ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 39,No. 2,Mar. 2024,pp. 348－360
DOI:10. 16337/j. 1004-9037. 2024. 02. 008

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于语义分割和融合残差U-Net的单视光学遥感影像三维重建方法

黄　桦[1]，朱宇昕[2]，章　历[3]，陈志达[4]，张乙志[1]，王　博[2]

(1. 浙江省测绘科学技术研究院,杭州 310012;2. 南京航空航天大学航天学院,南京 211106;3. 浙江艺佳地理信息技术有限公司,杭州 311700;4. 绍兴市上虞区自然资源监测中心,绍兴 312365)

摘　要：从单视遥感图像进行三维重建本身是一个解不唯一的非适定问题,往往需要大量的人工经验来补充缺失信息以构建完整三维模型。为了解决这一问题,提出了一种基于语义分割和融合残差U-Net的单视遥感影像三维重建方法。该方法包括语义分割和单视遥感影像高度估计两个阶段。语义分割阶段使用U-Net确定地物属性,在此基础上改进U-Net对遥感影像进行高度估计,并联合语义特征进行锚定高度回归以提高重建精度。针对改进U-Net,通过嵌入不同数量与通道的残差块,强化编码器的特征提取能力,并修改解码器输出层使其适应于高度回归任务,从而实现逐像素预测遥感影像的数字表面模型(Digital surface model, DSM)高度值。在公开的US3D数据集上得到了均方根误差(Root mean square error,RMSE)为2.751 m、平均绝对误差(Mean absolute error,MAE)为1.446 m的结果,重建结果均优于其余网络,证实该方法实现了基于单视遥感影像的三维估计,能够重建地物的分布结构。

关键词：语义分割;深度残差学习;融合残差U-Net;单视三维重建

中图分类号：TP391　　**文献标志码**:A

Three-Dimensional Reconstruction Method for Single-View Optical Remote Sensing Images Based on Semantic Segmentation and Residual U-Net Fusion

HUANG Hua[1], ZHU Yuxin[2], ZHANG Li[3], CHEN Zhida[4], ZHANG Yizhi[1], WANG Bo[2]

(1. Zhejiang Institute of Surveying and Mapping Science and Technology, Hangzhou 310012, China; 2. College of Aerospace Science, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China; 3. Zhejiang Yijia Geographic Information Technology Co. Ltd., Hangzhou 311700, China; 4. Shaoxing Shangyu District Natural Resources Monitoring Center, Shaoxing 312365, China)

Abstract: Three-dimensional (3D) reconstruction from single-view remote sensing images is an unsolvable problem, which often requires a lot of manual experience to supplement the missing information to construct a complete 3D model. To solve this problem, a 3D reconstruction method of single-view remote sensing image based on semantic segmentation and fusion residual U-Net is proposed. The method includes two stages: Semantic segmentation and height estimation of single-view remote sensing images. In the semantic segmentation stage, U-Net is used to determine the property of ground objects. On this basis, U-Net is improved to estimate the height of remote sensing image. The anchoring height regression

基金项目:浙江省基础公益研究计划(LTGS23D010003)。
收稿日期:2023-10-18;**修订日期**:2024-02-25

is combined with semantic features to improve the reconstruction accuracy. Specifically, in order to improve U-Net, the feature extraction capability of encoder is enhanced by embedding residual blocks with different numbers and channels, and the decoder output layer is modified to adapt to the height regression task, so as to achieve pixel-to-pixel prediction of digital surface model (DSM) height values of remote sensing images. The results of root mean square error (RMSE) of 2.751 m and mean absolute error (MAE) of 1.446 m are obtained on the published US3D data set, and the reconstructed results are superior to those of other networks, confirming that the method can realize 3D estimation based on single-view remote sensing images and can reconstruct the distribution structure of ground objects.

Key words: semantic segmentation; deep residual learning; residual U-Net fusion; single-view 3D reconstruction

引　言

在遥感技术和地理信息科学中,三维重建技术通过从遥感数据中提取三维地貌和地表结构信息,能够为地形分析与研究、城市规划设计、资源管理及交通规划等提供关键的数据支持。在收集大范围地理区域的光学遥感影像时,通常使用推扫成像技术,这种连续成像方式具有高速度、大面积覆盖和时间分辨率较高等优势,但往往只能提供单视图。与多视图重建不同,从单视图重建光学遥感影像可以被描述为单视深度估计问题,这需要一定的单目深度线索,包括纹理变化、遮挡、阴霾、光线和阴影等[1],而单视遥感影像通常只提供地物的二维投影信息,同时伴随着视角限制、缺乏标定信息、光照和阴影等困难。与其他图像相比,遥感图像具有更复杂的光谱特征,不同高度的物体可能由于相似的材料而具有相似的外观,导致从单幅图像可能会产生不匹配的特征。为了克服这些难点,学者们使用各种计算机视觉、图像处理和深度学习技术,以提高单视图遥感影像的三维重建效果。

传统的单视图深度估计算法通常依赖于图像内部的纹理几何信息,以推断相机的位置或局部距离信息。王光辉[2]提出了通过单幅图像进行平面测量的3种新方法,包括单应矩阵、欧式度量信息与长度变换关系。杨敏等[3]充分利用了人造结构场景中大量存在的平行性和正交性几何约束进行单视角三维重构。丁伟利等[4]则利用空间三方向正交的平行线束提取三维信息。这些方法使得通过单帧图像推断深度值成为可能,但受到严格的应用场景限制,且一次仅能计算图像中局部区域或像素点对间的相对距离,预测精度较低。其他如李健等[5-6]常用纹理和阴影针对单幅图像进行三维重建,该方法虽然具有更好的泛化性,但对光照和灰度要求较高,难以高质量重建真实图像。

此外,由于需要使用全局概率图模型才能对所选局部特征加以优化,因此早期的统计学习方法建立在人工选择、标注的特征和统计模型的基础上。Saxena等[7]使用了一个经过区别训练的马尔可夫随机场,结合了多尺度局部和全局图像特征,并将各个点的深度建模为不同点处的深度之间的关系。Liu等[8]将单目深度估计公式化为离散连续优化问题。Karsch等[9]提出的DepthTransfer方法利用全局特性,创新性地把有标签数据与待预测样本进行点到点的对应,迁移深度信息形成深度估计结果。早期基于图模型和分类器等统计模型的算法精度低、特征表示能力不足、缺乏上下文理解,很快被基于深度学习网络的深度估计算法取代。

基于深度学习网络的单幅图像深度信息估计包括卷积神经网络(Convolutional neural network, CNN)和生成对抗网络(Generative adversarial network, GAN)。Mou等[10]提出了一种全卷积-反卷积网络,以端到端方式学习单幅图像的高度信息。Zhang等[11]提出了一种多路径融合网络,引入多路径特征融合模块组合通过高效递归细化网络提取的多尺度特征,有效地利用不同抽象级别的信息。

Amirkolaee等[12]提出了一种深度卷积编码器-解码器网络,通过结合全局和局部特征来恢复物体的精确几何形状。Li等[13]提出将高度值划分为间距递增区间,并将回归问题转化为有序回归问题,使用有序损失进行网络训练。Liu等[14]提出了一种基于CNN的方法实现单幅图像的高度估计,并利用学习到的高度图来推断三维建筑的形状和二维建筑的足迹。Xing等[15]提出了一种渐进式学习网络,利用注意机制聚合高低特征,并通过渐进式细化模块逐步细化预测的高度图。随着生成对抗网络在计算机视觉各个领域的广泛应用,最近的一些研究将生成对抗网络用于图像到图像的翻译,从图像中生成相应的高度图。Ghamisi等[16]提出使用条件生成对抗性网络,其架构基于具有跳过连接和图像补丁规模的编码器-解码器网络。Paoletti等[17]基于变分自动编码器和生成对抗性网络来实现高程的预测。

综上,传统单视图深度估计方法对图像纹理、阴影等信息需求较高,早期统计模型的算法精度低、受特征表示限制、缺乏全局信息整合。对于单视图像的高度估计,需要考虑上下文信息,包括遮挡、插值、纹理梯度和纹理变化等。与计算机视觉领域用于深度估计的图像相比,遥感图像通常是正射投影的,这导致可获得的上下文信息有限。此外,遥感图像中有限的空间分辨率、相对较大的覆盖面积和微小的地物也给高度估计带来一定困难。因此,本文提出了一种基于语义分割和融合残差U-Net的单视光学遥感影像三维重建方法,结合地物属性和高度估计值实现地物场景的三维重建。

1 本文方法

本文方法流程图如图1所示,输入数据为24位RGB原始遥感影像,由于零分量分析(Zero-phase component analysis, ZCA)变换有利于增强高频成分,提升网络性能,首先对输入数据进行ZCA处理。为了减小数字表面模型(Digital surface model, DSM)缺少地物属性导致的高度估计误差,使用U-Net对遥感影像进行语义分割,帮助后续高度估计模型提高对地理区域的理解。在高度估计任务中,设计Res-Unet,强化网络特征提取能力,并结合语义分割结果进行锚定高度回归,实现遥感图像三维模型重建,输出影像对应区域的数字表面模型即DSM。本文中,Res-Unet使用U-Net作为基础框架,因此不对U-Net做额外说明。

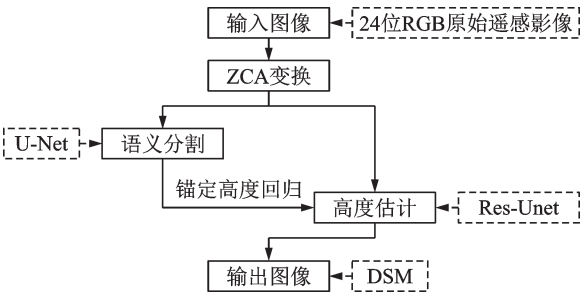


图1 本文方法流程图

Fig.1 Flowchart of the proposed methodologies

1.1 ZCA变换

物体边界在深度预测中起着重要作用。因此,使用带有数据归一化的边界增强或边缘锐化技术可以改善估计的深度值。为了在有噪声的数据中保持边缘并最小化沿边缘的视觉误差,本文使用ZCA方法。ZCA变换可以减小数据的相关性来减弱噪声的影响,同时保留数据的重要特征,提高数据的可解释性和分类性能。增强高频成分(如边缘)可以在应用卷积层的同时提取各种鲁棒特征。协方差矩阵计算为

$$\sum \boldsymbol{y}=\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i\cdot\boldsymbol{x}_i^{\mathrm{T}} \tag{1}$$

式中:$\boldsymbol{x}_i$为向量化的第i个归一化图像块;$\sum \boldsymbol{y}$为协方差矩阵;N为图像块的个数。从而进一步计算协方差矩阵的特征值和特征向量。ZCA白化变换为

$$\boldsymbol{x}_{\mathrm{ZCA}}=\boldsymbol{U}\begin{pmatrix}\frac{1}{\sqrt{\lambda_1+\varepsilon}} & 0 & \cdots & 0\\ 0 & \frac{1}{\sqrt{\lambda_2+\varepsilon}} & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \frac{1}{\sqrt{\lambda_n+\varepsilon}}\end{pmatrix}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x} \tag{2}$$

式中:λ为特征值;ε的作用是防止在分母上产生零值;$\boldsymbol{U}$为特征向量。

1.2　基于锚点的高度回归

不同地物之间的高度方差太大,用卷积神经网络进行回归效果并不理想。然而,对于同一类地物,其高度分布方差较为稳定。因此,引入锚定机制来获得更为稳定的高度回归。

在一张图像中,c类目标的高度H_c可以表示为

$$H_c=\delta_c\cdot s+\mu_c \tag{3}$$

式中:δ_c和μ_c分别表示c类地物高度的标准差和均值;s表示解码器预测的比例因子。本文在训练锚定高度回归网络时,分别计算训练集每张图像中各类地物的标准差和均值。

s的回归目标t可以通过式(4)来计算,其中目标类别由U-Net对输入图像进行语义分割得到。

$$t=\frac{H_c-\mu_c}{\delta_c} \tag{4}$$

1.3　基于深度残差网络的单视三维估计模型

为了获得更精细的结果,需要网络在保留高级别语义信息的同时利用低层次的细节。然而有限的训练样本难以训练出这样的网络,这时可以采用预训练的网络,并在目标数据集上微调[18];或者如U-Net[19]一样,使用广泛的数据增强。U-Net作为早期的一种通用型神经网络编码器部分较为简单,可能会丢失一些复杂的特征信息,但其架构有助于缓解训练问题。U-Net的跳跃连接结构为信息传播创造了一条路径,使得低级特征可以更容易地复制到相应的高级特征上。而残差神经网络(Residual network,ResNet)[20]可以通过堆叠多个残差块,从多维特征图中提取更丰富的、更高层次的特征表示,提供更强的感受野,帮助提高任务的性能。

本文使用的Res-Unet网络结构如图2所示,这种结构设计能够充分地挖掘可见光数据特征,同时利用ResNet的深度特征提取能力以及UNet的分辨率保留能力,在一定程度上提高模型对遥感图像高度的预测性能,从而更好地将单视光学遥感影像与DSM真值耦合。本文将证明通过使用残差结构代替普通单元,可以进一步提高网络性能。

1.3.1　卷积编码器

增加神经网络的深度可以增强网络的表示能力和特征抽取能力,提高网络性能,但也可能会出现网络退化问题。为了克服这一问题,He等[20]提出ResNet,在解决了网络退化问题的同时,残差单元通过局部特征和全局特征的融合来改善网络长距离信息依赖问题。因此,在编码步骤采用移除全连接层的ResNet34架构(本文采用了ImageNet预训练的权重)。

如图2所示,网络的输入为512像素×512像素的RGB图像。单视遥感影像经过卷积和池化层后,通过4组残差模块,每个模块中包含3个64通道、4个128通道、6个256通道、3个512通道的残差块。对于残差块中的归一化层、激活层以及卷积层,He等[21]详细讨论了不同组合的影响,在本文中采用如图3所示的完整预活化设计。残差块中包含2个相同输出通道的3×3卷积层,随后紧跟归一化与激活层,

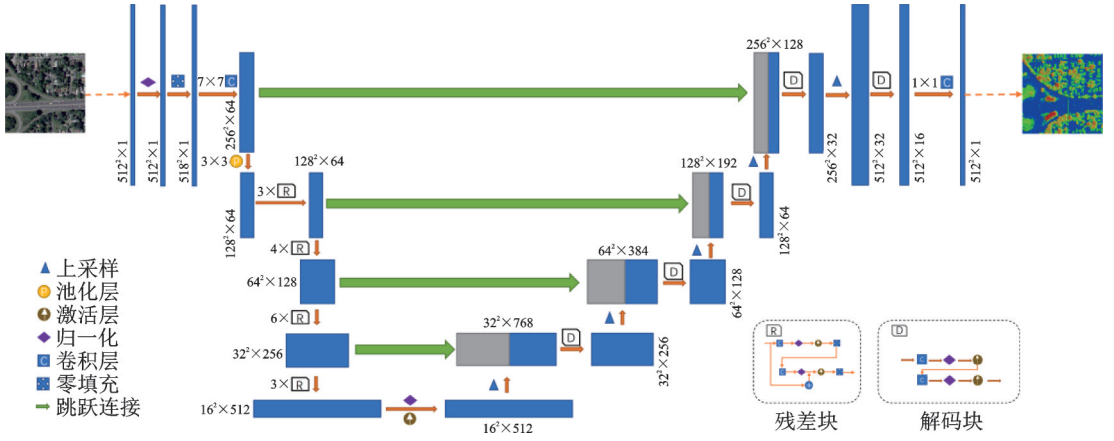


图 2 Res-Unet网络结构

Fig.2 Res-Unet network structure

输入跳过残差块中的卷积运算，直接与激活层输入相加，为保证图像尺寸匹配，在第2次卷积前增加零填充层。这4组残差模块可以在保证网络能够提取充足图像特征的前提下，有效地避免网络梯度消失和网络退化问题。

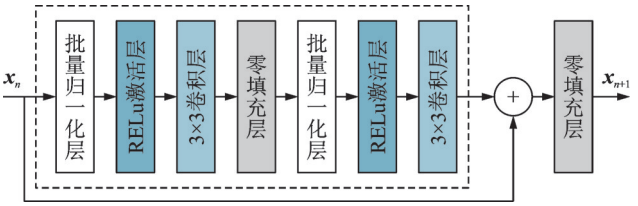


图 3 残差块结构

Fig.3 Residual block structure

1.3.2 卷积解码器

解码器包括4个阶段，每个阶段的反卷积层都使用一个上采样因子为2的卷积核，逐步增加特征图的分辨率以恢复更丰富的细节信息。在上采样步骤中，解码器会通过跳跃连接，将当前层的特征图与相应编码层级特征图进行融合，以捕获不同尺度的特征信息，这可以使解码器有效地传递编码器中的信息，以提高特征重建的质量，并且这个过程完全卷积，使得它可以处理任意大小的图像，具有灵活性和适用性。在最后的输出层使用1通道的3×3卷积核并添加全连接层，将特征图映射为最终的高度估计图，且输出图的尺寸将与原始单视遥感影像相匹配。

基于上述网络结构，给出本文方法整体网络结构图，如图4所示，本文主要采用U-Net语义分割网络与Res-Unet锚定高度回归网络，二者具有相同的解码器结构，但本文所使用的Res-Unet则在编码器部分结合残差结构，提高网络局部特征和全局特征的表达能力，改善网络长距离信息依赖问题。

1.4 损失函数设计

数据集中地面和水体的分布相对均匀，特征和边界更容易被学习，且与其他类别相互独立，受其他类别地物干扰较少，更容易被区分，而道路和桥梁在图像数据中可能会出现前景（道路和桥梁）-背景（其他地物）像素不平衡的问题，导致网络更容易识别前景，但难以对背景进行正确分类。因此，引入RetinaNet[22]中的α-平衡型焦点损失来增强网络识别不平衡类别的性能，其表达式为

$$\mathrm{FL}(p,y)=-\alpha y(1-p)^{\gamma}\ln p-(1-\alpha)(1-y)p^{\gamma}\ln(1-p) \tag{5}$$

式中：p表示模型预测的类别概率；y表示实际类别标签；α为平衡因子，用于调整类别的重要性；γ用于调整焦点损失的聚焦度。本文取$\alpha=0.25$，$\gamma=0.2$。

对于高度估计任务，本文采用剔除异常值的均方误差（Mean squared error，MSE）作为损失函数，以

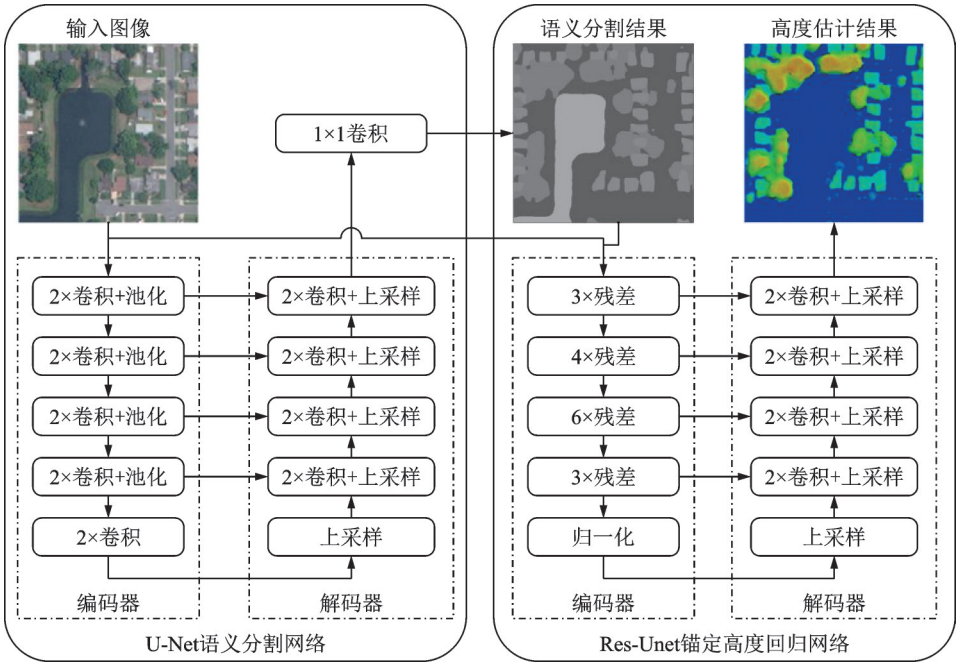


图4　整体网络结构图

Fig.4　Diagram of the overall network structure

减少异常高度值导致的误差。具体而言,创建一个掩码mask_true用于将有效的高度值标记为1,并将掩码与均方误差相乘以去除异常值,其表达式为

$$\mathrm{NO\_NAN\_MSE}=\frac{\sum_{i=1}^{n}\mathrm{mask\_true}\cdot(y_i-\hat{y}_i)^2}{\max\left(\sum_{i=1}^{n}\mathrm{mask\_true},1\right)} \tag{6}$$

式中:y_i为真值;$\hat{y}_i$为预测值;n为图像的像素点数目。

2　实验验证

2.1　数据来源与特点

2.1.1　数据来源

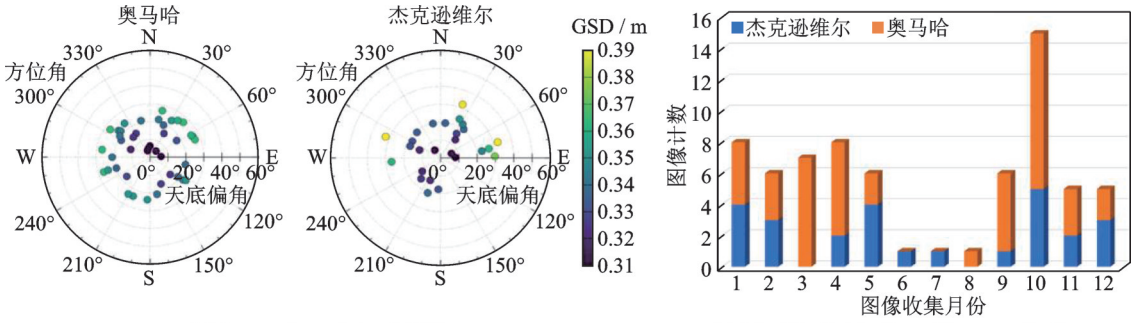
采用Worldview-3提供的Urban Semantic 3D(US3D)数据集,这是一个用于城市场景理解和分析的大规模公共数据集,覆盖约100 km^{2}的佛罗里达州杰克逊维尔和内布拉斯加州奥马哈两大城市。数据集包括这两个城市的RGB卫星图像、地面真实值以及语义分割标签,提供高质量的语义标注和真实场景,为帮助开发和评估城市场景理解算法、建筑信息建模等领域提供支持。

2.1.2　数据特点

多日期卫星图像:WorldView-3的RGB全色图像源数据包括2014—2016年在佛罗里达州杰克逊维尔收集的26幅图像,以及2014—2015年在内布拉斯加州奥马哈收集的43幅图像。地面采样距离(Ground sample distance, GSD)约为35 cm。

航空激光雷达数据:由激光雷达导出的数据提供地面真实几何值作为训练数据,脉冲间距(Aggregate nominal pulse spacing, ANPS)约为80 cm。

语义标签:语义类别包括建筑物、高架道路和桥梁、植被、地面和水等。图5统计了收集数据时图像视点、地面采样距离以及季节分布情况。



(a) Image viewpoint and GSD (b) Monthly distribution of image collection

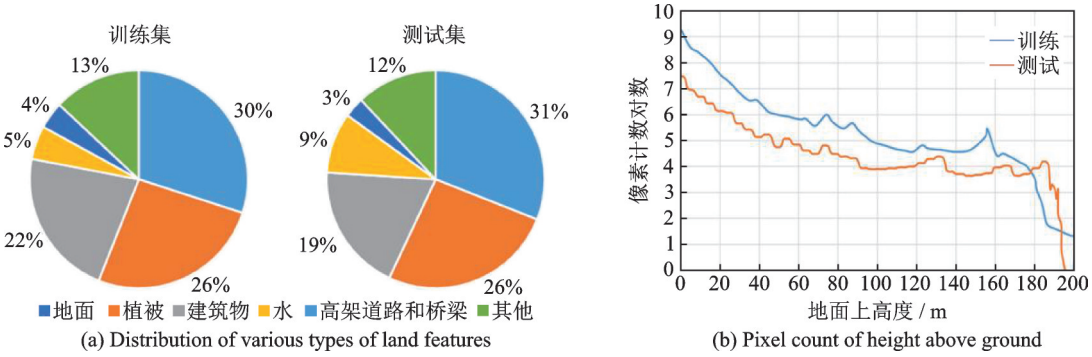
图5 图像采集情况

Fig.5 Image acquisition situation

2.2 数据增强与数据选择

受设备限制,本文将原始数据裁剪为512像素,同时由于数据的获取难度较大,采用以下数据增强操作,以提高网络的收敛速度:(1)概率为0.5的随机水平或垂直翻转;(2)概率为0.5的随机旋转90°;(3)概率为0.5的随机图像转置。

在机器学习模型训练和测试时,需要使用独立的数据集进行评估。因此,采用分层抽样的方式尽可能保持数据分布的一致性,减少样本选择的偏差,提高训练的可靠性和有效性。最终划分训练集和测试集的图像数据统计如图6所示。同时,对地面上高度数据进行逐像素计数,验证训练集与测试集在高度值分布上也具有一致性。



(a) Distribution of various types of land features (b) Pixel count of height above ground

图6 训练集与测试集统计

Fig.6 Statistics of training and testing datasets

2.3 实验设计与评价指标

基于上述划分好的数据集,神经网络的训练在具有12 GB内存的NVIDIA Tesla K80上进行,训练集1 780对图片,测试集400对图片。语义分割网络U-Net训练集由RGB遥感影像及其对应语义分割标签构成,本文高度估计网络Res-Unet训练集由RGB遥感影像及其对应航空激光雷达数据构成。二者分别单独训练,均采用Adam优化器进行参数更新,超参数选用该算法默认值,即$\beta_1=0.9$,$\beta_2=0.999$,训练批次大小设置为4,学习率为0.000 1,共训练400轮(Epoch)。

评价指标选取均方根误差(Root mean square error,RMSE)和平均绝对误差(Mean absolute error,MAE),用于度量模型预测值与实际值之间的误差,其表达式分别为

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{7}$$

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right| \tag{8}$$

2.4　实验结果与分析

2.4.1　本文实验

如图7所示,图像中的高频信息包括不同区域之间明显的界限(建筑物轮廓、道路边缘)以及小尺度结构(车辆、树木)等。图像在经过ZCA变换后,图像中的高频信息得到增强,有助于语义分割模型更容易地检测和区分不同的对象。特别是图像经过ZCA变换后,预测的高度值也更加接近真值。

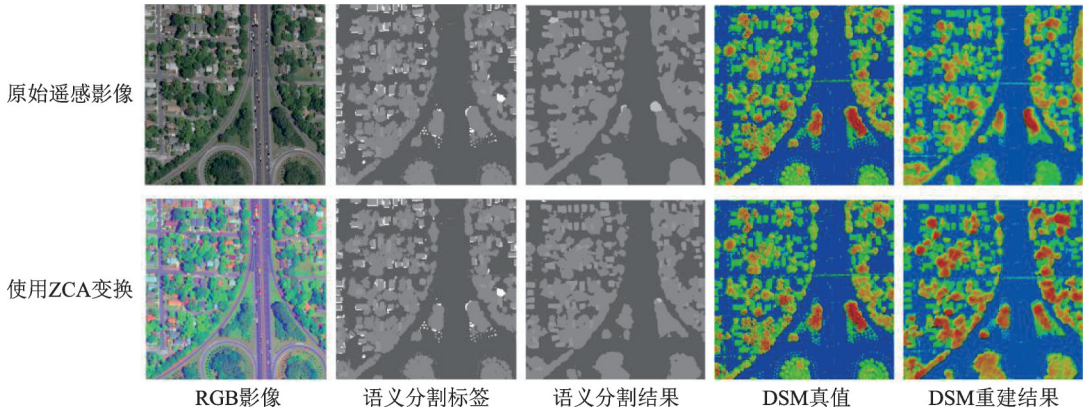


图7　ZCA变换结果对比

Fig.7　Comparison of ZCA transformation results

表1为上述实验定量分析结果,其中mIoU为平均交并比,mIoU3为带有阈值的平均交并比(仅统计超过阈值部分的交并比,此处阈值设置为0.3)。可以看出,使用ZCA变换后语义分割效果有所提升,且RMSE和MAE分别降低12.4%和12.7%,进而验证了ZCA变换对于语义分割和高度估计任务的有效性。

表1　ZCA变换对比

Table 1　Comparision of ZCA transformation

条件	地物类别	mIoU	mIoU3	RMSE/m	MAE/m
不使用ZCA变换	地面	0.85	0.84	3.143	1.657
	植被	0.62	0.23		
	建筑物	0.72	0.41		
	水域	0.72	0.39		
	高架桥	0.73	0.18		
使用ZCA变换	地面	0.87	0.83	2.751	1.446
	植被	0.65	0.31		
	建筑物	0.78	0.45		
	水域	0.68	0.52		
	高架桥	0.76	0.53		

Res-Unet在US3D测试集上的DSM重建结果如图8所示，图中分别选取水域、密集建筑物、植被以及高架桥4块区域。地物高度值用热力图表示，热力图中蓝色区域表示高度值较小，而红色区域表示高度值较大。从图8可以看出，Res-Unet结构充分发挥特征融合的优势，DSM重建结果的高度预测值与真值范围较接近。但由于树木结构的多样性以及遮挡效应，影响网络对植被实际高度进行估计，同时对比于不采用语义分割的直接高度回归，本文的锚定高度回归方法与DSM真值误差更小。

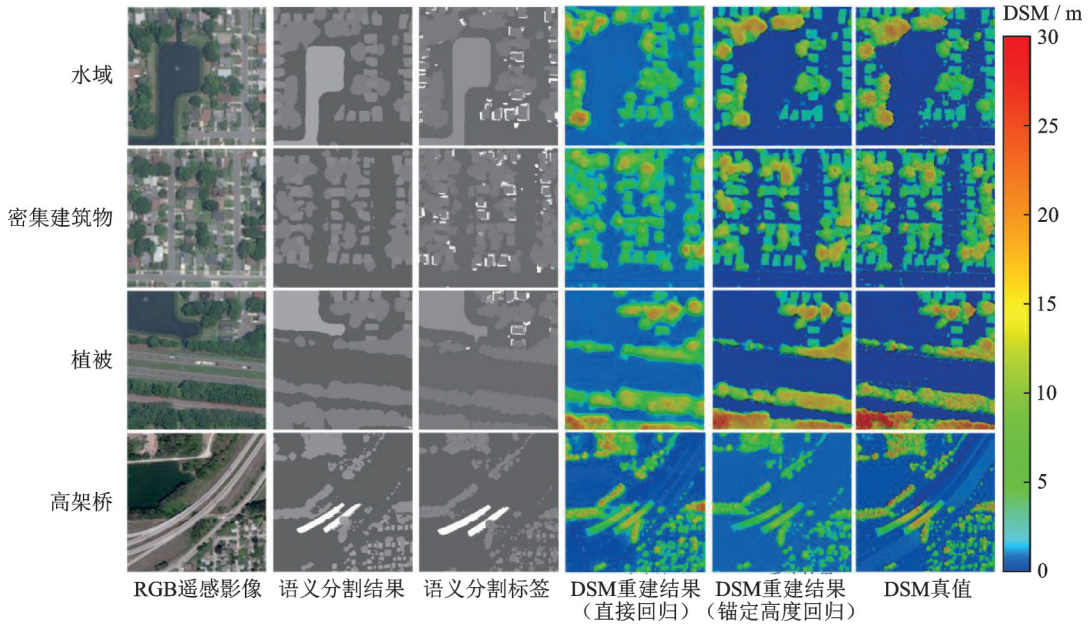


图8 DSM真值及重建结果

Fig.8 Ground truth and reconstruction results

2.4.2 消融实验

本文还采用类似Res-Unet嵌入结构的其他网络[23]（使用嵌入结构-Baseline的命名规则）进行消融实验。如图9所示，选取VGG16-Unet、Res-FPN、Res-PSPNet与本文网络进行对比实验。由测试结果可知，Res-Unet的三维重建效果优于其余网络。

（1）VGG16-Unet由于没有残差融合结构，在编码阶段提取语义特征的能力相较于Res-Unet较弱，尤其是对于细节信息的捕获，无法有效地保留原始图像的细节特征，出现地物结构缺失的现象，且无法捕捉足够的上下文信息，网络架构不足以处理复杂的高度值，导致重建后各类地物高度值均与真值差距较大。

（2）Res-FPN虽然也对多尺度特征图进行融合，但在解码部分仅使用插值来恢复高层特征图的分辨率，限制了其在细节信息方面的表现，而在相同任务上，本文网络使用卷积，通过网络自学习的方式进行上采样，这允许网络更好地恢复细节信息，从而提高网络性能。

（3）Res-PSPNet通过金字塔池化层来融合多尺度信息，没有建立编码器与解码器之间的连接，这使得它对于细节部分不够敏感。

本文网络采用的残差融合结构增强了网络的表达能力并减轻梯度消失问题，同时跳跃连接在不同层级上捕获多尺度信息，减少信息在编码器和解码器之间的丢失，保留原始图像的细节特征和结构信

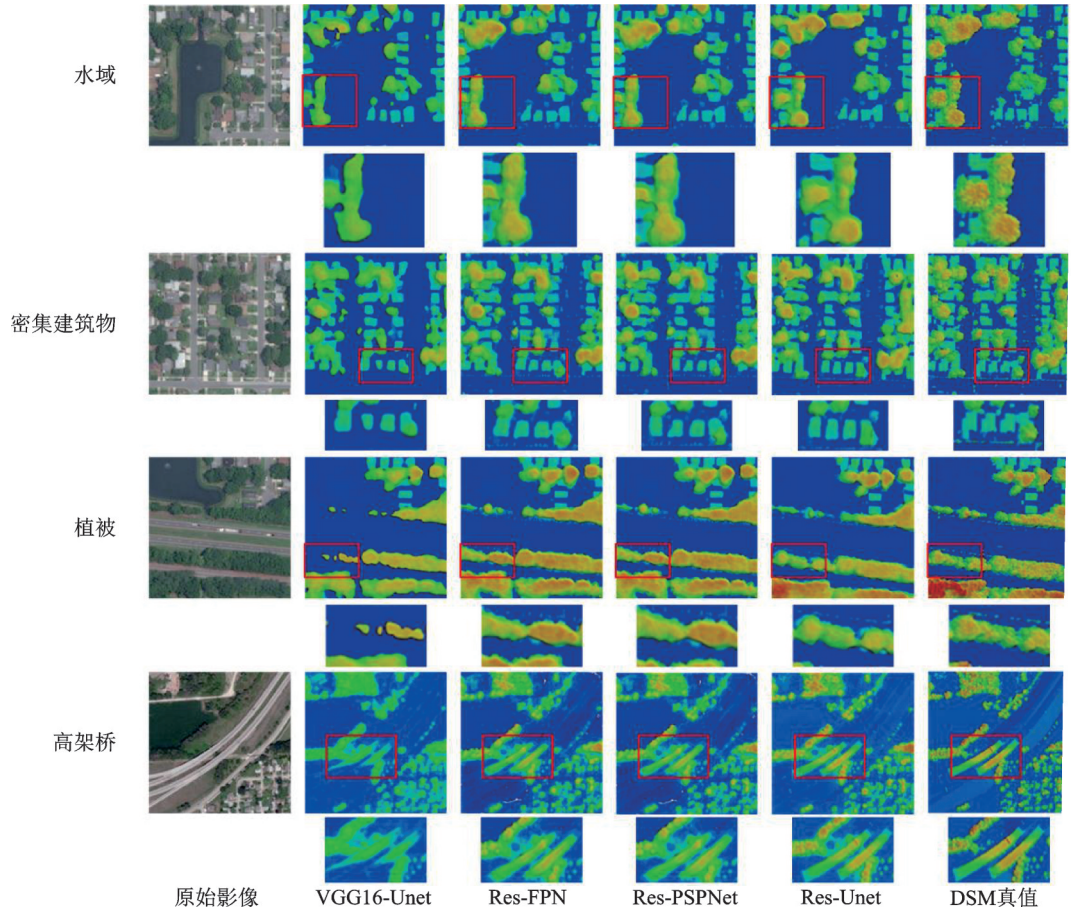


图9　消融实验结果

Fig.9　Ablation experiment results

息,从而提高网络重建精度。

表2、3中消融实验结果表明,本文网络结构能够有效降低结果与真值之间的误差,在US3D测试集上取得了RMSE为2.751 m、MAE为1.446 m的实验结果,与VGG16-Unet、Res-FPN、Res-PSPNet网络相比,RMSE平均降低38.4%,MAE平均降低36.5%,证明该网络能够结合DSM实现遥感影像的单视三维估计,具有地物分布结构重建能力。

表2　消融实验结果(语义分割)

Table 2　Ablation experiment results (semantic segmentation)

类别	mIoU	mIoU3
地面	0.87	0.83
植被	0.65	0.31
建筑物	0.78	0.45
水域	0.68	0.52
高架桥	0.76	0.53

表3　消融实验结果(高度估计)

Table 3　Ablation experiment results (height estimation)

对比网络		评价指标	
Baseline	嵌入结构	RMSE/m	MAE/m
UNet	VGG16	5.256	2.582
FPN	ResNet34	4.777	2.366
PSPNet	ResNet34	3.685	1.973
UNet	ResNet34	2.751	1.446

2.4.3 对比实验

Mou等[10]提出了一个完整的卷积-反卷积网络架构IM2HEIGHT,通过端到端训练来模拟单视遥感图像和高度图之间的模糊映射。其中卷积子网络将输入的遥感图像转换为高级多维特征表示,反卷积子网络将从卷积子网络中提取的特征生成高度图。此外,引入跳跃链接保留高度图精细边缘细节。Chen等[24]用多尺度结构进行高度估计。在残差金字塔解码器的上层表达全局场景结构,在下层表达局部结构。同时在每一层使用残差细化模块来预测残差映射,以逐步添加更精细的结构。此外,引入自适应密集特征融合模块充分融合利用多尺度图像的特征。本文也与上述两种方法进行了对比实验,实验结果如图10所示。本文方法通过锚定高度回归的思想,减少地物属性与高度不匹配的情况,更好地保留了地物分布结构。同时,如表4所示,本文方法在更少的参数与更快的速度下,实现了地物高度高精度估计。

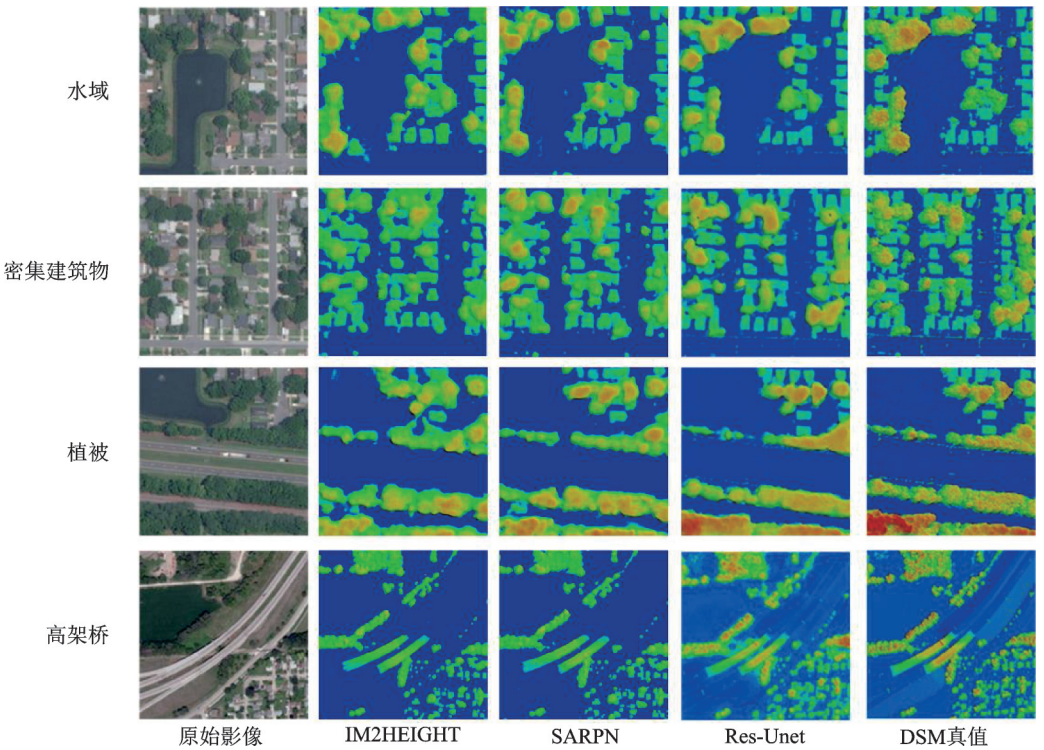


图10 对比实验结果

Fig.10 Comparative experiment results

表4 对比实验量化结果

Table 4 Quantitative results of comparative experiment

网络	RMSE/m	MAE/m	网络参数量/10^6	每秒浮点运算次数/10^9	测试时间/s
IM2HEIGHT	3.491	1.738	7.360	125.487	18.653
SARPN	3.045	1.546	38.651	75.308	15.922
Res-Unet	2.751	1.446	2.269	68.364	16.720

3 结束语

单视遥感影像无法直接提供CNN进行三维重建时所需的高度信息,因此提出了一种基于语义分

割和融合残差U-Net的单视光学遥感影像三维重建方法。该方法在使用U-Net对遥感影像进行语义分割的基础上,以U-Net为基本框架进行改进,使其用于影像的高度估计。在编码阶段,在原有结构上嵌入残差结构,对经过ZCA变换后的遥感影像进行特征提取;在解码器中使用跳跃连接以恢复细节信息。最后,联合语义分割结果进行锚定高度回归,以获得遥感影像的三维重建结果。本文方法使用DSM数据与单视光学遥感影像实现遥感影像的三维估计,因而有效解决了单视遥感影像无法为CNN提供充分信息的问题。在US3D数据集上进行测试,得到重建结果与真值的RMSE为2.751,MAE为1.446。同时,进行VGG16-Unet、Res-FPN和Res-PSPNet的消融实验,以及与IM2HEIGHT和SARPN的对比实验,本文网络实验结果均优于其他网络,证明该方法能够基于语义分割和残差U-Net实现单视光学遥感影像的三维估计。

参考文献:

[1] EIGEN D, FERGUS R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture[C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2015: 2650-2658.

[2] 王光辉. 基于图像的量测及三维重建研究[D]. 北京: 中国科学院自动化研究所, 2004.
WANG Guanghui. Research on image-based measurement and 3D reconstruction[D]. Beijing: Institute of Automation, Chinese Academy of Sciences, 2004.

[3] 杨敏, 沈春林. 未标定单幅结构场景图像的三维重构[J]. 中国图象图形学报: A 辑, 2004, 9(4): 456-459.
YANG Min, SHEN Chunlin. Three dimensional reconstruction of uncalibrated single structured scene images[J]. Chinese Journal of Image and Graphics: Volume A, 2004, 9 (4): 456-459.

[4] 丁伟利, 朱枫, 郝颖明. 基于单幅建筑物图像的三维信息提取[J]. 仪器仪表学报, 2008, 29(9): 1965-1971.
DING Weili, ZHU Feng, HAO Yingming. 3D information extraction based on a single building image[J]. Journal of Instrumentation, 2008, 29 (9): 1965-1971.

[5] 李健,杨苏,刘富强,等. RGBD融合明暗恢复形状的全视角三维重建技术研究[J].数据采集与处理, 2020, 35(1): 53-64.
LI Jian, YANG Su, LIU Fuqiang, et al. Full view 3D reconstruction by fusing RGBD and shape from shading[J]. Journal of Data Acquisition and Processing, 2020, 35(1): 53-64.

[6] 李健,李丰,何斌,等.单Kinect+回转台的全视角三维重建[J].数据采集与处理,2019,34(2): 205-213.
LI Jian, LI Feng, HE Bin, et al. Single Kinect and rotating platform for full-view 3D reconstruction[J]. Journal of Data Acquisition and Processing, 2019, 34(2): 205-213.

[7] SAXENA A, CHUNG S H, NG A Y. Learning depth from single monocular images[C]//Proceedings of the 18th International Conference on Neural Information Processing Systems. [S.l.]: NIPS, 2005: 1-8.

[8] LIU M, SALZMANN M, HE X. Discrete-continuous depth estimation from a single image[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 716-723.

[9] KARSCH K, LIU C, KANG S B. Depth transfer: Depth extraction from video using non-parametric sampling[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(11): 2144-2158.

[10] MOU L, ZHU X X. IM2HEIGHT: Height estimation from single monocular imagery via fully residual convolutional-deconvolutional network[EB/OL]. (2018-02-28)[2023-09-30]. https://arxiv.org/abs/1802.10249v1.

[11] ZHANG Y, CHEN X. Multi-path fusion network for high-resolution height estimation from a single orthophoto[C]//Proceedings of the 2019 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2017. Shanghai, China: IEEE, 2019: 186-191.

[12] AMIRKOLAEE H A, AREFI H. Height estimation from single aerial images using a deep convolutional encoder-decoder network[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 149: 50-66.

[13] LI X, WANG M, FANG Y. Height estimation from single aerial images using a deep ordinal regression network[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 19: 1-5.

[14] LIU C J, KRYLOV V A, KANE P, et al. Im2elevation: Building height estimation from single-view aerial imagery[J].

Remote Sense, 2020, 12: 2719.

[15] XING S, DONG Q, HU Z. Gated feature aggregation for height estimation from single aerial images[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 1-5.

[16] GHAMISI P, YOKOYA N. IMG2DSM: Height simulation from single imagery using conditional generative adversarial net[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15: 794-798.

[17] PAOLETTI M, HAUT J, GHAMISI P, et al. U-IMG2DSM: Unpaired simulation of digital surface models with generative adversarial networks[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 18: 1288-1292.

[18] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2015: 3431-3440.

[19] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// Proceedings of Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Munich, Germany: Springer, 2015: 234-241.

[20] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.[S.l.]: IEEE, 2016: 770-778.

[21] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks[C]// Proceedings of Computer Vision—ECCV 2016. Amsterdam, the Netherlands: Springer, 2016: 630-645.

[22] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2017: 2980-2988.

[23] 卢宏涛, 罗沐昆. 基于深度学习的计算机视觉研究新进展[J]. 数据采集与处理, 2022, 37(2): 247-278.
LU Hongtao, LUO Mukun. Survey on new progresses of deep learning based computer vision[J]. Journal of Data Acquisition and Processing, 2022, 37(2): 247-278.

[24] CHEN X T, CHEN X J, ZHA Z J. Structure-aware residual pyramid network for monocular depth estimation[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. [S.l.]: ACM, 2019: 694-700.

作者简介:

黄桦(1982-),男,高级工程师,研究方向:激光点云分类、三维GIS开发应用,E-mail: 61293447@qq.com。

朱宇昕(2000-),男,博士研究生,研究方向:三维重建,E-mail: zhu yuxin@nuaa.edu.cn。

章历(1980-),女,工程师,研究方向:航天摄影测量。

陈志达(1981-),男,正高级工程师,研究方向:航天摄影测量。

张乙志(1988-),男,高级工程师,研究方向:遥感影像处理。

王博(1988-),**通信作者**,男,副教授,研究方向:卫星摄影测量与计算机视觉,E-mail: wangbo_nuaa@nuaa.edu.cn。

(编辑:张黄群)