北京大学学报(自然科学版) 第 55 卷 第 6 期 2019 年 11 月
Acta Scientiarum Naturalium Universitatis Pekinensis, Vol. 55, No. 6 (Nov. 2019)
doi: 10.13209/j.0479-8023.2019.106

基于多任务学习的高分辨率遥感影像建筑实例分割

惠健[1,2] 秦其明[1,2,3,†] 许伟[1,2] 隋娟[1]

1. 北京大学遥感与地理信息系统研究所, 北京大学地球与空间科学学院, 北京 100871; 2. 空间信息集成与3S工程应用北京市重点实验室, 北京 100871; 3. 自然资源部地理信息系统技术创新中心, 北京 100871;
† 通信作者, E-mail: qmqinpku@163.com

摘要 针对基于深度神经网络的高分辨率遥感影像建筑物提取算法中将建筑物提取视为二分类问题(即将遥感影像中的像素点划分为建筑物与非建筑两类)而无法区分建筑物个体的局限性, 将基于 Xception module 改进的 U-Net 深度神经网络方法与多任务学习方法相结合进行建筑物实例分割, 在获取建筑物二分类结果的同时, 区分不同建筑物个体, 并选择 Inria 航空影像数据集对该方法进行验证。结果表明, 在高分辨率遥感影像的建筑物二分类提取方面, 基于 Xception module 改进的 U-Net 方法明显优于 U-Net 方法, 提取精度升高 1.4%; 结合多任务学习的深度神经网络方法不仅能够实现建筑物的实例分割, 而且可将二分类建筑物的提取精度提升约 0.5%。

关键词 多任务学习; 建筑物提取; 深度神经网络; 实例分割

Instance Segmentation of Buildings from High-Resolution Remote Sensing Images with Multitask Learning

HUI Jian[1,2], QIN Qiming[1,2,3,†], XU Wei[1,2], SUI Juan[1]

1. Institute of Remote Sensing and Geographic Information System, School of Earth and Space Sciences, Peking University, Beijing 100871; 2. Beijing Key Lab of Spatial Information Integration and 3S Application, Beijing 100871; 3. Geographic Information System Technology Innovation Center, Ministry of Natural Resources, Beijing 100871;
† Corresponding author, E-mail: qmqingpku@163.com

Abstract At present, building extraction from high-resolution remote sensing images using deep neural network is viewed as a binary classification problem, which divides the pixels into two categories, building and non-building, but it cannot distinguish individual buildings. To solve this problem, the U-Net modified with Xception module and multitask learning are combined to apply to the instance segmentation of buildings, which both acquires the binary classification and distinguishes the individual buildings. Inria aerial imagery is used as the research dataset to validate the algorithm. The results show that the binary classification performance of U-Net modified with Xception outperforms U-Net by about 1.4%. The multitask driven deep neural network not only accomplishes the instance segmentation of buildings, but also improves the accuracy by about 0.5%.

Key words multitask learning; building extraction; deep neural network; instance segmentation

建筑物提取是遥感影像信息获取的关键环节和研究热点, 近十年来, 相关研究成果已广泛应用于监测土地利用变化、城市扩张和灾害预警评估等方面, 对政府部门的政策制定和地理信息数据库的更新具有重要的参考意义[1–2]。随着深度学习理论的发展, 深度神经网络模型已被不同行业广泛使用, 并在计算机视觉任务中取得良好的表现[3–4]。人们在深度学习与遥感影像应用相结合方面进行探索, 验证使用深度神经网络, 尤其是卷积神经网络处理遥感数据的可行性, 并提出适用于高分辨率遥感影

国家重点研发计划(2017YFB0503905)资助
收稿日期: 2019–01–02; 修回日期: 2019–05–09

像建筑物提取的深度神经网络模型。

Mnih[5]提出基于块状区域的卷积神经网络和航空遥感影像的道路与建筑物提取方法。Alshehhi 等[6]改进 Minh 的模型, 用全局平均池化层(global average layer)替代全连接层(fully connected layer), 改善了建筑物与道路的预测精度。Maggiori 等[7]和 Huang 等[8]分别利用全卷积神经网络[9]及其变种进行建筑物提取, 消除由块状区域带来的不连续性, 同时提高建筑物提取精度。Wu 等[10]使用 U-Net 网络[11]提取建筑物, 并提出多约束方法, 增强深度神经网络的多尺度特征表示。除使用成熟的深度神经网络外, 一些学者结合遥感影像的特点(如多源数据和多尺度特性), 在改进现有模型的基础之上, 设计适用于遥感影像的深度神经网络模型。考虑到遥感影像的多尺度特征, Audebert 等[12]设计多核卷积层(multi-kernel convolution layer), 改进原有的 SegNet 网络模型[13], 提高了预测精度。Xu 等[14]和 Chen 等[15]利用 ResNet 网络[16]提取影像特征, 改善了全卷积神经网络对遥感影像中目标的分割精度。Pan 等[17]考虑遥感数据的多源特性, 提出融合 Lidar 数据与光学遥感数据的深度神经网络模型。

虽然基于高分辨率遥感影像与深度神经网络融合的建筑物提取结果表现良好, 但是利用改进深度神经网络结构来提高建筑物提取精度的研究尚有很大的发展空间, 还需进行深入的研究。

多任务学习指同时训练基于一组相同参数的多个任务。Zhang 等[18]通过增加 4 个辅助任务(包括脸部属性识别、头部姿势判别等), 提高了人脸特征点检测的精度。Bischke 等[19]同时训练建筑物二分类提取和遥感影像中建筑物内部点距离分类, 提升了深度卷积神经网络提取遥感影像建筑物的精度。Mou 等[20]通过增加辅助任务, 减少遥感影像中车辆边界的粘连, 提高目标提取的效果。由此可见, 将多任务与深度学习算法相融合, 可以有效地提高目标提取的精度。然而, 基于高分辨率遥感影像、多任务学习和改进深度神经网络结构的算法研究比较少见, 需要进一步探索和挖掘。

目前, 利用深度神经网络进行高分辨率遥感影像建筑物提取的研究中, 通常将遥感影像的像素点分为建筑物与非建筑物两类, 没有区分不同的建筑物个体。本文采用多任务学习方法, 在对遥感影像的像素点进行二分类的基础上, 增加辅助任务, 产生遥感影像中不同建筑物个体的像素点对应的高维特征向量, 使得同一建筑物个体的像素点对应的高维向量在高维空间中聚集, 不同建筑物个体的向量聚类中心彼此远离; 以二分类预测算法为掩膜滤除非建筑物像素点, 使用聚类算法对高维特征向量进行聚类, 完成不同建筑物个体的区分。同时, 本文引入 Xception module[21]改造 U-Net 神经网络, 以期提高深度神经网络的特征提取能力。

1 数据来源与研究区域

本文选用年法国国家信息与自动化研究所 2018 年发布的 Inria 航空影像数据集[22]作为研究数据。该数据集的空间分辨率为 0.3 m, 包含覆盖 5 个城市的 180 张影像, 每个城市拥有 36 张高分辨率遥感影像。由于需要区分建筑物个体, 本文选用美国奥汀区域(30°16′2″N, 97°45′50″W)的 36 张影像, 并采用 Maggiori 等[22]的划分方法, 将其中 31 张用于训练, 5 张用于测试。每张影像的像素点为 5000×5000, 覆盖范围约为 2.25 km^2。考虑到计算机性能, 以 384 个像素点为步长, 将测试影像原图裁剪为 416×416 个像素点, 并剔除没有建筑物的影像, 得到 4526 张训练样本和 762 张测试样本。为了生成实例分割对应的样本标签, 本文采用 scikit-image 中的函数, 由二分类真值图像生成对应的实例标记, 给不同建筑物分配不同的标签值。样本情况如图 1 所示, 图 1(c)中相同颜色的像素点表示属于同一个建筑物个体, 具有相同的标签值。

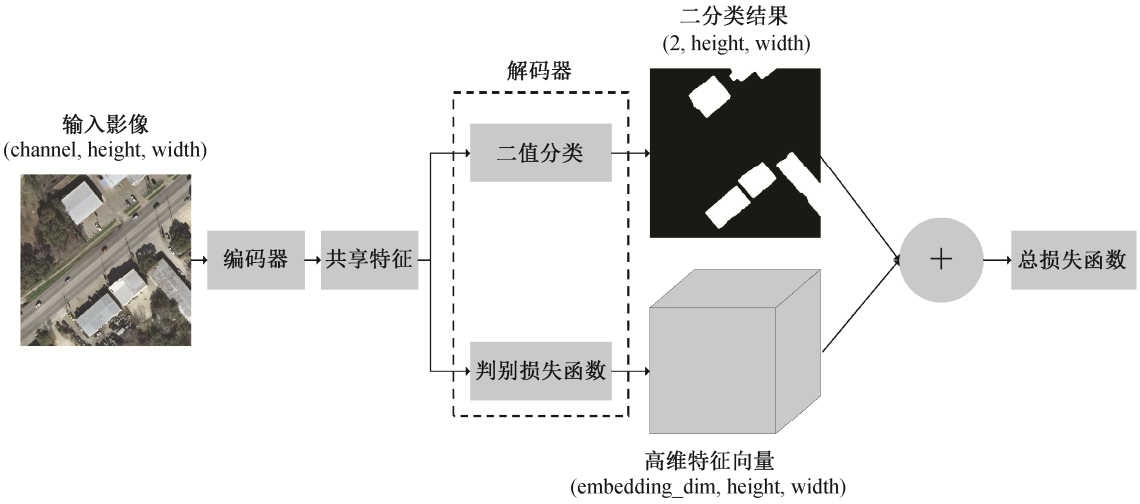
2 研究方法

本文提出基于 Xception module 的优化 U-Net 深度卷积网络模型。将 U-Net 中编码部分的连续卷积层用 Xception module 替换, 从而改善深度神经网络在高分遥感影像建筑物分类过程中特征提取的效果。同时, 为了区分遥感影像中的建筑物个体, 在 U-Net 所有具有“编码–解码”的结构中增加一个解码器, 用以生成遥感影像中建筑物像素点对应的高维向量, 即生成与遥感图像具有相同空间分辨率的三维矩阵, 其中第 3 个维度为高维特征向量。两个任务分别具有各自的“解码器”, 共享同一个“编码器”。通过共享特征、训练相近任务的方式, 提高模型所提取特征的表示能力与预测精度。两个任务各自的损失函数之和构成模型的总损失函数。使用梯度下降算法或其改进算法(本文采用 Adam 算法), 迭代更新其模型权重, 直到损失函数值收敛。本文使用

(a) 原图; (b) 二分类(真值); (c) 实例分割(真值)

图 1　Inria 航空遥感数据集

Fig. 1　Inria aerial image dataset



width 为影像宽度, height 为影像高度, channel 为影像波段数, embedding_dim 为高维向量的维度。下同

图 2　多任务实例分割模型

Fig. 2　Multi-task driven instance segmentation model

的模型如图 2 所示。

本文中模型的实例分割及高维特征向量的低维可视化工作流程如图 3 所示, 其中 p 代表像素点。首先输入遥感影像, 通过深度神经网络的正向(forward)运算后, 得到二分类预测结果(即建筑物与非建筑)及像素点分别对应的高维特征向量构成的多维矩阵。在进行建筑物的实例分割时, 使用深度神经网络建筑物二值分类结果作为掩膜, 滤除高维矩阵中所有非建筑物像素点对应的高维特征向量, 所有建筑物像素点(数目为 n)对应的高维特征向量(维度为 embedding_dim)组成一个含 n 个向量的向量集合。随后, 对该高维特征向量集合进行聚类, 得到最终的实例分割结果。图 3 左下角虚线框内展示判别损失函数解码器生成的特征图(feature map)(即图 3 中遥感影像所有像素点对应的高维特征向量), 以二分类结果为掩模, 滤除非建筑物像素点对应的向量, 最终得到所有建筑物像素点对应向量的过程。

2.1 U-Net 与 Xception module

U-Net 是一个经典的全卷积神经网络模型, 包括编码器和解码器两部分, 每个部分由连续的卷积层构成。在编码器中, 每两个卷积层之后会有一个池化层, 用于对特征图进行降尺度。与之相对应, 在解码器中同样存在一个上采样层, 用于提高特征图的分辨率。在编码过程中, 输入图像经过不断的卷积与池化, 得到不同尺度的特征图, 在此期间特征维度不断增加, 所学习的特征抽象程度不断提高, 分辨率不断降低。在相应的解码过程中, 特征维度降低, 分辨率增加, 最终得到与输入图像尺度相同

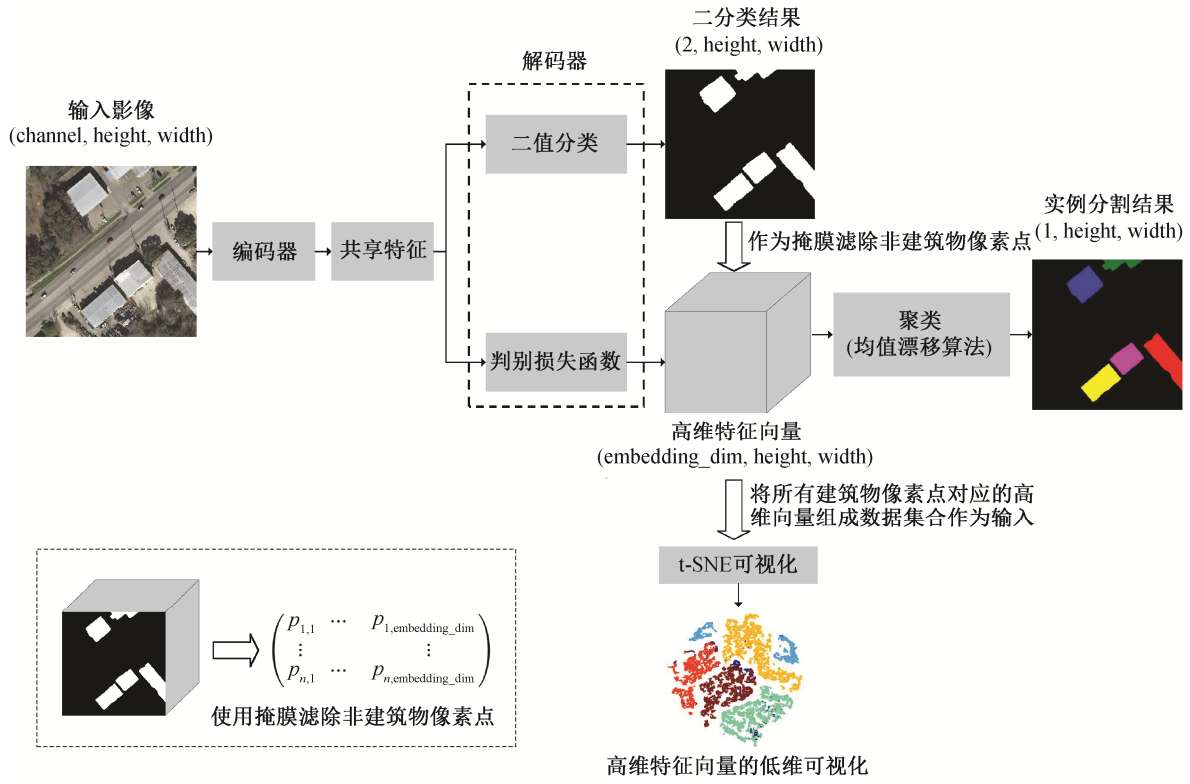


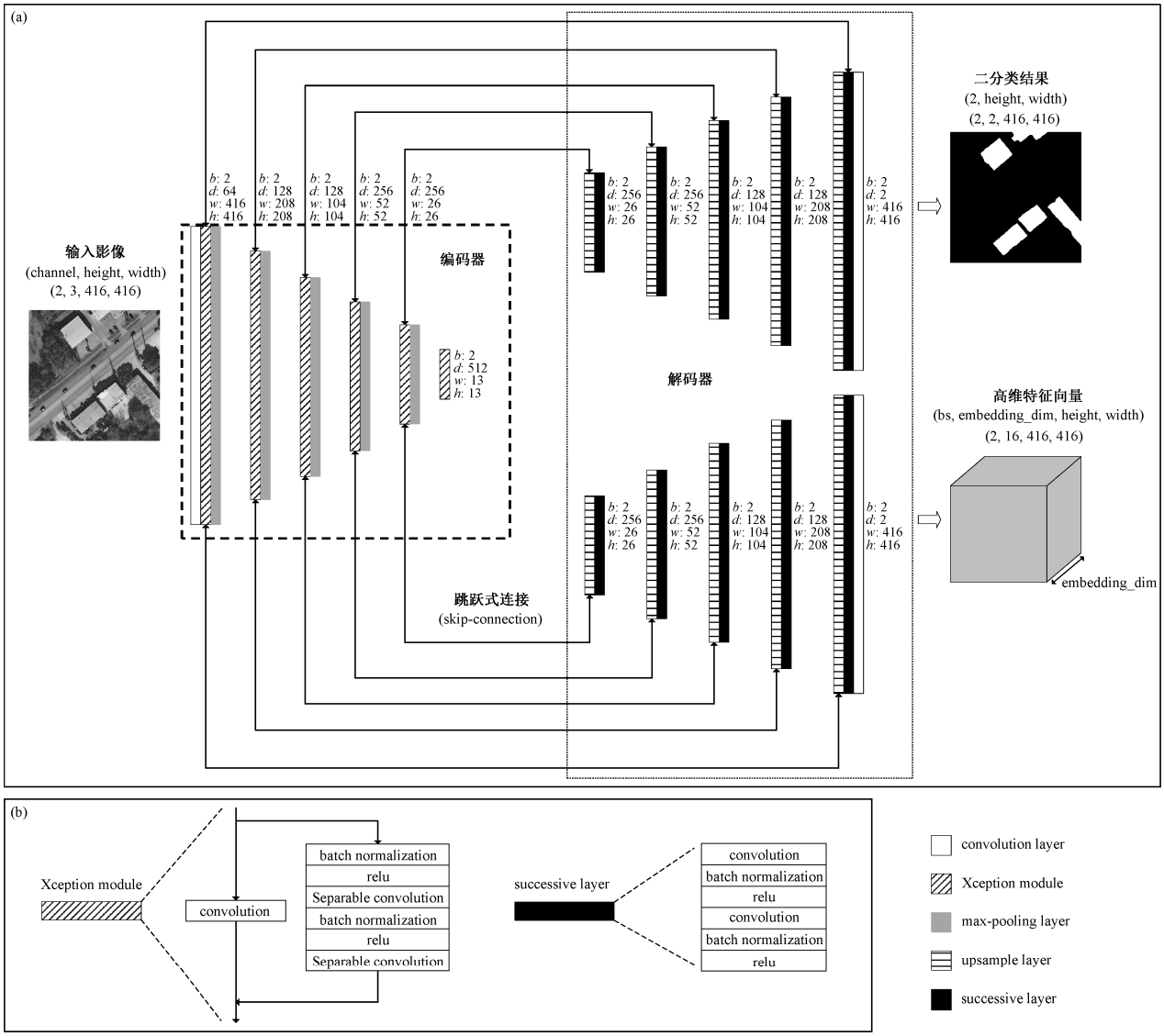
图 3 实例分割及高维特征向量的低维可视化总体流程

Fig. 3 Overall flow chart of instance segmentation and low-dimensional visualization of high-dimensional feature vectors

的分类结果。在编码器中，每次进行池化操作前，不同尺度的特征图都被保存，并通过跳跃式连接(skip-connection)的方式，将保存的特征图传递给对应尺度的解码器，为解码器提供不同尺度的特征信息。通过跳跃式连接的方式，恢复由于池化操作降低分辨率带来的空间信息损失，同时为解码器提供不同分辨率的特征。

Xception 是由 Chollet[21]于 2017 年提出的用于图像分类的深度卷积神经网络模型，在收敛速度和精度方面优于 esNet 等深度神经网络模型。Xception 中使用 Xception module 作为 Inception module 的改进版本，用空间相关性与波段相关性解耦的方式处理图像数据，明显提升分类精度[3,23–25]。Mahdianpari 等[26]验证了 Xception 处理遥感影像数据的有效性。由于遥感影像的多尺度特性[27–28]，同时考虑到内存限制以及建筑物形状较为规则，参照 Nowaczyński[29]的方法，本研究减少 U-Net 对应层的滤波器数目，增加网络的层数，同时，在 U-Net 的编码器部分引入 Xception module 结构，用 Xception module 替换掉卷积层，从而改善 U-Net 的分类精度。

本文网络结构如图 4 所示，bs 为批次数(batch size)。实验中，设 bs 为 2，height 和 width 为 416，embedding_dim 为 16。Max-Pooling 表示最大池化，upsample 表示上采样，此处使用 Conv2dTranspose，batchnorm 为批标准化，relu 为 ReLU 激活函数，separable conv 表示深度可分卷积。图 4 中卷积层(包括分离卷积层)具有相同参数设置，核步长为 1，核大小为 3，滤波器数目分别为(64，64，128，128，256，256，5 12 (编码器)及 256，128，128，64，64，2(二分类解码器)/16 (判别损失函数)；每个尺度的卷积层(或 Xception module)的输出都标明维度，b，d，h 和 w 分别代表特征图的批次数、维度、高度和宽度，两个解码器中所有维度一致，除最终卷积层的输出外，二分类的特征图维度为 2，判别损失函数解码器的特征图维度为 16(即 embedding_dim)。图 4 的网络结构部分包括一个编码器和两个解码器，其中二分类解码器生成大小为(bs，2，height，width)的特征图，通过 softmax 层后，生成建筑物和非建筑两类的概率分布，得到二分类结果，或者用于计算损失交叉熵损失值；判别损失函数解码生成大小为



(a) 网络结构; (b) Xception module 和 successive layer 的结构

图 4　网络结构

Fig. 4　Network architecture

(bs, embedding_dim, height, width)的特征图，即所有像素点对应的高维特征，这些特征向量将用于计算判别损失函数值。

2.2 多任务学习及判别损失函数

多任务学习是一种基于共享表示，把多个相关的任务放在一起学习的机器学习方法。将多任务学习方法应用于深度学习时，多个相关任务并行学习，对应的损失函数同时通过反向传播算法对网络参数进行更新。这样，多个任务能够通过共享特征而相互促进学习，提升精度及泛化效果。本文采用多任务学习方法，建立如下两个相关任务来完成遥感影像建筑物的实例分割：1) 对高分辨率遥感影像进行建筑物的二分类提取；2) 在高维特征空间中，生成遥感影像中建筑物像素点对应的高维特征向量，如图 2 所示。前者需要从网络中提取的特征能够区分建筑物与非建筑物，后者生成对应每个建筑物像素点的高维特征向量。在高维空间中，属于相同建筑物的像素点对应的特征向量在空间中聚集，不同建筑物像素点的特征向量彼此远离。

本文采用判别损失函数[30]训练深度神经网络，使其能够生成对应不同建筑物个体像素点的高维特征向量。当判别损失函数收敛时，使得生成的高维

特征向量能满足: 1) 具有相同标签的建筑物个体对应像素点的高维特征向量在特征空间中应该彼此聚集; 2) 具有不同标签的建筑物个体像素点的聚类中心在高维空间中应该彼此远离。判别损失函数如式(1)和(2)所示。

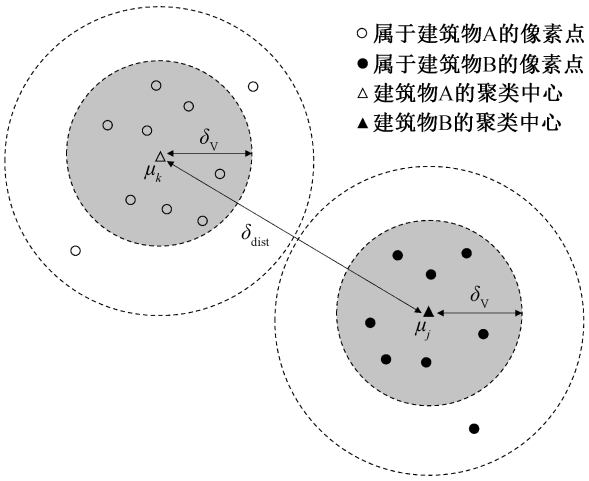
$$\begin{cases} L_{\text{var}} = \dfrac{1}{C}\sum_{m=1}^{C}\dfrac{1}{N_m}\sum_{i=1}^{N_m}\left[\mu_m - x_i - \delta_{\text{v}}\right]_+^2, \\ L_{\text{dist}} = \dfrac{1}{C(C-1)}\sum_{k=1}^{C}\sum_{\substack{j=1\\(k\neq j)}}^{C}\left[2\delta_{\text{d}} - \| \mu_k - \mu_j \|\right]_+^2, \\ L_{\text{reg}} = \dfrac{1}{C}\sum_{k=1}^{C}\| \mu_k \|。 \end{cases} \tag{1}$$

式中，L_{var} 表示相同建筑物个体像素点的聚类程度; L_{dist} 表示不同建筑物个体的聚类中心尽可能远离的程度; L_{reg} 表示对权重的正则化，提高其泛化能力; x_i 表示第 i 个建筑物像素点对应的高维向量; C 表示建筑物个数; N_m 表示不同建筑物对应的像素点个数; μ_m 表示建筑物个体的高维空间聚类中心; δ_{v} 表示方差; μ_k 和 μ_j 分别表示属于不同建筑物的聚类中心; $2\delta_{\text{d}}$ 表示两个建筑物个体的聚类中心最小距离。

$$L_{\text{discrimative}} = \alpha L_{\text{var}} + \beta L_{\text{dist}} + \gamma L_{\text{reg}}, \tag{2}$$

式中，$L_{\text{discrimative}}$ 为判别损失函数, α, β 和 γ 分别对应3个损失项的权重。

图 5 为一个判别损失函数的示意图，用二维平面代替高维空间。假设遥感影像中存在两个建筑物个体，经过多次迭代训练后，在图中可以看到属于相同建筑物个体的像素点聚集在其聚类中心周围，两个建筑物的聚类中心则相互远离。



μ_k和μ_j 表示建筑物个体的高维空间聚类中心，
δ_{v} 表示方差，δ_{dist} 表示聚类中心间的距离

图 5 判别损失函数示意图[31]

Fig. 5 Diagram of discriminative loss function[31]

本文中，基于深度神经网络的多任务学习模型的总损失函数为建筑物二分类的交叉熵损失函数与判别损失函数之和，计算公式如下:

$$L_{\text{total}} = \omega_1 L_{\text{discrimative}} + \omega_2 L_{\text{segmentation}}, \tag{3}$$

式中，ω_1 和 ω_2 为对应的权重。

本文建筑物实例分割方法的基础是对遥感影像进行建筑物与非建筑物二分类，以深度神经网络对高分辨率遥感影像的建筑物二分类预测值为掩膜，滤去所有非建筑物的像素点。将对应这些像素点的高维特征向量通过均值漂移算法(mean shift algorithm)进行聚类，得到的聚类中心数目即是建筑物数目，属于不同聚类中心的像素点对应不同的建筑物。根据判别损失函数的公式(式(1))可知，在理想情况下，当模型收敛时，属于不同建筑物个体的像素点对应的高维特征向量，在高维空间中以 δ_{v} 为方差，分布于聚类中心周围，表现为超球体。在均值漂移算法中，参数带宽(bandwidth)与高维特征向量的方差紧密相关，本文中设置为 $2\delta_{\text{v}}$。

3 实验结果与分析

为了排除其他干扰因素，本文中所有实验均采用相同的优化算法和数据扩增方法。以 p=0.5 对训练数据集进行水平翻转和镜像翻转。同时以均匀的概率采取 90°, 180°和 270°的角度，对训练数据进行旋转。实验中所有模型均采用 Adam 优化算法，其学习率为 0.0004，批尺寸为 2，训练次数为 200，k次迭代。对于多任务学习，判别损失函数中的 α=1, β=1, γ=0.001, δ_{v}=0.25, δ_{d}=18。两个任务的权重 ω_1, ω_2 均设置为 0.5。实验中的程序采用 PyTorch 深度学习框架，配置的显卡为 NVIDIA TITAN Xp, CPU 为 Intel i7-8700K。

3.1 网络模型构建

实验中，在不使用多任务学习的情况下，分别训练 U-Net 和本文提出的基于 Xception 的 U-Net，比较两种网络二分类提取建筑物的表现，结果如表 1 所示。本文提出的基于 Xception 的 U-Net 表现优于 U-Net，精度提升约 1.4%。

图 6 为不同网络结构的建筑物提取结果，显示

表 1　不同网络结构建筑物提取精度

Table 1　Measurements of building extraction by different network architectures

方法	精度/%	F1/%	交并比/%
U-Net	93.73	81.44	66.09
基于 Xception module 的 U-Net	95.14	82.14	66.74

测试样本在不同模型中的表现情况。与原始模型相比，本文改进后的U-Net模型产生更少的错分像素点。图6(d)中，少量建筑物像素点被误判为非建筑物(蓝色)，主要是由树木遮挡和阴影造成的。

3.2 实例分割

在得到建筑物二分类结果的基础上，使用预测结果(或二分类真值)作为掩膜，滤除所有非建筑物，仅保留建筑物像素点对应的高维特征向量，得到 n 个维度为 embedding_dim 的高维特征向量，如图3左下角虚线框内所示。以这个高维向量集合作为数据集，利用均值漂移算法对其进行聚类，得到建筑物像素点的实例标签。本文使用 scikit-learn 中的函数进行均值漂移算法的调用，考虑到函数中的带宽参数与判别损失函数中的 δ_{var} 相对应，聚类时未加入像素点的坐标信息(即图像平面的空间距离)。

图7展示建筑物分布较稀疏时实例分割的效果。图7(d)~(f)中，属于不同建筑物类别的像素点被赋予不同的标签，相同的建筑物像素点具有相同的颜色，黑色为背景颜色。实例分割真值中建筑物数目为8，两次聚类结果中建筑物数目分别为13和5。

从图7可以看出，属于不同建筑物个体的像素点在聚类之后，能够被明确地分开。图7中被红色圆圈标记的两部分，在二分类建筑物提取结果(图8(c))中并不存在，属于样本标记错误(原图(图8(a))中没有对应的建筑物)。在使用二分类样本标记作为掩膜的聚类分析结果(图8(e))中，标记错误区域产生多个聚类点，并且属于多个聚类中心的像素点混杂在一起，也说明建筑物的二分类提取任务与生成高维向量任务是相关的，即在二分类结果中没有建筑物像素点的区域，其像素点(图7中红色圆圈内)对应的高维向量杂散地分布，没有规律可循。因此，原图中建筑物数目实际上为6，使用预测二分类值作为掩膜，聚类后得到5个聚类中心，缺失的建筑物为原图(图7(a))左上角的小片建筑物区域。这是因为在二分类预测时，将该建筑物的像素点误判为非建筑物像素点，导致聚类时输入向量中缺失对应建筑物区域的高维特征向量，使得聚类中心缺失一个，造成建筑物数目预测的误差。

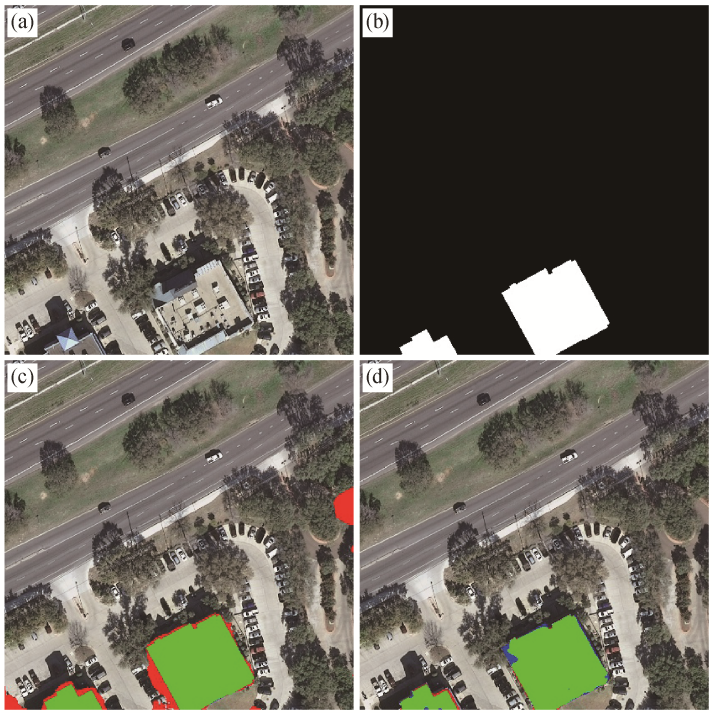
将本方法应用于建筑物较密集区域时，依然可以获得较好的建筑物二值提取和实例分割结果(图8)。实例分割真值中具有的建筑物数目为34，基于预测的二分类结果作为掩膜时，得到聚类中心32个，相差仅2个，即在建筑物较密集分布的情况下，该算法仍然能够有效地划分建筑物区域，提取建筑物个体。

表2给出建筑物实例分割精度，其中|DiC|(使用二分类真值为掩膜)表示以二分类真值(此处为建筑物提取)为掩膜进行聚类后，聚类中心数目与实例中建筑物真实个数之差的绝对值(absolute difference in count)[31]。从表2可以看出，在多任务学习的框架下同时训练两个相关任务，建筑物二分类的精度提高约0.5%，完成建筑物实例的区分。考虑到样本标记存在错误，预计在样本标记错误更少的数据集中本文方法的表现会更好。

3.3 可视化分析

t-SNE算法是Maaten等[32]提出的一种用于数据降维的机器学习算法，可以将高维数据映射到适合观察的2维或3维空间。本文将所有对应建筑物像素点的高维特征向量作为输入数据，使用t-SNE算法，对图7中实例分割产生的高维特征向量进行可视化分析，观察经过判别损失函数训练产生的高维向量的空间聚类情况。

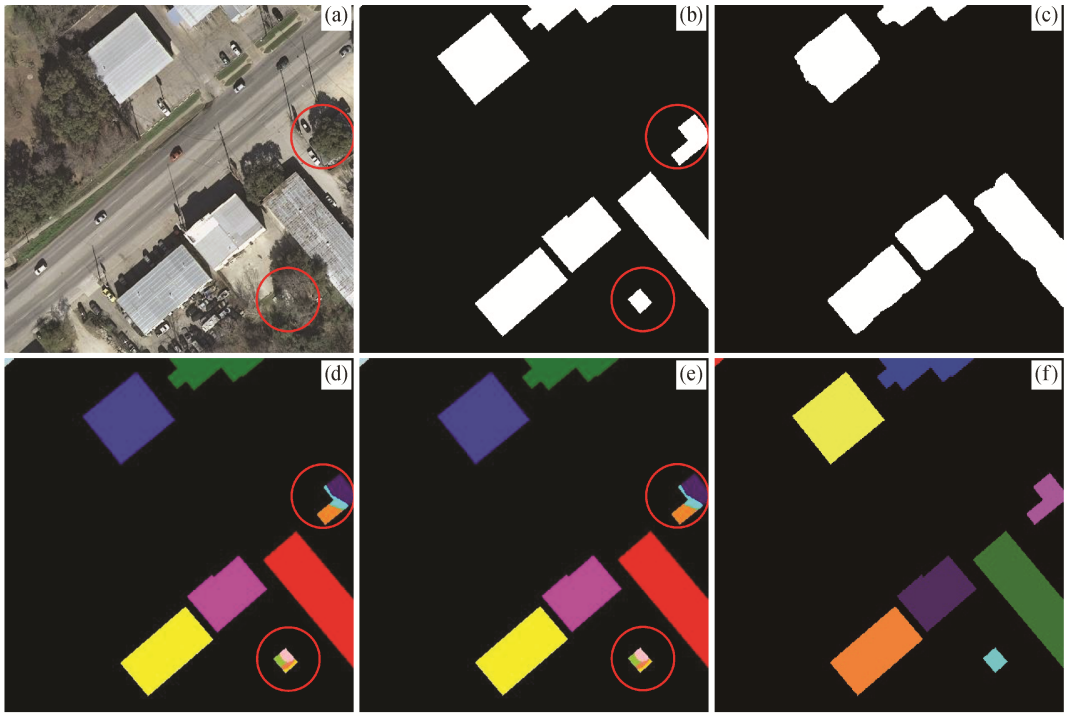
图9展示实例分割结果以及深度神经网络生成的高维特征向量的低维可视化结果。图9(a)表示用于聚类生成实例分割结果的高维特征向量，滤波后仅保留对应建筑物像素点的向量；图9(b)为实例分割结果，其中不同的建筑物个体具有不同的颜色(即不同的标签值)，非建筑物像素点为黑色；图9(c)为t-SNE可视化结果，将图9(a)中属于建筑物像素点的高维特征向量集合作为输入数据，得到其低维可视化图像，同一建筑物实例个体具有相同的颜色。图9(b)中共生成5个建筑物，在图9(c)中主要对应5种颜色，相同建筑物的高维向量聚集在一起，说明通过判别损失函数训练得到的高维特征向量学习到建筑物个体的语义信息。图9(c)中浅蓝色像素点属于同一建筑物个体，由于从高维投影至低维平面，因此形成两个聚集区，但仍然呈现聚集状态。从图9还可以看出，深蓝色像素点对应在真实标签中为非建筑物点而误分类为建筑物的像素点，混杂



(a) 原图; (b) 二分类(真值); (c) U-Net 预测值; (d) 改进后 U-Net 预测值。绿色: 真正, 表示建筑物像素点被正确分类; 红色: 假正, 表示非建筑物像素点被误判为建筑物(错分); 蓝色: 假负, 表示建筑物像素点被误判为非建筑物(漏分)

图 6 不同网络结构建筑物提取结果

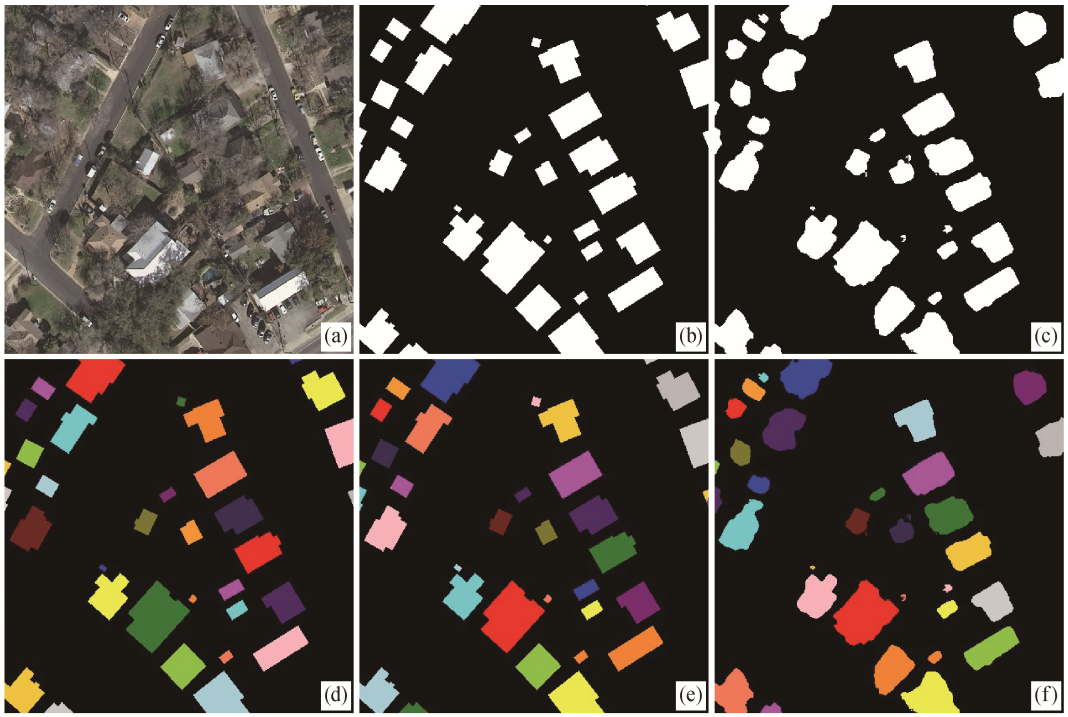
Fig. 6 Building extraction by different network architectures



(a) 原图; (b) 二分类真值; (c) 建筑物提取(二分类预测结果); (d) 实例分割真值; (e) 基于多任务学习二分类预测结果的实例分割结果(使用二分类真值); (f) 基于建筑物真实二分类标签的实例分割结果(使用预测值)

图 7 稀疏分布建筑物实例分割结果

Fig. 7 Instance segmentation of sparse buildings



(a) 原图; (b) 二分类真值; (c) 建筑物提取(二分类预测结果); (d) 实例分割真值; (e) 基于多任务学习二分类预测结果的实例分割结果(使用二分类真值); (f) 基于建筑物真实二分类标签的实例分割结果(使用预测值)

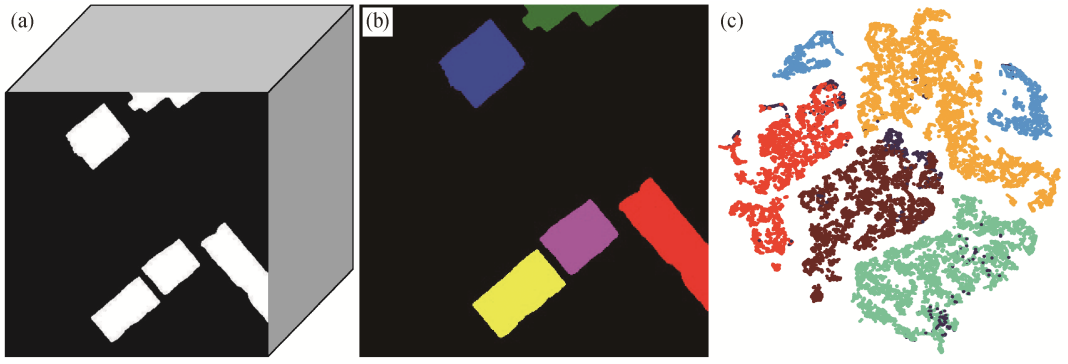
图 8　密集分布建筑物实例提取结果

Fig. 8　Instance segmentation of dense buildings

表 2　建筑物实例分割结果

Table 2　Measurements of building instance segmentation

方法	精度/%	F1/%	交并比/%	|DiC|(使用二分类真值为掩模)	|DiC|
二值分割	95.14	82.14	66.74		
实例分割	95.62	85.41	72.18	5.75	4.8



(a) 高维特征向量; (b) 实例分割(预测值); (c) t-SNE 可视化

图 9　实例分割及高维特征向量的低维可视化示例

Fig. 9　Example of instance segmentation and low-dimensional visualization of high-dimensional feature vectors

于不同簇团中，表明不同建筑物个体周围存在被误分类的非建筑物点，主要原因是二分类提取建筑物掩膜时，对一些建筑物边界产生错误的分类。

4　结论与展望

本文基于 Inria 航空影像，利用 Xception module

对 U-Net 深度神经网络进行改进，同时融合多任务学习算法，提高了基于高分遥感影像的建筑物提取和实例分割精度，得到的主要结论如下。

1) 对于高分遥感影像，使用基于 Xception module 的 U-Net 深度神经网络，建筑物二值提取的精度明显优于原始 U-Net 模型。

2) 通过多任务学习和聚类分析，进一步提升深度神经网络进行建筑物二值提取的精度，并实现基于高分遥感影像的建筑物实例分割。

后续工作中，我们将基于深度神经网络的目标识别算法，进一步提高建筑物识别与分割的精度。

参考文献

[1] 刘莹, 李强. 融合多特征的高分辨率遥感影像震害损毁建筑物检测. 测绘与空间地理信息, 2018, 41(6): 61–64

[2] 赵云涵, 陈刚强, 陈广亮, 等. 耦合多源大数据提取城中村建筑物——以广州市天河区为例. 地理与地理信息科学, 2018, 34(5): 7–13

[3] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, 2015: 1–9

[4] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks // Advances in Neural Information Processing Systems. South Lake Tahoe, 2012: 1097–1105

[5] Mnih V. Machine learning for aerial image labeling [D]. Toronto: University of Toronto, 2013

[6] Alshehhi R, Marpu P R, Woon W L, et al. Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 130: 139–149

[7] Maggiori E, Tarabalka Y, Charpiat G, et al. Convolutional neural networks for large-scale remote-sensing image classification. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(2): 645–657

[8] Huang Z, Cheng G, Wang H, et al. Building extraction from multi-source remote sensing images via deep deconvolution neural networks // Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International. Beijing: IEEE, 2016: 1835–1838

[9] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, 2015: 3431–3440

[10] Wu G, Shao X, Guo Z, et al. Automatic building segmentation of aerial imagery using multi-constraint fully convolutional networks. Remote Sensing, 2018, 10(3): 407

[11] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation // Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, 2015: 234–241

[12] Audebert N, Le Saux B, Lefèvre S. Semantic segmentation of earth observation data using multimodal and multi-scale deep networks // Asian Conference on Computer Vision. Taipei, 2016: 180–196

[13] Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481–2495

[14] Xu Y, Wu L, Xie Z, et al. Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. Remote Sensing, 2018, 10(1): 144

[15] Chen Q, Wang L, Wu Y, et al. Aerial imagery for roof segmentation: a large-scale dataset towards automatic mapping of buildings. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 147: 42–55

[16] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 770–778

[17] Pan X, Gao L, Marinoni A, et al. Semantic labeling of high resolution aerial imagery and Lidar data with fine segmentation network. Remote Sensing, 2018, 10(5): 743

[18] Zhang Z, Luo P, Loy C C, et al. Facial landmark detection by deep multi-task learning // European Conference on Computer Vision. Zurich: Springer, 2014: 94–108

[19] Bischke B, Helber P, Folz J, et al. Multi-task learning for segmentation of building footprints with deep neural networks [EB/OL]. (2017–09–18) [2018–10–26].

https://arxiv.org/abs/1709.05932

[20] Mou L C, Xiang Z X. Vehicle instance segmentation from aerial image and video using a multitask learning residual fully convolutional network. IEEE Transactions on Geoscience and Remote Sensing, 2018: 1–13

[21] Chollet F. Xception: deep learning with depthwise separable convolutions // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 1251–1258

[22] Maggiori E, Tarabalka Y, Charpiat G, et al. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark // IEEE International Symposium on Geoscience and Remote Sensing (IGARSS). Fort Worth, 2017: 3226–3229

[23] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 2818–2826

[24] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning // AAAI Conference on Artificial Intelligence. San Francisco, 2017: 4278– 4284

[25] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks // European Conference on Computer Vision. Amsterdam: Springer, 2016: 630–645

[26] Mahdianpari M, Salehi B, Rezaee M, et al. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. Remote Sensing, 2018, 10(7): 1119

[27] 郑卓, 方芳, 刘袁缘, 等. 高分辨率遥感影像场景的多尺度神经网络分类法. 测绘学报, 2018, 47(5): 620–630

[28] 林雨准, 张保明, 徐俊峰, 等. 多特征多尺度相结合的高分辨率遥感影像建筑物提取. 测绘通报, 2017(12): 53–57

[29] Nowaczyński A. Deep learning for satellite imagery via image segmentation [EB/OL]. (2017–04–12) [2018–10–26]. https://deepsense.ai/deep-learning-for-satellite-imagery-via-image-segmentatio

[30] De Brabandere B, Neven D, Van Gool L. Semantic instance segmentation with a discriminative loss function [EB/OL]. (2017–08–08) [2018–12–26]. https://arxiv.org/abs/1708.02551

[31] Scharr H, Minervini M, French A P, et al. Leaf segmentation in plant phenotyping: a collation study. Machine Vision & Applications, 2016, 27(4): 585–606

[32] Maaten L, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research, 2008, 9: 2579–2605