

第 18 届 IEEE 国际生物医学成像研讨会(ISBI)将于 2021 年 4 月 13 日至 16 日在法国尼斯举行

注意引导深度多实例学习用于早产儿视网膜病变分期

陈¹少斌, 张¹汝刚, 陈¹国珍, 赵²金凤, 王¹天福, 张²国明, 雷白英

¹ 深圳大学健康科学中心生物医学工程学院医学超声国家区域重点技术工程实验室, 广东省生物医学测量与超声成像重点实验室, 深圳 518060 (e-mail: leiby@szu.edu.cn)

² 深圳市眼科医院, 深圳市眼科重点实验室, 暨南大学第二附属医院, 深圳, 中国。

摘要

早产儿视网膜病变(ROP)是儿童获得性失明的最常见原因之一。ROP 分期是评价 ROP 严重程度对疾病控制和管理的重要一步。然而, ROP 分期仍然存在各种挑战, 因为 ROP 的模式与整个眼底图像相比相对模糊。此外, 数据集很小, 图像质量相当差。为了解决这些问题, 我们开发了一个多实例学习(MIL)网络, 它可以提取图像的特征, 这些特征可以通过全卷积网络(FCN)增强。FCN 生成的空间分数图(SSM)被裁剪成小块, 并馈送到所提出的 MIL 中进行进一步的特征学习。利用注意机制指导 MIL 池, 可以关注不同阶段的 ROP 特征, 提高分级效果。在内部 ROP 数据集上对所提出的网络进行了评估, 实验结果表明, 我们提出的方法在 ROP 阶段是有希望的。

索引术语-多实例学习, 早产儿视网膜病变分期, 全卷积网络。

1.介绍

早产儿视网膜病变(Retinopathy of prematurity, ROP)[1]是低出生体重儿常见的视网膜疾病, 是导致儿童失明的主要原因。1984 年, 来自 11 个国家的 23 名眼科医生共同制定了 ROP 国际分类(International Classification of ROP, 简称 ICROP)[2], 为 ROP 的分类提供了临床指导。根据视网膜在血管和非血管处的表现, 我们用五个阶段来表征 ROP 的程度。如果血管视网膜和无血管视网膜之间有一条“分界线”, 我们将其定义为第 1 阶段。这条分界线比较平坦, 上面有不寻常的血管分支。在第二阶段, “分割线”出现“脊”, 高度和宽度都有所增加。当纤维血管增生的症状出现在后侧神经管外或脊处时, 我们可以将其定义为第 3 期。第 4 期有部分视网膜脱离的症状, 第 5 期

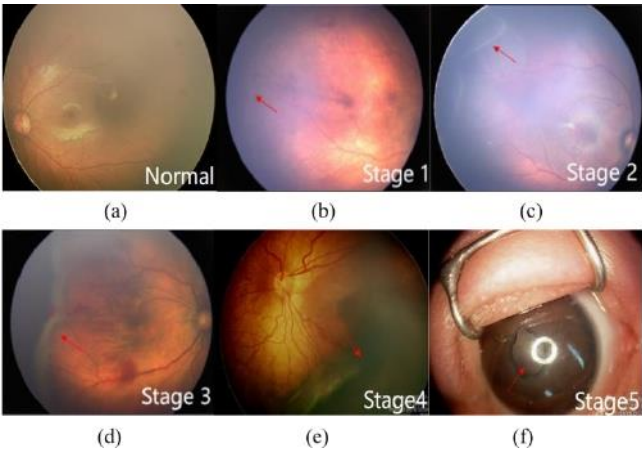


图 1 所示。正常(a)和不同阶段的 ROP。第 1 期(b) ROP 在有血管和无血管交界处出现轻微分界线(红色箭头)。阶段 2 (c)出现脊状突起, 分割线的宽度和高度增加。第三阶段(d)出现后部或脊部神经管外的纤维血管增生。第 4 期(e)出现部分视网膜脱离, 第 5 期(f)出现完全视网膜脱离。

发生完全性视网膜脱离。临床上, 第 4 期和第 5 期比较明显, 可通过双目间接检眼镜观察, 是诊断 ROP 的金标准方法[3]。然而, 第 1 期、第 2 期和第 3 期的外观相似, 因此诊断需要经验丰富的眼科医生和 RetCam3 等广角数字视网膜成像系统的图像记录。正常、阶段 1、阶段 2、阶段 3、阶段 4 和阶段 5 的 ROP 分别如图 1 所示。

ROP 的分期比较复杂, 原因如下:1)早期 ROP(1 期和 2 期), ROP 病灶外观不清晰, 与眼底整体图像相比, 病灶区域较小(见图 1);2)不同阶段 ROP 的外观相似;3)临床对 ROP 分期的评价主要依靠眼科医生对症状的主观解读, 需要专业知识, 耗时较长。为了解决这些挑战, 需要一种提取高级特征和小数据集鲁棒性的方法来完成 ROP 阶段任务。鉴于此, 我们设计了一个全卷积网络(fully convolutional network, FCN)[4]来提取 ROP 特征并生成空间评分图(spatial score map, SSM), 该图提供了 ROP 病变的像素级概率。由于 ROP 数据集很小, 为了获得鲁棒的分期性能, SSM 被裁剪成小块来增加数据集, 并被馈送到多实例学习(MIL)[5]网络中进行训练。然而, 尽管不同的 MIL 轮询方法

978-1-6654-1246-9/21/\$31.00©2021 IEEE

授权许可使用仅限于:阿伯丁大学。2023 年 9 月 25 日 13:24:07 UTC 下载自 IEEE Xplore。限制适用。

曾应用于提高网络性能，结果仍不尽人意。此外，对SSM和斑块大小的研究还不够深入。本文采用注意模块对提取的特征进行基于袋级池化的合并，得到ROP阶段的最终类别。比较不同的patch大小，优化网络性能。此外，为了扩大数据集并增加网络的鲁棒性，我们收集了来自多家医院的数据，并对三位眼科医生的数据进行了注释。实验结果表明，与传统模块相比，该方法在ROP阶段检测方面取得了较好的效果。总的来说，我们的贡献如下：

- 1)提出的框架包括两个主要模块，FCN模块完成端到端分割任务，MIL模块生成SSM，提供ROP病变的逐像素概率。
- 2)将SSM裁剪成不同大小的小块以适应MIL设置，有效地扩展了数据集的大小。
- 3)采用注意模块对基于袋级池化的特征进行合并，得到ROP阶段的最终类别。

2.方法

2.1.全卷积网络模块

我们提出的模型的架构如图2所示。引入FCN从眼底图像中提取局部特征，生成SSM。为了获得像素级输出(即SSM)，从我们的模型中移除卷积神经网络(CNN)的全连接(FC)层，然后进行上采样。该方法已被证明是一种有效的端到端像素级分割方法。我们使用SSM作为MIL输入，而不是原始图像。

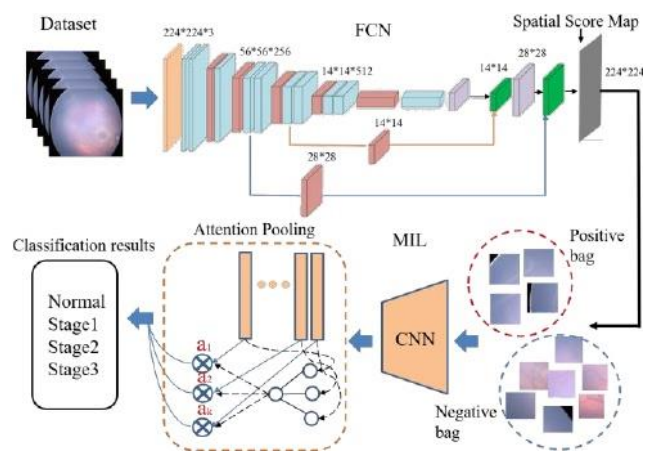


图2所示。我们的模型架构。该框架包括两个模块，FCN模块和MIL模块。

在本文中，首先将视网膜图像的原始尺寸降采样到 $224 \times 224 \times 3$ ，并经过四对卷积层和池化层。在全FC层，通道从 $\{4096; 4096; 1024\}$ 到 $\{1024; 1024; C\}$ ，其中C为任务中类别的个数，因此第一FC层的权值矩阵的大小为

$1024 \times 1024 \times 14$ 。对于卷积表示，第一个FC层的权重被重塑成一个大小为 $1024 \times 512 \times 7 \times 7$ 的4维张量。在FC层之后进行注意模块，特征图的大小恢复到 $224 \times 224 \times C$ (即SSM)。

2.2.多实例学习网络

深度学习用于医学图像分类面临两个常见的挑战:1)整个医学图像的兴趣区域(roi)相对较小，边界不明确;2)数据集规模小且不平衡。对于第一个挑战，roi对全局特征向量的影响是有限的，这严重影响了所得分类器的性能。MIL网络首先引入弱标记来解决药物活性预测问题[6]，并已成功用于检测医学图像分类问题。MIL是医学图像分类的自然学习场景，因为它允许在图像中局部自动检测目标模式，并对每个图像作为一个整体提出自动诊断。MIL是监督学习的延伸。它可以训练分类器使用这样的弱标记数据，这些数据位于这样的包中，而包中的单个实例没有标签。在MIL中，如果有一个正实例，则假定袋子是正确的，如果袋子中的所有实例都是负的，则假定袋子是负的。最初，MIL被公式化为一个二元分类问题，但也提出了多类扩展。MIL在医学成像学习中的另一个优势是它可以处理单张图像。

为了维持MIL假设并更好地处理ROP阶段任务中的小数据集问题，通过将FCN产生的SSM裁剪为大小相等的小的不重叠的小块，并将这些小块输入MIL进行训练，对网络进行了修改。这样，训练图像就被放大了几十倍。SSM是一个实例包，我们称之为包，其中的每个补丁也是一个实例。在本研究中，总共使用了5个补丁大小来评估网络性能。

2.3.细心的MIL池

在本研究中，我们使用MIL模块学习一个函数，该函数可以将输入数据集 $\{X_1, \dots, X_N\}$ 映射到相应的标签集 $\{Y_1, \dots, Y_N\}$ ，其中 $Y_i \in \{0, 1\}$ 。假设 X_k 是一个有许多实例的袋子，其中 $k \in \{1, \dots, N\}$ 。然后将袋子中的实例定义为 $\{x_{k1}, \dots, x_{kn}\}$ ，其中n是袋子 X_k 的实例总数，解决MIL问题的对称函数表示为：

$$f(X) = \theta(\eta_{x \in X} \varphi(x)), \tag{1}$$

其中，变量域中的变量是合适的变换， \square 是函数的置换不变性。 θ 是包中实例的评价函数。作为桥接实例到包的重要步骤，不同的应用程序对MIL池方法有不同的偏好。在本研究中， X_{kj} is的预测概率记为 =

($c = 1$), c 是我们的任务的类别和 $c \in \{1, \dots, c\}$ 。包 x 的聚合函数 ϕ_k 可以定义如下:

$$p_k^c = P(c_k = c | X_k) = \mathcal{F}(p_{k1}^c, \dots, p_{kn}^c). \tag{2}$$

本研究将找到聚合函数的最佳性能。

注意 MIL 池的目的是通过训练神经网络为实例分配权重。在我们提出的方法中，它被用于袋级。此外，权重之和必须为 1 才能不受袋子大小的影响。加权平均满足权重和嵌入都是 ϕ 函数的一部分的要求。设 $\phi = \{ \langle s_1:1 \rangle, \dots, \langle e:1 \rangle \}$ 是一袋 k 个嵌入，那么我们提出如下的 MIL 池:

$$z = \sum_{k=1}^K a_k h_k, \tag{3}$$

$$a_k = \frac{\exp \{W^T \tanh(Vk_k^T)\}}{\sum_{j=1}^K \exp \{W^T \tanh(Vk_j^T)\}}, \tag{4}$$

此外，我们利用双曲正切元素的非线性来包含适当梯度流的正负值。在本研究中，在每个实例上使用 max 池化和 Softmax 池化作为实例级 MIL 池化来比较 MIL 分类器的性能。

3.实验与结果

3.1. 数据集

我们的数据集包括 2015 年至 2018 年来自深圳眼科医院、深圳妇幼保健院和潍坊眼科医院的 1558 例 ROP 病例。在每次检查时，都会拍摄一张婴儿眼睛的标准 10 视图照片。我们只选择与所有眼科医生的眼罩一致的图像进行训练，将标记不一致的图像丢弃。最后，我们从 893 次 ROP 检查中得到 6209 张视网膜图像。为了评估我们方法的性能，我们使用了两个数据集来比较网络性能。最后，我们将图像分别划分为交叉验证集、训练集和测试集。数据集 I 和数据集 II 将图像分为四组，包括 Stage 1, Stage 2, Stage 3 和 Normal。然而，在数据集 I 中选择一个与 Normal 类的其他类相似的数据大小进行训练，而在训练过程中使用数据集 II 中的 2400 张 Normal 图像。在数据集 I 中，有 1000 个标记为 Normal, 360 个标记为 Stage 1。在数据集 II 中，在所有婴儿中，有 4000 个标记为 Normal，只有 360 个标记为 Stage1。此外，在数据集 I 和数据集 II 中，分别有 1064 个标记为阶段 2 和 785 个标记为阶段 3。由此可见，数据集是相当不平衡的。数据集 I 和数据集 II 用于验证所提出的不平衡数据学习的有效性。

3.2. 实验装置

我们在 NVIDIA TITAN XP GPU 上使用 Pytorch 框架，RAM 为 12 GB，可以加快我们的工作速度。为了节省计算资源，我们还调整了

训练数据集。通过将每个调整后的图像旋转 90 度，我们将数据集扩充到 4 倍。在我们的任务中，我们将最大 epoch 设置为 100，并将批处理大小设置为 10。初始学习率设置为 0.0001，然后在训练损失收敛时减少 0.9 的衰减。

为了评估我们提出的方法，我们使用了许多常见的评估标准，如特异性(SPEC)、曲线下面积(AUC)、灵敏度(SEN)、准确性(ACC)和 F1 分数(F1)。针对不同的模型，对 ROP 各阶段的总分类精度和测试精度进行了评价。

3.3.整体分期表现

为了强调 FCN 网络的有效性，我们使用了三种 MIL 池[7]算法，包括 Max 池化(MP)、Softmax 池化(SP)和注意力池化(AP)。我们评估了不同 MIL 池算法的性能，并与所提出的具有和不具有 FCN 的 MIL 网络进行了比较，用于特征提取和聚合过程中不同的 MIL 池。没有 FCN 特征提取的 MIL 意味着眼底图像被调整大小并直接裁剪成小块并输入到聚合过程中。如表 1 所示，在数据集 I 中，使用三个 MIL 池的所提出模型的阶段 1(S1)、阶段 2(S2)、阶段 3(S3)、Normal (NC)和 Total(T)的分类准确率(%)大大优于不使用 FCN 的模型。

结果表明，FCN 分割在提取眼底图像的高级特征方面具有较好的效果，极大地提高了 MIL 的分类性能。我们提出的注意机制成功地使神经网络能够更多地关注最可能的积极实例。通过这种方式，模型可以更具可解释性，并且可以从眼底图像中检测到关键信息。为了验证不平衡学习算法的性能，我们从数据集 II 中随机选择 1000 张正常眼底图像，并构建新的数据集 I。使用数据集 II 的表 1 中的模型性能比数据集 I 的性能要好得多。

3.4.分期性能对比

在这项研究中，我们使用了一些常见的模型，如 AlexNet[8]、VGG[9]、ResNet50[10]、ResNet101 和 Inceptionv4[11]来比较我们模型的性能。数据集 II 中的实验结果如表 2 所示。我们可以看到，不同模型之间的分期性能存在明显差异。我们的方法，FCN+MIL，在所有评估指标方面具有最佳性能。事实上，我们方法的所有评价指标都比其他所有模型好得多。我们的方法采用 FCN+MIL 框架和其他网络的接收者工作特征(ROC)曲线如图 4 所示。

表 1. 我们的模型与没有 FCN 的 MIL 网络的比较(%)。

	P	年代	无 FCN 的 MIL			FCN + MIL		
			ACC	AUC	F1	ACC	AUC	F1
数据集我	M P	S1	80.3	88.4	80.1	87.53	91.1	86.6
		s2	87.1	91.2	86.2	89.27	94.3	88.3
		s3 Ⅱ	87.2	91.5	86.7	89.3	94.6	88.4
		t	90.7	94.1	89.5	92.7	96.7	91.1
			88.3	93.5	87.4	91.5	95.8	90.1
	SP	S1	78.5	87.6	78.1	86.3	90.5	85.55
		s2	83.3	89.5	82.9	89.1	94.2	88.28
		s3 Ⅱ	85.5	90.5	85.0	88.9	93.7	88.1
		t	90.0	93.8	89.3	92.2	96.3	91.7
			88.1	93.1	87.6	90.9	95.1	89.8
	美联社	S1	81.4	89.2	81.9	89.3	93.8	88.63
		s2	85.2	90.7	84.7	90.1	94.7	89.24
		s3 Ⅱ	86.3	92.1	85.6	91.3	94.7	90.1
		t	91.9	95.3	91.0	93.5	96.8	92.6
			89.5	93.8	88.3	92.4	95.9	91.5
数据集 2	M P	S1	83.5	90.2	82.7	88.7	93.4	87.86
		s2	89.4	92.8	88.8	92.5	95.5	91.53
		s3 Ⅱ	90.4	93.1	89.6	92.7	95.7	91.6
		t	92.5	96.3	91.7	95.5	97.2	94.3
			89.3	93.5	88.4	93.2	96.3	92.1
	SP	S1	82.8	89.5	82.2	87.8	92.3	87.12
		s2	88.9	93.2	88.1	91.5	94.5	89.97
		s3 Ⅱ	89.7	93.7	87.8	91.2	95.0	89.7
		t	92.2	95.1	91.8	94.7	96.6	93.2
			88.4	93.1	87.6	92.8	95.8	91.3
	美联社	S1	85.3	90.7	84.5	90.2	94.7	89.2
		s2	90.3	95.2	89.5	94.1	96.8	93.0
		s3 Ⅱ	91.1	96.1	90.4	93.8	96.3	91.9
		t	95.4	98.3	94.2	96.7	98.5	95.2
			90.2	95.1	89.0	94.4	97.2	92.2

表 2. 所提模型和纯 MIL 网络的性能(%)。

方法	ACC	SEN	SPEC	AUC	F1
VGG16	86.53	84.22	85.33	86.4	86.31
AlexNet	87.21	88.33	84.54	84.5	83.64
ResNet50	79.46	76.87	80.27	81.5	78.75
Inception V4	87.23	88.35	84.52	84.3	83.62
MobileNet	89.42	88.32	88.91	89.8	87.53
Ours	94.43	92.71	94.23	97.2	92.27

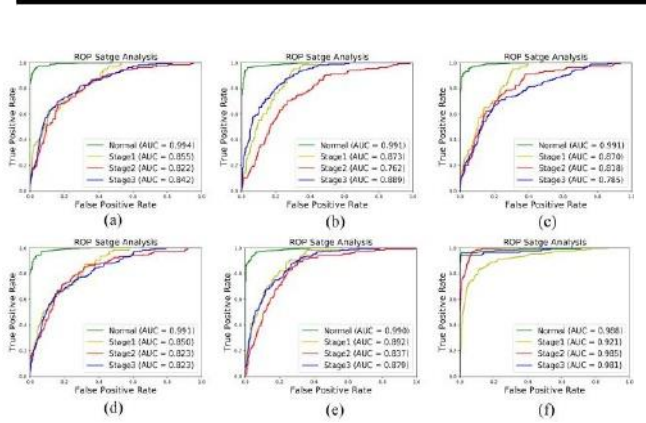


图 4所示。采用 FCN+ MIL 框架和比较网络，利用所提出的方法绘制了各类的 ROC 曲线。

4.结论

在本文中，我们提出了一种用于 ROP 分期的注意力引导 FCN+MIL 网络，其中 FCN 负责提取眼底图像的高级特征并生成 SSM。SSM 是与原始图像在相同空间位置的 ROP 病变的像素级概率，将其裁剪成不同大小的小块并输入 MIL

网络进行学习。为了进一步解决 ROP 阶段任务中轻微模式的关键问题，

利用注意机制指导 MIL 池，可以关注不同阶段的 ROP 特征，提高分级效果。实验结果表明，该方法是有利的，并有望实现 ROP。

5. 符合伦理标准利益冲突 作者声明他们对这项工作没有利益冲突。

6.鸣 谢

国家自然科学基金(No. 61871274、U1909209);广东医学图像处理重点实验室(No. 61871274、U1909209);广东珠江人才计划项目(2016ZT06S220)、深圳市孔雀计划项目(KQTD2016 053112051497、KQTD2015033016104926)、深圳市重点基础研究项目(JCYJ20180507184647636、JCYJ20170818094109846)。

7.参考文献

[1] A. Hellström, I. E. Smith, 和 O. J. T. I. Dammann, “早产儿视网膜病变”, vol. 382, no. 1. 9902, pp. 1445-1457, 2013。
[2] 王志强, 王志强, 王志强, “早产儿视网膜病变的国际分类再探讨”, 中华眼科杂志, 第 1 卷, 第 2 期。7, p. 991, 2005。

[3] 陈建军, 陈建军, 陈建军, 陈建军, 陈建军, “基于深度卷积神经网络的早产儿视网膜病变的自动诊断”, 中华眼科杂志, vol. 31, no. 3。7, pp. 803-810, 2018。
[4] 王晓明, 王晓明, “基于深度卷积神经网络的语义分割”, 《IEEE 会议》, 2015,pp. 391 - 391。
[5] 朱伟, 姜清, 王玉生, 谢晓霞, “基于稀疏标记的深度多实例网络全乳房 x 线照片分类”, 医学图像计算与计算机辅助干预, 2017,pp. 603-611: Springer。
[6] 张志刚, 李志刚, 李志刚, “基于多实例的正交正交矩形问题求解”, 中国机械工程, 第 6 卷, 第 6 期。1-2, pp. 31-71, 1997。
[7] 刘杰, 王大伟, 付慧, 徐艳, 刘杰, “基于深度学习的早期黄斑变性检测”, 计算机视觉学报, 2018,pp. 625-640: Springer。
[8] 张晓明, 张晓明, 张晓明, “基于深度卷积神经网络的图像分类”, vol. 26, no. 8。6, pp. 84-90, 2017。
[9] 张晓明, 张晓明, 张晓明。基于深度卷积神经网络的图像识别 [j] .中文信息学报, 2014。
[10] 何凯, 张晓明, 孙军, “基于深度残差学习的图像识别方法”, 《IEEE 会议论文集》, 2016,pp. 774 -778。
[11] 陈晓明, 陈晓明, 陈晓明, “残差连接对神经网络学习的影响”, 中国计算机科学, 2017,vol. 31, no. 11。1。