http://dx.doi.org/10.62762/TSCC.2025.972040

RESEARCH ARTICLE

LAE-GSDetect: A Lightweight Fusion Framework for Robust Small-Face Detection in Low-Light Conditions

Bo Chang[1,*], **Lichao Tang**[1,*], **Chen Hu**[1], **Mengxiao Zhu**[1], **Huijie Dou**[1] **and Kharudin Bin Ali**[2]

[1] Faculty of Electronic Engineering, Huaiyin Institute of Technology, Huaian 223003, China
[2] Faculty of Engineering Technology (Electrical and Automation), University College TATI, Kemaman 24000, Malaysia

Abstract

In response to the challenges of insufficient accuracy in face detection and missed small targets under low-light conditions, this paper proposes a detection scheme that combines image preprocessing and detection model optimization. Firstly, Zero-DCE low-light enhancement is introduced to adaptively restore image details and contrast, providing high-quality inputs for subsequent detection. Secondly, YOLOv11n is enhanced through the following improvements: a P2 small-target detection layer is added while the P5 layer is removed, addressing the original model's deficiency in detecting small targets and streamlining the computational process to balance model complexity and efficiency; the P2 upsampling is replaced with DySample dynamic upsampling, which adaptively adjusts the sampling strategy based on features to improve the accuracy of feature fusion; a lightweight adaptive extraction module (LAE) is incorporated to reduce the number of parameters and computational costs; finally, the detection head is replaced with GSDetect to maintain accuracy while reducing computational overhead. Experimental results show that the improved model reaches an mAP50 of 58.7%, which is an improvement of 10.2 percentage points compared with the original model. Although the computational complexity increases to 7.5 GFLOPs, the parameter count is reduced by 45%, offering a more optimal solution for face detection in low-light environments.

Keywords: low-light, face detection, image enhancement, lightweight, deep learning.

1 Introduction

With the swift advancement of intelligent surveillance [1], autonomous driving, face detection under complex illumination [2], and other fields, face detection technology—as core support for human–computer interaction and security protection—is increasingly applied in complex environments and has garnered significant attention [3, 4]. However, in low-light scenarios such as night surveillance and outdoor darkness, captured images often suffer from low visibility, blurred details, dense noise, and color distortion. These issues not only impair human visual perception but also severely degrade the performance of machine vision tasks,

Submitted: 30 September 2025
Revised: 22 November 2025
Accepted: 26 November 2025
Published: 18 December 2025

Vol. 2, **No.** 4, 2025.
10.62762/TSCC.2025.972040

***Corresponding author:**
✉ Bo Chang
changbo@hyit.edu.cn
✉ Lichao Tang
852541073@qq.com

Citation
Chang, B., Tang, L., Hu, C., Zhu, M., Dou, H. & Ali, K. B. (2025). LAE-GSDetect: A Lightweight Fusion Framework for Robust Small-Face Detection in Low-Light Conditions. *ICCK Transactions on Sensing, Communication, and Control*, 2(4), 250–262.

© 2025 ICCK (Institute of Central Computation and Knowledge)

which makes it difficult for traditional face detection models to accurately capture facial features. This often results in missed detections of small targets and an increased false detection rate, substantially limiting the technology's reliability in practical applications.

To address the aforementioned challenges in low-light face detection, existing studies have explored two primary technical approaches: independent enhancement optimization and direct detection model adaptation. On one hand, some studies focus solely on improving the quality of low-light images. For instance, traditional methods such as Retinex-based algorithms [5] and histogram equalization attempt to restore details by adjusting pixel intensity, yet they often struggle to balance noise suppression and feature preservation—either resulting in over-smoothed facial edges or amplified background clutter. Even advanced deep learning-based enhancement methods like EnlightenGAN [6] and Zero-DCE [7] (the baseline method adopted in this study) prioritize global visual naturalness, lacking targeted optimization for facial key regions (e.g., eyes, nose, and mouth) that are crucial for detection. Consequently, the enhanced images may still suffer from blurred local features, failing to provide effective input for subsequent face detection tasks.

On the other hand, other studies aim to enhance the robustness of detection models to low-light conditions without relying on pre-enhancement steps. For instance, some methods introduce multi-scale feature fusion modules to capture weak facial signals. However, these model-only optimizations have inherent limitations: in extreme low-light scenarios where image information is severely degraded, even the most advanced detection architectures struggle to extract discriminative features from raw low-light images. This results in persistently low recall rates for small faces and high false detection rates.

Inspired by these advances, this paper focuses on the face detection task in low-light scenarios and proposes a collaborative optimization method of "enhancement-detection". Firstly, the Zero-DCE algorithm is adopted to adaptively enhance low-light images, which preserves facial details while suppressing noise and improving image contrast. Secondly, aiming at the enhanced images, the YOLOv11n [8] model is improved and optimized: by integrating a dedicated detection layer for small targets, introducing dynamic upsampling and lightweight feature extraction modules, and replacing the detection head, the model's ability to capture tiny faces and distinguish features in low-light environments is strengthened. Experiments are conducted based on the DARK FACE [9] dataset to verify the comprehensive advantages of the proposed method in terms of accuracy, recall and inference efficiency. This study aims to provide a more practical solution for low-light face detection and promote the application of this technology in complex environments.

This paper adopts the "original Zero-DCE + improved YOLOv11n" scheme, whose core advantage lies in balancing "practicality" and "deployability": on one hand, the original Zero-DCE algorithm requires no reference images and has low computational overhead, enabling fast processing of low-light images and avoiding delays caused by complex enhancement algorithms; on the other hand, the lightweight optimization for the detection model (YOLOv11n) does not significantly increase model parameters, allowing the overall workflow to maintain high-efficiency inference characteristics. This scheme not only addresses the insufficient detection accuracy of the traditional "original Zero-DCE + basic model" pipeline but also avoids the pain points of end-to-end models (such as high complexity and difficult deployment), thereby providing a practical solution that is "easy to implement, highly adaptable, and real-time capable" for the application of low-light face detection technology on edge devices with limited resources.

2 Related Work

Low-light enhancement: Low-light images may be caused by insufficient ambient light during shooting or limitations of the camera's own exposure system. In the field of low-light image enhancement, a large number of relevant research results have been accumulated. Histogram equalization and its various variants are capable of expanding the dynamic range of images. The Retinex theory posits that an image can be decomposed into an illumination component and a reflectance component. Rooted in Retinex theory, most relevant studies [10] usually begin by estimating the illumination and reflectance components of low-light images, and then handle these two components in either a separate or concurrent way. In recent years, the advancement of such methods has largely depended on deep learning techniques. Jiang et al. [6] developed the EnlightenGAN approach, a solution built on a generative adversarial network

(GAN) framework that enables low-light image enhancement without the need for paired supervised data. It not only alleviates the problem of insufficient training data but also improves the visual quality of low-light images. Ma et al. [11] focused on creating a rapid, adaptable, and reliable low-light image enhancement method called SCI, aiming to address issues of subpar image quality and low clarity in low-light scenarios, while also satisfying the demands for processing speed, adaptability across different scenes, and stable performance in various low-light environments during practical applications. Guo et al. [7] introduced Zero-DCE, a method for low-light image enhancement to support detection tasks, demonstrating robust performance in various low-light environments practically.

Face Detection: Over the past few years, deep learning has attained significant achievements across various domains, such as object detection [12]. As an important branch of general object detection, face detection has also seen substantial progress. In addition to general object detection algorithms surveyed in [12], such as Faster R-CNN and YOLO [14], which have also been evaluated under challenging low-light conditions [13], specialized face detection algorithms such as DSFD [15], PyramidBox [16], and RetinaFace [17] have been developed, all demonstrating notable success, particularly on the widely-used WIDER FACE benchmark [26].

Low-light Face Detection: Based on the DARK FACE dataset, Wang et al. [18] proposed a combined high-low adaptation (HLA-Face) framework specifically designed for low-light face detection tasks. Through bidirectional low-tier adaptation and multi-task high-tier adaptation, this framework enables the adaptation of normal-light face detection models to low-light scenarios without low-light face annotations, achieving performance superior to traditional methods and close to that of fully supervised models using DARK FACE annotations. Yu et al. [19] proposed a single-stage face detection technique for extremely low-light scenarios, whose core lies in the collaboration of three modules—"image enhancement, detection optimization, and result fusion"—to achieve high-performance detection on the DARK FACE dataset.

3 Image Enhancement Method

In practical scenarios such as night monitoring and intelligent driving, low-light images often suffer from quality degradation issues including insufficient brightness, low contrast, and strong noise: dim frames mask key facial features, grayscale confusion blurs target boundaries, and noise amplification damages details. These problems ultimately cause detection models to fail in feature extraction, increase miss rates, and restrict the practical value of the system. As a preprocessing step, image enhancement technology can improve brightness and contrast while suppressing noise, providing clear and stable input for subsequent detection tasks. This paper compares three mainstream low-light enhancement methods, with visualization results shown in Figure 1.

Figure 1 demonstrates the differences between the methods from the perspectives of visual effect and task adaptability: EnlightenGAN improves brightness but tends to introduce color deviation due to training bias, which disrupts scene consistency and interferes with target discrimination; SCI achieves balanced brightness enhancement but has insufficient detail restoration, leading to blurring of facial edges and background textures; Zero-DCE uses pixel-level curve mapping to reasonably brighten images while accurately preserving details and color consistency, featuring clear facial contours, rich background details, and natural colors. Therefore, Zero-DCE is selected as the low-light image enhancement scheme in this paper due to its advantage of balancing brightness, details, and color.

4 Improved YOLOv11n Object Detection Algorithm

4.1 Overall Framework of the Improved Model

To address the challenges of detecting small faces, insufficient feature fusion, and high computational cost in low-light environments, this paper proposes an optimized YOLOv11n architecture with four key improvements. First, the feature pyramid structure is adjusted: a P2 feature layer dedicated to small-scale faces is added, while the P5 layer is removed. This balances model complexity and computational efficiency, enabling the model to better capture fine-grained details of small faces and avoid unnecessary computations from the less relevant P5 layer. Second, the conventional upsampling method in the P2 layer is replaced with the dynamic upsampling operator DySample [20], which generates content-aware sampling kernels to enhance feature fusion accuracy and retain precise spatial information for face detection. Third, some ordinary convolutions in the network are substituted with the lightweight

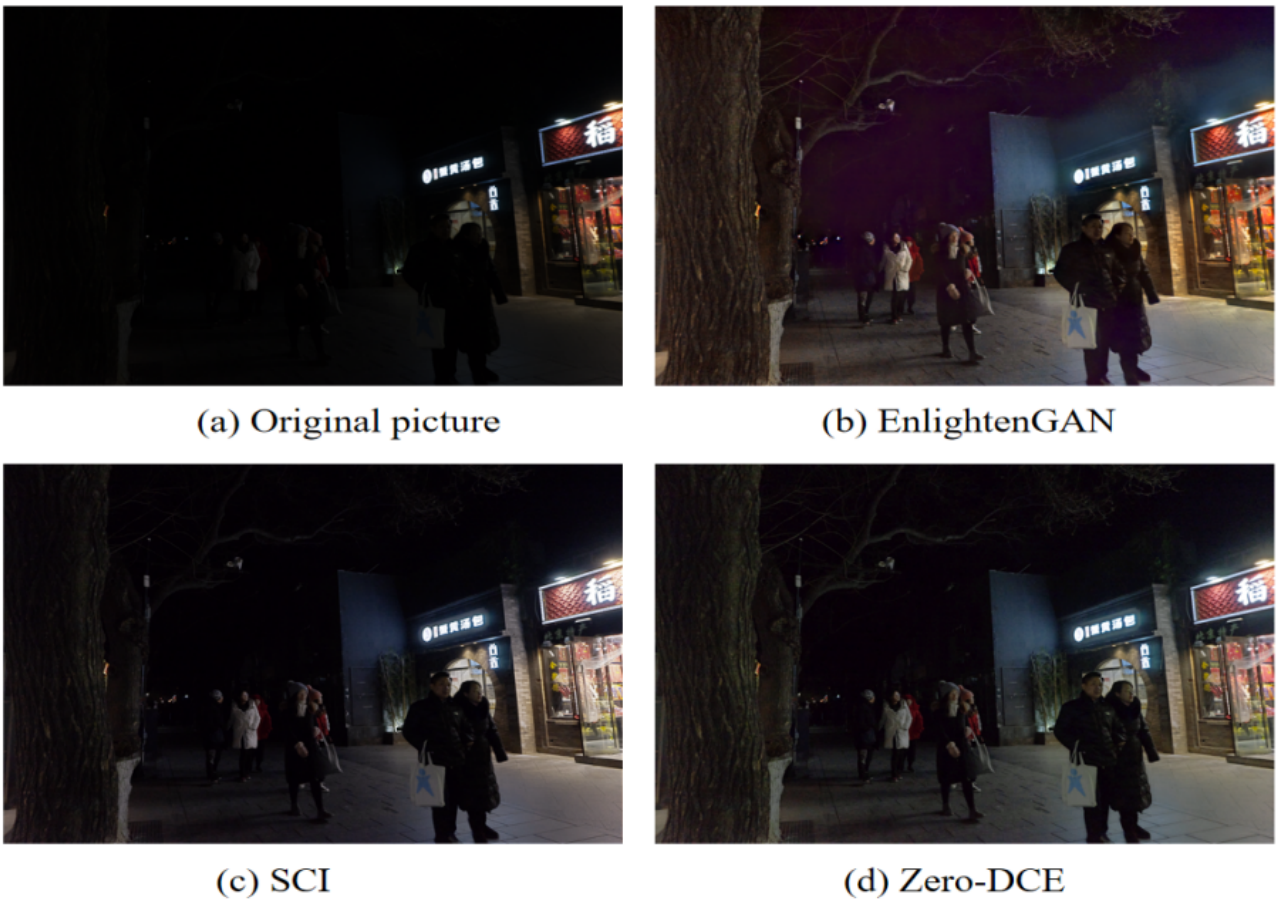


Figure 1. Visualization results of the original image and various image enhancement methods.

adaptive extraction module (LAE) [21], significantly reducing parameters and computational load while preserving strong feature representation capabilities by focusing on discriminative facial features. Finally, the original detection head is replaced with the lightweight GSDetect head, which reduces computational overhead via efficient parameter sharing while maintaining accuracy, rendering it applicable to real-time low-light face detection. The improved YOLOv11n's architecture is shown in Figure 2. Experiments on the DARK FACE dataset demonstrate that the optimized model outperforms the baseline YOLOv11n in small face detection accuracy and inference speed, verifying the improvements' effectiveness.

4.2 Adding a Small-Target Layer

In low-light face detection tasks, affected by insufficient lighting and environmental occlusion, small-sized faces under low-light conditions constitute a high proportion in the dataset. Even though the standard YOLOv11n multi-scale detection system is able to process targets of diverse sizes, it shows limitations in detecting small-sized faces under low-light conditions. Its P3, P4, and P5 output layers correspond to the detection of targets at different scales, with the P5 layer focusing on large targets, making it suitable for scenarios with large objects in the background. However, due to the high downsampling factor of the P5 layer, it struggles to retain sufficient detail when processing small-sized faces under low light, and also suffers from computational redundancy, resulting in suboptimal performance in detection tasks where low-light, small-sized faces are the core focus.

To address this, this paper optimizes the detection layer structure by adding a P2 detection layer specifically designed for small-sized faces under low-light conditions and eliminating the P5 large-target detection layer. The P2 layer increases the spatial definition, generating a 160×160 resolution feature map. As opposed to the 20×20 feature map from the P5 layer, it preserves more detailed information of small-sized faces under low light, thereby improving detection accuracy. Meanwhile, the P2 layer undergoes feature fusion with the P3 and P4 layers, enhancing the model's multi-scale feature extraction capability. This ensures precise capture of details for small-sized

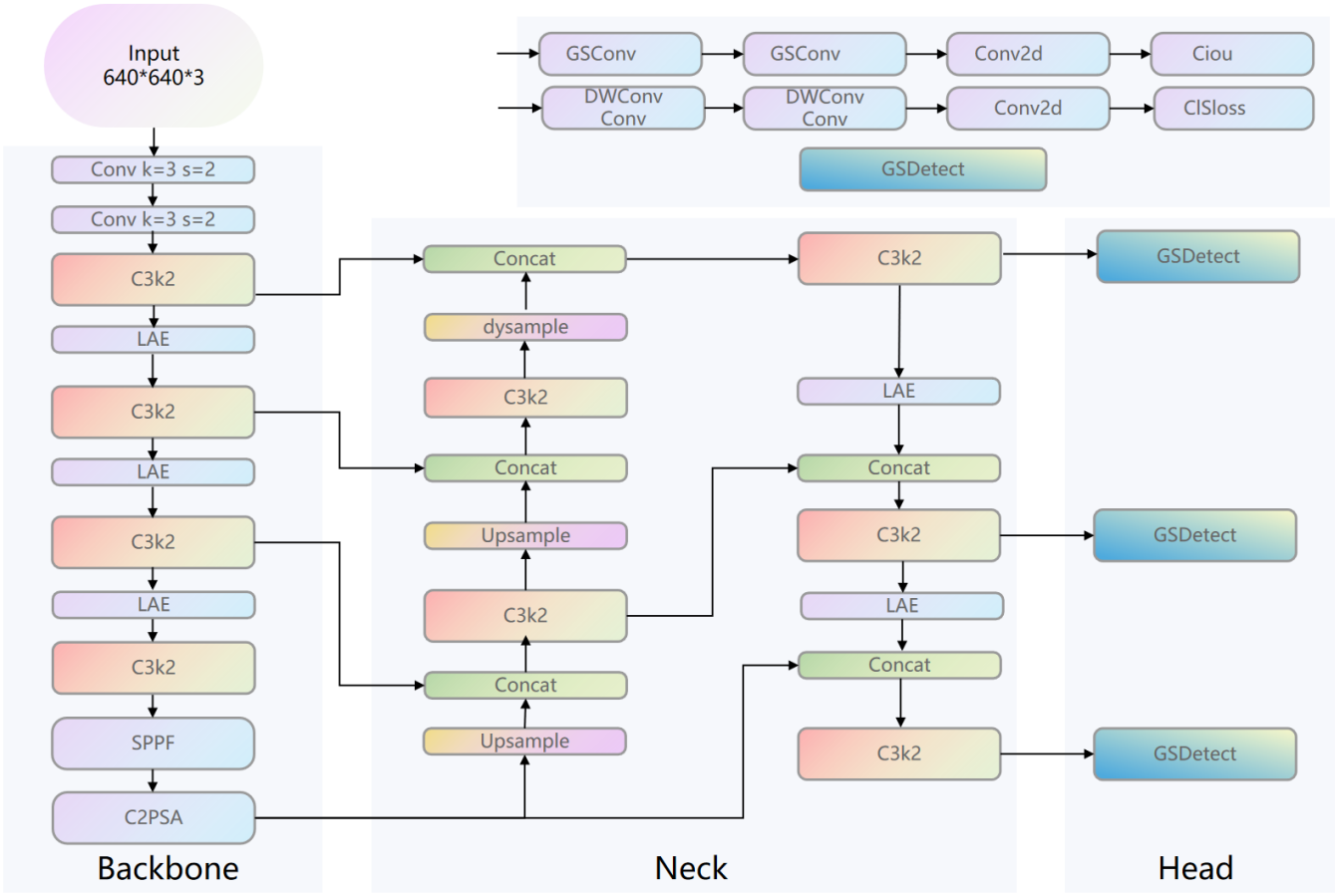


Figure 2. Architecture of the improved YOLOv11n network.

faces under low light while avoiding redundant computations. The structure is illustrated in Figure 3.

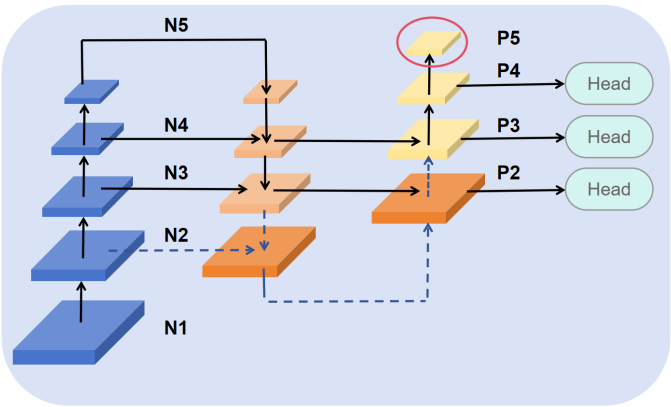


Figure 3. Schematic diagram of the improved detection layers.

4.3 Lightweight Dynamic Upsampling Operator

In low-light small target processing, static upsampling often struggles to accurately adapt to the irregular distribution of facial edges due to its fixed sampling patterns, leading to detail loss. In contrast, DySample employs dynamic point sampling, which can adaptively adjust the positions and weights of sampling points based on the actual distribution of edge details in the feature map, thereby enhancing the utilization efficiency of subtle contours and textures of faces under low-light conditions. To improve the capability of capturing edge details of small-sized faces in low-light scenarios, the dynamic upsampling operator DySample is introduced to replace the original static upsampling structure in the P2 layer. The dynamic upsampling based on sampling and the module design in DySample are illustrated in Figure 4.

As shown in Figure 4 (a), it resamples the given input feature map X of size $C \times H \times W$ and computes a new upsampled feature map via bilinear interpolation, resulting in an output size of $C \times SH \times SW$. The network upsampling can be defined as:

$$X' = \text{grid_sample}(X, S) \tag{1}$$

Figure 4 (b) illustrates two versions: the "static factor scope" and the "dynamic factor scope". Considering that facial image features under low light are prone to blurring, suffer from significant noise interference, and exhibit complex and variable lighting distributions,

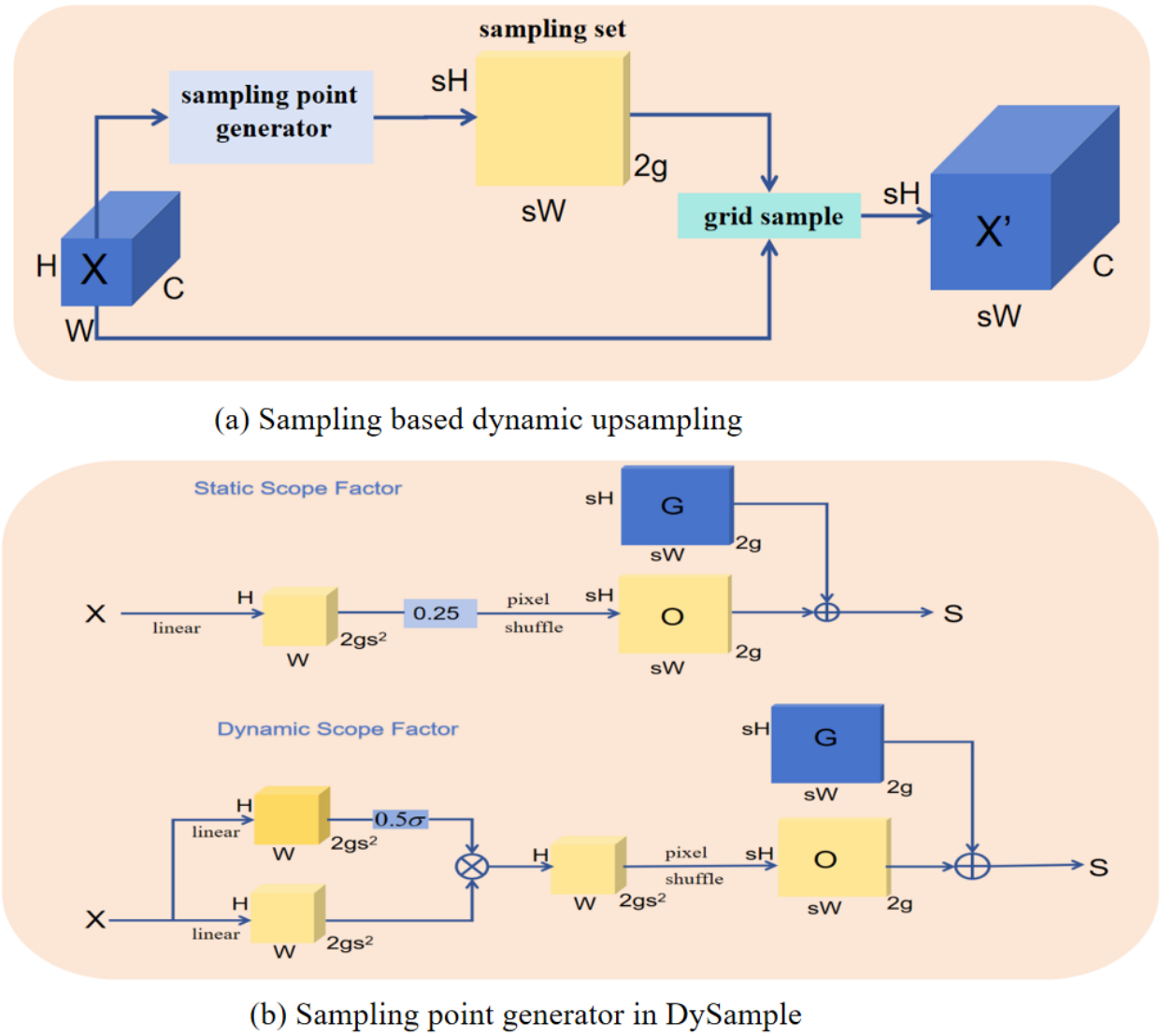


Figure 4. Dynamic upsampling and module architectures in DySample based on sampling. (a) Sampling based dynamic upsampling. (b) Sampling point generator in DySample.

the dynamic factor version can adaptively adjust processing strategies to capture facial information more accurately. Therefore, the dynamic factor version is selected.

The input feature map X with dimensions $H \times W \times C$ is first processed through two parallel linear layers, yielding branch features X_1 and X_2, both of size $H \times W \times 2gs^2$. The intermediate feature from the upper branch is passed through a sigmoid function σ and then scaled by 0.5 to produce a dynamic range factor 0.5σ. This factor is subsequently fused with the intermediate feature from the lower branch via element-wise multiplication, thereby adaptively modulating the features in the lower branch. The fused features then undergo a pixel rearrangement operation to generate the dynamic offset O. Finally, the offset O is added element-wise to the grid position G to form the final sampling set S. The fundamental implementation of this process can be defined as follows:

$$O = \text{linear}(X) \tag{2}$$

$$S = G + O \tag{3}$$

Compared with other dynamic upsamplers, DySample is designed from the perspective of point sampling, requiring no additional CUDA packages. It leverages highly optimized PyTorch operations to perform fast backpropagation, adding almost no extra training time or computational cost. As a result, DySample offers significant advantages in terms of inference speed, memory usage, and accuracy.

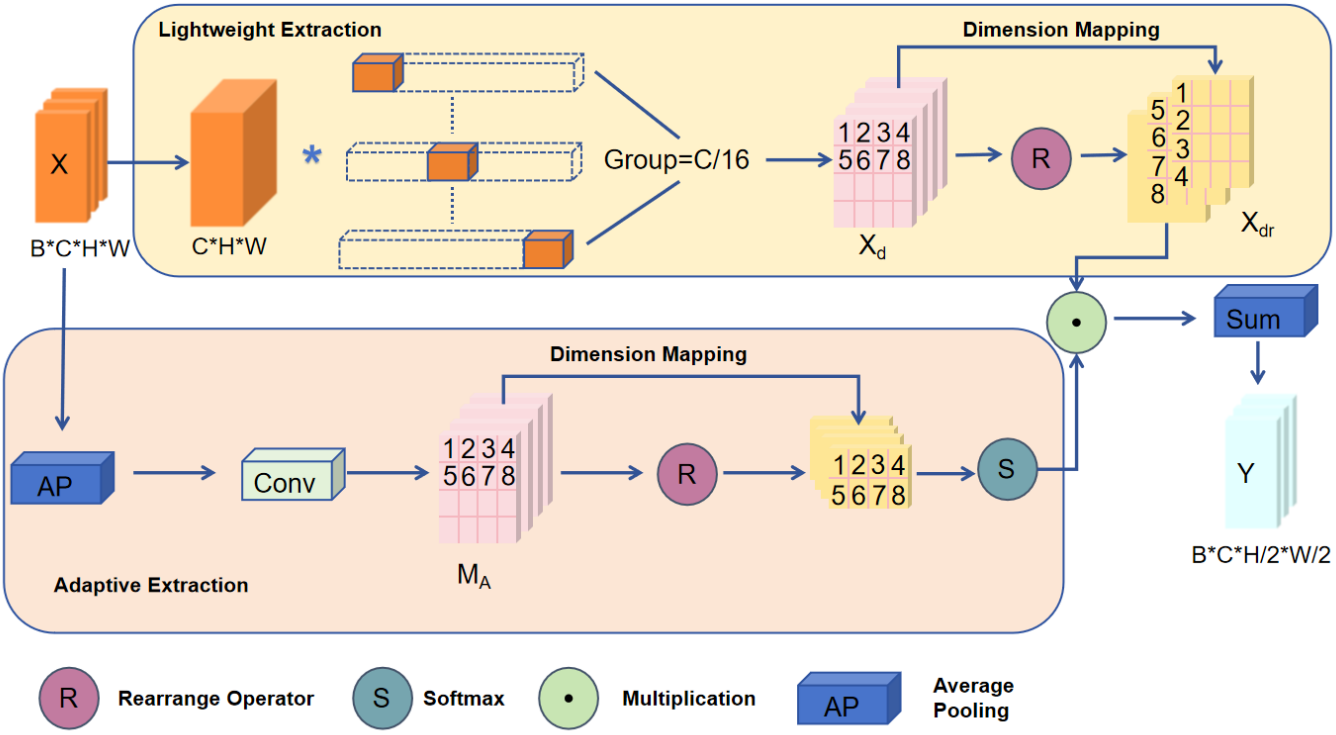


Figure 5. Architecture of the Lightweight Adaptive Extraction (LAE) module.

4.4 Lightweight Adaptive Extraction Module

To address the issues of excessive parameters and high computational cost resulting from stacked standard convolutional layers, this paper introduces the Lightweight Adaptive Extraction (LAE) module as a replacement for some standard convolutions in the YOLOv11n model, aiming to maintain detection accuracy while improving efficiency.

As illustrated in Figure 5, the LAE module employs a dual-branch parallel architecture that elegantly integrates the principles of parameter sharing and group convolution for efficient feature extraction and channel-wise information weighting. One branch is designed to transform spatial information (from height and width dimensions) into channel-wise representations. By utilizing group convolution, this branch reduces the number of parameters to 1/N of that in standard convolution, thereby substantially reducing computational complexity. The other branch extracts global features via average pooling and employs a lightweight convolutional layer to generate dynamic weights, which recalibrate the importance of different feature channels. This mechanism effectively mitigates the loss of edge details during downsampling. Finally, the module performs a weighted fusion of the outputs from both branches to form a comprehensive feature representation. The computational procedure is as follows:

The input feature map $X \in \mathbb{R}^{B\times C\times H\times W}$, where B denotes the batch size, C represents the number of channels, and H and W denote the height and width, respectively, is first processed through average pooling and a 1×1 convolution to generate the attention feature M_A.

$$M_A = \text{Conv}_{1\times1}[\text{AvgPool}(X)] \tag{4}$$

Through rearrangement and Softmax normalization, the attention weights Z are obtained:

$$Z = \text{Softmax}[\text{Rearrange}(M_A)] \tag{5}$$

where X is downsampled via grouped convolution to generate the downsampled feature X_d, which is then rearranged into X_{dr}:

$$X_d = G\text{Conv}(X) \tag{6}$$

$$X_{dr} = \text{Rearrange}(X_d) \tag{7}$$

Finally, the downsampled features are combined with the corresponding attention weights via a weighted summation to generate the output feature Y:

$$Y = \sum_{i=1}^{4}(X_{dr}^{i}\cdot Z^{i}) \tag{8}$$

4.5 Improved Detection Head

4.5.1 Grouped Separable Convolution

GSConv [22] (Grouped Separable Convolution) cleverly adopts a dual-branch architecture. It first splits the input channels into two parts: one branch uses standard convolution to extract channel-correlated features, while the other branch employs depthwise separable convolution to capture spatial details. The features from the two branches are then fused via concatenation, followed by a shuffle operation to break the channel isolation problem often caused by depthwise separable convolution. This allows full interaction between different channels, significantly reducing computational cost while effectively preserving feature representation capability. As a result, GSConv achieves a balance between lightweight design and accuracy, offering an innovative approach for efficient neural network design. The structure of GSConv is illustrated in Figure 6.

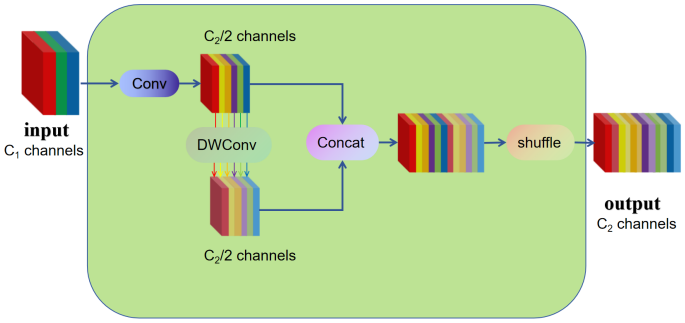


Figure 6. Schematic diagram of the GSConv structure.

4.5.2 Detection Head Improvements

In the optimization of the YOLOv11n detection head, replacing the first two conventional convolutions with GSConv yields multiple notable advantages. Unlike traditional convolutions that often incur high computational costs, GSConv features an innovative dual-branch architecture: one branch retains standard convolution to ensure robust core feature extraction, while the other employs depthwise separable convolution to drastically cut down on redundant calculations. This balanced design not only preserves the model's detection accuracy but also significantly lightens its computational burden, enabling more efficient real-time detection on edge devices or mobile platforms with constrained hardware resources—critical for computationally sensitive applications like multi-channel video surveillance and real-time autonomous driving perception. Moreover, GSConv integrates a unique channel shuffle mechanism, which effectively addresses the channel isolation limitation inherent in depthwise separable convolution. By facilitating sufficient cross-channel feature interaction, it overcomes the drawback of single-branch convolutions that struggle to capture inter-channel correlations. Compared with a single convolution mode, this mechanism allows the model to extract richer discriminative features, ranging from fine-grained details to global context, thereby enhancing its adaptability to complex scenarios such as crowded public spaces or variable lighting environments. The structural design of the improved detection head after this optimization is visually illustrated in Figure 7.

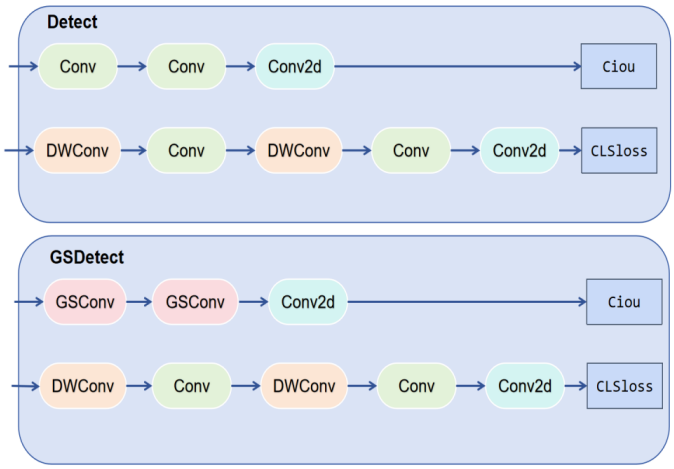


Figure 7. Schematic diagram of the improved detection head.

5 Experimental Results and Analysis

5.1 Dataset and Experimental Environment

The DARK FACE dataset is a widely employed benchmark for low-light face detection; therefore, this paper adopts it as the test dataset. This dataset provides 6,000 real-world low-light images captured in various settings such as at night, near teaching buildings, on streets, around bridges, over overpasses, and in parks. All these images are annotated with bounding boxes for faces and serve as the primary training and/or validation set. Additionally, it contains 9,000 unlabeled low-light images collected under the same conditions. It also features a unique set of 789 low-light/normal-light image pairs, captured under controlled real lighting conditions (though not necessarily containing faces), which can be utilized as part of the training data at the participants' discretion. A held-out test set comprising 4,000 low-light images with bounding box annotations for faces is also provided. Since the original test set labels are not publicly available, this paper randomly splits the

provided 6,000 labeled images into training, validation, and test sets in a 6:2:2 ratio. Note that this partitioning differs from the official competition split, which limits direct quantitative comparison with results reported in prior works that follow the official protocol.

Figure 8 presents the statistical analysis of face count and resolution conducted on the 6,000 images provided by the DARK FACE dataset. The experimental statistics on the distribution characteristics of the face dataset reveal that low-resolution faces (with dimensions below 300 pixels) overwhelmingly dominate the dataset, as shown in Figure 8 (a), indicating that small-resolution faces (small targets) constitute the majority. Figure 8 (b) demonstrates that the number of images containing 6 to 10 faces per image is the highest, while the count drops sharply when the number of faces exceeds 20. This reflects that most images contain a concentrated number of faces, with extreme multi-face scenarios being rare. Given that small-resolution faces (small targets) form the primary composition of the dataset, subsequent algorithms need to prioritize adaptation for small target detection. The detection performance will be evaluated using the mean Average Precision (mAP) metric.

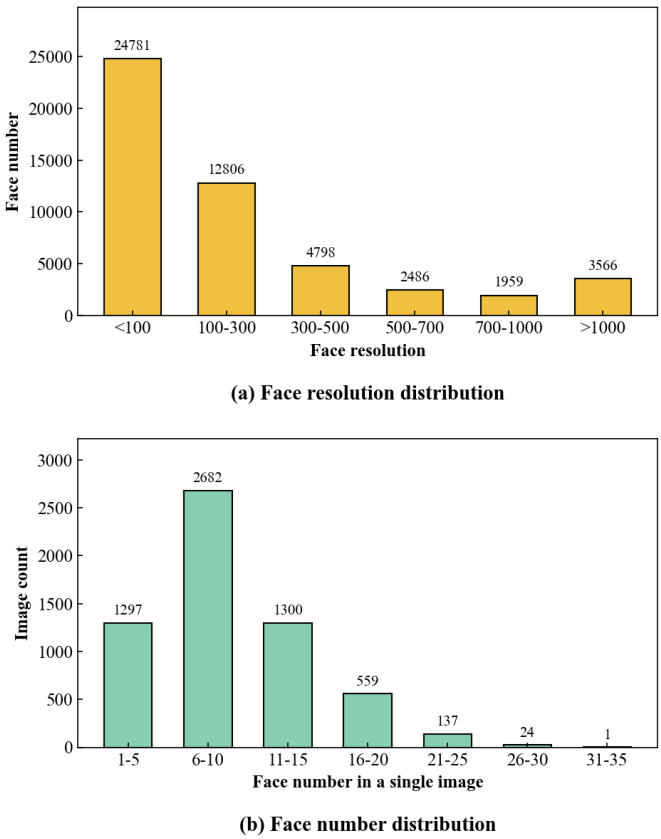


Figure 8. Distribution of face resolution and count in the DARK FACE dataset.

The experimental environment was configured as follows: the operating system was Ubuntu 22.04; the deep learning framework used was PyTorch 2.0.1; the programming language was Python 3.9; and the CUDA version was 11.8. Model training was accelerated using an NVIDIA GeForce RTX 4090 GPU. The training batch size was set to 8, the SGD optimizer was utilized with an initial learning rate of 0.01, and the training process was conducted for 300 epochs.

5.2 Evaluation Metrics

The performance of the network is evaluated using precision (P), recall (R), mean average precision (mAP), number of parameters (Params), and giga floating-point operations (GFLOPs). The relevant formulas are as follows:

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{9}$$

$$AP = \int_0^1 P(r)dr \tag{10}$$

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

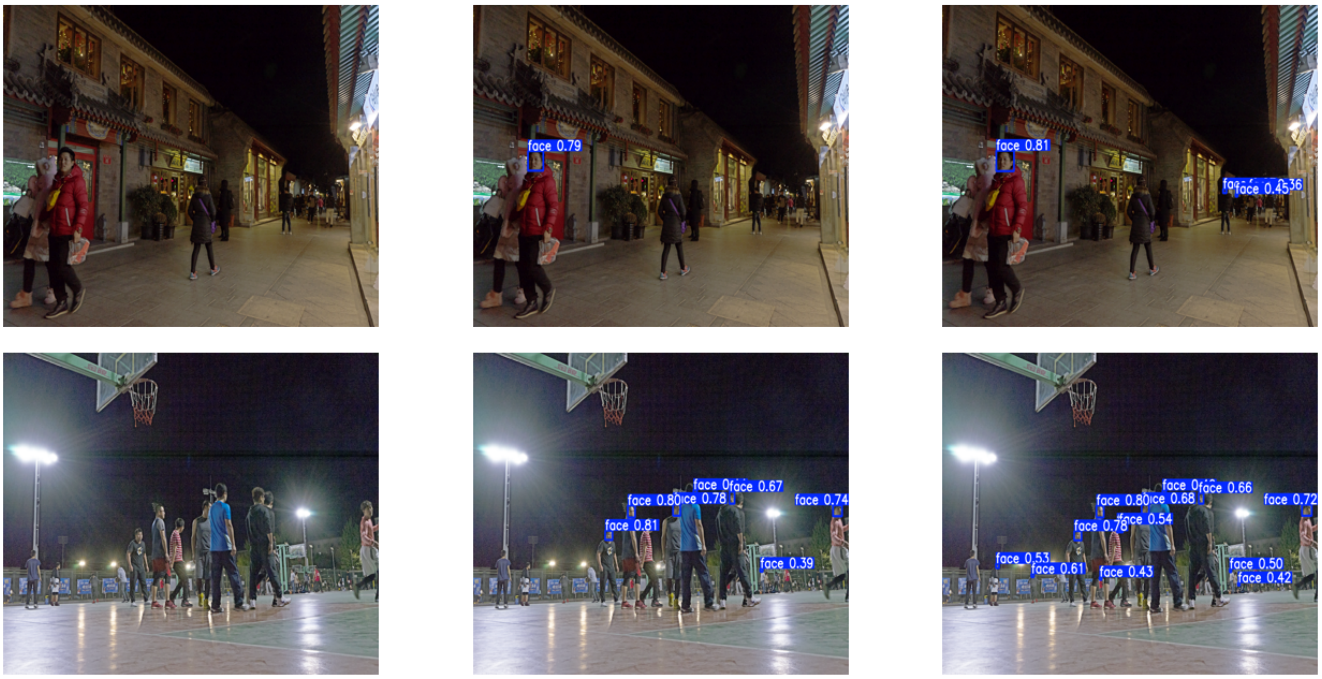
where TP refers to the count of correctly predicted positive samples, FN denotes the number of positive samples incorrectly predicted as negative, FP represents the number of negative samples mistakenly predicted as positive, and n is the total number of samples.

Table 1. Accuracy comparison of different image enhancement methods.

Enhancement method	mAP50(%)
—	45.7
EnlightenGAN	46.8
SCI	47.2
Zero-DCE	48.5

5.3 Comparison of Various Image Enhancement Approaches

In the third chapter, the visualization results of various image enhancement methods are presented. To further investigate the impact of different image enhancement techniques on model accuracy, we designed corresponding experiments, using the YOLOv11n model without any image enhancement



(a) Image to be detected (b) YOLOv11n (c) Improved model

Figure 9. Detection performance comparison between YOLOv11n and the improved model.

method as the benchmark for comparative analysis. The experimental results are shown in Table 1, where the "—" entry indicates that no image enhancement method was used. According to the data in Table 1, all three enhancement methods—EnlightenGAN, SCI, and Zero-DCE—improved the accuracy. Among them, Zero-DCE achieved the most significant optimization because it can accurately correct illumination and suppress noise, thereby providing clearer features for detection. SCI ranked second, demonstrating advantages in detail restoration. Although EnlightenGAN also brought improvements, the enhanced images tended to retain shadows, which to some extent interfered with detection. In summary, Zero-DCE shows better adaptability for low-light face detection tasks. Future work could focus on in-depth research into the collaborative optimization of Zero-DCE with detection models to further enhance performance.

5.4 Ablation Experiments

To validate the impact of different improvement modules on model performance, an ablation study was conducted on the experimental dataset, and the results are shown in Table 2. Here, "Params" denotes the number of parameters, and "FLOPs" represents the computational complexity. When no improvement modules were introduced, the baseline model (YOLOv11n with Zero-DCE enhancement but without any structural modifications) achieved an mAP50 of 48.5%, with parameters and FLOPs reaching 2.58×10^6 and 6.3×10^9, respectively. After adding

Table 2. Ablation experiments.

No.	Detection Layer	DySample	LAE	Detection Head	mAP50/%	Params/M	FLOPs/G
1	×	×	×	×	48.5	2.58	6.3
2	√	×	×	×	58.7	1.93	9.6
3	×	×	√	×	48.3	2.09	6.0
4	×	×	×	√	48.2	2.40	5.6
5	√	√	×	×	59.4	1.93	9.7
6	√	√	√	×	59.2	1.54	9.4
7	√	√	√	√	58.7	1.42	7.5

the detection layer, mAP50 increased to 58.7%, while parameters decreased to 1.93×10^6 due to structural simplification; however, FLOPs increased to 9.6×10^9 as a result of the additional computations introduced by this layer. When only the LAE module was utilized, mAP50 slightly dropped to 48.3%, parameters became 2.09×10^6, and FLOPs reduced to 6.0×10^9. When only the detection head was modified, mAP50 was 48.2%, parameters were 2.40×10^6, and FLOPs were 5.6×10^9. In the fifth experimental setup, on the basis of adding the detection layer, replacing the upsampling in the P2 layer with DySample led to more refined feature processing, improving mAP50 by an additional 0.7 percentage points to 59.4%, with FLOPs slightly rising to 9.7×10^9 and parameters remaining unchanged at 1.93×10^6. In the sixth setup, replacing standard convolutional layers with the LAE module on top of having the detection layer and DySample further enhanced the model's lightweight characteristics, reducing parameters to 1.54×10^6 and FLOPs to 9.4×10^9, while mAP50 remained stable at 59.2% despite minor fluctuations. In the seventh setup, replacing the detection head on the basis of the previous modules resulted in a slight decrease in mAP50 to 58.7%, but parameters and FLOPs were significantly reduced to 1.42×10^6 and 7.5×10^9, respectively. The optimized model demonstrates greater suitability for deployment on edge computing devices and effectively addresses the demands of complex low-light scenarios.

5.5 Comparative Experiments

To validate the advancement of the improved YOLOv11n model, it was compared with mainstream object detection models—YOLOv9t [23], YOLOv10n [24], the original YOLOv11n, YOLOv11s, and RT-DETR [25]—by training and conducting a comparative analysis on the DARK FACE dataset. The experimental results are shown in Table 3. As can be seen from the table, the improved YOLOv11n achieved a key mAP50 of 58.7%, demonstrating a significant improvement in detection accuracy compared to other models such as YOLOv9t, YOLOv10n, and the original YOLOv11n. Meanwhile, the improved YOLOv11n model has only 1.42M parameters and a computational cost of 7.5 GFLOPs. Compared to models like RT-DETR-l, which has 31.9M parameters and a computational cost of 103.4 GFLOPs, the improved YOLOv11n maintains superior parameter scale and computational overhead, effectively balancing detection accuracy and computational efficiency, thus exhibiting excellent performance equilibrium. It should be noted that, due to differences in dataset partitioning and training protocols, a direct quantitative comparison with specialized low-light face detection methods such as HLA-Face [18] and the single-stage approach of Yu et al. [19] is not conducted in this work, which represents a limitation to be addressed in future research.

Table 3. Comparative experimental results of different models.

Model	mAP50/%	Params/M	FLOPs/G
YOLOv9t	47.9	1.97	7.6
YOLOv10n	48.2	2.69	8.2
YOLOv11n	48.5	2.58	6.3
YOLOv11s	53.4	9.41	21.3
RT-DETR-l	50.6	31.9	103.4
RT-DETR-r50	49.7	41.9	125.6
Improved YOLOv11n	58.7	1.42	7.5

5.6 Visualization of Improved Results

Figure 9 presents a visual comparison of the detection results to intuitively demonstrate the performance advantages of the improved model. As shown in the Figure 9, compared to YOLOv11n, the improved model exhibits superior performance in the face detection task: the original YOLOv11n had a high missed detection rate, particularly for small, distant face targets in the scene, where numerous detections were missed. In contrast, the improved model detects a greater number of faces, effectively reducing the missed detection rate. For instance, in a street scenario, YOLOv11n detected only one face, whereas the improved model successfully identified faces located further in the distance. Similarly, in a basketball court scenario, the number of faces detected by the improved model is significantly higher than that detected by YOLOv11n. This indicates that the improved model can more comprehensively capture face targets within the image, demonstrating an overall more outstanding performance in the face detection task.

6 Conclusion

This paper addresses the challenges of low accuracy and high miss rates for small targets in face detection under low-light conditions by proposing a method that integrates Zero-DCE enhancement with an optimized YOLOv11n model. First, Zero-DCE is applied to low-light images to improve brightness and contrast, enhancing facial feature visibility and yielding an accuracy improvement of 2.8 percentage points compared with using the original

images. Subsequently, the YOLOv11n architecture is enhanced with several key modifications: a P2 layer dedicated to small-target detection, the DySample dynamic upsampling mechanism, a Lightweight Adaptive Extraction (LAE) module, and a GSDetect head. Evaluated on the DARK FACE dataset, the improved model achieves an mAP50 of 58.7%, an improvement of 10.2 percentage points over the original YOLOv11n, while reducing parameters by 45% and maintaining computational complexity at 7.5 GFLOPs. This solution achieves an effective balance between accuracy and efficiency, offering a practical and efficient approach for face detection under low-light conditions.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported in part by the National Natural Science Foundation of China under Grant 62205120; in part by the Key Research and Development Program of Tianjin, China under Grant 22YFZCSN00210.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Wang, W., Wang, X., Yang, W., & Liu, J. (2022). Unsupervised face detection in the dark. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(1), 1250-1266. [CrossRef]

[2] Cho, S. W., Baek, N. R., Kim, M. C., Koo, J. H., Kim, J. H., & Park, K. R. (2018). Face detection in nighttime images using visible-light camera sensors with two-step faster region-based convolutional neural network. *Sensors, 18*(9), 2995. [CrossRef]

[3] Jiang, B., Jiang, H., Zhang, H., Zhang, Q., Li, Z., & Huang, L. (2024). 4AC-YOLOv5: an improved algorithm for small target face detection. *EURASIP Journal on Image and Video Processing, 2024*(1), 10. [CrossRef]

[4] Zhu, C., Tao, R., Luu, K., & Savvides, M. (2018, June). Seeing Small Faces from Robust Anchor's Perspective. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5127-5136). IEEE. [CrossRef]

[5] Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., & Zhang, Y. (2023, October). Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 12470-12479). IEEE. [CrossRef]

[6] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., ... & Wang, Z. (2021). Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing, 30*, 2340-2349. [CrossRef]

[7] Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., & Cong, R. (2020, June). Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1777-1786). IEEE. [CrossRef]

[8] Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*. [CrossRef]

[9] Yang, W., Yuan, Y., Ren, W., Liu, J., Scheirer, W. J., Wang, Z., ... & Qin, L. (2020). Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Transactions on Image Processing, 29*, 5737-5752. [CrossRef]

[10] Fu, X., Zeng, D., Huang, Y., Liao, Y., Ding, X., & Paisley, J. W. (2016). A fusion-based enhancing method for weakly illuminated images. *Signal Processing, 129*, 82–96. [CrossRef]

[11] Ma, L., Ma, T., Liu, R., Fan, X., & Luo, Z. (2022, June). Toward Fast, Flexible, and Robust Low-Light Image Enhancement. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5627-5636). IEEE. [CrossRef]

[12] Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems, 30*(11), 3212-3232. [CrossRef]

[13] Shovo, S. U. A., Abir, M. G. R., Kabir, M. M., & Mridha, M. F. (2024). Advancing low-light object detection with you only look once models: An empirical study and performance evaluation. *Cognitive Computation and Systems, 6*(4), 119-134. [CrossRef]

[14] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016, June). You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 779-788). IEEE. [CrossRef]

[15] Li, J., Wang, Y., Wang, C., Tai, Y., Qian, J., Yang, J., ... & Huang, F. (2019, June). DSFD: Dual Shot Face Detector. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5055-5064). IEEE. [CrossRef]

[16] Tang, X., Du, D. K., He, Z., & Liu, J. (2018, September). PyramidBox: A Context-Assisted Single Shot Face Detector. In *European Conference on Computer Vision* (pp. 812-828). [CrossRef]

[17] Deng, J., Guo, J., Ververas, E., Kotsia, I., & Zafeiriou, S. (2020, June). RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5202-5211). IEEE. [CrossRef]

[18] Wang, W., Yang, W., & Liu, J. (2021, June). HLA-Face: Joint High-Low Adaptation for Low Light Face Detection. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 16190-16199). IEEE. [CrossRef]

[19] Yu, J., Hao, X., & He, P. (2021, October). Single-stage Face Detection under Extremely Low-light Conditions. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (pp. 3516-3525). IEEE. [CrossRef]

[20] Liu, W., Lu, H., Fu, H., & Cao, Z. (2023, October). Learning to Upsample by Learning to Sample. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 6004-6014). IEEE. [CrossRef]

[21] Yu, Z., Guan, Q., Yang, J., Yang, Z., Zhou, Q., Chen, Y., & Chen, F. (2024, December). Lsm-yolo: A compact and effective roi detector for medical detection. In *International Conference on Neural Information Processing* (pp. 30-44). Singapore: Springer Nature Singapore. [CrossRef]

[22] Li, H., Li, J., Wei, H., Liu, Z., Zhan, Z., & Ren, Q. (2024). Slim-neck by GSConv: A lightweight-design for real-time detector architectures. *Journal of Real-Time Image Processing, 21*(3), 62. [CrossRef]

[23] Wang, C. Y., Yeh, I. H., & Mark Liao, H. Y. (2024, September). Yolov9: Learning what you want to learn using programmable gradient information. In *European conference on computer vision* (pp. 1-21). Cham: Springer Nature Switzerland. [CrossRef]

[24] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., & Han, J. (2024). Yolov10: Real-time end-to-end object detection. *Advances in Neural Information Processing Systems, 37*, 107984-108011.

[25] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., ... & Chen, J. (2024, June). DETRs Beat YOLOs on Real-time Object Detection. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 16965-16974). IEEE. [CrossRef]

[26] Yang, S., Luo, P., Loy, C. C., & Tang, X. (2016, June). WIDER FACE: A face detection benchmark. In *2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5525-5533). IEEE. [CrossRef]



Bo Chang received her B.E. degree in Electronic Engineering from Shaanxi Institute of Technology, China, in 1994, and her M.E. degree in Communication Engineering from Southeast University, China, in 2007. She is currently an Associate Professor at Jiangsu Huaiyin Institute of Technology, China. Her main research interests include signal processing, detection technology, and applications of the Internet of Things. She is also engaged in teaching and research in communication engineering. (Email: changbo@hyit.edu.cn)



Lichao Tang received his B.E. degree in Electronic Information Technology from Huaiyin Institute of Technology, China, in 2022, and is currently pursuing the M.E. degree in Electronic Information at Huaiyin Institute of Technology, China. (Email: 852541073@qq.com)



Kharudin Bin Ali received the Diploma degree in industrial automation (mechatronics) and the bachelor's degree (Hons.) in engineering technology (mechatronics) from the TATI University College, in 2005 and 2012, respectively, the M.Sc. degree in mechatronic engineering from the University Malaysia Pahang, in 2017, and the Ph.D. degree from University Tenaga Nasional (UNITEN), Malaysia, in 2020. He is a Professional Technologist, in 2018. He has published a number of 33 papers in ISI and Scopus indexed journals and international conferences and supervised around 85 graduate and two postgraduate (master's) students. He also published one book in eddy current inspection.

His current research interests include non-destructive testing, sensor system design, embedded systems, the Internet of Things, and artificial intelligence. He has vast experience in industrial, teaching, and training since 2004. His industrial experience includes ESCATEC Mechatronics and TATI University College. The author has also received several awards, including the Best Paper Award at the ICON 2013 Conference and IGRAD 2018 Conference, and the Excellent Worker Award from TATI University College, in 2012 and 2018.(Email: kharudin@uctati.edu.my)