Theory and Practice of Science and Technology
2025, VOL. 6, NO. 10, 37-42
DOI: 10.47297/taposatWSP2633-456908.20250610

Research on Abstract Semantic Segmentation of Remote Sensing Data Based on UPerNet

Fei Sun[1], Run Wang[2,3], Xueming Li[4], Yurong Guan[1]

[1]Huanggang Normal University, School of Computer Science and Artificial Intelligence, Huanggang, Hubei 438000, China

[2]Hubei Key Laboratory of Resources and Eco-environmental Geology, Wuhan, Hubei, 430034, China

[3]Hubei Geological Environment Station, Wuhan, Hubei, 430034, China

[4]Huanggang Normal University, School of Foreign Studies, Huanggang, Hubei, 438000, China

ABSTRACT

In practical production and life scenarios, higher-level abstract semantics are typically of greater concern. This paper, based on deep learning methods, uses the abstract semantics of mining land degradation as an example and investigates the feasibility of directly segmenting high-level abstract semantics from GF-2 remote sensing data using the UPerNet model. Experimental studies on mining area data from three typical landform features in Hubei Province demonstrate that: 1) For abstract semantic classification of mining damage types, the UPerNet model achieves efficient and accurate segmentation, outperforming the U-Net model under identical experimental conditions; 2) In cases with limited and imbalanced datasets, UPerNet exhibits superior generalization ability when dealing with less distinguishable abstract semantics. These results confirm the feasibility of using the UPerNet model to extract higher-order abstract semantics from remote sensing appearance features, providing technical support for investigating mining land degradation and offering a reference for directly extracting high-level abstract semantics from remote sensing data in complex application scenarios.

KEYWORDS

UPerNet; Abstract Semantics; Deep Learning; Semantic Segmentation; U-Net

1 Introduction

Remote sensing data typically captures images of the Earth's surface from various platforms, providing a direct description of apparent ground features. Semantic segmentation techniques based on remote sensing data involve learning and training on these apparent features, making decisions based on their perceptual separability. However, in specific application scenarios, more abstract attributes are often assigned to these apparent features. Such attributes are generally high-level, abstract semantics that require integrated analysis of multi-source data. Therefore, directly extracting high-level abstract semantics from remote sensing data can leverage the advantages of remote sensing technology—such as real-time monitoring and large-area coverage—and holds significant importance. This paper aims to investigate the feasibility of extracting high-level abstract semantics from remote sensing data.

Mineral development often triggers ecological disturbances, such as ground subsidence, soil and water pollution, and vegetation destruction, which harm regional sustainable development. The Chinese government has consistently prioritized the sustainable development of mineral resources at the national level. In 2024, the *Notice on Launching the Pilot Program for a National Survey on Mining-Induced Land Damage* [1] was issued, emphasizing that conducting surveys on land damaged by mining is a foundational task for scientifically implementing ecological restoration of mining areas and holds significant importance. It is essential to comprehensively understand the extent of land damage caused by mineral resource development and to monitor in real time the excavation and compaction of land within mining areas. Compared to surface cover types in mining areas, describing damaged land types resulting from compaction and excavation involves more abstract, high-level semantics, and few studies directly utilize remote sensing data to achieve accurate extraction.

Intelligent methods represented by deep learning have been proven effective in remote sensing surveys of mining areas, such as location of settling ponds in mining areas [2-3], land classification in mining areas [4-5], environmental restoration of mining sites [6], lane departure monitoring for uncrewed vehicles in mines [7], scene recognition in coal mining areas [8], vegetation estimation in mining areas [9], and surface hazard monitoring in mining regions [10], etc. Semantic segmentation based on deep learning not only eliminates the cumbersome process of manual feature extraction through carefully designed network architectures [11] but also demonstrates superior segmentation performance on complex semantics [12]. This study focuses on typical open-pit mining land in Hubei Province, using Gaofen-2 (GF-2) satellite imagery as the primary data source, and conducts abstract semantic extraction of mined land damage based on artificial intelligence techniques. It further analyzes the feasibility of deep learning in extracting damaged abstract semantics, providing technical support for comprehensive surveys of mined land conditions.

2 Study Area and Data

2.1 Overview of the Study Area

The primary study areas are located in Jianshi County, southwest of Hubei Province; Yu'an County, in the central-western part of the province; and Wuxue City, in the east. Jianshi County is a typical mountainous county, characterized by numerous open-pit quarries that extract limestone, sandstone, and marble. Hilly and medium-to-low mountain landscapes characterize Yu'an County, serving as an important phosphate mining base in Hubei Province. Wuxue City is characterized by a landscape primarily composed of hills and plains, with open-pit mining areas primarily consisting of limestone quarries. In this study, 2017 mining land experimental sites were selected across these three regions (878 in Jianshi County, 582 in Yu'an County, and 557 in Wuxue City), with each mining site extended outward by a 300-meter rectangular buffer zone as the experimental area.

2.2 Data Sources

The remote sensing data source in this study is the sub-meter standard distribution data product from the Gaofen-2 satellite, which integrates 0.8-meter panchromatic (PAN) data with 3.2-meter multispectral (MS) data through registration and fusion, resulting in a data product with a spatial resolution of 0.8 meters that is suitable for quantitative analysis and research (Figure 1).

Figure 1 Typical remote sensing image of an open-pit mine

Based on the *National Pilot Survey on Mining-Induced Land Damage: Implementation Plan (Trial)*[1], this paper conducts indoor manual labeling and outdoor validation using historical land use data to produce mining-damaged land labels for each experimental area.

2.3 Abstract Semantic Description of Damage

Excavation-damaged land refers to land that has been damaged by surface excavation activities, including mining, sand extraction, and soil removal. Compacted land refers to land damaged by the compaction or occupation caused by stockpiling solid waste materials such as overburden, waste rock, mine slag, coal gangue, and tailings, as well as land occupied by mining-related buildings and structures.

Statistical analysis revealed an extreme imbalance in the number of patches and pixel proportions across the target mining area within the experimental regions (Table 1). Severe class imbalance, combined with the small and scattered distribution of targets, poses a significant bias threat to the training and learning of damage causation [13].

Table 1 Statistical overview of patches and pixels for primary damage types

Damage Type	Number of Patches	Proportion of Total Patches (%)	Number of Corresponding Pixels	Proportion of Total (Sample) Pixels (%)
Excavation Damage	3550	2.39%	101353676	13.42
Occupation Damage	8590	5.78%	86062370	11.39

3 Model and Methods

3.1 Model Selection

Mine area imagery typically exhibits characteristics such as complex land cover types, large-scale variations, and spectral confusion, often with imbalanced distributions. This study selects UPerNet [14] (Unified Perceptual Parsing for Scene Understanding) as the core segmentation model. The primary reasons are: 1) UPerNet integrates a Pyramid Pooling Module (PPM) with a Transformer encoder, enabling simultaneous enhancement of multi-scale features and effectively improving the segmentation performance for small targets in mining areas. 2) The model demonstrates strong robustness against spectral confusion. To address the severe spectral overlap commonly found in remote sensing data from mining areas, UPerNet leverages cross-resolution feature fusion to reduce interference from spectrally similar materials. 3) By employing feature channel pruning and optimization of pooling layer parameters, UPerNet achieves efficient real-time inference on a single GPU, making it particularly suitable for production environments with high demands for real-time performance.

To analyze the effectiveness of the UPerNet model, this paper selects the U-Net [15] model for performance comparison under the same experimental conditions.

3.2 Implementation Details and Parameter Optimization

This study crops the original data into image patches of 256 × 256 pixels, thereby establishing a uniform sample size. The actual training dataset, consisting of image patches containing damaged areas, comprises 11,439 samples. Of these, 90% are randomly selected for training, while the remaining 10% are used for testing to evaluate accuracy.

UPerNet and U-Net models utilize pre-trained models based on ImageNet [16], with ResNet-18 serving as the backbone network. As shown in Table 1, the total abstract semantic proportion of all damage types (compaction and excavation) is less than 25%. Therefore, given the extreme class imbalance, this study employs focal loss [17], which has been shown to significantly improve image classification performance on highly imbalanced datasets [18].

$$Focal\ Loss = -\sum_{i=1}^{N} \alpha \left(1 - p_i\right)^{\gamma} \log p_i \#(\text{Formula 1})$$

Here, α is the positive sample adjustment factor, γ is the parameter controlling the weighting distribution between complex and easy samples, and p is the model's predicted probability output.

This study employs the LR Finder and LDS (Low-Discrepancy Sequence Search) for hyperparameter optimization. Considering the hardware environment, the batch size is set to 32, and the learning rates for UPerNet and U-Net are 1.33e-04 and 1.17e-04, respectively.

The experimental hardware environment consists of an Intel Core i7 CPU, 64GB of memory, and an NVIDIA GeForce RTX 4070 with 12GB of memory.

3.3 Accuracy Evaluation

In this study, Pixel Accuracy (PA), F1 score, and Mean Intersection over Union (mIoU) are employed to compare the overall performance of models in abstract semantic segmentation; meanwhile, Precision, Recall, F1 score, and per-class Intersection over Union (IoU) are utilized as targeted evaluation metrics[12].

To balance the model's fitting and generalization capabilities, this study monitors the change in mIoU on the test dataset and halts training promptly when no improvement is observed over five consecutive epochs.

4 Experiments and Results

4.1 Abstract Semantic Segmentation of Damage Types Based on the UPerNet Model

4.1.1 Overall Performance Analysis of Abstract Semantic Segmentation for Damage Types

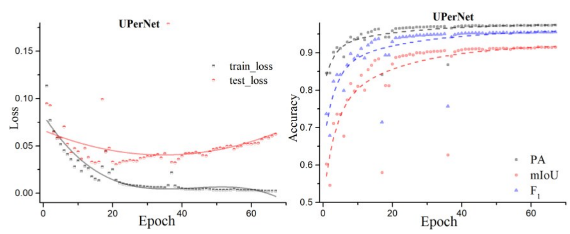

Figure 2 Loss convergence and accuracy performance of the UPerNet model for extracting abstract semantic segmentation in mining areas

Figure 2 illustrates the training loss of UPerNet and the overall performance of the model. UPerNet demonstrates significant performance improvement throughout the training process. The training loss monotonically decreases from an initial value of 0.11324 to 0.00208, while the test loss exhibits fluctuations but generally declines over time. Pixel accuracy (PA) increases from 84.55% to 97.34%, mean intersection over union (mIoU) rises from 0.6026 to 0.915, and the F1 score improves from 0.7365 to 0.9551, indicating that UPerNet effectively captures abstract semantic features for land degradation types in mining areas during training, thereby validating the effectiveness of both the model architecture and the training strategy.

However, overfitting was observed during the training. For instance, at epochs 17 and 36, the test loss abruptly increased to 0.099 and 0.179, respectively, while segmentation accuracy significantly declined, despite the training loss continuing to decrease. The results indicate a deterioration in the model's generalization capability.

4.1.2 Single-Class Performance Analysis of Damage Types Based on Abstract Semantic Categories

Regarding individual abstract semantic categories, the Intersection over Union (IoU) for the "excavation damage" semantic improved by 67.3% (from 0.544 to 0.9103) (Figure 3), recall increased by 11.3% (from 0.8515 to 0.9476), and precision improved by 59.5% (from 0.601 to 0.9585) (Figure 4). Correspondingly, for the "compression" semantic, IoU improved by 108.9%, recall increased by 58.4%, and precision rose by 58.3%. These results indicate that the model achieved high-quality segmentation performance for both abstract semantic categories. Notably, the model consistently outperformed its counterpart on the "excavation damage" semantic category compared to "compression," with a final IoU gap of 4.4%. The performance discrepancy may stem from class imbalance in the training data or inherent higher segmentation difficulty associated with the "compression" semantic category. Two distinct performance degradation phases occurred during training (at epoch 17 and epoch 36), indicating instability in model optimization progress. In later training stages (after epoch 40), the model's performance entered a stable plateau phase with minimal fluctuations across all metrics.

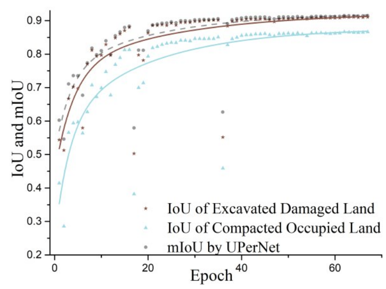

Figure 3 Accuracy performance of the UPerNet in extracting abstract semantic segmentation

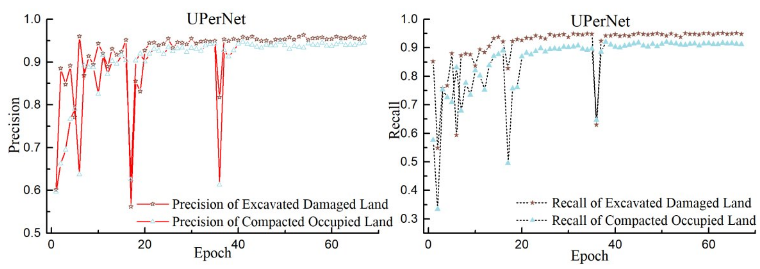

Figure 4 Loss convergence and accuracy performance of the UPerNet model for extracting abstract semantic segmentation in mining areas

4.2 Abstract Semantic Performance Comparison of Damage Types

By comparing the test errors and mIoU (Figure 5) between the UPerNet and U-Net models, UPerNet demonstrates a significant performance advantage, achieving a final mIoU of 0.915—3.64% higher than that of U-Net. In terms of efficiency, U-Net achieves high performance rapidly in the early stages but exhibits limited improvement in later phases. In contrast, UPerNet continues to improve throughout training, ultimately achieving superior segmentation accuracy. Regarding stability, U-Net shows relatively robust behavior with minimal fluctuations in test loss. Conversely, the optimization process of UPerNet exhibits instability. After Epoch 15, the U-Net enters a performance plateau. The test error curve of UPerNet suggests potential degradation in generalization capability during later training stages, necessitating a more sensitive early stopping strategy. These results indicate that while the U-Net model is suitable for resource-constrained application environments, the UPerNet model achieves superior performance when sufficient training resources are available.

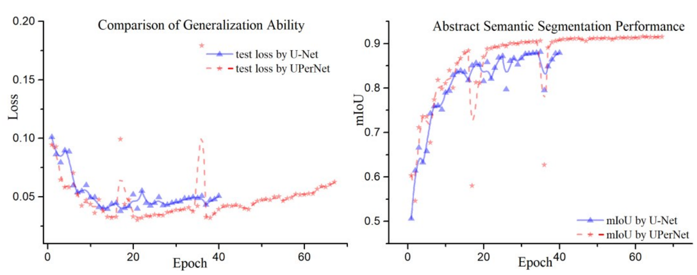

Figure 5 Comparison of Loss convergence and accuracy performance by UPerNet and U-Net for extracting abstract semantic segmentation in mining areas

Comparing the precision of the two models in extracting two types of abstract semantics (Table 2), UPerNet demonstrates a comprehensive performance advantage. The average IoU (excluding background) for UPerNet reaches 0.8886, which is 4.6% higher than that of U-Net. Notably, UPerNet excels in improving performance balance among abstract semantics. The IoU difference between excavation and compaction decreases from 6.65% in U-Net to 4.37%, while the F1-score difference reduces from 3.92% to 2.45%. The results indicate that UPerNet achieves a better balance in handling different abstract semantic categories, revealing its potential for addressing class imbalance issues and providing valuable reference for semantic segmentation applications in complex scenes.

Table 2 Accuracy performance of damaged land types by different optimal models

Model & Metric	mIoU Optimal Early Stopping Model								
Model	Epoch	Precision		Recall		IoU		F_1 value	
		Excav.	Occup.	Excav.	Occup.	Excav.	Occup.	Excav.	Occup.
U-Net	35	0.9399	0.9226	0.9278	0.8683	0.8758	0.8094	0.9338	0.8946
UPerNet	62	0.9561	0.9454	0.9502	0.9124	0.9104	0.8668	0.9531	0.9286

In this table, "Excav." and "Occup." are the abbreviations for Excavation damage and Occupation damage, respectively.

Figure 6 illustrates the generalization capability of UPerNet and U-Net on the reserved test area. From the visualized results, UPerNet demonstrates superior performance over U-Net in detailing the "occupation" semantics (as indicated by the red bounding box in the figure). However, for abstract semantic representation of "excavation," U-PerNet performs less effectively than U-Net, with the latter showing more accurate generalization along boundaries. Combined with the aforementioned quantitative analysis, UPerNet exhibits certain advantages in high-level abstract semantic segmentation tasks for remote sensing data, particularly demonstrating significant balancing capability for difficult-to-distinguish semantics in imbalanced scenarios.

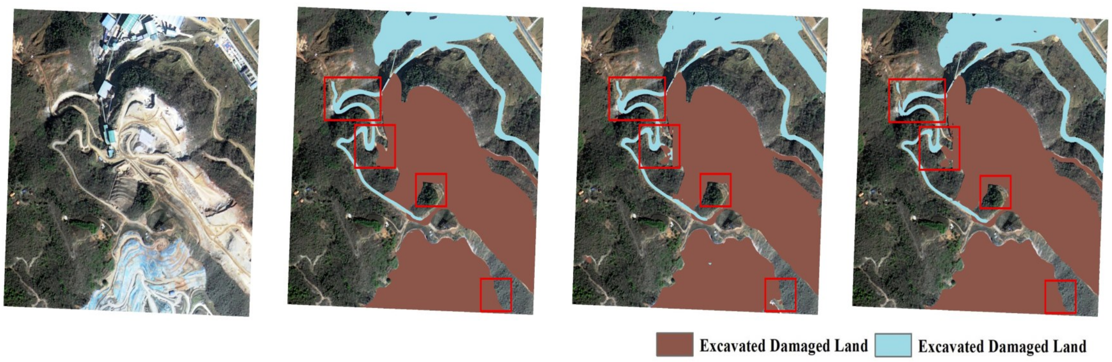

Figure 6 Generalization performance of UPerNet and U-Net for extracting abstract semantic segmentation by the mIoU-optimized models

5 Conclusion

Taking land degradation in mining areas as an example of abstract semantics, this study investigates the feasibility of directly segmenting high-level information from remote sensing data using the UPerNet model. The typical mountainous, hilly, and plain mining areas in Hubei Province were selected as the study sites, with the Gaofen-2 data employed for analysis. The results indicate that: (1) for abstract semantic categories related to mining damage, the UPerNet model achieves efficient and accurate segmentation, outperforming the U-Net model under identical experimental conditions; (2) in scenarios with limited and imbalanced datasets, UPerNet demonstrates superior generalization capability when distinguishing between less discriminative abstract semantic classes. In conclusion, deep learning models exemplified by UPerNet can accurately segment high-level abstract semantics based on remote sensing representations of land object characteristics, providing valuable reference for abstract semantic segmentation tasks in complex application scenarios.

Funding

This work was supported by the Science and Technology Project of Hubei Provincial Bureau of Geology (NO: KJ2024-89); the Natural Science Foundation of Hubei Province: (1) "Deep Intelligent Few-Shot Remote Sensing Classification under Visual Attention Mechanism" (NO: 2022CFB949); (2) "Deep Active Intelligence Interpretation of Imbalanced Remote Sensing Images under Quantitative-Spectral Collaboration"(NO: Q20212904)

References

[1] General Office of the Ministry of Natural Resources. Notice of the General Office of the Ministry of Natural Resources on carrying out the pilot work for the survey of land damaged by mining across the country[Z]. Beijing, 2024-07-18.

[2] Xiao S. Identification and Location Method of Sedimentation Tank in High Resolution Remote Sensing Image Based on the Improved Yolov3 Model [D]. Jiangxi University of Science and Technology, 2021.

[3] Li H, Xiao S, Wang X, *et al*. High-resolution remote sensing image rare earth mining identification method based on Mask R-CNN [J]. Journal of China University of Mining & Technology, 2020, 49 (06): 1215-1222.

[4] Zhang C, Li F, Li J, *et al*. Recognition of land use on open-pit coal mining area based on DeepLabv3+ and GF-2 high-resolution images [J]. Coal Geology & Exploration,2022, 50 (06): 94-103.

[5] Song R, Zheng H, Wang D, *et al*. Classification of features in open pit mining areas based on deep learning and high resolution remote sensing images [J]. China Mining Magazine,2022, 31 (07): 102-111.

[6] Zhu C, Chen Y, Lin J, *et al*. Empirical Study on Land Use Transformation of Rare Earth Mining Area in Hilly and Mountainous Areas of South China [J]. Journal of Gannan Normal University, 2022, 43 (06): 108-115.

[7] Jiao X. Research on Unstructured Road Recognition and Lane Departure Detection in Mining Area based on Machine Vision [D]. Xi'an University of Architecture and Technology, 2022.

[8] Zhao Y, Wei H, Dong J, *et al*. Opencast coal mine scene recognition based on sub-region multi-label learning [J]. National Remote Sensing Bulletin, 2022, 26 (09): 1849-1858.

[9] Ruan S, Jing W, Jing Y, *et al*. Detection on Vegetation Coverage Changes in Mining Areas Based on Improved DeepLabV3+[J]. Mining R & D, 2023, 43 (07): 143-151.

[10] Fan S, Gong M, Liu J, *et al*. Identification of Surface Hazards in Mining Areas Based on DeeplabV3+ Network and Attention Mechanism: Taking 30507 Working Face of the Tashan Coal Mine as an Example [J]. Science Technology and Industry, 2024, 24 (17): 278-286.

[11] Yuan X, Shi J, Gu L. A review of deep learning methods for semantic segmentation of remote sensing imagery[J]. Expert Systems with Applications, 2021, 169: 114417.

[12] Cheng J, Deng C, Su Y, *et al*. Methods and datasets on semantic segmentation for Unmanned Aerial Vehicle remote sensing images: A review [J]. ISPRS Journal of Photogrammetry and Remote Sensing,2024, 211: 1-34.

[13] Sun F, Fang F, Wang R, *et al*. An Impartial Semi-Supervised Learning Strategy for Imbalanced Classification on VHR Images[J].Sensors, 2020, 20 (22).

[14] Xiao T, Liu Y, Zhou B, *et al*. Unified Perceptual Parsing for Scene Understanding[C]. Computer Vision(ECCV 2018). Cham: Springer, 2018: 432-448.

[15] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]. Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015). Cham: Springer, 2015: 234-241.

[16] Russakovsky O, Deng J, Su H, *et al*. ImageNet Large Scale Visual Recognition Challenge [J]. International Journal of Computer Vision, 2015, 115 (3): 211-252.

[17] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[C]. 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 2999-3007.

[18] Dong J. Focal Loss Improves the Model Performance on Multi-Label Image Classifications with Imbalanced Data[C]. Proceedings of the 2nd International Conference on Industrial Control Network and System Engineering Research. Kuala Lumpur, Malaysia: Association for Computing Machinery, 2020: 18–21.