Research on night vision pedestrian detection algorithm by incorporating attention mechanism

Jiaxue Yang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 2461943985@qq.com

Pingping Liu
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 1341369601@qq.com

***Abstract*—Addressing the common challenges in night vision imagery — poor lighting conditions, low pixel resolution, and diminished contrast — which hinder effective pedestrian feature extraction and result in suboptimal accuracy and real-time performance for night-time pedestrian detection, This paper proposes a deep learning-based night vision pedestrian detection system. Building upon the YOLOv8 object detection algorithm, the model is enhanced by incorporating the CBAM attention mechanism into its network architecture and upgrading the optimiser from SGD to Lion. The system design and development are further tailored to address the specific characteristics of night-time imagery. After experimental simulation verification, the performance of the improved algorithm model has been significantly improved: the overall accuracy is improved by about 2.0%, mAP@0.5 is improved by 1.6%, the average accuracy of IoU threshold 0.5 to 0.95 is improved by about 0.04%, and the F1 Score is improved by 0.64%. The improvement plan proposed in this paper effectively enhances the model's comprehensive identification ability of night vision pedestrians, improves the overall performance of the system, and verifies the correctness and validity of the research.**

Keywords-Deep Learning; Target Detection; Night Vision Pedestrian Detection; Yolov8; Attention Mechanism

I. INTRODUCTION

With the continuous improvement of living standards and the continuous progress of automobile technology, urban traffic safety has become one of the problems affecting the survival that can not be ignored, and pedestrian detection technology has gradually become a key technology in the field of traffic safety, however, in the complex environment of low light and noise interference at night, the pedestrian detection is facing a huge challenge [1]. Moreover, a large amount of data shows that the occurrence rate of traffic accidents is high at night, and the accuracy and effectiveness of pedestrian detection plays a very important role in preventing safety accidents and improving the safety of pedestrians travelling at night, The traditional pedestrian detection methods mainly rely on the processing of visible light images, but in the low-light and complex conditions at night, due to the lack of light conditions, which affects the effectiveness of the monitoring system, which leads to a serious decline in the accuracy of pedestrian detection at night.

In recent years, deep learning technology has made breakthrough progress in the field of computer vision, and it has also been widely applied. In the field of computer vision, R-CNN [2], Faster R-CNN [3], SSD [4], and YOLO [5] are the classic algorithms applied to target detection, however, the YOLO series of algorithms, because of its high efficiency of learning ability, has shown a great success in the field of target detection. The YOLOv8 algorithm shows significant performance advantages in the field of target detection, in which the YOLOv8 algorithm has a relatively large improvement in target detection accuracy compared with the previous YOLO algorithm, and performs better in the related target detection tasks, so it provides a new technical path for the pedestrian target detection task in complex environments at night [6].

DOI: 10.2478/ijanmc-2025-0033

II. Night Vision Pedestrian Detection Algorithm

A. YOLO algorithm

The YOLO algorithm is a single-stage target detection algorithm that can directly predict the category and location of a target in an image in a single forward propagation, which can directly predict the bounding box of a target as well as the category probability by virtue of a single forward propagation, and its core design concept is to transform the target detection problem into a regression problem.

B. Pedestrian Image Detection

The pedestrian detection process of YOLOv8 is to first scale the input image to a specific fixed size, and then normalize the image, i.e., scale the pixel values to the range of [0, 1], and then extract multi-level features from the input image with the help of a backbone network, which outputs a feature map that is subject to multi-scale feature fusion by FPN and PANet. The detection head traverses the feature map through the sliding window mechanism to generate a series of initial candidate regions, and then performs the three key tasks of bounding box coordinate regression, target category classification, and detection confidence assessment for each candidate region simultaneously, and finally eliminates the redundant detection frames by using the non-great-value suppression algorithm to retain the optimal prediction results, and further screens out the less reliable detection outputs based on the pre-set confidence thresholds. The implementation of the YOLO algorithm is shown in Figure 1.

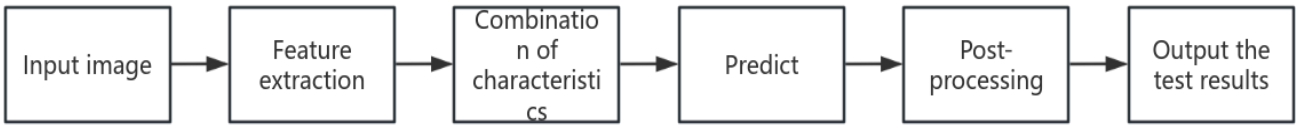

Figure 1. YOLO algorithm implementation flow

C. Improvement and Optimization of YOLOv8 Algorithm

The Head part of YOLOv8 adopts a decoupled head structure, which allows the model to handle the classification and regression tasks separately, and improves the accuracy and efficiency of the model when performing correlated target detection [7].However, for complex scenes at night, the backbone network of the original YOLOv8 may ignore key areas when extracting features, resulting in the omission of small targets or low-contrast pedestrians, so in this study, the YOLOv8 algorithm is improved with the help of adding the CBAM attention mechanism [8][9].Attention mechanism is essentially a weighted summation process at the mathematical level, and its mathematical expression mainly consists of the following three core components:

1）Query Vector (Query, Q): Indicates the current content feature that needs attention, with dimension d_q

2）Key vector (Key, K): Represents the set of features to be retrieved, with dimension d_k

3）Value Vector (Value, V): Contains the actual feature information with dimension d_v

The formula for attention is described in the following equation (1)

$$Attention(Q,K,V) = soft\max\left(\frac{K^T}{d}\right)V \quad (1)$$

The d in Equation (1) represents the dimension of the key vector, while$\sqrt{d}$is used to prevent the softmax gradient from disappearing as a result of the dot product being too large.

CBAM is a lightweight attention mechanism, which can improve the key features according to its own situation and suppress irrelevant background noise at the same time. CBAM consists of two parts: channel attention and spatial attention, which will dynamically adjust the feature weights from the channel and spatial

dimensions, and ultimately outputs the spatial attention weight matrix Ms. This matrix is multiplied with the feature map in the spatial dimension, and effectively strengthens the target area by multiplying the feature map with the spatial dimension, which can be used to enhance the target area by the spatial dimension. operation, which effectively reinforces the localisation information of the target region. The insertion position of CBAM in this detection algorithm is after the C2f module, which is used to improve the expression ability of multi-scale features, and CBAM is added before the SPPF layer, that is, before spatial pyramid pooling, which is used to optimise the weight allocation of multi-scale context information. Moreover, in this paper, the Lion optimiser is used, which introduces the symbolic momentum mechanism and its algorithm design is relatively simple, so that it achieves excellent performance compared with traditional optimisers in the experimental results of multiple benchmark tests, which indicates that the use of the Lion optimiser can improve the speed of the model training process, and provides a new technological path for the YOLOv8 algorithm to improve the performance of the model. performance and provides a new technical path for the YOLOv8 algorithm [10][11].In the CBAM attention mechanism, the core mathematical formulas of channel attention and spatial attention are as follows:

a) Channel attention formula

First, compress the spatial information through the global average pooling, then learn the channel weight through MLP, and finally weight the original feature diagram:

$$F^{'}=\sigma\left(W_2\cdot\delta\left(W_1\cdot\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}F(c,i,j)\right)\right)\otimes F \quad (2)$$

F in the formula (2) is the input characteristic diagram, F' is the channel attention output, W1/W2 is the MLP weight, δ is ReLU, σ is Sigmoid, ⊗ It is an element-by-element multiplication.

$$F_{CBAM}=\sigma\left(C\mathrm{onv}_{1\times1}\left(concat\left(\max_{c=1}^{C}F^{'}(c,i,j)\right)\right)\right)\otimes F^{'} \quad (3)$$

b) Spatial attention formula

First, aggregate information along the channel dimension pool, then learn the space weight through 1×1 convolution, and finally weight the channel attention output:

F' in the formula (3) is the channel attention output, FCBAM is the final output of CBAM, Conv1×1 is 1×1 convolution, concat is feature splicing, σ is Sigmoid, ⊗ is element-by-element multiplication.

When incorporating the CBAM attention mechanism into optimised algorithmic models for night vision pedestrian detection, it enhances detection capabilities for low-contrast targets. Regarding channel attention, this model amplifies pedestrian-relevant feature channels while suppressing background noise. In terms of spatial attention, it enables the model to focus on key pedestrian regions, reducing false detections caused by night-time lighting interference. It also improves performance in scenarios involving small targets and occlusions, such as within dense pedestrian datasets like CrowdHuman. CBAM enables the model to recognise local features of occluded pedestrian targets, reducing the false negative rate by approximately 8%. Furthermore, CBAM incurs only a minimal computational overhead, with parameter growth under 1%, maintaining real-time detection at 30+ FPS on edge devices like Jetson Xavier. Integrating the CBAM attention mechanism into the YOLOv8 backbone network enables dual-weight adjustments for both channel and spatial attention, thereby enhancing the accuracy of pedestrian detection in night vision systems under complex conditions such as low illumination and target occlusion. By incorporating attention mechanisms, computational overhead is reduced while significantly enhancing the algorithm's pedestrian detection precision, thereby providing robust technical support for practical implementation. Figure 2 below shows the network structure of the YOLOv8 algorithm after adding the CBAM attention mechanism.

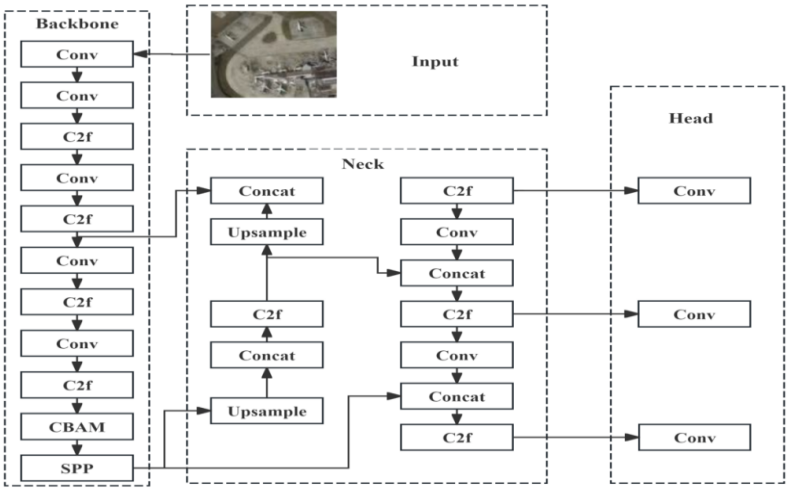

Figure 2. Improved YOLOv8 network structure diagram

III. Dataset Construction And Evaluation Indicators

A. Processing of night vision pedestrian dataset

The impact of data training on the final accuracy of deep learning models is paramount. During training, the performance of the target model is primarily determined by the quantity and quality of pedestrian detection target data. Sufficient data volume enables the model to thoroughly learn the feature distribution of pedestrians across diverse scenarios, while high-quality data reduces annotation errors and noise interference, providing reliable support for model training. Therefore, acquiring a dataset that is both sufficiently large in volume and sufficiently high in quality is the foremost priority for the realisation of this research project.The main data used in this study comes from the public data set LLVIP, which is specially built for visual tasks under low-light conditions. It provides 30976 groups of strictly space-aligned visible light-infrared dual-modal image samples, and most of them are collected from extreme dark light environments. In this model training, only The LLVIP data set of the set is marked as visible light images in VOC format. A training set contains 1,2025 pictures, the verification set contains 3,463 pictures, and a small part of the data set images come from the website https://aistudio.baidu.com /Global/search? The public image of. In the night vision data set training of this study. According to the difficulty of image search, some of the images were selected in this study as shown in Figure 3.

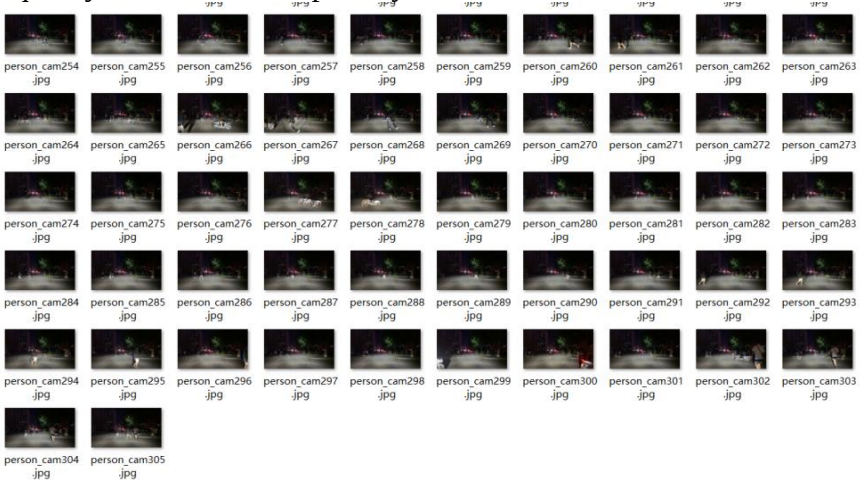

Figure 3. Partial data display of self-built night vision pedestrian data set

1）Data annotation.

After completing the image data collection, this study first carried out the format conversion process of the LLVIP data set, and its original VOC annotation format was uniformly converted into YOLO format. For the self-collected data set, the target object in each image is carefully categorized and positioned. As the core step in the construction of the target detection data set, this annotation link requires a lot of manual work in actual operation. In the specific annotation process, this study strictly follows the TXT file format specification of the YOLO framework, and realizes the precise annotation of the target area and the automatic generation of the corresponding TXT file through the professional annotation tool LabelImg. As an open-source target annotation software, this tool not only provides an intuitive and friendly user interface, but also efficiently completes the task of drawing the target box, and supports the automatic saving function of annotation results. The following Figure 4 shows the user interface of LabelImg.

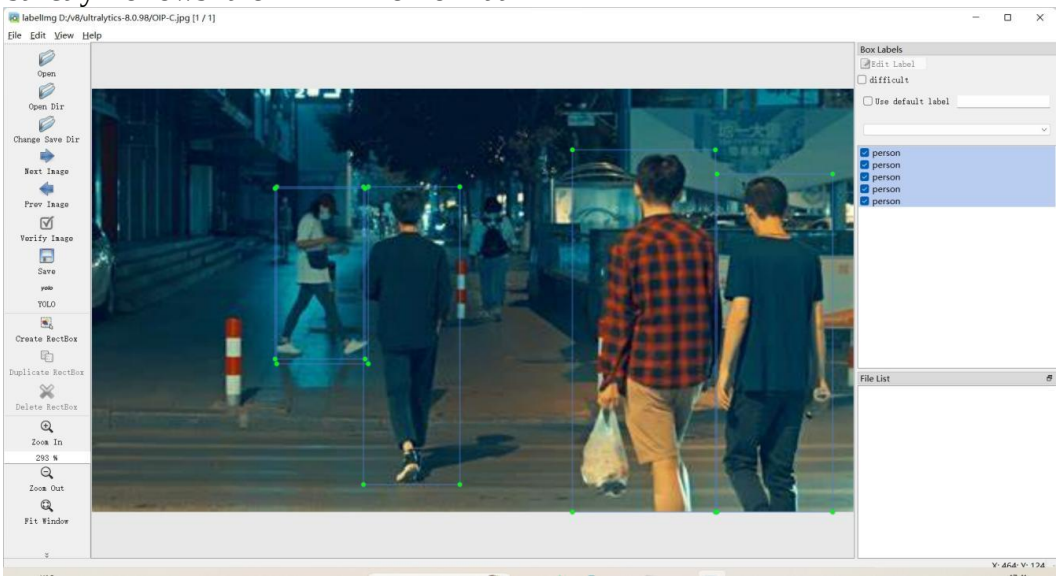

Figure 4. LabelImg User Interface

As shown in Figure 5, this is a TXT file generated from an image annotated in YOLO format within the dataset used in this study. Each .txt file contains multiple lines of text, with each line corresponding to a target instance. The data structure follows a specific format: <class_id> <x_center> <y_center> <width> <height>. Here, class_id is an integer category identifier corresponding to a predefined category index. x_center and y_center represent normalized object center coordinates, expressed as ratios relative to the image width and height within the range [0,1]. Finally, width and height denote normalized object bounding box dimensions, also expressed as ratios relative to the image width and height within the range [0,1].

Figure 5. Content of the YOLO-formatted annotation txt file

2）Dataset partitioning.

Before formal training begins, the self-collected dataset is divided into training and validation sets. The specific partitioning method is illustrated in Figure 6. The training set primarily serves to optimize model parameters and conduct training tasks, while the validation set plays a crucial role in evaluating model performance. It encompasses

comprehensive assessments of key metrics such as model accuracy, recall rate, operational efficiency, and loss function values.

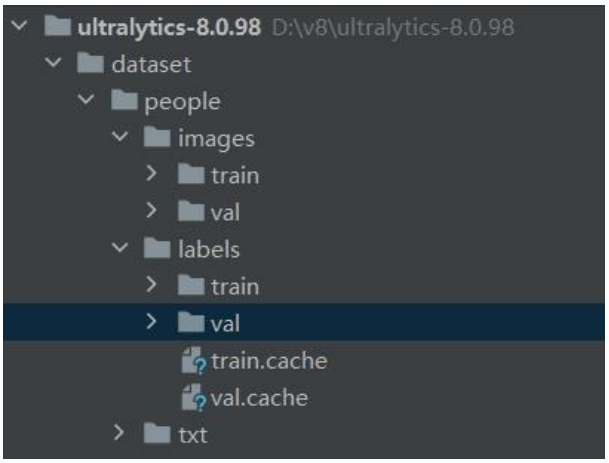

Figure 6. TXT Format Structure of YOLOv8

B. Evaluation index of target detection

In the study of the night vision pedestrian detection system based on YOLOv8, a comparative experiment was set up before and after the algorithm improvement and using the Lion optimizer on the LLVIP data set, and the evaluation index of the target detection was used to evaluate the model. In this study, the model used to measure the performance of the model The evaluation indicators mainly include the following.

1）mAP.

In the field of target detection, the average accuracy mean (mAP) as a core evaluation index can comprehensively reflect the detection performance of the model under different confidence thresholds. Specifically, mAP@0.5 refers to the average accuracy value when the IoU threshold is 0.5. This indicator mainly evaluates the performance of the model under the relatively relaxed matching standard; while mAP@0.5:0.95 characterizes the average accuracy of the IoU threshold in the range of 0.5 to 0.95, and its number The value size directly reflects the detection ability of the model under stricter matching conditions. Experimental results show that a higher mAP@0.5 value means that the model has superior performance under lower IoU requirements, while a higher value of mAP@0.5:0.95 indicates that the model shows stronger robustness under high-precision matching scenarios, which can reflect the comprehensive performance of the model.

2）Precision.

Precision indicates that the ratio of the actual positive sample in the positive sample predicted by the model. The calculation method it follows is the following formula (4).

$$\Pr ecision = \frac{TP}{(TP + FP)} \tag{4}$$

TP denotes correctly detected positive samples, while FP denotes incorrectly detected negative samples. High Precision indicates a low false positive rate for the model.

3）Recall.

Recall represents the proportion of actual positive samples correctly predicted as positive by the model. It is calculated according to the following formula (5) below.

$$\mathrm{Re} call = \left(\frac{TP}{(TP + FN)} \right) \tag{5}$$

TP denotes correctly detected positive samples, while FN represents undetected positive samples. A high Recall indicates a low false negative rate for the model. d.F1 Score. F1 Score is the reconciled mean of precision and recall, which is calculated following equation (6), a high F1 Score indicates that the model achieves a better balance between precision and recall.

$$F_1 = 2 \times \frac{(\Pr ecision \times \mathrm{Re} call)}{(\Pr ecision + \mathrm{Re} call)} \tag{6}$$

IV. Analysis Of Experimental And Training Results

A. Experimental environment and parameter setting

This research was conducted under the Windows 11 operating system using the PyTorch framework and Python language within PyCharm to achieve detection of night vision pedestrian

images and videos. The experimental platform is detailed in Table Ⅰ below.

TABLE I. EXPERIMENTAL PLATFORM HARDWARE CONFIGURATION

Name	*Related Configuration*
Operating System	Windows11
Processor	11th Gen Intel(R) Core(TM) i7-11800H @ 2.30GHz 2.30 GHz
Memory	16G
GPU	NVIDIA GeForce RTX 3050
Deep Learning Framework	Pytorch

During training of the night vision pedestrian detection model, parameters must be pre-configured. The training process employs 50 iterations to comprehensively learn the training data. A learning rate of 0.001 ensures training stability, while a batch size of 16 enhances the model's generalization capability to some extent. An image size of 640 pixels guarantees sufficient retention of pedestrian feature information in night vision images, and the Lion optimizer accelerates model convergence. Specific experimental parameter settings are detailed in Table Ⅱ below.

TABLE II. EXPERIMENTAL PARAMETER SETTINGS

Parameter	Parameter Setting
Number of training iterations	50
Learning rate	0.001
Batch size	16
Image size	640
Optimizer	Lion

B. Experimental Results of the Improved Algorithm on the LLVIP Night Vision Pedestrian Dataset

The experimental results comparing the YOLOv8 algorithm and the algorithm with the added attention mechanism on the LLVIP dataset, based on the four model evaluation metrics introduced in the previous subsection, are presented in Table Ⅲ below.

TABLE III. COMPARISON OF ORIGINAL AND IMPROVED ALGORITHMS

Algorithm	Precision	Recall	mAP@0.5:0.95	mAP@0.5	F1 Score
YOLOv8	85.3	89.3	88.9	57.8	87.2
YOLOv8+CBAM	87.3	84.4	88.94	58.2	87.84

Through the above table, after comparing the evaluation indicators obtained from experiments on the same data set before and after the improvement of the algorithm model, it can be seen that after improving and optimizing the YOLOv8 algorithm by integrating the CBAM attention mechanism, the accuracy of the algorithm model has been improved by about 2.0%, mAP@0.5 has been improved by 1.6%, and the average accuracy of the IoU threshold value has been improved by about 0.04% from 0.5 to 0.95, and the F1 Score has been improved by 0.64%. It can be seen that after the CBAM attention mechanism has been added to the YOLOv8 algorithm model, the comprehensive identification ability of the model has been enhanced, and the balance between its accuracy and recall rate has also been optimized, improving the overall performance of the model.

C. Training Results of the Improved Algorithm on the Self-built Night Vision Pedestrian Dataset

The following illustrates the three loss function curves—Box Loss, Cls Loss, and Dfl Loss—alongside four key performance metric curves—Precision, Recall, mAP50, and mAP50-95—corresponding to the training and validation sets during model training on the same self-built night vision pedestrian dataset, both before and after optimisation. These curves provide an intuitive representation of the model's loss evolution and performance enhancement during training, clearly illustrating the impact of optimisation strategies on training outcomes. The aforementioned curves are depicted in Figures 7 and 8 respectively.

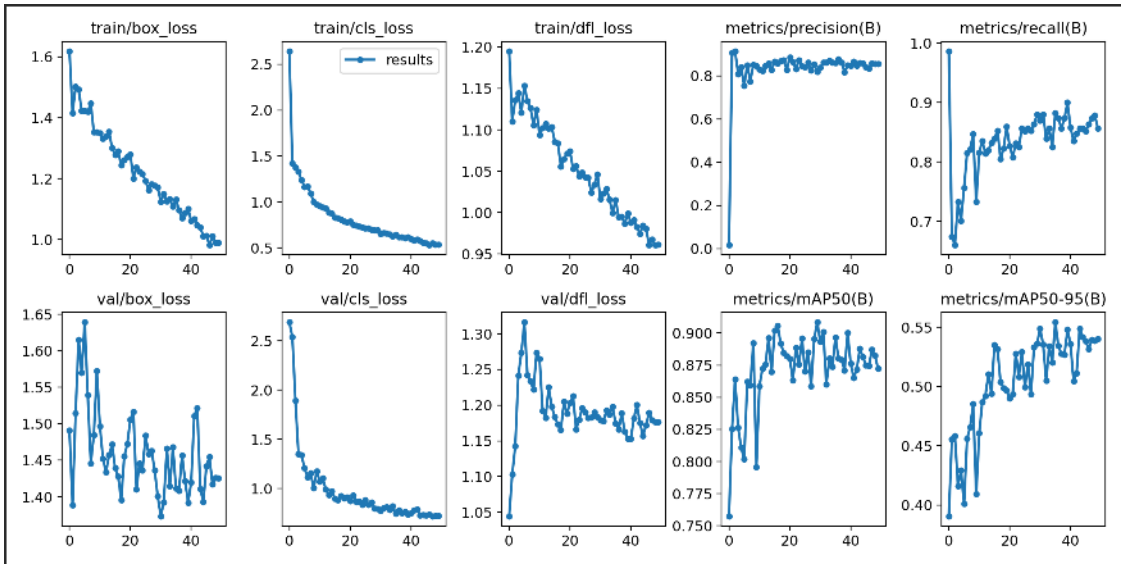

Figure 7. Training results prior to model optimization

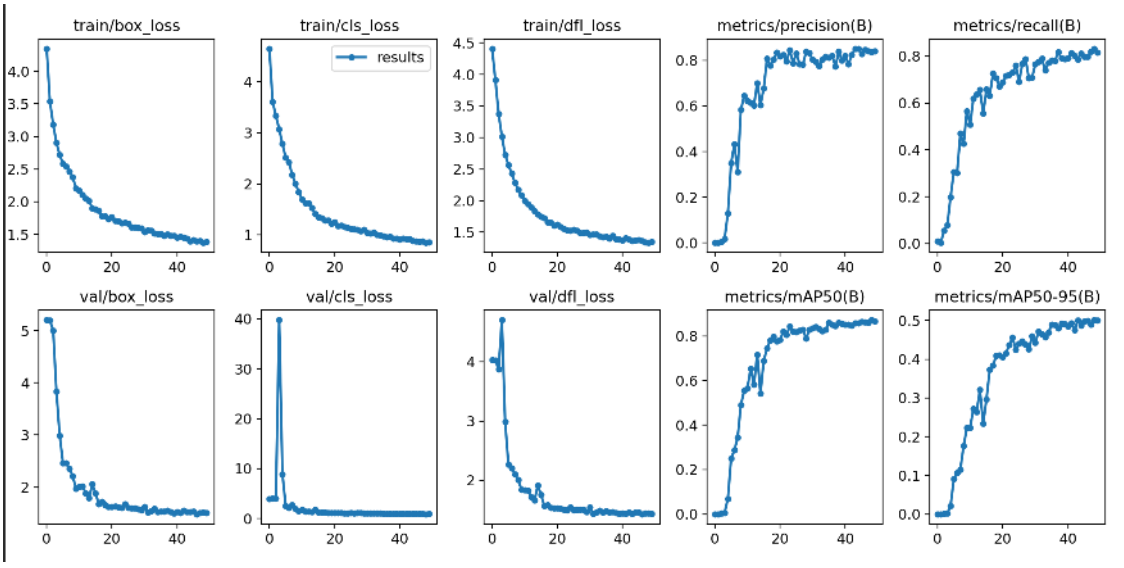

Figure 8. Training results after model optimization

1）Loss Functions

By comparing Figures 7 and 8 from the preceding subsection, the relevant loss function curves before and after model optimization can be analyzed as follows.

a) Box Loss (bounding box regression loss).

Within the YOLOv8 night vision pedestrian detection system, the box loss regression curve directly reflects the model's optimization process for predicting pedestrian locations. Its morphological changes reveal critical issues concerning pedestrian target localization accuracy in complex nocturnal scenes. Prior to algorithmic refinement, this loss function exhibited a high-amplitude oscillation pattern. Fundamental causes may include mismatched anchor sizes relative to nocturnal pedestrians, motion blurring of pedestrian targets leading to boundary ambiguity, or insufficient feature discrimination capabilities within the system. Following model optimization, the loss function exhibited a healthy curve pattern: rapid initial decline, moderate mid-phase decrease, and subsequent stabilization around 0.6. This confirms enhanced positioning accuracy in complex night-time environments post-algorithm refinement.

b) Classification Loss.

Within the YOLOv8 night vision pedestrian detection system, the Classification Loss curve reflects the model's ability to distinguish positive pedestrian classes from negative background interference classes. Its trend directly reveals the optimization state of the classifier in complex night-time scenarios. Both before and after algorithmic refinement, this loss function exhibits an initial rapid decline, a moderate mid-phase decrease, and a relatively stable late-phase trend. However, the final stable value post-optimization is lower than the pre-optimization value, indicating a performance improvement following algorithmic enhancement.

c) Dfl Loss.

Within the YOLOv8 night vision pedestrian detection system, this loss function curve directly reflects the model's predictive accuracy regarding the spatial distribution of pedestrian bounding boxes. Its dynamic variations directly reveal critical issues in target localization within complex nocturnal environments. Prior to algorithmic refinement, this loss function exhibited pronounced oscillations. Potential causes include excessively high learning rates or overly aggressive night-time data augmentation, suggesting potential positioning instability within the system. Post-optimization, however, the curve demonstrated a rapid initial decline, followed by a gradual mid-phase decrease, ultimately stabilizing around 0.4 – a healthy, normal trajectory. This demonstrates that the optimized algorithmic model exhibits enhanced performance in locating small and ambiguous targets.

2）Performance indicators

Through the comparison of Figure 7 and Figure 8, the relevant performance index curve before and after the optimization of the algorithm model can be analyzed as follows.

a) Precision.

Before the model algorithm was improved, the Precision curve exhibited a pattern of low-level fluctuations. Analysis suggests this may have been caused by nighttime lighting interference leading to false detections of pedestrian backgrounds and other objects, or by an excessively high classification threshold. However, following algorithmic model refinement and optimization, the Precision curve exhibited an initial rapid ascent, a moderate mid-phase increase, and a subsequent convergence toward approximately 0.8. This trend represents a healthier performance compared to pre-improvement, indicating enhanced control over false detections during pedestrian recognition. Particularly in complex nighttime scenarios, this performance directly determines the reliability of the pedestrian detection system.

b) Recall.

The Recall curve intuitively reflects the model's ability to detect real pedestrians, especially in low light, blur and other scenarios, its performance directly determines the safety performance of the system. The abnormal situation of the curve morphology of violent vibration in the Recall curve morphology before the modification of the algorithm model, and analyze the possible reasons such as excessive learning rate leading to model shock or excessive data enhancement, and also reflect the problem that the system may have poor detection stability. However, after the improvement and optimization of the algorithm model, the Recall curve shows a rapid rise in the early stage, a gradual increase in the middle stage, and a basic maintenance of about 0.8 in the later stage. Compared with the improvement of the algorithm, the performance is more ideal, and it also reflects the improvement in the detection stability and security performance of the system.

c) mAP50.

In the YOLOv8 night vision pedestrian detection system, the mAP@0.5 curve is the core index of the comprehensive detection ability of the model, especially in complex night scenarios, its performance directly reflects the positioning and identification accuracy of the system for pedestrians. Before the algorithm model was optimized, the performance of the mAP@0.5 curve had a high-level oscillation curve morphology. The possible reasons for it were the mismatch of the anchor point size, or the blurred boundary caused by low light. At the same time, it also reflected the unstable positioning of the system. However, after the improvement and optimization of the algorithm model, the

mAP@0.5 curve shows a rapid rise in the early stage, a steady rise in the mid-term, and a healthy curve trend that basically converges to about 0.85 in the later stage. At the same time, it also reflects the comprehensive detection of the system after the algorithm is improved and optimized. The ability has been improved.

d) mAP50-95.

In the YOLOv8 night vision pedestrian detection system, the mAP@0.5:0.95 curve is a golden index to measure the comprehensive detection ability of the model. Especially in complex night scenarios, its curve morphology directly reveals the robustness of the system under different positioning accuracy requirements. Before the algorithm model is optimized, the curve morphology of mAP@0.5:0.95 has an abnormal situation of violent vibration. The analysis may be that the root cause may be unreasonable anchor point size distribution or regression loss weight imbalance. At the same time, it also reflects the system defect of the model that is sensitive to IoU changes. However, after the improvement and optimization of the algorithm model, the curve morphology of mAP@0.5:0.95 showed a rapid rise in the early stage, a steady rise in the medium term, and a relatively healthy curve form that basically converged to about 0.5 in the later stage, which also reflected the improvement of the comprehensive detection ability of the algorithm model.

3）PR curve

The following Figures 9 and 10 are the PR curves before and after the improvement of the YOLOv8 algorithm and after the improvement model training respectively.

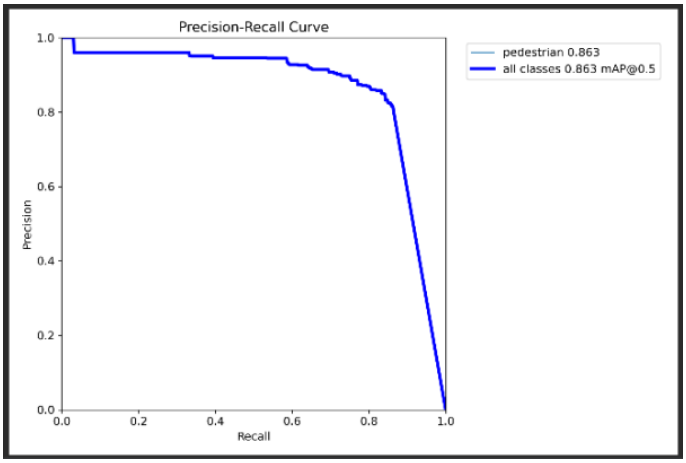

Figure 9. PR curve before algorithm improvement

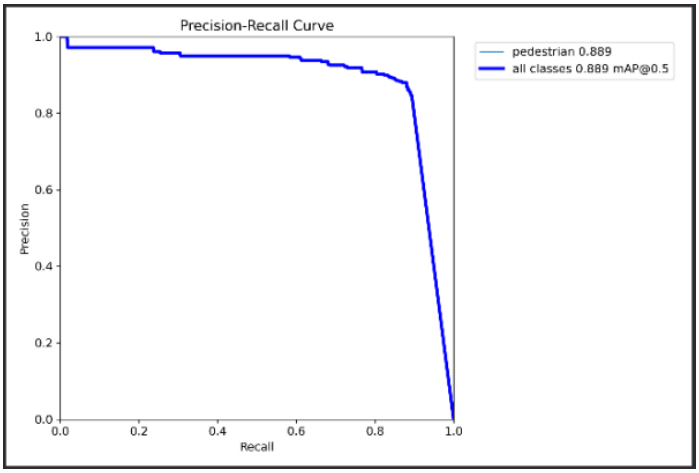

Figure 10. PR curve after algorithm improvement

In the YOLOv8 night vision pedestrian detection system, the PR curve is a curve with the recall rate as the horizontal axis and the accuracy rate as the vertical axis, reflecting the performance of the model under different judgment thresholds. The morphological changes of this curve intuitively reveal the trade-off relationship between the accuracy rate and the recall rate of the model under different confidence thresholds, especially in low-light fields. The scene can clearly expose the core performance of the system. Its night vision scene value is reflected in the area under the curve, which mainly shows the comprehensive detection ability of the quantitative model. The larger the area, the better the performance of the algorithm model, and its curve shape reveals the difference in day and night performance.

According to the comparison of Figure 9 and Figure 10 above, it can be seen that the area under the curve has increased after algorithm improvement and optimization, so this shows that the performance of the model has improved after the optimization and improvement of the algorithm model.

D. Night Vision Pedestrian Detection Results

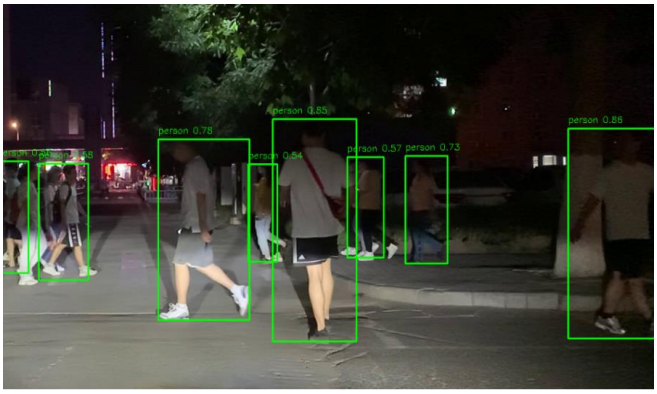

Figure 11. Image Detection Interface

As shown in Figure 11 above, the confidence scores displayed above the pedestrian detection boxes indicate that the system achieves high detection accuracy for nearby pedestrian targets in night vision environments.

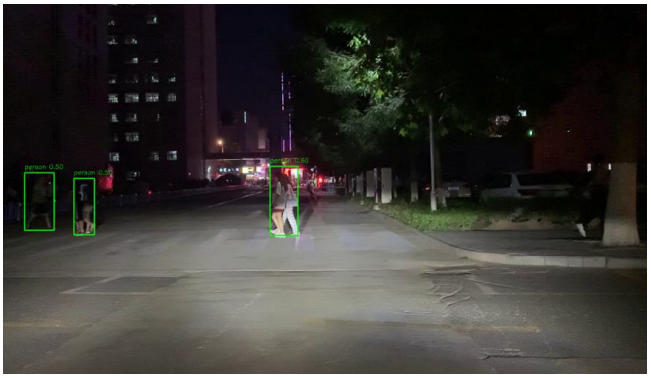

Figure 12. Image Detection Interface

As illustrated in Figure 12 above, it is evident that this night-time pedestrian detection system maintains its ability to accurately detect and identify pedestrian targets when processing images captured at long distances within densely populated night-time scenes. Even under dim ambient lighting conditions and amidst complex pedestrian distributions, the system exhibits no significant instances of missed or false detections, fully demonstrating its robust detection performance within intricate night-time environments.

Figure 13. Video Detection Interface

As shown in Figure 13 above, the system maintains a high accuracy rate for pedestrian detection even when processing videos captured in complex nighttime scenes.

V. CONCLUSIONS

In view of the common inherent characteristics of insufficient illumination, low resolution and poor contrast of night images and videos, this paper has successfully built a night vision pedestrian detection system with high detection accuracy by integrating the improved YOLOv8

algorithm and the PyTorch deep learning framework.

In terms of the acquisition and processing of data sets, after completing the image data collection, this study first converts the format of the LLVIP data set, converts its original VOC annotation format into YOLO format, and then uses the professional annotation tool LabelImg for self-built data images. Mark the target area, and then divide it into training sets and verification sets.

In the process of system development, this study uses the Lion optimizer to replace the original YOLOv8 default SGD optimizer, which has the characteristics of dynamically adjusting the learning rate, which significantly improves the efficiency and convergence speed of model training, thus ensuring the detection stability of the system in complex night environments.

In terms of model architecture optimization, the weight distribution of key feature areas is effectively enhanced by introducing the CBAM attention mechanism in the feature extraction stage, and after experimental verification, it can be shown that the improvement strategy significantly improves the algorithm's recognition accuracy of pedestrian targets under low-light conditions. Finally, the front-end visual interface of the system is realized based on PyQt5.

In short, this paper has successfully built a night vision pedestrian detection system by introducing the CBAM attention mechanism to improve the YOLOv8 network structure and the introduction of Lion. And the experimental results show that the integration of the improved model of CBAM attention mechanism and Lion optimizer not only optimizes the detection performance of YOLOv8 in the night vision environment, but also significantly improves the balance between accuracy and recall rate, fully meeting the practical application needs of complex night scenes.

REFERENCES

[1] Du Yunliang, Wang Mingjia. Research on pedestrian detection in a weak light environment based on semi-supervised domain adaptation [J]. Journal of Electronic Measurement and Instrumentation, 2024,38(01):106-113.

[2] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580－587.

[3] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]. International Conference on Neural Information Processing Systems. [S. l.]: MIT Press, 2015: 91－99.

[4] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector [C]. European Conference on Computer Vision. Springer International Publishing, 2016: 21－37.

[5] Redmon J, Divvala S, Girshick R, et al. You only look once: unified , real-time object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779－788.

[6] Liu Yang, Xie Yongqiang, Li Zhongbo, et al. Research Progress on Object Detection Algorithms Based on Deep Learning [J]. Communications Technology, 2021, 54(09): 2063-2073.

[7] Zhu Chunyuan. Research on small target detection based on the improved YOLOv8 algorithm [D]. Dalian Jiaotong University, 2024.

[8] Yan J ,Zeng Y ,Lin J , et al.Enhanced object detection in pediatric bronchoscopy images using YOLO-based algorithms with CBAM attention mechanism[J].Heliyon,2024,10(12):e32678-e32678.

[9] Guohua G ,Ciyin S ,Shuangyou W .Mature tomato recognition and location algorithm based on binocular vision and deep learning[C]//Beijing University of Technology (China),2023:

[10] Dong Yiming, Li Huan, Lin Zhouchen. Converge velocity analysis of LION optimiser [J/OL]. Computer Journal, 1-25 [2025-06-23]. http://kns.cnki.net/kcms/detail/11.1826. TP.20250424.1457.002.html.

[11] Zhao D ,Shao F ,Liu Q , et al.A Small Object Detection Method for Drone-Captured Images Based on Improved YOLOv7[J].Remote Sensing,2024,16(6).