Prompt-Based Vertebral Segmentation Using a Generative AI Approach in OVCF Spinal Radiographs

Po-Kai Su[1], Pei-Rong Jiang[1], Kai-Xuan Xu[2], Meng-Lei Su[3,4], Jiann-Her Lin[3,4], Hsin-Han Chiang[5], Hsiao-Chi Li[2,*]
[1] Master of Science in Artificial Intelligence Technology, National Taipei University of Technology, Taipei, Taiwan
E-mail: t111c71007@ntut.org.tw, t112c52010@ntut.org.tw
[2] Department of Electrical Engineering, National Taipei University of Technology, Taipei, Taiwan
E-mail: t112318155@ntut.org.tw, hcli@mail.ntut.edu.tw
[3] Department of Neurosurgery, Taipei Medical University Hospital, Taipei, Taiwan
[4] School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
[5] Department of Vehicle Engineering, National Taipei University of Technology, Taipei, Taiwan

***Abstract*—Osteoporotic vertebral compression fractures (OVCFs) are prevalent among elderly patients, with X-ray imaging serving as the primary diagnostic tool. However, challenges such as organ obstruction and poor contrast after vertebroplasty procedures complicate vertebral segmentation in spinal X-rays.**

This research introduces an innovative generative AI framework for vertebral segmentation in spinal X-ray images, combining YOLO-based detection with prompt-driven segmentation inspired by the Segment Anything Model (SAM). The system generates bounding boxes around vertebrae as segmentation prompts and employs an interpolation strategy to address potentially missed compressed vertebrae. By incorporating domain-specific knowledge of vertebral anatomy via the interpolation strategy, the framework enables accurate delineation of vertebral structures in cases of compression fractures.

The model achieves impressive performance metrics, including a Dice coefficient of 0.9389 ± 0.0026, an IoU of 0.8854 ± 0.0045, and a sensitivity of 0.9436 ± 0.0062. Validation was conducted using data from 305 patients and 813 spinal X-ray images sourced from Taipei Medical Hospital (2014-2024), with training based on 164 patients (531 images, 2014-2019) and validation on 141 patients (282 images, 2020-2024). This generative AI application effectively addresses clinical challenges in vertebral segmentation for OVCF patients, potentially enhancing the accuracy of diagnoses and treatment planning.

***Index Terms*—Osteoporotic Vertebral Compression Fracture (OVCF), Generative AI, Medical Image Segmentation, Prompt-based Augmentation, Vertebral Segmentation.**

I. Introduction

X-ray imaging has become the gold standard for the diagnosis and follow-up of osteoporotic vertebral compression fractures (OVCFs). However, several challenges arise when dealing with X-ray images of poor quality. First, anatomical obstructions and vertebroplasty (VP) procedures often result in reduced image contrast, making interpretation more difficult. Second, OVCF patients typically have low bone mineral density, and following VP treatment, the affected vertebrae are more prone to deformation. These factors make it challenging for clinicians to accurately assess vertebral body height and identify secondary compression fractures.

Moreover, patient variability and the lack of large-scale datasets further complicate research in this area. Most publicly available datasets are based on CT or MRI, with limited X-ray data specific to OVCF cases. Due to Institutional Review Board (IRB) constraints and data privacy concerns, many researchers are unwilling or unable to share their datasets, leading to a lack of domain-specific benchmarks and open leaderboards. This limits the generalizability and reproducibility of models, especially for OVCF patients who have undergone VP procedures, where clinical presentation and imaging can vary significantly between individuals.

II. Related Work

Kim et al. [1] presented an early deep learning approach using heatmap-based PoseNet to identify lumbar vertebrae (L1-L5) and M-Net for segmentation, but suffered from false-positive keypoints and poor L5 localization. Cheng et al. [2] adopted a two-stage method with YOLOv4 [3] detection followed by ResU-Net segmentation and random forest classification [4].

Most OVCF studies focus on classification tasks rather than segmentation [5]–[7]. Precise postoperative vertebrae segmentation remains underexplored despite its importance for treatment monitoring. Among the limited research addressing postoperative scenarios, Kónya et al. [8] specifically investigated segmentation challenges in patients with surgical implants, including cages, screws, and other hardware. Their comparative analysis of semantic versus instance segmentation approaches demonstrated that instance segmentation achieved better recognition performance in complex postoperative environments, where metallic artifacts and anatomical distortion pose significant challenges for automated analysis.

Most segmentation studies focus on traditional overlap metrics such as dice coefficient scores [9] [1], though few explore

979-8-3315-7206-8/25/$31.00

clinically meaningful evaluation approaches. Several studies demonstrate the clinical utility through segmentation-derived metrics. For instance, Kim et al. [10] used segmentation to calculate the Vertebral Compression Ratio (VCR) with strong physician agreement, Horng et al. [11] applied a region-of-interest-based method with Residual U-Net [12] for Cobb angle estimation in scoliosis patients, and Lin et al. [13] proposed a multitask framework (MRNet) for joint segmentation and key spinal parameters prediction.

Despite these advances, no standardized metric adequately captures the inherent complexity of OVCF datasets, which vary significantly in patient demographics, fracture severity, and surgical implants. This heterogeneity challenges existing approaches, highlighting the need for robust algorithms capable of handling diverse clinical presentations.

Some studies address vertebral separation through postprocessing. Shi et al. [14] proposed AGNet with morphological operations, but it relies on fixed parameters that may not generalize across datasets with varying pixel spacing. And, rule-based morphological methods can potentially distort vertebral anatomical integrity. Zhang et al. [15] combined Transformer-based Segmentation with adaptive postprocessing to tackle vertebral block adhesion through geometric analysis of length-to-width ratios and boundary angles. While effective on the AASCE 2019 dataset, this approach still relies on hand-crafted geometric rules and careful threshold tuning. With over-aggressive postprocessing, it potentially compromises clinical interpretability.

A persistent limitation across many studies is inadequate handling of fused or overlapping vertebrae from congenital anomalies, degenerative changes, or surgical fusion. Current approaches use implicit learning rather than explicit strategies for these challenging scenarios. BiLuNet [9] improves boundary delineation via a multi-path decoder, while iterative segmentation frameworks refine vertebrae individually, but both remain vulnerable to anatomical abnormalities [16]. Peng et al. [17] leveraged instance segmentation to reduce postprocessing reliance, while Du et al. [18] proposed MFENet with multi-scale convolutional modules and channel attention for low-contrast scenarios, but lack specific overlap evaluation.

No consensus exists for detecting fused or overlapping vertebrae. Most approaches require postprocessing, which can compromise performance. Therefore, we pursue precise vertebral identification without extensive postprocessing interventions.

III. Method

Inspired by prior work demonstrating the superiority of instance segmentation over semantic segmentation in distinguishing overlapping vertebrae [8], the workflow of our proposed framework is illustrated in Fig 1. There are two stages, including Stage 1: Detection and Localization, and Stage 2: Precise Segmentation. YOLOv8 plays as the backbone for vertebral bounding box detection, facilitating the extraction of individual vertebral regions through bounding boxes. However, certain vertebrae are occasionally missed in challenging cases. To address this limitation, and taking into account the continuous and structured nature of the spinal curvature, we propose a missing vertebrae candidate generation strategy to recover missing vertebral detections by leveraging spatial regularity along the spine.

In the second stage, we utilize prompts to perform precise segmentation. Unlike other models that rely on cropping techniques, our approach benefits from the full contextual information of the entire image, allowing the model to capture a more comprehensive understanding of global structures. After generating the image embedding from the original X-ray image with ViT-B, the previously obtained bounding boxes are used as prompts, which are iteratively fed into the SAM-like model to generate vertebral segmentation masks. This iterative mechanism enables precise and controllable segmentation, making it particularly suitable for vertebral-level processing in OVCF patients.

A. Missing Vertebrae Candidate Generation Strategy

The strategy used to interpolate missing vertebral bounding boxes is based on analyzing the vertical distribution of the detected boxes and filling in the gaps when a box is likely missing. To determine a missing vertebra candidate, the height of each bounding box and the vertical spacing (gap) between each pair of consecutive bounding boxes are calculated. If the box-gap exceeds a threshold, defined as 0.5 times the median box height in this work, the vertebra is likely missing in that region.

To estimate the missing bounding box, an interpolated position is calculated based on the coordinates of the two consecutive boxes (B_i and B_{i+1}) along the vertical axis with each bounding box defined as $B_i = (x_{left}, y_{top}, x_{right}, y_{bottom})$. At this stage, our goal is to generate a coarse prompt to compensate for missing detections. Although the interpolated box may not precisely align with the actual vertebra, we assume it provides sufficient guidance for the subsequent vertebral segmentation task.

The pseudocode for the interpolation algorithm is shown in Algorithm 1.

B. Loss Function

The total training loss combines cross-entropy and Dice losses to balance pixel-wise classification accuracy and overlap quality:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}}(\hat{y}, y) + \mathcal{L}_{\text{Dice}}(\hat{y}, y) \tag{1}$$

where $\hat{y}$ and y represent the predicted and groundtruth segmentation masks, respectively.

IV. Experimental Result

A. Dataset Description

The dataset was retrospectively collected from Taipei Medical University Hospital over a 10-year period (2014–2024) and categorized into two phases according to imaging format and hospital system differences. A summary is presented in Table I.

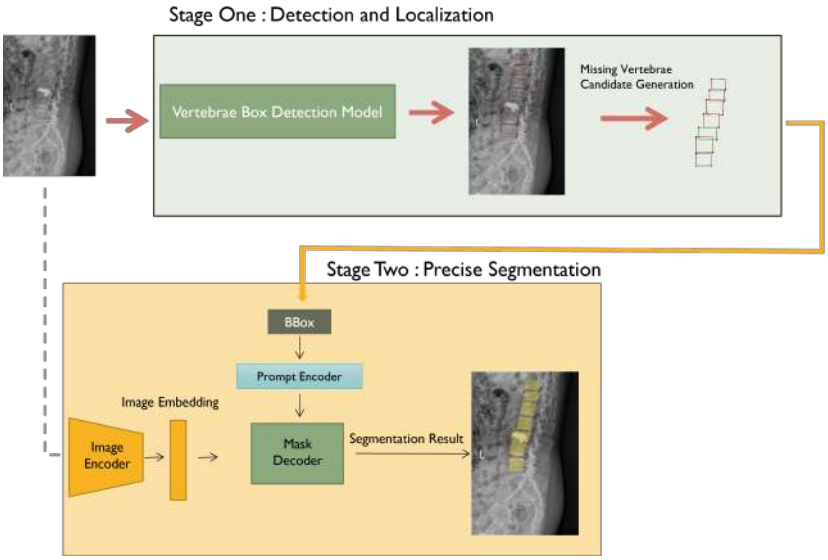

Fig. 1. Overview of the proposed iterative vertebrae segmentation pipeline.

Algorithm 1 Missing Vertebra Candidate Generation Strategy

Input: Sorted list of bounding boxes from top to bottom, i.e., B_1, ..., B_i, B_{i+1} where $B_i = (x^i_{left}, y^i_{top}, x^i_{right}, y^i_{bottom})$.
Output: List of original + interpolated bounding boxes.
for each pair of consecutive boxes B_i and B_{i+1} **do**
Compute the vertical gap

$$g \leftarrow y_{\text{top}}(B_{i+1}) - y_{\text{bottom}}(B_i)$$

if $g > 0.5\times$ median box-height **then**
$\tilde{x}_{\text{left}} \leftarrow \frac{x_{\text{left}}(B_i)+x_{\text{left}}(B_{i+1})}{2} - \delta$
$\tilde{x}_{\text{right}} \leftarrow \frac{x_{\text{right}}(B_i)+x_{\text{right}}(B_{i+1})}{2} + \delta$
$\tilde{y}_{\text{top}} \leftarrow y_{\text{bottom}}(B_i)$
$\tilde{y}_{\text{bottom}} \leftarrow y_{\text{top}}(B_{i+1})$
Add interpolated box:
$B_{\text{interp}} = (\tilde{x}_{\text{left}}, \tilde{y}_{\text{top}}, \tilde{x}_{\text{right}}, \tilde{y}_{\text{bottom}})$
end if
end for

Fig. 2. Developed annotation tool and sample annotation interface.

For image segmentation, we used LabelMe for polygon-based labeling with each structure annotated as either "Spine" or "Vertebroplasty", with an average of 20 points per polygon for higher precision. For vertebral height measurement, experienced medical doctors performed annotations using a dedicated web-based platform developed to address information security concerns. This platform records only measurement data while keeping images on local machines without online upload or storage.

For each vertebra, the doctor used six anatomical landmarks to measure body height (see Fig 2): upper left, lower left, bottom center, lower right, upper right, and top center, forming a hexagonal shape. The system automatically recorded data upon completion of each vertebra's annotation, with measurements saved in standardized JSON format and compressed into ZIP files for direct comparison with model predictions.

Phase I: Training Set

Phase I consisted of 531 BMP-format radiographic images collected from 164 patients (2014-2019), covering various clinical stages. All images were manually annotated for vertebral segmentation and used exclusively for model training.

Phase II: Testing Set

Phase II consisted of 282 DICOM-format radiographic images from 141 OVCF patients who underwent vertebroplasty

TABLE I
SUMMARY OF DATASET COMPOSITION

Phase	No. of Images	Annotated Vertebrae	Vertebroplasty (VP)
Training (Phase I)	531	4,759	651
Testing (Phase II)	282	2,537	342
Morphometric Subset	30	236	35

(2019-2024). These real-world clinical cases present higher complexity due to surgical artifacts, varying postures, and anatomical alterations. From 491 initial patients, images were selected based on: (1) DICOM data availability, (2) exclusion of complex surgical histories (hemilaminectomy, HIVD, ULBD, etc.), (3) postoperative imaging within 7 days, and (4) follow-up imaging within 6 months.

Morphometric Subset

A random subset of 30 patients from Phase II was selected for detailed morphometric analysis. Expert clinicians manually annotated 236 vertebrae to extract Anterior Body Height (ABH), Middle Body Height (MBH), and Posterior Body Height (PBH).

The complete dataset comprises 813 images and 7,296 vertebrae, with 993 vertebrae having undergone vertebroplasty (I). Phase I provides annotated training data with both normal and VP-treated vertebrae, Phase II includes complex post-vertebroplasty testing cases, and the morphometric subset enables quantitative shape analysis.

B. Evaluation Metrics and Clinical Correlation

Traditional metrics such as Dice score and Intersection over Union (IoU) often fail to evaluate a model's capability to accurately delineate individual vertebrae. To address this, we propose a modified Detection Rate metric inspired by object detection benchmarks [19]. A vertebra is considered "detected" if its IoU with groundtruth exceeds the threshold τ:

$$\text{Detection Rate}(\tau) = \frac{1}{N}\sum_{i=1}^{N}\begin{cases}1, & \text{if IoU}(V_i^{\text{pred}}, V_i^{\text{GT}}) > \tau \\ 0, & \text{otherwise}\end{cases} \tag{2}$$

where V_i^{pred} and V_i^{GT} denote the predicted and groundtruth segmentation masks of the i-th vertebra, respectively. This threshold accounts for cohesion issues where multiple vertebrae are incorrectly segmented as connected regions, causing reduced IoU with individual vertebrae and detection failure. This metric provides automatic, instance-based evaluation, particularly valuable for clinical applications where partial segmentation is insufficient.

For clinical validation, we conducted correlation analysis between automated measurements and physician assessments across ABH, MBH, and PBH using Mean Absolute Error (MAE), Intraclass Correlation Coefficient (ICC), Concordance Correlation Coefficient (CCC), and Pearson correlation coefficient.

C. One-Stage Segmentation Performance

We also compare our method against traditional one-stage segmentation approaches. Table II presents performance comparisons of various methods, including U-Net variants, DeepLabv3+, and YOLOv8.

Our method significantly outperforms traditional approaches across all metrics except sensitivity, where YOLOv8 achieves the highest score. The substantial improvement in detection rate (99.01% vs. 67.89-88.56% for other methods) demonstrates the superiority of our approach in accurately identifying and segmenting individual vertebrae.

D. Comparison with Two-Stage Segmentation Approaches

We evaluate our method against conventional two-stage pipelines combining object detection and segmentation. Following Cheng et al.'s methodology [2], we replace YOLOv4 with YOLOv8 for improved stability and fair comparison using the same detection backbone. After cropping bounding boxes, segmentation networks (U-Net [20], ResU-Net [4], DeepLabv3+ [21]) perform fine-grained segmentation within detected regions.

This pipeline helps segmentation models focus on relevant anatomical structures, reducing background noise and improving localization. Table III presents comprehensive comparisons across multiple evaluation metrics.

The results in Table III demonstrate that our proposed method achieves superior performance across most evaluation metrics. Specifically, our method attains the highest DICE score (0.9389 ± 0.0026), IoU (0.8854 ± 0.0045), accuracy (0.9901 ± 0.0004), sensitivity (0.9436 ± 0.0062), lowest average surface distance (0.8522 ± 0.0344), and highest detection rate (99.01 ± 0.24%).

E. Clinical Height Measurement Evaluation

To assess clinical applicability, we evaluated vertebral height measurements against physician expert assessments. Table IV presents correlation analysis between automated measurements from various two-stage models and expert groundtruth measurements.

The clinical evaluation results in Table IV clearly demonstrate the superior performance of our method compared to traditional U-Net-based approaches. Our method achieves the lowest Mean Absolute Error (MAE) and highest correlation coefficients across all three height measurements (ABH, MBH, PBH), indicating better agreement with physician assessments.

F. Qualitative Analysis

Fig. 3 provides a visual comparison of segmentation results from different two-stage methods. The qualitative analysis reveals that our method produces more accurate vertebral boundaries and better preserves anatomical structures compared to other approaches. The visualization demonstrates superior edge definition and reduced segmentation artifacts, which directly contribute to more accurate clinical measurements.

TABLE II
PERFORMANCE COMPARISON OF ONE-STAGE SEGMENTATION.

Method	DICE (↑)	IoU (↑)	ACC (↑)	SEN (↑)	SPE (↑)	ASD (↓)	DTR (↑)
U-Net [20]	0.9113 ± 0.0027	0.8390 ± 0.0044	0.9860 ± 0.0004	0.9004 ± 0.0090	0.9935 ± 0.0009	1.3704 ± 0.0455	88.5561 ± 0.8449
Res U-Net [4]	0.9054 ± 0.0036	0.8286 ± 0.0060	0.9846 ± 0.0007	0.9205 ± 0.0060	0.9902 ± 0.0012	1.5007 ± 0.0657	84.8930 ± 3.1287
DeepLabv3+ [21]	0.8969 ± 0.0014	0.8150 ± 0.0022	0.9831 ± 0.0003	0.9195 ± 0.0047	0.9887 ± 0.0006	1.6056 ± 0.0171	67.8922 ± 5.6217
YOLO v8 [22]	0.9025 ± 0.0018	0.8232 ± 0.0030	0.9831 ± 0.0004	**0.9675 ± 0.0022**	0.9845 ± 0.0006	1.5342 ± 0.1493	86.9786 ± 2.5476
Our method	**0.9389 ± 0.0026**	**0.8854 ± 0.0045**	**0.9901 ± 0.0004**	0.9436 ± 0.0062	**0.9942 ± 0.0007**	**0.8522 ± 0.0344**	**99.01 ± 0.24**

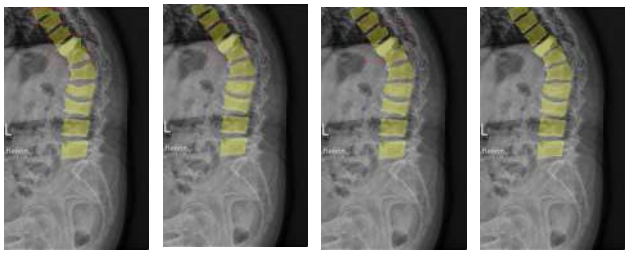

Fig. 3. Visual comparison of two-stage segmentation results for vertebral segmentation.

V. CONCLUSION

This study presents a novel generative AI framework for vertebral segmentation in OVCF X-ray images that addresses critical challenges in clinical spinal imaging. Our approach combines YOLO-based detection with prompt-driven segmentation inspired by the Segment Anything Model, enhanced by domain-specific adaptations. The framework achieves superior performance with a Dice coefficient of 0.9389 ± 0.0026, IoU of 0.8854 ± 0.0045, and detection rate of 99.01 ± 0.24%, outperforming both traditional and two-stage pipelines models. Importantly, our method demonstrates strong clinical validity through comprehensive correlation analysis with physician measurements across vertebral height parameters (ABH, MBH, PBH), with Pearson correlation coefficients ranging from 0.8111 to 0.9090. The key innovations for missing vertebrae candidate detection enable accurate segmentation even in challenging post-vertebroplasty cases where traditional methods often fail. The high detection rate and precise boundary delineation make this framework particularly suitable for monitoring compression fracture progression and treatment response in clinical workflows. This work represents a significant advancement in integrating foundation model concepts into medical image analysis, providing a pathway for developing robust, generalizable solutions for OVCF management. Future work should focus on prospective clinical validation and extension to other spinal pathologies.

ACKNOWLEDGMENTS

This work was supported by the National Science and Technology Council, Taiwan, under grant No. NSTC 113-2221-E-027-094.

REFERENCES

[1] K. C. Kim, H. C. Cho, T. J. Jang, J. M. Choi, and J. K. Seo, "Automatic detection and segmentation of lumbar vertebrae from x-ray images for compression fracture evaluation," *Computer Methods and Programs in Biomedicine*, vol. 200, p. 105 833, 2021, ISSN: 0169-2607.

[2] L.-W. Cheng, H.-H. Chou, Y.-X. Cai, *et al.*, "Automated detection of vertebral fractures from X-ray images: A novel machine learning model and survey of the field," en, *Neurocomputing*, vol. 566, p. 126 946, Jan. 2024, ISSN: 09252312.

[3] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[4] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94–114, 2020.

[5] S. Paik, J. Park, J. Y. Hong, and S. W. Han, "Deep learning application of vertebral compression fracture detection using mask r-cnn," *Scientific reports*, vol. 14, no. 1, p. 16 308, 2024.

[6] N. Hong, S. W. Cho, S. Shin, *et al.*, "Deep-learning-based detection of vertebral fracture and osteoporosis using lateral spine x-ray radiography," *Journal of Bone and Mineral Research*, vol. 38, no. 6, pp. 887–895, Apr. 2023, ISSN: 0884-0431. eprint: https://academic.oup.com/jbmr/article-pdf/38/6/887/55121936/jbmr4814.pdf.

[7] M. T. Löffler, A. Sekuboyina, A. Jacob, *et al.*, "A vertebral segmentation dataset with fracture grading," *Radiology: Artificial Intelligence*, vol. 2, no. 4, e190138, 2020.

TABLE III
PERFORMANCE COMPARISON OF TWO-STAGE SEGMENTATION.

Method	DICE (↑)	IoU (↑)	ACC (↑)	SEN (↑)	SPE (↑)	ASD (↓)	DTR (↑)
YOLOv8+U-Net [20]	0.9222 ± 0.0017	0.8566 ± 0.0029	0.9877 ± 0.0002	0.9082 ± 0.0079	0.9947 ± 0.0006	1.0787 ± 0.0219	94.53 ± 1.24
YOLOv8+DeepLabv3+ [21]	0.9222 ± 0.0019	0.8566 ± 0.0033	0.9877 ± 0.0003	0.9094 ± 0.0055	0.9946 ± 0.0004	1.0735 ± 0.0257	94.54 ± 0.97
YOLOv8+Res U-Net [4]	0.9226 ± 0.0032	0.8573 ± 0.0055	0.9878 ± 0.0004	0.9046 ± 0.0107	**0.9952 ± 0.0006**	1.0724 ± 0.0423	94.66 ± 0.77
Our method	**0.9389 ± 0.0026**	**0.8854 ± 0.0045**	**0.9901 ± 0.0004**	**0.9436 ± 0.0062**	0.9942 ± 0.0007	**0.8522 ± 0.0344**	**99.01 ± 0.24**

TABLE IV
EVALUATION OF DIFFERENT TWO-STAGE MODELS FOR MEASURING HEIGHTS WITH A PHYSICIAN EXPERT.

Method	Height	MAE (↓)	ICC (↑)	CCC (↑)	Pearson R (↑)	Pearson P
YOLOv8+U-Net [20]	ABH	2.8835	0.5790	0.5898	0.6645	< 0.0001
	MBH	2.5335	0.5876	0.5876	0.6458	< 0.0001
	PBH	4.0883	0.4121	0.4634	0.6377	< 0.0001
YOLOv8+DeepLabV3+ [21]	ABH	2.5289	0.6301	0.6356	0.6944	< 0.0001
	MBH	2.5285	0.5309	0.5314	0.5785	< 0.0001
	PBH	3.9368	0.4219	0.4692	0.6282	< 0.0001
YOLOv8+Res U-Net [4]	ABH	2.6269	0.5871	0.5937	0.6697	< 0.0001
	MBH	2.6070	0.5623	0.5645	0.6381	< 0.0001
	PBH	4.1188	0.3960	0.4464	0.6292	< 0.0001
Our Method	ABH	1.4181	0.8959	0.8963	0.9090	< 0.0001
	MBH	1.8540	0.7977	0.7980	0.8111	< 0.0001
	PBH	2.3408	0.7449	0.7595	0.8615	< 0.0001

[8] S. Kónya, T. S. Natarajan, H. Allouch, K. A. Nahleh, O. Y. Dogheim, and H. Boehm, "Convolutional neural network-based automated segmentation and labeling of the lumbar spine X-ray," en, *Journal of Craniovertebral Junction and Spine*, vol. 12, no. 2, pp. 136–143, Apr. 2021, ISSN: 0974-8237.

[9] V. L. Tran, H.-Y. Lin, H.-W. Liu, F.-J. Jang, and C.-H. Tseng, "Bilunet: A multi-path network for semantic segmentation on x-ray images," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 10 034–10 041.

[10] D. H. Kim, J. G. Jeong, Y. J. Kim, K. G. Kim, and J. Y. Jeon, "Automated vertebral segmentation and measurement of vertebral compression ratio based on deep learning in x-ray images," *Journal of digital imaging*, vol. 34, pp. 853–861, 2021.

[11] M.-H. Horng, C.-P. Kuok, M.-J. Fu, C.-J. Lin, and Y.-N. Sun, "Cobb angle measurement of spine from x-ray images using convolutional neural network," *Computational and mathematical methods in medicine*, vol. 2019, no. 1, p. 6 357 171, 2019.

[12] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.

[13] H.-Y. Lin, H.-W. Liu, *et al.*, "Multitask deep learning for segmentation and lumbosacral spine inspection," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022.

[14] W. Shi, T. Xu, H. Yang, *et al.*, "Attention gate based dual-pathway network for vertebra segmentation of x-ray spine images," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3976–3987, 2022.

[15] L. Zhang, J. Yang, D. Liu, *et al.*, "Spine x-ray image segmentation based on transformer and adaptive optimized postprocessing," in *2022 IEEE 2nd International Conference on Software Engineering and Artificial Intelligence (SEAI)*, IEEE, 2022, pp. 88–92.

[16] N. Lessmann, B. Van Ginneken, P. A. De Jong, and I. Išgum, "Iterative fully convolutional neural networks for automatic vertebra segmentation and identification," *Medical image analysis*, vol. 53, pp. 142–155, 2019.

[17] W. Peng, L. Li, L. Liang, *et al.*, "A convenient and stable vertebrae instance segmentation method for transforaminal endoscopic surgery planning," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 8, pp. 1263–1276, 2021.

[18] W. Du, Z. Liu, H. Fei, *et al.*, "Automatic segmentation of spine x-ray images based on multiscale feature enhancement network," *Medical Physics*, vol. 51, no. 10, pp. 7282–7294, 2024.

[19] T.-Y. Lin, M. Maire, S. Belongie, *et al.*, "Microsoft coco: Common objects in context," in *Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part v 13*, Springer, 2014, pp. 740–755.

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

[21] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *ECCV*, 2018.

[22] R. Varghese and S. M., "Yolov8: A novel object detection algorithm with enhanced performance and robustness," in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, 2024, pp. 1–6.