ISSN:3068-9376

Frontiers in Preventive Medicine

DOI: 10.64904/fpm25013

Original Research

A Deep Learning-Based Model for Auxiliary Diagnosis of Pediatric Pneumonia Detection

Linping Du [1], Zihao Fan [2], Gang He [2]*

[1]School of Clinical Medicine, Guizhou Medical University, Guiyang 550001, Guizhou, China

[2]School of Art and Media, Nanning College of Technology, Nanning 530100, GuangXi, China

Abstract

Early diagnosis of pediatric pneumonia is critical for reducing mortality rates, and automated detection based on chest X-rays is key to improving diagnostic efficiency. While deep learning-based object detection algorithms have shown promise in this task, existing methods often suffer from limited generalization and detection risks due to model complexity or loss of detail during feature extraction and upsampling processes. In this paper, we propose an enhanced YOLOv12n model specifically designed for pediatric pneumonia detection. We introduce the DA2C2f module, which leverages multi-level feature processing and deep convolution operations to improve the model's ability to capture complex features at different levels, effectively enhancing classification performance. Additionally, the DySample module utilizes dynamic upsampling to adaptively adjust the sampling points based on the local content of the input feature map, overcoming the limitations of traditional fixed upsampling methods and reducing image blurring and detail loss. Experimental results demonstrate that the proposed improved model outperforms competing models across key metrics, including recall rate, mean average precision (mAP), and model generalization. Specifically, our model achieves a 1.7% improvement in recall rate (85.4%), a 0.4% improvement in mAP@0.5 (84.4%), and a 0.4% improvement in mAP@0.5-0.95 (52.9%). Notably, in bacterial pneumonia detection, our model exhibits the highest average precision (AP) of 86.8%, a 2.2% improvement over YOLOv12n. Ablation studies further validate the critical role of the DA2C2f and DySample modules in enhancing model performance, particularly in improving recall rate while maintaining high detection accuracy. The improved YOLOv12n model holds significant potential for automated pediatric pneumonia diagnosis and provides valuable insights for future medical image analysis applications.

Keywords: YOLOv12n, Pediatric Pneumonia, Medical Imaging, DA2C2f Module, DySample Module

1. Introduction

Pediatric pneumonia is one of the leading causes of child mortality globally, particularly in low-and middle-income countries, where its high incidence and rapid progression continue to pose significant public health challenges [1]. Early diagnosis and timely treatment are key to reducing mortality rates [2]. Clinically, chest X-ray remains a crucial diagnostic tool for pneumonia. However, due to the complexity of pediatric imaging features and variations in the experience and diagnostic standards of physicians, misdiagnosis and delayed treatment are common. This issue is even more pronounced in primary healthcare or resource-limited settings.

With the advancement of artificial intelligence,

* *Corresponding author: linpingdu123@163.com*

Lianping Du and Zihao Fan are co-first authors.

deep learning has increasingly become a focus in the field of medical image analysis. Deep Convolutional Neural Networks (CNNs) are capable of automatically extracting high-level features from images through multi-layer non-linear transformations, thereby reducing dependence on human expertise and achieving superior performance in image classification and detection tasks [3]. Applying this approach to pediatric pneumonia detection can enhance diagnostic efficiency and accuracy, alleviate the burden on healthcare professionals, and contribute to the development of medical intelligence. Against this backdrop, object detection algorithms like YOLO (You Only Look Once) [4-14] have garnered significant attention. Unlike traditional detection methods, the YOLO family of algorithms frames the object detection task as an end-to-end regression problem, enabling real-time detection while maintaining high accuracy. The advantages of YOLO include: on one hand, rapid processing of large-scale chest X-ray images, and on the other, the ability to directly output lesion locations and categories, offering more intuitive and reliable support for pediatric pneumonia diagnosis. With the introduction of improved versions such as YOLOv5 and YOLOv7, detection speed and accuracy have been further enhanced, opening new possibilities for small target recognition and complex scene analysis in medical imaging.

Despite significant contributions to pediatric pneumonia detection, existing methods still exhibit limitations. For example, Chen et al. [15] proposed a transfer learning approach, pre-training the DenseNet121 model on a large adult chest X-ray dataset and fine-tuning it on a pediatric pneumonia dataset annotated according to WHO standards. However, the model's performance significantly declines on images with low annotation consistency. Kundu et al. [16] proposed a deep learning model based on a weighted average ensemble strategy, combining pre-trained models such as GoogLeNet, ResNet-18, and DenseNet-121, and dynamically adjusting weights based on four evaluation metrics (accuracy, recall, F1 score, and AUC) to improve the classification performance of pneumonia X-rays. However, its performance on the RSNA dataset was relatively low, indicating limited generalization ability across different data distributions and image qualities. Radočaj et al. [17] proposed using four CNN architectures (InceptionV3, InceptionResNetV2, DenseNet201, and MobileNetV2), combined with three convolution strategies (standard, multi-scale, and dilated convolution) and the Mish activation function, to differentiate bacterial, viral, and normal pediatric chest X-rays. However, the model's generalization was impacted by reliance on a single public dataset, potentially introducing image quality and population distribution biases. Habib et al. [18] proposed an ensemble method using fine-tuned CheXNet and VGG-19 CNNs to extract chest X-ray features, combined with data balancing techniques like random oversampling (ROS) to address class imbalance, and a random forest classifier for pneumonia classification. However, this method is computationally expensive and depends on multiple pre-trained networks and subsequent machine learning classifiers. Yao et al. [19] proposed a two-stage multimodal deep learning model (AMPNet) using X-ray and blood test data, leveraging a global-local attention mechanism for enhanced feature extraction and addressing class imbalance through a two-stage training strategy. However, the model's performance in differentiating viral from bacterial pneumonia was still suboptimal. Kaya et al. [20] proposed an ensemble CNN framework based on feature fusion, using hierarchical template matching for lung region segmentation, Chi-Square and mRMR for feature selection, and multiple machine learning classifiers for pneumonia detection. However, the model was only validated on a single public dataset (Kermany) and lacked verification on multi-source or real clinical data, limiting its generalization. Xin et al. [21] trained a deep learning model based on ResNet-50 using 5232 pediatric chest X-rays from the Guangzhou Women and Children's Medical Center for pneumonia classification (pneumonia/normal) but found a significant decline

in performance on external test datasets. Arun Prakash et al. [22] proposed a stacked ensemble learning method based on multi-model deep CNNs (MobileNet, DenseNet121, DenseNet169, DenseNet201), using CLAHE for image enhancement and multiple machine learning classifiers for final classification. However, the model's performance heavily depended on the choice of base CNN architectures and classifiers, limiting its generalization ability. Prakash et al. [23] proposed a pediatric pneumonia diagnosis method based on Xception network feature extraction and stacked ensemble learning, with Kernel PCA for dimensionality reduction and multiple machine learning classifiers for final classification, using stratified K-fold cross-validation to prevent overfitting. However, the model's performance was highly dependent on the combination of base classifiers, and its generalization ability was limited. Bal et al. [24] proposed a hybrid method for pediatric pneumonia detection based on deep learning feature extraction, using pre-trained CNN models combined with traditional machine learning classifiers. However, due to the small dataset size and limited test set, the model's generalization ability and reliability may be compromised.

Existing methods often face challenges in feature extraction and upsampling, leading to insufficient generalization and detection risks, particularly when handling complex and variable pediatric chest X-ray images. To address these issues, we propose an improved YOLOv12n model, specifically optimized for pediatric pneumonia detection. The main contributions of this paper include:

We propose an improved YOLOv12n model, specifically optimized for pediatric pneumonia detection, which significantly enhances detection performance while maintaining model lightweight.

We introduce the DA2C2f module, which improves the model's feature representation ability by utilizing multi-level feature fusion and deep convolution, effectively enhancing the model's ability to capture complex pneumonia features.

We introduce the DySample dynamic upsampling module, replacing traditional fixed upsampling methods, achieving content-aware feature map reconstruction, and significantly reducing detail loss and image blurring.

2. Method design

This study proposes an improved YOLOv12n [25] model for the automated detection of pediatric pneumonia. To address the limitations of existing methods in handling complex image features, we introduce two key modules: the DA2C2f module and the DySample module, both of which significantly enhance the model's performance. The DA2C2f module strengthens the model's feature representation capability through multi-level feature fusion and deep convolution operations, excelling particularly in capturing complex pneumonia characteristics. On the other hand, the DySample module employs dynamic upsampling, adaptively adjusting the sampling points based on the local content of the feature map, thereby reducing the loss of detail and blurring commonly encountered with traditional upsampling methods. This results in improved accuracy in image reconstruction.

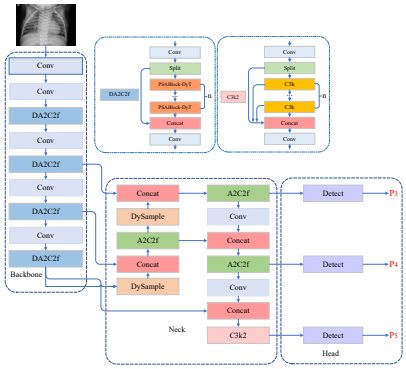


Figure 1. Schematic diagram of the improved YOLOv12n network architecture.

2.1 DA2C2f Module

To improve the performance of the YOLOv12n model in pediatric pneumonia detection, we introduce the DA2C2f module. Traditional YOLO models often face challenges when processing medical images, such as insufficient feature extraction and loss of details, which can lead to suboptimal detection accuracy. The DA2C2f module addresses these issues by enhancing the model's ability to capture complex pneumonia features through multi-level feature processing and deep convolution operations. This approach allows the model to better recognize fine details within pneumonia images, thereby improving detection accuracy.

As illustrated in Figure 2, the structure of the DA2C2f module is designed to improve the model's classification performance through multi-level feature processing. The input feature map first undergoes an initial feature extraction through a convolutional layer (Conv). This layer captures low-level features, such as edges and textures, which are crucial for subsequent processing. Next, the processed feature map passes through a series of Ablock-DYT modules, which are primarily responsible for further feature extraction and information enhancement. The design of the Ablock-DYT module combines deep convolution operations with dynamic adjustment mechanisms, enabling the model to capture more complex features at different levels and optimize them as needed. After passing through the Ablock-DYT [26] module, the processed data undergoes refinement through another convolutional layer, which strengthens the final feature representation. Finally, the network utilizes an Add operation to combine the output of the second convolutional layer with the earlier feature maps, resulting in the generation of a new output feature map.

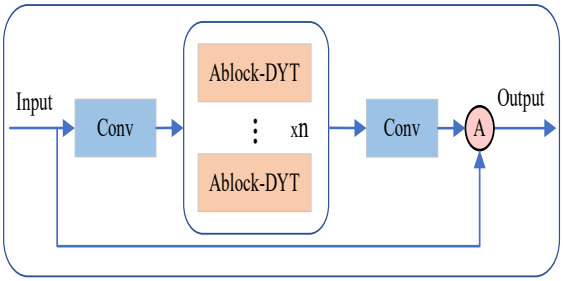


Figure 2. DA2C2f Module.

2.2 DySample Module

The DySample Module [27] is a dynamic upsampling method designed to enhance the quality of feature map reconstruction. The core concept of this module is to dynamically learn the sampling point coordinates within the input feature map, enabling content-aware upsampling operations. Unlike traditional fixed upsampling methods, such as bilinear interpolation or nearest-neighbor interpolation, DySample adjusts the sampling point locations based on the local content of the feature map. This dynamic adjustment allows the model to better preserve fine details and reduce blurring, ultimately improving the accuracy and clarity of the reconstructed feature map.

As illustrated in Figure 3, the input feature map X with a size of $H \times W \times C$ first passes through a linear layer, producing a new feature map O with a size of $H \times W \times 2gs^2$. This is followed by a pixel shuffle operation, which rearranges the feature map to yield a new tensor of size $sH \times sW \times 2g$, denoted as O. Concurrently, the original sampling grid G is element-wise added to the offset O, generating the final sampling set S. Finally, the grid sample operation is applied to the original input S using the obtained sampling grid X, where features are sampled from X according to the respective coordinates, ultimately generating the output feature map X' with a size of $sH \times sW \times C$.

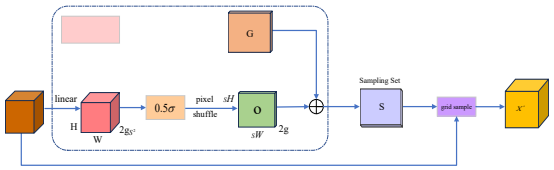


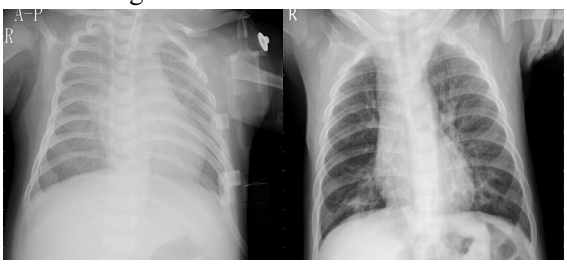
Figure 3. DySample Module

3. Experiment

3.1 Dataset

The pneumonia medical image dataset used in this study is sourced from the publicly available Chest X-ray Pneumonia Dataset [28], curated and released by Kermany et al. This dataset is widely used in deep learning research for pneumonia detection and classification. To ensure the quality of the data and minimize the impact of low-quality samples on model training and evaluation, we conducted a rigorous selection process, excluding images with low

resolution, blurring, or missing lesion areas. During the annotation process, we utilized the X-AnyLabeling tool to precisely label the target regions and categorized the images into two classes based on pathological types: (a) Bacterial Pneumonia and (b) Viral Pneumonia. In total, 2611 images were collected, and the data was split into training and testing sets at a 7:3 ratio, with the training set containing 1827 images and the testing set containing 784 images. Examples of pneumonia sample images are shown in Figure 4.



(a) (b)

Figure 4. Examples of pneumonia sample images.

3.2 Experimental Platform

The experiments in this study were conducted in the following hardware and software environment: For the deep learning tasks, we utilized the PyTorch 1.10.0 framework with Python 3.8 as the programming language, running on the Ubuntu 20.04 operating system. GPU acceleration was implemented through CUDA 11.3. In terms of hardware, the training process was carried out using a single NVIDIA RTX 4090 GPU (24GB), paired with an AMD EPYC 7T83 64-core processor (22 virtual cores) and 90GB of RAM. This configuration provided sufficient computational power and storage resources, enabling efficient model training and rapid iteration.

3.3 Hyperparameter Settings

During the model training phase, the input image resolution was standardized to 640×640 to ensure consistent input dimensions across all samples. The batch size was set to 64 to maximize the parallel computing capabilities of the GPU, thereby accelerating the training process. A total of 200 epochs were conducted, with the initial learning rate set to 0.01. To enhance the efficiency of data loading and preprocessing, 8 worker threads were enabled during data loading. For the optimizer configuration, the momentum coefficient was set to 0.937, and the weight decay coefficient was set to 0.0005. This configuration not only facilitated faster convergence but also helped mitigate the risk of overfitting by providing regularization.

3.4 Evaluation Indicators

In this study, the evaluation metrics used include F1 score, precision (P), recall (R), average precision (AP), and mean average precision (mAP) [29]. Additionally, the number of parameters (Parameters) was also considered. The formulas for these metrics are as follows:

$$\text{Precision} = \frac{T_p}{T_p + F_p}, \tag{1}$$

$$\text{Recall} = \frac{T_p}{T_p + F_N}, \tag{2}$$

$$\text{AP} = \int_0^1 P(R)\mathrm{d}R, \tag{3}$$

$$\text{mAP} = \frac{1}{n}\sum_{i=0}^{n} AP(i), \tag{4}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{5}$$

Where T_p represents the number of correctly detected targets; F_p represents the number of falsely detected targets; F_N represents the number of missed targets; n denotes the number of categories; and $AP(i)$ is the average precision of the i-th target class.

4. Experimental analysis

4.1 Algorithm Comparison Results

To evaluate the advantages of the proposed YOLO-based improved model over existing technologies, we conducted comparative experiments against several mainstream lightweight models, including YOLOv8n, YOLOv10n, YOLOv11n, and YOLOv12n. The detailed experimental results are presented in Table 1.

Table 1. Comparative experimental results of different algorithms.

Algorithms	Recall	mAP@0.5	mAP@0.5-0.95
YOLOv8n	82.6	82.3	51.2

YOLOv10n	80.9	82.5	49.7
YOLOv11n	80.1	83.1	52.2
YOLOv12n	83.7	84.0	52.5
Ours	**85.4**	**84.4**	**52.9**

Table 1 presents the comparative experimental results of different algorithms. The proposed method (Ours) achieves optimal performance across three core metrics—Recall, mAP@0.5, and mAP@0.5-0.95—outperforming all the listed YOLO series algorithms. Specifically, our method leads with a Recall of 85.4%, significantly surpassing the other models, indicating superior performance in reducing false negatives. Additionally, our model ranks first in both mAP@0.5, which measures detection accuracy, and mAP@0.5-0.95, which assesses robustness across various Intersection over Union (IoU) thresholds, outperforming the next best model, YOLOv12n, by 0.4% in both metrics. In comparison, other algorithms demonstrate varying strengths and weaknesses: YOLOv12n shows the second-best overall performance but falls notably behind in recall rate; YOLOv11n performs reasonably well in mAP@0.5-0.95, but suffers from the lowest recall; and YOLOv10n shows weak performance under stricter evaluation metrics. Overall, the experimental results decisively demonstrate that the proposed method excels in detection accuracy, offers comprehensive detection capabilities, and exhibits robust generalization performance.

Table 2. Comparison of various algorithms based on Average Precision (AP/%).

Algorithms / Classes	YOLO v8n	YOLO v10n	YOLO v11n	YOLO v12n	Ours
Bacterial Pneumonia	84.3	84.5	83.2	84.6	**86.8**
Viral Pneumonia	80.3	80.5	83.0	**83.5**	82.0

As shown in Table 2, the proposed method achieves an AP of 86.8% for the Bacterial Pneumonia category, significantly outperforming all comparison algorithms and demonstrating the highest recognition accuracy in this category. However, for the Viral Pneumonia category, our method, while superior to YOLOv8n and YOLOv10n, slightly lags behind YOLOv11n and YOLOv12n. This comparison indicates that our model has a distinct advantage in detecting Bacterial Pneumonia but still leaves room for improvement in recognizing Viral Pneumonia, reflecting the differences in how various algorithms learn features for different categories.

Table 3. Comparison of various algorithms based on Recall (R/%).

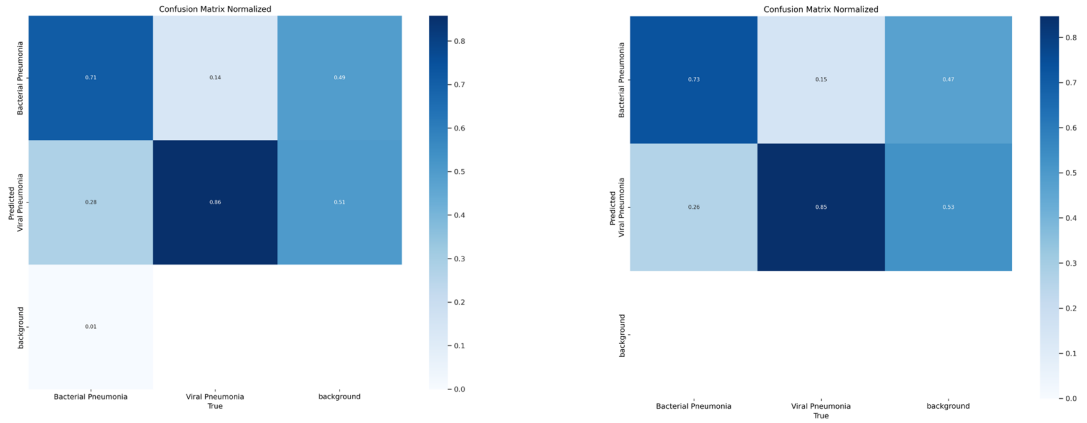
Algorithms / Classes	YOLO v8n	YOLO v10n	YOLO v11n	YOLO v12n	Ours
Bacterial Pneumonia	**82.3**	78.7	82.1	77.6	79.8
Viral Pneumonia	82.9	83.1	78.2	89.8	**91.0**

As shown in Table 3, the recall comparison results reveal that our method performs exceptionally well in the Viral Pneumonia category, achieving a Recall of 91.0%, significantly outperforming all comparison algorithms. This highlights the model's remarkably low false negative rate. Combined with the high Average Precision in Table 2, this demonstrates that the model's detection of this category is both comprehensive and reliable. However, for the Bacterial Pneumonia category, our method achieves a Recall of 79.8%, which, while outperforming YOLOv10n and YOLOv12n, is slightly lower than YOLOv8n and YOLOv11n. This, in conjunction with the highest Average Precision for this category in Table 2, suggests that our model's approach to Bacterial Pneumonia detection is more conservative and accurate. Overall, while our model represents a significant breakthrough in Viral Pneumonia detection, further optimization is needed to improve recall for Bacterial Pneumonia, ensuring a more balanced performance without compromising precision.

4.2 Result Visualization

To evaluate the performance of different models in the pneumonia classification task, we compared the prediction results of YOLOv12n and Ours for Bacterial Pneumonia, Viral Pneumonia, and Background categories. The confusion matrix provides a clear visualization of each model's classification accuracy and misclassification patterns. The matrix

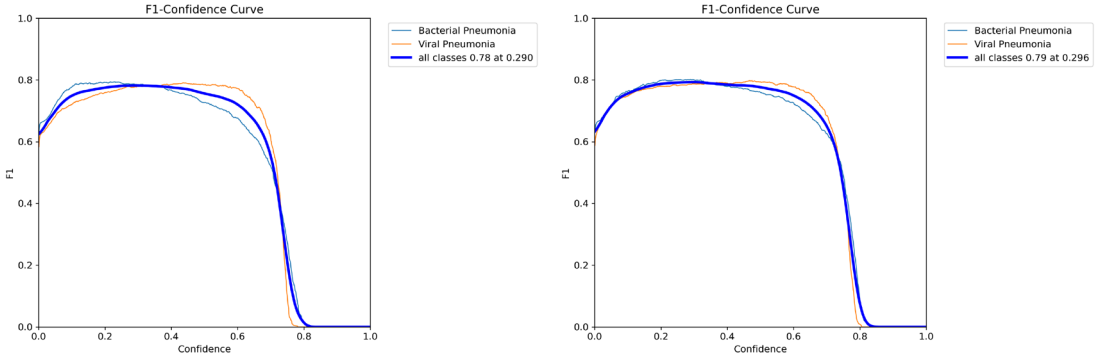
shown in the figure highlights the performance differences between the two models in handling these categories.



(a) (b)

Figure 5. Normalized confusion matrix:(a) YOLOv12n; (b) Ours.

Figure 5 presents a comparison of the performance of the two models in the classification task: YOLOv12n (Figure a) and Ours (Figure b). From the confusion matrices, it is evident that the Ours model demonstrates significant advantages in several aspects. First, for Bacterial Pneumonia, the accuracy of the Ours model is 0.73, which is notably higher than YOLOv12n's 0.71, with a lower misclassification rate for Viral Pneumonia. For Viral Pneumonia, while the accuracy of Ours is slightly lower, the misclassification rate for Bacterial Pneumonia is substantially reduced. More importantly, in Background prediction, the Ours model excels, with a marked decrease in the misclassification rate for the background class, indicating superior performance in effectively eliminating background interference. Overall, the Ours model outperforms YOLOv12n in reducing misclassifications, improving accuracy, and handling background interference, showcasing a stronger classification ability.



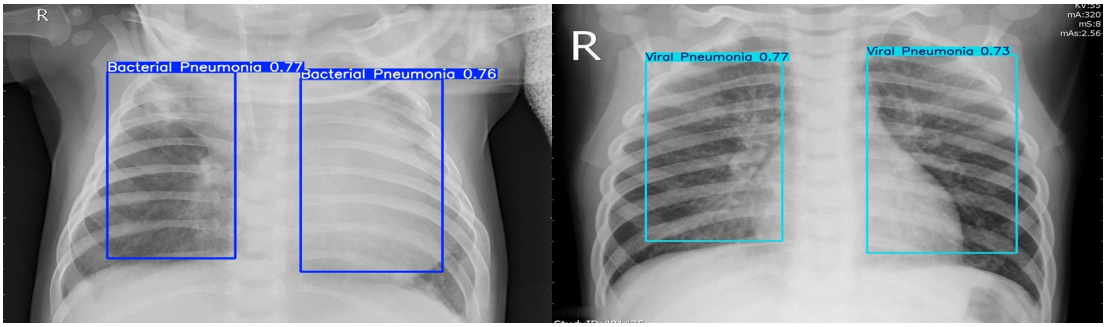
(a) (b)

Figure 6. F1-Confidence curve: (a) YOLOv12n; (b) Ours.

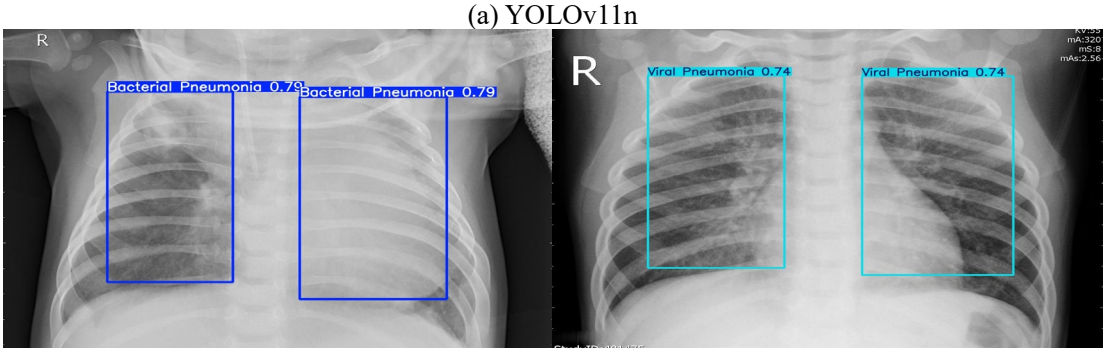
Figure 6 presents a comparison of the F1-Confidence curves between YOLOv12n (Figure a) and Ours (Figure b). As observed, at various confidence thresholds, the F1 score of YOLOv12n remains around 0.78, while our model shows an improvement with an overall F1 score of 0.79, demonstrating more stable performance. Although the F1 score curves for Bacterial Pneumonia and Viral Pneumonia are similar, with the Bacterial Pneumonia score slightly higher than Viral Pneumonia, our model consistently outperforms YOLOv12n at higher confidence levels. This indicates that, when higher confidence is achieved, our model is better at classification. Overall, our model outperforms YOLOv12n in both F1 score and classification accuracy.

To facilitate a better comparison of model performance in the pneumonia classification task, we present the detection results of YOLOv11n, YOLOv12n, and Ours on X-ray images. Each model was evaluated for Bacterial Pneumonia and

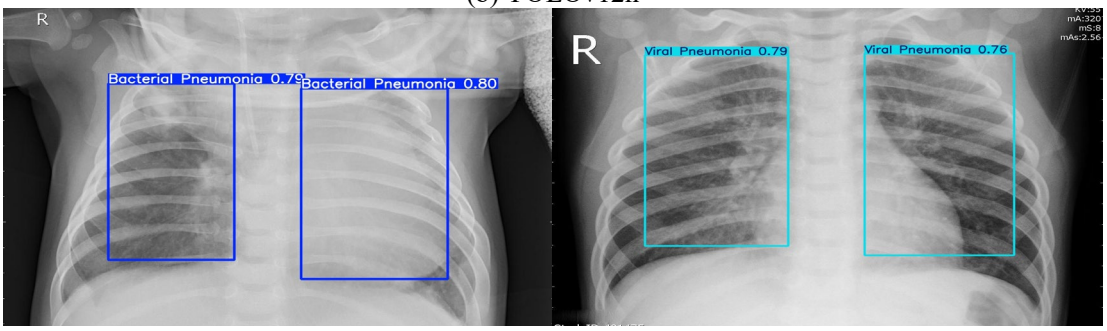
Viral Pneumonia classification, with confidence scores provided for each prediction. By comparing these results, we can clearly observe the differences in classification accuracy and confidence levels across the models.



(a) YOLOv11n



(b) YOLOv12n



(c) Ours

Figure 7. Detection results of different algorithms on the dataset: (a) YOLOv11n; (b) YOLOv12n; (c) Ours.

Figure 7 presents the classification results of YOLOv11n (Figure a), YOLOv12n (Figure b), and Ours (Figure c) on X-ray images. YOLOv11n shows confidence scores of 0.77 and 0.76 for Bacterial Pneumonia and 0.77 and 0.73 for Viral Pneumonia, demonstrating reasonable accuracy but relatively low confidence. YOLOv12n shows some improvement, with confidence scores of 0.79 and 0.79 for Bacterial Pneumonia and 0.74 for Viral Pneumonia, indicating enhanced classification performance. However, our model significantly outperforms the previous two, with confidence scores of 0.79 and 0.80 for Bacterial Pneumonia and 0.79 and 0.76 for Viral Pneumonia, showing more stable performance and higher confidence, particularly excelling in the detection of Bacterial Pneumonia. Overall, our model demonstrates higher confidence and stronger classification ability in both Bacterial Pneumonia and Viral Pneumonia detection.

4.3 Ablation Experiment

To further analyze the improvements of our model, we conducted ablation experiments, with the results presented in Table 4 under different experimental conditions. The experiments primarily compare the performance of the baseline YOLOv12n model with that of the model enhanced by different strategies (DA2C2f and DySample). Through three key metrics—Recall, mAP@0.5, and mAP@0.5-0.95—we assess the impact of each strategy on the model's performance.

Table 4. Ablation experiment results.

number	Experiments	Recall	mAP@0.5	mAP@0.5-0.95
1	YOLOv12n	83.7	84.0	52.5
2	YOLOv12n+DA2C2f	82.3	84.3	52.6
3	YOLOv12n+DA2C2f+DySample	**85.4**	**84.4**	**52.9**

Table 4 presents the ablation experiment results under different experimental conditions. The YOLOv12n model, as the baseline, demonstrates good performance with a Recall of 83.7, mAP@0.5 of 84.0, and mAP@0.5-0.95 of 52.5. After incorporating the DA2C2f strategy, although the Recall slightly decreases to 82.3, the mAP@0.5 improves to 84.3, and the mAP@0.5-0.95 shows a slight increase to 52.6. These results suggest that the DA2C2f strategy enhances the mAP@0.5, while also contributing to a modest improvement in mAP at higher IoU thresholds. Further addition of the DySample strategy leads to a significant improvement in Recall, rising to 85.4, with mAP@0.5 at 84.4 and mAP@0.5-0.95 at 52.9. This indicates that the DySample strategy notably boosts Recall, while also achieving a slight increase in mAP. These findings demonstrate that the DySample strategy effectively enhances Recall while maintaining performance on the mAP metrics, particularly in multi-class detection tasks.

5. Conclusion

This study proposes an improved YOLOv12n model for the automation of pediatric pneumonia detection. The model effectively addresses the issues of insufficient feature extraction and detail loss during upsampling, which are common in existing methods, by introducing two key modules: DA2C2f and DySample. The DA2C2f module enhances the model's ability to capture and express complex pneumonia features through multi-level feature fusion and deep convolution operations, while the DySample module employs a dynamic upsampling mechanism to adaptively reconstruct the feature map, significantly reducing detail loss and image blurring.

Experimental results on a publicly available pediatric chest X-ray dataset demonstrate that the proposed model outperforms the original YOLOv12n and other mainstream lightweight models in key metrics, including Recall, mAP@0.5, and mAP@0.5-0.95. Notably, the model achieved the highest average precision in detecting Bacterial Pneumonia. Ablation studies further confirm the effectiveness of the DA2C2f and DySample modules in enhancing model performance.

Although the proposed YOLOv12n-based model significantly improves pediatric pneumonia detection, there are several limitations that should be addressed in future research. Firstly, the model was primarily tested on a single dataset, and its generalization across diverse datasets with varying image qualities and demographic distributions needs further validation. Additionally, while the model performs well for bacterial pneumonia, its performance in detecting viral pneumonia could be improved. The model also involves complex modules, which may increase computational load, so future work could focus on optimizing it for resource-constrained environments. Finally, integrating other diagnostic modalities, such as CT scans or clinical biomarkers, could enhance detection accuracy. Addressing these limitations will be crucial for making the model more robust, efficient, and applicable in real-world clinical settings.

Fund This work was supported by the College Students' Innovation and Entrepreneurship Training Program of Nanning College of Technology under the project titled "Research on Real-time Healing Technology Based on the OpenCV Emotion Model" (Project No. S202513645075), and by the Scientific Research Foundation for Young and Middle-aged Teachers of Guangxi under the project "Research and Application of a Digital-Twin-Based Smart Campus Visualization and Integrated Management Platform" (Project No. 2025KY1105).

Conflict of Interest The authors have no conflicts to disclose.

Author Contributions Linping Du: conceptualization, formal analysis, methodology, and original draft writing. Zihao Fan: software, validation, writing-review & editing, and visualization. Gang He: funding acquisition, project administration, and writing-review & editing.

Data Availability The dataset generated during this study is available from the first author upon reasonable and justified request.

Received: October 27, 2025;
Accepted: October 30, 2025

References

[1] Mahtab, S., Blau, D. M., Madewell, Z. J., Ogbuanu, I., Ojulong, J., Lako, S., ... & Swart, P. J. (2024). Post-mortem investigation of deaths due to pneumonia in children aged 1–59 months in sub-Saharan Africa and South Asia from 2016 to 2022: an observational study. The Lancet Child & Adolescent Health, 8(3), 201-213.

[2] Guo, K., Cheng, J., Li, K., Wang, L., Lv, Y., & Cao, D. (2023). Diagnosis and detection of pneumonia using weak-label based on X-ray images: a multi-center study. BMC Medical Imaging, 23(1), 209.

[3] Mienye, I. D., Swart, T. G., Obaido, G., Jordan, M., & Ilono, P. (2025). Deep convolutional neural networks in medical image analysis: A review. Information, 16(3), 195.

[4] Redmon J. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[5] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[6] Farhadi A, Redmon J. Yolov3: An incremental improvement[C]//Computer vision and pattern recognition. Berlin/Heidelberg, Germany: Springer, 2018, 1804: 1-6.

[7] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arxiv preprint arxiv:2004.10934, 2020.

[8] Zhang Y, Guo Z, Wu J, et al. Real-time vehicle detection based on improved yolo v5[J]. Sustainability, 2022, 14(19): 12274.

[9] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arxiv preprint arxiv:2209.02976, 2022.

[10] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 7464-7475.

[11] Wang G, Chen Y, An P, et al. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J]. Sensors, 2023, 23(16): 7190.

[12] Wang C Y, Yeh I H, Mark Liao H Y. Yolov9: Learning what you want to learn using programmable gradient information[C]//European conference on computer vision. Cham: Springer Nature Switzerland, 2024: 1-21.

[13] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. arxiv preprint arxiv:2405.14458, 2024.

[14] Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements[J]. arxiv preprint arxiv:2410.17725, 2024.

[15] Chen, Yiyun, et al. "Deep learning for classification of pediatric chest radiographs by WHO's standardized methodology." PLoS one 16.6 (2021): e0253239.

[16] Kundu, Rohit, et al. "Pneumonia detection in chest X-ray images using an ensemble of deep learning models." PloS one 16.9 (2021): e0256630.

[17] Radočaj, Petra, and Goran Martinović. "Interpretable Deep Learning for Pediatric Pneumonia Diagnosis Through Multi-Phase Feature Learning and Activation Patterns." Electronics 14.9 (2025): 1899.

[18] Habib, Nahida, et al. "Ensemble of CheXNet and VGG-19 feature extractor with random forest classifier for pediatric pneumonia detection." SN Computer Science 1.6 (2020): 359.

[19] Yao, Dan, et al. "Accurate and intelligent diagnosis of pediatric pneumonia using X-ray images and blood testing data." Frontiers in bioengineering and biotechnology 11 (2023): 1058888.

[20] Kaya, Mahir. "Feature fusion-based ensemble CNN learning optimization for automated detection of pediatric pneumonia." Biomedical Signal Processing and Control 87 (2024): 105472.

[21] Xin, Kevin Z., David Li, and Paul H. Yi. "Limited generalizability of deep learning algorithm for pediatric pneumonia classification on external data." Emergency Radiology 29.1 (2022): 107-113.

[22] Arun Prakash, J., et al. "Pediatric pneumonia diagnosis using stacked ensemble learning on multi-model deep CNN architectures." Multimedia tools and applications 82.14 (2023): 21311-21351.

[23] Prakash, J. Arun, et al. "Stacked ensemble learning based on deep convolutional neural networks for pediatric pneumonia diagnosis using chest X-ray images." Neural Computing and Applications 35.11 (2023): 8259-8279.

[24] Bal, Ufuk, et al. "A deep learning feature extraction-based hybrid approach for detecting pediatric pneumonia in chest X-ray images." Physical and Engineering Sciences in Medicine 47.1 (2024): 109-117.

[25] Tian, Y., Ye, Q., & Doermann, D. (2025). Yolov12: Attention-centric real-time object detectors. *arXiv preprint arXiv:2502.12524*.

[26] Zhu, Jiachen, et al. "Transformers without normalization." Proceedings of the Computer Vision and Pattern Recognition Conference. 2025.

[27] Liu, W., Lu, H., Fu, H., & Cao, Z. (2023). Learning to upsample by learning to sample. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6027-6037).

[28] Kermany, Daniel S., et al. "Identifying medical diagnoses and treatable diseases by image-based deep learning." cell 172.5 (2018): 1122-1131.

[29] Ming Chen, Yixuan Xu, Wanxiang Qin, Yan Li, Jiyang Yu; Tomato ripeness detection method based on FasterNet block and attention mechanism. AIP Advances 1 June 2025; 15 (6): 065117. https://doi.org/10.1063/5.0280801