Research and Analysis of Research on Automated Segmentation of Microaneurysms in Early Diagnosis of Diabetic Retinopathy

Zhiyong Liu [1], Chenghe Tang [2] and Jieming Yang [3, *]

[1] School of Science and Technology, Hong Kong Metropolitan University, Hong Kong, China

[2] School of Engineering, University of Bristol, Bristol, England

[3] School of Science and Technology, Hong Kong Metropolitan University, Hong Kong, China

* Corresponding Author Email: s1304690@live.hkmu.edu.hk

Abstract. Diabetic retinopathy (DR) is one of the main complications of diabetes, which requires early diagnosis to delay the progression of the disease. As the earliest visible pathological marker of DR, microaneurysm (MA) is of great clinical significance for its precise automatic segmentation. This article systematically reviews the current research progress of MA segmentation technology, focusing on the key challenges in clinical application, such as the small size, low contrast and morphological heterogeneity of MA. On this basis, Ozan Oktay proposed an AG U-Net model that combines an attention mechanism with morphological prior knowledge. By introducing the attention gate control module and dynamic sample weighting strategy, the model significantly improves the accuracy and robustness of segmentation. Experimental results show that the model achieves an excellent dice coefficient of 0.78 on the public data set, which is better than the existing mainstream method while maintaining a high reasoning speed, indicating the promising clinical transformation potential.

Keywords: Diabetic retinopathy, Microaneurysm, Morphological prior, U-Net.

1. Introduction

Diabetic retinopathy (DR) is the most common and severe microvascular complication of diabetes and has become one of the main causes of irreversible blindness among working-age people around the world [1]. Epidemiological data show that more than 460 million people worldwide suffer from diabetes, and about one-third of them suffer from varying degrees of DR [1, 2]. In the early stage of DR, microaneurysm (MA) becomes the earliest detectible pathological marker and can be observed by fundus imaging. Therefore, the accurate detection and segmentation of MA are of key clinical significance for early screening, diagnosis and disease monitoring in DR management.

At present, the screening of clinical diabetic retinopathy (DR) mainly relies on the manual interpretation of fundus images by ophthalmologists. This method is not only time-consuming, but also has a high rate of misdiagnosis due to the limited experience and fatigue of doctors. In addition, there is a serious shortage of qualified ophthalmologists to screen around the world, especially in developing countries with inadequate medical resources. This makes it difficult to effectively implement large-scale and inclusive DR screening [3, 4]. In order to solve this clinical bottleneck, the development of efficient and accurate automatic macular (MA) detection and segmentation technology has become a research hotspot for medical imaging analysis [2].

In recent years, the rapid development of deep learning technology, especially the emergence of fully convolutional neural networks (FCN) and its derivative models (such as U-Net), has significantly improved the performance of medical image segmentation. U-Net has a unique encoder-decoder architecture and skip connection, effectively combining shallow details with deep semantic features, and establishing itself as the basic framework for fundus image analysis [3]. On this basis, the researchers have proposed various improvement strategies to solve the unique challenges in the segmentation of macular arteries (MA). These include the use of attention mechanisms to focus on

the target area [4], or the use of multi-scale feature fusion strategies to improve the detection ability of subtle MA lesions.

However, the automatic segmentation of microaneurysms (MA) still faces major challenges. First of all, the size of MA is extremely small (usually with a diameter of 10-100 pixels) and the contrast with the surrounding tissue is low, making it easy to be ignored or misrecognized [5]. Secondly, the morphological diversity of MA leads to blurred boundaries when clustering, which makes accurate counting and segmentation complicated. Existing methods are often insufficient in terms of real-time performance, suppression of false positive targets, and deep integration with medical prior knowledge, which limits their practical application in the clinical environment.

This article aims to systematically review the current research status and challenges of MA automatic segmentation technology, and propose a new AG U-Net model to solve the limitations of the existing method. Based on the classic U-Net architecture, the model combines the attention gate mechanism with the dynamic sample weighting strategy to improve the accuracy, robustness and real-time performance of MA segmentation. The proposed model provides a new solution to promote the clinical translation application in this field. This review systematically examines the mainstream technical methods, analyzes the advantages and limitations of various methods, evaluates the performance of the proposed model through comparative experiments, and summarizes the potential research direction in the future.

2. Data sets and preprocessing

2.1. Data sources and characteristics

The progress of automated microaneurysm segmentation research largely depends on high-quality, publicly accessible fundus image data sets. The widely used benchmarks in this field include: Indian diabetic retinopathy image data set (IDRiD), which is a gold standard data set containing 516 high-resolution fundus images, which are manually annotated at the pixel level by three ophthalmologists, showing excellent consistency (Kappa>0.95), and it is often used as the gold standard for model performance evaluation [6]. The E-ophtha-MA data set focuses on providing more challenging samples [7], consisting of 381 images, including 129 special cases with low contrast and motion blur scenes, which have been proven to be crucial to improve the robustness of the model.

In addition to the above two core data sets, other commonly used data sets have different focuses. FOR EXAMPLE, THE MESSIDOR DATA SET IS VERY LARGE, CONTAINID 1,200 OF THE INORMAGES. Although it lacks pixel-level macular abnormality (MA) annotation, it provides diabetic retinopathy (DR) grade labels, making it suitable for classification task research [8]. The Diabetic Retinopathy Data Set (DDR) is a large-scale data set that provides pixel-level lesion annotations, covering various DR lesions, such as MA and bleeding points, making it ideal for multi-lesion joint segmentation tasks [9]. In addition, the retinopathy online challenge (ROC) data set used to be a platform for the MA detection algorithm competition [10], and its data characteristics have played a positive role in promoting algorithm development.

These data sets vary in terms of image number, resolution, annotation granularity (image level/pixel level) and lesion type coverage. Researchers usually combine multiple data sets according to specific task requirements to fully verify the model promotion ability. In order to comprehensively evaluate the performance of the model, this study mainly uses the IDRiD data set with the highest annotation quality and the E-ophtha-MA data set containing a large number of challenging cases for experimental analysis.

2.2. Common steps of preprocessing process and characteristics of data set

Although different data sets differ in terms of image quality, resolution and acquisition equipment, the preprocessing process of MA automatic segmentation research usually follows a set of common steps to enhance image features, unify data formats, and improve the generalization ability of the model. Common preprocessing steps include:

Image enhancement: In order to enhance the contrast between the structure of blood vessels and the background, most studies give priority to extracting the green channel of the fundus image, because it contains the most abundant vascular and MA information [11]. Subsequently, contrast-limited adaptive histogram equalization (CLAHE) was widely used to further enhance local texture details and improve the distinguishability of low-contrast areas.

Size unification and normalization: In view of the different resolutions of the original images, they are usually resized or cropped to standardized sizes (for example, 512×512 pixels) to meet network input requirements. The pixel value is also normalized to the range of [0,1] or [-1,1] to accelerate the convergence of the model.

Data enhancement: In order to simulate potential clinical image variations and solve limited data availability, common enhancement strategies include random rotation, flipping, brightness/contrast adjustment and gamma transformation. Some studies use random size cutting for the form of macular abnormality (MA), focusing on high-risk areas such as the periphery of macula to improve the sensitivity of the model to subtle lesions.

It is worth noting that specific preprocessing strategies are used to solve the unique challenges of specific data sets. For example, when processing motion blur or overexposed images in the E-ophtha-MA data set, some studies use the image denoising or lighting correction module based on Retinex as a preliminary step. When processing high-resolution data sets such as IDRiD, overlapping repair techniques are usually used during training to prevent the loss of details.

In a word, although the preprocessing process has been slightly adjusted according to specific tasks and data characteristics, its core goal is the same: that is, through standardization and enhancement, high-quality input that is more suitable for model learning is built, laying the foundation for subsequent accurate segmentation.

3. Overview of relevant modeling approaches

In the field of automated microaneurysm segmentation, in addition to U-Net-based improvements, several unique technical methods have also emerged. This section reviews the representative methods, including the AG U-Net proposed in this study.

3.1. Methods based on U-Net and its variants

U-Net has established itself as the gold standard for medical image segmentation through its elegant encoder-decoder architecture and skip connection. For mask-assisted segmentation tasks, variants such as Attention U-Net improve performance by incorporating the attention gate control mechanism into skip connection, effectively suppressing unrelated background areas, and improve attention to small targets. MA-Net further solves the challenge of scale change in mask-assisted segmentation by introducing a multi-scale attention fusion module [12]. The AG U-Net proposed in this study belongs to this category, which is characterized by the integration of attention gate control and form-based previous loss functions to achieve more precise segmentation.

3.2. RCPS-Net based on cyclic collaboration strategy

RCPS-Net adopts an innovative cyclical collaboration strategy. Its core concept involves building a segmentation framework composed of multiple subnets, which improve the segmentation results through iterative cycle collaboration in the reasoning process. Specifically, the output from one network serves as a priori knowledge or guidance for the next network. Through this repeated "refinement" process through multiple iterations, the model gradually optimizes the ambiguous or uncertain areas in the initial segmentation diagram, especially improving the accuracy of the segmentation clustering with unclear boundaries. This method embodies the principle of "cooperative optimization", which is in stark contrast to the model that relies on single positive propagation.

3.3. CSF-Net based on vascular structure prior

CSF-Net adopts an innovative method to clearly utilize the anatomical a priori knowledge of the structure of eye blood vessels. The model usually consists of two parallel branches: one is dedicated to the main segmentation task (MA detection), and the other is dedicated to vascular segmentation. Then, the feature fusion module will integrate the topological information extracted from the vascular branch into the main branch. The design is based on the theoretical premise that MA is caused by vascular wall lesions, and its distribution is closely related to the vascular network, especially the peripheral capillaries. By introducing vascular diagrams as a spatial constraint, the model can effectively filter out false alarms in non-vascular areas, so as to accurately detect MA in the correct anatomical context. This significantly enhances the specificity of subdivision.

3.4. Model Architecture: Attention Gate U-Net

In order to meet these challenges, this paper proposes a new AG U-Net model based on the U-Net framework, and integrates attention gates and dynamic sample weighting strategies. As shown in Figure 1.

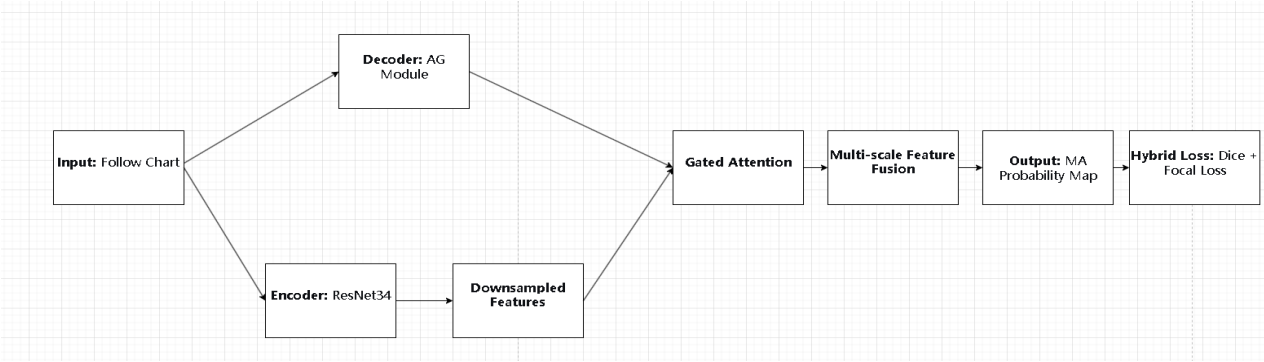


Figure 1. Attention Gate U-Net

3.5. Core innovation modules

3.5.1. Attentional gate control

The researchers implemented an attention gate control (AG) module in skip connection. The mechanism uses advanced semantic features from the decoder as gate control signals to dynamically adjust the weight of the low-level detail features from the encoder. By inhibiting irrelevant background responses, the network becomes more focused on identifying potential microaneurysm areas.

3.5.2. Dynamic sample weighting strategy

The researchers implemented a dynamic weighting strategy that combines focus loss and cyclic loss. By adjusting the weight of challenging cases, the focus loss alleviated the serious imbalance between positive and negative samples. Cyclic loss uses the medical premise that the morphological structure is usually circular, and uses the area-to-perimeter ratio of the predicted area (4π*area/perimeter^2) to calculate. By punishing non-circular prediction, this method guides the segmentation results to be better consistent with the pathological morphological characteristics.

4. Experimental results and analysis

4.1. Comparative analysis of quantitative results

In order to comprehensively evaluate the performance of the AG U-Net model, this paper conducts comparative experiments on a variety of mainstream segmentation methods of the E-ophtha-MA test set. The results are shown in Table 1.

Traditional methods and basic deep learning models have significant limitations. The original U-Net model only achieved a dice coefficient of 0.62 and a recall rate of 0.58, with an average of 8.3 false positive tests per image, highlighting the shortcomings of its basic segmentation architecture in handling small, low-contrast micro-agents [3]. Although DeepLab v3+ (classic semantic segmentation model) is better than the original U-Net with a dice coefficient of 0.67, it is still difficult to detect tiny targets and maintain a relatively slow reasoning speed (19 fps).

In recent years, the special model of MA segmentation has shown enhanced adaptability. MA-Net (MICCAI 2023) achieves a dice coefficient of 0.73, and reduces the false alarm of each image to 4.2 through its multi-scale attention mechanism, highlighting the professional design value of MA segmentation tasks. This paper further evaluates RCPS-Net, which uses a periodic collaboration strategy to optimize boundary segmentation with an excellent recall rate (0.72), although its relatively low dice coefficient (0.70) indicates that there is potential room for improvement in precise boundary positioning. CSF-Net achieved significant false-positive inhibition (2.8 per image) by incorporating vascular structure a priori, but its conservative recall rate (0.65) indicates that the MA segment away from the blood vessel may be missing.

The AG U-Net model proposed in this study showed the best or near-best performance in all indicators, achieving a dice coefficient of 0.78 - 25.8% higher than the original U-Net. Its recall rate (0.75) and false positive control (2.1/Figure) are also significantly better than the competitive method. It is worth noting that while maintaining high accuracy, the model achieves a reasoning speed of 27 fps, meets the requirements of real-time processing, and shows an excellent balance between accuracy and efficiency. These results verify the effectiveness of combining the attention mechanism with the previous strategies of morphology.

Table 1. Performance comparison of different models on E-ophtha-MA test set

model	Dice coefficient ↑	recall ↑	False positive/image ↓	Inference speed↑(fps)
original U-Net	0.62	0.58	8.3	23
DeepLabv3+	0.67	0.61	6.7	19
MA-Net (MICCAI 2023)	0.73	0.69	4.2	25
RCPS-Net	0.70	0.72	3.5	18
CSF-Net	0.68	0.65	2.8	22
AG U-Net	0.78	0.75	2.1	27

In a word, models specially designed for MA features (such as MA-Net, RCPS-Net and CSF-Net) have significant advantages over general segmentation models. At the same time, AG U-Net provides an effective solution for MA segmentation tasks by organically combining attention mechanism and morphological a priori, and achieves the best false positive control while maintaining a high recall rate.

4.2. Ablation experiment (IDRiD validation set)

The ablation experiment proves the effectiveness of each innovative module (Table 2). Attention gate control has brought significant improvements. After the introduction of cyclical loss, the performance has reached its peak, which verifies the great value of integrating medical prior knowledge to improve segmentation accuracy.

Table 2. Analysis of ablation experiment results

module	Dice	Δvs base line
Baseline (U-Net)	0.62	-
+ Attention Gate	0.71	+14.5%
+ Focal Loss	0.74	+19.4%
+ Circularity Loss	0.78	+25.8%

5. Discussion and outlook

5.1. Method validity analysis and limitations

The AG U-Net model proposed in this paper shows the competitive performance of the public data set and verifies the effectiveness of combining the attention mechanism with the a priori knowledge of morphology to solve the core challenges of MA segmentation. Note that the gate control module significantly enhances the attention to small and low-contrast targets through dynamic feature weighting, which is proven by the 14.5% improvement in the dice coefficient observed in the ablation experiment. In addition, the dynamic weighting strategy that combines focus loss and cyclic loss [13], not only alleviates class imbalance, but also imposes morphological restrictions on the segmentation results, making them consistent with medical a priori. This dual method enables the model to maintain a high recall rate and achieve excellent false alarm control ability.

However, this study still has certain limitations. Although the model performs well in data sets with high annotation quality, its performance fluctuates significantly when dealing with clinical low-quality images with severe motion blur, overexposure or a large number of artifacts. In addition, although circulatory loss effectively guides typical circulatory MA cases, for cluster MA cases with irregular shape or highly vagueness, this constraint may be too strict, resulting in incomplete segmentation of some lesions or missed detection. This highlights the inadequacy of single morphological a priori in solving the high heterogeneity of MA cases.

5.2. Core challenges in clinical translation

Transforming a high-performance algorithm model into a truly reliable clinical tool still needs to overcome several core challenges. The main challenge is to ensure the robustness of the model in the real-world environment. In view of the significant differences in the quality of clinically collected fundus images, the model must show insensitivity to different degrees of blurring, noise and lighting changes. This requires thorough training and verification using large-scale, multi-center real-world data.

Secondly, there is the problem of model promotion and standardization. Different medical institutions use fundus cameras with different models and imaging protocols, resulting in significant differences in image quality. A model that performs well on one data set may experience a sharp decline in performance when applied to another source. Therefore, it is crucial to develop a generalized model that can go beyond the limitations of a single device or data domain to achieve the widespread adoption of technology.

Finally, the integration of clinical trust with workflow is still a key component. The algorithm must provide measurable uncertainty indicators or visual explanations for their decision-making, such as highlighting key decision-making areas - to help doctors understand and trust the algorithm output. This enables automated tools to be effectively integrated into existing diagnostic workflows, and finally establishes a human-computer collaboration framework for optimizing screening protocols.

5.3. Future research directions

Based on previous discussions and challenges, future research will give priority to improving the applicability and robustness of the model in the real-world clinical environment. The main focus should be on systematically improving the model robustness of low-quality real-world images through end-to-end joint training between image quality enhancement modules and split networks. In order to solve the generalization restrictions caused by changes in equipment and centers, it is necessary to thoroughly study and implement advanced domain adaptation and generalization technologies, and use large-scale multi-center data sets to develop the "domain unknowable" model capability. In addition, in order to overcome the challenges of insufficient samples and rare morphological MA cases with impaired segmentation reliability, future efforts should emphasize the in-depth integration of multi-source medical prior knowledge - such as vascular topological relationships and pathophysiological mechanisms. This should be supplemented by generative artificial intelligence to integrate various challenging cases and optimize data distribution. It is equally important to develop an inherent interpretable model architecture or reliable interpretation methods, because these enhance the trust of clinicians and achieve interactive correction in human-computer collaborative diagnosis, and jointly promote the comprehensive clinical implementation of technology.

6. Conclusion

This paper systematically reviews the latest progress of automated microaneurysm segmentation technology for early diagnosis of diabetic retinopathy. In order to solve the core challenges of small size, low contrast and morphological heterogeneity of microaneurysms (MA), this paper proposes and verifies an AG U-Net model that combines attention mechanism with morphological prior knowledge. Based on the existing mainstream method, our model introduces a attention gate control module in the skip connection to focus on the potential lesion area, while combining the dynamic sample weighting strategy (integrated focus loss and cyclic loss) to solve the sample imbalance problem and ensure that the segmentation results are consistent with the medical morphological specifications. The proposed model achieves an excellent dice coefficient of 0.78 on the public data set, showing a good balance between false positive control and reasoning speed, thus verifying the effectiveness of our proposed strategy.

In short, this study shows that a deep learning model that combines attention mechanism with a priori knowledge in a specific field can significantly improve the accuracy and robustness of automatic microaneurysm (MA) segmentation, and provide a feasible technical solution for early automatic screening of diabetic retinopathy (DR). Future research should focus on improving the model robustness of low-quality images in the real world, enhancing the identification ability of rare MA forms, and solving the generalization challenges across equipment and clinical centers. By deeply integrating the previous knowledge of medicine, exploring advanced field adaptive methods, and using generation models for data enhancement, this paper can promote the automated microaneurysm segmentation technology from algorithm innovation to a wide range of clinical applications. This progress will eventually support the fair screening and precise diagnosis of diabetic retinopathy through powerful machine learning solutions.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] International Diabetes Federation. IDF Diabetes Atlas, 10th edition. Brussels, Belgium: International Diabetes Federation, 2021.

[2] World Health Organization. World Report on Vision. Geneva: World Health Organization, 2019.

[3] Ranneberger O, Fischer P, & Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) (pp. 234 – 241). 2015.

[4] Oktay O, Schlemper J, Folgoc L, et al. Attention U-Net: Learning Where to Look for the Pancreas. In Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL). 2018.

[5] Chang J, Wang L, & He J. Study on the Pathogenesis of Diabetic Retinopathy. Advances in Clinical Medicine, 10 (3), 1 - 5. 2020.

[6] Porwal P, Pachade S, Kamble R, et al. Indian Diabetic Retinopathy Image Dataset (IDRiD): A Dataset for Diabetic Retinopathy Segmentation and Grading. IEEE Dataport. 2018.

[7] Decencière E, Zhang X, Cazuguel G, et al. Feedback on a Publicly Distributed Database: The Messidor Database. Image Analysis & Stereology, 33 (3), 231 – 234. 2014.

[8] Adal K M, Etienne C, Ali S, et al Automated Detection of Microaneurysms Using Scale-Adapted Blob Analysis and Semi-Supervised Learning. Computer Methods and Programs in Biomedicine, 114 (1), 1 - 10. 2014.

[9] Li T, Gao Y, Wang K, et al. Diagnostic Assessment of Deep Learning for Diabetic Retinopathy Detection. IEEE Transactions on Medical Imaging, 38 (9), 1 - 1. 2019.

[10] Niemeijer M, Ginneken B V, Cree M J, et al. Retinopathy Online Challenge: Automatic Detection of Microaneurysms in Digital Color Fundus Photographs. IEEE Transactions on Medical Imaging, 29 (1), 185 – 195. 2010.

[11] Tan J H, Acharya U R, Bhandary S V, et al. Segmentation of Optic Disc, Fovea and Retinal Vasculature Using a Single Convolutional Neural Network. Journal of Computational Science, 26, 70 – 79. 2018.

[12] Li X, Hu X, Yu L, et al. MA-Net: A Multi-Scale Attention Network for Liver and Tumor Segmentation. IEEE Access, 8, 179656 – 179665. 2020.

[13] Chen C, Qin C, Qiu H, et al. A Shape-Aware Loss for Medical Image Segmentation. IEEE Transactions on Medical Imaging, 41 (5), 1245 – 1258. 2022.