Knowledge-Based Systems 250 (2022) 109109

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

CCA-Net: Clinical-awareness attention network for nuclear cataract classification in AS-OCT

Xiaoqing Zhang [a,b,*], Zunjie Xiao [b], Lingxi Hu [b], Gelei Xu [b], Risa Higashita [b,c,*], Wan Chen [d], Jin Yuan [d], Jiang Liu [a,b,e,f,*]

[a] Research Institute of Trustworthy Autonomous Systems, Southern University of Science and Technology, Shenzhen, 518055, China
[b] Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, 518055, China
[c] Tomey Corporation, Nagoya, 4510051, Japan
[d] State Key Laboratory of Ophthalmology, Sun Yat-sen University, 510060, Guangzhou, China
[e] Cixi Institute of Biomedical Engineering, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, 315201, China
[f] Guangdong Provincial Key Laboratory of Brain-inspired Intelligent Computation, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, 518055, China

ARTICLE INFO

Article history:
Received 15 September 2021
Received in revised form 21 December 2021
Accepted 19 May 2022
Available online 27 May 2022

Keywords:
Nuclear cataract classification
Clinical-awareness attention
AS-OCT image
Interpretation

ABSTRACT

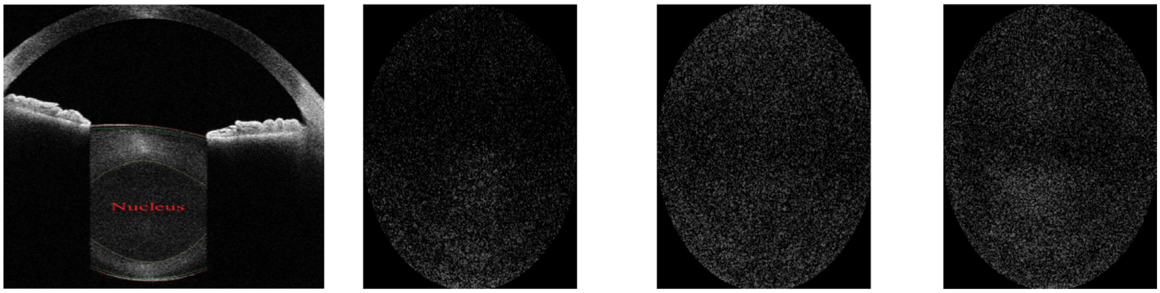
Nuclear cataract (NC) is the leading cause of vision impairment and blindness globally. NC patients can slow the opacity development with early intervention or recover vision with cataract surgery. Anterior segment optical coherence tomography (AS-OCT) images have been increasingly used for clinical NC diagnosis. Compared with other ophthalmic images, e.g., slit lamp images, AS-OCT images are vital for NC diagnosis due to their capability of clearly capturing the nucleus region. Moreover, clinical research has shown the high correlation and repeatability between NC severity levels and image features like mean, maximum, and standard deviation on AS-OCT images. This paper aims to incorporate the clinical features into convolutional neural networks (CNNs) to improve NC classification results and enhance the interpretation of the decision process. Thus, we propose a novel clinical-awareness attention network (CCA-Net) to classify NC severity levels automatically. In CCA-Net, we design a practical yet effective clinical-aware attention block, which not only uses the mixed pooling operator to extract clinical features from each channel but also applies the designed clinical integration operator to focus on salient channels. We conduct extensive experiments on one clinical AS-OCT image dataset and two publicly available ophthalmology datasets. The results demonstrate that the CCA-Net outperforms state-of-the-art attention-based CNNs and strong baselines. Moreover, we also provide in-depth analysis to explain the internal behaviors of our method, enhancing the interpretation ability of our method.

1. Introduction

Cataract is the leading ocular disease for blindness and visual impairment in the world. According to the World report on vision of World Health Organization (WHO) in 2019 [1], approximately 65.4 million patients suffer from moderate and severe vision impairment due to cataract. Moreover, the number of cataract patients would increase rapidly with the aging global population [2], since people over 60 have an 80% probability of cataract. Cataract surgery and early intervention are commonly used methods to improve the vision and life quality of patients [3]. According to the opacity location of cataract, it can be generally grouped into three types: nuclear cataract (NC), cortical cataract (CC), and posterior subcapsular cataract (PSC) [4]. NC is the commonest cataract type as well as an age-related ocular disease. Its clinical symptoms include the gradual clouding and progressive hardening of the nucleus region of the crystalline lens area. According to real clinical diagnosis requirements and opacity development of NC [5], it can be split into three stages based on Lens opacity classification system III (LOCS III) [6,7]: (1) Stage 0: Normal, there is no opacity on the nucleus region. (2) Stage 1: Mild NC (grade $=$ 1 or 2), the nuclear opacity is asymptomatic. (3) Stage 2: Severe NC (grade $\geq$ 3), the nuclear opacity is symptomatic. For patients with mild NC, clinical intervention can slow opacity development. It is necessary and essential for patients with severe NC to undergo cataract surgery and follow-up progress. Clinicians usually diagnose NC severity levels based on their experience and specialized knowledge based on slit-lamp images. However, such

* Corresponding authors.
E-mail addresses: 11930927@mail.sustech.edu.cn (X. Zhang), lisahigashita@gmail.com (R. Higashita), liuj@sustech.edu.cn (J. Liu).

https://doi.org/10.1016/j.knosys.2022.109109
0950-7051/



(a) AS-OCT (b) Normal (c) Mild (d) Severe

Fig. 1. 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.

a diagnosis mode could be error-prone and subjective. Since slit-lamp images have limitations in capturing clear images of the nucleus region, experienced clinicians who can adjust to such a shortage are scarce.

Anterior segment optical coherence tomography imaging (AS-OCT) technique is non-invasive, user-friendly, quick, and high-resolution, as shown in Fig. 1(a). Compared with other ophthalmic images like slit lamp images, AS-OCT images can clearly capture the whole lens area, including nucleus-, cortex-, and capsule- regions. Clinical research [8–11] has studied the correlation between clinical features (like mean, maximum, and standard deviation) from the nucleus region and NC severity levels on AS-OCT images. Such research found a high correlation between NC severity levels and clinical features [8,12] and also verified the repeatability with interclass and intraclass analysis. Fig. 1(b)–(d) provide three representative examples of three NC stages on AS-OCT images: normal (b), mild NC (c), and severe NC (d), which can help audiences understand the opacity information of NC on AS-OCT images easily and quickly.

Channel attention mechanisms have been demonstrated to provide the potential to boost the performance of convolutional neural networks (CNNs) in many learning tasks, e.g., computer vision tasks. One of the most representative examples is Squeeze-and-Excitation (SE) attention method [13]. It extracts channel-wise statistics information of feature maps from each channel with the global average pooling (GAP) operator. The extracted information in the SE block can be considered the mean feature from the nucleus region for NC clinically. Furthermore, other clinical features such as maximum and standard deviation can be viewed as different channel-wise information types, which can be obtained with other pooling operators.

Motivated by the link between channel-wise statistics information and clinical features of the nucleus region on AS-OCT images. This paper aims to fully leverage the potential of clinical prior knowledge to improve the NC classification by infusing them into attention-based CNN architecture design, which has not yet been studied by previous NC classification work. Thus, we propose a novel clinical-awareness attention network (CCA-Net) to predict NC severity levels on AS-OCT images. In the CCA-Net, this paper designs a practical yet effective clinical-awareness attention (CCA) block by infusing the clinical features. The proposed CCA block comprises of two main operators: *mixed pooling* and *clinical integration*. The mixed pooling operator extracts three clinical features from each channel with global average pooling (GAP), global max pooling (GMP), and global standard deviation pooling (GSP) methods, respectively. It is followed by the clinical integration operator, which produces per-channel recalibrated weights via channel-interaction operation and gating operator. The recalibrated weights are used to emphasize important channels and suppress less useful ones from each channel. A clinical AS-OCT image dataset with 16,201 images is collected to verify the effectiveness of our CCA-Net. Additionally, we use two publicly available ophthalmology image datasets to validate the general performance of our method. The results on the clinical AS-OCT image dataset and two public datasets show that our method outperforms strong baselines and previous state-of-the-art methods. Furthermore, we present the comprehensive visualization analysis to explain the inherent behaviors of CCA-Net with feature weight analysis and attention weight analysis, interpreting the decision-making of CCA-Net. We also utilize the class activation mapping (CAM) [14] to explain where and what CCA-Net focuses on through comparisons to strong baselines.

The contributions of this paper are three-fold:

- This paper proposes an attention-based CNN architecture, clinical-awareness attention network (CCA-Net) for automatic NC classification on AS-OCT images. In the CCA-Net, we design a practical yet effective clinical-awareness attention block, which not only extracts clinical features from each channel through the mixed pooling operator but also uses a clinical integration operator to highlight salient channels with produced attention weights.
- The experiment results on the clinical AS-OCT image dataset and two public ophthalmology image datasets demonstrate the superiority of our method through comparisons to strong baselines and previous state-of-the-art methods.
- To our best knowledge, we are the first to conduct a series of visualization analysis and ablation studies to analyze the internal behaviors and validity of our method in-depth by firstly using both feature weight analysis and attention weight analysis, interpreting the decision-making of our method for the NC diagnosis.

The rest of this paper is organized as follows: Section 2 gives a brief survey of this paper, including AS-OCT image-based ocular disease diagnosis and attention mechanisms. We introduce our CCA-Net detailedly in Section 3, comprised of SE attention block, CCA block and its variants, and implementation. In Section 4, datasets and experiment settings are presented. In Sections 5, 6, 7, we discuss the classification results and analyze the inherent behaviors of our method. Finally, we conclude the paper in Section 8.

2. Related work

In this section, we give a brief survey of this paper, comprised of AS-OCT image-based ocular disease diagnosis and attention block design.

2.1. AS-OCT image-based ocular disease diagnosis

AS-OCT is a new OCT technique that can capture an eye's whole anterior chamber structure in a non-invasive, quick, and high-resolution way. Recently, ophthalmologists and researchers have gradually used AS-OCT images for anterior segment ophthalmic disease diagnosis and scientific research purposes such as cornea and glaucoma [15]. [16,17] uses deep CNN models to segment the cornea structure automatically on AS-OCT images, which can be used to assist clinicians in diagnosing cornea diseases precisely and quickly. Fu et al. [18–20] applied AS-OCT images to diagnose angle-closure glaucoma with deep learning models. In addition, researchers also have begun to diagnose cataract by using AS-OCT images [8–10]. Wong et al. [8] studied the correlation between opacity information of NC and mean feature on AS-OCT images. The Spearman correlation coefficient results suggested a strong correlation between them. Wang et al. [12] also studied the correlation relationship between the opacity and mean and maximum features. Other clinical work [9–11] also got similar correlation relationship results by using clinical features like maximum and standard deviation. Motivated by the clinical research, Zhang et al. [21] developed a CNN model named GraNet to classify NC severity levels on AS-OCT images automatically; unfortunately, they obtained poor results. Following [21], [22] extracted image features from AS-OCT images for automatic NC classification and achieved over 75% of accuracy. Moreover, previous work also indicates that there exists a great improvement for AS-OCT image-based NC classification results.

2.2. Attention mechanisms

Attention mechanisms have been extensively studied and widely plugged into modern CNN models for improving the performance on different learning tasks such as classification tasks and segmentation tasks. They can be generally classified into channel attention mechanism and spatial attention mechanism. Typical examples including Squeeze-and-Excitation (SE) attention [13], self-attention [23], and criss-cross attention (CC) [24].

More related to this paper, SE attention method used both squeeze and excitation operators to model interdependencies among channels and generate attention weights for each channel accordingly. It first uses a GAP operator to compute channel-statistics information from each channel and then applies a simple network to adjust the relative weights and generate more informative outputs. Convolutional block attention module (CBAM) [25] and Bottleneck Attention Module (BAM) [26] extended the idea of SE block by further introducing another spatial attention block. Efficient Channel Attention block (ECA) module [27] proposes an adaptive function to learn cross-channel interaction. Style-based recalibration module (SRM) [28] not only introduces global average pooling operator but also utilizes prior knowledge to extract other channel-wise features from feature maps. In contrast to previous work include cataract classification and attention-based CNN architecture design, we incorporate the prior clinical knowledge of NC for into attention block by using three pooling methods to obtain clinical features for the first time, and design a channel-interaction method to model the interdependencies among channels and consider the relative importance of different clinical features. Furthermore, we use the softmax operator instead of the sigmoid operator to set the weights for each channel to highlight significant feature maps with the inter-comparison method.

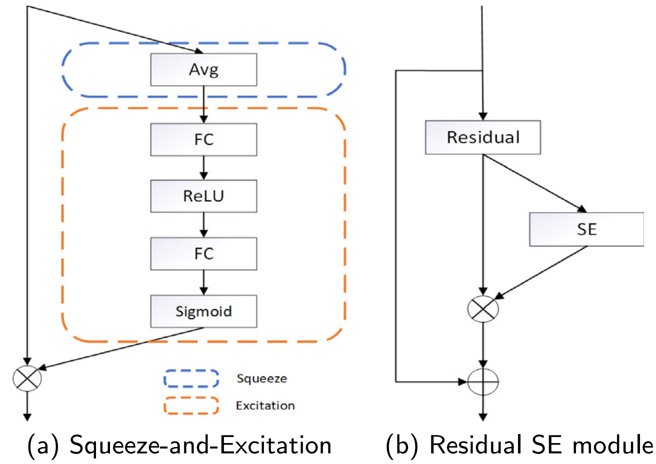


Fig. 2. The schema of (a) Squeeze-and-Excitation (SE) method; (b) Residual SE module: a SE block plugged into a residual block.

3. Methodology

We take the *clinical-awareness attention* (CCA) block as a computational unit that target at strengthening the capability of learned feature representations for CNNs. Given any intermediate feature tensor $X = [x_1, x_2, \ldots, x_C] \in R^{C\times H\times W}$ as the inputs and produces the augmented representation outputs $Y = [y_1, y_2, \ldots, y_C]$, where C is the number of channels; H and W denote the height and width of a feature map. To describe the proposed CCA block clearly, this paper firstly revisits the SE attention block.

3.1. Revisit squeeze-and-excitation attention

The standard convolution method cannot effectively model the inter-dependencies among channels, which place the same weights for each channel [13]. Since the standard convolution method cannot capture global feature representation information, global average pooling (GAP) can compensate for it.

SE attention block (as shown in Fig. 2) is comprised of two operators structurally: squeeze and excitation. It utilizes GAP to encode global representation information from each channel of feature maps while using a learnable network for reconstructing the inter-dependencies of channels dynamically. In the squeeze operator, the channel-wise statistics information z_c for cth channel is obtained by shrinking through spatial dimensions $H \times W$:

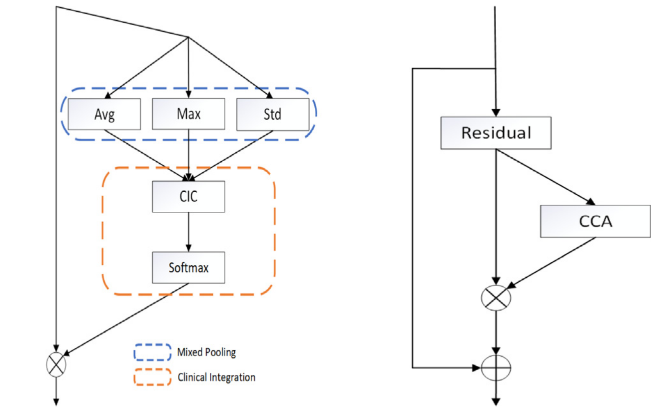
$$z_c = F_{gap}(U_c) = F_{sq}(U_c) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_c(i,j). \tag{1}$$

It is followed by the excitation operator, which uses a small and learnable fully-connected (FC) network to construct the dependencies of inter-channels and then applies a sigmoid operator to emphasize the significant channels, which can be formulated as follows:

$$\hat{X} = X \cdot \sigma(\hat{z}) = X \cdot \sigma(W_2 ReLU(W_1 z)), \tag{2}$$

where σ and ReLU indicate the sigmoid function and ReLU function; W_1 and W_2 indicate the learnable weights in FC layers; $\cdot$ indicates the channel-wise multiplication.

SE attention block has been proven to be an important component of CNNs due to its ability to enhance performance. However, it only considers global average channel-wise statistics



(a) Clinical-Awareness attention (b) Residual CCA

Fig. 3. The schema of clinical-awareness attention (CCA) method (a); Residual CCA module (b): a CCA block plugged into a residual block. **Avg, Max, and Std** indicate global average feature, global maximum feature, and global standard deviation feature.

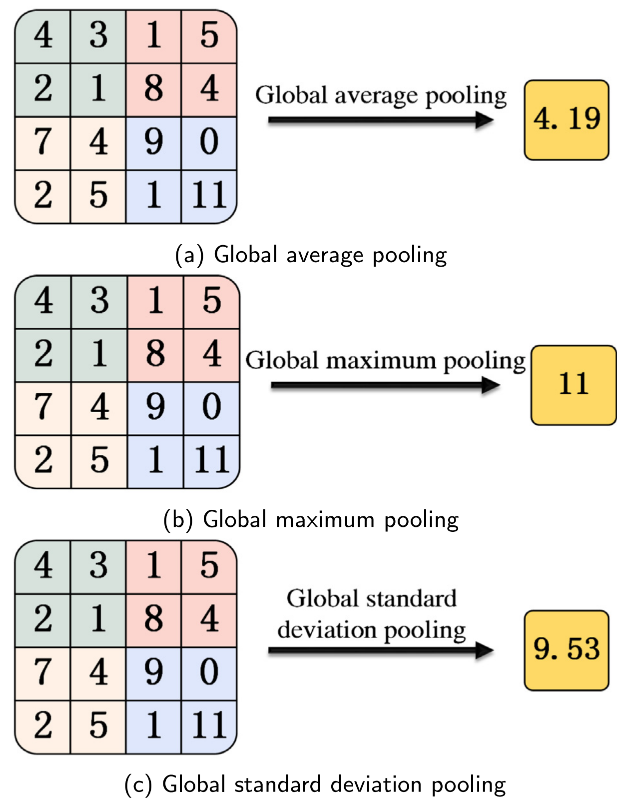


Fig. 4. Toy example comparisons of global average pooling (GAP), global maximum pooling (GMP), and global standard deviation pooling (GSP).

feature while ignoring other global channel-wise statistics features, e.g., maximum. Moreover, it uses the sigmoid operator to emphasize the important channels with the absolute weights but neglects the impacts of relative weights among channels. In contrast to the SE, we introduce a practical yet effective attention block, which takes into account both global feature representation information and the relative importance of channels.

3.2. Clinical-awareness attention blocks

Our clinical-Awareness attention (CCA) block incorporates clinical features into attention block design, consisting of two main operators: mixed pooling and clinical integration. The diagram of the CCA block can be seen in Fig. 3(a). Fig. 3(b) is the Residual CCA module, in which we combine a CCA block with a residual block.

3.2.1. Mixed pooling

The process of extracting global channel-wise statistics features from intermediate feature maps has been extensively studied in CNN and attention design [13,23] to improve the classification performance. In this paper, inspired by clinical research of NC [8,12], we extract three global channel-wise statistics features from each channel of feature maps through global average pooling (GAP), global max pooling (GMP), and global standard deviation pooling (GSP) operators accordingly: **Avg, Max, and Std**, which previous work has not been studied. Fig. 4 provides three examples for three pooling methods, aiming at helping audiences know the difference among these three pooling methods. Specifically, given cth feature map x_c, the mixed features can be obtained with the following equations:

$$\mu_c = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_c(i,j), \tag{3}$$

$$\delta_c = \sqrt{\frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}(x_c(i,j) - \mu_c)^2}, \tag{4}$$

$$Max_c = \max_{(i,j)\in x_c} x_c(i,j), \tag{5}$$

$$t_c = [\mu_c, \delta_c, Max_c], \tag{6}$$

where μ_c, δ, and Max_c indicate the average value (Avg), standard deviation value (Std), and max value (Max); $t_c \in R^3$ is the combination of three mixed pooling features, which concatenates with the channel axis. In Section 6, we will verify the superiority over the proposed mixed pooling compared to other pooling methods, demonstrating the effectiveness of extracted global channel-wise features.

The combination of GAP and GMP operators has been used in CBAM [25] for extracting different global channel-wise statistics features. However, they used a shared fully-connected network to construct the interdependencies of channels by sharing the same weights, which fails to highlight the difference between channel-wise statistics features. To address this problem, this paper proposes a clinical integration method.

3.2.2. Clinical integration

This paper converts clinical features into channel-wise weights with a channel-interaction operator. The weights are purposed to dynamically adjust the relative importance of mixed features and highlight the individual channels simultaneously.

To achieve these, this paper designs a practical *channel-interaction fully connected layer (CIC)* and adopts the softmax function as the gating operator. Taking the generated representation $T \in R^{N\times C\times 3}$ from the mixed pooling operator as the input, the clinical integration operator performs both the inter-channel dependency relationship modeling and channel-wise encoding with learnable weights $W = [w1, w2, \ldots, w_C] =\in R^{C\times C\times 3}$:

$$v_c = w_c t_{nc}, \tag{7}$$

where $\mathbf{V} \in R^{N\times C}$ indicates the learned features, $w_c \in R^{C\times 3}$ are learnable weights for the mixed pooling features from each

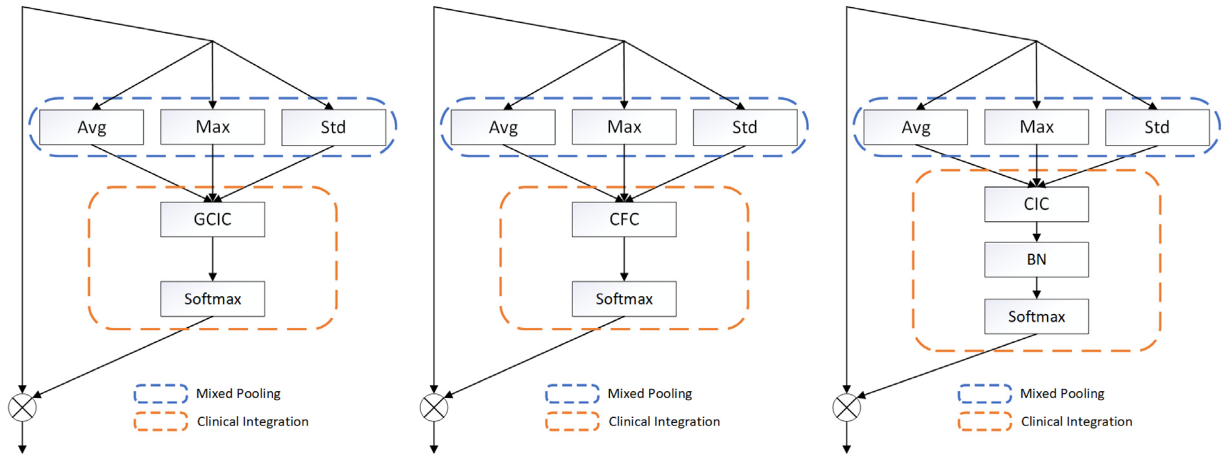


Fig. 5. Two variants of CCA block: CCA-Var1 (a) and CCA-Var2 (b). CCA-Var1: we replace the CIC layer with GCIC layer. CCA-Var2: we replace the CIC layer with CFC layer. CCA-BN (c): we use a batch normalization (BN) after CIC.

channel. This operator can be viewed as a channel-dependent (cross-channel) fully-connected layer with three nodes as inputs and a single node as the output.

It is followed by the softmax function as the gating operator, which is different from prior efforts [13,27,28]. It can be formulated as follows:

$$g_c = Softmax(v_c) = \frac{e^{v_c}}{\sum_{i=1}^{C} e^{v_i}}, \tag{8}$$

where $g_c \in R^{C\times1\times1}$ denotes the generated attention weights for each channel. Compared with the sigmoid operator, the softmax operator has the following advantages: Firstly, it constrains the attention weights and has a sum of 1, facilitating the learning of attention-based CNN models. Secondly, attention weights of other channels determine the particular channel-wise attention weight through softmax function, which can be taken as a weak interaction among channels, which enables neurons to highlight significant channels and suppress unimportant ones. Finally, the original input X is recalibrated by multiplying the attention weights:

$$y_c = x_c \cdot g_c, \tag{9}$$

where y_c is the augmented output. In Section 6.2, we will prove the effectiveness of the softmax operator, comparing with other gating functions such as sigmoid and tanh.

Discussion. To further exploit the effects of our CIC layer for both weighing features and highlight/suppress individual channels. We explore two variants of the CCA block, as shown in Fig. 5: (a) CCA-Var1: we replace the CIC layer with group channel-interaction fully connected layer **(GCIC)**; (b) CCA-Var2: we replace the CIC layer with channel-wise fully connected **(CFC)** layer. We also add a batch normalization (BN) [29] layer after the CIC layer (CCA-BN, as shown in Fig. 5(c)), and the effects is tested and discussed based on performance in Section 6.1.

Fig. 6 shows three examples of CIC, GCIC, and CFC for explaining their differences visually. Fig. 6(a) is an example of the CIC method, which can be viewed as a generalized linear model (GLM). CIC not only adjusts the relative importance of three feature representations but also considers the dependencies among channels. Fig. 6(b) is an example of the GCIC method, which can be taken as a local linear model (LLM). For the GCIC method, we split the number of weight matrices into multiple individual weight matrix groups and split feature representations from the previous layer into individual feature representation groups. The number of matrix groups and feature representation groups is equal, indicating we apply each weight matrix group to the feature representation group. In this paper, we set the number of groups to 8 in the GCIC. Although GCIC adjusts the relative importance of three feature representations, it only constructs the inter-dependencies among channels in each group. Fig. 6(c) is an example of the CFC method and can be considered as an individual linear model. Each weight matrix in the CFC is applied to its particular feature map to generate the learned feature. CFC only considers the relative importance of three feature representations without constructing the interdependencies among channels.

3.3. Implementation

This paper aims to fully leverage the potential of clinical features to improve the NC classification performance of CNN models by infusing advanced attention blocks with clinical prior knowledge. Therefore, this paper takes classic CNN architectures: ResNet18 and ResNet34 as examples to validate the advantages of the proposed clinical-awareness attention block. Fig. 3(b) provides an example of the Residual-CCA module, in which we plug a CCA block into a residual block. Because the residual connection method is skill at alleviating the gradient vanishing problem of deep network [30]. The CCA-Net is a stack of Residual-CCA modules. Softmax function is used as the classifier, which is a commonly used classifier in modern CNNs. We use the cross-entropy loss function as the loss function:

$$Loss_{CE} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log \hat{y}_i, \tag{10}$$

where y_i, $\hat{y}_i$, and N denote the ground truth, predicted labels, and the number of image instances.

4. Dataset and experiment setup

4.1. Dataset

This paper collected a clinical AS-OCT image dataset from a local health physical center through the CASIA2 ophthalmology device (Tomey Corporation, Japan). Fig. 1(a) presents an example of an AS-OCT image for the whole Anterior segment structure of the eye. Only the nucleus region is related to NC (red bold); hence, we use a deep segmentation network to acquire nucleus regions,

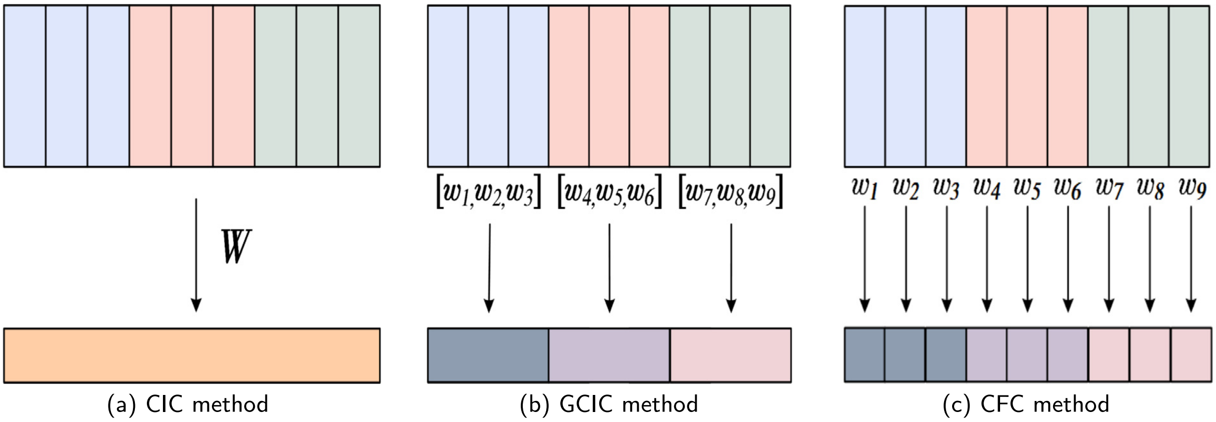


Fig. 6. Three examples of the connected methods: channel-interaction fully connected (CIC) method (a), group channel-interaction fully connected (GCIC) method (b), and channel-wise fully connected (CFC) method (c).

Table 1
The data distribution of the severity levels of NC on AS-OCT image dataset.

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

as shown in Fig. 1(b)–(d). The labels of NC are mapped from slit-lamp images due to the lack of a standard AS-OCT image-based cataract classification system. Three experienced ophthalmologists labeled slit-lamp images, confirming the label quality. The data collection of this paper is conducted according to the tenets of Helsinki Declaration. Because of the retrospective nature and fully anonymized usage of the dataset, we are exempted by the medical ethics committee to inform the patients.

The AS-OCT image dataset contains 543 participants (422 right eyes and 440 left eyes), and the average age is 61.30 $\pm$ 18.65 (range:14–95 years). 24 AS-OCT images are collected for each eye; under the guidance of experienced ophthalmologists, 4,487 images without complete lens regions are removed due to poorly opened eyelids. Thus, the available number of AS-OCT images for this paper is 16,201. We divide the AS-OCT image dataset into three disjoint subsets at the participant level: training, validation, and testing. This is because each participant usually has similar NC severity levels of both eyes. Table 1 shows the distribution of three NC severity levels on the AS-OCT image dataset. The size of the original AS-OCT image is 864 $\times$ 386, and we resize it into 224 $\times$ 224 as the input for the proposed CCA-Net and comparable CNNs.

Furthermore, we use two publicly available ophthalmology image datasets to verify the general performance of our CCA-Net: the ACRIMA dataset and the UCSD dataset. It is a fundus image dataset for glaucoma, comprised of 396 glaucomatous and 309 normal instances. More detailed introduction of the ACRIMA dataset can be found in [31]. We follow the same dataset split method as previous work adopted in the experiments. The UCSD dataset is an OCT image dataset comprised of the training and testing datasets. The training dataset has 108,312 images: 37,206 with choroidal neovascularization (CNV), 11,349 with diabetic macular edema (DME), 8,617 with drusen, and 51,140 normal. The testing dataset has 1000 images, and each category has the same number of images (250 images). The more detailed introduction of the dataset can be seen in [32]. In the experiments, we follow the dataset split method used in [33].

4.2. Metrics

Following cataract classification and glaucoma detection research [34–36], this paper uses five evaluation metrics to assess the classification performance of our method and baselines by following: Accuracy (ACC), precision (PR), specificity (SP), sensitivity (*Se*), F1 score, and kappa coefficient value (*Kappa*). ACC indicates the number of images that are correctly classified. F-score is an essential indicator for evaluating the overall performance of a method. *Se* is a significant evaluation measure for disease diagnosis clinically. Kappa coefficient is a vital measure for assessing diagnostic reliability clinically [34]. The above metrics are formulated as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, \tag{11}$$

$$Se = \frac{TP}{TP + FN}, \tag{12}$$

$$PR = \frac{TP}{TP + FP}, \tag{13}$$

$$SP = \frac{TN}{TN + FP}, \tag{14}$$

$$F1 = \frac{2 * PR * Se}{PR + Se}, \tag{15}$$

where TP, TN, FP, and FN indicate true positive, true negative, false positive, and false negative, respectively.

4.3. Baselines

This paper implements comprehensive experiments to demonstrate the effectiveness of our method in the following.

- **State-of-the-art attention methods.** We use the following channel attention methods for comparison: SE, CBAM [25], (BAM) [26], (SRM) [28], and ECA. These channel attention methods have achieved excellent performance on various classification tasks and can be easily plugged into modern CNNs, like ResNet18 and ResNet34. They also provide state-of-the-art baselines to verify the effectiveness of the proposed CCA and its variants.
- **Strong baselines.** To further demonstrate the superiority of our CCA-Nets, we use the following advanced CNNs and machine learning methods for comparison: (1) state-of-the-art CNN models: ResNet, GraNet, VGGNet, SGENet [37], and EfficientNet [38]. (2) Advanced machine learning methods: According to literature [22,39–41], we extract seventeen

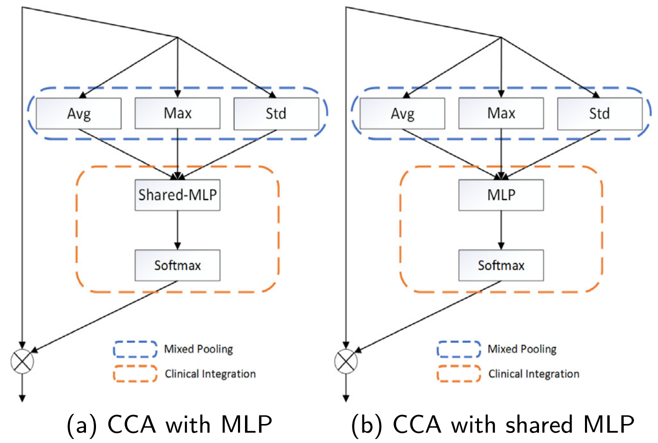


Fig. 7. Replacing the CIC with MLP and shared-MLP.

image features from the nucleus region on the AS-OCT images. Then we use machine learning methods to classify NC severity levels based on extracted features, such as random forest (RF), decision tree (DT), Gaussian Naive Bayes (NB), support vector machine (SVM), multiclass logistic regression (MLR), Adaboost, and GradientBoosting.

- **Comparison of different pooling integration methods.** This paper tests the practical benefits of the proposed CIC method and its variants through comparisons to other two pooling integration methods: a multi-layer perceptron (MLP) network of two fully connected layers (used in SE) and a shared MLP network (employed in CBAM). Fig. 7 presents examples of these two pooling integration methods based on the proposed CCA block. The difference between the MLP and the shared MLP is that: MLP sets different weights for three features while the shared MLP sets the same weights for them. Theoretically, the CIC is equivalent to MLP since both our CIC and MLP use learnable weights to set the relative importance of three feature representations for constructing the channel dependencies. However, MLP only uses a large kernel function to construct channel dependencies globally while ignoring the roles of different feature representations in each channel. This phenomenon has been discussed in detail [42]. Our CIC uses multiple kernel functions to model channels' inter-dependencies in a local–global manner. Our CIC uses multiple kernel functions to model channels' inter-dependencies in a local–global manner. Thus, it considers that different feature representations from each feature map play varying roles in every channel and interrelationship of channels, enabling CCA to focus more on significant channels and suppress less useful ones.
  Comparison of different gating operators. We demonstrate the superiority of softmax operator over other gating operators like sigmoid and sparsemax in Section 6.
- **Comparison of different pooling methods.** We compared our mixed pooling method with other six pooling method to valid the effectiveness of it.

4.4. Experiment setup

All methods, including deep learning methods and machine learning methods are implemented through Pytorch platform, scikit-learn package, and OpenCV. Experiments are run on a workstation with an NVIDIA TITAN V (11 GB RAM) GPU. We use the SGD optimizer with default settings (a momentum of 0.9 and a weight decay of 1e-5) to optimize the network in the training. We set the mini-batch size and training epochs to 16 and 150 for all deep learning methods in the training on the AS-OCT image dataset, respectively. The initial learning rate is set to 0.025 and decreased by a factor of 10 every 20 epochs. We set the fixed learning rate for 0.00025 when the training epochs are over 70. We follow the standard practice for data augmentation and perform the random flipping method (horizontal flipping and vertical flipping) and the random cropping method for the training set. We use the practical mean channel subtraction strategy to normalize the input images for training, validation, and testing. The results reported are based on the best classification results on the testing set.

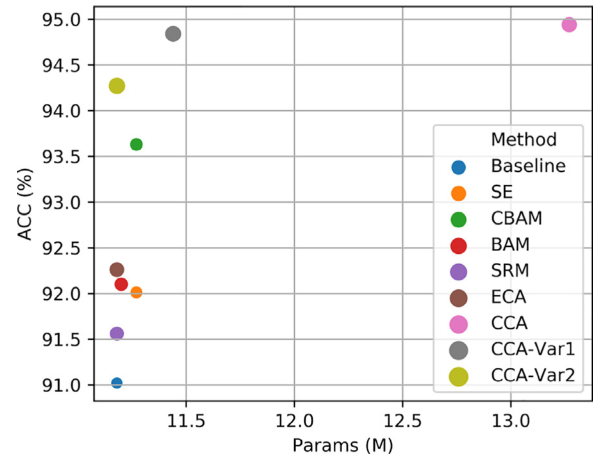


Fig. 8. Relationship between the performance of CCAs and the number of parameters in them, compared with advanced attention methods (baseline indicates ResNet18).

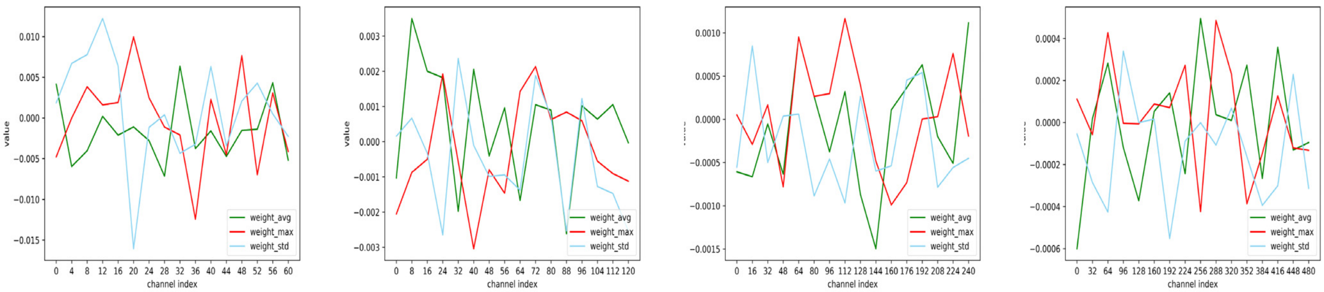
Table 2
Performance comparison and complexity comparison of state-of-the-art attention methods on the test dataset of AS-OCT images when taking ResNet18 and ResNet34. The best results in this table are labeled in **bold**.

Method	Backbone	ACC	F1	Kappa	Params	GFLOPs
ResNet18 [30]		91.02	90.98	83.43	11.18M	1.82
SE [13]		92.01	91.94	85.07	11.27M	1.82
CBAM [25]		93.63	93.15	88.26	11.27M	1.82
BAM [26]		92.10	92.59	85.32	11.20M	1.82
SRM [28]	ResNet18	91.56	91.30	84.75	11.18M	1.82
ECA [27]		92.26	94.95	85.62	11.18M	1.82
CCA		**94.94**	94.66	**90.70**	13.27M	1.82
CCA-Var1		94.84	**94.91**	90.20	11.44M	1.82
CCA-Var2		94.27	94.41	89.34	11.18M	1.82
ResNet [30]		88.57	88.57	78.27	21.29M	3.67
SE [13]		90.58	91.08	82.41	21.44M	3.67
CBAM [25]		92.61	92.63	86.08	21.45M	3.67
BAM [26]		93.35	93.33	87.29	21.31M	3.68
SRM [28]	ResNet34	91.75	91.48	84.23	21.29M	3.67
ECA [27]		91.25	91.58	83.62	21.29M	3.67
CCA		94.87	94.94	90.54	25.06M	3.68
CCA-Var1		94.68	94.67	90.22	21.76M	3.67
CCA-Var2		**95.13**	**95.06**	**90.93**	21.29M	3.67

5. Results and discussion

5.1. Comparison with state-of-the-art attention methods

We compare our CCA block and its two variants with other five advanced attention methods and baselines (ResNet18 and ResNet34), shown in Table 2. The results show that our CCAs consistently improve the NC classification performance on three evaluation measures. Specifically, *the CCA outperforms SRM and*



(a) Conv_2_2 (b) Conv_3_2 (c) Conv_4_2 (d) Conv_5_2

Fig. 9. Weight values of three global channel-wise statistics features along with channel index in three stages with CIC method: Avg (green), Max (red), and Std (blue).

Table 3
Significant analysis of CCA and its variants with with t-test method: CCA-Var1 and CCA-Var2.

Comparison	*p*-value
CCA vs. CCA-Var1	0.0001
CCA vs. CCA-Var2	0.009
CCA-Var1 vs. CCA-Var2	0.0002

ECA over absolute ***3%****, 2% in the accuracy, and obtains over* ***5%*** *in the kappa*, respectively; moreover, it uses the same computation cost as they did. We can also see that compared with SENet, CBAM, and BAM, the CCA uses slightly more parameters and obtains 2% improvement in the kappa and over 1.3% in the accuracy accordingly. Remarkably, two CCA variants also achieve over **94%** of accuracy, and CCA-Var2 gets the best accuracy with **95.13%** on the ResNet34 backbone, which demonstrate the effectiveness of our CCAs by infusing clinical prior knowledge.

Fig. 8 plots the performance relationship between the number of parameters for the CCAs and comparable attention methods on ResNet18. We conclude that the proposed CCAs keep a better balance between performance and model complexity through comparisons to previous attention methods, especially CCA-Var1 and CCA-Var2. Furthermore, these attention methods generally obtain better NC classification results with the ResNet18 than the ResNet34. This is because the available AS-OCT images for training the network are limited. In the following experiments, we use ResNet18 as the baseline to analyze the inherent behaviors of our CCA.

Table 3 presents the statistical significance of the NC classification results on the clinical AS-OCT image dataset for CCA-Net and its two variants (CCA-Var1-Net and CCA-Var2-Net) by using the Student's t-test method [43]. We can find the significant difference in the classification performance of CCA-Net and its variants (p-value <0.05), which indicates that the learned feature representations of these three deep networks are different.

5.2. Interpretation and visualization

5.2.1. Roles of different channel-wise statistics features

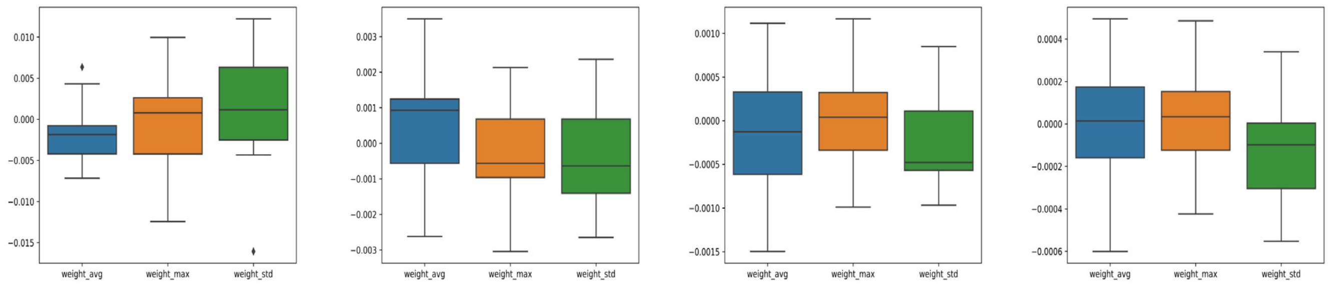
To investigate the roles of three different channel-wise statistics features in the CCA block, we analyze the weights for them in different level stages based on ResNet18: low (ConV_2_2), middle (Conv_3_2) and (Conv_4_2), and high (Conv_5_2). Fig. 9 shows weight distributions of these three features in the CCA block along with the channel index, respectively. The horizontal and vertical directions indicate the channel index and learned feature weights. Green color, red color, and blue color indicate the learned weights for three features: **Avg, Max, and Std**. We can see that averaged weight distributions for three features change with the depth of the network and the channel index. Two following reasons can account for them: (1) Three features play different roles in different stages, e.g., the Std has the higher weight value in the low-level stage while the Max and Avg have higher weight values in high-level stages; (2) different channels have varying effects on attention values, and weight initializations are various due to multiple kernel functions, leading to the difference of weights for three features along with the channel index.

We also calculate the averaged weight values for three channel-wise features in the proposed CIC method at all stages, as shown in Fig. 10. We find one interesting pattern about the roles of three features in the CCA block across the network depth: the Std plays a more important role than the Max and Average in the low-level stage, where Max and Avg have more significant effects in the high-level stage. The fluctuations of weights for Max and Avg are smaller than the weights for Std, which indicates that Max and Avg are more significant than Std for improving the NC classification performance. These results also agree with clinical findings and demonstrate the effectiveness of the proposed method. Figs. 11 and 12 further provide the averaged weight values for the three features in the GCIC and CFC methods. It can be seen that the weight distributions of three features in these two methods are similar to the distributions in Fig. 9, also proving the three features play different roles in the CCA block for improving the classification results.

5.2.2. Attention weight visualization of different attention methods

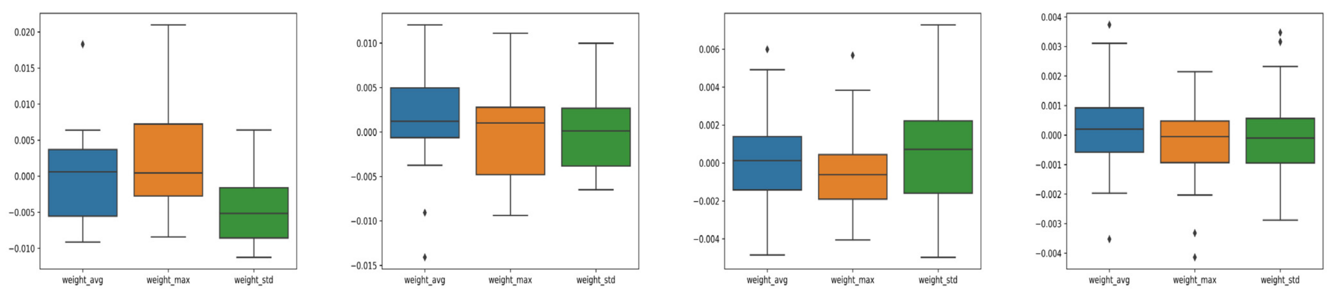
Although the proposed CCA shares similar aspects of feature recalibration with existing channel attention methods, like ECA and SE. However, we observe that the inherent characteristics of CCA are distinct from them based on attention weight analysis. Fig. 13 visualizes the attention weights of our CCA and three contrasted attention methods (SE, SRM, and ECA) for three NC severity levels on testing images at different stages accordingly. It can be seen that the difference of channels in the weight value changes in descending order with the depth of the network. This implies that more channels play essential roles in deep network layers, but there are redundancies among channels, as shown in Fig. 14. Interestingly, we find that the attention weight value fluctuation of our CCA is smaller than the SRM and ECA except for SE, indicating our CCA is easy to train. The fluctuation of attention weights in the SE is smooth, which places the approximate weights for all channels and cannot discriminate the difference between channels. The attention weight values of the CCA are smaller than the other three attention methods because we use the softmax function as the gating operator. All in all, we can conclude that the CCA highlights or suppresses channels more effectively through comparisons to other attention methods by infusing clinical prior knowledge into attention block design.

Fig. 14 further depicts the correlation matrices between channel weights generated by the CCA and the other three attention



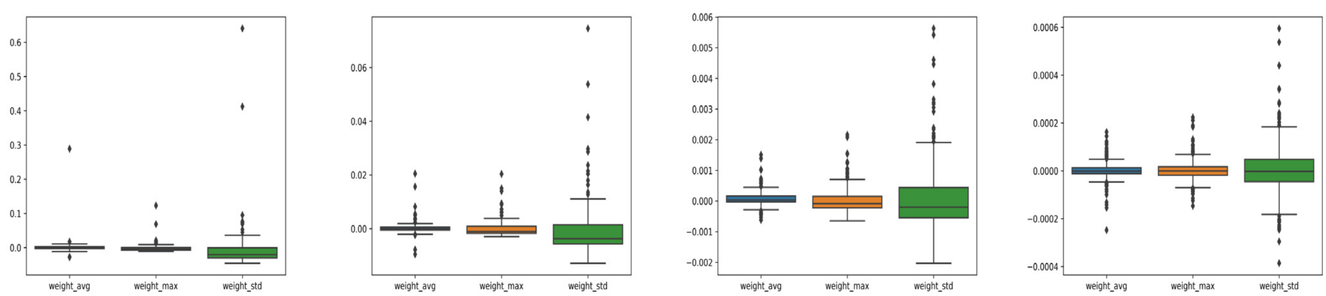
(a) Conv_2_2 (b) Conv_3_2 (c) Conv_4_2 (d) Conv_5_2

Fig. 10. The averaged weight values of three global channel-wise statistics features in all channels at three stages with the CIC method: Avg (blue), Max (orange), and Std (green). Standard deviation results are also plotted.



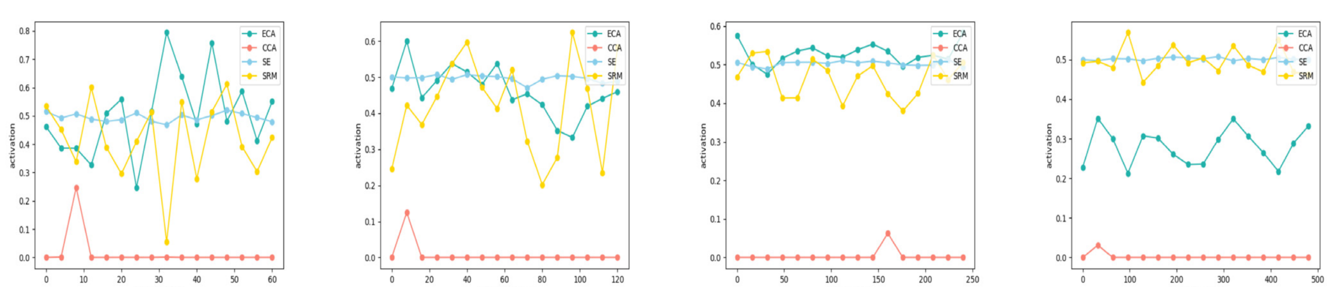
(a) Conv_2_2 (b) Conv_3_2 (c) Conv_4_2 (d) Conv_5_2

Fig. 11. The averaged weight values of three global channel-wise statistics features in all channels at three stages with the GCIC method: Avg (blue), Max (orange), and Std (green). Standard deviation results are also plotted.



(a) Conv_2_2 (b) Conv_3_2 (c) Conv_4_2 (d) Conv_5_2

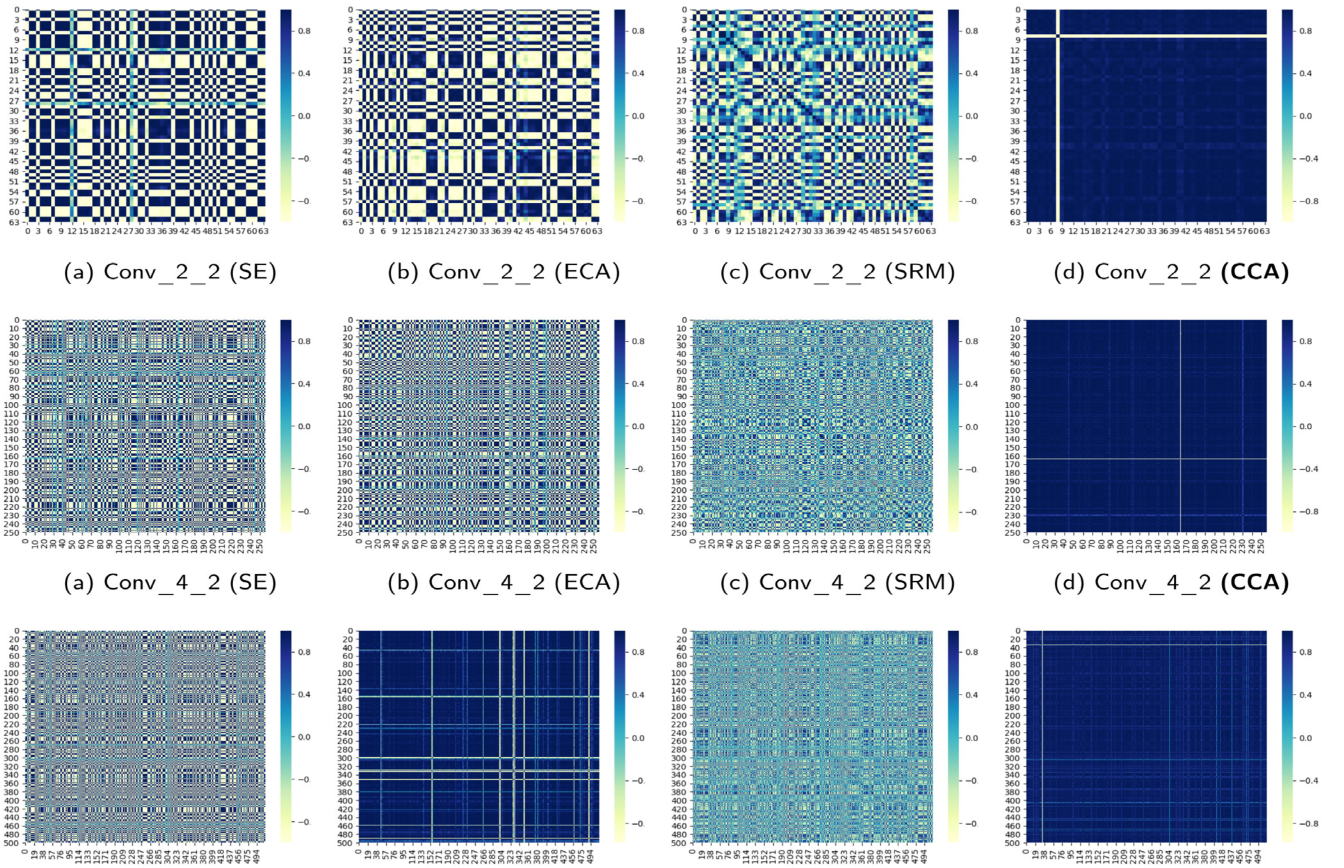
Fig. 12. The averaged weight values of three global channel-wise statistics features in all channels at three stages with the CFC method:Avg (blue), Max (orange), and Std (green). Standard deviation results are also plotted.



(a) Conv_2_2 (b) Conv_3_2 (c) Conv_4_2 (d) Conv_5_2

Fig. 13. The attention weight values of CCA and other three attention methods along with the channel index in three stages.

methods. As we expect, our CCA is different from the other three attention methods. There exist high-positive correlations between most channel weights in three-level stages. The SRM shows the lower correlation between channel weights through comparisons to the CCA and other two attention methods, while ECA exhibits high correlations between channels in the high-level



(a) Conv_2_2 (SE) (b) Conv_2_2 (ECA) (c) Conv_2_2 (SRM) (d) Conv_2_2 **(CCA)**

(a) Conv_4_2 (SE) (b) Conv_4_2 (ECA) (c) Conv_4_2 (SRM) (d) Conv_4_2 **(CCA)**

(a) Conv_5_2 (SE) (b) Conv_5_2 (ECA) (c) Conv_5_2 (SRM) (d) Conv_5_2 **(CCA)**

Fig. 14. Visualization of the correlation matrix between the channel weights. Conv_3_2, Conv_4_2, and Conv_5_2 represent low, middle, and high stages for SE, ECA, SRM, and CCA.

stage. Moreover, the total sum of channel weights in CCA is far smaller than SE, ECA, and SRM, since most channel weights are **close to zero due to the softmax function according to Fig. 13.** *It indicates that the channels with high-positive correlation are relatively useless in the CCA block compared with other three attention methods, proving that our method is more capable of highlighting important channels and inhibiting unimportant channels.*

5.2.3. Visualization of CAM

This paper also uses the CAM method to visualize the heat maps of the proposed CCA and comparable attention methods, as shown in Fig. 15. It provides three representative AS-OCT images of NC severity levels and their heat maps. For normal AS-OCT images, all methods accurately focus on feature representation location; however, our method focuses more on center and bottom regions for mild and severe NC than the other three attention methods. This is because these two regions contain more useful feature representations than other regions, according to the definition of opacity for NC. Compared to other state-of-the-art attention methods like SE, ECA, and SRM, we can conclude that the CCA is more capable of making the network model focus on salient feature representation that locates accurately. It can also explain why our CCA-Net obtains better classification results than advanced attention-based CNNs.

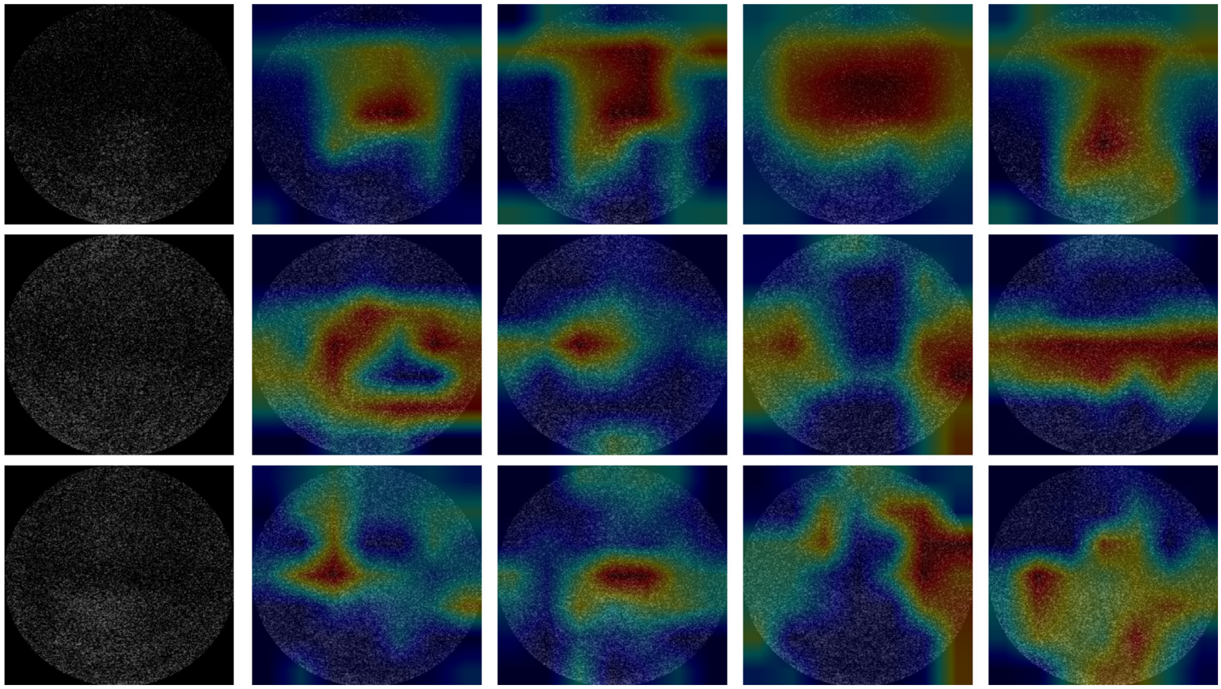
5.3. Comparison with strong baselines

Table 4 shows the NC classification results of our CCA-Nets and strong baselines (the best results of each evaluation measure are **bold**). It can be observed that our CCA-Nets significantly outperform state-of-the-art CNNs and machine learning methods.

CCA-Var2-Net-34 gets the best accuracy with 95.13% and the

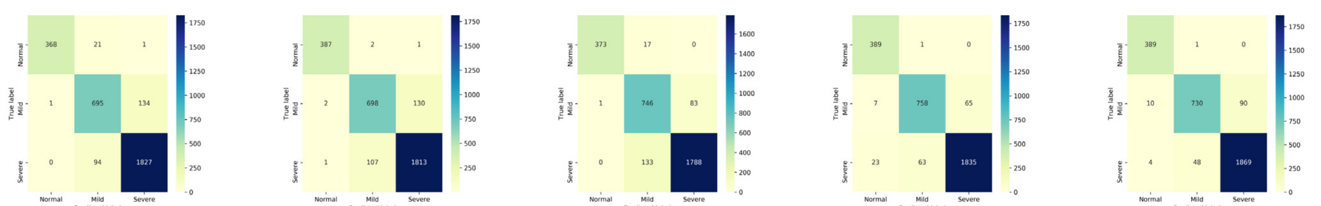
Table 4
NC classification result comparison of our CCA-Net and strong baselines. The best results in this table are labeled in **bold**.

Methods	ACC	F1	PR	Se	Kappa
RF	85.45	87.10	86.80	87.53	73.49
MLR	89.84	91.43	90.70	93.96	82.29
SVM	89.78	91.29	90.47	93.22	81.95
DT	80.70	80.66	82.17	79.34	62.75
NB	74.50	70.29	69.75	74.00	57.32
Adaboost	79.08	75.75	82.84	73.92	62.15
GradientBoosting	86.88	87.72	88.09	88.05	76.38
GraNet [21]	57.85	–	–	–	–
VGGNet19	89.91	91.15	90.48	93.52	82.35
GraNet	90.48	90.72	91.61	89.91	82.15
EfficientNet	91.50	91.38	91.71	91.11	84.31
ResNet18	91.02	90.98	91.35	90.69	83.43
CBAM-ResNet18	93.63	93.15	93.85	92.66	88.26
BAM-ResNet18	92.10	92.59	93.00	92.19	85.32
SRM-ResNet18	91.56	91.30	90.48	92.37	84.75
ECA-ResNet18	92.26	92.78	93.00	92.57	85.62
SENet-18	92.01	91.19	92.88	91.07	85.07
SGENet-18 [37]	92.55	92.80	92.85	92.87	86.34
CCA-Net-18	94.94	94.66	93.88	**95.53**	90.70
CCA-Var1-Net-18	94.84	94.91	**96.69**	93.63	90.20
CCA-Var2-Net-18	94.27	94.41	94.95	93.89	89.34
CCA-Net-34	94.87	94.94	95.13	94.79	90.54
CCA-Var1-Net-34	94.68	94.67	94.72	94.69	90.22
CCA-Var2-Net-34	**95.13**	**95.06**	95.21	95.00	**90.93**



(a) Inputs (b) SE (c) ECA (d) SRM (e) Ours (CCA)

Fig. 15. The CAM visualization of the proposed CCA and other state-of-the-art attention methods. Row 1 to row 3 denote the normal, mild NC, and severe NC.



(a) SENet-18 (b) ECA-ResNet-18 (c) SGENet-18 (d) CCA-Net-18 (e) CCA-Var2-Net-34

Fig. 16. Confusion matrices of CCA-Nets and other three representative attention-based networks.

best F1 with 95.06% accordingly. It exceeds channel attention-based CNNs, e.g., SRM, with over 2%, original CNNs with 3%, and feature extraction-based machine learning methods with over **5%**. Specifically, the CCA-Net-18 achieves the best sensitivity with **95.53%**. It outperforms comparable methods over 2%, indicating that our method is more capable of predicting NC severity levels accurately, which is a clinically significant diagnosis indicator. Our CCA-Nets achieve over 90% in the kappa except for CCA-Var2-Net-18, verifying the high reliability of our methods compared with other methods. As shown in Table 4, CNNs generally obtain better classification results than machine learning methods, which confirms that CNNs methods are more capable of extracting useful feature information than the combination of machine learning methods and feature extraction methods.

To see detailed classification results of our methods and comparable methods, Fig. 16 depicts the confusion matrices of CCA-Net-18, CCA-Var2-Net, SENet-18, ECA-ResNet18, and SGENet-18. It can see that our CCA-Nets achieve the best classification for three NC severity levels through comparisons to other three attention-based CNNs, respectively. However, the classification results of mild NC are worse than normal and severe NC, because the boundaries between them are not clear. Overall, these results demonstrate that CCA-Net's efficiency in classifying NC severity levels arising from prior clinical knowledge.

6. Ablation study

6.1. The impact of pooling integration methods

We test the effect of the proposed CIC method in the CCA block compared with other pooling integration methods. On top of our mixed pooling operator, we compared our CIC operator and its variants with other two pooling integration methods: a multi-layer perceptron (MLP) network of two fully connected layers (used in SE) and a shared MLP network (employed in CBAM), as shown in Fig. 7. Moreover, we also verify the effect of the batch normalization (BN) layer for the mentioned pooling integration methods by adding it after them, e.g., CIC layer with a BN layer, as shown in Fig. 5(c). To implement the MLP and the shared MLP based on mixed pooling, this paper concatenates the mixed pooling features along the channel axis for SE, and sums three mixed pooling features along the channel axis for CBAM. We also use the default configurations of CBAM and SE. We adopt the same experiment setting in Section 4 for all pooling integration methods on the AS-OCT image dataset.

Table 5 reports the classification results of our CIC, its two variants, and contrastive integration methods. The CIC and its variants achieve better classification performance than MLP and shared MLP, highlighting the superiority of the channel-interaction method over constructing the inter-dependencies of channels and

Table 5
Performance comparisons of different pooling integration methods when taking ResNet18. The best results in this table are labeled in **bold**.

Method	ACC	F1	Kappa
MP+MLP	93.03	93.15	87.04
MP+shared-MLP	93.89	94.03	88.61
MP+CIC	**94.94**	**94.66**	**90.70**
MP+GCIC	94.84	94.91	90.20
MP+CFC	94.27	94.41	89.34
MP+MLP+BN	92.93	92.97	86.59
MP+shared-MLP+BN	91.37	90.88	87.56
MP+CIC+BN	93.95	94.16	88.97
MP+GCIC+BN	93.35	93.54	87.86
MP+CFC+BN	92.94	92.74	87.31

Table 6
Comparisons of different gating operators based on CCA when taking ResNet18. The best results in this table are labeled in **bold**.

Operator	ACC	F1	Kappa
Sigmoid	93.28	93.36	87.30
Tanh	93.44	93.30	87.69
ReLU	89.05	88.90	78.80
Sparemax [44]	91.09	91.83	84.09
Softmax	94.94	94.66	90.70

Table 7
Comparison of different pooling methods based on the CCA block when taking ResNet18. The best results in this table are labeled in **bold**.

Method	ACC	F1	Kappa
ResNet18	91.02	90.98	83.43
GAP	93.28	91.33	87.43
GMP	93.03	92.58	87.00
GSP	92.52	92.61	86.33
GAP+GMP	94.21	94.27	89.03
GMP+GSP	94.05	92.49	88.94
GAP+GSP	93.31	93.91	87.86
GAP+GMP+GSP	**94.94**	**94.66**	**90.70**

Table 8
Performance comparison of glaucoma detection on the ACRIMA dataset. The best results in this table are labeled in **bold**.

Method	ACC	F1	Se	SP	AUC
VGG19 [31]	90.69	91.25	92.40	88.46	96.86
InceptionV3 [31]	90.00	90.56	92.16	0.8752	96.53
Xception [31]	89.77	90.51	93.46	85.80	96.05
ResNet50 [31]	90.29	90.76	91.05	89.43	96.14
SENet	96.48	96.40	96.15	95.82	99.37
SRM	95.78	95.71	95.71	95.64	99.07
ECA-Net	95.07	94.96	94.72	94.37	98.69
CCA-Net	**97.89**	**97.84**	**97.58**	**97.28**	**99.70**

adjusting the relative importance of three different channel-wise statistical features. The results also verify the effectiveness of the channel-connected method through comparisons to the fully-connected method for weighing mixed features. This is because the channel-connected method uses a local–global integration method and the fully-connected method adopts a global integration method, which agrees with the description in Sections 3 and 4.3.

Interestingly, all used integration methods with the BN method worsen the classification performance. The possible reason to account for classification results is that the softmax operator can be taken as a normalization method, which is conflicted with the BN method.

6.2. The impact of different gating operators

Table 6 provides the performance comparisons of five different gating operators. It can be seen that exchanging the tanh and the sigmoid for the softmax slightly worsens the NC classification results. However, when replacing the softmax with ReLu and sparsemax, the performance of CCA-Net dramatically drops below the ResNet-18 baseline. This indicates that the CCA block with softmax is more effective in adjusting the importance of channels than the other four gating operators. The careful construction of the gating operator is an important factor for the NC classification performance. Fig. 13 provides the attention weights in all channels for three severity levels of NC: Normal, Mild, and Severe. We can see that the softmax can adjust the relative importance of different channels effectively, highlighting significant channels and suppressing less useful channels.

6.3. The impact of pooling methods

Table 7 shows the classification results of seven pooling methods with the proposed clinical integration operator (except the ResNet-18 baseline). Compared with the baseline, each pooling component of mixed pooling boosts the performance. The combination of two pooling methods further brought significant improvements for NC classification on three metrics. Our mixed pooling method obtains the best classification results among seven pooling methods, demonstrating that the proposed mixed pooling outperforms the combination of GAP and GMP in CBAM and the combination of GAP and GSP in SRM. The results also verify that three features play different roles in the CCA block, agreeing with our expectation and clinical research.

7. Performance comparison on public datasets

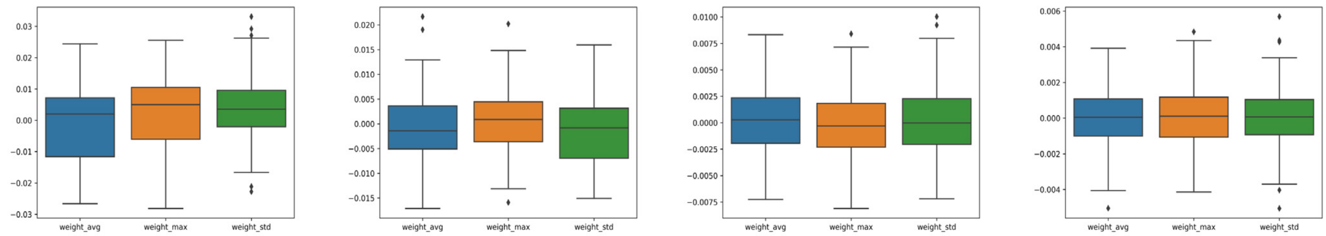
Table 8 reports the glaucoma detection results of our CCA-Net and other deep learning methods. The CCA-Net gets the best performance among all methods (**97.89%** of accuracy, and 97.58% of sensitivity). It increases over 1% through comparisons to other attention-based CNNs in terms of accuracy and sensitivity. All methods reached over 96% on AUC, and the proposed CCA-Net gets the highest AUC value.

Table 9 shows the classification results of our method, comparable attention-based CNNs, and previous classification results on the UCSD dataset. It can see that the proposed CCA-Net gets 96.09% accuracy and 94.06% sensitivity, respectively. It achieves about 1.5% improvement compared with state-of-the-art attention-based CNNs and previous work. The results on two public datasets also demonstrate the generalization ability of our method.

Figs. 17 and 18 provide attention weight visualization results of our CCA-Net on the ACRIMA dataset and the UCSD dataset. It can be observed that three global feature representations play different roles in AMD and glaucoma recognition, proving the effectiveness of our method.

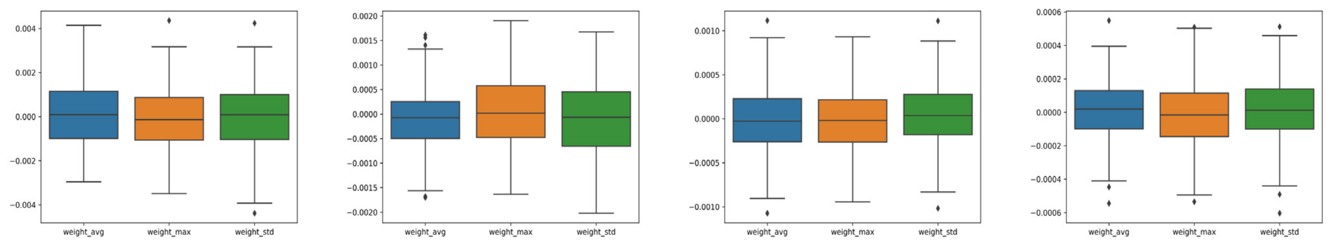
8. Conclusion and future work

This paper proposes a clinical-awareness attention network (CCA-Net) to predict AS-OCT image-based nuclear cataract severity levels. In the CCA-Net, we fully leverage the potential of clinical prior knowledge to construct a practical yet effective clinical-awareness attention (CCA) block for enhancing the performance by modeling the inter-relationships of channels in a local–global manner. The comprehensive experiments on the clinical AS-OCT image dataset and publicly available ophthalmology datasets show that our method performs better than strong baselines and previous state-of-the-art methods. Furthermore,



(a) Conv_2_2 (b) Conv_3_2 (c) Conv_4_2 (d) Conv_5_2

Fig. 17. The averaged weight values of three global channel-wise statistics features in all channels at three stages on the ACRIMA dataset by using CCA-Net: Avg (blue), Max (orange), and Std (green). Standard deviation results are also plotted.



(a) Conv_2_2 (b) Conv_3_2 (c) Conv_4_2 (d) Conv_5_2

Fig. 18. The averaged weight values of three global channel-wise statistics features in all channels at three stages on the USUD dataset by using CCA-Net: Avg (blue), Max (orange), and Std (green). Standard deviation results are also plotted.

Table 9
Performance comparison of the CCA-Net and state-of-the-art methods on the UCSD dataset. The best results in this table are labeled in **bold**.

Method	ACC	Se	F1	Kappa
LBP-SVM [33]	71.33	48.27	64.04	41.00
HOG-SVM [45]	78.90	66.20	–	–
MDFF [33]	93.93	91.76	91.46	83.00
VGG16 [33]	91.50	91.50	91.50	77.00
ResNet34 [32]	80.50	78.30	–	–
Inception [45]	90.30	90.00	–	–
LACNN [45]	90.20	88.10	–	–
LACNN-AlexNet [45]	91.20	86.80	–	–
LACNN-Inception [45]	93.00	91.60	–	–
SENet	94.16	90.00	91.49	91.23
SRM	94.20	89.74	91.30	91.29
ECA-Net	94.40	91.83	92.08	91.66
CCA-Net	**96.09**	**94.06**	**94.39**	**94.18**

we analyze the importance of clinical features with learnable weights visually and verify the effectiveness of inherent behaviors in CCA in-depth, enhancing the interpretation and understanding of our CCA-Net in automatic NC diagnosis. We hope it can motivate attention-based CNN architecture design to exploit prior knowledge fusion in other applications. We plan to develop lightweight CCA-Nets and deploy them on ophthalmic equipment in the future. More AS-OCT images would be collected to test the generation performance of designed models.

CRediT authorship contribution statement

Xiaoqing Zhang: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Zunjie Xiao:** Writing – review & editing, Investigation. **Lingxi Hu:** Writing – review & editing, Investigation. **Gelei Xu:** Writing – review & editing, Investigation. **Risa Higashita:** Conceptualization, Supervision, Data curation. **Wan Chen:** Supervision, Writing – review & editing, Investigation. **Jin Yuan:** Conceptualization, Writing – review & editing. **Jiang Liu:** Conceptualization, Supervision, Writing – review & editing, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by Guangdong Provincial Department of Education (2020ZDZX3043), Guangdong Provincial Key Laboratory (2020B121201001), and Shenzhen Natural Science Fund (JCYJ20200109140820699) and the Stable Support Plan Program (20200925174052004).

References

[1] W.H. Organization, et al., World Report on Vision, World Health Organization, 2019.
[2] M.J. Burton, J. Ramke, A.P. Marques, R.R.A. Bourne, H.B. Faal, The lancet global health commission on global eye health: vision beyond 2020, Lancet Glob. Health (suppl) (2021).
[3] Y.-C. Liu, M. Wilkins, T. Kim, B. Malyugin, J.S. Mehta, Cataracts, Lancet 390 (10094) (2017) 600–612.
[4] D. Lam, S.K. Rao, V. Ratra, Y. Liu, P. Mitchell, J. King, M.-J. Tassignon, J. Jonas, C.P. Pang, D.F. Chang, Cataract, Nat. Rev. Dis. Primers 1 (1) (2015) 1–15.
[5] M. Ozgokce, M. Batur, M. Alpaslan, A. Yavuz, A. Batur, E. Seven, H. Arslan, A comparative evaluation of cataract classifications based on shear-wave elastography and B-mode ultrasound findings, J. Ultrasound 22 (4) (2019) 447–452.
[6] W.L. Wong, X. Li, J. Li, C.-Y. Cheng, E.L. Lamoureux, J.J. Wang, C.Y. Cheung, T.Y. Wong, Cataract conversion assessment using lens opacity classification system III and wisconsin cataract grading system, Invest. Ophthalmol. Vis. Sci. 54 (1) (2013) 280–287, http://dx.doi.org/10.1167/iovs.12-10657.
[7] X. Zhang, Y. Hu, Z. Xiao, J. Fang, R. Higashita, J. Liu, Machine learning for cataract classification and grading on ophthalmic imaging modalities: A survey, Mach. Intell. Res. 19 (2022) 184–208, http://dx.doi.org/10.1007/s11633-022-1329-0.
[8] A.L. Wong, C.K.-S. Leung, R.N. Weinreb, A.K.C. Cheng, C.Y.L. Cheung, P.T.-H. Lam, C.P. Pang, D.S.C. Lam, Quantitative assessment of lens opacities with anterior segment optical coherence tomography, Br. J. Ophthalmol. 93 (1) (2009) 61–65, http://dx.doi.org/10.1136/bjo.2008.137653, URL https://bjo.bmj.com/content/93/1/61. arXiv:https://bjo.bmj.com/content/93/1/61.full.pdf.

[9] A. de Castro, A. Benito, S. Manzanera, J. Mompeán, B. Canizares, D. Martínez, J.M. Marín, I. Grulkowski, P. Artal, Three-dimensional cataract crystalline lens imaging with swept-source optical coherence tomography, Invest. Ophthalmol. Vis. Sci. 59 (2) (2018) 897–903.
[10] I. Grulkowski, S. Manzanera, L. Cwiklinski, J. Mompeán, A. De Castro, J.M. Marin, P. Artal, Volumetric macro-and micro-scale assessment of crystalline lens opacities in cataract patients using long-depth-range swept source optical coherence tomography, Biomed. Opt. Express 9 (8) (2018) 3821–3833.
[11] N.Y. Makhotkina, T.T. Berendschot, F.J. van den Biggelaar, A.R. Weik, R.M. Nuijts, Comparability of subjective and objective measurements of nuclear density in cataract patients, Acta Ophthalmol. 96 (4) (2018) 356–363.
[12] W. Wang, J. Zhang, X. Gu, X. Ruan, Y. Liu, Objective quantification of lens nuclear opacities using swept-source anterior segment optical coherence tomography, Br. J. Ophthalmol. (2021) bjophthalmol–2020–318334.
[13] J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-excitation networks, IEEE Trans. Pattern Anal. Mach. Intell. (2019).
[14] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 2921–2929, http://dx.doi.org/10.1109/CVPR.2016.319.
[15] M. Ang, M. Baskaran, R.M. Werkmeister, J. Chua, D. Schmidl, V.A. Dos Santos, G. Garhöfer, J.S. Mehta, L. Schmetterer, Anterior segment optical coherence tomography, Prog. Retin. Eye Res. 66 (2018) 132–156.
[16] V.A. Dos Santos, L. Schmetterer, H. Stegmann, M. Pfister, A. Messner, G. Schmidinger, G. Garhofer, R.M. Werkmeister, CorneaNet: fast segmentation of cornea OCT scans of healthy and keratoconic eyes using deep learning, Biomed. Opt. Express 10 (2) (2019) 622–641.
[17] B. Keller, M. Draelos, G. Tang, S. Farsiu, A.N. Kuo, K. Hauser, J.A. Izatt, Real-time corneal segmentation and 3D needle tracking in intrasurgical OCT, Biomed. Opt. Express 9 (6) (2018) 2716–2732.
[18] H. Fu, M. Baskaran, Y. Xu, S. Lin, D.W.K. Wong, J. Liu, T.A. Tun, M. Mahesh, S.A. Perera, T. Aung, A deep learning system for automated angle-closure detection in anterior segment optical coherence tomography images, Am. J. Ophthalmol. 203 (2019) 37–45.
[19] H. Fu, Y. Xu, S. Lin, D.W.K. Wong, B. Mani, M. Mahesh, T. Aung, J. Liu, Multi-context deep network for angle-closure glaucoma screening in anterior segment OCT, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 356–363.
[20] H. Fu, F. Li, X. Sun, X. Cao, J. Liao, J.I. Orlando, X. Tao, Y. Li, S. Zhang, M. Tan, et al., Age challenge: Angle closure glaucoma evaluation in anterior segment optical coherence tomography, Med. Image Anal. 66 (2020) 101798.
[21] X. Zhang, Z. Xiao, H. Risa, W. Chen, J. Yuan, J. Fang, Y. Hu, J. Liu, A novel deep learning method for nuclear cataract classification based on anterior segment optical coherence tomography images, in: IEEE SMC, 2020.
[22] X. Zhang, J. Fang, Z. Xiao, H. Risa, W. Chen, J. Yuan, J. Liu, Classification algorithm of nuclear cataract based on anterior segment coherence tomography image, Comput. Sci. 49 (2022) 204–210, http://dx.doi.org/10.11896/jsjkx.201100085.
[23] X. Wang, R. Girshick, A. Gupta, K. He, Non-local neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7794–7803.
[24] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, W. Liu, Ccnet: Criss-cross attention for semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 603–612.
[25] S. Woo, J. Park, J.-Y. Lee, I. So Kweon, Cbam: Convolutional block attention module, in: ECCV, 2018, pp. 3–19.
[26] J. Park, S. Woo, J.-Y. Lee, I.S. Kweon, Bam: Bottleneck attention module, 2018, arXiv preprint arXiv:1807.06514.
[27] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-Net: efficient channel attention for deep convolutional neural networks, 2020 IEEE, in: CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2020.
[28] H. Lee, H.-E. Kim, H. Nam, Srm: A style-based recalibration module for convolutional neural networks, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1854–1862.
[29] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, PMLR, 2015, pp. 448–456.
[30] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[31] A. Diaz-Pinto, S. Morales, V. Naranjo, T. Köhler, J.M. Mossi, A. Navea, CNNs for automatic glaucoma assessment using fundus images: an extensive validation, Biomed. Eng. Online 18 (1) (2019) 1–19.
[32] D.S. Kermany, M. Goldbaum, W. Cai, C.C. Valentim, H. Liang, S.L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, Cell 172 (5) (2018) 1122–1131.
[33] V. Das, S. Dandapat, P.K. Bora, Multi-scale deep feature fusion for automated classification of macular pathologies from OCT images, Biomed. Signal Process. Control 54 (2019) 101605, http://dx.doi.org/10.1016/j.bspc.2019.101605.
[34] L. Cao, H. Li, Y. Zhang, L. Zhang, L. Xu, Hierarchical method for cataract grading based on retinal images using improved haar wavelet, Inf. Fusion 53 (2020) 196–208.
[35] J. Jiang, X. Liu, L. Liu, S. Wang, E. Long, H. Yang, F. Yuan, D. Yu, K. Zhang, L. Wang, et al., Predicting the progression of ophthalmic disease based on slit-lamp images using a deep temporal sequence network, PLoS One 13 (7) (2018) e0201142.
[36] H. Fu, Y. Xu, S. Lin, D.W.K. Wong, M. Baskaran, M. Mahesh, T. Aung, J. Liu, Angle-closure detection in anterior segment OCT based on multilevel deep network, IEEE Trans. Cybern. 50 (7) (2020) 3358–3366, http://dx.doi.org/10.1109/TCYB.2019.2897162.
[37] X. Li, X. Hu, J. Yang, Spatial group-wise enhance: Improving semantic feature learning in convolutional networks, 2019, arXiv preprint arXiv:1905.09646.
[38] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.
[39] C. Kulkarni, Evaluation of the relation between lens opacities classification system III grading and nuclear size by direct measurement, Taiwan J. Ophthalmol. 10 (2).
[40] H. Li, J.H. Lim, J. Liu, P. Mitchell, A.G. Tan, J.J. Wang, T.Y. Wong, A computer-aided diagnosis system of nuclear cataract, IEEE Trans. Biomed. Eng. 57 (7) (2010) 1690–1698.
[41] X. Zhang, Z. Xiao, R. Higashita, Y. Hu, W. Chen, J. Yuan, J. Liu, Adaptive feature squeeze network for nuclear cataract classification in as-oct image, Journal of Biomedical Informatics 128 (2022) 104037.
[42] Z. Li, Y. Zhang, S. Arora, Why are convolutional nets more sample-efficient than fully-connected nets? 2020, arXiv preprint arXiv:2010.08515.
[43] T.B. Chandra, K. Verma, B.K. Singh, D. Jain, S.S. Netam, Coronavirus disease (COVID-19) detection in chest X-ray images using majority voting based classifier ensemble, Expert Syst. Appl. 165 (2021) 113909.
[44] A.F.T. Martins, R.F. Astudillo, From softmax to sparsemax: A sparse model of attention and multi-label classification, in: Proceedings of the 33rd International Conference on International Conference on Machine Learning, Vol. 48, in: ICML'16, JMLR.org, 2016, pp. 1614–1623.
[45] L. Fang, C. Wang, S. Li, H. Rabbani, X. Chen, Z. Liu, Attention to lesion: Lesion-aware convolutional neural network for retinal optical coherence tomography image classification, IEEE Trans. Med. Imaging 38 (8) (2019) 1959–1970.