

文章编号 1004-924X(2023)05-0644-12

## 采用独立分支网络的多源遥感数据 自适应融合地物分类

戴莫凡<sup>1,2</sup>, 徐青<sup>1,2</sup>, 邢帅<sup>1\*</sup>, 李鹏程<sup>1</sup>

- (1. 战略支援部队信息工程大学地理空间信息学院, 河南郑州 450001;  
2. 智慧中原地理信息技术河南省协同创新中心, 河南郑州 450001)

**摘要:** 针对现有基于深度学习的地物分类方法大多面向遥感影像, 而对点云数据的空间信息利用不足, 特别是对点云和影像这种异源特征融合不够充分的问题, 提出了一种采用独立分支网络结构的多源遥感数据自适应融合地物分类方法。首先, 对配准好的 LiDAR 点云和遥感影像分别采用三维网络和二维网络提取各模态的空间几何特征和语义特征; 其次, 在点云空间对影像特征进行交叉模态采样和特征对齐得到基于点的多源特征; 最后, 采用一种基于注意力机制的非线性自适应特征融合方法实现二、三维语义特征的融合。实验结果表明, 本文方法通过网络训练能够实现自适应数据特征的多源遥感数据融合分类, 针对 ISPRS 多源遥感数据集的植被、建筑物和地面三类地物平均分类精度达到 85.87%, 相较三维点云语义分割的分类精度提高了 10.12%。本文提出的独立分支融合网络能够实现二、三维数据的交互学习与深度融合, 为遥感多源数据地物分类提供了一种新的思路。

**关键词:** LiDAR 点云; 遥感影像; 多源数据; 独立分支网络; 自适应特征融合  
**中图分类号:** P237 **文献标识码:** A **doi:** 10.37188/OPE.20233105.0644

## Semantic segmentation of multi-source remote sensing data self-adaptive fusion with independent branch network

DAI Mofan<sup>1,2</sup>, XU Qing<sup>1,2</sup>, XING Shuai<sup>1\*</sup>, LI Pengcheng<sup>1</sup>

- (1. *Institute of Geospatial Information, Strategic Support Force Information Engineering University, Zhengzhou 450001, China;*  
2. *Collaborative Innovation Center of Geo-information Technology for Smart Central Plains, Zhengzhou 450001, China*)  
\* *Corresponding author, E-mail: xing972403@163.com*

**Abstract:** Existing deep learning-based terrain classification methods are mainly for remote sensing imagery; however, the spatial information of point clouds is underutilized. Specifically, the fusion of heterogeneous features is insufficient for point clouds and imagery. To utilize multi-source features fully, we propose a self-adaptive fusion classification method of multi-source remote sensing data based on independent branch network in this study. First, three-dimensional (3D) and two-dimensional (2D) networks are used

收稿日期: 2022-08-08; 修订日期: 2022-09-10.

基金项目: 国家自然科学基金项目 (No. 41876105, No. 41371436, No. 42001338); 河南省自然基金项目 (No. 202300410535); 智慧中原地理信息技术河南省协同创新中心和时空感知与智能处理自然资源部重点实验室基金项目 (No. 212108)

to extract the semantic features of registered LiDAR point clouds and remote sensing imagery. From the 3D space, the features of imagery are then sampled and aligned with those of point clouds. Finally, a non-linear self-adaptive feature fusion module is proposed to realize the fusion of multi-source semantic features. The experimental results indicate that the proposed method achieves an average classification accuracy of 85.87% on the vegetation, building, and ground of the ISPRS multi-source remote sensing dataset. Through network training, multi-source remote sensing data can be more data feature-adaptive fused and classified; further, the accuracy is significantly improved by 10.12% compared with the 3D classification result. The proposed independent branch fusion network can realize interactive learning and deep fusion of 2D and 3D data, and it provide a new idea for terrain classification based on remote sensing multimodal data fusion.

**Key words:** LiDAR point cloud; remote sensing imagery; multi-source data; independent branch network; self-adaptive feature fusion

## 1 引言

基于遥感数据的地物精确分类是对地观测的重要研究内容之一,在地理国情监测<sup>[1]</sup>、智慧城市<sup>[2]</sup>、森林资源调查<sup>[3]</sup>、三维重建<sup>[4]</sup>等领域发挥着重要作用。近些年,随着遥感影像光谱分辨率和时间分辨率的提高,大量基于深度学习的地物分类方法被陆续提出<sup>[5-7]</sup>,将卷积神经网络应用于遥感影像可以显著提高深度特征的提取能力。然而单一数据源由于缺乏丰富多样的信息,仍面临某些地物类别难以实现准确分类的情况,联合利用多源遥感数据是突破这一瓶颈的重要解决途径<sup>[8-9]</sup>。激光雷达(Light Detection and Ranging, LiDAR)技术具有快速、主动、可穿透性强的特点,其获取的点云数据结构稳定,能够客观真实地表达场景的复杂几何信息,成为高精度地面三维信息的重要数据源<sup>[10]</sup>。目前,已有研究通过融合多源遥感数据的光谱和空间特性实现地物分类<sup>[11-13]</sup>,但受限于不同传感器及数据结构差异,多模态的遥感数据融合研究仍有很多挑战<sup>[14]</sup>。

对影像和点云的融合分类研究主要分为在输入层的前期融合、特征层的中间融合和决策层的后期融合三类。前期融合是最简单也是早期常见的策略,在特征提取前直接将影像上像点的RGB值与二维深度图像上像点的深度值进行融合<sup>[15]</sup>,但受限于多模态数据较显著的结构差异,其互补特性难以在输入端被充分体现,冗余信息还需通过后续特征提取方法进一步剔除;后期融合则避免了这种结构差异,但由于在特征提取过

程中多模态的特性没有得到交互,分类误差受到单模态特征提取方法影响较大。因此在特征层对多模态数据进行交互是更为合理和灵活的融合策略,也是目前采用最多的融合方法。

在特征层融合影像和LiDAR点云的方法通常以影像为基准,通过将点云转化为数字地面模型(Digital Surface Model, DSM),从影像和DSM中提取空间特征,然后采用支持向量机(Support Vector Machine, SVM)、随机森林(Random Forest, RF)和多项式逻辑回归(Multinomial Logistic Regression, MLR)等分类器对叠加的空间和光谱特征进行联合分类<sup>[16-17]</sup>。近些年,随着深度学习模型在海量数据上表现出了强大的特征提取能力,研究多通过对深度特征的概括提取和对数据结构信息重构耦合,从而实现多源数据在特征级层次的融合,并取得了可观的解译精度<sup>[18]</sup>。其中,Hong等<sup>[19]</sup>在2021年首次提出了一个通用的和统一的遥感深度学习框架,对后续研究具有指导意义。

然而,目前提出的特征层融合方法大多以二维影像为基准,忽略了点云的三维空间信息,或往往仅包含点云的高程信息<sup>[20]</sup>,而其独有的三维优势结构信息由于在特征输入前被投影至二维平面造成了严重的丢失,例如植被点云最明显的空间几何形态特征在二维平面上则表现较差。同时,由于多模态数据间的高度信息耦合及数据结构差异,现有以特征串联的融合方式会导致信息融合不够充分,甚至破坏单模态的特征学习过程,分类精度仍有待提高。

因此,针对上述问题,本文提出了一种采用独立分支的自适应二、三维数据特征融合的语义分割网络。采用独立分支的特征提取结构进一步兼顾了点云的空间几何结构,同时不干扰影像纹理信息的提取,在最大程度利用了二、三维数据特性的同时,实现不同模态间的交互。自适应特征融合模块能够在网络训练过程中动态优化融合异源特征的权重,确保两种类型的特征有效融合,从而使模型具备更强的泛化能力。

## 2 多源遥感影像融合方法

### 2.1 基于二维网络的特征融合方法

早期嵌入式传感器可以同时感知光学和深度信息,处理 RGB-D 数据对计算机视觉领域有着重要的意义,因此大量融合研究集中于对点云深度图的利用,这些方法同时也对后续基于二维网络的特征融合方法有指导意义。其中,基于深度学习的融合方法多以平行的分支网络结构为主,Eitel 等<sup>[21]</sup>利用两个平行卷积神经网络(Convolution Neural Network, CNN),在相同的完全连接层中对深度特征进行简单融合,实现 RGB-D 数据的分类。FuseNet<sup>[22]</sup>采用交叉融合算法将双支的 SegNet 网络架构合并到编码器部分,应用于 RGB-D 数据的语义分割。在计算机视觉领域,Liu 等人<sup>[23]</sup>提出了一种自适应门控融合生成对抗网络(Generative Adversarial Network, GAN),其采用两个数据的编解码网络,通过 RGB 特征引导深度特征的学习过程实现跨模式融合。

随着激光雷达的发展,大范围区域的高精度高程信息可以通过转化为 DSM 数据与影像相结合。Piramanayagam 等<sup>[15]</sup>采用基于 FCN-32s 的深度神经网络,以合并点云和 DSM 的特征,用于语义分割任务。Zhang 等<sup>[24]</sup>通过分析 FCN 中各层的敏感性和贡献从而创建了最佳的层融合架构。Huang 等人<sup>[4]</sup>采用改进的残差网络提取特征,并引入门控残差细化网络,将来自不同模式的原始数据连接到几个通道中,实现特征的交互。

但通过简单的级联实现多模态特征的融合可能会进一步增加模型的维数,并且由于缺乏大量的标记数据,模型可能会出现维数诅咒,影响最终的多模态数据分类性能。这些因素都会在

一定程度上影响光谱和 LiDAR 数据的协同分类精度。区别于传统的基于元素简单求和或连接运算的特征融合方法,Hosseinpour 等<sup>[25]</sup>提出了一种新的交叉模态融合网络,通过将每个模态信息送入独立的编码器中,并在编码器-解码器结构上,将具有高语义定义的高层次特征引入到低层次特征中实现对 RGB 和 DSM 数据的融合。

现有研究的结果表明,点云的高度信息与影像信息在二维神经网络中结合能够有效提高地物的提取精度<sup>[21]</sup>,但不同传感器获取的数据本身存在较大的差异,而现有特征层线性融合方式会导致在网络训练中产生不相关的特征而无法有效地利用模态的互补性。尽管多尺度信息对于影像提取上下文特征十分重要,但仅为了实现网络的交互而对不同分辨率间特征进行融合,往往会忽略了底层空间信息和高层语义信息间的联系,损坏原有特征的提取过程。同时,基于二维神经网络的多模态融合方法虽然可以借助传统特征有效提高分类精度<sup>[15]</sup>,但缺点是会造成三维空间结构信息的丢失,在具有复杂空间结构的地物或地形起伏变化较大处的点云上表现较差。

### 2.2 基于三维网络的特征融合方法

近几年,基于深度学习的点云分类方法迅速发展,并凭借其独有的三维结构优势展现出良好的应用潜力。相较于早期将点云投影<sup>[26]</sup>或转化为二维影像的多视图<sup>[27]</sup>表征后采用二维卷积的方法,基于体素<sup>[28-29]</sup>的方法将无序点结构转化为三维网格并采用三维卷积实现分类。早期很多融合方法也采用同样思路,但都面临着固有空间结构和几何信息丢失并存在大量冗余计算的问题,而较难应用于大范围遥感场景。Qi 等<sup>[30]</sup>在 2017 年提出的 PointNet 是第一个能够直接输入并实现点云分类的网络,相较于将点云转化为二维深度图的形式,三维的表征方式可以更直观地还原真实的物理场景,并最大程度保留原有空间特性,为后续研究进一步挖掘并利用点的空间邻域关系提供了一个基准框架。近些年的大量研究也是在该网络基础上进行改进<sup>[31]</sup>,通过提取更丰富的边缘和局部特征进一步提高了点云的分类精度。

在特征融合方面,多视图的方法多采用二维网络进行特征提取<sup>[32-33]</sup>, FusionNet<sup>[34]</sup> 和

PVCNN<sup>[35]</sup>采用体素与点相融合的方法提高了对RGB-D数据的特征提取能力,通过体素分支提取粗粒度特征进行辅助。在影像与三维点云数据融合方面,Khaled等<sup>[33]</sup>联合RGB影像和LiDAR点云数据实现了三维语义分割,其突出贡献是将RGB影像转换为极坐标网格下的映射表示,建立了特征级的融合结构。Li等<sup>[36]</sup>利用机载多光谱LiDAR点云自带的光谱数据进行特征补充,同时通过在图卷积模型中引入几何矩描述点云的几何特征,能够进一步提高精度。Widyaningrum等<sup>[37]</sup>结合正射影像的RGB信息,通过将二维色彩信息投影至点云上,实现特征的融合。Poliyapram等<sup>[38]</sup>则提出了端到端的逐点激光雷达和图像的多模态融合网络PMNet,分别采用二维和三维规则网络对影像和点云进行特征提取,通过特征连接和分类网络,实现对机载LiDAR点云的分类。

### 2.3 存在的问题

从现有融合方法中可以看出,多模态遥感数据的融合研究普遍存在以下难点:

(1)数据表征不同:二维影像数据是三维场景经过投影后的规则像素表征,地物分辨率有限,而三维点云数据则是无序、不规则的离散点,三维坐标信息准确,但点密度低;

(2)特征表达差异大:对影像而言,光谱特征和上下文空间特征是准确进行地物表达的重要信息,研究重点在于通过挖掘像素间的相关性实现形态特征和纹理特征的提取,而点云是真实的三维场景,离散三维点间的几何关系和空间结构更为重要;

(3)特征融合难度高:为了提取上下文空间特征,规则的影像数据多以层次化卷积作为特征提取网络的骨干结构,而点云数据由于不规则的几何结构,无法采用二维卷积的网络架构,而多采用基于点的多层感知机和图神经网络。

然而,现有研究无论是基于二维神经网络还是三维网络,重点多放在如何设计复杂的网络结构上,尽管广泛使用的基于串联的融合方式在特征表示方面取得了成功,但在融合不同属性,尤其是异构特征上的能力仍然有限,简单的线性特征融合方法仅提供特征映射的固定线性聚合,会误导模型关注与模态不相关的特征,影响单模态

的特征提取能力,甚至降低模型的表示。同时,融合网络架构过度依赖某一模态,往往会造成对单模态信息的忽略,例如将点云转化为二维DSM或深度图的方式会受限于投影,难以捕捉点云的局部特征,并存在与影像不一致的尺度问题;而将RGB信息直接赋予三维点云,同样面临投影后的影像信息与坐标信息不对称的问题,目前对于激光雷达点云与影像的融合研究还未有突破性进展。

## 3 独立分支的自适应特征融合网络

### 3.1 采用独立分支的特征提取网络

模态融合的优化目标是使最终的分类网络能够获得不同模态间更多的共同信息,但通常也会导致网络仅使用了两种模式之间共同信息,而丢弃了每个传感器独有的信息。因此,为了在特征层保留每个模态独有的信息,本文对点云和影像分别采用了独立分支的语义特征提取网络,如图1所示,其中对于每个样本点,它同时来自一个二维影像和一个三维点云,三维真实标签作为监督分类的固定信息,当影像标签存在时,也可以作为辅助监督信息。

#### 3.1.1 二维影像语义分割网络

考虑到遥感影像普遍存在的同物异谱和异物同谱的特点,本文采用了改进的多尺度特征融合全卷积神经网络<sup>[36]</sup>(Fully Convolutional Networks, FCN)用于遥感影像的特征提取。基础的FCN网络能够接受任意尺寸的图像输入,并采用全卷积化、上采样和跳跃结构实现逐像素的图像分类。其中跳跃结构实现了不同层次特征图的融合,能够更好地兼顾图像低层次的细节局部特征和高层次的语义特征。本文采用在三种不同分辨率的融合方法中,分类效果最好的FCN-8s网络作为二维语义特征提取网络。采用独立分支的多源遥感数据自适应融合分类网络结构如图1。

当影像的二维标签存在时,其可用于影像语义分割网络的监督分类,作为增强三维网络训练的辅助信息。通过采样可得到三维点云集 $T = \{t_i(x_i, y_i, z_i)\}_{i=1}^N$ ,其中任一三维点 $t_i$ 所在的像素位置 $t_s(x_s, y_s)$ ,如公式(1)所示:

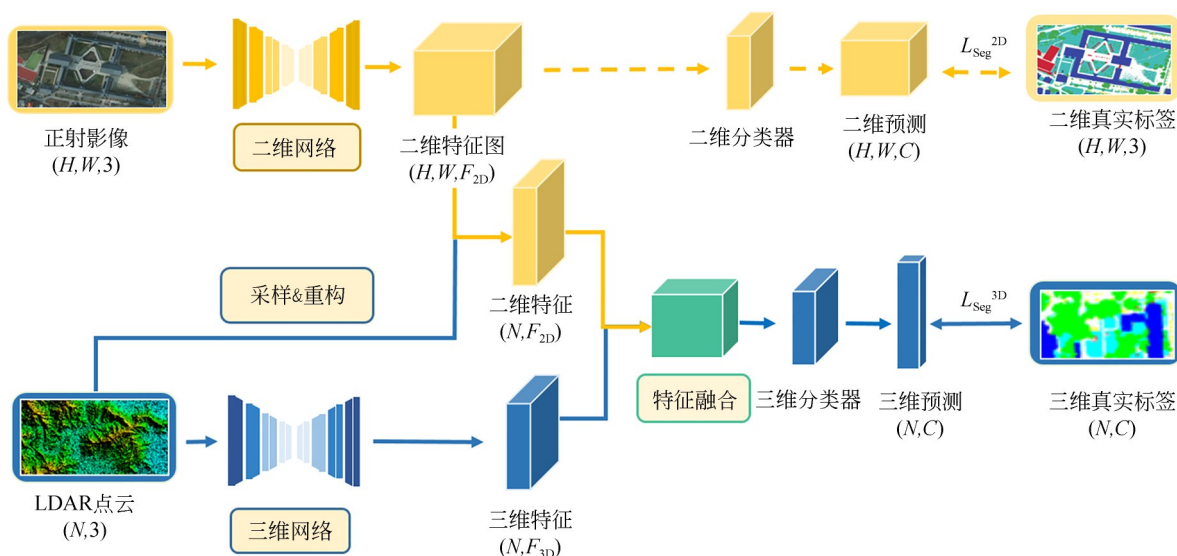


图 1 采用独立分支的多源遥感数据自适应融合分类网络结构

Fig. 1 Self-adaptive fusion classification network structure of multi-source remote sensing data with independent branches

$$\begin{aligned} x_s &= \text{int}((x_i - x_0)/dp), \\ y_s &= \text{int}((y_i - y_0)/dp), \end{aligned} \quad (1)$$

其中:  $x_i$  和  $y_i$  表示三维点的平面几何坐标,  $x_0$  和  $y_0$  表示影像所在区域的平面方向最小几何坐标,  $x_s$  和  $y_s$  表示三维点经采样后所在的像素坐标,  $dp$  为像素分辨率,  $\text{int}(\cdot)$  为取整函数。

此时, 对于分成  $C$  个类别的三维点云  $T_{s_i} = \{t_s(x_{s_i}, y_{s_i})\}_{i=1}^N$  的二维监督分类损失表示为:

$$L_{\text{Seg}}^{2D}(t_s, t_s'_{2D}) = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C t_s'_{2D}{}^{(n,c)} \log P_{t_s}{}^{(n,c)}, \quad (2)$$

其中:  $t_s'_{2D}$  表示经采样得到点所在像素的真实二维标签,  $P_{t_s}{}^{(n,c)}$  表示模型预测点  $t_s$  属于类别  $c$  的概率。

### 3.1.2 LiDAR点云语义分割网络

针对大尺度的机载 LiDAR 点云, 最近很多研究通过网络的改进探索局部结构来学习其多层次特征表示<sup>[39]</sup>。同样, 本文在现有研究的基础上, 采用作者之前所提出的一种融合几何卷积的神经网络<sup>[40]</sup>用于机载 LiDAR 点云的特征提取。该网络通过层次化卷积编码点的空间几何结构, 并与全局信息聚合能够提取多尺度逐点深度特征, 实现对于大区域点云复杂几何结构的获取。

同时, 采用点云分区、类别均衡等处理方法, 增强模型对机载点云的适用性, 在数据输入层面

保留原始的三维空间结构并直接输出逐点的地物提取结果, 此时三维的监督分类损失函数为:

$$L_{\text{Seg}}^{3D}(t, t'_{3D}) = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C t'_{3D}{}^{(n,c)} \log P_t{}^{(n,c)}, \quad (3)$$

其中,  $t'_{3D}$  表示经采样得到的点  $t$  的真实三维标签。

### 3.2 自适应特征融合

考虑到尽管使用了独立分支网络用于多模态的特征提取, 但由于点云和影像位于不同的度量空间, 直接融合 3D 点云与 2D 图像仍然困难。LiDAR 点云的三维坐标准确, 包含更多的结构信息, 但点密度低; 遥感影像覆盖范围广, 光谱信息丰富, 但地物分辨率有限, 几何结构信息不足, 如建筑物侧面等, 因此在两类数据特征融合时, 考虑三维信息和三维特征对二维具有较好的兼容性, 尤其是植被的多层空间表达, 采用将图像的二维特征采样至三维点集上的方式实现特征的对齐。具体而言, 本文在二维和三维网络输出概率前, 通过式(1)的采样过程从二维特征图中得到与点云中  $N$  个点所对应的  $N$  个像素, 从而得到  $N$  个点的影像特征, 将两个分支网络的语义特征输出保持相同的维度, 此时  $N$  个三维点既包含了二维影像的语义特征又包含了点云的深度几何特征和语义特征。

其次, 在特征融合策略上, 简单的拼接或相加的融合方法仅是对特征映射的固定线性聚合,

当场景改变时,无法保证模型的泛化能力。受人类视觉系统视觉感知的启发,注意力机制能够根据目标的重要性自适应分配相应的权重,因此点云和影像提取的语义特征的权重也应按照对分类性能的贡献重新分配。本文受图像中注意力

特征融合方法(Attention Feature Fusion, AFF)<sup>[41]</sup>的启发,提出了一种面向异源特征的自适应融合方法,如图2所示,能够依据点云的三维标签,动态学习并以非线性的方式优化点云二、三维特征的融合过程。

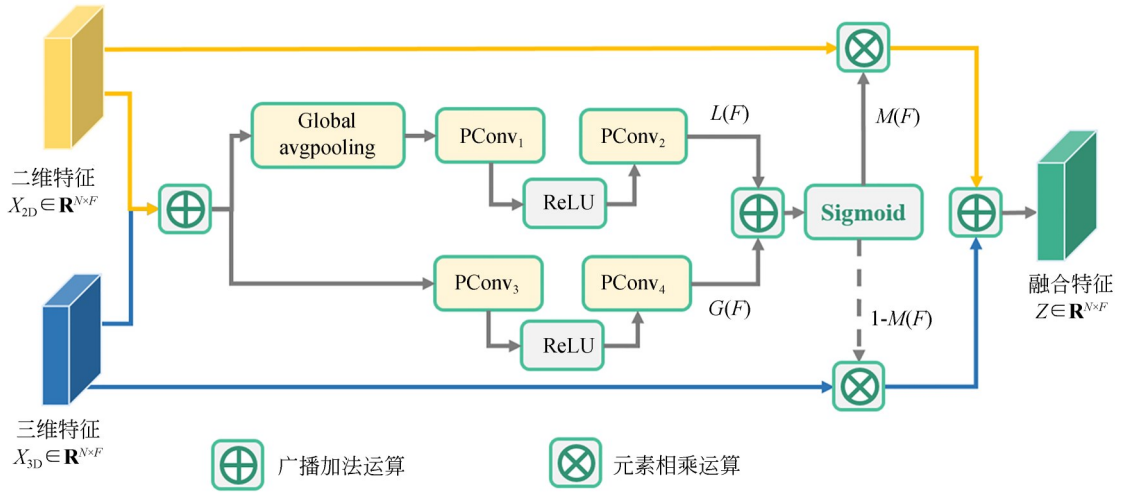


图2 自适应特征融合模块结构

Fig. 2 Self-adaptive feature fusion module structure

具体而言,给定要融合的点云二维和三维特征  $X_{2D}, X_{3D} \in \mathbb{R}^{N \times F}$ , 首先通过元素求和操作得到聚合特征。其次,利用逐点卷积实现每个点的跨通道信息融合,并通过瓶颈结构计算局部通道特征  $L(F) \in \mathbb{R}^{N \times F}$  和全局通道特征  $G(F) \in \mathbb{R}^{N \times F}$ :

$$L(F) = B\left(PCov_2\left(\delta\left(B\left(PCov_1(X)\right)\right)\right)\right), \quad (4)$$

$$G(F) = B\left(PCov_4\left(\delta\left(B\left(PCov_3(g(X))\right)\right)\right)\right), \quad (5)$$

其中: $g(\bullet)$ 表示全局平均池操作, $\delta(\bullet)$ 表示ReLU激活函数, $B(\bullet)$ 表示批标准化处理(Batch Normalization, BN), $PCov_i(\bullet)$ 表示各层具有不同输入输出通道数的逐点卷积运算。

值得注意的是,两个特征均与初始特征具有相同的维度,可以保留和突出底层特征中的细微细节,而后计算得到两类特征的注意力权重  $M(F)$ :

$$M(F) = \sigma(L(F) \oplus G(F)), \quad (6)$$

其中, $\oplus$ 表示逐元素相加运算。此时,融合的特征映射  $Z \in \mathbb{R}^{C \times F}$  可以表示为:

$$Z = M(F) \otimes X_{2D} + (1 - M(F)) \otimes X_{3D}, \quad (7)$$

其中, $\otimes$ 表示逐元素乘法运算,融合权重  $M(F)$  和  $1 - M(F)$  均由0到1之间的实数组成。此时网络能够对  $X_{2D}$  和  $X_{3D}$  的权重软选择或加权平均,在训练过程中实现动态优化,确保异源特征有效融合,从而得到对数据噪声更鲁棒的软标签,使模型具备更强的泛化能力。

## 4 实验结果及分析

### 4.1 数据集及评价指标

本文试验采用 ISPRS 提供的 Vaihingen 多源数据集<sup>[42]</sup>,包括分辨率为0.09 m的原始正射影像数据和平均点云密度为  $6.7 \text{ pts} \cdot \text{m}^{-2}$  的机载 LiDAR 点云数据。试验前对两块同时具有点云和影像的区域进行预处理,对其中的点云与影像数据进行配准与采样,得到两幅尺寸分别为  $2006 \times 3007$  和  $1919 \times 2569$  的训练影像和测试影像,以及相应分别包含 348 702 个点的训练点云和 174 145 个点的测试点云,试验区域如图3所示。同时,由于原始点云和影像数据的类别标注不统一,需对其进行类别分级及类别对齐,处理前后数据集的类别信息如表1统计所示。

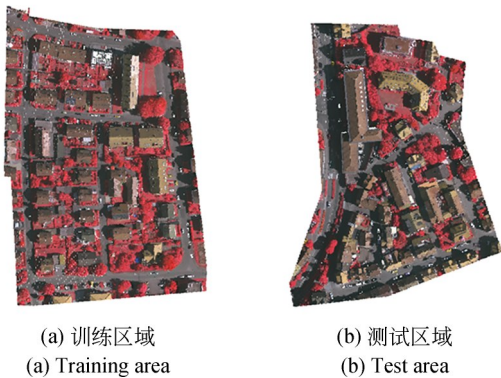


图3 Vaihingen数据训练集、测试集分布  
Fig. 3 Distribution of the Vaihingen training and test sets

试验均在 CPU Intel i7-9750H、Nvidia GeForce RTX3090 显卡硬件平台和基于 Ubuntu 18.04 系统下的 Cuda11.1、Python3.8 和 Py-

torch1.4 的软件平台下进行。输入批次 (Batch) 为 16, 迭代次数 (Epoch) 为 200, 初始学习率为 0.005, 学习率衰减系数为 0.5, 衰减步长为 20 000, 选择随机梯度下降和 L2 正则化的优化方法。精度评价指标包括精确度 (Accuracy)、总体精度 (Overall Accuracy, OA) 和平均交并比 (Mean Intersection over Union, mIoU)。

## 4.2 实验结果分析

### 4.2.1 多源数据融合实验结果分析

为验证本文融合方法在点云语义分割上的有效性, 实验首先从数据融合角度与目前主流的四种三维点云分类网络(即无影像数据条件下)进行对比, 评估多源数据融合的必要性以及本文采用的独立分支结构的性能, 实验结果如表 2 所示。

表 1 Vaihingen 实验数据集统计信息

Tab. 1 Statics of Vaihingen experimental dataset

类别 (Level_1)	影像类别 (Level_2)	点云类别 (Level_2)	训练样本	测试样本
建筑物类	建筑物	屋顶	112 377	51 520
		建筑物立面		
植被类	树木	树木	132 394	62 383
	低矮植被	低矮植被		
	/	灌木		
地面类	无法穿透的表面	无法穿透的表面	87 336	50 825
	车辆	车辆		
其他类	噪声/背景	栅栏/树篱	16 595	9 417
	/	电力线		

表 2 多源遥感数据融合分类精度

Tab. 2 Classification accuracy of multi-source remote sensing data fusion

Method	IoU			mIoU	OA	
	Ground	Vegetation	Building			
无影像条件	PointNet++ <sup>[31]</sup>	62.28%	52.31%	69.87%	41.10%	62.23%
	PointSIFT <sup>[43]</sup>	69.31%	50.92%	72.03%	48.65%	68.85%
	DGCNN <sup>[44]</sup>	76.50%	57.20%	79.72%	54.04%	76.42%
	RandLA-Net <sup>[45]</sup>	74.81%	59.52%	84.96%	56.93%	75.75%
<b>Ours</b>	<b>79.47%</b>	<b>75.52%</b>	<b>90.14%</b>	<b>68.57%</b>	<b>85.87%</b>	

注: 加粗为最优结果

本文方法在三类地物分类精度及总体精度上均获得了最佳结果, 总体精度达到 85.87%。在三维点云分类网络中, DGCNN 和 RandLA-Net 更适用于大场景点云数据, 其中 DGCNN 通

过对点云的不同尺度特征聚合, 提高了对地面点和植被的分类精度, RandLA-Net 方法从局部特征增强的角度, 进一步提高了整体的分类精度, 更有助于网络对边界进行识别, 其中建筑物的分

类精度超过80%。本文方法相较单独对点云分类的方法,通过独立分支网络引入影像数据特征极大丰富了网络的可用信息,显著提高了对植被和建筑物的提取能力,相比于RandLA-Net,精度分别提高了约16%和5%,整体分类精度和平均交并比分别提高了10.12%和11.64%,证明了纹理信息在地物分类中的重要性,同时独立分支结构能够极大程度保留多源数据的独有信息,显著提高了特征融合。

从图4显示的定性分类结果对比中也可以看到,测试区域在本文方法下的分类结果变得清晰可靠,整体上建筑物边界和主干道路边界变得整齐。在复杂区域的分类效果显著提高,当建筑物尺寸较小,且与地面点和植被点交杂时(黑色框区域),建筑物和道路的边界变得清晰,其他类点的干扰较少,相比于没有影像支持下的分类结果,本文方法显著提高了复杂点云分布区域的分类精度。

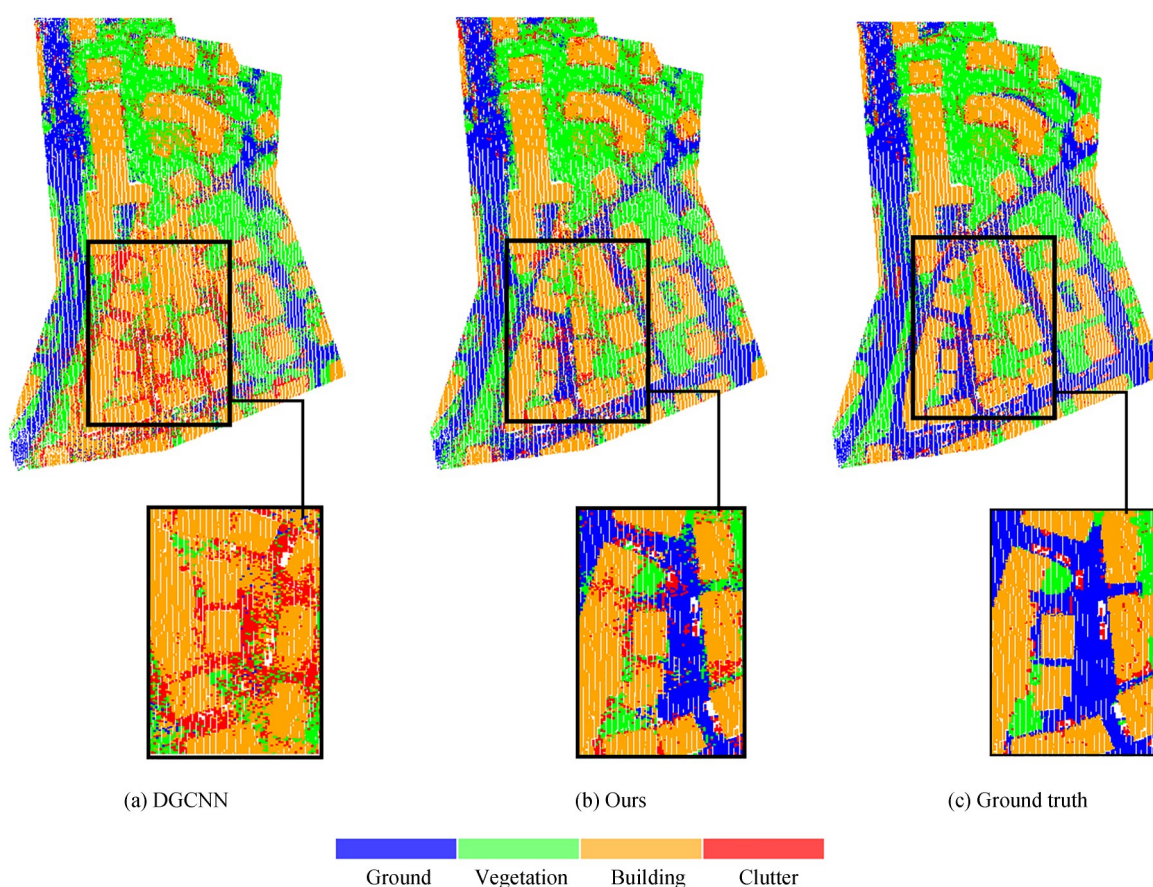


图4 测试区域分类结果

Fig. 4 Test area classification results

#### 4.2.2 自适应特征融合模块

赋予点云影像信息的自适应特征融合模块能够对提取的异源深度特征进行权值标定,进一步提高特征的识别能力。为了验证其有效性,实验从现有的独立分支网络结构中删除了该模块来进一步对比不同的特征融合策略。具体而言,在对齐二、三维语义特征后,分别采用将二维语义特征直接相加(3D-ADD)和拼接(3D-CAT)至

三维点特征上的两种线性特征融合方法代替基于注意力机制的自适应特征融合模块,进而评估本文自适应融合策略对异源特征的适用性能,实验结果如表3所示。

从表3中可以看出,在三类主要地物上,直接对两类深度特征相加方法的分类结果最差,不加区分地对特征进行融合会破坏各自提取特征的过程,采用拼接的线性特征堆叠方法尽管可以避

表 3 自适应特征融合方法性能  
Tab. 3 Performances of self-adaptive feature fusion methods

Method	IoU			mIoU	OA
	Ground	Vegetation	Building		
3D-ADD	67.18%	59.38%	74.61%	54.35%	68.53%
3D-CAT	76.30%	72.19%	86.21%	65.31%	82.65%
Ours	<b>79.47%</b>	<b>75.52%</b>	<b>90.14%</b>	<b>68.57%</b>	<b>85.87%</b>

注:加粗为最优结果

免独有特征的破坏,但会造成特征的冗余;而自注意模块在一定程度上可以提高分类精度,相较于线性拼接的方式分类精度提高了约3%。由此可以得出结论,使用非线性自适应特征融合模块可以更有效地融合异源特征,以获得更好的分类结果。

## 5 结 论

针对现有基于深度学习的点云影像融合分类方法对异源信息利用不够充分的问题,本文提出了一种采用独立分支网络结构的多源遥感数据自适应融合分类方法。具体而言,首先采用了具有2D和3D分支的双流架构,用于分别提取特定于遥感影像的空谱特征和LiDAR点云的几何特征,保证网络学习到的特征不限于模态学习到

的共同信息,避免了独有信息的丢失;其次在对齐二、三维语义特征基础上,将二维语义特征采样重构至三维点集上实现异源特征对齐,最后采用一种基于注意力的非线性特征融合方法,能够使网络在训练中自适应地动态优化异源特征的融合过程,从而具备更强的泛化能力。

在Vaihingen多源数据集上的实验结果表明,本文模型对三类地物的平均分类精度达到85.15%,相较仅使用三维网络方法精度提高了10.12%,相较简单的线性融合策略提高了17.34%;同时本文的可视化结果也显示了模型在细节上的较强分类能力,特别是在高树环绕的建筑和屋顶结构复杂的多层建筑等困难情况下仍然取得了优秀的性能。在未来的工作中,我们将进一步考虑跨模态的融合方法,并尝试在我们的框架中处理更复杂的城市点云。

## 参考文献:

- [1] 毕凯, 桂德竹. 浅谈地理国情监测与基础测绘[J]. 遥感信息, 2014, 29(4): 10-15.  
BI K, GUI D Z. Brief probe into national geographic conditions monitoring and fundamental surveying and mapping [J]. *Remote Sensing Information*, 2014, 29(4): 10-15. (in Chinese)
- [2] 李德仁, 姚远, 邵振峰. 智慧城市中的大数据[J]. 武汉大学学报·信息科学版, 2014, 39(6): 631-640.  
LI D R, YAO Y, SHAO Z F. Big data in smart city[J]. *Geomatics and Information Science of Wuhan University*, 2014, 39(6): 631-640. (in Chinese)
- [3] 李华玉, 陈永富, 陈巧, 等. 基于遥感技术的森林树种识别研究进展[J]. 西北林学院学报, 2021, 36(6): 220-229.  
LI H Y, CHEN Y F, CHEN Q, et al. Research progress of forest tree species identification based on remote sensing technology [J]. *Journal of Northwest Forestry University*, 2021, 36(6): 220-229. (in Chinese)
- [4] HUANG J F, ZHANG X C, XIN Q C, et al. Automatic building extraction from high-resolution aerial images and LiDAR data using gated residual refinement network[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019, 151: 91-105.
- [5] BERTASIUS G, TORRESANI L, YU S X, et al. Convolutional random walk networks for semantic image segmentation[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA. IEEE, 2017: 858-866.
- [6] GARCIA-GARCIA A, ORTS-ESCOLANO S,

- OPREA S, *et al.* A review on deep learning techniques applied to semantic segmentation[J]. *arXiv preprint arXiv:1704.06857*, 2017.
- [7] PANBOONYUEN T, JITKAJORNWANICH K, LAWAWIROJWONG S, *et al.* Road segmentation of remotely-sensed images using deep convolutional neural networks with landscape metrics and conditional random fields[J]. *Remote Sensing*, 2017, 9(7): 680.
- [8] ELSHEHABY A R, TAHA L G E D. A new expert system module for building detection in urban areas using spectral information and LIDAR data[J]. *Applied Geomatics*, 2009, 1(4): 97-110.
- [9] PEDERGNANA M, MARPU P R, DALLA MURAM, *et al.* Classification of remote sensing optical and LiDAR data using extended attribute profiles [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2012, 6(7): 856-865.
- [10] MALLET C, BRETAR F. Full-waveform topographic lidar: state-of-the-art [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2009, 64(1): 1-16.
- [11] RASTI B, GHAMISI P, GLOAGUEN R. Hyperspectral and LiDAR fusion using extinction profiles and total variation component analysis [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(7): 3997-4007.
- [12] GHAMISI P, RASTI B, YOKOYA N, *et al.* Multisource and multitemporal data fusion in remote sensing: a comprehensive review of the state of the art[J]. *IEEE Geoscience and Remote Sensing Magazine*, 2019, 7(1): 6-39.
- [13] CHEN Y S, LI C Y, GHAMISI P, *et al.* Deep fusion of remote sensing data for accurate classification[J]. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(8): 1253-1257.
- [14] DEBES C, MERENTITIS A, HEREMANS R, *et al.* Hyperspectral and LiDAR data fusion: outcome of the 2013 GRSS data fusion contest [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(6): 2405-2418.
- [15] PIRAMANAYAGAM S, SABER E, SCHWARTZKOPF W, *et al.* Supervised classification of multisensor remotely sensed images using a deep learning framework [J]. *Remote Sensing*, 2018, 10(9): 1429.
- [16] LI H, GHAMISI P, SOERGEL U, *et al.* Hyperspectral and LiDAR fusion using deep three-stream convolutional neural networks [J]. *Remote Sensing*, 2018, 10(10): 1649.
- [17] ZHANG M M, LI W, DU Q, *et al.* Feature extraction for classification of hyperspectral and LiDAR data using patch-to-patch CNN [J]. *IEEE Transactions on Cybernetics*, 2020, 50(1): 100-111.
- [18] XU X D, LI W, RAN Q, *et al.* Multisource remote sensing data classification based on convolutional neural network [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(2): 937-949.
- [19] HONG D F, GAO L R, YOKOYA N, *et al.* More diverse means better: multimodal deep learning meets remote-sensing imagery classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(5): 4340-4354.
- [20] 赵传, 张保明, 余东行, 等. 利用迁移学习的机载激光雷达点云分类[J]. *光学精密工程*, 2019, 27(7): 1601-1612.
- ZHAO C, ZHANG B M, YU D H, *et al.* Airborne LiDAR point cloud classification using transfer learning [J]. *Opt. Precision Eng.*, 2019, 27(7): 1601-1612. (in Chinese)
- [21] EITEL A, SPRINGENBERG J T, SPINELLO L, *et al.* Multimodal deep learning for robust RGB-D object recognition[C]. 2015 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Hamburg, Germany. IEEE, 2015: 681-687.
- [22] HAZIRBAS C, MA L N, DOMOKOS C, *et al.* FuseNet: incorporating depth into semantic segmentation via fusion-based CNN architecture [C]. *Asian Conference on Computer Vision*. Springer, Cham, 2017: 213-228.
- [23] LIU Z Y, ZHANG W, ZHAO P. A cross-modal adaptive gated fusion generative adversarial network for RGB-D salient object detection [J]. *Neurocomputing*, 2020, 387: 210-220.
- [24] ZHANG W K, HUANG H, SCHMITZ M, *et al.* Effective fusion of multi-modal remote sensing data in a fully convolutional network for semantic labeling [J]. *Remote Sensing*, 2017, 10(2): 52.
- [25] HOSSEINPOUR H, SAMADZADEGAN F, JAVAN F D. CMGFNet: a deep cross-modal gated

- fusion network for building extraction from very high-resolution remote sensing images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, 184: 96-115.
- [26] ALSHEHHI R, MARPU P R, WOON W L, *et al.* Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2017, 130: 139-149.
- [27] HU X Y, YUAN Y. Deep-learning-based classification for DTM extraction from ALS point cloud [J]. *Remote Sensing*, 2016, 8(9): 730.
- [28] SU H, MAJI S, KALOGERAKIS E, *et al.* Multi-view convolutional neural networks for 3D shape recognition [C]. 2015 *IEEE International Conference on Computer Vision. Santiago, Chile.* IEEE, 2015: 945-953.
- [29] WU Z R, SONG S R, KHOSLA A, *et al.* 3D ShapeNets: a deep representation for volumetric shapes [C]. 2015 *IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA.* IEEE, 2015: 1912-1920.
- [30] QI C R, SU H, MO K, *et al.* PointNet: deep learning on point sets for 3D classification and segmentation [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA.* IEEE, 2017: 77-85.
- [31] QI C R, YI L, SU H, *et al.* PointNet++: deep hierarchical feature learning on point sets in a metric space [J]. *Advances in Neural Information Processing Systems*, 2017, 30.
- [32] ZHOU Y, SUN P, ZHANG Y, *et al.* End-to-end multi-view fusion for 3d object detection in lidar point clouds [C]. *Conference on Robot Learning.* PLMR, 2020: 923-932.
- [33] EIMADAWI K, RASHED H, EISALLAB AEL, *et al.* RGB and LiDAR fusion based 3D Semantic Segmentation for Autonomous Driving [C]. 2019 *IEEE Intelligent Transportation Systems Conference. Auckland, New Zealand.* IEEE, 2019: 7-12.
- [34] ZHANG F H, FANG J, WAH B, *et al.* Deep FusionNet for point cloud semantic segmentation [C]. *European Conference on Computer Vision. Spring, Cham*, 2020: 644-663.
- [35] LIU Z, TANG H, LIN Y, *et al.* Point-Voxel CNN for Efficient 3D Deep Learning [C]. *Advances in Neural Information Processing Systems.* 2019, 32.
- [36] 王朝莹, 邢帅, 戴莫凡. 遥感影像与LiDAR点云多尺度深度特征融合的地物分类方法[J]. *测绘科学技术学报*, 2021, 38(6): 604-610, 617.
- WANG Z Y, XING S, DAI M F. A method of ground object classification based on multi-scale deep feature fusion of remote sensing image and LiDAR point cloud [J]. *Journal of Geomatics Science and Technology*, 2021, 38(6): 604-610, 617. (in Chinese)
- [37] WIDYANINGRUM E, BAI Q, FAJARI M K, *et al.* Airborne laser scanning point cloud classification using the DGCNN deep learning method [J]. *Remote Sensing*, 2021, 13(5): 859.
- [38] POLIYAPRAM V, WANG W M, NAKAMURA R. A point-wise LiDAR and image multimodal fusion network (PMNet) for aerial point cloud 3D semantic segmentation [J]. *Remote Sensing*, 2019, 11(24): 2961.
- [39] 王庆超, 付光远, 汪洪桥, 等. 多核融合多尺度特征的高光谱影像地物分类[J]. *光学精密工程*, 2018, 26(4): 980-988.
- WANG Q C, FU G Y, WANG H Q, *et al.* Fusion of multi-scale feature using multiple kernel learning for hyperspectral image land cover classification [J]. *Opt. Precision Eng.*, 2018, 26(4): 980-988. (in Chinese)
- [40] 戴莫凡, 邢帅, 徐青, 等. 多特征融合与几何卷积的机载LiDAR点云地物分类[J]. *中国图象图形学报*, 2022, 27(2): 574-585.
- DAI M F, XING S, XU Q, *et al.* Semantic segmentation of airborne LiDAR point cloud based on multi-feature fusion and geometric convolution [J]. *Journal of Image and Graphics*, 2022, 27(2): 574-585. (in Chinese)
- [41] DAI Y M, GIESEKE F, OEHMCKE S, *et al.* Attentional feature fusion [C]. 2021 *IEEE Winter Conference on Applications of Computer Vision. Waikoloa, HI, USA.* IEEE, 2021: 3559-3568.
- [42] ROTTENSTEINER F, SOHN G, JUNG J, *et al.* The isprs benchmark on urban object classification and 3d building reconstruction [J]. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* I-3 (2012), Nr. 1,

2012, 1(1): 293-298.

- [43] JIANG M, WU Y, ZHAO T, *et al.* Pointsift: A sift-like network module for 3d point cloud semantic segmentation [J]. *arXiv preprint arXiv: 1807.00652*, 2018.
- [44] WANG Y, SUN Y B, LIU Z W, *et al.* Dynamic graph CNN for learning on point clouds[J]. *ACM*

*Transactions on Graphics*, 2019, 38(5): 1-12.

- [45] HU Q Y, YANG B, XIE L H, *et al.* RandLANet: efficient semantic segmentation of large-scale point clouds[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA. IEEE, 2020: 11105-11114.

#### 作者简介:



戴莫凡(1998—),女,辽宁沈阳人,博士研究生,2020年于中国人民解放军战略支援部队信息工程大学获得学士学位,主要从事数字摄影测量、激光雷达点云数据处理等方面的研究。E-mail: daidai\_1115@163.com

#### 通讯作者:



邢帅(1979—),男,河南信阳人,博士,副教授,博士生导师,2001年、2004年、2008年于中国人民解放军信息工程大学分别获得学士、硕士、博士学位,主要从事激光雷达数据处理、数字摄影测量、海洋测绘等方面的研究。E-mail: xing972403@163.com