Journal of Biomedical Informatics 128 (2022) 104037

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Original Research

Adaptive feature squeeze network for nuclear cataract classification in AS-OCT image

Xiaoqing Zhang [a,2,*], Zunjie Xiao [a,2], Risa Higashita [b], Yan Hu [a], Wan Chen [c], Jin Yuan [c], Jiang Liu [a,d,e,f,1]

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

ARTICLE INFO

Keywords:
AS-OCT image
Nuclear cataract classification
Global adaptive pooling
Squeeze block
Adaptive feature squeeze network

ABSTRACT

Nuclear cataract (NC) is an age-related cataract disease. Cataract surgery is an effective method to improve the vision and life quality of NC patients. Anterior segment optical coherence tomography (AS-OCT) images are noninvasive, reproductive, and easy-measured, which can capture opacity clearly on the lens nucleus region. However, automatic AS-OCT-based NC classification research has not been extensively studied. This paper proposes a novel convolutional neural network (CNN) framework named Adaptive Feature Squeeze Network (AFSNet) to classify NC severity levels automatically. In the AFSNet, we construct an adaptive feature squeeze module to dynamically squeeze local feature representations and update the relative importance of global feature representations, which is comprised of a squeeze block and a global adaptive pooling operation. We conduct comprehensive experiments on a clinical AS-OCT image dataset and a public OCT images dataset, and results demonstrate our method's effectiveness and superiority over strong baselines and previous state-of-the-art methods. Furthermore, this paper also demonstrates that CNNs achieve better NC classification results on the nucleus region than the lens region. We also adopt the class activation mapping (CAM) technique to localize the discriminative regions that CNN models learned, which enhances the interpretability of classification results.

1. Introduction

Cataracts are the commonest cause of reversible loss of useful vision and blindness worldwide [1], which are the loss of crystalline lens transparency due to opacity. Cataract patients can improve their vision and life quality with early intervention and cataract surgery. NC is one of the most common cataract types, and its manifestations include the gradual clouding and progressive hardening of the nucleus region in the crystalline lens region. According to the real requirements of clinical NC diagnosis [35], NC can be categorized into three different development stages based on Lens Opacities Classification System III (LOCS III) [8]. Stage 0: Normal (non-nuclear cataract), without nuclear opacity. Stage 1: Mild cataract (NC grade $= 1$ or $= 2$), is asymptomatic. Stage 2: Moderate cataract (NC grade $= 3$), is symptomatic. Stage 3: Severe cataract (NC grade > 3), is symptomatic severely. Clinical intervention, such as Kary Uni eye drops, slowing NC development progress for mild NC patients. For patients with moderate NC, clinical progress follow-up is needed. Patients with severe cataract should undergo cataract surgery.

In the past years, clinicians have utilized several ophthalmic images (like fundus images and slit lamp images) for early cataract screening and clinical NC diagnosis, E.g., they usually compare tested slit-lamp images with standard slit-lamp images from LOCS III and then give the NC diagnosis results. However, this diagnosis mode is subjective and error-prone; and highly relies on the experience of clinicians. To help

* Corresponding author.
E-mail addresses: 11930927@mail.sustech.edu.cn (X. Zhang), liuj@sustech.edu.cn (J. Liu).
[1] Principal corresponding author.
[2] Xiaoqing Zhang and Zunjie Xiao contribute this work equally.

https://doi.org/10.1016/j.jbi.2022.104037
Received 20 June 2021; Received in revised form 24 January 2022; Accepted 21 February 2022
Available online 1 March 2022

clinicians diagnose NC accurately and objectively, researchers have developed many machine learning methods for automated cataract classification. Xu et al. [47,46] proposed the group sparsity regression (GSR) and similarity weighted linear reconstruction (SWLR) methods for automated NC classification based on slit lamp images and achieved good results. Xu et al. [45] applied the deep learning method to cataract classification by using fundus images.

Anterior segment optical coherence tomography (AS-OCT) image is a new ophthalmic image, which is easy-measured, objective, noninvasive, quick, user-friendly, and high-resolution. Moreover, it can capture the nucleus region clearly compared with other ophthalmic images like fundus image can not, which is vital for NC diagnosis clinically. Clinicians have increasingly used AS-OCT images for ocular disease diagnosis and scientific research purposes in recent years. Literature [11,25] uses the deep CNN models to segment corneal area automatically in the AS-OCT images, which is helpful for corneal disease diagnosis. Fu et al. [44,13,15,14] applied deep multi-level CNNs to automatically classify angle-closure glaucoma on AS-OCT images and obtained over 90% accuracy. Furthermore, clinical research works also have studied the opacity correlation between NC severity levels and the lens region in the AS-OCT image. Wong et al. [41] first studied the built the opacity relationship between NC severity levels and average density of the nucleus region based on Spearman's correlation coefficient method, and statistical results showed that the correlation between them is strong. Wang et al. [40] evaluated correlation relationship between the average density of the nucleus region and hard nuclear cataract, and the results show it is potential to diagnose hard NC objectively on AS-OCT images. Followed by [41], research works [27,4,5,18,34] also achieved similar correlation relationship results between NC severity levels and average density of the nucleus region on AS-OCT images. Fig. 1 provides three representative NC severity levels on AS-OCT images. Overall, these clinical AS-OCT image-based classification works provide clinical support for automated NC classification and can be a potential contribution to cataract surgery planning and clinical NC diagnosis.

Motivated by clinical research, Zhang et al. [53] utilized a CNN model named GraNet to classify NC severity levels by using the whole lens area on AS-OCT images, but they only achieved about 58% of accuracy. It also indicates that there exists great improvement room for automated NC classification. This paper proposes an efficient convolutional neural (CNN) network named adaptive feature squeeze Network (AFSNet) to classify NC severity levels on AS-OCT images automatically. The AFSNet consists of a *backbone network* (e.g., VGG19) and a novel convolution module named *adaptive feature squeeze (AFS) module*. The AFS is comprised of a squeeze block and a global adaptive pooling layer (GDP). The motivation to design the AFS module based on two reasons: 1) the last convolutional layer of CNNs has massive feature maps, which not only contains useful feature representations but also has redundant feature representations. It is a significant factor in affecting NC classification results and making deep CNN models easily fit in the training. 2) In modern CNNs, GAP usually replaces fully connected layers to follow the last convolutional layer. As a result, the GAP calculates the global average value in the pooling region, ignoring the relative importance of local features in the pooling region. We use a clinical AS-OCT image dataset with 7,919 images and a publicly available OCT image dataset [26] to verify the effectiveness of our method, and the experimental results show that the AFSNet achieves the best accuracy with **86.79%** and outperforms advanced deep learning methods and classical machine learning methods over **2.73%** on the clinical AS-OCT dataset as well as demonstrate the general performance of the AFSNet on the public dataset. We also compare the NC classification performance of CNN models between the nucleus region and the whole lens region by using AS-OCT images, respectively. Furthermore, this paper uses the class activation mapping (CAM) technique [57] to visualize the discriminative regions that CNN models localized, improving the explainability of our method.

The main contributions of this paper can be summarized as follows:

- This paper proposes an efficient convolutional neural network named adaptive feature squeeze network (AFSNet) for automatic NC classification on AS-OCT images. In our AFSNet, we construct an adaptive feature squeeze (AFS) module to dynamically squeeze local feature representations based on the proposed squeeze block and update the relative importance of global feature representations through the global adaptive pooling layer.
- The experiments are conducted on an AS-OCT image dataset and a public OCT image dataset, and the results demonstrate that our AFSNet achieves state-of-the-art classification results through comparisons to strong baselines and previous works. Furthermore, we also empirically demonstrate that our global adaptive pooling works better than other advanced pooling methods.
- We verify that CNN models obtain better classification results on the nucleus region than the whole lens regions, which is consistent with clinical NC classification works. Furthermore, we apply the CAM technique to visualize the discriminative regions of CNN models, enhancing the interpretability of classification results.

2. Related work

2.1. Automatic cataract classification based on different ophthalmic images

Over the years, researchers have made great efforts in automatic cataract classification based on different ophthalmic images. [29,52] presents a slit lamp image-based NC classification framework and obtains a 0.36 mean error. Xu et al. [47] combines the bag of words method with the group sparsity regression method for NC classification on slit lamp images. Cheng [6] further improved the NC classification results with a sparse range-constrained learning method by using slit lamp image-based NC classification. Literature [44] proposes an end-to-end deep learning framework for NC classification automatically by using Faster R-CNN and achieves 84.7% accuracy. Wu et al. [43] constructed a deep learning-based platform for slit lamp image-based cataract

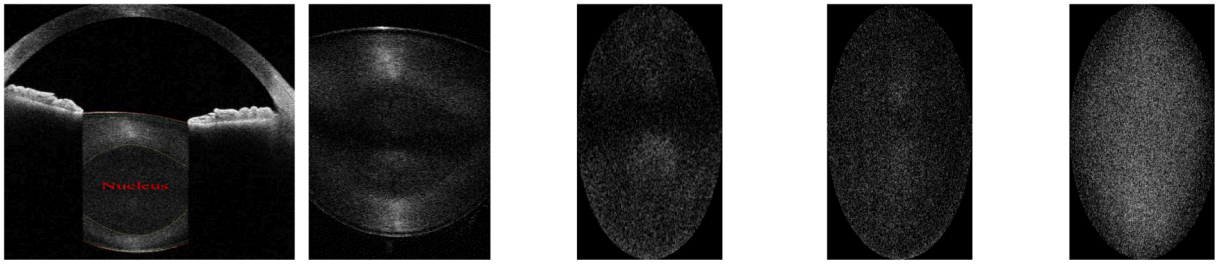


Fig. 1. Three NC severity levels based on AS-OCT image (a). Mild cataract (c) with slight opacity but is asymptomatic. Moderate cataract (d) with moderate opacity and is symptomatic. Severe cataract (e) with severe opacity and is symptomatic obviously.

screening. Xu et al. [45] proposed a global–local CNN framework for cataract screening on fundus images, in which they used different region information from fundus images. In [49,58], researchers also used a deep convolutional neural network to fundus image-based cataract screening and achieve good screening results. [2] uses ultrasonic images to classify cataract with a support vector machine (SVM), but images were collected from animals. Recently, researchers have begun to use machine learning methods to classify NC on AS-OCT images. Compared with other ophthalmic images, AS-OCT images are quick and reproductive which can capture the nucleus region clearly. [51] proposes a machine learning based NC classification on AS-OCT images and achieves over 74% accuracy. Zhang et al. [53] proposed a GraNet to predict NC severity levels, but they achieved poor performance, indicating there is much improvement room for NC classification on AS-OCT images.

2.2. Recent advances in CNNs

Recently, CNNs have achieved surpassing performance in many fields such as image classification, image segmentation, and natural language processing (NLP), etc. Researchers have proposed various methods to construct powerful CNN architectures, such as residual connection in ResNet [20], pointwise convolution method in Network in Network (NIN) [33], group convolution method in [50,55,54], and attention mechanism in squeeze-and-excitation network (SENet) [32].

More related to this paper, SqueezeNet [23] proposed a fire module to design the lightweight CNN architecture. The fire module consists of a squeeze convolution layer and an expand layer, which is proposed to reduce the number of parameters for deep CNNs. R. Girdhar et al. [16] proposed an attentional pooling method to enhance the action recognition performance. It combines the bottom-up attention method with the top-down attention method to focus on spatial feature representation for classification performance. Interleaved group convolutional neural network (IGCNet) [50], EfficientNet [38], and SKNet [32] have demonstrated that group convolution can reduce convolution redundancy as well as improve the performance of CNNs experimentally. In contrast to prior efforts, we aim at extracting useful feature representations by constructing the feature representation transformation which previous works [30,31] have discussed. To achieve this, we present a squeeze block to generate powerful feature representations and introduce a global adaptive pooling layer to adjust the relative importance of every feature representation to enhance the diversities of global feature representations.

3. Materials and methods

This section introduces the clinical AS-OCT image dataset and then illustrates the AFSNet in detail, presented in Fig. 2.

3.1. Materials

In this paper, we collect a clinical AS-OCT image dataset to evaluate the proposed method. The AS-OCT images are collected through the CASIA2 ophthalmology device (Tomey Corporation, Japan). The AS-OCT images can capture the whole anterior chamber structure (Fig. 1 (a)) of the eye includes the lens region. The lens region is comprised of the nucleus-, cortex-, and subcapsular-, as shown in Fig. 1(b); however, only the nucleus region is associated with NC according to clinical research work. We use a deep segmentation network to crop nucleus regions automatically, as shown in Fig. 1(c)(d)(e).

The AS-OCT image dataset contains 335 participants (335 eyes), and the average age is 69.40±9.97 (range:14–94 years). The NC severity levels for AS-OCT images are mapped from slit lamp images based on LOCS III. Since there is no gold standard for an NC classification system built on AS-OCT images. We collect 20 AS-OCT images from each eye and excluded 821 AS-OCT images without complete lens regions due to poorly opened eyelids; hence, the number of available AS-OCT images is 7,919. We split the AS-OCT image dataset into two disjoint subsets based on the participants: the training dataset and the testing dataset. Considering the NC severity levels of two eyes in each subject are similar. The number of AS-OCT images in the training dataset is 5655, and the number of AS-OCT images in the testing dataset is 2264. Table 1 summarizes the NC severity level distribution on the AS-OCT image dataset.

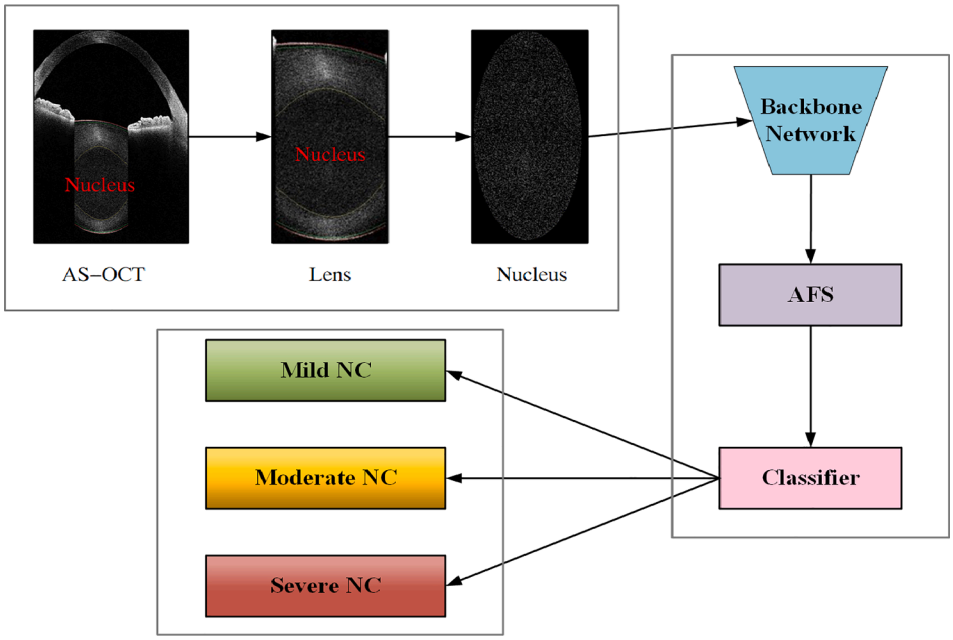


Fig. 2. The overall architecture of our adaptive feature squeeze network (ASFNet), which is comprised of a backbone network and a AFS module. First, we extract the nucleus region from AS-OCT images, which fed into the ASFNet directly; then AFSNet learns high-level feature representation from the original AS-OCT images in training; finally, the AFSNet generates the predicted NC severity levels.

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

NC level	Mild	Moderate	Severe
Training dataset	1996	2216	1443
Testing dataset	922	947	395
Total	2918	3163	1838

3.2. Adaptive feature sequeeze network

Fig. 2 shows our AFSNet architecture, which is able to generate NC severity levels automatically. Firstly, the backbone network, e.g., VGG16, extracts the high-level features from AS-OCT images. Then our AFS module squeezes local high-level features and updates the relative importance of the feature maps from the last convolutional layer dynamically through the proposed squeeze block and global adaptive pooling layer. Finally, the softmax classifier produces the predicted NC severity levels.

3.2.1. Global adaptive pooling layer

In modern CNNs, the global average pooling (GAP) layer usually replaces fully connected (FC) layers and follows the last convolutional layer. GAP is parameter-free as a structural regularizer, which is capable of alleviating over-fitting and achieving competitive performance. We use the following equation to represent the GAP operation through the following equation:

$$\mu_{kij} = \frac{1}{|R_{ij}|} \sum_{(i,j)\in R_{ij}} x_{kij}, \tag{1}$$

where μ_{kij} is the output of the *kth* feature map through the GAP operator, $|R_{ij}|$ denotes the pooling region size, and x_{kij} denotes every pixel feature in the pooling region $|R_{ij}|$.

The GAP operator treats every pixel feature representation equally and does not consider each pixel's relative importance in the pooling region, which is easily ignored by most previous works. To alleviate this problem, this paper designs a global adaptive pooling (GDP) operation to dynamically set the relative importance for each pixel feature in the pooling region by introducing learnable weights. The GDP can be wrote as follows:

$$\overline{\mu}_{kij} = \sum_{(i,j)\in R_{ij}} w_{kij} x_{kij}, \tag{2}$$

where $\overline{\mu}_{kij}$ is the output of the *kth* feature map, and w_{kij} denotes the learnable weights for each pixel feature in the *kth* feature map, which updates the relative importance of each pixel feature adaptively in the training. Given the sum of weights is 1 for all features in every feature map, thus, this paper presents a pooling region-based initialization method, in which we set initialized weight for each feature equaling to the inverse of the pooling region area, e.g., the height and width of a pooling region are 7 and 7, and the initialized weight of each feature is $\frac{1}{49}$. The relative weights of features in each feature map are updated dynamically in the training, and we can get final weights for all features.

According to dropout technique [37] and global max pooling (GMP) method, we explore three variants of GDP operation, named GDP-B, GDP-C, and GDP-D, and the original GDP named GDP-A, as shown in Fig. 3.

GDP-B uses a batch normalization (BN) [24] layer after the GDP layer.

GDP-C consists of an original GDP (GDP-A) layer and a GMP layer.

GDP-D uses a dropout layer to follow the combination of the GDP layer and the GMP layer. Following [48], we set dropout rate to 0.5 in this paper.

Furthermore, to test the effects of weight initialization methods for the GDP, we conduct comparable experiments and will discuss them in Section 5.2.

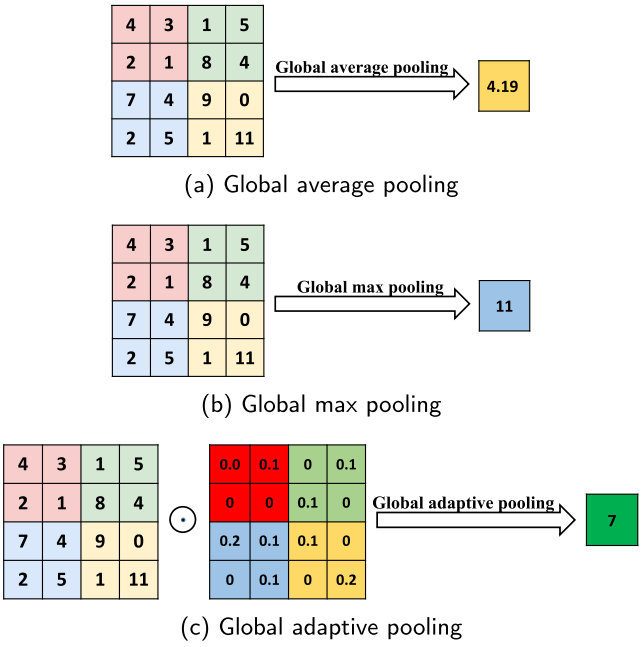


Fig. 3. Toy example illustrating three pooling operations: global average pooling (a), global max pooling (b), and global adaptive pooling (c).

3.2.2. Squeeze block

Another problem of CNNs is also easy to be neglected is that the last convolutional layer usually generated massive feature maps, which fed into the GAP layer or FC layers directly. However, not all feature maps contain useful feature representation information, and some feature maps have redundant feature map information. This paper questions that whether we construct a convolution block to squeeze feature map representation information, to both reduce feature redundancy and improve the performance.

To answer this question, we propose a simple yet effective convolution block named Squeeze block built on the pointwise convolution method. It is comprised of two pointwise convolution layers. The pointwise convolution method (1×1 conv.) has become an essential component in CNNs [33], which can cluster correlated feature representations from previous feature maps. Fig. 4(a) shows an example of the proposed squeeze block named Squeeze-1. In the Squeeze-1, given the generated feature maps from the last convolutional layer, two continuous pointwise convolution layers are applied to squeeze and extract useful feature representations. The number of convolution kernels in the first pointwise convolution layer changes with the classification

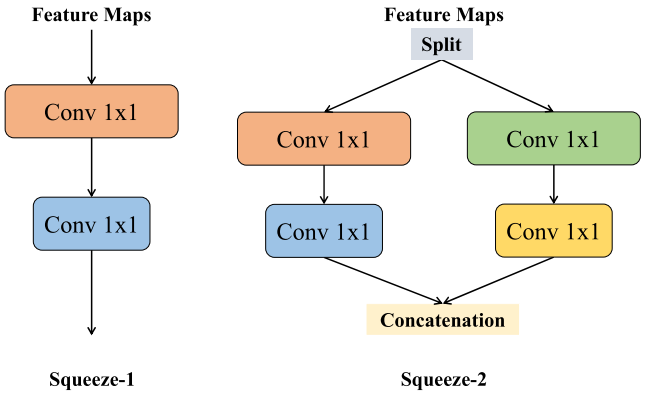


Fig. 4. Architectures of two squeeze blocks: squeeze-1 and squeeze-2.

performance. In contrast, the number of convolution kernels in the second pointwise layer is equal to the number of labels.

To further alleviate the convolution redundancy problem, we present a variant of the Squeeze-1 block by introducing the group convolution method named Squeeze-2, as illustrated in Fig. 4(b). This is because modern CNN architectures (e.g., ShuffleNet [39,56]) have demonstrated that group convolution method is capable of reducing convolution redundancy while improving the classification performance of CNNs. Following previous works, we split the generated feature maps into two independent convolution partitions equally In the Squeeze-2 block. Each convolution partition also two pointwise convolution layers, but the number of convolution kernels in the first pointwise convolution layer is half the number of convolution kernels. At the same time, the number of convolution kernels in the second pointwise convolution layer is also equal to the number of labels. In the experiments, we change the number of convolution kernels in the first pointwise convolution layer to demonstrate the generation ability of our squeeze blocks on state-of-the-art CNNs.

Discussion: The fire module in SqueezeNet is similar to our proposed Squeeze block. The core motivation between our Squeeze and the fire module is that the proposed squeeze block is used to extract and cluster informative feature representations for enhancing the classification performance while the fire module is used to construct lightweight CNN model. Moreover, we summarize the difference between our method and the fire module as follows:

- Our squeeze block uses the group convolution method and pointwise convolution method (1 × 1 Conv) for enhancing the feature representation difference of feature maps. The fire module adopts the multi-receptive field convolution method (1 × 1 Conv and 3 × 3 Conv) for learning feature representations from different receptive fields.
- Our squeeze block consists of two pointwise convolution layers, and the number of convolution kernels in each pointwise convolution layer decreased continuously. The core motivation to design the squeeze block is to aim to cluster feature representations to improve the difference between feature maps by decreasing the convolution kernels. We use the same convolution kernel size for two convolutional layers rather than mix convolution kernels, e.g., 1x1 Conv and 3x3 Conv. It is because we find that 1x1 Conv (pointwise convolution) is more able to cluster and learn feature representations of feature maps through comparisons to other convolution kernel sizes like 3x3 Conv and 5x5 Conv, which existing state-of-the-art CNNs have demonstrated. The fire module comprises a squeeze convolution layer and an expand layer. The squeeze convolution layer (1x1 Conv) in the fire module is used to limit the number of input channels to the expand layer, which is a significant factor in reducing the number of parameters of CNNs. The expand layer adopts a mix of 1x1 and 3x3 convolution kernels to learn varying feature representations from different receptive fields; moreover, it replaces a part of 3x3 convolution kernels with 1x1 convolution kernels for reducing parameters, which is another method to reduce the number of parameters.
- Our squeeze block is designed to learn informative feature representations only at high-level stage of CNNs; the fire module is utilized to construct light-weight CNNs at three different stages: low-level, medium-level, and high-level.

3.2.3. Adaptive feature squeeze module

Considering the advantages of GDP layer and squeeze block, this paper proposes an adaptive feature squeeze (AFS) module, which is capable of squeezing useful feature representations and adjusting relative importance of pixel feature representations in the pooling region. Fig. 5 depicts the paradigm of our adaptive feature squeeze block, comprised of a squeeze block and a global adaptive pooling layer. In the AFS block, the generated feature maps from the last convolutional layer

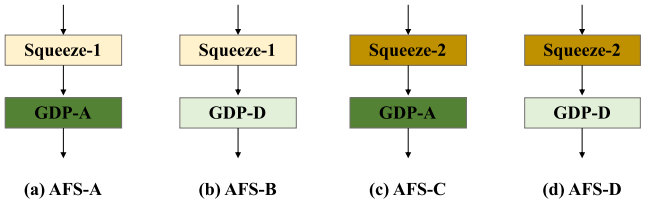


Fig. 5. Architecture of four AFS modules: AFS-A, AFS-B, AFS-C, and AFS-D.

are firstly fed into the squeeze block, followed by GDP layer, which generate the final outputs.

To study which components of the AFS module affect the classification results, we explore four AFS module topologies: AFS-A, AFS-B, ASF-C, and AFS-D, as shown in Fig. 5.

AFS-A is comprised of a Squeeze-1 block and a GDP-A layer.
AFS-B also uses the Squeeze-1 block but the pooling layer is GDP-B.
AFS-C uses the Squeeze-2 block and a GDP-A layer.
AFS-D adopts the Squeeze-2 block and a GDP-D layer.

3.2.4. Network architecture

Fig. 2 provides an example of AFSNet architecture, comprised of a backbone network (e.g., VGG16) without a GAP layer, an ASF module, and the classifier. The softmax function is used as the classifier, which generates three predicted NC severity levels. Moreover, cross-entropy (CE) loss is used as the loss function, which can be formulated as:

$$Loss_{CE} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log \widehat{y}_i, \tag{3}$$

where $y_i, \widehat{y}_i$, and N denote the ground truth, predicted labels, and the number of AS-OCT images in each iteration.

4. Metrics and experiment setting

4.1. Evaluation measures

Following previous cataract classification works [3], this paper uses five commonly used evaluation measures to assess our method and strong baselines: accuracy (ACC), sensitivity (Sen), precision (PR), F1 score, and kappa coefficient value. ACC, F1, and kappa are three major metrics to evaluate the overall performance of methods. ACC represents the number of AS-OCT images that are correctly classified. F1 is an essential indicator for evaluating the overall performance of a method. Sen is a vital evaluation measure for NC diagnosis clinically, indicating the number of each NC severity level are correctly classified, and macro sensitivity is used in this paper. Macro sensitivity is the arithmetic mean of the sensitivity values for all NC severity levels, which accounts for the label imbalance problem. Kappa is a useful measure to evaluate diagnostic reliability [9,3], as shown in Table 2. It is applied extensively in clinical studies and skill assessments. We can obtain a kappa coefficient value based on the confusion matrix. These evaluation measures can be represented as follows:

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

Table 2
An example of confusion matrix.

Predicted results	Ground truth			
	Class 1	Class 2	…….	Class n
Class 1	b_{11}	b_{12}		b_{1n}
Class 2	b_{21}	b_{22}		b_{2n}
……			……	
Class n	b_{n1}	b_{n2}		b_{nn}

$$Sen = \frac{TP}{TP + FN}, \tag{5}$$

$$PR = \frac{TP}{TP + FP}, \tag{6}$$

$$F1 = \frac{2 \times PR \times Sen}{PR + Sen}, \tag{7}$$

where TP, FP, TN and FN denote the numbers of true positives, false positives, true negatives and false negatives, respectively.

$$kappa = \frac{p_o - p_e}{1 - p_e}, \tag{8}$$

where p_o and p_e can be computed through the following equations:

$$p_o = \frac{\sum_{i=1}^{i=n} b_{i,i}}{\sum_{i,j=1}^{i,j=n} b_{i,j}}, \tag{9}$$

$$p_e = \frac{\sum_{j=1}^{j=n} b_{1,j} \times \sum_{i=1}^{i=n} b_{i,1} + \ldots + \sum_{j=1}^{j=n} b_{n,j} \times \sum_{i=1}^{i=n} b_{i,n}}{\sum_{i,j=1}^{i,j=n} b_{i,j} \times \sum_{i,j=1}^{i,j=n} b_{i,j}}. \tag{10}$$

4.2. Baselines

To evaluate the performance of our methods on NC classification task comprehensively, we conduct a series of comparable experiments.

1 **Pooling operations.** To demonstrate the effectiveness of the global adaptive pooling operation (GDP), global average pooling operation (GAP) [33], global max pooling operation (GMP), and mixed pooling operation (MP) [48] are used for comparison based on advanced CNN models, e.g., VGGNets, ResNets, and SENets. MP is comprised of a GAP and a GMP.

Furthermore, we also test the effects of initial weight initialization methods on our GDP, Kaiming initialization methods [19], Glorot initialization methods [17], constant one initialization method, and the proposed pooling region-based initialization method are adopted. The constant one initialization method denotes that initial weights are constant one. As previously introduced, the pooling region-based initialization method represents the initial weights equaling to the inverse of the pooling region.

2 **Explored squeeze blocks and ASF modules.** This paper also implements a series of comparable experiments to explore which factors affect the performance of Squeeze blocks and ASF modules and discuss the results. Following previous works [53,7], we only change the number of convolution kernels in the first convolutional layer of squeeze blocks, and the number of convolution kernels in the second convolutional layer is equal to the number of NC severity levels.

3 **Baselines.** According to previous NC classification works [46,51] and clinical AS-OCT image-based research works, this paper also extracts eighteen image features from the nucleus region on AS-OCT images: mean density, maximum density, median, entropy, and uniformity. Mean density is the average of the nucleus region in the AS-OCT images; Entropy and uniformity are extracted from the histogram, which can measure the uncertainty of the histogram. Based on extracted features, we use five classical machine learning methods to classify NC severity levels automatically: support vector machine (SVM), random forest (RF), decision tree (DT), Adaboost, and multi-class logistic regression (MLR). To further test the effectiveness of the proposed methods, state-of-the-art CNN models: ResNets [21], SENet [22], VGGs [36], EfficientNet [38], SKNet [32], and CBAM [42].

4 **NC classification results on the nucleus region and the lens region.** We use five advanced CNN models to compare NC classification results of the nucleus region and the lens region. Class activation mapping (CAM) technique [57] is used to highlight what the discriminative regions CNNs localize on the nucleus region and the lens region, respectively, which can help to explain classification results of CNNs on these two regions.

4.3. Experiment setting

This paper uses Pytorch, Python, OpenCV, and sci-kit-learn software packages to implement the proposed methods and strong baselines. We train all CNNs by using stochastic gradient descent (SGD) optimizer with the default setting, and the batch size is set to 16. The training epoch size is set to 100. The initial learning rate is set to 0.1 and decreased by ten every 30 epochs, and we set the fixed learning rate to 0.00025 when epochs over 90. We randomly flip and rotate images in training and normalize AS-OCT images with channels' means and standard deviations for data augmentation. We use OpenCV package to extract eighteen image features from the nucleus region on AS-OCT images, and five machine learning methods use the default setting. We conduct all methods on a server with an NVIDIA TITAN V (11 GB) GPU and train all deep learning methods from scratch.

5. Results and discussion

5.1. Performance comparison of different pooling operations

Table 3 shows NC classification results of our GDP operation and other three pooling operations based on ResNets, VGGNets, and SENets, respectively. We can see that our GDP consistently improves the performance than state-of-the-art pooling operations. The GDP obtains the best NC classification results with **85.16%** of accuracy, **85.17%** of F1, and **76.37%** of kappa value based on VGGNet19, respectively. Remarkably, it outperforms the other pooling operations on VGG19, e.g., GAP, GMP, and MP operations by above absolute 2.00%, **2.19%**, and 1.6% in the kappa value correspondingly. Furthermore, it also obtains similar improvements in accuracy and F1, respectively. These results demonstrate the effectiveness of the GDP, proving that our GDP is more capable of learning pixel feature information than parameter-free pooling functions e.g., GAP by introducing learnable parameters. This is because pixel feature information of each generated feature map of the last convolutional layer is different, while comparable pooling operations cannot learn pixel feature information by adjusting weights dynamically.

Fig. 6 presents twelve representative weight distributions of feature maps from the last convolutional layer. We can see that the weights of all features in a feature map are different, which suggests that every feature plays a different role in global feature representation generation. It also demonstrates that our method enables the CNNs to obtain better classification results by generating differentiated global feature representations, as well as proving the effectiveness of the proposed pooling methods.

In the following experiments, we use *VGG16 and VGG19* as baselines to test classification results of the GDP and its variants performance. Because our GDP gets better improvements on these two CNNs through comparisons to other pooling operations.

Table 4 and Fig. 7 show the classification performance of GDP-A and its three variants on VGG19 and VGG16. In Fig. 7, the horizontal axis represents the GDP and its three variants, and the vertical axis represents kappa values, respectively. It can be seen that GDP-D achieves the best NC classification results on VGG16 (e.g., **85.56%** of accuracy). GDP-A achieves the second-best classification performance on VGG19 (e.g., **85.15%** of accuracy). GDP-B obtains the worst NC classification of four pooling operations, which shows the BN method can not enhance the performance of the proposed GDP. GDP-A and GDP-D outperform GDP-C

Table 3
Performance comparison of pooling operations based on state-of-the-art CNN models. (The best results are marked in **bold**.)

	GAP			GMP			MP			GDP		
	ACC	F1	Kappa	ACC	F1	Kappa	ACC	F1	Kappa	ACC	F1	Kappa
ResNet18	84.06	84.13	74.49	83.79	83.70	74.20	84.28	84.04	75.09	84.41	84.36	75.26
ResNet34	84.01	84.00	74.50	83.53	83.49	73.77	84.72	84.73	75.64	84.85	84.73	75.89
VGG16	83.70	83.62	74.19	83.75	83.78	74.09	84.06	83.89	74.67	84.98	84.88	76.25
VGG19	83.83	83.77	74.37	83.70	83.62	74.18	84.10	83.98	74.77	**85.16**	**85.17**	**76.37**
SENet18	84.01	83.97	74.55	83.92	83.93	74.38	84.13	83.97	74.90	84.19	84.13	74.90
SENet34	83.61	83.54	73.93	83.48	83.47	73.73	84.50	84.35	75.35	84.81	84.75	75.81

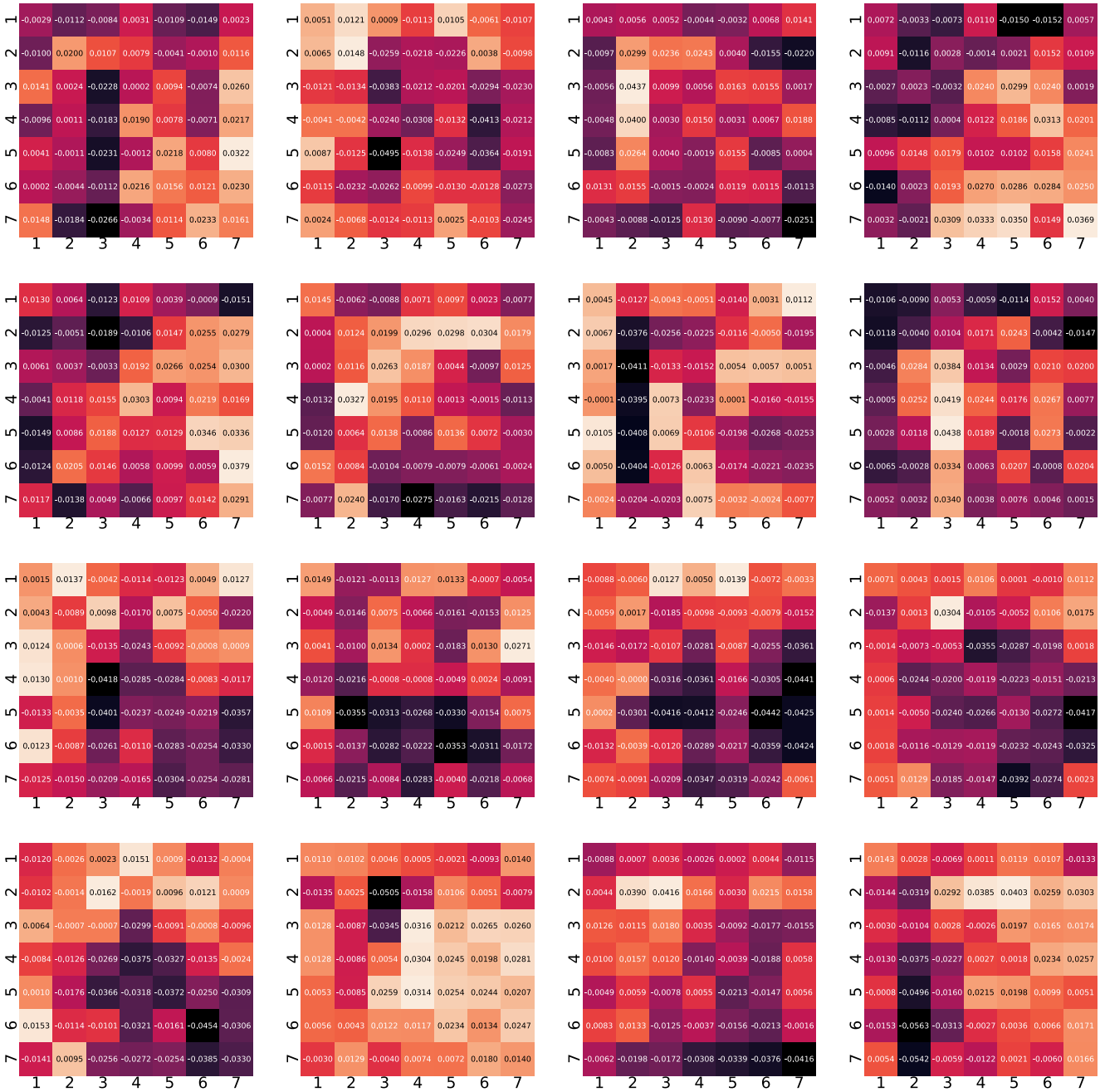


Fig. 6. Representative weight distributions are produced by our GDP method for feature maps from the last convolutional layer.

on VGG16 shows that maximum pixel representation information can improve the performance but not all maximum pixel representations have positive effects. Interestingly, GDP-A and GDP-D get better performance than GDP-C to show that the dropout technique can improve the performance of the GDP operation. However, the dropout rate is difficult to set, making the classification performance of GDP-D fluctuate vastly on different CNN models. Overall, the GDP and its variants achieve different degrees of improvement through comparisons to the GAP

Table 4
Performance of GDP and its variants based on VGG16 and VGG19. (The best results are marked in **bold**.)

Methods	VGG16		VGG19	
	ACC	F1	ACC	F1
GDP-A	84.98	84.88	85.16	85.17
GDP-B	84.32	84.14	84.19	84.05
GDP-C	84.50	84.31	85.03	84.96
GDP-D	**85.56**	**85.52**	84.50	84.41

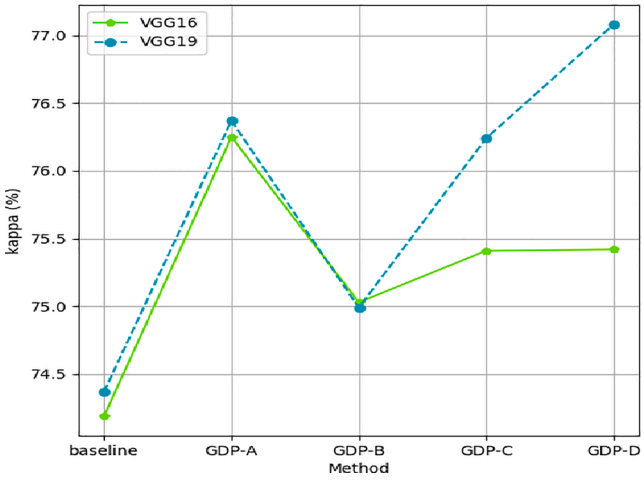


Fig. 7. Performance comparison of GDP and its variants based on VGG16 and VGG19 by using kappa coefficient value.

on CNN models, proving the general ability of our GDP operation. In the future, it is potential to use the GDP instead of the GAP to enhance the classification performance in CNNs on other learning tasks.

5.2. Effects of different weight initialization methods for global adaptive pooling layer

Table 5 compares the effects of six different weight initialization methods for our GDP operation on the AS-OCT image dataset (GDP denotes the GDP-A). We can see that two constant initialization methods achieve better classification performance than the Glorot initialization and Kaiming initialization methods. Specifically, our pooling region-based initialization method achieves the best classification results. Two reasons can account for the classification results. 1) Glorot initialization methods and Kaiming initialization methods set weights for pixels in the pooling region randomly. On the other hand, two constant initialization methods initially set the same weights for all pixels in the pooling region and dynamically change the relative importance of each pixel in the training. 2) AS-OCT image dataset used for this learning task is small; it is challenging to train proper weights for pixel feature representations in each pooling region.

5.3. Performance comparison of explored squeeze blocks

Table 6 reports the NC classification performance of two explored squeeze blocks through four classical CNN models: VGG16, VGG19, ResNet18, and SENet18 according to classification results from Table 3. As previously discussed, the core difference between the two squeeze blocks: Squeeze-1 only has one branch, while Squeeze-2 has two branches, respectively. Each squeeze block has two pointwise convolution layers. We only change the number of convolution kernels in the first layer of squeeze blocks.

It can be seen that Squeeze-2 obtain the best accuracy of **86.09%** on VGG16 (**the number of convolution kernels is 128**), and Squeeze-1 obtain the best accuracy of **85.91%** on VGG19 (**the number of convolution kernels is 256**). When the number of convolution kernels changes, four advanced CNN models with two squeeze blocks achieve the best accuracy accordingly, e.g., ResNet18 with Squeeze-1 block achieves the best accuracy of 85.38% when the number of convolution kernels is set to 256; SENet18 with Squeeze-2 block achieves the best accuracy of **85.20%** when the number of convolution kernels is set to 64. The results show that it is challenging to set the optimal number of convolution kernels in the second convolutional layer for achieving competitive performance.

In the following experiments, we set the number of convolution kernels for Squeeze-1 and Squeeze-2 to 256 and 128, respectively. Thus, AFS-A and AFS-B built on Squeeze-1 block have 256 convolution kernels, and AFS-C and AFS-D built on Squeeze-2 have 128 convolution kernels.

Fig. 7, and Fig. 8(b) present the best kappa values and the best F1 values of two explored squeeze blocks and four state-of-the-art baselines. The results show that our squeeze blocks outperform baselines with the lowest gain of 0.59% and the highest gain of **4.15%** on two evaluation measures. Furthermore, squeeze blocks obtain higher improvements on VGGNets than ResNet18 and SENet18 based on Table 6 and Fig. 5. Two reasons can explain these results: 1) the AS-OCT image dataset used for this learning task is not large enough; 2) the generated feature map sizes of the final convolutional layer in VGGNets are smaller than ResNet18 and SENet18, which suggests that small feature maps contain more useful feature representation information than large feature representation information. Overall, the results of Table 6 showing that the effectiveness of our squeeze blocks, proving that the pointwise convolution method and group convolution method are both capable of learning useful feature representation information and reducing redundant feature representation information.

5.4. Performance comparison of AFS modules

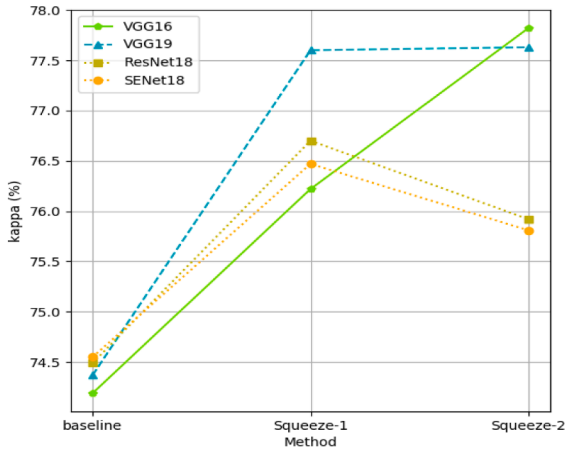
Table 7 presents the NC classification results of four explored ASF blocks, GDP, and squeeze blocks on VGG16 and VGG19. Baselines denote VGG16 and VGG19 with GAP. We can see that our methods achieve better NC classification results than two baselines on five evaluation measures, demonstrating the effectiveness of the proposed methods. ASF-A gets the best classification results with 86.79% of accuracy, 86.77% of F1, 79.02% of kappa based on VGG19, which outperforms the original VGG19 by above absolute **2.96%, 3%, and 4.65%**

Table 5
Effects of six different weight initialization methods for adaptive pooling operation. (The best results are marked in **bold**.)

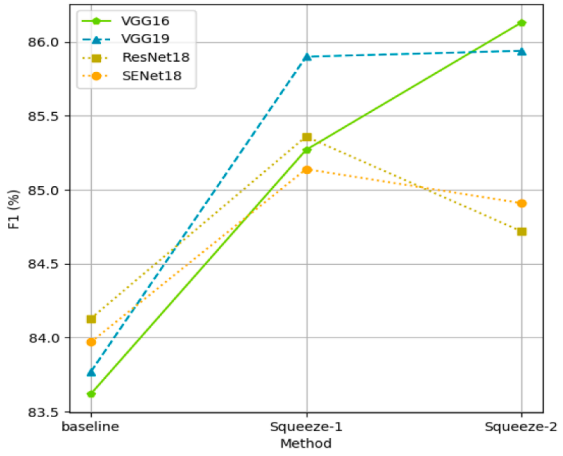
Weight initialization methods	VGG16				VGG19			
	Sen	F1	ACC	Kappa	Sen	F1	ACC	Kappa
Glorot's uniform	85.50	83.88	83.88	74.23	86.50	84.01	84.05	74.66
Glorot's normal	87.46	84.60	84.67	75.72	87.26	84.36	84.45	75.37
Kaiming's uniform	85.89	82.57	82.69	72.58	85.06	82.25	82.29	71.89
Kaiming's normal	86.66	83.62	83.70	74.15	85.99	82.99	83.04	73.07
Constant One	85.50	83.88	83.88	74.23	86.72	84.33	84.50	75.29
Pooling region-based initialization	**87.73**	84.88	84.98	76.25	87.16	**85.17**	**85.16**	**76.37**

Table 6
Accuracy comparison of explored squeeze blocks on four advanced CNN models. (The best results of each column is marked in **bold**.)

Convolution kernels	VGG16		VGG19		ResNet18		SENet18	
	Squeeze-1	Squeeze-2	Squeeze-1	Squeeze-2	Squeeze-1	Squeeze-2	Squeeze-1	Squeeze-2
512	84.85	84.41	85.51	85.65	84.63	84.36	84.06	84.28
256	84.54	85.07	**85.91**	85.78	**85.38**	84.41	84.63	84.32
128	**85.16**	**86.09**	84.01	**85.95**	84.45	**84.81**	84.45	84.28
64	84.81	84.76	84.50	85.16	84.54	84.63	**85.20**	84.63
32	84.50	85.03	84.81	84.28	85.07	84.50	84.94	**84.85**



(a) Kappa values of baselines and squeeze blocks.



(b) F1 scores of baselines and squeeze blocks.

Fig. 8. Kappa values and F1 scores of baselines and explored squeeze blocks on four advanced CNNs.

in the accuracy, F1, and kappa, respectively. The results also show that AFS blocks take advantage of both the squeeze block and the global adaptive average pooling method, which can alleviate the convolution redundancy problem and set relative importance for pixel feature representations in feature maps dynamically.

According to Table 7, it can be concluded that not all ASF modules get better classification results than individual squeeze blocks and GDP methods on two VGG networks. The following reasons can explain the results: 1) both GDP and Squeeze block are parameterized methods. It is challenging to obtain the proper parameters for weights of features in the training; 2) the dataset used in this paper is small. Overall, the average improvements of the ASF modules are higher than the individual GDP layer and squeeze blocks, showing the superiority of the ASF modules through comparisons to the results of Table 3 and Table 6.

5.5. Performance comparison with state-of-the-art methods

Table 8 provides NC classification results of advanced deep learning methods and machine learning methods. It can be observed that deep learning methods generally obtain better classification results than machine learning methods. This is because deep learning methods have powerful feature extraction ability based on the hierarchical structures, and they contain massive parameters compared with traditional machine learning methods. One possible reason that machine learning methods cannot achieve good results is that we do not extract enough useful features for machine learning methods and will design the improved feature extraction methods in the future work.

Our AFSNet-A achieves the best NC classification results: **86.79%** in the accuracy, **88.96%** in the sensitivity, **86.77%** in the F1 score, and **79.02%** in the kappa value. It outperforms state-of-the-art CNN models, e.g., SKNet, SENet with about 2.7% in the accuracy, and obtains the highest improvement is **5.27%** in the kappa value. The proposed GDP and squeeze block based on VGG16 also get better performance than advanced CNNs, proving the effectiveness of our methods. Furthermore, literature [53] achieves poor NC classification results using CNN, mainly because they use AS-OCT images of the whole lens region as inputs. Other types of cataracts in the crystalline lens can affect CNN models to distinguish different severity levels of NC.

Table 9 presents the detailed classification results of our AFSNet-A for three NC severity levels. The right-most column shows the number of test images in each NC severity level. Furtherly, Fig. 9 shows the

Table 7
Performance comparison of explored squeeze blocks. (The best results based on two state-of-the-art CNN models are marked in **bold**.)

Method	VGG16					VGG19				
	ACC	F1	PR	Kappa	Sen	ACC	F1	PR	Kappa	Sen
Baseline	83.70	83.62	86.53	74.19	86.53	83.83	83.77	86.63	74.37	86.83
GDP-A-only	84.98	84.88	85.34	76.25	87.73	85.16	85.17	85.59	76.37	87.16
GDP-D-only	85.56	85.52	85.51	77.08	87.85	84.50	84.41	84.92	75.42	87.09
Squeeze-1-only	85.16	85.27	85.56	76.22	85.90	85.91	85.90	85.97	77.60	88.21
Squeeze-2-only	86.09	**86.13**	**86.24**	77.82	87.92	85.95	85.94	86.65	77.63	87.67
AFS-A	85.58	85.53	87.26	77.07	87.85	**86.79**	**86.77**	**88.69**	**79.02**	**88.96**
AFS-B	85.47	85.37	85.57	76.96	87.87	85.47	85.35	86.12	76.94	88.10
AFS-C	**86.09**	86.05	86.03	**77.93**	**88.28**	85.73	85.78	86.36	77.19	87.16
AFS-D	85.07	84.99	85.27	76.23	87.71	84.94	84.89	84.92	76.09	87.23

Table 8
NC classification results of machine learning methods and deep learning methods. (The best results are marked in **bold**.)

Methods	Sen	ACC	F1	Kappa
SVM	81.98	77.78	77.51	65.35
RF	82.10	77.87	77.72	65.60
Adaboost	80.16	75.27	74.15	61.62
DT	81.09	76.63	76.38	63.60
MLR	82.51	78.40	78.24	66.37
GraNet [53]	–	57.85	57.00	–
VGG16	86.53	83.70	83.62	74.19
VGG19	86.83	83.83	83.77	74.37
SENet18	86.35	84.01	83.97	74.55
SqueezeNet	86.61	83.83	83.74	74.36
SENet34	86.02	83.61	83.54	73.93
ResNet18	85.23	84.06	84.13	74.49
ResNet34	85.81	84.01	84.00	74.50
EfficientNet	85.06	82.33	82.25	71.99
BAM-ResNet18	85.90	83.97	83.91	74.43
CBAM-ResNet18	85.90	83.97	83.91	74.43
Attn.Pooling + VGG19 [16]	85.46	83.16	83.12	73.20
Attn.Pooling + VGG16 [16]	85.64	83.78	83.81	74.12
SKNet	85.28	83.56	83.53	73.75
GDP-D + VGG16	87.85	85.56	85.52	77.08
Squeeze-2 + VGG16	87.92	86.09	86.13	77.82
AFSNet-C + VGG16	88.28	86.09	86.05	77.93
AFSNet-A + VGG19	**88.96**	**86.79**	**86.77**	**79.02**

Table 9
Detailed classification results for AFSNet-A based on VGG19.

class	PR	Sen	F1	Images
Mild	84.18	86.01	85.09	922
Moderate	85.14	82.89	84.00	947
Severe	96.75	97.98	97.36	395

confusion matrix of our AFSNet-A. The experiment results reported how a method performs in a realistic scenario. The sensitivity values of three NC severity levels are 86.01%, 82.89%, and 97.98%, accordingly. Compared with sensitivity values of mild and severe NC severity levels, ASNet-A obtains the low sensitivity value of moderate NC, indicating that it is easy to misclassify moderate NC images as mild NC is presented in Fig. 9. This is because the margin between moderate and mild NC is difficult to define clinically. Correctly classifying mild and moderate NC is significant for clinical cataract diagnosis, which can help clinicians diagnose NC accurately and objectively and let NC patients take effective therapeutic measures in advance.

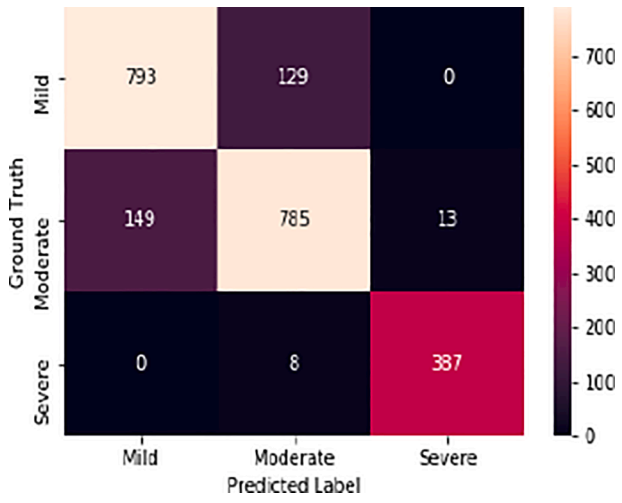


Fig. 9. The confusion matrix of ASFNet-A.

5.6. Performance comparison of deep learning methods on both nucleus region and lens region

Table 10 presents the classification results of five CNN models on the nucleus region and the lens region. We can conclude that five CNN models achieve better performance on the nucleus region than the lens region and improve with at least an absolute 3.23% in the ACC (range:3.23%-7.83%). Expect for our AFSNet (AFSNet-A), F1 scores and kappa values of the other four CNN models are lower than 80% and 70%, respectively, demonstrating the effectiveness of the AFSNet.

To see what discriminative feature representation that CNNs learned from these two regions, Fig. 10 shows discriminative region localization of heat maps with the CAM technique. It includes the nucleus region (left) and the lens region (region) on AS-OCT images. Fig. 10 has six rows, and the first row shows original AS-OCT images of three NC severity levels for each region. The other five rows represent the discriminative regions that five CNNs learned include our AFSNet. It can be seen that the AFSNet paid more attention to the accurate region information for two regions on AS-OCT images through comparisons to other four CNN models (e.g., VGG19 and SENet). The cortex region has the most negative effects on the classification results, explaining why the deep learning methods obtain poor results of lens region on AS-OCT images. Moreover, the other four CNN models localize the wrong regions for mild and moderate NC severity levels on both regions. This is because the margin between two NC severity levels is ambitious. In contrast, our AFSNet localizes the accurate region information on AS-OCT images but does not localize all correct regions, illustrating why our method gets the poor results of mild and moderate NC severity levels is shown in Table 9.

5.7. Validation

To further validate the performance of our AFSNet, we conduct extensive experiments on the public UCSD dataset. The UCSD dataset is an OCT image dataset, which comprises four different classes: choroidal neovascularization (CNV), diabetic macular edema (DME), drusen, and normal. The number of images for CNV, DME, drusen, and normal in the training dataset are 37,206, 11,349, 8,617, and 51,140, respectively. In the testing dataset, each class has 250 OCT images. The more detailed introduction of the dataset can be seen in [26]. In the experiments, we follow the dataset split method and the data preprocessing method in [10]. We train our AFSNet and SENet from scratch and use SGD as the optimizer with the default setting. The batch size and training epochs are set to 32 and 100, respectively. We set the initial learning rate to 0.025 and decreased it by a factor of five every twenty epochs.

As presented in Table 11, our AFSNet obtains the best classification results among all methods on the validation dataset. The results reported on the validation dataset are averaged with five runs of our method. Remarkably, the AFSNet achieves above 1.37% gain of accuracy, **4.12%** gain of kappa, and **1.74%**gain of F1 than previous advanced methods, indicating the effectiveness of our method is not limited to the clinical AS-OCT image dataset.

Table 12 shows the classification comparison of our proposed AFSNet and existing methods on the testing USUD dataset. We also can see

Table 10
NC classification results of the nucleus region and the lens region with five state-of-the-art CNN architectures. (The best results are marked in **bold**.)

Region	Nucleus			Whole lens		
	ACC	F1	Kappa	ACC	F1	Kappa
VGGNet19	83.83	84.14	74.37	79.55	78.94	67.74
ResNet18	84.06	84.13	74.49	76.23	76.18	62.06
SENet18	84.01	83.97	74.55	79.00	78.82	67.00
SKNet	83.56	83.53	73.75	80.23	79.87	68.98
ASFNet	**86.79**	**86.77**	**79.02**	82.05	81.86	71.36

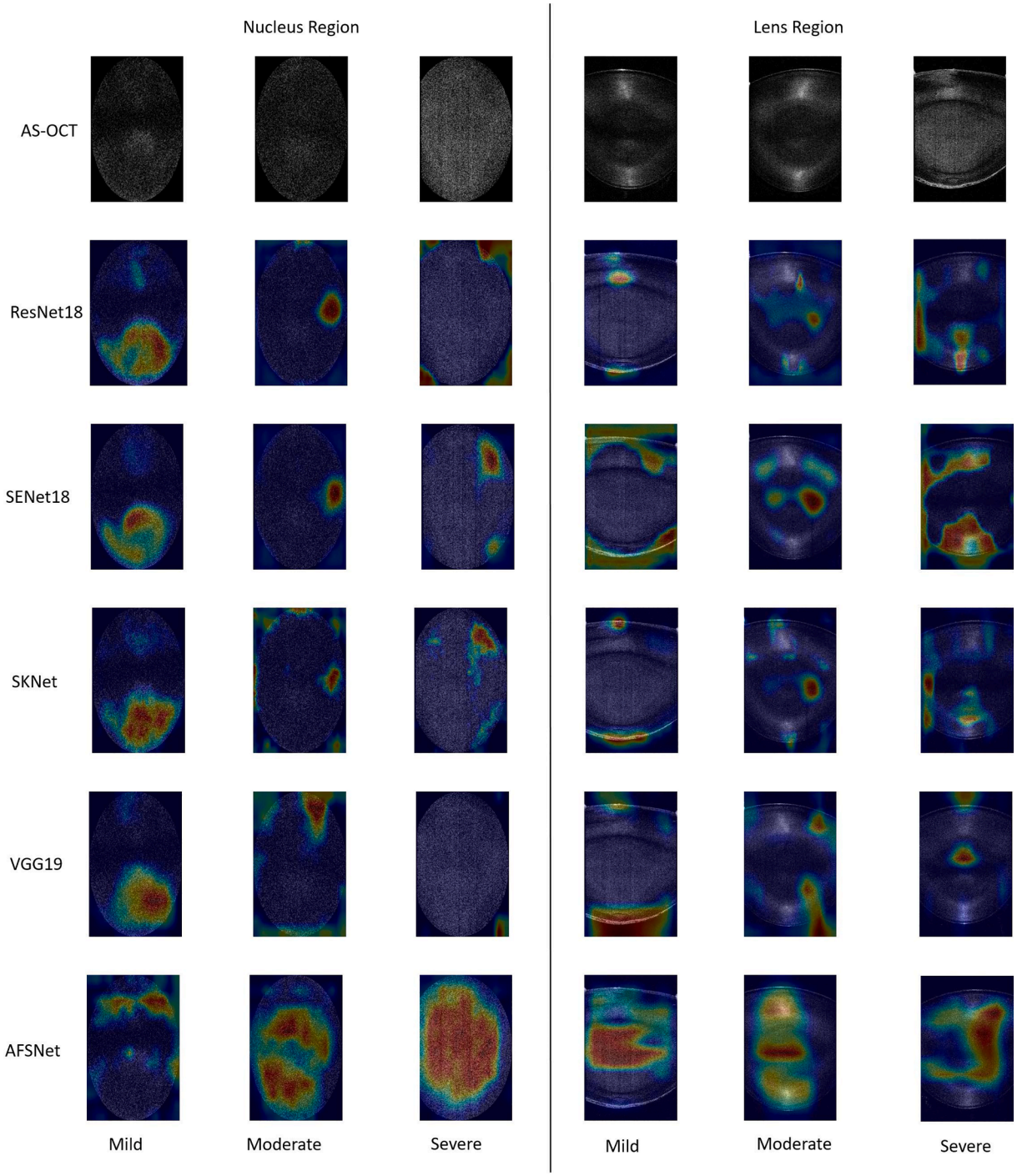


Fig. 10. Examples of the CAMs generated from five advanced CNN models for the nucleus region and the lens region on AS-OCT images. Each region presents three NC severity levels. The maps highlight the discriminative image regions that CNN models learned for specific NC severity levels.

that vanilla CNNs and attention-based CNNs obtain over 90% accuracy, and AFSNet gets the best classification performance, demonstrating its efficacy.

6. Conclusion and future work

This paper proposes an adaptive feature squeeze network (AFSNet) to automatically classify different nuclear cataract severity levels on AS-OCT images. In the AFSNet, we present an efficient adaptive feature squeeze (AFS) module to boost the classification results, comprised of a squeeze block and a global adaptive pooling layer. The experimental results demonstrate that our AFSNet achieves better NC classification performance than strong deep learning baselines and machine learning methods on the clinical AS-OCT image and public OCT datasets. Furtherly, we also demonstrated that deep learning methods obtain better classification results on the nucleus region than the lens region and use the class activation mapping (CAM) technique to visualize and locate what discriminative region information CNN models learned from the two regions, which enhances the Interpretability of our CNN models.

In the future, we will improve our methods and plan to deploy them on real ophthalmic equipment for clinical cataract diagnosis and cataract surgery planning. The code of this paper will be released: https

Table 11
Performance comparison of the AFSNet and state-of-the-art methods on validation dataset of the USUD dataset. The best results in this table are labeled in **bold**.

Method	ACC	Sen	F1	Kappa
HOG-SVM [10]	85.17	75.36	83.14	62.00
LBP-SVM [10]	71.33	48.27	64.04	41.00
Single CNN [10]	95.50	94.65	93.77	88.00
MDFF [10]	93.93	91.76	91.46	83.00
VGG [10]	91.50	91.50	91.50	77.00
Ensemble [10]	95.26	92.84	93.27	87.00
ResNet34 [26]	19.50±1.0	77.9±0.4	–	–
Inception [12]	87.80±2.5	85.50±1.5	–	–
LACNN [12]	90.10±1.4	86.80±1.3	–	–
LACNN-AlexNet [12]	89.40±1.8	85.50±1.3	–	–
LACNN-Inception [12]	92.10±0.9	89.30±1.6	–	–
SENet	94.16±1.5	90.00±1.3	91.49±1.6	91.23±1.2
AFSNet	**96.87**±1.2	**95.08**±1.15	**95.51**±1.4	**95.35**±1.3

Table 12
Performance comparison of the AFSNet and state-of-the-art methods on the testing dataset of the USUD dataset. The best results in this table are labeled in **bold**.

Method	ACC	Sen	F1	Kappa
HOG-SVM [10]	94.93	84.80	84.36	59.00
LBP-SVM [10]	52.20	52.20	42.50	42.00
MDFF [10]	99.60	99.60	99.60	99.00
VGG [10]	91.50	91.50	91.50	77.00
Ensemble [10]	99.30	99.30	99.30	98.00
ResNet34 [26]	96.60	97.80	–	–
CNN [28]	98.60	97.80	–	–
SENet	98.30±0.6	98.30±0.6	98.31±0.5	97.73±0.2
AFSNet	**99.80**±0.1	**99.80**±0.1	**99.80**±0.1	**99.70**±0.1

://github.com/TommyLitlle/AS-OCT-AFSNet.

CRediT authorship contribution statement

Xiaoqing Zhang: Conceptualization, Methodology, Software, Writing original draft, Writing review & editing. **Zunjie Xiao:** Writing review & editing Investigation. **Risa Higashita:** Conceptualization, Supervision, Data curation. **Yan Hu:** Conceptualization. Wan Chen: Supervision, Writing review & editing Investigation. **Jin Yuan:** Conceptualization, Writing review & editing. **Jiang Liu:** Conceptualization, Supervision, Writing review & editing, Project administration, Funding acquisition. Xiaoqing Zhang and Zunjie Xiao equally contributed to this paper.

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.

Acknowledgements

This work was supported in part by Guangdong Provincial Department of Education (2020ZDZX3043), Guangdong Provincial Key Laboratory (2020B121201001), National Natural Science Foundation of China (8210072776), and Shenzhen Natural Science Fund (JCYJ20200109140820699 and the Stable Support Plan Program 20200925174052004).

References

[1] R.R. Bourne, S.R. Flaxman, T. Braithwaite, M.V. Cicinelli, A. Das, J.B. Jonas, Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis, The Lancet Global Health 5 (2017) e888–e897.
[2] M. Caixinha, J. Amaro, M. Santos, F. Perdigão, M. Gomes, J. Santos, In-vivo automatic nuclear cataract detection and classification in an animal model by ultrasounds, IEEE Trans. Biomed. Eng. 63 (2016) 2326–2335.
[3] L. Cao, H. Li, Y. Zhang, L. Zhang, L. Xu, Hierarchical method for cataract grading based on retinal images using improved haar wavelet, Informat. Fusion 53 (2020) 196–208.
[4] A. de Castro, A. Benito, S. Manzanera, J. Mompeán, B. Canizares, D. Martínez, Marín, Three-dimensional cataract crystalline lens imaging with swept-source optical coherence tomography, IOVS 59 (2018) 897–903.
[5] D. Chen, Z. Li, J. Huang, L. Yu, S. Liu, Lens nuclear opacity quantitation with long-range swept-source optical coherence tomography: correlation to locs iii and a scheimpflug imaging-based grading system, Br. J. Ophthalmol. 103 (2019) 1048–1053.
[6] J. Cheng, Sparse range-constrained learning and its application for medical image grading, IEEE Trans. Med. Imaging 37 (2018) 2729–2738.
[7] F. Christian, Baumgartner, Konstantinos, Kamnitsas, Jacqueline, Matthew, P. Tara, Fletcher, Sandra, Smith, Sononet: Real-time detection and localisation of fetal standard scan planes in freehand ultrasound. IEEE Trans. Med. Imaging. (2017).
[8] L.T. Chylack, J.K. Wolfe, D.M. Singer, M.C. Leske, M.A. Bullimore, I.L. Bailey, The lens opacities classification system iii, Arch. Ophthalmol. 111 (1993) 831–836.
[9] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Measur. 20 (1960) 37–46.
[10] 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, https://doi.org/10.1016/j.bspc.2019.101605.
[11] 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. Exp. 10 (2019) 622–641.
[12] 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. Imag. 38 (2019) 1959–1970.
[13] 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.
[14] 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.
[15] 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.
[16] R. Girdhar, D. Ramanan, Attentional pooling for action recognition, in: NIPS, 2017.
[17] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.
[18] 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 (2018) 3821–3833.
[19] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1026–1034. https://doi.org/10.1109/ICCV.2015.123.
[20] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016a, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.
[21] 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, 2016b, pp. 770–778.
[22] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[23] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, K. Keutzer, Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size, 2016. arXiv preprint arXiv:1602.07360.
[24] 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.
[25] 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 (2018) 2716–2732.
[26] 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 (2018) 1122–1131.
[27] Y.N. Kim, J.H. Park, H. Tchah, Quantitative analysis of lens nuclear density using optical coherence tomography (oct) with a liquid optics interface: correlation between oct images and locs iii grading, J. Ophthalmol. 2016 (2016).
[28] F. Li, H. Chen, Z. Liu, X. Zhang, Z. Wu, Fully automated detection of retinal disorders by image-based deep learning, Graefe's Arch. Clin. Exp. Ophthalmol. 257 (2019) 495–505.

[29] 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 (2010) 1690–1698.
[30] H.X. Li, L.D. Xu, Feature space theory — a mathematical foundation for data mining, Knowl.-Based Syst. 14 (2001) 253–257, https://doi.org/10.1016/S0950-7051(01)00103-4. URL: https://www.sciencedirect.com/science/article/pii/S0950705101001034.
[31] H.X. Li, L.D. Xu, J.Y. Wang, Z.W. Mo, Feature space theory in data mining: transformations between extensions and intensions in knowledge representation, Expert Syst. 20 (2003) 60–71, https://doi.org/10.1111/1468-0394.00226. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0394.00226. arXiv: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-0394.00226.
[32] X. Li, W. Wang, X. Hu, J. Yang, Selective kernel networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019b, pp. 510–519.
[33] M. Lin, Q. Chen, S. Yan, etwork in network. ICLR, 2014.
[34] 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 (2018) 356–363.
[35] M. Ozgokce, M. Batur, M. Alpaslan, A comparative evaluation of cataract classifications based on shear-wave elastography and b-mode ultrasound findings, J. Ultrasound 22 (2019) 447–452.
[36] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014. arXiv preprint arXiv:1409.1556.
[37] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Machine Learning Res. 15 (2014) 1929–1958.
[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] Z. Ting, Q. Guo-Jun, X. Bin, W. Jingdong, Interleaved group convolutions for deep neural networks, ICCV, 2017.
[40] W. Wang, J. Zhang, X. Gu, X. Ruan, X. Chen, X. Tan, G. Jin, L. Wang, M. He, N. Congdon, et al., Objective quantification of lens nuclear opacities using swept-source anterior segment optical coherence tomography, Br. J. Ophthalmol. (2021).
[41] 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 (2009) 61–65, https://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.
[42] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
[43] X. Wu, Y. Huang, Z. Liu, W. Lai, E. Long, K. Zhang, J. Jiang, D. Lin, K. Chen, T. Yu, et al., Universal artificial intelligence platform for collaborative management of cataracts, Br. J. Ophthalmol. 103 (2019) 1553–1560.
[44] C. Xu, X. Zhu, W. He, Y. Lu, X. Li, Fully deep learning for slit-lamp photo based nuclear cataract grading, in: MICCAI, 2019a.
[45] X. Xu, L. Zhang, J. Li, Y. Guan, L. Zhang, A hybrid global-local representation cnn model for automatic cataract grading, IEEE JBHI, 2019b.
[46] Y. Xu, L. Duan, D.W.K. Wong, T.Y. Wong, J. Liu, Semantic reconstruction-based nuclear cataract grading from slit-lamp lens images, in: MICCAI, Springer, 2016, pp. 458–466.
[47] Y. Xu, X. Gao, S. Lin, D.W.K. Wong, J. Liu, D. Xu, Automatic grading of nuclear cataracts from slit-lamp lens images using group sparsity regression, in: MICCAI, Springer, 2013, pp. 468–475.
[48] D. Yu, H. Wang, P. Chen, Z. Wei, Mixed pooling for convolutional neural networks, in: International Conference on Rough Sets and Knowledge Technology, 2014.
[49] L. Zhang, J. Li, H. Han, B. Liu, J. Yang, Q. Wang, et al., Automatic cataract detection and grading using deep convolutional neural network, in: 2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC), IEEE, 2017a, pp. 60–65.
[50] T. Zhang, G.J. Qi, B. Xiao, J. Wang, Interleaved group convolutions, in: Proceedings of the IEEE International Conference on Computer Vision, 2017b, pp. 4373–4382. https://doi.org/10.1109/ICCV.2017.469.
[51] X. Zhang, J. Fang, Z. Xiao, H. Risa, W. Chen, J. Yuan, J. Liu, Research on classification algorithms of nuclear cataract based on anterior segment coherence tomography image, Comput. Sci. (2021).
[52] 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, 2020, arXiv preprint arXiv:2012.04830.
[53] X. Zhang, Z. Xiao, R. Higashita, 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: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2020, pp. 662–668. https://doi.org/10.1109/SMC42975.2020.9283218.
[54] X. Zhang, F. Yang, Y. Hu, Z. Tian, W. Liu, Y. Li, W. She, Ranet: Network intrusion detection with group-gating convolutional neural network, J. Network Comput. Appl. 198 (2022) 103266.
[55] X. Zhang, H. Zhao, S. Zhang, R. Li, A novel deep neural network model for multi-label chronic disease prediction, Front. Genet. 10 (2019) 351, https://doi.org/10.3389/fgene.2019.00351.
[56] X. Zhang, X. Zhou, M. Lin, J. Sun, Shufflenet: An extremely efficient convolutional neural network for mobile devices, in: CVPR, 2018, pp. 6848–6856.
[57] 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. https://doi.org/10.1109/CVPR.2016.319.
[58] Y. Zhou, G. Li, H. Li, Automatic cataract classification using deep neural network with discrete state transition, IEEE Trans. Medical Imaging (2019).