

基于 DSNet 的遥感影像语义分割方法

史芳行¹, 周林娥², 朱大明^{1*}, 付志涛¹

¹昆明理工大学国土资源工程学院, 云南 昆明 650093;

²曲靖职业技术学院, 云南 曲靖 655000

摘要 针对遥感影像语义分割数据的类别不平衡导致传统神经网络模型趋向于忽略困难样本、推理结果存在空洞和分割精度下降等问题, 提出了一种钻形神经网络语义分割方法。首先, 定义新的桥接模块用于融合浅层和深层特征信息, 使更多建筑细节特征能被网络捕捉; 其次, 利用多重损失函数在深度学习分割模型训练中加强对困难样本信息的提取; 最后, 平衡类别训练差异, 多层次提取遥感影像中的地物信息, 提高分割精度。实验结果表明, 所提方法的平均交并比达到 0.849, 建筑物漏识率和错误率较少, 分割精度相比现有方法有一定的提高。

关键词 遥感; 深度学习; 语义分割; 类别不平衡; DSNet

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP212901

DSNet-Based Remote Sensing Image Semantic Segmentation Method

Shi Fangxing¹, Zhou line², Zhu Daming^{1*}, Fu Zhitao¹

¹Faculty of Land and Resources Engineering, Kunming University of Science and Technology, Kunming 650093, Yunnan, China;

²Qujing Vocational and Technical College, Qujing 655000, Yunnan, China

Abstract In view of the problems that the traditional neural network model tends to ignore difficult samples due to the unbalanced classification of remote sensing image semantic segmentation data, and the reasoning results are hollow and the segmentation accuracy decreases, a drill-shaped neural network semantic segmentation method is proposed. First, a new bridge module is defined to fuse the shallow and deep feature information, thus more building details can be captured by the network; second, in the deep learning segmentation model training, the multi loss function is used to improve the extraction of difficult sample information; finally, to balance the differences of category training, the feature information is extracted from remote sensing images at multiple levels, and the segmentation accuracy is improved. The experimental results show that the average intersection to union ratio of the proposed method reaches 0.849, the building missing rate and wrong recognition rate are less, and the segmentation accuracy is improved compared with the existing methods.

Key words remote sensing; deep learning; semantic segmentation; category imbalance; DSNet

1 引言

计算机软硬件的发展影响着图像分割方法的发展, 受限于当时计算机软硬件条件, 传统的图像分割方法只能提取到图像低层特征, 需要人工干预, 且分割精度不理想。随着 GPU 加速技术的诞生, 计算机的计算能力大幅提升, 推动了深度学习技术的快速推广, 基于深度学习的图像语义分割方法迎来雨后春笋般的新发展, 并在各行各业中得到广泛应用^[1]。Long 等^[2]提出了全卷积网络 (FCN) 模型, 首次将卷积神经网络技术

应用于图像语义分割。剑桥大学语义分割开发团队提出的 SegNet^[3]通过上采样层和反卷积层解码特征, 缓解了内存压力和模型复杂度。Liu 等^[4]提出了 ParseNet, ParseNet 通过全局平均池化得到全局特征与局部特征, 综合利用上下文信息。Lin 等^[5]提出了 RefineNet, RefineNet 由基于残差连接设计的多个组件构成, 可有效弥补下采样过程中的丧失信息。Zhao 等^[6]提出了金字塔场景解析网络 (PSPNet), 该网络在 ResNet 的基础上加入扩张卷积提取图像特征, 融合多标准特征解决上下文关系推断能力不足等问题。Fu

收稿日期: 2021-11-08; 修回日期: 2021-12-05; 录用日期: 2022-01-07; 网络首发日期: 2022-01-17

基金项目: 国家自然科学基金 (41961053)

通信作者: *634617255@qq.com

等^[7]提出了双重注意网络(DANet),该网络通过ResNet主干网络提取特征图及聚合输出特征图的方式建立全局上下文依赖关系,增强特征表示。Liang等^[8]提出的图的长短期记忆网络(Graph-LSTM)将传统LSTM数据延伸到图结构中,利用超像素及其空间连接自适应构造无向图拓扑结构,保留了目标的边界信息,达到增强特征表示的效果。Luc等^[9]首次将生成对抗网络(GAN)应用于图像语义分割,将预测图和真实标签图输入判别网络中判断真假,以提高预测值与标签值之间的一致性。Ronneberger等^[10]提出的UNet模型采用U型对称结构,通过低层细节信息和高层分类信息的有效融合实现更精确的分割。在UNet中加入注意力机制、多任务学习、加权交叉熵损失函数的改进模型可适应不同分割场景,提高分割准确度^[11-13]。Zhou等^[14-16]在UNet结构的基础上提出UNet++,UNet++在UNet结构中加入6个密集卷积核,通过嵌套跳接减小浅层特征与深层特征之间的语义鸿沟,UNet的改进模型获得了更高的分割准确度和训练效率。

现有语义分割网络模型大多通过有效提取高低层次特征信息和融合网络中的多路径特征来增加分割准确度。在训练数据不平衡的情况下,现有模型通常倾向于占比较大的负例,不平衡负例能左右模型训练过程,交叉熵函数会削弱对困难样本获得的知识,使得现有模型在不平衡数据上预测效果不理想。为此,本文提出了一种钻形神经网络模型,该模型利用桥接模块提高特征提取效率,在网络中加入多重损失函数来提升对数据稀疏区的敏感程度,降低简单样本学习度,加强训练时模型对困难样本信息的获取,提取更多细节知识,从而提高遥感影像语义分割准确度。

2 基本原理

2.1 UNet++

UNet^[10]被提出后广泛地用于各类影像的分割,因其优秀的性能和分割效果使得UNet成为影像分割中最经典的网络之一。但是随着分割场景和分割数据的不断变化,UNet已不能满足新的分割要求。首先,每个用于训练的数据集有着不同的最优深度,通过UNet获取最优深度需要宽泛的架构搜寻和差异模型反复集成测试;其次,UNet中的跳接属于限制性融合方案,强制融合深层和浅层特征图。针对这些问题,Zhou等^[14]提出了一种扩充嵌套特征融合的UNet++,UNet++综合利用长短连接架构克服限制性跳接,更加立体地捕捉分割数据特征,通过扩充特征融合的方式融合多层次信息,利用更浅的网络结构使多尺度特征图具有相似可融性,采用多路损失估量模型预测值与真实值的不一致程度。UNet++整体模型呈实心V形,如图1所示,蓝色圈代表编码和解码,紫色圈表示新跳接路径上的稠密卷积块(Dense block),绿色箭线代表水平层的稠密卷积网络(DenseNet)结构^[17]。

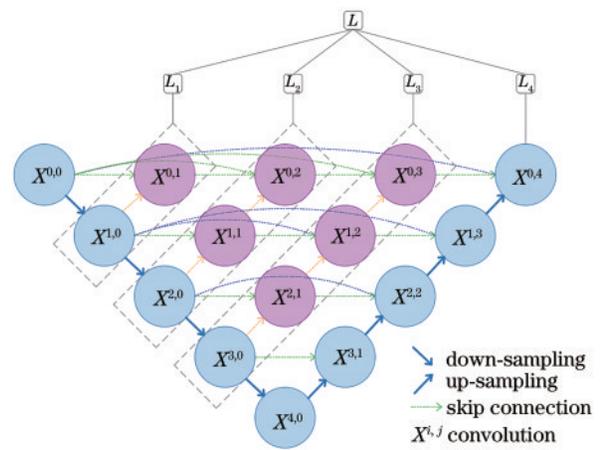


图1 UNet++结构

Fig. 1 UNet++ structure

图1中:左边 $X^{0.0} \sim X^{4.0}$ 代表最大值池化的下采样编码,每路下采样的下一节点和上一节点进行融合,形成多路损失结构;右边 $X^{4.0} \sim X^{0.4}$ 代表反卷积上采样解码,每路上采样的下一节点和同层跳接获得的稠密卷积块进行融合,累积全部初步特征图,在多语义级别生成完整分辨率的特征映射,有效融合底层和高层信息,进一步提高语义分割准确度。

2.2 数据类别不平衡

在语义分割任务中,由于计算机显存有限,难以将尺寸巨大的遥感影像送入模型训练,必须对影像进行裁剪,经过裁剪的数据存在类别不平衡问题,部分裁剪后的图像不同类别的训练样本差别很大,图像前景背景极度不平衡,一些待分割的图像占比整幅样本很少,如图2所示。假设训练集中背景占比99%,建筑物占比1%,传统神经网络模型会选择放弃建筑物预测,把所有样本分为背景而获得99%的训练分割准确率。过多的类别不平衡数据会促使网络模型提取样本数目较多、容易训练的像素,忽略样本数目较少、训练困难的像素,模型泛化能力无法得到提高,使得预测数据出现空洞和小目标的漏检。

2.3 DSNet

在UNet中,4次下采样对密集数据来说难以充分提取图像特征,而对于稀疏数据而言4次下采样又会提取冗余特征,导致预测过度。在UNet++模型中,每个跳跃路径均使用密集卷积块,先前特征图都累积到当前节点,使得参数量较大。面对难易不平衡数据时,两个模型都更多学习容易样本,忽略困难样本,使得在不平衡数据上的语义分割准确度不佳。基于此,本研究提出了一种钻形网络(DSNet),通过对UNet++结构的调整,去掉原网络中间的3个密集卷积核,结合新优化器,提高网络模型训练速度和缩减网络参数量,从而达到更好的分割效果。图3中:左边蓝色圈($X^{0.0} \sim X^{4.0}$)代表下采样编码,每个模块包含ReLU、Conv、BatchNorm等3个操作,通过多重损失函



图 2 类别不平衡数据

Fig. 2 Category imbalance data

数捕捉语义分割数据中的困难样本信息;顶部紫色圈代表桥接卷积块($X^{0,1} \sim X^{0,3}$),将编码器中的特征图通过上采样累积到桥接卷积块,得到第 2 路特征图,存储困难样本信息,桥接设计有利于优化器的优化;蓝色箭线传递困难样本信息,橙色平箭线连接同层中下采样简单信息,橙色斜箭线连接桥接卷积块中困难信息;右边 $X^{3,1} \sim X^{0,4}$ 为反卷积上采样解码,在解码器节点中通过反馈融合同层编码特征图和桥接累积特征图得到全局特征图。DSNet 参数设置如表 1 所示。

同时改进损失函数,使用 Focal 损失函数适应建筑语义分割模型,图 4 上部圆角方形代表 DSNet 输出的多路复合损失。所提方法舍去了 UNet++ 用于剪枝的深监督机制,将各分支的输出结果融合输出分类结果,最终输出结果包含所有分支的信息,并且网络会自动地根据损失函数来调整各路特征的权重,捕捉不

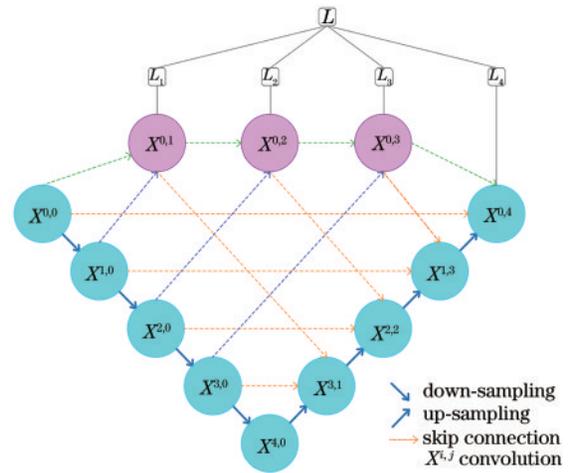


图 3 DSNet 结构

Fig. 3 DSNet structure

同深度的特征细节。

表 1 DSNet 参数设置

Table 1 DSNet parameter setting

Layer	Input size	Operation	Output size
$X^{0,0}$	$3 \times 512 \times 512$	ReLU	$32 \times 512 \times 512$
		32, Conv(3×3) BatchNorm	
$X^{1,0}$	$32 \times 512 \times 512$	ReLU	$64 \times 256 \times 256$
		64, Conv(3×3) BatchNorm	
$X^{2,0}$	$64 \times 256 \times 256$	ReLU	$128 \times 128 \times 128$
		128 Conv(3×3) BatchNorm	
$X^{3,0}$	$128 \times 128 \times 128$	ReLU	$256 \times 64 \times 64$
		256, Conv(3×3) BatchNorm	
$X^{4,0}$	$256 \times 64 \times 64$	ReLU	$512 \times 32 \times 32$
		512, Conv(3×3) BatchNorm	

续表

Layer	Input size	Operation	Output size
$X^{3,1}$	$512 \times 32 \times 32$	ReLU	$256 \times 64 \times 64$
		256, Conv(3×3)	
		BatchNorm	
$X^{2,2}$	$256 \times 64 \times 64$	ReLU	$128 \times 128 \times 128$
		128, Conv(3×3)	
		BatchNorm	
$X^{1,3}$	$128 \times 128 \times 128$	ReLU	$64 \times 256 \times 256$
		64, Conv(3×3)	
		BatchNorm	
$X^{0,1}, X^{0,2}, X^{0,3}, X^{0,4}$	$32 \times 512 \times 512$	ReLU	$32 \times 512 \times 512$
		32, Conv(3×3)	
		BatchNorm	

3 分析与讨论

3.1 数据获取与增广

所用数据集来自于 Google Earth 上截取的 RGB 影像,影像为 TIFF 格式,尺寸为 $8200 \text{ pixel} \times 7600 \text{ pixel}$,地面分辨率为 14 cm ,覆盖地表面积为 $(1100 \times 1100) \text{ m}^2$,影像包含山区乡村建筑、农业用地、山林草地等多类乡村用地类型。由于 GPU 内存限制,无法使用大图像训练,因此图像必须裁剪为小块才能送入模型。数据集仅标注建筑地物与非建筑地物 2 个语义类,白色标注为建筑物,其他地物为黑色背景,数据集按照 1:9 的比例划分验证集和训练集。

在深度学习中,训练模型的效果和样本数量成正比,数据的多少决定了模型的泛化能力,只对图像进行裁剪不能满足所需样本数据量,所以需要数据集进行数据增广处理。数据增广主要有两方面作用:1)增加数据量,提高模型的泛化能力;2)提升数据多样性,增强模型鲁棒性。对于遥感图像,常用数据增广方法有旋转、翻转、缩放、剪裁、平移、添加噪声、颜色抖动等,通过数据增广可以获得任意期望数据量,深入捕捉图像信息,而且用较小的图像训练卷积神经网络可以大大减少训练期间的 GPU 内存和计算成本。本研究使用数据增广获得 35000 对原图数据和标签数据,影像格式为 PNG,并随机抽取 3500 对原图和标签用于模型验证,所有样本数据的尺寸均为 512×512 。

3.2 实验评价指标

为验证所提方法的语义分割有效性和优势,采用像素准确率(PA)、平均像素准确率(MPA)、平均交并比(mIoU)等 3 个评价指标进行对比分析,其计算公式分别为

$$R_{PA} = \frac{1}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}}, \quad (1)$$

$$R_{MPA} = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}}, \quad (2)$$

$$R_{mIoU} = \frac{1}{k+1} \frac{\sum_{i=0}^k p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}}, \quad (3)$$

式中: p_{ij} 代表真实值为 i ,被预测为 j 的数量; $k+1$ 代表分割类别数; p_{ii} 为真正,代表正确分为该类的像素数量; p_{ji} 为假正,表示他类被分为该类的像素数; p_{ji} 为假负,表示该类被误分为他类的像素数量;式(1)~(3)先将背景和建筑物分别作为正样本求出评价指标,再取平均值便可得到所有类的平均评价指标,其中正样本为建筑物。

3.3 多重损失函数

在语义分割中存在大量前景与背景类别不平衡问题,使用单一损失函数往往趋向于捕捉建筑物占比更大的样本,而建筑物占比小背景占比大的样本容易被损失函数过滤掉,为解决该问题,本研究以 Focal 损失为主函数,Dice 损失为辅函数的多重损失函数优化网络模型。

Dice 损失^[18]是一种计算集合相似度函数,最初用在医学病理影像分割中,现被广泛应用于各类图像分割领域,Dice 损失常用于计算预测值和真实值之间的相似度,数值越小越相似,预测效果越理想。Dice 损失可以在初期加快收敛,提升模型训练效率。2020 年, Lin 等^[19]在 RetinaNet 中提出 Focal 损失函数,Focal 损失在样本极不平衡的单一目标检测任务中表现优异。Focal 损失可以较好地解决语义分割中正负样本比例严重失衡和难分类样本学习问题,降低大量简单正样本在训练中所占的权重,让模型趋于挖掘困难样本,捕捉 Dice 损失函数之外的困难样本中的信息,在模型后期提高模型鲁棒性和准度,有利于获得最优模型。复合损失函数可以从不同角度捕捉预测过程中的不足和损失,在定位全局最优的结果下得到局部最优,提高泛化性。损失函数的计算公式为

$$L_{DiceLoss} = 1 - \frac{2|X \cap Y| + C_{smooth}}{|X| + |Y| + C_{smooth}}, \quad (4)$$

$$L_{FocalLoss}(p_t) = -\alpha_t (1 - p_t)^\gamma \log p_t, \quad (5)$$

$$L_{ALL} = L_{FocalLoss}(p_t) + \lambda \times L_{DiceLoss}, \quad (6)$$

式中: $|X|$ 表示标签值像素个数; $|Y|$ 表示预测值像素个数; $|X \cap Y|$ 表示标签值和预测值交集的像素数; C_{smooth}

为防止分母为 0 的值; $-\log p_i$ 为初始交叉熵损失函数; α 为类别区间 (0,1 二分类) 的权重参数; $(1 - p_i)^\gamma$ 为简单/困难样本调节因子, γ 为聚焦参数; λ 为经验参数, 用于调节两个损失函数之间的权重。

实验结果表明, 经验参数 λ 取 0.9 时效果最优。不同损失函数的对比实验如图 4 所示, 随着训练次数的增加和困难样本信息的积累, 融合 Focal 损失的多重损失函数对比 Binary CrossEntropy (BCE) 损失函数曲线更快到达极值, 使用 Focal 损失的多重损失函数波动幅度较小, 训练后期逐渐趋于稳定并逼近全局最优。

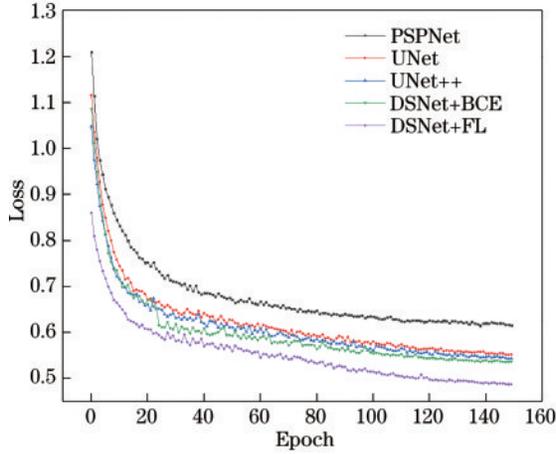


图 4 不同损失函数收敛效果对比

Fig. 4 Comparison of convergence effects of different loss functions

3.4 优化器

2014 年, 由 Kingma 等^[20]提出的 Adam 优化算法结合 AdaGrad 和 RMSProp 两种算法的优点, 考虑梯度均值和梯度未中心化方差来更新步长, 算法实现简单高效, 成为许多经典语义分割模型默认工作性能比较优秀的优化器。虽然该优化器收敛速度快, 但 Adam 存在参数过拟合及模型收敛不佳等问题。为此, Guger 等^[21]在提出了 AdamW 优化器, 在最终的参数更新时加入参数本身, 强化 L2 正则项作用, 使参数量尽可能缩减, 搭配 Amsgrad 可以加快收敛速度。AdamW 收敛优化效果相比 Adam 更快, 训练过程中损失值更小, 且数据曲线较为平滑, 最终结果比 Adam 更优, 如图 5 所示。

3.5 结果分析

实验的硬件环境如下: CPU 型号为 Intel i7-9750H; 内存为 16 GB; 显卡为 NVIDIA GeForce GTX 1660 Ti 6 GB; GPU 加速库采用 CUDA11.1。深度学习框架采用 Pytorch 和公开的数据增广库 Albumentations。实验迭代次数为 150 个 epoch, 在训练使用 Early Stopping 机制控制 epoch 训练次数, Early Stopping 机制可以判断在过程中连续训练 n 个 epoch 的评价指标不增加, 则提前终止模型训练, 节约时间成

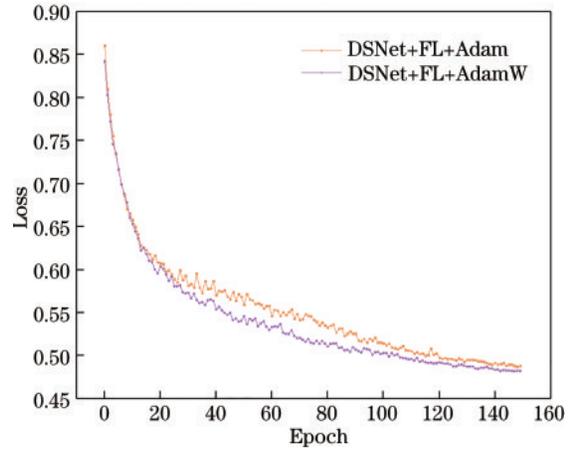


图 5 不同优化器收敛效果对比

Fig. 5 Comparison of convergence effects of different loss functions

本, 提高训练效率。学习率确定为 1×10^{-4} , 训练损失最小, 准确率最高, 更容易获得最优值。

表 2 展示了消融实验的结果, 从表 2 可以看出, 多重损失函数和新的优化器能够提高网络性能, 加入 Focal 损失和 AdamW 后的网络对于建筑物稀疏地区的图像分割结果有显著提升, 通过 AdamW 更新和计算影响模型训练和模型输出的网络参数, 使其逼近或达到最优值, 从而使 Focal 损失函数最小化, 因数据类别不平衡导致的困难样本分割准确度低的问题得到改善, 二者的搭配显著提高了建筑物语义分割准确度, 验证了所提方法的有效性。

表 2 消融实验

Table 2 Ablation experiment

Model	PA	MPA	mIoU
DSNet+BCE+Adam	0.901	0.954	0.831
DSNet+FL+Dice+Adam	0.939	0.962	0.839
DSNet+FL+Dice+AdamW	0.941	0.977	0.849

为了验证所提方法的效果, 将其语义分割结果同 3 种经典的语义分割模型网络 PSPNet 模型、UNet 模型、UNet++ 模型进行对比, 具体结果如表 3 所示。从表 3 可以看出, 相较于其他网络模型, 所提方法在像素准确率、平均像素准确率及平均交并比方面均有提高, 分割效果更好。

表 3 模型预测评价指标对比

Table 3 Comparison of model prediction and evaluation indexes

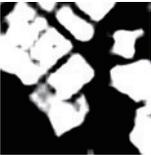
Model	PA	MPA	mIoU
PSPNet	0.898	0.913	0.797
UNet	0.913	0.956	0.812
UNet++	0.928	0.961	0.838
Proposed model	0.941	0.977	0.849

同时, 为检测模型的分割效果, 选择多处影像来检测网络的性能, 具体结果如表 4 所示。白色表示分割

得到的建筑物区域,与标签对比可知,各个网络基本均能分割出建筑物。但 PSPNet 的分割结果中获得的分割区域明显不足,且建筑物边界粗糙;而在 UNet 模型中,采用编码-解码结构,通过拼接特征向量获得低层个高层语义信息,在分割结果中建筑物边界提取效果明显;UNet++ 模型通过密集连接,更加立体融合各层特征信息,在分割结果中对小目标建筑物检测效果

有所改善,但仍存在空洞现象。所提模型使用新的网络结构,同时加入 Focal 损失及 AdamW,可学习到更多边界零碎信息、平衡类别训练差异、多层次提取遥感影像中的地物信息,分割效果得到提升。所提方法预测图空洞最少,提取的房屋与标签值更为接近,建筑物之间的划分界线最清晰,房屋边界较为光滑。

表 4 DSNNet 模型分割效果对比
Table 4 Segmentation effect comparison of DSNNet model

	Image	Label	PSPNet	UNet	UNet++	Proposed model
1						
2						
3						
4						

此外,数据集的标注精度对模型的预测效果有一定影响。标注精度和预测效果呈正相关,标注精度越高,预测效果越好,评估指标值越高。本研究数据源自高分遥感卫星影像,图片分辨率不如一些近地卫星影像的分辨率高,标注精度稍低。即便如此,所提 DSNNet 模型仍然比经典网络模型效果好,说明 DSNNet 对于标签有较好的容错能力,可获得准确度更高的推理结果。

4 结 论

针对 UNet 和 UNet++ 神经网络模型对类别不平衡数据特征提取不足的问题,提出了 DSNNet 神经网络模型,该模型通过改变模型结构,设计新的桥接让更多困难样本特征图信息被多重损失函数捕获,解决了训练数据中难易样本数量不平衡问题,提升对数据稀疏区域的学习能力,使用早停法约束模型训练过程,有效提升训练效率,降低时间复杂度。实验结果表明,DSNet 的评价指标均有提高,预测图与真实标签误差最小。虽然数据集的标注精度对模型特征提取有一定影响,进而影响预测效果。但总体上,DSNet 模型有着更好的效果,说明 DSNNet 对于稀疏数据标签的提取能力更强,分割准确度较高,是一种行之有效的遥感语义分

割网络模型。

参 考 文 献

- [1] 徐辉, 祝玉华, 甄彤, 等. 深度神经网络图像语义分割方法综述[J]. 计算机科学与探索, 2021, 15(1): 47-59.
Xu H, Zhu Y H, Zhen T, et al. Survey of image semantic segmentation methods based on deep neural network[J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(1): 47-59.
- [2] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3431-3440.
- [3] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [4] Liu W, Rabinovich A, Berg A C. ParseNet: looking wider to see better[EB/OL]. (2015-06-15)[2021-02-03]. <https://arxiv.org/abs/1506.04579>.
- [5] Lin G S, Milan A, Shen C H, et al. RefineNet: multi-path refinement networks for high-resolution semantic segmentation[C]//2017 IEEE Conference on Computer

- Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5168-5177.
- [6] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6230-6239.
- [7] Fu J, Liu J, Tian H J, et al. Dual attention network for scene segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3141-3149.
- [8] Liang X D, Shen X H, Feng J S, et al. Semantic object parsing with graph LSTM[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 125-143.
- [9] Luc P, Couprie C, Chintala S, et al. Semantic segmentation using adversarial networks[EB/OL]. (2016-11-25)[2021-02-05]. <https://arxiv.org/abs/1611.08408>.
- [10] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [M]// Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [11] Oktay O, Schlemper J, Folgoc L L, et al. Attention U-Net: learning where to look for the pancreas[EB/OL]. (2018-04-11) [2021-02-03]. <https://arxiv.org/abs/1804.03999>.
- [12] 刘尚旺, 崔智勇, 李道义. 基于UNet网络多任务学习的遥感图像建筑地物语义分割[J]. 国土资源遥感, 2020, 32(4): 74-83.
Liu S W, Cui Z Y, Li D Y. Multi-task learning for building object semantic segmentation of remote sensing image based on UNet network[J]. Remote Sensing for Land & Resources, 2020, 32(4): 74-83.
- [13] 沈言善, 王阿川. 基于深度学习的遥感图像地物分割方法[J]. 液晶与显示, 2021, 36(5): 733-740.
Shen Y S, Wang A C. Remote sensing image feature segmentation method based on deep learning[J]. Chinese Journal of Liquid Crystals and Displays, 2021, 36(5): 733-740.
- [14] Zhou Z W, Siddiquee M M R, Tajbakhsh N, et al. UNet++ : redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE Transactions on Medical Imaging, 2020, 39(6): 1856-1867.
- [15] 袁伟, 周甜, 奚宗顺, 等. MUNet: 一种多尺度自适应的遥感语义分割深度学习网络[J]. 测绘科学技术学报, 2020, 37(6): 581-588.
Yuan W, Zhou T, Xi Z S, et al. MUNet: a multi-branch adaptive deep learning network for remote sensing image semantic segmentation[J]. Journal of Geomatics Science and Technology, 2020, 37(6): 581-588.
- [16] 张永洪, 席梦丹. 带洞型U-Net++网络在遥感影像中建筑物的提取方法[J]. 测绘地理信息, 2021, 46(S1): 82-86.
Zhang Y H, Xi M D. A method for extracting buildings from remote sensing images with hole U-Net++ network[J]. Journal of Geomatics, 2021, 46(S1): 82-86.
- [17] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2261-2269.
- [18] Milletari F, Navab N, Ahmadi S A. V-net: fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3D Vision (3DV), October 25-28, 2016, Stanford, CA, USA. New York: IEEE Press, 2016: 565-571.
- [19] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [20] Kingma D P, Ba J. Adam: a method for stochastic optimization[EB/OL]. (2014-12-22) [2021-02-03]. <https://arxiv.org/abs/1412.6980>.
- [21] Gugger S, Howard J. Adam W and super-convergence is now the fastest way to train neural nets[EB/OL]. [2021-02-03]. <https://www.fast.ai/2018/07/02/adam-weight-decay/>.