

多层次特征优化融合的遥感图像分割网络

菅永胜, 朱大明*, 付志涛, 文诗雅

昆明理工大学国土资源工程学院, 云南 昆明 650093

摘要 为了从高分辨率遥感图像中准确地分割出地物目标, 提出了一种多级特征优化融合的遥感图像分割网络 (MRFNet), 着重将特征提取骨架网络中不同层级的特征图进行融合, 通过融合网络特征图中不同种类的信息来对输出特征图信息进行合理有效的提取和分析。同时使用了逐层的多尺度编码解码模块来细化与高层特征图进行融合的浅层特征图, 将不同种类的信息经过优化以后汇聚到高层特征图。然后采用空洞卷积金字塔对高层特征图提取不同感受野的信息, 优化了语义分割的输出特征图。通过在 ISPRS Vaihingen 数据集上进行实验, 该网络算法的总体精度达到了 90.34%, 与经典语义分割网络相比, 有效提升了遥感图像目标的检测精度。同时为了证明算法的泛化性, 在 ISPRS Potsdam 数据集上进行了泛化实验取得了 91.47% 的总体精度, 证明了该算法的有效性。

关键词 遥感; 语义分割; 多尺度编码解码; 特征融合

中图分类号 TP753

文献标志码 A

DOI: 10.3788/LOP212864

Remote Sensing Image Segmentation Network Based on Multi-Level Feature Refinement and Fusion

Jian Yongsheng, Zhu Daming*, Fu Zhitao, Wen Shiya

Faculty of Land Resource Engineering, Kunming University of Science and Technology,
Kunming 650093, Yunnan, China

Abstract To accurately segment ground objects from a high-resolution remote sensing image, we propose a remote sensing image segmentation network based on multi-level feature optimization fusion that focuses on the fusion of feature maps at different levels in the feature extraction skeleton network, performs reasonable and effective extractions, and analyzes output feature map information by fusing different types of information in the network feature map. Simultaneously, layer-by-layer multi-scale coding and decoding modules are used to refine the shallow feature map that merges with the high-level feature map, and the different types of information are optimized to the high-level feature map. The hollow convolution pyramid is then used to extract the information of different receptive fields on the high-level feature map, and the output feature map of semantic segmentation is optimized. When conducting experiments on the ISPRS Vaihingen dataset, the overall accuracy of the proposed network reaches 90.34%, which effectively improves the accuracy of remote sensing image target detection when compared with the classical semantic segmentation network. Moreover, to prove the generalization of the proposed algorithm, a generalization experiment on the ISPRS Potsdam dataset is conducted; the overall accuracy of this algorithm reaches 91.47%, proving its effectiveness.

Key words remote sensing; semantic segmentation; multi-scale encoding and decoding; feature fusion

1 引言

语义分割是计算机视觉领域的一个经典问题, 同时也在遥感图像处理中有很大的作用, 对地物分析起到了很大的帮助^[1]。传统的地物图像分类需要耗费大量的人力物力去实地考察得到结果。但是基于深度学

习图像识别的算法^[2], 可以标记部分样本进行模型学习, 然后使用模型对大量的样本进行识别。完全卷积网络 (FCN)^[3] 提出了编码解码的结构进行图像分割, 采用了全卷积的方式, 丢弃了图像分类的全连接层, 同时也提出了多级特征融合的想法, 为后面的工作开创了方向。U-Net^[4] 目前在遥感影像分割领域当中占据

收稿日期: 2021-11-03; 修回日期: 2021-12-08; 录用日期: 2021-12-21; 网络首发日期: 2021-12-30

基金项目: 国家自然科学基金(41961053, 41961039)

通信作者: *634617255@qq.com

统治地位,由于U-Net强化了多级的特征融合,结合了不同层级的编码解码阶段特征图进行特征融合优化特征图,取得了很好的效果。Deeplab系列算法^[5-8]提出了空洞卷积空间金字塔结构,采用了不同空洞率的空间卷积进行不同感受野的特征信息聚拢,以此优化特征图。

上述方法均为自然场景的深度学习分割算法,对于遥感影像而言,Mnih等^[9]使用深度学习来进行遥感影像图像分割的处理,并取得了很好的效果。对于遥感影像分割需要对地物目标的边界进行精细化的分割,但是边界的拟合问题是深度学习图像分割中一个难以解决的问题^[10]。Singh等^[11]较好地解决了小目标物体的分割问题即直接放大图像尺寸,采用了图像金字塔作为输入,但是遥感图像尺寸通常是上千万个像素,因此这种方法会耗费大量的计算资源。PerceptualGAN^[12]使用生成对抗网络对小目标生成一个超分辨特征叠加在原来的小目标的特征图上,以此增强小目标特征表达。但是这种方法会增加网络的复杂程度。Lin等^[13]提出了一种尺度感知网络,通过对像元进行重采样,使其适应不同尺度的目标,并引入了空间注意力,该方法适合具有尺度变化的目标分割。在当前发展阶段中,很多有关遥感图像语义分割算法是基于U-Net解码部分的改进。同时,也有Deeplab、CCNet^[14]、DANet^[15]等采用增强图像上下文信息或者建立像素相互关系来提升图像分割的效果。近年来注意力机制的提出,很多研究者采用注意力机制用于遥感影像分割算法。RANet^[16]采用了区域注意力机制,通过在不同区域的像素之间配置信息通路,使区域交互能够交换区域上下文以增强图像中的所有像素。SCAttNet^[17]通过将通道注意力和空间注意力采用并行构建的方式,获取了特征图关于通道上和空间上的像素关系,优化了特征图处理。SSAtNet^[18]提出了金字塔注意力池化的模块对特征图做自适应优化。

为了提高尺度差异性巨大的遥感图像场景目标的识别效果,针对以上出现的问题,本文提出了一种多层次特征优化融合的遥感图像分割网络,将特征提取骨干网络中不同层级的特征图进行融合,通过融合网络特征图中不同种类的信息对输出特征图信息进行合理有效的提取和分析;本文使用了逐层的多尺度编码解码模块来细化与高层特征图进行融合的浅层特征图,将不同种类的信息经过优化以后汇聚到高层特征图;采用金字塔空洞卷积对高层特征图提取不同感受野的信息,优化了语义分割的输出特征图。

2 多层次特征优化融合网络(MFRNet)

2.1 遥感图像分割算法流程

对于深度学习算法而言,通常会提前设计算法流程,然后根据具体的算法流程进行数据处理和训练。针对于本文的遥感深度学习图像分割算法,设计了特

定的算法流程来验证实验,具体步骤如下:

1) 获取数据确定数据信息。本实验选用了遥感图像分割领域的经典数据集ISPRS Vaihingen和Potsdam数据作为实验数据。Vaihingen数据集包括近红外、红和绿3个波段。Potsdam数据集包括近红外、红、绿和蓝4个波段。通过这样的设定,一定程度也证明了本文提出的算法在不同波段的遥感图像上均能取得很好的分割效果。

2) 确定原始数据集图像尺寸,并对数据进行切片生成训练集、验证集和测试集。原始的Vaihingen图像尺寸约为2000 pixel×3000 pixel,Potsdam图像尺寸为6000 pixel×6000 pixel。通常无法直接将这种大尺寸的数据直接输入到网络进行训练,因此实验中也是选用了固定的尺寸对原始影像进行滑窗式的切割,切割尺寸为512 pixel×512 pixel,滑窗步长为256,其中超出原始图像边缘的部分,选用了边缘镜像填充原始图像来补偿超出的区域。其中,测试集按照ISPRS官方的设定去选取,验证集则从训练集中随机进行挑选。

3) 算法框架设计。针对于尺度差异性巨大的遥感图像分割。本文基于经典的语义分割框架,设计了一种新的多尺度编码解码特征优化模块,通过将这个模块嵌入到特征提取部分的不同阶段,以此优化了遥感图像分割的特征提取部分,改进了分割效果。

4) 算法测试部分设计。选用普遍科学的训练和测试方式,在训练集上进行训练,同时在验证集上计算验证集精度来评价训练过程。通过训练一定的次数之后,观察到可视化的验证集精度曲线趋于平缓,则选用验证集精度最高的模型进行测试,并将计算测试的相关精度评价指标作为本文的实验结果。

2.2 多层次特征优化融合的遥感图像分割算法网络结构

多层次相邻特征融合网络主要是采用了融合多个层级的特征图,从而实现语义分割网络中编码部分的信息融合。对于遥感影像而言,通常语义类别比较单一,并没有非常丰富的场景语义信息,因此选取了ResNet50^[19]作为骨干网络。选取512 pixel×512 pixel×3 pixel的图像大小作为网络的输入尺寸,图1中F1、F2、F3、F4是ResNet50的网络骨架部分的多级特征图。ResNet50网络结构中存在着逐级池化下采样的模块,池化计算过程会降低特征图的维度,同时逐级加深特征图,导致ResNet50的不同层级的特征图具有不同的信息比重。浅层的特征图比如F1、F2会具有较为清晰的物体轮廓边缘信息,此部分内容对于解码部分恢复物体的边缘轮廓有所帮助。而F4、F3为深层的特征图,此部分特征图信息由于多级的池化作用,像素的感受野比较大,同时具备充分的语义信息,语义信息有助于解码部分对物体的类别进行区分,使得网络结构能对目标区域进行正确的分类。如果能实现浅层特征图信息和深层特征图信息相互配合,将

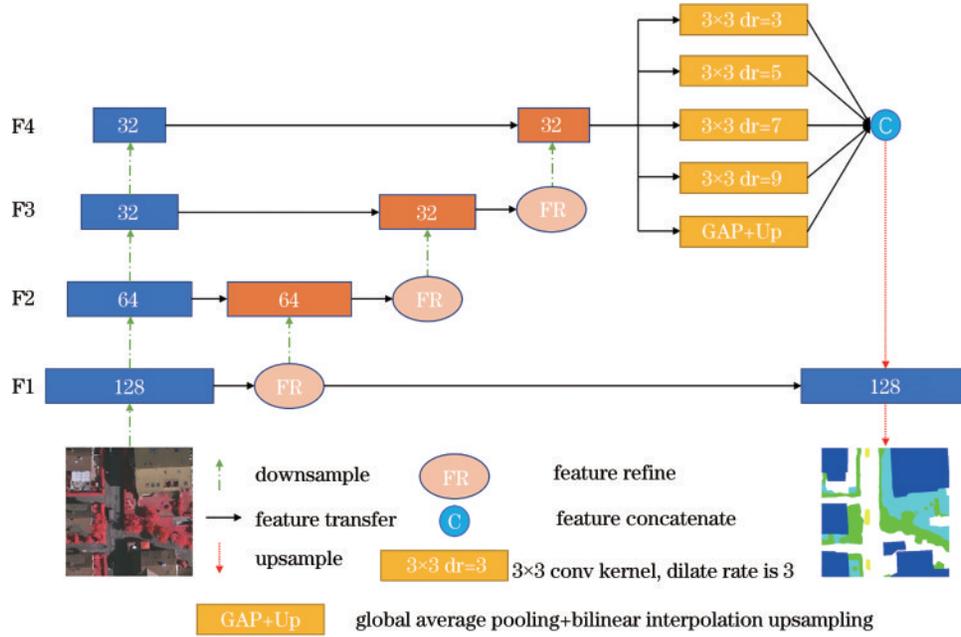


图 1 多层次特征优化融合网络

Fig. 1 Multi-level feature refinement and fusion network

有助于分割掩码从类别和边缘轮廓上都能比较好地贴合真实标签。因此,以 ResNet50 骨架网络的 4 个模块为基础设计了多层次特征优化融合的编码部分网络。采用了使用相邻层级的模块进行特征融合,以获取相邻的不同层级的融合特征。同时,为了加强浅层纹理特征的代表能力,设计了对浅层的纹理特征进行编码解码计算的再优化。

如图 1 所示,本文以 ResNet50 的 4 个阶段的特征图作为初始框架来进行扩展,分别为 F_1 、 F_2 、 F_3 、 F_4 ,其中,对于输入尺寸为 $512 \text{ pixel} \times 512 \text{ pixel}$ 的图像尺寸,得到的特征图尺寸分别为 4 个级别的特征图尺寸分别为 $128 \text{ pixel} \times 128 \text{ pixel}$ 、 $64 \text{ pixel} \times 64 \text{ pixel}$ 、 $32 \text{ pixel} \times 32 \text{ pixel}$ 、 $32 \text{ pixel} \times 32 \text{ pixel}$ 。针对遥感图像的图像特征,通过调整初始骨架网络的下采样的倍数,将最后一级特征图尺寸调整为 $32 \text{ pixel} \times 32 \text{ pixel}$ 。由此可得到 4 级具有不同信息侧重的特征图模块。

2.3 多尺度编码解码特征优化模块

通常来说,不同层级的特征图所具有的信息有不同的侧重,本次实验中以高层特征图为基础,通过多级的浅层特征图信息进行多级下采样以及特征融合,然后与高层级特征图进行信息聚合。为了细化不同层级的特征图,采用了不同尺度的下采样和上采样结构进行特征优化。如图 2 所示,对于输入特征图,选择了不同尺度的编码解码。分别对特征图进行 8 倍、4 倍、2 倍的均值池化计算,然后采用卷积计算细化特征图,同时添加 ReLU 非线性激活函数,最后采用反卷积上采样计算将缩小的特征图进行 8 倍、4 倍、2 倍放大尺寸,对特征图进行了不同级别的编码解码计算。下采样池化计算可以剔除特征图中的噪声信号,一定程度对特征

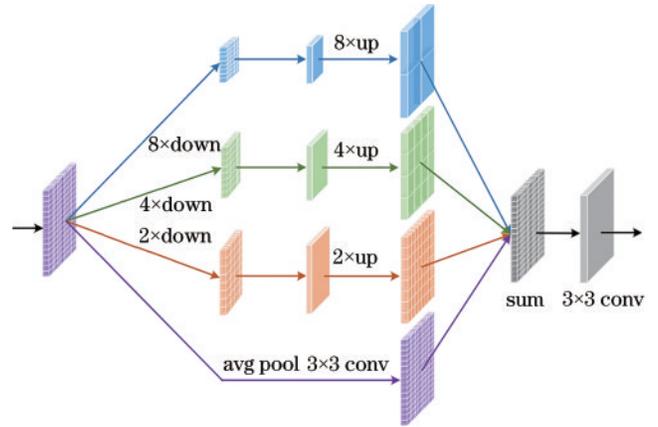


图 2 多尺度编码解码特征优化模块

Fig. 2 Multi-scale encoding-decoding feature refinement module

图信号进行了特征选择,增强了遥感影像的显著性的目标区域的特征。反卷积上采样计算可以对下采样的特征图进行一定程度的信息还原,主要是保留了显著性的目标区域的特征信息。池化计算一定程度丢弃了特征图的边界信息,为了补充特征信息,采用一个 1×1 卷积以及 ReLU 激活函数对特征图进行非线性变化得到简单的特征图映射。然后将 4 个进行不同种类编码解码计算的特征图进行相加特征融合,得到了优化的特征图。通过上述的计算方式,对特征图进行了特征信息筛选和融合,一方面加强了遥感影像目标的显著性信息,另一方面也通过池化计算弱化了特征图噪声信息。

如图 1 所示,通过进行逐层的多尺度编码解码特征优化模块,然后与相邻层级特征图进行融合,融合计算选用下采样和 concat 计算。通过不同层级的特征融合,将输出特征图进行了多级的信息优化,最终特征图

信息汇聚于顶层特征图。然后采用了空洞空间卷积池化金字塔(ASPP)进行顶层的特征图优化,通过对特征图进行不同尺度的空洞卷积计算,获取了不同范围的感受野特征,同时使用全局均值池化获取了特征图的全局信息。然后通过将不同范围的特征图进行concat融合,对顶层特征图进行优化。为了对网络的预测结果进行优化,对顶层的特征图补充了底层的特征图信息,底层特征图信息含有较多的边缘纹理信息,对分割结果能进行精细化处理,使得分割边缘更加贴近真实结果。

3 实验及结果分析

3.1 实验环境及设备

本次实验选用了Linux服务器系统,选用了NVIDIA Tesla K80显卡为计算资源,CPU为3×Xeon E5-2678 v3,内存为16 G,显存为12 G。深度学习算法开发框架为Pytorch1.7.0。

表1 Vaihingen训练样本像素类别分布情况

Table 1 Vaihingen training sample pixel category distribution

Class	Impervious surface	Building	Low vegetation	Tree	Car	Background
Pixel percentage / %	29.03	26.69	20.06	22.08	1.21	0.93

由于数据集的分类精度需要把类别边界的影响考虑在内,因此在官方的评价指标中,由于类别边界受到侵蚀,侵蚀圆盘半径达到3 pixel,这部分应该去除掉,不纳入到计算中。实验中,按照数据集官方的设置,将33张数据分为16张训练集和17张测试集。在实验过程中,本文将16张训练集分出2张图像作为验证集。在遥感图像分割中,通常是将高分辨率的遥感图像按照一定的规则做切割操作,本文以256 pixel为重叠区域将数据集切割成512 pixel×512 pixel大小的图片作为训练集。

3.3 数据处理过程

对于遥感影像来说,图像尺寸通常较大,受限于显卡的计算能力,一般将遥感影像数据进行切片处理。本次实验中,对原始影像进行切片处理,对于超过边缘的部分,选择了镜像填充,通过这个方法来扩充数据。最终得到877张训练集图片和169张验证集图片。同时,为了继续增强训练数据的多样性,在训练过程中,对训练集样本进行了几何变换的数据增强。在神经网络训练中,增广并不意味着数量越多越好,数据增广可能产生很大的随机特征。因此,采用以下基础的几何变换方式来进行数据增广:

1) 翻转变换。把图像按照水平或者垂直方向进行翻转处理;

2) 随机旋转变换。随机旋转图像的某个角度;

3) 大小缩放变换。按照一定比例来调整图像;

4) 上下左右平移变换。从水平或者垂直方向来对图像进行平移处理;

3.2 数据集

选用Vaihingen和Potsdam数据集^[20]为研究对象,将MFRNet与UNet、RefineNet^[21]、PSPNet^[22]和Deeplab v3+等经典的图像语义分割网络进行对比。Vaihingen数据集一共包含有33张高分辨率遥感影像,主要是处于城市区域。数据特征是3个波段数据构成,分别是近红外波段(IR)、红色波段(R)和绿色波段(G)。Vaihingen数据集每张图像的尺寸大小不一样,由于是同一个机器拍摄的影像,因此每张影像都有相同空间分辨率,可以获得相应的数字地表模型。其中,数据集中涉及到16个真实的标签图形,划分为6大类别。为了证明本文提出算法的泛化性,基于Potsdam数据集进行了实验,Potsdam数据集与Vaihingen数据集包含的地物类别相同。与Vaihingen数据集不同的是,Potsdam数据集采用了近红外波段(IR)、红色波段(R)、绿色波段(G)和蓝色波段(B)。Vaihingen数据集中,训练样本构成详情如表1所示。

5) 随机裁切。随机选择图像上的一部分进行裁切,得到局部图像;

6) 对比度变换。依据图像调整对比度。

3.4 分割结果精度比较以及可视化分析

为了证明本文提出的多层特征优化融合的遥感图像分割网络的有效性,基于ISPRS Vaihingen数据集和Potsdam数据集进行了大量的实验。首先,基于ISPRS提供的Vaihingen数据集进行了实验,通过对比本文提出的算法与一些经典的语义分割算法,实验结果证明,本文提出的算法能对高分辨率遥感影像中的地物进行有效的分割,同时分割结果能对轮廓部分进行有效的拟合。同时,影像分割可视化的结果,也能显示出本文算法对不同的遥感影像地物类别均能有较好的分割结果。实验选用像素精度(A_{pixel})和 F_1 分数作为图像分割效果的评价指标。计算方法为:假设数据一共有 $k+1$ 个类别,包括 k 个目标类别和1个背景类别,同时假设像素标签为 i 类别,预测结果为 j 类别,那么 p_{ii} 为真阳性, p_{ij} 为假阳性, p_{ji} 为假阴性。通过上述描述,可以定义精度计算为

$$A_{\text{pixel}} = \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}}, \quad (1)$$

$$P = \frac{S_{\text{TP}}}{S_{\text{TP}} + S_{\text{FP}}}, \quad (2)$$

$$R = \frac{S_{\text{TP}}}{S_{\text{TP}} + S_{\text{FN}}}, \quad (3)$$

$$F_1 = 2 \times \frac{P \times R}{P + R}, \quad (4)$$

式中： S_{TP} (true positive) 为真正例； S_{FP} (false positive) 为假正例； S_{FN} (false negative) 为假反例； P 为精确率，是指预测为真正比数与预测为正例总量

的比率； R 为召回率，表示预测正确的正例数与真正的正例数的比率； F_1 用来衡量精确率和召回率，是两者的调和均值。由上述公式计算，在 Vaihingen 数据集上实验结果的精度计算，如表 2 所示（表中 IoU 为交并比）。

表 2 Vaihingen 数据集目标分类精度
Table 2 Object classification accuracy on Vaihingen dataset

unit: %

Method	IoU					Mean IoU	Mean F_1	Overall accuracy
	Impervious surface	Building	Low vegetation	Tree	Car			
PSPNet	77.26	87.89	65.67	77.48	55.86	72.83	85.08	87.83
UNet	77.45	84.42	66.03	76.24	58.65	73.29	85.57	87.97
RefineNet	82.14	87.12	66.69	74.57	56.87	73.48	85.82	88.16
Deeplab v3+	83.77	89.92	67.93	78.67	62.46	76.55	87.90	89.65
MRFNet (proposed)	84.37	90.15	68.04	79.21	63.49	77.05	88.24	90.34

对于遥感影像地物分类算法而言，通常需要对分割算法的结果进行可视化分析，通过可视化分析也能更进一步地对算法进行优化和效果提升。如图 3 所示，选取部分遥感影像切片的预测结果进行可视化展示，以此对比不同算法的不同结果。从图 3 中标记的

方框内可以看出本文提出的多层级特征优化融合算法的结果更加贴合真实的标签，在建筑物、树木和汽车等类别中，均能体现出更加优越的分割结果。MRFNet 通过使用逐层的多尺度编码解码，细化了浅层的特征图，使得在小目标比如汽车类别上更具优势。同时，通

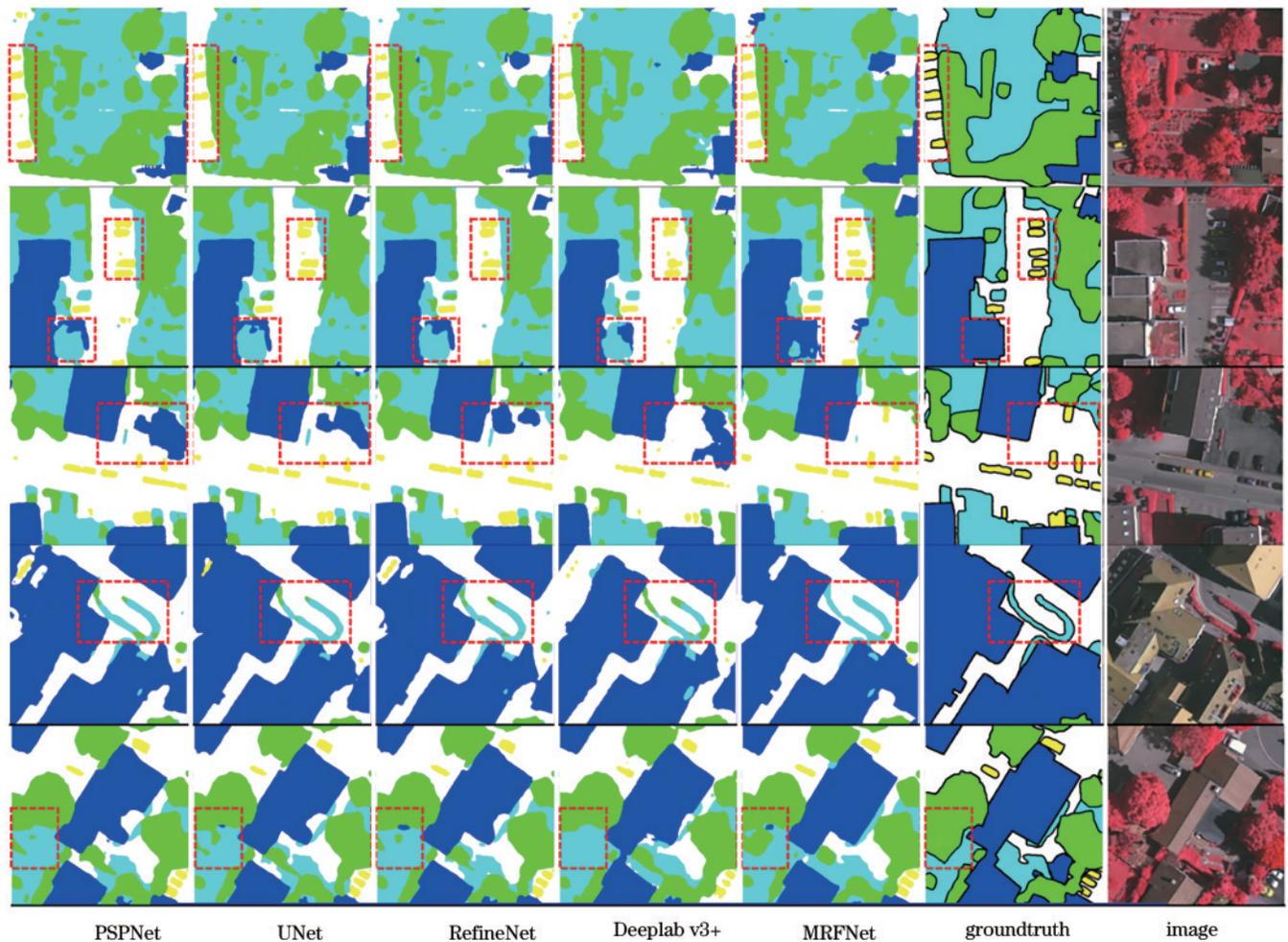


图 3 MRFNet 与经典的语义分割算法在 Vaihingen 数据集上的对比
Fig. 3 Comparison of MRFNet and classic semantic segmentation algorithms on Vaihingen dataset

过将不同层级的特征图信息汇聚到高层特征图上,丰富了高层特征图所含有的语义信息,使得在目标分类上具有更好的识别能力。

同时,为了证明本文提出算法在普遍的遥感影像数据集上均能取得较好的分割结果,选用 MRFNet 模型基于 Potsdam 数据集也进行了实验,实验结果如表 3 所示。同时对实验的分割效果进行了可视化处理,如

图 4 所示为 Potsdam 分割结果的全局效果图。结果表明,本文提出的算法在全局可视化效果上,没有体现出拼接的痕迹,表明本文算法在大尺寸的高分辨率遥感影像上均能体现出非常好的分割效果,可满足实际应用。在 Potsdam 数据集上的实验证明了本文算法具有较好的泛化性,在不同的高分辨率遥感数据集上,均有很好的分割效果。

表 3 Potsdam 数据集目标分类精度
Table 3 Object classification accuracy on Potsdam dataset

unit: %

Method	IoU					Mean IoU	Mean F_1	Overall accuracy
	Impervious surface	Building	Low vegetable	Tree	Car			
MRFNet (proposed)	84.37	90.15	68.04	79.21	63.49	77.05	88.24	91.47

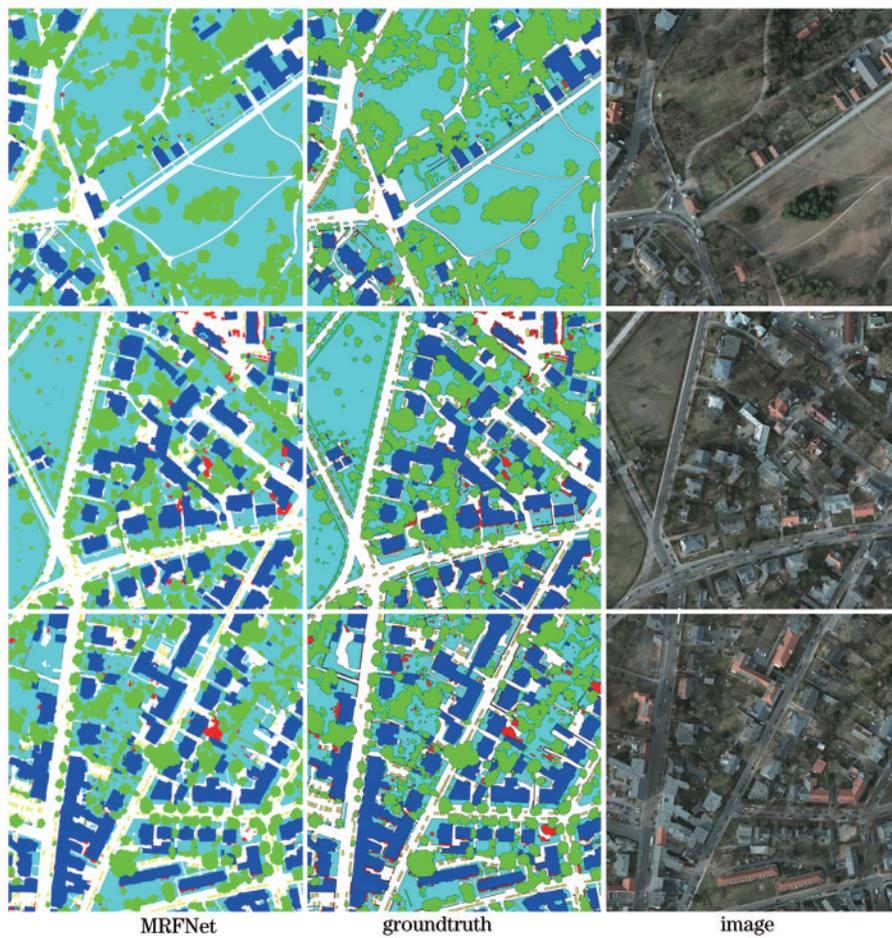


图 4 MRFNet 在 Potsdam 数据集上的分割结果

Fig. 4 Segmentation results of MRFNet on Potsdam dataset

4 结 论

基于 ResNet50 不同层级的特征图所具有不同含义的特征信息这一特点,设计了 MRFNet。对骨架网络上提取出来的不同层级的特征图进行了特征融合。同时,为了更好地细化不同阶段的特征图,对特征图进行了不同尺度的编码解码计算。选用了多个尺度的编码解码,然后进行不同尺度的融合,细化了特征图。通

过观察实验结果可看出,提出的方法在分割效果和精度上均已超过了现存的经典语义分割算法。同时,为了证明提出算法在不同的数据集上都能取得很好的分割效果,在 ISPRS 的 Vaihingen 数据集和 Potsdam 数据集上均进行了实验,实验结果表明,提出算法在不同的高分辨率遥感图像数据集上,均能取得很好的分割结果和精度指标。

参 考 文 献

- [1] Dong S J, Chen Z C. A multi-level feature fusion network for remote sensing image segmentation[J]. *Sensors*, 2021, 21(4): 1267.
- [2] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [3] 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.
- [4] 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.
- [5] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[EB/OL]. (2014-12-22) [2021-06-08]. <https://arxiv.org/abs/1412.7062>.
- [6] 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*, 2018, 40(4): 834-848.
- [7] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-06-17) [2021-05-06]. <https://arxiv.org/abs/1706.05587>.
- [8] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[EB/OL]. (2018-02-07) [2021-04-05]. <https://arxiv.org/abs/1802.02611>.
- [9] Mnih V, Hinton G. Learning to label aerial images from noisy data[C]//ICML'12: Proceedings of the 29th International Conference on International Conference on Machine Learning, Edinburgh, Scotland, UK. New York: ACM Press, 2012: 203-210.
- [10] Mi L, Chen Z Z. Superpixel-enhanced deep neural forest for remote sensing image semantic segmentation[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 159: 140-152.
- [11] Singh B, Davis L S. An analysis of scale invariance in object detection-SNIP[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 3578-3587.
- [12] Li J N, Liang X D, Wei Y C, et al. Perceptual generative adversarial networks for small object detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1222-1230.
- [13] Lin J B, Jing W P, Song H B. SAN: scale-aware network for semantic segmentation of high-resolution aerial images[EB/OL]. (2019-07-06) [2021-08-05]. <https://arxiv.org/abs/1907.03089>.
- [14] Huang Z L, Wang X G, Huang L C, et al. CCNet: criss-cross attention for semantic segmentation[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 603-612.
- [15] 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.
- [16] Shen D G, Ji Y F, Li P, et al. Ranet: region attention network for semantic segmentation[C]//Advances in Neural Information Processing Systems 33, December 6-12, 2020, Virtual. [S.l.: s.n.], 2020.
- [17] Li H F, Qiu K J, Chen L, et al. SCAAttNet: semantic segmentation network with spatial and channel attention mechanism for high-resolution remote sensing images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 18(5): 905-909.
- [18] Zhao Q, Liu J H, Li Y W, et al. Semantic segmentation with attention mechanism for remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5403913.
- [19] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [20] Gerke M. Use of the stair vision library within the ISPRS 2D semantic labeling benchmark (Vaihingen)[J]. 2014. [10.13140/2.1.5015.9683](https://doi.org/10.13140/2.1.5015.9683).
- [21] 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.
- [22] 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.