advanced the state of the art for thin structure reconstruction from cross-sections, our experiments in applying it to medical imaging data revealed a critical limitation: the quality of the final reconstruction remained fundamentally bounded by the accuracy of the initial 2D contour segmentation. The reconstruction algorithm only works on a single scene input, which due to the complexity of highly branching structures, can be slow. This observation led to our third contribution, VesselSDF, a comprehensive two-stage framework specifically designed for vascular network reconstruction from sparse medical CT scans. VesselSDF addresses both segmentation and reconstruction challenges within a unified pipeline, treating vessel segmentation as a continuous SDF regression problem rather than discrete voxel classification. The framework employs a 3D U-Net for initial binary vessel occupancy prediction, followed by a refinement stage that transforms this occupancy into a correctly scaled SDF using geometric constraints. VesselSDF introduces an adaptive Gaussian regularization technique that enforces smoothness based on proximity to vessel surfaces, ensuring global coherence while preserving fine details at vessel boundaries. Additional innovations include anisotropic Eikonal regularization to account for different spatial resolutions along the axial dimension and a surface regularization term that suppresses floating artifacts. Through quantitative and qualitative evaluations on challenging clinical vessel datasets, we demonstrated that VesselSDF outperforms state-of-the-art methods in preserving vessel geometry and connectivity, even in the

presence of significant inter-slice gaps.

Collectively, these contributions establish a coherent progression in addressing the challenges of 3D reconstruction using neural SDFs. GeoGen demonstrated the fundamental advantages of SDF representations for general surface reconstruction; CrossSDF developed specialized techniques for preserving thin structures and ensuring topological consistency; and VesselSDF integrated segmentation and reconstruction into a unified framework optimized for complex vascular networks. Throughout this progression, we have consistently demonstrated how careful formulation of geometric constraints and regularization strategies can overcome the inherent limitations of neural network-based approaches when applied to challenging reconstruction tasks. Our work highlights the versatility and effectiveness of neural SDF representations while emphasizing the importance of domain-specific adaptations to address particular reconstruction challenges.

From a methodological perspective, a key insight that emerged across our projects is the critical importance of appropriate geometric regularization in neural SDF learning. In GeoGen, we found that an SDF depth map consistency loss was essential for producing high-quality surfaces; in CrossSDF, the symmetric difference loss proved crucial for handling sparse cross-sectional data; and in VesselSDF, the adaptive Gaussian regularizer played a central role in balancing detail preservation with global smoothness. These findings suggest that the success of neural SDF approaches depends not just on network architecture or representation choice, but on carefully designed optimization objectives that incorporate appropriate geometric priors.

Another significant insight from our research concerns the handling of data sparsity and anisotropy. Medical imaging data often exhibits significant disparity between in-plane and through-plane resolution, creating unique challenges for 3D reconstruction. Through our work in CrossSDF and VesselSDF, we have demonstrated effective strategies for addressing this anisotropy, including adaptive sampling approaches, geometric priors, and regularization techniques that account for different spatial resolutions along different dimensions. These contributions have important implications beyond the specific applications explored in this thesis, potentially benefiting any domain dealing with anisotropic or sparse volumetric data, such as LiDAR-based environmental reconstruction where point clouds exhibit highly variable sampling density (Shinde and Durbha, 2023; Vizzo et al., 2021), geological cross-sectional analysis for subsurface modeling (Shang et al., 2023; Hassanzadeh et al., 2022), autonomous vehicle perception systems using sparse sensor arrays (Wei et al., 2023), oceanographic bathymetry map-

ping with irregularly sampled sonar data (Irisawa and Iiyama, 2024), and archaeological site reconstruction from limited excavation profiles (Nobles and Roosevelt, 2021).

5.2 Limitations

Despite the advances presented in this thesis, several technical limitations warrant discussion and suggest avenues for future research. 3D reconstruction approaches including Sawdayee et al. (2023) and Walker et al. (2025) rely on supervised learning with explicit ground truth data, which fundamentally constrains their applicability in scenarios with limited annotation resources. Although VesselSDF's network is able to reconstruct vessels from sparse CT slices unlike binary voxel classification methods (Chen et al., 2021; Oktay et al., 2018), its optimization requires paired volume-segmentation data, limiting its application to datasets with expert annotations.

The computational complexity of our methods presents significant practical challenges for deployment. Our SDF-based approaches necessitate volumetric grid calculations at inference time, which scale cubically with resolution. CrossSDF's uses a hash grid encoding which is computationally expensive, a symmetric difference loss and VesselSDF's distance-weighted regularization, while effective for capturing detail, impose additional computational overhead during both training and inference that limits real-time applications. This computational burden makes our approaches less suitable for time-sensitive clinical scenarios or deployment on resource-constrained devices.

Handling topological complexity remains challenging with our current methods. Our implicit SDF representation lacks explicit topological constraints. Although we partially address this through Eikonal regularization and adaptive sampling strategies, we cannot guarantee topological correctness, particularly for fine vascular structures where the zero-level set can develop spurious components or breaks when the SDF gradient is poorly conditioned. This limitation is especially pronounced when reconstructing thin structures with complex branching patterns, such as those found in hepatic vessel networks.

Anisotropic sampling patterns present significant challenges to our methods. CrossSDF's symmetric difference-based loss operates between slice planes, making it sensitive to inter-slice resolution variations and potentially leading to aliasing artifacts when slice thickness significantly exceeds in-plane resolution. While our contour-aware sampling mitigates this effect to some degree, the method still struggles with highly anisotropic data commonly encountered in clinical settings, where in-plane resolution may be an

order of magnitude finer than slice thickness. Our methods lack mechanisms to progressively refine reconstructions and selectively denoise different frequency components of the signal. This fixed optimization approach can struggle to resolve ambiguities in complex vessel structures, particularly when working with incomplete or noisy observations.

Recent advances in the field have begun addressing several of these limitations. Zhang et al. (2024) work on graph attention-guided diffusion models demonstrates that diffusion-based approaches can better accommodate incomplete annotations through their inherent stochasticity. Their 2D diffusion architecture with dynamic conditioning operates on 2D slices rather than full 3D volumes and leverages modern diffusion sampling techniques to reduce denoising steps from 1000 to approximately 30, dramatically accelerating inference. Their graph-attention mechanisms can explicitly enforce connectivity through multiscale node attention, achieving significantly better connectivity scores than other state-of-the-art methods. Their local ensemble module effectively compensates for 3D anisotropy by integrating neighboring features, making it robust to varying slice thicknesses. These advancements suggest promising directions for improving SDF-based vessel reconstruction methods in future work.

5.3 Future Work

Building on the foundations established in this thesis, several promising directions for future research emerge. We summarize two identified directions in this section.

5.3.1 Generative CrossSDF for Shape Completion

Building on the foundations established in CrossSDF, a compelling direction for future research lies in developing a generative approach to the completion of 3D shapes from limited cross-sectional data, as recent methodologies have shown with different input modalities (Zhang et al., 2024). The current CrossSDF method excels at reconstructing thin structures from cross-sections by employing a novel symmetric difference loss, adaptive sampling, and hybrid encoding. However, it remains fundamentally deterministic and requires a sufficient number of cross-sections to produce accurate reconstructions. As demonstrated in the ablation study (Figure 3.12), performance deteriorates significantly when the number of input slices drops below 50, revealing a clear limitation in extremely sparse data scenarios.

A generative extension of CrossSDF could overcome this limitation by using a pre-trained 2D diffusion model to generate plausible complete slice geometries from highly limited observations. The technical approach would address several interconnected challenges. First, rather than training a 3D diffusion model from scratch, which would be prohibitively expensive given the dimensionality of volumetric data, we could adapt pre-trained 2D diffusion models to work with SDF representations. This would require bridging the substantial domain gap between natural images and signed distance fields, whose mathematical properties (continuous distance values with sign changes at boundaries) differ fundamentally from photographic content.

To implement this, we could develop a multi-stage architecture where each cross-sectional slice is independently processed through a Vision Transformer encoder (Ye et al., 2023), producing feature tokens that capture the local geometric information. These features would then be combined using cross-attention mechanisms to establish spatial relationships between slices, addressing the challenge of coherent 3D structure inference. The encoding process could incorporate truncated SDF values to better match the statistical distribution expected by pre-trained diffusion models, while a specialized decoding process would ensure the output maintains valid SDF properties, including the Eikonal constraint that CrossSDF enforces through regularization.

For extremely thin structures, such as the vasculature networks in the Alveolis and Pulmonary datasets (Antonelli et al., 2022) that present significant challenges for CrossSDF, the generative approach could be enhanced with specific priors about branching structures and anatomical connectivity. This could be implemented through conditional control mechanisms that incorporate domain knowledge from medical imaging. Specifically, a ControlNet inspired (Zhang et al., 2023) architecture could extract topological features from sparse cross-sections and use these as conditioning signals during the diffusion process. For instance, centerline extraction algorithms (Yang, 2011) could identify the main branches from available slices, providing skeletal priors that the diffusion model would be required to honor. Additionally, anatomical constraints could be encoded as text prompts. These text prompts would not only specify the anatomical region but also incorporate quantitative parameters derived from statistical analysis of similar structures in anatomical atlases, ensuring that generated completions adhere to biologically plausible configurations even in regions far from observed data.

The computational efficiency of this pipeline would critically depend on the optimized SDF computation. The generative model would be trained to minimize both

reconstruction fidelity to the available cross-sections and adherence to the geometric constraints that make SDFs effective 3D representations. Such a system would dramatically extend the capabilities of CrossSDF, enabling applications in scenarios where data acquisition is severely limited. By combining the geometric understanding of CrossSDF with the generative power of diffusion models, we could realize the promise of detailed, anatomically accurate 3D reconstruction, from extremely sparse observations.

5.3.2 Medial-Fields for Enhanced Reconstruction

An additional promising research direction involves leveraging the concept of the medial axis to improve connectivity and robustness in SDF-based 3D reconstructions. Although our methods, particularly CrossSDF and VesselSDF, have demonstrated strong performance in recovering surfaces and thin structures, ensuring global connectivity remains challenging when dealing with highly fragmented or sparse data. The medial axis, a set of points equidistant from boundaries (Rebain et al., 2021), offers an interpretable geometric representation that can guide the network toward consistent, well-connected shapes, especially in vascular or branching structures where isolated floaters or missed segments can severely degrade reconstruction quality.

This approach would involve designing a pipeline that extracts a discrete approximation of the medial axis directly from the network's predicted SDF by identifying regions where SDF gradients exhibit notable discontinuities. A learning-based approach could then simultaneously predict both an SDF and a corresponding medial field through a dual-headed architecture, with the medial field serving as an auxiliary signal to enforce connectivity. Areas where the predicted skeleton remains disconnected would trigger a penalty, nudging the SDF predictions toward more coherent surfaces. To ensure connectivity, the system would penalize scenarios where predicted SDF values are excessively large in regions that should be interior based on medial axis information. An iterative refinement process could then identify locations where the skeleton and distance field disagree, adjusting both the SDF and medial field predictions to reduce topological errors. This approach holds particular promise for medical imaging applications, where enforcing connectivity can ensure that vascular networks remain globally consistent even with significant gaps between cross-sectional scans.

5.3.3 Conclusion

In conclusion, the methodologies developed in this thesis provide a solid foundation for addressing complex 3D reconstruction challenges across various domains. By systematically addressing the limitations of traditional approaches and leveraging the powerful representational capabilities of neural signed distance fields, we have demonstrated significant advances in the reconstruction of general geometry, thin structures, and complex vascular networks. The proposed generative CrossSDF extension would enable high-fidelity shape completion from extremely sparse observations, while the medial-field approach would significantly enhance topological accuracy by enforcing structural connectivity constraints through dual SDF and medial axis representations. Together, these advancements would address critical limitations in current methods, particularly for thin structures like vasculature where disconnections and topological errors remain problematic. By combining generative capabilities with geometric guarantees, these future directions offer the potential to expand the applicability of neural SDF-based reconstruction to scenarios with even more challenging data constraints and complex anatomical structures.

Appendix A

Geometry-Aware Generative Modeling via Signed Distance Fields

A.1 Implementation Details

The foundation of our model relies on the official implementation of Enhanced Generative 3D Models (EG3D) (Chan et al., 2022). We utilized R1 regularization, assigning a gamma = 1 for the synthetic humans and FFHQ dataset based on the input image size of 512 x 512 and batch size of 32 across 8 v100 GPUs, following the same hyperparameter tuning of EG3D. For ShapeNet Cars, we adopted a gamma value of 0.3 based on the 128 x 128 resolution and batch size of 32 (Fu et al., 2022). Our model employs the same architecture as StyleGAN2 (Or-El et al., 2022), composed of a mapping network with 8 hidden layers, and output convolutions yielding 96 feature maps. Following the EG3D protocol, these are then reshaped into 3 planes of 256 x 256 x 32 (Chan et al., 2022).

A.1.1 GeoGen Training

During the initial training of GeoGen for the FFHQ and Synthetics dataset, the model was trained end-to-end, a process that necessitated unique handling of the SDF depth consistency loss. For the first 10,000 epochs, we set the beta value for the Laplace density distribution to 0.1 and refrained from making it learnable, as our end-to-end model would not have been able to learn the best beta value at this stage (Fu et al., 2022). This approach allowed the model to first learn the optimal geometry and SDF depth map. In contrast, StyleSDF had to introduce a two-stage training process precisely

because their pipeline was not trained end-to-end. They consistently used a learnable beta parameter for the Laplace density distribution throughout their training, as their method required more flexibility in the control of the SDF consistency loss.

The Laplace beta value plays a crucial role in the SDF network as it controls the shape of the Laplace distribution, influencing how the model penalizes deviations from the expected SDF values. A lower beta value produces a wider distribution, allowing for a larger spread of SDF values, and a higher beta value tightens the distribution, constraining the SDF values more strictly. This ability to control the distribution of SDF values enables fine-tuning of the model's sensitivity to inconsistencies in the SDF depth, a key aspect of the learning process. After the generator in our model showed improvement in rendering, depth maps, and underlying geometry, we activated the SDF constraint for depth map regularization and introduced the learnable beta parameter for the remaining 10,000 epochs. This allowed us to dynamically adapt the SDF consistency loss and fine-tune the model's learning of SDF depth.

Both EG3D and GeoGen models underwent training for 20,000 epochs for the FFHQ and Synthetics data, while for the ShapeNet dataset, training was conducted for 10,000 epochs. The batch size for all models was 18, with the discriminator's learning rate at 0.002 and the generator's at 0.0025. The training was carried out using 4 NVIDIA P100, while an RTX 2080 and RTX 4090 were used for inference during inversions and sample generation. Our end-to-end training approach, including the specific handling of the Laplace beta value, was central to our method's effectiveness in learning SDF depth. It allowed us to combine the flexibility needed in the early stages of learning with the precision required in later stages, reflecting a sophisticated understanding of the role that SDF plays in the generative process.

A.1.2 SDF, Color Network and Surface Rendering

The resulting embedding from the augmented spatial representation is fed into the SDF (Signed Distance Field) network. This network utilizes the embedded position to query the SDF value at a specific point, which gives precise information regarding the distance to the nearest surface within the 3D space. The understanding of these distances is crucial in the reconstruction of 3D objects, as it provides detailed insights into the geometry and the underlying complexities of the data being modeled.

Once the SDF network receives and processes the embedded position, the computed SDF values are further handled by the color network. This auxiliary network takes the

SDF values and translates them into the corresponding color values for the rendered 3D object. The direct utilization of SDF values as input for the color network establishes a coherent link between the geometric structure and visual appearance of the object. Both the SDF and color networks are built with a single hidden layer comprising 64 hidden units and leverage a soft plus activation function. This structure ensures smooth transitions and optimal gradient flow within the networks. For the transformation of the SDF into tangible density, a specific surface rendering technique has been applied. The sampling strategy is carefully chosen and tailored to different datasets, such as using 48 uniformly spaced samples and 48 importance samples per ray for the FFHQ dataset, and 64 of each for ShapeNet cars and Synthetics data.

In combination, these elements forge an intricate pipeline that integrates spatial features and coordinates, through a positional encoder, with the SDF and color networks. The methodology's architecture ensures a nuanced and true-to-life representation across a multitude of datasets. The implementation of a positional encoder has further enhanced the SDF network's capacity to grasp and replicate complex 3D geometries. The employment of SDF networks for surface rendering has led to a more sophisticated and resilient interpretation of various datasets.

A.1.3 Reconstruction of Pseudo Ground Truth Meshes

To reconstruct pseudo ground truth meshes we use Planar Prior Assisted PatchMatch Multi-View Stereo (ACMP) (Xu and Tao, 2020) and Poisson surface reconstruction (Kazhdan et al., 2006). Our pseudo ground-truth generation protocol involves two stages: first, ACMP (Planar Prior Assisted PatchMatch Multi-View Stereo) estimates depth maps from multiple viewpoints by embedding planar models into PatchMatch MVS via a probabilistic graphical model, enabling recovery of depth information even in low-textured regions. Second, Poisson surface reconstruction transforms the oriented point clouds into watertight meshes by casting surface reconstruction as a spatial Poisson problem, providing resilience to data noise while maintaining computational efficiency. It is important to note that evaluation errors may originate from both our reconstruction method and limitations in this pseudo ground-truth protocol, particularly in regions with complex geometry or insufficient multi-view coverage.

A.1.4 Results without Positional Encoder

Here we explore causes behind the collapse of the GeoGen model, specifically when trained without the aid of positional encoding in the context of Neural Radiance Fields (NeRF) and GAN training. The absence of positional encoding can lead to several critical issues (see Figure A.2). Firstly, in GAN training, the phenomenon of mode collapse becomes more pronounced. This is where the generator starts producing a limited variety of outputs, failing to capture the complex data distribution. Secondly, the intrinsic characteristics of NeRF, which rely heavily on precise spatial information to render 3D scenes accurately, are compromised without positional encoding. This results in the model's inability to effectively learn and represent high-frequency details, leading to a loss of detail and realism in the generated images. Lastly, positional encoding plays a vital role in stabilizing the training process by providing a more detailed and nuanced understanding of spatial relationships in the data. Its absence can result in unstable training dynamics, ultimately causing the model to collapse, particularly evident in our observations post epoch 11000. This highlights the essential nature of positional encoding in maintaining the stability and efficacy of models like GeoGen, especially in complex applications involving synthetic human images and 3D rendering.

A.2 Datasets

A.2.1 FFHQ and Rebalanced FFHQ

Our modeling framework originally utilized the in-the-wild version of the FFHQ dataset, a comprehensive collection of uncropped, original PNG human images sourced from Flickr, as documented by Karras et al. (2019) (Karras et al., 2019). To adapt these images for our purposes, we employed a sophisticated face detection and pose-extraction system (Chan et al., 2022), allowing us to determine the face area and label each image with its corresponding pose. The images were then cropped to approximate the dimensions of the original FFHQ dataset. We assumed fixed camera intrinsics for all images, with a focal length 4.26 times the image width, mimicking a standard portrait lens (Chan et al., 2022). After removing a small number of images where face detection proved unsuccessful, our final dataset comprised 69,957 images.

In our reporting, we include the 2D performance metrics of models trained on the Rebalanced FFHQ dataset, particularly focusing on the outcomes from NVIDIA-trained models. The Rebalanced FFHQ dataset, known for its broader diversity in facial

orientations, plays a crucial role in enhancing the model's capability to understand and replicate human facial features from various angles. This dataset is especially valuable for models that need to handle a wide range of facial geometries, such as those used in advanced image generation and recognition tasks.

While we present these metrics to showcase the performance improvements facilitated by the Rebalanced FFHQ dataset, it's important to note a limitation in the available data. NVIDIA, the entity responsible for training these models, has not provided detailed information regarding the number of epochs, specific training methodologies, or other intricate details of the training process. This lack of detailed training information could potentially impact the reproducibility and further optimization of these models.

Understanding the training duration (measured in epochs) and the specific methodologies employed is crucial for comprehensively evaluating a model's performance and for making informed comparisons with other models. The absence of this information leaves a gap in fully understanding how the Rebalanced FFHQ dataset impacts model performance compared to the original FFHQ dataset. Despite this, the reported 2D metrics still offer valuable insights into the enhanced capabilities of models trained on the Rebalanced FFHQ dataset, highlighting their improved proficiency in handling diverse facial features and orientations.

A.2.2 ShapeNet V1

We utilized the ShapeNet V1 Cars dataset for additional validation, rigorously comparing methodologies on a specific subset that includes 128 renderings of synthetic cars (Chang et al., 2015). This carefully curated dataset offers a robust platform for assessing performance across various viewing angles, enabling a comprehensive evaluation of 3D reconstruction and rendering techniques.

The ShapeNet dataset, as employed in our setup, builds on prior research and consists of 2,100 car images captured from 50 different perspectives (Chang et al., 2015). The multi-angle images provide an ideal scenario to analyze geometric consistency, shadow rendering, and surface texturing. Similar to the preprocessing applied to the FFHQ dataset, our approach to the ShapeNet data followed established protocols, maintaining the integrity and original characteristics of the images. Unlike other methodologies that might use augmentation or mirror images, we consciously chose not to apply these techniques to preserve the authenticity of the data and ensure a more accurate assessment of the models' performance (Chang et al., 2015).

A.2.3 Synthetic Humans

Our training model also harnessed our proprietary synthetic human dataset. This extensive collection encompasses 200,000 images, representing 20,000 unique identities. Each of these identities is portrayed from only 10 viewpoints, a stark contrast to the Rodin model where each identity was rendered from 300 diverse viewpoints (Wang et al., 2023). Despite the significant reduction in viewpoints per identity in our dataset, our model produces high quality outputs in terms of geometry and rendering (An et al., 2023). Our training approach proves that strong performance can be achieved with a more limited number of viewpoints.

A.2.4 Pivotal Tuning Inversion

In the context of our work with Pivotal Tuning Inversion (PTI), a specialized process to invert generative models like StyleGAN, we adopt a meticulous procedure to enhance the accuracy and efficiency of the inversion.

Initially, we utilize an off-the-shelf face detection solution to accurately locate and extract face regions within the test images. This process allows for precise alignment and ensures that the features of interest are adequately centered and scaled. The extracted regions are then cropped and resized to a consistent resolution of 512x512 pixels, facilitating uniform processing and analysis across different images.

Following this preprocessing stage, we implement the PTI methodology as delineated by Tov et al. (Tov et al., 2023). This approach consists of two main stages:

1. **Fine-Tuning of Generator Weights.** Subsequent to the initial latent code optimization, we proceed with an additional 500 iterations dedicated to fine-tuning the generator's weights. This phase is pivotal in refining the subtle details and enhancing the realism of the generated images. By adjusting the generator's parameters, we align the synthetic outputs more closely with the underlying distribution of the real data, improving both the fidelity and the perceptual quality of the inversions.

2. **Latent Code Optimization.** For the first 500 iterations, we focus on the optimization of the latent code, a compact representation within the model's latent space that encodes the essential features of the target image. Utilizing gradient-based optimization techniques, we iteratively refine the latent code to minimize the

discrepancy between the generated image and the target. This stage ensures that the inverted model captures the essential characteristics of the face.

The combination of these two stages offers a robust and precise inversion process, enabling us to generate high-quality, detailed images that faithfully represent the original inputs. The PTI methodology, by explicitly separating the optimization of the latent code and the fine-tuning of the generator, provides a nuanced control over the inversion process, yielding superior results in terms of both accuracy and visual appeal.

A.2.5 Justifying the Limitations in GAN Inversion

In the field of Generative Adversarial Networks (GANs), particularly with advanced models like EG3D, the accuracy of GAN inversion can be inconsistent. This inconsistency can be attributed to several factors, encompassing both the inherent characteristics of the generative model and the methodologies used in the inversion process.

Firstly, the architecture and complexity of the GAN model play a crucial role. A model with limitations in its design may not capture a broad range of features effectively, leading to challenges in accurately reproducing certain types of images during inversion. For example, if the model's architecture does not account for a wide variety of facial orientations, it may struggle with accurately inverting images that fall outside of its trained norm.

Additionally, the scope and diversity of the training data are critical. A model trained on a dataset with limited variety, such as one primarily consisting of front-facing images, may not perform well in inverting images with diverse or unusual orientations. The quality and diversity of the training data directly influence the model's ability to handle a wide range of inversion tasks.

Furthermore, the model's resolution and detail capabilities are also significant. Models that generate lower-resolution images or lack fine detail might fail to accurately capture nuances in the inversion process, resulting in less precise or realistic inversions.

On the side of inversion methodologies, the efficiency of the algorithm and its approach to navigating and manipulating the latent space of the GAN are key factors. The choice of loss functions and regularization techniques within the inversion method can greatly affect the match quality between the inverted image and the original. Computational constraints can also limit the effectiveness of more resource-intensive, yet potentially more accurate, inversion methods.

In summary, the limitations in GAN inversion accuracy can be attributed to a complex interplay of factors related to both the generative model's characteristics and the inversion techniques used. Understanding and addressing these factors is crucial for improving the accuracy and reliability of GAN inversions.

A.2.6 Evaluation Metrics

Evaluating the quality and performance of generated images is paramount in understanding the effectiveness of generative models. To this end, we employed the Fréchet Inception Distance (FID) and Kernel Inception Distance (KID), calculating these metrics for 50,000 generated images against all training images for both FFHQ and synthetic humans datasets. The calculations were performed using the implementation provided in the StyleGAN2 codebase (Karras et al., 2019), ensuring consistency with commonly accepted standards.

Our GeoGen model's KID scores were found to be 100 times lower than those of comparative models, an unexpected result that warrants careful consideration. One possible hypothesis for this abnormality might be an alignment of specific features or particularities in the convergence behavior during the training of our model. It could also be related to the choice of hyperparameters or the data preprocessing steps that were unique to our experiment. However, these hypotheses are subject to further investigation, and the exact reason behind the unusually low KID score remains an intriguing question for future research.

Alongside the 2D image quality evaluation, we also assessed 3D geometry comparisons, adopting the Efficient Geometry Aware 3D Network (EG3D) (Chan et al., 2022) for evaluation. Our GeoGen model showed promising results relative to the EG3D model, as indicated by these metrics, both in terms of 2D image quality and 3D Chamfer distance metrics. The overall evaluation paints a comprehensive picture of our model's capabilities, but the abnormally low KID score serves as a reminder that there may always be underlying complexities and subtleties that require further exploration and understanding.

A.2.7 3D Reconstruction Metrics

The assessment of 3D geometry is a critical aspect of our evaluation, as it reflects the ability of the generative models to faithfully reconstruct and represent the intricate geometric details of the subjects. Table 2 from the paper presents a comprehensive com-

parison of different 3D reconstruction metrics for generative models on ShapeNet *Cars* and Synthetic Human *Heads*. The selected metrics include Overall Chamfer Distance, Mean Squared Error (MSE), Hausdorff Distance (HD), Earth Mover's Distance, and Mean Surface Distance (MSD).

These metrics were chosen for their ability to capture various aspects of geometric fidelity. Chamfer Distance provides a measure of dissimilarity between two point sets, emphasizing both the precision and recall of the reconstructed surfaces. MSE offers insights into the mean differences between corresponding points, focusing on local accuracy. HD measures the maximum distance from a point in one set to the nearest point in the other set, highlighting global discrepancies. Earth Mover's Distance quantifies the minimum amount of work to transform one point set into the other, capturing overall distribution alignment. Lastly, MSD focuses on the mean distance between surfaces, reflecting surface smoothness and consistency.

In the process of evaluating these metrics, we scaled the generated and ground-truth meshes to fit within a unit sphere to ensure a consistent basis for comparison. We then randomly sampled 20,000 points from the meshes, repeating this process 20 times, in order to compute the mean and standard deviation of the metrics. This methodology allowed us to capture a comprehensive and statistically robust representation of the geometric quality, eliminating potential biases related to specific sampling patterns or scaling discrepancies.

The results, as shown in Table 2 of the main paper indicate that GeoGen demonstrates superior results, reflecting its ability to represent finer geometric details. The table also includes comparisons with GeoGen without SDF and DL constraints, allowing for an understanding of how specific components and constraints influence model performance. The best-performing methods for each dataset are highlighted in bold, striking a balance between quantitative performance and perceptual realism. The rigorous evaluation of these 3D metrics underscores the effectiveness of our approach and contributes to a nuanced understanding of generative modeling for complex geometric structures.

A.3 Additional Qualitative Results

in Figure 2.7 we present a comparison of synthetic human avatar meshes across EG3D (Chan et al., 2022) and GeoGen. It is qualitatively evident that our model, leveraging the capabilities of the Signed Distance Field (SDF) network with SDF depth

consistency loss, surpasses both EG3D and StyleSDF (as shown in the main paper) in reconstructing detailed facial features, including the ears, nose, hair, and eyes.

Additionally, we demonstrate the ability of the GeoGen model in 3D reconstruction on the ShapeNet cars dataset in Figure 2.9 and Figure A.1 where it successfully reproduces granular details on the surface of the cars. We also show inversion results in Figure A.3. This distinction is further highlighted by contrasting the rendering qualities of the generated synthetic samples from the EG3D and GeoGen models, displayed in Figure A.3, against some ground truth samples. Unlike the EG3D model (Chan et al., 2022), which exhibits a lack of granular details, our model's implementation of a more advanced SDF network, combined with robust SDF constraints and feature storage within a triplane, yields more precise and refined reconstructions. Thus, our approach consistently and effectively bridges the gap between visual perception and geometric representation, outperforming other techniques in 3D reconstruction fidelity. That is also visible in Figure 2.9 and Figure A.1 where GeoGen is able to better reconstruct the surface of synthetic faces using a GAN inversion technique (Chan et al., 2022).

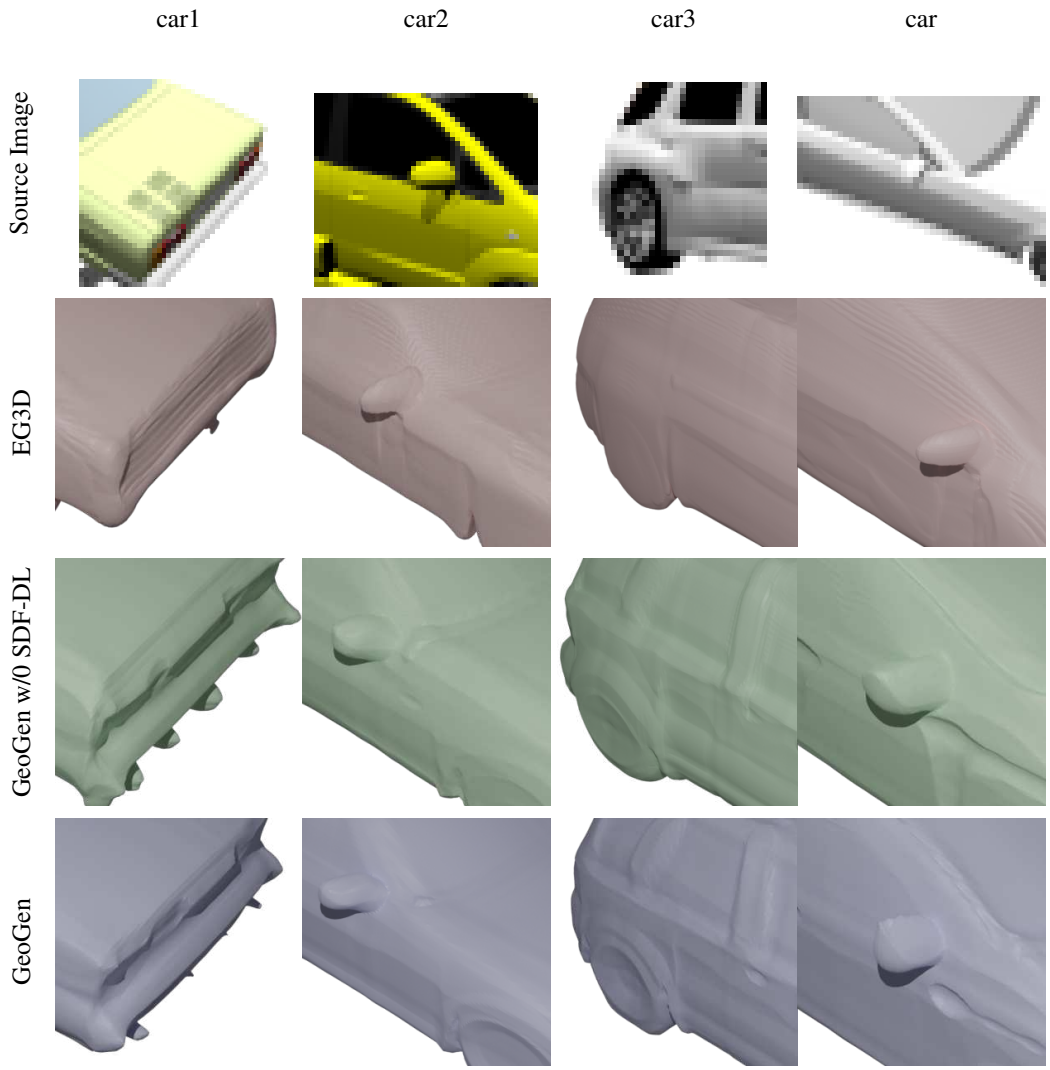


Figure A.1: A detailed comparison between EG3D and GeoGen in the context of ShapeNet cars inversion of meshes, emphasizing the differences in the geometric representation and rendering capabilities of both methods. The samples underscore the advanced efficacy of GeoGen in capturing and reconstructing intricate geometric details within the car models, even at granular levels. This superiority is attributed to the integration of the Signed Distance Field (SDF) network along with the SDF depth consistency loss within GeoGen's architecture. The SDF approach provides a continuous and differentiable representation of the car's surface, enabling more precise and robust alignment with the observed data. This contributes to better capturing of fine geometrical nuances and results in more accurate reconstructions. Conversely, the EG3D (Chan et al., 2022) method's rendered meshes reveal a deficiency in portraying granular details, leading to a more approximate and less nuanced depiction of the vehicles.

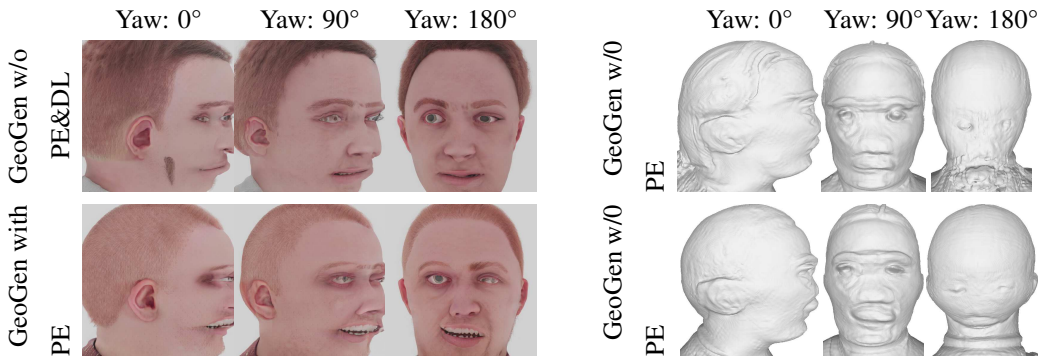


Figure A.2: This caption accompanies a series of synthetic images generated by the GeoGen model operating without a positional encoder. The figures on the left illustrate the model's output at different yaw angles, showcasing its ability to render facial features from various perspectives. On the right, the corresponding mesh structures are displayed, providing a deeper insight into the model's geometric rendering capabilities. These results were captured prior to the point of model collapse, highlighting the model's performance and limitations in the absence of positional encoding. This comparison not only demonstrates the visual output of the model but also underscores the critical role of positional encoding in maintaining structural integrity and realism in the generated images and meshes.

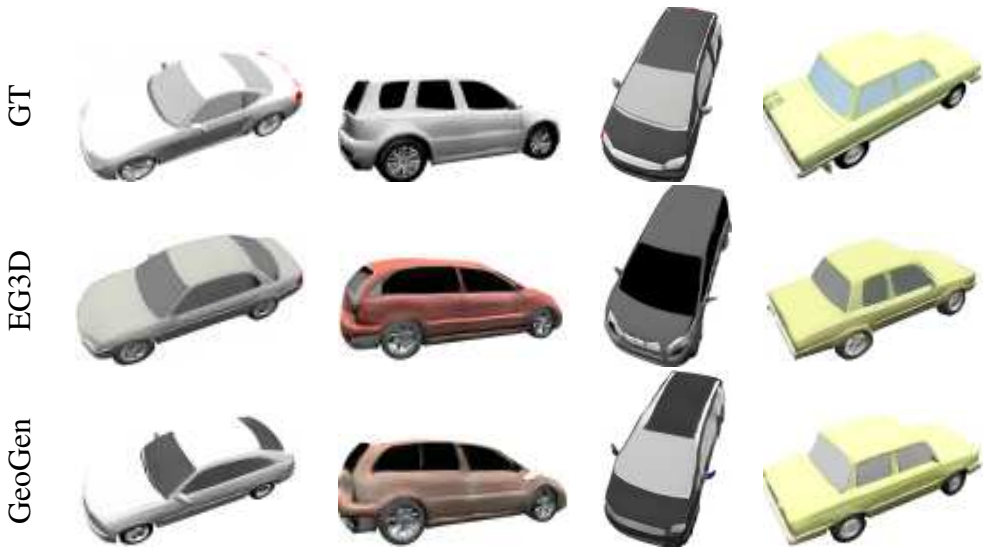


Figure A.3: Comparison of EG3D and GeoGen inversion results using held-out images from the ShapeNet Car test set. GeoGen results more closely resemble the input ground truth image (GT).

Appendix B

3D Reconstruction of Thin Structures From Cross-Sections

B.1 Additional Results

B.1.1 CT Scan Results

We present qualitative reconstructions of real anatomical structures with two vessel structures from the IRCADb-01 dataset (Soler et al., 2010) and two from the Medical Decathlon dataset (Antonelli et al., 2022), using pixel-level human-annotated segmentations. We evaluate CrossSDF on hepatic vessels from slices spaced 5mm apart, and liver vessels at 1mm apart. We display the results across different methods in Figure 3.11 which demonstrates our model's ability to handle clinically relevant challenges. As no ground truth 3D geometry is available, we created training and test splits by withholding $\approx 10\%$ of the slices and reported the 2D intersection-over-union (IoU) on the test set. To do so, we use integer division to compute a slice skipping frequency to withhold cross-sections. For example, for the hepatic vessel of patient 31, we remove every $61//10 = 6^{th}$ cross-section to compute 2D IoU. The quantitative results can be found in Table B.1.

Similarly to the synthetic scenes present in the main paper, OReX (Sawdayee et al., 2023) generally results in overly smooth geometry. In some cases, such as the thin and highly-branched contours of Patient 33's Hepatic Vessel, OReX fails to converge to a good reconstruction. Screened Poisson (Kazhdan and Hoppe, 2013) handles these cases more faithfully but is prone to breakages from the skipped slices, which is evident both qualitatively and quantitatively. Finally, we note CT scans with sparse slice thicknesses

are particularly difficult to reconstruct, highlighting our model's robustness in such scenarios (Ahmed et al., 2024).

Method	**Hepatic Vessel**		**Liver**	
	Patient 31 (61 Slices)	Patient 33 (71 Slices)	Patient 13 (149 Slices)	Patient 14 (75 Slices)
CrossSDF (Ours)	**0.79**	**0.63**	**0.83**	**0.88**
OReX (Sawdayee et al., 2023)	0.71	0.09	0.77	0.81
Poisson (Kazhdan and Hoppe, 2013)	0.24	0.55	0.21	0.84

Table B.1: Results on real CT scans. We report the average 2D IoU on a held out set of evaluation slices.

B.1.2 Additional Synthetic Results

B.1.2.1 Additional Comparisons

We display further qualitative comparisons in Figures 3.6 and 3.7. In Figure B.1 we also show results from Bermano et al. (2011), a non-neural method that can handle arbitrary cross-section orientations, and works by a barycentric blending of indicator functions in the cells of the planar arrangements of the cross-sections. As can be seen in the figure, their approach is susceptible to laddering artifacts, and fails to run on more complex scenes due to poor scalability.

B.1.2.2 No 2D SDF Labels

Although we supervise our model with 2D SDF labels, our symmetric difference loss prevents the neural field from learning each 2D SDF beyond its interior/exterior classification. For this reason, the optimization could be driven using indicator function labels (e.g. '1' for exterior and '-1' for interior). In Figure B.2 we present the result of doing so on the Balloon Dog scene (denoted as No 2D SDF Labels). We find that this leaves the predictable laddering artifacts along each contour. We attribute this to producing a discontinuous loss surface at the classification boundary. Conversely, our use of an L2 loss on 2D distance labels provides a smooth guide to this boundary.

B.1.2.2.1 Standard Hash-grid In addition to the baselines reported in the main paper, we conduct a further ablation by removing all of the key components of our CrossSDF approach; i.e. the adaptive encoding, Fourier features, and symmetric difference loss. We denote this as the 'Standard Hash-grid' baseline. The quantitative results for the

Method	Elephant	
	CD (↓)	**HD** (↓)
CrossSDF (full model)	**0.58**	**7.1**
Standard Hash-grid	1.51	15.0

Table B.2: Ablation of CrossSDF with all the components removed on the Elephant scene using aligned planes.

Elephant scene are presented in Table Table B.2. Removing these components leads to a degradation in performance, with the full model achieving the lowest CD and HD values.

B.1.3 Supplementary Video

A supplementary video is available on our website at `https://iamsalvatore.io/cross_sdf/`, showcasing side-by-side comparisons of reconstructions from our method with the baselines. The video also illustrates the optimization process for a different scene, providing deeper insights into our approach.

B.2 Implementation Details

We implemented our solution in PyTorch and used the accelerated tiny-cuda-nn (Müller, 2021) implementation for our hash-encoding module. All scenes were trained using a single 24GB NVIDIA RTX 4090. We train our network for a total of 500 epochs. The running time is commensurate with the complexity of the scene, taking as little as 20 minutes in the simplest cases (e.g. Elephant in Figure 3.7) to 5 hours in the most complex scenes (e.g. Alveolis in Figure 3.6). Our method does not have to create a (cubic-complexity) arrangement of cross-sectional planes, unlike (Bermano et al., 2011; Zou et al., 2015), and thus we scale better with the input complexity, only depending on the number of samples. Our network hyper-parameters are not tuned for each dataset. We use the same weights for losses and hash encoding resolution for all the datasets displayed. The mesh is extracted using marching cubes at 512^3 at the end of training.

B.2.1 Model Architecture

In Table B.3 we provide our hyper-parameters and network specifications. We use geometric initialization (Yariv et al., 2020) to start the model as a sphere at the beginning of training. The neural networks are trained using the ADAM optimizer. The learning rate is initialized to 5×10^{-4}, and reduced by a factor of 0.9 every 10 epochs.

Parameter	**Value**
SDF Network	
M_{SDF} layers	1
M_{SDF} width	256
M_{hash} layers	1
M_{hash} width	128
M_{RFF} layers	1
M_{RFF} width	128
RFF scale α	0.1
Softplus β	100
Hash-Grid & Fourier	
Grid levels	16
Min. grid res.	2^5
Max. grid res.	2^{10}
Grid feat. dim.	4
Dictionary size	2^{22}
RFF Gauss. var.	1.0
Sampling & Reg.	
Batch Size	2^{17}
Min. Surface β	100
Contour samples	50
Eikonal λ	1×10^{-3}
Min. surf. λ	5×10^{-2}
Weight decay	2×10^{-3}

Table B.3: Compact hyperparameter and network specifications.

B.2.2 Data Pre-Processing

B.2.2.1 Contouring

For our synthetic dataset, we extract contours (as polylines) by taking the mesh and cutting plane parameters, and use this to compute a set of contours that results from

cutting the mesh with the plane (e.g. each vertex has two connecting edges). For medical CT scans, human-annotated segmentations come in pixelized binary mask format, with slices along one axis. For each mask we generate a contour by running marching squares at 512 resolution.

B.2.3 Sampling

For the planar sampling Ω_{pl}, we employ a combination of uniform sampling, on-contour sampling, and fixed-radius sampling, as described below:

- **On-Contour Sampling:** (Ω_{on}): We uniformly sample 25 points along each edge of the contour.
- **Fixed-Radius Sampling:** For each on-contour sample $\mathbf{x} \in \Omega_{\text{on}}$, we take two additional samples, each located a fixed distance ε from the edge in the perpendicular direction. One sample is positioned outward, perpendicular to the contour, while the other is positioned inward.
- **Uniform Sampling:** In each slicing plane, we perform uniform sampling, generating 10,000 samples per plane.
- **Adaptive Contour Sampling:** We sample a bounding box around each contour until at least 50 interior samples are gathered. This ensures thin structures with low cross-sectional area are captured.

Following a similar approach to OReX, we compute labels for the newly generated samples at predetermined intervals of 0, 50, 100, 200, and 300 epochs as part of a pre-processing step. During each re-sampling phase, the fixed radius ε is gradually reduced over time with values $\varepsilon = 2^{-5}, 2^{-6}, 2^{-7}, 2^{-8}, 2^{-8}$. The label pre-processing step is efficient, taking less than three minutes for all tested scenes.

B.2.4 Baselines

For the point cloud reconstruction baseline methods, we use the on-contour samples Ω_{on} as a dense point cloud along each contour. For POCO (Boulch and Marlet, 2022), we use the model pre-trained on the ABC 10K dataset (Koch et al., 2019), which produced the best results relative to the alternative pre-trained models available.

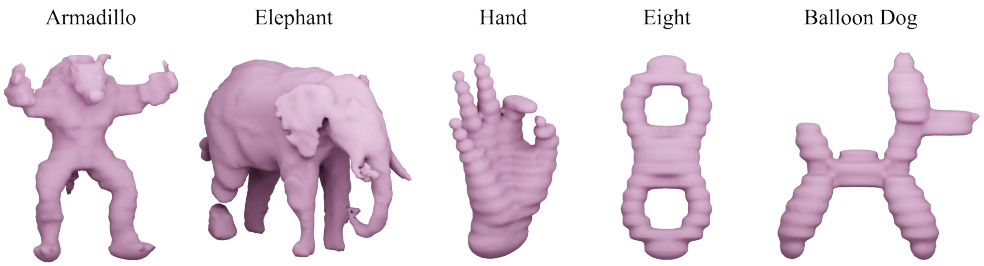


Figure B.1: Additional qualitative results from the method outlined in Bermano et al. (2011) using our synthetic dataset. Note the extreme laddering artifacts.

Figure B.2: Additional qualitative ablation results on the Balloon Dog scene.

Appendix C

Distance Field Priors for Vascular Network Reconstruction

C.1 Additional Results

C.1.1 Data Pre-Processing

C.1.1.1 CT Scans with Hounsfield Windowing and Vessel Enhancement

We provide further visualizations on real CT scans from both the Medical Segmentation Decathlon (Task 08) dataset (Antonelli et al., 2022) and the IRCADb dataset (Soler et al., 2010). In the main paper, we leveraged Hounsfield unit (HU) windowing to appropriately highlight the vessel intensity range (e.g. [100, 300] HU). Here, we offer side-by-side comparisons of raw slices (Figure C.1) vs. HU-windowed slices (Figure C.1). To enhance smaller branches, we applied the Frangi filter (Shi et al., 2023) which further improves the visibility of thin vessels.

C.1.1.2 Removing Empty Slices

As noted in the main paper, certain volumes contain slices without any annotated vessels, which can lead to ambiguities when training our Signed Distance Field (SDF) regressor. In such cases, including these empty slices can degrade performance by enforcing a large zero distance in places where no structure is present. Thus, we removed these vessel-free slices before generating SDF labels. By discarding slices with no vessel information, our approach avoids misleading supervision and maintains consistent geometry in the reconstructed vascular surface.

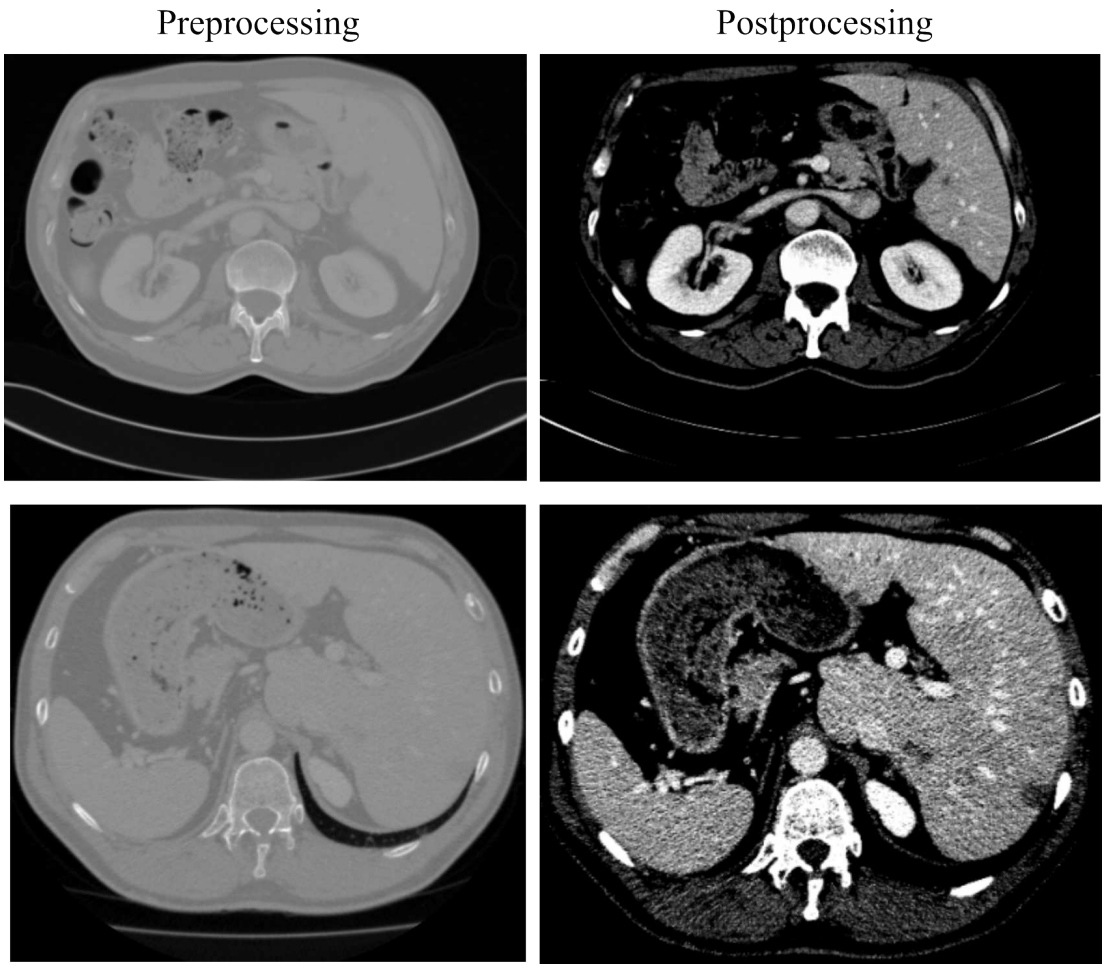


Figure C.1: Visualization of raw CT scan slices (left) and after highlighting vessels with frangi filter and Hounsefield unit with intensity ranges of (e.g. [100, 300].

C.1.2 Reconstruction Results

Figure 4.4 presents additional 3D reconstructions from the Medical Segmentation Decathlon dataset, highlighting the ability of our method to preserve thin hepatic branches. In particular, we note that VesselSDF avoids many of the floating or disconnected segments observed when directly training on binary occupancy alone. The Gaussian regularization (see main paper, Section 3.4) further mitigates noisy artifacts in regions far from vessel boundaries, preserving detail near the actual vessel surface.

We computed the 3D signed distance fields (SDF) from the binary segmentation volumes through a three-step process. First, we identified and removed axial slices without any annotated vessels to avoid supervising the network with purely empty slices. Second, we applied the Frangi vesselness filter (Shi et al., 2023) to each slice, using the recommended scale range for hepatic vessels ($\sigma \in [1.0]$). This aided the segmentation backbone by emphasizing fine tubular structures. Third, for each voxel within the remaining 3D volume, we computed the Euclidean distance to the nearest boundary voxel of the vessel, assigning negative distances to voxels identified as vessel interior and positive distances to voxels outside the vessel region.

C.1.3 Training Schedule and Hyperparameters

All models were trained using PyTorch with the Adam optimizer. We summarize our key hyperparameters in Table C.1. Unless otherwise noted, these values remain consistent for all experiments in both the Decathlon and IRCADb data:

C.1.4 Architecture

Our VesselSDF model employs a two-stage architecture as described in the main paper. The first stage, a 3D U-Net with attention gates, produces a binary occupancy estimate. The second stage refines this to a continuous SDF. We designed the second network with two encoder-decoder blocks and skip connections, detaching the gradient flow from the occupancy stage to prevent the SDF constraints from dominating the binary segmentation. For each volume, we extracted overlapping 3D patches of size $512 \times 512 \times 16$ for both training and inference.

Parameter	Value
Model Architecture	
3D U-Net depth	4 levels
SDF Encoder Decoder	4 levels
Attention Gate Type	3D Spatial + Channel
SDF extraction resolution	512^3
Optimization	
Learning rate	5×10^{-4}
Batch size	16 (volumetric patches)
Loss Coefficients	
λ_s (SDF loss)	1.0
λ_o (Occupancy)	1.0
λ_e (Eikonal)	0.01
λ_g (Gaussian)	0.01
λ_r (Surface)	0.1
Preprocessing	
Frangi filter scales	$[1.0]$

Table C.1: Summary of VesselSDF Hyperparameters and Architecture.

C.1.5 Inference

At inference time, we run both stages sequentially on the entire volume. The final SDF is transformed into a mesh via Marching Cubes at an isovalue of 0.0 at a resolution of 512^3. This approach enables us to maintain high-quality surface details while reconstructing continuous vessel structures across the entire volume. The resulting mesh exhibits smooth, anatomically plausible vessel geometry with well-preserved branching structures.

C.2 Ablation Studies

Our ablation studies, presented in Table 2 of the main paper, quantify the contribution of each component in the VesselSDF framework. When removing the SDF refinement stage (w/o SDF refinement), we observed a significant decrease in performance across multiple metrics, with Dice scores dropping from 0.72 to 0.69 and Jaccard Distance increasing from 0.48 to 0.52. This confirms that the geometric refinement provided by our SDF stage is essential for capturing accurate vessel topology.

Similarly, bypassing the binary occupancy prediction (w/o Binary Occupancy) led to even greater performance degradation, with Dice scores falling to 0.65 and Chamfer Distance increasing to 0.75. This validates our two-stage approach, showing that separating vessel detection from geometric refinement significantly improves reconstruction quality. Notably, while removing the Gaussian regularization (w/o Gaussian Loss) maintained similar Dice scores (0.72), it introduced unwanted surface artifacts as indicated by the increased Chamfer Distance (0.70 vs. 0.68). These results demonstrate that our adaptive Gaussian regularization effectively suppresses noise in regions distant from vessel surfaces while preserving fine details near the boundary.

The full VesselSDF model achieved the best performance across all metrics, with SDF refinement particularly enhancing vessel continuity as shown by the improved reconstruction metrics. This comprehensive ablation confirms that each component of our architecture contributes meaningfully to the final reconstruction quality, with particular emphasis on the importance of our two-stage approach and adaptive regularization strategy.

C.3 Evaluation Metrics

Our evaluation framework incorporates a diverse set of metrics designed to comprehensively assess the quality of vascular reconstructions. For thorough evaluation, we employ both volumetric and surface-based measures that collectively capture different aspects of reconstruction fidelity.

In terms of volumetric assessment, we first convert the continuous signed distance field (SDF) to a binary volumetric representation by thresholding at zero, where negative values indicate vessel interior and positive values represent the exterior. This conversion is necessary for comparison with ground truth binary annotations and for computing overlap-based metrics.

$$\text{Dice}(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|}, \tag{C.1}$$

where X and Y represent the predicted and ground truth volumes, respectively. The Dice coefficient (Equation (C.1)) serves as our primary overlap measure, quantifying the similarity between predicted and ground truth vessels. We further complement this with the Intersection over Union (IoU),

$$\mathrm{IoU}(X,Y) = \frac{|X \cap Y|}{|X \cup Y|}, \tag{C.2}$$

which offers a stricter assessment of volumetric overlap. Additionally, we employ the Jaccard Distance (JD) to evaluate topological coherence between the predicted and ground truth vessel networks. While not always formally defined, it can be expressed as

$$\mathrm{JD}(X,Y) = 1 - \mathrm{IoU}(X,Y),$$

with lower values indicating superior preservation of vessel topology and connectivity.

Beyond volumetric analysis, we place particular emphasis on surface-based metrics that directly evaluate the geometric accuracy of the reconstructed vessels. The Chamfer Distance (CD) computes the average bidirectional point-to-surface distance between the predicted and ground truth vessel surfaces:

$$\mathrm{CD}(S_1,S_2) = \frac{1}{|S_1|}\sum_{x\in S_1}\min_{y\in S_2}\|x-y\|_2 \; + \; \frac{1}{|S_2|}\sum_{y\in S_2}\min_{x\in S_1}\|y-x\|_2, \tag{C.3}$$

where S_1 and S_2 are point sets sampled from the predicted and ground truth surfaces, respectively. This provides a global assessment of surface reconstruction accuracy. We further incorporate the Hausdorff Distance (HD),

$$\mathrm{HD}(S_1,S_2) = \max\Big\{\max_{x\in S_1}\min_{y\in S_2}\|x-y\|_2,\ \max_{y\in S_2}\min_{x\in S_1}\|y-x\|_2\Big\}, \tag{C.4}$$

which captures the maximum deviation between surfaces and is particularly effective at penalizing structural discontinuities such as disconnected vessel fragments.

The combination of these metrics allows us to evaluate vessel reconstruction quality thoroughly. While volumetric metrics provide insight into overall segmentation accuracy, surface metrics directly assess the geometric fidelity that is crucial for clinical applications such as surgical planning and hemodynamic simulation. As demonstrated in Table 1 of the main paper, *VesselSDF* achieves especially notable improvements in surface-based metrics, confirming its effectiveness in preserving the detailed geometry and connectivity of thin vascular structures. This comprehensive evaluation approach provides strong evidence for the structural coherence and anatomical plausibility of our reconstructions, especially when compared to existing methods that struggle with thin vessel branches and complex topologies.

Bibliography

Ahmed, M., Garzanich, M., Melaragno, L. E., Nyirjesy, S., Windheim, N. V., Marquardt, M., Luttrull, M., Quails, N., and VanKoevering, K. K. (2024). Exploring ct pixel and voxel size effect on anatomic modeling in mandibular reconstruction. *3D Printing in Medicine*.

Al Moussawi, A., Galusinski, C., and Nguyen, C. (2015). 3d reconstruction of blood vessels. *Engineering with computers*.

Alblas, D., Brune, C., Yeung, K. K., and Wolterink, J. M. (2022). Going off-grid: Continuous implicit neural representations for 3d vascular modeling. In *Statistical Atlases and Computational Models of the Heart*.

Alt, H. and Guibas, L. J. (2000). Discrete geometric shapes: Matching, interpolation, and approximation. In *Handbook of computational geometry - Chapter 3*.

An, S., Xu, H., Shi, Y., Song, G., Ogras, U. Y., and Luo, L. (2023). Panohead: Geometry-aware 3d full-head synthesis in 360°. In *Conference on Computer Vision and Pattern Recognition*.

Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., Litjens, G., Menze, B., Ronneberger, O., Summers, R. M., et al. (2022). The medical segmentation decathlon. *Nature Communications*.

Atzmon, M. and Lipman, Y. (2020). Sal: Sign agnostic learning of shapes from raw data. In *Computer Vision and Pattern Recognition*.

Bajaj, C. L., Coyle, E. J., and Lin, K.-N. (1996). Arbitrary topology shape reconstruction from planar cross sections. *Graphical models and image processing*.

Barequet, G. and Sharir, M. (1994). Piecewise-linear interpolation between polygonal slices. In *Annual symposium on Computational geometry*.

Barequet, G. and Vaxman, A. (2009). Reconstruction of multi-label domains from partial planar cross-sections. *Computer Graphics Forum*.

Bermano, A., Vaxman, A., and Gotsman, C. (2011). Online reconstruction of 3d objects from arbitrary cross-sections. *ACM Transactions on Graphics*.

Bińkowski, M., Sutherland, D. J., Arbel, M., and Gretton, A. (2018). Demystifying mmd gans. In *International Conference on Learning Representations*.

Bogensperger, L., Narnhofer, D., Ilic, F., and Pock, T. (2023). Score-based generative models for medical image segmentation using signed distance functions. In *DAGM*.

Boissonnat, J.-D. and Memari, P. (2007). Shape reconstruction from unorganized cross-sections. In *Symposium on geometry processing*.

Boulch, A. and Marlet, R. (2022). Poco: Point convolution for surface reconstruction. In *CVPR*.

Castellan, C. and Dastarac, D. (2000). Use of computed tomography slices 3d-reconstruction as a powerful tool to improve manufacturing processes on aeroengine components. In *AIP Conference Proceedings*. American Institute of Physics.

Champetier, J., Le Bas, J.-F., Haouari, H., and Chaffanjon, P. (1992). Magnetic resonance imaging of the liver by oblique sections. *Surgical and Radiologic Anatomy*.

Chan, E. R., Lin, C. Z., Chan, M. A., Nagano, K., Pan, B., De Mello, S., Gallo, O., Guibas, L. J., Tremblay, J., Khamis, S., et al. (2022). Efficient geometry-aware 3d generative adversarial networks. In *Conference on Computer Vision and Pattern Recognition*.

Chan, E. R., Monteiro, M., Kellnhofer, P., Wu, J., and Wetzstein, G. (2021). pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis. In *Conference on Computer Vision and Pattern Recognition*.

Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al. (2015). Shapenet: An information-rich 3d model repository. *arXiv:1512.03012*.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., and Zhou, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. In *MICCAI*.

Chen, Z. and Zhang, H. (2019). Learning implicit fields for generative shape modeling. In *Conference on Computer Vision and Pattern Recognition*.

Chetan, A., Yang, G., Wang, Z., Marschner, S., and Hariharan, B. (2023). Accurate differential operators for hybrid neural fields. *Arxiv*.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). 3d u-net: learning dense volumetric segmentation from sparse annotation. In *MICCAI*.

Deng, J., Guo, J., Xue, N., and Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In *Conference on Computer Vision and Pattern Recognition*.

Deng, Y., Yang, J., Xiang, J., and Tong, X. (2022). Gram: Generative radiance manifolds for 3d-aware image generation. In *Conference on Computer Vision and Pattern Recognition*.

Dhariwal, P. and Nichol, A. (2021). Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*.

Dima, A., Zimmer, V., Menten, M., and et al. (2023). 3d arterial segmentation via single 2d projections and depth supervision in contrast-enhanced ct images. In *MICCAI*.

Erler, P., Guerrero, P., Ohrhallinger, S., Mitra, N. J., and Wimmer, M. (2020). Points2surf learning implicit surfaces from point clouds. In *ECCV*.

Esposito, S., Rebain, D., Onken, A., Li, C., and Aodha, O. M. (2025). Vesselsdf: Distance fields priors for vascular network reconstruction. *arxiv*.

Esposito, S., Xu, Q., Kania, K., Hewitt, C., Mariotti, O., Petikam, L., Valentin, J., Onken, A., and Aodha, O. M. (2024). Geogen: Geometry-aware generative modeling via signed distance functions. *Computer Vision and Pattern Recognition Workshops*.

Fu, Q., Xu, Q., Ong, Y. S., and Tao, W. (2022). Geo-neus: Geometry-consistent neural implicit surfaces learning for multi-view reconstruction. *Advances in Neural Information Processing Systems*.

Gadelha, M., Maji, S., and Wang, R. (2017). 3d shape induction from 2d views of multiple objects. In *Conference on 3D Vision*.

Gao, J., Shen, T., Wang, Z., Chen, W., Yin, K., Li, D., Litany, O., Gojcic, Z., and Fidler, S. (2022a). Get3d: A generative model of high quality 3d textured shapes learned from images. *Advances In Neural Information Processing Systems*.

Gao, K., Gao, Y., He, H., Lu, D., Xu, L., and Li, J. (2022b). Nerf: Neural radiance field in 3d vision, a comprehensive review. *arXiv:2210.00379*.

Geiger, B. (1993). *Three-dimensional modeling of human organs and its application to diagnosis and surgical planning*. PhD thesis, INRIA.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems*.

Gropp, A., Yariv, L., Haim, N., Atzmon, M., and Lipman, Y. (2020a). Implicit geometric regularization for learning shapes. In *ICML*.

Gropp, A., Yariv, L., Haim, N., Atzmon, M., and Lipman, Y. (2020b). Implicit geometric regularization for learning shapes. In *ICML*.

Guo, C. (2023). 3d sa-unet: 3d spatial attention unet with 3d aspp for white matter hyperintensities segmentation. *arXiv:2309.08402*.

Hassanzadeh, A., Vázquez-Suñé, E., Corbella, M., and Criollo, R. (2022). An automatic geological 3d cross-section generator: Geopropy, an open-source library. *Environmental Modelling & Software*.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *International Conference on Computer Vision*.

Henderson, P., Tsiminaki, V., and Lampert, C. H. (2020). Leveraging 2d data to learn textured 3d mesh generation. In *Conference on Computer Vision and Pattern Recognition*.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems*.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*.

Huber, D. J., Mueller, E., and Heubes, P. (1985). Oblique magnetic resonance imaging of normal structures. *American journal of roentgenology*.

Irisawa, N. and Iiyama, M. (2024). High-resolution bathymetry by deep-learning based point cloud upsampling. *IEEE Access*.

Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., and Maier-Hein, K. H. (2021). nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*.

Jacobs, S., Grunert, R., Mohr, F. W., and Falk, V. (2008). 3d-imaging of cardiac structures using 3d heart models for planning in heart surgery: a preliminary study. *Interactive cardiovascular and thoracic surgery*.

Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., and Aila, T. (2021). Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems*.

Karras, T., Laine, S., and Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Conference on Computer Vision and Pattern Recognition*.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). Analyzing and improving the image quality of stylegan. In *Conference on Computer Vision and Pattern Recognition*.

Kazhdan, M., Bolitho, M., and Hoppe, H. (2006). Poisson surface reconstruction. In *Eurographics symposium on Geometry processing*.

Kazhdan, M. and Hoppe, H. (2013). Screened poisson surface reconstruction. *ACM Transactions on Graphics*.

Kingma, D. P. and Welling, M. (2014). Auto-encoding variational bayes. In *International Conference on Learning Representations*.

Koch, S., Matveev, A., Jiang, Z., Williams, F., Artemov, A., Burnaev, E., Alexa, M., Zorin, D., and Panozzo, D. (2019). Abc: A big cad model dataset for geometric deep learning. In *Computer Vision and Pattern Recognition*.

Li, J. et al. (2023a). A hybrid network integrating convolution and transformer for thymoma segmentation. In *Intelligent medicine*.

Li, Z., Müller, T., Evans, A., Taylor, R. H., Unberath, M., Liu, M.-Y., and Lin, C.-H. (2023b). Neuralangelo: High-fidelity neural surface reconstruction. In *Computer Vision and Pattern Recognition*.

Liu, L., Bajaj, C., Deasy, J. O., Low, D. A., and Ju, T. (2008). Surface reconstruction from non-parallel curve networks. In *Computer Graphics Forum*.

Liu, L., Gu, J., Zaw Lin, K., Chua, T.-S., and Theobalt, C. (2020). Neural sparse voxel fields. *NeurIPS*.

Ma, B., Han, Z., Liu, Y.-S., and Zwicker, M. (2021a). Neural-pull: Learning signed distance function from point clouds by learning to pull space onto surface. *ICML*.

Ma, B., Han, Z., Liu, Y.-S., and Zwicker, M. (2021b). Neural-pull: Learning signed distance functions from point clouds by learning to pull space onto surfaces. In *ICML*.

Ma, B., Zhou, J., Liu, Y.-S., and Han, Z. (2023). Towards better gradient consistency for neural signed distance functions via level set alignment. In *Computer Vision and Pattern Recognition*.

Maas, K. W., Pezzotti, N., Vermeer, A. J., Ruijters, D., and Vilanova, A. (2023). Nerf for 3d reconstruction from x-ray angiography: Possibilities and limitations. In *Eurographics Workshop on Visual Computing for Biology and Medicine*.

Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., and Geiger, A. (2019). Occupancy networks: Learning 3d reconstruction in function space. In *Conference on Computer Vision and Pattern Recognition*.

Mildenhall, B., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., and Martin-Brualla, R. (2020). Nerf: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision*.

Moccia, S., De Momi, E., El Hadji, S., and et al. (2018). Blood vessel segmentation algorithms—review of methods, datasets and evaluation metrics. *Comput Methods Programs Biomed*.

Müller, T. (2021). tiny-cuda-nn.

Müller, T., Evans, A., Schied, C., and Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*.

Nam, J., Park, J., Lee, Y., and et al. (2021). Automatic pulmonary vessel segmentation on noncontrast chest ct. *Eur Radiol*.

Nguyen-Phuoc, T., Li, C., Theis, L., Richardt, C., and Yang, Y.-L. (2019). Hologan: Unsupervised learning of 3d representations from natural images. In *International Conference on Computer Vision*.

Nguyen-Phuoc, T. H., Richardt, C., Mai, L., Yang, Y., and Mitra, N. (2020). Blockgan: Learning 3d object-aware scene representations from unlabelled images. *Advances in Neural Information Processing Systems*.

Niemeyer, M. and Geiger, A. (2021). Giraffe: Representing scenes as compositional generative neural feature fields. In *Conference on Computer Vision and Pattern Recognition*.

Nobles, G. R. and Roosevelt, C. H. (2021). Filling the void in archaeological excavations: 2d point clouds to 3d volumes. *Open Archaeology*.

Novello, T., Schardong, G., Schirmer, L., da Silva, V., Lopes, H., and Velho, L. (2022). Exploring differential geometry in neural implicits. *Computers and Graphics*.

Oechsle, M., Peng, S., and Geiger, A. (2021). Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *International Conference on Computer Vision*.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., et al. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

Or-El, R., Luo, X., Shan, M., Shechtman, E., Park, J. J., and Kemelmacher-Shlizerman, I. (2022). Stylesdf: High-resolution 3d-consistent image and geometry generation. In *Conference on Computer Vision and Pattern Recognition*.

Park, J. J., Florence, P., Straub, J., Newcombe, R., and Lovegrove, S. (2019). Deepsdf: Learning continuous signed distance functions for shape representation. In *Conference on Computer Vision and Pattern Recognition*.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Conference on Computer Vision and Pattern Recognition*.

Rebain, D., Li, K., Sitzmann, V., Yazdani, S., Yi, K. M., and Tagliasacchi, A. (2021). Deep medial fields. *arXiv:2106.03804*.

Roich, D., Mokady, R., Bermano, A. H., and Cohen-Or, D. (2022). Pivotal tuning for latent-based editing of real images. *Transactions on Graphics*.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*.

Sawdayee, H., Vaxman, A., and Bermano, A. H. (2023). Orex: Object reconstruction from planar cross-sections using neural fields. In *Computer Vision and Pattern Recognition*.

Schwarz, K., Liao, Y., Niemeyer, M., and Geiger, A. (2020). Graf: Generative radiance fields for 3d-aware image synthesis. *Advances in Neural Information Processing Systems*.

Schwarz, K., Sauer, A., Niemeyer, M., Liao, Y., and Geiger, A. (2022). Voxgraf: Fast 3d-aware image synthesis with sparse voxel grids. *Advances in Neural Information Processing Systems*.

Shaker, A. M., Maaz, M., Rasheed, H., Khan, S., Yang, M.-H., and Khan, F. S. (2024). Unetr++: Delving into efficient and accurate 3d medical image segmentation. *Transactions on Medical Imaging*.

Shang, H., Shen, Y.-G., Li, S., Li, A., and Zhang, T. (2023). An automated mapping method of 3d geological cross-sections using 2d geological cross-sections and a dem. *ISPRS Int. J. Geo-Information*.

Shen, H., Wang, R., Zhang, J., and McKenna, S. J. (2017). Boundary-aware fully convolutional network for brain tumor segmentation. In *MICCAI*.

Shhadi, A. and Barequet, G. (2023). Topology-Controlled Reconstruction from Partial Cross-Sections.

Shi, G., Lu, H., Hui, H., and Tian, J. (2023). Benefit from public unlabeled data: A frangi filtering-based pretraining network for 3d cerebrovascular segmentation.

Shi, P., Hu, J., Yang, Y., Gao, Z., Liu, W., and Ma, T. (2024). Centerline boundary dice loss for vascular segmentation. In *MICCAI*.

Shin, M., Seo, Y., Bae, J., Choi, Y. S., Kim, H., Byun, H., and Uh, Y. (2023). Ballgan: 3d-aware image synthesis with a spherical background. In *International Conference on Computer Vision*.

Shinde, R. C. and Durbha, S. S. (2023). Deep convolutional compressed sensing-based adaptive 3d reconstruction of sparse lidar data: A case study for forests. *Remote Sensing*.

Shit, S., Paetzold, J. C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J. P., Bauer, U., and Menze, B. H. (2021). cldice-a novel topology-preserving loss function for tubular structure segmentation. In *Computer Vision and Pattern Recognition*.

Sitzmann, V., Martel, J., Bergman, A., Lindell, D., and Wetzstein, G. (2020). Implicit neural representations with periodic activation functions. *NeurIPS*.

Soler, L., Hostettler, A., Agnus, V., Charnoz, A., Fasquel, J., Moreau, J., Osswald, A., Bouhadjar, M., and Marescaux, J. (2010). 3d image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database. *IRCAD Tech. Rep.*

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. (2021). Score-based generative modeling through stochastic differential equations. In *ICLR*.

Takikawa, T., Litalien, J., Yin, K., Kreis, K., Loop, C., Nowrouzezahrai, D., Jacobson, A., McGuire, M., and Fidler, S. (2021). Neural geometric level of detail: Real-time rendering with implicit 3D shapes. In *Computer Vision and Pattern Recognition*.

Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J., and Ng, R. (2020). Fourier features let networks learn high frequency functions in low dimensional domains. *NeurIPS*.

Tetteh, G., Efremov, V., Forkert, N., and et al. (2020). Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3d angiographic volumes. *Frontiers in Neuroscience*.

Tov, E., Doe, J., and Smith, A. (2023). Pivotal tuning inversion. *Computer Graphics Forum*.

Vizzo, I., Chen, X., Chebrolu, N., Behley, J., and Stachniss, C. (2021). Poisson surface reconstruction for LiDAR odometry and mapping. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*.

Walker, T., Esposito, S., Rebain, D., Vaxman, A., Onken, A., Li, C., and Aodha, O. M. (2025). Crosssdf: 3d reconstruction of thin structures from cross-sections. *Computer Vision and Pattern Recognition*.

Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., and Wang, W. (2021). Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *Advances in Neural Information Processing Systems*.

Wang, T., Zhang, B., Zhang, T., Gu, S., Bao, J., Baltrusaitis, T., Shen, J., Chen, D., Wen, F., Chen, Q., et al. (2023). Rodin: A generative model for sculpting 3d digital avatars using diffusion. In *Conference on Computer Vision and Pattern Recognition*.

Wang, Y., Skorokhodov, I., and Wonka, P. (2022). Hf-neus: Improved surface reconstruction using high-frequency details. *Advances in Neural Information Processing Systems*.

Wang, Z., Wang, P., Wang, P.-S., Dong, Q., Gao, J., Chen, S., Xin, S., Tu, C., and Wang, W. (2024). Neural-imls: Self-supervised implicit moving least-squares network for surface reconstruction. *Transactions on Visualization and Computer Graphics*.

Wei, M., Zhu, M., Zhang, Y., Wang, J., and Sun, J. (2023). Real-time depth completion based on LiDAR-stereo for autonomous driving. *Frontiers in Neurorobotics*.

Wilson, N. M., Ortiz, A. K., and Johnson, A. B. (2013). The vascular model repository: a public resource of medical imaging data and blood flow simulation results. *Journal of medical devices*.

Wittmann, B., Wattenberg, Y., Amiranashvili, T., Shit, S., and Menze, B. (2024). vesselfm: A foundation model for universal 3d blood vessel segmentation. *arXiv:2411.17386*.

Wood, E., Baltrušaitis, T., Hewitt, C., Dziadzio, S., Cashman, T. J., and Shotton, J. (2021). Fake it till you make it: face analysis in the wild using synthetic data alone. In *International Conference on Computer Vision*.

Xie, Y., Takikawa, T., Saito, S., Litany, O., Yan, S., Khan, N., Tombari, F., Tompkin, J., Sitzmann, V., and Sridhar, S. (2022). Neural fields in visual computing and beyond. In *Computer Graphics Forum*.

Xu, J., Dong, A., Yang, Y., Jin, S., Zeng, J., Xu, Z., Jiang, W., Zhang, L., Dong, J., and Wang, B. (2025). Vsnet: Vessel structure-aware network for hepatic and portal vein segmentation. *Medical Image Analysis*.

Xu, Q. and Tao, W. (2020). Planar prior assisted patchmatch multi-view stereo. In *AAAI Conference on Artificial Intelligence*.

Yagis, E., Aslani, S., Jain, Y., Zhou, Y., Rahmani, S., Brunet, J., Bellier, A., Werlein, C., Ackermann, M., Jonigk, D., et al. (2024). Deep learning for 3d vascular segmentation in phase contrast tomography. *Research Square*.

Yang, Guanyu, e. a. (2011). Automatic centerline extraction of coronary arteries in coronary computed tomographic angiography. *The International Journal of Cardiovascular Imaging, vol. 28, no. 4*.

Yariv, L., Gu, J., Kasten, Y., and Lipman, Y. (2021). Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*.

Yariv, L., Kasten, Y., Moran, D., Galun, M., Atzmon, M., Ronen, B., and Lipman, Y. (2020). Multiview neural surface reconstruction by disentangling geometry and appearance. In *NeurIPS*.

Yariv, L., Puny, O., Neverova, N., Gafni, O., and Lipman, Y. (2024). Mosaic-sdf for 3d generative models.

Ye, H., Zhang, J., Liu, S., Han, X., and Yang, W. (2023). Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721*.

Yifan, W., Wu, S., Oztireli, C., and Sorkine-Hornung, O. (2021). Iso-points: Optimizing neural implicit surfaces with hybrid representations. In *Computer Vision and Pattern Recognition*.

Zhang, L., Rao, A., and Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. *https://arxiv.org/abs/2302.05543*.

Zhang, X., Broersen, A., van Erp, G., Pintea, S. L., and Dijkstra, J. (2024). A graph attention-guided diffusion model for liver vessel segmentation. *arXiv preprint arXiv:2411.00617*.

Zhang, Z. (2024). Efficient slice anomaly detection network for 3d brain mri volume. In *MICCAI*.

Zhao, F., Chen, Y., Hou, Y., and He, X. (2019). Segmentation of blood vessels using rule-based and machine-learning-based methods: A review. *Multimedia Systems*.

Zhou, H.-Y., Guo, J., Zhang, Y., Han, X., Yu, L., Wang, L., and Yu, Y. (2023). nnformer: Volumetric medical image segmentation via a 3d transformer. *Transactions on Image Processing*.

Zou, M., Holloway, M., Carr, N., and Ju, T. (2015). Topology-constrained surface reconstruction from cross-sections. *ACM Transactions on Graphics*.