CCC 2025

Proceedings of the Creative Construction Conference (2025)
Edited by: Miroslaw J. Skibniewski & Miklós Hajdu & Žiga Turk
https://doi.org/10.22260/CCC2025/0025

IDENTIFICATION OF SURFACE STRUCTURAL-LAYER IN ROAD CONSTRUCTION USING LOW-ALTITUDE UAV IMAGERY

Jiayi Li[1], **Zhiliang Ma[1]**, Guangtai Lin[2]

1 School of Civil Engineering, Tsinghua University, Beijing, China,100084
2 Guangxi Beidou Low-Altitude Economy Investment Co., Ltd., Nanning, Guangxi, China, 530000

Abstract

Monitoring and assessing the progress and quality in road construction are crucial aspects of infrastructure development. Among these, the identification of surface structural-layer serves as a key task for progress control and quality management. Traditional manual inspection methods are inefficient and costly, especially given the spatially linear and narrow nature of road construction areas, making them inadequate for smart construction management. The application of low-altitude UAV imagery offers a new opportunity to address these challenges. However, its effective utilization hinges on the precise identification of surface structural-layer during construction. Conventional segmentation techniques often struggle with noise interference, leading to limited accuracy in identifying surface structural-layer. To overcome this, this paper proposes an improved DeepLabV3+ model for the identification of surface structurallayer in road construction. The model enhancement involves integrating a ResNet-50 backbone to bolster feature representation and extraction capabilities, thereby mitigating overfitting risks. Additionally, a CBAM (Convolutional Block Attention Module) is incorporated into the ASPP (Atrous Spatial Pyramid Pooling) module to enhance model performance, particularly in capturing fine details and boundary information. A self-constructed dataset is utilized, partitioned into training, testing, and validation sets in an 8:1:1 ratio. The performance of the proposed model is evaluated using four key metrics: overall accuracy, mean accuracy, frequency-weighted accuracy, and mean intersection over union. Comparative analyses demonstrate the superiority of the improved DeepLabV3+ model over the baseline in terms of segmentation accuracy. This paper provides a new method for the monitoring and assessment of road construction progress and quality, advancing the application of low-altitude UAV technology in intelligent construction practices.

Peer-review under responsibility of the scientific committee of the Creative Construction Conference 2025.

Keywords: Low-altitude UAV; DeepLabV3+; surface structural-layer in road construction; semantic segmentation

1. Introduction

Monitoring and assessing the progress and quality in road construction are essential components of the full lifecycle management of road infrastructure. The road structural layers typically include the surface structural-layer, base course, and subgrade, among others. Among these, the construction of the surface structural-layer is critical due to factors such as complex construction processes, stringent quality control standards, and significant impact on the overall structural performance of the road, making it a key task of progress control and quality management [1]. Precise identification of surface structural-layer is also an essential component in monitoring and evaluating the progress and quality of its construction.

Due to the significant spatial linear extension characteristics of road projects, construction areas often exhibit narrow and elongated strip-like distributions. This characteristic presents significant challenges to traditional methods of monitoring progress and quality [2]. Currently, monitoring and assessing the progress and quality of surface structural-layer during construction primarily rely on traditional manual inspection methods. These methods involve technicians conducting on-site sampling inspections using

Corresponding author email address:

tools such as rulers and levels. However, traditional manual inspections are inefficient and costly, making it challenging to achieve continuous, automated monitoring throughout the construction process. Additionally, the inspection data often exhibit significant delays, failing to meet the requirements for timely, comprehensive, and precise monitoring demanded by modern intelligent construction management. In this context, low-altitude UAV imagery technology, as a new generation of information technology tool, provides a new opportunity for the digitalization and intelligence of road construction management [3].

In recent years, the rapid development of deep learning technologies has provided strong technical support for the analysis and application of image data in the road sector. Notably, significant progress has been made in tasks such as road driving area recognition and road scene semantic segmentation, particularly in the field of autonomous driving [4]. However, the research on the application of deep learning technologies in utilizing low-altitude UAV imagery for road surface structural-layer remains limited.

To address the research gap in this field, this paper proposes an improved DeepLabV3+ model. Specifically, ResNet-50 is employed as the backbone network, and the CCBAM is incorporated into the ASPP module.

2. Related works

In recent years, with the rapid development of UAV technology, the application of low-altitude UAV imagery data in monitoring and assessing road construction progress and quality has gradually gained attention [5]. UAV can capture a large amount of image data in a short period of time, which is particularly advantageous for monitoring tasks involving long-duration flights or extensive areas. Traditional manual processing methods or computer vision techniques often struggle to efficiently handle such large-scale imagery data. The development of machine learning including deep learning technologies has provided strong technical support for the large-scale and efficient processing of UAV imagery data in the field of road construction. Reference [6] applied the MLR (Multivariate Linear Regression) algorithm to process UAV imagery of road subgrades during construction, aiming to achieve rapid prediction of soil moisture content in the subgrade. However, the study was conducted in a test field rather than a real road construction scenario. Reference [7] utilized the YOLOv4 and Deep SORT algorithms to monitor road construction safety factors, including construction personnel, vehicles, safety signs, and guardrails. Reference [5] applied CSN (Convolutional Siamese Networks) and CL (Contrastive Loss) functions to identify changes in road construction. By comparing feature vector differences between images taken at different times, the method detects variations in construction areas. In the method, the Siamese network extracts meaningful feature representations from the images, while the contrastive loss function adjusts the network parameters through backpropagation to ensure that extracted features are sensitive to changes and robust to unchanged areas. Nevertheless, this paper did not perform semantic segmentation to extract road information from UAV imagery. In contrast, it employs semantic segmentation algorithms to extract key objects from UAV imagery of construction sites, thereby eliminating irrelevant information, and significantly enhances the efficiency and accuracy of image processing, facilitating automated construction progress monitoring, quality inspection, and decision support [8].

Deep learning models for semantic segmentation include FCN, E-Net, U-Net, and the DeepLab series [9]. Among these, the DeepLabV3+ model is widely used in semantic segmentation tasks due to its excellent multi-scale contextual information processing capabilities [10]. Despite the excellent performance of the DeepLabV3+ model in semantic segmentation tasks, it still faces challenges such as insufficient detail capture and a strong dependence on large-scale datasets. To address these limitations, existing studies have introduced stronger feature extraction backbones, such as ResNet, to enhance the model's feature representation and segmentation accuracy. Additionally, attention mechanisms, like SE and CBAM, have been integrated to improve the model's ability to model detailed features, thereby effectively enhancing overall performance [11]. Literature [8] referenced and applied an improved DeepLabV3+ algorithm combined with a semi-global stereo matching algorithm for road boundary segmentation and image object boundary matching, subsequently constructing a road width measurement system. However, the paper focuses on the operation and maintenance phase rather than the construction phase of the road. In addition, the paper is based on image data captured by vehicle-mounted cameras, rather than UAV imagery.

To date, no applications of deep learning-based UAV imagery analysis methods for highprecision identification of surface structural-layer during construction have been reported, particularly concerning the delineation of structural layer boundaries. Besides, the boundaries of surface structural-layer in construction scenarios are often obscured by factors such as construction materials, roadside vegetation, and building shadows, making accurate boundary detection a significant challenge in current research. Such factors justify our study.

3. Data Preparation

High-quality data preparation is not only the foundation of model training but also a critical factor influencing model performance. Therefore, four aspects, i.e., the characteristics of road surface structural-layer under construction, data acquisition, data preprocessing, and dataset construction are included in this section. Based on the analysis of data characteristics, it elaborates on acquisition parameters, augmentation procedures, filtering methods, annotation strategies, and dataset construction approaches, ensuring that the subsequent model training is scientifically grounded and practically feasible at the data level.

3.1. Road Surface structural-layer and Their Characteristics Under Construction

The fundamental components of a road include the surface structural-layer, subgrade, bridges and culverts, tunnels, drainage systems, protective works, and traffic service facilities [12]. Among them, the surface structural-layer is a critical component of the road structure. Its construction progress directly influences the advancement of the entire road project, making it a key path in the construction schedule. Additionally, the quality of its construction directly affects the road's strength, durability, and driving safety.

The primary construction materials for surface structural-layer include cement concrete and asphalt concrete [13]. These materials possess distinctive surface characteristics, rendering surface structural-layer with prominent visual features in UAV imagery, thereby facilitating effective differentiation from non-road areas. However, due to the similarity in construction materials between surface structural-layer and other road structures, UAV imagery data often exhibit similarities in grayscale, colour, texture, and shape. Consequently, more advanced feature extraction techniques are required to achieve precise differentiation.

Data Acquisition

This study collected a raw UAV imagery dataset collected on December 19, 2022, during the construction of the Baxiang Interchange section of the Baxiang Road in Dahua Yao Autonomous County, Hechi City, Guangxi Zhuang Autonomous Region, China (hereinafter as "Guangxi Baqiang Road project"). The UAV model employed was the DJI Matrice 300 RTK, equipped with the Share UAV PSDK 102S V3 camera. The flight parameters were set as follows: terrain-following flight at an altitude of 180 meters, forward overlap of 70%, side overlap of 80%, and GSD (Ground Sampling Distance) better than 3 cm/pixel. During the image acquisition, the ambient temperature was approximately 18°C, with clear weather and a southeast wind at level 2.

3.2. Data Preprocessing

This study utilized a raw UAV imagery dataset collected during the construction phase of the

Guangxi Baqiang Road project. The dataset comprises 523 images with a resolution of 6000×4000 pixels. After excluding images that do not contain surface structural-layer (such as those of farmland and forested areas), 160 images meeting the criteria for construction surface structural-layer were selected. However, the limited number of 160 images is insufficient for training deep learning models. To address this issue, a Python script was developed to perform semi-automatic data augmentation on these 160 original images. The specific process of data augmentation is as follows:

(1) Starting from the top-left corner, crop 160 images with dimensions of 512×512 pixels.

(2) Starting from the bottom-right corner, crop 160 images with dimensions of 512×512 pixels.

(3) Starting from the centre of the image, crop 160 images with dimensions of 512×512 pixels.

(4) Apply a 90° clockwise rotation to the 480 cropped images obtained from the above steps.

Due to the loss of surface structural-layer information in some cropped images or the presence of minimal information only at the boundaries, it was necessary to perform a secondary screening of the 960 UAV images after preliminary processing. Ultimately, 373 valid images were selected for the construction of the subsequent self-constructed dataset.

3.3. Dataset Construction

To enhance the efficiency of semantic segmentation annotation in the self-constructed dataset, this study employed the semantic segmentation AI tool AnyLabeling software. The initial intelligent semantic segmentation annotations were performed using its built-in Segment Anything 2 (Hiera-Large) model. However, the annotation results obtained using this software exhibited numerous errors in the boundary regions, necessitating manual corrections for each image to ensure the acquisition of high-precision annotated data.

During the manual correction of annotations, various interference factors arise, such as construction vehicles, personnel, material stockpiles, dirt-covered pavements, tree obstructions, roadside buildings, and shadows cast by trees. To ensure the accuracy of semantic segmentation annotations for the surface structural-layer, the following strategies were implemented to address these interference factors:

(1) Annotate interference objects within the boundary of the surface structural-layer as part of the surface structural-layer. Construction vehicles, personnel, and temporary building materials within the surface structural-layer may obscure parts of the pavement but do not alter the road's structural integrity. Therefore, classifying these interference objects as part of the surface structural-layer helps prevent misclassification and ensures accurate segmentation of the actual pavement area.

(2) Annotate interference objects that completely obscure the boundary of the surface structural-layer as non-surface structural-layer areas. Interference objects that entirely block the boundary of the surface structural-layer make it impossible to accurately infer the actual boundary. Labelling these areas as part of the surface structural-layer could lead the model to learn incorrect boundary features, thereby reducing the accuracy of semantic segmentation.

(3) The area within the boundary line of the surface structural-layer is annotated as the surface structural-layer region, while the area outside the boundary is labelled as the non-surface structural-layer region. Despite the blurred boundaries caused by obstructions such as trees and building shadows, partial information of the surface structural-layer remains identifiable. This approach helps retain valid surface structural-layer information, enhances the model's ability to recognize the surface structural-layer, and improves its handling of interference objects, thereby ensuring the accuracy of semantic segmentation.

The 373 annotated images obtained from the above process were exported as JSON format label files. Subsequently, a Python script was developed to batch convert these files into grayscale PNG images. This process resulted in 373 original images of 512×512 pixels, 373 corresponding grayscale images, and 373 JSON files containing the annotation information.
These datasets were then divided into training, validation, and test sets in an 8:1:1 ratio.

4. Improvement of DeepLabV3+

DeepLab is a representative series of semantic segmentation models built on convolutional neural networks. It achieves high-precision, pixel-level segmentation by integrating modules such as atrous convolution and ASPP into the Fully Convolutional Network (FCN) framework.

Specifically, DeepLabV1 introduced atrous convolution into the FCN architecture. Building on this, DeepLabV2 proposed the ASPP module. Further, DeepLabV3/V3+ optimized the dilationrate scheduling of atrous convolution and introduced an encoder–decoder structure.

After multiple iterations, the latest DeepLabV3+ model incorporated the encoder–decoder architecture, which not only enhances feature representation, but also significantly improves boundary-detail recovery, thereby boosting overall segmentation performance [14-17].

Specifically, DeepLabV3+ enhances segmentation accuracy, particularly along object boundaries, by integrating high-level semantic features extracted by the encoder with low-level spatial features. Its overall architecture employs a plug-and-play backbone design (DCNN, Deep Convolutional Neural Network), allowing the selection of Xception, ResNet, or MobileNet according to application needs [12]. To expand the receptive field without reducing feature-map resolution, DeepLabV3+ introduces atrous convolution in the final stage of the chosen backbone. The ASPP module in the encoder comprises 1×1 convolutions, three 3×3 convolutions with different dilation rates, and an image pooling layer. These components are designed for feature dimensionality reduction, multi-scale contextual information extraction, and global context acquisition, respectively. The decoder processes low-level features through 1×1 convolutions, fuses them with high-level semantic feature maps (upsample by 4), refines the features via multiple 3×3 convolutions, and performs another fourfold upsampling to output precise segmentation predictions. The overall architecture of the DeepLabV3+ network structure is illustrated in Figure 1.

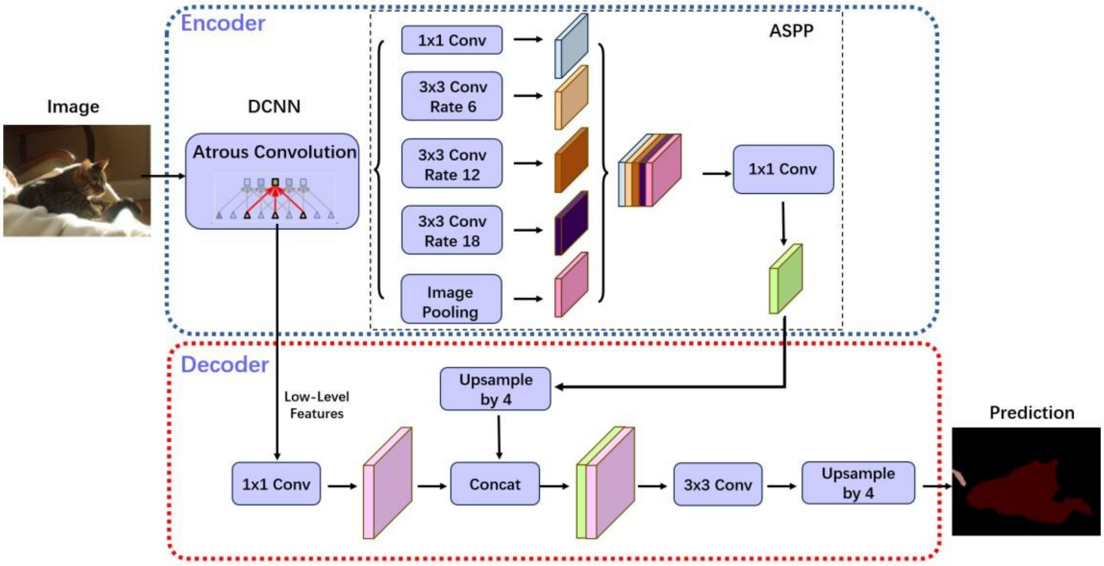

Fig. 1. Network architecture of the DeepLabV3+ network structure [17].

This study improves upon the baseline DeepLabV3+ model [18]. Inspired by the approaches in [9] and [11], the original backbone network is replaced with ResNet-50 to enhance the model's representation and feature extraction capabilities while reducing the risk of overfitting [9]. In addition, to further enhance the model’s performance, the CBAM module is integrated into the ASPP module located beyond the backbone network within the encoder [19]. The architecture of the improved DeepLabV3+ model is depicted in Figure 2. The following explains the rationale.

4.1. Backbone Network and Its Improvements

The surface structural-layer of roads often possesses complex boundaries and intricate details. Accurately segmenting these boundaries is crucial for precise identification of the distribution and morphology of surface structural-layer during construction. To improve segmentation accuracy, employing ResNet-50 as the backbone network is an effective approach. ResNet-50, through its deep network structure and residual connections, strengthens the model's ability to extract detailed features, particularly excelling in capturing complex boundaries and fine details.

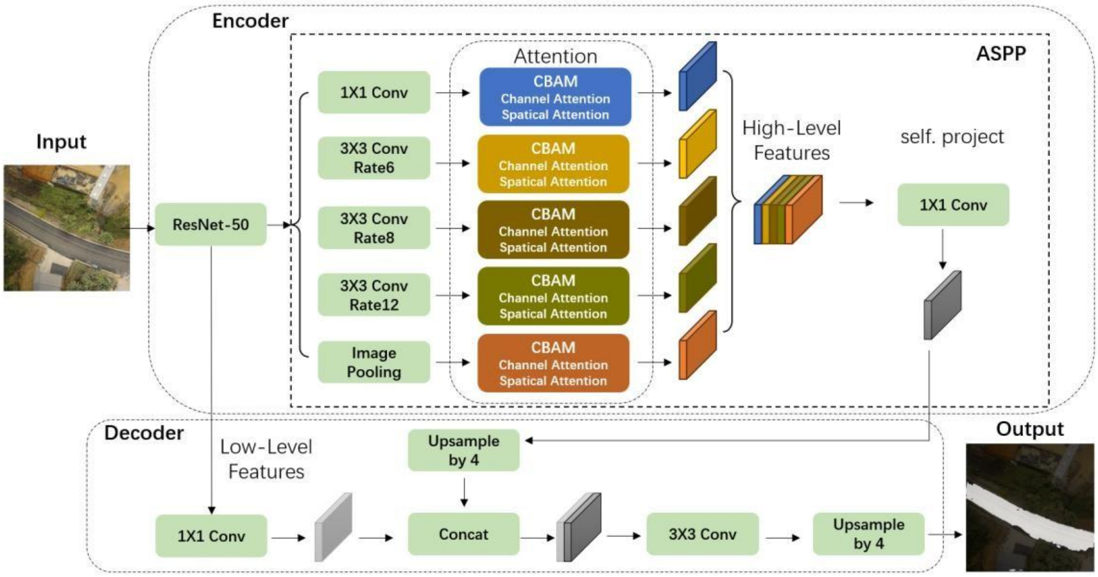

Fig. 2. Improved DeepLabV3+ network structure.

In traditional convolutional neural networks, information is processed through sequential layerwise propagation, which can lead to the vanishing gradient problem, especially in deeper networks. To mitigate this issue, ResNet introduces residual connections. The core idea of a residual connection is expressed by the formula:

$$y=F(x, \{W_i\})+x$$

In this equation, x represents the input feature map, F (x, {Wi}) denotes the feature map obtained after processing through convolutional layers and activation functions, and y is the output feature map. The residual connection facilitates the direct transmission of information between different layers, ensuring that gradients can effectively propagate to deeper layers, thereby addressing the vanishing gradient problem commonly encountered in traditional convolutional neural networks.

ResNet-50 is a variant of the ResNet series, consisting of 50 layers [20]. It retains the advantages of residual connections and enhances model performance through deeper network layers. In the semantic segmentation task of surface structural-layer, ResNet-50 effectively extracts detailed features through its deep convolutional layers and residual connections, particularly excelling in processing boundaries and small objects, thereby improving segmentation accuracy. Leveraging these advantages, employing ResNet-50 as the backbone network is expected to significantly enhance the segmentation accuracy of complex boundaries in surface structural-layer during construction, improving the model's performance in handling details and interference factors.

4.2. ASPP Module and Its Improvements

The key to identifying surface structural-layer during construction lies in the high-precision semantic segmentation of their boundaries. To further enhance the segmentation accuracy, attention mechanism modules are integrated into the DeepLabV3+ baseline model. Attention mechanisms help the model focus on important features while suppressing irrelevant ones, thereby improving its representational capacity. Common attention mechanisms include SE, ECA, CA, and CBAM [19]. Among these, CBAM is a lightweight and efficient mechanism that enhances model performance in complex image scenarios by optimizing feature map representations in convolutional neural networks. CBAM consists of two sequential modules: the Channel Attention Module (CAM) and the Spatial Attention Module (SAM), which focus on extracting key features and important regions within feature maps, respectively [21]. Incorporating CBAM into the DeepLabV3+ model enables precise attention to the boundaries and details of surface structural-layer during construction, thereby being expected to improve the accuracy of semantic segmentation, especially in fine-grained boundary recognition.

For this purpose, this paper retains the original multi-scale atrous convolution and image-level pooling structures in the ASPP module of DeepLabV3+, and introduces the CBAM attention mechanism to further boost the representational power of the feature maps. Specifically, the input image is first processed by five parallel branches:

(1) 1×1 convolution branch, to preserve the original channel information;

(2) Three 3×3 atrous convolution branches (with dilation rates of 6, 8, and 12), to capture context at multiple scales;

(3) Image-level global average pooling branch (Image Pooling), which applies global average pooling over the entire feature map, then a 1×1 convolution for channel compression, and upsamples the result back to the original feature-map size.

Subsequently, the five multi-scale features are then fed into the CBAM module, where CAM and SAM are applied sequentially. Through weight mapping along the channel and spatial dimensions, the model adaptively emphasizes semantically important regions while suppressing irrelevant noise. After attention weighting, the five feature maps are concatenated along the channel dimension, forming a "large" feature map (High-Level Features) that integrates multi-scale information with emphasized key areas. Finally, this feature map is passed through a 1×1 convolution (self. project) to unify the channel count and integrate information, outputting the final high-level semantic features to be passed to the decoder. It is important to note that the five branches operate in parallel, with each branch used to extract features at different scales and undergo attention weighting via CBAM. Although the CBAM modules are the same and perform the same function, their input and output data are different.

5. Experiments and Analysis

To validate the performance of the improved DeepLabV3+ model in semantic segmentation of surface structural-layer from UAV imagery during construction, this paper utilized the afore mentioned self-constructed dataset for training and evaluation.

5.1. Experimental Hardware and Software Environment

The training and evaluation processes were conducted on the AutDL computing cloud platform provided by Vistop Cloud (Nanjing) Technology Co., Ltd. The specific hardware and software environment are as follows:

(1) Software environment

The operating system used in this experiment was Ubuntu 20.04. Ubuntu 20.04, with the deep learning framework PyTorch 1.10.0, Python version 3.8, and CUDA version 11.3.

(2) Hardware configuration

The system used in this experiment was equipped with an NVIDIA GeForce RTX 4090D GPU, featuring 14,592 CUDA cores and 24 GB of GDDR6X memory, delivering exceptional computational power for large-scale deep learning model training and inference. The CPU was an Intel Xeon Platinum 8474C processor, operating at a base frequency of 2.1 GHz with a turbo boost up to 3.8 GHz, comprising 48 cores and 96 threads, facilitating efficient multi-threaded computations for data preprocessing tasks. The system was further supported by 80 GB of RAM, ensuring stable operation during extensive dataset training. Storage configuration included a 30 GB system disk and a 50 GB data disk, providing ample space for storing training data and model checkpoints.

5.2. Model Performance Evaluation Metrics

This paper utilized the improved DeepLabV3+ model to train and evaluate performance on a custom semantic segmentation dataset of surface structural-layer captured during the construction phase of the Guangxi Baqiang Road project.

The evaluation metrics for model performance include overall accuracy, mean accuracy, frequency-weighted accuracy, and mean intersection over union. Overall Accuracy (Overall Acc) represents the ratio of correctly classified pixels to the total number of pixels, providing an assessment of the model's overall classification performance across all categories. Mean Accuracy (Mean Acc) is the average of the accuracies for each class, calculated as the ratio of correctly classified pixels to the total number of pixels in that class. Frequency-Weighted Accuracy (FreqW Acc) is the weighted average of class accuracies, with weights corresponding to the frequency of each class in the dataset, thereby accounting for class distribution differences and giving higher weight to more frequent classes. Mean Intersection over Union (Mean IoU) is a critical metric for evaluating segmentation performance, indicating the ratio of the intersection to the union of the predicted and ground truth regions for each class. Mean IoU is the average of the IoU values across all classes.

These metrics are commonly used in combination to comprehensively assess the performance of semantic segmentation models, considering the model's processing capability across different classes and its overall performance.

5.3. Experimental Results and Analysis

To evaluate the performance of the improved DeepLabV3+ model, training experiments on both the baseline DeepLabV3+ model and the enhanced version was conducted. An ablation study was performed to compare the effects of the improvements. The optimal model parameters were assessed using four evaluation metrics: Mean IoU, Overall Acc, Mean Acc and FreqW Acc. All model training and evaluation were conducted in the same hardware environment, and the corresponding software environment was utilized according to the model configurations. Each model was trained using a custom semantic segmentation dataset of surface structurallayer images collected during the construction phase of the Guangxi Baqiang Road project.
The detailed data are presented in Table 1.

Table 1. Comparison and Performance Analysis of the DeepLabV3+ Baseline Model and the Improved DeepLabV3+ Model.

NO.	Overall Acc	Mean Acc	FreqW Acc	Mean IoU
DeepLabV3+	0.980276	0.978326	0.962549	0.928944
Improved DeepLabV3+	0.986225	0.969868	0.973090	0.947663

According to the data in Table 1, there are differences in the evaluation metrics between the baseline DeepLabV3+ model and the improved DeepLabV3+. Compared with the baseline, the improved DeepLabV3+ achieves a 1.92 percent increase in the primary metric, Mean IoU, indicating more effective performance in fine-grained segmentation and handling of complex boundary regions. In addition, the improved model also shows greater gains in Overall Accuracy and Frequency-Weighted Accuracy, with increases of 0.61 percent and 1.10 percent, respectively. However, Mean Accuracy decreases by 0.86 percent, possibly due to slight drops in classification accuracy in shadowed areas, regions occluded by trees, or small interfering objects. Overall, the improvements across multiple metrics demonstrate the superior performance of the improved DeepLabV3+ in high-precision semantic segmentation tasks.

The visualized images of the test set from the self-constructed semantic segmentation dataset of the surface structural-layer during the construction phase of the Guangxi Baqiang Road project, using the improved DeepLabV3+ model are shown in Figure 3. These include the original images, annotated grayscale images, output result images, and overlay images. Based on the visual analysis of these images, the misclassified pixels in the UAV imagery are predominantly concentrated near the boundaries of the surface structural-layer under construction. These misclassifications are largely influenced by interference factors such as tree obstructions and shadows, while the interference from vehicles with relatively clear boundaries is comparatively minimal.

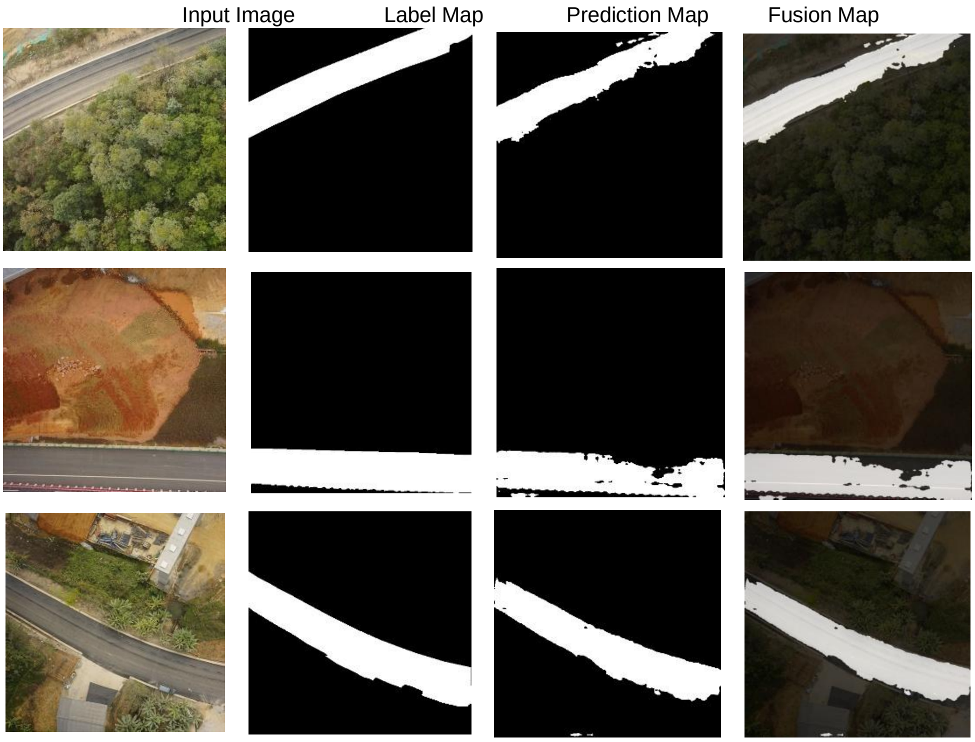

Fig. 3. Semantic segmentation results of surface structural-layer during construction, achieved using the improved DeepLabV3+ model.

6. Discussion and Conclusions

This paper addresses the demand for high-precision identification of surface structural-layer in road construction. An improved DeepLabV3+ model was proposed for ow-altitude UAV imagery semantic segmentation, to provide an intelligent solution for monitoring and evaluating construction progress and quality. Compared to traditional manual inspection approaches, lowaltitude UAV imagery offers advantages such as high efficiency, low cost, and wide-area coverage. When combined with deep learning algorithms, it significantly enhances the efficiency and accuracy of information acquisition and processing during the construction phase, thereby expanding the application scope of intelligent construction technologies. In terms of model design, this paper introduces two key improvements to the DeepLabV3+ model. First, the backbone network is replaced with ResNet-50 to enhance the model’s ability to extract finegrained features and to reduce the risk of overfitting. Second, the CBAM attention mechanism is integrated into the ASPP module to improve the model’s focus on boundaries and critical regions, thereby enhancing the overall performance of semantic segmentation.

Through comparative experimental analysis, selecting DeepLabV3+ as the baseline model is deemed reasonable, and incorporating the CBAM module into the ASPP section effectively enhances the model's performance. In the improved DeepLabV3+ model, the key performance metrics—Mean IoU, Overall Acc, and FreqW Acc are increased by 1.92 percent, 0.61 percent, and 1.10 percent, respectively.

In conclusion, the proposed method has achieved high-precision semantic segmentation of surface structural-layer during construction, thereby providing data support for intelligent construction management.

However, this method has currently conducted a small-scale experiment only on the construction-phase surface structural-layer imagery from the Guangxi Baqiang Road project. Future work will focus on optimizing data acquisition strategies to expand the scale of highquality datasets, thereby further enhancing the performance of the improved DeepLabV3+ model and providing foundational support for the monitoring and assessment of constructionphase surface structural-layer progress and quality.

Acknowledgements

This research was supported by the "Key Technologies and Applications for Fast, Accurate, and Optimal Decision-Making and Budgeting in Highway Construction Based on Low-Altitude UAVs," a project funded by the Guangxi Scientific Research and Technology Development Program (No. Guike AB25069267).

References

[1]. Ministry of Transport of the People's Republic of China, "Inspection and Evaluation Quality Standards for Highway Engineering Section 1 Civil Engineering," JTG F80/1-2017, China Communications Press, Beijing, 2017.
[2]. G. RenQiang, Z. Xianfeng, S. Quan,S. Min, Z. Qingzhan, “Research on road pavement monitoring and smoothness evaluation methods based on UAV lidar data,” Journal of Basic Science and Engineering, 2018, pp. 681-696, 2018, doi:10.16058/i.issn.1005-0930.2018.04.001.
[3]. Z. Chendong, “Research on safety-quality-progress monitoring technology of asphalt pavement construction based on UAV,” School of Transportation Southeast University, 2023, doi: 10.27014/d.cnki.gdnau.2023.001 870.
[4]. K. Michat, R. Szymon and W. Stawomir, “Image analysis in autonomous vehicles: a review of the latestai solutions and their comparison,” Applied Sciences, 2024, doi.org/10.3390/ app14188150.
[5]. H. Dongyeob, L. Suk Bae, S. Mihwa and C. Jun Sang, “Change detection in unmanned aerial vehicle images forprogress monitoring of road construction,” Buildings, 2021, doi.org/10.3390/ buildings 11040150.
[6]. L. Kicheol, P. Jeongjun and H. Gigwon, “Prediction of water content in subgrade soil in road construction using hyperspectral information obtainedthrough UAV,” Applied Sciences, 2024, doi.org/10.3390/ app 14031248.
[7]. Z. Chendong, Z. Junging, B. Tianxiang and G. Xiaofei, “Factors via UAV monitoring and identification of road construction safety,” Sensors, 2022 , doi.org/10.3390/s22228797.
[8]. D. Yigeng, X. Jianhua, P. Liang, W. Xian , D. Zihao and C. Jiabin, “A vision based road width measurement method,” Journal of Chinese Inertial Technology, pp. 218-223, 2022, doi.10.13695/i.cnki. 12-1222/03.2022. 02. 012.
[9]. M. Yan and G. Kezierbieke, “Research review of image semantic segmentation method in high-resolution remote sensing image interpretation,” Journal of Frontiers of Computer Science and Technology, pp. 15261548, 2023, doi: 10.3778/i.issn.1673-9418.2211015.
[10].H. Wenke,T. Fei,W. Zidan and F. Li, “Image Segmentation based on deep learning: a survey,” Computer Science, pp. 107-116, 2023, doi: 10.11896/isikx.230900002.
[11].L. Rongrong and H. Dongzhi, “Semantic segmentation based on DeepLabV3+ and attention mechanism,” 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference, 2021, doi: 10.1109/IMCEC51613.2021.9482207.
[12].Y. Jiahuan, “Road engineering,” Tsinghua University Press, Beijing, 2023, ISBN 978-7-302-59945-6.
[13].W. Kuangda, “Road engineering,” China Building Industry Press, Beijing, 2012, ISBN 978-7-112-14156-2.
[14].C. Liang-Chieh, P. George, K. lasonas, M. Kevin, and Y. Alan L., “Semantic image segmentation with deep convolutional nets and fully connected CRFs,” arXiv:1412.7062, 2016, doi.org/ 10. 48550/ arXiv.1412.7062
[15].C. Liang-Chieh, P. George, K. lasonas, M. Kevin, and Y. Alan L., “DeepLab: semantic Image segmentation with deep convolutional nets, atrous convolution and fully connected CRFs,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 834-848, 2018, doi: 10.1109/TPAMI. 2017.2699184.
[16].C. Liang-Chieh, P. George, S. Florian and A. Hartwig, “Rethinking atrous convolution for semantic Image segmentation,” arXiv:1706.05587, 2017, doi: 10.48550/arXiv.1706.05587.
[17].C. Liang-Chieh, Z. Yukun, P. George, S. Florian and A. Hartwig, “Encoder-Decoder with atrous separable convolution for semantic image segmentation,” arXiv:1802.02611, 2018, doi: 10.48550/ arXiv.1802.02611.
[18].VainF, "DeepLabV3Plus-Pytorch," GitHub, 2020. Available: https://github.com/VainF/DeepLabV3Plus-Pytorch.
[19].G. Jinmei, Z. W. Nurshazwani Wan Zakaria, W. Bisheng and M.Azmi Bin Ayub, “DeeplabV3+ model with CBAM and CSPM attention mechanism for navel orange defects segmentation,” International Journal of Advanced Computer Science and Applications, pp. 199-207, 2024.
[20].W. Sanghyun, P. Jongchan,L. Joon-Young and K. In So, “CBAM: convolutional block attention module,” 15th European Conference on Computer Vision, pp. 3-19, 2018, doi: 10.1007/978-3-030-01234-21.