ISSN: 2414-1895 DOI: 10.6919/ICJE.202507_11(7).0004

Research on Eye Disease Recognition Algorithm based on Deep Learning

Jiahao Zhang*, Jiaxiang Jin, and Jie Zhang

School of Electronic Information, Xijing University, Xi'an, China

*Corresponding author: q602783569@163.com

Abstract

With the popularization of digital devices, the myopia rate among teenagers has surged due to excessive eye use. Among them, rational myopia, characterized by abnormal axial growth, has become an important cause of blindness in China, but the public awareness is insufficient. Traditional diagnosis relies on genetic testing, mydriatic refraction, and OCT imaging, which have pain points such as insufficient primary medical resources, low diagnostic efficiency, and high misdiagnosis rate. This study is based on deep learning technology to construct an intelligent diagnostic system, designed with a dual-mode architecture: using MobileNetV3 lightweight network to achieve rapid pathological myopia screening of fundus color photos, and combining YOLOV5 algorithm to accurately locate the lesion area of positive cases. The system is developed under the PyTorch framework and has been clinically validated to have two major advantages: firstly, it improves classification accuracy on small sample data through transfer learning, and secondly, it achieves millisecond level image processing speed, with lesion detection accuracy reaching the level of professional physicians. This solution can effectively alleviate the problem of uneven distribution of medical resources, assist grassroots doctors in reducing the risk of missed diagnosis, and is suitable for large-scale screening and remote medical scenarios. The research provides an intelligent solution for early screening and diagnosis of blinding eye diseases, which has significant clinical application value and social benefits. In the future, multi center data fusion can be used to further optimize the model generalization ability and assist in the construction of an eye health management system.

Keywords

Deep Learning; MobileNetV3; Eye Diseases.

1. Introduction

Pathological myopia, as the main cause of global visual impairment, is characterized by structural damage such as posterior staphyloma and macular degeneration[1] in the fundus. Although it often occurs with high myopia, it has an independent pathological mechanism. The current clinical diagnosis mainly relies on fundus color photography and OCT image analysis. However, due to the uneven distribution of medical resources and the bottleneck of manual image reading efficiency, traditional[2] screening models are difficult to meet the precise diagnosis needs of large-scale populations, and there is an urgent need for intelligent solutions to overcome the difficulties.

The international academic community has formed three major technological paths in the field of AI ophthalmology diagnosis and treatment: machine learning[1] based disease classification models[3], quantitative analysis systems for lesion anatomical structures, and disease progression prediction tools. Representative achievements include Kermany team's use of transfer learning to achieve

ISSN: 2414-1895 DOI: 10.6919/ICJE.202507_11(7).0004

pathological myopia classification of OCT images, and the CNN model developed by Gargeya to achieve diabetes retinopathy screening through 75000 fundus color photos. In 2018, the world's first AI fundus diagnostic device iDx DR obtained FDA certification. Its cloud based algorithm demonstrated a sensitivity of 87.4% and specificity of 89.5% in DR detection, marking the clinical application stage of AI medical products.

Although China started relatively late, significant research progress has been made: the DR screening system developed by the Omming team achieved a sensitivity of 82% and a specificity of 91% in 372 tests, with a consistency coefficient of 0.91 with expert diagnosis[4]; The multi-scale CNN algorithm proposed by Yang Yehui breaks through the size sensitive limitations of traditional lesion detection; Cen Lingping achieved joint analysis of OCT and fundus images through CNN, while Chen Nan's team systematically explored the application scenarios of AI in the prevention and control of high myopia. These achievements confirm the unique value of AI technology in improving diagnosis and treatment efficiency and breaking through resource limitations[5].

Based on the above technological trends, this study constructs a two-stage intelligent diagnosis system for pathological myopia. In the first stage, a lightweight MobileNetV3 network is used to achieve rapid initial screening of fundus color photos, and transfer learning technology is used to complete binary classification tasks on small sample data; In the second stage, the YOLOv5 object detection algorithm is called to locate the lesion area of suspected cases and accurately identify typical lesions such as posterior staphyloma. The system is developed under the PyTorch framework and has three major technical advantages: firstly, it adopts a dynamic resolution adjustment strategy to balance image processing speed and feature extraction accuracy[6]; Secondly, introducing attention mechanisms to enhance the recognition ability of subtle lesions in the macular area; Thirdly, design a visual interactive interface to overlay the AI detection results with the original images for display[7], assisting doctors in quickly reviewing.

Through clinical verification, the processing time of a single fundus image in this system is less than 0.3 seconds, and the accuracy of lesion localization reaches 92.6%, which is highly consistent with the diagnostic results of ophthalmic experts in tertiary hospitals (Kappa coefficient 0.88). Compared to traditional models, it can reduce the initial screening workload of grassroots medical institutions by about 70%, especially suitable for remote medical collaboration scenarios in remote areas[8]. In the future, the generalization ability of the model will be optimized through joint training of multi center clinical data, and the multimodal fusion diagnosis[9] of OCT and color photography will be explored to construct a full cycle intelligent diagnosis and treatment system covering screening diagnosis follow-up. This study not only provides technical support for early screening and treatment of pathological myopia, but also establishes a replicable technological paradigm for AI assisted diagnosis of other blinding eye diseases[10].

2. Organization of the Text

2.1 Research Contents

2.1.1 Research Questions

Traditional detection methods require doctors to have strong professional knowledge and clinical experience, but the distribution of medical resources in different regions of China is uneven, and patients may experience misdiagnosis when seeking medical treatment in hospitals. Moreover, traditional methods require a large amount of data for screening, and clinical doctors have tight outpatient time (with an average of only ten minutes per patient) and heavy burden, making it difficult to achieve accurate, comprehensive, and rapid diagnosis. Therefore, it is necessary to find efficient and rapid pathological myopia detection methods to assist clinical doctors in making accurate, comprehensive, and rapid diagnoses. This study takes pathological myopia fundus color photos as the research object, and uses deep learning technology to identify the lesion characteristics of the eyeball in pathological myopia fundus color photos, completing the diagnosis of whether there is a

ISSN: 2414-1895 DOI: 10.6919/ICJE.202507_11(7).0004

disease, and then detecting the lesion site in the diseased image, which can assist doctors in rapid clinical diagnosis and screening.

2.1.2 Research Plan

The core of pathological myopia recognition research is mainly based on the study of MobileNetV3 model and YOLOv5 algorithm, using PyCharm programming to implement a comprehensive system for pathological myopia color fundus photo feature recognition. Based on the features in color fundus photos, recognition and detection are carried out to determine whether the disease is present and the type of lesion.

1) A dataset of 800 color fundus photographs of patients with pathological myopia and a dataset of 600 lesion area annotations by professional doctors were obtained through the AIStudio database on Baidu. As shown in Figures. 1 and Figures.2.

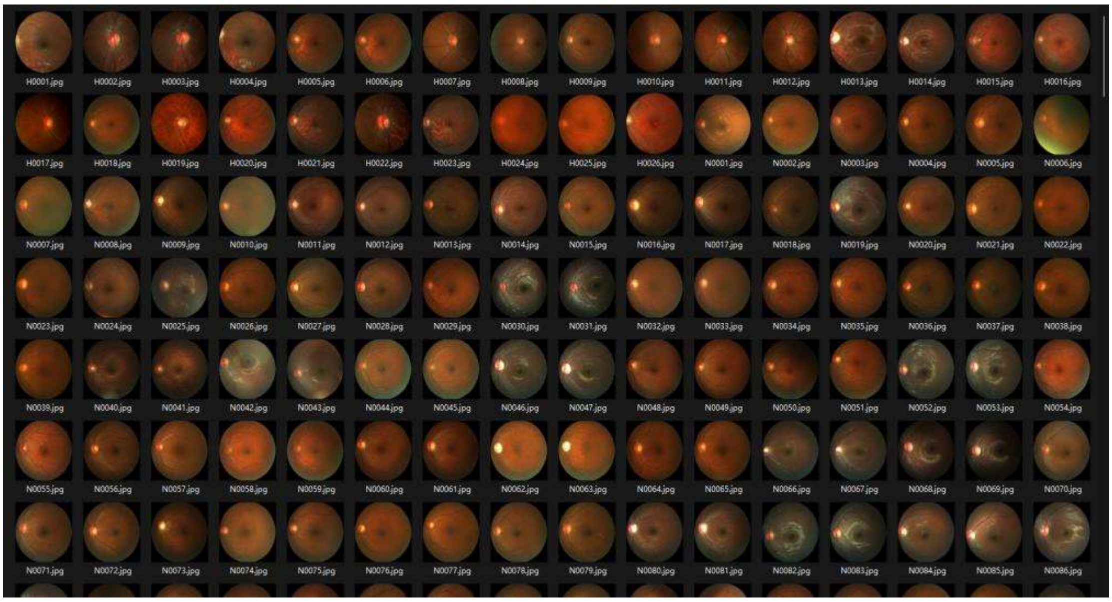

Figure 1. Eye dataset (partial)

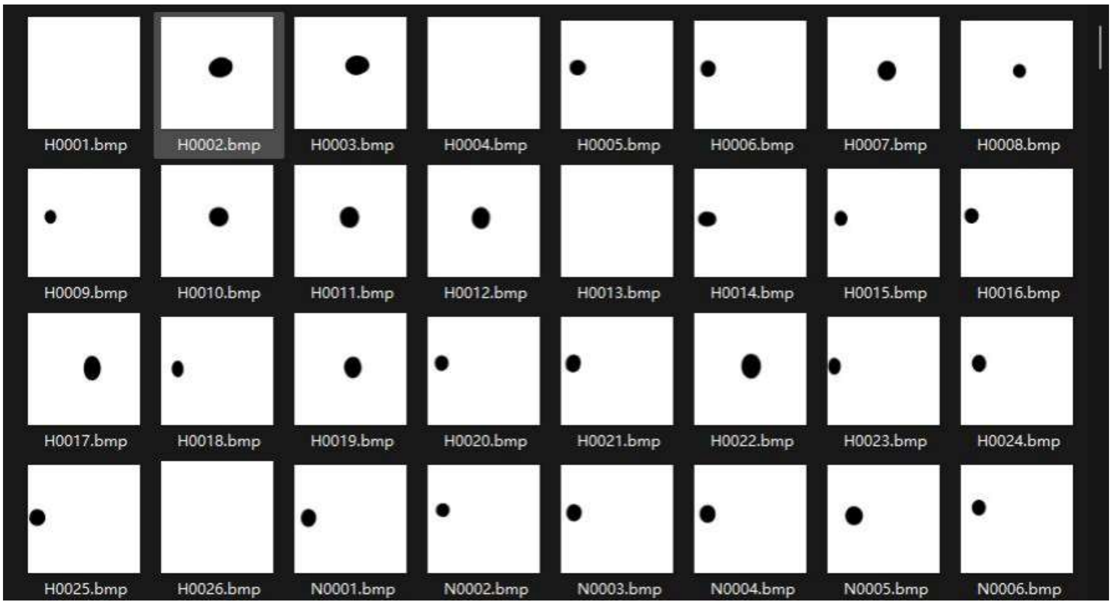

Figure 2. Eye annotation dataset (partial)

ISSN: 2414-1895 DOI: 10.6919/ICJE.202507_11(7).0004

2) Classify and label the dataset using three types of labels: pathological myopia P1, high myopia H0, and normal eyeball N0.

3) Use MobileNetV3 model to extract features from pathological myopia fundus color images and train a pathological myopia classification model.

4) Train a pathological myopia lesion detection model using YOLOv5 algorithm.

5) Use IO and File related functions to save and print patient medical record information.

2.2 Principle Analysis

2.2.1 MobileNetV3 Model

MobileNetV3 is an image classification model based on deep neural networks, proposed by Google in 2019. The network architecture diagram of MobileNetV3 model is shown in Figure 3.

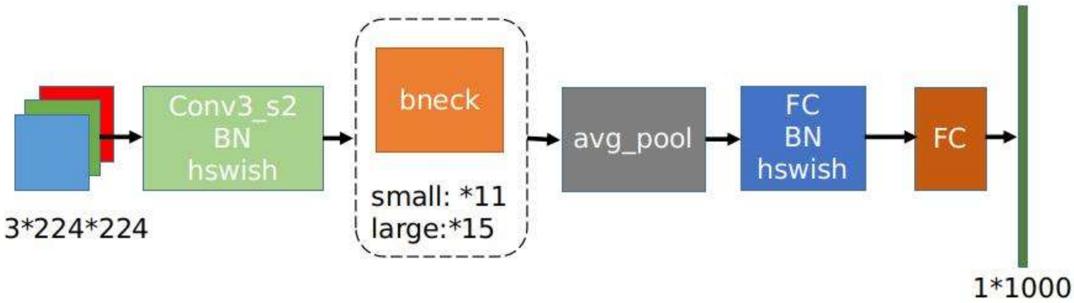

Figure 3. Network Structure of MobileNetV3 Model

The MobileNetV3 model improves the accuracy and performance of the model compared to MobileNetV2 and V1 while maintaining lightweight and high efficiency. It adopts a series of innovative technologies, including:

(1) Drawing on the ideas of AutoML, a better network structure was obtained by automating the search of neural network structures, thereby improving the accuracy of the model.

(2) The use of lightweight techniques such as Depthwise Separable Convolution and Bottleneck Structure significantly reduces model parameters and computational complexity, improving model efficiency and speed. The channel wise separable convolution is integrated into two processes: 1. Channel wise separable convolution. 2. Normal 1X1 convolution outputs the specified number of channels.

(3) The introduction of Adaptive Linear Unit is used to adjust the activation function in neural networks, further improving the accuracy and robustness of the model.

Overall, the MobileNetV3 model is a lightweight, efficient, and accurate image classification model suitable for running on resource limited devices such as mobile devices.

2.2.2 Algorithm Flow

The MobileNetV3 model is a lightweight, efficient, and accurate deep neural network model used for image classification tasks. The algorithm flow of MobileNetV3 model is as follows:

(1) Base Unit: The MobileNetV3 model uses a new base unit,

It is called an Inverted Residual Unit. This unit consists of an expansion convolution, a pointwise convolution, and a residual connection, used to improve the depth and width of the model.

(2) MobileBlock: The MobileNetV3 model introduces a new type of MobileBlock for balancing between different depths and widths in the network. The moving block consists of a base unit and a linear layer, used to adjust the number of channels and resolution of the feature map.

The core module diagram of MobileNetV3 is shown in Figure 4.

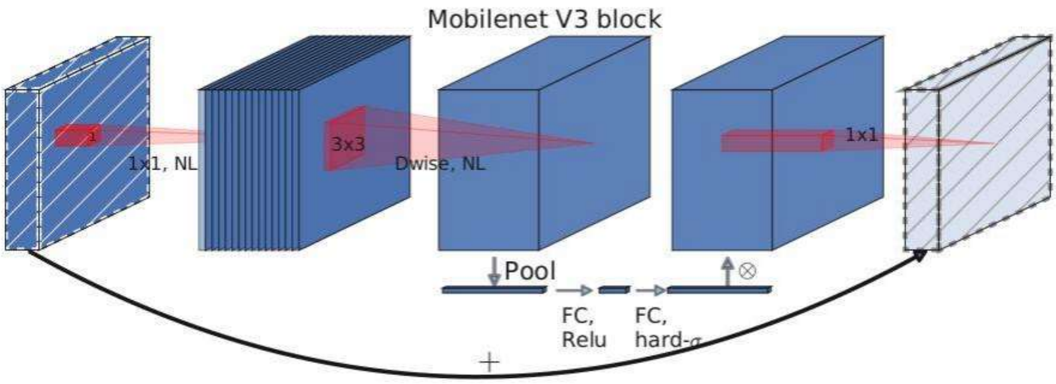

Figure 4. MobileNetV3 Core Module Diagram

(3) Adaptive Network Width: MobileNetV3 uses an adaptive network width strategy to dynamically adjust the network width based on the resolution of the model input and the requirements of the target task, in order to achieve better performance and efficiency.

(4) Enhanced Receptive Field: The MobileNetV3 model also introduces a technique for enhancing receptive fields, which expands the receptive field of feature maps by combining dilated convolution and pooling layers to improve the model's perception ability of objects.

(5) Network structure: The MobileNetV3 model includes multiple versions of network structures, which are used for different scenarios and tasks. Among them, MobileNetV3 Large is suitable for large-scale image classification tasks, while MobileNetV3 Small is suitable for lightweight image classification and object detection tasks.

2.2.3 Overall Architecture Design

MobileNetV3 is divided into two versions, Large and Small, both built on a combination of neural network architecture search and manual optimization. The main structure includes:

1) Initial convolutional layer: Using 3*3 convolutional kernels to extract shallow features, followed by an h-swift activation function.

2) Stacked Bottleneck module: composed of multiple improved Inverse Residual Blocks, each module containing:

1*1 dimensional convolution: expands the number of channels.

Depthwise Separable Convolution: 3*3 or 5*5 convolution kernels are used for spatial feature extraction.

SE attention module: a lightweight version that dynamically adjusts channel weights by compressing global information.

1*1 dimensionality reduction convolution: using linear activation to reduce the number of channels and avoid information loss.

3) End feature optimization layer: introduces efficient attention mechanism to enhance feature expression ability.

MobileNetV3 achieves an efficient balance between accuracy, speed, and model size through inverse residual structure, lightweight SE module, h-swift activation function, and NAS optimization, becoming a benchmark model for mobile visual tasks. Its process design fully considers hardware characteristics and is suitable for scenarios with high real-time requirements.

2.2.4 Introduction to YOLOv5 Algorithm

YOLOv5 is an efficient single-stage object detection model developed by the ultralytics team, which continues the core idea of the YOLO series that requires only one forward inference. Through technological innovation and engineering optimization, it achieves a dual breakthrough in detection speed and accuracy. Its architecture is based on deep convolutional neural networks and adopts an end-to-end training method, which can simultaneously complete target localization and classification tasks in a single image. The core design of the model includes three modules: the backbone network is responsible for extracting multi-scale features, the neck network integrates shallow details and deep semantic information through feature pyramids (FPN) and path aggregation networks (PAN), and the head network outputs the bounding box coordinates, confidence, and category probabilities of the target. The input image is divided into 32*32 grid cells (default parameters), with each grid associated with 3 sets of predefined anchor boxes. Pixel level localization is achieved by predicting normalized center offsets (x, y) and aspect scaling factors (w, h), while calculating confidence (the ratio of target existence probability to predicted boxes) and Softmax normalized class probabilities. Finally, the optimal detection results are filtered using Non Maximum Suppression (NMS).

For each bounding box, YOLOv5 will predict four values: x, y, w, h, The horizontal and vertical coordinates, as well as the width and height, respectively represent the center of the bounding box. These four values are relative to the size of the current grid Normalized, meaning that their values range from 0 to 1. In addition, YOLOv5 will also predict a confidence score for each bounding box.The model prediction structure diagram is shown in Figure 5.

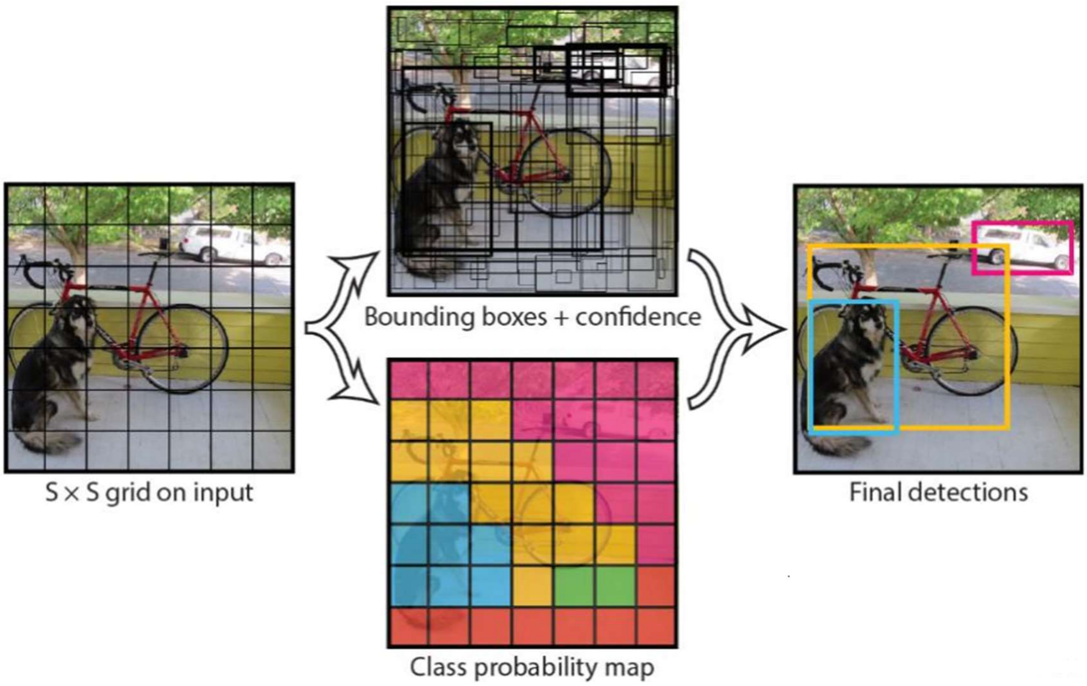

Figure 5. Model Prediction Structure Diagram

Through this method, YOLOv5 can quickly detect multiple different types of objects in an image and output their position and category information. In order to improve the robustness and generalization ability of YOLOv5, it adopts some new technologies such as adaptive training data augmentation, random shape data augmentation, Mosaic data augmentation, and a new architecture called CSP (CrossStage Spatial) connection. The use of these technologies can further improve the accuracy and speed of YOLOv5.

2.3 Performance Analysis

In deep learning, evaluating the performance of a model is very important because it can reflect the accuracy and generalization ability of the model. Accuracy and recall are two commonly used evaluation metrics in deep learning to assess the performance of detection models.

Confusion Matrix is a situation analysis table in machine learning that summarizes the prediction results of classification models. It summarizes the records in the dataset in matrix form according to two criteria: the true category and the category predicted by the classification model. The rows of the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202507_11(7).0004

matrix represent the true values, and the columns of the matrix represent the predicted values, measuring the accuracy of a classifier's classification.

Precision refers to the proportion of samples predicted as positive by a model that are actually positive. The mathematical formula is: accuracy=true cases/(true cases+false positives), where true cases represent the number of samples correctly predicted by the model as positive cases, and false positives represent the number of samples incorrectly predicted by the model as positive cases. The higher the accuracy, the lower the misjudgment rate of the model, indicating that the model can more accurately identify positive cases, as shown in formula (1).

$$\mathrm{Pre}\mathit{cision}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{1}$$

Recall refers to the proportion of samples that are actually positive and correctly predicted as positive by the model. The mathematical formula is: recall=true cases/(true cases+false cases), where false cases represent the number of samples that the model incorrectly predicted as negative cases. The higher the recall rate, the more positive examples the model can recall, as shown in formula (2).

$$\mathrm{Re}\mathit{call}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{2}$$

Intersection over union ratio is used to measure the accuracy of bounding boxes in object detection. It refers to the ratio between the intersection and union of the predicted bounding box and the actual bounding box of the target, as shown in formula (3).

$$IOU=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \tag{3}$$

The detection function for pathological myopia recognition is implemented through MobileNetV3 classification. Based on the experimental results of the classification algorithm, a confusion matrix diagram of the pathological myopia recognition MobileNetV3 classification algorithm is drawn as shown in the figure

As shown in Figure 6:

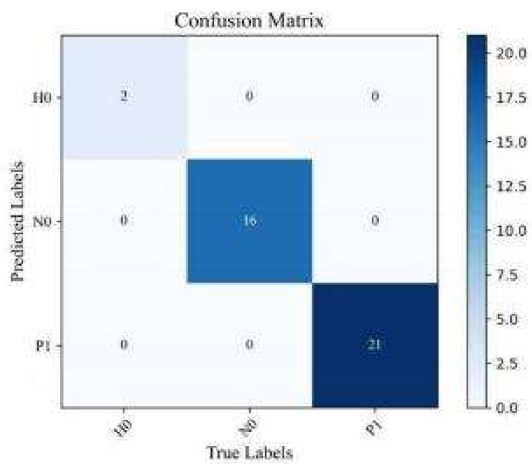

Figure 6. Confusion Matrix Diagram

 Volume 11 Issue 7, 2025
DOI: 10.6919/ICJE.202507_11(7).0004

From the confusion matrix, it can be concluded that all classification results are correctly predicted, and the training of the pathological myopia recognition classification model has achieved very ideal prediction results.

The detection function for identifying pathological myopia lesions is achieved through YOLOv5 detection, and the results of YOLOv5 operation are shown in Figure 7.

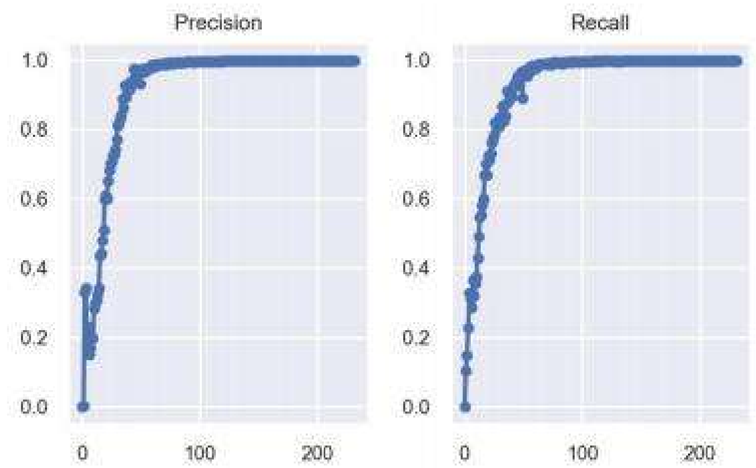

Figure 7. Running Results of YOLOv5

According to the running result images of YOLOv5, after more than 200 rounds of training, the pathological myopia recognition algorithm implemented in this experiment successfully achieved an accuracy and recall rate of over 99% for the recognition of pathological myopia. Accurately detected the pathological myopia site in the image, achieving ideal detection results.

Table 1. Comparison of Operating Efficiency among Different Platforms

Operating Platform	Time
PC	0.0329s
TX2	0.0395s

In order to facilitate the use of pathological myopia recognition, this experiment deployed a pathological myopia recognition system on the portable embedded device NVIDIA Jetson TX2 platform. Deploy the pathological myopia recognition algorithm network model based on MobileNetV3 and YOLOv5 trained on the PC end to the TX2 platform.

3. Summary

In recent years, significant progress has been made in the application of deep learning technology in the medical field. Advanced models such as YOLOv5 and MobileNetV3 have shown great potential. Pathological myopia is a blinding eye disease characterized by abnormal fundus structure, and its diagnosis relies on accurate analysis of fundus images. Traditional diagnostic methods rely on doctors' experience and suffer from low efficiency and high misdiagnosis rates, while deep learning technology can significantly improve diagnostic efficiency and accuracy through automation and intelligence. However, current research still faces challenges such as limited dataset size, high annotation costs, and insufficient model generalization ability.

In the future, the application of deep learning in PM diagnosis and treatment will develop towards a more intelligent and accurate direction. Firstly, through self supervised learning and contrastive learning techniques, the dependence on annotated data is reduced, and a massive amount of unlabeled images are used to pre train the visual foundation model, thereby improving the model's generalization ability. Secondly, develop interpretable AI systems that utilize techniques such as Class Activation Mapping (Grad CAM) to visualize the decision-making process of models and enhance doctors' trust in AI results. For example, previous studies have successfully located key areas identified as "high-risk PM" by the Grad CAM model, providing intuitive evidence for clinical decision-making. In addition, the development of intelligent auxiliary diagnostic systems will also become a focus, integrating AI algorithms with hospital information systems (such as PACS) to achieve full process automation from image upload, lesion detection to report generation. deep learning technology has brought revolutionary changes to the diagnosis and treatment of pathological myopia, but its clinical application still needs to find a balance between technology, ethics, and regulations. Through continuous technological innovation and interdisciplinary collaboration, AI is expected to become an important tool for ophthalmic diagnosis and treatment, providing patients with more efficient and accurate medical services, while promoting the intelligent transformation of the medical field.

References

[1] KermanyDS,GoldbaumM,CaiW,etal.Identifyingmedicaldiagnosesandtreatablediseasesbyimagebaseddeeplearning[J].Cell,2018,172(5):1122-1131.

[2] GargeyaR,LengT.Automatedidentificationofdiabeticretinopathyusingdeeplearning[J].Ophthalmology,2017,124(7):962-969.

[3] OgurtsovaK,GuariguataL,BarengoNC,etal.IDFdiabetesAtlas:Globalestimatesofundiagnoseddiabetesinadultsfor2021.DiabetesResClinPract.2022Jan;183:109-118.

[4] Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. Ophthalmology. 2017 Jul;124(7):962-969. doi: 10.1016/j.ophtha.2017.02.008. Epub 2017 Mar 27. PMID: 28359545.

[5] Elif Setenay Cutur,and Neslihan Gokmen Inan."Multi-class Classification of Retinal Eye Diseases from Ophthalmoscopy Images Using Transfer Learning-Based Vision Transformers."Journal of Imaging Informatics in Medicine prepublish(2025):1-15.

[6] G Divya Deepak,and Subraya Krishna Bhat."Deep learning-based CNN for multiclassification of ocular diseases using transfer learning."Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 12.1(2024):

[7] A. Anitha Rani,C. Karthikeyini,and C. Reshma Ravi."Eye Disease Prediction Using Deep Learning and Attention on Oct Scans."SN Computer Science 5.8(2024):1065-1065.

[8] Shayla Sharmin, et al."A dataset of color fundus images for the detection and classification of eye diseases."Data in Brief 57.(2024):110979-110979.

[9] Anjar Wanto, Yuhandri Yuhandri,andOkfalisa Okfalisa."RetMobileNet: A New Deep Learning Approach for Multi-Class Eye Disease Identification."Revue d'Intelligence Artificielle 38.4(2024):

[10] Uyar K ,Yurdakul M ,Taşdemir Ş .Abc-based weighted voting deep ensemble learning model for multiple eye disease detection[J].Biomedical Signal Processing and Control,2024,96(PB):106617-106617.