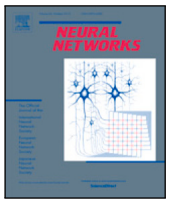
Neural Networks 182 (2025) 106895

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet



Full length article

MIU-Net: Advanced multi-scale feature extraction and imbalance mitigation for optic disc segmentation

Yichen Xiao [a,b,c,d,1], Yi Shao [a,b,c,d,1], Zhi Chen [a,b,c,d,1], Ruyi Zhang [a], Xuan Ding [a,b,c,d], Jing Zhao [a,b,c,d], Shengtao Liu [a,b,c,d], Teruko Fukuyama [a,b,c,d], Yu Zhao [a,b,c,d], Xiaoliao Peng [a,b,c,d], Guangyang Tian [e], Shiping Wen [e,*], Xingtao Zhou [a,b,c,d,*]

[a] Eye Institute and Department of Ophthalmology, Eye and ENT Hospital, Fudan University, Shanghai, 200031, China
[b] NHC Key Laboratory of Myopia (Fudan University), Key Laboratory of Myopia, Chinese Academy of Medical Sciences, Shanghai, 200031, China
[c] Shanghai Research Center of Ophthalmology and Optometry, Shanghai, 200031, China
[d] Shanghai Key Laboratory of Visual Impairment and Restoration, Shanghai, 200031, China
[e] Australian AI Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia

ARTICLE INFO

Keywords:
Optic disc segmentation
MIU-net
MFE module
Dual attention module
Focal loss

ABSTRACT

Pathological myopia is a severe eye condition that can cause serious complications like retinal detachment and macular degeneration, posing a threat to vision. Optic disc segmentation helps measure changes in the optic disc and observe the surrounding retina, aiding early detection of pathological myopia. However, these changes make segmentation difficult, resulting in accuracy levels that are not suitable for clinical use. To address this, we propose a new model called MIU-Net, which improves segmentation performance through several innovations. First, we introduce a multi-scale feature extraction (MFE) module to capture features at different scales, helping the model better identify optic disc boundaries in complex images. Second, we design a dual attention module that combines channel and spatial attention to focus on important features and improve feature use. To tackle the imbalance between optic disc and background pixels, we use focal loss to enhance the model's ability to detect minority optic disc pixels. We also apply data augmentation techniques to increase data diversity and address the lack of training data. Our model was tested on the iChallenge-PM and iChallenge-AMD datasets, showing clear improvements in accuracy and robustness compared to existing methods. The experimental results demonstrate the effectiveness and potential of our model in diagnosing pathological myopia and other medical image processing tasks.

1. Introduction

Myopia, is a prevalent condition characterized by the elongation of the eyeball, resulting in light rays focusing in front of the retina rather than directly on its surface. This leads to blurred vision when looking at distant objects, necessitating the use of corrective lenses or surgery (Patil, Shetty, Kale, Patil, & Sharma, 2024). Although often considered a minor inconvenience, myopia can escalate into pathological myopia, a severe form of the condition associated with significant ocular complications such as retinal detachment, macular degeneration, and glaucoma. These potential complications pose a substantial threat to vision and highlight the necessity for early and accurate diagnosis (Li, Liu et al., 2023; Zhang et al., 2023).

Pathological myopia presents distinct alterations in the ocular fundus, observable through fundus photography. These alterations encompass peripapillary atrophy, posterior staphyloma, and diffuse and patchy atrophy, posing significant challenges to the identification and segmentation of the optic disc (Park, Ko, Park, Kim, & Choi, 2022; Wan et al., 2024). The optic disc, serving as the exit point for the optic nerve, precise segmentation of the optic disc plays an important role in the diagnosis of ophthalmic diseases, as it allows for accurate assessment of optic disc morphological changes, enhances retinal image analysis, facilitates disease progression monitoring, and aids in diagnosis and treatment decisions, ultimately improving diagnostic accuracy and patient management (Han, Liu, Chen, & He, 2022; Lupon, Nolla, &

* Corresponding authors.
E-mail addresses: yichenxiao@fudan.edu.cn (Y. Xiao), hsuanyilee@email.ncu.edu.cn (Y. Shao), Zhi.Chen@fdeent.org (Z. Chen), 21301050117@m.fudan.edu.cn (R. Zhang), 17211260004@fudan.edu.cn (X. Ding), zhaojing_med@fudan.edu.cn (J. Zhao), 505560283@qq.com (S. Liu), Teru5255@hotmail.com (T. Fukuyama), yu.zhao@fdeent.org (Y. Zhao), 22111260012@m.fudan.edu.cn (X. Peng), guangyang.tian@student.uts.edu.au (G. Tian), shiping.wen@uts.edu.au (S. Wen), xingtaozhou@fudan.edu.cn (X. Zhou).
[1] The three authors contribute equally to this work.

https://doi.org/10.1016/j.neunet.2024.106895
Received 9 September 2024; Received in revised form 7 October 2024; Accepted 5 November 2024
Available online 14 November 2024

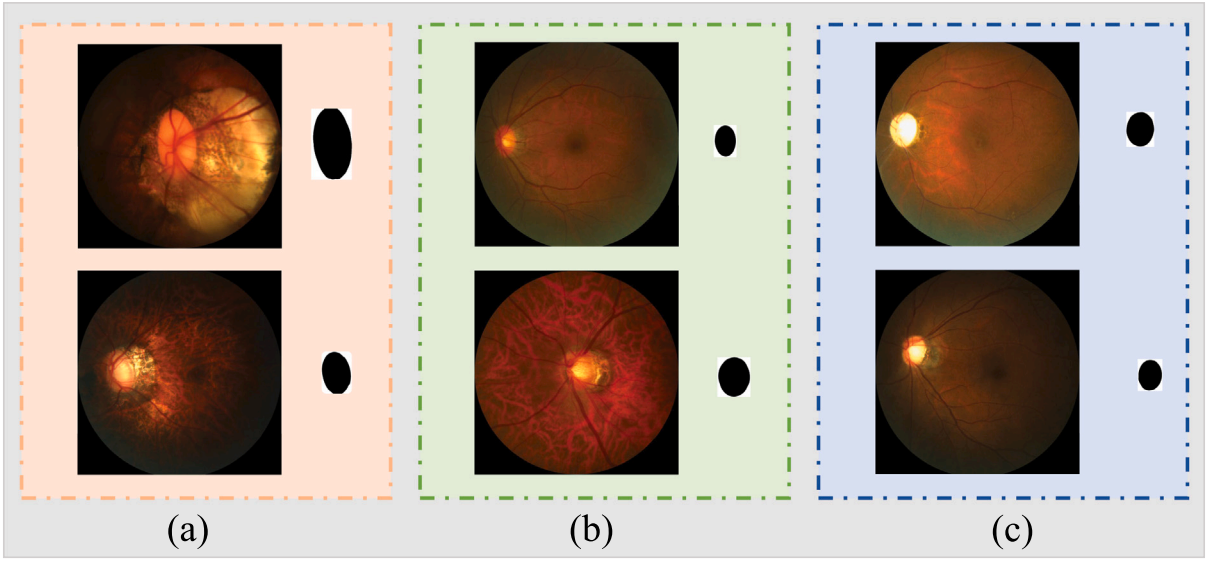


Fig. 1. (a), (b), and (c) show the fundus images and corresponding masks of pathological myopia, high myopia, and normal eyes, respectively.

Cardona, 2024). Fig. 1 shows the fundus images and their corresponding masks under different health conditions. From the figure, it can be seen that the optic disc in pathological myopia has already undergone pathological changes, which are significantly different from those in normal eyes.

Fundus photography, a cornerstone of ocular diagnostics, provides a detailed view of the retina and its structures through a non-invasive imaging technique. This method is instrumental in detecting and monitoring diseases that affect the back of the eye, making it a significant tool in routine eye examinations and supporting the early detection and management of serious eye conditions (Li, Foo et al., 2023; Niu et al., 2024). Advancements in automated image processing technologies have revolutionized medical imaging by providing consistent, fast, and scalable analysis compared to manual segmentation. Automated methods such as those based on the U-Net architecture eliminate variability due to individual expertise and fatigue, handle large volumes of data efficiently, and make diagnostic services more accessible, especially in underserved areas.

This paper introduces the enhanced MIU-Net model, which is tailored for precise optic disc segmentation in the context of pathological myopia. Our model integrates the multi-scale feature extraction (MFE) module, the dual attention module, the focal loss function, and data augmentation techniques, which significantly improve the performance of myopic optic disc segmentation compared to other methods. The contributions of this paper are manifold:

1. The integration of the MFE module allows for multi-scale feature extraction, enabling the model to capture a broader range of contextual information and finer details, which significantly improves segmentation performance across different pathological scales.
2. The dual attention module enhances segmentation clarity and precision by prioritizing important features in both the channel and spatial dimensions, thereby improving overall accuracy.
3. The use of the focal loss function effectively addresses the class imbalance issue commonly found in medical image segmentation, particularly when background pixels greatly outnumber optic disc pixels, resulting in improved accuracy. Additionally, data augmentation techniques are applied to mitigate the problem of limited data and further enhance model performance.
4. Experimental evaluations on the iChallenge-PM and iChallenge-AMD datasets demonstrate substantial improvements, highlighting the practical effectiveness and potential of the proposed model in medical image processing tasks.

2. Related work

2.1. Traditional segmentation methods

Optic disc segmentation traditionally relies on image processing techniques such as edge detection, region growing, thresholding, and mathematical morphological operations. These methods generally require explicit prior knowledge and depend on high contrast and clear boundaries within the images. For instance, Hoover, Kouznetsova, and Goldbaum (2000) developed a method that utilizes edge detection to identify candidate areas for the optic disc, followed by morphological operations to fine-tune the disc boundaries. Although effective, this method relies heavily on the image's high contrast and clear boundaries. Simultaneously, Foracchia, Grisan, and Ruggeri (2005) introduced a technique based on mathematical morphology and grayscale transformation, which shows improved robustness in handling fundus images with uneven illumination. Marín, Aquino, Gegúndez-Arias, and Bravo (2010) further developed a method based on elliptical fitting and morphology, enhancing the accuracy of optic disc localization across various backgrounds. These methods share common traits of requiring precise parameter adjustments and a high dependency on image quality.

As technology advanced, researchers began exploring more automated methods for optic disc segmentation. Mendonca and Campilho (2006) proposed a method based on dynamic thresholding and edge tracking that adapts to changes in image illumination and accurately segments the optic disc area. While these methods can achieve good segmentation results under specific conditions, they typically require precise parameter adjustments and high-quality images, limiting their application in complex or low-quality image scenarios.

2.2. Machine learning segmentation methods

As machine learning technology has evolved, methods for optic disc segmentation have transitioned from traditional image processing techniques to more advanced algorithms (Rauf, Gilani, & Waris, 2021; Tong et al., 2023). Compared to conventional methods, machine

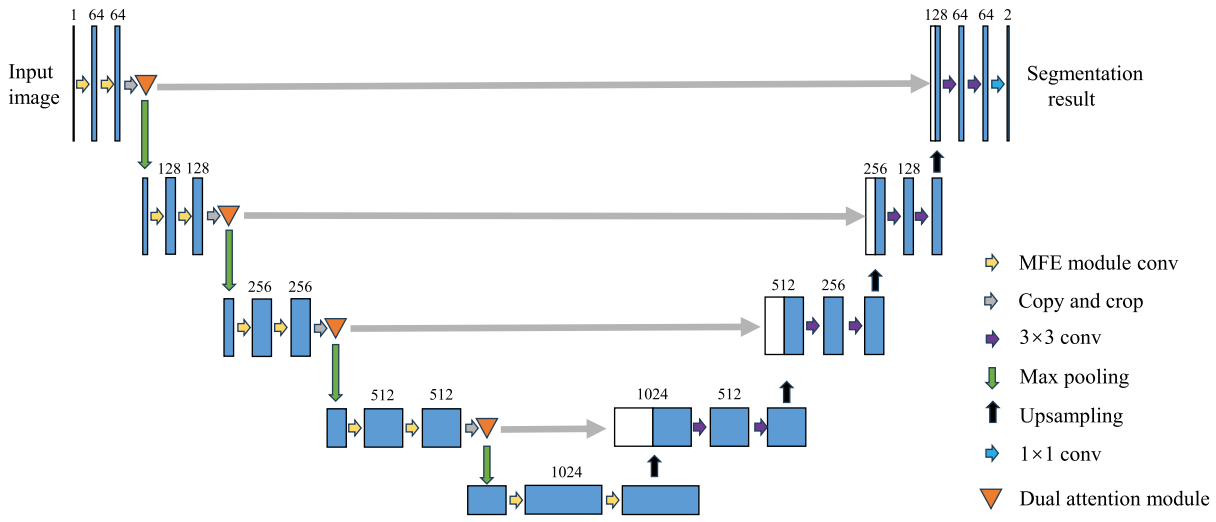


Fig. 2. Schematic diagram of MIU-Net, the meanings of the symbols in the diagram are explained in detail on the lower right.

learning approaches can automatically recognize and extract complex image features by learning from large datasets, significantly reducing reliance on manual feature engineering and image quality, thereby enhancing segmentation accuracy. Li and Chutatape (2004) developed a method based on image techniques combining Graph Cut technology and Support Vector Machines (SVM) for the automatic recognition and segmentation of the optic disc area. Cheng et al. (2013) proposed a method based on superpixel classification that uses the k-means clustering algorithm to optimize the initial stages of image segmentation and integrates gradient-boosted random forest classifiers to enhance the accuracy of optic disc detection. This method has been validated across multiple public fundus image databases, demonstrating superior performance compared to traditional approaches. Additionally, Zhang et al. (2016) utilized the AdaBoost algorithm to enhance the feature selection process, improving the accuracy and speed of segmentation. Furthermore, Morales (Morales, Naranjo, Angulo, & Alcañiz, 2013) adopted a combination of decision trees and K-Nearest Neighbours (K-NN) techniques, which effectively handle the segmentation of the optic disc and retinal vessels without additional computational burden.

However, despite many advantages over traditional image processing techniques, machine learning methods exhibit certain limitations when dealing with extremely complex or large-scale datasets compared to deep learning approaches. Machine learning typically requires predefined feature extraction and complex feature engineering, which can be circumvented by the automatic feature learning inherent in deep learning methods. Moreover, deep learning approaches often achieve more complex data representations through their deep network structures, thus performing better across a broader range of applications.

2.3. Deep learning segmentation methods

The advent of deep learning has significantly transformed the field of optic disc segmentation, introducing methodologies that substantially surpass previous machine learning and traditional image processing techniques. Deep learning methods, particularly those utilizing convolutional neural networks (CNNs), have demonstrated exceptional capabilities in learning hierarchical representations directly from raw image data, thus eliminating the need for manual feature selection and engineering (Choi et al., 2021; Yang et al., 2020).

One of the seminal architectures in this domain is U-Net, initially introduced by Ronneberger, Fischer, and Brox (2015). This architecture features a symmetric encoder–decoder structure that excels in medical image segmentation tasks by efficiently capturing both contextual information and fine details through skip connections. The U-Net has been extensively adapted and improved for the specific challenges of optic disc segmentation, demonstrating robust performance across diverse datasets. Some researchers then integrated the attention mechanism into the U-Net architecture. For instance, the Attention U-Net, proposed by Oktay et al. (2018), modifies the U-Net architecture to include an attention gate that suppresses irrelevant regions in an input image while highlighting salient features useful for a specific task, thereby enhancing the precision of segmentation. Following U-Net, advancements such as the DeepLab model introduced by Chen, Papandreou, Schroff, and Adam (2017), which incorporates atrous convolutions and spatial pyramid pooling, have further refined segmentation capabilities. This approach allows for more flexible receptive field sizes and better incorporation of context, which is crucial for the accurate delineation of the optic disc.

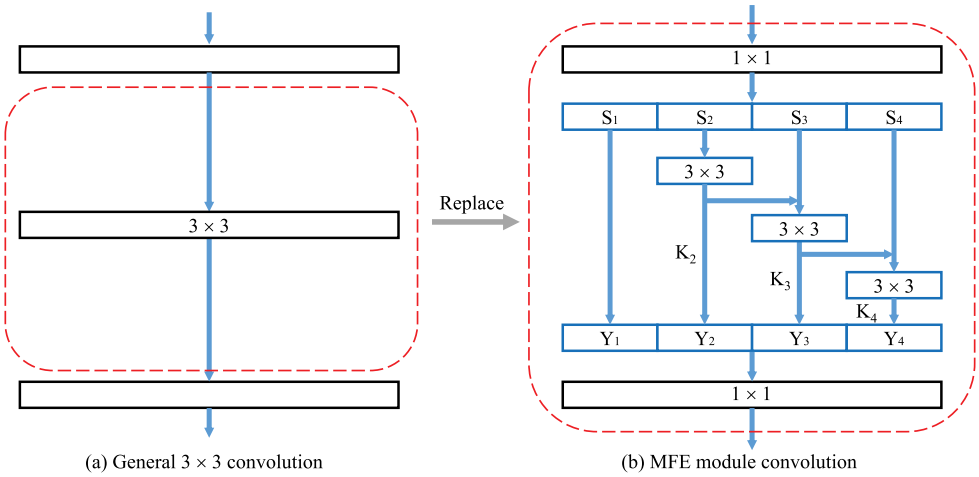
Furthermore, incorporating deep learning models tailored for specific imaging conditions, Fu et al. (2018) developed a deep learning system that uses a path-based CNN architecture to address the challenges posed by variations in imaging conditions, such as differences in illumination and ocular pathologies. This approach significantly enhances the adaptability and accuracy of optic disc segmentation under varied clinical environments. Additionally, Transformer has achieved success across various fields, and several studies have applied them to medical image segmentation task. For example, Mehmood, Alsharari, Iqbal, Spence, and Fahim (2024) proposed RetinaLiteNet, a lightweight deep learning model for optic disc segmentation. The model incorporates a multitask learning approach using an encoder–decoder structure. In the encoder, convolutional layers combined with multi-head attention capture both fine-grained local features and long-range dependencies. The decoder further enhances feature extraction by integrating skip connections and a convolutional block attention module (CBAM). Tested on DRIVE and IOSTAR datasets, RetinaLiteNet demonstrates competitive performance with minimal computational requirements, making it ideal for use in resource-limited environments.

These deep learning strategies collectively mark a significant advancement in the field of optic disc segmentation, offering superior performance in terms of accuracy, efficiency, and automation.

3. Proposed method

3.1. Overall structure of MIU-Net

The MIU-Net architecture represents a sophisticated synthesis of depth and precision, structured to address the intricate challenges of optic disc segmentation in fundus images. At its core, the MIU-Net harnesses the power of the U-Net architecture, renowned for its



(a) General 3 × 3 convolution (b) MFE module convolution

Fig. 3. Diagram of general convolution and MFE module convolution ($d = 4$).

proficiency in medical image segmentation. However, to specifically tackle the nuanced features of the optic disc, MIU-Net goes several steps further by incorporating two pivotal enhancements: the MFE modules and dual attention module. Fig. 2 shows the structure of MIU-Net.

MFE modules are ingeniously embedded within the encoder of MIU-Net. Each stage of the encoder is equipped with a MFE module that acts as a multi-scale processing unit. These modules dissect the incoming feature maps into subsets, process each subset at a different scale, and then judiciously reintegrate them. This process ensures that the features extracted encapsulate a rich, multi-scale contextual understanding, pivotal for delineating the diverse morphological variations of the optic disc across patients with pathological myopia.

Complementing the depth of the MFE modules, the dual attention module is meticulously integrated at each level of the encoder. The dual attention module, comprising both channel and spatial attention, refines the feature maps in a targeted fashion. It selectively amplifies salient features that are crucial for identifying the optic disc while simultaneously diminishing the influence of less pertinent areas. This attentional focus is especially beneficial in enhancing the contrast and definition of the optic disc against the complex backdrop of the retinal fundus, ensuring that the encoder outputs feature maps with heightened relevance.

3.2. MFE module

We show the difference between general convolution and MFE module convolution in Fig. 3. As shown in the red box (b) in Fig. 3, the MFE module employs a novel strategy for constructing hierarchical connections, a departure from the traditional single-scale processing found in general convolution layers, as depicted in red box (a) (Gao et al., 2019).

The MFE module functions by partitioning the input feature map into distinct subsets $S_1, S_2, S_3, \ldots, S_i$, where $i \in \{1, 2, \ldots, d\}$, each representing a dimension. Before this partitioning, an initial 1×1 convolution is applied to the input feature map. This first 1×1 convolution is responsible for reducing the number of channels in the input, lowering computational complexity while ensuring that important features are retained for the multi-scale processing in subsequent steps. These subsets are then intricately processed in sequence, except for S_1, each subset S_i undergoes a 3×3 convolution, denoted $\mathbf{K}_i()$, with its output Y_i being element-wise added to the next subset before it is passed through another convolution. This elegant cascading operation ensures that each convolved output contains aggregated contextual information from all previous scales in the sequence. We can use a formula to express this calculation process. In this paper, the parameter d which value is larger with the richer multi-scale features, is set to control the dimension of multi-scale features. We will discuss this in the subsequent experiments.

$$Y_i = \begin{cases} S_i & \text{if } i = 1; \\ \mathbf{K}_i(S_i) & \text{if } i = 2; \\ \mathbf{K}_i(S_i + Y_{i-1}) & \text{if } 2 < i \leq d. \end{cases} \tag{1}$$

Once the outputs $Y_1, Y_2, Y_3, \ldots, Y_i$ are generated from the convolutions, they are progressively combined to form a deep multi-scale feature map. At this point, a second 1×1 convolution is applied to fuse these outputs into a unified feature representation. This final 1×1 convolution helps merge the multi-scale features into a single output feature map, ensuring that the rich contextual information gathered from different scales is effectively integrated and that the output matches the required number of channels for further processing. This structure allows for a comprehensive feature representation, capturing an extensive range of contextual details within a singular architectural layer. Such a multi-faceted feature integration is especially beneficial for tasks demanding high levels of detail recognition, such as the segmentation of the optic disc, where variability in size, shape, and boundary definition requires an algorithm capable of discerning patterns across a spectrum of scales.

Within the MIU-Net model, the MFE module serves as a foundational element that enriches the network's feature learning capabilities. By producing outputs Y_i that each carry the integrated knowledge of all preceding convolutions, the module provides a robust foundation for precise segmentation outcomes. The MFE module differentiates itself from the general 3×3 convolution by offering the network a methodology to incrementally elaborate on the feature context, ultimately leading to a segmentation model that is both powerful and adept at accommodating the diverse challenges presented by retinal images affected by pathological myopia.

3.3. Dual attention module

In the MIU-Net architecture, the dual attention module plays a critical role in refining the feature representations produced by the MFE modules. Dual attention module, depicted in Fig. 4, systematically focuses the network's processing capacity on relevant patterns within the fundus images by emphasizing salient features and suppressing less informative ones. This module operates in two sequential attention stages: channel and spatial.

The channel attention mechanism first employs global average pooling and global max pooling to compress spatial information, resulting in two distinct feature descriptors. These descriptors capture the global statistics of the feature maps from two different perspectives, ensuring

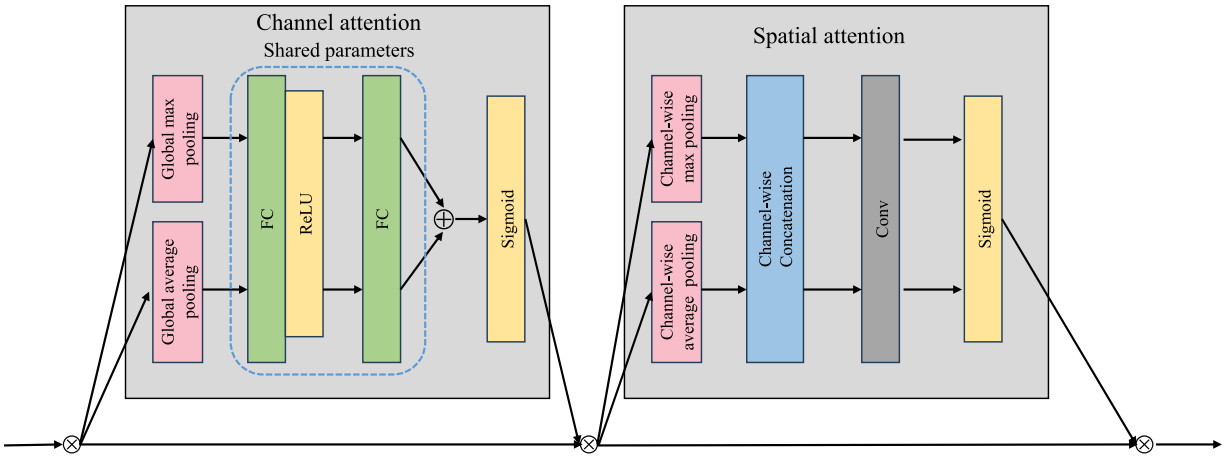


Fig. 4. Diagram of the dual attention module.

a comprehensive understanding of feature relevance. They are then fed through shared multi-layer perceptrons (MLPs) with one hidden layer, where ReLU activation is used after the first fully connected (FC) layer, fostering non-linear interactions between channel-wise features. The two processed signals are merged by element-wise addition and passed through a sigmoid activation function to obtain the final channel attention map. This map selectively enhances certain channels while diminishing others, according to the content of the feature map, effectively enabling the network to focus on more meaningful features for the optic disc segmentation task.

Subsequently, the spatial attention mechanism consolidates the refined feature map by applying channel-wise max pooling and average pooling, resulting in two distinct spatial context descriptors. These descriptors are concatenated and convolved with a standard convolutional layer, which captures the cross-channel interaction, highlighting where the network should focus spatially. The output of this convolution is then passed through a sigmoid function to generate the spatial attention map. This map emphasizes specific spatial regions of the feature map, further enhancing the details that are pertinent for the segmentation of the optic disc while suppressing the less relevant areas.

The incorporation of the dual attention module into MIU-Net follows the MFE modules within the encoder. It ensures that as the multi-scale features are hierarchically integrated within the MFE blocks, the subsequent feature refinement is attentively guided, honing the focus of the network on the optic disc. The ability of the dual attention module to highlight discriminative features effectively addresses the challenge of optic disc segmentation in pathological myopia, where variable disc appearances and similar looking structures may complicate the segmentation process. Thus, the dual attention module significantly contributes to the MIU-Net's performance by delivering a more targeted and context-aware feature processing, which is crucial for achieving high-precision segmentation in medical imaging tasks.

3.4. Focal loss

In training the MIU-Net model for optic disc segmentation, the selection of an appropriate loss function is paramount, particularly in addressing the significant class imbalance issue inherent in the task. The cross-entropy loss is commonly used in deep learning area. It measures the performance of a classification model whose output is a probability value between 0 and 1. The cross-entropy loss can be defined as:

$$L_{ce}(p_t) = -\log(p_t) \tag{2}$$

where p_t denotes the predicted probability for the ground truth class. For a binary classification scenario like optic disc segmentation, p_t equals the model's estimated probability for the optic disc class if the ground truth is 'optic disc'.

However, the cross-entropy loss does not address class imbalance, leading to a model biased towards the majority class. To this end, focal loss is employed due to its effectiveness in modulating the contribution of each sample to the loss based on the classification difficulty, thereby providing a solution to the imbalance challenge. Focal loss modifies the standard cross-entropy loss so that it reduces the relative contribution to the loss of easy examples Lin, Goyal, Girshick, He, and Dollár (2017). This allows the model to focus training more on hard, or easily misclassified examples. The focal loss function can be expressed as:

$$L_{fl}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{3}$$

where α_t is a balancing factor that adjusts the importance given to the background class. γ is the focusing parameter that scales the contribution of each sample to the loss based on how well it is classified. When γ is higher, the suppression of loss for well-classified examples is more pronounced. In this paper, γ is set to 2, which has been found effective in many segmentation tasks to adequately focus on hard-to-classify examples, we follow this convention. We will discuss the impact of α_t on the segmentation performance of the model in subsequent experiments and determine the optimal value.

For myopic optic disc segmentation, focal loss specifically alleviates the domination of the majority class (background) by diminishing the loss contribution from the numerous easy-to-classify background pixels. This redirection of focus enables the MIU-Net to concentrate on the more complex patterns and nuances of the optic disc, which are crucial for achieving an accurate segmentation. The model is trained to become sensitive to the variations and subtleties present in the optic disc, thereby boosting its discriminative power between the disc and the surrounding tissue.

4. Experiments

4.1. Dataset description

In order to verify the validity of the proposed method, two public datasets are used in this paper: iChallenge-PM (Fang et al., 2023) and iChallenge-AMD (Fang et al., 2022). iChallenge-PM comprises 1200 high-definition fundus photographs, which include a diverse representation of myopic pathologies. Each image is meticulously annotated to delineate the optic disc and optic cup boundaries, providing a reliable ground truth for segmentation model. Similar to the iChallenge-PM, the iChallenge-AMD dataset also includes 1200 annotated fundus images from patients at various stages of Age-related Macular Degeneration (AMD), expert annotations are provided to mark the regions affected by AMD-related changes, crucial for developing segmentation algorithms.

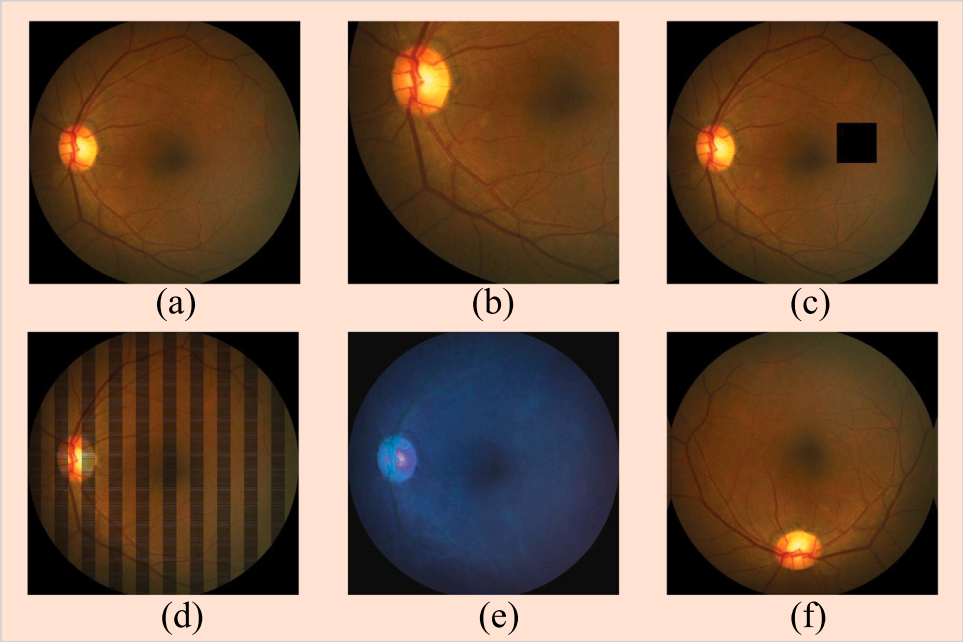


Fig. 5. The effect of different data augmentation methods applied to the original image. (a) is the original image, (b), (c), (d), (e), and (f) are cropped image, randomly erased image, grid-dropout image, colour-Jittered image, and rotated image respectively.

4.2. Implementation details

For both datasets, adjusting the images to a uniform resolution 512 × 512 to maintain consistency in input size for the MIU-Net. Moreover, brightness and contrast normalization are applied across the dataset to counteract variations in image acquisition conditions. The data is divided into 80% training and 20% test sets.

In medical image analysis, especially in the field of ophthalmology, annotated images of the optic disc are often very limited. Collecting and annotating these images require considerable time and effort from experts, resulting in typically small datasets. However, training deep learning models usually requires large amounts of data to avoid overfitting and improve generalization. To mitigate the issue of insufficient training samples, this paper adopts various data augmentation techniques to increase the diversity of the training set, helping the model learn more features and avoid overfitting. Additionally, data augmentation can simulate diverse variations in optic disc morphology, lighting, and noise, enhancing the model's generalization ability to effectively segment the optic disc in images from different patients and under different imaging conditions. Examples of data augmentation images are shown in Fig. 5.

The training and test of MIU-Net occur on a workstation equipped with an NVIDIA GeForce RTX 3080 GPU, which provides the computational power necessary for deep learning tasks. The system also includes an Intel Core i9 processor and 32 GB of RAM to facilitate efficient data handling and parallel processing tasks. Experiments are conducted using Python 3.10 and Pytorch 2.1 as the primary frameworks for deep learning model development. These tools are selected for their robust support for CNNs and ease of use in deploying complex models. The model is trained using the Adam optimizer with an initial learning rate of 0.001. The learning rate is scheduled to decay by a factor of 0.1 every 20 epochs to allow for more refined updates as the training progresses. The total number of training epochs is set to 100, providing sufficient time for convergence while avoiding overfitting. We use a batch size of 16, which is selected to balance between optimizing GPU memory usage and maintaining training speed. In terms of early stopping criteria, we monitor the loss during training and apply early stopping if the loss does not improve for 10 consecutive epochs, which helps prevent overfitting.

4.3. Baseline models and evaluation metrics

To validate the effectiveness of the MIU-Net in optic disc segmentation, it is essential to compare its performance against a selection of well-established baseline models. Here are five suitable baseline models: U-Net (Ronneberger et al., 2015), SegNet (Badrinarayanan, Kendall, & Cipolla, 2017), DeepLabv3+ (Chen, Zhu, Papandreou, Schroff, & Adam, 2018), Attention U-Net (Oktay et al., 2018) and TransUNet (Chen et al., 2021). For evaluating the performance of MIU-Net in optic disc segmentation, it is essential to select metrics that accurately reflect the efficacy of the model across various aspects of segmentation quality. We select three appropriate metrics that can be used to evaluate the performance: Accuracy (Acc), Intersection over Union (IoU), and the $F1-score$.

Acc measures the proportion of true results among the total number of cases examined. It is a straightforward metric that indicates the overall correctness of the segmentation results. Also known as the Jaccard Index, IoU is a common metric for the evaluation of object detection and segmentation models. It measures the overlap between the predicted segmentation and the ground truth. The F1-score is the harmonic mean of precision and recall, offering a balance between these two metrics. It is particularly useful when the classes are imbalanced, as it does not bias the model performance towards the majority class. These metrics can be defined as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{5}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \tag{6}$$

where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives.

4.4. Results and analysis

The segmentation results of different datasets are shown in Tables 1 and 2. The proposed MIU-Net achieves superior performance in optic disc segmentation on the iChallenge-PM dataset, as evidenced by

Table 1
Comparison of optic disc segmentation results on iChallenge-PM dataset by different methods.

Dataset	Method	Acc	IoU	F1-score
iChallenge-PM	U-Net	94.21	85.86	91.79
	SegNet	93.16	84.36	90.46
	DeepLabv3+	95.69	88.74	93.78
	Attention U-Net	95.75	88.43	94.48
	TransUNet	95.42	87.96	93.24
	MIU-Net (Proposed)	**96.44**	**89.23**	**94.93**

Table 2
Comparison of optic disc segmentation results on iChallenge-AMD dataset by different methods.

Dataset	Method	Acc	IoU	F1-score
iChallenge-AMD	U-Net	94.36	84.31	91.32
	SegNet	93.18	82.43	90.93
	DeepLabv3+	94.43	84.27	91.47
	Attention U-Net	94.97	84.69	91.65
	TransUNet	93.88	83.65	91.07
	MIU-Net (Proposed)	**95.44**	**85.43**	**91.83**

the highest scores in Accuracy (96.44%), IoU (89.23%), and F1-score (94.93%). This success can be attributed to several key innovations. The integration of the MFE module allows for enhanced multi-scale feature extraction, capturing both global context and local details crucial for accurate segmentation. Additionally, the dual attention module enhances the model's focus on important features by applying attention mechanisms at both channel and spatial levels. This targeted focus helps in distinguishing the optic disc from the background more precisely. Furthermore, the use of focal loss effectively addresses class imbalance, ensuring that the minority class receives adequate attention during training, leading to improved segmentation performance. Data augmentation improves the model's robustness and generalization performance.

Compared to other models, MIU-Net shows significant improvements over U-Net, SegNet, DeepLabv3+, Attention U-Net, and TransUNet. For example, while DeepLabv3+ achieves an IoU of 88.74% and an F1-score of 93.78%, MIU-Net surpasses these results with an IoU of 89.23% and an F1-score of 94.93%, highlighting its superior performance in accurately segmenting the optic disc. MIU-Net builds upon U-Net's foundational architecture and incorporates advanced modules, allowing it to overcome the limitations of traditional models. These enhancements make MIU-Net a more powerful and precise tool for medical image segmentation tasks.

On the iChallenge-AMD dataset, MIU-Net continues to demonstrate outstanding performance, achieving an accuracy of 95.44%, which is higher than Attention U-Net and DeepLabv3+, with accuracies of 94.97% and 94.43%, respectively. MIU-Net also achieves the highest IoU of 85.43%, reflecting its ability to capture optic disc structures accurately in the presence of age-related macular degeneration. Furthermore, with an F1-score of 91.83%, MIU-Net shows a balanced performance, effectively minimizing both false positives and false negatives. These results confirm MIU-Net's robustness and precision for optic disc segmentation, particularly in challenging cases like AMD-affected fundus images.

Across both datasets, MIU-Net consistently presents top-tier performance, particularly in the IoU metric, which is a stringent measure of segmentation accuracy. The high IoU scores corroborate the architectural enhancements of MIU-Net, specifically the integration of MFE module and dual attention module, which contribute to its precise delineation of the optic disc. The accuracy and F1-score metrics across both datasets indicate that MIU-Net excels in overall classification correctness and in balancing the trade-off between precision and recall, suggesting its potential for effective deployment in clinical settings for aiding in the diagnosis and treatment planning of ocular conditions. In summary, the consistent results across different datasets highlight the model's adaptability and generalization strength, affirming its value in advancing medical imaging diagnostics.

Table 3
The segmentation performance of MIU-Net after removing various modules.

Dataset	Method	Acc	IoU	F1-score
iChallenge-PM	MIU-Net with cross-entropy loss	94.69	88.21	92.32
	MIU-Net without MFE module	95.48	88.35	93.94
	MIU-Net without dual attention module	95.24	87.76	93.62
	MIU-Net without data augmentation	94.83	88.46	92.72
	MIU-Net (Proposed)	**96.44**	**89.23**	**94.93**

4.5. Parameter research

To investigate the impact of different parameters on segmentation performance, we conduct a comparative study using the iChallenge-AMD dataset. The results, depicted in Fig. 6(a), show the effect of varying the MFE module dimension d. As d increases from 2 to 4, both IoU and F1-score improve, reaching their peak values of 85.43% and 91.83%, respectively, at $d = 4$. This improvement is due to the enhanced multi-scale feature extraction capability provided by higher d values, which allows the model to capture more detailed and diverse features of the optic disc. However, further increasing d to 5 results in a slight decrease in performance. This indicates that an optimal dimension of d is 4, which provides the best balance between detailed multi-scale feature extraction and model complexity. Beyond this point, the added complexity does not yield significant performance gains and may introduce noise, thereby slightly degrading the model's accuracy.

Similarly, Fig. 6(b) illustrates the influence of the balancing factor α_t on segmentation results. As α_t increases from 0.1 to 0.5, there is a noticeable improvement in both IoU and F1-score, with the highest values of 85.43% and 91.83% achieved at $\alpha_t = 0.5$. This improvement is attributed to the better balance between the positive and negative class contributions, which helps the model to focus equally on both the optic disc and the background. Increasing α_t beyond 0.5 to 0.75 results in a slight decline in performance. This suggests that while moderate balancing improves model performance by ensuring that minority class examples (optic disc) receive adequate attention, overly high α_t values may lead to overemphasis on the minority class, reducing overall performance due to insufficient learning from the majority class (background). Thus, an optimal α_t value is 0.5, providing the best balance and enhancing segmentation accuracy and robustness.

4.6. Ablation study

In Table 3, the ablation study presented provides insights into the contribution of specific components within the MIU-Net model for optic disc segmentation, conducted on the iChallenge-PM dataset. Each row represents a variant of the MIU-Net with a key component removed or replaced, and the impact on performance is measured using accuracy, IoU, and F1-score.

Replacing the focal loss with the standard cross-entropy loss results in lower performance across all metrics compared to the proposed MIU-Net. This change leads to a slight decrease in accuracy, IoU, and F1-score, which implies that focal loss's ability to handle class imbalance is significant for the task of optic disc segmentation. Omitting the MFE module causes a noticeable drop in IoU to 88.35% from the proposed model's 89.23%, while maintaining a relatively high accuracy and F1-score. This indicates that while the overall classification accuracy remains high, the precision of the segmentation boundary, as quantified by IoU, relies heavily on the multi-scale feature extraction capabilities of the MFE module. Removing the dual attention module results in a lower IoU of 87.76% and a slight reduction in the F1-score to 93.62%, compared to the proposed MIU-Net. This suggests that

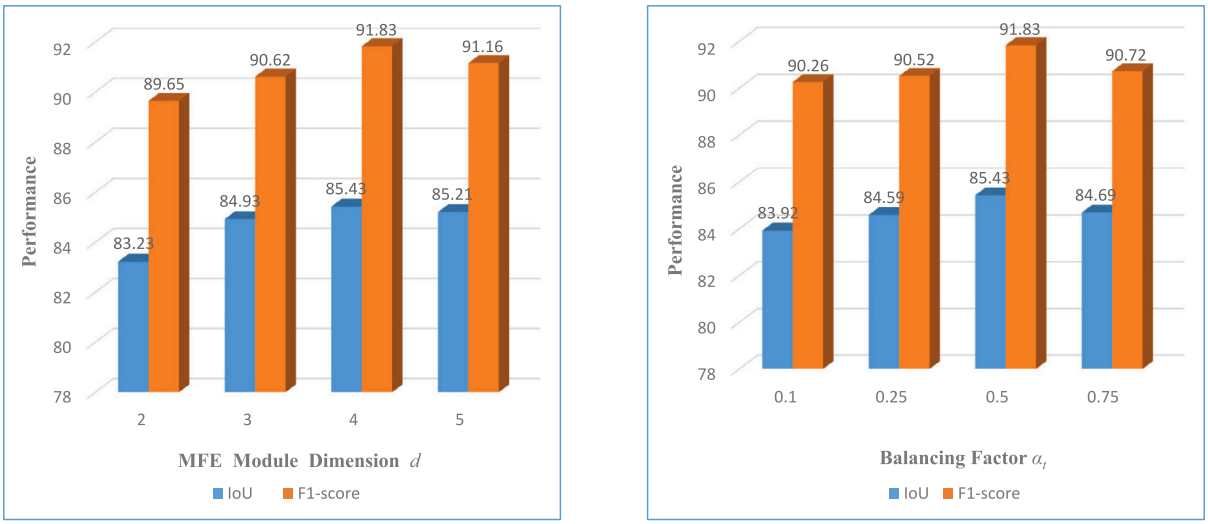


Fig. 6. (a) and (b) show the impact of MFE module dimension d and balancing factor α_t, respectively.

the dual attention module's role in focusing the network's attention on salient features significantly enhances the segmentation's precision. From the results in Table 3, we can clearly observe that the model without data augmentation performs poorly across all evaluation metrics, which demonstrates the significant role of data augmentation in improving the performance of optic disc segmentation model, especially in cases with insufficient data.

Based on these findings, we can conclude that each component of MIU-Net plays a vital role in its segmentation performance, and removing any of them leads to noticeable performance degradation. Specifically, the MFE module is essential for precise boundary detection, as evidenced by the drop in IoU when it is excluded. The dual attention module also proves to be important for improving segmentation accuracy, as indicated by the reduction in both IoU and F1-score. The focal loss function effectively handles the class imbalance issue inherent in optic disc segmentation, with cross-entropy loss leading to lower scores across all metrics when used as a replacement. Data augmentation addresses the issue of limited training data, making the model more resilient to variations in optic disc shape. It also helps the model adjust to different lighting conditions and noise, significantly enhancing its generalization capability and segmentation accuracy.

5. Conclusion

This study introduced MIU-Net, a novel deep learning architecture for the segmentation of the optic disc in retinal fundus images, particularly those affected by pathological myopia and age-related macular degeneration. The MIU-Net model incorporates advanced components such as the MFE module, the dual attention module, the focal loss function, along with data augmentation that collectively enhance its performance on the challenging task of accurately segmenting the optic disc.

The ablation study and comprehensive experiments conducted on the iChallenge-PM and iChallenge-AMD datasets have demonstrated the effectiveness of MIU-Net. The results indicate that MIU-Net outperforms several established segmentation models, achieving superior accuracy, IoU, and F1-score metrics. The integration of MFE modules significantly improves the model's ability to capture multi-scale information, while the dual attention module effectively focuses the model's attention on the most relevant features for segmentation tasks. The use of focal loss addresses the critical issue of class imbalance in the datasets, ensuring that the model does not bias its predictions towards the more prevalent class. Additionally, the use of data augmentation alleviates the issue of insufficient data and improves the model's robustness, leading to stronger generalization performance.

In conclusion, MIU-Net stands as a testament to the potential of integrating various enhancements within a U-Net based architecture. Its design philosophy and proven efficacy offer a path forward for future research in medical image analysis, opening avenues for even more sophisticated models that could further streamline the diagnostic process for ophthalmologists and enhance patient outcomes.

CRediT authorship contribution statement

Yichen Xiao: Visualization, Validation, Resources, Methodology, Formal analysis, Conceptualization. **Yi Shao:** Validation, Resources, Investigation, Formal analysis, Data curation, Conceptualization. **Zhi Chen:** Methodology, Investigation, Funding acquisition, Formal analysis. **Ruyi Zhang:** Resources, Investigation, Formal analysis. **Xuan Ding:** Resources, Investigation. **Jing Zhao:** Formal analysis, Conceptualization. **Shengtao Liu:** Validation, Formal analysis. **Teruko Fukuyama:** Investigation, Formal analysis. **Yu Zhao:** Resources, Investigation. **Xiaoliao Peng:** Resources, Investigation. **Guangyang Tian:** Writing – original draft, Software, Methodology. **Shiping Wen:** Writing – review & editing, Supervision. **Xingtao Zhou:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

There is no conflict of interest in this paper.

Data availability

No data was used for the research described in the article.

References

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder–decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(12), 2481–2495.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., & … Zhou, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306.

Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder–decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision* (pp. 801–818).

Cheng, J., Liu, J., Xu, Y., Yin, F., Wong, D. W. K., Tan, N. M., & . Wong, T. Y. (2013). Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE Transactions on Medical Imaging*, *32*(6), 1019–1032.

Choi, K. J., Choi, J. E., Roh, H. C., Eun, J. S., Kim, J. M., Shin, Y. K., & . Kim, S. J. (2021). Deep learning models for screening of high myopia using optical coherence tomography. *Scientific Reports*, *11*(1), 21663.

Fang, H., Li, F., Fu, H., Sun, X., Cao, X., Lin, F., & . Xu, Y. (2022). Adam challenge: Detecting age-related macular degeneration from fundus images. *IEEE Transactions on Medical Imaging*, *41*(10), 2828–2847.

Fang, H., Li, F., Wu, J., Fu, H., Sun, X., Orlando, J. I., & . Xu, Y. (2023). PALM: Open fundus photograph dataset with pathologic myopia recognition and anatomical structure annotation. arXiv preprint arXiv:2305.07816.

Foracchia, M., Grisan, E., & Ruggeri, A. (2005). Luminosity and contrast normalization in retinal images. *Medical Image Analysis*, *9*(3), 179–190.

Fu, H., Cheng, J., Xu, Y., Wong, D. W. K., Liu, J., & Cao, X. (2018). Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Transactions on Medical Imaging*, *37*(7), 1597–1605.

Gao, S. H., Cheng, M. M., Zhao, K., Zhang, X. Y., Yang, M. H., & Torr, P. (2019). Res2net: A new multi-scale backbone architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(2), 652–662.

Han, X., Liu, C., Chen, Y., & He, M. (2022). Myopia prediction: a systematic review. *Eye*, *36*(5), 921–929.

Hoover, A. D., Kouznetsova, V., & Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, *19*(3), 203–210.

Li, H., & Chutatape, O. (2004). Automated feature extraction in color retinal images by a model based approach. *IEEE Transactions on Biomedical Engineering*, *51*(2), 246–254.

Li, Y., Foo, L. L., Wong, C. W., Li, J., Hoang, Q. V., Schmetterer, L., & . Ang, M. (2023). Pathologic myopia: Advances in imaging and the potential role of artificial intelligence. *British Journal of Ophthalmology*, *107*(5), 600–606.

Li, M., Liu, S., Wang, Z., Li, X., Yan, Z., Zhu, R., & Wan, Z. (2023). MyopiaDETR: End-to-end pathological myopia detection based on transformer using 2D fundus images. *Frontiers in Neuroscience*, *17*, Article 1130609.

Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980–2988).

Lupon, M., Nolla, C., & Cardona, G. (2024). New designs of spectacle lenses for the control of myopia progression: A scoping review. *Journal of Clinical Medicine*, *13*(4), 1157.

Marín, D., Aquino, A., Gegúndez-Arias, M. E., & Bravo, J. M. (2010). A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Transactions on Medical Imaging*, *30*(1), 146–158.

Mehmood, M., Alsharari, M., Iqbal, S., Spence, I., & Fahim, M. (2024). RetinaLiteNet: A lightweight transformer based CNN for retinal feature segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2454–2463).

Mendonca, A. M., & Campilho, A. (2006). Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Transactions on Medical Imaging*, *25*(9), 1200–1213.

Morales, S., Naranjo, V., Angulo, J., & Alcañiz, M. (2013). Automatic detection of optic disc based on PCA and mathematical morphology. *IEEE Transactions on Medical Imaging*, *32*(4), 786–796.

Niu, Y. N., He, H. L., Chen, X. Y., Ling, S. G., Dong, Z., Xiong, Y., & . Jin, Z. B. (2024). A novel grading system for diffuse chorioretinal atrophy in pathologic myopia. *Ophthalmology and Therapy*, 1–14.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., & . Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

Park, S. J., Ko, T., Park, C. K., Kim, Y. C., & Choi, I. Y. (2022). Deep learning model based on 3D optical coherence tomography images for the automated detection of pathologic myopia. *Diagnostics*, *12*(3), 742.

Patil, Y., Shetty, A., Kale, Y., Patil, R., & Sharma, S. (2024). Multiple ocular disease detection using novel ensemble models. *Multimedia Tools and Applications*, *83*(4), 11957–11975.

Rauf, N., Gilani, S. O., & Waris, A. (2021). Automatic detection of pathological myopia using machine learning. *Scientific Reports*, *11*(1), 16570.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Vol. 18*, In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, munich, Germany, October (2015) 5-9, proceedings, part III* (pp. 234–241). Springer International Publishing.

Tong, H. J., Huang, Z. M., Li, Y. L., Chen, Y. M., Tian, B., Ding, L. L., & Zhu, L. L. (2023). Machine learning to analyze the factors influencing myopia in students of different school periods. *Frontiers in Public Health*, *11*, Article 1169128.

Wan, C., Fang, J., Li, K., Zhang, Q., Zhang, S., & Yang, W. (2024). A new segmentation algorithm for peripapillary atrophy and optic disk from ultra-widefield photographs. *Computers in Biology and Medicine*, *172*, Article 108281.

Yang, Y., Li, R., Lin, D., Zhang, X., Li, W., Wang, J., & . Lin, H. (2020). Automatic identification of myopia based on ocular appearance images using deep learning. *Annals of Translational Medicine*, *8*(11).

Zhang, J., Dashtbozorg, B., Bekkers, E., Pluim, J. P., Duits, R., & ter Haar Romeny, B. M. (2016). Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. *IEEE Transactions on Medical Imaging*, *35*(12), 2631–2644.

Zhang, Y., Li, Y., Liu, J., Wang, J., Li, H., Zhang, J., & Yu, X. (2023). Performances of artificial intelligence in detecting pathologic myopia: a systematic review and meta-analysis. *Eye*, *37*(17), 3565–3573.