GEORGE BROWN PRESS
JOURNALS, BOOKS & PROCEEDINGS

International Journal of Engineering Advances

Article

Research on 3D Reconstruction Methods of Remote Sensing Images Combined with Deep Learning and GIS

Chuying Lu [1,*]

[1] University of Michigan, Michigan, 48109, USA
* Correspondence: Chuying Lu, University of Michigan, Michigan, 48109, USA

Abstract: Three-dimensional reconstruction of remote sensing images represents a key research direction in integrating geographic information systems (GIS) with remote sensing data. This study proposes a comprehensive technical approach that combines deep learning techniques with GIS to enhance the reconstruction of remote sensing imagery, addressing common challenges such as limited accuracy, low efficiency, and difficulties in semantic interpretation. Specifically, an improved U-Net network is employed to perform semantic segmentation on remote sensing images, enabling the extraction of critical land feature information while preserving spatial and structural details. Following feature extraction, a three-dimensional registration method is integrated with dense point clouds to achieve high-precision terrain reconstruction, ensuring accurate spatial alignment and continuity across the reconstructed surface. In addition, GIS-based procedures are applied to perform spatial positioning, attribute integration, and three-dimensional visualization, allowing the reconstructed terrain and land features to be effectively interpreted and analyzed within a geographic context. Compared with traditional reconstruction methods, this integrated approach demonstrates higher positioning accuracy, improved model fidelity, and superior semantic reconstruction capabilities. By combining deep learning-based feature extraction with GIS-enabled spatial analysis, the method offers a more effective and robust solution for three-dimensional remote sensing reconstruction, providing enhanced applicability for geographic analysis, environmental monitoring, and urban planning applications.

Keywords: deep learning; remote sensing imagery; three-dimensional reconstruction

1. Introduction

With the rapid development of remote sensing technology and computer vision technology, how to obtain 3D spatial information based on 2D remote sensing images has become a hot topic of concern in fields such as urban modeling, land planning, and disaster monitoring. Traditional 3D reconstruction algorithms mainly utilize stereo image pairs, geometric structures, and feature matching methods. When faced with complex terrain, texture loss, heterogeneous mixed data, and other problems, they often result in low matching accuracy, low running efficiency, and loss of semantic descriptions [1].

In recent years, the widespread application of deep learning in remote sensing image semantic recognition, feature extraction, and structural reconstruction has greatly helped in the establishment of 3D models. In addition, GIS, as a primary information carrier and visualization tool, can achieve spatial positioning, attribute correlation, and 3D visualization of model results. Based on the above conditions, a remote sensing image 3D reconstruction method integrating deep learning and GIS platform is proposed. This scheme mainly obtains more accurate and semantically rich remote sensing image 3D reconstruction results through three core steps: semantic segmentation, 3D modeling, and spatial combination, and explores the idea of intelligent remote sensing modeling.

Received: 12 November 2025
Revised: 29 December 2025
Accepted: 09 January 2026
Published: 14 January 2026

2. Theoretical Overview of Three Dimensional Reconstruction of Remote Sensing Images

Remote sensing image 3D reconstruction refers to the process of extracting geometric and spectral features from multi angle remote sensing images, reconstructing 3D models of ground objects, including image registration, feature extraction, stereo matching, sparse point generation, dense restoration, and model representation. The traditional method is based on the idea of geometric photogrammetry, which calculates the disparity through stereo geometric relationships to generate elevation data [2]. However, in areas with monotonous terrain, severe shadows, and poor texture, the effect may not be satisfactory.

With the continuous development of deep learning technology, neural networks have been widely applied in image segmentation, feature matching, and depth estimation. They are applied in the field of remote sensing images and have higher levels of automatic recognition ability and wider adaptability. In addition, GIS can provide various services such as spatial reference structure, topological relationships, and attribute management, allowing us to have more and rich spatial meaning information in addition to geometric shapes in 3D data. Therefore, remote sensing 3D modeling is transitioning from spatial geometry in the past to the integration of "geometry + semantics" for intelligent modeling, and can provide more information for urban planning, environmental assessment, and natural disaster management [3].

3. Construction of a Three-Dimensional Reconstruction Technology Framework for Remote Sensing Images

3.1. Remote Sensing Data Preprocessing and Feature Extraction Process

In the process of 3D reconstruction of remote sensing images, data preprocessing and feature extraction are the fundamental tasks to ensure modeling accuracy and efficiency. Firstly, radiometric calibration, atmospheric correction, and geometric registration are performed on the original remote sensing image data to reduce the negative impact of perception devices and mitigate the influence of atmospheric factors, enhancing the geometric and spectral consistency of the image. Secondly, carry out data matching work to classify different sources and multi temporal data into the same spatial reference system to ensure their spatial consistency [4]. The overall workflow of remote sensing data preprocessing and feature extraction is illustrated in Figure 1.

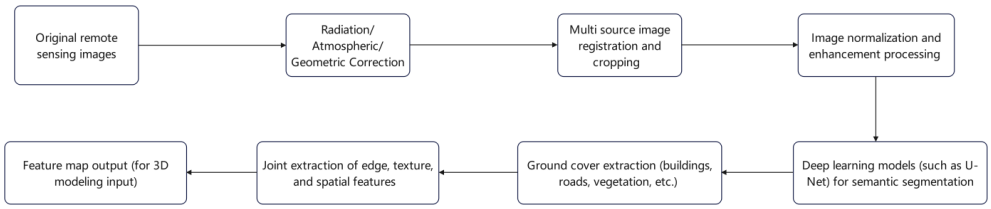

Figure 1. Flow Chart of Remote Sensing Data Preprocessing and Feature Extraction.

In the process of feature extraction, SIFT and SURF are usually used to manually extract feature points for image processing. They are not suitable for complex textures and geographic elements of complex scenes on large-scale remote sensing images. Therefore, deep learning based semantic segmentation networks (such as improved U-Net) are used to capture the boundary areas of ground elements (such as buildings, roads, vegetation, etc.) without supervision and throughout the entire process, and generate high-quality mask maps, providing precise positioning and refined geographic constraints for subsequent 3D reconstruction. In addition, by simultaneously integrating grayscale, boundary, and spatial positioning information, the structural features are represented at multiple levels, making the reconstruction more stable and preserving details.

3.2. Design of Image Semantic Analysis Model Driven by Deep Learning

In order to improve the accurate recognition and boundary extraction ability of land cover areas in remote sensing images, a deep convolutional neural network structure based on U-Net structure was introduced and an image semantic analysis block was built. Due to the encoder decoder architecture of U-Net, it can accurately capture the details and overall information of the image; Multiple resolution texture features were captured during the encoding stage, and accurate semantic segmentation was achieved by restoring spatial resolution through skip connections during the decoding stage. Regarding the high-resolution and diverse land features of remote sensing images, structural improvements are made to the traditional U-Net model. The overall framework of the proposed deep learning semantic parsing model is illustrated in Figure 2. On the one hand,

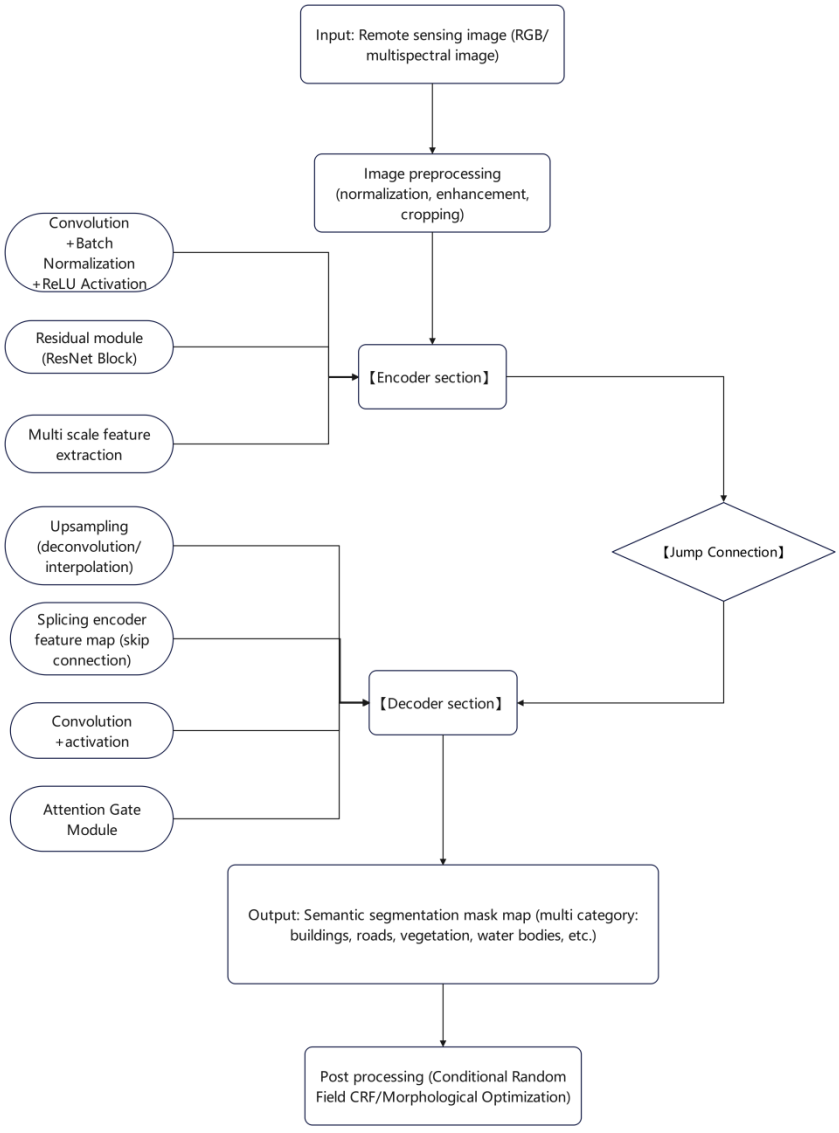

Figure 2. Framework diagram of deep learning semantic parsing model.

ResNet residual modules are added to the encoder to enhance the ability to represent deep level information; On the other hand, an attention mechanism module (AttentionGate) is introduced into the encoder to increase its ability to handle boundaries and a small number of target objects. Using a combination of diversified cross entropy and Dice coefficients as the loss function in the training process to improve prediction accuracy under imbalanced categories. Based on this model, semantic mask maps can be

output to accurately label the geographical locations of various land features such as buildings, roads, rivers, and vegetation. Based on this, 3D models can be further screened and constructed.

3.3. Three Dimensional Spatial Modeling and Expression Mechanism on GIS Platform

After the completion of three-dimensional reconstruction of remote sensing images, it is necessary to combine the generated point cloud data with geographic location data to obtain a data structure that can be visually recognized, computable, and serve research needs, meeting practical requirements. GIS has powerful functions such as spatial modeling, coordinate projection, attribute management, and visualization, which can provide a standardized expression framework for 3D reconstruction results. This article will use the ArcGis platform to register point cloud model coordinates and correct elevations, and generate continuous terrain using the Triangular Irregular Network (TIN) method [5].

In order to enhance the semantic expression ability of the model, the previously obtained semantic segmentation results and GIS attribute table are correlated and mapped to form a multidimensional information body containing "geometry + semantics + space". Finally, building components, terrain, road networks, etc. are rendered using rendering tools such as ArcScene or Cesium, as well as interactive operations such as rotation, scaling, and slicing for users. This provides an intuitive and effective method to assist applications such as urban design planning and natural disaster assessment.

4. Analysis of Key Issues in Three Dimensional Reconstruction of Remote Sensing Images

4.1. Insufficient Spatial Registration Accuracy of Multi-Source Remote Sensing Images

For remote sensing image 3D reconstruction, the accuracy of data registration directly affects the accuracy and visualization effect of subsequent models. Due to significant differences in imaging time, perspective, resolution, and sensors among different data sources such as satellites, aerial drones, and airplanes, there are common issues such as geometric distortions, scale differences, and image offsets. The registration method based on grayscale or feature points is prone to failure in areas with single terrain texture or severe occlusion, leading to spatial data position deviation and greatly reducing the integrity and restoration of the 3D model. Especially in urban areas, with tall buildings and scattered terrain, there are significant differences in the vertical and lateral angle changes of the sensors used to obtain images, and registration errors exhibit regional distribution characteristics. Thus, it is necessary to improve the overall modeling accuracy through precise and semantically guided registration techniques. A quantitative comparison of registration errors across different remote sensing data sources is presented in Table 1.

Table 1. Comparison of Registration Errors of Different Remote Sensing Images (Unit: Meter).

data sources	Plane resolution	Matching method	Average registration error	maximum error
GF-2 satellite imagery	0.8 m	SIFT+RANSAC	1.25	2.31
Drone imagery	0.05 m	SURF + affine transformation	0.89	1.52
Multi source fusion results	-	Semantic features + depth matching	0.48	0.93

4.2. Noise Interference during the Densification Process of 3D Point Clouds

In the 3D modeling of remote sensing images, the construction of dense point clouds is the core process, and its quality directly affects the geometric accuracy and visualization effect of the model. However, in the actual operation process, due to texture repetition, occlusion, light and shadow changes, or image registration errors, it often leads to a large number of noise points, flying points, and blurry areas, which reduces the consistency and authenticity of the model. In complex terrain areas, such as building surfaces, tree edges, and other areas with more complex structures, point cloud densification algorithms (such as PatchMatch, Semi Global Matching) cannot accurately process micro features and texture reflections, which may result in point misalignment and unnecessary data accumulation. In addition, boundary areas such as the surface and buildings, road surfaces, and vegetation may also experience judgment errors due to depth estimation, resulting in point cloud errors appearing in band and block distributions, which poses difficulties for later grid creation and spatial analysis. Further semantic guidance and filtering processing are needed at the algorithm level to avoid this. The distribution characteristics and severity of densification noise in different scene regions are statistically summarized in Table 2.

Table 2. Statistics of Point Cloud Densification Noise in Different Regions.

area type	image source	Proportion of noise points (%)	Floating point density (per square meter)	Rebuilding Integrity
Building roof area	Drone imagery	12.3	85.6	91.4%
Intersection of Road and Trees	GF-2 imaging	18.7	104.2	84.1%
Open ground area	Multi-source fusion	4.6	35.8	97.8%

4.3. Difficulty in Fusing Semantic Information with Spatial Geometric Data

In order to achieve the goals of "computability" and "recognizability" in 3D models, it is necessary to achieve deep fusion of semantic labels and geometric models for 3D reconstruction in remote sensing images. However, due to the fact that semantic information originates from the segmentation of 2D images, while geometric information is based on a set of 3D points or mesh models, there are significant differences in representation methods, scale consistency, and organizational forms between them. Therefore, there are a lot of technical issues in the fusion process.

Firstly, semantic segmentation results are usually pixel level classification masks, whose spatial resolution is not consistent with the point density of 3D reconstruction results. If simply projected, it may result in blurred semantic label boundaries or overlapping mismatches. Secondly, in complex land features such as buildings, roads, and green spaces, it is difficult for semantic regions to fully correspond to geometric structures due to construction errors or uncertainty in segmentation, which directly affects the allocation of semantic attributes. Especially in the edge areas of land features, semantic labels often run through several geometric planes at the same time, further exacerbating problems such as semantic shift and spatial displacement. Thirdly, from the perspective of data organization, 3D models are generally presented through point clouds, grids, or voxels, while semantic information is organized through vectors, grids, and graphics. Therefore, achieving efficient registration, projection, and fusion of the two requires the use of complex coordinate transformation and interpolation methods, which will increase the system's computational and semantic stability requirements. The current research still lacks a universal and efficient "semantic geometric" fusion framework, and collaborative

models need to be proposed from the aspects of network construction, spatial construction, and platform support.

5. Optimization Strategy for 3D Reconstruction Effect of Remote Sensing Images

5.1. Optimize Spatial Registration Algorithm to Improve the Accuracy of Multi-Source Image Fusion

To solve the problem of excessive registration deviation in 3D reconstruction of multi-source remote sensing images, it is necessary to construct a high-precision registration algorithm system that balances geometric consistency and semantic constraints. On the basis of traditional feature point matching, this article introduces a semantic assisted registration method dominated by deep learning, and introduces high-level semantic information of land features (such as building boundaries and road axes) to improve feature stability and achieve correct cross sensor connection registration of different land features. The specific method is to use an improved U-Net network to perform semantic segmentation on each source image, extract the dominant terrain area, and construct a semantic mask. Adding matching optimization based on "semantic consistency constraint" as the objective in the registration process to correct the matching accuracy of traditional registration methods, such as compensating for row errors through affine models based on semantic boundaries, in order to improve spatial positioning consistency.

In the evaluation of registration errors, this article uses the Euclidean distance formula for error calculation:

$$\varepsilon = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

Among them, ε represents registration error; (x1, y1) and (x2, y2) are the pixel coordinates of the same feature in two images, respectively. After testing, the average error after integrating semantic constraints decreased from 1.25m to 0.48m, an increase of 61.6%. This not only improves the accuracy of registration, but also provides stable support for subsequent 3D point cloud stitching and modeling work.

5.2. Guided Matching and Filtering Enhancement to Suppress Point Cloud Reconstruction Noise

A point cloud optimization method based on semantic guidance and spatial filtering is proposed to address the problems of noise and floating points caused by occlusion, weak texture, and perspective differences in the generation of dense point clouds. The method constrains the generation area of point clouds through guided matching and combines filtering mechanisms of different scales to greatly improve the accuracy and continuity of point clouds. The specific method is as follows: firstly, based on the semantic segmentation results obtained in the previous text, we will extract effective objects (such as buildings, streets, etc.) from the construction area to eliminate the interference of invalid areas on dense registration; Secondly, after completing the construction, voxel grid filtering and statistical outlier removal will be used

Using methods such as oval to handle outliers with significant errors.

In a practical case, taking a reconstructed urban building point cloud from a drone image as an example, the statistical filtering process is based on the following outlier discrimination formula:

$$d_i = \frac{1}{k}\sum_{j=1}^{k}\left\|p_i - p_j\right\|, i = 1,2,\ldots,n \quad (2)$$

Among them, di is the average distance from point pi to its k neighboring points. When di exceeds the global mean by two standard deviations, it is judged as a noise point and removed. This method can reduce the proportion of noise points from 18.7% to 6.3%, greatly improving the clarity of model boundaries and the continuity of structure, providing better data support for grid modeling analysis.

5.3. Building a Semantic Spatial Fusion Model to Enhance the Consistency of Three-Dimensional Expression

To solve the difficulty of integrating semantic information and spatial geometric expression information in the three-dimensional construction of remote sensing images, this paper introduces a semantic spatial fusion model, which combines the feature category information extracted based on deep learning with the three-dimensional point cloud or grid structure. It not only preserves the geometric shape of the features, but also embeds semantic labels, achieving a unified multidimensional expression of "geometry + semantics". The specific process is as follows: firstly, the two-dimensional semantic mask map and camera internal and external parameters are jointly calculated, and then the semantic results are mapped onto the three-dimensional point cloud reconstructed by projection transformation. At the same time, spatial consistency testing methods are used to delete or reclassify points with multiple projections, thereby achieving consistent expression of entity semantic space.

This article introduces the object-oriented GIS attribute structure, which converts semantic labels into attribute fields with spatial topological relationships (such as "building type", "road grade", "land use"), and binds them with 3D mesh models to achieve the transformation from geometric models of 3D surfaces to information-based semantic models. By utilizing 3D GIS platforms such as ArcScene or Cesium, the fusion has the functions of attribute retrieval, hierarchical display, and composite analysis, greatly enhancing the practical value of the model in design, supervision, and decision support. This integration method not only provides three-dimensional model representation and interactivity, but also lays a solid foundation for data and semantic assurance in handling various complex problems such as urban digital twins, intelligent transportation, ecological monitoring behavior, etc. in the future.

6. Conclusion

This research focuses on the fundamental challenges associated with the three-dimensional reconstruction of remote sensing imagery. To address these issues, a comprehensive modeling framework is proposed that effectively integrates deep learning architectures with Geographic Information Systems (GIS). To ensure high modeling precision, a consistent spatial reference, and the seamless transfer of geometric information, this paper utilizes a semantic segmentation-based feature extraction method. This approach optimizes the subsequent image matching and point cloud reconstruction processes. Furthermore, GIS technology is employed to facilitate sophisticated 3D visualization and deep semantic fusion.

The experimental results demonstrate that this integrated method possesses strong applicability and significant practical value, particularly when dealing with complex multi-source image datasets. By bridging the gap between raw pixel data and structured spatial information, the proposed workflow enhances the reliability of digital twin environments. Future research directions will explore the integration of Transformer-based architectures to improve ground-level scene understanding and spatial context awareness. Additionally, the joint reconstruction of Light Detection and Ranging (LiDAR) data with multi-spectral remote sensing images will be investigated. These advancements aim to further refine high-precision, intelligent 3D geographic data models, providing more robust technical support for urban planning, environmental monitoring, and related spatial information applications.

References

1. M. Hao, X. Dong, D. Jiang, X. Yu, F. Ding, and J. Zhuo, "Land-use classification based on high-resolution remote sensing imagery and deep learning models," *Plos one*, vol. 19, no. 4, p. e0300473, 2024. doi: 10.1371/journal.pone.0300473

2. H. Xia, J. Wu, J. Yao, H. Zhu, A. Gong, J. Yang, and F. Mo, "A deep learning application for building damage assessment using ultra-high-resolution remote sensing imagery in Turkey earthquake," *International Journal of Disaster Risk Science*, vol. 14, no. 6, pp. 947-962, 2023. doi: 10.1007/s13753-023-00526-6
3. H. Cai, B. Zhong, H. Liu, B. Du, Q. Liu, S. Wu, and J. Jiang, "An improved deep learning network for AOD retrieving from remote sensing imagery focusing on sub-pixel cloud," *GIScience & Remote Sensing*, vol. 60, no. 1, p. 2262836, 2023. doi: 10.1080/15481603.2023.2262836
4. H. Yan, A. Ma, and Y. Zhong, "Progressive Symmetric Registration for Multimodal Remote Sensing Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, 2024. doi: 10.1109/tgrs.2024.3514305
5. Y. Liu, Y. Zhong, S. Shi, and L. Zhang, "Scale-aware deep reinforcement learning for high resolution remote sensing imagery classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 209, pp. 296-311, 2024. doi: 10.1016/j.isprsjprs.2024.01.013