Modeling and Visualizing Vertebral Anatomy from Freehand 3D Ultrasound

by

Sen Li

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
Ph.D.

MONTREAL, AUGUST 9, 2024

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Mrs. Catherine Laporte, Thesis supervisor
Department of Electrical Engineering, École de Technologie Supérieure

Mrs. Farida Cheriet, Thesis Co-Supervisor
Department of Computer Engineering and Software Engineering, Polytechnique Montréal

Mr. Simon Drouin, Chair, Board of Examiners
Department of Software Engineering and IT, École de Technologie Supérieure

Mr. David Labbé, Member of the Jury
Department of Software Engineering and IT, École de Technologie Supérieure

Mr. Hassan Rivaz, External Independent Examiner
Department of Electrical and Computer Engineering, Concordia University

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON JULY 18, 2024

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGEMENTS

My journey toward a Ph.D. has been challenging yet rewarding, and I am grateful for the unwavering support and encouragement I received from my supervisors, colleagues, friends, and family. While it's difficult to express my sincere gratitude to them in words, I would like to take this opportunity to acknowledge their contributions.

First and foremost, I express my deepest gratitude to my principal supervisor, Professor Catherine Laporte, for her continuous guidance, motivation, and inspiration throughout my Ph.D. studies. Her invaluable advice and discussions during weekly meetings have positively impacted my research and career trajectory. I am also grateful for her support during the COVID-19 pandemic, where her instructions and supervision helped me adapt to remote working tools and maintain my research progress. I also thank my co-supervisors, Professor Farida Cheriet, for her valuable comments and feedback on my research manuscripts. I would also like to thank my Ph.D. assessment jury for their constructive feedback and support.

I thank the Natural Sciences and Engineering Research Council of Canada (NSERC) and Optimisation du Potentiel Synergique et Interinstitutionnel D'équipes Interdisciplinaires Académiques Novatrices (OPSIDIAN) for awarding me the merit scholarship that enabled me to pursue my Ph.D. project. I also want to thank Digital Research Alliance of Canada (DRAC) for providing high-performance computing resources. Additionally, I would like to thank the anonymous reviewers who provided valuable feedback on my research papers and helped improve my work's quality.

I appreciate Professor Carole Fortin and Philippe Paquette for their valuable guidance on spinal ultrasound data collection and lamina landmark labeling, as well as Lou Gauthier and Carine-Cindy Nguefack-Tonleu for their assistance during the data acquisition and manual labeling. I express gratitude to my colleagues at ÉTS: Sahba Changizi, Eija Aalto, Suzon Olory, Thanh-Dung Le, Toufik Bouras, Mario Francisco Munoz and many others for their support, encouragement, and the memories we have shared.

Finally, I am indebted to my parents for their unconditional love, support, and sacrifices. Their emotional and financial support have been invaluable during my Ph.D. journey. Special thanks to my wife, who has been with me every step of the way, providing unwavering support and encouragement throughout my Ph.D. journey. Without their love and dedication, this achievement would not have been possible, and I hope my accomplishments will make them proud.

I would like to thank everyone who made my Ph.D. journey a fulfilling and memorable experience. Thank you all for your support.

Modélisation et Visualisation de l'Anatomie Vertébrale à partir de l'Échographie 3D Libre

Sen Li

RÉSUMÉ

L'échographie à main levée (FH 3D US) émerge comme une alternative viable et non invasive à l'imagerie par rayons X pour l'évaluation des déformations de la colonne vertébrale telles que la scoliose idiopathique de l'adolescent, qui est la forme la plus courante de scoliose chez les adolescents et se caractérise par des courbures dépassant 10 degrés. Contrairement aux rayons X, qui dépendent des rayonnements ionisants et présentent des risques en cas d'utilisation fréquente, le FH 3D US offre une solution d'imagerie non ionisante, économique et en temps réel qui améliore la sécurité des jeunes patients nécessitant des suivis réguliers. Cette technologie permet également une imagerie dynamique et portable, et peut évaluer les patients dans leur posture debout naturelle, reflétant ainsi la courbure de la colonne vertébrale.

Toutefois, la nature intrinsèque de échographie présente des défis dans l'analyse de la colonne vertébrale. Des facteurs tels que les artefacts, les variations de contraste des tissus et la difficulté à maintenir un alignement constant de la sonde lors de l'acquisition des images par balayage manuel peuvent compliquer l'interprétation des images échographiques. Cette thèse se concentre sur le développement de méthodes novatrices pour l'extraction automatique de points de repère à partir des données de l'FH 3D US afin de faciliter l'analyse non invasive de l'anatomie de la colonne vertébrale.

Une contribution essentielle de ce travail est l'identification de courbes de lames en 3D appariées comme une représentation novatrice et robuste de la forme de la colonne vertébrale. Ces courbes sont constituées de paires de points de repère localisés sur les lames, créant une représentation robuste et riche en informations. Chaque point de repère correspondant à une lame lombaire est également utilisé pour guider un modèle d'apprentissage profond afin de cibler et d'extraire précisément la surface osseuse vertébrale. Pour établir le fondement d'une analyse automatisée, deux protocoles de marquage manuel standardisés ont été conçus et validés par des experts de l'échographie spinale. Ces protocoles assurent la précision et la cohérence dans l'identification des repères ponctuels appariés sur les lames pour la représentation de la forme de la colonne vertébrale en 3D et pour la délimitation de la surface osseuse vertébrale.

Sur la base de ces fondements, un cadre basé sur un Réseau Neuronal Convolutif (CNN) a été développé pour l'extraction automatique des repères de lames appariés à partir d'images d'échographie. Nous avons relevé des défis tels que l'identification des critères optimaux pour la reconnaissance des lames, l'optimisation des hyper-paramètres pour un entraînement robuste, et la sélection de modèles de CNN appropriés. La performance du cadre de travail basé sur CNN a été évaluée à l'aide de la validation croisée K-Fold sur les données de trois participants. Les résultats ont montré une erreur de distance moyenne de 2,1 ± 1,3 mm et de 1,8 ± 1,2 mm (3 mm est acceptable pour l'évaluation de la scoliose) pour les repères des lames gauche et droite,

respectivement. Ce travail a démontré la faisabilité de l'extraction des repères à partir d'images d'échographie 2D individuelles, ouvrant ainsi la voie à des applications en temps réel.

Pour améliorer davantage la robustesse et la fluidité de l'extraction des courbes de lames, nous avons étendu le cadre basé sur le CNN au Réseau Convolutif Récurrent de Localisation Séquentielle (SL-RCN), en tenant compte de la nature séquentielle en 3D des données FH 3D US. Cette avancée permet l'intégration des contraintes de forme de la colonne vertébrale en 3D dans le processus d'extraction, conduisant à des représentations plus précises. Une validation croisée à 7 volets est effectuée sur les données de 7 participants, en utilisant la stratégie de laisser-un-participant-de-côté. En comparaison avec le cadre basé sur le CNN, le SL-RCN génère des erreurs de distance moyenne gauche/droite réduites de 1,62/1,63 mm à 1,41/1,40 mm, et des distances de Fréchet discrètes normalisées de 0,591/0,639 à 0,428/0,457. Les résultats de l'expérience ont démontré l'efficacité du SL-RCN dans l'extraction de courbes de repères de lames appariées précises et fluides, et des études d'ablation ont vérifié l'utilité des composants architecturaux composant le SL-RCN.

Enfin, cette thèse explore l'application du Modèle Segment Anything (SAM) en Zero-Shot pour la segmentation des surfaces vertébrales dans les images d'échographie, visant à obtenir une visualisation vertébrale complète. La performance des méthodes de segmentation automatisées basées sur la version automatisée du SAM a été évaluée, et une méthode novatrice exploitant le SAM stimulé par des repères et la distribution d'intensité des images a été proposée pour une extraction précise des surfaces osseuses vertébrales. Cette méthode est spécifiquement adaptée aux images d'échographie transversales suboptimales, et au défi présenté par l'invisibilité des surfaces verticales dans les images d'échographie spinale transversales. Les masques d'ombre acoustique sous la surface osseuse extraite ont été évalués par rapport aux masques étiquetés manuellement, atteignant une moyenne d'Intersection over Union supérieure à 0,92. Cette approche a montré des résultats prometteurs dans la reconstruction de maillages 3D de vertèbres lombaires.

En conclusion, cette thèse présente des méthodes novatrices pour l'extraction automatique des caractéristiques vertébrales à partir des données FH 3D US, ouvrant la voie à l'analyse non invasive des déformations de la colonne vertébrale et facilitant l'application clinique de l'imagerie par ultrasons de la colonne vertébrale. Les méthodes proposées pour extraire les courbes de lamina en 3D appariées et les surfaces osseuses vertébrales, couplées aux protocoles de marquage manuel standardisés, contribuent de manière significative à l'avancement du FH 3D US en tant qu'outil viable pour l'imagerie diagnostique et interventionnelle de la colonne vertébrale.

Mots-clés: anatomie vertébrale, échographie, apprentissage profond, repère vertébral, forme de la colonne

Modeling and Visualizing Vertebral Anatomy from Freehand 3D Ultrasound

Sen Li

ABSTRACT

Freehand 3D Ultrasound (FH 3D US) is emerging as a viable, non-invasive alternative to X-ray imaging for assessing spinal deformities like Adolescent Idiopathic Scoliosis (AIS), which is the most common form of scoliosis in adolescents and is characterized by curvatures exceeding 10 degrees. Unlike X-rays, which rely on ionizing radiation and pose risks with frequent use, FH 3D US offers a radiation-free, cost-effective, and real-time imaging solution that enhances safety for young patients requiring regular follow-ups. This technology also allows for dynamic imaging and portability, and it can assess patients in their natural standing posture, thereby reflecting the spine curvature.

However, the inherent nature of ultrasound imaging presents challenges in spinal analysis. Factors such as image artifacts, variations in tissue contrast, and the difficulty of maintaining consistent probe alignment during freehand scanning can complicate the interpretation of ultrasound images. This thesis focuses on developing novel methods for the automatic extraction of vertebral features from FH 3D US data to facilitate non-invasive analysis of spine anatomy.

A key contribution of this work is the identification of paired 3D lamina curves as a novel and robust representation of spinal shape. These curves are comprised of pairwise anatomical landmarks of the laminae, creating a robust and information-rich representation. Each lumbar lamina landmark is further employed to guide a deep learning model in accurately targeting and extracting the vertebral bone surface. To establish a foundation for automated analysis, two standardized manual labeling protocols were designed and validated by spinal ultrasound experts. These protocols ensure accuracy and consistency in identifying pairwise point landmarks on the laminae for 3D spine shape representation and for delineating the vertebral bone surface.

Building upon this foundation, a Convolutional Neural Network (CNN)-based framework was developed for the automatic extraction of pairwise lamina landmarks from ultrasound images. We addressed challenges such as identifying optimal criteria for laminae structure identification, optimizing hyper-parameters for robust training, and selecting suitable CNN models. The performance of the CNN-based framework was assessed using K-Fold cross-validation on data from three participants. The results showed a mean distance error of 2.1 ± 1.3 mm and 1.8 ± 1.2 mm (3mm is acceptable for scoliosis assessment) for left and right lamina landmarks, respectively. This work demonstrated the feasibility of lamina landmark extraction from individual 2D ultrasound images, paving the way for real-time applications.

To further enhance the robustness and smoothness of lamina curve extraction, we extend the CNN-based framework to Sequential Localization Recurrent Convolutional Network (SL-RCN), accommodating the 3D sequential nature of FH 3D US data. This advancement allows for the integration of 3D spine shape constraints into the extraction process, leading to more

accurate representations. A 7-fold cross-validation is conducted on data from 7 participants, employing the leave-one-participant-out strategy. In contrast to the CNN-based framework, SL-RCN generates reduced left/right mean distance errors from 1.62/1.63mm to 1.41/1.40mm, and normalized discrete Fréchet distances from 0.591/0.639 to 0.428/0.457. The experiment results demonstrated the effectiveness of SL-RCN in extracting accurate and smooth paired lamina landmark curves and ablation studies verified the utility of the architectural components comprising SL-RCN.

Finally, this thesis explores the application of the Segment Anything Model (SAM) Zero-Shot for segmenting vertebral surfaces in ultrasound images, aiming to capture a complete vertebral visualization. The performance of SAM's automated and prompt-based segmentation methods was evaluated, and a novel method leveraging landmark-prompted SAM and image intensity distribution was proposed for accurate extraction of vertebral bone surfaces. This method is specifically tailored for suboptimal transverse ultrasound images, addressing the limitations of invisible vertical edges in transverse spinal ultrasound images. The acoustic shadow masks beneath the extracted bone surface were evaluated against manually labeled masks, achieving a mean Intersection over Union above 0.92. This approach demonstrated promising results in reconstructing 3D meshes of lumbar vertebrae.

In conclusion, this thesis presents novel methods for automatic extraction of vertebral features from FH 3D US data, paving the way for non-invasive analysis of spine deformities and facilitating the clinical application of spinal ultrasound imaging. The proposed methods for extracting paired 3D lamina curves and vertebral bone surfaces, coupled with the standardized manual labeling protocols, contribute significantly to the advancement of FH 3D US as a viable tool for spine imaging and intervention.

Keywords: vertebral anatomy, ultrasound, deep learning, vertebral landmark, spine shape

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ABBREVIATIONS

Name	Description	Pages
AIS	Adolescent Idiopathic Scoliosis	1, 9, 10, 17, 39, IX, XI
APM	Asymmetrical Peak Model	68, 70, 71, 74
BCE	Binary Cross Entropy, or Binary CE	42–44, 46, 47, 51, 82
BiGRU	Bidirectional GRU	49, 58
BiLSTM	Bidirectional LSTM	49, 58
CT	Computed Tomography	12
CAE	Conservative Adaptive Erasure	53, 54
CE	Cross Entropy	42–44, 46, 47, 51, 82, XVII
CNN	Convolutional Neural Network	3, 4, 26, 40, 41, 43–49, 51–54, 56, 58, 61–63, 67, 77, 78, 81–84, VII–X, XV
CoL	Center of Lamina	39
DG	Data Group	43–45, 48, 52–54, XIII
DLT	Direct Linear Transformation	12
FH 3D US	Freehand 3D Ultrasound	1, 2, 4, 14–17, 23, 24, 30, 39, 48, 49, 52, 63, 65–67, 76, 77, 79, 81, 83, 84, 89, VII–X, XIII, XV
GRU	Gated Recurrent Unit	49, 58, 62, XVII
H muscle	Horizontal trapezius muscle	28, 31
H-IPC	Horizontal IPC	29, 31, 68–71, XVI
ILR	Inter-Lamina Region	87, 88
IoU	Intersection over Union	67, 73, 76, XIII
IPC	Intensity Projection Curve	29–31, XI, XV, XVII, XVIII
LR	Lamina Region	87, 88
LS-TSE	**L**andmark prompted **S**AM - **T**rio **S**urface **E**xtraction	26, 38, 66, 67, 72–74, 76, 79, 80, 82–84, XIII, XVI
LSTM	Long Short-Term Memory	49, 58, 62, XVII

Name	Description	Pages
MDE	Mean Distance Error	45, 47, 48, 58–62, 78
MLP	Multilayer Perceptron	51, 61, 62
MSD	Mean Surface Distance	67, 73, 76
MSE	Mean Squared Error, or Mean SE	42–44, 46, 47, 51, 82
NDFD	Normalized Discrete Fréchet Distance	36, 49, 56, 58, 59, 61–63, 78
PCA	Principal Component Analysis	69, 70, XVI
PLV	Projected Laminae Vector	79, 89, 90, XII
PMC	Plane of Maximum Curvature	13
RCN	Recurrent Convolutional Network	49, 62, XVIII
RNN	Recurrent Neural Network	49, 51, 54, 55, 58, 61–63
RoI	Region of Interest	28, 30, 31, 53, 68, 69, XV, XVI
RVIBS	Rough Visual Intensity Brightness Score	27, 31, XI
SAM	Segment Anything Model	5, 66–68, 71, 72, 74, 76, 79, VIII, X, XVI
SE	Squared Error	42, XVIII
SL-RCN	Sequential Localization RCN	4, 26, 49–53, 56–59, 61–63, 78, 81–84, VIII–X, XIII, XV
SPL	Spinous Process Landmark	29, 31, 68–71, 73, 76
V muscle	Vertical muscle contiguous to the spinous process	28, 31
V-IPC	Vertical IPC	29, 31, 71
ViT	Vision Transformer	66
VPI	Volume Projection Imaging	15–19, 77, 79, 90, XI, XVIII
VPI-SP	VPI - Spinous Process	15, 16, 18
VPI-TP	VPI - Transverse Process	15, 16, 18
WI-IPC	Whole-Image IPC	29, 30

INTRODUCTION

Adolescent idiopathic scoliosis (AIS) is a three-dimensional spinal deformity occurring in adolescents, characterized by a curvature exceeding 10 degrees when viewed from the front. The term "idiopathic" signifies that the cause of this deformity remains unknown. AIS is the most prevalent form of scoliosis among adolescents, often displaying a high rate of curve progression. While early-stage AIS patients frequently follow a benign course not necessitating treatment, the inability to reliably predict curve progression makes regular follow-up essential.

X-ray imaging, including low-dose EOS, is the most frequently used medical imaging modality in current clinical spine exams, while ultrasound has the potential to be advantageous in regular follow-up exams for adolescents by providing real-time, non-invasive depictions of spinal anatomy, as well as facilitating continuous monitoring of spinal conditions over an extended period without exposing patients to radiation. Additionally, Freehand 3D Ultrasound (FH 3D US) distinguishes itself by its convenient portability, minimal expenditure, and the ability to be performed on patients in a standing posture (Vavruch & Tropp, 2016).

In scoliosis evaluation, ultrasound imaging offers a promising radiation-free alternative to X-rays. Although ultrasound images generally display only the posterior surface of vertebrae with less clarity compared to X-rays, innovative methods have been developed to estimate measures analogous to the Cobb angle, such as the center of laminae method (Chen, Lou, Zhang, Le & Hill, 2013) and the volume projection imaging method (Cheung, Zhou & et al., 2015). However, these methods either measure deformity on the coronal slice or a non-planar reslice are unable to provide information about axial plane vertebral rotation and are not fully automated.

Interpreting ultrasound images of the spine poses challenges due to variability in image contrast and difficulties in aligning the probe orthogonally to the bone surface, which affects the visibility of critical features like the spinous process. Additionally, localizing vertebral landmarks in transverse ultrasound sequences is challenging due to the complex morphological

characteristics of vertebrae, variations in brightness contrast of the paraspinal muscle response, and indistinct ultrasonic wave echoes along the bony surface (Feldman, Katyal & Blackwood, 2009; Hacihaliloglu, 2017). Ultrasound imaging artifacts further complicate the differentiation between spinous and interspinous process structures. Precise detection and localization of lamina landmarks in ultrasound images pose significant challenges due to the complex anatomy of the vertebrae, variations in tissue contrast, and ambiguity in ultrasound reflections from bone surfaces. Currently, there is a lack of tested lamina landmark recognition algorithms specifically designed for poor-quality 2D ultrasound scans or interlaminar scans. Moreover, defining single-point landmarks that accurately represent the lamina is a task without clear guidelines regarding the most relevant part of the visible lamina bone surface in ultrasound images.

This thesis aims to capture the intricate anatomy of the vertebrae using FH 3D US imaging. The main focus is on modeling and extracting relevant vertebral features that can effectively represent the shape of the spine. The ultimate goal is to perform non-invasive analysis of spine deformities and provide real-time guidance during interventions by leveraging the extracted features from FH 3D US data. The specific objectives were:

- **Identify Key Vertebral Features and Establish Ground Truth**: Identify an appropriate vertebral feature for a novel spine shape representation suitable for advanced 3D analyses. Acquire spinal FH 3D US data and establish ground truth labels for these identified vertebral features.
- **Develop a Feature Extraction Method at 2D Image Level**: Develop an AI-powered method for fast feature extraction from individual images to automate spine shape representation, explore ultrasound settings, and conduct a grid search to find a feasible configuration for an AI agent.
- **Evolve to Sequential 3D Spine Shape Learning**: Enhance the 2D image-based feature extraction method to accommodate sequential 3D spine shape information, thereby integrating

capabilities for 3D sequence-based feature extraction, followed by comprehensive validation and testing.

- **Automate Vertebral Surface Extraction**: Develop an automated method for vertebral surface extraction that leverages the previously extracted vertebral features, and subsequently validate and test this method.

This thesis is organized into six chapters. Chapter 1 presents the background and literature review. It contains an introduction to the spine, scoliosis, and a survey of the state-of-the-art in spinal ultrasound imaging. The following four chapters address the specific objectives. Chapter 6 discusses the findings and potential applications of the methods developed in this thesis.

Contributions 1: paired 3D lamina curves

Chapter 2 presents a comprehensive overview of the data acquisition, image interpretation, and manual labeling processes. Contributions 1, concerning the identification of paired 3D lamina curves from spinal ultrasound sequences, are detailed in this chapter. The representation of spinal shape using paired 3D lamina curves can fuse the advantages of both spinous process landmarks in the analysis of scoliotic deformities and transverse process landmarks in the assessment of axial vertebral rotation while mitigating deviation by virtue of its longer distance in the superior-inferior direction as compared to transverse process landmarks. Two standardized manual labeling protocols are proposed and have undergone validation by two medical professionals with strong expertise in vertebral ultrasound: one for identifying pairwise point landmarks on the laminae to create a 3D representation of spinal curvature, and the other for labeling the vertebral bone surface, including the left lamina, spinous process, and right lamina.

Contributions 2: CNN-based framework for landmark extraction

A deep learning framework is developed for automatic lamina landmark extraction. Chapter 3 proposes a convolutional neural network (CNN)-based framework for extracting pairwise lamina

landmarks from 2D ultrasound images, aimed at achieving efficient live landmark extraction. We explore depth values in ultrasound settings and conduct grid searches to identify a feasible configuration for the proposed framework. The experiments validated the effectiveness of the proposed framework by overcoming several primary barriers, including identifying the optimal criterion for laminae structure identification from a range of loss functions, optimizing hyper-parameters to guarantee robust training, and assessing the performance of a range of foundational CNN models. This work has been published in a peer-reviewed conference paper:

- Sen Li, Lou Gauthier, Farida Cheriet, Carole Fortin, Philippe Paquette, Catherine Laporte, "Lamina landmark detection in ultrasound images: a preliminary study," Proc. SPIE 12470, Medical Imaging 2023: Ultrasonic Imaging and Tomography, 1247011 (10 April 2023); https://doi.org/10.1117/12.2653277.

Contributions 3: SL-RCN for sequential curve extraction

Chapter 4 extends the CNN-based framework to accommodate the sequential nature of FH 3D US data, allowing implicit integration of spine shape constraints into lamina extraction. An upgraded framework, Sequential Localization Recurrent Convolutional Network (SL-RCN), is tailed for smooth 3D sequential lamina curve extraction. The experiments evaluated the inter- and intra-observer variability of the manual labeling protocol, validated the efficacy of the SL-RCN for smooth lamina curve extraction, and verified the utility of the architectural components comprising SL-RCN through ablation studies. This work has resulted in the following peer-reviewed publication:

- S. Li, F. Cheriet, L. Gauthier and C. Laporte, "Automatic 3D Lamina Curve Extraction from Freehand 3D Ultrasound Data using Sequential Localization Recurrent Convolutional Networks," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, doi: 10.1109/TUFFC.2024.3385698.

Contributions 4: vertebral bone surface extraction

Chapter 5 explores the application of the Segment Anything Model (SAM) Zero-Shot, which has shown promise in other medical imaging contexts, for segmenting vertebral surfaces in ultrasound images obtained from both a spine phantom and human subjects. The performance of SAM's automated mask generation ("SAM Auto") and prompt-based segmentation methods are compared. A novel method leveraging landmark prompted SAM and image intensity distribution is proposed for extracting vertebral bone surfaces, tailored specifically for suboptimal transverse ultrasound images. This work has been published in a peer-reviewed conference paper:

- S. Li, F. Cheriet, and C. Laporte, "SAM Zero-Shot detection of vertebral surfaces from freehand 3D ultrasound", Proceedings of the Annual Meeting of the Canadian Medical and Biological Engineering Society, vol. 46, Jun. 2024.

CHAPTER 1

BACKGROUND AND LITERATURE REVIEW

This chapter aims to provide a research background and literature review in the field. Section 1.1 delves into the anatomical and pathological intricacies of the human vertebral column, establishing a solid foundation for comprehending spinal structure, abnormalities, and clinical evaluations. Section 1.2 discusses the techniques employed for analyzing spine deformities, extracting spine features from ultrasound images, and facilitating the reconstruction of the 3D surface of the vertebrae.

1.1 Background

1.1.1 The vertebral column

The vertebral column is the skeletal structure extending from the skull to the pelvis in humans. The structure is divided into five distinct regions, namely cervical, thoracic, lumbar, sacrum, and coccyx, with each region fulfilling specific functions and comprising varying numbers of vertebrae. The shape of the vertebrae differs among regions to fulfill specific roles, and they are named according to the first letter of their respective region, along with a regional number in ascending order.

The uppermost region, referred to as the cervical region, corresponds to the anatomical area of the neck and comprises seven vertebrae known as cervical vertebrae (C1 to C7). Directly adjacent to the cervical region is the thoracic region, which encompasses twelve vertebrae labeled T1 to T12 and is functionally connected to the rib cage. Situated further below the thoracic region is the lumbar region, encompassing five vertebrae labeled L1 to L5, playing a pivotal role in providing essential support to the lower back.

In a healthy human spine, the anterior-posterior view typically appears as a straight line, while the lateral view exhibits an S-shaped curve. The S-shaped curve comprises four distinct curvatures that can be easily distinguished based on the separation of different spinal regions. Two of these

curvatures (cervical and lumbar) are posteriorly concave, curving inward, while the other two (thoracic and sacral) are posteriorly convex, curving backward. Functioning akin to a spring, the S-shaped curve enhances the power, flexibility, and shock-absorbing capacity of the vertebral column.

Vertebrae

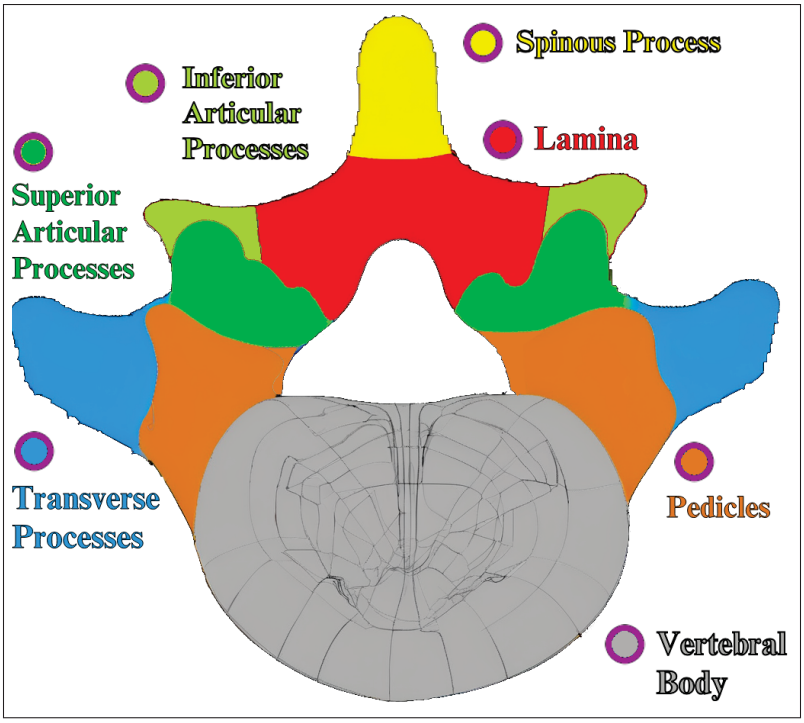

Figure 1.1 A Typical Vertebra

The 24 (cervical, thoracic, and lumbar) vertebrae share similar features. As illustrated in Figure 1.1, the general characteristics of a vertebra include a round cylinder-shape portion that is called the **body** (or the centrum). The body of the vertebra supports the weight, so an inferior vertebra generally has a larger and thicker body than its superior vertebrae. The vertebral foramen is the opening posterior to the body. The spinal cord is housed and protected by all of the vertebrae foramina together, whose combination is called the vertebral canal. The projection posterior to the vertebral foramen is the **spinous process**. The two lateral projections on the transverse plane are the **transverse processes**. The two symmetric regions that sit between the spinous process and the two transverse processes are **laminae**. The two connected regions between the two

transverse processes and the body are called the **pedicles**. The two pairs of laminae and pedicles together form the **vertebral arch**, with two pedicles being the two pillars and two laminae being the roof of the arch. At the intersection region between a pedicle and a lamina, there is one flat projection facing to the top of the vertebra and one facing to the bottom. The paired flat projections are called **superior articular process** and **inferior articular process** respectively.

1.1.2 Scoliosis

Scoliosis Categories	Age (years) of Scoliosis Occurrence
Congenital	At birth
Infantile	0 ~ 3
Juvenile	4 ~ 10
Adolescent	11 ~ 18
Adult	> 18 (may be degenerative)

Table 1.1 Scoliosis Classification

According to (Marieb & Hoehn, 2013), "Scoliosis, literally, 'twisted disease', is an abnormal *lateral* curvature that occurs most often in the thoracic region." This deformity is easily identifiable through a posterior-anterior radiograph, where the anterior view of the spine typically exhibits an "S" or "C" shape rather than a straight line. The majority of scoliosis cases are classified as idiopathic, indicating that their causes cannot be determined (Kim *et al.*, 2010). Scoliosis can be broadly classified into five categories based on the age of occurrence, as presented in Table 1.1. Among these categories, adolescents experience the most rapid progression of spinal curvature during their growth. As skeletal maturity advances, typically around the age of 25, the rate of curve progression tends to decrease.

1.1.2.1 Adolescent Idiopathic Scoliosis (AIS)

Adolescent Idiopathic Scoliosis (AIS) is the most prevalent type of scoliosis among affected individuals, with an estimated occurrence of 4 out of 100 adolescents (Scoliosis Research Society, 2019). AIS patients, who represent the majority of scoliosis cases, commonly experience a high rate of curve progression. Long-term follow-up studies on 194 untreated AIS patients over an

average of 39.3 years (Weinstein, Zavala & Ponseti, 1981) revealed that 72% of the 161 living patients had curves exceeding 50 degrees. Short-term studies conducted by various researchers (Brooks, Azen, Gerberg, Brooks & Chan, 1975; Rogala, Drummond & Gurr, 1978; Soucacos *et al.*, 1998; Wong & Tan, 2010) have also demonstrated a correlation between higher curve magnitudes and increased progression rates.

AIS patients undergo rapid spinal growth, resulting in the fastest curve progression. Cheung, Cheung, Samartzis & Luk (2018) found a positive correlation between growth rate and curve progression rate for body height and arm span. While most early-stage AIS patients with mild deformities have a benign clinical course that does not require treatment (North American Spine Society Patient Education Committee, 2020), the lack of a reliable method for prospectively identifying patients with progressive deformities poses a risk of unnecessary treatment (Ogilvie, 2020). Therefore, regular follow-up spine exams, typically every six months, play a crucial role in monitoring AIS patients. Depending on the rate of progression, the optimal observation interval for AIS patients may be as short as four months (Kim *et al.*, 2010). The traditional "watch and wait" approach is commonly employed for AIS patients with Cobb angles (defined in the following Section 1.1.2.2) below 25 degrees (Weinstein, Dolan, Cheng, Danielsson & Morcuende, 2008; Kim, 2014).

1.1.2.2 2D scoliosis assessment - Cobb angle

In order to capture the unique and complete shape of the spinal deformity for each individual, enabling accurate communication and treatment planning, the diagnosis of scoliosis deformity and its clinical assessment rely on a set of important terminologies (Stokes, 1994), including:

- **Vertebra centroid**: The point half way between the centers of the two endplates of a vertebra.
- **Vertebral body line**: The 3D curved line that passes through the centroids of the vertebral bodies.
- **Apical vertebra**: The vertebra or disc that has the greatest y coordinate in the global coordinate system.

- **Cephalad end vertebra**: The first vertebra in the cephalad direction from a curve apex whose superior surface is angled maximally toward the concavity of the curve, as measured in the posterior-anterior spinal projection.
- **Caudad end vertebra**: The first vertebra in the caudad direction from a curve apex whose inferior surface is angled maximally toward the concavity of the curve, as measured in the posterior-anterior spinal projection.
- **Cobb angle**: Angle between lines drawn on endplates of the end vertebrae.

The Cobb angle, derived from X-ray images, is currently considered the gold standard for evaluating the severity of scoliosis (Horng, Kuok, Fu, Lin & Sun, 2019). The measurement process involves identifying the two most tilted vertebrae: the cephalad end vertebra and the caudad end vertebra, within a specific curvature. Two main lines are extended along the endplates of these vertebrae, with one following the superior endplate of the cephalad end vertebra and the other along the inferior endplate of the caudad end vertebra. The angle formed at the intersection of these lines is the Cobb angle. In cases where space is limited, shorter lines perpendicular to the main lines can be drawn, and the angle between these shorter lines is equal to the Cobb angle.

The Cobb angle provides information solely on vertebral rotation within the 2D coronal plane, disregarding the 3D nature of scoliosis as a spinal deformity. Biplanar X-ray images with 3D spine reconstruction can be used to assess the 3D curvature of the vertebral body line. Section 1.1.2.3 will introduce scoliosis assessment using a reconstructed 3D spine model from multi-view X-rays.

1.1.2.3 3D evaluation of spine deformity from X-rays

Orthopedic research has made significant advances in the evaluation and treatment of spinal deformities, such as scoliosis, where accuracy in diagnosis and intervention planning is critical. The introduction of 3D reconstruction technology has revolutionized the field, providing detailed insights into the complex anatomy involved, thus leading to the cornerstone of 3D evaluation of spine deformity from X-rays.

Spine 3D reconstruction from multi-view X-rays

A pivotal technique employed for 3D reconstruction multi-view imaging is the combination of Direct Linear Transformation (DLT) and Stereo Reconstruction. DLT determines the 3x4 perspective transformation matrix of each view, which is a composite of the intrinsic parameters (focal length, principal point, skew) and extrinsic parameters (camera position and orientation) of a pinhole camera system (Li, 2016). Stereo Reconstruction enables the translation of 2D image landmark coordinates into their 3D world space counterparts through calibrated triangulation (Semeniuta, 2021) based on multi-view transformation matrices.

Delorme *et al.* (2003) introduced an innovative method to construct personalized, pathologically deformed 3D models of the human skeletal trunk, encompassing the spine, rib cage, and pelvis, from multi-view 2D planar radiographic images. These images consist of a postero-anterior radiograph (PA0), an angled postero-anterior radiograph (PA20), and a lateral radiograph (LAT). Their approach utilizes known close-range photogrammetry techniques, DLT and Stereo Reconstruction, to achieve 3D reconstruction of anatomical landmarks. Additionally, they employ 3D geometrical free-form deformation (elastic transformation) to deform high-resolution CT-scan reconstructed vertebra primitives into personalized low-resolution 3D data of the entire trunk, using a nonstructured set of control points. By merging minimal personalized information from 2D radiographic images with high-resolution generic 3D information from CT-scan reconstruction, this technique creates reconstructed high-resolution 3D geometrical models of the human skeletal trunk, aiding in the clinical assessment of spinal deformities.

Scoliosis evaluation in 3D

In order to accurately describe the vertebral body line in 3D and measure the 3D deformity of its curvature, we present definitions (Stokes, 1994) for new terminologies (bold) and clinical indices (italic bold):

- ***Apical vertebra axial rotation***: The vertebral transverse plane angulation (vertebral axial rotation) of the apical vertebra of a curve.

- **Plane of maximum curvature (PMC)**: The vertical plane that shows the greatest spinal curvature by a specified method (e.g., Cobb) when the specified part of the spine (e.g., curve bounded by end vertebrae) is projected onto it.
- ***Angle of plane of maximum curvature***: Angle θ_{PMC}, between the plane of maximum curvature and the global XZ plane.
- ***Curvature angle***: Angle measured on a specified plane with a specified method and with defined end vertebrae. E.g., Frontal Cobb Curvature Angle (scoliosis angle), Lateral Cobb Curvature Angle (lordosis angle, kyphosis angle), Maximum Cobb Curvature Angle (angle measured by the Cobb method in the plane of maximum curvature), Scoliosis Lateral Cobb Constrained Curvature Angle (constrained curvature angle of a scoliotic curve (frontal plane) projected into the lateral plane bounded by the frontal Cobb end vertebrae), etc.
- ***Constrained Cobb angle***: Cobb angle measured in a specified plane in which the end vertebrae have been defined in a plane other than in the plane of evaluation.

The radiologist initially identifies the two most tilted end vertebrae from the 3D spine reconstruction. Subsequently, the spine's 3D reconstruction is projected onto different vertical planes by rotating a vertical plane around the z-axis, transitioning between the Sagittal and Coronal Planes. Through this projection, multiple Cobb angles are obtained on various vertical planes, facilitating the identification of the PMC (Wu, Chu, He & Wong, 2020a; Wu, He, Chu & Wong, 2021), which exhibits the highest curvature. The orientation of the PMC, denoted as θ_{PMC}, provides crucial insights into the necessary corrective measures to bring the spine curvature to the sagittal plane, a process known as sagittalization. By integrating diverse curvature angles and determining the PMC, these clinical indices can enhance the accuracy of 3D scoliosis curvature assessment and serve as valuable guidance for surgical interventions aimed at restoring a more natural spine curvature.

Delorme *et al.* (2003) conducted a 3D evaluation of scoliosis using 3D spine models reconstructed from multi-view radiographic images of fifteen patients with varying scoliotic deformities. The creation of geometric models from these radiographs facilitated the measurement of diverse geometric indices related to the spine, rib cage, and pelvis. These measurements play a vital

role in accurately assessing the severity of deformities. The accuracy of these models, with an acceptable error margin of 3.3 ± 3.8 mm, indicates their suitability for customizing biomechanical simulations in orthopedic treatment strategies.

1.2 Spine ultrasound imaging

X-ray imaging, including low-dose EOS, is commonly used in clinical spine exams to evaluate the deformity in a standing position. However, the use of X-ray imaging raises concerns about radiation exposure, particularly for adolescents. To mitigate this risk, the EOS medical imaging system has been developed (McKenna *et al.*, 2012), offering reduced radiation doses for monitoring scoliosis and related orthopedic conditions. Nevertheless, ultrasound imaging provides a radiation-free and cost-effective alternative, making it an even better option for regular spine exams, especially for adolescents. Ultrasound not only provides real-time, non-invasive depictions of spinal anatomy but also facilitates continuous monitoring of spinal conditions over an extended period. Suzuki, Yamamuro, Shikata, Shimizu & Iida (1989) showed that ultrasound is a promising radiation-free option for regular follow-up exams by measuring vertebral rotation angles, which exhibit a strong linear relationship with the Cobb angle.

FH 3D US is particularly noteworthy due to its portability, low cost, and the ability to be performed on patients in a standing posture (Vavruch & Tropp, 2016). Researchers have demonstrated the utility of ultrasound imaging in outlining the spinous process and measuring vertebral rotation. Purnama & et al. (2010) proposed a framework for FH 3D US that enables human spine volume reconstruction after ultrasound image acquisition, demonstrating the feasibility of imaging the human spine using a FH 3D US system.

In a FH 3D US imaging system, patients are required to maintain a standing position during the FH 3D US scanning to accurately reflect the full extent of the deformity in the FH 3D US data. To assist patients in controlling their posture, two boards with posture supporters, one chest board, and one hip board are used. An eye spot helps ensure a proper standing posture, mitigating potential body displacement caused by probe pressure during scanning. A position

sensor is employed to measure the probe position and reconstruct the FH 3D US data in 3D. Two types of position sensors are commonly used: magnetic and optical. Optical position sensors rely on uninterrupted sight between cameras and target objects, often necessitating more space. On the other hand, magnetic position sensors offer greater flexibility in terms of space requirements; however, they require the environment to be free of ferromagnetic materials.

The calibration of a freehand 3D ultrasound system aims to determine the parameters necessary for the world space 3D reconstruction of each 2D ultrasound frame. This 3D reconstruction essentially maps every single 2D pixel (u, v) to its corresponding 3D voxel (x, y, z), which involves two pixel-to-millimeter scaling factors (s_u, s_v) and two transformation matrices: one T_I^P from the B-scan image coordinate system (CS) $\boldsymbol{I}$ to the mobile probe CS $\boldsymbol{P}$, and the other one T_P^W from $\boldsymbol{P}$ to the stationary position sensor CS $\boldsymbol{W}$ (also referred to as the world CS). The overall transformation is expressed as the multiplication of the homogeneous transformation matrices $T_P^W\ T_I^P\ (s_u u,\ s_v v,\ 0,\ 1)^T$. Here, T_P^W is collected from the position sensor, thus the calibration of the FH 3D US is to determine s_u, s_v, and T_P^W.

The calibration process is typically conducted by scanning a phantom, an object with precise and predefined geometric dimensions (Anagnostoudis & Jan, 2005). There are four commonly used types of phantoms: the point phantom (or cross-wire phantom), the stylus, the plane phantom, and the 2D alignment phantom. The choice of phantom depends on the desired accuracy, ease of use, reliability, and speed of calibration, with no single phantom universally outperforming the others (Hsu, Prager, Gee & Treece, 2009). By utilizing the known geometry of the phantom in conjunction with measurements derived from the collected scans, we can solve for the calibration parameters using the Levenberg-Marquardt algorithm.

1.2.1 Volume Projection Imaging (VPI)

Using the FH 3D US system, Cheung *et al.* (2015) proposed two methods for scoliosis assessment: volume projection imaging (VPI) - spinous process (VPI-SP) and VPI - transverse processes (VPI-TP). The process involves collecting a sequence of FH 3D US images that exhibit a similar

lateral curvature to the patient's actual spine. These images are then used to reconstruct a 3D ultrasound volume through interpolation and hole filling. The next step is to obtain a reslice for scoliosis severity analysis.

There are two reslicing methods - planar reslicing and non-planar reslicing - along with their corresponding reslices. Planar reslicing, which passes through multiple vertebral bodies due to the S-shaped curvature (as discussed in Section 1.1.1) of the human spine, results in black areas (acoustic shadow) that lack sufficient vertebral information. On the other hand, the narrowband non-planar reslice captures spinal process information across the entire spine from the 3D ultrasound volume. An optimal non-planar reslice exhibits intact and symmetric transverse processes with a high-luminance wave response and maintains a consistent distance parallel to the spinal cord, ensuring a robust and evenly distributed acoustic shadow curve of the spinal processes. To improve the acquisition of the non-planar reslice, Lee *et al.* (2021) proposed using the skin-to-lamina distance derived from a sagittal profile, bringing the non-planar reslice closer to the desired configuration.

Both the VPI-SP and VPI-TP methods aim to locate the most tilted vertebrae and reproduce the Cobb angle in the narrowband non-planar reslices using two lines. The VPI-TP method assumes that the line passing through the tips of the left/right transverse processes is parallel to the vertebral endplate, while the VPI-SP method assumes that the vertical plane passing through a spinous process in the superior-inferior direction is perpendicular to its corresponding vertebral endplate plane. In the VPI-SP method, two short lines along the center of their corresponding spinous-process acoustic shadow are drawn to calculate the spine curvature. For the VPI-TP method, two lines are determined based on the tips of paired transverse processes. It is important to consider the depth on the posterior-anterior axis when using the VPI-TP method, as the transverse processes in different regions have different posterior-anterior axis locations.

Scolioscan is a dedicated FH 3D US system designed specifically for clinical scoliosis assessment using the VPI method. Zheng *et al.* (2016) conducted an experiment to assess the reliability and validity of spine deformity measurement with Scolioscan. Their findings were compared to

gold standard Cobb angle measurements obtained from radiographs of AIS patients. The study demonstrated good intra-operator and inter-operator reliability for scoliosis scanning using the VPI method. In a comprehensive study involving 952 patients, Wong *et al.* (2019) meticulously evaluated the Scolioscan system and VPI method, using EOS radiography as the reference standard. They quantitatively demonstrated that FH 3D US imaging is a viable radiation-free alternative to radiography for regular follow-up evaluations during the treatment of scoliosis, particularly for AIS patients.

The VPI-based methods, while promising for AIS follow-up using FH 3D US imaging, exhibit several limitations. Firstly, the angle measured on the non-planar reslice is theoretically biased compared to the Cobb angle measured from coronal-plane X-ray imaging. Figure 1.2 illustrates the contrast between the world space and the vertebral space. While the Cobb angle is typically measured in the coronal plane of the world space, the VPI-based methods involve a non-planar reslice composed of stacked vertebral segments parallel to the Y_vZ_v plane in vertebral space. This difference introduces a bias when projecting the rotated non-planar reslice from vertebral space to the coronal plane in the world space. The bias increases with greater degrees of vertebral rotation, potentially affecting the accuracy of the VPI-based methods.

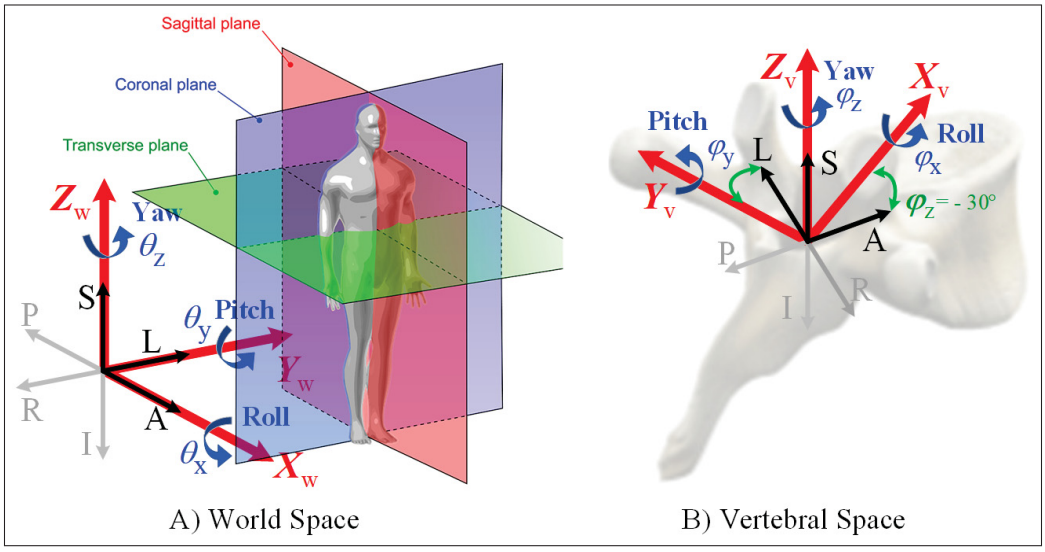

Figure 1.2 World Space vs Vertebral Space

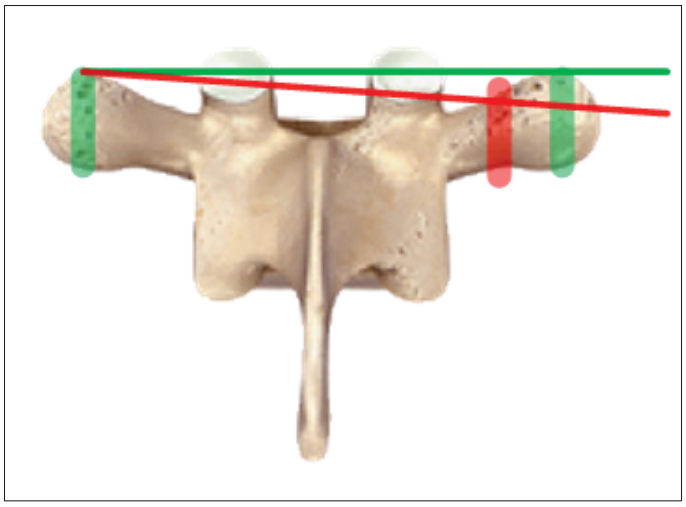

Figure 1.3 Asymmetrical reslice when axial-plane vertebral rotation happens

In addition to the aforementioned bias, the VPI-based methods inherit the limitation of the Cobb angle, as they are unable to provide information about axial-plane vertebral rotation.

In practice, more bias arises during the determination of the two most tilted vertebral curvature values from the non-planar reslice. The accuracy of scoliosis assessment using the VPI-TP method is contingent upon the symmetrical alignment of the left and right transverse processes within the non-planar reslice obtained from the ultrasound volume. Figure 1.3 illustrates this dependency: the intended representation of vertebral position is indicated by a green horizontal line, but in the presence of axial-plane vertebral rotation, asymmetry leads to a deviating representation, marked by a red line. Thus, the introduction of asymmetry in the VPI-TP method compromises the accuracy of scoliosis assessment.

Automated scoliosis assessment based on the VPI-SP method has been studied (Zhou & Zheng, 2015; Zhou, Jiang, Lai & Zheng, 2017; Zhou, Li, Zhou, Jiang & Zheng, 2020). These studies assume that the spinous-process acoustic shadow in the coronal plane reslice from an ultrasound volume generates a smooth curve. However, it should be noted that each vertebra contributes only one vertebral slice to the spine curvature in the non-planar reslice, and the true representation of its spinous process in the ultrasound volume non-planar reslice is a segment rather than a

smooth curve. Additionally, the shadows cast by different vertebrae may exhibit varying widths and degrees of asymmetry due to axial-plane vertebral rotation. These factors can significantly undermine the accuracy of vertebral curvature estimation. In summary, axial-plane vertebral rotation can not only introduce bias but also increase variability and reduce reliability in scoliosis assessments using the VPI-based methods.

1.2.2 Vertebral landmark localization

Localizing vertebral landmarks in transverse ultrasound sequences is challenging due to the complex morphological characteristics of vertebrae, variations in brightness contrast of the paraspinal muscle response, and indistinct ultrasonic wave echoes along the bony surface (Feldman *et al.*, 2009; Selbekk *et al.*, 2013; Jain & Taylor, 2004; Hacihaliloglu, 2017). Ultrasound imaging artifacts further complicate the differentiation between spinous and inter-spinous process structures. These difficulties are compounded by appearance variations caused by probe orientation, adding complexity to ultrasound image interpretation. Baum *et al.* (2018) proposed to facilitate manual identification of vertebral landmarks by incorporating a visual aid overlay within the ultrasound image space. Kamali *et al.* (2017) and Pinter *et al.* (2018) segment the bone contour in real-time from longitudinal scans and register it between the ultrasound space and a generic spine model, enabling 3D visualization of the spine. Although facilitating the visualization of transverse processes and lamina areas in 3D space, their methods are deficient in quantifying the 3D spine shape.

Berton, Cheriet, Miron & Laporte (2016b) proposed a segmentation method using linear discriminant analysis to extract the spinous process landmark in transverse scans, but this method was not tested for laminae landmarks, nor interspinous process views. Zeng *et al.* (2022, 2021) labeled the inter-spinous process structures as “fake” spinous process landmarks. Brignol, Gueziri, Cheriet, Collins & Laporte (2020) labeled the laminae landmarks below the bone surface wave response, and proposed an automated landmark localization method from high-quality transverse ultrasound scans through trend analysis on cumulative intensity curves of axial projections. However, it is sensitive to noise, artifacts, and thick bone responses. Its

effectiveness is contingent on preserving the symmetrical W-shape wave response and acoustic shadow clarity (Li *et al.*, 2023). Deng & Huang (2019) presented a Single Shot MultiBox Detector, which leverages deep learning techniques, to achieve the automatic detection of both spinous process and transverse process from transverse scans. Nevertheless, their bounding box encompasses both the lamina and transverse process regions. In the case of the laminae region, whose morphological characteristics exhibit a high degree of variability in practical scenarios, the utilization of a multiple anchor-box representation may not be able to effectively capture the relevant information. In conclusion, the extraction of sequential lamina landmarks from transverse ultrasound sequences remains an ill-defined problem.

1.2.3 Spine 3D surface reconstruction

The 3D reconstruction of the spinal tissue-bone interface holds notable clinical importance in terms of diagnosing and prognosticating spine deformities as well as providing guidance during intra-operative imaging procedures. Ungi *et al.* (2020) proposed to use U-Net to segment the bone surface in paramedian sagittal ultrasound images for the 3D reconstruction of left/right transverse process. The spine shape represented by paired transverse processes in 3D provided assistance in visualization and scoliosis measurement. Tang *et al.* (2021) introduced a U-Net-based approach for extracting the bone surface in order to reconstruct a phantom lumbar spine, where their method produces non-contiguous masks encompassing all discernible bone surface echo responses, including the anterior complex.

The extraction of the spinous process bone surface presents a significant challenge in spinal ultrasound imaging. This challenge arises from the underlying physical principles governing ultrasound wave propagation. In this process, bone surfaces that are perpendicular to the direction of the ultrasound wave generate the most pronounced response, resulting in the formation of a conspicuous high-luminance (white) horizontal curve within ultrasound images. While ultrasound scans effectively capture distinct horizontal edges, they tend to obscure vertical edges, impeding the accurate visualization of these features. A manual contour-tracing based method is proposed (Nguyen, Vo, Le & Lou, 2015; Jiang & Zhang, 2016) for 3D spinal surface

reconstruction, where the contour line effectively connects the left lamina, spinous process, and the right lamina, allowing for a potentially accurate representation of individual vertebra outlines in the reconstructed 3D spinal surface. However, this approach is labor-intensive.

A precise 3D spine model is essential for comprehending spinal structure, with the spinous process serving a critical role in providing 3D positional information and facilitating vertebral rotation analysis, particularly in the context of ultrasound-guided spine interventions. Accurate online 3D spinal surface reconstruction from ultrasound sequences can improve the interpretation of vertebral structure in ultrasound images, streamline sonographer training, enhance understanding of probe operation, and ultimately optimize the quality of acquired ultrasound sequences.

In the current literature, the existing methods for scoliosis assessment require manual work and are subject to theoretical biases. While vertebral landmarks can be utilized for 3D anatomical analysis, there is a lack of an automated and accurate method for extracting these landmarks, especially in sub-optimal images. Additionally, visualizing the spine remains challenging, particularly in determining the dimensions and orientation of the spinous process. In this thesis, we will address these challenges in spine anatomy analysis — both quantitatively and visually — in the subsequent chapters.

CHAPTER 2

DATA ACQUISITION, INTERPRETATION, AND LABELING

This chapter provides a comprehensive overview of the data acquisition, interpretation, and labeling processes used in this research. Section 2.1 details the data acquisition process using the FH 3D US system, including the equipment, operational techniques, and ethical protocols involved. It highlights the advantages of the system and the nature of the acquired data, while also providing an overview of detailed experimental procedures such as machine calibration, timing, and data collection speed. Section 2.2 delves into the interpretation of transverse spinal ultrasound images, discussing the interpretation of soft tissue layers, the challenges of delineating the spinous process outline, and the use of intensity-based image interpretation to assist in localizing the spinous process. Section 2.3 describes two manual labeling protocols: one for identifying pairwise point landmarks on the laminae to create a 3D representation of spinal curvature, and the other for labeling the vertebral bone surface, including the left lamina, spinous process, and right lamina.

2.1 Data acquisition using freehand 3D ultrasound

3D ultrasound systems can be classified into three types based on different data acquisition procedures (Prager, Ijaz, Gee & Treece, 2010), namely the 2D array transducer system, the mechanical 1D transducer system, and the FH 3D US system. Out of the three different 3D ultrasound systems, the FH 3D US system is the only one that can conveniently provide imaging for large body structures like the spine; it is portable, compact and has the lowest cost. FH 3D US captures 3D ultrasound sequences with each frame comprising a distinct 2D image and its corresponding transformation matrix obtained from a 3D position sensor.

Figure 2.1 demonstrates the acquisition of spinal ultrasound sequences in 3D. As the sonographer operates the ultrasound probe along the spine shape, data frames are collected at a frame sampling speed of 10 f/s and stored in two sequences: one with ultrasound images and the other with transformation matrices. The matrices map the images from the transducer to the

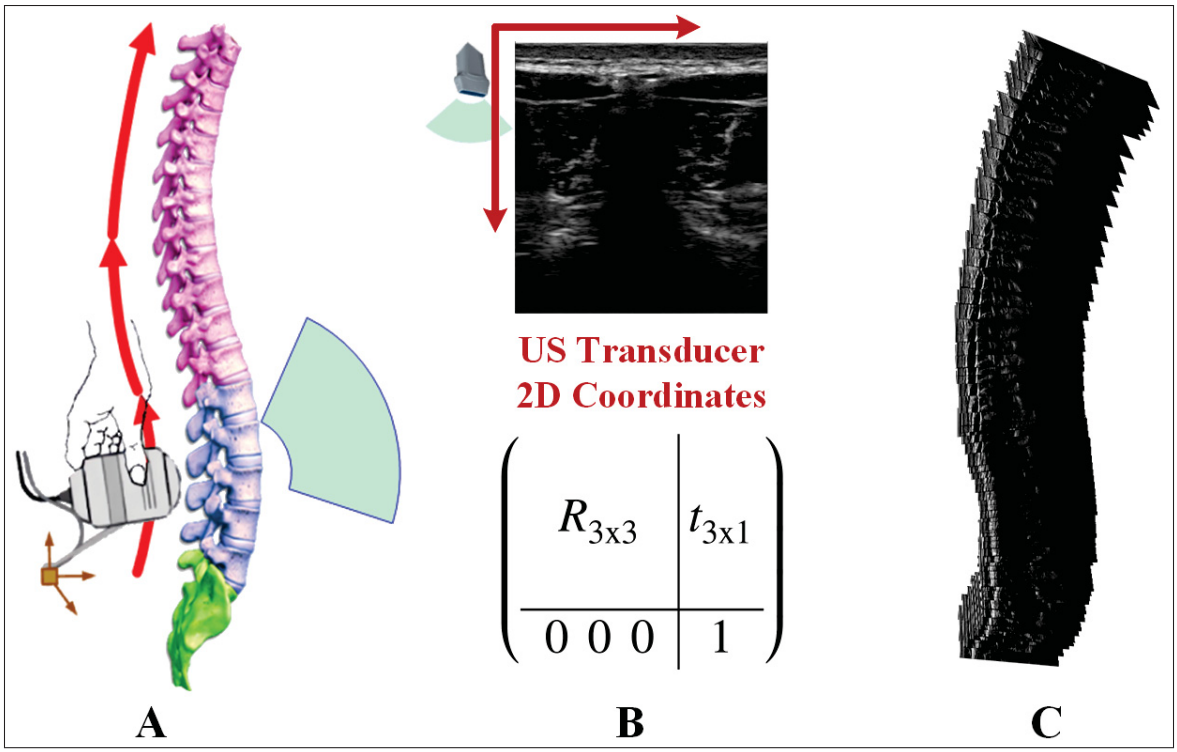

Figure 2.1 Spinal ultrasound sequence acquisition using FH 3D US:
A) Transverse spinal scanning protocol for the spine,
B) Image and 4x4 transformation matrix,
C) 3D ultrasound sequence in world coordinates

world coordinate system, enabling the rendering of a 3D spinal ultrasound sequence in world coordinates. While interpolation and hole-filling can reconstruct the 3D sequence into a volume, our work directly analyzes the sequential data to mitigate resolution loss and compounding errors caused by interpolation and diverse probe operation pressures. Subsequently, we will henceforth use the term “ultrasound sequences” to specifically denote the 3D ultrasound sequences acquired through the utilization of a FH 3D US system.

The collection of spinal ultrasound data from healthy adult volunteers was approved by the Research Ethics Committee of the École de technologie supérieure, with reference ID H20210604. A total of 38 individuals (16 males, 22 females), aged between 18 and 30 years, were recruited. The Ultrasonix Sonix Tablet, equipped with the integrated magnetic position sensor SonixGPS and a linear probe L14-5/38 GPS (Guidance Positioning System), was used to acquire spinal

sequences across the thoracic and lumbar regions in a standing posture. The Tracked Pointer Calibration method (Muratore & Galloway, 2001) was utilized to calibrate the SonixGPS sensor to the ultrasound transducer. As shown in Figure 2.2, the participant adopted an erect stance upon a positioning platform featuring a height-adjustable armrest, with the SonixGPS sensor positioned to the right side of the participant. The sonographer manipulated the ultrasound probe in an inferior-to-superior direction along the participant's spine, maintaining the probe's imaging plane perpendicular to the skin, following the spine's curvature, and keeping it as parallel to the vertebral transverse plane as possible throughout the data acquisition.

Figure 2.2 3D ultrasound sequence acquisition

The data acquisition was conducted within a 30-minute timeframe, involving clothing changes, an experimental briefing, ultrasound sequence collection, and rest intervals. Each sequence collection, covering the L5 to T1 vertebrae, took approximately 1 minute to complete. Two sonographers alternated in sequence collection, obtaining 6-10 sequences for each participant. The unprocessed 3D ultrasound sequences were then transferred from the Sonix Tablet to a desktop computer using the PLUS toolkit (Lasso *et al.*, 2014), which is integrated within the 3D Slicer software. For each participant, the selection of sequences for further research experiments was narrowed down to the best-quality sequence from each sonographer, totaling two sequences per participant. The selection criteria focused on ensuring consistent skin contact, clear visibility of symmetrical vertebral structures, and the central positioning of the spinous process within the scan.

The collected ultrasound sequences were labeled and utilized in subsequent research experiments. In Chapter 3, six sequences of laminae landmarks from three participants were manually labeled to validate the performance of our proposed CNN-based framework and assess its adaptability to varying ultrasound depth values. In Chapter 4, ten sequences of laminae landmarks from ten participants with a fixed generalized ultrasound depth value were labeled to evaluate the upgraded SL-RCN. In Chapter 5, fifteen lumbar vertebral bone surfaces from three participants were labeled to evaluate our proposed LS-TSE method.

2.2 Interpretation of transverse spinal ultrasound images

A high-quality spinal ultrasound sequence is expected to provide accurate anterior-posterior projected 2D curves of vertebral landmarks, which are essential for conducting curvature analysis in scoliosis assessments. Each 2D scan within the sequence ideally presents clear and identifiable vertebral landmarks. Effective interpretation of transverse spinal ultrasound scans not only enhances the quality and reliability of the collected data but also streamlines sonographer training and optimizes the data acquisition process in terms of both speed and image quality. Moreover, a comprehensive understanding of image intensity analysis can unveil intricate patterns within the images, thereby paving the way for potential methodologies in both manual and computerized

image feature extraction and precise localization of vertebral structures. Note that the term "transverse spinal ultrasound scan" refers to planes parallel to the X_vY_v plane in vertebral space, as denoted in Figure 1.2. As introduced in Section 2.1, the spinal transverse ultrasound images were obtained from healthy adults using the Ultrasonix Sonix Tablet with a linear probe L14-5/38 GPS. These images typically encompass the left lamina, spinous process, and right lamina. In this section, we interpret the soft tissue layers visible in the transverse ultrasound images of the vertebral column. Additionally, we propose a straightforward visual scoring system, Rough Visual Intensity Brightness Score (RVIBS), to assist in intensity-based image interpretation.

2.2.1 Soft tissue layers

During transverse spinal ultrasound imaging, the ultrasound wave propagates through the following soft tissues:

- Skin: contains connective tissue, blood vessels, and nerve endings, leading to a low ultrasound wave reflection;
- Subcutaneous Tissue: also known as Superficial Fascia, consists of loose connective tissue, adipose tissue, blood vessels, and nerves, which can produce a moderate reflection of ultrasound waves;
- Prevertebral Deep Fascia: a dense layer of connective tissue that surrounds and separates muscle groups, contributing to a relatively stronger reflection of ultrasound waves compared to the subcutaneous tissue;
- Trapezius Muscles: contain more organized and denser tissue compared to the previous layers, resulting in a strong reflection of ultrasound waves along the direction of the fascicular structure.

Figure 2.3 illustrates an axial-plane anatomical diagram in conjunction with an axial spinal ultrasound image, showcasing annotations of the soft tissues traversed by the ultrasound wave. The annotations are positioned only on the left half of the ultrasound image, providing a reference for intensity contrast on the right half. Two additional points of contrast warrant attention. Firstly, a thin film with strong intensity brightness is observed near the surface of the ultrasound

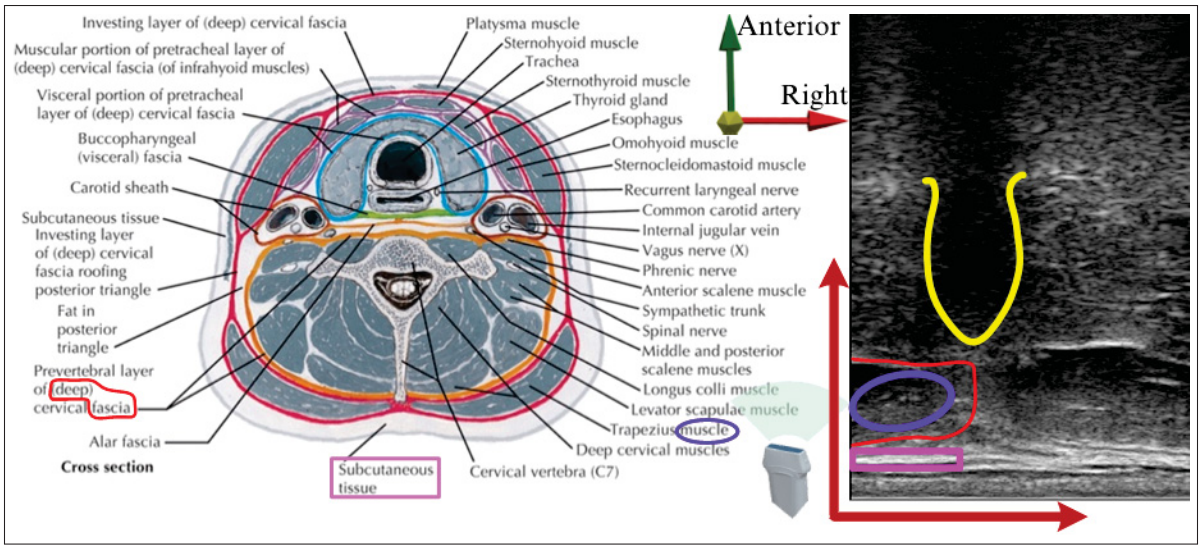

Figure 2.3 Soft tissue interpretation in axial spinal ultrasound images

transducer, i.e., the very top surface of the skin layer. Secondly, the left trapezius muscle marked in a blue ellipse reveals a darkened appearance in the axial ultrasound image, which is attributed to the direction of muscle fibers. The fascicular structure of the muscles becomes more discernible when visualized in the longitudinal view (Pillen, Boon & Van Alfen, 2016).

As the ultrasound wave propagates deeper from the skin to the vertebra, the axial ultrasound image reveals five vertical layers: skin, subcutaneous tissue, horizontal trapezius muscle (H muscle), spinous process (composed of vertical muscle (V muscle) wave response and spinous process acoustic shadow), and vertebral body (situated below the laminae). Note that the depth and image intensity values of the three soft tissue layers (prior to the spinous process) vary due to differences in tissue thickness, fat distribution, and muscle composition. By focusing on a region of interest (RoI) vertically from the skin to the laminae bone surface response, it becomes possible to disregard the vertebral body layer, thereby bypassing the thick bone response and the unexpected anterior complex.

2.2.2 Axial intensity projection curves (IPCs)

In spinal ultrasound imaging, the vertical bone surface contiguous to the spinous process predominantly remains indiscernible, primarily due to the inherent limitations in capturing vertical edges. Consequently, achieving precise delineation of the spinous process outline poses a formidable challenge, as only the uppermost horizontal surface of the spinous process manifests as a perceptible short bright segment. However, even this limited segment is sometimes nearly or entirely invisible because of ultrasound artifacts. These artifacts further exacerbate the difficulty in accurately capturing and characterizing the spinous process in spinal ultrasound imaging. In the task of localizing the spinous process, two key features offer positional information: the horizontal high-luminance curves corresponding to the laminae and the spinous process acoustic shadow. However, due to the intrinsic inability to capture vertical edges in ultrasound images, the spinous process acoustic shadow often blends with the ultrasound response of surrounding soft tissues, particularly in sub-optimal quality images. This blending results in a voluminous spinous process acoustic shadow with an unpredictable shape.

The intensity projection method involves projecting the intensity values of ultrasound image pixels onto the horizontal and vertical axes, resulting in the conversion of a single 2D intensity matrix into two 1D cumulative intensity projection curves (IPCs). This method leverages the ultrasound wave response (represented by white pixels) and the acoustic shadow (represented by black pixels), capitalizing on their potential for image pattern recognition and landmark localization. Brignol *et al.* (2020) employed the whole-image IPCs (WI-IPCs) for identifying three key vertebral landmarks, namely the spinous process and the left/right lamina, from transverse ultrasound scans. The column address (x_{SPL}) of the Spinous Process Landmark (SPL) was determined by finding the midpoint between two zero-crossings adjacent to the minimum point on the linearly detrended vertical intensity projection curve (V-IPC), and the row address (y_{SPL}) of the SPL was obtained by identifying the first minimum value on the linearly detrended horizontal intensity projection curve (H-IPC) after the zero-crossing.

However, their proposed method assumes certain conditions, including the presence of a pristine and symmetrical vertebral W-shape, a clear acoustic shadow beneath the W-shape, and the uniformity of isotropic biological tissue. These assumptions contribute to the maintenance of a seamless boundary between the tissue and the acoustic shadow beneath the bone structure. In practical scenarios, FH 3D US scans are generally asymmetric and are commonly prone to various sources of interference such as noise, ultrasound artifacts, and thick bone responses. As a result, the accuracy of WI-IPCs analysis heavily relies on the quality of the transverse scan, particularly in terms of preserving symmetrical and uniform wave responses as well as ensuring clarity in the acoustic shadow beneath the W-shape (Li *et al.*, 2023). It should be noted that the method's performance is uncertain when faced with asymmetrical wave responses of soft tissue or unexpected observations of the anterior complex, which can easily occur when a 2D scan is axial-rotated, leading to unreliable landmark localization.

To minimize the influence of thick bone responses, unexpected observations of the anterior complex, and unforeseen imaging anomalies, such as dark areas resulting from insufficient

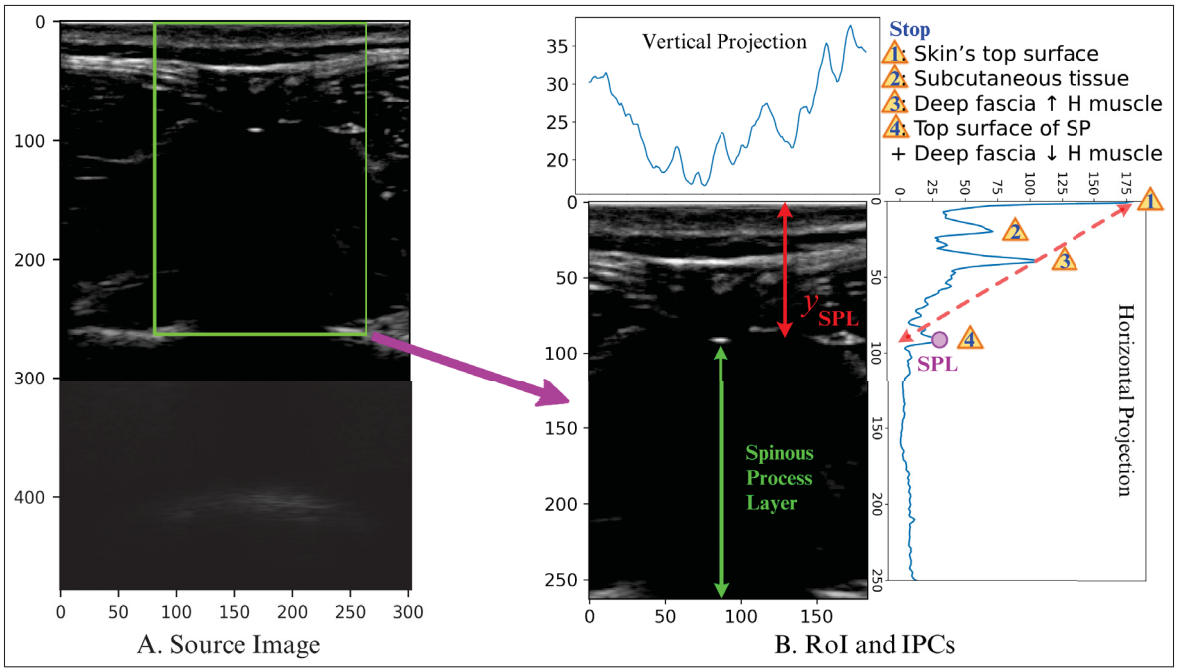

Figure 2.4 Axial intensity projection curves (IPCs) from RoI of transverse spinal ultrasound images

transducer-skin contact, our study concentrates on analyzing IPCs within a RoI that specifically targets the spinous process region. Figure 2.4 presents an example of IPCs within a manually determined RoI vertically spanning from the skin to the lamina, targeting the spinous process. The H-IPC exhibits peaks and valleys that reflect the composition of the different tissue layers encountered by the ultrasound beam. Simultaneously, the V-IPC depicts a horizontal pattern tracing the shape of the spinous process. In the subsequent section (Section 2.2.3), we introduce a rough visual scoring method designed to assist in the interpretation of the H-IPC.

2.2.3 Rough Visual Intensity Brightness Score (RVIBS)

To simplify the terminology, we utilize "H muscle" and "V muscle" to represent the trapezius muscle and the paraspinal muscles surrounding the spinous process, respectively. Vertically from the skin to the lamina bone response, we assign RVIBSs ranging from -1 to 3 to the four layers and two special thin films, the top surface of the skin, and the prevertebral deep fascia. Specifically, the scores are as follows: skin = 2, subcutaneous tissue = 3, H muscle = 1, spinous process = V muscle (2) + acoustic shadow (-1) = 1, top surface of the skin = 3, and prevertebral deep fascia = 3. According to the gross RVIBS, both the spinous process layer and the H muscle layer share the same minimum RVIBS level.

The sense of RVIBS is explicitly reflected by the H-IPC in Figure 2.4. Starting from the origin and extending along the positive direction of the depth axis, the curve exhibits a consistent downward trend attributed to the ultrasound wave attenuation. It reaches four peaks as it passes through four distinct stops characterized by the highest RVIBS value 3, subsequently transitioning from a downward trend to a horizontal advancement. These four RVIBS-3 stops, in order of the ultrasound propagation depth, correspond to the following regions: the top surface of the skin, subcutaneous tissue, deep fascia located above the H muscle, and the combined area comprising the top surface of the spinous process and the deep fascia beneath the H muscle. The vertical coordinate of the SPL y_{SPL} can be localized at the end of the fourth stop, which will be detailed in Section 5.2.2.

2.3 Labeling protocols on transverse spinal ultrasound images

In this section, we present two devised manual labeling protocols applied to transverse spinal ultrasound images. One protocol provides formal guidelines for labeling the most representative pairwise point landmarks representing the laminae, specifically focusing on the most relevant part of the visible lamina bone surface in ultrasound images. A systematic 3D sequential labeling approach is designed towards a pair of smooth 3D lamina curves that can represent the vertebrae in a comprehensive manner. The other protocol involves labeling the vertebral bone surface, encompassing the left lamina, spinous process, and the right lamina. These labeling protocols underwent validation by Professors Carole Fortin and Philippe Paquette, two medical professionals with strong expertise in vertebral ultrasound.

2.3.1 Lamina landmark labeling protocol

In the current literature, extracting single-point landmarks representing the laminae is an ill-defined task, as there is no formal guideline as to the most relevant part of the visible lamina bone surface in ultrasound images. We propose a protocol to solve this issue.

Figure 2.5 Row II presents a subjective preference diagram tailored to assist manual lamina landmark localization in sub-optimal quality transverse ultrasound images, where only a portion of the lamina bone surface is visible. In accordance with the physics of ultrasound wave propagation, the bone surface that is perpendicular to the direction of the wave generates the strongest response, resulting in a horizontal curve of high luminance (white) in the image. The preference diagram assigns the highest priority to the pair of points located slightly higher than the flat bottom of the left/right laminae regions, close to the spinous process and symmetrical with respect to the center shadow. It subsequently monotonically decreases the priorities on both sides. These preferred landmarks, typically located in the middle of high luminance horizontal bone surface curves, correspond to the paired green dots in Row I. The diagram also assigns lower priority to the points along the left/right edge of the bone surface curve, either inner (purple) or outer (red) relative to the spinous process, which are considered as less preferred

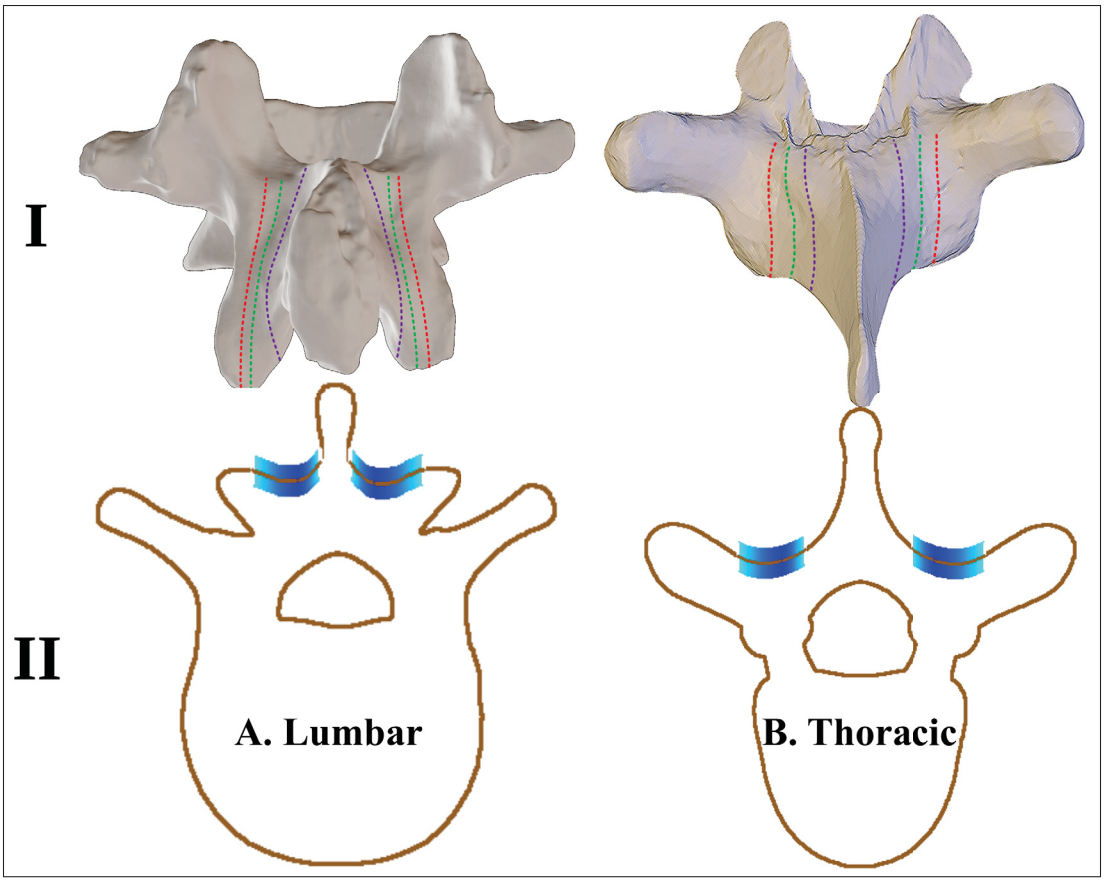

Figure 2.5 3D Lamina Landmark Labeling Protocol

landmark candidates. The color in the diagram does not convey quantitative information but serves as a visual guide to select lamina landmarks that lie as close as possible to the region in dark blue on the figure. Some positional deviation is tolerated to favor the creation of smooth laminae curves over jagged ones, ensuring the effective localization of these landmarks.

We define a left/right lamina curve as the curve that optimally fits all left/right lamina landmarks across a 3D ultrasound sequence. Figure 2.5 row I illustrates the anticipated paired vertebral laminae curves for individual lumbar vertebra and thoracic vertebra, which are comprised of paired sequences of lamina landmarks from the same vertebra. The green dotted curves represent the optimal pair, while the purple and red dotted curves correspond to the inner and outer boundary pairs, respectively; any pair of curves within the inner and outer boundaries may be considered acceptable, provided that their central curve lies parallel to the spinal cord.

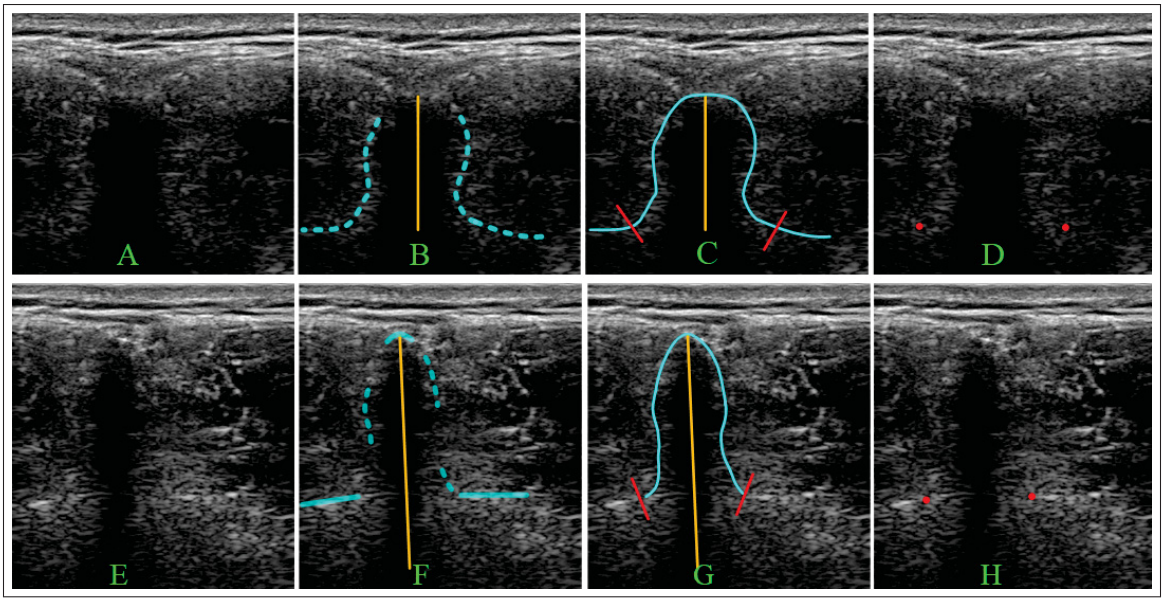

Figure 2.6 Practical implementation of manual 2D landmark labeling based on visual guidance

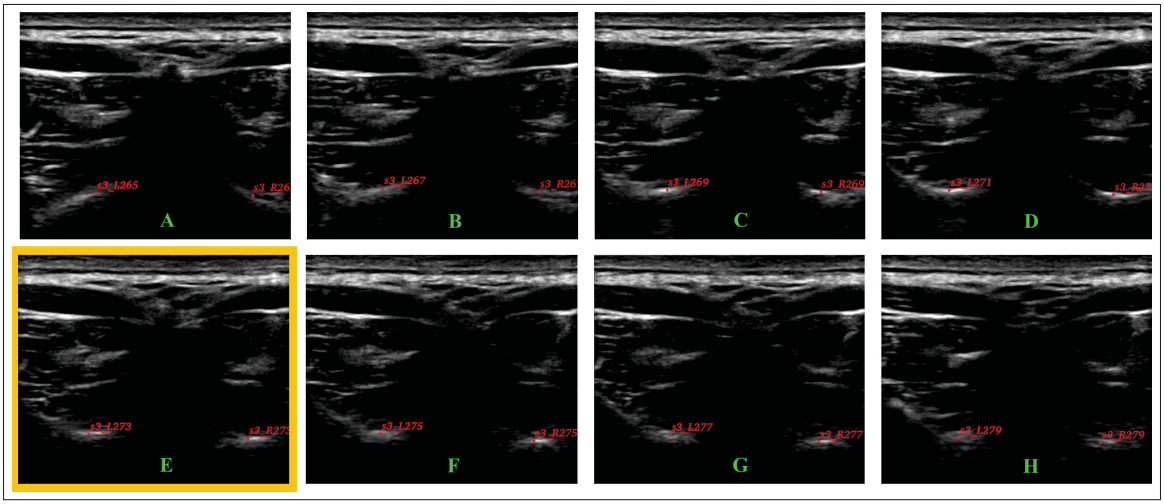

Figure 2.7 Manual landmark localization over an image sub-sequence covering one vertebra

Figure 2.6 illustrates a practical example of manual lamina landmark labeling for lumbar and thoracic vertebrae. The orange line represents the center of the spinous process, while the cyan

curves depict a virtual bone surface spanning the left and right lamina regions. Landmarks are annotated on the cyan bone surface contour by referencing the orange line.

Depending on tissue thickness, fat distribution, muscle composition, and, most importantly, the probe operation by the sonographer, the ultrasound sequence corresponding to one vertebra may exhibit various patterns of asymmetrical high-luminance laminae curves. Consequently, simply selecting the best lamina landmark pair on each frame independently leads to jagged laminae curves in 3D. In order to fine-tune and align the laminae landmark pairs, thereby achieving smoother vertebral cardinal splines, we propose the notion of **anchor frame**, and its extracted landmark pair is noted as the **anchor landmark pair**. We determine the anchor frame from comparably high-quality frames and subjectively choose the one whose anchor landmark pair effectively guides the alignment of both the anterior and posterior parts of vertebral landmark pairs during fine-tuning, while ensuring that the fine-tuned landmarks remain consistent with the visual guidance. The anchor landmark pair plays a vital role in ascertaining the lateral span of the smoothed pair of 3D lamina curves. Figure 2.7 demonstrates a labeled sub-sequence of

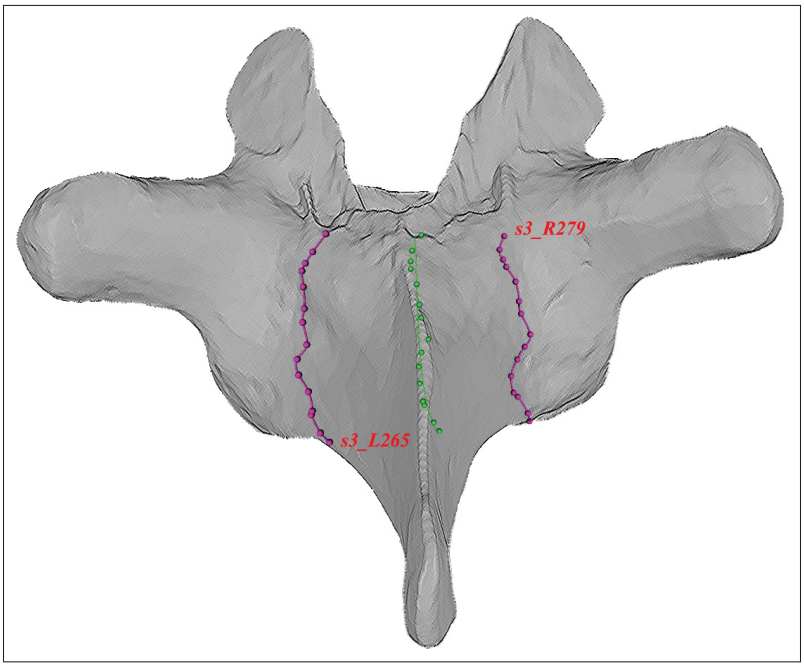

Figure 2.8 Vertebral 3D lamina curves
derived from an ultrasound sub-sequence

thoracic frames belonging to one vertebra, conforming to the labeling protocol and refined using the anchor frame. Figure 2.8 illustrates the extracted 3D lamina curves of the labeled thoracic sub-sequence, with the addition of middle points to indicate the spinal cord.

The Normalized Discrete Fréchet Distance (NDFD), which is detailed in Chapter 4, Section 4.2.4, serves as the primary metric for evaluating the disparity of the extracted 3D lamina curves. Additionally, this chapter includes an analysis of the intra- and inter-variation of the proposed sequential lamina landmark labeling protocol, presented in Section 4.3.2.

2.3.2 Vertebral bone surface labeling protocol

In practice, particularly when dealing with sub-optimal ultrasound images, the identification of vertical bone surfaces adjacent to the spinous process is notably challenging. This difficulty arises primarily due to the inherent limitations of ultrasound scanning in capturing clear vertical edges. Figure 2.9 demonstrates the invisibility of the border of the spinous process acoustic shadow, thereby highlighting the challenging nature of extracting the vertebral bone surface, even for experienced human observers.

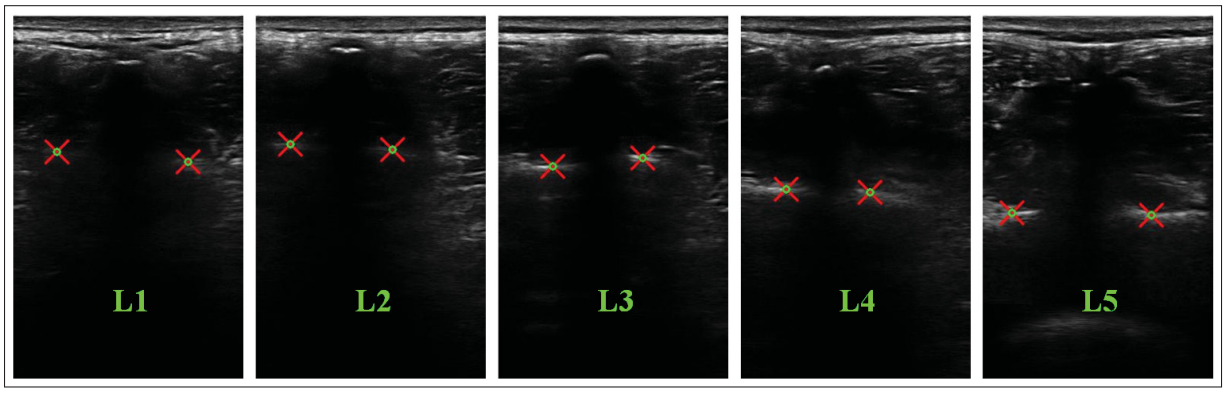

Figure 2.9 Raw ultrasound image samples of various lumbar vertebral levels from a participant with pairwise laminae landmarks labeled according to the protocol described in Section 2.3.1

The image gradient is an essential tool in digital image processing, representing the directional change in intensity or color at each pixel. It is mathematically expressed as a vector of partial

derivatives in the x and y directions. A gradient image can be generated to represent the image gradient, where the intensity of each pixel in the new image is typically calculated based on the magnitude of the gradient vector in the original image. In ultrasound images, this aids in delineating anatomical structures by highlighting boundaries where acoustic impedance changes. These changes typically occur at interfaces between different types of tissues or between tissues and bone structures, areas where the gradient magnitude is high due to rapid intensity variation.

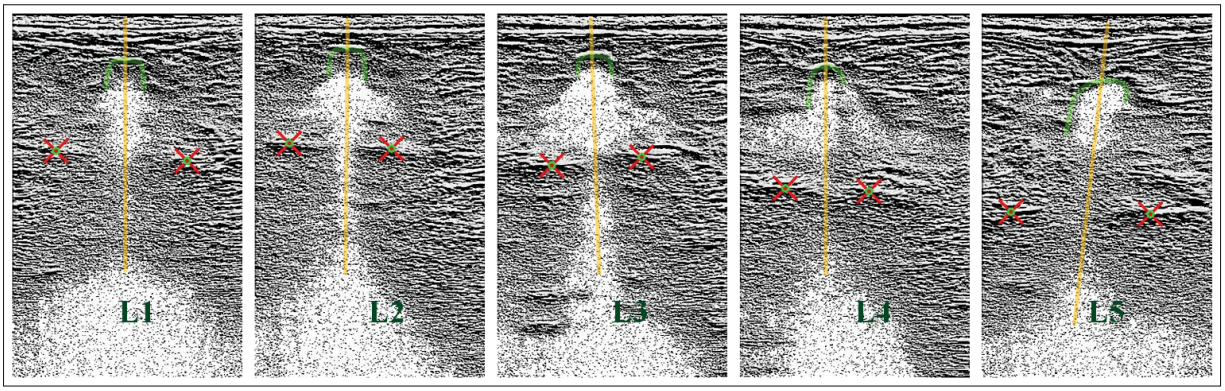

Figure 2.10 Gradient Images of Figure 2.9

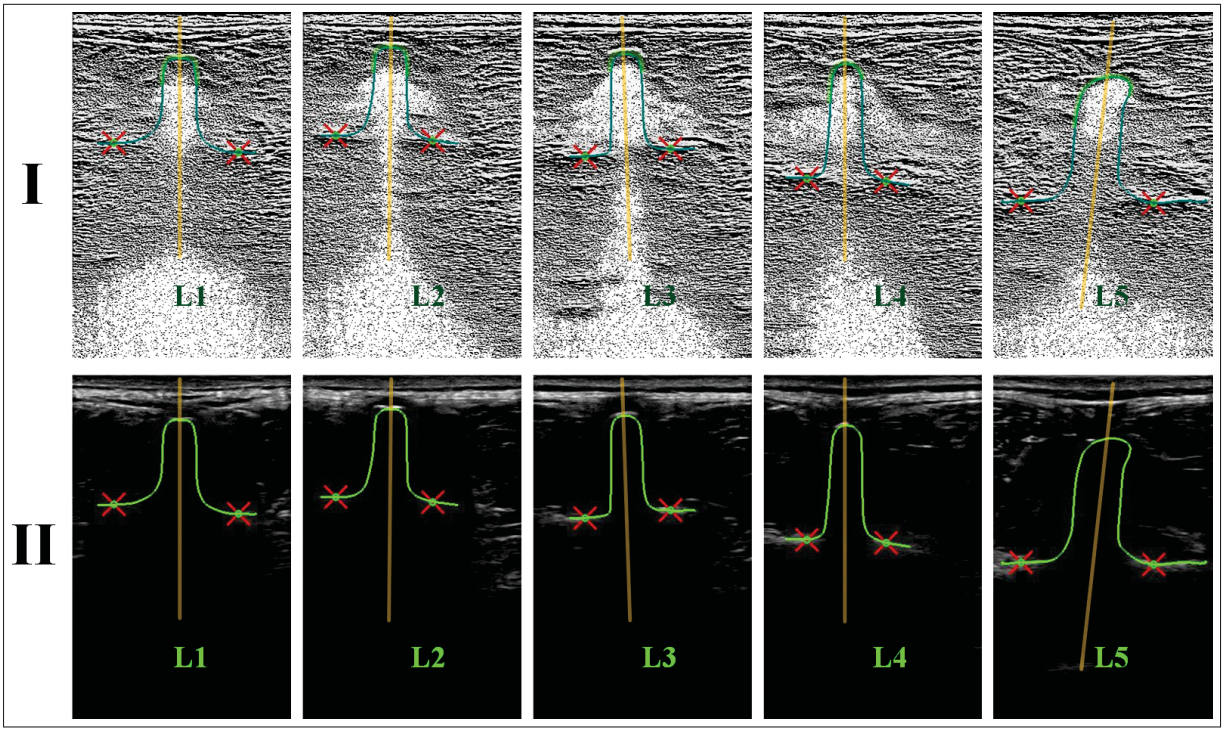

Figure 2.11 Manually Labeled Vertebral Bone Surfaces

Figure 2.10 shows the gradient images derived from the raw lumbar ultrasound image samples. In contrast to the original ultrasound images, the gradient images provide enhanced clarity and visibility of the spinous process structure. Notably, the center of the spinous process (indicated by the yellow straight line) becomes discernible, and the vertical boundary of the spinous process becomes partially observable (indicated by the light green curve). The vertebral bone surface can be identified by leveraging the pattern reflected in the image gradient. The left and right lamina bone surfaces are determined along the horizontal edges where pairwise laminae landmarks are located. Taking the yellow straight centerline as a reference, the appropriate vertical boundary of the spinous process is traced following the image gradient pattern. To achieve a continuous and smooth bone surface, a smooth curve adhering to the image gradient pattern connects the partially observable boundaries of the spinous process with the identified left and right lamina bone surfaces. The integration of these components enables the delineation of a continuous and smooth vertebral bone surface, as depicted in Figure 2.11.

One objective of this thesis is to develop an automated method for vertebral surface extraction. In Chapter 5, we introduce the LS-TSE method to achieve this objective by leveraging the extracted laminae landmarks. To assess the efficacy of the LS-TSE method, we propose this vertebral bone surface labeling protocol designed to generate ground truth data. This protocol is crucial for evaluating the accuracy of the LS-TSE method for automated vertebral surface extraction.

CHAPTER 3

LAMINA LANDMARK DETECTION FROM 2D ULTRASOUND IMAGES USING A DEEP LEARNING FRAMEWORK

3.1 Introduction

Vertebral landmarks (laminae, spinous process, and transverse process, etc.), are the meta-elements that make up the spine representation (Chen *et al.*, 2013; Cheung *et al.*, 2015). These landmarks provide valuable spinal shape information and distinct ultrasound imaging features, enabling anatomical structure-based research in various applications such as epidural anesthesia (Grau, Leipold, Conradi, Martin & Motsch, 2002; Ghosh, Madjdpour & Chin, 2016; Sadeghi, Patel & Carvalho, 2021a), image-to-image registration (Glocker, Zikic & Haynor, 2014; Che, Mathai & Galeotti, 2017; Gueziri, Drouin, Yan & Collins, 2019c), the severity assessment of spine deformity (Chen, Lou & Le, 2011; Cheung, Law & Zheng, 2013), axial vertebral rotation measurement (Vo, Lou & Le, 2015), etc. Among these landmarks, laminae are particularly important for analyzing spine deformities in three dimensions due to their interconnectedness across different vertebral levels through articular processes. In the context of FH 3D US imaging, laminae landmarks hold essential importance as they are directly attached to the spinal cord and are symmetrically distributed around the shadow of the SP in transverse ultrasound scans.

Chen *et al.* (2013) proposed the Center-of-Lamina (CoL) method for assessing the coronal curvature of children with AIS; the reliability and validity of using this CoL method have been statistically evaluated and perfectly verified over the years for both scoliosis deformity assessment and vertebral rotation measurements (Young, Hill, Zheng & Lou, 2015; Wang, Li, Lou & Wong, 2015; Chen, Le & Lou, 2016; Wu, Liu & Wong, 2020b). Sayed, Khodaei, Hill & Lou (2022) also proved the reliability of measuring kyphotic angles on ultrasound images using the CoL method for identifying spinous process.

Nevertheless, the automated detection and precise localization of lamina landmarks in ultrasound images present significant challenges. These challenges include the complex anatomy of the

vertebrae, variations in contrast across different tissues, and ambiguity in ultrasound reflections from bone surfaces. Currently, there is a lack of laminae landmark recognition algorithms that have been tested on poor-quality 2D ultrasound scans or interlaminar scans. In pursuit of advanced automated intelligent image analysis, we want a lamina landmark extraction algorithm capable of detecting the presence of laminae and predicting the coordinates of the most representative lamina points.

In this chapter, we propose a novel deep learning framework designed to automatically detect the presence of laminae and accurately estimate their landmark coordinates from 2D transverse ultrasound images. We explore depth values in ultrasound settings and conduct grid searches to identify a feasible configuration for the proposed framework. Our experiments validated the effectiveness of our proposed framework by overcoming several primary barriers. These included identifying the optimal criterion for laminae structure identification from a range of loss functions, optimizing hyper-parameters to guarantee robust training, and assessing the performance of a range of foundational CNN models.

3.2 Methods

3.2.1 Automated lamina landmark detection

Deep learning methods are widely used in object detection applications (Ren, He, Girshick & Sun, 2015; Handalage & Kuganandamurthy, 2021). These assess whether a certain object is present with a confidence probability and mark out the object localization by predicting an anchor box if the confidence is larger than a threshold. Inspired by this approach, we propose a laminae landmark detection framework that predicts the visibility of laminae structures and localizes pairwise landmark coordinates, with the anchor box replaced with a point.

Figure 3.1 shows a CNN-based (e.g., VGG, ResNet, DenseNet, etc.) framework to process an entire image in a single pass while concurrently addressing both a classification task and a prediction task. Specifically, it detects the visibility of laminae and localizes their pairwise

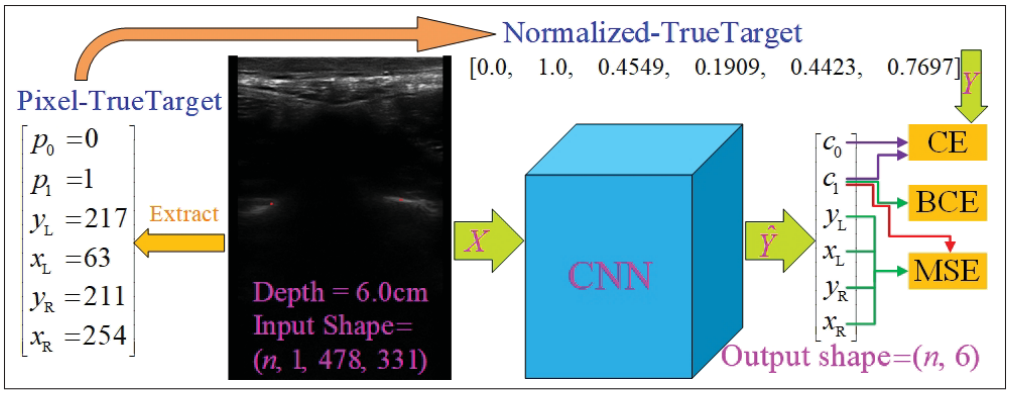

Figure 3.1 CNN-based laminae landmark detection framework

landmarks if the laminae are visible, integrating these functions into a unified regression problem. The input sample X to the framework is a 2D grayscale ultrasound image. The output prediction $\hat{Y}$ is a tuple of six variables, with two variables (c_0/c_1) indicating the absence or presence of visible laminae, and four variables $(x_L, y_L/x_R, y_R)$ presenting the image coordinates of their landmarks. Both X and the true target (Y) are normalized before training; detailed data processing steps are discussed in Section 3.2.2. This approach seamlessly translates raw image pixels into class confidence and landmark coordinates in a simultaneous operation.

In this framework, the mini-batch objective function is a weighted combination of the classification loss $\mathfrak{L}$c and the landmark coordinate losses $(\mathfrak{L}_{l\_L} + \mathfrak{L}_{l\_R})$:

$$\mathfrak{L}_l = \frac{1}{P}\sum_{m=1}^{P}(x_Y^m - x_{\hat{Y}}^m)^2 + (y_Y^m - y_{\hat{Y}}^m)^2, \quad \forall m \text{ where } c_Y^m = 1 \tag{3.1}$$

$$\mathfrak{L} = \mathfrak{L}_c + w\,(\mathfrak{L}_{l\_L} + \mathfrak{L}_{l\_R}) \tag{3.2}$$

where m is the index of image samples, P is the number of sampled positive frames in a mini-batch, $\mathfrak{L}_l$ is a single landmark coordinate loss, subscripts L/R refer to left/right lamina, w denotes empirically determined weights, and $\mathfrak{L}$ denotes the final loss of each mini-batch. The classification loss $\mathfrak{L}_c$ is induced by incorrect identification of the laminae structure. The localization loss $\mathfrak{L}_l$ results from the discrepancy between true and predicted landmark coordinates.

Regarding the determination of the classification component, Cross Entropy (CE) and Binary Cross Entropy (BCE) are well-known criteria for assessing the alignment between predicted probabilities and actual categories, represented by equations 3.3 and 3.4, respectively. CE is typically used in multi-class settings with a softmax activation function in the output layer to ensure that the output probabilities sum to one, making them interpretable as class probabilities. On the other hand, BCE, which simplifies to a form of CE for binary classification, uses a sigmoid activation function to output a single probability for the positive class. In a binary classification scenario, both CE and BCE can effectively fulfill the same role, with CE adjusted in the output layer and loss function formulation to handle two output probabilities using one-hot encoding. Considering that the landmark coordinate prediction loss $\mathfrak{L}_l$ is calculated using the Mean Squared Error (MSE), as shown in equation 3.1, it is worth considering mean Squared Error (SE) (equation 3.5) as a potential alternative to CE and BCE for the classification loss component $\mathfrak{L}c$ in the weighted combination of the total loss function.

$$CE = -\sum_{i=0}^{S-1} c_i^Y \log\left[Softmax(c_i^{\hat{Y}})\right] \tag{3.3}$$

$$BCE = -c_1^Y \log\left[Sigmoid(c_1^{\hat{Y}})\right] - (1 - c_1^Y) \log\left[1 - Sigmoid(c_1^{\hat{Y}})\right] \tag{3.4}$$

$$SE = \left[c_1^Y - Sigmoid(c_1^{\hat{Y}})\right]^2 + \left[c_1^Y - Sigmoid(c_1^{\hat{Y}})\right]^2 \tag{3.5}$$

In equation 3.3, S is the number of classes. In the validation experiments, we will use the MSE method to determine the loss from coordinate prediction and rank the effectiveness of using CE, BCE, and MSE methods for the determination of the classification loss.

3.2.2 Data processing

In this section, we provide an overview of the data source and outline the data processing steps undertaken for the subsequent two-phase experiments. As indicated in Table 3.1, our raw dataset comprises six sequences of 2D ultrasound images acquired from three participants. The “Depth” parameter represents the vertical field of view during ultrasound scanning, with an increase

in depth resulting in narrower ultrasound images, as depicted in the "Valid-Shape" column. To ensure consistency, all images were reshaped to a uniform size of (478, 331) by applying zero-padding on either side for sequences with a depth greater than 5.5cm.

Table 3.1 Description of data source and Data Group: PTCP is short for participant, n_T is the number of total frames in a sequence and r_P is the ratio of positive frames, DG is short for Data Group, and Val- is short for validation

PTCP	Sequence 1				Sequence 2				DG-2 Val-Set	
	n_T	r_P	Depth	Valid-Shape	n_T	r_P	Depth	Valid-Shape	n_T	r_P
1	623	13.32%	5.5cm	(478, 331)	607	16.80%	6.5cm	(478, 281)	1230	15.04%
2	1097	35.28%	6.0cm	(478, 303)	1134	30.25%	6.0cm	(478, 303)	2231	32.72%
3	627	25.52%	6.5cm	(478, 281)	410	14.88%	6.5cm	(478, 281)	1037	21.31%

The lamina landmarks were manually labeled following the protocol described in Section 2.3.1 and were subsequently validated by two spinal ultrasound experts. The coordinates of these landmarks were extracted and formatted into a 6-element vector, as exemplified in the left portion of Figure 3.1. Prior to training the deep learning agent, both the raw (unlabeled) images and the true targets were normalized. This involved normalizing the gray-scale image data and the laminae landmark coordinates to ensure they fell within the range of [0, 1].

Two distinct Data Groups (DGs) were created for K-Fold Cross Validation in the two-phase experiments. DG-1 was constructed by randomly shuffling the images from the six sequences into five equally sized folds. On the other hand, DG-2 was generated using a leave-one-participant-out strategy, resulting in a 3-fold validation set where each fold contained the two sequences from a single participant. In the validation experiments, the phase-one experiments utilize DG-1 to confirm the convergence of our proposed framework and rank the effectiveness of three loss criteria (CE, BCE, and MSE) in a scenario that exclusively classifies the visibility of laminae. The phase-two experiments employ DG-2 to conduct a grid search for the optimal hyper-parameters and subsequently validate a range of CNN models.

3.3 Experiments and results

In this section, we present the qualitative and quantitative validation results obtained from the preliminary lamina landmark identification and localization experiments conducted on two DGs. The VGG (Simonyan & Zisserman, 2014) and ResNet (He, Zhang, Ren & Sun, 2015) series, which are foundational CNNs with streamlined architectures, represent milestone achievements in the evolution of neural networks. Consequently, our experiments are centered around these two series.

Table 3.2 Classification Criteria Comparison:
5-Fold cross-validation on DG-1

	CE	BCE	MSE
mean F1 ± std	0.955 ± 0.010	0.961 ± 0.008	0.951 ± 0.012

Primarily, three experiments were conducted to compare classification loss functions using DG-1 and ResNet34 in a scenario focused solely on classification, where the landmark coordinate prediction error is consistently zero. Given the significant imbalance (there are far fewer images with visible laminae than images without) in the dataset, the F1 score was selected as the metric to evaluate classification performance, as accuracy may provide a misleading representation. Table 3.2 presents the results from a 5-fold cross-validation to assess the three classification loss functions. As DG-1 was created by randomly shuffling images from all participants, the similarity between image features across the folds (training and validation sets) was high. This resulted in a mean F1 score of 0.95 or higher across the three experiments, meanwhile demonstrating the convergence of our framework. Among the three loss functions tested, BCE achieved the highest average F1 score with the least variation. Consequently, BCE will be employed as the classification loss method in subsequent experiments.

Table 3.3 shows the cross-validation results from lamina landmark detection experiments using DG-2 (3-folds, each fold corresponds to one participant) and the BCE classification loss function. Due to the limited number of participants and the high inter-subject variability in spine shape, image shape, sequence length, and image quality, the mean and deviation of the F1 may not be

Table 3.3 Grid Search 3-fold Cross Validation Results using DG-2: Column "Exp." is the index of experiments, w is the loss function weight parameter from equation 3.2, the rightmost two columns show the MDE with the standard deviation of true-positive samples of left and right lamina respectively

Exp.	CNN Model	w	Performance F1 score			True-positive MDE ± std	
			PTCP-1	PTCP-2	PTCP-3	Left	Right
1	ResNet34	1	0.653	0.773	0.591	2.0 ±1.3	2.4 ±1.4
2	ResNet34	10	**0.689**	0.792	0.619	2.3 ±1.4	1.9 ±1.5
3	ResNet34	20	0.661	**0.798**	0.611	**1.9 ±1.5**	2.0 ±1.5
4	ResNet34	**30**	0.680	0.796	0.648	2.1 ±1.3	**1.8 ±1.2**
5	ResNet34	40	0.678	0.767	**0.667**	2.5 ±1.7	2.2 ±1.7
6	VGG16_bn	30	0.593	0.667	0.660	2.3 ±1.9	2.9 ±2.1
7	VGG19_bn	30	0.627	0.660	**0.705**	2.3 ±1.4	2.7 ±1.8
8	ResNet18	30	0.631	0.792	0.621	2.3 ±1.3	1.9 ±1.1
9	ResNet50	30	0.667	0.790	0.643	3.5 ±2.1	3.3 ±1.9
10	ResNet101	30	0.667	0.765	0.627	3.3 ±2.0	3.1 ±1.8
11	ResNet152	30	**0.708**	0.776	0.615	2.8 ±1.9	2.2 ±1.9
Brignol et al.			N/A			7.0 ±4.4	6.0 ±4.6

necessarily informative; therefore, we listed F1 values for all three folds. Exp. #1-5 compared different values of w, with $w = 30$ consistently yielding good F1 values across folds, and landmark mean distance error (MDE) in mm with the lowest standard deviation. Using $w = 30$ for Exp. #6-11, two VGG models with batch normalization and the ResNet series models were tested, and their overall F1 performance and coordinate prediction accuracy did not surpass those of Exp. #4. As an additional experiment, we tested Brignol et al.'s method (Brignol *et al.*, 2020) on the same data for landmark localization comparison. While the MDE results reported should be interpreted with care because of differences in labeling protocol between the approaches, the much lower standard deviations achieved by our method suggest that it is more robust and achieves more consistent results than Brignol et al.'s.

Figure 3.2 shows the 3D lamina curve comparisons using three lamina landmark extraction methods for sequence 1 of three participants: manual labeling, Brignol et al.'s method, and our proposed CNN framework (Exp. #4). Our proposed CNN agent identified a higher number of positive frames compared to manual labeling, indicating a potential superiority of AI agents over

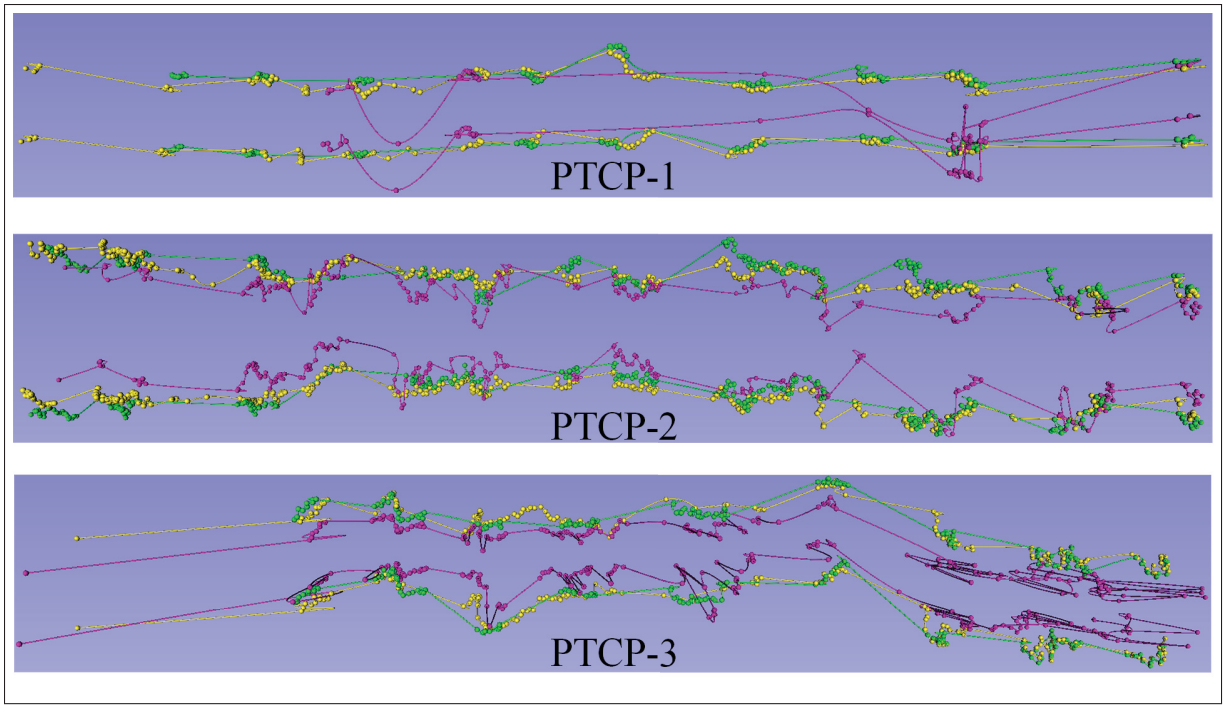

Figure 3.2 Lamina landmark localization comparison in the coronal plane

human vision when dealing with low-contrast images. By using the green lamina spline curves as a reference, a comparative analysis of the other two color curves reveals that the yellow curves, generated by our proposed CNN agent, exhibit closer alignment with the reference as well as a smoother trajectory compared to the purple curves generated using Brignol et al.'s method.

3.4 Discussion

In this study, we executed experiments in two phases. The first phase assessed the effectiveness of CE, BCE, and MSE in identifying laminae structures, while the second phase evaluated the performance of the VGG and ResNet series in detecting lamina landmarks.

Phase one primarily focused on classification tasks, which inherently involve probabilistic outcomes. The MSE criterion treats classification as a regression problem, resulting in less efficient learning due to the linear behavior of its gradient. This behavior fails to adapt well to the categorical nature of classification. Conversely, CE and BCE criteria impose a logarithmic

penalty for incorrect predictions, promoting faster and more stable convergence of the models by placing emphasis on accurate probability estimation. The results of our phase-one experiments confirmed that the CE and BCE criteria, due to their superior efficacy in managing classification tasks, are preferred over MSE. Furthermore, BCE demonstrated superior accuracy and robustness compared to CE. This is attributed to the fact that CE requires adjustments to the model's output layer to manage two probability outputs and utilizes one-hot encoding, whereas BCE deals directly with a single probability output for the positive class and a binary label. As a result, BCE is more straightforward and computationally efficient for binary classification due to its more direct approach to handling binary data.

Phase two validated the performance of VGG and ResNet series in transverse ultrasound image analysis. Both VGG and ResNet architectures have significantly influenced the field of deep learning, especially in computer vision. VGG demonstrated the efficacy of deep and uniform architectural patterns, while ResNet introduced residual learning, enabling the construction of even deeper networks by addressing the issue of vanishing gradients. ResNet's design incorporates skip connections and replaces fully connected layers with global average pooling, enhancing computational efficiency and reducing memory usage compared to VGG. Nevertheless, VGG networks performed commendably in the landmark detection task, with its MDE slightly exceeding that of ResNet34 (Exp. #4) by less than 1mm.

However, the current CNN-based framework may not be ideal for achieving the objective of providing smooth 3D lamina landmark curves that can then be used for 3D spinal deformity analysis. The framework's architecture is primarily designed for the analysis of single 2D images, and thus inherently lacks 3D inferential capabilities to account for the spatial relationships between the 2D images. Consequently, the predicted sequential lamina landmarks are independent of each other across frames and fail to naturally form smooth curves in 3D space. Furthermore, the MDE metric, which applies to individual 2D ultrasound images, is insufficient for capturing differences in network performance regarding the smoothness of their extracted 3D lamina curves. To effectively extract smooth 3D lamina curves from ultrasound sequences, an enhanced

AI framework is required, accompanied by a new performance metric specifically designed to assess the network's efficiency in extracting smooth 3D lamina curves.

3.5 Conclusions

In this chapter, we proposed the first lamina detection method tailored for transverse spinal ultrasound images, leveraging a CNN-based deep learning framework to simultaneously identify the presence of visible laminae and extract pairwise lamina landmarks. The framework was validated on low-quality ultrasound sequences covering both the lumbar and thoracic regions, demonstrating its potential applicability to medical landmark localization in 2D images beyond spinal ultrasound.

For the validation of the proposed framework, two distinct DGs were created, facilitating K-Fold cross-validation of our automated landmark localization framework. Results from DG-1 experiments confirmed the convergence of the classification component, while DG-2 experiments provided leave-one-participant-out validation results. The achieved physical MDEs of (2.1 ± 1.3)mm and (1.8 ± 1.2)mm for the left and right lamina in true-positive images highlight the accuracy and reliability of our proposed method. Overall, our study establishes a promising approach for automated lamina detection and landmark localization, with potential applications in spinal ultrasound and beyond.

In the subsequent chapter, a more detailed evaluation of the CNN-based deep learning framework will be presented, featuring cross-validation with seven participants and a test experiment involving three participants. An enhanced version of the framework will also be introduced, which accounts for the sequential nature of the FH 3D US data and incorporates the 3D position measurements acquired by the position sensor as contextual information. Additionally, advanced techniques such as data augmentation and multi-model embedding will be employed to further refine and improve the framework's performance.

CHAPTER 4

AUTOMATIC 3D LAMINA CURVE EXTRACTION FROM FREEHAND 3D ULTRASOUND DATA USING SEQUENTIAL LOCALIZATION RECURRENT CONVOLUTIONAL NETWORKS

4.1 Introduction

In the last chapter, we established the 2D labeling criteria pertaining to the visible lamina bone surface, and described a machine learning method for automating paired laminae landmark extraction from transverse ultrasound images. Nevertheless, that method based on 2D image analysis using CNNs is spatially constrained, as it fails to consider contextual relationships in the superior-inferior direction and exploit the full potential of 3D sequences obtained from FH 3D US. Given the complex nature of the spine, the pursuit of smooth and sequential 3D landmark curve extraction from ultrasound sequences holds the potential for achieving greater accuracy in identifying lamina structure than 2D landmark extraction, while simultaneously providing smoother lamina landmark curves. Recurrent convolutional networks (RCNs) can be a good solution, which leverages CNNs for image processing and the strengths of Recurrent neural networks (RNNs) (e.g., Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Bidirectional LSTM (BiLSTM), Bidirectional GRU (BiGRU)) for sequence analysis, as successfully applied in diverse domains (Donahue *et al.*, 2017; Poudel, Lamata & Montana, 2017; Jafari *et al.*, 2018; Bai *et al.*, 2018; Abdi *et al.*, 2017; Huang, Bridge, Noble & Zisserman, 2017; Xu, Csapo, Feng & Mi, 2020).

In this chapter, we extend our prior work to accommodate the sequential nature of FH 3D US data, allowing implicit integration of spine shape constraints into lamina extraction. Thereupon, we introduce Sequential Localization Recurrent Convolutional Networks (SL-RCN), a novel framework tailored for 3D lamina landmark curve extraction from tracked 3D ultrasound sequences. The Normalized Discrete Fréchet Distance (NDFD) is employed as the main metric to evaluate the disparity of the extracted 3D lamina curves. An evaluation through K-Fold cross-validation on integral ultrasound sequences covering thoracic and lumbar regions

demonstrates the superior performance of SL-RCN in accurately identifying lamina structures and extracting 3D lamina curves.

Section 4.2 elaborates on the 3D lamina curve extraction method. Section 4.3 thoroughly presents the evaluation results of both manual landmark labeling and 3D curve extraction. Section 4.4 discusses the AI model structure, contribution to spine ultrasound imaging, and potential impact on other areas. And Section 4.5 summarizes this study and outlines future work.

4.2 Methods

4.2.1 SL-RCN

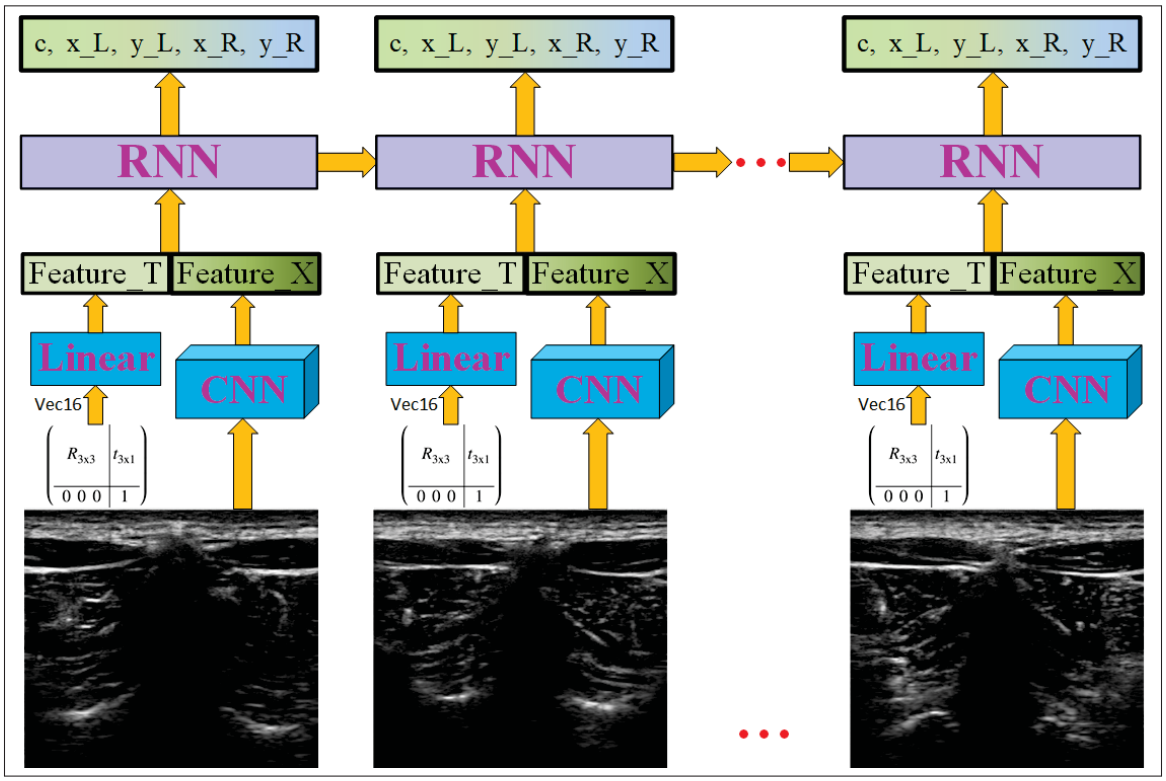

Figure 4.1 SL-RCN for 3D lamina curve extraction from ultrasound

As shown in Figure 4.1, both the 2D ultrasound image and its corresponding transformation matrix are the learning material for SL-RCN, and a tuple $\hat{Y} = (c, x_L, y_L, x_R, y_R)$ is predicted for each frame from the vertical output of an RNN under a many-to-many architecture. Variable c is the confidence of laminae structure detection with 0.5 being the threshold value. x and y denote the column and row lamina landmark coordinates, and their subscripts L and R refer to the left and right lamina, respectively. The input feature vector of the RNN consists of two parts, the image feature X extracted by the CNN from a 2D ultrasound image and the transformation matrix feature T augmented from a flattened 4×4 matrix by a linear multilayer perceptron (MLP) layer. Prior to training the AI agent, normalization is applied to both the input images and the manually labeled ground truth Y.

Based on the findings of our preliminary study in Chapter 3, it has been established that the BCE criterion yields more accurate and reliable predictions than the CE and MSE criteria for the classification task. Consequently, we have chosen to employ the BCE criterion for identifying the laminae structures in our SL-RCN. This choice results in a reduction of one variable in the output tuple of SL-RCN compared to the output tuple of the CNN-based framework (Figure 3.1).

In terms of the composition of the objective function, SL-RCN and the CNN-based framework exhibit similar structures. They both employ a weighted combination of the classification loss and the landmark coordinate losses. Furthermore, the calculation method for the landmark coordinate losses remains unchanged. This implies that equations 3.1 and 3.2 continue to be applicable to the SL-RCN. Focusing on the BCE criterion, the mini-batch classification loss $\mathfrak{L}_c$ can be represented by

$$\mathfrak{L}_c \;=\; -\frac{1}{K}\sum_{m=1}^{K} c_Y^m \log\Big[Sigmoid(c_{\hat{Y}}^m)\Big] \;+\; (1 - c_Y^m)\log\Big[1 - Sigmoid(c_{\hat{Y}}^m)\Big] \tag{4.1}$$

where K is the total ultrasound frames contained in a mini-batch. The objective of the SL-RCN is to fit the neural network for a pair of 3D curves, which is achieved by minimizing a weighted sum of $\mathfrak{L}_c$ and $\mathfrak{L}_l$ sequentially using transfer learning (CNN pre-trained with ImageNet data (Huh, Agrawal & Efros, 2016)) and mini-batch stochastic gradient descent.

4.2.2 Source 3D ultrasound sequences

Table 4.1 Description of manually labeled FH 3D US data:
PTCP is short for "participant",
data with IDs 1∼7 for k-fold cross-validation,
8∼10 for testing

PTCP ID	1	2	3	4	5
Total Frames p	503	641	383	612	351
Positive Ratio r	70.97%	50.39%	74.93%	66.18%	64.39%
PTCP ID	6	7	8	9	10
Total Frames p	533	444	262	336	253
Positive Ratio r	69.23%	85.81%	56.11%	51.49%	46.24%

Ten unprocessed 3D ultrasound sequences, each belonging to one participant, were acquired as source data for the subsequent experiments. During data acquisition, a field of view parameter of 6cm was used as the general depth value for ultrasound scanning, resulting in each gray-scale image having a uniform size of (478, 303) pixels. Ultrasound images featuring discernible horizontal high-luminance laminae bone surfaces were labeled as positive frames, whereas the remaining images were categorized as negative frames. The lamina landmarks were manually labeled following the protocol described in Section 2.3.1 and were subsequently validated by two spinal ultrasound experts. Sequences belonging to participants 8-10 designated as the test set, underwent sparse labeling, involving the omission of every other frame. The total number of labeled frames p in each participant's sequence is documented in Table 4.1. Row "Positive Ratio r" in Table 4.1 provides the number of frames labeled as positive, describing the quality of each participant's 3D ultrasound sequences.

4.2.3 Data processing

In this section, we describe the creation of multiple Data Groups (DGs) following the leave-one-participant-out method for K-Fold Cross Validation using data augmentation techniques, with the aim of conducting ablation experiments to validate the effectiveness of the proposed SL-RCN as well as comparing with our previous CNN framework (Li *et al.*, 2023). Prior to data augmentation and DG creation, it is noteworthy that the original images were initially

down-sampled by a factor of four, from (478, 303) to (119, 75), to alleviate computational resource demands. Section 4.2.3.1 outlines the data augmentation approach and DGing used for the CNN framework evaluation, and subsequently, Section 4.2.3.2 elaborates on the process of sampling video sub-sequences from augmented data for training and validating the SL-RCN.

4.2.3.1 Data augmentation

The addition of noise to images enhances the resilience of machine-learning models by reducing their sensitivity to minor input fluctuations. Gaussian noise introduces random fluctuations that mimic real-world noise in images. It has the potential to disrupt spurious patterns or correlations that machine learning models can erroneously latch onto, thereby increasing their robustness to ultrasound speckle noise. Furthermore, Gaussian noise complements various augmentation techniques, such as rotation, flipping, cropping, etc. In this study, a monovariate, zero-mean Gaussian noise with identical distribution was applied to the pixel intensities, with a standard deviation of 0.04.

Image occlusion and cropping are effective techniques for data augmentation, preventing overfitting and enhancing generalization in AI agents. While occlusion disrupts canonical features by randomly blocking image portions (e.g., Cutout (Devries & Taylor, 2017) and Random Erase (Zhong, Zheng, Kang, Li & Yang, 2017)), cropping enables selective focus on specific RoIs. This study proposes Conservative Adaptive Erasure (CAE) for transverse spinal ultrasound imaging, a novel data augmentation method combining the benefits of image occlusion and cropping. The aim is to guide the AI agent's attention towards horizontal bone surface structures surrounding the spinous process, including the left lamina, top surface of the spinous process, right lamina, and potential anterior complex beneath the spinous process acoustic shadow. By preserving the features of these high-luminance structures, the analysis primarily relies on the bone surface's wave response while minimizing predicted landmarks in the wave response of soft tissues or even acoustic shadows. In CAE, negative frames were subjected to random cropping of a rectangular region measuring 40% of the image's width and 40% of its height, specifically in the bottom region, either on the left or right corner, with

the cropped region being masked to zero. Correspondingly, positive frames were subjected to lateral cropping on either the left or right side of the image, whereby a crop of 40% width was taken from either the left or right side of the image, spanning from the image's side edge to 40% of the distance between the side edge and the column address of the side lamina landmark. The selection of the lateral side to be cropped is determined by the direction opposite to that of the middle point between the paired laminae landmarks' lateral tendencies. The proposed CAE approach offers a comprehensive means to accurately analyze and identify vertebral bone surface characteristics.

After data augmentation, for each participant, there are three image sequences (as shown in Figure 4.2) that share identical lamina landmark and transformation sequences, which include an original image sequence, a Gaussian-noised image sequence, and a cropped Gaussian-noised image sequence. The three image sequences with shared lamina landmarks form the X and Y of the DG utilized for assessing the efficacy of the CNN framework.

4.2.3.2 Video sampling

RNNs are specifically designed to operate on sequential data where the current output is influenced by both the previous inputs and the current input. As such, the input of RNNs must always be a sequence, as this enables the model to effectively capture and analyze the intricate temporal dependencies and interrelationships that exist between the individual data points. For the purpose of disambiguating between the 3D ultrasound sequences and the input sequences

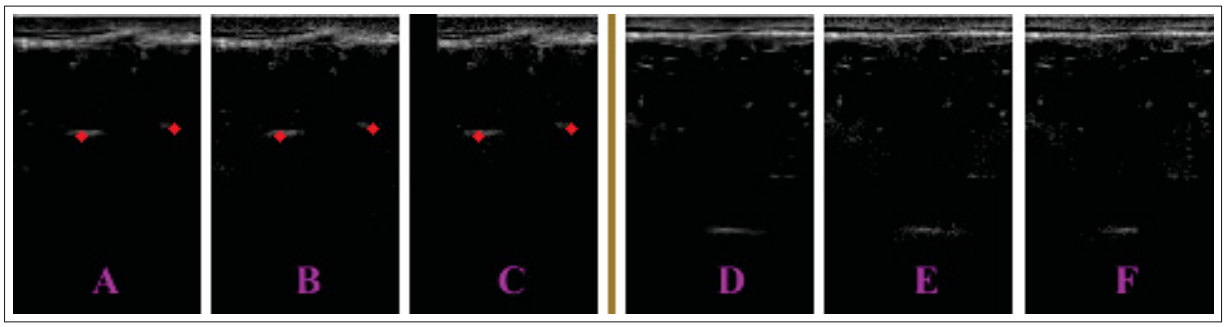

Figure 4.2 Augmented images with shared lamina landmarks

utilized by RNNs, we shall henceforth designate **an input sequence** for RNNs as **a video** and its sequence length as **video frame length** q.

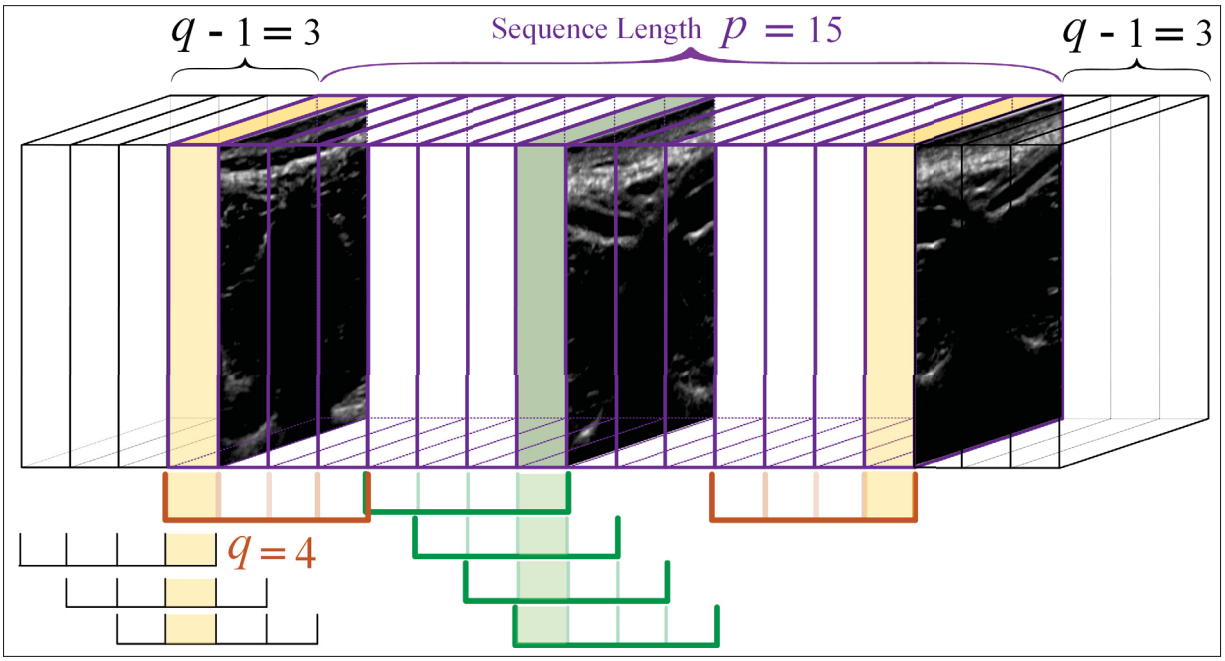

Figure 4.3 Example of balanced video sampling

Given the limited number (merely six original sequences, with one sequence reserved for validation purposes) of sequences available for RNN agent training, it is imperative to facilitate the meticulous learning of temporal dependencies from all possible sub-sequences (a sub-sequence is a subset of q consecutive frames sampled from its source sequence). To this end, we allow video overlapping during the sampling process for training set and slide the sampling window one frame at a time. This video-sampling approach, however, poses the challenge of an imbalanced training set. As shown in Figure 4.3 , the center frames (in green skin), whose distance from the sequence's edge is greater than the video frame length q, have been sampled q times through q videos. Conversely, the outer frames (in yellow skin) closer to the edge with a distance d less than q frames from the edge can only be sampled d times, leading to an unequal distribution. To address the issue of the imbalanced training set, we supplement $q-1$ blank (blacked-out) frames to the head and tail of the ultrasound sequences before the training set video sampling. We then generate an additional sequence of masks, which helps to identify

valid frames and blank frames for both overlapped training videos and non-overlapped validation videos.

In summary, for each participant with an original 3D ultrasound sequence length of p, by video-sampling from the three augmented 3D ultrasound sequences, we obtain $3 \times (p + q - 1)$ overlapping training videos and $3 \times \text{ceil}(p/q)$ non-overlapping validation videos. This sampling process is employed on four distinct sequences, namely the image sequence X, the lamina landmark sequence Y, the transformation sequence T, and the mask sequence M.

4.2.4 Evaluation method

The experiment is designed to evaluate both the manual sequential landmark labeling and the automated 3D lamina curve extraction approach utilizing SL-RCN, both qualitatively and quantitatively. The evaluation of the sequential lamina landmark labeling protocol proposed in Section 2.3.1 was carried out by two observers. The more experienced observer performed the sequential landmark labeling for all seven participants, and twice on a single participant for the intra-observer variability analysis at an interval of 5 months between the two labeling sessions. The ground truth was established using this observer's labeling results. The other observer performed the labeling for three participants for the evaluation of inter-observer variability.

A parameter grid search of 3D curve extraction utilizing the novel SL-RCN approach is firstly conducted based on a 7-fold cross-validation study implementing the leave-one-participant-out strategy, using the CNN framework from Chapter 3 as the baseline. The evaluation of the results pertaining to the identification of lamina and the extraction of landmark coordinates involves three intertwined metrics: F1 score, physical distance between the ground-truth and predicted landmarks, and NDFD, as detailed in the following. Consequently, using the four metrics, an analysis of the intra- and inter-observer variabilities of manual sequential landmark labeling is carried out, and an ablation study is conducted to elucidate the individual contributions of each component of SL-RCN towards the 3D curve extraction process.

In order to quantitatively evaluate the effectiveness of our proposed SL-RCN in extracting 3D lamina curves, a metric is required to measure the disparity or similarity between the predicted 3D curve and the manually labeled one. The Fréchet Distance metric gauges the similarity of two curves by identifying the longest shared path between them, thereby accounting for differences in both shape and parameterization of the curves. Given two polygonal curves A: [1, a] $\rightarrow R^n$ and B: [1, b] $\rightarrow R^n$, where n is a positive integer, a coupling C between A and B is a sequence (Wien, Eiter, Eiter, Mannila & Mannila, 1994)

$$C = (A(u_1), B(v_1)),\ (A(u_2), B(v_2)),\ \ldots,\ (A(u_m), B(v_m)).$$

Thus, the coupling respects the order of the points in A and B. The discrete Fréchet Distance is the minimum, over all possible couplings C, of the maximum distance between any coupled points from A and B in C; it can be expressed as:

$$\mathcal{F}(A, B) \ = \ \min_{\forall C}\Big\{ \max_{i=[1,\, m]} || \ A(u_i) - \ B(v_i) \ || \ \Big\}. \tag{4.2}$$

Nevertheless, this metric does not consider the variations in the starting points, orientations, and scales of the curves, which can lead to difficulties in comparing curves that have similarities but different starting points or orientations. By applying Procrustes analysis to align the curves before calculating Fréchet Distance, the effect of differences in rotation, scaling, and translation between the curves is minimized, thus improving the accuracy of 3D curve comparison. At the same time, Procrustes analysis necessitates that two curves undergoing alignment possess an equal number of landmarks, and the landmarks must be mutually correspondent across the curves. Through screening the true-positive landmarks and forming a true-positive curve from both the ground truth and predicted landmark sequences, a **normalized** Fréchet Distance can be calculated by comparing two Procrustes-aligned true-positive curves that have the same curve length.

In this study, we solely apply rotation and translation for the curve alignment while the two true-positive curves share the same scaling, and employ **normalized discrete** Fréchet Distance

(NDFD) as a metric to evaluate the 3D curve disparity. All the NDFD results presented in this study have been multiplied by a factor of 100 to reduce the number of fractional digits.

4.3 Results

4.3.1 3D lamina curve extraction

Building on the foundation of our prior research, we maintain the utilization of ResNet34 as the CNN model in both our established CNN framework and the image feature extractor within the novel SL-RCN architecture. In our primary testing, a two-layer RNN steadily generated convergent output and outperformed a 1-layer RNN on the task of 3D curve extraction. Consequently, we selected 2-layer RNNs for the subsequent evaluation experiments, with $w = 30.0$. These weights were chosen in accordance with the grid search results from Section 3.3.

Table 4.2 Grid Search 7-fold cross validation results The first column on the left denotes the experiment number: The number of the augmented transformation features #T is 512, "MDE" indicates the mean distance error of true-positive frames in millimeters, "L/R" denotes "Left/Right lamina"

#	RNN	q	F1 score	MDE_L	MDE_R	NDFD_L	NDFD_R
1	**BiLSTM**	**64**	0.841±0.060	**1.41±0.85**	**1.40±0.91**	**0.428±0.393**	**0.457±0.399**
2	BiLSTM	128	0.835±0.056	1.43±0.85	1.45±0.92	0.454±0.411	0.460±0.377
3	BiLSTM	256	0.828±0.047	1.41±0.89	1.47±0.93	0.513±0.499	0.526±0.531
4	LSTM	64	0.841±0.047	1.42±0.90	1.47±0.92	0.466±0.466	0.512±0.479
5	GRU	64	**0.848±0.049**	1.41±0.91	1.44±0.93	0.526±0.586	0.537±0.545
6	BiGRU	64	0.837±0.070	1.48±0.90	1.45±0.87	0.501±0.535	0.465±0.389
Previous Study:			0.814±0.087	1.62±1.32	1.63±1.14	0.591±0.449	0.639±0.402

Table 4.2 specifies grid search cross validation results on 3D lamina curve extraction. Experiment (exp.) #1-3 grid searched the video frame length q and exp. 4-6 tested various RNN models. As highlighted in row exp. #1, a BiLSTM model with $q = 64$ generates the best evaluation result on the 3D lamima curve extraction. A Wilcoxon signed-rank test was performed to compare the left/right NDFD values between the previous 2D image analysis method and the newly proposed

SL-RCN (exp. #1), yielding p-values of 0.0469 and 0.0156, respectively. The results indicate a statistically significant difference between the two methods at significance level 0.05.

Table 4.3 Deep learning (DL) model evaluation on 3D lamina curve extraction: The training set is from PTCPs ID 1∼7, and test set 8∼10; SL-RCN employs parameters from exp. #1.

DL Method	F1 score	MDE_L	MDE_R	NDFD_L	NDFD_R
Previous	0.745±0.059	1.66±1.43	1.93±1.44	0.638±0.425	0.661±0.441
SL-RCN	0.748±0.043	1.57±0.97	1.75±1.12	0.454±0.205	0.508±0.298

We acknowledge the limitation of our study, which involves a relatively small number of participants. Therefore, after the 7-fold cross-validation, we conducted a test experiment. Table 4.3 presents the test-set evaluation of the two methods, trained on 7 participant sequences and tested on 3 sparse (skipped every other frame) participant sequences. From the training set to the test set, as shown in Table 4.1, the average participant Positive Ratio r has decreased from 2348 / 3467 = 67.72% to 437 / 851 = 51.35%. This decrease indicates a reduction in image quality and explains the drop in the F1 score from the validation results in Table 4.2 to the test-set evaluation results in Table 4.3. However, focusing on the extracted left/right lamina curves, for both AI frameworks of our previous study and SL-RCN, the increase in the mean NDFD value is less than 0.05. The minimal increase of the NDFD proves the effectiveness of model training with data acquired from limited participants, and the results of our test-set evaluation in Table 4.3 effectively demonstrate SL-RCN's capability in extracting paired smooth lamina landmark curves.

Figure 4.4 demonstrates the comparison of 3D lamina curves, intercepting and zooming in on the sub-sequence of lumbar vertebrae L3 and L4, extracted using the previous 2D image analysis framework, the newly proposed SL-RCN, and the manual labeling method, respectively. The interweaving of three distinct colored curves affirms the effectiveness of artificial intelligence in automated extraction of 3D lamina curves. Using the green curve as the reference track, the yellow curve (SL-RCN) exhibits superior adherence to the reference track compared to the

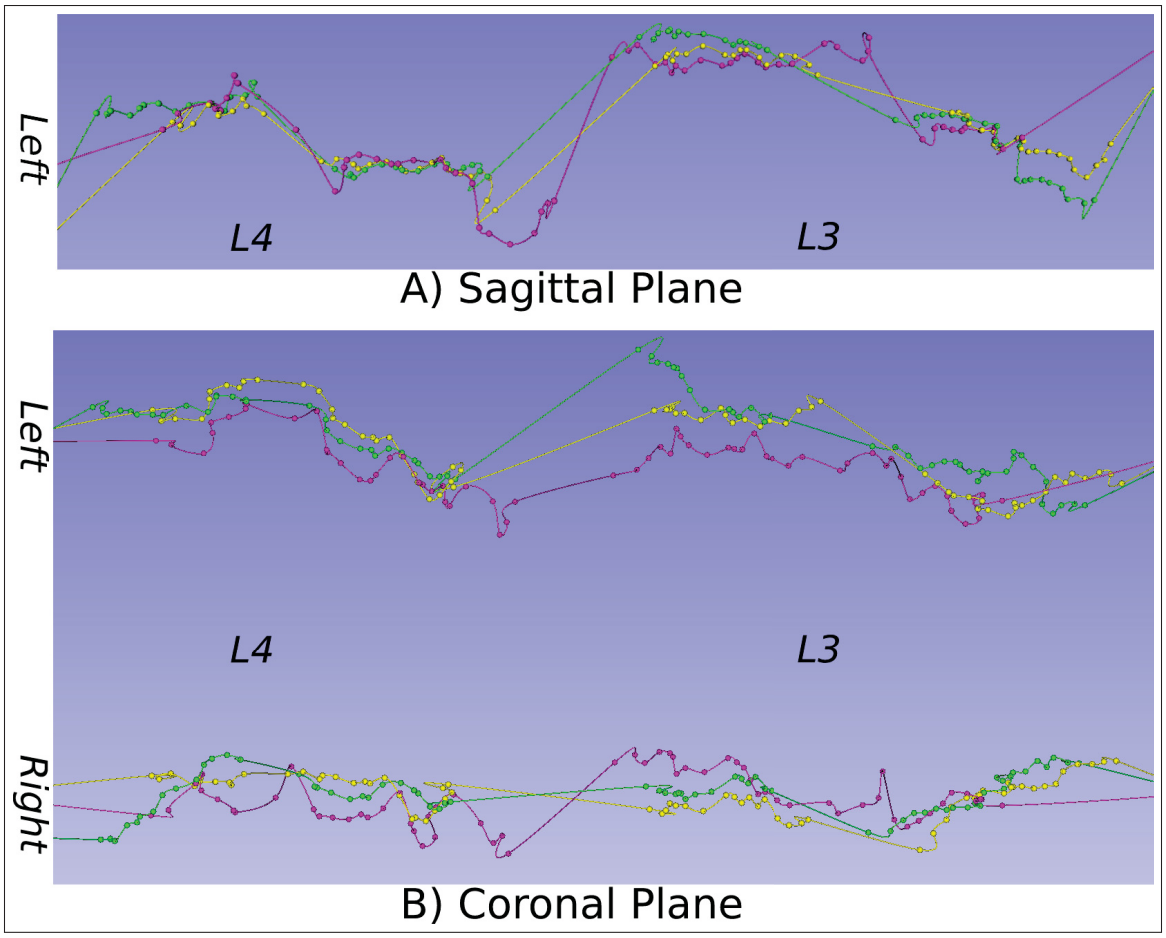

Figure 4.4 3D lamina curve extraction comparison

purple curve (2D image analysis), which exhibits more pronounced deviations from the reference track.

4.3.2 Intra- and inter-observer variability

Table 4.4 Evaluation of the proposed sequential lamina landmark labeling protocol

	PTCP ID	F1 score	MDE_L	MDE_R	NDFD_L	NDFD_R
Intra-	7	0.972	0.65±0.34	0.67±0.33	0.227	0.207
Inter-	3, 5, 7	0.885±0.097	1.34±1.12	1.30±1.14	0.356±0.155	0.560±0.328

As shown in Table 4.4, the intra-observer variability demonstrates a high F1 score, along with MDE less than 1mm. The inter-observer variability indicates a relatively lower F1 score, accompanied by an MDE of approximately 1.3mm on average, and the absence of a statistically significant distinction is supported by a Student's t-test with p-values of 0.56. We hypothesize that the lower F1 score can be attributed to a conservative decision by one of the raters while assessing low-contrast images. By comparing the NDFD values with the 3D curve extraction results presented in Table 4.2 it was evident that both CNN and SL-RCN proved their capability on 3D lamina curve extraction, with SL-RCN exhibiting a higher level of similarity to a human observer.

4.3.3 Ablation study

The SL-RCN architecture is composed of a RNN, a CNN, and a MLP feature augmenter for T, and its final output, as predicted by the RNN, is determined by two separate data streams, specifically the transformation features T and the image features X. Due to the dominance of the image features X in the lamina structure identification and landmark localization, which is extracted from the rich image feature map, out of the two features X and T, only the linear MLP feature augmenter and features T were ablated in this section. This study evaluates the effectiveness of the 4x4 transformation matrix in enhancing the 3D lamina curve extraction from 2D image sequence processing to 3D sequence analysis. Table 4.5 presents the results of ablation experiments conducted with a video sequence length $q = 64$. The exp. number is continued from Table 4.2 , and exp. #1 is carried over as the golden baseline for the ablation contrast.

Table 4.5 SL-RCN ablation 7-fold cross validation

#	Model	#T	F1 score	MDE_L	MDE_R	NDFD_L	NDFD_R
1	SL-RCN	512	0.841±0.060	1.41±0.85	1.40±0.91	0.424±0.389	0.451±0.388
7	↓ #T	256	0.850±0.049	1.45±0.83	1.45±0.90	0.496±0.522	0.490±0.399
8	× Linear	16	0.844±0.070	1.52±0.89	1.57±0.92	0.408±0.342	0.537±0.436
9	× T	0	0.836±0.059	1.48±0.87	1.48±0.92	0.449±0.396	0.546±0.483

Linear MLP. Exp. #7 and #8 aimed to evaluate the effectiveness of the component feature augmenter for T by decreasing the number of features T and completely removing feature augmentation, respectively. The F1 score is relatively unaffected; however, both the MDE and curve disparity demonstrate a discernible upward trend. We hypothesize that the feature augmenter can enhance the performance of SL-RCN in 3D lamina landmark localization by effectively smoothing the extracted 3D lamina curve.

Transformation Matrix. Exp. #9 is designed to assess the effectiveness of an image-only RCN in extracting 3D lamina curves. The curve disparity is greater than that of exp. #8, yet it still represents an advancement over the results achieved by our previous CNN framework. We hypothesize that this outcome implies a preferable pair of smooth lamina curves moving within a range akin to the one depicted in Figure 2.5. Exp. #7-9 together show the importance of the extended transformation features T in accurately extracting a pair of smooth 3D lamina curves from 3D ultrasound sequences.

4.4 Discussion

According to the evaluation results, the SL-RCN model exhibited a notable reduction in true-positive MDE and NDFD in comparison to the conventional CNN framework, despite a minor fluctuation in the F1 score. It is hypothesized that the improvement of NDFD was a result of taking into account contextual ultrasound frames, including images and 3D information from the transformation matrix; concomitantly, the smoother lamina landmark curves generated by the RNN reduce both the MDE and the standard deviation.

The integration of the transformation matrix feature T enhances the process of extracting 3D lamina curves, aligning with the anticipated outcome. Additionally, the inclusion of the feature augmenter facilitates the interpretation of spatial features derived from the transformation matrix, thereby augmenting the overall performance of the SL-RCN model.

In line with the accumulated expertise in the field (Cahuantzi, Chen & Güttel, 2021), LSTM model exhibits a marginal superiority over GRU in terms of overall performance; in the meantime,

the bidirectional RNN demonstrated a notable decrease in NDFD. The length of the video frames q determines the field of view for the SL-RCN agent to comprehend the spine morphology along the superior-inferior axis. Nevertheless, increasing the value of the parameter q leads to an increment in the number of overlapped frames at a quadratic rate, thereby quadratically increasing the CPU memory cost, posing a more challenging task (e.g., $q > 256$) of overfitting in the training process of SL-RCN. In this study, the test set is limited to three participants. Future work should involve testing the method on a larger data set, and eventually on a data set including scoliotic patients, to evaluate its generalizability.

4.5 Conclusion

In spinal 3D ultrasound sequence analysis, a pair of 3D lamina curves possess high qualifications for representing the spine shape, owing to its close proximity to the spinal cord, capacity to balance the lateral vision center in the transverse plane, and the extensive superior-inferior space spanning each vertebra. In this chapter, we developed SL-RCN, a deep learning model specifically tailored for 3D ultrasound sequence analysis, to automatically extract a pair of smooth 3D lamina curves. This novel deep learning framework underwent evaluation experiments on FH 3D US data acquired from ten healthy adult human participants, encompassing both lumbar and thoracic regions. We assess the performance of the AI models through 7-fold cross-validation using a leave-one-participant-out strategy. Our test-set experiment provides evidence of the absence of overfitting. The results demonstrate that our newly proposed SL-RCN outperforms the previous CNN framework comprehensively, and the ablation study provides further evidence of the significance of the transformation matrix feature T in achieving accurate extraction of smooth 3D lamina curves.

CHAPTER 5

VERTEBRAL SURFACE EXTRACTION FROM FREEHAND 3D ULTRASOUND USING SAM ZERO-SHOT

5.1 Introduction

In the preceding chapters, we identified a pair of 3D lamina landmark curves as an innovative representation of spine shape, established a standardized protocol for sequential manual lamina landmark labeling, and developed deep learning frameworks to extract these landmarks from ultrasound sequences. Building on this foundation, this chapter aims to further our research by developing an automated method for vertebral surface extraction. This method will utilize the laminae landmarks extracted by our previously developed deep learning models, taking one step forward toward automated 3D reconstruction of the spine. Such advancements are crucial for enhancing spine visualization and improving deformity analysis through the application of FH 3D US.

The accurate extraction of the 3D bone surface from FH 3D US image sequences is crucial for a comprehensive interpretation of spinal structures, geometric analysis of spine deformity, and for guiding ultrasound-assisted spine interventions (Gueziri, Drouin, Yan & Collins, 2019b; Gueziri, Santaguida & Collins, 2020). However, the vertical bone surface adjacent to the spinous process typically remains invisible in ultrasound scans. The automated extraction of the spinous process surface poses a significant challenge in spinal ultrasound imaging due to the inherently poor contrasts and unfavorable bone geometry present in these images. This challenge is rooted in the physics of ultrasound image formation. Bone surfaces that are perpendicular to the ultrasound wave direction tend to generate the most pronounced response, manifesting as a conspicuous high-luminance (white) horizontal curve within the ultrasound images. While ultrasound scans effectively capture distinct horizontal edges, they tend to obscure vertical edges, thus impeding the accurate visualization of spinous process features.

Despite these challenges, a precise 3D spine model is essential for comprehending spinal structure, with the spinous process providing key 3D positional information and facilitating

vertebral rotation analysis, particularly in the context of ultrasound-guided spine interventions. Accurate online 3D spinal surface reconstruction from ultrasound sequences could facilitate interpretation of vertebral structure in ultrasound images, streamline sonographer training, and ultimately optimize the quality of acquired ultrasound sequences.

The utilization of deep learning models for medical image segmentation has shown promising results in recent years, particularly with the introduction of vision transformers (ViTs) (Dosovitskiy *et al.*, 2020). The Segment Anything Model (SAM) (Kirillov *et al.*, 2023), a state-of-the-art ViT, has demonstrated impressive capabilities in segmenting various objects within 2D medical images. SAM employs a massive training dataset and leverages the ViT neural architecture to achieve accurate segmentations. It incorporates the use of prompts, such as foreground/background points and bounding boxes, to guide the segmentation process and accommodate the ambiguity in the target region's size or shape. SAM's prompt-based functionalities offer versatile options for precise and controlled segmentation in medical imaging applications (Mattjie *et al.*, 2023; Shi *et al.*, 2023).

The present work aims to develop an automated method to extract a smooth vertebral bone surface, including the spinous process, from FH 3D US data. In this chapter, we investigate the application of the SAM | specifically SAM Zero-Shot, which has shown potential for other medical imaging applications | for segmenting vertebral surfaces in ultrasound images obtained from a spine phantom and human subjects. We compare the performance of SAM's automated mask generation (denoted as "SAM Auto") and prompt-based segmentation methods. Additionally, we propose a novel vertebral bone surface extraction method, **L**andmark prompted **S**AM - **T**rio **S**urface **E**xtraction (LS-TSE), specifically designed for transverse ultrasound images. The LS-TSE method combines the advantages of AI-generated laminae landmarks for precise targeting of the left lamina, right lamina, and spinous process (indicated by the letter "L"), pre-trained SAM for controlled segmentation of ultrasound response from left/right laminae surface using prompts, (indicated by the letter "S"), and asymmetrical peak model for spinous process bone surface approximation with a trio fitting design based on the extracted curves of the left and right lamina bone surfaces (indicated by the letter "T").

5.2 Materials and methods

To evaluate the potential of SAM Zero-Shot, both with and without prompts, our experiments employ FH 3D US data acquired from a spine phantom (Blue Phantom, Kirkland, WA) and three ultrasound sequences from participants #1, #3, and #5 from Table 4.1. In total, 508 pairs of lamina landmarks were identified and localized in these images by manual labeling, and 531 by the CNN-based framework from Chapter 3. This section elaborates our proposed LS-TSE method to extract 2D curves of the vertebral bone surface from transverse ultrasound frames, which is designed to handle low-quality ultrasound data. The efficacy of LS-TSE in generating acoustic shadow (area below its extracted vertebral bone surface) masks is assessed against manually labeled masks on frames where laminae are visible, using the Intersection over Union (IoU) metric. Subsequently, the accuracy of vertebral (B-Spline) surface extraction using LS-TSE is evaluated using the Mean Surface Distance (MSD) metric.

The **T**rio **S**urface **E**xtraction (TSE) determines the transverse vertebral surface by joining the left lamina, the spinous process, and the right lamina. The detection of the left/right lamina surfaces is detailed in Section 5.2.1, and that of the spinous process surface in Section 5.2.2.

5.2.1 Lamina surface detection

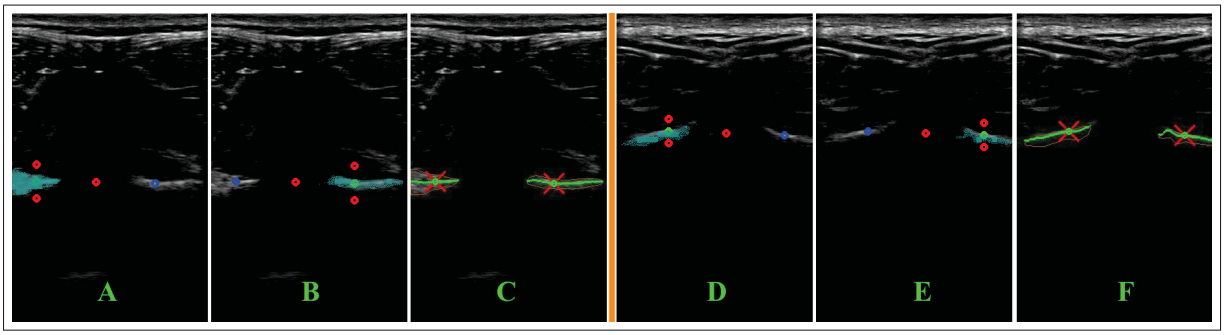

Figure 5.1 Laminae bone surface extraction using Lamina landmark prompted SAM Zero-Shot

The lamina landmarks extracted from the ultrasound images are used as point prompts to SAM Zero-Shot towards the extraction of left and right lamina segmentation masks, mask-L and

mask-R, respectively illustrated in Figure 5.1 A/D and Figure 5.1 B/E. To generate each mask, one foreground prompt point (the corresponding lamina landmark) and three background prompt points are employed, including the midpoint between two laminae landmarks and two points located slightly above and below each lamina landmark, at 0.9 and 1.1 times their depths with respect to the skin surface, respectively. The left/right lamina bone surface curve is obtained by tracing the curve that is vertically centered within its respective lamina mask, followed by a vertical translation of this curve to align it with the lamina landmark, thereby ensuring its accuracy.

5.2.2 Spinous process surface estimation

To approximate the outline of the spinous process in ultrasound transverse images, three essential elements need to be determined. These include a mathematical model expression that represents the shape of the spinous process, a suitable number of points for model fitting, and a bounding box that defines the range of the spinous process. The establishment of the left/right and bottom borders of the spinous process is achieved through the extraction of the laminae bone surface using Lamina landmark prompted SAM Zero-Shot. In this section, we will determine the top border of the spinous process by determining the vertical coordinate of the Spinous Process Landmark (SPL). Subsequently, the vertical intensity projection values will be employed to fit an Asymmetrical Peak Model (APM) for approximating the outline of the spinous process.

We propose to locate the vertical coordinate of SPL by investigating the H-IPC from an RoI determined by our identified pairwise laminae landmarks. As illustrated in Figure 5.2, the H-IPC is conceptualized as a two-stage connection, with the SPL serves as the watchman of the connection point between these two stages: stage one represents a consistent downward slope of the hillside encompassing the soft tissues prior to reaching the spinous process, while stage two corresponds to a horizontal expansion of the plain pertaining to the spinous process layer leading up to the laminae. The SPL is situated at the periphery of a low cliff (stop 4) that borders the plain, and descending from the cliff leads directly to the plain.

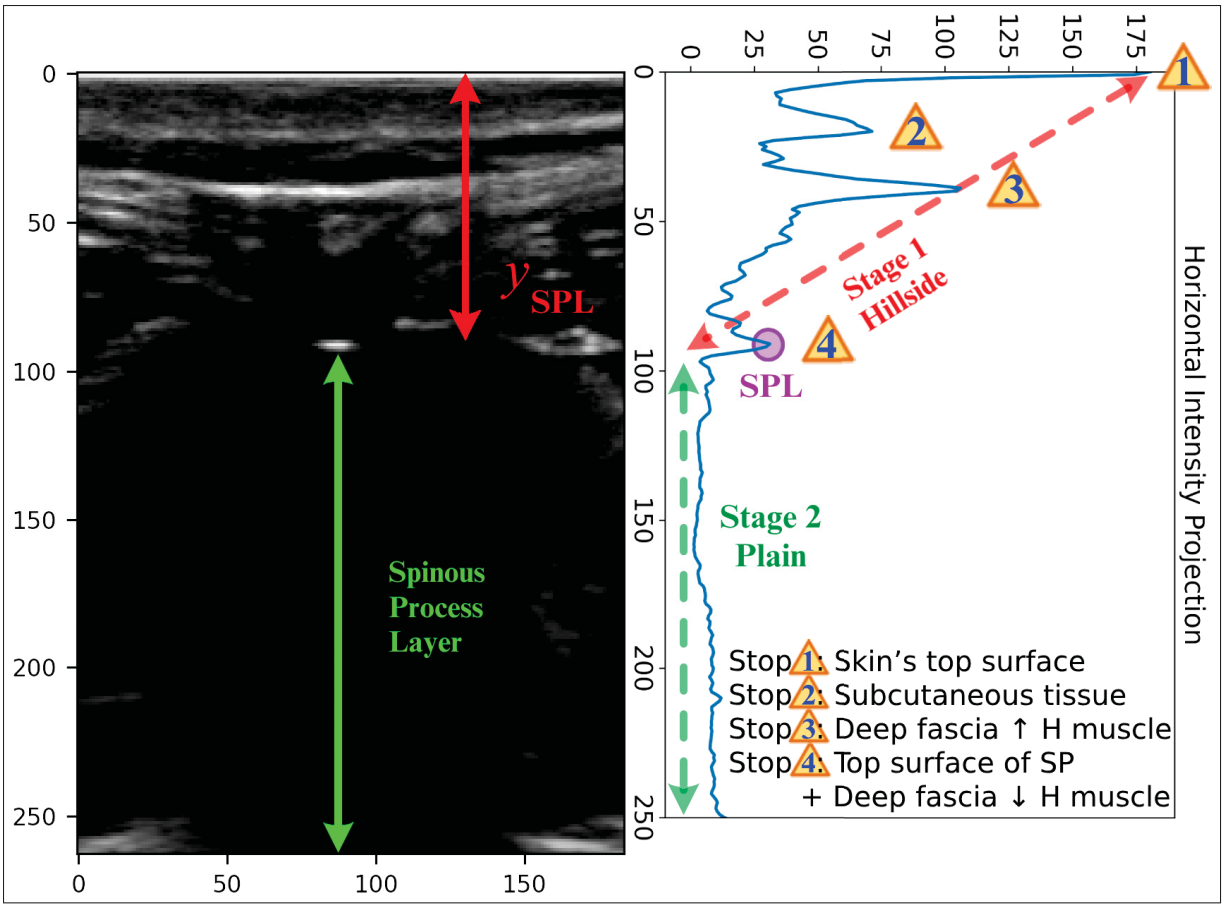

Figure 5.2 Interpretation of H-IPC in RoI

The vertical coordinate of the SPL can be determined by locating the connection point between the two stages, analogous to locating the fulcrum of a lopsided "seesaw." In order to localize the fulcrum of the "seesaw," we propose to reorient the H-IPC curve, balance the "seesaw" so that neither end touches the ground, and then fit it into a convex-shaped curve with the fulcrum toward the bottom of this curve. We apply Principal Component Analysis (PCA) to adjust the lopsided "seesaw" shaped curve by aligning the first principal component α with the main axis of variance, which corresponds to the longest dimension of the "seesaw." This alignment reorients the data, adjusting the curve's orientation so that it opens upward, thus elevating both ends of the seesaw. To simplify the convex curve fitting and the criteria to localize the SPL point, we perform variable standardization by removing the mean and scaling to unit variance

after applying PCA. Subsequently, the standardized PCA points are employed for the fitting of an APM, which is mathematically expressed as:

$$f(x; A, \mu, \sigma, \alpha) = A \cdot exp\Big[-\frac{1}{2}\big(\frac{x-\mu}{\sigma}\big)^2\Big] \cdot \big(1 + \alpha \cdot \frac{x-\mu}{\sigma}\big) \tag{5.1}$$

where A is the amplitude of the peak, μ is the center of the peak, σ is the standard deviation controlling the width of the peak, and α is the asymmetry parameter, which determines how much the peak is shifted asymmetrically. The parameters of the APM are determined using the Levenberg-Marquardt algorithm.

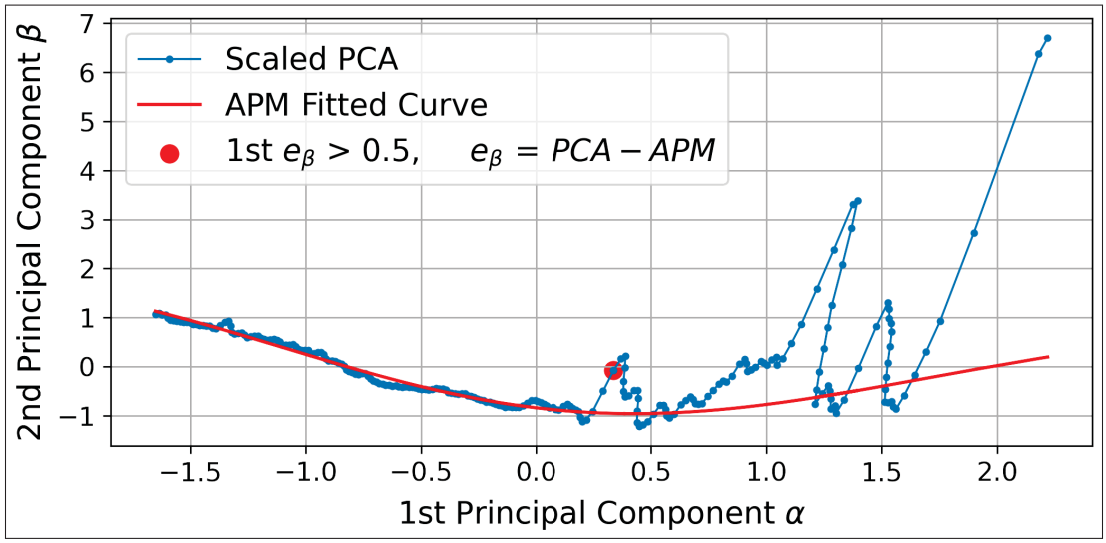

Figure 5.3 Scaled PCA of the H-IPC from Figure 5.2

As shown in Figure 5.3, the standardized principal components of the H-IPC depict a slightly lifted “seesaw” shape. The APM fitted curve aligns closely with the "plain stage," allowing us to determine the SPL point as the first point whose second principal component β exceeds the fitted APM point by more than 0.5, an empirical threshold in the PCA space that is typically smaller than the height difference from the four stops to their corresponding fitted APM points. Consequently, the vertical coordinate of the SPL y_{SPL} is ascertained from the original H-IPC by correlating it with the index of the SPL point identified from the PCA transformed data, as demonstrated in Figure 5.4 panel B.

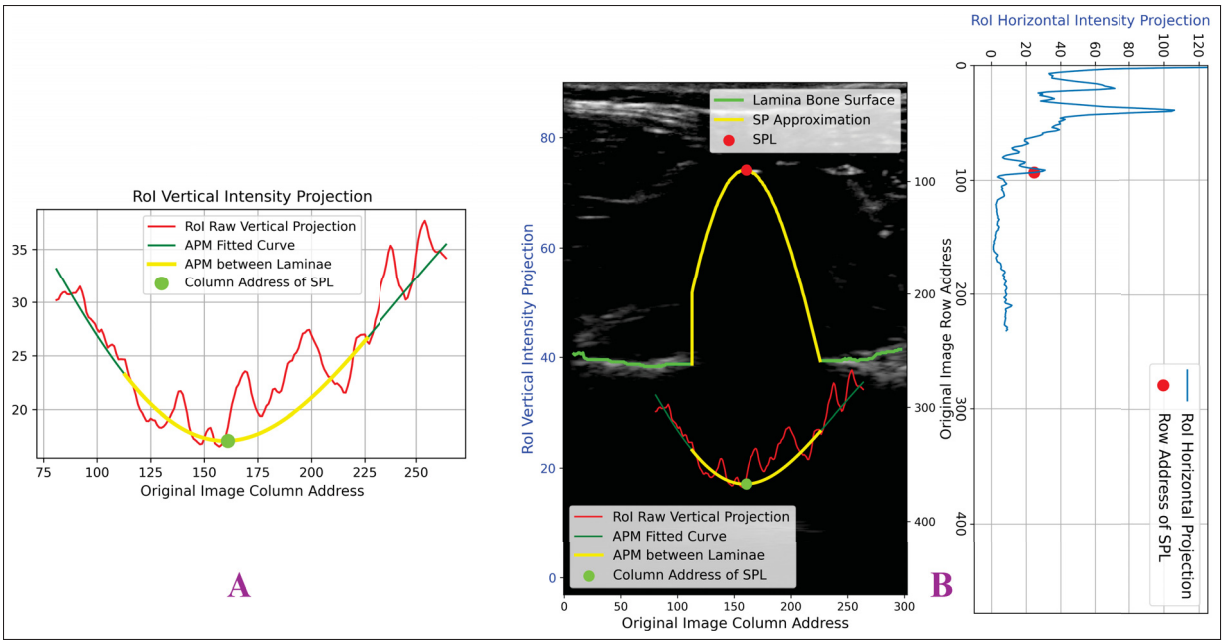

Figure 5.4 Spinous process surface estimation

The APM is again employed to model the shape of the spinous process. As a flexible yet simple skewed Gaussian curve, it can account for the complex anatomy of the spinous process caused by variations in ultrasound beam orientation. The V-IPC contains the acoustic shadow information outlining the spinous process. As depicted in Figure 5.4 panel A, an APM is fitted to the V-IPC to approximate the outline of the spinous process, leveraging the shape information of the spinous process embedded within the V-IPC. With the laminae surfaces identified by the lamina landmark prompted SAM Zero-Shot, and the vertical coordinate of the SPL established from the H-IPC, the height of the spinous process can be ascertained. As illustrated in Figure 5.4 panel B, the outline of the spinous process is approximated by a vertically translated, flipped, and stretched APM curve. Its potentially shorter "foot" is connected to its corresponding lamina bone surface with a vertical line.

5.3 Results

Figure 5.5 presents the segmentation results of SAM Zero-Shot on spinal ultrasound images. The "SAM Auto" method perfectly delineated the acoustic shadow in phantom ultrasound as

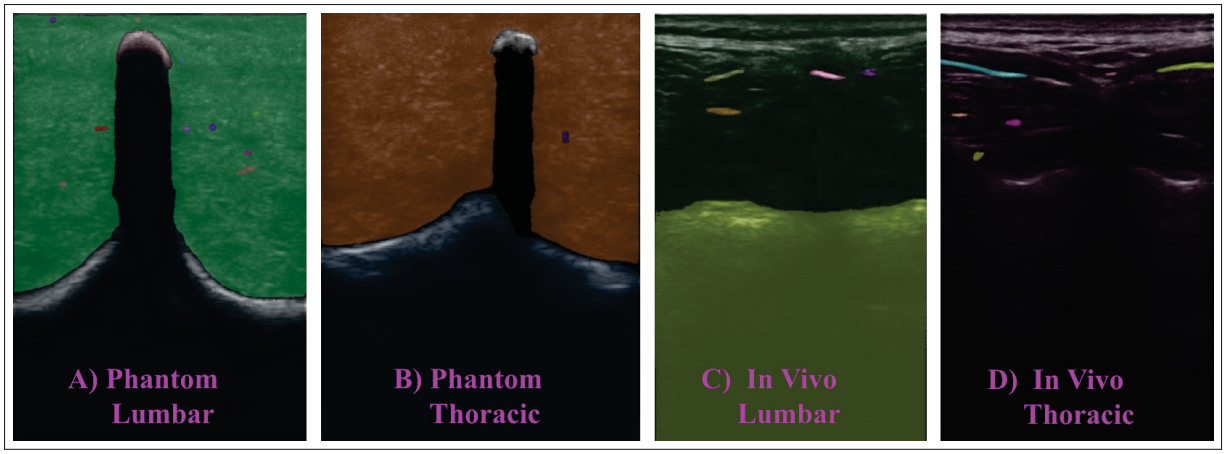

Figure 5.5 SAM Zero-Shot with no prompts in spinal ultrasound object segmentation

shown in panels A and B. However, it identified the acoustic shadow area beneath the laminae as an object in the human lumbar scan (panel C), and recognized no vertebral structure but only edges in the human thoracic scan (panel D). A more advanced method than "SAM Auto" is necessary for the task of vertebral surface extraction.

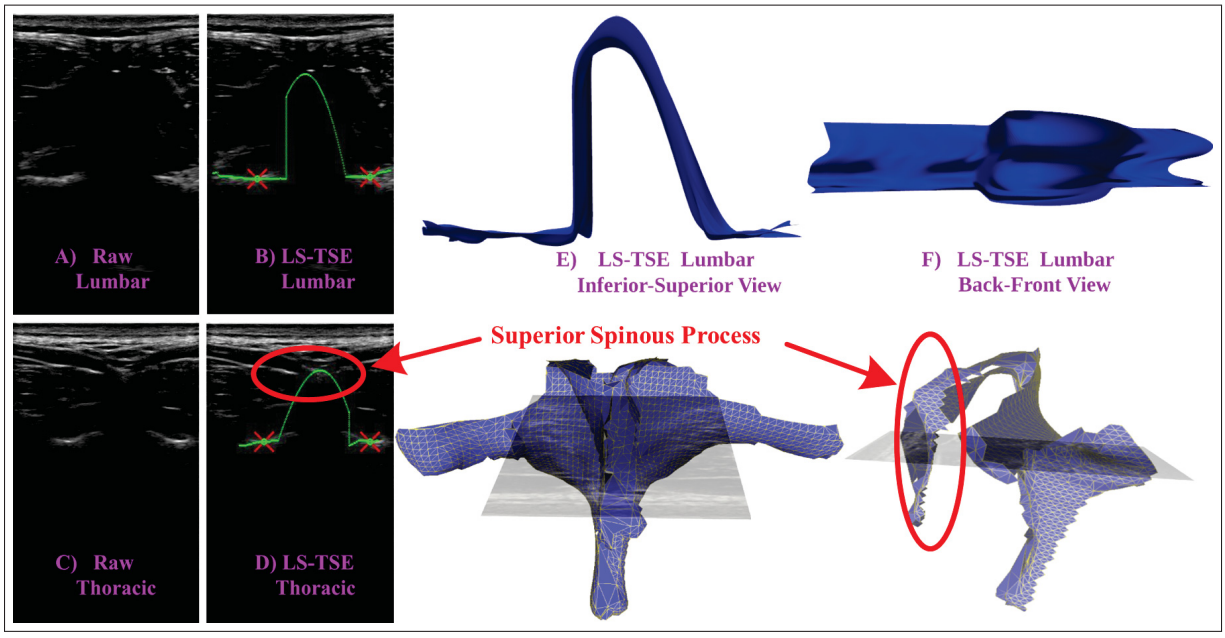

Figure 5.6 Vertebral surface extraction from human transverse spinal ultrasound using LS-TSE

Table 5.1 LS-TSE acoustic mask IoU against manual labeling

Method	PTCP #1	PTCP #3	PTCP #5
M-L	0.93	0.94	0.93
DL-L	0.94	0.94	0.92

Figure 5.6 panels A/B and C/D demonstrate bone surface extraction from individual lumbar and thoracic vertebra images, respectively. Panel B illustrates the ability of LS-TSE to extract the lumbar vertebral surface. Nevertheless, panel D shows a result where the detected SPL does not belong to the same vertebra level as the lamina bone surface. Instead, the SPL extracted in panel D actually belongs to the superior vertebral level. Thoracic spinal anatomy complexifies the task due to the physical overlap between consecutive vertebrae, and a surface model derived from these landmarks would not correctly represent the underlying anatomy. Consequently, in this preliminary study, our subsequent evaluation of the LS-TSE method focuses on the lumbar region.

Table 5.1 presents the assessment of LS-TSE in segmenting acoustic shadow masks under the lumbar vertebral bone surface. The IoU results suggest a high accuracy regardless of the method used to generate the lamina landmark prompts. This high tolerance for positional accuracy of the lamina landmarks demonstrates the strengths of LS-TSE, which combines AI-generated laminae landmarks for precise targeting of the laminae surfaces with the pre-trained SAM for controlled segmentation of the ultrasound bone surface response using prompts.

Figure 5.6 panels E and F show the 3D visualization of an extracted L5 vertebral surface obtained from participant #1. Figure 5.7 presents the point cloud comparison between the LS-TSE and manually extracted lumbar vertebral bone surfaces. The five lumbar vertebrae B-Spline meshes extracted by LS-TSE from participant #1 achieve an MSD error of 0.69 ± 0.15 mm, which would be adequate for image-guided intervention. The detailed surface distance values are reflected by the heatmap visualization in Figure 5.8.

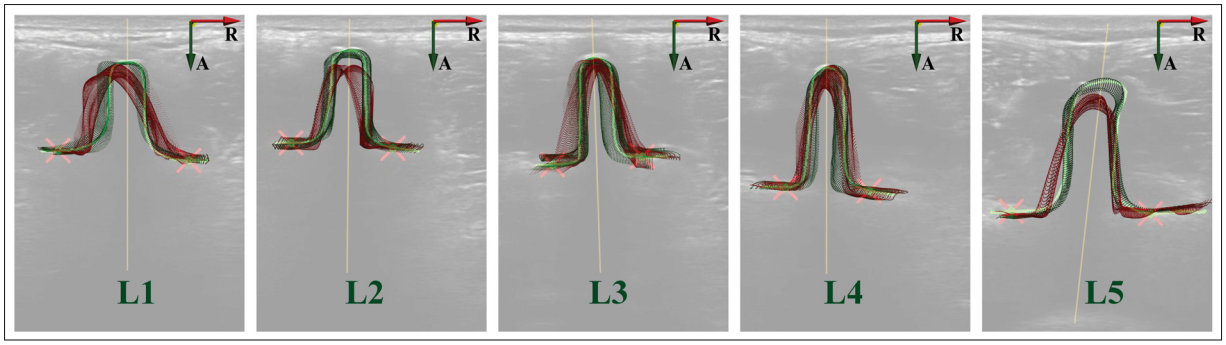

Figure 5.7 B-Spline 3D point cloud comparison between LS-TSE (red) and manually (green) extracted lumbar vertebral bone surfaces from participant #1, matched on the transparent background of Figure 2.11 row II

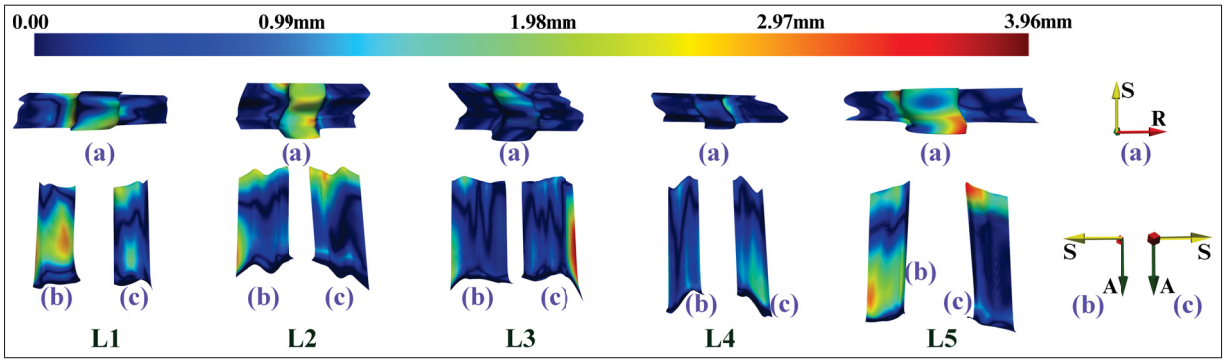

Figure 5.8 Heatmap Visualization of Surface Distance Error in mm

5.4 Discussion

The SAM Zero-Shot segmentation without prompts, while effective for phantom images, exhibits limited performance on human spine images. In contrast, our proposed LS-TSE extracts accurate surfaces for lumbar vertebrae in vivo. The primary surface distance error occurs either due to the thin head of the APM curve or because the visible lamina bone surface is considerably shorter than in reality. As shown in Figure 5.9, the two purple segments indicate surface distance errors exceeding 3mm, with the left one due to the shorter visible lamina and the right one due to the thin top of the spinous process. The cases of imbalanced visible laminae typically occur when the probe orientation deviates from the transverse plane in the corresponding vertebral space. This deviation is accompanied by non-zero transverse rotation ϕ_z or coronal rotation ϕ_x,

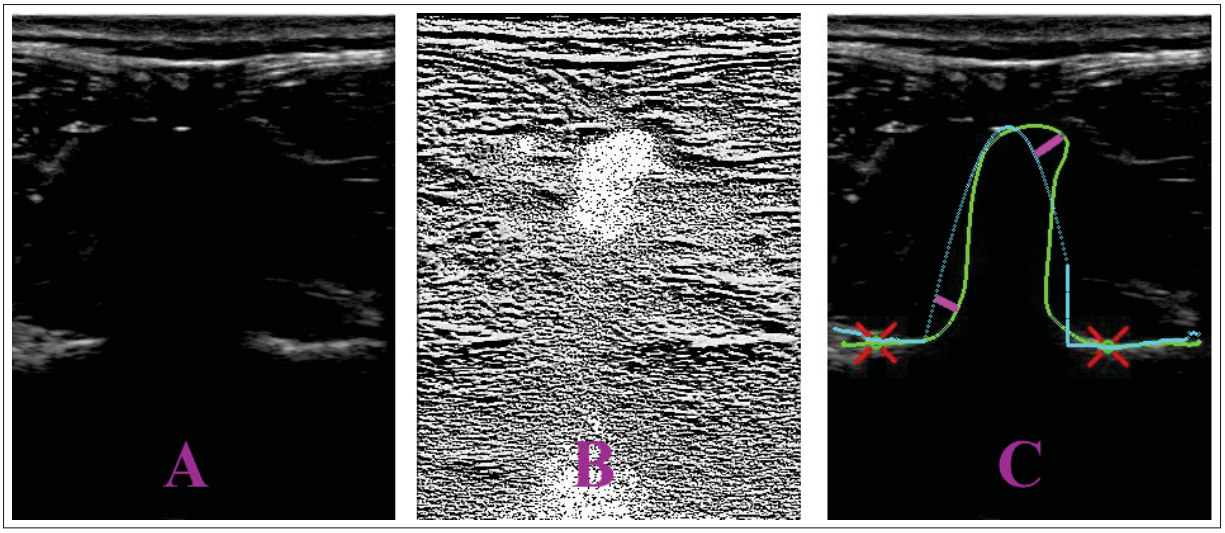

Figure 5.9 An Example Image in which Visible Left Lamina is Shorter Than in Reality

as illustrated in Figure 5.10. Greater accuracy could be achieved by maintaining the spinous process centered straight within the ultrasound image with balanced visible laminae during data acquisition.

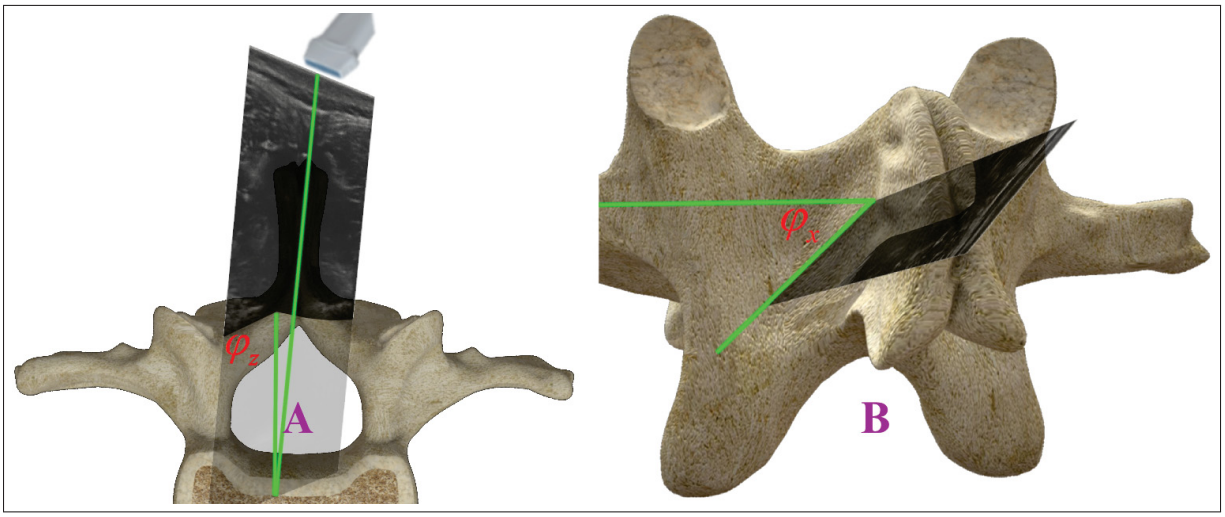

Figure 5.10 A Scanning Cases where Visible Lamina is Shorter Than in Reality

5.5 Conclusions

In this chapter, we explored the application of SAM Zero-Shot on the detection of vertebral surfaces from FH 3D US data. A novel method for tissue-bone separation utilizing landmark-guided SAM is proposed to achieve accurate 3D bone surface reconstruction from low-quality ultrasound image sequences. The resulting bone surface, represented as a 3D mesh, effectively reproduces a smooth outline of the vertebra, encompassing the left lamina, spinous process, and right lamina. To evaluate the performance of the proposed method, we used three FH 3D US sequences from three participants. The acoustic shadow masks beneath the extracted bone surface were evaluated against manually labeled masks, achieving a mean IoU above 0.92. Our evaluation on one human subject's data reveals a MSD of the extracted lumbar vertebral surfaces of 0.69±0.15 mm.

Lumbar vertebral surface detection offers greater value and reliability compared to limited anatomical landmark detection for X-ray to ultrasound registration in preoperative and intraoperative surgical procedures, and our LS-TSE method can enhance the accuracy of this registration process. Fully rendered 3D spine models derived from the LS-TSE generated 3D point clouds can be compared over time to track changes in spinal structure, monitor the effectiveness of treatment, and assess the progression of degenerative conditions. The technique can additionally foster advancements in medical pedagogy and augmented reality assistance guidance surgery. Future work will address SPL localization for thoracic vertebrae, and validation with data from patients with spinal deformities.

CHAPTER 6

GENERAL DISCUSSION

The laminae landmarks play a critical role in depicting the 3D shape of the spine. Sitting near the center of ultrasound transverse scans and being symmetrically positioned around the spinous process shadow, the paired laminae landmarks warrant further investigation and emphasis in research pertaining to characterization of spine shape utilizing FH 3D US transverse scanning. A novel representation of spinal shape using paired 3D lamina curves can fuse the advantages of both spinous process landmarks in the analysis of scoliotic deformities and transverse process landmarks in the assessment of axial vertebral rotation while mitigating deviation by virtue of its longer distance in the superior-inferior direction as compared to transverse process landmarks. A pair of 3D laminae curves spanning both the lumbar and thoracic regions can significantly enhance the VPI-based scoliosis deformity analysis by automating the acquisition of non-planar reslices, in addition to enabling accurate 3D reconstruction and registration of the spine column.

In the current literature, there is a lack of algorithms for recognizing laminae landmarks that have been validated on low-quality 2D ultrasound or interlaminar scans. In addition, the task of extracting single-point landmarks on the laminae remains poorly defined due to the absence of formal guidelines identifying the most pertinent part of the visible lamina bone surface in ultrasound images. In Chapter 2, we proposed two manual labeling protocols applied to transverse spinal ultrasound images. One protocol provides formal guidelines for labeling the most representative pairwise point landmarks representing the laminae, specifically focusing on the most relevant part of the visible lamina bone surface in ultrasound images. A systematic 3D sequential labeling approach is designed towards a pair of smooth 3D lamina curves that can represent the vertebrae in a comprehensive manner. The other protocol involves labeling the vertebral bone surface, encompassing the left lamina, spinous process, and the right lamina. In Chapter 3, a CNN-based deep learning framework is proposed to simultaneously identify the presence of visible laminae and extract pairwise lamina landmarks.

However, this CNN-based framework may not be ideal for generating smooth 3D lamina landmark curves essential for 3D spinal deformity analysis. This framework is primarily designed for analyzing individual 2D images, and therefore lacks the capability to infer 3D spatial relationships between these images. As a result, the sequentially predicted lamina landmarks are independent across frames and do not form naturally smooth curves in 3D space. Additionally, the MDE metric, which evaluates individual 2D ultrasound images, is inadequate for assessing the network's performance in terms of the smoothness of the extracted 3D lamina curves. To reliably extract smooth 3D lamina curves from ultrasound sequences, an advanced AI framework is required, along with a new performance metric specifically designed to evaluate the network's effectiveness in extracting smooth 3D lamina curves.

In Chapter 4, we introduced SL-RCN, an enhanced framework tailored to the sequential nature of FH 3D US data. This framework embeds 3D position measurements from a magnetic position sensor and accommodates contextual frames in its sequential inference. Moreover, advanced techniques such as data augmentation and transfer learning were applied to further refine and enhance the framework's performance. The Fréchet Distance considers the location and ordering of points along two curves by finding the minimum leash distance to walk along them continuously. We adopted the Normalized Discrete Fréchet Distance (NDFD) as the primary metric to evaluate the disparity of the extracted 3D lamina curves. We conducted an extensive evaluation of both the CNN-based model and the SL-RCN, involving cross-validation with seven participants and a test experiment with three participants. The outcomes clearly indicate that SL-RCN surpasses the CNN model in performance, with the ablation study further highlighting the critical role of the transformation matrix feature T in precisely extracting smooth 3D lamina curves.

This thesis addresses a significant topic in spinal imaging, which has the potential to have a substantial impact. Concretely, the extracted lamina landmarks can be projected onto the sagittal plane to aid in vertebral level identification (Hetherington, 2017; Brotherston, Desy & Ma, 2021) and kyphotic angle measurement (Sayed *et al.*, 2022) using ultrasound imaging. ANNEX I presents a preliminary sketch of a novel vertebral level identification method for every single

frame in 3D ultrasound sequences by analyzing sequential lamina landmarks, which does not require a known initial vertebral level. The paired 3D lamina curves provide coronal plane spinal curvature information, offering the potential to replace the VPI method (Cheung *et al.*, 2015; Wong *et al.*, 2019; Lee *et al.*, 2021) for scoliosis assessment and enable vertebral rotation analysis (Vo *et al.*, 2015; Chen *et al.*, 2016). ANNEX II outlines a possible scoliosis assessment method utilizing the Projected Laminae Vectors (PLVs), which leverages the extracted 3D lamina curves. The symmetrically distributed vertebral point clouds of lamina landmarks, centered around the spinous process, potentially support 3D reconstruction of the spinal column (Nguyen *et al.*, 2015) and 3D image registration (Effatparvar, Pierre & Sobczak, 2022) by combining FH 3D US with an articulated shape model (Boisvert, Cheriet, Pennec, Labelle & Ayache, 2008a). Furthermore, the extracted vertebral landmarks from ultrasound images have applications in epidural anesthesia (Ghosh *et al.*, 2016; Sadeghi, Patel & Carvalho, 2021b), image-to-image registration (Che *et al.*, 2017), and spinal intra-operative guidance (Vasudeva *et al.*, 2017a; Gueziri *et al.*, 2019b, 2020).

In Chapter 5, we explored the application of SAM Zero-Shot for detecting vertebral surfaces from FH 3D US data. We introduced LS-TSE, a prompt-based vertebral bone surface extraction method specifically tailored for low-quality transverse ultrasound images, and compared its performance with SAM's automated mask generation. While SAM Zero-Shot segmentation without prompts is effective for phantom images, it shows limited performance on human spine images. It is hypothesized that the uniform characteristics of isotropic biological tissue help maintain a distinct boundary between the tissue and the acoustic shadow beneath the bone structure, thus facilitating an optimal separation between tissue and bone. However, this scenario is highly unlikely to occur in vivo. In contrast, LS-TSE accurately extracts surfaces for lumbar vertebrae in vivo, benefiting from precise targeting of the laminae surfaces using landmark prompts and addressing the issue of invisible vertical edges in ultrasound through the design of trio surface extraction.

However, in images of thoracic vertebrae, the detected spinous process landmark often corresponds to a superior vertebral level rather than aligning with the lamina bone surface at the

same level. This issue arises due to the physical overlap between consecutive thoracic vertebrae, complicating the detection of the spinous process landmark. Theoretically, the LS-TSE method could also be applied to thoracic vertebrae, provided the spinous process landmark is correctly aligned with the lamina bone surface at the corresponding vertebral level.

When comparing the vertebral bone surfaces extracted using LS-TSE with manually labeled surfaces, significant surface distance errors occur when the visible lamina bone surface is much shorter than in reality. These discrepancies typically arise when the probe orientation deviates from the transverse plane in the corresponding vertebral space, resulting in non-zero transverse rotation ϕ_z or coronal rotation ϕ_x. This deviation causes the side of the spinous process facing the transducer to obstruct the emitted ultrasound signals. Greater accuracy can be achieved by ensuring the spinous process is centered and aligned within the ultrasound image, with balanced visible laminae during data acquisition.

CONCLUSION AND RECOMMENDATIONS

The purpose of this research was to develop and validate innovative methods for capturing and analyzing the intricate vertebral anatomy using FH 3D US imaging. Our study specifically focused on in vivo transverse spinal ultrasound sequences acquired from healthy adults. We successfully achieved four key objectives throughout this work.

The first objective was to identify key vertebral features and establish ground truth. These vertebral features are intended to provide an appropriate representation of spine shape suitable for further 3D analyses, such as spine deformity analysis, vertebral rotation measurement, vertebra reconstruction, and visualization. A pair of 3D lamina curves possess high qualifications for representing the spine shape, owing to their close proximity to the spinal cord, capacity to balance the lateral vision center in the transverse plane, and the extensive superior-inferior space spanning each vertebra. To establish the true lamina landmark curves, we initiated the ethical board protocol application, recruited participants, collected data, then developed software, and conducted manual labeling under standardized labeling protocols.

The second objective centered on developing and validating an AI-based method to extract essential features necessary for representing the spine shape from individual images. We acquired FH 3D US data under various ultrasound depth settings and conducted a grid search to identify a feasible configuration for a CNN-based framework. This framework is designed to automatically and efficiently extract pairwise laminae landmarks from transverse spinal ultrasound images.

The third objective aimed to enhance the learning of the 3D spine shape, improving the AI agent's ability to accurately extract spine representation features sequentially. We developed the SL-RCN model to reveal the 3D spine shape from FH 3D US data using the extracted smooth lamina curves. Its capability of sequential inference is enhanced through the integration of positional matrices and the accommodation of contextual frames.

The fourth objective was to develop a vertebral surface extraction method for visualizing vertebrae from 3D ultrasound sequences. We have proposed and validated the LS-TSE method for lumbar vertebral bone surface extraction, and implemented the 3D visualization by generating B-Spline meshes. Through systematic and methodical progression, we successfully met each of these objectives, contributing significantly to the field of spinal health imaging.

Contributions

Contribution 1: We recruited 38 healthy adult participants (16 males, 22 females) aged 18-30 years. We developed standardized manual labeling protocols for both pairwise laminae landmarks and vertebral bone surfaces from transverse spinal ultrasound images. For the purpose of lamina landmark extraction, sixteen ultrasound sequences were manually labeled. Of these, six sequences from three participants were utilized to assess the adaptability of the proposed CNN-based framework and its depth adaptability. The remaining ten sequences from ten participants, featuring a fixed generalized ultrasound depth value, were labeled to evaluate the performance of the proposed SL-RCN. Additionally, three ultrasound sequences from three different participants were labeled for the extraction of vertebral bone surfaces and the subsequent visualization of their 3D meshes built on the lamina curve representation.

Contribution 2: We developed a CNN-based framework designed to automatically and efficiently extract pairwise laminae landmarks from transverse spinal ultrasound images. Our experimental results demonstrated the superiority of BCE, over CE and MSE, as a robust loss function in binary classification tasks. We identified an optimal configuration for the proposed framework and established a promising approach for automated lamina detection and landmark localization, leveraging the foundational ResNet34 architecture.

Contribution 3: We enhanced the CNN-based framework by introducing SL-RCN, an innovative framework specifically designed for 3D lamina landmark curve extraction from tracked 3D

ultrasound sequences. This novel framework effectively accommodates the sequential aspect of FH 3D US data and integrates 3D positional information as contextual data. Our experimental results indicate that SL-RCN significantly outperforms the original CNN-based framework. Additionally, an ablation study highlights the critical role of the transformation matrix feature T in facilitating the precise extraction of smooth 3D lamina curves.

Contribution 4: We proposed the LS-TSE method to automatically extract lumbar vertebral bone surfaces from sub-optimal in vivo ultrasound sequences. The fully rendered 3D meshes, derived from 3D point clouds generated by the LS-TSE method, enable tracking changes in spinal structure, monitoring treatment efficacy, and assessing the progression of degenerative conditions. Additionally, this technique has the potential to enhance medical education and support augmented reality-assisted surgical procedures.

Recommendations

This thesis investigates the modeling and visualization of vertebral anatomy using FH 3D US imaging. Our experiments indicate that both the CNN-based framework and SL-RCN are capable of detecting 3D lamina landmarks. The CNN-based framework is proposed for individual image processing, which could be used in real-time feature detection, while the SL-RCN is designed for sequential data analysis, providing more accurate results for offline 3D spine shape analysis.

In this section, we first discuss potential directions for future investigation. Then, we provide general guidelines for data acquisition of spinal ultrasound images using FH 3D US. The following are several potential future research directions:

- **Detection of Single Lamina Landmarks from Transverse Ultrasound Images**: There are instances where only one side of the laminae is visible. A more powerful and robust AI agent is needed to identify a single lamina structure and localize its corresponding landmark.

- **Open-Source Software for Real-Time Lamina Landmark Detection**: The CNN-based framework could be integrated into a user interface for live ultrasound scanning, facilitating the interpretation of spinal ultrasound imaging and simplifying the training of sonographers.
- **Comprehensive Evaluation of SL-RCN**: This thesis assessed SL-RCN using ultrasound sequences from only ten healthy participants. A more comprehensive evaluation, involving a larger number of participants, including patients with spine deformities, is anticipated.
- **Per-Frame Vertebral Level Identification in Ultrasound Sequences**: ANNEX I presents a proof of concept for a novel per-frame vertebral level identification method, which leverages the extracted sequential lamina landmarks. This concept warrants further validation.
- **Scoliosis Assessment Using Projected Laminae Vectors from FH 3D US Data**: ANNEX II provides a proof of concept for a novel scoliosis assessment method, leveraging the extracted sequential lamina landmarks. This proof of concept deserves validation. In preliminary tests, a Sawbone's spine model could be used to simulate various spine deformities, and an RGBD camera could be employed to collect a 3D point cloud from each subject to generate ground truth data.
- **Evaluation of LS-TSE for Lumbar Vertebral Surface Extraction**: This thesis assessed LS-TSE using ultrasound sequences from only three healthy participants. A more extensive evaluation including patients with spine deformities is anticipated for statistical validation.

Based on our experimental findings, it is evident that the quality of the collected images significantly impacts feature extraction and the visualization of vertebral structures. To ensure high-quality data acquisition of spinal FH 3D US imaging, we propose the following guidelines:

1. Symmetry of W-Shape: Each 2D scan should display a symmetric W-shape. When the spinous process is visible, the left and right laminae (and transverse processes, if visible) should be evenly balanced with respect to the spinous process.
2. Centering the Spinous Process: In most 2D scans, the spinous process should be centered within the transducer's field of view to ensure maximum integrity of the W-shape. In

specific cases, such as the lumbar transverse interspinous view where the spinous process is not visible, the anterior complex should be centered instead.

3. Uniform Distribution of Frames: The frames within a sequence should be uniformly distributed along the probe's trajectory over the spine. This uniformity allows for 3D curves to be accurately fitted through the landmarks (such as the spinous process and left/right laminae) in 3D, thereby preventing overfitting.
4. Adequate Number of Scans per Vertebral Level: There should be a sufficient number of scans for each vertebral level to ensure that enough landmark points are captured to represent the basic outline of the vertebra.
5. Avoiding Intersection Between Frames: It is recommended to avoid any intersection between frames in a sequence to maintain clarity and integrity of the imaging data.

These guidelines are designed to optimize the quality of ultrasound images for effective diagnosis and monitoring of spine patients.

ANNEX I

VERTEBRAL LEVEL IDENTIFICATION

In the superior-inferior direction, the connectivity of laminar structures between different vertebrae is established through the articular processes. When ultrasound frames scan the area where two adjacent vertebrae overlap, they can be attributed to both the superior and inferior vertebral levels. To address this ambiguity, we propose a categorization scheme involving lamina regions (LRs) (each belonging to a distinct vertebral level) and inter-lamina regions (ILRs).

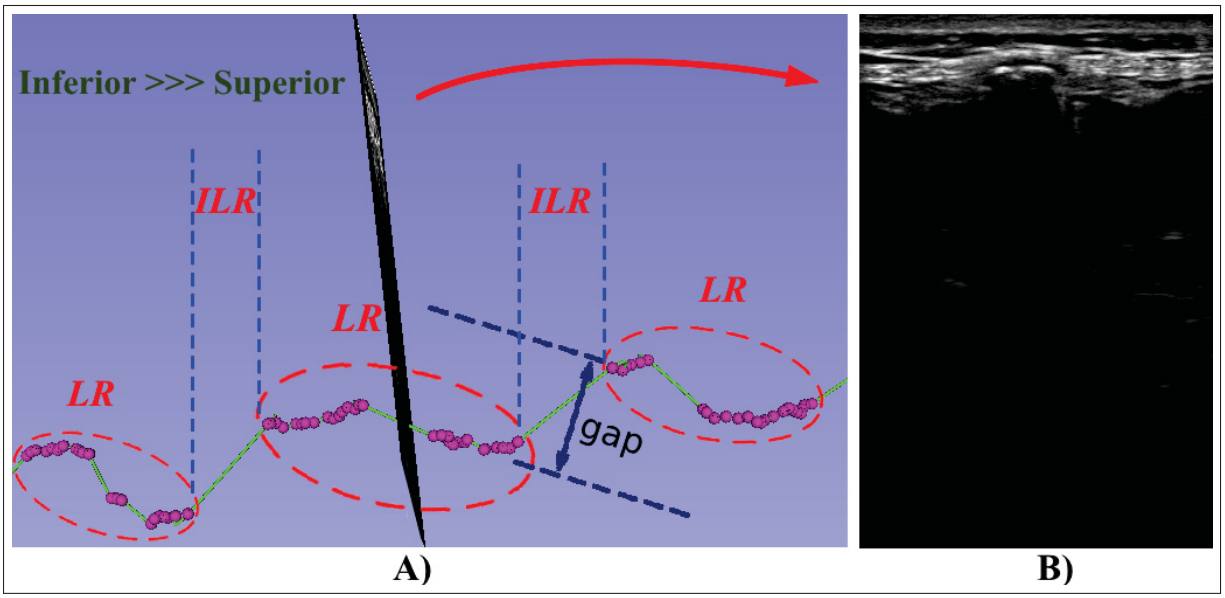

Figure-A I-1 LRs vs ILRs:
Panel A) shows an example of a step pattern of lamina landmarks in the sagittal plane,
Panel B) is an example 2D image with no visible laminae in the LR region

The presence of visible laminae in an image is intrinsically associated with a LR. Nonetheless, the physical properties governing the propagation of ultrasound waves engender the conflation of multiple echo responses emanating from bone surfaces located at disparate skin-depth values, thereby confounding the visual depiction of the bone surface in ultrasound images. In the context of spinal ultrasound scanning, the absence of visible laminae is not exclusive to frames that scan articular processes (i.e., ILR frames). Frames that scan stair-shaped bone surfaces

within a LR may also exhibit this feature. By leveraging the extracted lamina landmarks, an ultrasound sequence can be separated into alternating sub-sequences of lamina-present and lamina-absent frames. Thus, the task of vertebral level separation, based on localized lamina landmarks, involves grouping successive lamina-present sub-sequences into different LR, with the lamina-absent sub-sequence between two adjacent LR representing the ILR.

As illustrated in Figure I-1 panel A, the 3D lamina landmarks exhibit a regular step pattern along the inferior-superior (horizontal) direction in the sagittal plane, with a gradual decrease in the skin-bone depth value within each step. Each step is regarded as an individual LR, while the horizontal space between consecutive steps represents an ILR. The skin-bone depth difference between adjacent steps, referred to as the **gap**, serves as a characteristic attribute of an ILR. Consequently, consecutive lamina-present sub-sequences within the same step are grouped as a single LR, whereas each lamina-absent sub-sequence between two consecutive steps is identified as an ILR. Once the LR/ILR step pattern is established, the gap corresponding to the ILR between L5 and T12 is readily identifiable, typically showcasing a notable reduction in gap size from the lumbar to the thoracic region. This clear demarcation facilitates the determination of the lumbar-thoracic boundary, making vertebral level identification a straightforward task.

This method enables vertebral level identification from ultrasound sequences without requiring a known initial vertebral level. The primary components are the lamina landmarks (derived either manually or through AI-driven predictions) projected in the sagittal plane. This method has a high tolerance for positional accuracy of lamina landmarks, thanks to the step of landmark projection onto the sagittal plane, where the step pattern generally accepts all points within the lamina area.

ANNEX II

SCOLIOSIS ASSESSMENT USING PROJECTED LAMINAE VECTORS (PLVS)

Vertebral vectors have been employed for geometric spine shape representation from EOS X-ray images (Illés, Tunyogi-Csapó & Somoskeöy, 2010), enabling a convenient description of vertebral rotations and 3D deformity analysis. In this section, we introduce a method for scoliosis assessment from 3D ultrasound sequences using projected laminae vectors (PLVs).

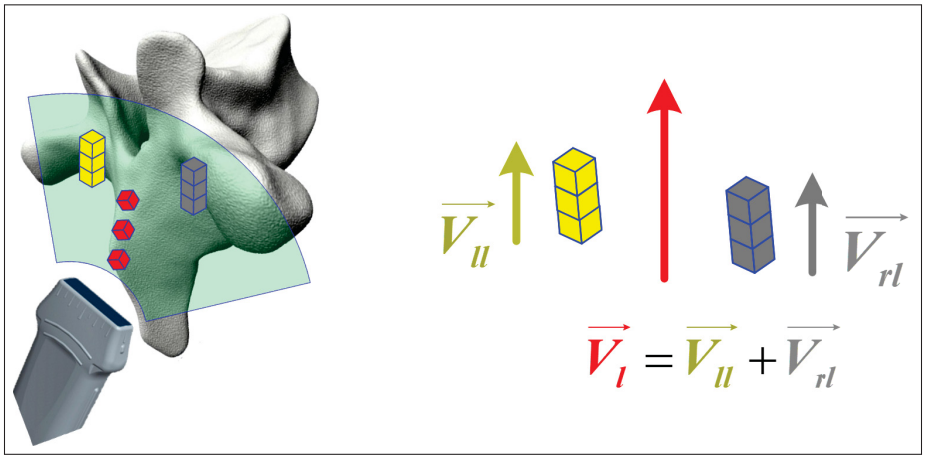

Figure-A II-1 Laminae Vectors from Spinal FH 3D US

Figure II-1 illustrates how a vertebra can be represented by a 3D lamina vector $\overrightarrow{V_l}$ using a vertebral sub-sequence of lamina landmarks extracted from a 3D ultrasound sequence. For the purpose of scoliosis assessment, the lamina vector $\overrightarrow{V_l}$ can then be projected onto the coronal plane. The 2D PLV $\overrightarrow{V_{pl}}$ is represented by

$$\overrightarrow{V_{pl}}(y_w, z_w) = \overrightarrow{V_{pll}} + \overrightarrow{V_{prl}}, \tag{A II-1}$$

where y_w and z_w denote variables in the World Space (refer to Figure 1.2). For each PLV $\overrightarrow{V_{pl}}(y_w, z_w)$, its vertebral roll angle θ_x is determined by

$$\theta_x = \arctan(y_w / z_w). \tag{A II-2}$$

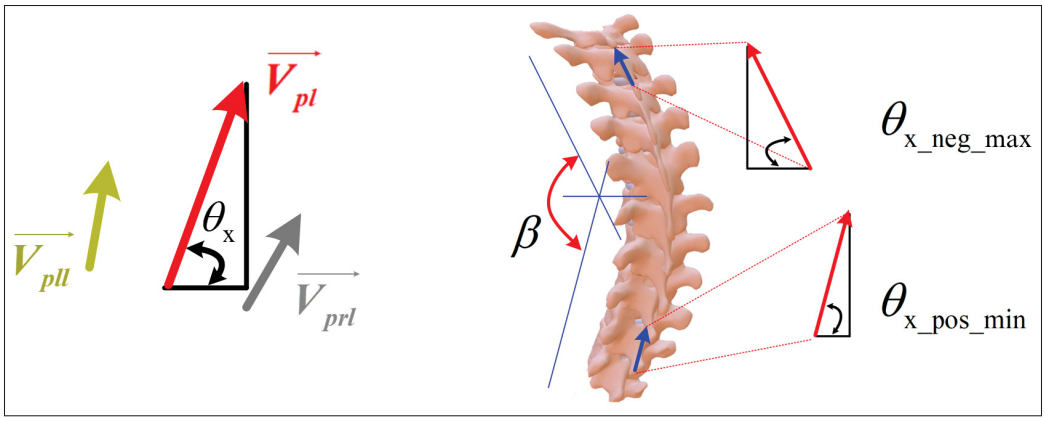

Figure-A II-2 PLV Method for Scoliosis Assessment

As illustrated in Figure II-2, the two most tilted vertebrae can be automatically identified by locating the two roll angles: $\theta_{x\_neg\_max}$ (the maximum negative θ_x) and $\theta_{x\_pos\_min}$ (the minimum positive θ_x). Subsequently, the severity of scoliosis deformity can be described by

$$\beta = \theta_{x\_pos\_min} - \theta_{x\_neg\_max}. \quad \text{(A II-3)}$$

This method utilizes pairwise lamina landmarks rather than spinous process landmarks, as their projection onto the posterior-anterior plane does not accurately represent spinal curvature under transverse plane vertebral rotation. Compared to the VPI-based methods (refer to Section 1.2.1), our proposed PLV method measures scoliosis deformity angle in world space, whereas VPI-based methods derive measurements from a non-planar reslice in vertebral space, which can introduce theoretical biases. Additionally, the PLV method is landmark-based, enabling 3D analysis of vertebral rotation through projections onto anatomical planes. In contrast, VPI-based methods lack the capability to provide information about axial-plane vertebral rotation. Moreover, our proposed PLV method can be executed fully automatically, whereas the VPI-based method necessitates manual intervention to assess depth values for obtaining the correct non-planar reslice.

BIBLIOGRAPHY

Abdi, A. H., Luong, C., Tsang, T., Jue, J., Gin, K., Yeung, D., Hawley, D., Rohling, R. & Abolmaesumi, P. (2017). Quality assessment of echocardiographic cine using recurrent neural networks: feasibility on five standard view planes. *MICCAI 2017*, pp. 302–310.

Acar, P., Battle, L., Dulac, Y., Peyre, M., Dubourdieu, H., Hascoet, S. & Vayssière, C. (2014). Real-time three-dimensional foetal echocardiography using a new transabdominal xMATRIX array transducer. *Archives of Cardiovascular Diseases*, 107(1), 4-9. doi: 10.1016/j.acvd.2013.10.003.

Anagnostoudis, A. & Jan, J. (2005, 05). Use of an electromagnetic calibrated pointer in 3D freehand ultrasound calibration.

Bai, W., Suzuki, H., Qin, C., Tarroni, G., Oktay, O., Matthews, P. M. & Rueckert, D. (2018). Recurrent neural networks for aortic image sequence segmentation with sparse annotations. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pp. 586–594.

Baum, Z., Ungi, T., Lasso, A., Church, B., Schlenger, C. & Fichtinger, G. (2018). Visual aid for identifying vertebral landmarks in ultrasound . *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*, 10576, 105760Z. doi: 10.1117/12.2292820.

Berton, F., Azzabi, W., Cheriet, F. & Laporte, C. (2015, 07). Automatic segmentation of vertebrae in ultrasound images. pp. 344-351. doi: 10.1007/978-3-319-20801-5_37.

Berton, F., Cheriet, F., Miron, M.-C. & Laporte, C. (2016a). Segmentation of the spinous process and its acoustic shadow in vertebral ultrasound images. *Computers in Biology and Medicine*, 72. doi: 10.1016/j.compbiomed.2016.03.018.

Berton, F., Cheriet, F., Miron, M.-C. & Laporte, C. (2016b). Segmentation of the spinous process and its acoustic shadow in vertebral ultrasound images. *Computers in Biology and Medicine*, 72. doi: 10.1016/j.compbiomed.2016.03.018.

Boisvert, J., Cheriet, F., Pennec, X., Labelle, H. & Ayache, N. (2008a). Geometric variability of the scoliotic spine using statistics on articulated shape models. *IEEE Transactions on Medical Imaging*, 4(27), 557–568. doi: 10.1109/tmi.2007.911474.

Boisvert, J., Cheriet, F., Pennec, X., Labelle, H. & Ayache, N. (2008b). Articulated spine models for 3-D reconstruction from partial radiographic data. *IEEE Transactions on Biomedical Engineering*, 55(11), 2565-2574. doi: 10.1109/TBME.2008.2001125.

Boundless.com. (2019). *Anatomy and Physiology (Boundless)*.

Brignol, A., Gueziri, H., Cheriet, F., Collins, D. & Laporte, C. (2020). Automatic extraction of vertebral landmarks from ultrasound images: a pilot study. *Computers in Biology and Medicine*, 122. doi: 10.1016/j.compbiomed.2020.103838.

Brooks, H., Azen, S., Gerberg, E., Brooks, R. & Chan, L. (1975). Scoliosis: a prospective epidemiological study. *J Bone Joint Surg Am*, 57(7), 968-972.

Brotherston, D., Desy, J. & Ma, I. W. (2021). Vertebral level identified by the intercristal line: Confirmation by ultrasound. *European Journal of Internal Medicine*, 86, 118-120. doi: https://doi.org/10.1016/j.ejim.2021.01.015.

Cahuantzi, R., Chen, X. & Güttel, S. (2021). A comparison of LSTM and GRU networks for learning symbolic sequences. *CoRR*, abs/2107.02248.

Chan, A., Coutts, B., Parent, E. & Lou, E. (2021). Development and evaluation of CT-to-3D ultrasound image registration algorithm in vertebral phantoms for spine surgery. *Annals of Biomedical Engineering*, 49(1), 310-321. doi: 10.1007/s10439-020-02546-5.

Che, C., Mathai, T. S. & Galeotti, J. (2017). Ultrasound registration: A review. *Methods*, 115, 128-143. doi: https://doi.org/10.1016/j.ymeth.2016.12.006. Image Processing for Biologists.

Chen, W., Lou, E. H. M. & Le, L. H. (2011). Using ultrasound imaging to identify landmarks in vertebra models to assess spinal deformity. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 8495-8498.

Chen, W., Lou, E. H. M., Zhang, P. Q., Le, L. H. & Hill, D. (2013). Reliability of assessing the coronal curvature of children with scoliosis by using US images. *Journal of Children's Orthopaedics*, 7, 521–529. doi: 10.1007/s11832-013-0539-y.

Chen, W., Le, L. H. & Lou, E. H. M. (2016). Reliability of the axial vertebral rotation measurements of adolescent idiopathic scoliosis using the center of lamina method on ultrasound images: in vitro and in vivo study. *European Spine Journal*, 25, 3265–3273. doi: 10.1007/s00586-016-4492-6.

CHERIET, F., DANSEREAU, J., PETIT, Y., AUBIN, C.-E., LABELLE, H. & DE GUISE, J. A. (1999). Towards the Self-Calibration of a Multiview Radiographic Imaging System for the 3D Reconstruction of the Human Spine and Rib Cage. *International Journal of Pattern Recognition and Artificial Intelligence*, 13(05), 761-779. doi: 10.1142/S0218001499000434.

Cheriet, F., Laporte, C., Kadoury, S., Labelle, H. & Dansereau, J. (2007). A Novel System for the 3-D Reconstruction of the Human Spine and Rib Cage From Biplanar X-Ray Images. *IEEE Transactions on Biomedical Engineering*, 54, 1356 - 1358. doi: 10.1109/TBME.2006.889205.

Cheung, C. J., Zhou, G. & et al., S. L. (2015). Ultrasound volume projection imaging for assessment of scoliosis. *IEEE Transactions on Medical Imaging*, 34(8), 1760-1768. doi: 10.1109/TMI.2015.2390233.

Cheung, C.-W. J., Law, S.-Y. & Zheng, Y.-P. (2013). Development of 3-D ultrasound system for assessment of adolescent idiopathic scoliosis (AIS): and system validation. *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6474-6477. doi: 10.1109/EMBC.2013.6611037.

Cheung, J. P. Y., Cheung, P. W. H., Samartzis, D. & Luk, K. D. (2018). Curve progression in adolescent idiopathic scoliosis does not match skeletal growth. *Clin Orthop Relat Res*, 2(476), 429-436. doi: 10.1007/s11999.0000000000000027.

Chin, K. J., Karmakar, M. K. & Peng, P. (2011a). Ultrasonography of the Adult Thoracic and Lumbar Spine for Central Neuraxial Blockade. *Anesthesiology*, 114, 1459–1485. doi: 10.1097/ALN.0b013e318210f9f8.

Chin, K. J., Karmakar, M. K. & Peng, P. (2011b). Ultrasonography of the adult thoracic and lumbar spine for central neuraxial blockade. *Anesthesiology*, 114(6), 1459-85. doi: 10.1097/ALN.0b013e318210f9f8.

CL, N. J. & JH., M. (1969). A study of vertebral rotation. *The Journal of Bone and Joint Surgery*, 51(2), 223–229.

coursehero.com. (2019). *Anatomy and Physiology*.

Delorme, S., Petit, Y., de Guise, J., Labelle, H., Aubin, C.-E. & Dansereau, J. (2003). Assessment of the 3-D reconstruction and high-resolution geometrical modeling of the human skeletal trunk from 2-D radiographic images. *IEEE Transactions on Biomedical Engineering*, 50(8), 989-998. doi: 10.1109/TBME.2003.814525.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. & Fei-Fei, L. (2009). Imagenet: a large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255.

Deng, Q. & Huang, Q. (2019). Detecting vertebra landmarks from ultrasound image using Single Shot MultiBox Detector. *2019 6th International Conference on Systems and Informatics (ICSAI)*, pp. 757-761. doi: 10.1109/ICSAI48974.2019.9010161.

Devries, T. & Taylor, G. W. (2017). Improved Regularization of Convolutional Neural Networks with Cutout. *CoRR*, abs/1708.04552.

Donahue, J., Hendricks, L. A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K. & Darrell, T. (2017). Long-term recurrent convolutional networks for visual recognition and description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 677-691. doi: 10.1109/TPAMI.2016.2599174.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *CoRR*, abs/2010.11929.

Downey, D. B., Fenster, A. & Williams, J. C. (2000). Clinical utility of three-dimensional US. *RadioGraphics*, 20(2), 559–571. doi: 10.1148/radiographics.20.2.g00mc19559.

Droste, R., Drukker, L., Papageorghiou, A. T. & Noble, J. A. (2020). Automatic probe movement guidance for freehand obstetric ultrasound.

Effatparvar, M. R., Pierre, M.-O. S. & Sobczak, S. (2022). Assessment and improvement of a novel ultrasound-based 3D reconstruction method: Registered for lumbar spine. *J. Med. Biol. Eng.*, 42(6), 790–799.

Esteban, J., Simson, W., Requena Witzig, S., Rienmüller, A., Virga, S., Frisch, B., Zettinig, O., Sakara, D., Ryang, Y.-M., Navab, N. & Hennersperger, C. (2018). Robotic ultrasound-guided facet joint insertion. *International journal of computer assisted radiology and surgery*, 13(6), 895—904. doi: 10.1007/s11548-018-1759-x.

et al., K. C. (2000). Final Report of the Technical Requirements for Image-Guided Spine Procedures Workshop. *Computer Aided Surgery*, 5(3), 180-215. doi: 10.3109/10929080009148889. PMID: 10964090.

Fan, X., Zhu, Q., Tu, P., Joskowicz, L. & Chen, X. (2023). A review of advances in image-guided orthopedic surgery. *Physics in Medicine & Biology*, 68(2), 02TR01. doi: 10.1088/1361-6560/acaae9.

Feldman, M. K., Katyal, S. & Blackwood, M. S. (2009). US artifacts. *RadioGraphics*, 29(4), 1179–1189. doi: 10.1148/rg.294085199.

Fenster, A. & Downey, D. B. (1996). 3-D ultrasound imaging: a review. *IEEE Engineering in Medicine and Biology Magazine*, 15(6), 41-51. doi: 10.1109/51.544511.

Fenster, A. & Downey, D. B. (2000). Three-dimensional ultrasound imaging. *Annual Review of Biomedical Engineering*, 2(1), 457-475. doi: 10.1146/annurev.bioeng.2.1.457.

Ghosh, S., Madjdpour, C. & Chin, K. (2016). Ultrasound-guided lumbar central neuraxial block. *BJA Education*, 16(7), 213-220. doi: https://doi.org/10.1093/bjaed/mkv048.

Gibby, J. T., Cvetko, S., Javan, R., Parr, R. & Gibby, W. A. (2020). Use of augmented reality for image-guided spine procedures. *European Spine Journal*, 1-10. Retrieved from: https://api.semanticscholar.org/CorpusID:220080092.

Glocker, B., Zikic, D. & Haynor, D. R. (2014). Robust registration of longitudinal spine CT. *Med Image Comput Comput Assist Interv*, 17, 251-8. doi: 10.1007/978-3-319-10404-1_32.

Glorot, X. & Bengio, Y. (2010, May). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 9(Proceedings of Machine Learning Research), 249–256.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*.

Grau, T., Leipold, R. W., Conradi, R., Martin, E. & Motsch, J. (2002). Efficacy of ultrasound imaging in obstetric epidural anesthesia. *Journal of Clinical Anesthesia*, 14(3), 169-175. doi: https://doi.org/10.1016/S0952-8180(01)00378-6.

Graumann, C., Fuerst, B., Hennersperger, C., Bork, F. & Navab, N. (2016). Robotic ultrasound trajectory planning for volume of interest coverage. *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 736-741.

Greiner, K. A. (2002). Adolescent idiopathic scoliosis: radiologic decision-making. *American family physician*, 65(9), 1817–1822.

Guerroumi, N., Playout, C., Laporte, C. & Cheriet, F. (2019, April). Automatic segmentation of the scoliotic spine from MR images. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 480-484. doi: 10.1109/ISBI.2019.8759413.

Gueziri, H.-E., Drouin, S., Yan, C. X. B. & Collins, D. L. (2019a). Toward real-time rigid registration of intra-operative ultrasound with preoperative CT images for lumbar spinal fusion surgery. *International journal of computer assisted radiology and surgery*, 14(11), 1933—1943. doi: 10.1007/s11548-019-02020-1.

Gueziri, H.-E., Drouin, S., Yan, C. X. B. & Collins, D. L. (2019b). Toward real-time rigid registration of intra-operative ultrasound with preoperative CT images for lumbar spinal fusion surgery. *Int J Comput Assist Radiol Surg*, 14, 1933–1943. doi: 10.1007/s11548-019-02020-1.

Gueziri, H.-E., Drouin, S., Yan, C. X. B. & Collins, D. L. (2019c). Toward real-time rigid registration of intra-operative ultrasound with preoperative CT images for lumbar spinal fusion surgery. *Int J Comput Assist Radiol Surg*, 14, 1933–1943. doi: 10.1007/s11548-019-02020-1.

Gueziri, H.-E., Santaguida, C. & Collins, D. L. (2020). The state-of-the-art in ultrasound-guided spine interventions. *Medical Image Analysis*, 65, 101769. doi: https://doi.org/10.1016/j.media.2020.101769.

Görres, J., Uneri, A., Silva, T., Ketcha, M., Reaungamornrat, S., Jacobson, M., Vogt, S., Kleinszig, G., Osgood, G., Wolinsky, J.-P. & Siewerdsen, J. (2017). Spinal pedicle screw planning using deformable atlas registration. *Physics in Medicine and Biology*, 62, 7. doi: 10.1088/1361-6560/aa5f42.

Hacihaliloglu, I. (2017). Ultrasound imaging and segmentation of bone surfaces: a review. *TECHNOLOGY*, 05(02), 74–80. doi: 10.1142/s2339547817300049.

Hacihaliloglu, I., Rasoulian, A., Rohling, R. N. & Abolmaesumi, P. (2013). Statistical Shape Model to 3D Ultrasound Registration for Spine Interventions Using Enhanced Local Phase Features. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, pp. 361–368.

Hacihaliloglu, I., Rasoulian, A., Rohling, R. N. & Abolmaesumi, P. (2014). Local Phase Tensor Features for 3-D Ultrasound to Statistical Shape+Pose Spine Model Registration. *IEEE Transactions on Medical Imaging*, 33(11), 2167-2179. doi: 10.1109/TMI.2014.2332571.

Handalage, U. & Kuganandamurthy, L. [Available online at ResearchGate]. (2021). Real-time Object Detection Using YOLO: A Review. Retrieved from: ResearchGate.

Harmouche, R., Cheriet, F., Labelle, H. & Dansereau, J. (2012). 3D registration of MR and X-ray spine images using an articulated model. *Computerized Medical Imaging and Graphics*, 36(5), 410-418. doi: https://doi.org/10.1016/j.compmedimag.2012.03.003.

He, K., Zhang, X., Ren, S. & Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, abs/1512.03385.

He, K., Zhang, X., Ren, S. & Sun, J. (2016). Identity mappings in deep residual networks. *CoRR*, abs/1603.05027.

Hetherington, J., Lessoway, V., Gunka, V., Abolmaesumi, P. & Rohling, R. (2017). SLIDE: automatic spine level identification system using a deep convolutional neural network. *Int J Comput Assist Radiol Surg*, 12(7), 1189-1198. doi: 10.1007/s11548-017-1575-8.

Hetherington, J. H. (2017). *Automated lumbar vertebral level identification using ultrasound*. (Ph.D. thesis, University of British Columbia).

Horng, M.-H., Kuok, C.-P., Fu, M.-J., Lin, C.-J. & Sun, Y.-N. (2019). Cobb Angle Measurement of Spine from X-Ray Images Using Convolutional Neural Network. *Computational and Mathematical Methods in Medicine*, 2019(6357171). doi: https://doi.org/10.1155/2019/6357171.

Hsu, P.-W., Prager, R. W., Gee, A. H. & Treece, G. M. (2009). Freehand 3D Ultrasound Calibration: A Review. In Sensen, C. W. & Hallgrímsson, B. (Eds.), *Advanced Imaging in Biology and Medicine: Technology, Software Environments, Applications* (pp. 47–84). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-68993-5_3.

Hu, J., Shen, L. & Sun, G. (2017). Squeeze-and-excitation networks. *CoRR*, abs/1709.01507.

Huang, G., Liu, Z. & Weinberger, K. Q. (2016). Densely connected convolutional networks. *CoRR*, abs/1608.06993.

Huang, Q. & Zeng, Z. (2017). A review on real-time 3D ultrasound imaging technology. *BioMed research international*, 6027029. doi: 10.1155/2017/6027029.

Huang, W., Bridge, C. P., Noble, J. A. & Zisserman, A. (2017). Temporal HeartNet: towards human-level automatic analysis of fetal cardiac screening video. *MICCAI 2017*, pp. 341–349.

Huh, M., Agrawal, P. & Efros, A. A. (2016). What makes ImageNet good for transfer learning? *CoRR*, abs/1608.08614.

Illés, T., Tunyogi-Csapó, M. & Somoskeöy, S. (2010). Breakthrough in three-dimensional scoliosis diagnosis: significance of horizontal plane view and vertebra vectors. *European Spine Journal*, 1(20), 135–143. doi: 10.1007/s00586-010-1566-8.

Illés, T., Lavaste, F. & Dubousset, J. (2019). The third dimension of scoliosis: the forgotten axial plane. *Orthopaedics & Traumatology: Surgery & Research*, 105. doi: 10.1016/j.otsr.2018.10.021.

Ioffe, S. & Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167.

Jafari, M. H., Girgis, H., Liao, Z., Behnami, D., Abdi, A., Vaseli, H., Luong, C., Rohling, R., Gin, K., Tsang, T. & Abolmaesumi, P. (2018). A unified framework integrating recurrent fully-convolutional networks and optical flow for segmentation of the left ventricle in echocardiography Data. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 29–37.

Jain, A. K. & Taylor, R. H. (2004). Understanding bone responses in B-mode ultrasound images and automatic bone surface extraction using a Bayesian probabilistic framework. *SPIE 5373, Medical Imaging 2004: Ultrasonic Imaging and Signal Processing*. doi: 10.1117/12.535984.

Jiang, C.-H., Shi, Y., Sun, Y.-M., Cai, M.-J., Wu, H.-L., Hu, L.-S., Yu, L.-M., Wang, P., Shen, J., Huang, Y.-C. & Yu, B.-S. (2023). Using a 3D Navigation Template to Increase the Accuracy of Thoracic Pedicle Screws in Patients with Scoliosis. *Bioengineering*, 10(7). doi: 10.3390/bioengineering10070756.

Jiang, W. & Zhang, J. (2016). 3D Surface Reconstruction of Spinal Ultrasound Images Based on 2D Contour Line. *Proceedings of the 2016 3rd International Conference on Materials Engineering, Manufacturing Technology and Control*, pp. 110-115. doi: 10.2991/icmemtc-16.2016.21.

JR., C. (1948). Outlines for the study of scoliosis. *The American academy of orthopaedic surgeons. Instructional course lectures.*, 5, 261-275.

Kamali, S., Ungi, T., Lasso, A., Yan, C., Lougheed, M. & Fichtinger, G. (2017). Localization of the transverse processes in ultrasound for spinal curvature measurement. *Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling*, 10135, 101350I. doi: 10.1117/12.2256007.

Karmakar, M. K. & Chin, K. J. (2012). Spinal Sonography and Applications of Ultrasound for Central Neuraxial Blocks. Retrieved from: https://www.nysora.com/techniques/neuraxial-and-perineuraxial-techniques/spinal-sonography-and-applications-of-ultrasound-for-central-neuraxial-blocks/.

Kim, H.-S. (2014). Evidence-based of nonoperative treatment in adolescent idiopathic scoliosis. *Asian Spine J*, 8(5), 695–702. doi: 10.4184/asj.2014.8.5.695.

Kim, H., Kim, H., Moon, E., Yoon, C., Chung, T., Song, H.-T., Suh, J., Lee, Y. & Kim, S. (2010). Scoliosis imaging: what radiologists should know. *Radiographics*, 30(7), 1823–1842. doi: 10.1148/rg.307105061.

Kingma, D. P. & Ba, J. L. (2015). Adam: a method for stochastic optimization. *ICLR*, abs/1412.6980.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P. & Girshick, R. (2023). Segment Anything.

Konieczny, M. R., Senyurt, H. & Krauspe, R. (2013). Epidemiology of adolescent idiopathic scoliosis. *Journal of Children's Orthopaedics*, 7(1), 3-9. doi: 10.1007/s11832-012-0457-4.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25* (pp. 1097–1105).

Lasso, A., Heffter, T., Rankin, A., Pinter, C., Ungi, T. & Fichtinger, G. (2014). PLUS: open-source toolkit for ultrasound-guided intervention systems. *IEEE Transactions on Biomedical Engineering*, (10), 2527-2537. doi: 10.1109/TBME.2014.2322864.

Lee, A. (2014). Ultrasound in obstetric anesthesia. *Seminars in Perinatology*, 38(6), 349-358. doi: https://doi.org/10.1053/j.semperi.2014.07.006. Special topics in obstetric anesthesia.

Lee, T. T.-Y., Lai, K. K.-L., Cheng, J. C.-Y., Castelein, R. M., Lam, T.-P. & Zheng, Y.-P. (2021). 3D ultrasound imaging provides reliable angle measurement with validity comparable to X-ray in patients with adolescent idiopathic scoliosis. *Journal of Orthopaedic Translation*, 29, 51-59. doi: https://doi.org/10.1016/j.jot.2021.04.007.

Lee, W. & Roh, Y. (2017). Ultrasonic transducers for medical diagnostic imaging. *Biomedical Engineering Letters*, 7(2), 91–97. doi: 10.1007/s13534-017-0021-8.

Lei, X., Zhang, Z. & Dong, P. (2018). Dynamic path planning of unknown environment based on deep reinforcement learning. *Journal of Robotics*, 1–10. doi: 10.1155/2018/5781591.

Levy, A. R., Goldberg, M. S., Mayo, N. E., Hanley, J. A. & Poitras, B. (1996). Reducing the lifetime risk of cancer from spinal radiographs among people with adolescent idiopathic scoliosis. *Spine*, 21(13), 1540–1547. doi: 10.1097/00007632-199607010-00011.

Li, M., Cheng, J., Ying, M., Ng, B., Zheng, Y.-P., Lam, T., Wong, W. & Wong, M. (2012). Could clinical ultrasound improve the fitting of spinal orthosis for the patients with AIS? *European spine journal*, 21, 1926-35. doi: 10.1007/s00586-012-2273-4.

Li, S. (2016). *Per-Pixel Calibration for RGB-Depth Natural 3D Reconstruction on GPU*. (Master's thesis, University of Kentucky). Retrieved from: https://uknowledge.uky.edu/ece_etds/93.

Li, S., Gauthier, L., Cheriet, F., Fortin, C., Paquette, P. & Laporte, C. (2023). Lamina landmark detection in ultrasound images: a preliminary study. *Medical Imaging 2023: Ultrasonic Imaging and Tomography*, 12470, 1247011. doi: 10.1117/12.2653277.

Li, Y., Liang, W., Zhang, Y. & Tan, J. (2018). Automatic global level set approach for lumbar vertebrae CT image segmentation. *BioMed Research International*, 2018(03), 1-12. doi: 10.1155/2018/6319879.

Litjens, G. J. S., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B. & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60-88.

Long, J., Shelhamer, E. & Darrell, T. (2015, June). Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431-3440. doi: 10.1109/CVPR.2015.7298965.

Marieb, E. N. & Hoehn, K. (2013). *Human Anatomy and Physiology* (ed. 9). Pearson.

Mattjie, C., de Moura, L. V., Ravazio, R. C., Kupssinskü, L. S., Parraga, O., Delucis, M. M. & Barros, R. C. (2023). Zero-shot performance of the Segment Anything Model (SAM) in 2D medical imaging: A comprehensive evaluation and practical guidelines.

McKenna, C., Wade, R., Faria, R., Yang, H., Stirk, L., Gummerson, N., Sculpher, M. & Woolacott, N. (2012). EOS 2D/3D X-ray imaging system: a systematic review and economic evaluation. *Health Technology Assessment*, 16(14). doi: 10.3310/hta16140.

McNally, D., Naish, C. & Halliwell, M. (2000). Intervertebral disc structure: observation by a novel use of ultrasound imaging. *Ultrasound in Medicine & Biology*, 26, 751-8. doi: 10.1016/S0301-5629(00)00209-X.

Milletari, F., Birodkar, V. & Sofka, M. (2019). *Straight to the point: reinforcement learning for user guidance in ultrasound*.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533.

Moiyadi, A. & Shetty, P. (2016). Direct navigated 3D ultrasound for resection of brain tumors: a useful tool for intraoperative image guidance. *Neurosurgical Focus*, 40, E5. doi: 10.3171/2015.12.FOCUS15529.

Mozaffari, M. H. & Lee, W.-S. (2017). Freehand 3-D ultrasound imaging: a systematic review. *Ultrasound in Medicine & Biology*, 43(10), 2099 - 2124. doi: 10.1016/j.ultrasmedbio.2017.06.009.

Muratore, D. M. & Galloway, R. L. (2001). Beam calibration without a phantom for creating a 3-D freehand ultrasound system. *Ultrasound in Medicine & Biology*, 27(11), 1557-1566. doi: https://doi.org/10.1016/S0301-5629(01)00469-0.

Nguyen, D. V., Vo, Q. N., Le, L. H. & Lou, E. H. (2015). Validation of 3D surface reconstruction of vertebrae and spinal column using 3D ultrasound data – A pilot study. *Medical Engineering & Physics*, 37(2), 239-244. doi: https://doi.org/10.1016/j.medengphy.2014.11.007.

Ning, Y., Gao, P., Sun, Y., Qi, X. & Hu, Y. (2017). A practical calibration method for spinal surgery robot. *2017 18th International Conference on Advanced Robotics (ICAR)*, 137-142. Retrieved from: https://api.semanticscholar.org/CorpusID:34087122.

North American Spine Society Patient Education Committee. (2020). Scoliosis, Adolescent Idiopathic. Retrieved from: https://www.spine.org/KnowYourBack/Conditions/Other/Scoliosis-Adolescent-Idiopathic.

Oakley, P. A. (2018). Is early treatment for mild adolescent idiopathic scoliosis superior over the traditional watch & wait approach? A case report with long-term follow-up. *Journal of physical therapy science*, 5(30), 680–684. doi: doi.org/10.1589/jpts.30.680.

Ogilvie, J. W. (2020). What's important: moving beyond idiopathic in adolescent idiopathic scoliosis. 3(102), 269-270. doi: 10.2106/JBJS.19.00433.

Pedrosa, J., Barbosa, D., Almeida, N., Bernard, O., Bosch, J. & D'hooge, J. (2016). Cardiac chamber volumetric assessment using 3D ultrasound - a review. *Current Pharmaceutical Design*, 22(105). doi: 10.2174/1381612822666151109112652.

Pillen, S., Boon, A. & Van Alfen, N. (2016). Chapter 42 - Muscle ultrasound. In *Neuroimaging Part II* (vol. 136, pp. 843-853). Elsevier. doi: https://doi.org/10.1016/B978-0-444-53486-6.00042-9.

Pinter, C., Travers, B., Baum, Z., Kamali, S., Ungi, T., Lasso, A., Church, B. & Fichtinger, G. (2018). Real-time transverse process detection in ultrasound. *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*, 10576, 105760Y. doi: 10.1117/12.2295605.

Poudel, R. P. K., Lamata, P. & Montana, G. (2017). Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. *Reconstruction, Segmentation, and Analysis of Medical Images*, pp. 83–94.

Prager, R. W., Ijaz, U. Z., Gee, A. H. & Treece, G. M. (2010). Three-dimensional ultrasound imaging. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 224(2), 193-223. doi: 10.1243/09544119JEIM586.

Purnama, K. E. & et al. (2010). A framework for human spine imaging using a freehand 3D ultrasound system. *Technol Health Care*, 18(1), 1–17. doi: 10.3233/THC-2010-0565.

R., P. & J., V. (1981). A study of scoliotic curve. The importance of extension and vertebral rotation. *Revue de Chirurgie Orthopedique et Reparatrice de l'Appareil Moteur*, 67(1), 25-34.

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016, jun). You only look once: unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788. doi: 10.1109/CVPR.2016.91.

Ren, S., He, K., Girshick, R. B. & Sun, J. (2015). Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497.

Rogala, E. J., Drummond, D. S. & Gurr, J. (1978). Scoliosis: incidence and natural history. A prospective epidemiological study. *J Bone Joint Surg Am*, 60(2), 173-176.

Roh, Y. (2014). Ultrasonic transducers for medical volumetric imaging. *Japanese Journal of Applied Physics*, 53, 07KA01. doi: 10.7567/JJAP.53.07KA01.

Ronneberger, O., Fischer, P. & Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241.

Roy, A. G., Navab, N. & Wachinger, C. (2018). Recalibrating fully convolutional networks with spatial and channel 'squeeze & excitation' blocks. *CoRR*, abs/1808.08127.

Sadeghi, A., Patel, R. & Carvalho, J. C. A. (2021a). Ultrasound-facilitated neuraxial anaesthesia in obstetrics. *BJA Educ*, 10, 369-375. doi: 10.1016/j.bjae.2021.06.003.

Sadeghi, A., Patel, R. & Carvalho, J. C. A. (2021b). Ultrasound-facilitated neuraxial anaesthesia in obstetrics. *BJA Educ*, 10, 369-375. doi: 10.1016/j.bjae.2021.06.003.

Sayed, T., Khodaei, M., Hill, D. & Lou, E. (2022). Intra- and inter-rater reliabilities and differences of kyphotic angle measurements on ultrasound images versus radiographs for children with adolescent idiopathic scoliosis: a preliminary study. *Spine Deform*, 10(3), 501–507.

Schönemann, P. H. (1966). A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31, 1-10. Retrieved from: https://api.semanticscholar.org/CorpusID:121676935.

Scoliosis Research Society. (2019). Adolescent idiopathic scoliosis. Retrieved from: https://etext.srs.org/patients-and-families/conditions-and-treatments/parents/scoliosis/adolescent-idiopathic-scoliosis.

Selbekk, T., Jakola, A. S., Solheim, O., Johansen, T. F., Lindseth, F., Reinertsen, I. & Unsgard, G. (2013). Ultrasound imaging in neurosurgery: approaches to minimize surgically induced image artefacts for improved resection control. *Acta Neurochirurgica*, 155(6), 973–980. doi: 10.1007/s00701-013-1647-7.

Semeniuta, O. (2021). Subset-based stereo calibration method optimizing triangulation accuracy. *PeerJ Comput. Sci.*, 7(e485). doi: 10.7717/peerj-cs.485.

Shi, P., Qiu, J., Abaxi, S. M. D., Wei, H., Lo, F. P.-W. & Yuan, W. (2023). Generalist Vision Foundation Models for Medical Imaging: A Case Study of Segment Anything Model on Zero-Shot Medical Segmentation. *Diagnostics*, 13(11), 1947. doi: 10.3390/diagnostics13111947.

Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556.

Soucacos, P. N., Zacharis, K., Gelalis, J., Soultanis, K., Kalos, N., Beris, A., Xenakis, T. & Johnson, E. O. (1998). Assessment of curve progression in idiopathic scoliosis. *European Spine Journal*, 7, 270–277.

Steenbergen, T., Geest, I., Janssen, D., Rovers, M. & Fütterer, J. (2019). Feasibility study of intraoperative cone-beam CT navigation for benign bone tumor surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 15, 3. doi: 10.1002/rcs.1993.

STOKES, I. A. F., BIGALOW, L. C. & MORELAND, M. S. (1986). Measurement of axial rotation of vertebrae in scoliosis. *Spine*, 11(3), 213–218. doi: 10.1097/00007632-198604000-00006.

Stokes, I. A. F. (1994). Three-dimensional terminology of spinal deformity. A report presented to the Scoliosis Research Society by the Scoliosis Research Society Working Group on 3-D terminology of spinal deformity. *Spine*, 19 2, 236-48. Retrieved from: https://api.semanticscholar.org/CorpusID:40375677.

Sutton, R. S. & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (ed. Second).

Suzuki, S., Yamamuro, T., Shikata, Shimizu, K. & Iida, H. (1989). Ultrasound measurement of vertebral rotation in idiopathic scoliosis. *The Journal of Bone and Joint Surgery*, 71-B(2), 252–255. doi: 10.1302/0301-620x.71b2.2647754.

Tang, S., Yang, X., Shajudeen, P., Sears, C., Taraballi, F., Weiner, B., Tasciotti, E., Dollahon, D., Park, H. & Righetti, R. (2021). A CNN-based method to reconstruct 3-D spine surfaces from US images in vivo. *Medical Image Analysis*, 74, 102221. doi: https://doi.org/10.1016/j.media.2021.102221.

Tao, R., Liu, W. & Zheng, G. (2022). Spine-transformers: Vertebra labeling and segmentation in arbitrary field-of-view spine CTs via 3D transformers. *Medical Image Analysis*, 75, 102258. doi: https://doi.org/10.1016/j.media.2021.102258.

The Editor of Encyclopaedia Britannica. (2018). *Vertebra column, Anatomy*. Retrieved from: https://www.britannica.com/science/vertebral-column.

Thomsen, M. & Abel, R. (2006). Imaging in scoliosis from the orthopaedic surgeon's point of view. *European Journal of Radiology*, 58(1), 41 - 47. doi: https://doi.org/10.1016/j.ejrad.2005.12.003. Imaging of Spine Disorders.

Ungi, T., Greer, H., Sunderland, K. R., Wu, V., Baum, Z. M. C., Schlenger, C., Oetgen, M., Cleary, K., Aylward, S. R. & Fichtinger, G. (2020). Automatic Spine Ultrasound Segmentation for Scoliosis Visualization and Measurement. *IEEE Transactions on Biomedical Engineering*, 67(11), 3234-3241. doi: 10.1109/TBME.2020.2980540.

Vasudeva, V. S., Abd-El-Barr, M., Pompeu, Y. A., Karhade, A., Groff, M. W. & Lu, Y. (2017a). Use of intraoperative ultrasound during spinal surgery. *Global Spine J.*, 7(7), 648–656.

Vasudeva, V. S., Abd-El-Barr, M., Pompeu, Y. A., Karhade, A., Groff, M. W. & Lu, Y. (2017b). Use of intraoperative ultrasound during spinal surgery. *Global Spine Journal*, 7(7), 648-656. doi: 10.1177/2192568217700100.

Vavruch, L. & Tropp, H. (2016). A comparison of Cobb angle: standing versus supine images of late-onset idiopathic scoliosis. *Pol J Radiol*, 81, 270–276. doi: 10.12659/PJR.895949.

Vo, Q. N., Lou, E. H. & Le, L. H. (2015). Measurement of axial vertebral rotation using three-dimensional ultrasound images. *Scoliosis*, 10(2), S7. doi: 10.1186/1748-7161-10-S2-S7.

Wang, Q., Li, M., Lou, E. H. M. & Wong, M. S. (2015). Reliability and validity study of clinical ultrasound imaging on lateral curvature of adolescent idiopathic scoliosis. *PLoS One*, 10(8), e0135264.

Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H. & Nandi, A. K. (2022). Medical image segmentation using deep learning: A survey. *IET Image Processing*, 16(5), 1243–1267. doi: 10.1049/ipr2.12419.

Weinstein, S., Zavala, D. & Ponseti, I. (1981). Idiopathic scoliosis: long-term follow-up and prognosis in untreated patients. *The Journal of bone and joint surgery. American volume*, 63(5), 702—712.

Weinstein, S. L., Dolan, L. A., Cheng, J. C. Y., Danielsson, A. & Morcuende, J. A. (2008). Adolescent idiopathic scoliosis. *The Lancet*, 371(9623), 1527-1537. doi: 10.1016/S0140-6736(08)60658-3.

Weiss, H.-R. (2011). The method of Katharina Schroth - history, principles and current development. *BioMed Central*, 6, 1748-7161.

Weiss, H.-R. & Goodall, D. (2008). The treatment of adolescent idiopathic scoliosis (AIS) according to present evidence. A systematic review. *European journal of physical and rehabilitation medicine*, 44, 177-93.

Weiss, H.-R., Weiss, G. & Petermann, F. (2009). Incidence of curvature progression in idiopathic scoliosis patients treated with scoliosis inpatient rehabilitation (SIR): an age- and sex-matched controlled study. *Pediatric rehabilitation*, 6, 23-30. doi: 10.1080/13638490310000952 88.

Wien, T. U., Eiter, T., Eiter, T., Mannila, H. & Mannila, H. (1994). *Computing discrete Fréchet distance*.

Winter, S., Pechlivanis, I., Dekomien, C., Igel, C. & Schmieder, K. (2009). Toward registration of 3D ultrasound and CT images of the spine in clinical praxis: design and evaluation of a data acquisition protocol. *Ultrasound in medicine & biology*, 35, 1773-82. doi: 10.1016/j.ultrasmedbio.2009.06.1089.

Wong, H.-K. & Tan, K.-J. (2010). The natural history of adolescent idiopathic scoliosis. *Indian J Orthop*, 1(44), 9-13. doi: 10.4103/0019-5413.58601.

Wong, Y., Lai, K. K., Zheng, Y., Wong, L. L., Ng, B. K., Hung, A. L., Yip, B. H., Chu, W. C., Ng, A. W., Qiu, Y., Cheng, J. C. & Lam, T. (2019). Is radiation-free ultrasound accurate for quantitative assessment of spinal deformity in idiopathic scoliosis (IS): a detailed analysis with EOS radiography on 952 patients. *Ultrasound in Medicine & Biology*, 45(11), 2866 - 2877. doi: https://doi.org/10.1016/j.ultrasmedbio.2019.07.006.

Wu, H.-D., Chu, W. C.-W., He, C.-Q. & Wong, M.-S. (2020a). Assessment of the plane of maximum curvature for patients with adolescent idiopathic scoliosis via computed tomography. *Prosthetics and Orthotics International*, 44(5), 298-304. doi: 10.1177/0309364620941297. PMID: 32693677.

Wu, H.-D., Liu, W. & Wong, M.-S. (2020b). Reliability and validity of lateral curvature assessments using clinical ultrasound for the patients with scoliosis: a systematic review. *European Spine Journal*, 29, 717-725. doi: 10.1007/s00586-019-06280-y.

Wu, H.-D., He, C., Chu, W. & Wong, M. S. (2021). Estimation of plane of maximum curvature for the patients with adolescent idiopathic scoliosis via a purpose-design computational method. *European Spine Journal*, 30. doi: 10.1007/s00586-020-06557-7.

Xu, K., Csapo, T. G., Feng, D. & Mi, H. (2020, Dec). Ultrasound tongue gestural sequence classification using convolutional auto-encoder and recurrent neural network. *Proceedings of the 12th International Seminar on Speech Production*.

Young, M., Hill, D. L., Zheng, R. & Lou, E. (2015). Reliability and accuracy of ultrasound measurements with and without the aid of previous radiographs in adolescent idiopathic scoliosis. *European Spine Journal*, 24, 1427–1433. doi: https://doi.org/10.1007/s00586-015-3855-8.

Z, Z., Q, Z., B, W., Y, Y., B, Q., Y, D. & Y, Q. (2011). Scoliotic posture as the initial symptom in adolescents with lumbar disc herniation: its curve pattern and natural history after lumbar discectomy. *BMC Musculoskelet Disord*, 12(216). doi: 10.1186/1471-2474-12-216.

Zagoruyko, S. & Komodakis, N. (2016). Wide residual networks. *CoRR*, abs/1605.07146.

Zeng, H., Ge, S.-H., Gao, Y.-C., Zhou, D.-S., Zhou, K., He, X.-M., Lou, E. & Zheng, R. (2021). Automatic segmentation of vertebral features on ultrasound spine images using stacked Hourglass network.

Zeng, H., Zhou, k., Ge, S., Gao, Y., Zhao, J., Gao, S. & Zheng, R. (2022). VertMatch: a semi-supervised framework for vertebral structure detection in 3D ultrasound volume.

Zhang, R., Xiao, X., Liu, Z., Li, Y. & Li, S. (2020). MRLN: Multi-Task Relational Learning Network for MRI Vertebral Localization, Identification, and Segmentation. *IEEE Journal of Biomedical and Health Informatics*, 24(10), 2902-2911. doi: 10.1109/JBHI.2020.2969084.

Zhao, X., Zhao, J., Sun, X.-J., Zhou, X.-H., Zhang, K. & Xie, Y.-Z. (2022). Optimizing Lumbar Pedicle Screw Trajectory Utilizing a 3D-Printed Drill Guide to Ensure Placement of Pedicle Screws Into Higher Density Bone May Improve Pedicle Screw Pullout Resistance. *World Neurosurgery*, 158, e459-e464. doi: https://doi.org/10.1016/j.wneu.2021.11.002.

Zheng, Y.-P., Lee, T. T.-Y., Lai, K. K.-L., Yip, B. H.-K., Zhou, G.-Q., Jiang, W.-W. & Lam, T.-P. (2016). A reliability and validity study for Scolioscan: a radiation-free scoliosis assessment system using 3D ultrasound imaging. *BioMed Central*, 11.

Zhong, Z., Zheng, L., Kang, G., Li, S. & Yang, Y. (2017). Random Erasing Data Augmentation. *CoRR*, abs/1708.04896.

Zhou, G., Jiang, W., Lai, K. & Zheng, Y. (2017). Automatic measurement of spine curvature on 3-D ultrasound volume projection image with phase features. *IEEE Transactions on Medical Imaging*, 36(6), 1250-1262. doi: 10.1109/TMI.2017.2674681.

Zhou, G.-Q. & Zheng, Y.-P. (2015, 10). Assessment of scoliosis using 3-D ultrasound volume projection imaging with automatic spine curvature detection. doi: 10.1109/ULTSYM.2015.0485.

Zhou, G.-Q., Li, D.-S., Zhou, P., Jiang, W.-W. & Zheng, Y.-P. (2020). Automating Spine Curvature Measurement in Volumetric Ultrasound via Adaptive Phase Features. *Ultrasound in Medicine & Biology*, 46(3), 828-841. doi: 10.1016/j.ultrasmedbio.2019.11.012.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N. & Liang, J. (2018). UNet++: a nested U-Net architecture for medical image segmentation. *CoRR*, abs/1807.10165.

Zhu, H., Quan, Q., Yao, Q., Liu, Z. & Zhou, S. K. (2023). UOD: Universal One-Shot Detection of Anatomical Landmarks. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pp. 24–34.

Zhu, Z., Zhao, Q., Wang, B., Yu, Y., Qian, B., Ding, Y. & Qiu, Y. (2011). Scoliotic posture as the initial symptom in adolescents with lumbar disc herniation: its curve pattern and natural history after lumbar discectomy. *BMC musculoskeletal disorders*, 12, 216. doi: 10.1186/1471-2474-12-216.