Article:

Shan, J., Jiang, W., Huang, Y. orcid.org/0000-0002-1220-6896 et al. (2 more authors) (2024) Unmanned Aerial Vehicle (UAV)-Based Pavement Image Stitching Without Occlusion, Crack Semantic Segmentation, and Quantification. IEEE Transactions on Intelligent Transportation Systems, 25 (11). pp. 17038-17053. ISSN 1524-9050

https://doi.org/10.1109/tits.2024.3424525

Unmanned Aerial Vehicle (UAV)-Based Pavement Image Stitching without Occlusion, Crack Semantic Segmentation and Quantification

Jinhuan Shan, Wei Jiang, Yue Huang, Dongdong Yuan, Yaohan Liu

Abstract—**Unmanned Aerial Vehicle (UAV)-based pavement distress detection offers efficient and safe advantages. However, obstructions from road vehicles and the slender shape of cracks in UAV images challenge accuracy. To address this, this study established specific flight parameters, proposed the Historical Best Matching Image (HBMI) approach for data loss due to obstructions, and created the UAV-Crack500 dataset with 500 finely annotated crack images. Three algorithms with different loss functions were investigated, finding that the U-Net network combined with our Completely Asymmetric Loss (CAL) achieved the best performance, resolving the issue of class imbalance. Morphological analysis of the semantically segmented images provided precise crack morphology features. In complex scenarios, errors in features like crack area, length, mean width, and maximum width remained within 16%. This study establishes a comprehensive UAV-based pavement distress detection system, overcoming obstructions for accurate assessment.**

Index Terms—**Pavement distress, Semantic segmentation, Crack quantification, Image stitching, Unmanned aerial vehicle (UAV)**

I. INTRODUCTION

TIMELY and effective maintenance is a crucial measure for preserving the functionality, safety, and durability of pavement [1], [2], [3]. Precision maintenance decisions rely heavily on regular pavement inspections and accurate pavement condition assessments. Traditional pavement inspection methods primarily involve manual visual surveys and pavement inspection vehicles [4], [5], [6]. Manual visual surveys rely on specialized personnel to observe and measure pavement distress, recording their condition manually. This approach suffers from low efficiency, low accuracy, traffic disruption, and safety concerns. Pavement inspection vehicles, on the other hand, collect road imagery using cameras and then employ manual or machine-based identification methods to detect distress. However, this method is costly and lacks automation. In summary, traditional inspection methods are ill-suited for large-scale and frequent inspections, which are at odds with the requirements of pavement maintenance [7], [8], [9].

To achieve large-scale lightweight pavement inspections, lightweight inspection equipment has been introduced into road inspections, such as vehicle-mounted cameras, mobile robot [10] and unmanned aerial vehicles [11], [12], [13]. Vehicle-mounted cameras conduct road inspections by equipping ordinary vehicles with cameras and GPS positioning devices. This method imposes relatively low demands on vehicles and can collect road image data through routine inspections. Subsequently, artificial intelligence methods are used for distress detection and localization. However, this approach is susceptible to traffic flow and lane changes, making it challenging to establish stable inspection routes. This results in redundant image data and continuously changing perspectives, posing significant difficulties for subsequent quantitative defect analysis. The lightweight mobile robot system is used for real-time detection of road cracks, but when dealing with complex traffic conditions, the robot needs enhanced sensing capabilities to ensure detection accuracy and operational safety. With the continuous improvement of UAV flight control and camera performance, UAVs have found widespread applications in industries such as agriculture, environmental monitoring, transportation, and surveying. Pavement distress detection based on UAVs effectively mitigates the influence of traffic on captured imagery.

In contrast to industries like agriculture, environmental monitoring, and surveying, pavement distress detection using UAVs is susceptible to vehicle obstructions, especially in densely trafficked areas, limiting data collection. According to literature research, there is currently limited research on UAV-based pavement distress detection. This is partly due to UAVs being an emerging technology and the requirements for safe flight altitudes for UAVs. In UAV-captured images, pavement cracks typically appear thin and occupy a small proportion of the overall image, making them vulnerable to being overshadowed by other image details or background noise. Pan et al. [14] employed support vector machines, artificial neural networks, and random forests to classify and detect cracks and potholes in UAV-captured road images, achieving an overall accuracy of 98.3%. Ali et al. [15] utilized CNN for

Manuscript received September 19, 2023; revised February 18, 2024, and June 9, 2024; accepted July 4, 2024. This work was jointly supported by the National Natural Science Foundation of China (Grant No. 52122809, Grant No. 52038001, Grant No. 52311530081). Jinhuan Shan would like to acknowledge the China Scholarship Council (CSC) for financial support (No. 202106567005). (Corresponding author: Wei jiang)

Jinhuan Shan, Wei Jiang, Dongdong Yuan and Yaohan Liu are with School of Highway, Chang'an University, Xi'an 710064, China (e-mail: jhshan@chd.edu.cn, jiangwei@chd.edu.cn).

Yue Huang is with Institute for Transport Studies (ITS), University of Leeds, Leeds LS2 9JT, UK (e-mail: y.huang1@leeds.ac.uk).

crack classification in UAV-captured road images and achieved a 92% accuracy through sliding window-based localization. Silva et al. [16] developed a pavement monitoring system based on UAVs and image processing techniques for pothole identification, achieving an accuracy of over 95%. Ji et al. [17] performed distress semantic segmentation on UAV-captured images using the DeepLabv3+ network and quantitatively assessed actual pavement conditions based on segmented pixel features. Zhu et al. [18] collected a dataset of 300 full-scale images (7952 × 5304 pixels^2) with UAVs and obtained a dataset (UAPD) containing 3151 images of size 512× 512 pixels^2 through cropping and annotation. This dataset covers six types of distress: transverse crack (TC), longitudinal crack (LC), alligator crack (AC), oblique crack (OC), pothole, and repair. They tested three classical object detection algorithms (Faster R-CNN, YOLOv3, and YOLOv4) and found that YOLOv3 achieved the best performance (56.6% mean average precision (MAP)). Zhong et al. [19] designed the W-segnet network for semantic segmentation experiments on the UAPD dataset, achieving better performance than semantic segmentation models like U-net, SegNet, and PSPNet. Zhang et al. [20] incorporated the Multi-level Attention Block (MLAB) into the YOLOv3 algorithm, improving its performance in defect detection on the UAPD dataset with a 7.66% mAP improvement. Hong et al. [21] enhanced the U-Net network by adding an improved encoder, a CBAM module, and a strategy for fusing long and short skip-connections, achieving precise segmentation of small and narrow cracks in UAV images compared to U-Net and other traditional networks. Currently, the majority of research efforts are predominantly directed towards enhancing the accuracy of pavement distress recognition, with a specific focus on the identification of distress in captured images.

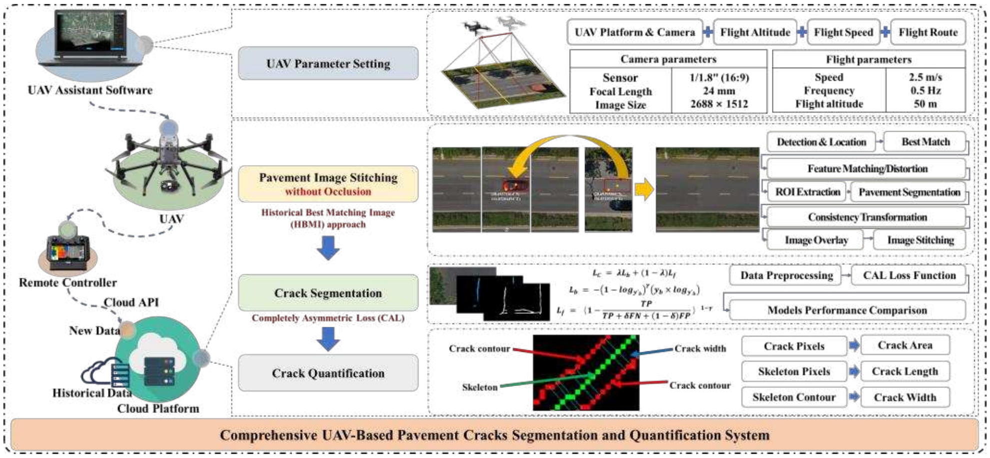
However, there is a paucity of studies that comprehensively investigate the statistical assessment and evaluation of the overall pavement distress condition along an entire road segment [22]. Image stitching technology allows for the creation of panoramic images covering an entire stretch of road. However, the presence of moving vehicles introduces the issue of ghosting in overlapping regions of stitched images. To date, there is a dearth of literature on ghosting elimination specifically in the context of pavement distress detection.

Pavement distress detection tasks mainly fall into three major categories: distress classification, object detection, and semantic segmentation [23]. Classification tasks are relatively straightforward, involving algorithms to assess whether an image contains distress and, if so, to classify the type of distress. Object detection involves identifying distress objects within an image and marking them with bounding boxes, thereby determining the location and type of distress [24]. Semantic segmentation tasks assign a class label to each pixel in an image, providing not only distress recognition but also accurate delineation of distress contours. Unlike object detection, semantic segmentation tasks require pixel-level understanding and classification while offering additional information, such as crack width, length, and orientation. This precise information supports accurate maintenance decision-making and pavement condition forecasting. However, the task of semantic segmentation for cracks presents unique challenges compared to other semantic segmentation endeavors, primarily because cracks do not exhibit well-defined boundaries. Furthermore, the relatively minor proportion of cracks within the overall imagery leads to lower accuracy in semantic segmentation tasks [25]. This discrepancy highlights the need for specialized approaches and methodologies to enhance the reliability of crack semantic segmentation techniques.

Building on the need to overcome the limitations present in current approaches to detecting pavement cracks, this study is poised to introduce a series of advancements. The contribution of this work can be summarized as follows:

1) We propose UAV flight parameters tailored to the requirements of pavement distress semantic segmentation tasks, optimizing the data acquisition process for improved segmentation outcomes.
2) We introduce the Historical Best Matching Image (HBMI) approach to effectively handle obstructions like vehicles in UAV road detection, improving accuracy and reducing the impact of these obstructions.
3) We propose the Completely Asymmetric Loss (CAL), addressing the class imbalance issue in UAV pavement image crack segmentation, and enhancing the crack detection accuracy.
4) We have developed a complete system that encompasses UAV-based parameter setting, obstruction-free image processing, and image segmentation and quantification, enabling intelligent, pixel-level detection across entire road segments for unobstructed analysis (Fig. 1).

The rest of the paper is organized as follows: Section 2 reviews the methods for crack detection in civil infrastructure, imbalanced image segmentation, and image stitching and ghosting elimination. In Section 3, UAV flight parameters based on crack semantic segmentation tasks was proposed. In Section 4, the Historical Best Matching Image (HBMI) approach was proposed for image stitching without vehicle obstructions. In Section 5, Dataset (UAV-Crack500) was established for UAV-based crack semantic segmentation, segmentation performance of slender cracks in UAV-captured images was enhanced by Completely Asymmetric Loss (CAL), and Crack features were extracted and quantified using morphological operations. In Section 6, we provided a summary of our current research and outlined directions for future work.



Camera parameters		Flight parameters	
Sensor Focal Length Image Size	1/1.8" (16:9) 24 mm 2688 × 1512	Speed Frequency Flight altitude	2.5 m/s 0.5 Hz 50 m

$$L_C = \lambda L_b + (1-\lambda)L_f$$

$$L_b = -(1-log_{y'_b})^{\gamma}(y_b \times log_{y'_b})$$

$$L_f = \left(1-\frac{TP}{TP+\delta FN+(1-\delta)FP}\right)^{1-\gamma}$$

Fig. 1. UAV-Driven Pavement Crack Analysis System

II. Related Works

Based on the issues mentioned in Section I regarding existing crack detection methods, this section primarily conducts a literature review from three perspectives: the challenges and solutions in crack detection within civil infrastructure, solutions for class imbalance, and solutions for image stitching and ghosting elimination in the context of crack detection.

A. Crack Detection in Civil Infrastructure

Crack detection plays a crucial role in the maintenance of civil infrastructure, including buildings, roads, tunnels, etc. [26], often involving three types of construction materials: asphalt [27], concrete [28], and metal [29]. Although these materials differ in texture and luster, and the causes and patterns of cracks vary, there are common challenges in the semantic segmentation of cracks, particularly in the case of slender cracks [30], [31]:

1) **Topological Complexity**: Slender cracks often present more complex patterns compared to their shorter, wider counterparts. The topology of cracks is intricate, with no clear, defined boundaries. Cracks differ in width and direction, and may intersect with other pavement features, challenging segmentation algorithms to accurately trace the entire length of the crack.
2) **Background Noise Diversity**: The environments in which civil infrastructure crack detection occurs are varied, with diverse and complex background noise (such as lighting, shadows, stains, etc.). Semantic segmentation models may struggle to maintain the continuity of long linear features, a problem exacerbated in the case of slender cracks where detection breaks can occur due to variations in lighting, shadows, and stains.
3) **Class Imbalance**: Class imbalance is an issue when the dataset contains significantly more instances of one category (e.g., non-cracked surfaces) over another (e.g., slender cracks). This imbalance can cause biases in machine learning models, making them more effective at identifying the majority class and less so at detecting the minority class, in this case, slender cracks.

To address the complexity of crack topology and the diversity of background noise, solutions can primarily be approached from two angles:

1) **Multi-scale Features/Models Fusion and Global Attention Mechanisms**: Multi-scale features fusion can improve the model's accuracy in segmenting cracks of different scales, while global attention mechanisms can endow the model with a comprehensive perspective, reducing the impact of noise on segmentation results [32]. Furthermore, the ensemble of different segmentation models can leverage the strengths of various models to improve crack segmentation performance [33].
2) **Data Augmentation**: The performance of models can be enhanced through training with a large variety of images featuring different crack shapes and backgrounds. Data augmentation mainly involves increasing the quantity of original data and employing data enhancement methods to expand the dataset.

The issue of class imbalance will be discussed in detail in Section B, along with strategies for image stitching and the removal of ghosting effects, which are crucial for enhancing the reliability and accuracy of crack detection in civil infrastructure.

B. Imbalanced Image Segmentation

Deep learning algorithms perform data-driven feature learning in an end-to-end manner, leading to remarkable advancements in the field of image segmentation. Presently, a plethora of classical semantic segmentation models, including but not limited to FCN (Fully Convolutional Network) [34], U-Net [35], DeepLab [36], and PSPNet [37], have achieved impressive results in a wide range of image segmentation tasks. These models not only exhibit outstanding performance but also provide effective means to deeply comprehend and process semantic information within images.

In convolutional neural networks (CNNs), learning abstract feature representations through successive convolution and pooling operations has become a predominant approach in image processing. Nevertheless, the continuous application of these operations can result in a decrease in feature resolution, leading to suboptimal performance in tasks involving dense predictions for small objects or object edges. Furthermore, the uneven distribution of pixel quantities among different classes in images can cause models to exhibit a bias towards predicting the majority class while neglecting minority classes, thereby impacting prediction accuracy.

To address these issues, researchers have made enhancements to classical semantic segmentation networks, including:

1) **Data Augmentation**: For minority classes, data augmentation techniques are employed to increase sample quantity through data transformations. This aids in improving the ability of model to learn from minority classes [38].
2) **Patch-Based Approaches**: These methods divide the image into smaller patches to allow the model to focus more on local regions, thus mitigating the impact of class imbalance issues and enhancing model performance [39]. Furthermore, Performance in semantic segmentation can be improved by initially conducting object detection to extract target regions, followed by applying semantic segmentation specifically to these identified areas. This approach leverages the precision of targeted detection to refine the subsequent segmentation process [40].
3) **Loss Function Improvements**: Specialized loss functions designed to handle class imbalance scenarios have been devised, such as Weighted Cross-Entropy Loss [41], Dice loss [42], Focal loss [43], and others. These loss functions assist the model in better handling imbalanced class data.
4) **Algorithmic Architecture Enhancements**: Techniques such as attention mechanisms and balanced sampling have been utilized to ensure appropriate attention to each class [44], [45].

In current research on semantic segmentation of UAV images, it is common for researchers to employ the patch approach, where the entire image is segmented into smaller patches. This not only conserves training resources but also alleviates the issue of sample imbalance through patch selection. This approach has emerged as an effective strategy for addressing small target detection and class imbalance issues. However, research focused on improving the accuracy of UAV road image crack segmentation through loss functions remains limited. This paper employs both patch-based techniques and loss function improvements to enhance the accuracy of pavement crack semantic segmentation.

C. Image Stitching and Ghosting Elimination

UAV-captured images are constrained by their capture range, resulting in incomplete coverage of the area. Consequently, comprehensive road condition assessment cannot be solely reliant on individual images. To address this challenge, the typical approach for UAV aerial images involves image stitching to present a holistic road landscape. The process of image stitching generally comprises a series of operations, including image registration, reprojection, stitching, and blending [46]. However, owing to the presence of dynamic elements such as moving vehicles in road images captured by UAVs, these elements can introduce issues, notably the occurrence of artifacts commonly referred to as ghosting.

Currently, in the field of image stitching, three predominant methodologies are employed to mitigate these issues:

1) **Seam-Driven Method**: This approach involves the detection of dynamic objects, the identification of optimal segmentation and stitching paths, and subsequent fusion of segmented images to eliminate ghosting artifacts. For example, Davis et al. [47] analyze intensity differences in images to locate segmentation lines with low intensity disparities for seamless transitions. Google Maps [48], on the other hand, utilizes an energy function to determine the optimal segmentation path and replaces pixels containing moving objects with corresponding pixels from another image, thus minimizing ghosting effects.
2) **Pixel Replacement Method**: This method detects dynamic objects and utilizes two-dimensional masks to replace pixels in one image with corresponding pixels from another image to eliminate redundant dynamic objects before image stitching. For instance, Murodjon et al. [49] detect moving objects by comparing absolute differences in overlapping regions of two stitched images and employ a two-dimensional mask-based approach for pixel replacement. Xue et al. [50] achieve ghosting object elimination through recognition of these artifacts and selection of image information sources based on predetermined rules.
3) **Deep Learning Method**: This approach employs deep learning techniques for the detection, removal, and content inpainting of objects in images. Among these techniques, Generative Adversarial Networks (GANs) are frequently utilized. The process involves creating unobstructed images using a generator within a pretrained GAN network, followed by enhancing the generator's capability to remove obstructions through adversarial learning with a discriminator [51], [52], [53]. Unlike the pixel replacement and seam-driven methods, which derive image pixels from real-world scenes, the images generated through deep learning are determined by the pixels surrounding the occluded area and the network's prior knowledge. Therefore, while this method effectively addresses image occlusions, the generated images may not accurately reflect the true pixels of the original occluded area, making it less suitable for applications such as crack detection in occluded regions.

In addition, Zhang et al. [54] have addressed the issue by employing a semantic segmentation algorithm to separate foreground and background images. They subsequently eliminate feature matching points in the foreground images and perform feature matching on the background images to achieve successful stitching of the background. However, it's important to note that ghosting artifacts may still persist in the foreground images.

It is noteworthy that there is currently a limited body of literature specifically addressing ghosting elimination in the context of pavement distress detection. This poses a considerable challenge for assessing the overall road condition through UAV aerial surveys. Furthermore, in areas with substantial vehicular traffic, even after applying ghosting elimination procedures, complete elimination of vehicle obstructions on the road may not be achievable. This, in turn, can potentially impact the accurate evaluation of pavement conditions. Currently, there is a dearth of literature addressing such challenges. However, drawing inspiration from the Pixel Replacement Method, this paper introduces a historical image-based pixel replacement method in Section 4.

III. UAV Parameter Setting

In order to acquire an adequate volume of road imagery data at a sufficiently high resolution, it is imperative to initially configure the flight parameters of the UAV. This entails establishing parameters such as flight altitude, image resolution, flight speed, and route planning, among others.

A. UAV Platform and Camera

The research employs the DJI M300 RTK, a recently introduced professional-grade unmanned aerial vehicle. This UAV boasts advanced features such as high-performance flight control, centimeter-level precision positioning, dual-battery endurance, IP45 high protection rating, substantial payload capacity, and heightened safety capabilities. Presently, it finds widespread applications in various domains, including construction, surveying, security monitoring, and emergency response.

The camera employed in this study is the Zenmuse H20N, an all-weather hybrid sensor payload that integrates dual thermal imaging cameras, wide-angle camera, zoom camera, and a laser rangefinder sensor. It incorporates state-of-the-art neural network-based noise reduction algorithms to overcome challenges in low-light conditions, ensuring robust focusing for clear and bright imagery. The visible light zoom camera boasts a 4-megapixel resolution and supports a 20x hybrid optical zoom. Leveraging the DJI M300 RTK platform and the Zenmuse H20N camera enables stable high-altitude flight and facilitates high-precision image acquisition. Additionally, the zoom camera allows for localized magnification, achieving pixel-level resolution for pavement distress analysis.

B. Flight Altitude

Upon selecting the camera and aerial platform, the determination of flight altitude is primarily influenced by three critical factors: safety considerations, image resolution, and coverage area.

1) **Safety Considerations:** In urban areas, dense high-rises and electromagnetic sources (i.e., urban buildings and cell towers) disrupt both GPS signals for UAV navigation and communication between UAVs and remote controllers. Higher flight altitudes reduce these urban interferences, enhancing UAV control and safety [55].
2) **Image Resolution:** Lower flight altitudes yield higher image resolutions. Furthermore, the use of zoom lenses allows achieving high resolutions even at moderate flight altitudes.
3) **Coverage Area:** To minimize the need for multiple aerial passes, it is desirable to cover a larger area of the road. However, it's essential to note that as the coverage area increases, the image resolution decreases due to the fixed size of the camera photosensitive components.

Considering that urban roadside mobile communication towers typically fall within the range of 30-50 meters in height, with altitudes exceeding 50 meters not causing disruptions to drivers and being higher than typical roadside infrastructure and residential buildings, a flight altitude greater than or equal to 50 meters was selected for this study.

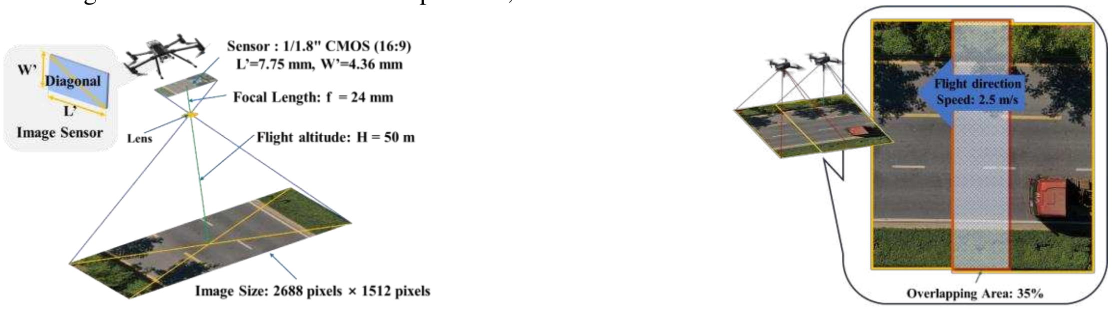
The Zenmuse H20N zoom camera features a standard image sensor size of 1/1.8 inches. By determining the coverage width and flight altitude, optical zoom is employed to maximize image resolution. In practical testing, it was observed that at a flight altitude of 50 meters, 4X optical zoom adequately covers a single three-lane direction.

With a chosen flight altitude of 50 meters, a 4X optical zoom with a focal length of 24 mm, and an image sensor size of 7.75 mm × 4.36 mm (16:9), the captured image dimensions are 2688 pixels × 1512 pixels (Fig. 2 (a)). This allows for the calculation of the actual image coverage area and the real-world dimensions represented by each pixel using optical principles (1).

$$scale = \frac{f}{H} = \frac{S_w}{G_w} \tag{1}$$

Where f is focal length of the camera; H is the flying height above ground level; S_w is the sensor width of the camera; G_w is the image footprint on the ground.

Based on the derived formula, the real image dimensions are determined to be 16 m × 9 m, with an approximate individual pixel size of 6 mm × 6 mm.



(a) Flight altitude and capturing range

(b) Flight speed and overlapping area

Fig. 2. UAV flight parameters

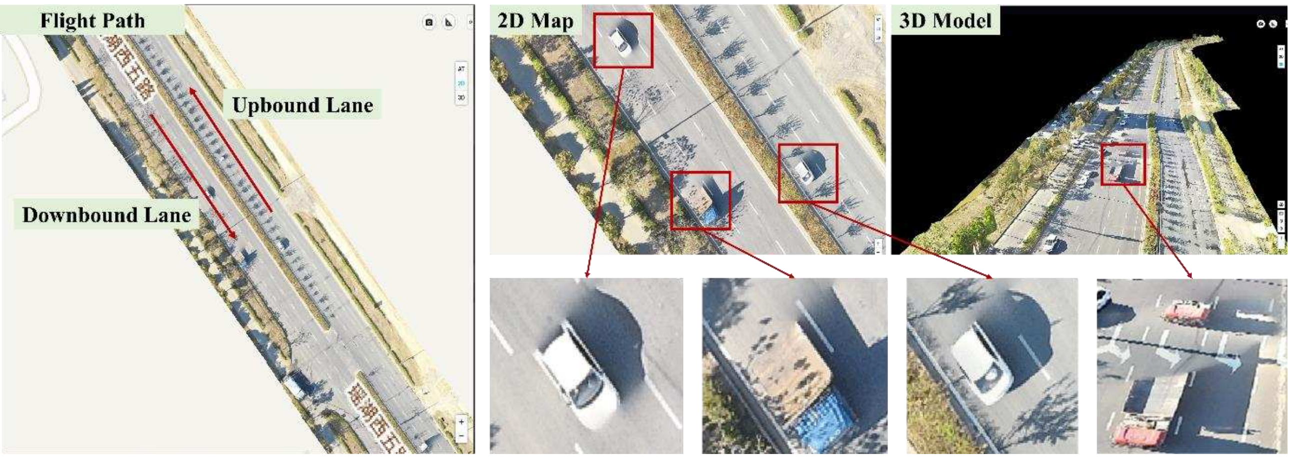


Fig. 3. Route planning, and ghosting artifacts in stitching via DJI Terra software.

C. Flight Speed

Flight speed primarily considers flight efficiency and photo overlap. To enhance the quality of subsequent image stitching, it is recommended that the overlap between stitched images falls within the range of 20% to 50%. Based on the capture frequency (2 seconds per image) and the image width (9 meters), a flight speed of 2.5 m/s is set, resulting in an approximate along-track photo overlap rate of 35% (Fig. 2 (b)).

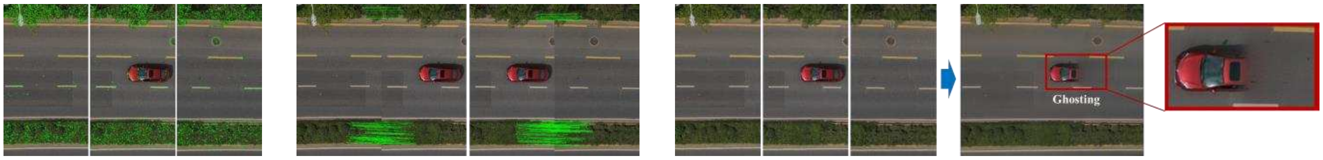
D. Route Planning

Utilizing the waypoint flight functionality of the DJI M300 RTK, flight waypoints are pre-set on the map. This study focuses on collecting single-directional lane images, and thus, waypoints are positioned along the centerline of the single-directional lane. Upon UAV activation, it autonomously cruises along the designated flight path, and upon mission completion, it returns automatically. The flight route is configured for both up and down lane traversal, capturing data from both directions, ensuring comprehensive coverage of both lanes (Fig. 3).

To document the capture range and image quality, DJI Terra is employed for the stitching of visible light images, which are then overlaid onto a map. This visualization effectively depicts the flight path and the collected data (Fig. 3). DJI Terra is a professional ground modeling software developed by DJI, designed for efficient, precise modeling, mapping, and analysis tasks. However, when applied to road scenarios, the dynamic nature of road vehicles can pose challenges. Issues such as modeling failures and ghosting were encountered during the road image stitching process, as evidenced in the 2D Map and 3D Model of Fig. 3. These 3D models were generated using DJI Terra's 3D photogrammetric reconstruction technology, which converts 2D images into three-dimensional representations using advanced photogrammetry techniques [56]. Consequently, for pavement inspection, the presence of vehicle occlusion and modeling artifacts may potentially impact distress detection. The following section will progressively address these issues of vehicle occlusion and stitching artifacts to achieve comprehensive pavement distress detection via UAV-based methods.

IV. Pavement Image Stitching without Occlusion

Common UAV image stitching primarily relies on algorithms that detect keypoint features in registered images (Fig. 4 (a)), compute invariant feature descriptors, and subsequently employ these invariant feature descriptors for image matching (Fig. 4 (b)). This approach is commonly and effectively employed when capturing static structures and environments. However, in the context of acquiring UAV-based pavement images, inevitable occlusion by road vehicles leads to the loss of pavement pixels. Additionally, due to the dynamic nature of moving vehicles, attempts to stitch pavement images can result in stitching failures or the creation of discontinuous artifacts (ghosting) (as depicted in Fig. 4 (c). This is due to the introduction of inconsistent features or changes in perspective by moving objects, rendering accurate alignment and stitching of image features challenging.



(a) Features detecting (b) Features matching (c) Image distorting and stitching

Fig. 4. Common method for image stitching

Considering the challenges, this paper introduces the **Historical Best Matching Image (HBMI) approach**. Its objective is to extract the historically best-matching region of the obscured road surface and fuse it with the current image to generate an unobstructed, true-to-life panoramic view of the road. The specific steps are outlined as follows (Fig. 5):

1) **Image capturing** (Fig. 5 (a)): Leveraging the UAV flight parameters from Section 3, pavement imagery is acquired, generating a sequence of road images with continuous overlapping region.
2) **Detection and location** (Fig. 5 (b)): Utilizing an object detection algorithm (in this study, YOLO v5), road vehicles are detected, and their geographical coordinates are mapped to determine vehicle locations, defining Regions of Interest (ROIs) for each image.
3) **Best match for historical captured image** (Fig. 5 (c)): Based on the ROI locations, the corresponding best-matching historical images are identified in the image repository. The best matching criteria are as follows:
   a) No occluding detection boxes (vehicles) in the historical image.
   b) The center of the image is within a specified distance of the ROI center, ensuring minimal distortion within the ROI.
   c) The historical image capture time is closest to the current image's, preserving the pavement state as closely as possible.
4) **Feature matching and distortion correction** (Fig. 5 (d)): The historical image that best matches the obstructed photo is identified using the best-matching criteria. Keypoint detection and feature matching are performed between this historical best-matching image and the obstructed image. Based on the feature matching results, a projection and mapping transformation is applied to the historical best-matching image to align it with the obstructed image.
5) **Region of Interest (ROI) extraction** (Fig. 5 (e)): Using the ROI bounding box coordinates, the target area is extracted from the image after the projection mapping transformation.
6) **Pavement Segmentation** (Fig. 5 (f)): To enhance the stitching quality, it is crucial to maintain consistent color tones between the stitched image and the obstructed image. As the previous step extracts pavement pixels, achieving color tone consistency with pavement of the obstructed image requires segmentation. Utilizing a semantic segmentation algorithm (in this study, DeepLabv3), asphalt pavement pixels within the image are segmented.
7) **Color Tone Consistency Transformation** (Fig. 5 (g)): The images of the target area (ROI) extracted in Step 5 are subjected to color correction to match the color tone distribution of the segmented pavement in Step 6 (histogram matching method is employed in this study). Histogram matching alters the histogram of an image, changing the grayscale of individual pixels, enhancing local contrast without affecting global contrast. This method is particularly useful for images where both the background and foreground are too bright or too dark. As shown in Fig. 5 (g), the color tone distribution of the target area after histogram matching (ROI-HM) aligns with that of the pavement.
8) **Image Overlay** (Fig. 5 (h)): The color-adjusted image ROI-HM from Step 7 is directly placed over target location of the obstructed image, achieving image fusion.
9) **Features Matching and Distortion** (Fig. 5 (i)): By repeating Steps (2)-(8), pavement pixels with vehicle occlusion in the UAV capture sequence are fused, and traditional keypoint detection and matching of the image sequence are performed. Subsequently, based on the matched features, a projection mapping transformation is applied.
10) **Image Stitching** (Fig. 5 (i)): After obtaining all the images transformed through projection mapping, these images are stitched together. In this study, the MultiBand Blender method is used, which is proposed by Matthew Brown and David G. Lowe [57]. It can decompose different frequency bands (high-frequency and low-frequency components) and blends each band individually. High-frequency components refer to areas of rapid intensity (brightness/gray scale) change, like edges or contours in an image. Low-frequency components, on the other hand, indicate areas where the intensity changes more gradually, typically found in larger uniform color blocks. This distinction is crucial in blending UAV-captured images to create seamless mosaics, efficiently handling overlapping and alignment challenges in aerial imagery.

Considering the gradual yet potentially rapid progression of pavement crack development, we recommend reducing UAV inspection intervals for areas with poor pavement conditions. This strategy aims to minimize discrepancies between historical and current data, ensuring timely maintenance interventions. Furthermore, where feasible in terms of time and budget, performing multiple data collections over consecutive days on the same road segment can effectively address crack development. This frequent monitoring enables a more comprehensive and detailed understanding of crack progression.

To further contrast the differences between our method and traditional approaches, this paper conducted ablation experiments comparing the presence or absence of historical best matching images with projection mapping and the consistency transformation of color tones within ROIs. While the use of histogram matching aligned the color tones with the background image, it resulted in some offset in pavement distress due to the absence of feature matching and mapping with the original image, as depicted in Fig. 6 (a). After employing feature matching and mapping, road distress in the image became aligned, meaning spatial pixels corresponded correctly. However, noticeable differences in color tones between the foreground image (ROI) and the background image were observed, which may introduce interference during the subsequent pavement distress recognition process, as shown in Fig. 6 (b). Finally, with the use of projection mapping and histogram matching, the foreground integrated well with the background, and pixel alignment was achieved, significantly restoring the original pavement condition, as illustrated in Fig. 6 (c).

In Fig. 6 (c), it can be observed that, even with the combination of histogram matching and projection mapping, there are still some differences in color tones at the edge. This is because during histogram matching, pavement pixel selection encompasses the entire image, and the grayscale distribution of the entire pavement pixels can be influenced by some pixels (e.g., patch-like repairs). Therefore, if further improvement in the fusion of edge pixels is desired, a region-based pixel color tone consistency approach can be employed. This involves extracting road surface pixels within a certain range from the region of interest (ROI) and then performing histogram matching with the segmented ROI image. However, it's important to note that this does not affect road distress detection. Once the model is trained, it will not erroneously detect the stitching traces during pavement distress detection.

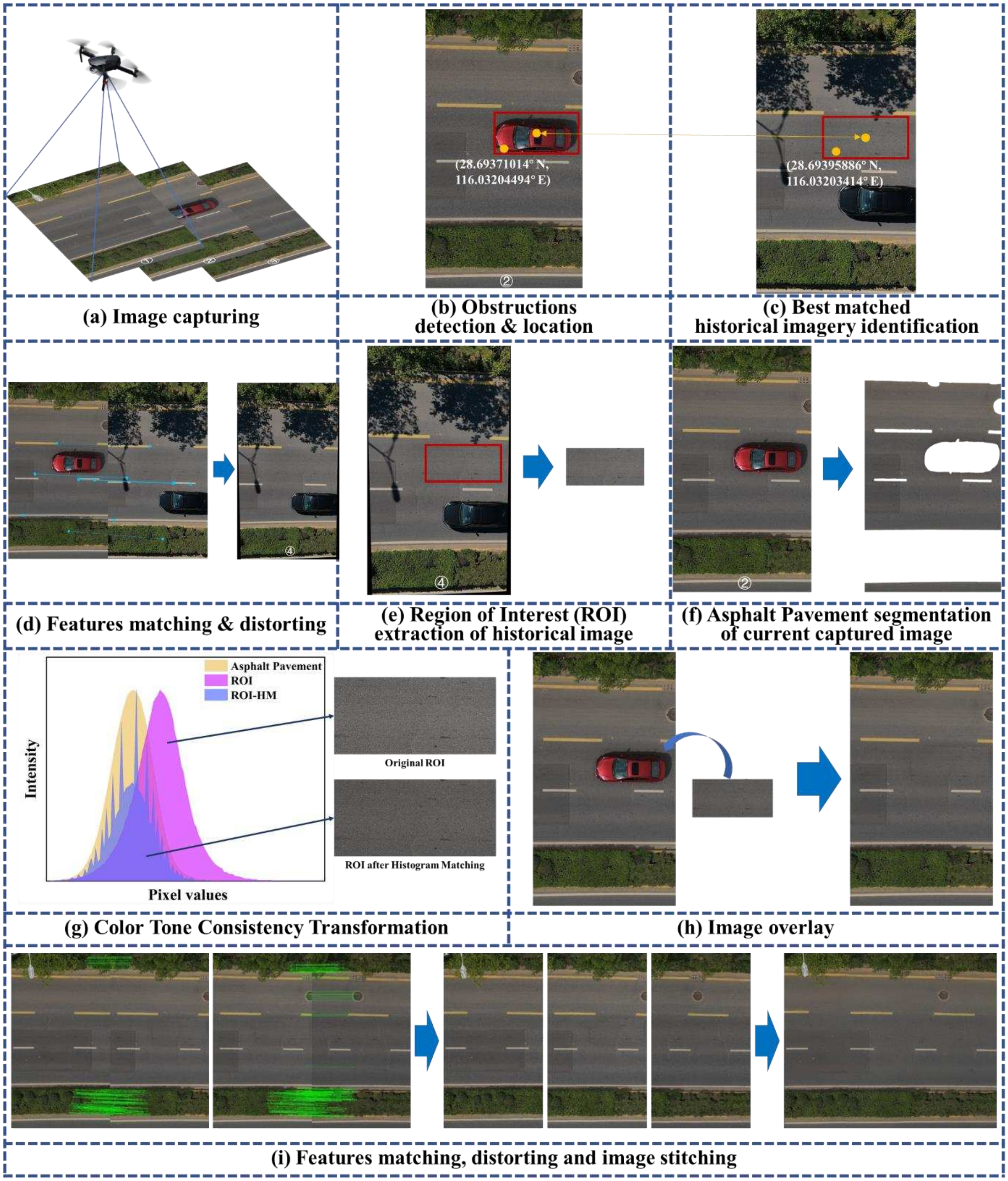
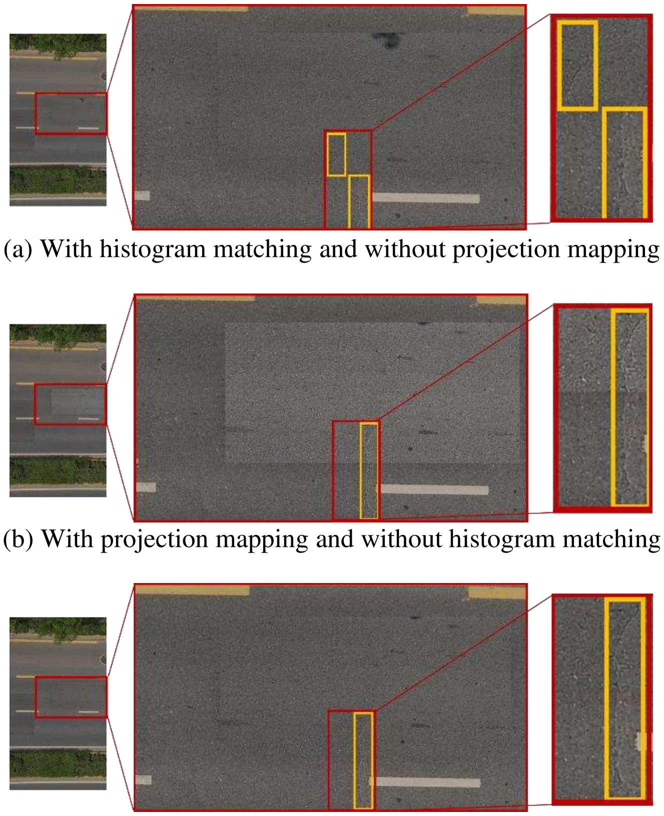

①
②
③
(28.69371014° N,
116.03204494° E)
②
(28.69395886° N,
116.03203414° E)
(a) Image capturing
(b) Obstructions detection & location
(c) Best matched historical imagery identification
④
④
②
(d) Features matching & distorting
(e) Region of Interest (ROI) extraction of historical image
(f) Asphalt Pavement segmentation of current captured image
Asphalt Pavement
ROI
ROI-HM
Intensity
Pixel values
Original ROI
ROI after Histogram Matching
(g) Color Tone Consistency Transformation
(h) Image overlay
(i) Features matching, distorting and image stitching

Fig. 5. Historical best matching image (HBMI) approach



(a) With histogram matching and without projection mapping

(b) With projection mapping and without histogram matching

(c) With histogram matching and projection mapping

Fig. 6. Ablation study of histogram matching and projection mapping

V. Crack Semantic Segmentation and Quantification

UAV-based pavement distress semantic segmentation encounters two primary challenges:

1) **Annotation Difficulty:** Current vehicle-mounted cameras provide high pixel accuracy, with resolutions typically reaching 1 mm × 1 mm per pixel. In contrast, in images captured by UAVs, each pixel represents a larger real-world size, posing challenges for pixel-level annotations due to the relatively small dimensions of cracks.
2) **Segmentation Challenge:** The larger pixel sizes in UAV images, combined with the slender characteristics of cracks, result in a minor proportion of the image being occupied by these distress features. Consequently, class imbalance issues arise during semantic segmentation. When training algorithms directly on UAV photos, it is difficult to accurately identify cracks, occasionally predicting entirely black segments (representing the background).

To address the annotation difficulty, this study meticulously configured the UAV flight path and zoom settings. Specifically, the UAV captured images separately for the up and down lanes while employing a 4X optical zoom, achieving a pixel accuracy of approximately 6 mm × 6 mm. This configuration facilitated the acquisition of crack pixels with enhanced precision. Furthermore, future enhancements in camera precision and photosensitive component sizes hold the potential to meet pixel-level detection requirements akin to those of vehicle-mounted cameras. To confront the segmentation challenge, this paper introduced image partitioning and integrated novel loss functions, resulting in a notable enhancement in semantic segmentation accuracy.

A. Data Preprocessing

The inherent imbalance in pixel distribution can introduce a bias in models, causing them to predominantly predict background areas while potentially neglecting regions of interest, thus compromising overall model performance. As illustrated in the 2688×1512 UAV-captured image (Fig. 7), cracks constitute a relatively small fraction, accounting for just 0.52% of total pixels. Notably, during testing, direct training on the original image consistently resulted in predictions classifying the entire image as background, effectively rendering it entirely black.

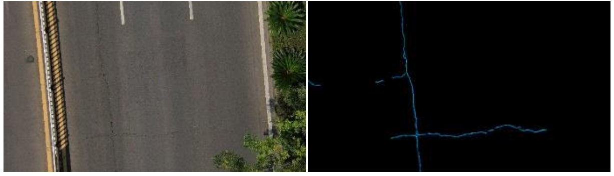


Fig. 7. Full image and ground truth

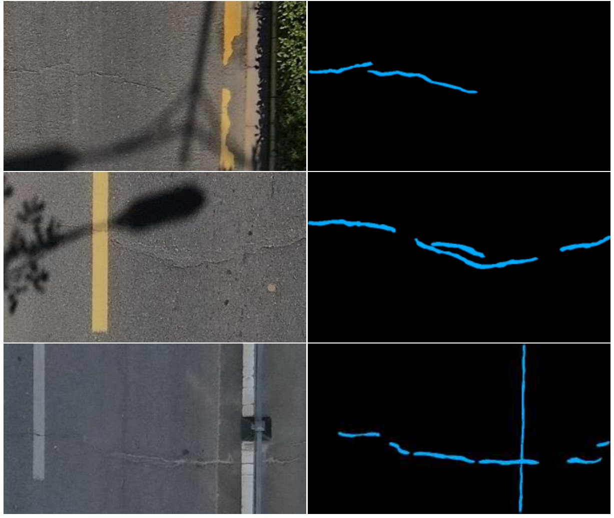


Fig. 8. Distracting scene image blocks

To address this challenge, this study adopted a segmentation strategy that involved partitioning the original image into 16 blocks (each measuring 672×378 pixels) and excluding blocks devoid of cracks. We primarily utilized EISeg for annotation, following a two-step approach. Initially, multiple individuals performed coarse crack annotations. Given the inherent variability in individual markings, these annotated JSON files (containing segmented area points) and images were subsequently sent to road maintenance experts for refined adjustments in the interactive EISeg interface. Subsequently, 500 representative crack images were subjected to meticulous fine-grained semantic annotation (named as UAV-Crack500). These images were then distributed into distinct sets: 250 images for training, 50 for validation, and 200 for testing. It is worth highlighting that the image selection process deliberately incorporated scenarios with potential sources of disturbance, such as the presence of road markings, shadows, curbstones, trees, and road dividers (Fig. 8).

B. Loss Function for Imbalanced Data

In the context of binary classification tasks, the prevailing loss function is Binary Cross-Entropy Loss (BCE), which is mathematically represented as follows (2):

$$L_{BCE} = -((y \times log_{y'}) + (1 - y) \times \log(1 - y')) \quad (2)$$

Where y denotes the actual ground truth, while y' signifies the predicted value.

However, the utility of Cross-Entropy loss functions primarily hinges on their capacity to minimize pixel-level discrepancies, treating each pixel with equal weight during loss calculation. When confronted with class imbalance, these functions tend to disproportionately favor pixels belonging to the majority class, resulting in a diminished capacity to delineate the minority class effectively.

Focal Loss emerges as a purpose-built solution designed to address the intricacies of class imbalance. It achieves this by downweighting easily classified samples, thereby shifting the focus toward samples that pose greater classification challenges. The formula defining Focal Loss is as follows (3):

$$L_F = -\alpha_t (1 - y')^{\gamma} (y \times log_{y'}) \quad (3)$$

Where α_t represents a balancing factor strategically employed to modulate the weighting of samples, distinguishing between those readily classified and those requiring greater scrutiny. In this paper, we experimented with different balancing factors α_t for foreground and background elements, setting the weight factor for foreground elements at 0.50 and 0.75. A factor of 0.50 indicates equal weights for foreground and background, while 0.75 increases the weight factor for the foreground elements. The factor γ exerts control over the extent of adjustment applied to sample weights.

Focal Loss, with its specialized design, sharpens the discernment of arduously classified samples, serving as an effective tool for mitigating class imbalance issues. Nevertheless, it imparts uniform parameters across all classes, which can suppress the contribution of rare classes and may not always outperform Binary Cross-Entropy Loss in practical applications.

Dice loss is employed to assess the similarity between the segmentation masks generated by the model and the actual ground truth masks. Unlike the traditional cross-entropy loss, which processes the predictive probability of each pixel or class individually, Dice loss optimizes the model by evaluating the spatial overlap between classes. This approach inherently mitigates the model bias caused by class imbalance, focusing on the holistic agreement between predicted and true segmentation areas rather than individual pixel accuracy. The formula defining Dice Loss is as follows (4):

$$L_{Dice} = 1 - \frac{2TP}{2TP + FN + FP} \quad (4)$$

Where *TP* represents True Positives, *TN* represents True Negatives, *FP* represents False Positives, and *FN* represents False Negatives.

This study introduces a novel Completely Asymmetric Loss (CAL), built upon the foundation of the Asymmetric Unified Focal Loss [58]. The formula is as follows (5):

$$L_C = \lambda L_b + (1 - \lambda) L_f \quad (5)$$

Where,

$$L_f = \left(1 - \frac{TP}{TP + \delta FN + (1 - \delta) FP}\right)^{1-\gamma} \quad (6)$$

$$L_b = -(1 - log_{y'_b})^{\gamma} (y_b \times log_{y'_b}) \quad (7)$$

L_f represents the loss function for foreground (positive) samples, while L_b signifies the loss function for background (negative) samples. λ is used to balance the weights of positive and negative samples in the loss function. In this paper, by controlling δ and γ, the background elements have been suppressed and the foreground elements have been enhanced, thus λ is set to 0.5, which means that the weights for both the positive and negative samples are equal. The parameter δ is employed to control the relative weight between positive and negative samples. When δ > 0.5, it indicates that the loss function is assigning a greater relative weight to false negatives (FN), which are actual positive samples that were incorrectly predicted as negative. This can lead to a higher penalty for misclassifying positive samples, effectively encouraging the model to be more sensitive to the detection of foreground elements. Therefore, in this paper, δ is set to 0.6. Additionally, the loss for rare classes is amplified through the exponent (1-γ). y'_b denotes the predicted value for the background, and y_b corresponds to the actual background label. The loss for the background is suppressed by the exponent γ. Additionally, γ also relatively enhances the weight of the difficult-to-classify samples within the background.

This approach is designed to address class imbalance issues by allowing for asymmetric treatment of positive and negative classes, tailored to the specific needs of the problem. It is noteworthy that γ can suppress the overall background loss while enhancing the weight of difficult-to-classify samples within the background; when 0<γ<1, (1-γ) enhances the overall foreground loss, but increases the weight for easy-to-classify samples even more. Therefore, the following text selects different values of γ for hyperparameter tuning (γ=0.2, 0.4, 0.6, 0.8).

C. Crack Semantic Segmentation

This study employed three widely recognized and effective semantic segmentation algorithms, namely U-Net, PSPNet, and DeepLabv3+, as benchmark algorithms for semantic segmentation of UAV images. The U-Net architecture consists of an encoder and a decoder, forming a "U"-shaped network structure. The encoder is responsible for capturing contextual information and extracting features, while the decoder restores the size and generates segmentation masks through upsampling. PSPNet leverages pyramid pooling to extract and fuse contextual information at different scales, and it achieves semantic feature extraction for objects of various sizes through multiscale feature fusion. DeepLabv3+ also adopts an encoder-decoder structure. In the encoder part, it combines dilated convolution and global average pooling techniques to extract and fuse multiscale features. The decoder part utilizes bilinear interpolation for upsampling, producing semantic segmentation results with image resolution.

This study employed Accuracy, Precision, Recall, F$_1$-score, and Intersection over Union (IoU) as evaluation metrics, calculated using the (8)-(12):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall} \quad (11)$$

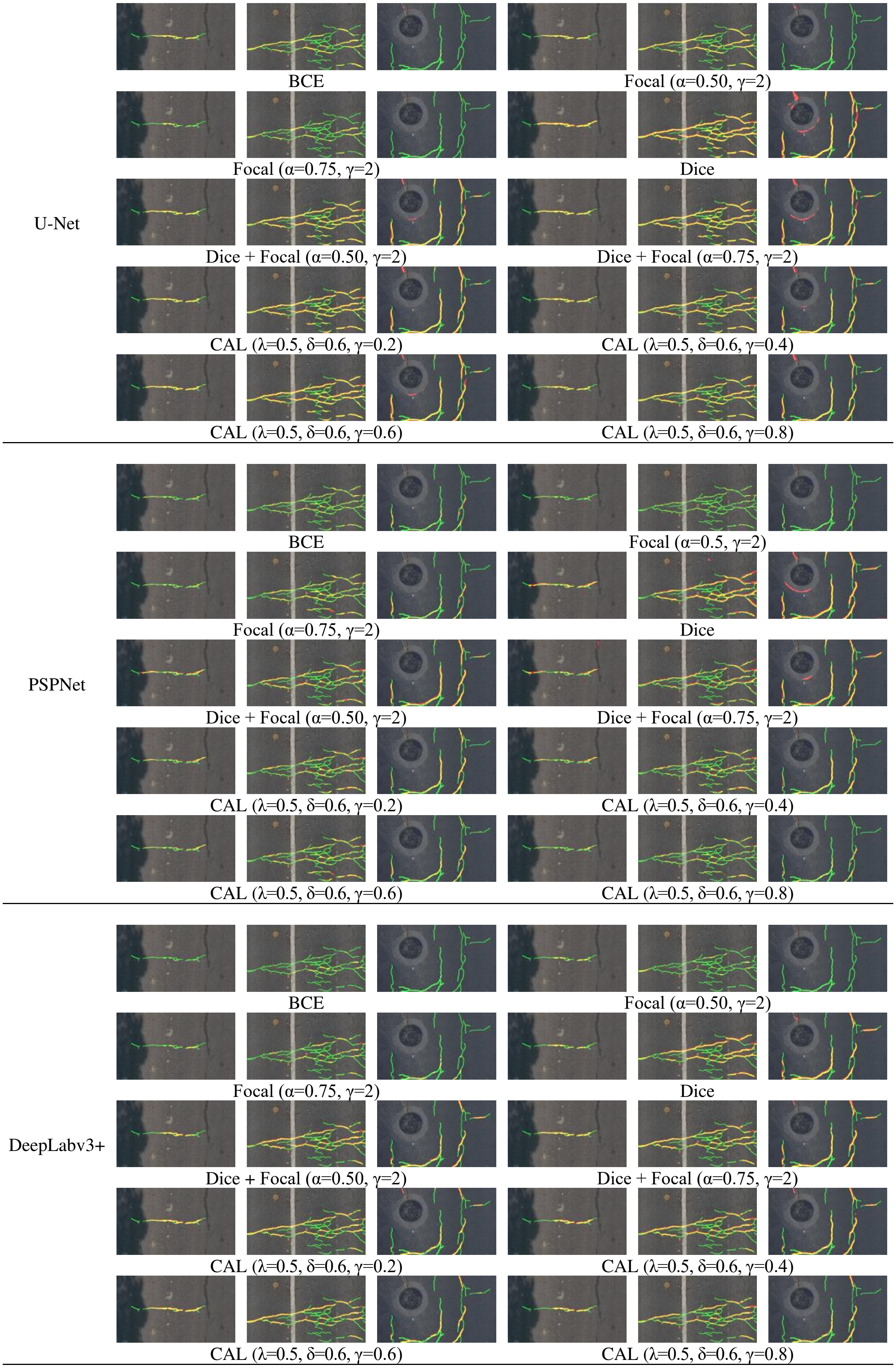
$$IoU = \frac{TP}{TP+FP+FN} \quad (12)$$

In this study, we leveraged the publicly accessible model zoo, Segmentron [59], hosted on GitHub, as our experimental platform. Additionally, we customized the loss function and applied color jitter augmentation with parameters set to brightness=0.1, contrast=0.1, saturation=0.1, and hue=0.1, to increase image diversity and thereby improve the model's performance. Similar to many studies in crack segmentation [33], [60], [61], this research employs a 1-pixel tolerance margin to mitigate the substantial label noise present in crack dataset images. This noise primarily arises from variations in human annotator judgments, human errors, or inaccuracies from computer-generated labels. By implementing this tolerance, the study permits slight discrepancies between the model's predictions and the ground truth, ensuring that these minor deviations do not negatively impact the assessment of the model's performance. This approach effectively acknowledges and makes allowances for the natural inconsistencies found in crack annotations. The data presented in the table clearly indicates that Focal Loss demonstrates limited effectiveness in the recognition of UAV imagery, and in some cases, may even produce detrimental outcomes. This phenomenon is largely ascribed to Focal Loss's concurrent suppression of both infrequent and background samples. Conversely, models employing Dice Loss observed a substantial enhancement in performance. However, the combination of Dice and Focal Loss did not yield further improvements in model efficacy. Our proposed Completely Asymmetric Loss, in comparison to Dice Loss, achieves an improvement in the F_1-score ranging from 0.1% to 0.7%. Moreover, with parameters set to λ=0.5, δ=0.6, and γ=0.2, our model secures a more favorable outcome in terms of the F_1-score. This underscores the value of an asymmetric approach in enhancing the prediction of rare samples. During the comparative analysis of different algorithms, it was noted that U-Net exhibited the best predictive performance. This observation can be attributed to the fact that the focus of this study was on the detection of narrow and slender cracks. Expanding the field of view and extracting features of varying sizes did not contribute significantly to the semantic understanding of these features; instead, it might introduce noise and interference.

TABLE I

THE PERFORMANCE OF MODELS ON THE TEST SET OF UAV-CRACK500 (%)

Model	Loss	Acc	Pr	Re	F_1	mIoU	Crack IoU
U-Net	BCE	98.38	96.57	19.82	32.88	70.21	42.03
	Focal (α=0.50, γ=2)	98.42	94.83	22.79	36.75	69.15	40.09
	Focal (α=0.75, γ=2)	98.22	**98.17**	9.78	17.79	70.03	41.85
	Dice	98.76	74.19	69.60	71.82	**76.92**	**55.18**
	Dice+Focal (α=0.50, γ=2)	**98.91**	84.36	60.86	70.71	76.03	53.45
	Dice+Focal (α=0.75, γ=2)	98.89	81.42	64.07	71.71	75.88	53.21
	CAL (λ=0.5, δ=0.6, γ=0.2)	98.88	80.11	65.47	**72.05**	75.90	53.27
	CAL (λ=0.5, δ=0.6, γ=0.4)	98.83	85.38	54.33	66.40	75.02	51.39
	CAL (λ=0.5, δ=0.6, γ=0.6)	**98.91**	84.74	60.61	70.67	75.95	53.24
	CAL (λ=0.5, δ=0.6, γ=0.8)	98.85	88.16	52.45	65.77	75.12	51.58
PSPNet	BCE	98.28	87.38	16.84	28.23	68.47	38.95
	Focal (α=0.50, γ=2)	98.18	**89.18**	9.46	17.10	68.35	38.70
	Focal (α=0.75, γ=2)	98.35	81.24	25.13	38.38	68.06	38.40
	Dice	98.46	66.98	61.82	64.30	71.73	45.38
	Dice+Focal (α=0.50, γ=2)	**98.50**	72.19	49.41	58.67	71.01	43.96
	Dice+Focal (α=0.75, γ=2)	98.46	70.32	50.09	58.50	71.02	44.00
	CAL (λ=0.5, δ=0.6, γ=0.2)	98.45	66.04	**63.96**	**64.98**	**72.66**	**47.16**
	CAL (λ=0.5, δ=0.6, γ=0.4)	98.43	76.84	35.66	48.71	68.99	40.07
	CAL (λ=0.5, δ=0.6, γ=0.6)	98.40	79.61	30.12	43.70	68.87	39.90
	CAL (λ=0.5, δ=0.6, γ=0.8)	98.33	82.96	22.27	35.12	68.73	39.50
DeepLabv3+	BCE	98.13	86.54	6.44	12.00	62.20	26.29
	Focal (α=0.50, γ=2)	98.14	**95.61**	5.85	11.02	64.19	30.34
	Focal (α=0.75, γ=2)	98.24	91.36	12.79	22.44	68.24	38.50
	Dice	98.50	68.64	**58.88**	63.38	72.36	46.33
	Dice+Focal (α=0.50, γ=2)	98.63	77.24	51.41	61.73	72.53	46.66
	Dice+Focal (α=0.75, γ=2)	98.65	80.22	48.68	60.59	72.39	46.28
	CAL (λ=0.5, δ=0.6, γ=0.2)	98.65	76.50	54.27	**63.49**	72.70	47.02
	CAL (λ=0.5, δ=0.6, γ=0.4)	98.67	77.60	53.69	63.47	72.78	47.17
	CAL (λ=0.5, δ=0.6, γ=0.6)	**98.69**	80.99	50.90	62.51	**72.84**	**47.21**
	CAL (λ=0.5, δ=0.6, γ=0.8)	98.59	78.91	45.94	58.07	71.35	44.33



U-Net

BCE

Focal (α=0.50, γ=2)

Focal (α=0.75, γ=2)

Dice

Dice + Focal (α=0.50, γ=2)

Dice + Focal (α=0.75, γ=2)

CAL (λ=0.5, δ=0.6, γ=0.2)

CAL (λ=0.5, δ=0.6, γ=0.4)

CAL (λ=0.5, δ=0.6, γ=0.6)

CAL (λ=0.5, δ=0.6, γ=0.8)

PSPNet

BCE

Focal (α=0.5, γ=2)

Focal (α=0.75, γ=2)

Dice

Dice + Focal (α=0.50, γ=2)

Dice + Focal (α=0.75, γ=2)

CAL (λ=0.5, δ=0.6, γ=0.2)

CAL (λ=0.5, δ=0.6, γ=0.4)

CAL (λ=0.5, δ=0.6, γ=0.6)

CAL (λ=0.5, δ=0.6, γ=0.8)

DeepLabv3+

BCE

Focal (α=0.50, γ=2)

Focal (α=0.75, γ=2)

Dice

Dice + Focal (α=0.50, γ=2)

Dice + Focal (α=0.75, γ=2)

CAL (λ=0.5, δ=0.6, γ=0.2)

CAL (λ=0.5, δ=0.6, γ=0.4)

CAL (λ=0.5, δ=0.6, γ=0.6)

CAL (λ=0.5, δ=0.6, γ=0.8)

Fig. 9. Typical scenarios with different color tones and various types of interference (yellow: TP; red: FP; green: FN)

Fig. 9 selected for the study depict typical scenarios with different color tones and various types of interference. From the predictive analysis, the application of Completely Asymmetric Loss boosts the model's crack prediction capabilities. In contrast to Dice Loss, which results in a higher incidence of background and noise pixels being incorrectly classified as cracks, our loss function achieves a more balanced trade-off between precision and recall.

Notably, the model demonstrates commendable robustness in complex environments, as evidenced by its ability to distinguish between actual cracks and repaired sections in the first image. However, the segmentation performance for cracks in shadowed areas was found to be suboptimal, indicating a need for future improvements. Enhancing the model's robustness across diverse scenarios will require the incorporation of multi-scenario data and structural optimizations of the model.

When employing DeepLabv3+ and PSPNet, the incidence of noise pixels being mistakenly predicted as cracks decreases. This improvement is attributed to the model's enhanced receptive field using atrous convolution. However, the introduction of atrous convolution and downsampling slightly diminishes the model's ability to accurately predict smaller crack pixels. By comparing the three different algorithms, it becomes apparent that utilizing U-Net yields superior results for predicting UAV road surface images, while DeepLabv3+ performs the least effectively. Consequently, in subsequent research, emphasis can be placed on enhancing predictive capabilities by refining algorithms based on U-Net.

D. Crack Quantification

After the semantic segmentation of cracks, rich pavement damage information can be obtained. In this paper, morphological operations were applied to the segmented crack images to obtain information such as crack area, crack length, crack maximum width, crack mean width, and crack width distribution, which provide crucial data support for road inspection (Fig. 10). The pseudo-code of morphological operations is shown in Algorithm 1.

1) **Crack Area**: To ascertain the proportion of cracks, the ratio of the total crack area to the total area of the original image was calculated. This was achieved by segmenting the images, which had been processed to remove occlusions and were stitched together, utilizing the U-Net + CAL (λ=0.5, δ=0.6, γ=0.2) algorithm introduced in Section 5 C. Through this process, pixels were classified as either crack pixels or background pixels, allowing for the calculation of the crack ratio and area.
2) **Crack Skeleton Extraction and Length**: To accurately predict the geometric features of cracks, this paper applied morphological algorithms (Scikit-Image library) to extract the centerline or skeleton of the cracks, converting the crack regions into pixelated lines, where the pixel count represented the total crack length (with each pixel equivalent to 6 mm, as calculated in Section 3).
3) **Crack Maximum Width and Mean Width**: To calculate the mean and maximum widths of cracks, for each point on the medial axis, the corresponding value in the distance transform image, which represents the shortest distance from the medial axis to the crack boundary, is utilized. Since the medial axis denotes the center of the crack, this distance to the boundary is doubled to determine the full width of the crack at that point. The mean crack width is then obtained by averaging these doubled distances across all points on the medial axis, while the maximum width is identified by selecting the highest value among these calculated widths.

Algorithm 1 Crack Quantification Using Morphological Image Processing Techniques.

```
Input: Grayscale image after crack segmentation
Output: Total Pixel, Skeleton length,
        Max & Mean Width of Crack
Binary image ← Apply threshold(Grayscale image)
CrackArea ← np.sum(Binary image)
Skeleton ← medial_axis(Binary image)
CrackLength ← np.sum(Skeleton)
Skeleton Points ← np.argwhere(Skeleton)
Contours ← cv2.findContours(Grayscale image)
Border Points ← np.argwhere(Contours)
for each point in Skeleton Points do:
    (y, x) ← skeleton point
    Distances ← cdist([(y, x)], Border Points)
    CrackWidth ← min(Distances) × 2
    CrackWidths.append(CrackWidth)
end
AvgWidth ← np.mean(CrackWidths)
MaxWidth ← np.max(CrackWidths)
```

In this research, three complex crack scenarios, as illustrated in Figure 10, were meticulously selected for a comprehensive quantitative assessment encompassing crack area, length, mean width, and maximum width. Additionally, the average outcomes for the entire dataset were systematically evaluated, with the results concisely compiled in Table II. The analysis highlighted that the pixel error was confined to within 16%, a discrepancy primarily linked to the challenges posed by long-distance photography and limitations in image resolution, which often lead to images depicting cracks in an elongated form, thereby obscuring the clarity of crack boundaries.

Looking forward, it is anticipated that advancements in camera lens technology will pave the way for the acquisition of higher-resolution images. Such developments are expected to substantially bolster the segmentation capabilities of the network, thereby elevating the precision in the quantification of cracks.

TABLE II
THE RESULTS OF PAVEMENT DISTRESS QUANTIFICATION

	Ground Truth (pixel)				U-Net + CAL (pixel)				Error (%)			
Image	1	2	3	Total (/image)	1	2	3	Total (/image)	1	2	3	Total (/image)
Area	2614	16791	10099	5205.94	2225	15872	11538	4964.48	14.88	5.47	14.25	4.64
Length	484	3120	1589	711.16	457	2893	1337	665.02	5.58	7.28	15.86	6.49
Mean width	3.86	3.69	4.20	5.51	4.38	4.21	4.62	5.57	13.40	14.00	9.99	1.15
Max width	10.00	10.24	10.20	9.77	8.94	10.00	11.31	9.49	10.56	2.34	10.94	2.81

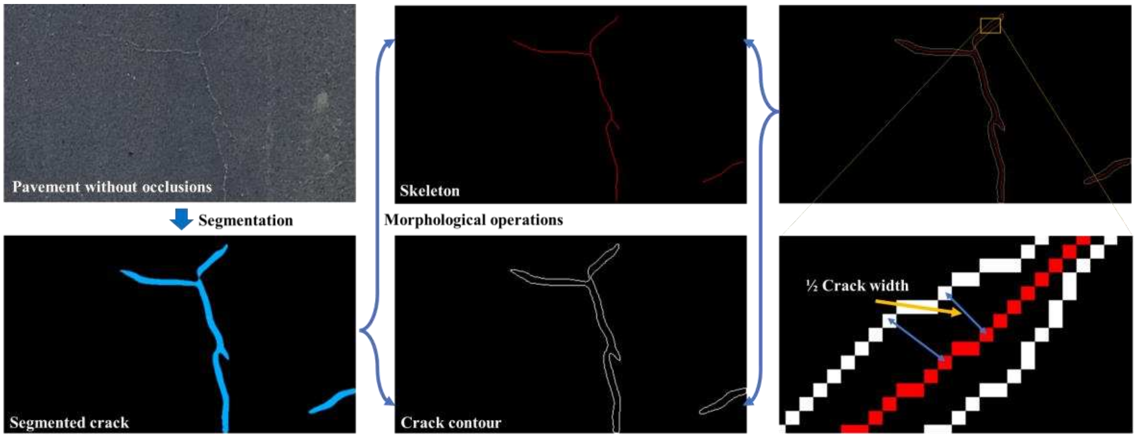


Fig. 10. Crack quantification

VI. CONCLUSION AND FUTURE RESEARCH

Unmanned aerial vehicles offer the potential for significantly enhanced efficiency and coverage in pavement defect detection. However, this task comes with its own set of challenges, including obstructions by road vehicles and the fine, slender features of cracks in the images. These challenges can severely impact the precision of pavement defect semantic segmentation. In this study, a systematic approach is taken to address these challenges and provide a comprehensive road coverage solution.

Firstly, precise UAV flight parameters tailored to the requirements of semantic segmentation are established. To tackle road vehicle obstructions, a novel method known as Historical Best Matching Image (HBMI) approach is introduced. This method effectively allows obstructed regions to be replaced with corresponding pixels, resulting in seamless and unobstructed image stitching for comprehensive road coverage.

Furthermore, to handle the intricate, slender characteristics of cracks, image segmentation and selection processes are implemented. These steps culminate in the creation of a dataset comprising 500 UAV images specifically designed for crack semantic segmentation (UAV-Crack500). Three semantic segmentation algorithms (U-Net, PSPNet, and DeepLabv3+) are systematically evaluated alongside the loss functions (Binary Cross-Entropy Loss, Focal Loss, Dice Loss, Dice + Focal Loss, and Completely Asymmetric Loss). The experimental findings unequivocally indicate that the U-Net algorithm paired with our Completely Asymmetric Loss (U-Net + CAL (λ=0.5, δ=0.6, γ=0.2)) delivers the most robust performance.

Finally, advanced morphological algorithms are employed to extract critical crack features, such as crack area ratio, crack length, crack maximum width, and crack mean width, from semantically segmented images. These results underscore the superior performance of the U-Net + CAL algorithm, particularly in complex scenarios. The precise quantification of crack features enhances the accuracy of road condition assessments.

While this study offers a comprehensive framework for pavement distress detection using UAVs, it is accompanied by certain limitations that warrant further investigation. Firstly, there is a need for the enhancement of UAV imaging accuracy to meet the stringent precision requirements of distress detection. Secondly, the establishment of a comprehensive pavement distress evaluation system remains an open challenge. Subsequent research efforts will be directed toward the detection of other distress types and the conduct of comparative analyses with existing technologies to achieve high-precision and efficient defect detection.

ACKNOWLEDGMENT

The authors would like to thank the Jiangxi Provincial Highway Engineering Testing Center for their technical support in collecting pavement distress data.

REFERENCES

[1] C. Torres-Machi, E. Pellicer, V. Yepes, and A. Chamorro, “Towards a sustainable optimization of pavement maintenance programs under budgetary restrictions,” *J. Clean. Prod.*, vol. 148, pp. 90–102, Apr. 2017, doi: 10.1016/j.jclepro.2017.01.100.

[2] W. Chen and M. Zheng, "Multi-objective optimization for pavement maintenance and rehabilitation decision-making: A critical review and future directions," *Autom. Constr.*, vol. 130, p. 103840, Oct. 2021, doi: 10.1016/j.autcon.2021.103840.

[3] W. Jiang *et al.*, "Research on Pavement Traffic Load State Perception Based on the Piezoelectric Effect," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 8, pp. 8264–8278, Aug. 2023, doi: 10.1109/TITS.2023.3264248.

[4] T. B. J. Coenen and A. Golroo, "A review on automated pavement distress detection methods," *Cogent Eng.*, vol. 4, no. 1, p. 1374822, Jan. 2017, doi: 10.1080/23311916.2017.1374822.

[5] A. Ragnoli, M. R. De Blasiis, and A. Di Benedetto, "Pavement Distress Detection Methods: A Review," *Infrastructures*, vol. 3, no. 4, Art. no. 4, Dec. 2018, doi: 10.3390/infrastructures3040058.

[6] Z. Du, J. Yuan, F. Xiao, and C. Hettiarachchi, "Application of image technology on pavement distress detection: A review," *Measurement*, vol. 184, p. 109900, Nov. 2021, doi: 10.1016/j.measurement.2021.109900.

[7] X. Sun, Y. Xie, L. Jiang, Y. Cao, and B. Liu, "DMA-Net: DeepLab With Multi-Scale Attention for Pavement Crack Segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18392–18403, Oct. 2022, doi: 10.1109/TITS.2022.3158670.

[8] Y. Yu, H. Guan, D. Li, Y. Zhang, S. Jin, and C. Yu, "CCapFPN: A Context-Augmented Capsule Feature Pyramid Network for Pavement Crack Detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 3324–3335, Apr. 2022, doi: 10.1109/TITS.2020.3035663.

[9] H. Li, D. Song, Y. Liu, and B. Li, "Automatic Pavement Crack Detection by Multi-Scale Image Fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2025–2036, Jun. 2019, doi: 10.1109/TITS.2018.2856928.

[10] G. Zhu *et al.*, "A lightweight encoder–decoder network for automatic pavement crack detection," *Comput.-Aided Civ. Infrastruct. Eng.*, vol. 39, no. 12, pp. 1743–1765, 2024, doi: 10.1111/mice.13103.

[11] P. R. T. Peddinti, H. Puppala, and B. Kim, "Pavement Monitoring Using Unmanned Aerial Vehicles: An Overview," *J. Transp. Eng. Part B Pavements*, vol. 149, no. 3, p. 03123002, Sep. 2023, doi: 10.1061/JPEODX.PVENG-1291.

[12] E. Ranyal, A. Sadhu, and K. Jain, "Road Condition Monitoring Using Smart Sensing and Artificial Intelligence: A Review," *Sensors*, vol. 22, no. 8, Art. no. 8, Jan. 2022, doi: 10.3390/s22083044.

[13] X. Lei, C. Liu, L. Li, and G. Wang, "Automated Pavement Distress Detection and Deterioration Analysis Using Street View Map," *IEEE Access*, vol. 8, pp. 76163–76172, 2020, doi: 10.1109/ACCESS.2020.2989028.

[14] Y. Pan, X. Zhang, G. Cervone, and L. Yang, "Detection of Asphalt Pavement Potholes and Cracks Based on the Unmanned Aerial Vehicle Multispectral Imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 10, pp. 3701–3712, Oct. 2018, doi: 10.1109/JSTARS.2018.2865528.

[15] L. Ali, N. K. Valappil, D. N. A. Kareem, M. J. John, and H. Al Jassmi, "Pavement Crack Detection and Localization using Convolutional Neural Networks (CNNs)," in *2019 International Conference on Digitization (ICD)*, Nov. 2019, pp. 217–221. doi: 10.1109/ICD47981.2019.9105786.

[16] L. A. Silva, H. Sanchez San Blas, D. Peral García, A. Sales Mendes, and G. Villarubia González, "An Architectural Multi-Agent System for a Pavement Monitoring System with Pothole Recognition in UAV Images," *Sensors*, vol. 20, no. 21, Art. no. 21, Jan. 2020, doi: 10.3390/s20216205.

[17] A. Ji, X. Xue, Y. Wang, X. Luo, and L. Wang, "Image-based road crack risk-informed assessment using a convolutional neural network and an unmanned aerial vehicle," *Struct. Control Health Monit.*, vol. 28, no. 7, p. e2749, 2021, doi: 10.1002/stc.2749.

[18] J. Zhu, J. Zhong, T. Ma, X. Huang, W. Zhang, and Y. Zhou, "Pavement distress detection using convolutional neural networks with images captured via UAV," *Autom. Constr.*, vol. 133, p. 103991, Jan. 2022, doi: 10.1016/j.autcon.2021.103991.

[19] J. Zhong, J. Zhu, J. Huyan, T. Ma, and W. Zhang, "Multi-scale feature fusion network for pixel-level pavement distress detection," *Autom. Constr.*, vol. 141, p. 104436, Sep. 2022, doi: 10.1016/j.autcon.2022.104436.

[20] Y. Zhang *et al.*, "Road damage detection using UAV images based on multi-level attention mechanism," *Autom. Constr.*, vol. 144, p. 104613, Dec. 2022, doi: 10.1016/j.autcon.2022.104613.

[21] Z. Hong *et al.*, "Highway Crack Segmentation From Unmanned Aerial Vehicle Images Using Deep Learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3129607.

[22] W. S. Qureshi *et al.*, "An Exploration of Recent Intelligent Image Analysis Techniques for Visual Pavement Surface Condition Assessment," *Sensors*, vol. 22, no. 22, Art. no. 22, Jan. 2022, doi: 10.3390/s22229019.

[23] W. Cao, Q. Liu, and Z. He, "Review of Pavement Defect Detection Methods," *IEEE Access*, vol. 8, pp. 14531–14544, 2020, doi: 10.1109/ACCESS.2020.2966881.

[24] S. Jiang and J. Zhang, "Real-time crack assessment using deep neural networks with wall-climbing unmanned aerial system," *Comput.-Aided Civ. Infrastruct. Eng.*, vol. 35, no. 6, pp. 549–564, 2020, doi: 10.1111/mice.12519.

[25] K. Zhang, Y. Zhang, and H.-D. Cheng, "CrackGAN: Pavement Crack Detection Using Partially Accurate Ground Truths Based on Generative Adversarial Learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 1306–1319, Feb. 2021, doi: 10.1109/TITS.2020.2990703.

[26] D. Ai, G. Jiang, S.-K. Lam, P. He, and C. Li, "Computer vision framework for crack detection of civil infrastructure—A review," *Eng. Appl. Artif. Intell.*, vol. 117, p. 105478, Jan. 2023, doi: 10.1016/j.engappai.2022.105478.

[27] D. Ai, G. Jiang, L. Siew Kei, and C. Li, "Automatic Pixel-Level Pavement Crack Detection Using Information of Multi-Scale Neighborhoods," *IEEE Access*, vol. 6, pp. 24452–24463, 2018, doi: 10.1109/ACCESS.2018.2829347.

[28] W. Wang and C. Su, "Automatic concrete crack segmentation model based on transformer," *Autom. Constr.*, vol. 139, p. 104275, Jul. 2022, doi: 10.1016/j.autcon.2022.104275.

[29] S. Wang, C. Liu, and Y. Zhang, "Fully convolution network architecture for steel-beam crack detection in fast-stitching images," *Mech. Syst. Signal Process.*, vol. 165, p. 108377, Feb. 2022, doi: 10.1016/j.ymssp.2021.108377.

[30] W. Choi and Y.-J. Cha, "SDDNet: Real-Time Crack Segmentation," *IEEE Trans. Ind. Electron.*, vol. 67, no. 9, pp. 8016–8025, Sep. 2020, doi: 10.1109/TIE.2019.2945265.

[31] C. Li *et al.*, "CrackCLF: Automatic Pavement Crack Detection Based on Closed-Loop Feedback," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 6, pp. 5965–5980, Jun. 2024, doi: 10.1109/TITS.2023.3332995.

[32] H. Chu, W. Wang, and L. Deng, "Tiny-Crack-Net: A multiscale feature fusion network with attention mechanisms for segmentation of tiny cracks," *Comput.-Aided Civ. Infrastruct. Eng.*, vol. 37, no. 14, pp. 1914–1931, 2022, doi: 10.1111/mice.12881.

[33] Z. Fan *et al.*, "Ensemble of Deep Convolutional Neural Networks for Automatic Pavement Crack Detection and Measurement," *Coatings*, vol. 10, no. 2, Art. no. 2, Feb. 2020, doi: 10.3390/coatings10020152.

[34] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440. Accessed: Sep. 04, 2023. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation." arXiv, May 18, 2015. Accessed: Sep. 04, 2023. [Online]. Available: http://arxiv.org/abs/1505.04597

[36] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs." arXiv, May 11, 2017. Accessed: Sep. 04, 2023. [Online]. Available: http://arxiv.org/abs/1606.00915

[37] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 6230–6239. doi: 10.1109/CVPR.2017.660.

[38] B. Bosquet, D. Cores, L. Seidenari, V. M. Brea, M. Mucientes, and A. D. Bimbo, "A full data augmentation pipeline for small object detection based on generative adversarial networks," *Pattern Recognit.*, vol. 133, p. 108998, Jan. 2023, doi: 10.1016/j.patcog.2022.108998.

[39] Y. Liu, Q. Ren, J. Geng, M. Ding, and J. Li, "Efficient Patch-Wise Semantic Segmentation for Large-Scale Remote Sensing Images,"

Sensors, vol. 18, no. 10, Art. no. 10, Oct. 2018, doi: 10.3390/s18103232.

[40] J. Liu *et al.*, "Automated pavement crack detection and segmentation based on two-step convolutional neural network," *Comput.-Aided Civ. Infrastruct. Eng.*, vol. 35, no. 11, pp. 1291–1305, 2020, doi: 10.1111/mice.12622.

[41] Ö. Özdemir and E. B. Sönmez, "Weighted Cross-Entropy for Unbalanced Data with Application on COVID X-ray images," in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Oct. 2020, pp. 1–6. doi: 10.1109/ASYU50717.2020.9259848.

[42] T. A. Soomro, A. J. Afifi, J. Gao, O. Hellwich, M. Paul, and L. Zheng, "Strided U-Net Model: Retinal Vessels Segmentation using Dice Loss," in *2018 Digital Image Computing: Techniques and Applications (DICTA)*, Dec. 2018, pp. 1–8. doi: 10.1109/DICTA.2018.8615770.

[43] M. S. Hossain, J. M. Betts, and A. P. Paplinski, "Dual Focal Loss to address class imbalance in semantic segmentation," *Neurocomputing*, vol. 462, pp. 69–87, Oct. 2021, doi: 10.1016/j.neucom.2021.07.055.

[44] S. Sang, Y. Zhou, M. T. Islam, and L. Xing, "Small-Object Sensitive Segmentation Using Across Feature Map Attention," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 6289–6306, May 2023, doi: 10.1109/TPAMI.2022.3211171.

[45] L. Wu *et al.*, "Data augmentation based on multiple oversampling fusion for medical image segmentation," *PLOS ONE*, vol. 17, no. 10, p. e0274522, Oct. 2022, doi: 10.1371/journal.pone.0274522.

[46] D. Ghosh and N. Kaabouch, "A survey on image mosaicing techniques," *J. Vis. Commun. Image Represent.*, vol. 34, pp. 1–11, Jan. 2016, doi: 10.1016/j.jvcir.2015.10.014.

[47] J. Davis, "Mosaics of scenes with moving objects," in *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No.98CB36231)*, Jun. 1998, pp. 354–360. doi: 10.1109/CVPR.1998.698630.

[48] A. Flores and S. Belongie, "Removing pedestrians from Google street view images," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, Jun. 2010, pp. 53–58. doi: 10.1109/CVPRW.2010.5543255.

[49] A. Murodjon and T. Whangbo, "A method for manipulating moving objects in panoramic image stitching," in *2017 International Conference on Emerging Trends & Innovation in ICT (ICEI)*, Feb. 2017, pp. 157–161. doi: 10.1109/ETIICT.2017.7977029.

[50] W. Xue, Z. Zhang, and S. Chen, "Ghost Elimination via Multi-Component Collaboration for Unmanned Aerial Vehicle Remote Sensing Image Stitching," *Remote Sens.*, vol. 13, no. 7, Art. no. 7, Jan. 2021, doi: 10.3390/rs13071388.

[51] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi, "EdgeConnect: Structure Guided Image Inpainting using Edge Prediction," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Oct. 2019, pp. 3265–3274. doi: 10.1109/ICCVW.2019.00408.

[52] J. Pyo, Y. G. Rocha, A. Ghosh, K. Lee, G. In, and T. Kuc, "Object Removal and Inpainting from Image using Combined GANs," in *2020 20th International Conference on Control, Automation and Systems (ICCAS)*, Oct. 2020, pp. 1116–1119. doi: 10.23919/ICCAS50221.2020.9268330.

[53] X. Zhang, D. Zhai, T. Li, Y. Zhou, and Y. Lin, "Image inpainting based on deep learning: A review," *Inf. Fusion*, vol. 90, pp. 74–94, Feb. 2023, doi: 10.1016/j.inffus.2022.08.033.

[54] G. Zhang *et al.*, "UAV Low-Altitude Aerial Image Stitching Based on Semantic Segmentation and ORB Algorithm for Urban Traffic," *Remote Sens.*, vol. 14, no. 23, Art. no. 23, Jan. 2022, doi: 10.3390/rs14236013.

[55] S. Javaid *et al.*, "Communication and Control in Collaborative UAVs: Recent Advances and Future Trends," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 6, pp. 5719–5739, Jun. 2023, doi: 10.1109/TITS.2023.3248841.

[56] S. Jarahizadeh and B. Salehi, "A Comparative Analysis of UAV Photogrammetric Software Performance for Forest 3D Modeling: A Case Study Using AgiSoft Photoscan, PIX4DMapper, and DJI Terra," *Sensors*, vol. 24, no. 1, Art. no. 1, Jan. 2024, doi: 10.3390/s24010286.

[57] M. Brown and D. G. Lowe, "Automatic Panoramic Image Stitching using Invariant Features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, Aug. 2007, doi: 10.1007/s11263-006-0002-3.

[58] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Comput. Med. Imaging Graph.*, vol. 95, p. 102026, Jan. 2022, doi: 10.1016/j.compmedimag.2021.102026.

[59] LikeLy-Journey, "PyTorch for Semantic Segmentation." Aug. 25, 2023. Accessed: Sep. 05, 2023. [Online]. Available: https://github.com/LikeLy-Journey/SegmenTron

[60] X. Weng, Y. Huang, and W. Wang, "Segment-based pavement crack quantification," *Autom. Constr.*, vol. 105, p. 102819, Sep. 2019, doi: 10.1016/j.autcon.2019.04.014.

[61] R. Augustauskas and A. Lipnickas, "Improved Pixel-Level Pavement-Defect Segmentation Using a Deep Autoencoder," *Sensors*, vol. 20, no. 9, Art. no. 9, Jan. 2020, doi: 10.3390/s20092557.

Jinhuan Shan received the B.S. degree from the Department of Civil Engineering, Zhejiang Sci-Tech University, China, in 2017. He is currently pursuing the Ph.D. degree in the Department of Road and Railway Engineering at Chang'an University, China.

His research interests include deep learning-based pavement distress detection, as well as green pavement materials and intelligent road infrastructure.

Wei Jiang received his M.S. and Ph.D. degrees from Chang'an University in 2008 and 2011, respectively.

He is currently a professor at the School of Highway. His research has focused on the theories and methods of green pavement materials and structures, as well as the green energy transformation of pavements.

Dr. Jiang led over 13 national and provincial projects, including the National Science Fund for Excellent Young Scientists, National Key R&D Program of China, the National Natural Science Foundation, the Fok Ying-Tong Education Foundation for Young Teachers in the Higher Education Institutions of China, etc.

Yue Huang obtained his Ph.D. degree from Newcastle University in 2007.

He is currently an associate professor at the Institute for Transport Studies (ITS), University of Leeds, UK. His research areas include life cycle assessment, pavement evaluation and recycling, and road safety.

Dr. Huang is a Chartered Engineer (CEng) and a Fellow of the Higher Education Academy (FHEA).

Dongdong Yuan received his M.S. and Ph.D. degrees from Chang'an university, Shaanxi, China, in 2018 and 2024, respectively. He is currently a lecturer in the Department of Road and Railway Engineering at Chang'an University.

His research focuses on thermoelectric asphalt pavement and modified asphalt.

Yaohan Liu received her B.S. degree from Shandong Jianzhu University, Shandong, China, in·2016. She is currently pursuing her Ph.D. degree in the Department of Road and Railway Engineering at Chang'an University, Shaanxi, China.

Her research interests include intelligent pavement monitoring and maintenance decision-making.