# ASDC-NET: ANTERIOR SEGMENT DISEASE CLASSIFICATION NETWORK BASED ON SLIT-LAMP IMAGES

*Yiming Ding[1], Weifang Zhu[1], Fei Shi[1], Ruiqi Ma[2], Xinjian Chen[1,3]*

[1]School of Electronics and Information Engineering, Soochow University, China
[2]Department of Ophthalmology, Fudan Eye & ENT Hospital, Shanghai, China
[3]State Key Laboratory of Radiation Medicine and Protection, Soochow University, China

2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI) | 979-8-3315-2052-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/ISBI60581.2025.10980943

**ABSTRACT**

Anterior segment diseases are common in clinical ophthalmology, but most studies focus on individual conditions, with relatively few addressing the broader classification of these diseases. In this paper, we propose an anterior segment disease classification network (ASDC-NET) to classify 5 types of diseases based on slit-lamp images. The proposed method designs a wavelet feature enhancement (WFE) module to extract frequency domain information, a multi-scale feature fusion (MSFF) module to combine features at different scales, and an adaptive channel aggregation (ACA) module to adapt to lesion variability. Evaluation results on the dataset including 1,250 slit-lamp images demonstrate the excellent classification performance of the proposed method, outperforming other state-of-the-art methods.

***Index Terms***— Slit-lamp image, Deep learning, Anterior segment diseases, Image classification

## 1. INTRODUCTION

The anterior segment is defined as the first third of the eye and includes the conjunctiva, cornea, anterior chamber, iris, pupil, ciliary body, and lens, which constitutes the path of light through the eye and the refractive system of the eye [1]. Pterygium, age-related cataract, corneal ulcers, conjunctivitis, eyelid masses are common anterior segment diseases in ophthalmology clinic, which may lead to vision impairment or other serious complications. The slit-lamp microscope, is one of the imaging devices frequently used to examine anterior segment diseases, which allows for a clear observation of the specific location, size, nature and depth of anterior segment lesions. Automated identification of anterior segment diseases based on slit-lamp images can assist in early diagnosis and timely treatment planning.

Artificial intelligence has made substantial progress in the medical field, with its early applications in ophthalmology primarily targeting the detection and diagnosis of retinal diseases. Recently, some studies have focused on anterior segment diseases of the eye. Zheng et al. [2] developed a model for aiding the diagnosis of pterygium using transfer learning, which distinguishes normal, monitoring-phase, and surgery-required from anterior segment images. Xu et al. [3] applied a stacked multi-feature approach to assess the severity of cataracts, using techniques like residual networks and gray-level co-occurrence matrices to grade the disease into six levels of severity. However, most current research remains centered on the classification of specific diseases, with fewer studies addressing the classification of various disease types. Li et al. [4] improved the diagnostic performance for anterior segment diseases (such as cataracts, keratitis, and pterygium) in slit-lamp images by segmenting anatomical structures and annotating pathological changes.

The Vision Transformer (ViT) [5], a model built on self-attention mechanisms, has been increasingly applied to medical image analysis. By capturing long-range dependencies, it enhances classification performance. Manzari et al. [6] designed the MedViT model, trained and tested on multiple publicly available medical image datasets (including CT, X-ray, ultrasound, and OCT images), and it demonstrated outstanding performance in several medical image classification tasks. Despite its potential, the model faces challenges, such as high computational complexity and a requirement for large amounts of labeled data, which limit its broader clinical adoption. Researchers are now exploring more efficient models, such as MobileViT [7], which aims to reduce computational demands while maintaining high performance. This line of research offers new perspectives for improving the efficiency and accuracy of medical image analysis. For example, Cao et al. [8] combined dilated convolution with standard convolution to extract multi-scale local features, optimizing the MobileViT module for classifying benign and malignant lung nodules. Varam et al. [9] tested various ViT models (including MobileViT) on gastrointestinal disease datasets to develop a model that improves the efficiency of medical image analysis while reducing computational resource requirements.

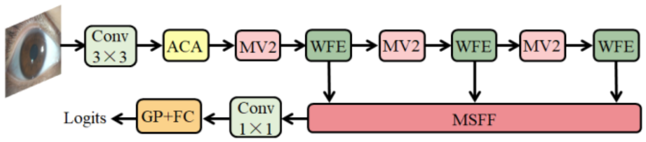


**Fig. 1.** Overview of the proposed network.

In this paper, we propose a novel MobileViT based network for 5 types of anterior segment disease classification in slit-lamp images(ASDC-NET). The main contributions of this work can be summarized as follows: (1) A wavelet feature enhancement (WFE) module is designed to capture frequency information, which is significant for image classification. (2) A multi-scale feature fusion (MSFF) module is designed to fully leverage both deep and shallow features, thereby enhancing the representational capability of the network. (3) Due to the various shapes and locations of lesions for different diseases, an adaptive channel aggregation (ACA) module is designed to assist the network in making fine adjustments to important regions on the feature map.

## 2. METHOD

### 2.1. Overall structure of the proposed network

The overall structure of the proposed ASDC-NET is shown in Figure 1, which is based on MobileViT. Considering that the frequency domain also contains useful information, a WFE module is designed to replace the original MobileViT block. Three WFE modules are used to extract feature information at different scales, and the MSFF module is employed to fuse the multi-scale feature information. An ACA module is designed to replace the original first MV2 block to address the situation where the shapes and locations of lesions vary across different diseases.

### 2.2. Wavelet Feature Enhancement (WFE) module

We consider introducing the wavelet transform into the upper branch, which is originally unprocessed and directly performed feature concatenation, to supplement the information from the frequency domain. This allows the branch not only to simply transmit unprocessed features but also to further extract and enhance features from the perspective of the frequency domain. Figure 2 shows the structure of the proposed wavelet feature enhancement (WFE) module. The input features in the upper branch sequentially pass through a discrete wavelet transform (DWT), a 3×3 convolution layer, and an inverse discrete wavelet transform (IDWT) to obtain convolutional features from different frequency bands. These features are then concatenated with the features processed by the local representation block and the global representation block, and

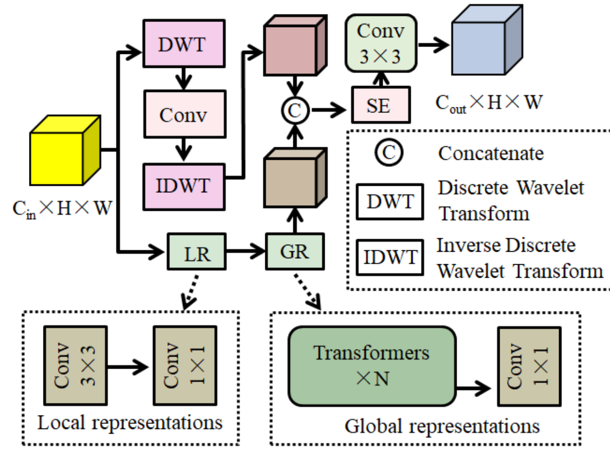


**Fig. 2.** Structure of Wavelet Feature Enhancement module.

the SE (Squeeze-and-Excitation) module [10] is used to adjust the channel weights. This combines information from both the frequency and spatial domains, enhancing the feature representation capabilities of the model.

### 2.3. Multi-scale Feature Fusion (MSFF) module

We posit that the features extracted by each WFE module have different receptive fields and semantic levels, and the features extracted by each block may contain valuable information. High-level features typically capture broader contextual information at the cost of some detail, while low-level features, though rich in detail, lack global information. Relying solely on features from a single level may lead to excessive sensitivity to certain specific patterns, while combining multi-level features allows the network to consider information from different levels simultaneously. Figure 3 illustrates the structure of the proposed multi-scale feature fusion (MSFF) module. The features of the first two WFE modules are downsampled independently and then concatenated with the features of the third WFE module. After concatenation, the SE module adjusts the feature weights, followed by RELU activation to introduce nonlinearity, and finally, adaptive modulation through element-wise multiplication. The multi-level fusion ensures that information from both shallow and deep layers participates in the final inference process, thereby reducing the risk of information loss. By fusing features from different levels, the network can better classify different object categories.

### 2.4. Adaptive Channel Aggregation (ACA) module

The first MV2 block in MobileViT is replaced with an adaptive channel aggregation module, as illustrated in Figure 4. This modification provides greater flexibility in handling complex details and local structures. Traditional convolution

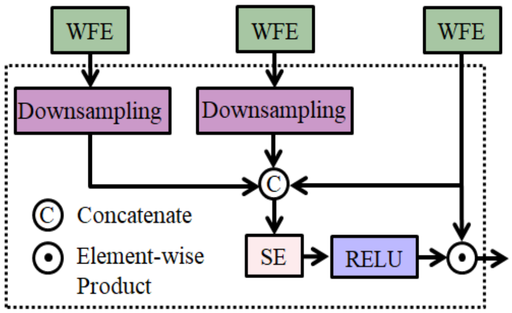


**Fig. 3.** Structure of Multi-scale Feature Fusion module.

operations are often limited by fixed local windows and sampling shapes, whereas AKConv (Adaptive Kernel Convolution) [11] allows the convolutional kernel to adjust its number of parameters and sampling shapes dynamically, enabling more precise adaptation to complex scenarios. By decomposing the channel number to 1, the learnable scaling factor (σ) allows the model to flexibly control the balance between the original and decomposed features, restoring the channel number to its original dimension, and finally outputting through a residual connection. Feature learning is enhanced through channel aggregation, improving the quality of feature representation and increasing the model's flexibility and generalization ability.

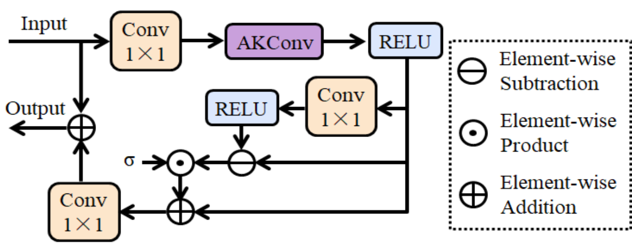


**Fig. 4.** Structure of Adaptive Channel Aggregation module.

## 3. EXPERIMENTS AND RESULTS

### 3.1. Dataset and implementation details

1,250 slit-lamp images from 1,250 eyes were collected from the Eye and ENT Hospital of Fudan University, which was approved by the Review Committee of the Hospital (No. 2023107) and adhered to the principles of the Helsinki Declaration. All subjects obtained informed consent. All data were categorized based on diagnoses made by professional ophthalmologists into pterygium, age-related cataract, corneal ulcers, conjunctivitis, eyelid masses and normal. The dataset was randomly divided into training, validation, and test sets in the ratio of 6:2:2. Table 1 shows the detailed data distribution.

The proposed method was performed on the public platform Pytorch and a NVIDIA RTX 3060 GPU of 12G memory. Since the dataset contains three image sizes (3648×2432, 4800×3200, and 5472×3648), the original images were resized to 512×512 using bilinear interpolation to ensure uniformity for training. To prevent overfitting, online data augmentation including horizontal flipping, random cropping and color perturbation were performed. The AdamW was used as the optimizer, initial learning rate was set to 0.0002, the batch size was set to 8, and the number of epochs was 120.

**Table 1.** The data distribution of the dataset.

| Categories | Training | Validation | Testing | Total |
|---|---|---|---|---|
| Pterygium | 115 | 38 | 38 | 191 |
| Age-related cataract | 224 | 75 | 74 | 373 |
| Corneal ulcers | 151 | 50 | 51 | 252 |
| Conjunctivitis | 73 | 24 | 24 | 121 |
| Eyelid masses | 85 | 28 | 28 | 141 |
| Normal | 102 | 35 | 35 | 172 |
| All | 750 | 250 | 250 | 1250 |

### 3.2. Evaluation metrics and results

In order to evaluate the classification performance of different classification methods comprehensively and fairly, weighted recall (W_R), weighted precision (W_P) and weighted F1 score (W_F1) were used to quantitatively analyze the experimental results.

**Table 2.** Quantitative comparison among different methods.

| Methods | W_R | W_P | W_F1 |
|---|---|---|---|
| ResNet | 0.8321 | 0.8438 | 0.8343 |
| EffcientNet | 0.8600 | 0.8659 | 0.8606 |
| DenseNet | 0.8597 | 0.8651 | 0.8589 |
| RepViT | 0.8368 | 0.8283 | 0.8304 |
| Baseline | 0.8680 | 0.8695 | 0.8678 |
| Baseline+WFE | 0.8919 | 0.8945 | 0.8918 |
| Baseline+MSFF* | 0.8734 | 0.8642 | 0.8652 |
| Baseline+ACA | 0.8881 | 0.8894 | 0.8872 |
| **Proposed** | **0.9120** | **0.9128** | **0.9121** |

* indicates that the input to the MSFF model is the original MobileViT block.

The proposed ASDC-NET is compared with other state-of-the-art methods such as ResNet[12], EffcientNet[13], DenseNet[14], and RepViT[15], all of which are trained using the same parameters. MobileViT is taken as the baseline in our experiments. As shown in Table 2, the proposed method demonstrates superior performance compared to the other networks. Compared with the baseline, the proposed ASDC-NET achieves an increase of 4.40% and 4.33% for the weighted recall and weighted precision, respectively. The experimental results demonstrate that ASDC-NET is more effective in addressing the problem of anterior segment disease classification. Additionally, the ablation experiments about the proposed WFE, MSFF and ACA modules have been performed. Compared to the baseline, ablation experiment results show the improvements across all three metrics, further confirming the effectiveness of the proposed modules. Figure 5 shows different types of diseases and their corresponding visual feature maps, indicating that our proposed model can accurately identify different categories of features.

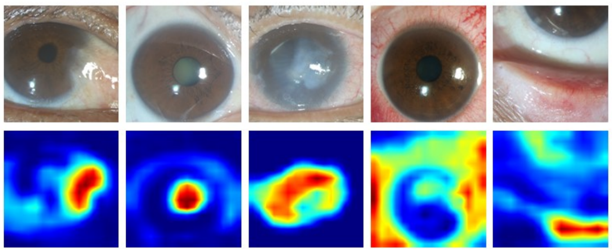

**Fig. 5.** Visualizations of feature map with different anterior segment diseases.

## 4. CONCLUSION

In this paper, we propose a novel classification network designed to distinguish various anterior segment diseases based on slit-lamp images, achieving commendable classification performance. The effectiveness of the proposed method is demonstrated through comparison and ablation experiments on the dataset. Given that slit-lamp images contain a wealth of color information, we believe that making better use of this information could lead to improved classification results. In future work, we will focus on utilizing this color information more effectively.

## 5. ACKNOWLEDGMENTS

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFA0701700, in part by the National Natural Science Foundation of China under Grant U20A20170, 62371326 and 62271337, in part by Natural Science Foundation of Jiangsu Province under Grant BK20231310.

## 6. REFERENCES

[1] Shao, Yi, Ying Jie, Zu-Guo Liu, Chinese Ophthalmic Imaging Study Groups, and Intelligent Medicine Branch of Chinese Medicine Education Association. "Guidelines for the application of artificial intelligence in the diagnosis of anterior segment diseases (2023)." International Journal of Ophthalmology 16, no. 9 (2023): 1373.

[2] Zheng, Bo, Yunfang Liu, Kai He, Maonian Wu, Ling Jin, Qin Jiang, Shaojun Zhu, Xiulan Hao, Chenghu Wang, and Weihua Yang. "Research on an intelligent lightweight- assisted pterygium diagnosis model based on anterior segment images." Disease Markers 2021, no. 1 (2021): 7651462.

[3] Xu, Xi, Linglin Zhang, Jianqiang Li, Yu Guan, and Li Zhang. "A hybrid global-local representation CNN model for automatic cataract grading." IEEE journal of biomedical and health informatics 24, no. 2 (2019): 556-567.

[4] Li, Wangting, Yahan Yang, Kai Zhang, Erping Long, Lin He, Lei Zhang, Yi Zhu et al. "Dense anatomical annotation of slit-lamp images improves the performance of deep learning for the diagnosis of ophthalmic disorders." Nature Biomedical Engineering 4, no. 8 (2020): 767-777.

[5] Dosovitskiy, Alexey. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).

[6] Manzari, Omid Nejati, Hamid Ahmadabadi, Hossein Kashiani, Shahriar B. Shokouhi, and Ahmad Ayatollahi. "MedViT: a robust vision transformer for generalized medical image classification." Computers in Biology and Medicine 157 (2023): 106791.

[7] Mehta, Sachin, and Mohammad Rastegari. "Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer." arXiv preprint arXiv:2110.02178 (2021).

[8] Cao, Keyan, Hangbo Tao, Zhiqiong Wang, and Xi Jin. "MSM-ViT: A multi-scale MobileViT for pulmonary nodule classification using CT images." Journal of X-Ray Science and Technology 31, no. 4 (2023): 731-744.

[9] Varam, Dara, Lujain Khalil, and Tamer Shanableh. "On-Edge Deployment of Vision Transformers for Medical Diagnostics Using the Kvasir-Capsule Dataset." Applied Sciences 14, no. 18 (2024): 8115.

[10] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132-7141. 2018.

[11] Zhang, Xin, Yingze Song, Tingting Song, Degang Yang, Yichen Ye, Jie Zhou, and Liming Zhang. "AKConv: Convolutional kernel with arbitrary sampled shapes and arbitrary number of parameters." arXiv preprint arXiv:2311.11587 (2023).

[12] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

[13] Tan, Mingxing. "Efficientnet: Rethinking model scaling for convolutional neural networks." arXiv preprint arXiv:1905.11946 (2019).

[14] Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. "Densely connected convolutional networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700-4708. 2017.

[15] Wang, Ao, Hui Chen, Zijia Lin, Jungong Han, and Guiguang Ding. "Repvit: Revisiting mobile cnn from vit perspective." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15909-15920. 2024.