Article

A Novel UAV-based Road Damage Detection Algorithm with Lightweight Convolution and Attention Mechanism

Liang Chen [1], Peishu Wu [1], Weilong Tan [2], Han Li [2], Haonan Chen [2], and Nianyin Zeng [1,] *

[1] School of Aerospace Engineering, Xiamen University, Fujian 361102, China
[2] College of Electrical Engineering and Automation, Fuzhou University, Fujian 350108, China
* Correspondence: zny@xmu.edu.cn

Received: 17 June 2025
Accepted: 25 September 2025
Published: 16 December 2025

Abstract: In this paper, a novel attention- and lightweight convolution-based road damage detection network (ALC-Net) is proposed to address the trade-off between accuracy and real-time performance in processing unmanned aerial vehicle (UAV) imagery. Specifically, a lightweight module that integrates ghost convolution with the squeeze-and-excitation (SE) attention mechanism is designed, which effectively reduces model parameters while enhancing detection accuracy. The focus module is introduced to perform downsampling and channel-wise concatenation of input images, thereby enriching feature diversity. Furthermore, a coordinate attention mechanism is incorporated to aggregate horizontal and vertical spatial information, emphasizing subtle road damage characteristics. The proposed ALC-Net is comprehensively evaluated on a UAV-captured road damage dataset, demonstrating superior detection performance compared to other state-of-the-art approaches. The contributions of key components in ALC-Net are also validated through ablation studies, confirming their ability to enhance feature extraction capabilities while reducing computational complexity. Additionally, experiments on non-UAV road damage datasets further reveal the robust generalization capability of ALC-Net, exhibiting substantial potential for broader applications.

Keywords: road damage detection; unmanned aerial vehicle (UAV); attention; lightweight

1. Introduction

Road damage detection involves localizing and identifying surface defects such as cracks and potholes via image or sensor data to assess road conditions and guide maintenance efforts [1,2]. Traditional methods rely on manual inspections or road inspection vehicles, which suffer from inefficiency and high costs. Therefore, intelligent detection algorithms that can be broadly categorized into vibration sensors-based, 3D sensors-based, and images-based [3] are developed. Within them, images-based methods have become mainstream due to the advancements in computer vision. But existing road damage images primarily derive from vehicle-mounted or handheld cameras, resulting in sophisticated data collection processes and inconsistent image quality. Along with the rapid development of low-altitude unmanned aerial vehicle (UAV), integrated schemes, which acquire high-resolution road images via UAVs and apply advanced detection algorithms to classify and localize damage, gradually stand out. Benefiting from the advantages of UAV, this paradigm enhances image quality, expands coverage and simplifies data collection. In addition, real-time accurate detection can also be realized through advanced object detection algorithms, which remains a critical research focus in academia.

In the research of object detection, accuracy and real-time performance are essential yet challenging to balance. For instance, hybrid models combining Transformers and convolutional neural networks (CNNs) have been proposed to address complex backgrounds and multi-scale targets, but sacrifice inference speed for improved accuracy [4]. Distributed edge-cloud collaborative frameworks [5] and edge-embedded lightweight algorithms with attention mechanisms have been introduced to accelerate inference, but destroy partial performance. And specific to the research of road damage detection, further unique challenges are posed due to complex morphologies, background interference, and small crack sizes. To address these issues, methods such as enhanced multi-target extraction via improved YOLOv8 [6] (You Only Look Once) and GAN-based texture synthesis for data augmentation [7] have been proposed. Although the detection effect is improved, additional convolutional or attention modules increases model complexity and computational burdens. Thus, lightweight techniques like depthwise separable convolution [5], tensor decomposition [8], and knowledge distillation [9,10] have also been researched, but they inevitably degrade accuracy and necessitate manual verification, leading to limited practical utility.

Building upon the discussions above, an innovative road damage detection model ALC-Net is proposed based on YOLOv11 in this paper, tailored for UAV-captured road scenes to achieve precise and efficient detection. Particularly, a focus module is introduced to downsample and concatenate input images, reducing interference from irrelevant high-resolution details while enriching feature diversity. Then, a coordinate attention mechanism [11] is integrated to aggregate global information and emphasize fine crack features. After that, ghost convolutions [12] are adopted to replace standard convolutions to minimize parameters and facilitate lightweight deployment, while mitigating performance degradation since their inherent feature. Additionally, a squeeze-and-excitation (SE) module [13] is incorporated to capture channel-wise dependencies, enhance global feature integration, and amplify critical ghost feature maps through adaptive channel recalibration. Consequently, the model can not only adapt to the lightweight deployment on the UAV, but also ensure robust and excellent detection performance. It is worth mentioning that YOLOv11 is selected as the baseline due to its well-balanced architecture that offers a strong trade-off between speed and accuracy in this task, outperforming models such as YOLOv8 and YOLOv10 and providing a robust yet straightforward foundation for integrating the proposed strategies.

The main contributions of this article are summarized as follows.

- A novel high-accuracy and lightweight detection network ALC-Net is proposed, specifically tailored for the challenging task of road damage detection in high-resolution UAV imagery with enhanced efficiency.
- A lightweight convolutional block is designed and implemented, integrating ghost convolution and SE attention mechanisms to achieve significant parameter reduction while simultaneously enhancing discriminative feature representation.
- Specific architectural components, namely the focus module and coordinate attention mechanism, are strategically incorporated, bolstering the feature extraction and spatial awareness capabilities of the network for robust damage detection.

The remainder of this paper is organized as follows. Related works are reviewed in Section Ⅱ. The proposed ALC-Net is comprehensively elaborated in Section Ⅲ. Experimental results are presented in Section Ⅳ, and conclusions are finally provided in Section Ⅴ.

2. Related Work

In this section, a comprehensive review of the current state-of-the-art object detection models relevant to our work is provided. Furthermore, research on personalized models tailored specifically for the unique challenges of road damage detection is also highlighted.

2.1. Overview of Object Detection Models

Object detection models, which aim to locate and classify objects within images, can be broadly categorized based on their detection and recognition strategies into two main types: two-stage and one-stage detectors. Two-stage detectors, such as the prominent R-CNN series including Faster R-CNN [14], Mask R-CNN [15], and Cascade R-CNN [16], partition the object detection process into two sequential stages. The initial stage involves proposing a sparse set of candidate object locations, which are often generated by a dedicated Region Proposal Network (RPN). The subsequent stage then classifies each proposed region and refines its bounding box. Faster R-CNN revolutionizes this approach by integrating the RPN into the neural network, significantly accelerating proposal generation compared to earlier methods like selective search. Mask R-CNN extends this by adding instance segmentation, while Cascade R-CNN improves localization accuracy through cascaded refinement stages with increasing IoU thresholds. Although two-stage detectors generally achieve high accuracy and robustness, their sequential nature and the processing of multiple proposals result in higher computational cost and slower inference speed. These characteristics typically render them less suitable for deployment in resource-constrained environments like UAVs.

In contrast, one-stage detectors directly predict object bounding boxes and class probabilities across the entire image in a single forward pass. This end-to-end approach eliminates the explicit region proposal stage, leading to significantly faster inference and simpler architectures. Representative algorithms include the widely-used YOLO series [17,18], RetinaNet [19] and SSD (Single Shot MultiBox Detector) [20]. In YOLO, a grid-based prediction method is introduced, where each grid cell is responsible for detecting objects centered within it. SSD improves upon this by utilizing multi-scale feature maps from different network layers and employing a set of predefined default boxes at various scales and aspect ratios to handle objects of different sizes effectively. Subsequent iterations of the YOLO series have continuously pushed the boundaries of speed-accuracy trade-off through architectural advancements like improved backbones, sophisticated feature fusion networks, and advanced training techniques [21,22]. In addition, research on intelligent optimization algorithms has also laid a solid foundation for further improving detection accuracy [23–29]. These developments have established one-stage detectors as compelling choices for applications that require high-speed performance.

Despite the impressive performance on a wide range of general object detection tasks has been achieved, these generic models often require task-specific adaptations to achieve optimal results in specialized domains with unique object characteristics or challenging imaging conditions [30–32]. One such challenging domain is object detection in high-resolution aerial images captured by UAVs. This task presents several unique difficulties, such as the high resolution of imagery covering vast areas, the prevalence of tiny objects occupying only a few pixels, significant object scale variations, and so on. To address these issues, extensive research has been conducted into specialized methods for UAV object detection [33–35]. For instance, to improve small-target perception while reducing parameters, the large-object detection heads in YOLOv5 have been replaced with specialized tiny-object prediction heads [36]. Similarly, a feature optimization fusion (FOF) module has been introduced in [37], which combines deformable [38] and partial convolutions [39], to refine multi-scale feature integration.

While these methods have shown promising results on general UAV image datasets, the objects detected are typically common categories like vehicles or buildings. The distinct visual properties of road damage, which differs significantly from these targets, mean that the generalization performance of these methods directly to UAV-based road damage detection is not guaranteed, highlighting the need for dedicated research in this specific application area.

2.2. Overview of Road Damage Detection Models

Road damage (see Figure 1), such as cracks (including longitudinal, transverse, and alligator patterns) and potholes, possesses characteristics vastly different from common objects. These damages exhibit diverse shapes, sizes ranging from hairline cracks to large surface defects, and often have low contrast against varied road textures. Notably, cracks frequently present extreme length-to-width ratios, posing a unique challenge for detection. Various techniques tailored to these attributes have been explored to specifically address road damage detection in recent research. Some methods adapt general object detection frameworks to better handle the multi-scale nature and diverse types of damage. For example, in [40], an improved YOLOv5 has been integrated with ensemble learning to boost detection capability for multi-scale damage and validated its performance on datasets including disaster-induced road damage. To address the critical issue of fine detail loss in deeper network layers, a cross-layer attention mechanism has been proposed in [41], which adaptively reinforces shallow features and ensures ample edge details for reference. Furthermore, to tackle the challenge of detecting damages with extreme aspect ratios, in [42], the standard downsampling in YOLOX has been replaced with the refined switchable atrous convolution [43], which allows the network to adaptively adjust its receptive field shape and size. Beyond these specific approaches, there are other innovations [44,45] available for the detection accuracy or model efficiency.

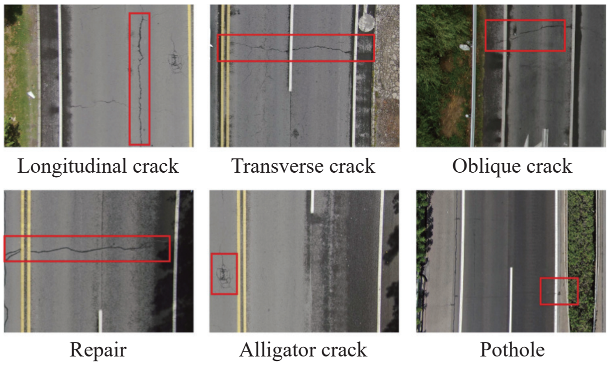

Figure 1. Types of road damage.

In summary, while significant progress has been made in both generic object detection and its application to challenging domains like UAV imagery, as well as in specialized road damage detection, a notable research gap exists concerning the task of detecting road damage from the specific viewpoint of low-altitude UAVs. A substantial portion of existing road damage research focuses on ground-level data, which differs significantly from aerial-view imagery in terms of perspective, lighting, and the scale and appearance of damage relative to the scene. Therefore, this study addresses this gap by focusing on constructing a real-time and efficient detection model specifically designed for low-altitude UAV aerial-view imagery to achieve rapid and accurate identification of various types of road damage, thereby contributing to the advancement of technology in this critical domain.

3. Methodology

In this section, the overall framework of the proposed ALC-Net is elaborated, and the critical improvements are also detailed, as well as their mechanisms and advantages for road damage detection.

3.1. Overall Framework of ALC-Net

The overall framework of the ALC-Net is illustrated in Figure 2, which builds upon the YOLOv11 architecture with targeted improvements for road damage detection. In general, the model is divided into three components: backbone, neck, and detection head. The input image sequentially passes through these components to generate detection results, including bounding boxes and class probabilities. Firstly, the backbone is constructed to extract multi-scale features from the input image. Specifically, the input image is first passed through the focus module for downsampling and stitching, and then followed by four convolution operations combined with the attention mechanism and the C3K2 module (a cross stage partial-style bottleneck module). It is noticeable that these operations incorporate coordinate attention, ghost convolution, SE attention, and spatial pyramid pooling fast (SPPF), enhancing focus on target regions and fine-grained damage. After that, the C2PSA module (cross stage partial with pyramid squeeze attention) is designed to further enhance features prepared for the next process.

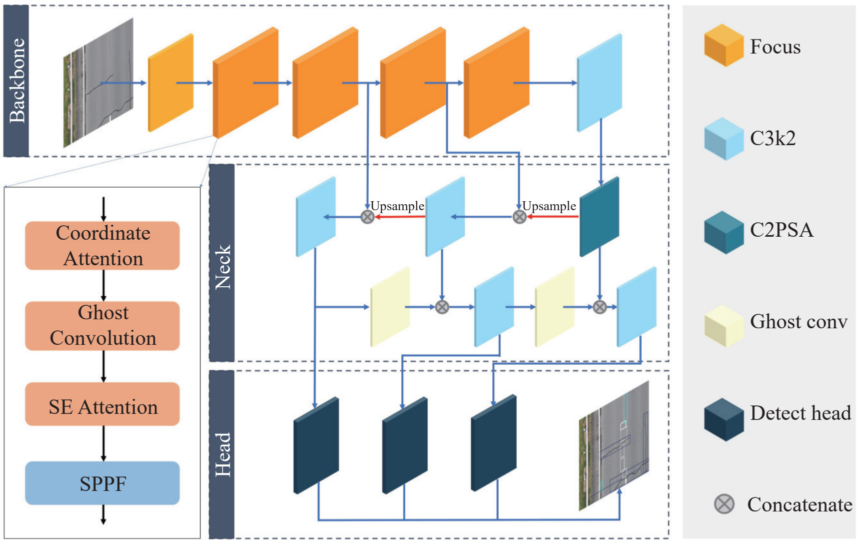

Figure 2. Overall framework of proposed ALC-Net. It is divided into three major parts: backbone, neck and head. The enhancement strategies are conducted on the backbone.

Secondly, the neck adopts a feature pyramid network (FPN) structure to fuse multi-scale feature maps from the backbone. In particular, features from the backbone are upsampled and convolved to generate three distinct scales, which are then concatenated channel-wise with corresponding features in the backbone to form fused multi-scale representations. Thirdly, the detection head consists of three detection units, which are designed for processing features from the neck at different scales and generating detection results, respectively. Finally, detections are aggregated and visualized on the original image.

3.2. Focus Module

Considering the tiny and localized characteristics of the road damage, the focus module is introduced and positioned at the first layer of the backbone, which enhances the perception capacity to subtle damage by strategically downsampling and reorganizing input images. The implementation details are shown in Figure 3, let the input image be $I \in R^{H \times W \times C}$. The module first performs sampling with a stride of 2 in both spatial dimensions, effectively partitioning I into non-overlapping 2 × 2 patches and extracting one pixel from each patch based on its position. This generates four sub-feature maps denoted as $F_0, F_1, F_2, F_3 \in R^{\frac{H}{2} \times \frac{W}{2} \times C}$, which are indicated by different colors in Figure 3 and can be formulated as:

$$\begin{aligned} F_0(i,j,c) &= I(2i,2j,c), \\ F_1(i,j,c) &= I(2i+1,2j,c), \\ F_2(i,j,c) &= I(2i,2j+1,c), \\ F_3(i,j,c) &= I(2i+1,2j+1,c) \end{aligned} \tag{1}$$

where $0 \leqslant i < H/2$, $0 \leqslant j < W/2$, $0 \leqslant c < C$. It can be seen that the four sub-feature maps are changed to one-quarter of the original image. Then, the four sub-feature maps are concatenated along the channel dimension to output the feature map $F_{out} \in R^{\frac{H}{2} \times \frac{W}{2} \times (4C)}$ enriched with fine details, which can be denoted as:

$$F_{out}(i,j,:) = \text{Concat}\left[F_0(i,j,:), F_1(i,j,:), F_2(i,j,:), F_3(i,j,:)\right] \tag{2}$$

With this operation, the focus module reorganizes the spatial information of the original image into the channel dimension. This processing converts the lost spatial information into channel information while achieving the same downsampling effect as the convolution with a stride of 2, which allows subsequent convolutional layers to process this local detail information in a larger receptive field, thereby enhancing the ability of the network to perceive small, local breakage. Compared with directly using stride-2 convolution, the focus module is computationally more efficient and better preserves edge and detail information, which is crucial for detecting road damage.

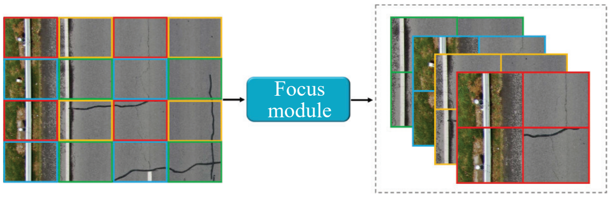

Figure 3. The structure of focus module. It partitions the input into four sub-feature maps, which are then concatenated along the channel dimension.

3.3. Coordinate Attention Mechanism

The coordinate attention mechanism can dynamically adjust the weight size of different positions in the multi-scale feature map, which is another vital module of ALC-Net. Road damage is manifested as local features in the feature map, and often presents transverse or longitudinal distribution patterns. Therefore, in order to enhance the sensitivity of the model to spatial-channel information, make it pay more attention to the critical area and improve the learning of directional feature patterns, coordinate attention is introduced, as illustrated in Figure 4.

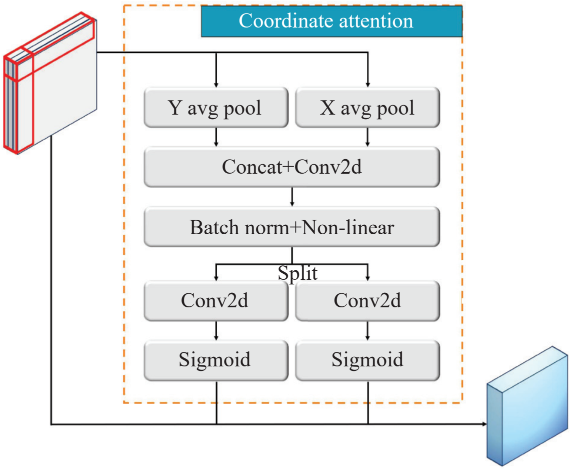

Figure 4. The structure of coordinate attention mechanism. It performs average pooling along horizontal and vertical axes, followed by convolutional transformations and sigmoid activation to generate spatial attention weights.

Assume the input feature map is $X \in R^{H \times W \times C}$, where H is height, W is width, and C is channel size. The workflow of coordinate attention mechanism is as follows. Firstly, average pooling including X Avg Pool and Y Avg Pool is performed along horizontal and vertical directions of the input feature to capture the long-range dependencies in two spatial directions:

$$Z_X = \frac{1}{W} \sum_{0 \leqslant w < W} X(:,w,:), Z_Y = \frac{1}{H} \sum_{0 \leqslant h < H} X(h,:,:) \tag{3}$$

where Z_X and Z_Y represent the pooling results in horizontal and vertical directions, respectively.

Then, two directional features are concatenated, and encoded via a 1 × 1 convolution, followed with the batch-normalization and ReLU operation. The obtained feature $\widetilde{f}$ associates the directional information, and can be denoted as follows:

$$\widetilde{f} = \mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{Conv}_{1\times1}\left(\mathrm{Concat}\left[Z_X, Z_Y\right]\right)\right)\right) \tag{4}$$

Next, the encoded feature $\widetilde{f}$ is split into two independent feature maps ($\widetilde{f}_X$ and $\widetilde{f}_Y$) along the spatial dimension. After that, through the 1 × 1 convolution and sigmoid normalization operation, the channel size is restored to C and the directional attention weights are obtained, denoted as:

$$S_X = \sigma\left(\mathrm{Conv}_{1\times1}(\widetilde{f}_X)\right),\ \ S_Y = \sigma\left(\mathrm{Conv}_{1\times1}(\widetilde{f}_Y)\right) \tag{5}$$

Finally, the input feature X is reweighted by element-wise multiplication with the obtained directional attention weights to realize the effective fusion of channels and spatial location information. Note that both S_X and S_Y are broadcast to match the spatial dimensions of X. The final output feature is:

$$Y_{CA} = X \cdot S_X \cdot S_Y \tag{6}$$

3.4. Ghost Convolution with SE Attention

To balance lightweight deployment and performance on UAVs, standard convolution layers in the network are replaced with ghost convolution modules, which is illustrated in Figure 5. Similarly, let the input feature map be $X \in R^{H\times W\times C}$. The ghost convolution operation primarily consists of two steps. Firstly, a conventional convolution with m output channels is applied to generate intrinsic feature maps $Y' \in R^{H\times W\times m}$. Then, they are mapped s times to obtain ghost feature maps, one of which is an identity mapping ($\Phi_{i,1}$), while the remaining are typically mapped using depthwise convolution ($\Phi_{i,j}$ for $j = 2, \cdots, s$). This generates $m \times (s-1)$ ghost feature maps:

$$\widetilde{y}_{i,j} = \Phi_{i,j}\left(y'_i\right) \tag{7}$$

where $i = 1, \cdots, m,\ j = 2, \cdots, s$.

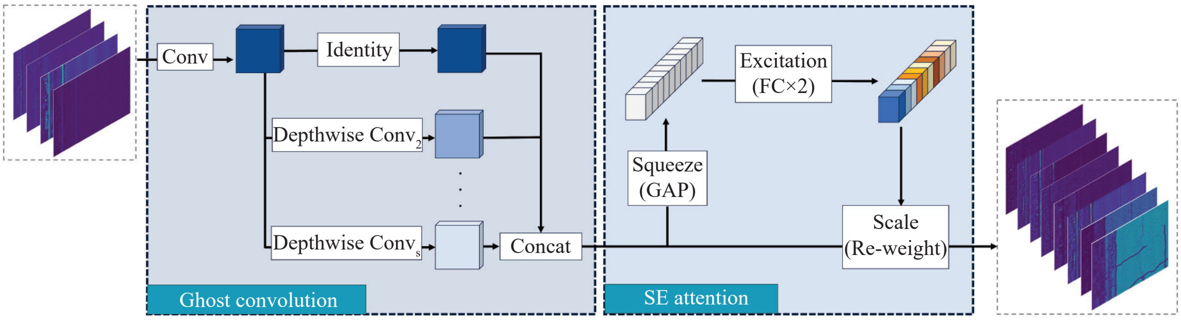

Figure 5. The structure of ghost convolution with squeeze-and-excitation attention. First, numerous ghost feature maps are generated by ghost convolution. Then, the redundant features are suppressed by the SE attention.

Finally, the intrinsic feature maps and all ghost feature maps are concatenated along the channel dimension to form the final output feature map, as expressed in following formula:

$$Y_{GC} = \mathrm{Concat}\left(\{y'_i\}_{i=1}^{m}, \{\widetilde{y}_{i,j}\}_{i=1,j=2}^{m,s}\right) \tag{8}$$

The purpose of ghost convolution is to generate abundant feature maps that are similar to the intrinsic feature map, which are believed to reveal the potential essential features in the image, so as to enhance the feature extraction ability of the model. To generate features of the same dimension size, ghost convolution requires significantly less parameters than the standard convolution, and the performance will not decrease excessively.

Although lightweighting can be achieved by the ghost convolution, feature redundancy is inevitable due to the numerous similar feature maps being generated. In addition, there are some feature maps that are unrelated to the feature essence, which may be a major factor leading to the decline in model performance. Based on these considerations, a SE module is appended after each ghost convolution, as also shown in Figure 5. Specifically, the features after ghost convolution are squeezed as a numerical value by global average pooling in each channel to represent respective global spatial information. Next, to comprehensively capture the dependency relationships from the squeezed information, an excitation operation consisting of two fully connected layers is conducted, and the attention weights of each channel is obtained. Finally, the scale operation is utilized to reweight the input feature map according to the extracted attention weights, which enhances the model to focus on the essential features by emphasizing informative channels and suppressing less relevant or redundant ones, thereby mitigating the potential performance drawbacks caused by ghost convolution. The overall process of SE module is formulated as follows:

$$S = \sigma(W_2 \cdot \mathrm{ReLU}(W_1 \cdot \mathrm{GAP}(Y_{GC}))) \tag{9}$$

$$\hat{Y} = Y_{GC} \cdot S \tag{10}$$

where $\mathrm{GAP}(\cdot)$ represents globally average pooling, W_1 and W_2 are the weights of linear layers, S is the result after squeeze and excitation operation, while $\hat{Y}$ is the reweighted output.

4. Experiments

In this section, the experimental results and analysis of the proposed ALC-Net on the benchmark datasets are presented. To begin with, the experimental settings are briefly described.

4.1. Experimental Settings

To ensure the fairness of the experiments, all of them are conducted on a personal computer with the following specifications: Windows 11 OS, NVIDIA GeForce RTX 4070 Ti SUPER GPU. The models are implemented in Python 3.11 programming language using PyTorch framework.

The primary benchmark dataset, UAV-PDD2023 [46], contains 2,440 high-resolution UAV-captured road damage images annotated with six damage classes: longitudinal cracks, transverse cracks, oblique cracks, alligator cracks, patching, and potholes. The dataset is then split into training, validation, and testing sets at an 8:1:1 ratio, as shown in Table 1. Additionally, to further evaluate the generalization performance, the RDD2022 dataset [47], which contains longitudinal cracks, transverse cracks, alligator cracks, and potholes, is also introduced. It is worth noting that only the subsets China drone, China motorbike, and United States in the RDD2022 are selected to conduct experiments. Among them, the images in China drone are UAV-captured, while the others are from the smartphones or vehicle-mounted cameras, ensuring the diversity of the data source. Similarly, three subsets are partitioned in an 8:1:1 ratio, as presented in Table 1.

Table 1 The distribution of datasets used in this paper

Dataset	Training set	Validation set	Testing set
UAV-PDD2023 [46]	1952	244	244
China drone [47]	1535	192	192
China motorbike [47]	1547	193	194
United States [47]	3844	480	481

Considering the constraints of computing resources and experimental fairness, training parameters are set uniformly, where epochs are 300, batch-size is 8, initial learning rate is 0.01 (decaying to 1% of the initial rate in final), and Adam optimizer is chosen. Before training, a series of preprocessing are conducted. Firstly, the images in UAV-PDD2023 and RDD2022 are resized to 1024 × 1024 and 320 × 320 to balance resolution and efficiency, respectively. Next, pixel values are normalized to accelerate convergence. Finally, data augmentation operations such as flipping, rotation, and scaling are adopted to mitigate overfitting. In order to quantitatively evaluate the model performance, the following metrics are adopted, including $Precision$, $Recall$, mAP_{50}, mAP_{50-95}. Specifically, $Precision$ represents the correct proportion of the targets being predicted as positive, while $Recall$ represents the proportion of the positive targets being correctly predicted, mAP_{50} indicates the mAP (mean average precision) over IoU (Intersection over Union) at 0.5, and mAP_{50-95} indicates mAP over IoU from 0.5 to 0.95 with an interval of 0.05. Meanwhile, the number of parameters ($Params$), frames per second (FPS) and giga floating-point operations per second ($GFLOPs$) are also adopted to assess the model complexity.

4.2. Comparison with SOTA Methods

To thoroughly evaluate the performance of the proposed ALC-Net, comparative experiments are conducted on the UAV-PDD2023 dataset against several state-of-the-art (SOTA) object detection methods. The selected counterparts include representative two-stage detectors such as Faster R-CNN [14] and Cascade R-CNN [16], as well as prominent one-stage detectors like RetinaNet [19] and various lightweight variants from the YOLO series (YOLOv5n, YOLOv6n [17], YOLOv8n, YOLOv10n [18], and YOLOv11n). The overall quantitative results comparing these models are presented in Table 2, with the optimal values for each metric highlighted in bold.

Table 2 Comparison results with SOTA models on UAV-PDD2023 dataset

Model	*Precision*(%)	*Recall*(%)	mAP_{50}(%)	mAP_{50-95}(%)
RetinaNet	46.4	48.4	56.5	33.0
Cascade R-CNN	68.2	64.8	63.7	41.6
Faster R-CNN	59.6	64.6	64.9	41.3
YOLOv5	67.5	53.7	57.8	29.3
YOLOv6	59.8	49.6	48.2	22.3
YOLOv8	69.2	60.7	66.9	38.1
YOLOv10	67.5	60.2	64.8	37.8
YOLOv11	72.7	62.0	67.4	36.9
ALC-Net	**76.0**	**70.1**	**75.0**	**42.1**

As demonstrated in Table 2, the proposed ALC-Net achieves superior performance across all evaluated metrics on the challenging UAV-PDD2023 dataset. Specifically, ALC-Net obtains a *Precision* of 0.760, a *Recall* of 0.701, a mAP_{50} of 0.750, and a mAP_{50-95} of 0.421. These results indicate that the ALC-Net not only accurately identifies road damage but also successfully detects a large proportion of the actual damage instances. The high mAP_{50} value reflects strong overall detection performance, while the mAP_{50-95} value further assesses the model ability to accurately localize damage with tight bounding boxes across varying strictness levels.

Comparing the performance of ALC-Net to the evaluated SOTA models, it significantly outperforms all counterparts. Notably, when compared to the suboptimal performing method in Table 2, ALC-Net achieves a substantial improvement: *Precision* is enhanced by 0.033, *Recall* by 0.053, mAP_{50} by a considerable 0.076, and mAP_{50-95} by 0.005. These quantitative findings robustly demonstrate the effectiveness of the proposed model for UAV-based road damage detection and underscore its superiority among numerous advanced general-purpose detection models.

To provide an intuitive qualitative comparison of detection performance, partial visualization results from ALC-Net and the baseline YOLOv11 are presented in Figure 6. These figures showcase challenging scenarios from the dataset. Upon visual inspection and comparison with YOLOv11, ALC-Net demonstrates superior detection quality. Specifically, ALC-Net is shown to generate fewer false positive detections and fewer false negative detections (see row 1 of Figure 6). Furthermore, ALC-Net exhibits better capability in detecting subtle or fine cracks, particularly when they appear in sophisticated scenes with complex backgrounds or challenging illumination conditions, which are often missed by the baseline (see row 2 in Figure 6). The visualizations also indicate that ALC-Net provides higher confidence scores and more accurate bounding boxes for identified crack categories, reflecting improved discriminative power for these critical damage types (see row 3 in Figure 6). The clear visual improvements observed in these examples align well with the quantitative metric results presented in Table 2, collectively providing strong evidence that verifies the superior detection performance of the proposed ALC-Net for UAV-based road damage detection.

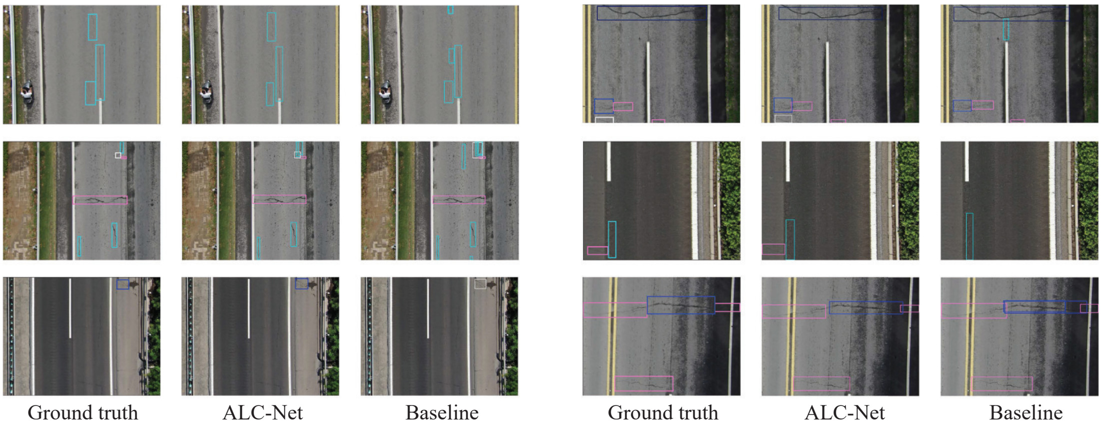

Figure 6. Qualitative comparison of detection results between the proposed ALC-Net and the baseline model on challenging images from the UAV-PDD2023 dataset.

4.3. Ablation Studies

To further validate the effectiveness of each module adopted in the model and analyze their contributions, ablation studies are performed on the UAV-PDD2023 dataset. In the experiments, the focus module, coordinate attention (CA), ghost convolution and SE attention are successive added to the baseline YOLOv11 for evaluation, and the results are shown in Figure 7. On the whole, the optimal performance is still achieved by the ALC-Net, which indicates the positive combination effect of each module that can collaboratively enhance the detection capacity. Comparing the results in detail, it can be seen that each module has made significant contributions. Firstly, the introduction of focus module on the baseline improves most metrics because the downsampling operation makes the fine-grained features prominent, which is convenient for the extraction and understanding by the model. Then, the coordinate attention mechanism is added, and mAP_{50} and mAP_{50-95} metrics both increase, indicating the promotion of its directional feature mining and fusion ability to the detection of road cracks, especially for transverse and longitudinal cracks detection. But the slight decline in $Precision$, $Recall$ also signifies the relatively weak effect of coordinate attention to detect other damage categories, which can be further researched in future.

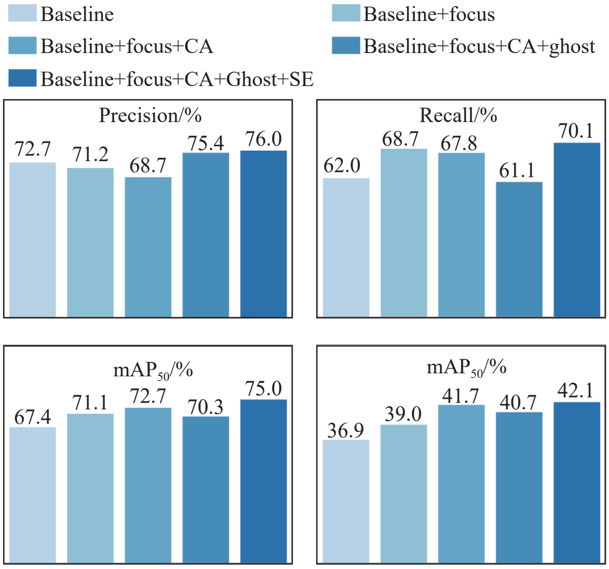

Figure 7. Results of ablation studies on UAV-PDD2023 dataset.

Subsequently, the ghost convolution is selected to replace the standard convolution module. Experiment results show that model performance reduces slightly and $Params$ reduces by 12% (see Table 4) simultaneously, which is consistent with the characteristics of this module. The outcomes of sacrificing tiny accuracy for model lightweighting are acceptable and as expected. Finally, the SE attention mechanism is added after ghost convolution, and each metric is remarkably improved. It demonstrates that the SE module is adaptive to the task of road damage detection, and also makes up for the shortcomings of the ghost convolution by adjusting the weight distribution of ghost feature maps and reduce the influence of the irrelevant feature maps.

4.4. Generalization Analysis

To validate the generalization capability of the proposed ALC-Net across diverse scenarios and data sources, three subsets from the RDD2022 dataset are selected for experimentation. The RDD2022 dataset provides images collected from various regions and under different conditions, offering a valuable testbed for generalization. The final results on these subsets are shown in Table 3.

Table 3 Comparison results with baseline on four datasets

Dataset	Model	$Precision$(%)	$Recall$(%)	mAP_{50}(%)	mAP_{50-95}(%)
UAV-PDD2023	YOLOv11	72.7	62.0	67.4	36.9
	ALC-Net	**76.0**	**70.1**	**75.0**	**42.1**
China drone	YOLOv11	66.2	44.2	49.3	25.8
	ALC-Net	**67.9**	**47.4**	**51.7**	**27.2**
China motorbike	YOLOv11	78.5	70.9	75.5	41.7
	ALC-Net	**83.6**	**72.2**	**79.1**	**47.4**
United States	YOLOv11	**65.0**	35.9	34.7	16.8
	ALC-Net	62.3	**37.2**	**35.6**	**17.3**

As presented in the table, the proposed ALC-Net exhibits robust performance on these diverse RDD2022 subsets. Notably, all evaluation metrics of ALC-Net show varying degrees of improvement compared to the baseline. This consistent performance across different data sources directly underscores the excellent generalization ability and adaptability of ALC-Net to diverse damage situations and imaging environments. This robustness reflects its practicability for deployment in varied real-world road inspection scenarios using UAVs.

Nevertheless, it is important to note that performance is relatively unsatisfactory on the China drone and United

States subset. This is likely due to the presence of complex background conditions, including significant illumination variations, shadows, and occlusion, which pose inherent challenges for detection models in uncontrolled environments. The results highlight that while ALC-Net generalizes well, challenging environmental factors remain an area for further investigation in future work to improve robustness in highly complex scenarios.

4.5. Cost Analysis

Beyond detection accuracy, the lightweight nature of the model is crucial for practical deployment on UAVs, which have limited computational resources. Therefore, a cost analysis is conducted to compare the efficiency of different models in terms of $Params$, FPS, and $GFLOPs$ on the UAV-PDD2023 dataset. The results are shown in Table 4.

Table 4 Cost comparison with SOTA methods on UAV-PDD2023 dataset

Model	$Params(M)$	FPS	$GFLOPs$
RetinaNet	38.0	41.9	183.0
Cascade R-CNN	69.4	12.4	190.0
Faster R-CNN	41.4	25.4	178.0
YOLOv5	4.2	217.4	7.1
YOLOv6	2.5	163.9	11.8
YOLOv8	3.0	188.7	8.1
YOLOv10	2.7	222.2	8.2
YOLOv11	2.6	**232.6**	6.3
ALC-Net	**2.3**	153.5	**5.8**

From the results, ALC-Net achieves the optimal (lowest) values for both $Params$ and $GFLOPs$, while ranking at an average level for FPS among the compared models. The minimal $Params$ count demonstrates a key advantage for effortless deployment on UAV platforms with limited memory. Similarly, the low $GFLOPs$ value signifies reduced computational cost and energy consumption per inference, which is vital for real-time processing on battery-powered UAVs.

A notable observation from Table 4 is the reduction in FPS for ALC-Net compared to the baseline. This result is primarily attributed to the introduction of the coordinate attention mechanism. Although this module is instrumental in enhancing detection accuracy by capturing long-range dependencies and directional features critical for identifying cracks, it introduces additional computational steps, including two spatial average pooling operations and subsequent convolutions. This deliberate design prioritizes a significant gain in accuracy over maintaining the absolute highest inference speed, striking a balance that remains highly practical for real-world UAV deployment.

While the average FPS indicates room for improvement in inference speed, the superiority in $Params$ and $GFLOPs$ highlights the lightweight efficiency of ALC-Net. For UAV deployment, minimizing model size and computational load ($Params$ and $GFLOPs$) are often more critical than achieving the absolute highest FPS, as it directly impacts deployability and energy efficiency. In summary, ALC-Net stands at the forefront among the tested models in terms of lightweight design, achieving a strong balance between detection performance and resource efficiency, thus confirming its practicality for UAV-based road damage inspection.

4.6. Heatmap Visualization Analysis

To provide qualitative insight into the ability of ALC-Net to spatially focus on relevant regions for road damage detection, attention heatmaps generated by the model are visualized. These heatmaps overlaid on the original images are presented in Figure 8, illustrating the areas where the network exhibits high activation. And the labels are also visualized for comparison analysis.

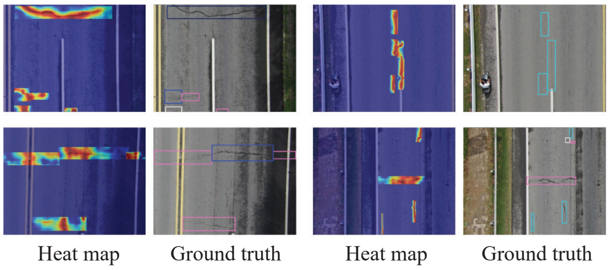

Figure 8. Heatmap visualization of the ALC-Net.

Upon examining the heatmaps, a strong and consistent correspondence between high activation areas and the visually apparent road damage features is clearly observed. High heatmap values, typically indicated by warmer colors, are precisely concentrated on the specific structures representing road damage. These activation patterns are shown to accurately trace the shapes and boundaries of the damage, indicating that the network attention is effectively directed towards the discriminative visual cues associated with road defects. Crucially, low activation is consistently observed in areas of the image that do not contain road damage, including intact road surfaces, surrounding vegetation, and irrelevant objects. This demonstrates that the spatial attention of network is effectively learned to be away from background clutter and non-damage features.

This precise spatial focusing ability, as visually confirmed by the heatmaps, is a key factor contributing to robust detection performance of ALC-Net. The capacity of the model to accurately localize the damage features by concentrating its attention on the most informative pixels directly facilitates accurate bounding box prediction and reduces potential false positives arising from misleading background elements. The results once again demonstrate that the architectural design of ALC-Net enables enhanced spatial awareness and feature learning capabilities, which are critical for accurate and reliable road damage detection in complex UAV imagery.

5. Conclusion

In this paper, an innovative ALC-Net has been proposed for road damage detection in UAV imagery. The model has integrated multiple strategies to enhance overall performance: the focus module and coordinate attention mechanism collectively improve feature representation by emphasizing image details and directional damage characteristics. By cascading ghost convolution with SE attention, the model has achieved an optimal balance between accuracy and computational efficiency, effectively mitigating performance degradation caused by lightweight design.

Extensive evaluations on the UAV-PDD 2023 dataset have demonstrated the superiority of ALC-Net, outperforming eight state-of-the-art models across all metrics. Ablation studies have validated the efficacy of individual modules, while tests on additional benchmark datasets (RDD2022 subsets) have confirmed robust generalization of the model to diverse scenarios and data sources. In future, works can focus on further improving detection accuracy through enhancing multi-scale feature fusion. Additionally, deploying ALC-Net on UAV hardware for real-world road inspection represents a promising direction for practical implementation.

Author Contributions: Liang Chen: Conceptualization, Methodology, Formal analysis, Investigation, Writing—original draft, Writing—review & editing, Software. **Peishu Wu:** Methodology, Formal analysis, Writing—original draft, Writing—review & editing. **Weilong Tan:** Formal analysis, Investigation, Writing—original draft. **Han Li:** Formal analysis, Writing—original draft, Writing—review & editing. **Haonan Chen:** Methodology, Investigation, Software. **Nianyin Zeng:** Conceptualization, Methodology, Writing—review & editing Funding acquisition.

Funding: This work was supported in part by the Independent Innovation Foundation of AECC under Grant ZZCX-2023-005, the Natural Science Foundation for Distinguished Young Scholars of the Fujian Province under Grant 2023J06010, and the National Key Research and Development Program of China under Grant 2024YFC3407000.

Data Availability Statement: This study used existing datasets that are openly available at https://zenodo.org/records/8429208 and https://github.com/sekilab/RoadDamageDetector.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Abu Dabous, S.; Ait Gacem, M.; Zeiada, W.; *et al.* Artificial intelligence applications in pavement infrastructure damage detection with automated three-dimensional imaging-A systematic review. Alexandria Eng. J., **2025**, *117*: 510−533. doi: 10.1016/j.aej.2024.11.081
2. Tafida, A.; Alaloul, W.S.; Zawawi, N.A.B.W.; *et al.* Advancing smart transportation: A review of computer vision and photogrammetry in learning-based dimensional road pavement defect detection. Comput. Sci. Rev., **2025**, *56*: 100729. doi: 10.1016/j.cosrev.2025.100729
3. Zhang, Y.C.; Liu, C. Real-time pavement damage detection with damage shape adaptation. IEEE Trans. Intell. Transport. Syst., **2024**, *25*: 18954−18963. doi: 10.1109/TITS.2024.3416508
4. Lu, W.J.; Lan, C.Z.; Niu, C.Y.; *et al.* A CNN-transformer hybrid model based on CSWin transformer for UAV image object detection. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., **2023**, *16*: 1211−1231. doi: 10.1109/JSTARS.2023.3234161
5. Yuan, Y.Z.; Gao, S.C.; Zhang, Z.T.; *et al.* Edge-cloud collaborative UAV object detection: Edge-embedded lightweight algorithm design and task offloading using fuzzy neural network. IEEE Trans. Cloud Comput., **2024**, *12*: 306−318. doi: 10.1109/TCC.2024.3361858

6. Zhu, J.Q.; Wu, Y.X.; Ma, T. Multi-object detection for daily road maintenance inspection with UAV based on improved YOLOv8. IEEE Trans. Intell. Transport. Syst. **2024**, *25*, 16548–16560. doi: 10.1109/TITS.2024.3437770
7. Chen, T.Y.; Ren, J.T. Integrating GAN and texture synthesis for enhanced road damage detection. IEEE Trans. Intell. Transport. Syst., **2024**, *25*: 12361−12371. doi: 10.1109/TITS.2024.3373394
8. Zeng, N.Y.; Li, X.Y.; Wu, P.S.; *et al*. A novel tensor decomposition-based efficient detector for low-altitude aerial objects with knowledge distillation scheme. IEEE/CAA J. Autom. Sin., **2024**, *11*: 487−501. doi: 10.1109/JAS.2023.124029
9. Min, X.L.; Zhou, W.; Hu, R.; *et al*. LWUAVDet: A lightweight UAV object detection network on edge devices. IEEE Internet Things J., **2024**, *11*: 24013−24023. doi: 10.1109/JIOT.2024.3388045
10. Wu, P.S.; Wang, Z.D.; Li, H.; *et al*. KD-PAR: A knowledge distillation-based pedestrian attribute recognition model with multi-label mixed feature learning network. Expert Syst. Appl., **2024**, *237*: 121305. doi: 10.1016/j.eswa.2023.121305
11. Hou, Q.B.; Zhou, D.Q.; Feng, J.S. Coordinate attention for efficient mobile network design. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021*; IEEE: New York, 2021; pp. 13708–13717. doi:10.1109/CVPR46437.2021.01350
12. Han, K.; Wang, Y.H.; Tian, Q.; *et al*. GhostNet: More features from cheap operations. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020*; IEEE: New York, 2020; pp. 1577–1586. doi:10.1109/CVPR42600.2020.00165
13. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018*; IEEE: New York, 2018; pp. 7132–7141. doi:10.1109/CVPR.2018.00745
14. Ren, S.Q.; He, K.M.; Girshick, R.; *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell., **2017**, *39*: 1137−1149. doi: 10.1109/TPAMI.2016.2577031
15. He, K.M.; Gkioxari, G.; Dollár, P.; *et al*. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017*; IEEE: New York, 2017; pp. 2980–2988. doi:10.1109/ICCV.2017.322
16. Cai, Z.W.; Vasconcelos, N. Cascade R-CNN: Delving into high quality object detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018*; IEEE: New York, 2018; pp. 6154–6162. doi:10.1109/CVPR.2018.00644
17. Li, C.Y.; Li, L.L.; Jiang, H.L.; *et al*. YOLOv6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv: 2209.02976, 2022. doi:10.48550/arXiv.2209.02976
18. Wang, A.; Chen, H.; Liu, L.; *et al*. YOLOv10: Real-time end-to-end object detection. In *Proceedings of the 38th International Conference on Neural Information Processing Systems, Vancouver BC Canada, 10–15 December 2024*; Curran Associates Inc.: Red Hook, 2024; p. 3429.
19. Lin, T.Y.; Goyal, P.; Girshick, R.; *et al*. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017*; IEEE: New York, 2017; pp. 2999–3007. doi:10.1109/ICCV.2017.324
20. Liu, W.; Anguelov, D.; Erhan, D.; *et al*. SSD: Single shot MultiBox detector. In *Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37. doi:10.1007/978-3-319-46448-0_2
21. Chen, H.N.; Wu, P.S.; Wen, W.M.; *et al*. DLA-Net: A dynamically learnable attention network for intelligent surface visual inspection of aero-engine blades. IEEE Trans. Instrum. Meas., **2025**, *74*: 3532114. doi: 10.1109/TIM.2025.3561440
22. Yue, X.L.; Chen, J.D.; Zhong, G.Q. Metal surface defect detection based on metal-YOLOX. Int. J. Network Dyn. Intell., **2023**, *2*: 100020. doi: 10.53941/ijndi.2023.100020
23. El-Shorbagy, M.A.; Bouaouda, A.; Nabwey, H.A.; *et al*. Bald eagle search algorithm: A comprehensive review with its variants and applications. Syst. Sci. Control Eng., **2024**, *12*: 2385310. doi: 10.1080/21642583.2024.2385310
24. Fang, W.H.; Shen, B.; Pan, A.Q.; *et al*. A cooperative stochastic configuration network based on differential evolutionary sparrow search algorithm for prediction. Syst. Sci. Control Eng., **2024**, *12*: 2314481. doi: 10.1080/21642583.2024.2314481
25. Li, H.; Liu, H.N.; Lan, C.B.; *et al*. SMWO/D: A decomposition-based switching multi-objective whale optimiser for structural optimisation of Turbine disk in aero-engines. Int. J. Syst. Sci., **2023**, *54*: 1713−1728. doi: 10.1080/00207721.2023.2209873
26. Li, H.; Wang, Z.D.; Zeng, N.Y.; *et al*. Promoting objective knowledge transfer: A cascaded fuzzy system for solving dynamic multiobjective optimization problems. IEEE Trans. Fuzzy Syst., **2024**, *32*: 6199−6213. doi: 10.1109/TFUZZ.2024.3443207
27. Sheng, M.M.; Ding, W.J.; Sheng, W.G. Differential evolution with adaptive niching and reinitialisation for nonlinear equation systems. Int. J. Syst. Sci., **2024**, *55*: 2172−2186. doi: 10.1080/00207721.2024.2337039
28. Xue, J.K.; Shen, B. A survey on sparrow search algorithms and their applications. Int. J. Syst. Sci., **2024**, *55*: 814−832. doi: 10.1080/00207721.2023.2293687
29. Zhang, T.H.; Liu, Q.X.; Liu, J.Y.; *et al*. Multiple-bipartite consensus for networked Lagrangian systems without using neighbours' velocity information in the directed graph. Syst. Sci. Control Eng., **2023**, *11*: 2210185. doi: 10.1080/21642583.2023.2210185
30. Li, H.; Wu, P.S.; Wang, Z.D.; *et al*. A generalized framework of feature learning enhanced convolutional neural network for pathology-image-oriented cancer diagnosis. Comput. Biol. Med., **2022**, *151*: 106265. doi: 10.1016/j.compbiomed.2022.106265
31. Liang, Y.P.; Tian, L.L.; Zhang, X.; *et al*. Multi-dimensional adaptive learning rate gradient descent optimization algorithm for network training in magneto-optical defect detection. Int. J. Network Dyn. Intell., **2024**, *3*: 100016. doi: 10.53941/IJNDI.2024.100016
32. Yuan, Z.F.; Li, Y.; Liu, Y.; *et al*. Unsupervised ship detection in SAR imagery based on energy density-induced clustering. Int. J. Network Dyn. Intell., **2023**, *2*: 100006. doi: 10.53941/IJNDI.2023.100006
33. Xu, J.H.; Fan, X.T.; Jian, H.D.; *et al*. YoloOW: A spatial scale adaptive real-time object detection neural network for open water search and rescue from UAV aerial imagery. IEEE Trans. Geosci. Remote Sens., **2024**, *62*: 5623115. doi: 10.1109/TGRS.2024.3395483
34. Ye, T.; Qin, W.Y.; Zhao, Z.Y.; *et al*. Real-time object detection network in UAV-vision based on CNN and transformer. IEEE Trans. Instrum. Meas., **2023**, *72*: 2505713. doi: 10.1109/TIM.2023.3241825
35. Zhang, Y.Z.; Wu, C.Y.; Zhang, T.; *et al*. Self-attention guidance and multiscale feature fusion-based UAV image object detection. IEEE Geosci. Remote Sens. Lett., **2023**, *20*: 6004305. doi: 10.1109/LGRS.2023.3265995
36. Jiang, L.J.; Yuan, B.X.; Du, J.W.; *et al*. MFFSODNet: Multiscale feature fusion small object detection network for UAV aerial images. IEEE Trans. Instrum. Meas., **2024**, *73*: 5015214. doi: 10.1109/TIM.2024.3381272
37. Lan, Z.Y.; Zhuang, F.Y.; Lin, Z.J.; *et al*. MFO-Net: A multiscale feature optimization network for UAV image object detection.

IEEE Geosci. Remote Sens. Lett., **2024**, *21*: 6006605. doi: 10.1109/LGRS.2024.3382090

38. Dai, J.F.; Qi, H.Z.; Xiong, Y.W.; *et al*. Deformable convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017*; IEEE: New York, 2017; pp. 764–773. doi:10.1109/ICCV.2017.89
39. Chen, J.R.; Kao, S.H.; He, H.; *et al*. Run, don't walk: Chasing higher FLOPS for faster neural networks. In *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023*; IEEE: New York, 2023; pp. 12021–12031. doi:10.1109/CVPR52729.2023.01157
40. Wang, S.X.; Jiao, H.Z.; Su, X.; *et al*. An ensemble learning approach with attention mechanism for detecting pavement distress and disaster-induced road damage. IEEE Trans. Intell. Transport. Syst., **2024**, *25*: 13667−13681. doi: 10.1109/TITS.2024.3391751
41. Yin, T.X.; Zhang, W.; Kou, J.Q.; *et al*. Promoting automatic detection of road damage: A high-resolution dataset, a new approach, and a new evaluation criterion. IEEE Trans. Autom. Sci. Eng., **2025**, *22*: 2472−2484. doi: 10.1109/TASE.2024.3379945
42. Li, J.; Qu, Z.; Wang, S.Y.; *et al*. YOLOX-RDD: A method of anchor-free road damage detection for front-view images. IEEE Trans. Intell. Transport. Syst., **2024**, *25*: 14725−14739. doi: 10.1109/TITS.2024.3389945
43. Qiao, S.Y.; Chen, L.C.; Yuille, A. DetectoRS: Detecting objects with recursive feature pyramid and switchable atrous convolution. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021*; IEEE: New York, 2021; pp. 10208–10219. doi:10.1109/CVPR46437.2021.01008
44. Khan, M.W.; Obaidat, M.S.; Mahmood, K.; *et al.* Real-time road damage detection and infrastructure evaluation leveraging unmanned aerial vehicles and tiny machine learning. IEEE Internet Things J., **2024**, *11*: 21347−21358. doi: 10.1109/JIOT.2024.3385994
45. Pham, S.V.H.; Van Tien Nguyen, K.; Le, L.H.; *et al*. Developing RTI IMS software to autonomously manage road surface quality, adapting to environmental impacts. IEEE Trans. Intell. Transport. Syst., **2024**, *25*: 18472−18484. doi: 10.1109/TITS.2024.3442949
46. Yan, H.H.; Zhang, J.F. UAV-PDD2023: A benchmark dataset for pavement distress detection based on UAV images. Data Brief, **2023**, *51*: 109692. doi: 10.1016/j.dib.2023.109692
47. Arya, D.; Maeda, H.; Ghosh, S.K.; *et al*. RDD2022: A multi-national image dataset for automatic road damage detection. Geosci. Data J., **2024**, *11*: 846−862. doi: 10.1002/gdj3.260

Citation: Chen, L.; Wu, P.; Tan, W.; *et al*. A Novel UAV-based Road Damage Detection Algorithm with Lightweight Convolution and Attention Mechanism. *International Journal of Network Dynamics and Intelligence*. 2025, 4(4), 100025. doi: 10.53941/ijndi.2025.100025