

# 基于多流融合网络知识提取的早产儿视网膜病变自动检测

沈<sup>a</sup>英山, 罗<sup>a</sup>志涛, 徐<sup>a</sup>木新, 梁<sup>a</sup>志浩, 范<sup>a</sup>小茂, 卢晓河

<sup>a</sup> 华南师范大学计算机科学学院, 广州 510631

<sup>b</sup> 南方医科大学珠江医院眼科, 广州 510282

## 条信息

关键词:

深度神经网络眼底图像  
知识蒸馏轻量级

早产儿视网膜病变

## 摘要

早产儿视网膜病变(ROP)是一种潜在的致盲眼病, 主要发生在低出生体重的早产儿。目前, rop 是全球儿童致盲的首要原因。有多种方法可用于检测或分期 ROP。然而, 很少有研究关注 ROP 的早期检测和需要治疗的 ROP (TR-ROP)。此外, 最近提出的大多数神经网络包含大量的神经网络参数。本研究旨在提出一种基于早期眼底图像的多流融合神经网络知识提取的轻量级 TR-ROP 检测神经网络。首先对多流融合神经网络进行高精度训练, 然后利用知识蒸馏将其知识转移到更小的模型中, 该模型适合部署到嵌入式 ROP 检测设备中。实验在完全由早期眼底图像组成的数据集上进行了五折交叉验证。该网络在 ROP 检测上的准确率、灵敏度和特异度分别为 0.9734、0.9456 和 0.9823, 在 TR-ROP 检测上的准确率、灵敏度和特异度分别为 0.9222、0.9516 和 0.8571, 均优于现有的先进方法。

## 1. 介绍

早产儿视网膜病变(ROP)是一种潜在的致盲眼病 [1,2,3]。主要发生于低出生体重早产儿 [4,5]。引起视网膜脱离, 导致终身视力损害和失明。

目前, ROP 是儿童失明的主要原因 [6,7,8]。2010 年, 中国早产儿超过 1400 万, 其中约 18.47 万早产儿发生 ROP, 约占 1.32%。约 12,300 名 ROP 早产儿出现轻度或中度视力障碍, 20,000 名早产儿出现重度视力障碍甚至失明 [9,10,11,12]。在大多数情况下, ROP 会在几个月内消失 [13,14]。然而, 如果不及时干预, 仍有超过 10% 的 ROP 早产儿会出现视力障碍或失明 [15,16,17]。因此, 对需要治疗的早产儿进行早期筛查具有重要意义。

需要处理的 ROP (TR-ROP)检测任务包括两个阶段:ROP 检测和 TR-ROP 检测。在第一阶段, ROP 检测主要基于电子病历系统中的临床数据和眼底图像。临床资料方面, 记录早产儿出生体重、胎龄、日增重等与 ROP 相关的特征, 建立 ROP 检测模型。Owen 等人根据早产儿出生体重、是否需要手术和母体镁预防来检测 ROP 进行了队列分析 [18]。Ying 等人基于出生体重和呼吸状态建立了 ROP 检测模型 e-ROP [19]。Pivodic 等人通过分析性别、胎龄和出生体重建立了严重 ROP 的检测模型 [20]。

作者。

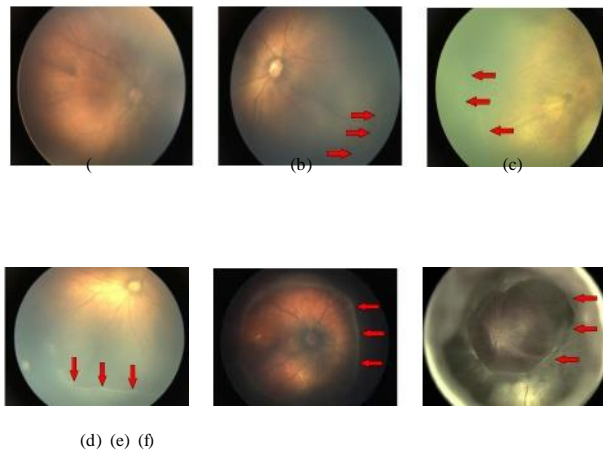


图 1:眼底图像示例。病变区域用红色箭头标记。(一)正常;(b)第一阶段的 ROP;(c)第二阶段的 ROP;(d)第三阶段的 ROP;(e)第四阶段的 ROP;(f)第 5 阶段的 ROP。

Binbaum 等人通过分析早产儿出生体重、胎龄和日增重, 提出了一种多因素 logistic 回归模型对 ROP 进行分类 [21]。此外, Binbaum 等人提出了一种名为 G-ROP 的混合建模方法来分析 ROP, 并取得了较好的结果 [22]。Eckert 等人基于线性回归方法建立了检测严重 ROP 的模型 [23]。这些 ROP 检测方法在其指定的数据集上都能获得令人满意的结果。然而, 由于不同地区早产儿的出生体重、胎龄等特征存在差异, 其 ROP 检测性能并不稳定。当上述 ROP 检测模型推广到不同地区的不同临床数据集时, ROP 检测性能大大降低 [24,25]。

此外，上述大多数 ROP 检测通常需要早产儿 39 天的体重记录。这一过程相当费力，且体重记录往往不完整，限制了检测模型在临床实践中的应用。

眼底图像是检测 ROP 的另一种方法。医生可以直接观察 rop 相关病变区域，并在一定程度上诊断是否需要治疗。根据国际 ROP 检测指南(ICROP)，ROP 症状可分为 5 个阶段，如图 1 所示[26]。第 4、5 期的 ROP 症状相当严重，明显需要治疗。然而，在第 1、2 和 3 期，ROP 症状较轻，甚至难以观察到。近年来，许多研究者致力于基于眼底图像的自动 ROP 检测方法。Brown 等人实现并验证了一种基于深度卷积神经网络的 ROP + 疾病自动诊断算法[27]。Tan 等人介绍了一种称为 ROP 的深度学习方法[28]。Hu 等人提出了一种新颖架构的卷积神经网络，通过每次检查来识别 ROP 疾病的存在及其严重程度[29]。Peng 等人开发了一种基于深度神经网络的五级 ROP 分期网络[30]。Agrawal 等人开发了一种集成框架来自动识别眼底图像中的区域[31]。此外，Agrawal 等人实现了基于 dl 的分割技术，从眼底图像中分割血管和脊[32]。Chen 等人提出了一个使用注意力机制和全卷积网络[33]的多实例学习网络。与基于临床数据的方法相比，这些基于眼底图像的方法可以获得更高的准确性和特异性。虽然这些方法在所有五个阶段的 ROP 检测中都能获得令人满意的结果，但当 ROP 处于早期 1、2 和 3 阶段，眼底图像中出现轻微或轻微的症状时，它们遇到了很大的挑战。此外，基于上述方法的 ROP 检测模型庞大，神经网络参数庞大。它们不适合部署到嵌入式 ROP 检测设备中。

对于第二阶段的 TR-ROP 检测，它筛选出需要治疗的 ROP 早产儿，以防止其病情在 ROP 疾病的连续病程中恶化。同时，TR-ROP 的检测主要基于早期眼底图像，因为所有处于 ROP 后期 4 期和 5 期的早产儿都明显需要治疗。TR-ROP 检测的任务是在早期阶段识别早产儿 ROP，这在已发表的文献中尚未报道。本研究旨在提出一种基于早期眼底图像的多流融合神经网络知识提取的轻型 TR-ROP 检测神经网络。TR-ROP 检测神经网络主要包括 ROP 检测和 TR-ROP 检测两个阶段。对于 ROP 检测，首先训练了一个称为 MS-ROP 的多流融合神经网络。

MS-ROP 包括三个模块:特征提取模块、特征融合模块和分类模块。特征提取模块包含 ResNet18、DenseNet121 和 EfficientNetB2 3 个并行深度神经网络，用于提取丰富多样的高级特征。然后，在特征融合模块中，将 3 个不同的高级特征堆叠成单一的高级特征，将多样化的信息融合在一起。最后，分类模块负责对融合后的特征进行分类。然而，MS-ROP 具有巨大的神经网络参数。为了克服这一挑战，提出了一种轻量级的神经网络 MN-ROP，利用知识蒸馏方法将知识从 MS-ROP 转移到它。对于 TR-ROP 检测，首先对 MS-ROP 进行转移和训练，并命名为 MS-TR-ROP。与 ROP 检测一样，提出了一种基于 MN-ROP 的轻量级神经网络 MN-TR-ROP。同样，采用知识蒸馏方法将知识从 MS-TR-ROP 转移到 MN- TR-ROP。综上所述，主要贡献总结如下：

- (i)对早期眼底图像进行 TR-ROP 检测。采用图像增强算法使病变区域明显，解决早期眼底图像中症状不明显的问题。
- (ii)提出一种多流融合神经网络(即 MS-TR-ROP)来自动检测眼底图像中的 TR-ROP。MS-TR-ROP 主要由特征提取模块和特征融合模块组成。为了提取丰富多样的高级特征信息，特征提取模块采用了 Resnet18、DenseNet121 和 EfficientNetB2 三流强大神经网络。为了进一步学习高级特征，提出了一种叠加方案的特征融合模块。
- (iii)为了减少 MS-TR-ROP 中权重参数的数量，提出了一种轻量级神经网络(即 MN-TR-ROP)，利用知识蒸馏从 MS-TR-ROP 中转移知识。MS- TR-ROP 有 26.83M 个参数，而提出的 MN-TR-ROP 只有 350 m 个参数，其中 MN-TR-ROP 的权重参数减少了约 86.95%。
- (iv)大量实验结果表明，轻量级神经网络 MN-TR-ROP 在 TR-ROP 检测中达到了最先进的性能，准确率为 0.922，灵敏度为 0.952。结果显示，MN-TR-ROP 成功地学习了症状的特征。

本文的其余部分组织如下。在第 2 节中，已经详细阐述了所提出的方法。在第三节中，介绍了实验设置。

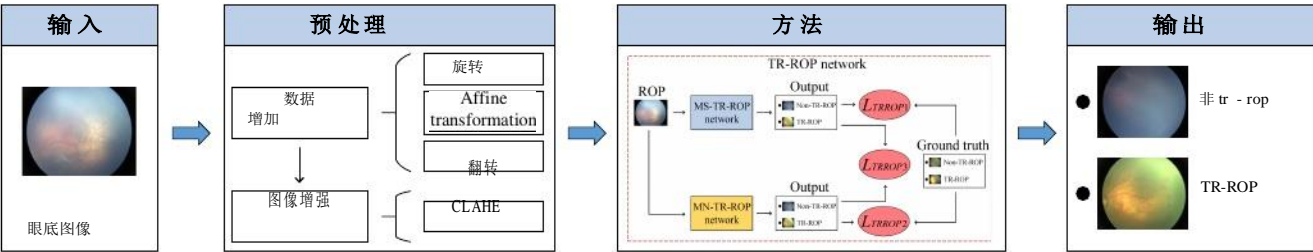


图 2:TR-ROP 检测流程示意图。输入为眼底图像。经过包括数据增强和图像增强策略的预处理步骤后，图像通过所提出的网络，最终获得检测结果的输出。“CLAHE”和“TR-ROP”分别代表对比度有限的自适应直方图均衡化算法和需要处理的 ROP。

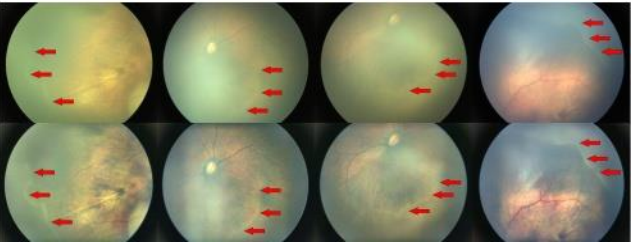


图 3:原始眼底图像和图像增强后的眼底图像示例。病变区域用红色箭头标记。第一行为原始眼底图像。第二行为一张张对应的图像

增强。

第四节详细介绍了实验结果。实验结果在第 5 节中进行了讨论。最后，在第 6 节给出了一个简短的结论。

2.方法

图 2 显示了 TR-ROP 检测的整个过程。主要包括图像预处理，提出

网络，以及性能评估指标。

2.1.预处理

ROP 和 TR-ROP 的早期检测主要面临以下两大挑战：一是不平衡。其次，早期眼底图像中的症状轻微或轻微，使得网络难以检测。第一个挑战可以克服

通过将少数族裔类别中的图片旋转 90 度和 180 度并保存下来，使类别比例接近 1:1。此外，在线数据增强

包括翻转、旋转和仿射变换的融合，增强模型泛化能力[34]。在细节微妙。因此，采用一种对比度有化算法(CLAHE)进行图像增强

在突出细节和限制噪声放大情况方面是有效的[35,36]。给出了采用 CLAHE 算法前后眼底图像的实例

图 3。很明显，应用 CLAHE 算法后，rop 相关病变区域变得更加明显。进一步将眼底图像大小调整为 256×256×3，并标准化为 [0,1]，以节省计算资源[37]。

2.2.该方法

如图 4 所示。TR-ROP 检测神经网络

网络包括两个阶段:ROP 检测和 TR-ROP 检测。这两个阶段具有相似的过程和网络结构。对于 ROP 检测，首先训练多流融合神经网络 MS-ROP。

然后，采用一种轻量级的神经网络精馏方法从 MS-ROP 中转移知识。对于 TR-ROP 检测，首先对 MS-ROP 进行转移和训练，命名为 MS-TR-ROP。与 ROP 检测一样，提出了一种轻量级神经网络 TR-ROP，其 MS-TR-ROP 将知识传递给 MN-TR-

2.2.1.MS-ROP 网络

MS-ROP 包括三个模块:特征提取模块、特征融合模块和分类模块

模块。

特征提取模块：特征提取模块包含 ResNet18、DenseNet121 和 EfficientNetB2 三个并行的深度神经网络，用于提取丰富多样的高级特征[38,39,40]。这三个并行深度神经网络中的每一个都能专注于干

特征融合模块:首先将上述神经网络中提取的 3 个不同特征叠加在一起，大小为 8 × 8 × 2944。它包含了

眼底图像中不同区域的信息[41,42,43]。然后特征图通过 1 × 1 的卷积层，灵活控制特征图的深度，减少参数。它还实现了跨通道信息合成，并添加了一个非线性特征。经过卷积，得到特征图的个数

通道减少到 256 个。

分类模块:将  $8 \times 8 \times 256$  大小的特征图通过自适应平均池化处理

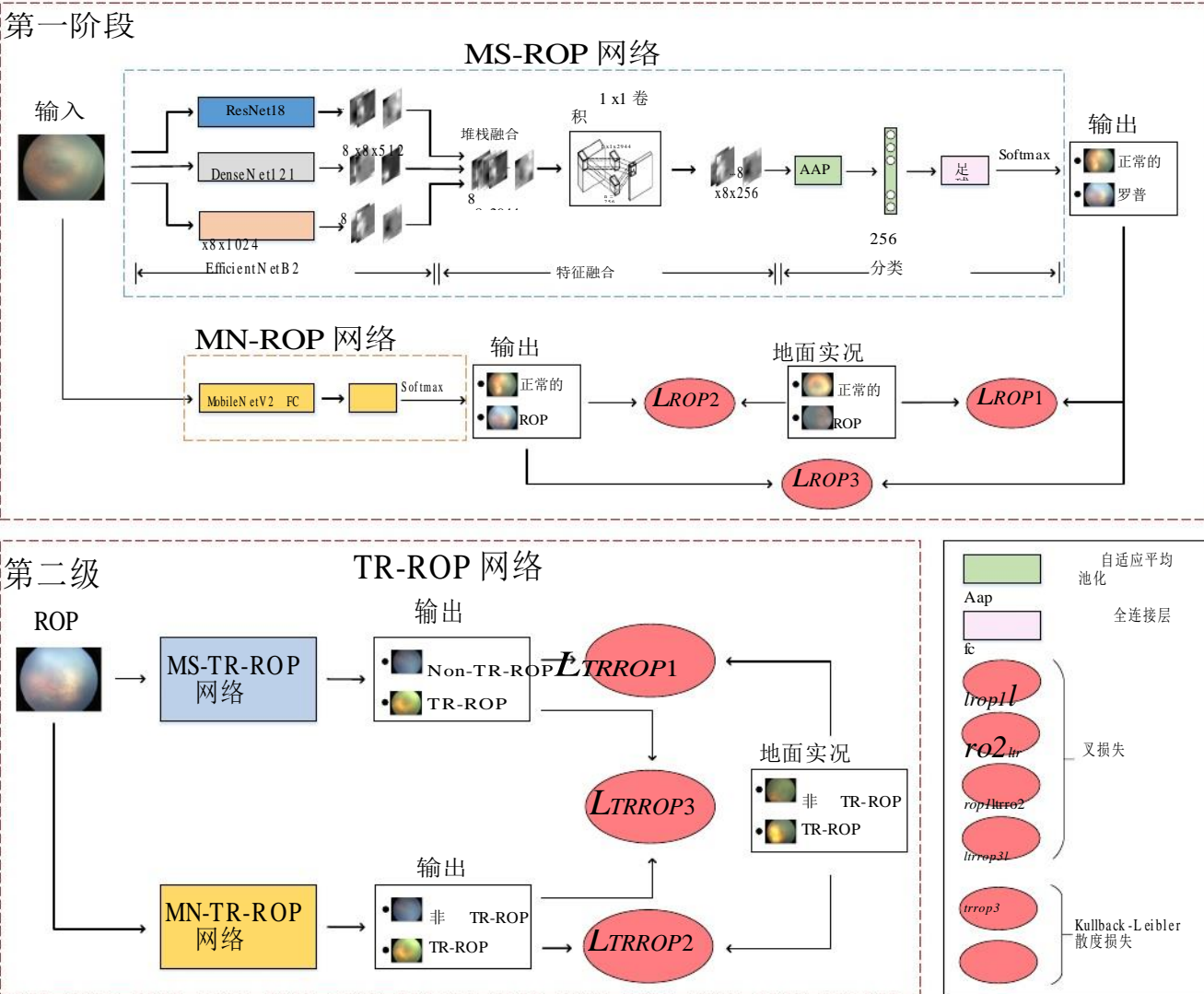


图 4: 本文提出的 ROP 检测方法 (正常/ROP) 和需要处理的 ROP 检测方法 (非 TR-ROP/TR-ROP) 示意图。所提出的方法, MS-ROP 网络和 MN-ROP 网络分别用红色、蓝色和橙色虚线框表示。第一阶段的输入是单张眼底图像。第二阶段的输入是单幅 ROP 眼底图像。

层。平均池化层不需要训练参数,从而避免过拟合,提高模型性能[44,45]。得到检测结果的输出,然后是一个全连接层和一个 softmax 层。

2.2.2.MN-ROP 网络

提出了一种轻量级神经网络 MN-ROP,利用知识蒸馏的方法将 MS-ROP 中的知识转移到 MN-ROP 中,以克服神经网络参数庞大的挑战。提出的 MN-ROP 是基于 Google 在 2017 年发布的 MobileNetV2 神经网络[46]。mn-rop 的主要优点是深度可分离卷积代替常规卷积。MobileNetV1 使用宽度

参数和操作数量的减少也导致了由于特征损失[47]导致的精度下降。因此,提出了 MobileNetV2 作为 MobileNetV1 的改进,以提高网络的表征能力。在 MobileNetV2 中引入线性瓶颈是为了防止非线性层 (ReLU) 丢失一些信息。此外,基于 ResNet 和 DenseNet 等一系列网络实现的性能,引入了倒残差。在 TR-ROP 检测中,提出的 MN-TR-ROP 具有与 MN-ROP 相似的网络结构。此外,在对 MN-TR-ROP 进行训练之前,先将 MN-ROP 的参数转移到



---

**算法 1:**多流融合网络中提取知识的算法。

---

**输入:**眼底图像数据集  $\mathcal{D}$ , ground truth  $\mathcal{Y}$ ,  
epoch  $\epsilon$ , episode  $i$ , 参数  $\theta$ , 温度系数  $T$

**输出:**检测结果  $\hat{Y}$

**1 开始**

2  $\epsilon \leftarrow 0, Y' \leftarrow []$

**3 培训阶段**
**4 重复**

采用 2.1 节中提到的图像增强算法;

6 获得 MS-ROP 的 softmax 分布

;

7 获得 MN-ROP 的 softmax 分布

;

8 通过式(3)计算  $\mathcal{L}_{ROP1}$ ; 用式(4)计算

9  $\mathcal{L}_{ROP2}$ ; 用式(5)计算总损失  $\mathcal{L}$ ; 通过

10 Adam 优化器更新 MN-ROP 的参数;

...

12  $\epsilon \leftarrow \epsilon + 1$ ;

13 **直至**  $\epsilon = \epsilon_{max}$  或激活提前停止标准;

**14 验证阶段**

15 获取验证集  $\mathcal{D}_{val} = \{x_i; i=1, \dots, N\}$ ;

16  $\epsilon \leftarrow 0$

**17 重复**

18 获取预测标签  $\hat{y}_i$  of 输入  $x_i$ ;

19 Append  $y_i$  to  $Y'$ ;

20  $i \leftarrow i + 1$ ;

21 **直到**  $\epsilon = \epsilon_{max}$ ;

**日结束**

---

### 2.2.3. 知识蒸馏和损失函数

算法 1 正式定义了所提方法的过程。一般来说, 越深的网络模型可以获得越高的性能。虽然 MN-ROP 参数较少, 但性能并不理想。知识蒸馏被证明可以有效地压缩模型, 同时保持尽可能好的性能[48,49]。Hinton 等人提出了一种通过与教师相关的软目标[50]进行知识转移的方法。学生模型的 softmax 分布需要与教师模型的 softmax 分布进行匹配, 这也被称为软目标。在提出的方法中, MS-ROP 是教师模型, 而 MN-ROP 是学生模型。广义 softmax 函数定义如下:

$$q_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)} \quad (1)$$

其中  $q_i$  denotes 概率,  $T$  表示温度系数,  $z_i$  denotes logits。一般情况下的 softmax 函数是  $T=1$  的特殊情况。

综上所述, 在 ROP 检测中, 本文提出的网络由三个损失函数组成:(1)MS-ROP 输出与地面真值之间的损失, 记为  $\mathcal{L}_{ROP1}$ ; (2) MN-ROP 输出与地面真值之间的损失, 记为  $\mathcal{L}_{ROP2}$ ; (3) MN-ROP 输出与 MS-ROP 输出之间的损失, 记为  $\mathcal{L}_{ROP3}$ 。  $\mathcal{L}_{ROP1}$  和  $\mathcal{L}_{ROP2}$  是 MN-ROP 提取知识的全部损失的一部分。损失函数定义如下:

$$\mathcal{L}_{ROP1} = \text{CrossEntropy}(Q_S, y_{true}) \quad (2)$$

$$\mathcal{L}_{ROP2} = \text{CrossEntropy}(Q_N, y_{true}) \quad (3)$$

$$\mathcal{L}_{ROP3} = KLdiv(Q_S, Q_N) \quad (4)$$

式中  $Q_S$  denotes MS-ROP 的 softmax 分布,  $Q_N$  denotes MN-ROP 的 softmax 分布,  $y_{true}$  denotes ROP 检测的地真值。  $\mathcal{L}_{ROP1}$  和  $\mathcal{L}_{ROP2}$  采用交叉熵损失函数计算,  $\mathcal{L}_{ROP3}$  采用 Kullback-Leibler 散度损失函数计算。

进一步定义知识提炼 MN-ROP 的总损失为:

$$\mathcal{L} = \alpha T^2 \times \mathcal{L}_{ROP3} + (1 - \alpha) \times \mathcal{L}_{ROP2} \quad (5)$$

式中  $\alpha$  表示温度系数,  $T$  表示  $\mathcal{L}_{ROP3}$  与  $\mathcal{L}_{ROP2}$  之间的比例。受 Hinton 的启发, 我们在实验[50]期间也将  $\alpha$  设为 5, 将  $T$  设为 0.95。

在 TR-ROP 检测中, 这三个损失函数与 ROP 检测中的损失函数相似, 定义如下:

$$\mathcal{L}_{TRROP1} = \text{CrossEntropy}(Q'_S, y'_{true}) \quad (6)$$

$$\mathcal{L}_{TRROP2} = \text{CrossEntropy}(Q'_N, y'_{true}) \quad (7)$$

$$\mathcal{L}_{TRROP3} = KLdiv(Q'_S, Q'_N) \quad (8)$$

其中  $Q'_S$  denotes softmax MS-TR-ROP 分布,

$Q'_N$  denotes MN-TR-ROP 的 softmax 分布

为 TR-ROP 检测中的接地真值。与 ROP 检测一样, 使用交叉熵损失函数  $\mathcal{L}_{TRROP1}$  计算  $\mathcal{L}_{TRROP1} < p$  和  $\mathcal{L}_{TRROP2} < p$ , 使用 Kullback-Leibler 散度损失函数计算  $\mathcal{L}_{TRROP3}$ 。

同理, MN-TR-ROP 提炼知识的总损失定义为:

$$\mathcal{L}' = \alpha T^2 \times \mathcal{L}_{TRROP3} + (1 - \alpha) \times \mathcal{L}_{TRROP2} \quad (9)$$

式中  $\alpha$  表示温度系数,  $T$  表示  $\mathcal{L}_{TRROP3}$  与  $\mathcal{L}_{TRROP2}$  之间的比例。与 ROP 检测一样, 我们在实验中设置  $\alpha$  为 5, 设置  $T$  为 0.95。



表 2

实验装置细节

优化器	↓	学习速率	体重衰变	时代	早期终止标准
亚当	0.9, 0.999	0.0001	0.0001	200	确认损失

表 3

用于训练和验证所提出的网络的数据集

褶皱	训练集				验证设置			
	正常的	罗普	Non-TR-ROP	TR-ROP	正常的	罗普	Non-TR-ROP	TR-ROP
1	1181	353	80	273	278	105	33	72
2	1156	378	87	291	303	80	26	54
3	1166	368	84	284	293	90	29	61
4	1161	373	85	288	298	85	28	57
5	1175	359	81	278	284	99	32	67

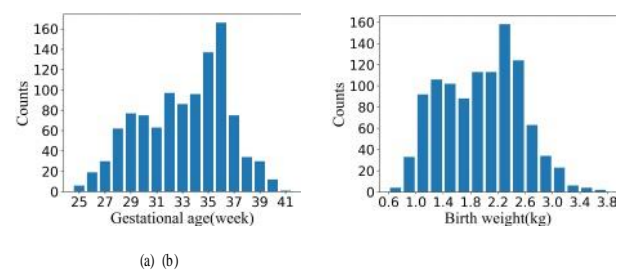


图 6:纳入研究婴儿的胎龄和出生体重直方图。(a)胎龄(周)。(b)出生体重(kg)。

使用 11GB RAM。ResNet18、DenseNet121、EfficientNetB2 和 MobileNetV2 网络使用在 ImageNet 数据集上预训练的权值进行初始化。使用 adam 优化器，初始学习率为 0.0001，权重衰减为 0.0001。对于 MS-ROP 和 MS-tr-ROP 的训练，最大 epoch 数设置为 200，批大小设置为 8。对于 MN-ROP 和 MN-TR-ROP 的训练，最大 epoch 数设置为 200，批大小设置为 16。根据验证丢失情况，采用早期停止策略，及时停止训练，防止模型过拟合。

所有实验均采用五折交叉验证策略。实验结果为采用五折交叉验证策略的 5 个结果的平均值。数据集被细分为 5 个近似相等的部分。在每一层中，网络在四层(80%)上进行训练，并在剩下的一层(20%)上进行验证，以评估性能。用于训练和验证所提出的网络的样本数量如表 3 所示。

#### 4.实验与结果

##### 4.1.ROP 检测结果

表 4 比较了我们提出的方法与最近流行的图像分类算法的性能。每个实验重复 5 次，并且

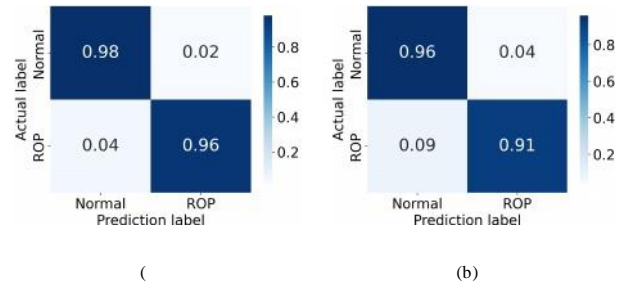
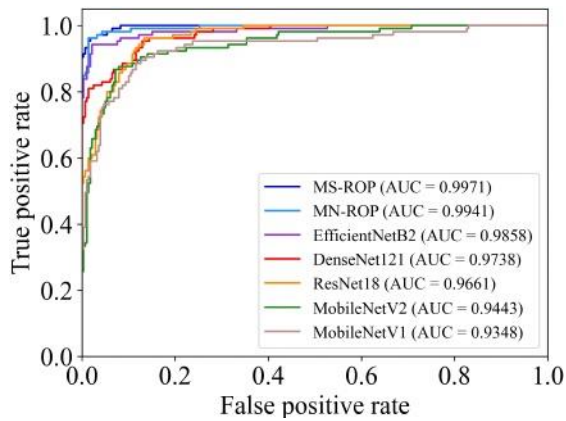


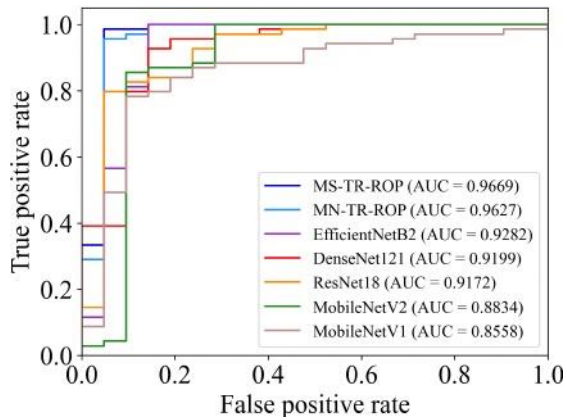
图 7:ROP 检测中 MN-ROP 和 MobileNetV2 网络的两个混淆矩阵。(a) MN-ROP。(b) MobileNetV2 神经网络。两个对角线元素表示在正常/ROP 中正确分类的图像的百分比。

计算平均值作为最终结果。如表 4 所示，与 ResNet18、DenseNet121、EfficientNetB2、MobileNetV1 和 MobileNetV2 网络相比，所提出的 MN-ROP 网络的准确率分别提高了 2.04%、1.52%、2.82%、4.91% 和 2.72%。与 ResNet18、DenseNet121、EfficientNetB2、MobileNetV1 和 MobileNetV2 网络相比，所提出的 MN-ROP 网络的灵敏度分别提高了 0.33%、5.04%、5.67%、9.80% 和 5.44%。其精确度为 0.9423，F1 值为 0.9435，特异度为 0.9823，均优于上述 5 个网络。多流融合神经网络 MS-ROP 在所有实验网络中表现最为突出，准确率为 0.9801，灵敏度为 0.9609，特异性为 0.9862。此外，还将其性能与最近流行的图像分类算法进行了比较，包括 ViT、ResNet-RS、SwinV2、NesT、HaloNetH2 和 DeiT[54,55,56,57,58,59]。实验结果进一步表明了所提方法的有效性。MN-ROP 的性能与 MS-ROP 非常接近，准确度降低 0.67%，灵敏度降低 1.53%，特异性降低 0.39%。然而，MS-ROP 的复杂性远远大于 MN-ROP。





(一)



(b)

图 8:ROP 检测和 TR-ROP 检测网络的 ROC 曲线。(a) ROP 检测(正常/ROP)。(b) TR-ROP 检测(非 TR-ROP/TR-ROP)。

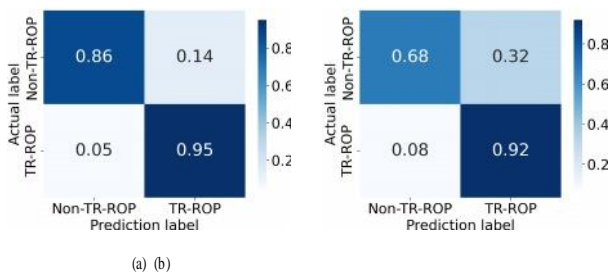


图 9 TR-ROP 检测中 MN-TR-ROP 和 MobileNetV2 网络的两个混淆矩阵。(a) MN-TR-ROP。(b) MobileNetV2 神经网络。两个对角线元素表示在非 TR-ROP/TR-ROP 中正确分类的图像的百分比。

如表 5 所示，MS-ROP 的神经网络参数为 26.83MBytes (M)，与 MN-ROP (3.50M)的数量形成鲜明对比。对于乘法累加操作(mac)，MN-ROP 的计算成本最低(0.33G)，而 MS-ROP 的计算成本要高得多(5.12G)

性能，MN-ROP 和 MobileNetV2 神经网络的混淆矩阵如图 7 所示。MN-ROP 在检测 ROP 方面优于 MobileNetV2。为了更详细地比较 ROC 曲线，图 8(a)显示，除了教师网络 MS-ROP 之外，本文提出的 MN-ROP 在所有实验神经网络中取得了最好的性能。MN-ROP 的 AUC 值分别比 EfficientNetB2、DenseNet121、ResNet18、MobileNetV2 和 MobileNetV1 网络高 0.83%、2.03%、2.80%、4.98% 和 5.93%。

## 4.2.TR-ROP 检测结果

表 6 显示了所提出的网络和其他不同网络在 TR-ROP 检测中的性能。与 ResNet18、DenseNet121、EfficientNetB2、MobileNetV1 和 MobileNetV2 网络相比，所提出的 MN-TR-ROP 网络的准确率分别提高了 4.44%、4.44%、5.55%、9.80% 和 7.78%。与 DenseNet121、MobileNetV1 和 MobileNetV2 网络相比，所提出的 MN-TR-ROP 网络的灵敏度分别提高了 4.84%、8.02% 和 3.22%。但是，它的灵敏度与 EfficientNetB2 网络相同，灵敏度低于 ResNet18 网络。该网络的精度为 0.9365,F1 值为 0.9440，特异度为 0.8571，均优于上述 4 种网络。多流融合神经网络 MS-TR-ROP 在所有实验网络中表现最为突出，准确率为 0.9444，灵敏度为 0.9677，特异性为 0.8929。与 ROP 检测类似，将其性能与最近流行的图像分类算法(包括 ViT, ResNet-RS, SwinV2, NesT, HaloNetH2 和 DeiT)进行比较，以验证所提出方法的有效性。与 MS-TR-ROP 相比，MN-TR-ROP 的准确性降低 2.22%，灵敏度降低 1.61%，特异性降低 3.58%。如前所述，MS-TR-ROP 的复杂性和计算成本远远大于 MN-TR-ROP。为了进一步分析检测性能，图 9 显示了 MN-TR-ROP 和 MobileNetV2 神经网络的混淆矩阵。如图 9 所示，MN-TR-ROP 在检测 TR-ROP 方面取得了优异的性能。同时，也能准确识别出非 TR-ROP 早产儿。并给出 ROC 曲线，如图 8(b)所示。MN-TR-ROP 的 AUC 值为 0.9627，分别比 EfficientNetB2、DenseNet121、ResNet18、MobileNetV2 和 MobileNetV1 网络高 3.45%、4.28%、4.55%、7.93% 和 10.69%。

## 5.讨论

类激活映射(class activation mapping, CAM)是一种帮助 cnn 可视化的工具，它被引入来进一步验证所提出网络的有效性[60]。将得到的 CAM 与原始图像叠加，从而得到

表 4

建议网络 and 不同网络在 ROP 检测中的性能 (正常 /ROP)

网络	精度	精度	l	灵敏度	特异性
ResNet18 [38]	0.9530	0.8909	0.9159	0.9423	0.9570
DenseNet1 21 [39]	0.9582	0.9495	0.9216	0.8952	0.9820
EfficientNet B2 [40]	0.9452	0.8980	0.8934	0.8889	0.9648
MobileNetV1 [47]	0.9243	0.8725	0.8599	0.8476	0.9532
MobileNetV2 [46]	0.9462	0.8863	0.8874	0.8912	0.9646
维特 [54]	0.9321	0.8679	0.8762	0.8846	0.9498
ResNet-RS [55]	0.9608	0.9245	0.9289	0.9333	0.9712
SwinV2 [56]	0.9504	0.9388	0.9064	0.8762	0.9784
巢 [57]	0.9661	0.9259	0.9390	0.9524	0.9712
HaloNetH2 [58]	0.9556	0.8929	0.9122	0.9324	0.9568
Deit [59]	0.9582	0.9320	0.9231	0.9143	0.9748
MS-ROP	0.9801	0.9565	0.9587	0.9609	0.9862
MN-ROP	0.9734	0.9423	0.9435	0.9456	0.9823

表 5

所提出的网络 and 不同网络的参数和计算成本

方法	参数(米)	mac (G)
ResNet-RS [55] Swin V2 [56]	11.69	1.82
HaloNetH2 [58]	7.98	2.88
Deit [59] MS-ROP	9.20	1.03
MN-ROP	4.20	0.57
	36.03	2.30
	88.12	7.70
	9.40	1.75

可以观察到网络更关注的地区。ROP 眼底图像的示例及其对应的类激活映射如图 10 所示。图 10(a)展示了所提方法中正确诊断的眼底图像示例。他们发现，所提出的网络可以集中在早期眼底图像中与 rop 相关的病变区域，也称为脊。图 10(b)展示了所提出方法中被误诊的眼底图像的例子。从图 10(b)中的第一个样本可以看到，虽然网络成功聚焦在细微脊上，但它更关注视盘。从第二个被误诊的样本来看，网络可以聚焦在细微脊上，但遗憾的是，它仍然更多地关注了更亮的区域。从第三和第四个误诊样本中，我们得出结论，光的反射被误认为脊，导致错误的诊断。

TR-ROP 检测对于预防早产儿失明至关重要。图 11 显示了非 TR-ROP 和 TR-ROP 眼底图像的例子。临床上，TR-ROP 的诊断是基于血管扭曲扩张的存在。然而，临床上很难观察到血管迂曲扩张。

此外，由于隐私政策，很难获取医疗数据并对其进行标注。在我们未来的工作中，受 GAN 和自动编码器的启发，我们可能会进行相关的尝试

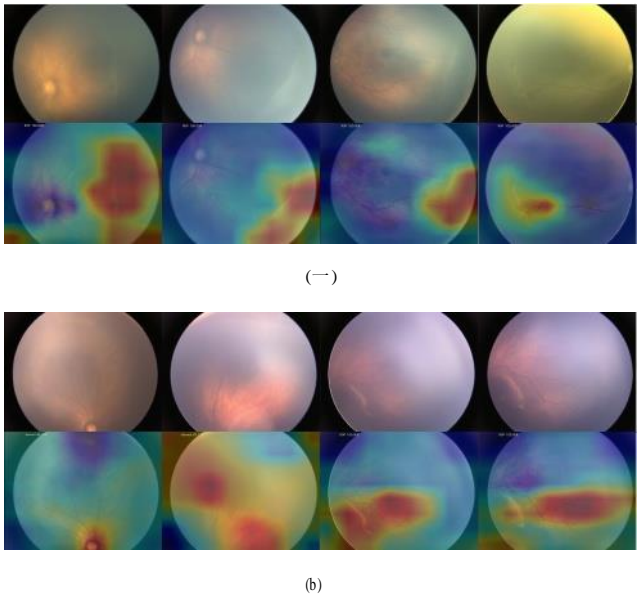


图 10:ROP 眼底图像及其类激活映射的示例。第一行是输入的眼底图像。第二行显示了它们对应的类激活映射。(a)正确诊断眼底图像的示例。(b)眼底图像误诊的示例。

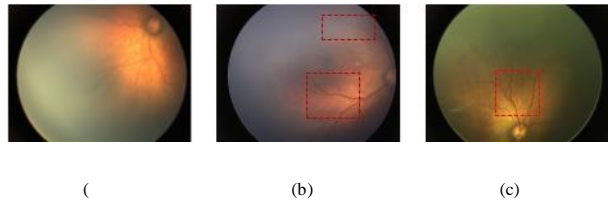


图 11:眼底 ROP 图像的示例。病变区域用红色虚线框标出。(a)非 tr - rop;(b) TR-ROP;(c) TR-ROP;

我们的 ROP 数据集的数据增强[61,62]。此外，为了进一步提供我们提出的方法的可解释性，XAI 将被嵌入我们未来的工作中[63]。

表 6

建议网络 and 不同网络在 TR-ROP 检测中的性能(非 TR-ROP/TR-ROP)

方法	精度	精度	精度	灵敏度	特异性
ResNet18 [38]	0.8778	0.8696	0.9160	0.9677	0.6923
DenseNet121 [39]	0.8778	0.9180	0.9105	0.9032	0.8214
EfficientNetB2 [40]	0.8667	0.8676	0.9077	0.9516	0.6786
MobileNetV1 [47]	0.8242	0.8971	0.8841	0.8714	0.6667
MobileNetV2 [46]	0.8444	0.8636	0.8906	0.9194	0.6865
维特 [54]	0.8571	0.9130	0.9075	0.9021	0.7143
ResNet-RS [55]	0.8901	0.9167	0.9296	0.9429	0.7364
SwinV2 [56]	0.8791	0.9403	0.9205	0.9016	0.8095
巢[57]	0.8913	0.9412	0.9276	0.9143	0.8152
HaloNetH2 [58]	0.8864	0.9275	0.9201	0.9128	0.7619
Deit [59]	0.8893	0.9041	0.9235	0.9437	0.7198
MS-TR-ROP	0.9444	0.9524	0.9600	0.9677	0.8929
MN-TR-ROP	0.9222	0.9365	0.9440	0.9516	0.8571

6.结论

本研究提出了一种从多流融合神经网络中提取知识的轻量级神经网络，用于早期眼底图像的 TR-ROP 检测。TR-ROP 检测神经网络包括 ROP 检测和 TR-ROP 检测两个阶段。多流融合神经网络 MS-ROP 由特征提取模块、融合模块和检测模块组成。利用知识蒸馏将 MS-ROP 中的知识转移到 MN-ROP 神经网络中。在训练和验证该网络时，我们采用了五折交叉验证策略。实验结果表明，所提出的 MN-ROP 在 ROP 检测和 TR-ROP 检测方面都优于流行的神经网络。MS-ROP 的性能最突出，而从 MS-ROP 中提取知识的 MN-ROP 的性能与之非常接近。可视化结果证明了所提网络的有效性。此外，所提出的网络包含非常少的神经网络参数，适合部署到嵌入式 ROP 检测设备中。该网络可以帮助眼科医生进行 ROP 的临床诊断和治疗。

参考文献

[1] Parag K Shah, Vishma Prabhu, Smita S Karandikar, Ratnesh Ranjan, Venkatapathy Narendran 和 Narendran Kalpana。早产儿视网膜病：过去、现在和未来。*世界临床儿科学杂志*，5(1):35,2016。

[2] Rebecca J Vartanian, Cagri G Besirli, John D Barks, Chris A Andrews, David C Musch。早产儿视网膜病变的筛查和治疗趋势*儿科*，139(1), 2017。

[3] Audina M Berrocal, Kenneth C Fan, Hasenin Al-Kharsan, catherine I Negron, Timothy Murray。早产儿视网膜病变:使用单中心方法筛查和治疗早产儿视网膜病变的进展*美国眼科杂志*，233:189 - 215,2022。

[4]张晓明，张晓明，张晓明，Michael F Chiang。詹姆斯·雷诺兹，张晓明。筛查性检查

早产儿视网膜病变*儿科*，142(6)，2018。

[5]罗奈尔德·m·汉森，安妮·莫斯科维茨，詹姆斯·d·阿库拉，安妮·b·富尔顿。早产儿视网膜病变中的神经视网膜。*视网膜与眼研究进展*，56:32-57,2017。

[6] S Grace Prakalapakorn, Lucy Greenberg, Erika M Edwards, Danielle EY Ehret。2008-2018 年早产儿视网膜病变筛查和治疗的趋势*儿科*，147(6)，2021。

[7]张晓明，张晓明，李晓明，Thomas M Lietman, Michael F Chiang，中国眼科协会，等。发展中国家眼科护理面临的挑战。*JAMA*，2014,32(5):644 - 644。

[8] Parveen Sen, Weichi Wu, Parijat Chandra, Anand Vinekar, Pradeep T Manchegowda, Pramod bende。早产儿视网膜病变的治疗:亚洲人的观点*眼科*，34(4):632-642,2020。

[9]张建军，张建军，张建军。2010 年区域和全球水平早产儿相关视力损害和早产儿视网膜病变的估计*儿科研究*，74(1):35-49,2013。

[10]李建军，李建军，李建军，李建军。早产儿视网膜病变:危险因素及其临床意义的综述*眼科调查*，63(5):618-637,2018。

[11]吴田，张莉，佟宇，曲毅，夏斌，穆德智。中国极低出生体重儿的早产儿视网膜病变:发病率和围生期危险因素*眼科研究与视觉科学*，59(2):757-763,2018。

[12] Juan Carlos Romo-Aguas, Ana González-H León, Miroslava Paolah Meraz-Gutiérrez, 和 María A Martínez-Castellanos。早产儿视网膜病变:基于国际筛查指南的异常值发病率报告*国际视网膜和玻璃体杂志*，5(1):1 - 6,2019。

[13] Kelly C Wade, Guishuang Ying, Agnieszka Baumritter, Alice Gong, Alex R Kemper, Graham E Quinn 等。早产儿发生早产儿视网膜病变风险低的相关因素*JAMA*，2013(2):160-166。

Girija Natarajan, Seetha Shankaran, Tracy L Nolen, Amaanti Sridhar, Kathleen A Kennedy, Susan R Hintz, Dale L Phelps, Sara B DeMauro, Waldemar A Carlo, Marie G Gantz 等人。通过治疗早产儿视网膜病变早产儿的神经发育结局*儿科*，144(2)，2019。

[15] André Moraes Freitas, Ricardo Mörschbacher, Mariana R Thorell 和 Ernani Luis Rhoden。早产儿视网膜病变的发病率和危险因素:一项回顾性队列研究 *国际视网膜与玻璃体杂志*, 4(1):1 - 8,2018。

[16]张继雄。早产儿视网膜病变发生发展的危险因素分析 *公共科学图书馆*, 14(7):e0219934, 2019。



[17]戴毅, 朱丽, 周业群, 吴艳秋, 陈冬梅, 王瑞全, 吴艳秋, 郭少青, 高玲玲, 徐萍, 王洋, 董珊珊, 刘宁, 吴艳秋, 陈浩明, 孙玉梅, 陈超, 张淑莲。中国超早产儿视网膜病变治疗的发生率 《*儿科和围产期流行病学*》, 2021 年。

[18] Leah A Owen, Margaux A Morrison, Robert O Hoffman, Bradley A Yoder 和 Margaret M DeAngelis。早产儿视网膜病变:预防和预测疾病的综合风险分析。 *公共科学图书馆*, 12(2):e0171467, 2017。

[19] 桂双英, Graham E Quinn, Kelly C Wade, Michael X Repka, Agnieszka Baumritter, Ebenezer Daniel, 等。远程医疗评估急性期早产儿视网膜病变(e-rop)研究中早产儿转诊视网膜病变发生的预测因素 *JAMA*, 2013(3):394 - 391。

[20]刘建军, 刘建军, 刘建军, 刘建军, 刘建军, 刘建军, 刘建军, 刘建军, 刘建军, 刘建军, 等。使用出生特征进行威胁视力的早产儿视网膜病变的个体风险预测。 *JAMA*, 2013(1):21-29,2020。

[21] Gil Binenbaum, 桂双英, Lauren A. Tomlinson, 早产儿视网膜病变(G-ROP)研究小组。费城儿童医院早产儿视网膜病变(CHOP ROP)模型的验证 *JAMA*, 2017,35(8):871-877。

[22]Gil Binenbaum, 詹姆斯·谢弗, 劳伦·汤姆林森, 英桂爽。早产儿视网膜病变的预测模型:出生后生长和 rop (g-rop)研究的主要结果 *中国儿童眼科学与斜视杂志*, 21(4):559 - 559,2017。

[23] GU Eckert, JB Fortes Filho, M Maia, RS Procianoy。极低出生体重早产儿早产儿视网膜病变的预测评分 *眼科*, 26(3):400-406,2012。

[24]李晓明, 杨晓明, 李晓明, 李晓明。早产儿视网膜病变预测的临床模型和算法:美国眼科学会的一份报告 *眼科学杂志*, 2016(4):444 - 444。

[25]李建平, 李建平, 李建平, Gerd Holmström, Kristina Tornqvist, Abbas Al Hawasi, Åsa Nilsson。在过去 10 年中, 早产儿视网膜病变的发生率增加, 并且存在显著的地区差异。 *眼科学报*, 96(2):142-148,2018。

[26]蒋 Michael F, Graham E Quinn, Alistair R Fielder, Susan R Ostmo, RV Paul Chan, Audina Berrocal, Gil Binenbaum, Michael F, J Peter Campbell, Antonio Capone Jr 等。早产儿视网膜病变的国际分类。 *眼科学杂志*, 28(10):551 - 568。

[27] 詹姆斯 M Brown, J Peter Campbell, Andrew Beers, Ken Chang, Susan Ostmo, RV Paul Chan, Jennifer Dy, Deniz Erdogmus, Stratis Ioannidis, Jayashree Kalpathy-Cramer, 等。利用深度卷积神经网络实现早产儿视网膜病变附加疾病的自动诊断。 *JAMA*, 36(7):803-810,2018。

[28] Zachary Tan, Samantha Simkin, Connie Lai, and Shuan Dai。深度学习算法用于早产儿视网膜病变附加疾病的自动诊断。 *转化视觉科学与技术*, 8(6):23 - 23,2019。

[29]胡俊杰, 陈圆圆, 钟杰, 鞠荣, 张一。利用深度神经网络对早产儿视网膜病变进行自动化分析。 *IEEE*, 38(1):269-279,2018。

[30]彭圆圆, 朱卫方, 陈凤, 王婷, 耿乐, 余凯, 周一, 王婷, 向道曼, 陈凤, 等。基于深度特征融合和有序分类策略的早产儿视网膜病变自动分期。 *IEEE 医学影像学报*, 2021。

[31] Ranjana Agrawal, Sucheta Kulkarni, Rahee Walambe and Ketan Kotecha。自动检测所有区域的辅助框架  
深度学习在早产儿视网膜病变中的应用 *数字影像*, 34(4):932-947,2021。

[32] Ranjana Agrawal, Sucheta Kulkarni, Rahee Walambe, Madan Deshpande, and Ketan Kotecha。利用深度学习对早产儿视网膜病变进行视网膜眼底图像分割的深入研究。 *多媒体工具与应用*, 81(8):11441-11460,2022。

陈少斌, 张汝刚, 陈国珍, 赵金凤, 王天福, 张国明, 雷柏英。注意力引导深度多实例学习在早产儿视网膜病变分期中的应用。 *2021 年 IEEE 第 18 届生物医学成像国际研讨会(ISBI)*, 1025-1028 页。IEEE, 2021。

[34]唐志强、高云和、Leonid Karlinsky、Prasanna Sattigeri、Rogerio Feris、Dimitris Metaxas。在线增强:用较少的领域知识进行在线数据增强。在 *欧洲计算机视觉会议*上, 第 313-329 页。施普林格, 施普林格国际出版, 2020 年。

[35] Sima Sahu, Amit Kumar Singh, SP Ghrera, Mohamed Elhoseny 等。一种基于 clahe 的视网膜眼底图像去噪和对比度增强方法。 *光学与激光技术*, 110:87-98,2019。

[36] Mohammed J Alwazzan, Mohammed A Ismael, 和 Asmaa N Ahmed。一种使用维纳滤波器和 clahe 增强彩色视网膜眼底图像的混合算法。 *数字影像学报*, 34(3):750-759,2021。

[37]谢灵曦, 田琦, 张博。简单的技术有意义:图像分类的特征池化和归一化。 *IEEE 视频技术学报*, 26(7):1251-1264,2015。

[38]何开明, 张翔宇, 任少卿, 孙健。用于图像识别的深度残差学习。 *IEEE 计算机视觉与模式识别会议论文集*, 770 - 778 页, 2016。

[39] Huang Gao, Liu Zhuang, Laurens Van Der Maaten, Kilian Q Weinberger。密集连接的卷积网络。 *IEEE 计算机视觉与模式识别会议论文集*, 第 4700-4708 页, 2017。

[40] Tan Mingxing 和 Quoc Le。Efficientnet:重新思考卷积神经网络的模型缩放。 *《机器学习国际会议》*, 第 6105-6114 页。PMLR, 2019 年。

[41]宋薇薇、李书涛、方乐源、陆婷。利用深度特征融合网络进行高光谱图像分类。 *IEEE 与遥感学报*, 36(6):373 - 3184,2018。

[42]李鲜, 丁明丽, Aleksandra Pižurica。通过双流卷积神经网络进行深度特征融合, 用于高光谱图像分类。 *IEEE 与遥感学报*, 58(4):2615-2629,2019。

[43]王品, 王嘉欣, 李永明, 李普飞, 李林宇, 蒋明峰。基于深度特征融合和增强路由的乳腺癌组织病理学图像自动分类。 *生物医学信号处理与控制*, 65:102341,2021。

[44]刘英禄, 张彦明, 张旭耀, 刘成林。用于图像分类的自适应空间池化。 *模式识别*, 55:58-67,2016。

[45]李海韵、朴晋亨、黄宰胤。用于超声图像中乳腺癌分割的多尺度网格平均池化通道注意力模块。 *IEEE*, 2016,35(7):344 - 344,2016。

[46] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen。Mobilenetv2:倒残差和线性瓶颈。 *IEEE 计算机视觉与模式识别会议论文集*, 第 4510-4520 页, 2018。

[47] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, 和 Hartwig Adam。Mobilenets:用于移动视觉应用的高效卷积神经网络。 *arXiv 预印本 arXiv:1704.04861*, 2017。

[48]王林和尹国珍。视觉智能的知识提炼与师生学习:综述与新展望。 *IEEE 模式分析与机器学习学报*,

- [49] 苟建平、余宝生、Stephen J Maybank、陶大成。知识蒸馏:一项调查。*计算机视觉学报*, 129(6):1789 - 1819,2021。
- [50] Geoffrey Hinton, Oriol Vinyals, 和 Jeff Dean。在神经网络中提取知识。*arXiv 预印本 arXiv:1503.02531*, 2015。
- [51] 李晓欣。中国早产儿视网膜病变筛查指南(2014)。*中华眼科杂志*, 50(12):933-935,2014。
- [52] 杨倩, 周晓红, 倪应钦, 单海东, 史文静, 尹雪蕾, 刘江琴, 黄欣。基于 5 年队列研究的优化中国早产儿视网膜病变筛查指南 *英国眼科杂志*, 105(6):819-823,2021。
- 早产儿视网膜病变早期治疗协作组等。早产儿视网膜病变的发病率和病程:早产儿视网膜病变早期治疗研究的结果 *中华儿科杂志*, 116(1):15 - 23,2005。
- Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, 翟晓华, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly 等。一张图片值 16x16 个字:变形金刚用于图像识别的尺度。*arXiv 预印本 arXiv:2010.11929*, 2020。
- [55] 刘建军, 刘建军, 刘建军, 刘建军。revisited resnets:改进的训练和缩放策略。《*神经信息处理系统进展*》, 34:22614-22627,2021。
- [56] 刘泽, 胡韩, 林宇通, 姚祝良, 谢振达, 魏以轩, 宁佳, 曹悦, 张, 董丽, 等。Swin 变压器 v2:放大容量和分辨率。*IEEE/CVF 计算机视觉与模式识别会议论文集*, 第 12009 - 119,2022 页。
- [57] 张自钊, 张韩, 赵龙, 陈婷, Serkan Ö. 嵌套分级变压器:迈向准确、数据高效、可解释的视觉理解。《*AAAI 人工智能会议论文集*》, 第 36 卷, 3417-3425 页, 2022 年。
- [58] Ashish Vaswani, Prajit Ramachandran, Aravind Srinivas, Niki Parmar, Blake Hechtman 和 Jonathon Shlens。缩放局部自我注意, 用于参数高效的视觉骨干。*IEEE/CVF 计算机视觉与模式识别会议论文集*, 12894 - 12904,2021。
- [59] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, 和 Hervé Jégou。训练数据高效的图像转换器 & 通过注意力进行提炼。《*国际机器学习会议*》, 第 10347-10357 页。PMLR, 2021 年。
- [60] Bolei Zhou、Aditya Khosla、Agata Lapedriza、Aude Oliva 和 Antonio Torralba。学习深度特征进行判别性本地化。*IEEE 计算机视觉与模式识别会议论文集*, 第 2921-2929 页, 2016。
- [61] Poonam Chaudhari, Himanshu Agrawal, 和 Ketan Kotecha。使用 mg-gan 对基因表达数据进行数据增强以改进癌症分类。*软计算*, 24(15):11381-11391,2020。
- [62] Jane Saldanha, Shaunak Chakraborty, Shruti Patil, Ketan Kotecha, Satish Kumar, and Anand Nayyar。使用变分自编码器进行数据增强以改进呼吸系统疾病分类。*Plos one*, 17(8):1 - 41,08 2022。
- [63] Gargi Joshi, Rahee Walambe, Ketan Kotecha。多模态深度神经网络的可解释性综述。*IEEE 接入*, 9:59800 - 59821,2021。