

糖尿病性视网膜病变的语义分割模型广泛训练方案

2024 年 4 月 28 日

摘要

糖尿病性视网膜病变是全球范围内造成视力障碍和法定失明的主要原因之一。随着糖尿病患者数量的不断增加，快速、准确地进行病变分析和诊断的需求变得日益迫切。借助深度学习方法，可以在眼底图像中准确地标记出病变区域，从而辅助医生进行更精准的诊断。针对这一问题，本团队以 DRAC2022 赛事提供的数据集为基础，构建了一套有效的模型，能够精准分割视网膜内微血管异常、非灌注区和新生血管等三类病变区域，尝试了多种基于 Unet++ 的模型框架，并引入了注意力机制等优化手段。经过本地测试和提交测试，这套模型系统的有效性得到了验证，并且在 DRAC2022：Task 1 Segmentation 中排名第 10 名（10/34）。Github 代码位于：<https://github.com/WinterSun0729/DRAC2022-AMDYES>。

关键词：糖尿病性视网膜病变，DRAC2022，深度学习，Unet++

Abstract

Diabetic retinopathy is one of the leading causes of visual impairment and legal blindness globally. With the increasing number of diabetic patients, the demand for rapid and accurate analysis and diagnosis of lesions is becoming increasingly urgent. Leveraging deep learning methods, we can accurately label lesion areas in fundus images to assist physicians in more precise diagnosis. To address this issue, our team has built an effective model based on the dataset provided by the DRAC2022 competition, which can accurately segment three types of lesion areas in the retina: microaneurysms, non-perfusion areas, and neovascularization. We have experimented with various Unet++-based model frameworks and introduced optimization techniques such as attention mechanisms. After local testing and submission testing, the effectiveness of this model system has been validated, ranking 10th out of 34 in DRAC2022: Task 1 Segmentation. Our code is publicly available at <https://github.com/WinterSun0729/DRAC2022-AMDYES>.

Keywords: Diabetic retinopathy, DRAC2022, Deep learning, Unet++

1 简介与意义/Introduction

1.1 项目意义和依据/Significance

眼底疾病，尤其是糖尿病性视网膜病变，是全球范围内导致视力障碍和法定失明的主要原因之一，其中 78% 的人有 15 年或更长糖尿病病史。随着近视患者数量的增加，对于快速、准确地进行糖尿病性视网膜病变分析和诊断的需求日益迫切。本项目旨在通过引入深度学习技术，特别是语义分割网络（如 nnU-Net⁽¹⁾、Unet++⁽²⁾ 等），实现对糖尿病性视网膜病变的自动化分割，为医生提供更迅速、准确的临床辅助工具，提高糖尿病性视网膜病变的早期诊断准确性。

糖尿病性视网膜病变的自动化分割对于眼科医生的临床工作具有重要的意义。通过借助深度学习方法，可以在眼底图像中准确地标记出病变区域，辅助医生进行更精准的诊断。这不仅可以提高诊断效率，还有助于早期发现糖尿病性视网膜病变的迹象，可以提供早期干预和定制化治疗，有助于防止病情的进一步恶化，最终降低患者视力受损的风险。

目前，糖尿病性视网膜病变的诊断主要依赖于专业医生对眼底图像的目视检查，这种方法受限于医生主观判断的影响，存在一定的误诊率和漏诊率。现有的自动化分割方案在处理糖尿病性视网膜病变时仍面

临一些挑战，例如病变区域的复杂形态、不同病变之间的相互影响等。因此，需要一种更为准确、稳健的算法来应对这些技术难题。

本团队的研究将以语义分割网络为基础，通过大量标注好的眼底图像数据集进行模型训练，将关注算法在分割糖尿病性视网膜病变方面的性能表现，力求在准确性和鲁棒性上取得平衡。本团队的工作流程将包括数据预处理、模型训练、验证和测试等环节，确保算法能够适用于不同类型的眼底图像，具备一定的泛化能力。最终，本团队的研究成果通过了本地测试和提交测试。

优点和特色方面，本团队的工作将充分利用深度学习技术在图像分析领域的优势，通过端到端的学习方式，使模型能够自动学习并理解眼底图像中的糖尿病性视网膜病变特征，避免了手动设计特征的复杂性。同时，本团队将关注模型的可解释性，以便医生能够理解算法的分割结果，增加医疗决策的可信度。

1.2 系统框架/Article Structure

本系统主要是基于 NestedUnet (Unet++) (2) 框架构建的。Unet++ 是对经典的 U-Net 模型的改进和扩展，旨在进一步提高语义分割任务的性能。它保留了 U-Net 的基本结构，同时引入了一些新的机制，包括密集连接和跨阶段连接，以增强模型的性能，从而适应医学图像处理领域的需求。Unet++ 的具体原理会在 3.3.2 节介绍。

图6展示了本系统的框架。首先，将大量的眼部医疗影像数据，包括视网膜底部图像等进行预处理，然后，采用了多种数据增广技术，包括水平翻转 (HorizontalFlip)、旋转 (Rotate)、随机裁剪 (RandomCrop)、高斯噪声 (GaussNoise)、随机亮度对比度调整 (RandomBrightnessContrast)、弹性变换 (ElasticTransform)、网格畸变 (GridDistortion) 等操作，以提高后续模型的训练效果。接下来采用 Unet++ 或 Unet 等模型进行训练，通过监督学习的方式，让模型学习特征并进行糖尿病性视网膜病变的分割。在训练过程中，本团队将使用交叉熵损失函数和优化算法来优化模型的参数。训练完成后，对测试集进行评估，包括计算模型的准确率、F1 分数等指标，以评估模型的性能。最后，将部署训练好的模型，用于实际的糖尿病性视网膜病变诊断，为临床医生提供可靠的辅助诊断工具，帮助提高疾病的早期检测率和诊断准确性。

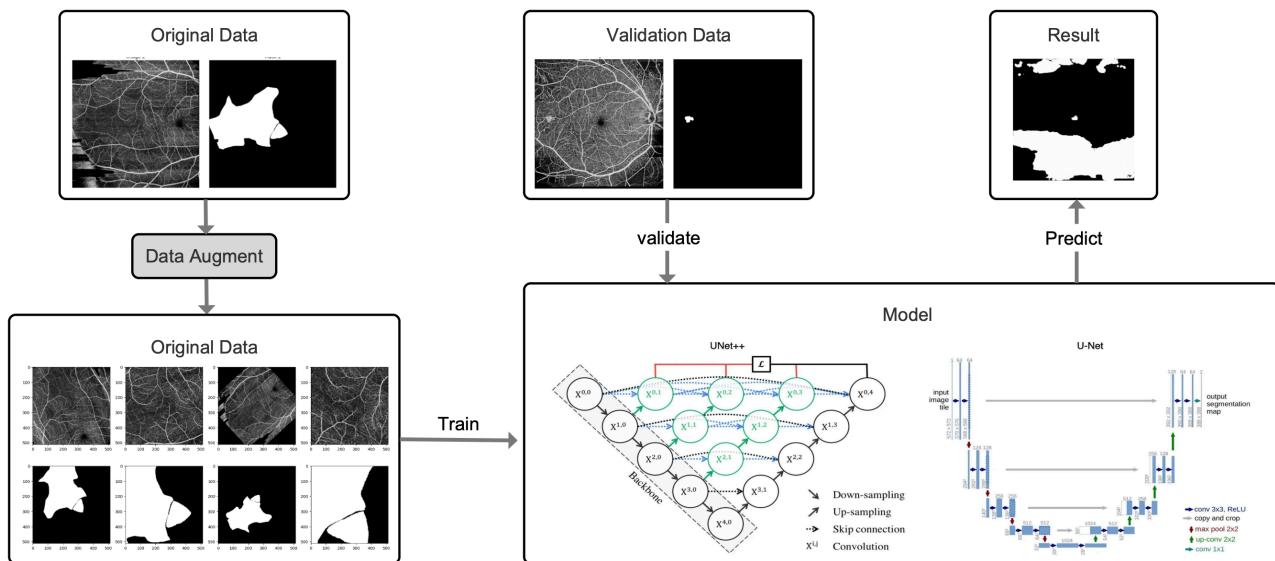


图 1: 系统总体框架

2 相关工作/Related Works

近年来，随着深度学习技术的不断发展，眼底图像分析领域取得了显著的进展。以下是关于糖尿病性视网膜病变检测、医疗影像语义分割网络及其临床应用的相关工作综述。

糖尿病性视网膜病变检测：糖尿病性视网膜病变是致盲的主要原因之一，其中 78% 的人有 15 年或更长糖尿病病史 (3)。糖尿病性视网膜病变通常引起血管结构的渐变变化和相应的异常。通过目视检查视网膜底

部图像以发现视网膜病变，如微动脉瘤（MAs）、视网膜内微血管异常（IRMAs）、非灌注区和新生血管，可诊断糖尿病性视网膜病变。这些病变的检测对糖尿病性视网膜病变的诊断至关重要。已有一些研究使用视网膜底部图像进行糖尿病性视网膜病变的诊断⁽⁴⁾。随着越来越受欢迎，光学相干断层扫描血管成像（OCTA）具有在微血管水平上以极大细节可视化视网膜和脉络膜血管的能力⁽⁵⁾。特别是，扫描源（SS）-OCTA 还可以额外评估脉络膜血管的状况。已有一些使用 SS-OCTA 对糖尿病性视网膜病变动性特征进行分级的工作⁽⁶⁾⁽⁷⁾。此外，超宽视网膜光学相干断层扫描血管成像（UW-OCTA）模式显示出在典型 OCTA 未能捕捉到的视网膜边缘的病变负担更高⁽⁸⁾。一些工作已经使用 UW-OCTA 进行糖尿病性视网膜病变分析⁽⁸⁾⁽⁹⁾。传统的糖尿病性视网膜病变分级诊断主要依赖于眼底摄影和荧光素血管造影（FFA），尤其是对于增生性糖尿病性视网膜病变，它严重危及视力健康。FA 主要用于检测新生血管的存在与否。眼底摄影难以检测早期或小的新生血管病变。FA 是一种侵入性的眼底成像技术，不能用于过敏、怀孕或肝肾功能不佳的患者。超宽 OCTA 可以非侵入性地检测糖尿病性视网膜病变新生血管的变化，因此它是帮助眼科医生诊断增生性糖尿病性视网膜病变的重要成像技术。然而，目前尚无能够利用 UW-OCTA 进行自动糖尿病性视网膜病变分析的工作。在糖尿病性视网膜病变分析过程中，首先需要评估 UW-OCTA 的图像质量，然后选择图像质量较好的图像。然后进行糖尿病性视网膜病变分析，如病变分割和增生性糖尿病性视网膜病变检测。

医疗影像语义分割网络：Tran 等人提出了一种多层 U-Net（Un-Net）网络模型，用于肝脏及肝脏肿瘤的 CT 图像分割。该模型在 U-Net 的基础上进行了改进，引入了多层结构（Un-Net），将卷积单元的输出特征作为跳跃连接，以更充分地利用卷积单元中的输出特征。实验结果表明，在 LiTS 和 3DIRCADb 两个公共数据集上，提出的模型在肝脏和肝脏肿瘤的分割任务中取得了优越的性能，Dice's 相似性系数（DSC）分别达到 96.38% 和 73.69%⁽¹⁰⁾。Azad 等人提出了 TransDeepLab，这是一种基于 Transformer 的医学图像分割模型，对 DeepLab v3+ 进行了扩展。该模型利用分层 Swin-Transformer 和移位窗口对 Atrous Spatial Pyramid Pooling (ASPP) 模块进行建模。通过引入 Transformer，TransDeepLab 更好地捕获了全局信息，相较于传统的 CNN 模型，在医学图像分割任务中取得了显著性能提升⁽¹¹⁾。Ibtehaz 和 Rahman 提出了 MultiResUNet，这是对传统 U-Net 架构的改进，用于多模态生物医学图像分割。通过在 U-Net 的基础上进行修改，提出的 MultiResUNet 模型在处理各种具有挑战性的数据集时取得了显著的性能提升。在五个不同的数据集上，MultiResUNet 相较于传统 U-Net 在性能上的相对提升分别为 10.15%，5.07%，2.63%，1.41% 和 0.62%。该模型在理想图像和挑战性图像方面均表现出色⁽¹²⁾。

图像分割的临床应用：Angelini 等人对胶质瘤的动态变化进行了广泛的研究，强调了计算机生长模型在临床应用中的重要性。他们提到了虚拟大脑的计算机生长，通过地图集和患者大脑 MRI 数据之间的注册，实现了真实的 3D 分段进化匹配。该文系统回顾了现有的计算任务，包括图像分割、图像配准以及硅生长模型。临床应用方面，作者强调了这些计算模型在胶质瘤治疗管理中的潜在重要性⁽¹³⁾。Luo 等人关注深度学习在口腔医学图像分割方面的临床应用。他们系统地回顾了基于深度学习的口腔医学图像分割方法，并分析了它们的优缺点。文中探讨了数据源、骨干网络和任务形式等主要方面。此外，文中还提到了这些深度学习方法在口腔医学领域的临床应用，强调了它们在医学图像分析中的潜在作用⁽¹⁴⁾。Luo 等人针对腹部器官分割任务建立了一个名为 WORD 的大规模数据集，并进行了相关研究。他们指出，全腹部器官分割对于诊断腹部病变、放射治疗和随访至关重要。研究中，提出的 WORD 数据集包含 150 个腹部 CT 卷（30495 张切片），每个卷有 16 个器官，具有精细的像素级注释和基于涂鸦的稀疏注释。该数据集为腹部多器官分割任务提供了新的基准，同时通过邀请三名经验丰富的肿瘤学家对模型预测进行修订，衡量了深度学习方法与专业医生之间的差距。这项工作不仅为腹部多器官分割提供了新的基准，还探索了在大规模数据集上的推理效率学习和基于涂鸦的注释有效学习⁽¹⁵⁾。

3 研究内容与方法 (或算法)/Contents and Methods(or Algorithm)

3.1 数据准备和产生

3.1.1 数据来源与分析

本项目的数据来源自 DRAC 2022: Diabetic Retinopathy Analysis Challenge 赛事方。在赛事网站 (<https://drac22.grand-challenge.org/>) 上下载获得数据集，内容包括了完整的、未经划分的训练集与比赛所用测试集。

在进行下一步的数据工作前,本团队确认了原版数据的具体格式。训练集的原图同遮罩尺寸为 1024*1024, 图片格式为单通道灰度格式的 PNG 图像。以上特征将会决定本团队后续模型的预处理及输入层设置。

分割任务数据集中包含了三个子任务, 分别为视网膜内微血管异常 Intraretinal Microvascular Abnormalities, 非灌注区 Nonperfusion Areas, 新生血管 Neovascularization。这三者的病理学意义在前文已经叙述, 在数据确认部分只需要通过阅读遮罩定性分析任务难易度(以指导后续训练调参过程), 以及可用数据集大小。

最终定性分析的结论是, 非灌注区的难度相对较低, 另外两个子任务的难度基本相当。为了更清晰地展示这一情况, 下方将对每一子任务举例说明。

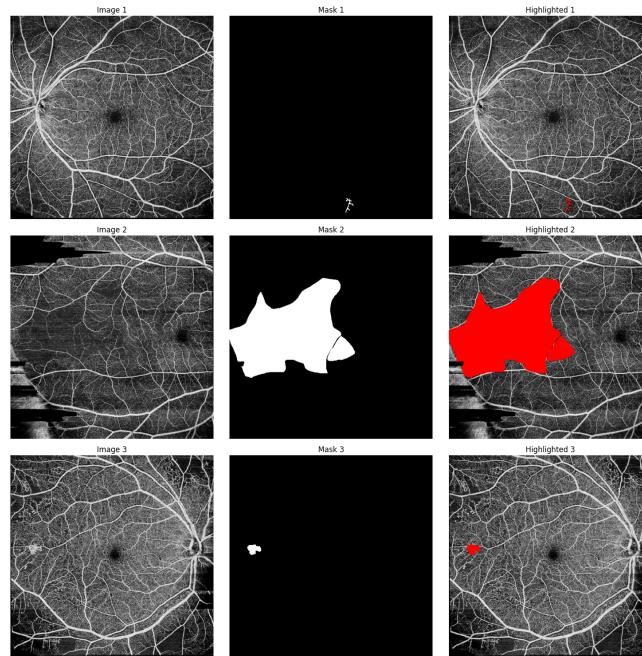


图 2: 子任务展示

而子任务数据集的情况如下表:

Task	Traning Set
Intraretinal Microvascular Abnormalities	86
Nonperfusion Areas	106
Neovascularization	35

表 1: Dataset 展示

3.1.2 数据预处理

在绝大多数情况下, 数据预处理的目的在于匹配模型与增强结果, 本团队也就这两方面情况分析数据预处理的需求。

首先, 原数据的数据类型为 1024x1024 的 png 单通道灰度图。根据这一特点, 本团队具体规划了模型的输入层形状, 使得最终训练模型能够与原数据在格式上匹配良好, 不需要格式上的预处理。

而后, 对数据集进行数据增广, 本团队认为非常有必要。这是由于医学影像的图像处理数据集往往相当有限, 不经过一定程度地增广很容易造成学习能力低, 预测效果差的情况。具体到类别上, 本团队通过工具, 向原数据集图像中引入了 HorizontalFlip, Rotate, RandomCrop, GaussNoise, RandomBrightnessContrast, Compose, Resize, Normalize, OneOf, ElasticTransform, GridDistortion, Affine 等一系列操作, 极大地扩充了数据集的可用性与有效性, 具体操作如下:

首先, 本团队使用了水平翻转 (HorizontalFlip) 和旋转 (Rotate) 操作, 这些操作可以增加数据的多样性, 从而使模型更加具有泛化能力。通过随机翻转和旋转图像, 可以生成更多变体的图像, 从而扩充训练

数据集。接着采用了随机裁剪（RandomCrop）操作，将输入图像裁剪为指定大小。这有助于模型学习到不同尺度下的特征，并增加模型对于输入图像尺度变化的鲁棒性。同时，引入了高斯噪声（GaussNoise）操作，以模拟真实世界中的噪声情况，从而增强模型的鲁棒性。随后使用了一系列仿射变换和弹性变换操作，包括弹性变换（ElasticTransform）、网格扭曲（GridDistortion）、仿射变换（Affine）和随机亮度对比度调整（RandomBrightnessContrast）。这些变换操作可以进一步增加数据的多样性，使模型更具有鲁棒性，并提高模型对不同光照和视角下图像的泛化能力。具体效果展示如图3。最后，对图像进行标准化处理（Normalize），将图像的像素值分布调整到标准正态分布。这有助于加速模型的收敛过程，并提高训练效果。

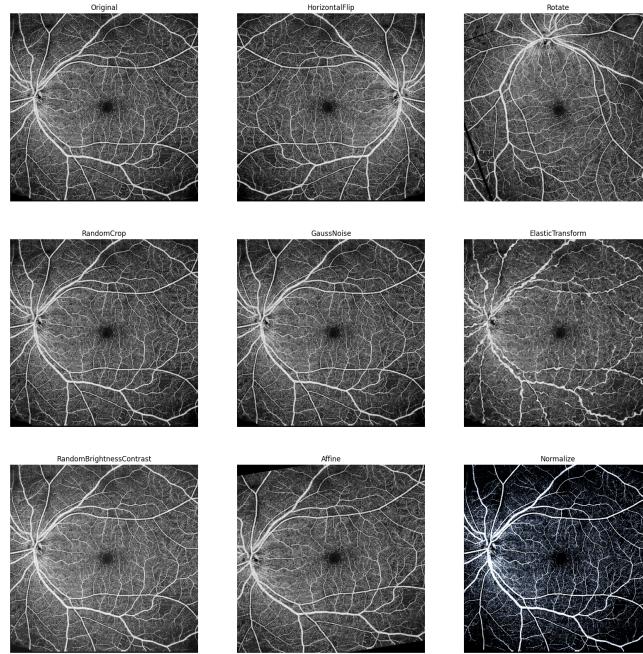


图 3: 数据增广效果

3.2 模型框架选择

基于糖尿病性视网膜病理图像特征，本团队分析该任务属于典型的语义分割问题。其目标在于将眼底图像中的病变区域基于整体图像位置精确分割，而不关注病变块的每个实例，这与语义分割任务的典型特征相符。基于此，本团队对语义分割模型进行了研究与探讨。

一般地，语义分割模型建立在 CNN 网络提取特征的基础上，但早期的语义分割模型陷于特征图过小无法保留信息与巨大计算量的困境，如 FCN 等原始模型。这些模型会对网络中的图片进行相当程度的压缩，导致空间信息严重损失，进一步地影响分割的准确性。

为了解决这一问题，后续的研究上，围绕 U-Net 架构作为解决方案，构造了一系列的模型。一种是以 NestedUnet (Unet++) (2) 为代表的架构，它使用金字塔状的结构，通过增加多个嵌套的 U-Net 模块来构建一个更深层次、更复杂的网络。这种结构使得模型能够更好地捕获不同尺度的特征信息，并且具有更强的特征表征能力；另一种是以 nnU 为代表的架构，它通过添加嵌套结构和采用更复杂的损失函数等方式，增强了模型在医学图像分割任务中的性能和泛化能力。

经过比较选择，本团队最终采用了 Unet++ 系列的架构来解决本次的语义分割任务。后续会详细说明该类架构的模型细节与独特优势。

3.2.1 nnU-Net 模型

1.2 节中提到，nnU 只是在 Unet 的基础上修改了激活函数和归一化的方法，具体改动是：

- ReLU 换 leaky ReLU (neg.slope 1e-2)
- Batch Norm 换 Instance Norm

nnU 架构包括三种不同的模型：2D U-Net、3D U-Net 和 U-Net 级联模型。这些模型都是基于 U-Net 架构的变体，用于医学图像分割任务。

2D U-Net 通常用于处理 2D 医学图像数据。虽然在处理 3D 医学图像分割任务时，使用 2D U-Net 可能不是最佳选择，因为无法充分利用沿着 z 轴方向的有价值信息。然而，有证据表明，对于具有各向异性特性的数据集，传统的 3D 分割方法性能下降。例如，在处理前列腺数据集时就存在这样的情况。

相比之下，3D U-Net 是处理 3D 医学图像数据的合适选择。理想情况下，可以在整个患者图像上进行训练。但由于 GPU 内存限制，通常只能在图像块（patch）上训练此体系结构。对于一些数据集，如脑肿瘤、海马和前列腺等，基于 patch 的训练效果较好。然而，对于一些大型结构，如肝脏等，由于图像尺寸过大，patch 训练会损失大量的上下文信息。

为了解决 3D U-Net 在处理大型图像尺寸的数据集时的缺陷，提出了 U-Net 级联模型。下图 3 展示了该级联模型的结构。该模型包括两个阶段：首先，对降采样的图像进行 3D U-Net 训练。其次，将第一个 U-Net 的分割结果上采样到原始体素间距，并作为一个热编码的附加输入通道传递到第二个 3D U-Net。第二个 3D U-Net 在全分辨率上进行训练，以提高分割性能。

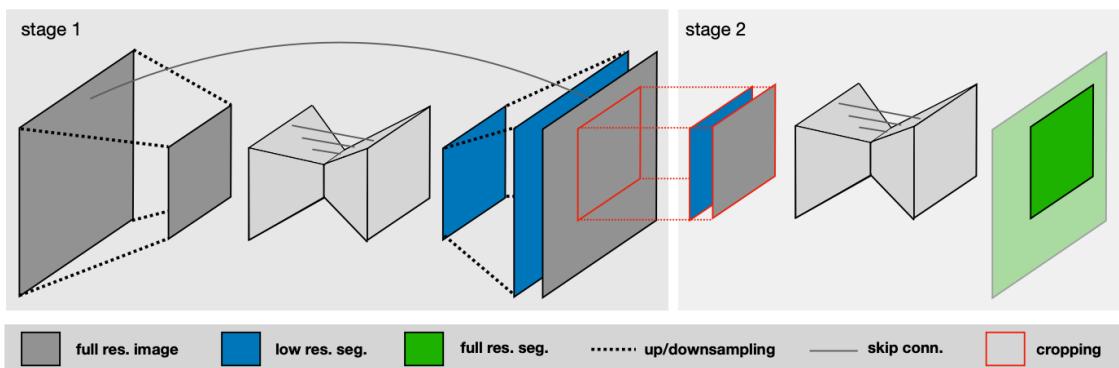


图 4: nnU-Net 级联模型结构 (1)

3.2.2 Unet++ 模型

Unet++ 是一种用于图像分割任务的深度学习架构，它是对传统 Unet 模型的改进和扩展。Unet++ 的设计灵感源自 Unet，但在网络结构上做出了一系列创新，以提高分割性能和网络的表达能力。UNet++ 由深度不同的 U-Net 组成，其解码器通过重新设计的跳跃路径在相同分辨率上密集连接，旨在解决 U-Net 的两个关键挑战：1) 最佳架构的深度未知；2) 跳跃连接设计过于限制。Unet++ 的核心思想是通过多层次的跳跃连接（skip connections）来促进信息的流动，并在网络的不同层次上捕获不同尺度的特征信息。与传统的 Unet 不同，Unet++ 在每个下采样（编码器）和上采样（解码器）路径上都添加了一个额外的分支，称为“Nested”分支。这个 Nested 分支由两个连续的卷积块组成，每个卷积块都具有多个卷积层和非线性激活函数，用于提取和表达更丰富的特征信息。通过这种方式，Unet++ 不仅可以更好地捕获局部和全局特征，还可以有效地避免信息丢失和梯度消失问题。下图 4 展示了 Unet++ 和 Unet 的架构对比。

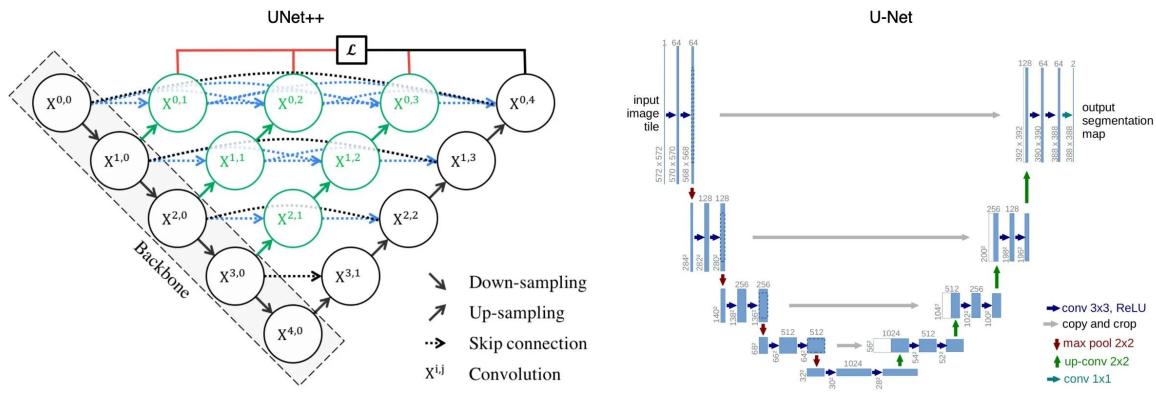


图 5: UNet++ 和 UNet 架构对比 (2)

如图 3 所示, UNet++ 在原始的 U-Net 上引入了 3 个改进:

- 重新设计的跳跃路径 (显示为绿色)
- 密集跳跃连接 (显示为蓝色)
- 深度监督 (显示为红色)

重新设计的跳跃路径: 在 UNet++ 中, 增加了重新设计的跳跃路径 (绿色显示), 以弥补编码器和解码器子路径之间的语义差别。这些卷积层的目的是减少编码器和解码器子网络的特征映射之间的语义差距。因此, 对于优化器来说, 这可能是一个更直接的优化问题。U-Net 采用跳跃连接, 直接连接编码器和解码器之间的特征映射, 导致把语义上不相似的特征映射相融合。然而, 在 UNet++ 中, 相同 dense block 的前一个卷积层的输出与较低层 dense block 对应的上采样输出进行融合。这使得已编码特征的语义级别更接近于等待在解码器中的特征映射的语义级别, 因此, 当接收到语义上相似的特征映射时, 优化更容易。跳跃路径上的所有卷积层使用大小为 3×3 的核。

密集跳跃连接: 在 UNet++ 中, 密集跳跃连接 (用蓝色显示) 实现了编码器和解码器之间的跳跃路径。这些 Dense blocks 是受到 DenseNet 的启发, 目的是提高分割精度和改善梯度流。密集跳跃连接确保所有先验特征图都被累积, 并通过每个跳跃路径上的 dense 卷积块而到达当前节点。这将在多个语义级别生成完整分辨率的特征映射。

深度监督: 在 UNet++ 中, 深度监督 (用红色显示) 通常通过在解码器的不同层级上添加附加的输出层和损失函数来实现。这些附加的输出层可以在不同的分辨率上生成预测结果, 并与原始标签进行比较, 从而计算相应的损失。通过引入深度监督, UNet++ 可以使网络在训练过程中更加稳定和可靠。这是因为即使在网络较深的层级上, 也可以及时地传播误差信号, 并且在每个阶段都可以对特定的特征进行监督学习。这样可以有效地避免梯度消失或梯度爆炸的问题, 并且有助于网络学习到更加鲁棒和准确的特征表示, 从而提高分割性能和泛化能力。

3.2.3 其余模型

在模型框架的选择中, 本团队还考虑了一系列的备用模型, 包括但不限于 Deeplab、SAMadapter 等选项。尽管后续以上框架由于各种原因未被纳入项目使用, 但本团队认为有必要说明以上框架的特点与遗憾, 以补充实验过程中的考量。

Deeplab 是与 unet 并列的另一大分割模型框架思路, 其核心思路是依赖空洞卷积技术以及空间金字塔池化设计来传递不同粒度的图像语义信息。但相比于 Deeplab, Unet 网络在小尺度细节上更有优势, 对于医学图像分析这一课题更加适应, 尤其在本课题的 task1 与 task3 中。

SAMadapter(16) 是基于 Segment Anything (SAM) 这一图像分割任务基础大模型, 所设计出的 Adapter 模块, 使得可以在不微调 SAM 网络的情况下适配各类下游任务。现有的一些工作已经可以说明 SAMadapter 具有有效处理医学图像分割的能力。SAMadapter 的核心特点在于引入了稀疏注意力模块和适配器模型。稀

疏注意力模块通过采用较少的注意力头和较小的局部注意力窗口，降低了计算复杂度。这种设计保持了模型的性能，同时减少了参数数量和计算成本，特别是在处理长序列数据时效果显著。适配器模型是 SAMadapter 的另一个重要组成部分，它是一种轻量级的模型组件，通常由几层线性变换和激活函数组成。适配器模型能够在保持预训练模型参数固定的同时，通过添加少量参数，将模型应用到特定的下游任务上。下图 5 展示了 SAMadapter 的基本架构。

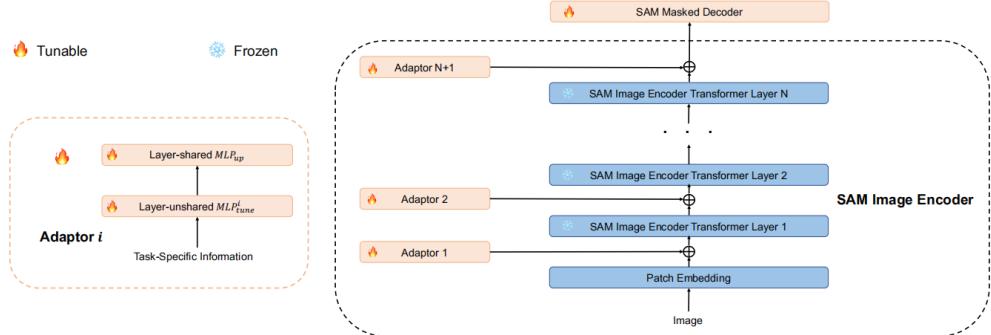


图 6: SAMadapter 基本架构 (16)

3.3 模型训练

本团队使用 Nvidia Tesla V100-32G 显卡通过 SAM-adapter 模型对 task1 训练 100 代左右，发现其展现出了大模型的优势——它的训练曲线平稳的多，但是由于资源消耗实在过大（30G 显存要求），训练时间极长且收益不高，本团队最终没有继续这个尝试。

本团队使用 3090 显卡通过 Unet 及 Unet++ 进行超过三千轮的训练尝试。值得一提的是，数据集在训练中表现并不稳定，在过程中选择具有高 dice 得分的权重继续训练能够提高结果期望，最佳的结果对于后面二者大约在 300 代左右出现。在 IMA (Intraretinal Microvascular Abnormalities) 和 NA (Nonperfusion Areas) 类别的分割中本团队使用 Unet++ 进行训练取得了良好的成果，在 NV (Neovascularization) 中使用 UNet 进行训练，取得了尽可能的最优结果。

3.3.1 损失函数与优化器的选择

目前主流的图像分割问题所使用的损失函数包括交叉熵损失 Cross Entropy Loss、Dice Loss 等。本团队经过讨论，选择了 Dice Loss 作为模型训练时的损失函数。

Dice Loss 是一种基于相似度度量的损失函数，用于衡量模型预测结果与真实结果之间的重叠程度。通过计算模型预测的掩码和真实掩码之间的 Dice 系数，即两个集合的重叠部分与它们的平均大小之比，来评判模型预测结果。一般地，可以如此表示 Dice Loss。

$$L_{dice} = 1 - \frac{2|X \cap Y|}{|X| + |Y|}$$

在本课题中，Dice Loss 相比于交叉熵损失的优势是明显的。交叉熵损失逐个像素地考察分类的准确性，在多类别的分割任务中有更好的表现。但这种分类方式会受到数据类别平衡的影响，换而言之，当数据本身类别不平衡，存在绝对主导的大类别与细小分割时，交叉熵模型的鲁棒性是得不到保证的。以本课题为例，在 task1 与 task3 中，健康部分的区域往往占据整张图片的绝大部分，使得交叉熵损失总是处于一个较好的水平，对预测的评判效果有限。而 Dice Loss 对像素数量的依赖更小，一定程度上增强了鲁棒性。

而优化器方面，本团队使用了 Adam 优化器，这也是在目前图像分割模型中被广泛使用的主流选择。Adam 是一种结合了自适应学习率调整和动量优化的特性的基于梯度的优化算法。其优势在于为每个参数维护了自适应的学习率，这种设计能更好适应医学影像分割这类复杂模型。另一方面，动量优化的技术使得 Adam 能积累之前的梯度信息，以此加速收敛过程。在有限的时间成本内获得较好较稳定的训练结果，这是本团队着重考量的。

3.3.2 训练参数

Architecture	Initial learning rate	Minimum learning rate	Scheduler	Metric	Optimizer	Loss Function	Input Size
Unet++	1E-04	1E-09	MultiStepLR	Dice	Adam	Dice Loss	[1024,1024]
Unet	1E-04	1E-09	CosineAnnealingLR	Dice	Adam	Dice Loss	[1024,1024]
SAMadapter	2E-04	1E-07	MultiStepLR	Dice	AdamW	Dice Loss	[1024,1024]
DeepLabV3+	5E-04	5E-09	\	Dice	Adam	Dice Loss	[1024,1024]

表 2: 各模型训练参数

在 scheduler 选择上，使用 Unet++ 时，本团队采用 MultiStepLR 调度器，它允许在训练的预定周期点显著降低学习率。具体来说，在总训练周期的四分之一时，将学习率降至初始学习率的 10%。这种方法有助于快速接近全局最优解，在初期利用较大的学习率进行大范围的搜索，然后通过减少学习率来细化搜索，从而提高训练的稳定性和最终模型的性能。

在使用 Unet 进行训练时采用的则是 CosineAnnealingLR 调度器，这种调度器会根据一个周期性的余弦函数降低学习率，模拟退火过程中的冷却行为。在每个周期的四分之一时刻，它将学习率调整为前一学习率的 60%。这种平滑的学习率减少方式有助于在训练过程中避免陷入局部最优解，同时保持一定的探索能力，从而在全局范围内寻找最优解。

当使用 Adam 优化器时，这两种调度器各有其优势。Adam 优化器本身就包含了对学习率的自适应调整能力，结合这些调度器可以进一步细化学习率的控制。Adam 优化器通过计算梯度的一阶矩估计和无偏二阶矩估计自动调整每个参数的学习率，与之结合的 MultiStepLR 或 CosineAnnealingLR 调度器可以在 Adam 的基础上，根据训练进度提供一个大范围的系统性调整。这样的组合使得网络训练更加高效，能在初期快速下降，在接近收敛时通过更精细的学习率调整避免过度摆动，从而提高模型的准确率和稳定性。

在模型验证阶段，本团队选择了与比赛评价标准相一致的 Dice 系数（Dice metric）作为性能评估的指标。这一指标通过计算两倍的预测和真实值之间交集的大小，除以预测和真实值之和来实现。其值越接近 1，表示模型的预测结果与实际情况越吻合，表现越优秀。

使用 Dice 系数作为评估标准的优势在于其对小区域的敏感性较高，能有效评估模型在捕捉边界细节方面的能力。这对于医学图像分析尤为重要，因为在这类应用中，边界的准确识别直接关系到诊断的准确性。此外，Dice 系数也支持更全面地理解模型在不同类别间的表现，帮助本团队在开发过程中进行更针对性的优化，从而提高模型的整体性能。其他具体参数如2所示

3.3.3 分割效果可视化

为了充分说明训练情况，以 task3 为例，图7展示了 500 次迭代中模型训练中 loss 的变化，可以看到训练损失随着周期增加而稳定下降，表示模型在训练数据集上的预测性能在不断提高。而验证损失在多个周期内波动较大，这表明模型在未见过的数据集上的泛化能力不够稳定。

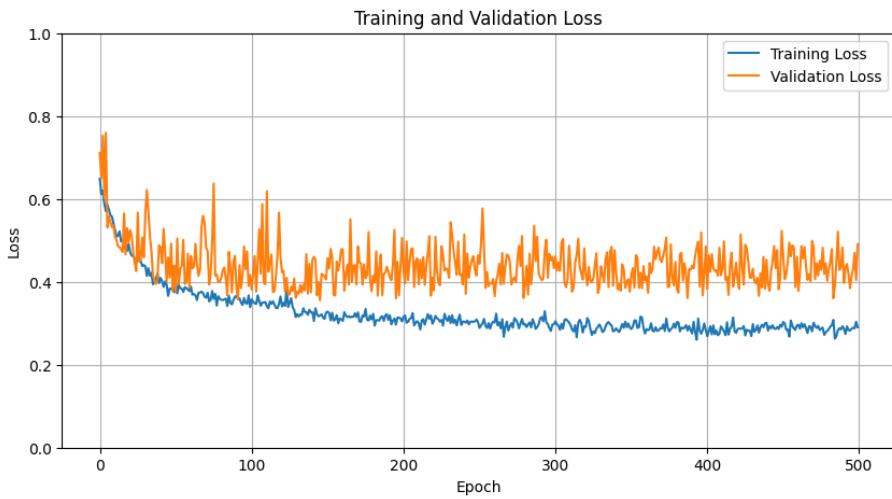


图 7: task3 训练过程中的 loss

4 实验结果与分析/Results and Analysis

4.1 效果展示/Result Demonstration

经由以上工作，本团队基于 Unet++ 与 Unet 框架构建模型，并用赛事方提供的数据集展开学习，训练得到适用于糖尿病性视网膜病变的病理图像分割模型。

最终本团队的模型对于赛事方要求的三个子任务，即视网膜内微血管异常 Intraretinal Microvascular Abnormalities，非灌注区 Nonperfusion Areas，新生血管 Neovascularization，分别展开训练，并进行了漫长的筛选工作获得各自对应的最佳权重 pth。下图展示了在划分得到的验证集上训练的结果，选取出来的较为代表性的样例足以说明，模型在病理图像分割上已然学习到了不错的结果。

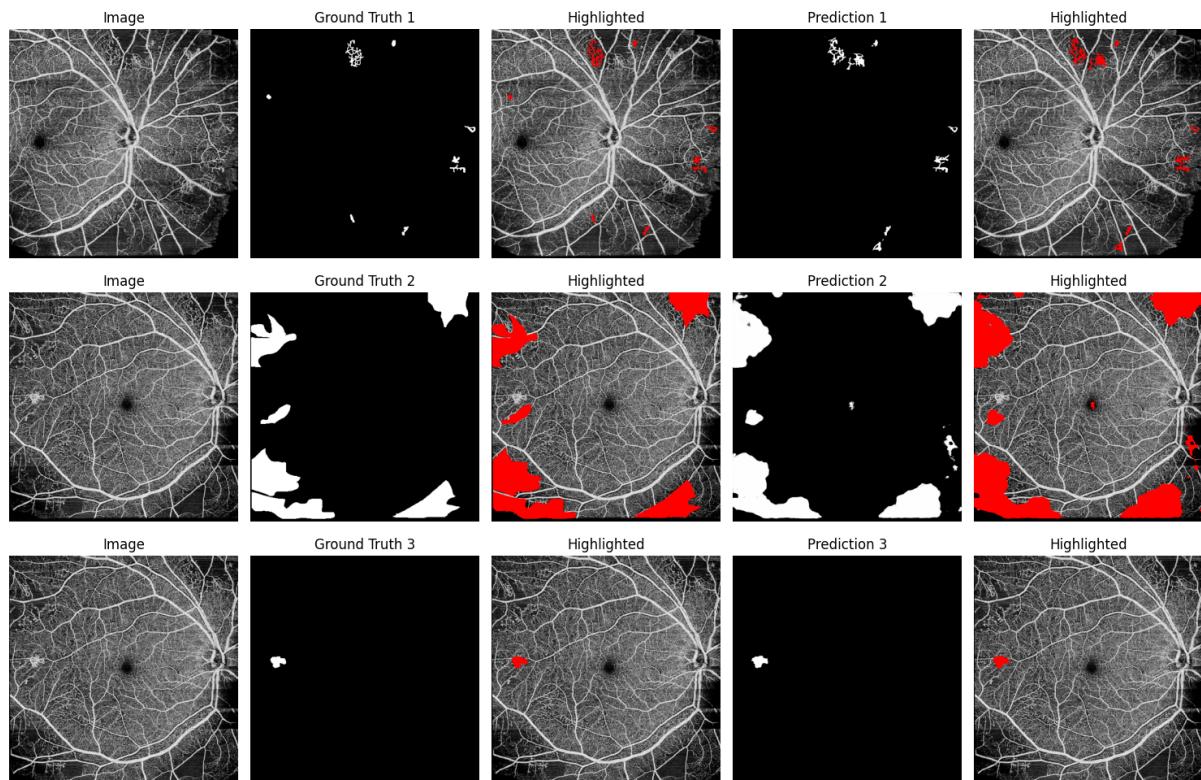


图 8: 验证集效果展示

随后，本团队使用以上模型对赛事方提供的测试集进行了预测工作，图9展示了其中具有代表性的结果。

另有表3展示了模型在测试集上预测结果的 Dice Loss (具体计算值由赛事方网站提供)。

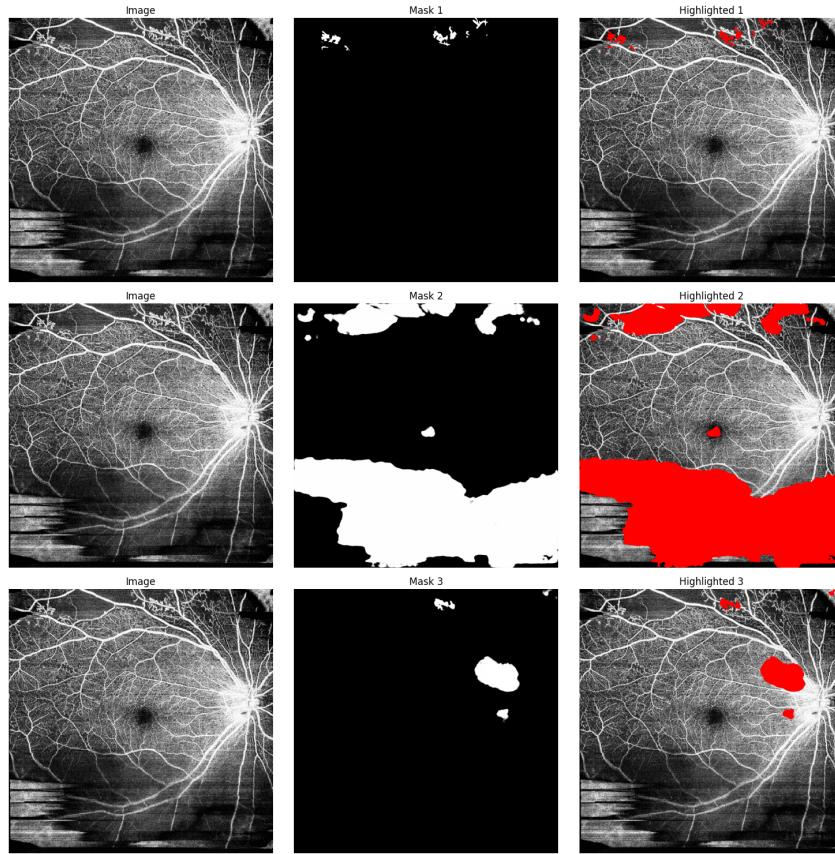


图 9: 测试集效果展示

Task1 Dice	Task2 Dice	Task3 Dice
0.3764	0.5849	0.4888

表 3: 最终比赛结果展示

4.2 实验结果对比/Comparison of results

Model	Task1 Dice	Task2 Dice	Task3 Dice
Unet	\	0.5526	0.4888
Unet++	0.2885	\	\
Unet++ with Attention Layer	0.3764	0.5849	0.4354
DeepLab	0.1373	0.5078	\
SAMadapter	0.3061	\	\

表 4: 不同模型结果对比

如表4所示，本团队尝试并比较了五种不同的模型在三个医学影像分割任务上的性能。结果显示，虽然每种模型的表现各有千秋，但 Unet++ 加入注意力层后，在多个任务上均表现出较高的 Dice 分数，尤其是在第二个任务上达到了 0.5849 的高分。这突显了注意力机制在提升模型分割精度方面的潜力。DeepLab 模型在任务二也表现良好，但在任务一的分数较低。SAMadapter 模型则在任务一表现不俗，展现了其处理复杂分割任务的能力。这些结果为未来选择合适的模型提供了宝贵的参考，有助于根据具体任务需求定制和优化模型。

5 特色与创新

5.1 通道注意力机制

通道注意力机制 (Channel Attention Module: CAM) 是这个项目的特色与创新之一。CAM 是一种轻量级的注意力机制，专门用于增强特征图中的通道间的信息交互，以提高模型性能。CAM 的架构如下：

- **平均池化层**: 首先，对输入的特征图进行全局平均池化操作，将特征图的空间维度压缩为 1×1 ，保留每个通道的平均值作为该通道的全局特征表示。
- **全连接层**: 然后，通过一层全连接层，将全局特征表示映射到一个新的向量空间。这里的全连接层采用了两层，中间包含 ReLU 激活函数，以增加模型的非线性能力。
- **Sigmoid 函数**: 最后，将全连接层的输出通过 Sigmoid 函数，将其值缩放到 $[0, 1]$ 的范围内，得到的输出作为每个通道的注意力权重。这些权重用于对原始特征图进行加权，以产生具有更强的语义表征的特征图。

通过引入 Channel Attention Module，模型能够动态地学习每个通道之间的关系，并且根据输入图像的内容自适应地调整注意力权重，从而增强了特征图的表示能力。这使得模型能够更好地捕获输入图像中的重要特征，提高了模型的语义分割性能，同时还减少了参数数量和计算成本。

5.1.1 采用 SAM 大模型进行医学影像分割

本团队创新性地在比赛中尝试了 SAM-Adapter 模型，这是一个针对大型预训练模型 SAM 进行的修改，特别是为了复杂的医学图像分割任务。这种方法将任务特定信息直接整合到 SAM 网络中，而无需对其进行全面的重新训练。该方法使用视觉提示来指导网络关注医学图像的相关特征，从而显著提高了分割准确性和模型的适应性。这种策略不仅保留了原始 SAM 模型的稳健性，还确保了高效的性能，这在医疗应用中至关重要，其中精确和可靠的图像分析是必需的。这可能代表了人工智能在医疗保健应用中的一次突破，为医学诊断和治疗计划提供了一个强大的工具。

6 总结/Conclusion

针对糖尿病性视网膜病变的病理图像分割课题，本团队以 DRAC2022 赛事方提供的数据集为基础，构建了一套能有效分割视网膜内微血管异常、非灌注区、新生血管三类病变区域的模型。本团队尝试了基于 Unet++ 为主的多种模型框架，并引入了注意力 Attention 等机制优化模型。最终训练得到的理想模型在验证集与测试集上均表现出了较好的成绩，被证明能够切实有效地完成病理图像分割任务。

参考文献

- [1] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajittra, S. Wirkert, and K. H. Maier-Hein, “nnu-net: Self-adapting framework for u-net-based medical image segmentation,” 2018.
- [2] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: A nested u-net architecture for medical image segmentation,” 2018.
- [3] M. Tian, S. Wolf, M. R. Munk *et al.*, “Evaluation of different swept-source optical coherence tomography angiography (ssocta) slabs for the detection of features of diabetic retinopathy,” *Acta ophthalmologica*, vol. 98, no. 1, 2019.
- [4] L. Dai, L. Wu, H. Li *et al.*, “A deep learning system for detecting diabetic retinopathy across the disease spectrum,” *Nat Commun*, vol. 12, p. 3242, 2021.
- [5] R. F. Spaide, J. G. Fujimoto, N. K. Waheed *et al.*, “Optical coherence tomography angiography,” *Progress in retinal and eye research*, vol. 64, pp. 1–55, 2018.
- [6] K. B. Schaal, M. R. Munk, I. Wyssmueller *et al.*, “Vascular abnormalities in diabetic retinopathy assessed with sweptsource optical coherence tomography angiography widefield imaging,” *Retina*, vol. 39, no. 1, pp. 79–87, 2019.
- [7] P. E. Stanga, A. Papayannis, E. Tsamis *et al.*, “New findings in diabetic maculopathy and proliferative disease by swept-source optical coherence tomography angiography,” *OCT Angiography in Retinal and Macular Diseases*, vol. 56, pp. 113–121, 2016.
- [8] Q. Zhang, K. A. Rezaei, S. S. Saraf *et al.*, “Ultra-wide optical coherence tomography angiography in diabetic retinopathy,” *Quantitative imaging in medicine and surgery*, vol. 8, no. 8, p. 743, 2018.
- [9] J. F. Russell, Y. Shi, J. W. Hinkle *et al.*, “Longitudinal wide-field swept-source oct angiography of neovascularization in proliferative diabetic retinopathy after panretinal photocoagulation,” *Ophthalmology Retina*, vol. 3, no. 4, pp. 350–361, 2019.
- [10] S. T. Tran, C. H. Cheng, and D. G. Liu, “A multiple layer u-net, un-net, for liver and liver tumor segmentation in ct,” *IEEE Access*, vol. 9, pp. 3752–3764, 2021.
- [11] R. Azad *et al.*, “Transdeeplab: Convolution-free transformer-based deeplab v3+ for medical image segmentation,” in *Predictive Intelligence in Medicine*. Springer, 2022.
- [12] N. Ibtehaz and M. S. Rahman, “Multiresunet : Rethinking the u-net architecture for multimodal biomedical image segmentation,” *Neural Networks*, vol. 121, pp. 74–87, 2020.
- [13] E. D. Angelini, O. Clatz, E. Mandonnet *et al.*, “Glioma dynamics and computational models: a review of segmentation, registration, and in silico growth algorithms and their clinical applications,” *Current Medical Imaging*, vol. 3, no. 4, pp. 262–276, 2007.
- [14] D. Luo, W. Zeng, J. Chen *et al.*, “Deep learning for automatic image segmentation in stomatology and its clinical application,” *Frontiers in Medical Technology*, vol. 3, p. 767836, 2021.
- [15] X. Luo, W. Liao, J. Xiao *et al.*, “Word: A large scale dataset, benchmark and clinical applicable study for abdominal organ segmentation from ct image,” *arXiv preprint arXiv:2111.02403*, 2021.
- [16] T. Chen, L. Zhu, C. Ding, R. Cao, S. Zhang, Y. Wang, Z. Li, L. Sun, P. Mao, and Y. Zang, “Sam fails to segment anything? – sam-adapter: Adapting sam in underperformed scenes: Camouflage, shadow, and more,” 2023.

竞赛结果截图

Grand Challenge Challenges Algorithms Reader Studies ...

Evaluation

ID
38bdb334-d7b6-48b9-9ef0-d2d1b15ed710

Submission ID
93c78de8-7f7d-4625-a86f-bc1b31654a2a

Method ID
83ffbe71-4b1b-4eae-b274-a6878032fcab

Status
Succeeded

User
 wintersun0729

Challenge
DRAC22

Phase
Task 1 Segmentation - Post Challenge Submission

Submission created
April 28, 2024, 9:12 p.m.

Result created
April 28, 2024, 9:12 p.m.

Position on leaderboard
10

图 10: 提交测试结果

Grand Challenge Challenges Algorithms Reader Studies ... Help ▾    

Task 1 Segmentation - Post Challenge Submission Leaderboard

Search: Additional metrics ▾ Hide additional metrics

#	User (Team)	Created	% mean DSC	% mean IoU	Class 1 DSC	Class 2 DSC	Class 3 DSC	Comment	% Publication	% Supplementary File
1st	 lhy-init	21 April 2024	0.5931	0.4461	0.4554	0.6911	0.6328	submit1		
2nd	 jianke0604	25 April 2024	0.5756	0.4265	0.4599	0.6483	0.6187			
3rd	 PQPQ	22 April 2024	0.5551	0.4015	0.4716	0.6235	0.5703			
4th	 gs63891 (galactic fleet)	7 Oct. 2022	0.5544	0.4059	0.4403	0.6426	0.5804			
5th	 liyihao19973 (LiTiM)	3 Oct. 2022	0.5387	0.3966	0.4079	0.6515	0.5566			
6th	 Linus (AIM-lab)	7 Oct. 2022	0.5290	0.3847	0.4312	0.6039	0.5518			
7th	 liuyushu0206	22 Dec. 2023	0.5149	0.3762	0.3655	0.6108	0.5683	nn		
8th	 sharryXR	27 April 2024	0.5129	0.3664	0.4141	0.6057	0.5190	5		
9th	 lsyragnarok (DIP)	7 Jan. 2023	0.5012	0.3552	0.4167	0.6018	0.4852			
10th	 wintersun0729 (AMDYes)	28 April 2024	0.4833	0.3428	0.3764	0.5849	0.4888			
11th	 official	11 May 2023	0.4750	0.3370	0.3802	0.5903	0.4544			
12th	 210227028 (DMTT)	1 Dec. 2022	0.4743	0.3527	0.3264	0.6445	0.4520	base_nnunet		
13th	 Jiaolr	27 April 2024	0.4655	0.3286	0.3764	0.5849	0.4354			

图 11: 排名榜单