DMINet: A lightweight dual-mixed channel-independent network for cataract recognition

Yu Chen[1], Xiao Wu[1], Qiuyang Yang[1], Yuhang Zhao[1], Xiaoqing Zhang[1], Yuan Jin[3], Risa Higashita[1,2], and Jiang Liu[1,4]

[1] Research Institute of Trustworthy Autonomous Systems and Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China
[2] TOMEY Corporation, Nagoya 4510051, Japan
[3] Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou 510060, China
[4] Guangdong Provincial Key Laboratory of Brain-inspired Intelligent Computation, Southern University of Science and Technology, Shenzhen 518055, China

Abstract. Cataracts are the leading causes for visual impairment globally, attracting attention from society. Over the years, researchers have developed many convolutional neural networks (CNNs) to recognize severity levels of cataract based on different ophthalmic image types. However, most existing works focused on improving the cataract recognition performance while ignoring deployment limitations on resource-constrained medical devices. To this problem, this paper proposes a novel dual-mixed channel-independent convolution (DMIConv), which takes advantages of the multiscale convolution kernel sizes in a depthwise convolutional layer and depthwise dilation convolutional layer. Moreover, we use the DMIConv as a drop-in replacement of the depthwise convolution to design a lightweight dual-mixed channel-independent network (DMINet) to recognize severity levels of cataract. To verify the effectiveness of our DMINet, extensive experiments are conducted on a clinical anterior segment optical coherence tomography (AS-OCT) dataset of nuclear cataract (NC) and a public OCT dataset. The results show that our proposed DMINet keeps a better tradeoff between the model complexity and the classification performance than state-of-the-art methods, e.g., DMINet outperforms MixNet by **3.34%** of accuracy while reduced by **4.87%** in parameters.

Keywords: cataract · DMIConv · classification · convolution · lightweight.

1 Introduction

Cataract is the first cause of blindness and the main cause of visual impairment in the world. With the global ageing population, the number of cataract patients will grow rapidly because the cataract is also an age-related ocular disease. Cataract patients can improve their vision and life quality with early intervention

and cataract surgery, which are two main methods to reduce the burden on society [1].

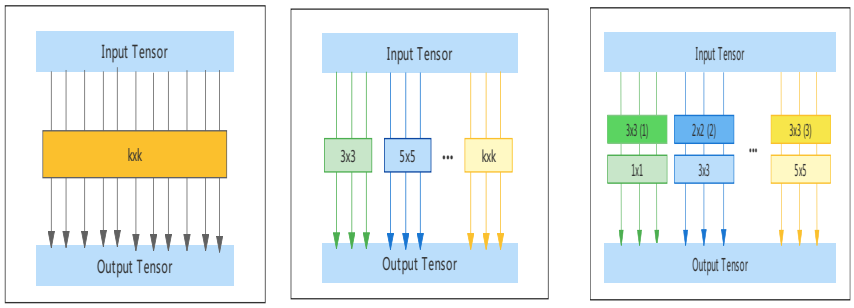
Over the years, scholars have developed a lot of advanced machine learning methods for automatic cataract screening and recognition based on different ophthalmic images [24], aiming at to detect cataract as soon as possible and improve the efficiency and precision of clinicians in clinical cataract diagnosis. For example, Li et al. [11] proposed a slit lamp image-based nuclear cataract (NC) grading framework, which is a combination of the lens detection, feature extraction, and grading. Xu et al. [19] used the bag-of-features (BOF) method to extract features for improving the grading results of NC based on slit images. Zhang et al. [21] used Watershed and Markov random fields (MRF) methods for posterior subcapsular cataract (PSC) detection based on anterior retro-illumination images.

With the advent of deep learning techniques, deep neural networks have widely been utilized to recognize the severity levels of cataract. [17] uses faster R-CNN to classify the severity levels of NC based on slit lamp images in a end-to-end manner. Zhang et al. [25] developed an adaptive feature squeeze network (AFSNet) to classify severity levels of NC based on anterior segment optical coherence tomography (AS-OCT) images. Xu et al. [18] proposed a hybrid global-local representation convolutional neural network (CNN) for fundus image-based cataract screening. Most deep learning-based cataract classification works mainly focused on the improvements of cataract classification and rare works aimed at developing lightweight networks for automatic cataract classification, which is easy to be deployed on resource-constrained medical devices and mobile devices.

In recent years, researchers have spent much effort on designing lightweight and efficient convolutional neural network (CNN) architectures for computer vision tasks. SqueezeNet [8] is the first research to focus on lightweight CNN architecture design. Following SqueezeNet, various advanced lightweight CNNs are presented to tackle different learning tasks [20, 15, 13] based on efficient convolution methods. Depthwise convolution method is a widely method to construct efficient CNNs, such as MobileNets [7, 13], ShuffleNets [22], NASNet [27], and EfficientNet [14]. Unlike the standard convolution method, the depthwise convolution (DW-Conv) applies a convolution kernel to each channel separately, reducing the computational burden by a factor of C (C denotes the number of channels). MixNet [15] proposed a mixed depthwise convolution (MixConv) method to prompt the DW-Conv, which adopts multiscale convolution kernel sizes in a depthwise convolutional layer by splitting channels into several groups and applying different kernel sizes to each group of channels. However, MixConv uses large convolution kernel size, increasing computational cost and parameters.

In order to achieve a good balance between the performance and the complexity of model, this paper develops a novel dual-mixed channel-independent convolution (DMIConv) method, which exploits the advantages of multiscale kernel sizes in a depthwise convolutional layer and depthwise dilation convolutional layer sequentially by partitioning channels into multiple groups according to the

number of different convolution kernel sizes. Furthermore, we treat the DMIConv as a simple drop-in replacement of the vanilla depthwise convolution method and then construct an effective yet lightweight dual-mixed channel-independent network (DMINet) to recognize severity levels of cataract. We conduct experiments on a clinical AS-OCT dataset of NC. Results show that DMINet better balances the model complexity and the classification performance through comparisons to other advanced lightweight CNNs. To further validate the effectiveness of our DMINet, a publicly available OCT is also adopted, and experimental results show that our DMINet achieves competitive performance compared with strong baselines and published methods.



Depthwise convolution MixConv Our proposed DMIConv

Fig. 1. Dual-mixed channel-independent convolution (DMIConv): Unlike depthwise convolution sets the same convolution kernel size for channels and MixConv applies different convolution kernel sizes to channels by group strategy, our DMIConv not only use mixed convolution kernel sizes but also exploits advantages of depthwise convolution and depthwise dilated convolution.

2 Method

As previously introduced, the key motivation for devising the dual-mixed channel-independent convolution (DMIConv) method is to improve both NC classification results and efficiency. In this section, we will first introduce DMIConv in detail.

2.1 Dual-mixed channel-independent convolution

This section first revisits the original depthwise convolution method, as shown in Fig. 1(a). Given any input feature map $D_H \times D_W \times M$ and generates the output feature map $D_H \times D_W \times N$ where D_H, D_W, M, and N represents the spatial height, spatial width, the number of input channels, and the number of output channels (Specifically, N is equal to M). In the depthwise convolution method,

we apply a filer with convolution kernel size $K \times K$ to each channel separately, which can be formulated as follows:

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} * F_{k+i-1,l+j-1,m} \tag{1}$$

The original depthwise convolution only uses the same convolution kernel size for all channels, e.g., the commonly-used 3×3 convolution kernel. It overlooks that multiscale convolution kernel sizes that can boost the classification performance. To alleviate this shortcoming of depthwise convolution, Tan et al. designed a variant of depthwise convolution method, termed MixConv. As shown in Fig. 1(b), MixConv divides channels into multiple groups based on the number of different convolution kernel sizes and then applies each convolution kernel size to each group. One shortcoming of MixConv is that it uses large convolution kernel sizes, which increases the computational cost.

To overcome the shortcoming of MixConv, this paper designs a dual-mixed channel-independent convolution method, as shown in Fig. 1(c), which is a combination of three efficient convolution methods: multiscale convolution kernel sizes, depthwise convolution (DW-Conv), and depthwise dilated convolution (DW-D-Conv). Like MixConv, we first adopt the group strategy to partition the channels into multiple groups according to different convolution kernel sizes. Specifically, input feature maps are partitioned into z groups: $< D_H \times D_W \times m_1, D_H \times D_W \times m_2, ..., D_H \times D_W \times m_z >$, where $m_1 + m_2 + ... + m_z = M$.

In each convolution kernel size group, we use a DW-D-Conv and a DW-Conv to continuously capture long-range and local patterns from input feature maps. Previous works have demonstrated that dilated convolution method can increase the receptive field size with smaller convolution kernel sizes through comparisons to regular convolution and pointwise convolution (Conv1 $\times$ 1) method. Hence, this paper uses a depthwise dilated convolution to build the long-range relationships among feature representations in a feature map. It is followed by a depthwise convolution, which is used to construct local relationships among feature representations in a feature map. Moreover, our proposed DMIConv can be taken as a simple alternative to depthwise convolution for designing lightweight CNN architectures.

Furthermore, this paper also designs three variants of DMIConv (DW-D-Conv+DW-Conv): DW-D-Conv+DW-Conv+Conv1 $\times$ 1, DW-Conv+Conv1 $\times$ 1 , and DW-D-Conv+Conv1 $\times$ 1 to test which affects the performance of our DMIConv. Conv1 $\times$ 1, also named pointwise convolution method, which is a widely-used method to construct lightweight CNNs [6, 13].

2.2 Network architecture

Fig 2 presents a general architecture of our DMINet based on the proposed DMIConv, which adopts the same network depth as MobileNetV1 used to ensure performance comparison fairness. In the DMINet, we first use the same stem as MixNet adopts, followed by a series of dual-mixed channel-independent

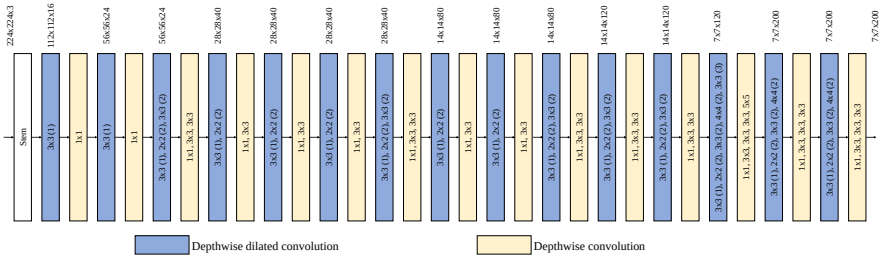


Fig. 2. A framework of the dual-mixed channel-independent network (DMINet), which is built on the proposed DMIConv.

convolutional layers, a global average pooling (GAP) layer, and a classifier layer. This paper uses the commonly-used softmax and cross entropy loss functions as the classifier and loss, respectively.

3 Metrics and experiment setting

3.1 Metrics

Following previous works on cataract classification tasks [23, 2], this paper uses four commonly-adopted evaluation measures to assess the classification performance of methods: accuracy (ACC), sensitivity (Sen), precision (PR), and F1 measure. ACC and F1 are usually utilized to evaluate the overall performance of a method. Sen is a very significant indicator for clinical disease diagnosis. The following equations can obtain these four evaluation measures:

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \tag{2}$$

$$Sen = \frac{TP}{TP + FP} \tag{3}$$

$$PR = \frac{TP}{TP + FN} \tag{4}$$

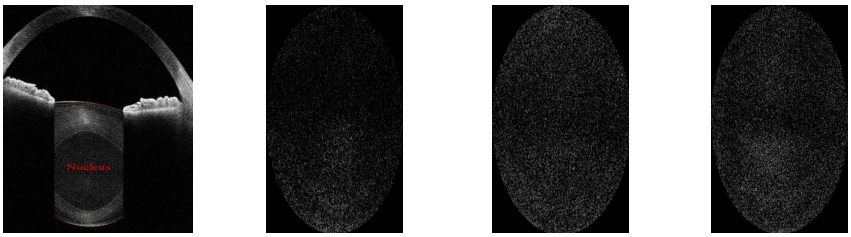
$$F1 = \frac{2 \times Sen \times PR}{Sen + PR} \tag{5}$$

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

3.2 Dataset

This paper uses two ophthalmic image dataset and one fundus image dataset to demonstrate the performance of the DMINet: a clinical AS-OCT dataset of NC and a public OCT dataset [10].

AS-OCT dataset. It is a clinical dataset of NC, which is collected by CASIA2 device (TOMEY Company, Japan). The dataset contains 534 participants: 422 right eyes and 440 left eyes, and the total number of AS-OCT images (as shown in Fig 3 is 16,201. It contains three severity levels of NC: normal, mild, and severe, and we split it into three disjoint datasets based on the participant level: training, validation, and testing. Table 1 lists three severity level distributions of NC for disjoint datasets.



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

Fig. 3. Three severity levels of nuclear cataract (NC) based on AS-OCT images (a). Normal (b) without opacity; Mild NC (c) with slight opacity but is asymptomatic; Severe NC (d) with opacity and is symptomatic.

Table 1. Detailed nuclear cataract level distributions on the AS-OCT dataset

Dataset	Normal	Mild	Severe	total
Training	1004	2872	5518	9394
Validation set	345	1306	1740	3391
Testing	254	664	2498	3416
Total	1603	4842	9756	16021

UCSD dataset. It is a publicly available OCT of age related macular degeneration (AMD) and diabetic macular edema (DME), which is collected from Medical Center Ophthalmology Associates, the Shanghai First People's Hospital, Shiley Eye Institute of the University of California San Diego, the California Retinal Research Foundation, and Beijing Tongren Eye Center. The training set contains 83,484 OCT images from 4,686 patients: 37,205 images of choroidal neovascularization (CNV), 11,348 images of diabetic macular edema (DME), 8,617 images of drusen, and 51,140 images of healthy. The testing dataset contains 1000 images and each class has the number of OCT images. In this paper, we follow the same training dataset splitting strategy in[25, 26].

Glaucoma dataset. It is a fundas image dataset of glaucoma, which is collected from Beijing Tongren Hospital. The dataset contains 4,854 fundus images: 3,143 images of normal people and 1711 images of patient. The distributions are shown in Table 2.

Table 2. Detailed glaucoma distributions on the dataset

Dataset	Normal	patient	total
Training	1885	1026	2911
Validation set	629	342	971
Testing	629	343	972
Total	3143	1711	4854

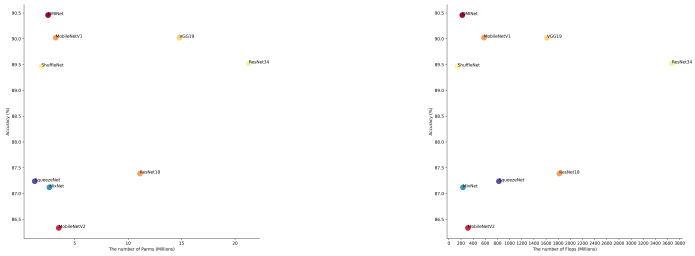
3.3 Experimental setting

We use Pytorch, python, scikit-learn, and OpenCV software packages to our proposed DMINet and comparable deep neural networks. This paper uses the stochastic gradient descent (SGD) method to optimize the learnable weights of deep neural networks with the default settings. The training epochs and batch size are set for 150 and 64 correspondingly. We set the initial learning rate for 0.0025, which is decreased by a factor of 5 every 25 epochs. This paper sets the final learning rate at 0.0001 when training epochs are more than 130. This paper follows the common data augmentation strategies to preprocess the training images (such as cropping and flipping), and resizes the images of two datasets into 224×224. All experiments are run on a server with one NVIDIA TITAN GPU, 10 GB DDR memory, and the operating system is Ubuntu.

4 Result analysis and discussion

4.1 Performance comparison on the AS-OCT dataset

This paper first compares DMINet with commonly-used CNNs and lightweight CNNs, as shown in Table 3. It can be seen that our DMINet achieves the best performance of NC recognition on four evaluation measures: 90.46% of accuracy, 90.98% of F1, 90.46% of precision, and 92.76% of sensitivity, respectively. For instance, with less or comparable parameters and flops (millions) than MixNet and SqueezeNet, it significantly improves **3.34%** and **3.22%** gains of accuracy. Fig. 4(a) lists the relationships between parameters and accuracy, and our DMINet achieves a better trade-off between the NC classification performance and memory requirement than other comparable CNNs. Fig. 4(b) shows the relationships between flops and accuracy; DMINet requires fewer flops than other CNNs except for ShuffleNet while obtaining the best NC classification performance. In general, the classification results demonstrate that our proposed



(a) (b)

Fig. 4. (a) Relationship between the performance of DMINet and the total number parameters (Millions), compared to state-of-the-art CNNs; (b) Relationship between the performance of DMINet and the total number flops (Millions), compared to state-of-the-art CNNs.

DMINet keeps a better trade-off between the classification performance and the complexity through comparisons to other CNNs, which is feasible to be deployed on medical devices.

Table 3. Performance comparison of DMINet, advanced CNNs, and lightweight CNNs on the AS-OCT dataset

Model	Accuracy	F1	PR	Sen	Parms (M)	Flops (M)
ResNet18	87.39	88.09	87.39	89.90	11.1	1818
ResNet34	89.52	90.03	89.52	91.60	21.3	3671
VGG19	90.02	90.59	90.02	92.67	14.8	1615
MobileNetV1	90.02	90.41	90.02	91.12	3.21	577.8
MobileNetV2	86.33	87.29	86.33	90.65	3.50	314.2
ShuffleNet	89.46	90.04	89.46	91.95	**1.88**	**145.9**
MixNet	87.12	87.95	87.12	90.69	2.62	232.0
SqueezeNet	87.24	88.03	87.24	90.22	1.25	823.4
DMINet	**90.46**	**90.98**	**90.46**	**92.76**	2.50	221

4.2 Performance comparison on the UCSD dataset

Table 4 lists the classification results of DMINet, original CNNs, state-of-the-art attention CNNs, and published works on the UCSD dataset. It can be observed that our DMINet achieved the best classification results of AMD and DME by using OCT images among machine learning methods and deep learning methods. Compared with HOG-SVM, DMINet achieved **25.50%** increase in accuracy and **46.88%** increase in sensitivity, respectively. Moreover, Our DMINet obtains slight improvements through comparisons to state-of-the-art attention CNNs. Overall, the classification results in Table 4 also demonstrate

that DMINet achieves a better tradeoff between the model complexity and the classification performance than machine learning methods and deep learning methods.

Table 4. Performance comparison of the DMINet and state-of-the-art methods on UCSD dataset

Method	ACC	Sen	F1
LBP-SVM [3]	71.33	48.27	64.04
HOG-SVM [4]	78.90	66.20	–
MDFF [3]	93.93	91.76	91.46
VGG16 [3]	91.50	91.50	91.50
ResNet34 [10]	80.50	78.30	–
Inception [4]	90.30	90.00	–
LACNN [4]	90.20	88.10	–
LACNN-Inception [4]	93.00	91.60	–
SENet [9]	94.16	90.00	91.49
SPANet [5]	94.11	89.83	91.32
BAM [12]	94.89	91.95	92.69
CBAM [16]	94.20	89.74	91.30
MPANet [26]	96.74	95.12	95.39
DMINet	96.86	95.15	95.45

4.3 Performance comparison on the Glaucoma dataset

Table 5 lists the classification results of DMINet, MobileV2 and three kind of variants of DMINet. It can be observed that our DMINet achieved the best classification results of AMD and DME by using OCT images among machine learning methods and deep learning methods. Compared with MobileV2, DMINet achieved **1.34%** increase in accuracy. And the DMINet achieves a better tradeoff between the model complexity and the classification performance.

Table 5. Performance comparison of DMINet, MobileV2, and three variants on the glaucoma dataset

Convolution method	ACC	Parms (M)	Flops (M)
DW-D-Conv+DW-Conv+Conv1 $\times$ 1	90.22	3.84	457.0
DW-Conv+Conv1 $\times$ 1	90.53	3.95	467.8
DW-D-Conv+Conv1 $\times$ 1	91.35	3.77	446.2
DW-D-Conv+DW-Conv (DMIConv)	**94.34**	2.50	**220.7**
MovileV2	93.00	**2.23**	313.0

4.4 Ablation study

To further examine the effectiveness of our DMIConv, Table 6 lists the classification performance of DMIConv and its three variant based on MobliNetV1 architecture. We can see that compared with the other three variants, our proposed DMIConv achieves higher accuracy while requiring fewer parameters and flops. The key difference between our method and other comparable convolution methods is that the other three comparable convolution methods adopt the pointwise convolution method (Conv1$\times$1), which uses the fully-connected way to build connections among channels but also increases feature map redundancies. It is the main reason why DMIConv gets better performance than its variants.

Table 6. Performance comparison of DMIConv and its variants on the AS-OCT dataset

Convolution method	ACC	Parms (M)	Flops (M)
DW-D-Conv+DW-Conv+Conv1 $\times$ 1	88.62	3.84	457.0
DW-Conv+Conv1 $\times$ 1	88.73	3.95	467.8
DW-D-Conv+Conv1 $\times$ 1	89.60	3.77	446.2
DW-D-Conv+DW-Conv (DMIConv)	90.46	**2.50**	**220.7**

5 Conclusion and future work

This paper proposes an efficient dual-mixed channel-independent convolution (DMIConv), which takes advantages of different convolution kernel sizes, depthwise dilated convolution, and depthwise convolution. We construct a lightweight dual-mixed channel-independent network (DMINet) based on the DMIConv method to recognize the severity levels of nuclear cataract. Experimental results on an AS-OCT dataset show that our proposed DMINet obtains significantly better classification results and efficiency than state-of-the-art lightweight CNNs. Moreover, the results of another public OCT dataset demonstrate the effectiveness of our method. In the future, we plan to deploy our DMINet on an actual medical device to further test its performance.

References

1. Bourne, R.R., Flaxman, S.R., Braithwaite, T., Cicinelli, M.V., Das, A., Jonas, J.B., Keeffe, J., Kempen, J.H., Leasher, J., Limburg, H., et al.: 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**(9), e888–e897 (2017)
2. Cao, L., Li, H., Zhang, Y., Zhang, L., Xu, L.: Hierarchical method for cataract grading based on retinal images using improved haar wavelet. Information Fusion **53**, 196–208 (2020)

3. Das, V., Dandapat, S., Bora, P.K.: Multi-scale deep feature fusion for automated classification of macular pathologies from oct images. Biomedical Signal Processing and Control **54**, 101605 (2019). https://doi.org/https://doi.org/10.1016/j.bspc.2019.101605
4. Fang, L., Wang, C., Li, S., Rabbani, H., Chen, X., Liu, Z.: Attention to lesion: Lesion-aware convolutional neural network for retinal optical coherence tomography image classification. IEEE transactions on medical imaging **38**(8), 1959–1970 (2019)
5. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. PAMI **37**(9), 1904–1916 (2015)
6. He, Z., Zhang, X., Cao, Y., Liu, Z., Zhang, B., Wang, X.: Litenet: Lightweight neural network for detecting arrhythmias at resource-constrained mobile devices. Sensors **18**(4), 1229 (2018)
7. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
8. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. arXiv preprint arXiv:1602.07360 (2016)
9. Jie, Hu, Li, Shen, Samuel, Albanie, Gang, Sun, Enhua, Wu: Squeeze-and-excitation networks. TPAMI (2019)
10. Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., et al.: Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell **172**(5), 1122–1131 (2018)
11. Li, H., Lim, J.H., Liu, J., Mitchell, P., Tan, A.G., Wang, J.J., Wong, T.Y.: A computer-aided diagnosis system of nuclear cataract. IEEE Transactions on Biomedical Engineering **57**(7), 1690–1698 (2010)
12. Park, J., Woo, S., Lee, J.Y., Kweon, I.S.: Bam: Bottleneck attention module. arXiv preprint arXiv:1807.06514 (2018)
13. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4510–4520 (2018)
14. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)
15. Tan, M., Le, Q.V.: Mixconv: Mixed depthwise convolutional kernels. arXiv preprint arXiv:1907.09595 (2019)
16. Woo, S., Park, J., Lee, J.Y., So Kweon, I.: Cbam: Convolutional block attention module. In: ECCV. pp. 3–19 (2018)
17. Xu, C., Zhu, X., He, W., Lu, Y., He, X., Shang, Z., Wu, J., Zhang, K., Zhang, Y., Rong, X., et al.: Fully deep learning for slit-lamp photo based nuclear cataract grading. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 513–521. Springer (2019)
18. Xu, X., Zhang, L., Li, J., Guan, Y., Zhang, L.: A hybrid global-local representation cnn model for automatic cataract grading. IEEE journal of biomedical and health informatics (2019)
19. Xu, Y., Gao, X., Lin, S., Wong, D.W.K., Liu, J., Xu, D., Cheng, C.Y., Cheung, C.Y., Wong, T.Y.: Automatic grading of nuclear cataracts from slit-lamp lens images using group sparsity regression. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 468–475. Springer (2013)

20. Yang, F., Zhang, X., Zhu, Y.: Pdnet: a convolutional neural network has potential to be deployed on small intelligent devices for arrhythmia diagnosis. Computer Modeling in Engineering & Sciences **125**(1), 365–382 (2020)
21. Zhang, W., Li, H.: Lens opacity detection for serious posterior subcapsular cataract. Medical & biological engineering & computing **55**(5), 769–779 (2017)
22. Zhang, X., Zhou, X., Lin, M., Sun, J.: Shufflenet: An extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6848–6856 (2018)
23. Zhang, X., Fang, J., Xiao, Z., Risa, H., Chen, W., Yuan, J., Liu, J.: Research on classification algorithms of nuclear cataract based on anterior segment coherence tomography image. Computer Science (2022). https://doi.org/10. 11896 / jsjkx. 201100085
24. Zhang, X., Hu, Y., Xiao, Z., Fang, J., Higashita, R., Liu, J.: Machine learning for cataract classification/grading on ophthalmic imaging modalities: a survey. Machine Intelligence Research (2022). https://doi.org/http://doi.org/10.1007/s11633-022-1329-0
25. Zhang, X., Xiao, Z., Higashita, R., Hu, Y., Chen, W., Yuan, J., Liu, J.: Adaptive feature squeeze network for nuclear cataract classification in as-oct image. Journal of Biomedical Informatics **128**, 104037 (2022)
26. Zhang, X., Xiao, Z., Li, X., Wu, X., Sun, H., Yuan, J., Higashita, R., Liu, J.: Mixed pyramid attention network for nuclear cataract classification based on anterior segment oct images. Health Information Science and Systems **10**(1), 1–12 (2022)
27. Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V.: Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8697–8710 (2018)