

文章编号: 2096-1618(2023)06-0673-08

# 基于改进 DeepLabV3p 的遥感图像中小目标分割方法

金芊芊, 罗建, 张晓倩, 杨梅, 李杨

(西华师范大学电子信息工程学院, 四川 南充 637009)

**摘要:**针对背景信息复杂、目标类别不均衡,遥感图像的中小目标在分割时常出现误检、漏检的问题,提出一种基于 DeepLabV3p 改进的遥感图像中小目标分割方法。采用 ResNet101 作为 DeepLabV3p 的骨干网络,提出多级感受野融合的 ASPP 模块,以获取更多感受野;添加 SE 注意力机制,使模型获得更加精准的通道信息;使用加权的 CrossEntropyLoss 和 LovaszSoftmaxLoss 损失函数进行训练,克服数据集目标不均衡的问题;使用全连接条件随机场对预测结果进行图像后处理,对模型输出进行精细化处理。实验结果表明,使用该方法对 DLRSD 数据集进行分割, mIOU 可达到 73.22%, 与基础网络相比提高了 3.78%, 有效提高了遥感图像中小目标的分割精度和准确率。

**关键词:** DeepLabV3p; 遥感图像; SE 注意力机制; ASPP; CRFs 全连接条件随机场; 混合损失函数

**中图分类号:** TP75

**文献标志码:** A

**doi:** 10.16836/j.cnki.jcuit.2023.06.009

## 0 引言

简单来说,遥感图像就是人们利用各类飞行器(包括无人机、航空卫星等)远距离对地观测所拍摄记录下来的图像,可以帮助我们实现对地观测的各项数据的分析。遥感图像的获取方式使其具有目标复杂多变、目标形态差异巨大、目标边缘不清晰等特征,因此在提高其分割精度和准确度时会遇到很多困难<sup>[1-2]</sup>。Maggiori 等<sup>[3]</sup>使用 FCN<sup>[4]</sup>全卷积神经网络来提取遥感图像的地面特征,相比于传统网络有较好的表现。Li 等<sup>[5]</sup>在 U-Net<sup>[6]</sup>基础上添加不对称卷积块提高了网络提取特征的能力。Zhao 等<sup>[7]</sup>提出的金字塔场景分析网络(PSPNet),应用金字塔池化模块获取更多的全局信息。Chen 等<sup>[8]</sup>提出的 DeepLabV3p 模型,使用可以扩充感受野的空洞卷积,减少网络中的池化操作,保留特征层的细节信息,还提出了 ASPP(atros spatial pyramid pooling)来提取不同尺度的信息,较大提升了模型的效果。DeepLabV3p 是当前较为流行的遥感图像分割算法之一。杨蜀秦等<sup>[9]</sup>在 DeepLabV3p 中添加多头自注意力机制来捕获无人机小麦倒伏图像中的倒伏特征;黄聪等<sup>[10]</sup>在 DeepLabV3p 中引入坐标注意力机制,并且采用联级特征融合的方法来更好地利用图像的语义特征信息。以上研究虽然对 DeepLabV3p 提出一定的改进,但是提升的效果有限,仍然难以提高对遥感图像中小目标的分割精度和准确度。

综上,找到一种更有效地综合利用多尺度信息、捕

捉细节信息的方式,是提高遥感图像分割精度的关键之一。因此,本文使用 DeepLabV3p 作为基础网络模型,并对其做了以下改进:(1)提出一种多级感受野融合的 ASPP 结构(ASPP with multi-level receptive field fusion, Rff-ASPP),获取更多感受野信息;(2)使用 SE 注意力机制<sup>[11]</sup>来聚焦目标区域提取全局信息,提高分割效率;(3)使用由 CrossEntropyLoss 和 LovaszSoftmaxLoss 损失函数<sup>[12]</sup>构成的混合损失函数替代原有的 CrossEntropyLoss,克服样本不均衡问题;(4)引入全连接条件随机场进行图像后处理,将神经网络输出的预测图做进一步处理,最终得到细节分割得更加出色的输出结果。

## 1 模型与改进

### 1.1 DeepLabV3p 模型

DeepLabV3p 是 DeepLab 系列语义分割网络模型的最新版本,是由 DeepLabV1<sup>[13]</sup>、DeepLabV2<sup>[14]</sup>、DeepLabV3<sup>[15]</sup>发展而来。它的整体结构是 encoder-decoder 结构,使用先前的 DeepLabV3 作为 encoder,然后在后面添加一个简单高效的 decoder 来进行不同尺度信息的融合。Encoder 把 Resnet101 作为骨干网络进行特征的初步提取,然后将初步提取的特征层送进 ASPP 模块进行加强特征提取,ASPP 通过不同大小采样率的空洞卷积来提取不同感受野的信息,最后拼接在一起得到加强特征层。Decoder 将骨干网络提取出来的低层特征和经过编码器处理的加强特征做进一步的特征融合,将不同尺度的细节信息和语义信息更好地结合

在一起,减少了信息的流失,最后再通过两次  $3\times 3$  的卷积对特征进行细化,通过 4 倍上采样恢复图像尺寸得到输出,DeepLabV3p 结构如图 1 所示。

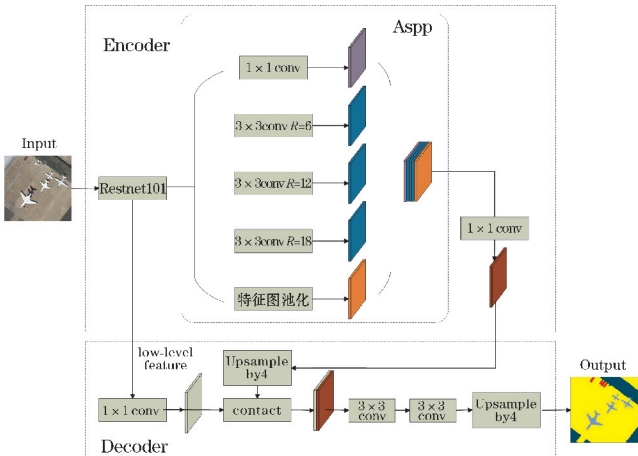


图 1 DeepLabV3p 模型结构图

1.2 改进的模型:RSC-DeepLabV3p

为提高 DeepLabV3p 模型对遥感图像中小目标的分割效果,本文在多个方面对模型进行改进,提出一种

改进的图像分割模型——RSC-DeepLabV3p。DeepLabV3p 的编码部分由 Restnet101 骨干网络和 ASPP 组成,为提高模型对多尺度信息的利用,对 ASPP 模块进行改进,提出一种新的多级感受野融合的 ASPP 结构。Rff-ASPP 将 ASPP 中通过不同采样率卷积获得的 5 个特征层,从上往下依次通过 contact 函数进行融合,使其在获得不同感受野信息的同时也实现了信息的共享,得到更大范围的感受野。在解码部分,添加 SE 注意力机制来聚焦目标区域,用来更好地提取全局信息,提高分割效率。

对损失函数也进行了优化,使用加权的 CrossEntropyLoss 和 LovaszSoftmaxLoss 损失函数相结合替代原有的多分类交叉熵损失函数。LovaszSoftmax 损失函数适用于样本不均衡的数据集,并且对于小目标比较友好,与其他损失函数一起加权使用时效果较好。为优化从改进 DeepLabV3p 卷积神经网络得到的预测图,引入全连接条件随机场进行细分,不仅可以提高模型分割后的结果,而且算法复杂度也较低,RSC-DeepLabV3p 模型结构如图 2 所示。

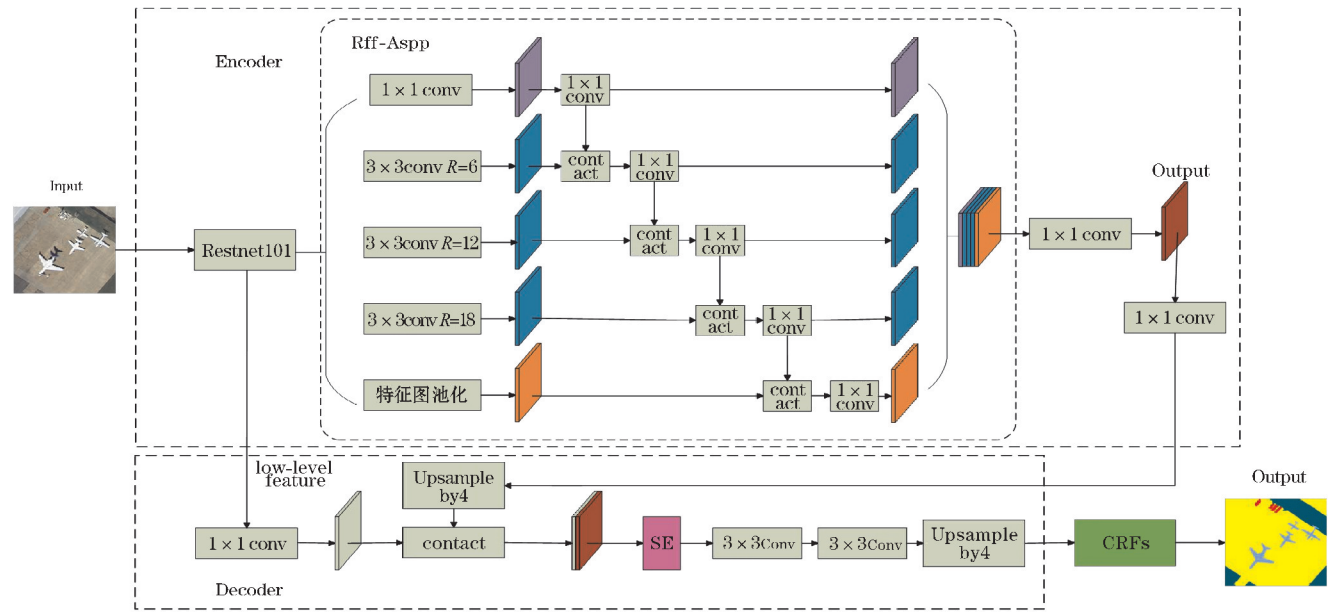


图 2 RSC-DeepLabV3p 模型结构图

1.3 SE 注意力机制

SE 注意力机制主要是仿照人类的眼睛对图像重要信息的“聚焦”,只是“聚焦”的是特征层中重要的通道特征。在模型训练过程中,SE 模块通过自适应学习找出比较重要的通道特征,并且赋予其更大的权重,使模型对目标特征进行关注。这不仅可以节省计算量,还保证其在各种网络架构中的通用性。其结构如图 3 所示。

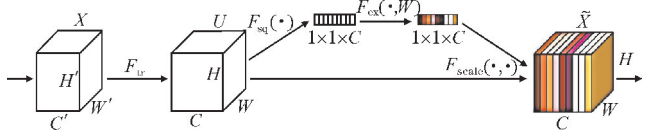


图 3 SE 注意力机制结构图

SE 注意力的实现主要有以下 3 个步骤:

(1) Squeeze:对获取到的多个特征图进行全局平均池化,把每个通道的特征压缩为一个数来表示,获取到每个通道的全局信息。其表达式:

$$Z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j)$$

式中,  $H$  为特征图的高,  $W$  为特征图的宽,  $c$  为特征通道数。

(2) Excitation: 通过两个全连接层和一个 sigmoid 激活函数, 实现权重的学习, 从而找出比较重要的特征信息, 减少对其他信息的关注。其表达式:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \text{ReLU}(W_1 z))$$

式中,  $\sigma$  代表 sigmoid 激活函数,  $W_1$ 、 $W_2$  代表两个全连接层。

(3) Reweight: 将前面所得的权重赋值到各特征图, 即将学习的权重与各个通道的原始特征相乘。其表达式:

$$\tilde{x}_c = \text{Fscale}(u_c, s_c) = s_c \cdot u_c$$

式中,  $s$  代表学习到的权重信息,  $u$  为原始特征图。

SE 注意力机制最终能实现通道特征的增强, 并且不改变其大小。因此, SE 注意力机制能够提高模型的精度, 并且不改变特征图的大小。这样, 在网络中添加 SE 模块不用添加参数或者改变结构, SE 常被用在各种网络模型中。本文方法将 SE 注意力模块添加到 decoder 中层卷积之前, 提高了模型对于全局特征的提取能力。

#### 1.4 改进 ASPP

ASPP 由多个不同采样率的空洞卷积并行卷积然后拼接组成, 目的是为进一步提取多尺度信息。ASPP 顶部使用  $1 \times 1$  的卷积提取特征层, 第二、三、四层则使用采样率为 6、12、18 的空洞卷积进行采样, 提取具有不同大小的感受野的特征层, 最后一层使用全局平均池化, 目的是提取全局信息; 然后, 通过 contact 函数拼接在一起; 最后, 通过  $1 \times 1$  的卷积调整特征层的形状。

本文的多级感受野融合的 ASPP 结构是在 FPN<sup>[16]</sup> 的启发下提出的, FPN 是 2017 年提出的一种网络。FPN 提出了一种新的想法, 就是将通过下采样获得的高层特征进行上采样然后与低层特征逐层相连接, 并且在每一层都进行预测, 这样就使得低层特征图得到了更充分的利用。改进 ASPP 的结构为沿着 ASPP 模块从上往下的方向, 首先对最顶层进行一个  $1 \times 1$  的卷积, 然后通过 contact 函数将该特征层和相邻特征层拼接在一起, 接着对拼接后的特征层重复刚才的操作, 直到最后的池化层。这样就实现了不同感受野的特征层的信息共享, 提高对感受野的利用。改进后的 ASPP 结构如图 4 所示。

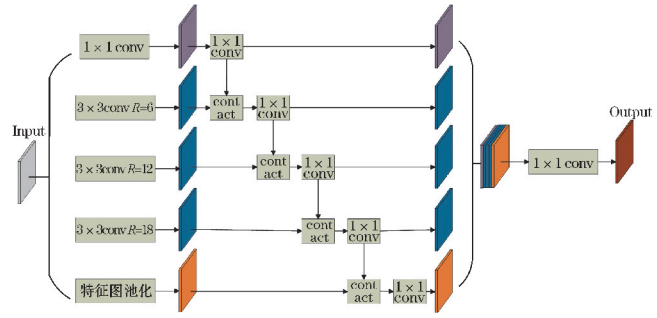


图4 改进的 ASPP (Rif-ASPP)

#### 1.5 优化损失函数

损失函数是用来对模型训练效果好坏进行评判的方法, 其值越小, 模型训练效果就越好。遥感图像包含的目标又多又杂, 并且类别十分不均衡, DeepLabV3p 模型原有的多分类交叉熵损失函数虽然可以处理多分类分割问题, 但是对于遥感图像这种背景像素所占比例较大且中小目标较多的图像, 训练效果较差。

为了达到更好的分割效果, 将 CrossEntropyLoss 和 LovaszSoftmaxLoss 损失函数加权结合来训练改进后的 DeepLabV3p 网络模型。Berman 等<sup>[12]</sup> 提出 LovaszSoftmaxLoss 用于样本不均衡的数据集, 并且对于小目标比较友好, 但是使用该损失函数优化时主要针对的是 image-level mIoU, 适用于微调的过程。因此 LovaszSoftmaxLoss 与其他损失函数一起加权使用时效果更优。

多分类交叉熵损失函数计算公式如下:

$$L_1 = -\frac{1}{p} \sum_{i=1}^p \lg f_i(y_i^*)$$

式中,  $p$  为输入的像素数量,  $y_i^*$  为像素  $i$  的类别,  $f_i(y_i^*)$  为像素  $i$  的概率估计。

Softmax 函数公式:

$$f_i(c) = \frac{e^{F_i(c)}}{\sum_{c' \in C} e^{F_i(c')}}, \forall i \in [1, p], \forall c \in C$$

式中,  $c$  为对象类别, 一个像素  $i$  的类别为  $\tilde{y}_i = \arg\max_{c \in C} F_i(c)$ 。

Jaccard 索引, 即 IOU 值, 图像语义分割中常使用 IOU 作为评价指标, 计算公式:

$$J_c(y^*, \tilde{y}) = \frac{|(y^* = c) \cap (\tilde{y} = c)|}{|(y^* = c) \cup (\tilde{y} = c)|}$$

LovaszSoftmaxLoss 的作者使用 Jaccard 索引作为损失函数, 因此最终的 LovaszSoftmaxLoss 表达式:

$$\Delta J_c(y^*, \tilde{y}) = 1 - J_c(y^*, \tilde{y})$$

式中,  $y^*$  为真实值,  $\tilde{y}$  为预测值。

最终的损失函数为 CrossEntropyLoss 和 LovaszSoftmaxLoss 加权结合的混合损失函数, 因此最终表达式:

$$\text{Loss} = a \cdot L_1 + b \cdot \Delta J_c(y^*, \tilde{y})$$

式中,  $a$ 、 $b$  为权重,  $a+b=1$ 。



1.6 全连接条件随机场

全连接条件随机场 CRFs 能够对模型预测的分割结果做进一步的处理,以改善分割图像中的边缘模糊问题<sup>[17]</sup>。一张图片由很多个像素点组成,CRFs 做的工作就是找出原始图片中各个像素点的位置信息以及颜色信息,结合改进 DeepLabV3p 模型的预测概率分布图来计算像素点归类概率,然后进行推理,找出每个像素最可能的类,推理次数不同,优化效果也就不同。其推理流程如图 5 所示。

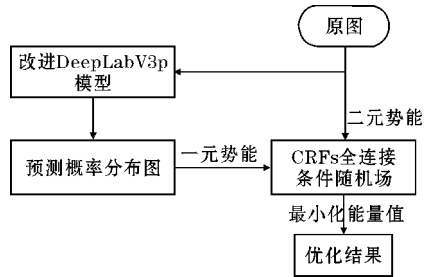


图 5 CRFs 流程图

全连接条件随机场符合吉布斯分布,如下:

$$p\left(x=X \mid I=\frac{1}{z(I)} e^{-E(X \mid I)}\right)$$

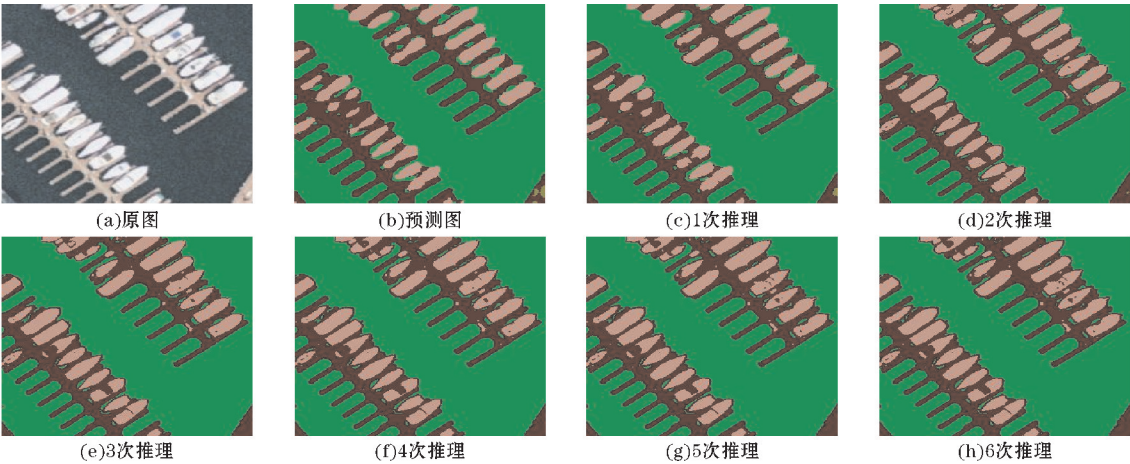


图 6 原图与不同推理次数优化图对比

2 实验结果与讨论

2.1 数据集与实验环境

使用 DLRSD 数据集,该数据集是武汉大学基于 UC Merced\_LandUse 数据集标注的,包括 17 个类别,共 2100 张图片,且大小均为 256×256,其中包括建筑物、飞机、船、汽车、树木等中小目标类别。按照 7 : 2 : 1 将数据集划分为训练集、验证集以及测试集,其中对训练集采用缩放、随机裁剪、随机水平翻转、归一化等数据预处理操作。

式中, $x$  是观测值, $E(X|I)$  是能量函数,该能量函数由一元势函数  $\psi_u(x_i)$  和二元势函数  $\psi_p(x_i, y_i)$  构成,如下:

$$E(X|I)=\sum_i \psi_u(x_i)+\sum_{i,j} \psi_p(x_i, y_j)$$

一元势函数  $\psi_u(x_i)$  通过模型输出的预测图计算像素点  $i$  的类别为  $x_i$  的概率,二元势函数  $\psi_p(x_i, y_i)$  则是通过原图中像素点的颜色值、像素点间的距离等来判断两个像素之间的关系,全连接条件随机场中的二元势函数计算每一个像素点和其他所有像素点之间的关系。通过推理,使得  $E(x)$  能量值越小,则最终相似像素点为相同类别的概率就越大。通过反复迭代最小化能量函数,可以得到最终的后处理结果。

本文将原始特征图与模型输出的预测图放入全连接随机场进行实验,进行一元势和二元势,然后通过设定推理次数来确定优化次数。利用全连接条件随机场对模型输出的预测图进行 6 次推理,推理后的对比结果如图 6 所示。由图 6 可以看出,前 3 次推理取得的效果较好,后 3 次由于过度推理,受颜色信息影响过大,反而优化效果不好。因此,本文选择使用 3 次推理的 CRFs 全连接条件随机场对模型预测图进行优化。

本文实验均在百度的 AI Studio 平台进行,选用 PaddlePaddle 2.3.0 框架,选用硬件环境如表 1 所示。

表 1 实验的硬件环境

硬件名称	硬件型号/大小
GPU	Tesla V100
Video Mem	32GB
CPU	4 Cores
RAM	32GB
Disk	100GB

使用 SGD 优化算法,PolynomialDecay 学习策略,实验的各项参数如表 2 所示。

表 2 实验的训练参数

参数名称	参数大小
learning_rate	0.005
End learning_rate	0
batchsize	4
iters	40000
momentum	0.9
weight_decay	4.0e-5

2.2 实验评价指标

使用 mIoU (交并比)、Dice (相似系数)、Kappa 系数、ACC (准确率) 作为模型性能的评价指标。这几个值越大,则模型的性能越好。各指标的计算公式如下:

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{t_p}{t_p + f_n + f_p}, \quad ACC = \frac{\sum t_p}{N}$$

$$Kappa = \frac{p_o - p_e}{1 - p_e}, \quad Dice(A, M) = \frac{2(A \cap M)}{|A| + |M|}$$

式中,  $k$  表示类别,  $k+1$  表示加上背景类,  $t_p$  表示预测值

属于该类别并且真实值也属于该类别的像素点数目,  $f_n$  表示预测是此类别但是真实类别是另一种的像素点数目,  $f_p$  表示预测为其他类别而真实值属于此类别的像素点数目;  $A$  表示预测图,  $M$  表示真实标签图;  $p_o$  为分类器的准确率,  $p_e$  为随机分类器的准确率;  $N$  表示总的像素点数目。

2.3 实验结果

2.3.1 对比实验结果与分析

基础 DeepLabV3p 与改进后的 DeepLabV3p 在验证集上的 mIoU 值和在训练集上的 loss 值曲线如图 7 所示,蓝色线代表本文提出的 RSC-DeepLabV3p 模型,黄色线代表 DeepLabV3p 模型。损失值随着迭代次数增加而降低,mIoU 值随之增加,当训练到 1000 个 Iters 时,mIoU 值与 loss 值均开始收敛。改进后的 DeepLabV3p 不仅在损失值和 mIoU 值曲线的收敛速度上表现更好,并且最终的损失值更低,mIoU 值更高。

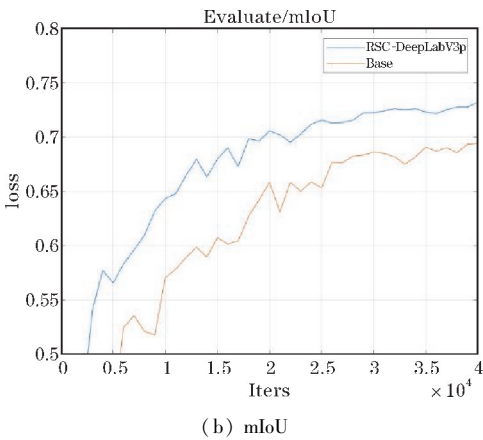
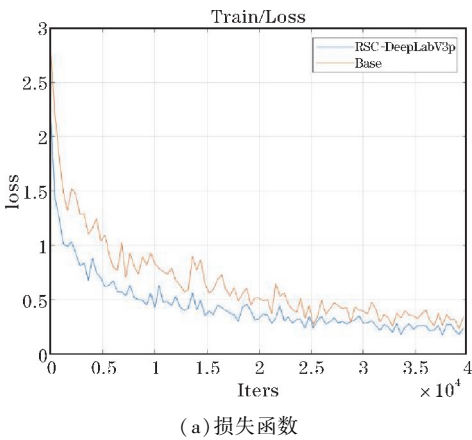


图 7 改进前后 DeepLabV3p 的损失函数和 mIoU 值曲线

为进一步证明改进的 DeepLabV3p 模型的确在遥感图像中小目标的分割上表现优秀,对当前常用的一些分割网络如 U-Net、FCN、PSPNet 也进行对比实验,实验环境、数据集、训练参数等实验条件均一致,实验结果如表 3 所示。由表 3 可知, Deeplabv3p 模型在 DLRSD 数据集上的综合表现均优于其他模型,而本文

改进的 DeepLabV3p 模型在各方面表现都优于 DeepLabV3p 模型, mIoU 为 73.22%、准确率为 86.48%、Kappa 系数为 84.44%、dice 为 82.03%,其中 mIoU 和 dice 较基础模型 DeeplabV3p 提高了 3.78%、2.58%。说明本文改进的方法分割效果较好,有利于遥感图像中小目标的分割。

为更直观地体现改进模型的性能,分别使用不同模型进行推理,并对其中部分结果进行可视化分析。图 8 的 4 组图片是各模型分别对包含飞机、住宅、港口等目标的遥感图像进行分割的结果。可以看出,与 U-Net、FCN 等相比,DeepLabV3p 模型在分割轮廓上表现更好。如图 8(a) 所示,DeepLabV3p 完整地分割出了港口上每一艘船的轮廓,其中穿插的道路的分割形状也比较完整,而 U-Net、FCN 分割边缘十分粗糙,FCN 只能分割出船只的大体位置,且部分道路分割不完整;

表 3 不同网络模型评估对比 单位: %

模型名	评价指标			
	mIoU	ACC	Kappa	DICE
U-Net	62.17	79.64	76.49	74.18
FCN	67.54	83.65	81.10	77.64
PSPNet	56.70	78.30	75.10	69.56
GCNet	57.49	78.77	75.69	70.41
Deeplabv3p	69.44	84.83	82.51	79.45
RSC-DeepLabV3p	73.22	86.48	84.44	82.03

与 DeepLabV3p 相比, RSC-DeepLabV3p 则表现更好, 本文模型在船只和道路的细节上分割得更加清楚, 每

一艘船中间间隔的道路也能够分割出来, 整体表现最好。

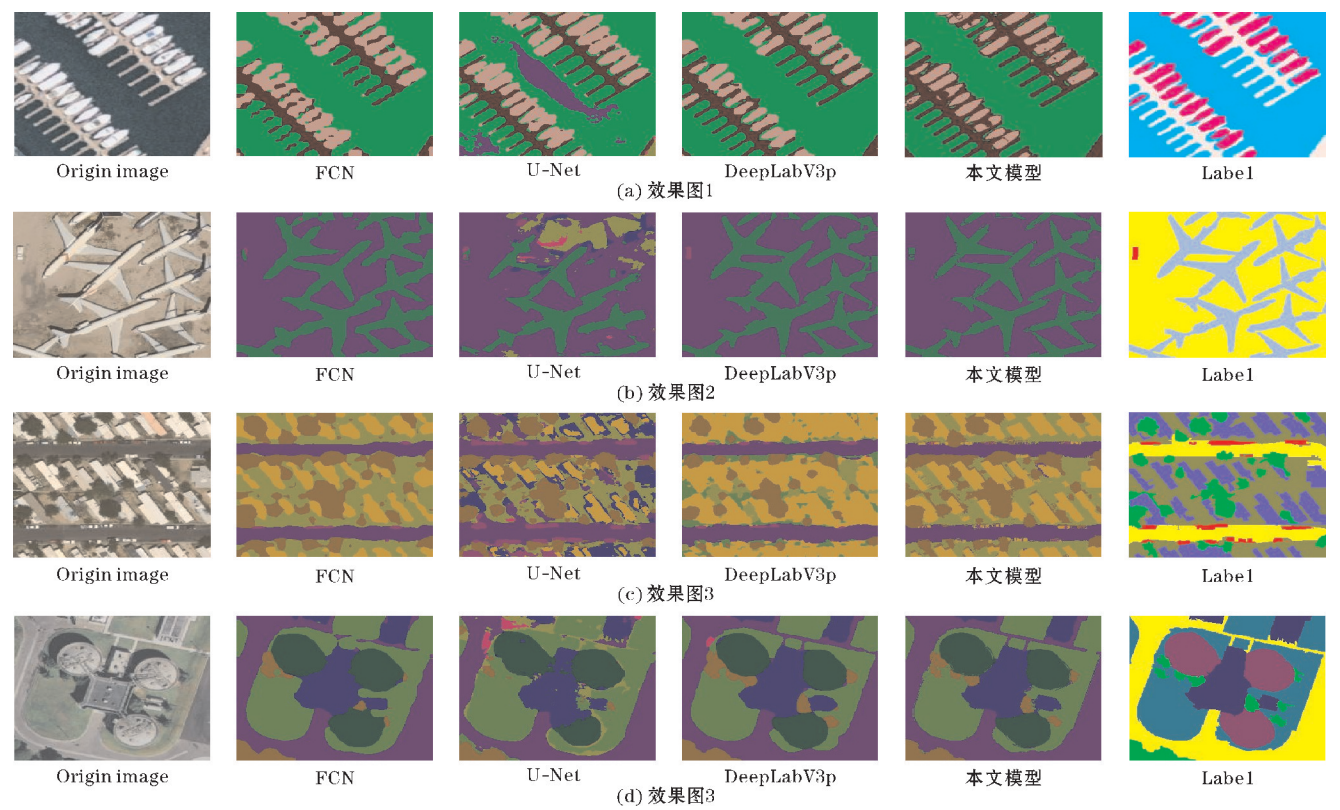


图8 不同模型推理可视化

2.3.2 消融实验结果及分析

为验证改进模块的有效性, 分别添加改进模块逐个进行实验, 以验证多级感受野融合的 ASPP、Mixed-Loss、SE 模块对模型性能的提升效果。由表 4 的实验结果可知, 各改进方案对模型性能均有提升。改进 ASPP, 能够提高不同感受野的利用率, mIoU 较原模型

提高了2.12%; 添加 SE 模块, 提高了模型在通道上的信息的聚焦, mIoU 较原模型提高了1.75%; 使用 MixedLoss, 模型收敛速度加快, 改善了样本不均衡的问题, mIoU 较原模型提高了2.26%。每添加一个改进模块, 模型的总体性能都有所提高, 最终结合所有改进模块的模型表现最好, mIoU 较基础模型提高3.78%。

表4 消融实验

模型	添加模块			评价指标			
	Rff-ASPP	MixedLoss	SE	mIoU/%	ACC/%	Kappa/%	DICE/%
Base				69.44	84.83	82.51	79.45
1	√			71.60	85.49	83.28	80.94
2		√		71.70	85.47	83.29	80.99
3			√	71.19	85.62	83.46	80.75
4	√	√		72.18	85.87	83.73	81.36
5	√		√	71.78	85.72	83.55	81.08
6		√	√	72.25	85.92	83.81	81.45
RSC-DeepLabV3p	√	√	√	73.22	86.48	84.44	82.03

3 结论

通过对比实验发现, 当前常用的语义分割模型在

遥感图像中小目标的分割上还存在不足, 比如边缘分割模糊, 细节分割不到位等, 不利于各项遥感工作的开展。本文针对这一问题, 使用在遥感图像分割中表现较好的 DeepLabV3p 模型作为基础模型, 对其进行多



方面的改进,得出以下结论:

(1)改进的多级感受野融合 ASPP 能够进一步对不同采样率卷积出来的特征层进行感受野融合,更大程度发挥了 ASPP 模块的作用。

(2)在解码器添加 SE 注意力机制能够使模型对通道信息中的目标信息更加专注,减少了其他通道信息对训练的干扰。

(3)使用加权的混合损失函数(LovaszSoftmaxLoss 和 CrossEntropyLoss),结合二者的长处,样本不均衡问题得到了改善。

(4)全连接条件随机场对模型预测图进行精细化处理,可以改善物体分割边缘粗糙的问题。

(5)融合以上的改进方案能够更大程度提高模型的性能,为遥感图像中小目标分割的边缘不清晰、细节不准确的问题提供了一种解决方案,本文模型最终 mIoU 提高3.78%。

当然,因为网络模型的训练过程十分复杂并且缺乏可解释性,本文的研究还不够深入,在后续工作中将重点研究如何在不增加复杂计算量的情况下实现更高精度的分割。在进行 CRFs 图像后处理时发现,推理次数较少时,优化效果不够明显,推理次数较多时又容易受到颜色信息的影响,因此后续还将对图像后处理算法进行研究和优化。

## 参考文献:

- [1] 董蕴雅,张倩. 基于 CNN 的高分遥感影像深度语义特征提取研究综述[J]. 遥感技术与应用, 2019,34(1):1-11.
- [2] 尹昊. 基于卷积神经网络的遥感图像分割方法研究[D]. 泰安:山东农业大学,2021.
- [3] Maggiori E, Tarabalka Y, Charpiat G, et al. Fully convolutional neural networks for remote sensing image classification[C]. 2016 IEEE international geoscience and remote sensing symposium (IGARSS). IEEE,2016;5071-5074.
- [4] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015;3431-3440.
- [5] Li R, Zheng S, Duan C, et al. Multistage attention ResU-Net for semantic segmentation of fine-resolution remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters. 2021,19:1-5.
- [6] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]. International Conference on Medical image computing and computer-assisted intervention. Springer, Cham,2015;234-241.
- [7] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017;2881-2890.
- [8] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. Proceedings of the European conference on computer vision(ECCV). 2018;801-818.
- [9] 杨蜀秦,王鹏飞,王帅,等. 基于多头自注意力 DeepLab v3+的无人机遥感影像小麦倒伏检测[J/OL]. 农业机械学报;1-11[2022-11-09].
- [10] 黄聪,杨璐,刘毅,等. 基于改进 DeepLabV3p 的遥感图像分割算法[J/OL]. 电子测量技术;1-7[2022-11-09].
- [11] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018;7132-7141.
- [12] Berman M, Triki A R, Blaschko M B. The lovasz-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018;4413-4421.
- [13] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv preprint arXiv:1412.7062,2014.
- [14] Chen L C, Papandreou 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. 2017,40(4):834-848.
- [15] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmen-

tation[J]. arXiv preprint arXiv:1706.05587,2017.

[ 16 ] Lin T Y,Dollár P,Girshick R,et al. Feature pyramid networks for object detection[ C ]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017:2117–2125.

[ 17 ] 高宇田,杨阳,赵广帅,等. 基于全连接条件随机场的图像语义分割算法研究[ J]. 长春理工大学学报(自然科学版),2021,44(2):60–66.

## Small Target Segmentation Method in Remote Sensing Image based on Improved DeepLabV3p

JIN Qianqian, LUO Jian, ZHANG Xiaoqian, YANG Mei, LI Yang  
(School of Electronic Information Engineering,Xihua Normal University,Nanchong 637009,China)

**Abstract:**Due to the complexity of background information and the imbalance of target categories, small and medium-sized targets in remote sensing images are often subject to false detection and missing detection in segmentation. In order to solve this problem, an improved segmentation method for small and medium-sized targets of remote sensing images is proposed, which is based on DeepLabV3p. The ASPP module of multi-level receptive field fusion is proposed to obtain more receptive fields. In the decoding part, Adds SE attention mechanism to enable the model to obtain more accurate channel information. The weighted Cross Entropyloss function and LovaszSoftmaxLoss function are used for training. Finally, the CRFs is used for image post-processing of the prediction results, and the model output is refined. The experimental results show that using this method to segment images in DLRSD dataset, the mIOU can reach 73.22%,which is 3.78% higher than that of the basic network, effectively improves the segmentation precision and accuracy of small and medium targets in remote sensing images.

**Keywords:**DeepLabV3p;remote sensing images;SE attention mechanism;ASPP;CRFs;mixed loss function