计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200700182

基于 U-Net 特征融合优化策略的遥感影像语义分割方法

王施云　杨　帆
河北工业大学电子信息工程学院　天津 300401
(18222953150@163.com)

摘　要　高分辨率遥感影像的空间分辨率高、地物信息丰富、复杂程度高、各类地物的大小尺寸不一,这为分割精度的提高带来了一定的难度。为提高遥感影像语义分割精度,解决 U-Net 模型在结合深层语义信息与浅层位置信息时受限的问题,文中提出了一种基于 U-Net 特征融合优化策略的遥感影像语义分割方法。该方法采用基于 U-Net 模型的编码器-译码器结构,在特征提取部分沿用 U-Net 模型的编码器结构,提取多个层级的特征信息;在特征融合部分保留 U-Net 的跳跃连接结构,同时使用提出的特征融合优化策略,实现了高层语义特征与底层位置特征的融合-优化-再融合。此外特征融合优化策略还使用空洞卷积获取了更多的全局特征,并采用 Sub-Pixel 卷积层代替传统转置卷积,实现了自适应上采样。所提方法在 ISPRS 的 Potsdam 数据集和 Vaihingen 数据集上得到了验证,其总体分割精度、Kappa 系数和平均交并比 mIoU 3 个评价指标在 Potsdam 数据集上分别为 86.2%,0.82,0.77,在 Vaihingen 数据集上分别为 84.5%,0.79,0.69;相比传统的 U-Net 模型,所提方法的 3 个评价指标在 Potsdam 数据集上分别提高了 5.8%,8%,8%,在 Vaihingen 数据集上分别提高了 3.5%,4%,11% 。实验结果表明,基于 U-Net 特征融合优化策略的遥感影像语义分割方法,在 Potsdam 数据集和 Vaihingen 数据集上都能达到很好的语义分割效果,提高了遥感影像的语义分割精度。

关键词:深度学习;特征融合;遥感影像;空洞卷积;语义分割
中图法分类号　TP391

Remote Sensing Image Semantic Segmentation Method Based on U-Net Feature Fusion Optimization Strategy

WANG Shi-yun and YANG Fan
School of Electronic and Information Engineering,Hebei University of Technology,Tianjin 300401,China

Abstract　Due to the high spatial resolution of high-resolution remote sensing images,rich ground objects information,high complexity,uneven distribution of target categories and different sizes of various ground objects,it is difficult to improve the segmentation accuracy. In order to improve the semantic segmentation accuracy of remote sensing images and solve the problem that U-Net model is limited when combining deep semantic information and shallow position information,a semantic segmentation method of remote sensing images based on U-Net feature fusion optimization strategy is proposed. This method adopts the encoder-decoder structure based on U-Net network. In the feature extraction part of the network,the encoder structure of U-Net model is used to extract the feature information of multiple layers. In the feature fusion part,the jump connection structure of U-Net is retained,and at the same time,the feature fusion optimization strategy proposed in this paper is used to realize the fusion-optimization-refusion of high-level semantic features and low-level location features. In addition,the feature fusion optimization strategy uses dilated convolution to get more global features,and uses Sub-Pixel convolutional layer instead of traditional transposed convolution to achieve adaptive upsampling. This method is validated on the Potsdam dataset and Vaihingen dataset of ISPRS. The three evaluation indexes,overall classification accuracy,Kappa coefficient and mIoU in the verification are 86.2%,0.82,0.77 on Potsdam dataset,and 84.5%,0.79,0.69 on Vaihingen dataset. Compared with the traditional U-Net model,the three evaluation indicators are increased by 5.8%,8%,8% on Potsdam dataset,and 3.5%,4%,11% on Vaihingen dataset. Experimental results show that the remote sensing image semantic segmentation method based on the U-Net feature fusion optimization strategy has achieved good semantic segmentation effects on both the Potsdam dataset and the Vaihingen dataset,which can improve the accuracy of semantic segmentation of remote sensing images.

Keywords　Deep learning,Feature fusion,Remote sensing image,Dilated convolution,Semantic segmentation

到稿日期:2020-07-28　返修日期:2020-09-19　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家重点研发计划智能机器人专项(2019YFB1312102);河北省自然科学基金(F2019202364)
This work was supported by the National Key R&D Program Intelligent Robot Special Project (2019YFB1312102) and Natural Science Foundation of Hebei Province (F2019202364).
通信作者:杨帆(yangfan@hebut.edu.cn)

1 引言

随着遥感传感器技术的发展，遥感影像的空间分辨率不断提高，更高空间分辨率的影像不仅包含大量的地物细节特征，而且复杂程度也更高[1]，存在目标类别分布不平衡、不同目标边缘易重叠、个别目标尺度较小、纹理细节难以分辨等问题，为高分辨率遥感影像的语义分割带来了一定的技术难度。

遥感影像语义分割是根据影像的颜色、几何、纹理等特征信息，按照一定的规则和算法，为每一个像素添加一个类别标签的过程[2]。传统遥感图像语义分割通常采用场论和集合论相结合的方法，以及机器学习的方法[3]。这类方法都是建立在精准的数学模型之上，具有严密的数学论证，但通常分割的准确率较低，难以推广到其他数据集。随着深度学习的不断发展，新的方法也不断出现。

2015 年，Long 等[4]提出全卷积神经网络(Fully Convolutional Networks，FCN)，首次实现了端到端的语义分割，开创了全卷积网络用于图像语义分割的先河，并且在自然场景图像分割中取得了较好的成绩。FCN 网络将全连接层转化为卷积层，并使用上采样操作，以产生预测图。该网络在测试集的精度上相比非深度学习的方案提高很多，并且大大降低了流程的复杂度。为提高 FCN 的分割精度，Yu 等[5]引入了空洞卷积，通过特征图的感受野来提升输出特征的稠密度，从而使目标分割更加精准。基于相似原理并结合条件随机场原理，Chen 等[6]提出 DeepLab V1 图像分割算法，目前已经发展到 DeepLab V3+，使目标分割的精度得到了进一步提高。

Ronneberger 等[7]在 MICCAI 会议上提出了 U-Net 模型。其采用的编码器(下采样)-译码器(上采样)对称结构和跳跃连接是一种非常经典的设计方法，跳跃连接将深层的语义信息与浅层的位置信息结合在一起，可以获取更多的位置信息，在当时的医学图像分割领域取得了最佳效果。相较于 FCN 模型，U-Net 模型具有结构简单、训练时间短、训练参数少等优点，被后续的研究者广泛使用并加以改进。

Yuan[8]利用 ResNet34 作为 U-Net 的编码器部分，通过使用 Jaccard 损失和二元交叉熵损失的加权和作为总损失函数，显著地提高了 U-Net 对道路分割提取的精度。Su 等[9]提出使用改进的 U-Net 对高分辨率遥感图像进行语义分割，该方法能够快速地分割图像，且对设备要求较低，但分割精度不够高。Berman 等[10]使用一个通用的方法来优化 Jaccard 损失，该优化方法显著提高了语义分割的准确性，同时模型的推理速度几乎不受影响。Shi 等[11]提出 Sub-Pixel 卷积层，其可以在不添加人为数值的情况下增加图像的分辨率。Maggiori 等[12]提出了一种高分辨率遥感影像分类网络结构，该结构引入多层感知机(Multi-Layer Perceptron，MLP)进行多分辨率特征图融合，以调和卷积神经网络特征图中的识别和定位矛盾。

针对 U-Net 模型采用跳跃连接方式在结合深层语义信息与浅层位置信息时受限的问题，本文提出了一种基于 U-Net 跳跃连接的特征融合优化策略，采用一种融合-优化-再融合的结构，结合空洞卷积和 Sub-Pixel 卷积层并引入多层感知机，以提高高分辨率遥感影像的语义分割精度。

2 相关背景知识

2.1 空洞卷积

VGG 网络证明了使用小卷积核叠加来取代大卷积核，可以实现与大卷积核同样的感受野大小并且减少参数，但叠加操作时感受野是线性增长的，其公式为 $(KernelSize-1)*layers+1$，而空洞卷积可以按指数级增长感受野。

由空洞卷积原理[5]可知：当卷积核为 3×3 的空洞卷积的扩张率 $rate=1$ 时，其卷积操作等同于标准卷积，感受野为 3×3；当扩张率 $rate=2$ 时，相当于在对输入特征做卷积操作时，每两个相邻的作用点之间都添加了 $rate-1$，即 1 个空洞，此时当前层的感受野为 5×5，叠加 $rate=1$ 的卷积作用时，感受野增大到 7×7；当扩张率 $rate=4$ 时，当前层感受野为 9×9，同理与上述两个空洞卷积叠加操作时，能达到 15×15 的感受野。不考虑叠加作用，空洞卷积当前层的感受野计算公式为：

$$rf=(rate-1)*(ksize-1)+ksize \tag{1}$$

叠加时感受野的计算公式为：

$$RF'=(RF-1)*stride+(rate*(ksize-1)+1) \tag{2}$$

其中，$ksize$ 是卷积核大小，$rate$ 是空洞卷积扩张率，$stride$ 是卷积操作步长，RF 是前一层的感受野大小，RF' 是叠加卷积操作后的感受野大小。

空洞卷积的好处是可扩大感受野，让每个卷积输出都包含较大范围的信息。在图像需要全局信息时，都可以应用空洞卷积，比如图像分割。

2.2 Sub-Pixel 卷积层

上采样也就是反卷积，在特征解码部分，主要用来将特征图恢复成输入图像的大小，本文方法中主要使用 Sub-Pixel 卷积层代替传统上采样层，通过较小的窗口得到更大的感受野。

转置卷积操作首先对低分辨率的特征图进行补零或双线性插值等操作来增加分辨率，再通过卷积层使得放大的特征图变精细。Sub-Pixel 卷积层则无需补充任何人为数值，可直接在低分辨率特征图上进行卷积操作，并对生成的特征图进行重新排列以实现分辨率的增加。在分辨率提高之前进行卷积操作，可以以较小的卷积核获得更多的上下文信息。Sub-Pixel 卷积层上采样操作可由式(3)表示：

$$PS(T)_{x,y,c}=T_{\lfloor x/r\rfloor,\lfloor y/r\rfloor,c\cdot r\cdot \mathrm{mod}(y,r)+c\cdot \mathrm{mod}(x,r)+c} \tag{3}$$

其中，低分辨率特征图 T 的形状为 $R^{H\times W\times C\cdot r^2}$，$r$ 为上采样倍数，经过重新排列操作 PS 后，得到形状为 $R^{rH\times rW\times C}$ 的高分辨率特征图。x,y,c 为高分辨率特征图的空间位置，则对应的低分辨率特征图的亚像素空间位置为 $\lfloor x/r\rfloor$，$\lfloor y/r\rfloor$，$c\cdot r\cdot \mathrm{mod}(y,r)+c\cdot \mathrm{mod}(x,r)+c$，其中 $\lfloor\cdot\rfloor$ 为向下取整，$\mathrm{mod}(\cdot)$ 为取余。

2.3 批标准化

传统网络通常采用随机梯度下降算法来训练网络，但训练过程中模型参数不断地改变，每一层的输入数据也随之改变，导致学习速度太慢、学习效果严重依赖于初始的数据分布、反向传播过程出现梯度爆炸和梯度消失等问题[13]，批数

据之间的差别过大也会影响训练速率[14]。2015 年，Google 提出了批标准化(Batch Normalization，BN)[15]的概念，在网络的每一层输入之前加入一个归一化层，对之前的数据作归一化处理后再输入当前层训练。

归一化公式为：

$$\bar{x}^{(k)}=\frac{x^{(k)}-E(x^{(k)})}{\sqrt{var(x^{(k)})}} \tag{4}$$

$$y^{(k)}=\gamma^{(k)}\bar{x}^{(k)}+\beta^{(k)} \tag{5}$$

其中，$E(x^{(k)})$代表所有训练集的平均值；$var(x^{(k)})$代表所有训练集的方差；$\bar{x}^{(k)}$代表某个训练值的 batch，这意味着不管使用哪一组 batch，样本的方差和均值都是固定的。另外，$\gamma^{(k)}$和 $\beta^{(k)}$是缩放和偏移系数，均为可学习的参数，可以让网络自己找到一个正态变换后不破坏原特征分布的平衡状态。

3 U-Net 跳跃连接的特征融合优化策略

本文提出的基于 U-Net 特征融合优化策略的网络模型如图 1 所示。

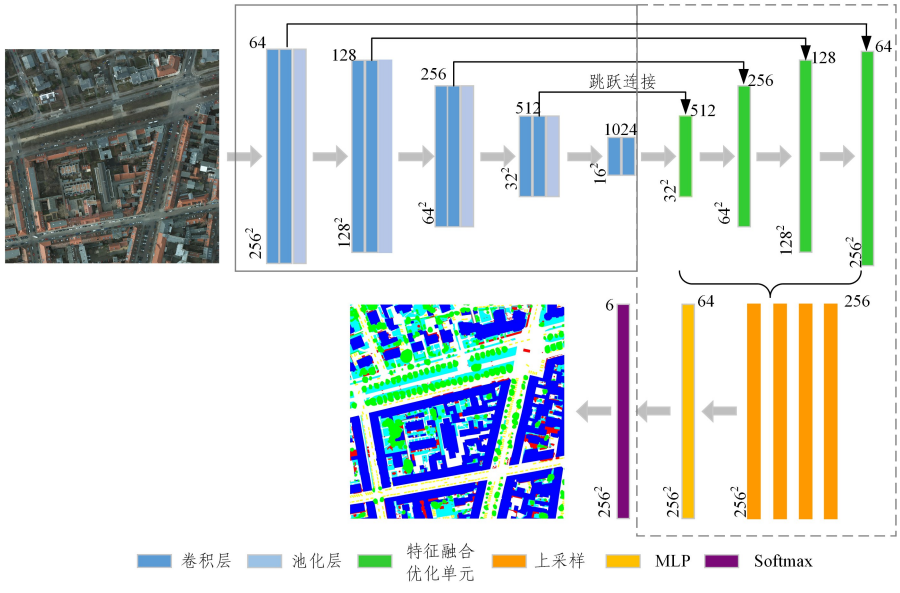


图 1 网络模型图(电子版为彩色)

Fig. 1 Network model diagram

图 1 中实线框是特征提取部分，虚线框是特征融合部分，融合后的特征图通过 Softmax 分类器得到分类结果。

3.1 基于 U-Net 模型的特征提取

实验主要采用 U-Net 模型的编码器部分对高分辨率遥感影像进行特征提取，表 1 为其网络模型结构，其对应图 1 的实线框内容。

表 1 U-Net 编码器网络结构

Table 1 U-Net encoder structure

Layer	Filter Size	Filter Number	Stride
Conv1_1	3	64	1
Conv1_2	3	64	1
Pool1	2	—	2
Conv2_1	3	128	1
Conv2_2	3	128	1
Pool2	2	—	2
Conv3_1	3	256	1
Conv3_2	3	256	1
Pool3	2	—	2
Conv4_1	3	512	1
Conv4_3	3	512	1
Pool4	2	—	2
Conv5_1	3	1024	1
Conv5_2	3	1024	1

U-Net 编码器共包含五层，其中除第五层外每一层都包含 2 个 3×3 的卷积层和 1 个 2×2 的池化层，使用 ReLU 作为激活函数，而第五层只包含 2 个卷积层和激活函数。在卷积操作中为了维持输出数据尺寸与输入数据一致，方便计算，设置 Padding 为 SAME 对输入数据进行零值填充，同时设置卷积操作的步长 $stride=1$，以达到保持数据大小不变只增加通道数的目的。池化操作采用的是池化窗口为 2×2 的最大池化，步长 $stride=2$，输入数据经过最大池化操作后，输出数据大小变为原来的 1/2，通道数不发生变化。将尺寸为 $h\times w\times d$ 的图像输入编码器，在经过一系列的卷积、池化操作后，输出尺寸为 $\frac{h}{16}\times\frac{w}{16}\times 1024$。

为了降低模型对初始权重的敏感度，加快训练速度，同时有效防止过拟合，在特征提取部分的卷积层与激活层之间加入了批标准化层，对数据进行归一化处理。

3.2 U-Net 特征融合优化策略

3.2.1 跳跃连接结构

对于原始的 U-Net 网络模型来说，其主要特点就是采用了跳跃连接的方式，将深层的语义信息上采样后与浅层的位置信息按深度连接在一起，并通过逐层连接的方式构成了一个优美的“U”型结构。

本实验在对特征融合部分的改进中保留了这种跳跃连接方式，然后提出了一种特征融合优化策略，细化每一层的跳跃连接融合特征，并实现多尺度的特征融合，达到了更好的分割效果。

3.2.2 特征融合优化策略

特征融合优化策略是一个特征融合-优化-再融合的过程，网络模型如图 1 中虚线框所示，包含特征融合优化单元以及上采样融合和多层感知机 MLP。特征融合优化单

元的结构如图 2 所示。

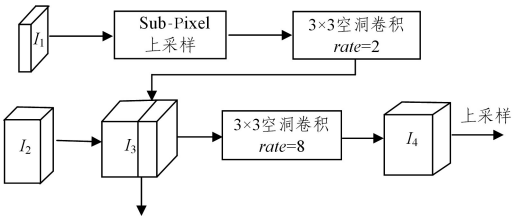


图 2 特征融合优化单元

Fig. 2 Feature fusion optimization unit

特征融合优化单元包含融合优化策略中的融合和优化过程。这个融合与 U-Net 的融合理念一致，是基于跳跃连接结构的相邻两个编码器输出特征的融合。假设 $\frac{W}{2}\times\frac{H}{2}\times C_1$ 的深层特征 I_1 与 $W\times H\times C_2$ 的浅层特征 I_2，是两个相邻层级特征提取的结果，分辨率相差一倍。在融合策略中，首先对特征 I_1 进行 Sub-Pixel 上采样，使其与特征 I_2 的分辨率相同，通过扩张率 $rate=2$ 的空洞卷积操作，对特征信息进行细化，之后与上一层的编码器输出特征 I_2 按通道的维度进行串联拼接，得到大小为 $W\times H\times C_3$ 的融合后的特征 I_3，根据 Sub-Pixel 的上采样原理可得 $C_3=\frac{C_1}{4}+C_2$。融合后的特征则继续通过上采样的操作增加分辨率，与更浅一层的编码器输出特征相融合，直至特征空间分辨率大小恢复到与输入图像一致。

优化过程则是对相邻层拼接结果，即 I_3 进行优化。用 $rate=8$ 的 3×3 空洞卷积对串联后的特征进行操作得到优化后的特征 I_4。这种方式可以通过较小的卷积核获得更大的感受野，从而获得更多的位置信息。

再融合则是对每一层的优化结果进行特征融合。将各层优化后的特征上采样至输入图像的尺寸大小，得到该层级最终提取到的特征信息。将这些特征图全部按深度拼接在一起，这样就更好地融合了更多的深层语义信息以及浅层位置信息，会尽可能地获取全局的上下文信息。

将再融合得到的特征图输入 MLP，如图 1 中的黄色模块所示，学习如何融合多分辨率的特征图。MLP 是拥有一个隐藏层的计算系统，其工作原理为：采用一个 1×1 的卷积层以及激活函数来拟合特征融合函数，将融合操作变成可以学习的过程。

3.3 分类器与损失函数的选择

实验采用的遥感影像包含 6 类地物，是多标签分类的语义分割，故采用 Softmax 分类器对特征图进行分类。Softmax 函数的作用是将每个类别所对应的输出分量归一化，使各个分量的和为 1，也就是将输出向量的分量值转化成了模型预测的输入像素元属于每个类别的概率。Softmax 函数的表达式为：

$$S_i=\frac{e^{v_i}}{\sum_{j=1}^{N}e^{v_j}} \tag{6}$$

其中，v_i 是像素元属于第 i 类的分量值。通过 e^{v_i} 操作将其转化成非负数，再对其进行归一化处理。

输出层经过 Softmax 函数作用之后，使用交叉熵 Cross Entropy 作为损失函数进行 loss 计算。Cross Entropy 的表达式如式(7)所示：

$$H_y(y')=-\sum_i y_i\log(y_i') \tag{7}$$

其中，y_i 是 label 的第 i 个值，y_i' 是经过 Softmax 归一化后输出的对应分量。由式(7)可知，当分类越准确时，y_i' 所对应的分量会越接近 1，从而 $H_y(y')$ 值也就会越小。

4 实验结果与分析

4.1 数据集与数据增强

实验使用的是 ISPRS WG III/4 竞赛中公开的两个高分辨率遥感影像数据集(Potsdam 数据集和 Vaihingen 数据集)，来对模型进行评估。

Potsdam 数据集共包括 38 幅图像，图像的空间分辨率为 5cm，像素大小均为 6000×6000。图像均提供了相应的标签，标签共分为 6 类，分别是不透水表面、建筑物、低矮植物、树、车以及其他。实验使用 RGB 三通道的图像，并选取其中的 16 幅图像作为训练集，3 幅图像作为验证集，4 幅图像作为测试集。

Vaihingen 数据集共包含 33 幅图像，图像空间分辨率为 9cm，像素大小不一，平均尺寸为 2494×2064。标签类别和数量与 Potsdam 数据集一致。实验使用三通道的 IRRG 图像，并选取其中的 16 幅图像作为训练集，5 幅图像作为验证集，12 幅图像作为测试集。

由于数据集的图像尺寸过大，为了适应硬件环境以及网络模型结构，对遥感影像进行剪裁切块处理。有实验表明[16]，当剪裁尺寸固定，块间覆盖率为 75%时能得到更好的实验结果，故本实验采用 75%的块间覆盖率对数据集图像进行 256×256 的切块处理。此外，为了防止过拟合，实验通过水平、垂直翻转，旋转 90°，180°，270°等操作，对数据集进行数据增强，结果如图 3 所示。

(a)原图 (b)水平翻转 (c)垂直翻转 (d)旋转 90° (e)旋转 180° (f)旋转 270°

图 3 数据增强

Fig. 3 Data enhancement

4.2 评价指标

实验采用总体精度 OA，F1-Sorce，Kappa 系数以及平均交并比 mIoU 这 4 个评价指标对网络分割性能进行评估。这些评价指标都是基于混淆矩阵计算的。

混淆矩阵设定了两个类别 Positive 和 Negative，分别简写为 P 和 N。矩阵中 TP 和 TN 分别代表模型预测成功的两个类别的数量，FP 是真实值为 N 模型预测为 P 的数量，FN 则是真实值为 P 模型预测为 N 的数量。以下评价指标都将依据这 4 个量做计算。

4.2.1 总体精度 OA

模型的性能越好，总体精度 OA 越高，其公式为：

$$OA=\frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

4.2.2 F1-Sorce

F1-Sorce 的公式如下：

$$F1\text{-}Score=\frac{2PR}{P+R} \tag{9}$$

其中，P 代表 Precision，是在模型预测为 P 的结果中，预测正确的比重。R 代表 Recall，是真实值为 P 的结果中，模型预测正确的比重。

$$Precision=\frac{TP}{TP+FP} \tag{10}$$

$$Recall=\frac{TP}{TP+FN} \tag{11}$$

F1-Score 指标综合了 Precision 与 Recall 的结果，取值范围从 0 到 1，1 代表模型的输出最好，0 代表模型的输出结果最差。

4.2.3 Kappa 系数

Kappa 系数的计算公式为：

$$Kappa=\frac{p_o-p_e}{1-p_e} \tag{12}$$

其中，p_o 是每一类正确分类的样本数之和除以总样本数，也就是总体分类精度。假设每一类的真实样本个数分别为 a_1，a_2，…，a_C，而预测出来的每一类的样本个数分别为 b_1，b_2，…，b_C，总样本个数为 n，则有：

$$p_e=\frac{a_1\times b_1+a_2\times b_2+\cdots+a_C\times b_C}{n\times n} \tag{13}$$

4.2.4 平均交并比 mIoU

交并比是实际分割区域与模型预测分割区域的交集和并集的比值，其计算式为：

$$IoU=\frac{TP}{TP+FP+FN} \tag{14}$$

$$mIoU=\frac{1}{n}\sum_{i=1}^{n}IoU \tag{15}$$

通常采用平均交并比 $mIoU$ 来评价语义分割模型的性能，其公式为式(15)，其中 n 是类别数，i 代表第 i 类。

4.3 实验环境

本实验的硬件条件为 Intel Xeon Silver 4114 CPU，64 G 内存，2.2 GHz 主频，NVIDIA GeForce GTX 1080 Ti GPU，显卡内存为 12GB。软件环境以 Tensorflow r1.15 为开发框架，在 Ubuntu20.04 的系统及 Python3.6 的环境中运行。

4.4 训练策略

以数据增强后的数据集作为实验的输入图像，采用 He_Normal 初始化方式对权重进行初始化。初始学习率设置为 10^{-4}，批尺寸设置为 10，权重衰减系数为 0.0001，使用交叉熵函数作为损失函数，采用 Adam 优化器对参数进行优化。

在测试阶段，由于验证集图像尺寸较大，需要将图像裁成略小的图块来适应设备的限制。因此对验证集采取滑窗的形式，将其切割成 256×256 大小的图片，为避免剪裁拼接出现边界现象，将重叠率设置为 75%。

4.5 性能分析

为验证模型网络的性能，将本文方法与 FCN，U-Net 等通用分割模型进行对比，同时将未使用空洞卷积的模型(Ours-DC)与本文方法作对比。

4.5.1 Potsdam 数据集上的对比

4 种方法在 Potsdam 数据集的测试集中得到的指标数据如表 2 所列。

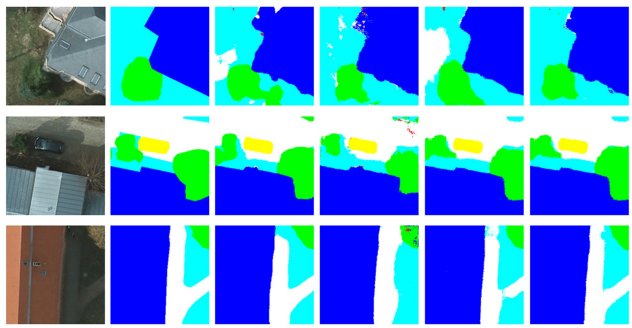
表 2 Potsdam 数据集在不同网络模型下的验证结果

Table 2 Validation results on Potsdam dataset under different network models

Method	F1-Socre/%					OA/%	Kappa	mIoU
	Impervious surfaces	Building	Low vegetation	Tree	Car			
FCN	86.0	89.9	77.2	81.1	86.9	82.8	0.76	0.72
U-Net	82.9	88.7	73.4	77.1	85.8	80.4	0.74	0.69
Ours-DC	84.2	88.6	78.1	82.3	87.8	83.4	0.78	0.73
Ours	**86.7**	**92.4**	**81.7**	**84.7**	**88.4**	**86.2**	**0.82**	**0.77**

从表 2 的对比数据可以看出，相比 U-Net 模型，FCN 在遥感影像数据集上有更好的表现。未使用空洞卷积的特征融合优化网络相比原始 U-Net 的分割结果，除建筑以外在其他类别的分割效果都有提升，总体的分割精度提升 3%，$Kappa$ 系数和平均交并比均提升 4%左右。而本文方法在各个评价指标上都表现出更好的效果：与未使用空洞卷积的结果相比，OA，$Kappa$ 系数和 $mIoU$ 分别提升了 2.8%，4%，4%；与 U-Net 模型相比，每一个类别的 $F1\text{-}Socre$ 指标都有至少 2%的提升，在低矮植物和树木的预测结果中更是有不少于 7%的提升，OA，$Kappa$ 系数，$mIoU$ 则分别提升了 5.8%，8%和 8%。

图 4 给出了本文方法与其他几种方法语义分割结果的比较。



(a)原始图像 (b)标签 (c)FCN (d)U-Net (e)本文方法-空洞卷积 (f)本文方法

注：白色代表不透水表面，蓝色代表建筑物，青色代表低矮植物，绿色代表树，黄色代表车，红色代表其他

图 4 Potsdam 数据集上的语义分割结果(电子版为彩色)

Fig. 4 Semantic segmentation results on Potsdam dataset

可以看出，FCN 模型错误地将低矮植物预测成了树木，而 U-Net 系列模型较好地区分了低矮植物和树木；本文方法的分割结果在形状上更接近标签结果，对于建筑边界的划分以及同类别大面积的区域划分也有更好的表现；FCN 将不透水表面预测成了低矮植物，使低矮植物与建筑连接在了一起；U-Net 虽然能正确预测了不同类别间的不透水表面，但没能正确预测同类间的不透水表面，而改进后的 U-Net 虽能正确预测以上两个问题，但显然使用空洞卷积的本文方法分割效果更佳。总体来说，本文方法在示例图中均有更好的表现。

4.5.2 Vaihingen 数据集上的对比

表 3 列出了本文方法与其他几种方法在 Vaihingen 数据集上得到的指标数据对比结果。

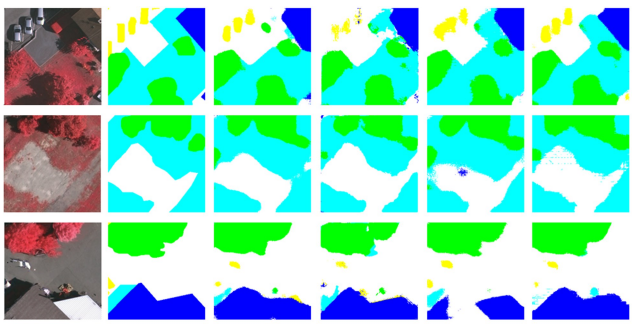
表 3 Vaihingen 数据集上不同网络模型下的验证结果

Table 3 Validation results on Vaihingen dataset under different network models

Method	*F1-Socre*/%					*OA*/%	*Kappa*	*mIoU*
	Impervious surfaces	Building	Low vegetation	Tree	Car			
FCN	84.1	85.9	76.2	83.7	65.5	82.1	0.76	0.62
U-Net	83.1	85.8	74.1	82.0	24.9	81.0	0.75	0.58
Ours-DC	85.8	88.8	**76.6**	83.8	49.7	83.6	0.78	0.64
Ours	**86.7**	**89.6**	76.5	**84.2**	**70.4**	**84.5**	**0.79**	**0.69**

由于 Vaihingen 数据集中车辆的数据量相对较少，因此分割效果整体偏差，但 FCN 在车辆分割中有相对较好的表现，*F1-Socre* 值为 65.5%，本文方法得到的 *F1-Socre* 值则是 70.4%，车辆分割效率相较于 FCN 有进一步提升。此外，本文方法在对不透水表面和建筑分割时也有不低于 2%的提升效果。从分割的整体效率看，本文方法在 *OA*，*Kappa* 系数和平均交并比 *mIoU* 3 个指标中得到的数据分别为 84.5%，0.79 和 0.69，对比 U-Net 模型分别提升了 3.5%，4% 和 11%。

在图 5 的语义分割图对比结果中，FCN 在对小尺度目标车辆分割时有较好的表现，但在不透水表面、建筑、低矮植物这些大尺度目标方面，往往不能完整分割出目标区域，无法清晰地划分出不同类别的边界；相比之下，本文方法在对低矮植物进行分割时得到的形状更完整，对不透水表面分割时得到结果更接近给定标签，对建筑物分割时边界更准确。因此，本文方法的整体分割效果比原始 U-Net 和未使用空洞卷积的融合策略更好。



(a)原始图像 (b)标签 (c)FCN (d)U-Net (e)本文方法-空洞卷积 (f)本文方法

注：白色代表不透水表面，蓝色代表建筑物，青色代表低矮植物，绿色代表树，黄色代表车，红色代表其他

图 5 Vaihingen 数据集语义分割结果(电子版为彩色)

Fig. 5 Semantic segmentation results on Vaihingen dataset

通过两个数据集的对比实验可以得到结论：相较于单纯的跳跃连接结构，特征融合优化单元保留了其深层语义信息与浅层位置信息结合的优点，同时又对融合后的特征通过空洞卷积操作以扩大感受野，细节优化后再融合，这种结构能更多地恢复影像的全局特征。本文用 Sub-Pixel 卷积层代替传统上采样，减少人为数值的添加；使用空洞卷积代替传统卷积，在相同大小的卷积核尺寸下提高感受野；使用多层感知机，能够使特征融合过程变为可学习的过程，提高特征融合效率。因此，基于 U-Net 特征融合优化策略的语义分割方法能够提高遥感影像的语义分割精度。

结束语 为提升 U-Net 模型对于遥感影像的语义分割精度，本文提出一种基于 U-Net 模型的特征融合优化策略，采用融合-优化-再融合方法对提取的特征进行融合，相比 U-Net 模型达到了更好的分割效果。下一步将研究多尺度的特征提取，进一步提升遥感图像的分割精度。

参 考 文 献

[1] WANG B, FAN D L. A Summary of the Research Progress of Deep Learning in Remote Sensing Image Classification and Recognition[J]. Bulletin of Surveying and Mapping, 2019, 503(2): 108-111, 145.

[2] QIN Y Q, CHI M M. High-resolution remote sensing image semantic segmentation method combined with scene classification data[J]. Computer Applications and Software, 2020, 37(06): 126-129, 134.

[3] WANG E D, QI K, LI X P, et al. Semantic segmentation method of remote sensing image based on neural network[J]. Acta Optica Sinica, 2019, 39(12): 93-104.

[4] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 3431-3440.

[5] YU F, KOLTUN V. Multi-Scale Context Aggregation by Dilated Convolutions[C] // International Conference on Learning Representations. San Juan, Puerto Rico, 2016.

[6] CHEN L C, PAPANDEROU 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 & Machine Intelligence, 2016, 40(4): 834-848.

[7] RONNEBERGER O, FISCHER P, BROX T, et al. U-net: Con-

volutional networks for biomedical image segmentation[J]. Medical Image Computing and Computer Assisted Intervention, 2015, 28(4): 234-241.

[8] YUAN J Y. Automatic building extraction in aerial scenes using convolutional networks[J]. arXiv: 1602.06564, 2016.

[9] SU J M, YANG L X, JING W P. Semantic segmentation method of high-resolution remote sensing image based on U-Net[J]. Computer Engineering and Applications, 2019, 55(7): 207-213.

[10] BERMAN M, TRIKI A R, BLASCHKO M B. The Lovász-softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: UT, 2018: 4413-4421.

[11] SHI W Z, CABALLERO J, HUSZAR F, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, 2016: 1874-1883.

[12] MAGGIORI E, TARABALKA Y, CHARPIAT G, et al. High-resolution aerial image labeling with convolutional neural networks [C]// IEEE Transactions on Geoscience and Remote Sensing. 2017: 7092-7103.

[13] ZHOU J Y, ZHAO Y M. Overview of Convolutiotnal Neural Networks in Image Classification and Target Detection[J]. Computer Engineering and Applications, 2017, 53(13): 34-41.

[14] PASCANU R, MIKOLOV T, BENGIO Y. On the difficulty of training recurrent neural networks[C]// Proceedings of the 30th International Conference on Machine Learning (CML2013). Atlanta, GA, USA, 2013: 1310-1318.

[15] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv: 1502.03167v3, 2015.

[16] XU Z J, YANG X B, HE L M, et al. Multiscale remote sensing semantic segmentation network[J/OL]. Computer Engineering and Applications: 1-9[2020-07-18]. http://kns.cnki.net/kcms/detail/11.2127.TP.20200423.1009.006.html.

WANG Shi-yun, born in 1994, postgraduate. Her main research interests include intelligent information processing and so on.

YANG Fan, born in 1966, Ph.D, professor, Ph.D supervisor. His main research interests include computer vision inspection technology, image processing and pattern recognition research.