



南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

本科生毕业设计（论文）

题 目： 基于虹膜 OCT 图像全局与局部分割
 框架的研究

姓 名： 周一凡

学 号： 11910311

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 教授

年 月 日

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

基于虹膜 OCT 图像全局与局部分割框架的研究

周一凡

(计算机科学与工程系 指导教师：刘江)

[摘要]：虹膜是眼睛中非常重要的组织之一。临床上可以通过分析眼前节光学相干断层成像 (Anterior segment Optical Coherence Tomography, AS-OCT) 中虹膜的形态与结构，来辅助青光眼、虹膜炎等疾病的诊断。然而，在光学相干断层成像中，虹膜尾端邻角膜、睫状肌，存在较为棘手的模糊区域。这给虹膜的准确分割带来挑战。本研究针对 AS-OCT 中的虹膜分割任务，提出了一个全局与局部分割的模型框架。该框架由全局分割与关键点定位网络、局部分割网络以及融合网络组成。全局分割与关键点定位网络由一个共享的编码器和两个独立的解码器组成，用于完成全局分割和关键点定位。关键点定位为局部分割网络生成关键区域子图。局部分割网络专注于易错区域的分割。融合网络根据全局和局部分割的结果生成最终输出。实验在 290 张图像的私有虹膜数据集和 1800 张图像的公开数据集 CAMUS 上进行。实验结果表明，该方法在存在相同难点的分割任务中对比 U-Net、CE-Net 和 U-Net++ 拥有更好的精准度。

[关键词]：深度学习；医学图像分割；全局与局部分割；注意力机制

[ABSTRACT]: The iris is one of the most important tissues in the eye. The morphology and structure of the iris on Anterior segment Optical Coherence Tomography (AS-OCT) is used clinically to assist in the diagnosis of glaucoma, iritis and other diseases. However, in AS-OCT images, there is a tricky blurring of the iris caudal to the cornea and ciliary muscle. This poses a challenge for accurate segmentation of the iris. In this study, a framework for global and local segmentation is proposed for the task of iris segmentation of AS-OCT images. The framework consists of a global segmentation and key point localization network, a local segmentation network and a fusion network. The global segmentation and key point localization network consists of a shared encoder and two independent decoders for global segmentation and key point localization. Key point localization generates key-point-patches for the local segmentation network. The local segmentation network focuses on the parts with higher difficulty, resulting in more accurate results for the final segmentation. The fusion network generates the final output based on the results of the global and local segmentation. The experiments were conducted on the private iris-290 dataset and a public CAMUS dataset with 1800 images. The results show that the method has better accuracy compared to U-Net, CE-Net and U-Net++.

[Keywords]: Deep learning; Medical image segmentation; Global and local segmentation; Attention mechanisms

目录

1. 背景介绍.....	6
1.1 虹膜分割任务.....	6
1.2 深度学习方法.....	7
1.3 主要工作.....	8
2. 方法设计.....	8
2.1 确定关键点.....	8
2.2 模型框架结构.....	9
2.2.1 全局分割与关键点定位网络.....	10
2.2.2 局部分割网络.....	10
2.2.3 融合网络.....	10
2.3 训练过程.....	11
3. 实验结果.....	11
3.1 数据集.....	11
3.2 实验设计.....	11
3.3 实验结果分析.....	12
4. 结论.....	14
参考文献.....	15
致谢.....	17

1. 背景介绍

1.1 虹膜分割任务

近年来，随着深度学习相关方法的发展，基于深度学习的医学图像分割任务在辅助医生临床诊断中有了更多的使用场景。医学图像分割任务的目的是，将医学图像中的目标组织或器官区域与背景区域区分开来。医学图像如核磁共振成像（MRI）、计算机断层扫描图像（CT）、光学相干断层成像（OCT）以及超声图像（US）等往往会在局部的小区域出现成像模糊^[1]。使得医学图像分割任务往往难度较大。

本研究の対象为眼前节光学相干断层成像（AS-OCT）中的虹膜分割任务。AS-OCT 图像如图 1 所示，图像是通过接收光源发出的红外光利用低相干光干涉的原理在参考平面与眼前节组织之间的反射信号的差异而得到的纵向二维图像^[2]。图像中包含了角膜、虹膜、睫状体以及晶状体的结构的纵向信息。图 1 中红色框中的结构为虹膜。由于虹膜为圆形环状结构，因此其纵向截面成像被分为了左右两个部分。然而在 AS-OCT 图像中，虹膜的上下表面往往都存在着较为明显的边界区分。对比之下，虹膜两侧尾端邻角膜、睫状肌，成像中存在模糊区域。同时，由于该部分面积较小，在使用深度学习方法进行下采样过程中容易丢失关键信息。这使得虹膜的准确分割颇具挑战。



图 1 眼前节光学相干断层成像（AS-OCT）

1.2 深度学习方法

在过去的研究中,基于边缘检测、统计形状模型以及谱聚类等传统的机器学习算法为图像分割任务的主要研究方向^[3]。随着深度学习相关的研究越来越深入,结合深度学习的相关方法在图像处理相关问题中的研究也逐渐增多,深度学习在图像分类等任务中的效果也初显成效。而图像分割问题也可以转化为像素级别的分类问题,因此将深度学习方法引入到图像分割任务中也是大势所趋。在 2014 年,全卷积网络^[4](FCN)结合了卷积神经网络相关方法,在图像分割任务中得到了应用。为了实现语义分割任务,全卷积网络中使用了反卷积来对高维度的特征图进行上采样,使得最终网络可以输出与输入图像相同的尺寸的预测图,以此来实现像素点级别的分割。全卷积网络是第一个将深度学习应用到图像分割中的方法。

然而,全卷积网络的分割精准度并不理想,尤其是在难度更大的医学图像中。在 2015 年,U-Net^[5]的提出使得深度学习方法在医学图像分割任务中的效果大幅提升。U-Net 使用了对称的下采样与上采样结构,并使用跳跃连接(skip connection)来将深层的特征与浅层特征相结合。使得网络能够同时关注到全局的整体信息与局部的纹理信息^[5]。同时,更多的上采样过程能够使网络得到更加精细的输出结果。在 U-Net 提出之后,很多的研究如 U-Net++^[6]、Attention U-Net^[7]以及 CE-Net^[8]等,都是在 U-Net 的基础结构之上,通过改变卷积层的结构、连接方式或加入空间注意力机制、通道注意力机制等方法来进一步提升分割效果。

注意力机制在基于深度学习的图像处理中有着重要的作用。注意力机制相关方法的目的是希望在模型的训练中加强对重要信息的关注与学习并且忽略无关或不重要的信息,使得模型能更快收敛同时拥有更好的性能^[9]。注意力机制的提出是受到人类的视觉系统的特点启发。当人类在理解视觉信息时,往往仅需关注关键的特征就能做出相应的判断。而在图像分割任务中,使模型增强关键特征的理解与利用在很多场景下能得到更好的性能表现。在无时序的图像分割任务中,常见的注意力方法可大致分为通道域注意力、空间域注意力以及通道与空间混合注意力三类。通道域注意力方法通过对应的注意力模块在通道维度对模型进行自适应性加权^[9],典型的方法有 SENet^[10]、EncNet^[11]以及 GCT^[12]等。而空间域注意力方法则是加强数据在空间维度的搜索与自适应加权,典型的方法有 STN^[13]、

DCN^[14]以及基于自注意力机制的 Transformer^[15]类方法。而通道与空间混合注意力机制集两种不同的注意力机制思想为一体,在通道与空间两个维度中加入注意力机制,典型的方法有 Residual Attention^[16]等。

1.3 主要工作

在虹膜的分割任务中,数据集中所有图像都有着较为相似的结构分布。通过统计 U-Net 模型的分割结果,可以得出虹膜分割在空间上的易错区域。受到注意力机制思想的启发,本研究设计了兼顾全局与局部的深度学习模型框架。在确定该任务的易错区域特征的情况下,可以通过训练定位模型来找到每个样本的局部易错区域。以此来对虹膜任务中的高难度区域进行针对性关注,从而提升最终的虹膜分割的精准度。研究的主要工作为:

- (1) 通过分析统计 U-Net 在虹膜分割任务中的结果,找到了在该任务中的易错区域。
- (2) 设计了兼顾全局与局部的深度学习模型框架。
- (3) 针对该模型框架设计了对比实验,并对实验结果进行了分析
- (4) 对该模型框架的适用性与局限性进行了探讨。

2. 方法设计

2.1 确定关键点

通过分析在虹膜分割数据集上训练的 U-Net 模型的结果,可以统计出模型预测在空间上的错误分布。图 2 为 U-Net 模型在虹膜分割任务中的错误分布热力图。由于不同图像中的虹膜的大小与位置有着一定差异,在统计每张图像的错误分布时对虹膜部分进行了对齐与归一化操作。在热力图中区域的颜色越亮,代表该区域的为 U-Net 输出结果错误越普遍的区域,也就是易错区域。从热力图中可以看出,无论是左边还是右边部分的虹膜,除开虹膜边界不可避免的细小误差以外,最明显的高亮区域分布在虹膜靠近睫状体一侧的尾端。因此本研究中将虹膜的两侧尾端作为图像中的关键区域。

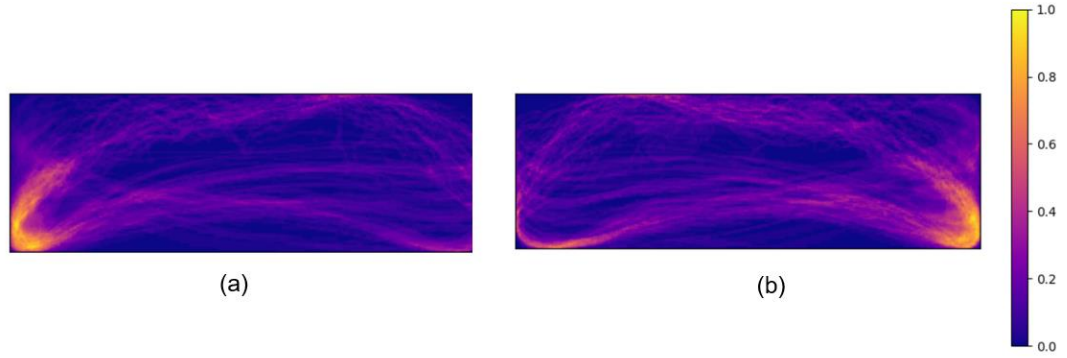


图2 U-Net 分割错误区域热力图：(a) 左侧虹膜 (b) 右侧虹膜

2.2 模型框架结构

根据 U-Net 模型对虹膜分割任务的结果的分析，总结得到该任务有着以下几个特点：

- (1) AS-OCT 图像中虹膜都具有相似的形状与位置特征。
- (2) 虹膜的上下表面都存在比较明显的边界。
- (3) U-Net 分割模型在对虹膜两侧与角膜、睫状肌相近的区域错误率偏高。

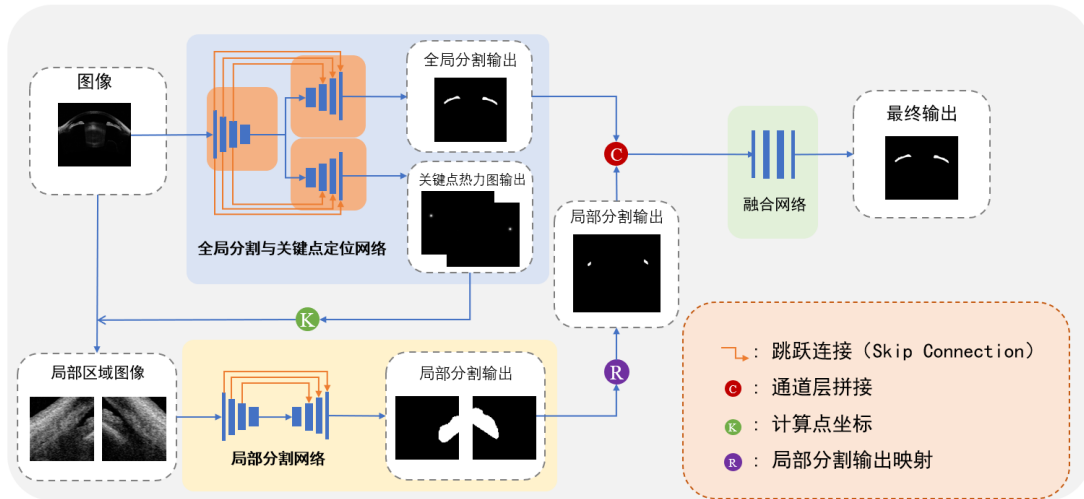


图3 模型设计框架图

基于该任务的以上特点，模型的框架设计如图3所示。分为全局分割与关键点定位网络、局部分割网络与融合网络三个部分。全局分割与关键点定位网络以原始图像作为输入，并输出全局的分割的预测与关键点的预测热力图。局部分割网络通过关键点定位网络输出得到关键点坐标，并以此坐标为基准裁切出局部图像作为输入。局部分割网络输出局部预测。融合网络以全局分割预测结果与局部

分割预测结果作为输入，输出最终的预测结果。其中的关键点为人工先验知识所得到 AS-OCT 图像的虹膜分割任务中的易错区域定位点。

2.2.1 全局分割与关键点定位网络

虽然虹膜在 AS-OCT 图像中的空间分布较为稳定的，但是在不同的图像中虹膜的位置还是有所差异。通过关键点定位网络可以得到图像中的虹膜的关键点的坐标，从而针对该区域进行更加精细的分割。关键点定位网络可以使用与分割网络结构相似的网络实现，使用人工标注的关键点生成的高斯热力图作为标签训练。受到多任务学习^[17]思想的启发，由于全局分割任务与关键点定位任务之间有着一定的相关性，通过在全局分割网络与关键点定位网络共享编码器，可以使得两个任务之间相互辅助来得到更好的效果。

全局分割与关键点定位网络由一个共享的编码器，与两个对应不同任务的解码器构成。编码器与两个解码器之间使用与 U-Net 相同的跳跃连接。对应全局分割任务的解码器输出使用全局分割的标签进行监督学习，对应关键点定位的网络的解码器输出使用关键点为中心的高斯热力图进行监督学习。由于数据集中包含了全局分割的标注信息，需要生成对应的关键点高斯热力图。热力图参照人工标注的关键点生成。

2.2.2 局部分割网络

局部分割部分采用了 U-Net 作为分割网络。局部分割网络的输入为以关键点坐标为中心裁切的关键区域。关键点坐标通过对关键点定位网络的输出计算最大值得到。关键区域的裁切大小为固定尺寸。通过分析数据集，将裁切的固定尺寸设置为 (100, 100)。同样裁切尺寸也使用在关键区域标注的裁切上。

2.2.3 融合网络

融合网络为整个模型框架最终的输出模块。在得到了全局分割网络与局部分割网络的输出后，需要将两者的结果进行结合得到最终的输出。然而，直接将两者的结果进行相加并不是一个合理的处理方法。因为这样会导致全局分割网络在关键区域不够精细的结果继续保留在最终的输出中。因此，融合网络采用了简单的三层卷积层来融合全局与局部分割的特征。

融合网络的输入为全局分割结果与局部分割结构拼接而成的两通道图片。由于局部分割网络的输出为裁切后的固定尺寸，因此在与全局分割结果拼接之前，还需要将局部分割的结果根据关键点坐标映射到空白背景上，使其拥有和全局分割结果相同的尺寸。融合网络输出通过全局分割标签进行监督学习。

2.3 训练过程

由于局部网络的训练过程依赖较为精准的关键点坐标信息，因此整个模型框架采用分阶段训练的方式。第一阶段训练全局分割与关键点定位网络，使用全局分割标签与关键点高斯热力图监督学习。当该部分的网络收敛之后，将该部分的参数冻结。第二阶段训练局部分割网络，使用上一阶段的关键点输出坐标对原图与标签进行裁切。第三阶段将第一阶段训练的全局分割与关键点定位网络的参数解冻，通过融合网络对整体的网络进行微调。

3. 实验结果

3.1 数据集

实验使用了私有的 Iris-290 数据集与公开的 CAMUS^[18]数据集。两个数据集都是对图像中形态与结构较为稳定的目标进行分割，且都存在小范围的模糊区域。Iris 数据集是由香港中文大学 (CUHK)、东京大学 (TU)、加州大学旧金山分校 (UCSF) 和中山眼科中心 (ZOC) 四个不同的数据中心从 2017 年到 2019 年收集的 AS-OCT 图像数据集。其中包含了 185 人在光亮与暗照的条件下的单眼 OCT 成像。每个单眼都包含了 18 张均匀分布在 360 度上的 B 型扫描 (B-scan) 成像。Iris-290 由在所有图片中挑选出的 290 张具有代表性的图片组成。CAMUS 为 450 个患者的心动图心尖四腔与两腔视图数据集，其中包含了多个分割任务。实验中选取了左心室心内膜分割任务 (LV_{Endo}) 进行了对比。在实验过程中，两个数据集的训练集与测试集的比例都是 8 : 2。

3.2 实验设计

为了证明模型框架的可靠性，实验选取了在医学图像分割中常用的 U-Net、CE-Net 和 U-Net++ 作为对比，在两个数据集上进行了实验。同时实验还对比了全

局分割网络与整体模型框架的输出结果。所有的代码都基于 Pytorch 框架实现，每一组实验都以初始值为 0.00005 的余弦退火学习率策略(Cosine Annealing LR)训练了 500 个轮回。训练中使用了二分类交叉熵损失函数 (BCE Loss) 作为约束。最终采用 Dice 指标和交并比 (IOU) 作为评价指标来量化模型的性能。表 1 为模型框架中的具体参数。

表 1 全局与局部分割网络模型各层参数

层数	SE	GSD	KPLD	L-Encoder	L-Decoder	融合网络
1	conv3-16	up-conv-128	up-conv-128	conv3-16	up-conv-128	conv3-16
2	maxpool	——	——	maxpool	——	——
3	conv3-32	up-conv-64	up-conv-64	conv3-32	up-conv-64	conv3-16
4	maxpool	——	——	maxpool	——	——
5	conv3-64	up-conv-32	up-conv-32	conv3-64	up-conv-32	conv1-1
6	maxpool	——	——	maxpool	——	——
7	conv3-128	up-conv-16	up-conv-16	conv3-128	up-conv-16	——
8	maxpool	——	——	maxpool	——	——
9	conv3-256	conv1-1	conv1-K	conv3-256	conv1-1	——
10	——	sigmoid	sigmoid	——	sigmoid	sigmoid

注：SE (Share Encoder) 为全局分割与关键点定位网络共享编码器，GSD (Global Segmentation Decoder) 为全局分割网络解码器，KPLD (Key Point Localization Decoder) 为关键点定位解码器，L (Local patch network) 为局部分割网络。

3.3 实验结果分析

表 2 模型在 Iris-290 与 CAMUS 数据集上的对比实验结果

模型	Iris-290		CAMUS	
	Dice	IOU	Dice	IOU
U-Net	0.9463	0.8981	0.8947	0.8118
CE-Net	0.9513	0.9073	0.9025	0.8246
U-Net++	0.9572	0.9180	0.9000	0.8203
GLS	0.9626	0.9280	0.9185	0.8511

从表2中可以看出,在 Iris-290 数据集中,U-Net模型 Dice 指标达到了 0.9463。CE-Net 与 U-Net++的模型结果率略优于 U-Net, Dice 指标分别为0.9513 和 0.9572。而全局与局部的分割模型 Dice 指标达到了 0.9626。在 CAMUS 数据集中, U-Net 模型 Dice 指标为 0.8974。CE-Net 与 U-Net++的模型结果较为相近,分别为 0.9025 和 0.9000。全局与局部的分割模型在该数据集上的指标则是达到了 0.9185。从实验结果可以得出,在两个不同的数据集上,全局与局部的分割模型的性能都要优于其它三个模型。

表 3 多任务全局分割网络对比实验结果

模型	Iris-290		CAMUS	
	Dice	IOU	Dice	IOU
U-Net	0.9463	0.8981	0.8947	0.8118
GS	0.9518	0.9074	0.9059	0.8280
GLS	0.9626	0.9280	0.9185	0.8511

从表3中可以看出,全局分割与关键点定位的全局分割模型在 Iris-290 数据集上 Dice 指标为 0.9518,在 CAMUS 数据集上的 Dice 指标为 0.9059。两个数据集上的结果都要优于 U-Net,同时要低于融合了局部分割结果的最终输出。可以看出,多任务的全局分割与关键点定位能够强化全局的分割效果,而局部分割结果的补充又能更进一步优化最终的输出结果。

从图4可视化结果可以看出,在 Iris-290 数据集中,全局与局部分割网络的分割结果在整体上更加接近分割标签。同时得益于局部分割网络的针对性分割,在虹膜与睫状肌区域相接的尾端有着更好的效果。在 CAMUS 数据集中,全局与局部分割网络在三个顶点的关键区域中的分割效果也优于其他三个网络。

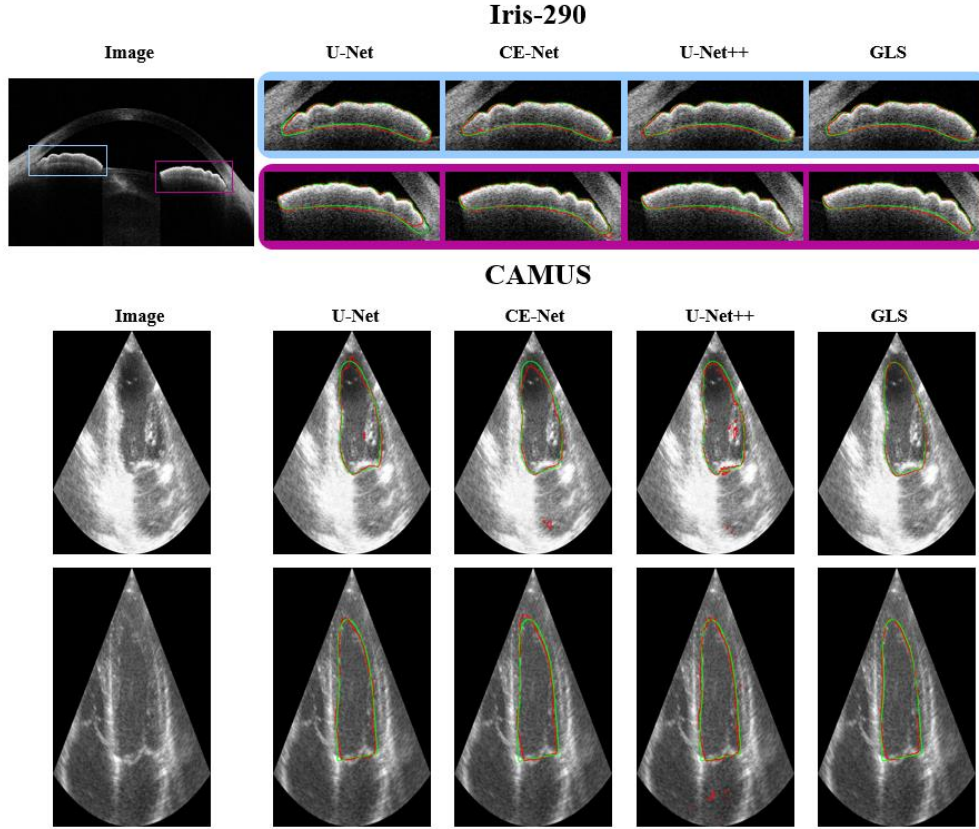


图 4 可视化分割结果图。绿色线为标注边界，红色线为预测边界

4. 总结

本研究讨论了 AS-OCT 图像中虹膜分割任务的难点。并针对 AS-OCT 图像的虹膜分割任务中的关键区域，设计了全局与局部的分割模型框架。框架由多任务的全局分割与关键点定位网络、负责局部精细分割的局部分割网络与最终的融合网络组成。同时通过分析对应的对比实验得出，该框架在 Iris-290 和 CAMUS 两个数据集上相较于 U-Net、CE-Net 以及 U-Net++ 有一定的提升。由于该框架中的局部分割网络仅能对一类相同的区域进行分割，目前该框架仅适用于图像结构稳定且有着相同的小区域易错区域的分割任务。当数据集中不存在明显的共性时，该框架便不再适用。在未来的工作中，可以尝试使用更加通用的关键点定位与分割网络，使该框架能够应用在更加复杂的医学图像分割任务中。

参考文献

- [1] Karimi, Davood, et al. “Accurate and Robust Deep Learning-Based Segmentation of the Prostate Clinical Target Volume in Ultrasound Images.”. *Medical Image Analysis*, July 2019, pp. 186–96.
- [2] 黄智宇,胡毅成,周传清等.扫频光学相干断层成像技术及其在眼科的应用[J].*眼科学报*,2021,36(01):55-65.
- [3] Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., & Nandi, A. K. . Medical image segmentation using deep learning: A survey. *IET Image Processing*,2022, 16(5), 1243–1267.
- [4] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. 2015. “Fully Convolutional Networks for Semantic Segmentation.” In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.
- [6] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Lecture Notes in Computer Science* (pp. 3–11).
- [7] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. 2018. Attention u-net: Learning where to look for the pancreas.
- [8] Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu. 2019. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE transactions on medical imaging* 38, 10 (2019), 2281–2292.
- [9] Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., ... Hu, S.-M. (2022). Attention Mechanisms in Computer Vision: A Survey. *Computational Visual Media*, 331–368.
- [10] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019: 2011-2023.

- [11] H. Zhang, K. J. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, and A. Agrawal, "Context encoding for semantic segmentation," in 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018. IEEE Computer Society, 2018, pp. 7151–7160.
- [12] Z. Yang, L. Zhu, Y. Wu, and Y. Yang, "Gated channel transformation for visual recognition," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11 794–11 803.
- [13] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial Transformer Networks. Computer Vision and Pattern Recognition. 2015.
- [14] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," 2017.
- [15] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, AidanN., ... Polosukhin, I. (2017). Attention is All you Need.
- [16] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 3156–3164
- [17] RUDER S. An Overview of Multi-Task Learning in Deep Neural Networks[Z]//arXiv: Learning. 2017.
- [18] S. Leclerc, E. Smistad, J. Pedrosa, A. Ostvik, et al." Deep Learning for Segmentation using an Open Large-Scale Dataset in 2D Echocardiography" in IEEE Transactions on Medical Imaging, vol. 38, no. 9, pp. 2198-2210, Sept. 2019.

致谢

在最后，我要感谢我的导师刘江教授给我在本科生阶段的生活以及学习上提供的指导与帮助。同时我也要感谢 risa 老师以及杨冰师兄对我的论文的指导。加入 iMed 团队近两年的时间里，我学习到了如何用课本中的知识去解决实际问题。通过创新实践的项目，我锻炼了与团队合作的能力。当然最珍贵的经历，就是让作为本科生的我们参与到实验室相关的科研项目中。在未来我会谨记刘老师对我们的教导，也不会忘记实验室的各位老师、师兄以及师姐给我帮助。我要感谢 iMed 团队里的各位伙伴，在 iMed 时间两年时间一定会是我本科生阶段最宝贵的一段回忆。