IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 43, No. 2, April, 2026, pp. 775-783
Journal homepage: http://iieta.org/journals/ts

An Attention-Guided Lightweight YOLO11n with Ghost Convolution for Pediatric Tooth Detection in Panoramic Radiographs

Jing Ning[1], Bo Liu[2], Biao Zhou[1], Jianing Zhang[3], Yue Ding[2], Yifang Li[2], Yuxing Gao[4*], Yongping Ma[1*]

[1] Department of Stomatology, Baoding Second Hospital, Baoding 071000, China
[2] College of Quality and Technical Supervision, Hebei University, Baoding 071002, China
[3] College of Stomatology, North China University of Science and Technology, Tangshan 063000, China
[4] School of Basic Medical Sciences, Hebei University, Baoding 071002, China

Corresponding Author Email: gaoyuxing0110@163.com; 3201165077@qq.com

https://doi.org/10.18280/ts.430218

Received: 18 October 2025
Revised: 30 January 2026
Accepted: 22 February 2026
Available online: 30 April 2026

Keywords:
pediatric tooth detection, lightweight YOLO11n, Ghost module, dual attention mechanism, Convolutional Block Attention Module, pediatric panoramic radiographs

ABSTRACT

Automated pediatric tooth detection in panoramic radiographs faces significant clinical challenges, including severe tooth overlapping, blurred anatomical boundaries, and the high computational demands of conventional models on edge devices. This study proposes a lightweight YOLO11n algorithm integrating Ghost modules and a dual attention mechanism. Specifically, GhostConv and C3Ghost modules are incorporated into the backbone and neck networks to eliminate feature redundancy and achieve deep lightweight reconstruction. To compensate for potential fine-grained feature loss, a Convolutional Block Attention Module (CBAM) is integrated prior to the detection head to enhance the network's focus on dental contours and suppress background noise. Evaluated on a clinical dataset comprising 1,037 annotated panoramic radiographs, the proposed model reduces parameters to 1.20M and model size to 2.8MB, representing nearly a 50% reduction compared to the baseline, while achieving a mean Average Precision (mAP@0.5) of 97.65%. This approach achieves an optimal tradeoff between computational efficiency and accuracy, providing a highly deployable solution for dental edge computing terminals.

1. INTRODUCTION

Accurate localization and identification of pediatric permanent teeth are crucial steps in clinical auxiliary diagnosis, orthodontic treatment planning, and individual developmental assessment in dentistry. As the most common two-dimensional imaging modality in stomatology, panoramic radiographs can comprehensively and intuitively display the spatial topological distribution of the maxilla, mandible, alveolar bone, and the entire dentition. However, in practical clinical scenarios, automated tooth identification for pediatric full-mouth panoramic radiographs still faces significant challenges. For pediatric patients in the mixed dentition stage, their images typically feature complex characteristics such as the transition between primary and permanent teeth, severe tooth crowding and overlapping, and blurred anatomical boundaries due to lower bone density; traditional manual tooth identification and numbering rely heavily on the clinical experience of dentists, which is not only time-consuming and labor-intensive but also highly subjective, making it prone to missed diagnoses and misjudgments during tedious image interpretation.

To overcome the limitations of manual interpretation and achieve automated tooth detection and numbering, early research primarily relied on a combination of traditional image processing techniques and machine learning classifiers, typically using hand-crafted operators such as Histogram of Oriented Gradients (HOG) and Active Shape Models (ASM) to extract image underlying features, combined with Support Vector Machines (SVM) for object classification. However, such methods are highly sensitive to illumination changes and equipment noise distribution in medical imaging, presenting significant limitations in robustness. In recent years, with the rapid development of deep learning, methods based on Convolutional Neural Networks (CNNs) have gradually become the mainstream [1]. Early deep learning works mostly adopted multi-stage pipeline architectures; for instance, Tuzoff et al. [2] first constructed a "detection-numbering" framework cascading Faster R-CNN with a VGG network, preliminarily considering the absolute position and sequential arrangement information of teeth. To avoid error accumulation caused by cascade structures, subsequent research gradually shifted towards end-to-end unified models. Görürgöz et al. [3] validated the performance of Faster R-CNN in periapical radiograph detection; Chen et al. [4] utilized Faster R-CNN for localization and proposed a post-processing method based on clinical prior rules to correct logical deviations in prediction results; Raeisi et al. [5] constructed a complex multi-branch model to accommodate both position regression and numbering classification requirements; Chung et al. [6] took a different approach, proposing an anchor-free method based on center point regression to achieve tooth localization and

numbering.

Despite the remarkable progress achieved by these deep learning models, traditional high-precision object detection algorithms usually possess a massive number of parameters, making them difficult to deploy on resource-constrained devices. Addressing the dual pain points of the aforementioned complex clinical features and limited computational power for engineering deployment, we propose an improved YOLO11n pediatric tooth localization and identification algorithm based on a lightweight architecture and a dual attention mechanism [7]. This study takes the latest standard YOLO11n as the baseline model. First, the Ghost module is innovatively introduced into the backbone and neck of the network [8], which effectively eliminates redundant feature maps by substituting part of conventional dense convolutions with low-cost linear transformation operations, thereby achieving a substantial reduction in model parameters and computational complexity. Building upon this, to compensate for the potential loss of fine-grained features caused by deep lightweighting operations, the Convolutional Block Attention Module (CBAM) channel and spatial dual attention mechanism are further integrated before the detection head [9]. This mechanism can guide the network to adaptively filter out low-bone-density artifact noise, precisely focusing computational resources on real dental crown contours and root edges, thereby significantly enhancing the model's bounding box regression precision and anti-interference capability in the context of complex mixed dentition under extremely low computational power consumption.

To verify the clinical effectiveness and deployment potential of the proposed algorithm, this study constructed a high-quality pediatric panoramic radiograph dataset. The dataset originates from 1,037 images of children aged 3–14 collected by the Second Hospital of Baoding. Professional physicians conducted meticulous tooth numbering annotations on tens of thousands of permanent teeth, covering complex clinical distributions such as normal dentition, supernumerary teeth, and missing teeth, providing solid data support for model training. Systematic evaluation and comparative experiments conducted on this basis show that the improved object detection network performs excellently in balancing high efficiency and high precision. Its parameter count (Params) is only 1.20 M, and the model volume is compressed to a tiny 2.8 MB, achieving a reduction of nearly 50% compared to the baseline model; simultaneously, the model's mean Average Precision (mAP@0.5) reached 97.65%, successfully outperforming the unlightweighted standard network. Combined with detailed ablation experiments and visual analysis of detection results, this study profoundly reveals the inherent complementary effects of the Ghost lightweight module and the CBAM attention mechanism in medical image processing. This not only proves the success of the proposed algorithm in achieving an optimal trade-off between computational efficiency and detection precision, but also provides a highly promising technical solution for the low-cost, large-scale deployment of future pediatric dental intelligent auxiliary diagnostic models on constrained edge computing terminals.

2. METHODS

2.1 YOLO11-Ghost-CBAM

Targeting the computational bottlenecks faced by the native YOLO11 network architecture when deployed on constrained hardware in primary dental clinics, as well as the issue of small targets being prone to missed detection in complex pediatric oral images (mixed dentition stage), we conducted a deep physical reconstruction and feature enhancement of the original network architecture. The main improvement strategies are divided into two aspects: an ultra-minimalist lightweight backbone design based on the Ghost module, and multi-scale feature refinement based on the CBAM attention mechanism. The overall network architecture of the improved YOLO11-Ghost-CBAM is shown in Figure 1.

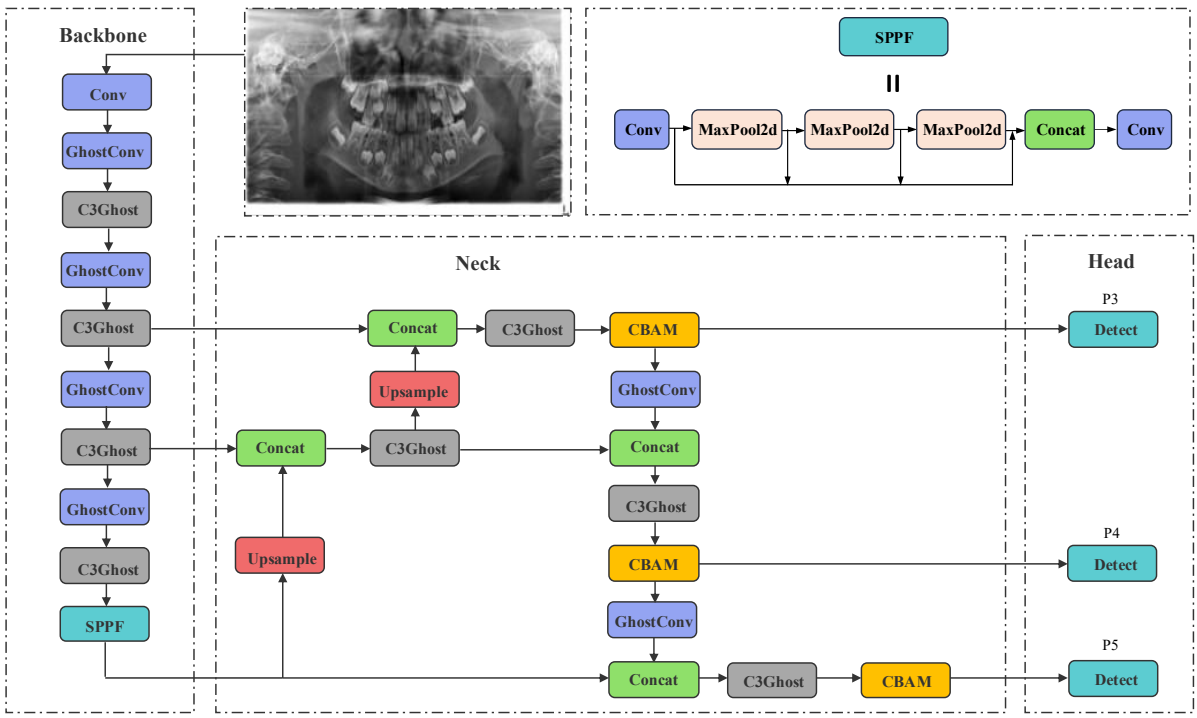

Figure 1. The network architecture of YOLO11-Ghost-CBAM

2.2 Lightweight design with Ghost module

In the native YOLO11 model, the number of deep network channels expands exponentially (the deepest feature map channel count reaches up to 1024), and the massive stacking of standard convolutions and C3k2 modules, while extracting rich global semantic information, also leads to severe feature map redundancy [10]. Extensive visualization research indicates extremely high similarity among feature maps generated by deep networks; this "feature redundancy" causes massive parameter and computational overload, severely restricting the model's deployment on medical terminals lacking high-performance dedicated graphics cards.

To this end, we first structural slimming of the network at the macroscopic topological level. The network's Depth Factor is compressed from the default value to 0.33 to reduce the repetitive stacking redundancy of feature extraction modules; simultaneously, targeting the characteristic that pediatric dental image categories are relatively simple and semantic complexity is lower than that of natural scene datasets, the maximum channel count of the deep network is forcibly truncated to 512. Since the computational complexity of convolutional layers is proportional to the square of the channel count, this single channel truncation strategy instantly slashes the Params of the deepest network layers by nearly 75%.

At the microscopic operator level, the lightweight Ghost module is introduced to replace traditional heavy standard convolutions. Traditional standard convolutions consume massive multiply-accumulate operations when generating feature maps, whereas the Ghost module abandons this expensive global mapping approach [11]. The feature generation comparison between standard convolution and the Ghost module is shown in Figure 2.

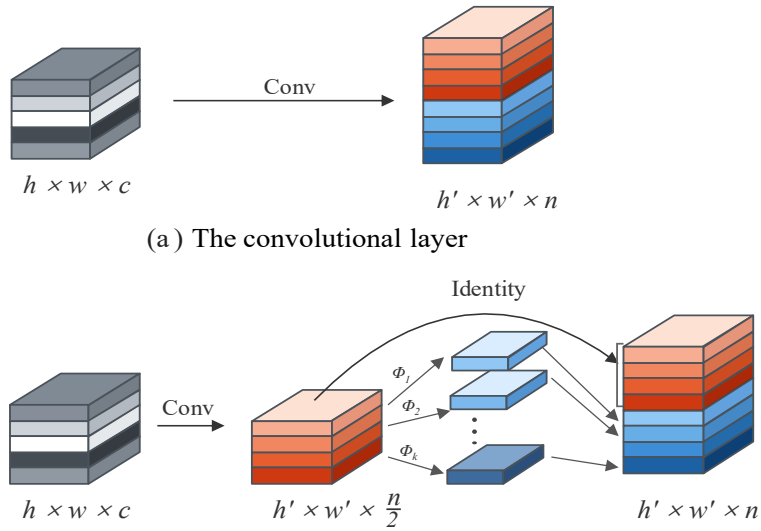

Figure 2. Feature generation process of (a) the standard convolution layer and (b) the Ghost module

As shown in Figure 2, traditional standard convolution needs to complete all feature mappings at once when processing input feature maps with dimensions $h \times w \times c$. The core idea of the Ghost module is divided into two stages: first, a small number of standard convolution kernels are used to extract the "intrinsic features" of the image; subsequently, based on these features, ultra-low-cost linear transformation operations $\Phi_1, \Phi_2, \ldots, \Phi_k$ (such as depthwise separable convolutions) are applied to generate the remaining "Ghost features", and finally, an identity mapping is used to concatenate the two along the channel dimension to achieve feature reuse.

Assuming the input feature map is $X \in R^{c \times h \times w}$, and the intrinsic feature map generated by conventional convolution in the first stage is $Y' \in R^{m \times h' \times w'}$, its process of generating "phantom features" using cheap linear operations can be represented as:

$$y_{ij} = \Phi_{i,j}(y'_i), \forall i = 1, \ldots, m, j = 1, \ldots, s-1 \quad (1)$$

where, y'_i is the i-th channel feature in the intrinsic feature map Y', $\Phi_{i,j}$ represents the j-th cheap linear operation (e.g., 3×3 depthwise convolution), and s is the channel expansion ratio. The final output feature map Y is composed of the concatenation of intrinsic features and phantom features. In actual network construction, the output of this module is further combined with Batch Normalization and the SiLU activation function, encapsulating it into a complete GhostConv basic operator [12].

Upon completing the lightweighting of microscopic operators, we further comprehensively replace the primary feature extraction modules in the Backbone and Neck networks with the novel C3Ghost module [13]. The native C3k2 module stacks a large number of heavy convolutions internally, while the C3Ghost module ingeniously combines the shunting thought of the CSP (Cross Stage Partial) architecture with the lightweight advantages of the Ghost module.

As shown on the left panel of Figure 3, for an input feature map with dimensions $h \times w \times c$, the C3Ghost module first splits it in parallel into two branches. Both branches pass through pointwise convolutional layers configured with $k = 1, s = 1, p = 0$ (i.e., 1×1 convolution, stride 1, zero padding); this operation, while integrating cross-channel information, precisely compresses the channel count of each branch to $n/2$, effectively avoiding the dimensionality disaster in subsequent operations. Subsequently, one of the branches enters the feature extraction backbone composed of n GhostBottleNecks.

As shown on the right panel of Figure 3, a single GhostBottleNeck is internally composed of two Ghost modules in series, and residual connections are introduced at both ends of the module; this ensures that deep networks will not experience vanishing gradients and feature degradation during forward propagation. Finally, the feature branch deeply processed by the bottleneck layer is concatenated with the unprocessed shortcut branch, and passes through another 1 × 1 convolutional layer for dimensional alignment and feature fusion. Through this synergistic design of cross-stage shunting and lightweight bottleneck layers, the network successfully strips away invalid computational redundancy, substantially improving the model's inference efficiency on clinical terminals.

2.3 Multi-scale refinement with Convolutional Block Attention Module

Although the large-scale application of Ghost modules and the channel truncation strategy endow the model with excellent lightweight characteristics, the degradation in the computational dimension inevitably weakens the network's representational capacity for local minute features. Especially in pediatric panoramic radiographs, patients are in the mixed dentition stage, and the small, unerupted permanent tooth germs often severely overlap with primary tooth roots, complex alveolar bone backgrounds, and soft tissues. In such extremely low-contrast and cluttered background scenarios, an extremely lightweight network is highly prone to target boundary blurring and missed detections of small targets.

To compensate for the precision loss caused by lightweighting, we precisely embed the plug-and-play, computationally inexpensive CBAM [14] at key nodes of the three multi-scale prediction branches (P3, P4, P5) outputted from the Neck layer to the Detect head. Unlike traditional attention mechanisms that only focus on a single dimension, CBAM sequentially infers attention weights along two independent dimensions: Channel and Spatial, achieving secondary refinement of deep features (Figure 4).

First, the Channel Attention Module focuses on solving the "what is the target" problem. Different feature channels often respond to different visual patterns; this module aggregates spatial information through global Max-Pooling and global Avg-Pooling, and feeds it into a shared Multi-Layer Perceptron (MLP) to adaptively learn the weight coefficients of each channel, thereby effectively enhancing valid channels representing tooth contours and filtering out irrelevant noise from soft tissues and alveolar bone backgrounds [15]. For input feature F, the calculation process of its channel attention weight $M_c(F)$ is shown in Formula (2):

$$M_c(F) = \sigma\left(MLP\left(AvgPool(F)\right) + MLP\left(MaxPool(F)\right)\right) \quad (2)$$

Subsequently, the feature map purified by channel weighting enters the Spatial Attention Module, which focuses on solving the "where is the target" problem. After performing aggregation pooling operations on the channel dimension, it extracts spatial dependencies through a 7 × 7 large receptive field convolution kernel to generate a spatial weight matrix, accurately calibrating the spatial distribution of tooth targets on the 2D image. Its calculation process is shown in Formula (3):

$$M_s(F) = \sigma\left(f^{7\times7}([AvgPool(F); MaxPool(F)])\right) \quad (3)$$

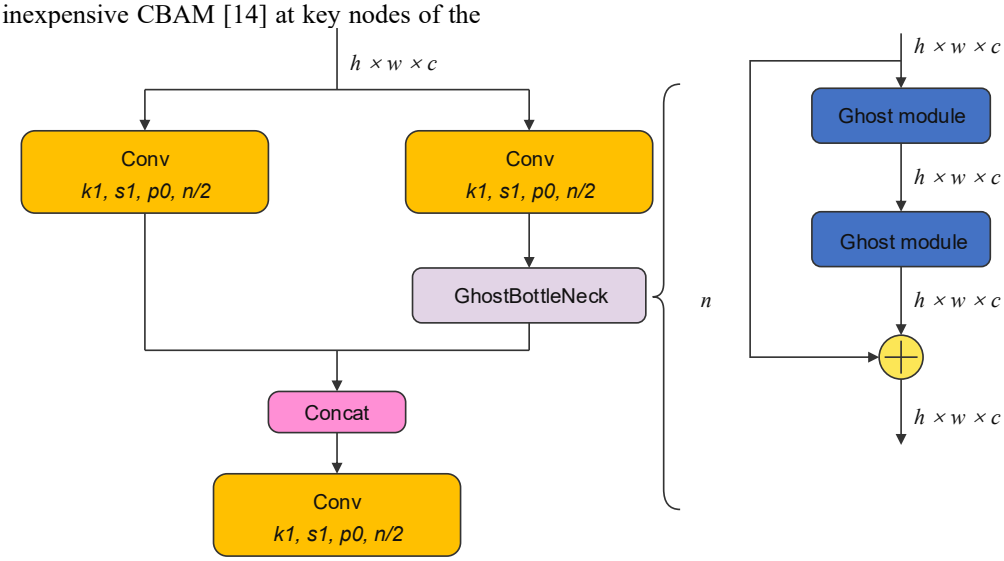

Figure 3. The network structure of the C3Ghost module

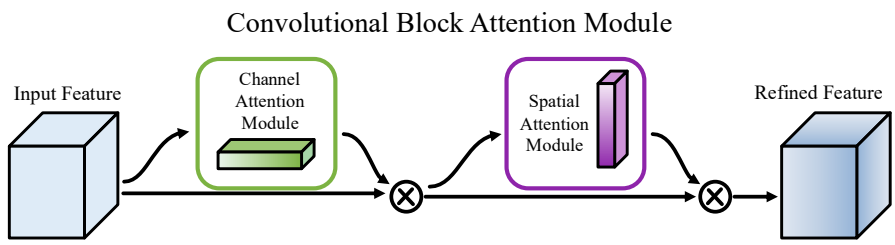

Figure 4. The Convolutional Block Attention Module (CBAM) dual attention mechanism

where, σ represents the Sigmoid activation function, [;] represents the feature concatenation operation, and $f^{7\times7}$ represents a convolution operation with a kernel size of 7 × 7.

Through the sequential filtering of the aforementioned channel and spatial dual attention mechanisms, the CBAM module enhances the network's response to the edge features of tiny and overlapping teeth via adaptive weighting in both spatial and channel dimensions, effectively suppressing the interference of background noise and soft tissue artifacts. Experiments show that this plug-and-play reinforcement strategy, at the cost of adding a very small number of parameters, perfectly compensates for the feature loss brought by the Ghost lightweight reconstruction, successfully achieving an optimal Accuracy-Efficiency Trade-off [16].

3. EXPERIMENT

3.1 Data collection and preprocessing

This study collected a total of 1,037 pediatric panoramic radiographs from patients aged 3–14 at the Stomatology Department of the Second Hospital of Baoding. All panoramic films excluded cases of obvious artifacts, edentulous jaws, and ectopic teeth, and each image was digitized at a resolution of 12.5 pixels per millimeter, with dimensions of approximately 3000 × (1200 to 1450) pixels. During the formal annotation stage, trained physicians used the X-AnyLabeling software to create minimum bounding boxes for each complete tooth (including crown and root) in the images, while simultaneously assigning accurate tooth numbers based on the FDI tooth numbering system [17]. After annotation, X-AnyLabeling automatically saved the label files containing bounding box coordinates, tooth numbers, and staging results uniformly into JSON and TXT formats, as shown in Figure 5.

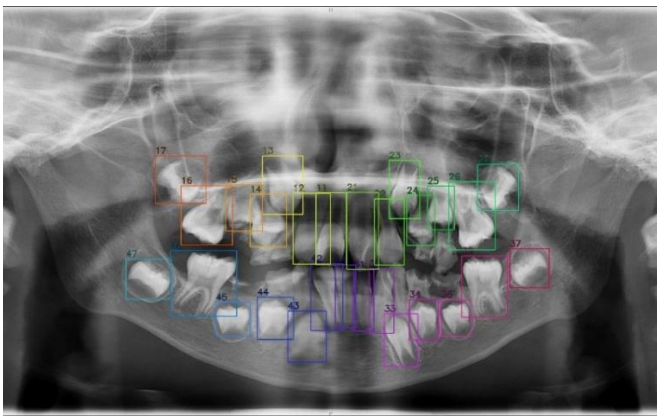

Figure 5. Results of pediatric permanent tooth numbering

3.2 Dataset Distribution

Based on previous data collection and software bounding box annotation, this study detailed the statistics and categorization of the 1,037 panoramic radiographs collected from the Second Hospital of Baoding. The constructed dataset covers a rich sample of permanent teeth in all quadrants of the full mouth; the specific quantitative distribution of FDI tooth numbers is shown in Table 1.

To comprehensively evaluate the generalization ability and robustness of the model, this study randomly divided the collected 1,037 pediatric panoramic radiographs into a training set (approx. 829 images), a validation set (approx. 104 images), and a test set (approx. 104 images) at a ratio of 8:1:1. The training set is used for the learning and updating of model parameters, the validation set is used for hyperparameter tuning and preventing overfitting during the training process, and the test set serves as completely unseen data for the final evaluation of the model's detection performance in real-world scenarios.

Table 1. Quantitative distribution of FDI tooth numbers

FDI	Quantity	FDI	Quantity	FDI	Quantity	FDI	Quantity
11	1033	21	1031	31	1026	41	1028
12	1027	22	1020	32	1010	42	1008
13	1027	23	1027	33	1034	43	1034
14	1032	24	1032	34	1033	44	1034
15	996	25	999	35	996	45	992
16	1036	26	1035	36	1036	46	1035
17	1001	27	1002	37	1024	47	1025
18	416	28	430	38	583	48	579

3.3 Data augmentation

Data augmentation is one of the important methods to enhance network performance and reduce the risk of deep learning model overfitting. During the data preparation stage, this study applied data augmentation methods to optimize the strictly screened and annotated sample set, thereby ultimately constructing a high-quality classification and detection dataset, which substantially improved the model's robustness and generalization ability when handling complex medical imaging data with strong heterogeneity and imbalanced distributions.

3.4 Implementation details of model training

To accelerate model training and enhance its generalization capacity and stability, the experimental execution environment was an Ubuntu 20.04 operating system with 32GB of RAM, equipped with a 16GB RTX 4060Ti graphics card and a high-performance multi-core CPU processor. The parallel computing architecture utilized Cuda 11.8 and the CuDNN acceleration library, and programming was done in Python 3.9, based on the PyTorch 1.12 deep learning framework, and integrated with the OpenCV library. In terms of hyperparameter configurations, the input image size was set to 640; all models were trained for 100 iterations (epochs), with a batch size set to 64. Optimization of network parameters used the Stochastic Gradient Descent (SGD) algorithm, with an initial learning rate set to 0.001 and a weight decay coefficient of 0.0001.

3.5 Metrics of performance test

To comprehensively evaluate the performance of the pediatric tooth localization and identification model, this study not only focused on the detection precision of the model but also fully considered the requirements for its deployment on clinical edge computing or lightweight devices. Therefore, this study selected Precision, Recall, Average Precision (AP), mAP, Floating Point Operations (FLOPs), Params, and model size as core evaluation metrics.

Precision refers to the proportion of actual positive targets among the targets predicted as positive by the model, and its calculation is shown in Formula (4). True Positives (TP)

represents the number of permanent tooth targets correctly identified by the model; False Positives (FP) and False Negatives (FN) respectively represent the number of backgrounds or primary teeth incorrectly identified as permanent teeth by the model, and the number of actual permanent teeth missed by the model.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

Recall refers to the proportion of targets correctly identified by the model among all actual positive targets, and its calculation is shown in Formula (5).

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

Average Precision (AP) equals the area under the Precision-Recall (P-R) curve, and a value closer to 1 indicates better localization performance of the model on that category; its calculation is shown in Formula (6).

$$AP = \int_0^1 Precision(Recall)\ dRecall \tag{6}$$

Mean Average Precision (mAP) is the average of the AP values across all sample categories; it is the most commonly used evaluation metric in the object detection field and can intuitively reflect the comprehensive detection performance of the current model, with its calculation shown in Formula (7). Where N is the total number of detected object categories.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{7}$$

FLOPs can objectively reflect the computational complexity of the model, and its calculation is shown in Formula (8).

$$FLOPs = 2 \times H \times W \times C_{in}(K^2 + 1) \times C_{out} \tag{8}$$

The lightweight degree of the model is primarily evaluated through the network Params, and its calculation is shown in Formula (9).

$$Params = C_{in} \times K^2 \times C_{out} \tag{9}$$

Additionally, model size can be used to directly measure the convenience of model deployment and its applicability across various clinical lightweight devices.

4. RESULTS AND ANALYSIS

4.1 Ablation analysis

This study compared the performance of the baseline model (standard YOLO11n) with network structures introducing different improvement modules on the same self-built pediatric panoramic radiograph test set, to verify the effectiveness of introducing Ghost modules (C3Ghost and GhostConv) into the backbone and neck networks, as well as introducing the CBAM attention mechanism before the detection head. The corresponding ablation experiment results are shown in Table 2.

Table 2. Performance metrics of ablation experiments

A	B	mAP@0.5(%)	Params (M)	Flops (G)	Model Size (MB)	FPS
-	-	97.60	2.60	6.3	5.5	141
√	-	97.50	1.20	5.3	2.7	107
-	√	97.77	2.35	6.1	5.6	109
√	√	97.65	1.20	5.3	2.8	101

Note: A denotes C3Ghost and GhostConv; B denotes the CBAM module.

As shown in Table 2, the introduction of different improvement modules caused the model to exhibit differentiated performance focuses between lightweighting and detection precision. Specifically, when only the Ghost lightweight module (Configuration A) was introduced, the model's Params was significantly reduced from 2.60 M to 1.20 M, and computational volume (FLOPs) and model volume dropped substantially to 5.3 G and 2.7 MB, respectively. This indicates that the Ghost structure effectively eliminated feature redundancy in conventional convolutions, achieving deep lightweighting. However, this degradation in computational dimension inevitably resulted in the loss of some fine-grained features, causing the mAP50 to drop slightly to 97.50% compared to the baseline model. Conversely, when only the CBAM attention mechanism was introduced before the detection head (Configuration B), benefiting from the adaptive weighting effects in both spatial and channel dimensions, the network's feature representation ability for key targets was significantly enhanced, and the mAP@0.5 reached a peak of 97.77%.

To achieve an optimal balance between computational efficiency and localization precision, the final model completely integrated the Ghost structure and the CBAM mechanism. Experimental results show that while maintaining extreme lightweight characteristics (1.20 M parameters, 2.8 MB model size, nearly a 50% compression compared to the baseline model), the improved YOLO11n effectively compensated for the precision loss caused by lightweighting with the help of the CBAM module. The final model's mAP50 rebounded to 97.65%, not only making up for the feature loss but also maintaining a high precision level equivalent to the benchmark model (97.60%). Although compared to the configuration introducing only the CBAM module, the final model's precision experienced a minor drop of 0.12% (from 97.77% to 97.65%), this was traded for a substantial reduction of nearly 50% in Params (from 2.35 M down to 1.20 M). Considering the computational and memory bottlenecks of clinical portable devices, this strategy achieved an effective compromise between deployment feasibility and high precision. Experiments show that this combined architecture, while stripping away network computational redundancy, better preserved shallow visual edge information and integrated deep anatomical semantics, providing a reliable solution for edge medical device deployment under stringent computational conditions.

4.2 Comparative analysis of different object detection algorithms

During the testing phase, this study evaluated the comprehensive performance of the proposed improved YOLO11n model against classic lightweight object detection

networks in the YOLO series, including YOLOv5n [18], YOLOv8n [19], YOLOv10n [20], standard YOLO11n, and YOLOv13n [21]. To ensure fairness and rigor in the comparison, all models underwent end-to-end training and evaluation on the exact same pediatric panoramic radiograph dataset, using identical hyperparameter configurations and hardware environments. The specific performance metric comparisons of these networks are detailed in Table 3.

Table 3. Comparison of performance metrics among different object detection algorithms

Model	mAP50 (%)	Params (M)	Flops (G)	Model Size (MB)
YOLOv5n	95.55	2.51	7.1	5.3
YOLOv8n	97.10	3.10	8.1	6.3
YOLOv10n	96.84	2.27	6.6	5.8
YOLO11n	97.60	2.60	6.3	5.5
YOLOv13n	**98.12**	2.45	6.2	5.5
Ours	97.65	**1.20**	**5.3**	**2.8**

As shown in Table 3, the improved YOLO11n model proposed in this study achieved a good balance between detection precision and lightweight deployment requirements. In terms of lightweight metrics, the Params of our model is 1.20 M, the computational volume is reduced to 5.3 G, and the model size is shrunk to 2.8 MB. Compared to the benchmark model YOLO11n (2.60 M parameters, 5.5 MB size), the proposed model achieved nearly 50% compression in both Params and volume. In terms of detection precision, although YOLOv13n obtained the highest mAP50 (98.12%), its Params (2.45 M) and model volume (5.5 MB) are relatively large, which is prone to increasing the risk of memory occupation and inference latency in strictly restricted edge computing devices. In contrast, under the premise of substantially cutting network redundancy, the improved model's mAP50 was still maintained at a high level of 97.65%. This precision not only maintained a performance comparable to the benchmark model YOLO11n (97.60%), but also outperformed classic networks like YOLOv5n, YOLOv8n, and YOLOv10n. This performance can be attributed to the architectural design: when processing complex, overlapping, and crowded pediatric mixed dentitions with lower bone density, the standard YOLO series with conventional dense convolutions is prone to introducing excessive background noise interference; however, the proposed model, through the combination of the Ghost lightweight module and the CBAM attention mechanism, enhanced the network's ability to extract local critical anatomical details from limited medical image data. This indicates that the proposed algorithm, while overcoming complex dention detection challenges, effectively lowered the hardware computational threshold, providing a highly feasible lightweight solution for the clinical deployment of edge medical devices (e.g., portable intraoral scanners).

4.3 Visual analysis

To intuitively assess whether deep lightweighting operations negatively impacted the model's actual localization performance, this study selected representative complex pediatric panoramic radiographs from the test set for visual analysis. Figure 6 visually demonstrates the bounding box detection effects of different models under real clinical scenarios. Among them, Figure 6(a) displays original images with clinical expert annotations (Ground Truth), Figure 6(b) shows detection results of the standard YOLO11n model, and Figure 6(c) presents the detection results of our improved YOLO11n model.

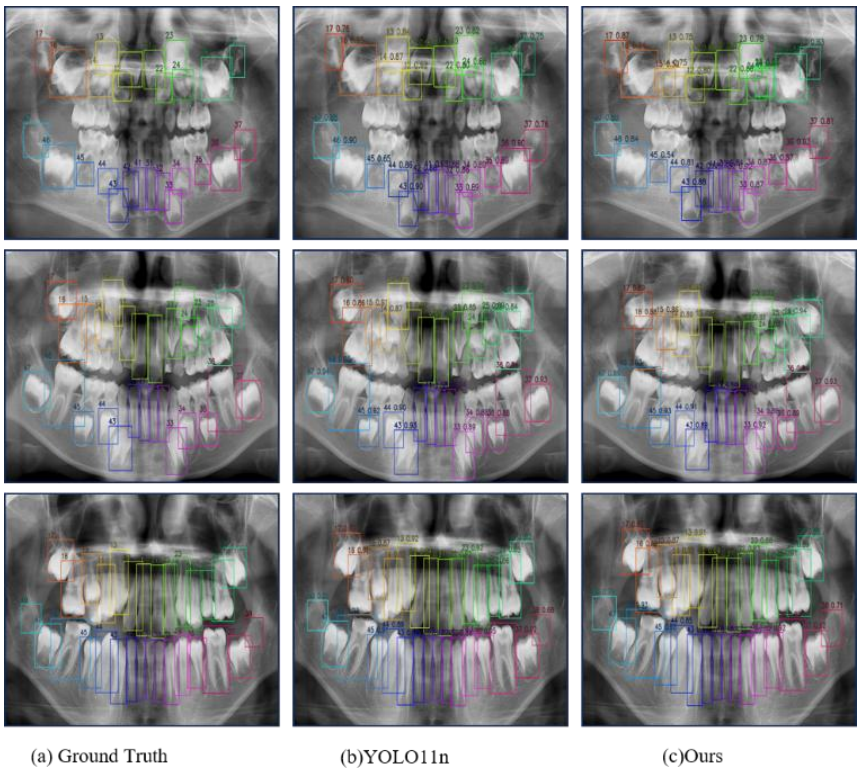

Figure 6. Comparison of detection effects. (a) Ground Truth; (b) YOLO11n; (c) Ours

Horizontal comparison clearly reveals that the standard YOLO11n model (Figure 6(b)) already demonstrates satisfactory localization capabilities against complex mixed dentition backgrounds, enabling relatively accurate bounding of target permanent teeth. Of critical significance is the improved YOLO11n model proposed herein (Figure 6(c)): despite undergoing deep lightweight reconstruction via Ghost modules, achieving nearly 50% reduction in parameters and model size relative to the baseline, its detection performance remains highly consistent with that of the standard YOLO11n. Visualization results indicate that the improved model (Figure 6(c)) precisely localizes all target teeth without exhibiting missed detections of small targets attributable to drastic network parameter reduction, nor false detections of adjacent primary teeth or bone artifacts. Moreover, its generated bounding boxes maintain tight, accurate fitting to actual tooth contours, with localization precision fully comparable to the unlightweighted baseline network.

These visualization results provide compelling empirical evidence that, owing to the effective compensation of key anatomical features by the CBAM dual attention mechanism, the proposed lightweight strategy achieves substantial reductions in hardware computational resource consumption and improved model inference efficiency while perfectly maintaining detection performance comparable to the baseline YOLO11n. This effectively realizes the "lightweight yet highly efficient" deployment of the algorithm under resource-constrained hardware conditions.

5. CONCLUSIONS

Targeting clinical auxiliary diagnosis difficulties in pediatric mixed dentition panoramic radiographs—such as tooth crowding, overlapping, and blurred boundaries due to low bone density—and to satisfy the lightweight deployment requirements of object detection algorithms on various medical edge computing devices, an improved YOLO11n algorithm is proposed for pediatric tooth localization and identification based on Ghost structures and the CBAM dual attention mechanism. By introducing GhostConv and C3Ghost modules into the backbone and neck of the network, this study effectively eliminated redundant computations generated by traditional conventional convolutions during feature extraction, achieving extreme lightweight reconstruction of the model. Experimental data show that the improved network has a Params of only 1.20 M, computational volume reduced to 5.3 G, and model size compressed to a tiny 2.8 MB. Compared to the standard baseline model YOLO11n, this model achieved a substantial reduction of nearly 50% in both parameters and volume, drastically reducing reliance on hardware computational power and clearing the computational obstacles for deploying the algorithm on lightweight devices.

At the same time, to compensate for the loss of fine-grained features that deep model lightweighting might bring, we specifically integrate the CBAM channel and spatial dual attention mechanism before the detection head. This mechanism successfully guided the network to adaptively focus on critical anatomical regions, such as dental crown edges and alveolar bone boundaries, in complex backgrounds, effectively suppressing interference from artifacts and irrelevant tissues. Test results demonstrate that the improved YOLO11n reached a mean Average Precision (mAP50) of 97.65%; not only did it successfully surpass the unlightweighted baseline model, but when handling difficult samples involving supernumerary teeth and primary tooth interference, it could still generate bounding boxes precisely fitting the real contours, truly achieving lightweighting without performance degradation. In summary, the proposed lightweight detection algorithm effectively realized an optimal balance between computational efficiency and detection precision in the complex task of pediatric tooth localization and identification. These research findings not only provide highly robust technical support for the intelligent auxiliary diagnosis of pediatric panoramic images, but also offer a highly promising and feasible lightweight solution for the large-scale deployment of such deep learning algorithms onto constrained hardware platforms like portable intraoral scanners and primary clinic terminals in the future.

ACKNOWLEDGMENT

This work was supported by the Natural Science Foundation of Hebei Province (Beijing-Tianjin-Hebei Basic Research Cooperation Special Project) (Grant No.: H2023104901).

REFERENCES

[1] Huang H, Xia T, Ren P. Partial channel network: Compute fewer, perform better. arXiv preprint arXiv:2502.01303, 2025.

[2] Tuzoff, D.V., Tuzova, L.N., Bornstein, M.M., Krasnov, A.S., Kharchenko, M.A., Nikolenko, S.I., Sveshnikov, M.M., Bednenko, G.B. (2019). Tooth detection and numbering in panoramic radiographs using convolutional neural networks. Dentomaxillofacial Radiology, 48(4): 20180051. https://doi.org/10.1259/dmfr.20180051

[3] Görürgöz, C., Orhan, K., Bayrakdar, I.S., Çelik, Ö., Bilgir, E., Odabaş, A., Aslan, A.F., Jagtap, R. (2022). Performance of a convolutional neural network algorithm for tooth detection and numbering on periapical radiographs. Dentomaxillofacial Radiology, 51(3): 20210246. https://doi.org/10.1259/dmfr.20210246

[4] Chen, H., Zhang, K., Lyu, P., Li, H., Zhang, L., Wu, J., Lee, C. (2019). A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. Scientific Reports, 9(1): 1-13. https://doi.org/10.1038/s41598-019-40414-y

[5] Raeisi, Z., Rokhva, S., Rahmani, F., Goodarzi, A., Najafzadeh, H. (2025). Multi-label diagnosis of dental conditions from panoramic x-rays using attention-enhanced deep learning. Oral and Maxillofacial Surgery, 29(1): 1-25. https://doi.org/10.1007/s10006-025-01463-y

[6] Chung, M., Lee, J., Park, S., Lee, M., Lee, C.E., Lee, J., Shin, Y. (2021). Individual tooth detection and identification from dental panoramic X-ray images via point-wise localization and distance regularization. Artificial Intelligence in Medicine, 111: 101996. https://doi.org/10.1016/j.artmed.2020.101996

[7] Alkhammash, E.H. (2025). Multi-classification using YOLOv11 and hybrid YOLO11n-mobilenet models: A fire classes case study. Fire, 8(1): 17. https://doi.org/10.3390/fire8010017

[8] Elhenidy, A.M., Labib, L.M., Haikal, A.Y., Saafan, M.M. (2025). GY-YOLO: Ghost separable YOLO for pedestrian detection. Neural Computing and Applications, 37(20): 14907-14933. https://doi.org/10.1007/s00521-025-11207-4

[9] Cheng, A., Xiao, J., Li, Y., Sun, Y., Ren, Y., Liu, J. (2024). Enhancing remote sensing object detection with K-CBST YOLO: Integrating CBAM and Swin-Transformer. Remote Sensing, 16(16): 2885. https://doi.org/10.3390/rs16162885

[10] Hidayatullah P, Syakrani N, Sholahuddin MR (2025). Yolov8 to yolo11: A comprehensive architecture in-depth comparative review. arXiv preprint arXiv:2501.13400, 2025.

[11] Tang, J., Xu, B., Li, J., Zhang, M., Huang, C., Li, F. (2025). Ghost-YOLO-GBH: A lightweight framework for robust small traffic sign detection via GhostNet and bidirectional multi-scale feature fusion. Eng, 6(8): 196. https://doi.org/10.3390/eng6080196

[12] Cao, J., Bao, W., Shang, H., Yuan, M., Cheng, Q. (2023). GCL-YOLO: A GhostConv-based lightweight YOLO network for UAV small object detection. Remote Sensing, 15(20): 4932. https://doi.org/10.3390/rs15204932

[13] Naufal, M.F., Kusuma, S.F. (2025). BloodCell-YOLO: Efficient detection of blood cell types using modified YOLOv8 with GhostBottleneck and C3Ghost modules. Journal of Information Systems Engineering and Business Intelligence, 11(1): 41-52. https://doi.org/10.20473/jisebi.11.1.41-52

[14] Bin Islam, S., Chowdhury, M.E.H., Hasan-Zia, M., Kashem, S.B.A., Majid, M.E., Kunju, A.K.A., Khandakar, A., Ashraf, A., Nashbat, M. (2025). VisioDECT: A novel approach to drone detection using CBAM-integrated YOLO and GELAN-E models. Neural Computing and Applications, 37(24): 20181-20204. https://doi.org/10.1007/s00521-025-11448-3

[15] Yu, H., Wang, J., Han, Y., Fan, B., Zhang, C. (2024). Research on an intelligent identification method for wind turbine blade damage based on CBAM-BiFPN-YOLOV8. Processes, 12(1): 205. https://doi.org/10.3390/pr12010205

[16] Sun, A., Mao, Z. (2026). YOLO-CFAEW: A YOLOv8-based detection network with CBAM attention, RFAConv module, and multi-head structure for engine casting defect detection. Journal of Intelligent Manufacturing. https://doi.org/10.1007/s10845-026-02801-x

[17] Tekin, B.Y., Ozcan, C., Pekince, A., Yasa, Y. (2022). An enhanced tooth segmentation and numbering according to FDI notation in bitewing radiographs. Computers in Biology and Medicine, 146: 105547. https://doi.org/10.1016/j.compbiomed.2022.105547

[18] Zhang, Y., Guo, Z., Wu, J., Tian, Y., Tang, H., Guo, X. (2022). Real-time vehicle detection based on improved YOLO v5. Sustainability, 14(19): 12274. https://doi.org/10.3390/su141912274

[19] Wang, G., Chen, Y., An, P., Hong, H., Hu, J., Huang, T. (2023). UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. Sensors, 23(16): 7190. https://doi.org/10.3390/s23167190

[20] Mao, M., Lee, A., Hong, M. (2024). Efficient fabric classification and object detection using YOLOv10. Electronics, 13(19): 3840. https://doi.org/10.3390/electronics13193840

[21] Liu, Z., Wang, J., Wu, H., Xue, F., Qin, Z., Sun, S., Guo, X., Zhao, F. (2026). Water-aware real-time detection of floating plastic debris via an enhanced YOLOv13 framework for aquatic pollution monitoring. Expert Systems with Applications, 313: 131552. https://doi.org/10.1016/j.eswa.2026.131552