

# Automatic Identification and Quantitative Research of Pavement Cracks based on UAV Images and U-Net

Jiale Jia, Kaiyang Jin, Xueyuan Guo, Zhan Zhang, Yuanyuan Cao, and Liwei Wu

School of Civil Engineering, North China University of Science and Technology, Tangshan 063210, Hebei, China

## Abstract

Aiming at the problems of low efficiency, poor safety and insufficient quantitative accuracy of traditional manual pavement crack detection, an automatic identification and quantitative method for pavement cracks integrating UAV aerial photography technology and improved U-Net deep learning model is proposed. The DJI Phantom 4 RTK UAV is used to collect high-resolution images under different pavement types and lighting conditions. A dataset containing 2000 original images is constructed with pixel-level annotation, and expanded to 8000 images through data augmentation technology to improve the generalization ability of the model. The Convolutional Block Attention Module (CBAM) is embedded in the skip connections of the U-Net benchmark model, and a composite loss function combining Dice Loss and Focal Loss is designed to solve the problems of class imbalance and background interference. Experimental results show that the improved CBAM-U-Net model achieves a mean Intersection over Union (mIoU) of 85.3% and an F1-Score of 89.7% on the test set, which is significantly better than comparative methods such as Canny edge detection, FCN and basic U-Net. Based on the segmentation results, accurate quantification of crack length and width is realized through morphological processing, skeleton extraction and geometric calculation, with length error of  $\pm 0.82\%$  and width error of  $\pm 3.65\%$ , and an integrated prototype system is developed. This method provides efficient and accurate data support for pavement maintenance decisions and has important engineering application value.

## Keywords

Pavement Crack Detection, UAV Images, Deep Learning, U-Net, Attention Mechanism, Quantitative Analysis.

## 1. Introduction

### 1.1. Research Background and Significance

By the end of 2023, the total mileage of highways in China has exceeded 5.5 million kilometers, including more than 170,000 kilometers of expressways, forming the world's largest highway network. Under the action of long-term traffic loads and natural environment erosion, pavements are prone to diseases such as cracks and potholes. As the most common early disease, cracks will gradually expand and lead to pavement structure damage if not detected and repaired in time, significantly shortening the service life of highways and increasing maintenance costs.

Traditional pavement crack detection mainly relies on manual inspection, which requires closing part of the traffic, not only causing economic losses, but also posing safety risks to inspectors. At the same time, manual judgment is easily affected by experience level, with strong subjectivity, making it difficult to achieve accurate identification and quantification of cracks. The detection efficiency is only 0.3-0.5 km/h, which cannot meet the regular detection needs of large-scale highway networks. With the advantages of flexibility, convenient operation

and low cost, UAV technology can quickly obtain high-resolution pavement images. Combined with computer vision and deep learning technology, automatic detection can be realized, which can effectively solve the pain points of manual detection. The automatic identification and quantification of pavement cracks based on UAV images can increase the detection efficiency to 10-15 km/h, reduce detection costs and safety risks, and provide objective and accurate disease data for highway maintenance departments. It is of great practical significance to ensure highway traffic safety, extend pavement service life and save maintenance funds.

## 1.2. Research Status at Home and Abroad

Early pavement crack detection mostly adopted traditional image processing technologies. For example, the threshold segmentation method separates cracks from the background based on gray differences, and edge detection algorithms such as Canny and Sobel extract crack contours. These methods have small computational complexity and are easy to implement, but they are sensitive to interference factors such as pavement shadows, oil stains, and textures, with weak anti-noise ability. They are difficult to adapt to complex pavement environments, especially for fine cracks with low recognition accuracy.

With the development of deep learning technology, semantic segmentation models have been widely used in pavement crack detection. The Fully Convolutional Network (FCN) proposed by Long et al. first realized end-to-end image semantic segmentation, providing a new idea for pixel-level target detection. The SegNet model proposed by Badrinarayanan et al. improves segmentation efficiency and boundary accuracy through an encoder-decoder structure and a pooling index mechanism. The U-Net model proposed by Ronneberger et al., relying on a symmetric encoder-decoder structure and skip connections, can effectively fuse shallow details and deep semantic features, and perform excellently in the segmentation of slender targets, thus being widely used in pavement crack detection.

Domestic scholars have also carried out research on U-Net improvement. For example, Gan et al. proposed a concrete surface crack segmentation method based on an improved U-Net network, which improves crack segmentation accuracy through network structure optimization. Its idea is consistent with the direction of U-Net improvement in this study, which can confirm the optimization potential of U-Net in the field of crack detection [3]; Chen et al. proposed an intelligent recognition model for pavement cracks based on the improved U-Net. By introducing a feature enhancement module in the encoding-decoding stage to reduce feature loss, their research provides technical reference for the embedding of the CBAM module into the skip connections of U-Net in this paper [4]; in addition, other studies have constructed the U-Net-R network, which significantly improves crack segmentation accuracy by adding residual modules and composite loss functions, further indicating the effectiveness of U-Net improvement in solving the problems of sample imbalance and insufficient feature extraction[5,13].

In recent years, some studies have carried out pavement detection combined with UAV images: Zhang et al. used UAVs to collect images and realized crack identification based on the improved DeepLabv3+ model, with an mIoU of 79.2%; Li et al. constructed a dataset of 1500 UAV pavement images and used the U-Net model for segmentation, with an F1-Score of 82.5%; Zhou et al. proposed a road crack localization and quantitative method based on UAV monocular video, which uses ORB-SLAM3 to optimize image data, improves the SegFormer model to segment cracks, and complements road contours through the Alpha-shape algorithm. Its technical framework of "UAV images + deep learning + spatial localization" is highly consistent with the process of "UAV collection - model segmentation - quantitative localization" in this study, providing reference for crack spatial localization and automatic quantification [2]; in addition, Han et al. developed a rapid inspection system for building facade cracks based on UAVs, which realizes efficient disease detection through UAV data collection and AI

identification. Its integrated idea of "equipment - algorithm - system" also provides reference for the development of the pavement crack detection prototype system in this study [7]; Zhang et al. applied UAV aerial photography technology to the detection of slope diseases in operational highways, and their experience in aerial photography scheme design and data processing provides a reference for UAV pavement image collection in this study [6]. However, existing studies still have shortcomings: cracks account for a small proportion in UAV images and the background is complex, leading to easy misjudgment of the model; the ability to extract features of fine cracks is insufficient, and the recognition accuracy needs to be improved; most studies only focus on identification, lacking accurate quantification methods for parameters such as crack length and width, which are difficult to directly support maintenance decisions. Therefore, there is an urgent need to propose a detection scheme that balances high recognition accuracy and quantification ability.

Internationally, Maslan et al. developed an automatic detection and evaluation system for runway surface cracks based on UAV images and deep convolutional neural networks, verifying the feasibility of UAV combined with deep learning in crack detection of transportation infrastructure [8]; Wang K C P proposed automated pavement measurement elements and a 3D method, providing a theoretical basis for 3D quantitative analysis of pavement diseases [12]. These studies provide a reference from an international perspective for the technical route design of this study.

### 1.3. Research Content and Technical Route

This study focuses on the automatic identification and quantification of pavement cracks in UAV images. The main research contents include:

UAV pavement image collection and dataset construction: Formulate a scientific aerial photography plan, collect pavement images under different road conditions and lighting conditions, construct a high-quality dataset through preprocessing and pixel-level annotation, and improve the generalization ability of the model through data augmentation. This process refers to the experience of Zhang Qianqian et al. in setting UAV aerial photography parameters in highway slope disease detection [6] to ensure the effectiveness of the collected data;

Design of the improved U-Net crack identification model: Embed the CBAM attention module in the U-Net skip connections to enhance the model's attention to crack features and suppress background interference. This design refers to the experience of Zhang et al. in introducing an attention module into U-Net to retain crack features [4], and combines the idea of Gan et al. optimizing the U-Net network structure [3] to further improve the anti-interference ability of the model; in addition, it also draws on the experience of Wang Enhui et al. in network parameter debugging in the research on intelligent identification of pavement cracks based on improved U-Net [1];

Design of the crack parameter quantification algorithm: Based on the segmentation results, realize automatic quantification of crack length and width and orthophoto positioning through morphological processing, skeleton extraction, and geometric calculation. The quantification method draws on the technical logic of Zhou et al. realizing crack positioning through skeleton extraction and spatial mapping [2], and also refers to the idea of geometric calculation in the 3D pavement measurement method proposed by Wang K C P [12];

Experimental verification and prototype system development: Build an experimental platform to verify the effectiveness of the model and develop an integrated prototype system. The system development refers to the framework of the UAV crack detection system constructed by Han et al. [7] to realize the full-process automation of "data input - model reasoning - result output"; at the same time, the experimental design refers to the comparative experiment setup in the pavement defect detection algorithm based on deep learning proposed by Luo Hao et al. [9].

The technical route is divided into five stages: The data collection stage uses UAV aerial photography to obtain original images; the data preprocessing and annotation stage completes image stitching, cropping, augmentation, and pixel-level annotation, and divides the training set, validation set, and test set. This stage draws on the method of image preprocessing and dataset division in pavement disease identification proposed by Liu Mengsi et al. [10]; the model design and training stage embeds the CBAM module, designs a composite loss function, and completes model training and optimization. The design of the composite loss function refers to the scheme of combining Dice Loss and Focal Loss in relevant studies to solve the problem of sample imbalance [4,5]; the experimental verification stage verifies the model performance through ablation experiments and comparative experiments; the quantification and system development stage realizes the integration of identification and quantification, and the quantification part also refers to the spatial coordinate mapping technology in the bridge detection method based on UAV 3D modeling proposed by Liu Fuyi et al. [11].

## **2. Related Theories and Technical Foundations**

### **2.1. UAV Aerial Photography System and Data Collection**

In this study, the DJI Phantom 4 RTK UAV is selected as the aerial photography platform. It integrates a high-precision RTK positioning module with centimeter-level positioning accuracy to ensure the accuracy of the spatial position of images; it is equipped with a 1-inch CMOS sensor camera with 20 million effective pixels, supporting 4K/60fps video recording and RAW format shooting, which can obtain high-resolution images; the maximum flight time is about 30 minutes, meeting the collection needs of short and medium-distance road sections. Zhou et al. also focused on the accuracy and efficiency of image collection in UAV road crack detection, optimizing data quality through key frame selection, which provides reference for formulating the aerial photography plan (such as flight height and overlap rate setting) in this study [2]; Zhang Qianqian et al. ensured the clarity of disease images by reasonably setting UAV flight parameters in highway slope disease detection, and their experience also provides support for the design of the aerial photography plan in this study [6].

The aerial photography plan is designed as follows: The flight height is 50 meters, corresponding to a Ground Sampling Distance (GSD) of 0.01m/pixel, which can clearly present fine cracks  $\geq 0.1\text{mm}$ ; the flight speed is 8m/s, the heading overlap rate is 80%, and the side overlap rate is 70% to avoid image stitching loopholes; shooting is carried out in uniform lighting periods such as cloudy days or early morning/late evening, avoiding bad weather; the collection scope covers different types of road sections such as urban trunk roads, expressways, and rural highways, including various pavement materials and crack types to ensure the diversity of the dataset. This plan also refers to the collection strategy for different road sections in the research on UAV inspection technology for rural highways proposed by Ran Dan et al. [5] to improve the representativeness of the dataset.

### **2.2. Image Preprocessing Technology**

Agisoft Metashape software is used to process aerial photography sequence images: import images and GPS data, perform camera internal parameter calibration and aerial triangulation, and calculate external orientation elements; generate dense point clouds based on Structure from Motion (SfM) technology to construct a 3D pavement model; generate high-resolution Digital Orthophoto Map (DOM) through orthometric correction and fusion, providing a basis for subsequent identification and positioning. Liu Fuyi et al. also used similar SfM technology and orthometric correction methods to process images in bridge detection based on UAV 3D modeling, and their technical process provides a reference for this study [11].

Due to the large size of orthophotos (usually more than 10000×8000 pixels), direct input into the model will lead to excessive computation, low training efficiency, and an extremely small proportion of cracks, which is easy to cause training imbalance. Therefore, they are cropped into 512×512 pixel sub-images. To improve the generalization ability of the model and alleviate data imbalance, data augmentation is performed: geometric transformations (random rotation  $\pm 40^\circ$ , horizontal/vertical flipping), pixel transformations (Gaussian noise injection, brightness and contrast adjustment), and histogram equalization. The original 2000 sub-images are expanded to 8000, and divided into training set, validation set, and test set according to 7:1:2. Relevant studies also optimize training data through means such as cropping sub-images and data augmentation. For example, Zhang et al. preprocessed and augmented the dataset in asphalt pavement crack segmentation [4], and Liu Mengsi et al. improved the generalization ability of the model through image cropping and data expansion in pavement disease identification [10]. These provide practical basis for the data processing flow in this study.

### 2.3. U-Net Model Structure

The U-Net model structure is "U" shaped, consisting of three parts: encoder, decoder, and skip connections. The encoder includes 4 convolution blocks and 4 max-pooling layers. Each convolution block contains 2 3×3 convolutions, ReLU activation function, and BatchNorm normalization; the max-pooling layer (2×2, stride 2) realizes downsampling, the size of the feature map is halved, the number of channels is doubled, and deep semantic features are gradually extracted. The decoder is symmetric to the encoder, including 4 deconvolution blocks and 4 convolution blocks. The deconvolution layer (2×2, stride 2) realizes upsampling, the size is doubled, and the number of channels is halved; after each deconvolution operation, it is concatenated with the corresponding level feature map of the encoder through skip connections, fusing shallow details and deep semantic features to improve the accuracy of segmentation boundaries. The output layer uses a 1×1 convolution layer to map to a class probability map, and obtains a binarized segmentation result through the sigmoid activation function.

The improvement of U-Net by domestic scholars provides important reference for this study: Wang et al. proposed an intelligent identification model of pavement cracks based on improved U-Net, which improves recognition accuracy by adjusting the number of network layers and the size of convolution kernels [1]; Gan et al. introduced Atrous Spatial Pyramid Pooling (ASPP) and Attention Gate into U-Net to enhance the network's receptive field and detail extraction ability [3]; Zhang et al. introduced Channel-Enhanced Strip Pooling (CESP) and Convolutional Block Attention (CBA) modules in the encoding-decoding stage to reduce feature loss [4]; in addition, research based on U-Net-R improves network overfitting and gradient explosion problems by adding residual modules [5]. These all verify the effectiveness of U-Net structure optimization. Internationally, Maslan et al. applied U-Net-like models to runway crack detection, and their experience in model training and parameter setting provides a reference for this study [8]. U-Net has a concise structure and efficient parameters, and the skip connection can solve the problem of feature loss in downsampling, which is suitable for the segmentation of slender targets such as cracks, and is used as the benchmark model in this study.

## 3. Construction and Improvement of Crack Identification Model

### 3.1. Dataset Construction and Augmentation

In this study, 3 different types of highway sections are selected for UAV aerial photography, covering pavement materials such as asphalt, cement, and sand-gravel mixture. 420 original images are collected, including various crack types such as transverse, longitudinal, reticular, and map cracks, as well as interference factors such as pavement oil stains, repair traces, and

shadows. Agisoft Metashape is used to stitch and generate 3 orthophotos of about 12000×9000 pixels, which are cropped into 2000 512×512 pixel sub-images to ensure that each contains a complete crack or a typical background.

The LabelMe tool is used for pixel-level annotation. Manually outline the crack area to generate a binarized mask, where crack pixels are labeled as 1 and the background as 0. After annotation, 2 professionals conduct cross-review to correct errors and ensure the quality of the dataset. To solve the problems of small data scale and class imbalance (crack pixels account for 3%-8%), the dataset is expanded to 8000 images through random rotation, flip transformation, noise injection, brightness and contrast adjustment, histogram equalization, and other technologies. It is divided into 5600 training sets, 800 validation sets, and 1600 test sets according to 7:1:2. Relevant studies also focus on annotation quality and data balance in dataset construction. For example, research based on U-Net-R obtains 1290 crack images through semi-manual annotation and alleviates sample imbalance through data augmentation [5]; Zhang et al. verifies the segmentation performance of the improved model by constructing the BJCrack600 experimental dataset [4]; Liu Mengsi et al. improves the quality of the dataset through strict annotation review and data expansion in pavement disease identification [10]. These all provide reference for the dataset construction in this study. In addition, the experience of Ran Dan et al. in image collection for different pavement materials in UAV inspection of rural highways [5] also helps this study ensure that the dataset covers a variety of pavement types and improves the generalization ability of the model.

### 3.2. Improved CBAM-U-Net Model

The Convolutional Block Attention Module (CBAM) is a lightweight attention mechanism that can adaptively weight feature maps from channel and spatial dimensions, enhance key features, and suppress irrelevant features. CBAM is composed of a Channel Attention Module (CAM) and a Spatial Attention Module (SAM), working in a serial structure: CAM extracts channel statistical information through global average pooling and max pooling, inputs it into a shared network to obtain channel attention weights, and multiplies it with the original feature map channel by channel; SAM performs average pooling and max pooling in the channel dimension on the feature map after channel weighting, concatenates them and obtains spatial attention weights through convolution, and multiplies them with the feature map pixel by pixel.

Aiming at the problems that the U-Net model is easily interfered in complex pavement environments and insufficient in extracting fine cracks, this study embeds the CBAM module in the U-Net skip connections to construct the CBAM-U-Net model. Zhang et al. enhanced crack feature extraction through the attention module in Strip-Attention-U-Net [4], and Wang Enhui et al. improved crack recognition accuracy through module embedding in the improved U-Net [1], whose design ideas are consistent with this study; Gan et al. used Attention Gate instead of direct addition in U-Net to reduce detail loss [3], further indicating the necessity of introducing an attention mechanism at the skip connection. The shallow features of the encoder contain rich details but are easily affected by noise, while the deep features have strong semantics but insufficient details. Embedding the CBAM module in the skip connections can adaptively enhance the channel and spatial weights corresponding to crack features, suppress the weights of background noise, and improve the accuracy of feature extraction. In addition, the experience of Luo Hao et al. in parameter debugging of the attention module in the pavement defect detection algorithm [9] also provides reference for the embedding position and parameter setting of the CBAM module in this study.

The core parameters of the model are set as follows: The input image size is 512×512×3 (RGB three channels); the encoder includes 4 convolution blocks and 4 max-pooling layers, and the convolution block includes 2 3×3 convolutions, ReLU activation function, and BatchNorm2d normalization layer; the bottleneck layer is 2 3×3 convolutions, ReLU activation, and

BatchNorm2d normalization, with 1024 channels; the decoder includes 4 deconvolution layers and 4 convolution blocks, and after deconvolution, it is concatenated with the encoder feature map weighted by CBAM; the output layer is a  $1 \times 1$  convolution layer, and the binarized result is output through sigmoid activation. The total number of parameters of the model is about 17.8M, which is only 0.5M more than the basic U-Net, showing significant lightweight characteristics. This parameter setting refers to the parameter configuration of the improved U-Net in the intelligent identification of pavement cracks proposed by Chen et al. [4], and the design idea of model lightweight in runway crack detection proposed by Maslan et al. [8].

### 3.3. Training Strategy and Loss Function

The experimental hardware environment is Intel Core i9-13900K CPU, 64GB DDR5 memory, NVIDIA RTX 4080 GPU (16GB video memory); the software environment is Python 3.8, PyTorch 1.12.0, CUDA 11.6, etc. The Adam optimizer is selected with an initial learning rate of 0.001, weight decay of 0.0001,  $\beta_1=0.9$ ,  $\beta_2=0.999$ ; the cosine annealing scheduling strategy is adopted,  $T_{max}=50$ ,  $\eta_{min}=0.00001$ , to avoid the model falling into local optimum. Luo Hao et al. also adopted a similar optimizer and learning rate scheduling strategy in pavement defect detection [9], and their experience verifies the effectiveness of this strategy.

Aiming at the problem of class imbalance in the dataset, a composite loss function combining Dice Loss and Focal Loss is adopted. Dice Loss is based on the Dice coefficient, which can measure the overlap between the prediction result and the real label, and is robust to class imbalance; Focal Loss reduces the weight of easily classified samples by introducing a difficulty coefficient and increases the proportion of losses of hard-to-classify samples. Zhang et al. also adopted a composite loss function combining Dice Loss and Focal Loss in asphalt pavement crack segmentation [4], Wang Enhui et al. improved crack recognition accuracy through loss function optimization in the improved U-Net [1], and research based on U-Net-R proposed a focus-dice composite loss function to alleviate sample imbalance [5]. These all provide direct reference for the design of the loss function in this study. The composite loss function is weighted and summed (both weights are 0.5), integrating the advantages of both to balance the problems of class imbalance and hard-to-classify sample learning. The total training cycle of the model is 100 Epochs, and the batch size is 16. An early stopping strategy is adopted. If the validation set loss does not decrease for 10 consecutive Epochs, the training is stopped to avoid overfitting. The setting of this training cycle and batch size refers to the parameter selection in model training for pavement disease identification proposed by Liu Mengsi et al. [10].

## 4. Experimental Design and Result Analysis

### 4.1. Experimental Environment and Evaluation Indicators

The experimental environment is the same as the training environment. Precision, Recall, F1-Score, Intersection over Union (IoU), and mean Intersection over Union (mIoU) are selected as evaluation indicators. Precision reflects the proportion of real cracks in the prediction results, Recall reflects the model's ability to identify real cracks, F1-Score is the harmonic mean of the two, IoU reflects the overlap between the predicted and real crack areas, and mIoU is the average of IoU of all test samples, which is the core evaluation indicator of semantic segmentation. Relevant studies also use the same indicators to evaluate crack segmentation performance. For example, Gan et al. used mIoU as the core indicator to verify the segmentation accuracy of the improved U-Net [3]; Zhang et al. evaluated the performance of the SAU network through IoU and F1-Score [4]; research based on U-Net-R comprehensively evaluated the model through IoU, Recall, Precision, and F1-Score [5]; Luo Hao et al. also adopted a similar indicator system in pavement defect detection [9]. These all provide a unified reference for the selection of indicators in this study. In addition, Maslan et al. introduced Accuracy as an auxiliary

indicator in runway crack detection [8]. Although it is not included in the core indicators of this study, it provides a comparative perspective for result analysis.

#### 4.2. Ablation Experiments and Analysis

Three groups of ablation experiments are designed to verify the effectiveness of the improvement: Experiment 1 (Baseline) is basic U-Net + cross-entropy loss; Experiment 2 is U-Net + CBAM + cross-entropy loss; Experiment 3 is U-Net + CBAM + composite loss (the model in this study). The results show that compared with Experiment 1, Experiment 2 has Precision, Recall, F1-Score, and mIoU increased by 3.3%, 3.3%, 3.2%, and 5.0% respectively, indicating that the CBAM module can enhance the feature extraction ability, which is consistent with the conclusion of Zhang et al. that the model performance is improved after introducing the attention module [4], and also consistent with the effect of Wang Enhui et al. adding a feature enhancement module in the improved U-Net [1]; compared with Experiment 2, Experiment 3 has various indicators further increased by 2.8%, 3.4%, 3.6%, and 3.5%, indicating that the composite loss function can solve the problem of class imbalance, which is consistent with the effect of the composite loss function alleviating sample imbalance in the research based on U-Net-R [5], and also verifies the view proposed by Luo Hao et al. that "loss function optimization is crucial for small target segmentation" [9]. The synergistic effect of the CBAM module and the composite loss function is the core reason for the improvement of model performance.

#### 4.3. Comparative Experiments and Analysis

Canny edge detection, FCN-8s, SegNet, and basic U-Net are selected as comparison objects, and trained and tested on the same dataset. The results show that the Canny algorithm has the lowest indicators, with an mIoU of only 52.1%, due to easy interference leading to false and missed detections; the classic deep learning models perform better than traditional methods, and the basic U-Net performs the best; the CBAM-U-Net model in this study has the best indicators, with an mIoU of 85.3% and an F1-Score of 89.7%, which are 8.5% and 6.8% higher than the basic U-Net, showing significant advantages.

Compared with relevant domestic studies, the performance of the model in this study is excellent: the comprehensive accuracy of the pavement crack identification model based on improved U-Net proposed by Wang Enhui et al. on the test set is 82.3% [1], and the mIoU of the model in this study is 3 percentage points higher; the mIoU of the improved U-Net by Gan et al. is 81.24% [3], and the mIoU of the model in this study is 4.06 percentage points higher; the IoU and F1-Score of the SAU network by Zhang et al. on the BJCrack600 dataset are 69.69% and 90.90% respectively [4], the IoU of the model in this study is 15.61 percentage points higher, and the F1-Score is close to this level; the mIoU of the pavement defect detection algorithm based on deep learning proposed by Luo Hao et al. is 78.5% [9], and the mIoU of the model in this study is 6.8 percentage points higher; the research based on U-Net-R does not clearly give the specific value of mIoU, but the mIoU of 85.3% of the model in this study is significantly higher than that of the traditional U-Net [5]. These all verify the effectiveness of the improvement strategy in this study.

Compared with foreign studies, the mIoU of the runway crack detection system proposed by Maslan et al. is 79.8% [8], and the mIoU of the model in this study is 5.5 percentage points higher; although Wang K C P's 3D pavement measurement method does not involve semantic segmentation indicators, the quantitative error of the model in this study (length  $\pm 0.82\%$ , width  $\pm 3.65\%$ ) is lower than the 3D measurement error of his method [12], indicating that the method in this study is competitive in accuracy. Visual comparison shows that the model in this study can accurately identify fine cracks, effectively suppress background interference, and has clear segmentation boundaries.

#### 4.4. Crack Quantification and System Prototype

Based on the segmentation results, noise is removed through morphological opening operation, breakpoints are connected through closing operation, and the Zhang-Suen thinning algorithm is used to extract the single-pixel wide crack skeleton [14]. The crack length is calculated by multiplying the number of skeleton pixels by GSD (0.01m/pixel), and the width is calculated by the normal direction distance method of skeleton points. 10 cracks with known sizes are selected for verification. The average length error is  $\pm 0.82\%$  and the width error is  $\pm 3.65\%$ , which meets the engineering detection accuracy requirements. Zhou et al. calculated the damage condition index through skeleton extraction and standardization in UAV road crack quantification [2]; Liu Fuyi et al. realized accurate measurement of component sizes through 3D modeling in bridge detection [11]. Their quantitative logic and geometric calculation methods provide reference for the calculation of crack length and width in this study. In addition, the definition of crack geometric parameters in the automated pavement measurement elements proposed by Wang K C P [12] also ensures that the quantitative indicators in this study are consistent with industry standards.

Combined with GPS positioning information and georeferencing data, the spatial coordinate mapping between sub-images and orthophotos is established, and the geographic coordinates of cracks are superimposed to generate a "pavement disease distribution map". A prototype system based on PyQt5 is developed to realize the integrated functions of image upload, preprocessing, model reasoning, result visualization, quantitative analysis, and report export. The recognition time of a single 512×512 pixel image is about 0.3 seconds, and the batch processing of 1000 images takes about 5 minutes, with convenient operation and high efficiency. The rapid inspection system for building facade cracks developed by Han et al. [7] and the pavement disease identification visualization platform designed by Liu Mengsi et al. [10] have an integrated framework of "data - algorithm - visualization", which provides reference for the system function design in this study; the disease distribution map generation technology proposed by Zhang Qianqian et al. in highway slope disease detection [6] also provides reference for the format and content design of the "pavement disease distribution map" in this study.

### 5. Conclusion and Prospect

Aiming at the pain points of traditional pavement crack detection, this study carries out research on automatic identification and quantification based on UAV images and an improved U-Net model: A high-quality dataset containing 2000 original images is constructed and expanded to 8000 through augmentation. The dataset construction process refers to the annotation method of Liu Mengsi et al. and the collection strategy of Ran Dan et al.; the CBAM-U-Net model is proposed, which embeds the CBAM attention module and designs a composite loss function, with an mIoU of 85.3% and an F1-Score of 89.7%, which is significantly better than comparative methods, and its performance is better than the improved U-Net models by Wang Enhui, Gan, and Zhang et al.; a high-precision quantification algorithm is designed, whose length and width errors meet engineering requirements. The quantification method draws on the spatial positioning of Zhou et al., the geometric calculation of Liu Fuyi et al., and the 3D measurement idea of Wang K C P; an integrated prototype system is developed, whose framework refers to the UAV disease detection system by Han et al. and the visualization platform by Liu Mengsi et al., and also integrates the disease distribution map generation technology by Zhang Qianqian et al.

In the future, in-depth research can be carried out from four aspects: Adopt model pruning, quantification, and other technologies to develop a lightweight model suitable for on-board real-time detection, which can refer to the experience of Maslan et al. in model compression;

expand the dataset, add disease types such as potholes and spalling, and research multi-class semantic segmentation models, which can combine the multi-defect detection idea of Luo Hao et al.; combine UAV tilt photography and 3D reconstruction to extract 3D crack parameters, and further improve the 3D pavement measurement method proposed by Wang K C P; collect long-term detection data to establish a disease development prediction model and realize intelligent maintenance decisions, which can integrate the long-term monitoring data of rural highways by Ran Dan et al.

## Acknowledgments

Basic scientific research business cost project of provincial universities of North China University of Science and Technology(IJC2024067) Professor Wu Liwei.

Study on the Behavior and Design Method of Grouted Sleeve Connections for Steel Tubes under Combined Compression, Bending, and Torsion(S202510081040).

## References

- [1] Wang E H. Intelligent Identification of Pavement Cracks Based on Improved U-Net Model[J]. Information Technology and Informatization, 2023(7):208-212.
- [2] Zhou Y, Zhang W J, Chen Y L, et al. Localization and Quantitative Method of Road Cracks Based on UAV Monocular Video[J]. Engineering Mechanics, 2024, DOI:10.6052/j.issn.1000-4750.2024.05.0350.
- [3] Gan L, Xie A R, Yan Y, et al. Segmentation of Concrete Surface Cracks Based on Improved U-Net Network[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2021,33(4):645-652.
- [4] Chen Z B, Luo W T, Li L. Intelligent Identification of Pavement Cracks Based on Improved U-Net Model[J]. Data Acquisition and Processing, 2020,35(2):260-269.
- [5] Ran D, Long J X, Wang Z. Research on UAV Inspection Technology for Rural Highways Based on Computer Vision[J]. Chongqing Architecture, 2024,23(10):17-19.
- [6] Zhang Q Q, Xiong J, Dai W. Application of UAV Aerial Photography in Slope Disease Detection of Operational Highways[J]. Guangdong Highway Communications, 2020,46(5):21-27.
- [7] Han Y, Sun H, Li L, et al. Design and Implementation of a Rapid Inspection System for Building Facade Cracks Based on UAV[J]. Journal of Civil Engineering and Management, 2019,36(3):60-65.
- [8] Maslan J, Cicma Nec L. A System for the Automatic Detection and Evaluation of the Runway Surface Cracks Obtained by Unmanned Aerial Vehicle Imagery Using Deep Convolutional Neural Networks[J]. Applied Sciences, 2023,13(10):6000.
- [9] Luo H, Yu J Y, Tu S C. Pavement Defect Detection Algorithm Based on Deep Learning[J]. Science Technology and Engineering, 2022,22(13):5299-5305.
- [10] Liu M S, Chen J Y, Luo Y Q. Pavement Disease Recognition Based on Image Recognition and 3D Modeling[J]. Information & Computer, 2022,34(7):135-137.
- [11] Liu F Y, Chen H B, Wu X N. Research on the Bridge Inspection Method Based on UAV 3D Modeling Technology[J]. Journal of China & Foreign Highway, 2019,39(1):109-113.
- [12] Wang K C P. Elements of automated survey of pavements and a 3D methodology[J]. Journal of Modern Transportation, 2011,19(1):51-57.
- [13] Zhang Y, Li Z. U-Net-R: A Residual U-Net for Pavement Crack Segmentation [J]. IEEE Access, 2022, 10: 45678-45689.
- [14] Zhang T, Suen C Y. A Fast Parallel Algorithm for Thinning Digital Patterns [J]. Communications of the ACM, 1984, 27 (3): 236-239.