Titre: Title:	Intraoperative Vertebra Segmentation From 3D Point Clouds for Radiation-Free Spinal Curve Tracking in Scoliosis Surgery
Auteur: Author:	Yu-Chi Kung
Date:	2025
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Kung, Y.-C. (2025). Intraoperative Vertebra Segmentation From 3D Point Clouds for Radiation-Free Spinal Curve Tracking in Scoliosis Surgery [Master's thesis, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/70103/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/70103/
Directeurs de recherche: Advisors:	Lama Séoud, Manuela Kunz, & Stefan Parent
Programme: Program:	génie informatique

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Intraoperative Vertebra Segmentation from 3D Point Clouds for Radiation-Free Spinal Curve Tracking in Scoliosis Surgery

YU-CHI KUNG
Département de génie informatique et génie logiciel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie informatique

Novembre 2025

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Ce mémoire intitulé :

Intraoperative Vertebra Segmentation from 3D Point Clouds for Radiation-Free Spinal Curve Tracking in Scoliosis Surgery

présenté par **Yu-Chi KUNG**
en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Farida CHERIET, présidente
Lama SÉOUD, membre et directrice de recherche
Manuela KUNZ, membre et codirectrice de recherche
Stefan PARENT, membre et codirecteur de recherche
Michael LEUNG, membre

DEDICATION

À ma famille à Taïwan,
à mes amis, mes labmates, mon superviseur,
et à vous tous que j'ai eu la chance de rencontrer à Montréal,
vous me manquerez.

ACKNOWLEDGEMENTS

This journey began with a meaningful encounter, one that might never have happened without Professor Farida Cheriet and my dear friend Imane Chefi. Their early encouragement and belief in me were the sparks that set everything in motion, and for that, I am truly grateful.

I am especially thankful to my supervisor, Professor Lama Séoud, who has been far more than an academic guide. She has been both a mentor and a friend, supporting me through challenges with patience, empathy, and insight. I also want to thank Professor Manuela Kunz, whose work and perspective inspired key ideas in this project, and Dr. Stefan Parent, an exceptional surgeon and generous mentor, for inviting me to access the OR, facilitating connections, and making possible a short-term internship with Orthofix. Special thanks to Soraya Barchi for her kind support throughout the process.

That internship in Toronto was a turning point. I'm grateful to Michael Leung for the opportunity and continuous encouragement, and to Roozbeh Shams, whose mentorship during data collection and problem-solving made a lasting impact. The experience not only expanded my technical knowledge but also strengthened my motivation to pursue a career in industry.

Closer to home, being part of VisionIC has been one of the most rewarding aspects of this journey. My sincere thanks to Philippe Baumstimler, Hugo Rodet, Doha Zrouki, Gaspar Faure, Étienne Lescarbeault, Victor Nogues, and Ghazal Ebrahimi; your support, both technical and personal, helped me navigate research and life abroad. I'm also grateful to Philippe Debanné and Hanna Sabir for assisting with Dataset P, saving me valuable time. Warm thanks to my friends at Magnu, especially Imane Chefi, as well as our dedicated intern Tibor Kubik, whose contributions greatly advanced my work.

This thesis was made possible by the generous support of the OPSIDIAN Foundation and the TransMedTech Institute, whose funding enabled me to carry out this research with focus and freedom.

Finally, to my family, though we may be far apart, your love and unwavering support have accompanied me every step of the way. Thank you for making this journey not only possible, but truly meaningful.

RÉSUMÉ

La scoliose se caractérise par une courbure anormale de la colonne vertébrale, pouvant avoir un impact significatif sur la posture, la mobilité et la qualité de vie des patients. Pour les cas les plus sévères, une intervention chirurgicale corrective devient nécessaire afin de réaligner la colonne vertébrale et de prévenir l'aggravation de la déformation, en stabilisant ou en fusionnant les vertèbres atteintes. Cette intervention s'effectue généralement en deux étapes clés. D'abord, la pose de vis pédiculaires pour ancrer les vertèbres ; ensuite, le réalignement progressif de la colonne pour corriger la déformation. Bien que les procédures chirurgicales assistées par ordinateur aient considérablement amélioré la précision et l'efficacité de la pose des vis pédiculaires, la deuxième étape, à savoir le réalignement de la colonne, repose encore largement sur l'expertise du chirurgien et son appréciation visuelle, et bénéficie de beaucoup moins de soutien technologique. De plus, la majorité des systèmes de navigation actuels s'appuie sur l'imagerie radiographique peropératoire, exposant ainsi les patients et le personnel médical aux rayonnements ionisants, ce qui soulève des préoccupations en matière de sécurité et d'usage à long terme.

Cette thèse propose une nouvelle approche d'assistance peropératoire sans recours aux rayonnements, basée sur l'utilisation de nuages de points 3D acquis de manière non-irradiante par un senseur à lumière structurée. Nous présentons un pipeline de segmentation fondé sur l'apprentissage profond, capable d'identifier en temps réel les structures vertébrales apparentes à partir des données 3D. Comparé aux méthodes existantes, notre approche basée sur le modèle Point Transformer V3, présente de meilleurs performances sur la tâche de segmentation des vertèbres, évaluée à partir de la base de données public SpineDepth. Pour pallier à la limitation des déplacements vertébraux dans ce jeu de données et pour améliorer la généralisation du modèle à différents contextes, nous avons développé trois bases de données semi-synthétiques additionnelles, intégrant une stratégie d'augmentation colorimétrique simulant diverses variations anatomiques et conditions d'imagerie rencontrées en chirurgie de la scoliose. Nous avons démontré que l'utilisation additionnelle de ces données semi-synthétiques pour l'entrainement du modèle permet, du moins qualitativement, une meilleure segmentation des vertèbres dans des acquisitions intra-opératoires réelles.

Ce travail établit un lien entre les recherches antérieures en modélisation anatomique préopératoire et les futurs systèmes d'évaluation du réalignement en temps réel, représentant une avancée notable vers une chirurgie de la scoliose assistée par l'image sans rayonnement ionisant et permettant un suivi en continu de l'alignement vertébral.

ABSTRACT

Scoliosis is characterized by an abnormal 3D curvature of the spine that can significantly affect a patient's posture, mobility, and quality of life. In severe cases, corrective surgery becomes necessary to realign the spinal column and prevent further progression of the curvature by stabilizing or fusing the affected vertebrae. This procedure is typically carried out in two critical stages: first, the placement of pedicle screws to anchor the vertebrae; second, the spinal realignment to gradually correct the deformity. Although computer-assisted surgical technologies have greatly improved the accuracy and efficiency of pedicle screw placement, the second phase, spinal realignment, remains largely dependent on the surgeon's expertise and visual judgment, and receives far less technological support. Furthermore, most existing navigation systems rely on intraoperative radiographic imaging, which exposes both patients and medical staff to ionizing radiation, raising concerns about safety and long-term use.

This thesis proposes a novel, radiation-free approach to intraoperative guidance using 3D point clouds acquired using a structured-light sensor. We present a deep learning-based segmentation framework capable of identifying exposed vertebral structures in real time, directly from 3D point clouds. At the core of our framework is the Point Transformer V3, which demonstrated superior performance on vertebrae segmentation over prior methods when evaluated on the public SpineDepth dataset. Because this dataset is limited in terms of vertebral displacement and, at the same time, to improve domain generalization, we developed three additional semi-synthetic datasets with a color-based augmentation strategy that simulates a range of anatomical and imaging variations encountered in scoliosis surgery. Incorporating these semi-synthetic datasets into model training noticeably improves vertebra segmentation in real intraoperative acquisitions.

This work builds a bridge between earlier research in preoperative anatomical modeling and future systems for real-time continous spinal alignment assessment, representing a critical step toward comprehensive, intelligent, and radiation-free image-guided scoliosis surgery.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

AIS	Adolescent Idiopathic Scoliosis
MRI	Magnetic Resonance Imaging
CT	Computed Tomography
AP	Antero-Posterior
PSF	Posterior Spinal Fusion
ASF	Anterior Spinal Fusion
IGSS	Image-Guided Surgery Systems
IGS	Image-Guided Systems
CAN	Computer-Assisted Navigation
OR	Operating Room
EM	Electromagnetic
OTI	Optical Topographic Imaging
MvIGS	Machine-vision Image-Guided Surgery
LiDAR	Light Detection and Ranging
CNN	Convolutional Neural Networks
AR	Augmented Reality
SLS	Structured Light Scanning
SVM	Support Vector Machines
MLP	Multilayer Perceptron
kNN	k-Nearest-Neighbor
PTv3	Point Transformer V3
AI	Artificial Intelligence
TLS	Terrestrial Laser Scanner
SDF	Signed Distance Field
DSC	Dice Similarity Coefficient
FPFH	Fast Point Feature Histograms
ICP	Iterative Closest Point
DICOM	Digital Imaging and Communications in Medicine
LOOCV	Leave One Out Cross Validation

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

Scoliosis, characterized by an abnormal 3D curvature of the spine, can significantly impact a patient's posture, mobility, and overall quality of life. In severe cases, surgery often becomes the only effective solution to stop the progression of spinal curvature and restore proper alignment. The surgical procedure is typically carried out in two critical stages: the first involves the placement of pedicle screws to anchor the vertebrae, and the second is the spinal realignment itself, where the curved spine is gradually corrected and stabilized.

Over the past decade, modern surgical technologies have greatly enhanced precision in the surgical environment. Computer-assisted navigation systems now routinely support pedicle screws placement, offering real-time visual guidance to improve both safety and accuracy. However, this technological support is largely confined to the first phase of the surgery. Once the screws are placed, the critical task of spinal correction is still primarily through the surgeon's expertise, visual assessment, and tactile feedback. Despite its clinical importance, this second phase, spinal realignment, remains one of the least supported steps in scoliosis surgery by current computer-assisted systems.

The research presented in this thesis emerges from an effort to bridge this gap. It is part of a broader, long-term vision to develop an automatic, intelligent, radiation-free system that assists throughout the entire surgical workflow, not only during pedicle screws placement, but also during spinal correction. At the heart of this vision is the use of 3D point clouds acquired intraoperatively using structured-light technologies, such as the 7D Surgical System. These imaging platforms offer real-time anatomical views without radiation, opening new possibilities for intraoperative guidance.

This research project builds on earlier efforts within our research team, which focused on the acquisition of the preoperative spine shape. More precisely, the work of Antonin Tranchon proposed a method for segmenting vertebrae from MRI scans to create 3D anatomical models with detailed posterior arches (see Figure 1.1(a)). It also presented a proof-of-concept for the registration of the preoperative spine model to a structured-light scan of the spine model that simulated the surgical condition, notably without other surrounding tissues (see Figure 1.1(b)). This registration aimed to assess the spinal alignment during the simulated procedure [2]. The current project presented in this thesis takes the next step by taking actual intraoperative data and automatically segmenting the vertebrae in the point cloud (see Figure 1.2). The ultimate goal is to create a closed-loop system that continuously monitors spine geometry during the surgery, without radiation, to evaluate the spinal alignment and

provide feedback to the surgeon in real time.

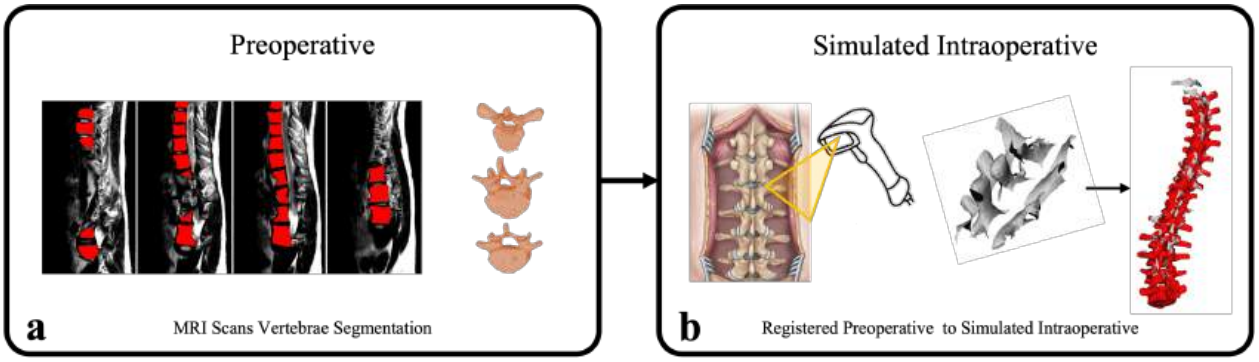

Figure 1.1 Overview of the previous work by Antonin Tranchon [2]. (a) Vertebrae segmentation from MRI scans to generate 3D anatomical models with detailed posterior arches. (b) Registration of the preoperative model to a structured-light scan simulating intraoperative conditions, without surrounding tissues. *Adapted from* Tranchon et al. (2024)

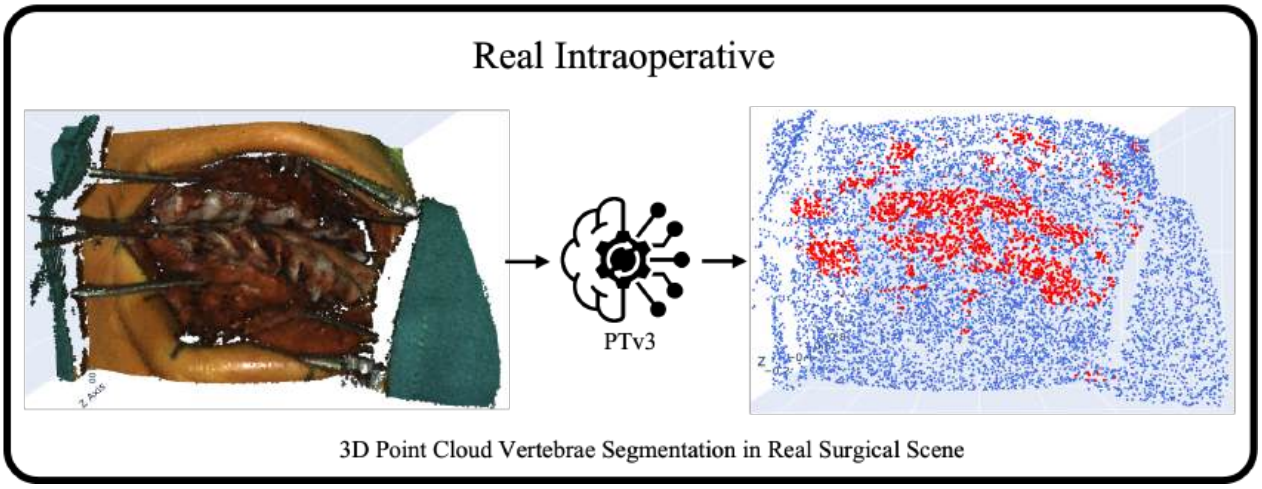

Figure 1.2 Overview of our work: automatic segmentation of vertebrae from intraoperative 3D point clouds.

The structure of this thesis reflects the layered progression of this research:

Chapter 2 introduces the foundational knowledge required to understand this work and provides an overview of the literature related to surgical assistance for scoliosis and the use of 3D data in intraoperative conditions.

Chapter 3 identifies the key limitations in current systems, outlines the motivation behind this study, and defines the central research objectives.

Chapter 4 details the approach taken to address these objectives, beginning with the selection of a deep learning model capable of capturing geometric structures for segmentation. It also describes datasets acquisition, processing pipelines, and the training strategy used to generalize the model across different domains, particularly given the limited accessibility of intraoperative data.

Chapter 5 presents the results of our experiments and qualitative evaluations, demonstrating how the proposed approach performs under real intraoperative conditions. The appendix further explores how segmentation outcomes can support future applications such as spinal alignment tracking via registration, along with other complementary experiments that highlight current limitations.

Finally, Chapter 6 synthesizes the findings, revisits the research contributions, and proposes future directions for clinical translation and intraoperative data collection.

CHAPTER 2 LITERATURE REVIEW

This chapter provides a structured review of the scientific literature and foundational concepts related to this research. Section 2.1 introduces Adolescent Idiopathic Scoliosis (AIS), the clinical condition underpinning this study's motivation. Section 2.2 reviews current surgical interventions for scoliosis, emphasizing the integration of image-guided surgical (IGS) systems in clinical workflows. Section 2.3 examines the fundamentals of three-dimensional (3D) data, including its representations, acquisition methods, and relevance in medical imaging. Section 2.4 explores segmentation methodologies, contrasting traditional 2D-based techniques with recent advances in 3D point cloud segmentation. Finally, Section 2.5 outlines the development of semi-synthetic surgical scenes, discussing their utility in augmenting training datasets and supporting the development of robust deep learning models.

2.1 Adolescent Idiopathic Scoliosis (AIS)

Scoliosis is characterized by an abnormal 3D curvature of the spine. It can develop at any age, but adolescent idiopathic scoliosis (AIS) is the most common type, affecting approximately 2% to 4% of adolescents. Although scoliosis occurs equally in males and females, females are up to ten times more likely to experience curve progression. In most cases, it is idiopathic, meaning that the underlying cause remains unknown.

Diagnosis and Clinical Evaluation

AIS is commonly diagnosed during adolescence through physical examinations and imaging. The curvature can develop in any region of the spine and is diagnosed through physical examination and imaging. A primary clinical screening tool is the Adam's forward bend test, where the patient bends forward at the waist while the clinician observes for rib cage asymmetry or a visible rib hump, indicating spinal rotation. Confirmation is achieved through standing full-spine X-rays, where a curvature in the frontal plane, measured by the Cobb angle, greater than 10° confirms a diagnosis of scoliosis (see Figure 2.1). In specific cases involving rapid progression, neurological symptoms, or early onset, magnetic resonance imaging (MRI) is performed to rule out associated neurological abnormalities such as tethered cord, syringomyelia, or spinal tumors [9]. When a scoliosis is confirmed, regular clinical follow-up every 6- to 12-months is recommended, with X-rays acquisition, to monitor the progression.

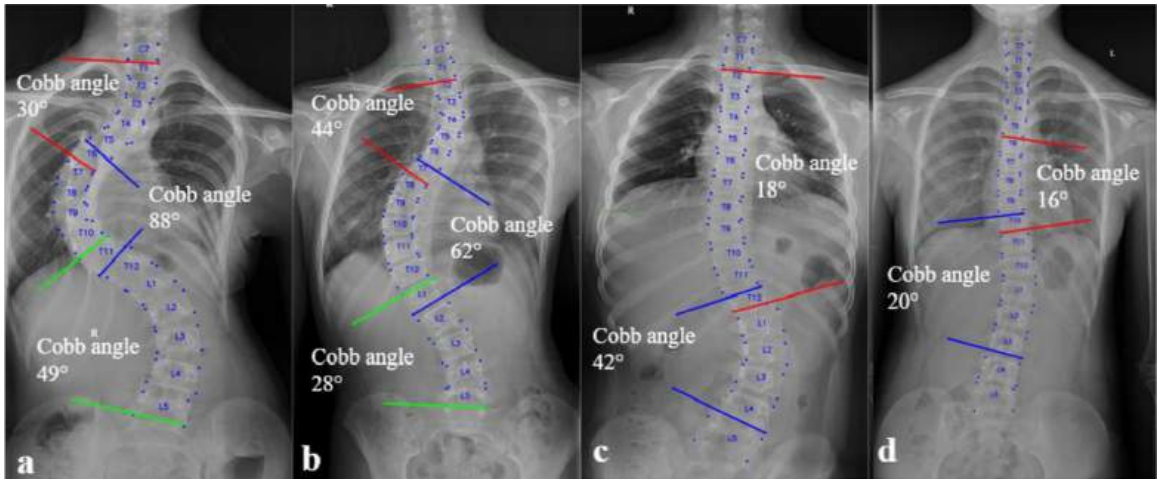

Figure 2.1 Visualization of spinal curvatures with varying Cobb angles on standing antero-posterior (AP) radiographs. *Adapted from* Sun et al. (2022) [3].

Treatment and Surgical Indications

Management strategies for AIS depend on its severity quantified by the Cobb angle, the location of the curvature (apical vertebra), and progression of the spinal curvature. Nonsurgical methods, such as bracing, may be effective in milder cases but can fail in up to 42.5% of patients [10]. Surgical intervention is recommended when the Cobb angle exceeds 45° – 50° due to the following concerns:

- Curves over 50° typically tend to progress even after skeletal maturity.
- Severe curvatures may compromise pulmonary function and result in respiratory complications.
- Progressive deformities become increasingly challenging to correct surgically.

The primary goal of AIS surgical treatment is to realign the spine in all planes and to maintain this correction. Modern surgical instrumentation, such as pedicle screws and rods, has significantly improved correction outcomes. Following successful fusion, patients often return to normal activities, including sports [11].

2.2 Scoliosis Surgery

This section reviews surgical approaches for AIS, introduces image-guided systems in spine surgery, and highlights their clinical benefits.

2.2.1 Surgical Procedures for AIS

Our research primarily investigates the surgical management of severe scoliosis, with a particular focus on AIS. Common surgical interventions include posterior spinal fusion (PSF), anterior spinal fusion (ASF), or a combination of both. Among these, PSF remains the most widely accepted technique for preventing curve progression following skeletal maturity [12]. This procedure involves a midline posterior approach to expose key spinal structures, including laminae, spinous processes, transverse processes, and facet joints.

In spinal fusion surgery, the curved vertebrae are joined together so they heal into a single, solid bone. This stops growth in the affected segment of the spine and prevents the curvature from worsening. To promote fusion, surgeons use a material called a bone graft. Small pieces of bone are placed between the vertebrae being fused, and over time, these pieces grow together, much like the healing process of a broken bone. To keep the spine properly aligned while the fusion occurs, metal rods are usually implanted. These rods are secured to the spine using screws, hooks, or wires. The number of instrumented vertebral levels depend on the curvature.

The surgical intervention includes two main steps: 1) pedicle screws insertion and 2) spinal realignement by various maneuvers such as rod derotation and direct vertebral rotation.

To verify the adequacy of deformity correction and implant positioning, intraoperative imaging is routinely employed before finalizing the procedure. Surgeons typically utilize 2D radiographs to obtain real-time anteroposterior (AP) and lateral views of the spine, enabling immediate assessment of rod contouring, pedicle screw trajectory, and overall spinal alignment. Moreover, radiographs play a critical role throughout the perioperative process, not only during surgery, but also for preoperative planning and postoperative follow-up, where they help detect hardware complications and monitor fusion progress [13]. Following confirmation of proper alignment and hardware positioning, additional procedures such as facetectomies and osteotomies may be performed to further enhance spinal flexibility and correction. Once optimal alignment is achieved, bone graft materials are placed to facilitate spinal fusion.

Although PSF is highly effective in achieving substantial deformity correction and long-term stability [14], it remains an invasive procedure associated with considerable blood loss, soft tissue disruption, and postoperative morbidity. Conventional pedicle screw placement relies heavily on anatomical landmarks, tactile feedback, and fluoroscopic imaging. However, these free-hand methods carry a notable risk of inaccuracies, with pedicle screw misplacement reported in up to 10% of cases and 1 in 300 patients may require revision surgery [15]. The risk is particularly elevated in patients with smaller stature or severely deformed vertebrae, where

narrowed pedicle anatomy increases the likelihood of neurovascular or visceral injury [16].

Given these challenges, the need for enhanced precision has driven the development of advanced guidance technologies. Intraoperative 3D imaging and navigation systems have emerged as effective solutions to improve the accuracy and safety of screw placement. These systems enhance visualization and intraoperative decision-making, resulting in fewer placement errors and improved surgical outcomes across spinal regions [17].

In the following section, we will explore the principles and integration of image-guided navigation systems in modern spine surgery.

2.2.2 Image-Guided Surgery Systems (IGSS)

Image-guided surgery (IGS), also known as computer-assisted navigation (CAN), refers to a set of intraoperative navigation techniques that use preoperative imaging data to guide surgical procedures. IGS systems are computerized platforms that integrate imaging modalities, such as CT, MRI, or 3D fluoroscopy, to provide real-time, three-dimensional visualization of anatomical structures and surgical tools. By offering accurate spatial information during procedures, particularly helpful when the anatomy of interest is unexposed, IGS significantly enhances surgical precision, minimizes intraoperative complications, and contributes to better clinical outcomes [18].

IGSS in Spine Surgery

In spine surgery, where critical neurovascular structures lie within millimeters of surgical landmarks, precision is paramount. Traditional methods, based on a surgeon's anatomical knowledge, tactile feedback, and fluoroscopy guidance, have served for decades but show limitations, especially in cases with severe deformities or atypical anatomy [19]. These techniques are also associated with increased radiation exposure and longer operative times, particularly in minimally invasive or multilevel procedures [11,20]. As image-guided systems become more widely adopted in routine spine surgery, it is essential to understand their technological evolution and current applications. A clear understanding of this evolution not only informs the effective adoption of current systems but also enables surgeons to anticipate limitations, optimize workflows, and align procedural protocols with emerging biomedical technologies [21]. The following subsection provides an overview of how IGSS have progressed, tracing its progression from early guidance tools to today's advanced navigation platforms.

Evolution of IGSS

The progression of IGSS in spine surgery has followed a clear trajectory: each generation sought to resolve the shortcomings of its predecessor, yet often introducing new limitations. Understanding this technological evolution not only clarifies the current landscape but also highlights the need for innovation in surgical navigation.

The earliest image guidance relied on plain radiographs in the late 19th century, offering basic anatomical reference but limited by their static, two-dimensional nature and lack of depth perception. The advent of 2D fluoroscopy in the mid-20th century provided real-time intraoperative imaging, improving decision-making and screw placement accuracy. However, its single-plane visualization required frequent repositioning of the C-arm and exposed patients and surgical staff to cumulative radiation.

To improve spatial resolution, 3D imaging systems such as cone-beam CT and intraoperative 3D fluoroscopy (e.g., Medtronic O-arm, Ziehm 3D C-arm) were introduced [22]. These enabled volumetric visualization and improved anatomical assessment. Yet, their adoption brought new challenges: complex setup workflows, reliance on non-sterile radiology staff, increased operative time, and the continued use of ionizing radiation. The physical size of these systems also limits maneuverability in the operating room (OR) and surgeon autonomy.

Concerns over radiation exposure, particularly for pediatric patients undergoing repeated imaging, led to the development of non-radiative technologies. These included electromagnetic (EM) tracking, optical topographic imaging (OTI), and surface mapping based on external landmarks [23, 24]. Although these approaches reduced radiation, they introduced their own limitations. EM tracking was prone to metal interference, while optical systems suffered from line-of-sight issues. Non-sterile camera placement outside the operative field further disrupted workflow by requiring indirect adjustments via OR staff.

Radiation remains a long-term concern. Studies report a fivefold increase in cancer incidence among patients with adolescent idiopathic scoliosis (AIS) over 25 years [25, 26]. Orthopedic residents, too, receive nearly double the radiation exposure of the general population [27].

In sum, while IGSS have transformed spinal surgery by increasing safety and accuracy, legacy systems have introduced significant challenges, including extended setup and surgical time, dependence on non-sterile staff, bulky equipment in the operating room, and ongoing radiation exposure. The convergence of these issues, radiation risk, workflow inefficiency, and reduced surgeon autonomy, paved the way for a new generation of image-guided systems. One such innovation is the 7D Surgical Machine-vision Image-Guided Surgery (MvIGS) system, which represents a paradigm shift in spinal and cranial navigation.

The 7D Machine-vision Image Guided Surgery (MvIGS) system leverages non-ionizing structured light and advanced machine vision algorithms to acquire high-resolution 3D surface scans of the surgically exposed spine [4, 11, 28]. Its proprietary **FLASH registration** algorithm aligns these intraoperative scans to a preoperative CT-derived spine model, enabling radiation-free anatomical navigation during spinal procedures (see Figure 2.2).

The registration process traditionally begins with the manual identification of three anatomical landmarks on each vertebra in the preoperative scan. During surgery, corresponding points are manually selected on the intraoperative surface scan. The 7D software computes an initial transformation matrix using these paired landmarks and subsequently refines it using its FLASH algorithm. Once complete, the overlaid preoperative model enables real-time surgical navigation without the use of intraoperative X-ray or CT.

Recent system enhancements have focused on improving registration efficiency and anatomical adaptability. These upgrades aim to facilitate automatic or semi-automatic registration of multiple vertebral levels from a single structured-light scan, addressing one of the primary challenges in spinal surgery, discrepancies between spinal geometry in the preoperative CT (typically acquired in a supine position) and the actual alignment during surgery (performed in a prone position). By allowing each vertebra to be independently registered, the system is designed to accommodate spinal flexibility and intraoperative correction maneuvers more effectively.

Despite these technical advancements, several limitations remain. First, the system still relies on a semi-manual workflow for registration initialization. Surgeons must manually click corresponding anatomical landmarks in both the preoperative and intraoperative datasets, which introduces variability and depends heavily on user expertise. This manual step can be particularly challenging in cases where anatomical landmarks are partially obscured or only limited portions of the vertebrae are exposed.

Second, although recent improvements enhance intraoperative tracking and reduce reliance on intraoperative imaging, they do not consistently eliminate the need for fluoroscopy. In complex spinal deformity cases, especially those involving significant anatomical variability or flexible spines, surgeons may lack sufficient confidence in the automatic surface-based registration and revert to C-arm fluoroscopy for confirmation. Observations from clinical practice indicate that fluoroscopic imaging is often used alongside the 7D system to ensure registration accuracy, particularly when navigating deeper structures or verifying alignment post-correction.

These remaining challenges suggest that while the 7D MvIGS platform provides a fast, radiation-free alternative to traditional image guidance systems, its current capabilities may

not yet fully replace fluoroscopic feedback in all surgical scenarios. Furthermore, the reliance on semi-manual landmark selection limits automation and may affect reproducibility, particularly in procedures requiring high precision across multiple spinal levels.

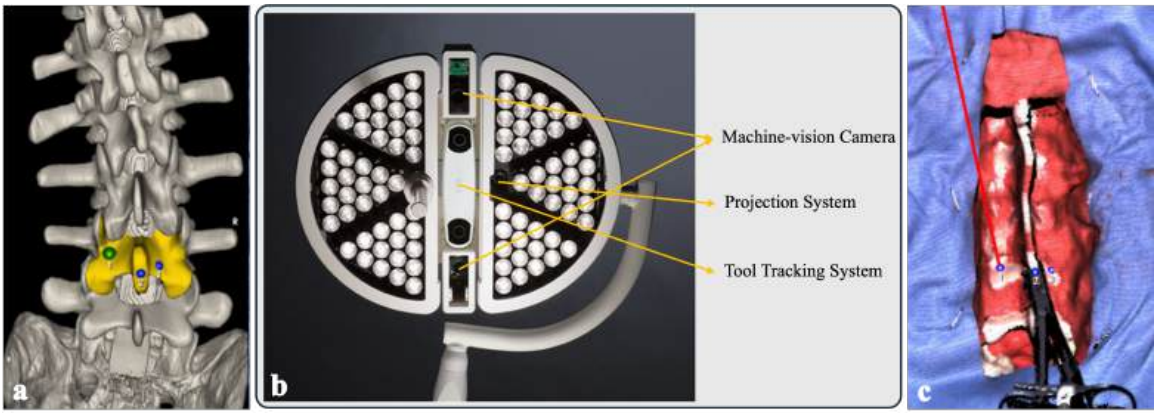

Figure 2.2 Registration workflow in 7D Surgical's Flash Registration system. (a) The process begins with a 3D rendering of the preoperative CT scan, where the target anatomy is defined by the yellow highlighted region. (b) A machine-vision camera captures a high-resolution 3D surface scan of the exposed anatomy using structured light. (c) The intraoperatively digitized surface of the spine is shown. The red Play-Doh simulates soft tissue to approximate surgical conditions. The 7D system aligns the captured surface scan with the preoperative model through its proprietary registration algorithm, enabling radiation-free navigation. *Adapted from* Faraji et al. (2020) [4]

2.3 3D Data in IGSS

Three-dimensional (3D) data forms the backbone of modern image-guided surgery systems (IGSS), supporting accurate spatial modeling, real-time navigation, and postoperative analysis. Unlike two-dimensional (2D) data, which provides only planar information, 3D data encodes geometry in terms of width, height, and depth, allowing for a comprehensive volumetric understanding of anatomical structures.

3D Data Representation

3D data can be represented in various formats, each tailored to specific use cases:

- Point clouds: unordered collections of spatial (x, y, z) coordinates representing surface geometry
- Surface meshes: interconnected vertices forming polygonal surfaces

- Voxels: 3D pixels capturing volumetric intensity (as in CT or MRI)
- Surfels: surface elements combining geometry, normals, and optional color

Among these, point clouds are especially valuable for surgical applications due to their compactness, direct representation of surface anatomy, and computational efficiency. They are ideally suited for registration, navigation, and tracking tasks, where fast and accurate surface alignment is crucial [29].

Beyond the operating room, 3D point clouds have become a foundational data type across fields such as computer vision, robotics, machine learning, and geographic information systems [30]. In medical imaging, point clouds derived from structured light, stereo vision, or volumetric imaging enable clinicians to visualize patient-specific anatomy in high spatial detail. This capability enhances diagnosis precision, supports surgical planning, and improves intraoperative guidance.

Point cloud data has also been widely adopted for medical training and simulation in recent years. By reconstructing realistic 3D models of human anatomy, surgeons can rehearse procedures, while students can explore complex anatomical relationships in an interactive, virtual environment [31].

3D Data Acquisition

The generation of 3D anatomical data in IGSS typically relies on one or more of the following techniques:

- Volumetric imaging (e.g., CT or MRI): provides dense voxel grids of internal anatomy.
- Stereo vision: estimates depth from images captured by spatially separated cameras.
- Photogrammetry: reconstructs 3D structure from multiple 2D views.
- Laser scanning (LiDAR): uses light pulses to measure distances via time-of-flight.
- Structured light scanning: projects a known pattern onto a surface and analyzes deformation to infer depth.

A prominent example of structured light scanning in a surgical context is the 7D Surgical Machine-vision Image-Guided Surgery (MvIGS) system. This platform combines structured infrared light projection with stereo vision cameras to generate dense, high-resolution point clouds of the exposed anatomy in real time [32]. The resulting surface scan is automatically

registered to the preoperative CT using the FLASH registration method, streamlining the workflow without requiring intraoperative X-ray or manual landmark-based alignment [4,33].

In addition to large console-based systems, hand-held structured light scanners have been investigated as portable, low-cost alternatives. For instance, Chan et al. [34] demonstrated the feasibility of using a hand-held structured light scanner to acquire intraoperative surface data for tissue classification. Such compact devices offer greater flexibility in constrained surgical environments and represent a promising direction for mobile 3D acquisition and navigation support.

2.4 Segmentation Techniques

Segmentation plays a central role in computer-assisted interventions, enabling the identification of anatomical structures or surgical instruments for real-time guidance, registration, and visualization. While traditionally developed for radiological imaging, segmentation approaches differ considerably depending on the imaging modality, intraoperative constraints, and clinical objectives. In this section, we differentiate between preoperative medical image segmentation and intraoperative surgical scene segmentation, with a focus on the evolution from 2D RGB-D techniques to native 3D point cloud segmentation.

2.4.1 Medical Image Segmentation

Medical image segmentation is primarily used in preoperative planning, targeting imaging modalities such as CT and MRI. Over time, techniques have progressed from rule-based methods to deep learning approaches, particularly convolutional neural networks (CNNs), which are now widely adopted to segment organs, spine, and tumors in 2D slices or 3D volumes [35–37]. These models perform well in preoperative settings where data can be processed offline and at high resolution. For instance, Tranchon et al. proposed a vertebra segmentation pipeline for adolescent idiopathic scoliosis (AIS) using whole-slice MRI and hybrid refinement. Their method accurately delineates spinal anatomy, especially the posterior vertebral arch, which is crucial for surgical planning. However, these solutions are not readily transferable to intraoperative environments due to constraints on imaging modalities and the need for real-time performance [2].

2.4.2 Surgical Scene Semantic Segmentation

Intraoperative segmentation presents unique challenges compared to preoperative imaging. The surgical field is dynamic, partially occluded, and often illuminated under non-standard

lighting conditions. Furthermore, intraoperative systems demand real-time performance to support tasks like instrument tracking and anatomy registration.

RGB-D Based Approaches

The rise of RGB-D sensors and surgical microscopes has led to a flow of CNN-based segmentation methods to process either RGB or RGB-D images [1, 38, 39]. RGB-D data combines standard 2D images with a corresponding depth channel, offering enhanced spatial cues. However, this format remains fundamentally 2.5D, lacking true volumetric understanding. Most segmentation models for RGB-D rely on 2D convolutional architectures like U-Net, which process depth as an auxiliary channel rather than as a fully spatial structure.

Recent studies illustrate the potential of RGB-D segmentation: Tanzi et al. proposed a real-time deep learning framework for semantic segmentation of intraoperative images to enhance 3D augmented reality (AR) overlays. Their model processes RGB endoscopic streams to segment tissue in real time for surgical navigation [38]. Similarly, Scheikl et al. developed a deep learning approach for semantic segmentation of organs and tissues in laparoscopic images, aiming to highlight critical anatomical areas during surgery and support AR-based manual navigation and planning [39]. More recently, Liebmann et al. presented a marker-less surgical navigation system for spine procedures, combining RGB-based segmentation with continuous pose tracking to enable automatic registration with preoperative data [1].

While effective, most RGB-D methods treat depth as a secondary input to 2D CNNs. This 2.5D approach lacks volumetric consistency, and reprojecting 3D information into 2D planes can introduce errors, especially in anatomically complex regions like the spine. The reliance on planar representations limits their ability to fully model surface continuity and geometric detail.

Point Cloud Segmentation in Orthopedic Surgery

To address the limitations of 2.5D techniques, we shift focus to native 3D point cloud segmentation. Point clouds are an unprojected representation of 3D space, directly capturing surface geometry and preserving spatial continuity, qualities critical for high-fidelity segmentation in image-guided surgery.

One promising technology for real-time point cloud acquisition is structured light scanning (SLS). For example, the 7D Surgical Machine-vision Image-Guided Surgery (MvIGS) system uses SLS combined with stereo vision cameras to reconstruct dense, radiation-free 3D point clouds of the exposed surgical field. These data provide a geometrically accurate, high-

resolution model of the patient's anatomy suitable for both segmentation and registration [4].

Further supporting this approach, Chan et al. demonstrated the potential of machine learning to classify tissues such as bone, cartilage, and ligament directly from structured light scans. Their pipeline extracted spatial and textural features from 3D surfaces and used Random Forests, SVMs, and simple feedforward neural networks to achieve classification accuracies of 80-90%. Importantly, their results suggest that incorporating spatial geometry improves performance, highlighting the untapped potential of native 3D data for intraoperative segmentation [34].

Despite its promise, direct point cloud segmentation in surgical scenes remains underexplored. Many previous methods voxelize or project the data into 2D to maintain compatibility with conventional CNNs, sacrificing spatial precision in the process.

2.5 Deep Learning on 3D Point Clouds

As outlined in the previous section, conventional 2D and RGB-D segmentation techniques struggle to accurately capture complex surgical scenes where fine surface detail and spatial continuity are crucial. Unlike images, which are represented as structured grids, point clouds are unordered and sparse, lacking an explicit neighborhood structure. These properties make them incompatible with conventional convolutional neural networks (CNNs) designed for image-based tasks. Early attempts to bridge this gap involved voxelizing point clouds or projecting them into multiple 2D views; however, these transformations often led to quantization errors and loss of geometric fidelity issues, which are problematic in the surgical domain, where precision is key.

2.5.1 PointNet and PointNet++

To overcome these challenges, the field has shifted toward direct learning from raw 3D point clouds using specialized architectures that maintain permutation invariance, geometric continuity, and spatial flexibility. This transition marked a significant breakthrough with the introduction of PointNet [5], the first deep neural network designed to operate directly on unordered point sets. PointNet processes individual points independently with shared MLPs, then aggregates global context via symmetric functions (e.g., max pooling), effectively capturing both local and global features. This architecture preserved the raw spatial properties of 3D data, setting a new baseline for tasks such as shape classification and part segmentation (see Figure 2.3).

However, PointNet's lack of local neighborhood aggregation limited its ability to capture spa-

tial relationships, which are essential for dense semantic segmentation. To address this, PointNet++ [6] extended the architecture by introducing hierarchical feature learning through local grouping and sampling. This allowed the model to capture multiscale geometric features while retaining PointNet's strengths in permutation invariance and efficiency. PointNet++ remains a cornerstone in 3D deep learning, widely adopted in medical applications such as dental modeling, organ segmentation, and orthopedic planning.

Despite these improvements, PointNet++ still faces key limitations when applied to more complex and high-resolution data. Its reliance on fixed-radius or k-nearest-neighbor (k-NN) queries for local grouping can lead to inconsistent performance across varying point densities and different scales. Additionally, the use of max pooling within local regions discards potentially informative contextual features, reducing the model's ability to learn about fine structure. Most notably, PointNet++ is unable to model long-range dependencies, which are crucial for capturing relationships across spatially distant and functionally connected regions (see Figure 2.4).

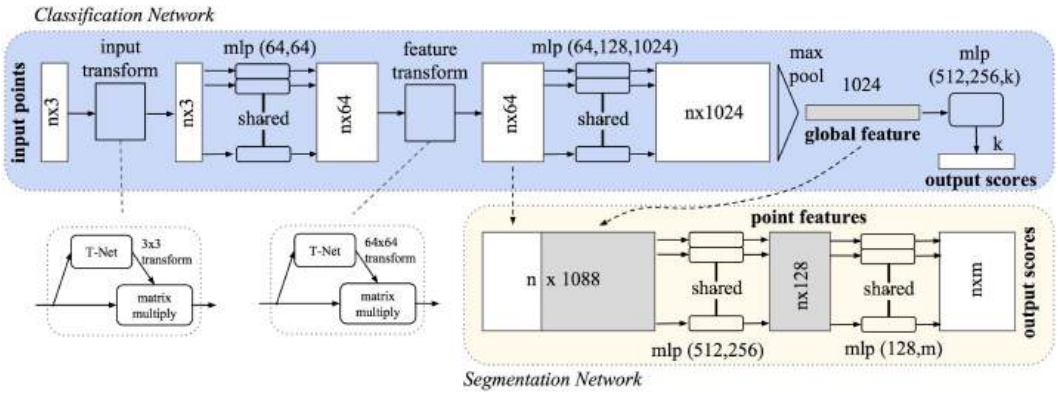

Figure 2.3 Architecture of PointNet [5]

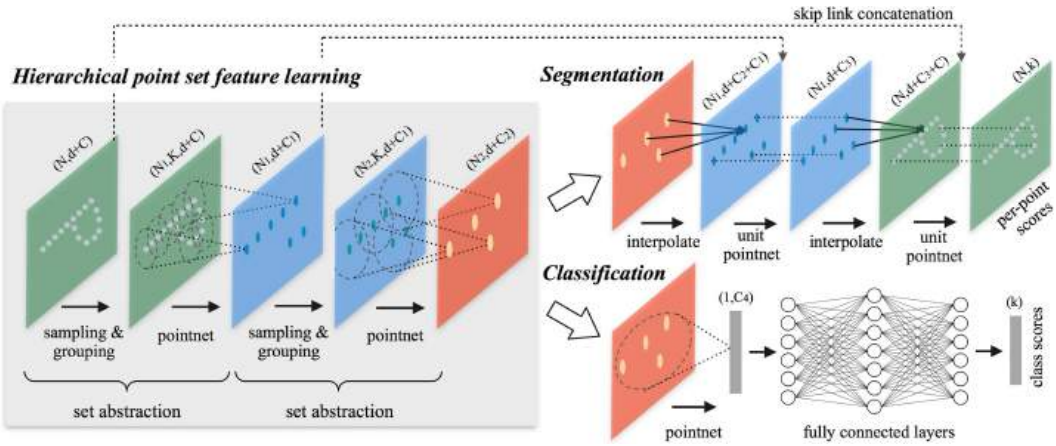

Figure 2.4 Architecture of PointNet++ [6]

2.5.2 Point Transformer V3: Self-Attention for Geometric Reasoning

Point Transformer V3 (PTV3) [7] represents a major advancement in deep learning for 3D data by directly addressing the shortcomings of earlier point-based models [5, 6]. Instead of relying on static neighborhood definitions and hand-crafted aggregation functions, it introduces a dynamic, learnable self-attention mechanism designed to capture complex geometric relationships within point clouds better.

The architecture begins with point cloud serialization, which transforms the unstructured point cloud data into a structured sequence using space-filling curves such as Z-order or Hilbert curves. This conversion enables efficient and scalable learning by imposing a spatial order on the data (see Figure 2.5).

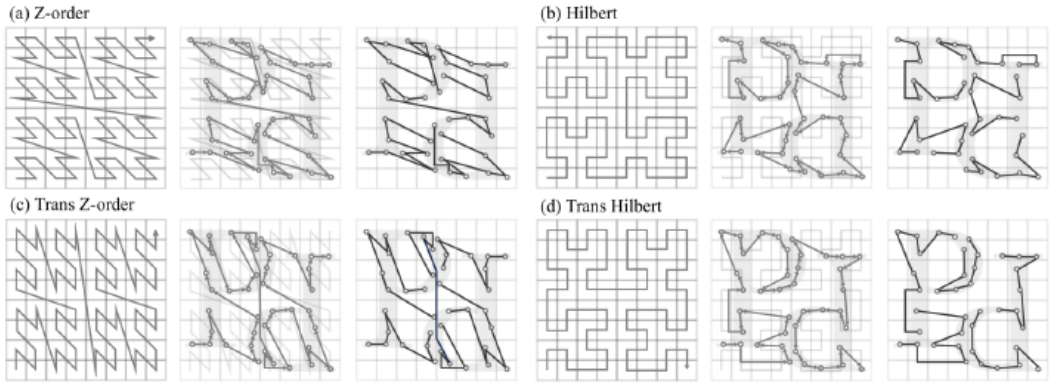

Figure 2.5 Point Cloud Serialization: Illustration of the process that transforms unstructured point clouds into structured sequences using space-filling curves (e.g., Z-order or Hilbert curve), enabling efficient processing and learning. [7]

Once serialized, the points are divided into patches through a patch grouping strategy. This grouping supports localized attention operations by allowing each patch to capture fine-grained geometric features within its spatial neighborhood (see Figure 2.6).

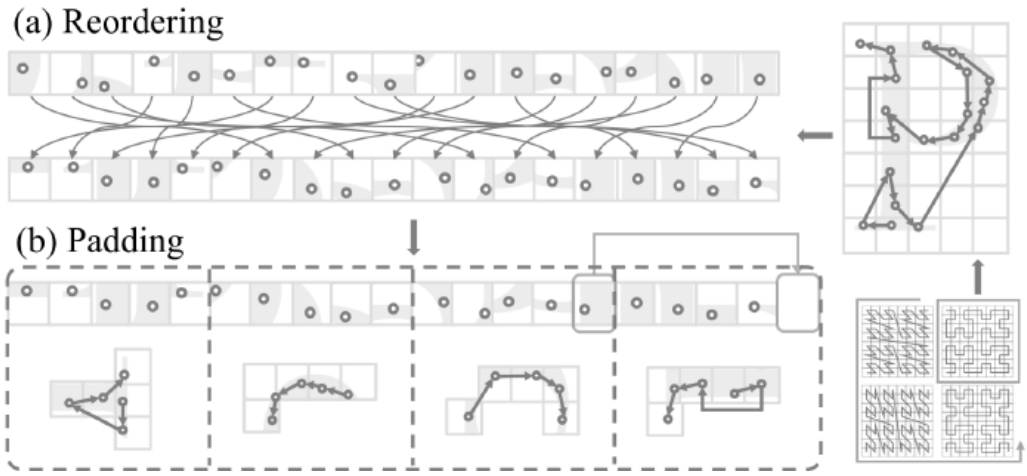

Figure 2.6 Patch Grouping: Visualization of the grouping strategy where serialized point cloud sequences are divided into patches. Points within each patch are aggregated for localized attention computation. [7]

To enrich the model's understanding of global context and reduce overfitting to rigid spatial patterns, Point Transformer V3 introduces a range of patch interaction strategies. These include shift-dilation, shift-patch, shift-order, and shuffle-order techniques. By dynamically adjusting how patches interact and share information, the model learns to integrate both local geometry and non-local semantic cues (see Figure 2.7).

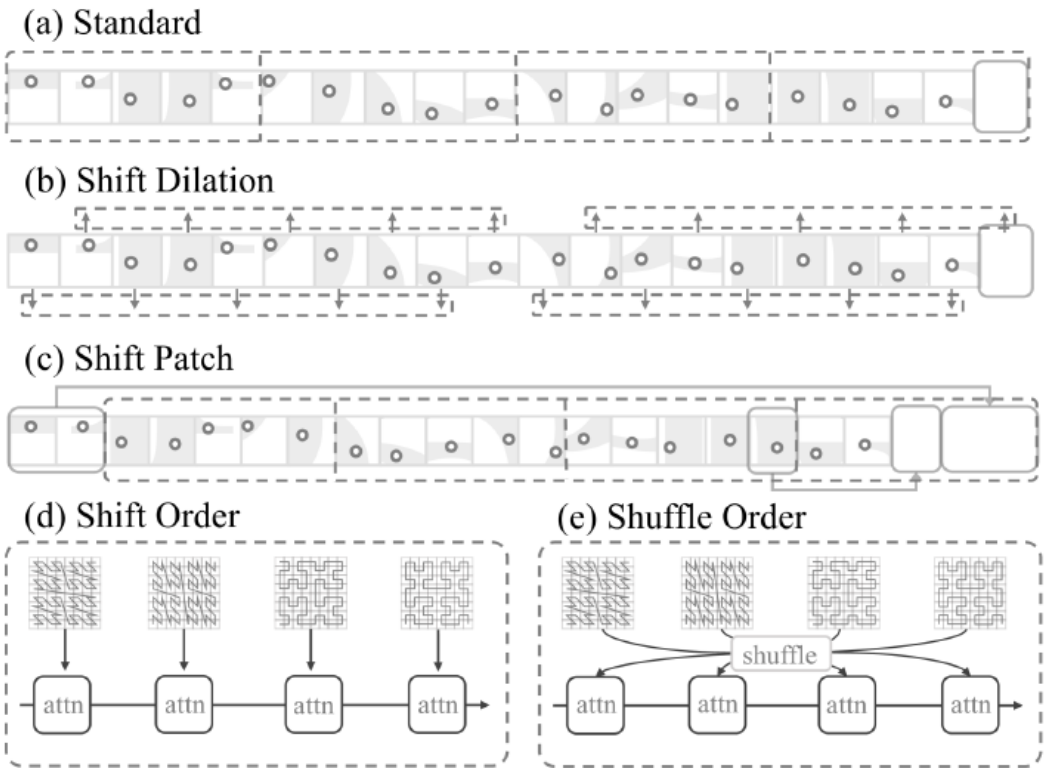

Figure 2.7 Patch Interaction Strategies: Comparison of various patch interaction methods, including shift-dilation, shift-patch, shift-order, and shuffle-order, which facilitate feature fusion across patches to capture both local and global context. [7]

At the core of the network are the Point Transformer Blocks, which implement a query-key-value attention framework. Unlike traditional methods that aggregate features solely by proximity, Point Transformer V3 incorporates relative positional encoding into its attention weights. This design allows the model to factor in not only feature similarity but also spatial distance and directionality, critical for capturing anatomical continuity and structural context in complex 3D forms.

The model's overall architecture adopts a hierarchical encoder-decoder format (see Figure 2.8). Self-attention layers are used in both the down-sampling and up-sampling stages, replacing traditional pooling operations. This attention-based design helps preserve spatial coherence and contextual detail across multiple scales, which is especially valuable in tasks involving complex geometries or partially occluded objects, for example, segmenting furniture or structural elements in cluttered indoor environments.

Beyond its architectural innovations, Point Transformer V3 demonstrates strong scalability and generalization capabilities. Its ability to handle large-scale point clouds with high accu-

racy has been validated in diverse applications, including indoor scene reconstruction, object detection in robotics, and outdoor LiDAR segmentation for autonomous driving. These results highlight the model's robustness under varying spatial resolutions and its effectiveness even when training data is limited or partially labeled.

Recently, its potential has extended into the medical domain. In a study focused on digital dentistry, researchers applied a Point Transformer V3-inspired architecture to detect anatomical landmarks on intraoral 3D scans [40]. Despite challenges such as small dataset sizes and high anatomical variability, their method effectively learned meaningful geometric and anatomical features from raw point cloud data. Their results from the 3DTeethLand Grand Challenge at MICCAI 2024 demonstrate that Point Transformer V3 can be adapted for fine-grained localization tasks in clinical settings, underscoring its versatility and growing relevance in medical image analysis.

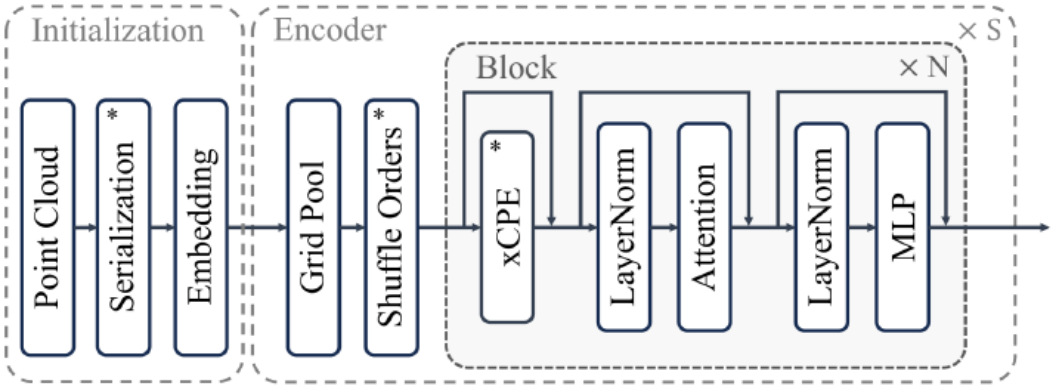

Figure 2.8 Overall Architecture of Point Transformer V3: Schematic diagram of the hierarchical encoder-decoder structure of PTv3, illustrating the integration of self-attention layers within down-sampling (feature abstraction) and up-sampling (feature propagation) stages. [7]

2.6 Intraoperative Data

Deep learning models typically require extensive annotated datasets. However, collecting such data during real surgeries remains challenging due to ethical, logistical, and technical constraints. In spine surgery, particularly posterior approaches, obtaining clear exposure of vertebral structures is limited, and key anatomical landmarks such as the spinous and transverse processes are often only partially visible. These limitations complicate the creation of reliable ground truth labels and hinder the scalability of model training based solely on clinical data.

A notable attempt to address this gap is the publicly available SpineDepth dataset [8]. It

was designed to support the development of deep learning algorithms for spinal shape reconstruction and intraoperative navigation. The dataset includes over 299,000 RGB-D frames captured during simulated pedicle screw placement procedures performed on ten cadaveric lumbar spine specimens. However, it is important to note that these specimens do not represent scoliosis cases. Data collection was conducted in a controlled surgical setting using two synchronized depth cameras. Each frame is paired with high-resolution vertebral meshes and pose information from an optical tracking system.

While valuable, this setup is challenging to replicate (see Figure 2.9). It relies on attaching optical markers to each vertebra and requires seamless calibration and integration between the tracking system and dual RGB-D cameras—conditions that are difficult to achieve outside of a specialized lab environment, particularly when using an artificial spine model such as a sawbone phantom.

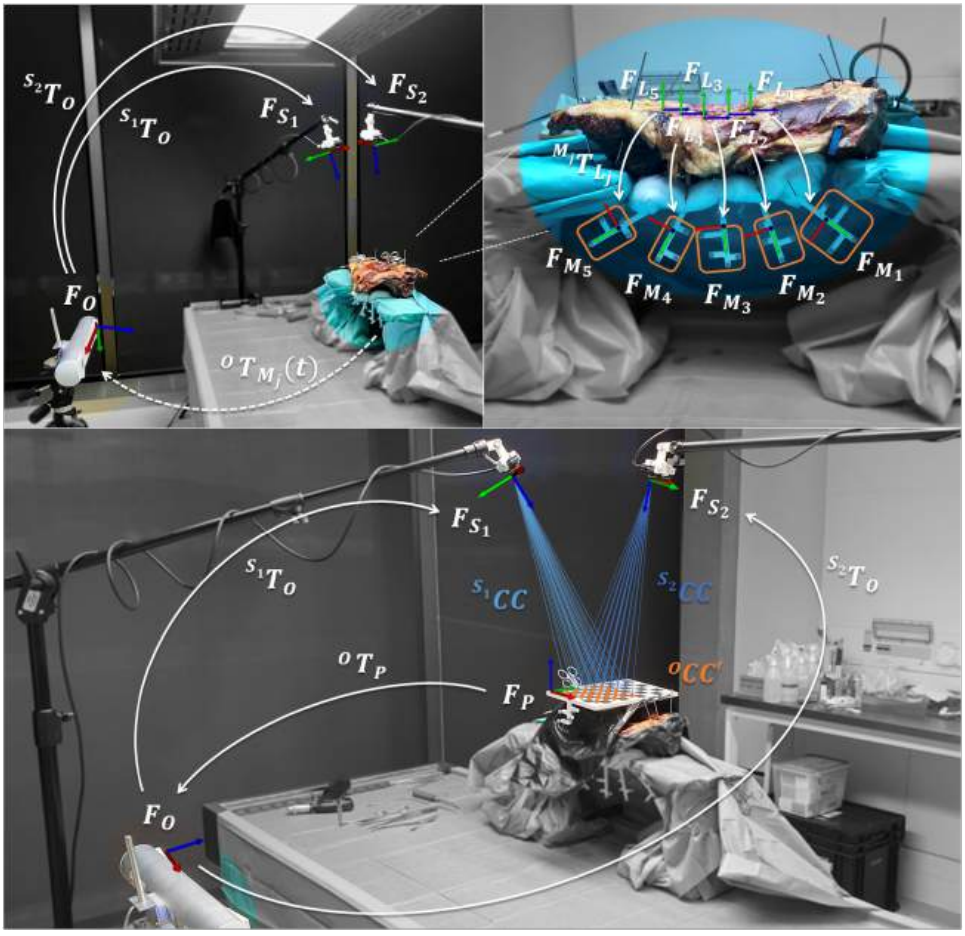

Figure 2.9 Experimental setup used in the SpineDepth dataset. *Adapted from* liebmann et al. (2021) [8].

While SpineDepth offers valuable intraoperative-like data with dense 3D annotations, it also comes with important limitations. First, because the specimens are cadaveric, they lack physiological movement and natural tissue compliance, which reduces realism compared to live surgical conditions. Second, due to tissue preservation processes, the color and texture of the vertebrae differ substantially from those observed during actual surgeries. This results in limited color contrast between vertebrae and background tissues, as confirmed by RGB distribution analyses. Consequently, models trained on SpineDepth may struggle to generalize to real surgical environments where tissue appearance is more complex and variable.

These challenges have motivated the integration of semi-synthetic or simulation-based data into model development pipelines. Synthetic data offers a controlled and scalable alternative that can replicate the spatial and visual complexity of surgical environments while enabling precise ground truth annotation. Prior work supports the feasibility of synthetic datasets for surgical AI tasks. For instance, Yoon et al. introduced a virtual surgery simulation framework to train segmentation networks on synthetic RGB images, effectively compensating for the lack of annotated intraoperative data [41]. Other studies, such as VisionBlender [42], have demonstrated the utility of Blender-generated scenes in creating labeled datasets for robotic and laparoscopic surgery. More recently, Pérez et al. introduced TLSynth [43], a Blender-based add-on that simulates terrestrial laser scanner (TLS) point clouds with realistic scanning noise and density, underscoring the practical value of synthetic point clouds in computer vision research. These precedents validate the use of simulation-based tools for generating annotated 3D datasets in domains with limited real-world data availability.

CHAPTER 3 RATIONALS AND OBJECTIVES

In the previous chapter, we reviewed key concepts related to Adolescent Idiopathic Scoliosis (AIS), focusing on the clinical workflow of posterior spinal fusion (PSF), a standard surgical intervention. This multi-stage procedure typically involves the fixation of pedicle screws, followed by deformity correction using contoured rods, and concludes with postoperative radiographic assessment to evaluate alignment and surgical outcomes.

Although PSF is an established and effective approach for severe AIS correction, it remains a complex and high-risk procedure. The surgery's success hinges on precise intraoperative decision-making, particularly during deformity correction, where suboptimal realignment can result in persistent deformity or necessitate revision surgery. Image-guided surgery systems (IGSS) have been introduced to enhance surgical precision, relying primarily on intraoperative imaging registered to preoperative scans. However, our synthesis of current literature and clinical practice reveals several critical limitations in existing IGSS strategies, particularly for real-time spine tracking:

- While many systems assist in pedicle screw placement, few offer dynamic, quantitative feedback on vertebral alignment during correction maneuvers. This limits the surgeon's ability to assess and adjust alignment in real time.
- Intraoperative fluoroscopy and CT scans are frequently used to update anatomical context, but they expose patients and surgical teams to significant levels of radiation. This restricts the frequency of updates and is particularly problematic in pediatric cases.
- Most current segmentation and registration methods rely on RGB-D inputs, which are limited in resolution and depth accuracy. However, new-generation surgical cameras, such as those from 7D Surgical, offer dense and accurate point clouds as direct output. These radiation-free devices are already being used in many operating rooms, including at CHU Sainte-Justine. Yet, current deep learning approaches do not take full advantage of this raw data, missing an opportunity for higher-resolution, real-time anatomical segmentation.
- Robust deep learning models, especially those designed for 3D segmentation, require large volumes of annotated intraoperative data. Such datasets are difficult to obtain due to ethical, logistical, and annotation challenges. In our case, only a small number

of usable real surgical recordings were available, and none had complete landmark annotations.

For this work, we made three hypotheses:

- The advanced attention mechanisms of the Point Transformer V3 architecture would outperform prior approaches in 3D point cloud segmentation of vertebrae, particularly when compared to previous studies that utilized RGB-D data [1].
- The SpineDepth dataset alone is insufficient to capture the anatomical and visual variability required for robust generalization to diverse intraoperative scenarios, especially concerning scoliosis cases.
- Supplementing real data with carefully designed semi-synthetic datasets, when combined with appropriate augmentation strategies, can significantly improve segmentation performance on real intraoperative data.

These challenges motivate the development of a novel, data-driven solution to improve intraoperative guidance for spinal deformity correction. While previous work in our team [2] has focused on vertebra segmentation in pre-operative MRI scans and provided a proof of concept for the registration from pre-operative to synthetic intraoperative data, this master's project aims at an automatic vertebrae segmentation from intraoperative structure light scan. Our proposed research introduces a workflow that leverages radiation-free 3D point clouds, state-of-the-art deep learning architectures, and a semi-synthetic data generation pipeline. Our specific objectives are:

- To identify and evaluate deep learning models capable of semantically segmenting critical vertebra landmarks, particularly the spinous and transverse processes, from raw 3D point clouds acquired intraoperatively.
- To address the limited availability of annotated intraoperative data, a Blender-based semi-synthetic data pipeline was developed.
- To train and evaluate the segmentation model across multiple data domains, including real cadaveric point clouds, semi-synthetic simulations, and phantom-based acquisitions, to assess its generalization capability. By systematically combining and testing different datasets, we investigate how synthetic data improves model robustness under varying anatomical presentations and sensor conditions.

Together, these objectives contribute to a robust framework for non-irradiating, anatomically informed surgical assistance. By leveraging cutting-edge 3D sensing technologies already present in the operating room and augmenting them with synthetic data grounded in clinical realism, our approach aims to deliver accurate, reproducible spinal measurements without added imaging or radiation exposure. This, in turn, supports faster, more informed intraoperative decisions, improves preoperative planning, and facilitates clearer communication between surgeons and patients, ultimately enhancing the accessibility of real-time spine alignment tracking.

CHAPTER 4 METHODOLOGY

This chapter details the methodology developed to achieve the three main research objectives. Our workflow is divided into three key stages. First, we investigate the performance of a deep learning model for semantic segmentation using raw 3D point cloud data. This includes preprocessing and annotation of an open-source dataset and training a transformer-based model for vertebrae landmarks segmentation. Second, we develop a semi-synthetic data generation pipeline to simulate intraoperative spinal exposures, including data collection using both physical models and Blender-based scene generation, and training experiments combining multiple data sources. Finally, we evaluate the trained model on real intraoperative data acquired from clinical scoliosis surgeries to assess the feasibility of deploying the method in a surgical context. Each stage is described in the sections that follow.

4.1 Semantic Segmentation on Public Dataset: SpineDepth

4.1.1 Dataset Overview

As an initial benchmark, we evaluated our model on the publicly available SpineDepth dataset [8]. This dataset was collected during simulated pedicle screw placement procedures performed on ten cadaveric lumbar spine (L1-L5) specimens in a controlled surgical environment. It contains over 299,000 RGB-D frames captured from multiple viewpoints using two synchronized depth cameras. Each frame includes accurately registered 3D vertebral meshes and corresponding pose data, obtained via a high-precision optical tracking system. The dataset is designed for the development and validation of deep learning methods targeting spinal shape reconstruction and intraoperative navigation.

4.1.2 Point Cloud Extraction and Annotation

To convert RGB-D data into 3D representations suitable for deep learning, we used the Stereolabs ZED Python API to extract point clouds from each frame, preserving both spatial coordinates and color information. To reduce computational complexity and focus on the relevant anatomy, we defined a bounding box centered on each vertebra to isolate the region of interest (ROI), namely, the spine specimen in each frame. This step effectively filtered out background points and preserved only those within the surgical field. Ground truth labels were generated by aligning the vertebral mesh models to each frame using the provided ground truth transformation matrices (see Figure 4.1).

To determine whether each point in the cloud lies inside or outside the vertebral surfaces, we applied a Signed Distance Field (SDF)-based annotation method. The SDF computes the shortest distance from any point $\mathbf{p} \in \mathbb{R}^3$ to the surface of a 3D object, with the sign indicating spatial relationship to the surface:

$$\mathrm{SDF}(\mathbf{p}) = \begin{cases} -\min_{\mathbf{q}\in\partial\mathcal{M}} \|\mathbf{p}-\mathbf{q}\|, & \text{if } \mathbf{p} \in \mathcal{M}_{\text{inside}} \\ \min_{\mathbf{q}\in\partial\mathcal{M}} \|\mathbf{p}-\mathbf{q}\|, & \text{if } \mathbf{p} \in \mathcal{M}_{\text{outside}} \end{cases} \tag{4.1}$$

where $\partial\mathcal{M}$ denotes the surface boundary of the vertebral mesh $\mathcal{M}$, and $\|\cdot\|$ is the Euclidean distance. A negative SDF value indicates the point lies inside the mesh, zero means it is on the surface, and a positive value denotes it is outside.

Points were labeled based on this signed distance: if the SDF value of a point was less than or equal to a predefined threshold (e.g., 0 mm), it was considered part of the corresponding vertebra and assigned its anatomical class (1–5); otherwise, it was labeled as background (class 0). This method provides a precise and geometry-aware way to annotate complex anatomical structures in point clouds (see Figure 4.1).

Given the scale of the dataset (over 299,000 frames), annotation was parallelized across multiple CPU cores to improve efficiency. Additionally, substantial disk storage was required to accommodate the large number of annotated point clouds, which were stored as NumPy arrays for rapid loading during training.

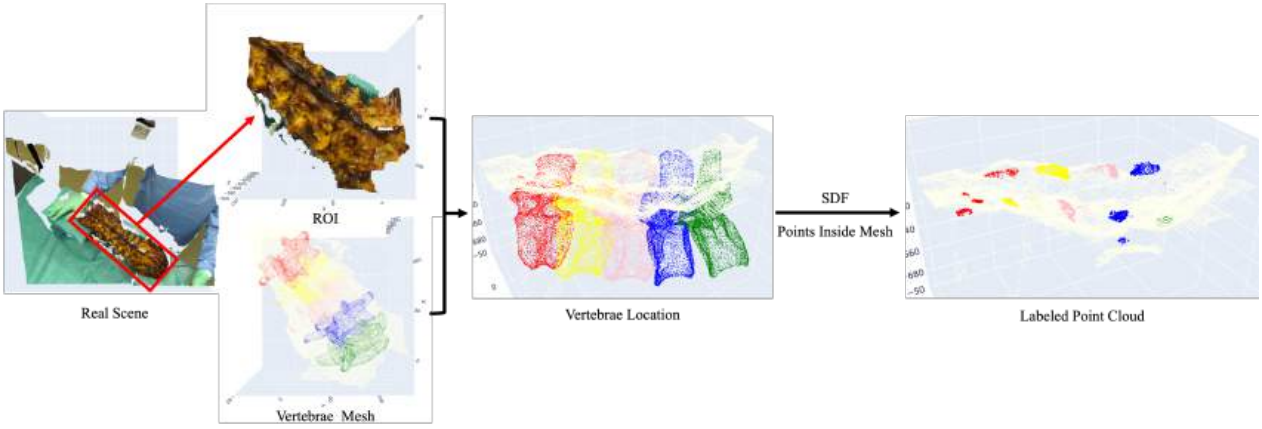

Figure 4.1 Overview of SpineDepth dataset preparation

4.1.3 Model Architecture and Training Strategy

We trained a Point Transformer V3 model [7], a state-of-the-art model designed for point cloud segmentation tasks. This architecture was adapted to classify vertebral landmarks in intraoperative 3D point clouds (see Figure 4.2). Specifically, the model was trained to identify exposed spinous and transverse processes within each frame. To enable per-point classification, we appended a final convolution layer followed by a softmax activation to the decoder output. Depending on the segmentation task, two model variants were trained: one for binary classification (vertebra vs. background) and another for multi-class segmentation, distinguishing five vertebral levels and one background.

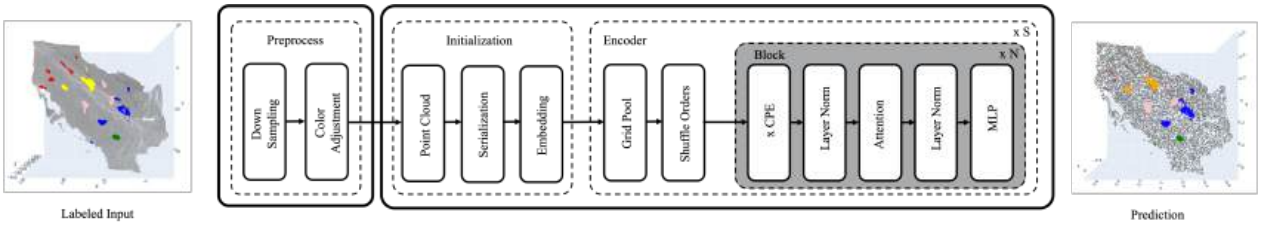

Figure 4.2 Modified Model Architecture.

Data Sampling and Class Imbalance Handling

In each frame, vertebrae points constitute approximately 20% of the point cloud, while background (non-vertebrae) points account for the remaining 80%. To respect this natural class distribution and manage memory constraints, we applied stratified downsampling to select 10,000 points per frame while preserving the original label ratio (see Figure 4.3).

To further address class imbalance during training, we employed a weighted focal loss. The focal loss down-weights well-classified examples and focuses training on hard, misclassified samples. The loss for a predicted logit $\hat{y} \in \mathbb{R}^C$ and target label $y \in \{0, 1, \ldots, C-1\}$, with per-class weights α_c, is defined as:

$$\mathcal{L}_{\text{focal}}(y, \hat{y}) = -\alpha_y (1 - p_y)^\gamma \log(p_y) \tag{4.2}$$

where: $p_y = \text{softmax}(\hat{y})_y$ is the predicted probability for the correct class, γ is the focusing parameter, set to $\gamma = 5.0$ in our experiments, since the dataset is highly imbalanced, α_y is the weight for class y, computed based on class frequency.

The class weights α_c were derived from the training set label distribution before sampling. Specifically, we first calculated the normalized frequency f_c for each class c, and then defined

the weight as:

$$\alpha_c = \left(\frac{\max(f)}{f_c} \right)^{1/3} \tag{4.3}$$

This formulation assigns higher weights to minority classes (e.g., vertebrae) and lower weights to majority classes (e.g., background), in a smoothed manner using a cube root to avoid excessive weighting. These weights are then applied dynamically in the loss computation for each sample, according to its class label.

Color Distribution Analysis and Augmentation

The texture and spatial characteristics of cadaveric anatomy differ from live surgical conditions. Due to tissue shrinkage and discoloration in fresh-frozen cadaveric specimens, both the color and geometry of the captured data can deviate significantly from real intraoperative scenes, making segmentation more challenging. An RGB color distribution analysis showed minimal separability between vertebrae and background classes (see Figure 4.4).

To improve model generalization, we applied color augmentation to vertebrae points during training. Specifically, in 50% of training batches, the RGB values of vertebrae points were randomly shifted toward a reference distribution resembling real lumbar surgical anatomy. The transformation strength was randomized for each point by interpolating between the original color and the target tone. This augmentation was applied only during training; meanwhile, all point clouds were normalized to ensure consistent input scale during both training and testing.

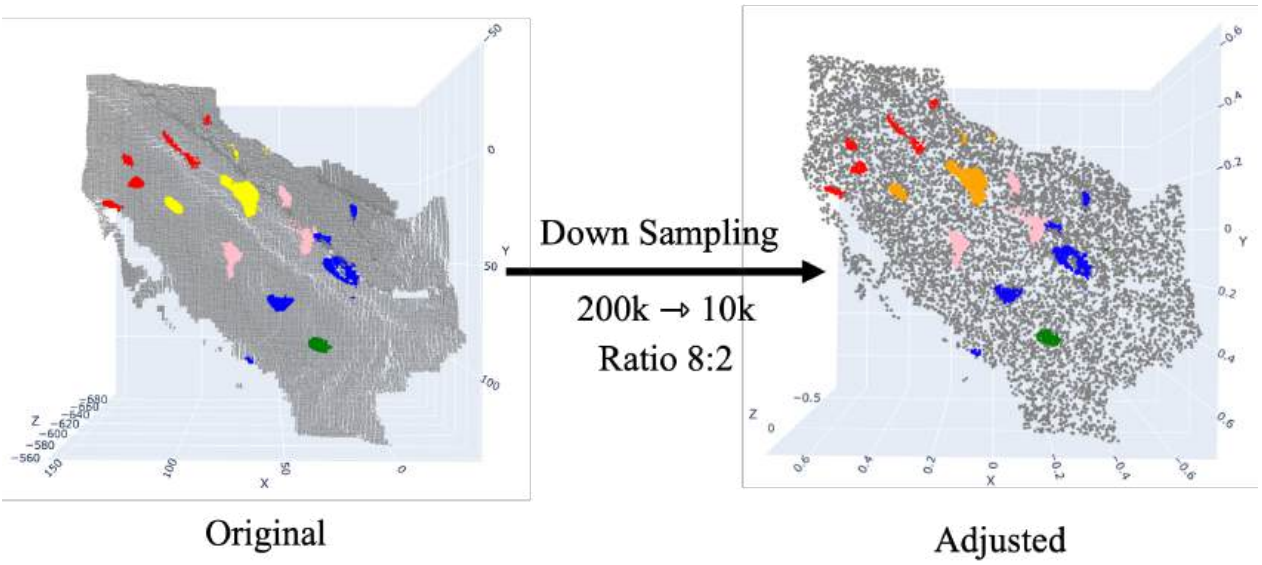

Figure 4.3 Visualization of the downsampling process, reducing the point cloud from approximately 200,000 points to 10,000 points. An 80/20 ratio was applied to preserve the original distribution between background and vertebrae points.

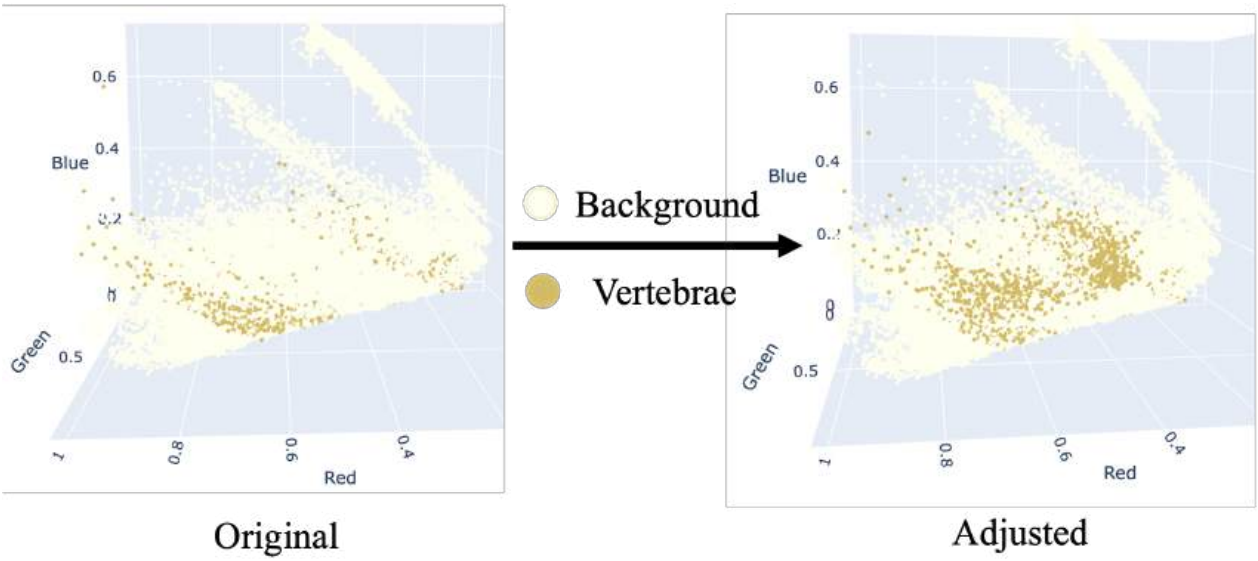

Figure 4.4 Visualization of color augmentation. The original RGB space distribution (left) is transformed to an adjusted version (right), where the two clusters, background and vertebrae, are more distinctly separated.

4.1.4 Cross-Validation and Evaluation

A leave-one-out cross-validation scheme was employed to ensure robust evaluation across the eight specimens in the SpineDepth dataset. Performance was measured using the Dice Similarity Coefficient (DSC), consistent with the metrics and evaluation protocol used in prior studies [1]. Our results were directly compared against RGB-D-based segmentation methods reported in the literature [1], allowing for a fair assessment of the benefits of using raw 3D point clouds as model input. A comparative analysis of segmentation accuracy is presented in Table 5.1.

4.2 Semantic Segmentation on Semi-Synthetic Dataset

The SpineDepth dataset, while valuable, consists of cadaver lumbar specimens that were rigidly fixated in a test setup with minimal to no physiological motion. As such, it does not reflect the dynamic anatomical variations observed in scoliosis cases, particularly in the thoracic region (see Section B). Furthermore, due to the limited availability of annotated intraoperative 3D point cloud data, we constructed a semi-synthetic dataset generation pipeline that integrates virtual simulations and physical acquisitions. This approach enables the creation of diverse and anatomically realistic point clouds, tailored to mimic the appearance and spatial characteristics of radiation-free 7D surgical outputs [11]. In this section, we describe the construction of three datasets used for model training and cross-dataset validation.

4.2.1 Semi-Synthetic Data Generation

Dataset A: Blender-Simulated Surgical Scenes Using Real Scoliosis Cases

This dataset was created using stereo-radiographic 3D reconstruction of real scoliosis patients' spine models and procedural scene generation in Blender to simulate realistic surgical exposures. The source data originates from a previous study [2], which obtained ethical approval from CHU Sainte-Justine to collect pre- and post-operative 3D spine models from 49 Adolescent Idiopathic Scoliosis (AIS) patients. These 3D models (4.5 a) were reconstructed from biplanar radiographs (posterior and lateral views). A trained physician manually annotated key anatomical landmarks on each vertebra (T1–T12, L1–L5), and the landmark coordinates were provided as labeled points in JSON format.

Using the STL spine models and landmark annotations, we generated virtual surgical scenes in Blender. The labeled landmarks were used to define procedural surfaces that partially cover the vertebrae, simulating varying levels of surgical exposure (4.5 b). Geometry node

operations were applied to distribute points over these defined surfaces, producing synthetic point clouds that mimic the visual field during posterior spinal fusion (4.5 c). Each scene was exported as a PLY file containing 3D spatial coordinates and surface normals.

To enhance visual realism, RGB color values were assigned to each point based on intensity distributions observed in real surgical point clouds from the CHU Sainte-Justine dataset. This step ensures that the synthetic point clouds not only capture anatomical geometry but also exhibit color characteristics that closely match real intraoperative imaging conditions.

Post-processing was conducted in Python (VSCode) to refine the raw Blender-exported point clouds. First, a signed distance field (SDF) was computed between the mesh surfaces of the vertebrae and the point cloud to identify and remove points that unrealistically intersected the vertebrae. Next, surface normals were used to determine the visibility of points from a posterior viewpoint, retaining only those that lie above the simulated soft tissue surfaces and are visible from the typical surgeon's field of view. The filtered point cloud was then annotated using the same SDF-based labeling strategy as in Section 4.1, assigning binary labels: vertebra (label=1) and background (label=0).

The complete pipeline for Dataset A is shown in Figure 4.5. The final point clouds are anatomically consistent and photorealistic, providing a robust testbed for assessing the impact of synthetic-only augmentation on model performance.

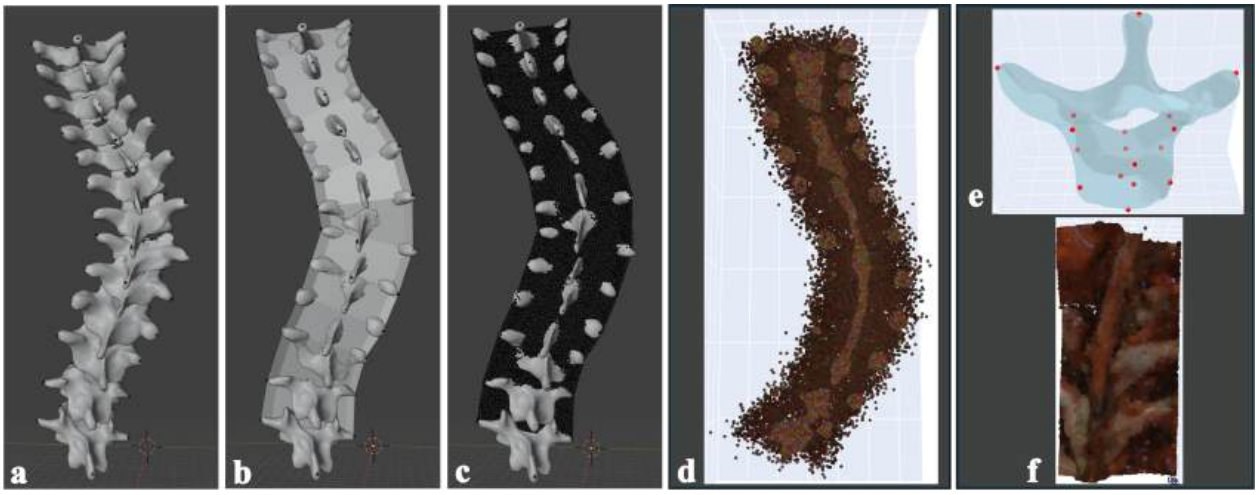

Figure 4.5 Dataset A: Blender-simulated surgical scene using real scoliosis cases. (a) 3D spine model reconstructed from biplanar X-ray of an adolescent idiopathic scoliosis (AIS) patient. (b) Procedural surface creation guided by vertebral landmarks (see e). (c) Geometry node operations applied to distribute surface points. (d) Assigning realistic color mapping based on intensity statistics derived from real surgical data (see f) using VS Code. (e) Annotated vertebral landmarks. (f) Color example of Vertebrae and surrounding tissues in intraoperative scene.

Dataset D: Physical Phantom Data Acquired with the 7D Surgical System

This dataset was acquired using a commercially available artificial spine phantom[1] and the 7D Surgical System. This system allows radiation-free acquisition of intraoperative point clouds and offers tools for real-time anatomical registration. The goal of this dataset was to simulate realistic spinal deformity correction scenarios using a physical phantom and to evaluate the performance of the segmentation model on point-cloud data acquired from the same device used in actual clinical surgeries.

The workflow began with preoperative CT imaging of the artificial spine. Using the 7D software interface, we manually selected three anatomical landmarks, one spinal process and two transverse processes, on each vertebra from T1 to T12 in the CT volume. These points, along with vertebral level labels, served as reference anchors to establish vertebral-specific coordinate frames.

To simulate intraoperative conditions, spinal curvature was manually introduced by adjusting the alignment of the phantom, and metallic components were added to emulate the presence of surgical instruments. The FLASH feature of the 7D system was then used to capture

[1] `https://www.sawbones.com/full-spine-solid-foam-vertebral-column-sacrum-1323.html`

high-resolution, radiation-free 3D surface scans of the phantom's posterior anatomy under various deformity states (see Figure 4.7). Each FLASH capture was followed by manual re-registration of the vertebrae: the same three anatomical landmarks were selected on the captured point cloud for each vertebra, approximately matching their counterparts from the CT reference.

To validate the accuracy of the transformation, a tracked probe was used to point to known anatomical locations on the phantom, confirming the overlay alignment on the 7D system interface. After verifying registration, further curvatures were introduced, and the process was repeated, allowing the generation of 48 point clouds across a range of spinal deformities. This iterative acquisition strategy ensured anatomical variability and realism while maintaining tight control over registration accuracy.

All transformation matrices and point cloud captures were logged by the 7D system and later exported with timestamps. The post-processing pipeline consisted of two main steps. First, the preoperative CT DICOM volumes were manually segmented into individual vertebrae (T1–T12), as shown in Figure 4.6. Due to slight anatomical differences between the CT model and the physical phantom, minor manual adjustments were necessary to accurately align the segmentations. Second, the recorded transformation matrices were applied to each vertebra mesh and used in conjunction with signed distance field (SDF) calculations to annotate each point in the 7D-captured clouds. Points inside the vertebral meshes were labeled as class 1 (vertebra), and all others as class 0 (background), as illustrated in Figure 4.8.

This dataset closely mirrors the point clouds captured intraoperatively by the 7D Surgical System and provides high-fidelity, radiation-free annotations. It plays a critical role in validating the segmentation model's ability to generalize across domains and surgical scenarios.

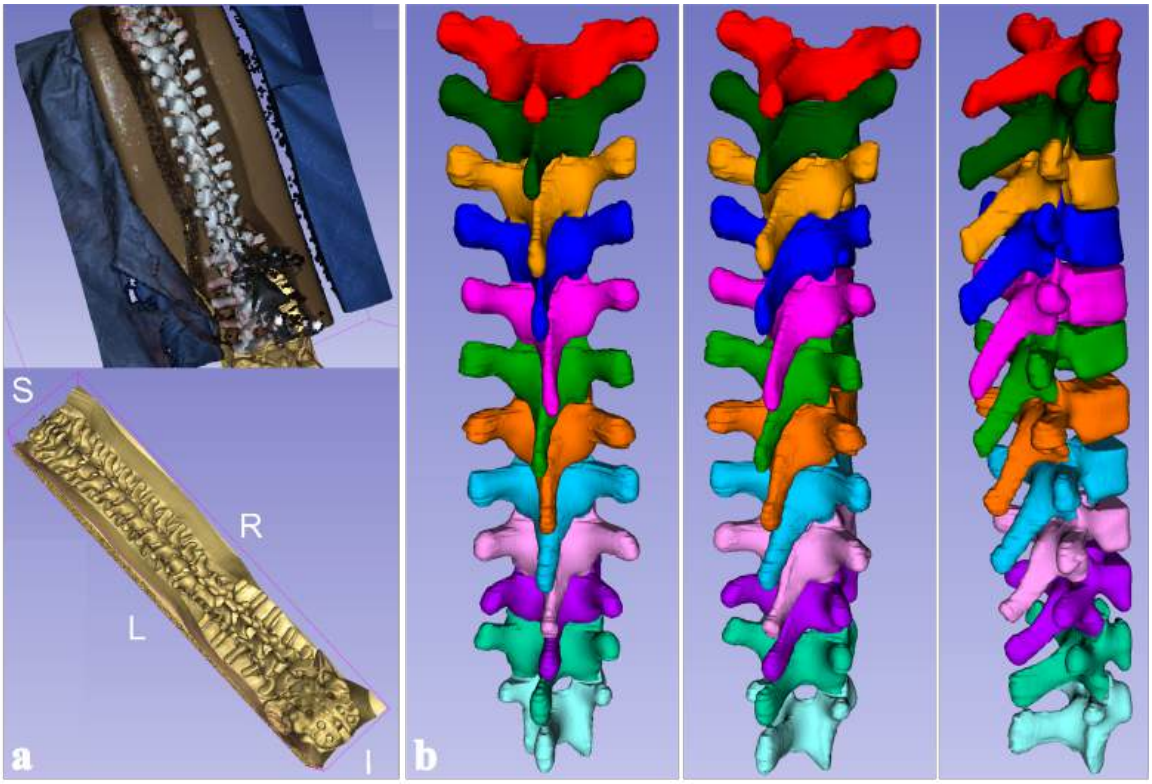

Figure 4.6 Dataset D: Manual segmentation of preoperative CT scans from the spine phantom in 7D Surgical System. Each vertebra (T1–T12) was segmented individually for accurate vertebral level alignment with the captured 3D point clouds.

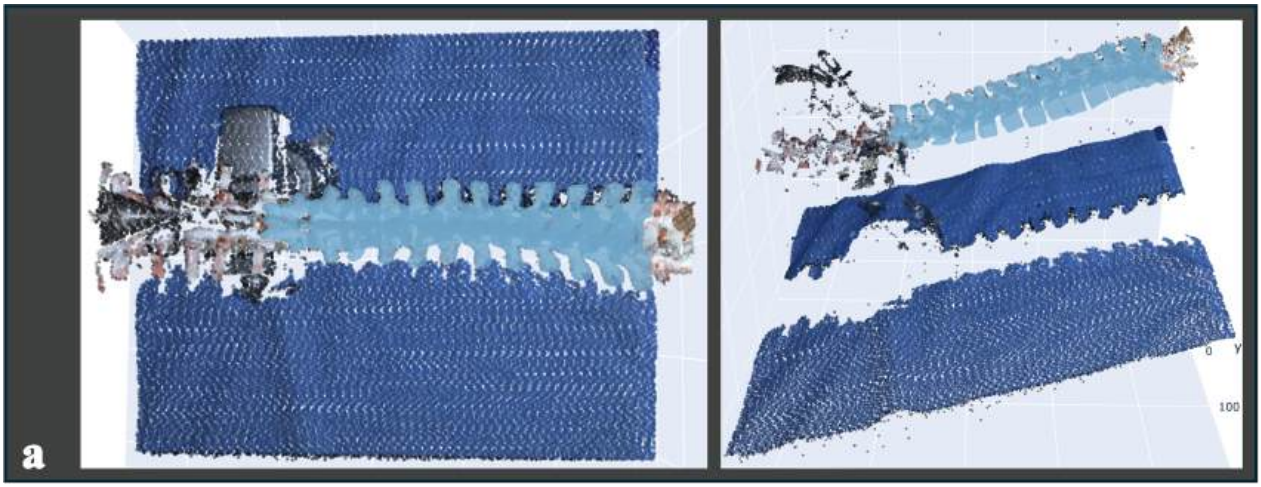

Figure 4.7 Dataset D: CT Registration vertebral models to 7D Surgical captured point clouds. Anatomical landmarks were matched to compute rigid transformations for each vertebra.

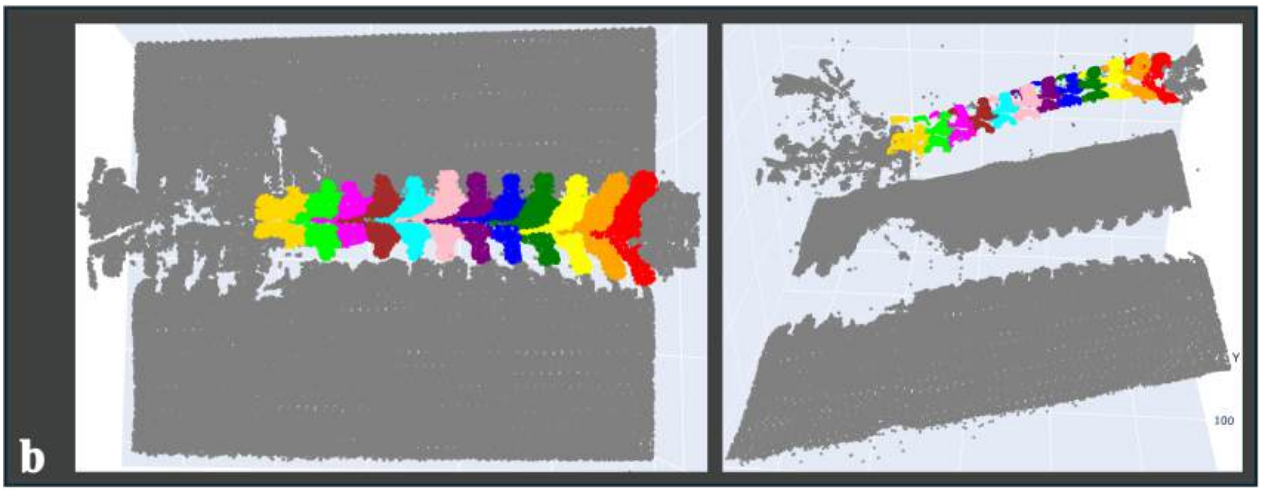

Figure 4.8 Dataset D: Illustration of the annotated point clouds captured from 7D Surgical using signed distance field (SDF) to assign vertebra (label = 1) and background (label = 0).

Dataset P: RGB-D Capture of Phantom Spine for Blender-based Simulation

The third dataset was developed using the same artificial spine phantom as in Dataset D but acquired using a consumer-grade RGB-D camera (Intel RealSense), which uses stereoscopic depth sensing technology. This setup was designed to evaluate the feasibility of low-cost sensors for generating semi-synthetic training data, particularly in environments where access to surgical-grade devices is limited.

To simulate varying anatomical conditions, the phantom spine was manually adjusted to represent six distinct spinal curvatures, ranging from neutral alignment to severe deformity. For each configuration, two black markers were manually painted on the laminae of each vertebra to serve as reference points for surface reconstruction in Blender. These markers were automatically detected via blob detection in the RGB images. Their corresponding 3D positions were computed using the RealSense camera's intrinsic parameters and stored in JSON format.

From a fixed posterior viewpoint, 20 RGB-D images were captured per curvature configuration. Each frame was reprojected into 3D space to generate raw point clouds using depth and RGB data. These captured point clouds were fused with anatomical surfaces generated in Blender using the marker locations as procedural guides, following the same approach described in Dataset A. The resulting hybrid scenes, composed of real sensor input and synthetic geometry, were exported as PLY files for further processing.

To determine visibility and assign labels, a ray-casting algorithm was applied to the merged

point clouds. Since the Blender-generated vertices did not contain color information, RGB values served as a proxy for semantic labeling: points with RGB values were assumed to correspond to vertebral surfaces and were labeled as vertebra (label = 1), while points without RGB values were treated as synthetic background and labeled as non-vertebra (label = 0).

This hybrid approach enabled cost-effective annotation by combining sensor-acquired depth data with procedurally generated meshes. It offers a practical and accessible alternative for generating annotated datasets without the need for specialized surgical navigation systems. The overall data pipeline is illustrated in Figure 4.9.

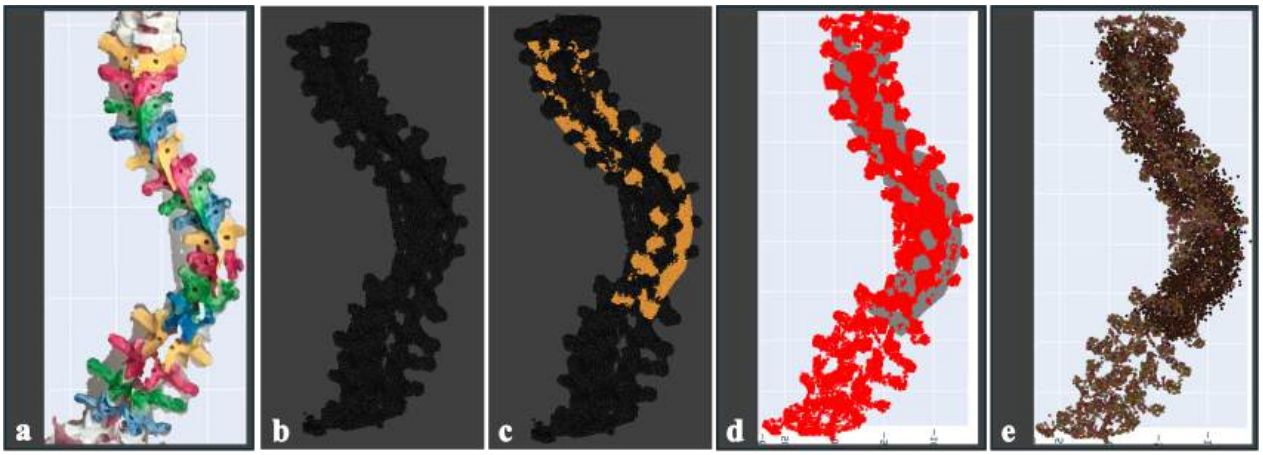

Figure 4.9 Dataset P: RGB-D–based acquisition and hybrid simulation of the spine phantom. (a) Raw point cloud captured using the Intel RealSense camera. (b) Imported point cloud and landmarks visualized in Blender. (c) Procedural surface generation based on black anatomical markers shown in (a). (d) Ray-casting is used to determine visibility and assign vertebra vs. background labels. (e) Colorization based on statistical color profiles used in Dataset A for visual realism.

Table 4.1 Summary of datasets used in this study.

Dataset	**Spine Source**	**Deformity Scoliosis**	**Acquisition Sensor**	**Soft Tissue**	**Vertebrae Levels**
S	Cadaveric	None	RealSense	Cadaveric	L1–L5
A	3D Model	Real Case	Simulated	Blender Simulated	T1–T12
D	Phantom	Simulated	7D Surgical	Physically Simulated	T1–T12
P	Phantom	Simulated	RealSense	Blender Simulated	T1–T12

4.2.2 Training Strategy and Data Augmentation

Building on findings from Section 4.1.3, we trained Point Transformer V3 on each semi-synthetic dataset independently to evaluate their standalone and complementary utility for semantic segmentation of vertebrae in intraoperative-like 3D point clouds. Each dataset required tailored augmentation strategies to mitigate overfitting and promote generalization across different anatomical and visual domains.

Dataset A: Blender-Simulated Surgical Scenes Using Real Scoliosis Cases

Data were split by patient ID (82% train, 16% validate, 2% test) to ensure subject-independent evaluation. Given that RGB values were procedurally generated to simulate surgical textures, we applied color augmentation to prevent the model from over-relying on synthetic appearance cues and to promote geometry-focused learning.

We implemented three augmentation modes applied randomly during training:

- **Generated mode:** RGB values mimic distributions from real surgical point clouds (as in Section 4.2.1).
- **Same mode:** All points are assigned a uniform color (ignoring RGB), removing color cues.
- **Swapped mode:** RGB values for vertebra and non-vertebra classes are reversed to confuse color-based learning intentionally.

All point clouds were normalized before being put into the model.

Dataset D: Physical Phantom Acquisitions with 7D Surgical System

We performed a split with stratification (75% train, 19% validate, 6% test) by deformity configuration and fixed random seed for reproducibility. Since this dataset closely resembles actual intraoperative point clouds, including soft tissue, metallic instruments, and realistic surface textures, we retained the original RGB values without applying color augmentation. Standard normalization was applied to all inputs.

Dataset P: RGB-D Capture of Phantom Spine for Blender-based Simulation

This dataset involved leave-one-out cross-validation (LOOCV) across six distinct curvature conditions. We used the same color augmentation scheme as for Dataset A to examine

how variations in color representation affect model performance across sensor domains. The objective was to evaluate the feasibility of RGB-D driven semi-synthetic datasets in promoting robust learning.

All datasets used a consistent input normalization pipeline to ensure uniform data scaling before model ingestion.

4.2.3 Cross-Dataset Training and Evaluation

To assess the generalization capability of our segmentation model across diverse anatomical and visual domains, we conducted systematic cross-dataset training and evaluation using combinations of the semi-synthetic datasets and the real SpineDepth dataset (see Table 4.1).

We explored 15 combinations:

- Individual: **S**, **A**, **D**, **P**
- Pairwise: e.g., **SA**, **SD**, **AD**
- Triplets: e.g., **SAD**, **SDP**, **ADP**
- Full combination: **SADP**

For each dataset combination, we applied dataset-specific pre-processing and augmentation strategies. SpineDepth dataset followed the class-balanced sampling and augmentation pipeline described in Section 4.1.3. Datasets A and P incorporated the color augmentation modes outlined in Section 4.2.2 to promote robustness against artificial texture bias. Dataset D, which captures high-fidelity intraoperative appearances, was used without any augmentation to preserve its realistic visual properties.

All models were trained with Point Transformer V3 using the same hyperparameters and normalization pipeline. For evaluation, each trained model was tested independently on all datasets using the Dice Similarity Coefficient (DSC) as the primary metric. Moreover, to interpret the results beyond simple mean DSC, two evaluation metrics were introduced (see Evaluation Metrics).

To identify the most generalizable training configurations, DSC scores were averaged across test sets. The top three combinations were selected for fine-tuning, where augmentation parameters and class weights were further adjusted to maximize performance across unseen domains.

This cross-dataset training strategy enabled us to quantify the contribution of each dataset to overall model robustness, to identify synergistic effects that arise from combining datasets,

and to better understand the transferability of learned features across both synthetic and real surgical domains.

Evaluation Metrics

1) Unseen-Drop To quantify how well the model generalizes to a dataset that was not included during training, the **unseen-drop** metric was defined:

$$\text{Unseen Drop} = \overline{\text{DSC}}_{\text{seen}} - \overline{\text{DSC}}_{\text{unseen}} \tag{4.4}$$

Here, $\overline{\text{DSC}}_{\text{seen}}$ is the average DSC of the domains used during training, and $\overline{\text{DSC}}_{\text{unseen}}$ is the performance on the held-out test domain. This metric captures the generalization gap and provides a direct measure of performance degradation when evaluating on entirely unseen data.

The use of a held-out domain for testing aligns with standard practices in domain generalization research, where generalization to out-of-distribution data is assessed through performance drop. Similar metrics have been adopted in prior works such as [44], where domain shifts are quantified through changes in predictive accuracy across unseen environments.

It is important to note that the **unseen-drop** metric is only applicable to training configurations that exclude at least one dataset. For the fully inclusive combination (SADP), where all four datasets are used for training, no domain remains unseen, and this metric cannot be computed.

2) Inter-Domain Standard Deviation To evaluate the consistency of the model's performance across datasets, the standard deviation of DSC values was computed:

$$\text{Inter-Domain STD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\text{DSC}_i - \overline{\text{DSC}})^2} \tag{4.5}$$

where $N = 4$ represents the number of test domains and $\overline{\text{DSC}}$ is the mean DSC across all domains. A lower standard deviation indicates that the model performs uniformly across datasets, suggesting stable generalization. Conversely, a high value reflects performance variance and possible overfitting to specific domains.

This type of variance-based evaluation is commonly used in domain generalization and fairness literature to assess robustness and consistency across multiple environments. For example, [45] uses cross-domain standard deviation as a key metric to evaluate domain general-

ization algorithms under the DomainBed benchmark.

4.3 Model Evaluation on Intraoperative Surgical Data

Given the ultimate objective of supporting real-time intraoperative assistance, our best-performing models were evaluated on two intraoperative point clouds captured by 7D Surgical at CHU Sainte-Justine. These acquisitions were performed toward the end of the procedures, after pedicle screw insertion but before the final radiographic imaging, providing a realistic yet radiation-free screenshot of the surgical field.

Building on the findings from Section 4.2.3, the best dataset combinations were selected and the model was retrained using the data augmentation strategies described in Section 4.2.2 and the class-balanced focal loss settings from Section 4.1.3 to enhance generalization and robustness in clinical settings.

To ensure compatibility with the model's input constraints, the same stratified downsampling procedure was conducted during training (see Section 4.1.3) to reduce the number of points per frame to 10,000 while preserving the original class distribution.

Due to the absence of ground truth labels in intraoperative data, this evaluation is purely qualitative. The side-by-side visualizations of the raw intraoperative scans were provided and the corresponding model predictions in Section 5.3, offering visual insight into the model's ability to localize vertebral structures in real surgical conditions.

This evaluation was designed to test the feasibility of applying transformer-based point cloud segmentation models in the operating room and highlights the promise of geometry-driven, real-time anatomical guidance in spinal procedures.

CHAPTER 5 RESULTS AND DISCUSSION

This chapter presents the experimental results, visualizations, and critical analysis of the segmentation framework described in Chapter 4. First, the training outcomes of the Point Transformer V3 model on the publicly available SpineDepth dataset are reported, including comparisons with prior work [1] and the extension from binary (vertebra vs. background) segmentation to six-class vertebral level segmentation. The development and evaluation of three semi-synthetic datasets are then detailed, supported by both qualitative visualizations and quantitative metrics. To assess domain generalization, cross-dataset experiments are conducted using data from cadaveric specimens (Section 4.1), real scoliosis cases with simulated tissue, and artificial spine models acquired with identical and different sensors, followed by manual correction (Section 4.2). Finally, the best-performing model is evaluated on a small, manually segmented intraoperative dataset (Section 4.3) to investigate its feasibility in real-world surgical settings.

5.1 Model's Training Results on Public Dataset: SpineDepth

The SpineDepth dataset [8], developed in the earlier work, consists of pose-annotated RGB-D recordings from mock spinal procedures conducted on ten human cadaveric specimens. For the purposes of this study, two specimens were excluded to maintain consistency and data quality. Specimen 1 was excluded due to limited anatomical exposure, which resulted in incomplete visualization of vertebral structures. Specimen 10 was excluded due to significant anatomical variation, specifically, a much smaller vertebral size compared to the other specimens, making it an outlier in both shape and scale. These same two specimens were also excluded in the original segmentation benchmark study [1], leaving eight specimens (see 5.1) for training and evaluation.

To ensure a fair comparison with prior work [1], the same experimental setup was adopted:

- **Evaluation protocol:** Leave-One-Out Cross-Validation (LOOCV) across eight folds, holding out one specimen per fold for testing.
- **Evaluation metric:** Dice Similarity Coefficient (DSC), with model performance reported using the median DSC across folds.

The segmentation results for both the original binary setting (vertebra vs. background) were evaluated, as illustrated in Figure 5.2 for one specimen, and our extended six-class version,

which differentiates between individual lumbar vertebral levels (L1–L5), shown in Figure 5.3. Table 5.1 summarizes the quantitative comparison, including the baseline from the previous RGB-D approach and our results with Point Transformer V3.

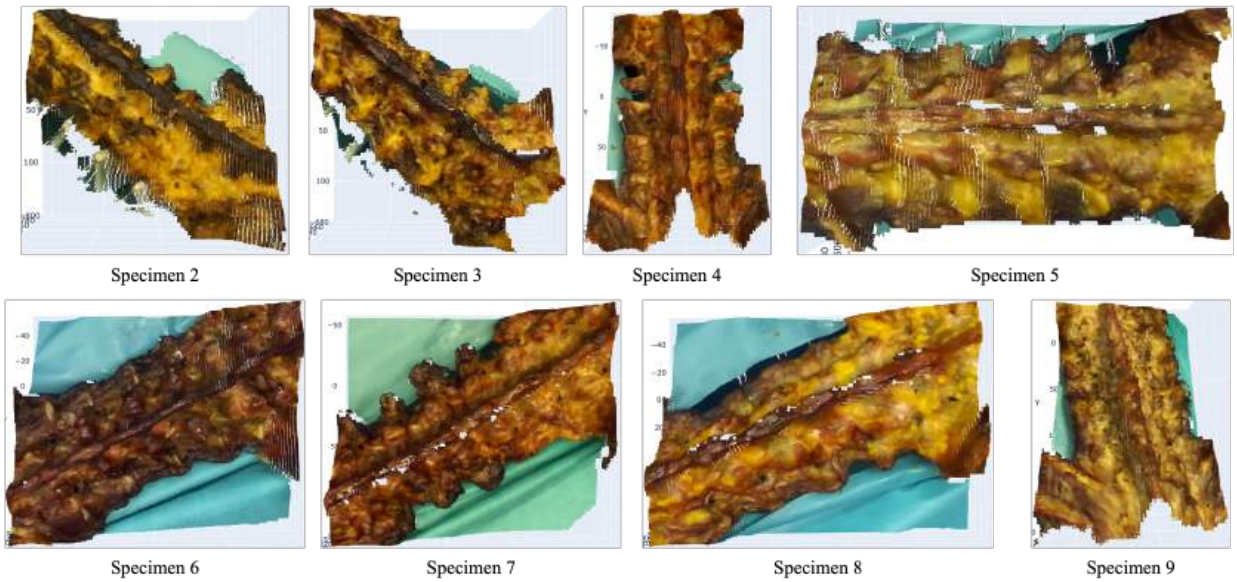

Figure 5.1 Visualization of the eight specimens used for training and evaluation.

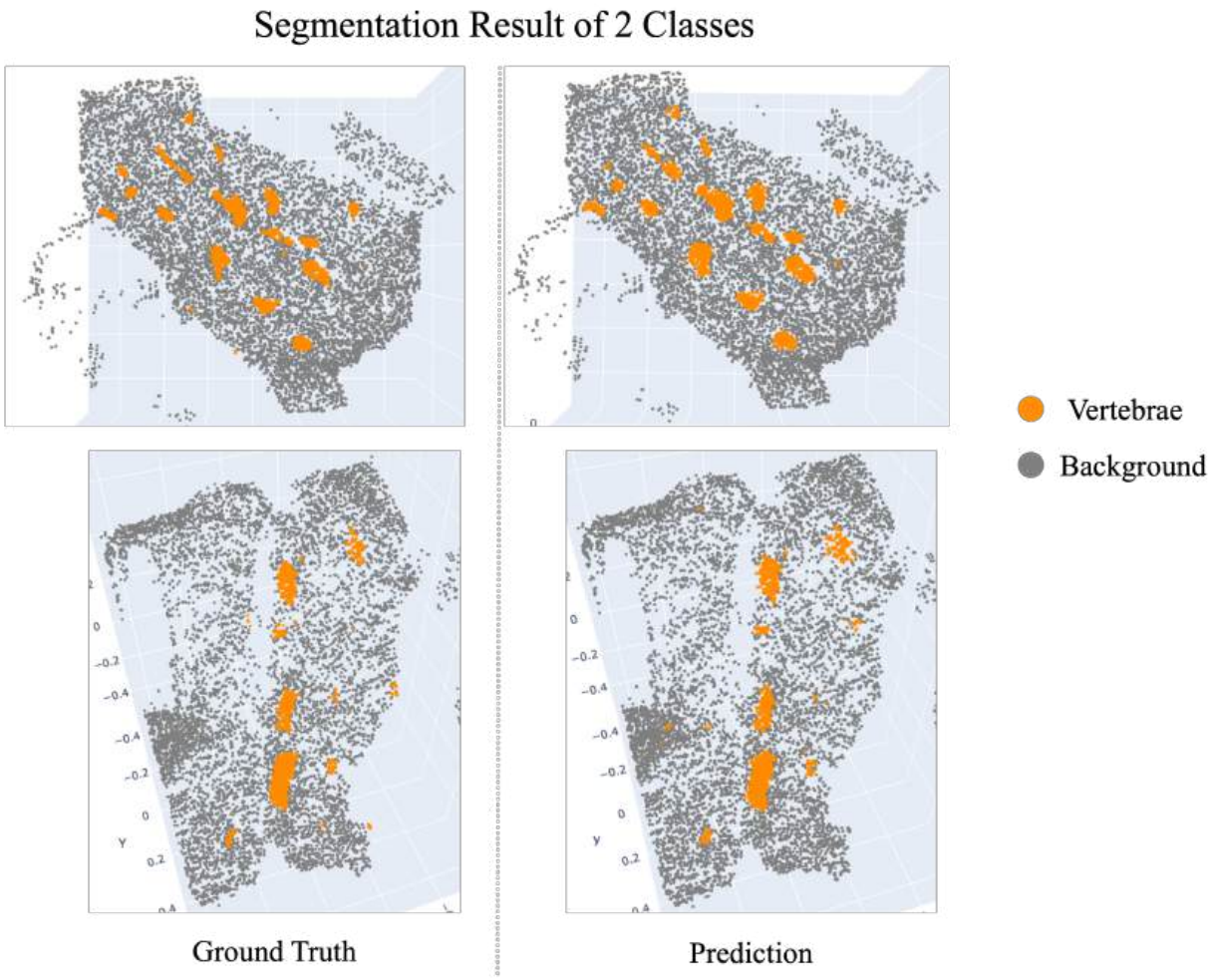

Figure 5.2 Qualitative results of Point Transformer V3 on the SpineDepth dataset (two-class segmentation: vertebra vs. background, DSC = 0.87). Each row shows a different frame from the same specimen. Column 1: ground truth; Column 2: model prediction.

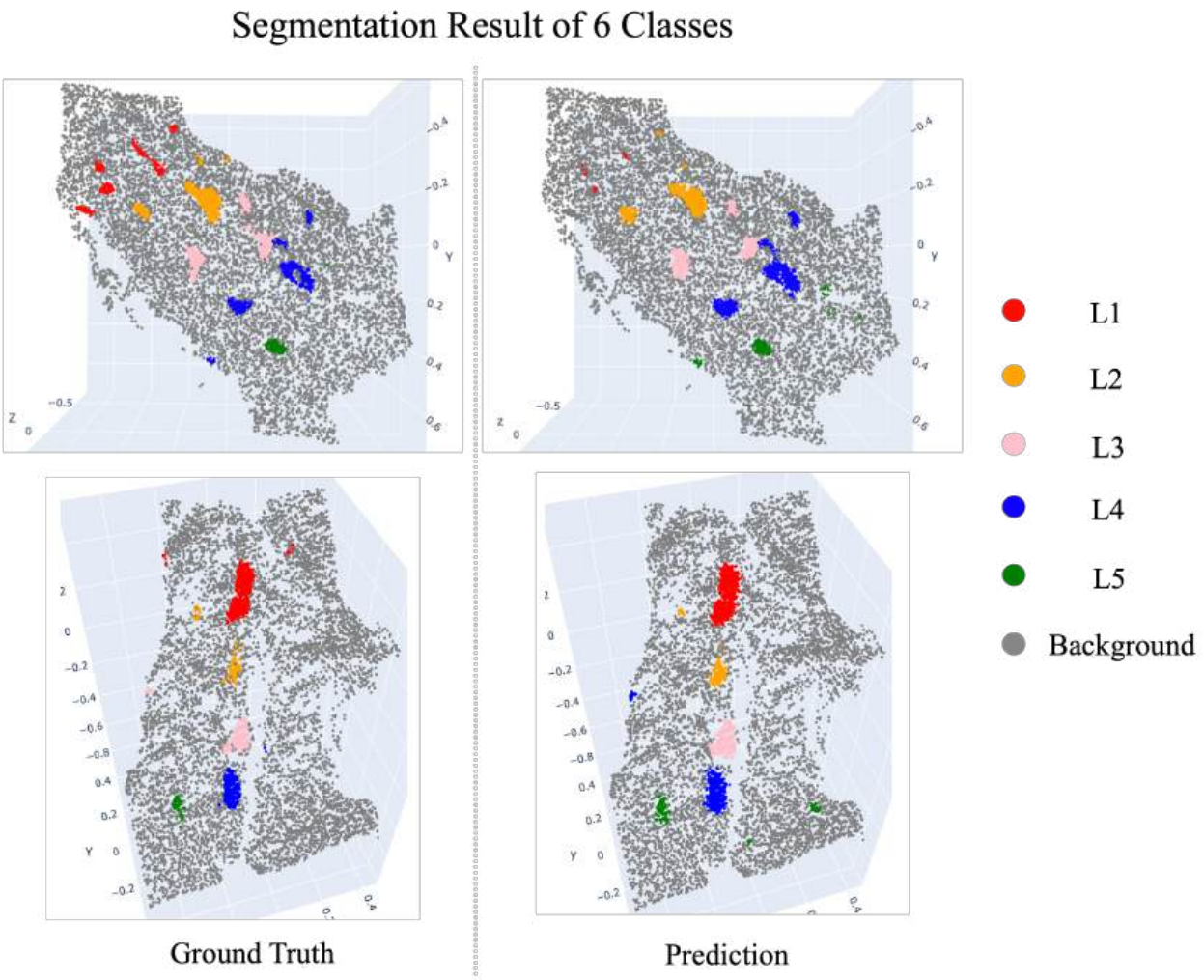

Figure 5.3 Qualitative results of Point Transformer V3 on the SpineDepth dataset (six-class segmentation: L1–L5 vertebrae + background, DSC = 0.74). Each row shows a different frame from the same specimen. Column 1: ground truth; Column 2: model prediction.

The baseline state-of-the-art method achieved a median DSC of 0.740 in the binary (2-class) setting, while the proposed Point Transformer V3 framework improved this to 0.845 (Table 5.1). The highest DSC obtained by the baseline was 0.760 for specimen 5, whereas our approach achieved the highest score of 0.870 for specimen 3. Although the six-class (multi-class) setting yielded a lower median DSC compared to the binary task, its accuracy remained comparable to the binary classifier from the baseline method, despite the increased difficulty of distinguishing between individual vertebral levels.

A non-parametric Wilcoxon signed-rank test [46] comparing the median DSC values for the binary segmentation results showed that the improvement achieved by the proposed method

was statistically significant ($p = 0.00781$). This confirms that the observed performance gains are unlikely to be due to chance, reinforcing the effectiveness of our approach over the existing state-of-the-art.

Table 5.1 Dice Similarity Coefficient (DSC) for semantic segmentation on the SpineDepth dataset, computed over the segmented lumbar spine region. In the baseline RGB-D approach [1], a U-Net-based model was trained using binary segmentation masks to isolate the lumbar anatomy from full RGB-D frames. In contrast, our Point Transformer V3 model was trained directly on pre-extracted regions of interest (ROIs) around the lumbar vertebrae. Results are shown for both binary (2-class) and multi-class (6-class) segmentation, with statistical comparison to the baseline method.

Dataset	**Specimen**	**RGB-D U-Net Based [1] (2 classes)**	**Point Cloud PTv3 (2 classes)**	**Point Cloud PTv3 (6 classes)**
SpineDepth	2	0.67	0.82	0.68
	3	0.74	0.87	0.75
	4	0.75	0.84	0.68
	5	0.76	0.85	0.71
	6	0.74	0.86	0.68
	7	0.74	0.84	0.72
	8	0.74	0.85	0.71
	9	0.67	0.83	0.72
	Median DSC	**0.74**	**0.845**	**0.72**

5.2 Results of Cross-Dataset Training and Evaluation

As outlined in Section 4.2, three semi-synthetic datasets were generated, each introducing variability in anatomy, acquisition modality, and scene realism, to support robust model training. These datasets include procedurally simulated data, RGB-D sensor captures, and radiation-free point clouds from the 7D Surgical System, collectively designed to reflect diverse intraoperative conditions.

Building on this foundation, a series of cross-dataset training experiments were conducted (see Section 4.2.3) to evaluate the generalizability of our semantic segmentation model. Specifically, the Point Transformer V3 model was trained on various combinations of four datasets (see Table 4.1). Each combination was evaluated on all four datasets independently, and the Dice Similarity Coefficient (DSC) was computed for each target domain. To further analyze

performance beyond mean DSC, two additional evaluation metrics were introduced (see Section 4.2.3). Our goal was to examine how well the model trained on one or more datasets can generalize to unseen domains, as well as to evaluate consistency across domains.

5.2.1 Cross-Dataset Training Results

The table below summarizes the DSC results for each dataset when the model is trained on specific dataset combinations. The mean DSC was also reported, the unseen-drop (when applicable), and the inter-domain standard deviation (see Evaluation Metrics):

Table 5.2 Dice Similarity Coefficient (DSC) performance of models trained on different dataset combinations. Each column reports the DSC evaluated on the corresponding test domain. **Mean DSC** is the average DSC across all test domains. **Unseen Drop** quantifies the performance degradation on the domain excluded from training, computed as the difference between the average DSC on seen domains and the average DSC on the held-out domain (see Eq. 4.4). **STD DSC** measures the standard deviation across the four test domains, reflecting consistency of model performance (see Eq. 4.5). Note: Unseen Drop is not applicable for combinations using all four datasets (SADP), as no domain is excluded.

Train Combo	**S** (25)	**A** (16)	**D** (9)	**P** (20)	**Mean DSC ↑**	**STD DSC ↓**	**Unseen Drop ↓**
S (175)	0.551	0.464	0.690	0.204	0.477	0.177	0.098
A (80)	0.172	0.986	0.217	0.664	0.510	0.336	0.635
D (48)	0.075	0.117	0.959	0.263	0.353	0.356	0.807
P (100)	0.146	0.933	0.217	0.985	0.571	0.390	0.553
SA	0.739	0.918	0.667	0.125	0.612	0.296	0.433
SD	0.488	0.527	0.958	0.299	0.568	0.244	0.310
SP	0.720	0.330	0.648	0.890	0.647	0.203	0.316
AD	0.146	0.983	0.957	0.732	0.704	0.674	0.531
AP	0.154	0.986	0.216	0.988	0.586	0.402	0.802
DP	0.126	0.793	0.962	0.979	0.715	0.348	0.511
SAD	0.443	0.952	0.925	0.421	0.685	0.253	0.352
SAP	0.473	0.959	0.645	0.977	0.763	0.213	0.158
SDP	0.529	0.597	0.929	0.975	0.757	0.197	0.214
ADP	0.144	0.984	0.962	0.989	0.770	0.361	0.830
SADP	0.684	0.928	0.911	0.927	0.862	0.103	N/A

5.2.2 Observations and Discussion

Several key insights can be drawn from the results (see Table 5.2):

- Training on all four datasets (SADP) resulted in the highest mean DSC (0.862) and

the lowest inter-domain standard deviation (0.103), demonstrating the effectiveness of aggregating diverse data sources. These results highlight the value of integrating diverse data sources, capturing variability in anatomy, imaging modality, and acquisition conditions, to build generalizable models.

- Training on individual datasets is insufficient for broad generalization. Models trained on individual datasets showed significant domain-specific biases. For example, the model trained only on Dataset S achieved a passable DSC on Dataset D (0.690) but performed poorly on Dataset P (0.204), showing limited transferability across domains. This suggests that single-domain training limits transferability to different clinical environments or data distributions.

- Although both Dataset A and Dataset P are synthetic with manually assigned colors, they differ in realism and structure. Models trained on Dataset P generalize well to both synthetic domains (P: 0.985, A: 0.933), likely due to its consistent signal, whereas models trained on Dataset A perform moderately on P (0.664). Combining both datasets (AP) boosts within-domain performance (A: 0.986, P: 0.988) but shows limited generalization to other datasets like S (0.154) and D (0.216), highlighting the persistent gap between synthetic and real data.

- Unseen-Drop values capture the cost of missing domain knowledge. When one dataset was excluded during training, the performance on that unseen domain typically dropped significantly. For example, leaving out Dataset S (ADP combination) caused the largest unseen-drop of 0.830, suggesting that Dataset S contains critical features or patterns not captured by the other datasets. This reinforces the importance of including representative samples from each domain when aiming for robust generalization.

- Inter-domain standard deviation provides complementary information to mean DSC, since mean DSC alone does not capture inconsistency across test domains. For instance, the AD model achieved a relatively high mean DSC (0.704), but with a high inter-domain STD (0.674), indicating that its good performance is unevenly distributed and possibly over-fitting to specific domains. In contrast, the SP model had a moderate mean DSC (0.647) with a lower STD (0.203), reflecting more consistent, though not optimal, performance across domains.

  These examples support the following intuition:

  - High mean DSC + low STD DSC $\rightarrow$ strong generalization across domains
  - High mean DSC + high STD DSC $\rightarrow$ domain-specific overfitting

  - Low mean DSC + low STD DSC → consistently poor performance
  - Low mean DSC + high STD DSC → unstable and domain-dependent performance

- Among the three-dataset combinations, ADP (mean DSC = 0.770), SAP (0.763), and SDP (0.757) performed relatively well. However, they still showed a noticeable performance gap (up to 0.1 DSC) compared to SADP, indicating that partial combinations improve generalization but do not fully match full training; comprehensive domain coverage is still critical for optimal performance.

These results demonstrate the importance of dataset diversity for robust generalization in 3D point cloud segmentation. They further highlight that mean accuracy, unseen-drop, and inter-domain standard deviation should be jointly considered to diagnose generalization strength and cross-domain reliability.

5.3 Generalization to Actual Intraoperative Data

As described in Section 4.3, the final stage of the evaluation focused on testing the model's applicability in real surgical settings. To this end, our best-performing model was evaluated on two intraoperative 7D system's point clouds captured during spinal procedures at CHU Sainte-Justine, providing realistic, radiation-free screenshots of the surgical field.

Building on the cross-dataset experiments in Section 5.2, the selected model was trained using the combined SADP dataset, which integrates both real and semi-synthetic data to promote robustness across domains. However, initial qualitative testing revealed that models trained without data augmentation, whether using only the SpineDepth dataset or the full SADP configuration, struggled to generalize to intraoperative point clouds. As illustrated in Figure 5.4, these models exhibited poor segmentation accuracy, underscoring a significant domain gap between training and real surgical data.

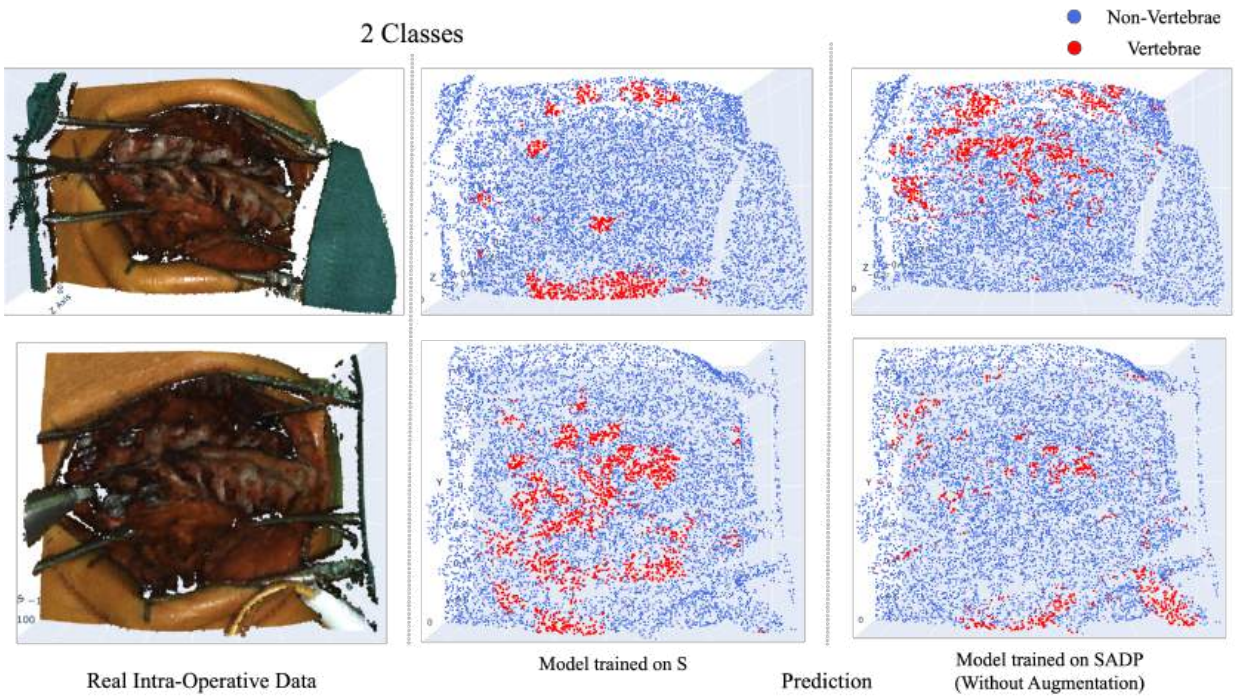

Figure 5.4 Qualitative segmentation results on real intraoperative point clouds. Models trained on the SpineDepth dataset and the SADP configuration without any data augmentation fail to generalize, highlighting a strong domain gap.

To mitigate this issue, a targeted color augmentation pipeline was developed (see Section 4.2.2) designed to reduce the model's reliance on color-based cues and encourage learning of geometry-based features. This was especially important for datasets such as Dataset A (Blender-simulated scoliosis cases) and Dataset P (RGB-D acquisition of the phantom spine), where color distributions are either synthetic or sensor-dependent.

Prior to applying these augmentations, a comparative color analysis was performed between the cadaveric SpineDepth dataset and a representative intraoperative case (Figure 5.5). The results showed substantial differences in the color profiles of vertebral points across domains, whereas non-vertebral (background) points exhibited more consistent distributions. This validated the need for augmentation strategies to increase color robustness and domain transferability.

To further evaluate the effectiveness of these augmentations, an additional semi-synthetic experiment was conducted, detailed in Appendix D. In this experiment, the realistic intraoperative color distributions were applied to a synthetic rendering of the SpineDepth specimen S3. The comparison between models trained on SpineDepth alone and the full SADP configuration revealed that color-aware generalization is significantly improved when diverse

domains are included during training.

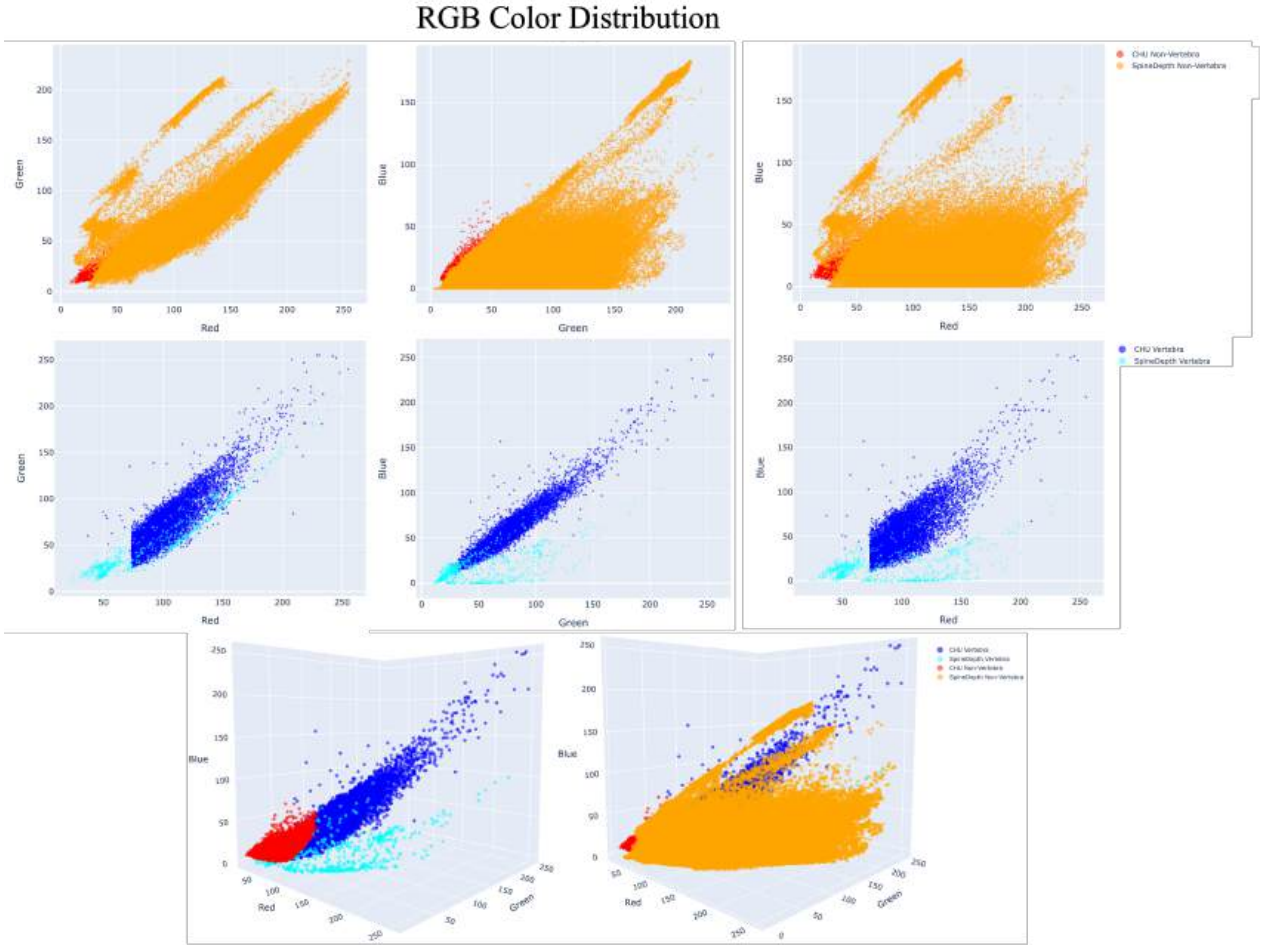

Figure 5.5 Color distribution analysis across domains. Top row: non-vertebra (background) color clusters. Middle row: vertebrae color clusters. Bottom row: visualization in RGB space. The color disparity between synthetic and intraoperative vertebrae supports the need for augmentation.

During training, one of the following augmentation modes was randomly applied to each training sample:

- **Generated mode:** RGB values were sampled from distributions observed in real intraoperative scenes to simulate realistic color textures.
- **Same mode:** Uniform RGB values were applied to all points in a scene, effectively removing texture information and emphasizing geometry.
- **Swapped mode:** RGB values for vertebra and non-vertebra points were reversed, disrupting appearance priors to challenge the model's robustness.

The augmentation strategy was tuned by varying the probabilities of each mode to find a balance between disrupting color priors and retaining anatomical cues. All training data were preprocessed using the same stratified downsampling and spatial normalization pipeline described in Chapter 4 to ensure consistency.

The final model, trained on the full SADP dataset with the optimized augmentation pipeline, was qualitatively evaluated on intraoperative data. As manual segmentation, the gold standard, is unavailable for intraoperative data, formal quantitative validation is precluded. Nevertheless, visual inspection reveals that the proposed method produces segmentations with superior anatomical fidelity and better alignment with expected anatomical structures when compared to the baseline. As shown in Figure 5.6, the model successfully localized exposed vertebral anatomy, demonstrating its capacity to generalize to real surgical environments and reinforcing its potential for clinical integration.

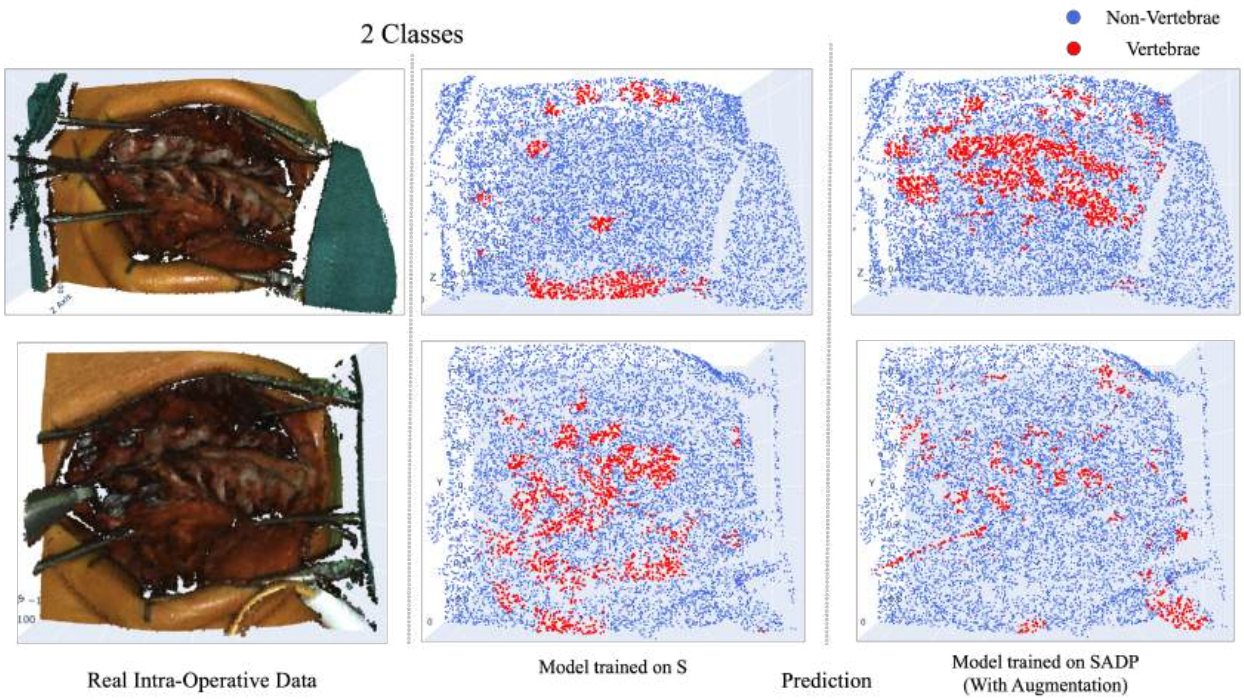

Figure 5.6 Qualitative segmentation results on real intraoperative point clouds. The model was trained on the SADP configuration with targeted data augmentation, improving generalization to real surgical scenes.

CHAPTER 6 CONCLUSION

6.1 Summary of Works

This work introduces a novel deep learning framework for real-time semantic segmentation of intraoperative 3D point clouds, enabling the identification of vertebral structures during scoliosis surgery. Our primary objective was to provide surgeons with accurate, radiation-free anatomical insights to support spinal alignment correction. To achieve this, we processed and annotated the public SpineDepth dataset. Additionally, we developed three semi-synthetic datasets to simulate a wide variety of spinal deformities and introduce different sensor characteristics, thereby getting closer to realistic surgical environments.

We selected the Point Transformer V3 architecture for this task based on our first hypothesis: that its advanced attention mechanisms would outperform prior approaches in point cloud segmentation. This was validated by a 14.19% improvement in results on the SpineDepth dataset compared to previous work, which utilized RGB-D data [1]. A non-parametric Wilcoxon signed-rank test on the binary segmentation results confirmed that this improvement was statistically significant ($p = 0.00781$), reinforcing the robustness of the performance gain and supporting the choice of PTv3 as the backbone of our framework.

Our second hypothesis stated that the SpineDepth dataset alone, while valuable, is insufficient to capture the anatomical and visual variability needed for robust generalization to intraoperative scenarios, especially for scoliosis. We confirmed this by observing limited transferability of models trained solely on this dataset.

To address this limitation, we proposed our third hypothesis: that supplementing real data with carefully designed semi-synthetic datasets, when combined with appropriate augmentation strategies, can improve segmentation performance on real intraoperative data. We supported this through cross-domain training experiments and qualitative evaluations on intraoperative point clouds, where our model demonstrated improved generalization and localization accuracy.

To our knowledge, this is the first study to apply deep learning-based segmentation directly to non-radiation intraoperative point cloud data acquired using a state-of-the-art surgical navigation system. Our key contributions include: 1) a complete pipeline for point cloud annotation and preprocessing, 2) a framework for generating diverse semi-synthetic training data, and 3) domain-aware augmentation strategies to bridge the gap between synthetic and real clinical data.

This research represents a significant step toward intelligent intraoperative assistance, providing a safer, radiation-free solution capable of continuously tracking and visualizing spinal curvature in real time. By delivering precise, ongoing feedback throughout the procedure, the system empowers surgeons with actionable insights to guide decision-making and enhance surgical outcomes.

6.2 Limitations

Despite the promising results, several limitations must be acknowledged. The current work lacks true intraoperative ground truth annotations, and evaluation on real surgical data was therefore limited to qualitative comparisons based on visual inspection. The cadaveric data used in the SpineDepth dataset, although valuable, does not reflect the complex and variable conditions encountered during live surgeries. These specimens were immobile, lacked soft tissue dynamics and scoliosis deformity, exhibited tissue discoloration, and were sourced exclusively from the lumbar spine region.

Similarly, the semi-synthetic datasets, while designed to mimic real intraoperative conditions, remain approximations. Factors such as bleeding, occlusion from instruments, lighting variability, and soft tissue deformation are not fully represented. The diversity of captured data is also limited. In actual surgeries, patient-specific anatomy, curvature severity, device usage, and exposure depth vary widely, and these variations play a critical role in model generalization. Further studies with broader and more realistic datasets are required to evaluate the performance in the operating room.

6.3 Recommendation for Future Work

Looking ahead, this research lays the groundwork for expanding real-time intraoperative assistance tools in spine surgery. A key direction for future work involves the development of a standardized intraoperative data collection protocol that includes annotated ground truth. Leveraging the capabilities of the 7D Surgical System, FLASH captures can be acquired at clinically meaningful time points during scoliosis procedures. By manually identifying anatomical landmarks and aligning them with preoperative CT scans, transformation matrices for each vertebra can be computed to enable accurate anatomical registration.

Our preliminary work with Dataset D confirms the feasibility of extracting both point clouds and transformation data from the 7D system for training and evaluation purposes. Continued intraoperative data collection using this pipeline will enrich the diversity and realism of training samples, ultimately enabling more robust and clinically relevant. Real surgical data

will also enable the inclusion of additional texture features, such as shininess and reflection, beyond standard RGB color information.

Beyond dataset expansion, another promising avenue for future research is the integration of segmentation results with real-time registration techniques to support dynamic intraoperative tracking. This would allow continuous assessment of spinal alignment throughout the procedure, offering surgeons immediate and radiation-free feedback. In Appendix A, we demonstrated how a subset of the model's predicted vertebral points can support successful registration from preoperative 3D CT to a predicted intraoperative point cloud. This alignment process also serves as a proxy to assess intraoperative spinal curvature. We're also actively collecting more intraoperative data at CHU Sainte Justine. The research protocol for this data collection includes perioperative radiographs taken at the end of surgery after spinal correction. These radiographs will serve as the ground truth for spinal alignment, allowing us to compare and validate the results of our complete registration pipeline.

Lastly, future efforts should focus on integrating this pipeline with commercial image-guided navigation systems and conducting prospective validation studies in real surgical settings. Such validation is critical to establishing the clinical utility, safety, and reliability of our approach. Building on the methods and findings presented in this thesis, there is strong potential to advance the development of intelligent, radiation-free surgical navigation tools that support safer and more precise scoliosis correction.

REFERENCES

[1] F. Liebmann, M. von Atzigen, D. Stütz, J. Wolf, L. Zingg, D. Suter, N. A. Cavalcanti, L. Leoty, H. Esfandiari, J. G. Snedeker *et al.*, "Automatic registration with continuous pose updates for marker-less surgical navigation in spine surgery," *Medical Image Analysis*, vol. 91, p. 103027, 2024.

[2] A. Tranchon, M. Kunz, and L. Séoud, "Preoperative mri whole-vertebrae segmentation in patients with severe adolescent idiopathic scoliosis," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2024, pp. 1–5.

[3] Y. Sun, Y. Xing, Z. Zhao, X. Meng, G. Xu, and Y. Hai, "Comparison of manual versus automated measurement of cobb angle in idiopathic scoliosis based on a deep learning keypoint detection technology," *European Spine Journal*, pp. 1–10, 2022.

[4] Z. Faraji-Dana, A. L. Mariampillai, B. A. Standish, V. X. Yang, and M. K. Leung, "Machine-vision image-guided surgery for spinal and cranial procedures," in *Handbook of Robotic and Image-Guided Surgery*. Elsevier, 2020, pp. 551–574.

[5] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.

[6] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in neural information processing systems*, vol. 30, 2017.

[7] X. Wu, L. Jiang, P.-S. Wang, Z. Liu, X. Liu, Y. Qiao, W. Ouyang, T. He, and H. Zhao, "Point transformer v3: Simpler faster stronger," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 4840–4851.

[8] F. Liebmann, D. Stütz, D. Suter, S. Jecklin, J. G. Snedeker, M. Farshad, P. Fürnstahl, and H. Esfandiari, "Spinedepth: A multi-modal data collection approach for automatic labelling and intraoperative spinal shape reconstruction based on rgb-d data," *Journal of Imaging*, vol. 7, no. 9, p. 164, 2021.

[9] J. P. Horne, R. Flannery, and S. Usman, "Adolescent idiopathic scoliosis: diagnosis and management," *American family physician*, vol. 89, no. 3, pp. 193–198, 2014.

[10] T. Kotwicki, J. Chowanska, E. Kinel, D. Czaprowski, M. Tomaszewski, and P. Janusz, "Optimal management of idiopathic scoliosis in adolescence," *Adolescent health, medicine and therapeutics*, pp. 59–73, 2013.

[11] K. B. L. Lim, I. S. X. Yeo, S. W. L. Ng, W. J. Pan, and N. K. L. Lee, "The machine-vision image guided surgery system reduces fluoroscopy time, ionizing radiation and intraoperative blood loss in posterior spinal fusion for scoliosis," *European Spine Journal*, vol. 32, no. 11, pp. 3987–3995, 2023.

[12] Z. Deniz Olgun and M. Yazici, "Posterior instrumentation and fusion," *Journal of children's orthopaedics*, vol. 7, no. 1, pp. 69–76, 2013.

[13] M. L. Goodwin, J. M. Buchowski, and D. M. Sciubba, "Why x-rays? the importance of radiographs in spine surgery," *The Spine Journal*, vol. 22, no. 11, pp. 1759–1767, 2022.

[14] B. S. Lonner, D. Kondrashov, F. Siddiqi, V. Hayes, and C. Scharf, "Thoracoscopic spinal fusion compared with posterior spinal fusion for the treatment of thoracic adolescent idiopathic scoliosis," *JBJS*, vol. 88, no. 5, pp. 1022–1034, 2006.

[15] L. V. Floccari, A. N. Larson, C. H. Crawford III, C. G. Ledonio, D. W. Polly, L. Y. Carreon, and L. Blakemore, "Which malpositioned pedicle screws should be revised?" *Journal of Pediatric Orthopaedics*, vol. 38, no. 2, pp. 110–115, 2018.

[16] J. M. Hicks, A. Singla, F. H. Shen, and V. Arlet, "Complications of pedicle screw fixation in scoliosis surgery: a systematic review," *Spine*, vol. 35, no. 11, pp. E465–E470, 2010.

[17] L. T. Holly and K. T. Foley, "Intraoperative spinal navigation," *Spine*, vol. 28, no. 15S, pp. S54–S61, 2003.

[18] S. Atallah, *Digital surgery*. Springer, 2021.

[19] F. Sommer, J. L. Goldberg, L. McGrath, S. Kirnaz, B. Medary, and R. Härtl, "Image guidance in spinal surgery: a critical appraisal and future directions," *International Journal of Spine Surgery*, vol. 15, no. s2, pp. S74–S86, 2021.

[20] I. Hussain, M. Cosar, S. Kirnaz, F. A. Schmidt, C. Wipplinger, T. Wong, and R. Härtl, "Evolving navigation, robotics, and augmented reality in minimally invasive spine surgery," *Global Spine Journal*, vol. 10, no. 2_suppl, pp. 22S–33S, 2020.

[21] J. P. Wilson Jr, L. Fontenot, C. Stewart, D. Kumbhare, B. Guthikonda, and S. Hoang, "Image-guided navigation in spine surgery: from historical developments to future perspectives," *Journal of Clinical Medicine*, vol. 13, no. 7, p. 2036, 2024.

[22] J. Silbermann, F. Riese, Y. Allam, T. Reichert, H. Koeppert, and M. Gutberlet, "Computer tomography assessment of pedicle screw placement in lumbar and sacral spine: comparison between free-hand and o-arm based navigation techniques," *European Spine Journal*, vol. 20, pp. 875–881, 2011.

[23] W. Birkfellner, J. Hummel, E. Wilson, and K. Cleary, "Tracking devices," in *Image-guided interventions: technology and applications*. Springer, 2008, pp. 23–44.

[24] A. J. Butler, M. W. Colman, J. Lynch, and F. M. Phillips, "Augmented reality in minimally invasive spine surgery: early efficiency and complications of percutaneous pedicle screw instrumentation," *The Spine Journal*, vol. 23, no. 1, pp. 27–33, 2023.

[25] J. Cool, G. Streekstra, J. Van Schuppen, A. Stadhouder, J. Van den Noort, and B. Van Royen, "Estimated cumulative radiation exposure in patients treated for adolescent idiopathic scoliosis," *European Spine Journal*, vol. 32, no. 5, pp. 1777–1786, 2023.

[26] C. O. R. R. Group *et al.*, "Quantification of radiation exposure in canadian orthopaedic surgery residents," *JBJS Open Access*, vol. 9, no. 3, pp. e23–00 170, 2024.

[27] A. S. Narain, F. Y. Hijji, K. H. Yom, K. T. Kudaravalli, B. E. Haws, and K. Singh, "Radiation exposure and reduction in the operating room: perspectives and future directions in spine surgery," *World journal of orthopedics*, vol. 8, no. 7, p. 524, 2017.

[28] G. M. Malham and N. R. Munday, "Comparison of novel machine vision spinal image guidance system with existing 3d fluoroscopy-based navigation system: a randomized prospective study," *The Spine Journal*, vol. 22, no. 4, pp. 561–569, 2022.

[29] G.-D. Chen and F.-F. Wang, "Medical data point clouds reconstruction algorithm based on tensor product b-spline approximation in virtual surgery," *Journal of Medical and Biological Engineering*, vol. 37, pp. 162–170, 2017.

[30] J. Yin, J. Fang, D. Zhou, L. Zhang, C.-Z. Xu, J. Shen, and W. Wang, "Semi-supervised 3d object detection with proficient teachers," in *European Conference on Computer Vision*. Springer, 2022, pp. 727–743.

[31] D. Krawczyk and R. Sitnik, "Segmentation of 3d point cloud data representing full human body geometry: A review," *Pattern Recognition*, vol. 139, p. 109444, 2023.

[32] J. Geng, "Structured-light 3d surface imaging: a tutorial," *Advances in optics and photonics*, vol. 3, no. 2, pp. 128–160, 2011.

[33] D. D. Frantz, S. E. Leis, S. Kirsch, and C. Schilling, "System for determining spatial position and/or orientation of one or more objects," Sep. 11 2001, uS Patent 6,288,785.

[34] B. Chan, J. F. Rudan, P. Mousavi, and M. Kunz, "Intraoperative integration of structured light scanning for automatic tissue classification: a feasibility study," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, pp. 641–649, 2020.

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[36] M. A. Mazurowski, H. Dong, H. Gu, J. Yang, N. Konz, and Y. Zhang, "Segment anything model for medical image analysis: an experimental study," *Medical Image Analysis*, vol. 89, p. 102918, 2023.

[37] X. Fan, Q. Zhu, P. Tu, L. Joskowicz, and X. Chen, "A review of advances in image-guided orthopedic surgery," *Physics in Medicine & Biology*, vol. 68, no. 2, p. 02TR01, 2023.

[38] L. Tanzi, P. Piazzolla, F. Porpiglia, and E. Vezzetti, "Real-time deep learning semantic segmentation during intra-operative surgery for 3d augmented reality assistance," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 9, pp. 1435–1445, 2021.

[39] P. M. Scheikl, S. Laschewski, A. Kisilenko, T. Davitashvili, B. Müller, M. Capek, B. P. Müller-Stich, M. Wagner, and F. Mathis-Ullrich, "Deep learning for semantic segmentation of organs and tissues in laparoscopic surgery," in *Current directions in biomedical engineering*, vol. 6, no. 1. De Gruyter, 2020, p. 20200016.

[40] T. Kubík, O. Kodym, P. Šilling, K. Trávníčková, T. Mojžiš, and J. Matula, "Leveraging point transformers for detecting anatomical landmarks in digital dentistry," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2025, pp. 216–228.

[41] J. Yoon, S. Hong, S. Hong, J. Lee, S. Shin, B. Park, N. Sung, H. Yu, S. Kim, S. Park *et al.*, "Surgical scene segmentation using semantic image synthesis with a virtual surgery environment," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 551–561.

[42] J. Cartucho, S. Tukra, Y. Li, D. S. Elson, and S. Giannarou, "Visionblender: a tool to efficiently generate computer vision datasets for robotic surgery," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 9, no. 4, pp. 331–338, 2021.

[43] E. Pérez, A. Sánchez-Hermosell, and P. Merchán, "Tlsynth: A novel blender add-on for real-time point cloud generation from 3d models," *Remote Sensing*, vol. 17, no. 3, p. 421, 2025.

[44] D. Li, Y. Yang, Y.-Z. Song, and T. M. Hospedales, "Deeper, broader and artier domain generalization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 5542–5550.

[45] I. Gulrajani and D. Lopez-Paz, "In search of lost domain generalization," in *International Conference on Learning Representations (ICLR)*, 2020.

[46] I. C. A. Oyeka, G. U. Ebuh *et al.*, "Modified wilcoxon signed-rank test," *Open Journal of Statistics*, vol. 2, no. 2, pp. 172–176, 2012.

[47] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *2009 IEEE international conference on robotics and automation*. IEEE, 2009, pp. 3212–3217.

[48] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-d point sets," *IEEE Transactions on pattern analysis and machine intelligence*, no. 5, pp. 698–700, 1987.

APPENDIX A PREOPERATIVE CT REGISTRATION TO MODEL PREDICTIONS

To evaluate whether the model's segmented vertebrae points are sufficient for downstream tasks such as spinal alignment assessment, we investigated their utility for registering preoperative 3D CT scans. Specifically, we examined the minimum number of predicted vertebral points required to achieve accurate registration of the preoperative CT to the intraoperative scene. To identify this minimal subset, we applied a region growing algorithm starting from high-confidence predictions located on the spinous and transverse processes, anatomical landmarks known for their prominence and stability, gradually expanding the region until registration accuracy converged.

The registration pipeline was composed of two main stages: a coarse global alignment followed by local refinement.

1. Global Alignment via FPFH and Arun's Method. We first computed local geometric features using the Fast Point Feature Histograms (FPFH) [47]. FPFH characterizes the neighborhood of a point p based on pairwise angular relationships with its neighbors:

$$\text{FPFH}(p) = \frac{1}{|\mathcal{N}(p)|} \sum_{q \in \mathcal{N}(p)} \text{SPFH}(p, q)$$

where $\mathcal{N}(p)$ denotes the set of neighboring points around p, and $\text{SPFH}(p, q)$ captures angular features such as differences in normals and point positions.

We then established correspondences between the up-sampled-predicted points and the CT mesh vertices using nearest-neighbor matching in FPFH feature space. From these correspondences, an initial rigid transformation $\mathbf{T}_{\text{init}} = [\mathbf{R}|\mathbf{t}]$ was computed using Arun's method [48]. Given two sets of corresponding points $\{\mathbf{x}_i\}$ and $\{\mathbf{y}_i\}$, the method minimizes:

$$\min_{\mathbf{R},\mathbf{t}} \sum_i \|\mathbf{y}_i - (\mathbf{R}\mathbf{x}_i + \mathbf{t})\|^2$$

The optimal rotation $\mathbf{R}$ is found via Singular Value Decomposition (SVD) of the centered covariance matrix, and the translation $\mathbf{t}$ aligns the centroids.

2. Local Refinement via Iterative Closest Point (ICP). To further refine the alignment, we applied the ICP algorithm, which iteratively minimizes the Euclidean distance

between closest point pairs:

$$\min_{\mathbf{R},\mathbf{t}} \sum_i \|\mathbf{y}_i - (\mathbf{R}\mathbf{x}_i + \mathbf{t})\|^2, \quad \text{where } \mathbf{y}_i = \text{NN}(\mathbf{x}_i)$$

Here, $\text{NN}(\mathbf{x}_i)$ denotes the closest point on the target mesh for each source point. ICP refines $\mathbf{R}$ and $\mathbf{t}$ until convergence, using the previously computed transformation from Arun's method as initialization.

This approach allows us to evaluate not only the geometric accuracy of the model's segmentation but also its applicability in clinical workflows requiring rigid registration between preoperative and intraoperative data for assessment of the spinal alignment (see Figure A.1).

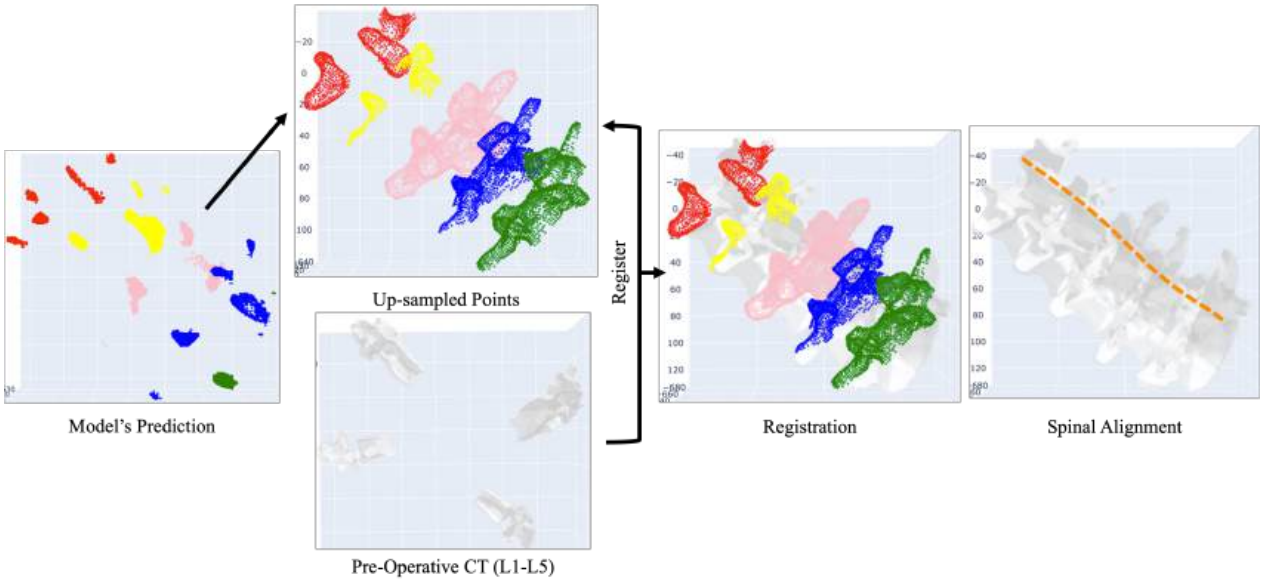

Figure A.1 Overview of the registration pipeline aligning preoperative CT vertebrae to the model's predicted segmentation.

APPENDIX B SPINEDEPTH: VERTEBRAL DISPLACEMENT

To better understand the temporal characteristics of the SpineDepth dataset, we analyzed the displacement of the vertebral centroids (L1-L5) over time. For each frame in a representative recording (frame rate: 15 fps, total frames: 285), we computed the centroid of each vertebra and tracked its position across the sequence in 3D space.

Figure B.1 illustrates the locations of vertebral centroids from frame 0 to frame 285 in three orthogonal planes (XY, YZ, XZ) and in full 3D space. The visualization clearly demonstrates that vertebral movement is minimal, reflecting the static nature of cadaveric specimens used in SpineDepth. This limited displacement emphasizes the lack of natural physiological motion, making the dataset less representative of real intraoperative scenarios involving scoliosis.

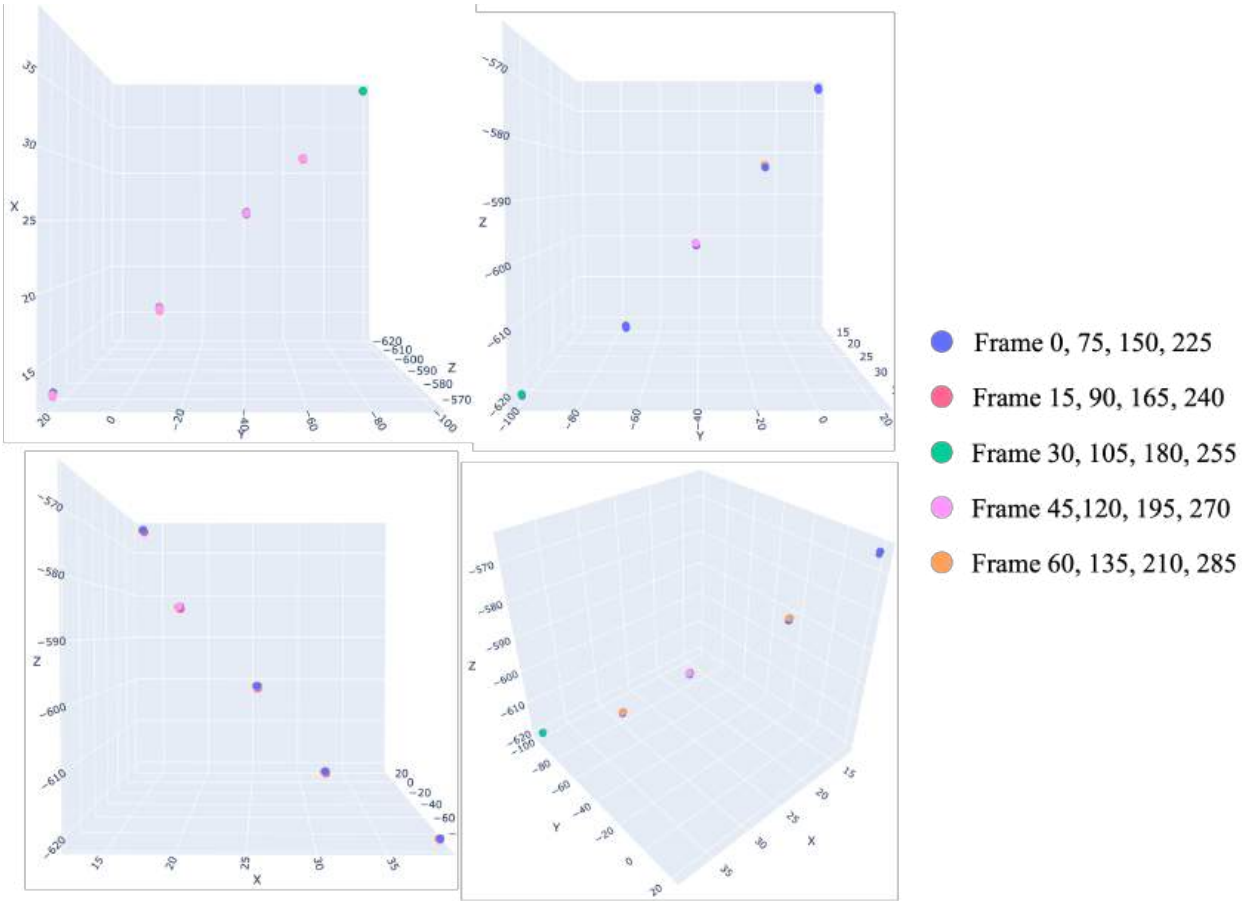

Figure B.1 Displacement of vertebral centroids (L1–L5) across 285 frames (15 fps) in the XY, YZ, XZ planes and 3D (XYZ) space. The minimal movement illustrates the static nature of the cadaveric setup in SpineDepth.

APPENDIX C SPINEDEPTH DATASET REDUCTION ANALYSIS

Given the low frame-to-frame variability observed in SpineDepth, we conducted additional experiments to assess the feasibility of reducing the training set size without compromising performance.

For 2-class segmentation (vertebra vs. background), we reduced the training dataset from the original 9,421 frames to just 175 frames, selecting 25 representative frames per specimen. The model was retrained using the same architecture and hyperparameters. For 6-class segmentation (individual vertebrae L1–L5 + background), we found that using 300 frames per specimen was sufficient to reach performance levels comparable to training on the full dataset.

As evaluated by the Dice Similarity Coefficient (DSC), the performance degradation in both scenarios was minimal (see Table C.1). These results suggest a high degree of temporal redundancy in SpineDepth, allowing for a substantial reduction in data volume during training without meaningful loss in segmentation quality.

Table C.1 Dice Similarity Coefficient (DSC) for Specimen 3 under different training set sizes using Point Transformer V3.

Training Set Size	**2-Class DSC**	**6-Class DSC**
All (9,421 frames)	0.876	0.752
(25 frames) × 7	0.872	0.603
(500 frames) × 7	—	0.750

APPENDIX D SEMI-SYNTHETIC EXPERIMENT ON SPINEDEPTH DATASET

As described in Section 5.3, we conducted an additional experiment using the SpineDepth dataset to assess how synthetic exposure and intraoperative color information affect model performance. Specifically, we applied the same semi-synthetic generation method used for Dataset A to one SpineDepth specimen (S3), introducing realistic surgical scene context (see Figure D.1). We then assigned colors based on real intraoperative RGB distributions, as analyzed in Figure 5.5.

The semi-synthetic S3 sample was evaluated using two models: one trained solely on the SpineDepth dataset, and another trained on the full combination of datasets (SADP). As shown in Figure D.1, the model trained exclusively on SpineDepth failed to generalize well to the synthetic-intraoperative setting with realistic color assignment. In contrast, the SADP-trained model performed significantly better, demonstrating improved adaptability to variations in visual appearance and scene context.

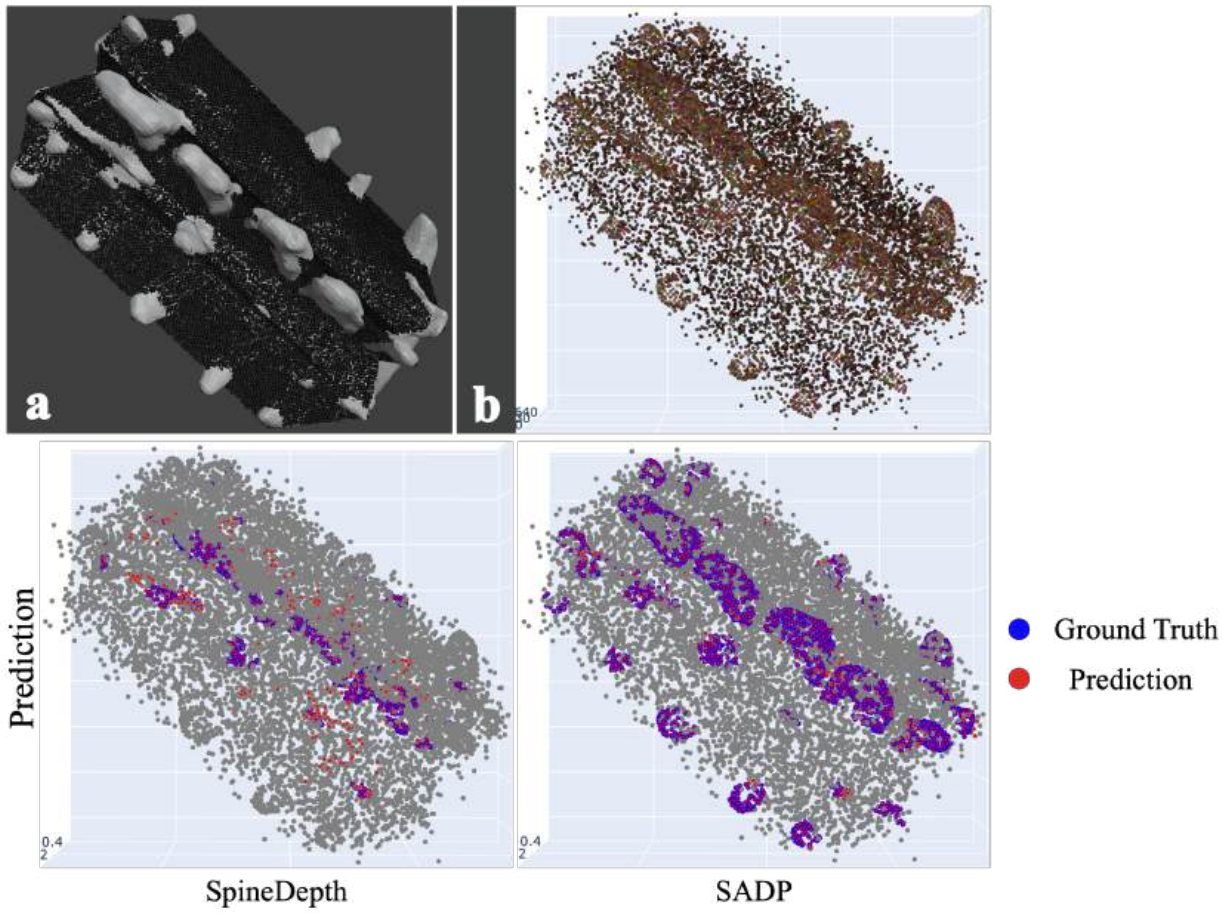

Figure D.1 Semi-synthetic experiment on SpineDepth specimen S3. (a) Blender rendering pipeline used to simulate the surgical environment. (b) Color assignment using intraoperative RGB statistics. Bottom row: predictions from two models, trained on SpineDepth only (left) and trained on the full SADP combination (right). The SADP model demonstrates greater robustness to color and context shifts.