

Research on Auxiliary Diagnosis of Thoracic Fractures Based on Bidirectional Feature Pyramid Network

He Li, Xinyuan Zhao*

Xinjiang normal university; Urumqi; Xinjiang; 830000; China

Received: September 17, 2025

Revised: September 19, 2025

Accepted: September 21, 2025

Published online: September 21, 2025

To appear in: *International Journal of Advanced AI Applications*, Vol. 1, No. 7 (November 2025)

* Corresponding Author:
Xinyuan Zhao
(162100@qq.com)

Abstract. Thoracic fractures are a common occurrence in chest injuries, and traditional detection methods are prone to high rates of missed and false detections due to the complexity of thoracic fractures. However, detection methods based on deep learning technology have better detection performance and speed advantages. BiFPN enhances feature expression ability by dynamically adjusting the importance of each layer's features through bidirectional multiple fusion and utilizing a learnable weighting mechanism. This article uses a network model called YOLOv12+BiFPN for thoracic spine fracture detection, replacing the original feature fusion part in YOLOv12+BiFPN with the BiFPN structure. BiFPN allows multiple bidirectional information transfers between feature maps at different levels, improving the recognition ability of fracture areas of different sizes. It adds up-sampling and down-sampling paths to achieve bidirectional feature flow, designs weighted fusion nodes, and uses learnable weights to achieve adaptive feature combination. This module, with its efficient multi-scale feature fusion capability, not only significantly enhances the performance of object detection and classification, but also supports faster training and inference processes, demonstrating outstanding performance in processing large-scale image data. The improvement plan increased the mAP_{0.5} index from 0.8612 to 0.9354, and achieved a robust increase of 0.41% to 1.2% in mIoU, while the inference speed only slightly decreased by 3%. The YOLOv12+BiFPN network proposed in this paper greatly improves the accuracy of thoracic fracture detection, significantly promoting the improvement of diagnostic efficiency and precision.

Keywords: Thoracic vertebral fracture; Deep learning; YOLOv12; BiFPN model

1. Introduction

Thoracic fractures are among the most common types of chest injuries encountered in clinical practice, particularly in trauma care and emergency medicine [1]. These fractures can result from various causes such as high-energy impacts including traffic accidents, falls from heights, or blunt force trauma to the chest region. Timely and accurate diagnosis and appropriate management of thoracic fractures are critically important because delays or errors can lead to severe complications, including spinal cord injury [2-4], neurological deficits, chronic pain, and long-term functional impairment. Consequently, a thorough understanding of thoracic fracture characteristics and the implementation of effective diagnostic and therapeutic strategies are essential for improving patients' clinical outcomes and their quality of life. Historically, the primary modality for diagnosing thoracic fractures has been imaging, with computed tomography (CT) scans playing a vital role due to their high resolution and ability to generate cross-sectional images of the chest. CT imaging is considered the gold standard in assessing bony injuries of the thoracic spine because it offers excellent visualization of fracture details, displacement, and involvement of adjacent structures. Despite these advantages, CT scans generate a large volume of image slices — often hundreds per scan — which must be carefully reviewed by radiologists or orthopedic specialists. This manual review process is time-consuming and can be subject to human error, especially under conditions of heavy clinical workload or limited availability of experienced personnel. Moreover, the complexity of thoracic fracture presentations further complicates the diagnostic process. Some fractures may be subtle or share imaging features with other chest injuries such as rib fractures, lung contusions, or degenerative spinal changes [5]. This overlap increases the risk of missed or incorrect diagnoses, which can have serious implications for treatment planning. Misdiagnosis can delay the initiation of appropriate therapy, potentially worsening patient prognosis. In emergency settings, the need for rapid, reliable evaluation is even more pronounced, underscoring the limitations of relying solely on conventional image interpretation. With the rapid advancement of artificial intelligence and machine learning technologies, deep learning techniques have emerged as promising tools to address these challenges in medical image analysis. Deep learning models, especially convolutional neural networks (CNNs), have demonstrated remarkable performance in recognizing complex patterns and features within medical images that may not be easily discernible to the human eye [6, 7]. Applied to thoracic fracture detection, deep learning algorithms can assist physicians by automatically identifying and localizing fracture sites on CT images, thereby reducing the workload and potential for diagnostic error. One particularly

effective approach in this realm is the use of object detection algorithms tailored to medical imaging. The You Only Look Once (YOLO) family of models is among the leading frameworks for rapid, accurate object detection. Its real-time processing capability makes it especially suitable for emergency medical applications where time is critical. The integration of YOLOv12, an improved version with enhanced detection accuracy and speed, with advanced network architecture optimizations such as Bi-directional Feature Pyramid Network (BiFPN) further enhances the model's ability to capture features at multiple scales and improve the representation of fracture-related image details. The BiFPN architecture excels at fusing information from different levels of the neural network, allowing for more precise localization and classification of fractures within complex CT slice data. By combining YOLOv12 with BiFPN, the detection algorithm can efficiently process large volumes of thoracic CT images and provide automated preliminary diagnoses. This approach not only increases diagnostic accuracy by reducing false negatives and false positives but also accelerates the workflow by enabling real-time or near-real-time image interpretation. As a result, emergency and trauma care teams can make quicker, more informed decisions regarding patient management, potentially reducing morbidity associated with thoracic fractures. The application of deep learning-based detection systems also addresses broader systemic issues in healthcare. In many regions, the shortage of specialized radiologists and uneven distribution of medical resources results in delayed diagnosis and treatment, particularly in rural or underserved areas. Automated fracture detection tools can help bridge this gap by providing consistent, objective analysis irrespective of local expertise levels. This democratization of diagnostic capability can lead to more equitable healthcare delivery and alleviate pressure on overburdened healthcare professionals. Furthermore, the use of AI-assisted diagnostic tools can contribute to overall cost reduction in medical care. Faster and more accurate diagnoses reduce the need for repeat imaging studies and unnecessary hospital stays. They also enable timely interventions that mitigate complications, thereby decreasing long-term treatment costs. Hospitals benefit from streamlined workflows, and patients experience fewer delays, improved outcomes, and enhanced satisfaction with their care.

In conclusion, the integration of deep learning methods such as the YOLOv12 model combined with BiFPN optimization represents a significant advancement in thoracic spine fracture detection. This technology harnesses the power of artificial intelligence to transform traditional diagnostic practices, offering greater accuracy, speed, and reliability. As deep learning continues to evolve and become more embedded in clinical workflows, its role in improving medical imaging analysis and patient care is poised to grow increasingly influential.

The ongoing research and development in this area promise not only to enhance the management of thoracic fractures but also to provide a model for the application of AI across a broad spectrum of medical conditions, ultimately contributing to higher standards of healthcare quality and efficiency.

2. Methodology

2.1 ADD BIFPN (Bidirectional Feature Pyramid Network) module

The introduction of BiFPN (Bidirectional Feature Pyramid Network) module has significant advantages for thoracic spine fracture detection algorithm, which are reflected in the following aspects:

Efficient multi-scale feature fusion BiFPN effectively integrates feature maps of different levels through a bidirectional feature fusion mechanism, achieving complementarity between high-resolution detail information and high semantic contextual information. This is particularly crucial for the detection of lesions with diverse shapes and sizes, such as thoracic fractures, which can improve the ability to identify small fractures and complex structures.

Adaptive feature weighting. BiFPN adopts a learnable weight mechanism to dynamically adjust features at different scales, enabling the model to adaptively focus on feature layers that contribute more to the detection task, improving the specificity and effectiveness of feature expression, and thereby enhancing the accuracy of fracture localization and classification [8].

Balance between computational efficiency and performance. Compared to traditional feature fusion methods, BiFPN design has higher computational efficiency and can achieve faster feature transfer and fusion, which is particularly important for processing large amounts of CT slice data. It ensures detection speed without sacrificing detection accuracy and is suitable for real-time clinical application needs.

Enhance the model's generalization ability. Multi scale feature fusion enables the model to capture multi-level and multi angle lesion manifestations, which helps improve adaptability to individual differences in fracture manifestations among different patients, thereby enhancing the robustness and generalization ability of the overall model.

The introduction of the BiFPN module not only optimizes feature expression and information flow paths [9], but also substantially improves the detection performance and operational efficiency of the thoracic spine fracture detection model, making it an important component in enhancing the performance of deep learning fracture detection systems. The overall structure of the BiFPN module is shown in Figure 1.

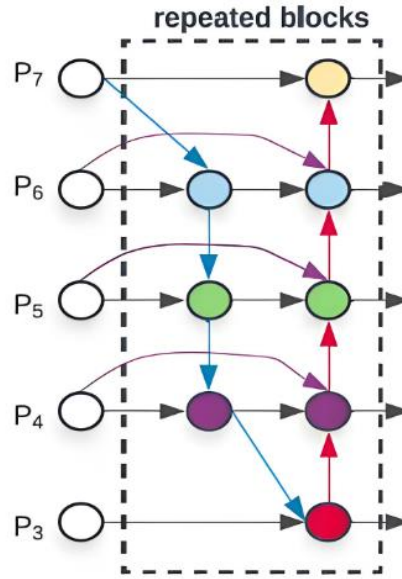


Figure 1. BIFPN (Bidirectional Feature Pyramid Network) module

2.2 Adaptive weight adjustment mechanism

The distribution of fracture areas of different scales is uneven in medical imaging, and there are significant morphological differences. Fixed weight fusion may lead to insufficient feature expression [10]. The adaptive weight mechanism dynamically provides weights for each layer and channel through model learning, optimizing the feature fusion process. Introducing learnable weight parameters into the weighted fusion nodes in BiFPN (usually through ReLU non negativity and normalization to ensure stability); Combining attention mechanism, adaptively adjusting feature response for channel and spatial dimensions; Gate Unit can be used to enhance the amplification of key features [11], suppress irrelevant noise, and dynamically highlight the key feature maps and channels for detecting fracture areas; Reduce noise interference and improve model localization and classification accuracy; More flexible adaptation to different patients and changes in imaging conditions [12].

3. Results

3.1 Dataset collection

In this study on thoracic spine fracture detection, the dataset used is the publicly available RibFrac database, which covers 400 patients with a total of 3321 thoracic spine fracture lesions. There are 319 male patients and 81 female patients, with a wide age distribution ranging from 4 to 79 years old. The dataset is divided into 280 fracture positive CT images for model training, 40 for validation (including 10 cases without fractures), and 80 unlabeled test CT images. In terms of image acquisition, the number of thoracic CT slices per patient ranges from 350 to 630,

with slice thickness controlled between 1mm and 1.25mm, ensuring high uniformity and stability of image data in spatial resolution. All training and validation images are accurately annotated by senior radiologists, covering the localization of fracture areas and four types of fracture classification labels based on clinical standards. The annotation quality is excellent, providing solid data support for the supervised learning of the model. In order to eliminate potential respiratory artifact interference, the data preprocessing stage strictly screened and removed the images, effectively improving the authenticity of the data and the accuracy of the experiment. Overall, this dataset has high quality, diversity, and clinical relevance, fully meeting the needs of training and evaluating intelligent detection algorithms for thoracic fractures in this study, and laying a solid foundation for model construction.

3.2 experimental setup

This chapter's experimental design revolves around three model structures: the baseline YOLOv12+BiFPN network, and two improvement schemes based on this architecture that introduce BiFPN module optimization and attention mechanism enhancement. Through systematic comparative experiments, comprehensively evaluate the improvement effect of each scheme on model performance, and ultimately select the attention module with the best performance as the final optimization strategy. This series of improvements aims to verify that the algorithm model proposed in this article has superior detection ability and generalization performance in thoracic spine fracture detection tasks. The experimental evaluation indicators include recall, precision, and the average accuracy (mAP_{0.5}) when the intersection to union ratio threshold is 0.5. mAP_{0.5} is the most core indicator, which directly reflects the overall detection accuracy of the model; The recall rate is the second highest, measuring the model's ability to detect fracture lesions; The accuracy ranks last and assists in evaluating the accuracy of the detection results [13]. The dataset used is a fracture CT image library constructed by ourselves in this article, which is reasonably divided into training set, validation set, and testing set to ensure the rigor and representativeness of the experiment. During the experiment, the model undergoes end-to-end training on the training set to obtain the optimal weight parameters, which are then used for performance validation on the test set. In terms of software and hardware environment configuration, the batch size for training is set to 16, the number of iterations is controlled at 200, and the initial learning rate is set to 0.001. By combining optimization algorithms and learning rate adjustment strategies, the model is ensured to converge stably and train efficiently. The overall experimental process is rigorous and the evaluation system is complete, fully verifying the effectiveness and practical value of the

proposed improvement plan.

3.3 experimental result

(1) Improvement of BIFPN module

Enhance the interactivity of feature pathways. By improving the feature flow mechanism, the combination between multi-level features is smoother and more comprehensive, promoting the deep integration of fine-grained features and high-level semantic information, and enhancing the model's ability to capture diverse fracture manifestations.

Table 1. Table of indicators before and after improving BIFPN module.

Evaluation	YOLOv12+BiFPN	After Improving The BIFPN Module
mAP_0.5	0.8900	0.8970
Precision	0.8520	0.8510
Recall	0.8910	0.9050

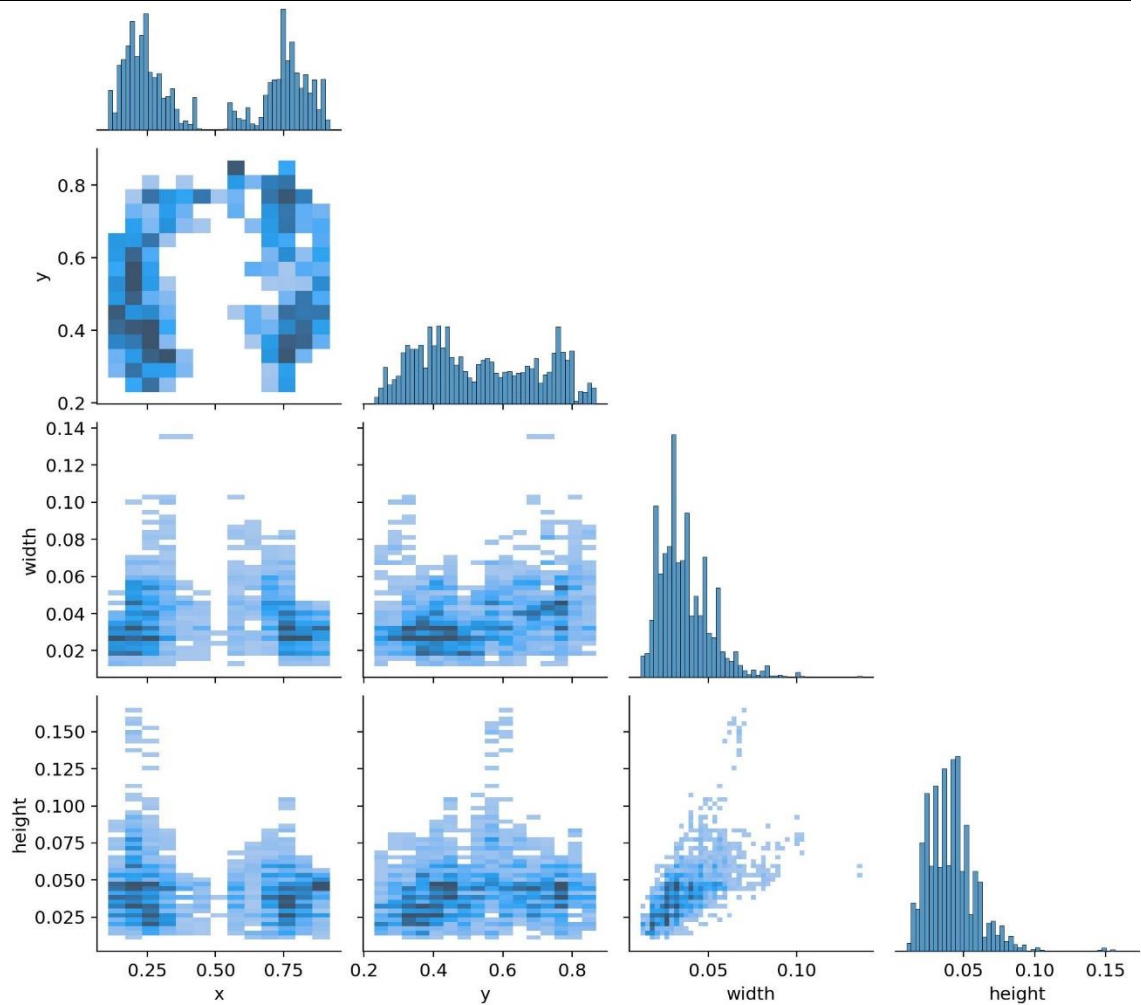


Figure 2. Pair plot (pairwise variable relationship diagram)

Implement dynamic weight adjustment. Introduce a more flexible weight allocation strategy to enable the model to automatically identify and strengthen key feature layers, weaken redundant or irrelevant information, thereby improving the accuracy and reliability of recognition.

Improve computational efficiency. Optimize the design of BiFPN structure, reduce redundant computation and transmission steps, so that the module can maintain efficient feature fusion while reducing computational resource consumption, improving the inference speed of the model, and meeting practical application requirements. The results of improving the BiFPN module are shown in Table 1.

Enhance the universality of the model. By using more reasonable feature fusion methods, the adaptability of the model under different case and image quality conditions can be improved, making it more robust and stable. The improvement of the BiFPN module aims to achieve refinement and efficiency in feature fusion, thereby promoting comprehensive optimization of detection models in terms of accuracy, execution efficiency, and generalization ability. The obtained pairwise variable relationship diagram is shown in Figure 2.

(2) Add attention mechanism

Adding attention mechanism, in computer vision tasks, attention mechanism is mainly applied in tasks such as image description generation, image classification, and object detection. Through attention mechanism, the model can dynamically select and focus on important areas in the image, thereby better understanding and processing the content of the image. Thereby improving the accuracy and robustness of object detection. In this experiment, different attention mechanisms will be added to the original YOLOv12+BiFPN network for comparative experiments, and the attention module with the best model promotion effect will be selected as the final improvement plan. The results of improving the add attention mechanism are shown in Table 2

Experimental data shows that both improvement strategies proposed in this paper exhibit significant performance improvements. It is worth noting that when multiple strategy combinations are used for optimization, the model performance exhibits a synergistic enhancement effect, and the improvement is significantly better than the effect of a single improvement strategy. This fully validates the effectiveness and complementarity of each improved module, and the resulting improved YOLOv12+BiFPN fusion model demonstrates excellent accuracy and clinical practical value in rib fracture detection tasks. From this, it can be seen that improving the network does indeed have a good effect on the model's prediction of

rib fractures. Verified the correctness and practicality of the improvement measures proposed in this article.

Table 2. Table of changes in various indicators before and after adding each attention module.

Moule	MAP_0.5	Precision	Recall
YOLOv12+BiFPN	0.8850	0.8520	0.8800
YOLOv12+BiFPN+GAM	0.8950	0.8620	0.8900
YOLOv12+BiFPN+SA	0.8890	0.8580	0.8850

4. Conclusion

Although the YOLOv12+BiFPN fusion model proposed in this article has achieved significant results in thoracic fracture detection tasks, there is still room for further improvement and expansion. Future work can be explored in depth from the following aspects:

(1) Multimodal data fusion.

Explore multimodal fusion strategies by combining medical imaging data (such as MRI, X-ray) or clinical auxiliary information other than CT images to enhance the model's ability to identify complex lesions and the comprehensiveness of diagnosis.

(2) Introduction of small sample and unsupervised learning methods.

In response to the scarcity of clinical annotated data, advanced technologies such as semi supervised, self supervised, or transfer learning are attempted to enhance the generalization ability and robustness of the model under limited annotated samples.

(3) Model lightweighting and accelerated optimization.

On the premise of ensuring detection accuracy, further optimize the model structure and inference efficiency to meet the real-time and deployable requirements of clinical rapid diagnosis, and promote the application of the model in practical medical equipment.

(4) Abnormal detection and rare pathological identification.

Expand the research on the recognition of rare and complex fracture types using extended models, combined with anomaly detection techniques, to improve the sensitivity and accuracy of the model in diverse pathological conditions.

(5) Clinical validation and cross center promotion.

Strengthen clinical trials with multiple centers, devices, and populations, evaluate the stability and applicability of the model, and lay a solid foundation for achieving widespread clinical applications.

With the continuous advancement of artificial intelligence technology and the increasing richness of medical imaging data, algorithm optimization and application expansion combined with clinical practical needs will provide more accurate and efficient technical support for intelligent diagnosis of fractures, promoting the greater value of intelligent auxiliary diagnosis systems in the field of orthopedics.

References

- [1] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.
- [2] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. [J]. CoRR, 2014, abs/1409.1556.
- [3] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [5] Ling F, Jingchen M, Yacheng R, et al. Automatic detection of lung nodules: false positive reduction using convolution neural networks and handcrafted features[C]. Medical Imaging, SPIE, 2017.
- [6] Zhiyi L, Yejie Z. Development paradigm of artificial intelligence in China from the perspective of digital economics [J]. Journal of Chinese Economic and Business Studies, 2022, 20 (2): 207-217.
- [7] Yudong Z, Jin H, Shuwen C. Medical Big Data and Artificial Intelligence for Healthcare [J]. Applied Sciences, 2023, 13 (6): 3745-3745.
- [8] Ahmed M R, Zhang Y, Liu Y, et al. Single Volume Image Generator and Deep Learning-based ASD Classification[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 24(11): 3044-3054.
- [9] Lin T H, Jhang J Y, Huang C R, et al. Deep Ensemble Feature Network for Gastric Section Classification[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 25(1): 77-87.
- [10] Zhao Y, Liu Y, Kan Y, et al. Spatial-Frequency Non-local Convolutional LSTM Network for pRCC Classification[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 22-30.
- [11] Li X, Shen L, Xie X, et al. Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection[J]. Artificial intelligence in medicine, 2020, 103: 101744.
- [12] Li Z, Zhang S, Zhang J, et al. MVP-Net: Multi-view FPN with position-aware attention for deep universal lesion detection[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 13-21.
- [13] Tao Q, Ge Z, Cai J, et al. Improving deep lesion detection using 3d contextual and spatial attention[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 185-193.