

基于半监督学习的糖尿病视网膜病变自动评级和分割

居逸文

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2024年4月22日; 录用日期: 2024年5月23日; 发布日期: 2024年5月31日

摘要

糖尿病视网膜病变(DR)是导致永久性失明的主要原因之一, 尤其是在中老年人中。早期发现DR对于预防视力丧失至关重要。疾病严重程度分级可以看作是一个分类问题, 只需要图像级的标注, 而病变分割则需要更强的像素级标注。然而逐像素数据标注非常耗时, 需要专家。本文提出了一种协同学习的方法来联结优化DR评级和病灶检测。首先利用小部分逐像素数据标注数据训练一个多病变标签语义分割模型, 接着基于最初对大量图像级注释数据预测的病变图, 设计了一个疾病注意力评级模型。同时, 疾病注意力模型可以利用特定类的信息对病灶图进行细化, 以半监督的方式对分割模型进行微调。生成对抗性架构也被应用在训练过程中。接着在三个公共数据集上对本文的方法进行验证, 并与一些先进的模型进行对比, 取得较好的性能。

关键词

糖尿病视网膜病变, 评级, 分割, 生成对抗, 半监督, 疾病注意力

Automated Diabetic Retinopathy Grading and Segmentation Based on Semi-Supervised Learning

Yiwen Ju

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 22nd, 2024; accepted: May. 23rd, 2024; published: May. 31st, 2024

Abstract

Diabetic retinopathy (DR) is one of the leading causes of permanent blindness, especially among middle-aged and elderly individuals. Early detection of DR is crucial for preventing vision loss. The severity grading of the disease can be seen as a classification problem, requiring only image-level annotations, while lesion segmentation requires more detailed pixel-level annotations. However, pixel-wise data annotation is time-consuming and requires expertise. In this paper, we propose a collaborative learning approach to integrate the optimization of DR grading and lesion detection. Firstly, we train a multi-label semantic segmentation model using a small subset of pixel-wise annotated data, then we design a disease attention grading model based on the initial predictions of lesion maps from a large amount of image-level annotated data. Meanwhile, the disease attention model can refine lesion maps using specific class information to fine-tune the segmentation model in a semi-supervised manner. Generative adversarial architectures are also applied in the training process. We then validate our approach on three public datasets and compare it with some state-of-the-art models, achieving better performance.

Keywords

Diabetic Retinopathy, Automated Grading, Segmentation, Generative Adversarial, Semi-Supervised, Disease Attention

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

糖尿病视网膜病变(DR)是糖尿病引起的血管损伤最严重的并发症之一。如图 1 所示,眼科医生根据微动脉瘤、渗出物和出血等各种病变的存在来评估 DR。眼底图像涉及三个 DR 相关病变,箭头表示不同的 DR 病变。

DR 自动诊断已被广泛探索并应用于各种实用的计算机辅助医疗系统中。DR 逐像素的病变分割[1] [2] [3] [4]和严重程度评级[5] [6] [7]是 DR 自动诊断的两个主要的基本问题。

考虑到眼底图像逐像素的标签数据是昂贵的,需要眼科专家非常耗时的标注。所以对于需要大量标签的全监督学习的分割任务[8] [9] [10]来说,分割医学图像是不切实际的。而在医学图像分割中,精确性是至关重要的,因为错误的分割结果可能会导致误诊或错误的治疗决策。

在早期的研究中,半监督方法主要将传统机器学习方法应用于医学图像数据集中的无标签数据。例如,Bai 等人[11]提出了一种基于自训练的肝脏 MR 图像分割方法,通过自监督学习特征和预测解剖位置,在标签数据较少的情况下取得了和 U-net 网络相近的效果。Filipovych 等人[12]将半监督 SVM 应用于解决阿尔茨海默病中轻度认知障碍的分类问题。

近年来,医学图像领域的半监督学习方法大多基于深度学习,主要包括基于对抗性网络和基于一致性正则化的方法。Lecouat 等人[13]提出了一种基于图像块的半监督方法,用于克服糖尿病性视网膜病变分类中精细度不够的问题,该方法能够利用极少的标签数据取得很好的 AUC 值。Diaz-Pinto 等人[14]将生成对抗网络引入到青光眼图像数据集上,用于自动分类青光眼图像,该方法能够利用标记和未标记数

据合成图像，并且能够自动提取标签。Perone 等人[15]则通过将基于一致性的平均教师模型应用到 MRI 图像数据集中进行验证，实验证明在不同大小的数据集上都有着不错的提升。

近年来，GAN 被广泛应用到视网膜眼底图像的自动生成中。Costa 等人[16]首先使用对抗性自编码器

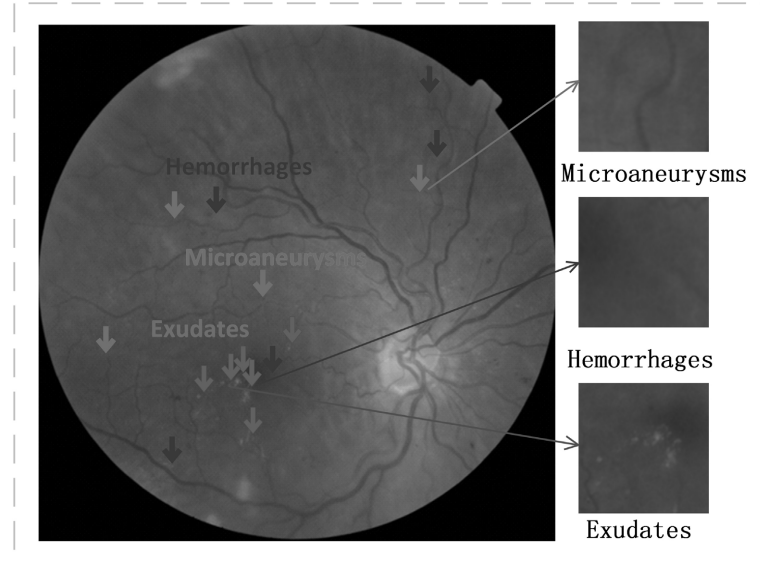


Figure 1. A fundus image with microaneurysms, exudates and hemorrhages

图 1. 眼底图像显示微动脉瘤，渗出和出血

进行视网膜眼底血管树生成，随后以生成的视网膜眼底血管树为输入，生成视网膜眼底图像。这种方法能够合成更多形态学不同的血管树，增加生成的视网膜眼底图像的多样性。然而，在合成视网膜眼底图像时没有考虑 DR 相关病变信息，导致生成的视网膜眼底图像中 DR 相关病变信息不明显。Zhao 等人[17]在生成器中引入风格迁移技术以增加生成的视网膜眼底图像的多样性。该方法能够以单一的血管树为输入生成多个视网膜眼底图像。然而，生成的视网膜眼底图像的 DR 病变等级设定与输入图像相同，导致生成的视网膜眼底图像 DR 病变等级标签不准确。Niu 等人[18]首先从参考图像中提取病理描述符，随后使用病理描述符指导生成器生成视网膜眼底图像。该方法可以操纵病变的位置和数量，但不能精确控制 DR 病变等级。Tavakkoli 等人[19]使用基于 CGAN 的两阶段生成器，以视网膜眼底图像作为输入生成荧光素血管造影图像。Lim 等人[20]以 StyleGAN [21]作为生成器、ResNet-50 [22]作为判别器，以潜在变量为输入生成视网膜眼底图像。他们首先用所有 DR 病变等级的视网膜眼底图像训练 GAN 模型，随后使用 DR 病变等级较高的视网膜眼底图像对模型进行微调，最后用训练好的判别器，对生成的视网膜眼底图像进行筛选。Yang 等人[23]以残差结构对 CycleGAN 进行改进，对不同眼底相机品牌的视网膜眼底图像进行域适应。

针对上述问题，本文设计了一种基于半监督的糖尿病视网膜病变自动评级和分割，主要贡献如下：

(1) 提出了一种用于逐像素分割的多病灶掩模生成器。设计了一个基于 Token-MLP 模块的 U 型网络和一个联合目标函数；

(2) 对于图像级别的注释数据，设计了一种基于 EfficientNet 的疾病注意力评级模型，该模型仅采用弱监督的类别特定信息即可自动预测病灶图；

(3) 病灶分割和疾病分级任务在端到端模型中进行优化。

2. 模型设计

像素级标注图像 X^p 和图像级标注图像 X^l ，本文方法的最终目标是协同优化病变分割模型 $G(\cdot)$ 和疾

病分级模型 $C(\cdot)$ ，二者共同提高彼此的精度。为了训练分割模型，目标是通过以下函数最小化预测的病变图与地面真实掩膜之间的差异：

$$\min_G \sum_{l=1}^L L_{Seg} \left(G(X^P), G(X^I), s_l^P, \tilde{s}_l^P \right) \quad (1)$$

这里， s_l^P 表示像素级标注图像的地面真实情况，而 \tilde{s}_l^P 则是由病变注意力分级模型学习得到的图像级标注图像的伪掩蔽。 L 是与特定疾病相关的病变种类的总数。疾病分级模型的优化函数定义为：

$$\min_{C.att} L_{Cls} \left(C(X^I) \cdot att \left(G(X^I) \right), y^I \right) \quad (2)$$

其中， $att(\cdot)$ 表示病变注意力模型， y^I 是图像级标注数据的疾病严重性分类标签。注意，在公式 1 中， \tilde{s}_l^P 等于 $att \left(G(X^I) \right)$ 。 L_{Seg} 和 L_{Cls} 的详细定义在第 4.2 节和第 4.3 节中分别解释。因此，要协同学习这两个任务，最重要的因素是如何设计和优化 $G(\cdot)$ 、 $C(\cdot)$ 和 $att(\cdot)$ 。

提出的网络架构概述如图 2 所示，它由两部分组成。对于第一部分，将少量的 X^P 作为输入，以全监督的方式训练一个多病变掩膜生成器。一旦它预训练完成，剩余的大规模 X^I 也会通过生成器。一个通过对抗性训练损失进行优化的鉴别器被设计用来区分这两种类型的数据。对于第二部分， X^I 及其最初预测的病变图被采用来学习一个病变注意力模型，该模型仅使用疾病分级标签。病变注意力分级模型提高了分类准确性。此外，生成的病变注意力图可以作为伪掩蔽，在半监督的方式下为大规模未标记数据改进病变掩膜生成器。

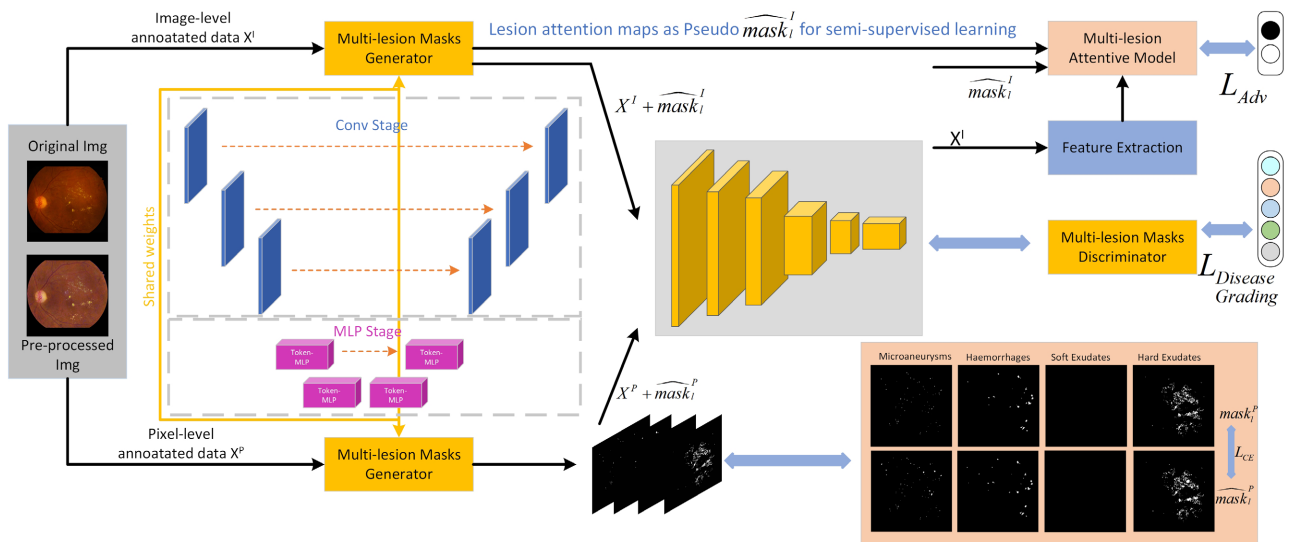


Figure 2. Overall network architecture diagram.

图 2. 总体网络结构图

2.1. 对抗性多病变掩膜生成器

训练一个语义分割模型往往需要大部分像素级别标注的数据。然而，在医学影像中标注的成本极高，不得不找到一种有效利用有限标注数据来训练模型的方法。本文采用了一种多病变掩膜生成器，来自于 U 型网络，并且嵌入了一种 MLP 的结构来完成任务。U 型网络最初用于电子显微镜堆栈中，神经结构的分割，其采用了编解码的结构，由全卷积网络构建。跳跃连接的作用是在神经网络的不同层之间建立直接连接，使得网络能够跨越较长的路径传递信息，有助于解决梯度消失的问题。这种设计可以在解码的

过程中更好的保留边缘和纹理细节并且加快收敛速度。

首先在 U 型网络中，嵌入了标记化的 MLP 模块并修改成一个多病变掩膜生成器。如图 3(a)所示，在标记化的 MLP 块中，首先移动特征并将它们投影到 token 中。为了标记化，首先使用 3×3 的卷积核，并将通道数更改为 E，其中 E 是嵌入维度(token 数量)，是一个超参数。然后，将这些令牌传递到一个移动的 MLP(沿宽度)，其中 MLP 的隐藏维度是一个超参数 H。接下来，特征通过一个深度可分离卷积层(DWConv)传递。在这个块中使用 DWConv 有两个原因：1) 它有助于编码 MLP 特征的位置信息，在 MLP 块中的卷积层足以编码位置信息，并且实际上表现比标准的位置编码技术更好。像 ViT 中的位置编码技术需要在测试和训练分辨率不同时进行插值，通常会导致性能降低。2) DWConv 使用的参数数量较少，因此提高了效率。然后，使用一个 GELU 激活层。本文使用 GELU 而不是 RELU，因为它是一种更平滑的替代方案，并且被发现性能更好。然后，将特征通过另一个移动的 MLP (沿高度)传递如图 3(b)所示，将维度从 H 转换为 O。在这里使用一个残差连接，并将原始的 token 作为残差添加。然后，应用一个层归一化(LN)并将输出特征传递到下一个块。因为沿着 token 进行归一化更有意义。

图 3 中左边的部分显示了分割模型的示意图。总体而言，编码器和解码器共包括九个特征映射模块。除了编码器的前两个模块和解码器的后两个模块使用常规卷积操作外，其余模块都替换为标记化的 MLP 模块。每个元组由两个可分离卷积组成，后跟批量归一化、ReLU 激活、最大池化和 1×1 卷积的快捷方式。空间卷积核大小为 3×3 ，并设置相同的填充。在解码器部分，每个元组之前都使用了上采样和跳跃连接。最后，添加了 L 个带有 Sigmoid 激活的卷积层，以生成 L 个不同的病变掩膜。

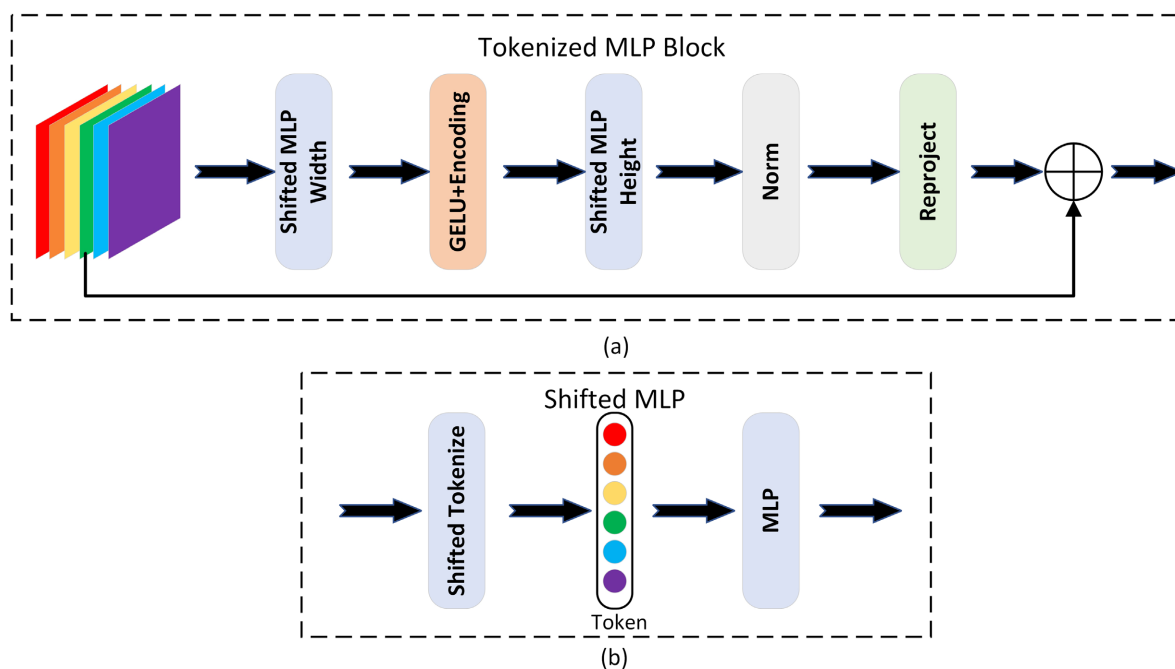


Figure 3. Architecture of token-MPL model

图 3. 标记化的 MLP 结构图

为了优化病变掩膜生成器，同时使用像素级标注数据和图像级标注数据。对于像素级标注的病变掩膜，使用二元交叉熵损失函数 L_{CE} 最小化预测与地面真实之间的距离。基于第 2.2 节介绍的病变注意力模型，还获得了图像级标注数据的伪掩膜地面真实，以优化 L_{CE} 。此外，为了利用没有像素级标注的数据生成更好的病变掩膜，添加了一个多病变鉴别器 D，通过生成对抗网络(GAN)架构进行训练。传统的 GAN

包括一个生成网络和一个判别网络，它们进行竞争性的极小 - 极大博弈。通常会使用来自均匀分布或高斯分布的潜在随机向量 z 作为生成器的输入来合成样本。然后，鉴别器的目标是区分真实数据 x 和生成的样本。其本质目标是使 $p_z(z)$ 收敛到目标真实数据分布 $p_{\text{data}}(x)$ 。在本文中，并不是从随机噪声中生成样本，而是将由像素级标注数据的生成器预测的病变图作为真实数据分支，将由图像级标注数据生成的病变图作为假样本分支。用于优化病变分割任务的总损失可以定义为：

$$\begin{aligned} L_{\text{Seg}} &= L_{\text{Adv}} + \lambda L_{\text{CE}} \\ &= E \left[\log \left(D \left(X^P, G \left(X^P \right) \right) \right) \right] \\ &\quad + E \left[\log \left(1 - D \left(X^I, G \left(X^I \right) \right) \right) \right] \\ &\quad + \lambda E \left[-s \cdot \log G \left(X^{(P,I)} \right) - (1-s) \cdot \log \left(1 - G \left(X^{(P,I)} \right) \right) \right] \end{aligned} \quad (3)$$

其中， s 是 s_i^P 和 s_i^I 的简明表示，分别对应像素级和图像级标注数据的地面真实情况。 λ 是两个目标函数的平衡权重。预测的多病变掩模与输入图像连接起来，然后作为鉴别器 D 的输入，该鉴别器具有五个卷积映射元组。每个元组由两个卷积层组成，卷积核大小为 3，以及一个步长为 2 的最大池化层，用于逐渐编码增大的感受野的上下文信息。对于每个元组，采用了 ReLU 激活和批量归一化。在 D 的末尾使用了全局平均池化，然后是一个密集连接和 Sigmoid 激活函数，如果预测的病变图是真实标签或伪标签则输出。

2.2. 疾病注意病变注意力疾病分级

为了评估疾病的严重程度，专家通常通过观察疾病特征的详细病变迹象进行诊断。采用经典的深度分类模型可以实现基本性能，但准确性有限。视觉注意力模型以类似人类的方式处理识别任务，自动提取特定于任务的区域，并忽略不相关的信息以提高性能。然而，大多数传统的注意力模型都是为一般物体图像提出的，并且只需要预测粗略的注意力图。注意力机制通常是使用高级特征设计的。对于医学图像，在病变区域非常小，且预期以像素级方式关注的情况下，本文的模型还采用具有高分辨率的低级特征图来指导注意力模型的学习。此外，对于那些仅具有图像级疾病分级注释的图像，本文的病变注意力模型可以生成像素级注意力图，然后将其用作病变分割模型中的半监督学习的伪标签。

病变注意力疾病分级模型如图 4 所示，是由一个主分支用于提取和分类输入疾病图像的特征，并由 L 个分支用于学习 L 个病变的注意力模型。不使用最初由分割模型预测的病变掩模直接关注分类模型，因为像素级标注的医学图像通常非常少，因此最初预测的掩模太弱无法使用。此外，可以利用图像级分级标签来提供区分性的本地化信息，以改进病变注意力图。

在本文的方法中，疾病分级模型 $C(\cdot)$ 和病变注意力模型 $\text{att}(\cdot)$ 紧密集成。首先采用一个带有基本卷积神经网络的疾病分类模型，仅使用输入图像进行分级学习。一旦它被预训练，分别表示低级和高级特征表示的 f^{low} 和 f^{high} 可以作为像素级和类别级别的指导，用于学习注意力模型。此外，还将最初预测的病变图，表示为 $m_{i=1}^L$ ，编码为注意力模型的输入。整体表达式由以下方程定义：

$$a_{i=1}^L = \text{att} \left(f^{\text{low}}, f^{\text{high}}, m_{i=1}^L \right) \quad (4)$$

其中，输出 $a_{i=1}^L$ 是注意力图，对于不同病变区域给出高响应，这些区域特征化了该疾病。提出的注意力机制由两个步骤组成。第一步是利用像素级病变特征，通过融合来自输入图像和最初预测的病变掩模的编码低级嵌入。对于第 1 个病变，可以通过以下方程获得一个关注特征的中间状态：

$$f_i^{(\text{low_att})} = \text{ReLU} \left(W_i^{\text{low}} \text{concat} \left(m_i, f_i^{\text{low}} \right) + b_i^{\text{low}} \right) \quad (5)$$

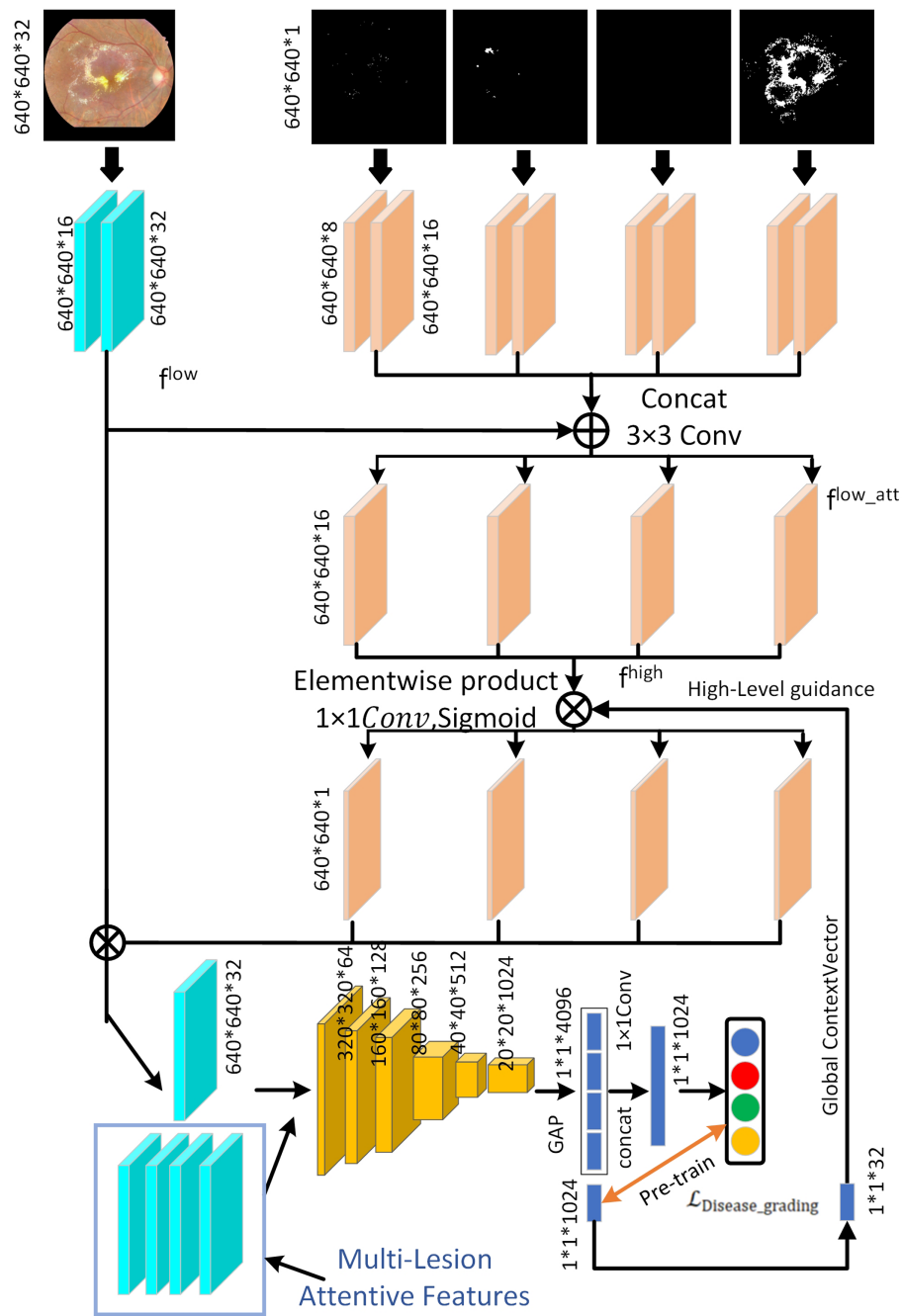


Figure 4. Disease-specific lesion grading module
图 4. 疾病注意病变评级模块

其中, $concat(\cdot)$ 表示沿通道方向的拼接。对于第二步,使用全局上下文向量与低级注意力特征进行相关,并进一步生成病变图,如下所示:

$$a_i = Sigmoid\left(W_i^{high} \left[f_i^{(low-att)} \odot f^{high} \right] + b_i^{high} \right) \quad (6)$$

其中, \odot 表示逐元素乘法。全局上下文向量 f^{high} 具有与 $f_i^{low-att}$ 相同的通道维度,通过对基本预训练分类模型的顶层特征进行 1×1 卷积计算得到。这种高级别的指导包含丰富的类别信息,以加权低级特征并细

化精确的病变细节。注意, W_l^{low} 、 W_l^{high} 和偏置项是第 1 个病变的可学习参数。

基于 L 个病变注意力图, 分别对低级图像特 f^{low} 进行逐元素乘法, 并使用这些注意力特征微调预训练的疾病分类模型。所有病变注意力特征共享与分级模型相同的权重, 并将输出特征向量拼接以学习最终表示。疾病分级的目标函数 L_{cls} 采用了由于不平衡数据问题而采用的焦点损失。与此同时, 精细化的多病变注意力图被用作伪标签, 以半监督的方式与分割模型共同训练。

3. 实验结果与分析

3.1. 实验数据集

为了评估本文所提出的算法的性能, 本文主要用到的数据集是 IDRiD 数据集, EyePACS 数据集和 Messidor1 数据集。

IDRiD 数据集[24]是唯一提供像素级多病变注释的 DR 数据集。它包含 81 张具有 DR 症状的彩色眼底图像, 并分为 54 张用于训练和 27 张用于测试。医学专家使用二进制掩模对包括微小动脉瘤、出血、硬性渗出和软性渗出在内的病变进行了注释。IDRiD 还有一个图像级别注释集, 包含 413 张训练图像和 103 张测试图像, 这些图像只有严重程度分级标签。使用病变分割集以完全监督的方式训练多病变掩模生成器。然后, 使用分级集来学习用于分类和半监督分割的病变注意模型。

EyePACS 数据集[25]包含 35,126 张训练图像和 53,576 张测试图像。分级协议与 IDRiD 数据集相同, 有五个 DR 类别。然而, 从该数据集收集的图像是由不同类型的相机在各种光照条件下拍摄的, 并且注释质量较差。由于该数据集只有图像级别的分级标签, 主要用它来训练病变注意的疾病分级模型。

Messidor1 数据集[26]包含 1200 张眼底图像, 但其分级标准与前两个数据集不同, 只有 4 个级别。等级 0 和 1 被标记为不需转诊, 而等级 2 和 3 被视为需要转诊。除等级 0 外的所有等级都表示 DR 的异常情况。遵循中使用的评估协议, 仅采用此数据集来测试在 EyePACS 上训练的模型。

3.2. 评价指标和训练策略

为了定量评估病变分割任务的性能, 计算了受试者工作特征曲线(Receiver Operating Characteristic Curve, ROC Curve)和精确召回曲线(Precision-Recall, PR Curve)曲线下的曲线下面积(AUC)值。此外, 为了评估 DR 分级模型的精度, 除了正常的分类准确率外, 还引入了一个二次加权 kappa 指标。

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \quad (7)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (10)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (11)$$

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (12)$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

ROC 曲线的横坐标为 FPR (False Positive Rate), 如式 12 和 13 所示, 纵坐标为 TPR (True Positive Rate), 如式所示 PR 曲线的横坐标为召回率, 纵坐标为精确率。其中 TP (True positive) 表示真阳性即标签为 1 预测也是 1, FN (False negative) 表示假阴性即标签为 1 预测为 0, FP (False positive) 表示假阳性即标签是 0 预测为 1, TN (True negative) 表示真阴性即标签为 0 预测也为 0。

模型的训练方案分为两个阶段。在第一步中, 使用像素级标注数据通过 L_{CE} 预训练多病变分割模型, 并使用图像级标注数据通过 L_{cls} 预训练基本疾病严重程度分类模型。两者都是以全监督的方式训练的。采用了 ADAM 优化器, 学习率为 0.0002, 动量为 0.5。对于预训练分割模型, 批量大小设置为 32, 训练 60 个 epochs; 而对于分级模型, 批量大小为 128, 训练 30 个 epochs。一旦预训练完成, 可以获得最初预测的病变掩模以及输入图像的低级和高级特征表示, 从而可以同时训练病变注意力模型进行半监督分割, 并进一步提高分级性能。在这个阶段, 添加了 L_{Adv} 进行半监督学习, 以及病变注意力模块进行疾病分级。整个模型以端到端的方式进行微调。方程 3 中的 λ 设置为 10, 这是得到最佳性能的值。微调时, 批量大小设置为 16, 训练 50 个 epochs。所有实验在一台 NVIDIA Geforce GTX 4090 Ti GPU 上进行训练, 显存为 12 G。

3.3. 关于预测精度的对比实验与结果分析

在评估定量病变分割精度和 DR 分级准确性之前, 首先在 IDRID 数据集上定性展示了病变注意模型在半监督分割上的有效性。IDRID 数据集具有真实的标签。图 5 比较了仅采用有限像素级注释数据的预训练模型和使用大规模图像级别注释数

据进行半监督训练的最终模型对四种不同病变的分割结果。对于预训练模型, 失败的情况通常是病变模式的漏检(假阴性)。在一些小区域也出现了误报(假阳性)。通过使用图像级别注释数据进行半监督分割, 所有病变的结果都明显得到了改善。病变分割对提高 DR 分级的有效性由第 3.3 节的消融研究进行评估。

首先粗略地标记了一些区域, 以比较仅使用像素级别病变注释进行预训练的初始模型和使用大规模图像级别注释数据进行半监督训练的模型。

为了评估病变分割对于 DR 分级的有效性以及注意模型对半监督学习的改进, 将三种基线模型与提出的最终模型进行比较。Ori: 首先研究病变分割模型是否可以提高 DR 分级的准确性。在这个基线中, 不使用病变注意特征, 而是直接在预处理的眼底图像上训练分级模型。Lesion (预训练): 作为对比, 本文还测试了一个仅在有限的像素级别病变注释数据上预训练的基线模型。最初生成的多病变掩模在图像特征图上加权, 用于训练分级模型, 而没有使用病变注意模型。Lesion (半监督): 本文还探讨了通过病变注意模型对半监督学习的改进。在这个基线中, 只采用交叉熵损失来学习病变分割模型。Lesion (半监督 + Adv): 最终的方法将对抗性训练架构集成到病变分割目标函数中。

表 1 显示了不同方法的分类准确度和 kappa 分数。在 IDRID 数据集上, 与不使用病变掩模信息的基本分类模型相比, 仅在像素级别注释数据上预训练的初始分割模型可以将分级的准确度提高 2.67%, kappa 分数提高 3.8%。通过使用图像级别注释数据进行半监督学习, 可以实现更显著的改进。特别是, kappa 分数的大幅增加 5.86% 证明了提出的病变注意模型可以有效地精细化病变图, 并因此改善分级结果。

此外, 对抗性训练架构也可以使最终结果受益, 分类准确度进一步提高 1.18%, kappa 分数提高 1.55%。由于 EyePACS 数据集只有图像级别的注释, 采用在 IDRID 数据集上预训练的完全监督模型。与在 IDRID 数据集上获得的性能结果相比, 本文的模型的每个组件对于分级任务都具有积极的贡献, 相对于其他方法。

为进一步评估模型, 还在 Messidor1 数据集上进行了实验。遵循[27]中的评估方法和协议, 比较了

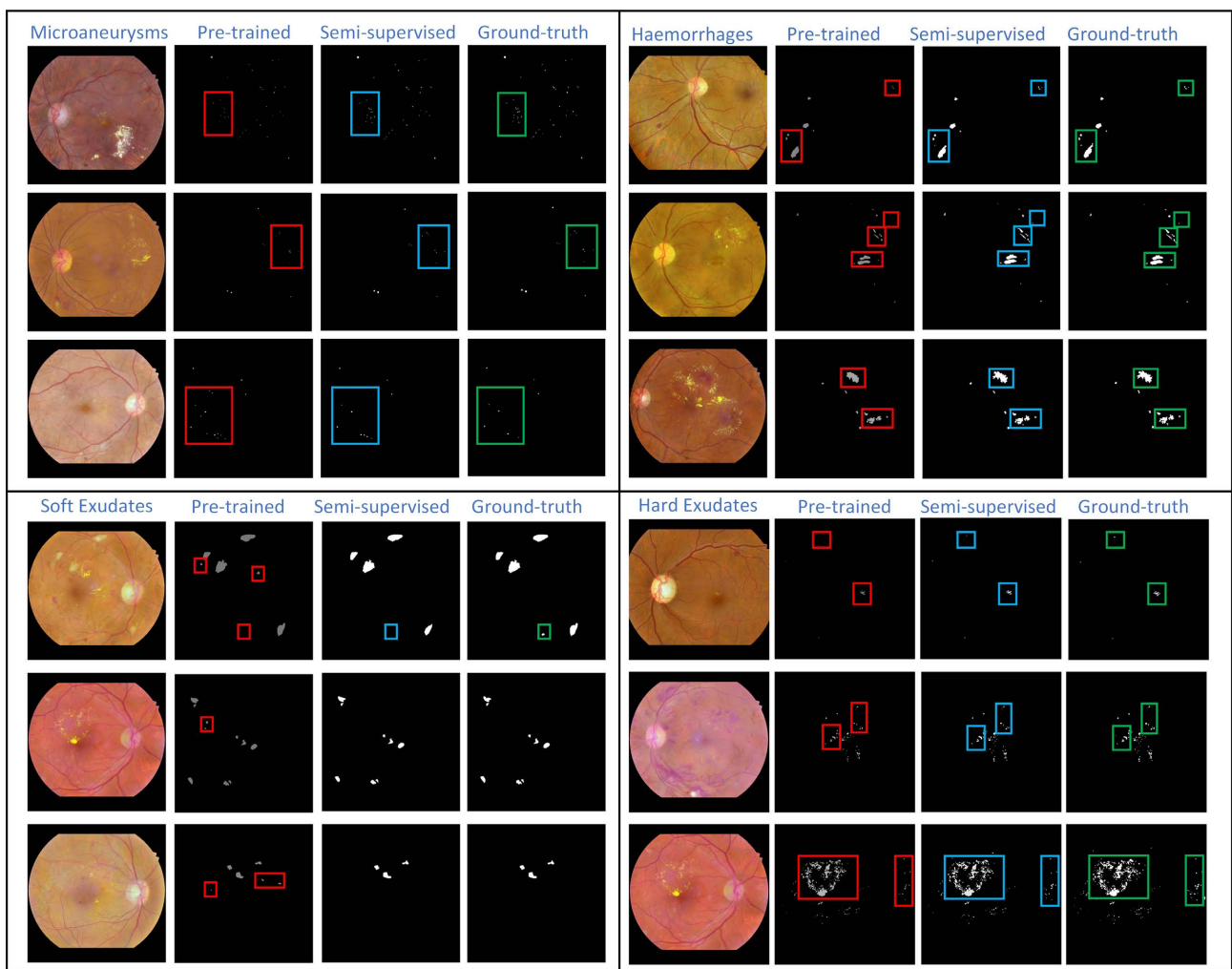


Figure 5. Qualitative multi-lesion segmentation result

图 5. 定性多病变分割结果

Table 1. Ablation experiments on the IDRID dataset and EyePACS

表 1. 在 IDRID 数据集和 EyePACS 的消融实验

数据集方法	IDRID		EyePACS	
	Acc	Kappa	Acc	Kappa
Ori	83.2	78.3	85.4	83.4
Lesion (Pretrained)	87.3	83.0	86.0	84.4
Lesion (Semi)	90.1	88.9	87.9	86.1
Lesion (Semi + Adv)	92.2	90.8	89.1	87.2

ROC 曲线下面积(AUC)和正常和转诊分类的准确性, 如表 2 所示。对于两种实验设置, 具有病变注意模型、半监督分割和对抗性训练架构的提出方法达到了最佳性能。由于 Messidor1 数据集的图像质量接近于 IDRID, 即使是基于预训练病变的模型, 也可以与基本的整体分类模型相比获得显著的增益。

除了疾病分级性能的提升外, 还调查了基于病变伪标签的半监督分割的有效性。在具有真实标签的 IDRID 数据集上评估分割性能。评估的四种不同病变, 包括微动脉瘤、出血、硬渗出和软渗出, 是糖尿

病视网膜病变的主要表现。通过 ROC 曲线、PR 曲线和相应的 AUC 值进行评估。探索了最终模型的每个提出的组件与三个基线：使用正常卷积元组的预训练分割模型、基于标记化的 MLP 模块的模型以及不带对抗性训练架构的半监督学习组件。

Table 2. Ablation experiments on the Messidor1 dataset

表 2. 在 Messidor1 数据集上的消融实验

数据集方法	Referral		Normal	
	AUC	Acc	AUC	Acc
Ori	93.4	90.2	88.9	87.8
Lesion (Pretrained)	95.3	90.0	91.9	90.1
Lesion (Semi)	97.1	93.0	93.7	91.8
Lesion (Semi + Adv)	97.6	93.9	94.3	92.2

在表 3 中列出了详细的 AUC 值。表格中 CE1 和 CE2 分别表示分割模型仅采用普通的卷积层和采用了标记化的 MLP 层。基于标记化的 MLP 模块的病变分割模型在四种不同病变上一贯优于基于普通卷积的版本。ROC 曲线和 PR 曲线的 AUC 平均分别提高了 1.02% 和 1.92%，证明可移动的 MLP 模块及深度可分离卷积确实可以改善分割结果。通过利用更多的图像级别注释数据，生成伪标签进行半监督分割，观察到了明显的改善，PR 曲线的 AUC 平均提高了 2.16%。此外，用于半监督学习的对抗性训练架构可以稍微进一步提高分割精度。

Table 3. Performance comparison of multi-lesion segmentation on the IDRID dataset

表 3. 多病变分割在 IDRID 数据集上的性能比较

病灶方法	硬渗出液		软渗出液		微动脉瘤		出血	
	ROC	PR	ROC	PR	ROC	PR	ROC	PR
CE1 (Conv)	96.2	82.6	94.4	68.2	95.1	46.3	94.4	64.6
CE2 (Token-MLP)	96.8	84.6	95.4	71.6	96.5	47.3	95.4	65.8
CE2 + semi	98.9	87.6	97.1	73.4	97.8	48.9	97.0	68.1
CE2 + semi + Adv	99.4	88.7	99.4	74.1	98.3	49.6	97.8	69.4

如表 4 显示了不同病变的 PR 曲线的 AUC 得分的前三名，以及来自其他视觉任务的两种半监督分割方法 AdvSeg 和 ASDNet 的表现。尽管本文的方法在微动脉瘤检测方面的性能略低于当前的顶级模型，但其他三种病变都取得了适度的改善。软渗出病变取得了特别大的提升，提高了 4.12%。此外，的模型在 PR 的 AUC 上的平均增加分别比 AdvSeg 和 ASDNet 高出了 6.89% 和 5.22%。

为了使本文的方法更加令人信服，将其与最先进的 DR 分级模型进行比较。组合了 CKML 和带有额外核心的 VNX 旨在采用多个滤波器尺寸来学习细粒度的判别特征。

Zoom-in-Net 提出了一个门控注意力模型，并结合了三个子网络来对整体图像、高分辨率裁剪和门控区域进行分类。AFN 具有统一病变检测和 DR 分级的类似思想。然而，所使用的注意模型仅是类驱动的，无法学习精确的语义病变图。此外，还邀请了专家对 Messidor1 数据集进行分级。

表 5 比较了不同方法的结果。在 Messidor1 数据集的可引用/不可引用和正常/异常设置中，与其他方法相比，本文的方法可以获得 ROC 曲线下面积(AUC)和分级准确性的最高得分。值得一提的是，本文的

方法在转诊和正常设置的 AUC 上分别比专家高出 3.6% 和 2.1%。

Table 4. Comparative experiments on segmentation performed by our method on the IDRID dataset

表 4. 本文的方法在 IDRID 数据集上分割的对比实验

病灶方法	硬渗出液		软渗出液		微动脉瘤		出血	
	PR	ROC	PR	ROC	PR	ROC	PR	ROC
VRT(1 st)	0.7127	-	0.6995	-	0.4951	-	0.6804	-
PATech (2 nd)	0.8850	-	-	-	0.4740	-	0.6490	-
iFLYTEK (3 rd)	0.8741	-	0.6588	-	0.5017	-	0.5580	-
DRUNet [24]	-	-	-	-	-	0.9820	-	-
SESV [25]	-	-	-	-	0.5099	-	-	-
L-seg [26]	0.7945	-	0.7113	-	0.4267	-	0.6374	-
AdvSeg [4]	0.8032	0.9456	0.6756	0.9318	0.4782	0.9612	0.5923	0.9256
ASDNet [27]	0.8095	0.9502	0.6924	0.9489	0.4782	0.9692	0.6285	0.9324
Ours	0.8876	0.9941	0.7424	0.9943	0.4983	0.9832	0.6943	0.9784

Table 5. Comparative experiments on grading performed by our method on the Messidor1 dataset

表 5. 在 Messidor1 数据集上评级的对比实验

数据集分类方法	Messidor1			
	Referral		Normal	
	AUC	Acc	AUC	Acc
VNXK [28]	88.7	89.3	87.0	87.1
CKML [28]	89.1	89.7	86.2	85.8
Expert [29]	94.0	-	92.2	-
Zoom-in-Net [30]	95.7	91.1	92.1	90.5
AFN [23]	96.8	-	93.5	-
Ours	97.8	94.3	94.5	92.7

4. 结论

在本文中，提出了一种协同学习的方法，用于医学影像的半监督病变分割和疾病分级。病变标签被用来指导分类模型并提高分级的准确性，同时利用特定类别的标签的病变注意力模型也使得分割结果受益。大量实验表明，本文方法在 IDRID 数据集上的分割 PR 和 ROC 指标，优于目前大多数先进的方法。以及在 Messidor1 数据集上的分类效果领先于很多传统的 CNN 网络模型和半监督模型。在消融模块中，在 IDRID 数据集以及 EyePACS 数据集，我们的每个模块都体现了优秀的性能。

参考文献

- [1] Tan, J.H., Fujita, H., Sivaprasad, S., *et al.* (2017) Automated Segmentation of Exudates, Haemorrhages, Microaneurysms Using Single Convolutional Neural Network. *Information Sciences*, **420**, 66-76. <https://doi.org/10.1016/j.ins.2017.08.050>
- [2] Guo, S., Li, T., Wang, K., *et al.* (2019) A Lightweight Neural Network for Hard Exudate Segmentation of Fundus Im-

- age. *Proceedings of the 28th International Conference on Artificial Neural Networks*, Munich, 17-19 September 2019, 189-199. https://doi.org/10.1007/978-3-030-30508-6_16
- [3] 傅迎华, 李震, 张雨鹏. 一种融合残差模块的 U-Net 对眼底图像渗出液的检测[J]. 小型微型计算机系统, 2021, 42(7): 1479-1484.
- [4] Guo, S., Li, T., Kang, H., *et al.* (2019) L-Seg: An End-to-End Unified Framework for Multi-Lesion Segmentation of Fundus Images. *Neurocomputing*, **349**, 52-63. <https://doi.org/10.1016/j.neucom.2019.04.019>
- [5] Dai, L., Sheng, B., Wu, Q., Li, H., Hou, X., Jia, W. and Fang, R. (2017) Retinal Microaneurysm Detection Using Clinical Report Guided Multi-Sieving CNN. *Proceedings of the 20th International Conference on Medical Image Computing and Computer Assisted Intervention-MICCAI 2017*, Quebec City, 11-13 September 2017, 525-532. https://doi.org/10.1007/978-3-319-66179-7_60
- [6] Yang, Y., Li, T., Li, W., Wu, H., Fan, W. and Zhang, W. (2017) Lesion Detection and Grading of Diabetic Retinopathy via Two-Stages Deep Convolutional Neural Networks. *Proceedings of the 20th International Conference on Medical Image Computing and Computer Assisted Intervention-MICCAI 2017*, Quebec City, 11-13 September 2017, 533-540. https://doi.org/10.1007/978-3-319-66179-7_61
- [7] Zhou, K., Gu, Z., Liu, W., Luo, W., Cheng, J., Gao, S. and Liu, J. (2018) Multicell Multi-Task Convolutional Neural Networks for Diabetic Retinopathy Grading. *Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, 18-21 July 2018, 2724-2727. <https://doi.org/10.1109/EMBC.2018.8512828>
- [8] Li, Q., Arnab, A. and Torr, P.H.S. (2018) Weakly- and Semi-Supervised Panoptic Segmentation. *Proceedings of the 15th European Conference on Computer Vision-ECCV 2018*, Munich, 8-14 September 2018, 106-124. https://doi.org/10.1007/978-3-030-01267-0_7
- [9] Fan, R., Hou, Q., Cheng, M.-M., Yu, G., Martin, R.R. and Hu, S.-M. (2018) Associating Inter-Image Salient Instances for Weakly Supervised Semantic Segmentation. *Proceedings of the 15th European Conference on Computer Vision-ECCV 2018*, Munich, 8-14 September 2018, 371-388. https://doi.org/10.1007/978-3-030-01240-3_23
- [10] Wang, X., You, S., Li, X. and Ma, H. (2018) Weakly-Supervised Semantic Segmentation by Iteratively Mining Common Object Features. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 1354-1362. <https://doi.org/10.1109/CVPR.2018.00147>
- [11] Bai, W., Chen, C., Tarroni, G., *et al.* (2019) Self-Supervised Learning for Cardiac MR Image Segmentation by Anatomical Position Prediction. *Proceedings of the 22nd International Conference on Medical Image Computing and Computer Assisted Intervention*, Shenzhen, 13-17 October 2019, 541-549. https://doi.org/10.1007/978-3-030-32245-8_60
- [12] Filipovych, R., Davatzikos, C. and Alzheimer's Disease Neuroimaging Initiative (2011) Semi-Supervised Pattern Classification of Medical Images: Application to Mild Cognitive Impairment (MCI). *NeuroImage*, **55**, 1109-1119. <https://doi.org/10.1016/j.neuroimage.2010.12.066>
- [13] Lecouat, B., Chang, K., Foo, C.S., *et al.* (2018) Semi-Supervised Deep Learning for Abnormality Classification in Retinal Images. arXiv Preprint arXiv:1812.07832.
- [14] Diaz-Pinto, A., Colomer, A., Naranjo, V., *et al.* (2019) Retinal Image Synthesis and Semi-Supervised Learning for Glaucoma Assessment. *IEEE Transactions on Medical Imaging*, **38**, 2211-2218. <https://doi.org/10.1109/TMI.2019.2903434>
- [15] Perone, C.S. and Cohen-Adad, J. (2018) Deep Semi-Supervised Segmentation with Weight-Averaged Consistency Targets. *Proceedings of the 4th International Workshop on Deep Learning in Medical Image Analysis and the 8th International Workshop on Multimodal Learning for Clinical Decision Support*, Granada, 20 September 2018, 12-19. https://doi.org/10.1007/978-3-030-00889-5_2
- [16] Costa, P., Galdran, A., Meyer, M.I., *et al.* (2017) End-to-End Adversarial Retinal Image Synthesis. *IEEE Transactions on Medical Imaging*, **37**, 781-791. <https://doi.org/10.1109/TMI.2017.2759102>
- [17] Zhao, H., Li, H., Maurer-Stroh, S., *et al.* (2018) Synthesizing Retinal and Neuronal Images with Generative Adversarial Nets. *Medical Image Analysis*, **49**, 14-26. <https://doi.org/10.1016/j.media.2018.07.001>
- [18] Niu, Y., Gu, L., Lu, F., Lv, F., Wang, Z., *et al.* (2019) Pathological Evidence Exploration in Deep Retinal Image Diagnosis. *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*, Honolulu, 27 January-1 February 2019, 1093-1101. <https://doi.org/10.1609/aaai.v33i01.33011093>
- [19] Tavakkoli, A., Kamran, S.A., Hossain, K.F., *et al.* (2020) A Novel Deep Learning Conditional Generative Adversarial Network for Producing Angiography Images from Retinal Fundus Photographs. *Scientific Reports*, **10**, Article No. 21580. <https://doi.org/10.1038/s41598-020-78696-2>
- [20] Lim, G., Thombre, P., Lee, M.L., *et al.* (2020) Generative Data Augmentation for Diabetic Retinopathy Classification. *Proceedings of the 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*, Baltimore,

- 9-11 November 2020, 1096-1103. <https://doi.org/10.1109/ICTAI50040.2020.00167>
- [21] Karras, T., Laine, S. and Aila, T. (2019) A Style-Based Generator Architecture for Generative Adversarial Networks. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 4396-4405. <https://doi.org/10.1109/CVPR.2019.00453>
- [22] He, K., Zhang, X., Ren, S., *et al.* (2016) Deep Residual Learning for Image Recognition. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [23] Yang, D., Yang, Y., Huang, T., *et al.* (2020) Residual-CycleGAN Based Camera Adaptation for Robust Diabetic Retinopathy Screening. *Proceedings of the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention*, Lima, 4-8 October 2020, 464-474. https://doi.org/10.1007/978-3-030-59713-9_45
- [24] Lin, Z., Guo, R., Wang, Y., *et al.* (2018) A Framework for Identifying Diabetic Retinopathy Based on Anti-Noise Detection and Attention-Based Fusion. *Proceedings of the 21st International Conference on Medical Image Computing and Computer Assisted Intervention-MICCAI 2018*, Granada, 16-20 September 2018, 74-82. https://doi.org/10.1007/978-3-030-00934-2_9
- [25] Kou, C., Li, W., Liang, W., *et al.* (2019) Microaneurysms Segmentation with a U-Net Based on Recurrent Residual Convolutional Neural Network. *Journal of Medical Imaging*, **6**, Article 025008. <https://doi.org/10.1117/1.JMI.6.2.025008>
- [26] Xie, Y., Zhang, J., Lu, H., *et al.* (2020) SESV: Accurate Medical Image Segmentation by Predicting and Correcting Errors. *IEEE Transactions on Medical Imaging*, **40**, 286-296. <https://doi.org/10.1109/TMI.2020.3025308>
- [27] Hung, W.C., Tsai, Y.H., Liou, Y.T., *et al.* (2018) Adversarial Learning for Semi-Supervised Semantic Segmentation. arXiv Preprint arXiv:1802.07934
- [28] Vo, H.H. and Verma, A. (2016) New Deep Neural Nets for Fine-Grained Diabetic Retinopathy Recognition on Hybrid Color Space. *Proceedings of the 2016 IEEE International Symposium on Multimedia (ISM)*, San Jose, 11-13 December 2016, 209-215. <https://doi.org/10.1109/ISM.2016.0049>
- [29] Sánchez, C.I., Niemeijer, M., Dumitrescu, A.V., *et al.* (2011) Evaluation of a Computer-Aided Diagnosis System for Diabetic Retinopathy Screening on Public Data. *Investigative Ophthalmology & Visual Science*, **52**, 4866-4871. <https://doi.org/10.1167/iovs.10-6633>
- [30] Wang, Z., Yin, Y., Shi, J., Fang, W., Li, H. and Wang, X. (2017) Zoom-in-Net: Deep Mining Lesions for Diabetic Retinopathy Detection. *Proceedings of the 20th International Conference on Medical Image Computing and Computer Assisted Intervention-MICCAI 2017*, Quebec City, 11-13 September 2017, 267-275. https://doi.org/10.1007/978-3-319-66179-7_31