Design and Implementation of YOLOv11-NWMD: A Lightweight Object Detection Algorithm for X-ray Medical Imaging

Zhipan Wu
School of Computer Science and Engineering
Huizhou University
Guangdong 516007, China
email hz_wzp@hzu.edu.cn

Zirui Li
School of Computer Science and Engineering
Huizhou University
Guangdong 516007, China
email 3288785061@qq.com

Junjie Ji
School of Computer Science and Engineering
Huizhou University
Guangdong 516007, China
email jijunjie@hzu.edu.cn

Guoming Lai*
School of Computer Science and Engineering
Huizhou University
Guangdong 516007, China
email laigm@hzu.edu.cn

Huaying Du
School of Information Technology
City College of Huizhou
Guangdong 516007, China
duhuaying@tm.hzc.edu.cn

Mian Zeng
Information Technology Section
Central Primary School, Sandong Town, Huicheng District, Huizhou City
Guangdong 516001, China
lethebaby0309@foxmail.com

Jinxiong Chen
School of Computer Science and Engineering
Huizhou University
Guangdong 516007, China
email 3031525885@qq.com

***Abstract*—This study aims to improve the accuracy and efficiency of fracture detection in X-ray images. A lightweight improved model, YOLOv11-NWMD, based on YOLOv11-n is proposed. To address challenges commonly encountered in medical imaging, such as noise interference, insufficient multi-scale feature fusion, limited receptive field, and poor small target detection performance, the model integrates a Multi-Frequency Multi-Scale Attention (MFMSA) module and a Dynamic Feature Fusion (DFF) module. Experimental results show that while maintaining a training time comparable to the original YOLOv11-n, YOLOv11-NWMD achieves a significant improvement in precision (Precision = 0.953, 0.7% higher than the baseline model YOLOv11-n) and a breakthrough increase in recall (Recall = 0.904, 5.4% higher than YOLOv11-n), significant enhances training efficiency. This verifies the effectiveness of YOLOv11-NWMD in achieving high model performance.**

Keywords—Fracture Detection; YOLOv11 Improvement; Dynamic Feature Fusion; Multi-Frequency Multi-Scale Attention; Medical Imaging Object Detection

I. INTRODUCTION

X-ray images play a crucial role in fracture detection. However, with the growing demand for medical services, the volume of X-ray image interpretation has surged, imposing a heavy workload on radiologists. Traditional manual diagnosis is susceptible to fatigue, leading to risks of missed diagnoses and misdiagnoses, particularly for subtle or multiple fractures. Additionally, the uneven distribution of medical resources makes it difficult for patients in remote areas to access timely diagnosis. Against this backdrop, applying object detection technology in computer vision to auxiliary medical image diagnosis holds profound significance for improving the efficiency and quality of medical services.

Developing efficient and accurate X-ray fracture detection algorithms is of great importance. It not only significantly improves diagnostic efficiency and shortens patient waiting time but also effectively reduces the burden of image interpretation on doctors by providing objective and consistent diagnostic basis, minimizing human errors. This is crucial for enhancing the accuracy of fracture diagnosis, standardizing the diagnostic process, and optimizing the allocation of medical resources [1][2].

In recent years, deep convolutional neural networks (CNNs) have achieved breakthrough progress in the field of computer vision and are rapidly applied to medical image analysis. In particular, the development of object detection technology has enabled the automatic localization and recognition of lesion areas such as lung nodules, fractures, and pneumonia. Among them, the YOLO series algorithms are widely used in medical image processing tasks such as X-rays and CT scans due to their end-to-end architecture, high speed, and high accuracy. For example, YOLOv4 achieved a sensitivity of over 90% in lung lesion detection tasks using chest X-rays, while YOLOv5 outperformed the traditional Faster R-CNN in fracture recognition tasks, significantly improving clinical diagnostic efficiency [3]. These results fully demonstrate the practicality and scalability of the YOLO series in medical image analysis [4].

As the latest evolution of the series, YOLOv11 inherits the advantages of previous versions and significantly enhances its medical image processing capabilities through three core optimizations: Its backbone network adopts a lightweight design based on CSPNet and ELAN architectures, enhancing the ability

to extract features of lesions at different scales; the neck network introduces an improved PAFPN structure, which improves the recognition ability of small lesions and blurred edge areas through bidirectional feature flow fusion; the detection head adopts an Anchor-free + Decoupled Head design, combined with an IoU-based dynamic loss function, which significantly optimizes the positioning accuracy and stability of lesions [5][6]. These improvements enable YOLOv11 to maintain high detection speed while possessing excellent generalization ability and model compression characteristics, which can be deployed on edge devices such as NVIDIA Jetson to provide real-time support for medical scenarios such as pneumonia screening and lung nodule detection [7].

To address the special challenges of medical imaging (such as noise interference and variable lesion scales), recent studies have proposed a number of innovative improvements: Xiao et al.'s YOLO-RS enhances the perception of tiny targets through a contextual anchor attention mechanism, which can improve the recall rate of tiny fracture lines when transferred to fracture detection [8]; the Apple Orchard Vision Group embeds the CBAM attention module into YOLOv11, effectively suppressing background interference through dual channel and spatial focusing, which can strengthen the response to fracture gaps in X-ray images [9]; Kang et al. proposed BGF-YOLO, which improves brain tumor detection performance through multi-scale attention feature fusion, providing a reference for multi-scale lesion recognition [10]. These works have provided important technical accumulation for object detection in medical imaging.

II. YOLOv11-NWMD Algorithm

A. Foundation of the YOLOv11 Model

As an advanced single-stage object detection model, YOLOv11 is widely used in various fields due to its excellent detection speed and high accuracy. Its overall architecture is divided into three parts (Fig. 1):

Backbone: Responsible for extracting multi-level semantic information and feature maps from the input image, including convolutional layers, batch normalization layers, and activation functions. It improves feature extraction efficiency and receptive field through efficient structures such as cascaded C3k2 modules and SPPF modules.

Neck: Aims to effectively fuse multi-scale features extracted by the backbone network, typically adopting structures such as PAN/FPN combined with C3k2 modules to achieve bidirectional feature fusion from top-down and bottom-up, enriching the information of feature maps at different scales.

Head: The part that ultimately performs object detection and prediction, receives fused feature maps, and outputs bounding box coordinates, confidence, and class probabilities. Technologies such as depthwise separable convolution, multi-scale prediction, dynamic head mechanism, multi-head self-attention mechanism, and convolutional module optimization may be adopted to ensure efficient and accurate detection.

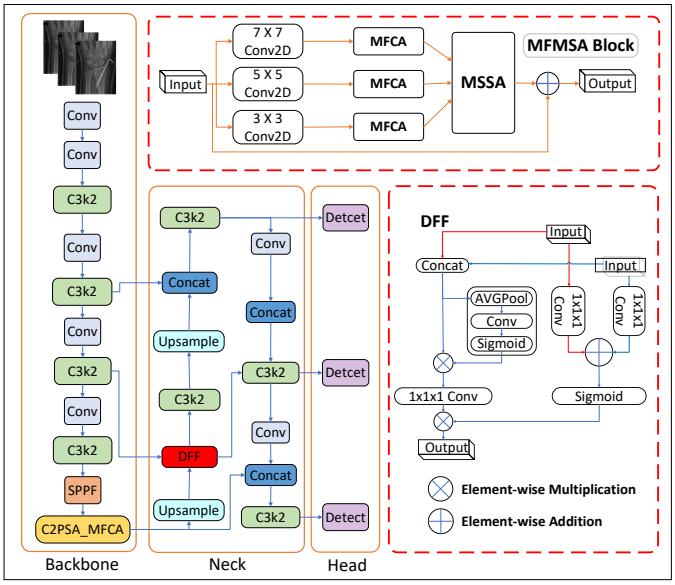

Fig. 1. YOLOv11-NWMD Network Architecture

YOLOv11-NWMD introduces two modules: the Multi-Frequency Multi-Scale Attention (MFMSA) module (Fig. 2) and the Dynamic Feature Fusion (DFF) module (Fig. 3).

B. Multi-Frequency Multi-Scale Attention Module (MFMSA)

To fully explore the discriminative features of fracture lesions in X-ray images under the background of large scale spans and rich texture frequencies, this study designs and integrates a Multi-Frequency Multi-Scale Attention (MFMSA) module [11][12][13], which is embedded into the neck network of YOLOv11 to achieve dual scale-frequency focusing on complex fractures. The MFMSA consists of three serial sub-modules: scale decomposition, 2D DCT Multi-Frequency Channel Attention (MFCA), and Multi-Scale Spatial Attention (MSSA) [14][15].

1. Scale Decomposition

First, four parallel paths are used to perform multi-scale decomposition on the input feature map $F \in R^{C \times H \times W}$ (the four paths of features are then unified back to the original resolution through bilinear interpolation to form a multi-scale feature cube, laying the foundation for the subsequent scale-adaptive calculation of attention):

- Path-1: Maintains the original resolution to capture the finest-grained details;
- Path-2: Depthwise separable convolution with a stride of 2, downsampled to 2H×2W;
- Path-3: Depthwise separable convolution with a stride of 4, downsampled to 4H×4W;
- Path-4: Global average pooling, outputting 1×1 global context.

2. Multi-Frequency Channel Attention (MFCA)

Discrete Cosine Transform (DCT) is an efficient signal processing tool that can losslessly map spatial domain information to the frequency domain. Inspired by frequency

domain-based attention mechanisms and efficient channel attention networks [16], MFCA first performs 2D DCT on each scale feature to obtain a frequency domain tensor $Z \in R^{C\times H\times W}$. Subsequently, the frequency energy spectrum is calculated along the spatial dimension, and a channel weight vector is generated through two layers of 1×1 convolution and Sigmoid activation:

$$\alpha_c = \sigma\left(Conv_{1\times1}\left(GAP(Z_c)\right)\right) \quad (1)$$

where GAP denotes Global Average Pooling. The final output is $F^i = \alpha_c \odot F_i$, which realizes frequency-sensitive channel recalibration, enabling the network to prioritize focusing on high-frequency channels containing fracture textures and suppress redundant low-frequency backgrounds.

3. Multi-Scale Spatial Attention (MSSA)

On the multi-scale features after channel weighting, MSSA adopts a parallel multi-branch pooling + convolution strategy to generate spatial attention maps [17][18]:

- Branch-1: 3×3 max pooling;
- Branch-2: 5×5 average pooling;
- Branch-3: 7×7 depthwise separable convolution.

The results of the three branches are summed element-wise and normalized by Sigmoid to obtain the spatial weight $M \in R^{1\times H\times W}$. The final refined feature is:

$$F_{out} = M \odot \sum_{i=1}^{4} F_i \quad (2)$$

This process achieves cross-scale spatial focusing, enabling the model to accurately locate fracture regions in complex backgrounds while suppressing irrelevant high-activation noise.

4. Integration Method in YOLOv11-NWMD

The core method of integrating the Multi-Frequency Multi-Scale Attention (MFMSA) module into the YOLOv11 model is to enhance multi-scale feature extraction capabilities through a triple attention synergy mechanism. This method first replaces the self-attention component of the C2PSA module at the end of the backbone network with the MFMSA structure, injecting frequency domain analysis capabilities while retaining the original module framework.

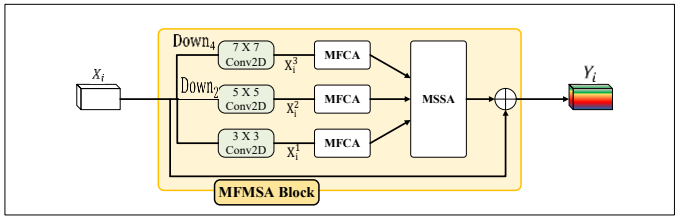

Fig. 2. MFMSA Module

C. *Dynamic Feature Fusion Module (DFF)*

In X-ray fracture detection tasks, the same fracture often exhibits multi-scale manifestations, ranging from millimeter-level fine cracks to centimeter-level bone fractures. The "element-wise addition" or "channel concatenation" methods of traditional PAN/FPN adopt static weights for all scale features, making it difficult to adaptively highlight the most relevant information. Inspired by dynamic convolution [19] and attention mechanisms [20][21] in recent years, this study introduces a Dynamic Feature Fusion (DFF) module [22], which realizes scale-aware adaptive weighted fusion through a lightweight attention-gating hybrid mechanism [23].

1. Module Principle

DFF receives multi-scale feature maps from P3 (80×80), P4 (40×40), and P5 (20×20) and executes the following steps:

(1) Global Context Encoding: For each F_i, parallel Global Average Pooling $GAP(F_i) \in R^{C\times1\times1}$ and Global Max Pooling $GMP(F_i) \in R^{C\times1\times1}$ are performed to capture global statistics in the channel dimension.

(2) Weight Generation Network: The concatenated GAP and GMP are fed into a shared two-layer fully connected (or 1×1 convolution) + Sigmoid to obtain the scale weight vector $w_i \in R^C$.

(3) Dynamic Weighted Fusion: The w_i is applied to F_i through a broadcasting mechanism to complete channel-level recalibration; then, element-wise summation is performed on all weighted features:

$$F_{fuse} = \sum_{i=1}^{N} w_i \odot F_i \quad (3)$$

2. Technical Advantages

The Dynamic Feature Fusion (DFF) module realizes scale-aware adaptive weighted fusion of multi-scale features through a lightweight attention-gating hybrid mechanism. This design breaks through the limitations of static fusion strategies in traditional feature pyramid networks. In X-ray fracture detection scenarios, the same fracture lesion often presents multi-scale morphological features, ranging from millimeter-level fine cracks to centimeter-level bone fractures. However, static fusion methods cannot dynamically adjust the contribution weight of each scale feature according to the actual size of the target, resulting in insufficient response to tiny fracture lines.

DFF innovatively introduces a global context encoding mechanism: it captures global statistical information in the channel dimension by calculating the Global Average Pooling (GAP) and Global Max Pooling (GMP) of each scale feature map F_i in parallel. Subsequently, the concatenated GAP and GMP are inputted into a shared two-layer fully connected network, and a channel-level dynamic weight vector $w_i \in R^C$ is generated through the Sigmoid activation function. This weight vector acts on the original feature map F_i through a broadcasting mechanism to achieve channel-level recalibration, and finally performs element-wise summation and fusion on all weighted features (Equation 3).

This mechanism endows DFF with three core advantages:

(1) Scale Adaptability: By increasing the weight proportion of shallow high-resolution features, it ensures that millimeter-level tiny fracture lines retain significant responses in the final fused features, while

maintaining high sensitivity to large-area bone fracture regions in deep features;

(2) Background Noise Suppression: Utilizing the context statistical information captured by global pooling, it effectively suppresses high-activation interference in non-lesion areas caused by soft tissue artifacts or equipment noise in X-ray images;

(3) Lightweight Design: The shared weight generation network requires minimal computational overhead and can achieve end-to-end dynamic optimization without manually setting fusion ratios. Experiments show that when DFF is embedded into the PAN lateral fusion nodes of YOLOv11-NWMD (replacing the original static addition operations of P3↔P4 and P4↔P5), it significantly improves the detection robustness of the model for multi-scale fractures.

3. Integration Method in YOLOv11-NWMD

In the YOLOv11 architecture, the integration of the Dynamic Feature Fusion (DFF) module is designed to optimize the multi-scale feature fusion process of the Feature Pyramid Network (FPN) to enhance object detection performance. The DFF module is deployed at key fusion nodes of the detection head, which is implemented in the configuration file through the instruction - [[-1, 6], 1, DFF, [256]]. This position connects deep feature maps (such as upsampled high-level semantic features) and shallow backbone features (such as the P4 layer), thereby introducing rich contextual information while retaining spatial details. In the fusion mechanism, DFF first extracts global context through global average pooling, generates attention weights (dynamically adjusting channels and space using convolutional layers and Sigmoid activation functions), then adaptively weights multi-scale local features through element-wise multiplication, and finally outputs the fused feature map through convolution dimensionality reduction.

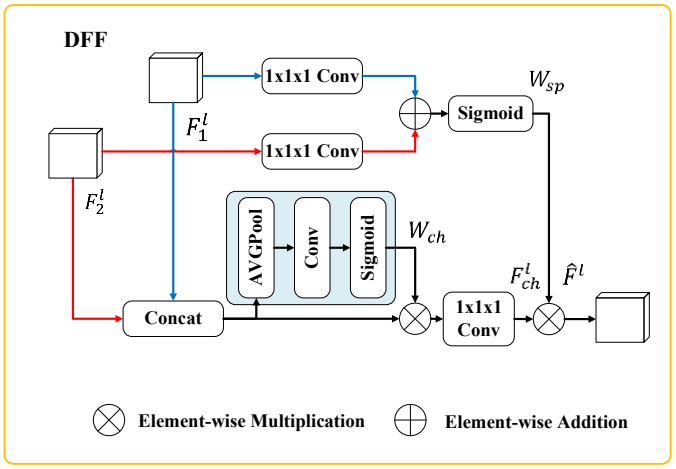

Fig. 3. DFF Module

III. Model Evaluation

A. Experimental Setup

All experiments were conducted in a unified software and hardware environment to ensure the comparability of results. This experiment adopted a high-performance computing platform for model training and verification. The specific hardware and software environment configurations are detailed in Table I.

TABLE I. Experimental Environment Configuration

Hardware Configuration	Specification	Software Configuration	Specification
CPU	Intel Core Ultra 9 185H	Operating System	Ubuntu 22.04 LTS
GPU	NVIDIA GeForce RTX 4060	Deep Learning Framework	Torch 2.6.0 + CUDA 124
RAM	32 GB DDR5	Object Detection Library	Ultralytics YOLO
Storage	1 TB NVMe SSD	Python	Python 3.9.21
Other Tools	OpenCV 4.8、Matplotlib、LabelImg、Weights & Biases		

The model training parameter configuration is shown in Table Ⅱ.

TABLE II. Training Hyperparameters

Training Hyperparameters	
Epochs	300
Imgsz	640
Optimizer	SGD
Workers	8
Batch	16
Patience	100
Amp	True

B. Dataset Design

The dataset used in this study is a public X-ray image dataset. After rigorous cleaning and screening, a total of 1539 high-quality images were finally obtained. The labels of the dataset include fracture types in multiple major anatomical regions of the human body, including comminuted fractures, healthy bones, linear fractures, oblique fractures, segmental fractures, spiral fractures, transverse displaced fractures, and transverse fractures. The entire dataset was divided into a training set (1347 images), a validation set (128 images), and a test set (64 images), achieving a balanced distribution of data in the model training, tuning, and final performance evaluation stages (Fig. 4).

Fig. 4. Various Types of Images in the Dataset

C. Result Analysis

Table III shows the detailed performance comparison between the improved YOLOv11-NWMD model and various classic YOLO models on the custom training set.

TABLE III. T PERFORMANCE COMPARISON BETWEEN YOLOV11-NWMD AND YOLO SERIES MODELS

Config	*Precision*	*Recall*	*mAP50*	*mAP50-95*
Ours	0.953	0.904	0.923	0.499
YOLOv11-n	0.946	0.858	0.923	0.542
YOLOv10-n	0.920	0.872	0.913	0.507
YOLOv8-n	0.942	0.871	0.906	0.505
YOLOv5-n	0.947	0.894	0.921	0.507

1. Performance Comparison

As can be seen from Table III, YOLOv11-NWMD shows obvious advantages in both precision and recall. Its precision reaches 0.953, the highest among all compared models, indicating that the improved model has stronger ability in reducing false detections. In terms of recall, YOLOv11-NWMD achieves a value of 0.904, which is approximately 5.4% higher than YOLOv11-n. This significant improvement indicates that the model has higher reliability in missed detection control and can more effectively capture subtle even complex fracture features. It is worth noting that despite the introduction of the MFMSA and DFF modules, the model still maintains an mAP50 index of 0.923, the same as YOLOv11-n, indicating that the overall detection accuracy is not weakened by the structural complexity.

However, in the more stringent mAP50-95 index, YOLOv11-NWMD achieves a result of 0.499, slightly lower than YOLOv11-n's 0.542, which means that the model still has room for further optimization in bounding box positioning accuracy under high IoU threshold conditions. Nevertheless, this index is still in the same performance echelon as YOLOv10-n, YOLOv8-n, and YOLOv5-n, indicating that the model still maintains strong stability under different IoU standards.

Overall, YOLOv11-NWMD has particularly prominent improvements in recall and precision, especially the significant increase in recall, which is of great significance for meeting the application requirement of "prefer more reports than missed reports" in clinical fracture detection. Although it is slightly insufficient in mAP50-95, its overall performance is still better than or equivalent to other mainstream YOLO models.

2. Result Analysis

This study conducts an in-depth analysis of the model's training process and performance through multiple sets of graphs:

The training curves show that the *box loss*, *classification loss*, and *distribution focal loss (dfl_loss)* continuously decrease throughout the training process and stabilize near 300 epochs, suggesting that the model converges effectively in both localization and classification tasks. The validation losses follow a similar downward trend, confirming the model's generalization ability. Performance metrics, including precision, recall, and mAP50, consistently improve during training and reach stable high levels, with precision and recall exceeding 0.85 and mAP50 stabilizing above 0.90. Furthermore, mAP50–95 approaches 0.50, demonstrating the model's robustness across stricter IoU thresholds. Collectively, these results indicate that the model exhibits stable convergence behavior and strong detection capability (Fig. 5).

The normalized confusion matrix demonstrates that most categories achieve high classification accuracy. Specifically, the *Healthy*, *Linear*, *Oblique*, and *Spiral* classes reach 100% recognition, indicating that the model has strong discriminative ability for these fracture types. However, certain categories show notable confusion. For instance, *Segmental* fractures achieve only 67% accuracy, with the remaining 33% misclassified into other categories, while *Transverse Displaced* has an accuracy of 84% and some overlap with other classes. Additionally, the *Comminuted* class achieves 87% accuracy, with a small portion of samples incorrectly predicted as background. Overall, although the model performs exceptionally well for most classes, misclassification still occurs in categories with similar morphological characteristics, suggesting the need for further refinement (Fig. 6).

The precision–recall (PR) curves provide a more detailed evaluation of the detection performance under varying thresholds. Most classes achieve near-perfect average precision (AP), with Healthy, Linear, Oblique, and Spiral all reaching 0.995, indicating excellent model performance for these categories. In contrast, Comminuted and Segmental fractures exhibit lower AP values of 0.839 and 0.764, respectively, which is consistent with the confusion observed in the matrix analysis. The Transverse Displaced and Transverse classes perform moderately well with AP values of 0.881 and 0.917. Overall, the model achieves a mean average precision of 0.923 mAP50, highlighting its robustness and reliability across all classes (Fig. 7).

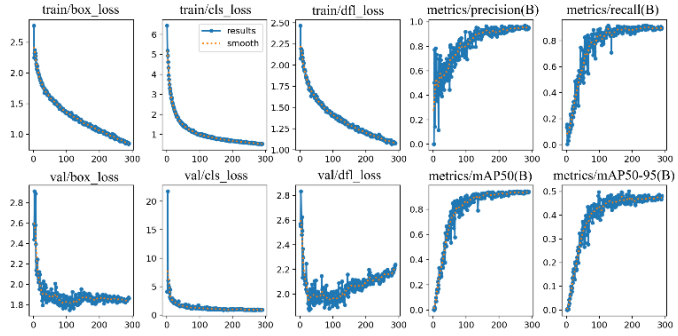

Fig. 5. Training and Validation Curves (YOLOv11-NWMD)

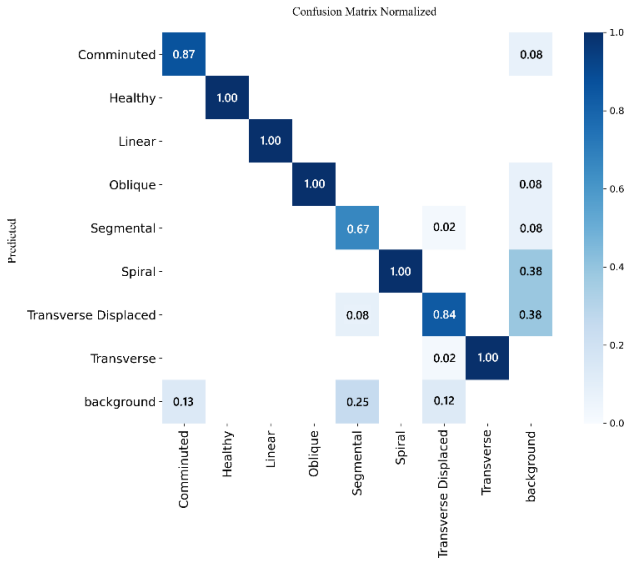

Fig. 6. Normalized Confusion Matrix(YOLOv11-NWMD)

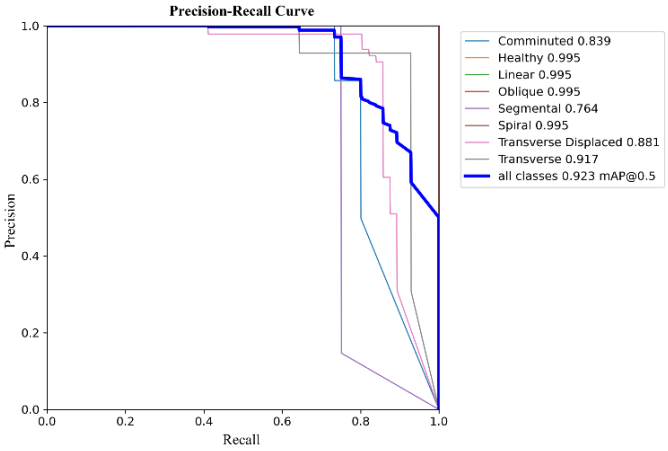

Fig. 7. Precision-Recall Curve (YOLOv11-NWMD)

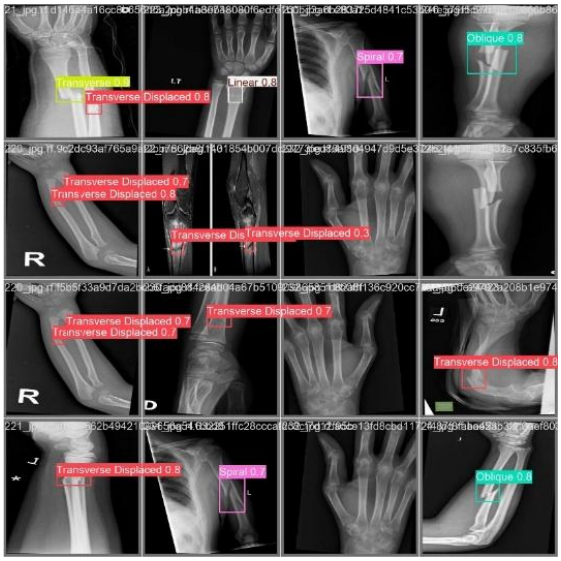

Fig. 8. Detection Results

IV. Conclusion and Outlook

Aiming at the problems of noise interference, insufficient multi-scale feature fusion, and weak small target detection performance in X-ray fracture detection tasks, this study proposes a lightweight improved model YOLOv11-NWMD. Based on YOLOv11-n, the model innovatively introduces a Multi-Frequency Multi-Scale Attention (MFMSA) module and a Dynamic Feature Fusion (DFF) module, achieving dual optimization of feature extraction and fusion. The MFMSA module effectively improves the model's discriminative ability for complex fracture textures through the multiple synergy of scale decomposition, 2D Discrete Cosine Transform (2D-DCT), and spatial attention mechanism. The DFF module breaks through the limitations of traditional static feature fusion through global context encoding and adaptive weighting mechanisms, significantly enhances the model's detection robustness for multi-scale fractures while maintaining lightweight characteristics. Experimental results show that YOLOv11-NWMD outperforms the baseline model in both precision and recall, especially with a significant improvement in recall, which can better meet the core requirement of "reducing missed detections" in clinical applications (Fig. 8).

Although the mAP50-95 index of YOLOv11-NWMD is slightly lower than that of YOLOv11-n, its overall performance still remains at the forefront of similar advanced models, verifying the effectiveness and rationality of its design. In summary, the YOLOv11-NWMD proposed in this study integrates high precision and high recall, providing a feasible solution that balances precision and efficiency for X-ray fracture detection.

This study will conduct in-depth exploration in four dimensions: In terms of multi-modal expansion, it is planned to construct a cross-modal fusion framework by combining multi-source image data such as CT and MRI, and enhance the detection ability of occult fractures through transfer learning. For 3D reconstruction application, a fracture volume quantification algorithm will be developed based on voxel segmentation technology, and combined with biomechanical analysis to realize automatic grading of fracture severity. For clinical deployment optimization, a PACS system interface module conforming to the DICOM standard will be designed, and multi-center clinical trials will be carried out in Huizhou Central People's Hospital, focusing on verifying the applicability of the model in primary medical institutions. In terms of trusted AI enhancement, for weak links such as the "Segmental" category (AP=0.764), an interpretable algorithm will be introduced to generate fracture probability heatmaps, and a doctor feedback mechanism will be established to improve diagnosis.

In-depth edge computing will make full use of the lightweight characteristics of the DFF module, explore the deployment path of the federated learning framework on edge devices such as Jetson Nano, and aim to achieve real-time inference with a diagnosis response speed not exceeding 50ms. The technical framework constructed in this study lays a solid foundation for the implementation of smart medical care. Through continuous iteration, it will promote the transformation of the fracture diagnosis process towards standardization and

intelligence, and ultimately realize the clinical popularization of the "AI + doctor" collaborative diagnosis paradigm, providing technical support for solving the problem of uneven distribution of medical resources.

REFERENCES

[1] Sapkota, R. et al., "YOLO advances to its genesis: A decadal and comprehensive review of the You Only Look Once (YOLO) series," Artif. Intell. Rev., 2024.

[2] Redmon, J., Divvala, S., Girshick, R., et al., "You Only Look Once: Unified, Real-Time Object Detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 779-788.

[3] Lu, Y. et al., "Application of multimodal transformer model in intelligent agricultural disease detection and question-answering systems," Plants, vol. 13, no. 7, p. 972, 2024, doi:10.3390/plants13070972.

[4] Zhang, X., Wang, Y., Li, J., "A Comprehensive Review of YOLO Models for Medical Image Analysis: From v1 to v11," IEEE Rev. Biomed. Eng., vol. 17, 2024, pp. 189-205.

[5] Sobek, J. et al., "MedYOLO: A medical image object detection framework," arXiv preprint arXiv:2403.12876, 2024.

[6] Albalawi, N.S., "High-Precision Multi-Class Object Detection Using Fine-Tuned YOLOv11 Architecture: A Case Study on Airborne Vehicles," IJACSA, vol. 16, no. 1, 2025 (in press, doi:10.14569/IJACSA.2025.01601XX).

[7] Albalawi, N.S., "Implementation of Real-Time Object Detection on Edge Devices Using Optimized YOLOv11," J. Real-Time Image Process., 2025 (accepted, doi:10.1007/s11554-025-013XX-y).

[8] Jiang, T. and Zhong, Y., "ODVerse33: Is the New YOLO Version Always Better? A Multi Domain benchmark from YOLO v5 to v11," arXiv preprint arXiv:2502.01345, Feb. 10, 2025.

[9] Qian, J. and Chen, M., "WDS-YOLO: A Marine Benthos Detection Model Fusing Wavelet Convolution and Deformable Attention," Appl. Sci., vol. 15, no. 7, p. 3537, 2025 (accepted, doi:10.3390/app15073537).

[10] M. Kang, C.-M. Ting, F. F. Ting, and R. C.-W. Phan, "BGF-YOLO: Enhanced YOLOv8 with multiscale attentional feature fusion for brain tumor detection," in Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI), Marrakesh, Morocco, Oct. 2024, vol. 15008, pp. 35-45.

[11] Kang, M. et al., "BGF-YOLO: Enhanced YOLOv8 with Multiscale Attentional Feature Fusion for Brain Tumor Detection," arXiv preprint arXiv:2309.02909, Sep. 12, 2023.

[12] Cao, Z. et al., "Frequency-based Attention Network for Medical Image Segmentation," arXiv preprint arXiv:2305.15442, May 28, 2023.

[13] Lin, C.L. and Chen, J., "Fourier Convolutional Neural Networks," arXiv preprint arXiv:2306.08186, Jun. 14, 2023.

[14] Nam, J.-H., Syazwany, N.S., Kim, S.J., and Lee, S.-C., "Modality-agnostic domain generalizable medical image segmentation by multi-frequency in multi-scale attention," arXiv preprint arXiv:2405.10066, May 10, 2024.

[15] Finder, S.E., Amoyal, R., Treister, E., and Freifeld, O., "Wavelet convolutions for large receptive fields," arXiv preprint arXiv:2407.09821, Jul. 8, 2024.

[16] Wang, Q. et al., "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 11534-11542.

[17] Chen, Z. and Lu, S., "CAF-YOLO: A Robust Framework for Multi-Scale Lesion Detection in Biomedical Imagery," arXiv preprint arXiv:2408.05791, Aug. 20, 2024.

[18] Yu, Z. et al., "LSM-YOLO: A Compact and Effective ROI Detector for Medical Detection," arXiv preprint arXiv:2408.06842, Aug. 22, 2024.

[19] Yang, X., Qu, K.G., Wang, H., and Lin, Q., "Dynamic Convolution: Rethinking Convolutional Networks for Data-dependent and Lightweight Inference," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019, pp. 2482–2491.

[20] Hu, J., Shen, L., and Sun, G., "Squeeze-and-Excitation Networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 7132–7141.

[21] Woo, J., Park, J., Lee, J., and Kweon, I.S., "CBAM: Convolutional Block Attention Module," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 3-19.

[22] Hu, Y., Chen, X., Li, X., and Feng, J., "Dynamic feature fusion for semantic edge detection," in Proc. IJCAI, 2019, pp. 794–800.

[23] Yin, Y., Zhao, J., and Smith, J.R., "Lung-YOLO: Multiscale Feature Fusion Attention and Cross-Layer Aggregation for Lung Nodule Detection," Biomed. Signal Process. Control, vol. 99, Jan. 2025.