



## 基于深度学习的医学影像分割研究综述

曹玉红<sup>1</sup>, 徐海<sup>2</sup>, 刘荪傲<sup>2</sup>, 王紫霄<sup>2</sup>, 李宏亮<sup>3\*</sup>

(1. 中国电子学会, 北京 100036; 2. 中国科学技术大学 信息科学技术学院, 合肥 230026;

3. 中国科学院大学 工程科学学院, 北京 100049)

(\* 通信作者电子邮箱 lihongliang@ucas.ac.cn)

**摘要:**医学影像分割是计算机辅助诊断中的一项基础且关键的任务, 目的在于从像素级别准确识别出目标器官、组织或病变区域。不同于自然场景下的图像, 医学影像往往纹理复杂, 同时受限于成像技术和成像设备, 医学影像噪声大, 边界模糊而不易判断。除此之外, 对医学影像进行标注极大依赖于医疗专家的认知和经验, 因此可用于训练中的标注数据少且存在标注误差。由于上述的医学影像边缘模糊不清、训练数据较少和标注误差较大等特点, 基于传统图像分割算法搭建的辅助诊断系统难以满足临床应用的要求。近年来随着卷积神经网络(CNN)在计算机视觉和自然语言处理领域的广泛应用, 基于深度学习的医学影像分割算法取得了极大的成功。首先概述了近几年基于深度学习的医学影像分割的研究进展, 包括这些医学影像分割算法的基本结构、目标函数和优化方法。随后针对医学影像标注数据有限的问题, 对目前半监督条件下医学影像分割的主流工作进行了整理归纳和分析。此外, 还介绍了针对标注误差进行不确定性分析的相关工作。最后, 总结分析了深度学习医学影像分割的特点并展望了未来的研究趋势。

**关键词:**医学影像分割; 深度学习; 卷积神经网络; 半监督学习; 不确定性估计

**中图分类号:**TP391.7 **文献标志码:**A

### Review of deep learning-based medical image segmentation

CAO Yuhong<sup>1</sup>, XU Hai<sup>2</sup>, LIU Sun'ao<sup>2</sup>, WANG Zixiao<sup>2</sup>, LI Hongliang<sup>3\*</sup>

(1. Chinese Institute of Electronics, Beijing 100036, China;

2. School of Information Science and Technology, University of Science and Technology of China, Hefei Anhui 230026, China;

3. School of Engineering Sciences, University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** As a fundamental and key task in computer-aided diagnosis, medical image segmentation aims to accurately recognize the target regions such as organs, tissues and lesions at pixel level. Different from natural images, medical images show high complexity in texture and have the boundaries difficult to judge caused by ambiguity, which is the fault of much noise due to the limitations of the imaging technology and equipment. Furthermore, annotating medical images highly depends on expertise and experience of the experts, thereby leading to limited available annotations in the training and potential annotation errors. For medical images suffer from ambiguous boundary, limited annotated data and large errors in the annotations, which makes it is a great challenge for the auxiliary diagnosis systems based on traditional image segmentation algorithms to meet the demands of clinical applications. Recently, with the wide application of Convolutional Neural Network (CNN) in computer vision and natural language processing, deep learning-based medical segmentation algorithms have achieved tremendous success. Firstly the latest research progresses of deep learning-based medical image segmentation were summarized, including the basic architecture, loss function, and optimization method of the medical image segmentation algorithms. Then, for the limitation of medical image annotated data, the mainstream semi-supervised researches on medical image segmentation were summed up and analyzed. Besides, the studies related to measuring uncertainty of the annotation errors were introduced. Finally, the characteristics summary and analysis as well as the potential future trends of medical image segmentation were listed.

**Key words:** medical image segmentation; deep learning; Convolutional Neural Network (CNN); semi-supervised learning; uncertainty estimation

收稿日期:2020-10-21;修回日期:2021-01-13;录用日期:2021-01-19。

基金项目:国家自然科学基金资助项目(62022076, U19A2057, 61976008)。

**作者简介:**曹玉红(1968—),女,河北平山人,高级工程师,博士,主要研究方向:计算机视觉、人工智能;徐海(1993—),男,江西乐平人,博士研究生,主要研究方向:计算机视觉、人工智能、语义分割;刘荪傲(1997—),男,山东青岛人,硕士研究生,主要研究方向:计算机视觉、人工智能、语义分割;王紫霄(1999—),男,河北石家庄人,硕士研究生,主要研究方向:计算机视觉、人工智能、半监督学习;李宏亮(1987—),男,河南安阳人,工程师,博士,主要研究方向:计算机视觉、计算成像。



## 0 引言

随着医学影像成像技术和成像设备的快速发展和普及,全球每天产生大量的医学影像数据,借助计算机进行医学影像分析在临床诊断、手术方案制定中的重要性日益凸显<sup>[1]</sup>。其中,医学影像分割能够有效地提取目标区域的形状和空间信息,是进行医学影像定量分析的关键步骤之一<sup>[2]</sup>,目的是以机器视觉方式自动从医学图像中逐像素地识别出目标区域(器官、组织或病灶)。早期的医学影像分割系统主要基于传统的图像分割算法搭建,如基于边缘检测的分割算法<sup>[3]</sup>、基于阈值的分割算法<sup>[4]</sup>和基于区域的分割算法<sup>[5]</sup>。但医学图像通常具有对比度低、组织纹理复杂、边界区域模糊等特点,极大地限制了此类图像分割算法的效果和应用场景。随后,针对特定任务设计手工特征的分割算法很长一段时间成为了医学影像分割的研究主流<sup>[6]</sup>,然而手工特征的设计极大依赖医生的专业先验知识,而且往往泛化能力差,无法迁移到新的任务场景下。因此在实际应用中基于传统图像分割技术的医学影像分割系统仍然不够成熟,无法获得令人满意的分割效果。

近年来,随着计算机技术和人工智能的快速发展,卷积神经网络(Convolutional Neural Network, CNN)<sup>[7-9]</sup>强大的建模能力被广泛研究,相比传统的算法,基于卷积神经网络的深度学习算法在图像处理各领域带来了突破性的进展,如图像分类<sup>[10]</sup>、语义分割<sup>[11]</sup>等,基于深度学习的图像分割算法也被引入到医学影像分割<sup>[12-15]</sup>中。深度学习算法的自动提取特征能力有效地克服了传统医学图像分割算法过多依赖医疗专家先验认知这一弊端,且深度学习算法的可移植性高,借助迁移学习能够快速地拓展到不同的任务场景下。

尽管深度学习在图像分割中取得了突破性的提升,医学影像区别于自然场景图像的特点决定了医学影像分割需要面临更大的挑战,主要来源于以下几个方面:

### 1) 医学影像边缘模糊不清。

受限于成像技术,医学影像相较于自然图像往往有对比度低、噪声较大的特点,同时医疗影像中组织纹理复杂,边界模糊不易区分。此时如何提升网络模型的抗干扰能力和鲁棒性,以及对边界附近像素的准确性是一个非常巨大的挑战。

### 2) 医学影像标注数据少。

医学影像数据获取困难(尤其对于罕见疾病),同时图像分割任务训练过程中需要对图像每个像素的类别进行标注,而且医学影像标注对医疗专业知识依赖性强,因此,获取足够的标注样本是极度耗时耗力的。如何在有限的标注样本下,减轻训练分割模型时对像素级标注的依赖,是医学影像分割的另一挑战。

### 3) 医学影像标注误差大。

医学影像病变形态学上高度异质化,使得标注过程极大依赖于医疗专家的认知和经验,而考虑到标注医生主观标准上的不确定性和不同专家客观上的认知差异化,标注过程中漏标、误标不可避免,标注的准确度并不完全可靠。如何在有限的医疗标注资源下,对模型不确定性的准确量化,是当前面临的又一挑战。

综上所述,深度学习在医学影像分割中具有广阔的应用前景,但同时也面临巨大的挑战。

## 1 相关工作

随着深度学习的崛起,研究人员将应用于自然图像的分割算法<sup>[11,16-17]</sup>引入到医学领域。其中最具代表性的研究工作是全卷积网络(Fully Convolutional Network, FCN)<sup>[11]</sup>,FCN实现了不改变图像尺寸的情况下对分割网络进行端到端(End to End)的训练,并较传统方法取得了显著性的提升。伴随着FCN的成功,研究人员开始关注如何针对医疗影像的特点对分割网络进行改进,考虑到医疗图像具有丰富的空间信息(如复杂的纹理结构),而网络下采样过程容易丢失空间信息,基于编码-解码(Encoder-Decoder)的网络结构开始崭露头角。中国科学院慈溪医工所团队<sup>[12]</sup>结合具有对称结构的编解码网络对视网膜血管进行了精细化分割,并基于分割结果量化分析了健康人群视网膜和阿兹海默症患者之间的差异。实际上医学影像数据大多数为3D的容积数据(如CT(Computed Tomography)、MRI(Magnetic Resonance Imaging)数据),为了保留不同层间的位置关系,Cicek等<sup>[18]</sup>通过将二维卷积层替换为三维卷积层构建了3D U-Net,实现了3D数据的端到端处理。随着基础模型的完善,人们开始更多地考虑如何优化分割的效果,如引入注意力机制来优化特征,以达到减小类内差异同时增大类间差异的目的。中国科学技术大学Xie等<sup>[15]</sup>根据肿瘤位置关系提出级联的注意力分割网络,有效提高了脑胶质瘤区域分割精度。此外,研究人员尝试从目标函数、增大感受野、解决类别不平衡等多种角度对分割模型进行优化。

尽管深度学习神经网络相比传统算法表现出了显著的进展,但在实际应用中医学图像标注过程耗时耗力,限制了深度学习算法在该领域的进一步发展。相对地直接获取大量的医学影像数据较为容易,因此为了减轻对标注的依赖、降低成本,半监督学习算法得到了广泛的关注和研究。半监督医学影像分割的核心是如何利用未标注的数据,基于自训练(Self-training)和协同训练(Co-training)的算法是此领域最常见的半监督分割算法,此类方法通过为未标注数据生成伪标签(Pseudo Label)并优化更新方式进行迭代训练。半监督学习中为了能够使用少量标注数据训练出更加鲁棒的模型,提出了对未标注数据添加扰动并对预测一致性进行约束的方法<sup>[19-20]</sup>,如基于均值教师(Mean Teacher, MT)的半监督方法<sup>[20]</sup>和基于几何变换一致性的方法<sup>[21]</sup>。此外,研究人员开始考虑更多样的利用未标注数据的方式,如基于图(Graph)进行正则化<sup>[22-23]</sup>,基于生成对抗网络(Generative Adversarial Network, GAN)<sup>[24]</sup>来生成更多的可用于训练的数据也是提升半监督分割效果的方法之一。

由于标注医生主观标准上的不确定性和不同专家客观上的认知差异化,标注的准确度并不完全可靠,因此对医学影像分割中的预测结果给出定量的不确定性度量是辅助诊断的重要补充,近期关于医学影像分割的不确定性也开始引起新的研究热潮。根据不确定性的分布类型角度,Swiler等<sup>[25]</sup>将其分为认知不确定性(Epistemic uncertainty)和随机不确定性(Aleatoric uncertainty)。认知不确定性是指模型认知上的不确定性,研究者根据对模型不确定性评估的方式不同,将其大致分为两大类,即深度模型集成(Deep model ensemble)<sup>[26]</sup>和深度贝叶斯网络(Deep Bayesian Neural Network)<sup>[27]</sup>。随机不确定性指的是观测中固有的噪声,这部分不确定性来源于医疗设备成像的数据本身噪声以及标注存在的不可控误差。



## 2 全监督医学图像分割

近年来,卷积神经网络<sup>[7-8]</sup>已经成为处理图像分割任务的主流方法,并被广泛拓展到医学图像分割当中。卷积网络能够通过学习特定的卷积核提取丰富的图像特征,从而生成有效、准确的分割结果。受限于计算资源,卷积网络通常由多个小尺寸的卷积层堆叠而成,并在此过程中进行下采样操作以减小图像的空间尺寸,从而逐步扩大卷积核的感受野,实现由浅到深、由局部到整体的多级特征提取。

全监督学习是医学影像分割任务最基本、应用最广泛的方法。全监督的语义分割要求提供像素级的标注作为训练参考,对于训练数据量以及标注具有较高的要求。尽管医学影像数获取困难,数据集构建成本高,但是为了满足医学领域的巨大需求,目前已经出现了许多公开的医学图像数据集,保证了全监督医学图像分割研究的充分发展。

### 2.1 网络结构

医学图像在数据结构上与自然图像类似,同时,医学图像也存在与自然图像明显不同的特性,如空间尺寸、目标大小、成像质量等。基于这些特性,研究者对自然图像的网络结构进行改进,构建更适用于医学领域的模型。总体来说,目前用于医学分割的网络都沿用了编码器-解码器的对称结构,并在此基础上强化图像特征的提取。本节将首先介绍编码器-解码器的一系列经典网络结构,随后介绍改进模块,如注意力机制与新型卷积等,最后将介绍针对特定任务使用的模型级联策略。

#### 2.1.1 编码器-解码器结构

与图像分类任务不同,分割任务要求生成与输入图像尺寸一致的像素级分割结果,因此无法直接将分类任务的网络结构应用于分割任务。全卷积网络<sup>[11]</sup>通过将全连接层替换为卷积层,实现了基于卷积神经网络的图像分割。由于在特征提取过程中存在下采样操作,在生成分割结果时需要通过插值计算进行上采样。在此基础上,Ronneberger等<sup>[14]</sup>提出了用于细胞分割的U-Net,这一结构随后被广泛应用于各种医学图像分割任务中。U-Net包括用于特征提取的编码器,以及与之对称、用于恢复空间分辨率并生成分割结果的解码器,具体结构如图1所示。由于网络整体形状类似于字母U,故被称作U-Net。U-Net的编码器部分通过堆叠 $3 \times 3$ 卷积层与激活函数实现特征提取,并通过 $2 \times 2$ 最大池化层降低分辨率,每次将空间尺寸减半并加倍通道数。在解码器部分,使用 $2 \times 2$ 的转置卷积恢复空间分辨率,并通过跳跃连接(Skip Connection)将上采样后的特征与编码器部分同层的特征进行级联(concatenation),作为后续卷积层的输入。

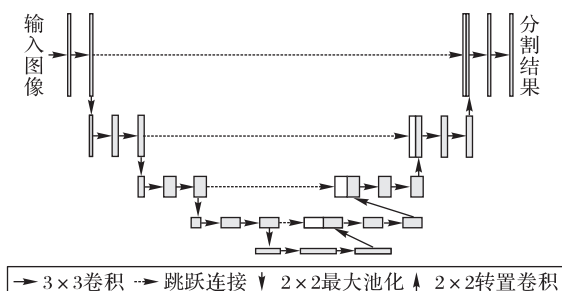


图1 U-Net的网络结构

Fig. 1 Network structure of U-Net

U-Net最初被设计用于2D图像的细胞分割,而很多医学图像数据实际为3D的容积数据。尽管可以将容积数据拆分为

2D图像序列进行处理,但这种方式忽视了不同层间的位置关系,并且往往不同位置的图像差别较大,不利于网络学习通用特征。因此Cicek等<sup>[18]</sup>通过将二维卷积层替换为三维卷积层构建了3D U-Net,实现了3D数据的端到端处理。结合深度学习领域的相关研究,Milletari等<sup>[13]</sup>提出了V-Net以更好地处理容积数据。相比于3D U-Net,V-Net的改进包括:1)使用更有效的激活函数PReLU(Parametric Rectified Linear Unit);2)使用步长为2的 $2 \times 2$ 卷积代替最大池化(Max Pooling)实现下采样;3)在卷积层引入了残差连接以提升学习效果。

U-Net通过跳跃连接实现了不同层级的特征融合,提高了分割精度。Zhou等<sup>[28]</sup>进一步对多层特征的融合方式进行改进,提出了U-Net++,结构如图2所示。U-Net++将U-Net中简单的跳跃连接替换为卷积层,并且在同分辨率下的不同卷积层、相邻分辨率下的卷积层间添加跳跃连接,从而形成密集连接以强化特征融合。为了保证网络的充分学习,U-Net++还添加了多个中继监督层同时计算损失函数。得益于此,U-Net++可以根据算力情况通过剪枝减小模型规模,而性能仅有小幅度下降。

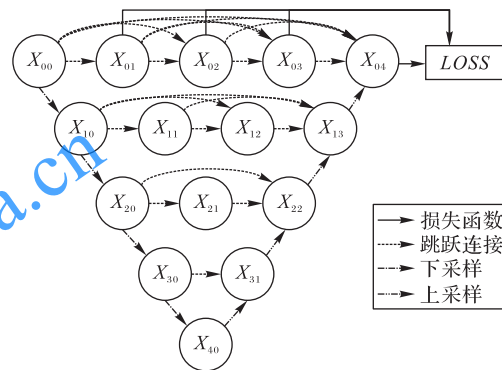


图2 U-Net++的网络结构

Fig. 2 Network structure of U-Net++

#### 2.1.2 注意力机制

随着对神经网络研究的不断深入,注意力机制逐渐得到了广泛的应用,这一机制在人类视觉系统中同样至关重要。概括地说,注意力机制通过计算注意力权重,对特征进行重新加权,以达到强化有效特征、抑制无效特征的目的。根据应用位置的不同,可以分为空间注意力与通道注意力。通道注意力的典型代表为SENet(Squeeze-and-Excitation Network)<sup>[29]</sup>。SENet提出了压缩-激发(Squeeze-and-Excitation, SE)模块对对不同通道的特征进行加权,如图3所示。该模块通过全局平均池化(Average Pooling)将尺寸为 $C \times H \times W$ 的输入特征压缩为 $C \times 1 \times 1$ ,再通过全连接层计算通道注意力权重,与输入通道相乘得到加权后的权重。此模块的优点在于计算量小且即插即用,因此常被应用在医学任务中作为对U-Net的改进<sup>[30]</sup>,取得了较好的效果。

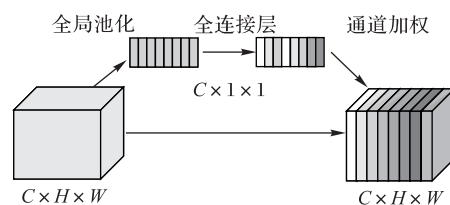


图3 SE模块结构

Fig. 3 Structure of SE module



空间注意力则以 Non-local<sup>[31]</sup>的一系列工作为代表。区别于通道注意力,空间注意力给每个像素计算注意力图以实现全图范围的特征提取,从而有效地弥补了卷积操作因感受野有限导致的全局特征提取能力的不足。标准的 Non-local 空间注意力计算流程如图 4 所示。给定输入特征  $X = \{x_1, x_2, \dots, x_{HW}\}$ , 首先计算像素间的特征相似度:

$$f(x_i, x_j) = \sigma(\theta(x_i)^T \phi(x_j)) \quad (1)$$

其中  $\theta(\cdot), \phi(\cdot)$  为线性变换, 由  $1 \times 1$  卷积实现。  $\sigma(\cdot)$  为 softmax 函数, 用于将相似度归一化:

$$\sigma(\theta(x_i)^T \phi(x_j)) = \frac{\exp(\theta(x_i)^T \phi(x_j))}{\sum_l \exp(\theta(x_i)^T \phi(x_l))} \quad (2)$$

至此,可以得到每个像素的全图注意力图。随后进行特征重加权:

$$y_i = \sum_j f(x_i, x_j) g(x_j) \quad (3)$$

在图 4 的模块中,还额外加入了残差连接以保证训练过程的稳定性。空间注意力优越的长距离特征提取能力同样也可应用在医学图像分割当中,例如 He 等<sup>[32]</sup>表明,引入空间注意力可以有效地提升医学图像分割网络对于对抗攻击的鲁棒性。

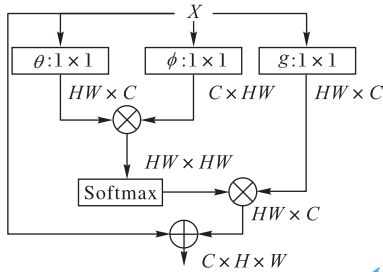


图 4 Non-local 模块结构

Fig. 4 Structure of Non-local module

注意力机制的核心是通过计算注意力图实现特征的重加权,遵循这一思想,可以根据特定的医学图像分割任务设计与上文不同的注意力计算模块。以脑胶质瘤的分割为例,在脑部肿瘤分割挑战赛(BraTS)<sup>[33]</sup>数据集中,胶质瘤被划分为三个等级:全肿瘤(Whole Tumor, WT)、肿瘤核心(Tumor Core, TC)和增强肿瘤(Enhancing Tumor, ET),而三者存在包含关系,即  $ET \subset TC \subset WT$ 。因此,可以通过依次分割三个等级的肿瘤实现由粗到精的精细分割。

OMNet(One-pass Multi-task Network)<sup>[34]</sup>将这种逐级分割的思想引入通道注意力中,根据前一级肿瘤的分割情况调整通道重要性,用于强化下一级肿瘤的分割效果。而 DCAN(Deep Cascaded Attention Network)<sup>[15]</sup>则以此改进空间注意力,根据前一级的分割结果对背景区域的像素进行抑制,使下一级肿瘤的分割更集中在前一级的分割区域。

### 2.1.3 改进卷积计算

标准卷积的问题在于感受野有限且固定,导致其无法有效地提取全局信息。为了增大感受野,需要堆叠多层卷积层并通过下采样操作降低空间分辨率。然而,这种操作仍存在局限性,因此出现了许多对卷积运算的改进工作,例如空洞卷积<sup>[35]</sup>与可变形卷积(图 5)<sup>[36]</sup>。空洞卷积的优势在于可以在不进行下采样、不增加参数数量的前提下扩大卷积运算的感受野,

从而可以在更高的分辨率下进行特征提取,避免因下采样造成的空间信息损失,而将空洞卷积整合到 U-Net 的编码器结构中已经被证明对医学图像分割同样具有提升效果<sup>[37-38]</sup>。

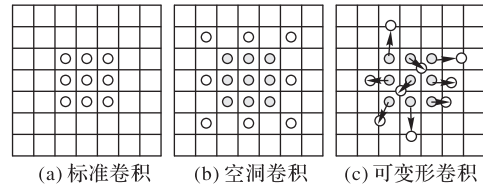


图 5 标准卷积、空洞卷积和可变形卷积示意图

Fig. 5 Schematic diagram of standard convolution, dilated convolution and deformable convolution

空洞卷积仅是在标准卷积的基础上增加空洞以扩大计算范围,因此与后者同样是计算位置固定的卷积操作。然而对于不同的像素,模型希望卷积核能够根据像素之间的相关性自适应地选择计算位置,从而实现更有效的特征提取。为了实现这一目标,可变形卷积通过额外的偏移预测分支,为输入特征的每个像素计算卷积计算时的偏移量,使特征提取更集中、高效。这一运算同样可以应用于医学图像分割中,例如 Guo 等<sup>[39]</sup>提出了使用可变形卷积进行多模态器官分割,并通过在偏移预测中引入全局信息进一步强化了可变形卷积的特征提取能力。

如前文所述,与自然图像不同,相当一部分医学图像(如磁共振影像)实际上为三维容积数据。尽管可以使用三维卷积网络直接计算,但相较于二维网络,三维网络的参数量呈指数级增加,限制了其推广应用。而如果使用二维网络计算,则会完全忽略一个维度的信息,影响分割效果。为了缓解这一问题,WNNet(Whole tumor Network)<sup>[37]</sup>提出使用二维卷积提取平面信息,并随后使用一维卷积提取第三个维度的信息。同时如图 6 所示,医学三维影像对于三个维度的切面具有明确的定义,即冠状面(Coronal)、矢状面(Sagittal)和横断面(Axial),每个切面能够显示的医学信息有所不同。由于不对称的卷积结构对三个维度的提取能力不同,WNNet 提出多视角训练策略,即将三维数据以三个方向输入网络分别训练,但也导致了计算时间的加倍。区别于 WNNet, MFNet(Multi-direction Fusion Network)<sup>[38]</sup>在将三维卷积拆分为伪三维卷积<sup>[40]</sup>的基础上提出了多方向融合模块,如图 7 所示。该模块使用三支并行的计算分支,每个分支从不同方向将  $3 \times 3 \times 3$  卷积拆分为  $3 \times 3 \times 1$  与  $1 \times 3 \times 3$  卷积。相较于 WNNet,该方法同时从三个方向提取特征并进行融合,避免了多次训练与推理的额外计算开销。

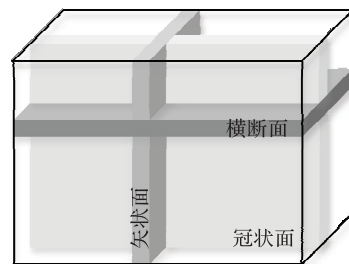


图 6 医学影像的切面划分

Fig. 6 Section division of medical image

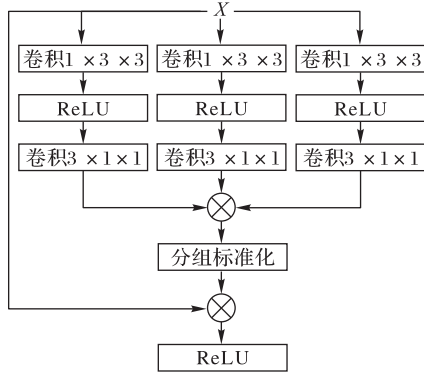


图 7 多方向融合模块结构  
Fig. 7 Structure of multi-directional fusion module

### 2.1.4 模型级联

对于脑胶质瘤分割一类的医学分割任务,由于存在由粗到精的分割过程,除了使用一个模型完成一次性分割,另一种经典而有效的处理方式是将多个模型级联起来,每个模型分别完成一个分割子任务,并根据分割结果为下一个任务提供范围更小的感兴趣区域,图 8 展示了级联模型分割的基本流程。例如,Wang 等<sup>[37]</sup>使用三个模型进行胶质瘤的分割,第一个模型预测全肿瘤,根据预测结果计算包围全肿瘤的矩形框,在输入容积数据上将该部分裁剪出来,送入第二个模型预测肿瘤核心。最后,根据预测的肿瘤核心使用模型 3 分割增强肿瘤。

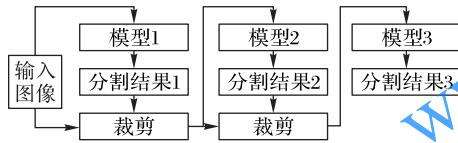


图 8 级联模型的基本流程  
Fig. 8 Basic flow of cascade model

与单模型分割相比,级联模型可以根据预测结果逐步缩小感兴趣区域,从而减少过度分割的情况。但由于使用多个模型,受算力限制,每个模型的规模往往无法与单模型相同。此外,由于后续分割直接依赖于前一级的分割结果,因此对分割准确性有很高的要求。为了保证后续分割的效果,级联模型通常采取分步训练的策略,以保证在增加更精细分割任务时能够提供较好的粗分割结果。虽然随着对卷积神经网络的研究,不断有更有效的单模型分割方法被提出,但基于简单模型的级联方法仍表现出十分出色的效果,例如Jiang 等<sup>[41]</sup>通过两个 U-Net 的级联模型赢得了 2019 年脑肿瘤分割挑战的第一名。因此,对于追求准确性与实用性的医学影像分割来说,级联模型是与单模型同样值得关注的方法。

### 2.2 损失函数

在全监督学习中,损失函数直接决定了网络的训练目标。对于图像分割任务而言,最常用的损失函数为交叉熵损失,这一损失被广泛应用于自然图像分割任务中。而医学图像相较于自然图像又存在其独特性,主要在于前景与背景类别的严重不平衡。因此,许多工作着眼于损失函数的改进,以提高分割模型在医学图像上的性能。此外,针对特定的医学场景,多任务学习也经常受到关注。本节将分别对目前常用的损失函数进行介绍。

#### 2.2.1 交叉熵损失

交叉熵(Cross Entropy)损失是图像分割任务中应用最广泛的损失函数,并同时适用于二分类和多分类任务。在医学图像分割中,任务往往定义为二分类任务,即将像素划分为前景(正例)与背景(负例)区域。用于二分类任务的交叉熵损失可以写为:

$$L_{ce} = -y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i) \quad (4)$$

其中: $p_i$ 为网络预测第  $i$  个样本为前景的概率, $y_i$ 为标注图中对应样本的标签,前景为 1,背景为 0。交叉熵损失均衡地考虑了全部像素的影响,而分割任务的难点在对边界部分的准确分割。为此,U-Net<sup>[14]</sup>提出为交叉熵计算增加权重,以强化对特定像素的学习。权重的大小受像素与分割边界的距离控制,更靠近边界的像素具有更高的权重。类似地,Guo 等<sup>[39]</sup>提出根据距离变换计算像素级权重图,同样可以加强对于边界部分的分割效果。

在标准的交叉熵损失中,正样本和负样本对损失函数具有平等的影响权重。然而对于医学图像分割任务,前景类别如目标器官、病变区域往往仅占整个图像的一小部分,意味着前景像素与背景像素的数量存在严重的不平衡;同时,大量背景像素可以被很简单地分割出来,导致训练时存在大量的简单负样本,严重影响了模型的学习效果。对于这类任务,一个可行的选择是使用 Focal Loss<sup>[42]</sup>取代交叉熵:

$$L_{focal} = -y_i(1 - p_i)^\gamma \log(p_i) + (1 - y_i)p_i^\gamma \log(1 - p_i) \quad (5)$$

相比交叉熵损失,Focal Loss 增加了权重调节项  $(1 - p_i)^\gamma$  与  $p_i^\gamma$ ,其中  $\gamma$  是指数形式的权重因子,依据正确预测的概率对样本进行指数加权。如果网络对于像素属于前景或背景的预测概率接近 1,权重调节项则会接近 0,从而自适应地降低了简单样本的权重,保证了网络在训练过程中更关注于对难样本的学习。

#### 2.2.2 Dice 损失

在评估医学图像分割任务的性能时,Dice 系数为一个常用的指标:

$$Dice = \frac{2|P \cap Y|}{|P| + |Y|} \quad (6)$$

式(6)中, $P$  表示预测结果, $Y$  表示标注图。在评估 Dice 系数时,通常只关注前景的分割结果,因此对于二分类任务,更常用的 Dice 系数计算公式为:

$$Dice = \frac{2 \sum_i p_i y_i}{\sum_i p_i + \sum_i y_i} \quad (7)$$

其中: $p_i \in (0, 1)$  为前景预测概率, $y_i \in \{0, 1\}$  为二值标签。V-Net<sup>[13]</sup>提出了基于 Dice 系数的 Dice Loss:

$$L_{dice} = 1 - \frac{2 \sum_i p_i y_i + 1}{\sum_i p_i + \sum_i y_i + 1} \quad (8)$$

式(8)中的拉普拉斯平滑项(即分子分母同时加+1)避免了分母为 0 的情况,同时也定义了预测结果与标注图均不存在前景标签时的 Dice 系数为 1。相比交叉熵损失,Dice Loss 直接基于分割的评价指标对网络进行优化,同时避免了前景与背景像素数量不均衡的问题。Dice Loss 的局限性在于只适用于二分类情况。Sudre 等<sup>[43]</sup>提出了广义 Dice Loss,将其扩展到了多类别,并统计各类别标签数以增大标签少的类别的权



重,从而实现不同类别的平衡。

### 2.2.3 多任务损失

为了增强分割网络的特征提取能力,研究者在设计损失函数时,除了最终的分割损失之外,还可以根据任务特点设计额外的预测分支以组成多任务损失。Ren等<sup>[44]</sup>设计了用于医学图像分割的多级任务分解,除了分割任务之外,还增加了类别和场景预测任务分支。网络在生成像素级分割结果的同时,预测整张图像中存在的目标类别种类,以及更高层级的任务类型。此外, Ren等还设计了一种同步正则化以加强不同任务之间的联系,最终达到提升分割精度的效果。Guo等<sup>[45]</sup>同样设计了类别级别的预测任务,但将类别存在性的预测精度提高到网络下采样后的分辨率,不同于之前全图级别的预测。

### 2.3 精度优化

尽管关于改进网络模型的工作不断出现,但以U-Net为代表的经典网络仍然具有相当的竞争力,在众多医学图像分割比赛中具有重要地位。例如,2019年脑胶质瘤分割比赛的第一名使用的是两个级联的U-Net结构<sup>[41]</sup>。而nnU-Net<sup>[46]</sup>则在使用U-Net结构的基础上使用了更有效的训练设置,在多个医学分割比赛中名列前茅。因此,对于医学图像分割而言,训练设置同样是至关重要的一部分。本节旨在介绍一些不依赖于模型结构的通用训练技巧,以提升最终的分割性能。

#### 2.3.1 数据增广

为了避免过拟合,同时增强网络对于各种变化的鲁棒性,训练数据增广是模型训练不可缺少的操作。由于医学图像及标注获取的困难性,医学数据集规模往往远小于自然图像数据集,因此更容易出现过拟合现象。常用的训练数据增广方法包括随机缩放、随机裁剪、随机旋转、随机翻转、随机噪声等。更进一步的复杂增广方法则包括空间与灰度变换,如弹性形变<sup>[14]</sup>、B样条插值<sup>[13]</sup>、伽马校正<sup>[46]</sup>等。为了减少读取开销,数据增广通常是在训练过程中实时进行的,在实际使用中可根据数据规模和算力情况灵活选择。

除了训练数据增广之外,测试时同样经常进行数据增广以强化分割效果。测试数据增广通常包括多尺度缩放以及镜像翻转<sup>[41,46]</sup>,并将多种增广后的预测结果取平均值作为最终预测结果。与单尺度预测相比,增广预测通常表现出更精确、更稳定的分割性能。

#### 2.3.2 模块优化

随着深度学习研究的发展,不断有更有用的通用网络模块被提出,并可以整合到U-Net的编码器-解码器结构中。例如,在卷积层与激活函数之间加入批标准化层(Batch Normalization)<sup>[47]</sup>,可以使网络收敛速度更快、鲁棒性更好、效果更出色。然而批标准化的性能直接受批尺寸影响,在批尺寸很小时效果不理想。对于医学影像分割中常见的三维卷积网络,由于其本身计算开销较大,批尺寸通常严重受限(往往为1或2),此时引入批标准化并不合适。对于这类网络,使用计算不依赖于批尺寸的标准化方法,例如分组标准化<sup>[48]</sup>、样本标准化<sup>[49]</sup>和层标准化<sup>[50]</sup>,往往可以达到更好的效果。图9给出了四种标准化的计算方式示意图。其中,分组标准化的分组数为超参数,可以根据实际训练情况进行调整。当分组数为1时,分组标准化变为层标准化;当分组数等于通道数时,等价于样本标准化。

除了标准化层,对激活函数的改进同样值得关注。标准

的ReLU(Rectified Linear Unit)激活函数仅在输入大于0时保留激活值,而完全忽略了输入为负值的情况。作为改进,LeakyReLU<sup>[51]</sup>在ReLU的基础上为负值区域保留了较小的固定斜率,避免了完全失活的情况。PReLU<sup>[52]</sup>将负值区域的固定斜率改为可学习的参数,进一步地强化了激活函数的表示能力。

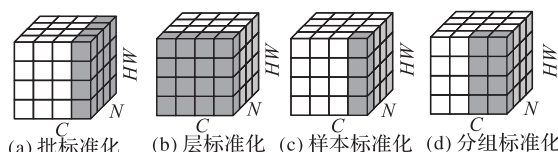


图9 标准化方法示意图

Fig. 9 Schematic diagram of normalization

#### 2.3.3 模型融合

多模型融合是医学图像分割比赛中的常用技巧,由于训练的随机性,单个模型容易陷入局部最优解,而整合多个模型的预测结果通常可以提高整体分割效果,增强分割的鲁棒性。多模型融合的方式可以是:1)对训练数据进行多折划分,多次训练同一个模型<sup>[53-54]</sup>;2)选用多种模型,分别进行训练<sup>[46,55]</sup>。类似于测试数据增广,最终结果由多个模型的预测平均得到(图10)。

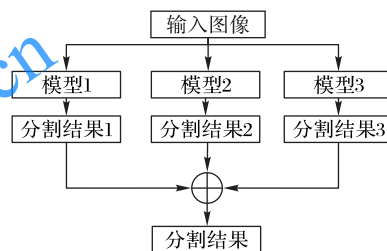


图10 模型融合的基本流程

Fig. 10 Basic flow of model fusion

#### 2.3.4 后处理

在得到网络生成的分割图后,还可以通过后处理方法进一步对分割结果进行细化,例如使用条件随机场(Conditional Random Field, CRF)<sup>[35]</sup>来平滑分割图的边界,去除分割噪声。针对特定的医学任务,还可以根据先验知识设计后处理方法,以弥补网络分割的不足。阈值化<sup>[34,41]</sup>作为一种较为常见的后处理手段,目的是去除灰度不满足阈值的像素,或者去除体积小于阈值的连通区域。OMNet<sup>[34]</sup>对于脑肿瘤分割任务进一步提出了基于体素灰度的聚类方法,以减少对于增强肿瘤的误分类情况。

## 3 半监督医学图像分割

尽管深度学习相比传统算法表现出了显著的进展,但其在训练过程中需要大量的标注数据作为支撑。在实际应用中医学图像语义复杂且常包含3D信息,标注过程耗时耗力,限制了深度学习算法在该领域的进一步发展。相对地直接获取大量的医学影像数据较为容易,因此为了减轻对标注的依赖、降低成本,半监督学习算法得到了广泛的关注和研究。

半监督学习除了使用少量数据  $X_L = (x_l)_{l \in [1, N]}$  和对应的标注  $Y_L = (y_l)_{l \in [1, N]}$  外,还引入了大量的未标记数据  $X_U = (x_u)_{u \in [N+1, M]}$  辅助训练,在研究中通常将已有数据集的部分



标签丢弃来模拟该情况。半监督学习在应用时的一个必要条件是数据的分布  $p(\mathbf{x})$  包含后验分布  $p(\mathbf{y}|\mathbf{x})$  的相关信息,这在多数情况下都是成立的,但是在训练前无法得知两者间的关系,因此如何有效地从中提取出关于后验分布的信息是半监督学习方法的关键。目前的方法通常遵循三个基本假设来描述  $p(\mathbf{x})$  与  $p(\mathbf{y}|\mathbf{x})$  的关系:平滑假设(smoothness assumption)、低密度假设(low-density assumption)和流形假设(manifold assumption)。平滑假设认为两个在输入空间中相近的数据点应有相似的标签,低密度假设认为分类时的决策边界应尽可能地穿过数据稀疏的区域,也称为聚类假设,流形假设认为在同一低维流形中的数据点应有相同的标签。

本章将介绍目前医学分割领域中各类半监督算法中的代表性工作。

### 3.1 基于自训练和协同训练的算法

自训练算法和协同训练算法均通过流形假设来利用已标记数据传播信息生成伪标签并进行迭代优化,已有很多研究将此思想应用于医学影像分割,文献[54, 56-57]等方法采用自训练的分割算法,这些算法仅使用单一模型完成训练过程。相对地,文献[58-60]等方法使用的协同训练算法利用两个或以上的模型共同完成训练优化。

自训练算法是最常见的半监督学习算法之一,它使用单一的模型,通过为无标记数据预测伪标签,进而在学习伪标签并重新预测更新的迭代过程中增强网络的泛化能力。以  $L_S$ 、 $L_U$  表示常用损失函数(如交叉熵),  $\hat{\mathbf{y}}_i$  表示伪标签,则此方法训练时的优化目标可表示如下:

$$L = \sum_{i=1}^N L_S(f(\mathbf{x}_i), \mathbf{y}_i) + \sum_{i=N+1}^{N+M} L_U(f(\mathbf{x}_i), \hat{\mathbf{y}}_i) \quad (9)$$

从优化方式可以看出,此类方法的缺陷是需要依赖于生成伪标签的质量,当网络学习到错误的标记后可能会不断将其放大从而影响最终性能。由于自训练算法仅参考了网络本身提供的信息,预测结果中的信息量有限且通常伴随着部分误判,特别是对于语义歧义性高、边缘模糊的医学影像,生成伪标签的质量并不稳定,因此目前对自训练算法的研究主要集中在如何在嘈杂的伪标签中进行学习。

一个改进的思路是对分割结果进行后处理以精炼提升伪标签的质量,Bai等<sup>[56]</sup>将条件随机场的后处理方法与自训练算法结合并应用于心室MRI的分割任务中。该方法首先学习已标记数据,然后对于未标记数据进行分割,之后使用了CRF来精炼分割结果并使用优化后的分割图来指导下一轮的迭代,最终有效地提升了分割的质量。相似地,Tang等<sup>[57]</sup>则使用了水平集(level set)的方法来作为后处理精炼伪标签。另外Rajchl等<sup>[54]</sup>也基于自训练的方法并额外使用了边框级的弱标注辅助监督过程。

自训练算法通过网络本身的预测来分配标签,可以看作运用流形假设将学到的标签传播至相似的数据上,从而学习了所有数据在其特征空间上的分布特点,并且在优化损失函数(如交叉熵)的同时隐式地使决策边界远离高密度数据区域,根据低密度假设最终学到了更加合理决策边界,进而提升了网络的鲁棒性。

协同训练算法将自训练算法进行了扩展,为了降低单一模型预测带来的局限性提出使用多个预训练的模型以综合预测伪标签,通过模型间的融合来提升伪标签的质量。需要注

意的是,协同训练需要使不同的模型在预训练过程中相互独立以提取不同的知识,实现时通常需要将数据集进行额外的划分保证子集间存在差异性或利用同一数据的不同视图,这样在随后的训练阶段就可以通过在未标记数据上的预测来传播每个模型学到的知识达到相互补充的效果,最终得到更加鲁棒的网络。

Zhou等<sup>[60]</sup>基于协同训练的方法定义了额外的学生模型来学习融合后的伪标签。为了获取独立的子数据集,Zhou等利用器官分割中3D医学影像数据可以分解为不同的轴向视图(矢状面、冠状面和轴向)的特点,在不同的轴上对3D数据进行切片构造子数据集并使用2D分割网络进行预训练得到3个教师模型。融合阶段通过“投票”的方式选择伪标签,对于预测一致的像素直接保留结果,而对不一致的部分则取置信度得分最高的标签作为伪标记。最后使用一个新的学生网络在扩充后的数据集上进行训练。

另一种常见的协同训练方式没有使用学生模型,而是使用了相互指导的学习策略,即每个模型使用其他模型融合得到的伪标签进行训练,从而直接学习互补的知识。在此基础上为了进一步过滤噪声数据,Xia等<sup>[59]</sup>提出了基于不确定性的融合生成策略,通过添加Dropout利用贝叶斯深度网络估计预测的不确定性,进而在融合阶段以加权的方式生成更可信的伪标签。

Peng等<sup>[58]</sup>使用多个模型预测的均值作为伪标签,同时为了使模型学习到更多互补的知识,引入了对抗样本以捕捉不同模型间的差异。此方法额外定义了差异损失函数,针对每个模型  $f_i$  对输入  $\mathbf{x}$  进行调整生成对应的对抗样本  $g_i(\mathbf{x})$ ,如图11,其中无标签的对抗样本由虚拟对抗训练(Virtual Adversarial Training, VAT)生成,有标签的对抗样本则使用快速梯度法(Fast Gradient Sign Method, FGSM)生成,进而在其他模型的指导下优化使其对于对抗样本更加鲁棒。

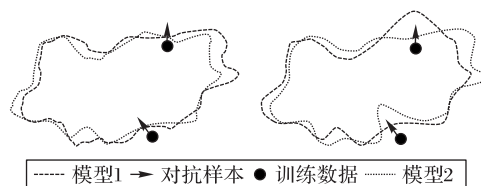


图 11 对抗样本示意图

Fig. 11 Schematic diagram of adversarial samples

还有一些方法对学习伪标签的过程进行了调整,通过引入额外的约束以提高伪标签的利用效率,Kervadec等<sup>[61]</sup>针对伪标签不可靠的问题提出了课程半监督学习(curriculum semi-supervised learning),此方法通过学习更加宽松的区域表达来提升网络的泛化性能。具体地,课程半监督学习框架定义了一个辅助分类网络预测输入图像中前景部分区域的大小  $R$ ,进而在网络分割无标记数据时统计输出结果的前景区域大小并将其限制在  $R$  的附近  $(1 - \lambda R, 1 + \lambda R)$ ,优化时将超出的部分作为正则惩罚项加入到损失函数中,从而避免了利用错误的像素级预测作为伪标签带来的影响。最终通过左心室分割任务展现了其算法的优势。

使用带噪的伪标签容易造成模型退化而约束后的伪标签又无法提供足够的信息量,为了平衡两者间的矛盾,Min等<sup>[62]</sup>定义了深度注意力网络(Deep Attention Network, DAN)以自适应地发现 and 纠正噪声标签中错误的信息,并且提出了分级



蒸馏的方法生成更加可靠的伪标签,最终在多个医学分割任务上有效地提升了网络的性能。整个框架的训练过程分为三步,首先使用 DAN 在有标记数据下进行预训练,然后通过分级蒸馏的方式为无标记数据生成伪标签,最后使用所有的数据和标签重新训练模型。其中 DAN 模型在训练时使用两个学生网络同时学习相同的数据,并根据模型间的预测和内部特征的关联筛选出可靠的梯度部分执行反向传播,使其对错误标签拥有一定的纠正能力。此外在生成伪标签时,融合了数据蒸馏与模型蒸馏的特点,通过将模型蒸馏中每个模型的预测替换为每个模型在多种数据变换下的预测将两种方式分层次地结合起来,如图 12,从而进一步提升伪标签的质量。

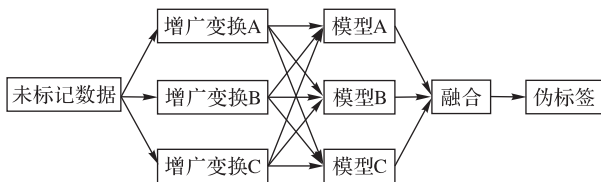


图 12 分级蒸馏示意图

Fig. 12 Schematic diagram of hierarchical distillation

### 3.2 基于一致性正则的算法

根据平滑假设,对数据进行扰动后应该得到一致的输出结果,然而通常训练得到的卷积神经网络无法保证这种变换不变性,从而泛化性能较差。半监督学习中为了能够使用少量标注数据训练出更加鲁棒的模型,提出了对数据扰动前后的一致性进行约束的方法,实现上常通过定义额外的子任务提取对应的不变性以辅助优化网络。一些代表性的研究方法包括使用均值教师的半监督方法,如 MT<sup>[19]</sup>、UAMT (Uncertainty Aware Mean Teacher)<sup>[20]</sup>。还有基于几何变换一致性的方法,包括 TCSM (Transformation Consistent Self-ensembling Model)<sup>[21]</sup>、semiTC (semi-supervised Transformation-Consistent network)<sup>[63]</sup>,以及两种方法的结合 TCSMv2<sup>[64]</sup>等。

在文献 [65] 中  $\Pi$ -Model 和 Temporal Ensembling 的启发下,均值教师算法<sup>[66]</sup>对两者的思想进行了融合,Perone 等<sup>[19]</sup>基于此方法在脊髓灰质分割任务上进行了实验,整体的训练框架如图 13。首先在初始化时定义了相同结构的教师模型  $f_t$  和学生模型  $f_s$ ,其中教师模型仅通过学生模型每次迭代参数的指数滑动平均 (Exponential Moving Average, EMA) 更新以融合不同时期的训练成果,泛化能力更强。训练时对于同一数据在添加不同噪声  $\eta, \eta'$  后分别让教师和学生模型进行预测,将两者分割结果的均方差作为辅助损失优化学生模型,此一致性损失既包含了与时序融合后模型预测的一致性,又含有不同噪声扰动下的不变性,最终整体损失函数如下:

$$L = \sum_{i=1}^N L_S (f_s(x_i), y_i) + \sum_{i=1}^{M+N} L_U (f_s(x_i, \eta), f_t(x_i, \eta')) \quad (10)$$

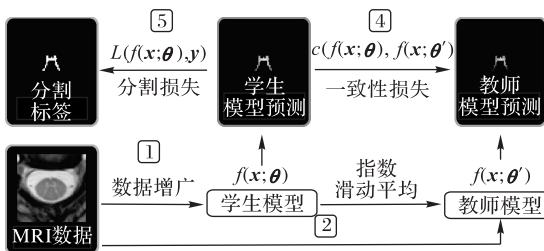


图 13 均值教师分割算法

Fig. 13 Mean teacher segmentation method

均值教师算法可以看作利用扰动不变性的同时融入了伪标签的思想,由于进行了时序上的融合,教师模型的预测更加稳定并可以作为标签指导学生模型的更新方向。

Yu 等<sup>[20]</sup>从不确定性的角度对教师模型的预测进行了筛选,增加了蒙特卡罗 Dropout (Monte Carlo Dropout) 用于衡量教师模型预测的不确定度,进而根据阈值选取低不确定度的部分计算一致性损失,最终模型的精度在左心室分割任务中相比原始均值教师方法得到了进一步的提升。

Li 等<sup>[21]</sup>则从旋转不变性的角度利用未标记数据,提出了 TCSM 用于皮肤病变分割,主要思想为对同一图像经过旋转变换后分割结果也应具有同样的变换。在训练过程中网络对同一数据预测两次,其中分别在预测前后执行同样的变换  $T$  (如旋转、翻转等),之后优化结果间的均方误差  $\|T(f(x)) - f(T(x))\|^2$ 。更进一步,为了提升模型在未标记数据上的预测质量,TCSM 在融合了均值教师方法后提出了 TCSMv2<sup>[64]</sup>,改进后的算法在计算不变性损失时引入了教师模型获取更加可靠指导。

类似地,Bortsova 等<sup>[63]</sup>使用弹性形变前后的一致性设计一致性损失。对于同一输入数据随机执行弹性形变  $t_1, t_2$  后分别预测得到  $y_1, y_2$ ,随后调整第一次输出的结果与第二次相匹配:  $y_1' = (t_2 \circ t_1^{-1})y_1$ ,最后在使用均方误差约束二者间的形变一致性  $\|y_2 - y_1'\|^2$ 。

此外还有一类思路使用重建的方法,即约束从编码器输出的特征中还原的图像应与真实的图像相似,进而强化编码器的特征提取能力。Chen 等<sup>[67]</sup>利用重建的方法构建了多任务注意力机制半监督学习 (Multi-task Attention-based Semi-Supervised Learning, MASSL) 框架辅助训练,总体框架如图 14。具体来说在经编码器得到深层特征后,除了执行分割任务外,定义了重建解码器预测前景和背景部分的输入图像,再与二值分割结果相乘后和真实的前背景图计算均方误差。

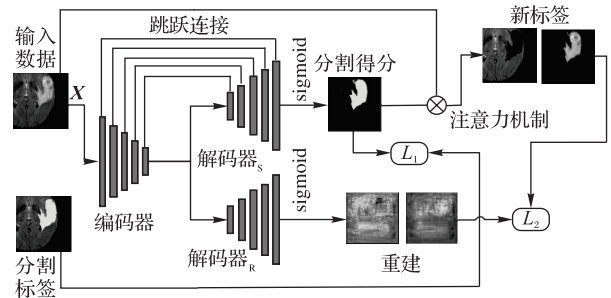


图 14 MASSL 网络框架

Fig. 14 Network structure of MASSL

### 3.3 基于图的算法

基于图的算法在特征空间中的数据点上建立加权无向图  $G = (V, E)$ ,其中  $V$  表示数据点,图结构中的边用于描述样本之间局部相似性,相连的样本相似度高,因而根据流形假设信息可以沿着图的边进行传播,最后将图上所有数据点划分到不相交的子集中完成分类过程。其中的代表性算法包括 Baur 等<sup>[68]</sup>提出的随机特征嵌入的半监督学习算法,以及使用图正则化的<sup>[22-23]</sup>等方法。

嵌入半监督学习 (semi-supervised embedding) 算法<sup>[69]</sup>通过减小相似数据距离、增大无关数据距离的方式利用图中数据的分布进行优化,但对于像素级的分割任务此方法计算开支





较大,因此 Baur 等<sup>[68]</sup>对算法进行了调整,在多发性硬化病变分割任务中提出了随机特征嵌入(Random Feature Embedding, RFE)的思想,针对图像中大量的像素进行了采样,只使用部分像素参与计算,从而能够在像素级分割结果上进行优化。

图正则化的方法使用图平滑(Graph-smooth Regularization)的思想来标记额外的数据,其中使用图拉普拉斯算子衡量节点间的相似性,并作为正则器优化图的平滑性。

在脑部 MRI 肿瘤分割中, Song 等<sup>[22]</sup>提出了一种基于图正则化的归纳学习方法,使用隐变量来生成最终预测:  $x \rightarrow t \rightarrow y$  并基于高斯随机场(Gaussian Random Field, GRF)对潜在变量  $t$  进行建模,之后使用图拉普拉斯算子衡量节点间的相似性并作为正则器对其进行优化。

Xu 等<sup>[23]</sup>将其与深度神经网络结合辅助神经元分割,通过计算边的权重  $w$  来量化描述相邻数据的相似性进而通过边传播标签至相邻的数据。实现时首先将影像数据分解为多个超像素  $v$ , 然后使用  $\epsilon$ -最近邻图( $\epsilon$ -nearest neighboring graph)来描述所有超像素的关系,图中边的权重  $w$  使用 RBF 核(Radial Basis Function kernel)来计算,即  $w_{i,j} = \exp(-\|v_i - v_j\|_2^2)$  并忽略小于  $\epsilon$  的边。形成图后利用均方差损失来约束图中相连的数据标签相似:  $R = \sum w_{i,j} \|y_i - y_j\|_2^2$ 。在得到区域级的预测后,进一步使用全局关联和形状先验的方式取得细化的分割结果。

### 3.4 基于生成对抗网络的算法

生成对抗网络(Generative Adversarial Network, GAN)<sup>[70]</sup>是一种基于对抗的学习生成模型算法,包含生成器(generator)与判别器(discriminator),其中生成器用于数据的生成,判别器用于结果的评估。训练时判别器学习如何将生成器产生的数据与真实数据区分开,而生成器学习如何产生可以迷惑判别器的数据。在对抗中两个模型的能力均能够得到强化提升。由于在相互对抗的过程中不需要数据本身的标签,GAN 在半监督学习中得到了大量的应用与改进,基于对抗过程设计思路的不同包括 Chaitanya 等<sup>[71]</sup>和 Mondal 等<sup>[72]</sup>的生成数据的方法, Zhang 等<sup>[73]</sup>、Nie 等<sup>[74]</sup>和 Zhou 等<sup>[75]</sup>的评估分割结果的算法,以及 Ross 等<sup>[76]</sup>定义额外对抗任务训练特征提取能力的方法。

缺少数据是需要进行半监督学习的主要原因,而 GAN 的生成器本身就具有生成数据的能力。Chaitanya 等<sup>[71]</sup>从该角度出发,提出将 GAN 中的生成器用于合成虚假影像与标签以缓解数据不足的困难。对于生成器  $G$ , 输入标记数据  $X_L$  和随机生成的向量  $z$ , 输出变形场以扭曲输入图像得到新的数据  $X_G$ 。另外定义了判别器  $D$  用于区分生成数据  $X_G$  与真实数据  $X_L \cup X_U$ , 对抗训练时提升生成器在分类器上的得分  $L_G = \log(1 - D(G(X_L, z)))$ , 分类器损失  $L_D = \log(D(X_L \cup X_U)) - \log(1 - D(G(X_L, z)))$ , 对抗学习后将新生成的数据加入分割网络  $S$  的训练中。具体地,研究了两种数据生成方式:变形场生成器和加性强度场生成器,如图 15, 变形场生成器通过产生变形场  $v$  同时扭曲输入图像与标签进行增广,而加性强度场生成器输出强度信号  $\Delta I$  通过与输入图像相加并保留标签实现数据增广。在心脏 MRI 分割数据集上的实验结果验证了 GAN 可以作为一种强大的数据增广方式扩充缓解数据不足的问题。

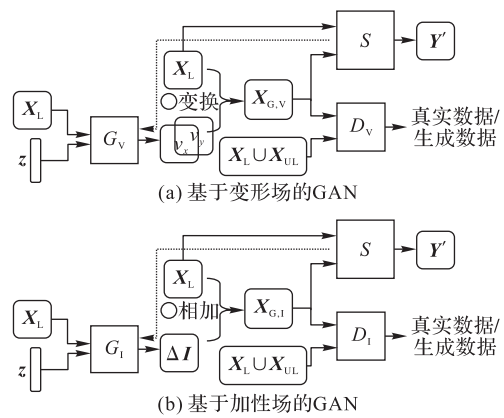


图 15 基于 GAN 的数据增广方式

Fig. 15 Data augmentation methods based on GAN

Mondal 等<sup>[72]</sup>同样使用对抗网络来生成虚假的数据辅助分割,不同于使用二分类判别器的评估方式,此方法将分割网络作为判别器引入,对于  $K$  类分割任务,将虚假数据定义为新的  $K+1$  类,在对抗学习过程中,分割网络希望能将虚假数据分类至第  $K+1$  类,同时将无标记数据分入原先的  $K$  类中,生成器  $G$  则相反。为了保证训练的稳定性,使用特征匹配的方式训练生成器,定义  $L_G = \|E(f(x)) - E(f(G(z)))\|^2$  使其在特征空间上与常规数据相似,其中  $f(x)$  表示分类器内部层的输出。在脑部 MRI 数据集上的实验结果表明此方法可以使用极少的标记数据获取较高的性能。

另一类方法结合了伪标签的思路,将分割网络作为生成器来产生分割图,进而将分类器(如 ResNet<sup>[7]</sup>)作为对抗网络中的判别器用于评估分割网络预测的伪标签质量,从而监督分割网络生成更真实的预测结果。

其中代表性的算法是 Zhang 等<sup>[73]</sup>提出的深度对抗网络(Deep Adversarial Network, DAN)框架, DAN 将对抗网络应用于腺体分割与真菌分割任务,首先在有标记数据上预训练分割网络,在加入无标记数据后定义了判别网络来评价分割网络的预测质量,使其在训练过程中判断分割结果是否来源于训练过的有标记数据,最后固定训练好的判别器,鼓励分割网络欺骗评价网络,使其对所有数据的分割结果都判定为有标记数据,以此促使分割网络从对抗学习的过程中提高预测的质量,整个训练框架如图 16。

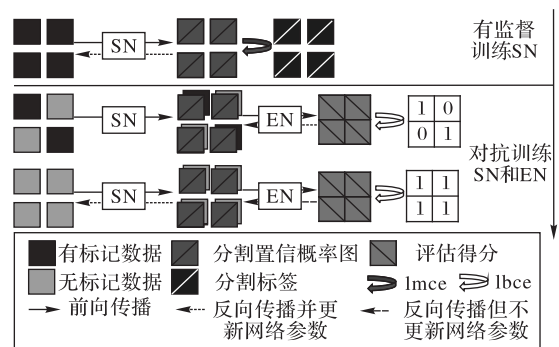


图 16 DAN 训练框架

Fig. 16 Training framework of DAN

Nie 等<sup>[74]</sup>根据分割任务的特点进一步细化了评价网络的目标,将判别器同样改为二分类的分割网络,使其产生像素级的质量评估,最后对无标记数据选取高质量的分割区域作为



伪标签参与训练,之后根据自训练的方法迭代地优化模型。

Zhou 等<sup>[75]</sup>探讨了使用了图像级标注的弱监督情形下伪标签的优化,仍然使用判别器评估图像级分割网络的真伪,此

外还定义了使用图像级标注预训练的分类网络,利用其产生的注意力特征辅助优化原分割网络的结果得到新的伪标签用于监督分割网络训练。

表 1 半监督医学影像分割方法汇总

Tab. 1 Summary of semi-supervised medical image segmentation methods

方法	描述	模型
自训练算法	使用网络本身的预测作为伪标签,之后将无标记数据加入数据集训练并重新预测伪标签,迭代以上过程直至收敛	Bai 等 <sup>[56]</sup> 、Tang 等 <sup>[57]</sup> 、Rajchl 等 <sup>[54]</sup> 、Kervadec 等 <sup>[61]</sup>
协同训练算法	将自训练中伪标签的生成方式扩展为通过多个模型预测并融合得到伪标签,进而迭代训练	Peng 等 <sup>[58]</sup> 、Zhou 等 <sup>[60]</sup> 、Xia 等 <sup>[59]</sup> 、Min 等 <sup>[62]</sup>
基于一致性正则的算法	利用对输入添加扰动或变换后输出应当具有一致性的先验条件,通常以辅助任务的方式添加正则化约束,对伪标签的依赖较小,主要关注于提升网络的泛化能力	Perone 等 <sup>[19]</sup> 、Yu 等 <sup>[20]</sup> 、Li 等 <sup>[21,64]</sup> 、Bortsova 等 <sup>[63]</sup> 、Chen 等 <sup>[67]</sup>
基于图的算法	将数据映射至图上,通过边描述数据间的相似性,进而使信息通过图结构传播并生成伪标签	Baur 等 <sup>[68]</sup> 、Song 等 <sup>[22]</sup> 、Xu 等 <sup>[23]</sup>
基于生成对抗网络的算法	构造辅助任务并利用 GAN 对抗学习,利用对抗样本提升网络的性能;根据辅助任务设计思路的不同可以达到精炼分割结果或补充额外数据的目的	Chaitanya 等 <sup>[71]</sup> 、Mondal 等 <sup>[72]</sup> 、Zhang 等 <sup>[73]</sup> 、Nie 等 <sup>[74]</sup> 、Zhou 等 <sup>[75]</sup> 、Ross 等 <sup>[76]</sup>

此外还有的研究在额外的辅助任务上执行对抗训练,从而间接提升网络的特征提取能力。Ross 等<sup>[76]</sup>在内窥镜器官分割任务中定义了从灰度图中还原的着色任务。训练流程如图 17,与其他半监督学习的流程不同,该方法首先在无标记数据上学习得到预训练模型。具体地,考虑到此类外科分割任务中的数据为彩色影像的特点,先将其转换至 Lab 颜色空间,再使用分割网络预测彩色部分分量,相对的判别网络负责区分输入图像为原始彩色分量还是分割网络输出的结果。得到预训练模型后,再使用有标记数据对分割网络的最后一层进行调整以实现分割任务。实验结果表明,着色可以为分割提供一定的特征提取能力,在只有很少的标记数据时效果明显。

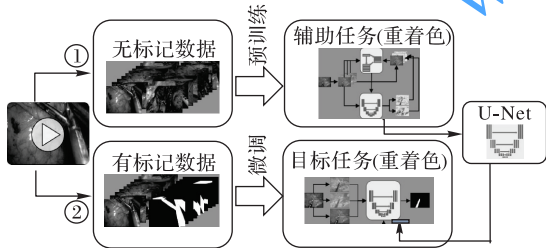


图 17 基于重着色的训练流程  
Fig. 17 Training framework based on re-colorization

### 4 医学影像分割中的不确定性度量

基于深度学习的医学影像分割在医学分割各任务中均取得了显著的成果,但是获取十分准确而且可靠的分割结果对大多数任务来说仍然具有挑战,尤其是在目标的边缘区域。因此,在给出分割的结果的同时对预测结果不确定度进行定量分析对理解分割结果的可靠性有重大意义<sup>[77]</sup>,例如,不确定性度量可以用来指示出潜在可能的误分割区域,从而指导医生对模型不确定度高的部分进行复查。

早期关于深度学习网络不确定性度量的研究主要集中在图像分类和检测这些粗粒度的预测任务中,随后,研究者 Kendall 等<sup>[78]</sup>在 2015 年将其推广到需要对逐像素预测结果进行不确定性度量的图像分割领域。文献[26]根据不确定性的分布类型角度,将不确定性分为认知不确定性(Epistemic uncertainty)和随机不确定性(Aleatoric uncertainty)。

### 4.1 模型认知不确定性

认知不确定性也称为模型不确定性,指的是系统原则上具备某种认知能力,但是受限于标注数据量、训练策略以及评价体系,从而导致的模型认知上的不确定性。可以通过提供额外的训练数据和改进模型训练策略来减轻和消除这种不确定性。在有限的医疗标注资源下,对模型不确定性的准确量化,是对当前医疗智能诊断系统的重要补充。模型不确定性的核心是获得模型参数改变时预测结果的分布,而传统卷积神经网络参数固定只能得到一次预测结果。研究者根据对模型不确定性评估的方式不同,将其大致分为两类,即深度模型集成(Deep model ensemble)<sup>[26]</sup>和深度贝叶斯网络(Deep Bayesian Neural Network)<sup>[79]</sup>。

#### 4.1.1 深度模型集成

早期深度模型集成的方式主要采用生成多个训练模型来近似预测分布,如图 18。文献[26]改变模型初始化参数从而获得不同初始化条件下的训练模型,进一步用获得的多个模型下的预测集起来表征模型不确定性:

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{M} \sum_{m=1}^M p_{\theta_m}(\mathbf{y}|\mathbf{x}, \theta_m) \tag{11}$$

文献[80]对模型集成方式进行简化,用多假设预测(Multiple Hypothesis Prediction)替换模型原来的单一假设预测(Single Hypothesis Prediction)。根据不同假设输出损失将标签空间进行沃罗伊诺划分(Voronoi tessellation),更新过程计算最近标签空间的损失进行反向传播。

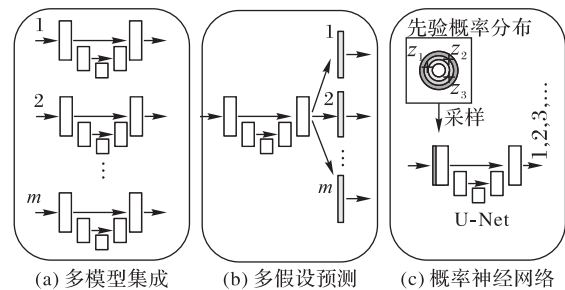


图 18 深度模型集成方式

Fig. 18 Deep model integration method

上述模型集成的方法能够得到一致的输出预测,但是受



限于固定的集成模型数目,无法得到更多假设数目或者连续空间假设下的输出结果,同时训练队计算资源的消耗也大大增加。随后, Kohl 等<sup>[81]</sup>在 2019 年 NIPS (advances in Neural Information Processing Systems)会议上提出概率神经网络,设计先验网络获得输入在潜在空间(Latent space)下的分布,然后通过计算 KL 散度(Kullback-Leibler Divergence, KLD)与后验网络下标签在潜在空间的分布对齐,以获得连续空间下的预测结果,从而获得模型预测的不确定性。Baumgartner 等<sup>[82]</sup>通过对多尺度下特征空间进行层级化(Hierarchical)建模,进一步提升了模型在连续空间下预测的精细化程度。

#### 4.1.2 深度贝叶斯网络

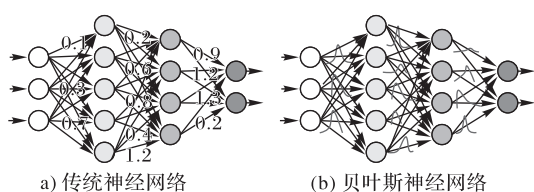


图 19 两种神经网络的区别

Fig. 19 Difference between two neural networks

传统的神经网络模型可以视为一个条件分布模型  $P(y|x, w)$ : 输入为  $x$ , 模型参数  $w$ , 输出预测  $y$  的分布。网络的学习过程是对模型参数  $w$  的最大似然估计:

$$w^{MLE} = \arg \max_w \log P(\mathcal{D}|w) \quad (12)$$

其中  $\mathcal{D}$  代表训练数据。此种优化过程下模型的参数  $w$  是固定的取值, 以此得到输出  $y$  的预测也是固定的, 无法体现不确定性。而深度贝叶斯网络<sup>[78]</sup>为神经网络的参数引入概率分布, 如图 19, 根据输入数据的分布去学习网络参数的后验概率分布, 建立基于模型参数概率分布的预测期望以度量不确定性:

$$P(y|x) = E_{P(w|\mathcal{D})} [P(y|x, w)] \quad (13)$$

而根据贝叶斯理论, 模型参数后验概率  $P(w|\mathcal{D})$  是无法直接求解的, 因为:

$$P(w|\mathcal{D}) = \frac{P(\mathcal{D}|w)P(w)}{P(\mathcal{D})} \quad (14)$$

式(14)中各项均无法直接求解。为了将求  $w$  后验分布的问题转化为更好求解的优化问题, 研究者们引入变分推断<sup>[83]</sup> (variational) 的思想, 这类方法基于由一组参数  $\theta$  控制的先验假设概率分布  $q(w|\theta)$  去逼近待求解的模型参数真实后验概率分布  $p(w|\mathcal{D})$ , 转化为如基于高斯先验假设的  $(\mu, \theta)$  的参数优化问题。这个问题可以通过最小化假设先验分布和真实后验分布的 KL 散度进行求解,

$$\theta^* = \arg \min_{\theta} D_{KL} [q(w|\theta) || P(w|\mathcal{D})] \quad (15)$$

上面优化目标中的 KL 散度可以分解成  $D_{KL} [q(w|\theta) || P(w)]$  和  $E_{q(w|\theta)} [\log P(\mathcal{D}|w)]$  两项之差。文献[84]中用蒙特卡洛采样法去近似 KL 散度中的积分项求解, 而以一定概率随机关闭模型中参数的 Dropout 策略<sup>[27, 85]</sup>可以结合蒙特卡洛采样以达到变分贝叶斯近似的目的, 同时降低模型训练的复杂度。

#### 4.2 随机不确定性

随机不确定性指的是观测中固有的噪声, 这部分不确定性来源于医疗设备采集成像的数据本身噪声以及标注存在的不可控误差, 不能通过获取更多的数据来减轻这种不确定性。对医学影像分割过程中随机不确定性的量化能标识出分割不

确定性很高的区域, 有助于辅助医生的判断。其中按照随机不确定性的分布主要可以分为输入不确定性和输出不确定性。

输入不确定性指的是由于成像设备的限制导致的医学影像的模糊性, 进而导致标注结果受到医疗专家主观认知以及客观差异化的影响, 造成误标、漏标的情况, 如图 20 所示, 图中不同的轮廓表示了不同专家的标注结果。Joskowicz 等<sup>[86]</sup>通过多轮次标注对标签的差异性进行统计建模, 界定差异范围 (variability range) 对输入不确定性统计分析, 对不确定性高的样例或者区域可以进一步重新标注以达到纠错的目的。而实际应用过程中, 不确定性统计建模的方式耗时耗力, 无法满足动态高效的需求, 因此基于模型输出分布的输出不确定性被较广泛研究。文献[78]对输入分别赋予同方差 (Homoscedastic) 噪声和异方差 (Heteroscedastic) 噪声以对模型输出的分布建模, 达到对随机不确定性量化的效果。对输出分布的研究也可以利用测试过程对数据增广的方式达到, Wang 等<sup>[87]</sup>采用几何变换和颜色空间变换对测试过程中输入数据进行增广, 观测输出空间的差异性, 进而推断数据本身的观测固有噪声。

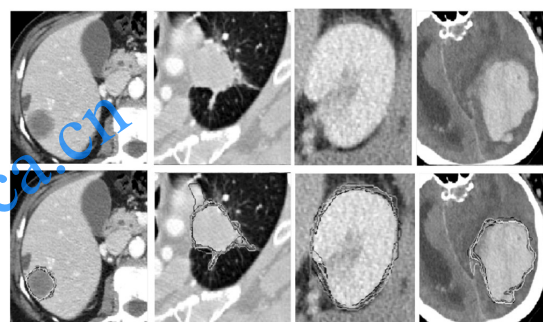


图 20 随机不确定性示意图

Fig. 20 Schematic diagram of random uncertainty

### 5 基于深度学习的医学影像分割趋势展望

医学影像分割在医疗智能辅助诊断中的价值明显, 尽管基于深度学习在医学影像分割领域中已经取得了显著的进展, 但基于深度学习思想和方法建立更加精确、高效、鲁棒的分割模型仍然值得更深入的研究。目前医学影像分割质量的提升主要得益于网络模型在图像表征学习能力上的优势, 以及现有计算技术下处理大规模数据的高效性。多数医学影像任务场景下目前的分割算法还达不到符合医疗应用的要求, 算法要求标注数据量大且重复标注多, 另外分割结果单一, 有效信息少。未来医学影像分割需要在以下几个方向开展更深入的研究。

#### 5.1 优化网络结构

医学影像中的组织不像自然图像具有清晰的边缘、纹理和颜色, 因此病变和健康组织的视觉纹理很难被区分开, 病变区域附近的背景冗余信息会严重干扰目标视觉特征的表达能力。而由于病变组织的多变性和复杂性, 类别内的样本纹理也存在着巨大的差异。因此, 医学影像分割会面临较小的类间区分性和较大的类内差异性。如何针对医学影像的特点, 设计能够将网络注意力侧重在目标区域, 且优化网络的特征表达, 使得提取到的目标区域更加紧凑, 和背景特征之间的距离尽可能大, 是一个亟须解决的问题。



## 5.2 不确定性度量

由于医疗业务数据特点(数据模糊、标注不准确)和应用场景的特殊性,对模型的鲁棒性和精度要求很高,目前的医学影像分割算法通常仅能给出单一的分割结果,有用信息量少。医生希望模型给出预测结果的同时,对结果的不确定性也能给出量化,这样的话医生就可以将精力重点放在模型不确定度高的地方,减少重复劳动。因此亟须在已有分割模型的预测基础上,增加关于分割网络不确定性的研究;同时如何结合不确定性的量化指标,优化模型训练过程,提升模型的分割性能,值得进一步的探索。

## 5.3 智能数据标注策略

目前,医学影像智能分析算法多以纯数据驱动的方式进行训练,造成模型泛化能力受标注数据质量的严重影响,过拟合严重。数据标注主要问题包括样本分布不均衡、标注差异化、同质样本冗余、样本孤立点等。针对这些问题,需要提出高效的与数据交互驱动的数据标注策略,使得模型训练过程中能够主动挑选出高价值的数据样本,交给医疗专家进行标注,从而减少重复标注工作,优化标注流程,达到海量样本空间下模型高效学习的目的。

## 5.4 无标注数据的利用

随着医学技术的发展,医学影像数据将会更加庞大,而医疗资源无法对所有数据进行标注,因此未来半监督学习算法还有很大的发展空间,并将获得更多的关注与研究。目前在半监督医学分割领域中还存在一些问题,首先是现有算法的性能上距离全监督学习的效果还有很大距离<sup>[88]</sup>,原因主要在于无标记数据中的信息难以被利用。一方面仅靠三个基本假设来定义数据分布与后验间的关系并不准确,使用某些特定先验的半监督学习策略在其他分布的数据上会造成一定的性能下降。另一方面尽管大多数算法对无标记数据中的信息进行了筛选约束,训练过程中仍不可避免地学习到错误的信息,从而导致了潜在的性能下降。综上所述,如何在医学影像分割任务中提出新的半监督学习算法,更深入挖掘未标注数据的有用信息,是研究的一个重要方向。

## 6 结语

医学影像分割是计算机辅助诊断中的重要一环,在过去几年随着深度学习的迅速发展得到广泛的关注。本文充分总结了基于深度学习的医学影像分割的研究进展。首先,本文重点介绍了医学影像分割深度学习模型的基本框架,并对比分析了基础网络结构的发展过程、用于优化的目标函数和用于提升模型性能的各种方法。随后本文针对医学影像中标注获取困难的问题,重点讨论了半监督条件下医学影像分割的发展现状,对半监督分割方法进行了归纳整理。还对医学影像分割中分割的不确定性研究这一较为新兴的研究方向进行了分析,论述了医学图像模糊、标注噪声大的不确定性分析的重要意义,并对比了主流的模型不确定性和随机不确定性的研究方法。最后,本文对深度学习在医学影像分割中的发展方向进行了展望,深度学习的进步也将推动着医学影像分割向更深、更广的领域发展。

### 参考文献 (References)

[1] DOI K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential [J]. *Computerized Medical Imaging and Graphics*, 2007, 31(4/5): 198-211.

[2] PHAM D L, XU C Y, PRINCE J L. Current methods in medical image segmentation [J]. *Annual Review of Biomedical Engineering*, 2000, 2:315-337.

[3] MUTHUKRISHNAN R, RADHA M. Edge detection techniques for image segmentation [J]. *International Journal of Computer Science and Information Technology*, 2011, 3(6): 259-267.

[4] OTSU N. A threshold selection method from gray-level histograms [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, 9(1): 62-66.

[5] KAGANAMI H G, ZOU B J. Region-based segmentation versus edge detection [C]// *Proceedings of the 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. Piscataway: IEEE, 2009: 1217-1221.

[6] PATIL D D, DEORE S G. Medical image segmentation: a review [J]. *International Journal of Computer Science and Mobile Computing*, 2013, 2(1): 22-27.

[7] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]// *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2016: 770-778.

[8] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015-04-10) [2020-10-10]. <https://arxiv.org/pdf/1409.1556.pdf>.

[9] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions [C]// *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2015: 1-9.

[10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// *Proceedings of the 25th International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc., 2012: 1097-1105.

[11] LONG J, SHEHMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]// *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2015: 3431-3440.

[12] MA Y H, HAO H Y, XIE J Y, et al. ROSE: a retinal OCT-angiography vessel segmentation dataset and new model [J]. *IEEE Transactions on Medical Imaging*, 2021, 40(3): 928-939.

[13] MILLETARI F, NAVAB N, AHMADI S A. V-Net: fully convolutional neural networks for volumetric medical image segmentation [C]// *Proceedings of the 4th International Conference on 3D Vision*. Piscataway: IEEE, 2016: 565-571.

[14] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [C]// *Proceedings of the 2015 International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 9351. Cham: Springer, 2015: 234-241.

[15] XU H, XIE H T, LIU Y Z, et al. Deep cascaded attention network for multi-task brain tumor segmentation [C]// *Proceedings of the 2019 International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 11766. Cham: Springer, 2019: 420-428.

[16] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation [EB/OL]. (2017-12-05) [2020-10-10]. <https://arxiv.org/pdf/1706.05587.pdf>.



- [17] CHEN L C, ZHU Y K, PAPANDEOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]// Proceedings of the 2018 European Conference on Computer Vision, LNCS 11211. Cham: Springer, 2018: 833-851.
- [18] ÇIÇEK Ö, ABDULKADIR A, LIENKAMP S S, et al. 3D u-net: learning dense volumetric segmentation from sparse annotation [C]// Proceedings of the 2016 International Conference on Medical Image Computing and Computer-Assisted Intervention, LNCS 9901. Cham: Springer, 2016: 424-432.
- [19] PERONE C S, COHEN-ADAD J. Deep semi-supervised segmentation with weight-averaged consistency targets [C]// Proceedings of the 2018 International Workshop on Deep Learning in Medical Image Analysis/ 2018 International Workshop on Multimodal Learning for Clinical Decision Support, LNCS 11045. Cham: Springer, 2018: 12-19.
- [20] YU L Q, WANG S J, LI X M, et al. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation [C]// Proceedings of the 2019 International Conference on Medical Image Computing and Computer-Assisted Intervention, LNCS 11765. Cham: Springer, 2019: 605-613.
- [21] LI X M, YU L Q, CHEN H, et al. Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model [C]// Proceedings of the 2018 British Machine Vision Conference. Durham: BMVA Press, 2018: No. 0162.
- [22] SONG Y Q, ZHANG C S, LEE J G, et al. Semi-supervised discriminative classification with application to tumorous tissues segmentation of MR brain images [J]. Pattern Analysis and Applications, 2009, 12(2): 99-115.
- [23] XU K, SU H, ZHU J, et al. Neuron segmentation based on CNN with semi-supervised regularization [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2016: 1324-1332.
- [24] YI X, WALIA E, BABYN P. Generative adversarial network in medical imaging: a review [J]. Medical Image Analysis, 2019, 58: No. 101552.
- [25] SWILER L P, GIUNTA A A. Aleatory and epistemic uncertainty quantification for engineering applications: SAND2007-2670C [R]. Albuquerque, NM: Sandia National Laboratories, 2007.
- [26] LAKSHMINARAYANAN B, PRITZEL A, BLUNDELL C. Simple and scalable predictive uncertainty estimation using deep ensembles [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2017: 6405-6416.
- [27] GAL Y, GHAHRAMANI Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning [C]// Proceedings of the 33rd International Conference on Machine Learning. New York: JMLR. org, 2016: 1050-1059.
- [28] ZHOU Z, RAHMAN SIDDIQUEE M M, TAJBAKSH N, et al. UNet++: a nested U-Net architecture for medical image segmentation [C]// Proceedings of the 2018 International Workshop on Deep Learning in Medical Image Analysis/ 2018 International Workshop on Multimodal Learning for Clinical Decision Support, LNCS 11045. Cham: Springer, 2018: 3-11.
- [29] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [30] ZHOU C H, CHEN S C, DING C X, et al. Learning contextual and attentive information for brain tumor segmentation [C]// Proceedings of the 2018 International MICCAI Brainlesion Workshop, LNCS 11384. Cham: Springer, 2018: 497-507.
- [31] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7794-7803.
- [32] HE X, YANG S B, LI G B, et al. Non-local context encoder: robust biomedical image segmentation against adversarial attacks [C]// Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2019: 8417-8424.
- [33] MENZE B H, JAKAB A, BAUER S, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) [J]. IEEE Transactions on Medical Imaging, 2014, 34(10): 1993-2024.
- [34] ZHOU C H, DING C X, WANG X C, et al. One-pass multi-task networks with cross-task guided attention for brain tumor segmentation [J]. IEEE Transactions on Image Processing, 2020, 29: 4516-4529.
- [35] CHEN L C, PAPANDEOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [36] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 764-773.
- [37] WANG G T, LI W Q, OURSELIN S, et al. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks [C]// Proceedings of the 2017 International MICCAI Brainlesion Workshop, LNCS 10670. Cham: Springer, 2017: 178-190.
- [38] LIU S A, XU H, LIU Y Z, et al. Improving brain tumor segmentation with dilated pseudo-3D convolution and multi-direction fusion [C]// Proceedings of the 2020 International Conference on Multimedia Modeling, LNCS 11961. Cham: Springer, 2020: 727-738.
- [39] GUO X N, XIE H T, XU H, et al. Global context and boundary structure-guided network for cross-modal organ segmentation [J]. Information Processing and Management, 2020, 57 (4) : No. 102252.
- [40] QIU Z F, YAO T, MEI T. Learning spatio-temporal representation with pseudo-3D residual networks [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 5534-5542.
- [41] JIANG Z Y, DING C X, LIU M F, et al. Two-stage cascaded U-Net: 1st place solution to BraTS challenge 2019 segmentation task [C]// Proceedings of the 2019 International MICCAI Brainlesion Workshop, LNCS 11992. Cham: Springer, 2019: 231-241.
- [42] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2999-3007.
- [43] SUDRE C H, LI W Q, VERCAUTEREN T, et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations [C]// Proceedings of the 2017 International



- Workshop on Deep Learning in Medical Image Analysis/ 2017 International Workshop on Multimodal Learning for Clinical Decision Support, LNCS 10553. Cham: Springer, 2017: 240-248.
- [44] REN X H, AHMAD S, ZHANG L C, et al. Task decomposition and synchronization for semantic biomedical image segmentation [J]. *IEEE Transactions on Image Processing*, 2020, 29: 7497-7510.
- [45] LIU S A, GUO X N. Improving brain tumor segmentation with multi-direction fusion and fine class prediction[C]// *Proceedings of the 2019 International MICCAI Brainlesion Workshop*, LNCS 11992. Cham: Springer, 2019: 349-358.
- [46] ISENSEE F, PETERSEN J, KLEIN A, et al. nnU-Net: self-adapting framework for U-Net-based medical image segmentation [EB/OL]. (2018-09-27) [2020-10-10]. <https://arxiv.org/pdf/1809.10486.pdf>.
- [47] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]// *Proceedings of the 32nd International Conference on Machine Learning*. New York: JMLR.org, 2015: 448-456.
- [48] WU Y X, HE K M. Group normalization [C]// *Proceedings of the 2018 European Conference on Computer Vision*, LNCS 11217. Cham: Springer, 2018: 3-19.
- [49] ULYANOV D, VEDALDI A, LEMPITSKY V. Instance normalization: the missing ingredient for fast stylization [EB/OL]. (2017-11-06) [2020-10-10]. <https://arxiv.org/pdf/1607.08022.pdf>.
- [50] BA J L, KIROS J R, HINTON G E. Layer normalization [EB/OL]. (2016-07-21) [2020-10-10]. <https://arxiv.org/pdf/1607.06450.pdf>.
- [51] MAAS A L, HANNUN A Y, NG A Y. Rectifier nonlinearities improve neural network acoustic models [C]// *Proceedings of the 30th International Conference on Machine Learning*. New York: JMLR.org, 2013: 3.
- [52] HE K M, ZHANG X Y, REN S Q, et al. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification [C]// *Proceedings of the 2015 IEEE International Conference on Computer Vision*. Piscataway: IEEE, 2015: 1026-1034.
- [53] MYRONENKO A. 3D MRI brain tumor segmentation using autoencoder regularization [C]// *Proceedings of the 2018 International MICCAI Brainlesion Workshop*, LNCS 11384. Cham: Springer, 2018: 311-320.
- [54] RAJCHL M, LEE M C H, OKTAY O, et al. DeepCut: object segmentation from bounding box annotations using convolutional neural networks [J]. *IEEE Transactions on Medical Imaging*, 2017, 36(2): 674-683.
- [55] KAMNITSAS K, BAI W J, FERRANTE E, et al. Ensembles of multiple models and architectures for robust brain tumour segmentation [C]// *Proceedings of the 2017 International MICCAI Brainlesion Workshop*, LNCS 10670. Cham: Springer, 2017: 450-462.
- [56] BAI W J, OKTAY O, SINCLAIR M, et al. Semi-supervised learning for network-based cardiac MR image segmentation [C]// *Proceedings of the 2017 International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 10434. Cham: Springer, 2017: 253-260.
- [57] TANG M, VALIPOUR S, ZHANG Z C, et al. A deep level set method for image segmentation [C]// *Proceedings of the 2017 International Workshop on Deep Learning in Medical Image Analysis/ 2017 International Workshop on Multimodal Learning for Clinical Decision Support*, LNCS 10553. Cham: Springer, 2017: 126-134.
- [58] PENG J Z, ESTRADA G, PEDERSOLI M, et al. Deep co-training for semi-supervised image segmentation [J]. *Pattern Recognition*, 2020, 107: No. 107269.
- [59] XIA Y D, YANG D, YU Z D, et al. Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation [J]. *Medical Image Analysis*, 2020, 65: No. 101766.
- [60] ZHOU Y Y, WANG Y, TANG P, et al. Semi-supervised 3D abdominal multi-organ segmentation via multi-planar co-training [C]// *Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision*. Piscataway: IEEE, 2019: 121-140.
- [61] KERVADEC H, DOLZ J, GRANGER É, et al. Curriculum semi-supervised segmentation [C]// *Proceedings of the 2019 International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 11765. Cham: Springer, 2019: 568-576.
- [62] MIN S B, CHEN X J. A robust deep attention network to noisy labels in semi-supervised biomedical segmentation [EB/OL]. (2018-07-31) [2020-10-10]. <https://arxiv.org/pdf/1807.11719v1.pdf>.
- [63] BORTSOVA G, DUBOST F, HOGEWEG L, et al. Semi-supervised medical image segmentation via learning consistency under transformations [C]// *Proceedings of the 2019 International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 11769. Cham: Springer, 2019: 810-818.
- [64] LI X M, YU L Q, CHEN H, et al. Transformation-consistent self-ensembling model for semisupervised medical image segmentation [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(2): 523-534.
- [65] LAINE S, AILA T. Temporal ensembling for semi-supervised learning [EB/OL]. (2017-03-15) [2020-10-10]. <https://arxiv.org/pdf/1610.02242.pdf>.
- [66] TARVAINEN A, VALPOLA H. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results [C]// *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc., 2017: 1195-1204.
- [67] CHEN S, BORTSOVA G, GARCÍA-UCEDA JUÁREZ A, et al. Multi-task attention-based semi-supervised learning for medical image segmentation [C]// *Proceedings of the 2019 International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 11766. Cham: Springer, 2019: 457-465.
- [68] BAUR C, ALBARQOUNI S, NAVAB N. Semi-supervised deep learning for fully convolutional networks [C]// *Proceedings of the 2017 International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 10435. Cham: Springer, 2017: 311-319.
- [69] WESTON J, RATLE F, MOBAHI H, et al. Deep learning via semi-supervised embedding [M]// MONTAVON G, ORR G B,



- MÜLLER K R. *Neural Networks: Tricks of the Trade*, LNCS 7700. 2nd ed. Berlin: Springer, 2012: 639-655.
- [70] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]// *Proceedings of the 27th International Conference on Neural Information Processing Systems*. Cambridge: MIT Press, 2014: 2672-2680.
- [71] CHAITANYA K, KARANI N, BAUMGARTNER C F, et al. Semi-supervised and task-driven data augmentation [C]// *Proceedings of the 2019 International Conference on Information Processing in Medical Imaging*, LNCS 11492. Cham: Springer, 2019: 29-41.
- [72] MONDAL A K, DOLZ J, DESROSIERS C. Few-shot 3D multi-modal medical image segmentation using generative adversarial learning[EB/OL]. (2018-10-29) [2020-10-10]. <https://arxiv.org/pdf/1810.12241.pdf>.
- [73] ZHANG Y Z, YANG L, CHEN J X, et al. Deep adversarial networks for biomedical image segmentation utilizing unannotated images[C]// *Proceedings of the 2017 International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 10435. Cham: Springer, 2017: 408-416.
- [74] NIE D, GAO Y Z, WANG L, et al. ASDNet: attention based semi-supervised deep networks for medical image segmentation [C]// *Proceedings of the 2018 International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 11073. Cham: Springer, 2018: 370-378.
- [75] ZHOU Y, HE X D, HUANG L, et al. Collaborative learning of semi-supervised segmentation and classification for medical images [C]// *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2019: 2074-2083.
- [76] ROSS T, ZIMMERER D, VEMURI A, et al. Exploiting the potential of unlabeled endoscopic video data with self-supervised learning[J]. *International Journal of Computer Assisted Radiology and Surgery*, 2018, 13(6): 925-933.
- [77] MERTIKOPOULOS P, MOUSTAKAS A L. Learning in an uncertain world: MIMO covariance matrix optimization with imperfect feedback[J]. *IEEE Transactions on Signal Processing*, 2016, 64(1): 5-18.
- [78] KENDALL A, GAL Y. What uncertainties do we need in bayesian deep learning for computer vision?[C]// *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc., 2017: 5580-5590.
- [79] GAL Y, ISLAM R, GHAHRAMANI Z. Deep Bayesian active learning with image data [C]// *Proceedings of the 34th International Conference on Machine Learning*. New York: JMLR.org, 2017: 1183-1192.
- [80] RUPPRECHT C, LAINA I, DIPIETRO R, et al. Learning in an uncertain world: representing ambiguity through multiple hypotheses [C]// *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Piscataway: IEEE, 2017: 3591-3600.
- [81] KOHL S, ROMERA-PAREDES B, MEYER C, et al. A probabilistic U-Net for segmentation of ambiguous images [C]// *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc., 2018: 6965-6975.
- [82] BAUMGARTNER C F, TEZCAN K C, CHAITANYA K, et al. PHiSeg: capturing uncertainty in medical image segmentation [C]// *Proceedings of the 2019 International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 11765. Cham: Springer, 2019: 119-127.
- [83] LOPEZ R, REGIER J, JORDAN M I, et al. Information constraints on auto-encoding variational Bayes[C]// *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc., 2018: 6117-6128.
- [84] BLUNDELL C, CORNEBISE J, KAVUKCUOGLU K, et al. Weight uncertainty in neural networks [C]// *Proceedings of the 32nd International Conference on Machine Learning*. New York: JMLR.org, 2015: 1613-1622.
- [85] KINGMA D P, SALIMANS T, WELLING M. Variational dropout and the local reparameterization trick [C]// *Proceedings of the 28th International Conference on Neural Information Processing Systems*. Cambridge: MIT Press, 2015: 2575-2583.
- [86] JOSKOWICZ L, COHEN D, CAPLAN N, et al. Automatic segmentation variability estimation with segmentation priors [J]. *Medical Image Analysis*, 2018, 50: 54-64.
- [87] WANG G T, LI W Q, AERTSEN M, et al. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks [J]. *Neurocomputing*, 2019, 338: 34-45.
- [88] SINGH A, NOWAK R D, ZHU X J. Unlabeled data: now it helps, now it doesn't [C]// *Proceedings of the 21st International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc., 2009: 1513-1520.

This work is partially supported by the National Natural Science Foundation of China (62022076, U19A2057, 61976008).

**CAO Yuhong**, born in 1968, Ph. D., senior engineer. Her research interests include computer vision, artificial intelligence.

**XU Hai**, born in 1993, Ph. D. candidate. His research interests include computer vision, artificial intelligence, semantic segmentation.

**LIU Sun'ao**, born in 1997, M. S. candidate. His research interests include computer vision, artificial intelligence, semantic segmentation.

**WANG Zixiao**, born in 1999, M. S. candidate. His research interests include computer vision, semantic segmentation, semi-supervised learning.

**LI Hongliang**, born in 1987, Ph. D., engineer. His research interests include computer vision, computational imaging.