

Article

Research on Pupil Center Localization Detection Algorithm with Improved YOLOv8

Kejuan Xue ¹, Jinsong Wang ^{1,2,*} and Hao Wang ¹

¹ School of Optoelectronic Engineering, Changchun University of Science and Technology, Changchun 130022, China

² Zhongshan Institute of Changchun University of Science and Technology, Zhongshan 528437, China

* Correspondence: soldier_1973@163.com; Tel.: +86-13604309217

Abstract: Addressing issues such as low localization accuracy, poor robustness, and long average localization time in pupil center localization algorithms, an improved YOLOv8 network-based pupil center localization algorithm is proposed. This algorithm incorporates a dual attention mechanism into the YOLOv8 backbone network, which simultaneously attends to global contextual information of input data while reducing dependence on specific regions. This improves the problem of difficult pupil localization detection due to occlusions such as eyelashes and eyelids, enhancing the model's robustness. Additionally, atrous convolutions are introduced in the encoding section, which reduce the network model while improving the model's detection speed. The use of the Focaler-IoU loss function, by focusing on different regression samples, can improve the performance of detectors in various detection tasks. The performance of the improved YOLOv8 algorithm was 0.99971, 1, 0.99611, and 0.96495 in precision, recall, MAP50, and mAP50-95, respectively. Moreover, the improved YOLOv8 algorithm reduced the model parameters by 7.18% and the computational complexity by 10.06%, while enhancing the environmental anti-interference ability and robustness, and shortening the localization time, improving real-time detection.

Keywords: positioning of pupil center; YOLOv8; dual attention mechanism; hollow convolution; Focaler-IoU



Citation: Xue, K.; Wang, J.; Wang, H. Research on Pupil Center Localization Detection Algorithm with Improved YOLOv8. *Appl. Sci.* **2024**, *14*, 6661. <https://doi.org/10.3390/app14156661>

Academic Editor: Christos Bouras

Received: 24 June 2024

Revised: 27 July 2024

Accepted: 28 July 2024

Published: 30 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Pupil localization plays an important role in identity verification, disease recognition, visual focus of attention (VFOA) tracking, cognitive assessment for dementia, and detection of driving fatigue [1]. However, factors such as poor image quality and occlusion caused by eyelids, eyelashes, and blinking, as well as the pupil shape not being a perfect circle [2], have led to a decrease in the accuracy and robustness of pupil center localization, making it still challenging to achieve high-precision and real-time localization.

Existing research has introduced numerous pupil localization detection techniques, both domestically and internationally. These encompass methods such as gradient vectors, Hough circle transform, ellipse fitting, and other traditional image processing techniques [3–8]. Vranceanu R et al. made use of gray-level details from human eyes [9], positioning the eye area via a gray-level projection function. This approach boasts low computational requirements but comes with lower localization precision. Zhang Hongwei et al. [10] proposed a refined pupil recognition algorithm based on Hough circle transform. This advancement attained pupil center localization by confining the detection radius after conducting Canny edge detection and employing Hough transform to determine the optimal fitting circle. This advancement tackles the high computational complexity and poor real-time performance of traditional Hough transform methods, yet it necessitates a resetting of radius parameters, leading to increased manual dependency with the change in image datasets. Loy et al. [11] introduced a swift radial symmetry transform algorithm, essentially refining Hough transform by transitioning the mapping space from parametric space to

digital images, thus successfully reducing the algorithm's transformation dimensions and addressing the issue of increased algorithmic complexity due to the high dimensionality of Hough transform's parametric space, which often hinders real-time performance.

With the advancement and refinement of deep learning techniques, their application in fields such as image classification, segmentation, and object tracking has yielded outstanding outcomes. In pupil localization, researchers have employed convolutional neural networks (CNNs) and the YOLO series networks. Instances include the utilization of eye images as inputs for CNNs to directly predict pupil centers in [12–15], and pupil localization via the YOLOv3 model coupled with the starburst ray method in [16]. Boosted-oriented edge optimization (BORE) [17] works as a booster algorithm that can be used to increase the accuracy of classification and regression problems. BORE algorithm-based pupil localization methods operate beyond 500 Hz [18]. Dealing with low-quality VS eye images is extremely difficult. However, the symmetric and expansive structure of U-Net helps achieve precise pupil localization while maintaining contextual information. Moreover, it generates greater accuracy in pupil localization for highly challenging VS eye images while letting the algorithm run in real time [19–21]. Compared to traditional approaches, object tracking networks exhibit notable improvements but are often less robust and may result in fluctuations in pupil center positions.

Deep learning methods extract deep semantic information from images, known as deep features [22,23]. Their essence lies in transforming target recognition into a regression task. For images under varying conditions, including noise, these methods provide rapid target recognition while maintaining high detection accuracy, thereby demonstrating overall excellence. Compared to traditional methods, deep learning offers superior expression and representation of targets within images. Furthermore, in scenarios involving changes in target shapes, lighting conditions, or occlusions, deep learning-based object detection algorithms display greater robustness. However, this approach demands significant computational resources, extensive training data, and tackling issues such as overfitting.

To enhance pupil positioning accuracy, this paper introduces a pupil center localization algorithm grounded in an enhanced YOLOv8n network. This algorithm integrates a dual attention mechanism into the YOLOv8n backbone network, effectively aggregating and disseminating informative global features across the spatiotemporal space of images. This approach facilitates access to spatial features in subsequent convolutional layers. While emphasizing global contextual information from input data, it mitigates dependence on specific regions, thereby enhancing model robustness. Furthermore, the incorporation of dilated convolutions, replacing standard convolutions in the original network, lightens the model and accelerates the detection speed. The employment of the Focaler-IOU loss function, focusing on diverse regression samples, strengthens the detector's performance across various detection tasks. The application of this algorithm in a custom-developed multi-parameter eyeball measurement system significantly reduces measurement time and elevates the accuracy, real-time performance, and robustness of optical eyeball biometric measurements, thereby contributing to the intelligent evolution of ophthalmic instrumentation.

The structure of this paper mainly includes the following: Firstly, the YOLOv8 model is introduced and compared with the YOLOv5 model; secondly, the improved YOLOv8n algorithm is proposed, which mainly includes the integration of a dual attention mechanism in the YOLOv8n backbone network to enhance the ability to extract human eye pupil features, analysis of the impact of difficult samples and easy-to-handle sample distribution on regression results, and the proposal of a new method, Focaler-IOU, which can improve the detector's performance in different detection tasks by focusing on different regression samples, introducing the void convolution to not only reduce the network model but also improve the detection speed of the model; thirdly, our experimental results and analysis are presented, where the experimental dataset, experimental configuration, and evaluation index are introduced, and through the comparative experiments we verify the functional

indicators of the proposed improved algorithm, including different attention mechanisms and different models for comparison. Finally, the research contents are summarized.

2. YOLOv8 Model

YOLOv8 is a new network released by Ultralytics in early 2023, proposed based on the YOLOv5 model. It comprises four models: YOLOv8l, YOLOv8m, YOLOv8s, and YOLOv8n [24]. Among them, the YOLOv8n network is the one with the lowest complexity in the YOLOv8 series. It maintains high detection accuracy while having faster inference speed and easier satisfaction of lightweight requirements. Hence, improvements have been made to the model based on YOLOv8n, with its structure shown in Figure 1.

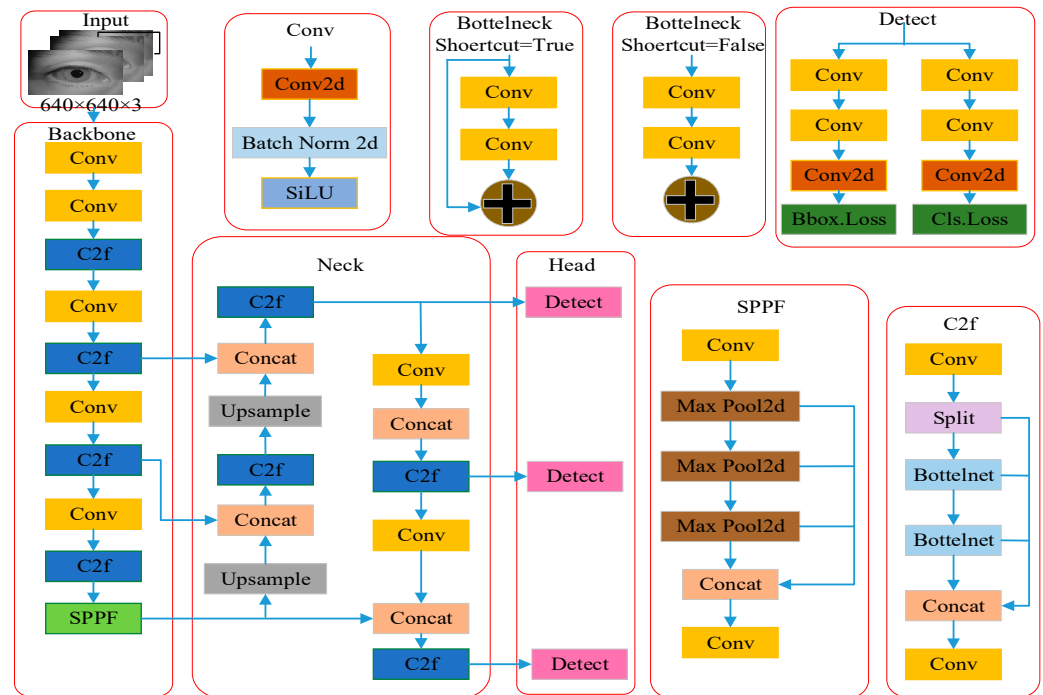


Figure 1. YOLOv8 network architecture.

The YOLOv8 network structure [25–27] comprises four parts: input, backbone, neck (for feature enhancement), and prediction. Firstly, the backbone module continues to utilize the CSP concept for feature propagation, while introducing the C2f module (CSP bottleneck with two convolutions) module to replace the C3 (CSP bottleneck with three convolutions) module in the YOLOv5 model. Compared to the YOLOv5 model released in 2020, the C2f module employs a multi-branch flow design, providing the model with richer gradient information, enhancing its feature extraction ability, and improving the network learning efficiency. The SPPF module is retained, which can obtain rich gradient flow information while optimizing the network structure. Secondly, the neck module maintains the FPN + PAN structure, enabling multi-scale feature fusion capabilities. Finally, in the prediction section, the original coupled head is changed to a decoupled head, introducing the decoupled head to separate classification and regression into two separate structures, improving the model's convergence ability. A schematic of the YOLOv8 improvement module is shown in Figure 2. YOLOv8 utilizes an anchor-free network, which improves the imbalance issue between positive and negative samples through dynamic allocation of positive and negative samples.

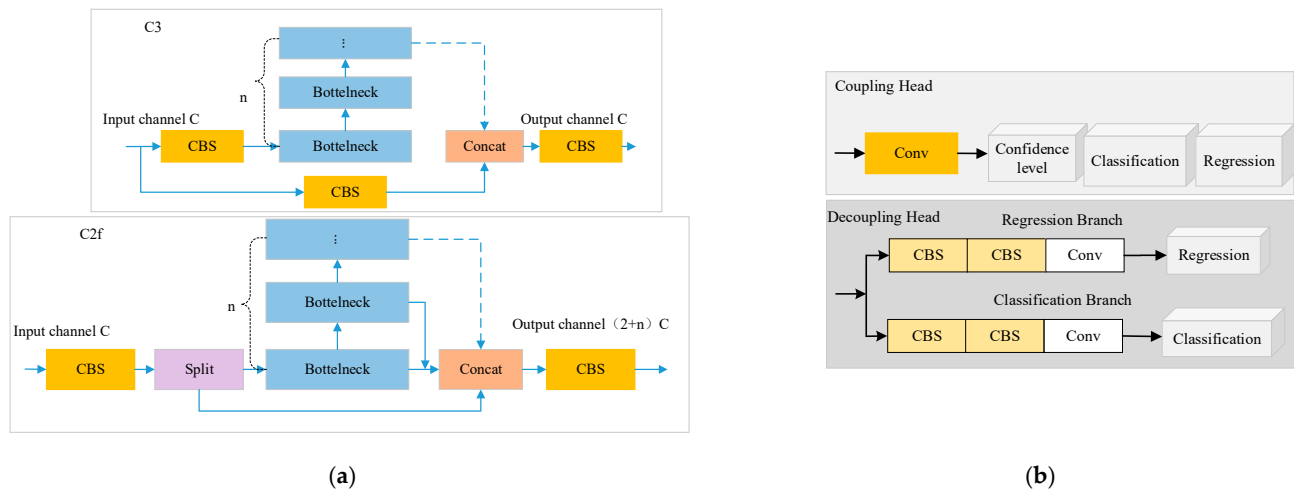


Figure 2. Improvement points of YOLOv8n: (a) C3 and C2f modules; (b) coupling head and decoupling head.

3. YOLOv8n Model Improvement

3.1. Dual Attention Mechanism

Adding an attention mechanism to the YOLOv8n backbone network, in deep learning, is a method that imitates the principle of human attention allocation. It can help neural networks to automatically learn weights and focus key information when processing input sequences. Through this approach, neural networks can more effectively capture long-range dependencies in input sequences. The core idea of attention mechanisms is to assign a weight value to each element in the input sequence, which will determine the level of attention that the model pays when processing the input sequence. The weight value is calculated by a learnable function, which is usually a neural network. This model adopts a dual attention (DA) mechanism, which is a dual-attention network that introduces a self-attention mechanism [28]. In order to enhance the ability to extract human pupil features, a DA mechanism was integrated before the SPPF layer in the YOLOv8 backbone network. The structural modifications before and after the improvement are shown in Figure 3.

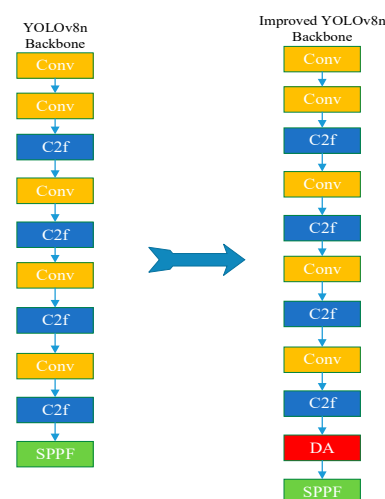


Figure 3. Improvement points of YOLOv8n.

DA adopts a self-attention mechanism to capture rich semantic information, consisting of two parallel attention modules. The dependency relationship between any two positions in the feature map is obtained through the position attention module (PAM), and the dependency relationship between any two channels is obtained through the channel

attention module (CAM). Finally, the output of the two attention modules is transformed using convolutional layers, and an element-by-element summation is performed to achieve feature fusion. The structure is shown in Figure 4.

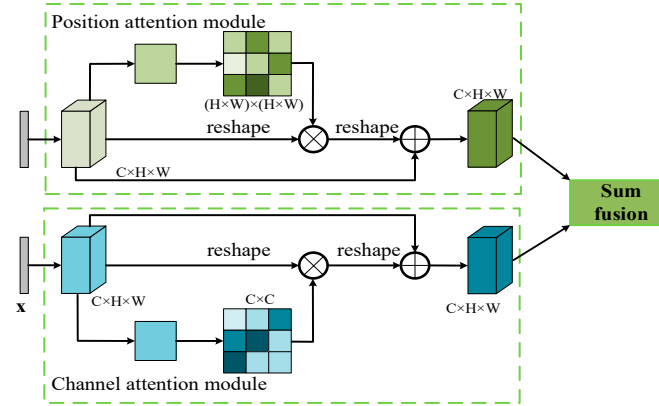


Figure 4. DA module.

3.1.1. Position Attention Module

The upper part of Figure 4 is a compact position attention module, as shown in Figure 5. By utilizing the correlation between two feature points, broader contextual information is encoded into local features, thereby mutually enhancing their respective feature representation capabilities. This module is used to extract spatial information. The relationship between each pixel and several aggregation centers was constructed through pyramid pooling (pooling of different sizes), and these pooling features were concatenated to form a self-attention inner product, which to some extent reduced the computational complexity and memory consumption.

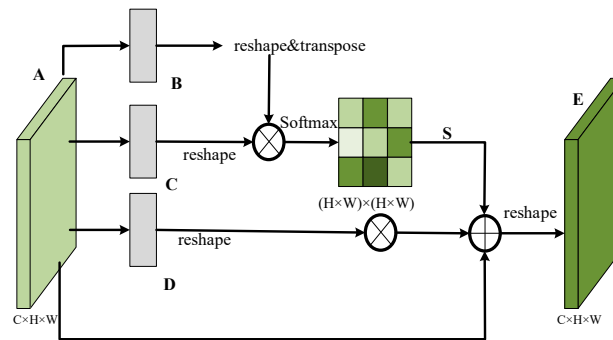


Figure 5. Structural diagram of compact position attention module.

Firstly, the original features $A \in \mathbb{R}^{C \times H \times W}$ are dimensionally reduced through a convolutional layer with BN and ReLU layers, obtaining the features $B \in \mathbb{R}^{C \times H \times W}$, $C \in \mathbb{R}^{C \times H \times W}$, and $D \in \mathbb{R}^{C \times H \times W}$, and changing the dimensions of features B , C , and D to $\mathbb{R}^{N \times C}$, $\mathbb{R}^{C \times N}$, and $\mathbb{R}^{C \times N}$, respectively, where $N = H \times W$. Secondly, by multiplying the transpose matrix of feature B and the C matrix to obtain the correlation strength matrix between any two point features, the softmax layer is used to obtain the spatial attention mapping of each position to other unknown features.

$$s_{ji} = \frac{\exp(B_i \bullet C_j)}{\sum_{i=1}^N \exp(B_i \bullet C_j)} \quad (1)$$

where s_{ji} represents the similarity between position i and position j , and the larger the value of s_{ji} , the more similar the two features are. Finally, transpose S and multiply it with the D

matrix, and then add it element by element with the original feature A to obtain the final output $E \in \mathbb{R}^{C \times H \times W}$, which is

$$E_j = \alpha \sum_{i=1}^N (s_{ji} D_i) + A_j \quad (2)$$

where α represents the scale coefficient, with an initial value of 0. By gradually learning, larger weights can be obtained.

3.1.2. Channel Attention Module

The lower part of Figure 4 is a compact channel attention module, which integrates the relevant features between all channel mappings, explicitly models the dependency relationships between channels, enhances the response ability of specific semantics under channels, and is used to extract channel information. Its structure is shown in Figure 6.

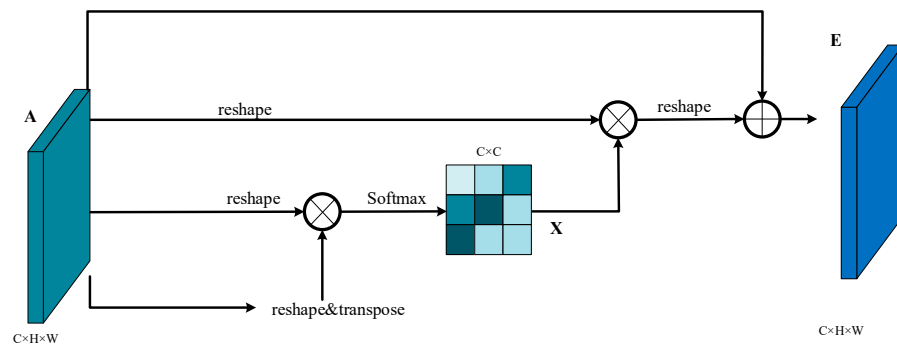


Figure 6. Structural diagram of compact channel attention module.

Unlike the upper part, the three feature maps below have not undergone convolutional dimensionality reduction. Instead, their dimensions are directly transformed into $\mathbb{R}^{C \times N}$, $\mathbb{R}^{N \times C}$, and the correlation strength of the two channels is obtained by multiplying the matrices. Then, the channel attention map $X \in \mathbb{R}^{C \times C}$ is obtained through the softmax layer, which is

$$x_{ji} = \frac{\exp(A_i \bullet A_j)}{\sum_{i=1}^C \exp(A_i \bullet A_j)} \quad (3)$$

where x_{ji} represents the impact of channel i on channel j .

Finally, the response value x_{ji} of the channel attention mapping is used as a weighting, and the dimension-transformed $A \in \mathbb{R}^{C \times N}$ is weighted and fused to obtain the final output $E \in \mathbb{R}^{C \times H \times W}$, which is

$$E_j = \beta \sum_{i=1}^N (x_{ji} A_i) + A_j \quad (4)$$

where β is the scale coefficient, with an initial value of 0. By gradually learning, larger weights can be obtained.

In order to further obtain the features of global dependency relationships, the output results of the two modules are added and fused to obtain the final features for pixel classification.

3.2. The Focaler-IoU Damage Function

In the field of object detection, bounding box regression is crucial, with the accuracy of object detection largely dependent on the loss function for bounding box regression. Existing studies enhance regression performance by leveraging the geometric relationships between bounding boxes, yet they overlook the influence of the distribution of difficult- and easy-to-handle samples on bounding box regression. This paper analyzes the impact of the distribution of difficult- and easy-to-handle samples on regression outcomes and introduces a novel method, Focaler-IoU.

Focaler-IoU enhances detector performance in diverse detection tasks by focusing on distinct regression samples. It reconstructs the *IoU* loss using linear interval mapping to address the distribution of samples, analyzing and considering the impact of sample distribution on regression outcomes, an aspect often overlooked in traditional *IoU* loss functions. Through this innovative method, Focaler-IoU addresses the deficiencies of existing bounding box regression techniques, thereby significantly enhancing detection performance across various detection tasks. The specific expression is as shown in (5):

$$IoU^{focaler} = \begin{cases} 0 & IoU < d \\ \frac{IoU-d}{u-d} & d < IoU < u \\ 1 & IoU > u \end{cases} \quad (5)$$

where *IoU* represents the intersection-to-union ratio, which is the ratio of the area of the overlapping part of two bounding boxes to the area of their union part, while *d* is the lower threshold and *u* is the higher threshold.

This design allows the loss function to be sensitive to *IoU* values within a certain range, allowing it to focus more on samples with moderate overlap between predicted bounding boxes and true bounding boxes, which are neither too difficult nor too easy. This helps the model better learn to extract features from moderately difficult samples, rather than just focusing on the easiest or most difficult samples.

3.3. Dilated Convolution

The traditional convolution process involves traversing the input image from top to bottom and from left to right through a sliding window, as shown in Figure 7. Hollow convolution (also known as dilated convolution or dilated convolution) is different from traditional convolution in that it inserts several zeros between convolution kernels to expand the coverage range of the kernel. Hollow convolution can aggregate larger-scale information, and under the same kernel size, hollow convolution has a larger perceptual domain, which can reduce information loss.

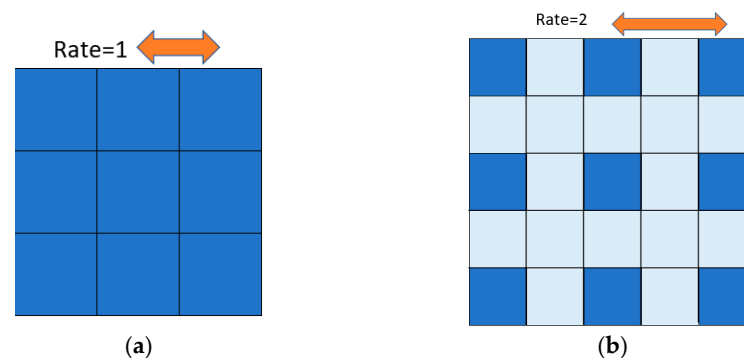


Figure 7. Ordinary convolution and dilated convolution: (a) ordinary convolution; (b) dilated convolution.

In the algorithm network structure of this article, the encoding part introduces dilated convolution. The first and second layers of the network use two 3×3 dilated convolutions, while the third and fourth layers use a single 3×3 dilated convolution with an expansion rate of 2. At this point, the perceptual domain becomes 5×5 , which is equivalent to a 3×3 dilated convolution and a 5×5 ordinary convolution. Continuous concatenation of several dilated convolutions enables the network to extract more comprehensive features and obtain richer information, thereby improving the segmentation accuracy.

4. Experimental Results and Analysis

4.1. Datasets

This article uses the CASIA-Iris V4 database of the Chinese Academy of Sciences for validation. CASIA-Iris V4 is an extension of CASIA-Iris V3 and contains six subsets.

The three subsets from CASIA-Iris V3 are CASIA-Iris-Interval, CASIA-Iris-Lamp, and CASIA-Iris-Twins. The three new subsets are CASIA-Iris-Distance, CASIA-Iris-Thousand, and CASIA-Iris-Syn. CASIA-Iris V4 contains a total of 54,601 iris images from more than 1800 genuine subjects and 1000 virtual subjects. All Iris images are 8-bit gray-level JPEG files, collected under near-infrared illumination or synthesized. Some statistics and features of each subset are given in Table 1. The six datasets were collected or synthesized at different times, and CASIA-Iris-Interval, CASIA-Iris-Lamp, CASIA-Iris-Distance, and CASIA-Iris-Thousand may have a small inter-subset overlap in subjects. A total of 10,000 iris images were selected as experimental subjects, including images of eyelashes, eyelid occlusion, hair occlusion, individual differences, and white spot interference caused by corneal reflection. The improved YOLOv8 algorithm for the target recognition process is shown in Figure 8.

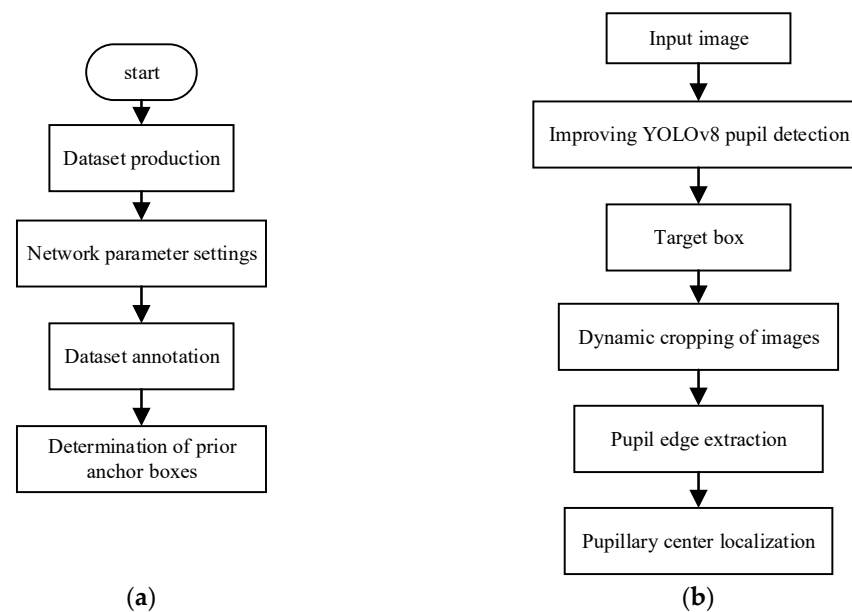


Figure 8. Process diagram of pupil center positioning method: (a) Algorithm development process. (b) Iris detection process.

Firstly, we created corresponding labels for the above images as training and validation sets, and we trained and validated the model proposed in this paper. The dataset and labels are shown in Figure 9.

The YOLOv8n model was used to train the above dataset to obtain the best model file and corresponding validation set data information, as shown in Figure 10. Figure 10 shows the model training results of the graph, where the loss function is used to measure the degree of difference between the prediction and the true value of the model, which to a great extent determines the performance of the model. Figure 10 shows the model training results, includes the location loss box_loss, the loss of classification cls_loss and the evaluation indicators. The error between the prediction frame and the calibration frame (GIoU); the smaller the location, the more accurate it is. The training set reaches 0.27726, and the verification set reaches 0.29232. The calculation of the anchor frame and the corresponding calibration classification is correct, and the smaller the classification, the more accurate it is, where the training set reaches 0.14171 and the verification set reaches 0.13225. The loss of confidence obj_loss gives the confidence of the calculation network, where the smaller the ability to determine the target, the more accurate it is; the training set reaches 0.78801, and the verification set reaches 0.78403. The evaluation indicators: precision, recall, mAP, and mAP@0.5:0.95; the evaluation indicators of the improved YOLOv8n proposed in this paper are 0.99971, 1, 0.995, and 0.96495, respectively. The data visualization results and the visualization results of the training data are shown in Figures 11 and 12, respectively.

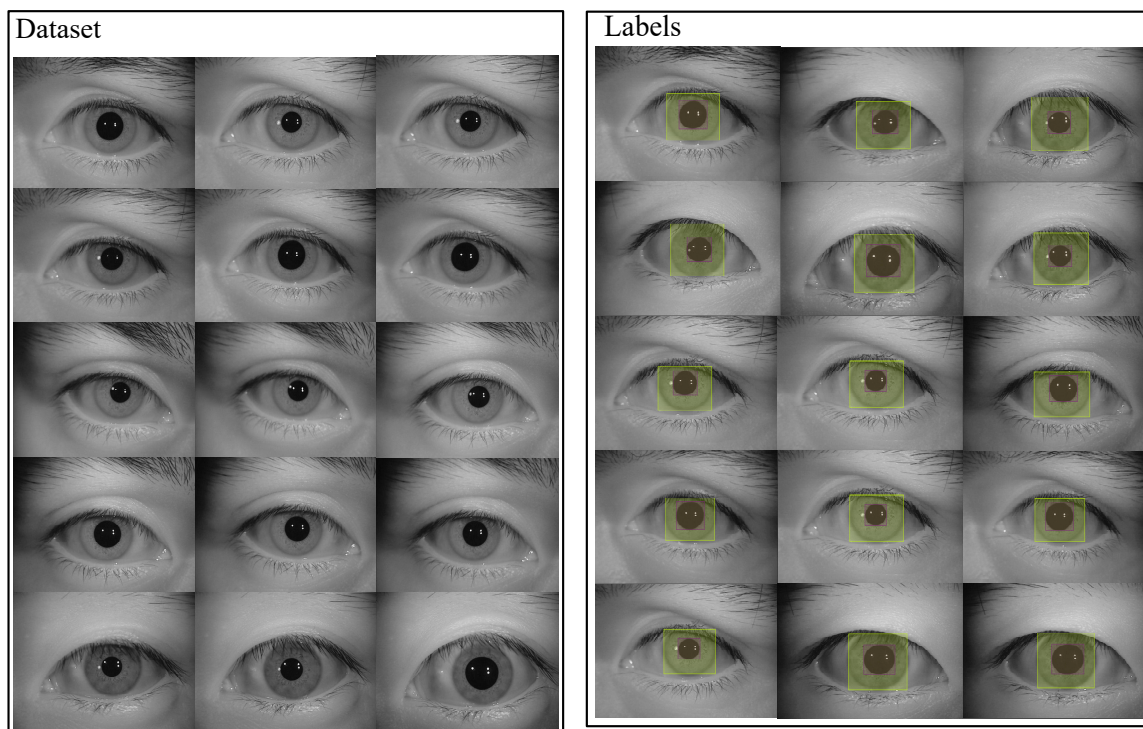


Figure 9. Datasets and labels.

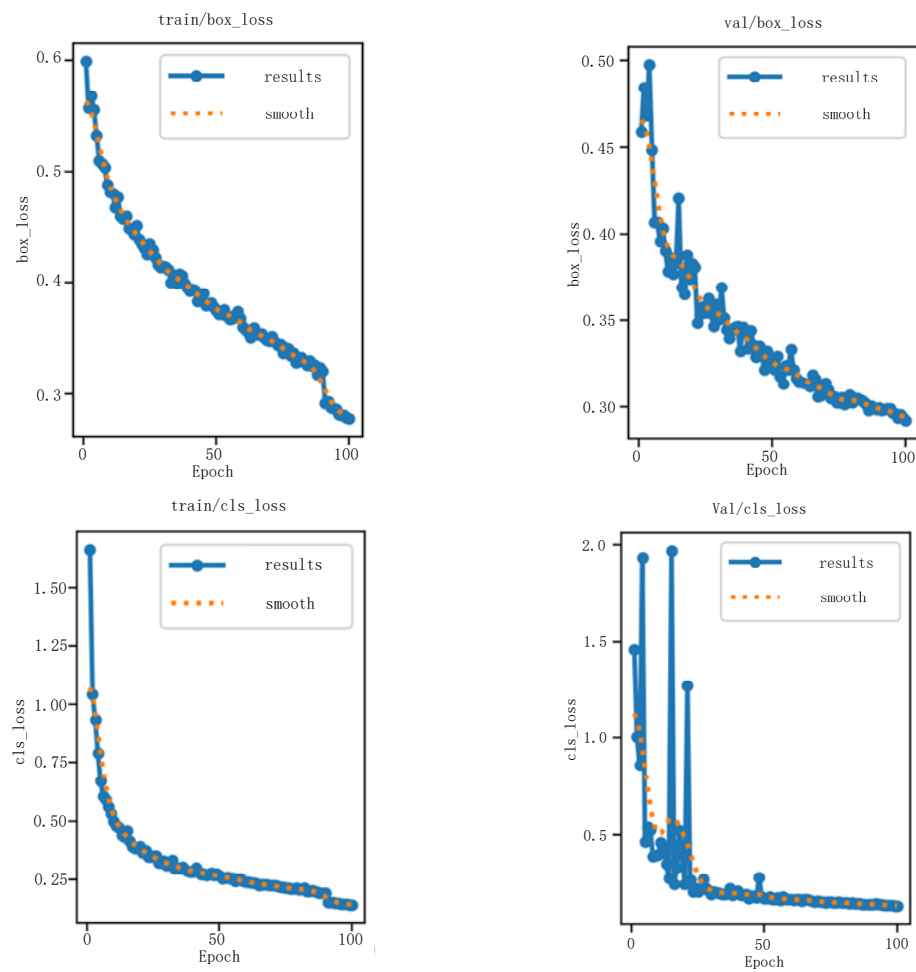


Figure 10. Cont.

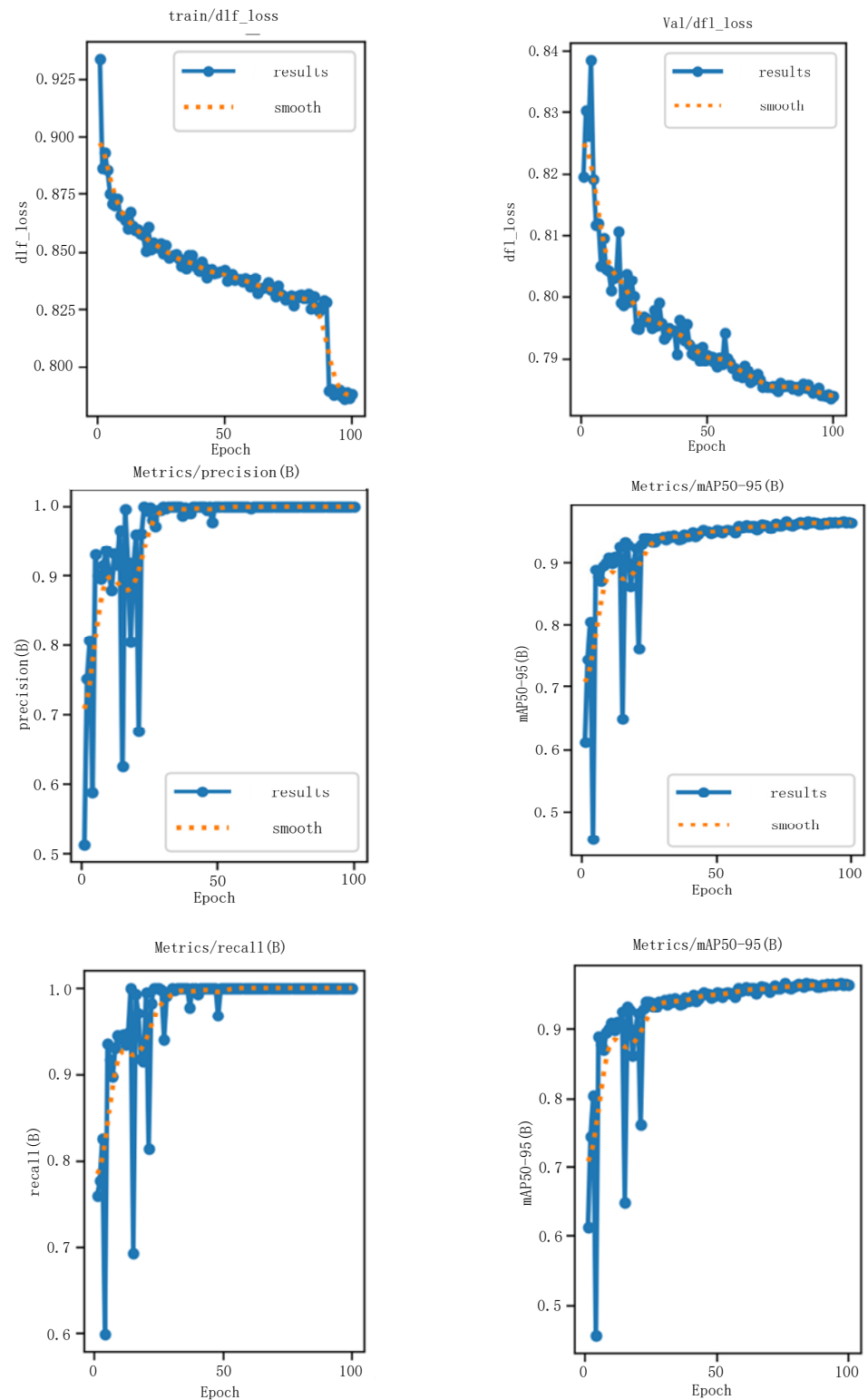


Figure 10. Model training results.

We selected the images to be predicted in the test set and put them into the model for prediction. The following recognition results and corresponding segmentation images were obtained, as shown in Figure 13. More recognition results are shown in Appendix A.

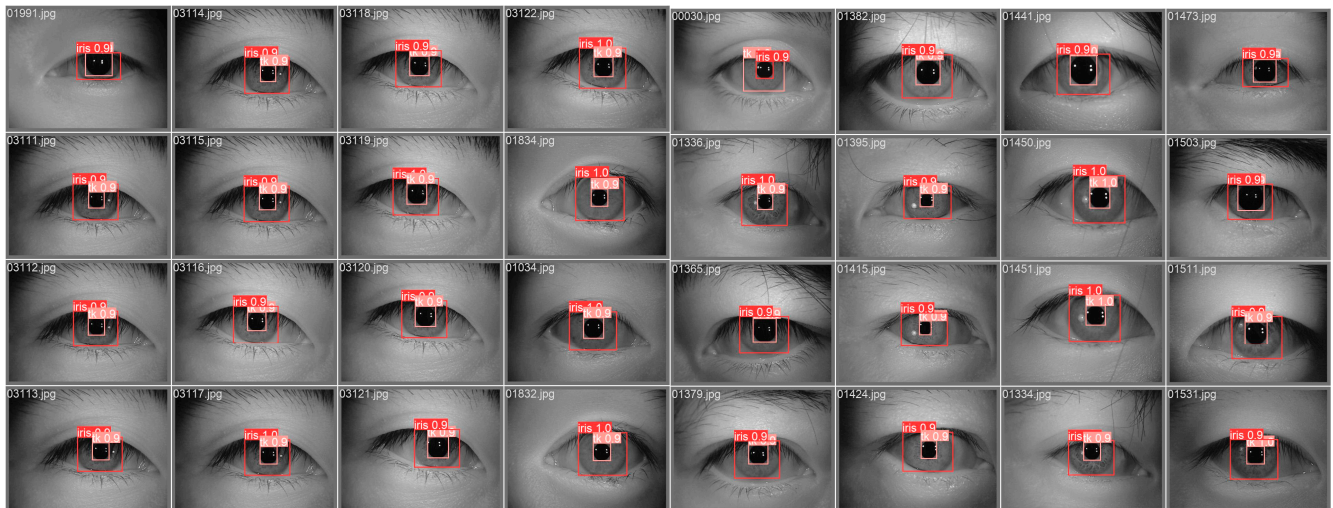


Figure 11. Validation results.

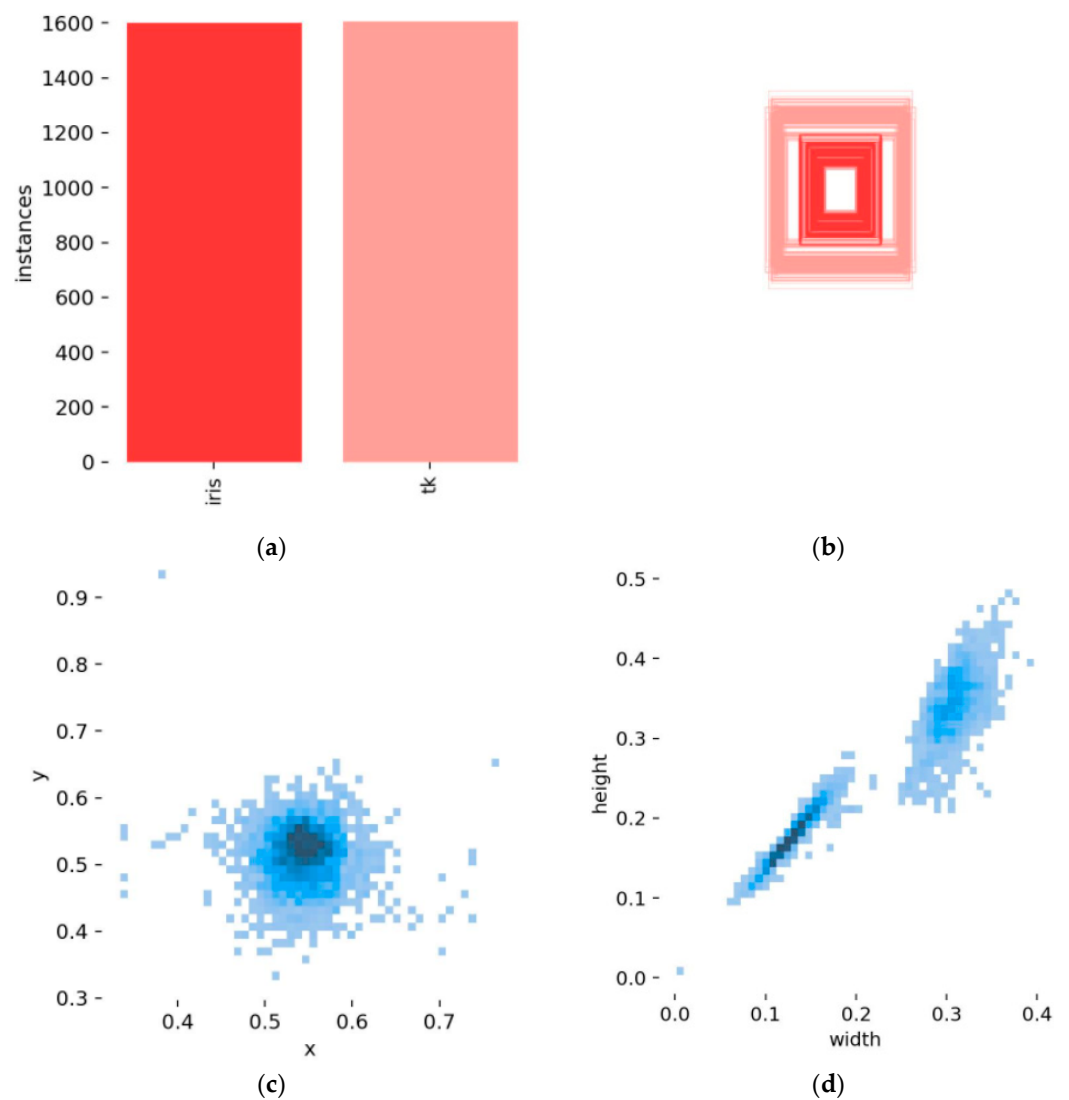


Figure 12. Visualization results of the training data: (a) Histogram illustrating the number of instances for each category. (b) Distribution of bounding boxes for all data. (c) Histogram of x and y variables, displaying the spatial distribution of the dataset. (d) Histogram of width and height variables, illustrating the dataset's distribution.

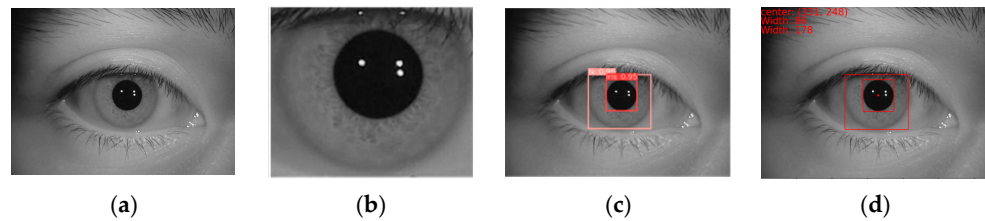


Figure 13. Recognition results and segmented images: (a) original image; (b) ROI image; (c) target recognition frame selection image; (d) target recognition result image.

4.2. Experimental Configuration and Evaluation Indicators

The experimental system used in this article was the Windows 11 operating system, and the experimental computer was configured with an NVIDIA GeForce RTX 4060 graphics card with 8 GB of memory. The CPU was an Intel (R) Core (TM) i7-13700H, and the model was deployed under the Python framework. The Python framework of the model was PyTorch 2.0, and the programming language was Python 3.9, (video training version) CUDA11.7. The hyperparameters of the training model were set as follows: the size of the training batch was Batch_size = 96, the training time of the model was epoch = 100, and the initial learning rate was 0.001. PyTorch 2.0 offers the same eager-mode development and user experience, while fundamentally changing and supercharging how PyTorch operates at the compiler level under the hood. PyTorch 2.0 introduces the torch.compile feature, which takes PyTorch's performance to new heights and begins moving some of PyTorch back from C++ into Python. This is an important new direction for PyTorch—hence, 2.0. torch.compile is a fully additive (and optional) feature, so 2.0 is by definition 100% backward-compatible. Underpinning torch.compile are new technologies—TorchDynamo, AOTAutograd, PrimTorch, and TorchInductor.

This experiment uses precision, recall, and mean average precision (mAP) as key indicators to evaluate the performance of the model. Precision measures the accuracy of model detection (i.e., positive predictive value), while recall evaluates the comprehensiveness of model detection (i.e., sensitivity). In the mean accuracy mAP , m represents the mean, and aa is the average accuracy for this type of sample when the IOU threshold of the confusion matrix is set to 0.5. $mAP@0.5$ takes the average accuracy of samples from all categories, reflecting the trend of model accuracy with respect to recall. The higher the value, the easier it is for the model to maintain high accuracy under high recall, and $mAP@0.5\sim0.95$ represents the average mAP value at different thresholds (from 0.5 to 0.95, with a step size of 0.05). The formula for parameter measurement is shown in (6) to (10):

$$I_{IOU} = \frac{A \cap B}{A \cup B} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$A_P = \int_0^1 P(r) dr \quad (9)$$

$$mAP = \frac{\sum_{i=1}^N A_{P,i}}{N} \quad (10)$$

where A and B represent the predicted and actual bounding box sets in the given expression, respectively. TP (true positive), FP (false positive), and FN (false negative) represent the number of correctly predicted, incorrectly predicted, and missed safety helmet targets, respectively. $P(r)$ represents a smooth and precise recall curve and integrates it to obtain the area under the smooth curve. N represents the number of detection categories, and $A_{P,i}$

represents the accuracy of class i , where i is the index. The proximity between the predicted boundary and the actual bounding box is determined by the IOU , which is set to 0.5.

4.3. Comparative Experiment on Attention Mechanisms

In order to improve the detection accuracy of the model, an attention mechanism was added to the original YOLOv8n for comparative experiments. We compared the current mainstream attention mechanisms that combine spatial and channel attention (CBAM), squeeze and excitation (SE), channel attention (CA), and the dual attention mechanism used in this paper, which are suitable for object detection tasks. The specific experimental results are shown in Table 2. According to Table 2, the DA mechanism proposed in this study significantly improved the detection accuracy, with the highest $mAP@0.5$, at 94.013%, while $mAP@0.5:0.95$ was the highest at 60.251% (calculation can be found in Formula (10)).

Table 2. Experimental comparison of different attention mechanisms.

Algorithm	P	R	mAP@0.5	mAP@0.5:0.95
YOLOv8n	0.92411	0.8587	0.9221	0.58215
YOLOv8n + SE	0.91023	0.84991	0.91372	0.56497
YOLOv8n + CBAM	0.9227	0.86261	0.92526	0.58444
YOLOv8n + CA	0.92902	0.86654	0.93143	0.59625
YOLOv8n + DA	0.93011	0.87105	0.94013	0.60251

4.4. Comparative Experiments of Different Models

To validate the proposed improved YOLOv8n algorithm, we trained the model using the same training set and hyperparameters as the original YOLOv8n algorithm. The comparison of the model evaluation index curves is shown in Figure 14. Figure 14a shows the change curve of precision. The improved YOLOv8 algorithm has relatively small fluctuations in the curve and relatively high values, indicating that the improved YOLOv8 algorithm has good performance in model precision. Figure 14b shows the recall change curve. During the model training process, the improved YOLOv8 algorithm has a higher curve value, indicating its strong ability to predict positive samples correctly. Figure 14c,d show the variation process of the average precision mean curve. The improved YOLOv8 algorithm has less volatility and a more stable model. By comparison with the various evaluation indicators of the benchmark model, the results show that the improved YOLOv8 algorithm has better performance and meets the requirements for effective monitoring of human pupils.

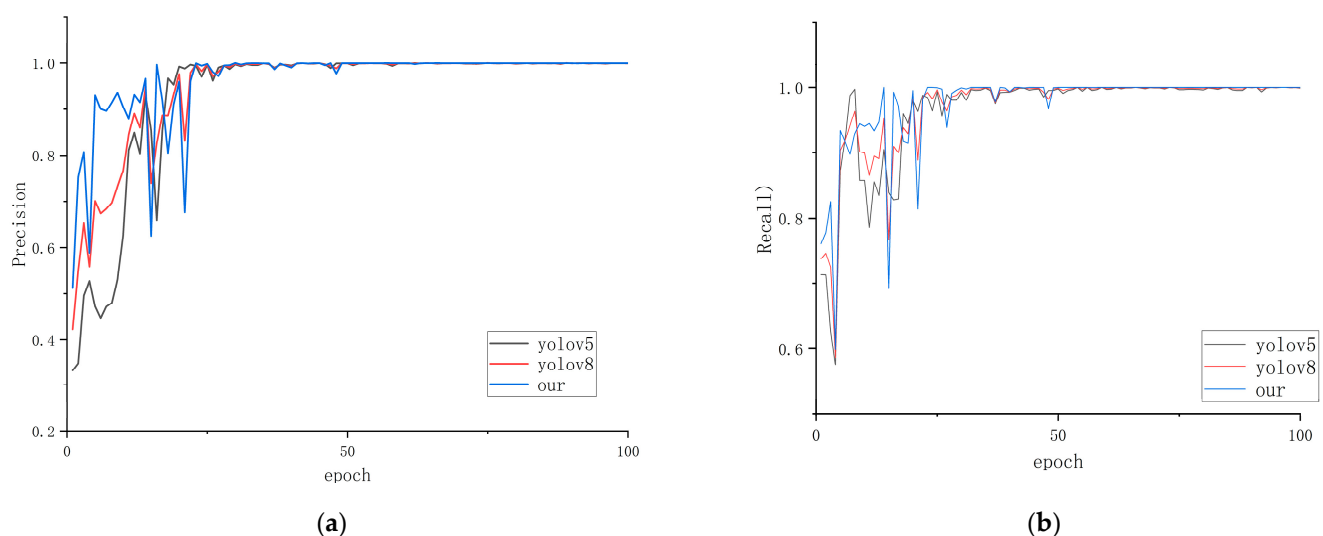


Figure 14. Cont.

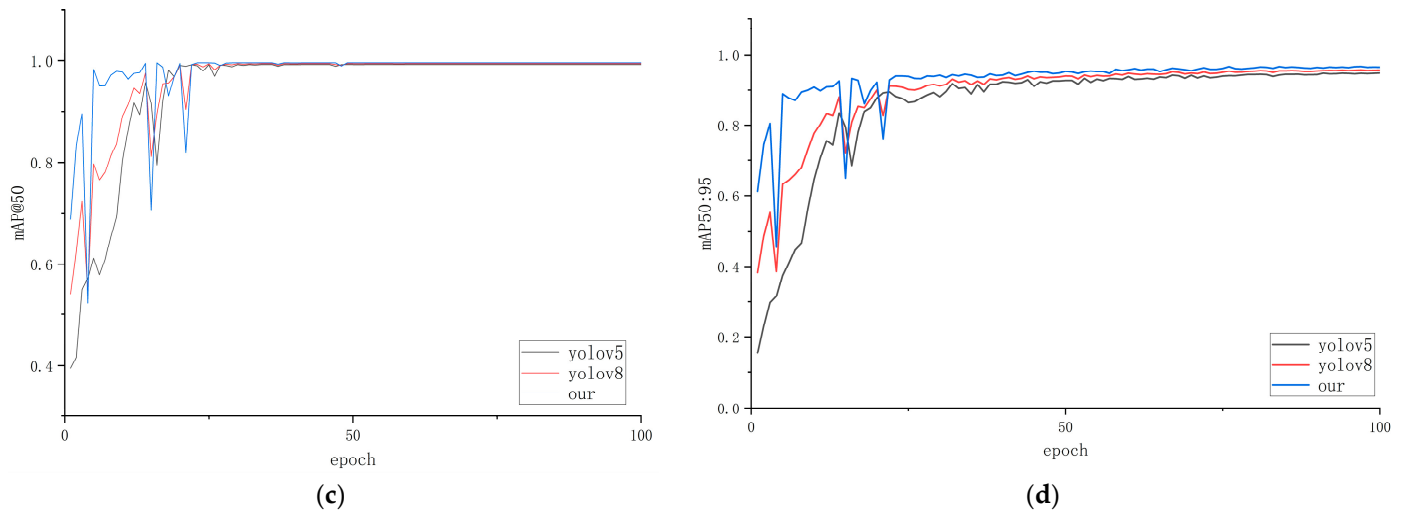


Figure 14. Comparison of model evaluation index curves: (a) precision; (b) recall; (c) mAP@0.5; (d) mAP@0.5~0.95.

5. Discussion

In order to further verify the performance of this algorithm, a new dataset was created, which was selected from the original CASIA-Iris-Interval v4, including different image noise (such as eyelids, eyelashes, hair, and other occlusions), with a total of 10,500 images, including part of the training set and part of the test set. Through the algorithm proposed in this paper, the test results show that the improved YOLOv8 algorithm, in terms of evaluation accuracy, recall rate, mAP50, mAP50-95, and the original test set of 1000 image recognition results, is almost the same; therefore, the improved YOLOv8 algorithm proposed in this paper improves the robustness. The specific experimental results are shown in Figure 15.

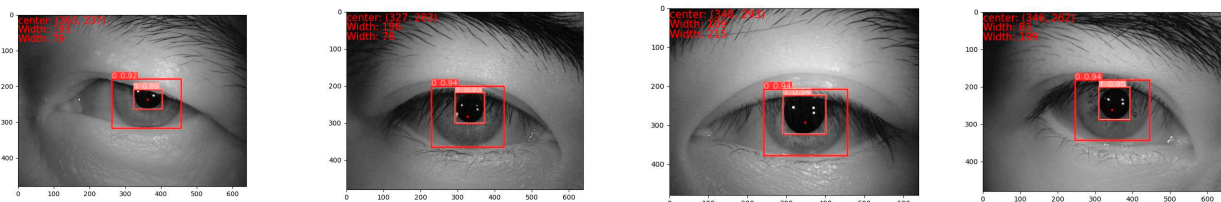


Figure 15. Partial image recognition result map containing noise.

Table 3 shows the experimental results of the algorithm compared with the YOLOv8 and YOLOv5 models. Compared with the YOLOv8 algorithm, the improved YOLOv8n algorithm achieved improvements of 0.027%, 0.058%, 0.167%, and 0.868% in precision, recall, mAP50, and mAP50-95, respectively. Moreover, under various environmental experiments, the improved YOLOv8n algorithm reduced the model parameters by 7.18%. The size of the model parameters directly determines the complexity of the model and storage space requirements, and in the model training and reasoning process, the size of the parameters directly determines the amount of calculation required, so reducing the size of the model parameters can reduce the consumption of computing resources and improve the efficiency of computing, enabling the model to complete the training and reasoning tasks faster on resource-constrained devices. The model parameters also affect the generalization ability and robustness of the model. Too many parameters may lead to overfitting and reduce the generalization ability of the model, while reducing the number of parameters will help to prevent overfitting and improve the robustness of the model. Table 4 shows the partial image processing times; the results show that the average positioning time is 18.3667 ms to meet the real-time pupil recognition requirements.

Table 3. Experimental evaluation results.

Algorithm	P	R	mAP@0.5	mAP@0.5:0.95	Parameters	FLOPs
YOLOv5	0.99913	0.99881	0.99275	0.9480	3006038	15.8
YOLOv8n	0.99944	0.99942	0.99445	0.95673	3006038	8.1
Our	0.99971	1	0.99611	0.96495	2790204	5.7
	+0.027%	+0.058%	+0.167%	+0.868%	−7.18%	−29.6%

Table 4. Detection time of partial images.

NO	Time/ms
1	15
2	15
3	15
4	17
5	16
6	15
7	17
8	15
9	21
10	27
11	22
12	22
13	15
14	21
15	18
16	14
17	15
18	21
19	21
20	15
21	21
22	21
23	15
24	15
25	21
26	21
27	21
28	19
29	24
30	16
Average	18.3667

In order to further verify the detection and location effect of this algorithm, it was qualitatively compared with the traditional basic pupil location algorithm; the detection results of the traditional radial symmetry transform algorithm and the location algorithm based on gradient means are shown in Figure 16.

The results show that there are other areas of radial symmetry in the image, such as speckles, eyelids, and so on, in addition to the pupil with radial symmetry; moreover, the gradient value at the edge of these areas is high, and the contribution of radial symmetry is higher than that of the pupil area. However, the algorithm based on gradient means can eliminate the high-gradient region and effectively avoid the influence of eyelid and speckle reflection noise. The improved YOLOv8 algorithm proposed in this paper has high accuracy because it uses a dual attention mechanism, which not only focuses on the global contextual information of the input data but also reduces the dependence on a specific region and improves the feature fusion ability of the model. Secondly, the introduction of empty convolution to replace the ordinary convolution in the original network, without sacrificing the detection accuracy, improves the detection speed of the model and can meet the real-time requirements of pupil target detection.

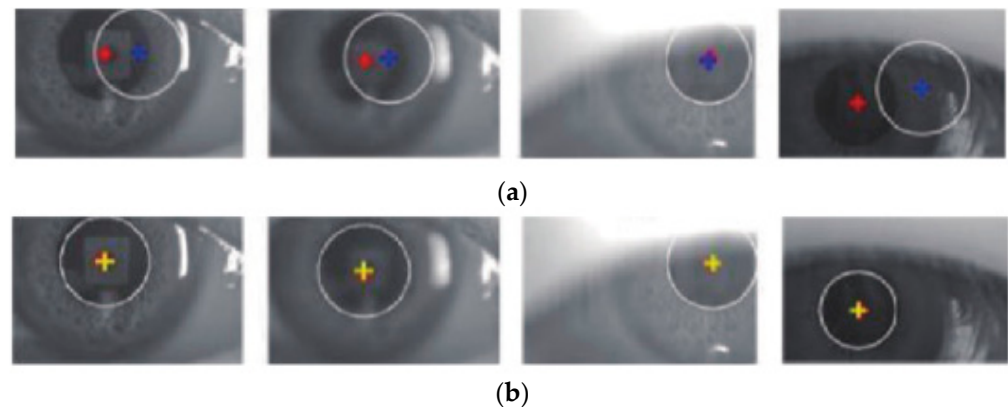


Figure 16. Location algorithm based on gradient means: (a) Traditional radial symmetry transformation algorithm. (b) Location algorithm based on gradient means. + is the artificial marking of the center of the pupil; + and + are the algorithm locates the center point of the pupil.

The improved algorithm proposed in this paper can accurately identify the location of eyelashes, eyelids, hair occlusions, incomplete pupils (upper or lower), left or right deviations, and excessively small pupils; however, for cases affected by strong light, the reflection of strong light on the surface of the eyeball makes the boundary color of the object (pupil and iris) blurred, as shown in Figure 17, which affects the feature extraction and recognition of the boundary part, making it difficult to recognize and locate the pupil. Improvements can be made by increasing the number of training samples that are affected by intense light, and by increasing the number of training categories.

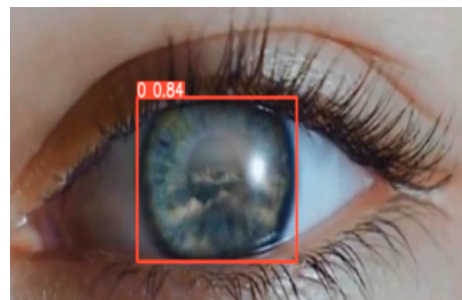


Figure 17. Images in bright light.

6. Conclusions

In order to improve the accuracy and speed of the optical biological parameter detection system for the human eye, this paper proposes a pupil target detection method based on improved YOLOv8n. Firstly, a dual attention mechanism is adopted, which not only focuses on the global contextual information of input data but also reduces the dependence on specific regions, improving the feature fusion ability and robustness of the model. Secondly, introducing dilated convolutions to replace ordinary convolutions in the original network improves the detection speed of the model without sacrificing detection accuracy. Finally, using Focaler-IoU as the damage function, by focusing on different regression samples, the performance of detectors in different detection tasks can be improved, resulting in improved bounding box regression performance of the detection model. The experimental results of the improved model on the CASIA-Iris dataset show that, compared with the original YOLOv8n, the method proposed in this paper has improved average detection accuracy and speed. The performance of the improved YOLOv8n algorithm was 0.99971, 1, 0.99611, and 0.96495 in precision, recall, MAP50, and mAP50-95, respectively. Moreover, the improved YOLOv8n algorithm reduced the model parameters by 7.18% and the computational complexity by 10.06%. The average processing time of each image was 18.37 ms. Compared with some mainstream object detection methods at present, this

algorithm also achieved certain advantages in several evaluation indicators, and it was able to effectively perform the object detection task for human pupils. Although the improved algorithm proposed in this paper can realize the recognition of the pupil and iris in the presence of eyelashes, eyelids, occlusions, and other circumstances, the effect of strong light on human eye recognition will be the focus of future research work.

Author Contributions: Conceptualization, K.X. and J.W.; methodology, K.X. and H.W.; software, K.X. and H.W.; validation, K.X. and H.W.; formal analysis, K.X.; investigation, K.X.; resources, K.X. and H.W.; data curation, K.X. and H.W.; writing—original draft preparation, K.X.; writing—review and editing, K.X. and H.W.; visualization, K.X.; supervision, K.X. and J.W.; project administration, K.X.; funding acquisition, J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Jilin Provincial Science and Technology Development Program Key R&D Project (20210203156SF), the Jilin Province Key Science and Technology Research and Development projects (20180201025GX), and the Jilin Provincial Department of Education's 13th Five-Year Plan Science and Technology Project (JJKH20190599KJ).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

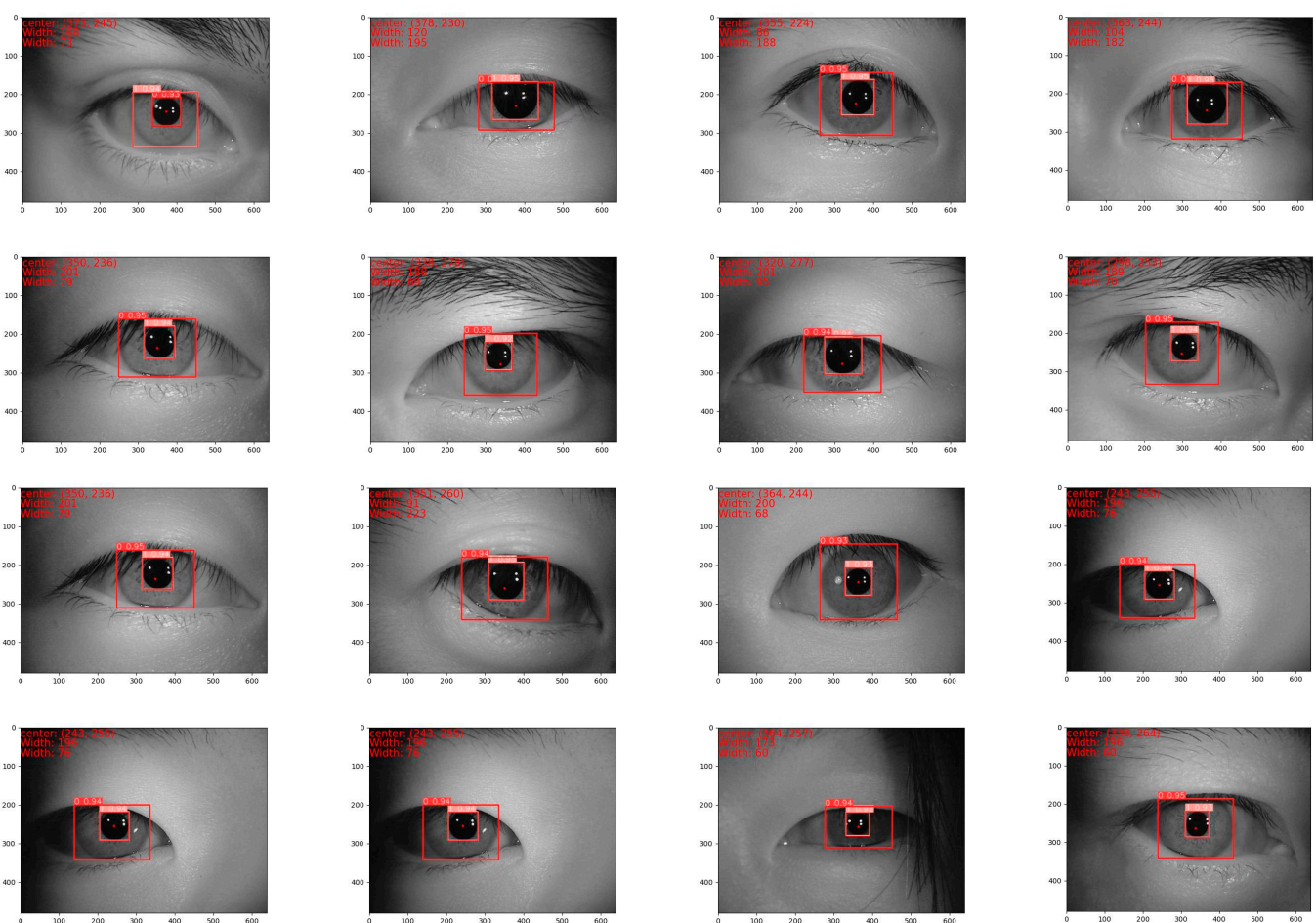


Figure A1. Cont.

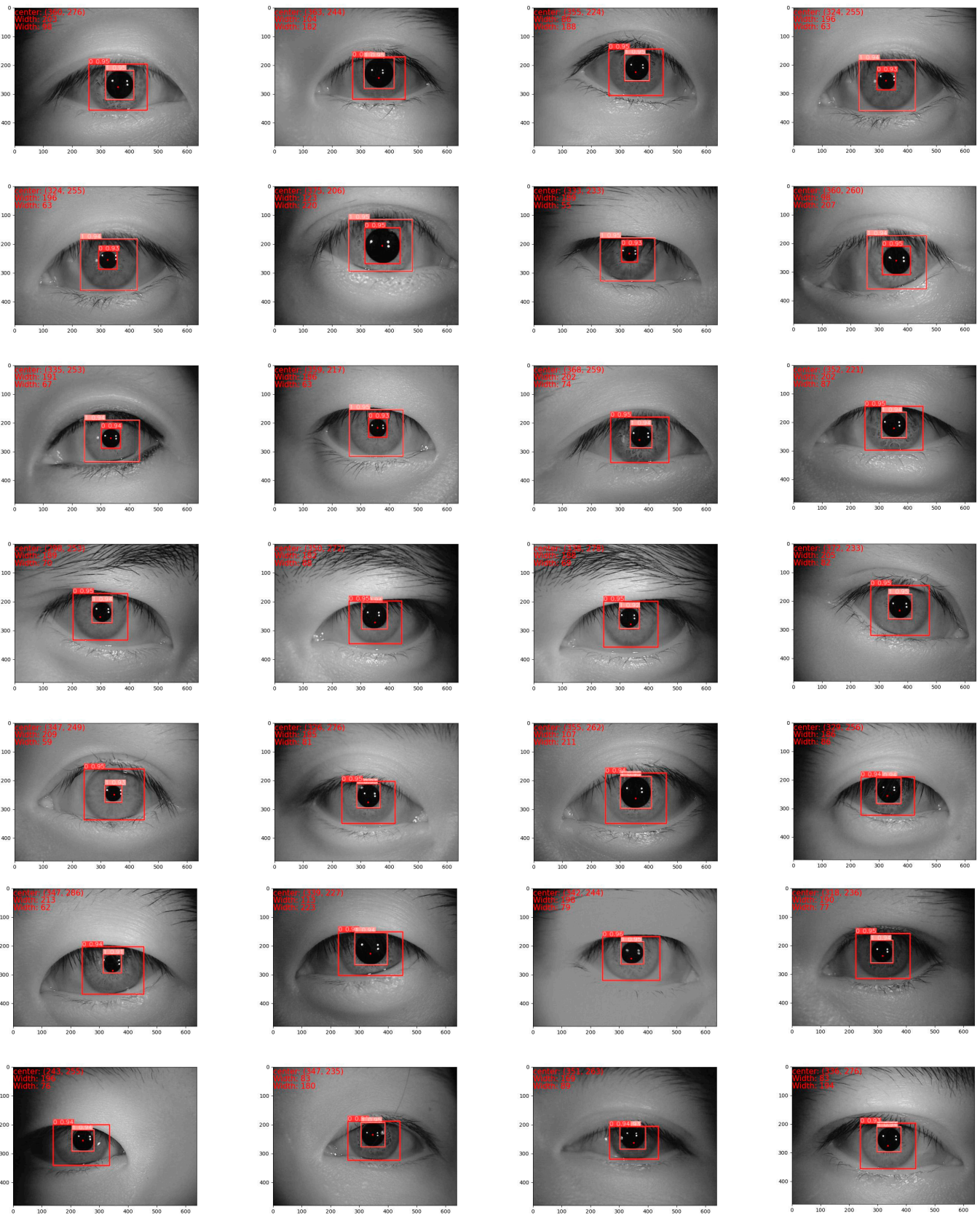


Figure A1. Cont.

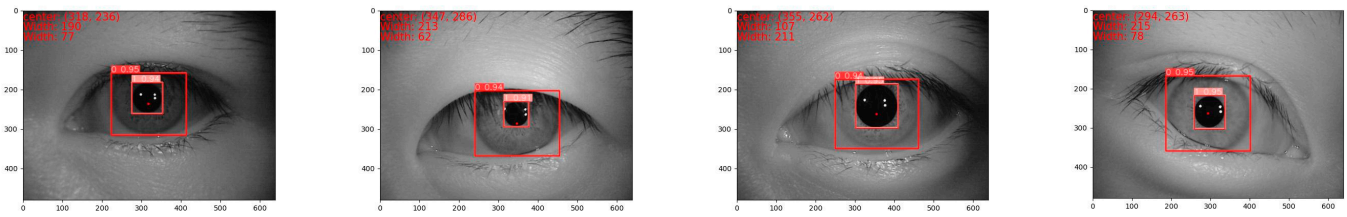


Figure A1. Diagram of part of the experimental results.

References

- Massé, B.; Ba, S.; Horaud, R. Tracking gaze and visual focus of attention of people involved in social interaction. *IEEE Trans. Pattern. Anal. Mach. Intell.* **2018**, *40*, 2711–2724. [\[CrossRef\]](#)
- Rathnayake, R.; Madhushan, N.; Jeeva, A.; Darshani, D.; Subasinghe, A.; Silva, B.N.; Wijesinghe, L.P.; Wijenayak, U. Current Trends in Human Pupil Localization: A Review. *IEEE Access* **2023**, *11*, 115836–115853. [\[CrossRef\]](#)
- Jan, F.; Min, N. An effective iris segmentation scheme for noisy images. *Biocybern. Biomed.* **2020**, *40*, 1064–1080. [\[CrossRef\]](#)
- Poulopoulos, N.; Psarakis, E. DeepPupil net: Deep residual network for precise pupil center localization. In Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022), Setúbal, Portugal, 6–8 February 2022; Scitepress: Setúbal, Portugal, 2022; pp. 297–304.
- Timm, F.; Barth, E. Accurate Eye Centre Localisation by Means of Gradient. *Visapp* **2011**, *11*, 125–130.
- Nugroho, R.H.; Nasrun, M.; Setianingsih, C. Lie Detector with Pupil Dilation and Eye Blinks Using Hough Transform and Frame Difference Method with Fuzzy Logic. In Proceedings of the 2017 International Conference on Control, Electronics, Renew—able Energy and Communication (ICCREC), Yogyakarta, Indonesia, 26–28 September 2017.
- Wang, P.; Wen, H.T.; Wang, S. High-precision Pupil Center Positioning Method Based on Near-eye Infrared Image. *J. Harbin Univ. Sci. Technol.* **2022**, *27*, 38–46.
- Cai, H.Y.; Shi, Y.; Lou, S.L.; Wang, Y.; Chen, W.G.; Chen, X.D. Pupil location algorithm applied to infrared ophthalmic disease detection. *Chin. Opt.* **2021**, *14*, 605–614.
- Vranceanu, R.; Florea, C.; Florea, L.; Vertan, C. Gaze direction estimation by component separation for recognition of Eye Accessing Cues. *Mach. Vis. Appl.* **2015**, *26*, 267–278. [\[CrossRef\]](#)
- Zhang, H.; Wang, S.; Li, X.; Ma, H.; Cui, J.; Zhu, J.; Li, M.; Gao, Y.; Han, L.; Zhou, B. Research and implementation of pupil recognition based on Hough transform. *Chin. J. Liq. Cryst. Disp.* **2016**, *31*, 621–625. [\[CrossRef\]](#)
- Loy, G.; Zelinsky, A. Fast radial symmetry for detecting points of interest. *IEEE Trans. Pattern Anal.* **2003**, *25*, 959–973. [\[CrossRef\]](#)
- Ning, X.; Mu, L. Research on pupil localization algorithm based on gradient direction constraint. *Foreign Electron. Meas. Technol.* **2021**, *40*, 115–121.
- Fuhl, W.; Santini, T.; Kasneci, G.; Kasneci, E. Pupil Nte: Convolutional neural networks for robust pupil detection. *Rev. Odontol. Unesp* **2016**, *19*, 806–821.
- Chinsatit, W.; Saitoh, T. CNN-based pupil center detection for wearable gaze estimation system. *Appl. Comput. Intell. Soft Comput.* **2017**, *2017*, 8718956. [\[CrossRef\]](#)
- Ma, Q.; Han, H.; Han, X.; Zhang, Z. Pupil center location based on star ray method. *Comput. Eng. Des.* **2021**, *42*, 1409–1417.
- Fuhl, W.; Eivazi, S.; Hosp, B.; Eivazi, A.; Rosenstiel, W.; Kasneci, E. BORE: Boosted-oriented edge optimization for robust, real time remote pupil center detection. In Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, Warsaw, Poland, 14–17 June 2018. [\[CrossRef\]](#)
- Hosp, B.; Eivazi, S.; Maurer, M.; Fuhl, W.; Geisler, D.; Kasneci, E. RemoteEye: An open-source high-speed remote eye tracker: Implementation insights of a pupil- and glint-detection algorithm for high-speed remote eye tracking. *Behav. Res. Methods* **2020**, *52*, 1387–1401. [\[CrossRef\]](#)
- Donuk, K.; Hanbay, D. Pupil center localization based on mini U-net. In Proceedings of the International Artificial Intelligence and Data Processing Symposium, (IDAP), Sofia, Bulgaria, 22 May 2022; pp. 185–191.
- Chen, G.; Dong, Z.; Wang, J.; Xia, L. Pupil localization algorithm based on improved U-Net network. *Electronics* **2023**, *12*, 2591. [\[CrossRef\]](#)
- Song, B.; Du, W.; Duan, N.; Li, X. Research on pupil location algorithm of non-contact tonometer. *Electron. Meas. Technol.* **2022**, *45*, 112–117.
- Sun, Y.; Liu, W.; Jiang, M. Pupil location algorithm based on Attention Gate and dilated convolution. *Electron. Meas. Technol.* **2023**, *46*, 126–132.
- Xu, D.; Wang, L.; Li, F. Review of Typical Object Detection Algorithms for Deep Learning. *Comput. Eng. Appl.* **2021**, *57*, 10–25.
- Li, X. Research on Pupil Center Location and Tracking Based on Near-Eye Infrared Video Images. Master's Thesis, North University of China, Taiyuan, China, 2024.
- Wang, X.; Gao, H.; Jia, Z. Improved road defect detection algorithm based on YOLOv8. *Comput. Eng. Appl.* **2024**, *13*, 2413.
- Zhang, L.; Sun, Z.; Tao, H.; Hao, S.; Yan, Q.; Li, X. Research on real-time monitoring method of mine personnel protective equipment with improved YOLOv8. *Coal Sci. Technol.* **2024**, 1–12. Available online: <https://link.cnki.net/urlid/11.2402.td.20240527.1700.004> (accessed on 27 July 2024).

26. Lin, B.Y. Safety Helmet Detection Based on Improved YOLOv8. *IEEE Access* **2024**, *12*, 28260–28270. [[CrossRef](#)]
27. Tian, P.; Mao, L. Improved YOLOv8 Object Detection Algorithm for Traffic Sign Target. *Comput. Eng. Appl.* **2024**, *60*, 202–212.
28. Hu, M.; Wang, R.; Zhang, W.; Zhang, Q. Multi-scale Referring Image Segmentation Based on Dual Attention. *J. Comput.-Aided Des. Comput. Graph.* **2024**, 1–10. Available online: <https://link.cnki.net/urlid/11.2925.tp.20240531.1456.002> (accessed on 27 July 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.