

# CHANNEL-WISE AND SPATIAL FEATURE RECALIBRATION NETWORK FOR NUCLEAR CATARACT CLASSIFICATION

Xiaoqing Zhang<sup>1</sup>, Gelei Xu<sup>1</sup>, Junyong Shen<sup>1</sup>, Zunjie Xiao<sup>1</sup>, Qiuyang Yan<sup>1</sup>, Jin Yuan<sup>2</sup>,  
Risa Higashita<sup>1,3\*</sup>, Jiang Liu<sup>1,4\*</sup>

<sup>1</sup>Research Institute of Trustworthy Autonomous Systems and Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China

<sup>2</sup>State Key Laboratory of Ophthalmology, Sun Yat-sen University, Guangzhou, China

<sup>3</sup>Tomey Corporation, Japan

<sup>4</sup>Cixi Institute of Biomedical Engineering, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, China

Email: risa@mail.sustech.edu.cn, liuj@sustech.edu.cn

## ABSTRACT

Nuclear cataract (NC) is a prior age-related disease for blindness and vision impairment globally. Anterior segment optical coherence tomography (AS-OCT) image is a new ophthalmology image, which can capture the lens nucleus region clearly compared with other ophthalmic images, e.g., slit lamp images. Clinical research has suggested that features e.g., mean from AS-OCT images have varying correlations with NC severity levels. However, existing convolutional neural network (CNN) based NC classification works have not incorporated the clinical features into the network design to improve the performance. To this end, we propose a novel channel-wise and spatial feature recalibration network (CSFR-Net) to predict NC severity levels automatically, which is built on a stack of channel-wise and spatial feature recalibration (CSFR) modules. In each CSFR module, we construct a channel-wise feature recalibration block and a spatial feature recalibration block to recalibrate intermediate feature maps dynamically. This feature recalibration strategy enables CSFR-Net to highlight feature representations and suppress unnecessary ones in a global-and-local manner. We conduct extensive experiments on a clinical AS-OCT image dataset and CIFAR benchmarks. The results show that our CSFR-Net achieves better performance than state-of-the-art methods with less model complexity.

**Index Terms**— Nuclear cataract classification, channel-wise and spatial feature recalibration, attention, AS-OCT

## 1. INTRODUCTION

Age-related cataract is the leading ocular disease for blindness and vision impairment globally. According to World

Health Organization (WHO) [1], it is estimated that 65.4 million cataract patients are suffering from moderate and severe vision impairment. Nuclear cataract (NC) is one of the most common age-related cataract types. Its symptoms include the gradual clouding and progressive hardening in the nucleus region of the crystalline lens. According to clinical diagnosis requirements [2], we roughly group NC into three severity levels based on Lens Opacities Classification System III (LOCS III): normal, mild, and severe. Normal people do not take any therapeutic measures; mild NC patients can take clinical intervention to slow the opacity development progress; severe NC patients should undergo cataract surgery or take clinical follow-up. Clinically, ophthalmologists usually adopt slit lamp images to diagnose NC based on their experience and professional knowledge. However, this diagnosis mode is error-prone and subjective. Since slit lamp images have limitations in clearly capturing the nucleus region.

AS-OCT is a new ophthalmology imaging, which is capable of capturing the lens area, comprised of nucleus-, cortex-, and capsule- regions compared with other ophthalmic images, e.g., slit lamp images. Recent research has studied the correlation between clinical features like mean and maximum on AS-OCT images and NC severity levels based on the Spearman correlation coefficient method. The statistical results showed a high correlation and repeatability between NC severity levels and clinical features. Apart from clinical research, researchers also have developed artificial intelligence (AI) techniques for automatic NC classification on AS-OCT images. [3] proposes a CNN model named GraNet for AS-OCT image-based NC classification, but they achieved poor performance.

Recently, attention mechanisms [4, 5] have demonstrated that they are able to improve the performance of deep networks in various fields, e.g., computer vision. The intuition behind learning attention weights is to make the net-

\*Corresponding author

work know where to pay attention. Squeeze-and-excitation (SE) [6] attention is one of the most prominent attention methods. It builds channel dependencies among channels and presents promising performance by increasing negligible cost. SE is succeeded by the convolutional block attention module (CBAM) [7], which combines channel attention with spatial attention for further enhancing the classification results. It utilizes global average pooling (GAP) and global max pooling (GMP) methods to produce global and local statistics information from both channel and spatial perspectives. Interestingly, statistics-wise information in CBAM can be taken as a representation form for clinical features of NC: *mean and max*. CBAM only treats these two statistics features equally with a shared multi-layer perceptron (MLP); however, these features have different correlation coefficients with NC severity levels, suggesting they make different contributions to NC diagnosis. Unlike aforementioned methods achieved better performance by constructing complex models but did not consider clinical prior knowledge infusion. This paper questions whether *we can incorporate clinical prior knowledge into attention design to boost NC classification results in a more efficient manner?*

To answer this question, we propose a channel-wise and spatial feature recalibration network for automatic NC classification, namely CSFR-Net, which not only adaptively learns to reweigh local features but also recalibrates the relative importance of channels. In the CSFR-Net, we construct an effective channel-wise and spatial feature recalibration (CSFR) module, consisting of a channel-wise feature recalibration attention block and a spatial feature recalibration block for focusing on meaningful feature representations in a global-local way. To demonstrate the effectiveness of our CSFR-Net, we conduct comprehensive experiments on a clinical AS-OCT image dataset and CIFAR benchmarks. The experimental results show our CSFR-Net achieves better performance than state-of-the-art methods with lower model complexity, e.g., CSFR-Net obtains **2.39%** gain of accuracy compared with CBAM.

## 2. RELATED WORK

**AS-OCT based ocular disease diagnosis.** Over the past years, clinicians and researchers have increasingly used AS-OCT images for ocular disease diagnosis and scientific research purposes. [8] proposes a deep segmentation network named CorneaNet to crop the cornea structure automatically from the anterior segment structure based on AS-OCT images. Wong et al. [9] had studied the correlation between the opacity of NC and image features on AS-OCT images, and the statistical results indicated that there is a strong correlation between them. [10] also studies the correlations between NC severity levels and image features like mean and max, and they got similar correlation results. Motivated by clinical research, Zhang et al. [11, 12] proposed a machine learning-

based NC classification framework based on AS-OCT images. Xiao et al. [13] proposed a gated channel attention network (GCA-Net) for NC classification and achieved good performance.

**Attention mechanism.** It is well-known that humans selectively concentrate on parts of the given inputs for grasping important information rather than processing the complete input information. Over the years, researchers have widely incorporated this attention mechanism into existing CNN architectures and verified its effectiveness in a wide range of learning tasks such as image classification, text classification, image segmentation, etc. Guo et al. [5] proposed a spatial pyramid attention network (SPANet) for image classification by combining structural information with the pyramid pooling method. Hu et al. [14] constructed a channel-wise and spatial feature modulation network for single image super-resolution. SE attention block was introduced to build the dependencies among channels. CBAM used channel-wise attention and a spatial attention block to emphasize/ suppress feature representation with spatial and channel axes more related to this paper. In each attention block, they use both GAP and GMP to extract spatial statistics and channel-wise features. In concurrent with CBAM, Park et al. [15] presented a bottleneck attention module (BAM) to make CNNs focus on salient feature representation through the channel and spatial pathways concurrently. Unlike CBAM, we reformulate both channel-wise and spatial recalibration in terms of infusing clinical priors without using both channel and spatial relationships and lowering model complexity.

## 3. METHOD

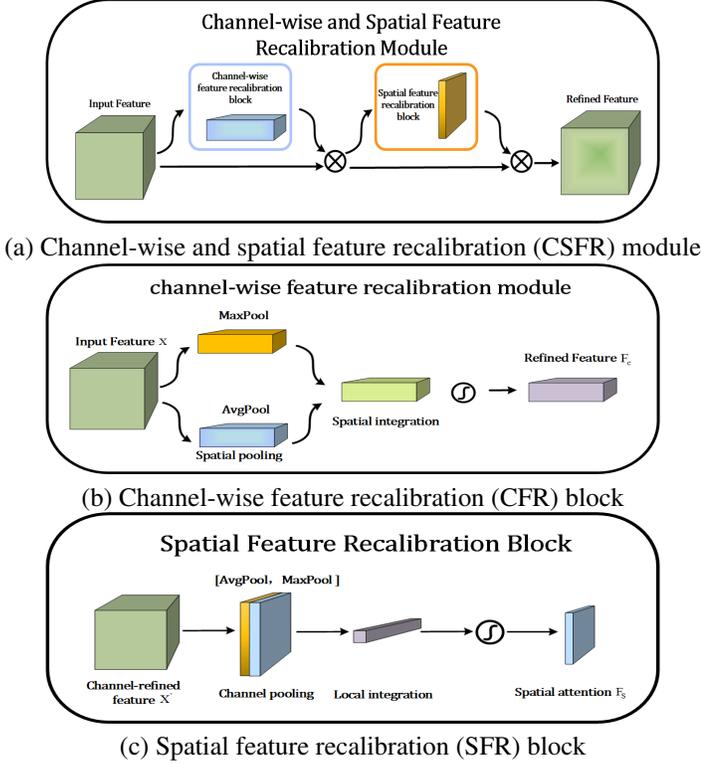
### 3.1. Channel-wise and spatial feature recalibration module

Fig. 1(a) shows the overall framework of our channel-wise and spatial feature recalibration (CSFR) module, which consists of two sequential feature recalibration blocks: a channel-wise feature recalibration (CFR) block and a spatial feature recalibration (SFR) block. Given the intermediate feature maps  $X \in R^{N \times C \times H \times W}$ , where  $N$ ,  $C$ ,  $H$ , and  $W$  represent the number of images in each batch size, the number of channels, the height, and width of the feature map, respectively. CSFR first generates 1D channel attention map  $F_c \in R^{N \times C \times 1 \times 1}$  then produces a 2D spatial attention map  $F_s \in R^{N \times 1 \times H \times W}$  sequentially. The above processes can be expressed:

$$X' = F_c \otimes X, \quad (1)$$

$$X'' = F_s \otimes X', \quad (2)$$

where  $\otimes$ ,  $X'$ , and  $X''$  represent element-wise multiplication operation, the refined feature map generated by CFR, and the final refined feature map. We introduce two attention blocks detailedly in the following.



**Fig. 1.** The overview of channel-wise and spatial feature recalibration (CSFR) module. The module consists of two attention blocks: channel-wise feature recalibration block (CFR) and a spatial feature recalibration (SFR) block. The intermediate feature maps is adaptively recalibrated with the CSFR module.

**Channel-wise feature recalibration block.** Clinical NC research has shown that mean and max features had strong correlations with NC severity levels, but the effects of two features for NC diagnosis are different. Moreover, these features can be viewed as other forms of CNN representations. Motivated by the relationship between the clinical features and CNN representation, this paper introduces a channel-wise feature recalibration block (CFR) to leverage the potential of clinical priors for enhancing performance by emphasizing/inhibiting channels. It consists of two main components as shown in Fig. 1(b): *spatial pooling* for extracting two spatial statistics features from feature maps with GAP and GMP as global-clinical features, and *spatial integration* for producing channel-independent weights adaptively by exploiting spatial statistics with the channel-wise operation. The overall process can be obtained in the following:

$$T = [GAP(X), GMP(X)] \in R^{N \times C \times 2}, \quad (3)$$

$$F_c = \sigma(BN(w_c \cdot t_{nc})), \quad (4)$$

where  $t_{nc} \in R^2$  represents the extracted spatial statistics information for instance  $n$  and channel  $c$ ,  $w_c$  represents the

learnable parameters for channel  $c$ , BN represents batch normalization layer, and  $\sigma$  denotes sigmoid function.  $w_c \cdot t_{nc}$  represents channel-wise operation, which can be viewed as a channel-independent connection layer with two nodes as inputs and one individual node as the output.

**Spatial feature recalibration block.** Following the CFR block, we utilize a spatial feature recalibration (CSR) block to emphasize/suppress local features, which is a complementary to the CFR. CSR also has two main components (as shown in 1(c): *channel pooling and local integration*. We use the channel pooling operator to extract local statistics features—mean and max for all feature maps along with channel axis as local clinical features. This paper implements them with channel average pooling and channel max pooling methods for producing two feature maps:  $M_{avg} \in R^{1 \times H \times W}$  and  $M_{max} \in R^{1 \times H \times W}$ . Specifically, this paper transforms these two 2D maps into two 1D maps:  $M'_{avg} \in R^{HW \times 1}$  and  $M'_{max} \in R^{HW \times 1}$  and construct a local feature vector, which can be written as follows:

$$M_s = [M'_{avg}, M'_{max}], \quad (5)$$

like CFR, this paper uses a local integration operator to adjust the relative importance of features as to emphasize/suppress every local feature, which can be obtained:

$$z_{ns} = \sigma(BN(w_s \cdot M_s)), \quad (6)$$

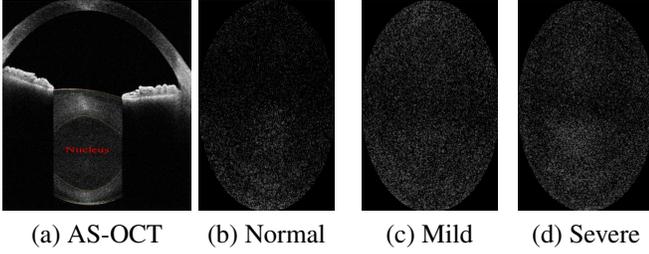
where  $Z \in R^{N \times HW \times 1}$  and  $w_s \in R^{HW \times 2}$  indicate the encoded local feature representations and learnable weight parameters. Finally, we obtain a 2D spatial attention map by reshaping  $z_s$  into  $F_s \in R^{N \times 1 \times H \times W}$  and get the final refined output representations with Eq.2.

### 3.2. Implementation

This paper aims to incorporate clinical prior knowledge into attention design for augmenting the feature representations of CNNs. We take two ResNets as backbones: ResNet18 and ResNet34, to demonstrate the advantages of our CFSR over advanced attention methods. We plug a CSFR module into a residual block termed Residual-CSFR module, and CSFR-Net is built on a stack of Residual-CSFR modules.

## 4. DATASET AND EXPERIMENT SETTINGS

**Clinical AS-OCT dataset.** We collected a clinical AS-OCT image dataset using a CASIA2 ophthalmology device (TOMEY Corporation, Japan), namely CASIA2 AS-OCT. The dataset comprises 422 right eyes and 440 left eyes from 543 participants—available images for 16,201: 1603 normal, 4,842 mild NC, and 9,756 severe NC. We use a deep segmentation network to crop the nucleus region from the anterior segment structure based on AS-OCT images. Since lacking



**Fig. 2.** Three severity levels of nuclear cataract (NC) based on AS-OCT images(a). Normal (b) without opacity; Mild NC (c) with slight opacity but is asymptomatic; Severe NC (d) with opacity and is symptomatic.

**Table 1.** NC severity level distribution on CISIA AS-OCT dataset.

	Normal	Mild	Severe
Training	896	3219	5504
Validation	317	793	2331
Testing	390	830	1921
Total	1603	4842	9756

a standard AS-OCT image-based cataract classification system, we map labels of AS-OCT images from slit lamp images. Three experienced ophthalmologists labeled slit lamp images based on LOCS III, confirming label quality. Fig. 3 provides three examples of NC severity levels on AS-OCT images. We also split AS-OCT images into three disjoint datasets at the participant level: training, validation, and testing, which confirms AS-OCT images of each participant fall into the same sub-datasets, as shown in Table 1. The original sizes of images are various, and this paper resizes them into  $224 \times 224$  as the inputs for deep networks.

**Evaluation metrics.** This paper uses the following measures to assess the overall performance of methods: accuracy (ACC), precision (PR), sensitivity (Sen), F1 measure, and kappa coefficient value.

**Experiment settings.** We implement CSFR-Net and competitive methods with the Pytorch tool and train all models using a stochastic gradient descent (SGD) optimizer with the default setting. Epochs and batch size are set to 150 and 32 accordingly. We set the initial learning rate (lr) to 0.025 and decrease it by a factor of 5 every 20 epochs. We set a fixed lr value to 0.00035 for all models when training epochs over 100. In the training, we follow standard data augmentation methods like the random flipping and the random cropping methods for training images. The practical mean channel subtraction is used to normalize for training, validation, and testing datasets. We run all methods on a workstation with an NVIDIA TITAN V (11GB RAM) GPU.

**Table 2.** Performance comparison and complexity comparison of CSFR and state-of-the-art attention methods on the CISIA2 AS-OCT dataset when taking ResNet18 and ResNet34.

Method	Backbone	ACC	F1	Kappa	Params
ResNet18 [16]	ResNet18	90.86	91.73	83.59	11.18M
SE [6]		91.05	89.30	83.72	11.27M
CBAM [7]		91.50	92.05	84.19	11.27M
BAM [15]		92.10	92.59	85.32	11.20M
ECA [4]		90.64	84.37	82.38	11.18M
CSFR		<b>93.89</b>	<b>93.64</b>	<b>88.86</b>	11.20M
ResNet34 [16]	ResNet34	91.95	90.54	85.01	21.29M
SE [6]		91.40	91.33	84.35	21.44M
CBAM [7]		92.10	90.82	85.09	21.45M
BAM [15]		93.35	93.33	87.29	21.31M
ECA [4]		91.25	91.58	83.62	21.29M
CSFR		<b>94.62</b>	<b>94.48</b>	<b>90.00</b>	21.32M

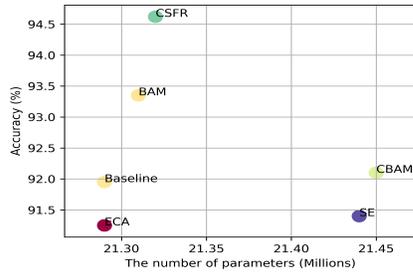
## 5. RESULTS AND DISCUSSION

### 5.1. Comparisons with state-of-the-art attention methods

We compare our method with the other four state-of-the-art attention methods on CISIA2 AS-OCT dataset by using ResNets (ResNet18 and ResNet34) as backbone architectures, and Table 2 summarizes the classification results. Our CSFR continuously improves the classification performance over strong attention methods by using similar model complexity. Specifically, CSFR outperforms CBAM by above **4.67%** in the kappa and **2.39%** in the accuracy, although CBAM uses more parameters than CSFR. Compared to BAM, CSFR achieves over **2.6%** gain of kappa value by using competitive parameters, which confirms the superiority of our method. Fig. 3 plots a comparison of our CSFR and comparable attention methods in terms of accuracy and the number of model parameters. It can be seen that our method obtains a better trade-off between accuracy and the number of parameters, demonstrating it is more efficient in utilizing parameters through comparisons to other attention methods and baselines.

### 5.2. Comparisons with strong baselines

Table 3 provides the NC classification results of our CSFR-Net and strong baselines, including classical machine learning methods and advanced CNN models. Following [11], we also extract 18 features from the nucleus region on AS-OCT images, then apply five commonly-used machine learning methods to predict NC severity levels based on extracted features: support vector machine (SVM), random forest (RF), and GradientBoosting. It can be observed that our method attains the best results on five evaluation measures among all methods. Remarkably, CSFR-Net outperforms machine learning methods with above **4.78%** in the accuracy,



**Fig. 3.** Relationship between the performance of CSFR and the model complexity of it, compared to state-of-the-art attention methods (baseline indicates ResNet34).

**Table 3.** NC classification result comparison of our CSFR-Net and strong baselines.

Methods	ACC	F1	PR	Sen	Kappa
RF	85.45	87.10	86.80	87.53	73.49
SVM	89.78	91.29	90.47	93.22	81.95
GradientBoosting	86.88	87.72	88.09	88.05	76.38
GraNet	90.48	90.72	91.61	89.91	82.15
EfficientNet [17]	91.50	91.38	91.71	91.11	84.31
ResNet34	91.95	90.54	92.51	89.13	85.01
GCA-Net	93.19	93.48	93.99	93.01	87.30
CBAM	92.10	90.82	93.18	88.87	85.09
BAM	93.35	93.33	93.58	91.68	87.29
ECA	91.25	91.58	92.28	90.94	83.62
SENet	91.40	91.33	90.11	92.69	84.35
CSFR-Net	<b>94.62</b>	<b>94.48</b>	<b>95.07</b>	<b>93.93</b>	<b>90.00</b>

**3.05%** in the F1, and **7.71%** in the kappa. respectively. It also achieves better performance than advanced attention-based CNNs and original CNNs, e.g., CSFR-Net obtains over **2.07%** improvement of accuracy than SGENet. Overall, the results demonstrate the effectiveness of incorporating clinical prior knowledge into attention design for NC classification in a global-local way.

### 5.3. Results of attention block arrangements

In this section, we empirically validate the effectiveness of our design choice and adopt ResNet34 as the backbone architecture according to classification results in Table 1. Table 4 provides the classification performance of four different ways of arranging attention submodules: CFR-spatial, channel-SFR, channel-spatial (employed in CBAM), and CFR-CSR (ours). Since the attention methods in each module are different, the combination and order may affect the NC classification performance. We can observe that all arranging attention methods outperform using only single attention methods, verifying that using both attention methods surely improve the NC classification results. Moreover, CFR-SFR achieves

**Table 4.** Comparison of different attention combinations when taking ResNet34.

Design	ACC	F1	Kappa
Channel (SE)	91.40	91.33	84.35
Spatial	91.25	90.11	83.93
CFR	92.00	90.43	84.66
SFR	91.44	90.10	84.25
CFR+spatial	93.16	92.65	87.54
channel+SFR	93.57	93.18	88.17
channel+spatial (CBAM)	92.10	90.80	85.09
CFR+SFR (CSFR)	<b>94.62</b>	<b>94.48</b>	<b>90.00</b>

better performance than other three arranging attention methods, showing the effectiveness of the proposed two attention methods in CSFR module.

### 5.4. Validation on CIFAR benchmarks

To further evaluate the general performance of our CSFR-Net, we conduct a series of experiments on two CIFAR benchmarks [18]. These two datasets consist of 50,000 training and 10,000 test images correspondingly, with  $32 \times 32$  pixel each. For data augmentation, this paper follows the standard practice [5] and pads each image by 4 pixels with value zero, then randomly crops the padded image back to the original image size. The model evaluation is performed on original images. Additionally, we use mean channel subtraction to normalize input data for facilitating training. We train networks with SGD for 200 epochs by setting a batch size to 128. The initial learning rate is set to 0.1, which is divided by a factor of 10 every 40 epochs. As presented in Table 5, CSFR-Net consistently improves performance on CIFAR benchmarks through comparisons to strong attention-based networks and two baselines (ResNet18 and ResNet50). Remarkably, our CSFR-Nets outperforms SPANet with over **3%** while reducing more than half number of parameters on the CIFAR 100 dataset. Compared with CBAM and BAM, CSFR-Net obtains about 1% gain by using slightly fewer parameters on the CIFAR 100 dataset. The results of Table 5 suggest the effectiveness of CSFR-Net is not constrained to the CASIA2 AS-OCT dataset.

## 6. CONCLUSIONS

This paper presents a channel-wise and spatial feature recalibration network (CSFR-Net) by introducing clinical prior knowledge for NC classification. In the CSFR-Net, we construct a lightweight yet efficient channel-wise and spatial feature recalibration module to adjust intermediate feature maps in a global-local manner. Experiments on the clinical AS-OCT dataset and CIFAR datasets demonstrate the effectiveness of our method. It keeps a better trade-off between the

**Table 5.** Accuracy on CIFAR benchmarks with ResNet18 and ResNet50 as backbones and complexity comparison.

Method	Backbone	CIFAR-10		CIFAR-100	
		ACC	Params	ACC	Params
Baseline	ResNet18	93.02	11.17M	74.56	11.22M
SENet [6]		94.84	11.27M	75.19	11.32M
BAM [15]		95.20	11.20M	78.09	11.24M
CBAM [7]		95.19	11.26M	77.82	11.31M
SPANet [5]		95.00	12.13M	75.56	12.18M
ECA [4]		93.12	11.18M	74.43	11.23M
CSFR-Net		<b>95.54</b>	11.18M	<b>78.94</b>	11.23M
Baseline	ResNet50	93.62	23.52M	78.51	23.71M
SENet [6]		95.35	26.06M	79.28	26.64M
BAM [15]		95.54	23.88M	80.00	24.06M
CBAM [7]		95.70	26.05M	80.13	26.24M
SPANet [5]		95.63	51.17M	78.21	51.36M
ECA [4]		94.00	23.53M	78.07	23.71M
CSFR-Net		<b>95.86</b>	23.56M	<b>81.32</b>	23.75M

performance and model complexity through comparisons to state-of-the-art methods.

## Acknowledgements

This work was supported part by Guangdong Provincial Department of Education (2020ZDZX3043, SJJG202002), and Guangdong Provincial Key Laboratory (2020B121201001).

## 7. REFERENCES

- [1] World Health Organization et al., “World report on vision,” 2019.
- [2] Mesut Ozgokce, Muhammed Batur, Muhammed Alpaslan, and Alpaslan Yavuz, “A comparative evaluation of cataract classifications based on shear-wave elastography and b-mode ultrasound findings,” *Journal of ultrasound*, vol. 22, no. 4, pp. 447–452, 2019.
- [3] Xiaoqing Zhang, Zunjie Xiao, Higashita Risa, Wan Chen, Jin Yuan, Jiansheng Fang, Yan Hu, and Jiang Liu, “A novel deep learning method for nuclear cataract classification based on anterior segment optical coherence tomography images,” in *SMC. IEEE. IEEE*, 2020.
- [4] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu, “Eca-net: efficient channel attention for deep convolutional neural networks, 2020 ieee,” in *CVPR. IEEE*, 2020.
- [5] Jingda Guo, Xu Ma, Andrew Sansom, Mara McGuire, Andrew Kalaani, Qi Chen, Sihai Tang, Qing Yang, and Song Fu, “Spanet: Spatial pyramid attention network for enhanced image recognition,” in *ICME. IEEE*, 2020, pp. 1–6.
- [6] Jie, Hu, Li, Shen, Samuel, Albanie, Gang, Sun, Enhua, and Wu, “Squeeze-and-excitation networks,” *TPAMI*, 2019.
- [7] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, “Cbam: Convolutional block attention module,” in *ECCV*, 2018, pp. 3–19.
- [8] Valentin Aranha Dos Santos, Leopold Schmetterer, Hannes Stegmann, Martin Pfister, and Alina Messner, “Corneanet: fast segmentation of cornea oct scans of healthy and keratoconic eyes using deep learning,” *BOE*, vol. 10, no. 2, pp. 622–641, 2019.
- [9] A L Wong, C K-S Leung, R N Weinreb, A K C Cheng, C Y L Cheung, and Lam, “Quantitative assessment of lens opacities with anterior segment optical coherence tomography,” *British Journal of Ophthalmology*, vol. 93, no. 1, pp. 61–65, 2009.
- [10] Natalia Y Makhotkina, Tos TJM Berendschot, Frank JHM van den Biggelaar, Alexander RH Weik, and Rudy MMA Nuijts, “Comparability of subjective and objective measurements of nuclear density in cataract patients,” *Acta ophthalmologica*, vol. 96, no. 4, pp. 356–363, 2018.
- [11] Xiaoqing Zhang, Jiansheng Fang, Zunjie Xiao, Higashita Risa, Wan Chen, Jin Yuan, and Jiang Liu, “Research on classification algorithms of nuclear cataract based on anterior segment coherence tomography image,” *Computer Science*, 2021.
- [12] Xiaoqing Zhang, Zunjie Xiao, Risa Higashita, Yan Hu, Wan Chen, Jin Yuan, and Jiang Liu, “Adaptive feature squeeze network for nuclear cataract classification in as-oct image,” *Journal of Biomedical Informatics*, p. 104037, 2022.
- [13] Zunjie Xiao, Xiaoqing Zhang, and Higashita Risa, “Gated channel attention network for cataract classification on as-oct image,” 2020.
- [14] Yanting Hu, Jie Li, Yuanfei Huang, and Xinbo Gao, “Channel-wise and spatial feature modulation network for single image super-resolution,” *TCSVT*, vol. 30, no. 11, pp. 3911–3927, 2019.
- [15] Jongchan Park, Sanghyun Woo, Joon-Young Lee, and In So Kweon, “A simple and light-weight attention module for convolutional neural networks,” *IJCV*, vol. 128, no. 4, pp. 783–798, 2020.
- [16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” in *CVPR*, 2016, pp. 770–778.
- [17] Mingxing Tan and Quoc Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *ICML. PMLR*, 2019, pp. 6105–6114.
- [18] Alex Krizhevsky, Geoffrey Hinton, et al., “Learning multiple layers of features from tiny images,” 2009.