



题 目:	基于非对称异构核重构网络模型的 致盲性疾病识别
姓 名:	巫 晓
学 号:	11912803
系 别:	计算机科学与工程系
专 业:	计算机科学与技术
指导教师:	刘 江

1

诚信承诺书

1. 本人郑重承诺所呈交的毕业设计(论文),是在导师的指导下,独立进行研究工作所取得的成果,所有数据、图片资料均真实可靠。

2. 除文中已经注明引用的内容外,本论文不包含任何其他人或集体已经发表或撰写过的作品或成果。对本论文的研究作出重要贡献的个人和集体,均已在文中以明确的方式标明。

3. 本人承诺在毕业论文(设计)选题和研究内容过程中没有抄袭他人研究成果和伪造相关数据等行为。

4. 在毕业论文(设计)中对侵犯任何方面知识产权的行为,由本人承担相应的法律责任。

作者签名:

_____年____月____日

基于非对称异构核重构网络模型的致盲性 疾病识别

巫 晓

(计算机科学与工程系 指导教师: 刘江)

[摘要]: 致盲性疾病会长期影响患者视力和正常工作生活, 给社会带来负担, 在全球范围内引起了广泛关注。眼科医生通常使用彩色眼底图像或是 OCT 图像进行致盲性疾病的诊断工作, 容易受主观因素影响。近年来, 研究者提出了许多先进的卷积神经网络来辅助眼科医生高效诊断致盲性疾病。但是, 大部分方法专注于构建更加复杂的神经网络模型从而提高准确率, 忽视了眼科医疗设备本身硬件性能带来的限制, 使得模型难以部署。因此, 本文基于多尺度卷积核和异构卷积核提出了多尺度异构卷积模块 (multi-scale heterogeneous convolution block, MHCBLOCK)。该模块使用多尺度卷积核获取不同分辨率下的病理信息, 并使用异构卷积核捕获垂直和水平方向的病理信息。基于 MHCBLOCK, 本文还构建了一个轻量高效的多尺度异构卷积网络 (multi-scale heterogeneous convolution network, MHCNet) 来进行致盲性疾病的识别分类工作。训练完成后, 本文将模块中同尺度的异构卷积核融合, 使得模型准确率提升的同时仍能保持较少的参数量。为验证本文所提方法的有效性, 本文分别在四个临床数据集上进行大量实验, 分别是 AS-OCT 数据集、UCSD 数据集、ISIC2018 数据集和

BTC 数据集。实验结果表明，MHCNet 相较于经典的轻量级网络如 EfficientNet 和重构卷积网络如 ACNet 在准确率与模型复杂度之间有着更好的平衡。例如，在 AS-OCT 数据集上，MHCNet 在准确率方面分别比 EfficientNet 和 ACNet 提升了 3.53%和 3.55%，但模型参数量分别降低了 10.8%和 82.4%。

[关键词]： 轻量化；多尺度；异构卷积；致盲性疾病；医疗诊断

[ABSTRACT]: Blinding diseases will affect patients' vision and regular work and life for a long time and burden society, which has attracted wide attention worldwide. Ophthalmologists usually use color fundus images or OCT images to diagnose blinding diseases, which are easily affected by subjective factors. In recent years, researchers have proposed a lot of state-of-the-art convolutional neural networks to help ophthalmologists efficiently diagnose blinding diseases. However, most of the methods focus on building more complex neural networks to improve accuracy, ignoring the constraints brought by the hardware of ophthalmic medical equipment itself, which makes it difficult to implement. Thus, we propose a multi-scale heterogeneous convolution block (MHCBlock) based on the multi-scale convolution kernel and the heterogeneous convolution kernel. Our MHCBlock obtains pathological information in different resolutions with a multi-scale convolution kernel and uses a heterogeneous convolution kernel to capture horizontal and vertical square pathological information. Based on the MHCBlock, we also propose a lightweight yet efficient multi-scale heterogeneous convolution network (MHCNet) to classify blinding diseases. After training, we merge the heterogeneous convolution kernel in the same scale per block, which improves the model accuracy while maintaining fewer parameters. To verify the effectiveness of our method, we conduct abundant experiments on four clinical data sets. They are AS-OCT, UCSD, ISIC2018, and BTC. The experimental results show that

MHCNet achieves a better balance of accuracy and model complexity compared with various lightweight networks such as EfficientNet, and reparameterization networks such as ACNet. For example, on AS-OCT dataset, MHCNet improves 3.53% and 3.55% in terms of accuracy compared with EfficientNet and ACNet, but reduces 10.8% and 82.4% in terms of model parameters.

[Keywords]: Lightweight; Multi-scale; Heterogeneous convolution; Blinding diseases; Medical diagnosis

目录

1. 引言.....	9
2. 相关工作.....	11
2.1 异构卷积.....	11
2.2 多尺度卷积.....	12
2.3 重参数化.....	12
3. 多尺度异构卷积网络.....	13
3.1 异构卷积.....	13
3.2 多尺度卷积.....	15
3.3 重参数化.....	16
3.4 多尺度异构卷积网络.....	16
4. 实验设置与数据集介绍.....	17
4.1 AS-OCT 数据集.....	18
4.2 UCSD 数据集.....	18
4.3 ISIC2018 数据集.....	19
4.4 BTC 数据集.....	19
4.5 评价指标.....	19
4.6 实验设置.....	20
5. 实验结果与分析.....	20
5.1 AS-OCT 数据集的分类结果分析.....	20
5.2 UCSD 数据集的分类结果分析.....	22

5.3 其他疾病数据集分类结果分析.....	23
5.4 消融实验.....	23
6. 结语.....	25
参考文献.....	26
致谢.....	30

1. 引言

由于多数致盲性疾病的发病率随年龄增长而增加,在全球社会步入老龄化的当下,因致盲性疾病而导致视力损伤的患者数量也在不断上升,引起了全世界广泛关注。根据世界卫生组织(world health organization, WHO)发表的世界视觉报告,至2020年,全球至少有22亿患者饱受视力损伤的困扰,其中至少有10亿患者是本可以通过预防或治疗来避免或改善其视力损伤^[1]。其中,青光眼、白内障、糖尿病视网膜病变和年龄性黄斑变性的患者占绝大多数。在疾病早期发现并接受治疗能够有效改善患者视力并延缓疾病进程,减小社会负担。但是,由于眼科医生资源的稀缺^[2],一位临床医生往往要面对多位患者,这使得医生需要长时间工作,从而导致其判断能力下降。因此,如何准确客观地诊断致盲性疾病是临床治疗中亟需解决的问题和挑战。

为了帮助临床医生高效地诊断致盲性疾病,近几年来研究者提出了许多基于机器学习或深度学习的致盲性疾病诊断算法。Fu等^[3]提出了一个多上下文神经网络对AS-OCT图像下的闭角青光眼进行自动诊断。Zhang等^[4]在AS-OCT图像上提出了一个混合金字塔注意力网络用于核性白内障自动分类。Burlina等^[5]基于彩色眼底图像使用深度神经网络进行年龄性黄斑变性的自动分级。Xiao等^[6]提出一个多任务卷积神经网络用于三维AS-OCT影像下的白内障自动检测。虽然现有研究已经取得了可观的进展,但是大部分已有工作都专注于构建复杂的神经网络模型以提高准确率,忽视了眼科医疗设备本身硬件资源的局限性。不仅如此,这些卷积神经网络模型大都使用单个卷积尺度,无法有效利用不同尺度下的病理信息。

近年来,高效轻量的网络结构设计受到了工业界和学术界的广泛关注。研究者已经提出了多种轻量的网络结构如: MobileNets^[7-9]、IGCNets^[10-12]、ESPNets^[13-14]、ShuffleNets^[15-16]、MixNet^[17]和EfficientNet^[18]等等。深度可分离卷积(depthwise separable convolution)是一种轻量级神经网络中十分常见的卷积方式,由深度卷积(depthwise convolution)和逐点卷积(pointwise convolution)组成。它通过深度卷积提取每个通道的空间信息,然后再由逐点卷积进行通道间的信息融合。相较于原始的三维卷积,深度可分离卷积拥有更少的参数量和计算量。在此基础上,部分网络如MixNet考虑到逐点卷积仍有较大的参数量和计算量,删去了逐点卷

积，仅使用深度卷积来进行特征提取，进一步减少了模型复杂度。轻量级网络如 MobileNets^[7-9] 通常使用 3×3 的小卷积核，导致其仅关注于低分辨率，而忽视了高分辨率下的病理信息。最近的研究工作^[18-20] 已经证明了更大尺度如 5×5 、 7×7 的卷积核有着提高模型性能的潜力。但是，大部分已有的轻量级神经网络模型仅使用小卷积核，忽视了高分辨率下的病理信息。我们希望能够同时获取高低分辨率下的病理信息，提升模型性能，并且在模型性能和模型参数量之间取得较好的平衡。

另一方面，为了在不增加模型复杂度的情况下提升模型性能，研究者基于卷积神经网络中各模块之间的关系提出了多种重参数化方法与网络。Sergey 等^[21] 提出了一种狄拉克权重重参数化方法以及 DiracNets，将原有卷积操作进行拓展。Ding 等^[22] 提出了使用一维卷积核增强方形卷积核骨架的重参数化方法，取得较好效果，但是仅在三维卷积核中进行实验。RepVGG^[23] 中提出了一种类似于 DiracNets 的重参数化方法，但是实现方法不同并且效果更好。大部分已有的重参数化方法关注于模型结构之间的关系，没有考虑轻量化模型的需求，而我们希望能将重参数化方法应用于轻量级模型，在不提升模型复杂度的同时提升模型性能，便于部署在医疗设备上。

本文提出多尺度异构卷积模块 (MHCBLOCK)，是一个引入了多尺度卷积核和异构卷积核的即插即用模块，该模块拥有以下优点：

- MHCBLOCK 是一个即插即用模块，拥有唯一的输入和输出，不需要在模块外增加分支。
- MHCBLOCK 使用多尺度卷积核，能够同时获取高分辨率和低分辨率病理信息。
- MHCBLOCK 使用异构卷积核捕获垂直和水平方向的病理特征，并且在推理时通过重参数化方法将异构卷积核融合，提升模型性能且不增加模型推理复杂度。

对于传统的多分支模型如 Inceptions^[24-25]，虽然它们取得了比较好的模型性能，但是多分支会带来大量的计算开销，使得其对于资源有限的医疗设备不太友好。考虑到多分支模型的参数仅在训练时变化，我们可以通过重参数化的方法对模型训练和模型推理进行解耦，改变模型推理结构，在保持多分支训练学习到的信息的同时降低模型复杂度，提升模型推理速度。MHCBLOCK 中的异构卷积核存

在多分支结构，我们在推理时通过重参数化方法将其融合为单个方形卷积，避免了多分支结构在推理过程中的大量开销。

我们将本文主要贡献总结如下：

- 本文提出了 MHCBlock，一个新颖的轻量级神经网络模块。MHCBlock 可以通过相应深度卷积核直接构建，不需要其他的超参数。此外，MHCBlock 既能通过多尺度卷积核同时捕获不同分辨率病理信息，又能通过异构卷积核同时捕获垂直和水平方向的病理信息，增强病理特征学习能力。
- 本文基于 MHCBlock 构建了一个轻量且高效的神经网络模型，MHCNet，来对致盲性疾病进行快速诊断。并且，我们将 MHCNet 在四个疾病数据集上进行了充分实验。实验结果表明，相较于其他先进的轻量级网络和重参数化网络，MHCNet 能够在模型性能与模型复杂度之间取得更好的平衡。
- 本文通过消融实验验证了 MHCBlock 中异构卷积核的有效性，并且验证了其在其他疾病数据集上的泛化能力。

2. 相关工作

2.1 异构卷积

异构卷积通常被用作为原始方形卷积的替代品，其通过不同形状的卷积核组合达到近似于方形卷积核的效果，并且能够减小模型参数量和计算量。在部分工作中^[25-28]，研究者将 $K \times K$ 的方形卷积核分解一个 $K \times 1$ 垂直卷积核和一个 $1 \times K$ 的水平卷积核，保持较好性能的同时减少了参数量和计算量。其原理是：当一个二维矩阵的秩等于 1 时，其可以被等价分解为两个一维向量。但是，深度学习中大部分的卷积核秩并不为 1，直接进行卷积核分解的话会导致大量的信息丢失^[28]。针对这个问题，Jaderberg 等^[27]通过滤波器重构优化的方法保证了卷积核的低秩；Denton 等^[26]则使用基于奇异值分解(singular value decomposition, SVD)的方法找到了方形卷积核的低秩近似解；Jin 等^[28]提出了一个前馈加速的扁平卷积神经网络，使用一维的垂直卷积核和水平卷积核代替了方形卷积核，通过模型约束来学习一维卷积核参数，三者均达到了较好的效果，并且获得了至少两倍的卷积层加速。在 InceptionV3^[25]中，作者将 7×7 的方形卷积核分解成 7×1 、 1×7 卷积模

块,大幅减少了模型参数,但是上述异构卷积模块在低层特征图上的表现效果不佳。Li 等^[29]使用类似 Inception 的多分支结构,通过增强异构卷积模块获取长程和短程特征,在实时语义分割任务上获得了较好的效果。Guo 等^[30]使用 3×1 和 1×3 的异构卷积核构建卷积模块,在保持高速检测的同时能够有较好的准确度。SLaKNet^[31]使用 51×5 、 5×51 和 5×5 的异构卷积模块将模型尺度扩大到了 51×51 ,并取得了与其他先进的神经网络模型相近的性能。

相较而言,MHCBlock 中的异构卷积模块仅用于在训练过程,通过异构卷积核捕获水平和垂直方向上的病理特征,并在推理过程中,将其所学到的信息融合成单个方形卷积核,减小推理时模型参数数量与计算量。

2.2 多尺度卷积

传统的神经网络模型^[32-33]通常使用小卷积核来进行特征提取,导致模型仅关注于低分辨率特征,忽略了高分辨率特征对模型性能提升的潜力。为了能够同时捕获高低分辨率特征,许多工作^[13, 17, 34]引入了多尺度卷积,通过不同大小的卷积核提取不同分辨率下的图像特征。Zhao 等^[34]提出了 PSPNet,通过多个不同大小的卷积核进行特征提取,在场景解析任务中获得较好结果。Mehta 等^[13]则在 PSPNet 的基础上将多尺度卷积核替换为了多尺度空洞卷积核,进一步增大了特征提取范围,但是致盲性疾病的病理信息一般是连续的,使用空洞卷积会破坏其连续性。Tan 等^[17]提出了一种混合深度卷积 (mixed depthwise convolution, MixConv),通过多个尺度的深度卷积核并行进行卷积操作,分别获取高低分辨率下的信息。与 MixConv 类似,MHCBlock 也通过多尺度深度卷积来获取不同分辨率下的病理信息。

2.3 重参数化

传统的深度学习模型在训练和推理时结构保持一致,而重参数化将模型训练与模型推理解耦,将模型训练时学习到的参数进行等价或近似等价的转换并放入新的模型结构中,从而在保持模型性能的同时,减小模型推理开销。DiracNet^[21]和 RepVGG^[23]提出了一种类似的重参数化方法,通过一个普通卷积、逐点卷积和残差分支分别获取特征图特征,并且在推理时将所有分支融合为单个方形卷积,



图1 (a) 为原始的深度卷积核, (b) 为 MHCBlock 使用的异构卷积核模块图

在不增加模型参数和计算量的情况下提升模型性能。ACNet^[22]提出非对称卷积模块 (asymmetric convolution block, ACB), 通过 3×1 、 1×3 的异构卷积核增强原有的 3×3 卷积核的分类性能, 在推理时, ACB 通过重参数化的方式将三个异构卷积核融合为单个方形卷积核, 不增加模型推理开销。Guo 等^[35]在视觉目标跟踪任务中使用了同样的方法, 并且取得较好的结果。Ding 等^[36]提出了一个多分支模块 (diverse branch block, DBB), 通过不同组合的普通卷积、逐点卷积和平均池化操作获取不同特征、丰富特征空间, 并在推理时通过重参数化转换为单个方形卷积, 不增加格外的模型复杂度。RepLKNet^[37]同样通过重参数化的方法将不同大小的卷积核分支融合, 使得其卷积尺度达到了 31×31 , 并取得了与其他先进模型相近的性能。同样的, 我们提出的 MHCBlock 也使用重参数化的思想, 在训练结束后将不同尺度下的异构卷积核融合, 整合不同异构卷积核学到的信息, 减小模型推理开销。

3. 多尺度异构卷积网络

3.1 异构卷积

图 1 (a) 是一个原始的 $K \times K$ 深度卷积核, 本文将其替换为一个如图 1 (b) 所示的包含 $K \times K$ 、 $K \times 1$ 和 $1 \times K$ 的异构卷积核模块, 三个卷积核均为深度卷积核。在该模块中, 我们令三个异构卷积核的输出特征图相加得到新的输出特征图。对于其卷积过程, 输入的特征图为 $X_{H,W,M}$; 输出的特征图为 $Y_{H,W,M}$; 三个异构卷积核

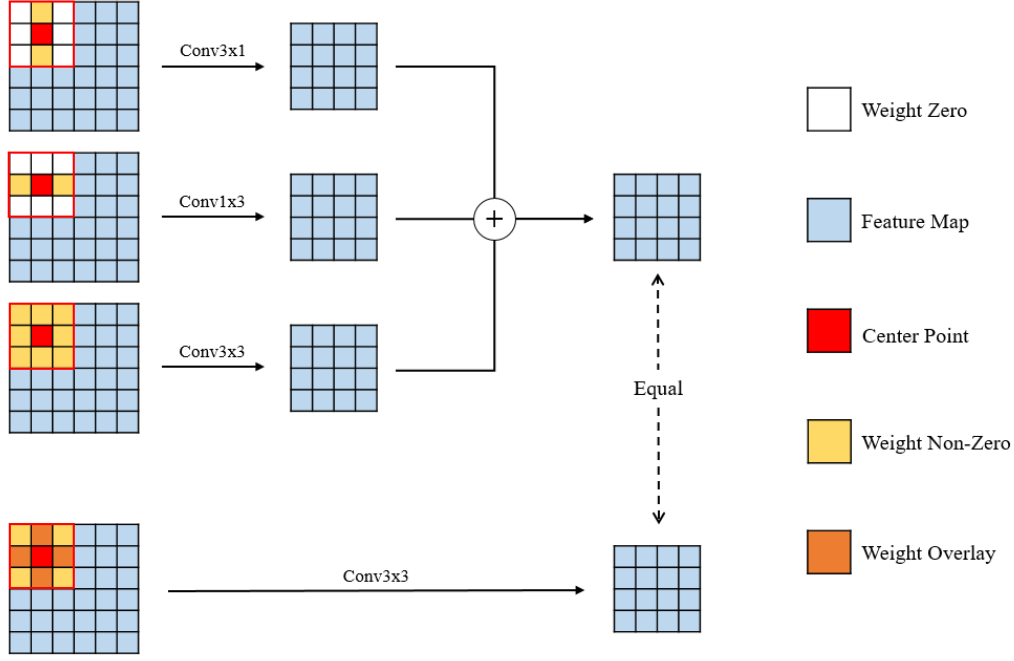


图2 异构卷积核的可加性。其中，红色方框为卷积核范围；蓝色方格为特征图；红色方格为中心点；黄色方格为有参数的卷积核部分；白色方格为参数为零的卷积核部分；棕色为卷积核重叠部分。显然，图中分别使用 3×1 、 1×3 、 3×3 卷积核和三个异构卷积核相加后的 3×3 卷积核卷积效果等价

分别为 $W_{K,K}$ 、 $W_{K,1}$ 和 $W_{1,K}$ ，数量为 M 组。其中， H 为特征图高度； W 为特征图宽度； M 为特征图通道数； K 为卷积核大小。对于输出特征图第 i 个通道的特定点 $y_{o,v,i}$ ，我们可以得到如下关系：

$$y_{o,v,i}^S = \sum_{h=1}^K \sum_{w=1}^K W_{h,w}^{i,S} * X_{o-\frac{K+1}{2}+h, v-\frac{K+1}{2}+w} \quad (1)$$

$$y_{o,v,i}^V = \sum_{h=1}^K W_{h,1}^{i,V} * X_{o-\frac{K+1}{2}+h, v-\frac{K+1}{2}+1} \quad (2)$$

$$y_{o,v,i}^H = \sum_{w=1}^K W_{1,w}^{i,H} * X_{o-\frac{K+1}{2}+1, v-\frac{K+1}{2}+w} \quad (3)$$

$$y_{o,v,i} = y_{o,v,i}^S + y_{o,v,i}^V + y_{o,v,i}^H \quad (4)$$

其中， $*$ 为卷积操作； $W_{h,w}^{i,S}$ 、 $W_{h,1}^{i,V}$ 和 $W_{1,w}^{i,H}$ 分别代表第 i 组 $K \times K$ 、 $K \times 1$ 和 $1 \times K$ 卷积核的参数； $y_{o,v,i}^S$ 、 $y_{o,v,i}^V$ 和 $y_{o,v,i}^H$ 分别代表相应卷积核的卷积结果。由于输入特

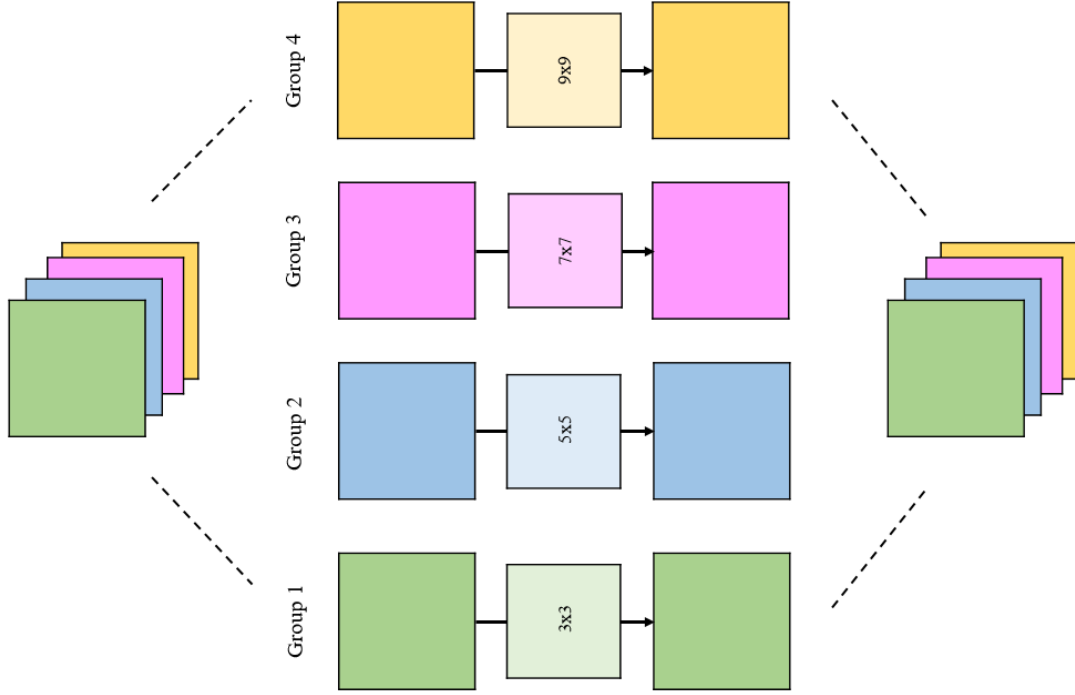


图3 图中 $K \times K$ 代表尺度为 K 的异构卷积模块。当把输入特征图分为四组时，使用的异构卷积核模块尺度分别为 3、5、7、9

征图与卷积方式均相同，我们可以将 $K \times 1$ 和 $1 \times K$ 的一维卷积核分别视为只有其核中心点所在行、列为非零参数的 $K \times K$ 方形卷积核，如图 2 所示。因此，我们可以很容易地得到上述公式等效于公式 (5)。

$$y_{o,v,i}^S = \sum_{h=1}^K \sum_{w=1}^K (W_{h,w}^{i,H} + W_{h,w}^{i,V} + W_{h,w}^{i,S}) * X_{o-\frac{K+1}{2}+h, v-\frac{K+1}{2}+w} \quad (5)$$

因此，我们可以得到结论：当输入特征图与卷积方式均相同时，在同一尺度下的异构卷积核具有可加性。

3.2 多尺度卷积

我们在 MHCBlock 中使用多尺度卷积模块来同时捕获高低分辨率下的病理信息，每个尺度包含一个如图 1 (b) 所示的异构卷积核模块。令输入的特征图大小为 $X_{C,H,W}$ ，其中， C 为特征图通道数， H 和 W 分别为特征图高度和宽度。

MHCBlock 中的多尺度卷积模块将输入特征图按通道平均分成 K 个组，每组使用不同大小的异构卷积模块来提取不同尺度下的病理特征。一般情况下，每个组

为 $\frac{C}{K}$ 个通道；当 C 不能被 K 整除时，最后一个组为 $\lceil \frac{C}{K} \rceil$ 个通道，其他组为 $\lfloor \frac{C}{K} \rfloor$ 个通道。在多尺度卷积模块中，我们分别使用尺度为 $2i+1$ 的异构卷积核模块进行卷积运算。如图 2 所示，当 K 等于 4 时，使用的异构卷积核模块尺度分别为 3、5、7、9。

3.3 重参数化

本文通过重参数化的方法来减小模型推理复杂度。在模型训练完成后，我们首先将 MHCBLOCK 中的异构卷积核与其后面的批归一化 (batch normalization, BN)^[38]层融合，然后将异构卷积模块中的 $K \times K$ 、 $K \times 1$ 和 $1 \times K$ 卷积核融合为单个 $K \times K$ 卷积核以减小模型参数量，加快模型推理速度。

对于异构卷积核与 BN 层的融合，我们以 $K \times K$ 的深度卷积核为例进行说明。令输入的特征图为 $X_{H,W,M}$ ；输出的特征图为 $Y_{H,W,M}$ ；则卷积核为 $W_{K,K}$ ，数量为 M ，其意义与 3.1 节中一样。对于输出特征图的第 i 个通道，其相应的计算表达式如公式 (6) 所示。

$$Y_{:,i} = X_{:,i} * W_{:,i}^i + Bias_i \quad (6)$$

其中， $*$ 仍然代表卷积操作， $X_{:,i}$ 和 $Y_{:,i}$ 分别代表输入、输出特征图的第 i 个通道， $W_{:,i}^i$ 为第 i 个深度卷积核， $Bias_i$ 为第 i 个深度卷积核的偏置。对于特征图的第 i 个通道，BN 层相应的计算表达式如公式 (7) 所示。

$$Y_{:,i} = \frac{X_{:,i} - \mu_i}{\sigma_i} \times \gamma_i + \beta_i \quad (7)$$

其中， μ_i 表示输入特征图第 i 通道的平均值， σ_i 则是其标准差。结合公式 (6) 和 (7) 可以得到公式 (8)。

$$Y_{:,i} = (X_{:,i} * W_{:,i}^i - \mu_i) \times \frac{\gamma_i}{\sigma_i} + \beta_i \quad (8)$$

从公式 (8) 中不难看出，深度卷积层自身的偏置不会影响到 BN 层的输出结果。因此，本文去除深度卷积层的偏置以进一步缩小模型参数。同时，由于 BN 层的 γ_i 、 μ_i 、 σ_i 和 β_i 在推理时固定，我们可以在训练完成后将 BN 层与深度卷积层融合，加快模型推理速度。其表达式如公式 (9) 所示。

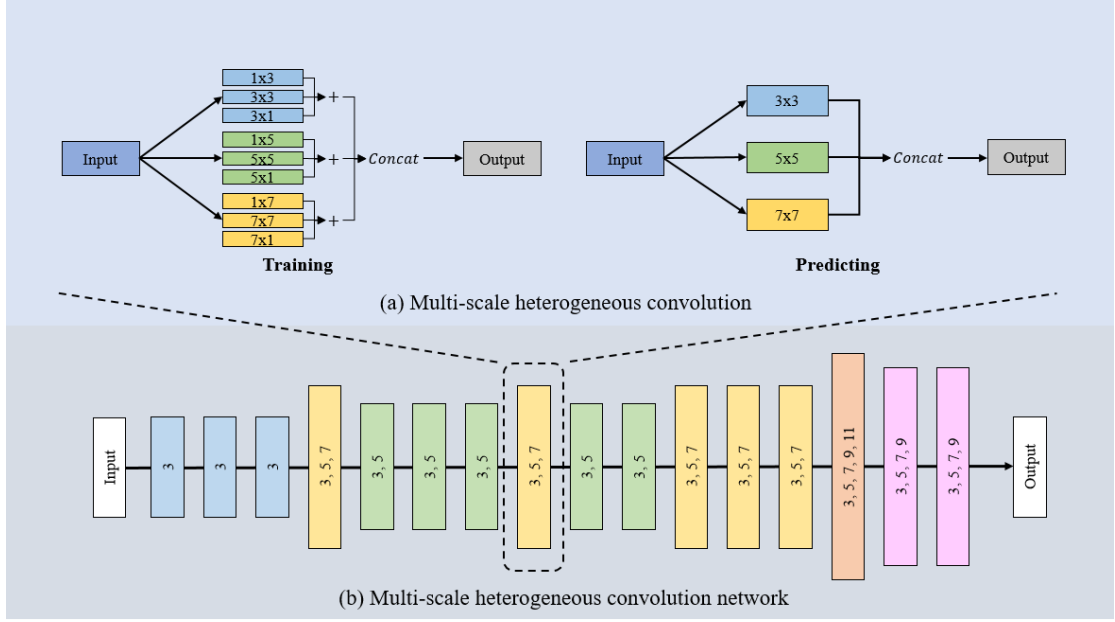


图4 图中，(a)为 MHCBlock 在训练时和推理时的结构，推理时 MHCBlock 将异构卷积核通过重参数化融合，减小模型推理复杂度；(b)为 MHCNet 整体架构。

$$Y_{:,i} = X_{:,i} * (W_{:,i}^i \times \frac{\gamma_i}{\sigma_i}) + (\beta_i - \mu_i \times \frac{\gamma_i}{\sigma_i}) \quad (9)$$

其次，根据 3.1 节中证明的同尺度异构卷积核可加性，我们在 MHCBlock 训练完成后，对于同一尺度下的异构卷积核进行融合。我们将 $K \times 1$ 和 $1 \times K$ 深度卷积核拓展为 $K \times K$ 深度卷积核，并将其与原有 $K \times K$ 深度卷积核相加，然后进行推理。通过重参数化，模型在推理时只需要经过一个 $K \times K$ 的深度卷积核，大幅减少了模型推理所需的时间和参数量。

3.4 多尺度异构卷积网络

为了验证多尺度异构卷积模块的有效性，本文使用与 MixNet^[17] 相同的模型骨架 (MixNet-s) 构建了我们的多尺度异构卷积网络 (MHCNet)，如图 4 所示。该模型将 MixNet-s 中所有 MixConv 模块替换为多尺度异构卷积模块，并沿用其多尺度分组策略。

4. 实验设置与数据集介绍

本文使用一个私有的临床 AS-OCT 核性白内障数据集和三个公开的疾病数

数据集对我们提出的 MHCNet 进行实验验证。三个公开数据集分别为 UCSD 的年龄性黄斑变性数据集、2018 年 ISIC 比赛的皮肤病数据集和一个脑肿瘤数据集。

4.1 AS-OCT 数据集

本文使用的 AS-OCT 数据集是私有的自建临床核性白内障数据集，数据来源是本地一家三甲医院。所有图像均由日本 TOMEY 公司生产的 CASIA2 眼科仪器采集。该数据集共包含了 543 名受试者和 882 只眼睛。其中，右眼有 442 只，左眼有 440 只，受试者的年龄分布为 61.30 ± 18.65 岁。所有受试者在数据采集前都知情该数据用途。

由于缺乏针对 AS-OCT 图像的白内障临床诊断金标准，数据集中受试者的图像标签由对应的裂隙灯图像标签映射得到。首先病人将同时拍摄 AS-OCT 图像和裂隙灯图像，然后由三名经验丰富的临床医生根据 LOCS III^[39] 白内障分类系统对病人的裂隙灯图像进行核性白内障严重级别评定，最后映射到 AS-OCT 图像上得到相应的图像标签，保证了标签的质量和可靠性。

由于同一个受试者两只眼睛的白内障级别相似，本文以受试者为基本单位，将数据集随机分为训练集、验证集和测试集，并确保受试者的两只眼睛不会出现在不同数据集中。该数据集共有 16,201 张 AS-OCT 图像。表 1 为数据集中不同白内障级别 AS-OCT 图像的数量分布。

表 1 AS-OCT 数据集数据分布表

	Normal	Mild	Severe	Total
Training	1004	2872	5518	9394
Validation	345	1306	1740	3391
Testing	254	664	2498	3416
Total	1603	4842	9756	16201

数据来源：iMED 团队

4.2 UCSD 数据集

该数据集为加州大学圣地亚哥分校收集整理的年龄性黄斑变性 OCT 公开数

数据集，数据集图像主要分为正常、沉积物、糖尿病黄斑水肿和脉络膜新生血管四类，并且划分了相应的训练集、验证集和测试集，数据集共有 84484 张图像。

4.3 ISIC2018 数据集

该数据集是 2018 年 ISIC 比赛的皮肤病数据集^[40-41]，数据集包含了七类皮肤病的皮肤镜图像，分别是黑素瘤、黑素细胞痣、基底细胞癌、光化性角化病、良性角化病、皮肤纤维瘤和血管受损。我们仅使用该比赛公开的训练集和验证集进行实验，共 10208 张。

4.4 BTC 数据集

该数据集是一个公开的脑肿瘤核磁共振成像数据集，共有 3264 张图像。数据集中图像被分成四类，分别为正常、神经胶质瘤肿瘤、脑膜瘤肿瘤和垂体肿瘤。该数据集已经预先进行了训练集和测试集的划分，本文沿用其提供的划分方法进行实验。

4.5 评价指标

基于已有的致盲性疾病识别工作^[42-43]，本文使用三种常见的评价方法来评估 MHCNet 以及其他先进网络的分类性能：准确率 (accuracy, ACC)、敏感度 (sensitivity, SEN) 和 F1 分数 (f1 score)。准确率表示模型分类正确的图像比例，敏感度表示模型对目标类别的敏感程度，F1 分数则是评估模型能力的重要指标。这些评估方法的计算表达式如下：

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (10)$$

$$SEN = \frac{TP}{TP + FN} \quad (11)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (12)$$

此外，我们还通过模型参数数量和模型计算量 (multiply-accumulate operation, MAC) 来评估模型的复杂度。

4.6 实验设置

本文使用 PyTorch、scikit-learn 和 OpenCV 等工具实现 MHCNet 以及其他卷积神经网络模型，并且所有实验均在一个配有 NVIDIA TITAN V (12GB RAM) GPU 的 Ubuntu 服务器上进行。对于模型的训练过程，我们使用交叉熵损失函数（cross entropy loss function）和随机梯度下降（stochastic gradient descent, SGD）来进行模型参数的更新，使用 0.9 的随机梯度下降动量和 $1e-5$ 的权重衰减，批量大小为 32。数据集训练的学习率和训练轮数默认为 0.01 和 200，并且在 100 轮以后，每训练 10 轮，学习率降低 10 倍。

5. 实验结果与分析

在这一节中，我们分别在四个数据集上比较了 MHCNet 与先进的基线模型、轻量级模型和重参数化模型的性能，并且在 AS-OCT 数据集和 UCSD 数据集上比较了 MHCNet 与其它模型的复杂度，比较的模型有 ResNet^[33]、VGGNet^[32]、ResNeXt^[44]、IGCNet^[10]、EfficientNet^[18]、ESPNetV2^[14]、MobileNetV2^[8]、ShuffleNetV1/V2^[15-16]、ACNet^[22]、RepVGG^[23] 和 MixNet^[17]。此外，我们还在 ISIC2018 数据集上对 MHCNet 及其变体进行了消融实验。

5.1 AS-OCT 数据集的分类结果分析

表 2 为 MHCNet 与先进的神经网络模型在 AS-OCT 数据集上的分类结果与模型复杂度的对比。不难看出，MHCNet 在准确率、敏感度和 F1 分数上都取得了最好的结果，并且在模型参数量和模型计算量上优于三个基线模型、两个重参数化模型和大部分轻量级模型。例如，相较于 RepVGG，MHCNet 减少了约 63.1% 的参数量和 81.3% 的计算量，但是在准确率、敏感度和 F1 分数上分别提升了 4.90%、8.62% 和 2.96%。对比于轻量级模型，MHCNet 在保持较低模型复杂度的同时获得了优于其他模型的准确率。例如，IGCNetV1 虽然参数量仅约为 MHCNet 的 17.7%，但 MHCNet 在计算量上减少了约 69.6% 并且取得了 6.62% 的准确率提升；ESPNetV2 虽然计算量仅约为 MHCNet 的 39.6%，但 MHCNet 在准确率方面提高了 2.78%。此外，图 5 展示了 MHCNet 与其他模型在 AS-OCT 数据集上的性能表

现与模型复杂度。横坐标为模型参数量，纵坐标为模型准确率，圆盘大小则代表了模型计算量，圆盘越大则代表模型计算量越大，反之则越小。可以看出，MHCNet 在性能与效率的平衡上取得了最好的结果。

表 2 AS-OCT 数据集分类结果以及模型复杂度比较

Methods	ACC	SEN	F1	Params(M)	MACs(M)
ResNet18	87.24	84.60	83.70	11.178	1824
VGGNet16	82.26	76.51	77.96	14.725	15401
ResNeXt50	84.88	79.75	80.33	23.016	4286
IGCNetV1	82.95	76.31	73.83	0.460	840
EfficientNet	86.04	85.10	81.67	2.910	1412
ESPNetV2	86.89	82.39	84.51	0.645	101
MobileNetV2	81.75	76.64	72.36	2.228	326
ShuffleNetV1	80.06	70.27	74.66	0.883	147
ShuffleNetV2	86.86	77.63	82.48	5.324	577
ACNet	86.02	61.07	57.91	14.725	15401
RepVGG	84.67	79.08	82.61	7.032	1360
MixNet-s	85.20	82.62	75.87	2.602	263
MHCNet	89.57	87.70	85.57	2.596	255

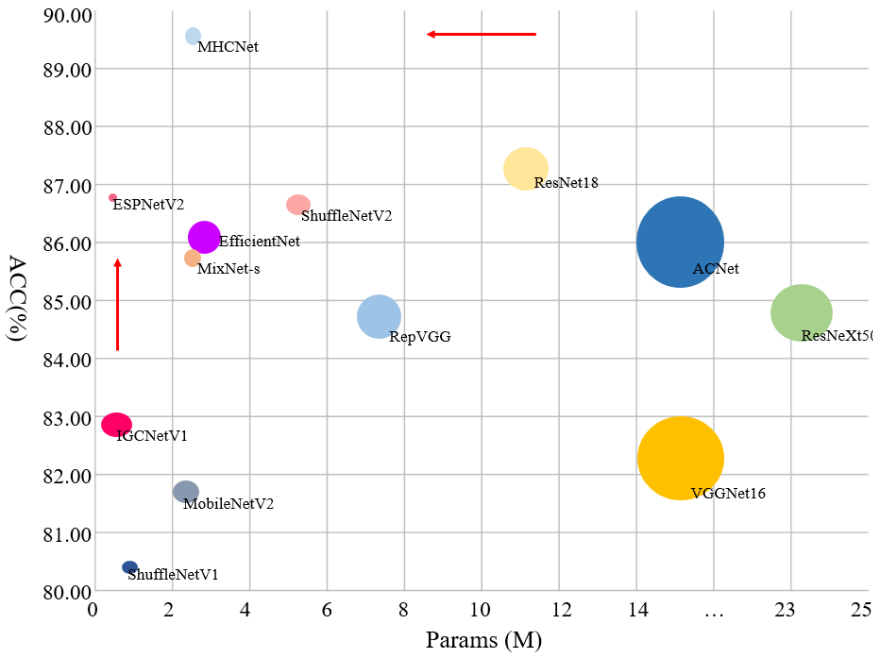


图 5 AS-OCT 数据集结果和模型复杂度比较，圆盘越大则模型计算量越大，反之则越小

5.2 UCSD 数据集的分类结果分析

在 UCSD 数据集上, MHCNet 及其他竞争模型的年龄性黄斑变性分类结果如表 3 所示。可以看出,其他模型在该数据集上都有较好的表现,例如 VGGNet16、ResNext50 和 ACNet 都达到了 94.86%的准确率, 高于使用的轻量级网络, 但仍然不及 MHCNet 95.26%的准确率。对于这个结果, 我们有两种可能的解释: (1) MHCNet 使用了不同尺度的卷积核, 能够同时捕获多尺度病理信息, 相对的, ResNet18 等模型仅仅使用了小卷积核来获取局部特征, 而这种方式仅仅适合精细的病理特征; (2) MHCNet 使用了异构卷积模块, 能够更好地捕获垂直方向和水平方向的病理特征, 更有利于疾病分类任务; (3) VGGNet16、ResNext50 和 ACNet 的参数量分别是 MHCNet 的 5.69 倍、8.86 倍和 5.69 倍, 大量的参数保证了模型性能, 但是其中有较多的冗余参数, 限制了模型性能的进一步提升, 相比而言, MHCNet 参数量更少, 而且更能有效地利用参数。同时, 相较于大部分轻量级模型, MHCNet 使用更少的参数量和计算量取得了更好地准确率。例如, MHCNet 仅使用了 EfficientNet 89.2%的参数但是性能提高了 1.61%。

表 3 UCSD 数据集分类结果以及模型复杂度比较

Methods	ACC	SEN	F1	Params(M)	MACs(M)
ResNet18	94.05	94.03	93.92	11.179	1824
VGGNet16	94.86	94.83	94.79	14.725	15401
ResNeXt50	94.86	94.83	94.79	23.018	4286
IGCNetV1	93.35	93.28	93.15	0.472	840
EfficientNet	93.65	93.65	93.57	2.910	1412
ESPNetV2	93.65	93.65	93.50	0.646	101
MobileNetV2	94.66	94.67	94.60	2.229	326
ShuffleNetV1	94.36	94.37	94.29	0.884	147
ShuffleNetV2	93.85	93.82	93.76	5.326	577
ACNet	94.86	94.86	94.79	14.725	15401
RepVGG	94.15	94.17	94.10	7.034	1360
MixNet-s	94.46	94.46	94.39	2.604	263
MHCNet	95.26	95.23	95.18	2.597	255

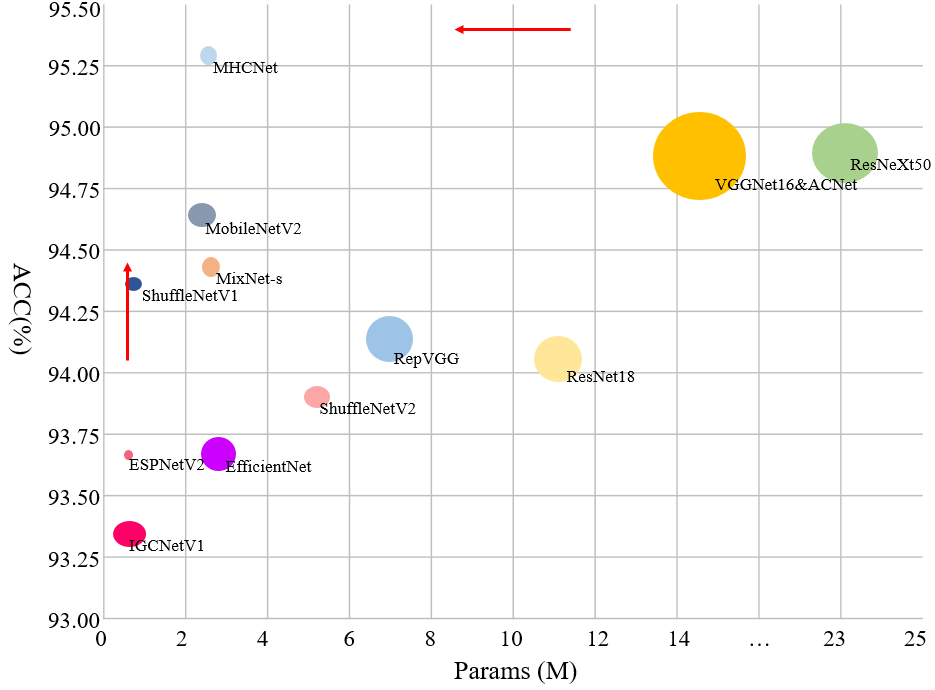


图 6 UCSD 数据集结果和模型复杂度比较，圆盘越大则模型计算量越大，反之则越小

此外，图 6 展示了 MHCNet 与其他先进的神经网络模型之间的性能、模型复杂度的对比。不难看出，MHCNet 在模型性能与模型复杂度上有着最好的平衡。

5.3 其他疾病数据集分类结果分析

除了致盲性疾病数据集，我们还使用了皮肤病数据集和脑肿瘤数据集以验证模型的鲁棒性与泛化能力，其分类结果如表 4 所示。在这两个数据集上，MHCNet 均取得了最好的分类性能。在 ISIC2018 皮肤病数据集上，相较于其他先进的神经网络模型，MHCNet 在准确率、敏感度和 F1 分数上最高分别取得了 3.12%、3.12%和 5.90%的提升；在 BTC 脑肿瘤数据集上，MHCNet 则最高分别取得了 5.73%、6.06%和 6.78%的提升。

5.4 消融实验

在这一节中，我们设计了多个 MHCBlock 的变种来进行消融实验，以验证其有效性。我们使用与 5.3 节中与 ISIC2018 数据集实验相同的实验设置。在 MHCNet 的基础上，我们分别设计了没有异构卷积核和 BN 层的、没有垂直卷积

表 4 ISIC2018 数据集与 BTC 数据集分类结果比较

Methods	ISIC2018			BTC		
	ACC	SEN	F1	ACC	SEN	F1
ResNet18	78.65	78.65	77.39	85.16	84.26	84.71
VGGNet16	80.21	80.21	78.61	85.42	84.51	84.73
ResNeXt50	78.13	78.13	74.85	82.81	82.01	81.91
IGCNetV1	79.17	79.17	77.60	80.73	79.74	79.14
EfficientNet	80.21	80.21	79.42	85.42	84.30	84.86
ESPNetV2	79.17	79.17	76.79	82.81	81.63	81.94
MobileNetV2	79.69	79.69	77.82	85.68	84.97	85.28
ShuffleNetV1	80.21	80.21	77.94	82.29	81.01	81.62
ShuffleNetV2	80.21	80.21	78.71	83.33	82.55	82.87
ACNet	80.21	80.21	79.06	85.16	84.26	84.71
RepVGG	78.65	78.65	77.87	85.68	85.15	85.38
MixNet-s	78.65	78.65	77.01	83.59	82.38	82.69
MHCNet	81.25	81.25	80.75	86.46	85.80	85.92

核的、没有水平卷积核的、没有 BN 层的 MHCBLOCK 和原本的 MHCBLOCK 进行实验，实验结果如表 5 所示。可以看到，MHCBLOCK 在缺失 BN 层时，模型计算量会下降，但是分类性能也会大幅下降，这是由于 BN 层能够将数据归一化到合适的尺度，便于模型训练和推理；在缺少某一个异构卷积核时，性能也会小幅下降，例如没有垂直卷积核的 MHCBLOCK 在准确率、敏感度和 F1 分数上分别下降了 1.56%、1.56%和 2.70%，说明异构卷积核能够更好地捕获垂直和水平方向上的病理特征，提升模型性能。

表 5 AS-OCT 数据集分类结果以及模型复杂度比较

Methods	Hor	Ver	BN	ACC	SEN	F1	Params(M)	MACs(M)
MHCNet				64.06	64.06	50.03	2.578	238
MHCNet	√		√	79.69	79.69	78.05	2.602	255
MHCNet		√	√	79.69	79.69	78.95	2.602	255
MHCNet	√	√		64.06	64.06	50.03	2.576	238
MHCNet	√	√	√	81.25	81.25	80.75	2.602	255

6. 结语

本文提出了一个即插即用的多尺度异构卷积模块 (MHCBlock)，相较于传统的三维卷积核和深度可分离卷积核，该模块能够通过多尺度卷积核捕获不同尺度下的病理信息，并且通过异构卷积核捕获垂直和水平方向的病理信息，同时还通过重参数化的形式将 BN 层与卷积层融合、将异构卷积核融合，在不降低模型性能的前提下减小模型参数量，提升模型推理速度。此外，我们使用 MHCBlock 构建了多尺度异构卷积网络 (MHCNet)，并在一个私有数据集和三个公开数据集上进行了充分实验。结果表明，MHCNet 相较于其他先进的卷积神经网络，如 ACNet 和 MobileNetV2，能够更好地平衡模型性能与模型复杂度。同时，我们还设计了消融实验验证了异构卷积核和 BN 层对模块分类性能的提升。可以看出，MHCNet 有着能够部署在资源有限的医疗设备上面的潜力，例如 OCT 设备和眼底图像设备。

在未来，我们将收集更多致盲性疾病数据集以验证 MHCNet 对其它致盲性疾病的泛化能力，并且将 MHCNet 部署到真实的嵌入式医疗设备上以进一步测试模型性能。

参考文献

- [1] World Health Organization. World report on vision[J]. 2019.
- [2] Resnikoff S, Lansingh V C, Washburn L, et al. Estimated number of ophthalmologists worldwide (International Council of Ophthalmology update): will we meet the needs?[J]. British Journal of Ophthalmology, 2020, 104(4): 588-592.
- [3] Fu H, Xu Y, Lin S, et al. Multi-context deep network for angle-closure glaucoma screening in anterior segment OCT[C]//Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11. Springer International Publishing, 2018: 356-363.
- [4] Zhang X, Xiao Z, Li X, et al. Mixed pyramid attention network for nuclear cataract classification based on anterior segment OCT images[J]. Health Information Science and Systems, 2022, 10(1): 3.
- [5] Burlina P M, Joshi N, Pekala M, et al. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks[J]. JAMA ophthalmology, 2017, 135(11): 1170-1176.
- [6] Xiao Z, Zhang X, Higashita R, et al. A 3D CNN-based multi-task learning for cataract screening and left and right eye classification on 3D AS-OCT images[C]//Proceedings of the 2021 International Conference on Intelligent Medicine and Health. 2021: 1-7.
- [7] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.
- [8] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.
- [9] Howard A, Sandler M, Chu G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1314-1324.
- [10] Zhang T, Qi G J, Xiao B, et al. Interleaved group convolutions[C]//Proceedings of the IEEE international conference on computer vision. 2017: 4373-4382.
- [11] Xie G, Wang J, Zhang T, et al. Interleaved structured sparse convolutional neural

networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8847-8856.

[12] Sun K, Li M, Liu D, et al. Igcv3: Interleaved low-rank group convolutions for efficient deep neural networks[J]. arXiv preprint arXiv:1806.00178, 2018.

[13] Mehta S, Rastegari M, Caspi A, et al. Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation[C]//Proceedings of the european conference on computer vision (ECCV). 2018: 552-568.

[14] Mehta S, Rastegari M, Shapiro L, et al. Espnetv2: A light-weight, power efficient, and general purpose convolutional neural network[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 9190-9200.

[15] Zhang X, Zhou X, Lin M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6848-6856.

[16] Ma N, Zhang X, Zheng H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 116-131.

[17] Tan M, Le Q V. Mixconv: Mixed depthwise convolutional kernels[J]. Proceedings of the British Machine Vision Conference (BMVC), 2019: 116.1-116.13.

[18] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//International conference on machine learning. PMLR, 2019: 6105-6114.

[19] Cai H, Zhu L, Han S. Proxylessnas: Direct neural architecture search on target task and hardware[C]//International Conference on Learning Representations. 2019.

[20] Tan M, Chen B, Pang R, et al. Mnasnet: Platform-aware neural architecture search for mobile[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 2820-2828.

[21] Zagoruyko S, Komodakis N. Diracnets: Training very deep neural networks without skip-connections[J]. arXiv preprint arXiv:1706.00388, 2017.

[22] Ding X, Guo Y, Ding G, et al. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1911-1920.

- [23] Ding X, Zhang X, Ma N, et al. Repvgg: Making vgg-style convnets great again[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13733-13742.
- [24] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [25] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.
- [26] Denton E L, Zaremba W, Bruna J, et al. Exploiting linear structure within convolutional networks for efficient evaluation[J]. Advances in neural information processing systems, 2014, 27.
- [27] Jaderberg M, Vedaldi A, Zisserman A. Speeding up convolutional neural networks with low rank expansions[J]. arXiv preprint arXiv:1405.3866, 2014.
- [28] Jin J, Dundar A, Culurciello E. Flattened convolutional neural networks for feedforward acceleration[J]. arXiv preprint arXiv:1412.5474, 2014.
- [29] Li Y, Li X, Xiao C, et al. EACNet: Enhanced asymmetric convolution for real-time semantic segmentation[J]. IEEE signal processing letters, 2021, 28: 234-238.
- [30] Yang Z, Ma X, An J. Asymmetric convolution networks based on multi-feature fusion for object detection[C]//2020 IEEE 16th international conference on automation science and engineering (CASE). IEEE, 2020: 1355-1360.
- [31] Liu S, Chen T, Chen X, et al. More convnets in the 2020s: Scaling up kernels beyond 51x51 using sparsity[J]. arXiv preprint arXiv:2207.03620, 2022.
- [32] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]// International Conference on Learning Representations (ICLR 2015). 2015.
- [33] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [34] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.
- [35] Shaozhe G, Yong L, Youshan Z, et al. A Asymmetric Attention Siamese Network for Visual Object Tracking[C]//2021 2nd International Conference on Big Data and Informatization Education (ICBDIE). IEEE, 2021: 163-168.

- [36] Ding X, Zhang X, Han J, et al. Diverse branch block: Building a convolution as an inception-like unit[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 10886-10895.
- [37] Ding X, Zhang X, Han J, et al. Scaling up your kernels to 31x31: Revisiting large kernel design in cnns[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 11963-11975.
- [38] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International conference on machine learning. pmlr, 2015: 448-456.
- [39] Chylack L, Wolfe J, Singer D, et al. The lens opacities classification system III[J]. Archives of ophthalmology, 1993, 111(6): 831-836.
- [40] Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions[J]. Scientific Data, 2018, 5.
- [41] Codella N, Gutman D, Celebi M, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)[C]//2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). 2018:168-172.
- [42] Cao L, Li H, Zhang Y, et al. Hierarchical method for cataract grading based on retinal images using improved Haar wavelet[J]. Information Fusion, 2020, 53: 196-208.
- [43] Fu H, Xu Y, Lin S, et al. Angle-closure detection in anterior segment OCT based on multilevel deep network[J]. IEEE transactions on cybernetics, 2019, 50(7): 3358-3366.
- [44] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1492-1500.

致谢

首先，我要感谢父母对本人学习生活的支持，在我失败的时候给予我鼓励，在我成功的时候给予我肯定，让我有信心一路前行。然后，我要感谢刘江教授和章晓庆博士对本人学业的指导和关心，他们带领我打开了科研的大门，在我遇到困难时给予帮助。最后，我要感谢南方科技大学以及南方科技大学计算机科学与工程系的各位老师对我的辛勤培养，还要感谢同学们的帮助和鼓励。