增强特征提取和解释的遥感图像语义分割模型

于攀琳,吴　旭*,张凌云,刘子涵
(成都理工大学 计算机与网络安全学院,四川 成都 610059)

摘　要:DeepLab V3+是一种具有 Encoder-Decoder 结构的语义分割模型,因其逐像素分类的特点适用于处理遥感图像的土地覆盖分类问题。然而,下采样过程导致的特征图损失,会使连续大尺度地物的内部出现不连续无标签空洞区域,并且双线性插值算法会丢失分割边缘细节。针对上述问题,该文提出了一种基于 DeepLab V3+改进的 V3plus-EN-TC 模型。将骨干网络替换为特征提取能力更强的 EfficientNet,引入 SE 模块和倒置残差链接,增强 Encoder 对通道信息和多尺度空间信息的感知与提取能力;融合三个层次的特征,并且采用转置卷积和双线性插值结合的上采样方法,提高 Decoder 的特征解释能力,抑制空洞区域的出现,提高边缘精度;利用 DiceFocal 联合损失函数,解决样本分布不平衡问题,并且进一步聚焦混合像元。改进模型 V3plus-EN-TC 在预处理后的遥感数据集 GID 上,较 FCN、U-Net、SegNet、PSPNet、CBAM-DeepLab V3+、CRF-DeepLab V3+等模型,空洞区域显著减少,模型精度提升。改进模型的平均交并比、F_1 分数、平均像素精度分别达到了 84.74%、88.39%、86.64%。

关键词:遥感图像;语义分割;DeepLab V3+;EfficientNet;转置卷积;损失函数
中图分类号:TP753　　**文献标识码**:A　　**文章编号**:1673-629X(2025)07-0008-08
doi:10.20165/j.cnki.ISSN1673-629X.2025.0039

A Semantic Segmentation Model for Remote Sensing Images with Enhanced Feature Extraction and Interpretation

YU Pan-lin, WU Xu*, ZHANG Ling-yun, LIU Zi-han
(School of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu 610059, China)

Abstract: DeepLab V3+ is a semantic segmentation model with an Encoder-Decoder structure. Due to its characteristic of per-pixel classification, it is suitable for dealing with the land cover classification problem of remote sensing images. However, the loss of feature maps caused by the downsampling process will lead to the appearance of discontinuous unlabeled void areas inside continuous large-scale ground objects, and the bilinear interpolation algorithm will lose the details of segmentation edges. In response to the above problems, we propose a V3plus-EN-TC model improved based on DeepLab V3+. The backbone network is replaced with EfficientNet, which has a stronger feature extraction ability. The SE module and inverted residual connections are introduced to enhance the Encoder's ability to perceive and extract channel information and multi-scale spatial information. Features at three levels are fused, and an upsampling method combining transposed convolution and bilinear interpolation is adopted to improve the feature interpretation ability of the Decoder, suppress the appearance of void areas, and improve the edge accuracy. The DiceFocal combined loss function is utilized to solve the problem of unbalanced sample distribution and further focus on mixed pixels. On the preprocessed remote sensing dataset GID, compared with models such as FCN, U-Net, SegNet, PSPNet, CBAM-DeepLab V3+, and CRF-DeepLab V3+, the improved model V3plus-EN-TC has significantly fewer void areas and improved model accuracy. The mean intersection over union, F_1 score, and mean pixel accuracy of the improved model reach 84.74%, 88.39%, and 86.64% respectively.

Key words: remote sensing image; semantic segmentation; DeepLab V3+; EfficientNet; transposed convolution; loss function

0　引　言

精确的土地覆盖分类结果能够为生态经济发展和城市规划提供数据支持[1-2]。语义分割通常采用像素级(端到端)分割方法来处理遥感图像的土地覆盖分

收稿日期:2024-10-31　　**修回日期**:2025-03-05
基金项目:四川省科学研究项目(25NSFSC1726)
作者简介:于攀琳(1998-),女,硕士研究生,研究方向为人工神经网络与遥感;通讯作者:吴　旭(1979-),男,讲师,博士,研究方向为计算机与地质交叉学科。

类任务[3]。在深度学习出现之前,研究人员普遍采用机器学习方法来处理分割任务[4-5],但这些方法往往忽略了遥感图像的空间复杂性以及通道相关性。例如,草地上不同植物物种可能表现出光谱差异性,而草地和森林可能表现出光谱相似性,这一现象在遥感领域被称为“同类不同谱,同谱不同类”。为了打破机器学习的局限性,研究人员提出使用深度学习技术进行遥感图像土地覆盖分类任务,如多层感知器(MLP)、卷积神经网络(CNN)、循环神经网络(RNN)和 Transformer。这些方法具有更高的鲁棒性、容错性和特征处理能力[6]。特别是基于 CNN 的 U-Net、SegNet、PSPNet、DeepLab 等模型,由于它们强大的特征提取能力,被广泛应用于遥感图像的处理[7-8]。

近年来,DeepLab V3+成为了语义分割领域的研究热点之一。DeepLab V3+[9]整体采用 Encoder-Decoder 架构,并使用深度可分离卷积来加速训练过程,引入空洞空间金字塔池(Atrous Spatial Pyramid Pooling,ASPP)模块,通过不同的空洞率的组合来增强对不同尺度特征的识别能力。图像首先经过 Encoder 的深度卷积神经网络(DCNN)和 ASPP 模块提取多尺度特征。输出再进入 Decoder 与提取自 DCNN 的浅层特征 Concat 融合,使用双线性插值算法上采样恢复成原始图像大小。该模型在大多数语义分割数据集上的表现都优于其他模型[10-11],但是其结构是针对一般图像的分割来设计的,想要在遥感图像上获得较高的精度,还需要对模型进一步优化。

Chen 等人[12]将深度可分离的 MobileNet 作为 DeepLab V3+的骨干网络,引入混合空洞卷积模块和注意力机制。Quan 等人[13]基于 DeepLab V3+提出一种结合 U-Net 融合浅层特征的模型。Zheng 等人[14]基于注意力优化机制改进数据增强策略,并且优化 ASPP 模块空洞率的组合。上述改进考虑了模型复杂度、特征图分辨率、空间信息与通道信息的关联性、数据可靠性等影响模型精度的因素,证明增强了特征提取能力的模型能够达到更高的整体精度。然而,其分割结果中仍然存在空洞区域,以及粗糙的边缘像素划分问题。

为了实现更精确的土地覆盖分类,该文提出一种基于 DeepLab V3+模型改进的人工神经网络模型 V3plus-EN-TC。主要贡献包括:(1)提出一个增强 Encoder 特征提取能力和 Decoder 解释能力的人工神经网络模型用于遥感图像语义分割任务;(2)采用更针对样本分布不平衡问题的联合损失函数来训练模型;(3)在预处理后的遥感数据集上比较其他模型与文中模型的精度衡量指标,数据结果表明文中模型在遥感图像上较原模型表现更好。

1　方　法

1.1　改进模型 V3plus-EN-TC

V3plus-EN-TC 整体架构如图 1 所示。本研究首先将骨干网络替换为 EfficientNet,使 Encoder 具有更强的特征提取能力,并针对遥感图像的特点对模块进行优化。引入压缩奖励(Squeeze and Excitation,SE)机制,让模型着重关注携带重要信息的通道,骨干网络使用空洞卷积来增大感受野,利用倒置残差连接保持特征的维度。此外,提取出浅层和中间层特征参与后续计算。然后,在 Decoder 部分融合更多层级的特征图,采用转置卷积与双线性插值结合的方法对融合后的特征图进行上采样。使模型能够学习到更多尺度特征之间的联系,减少空洞区域的产生,细化边缘混合像元的分类,提高 Decoder 的特征解释能力。最后,选择 DiceFocal 联合损失函数作为模型训练的依据。Dice 损失减小样本中标签分布不平衡对模型训练过程的影响;Focal 损失让模型更关注不易分割的混合像元。

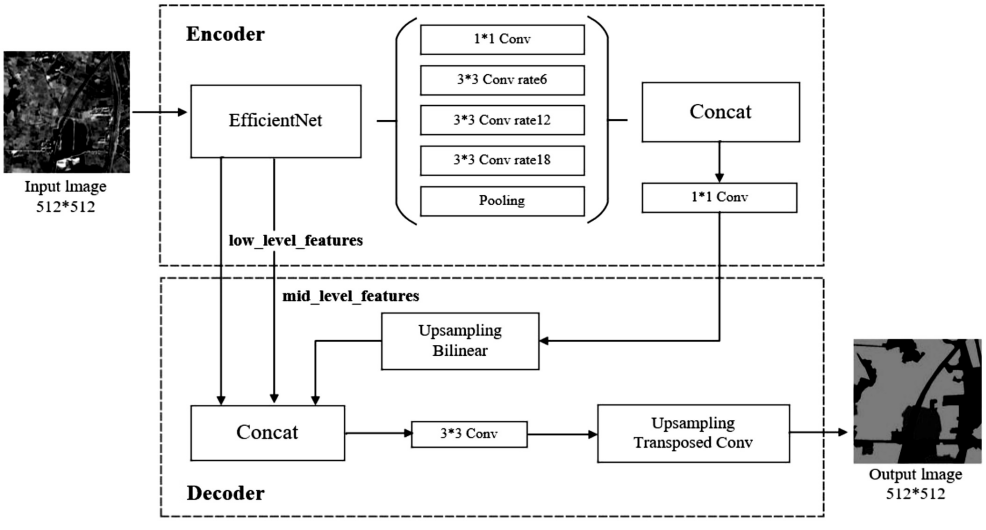

图 1　V3plus-EN-TC 整体架构

1.2 骨干网络

骨干网络的作用是从输入图像中学习和提取特征。原模型主干网络使用的是改进的 Xception,其可以被看作是带有残差连接的深度可分离卷积 block 的堆叠。但是与传统的深度可分离卷积不同的是,Xception 交换了逐点卷积和逐通道卷积的顺序,先使用 1 * 1 卷积处理跨信道信息,再使用不同卷积核的组合处理空间信息。此外,骨干网络还结合批归一化(Batch Normalization,BN)和 ReLU 激活函数,在一般图像上实现了超过 85% 的平均交并比(MIoU)[15],然而当其应用于遥感图像时,模型精度有所下降。问题主要在两个方面:本应该连续的大尺度地物内部出现了离散的小块无标签空洞区域;处于地物边缘的混合像元包含多个类别的地物,辨别这些像元的类别一直是土地覆盖分类任务中的一个挑战[16]。因此,在深度卷积层中学习不同尺度的特征对于增强分割性能至关重要,骨干网络提取不同尺度特征的能力是影响模型精度的关键因素之一[17]。

目前国内外研究中 DeepLab V3 常用的主干网络有 ResNet 和 MobileNet。赵玉刚等人[18]以 ResNeSt 作为主干,获得了更高的分割精度。马静等人[19]将 MobileNet 作为主干,减少训练时间和模型计算量。近年来,EfficientNet[20]因其强大的信息提取能力,也逐渐被应用于遥感图像的处理。梁伟[21]将 YOLO-V4 的主干网络替换为 EfficientNet 来优化小目标的检测精度,刘浪等人[22]采用 EfficientNet-B0 作为主干网络来检测不同场景下的舰船,这些模型在遥感数据集上均获得了比原模型更好的结果。

为了提高原模型在遥感数据集上的表现,该文将原骨干网络替换为 EfficientNet。该网络结合 ResNet 的残差连接和 MobileNet 的深度可分离卷积,主要组成部分是倒置线性瓶颈(MBConv)block,其结构如图 2 所示,减少了参数量和浮点运算量。该 block 引入空洞卷积、SE 模块以及倒置残差连接。空洞卷积增大感受野,降低多重卷积核带来的计算开销,并且不同空洞率的组合,让 DCNN 可以提取到多尺度特征。SE 模块是一种通道注意力机制,给各个通道分配权重,使模型能够更关注重要的通道。其计算过程如图 3 所示。先用卷积 F_{tr} 调整特征图的宽、高和通道数(分别由 W、H 和 C 表示);再由全局平均池化 F_{sq} 将特征映射压缩成长度为 C 的包含上下文信息的向量;然后由两个全连接层组成的 F_{ex},先将向量的 C 个通道压缩为 $\frac{C}{r}$ 个通道,再使用 Sigmoid 激活函数,并将通道数恢复为 C 以获得每个通道的注意力权重;最后 F_{scale} 将权重应用于每个通道的特征,从而获得输出。一般残差连接是先降维再升维,倒置残差连接是先升维再降维,可以在解决梯度消失和爆炸问题的同时,维持特征的维度和多样性。这些模块的加入目的是使 DCNN 具有更强的多尺度特征感知能力和特征提取能力[23]。

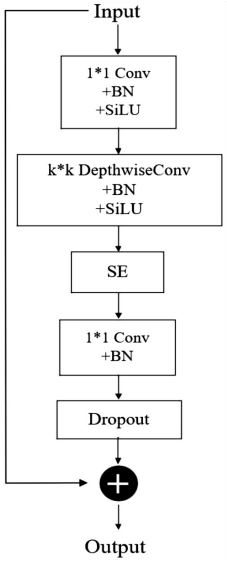

图 2 MBConv block 结构

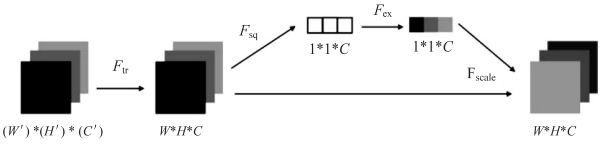

图 3 SE 模块计算过程

1.3 结合转置卷积与双线性插值的上采样

上采样是将图像放大的过程,常用的上采样方法包括插值法、转置卷积、反池化等。插值法计算量小于其他两种方法,是人工神经网络中最常用的上采样方法[24]。DeepLab V3+的 Decoder 使用双线性插值将特征图还原成输入的大小,但是双线性插值会使上采样后的图像边缘变得模糊,增大模型分割混合像元的难度。转置卷积与反池化都是基于数据的方法,能够充分考虑到空间和通道信息,而转置卷积比反池化更能注意到复杂的细节特征,更适用于处理遥感图像。因此,该文选择转置卷积作为 Decoder 的第二次上采样操作,将融合多层级特征后的特征图还原至输入大小。此外,增加中间层特征参与融合,提供更多层级的空间特征。原模型融合骨干网络的浅层特征图和 ASPP 输出的特征图,Decoder 无法解译出足够的信息,改进后的 Decoder 一共融合三种特征图:从 EfficientNet 提取的 256 * 256 * 48 浅层特征和 64 * 64 * 48 的中间层特征,以及 ASPP 输出的 16 * 16 * 256 的特征图。先使用双线性插值全部上采样到统一大小,然后沿通道拼接成 256 * 256 * 352 的新特征图。最后使用转置卷积处理该特征图,将尺寸扩大到同输入图像一致。转置卷积输出的计算公式如下:

$$H_{out} = H_{in} * V_{stride} - 2 * V_{padding} + V_{kernelsize} \quad (1)$$

$$W_{out} = W_{in} * V_{stride} - 2 * V_{padding} + V_{kernelsize} \quad (2)$$

其中：V_{stride}、$V_{padding}$ 和 $V_{kernelsize}$ 分别表示 stride、padding 和 kernelsize；H_{out}、H_{in}、W_{out}、W_{in} 分别表示特征图的输出高度、输入高度、输出宽度、输入宽度。

1.4　DiceFocal 联合损失函数

损失函数的选择在模型训练过程中起着十分关键的作用，直接影响模型收敛的速度。首先，由于图像注释中存在背景像素，计算损失时背景像素会使损失值无效，并且遥感图像样本中普遍存在的类别分布不平衡的问题。该文选择 Dice 损失作为联合损失函数的一部分。该损失函数利用标注和预测结果的交集和并集进行计算，有效降低了背景像素的影响[25]。其次，为了让模型更好地分割容易被错分漏分的小尺度地物，该文选择将 Focal 损失作为联合损失函数的一部分。该损失函数基于交叉熵，通过一个动态因子来降低训练过程中易分割样本的权重。DiceFocal 联合损失函数包括 Dice 损失和 Focal 损失两个部分。

$$L_{DiceFocal} = L_{Dice} + L_{Focal} \quad (3)$$

Dice 损失计算公式如下：

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} l_i y_i}{\sum_{i=1}^{N} l_i + \sum_{i=1}^{N} y_i} \quad (4)$$

Focal 损失计算公式如下：

$$L_{Focal} = -\frac{1}{N}\sum_{i=1}^{N}[\alpha f_1(l_i, y_i) + (1-\alpha) f_2(l_i, y_i)] \quad (5)$$

其中，公式 5 中的 f_1、f_2 计算公式分别是：

$$f_1(l_i, y_i) = l_i (1 - y_i)^2 \log y_i \quad (6)$$

$$f_2(l_i, y_i) = (1 - l_i) y_i^2 \log(1 - y_i) \quad (7)$$

其中：λ 的值决定两个损失函数所占的比重，该文根据经验将其设定为 0.2；l_i 表示标注类别，y_i 表示预测类别；α 是 Focal 的动态因子，根据经验将其设定为 0.4。

2　实验设计

2.1　数据集介绍以及预处理

本研究所使用的数据集是由武汉大学制作的公开的高分辨率遥感数据集 GID[26]，该数据集覆盖面广、分布广、空间分辨率高，专门针对土地覆盖分类任务。数据集中的图像由 GF-2 卫星拍摄，涵盖中国 60 多个城市，包含三个可见光通道（红、绿、蓝）和一个近红外通道（NIR）。GID 包含 150 幅遥感图像，每幅图像的尺寸为 7 300 * 6 908 像素。还包含相对应的标注图像，标注了五个土地覆盖类型：

农田（Farmland）、建筑物（Building）、水（Water）、森林（Forest）和草地（Meadow）。图 4 展示了数据集中部分原始图像和标注图像。

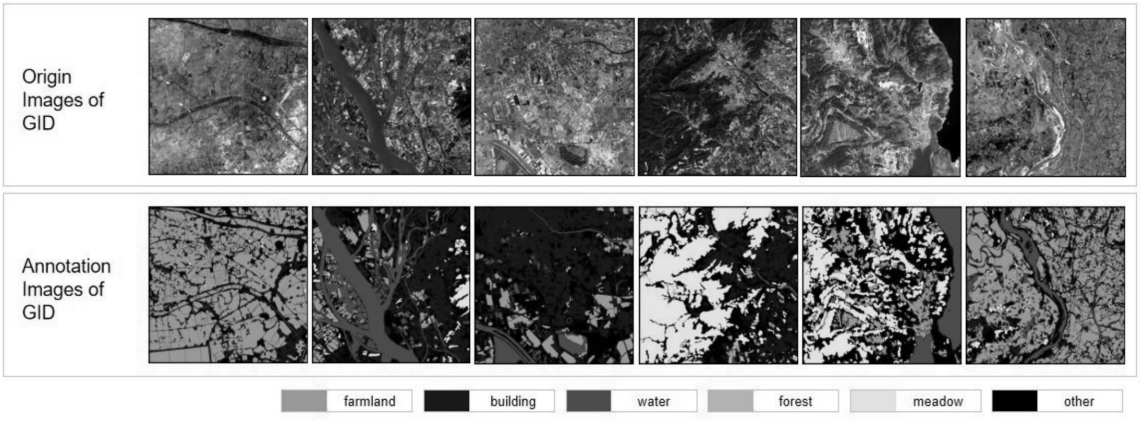

图 4　原始图像和标注图像示例

本研究删除了原始数据中存在的空白边缘，然后随机裁剪成 512 * 512 的大小，一共裁剪出15 000组原图像和标注图像。为了减少 GPU 性能的浪费，从 15 000组中删除背景像素超过全图 50% 的组，最终得到 11 269 组图像。其中，6 000 组作为训练集，3 000 组作为验证集，2 269 组作为测试集。由于随机裁剪图像中存在的重叠区域过多会影响模型的训练，本研究随机选择训练集中的数据进行图像增强，25% 的图像水平翻转，25% 的图像高斯模糊，25% 的图像顺时针旋转 90°。

在对数据集进行深入分析时发现，建筑物样本数量相对较少，而森林和草地样本数量相对较多，农田和水的样本数量处于中间水平。不同土地覆盖类型的样本分布呈现出明显的不均匀性，会导致模型在学习过程中对样本数量较多的类别产生偏向，而对样本数量较少的类别学习不够充分，从而影响模型对各类别土地覆盖的分类准确性和泛化能力。

2.2　模型训练设置

将初始学习率设置为 0.007，训练过程中采用了等间隔余弦退火算法来降低学习率，直到达到最小值

0.000 07。优化器使用带动量的 SGD,动量值设置为 0.75,减少训练时间,并且利用动量来防止模型收敛到局部最优值。

2.3 评价指标

为了定量地评价改进模型的分割精度,本研究使用 MIoU、F_1 分数和平均像素精度(MPA)这 3 个评价指标。这些指标在混淆矩阵的基础上进行计算,需要知道预测为真的正样本 TP、预测为假的正样本 FP、预测为真的负样本 TN 和预测为假的负样本 FN 的值。

神经网络语义分割领域中,MIoU 是一个非常重要的指标。它是将预测结果和标注结果的交集除以每个类别的并集,然后将这些比率相加并计算平均值来获得,计算公式如下:

$$S_{\mathrm{MIoU}} = \frac{1}{K}\sum_{i=1}^{k}\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}} + N_{\mathrm{FP}} + N_{\mathrm{FN}}} \tag{8}$$

其中,S_{MIoU} 表示 MIoU,N_{TP} 、N_{FP} 和 N_{FN} 分别表示 TP、FP 和 FN 的数量,k 表示地物类别的数量,文中 $k=5$。

F_1 分数是统计学中用于衡量二分类(或多任务二分类)模型精确度的指标,是模型的精确度和召回率的调和平均值,同时兼顾了分类模型的查准率和查全率,适用于样本不平衡的情况。本研究将像素分类的正确和错误视作二分类来计算 F_1 分数。F_1 分数首先需要计算查准率和查全率,完整的计算公式如下:

$$S_{\mathrm{Precision}} = \frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}} + N_{\mathrm{FP}}} \tag{9}$$

$$S_{\mathrm{Recall}} = \frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}} + N_{\mathrm{FN}}} \tag{10}$$

$$F_1 = \frac{2 * S_{\mathrm{Precision}} * S_{\mathrm{Recall}}}{S_{\mathrm{Precision}} + S_{\mathrm{Recall}}} \tag{11}$$

其中,$S_{\mathrm{Precision}}$ 和 S_{Recall} 表示查准率和查全率。

像素精度(PA)是另一种基于混淆矩阵的计算,它表示的是特定类别的正确分类像素占像素总数的比例。MPA 是所有类别的像素精度的平均值,其计算公式如下:

$$S_{\mathrm{MPA}} = \frac{1}{K}\sum_{i=1}^{k}\frac{N_{\mathrm{TP}} + N_{\mathrm{TN}}}{N_{\mathrm{TP}} + N_{\mathrm{TN}} + N_{\mathrm{FP}} + N_{\mathrm{FN}}} \tag{12}$$

其中,S_{MPA} 表示像素精度的平均值。

3 结果与分析

3.1 对比实验

本研究比较了八种语义分割模型在 GID 上的表现,包括 FCN、U - Net、SegNet、PSPNet、CBAM - DeepLab V3+[27]、CRF-DeepLab V3+[28],以及文中模型 V3plus-EN-TC。表 1 展示了每个模型的 MIoU、F_1 分数和 MPA,表 2 列出了五种土地覆盖类型各自的 IoU。

表 1 不同模型分割结果 %

网络名称	MIoU	F_1	MPA
FCN	62.66	69.72	68.19
U-Net	67.63	73.09	70.54
SegNet	69.93	76.96	75.73
PSPNet	72.01	78.52	73.49
DeepLab V3+	76.74	82.53	80.12
CBAM-DeepLab V3+	79.61	84.82	82.27
CRF-DeepLab V3+	77.12	82.97	80.43
V3plus-EN-TC(our)	84.74	88.39	86.64

表 2 各个类别的 IoU %

网络名称	Farm-Land	Building	Water	Forest	Meadow
FCN	60.18	71.20	77.44	50.51	53.97
U-Net	64.03	73.11	72.78	68.23	60.00
SegNet	64.78	75.26	82.04	66.65	60.92
PSPNet	71.26	84.31	82.95	65.80	55.73
DeepLab V3+	75.17	87.53	86.12	66.24	68.64
CBAM-DeepLab V3+	79.41	90.76	90.59	68.15	69.14
CRF-DeepLab V3+	76.68	89.88	85.26	67.05	66.73
V3plus-EN-TC(our)	86.45	94.36	92.72	71.89	78.28

FCN 是较早提出的一种网络结构相对简单的方法,其性能较差。相比之下,U-Net 和 SegNet 能够融合不同尺度特征,分割遥感图像可以获得更高的精确度。由于混合像元的复杂性,模型需要融合多尺度特

征。PSPNet 引入空间金字塔(SPP)模块,提取更丰富的特征,改善空间分布感知。DeepLab V3+在 SPP 的启发下设计了 ASPP,该模块使用空洞卷积扩大感受野,并且当有充足的实验数据时,其网络结构更有助于拟合训练数据。CBAM-DeepLab V3+在 DeepLab V3+的主干网络引入混合注意力机制来提高 Encoder 对特征的识别和学习,CRF-DeepLab V3+使用条件随机场(Conditional Random Field,CRF)辅助 Decoder 对特征的解释。二者的评价指标数据都优于原模型,证明了提高特征提取和解释的能力能够提升原模型在遥感图像上的表现。优化模型 V3plus-EN-TC 既采用注意力机制来增强 Encoder 的特征提取能力,又结合转置卷积获得了更强的 Decoder 特征解释能力,最终分割结果优于其他模型,MIoU、F_1 分数、MPA 分别达到了 84.74%、88.39%、86.64%,其中,对建筑物和水体的分割精度最高,两种地物类型的 IoU 分别是 94.36% 和 92.72%

3.2　消融实验

为了验证所做改进的有效性,本研究进行了消融实验。表 3 列出了原模型、两组消融模型、改进模型之间的对比。表 4 展示了每个类别的 IoU。其中,V3plus-EN 表示替换骨干网络为 EfficientNet,V3plus-TC 表示使用结合转置卷积的上采样方法,V3plus-EN-TC 是最终改进模型。

由于 SE 模块、空洞卷积和倒置残差连接的引入增强了 Encoder 的特征提取和融合能力,新主干网络的表现优于其他骨干网络。但是经过反复实验发现,由于网络层数过深,结构过于复杂,模型在训练过程中发生了过拟合,因此该文去掉了 block 中一部分 SE 模块,并且将 Drop 率设置为 0.5,最终得到消融模型 V3plus-EN。由表 3、表 4 可以看出,该消融模型在原模型基础上有较好的提升。

表 3　消融模型分割结果的 MIoU、F_1 和 MPA 比较　%

网络名称	MIoU	F_1	MPA
DeepLab V3+	76.74	82.53	80.12
V3plus-EN	81.68	85.38	84.01
V3plus-TC	78.59	84.22	81.19
V3plus-EN-TC(our)	84.74	88.39	86.64

表 4　消融模型的五种类别的 IoU 比较　%

网络名称	Farm-land	Building	Water	Forest	Meadow
DeepLab V3+	75.17	87.53	86.12	66.24	68.64
V3plus-EN	81.03	92.14	90.69	71.41	73.13
V3plus-TC	76.98	88.92	87.21	65.00	73.84
V3plus-EN-TC(our)	86.45	94.36	92.72	71.89	78.28

此外,融合更多层次的特征图后,结合转置卷积和双线性插值两种方法对融合后的特征图进行上采样,得到的消融模型 V3plus-TC 也比原模型有更好的精度评价。

改进模型 V3plus-EN-TC 的 MIoU、F_1 分数和 MPA,相较原模型分别增加了 8.00 百分点、5.86 百分点和 6.52 百分点,说明提高 Encoder 的特征提取能力和 Decoder 的特征解释能力能够提升语义分割模型在遥感图像上的分割精度。

3.3　损失曲线对比

该文对比了改进模型在三种损失函数下的损失曲线,分别是 Dice、Focal,以及该文使用的 DiceFocal 联合损失函数。损失随 epoch 变化的曲线如图 5 所示。其中:浅灰色曲线代表的是本文训练模型所使用的 DiceFocal 联合损失函数;黑色曲线表示使用 Dice 损失,由于它计算的是重叠区域,损失值变化很不稳定,200 个 epoch 之后值依然较大;深灰色曲线表示使用 Focal 损失,收敛速度快于前者,损失值也更小。浅灰色曲线表示使用 DiceFocal 联合损失函数,它比单独使用 Dice 或者 Focal 的效果都要更稳定,更早收敛,更趋近于理想训练结果。

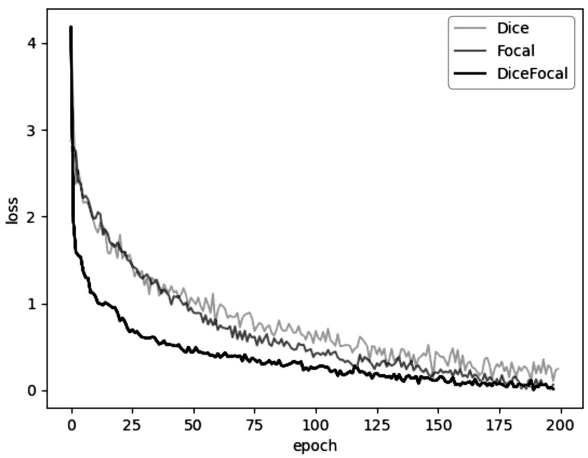

图 5　不同损失函数下的损失曲线

3.4 分割结果对比

图 6 展示的是其他 DeepLab V3+优化模型与文中模型的消融版本的分割结果,图中(a)列是 CBAM-DeepLab V3+的结果,(b)列是 CRF-DeepLab V3+的结果,(c)列是原模型的结果,(d)列是 V3plus-EN 的结果,(e)列是 V3plus-TC 的结果,(f)列是改进模型 V3plus-EN-TC 的结果。可以明显看出,原模型应用于遥感图像时存在较多空洞区域。(a)列和(d)列模型是在骨干网络中加入了能够结合通道信息和空间信息的注意力机制,相较原模型准确率有很大提升。(b)列、(c)列与(e)列分别是用条件随机场、双线性插值和转置卷积来解释融合后的特征,可以看出转置卷积更能还原小尺度的边缘细节。在文中模型的分割结果中,空洞现象显著减少,有效抑制大尺度地物中间空洞区域的出现,还展现出更优秀的边缘细节,尤其是在识别建筑物和水体方面。

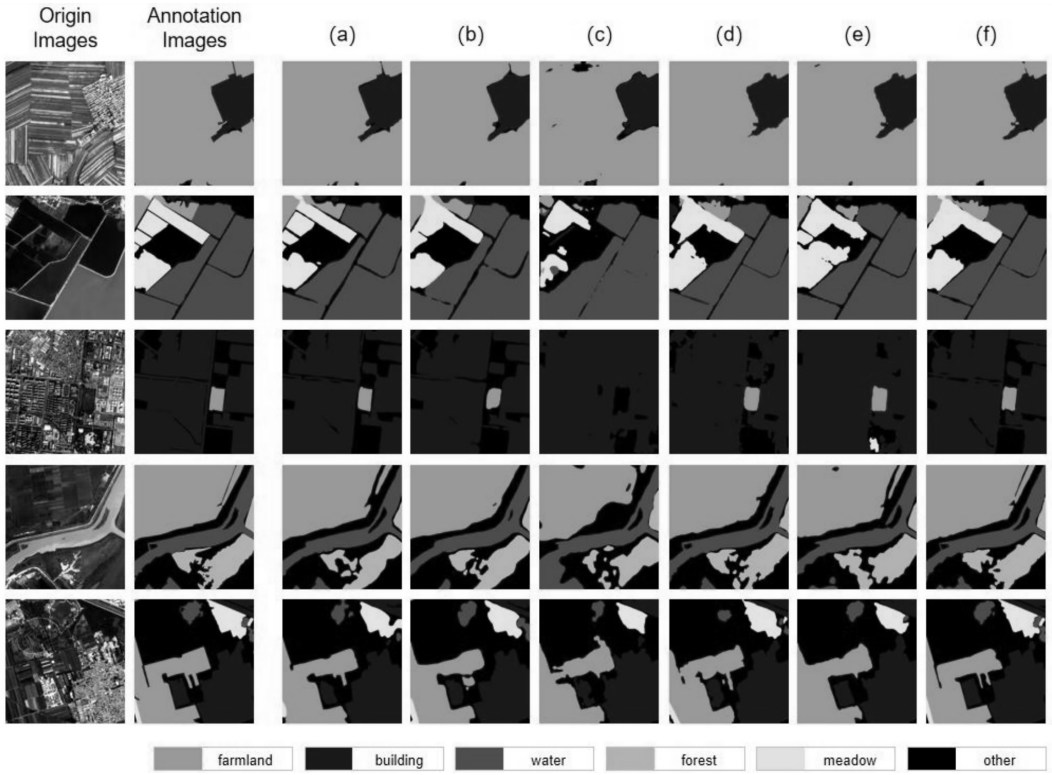

图 6 分割结果对比

4 结束语

基于 DeepLab V3+模型,该文提出了一个针对遥感图像改进的语义分割模型,用 EfficientNet 替换原本的主干网络,引入 SE 模块、空洞卷积、倒置残差连接,融合更多层次的特征,使用双线性插值和转置卷积结合的上采样方法,并且使用 DiceFocal 联合损失函数进行训练。使模型具有更强的特征提取和解释能力,能够有效处理复杂的空间信息和通道信息,提高遥感图像语义分割的精度,减少大尺度地物内部空洞区域的出现。在 GID 遥感数据集上的实验结果证明了改进模型的有效性,MIoU、F_1 分数和 MPA 分别达到 84.74%、88.39% 和 86.64%,与原模型相比分别提高了 8.00 百分点、5.86 百分点和 6.52 百分点。

然而,提出的改进模型仍然存在进一步优化的空间。首先,模型用于其他领域数据集的泛化性能仍需评估。此外,该模型需要大量的训练时间。未来的研究可能会更加关注提高模型的训练速度。

参考文献:

[1] 汤泊川,帕力旦·吐尔逊,柏洁馨,等. 结合 CNN 和 Transformer 的遥感图像土地覆盖分类方法[J]. 微电子学与计算机,2024,41(4):64-73.

[2] NALLA S,TOTAKURA M,PIDIKITI D,et al. Monitoring urban growth using land use land cover classification[J]. Lecture Notes in Networks and Systems,2023,615:275-283.

[3] 林云浩,王艳军,李少春,等. 一种耦合 DeepLab 与 Transformer 的农作物种植类型遥感精细分类方法[J]. 测绘学报,2024,53(2):353-366.

[4] MORADKHANI K,FATHI A. Segmentation of waterbodies in remote sensing images using deep stacked ensemble model[J]. Applied Soft Computing,2022,124:109038.

[5] LI L,ZHU Z,WANG C. Multiscale entropy-based surface complexity analysis for land cover image semantic segmentation[J]. Remote Sensing,2023,15(8):2192.

[6] CHENG X,SUN Y,ZHANG W,et al. Application of deep learning in multitemporal remote sensing image classification

[J]. Remote Sensing,2023,15(15):3859.

[7] CHEN B,ZOU X,ZHANG Y,et al. Leformer:a hybrid CNN-transformer architecture for accurate lake extraction from remote sensing imagery[C]//IEEE international conference on acoustics speech and signal processing. [s. l.]:IEEE,2024:5710-5714.

[8] LIU B,WU H,BAO X,et al. LPCUNet:a lightweight pure CNN UNet for efficient urban scene remote sensing semantic segmentation[C]//2023 4th international conference on computer vision image and deep learning. [s. l.]:IEEE,2023:57-61.

[9] CHEN L C,ZHU Y,PAPANDREOU G,et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[J]. Lecture Notes in Computer Science,2018,11211:833-851.

[10] ZHANG Y,ZHANG Y,ZHANG Q. Semantic segmentation of traffic scene based on DeepLabv3+ and attention mechanism[C]//2023 3rd international conference on neural networks information and communication engineering. [s. l.]:IEEE,2023:542-547.

[11] SUN J,ZHOU J,HE Y,et al. RL-DeepLabv3+:a lightweight rice lodging semantic segmentation model for unmanned rice harvester[J]. Computers and Electronics in Agriculture,2023,209:107823.

[12] CHEN Hui,QIN Yuanshou,LIU Xinyuan,et al. An improved DeepLabv3+ lightweight network for remote-sensing image semantic segmentation[J]. Complex & Intelligent Systems,2023,10(2):2839-2849.

[13] QUAN B,LIU B,FU D,et al. Improved deeplabv3 for better road segmentation in remote sensing images[C]//Proceedings - 2021 international conference on computer engineering and artificial intelligence. [s. l.]:[s. n.],2021:331-334.

[14] ZHENG K,WANG H,QIN F,et al. An improved land use classification method based on DeepLab V3+ under GauGAN data enhancement[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2023,16:5526-5537.

[15] GUO Q,XU Y. Image semantic segmentation model based on CBAMUNet[C]//Proceedings of SPIE - the international society for optical engineering. [s. l.]:[s. n.],2024:68-77.

[16] LU K,MA Z,HUO P,et al. Mixed pixel saturability based area estimation model on remote sensing image[C]//2023 IEEE 6th international conference on pattern recognition and artificial intelligence. Haikou:IEEE,2023:751-757.

[17] HU L,ZHOU X,RUAN J,et al. ASPP+-LANet:a multi-scale context extraction network for semantic segmentation of high-resolution remote sensing images[J]. Remote Sensing,2024,16(6):1-13.

[18] 赵玉刚,刘文萍,周　焱,等. 基于注意力机制和改进 DeepLabV3+的无人机林区图像地物分割方法[J]. 南京林业大学学报:自然科学版,2024,48(4):93-103.

[19] 马　静,郭中华,马志强,等. 基于轻量化的 DeepLabV3+遥感图像地物分割方法[J]. 液晶与显示,2024,39(8):1001-1013.

[20] TAN M,LE Q V. EfficientNet:rethinking model scaling for convolutional neural networks[C]//36th international conference on machine learning. Long Beach:ICML,2019:10691-10700.

[21] 肖振久,杨玥莹,孔祥旭. 基于改进 YOLOv4 的遥感图像目标检测方法[J]. 激光与光电子学进展,2023,60(6):407-415.

[22] 刘　浪,刘国栋,刘　佳. 基于改进 EfficientDet 算法的可见光遥感舰船目标检测[J]. 现代电子技术,2022,45(22):28-32.

[23] YIN H,YANG C,LU J. Research on remote sensing image classification algorithm based on EfficientNet[C]//2022 7th international conference on intelligent computing and signal processing. Xi′an:ICSP,2022:1757-1761.

[24] ISLAM R,HOSSEN S,ARIFUL ISLAM S M,et al. BI-TLM:bilinear interpolation with transfer learning model for breast cancer classification[C]//2023 6th international conference on electrical information and communication technology. Khulna:[s. n.],2023:1-5.

[25] MING Q,XIAO X. Towards accurate medical image segmentation with gradient-optimized dice loss[J]. IEEE Signal Processing Letters,2024,31:191-195.

[26] TONG X Y,XIA G S,LU Q,et al. Land-cover classification with high-resolution remote sensing images using transferable deep models[J]. Remote Sensing of Environment,2020,1:237.

[27] CHANG H,GUO S,ZHANG H,et al. Apple planting area extraction based on improved DeepLab V3 +[J]. Nongye Jixie Xuebao/Transactions of the Chinese Society for Agricultural Machinery,2023,54:206-213.

[28] WANG Z,FAN B,TU Z,et al. Cloud and snow identification based on DeepLab V3+ and CRF combined model for GF-1 WFV images[J]. Remote Sensing,2022,14(19):4880.