引文格式：赵斌,王乾. 基于深度特征聚合网络的高分辨率遥感图像语义分割[J]. 地理空间信息, 2025, 23(1): 80-84. doi:10. 3969/j.issn. 1672-4623.2025.01.017

基于深度特征聚合网络的高分辨率遥感图像语义分割

赵　斌[1]，王　乾[2]
（1. 广州市市政工程设计研究总院有限公司，广东 广州 510663；2. 山东黄金集团蓬莱矿业有限公司，山东 烟台 265621）

摘　要：高分辨率遥感图像语义分割在土地规划和智慧城市领域具有广泛应用。为准确快速地完成遥感图像语义分割，提出了一种基于深度特征聚合网络的语义分割方法。首先利用基础特征表示层获取不同分辨率的特征图信息，再利用中间特征聚合层和深层特征聚合层融合不同分辨率和尺度的特征图，最后通过特征聚合模块进行特征拼接，并输出精确的语义分割结果。在高分辨率遥感图像数据集 Vaihingen 和 Postdam 上的实验结果显示，该方法能精确分割不同场景下的遥感图像，优于现有多种遥感图像语义分割方法。

关键词：遥感图像；语义分割；特征提取；特征融合；深度学习
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2025）01-0080-05

Semantic Segmentation of High-resolution Remote Sensing Images Based on Deep Feature Aggregation Network

ZHAO Bin[1], WANG Qian[2]
(1. Guangzhou Municipal Engineering Design and Research Institute Co., Ltd., Guangzhou 510663, China;
2. Shandong Gold Group Penglai Mining Co., Ltd., Yantai 265621, China)

Abstract: Semantic segmentation of high-resolution remote sensing images is widely used in land planning and smart city. To complete the semantic segmentation of remote sensing images accurately and quickly, we proposed a semantic segmentation method based on deep feature aggregation network. Firstly, we used the basic feature representation layer to obtain the feature map information of different resolutions. Then, we used the middle feature aggregation layer and the deep feature aggregation layer to fuse the feature maps of different resolutions and scales. Finally, we used the feature aggregation module to perform feature splicing and output accurate semantic segmentation results. Experimental results on datasets Vaihingen and Postdam show that the proposed method can accurately segment remote sensing images in different scenes, and outperform existing methods.

Key words: remote sensing image, semantic segmentation, feature extraction, feature fusion, deep learning

遥感图像语义分割作为遥感数据定量分析的一种手段，在地理测绘[1]、农业生产[2]、智慧城市[3]等建设发展中发挥着重要作用，但准确快速地进行遥感图像语义分割面临诸多困难和挑战。现有遥感图像语义分割方法多依赖手工特征提取和有监督分类器，如基于阈值的遥感图像分割方法首先设计多个特征提取器获取不同尺度特征信息，再利用分层阈值法得到遥感图像分割结果；基于 K 均值聚类的遥感图像分割方法利用离散聚类对不同地物目标区域像素进行划分，再利用粒子群优化算法优化聚类迭代过程；基于模糊 C 均值聚类的遥感图像分割方法通过分层优化策略对地物目标像素进行细粒度切分，再利用逐层迭代方法优化聚类结果，但这些方法需要特定的特征提取方法和先验知识，所建立的语义分割模型的泛化能力较差[4-8]。得益于卷积神经网络（CNN）[9]强大的特征提取能力，深度学习在图像处理领域得到了广泛应用。基于 CNN 的方法能自动从输入图像中学习空间特征和拓扑关系，并根据学习的特征完成语义分割。注意力机制能帮助模型更加准确地捕获遥感图像中不同区域之间的关系和重要性，从而提升语义分割的性能和精度。视觉大模型基于深度学习和 CNN，具有强大的特征表示和学习能力，能够更好地理解遥感图像中的语义信息。鉴于 CNN、注意力机制和视觉大模型在语义分割中的优势，许多学者将其移植到遥感图像语义分割中，并取得了较好的进展，如 Li R R 等[10]提出包含多路径结构的 DeepUNet 模型，利用收缩路径提取遥感图像的多尺度特征信息，利用扩张路径提升特征图分辨率优化边缘区域分割结果；Pan X 等[11]构造了密集金

项目来源：国家自然科学基金资助项目（51375376）。
第一作者简介：赵斌（1970—），本科，工程师，从事工程测量工作，E-mail：zhaobin20220715@163.com。

字塔结构，通过不同通道特征信息的交互融合来增强模型的特征表示能力；Zhao J等[12]在语义分割模型中引入了条件随机场的学习和推理过程，并以循环神经网络的形式嵌入到FCNs模型中提升细粒度的语义分割效果；Zhang R等[13]将DeepLabV3模型得到的初始语义分割结果映射到三维点云数据，有效提升了对建筑物区域的语义分割精度；刘立等[14]结合深度学习和遥感地学特征构造了一种语义分割模型，利用连通相似性指数来评估构造的语义分割算法的有效性和分割精度；李宇等[15]通过CNN模型提取遥感图像的细节和高程信息，并利用条件随机场模型进行参数推理实现端到端的遥感图像语义分割；沈骏翱等[16]结合深度学习和注意力机制，通过挖掘遥感图像的形状、结构、纹理和色调等信息提升模型的感知能力，实现对遥感图像中水体区域的分割；刘尚旺等[17]结合U-Net模型和多任务学习，通过并行学习地物目标区域和边界区域的特征信息实现不同场景下的遥感图像语义分割。LANet利用补丁注意力机制和注意力嵌入模块获取局部和全局特征，以提升遥感图像语义分割精度。AERFC通过自适应调整不同卷积核感受野来获取遥感图像多尺度特征，以实现复杂场景下的遥感图像语义分割。AFNet通过尺度特征注意力和相邻得分图优化模块来增强遥感显著性特征表示。

1　研究方法

为充分利用遥感图像中包含的细节和空间上下文信息，提高语义分割的精度和速度，本文提出了一种基于深度特征聚合网络（DFANet）的遥感图像语义分割方法。DFANet包括基础特征表示层、中间特征聚合层、深层特征聚合层和特征聚合模块。DFANet的3级特征表示能精细调整特征信息，以弥补上采样和下采样产生的特征信息损失。特征聚合模块优化了不同特征提取层信息，使模型具有更好的特征学习能力。DFANet进行语义分割的过程可分为两个阶段：①特征编码器，通过基础特征表示层的卷积和池化操作得到不同分辨率的特征图；②特征解码器，通过上采样操作来恢复特征图分辨率尺寸。

1.1　基础特征表示层

基础特征表示层是以CNN为框架的浅层多尺度特征提取过程，与U-Net模型的区别在于取消了跳跃连接操作，通过中间特征聚合层、深度特征聚合层和特征聚合模块替代跳跃连接。除此之外，基础特征表示层增加了批归一化操作，以减少反向传播导致的网络参数偏移影响。基础特征表示层的计算公式为：

$$O_x = f(F_x;\theta)$$

式中，O_x 为基础特征表示层的输出；f 为卷积、池化和激活操作；F_x 为不同分辨率特征图；θ 为通过 f 学习的模型参数。

1.2　中间和深层特征聚合层

中间和深层特征聚合层（图1）在减少特征信息损失的同时，增强了基础特征表示层的特征表达能力。中间特征聚合层用于提取基础特征表示层的中间层特征，再将其传递到深层特征聚合层，同时完成特征聚合。基础特征表示层包括3次下采样和1次上采样操作，中间特征聚合层将对每次采样获取的基础层特征进行多次卷积操作。中间特征聚合层包括5次卷积操作，涵盖了基础特征表示层中所有上采样和下一层的下采样过程，并在第二次卷积操作后，通过裁剪和通道拼接的方式对基础特征表示层中获取的信息进行补充，增强模型特征表达能力。中间特征聚合层的计算过程为：

$$f_2 = M_1(f_1;\theta_1)$$
$$f_3 = M_2(f_2;\theta_2)\odot F_1$$
$$f_4 = M_3(f_3;\theta_3)\odot F_2$$
$$f_5 = M_4(f_4;\theta_4)\odot F_3$$
$$f_6 = M_5(f_5;\theta_5)\odot F_4$$

式中，$\odot$ 为对应元素相乘；M 为中间层卷积；θ 为卷积核参数；F 为聚合后特征。

为减少上采样和下采样操作导致的特征信息损失，中间特征聚合层采用裁剪和级联操作将聚合特征 f_3~f_6 迁移到基础特征表示层的上采样和下采样特征 B_4~B_7。

深层特征聚合层用于挖掘深层特征信息，提取遥感图像包含的固有特征和拓扑结构信息，位于模型底层。深层特征聚合层包括4次聚合操作，除第一次聚合操作之外，其他聚合操作都会输出聚合特征 AF。尽管深度聚合操作采用了通道拼接，但并不是只叠加通道级的原始特征。f_2 经过1次卷积操作后与 f_1 叠加，能有效避免通道拼接产生的特征不匹配问题。此外，经过深层特征聚合操作后，模型能进一步提取深层特征信息。深层特征聚合层的计算过程为：

$$J_1 = f_1 \oplus f_2$$
$$J_2 = J_1 \oplus f_3$$
$$J_3 = J_2 \oplus f_4$$
$$J_4 = J_3 \oplus f_5$$

式中，$\oplus$ 为通道拼接操作；J 为深层聚合特征。

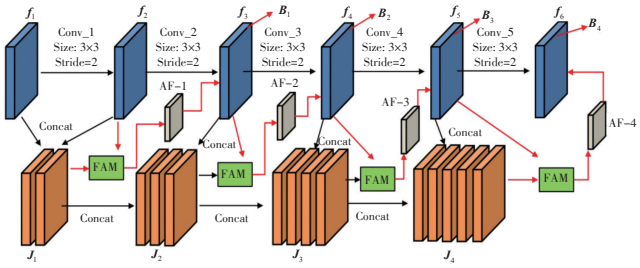


图1 中间和深层特征聚合层结构

注：Conv 为卷积操作；Concat 为通道拼接；f 为中间层特征；B 为基础层特征；J 为聚合层特征。

1.3 特征聚合模块

特征聚合模块用于融合中间聚合层和深层聚合层特征，并对特征进行加权调整，丰富特征信息。特征聚合模块以 f 和 J 作为输入，通过卷积操作得到特征 f_c 和 J_c。为补充特征维度、丰富语义特征信息，特征聚合模块通过通道拼接和相应元素相加得到聚合特征 A_1，再通过卷积和 SoftMax 函数运算得到聚合特征 A_3，使 A_3 具有 f 和 J 的信息内容。经过多次卷积操作，A_3 还包含了更深层次的语义信息。基于 A_1，A_3 进一步包含了更加抽象的权重特征信息。根据 A_3 包含的特征信息，特征聚合模块通过对应元素相乘的方式对 f_c 和 J_c 进行加权，从而进一步增强模型的特征表示能力；再通过原始相加和1×1卷积进行特征融合，获取聚合特征 AF。特征聚合模块中 A_1 和 A_3 的计算过程为：

$$A_1=(f_c(f;\theta_1)+J_c(j;\theta_2))\oplus J_c\oplus f_c$$

$$A_3=\frac{\exp(A_2(A_1;\theta_3))}{\sum\exp(A_2(A_1;\theta_3))}$$

式中，$\oplus$ 为特征通道级联；θ_1、θ_2 分别为 J_c 和 f_c 的卷积参数；θ_3 为 A_2 的卷积参数。

得到 A_3 之后先赋予 f_c 和 J_c 权重，再将其加入加权特征中，调整卷积操作所需的通道数量，使 AF 的通道数量与中间层相同。在上述操作中，A_1 不仅聚合了 f 和 J 特征信息，还引入了中间层和深层特征聚合层的特有特征信息。与此同时，通过卷积操作从 A_1 中提取特征 A_2，进一步提取特征和调整通道数量，为后续的特征加权提供特征权重。A_3 采用SoftMax函数激活 A_2，以确保 A_3 具有特征区分能力。A_3 对基础特征表示层和深层特征聚合层的加权表示为：

$$A_4=A_3\odot f_c+A_3\odot J_c$$

$$A_5=f_{A_4}(A_4;\theta_{A_4})$$

式中，$\odot$ 为对应元素相乘；f_{A_4} 为 A_4 和 A_5 之间的卷积操作；θ_{A_4} 为 f_{A_4} 的卷积核参数。

2 实验结果与分析

2.1 实验数据集

为验证本文方法的有效性，采用ISPRS提供的Vaihingen和Potsdam高分辨率遥感图像数据集进行语义分割实验，包括植物、道路、建筑、水体和背景5个类别。Vaihingen数据集主要包含真彩色影像、红外影像和地物分类等信息，而Potsdam数据集则提供了RGB（红—绿—蓝）彩色影像、红外四频带影像、卫星高程模型（DEM）和各种语义注释。图像平均尺寸为2 494×2 064，共37幅原始图像（图2）。实验过程中选取22幅图像作为训练数据集，6幅图像作为验证集，其余9幅图像作为测试集。



图2 遥感图像样本和对应的真实标注

2.2 实验参数设置

实验采用智能计算工作站进行算法部署，硬件配置为CPU Intel Xeon E5-2643v3，GPU 2*Titan RTX 24 GB，RAM 32 GB，Hard disk SAMSUNG 512 G，Operating system Ubuntu 18.02；采用PyTorch 1.2.0深度学习开源库作为深度学习代码编写工具。在模型训练过程中，选取随机梯度下降算法作为优化器，设置动量为0.9、初始学习率为0.01，最小学习率为0.001；选取交叉熵作为损失函数，余弦退火作为学习率衰减方法，模型迭代次数设置为2 000次，每200次迭代重置学习率。学习率衰减计算公式为：

$$lr=lr_{\min}+\frac{1}{2}(lr_{\max}-lr_{\min})(1+\cos(\frac{T_{\text{cur}}}{T_{\max}}\Pi))$$

式中，$lr_{\min}$、$lr_{\max}$ 分别为最小、最大学习率；T_{cur} 为当前迭代次数；$T_{\max}$ 为最大迭代次数。

为验证本文方法的有效性，采用F1_score、敏感度（SE）、特异度（SP）和准确度（PA）从定性和定量两个方面对模型进行评估，具体计算公式为：

$$\text{F1_score}=\frac{2TP}{2TP+FP+FN}$$

$$PA=\frac{TP+TN}{TP+FP+FN+TN}$$

$$SE=\frac{TP}{TP+FN}$$

$$SP=\frac{TP}{TN+FP}$$

式中，*TP* 为真阳性像素数量；*FP* 为假阳性像素数量；*TN* 为真阴性像素数量；*FN* 为假阴性像素数量。

2.3　实验结果分析

本文将DFANet与多个语义分割模型进行对比分析，以评估其有效性和优势。为确保实验的公平性，不同语义分割方法与DFANet均采用相同的训练数据集、训练策略和图像预处理方法。

2.3.1　Vaihingen数据集结果

DFANet与其他语义分割模型在Vaihingen数据集上的定量对比结果见表1，可以看出，DFANet在*PA*和F1_score指标上均优于其他语义分割模型，由于FCN、SegNet和U-Net模型的结构相对简单，因此无法充分提取遥感图像中包含的语义特征信息，且不同特征之间的相互作用较弱，使得分割结果较差；虽然PSPNet和Segmenter的分割精度与DFANet接近，但其计算参数、训练时间和测试时间均高于DFANet。DFANet与其他模型的定性比较结果见图3，可以看出，DFANet能够完全分割不同类别的地物目标，对于水体、道路等小区域的分割效果较好，而其他模型存在不同程度的误分割现象。

表1　DFANet与其他语义分割模型在Vaihingen数据集上的量化比较结果

方法	植物	道路	建筑	水体	背景	*PA*	F1_score	Parameter (1×10^6)
FCN	72.3	78.5	81.6	82.5	84.9	85.2	78.2	34.88
SegNet	71.6	79.2	82.7	81.7	86.8	85.8	77.1	28.56
U-Net	74.2	81.3	84.5	82.6	85.2	86.3	79.4	14.56
PSPNet	79.3	85.4	86.2	84.2	87.3	86.9	81.6	8.35
Segmenter	88.7	87.9	88.9	87.1	88.2	88.4	85.4	52.66
DFANet	91.2	90.3	92.4	91.5	87.7	90.3	89.2	8.14

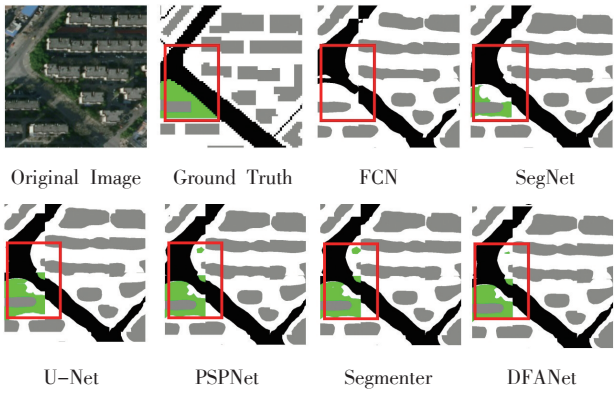


图3　DFANet与其他语义分割模型的分割结果比较一

2.3.2　Potsdam数据集结果

DFANet与其他语义分割模型在Potsdam数据集上的定量对比分析结果见表2，可以看出，DFANet在*PA*和F1_score指标上均优于其他模型，较弱的特征提取能力和遥感图像复杂背景的干扰导致FCN、SegNet和U-Net的分割精度较低；PSPNet和Segmenter对于小区域地物目标的细节特征提取能力有限，使其在Potsdam数据集上的分割精度不如DFANet；而DFANet对轮廓和细节区域的语义分割效果较好（图4）。除此之外，现有的语义分割模型在植物和建筑物的分割结果上出现了混淆问题，而DFANet能够取得较好的结果。其原因在于DFANet采用深度特征聚合层和特征聚合模块有效提升了模型的特征提取能力，抑制了背景特征信息对模型分割精度的影响。

表2　DFANet与其他语义分割模型在Potsdam数据集上的量化比较结果

方法	植物	道路	建筑	水体	背景	*PA*	F1-score	Parameter (1×10^6)
FCN	73.4	80.2	79.3	81.4	83.8	83.1	77.4	34.88
SegNet	72.6	81.3	82.5	79.6	81.8	84.2	76.8	28.56
U-Net	74.5	82.1	82.8	80.6	82.4	85.6	77.2	14.56
PSPNet	81.2	83.4	84.2	82.1	84.6	86.8	83.4	8.35
Segmenter	84.8	86.2	87.5	85.2	87.2	88.1	86.2	52.66
DFANet	89.4	90.5	91.5	88.9	91.2	89.7	87.8	8.14

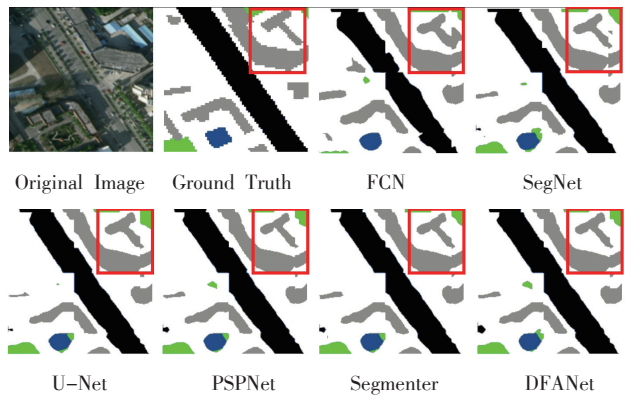


图4　DFANet与其他语义分割模型的分割结果比较二

3　结　语

为实现对遥感图像的准确语义分割，本文提出了一种基于DFANet的语义分割方法。该方法能获取丰富的多尺度特征，增强不同特征之间的信息交互，分割遥感图像的不同地物目标区域。DFANet的基础特征表示层能获取遥感图像的浅层特征信息，中间和深层聚合层能挖掘遥感图像的多尺度特征和语义特征信息，特征聚合模型可实现细粒度的特征融合。在公开的高分辨率遥感图像数据集Vaihingen和Potsdam上的定量和定性比较结果证明了本文方法的有效性和优势。未来将进一步关注如何将遥感图像包含的上下文信息融入语义分割模型的构造过程，实现对模型参数和结构的优化，并提升模型训练速度和分割精度。

参考文献

[1] 胡庆武,艾明耀,付建红. 基于测绘遥感地理信息工程的地理国情监测专业建设研究[J]. 测绘通报,2016(10):129-132

[2] 蒋瑞瑞,甘甫平,郭艺,等. 土壤水分多源卫星遥感联合反演研究进展[J]. 自然资源遥感,2024,36(1):1-13

[3] 赖积保,康旭东,鲁续坤,等. 新一代人工智能驱动的陆地观测卫星遥感应用技术综述[J]. 遥感学报,2022,26(8):1 530-1 546

[4] Shelhamer E,Long J,Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(4):640-651

[5] 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):2 481-2 495

[6] Wang X H, Wang X Y, Zhao K Y, et al. FSL-Unet: Full-scale Linked Unet with Spatial-spectral Joint Perceptual Attention for Hyperspectral and Multispectral Image Fusion[J]. IEEE Transactions on Geoscience and Remote Sensing,2022(60):1-14

[7] Zhang Z D, Wang X R, Jung C. DCSR: Dilated Convolutions for Single Image Super-resolution[J]. IEEE Transactions on Image Processing,2019,28(4):1 625-1 635

[8] Yu L J, Zeng Z X, Liu A. A Lightweight Complex-valued DeepLabv3+ for Semantic Segmentation of PolSAR Image[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2022(15):930-943

[9] Pang Y W, Sun M L, Jiang X H, et al. Convolution in Convolution for Network in Network[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(5): 1 587-1 597

[10] Li R R, Liu W J, Yang L. DeepUNet: A Deep Fully Convolutional Network for Pixel-level Sea-land Segmentation[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2018,11(11):3 954-3 962

[11] Pan X, Gao L, Zhang B, et al. High Resolution Aerial Imagery Semantic Labeling with Dense Pyramid Network[J]. Sensors,2018,18(11):3 774

[12] Zhao J, Zhong Y, Zhang L. Detail-preserving Smoothing Classifier Based on Conditional Random Fields for High Spatial Resolution Remote Sensing Imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(5):2 440-2 452

[13] Zhang R, Li G, Li M, et al. Fusion of Images and Point Clouds for the Semantic Segmentation of Large-scale 3D Scenes Based on Deep Learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing,2018(143):85-96

[14] 刘立,董先敏,刘娟. 顾及地学特征的遥感影像语义分割模型性能评价方法[J]. 自然资源遥感,2023,35(3):80-87

[15] 李宇,肖春姣,张洪群,等. 深度卷积融合条件随机场的遥感图像语义分割[J]. 国土资源遥感,2020,32(3):15-22

[16] 沈骏翱,马梦婷,宋致远,等. 基于深度学习语义分割模型的高分辨率遥感图像水体提取[J]. 自然资源遥感,2022,34(4):129-135

[17] 刘尚旺,崔智勇,李道义. 基于Unet网络多任务学习的遥感图像建筑地物语义分割[J]. 国土资源遥感,2020,32(4):74-83

(上接第61页)

地理科学,2017,37(8):1 218-1 225

[15] 王权,李阳兵,刘亚香,等. 基于地形梯度的岩溶槽谷区土地利用空间格局分析[J]. 生态学报,2019,39(21):7 866-7 880

[16] 范勋承. 基于GIS的松溪县土地利用空间格局分析[J]. 科学技术创新,2021(8):1-2

[17] 赵莉莉,何东进,覃德华,等. 2000—2014年闽东海岸带土地利用动态变化[J]. 热带作物学报,2017,38(5):803-810

[18] 陈柯欣. 基于DEM的土地利用空间格局分析:以大连市普兰店为例[J]. 软件,2019,40(2):111-115

[19] 赵晶,徐建华,梅安新,等. 上海市土地利用结构和形态演变的信息熵与分维分析[J]. 地理研究,2004,23(2):137-146

[20] 周子英,段建南,梁春凤. 长沙市土地利用结构信息熵时空变化研究[J]. 经济地理,2012,32(4):124-129

[21] 王佳楠,张志. 巴音河流域土地利用变化及驱动力分析[J]. 地理空间信息,2023,21(2):35-40

[22] 李娜,谢德体,王三. 基于区域分化下重庆市土地利用结构效率评价:综合运用信息熵和Malmquist指数[J]. 西南大学学报(自然科学版),2018,40(6):115-123

[23] 王玉明,王瑞康. 城市化时期郊区土地利用结构信息熵上升的原因[J]. 地理学报,2018,73(9):1 647-1 657

[24] 土地利用现状分类:GB/T 21010-2017[S].

[25] 陈利顶,杨爽,冯晓明. 土地利用变化的地形梯度特征与空间扩展:以北京市海淀区和延庆县为例[J]. 地理研究,2008,27(6):1 225-1 234

[26] 马士彬,张勇荣,安裕伦. 山区城市土地利用动态空间分布特征:以贵州省六盘水市为例[J]. 自然资源学报,2012,27(3):489-496

[27] 翟香,兰安军,廖艳梅,等. 地形梯度对土地利用格局时空特征的影响:以毕节市为例[J]. 湖南师范大学自然科学学报,2022,45(6):70-80