Lumbar Spine Localization in X-ray Images Using Object Detection with Advanced Bounding Box Techniques for Enhanced Medical Diagnostics

Sittisak Saechueng
Faculty of Informatics
Burapha University
Chonburi, Thailand
sittisak.sa@go.buu.ac.th

Ponlawat Chophuk*
Faculty of Informatics
Burapha University
Chonburi, Thailand
ponlawat.ch@go.buu.ac.th

2025 17th International Conference on Knowledge and Smart Technology (KST) | 979-8-3315-2040-3/25/$31.00 ©2025 IEEE | DOI: 10.1109/KST65016.2025.11003280

Abstract—**Accurate localization of lumbar spine corners in lateral X-ray images remains a challenging task due to poor image quality, overlapping anatomical structures, and noise. This paper presents an enhanced YOLOv5-based method that significantly improves detection accuracy by integrating advanced preprocessing, outlier elimination, and nearest neighbor refinement techniques. The proposed method achieves an average error of 3.51 mm, a substantial improvement compared to the 5.56 mm error of the competitor's model, highlighting its precision in detecting lumbar vertebrae corners. Additionally, the method attains an mAP of 97.7%, with precision, recall, and F1-scores of 96.7%, 96.6%, and 96.6%, respectively. These results demonstrate the method's precision and effectiveness in detecting lumbar vertebrae corners.**

Keywords- lumbar spine localization, x-ray imaging, object detection, advanced bounding box techniques

I. INTRODUCTION

The accurate localization of lumbar spine corners in lateral X-ray images plays a pivotal role in diagnosing conditions such as right laterolisthesis and left laterolisthesis[1]. These conditions require precise identification of the corners of lumbar vertebrae (L1-L5) for effective prediction and treatment. However, achieving such precision is a challenging task due to the inherent nature of X-ray images, which often suffer from poor image quality, overlapping anatomical structures, and the presence of artifacts resembling bone corners. These issues can result in the detection of outliers or false positives, leading to inaccurate diagnostic outcomes.

To address these challenges, past research has employed traditional image processing techniques, such as polyline simplification for edge segmentation [2], Otsu thresholding for intensity-based segmentation [3], and template matching for modeling intervertebral regions [4]. While these methods have shown promise in controlled scenarios, their effectiveness diminishes when applied to X-ray images due to the high levels of noise and ambiguity. Consequently, deep learning-based object detection methods, such as the Single Shot Multibox Detector (SSD) and Region-Based Fully Convolutional Network (R-FCN), have been introduced to enhance feature representation and localization accuracy [5-6]. Among the state-of-the-art object detection models, YOLOv5 has emerged as a preferred choice for lumbar spine localization due to its lightweight design, fast inference speed, and high accuracy. Comparative studies demonstrate that YOLOv5 outperforms models such as MobileNetV1, ResNet50V1, and EfficientDet D0, achieving superior mean average precision [7]. However, despite its advantages, YOLOv5 still faces challenges in eliminating false positives and accurately localizing vertebral corners critical for diagnosing lumbar spine-related conditions.

This study proposes an enhanced YOLOv5-based framework tailored for real-time and accurate localization of lumbar spine corners. The approach integrates robust preprocessing steps, advanced outlier elimination techniques, and refinement processes using nearest neighbor analysis to address the limitations of existing methods. The proposed method achieves notable advancements in the accurate localization of lumbar spine corners in lateral X-ray images, outperforming conventional models.

This paper is structured as follows: Section II analyzes the problem and challenges in lumbar spine localization. Section III outlines the proposed methodology. Section IV discusses the experimental results. Section V concludes with key findings and potential future research directions.

II. PROBLEM ANALYSIS

The application of YOLOv5 for lumbar spine corner detection in lateral X-ray images offers a fast and effective approach for identifying vertebral points. However, as shown in Figure 1, the detection process frequently includes incorrect points (red markers) alongside correct points (green markers). These inaccuracies stem from YOLOv5's reliance on recognizing visual patterns that resemble the target features, such as vertebral corners. In X-ray images, overlapping anatomical structures, noise, and artifacts that mimic true bone corners often lead to false positives. Consequently, while YOLOv5 effectively identifies many correct points, the presence of incorrect detections undermines its reliability for clinical

979-8-3315-2040-3/25/$31.00 ©2025 IEEE

applications. To address these challenges, advanced refinement techniques, including outlier elimination and contextual analysis, are essential to enhance detection accuracy and ensure the outputs are dependable for precise medical diagnoses.

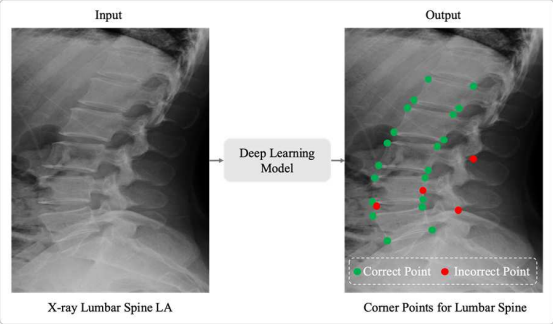

Fig. 1. Problem analysis of YOLOv5 in lumbar spine detection.

III. METHODOLOGY

The proposed methodology for accurately detecting corner points of the lumbar spine (L1 to L5) begins with lateral (LA) view X-ray images as input, as illustrated in Figure 2. During the preprocessing stage, the images are resized, and bounding boxes are created to facilitate object detection. The object detection phase identifies lumbar spine bounding boxes along with corner bounding boxes for the lumbar spine. In the postprocessing stage, outliers are eliminated to enhance accuracy, and nearest neighbor corner point detection is applied to precisely determine the corner locations. The final output provides accurate corner coordinates for the lumbar vertebrae, enabling more effective analysis and diagnosis.

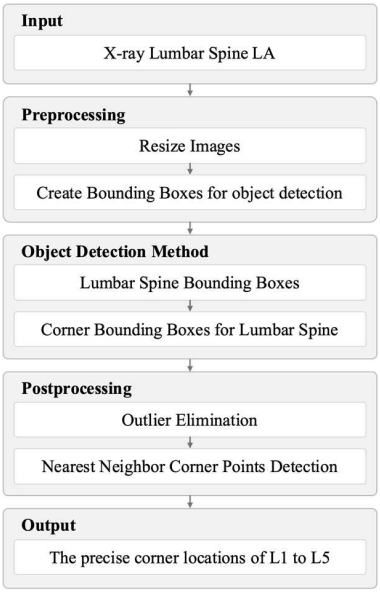

Fig. 2. A workflow of a proposed methodology.

A. Dataset

The BUU-LSPINE dataset includes lateral (LA) views of X-ray images, which are crucial for diagnosing spondylolisthesis as they reveal vertebral displacement in the sagittal plane, potentially affecting spinal canal compression or nerve damage. The dataset contains 3,600 LA images, all standardized to face the left side for consistency. Each image is annotated with detailed ground truth, including vertebral positions (corner points of L1–L5 and S1a), spondylolisthesis diagnosis (e.g., anterolisthesis and retrolisthesis), and Lumbosacral Transitional Vertebrae (LSTV) conditions. Patient ages range from 6 to 97 years, and the images were collected from 2000–2021 at Burapha University Hospital [7].

B. Preprocessing

The preprocessing stage consists of two main steps that are essential for preparing the input images for YOLO model training. First, the input images are resized to a standardized size of 640x640 pixels. This resizing ensures that all images in the dataset have consistent dimensions, which is critical for the YOLO model to process them effectively and achieve reliable predictions. The resized images maintain the structural integrity of the labeled regions, allowing the model to focus on the relevant features without being affected by variations in image dimensions.

Second, bounding boxes are created for object detection, involving two distinct sets. The first set, A, represents the bounding boxes for each of the five lumbar vertebrae, which are used to identify and isolate the lumbar spine from surrounding areas or noise, as defined in Equation (1). These bounding boxes help the model eliminate irrelevant objects and focus only on the regions of interest. The first set, E, consists of bounding boxes for the corner points of the lumbar vertebrae, as defined in Equation (2). These corner bounding boxes are used to precisely calculate the coordinates of each corner, enabling the detection of abnormalities such as spondylolisthesis. By completing these steps, the preprocessing stage ensures that the input data is structured, accurate, and optimized for the YOLO model to perform efficient and precise object detection and analysis.

$$A = \{F_1, F_2, F_3, F_4, F_5\}. \quad (1)$$

Here, $A, F_1, F_2, F_3, F_4, F_5$ represent the set of bounding boxes for the lumbar spine and the 1st to 5th lumbar vertebrae, respectively.

$$E = \{B_1, B_2, B_3, B_4, \ldots, B_{20}\}. \quad (2)$$

Here, $E, B_1, B_2, B_3, B_4, \ldots, B_{20}$ represent the set of corner bounding boxes for lumbar spine and the 1st to 20th corner of each lumbar vertebra , respectively.

C. Object Detection Method

The object detection method employs the YOLOv5 model to detect and generate bounding boxes for the corner points of the lumbar vertebrae. During testing on lateral X-ray images, the model effectively identifies the majority of corner points, demonstrating its capability to localize regions of interest. However, challenges arise due to the inherent complexities of

medical imaging. These challenges include overlapping anatomical structures, noise from surrounding tissues, and similar visual patterns in nearby regions, which can confuse the model. As a result, the output may include additional bounding boxes, referred to as outliers as shown in Figure 3, that do not correspond to actual corner points. This method identifies corner points corresponding to vertebrae from F_1 to F_5 with notable accuracy. In the image, green bounding boxes indicate correctly detected points, while red boxes represent incorrect detections or false positives. The model demonstrates strong performance in localizing regions of interest on lateral X-ray images. However, challenges arise due to complexities inherent in medical imaging, such as overlapping anatomical structures, noise from surrounding tissues, and visually similar patterns in nearby areas. These factors contribute to the generation of false-positive bounding boxes, marked in red in the figure, which do not correspond to actual vertebral corners. For instance, false detections may occur in regions where changes in bone density or non-vertebral structures resemble corner-like features, as seen with B_4, B_9, B_{13}, B_{17}, B_{18}, and B_{20}.

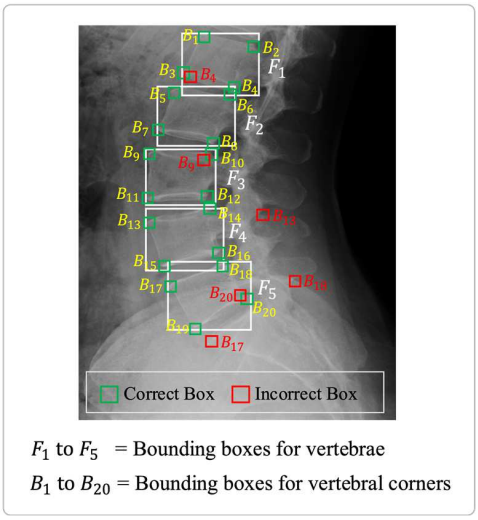

Fig. 3. Object detection method.

D. Postprocessing

In the postprocessing stage, various techniques are employed to refine the results from the initial detection process, ensuring accuracy and reliability. The centroids of bounding boxes are first calculated to pinpoint key features within the image. Subsequently, outliers within and outside the bounding boxes are removed using padding regions and distance equations to filter irrelevant data. Finally, the process delves deeper to identify the true corners of the lumbar spine by analyzing the refined centroids.

$$(x_c, y_c) = (x_{\text{norm}} \cdot W, y_{\text{norm}} \cdot H) \tag{3}$$

Here, (x_c, y_c) represent the center of the box in pixels, where $(x_{\text{norm}}, y_{\text{norm}})$ are the normalized coordinates provided by the model. and W, H are the width and height of the image in pixels.

The method described focuses on the removal of outside outliers in a postprocessing stage using padded bounding boxes. The centroids of the bounding boxes are determined using the calculation provided in Equation (3), and their validity is determined using mathematical constraints involving horizontal (P_x) and vertical (P_y) padding distances. These padding distances extend the bounding box boundaries to ensure robustness by accounting for minor deviations in centroid placement. Centroids falling outside the padded regions are classified as outside outliers and systematically removed. This process, as depicted in Figure 4 and Equation (4), retains only valid centroids within the extended boundaries (yellow dots) while eliminating outliers (cross markers).

$$(x_{min} - P_x \le x_c \le x_{max} + P_x) \wedge (y_{min} - P_y \le y_c \le y_{max} + P_y). \tag{4}$$

Here, $x_{min}, x_{max}, y_{min}, y_{max}$ are minimum and maximum coordinates of the bounding box. P_x and P_y are Horizontal and vertical padding distances.

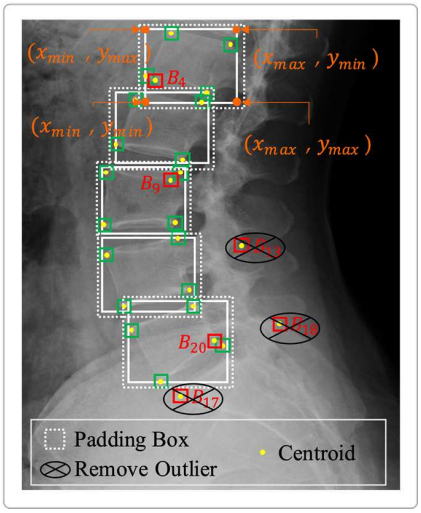

Fig. 4. External outlier removal using padded boxes.

The method for identifying the corner points of each vertebra (L1, L2, L3, L4, L5) is shown in Figure 5. This process involves calculating the Euclidean distance, as defined in Equation (5), between two key points: (1) the corner points (orange dots) of the white bounding boxes, which represent the corners of the vertebra region, and (2) the centroids of smaller yellow bounding boxes, which represent potential candidates for the correct corner points. For a yellow centroid to be valid, it must be located inside the corresponding white bounding box. Once the Euclidean distances are calculated for each candidate point, the selection of the correct corner point is based on two conditions. The first condition is to choose the point with the shortest distance to the centroid, ensuring that the selected point is spatially close to the true corner of the vertebra. The second condition is to verify that the selected point matches the target label, as specified in

Equation (6). This ensures that the identified corner point corresponds to the correct anatomical region or labeled structure of the vertebra. By applying these two conditions, the algorithm removes outliers and ensures accurate identification of the vertebra's corners for further analysis.

$$D_{i,j} = \sqrt{\left(x_{c_i} - x_{\text{corner},j}\right)^2 + \left(y_{c_i} - y_{\text{corner},j}\right)^2} \quad (5)$$

Here, $D_{i,j}$ represents the Euclidean distance between the center point of box i (x_{c_i}, y_{c_i}) and the j^{th} corner $(x_{\text{corner},j}, y_{\text{corner},j})$. i =1,2,…,20, representing the index of the boxes. j = 1, 2, 3, 4, representing the corners of the box: bottom-left corner (x_{min}, y_{min}), top-left corner (x_{min}, y_{max}), bottom-right corner (x_{max}, y_{min}), top-right corner (x_{max}, y_{max}).

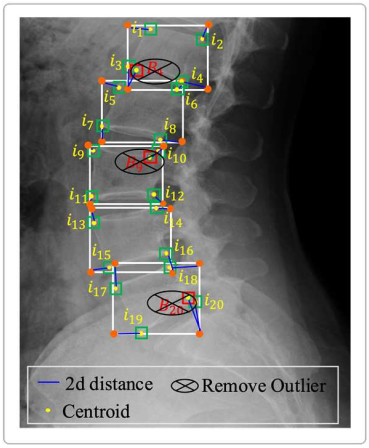

Fig. 5. Internal outlier removal using shorter distances.

$$i_* = \arg\min_i \{D_{i,j} \mid \text{Label}_i = \text{Target Label}\} \quad (6)$$

Here, i_* represents the index of the point i that minimizes the distance $D_{i,j}$, subject to the condition that the label of the point Label_i matches the target label.

IV. Result

The experimental results demonstrate that the YOLOv5 model proposed in this study outperforms previous work, achieving average values of mAP50, Precision, Recall, and F1-score at 0.977, 0.967, 0.966, and 0.966, respectively. This highlights the model's accuracy and effectiveness in detecting lumbar vertebrae features compared to prior studies. The proposed YOLOv5-based method achieves a notable advancement in lumbar spine corner detection, reducing the average error to 3.51 mm, outperforming the competitor's model at 5.56 mm, as shown in Table I. This significant improvement underscores the effectiveness of the enhanced preprocessing, outlier elimination, and refinement techniques. The dataset was divided into a training set and testing set in an 80:20 ratio for experimentation. Table II presents a detailed evaluation of the corner point extraction performance for the lumbar spine using the proposed YOLOv5-based method. It shows the average error in millimeters for each lumbar vertebra (L1 to L5), highlighting significant improvements in detection accuracy. The results demonstrate that the proposed model achieves lower errors compared to previous approaches. Notably, the smallest average error is 0.95 mm for L4, indicating high precision. However, the highest error is observed at the L1 vertebra, with an average error of 9.99 mm. This higher error rate is primarily due to certain areas of the X-ray images being unclear, particularly in the upper lumbar region, where image quality tends to degrade. Additionally, overlapping anatomical structures and artifacts that resemble vertebral corners further complicate accurate detection, leading to false positives. Despite these challenges, the overall results emphasize the model's effectiveness for accurate lumbar spine corner localization.

TABLE I. Corner Point Extraction Comparison

Ref.	View	Average Error (Millimeter)
[7]	LA	5.56
Proposed	LA	3.51

TABLE II. Lumbar Spine Corner Point Extraction

Lumbar spine	Corner	Average Error (Millimeter)
L1	i_1	8.62
	i_2	9.99
	i_3	4.30
	i_4	3.82
L2	i_5	5.83
	i_6	2.21
	i_7	5.26
	i_8	3.51
L3	i_9	2.50
	i_{10}	2.96
	i_{11}	1.22
	i_{12}	2.12
L4	i_{13}	3.48
	i_{14}	0.95
	i_{15}	2.21
	i_{16}	2.83
L5	i_{17}	2.31
	i_{18}	1.67
	i_{19}	2.61
	i_{20}	1.77

V. Conclusion

In this study, a refined YOLOv5-based framework for lumbar spine corner detection in lateral X-ray images was proposed, achieving superior performance over conventional methods. The approach integrated advanced preprocessing, outlier elimination, and refinement techniques using nearest neighbor analysis, addressing challenges like overlapping anatomical structures, noise, and false positives. Experimental results demonstrated an average error reduction to 3.51 mm compared to 5.56 mm by competing models, with

consistently high accuracy across all lumbar vertebrae (L1–L5). Future work could explore expanding this approach to other anatomical regions and improving real-time detection capabilities for broader medical applications.

ACKNOWLEDGMENT

This work was supported by the Faculty of Informatics, Burapha University, Chonburi, Thailand. This study was conducted under the approval of the Institutional Review Board (IRB), Approval No. HS024/2565(E3).

REFERENCES

[1] G. M. Trinh, H. C. Shao, K. L. C. Hsieh, C. Y. Lee, H. W. Liu, C. W. Lai, et al., "Detection of lumbar spondylolisthesis from X-ray images using deep learning network," *J. Clin. Med.*, vol. 11, no. 18, p. 5450, 2022.

[2] S. Ebrahimi, L. Gajny, W. Skalli, and E. Angelini, "Vertebral corners detection on sagittal X-rays based on shape modelling, random forest classifiers and dedicated visual features," *Comput. Methods Biomech. Biomed. Eng.: Imaging Vis.*, vol. 7, no. 2, pp. 132-144, 2019.

[3] M. Liaskos, M. A. Savelonas, P. A. Asvestas, M. G. Lykissas, and G. K. Matsopoulos, "Bimodal CT/MRI-based segmentation method for intervertebral disc boundary extraction," *Information*, vol. 11, no. 9, p. 448, 2020.

[4] M. Benjelloun, S. Mahmoudi, and M. A. Larhmam, "Template matching approaches applied to vertebra detection," in *Advances in Image Segmentation*, 2012.

[5] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, pp. 261-318, 2020.

[6] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," *Digit. Signal Process.*, vol. 126, p. 103514, 2022.

[7] P. Klinwichit, W. Yookwan, S. Limchareon, K. Chinnasarn, J. S. Jang, and A. Onuean, "BUU-LSPINE: A Thai open lumbar spine dataset for spondylolisthesis detection," *Appl. Sci.*, vol. 13, no. 15, p. 8646, 2023.