

# 增强细节的 RGB-IR 多通道特征融合语义分割网络

谢树春<sup>1</sup>, 陈志华<sup>1</sup>, 盛斌<sup>2</sup>

(1. 华东理工大学 信息科学与工程学院, 上海 200237; 2. 上海交通大学 电子信息与电气工程学院, 上海 200240)

**摘要:** 现有基于深度学习的语义分割方法对于遥感图像的地物边缘分割不准确, 小地物分割效果较差, 并且 RGB 图像质量也会严重影响分割效果。提出一种增强细节的 RGB-IR 多通道特征融合语义分割网络 MFFNet。利用细节特征抽取模块获取 RGB 和红外图像的细节特征并进行融合, 生成更具区分性的特征表示并弥补 RGB 图像相对于红外图像所缺失的信息。在融合细节特征和高层语义特征的同时, 利用特征融合注意力模块自适应地为每个特征图生成不同的注意力权重, 得到具有准确语义信息和突出细节信息的优化特征图。将细节特征抽取模块和特征融合注意力模块结构在同一层级上设计为相互对应, 从而与高层语义特征进行融合时抑制干扰或者无细节信息的影响, 突出重要关键细节特征, 并在特征融合注意力模块中嵌入通道注意力模块, 进一步加强高低层特征有效融合, 产生更具分辨性的特征表示, 提升网络的特征表达能力。在公开的 Postdam 数据集上的实验结果表明, MFFNet 的平均交并比为 70.54%, 较 MFNet 和 RTFNet 分别提升 3.95 和 4.85 个百分点, 并且对于边缘和小地物的分割效果提升显著。

**关键词:** 遥感图像; 深度学习; 语义分割; RGB-IR 多通道; 细节特征抽取; 特征融合注意力

开放科学(资源服务)标志码(OSID):



中文引用格式: 谢树春, 陈志华, 盛斌. 增强细节的 RGB-IR 多通道特征融合语义分割网络[J]. 计算机工程, 2022, 48(10): 230-237, 244.

英文引用格式: XIE S C, CHEN Z H, SHENG B. Detail-enhanced RGB-IR multichannel feature fusion network for semantic segmentation[J]. Computer Engineering, 2022, 48(10): 230-237, 244.

## Detail-Enhanced RGB-IR Multichannel Feature Fusion Network for Semantic Segmentation

XIE Shuchun<sup>1</sup>, CHEN Zhihua<sup>1</sup>, SHENG Bin<sup>2</sup>

(1. College of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China;

2. School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

**[Abstract]** Existing semantic segmentation methods based on deep learning are inaccurate for the edge segmentation of remote sensing images. The segmentation effect of small ground objects is poor, and the quality of RGB images seriously affects the segmentation effect. This study proposes a detail-enhanced RGB-IR multichannel feature fusion network for semantic segmentation, named MFFNet. The detail feature extraction module is used to obtain detailed features of RGB and infrared images. These are fused to generate a more distinctive feature representation and offset the missing information of RGB images relative to infrared images. While fusing the detail and high-level semantic features, the feature fusion attention module is used to adaptively generate different attention weights for each feature map to obtain an optimized feature map with more accurate semantic information and prominent detail information. The detail feature extraction and feature fusion attention module structures are designed to correspond to each other at the same level and suppress the influence of interference or irrelevant detail information when fusing high-level semantic features, highlight key detail features, and embed the channel attention module in the feature fusion attention module. This strengthens the effective fusion of high- and low-level features and generates a more discriminative feature representation, thereby improving the feature expression ability of the network. Experiments on the public Postdam dataset show that the mean intersection over union of MFFNet is 70.54%, which is an improvement of 3.95 and 4.85 percentage points compared with that of MFNet and RTFNet, respectively, and the segmentation effect for edges and small ground objects is significantly improved.

**[Key words]** remote sensing image; deep learning; semantic segmentation; RGB-IR multichannel; detail feature extraction; feature fusion attention

DOI: 10.19678/j.issn.1000-3428.0063316

**基金项目:** 国家自然科学基金面上项目“光照一致的立体视频编辑与合成技术研究”(61672228); 装备预研教育部联合基金“基于遥感数据的目标识别预警和关联决策技术研究”(6141A02022373)。

**作者简介:** 谢树春(1996—), 男, 硕士研究生, 主研方向为数字图像处理、计算机图形学; 陈志华、盛斌, 教授、博士。

**收稿日期:** 2021-11-23 **修回日期:** 2022-01-06 **E-mail:** czh@ecust.edu.cn

## 0 概述

遥感图像中包含非常丰富的地物信息, 遥感图像的利用价值在于可对其进行重要信息的提取, 但处理过程也非常复杂。遥感图像语义分割是提取遥感图像重要信息的前提, 也是学术界和工业界的研究难点。遥感图像覆盖范围广, 地物信息复杂多样, 存在很多的小地物类别, 使得分割难度加大, 并且存在类间相似性和类内差异性, 进一步加大了分割难度。

全卷积神经网络是目前实现图像语义分割的主流方法。基于全卷积神经网络提出的 FCN<sup>[1]</sup> 是深度学习应用在图像语义分割的代表方法, 其作为一种端到端的分割方法, 应用于图像语义分割领域时得到了很好的效果。SegNet<sup>[2]</sup> 和 U-Net<sup>[3]</sup> 是对 FCN 的改进, SegNet 引入了更多的跨层连接, U-Net 在上采样阶段依然保留有大量的通道, 使得网络可以将上下文信息向更高层分辨率传播。ERFNet<sup>[4]</sup> 使用残差连接来加速特征学习以及消除梯度消失的现象, 并使用深度可分离卷积来减少网络的参数数量, 提高模型推算速度。SKASNet<sup>[5]</sup> 构建了一个新的残差模块, 通过调节感受野的大小获得多尺度信息。DeepLabv3+<sup>[6]</sup> 引入语义分割常用的编解码结构并使用可任意控制编码器提取特征的分辨率, 通过空洞卷积平衡精度和耗时。现有的遥感图像语义分割方法主要对上述模型进行微调与改进。文献[7-8] 将基于 U-Net 改进的网络结构用于遥感图像上进行语义分割时获得了可观的效果。RWSNet<sup>[9]</sup> 将 SegNet 和随机游走相结合, 缓解了分割对象边界模糊的问题。

近年来, 研究者设计了很多用于提高语义分割网络性能的模块, 如受到广泛关注的注意力机制。注意力机制可以在网络训练过程中加强对一些重要特征区域或者重要特征通道的注意力, 提升网络对特征的表达能力。在 SENet<sup>[10]</sup> 中, 压缩、激励和重标定三个部分组成注意力机制, 使网络利用全局信息有选择地增强有用特征通道并抑制无用特征通道, 实现特征通道自适应校准。CBAM<sup>[11]</sup> 将注意力机制同时运用在通道和空间两个维度上来提升网络模型的特征提取能力。卷积神经网络中的卷积单元每次只关注邻域卷积核大小的区域, 是局部区域的运算。文献[12] 提出了 Non-local Neural Networks 用于捕获长距离关系。文献[13] 在特征提取网络中加入注意力模块来减少分割精度损失。文献[14] 基于 U-Net 改进通过注意力机制以提高模型的灵敏度, 并抑制无关特征区域的背景影响。文献[15] 通过全局注意力金字塔与通道注意力解码器来解决地物小和类内尺度存在差异的问题。

特征融合也是一种提高分割性能的流行方法。高层语义特征具有大的语义结构, 但对小结构丢失

严重, 低层细节特征保留了丰富的细节信息, 但语义类别信息很差。文献[16-17] 通过设计一个优秀的特征融合方法进一步提高了网络的分割性能。FPN<sup>[16]</sup> 最初用于目标检测任务, 但是也可以应用于语义分割, 通过按元素相加的方式来融合全局和局部特征, 而 PSPNet<sup>[17]</sup> 特征融合更强调全局特征, 文献[18] 则提出了一种增强特征融合的解码器来提高语义分割模型的性能。遥感图像语义分割网络需要设计优异的特征融合方法来加强高低层特征的融合, 对此, 文献[19] 通过高层语义特征和低层细节特征融合来提高模型的分割准确率, 文献[20] 设计了自适应融合模块 (AFM)。一些通过结合边缘检测<sup>[21]</sup> 和融入深度信息<sup>[22-23]</sup> 的网络模型也能一定程度上提升语义分割的性能。此外, 光照不足的条件也会导致 RGB 图像质量下降。红外图像可以很好地弥补光照不足等问题, 捕捉到更多 RGB 图像所缺失的信息。基于 RGB-IR (RGB 图像和相对应的 Infrared 图像按通道维度叠加后得到 RGB-Infrared 图像) 的语义分割模型 MFNet<sup>[24]</sup>、RTFNet<sup>[25]</sup> 通过融合 RGB 和红外信息来克服光照不足以及天气条件恶劣等问题, 提高了语义分割的性能。

现有基于 RGB-IR 的语义分割模型没有很好地将 RGB 和红外信息充分融合, 也较少提取到 RGB 图像相对于红外图像所缺失的信息。本文提出一个细节特征抽取模块来提取 RGB 图像和红外图像的细节特征信息同时进行融合, 生成更具区分性的特征表示并弥补 RGB 图像相对于红外图像所缺失的信息。此外, 提出一种特征融合注意力模块来有效融合细节特征和高层语义特征, 得到具有更准确语义信息的优化特征图。基于以上模块, 构建增强细节的 RGB-IR 多通道特征融合语义分割网络 MFFNet, 通过融合 RGB 图像和红外图像, 解决现有方法地物边缘分割不准确、小地物分割效果差的问题, 同时提升光照不足、恶劣天气条件情况下的分割效果。

## 1 RGB-IR 多通道特征融合语义分割网络

### 1.1 细节特征抽取模块

为了解决上文提到的遥感图像语义分割存在的难题, 并提高模型的分割性能, 需要提取更多的图像细节特征, 以便后续融合到高层语义特征中来进一步丰富细节信息。此外, 需要将抽取到的 RGB 和红外图像的细节特征进行深层次融合, 生成更具分辨性的特征表示, 弥补 RGB 图像相对于红外图像所缺失的信息, 提高模型的特征表达能力, 进而提升模型的分割性能。本文提出由注意力模块构成的细节特征抽取模块, 如图 1 所示。

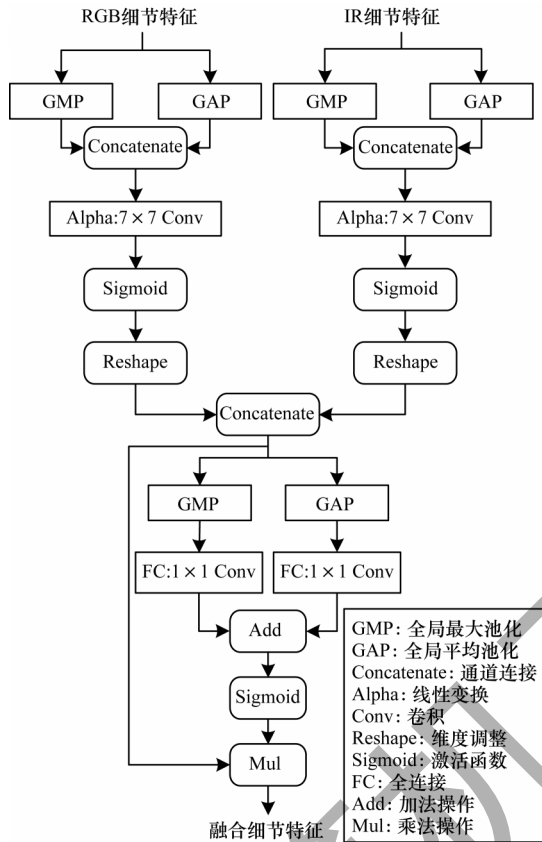


图1 细节特征抽取模块

Fig.1 Detail feature extraction module

细节特征抽取模块首先对某一阶段的特征图 $X$  ( $X$ 是从RGB或红外图像中提取到的特征图)分别进行全局平均池化操作和全局最大池化操作,然后对得到的结果进行拼接操作,再进行Alpha线性变换得到Alpha特征,之后通过一个Sigmoid激活函数来得到注意力权重以加强对重要特征区域的注意力,最后和特征图 $X$ 相乘得到优化后的特征图 $Y$ 。由于细节特征抽取模块是接在低层卷积层后的,因此 $Y$ 包含了非常丰富的细节信息,并且一些重要的细节特征也是被加强的,此计算过程和文献[11]中的空间注意力相似,计算公式如下:

$$Y = \frac{1}{1 + \exp(-(\mathbf{W}_a([\text{AvgPool}(X); \text{MaxPool}(X)]))})} \cdot X \quad (1)$$

其中: $X$ 为输入特征图; $\mathbf{W}_a$ 是可学习的权重矩阵,通过空间域的 $7 \times 7$ 卷积实现;AvgPool和MaxPool分别为全局平均池化操作和全局最大池化操作。

分别对同一阶段RGB和红外图像中提取到的特征图 $X_{\text{rgb}}$ 、 $X_{\text{ir}}$ 进行上述计算得到 $Y_{\text{rgb}}$ 、 $Y_{\text{ir}}$ ,然后再对这两个优化后的细节特征图采用拼接操作进行融合,再通过通道注意力来自适应地为通道重新分配不同的权重,以优化融合后的细节特征图,最终得到融合细节特征图 $Z$ 。此过程的计算公式如下:

$$Z = \sigma(f_{c \rightarrow c/r}^{1 \times 1}(f_{c/r \rightarrow c}^{1 \times 1}(\text{AvgPool}([Y_{\text{rgb}}; Y_{\text{ir}}]))) + f_{c \rightarrow c/r}^{1 \times 1}(f_{c/r \rightarrow c}^{1 \times 1}(\text{MaxPool}([Y_{\text{rgb}}; Y_{\text{ir}}])))) \quad (2)$$

其中: $\sigma$ 为Sigmoid激活函数; $f_{c \rightarrow c/r}^{1 \times 1}$ 为2D卷积操作,卷积核大小为 $1 \times 1$ ,通道数从 $c$ 减为 $c/r$ ;  $f_{c/r \rightarrow c}^{1 \times 1}$ 为2D卷积操作,卷积核大小为 $1 \times 1$ ,通道数从 $c/r$ 增加到 $c$ , $r$ 为减少率;AvgPool和MaxPool分别为全局平均池化操作和全局最大池化操作。

至此,已经从RGB图像和红外图像中抽取到了细节特征信息,并且得到了融合后的细节特征图。然后需要把这些融合后的细节特征图整合到高级语义特征中来增加丰富细节信息,以优化网络的特征表达能力,从而提高模型的灵敏度。

## 1.2 特征融合注意力模块

本文提出的特征融合注意力模块不像其他网络那样简单地将低层细节特征和高层语义特征进行相加或者拼接,这样做会把干扰或者无关信息同时也融合到高层语义特征中,并且不能很好地融合高低层特征。本文把通过细节特征抽取模块得到的RGB和红外图像融合后的细节特征通过特征融合注意力模块来融合进高层语义特征,从而在和高层语义特征进行融合时抑制干扰或者避免无细节信息的影响,突出重要关键细节特征。此外,本文在特征融合注意力模块中嵌入通道注意力模块,产生更具分辨性的特征表示,以提高网络的灵敏度。

特征融合注意力模块如图2所示。融合高低层特征的操作一般有拼接操作和相加操作。首先采用拼接操作来结合高低层特征,并通过一个卷积核大小为 $1 \times 1$ 的卷积层来减少通道数,提高模型的推理速度,然后经过一个卷积核大小为 $3 \times 3$ 的卷积层,最后通过一个通道注意力机制生成新的特征图 $X_{\text{fuse}}$ 。

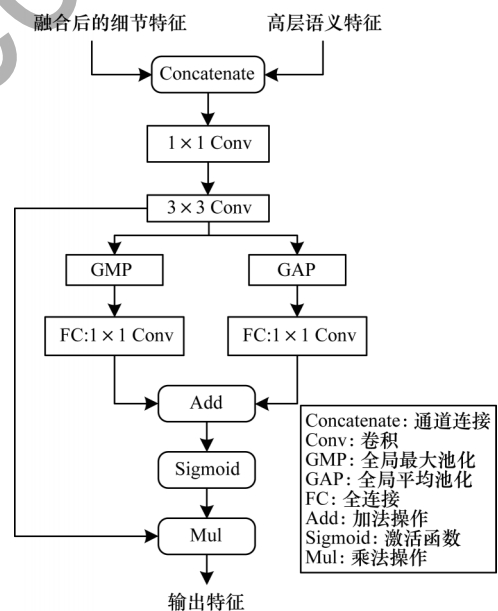


图2 特征融合注意力模块

Fig.2 Feature fusion attention module



特征融合注意力模块的计算公式如下:

$$\begin{aligned} X_{\text{fuse}} = & M_c(f^{3 \times 3}(f_{c \rightarrow c/r}^{1 \times 1}([X_1; X_2])))f^{3 \times 3}(f_{c \rightarrow c/r}^{1 \times 1}([X_1; X_2])) \quad (3) \\ M_c(X) = & \sigma(f_{c \rightarrow c/r}^{1 \times 1}(f_{c/r \rightarrow c}^{1 \times 1}(\text{AvgPool}(X)))) + \\ & f_{c \rightarrow c/r}^{1 \times 1}(f_{c/r \rightarrow c}^{1 \times 1}(\text{MaxPool}(X))) \quad (4) \end{aligned}$$

其中:  $X$  为输入特征图;  $\sigma$  为 Sigmoid 激活函数;  $f_{c \rightarrow c/r}^{1 \times 1}$  为 2D 卷积操作, 卷积核大小为  $1 \times 1$ , 通道数从  $c$  减为  $c/r$ ;  $f_{c/r \rightarrow c}^{1 \times 1}$  为 2D 卷积操作, 卷积核大小为  $1 \times 1$ , 通道数从  $c/r$  增加到  $c$ ,  $r$  为减少率; AvgPool 和 MaxPool 分别为全局平均池化操作和全局最大池化操作;  $X_1$  为细节分支生成的低层细节特征图;  $X_2$  为高层特征图;  $f^{3 \times 3}$  为 2D 卷积操作, 卷积核大小为  $3 \times 3$ , 此卷积操作

后跟随有 BatchNorm 操作和 ReLu 操作。

特征融合注意力模块融合细节特征抽取模块得到的 RGB 和红外图像融合后的细节特征和高层语义特征, 在每一次上采样阶段前都采用特征融合注意力模块进行特征融合来丰富细节信息和上下文信息, 保证像素语义分类准确, 同时优化小地物的分割效果, 进一步提高模型的分割准确率, 使网络模型更好地定位到边界。

### 1.3 多通道特征融合网络

本文基于细节特征抽取模块和特征融合注意力模块, 提出一种增强细节的 RGB-IR 多通道特征融合语义分割网络 MFFNet, 如图 3 所示。

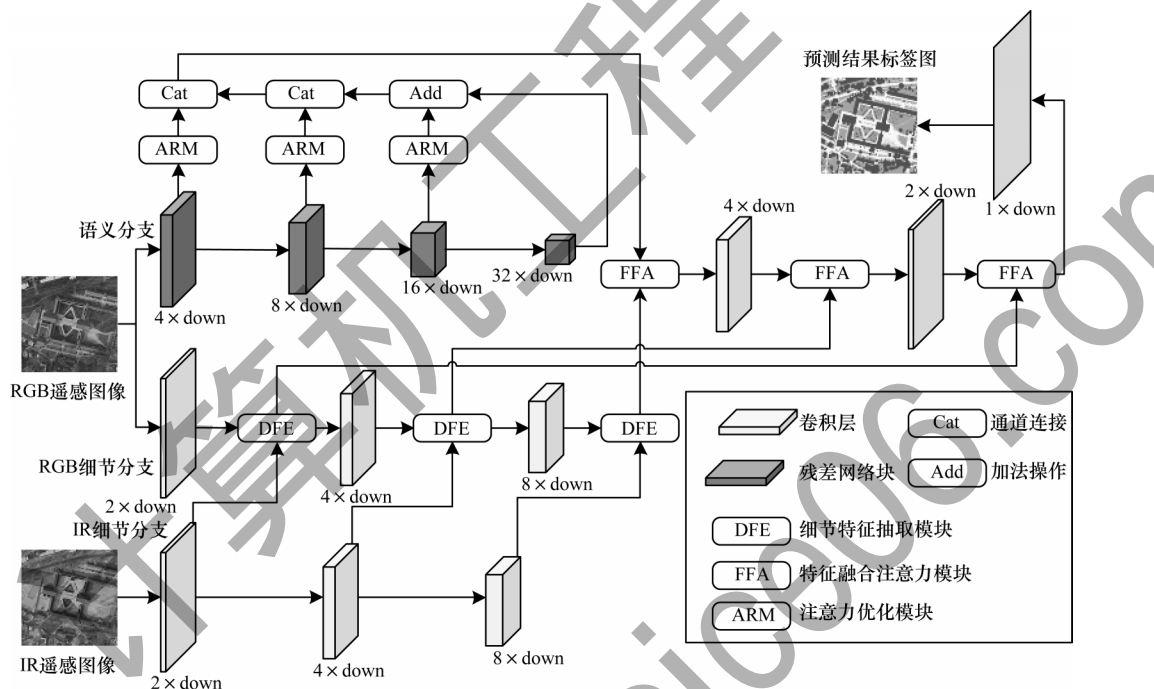


图3 MFFNet网络结构

Fig.3 Network structure of MFFNet

MFFNet包括细节分支和语义分支这两个分支。细节分支通过细节特征抽取模块从RGB图像和红外图像中抽取到细节特征信息, 并且得到融合后的细节特征。语义分支使用轻量级的残差网络 ResNet18 作为主干网络, 从而进行快速下采样以提取高层语义特征。得益于 BiSeNet<sup>[26]</sup> 的启发, 本文在语义分支中还利用了一个注意力优化模块来优化输出特征, 注意力优化模块结构如图 4 所示。最后, 在 MFFNet 的上采样阶段把融合后的细节特征通过特征融合注意力模块整合到高级语义特征中来增加丰富细节信息, 以优化网络的特征表达能力, 从而提高模型的灵敏度。

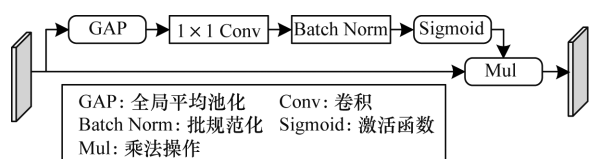


图4 注意力优化模块

Fig.4 Attention optimization module

### 1.4 损失函数

为了更好地指导模型训练进而提高地物边界的分割效果以及模型整体的分割性能, 受文献[27]的启发, 本文在遥感图像语义分割常用的交叉熵损失函数基础上加权边界损失<sup>[27]</sup>和 Jaccard 损失。在损失函数中, 加权边界损失可以指导模型训练进一步生成更好的地物边界分割效果。通过在损失函数中加权 Jaccard 损失直接指导模型训练, 能够有效提高模型整体的分割性能。

交叉熵损失函数是目前流行的语义分割任务中使用的损失函数, 用于指导模型进行训练。交叉熵损失函数  $E_{\text{loss}}$  的定义如下:

$$E_{\text{loss}} = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C q_c^{(n)} \cdot \text{lb } p_c^{(n)} \quad (5)$$

其中:  $N$  是小批量样本的数量;  $p_c^{(n)}$  是样本  $n$  分类为  $c$  类别的 softmax 概率;  $q_c^{(n)}$  是以 one-hot 编码时相应样本类别的标签;  $C$  是所有类别数。

交叉熵损失函数通过对所有像素的求和计算得出,不能很好地反映不平衡类。中位数频率平衡加权交叉熵损失函数考虑到了不平衡类问题,通过在训练集中统计类别的中位数频率和实际类别频率的比率来进行加权损失。中位数频率平衡加权交叉熵损失函数的定义如下:

$$M_{\text{loss}} = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C w_c \cdot q_c^{(n)} \cdot \ln p_c^{(n)} \quad (6)$$

$$w_c = \frac{\text{median}(f_c | c \in C)}{f_c} \quad (7)$$

其中:  $w_c$  是类别  $c$  的权重;  $f_c$  是类别  $c$  的像素的频率;  $\text{median}(f_c | c \in C)$  是所有  $f_c$  的中位数。

边界损失函数建立在边界度量边界  $F_1$  得分的基础上,因此,应先定义边界准确率和边界召回率。边界准确率  $P$  和边界召回率  $R$  分别定义如下:

$$P = \frac{1}{|B_p|} \sum_{x \in B_p} [[d(x, B_g) < \theta]] \quad (8)$$

$$R = \frac{1}{|B_g|} \sum_{x \in B_g} [[d(x, B_p) < \theta]] \quad (9)$$

其中:  $B_p$  表示预测边界;  $B_g$  表示真实标签边界;  $\theta$  是预定义的阈值,实验时默认取 3;  $[[\cdot]]$  表示逻辑表达式的指示函数。

边界度量边界  $F_1$  得分和边界损失函数  $B_{\text{loss}}$  定义如下:

$$F_1 = \frac{2PR}{P+R} \quad (10)$$

$$B_{\text{loss}} = 1 - F_1 \quad (11)$$

Jaccard 损失函数  $J_{\text{loss}}$  定义如下:

$$J_{\text{loss}} = 1 - \frac{|y_p \cap y_g|}{|y_p \cup y_g|} \quad (12)$$

其中:  $y_p$  和  $y_g$  分别表示预测标签和真实标签。

总的损失函数  $L_{\text{loss}}$  定义如下:

$$L_{\text{loss}} = aM_{\text{loss}} + bB_{\text{loss}} + cJ_{\text{loss}} \quad (13)$$

其中:  $a$ 、 $b$  和  $c$  分别是中位数频率平衡加权交叉熵损失、边界损失和 Jaccard 损失相应的权重系数。

## 2 实验与分析

### 2.1 数据集

实验使用的测试基准数据集是由国际摄影测量与遥感协会 (ISPRS) 组织发布的 Postdam 数据集。摄影测量学的研究方向之一是从机载传感器获取的数据中自动提取城市物体。这项任务的挑战性在于,在高分辨率的图像数据中,诸如建筑物、道路、树木和汽车之类的地面物体,同类对象有着非常不同的外观,这导致了较大的组内差异,而组间差异却很小。Postdam 数据集包括 6 种地面物体:不透水地面(例如道路),建筑物,低矮植被、树木,汽车,杂物。Potsdam 数据集包含 38 张高分辨率的 RGB 和 IR 遥

感图像,图像分辨率大小均为  $6\,000 \times 6\,000$  像素。图 5 所示为 Postdam 数据集的部分示例图。

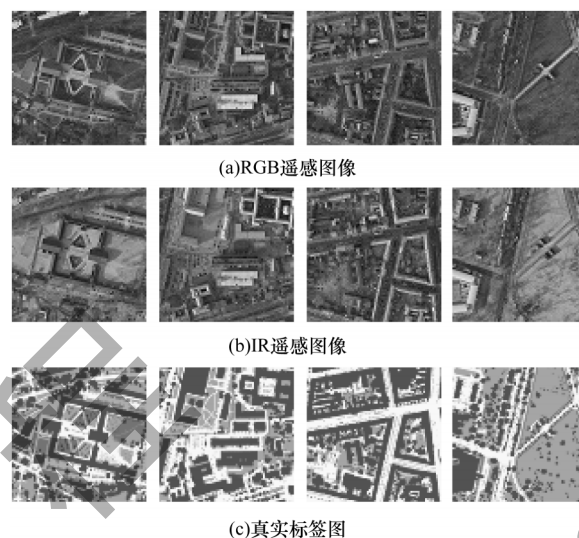


图 5 Postdam 数据集的部分示例图

Fig.5 Part of sample images in Postdam data set

### 2.2 评价指标

平均交并比 (Mean Intersection over Union, MIoU) 是语义分割的标准评价指标,整体准确率、精确率、召回率和 F1 分数是遥感图像语义分割最常用的评价指标。本文使用平均交并比、整体准确率、精确率、召回率和 F1 分数来度量本文提出的模型。平均交并比是对每一类预测的结果和真实值的交集与并集的比值求和平均的结果,交并比 (Intersection over Union, IoU) 利用混淆矩阵得到,计算公式如下:

$$I_{\text{IoU}} = \frac{T_p}{T_p + F_p + F_n} \quad (14)$$

其中:  $T_p$  代表真阳性,表示某一给定类别中被正确分类的像素数;  $F_p$  代表假阳性,表示被错误分类到特定类别的其他类别的像素数;  $F_n$  表示假阴性,表示一个给定类别被错误分类为其他类别的像素数。

整体准确率是正确标记的像素总数除以像素总数。精确率  $P_{\text{Precision}}$ 、召回率  $R_{\text{Recall}}$  以及 F1 分数  $F_1$  利用混淆矩阵得到,计算公式如下:

$$P_{\text{Precision}} = \frac{T_p}{T_p + F_p} \quad (15)$$

$$R_{\text{Recall}} = \frac{T_p}{T_p + F_n} \quad (16)$$

$$F_1 = 2 \times \frac{P_{\text{Precision}} \times R_{\text{Recall}}}{P_{\text{Precision}} + R_{\text{Recall}}} \quad (17)$$

### 2.3 实验结果与分析

本文模型使用开源库 PyTorch 1.7.1 和 torchvision 0.8.2 实现,实验使用 NVIDIA 公司的 GeForce RTX 090 GPU, 24 GB 的内存, CUDA 的版本是 11.2。本文提出的模型是轻量级的,在训练时设置 mini-batch 大小为 48,使用 Adam 作为优化算法应对梯度下降问题,学习率大

小设置为  $5 \times 10^{-4}$ , 权重衰减因子设置为  $2 \times 10^{-4}$ , 学习率衰减因子设置为 0.1, 每训练 120 个 epoch 调整学习率, 共训练 200 个 epoch。

为了验证本文提出的 MFFNet 模型对 RGB 遥感图像和红外遥感图像融合的有效性, 以及是否能够提高小地物和边界的分割效果, 在公开的 Potsdam 数据集上进行实验。Potsdam 数据集被广泛用于评估遥感图像语义分割模型的性能, 包含 38 张高分辨率的 RGB 遥感图像和相对应红外遥感图像, 每张图像分辨率大小为  $6\,000 \times 60\,000$  像素。本文将该数据集图像分为 20 张训练图像、10 张验证图像和 8 张测试图像, 然后进行数据预处理, 通过裁剪 20 张训练图像, 每张图像都用滑动窗口的方法进行裁剪, 步长为滑动窗口的大小, 获得 225 张  $400 \times 400$  像素的图像, 共得到 4 500 张训练图像, 然后再进行数据增强操作 (包括旋转、模糊、添加噪声等) 扩充一倍训练数据集, 最后共得到 9 000 张  $400 \times 400$  像素的训练图像。使用同样的滑动窗口方法裁剪验证集图像和测试集图像, 得到 2 250 张  $400 \times 400$  像素的验证集图像和 1 800 张  $400 \times 400$  像素的测试集图像, 相对应的红外遥感图像也以同样的方式进行裁剪。

本文使用平均交并比、整体准确率、精确率、召回率和 F1 分数来评估 MFFNet, 实验结果如表 1 所示, 其中, 加粗数据表示最优值, 3c 表示网络是三通道, 输入只有 RGB 图像, 4c 是将 RGB 和 IR 通道叠加作为输入, 对比实验的网络模型中 RTFNet 采用残差网络 ResNet50 作为主干网络, DeepLabv3+ 和 PSPNet 采用残差网络 ResNet101 作为主干网络。对比表 1 所有 RGB-IR 四通道作为输入的网络模型实验结果可以看出, 本文提出的 MFFNet 模型在上述的各个评价指标上都是最优的, 对于语义分割的标准评价指标平均交并比, MFFNet 较对比模型中最优的模型提升了 2.72 个百分点, 在其他各个评价指标上, MFFNet 较对比模型中最优的模型也都有很大的提升: 整体准确率提升 1.14 个百分点, 精确率提升 3.69 个百分点, 召回率提升 0.04 个百分点, F1 分数提升 2.04 个百分点。此外, 对比表 1 所有 RGB-IR 四通道作为输入的网络模型实验结果可以看出, 本文提出的 MFFNet 模型不仅仅是对于整体的分割效果是最好的, 而且对于小物体类别车的分割效果在每个评价指标上也是最优的, 相对于对比实验中最优的模型而言有非常大的提升: 交并比提升 7.3 个百分点, 精确率提升 9.52 个百分点, F1 分数提升 4.6 个百分点。

表 1 Potsdam 数据集上不同模型的性能对比  
Table 1 Performance comparison of different models in Potsdam data set %

网络模型	平均交并比	整体准确率	精确率	召回率	F1 分数	车类别交并比	车类别精确率	车类别 F1 分数
ERFNet <sup>[4]</sup> (4c)	67.82	84.07	78.06	81.25	79.37	73.79	76.28	84.92
ERFNet <sup>[4]</sup> (3c)	68.20	84.38	78.67	81.17	79.69	73.94	76.23	85.02
SegNet <sup>[2]</sup> (4c)	65.90	82.77	76.16	80.39	77.95	71.04	73.82	83.07
SegNet <sup>[2]</sup> (3c)	66.46	82.83	76.89	80.45	78.45	73.46	76.06	84.70
PSPNet <sup>[17]</sup> (4c)	67.22	84.18	77.84	81.13	79.16	68.23	70.85	81.11
PSPNet <sup>[17]</sup> (3c)	67.19	84.08	77.73	80.75	78.94	69.93	72.51	82.30
DeepLabv3+ <sup>[6]</sup> (4c)	67.29	84.10	77.66	80.59	78.89	72.24	74.77	83.88
DeepLabv3+ <sup>[6]</sup> (3c)	67.55	83.85	78.14	80.86	79.27	73.12	75.75	84.48
U-Net <sup>[3]</sup> (4c)	61.17	79.57	73.65	76.16	74.29	66.81	71.56	80.10
U-Net <sup>[3]</sup> (3c)	61.08	79.54	73.66	74.90	73.96	68.71	73.54	81.45
MFNet <sup>[24]</sup>	66.59	83.19	77.29	79.69	78.34	74.49	77.63	85.38
RTFNet <sup>[25]</sup>	65.69	82.81	76.78	79.46	77.87	69.11	72.87	81.73
FuseNet <sup>[23]</sup>	63.31	81.19	74.04	78.16	75.67	66.73	70.92	81.73
MFFNet	70.54	85.32	81.75	81.29	81.41	81.79	87.15	89.98

从表 1 中还可以看到, 在对比模型中, 除 PSPNet 和 UNet 外, 其他模型直接把 RGB 三通道 (3c) 图像改为 RGB-IR 四通道 (4c) 图像作为网络输入, 不仅不能改善反而还降低了网络模型的分割效果, PSPNet 和 UNet 直接把 RGB 三通道 (3c) 图像改为 RGB-IR 四通道 (4c) 图像作为网络输入, 在整体分割性能上虽然有略微的一点提升, 但对于小地物类别车的分割效果却受到大幅影响。

图 6 和图 7 为在 Potsdam 数据集上的部分实验结

果图, 从中可以清楚地看到, 对比模型不能很好地分割小地物类别车, 小地物的边缘分割也是不准确的, 并且小区域的分割效果也很差。本文提出的 MFFNet 模型对小地物的分割效果明显优于对比模型, 小地物的分割效果很好, 不存在边缘分割不准确的情况, 并且对于小区域的分割效果要好很多。由此可以证明, 本文模型不仅可使遥感图像整体的分割效果有很大的提升, 对于图像中小地物的分割, 效果的提升也是非常明显的。



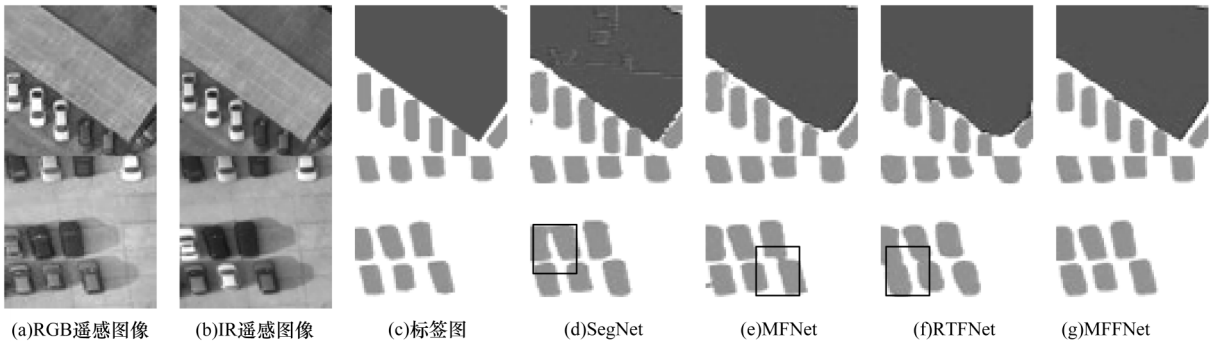


图6 Potsdam数据集上的实验的结果图1

Fig.6 Experimental result images 1 in Potsdam dataset

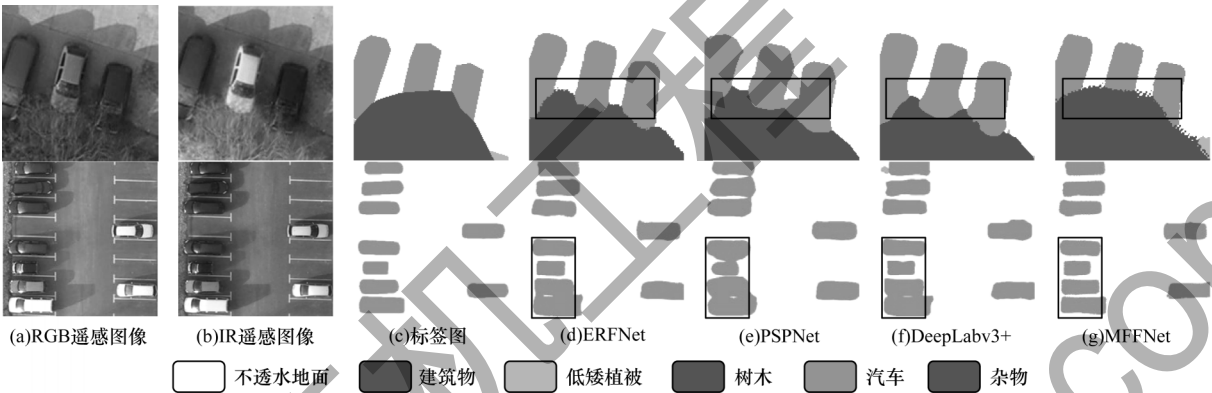


图7 Potsdam数据集上的实验的结果图2

Fig.7 Experimental result images 2 in Potsdam dataset

为了进一步说明本文提出的模型能够有效地整合RGB图像和红外图像的信息,在Postdam数据集上进行消融实验,将RGB和RGB-IR分别作为MFFNet网络输入。将RGB作为网络输入时,微调MFFNet网络,去掉IR细节分支,整体分割性能对比如图8所示,小地物车类别分割性能对比如图9所示,其中无填充的柱状图是RGB图像作为网络输入的实验结果,有填充的柱状图是RGB-IR图像作为网络输入的实验结果。在表2中,3c表示网络是三通道输入只有RGB图像,4c是将RGB和IR通道叠加作为输入。从表2中数据的比较可以清楚地看出,本文提出的模型对红外图像融合具有有效性,对于整体的分割效果和小地物的分割性能均较优。

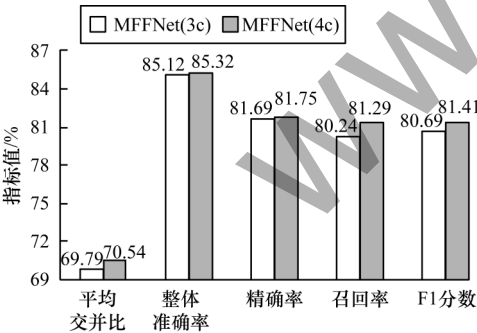


图8 RGB和RGB-IR分别作为MFFNet网络输入的整体分割性能

Fig.8 Overall segmentation performance when RGB and RGB-IR as input to the MFFNet network respectively

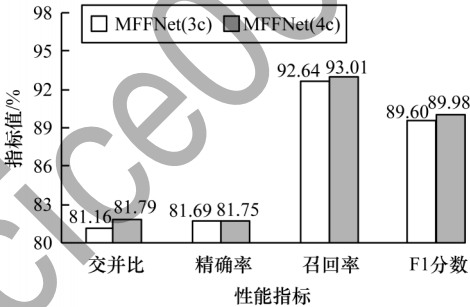


图9 RGB和RGB-IR分别作为MFFNet网络输入的车类别分割性能

Fig.9 Car category segmentation performance when RGB and RGB-IR as input to the MFFNet network respectively

表2 RGB和RGB-IR分别作为MFFNet网络输入的具体性能对比

Table 2 Specific performance comparison when RGB and RGB-IR as input to the MFFNet network %

评价指标	MFFNet(3c)	MFFNet(4c)
平均交并比	69.79	70.54
整体准确率	85.12	85.32
精确率	81.69	81.75
召回率	80.24	81.29
F1 分数	80.69	81.41
车类别交并比	81.16	81.79
车类别精确率	81.69	81.75
车类别召回率	92.64	93.01
车类别F1 分数	89.60	89.98

### 3 结束语

本文构建增强细节的 RGB-IR 多通道特征融合语义分割网络 MFFNet, 以解决遥感图像语义分割中存在的问题。提出一种能够有效融合 RGB 图像和红外图像的细节特征抽取模块, 从而获取丰富的融合细节信息, 并提出一种新的特征融合方法——特征融合注意力模块, 将细节特征抽取模块提取到的融合细节特征充分融合进高级语义特征中, 以优化网络的表达能力, 提高模型的灵敏度。在 Postdam 数据集上的实验结果证明了该模型的有效性。下一步将结合神经架构搜索 (Neural Architecture Search, NAS) 技术优化细节特征融合模块的结构, 加强 RGB 图像和红外图像细节特征信息的整合, 提高模型的分割性能, 同时降低模型的复杂度。

### 参考文献

- [1] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 3431-3440.
- [2] 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.
- [3] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin, Germany: Springer, 2015: 234-241.
- [4] ROMERA E, ÁLVAREZ J M, BERGASA L M, et al. ERFNet: efficient residual factorized ConvNet for real-time semantic segmentation[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(1): 263-272.
- [5] 谭镭, 孙怀江. SKASNet: 用于语义分割的轻量级卷积神经网络[J]. 计算机工程, 2020, 46(9): 261-267.
- [6] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 833-851.
- [7] DONG R S, PAN X Q, LI F Y. DenseU-net-based semantic segmentation of small objects in urban remote sensing images[J]. IEEE Access, 2019, 7: 65347-65356.
- [8] CUI B E, CHEN X, LU Y. Semantic segmentation of remote sensing images using transfer learning and deep convolutional neural network with dense connection[J]. IEEE Access, 2020, 8: 116744-116755.
- [9] JIANG J, LYU C J, LIU S Y, et al. RWSNet: a semantic segmentation network based on SegNet combined with random walk for remote sensing[J]. International Journal of Remote Sensing, 2020, 41(2): 487-505.
- [10] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 7132-7141.
- [11] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 3-19.
- [12] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 7794-7803.
- [13] 程晓悦, 赵龙章, 胡穹, 等. 基于密集层和注意力机制的快速语义分割[J]. 计算机工程, 2020, 46(4): 247-252, 259.
- [14] CHENG X Y, ZHAO L Z, HU Q, et al. Fast semantic segmentation based on dense layer and attention mechanism[J]. Computer Engineering, 2020, 46(4): 247-252, 259. (in Chinese)
- [15] GUO M Q, LIU H, XU Y Y, et al. Building extraction based on U-Net with an attention block and multiple losses[J]. Remote Sensing, 2020, 12(9): 1400.
- [16] WANG S Q, ZHANG C, WU M. Accurate semantic segmentation in remote sensing image[C]//Proceedings of International Conference on Computing and Pattern Recognition. New York, USA: ACM Press, 2019: 173-178.
- [17] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 936-944.
- [18] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 2881-2890.
- [19] 马震环, 高洪举, 雷涛. 基于增强特征融合解码器的语义分割算法[J]. 计算机工程, 2020, 46(5): 254-258, 266.
- [20] MA Z H, GAO H J, LEI T. Semantic segmentation algorithm based on enhanced feature fusion decoder[J]. Computer Engineering, 2020, 46(5): 254-258, 266. (in Chinese)
- [21] WANG E D, JIANG Y M, LI Y, et al. MFCSNet: multi-scale deep features fusion and cost-sensitive loss function based segmentation network for remote sensing images[J]. Applied Sciences, 2019, 9(19): 4043.
- [22] SHANG R H, ZHANG J Y, JIAO L C, et al. Multi-scale adaptive feature fusion network for semantic segmentation in remote sensing images[J]. Remote Sensing, 2020, 12(5): 872.
- [23] 王囡, 侯志强, 赵梦琦, 等. 结合边缘检测的语义分割算法[J]. 计算机工程, 2021, 47(7): 257-265.
- [24] WANG N, HOU Z Q, ZHAO M Q, et al. Semantic segmentation algorithm combined with edge detection[J]. Computer Engineering, 2021, 47(7): 257-265. (in Chinese)
- [25] 张娣, 陆建峰. 基于双目图像与跨级特征引导的语义分割模型[J]. 计算机工程, 2020, 46(10): 275-281, 288.
- [26] ZHANG D, LU J F. Semantic segmentation model based on binocular images and guidance of cross-level features[J]. Computer Engineering, 2020, 46(10): 275-281, 288. (in Chinese)

(下转第 244 页)



(上接第 237 页)

- [23] HAZIRBAS C, MA L, DOMOKOS C, et al. FuseNet: incorporating depth into semantic segmentation via fusion-based CNN architecture [C]//Proceedings of Asian Conference on Computer Vision. Berlin, Germany: Springer, 2017: 213-228.
- [24] HA Q S, WATANABE K, KARASAWA T, et al. MFNet: towards realtime semantic segmentation for autonomous vehicles with multi-spectral scenes [C]//Proceedings of International Conference on Intelligent Robots and Systems. Washington D. C. , USA: IEEE Press, 2017: 5108-5115.
- [25] SUN Y X, ZUO W X, LIU M. RTFNet: RGB-thermal fusion network for semantic segmentation of urban scenes[J]. IEEE Robotics and Automation Letters, 2019, 4(3): 2576-2583.
- [26] YU C Q, WANG J B, PENG C, et al. BiSeNet: bilateral segmentation network for real-time semantic segmentation [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 334-349.
- [27] BOKHOVKIN A, BURNAEV E. Boundary loss for remote sensing imagery semantic segmentation [C]//Proceedings of International Symposium on Neural Networks. Berlin, Germany: Springer, 2019: 388-401.

编辑 金胡考