RESEARCH   **Open Access**

YOLO-LRDD: a lightweight method for road damage detection based on improved YOLOv5s

Fang Wan[1], Chen Sun[1], Hongyang He[2*] , Guangbo Lei[1], Li Xu[1] and Teng Xiao[1*]

*Correspondence:
HXH228@student.bham.ac.uk;
xiao@hbut.edu.cn

[1] School of Computer Science,
Hubei University of Technology,
Wuhan 430068, China
[2] College of Engineering
and Physical Sciences, The
University of Birmingham,
Edgbaston, Birmingham B15
2TT, UK

Abstract

In computer vision, timely and accurate execution of object identification tasks is critical. However, present road damage detection approaches based on deep learning suffer from complex models and computationally time-consuming issues. To address these issues, we present a lightweight model for road damage identification by enhancing the YOLOv5s approach. The resulting algorithm, YOLO-LRDD, provides a good balance of detection precision and speed. First, we propose the novel backbone network Shuffle-ECANet by adding an ECA attention module into the lightweight model ShuffleNetV2. Second, to ensure reliable detection, we employ BiFPN rather than the original feature pyramid network since it improves the network's capacity to describe features. Moreover, in the model training phase, localization loss is modified to Focal-EIOU in order to get higher-quality anchor box. Lastly, we augment the well-known RDD2020 dataset with many samples of Chinese road scenes and compare YOLO-LRDD against several state-of-the-art object detection techniques. The smaller model of our YOLO-LRDD offers superior performance in terms of accuracy and efficiency, as determined by our experiments. Compared to YOLOv5s in particular, YOLO-LRDD improves single image recognition speed by 22.3% and reduces model size by 28.8% while maintaining comparable accuracy. In addition, it is easier to implant in mobile devices because its model is smaller and lighter than those of the other approaches.

Keywords: Road damage detection, Lightweight, Object detection, YOLOv5s, Deep learning

1 Introduction

Nowadays, more than 80 countries and territories around the world have highways with a total distance of more than 230,000 km in operation. The USA, with a total length of 88,000 km of highways, has completed an interstate highway grid with a core of interstate highways. While China also has a huge network of highways and roads where cracks in the road surface form and infiltration of rainwater accelerates the expansion of defects creating traps for moving vehicles. If not timely detection and access to road damage information and repair damaged roads, poor road conditions can lead to excessive wear and tear on vehicles and can increase the likelihood of traffic accidents, leading to additional financial losses. According to data, poor road conditions are responsible

for 16% of traffic accidents [1]. To protect the lives of pedestrians and reduce property damage, it is urgent to address the issue of road damage detection.

Current methods of detecting road damage are classified into three categories, manual inspection, automated inspection, and image processing techniques. Pavement inspection in developing countries usually relies on manual inspection, but traditional manual inspection suffers from poor safety, low efficiency, high costs, and relies on the experience of the inspector, which can lead to inconsistent judgment. With the development of technology, the use of automated road inspection is gradually increasing, such as road inspection vehicles equipped with infrared or sensor equipment [2, 3], but the complexity of the road environment makes it difficult for automated inspection equipment to meet the needs of practical engineering in terms of recognition accuracy and speed, and such equipment requires high hardware costs so that the corresponding inspection costs are also higher. Image processing technology has the advantages of high efficiency and low cost, and recognition accuracy is also gradually increasing with the development of technology. As a result, many researchers have used image processing techniques to detect pavement damage [4–6]. Traditional image processing techniques usually use manually selected features, such as color, texture, and geometric features, to segment pavement defects and then use machine learning algorithms for classification and matching to achieve the detection of pavement damage. For example, Fernaldez et al. [7] first preprocessed the crack image of road to emphasize the main characteristics of the crack, and then applied a decision tree heuristic algorithm to perform the final classification of the image. However, due to the complexity of the road environment, traditional image processing methods cannot achieve the requirements for model generalization capability and robustness in practical engineering through manually designed feature extraction. Compared with traditional image processing techniques, image processing techniques based on deep learning theory have been widely used in pavement defect detection with higher accuracy, faster speed, and embeddability [8].

The object detection system has been variedly used in military and health sectors for efficient assistance in various fields [9]. Deep learning-based models are increasingly being widely used under their powerful feature extraction capabilities, such as convolutional neural networks [10] being widely used in tasks such as image classification [11], object detection [12], and semantic segmentation [13]. The current object detection networks for road damage are generally divided into two categories, one with a two-stage model based on candidate regions, for example. Xu et al. [14] propose a novel tunnel defect inspection method based on the Mask R-CNN. To improve the accuracy of the network, they endow it with a path augmentation feature pyramid network (PAFPN) and an edge detection branch. Wang [15] detected and classified damaged roads on the faster R-CNN-based network model and to address the problem of an unbalanced distribution of data across different defect classes, proposed to introduce data augmentation techniques before training to obtain an average F1-Score score of 62.5%. Another type of regression-based is single-stage network. Jeong et al. [16] applied test-time data augmentation (TTA) on a YOLOv5x-based model, which generates a large number of new images for data augmentation by horizontally flipping each training image, increasing the image resolution, etc., and adding the existing images together with the augmented images to the trained u-YOLO. The model scored 67% in F1 and won the first place in

Wan *et al. EURASIP Journal on Advances in Signal Processing*   (2022) 2022:98

Page 3 of 18

the Global Road Damage Detection Challenge (GRDDC) competition, but the detection speed was not satisfactory and not real time. Wang et al. [17] targeted the characteristics of road damage with elongated and microminiature and used the model based on The YOLOv3 model combining low-level features with high-level features and improving the loss function to improve the detection accuracy. However, this model is only highly accurate in detecting transverse or longitudinal cracks, but in reality, road damage types are often very diverse and this proposed method is not universal for road damage detection.

Recently, many researchers are dedicated to proposing lightweight road damage detectors. Shim et al. [18] designed a lightweight semantic segmentation network. They optimized the parameters of the model but did not consider whether the detection speed of the model has an impact. Sheta et al. [19] developed a lightweight convolutional neural networks model to detect pavement cracks, which architecture performs well in detecting cracks. However, the usage scene is too simple to adapt to the multiple damage types on the road. Guo et al. [20] improved YOLOv5s model used to detect various road surface diseases, which can improve the accuracy of object detection. However, there are models with a low degree of lightweight, which is more difficult to meet the requirements of embedded devices than other lightweight models. The above methods have made a reasonable contribution to lightweight models in the road damage detection field. Unfortunately, the lightweight models designed by these studies do not have a good balance between detection precision and detection speed.

Previous datasets on road defects suffer from unclear labeling, sparse defect categories, or unbalanced sample sizes of defect types. Studies have shown [21] that the quality of the dataset and the number of sample distributions play a crucial role in the performance of the network model. Although the dataset used in this paper can solve some of these problems, the problem of the unbalanced number of samples for different defects still exists. To address this problem, Shim et al. [22] proposed technology that includes a super-resolution and semi-supervised learning method based on a generative adversarial network, which can improve road image quality and enhance detection performance. Maeda et al. [23] used a data augmentation method combining PG-GAN with the Poisson hybrid method to increase the pothole data, and this method improved the F1-Score by 5% on pavement pothole detection. The above findings suggest that data augmentation techniques can effectively improve the network's ability to extract features from samples.

Nowadays, the existing deep learning model cannot meet the requirements of detection and real-time road damage detection. So, it is urgent to find a suitable model that can improve the detection precision and reduce the complexity of the model. YOLOv5 is an advanced single-stage object detection model, which can realize real-time detection with high prediction precision. YOLOv5 is an advanced single-stage object detection model, which can realize real-time detection with high prediction accuracy. Its mAP can reach up to 72% in COCO 2017 val set. The YOLO model gradually increases according to the network depth and the dimension of the feature map and is divided into YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. However, although the YOLOv5s method has high precision for road damage detection tasks, it still has the disadvantage of a large number of calculations. Therefore, we proposed the YOLO-LRDD model

Wan *et al. EURASIP Journal on Advances in Signal Processing*   (2022) 2022:98

Page 4 of 18

to improve the prediction precision and achieve real-time detection while reducing the complexity of the model based on the improved YOLOv5s model.

In our approach, we propose a lightweight model that is more suitable for road damage detection. In order to achieve the goal of achieving a good balance between detection precision and speed while keeping the model lightweight. Overall, this study makes four contributions: (1) We propose YOLO-LRDD, a novel lightweight model for road damage detection. (2) We design a new backbone network, which can effectively alleviate the situation that the model is lightweight and the detection precision reduced. (3) In order to enhance the feature description ability of the model and obtain high-quality anchor boxes, we replace the original YOLOv5s neck network and position loss function. (4) The original road damage dataset has been expanded with local road damage samples in China. The rest of this article is organized as follows. Section 2 describes the proposed YOLO-LRDD detection scheme in detail. Section 3 reports and discusses the experimental results. Section 4 concludes this paper. Section 5 describes the importance of our work to the world scientific community and directions for future work.

2 Method

2.1 The overview of YOLO-LRDD

In view of the insufficient number of road samples about China in the road damage detection dataset, we have expanded the dataset so that the model can be applied in road detection in China. The currently used YOLOv5s network applied to road damage detection has the disadvantages of large model parameters and slow detection speed. To solve these problems, we propose a lightweight backbone network called Shuffle-ECANet, which combines ShuffleNetV2 [24] with the ECA-Net [25] attention mechanism, it can make the model lightweight, and the detection speed is much faster than the original YOLOv5s. However, after we replaced the backbone network of YOLOv5s, we found that the detection accuracy of the model decreased, especially when faced with areas with insignificant damage features. Our analysis is that the feature pyramid network of YOLOv5s uses PANet [26], and it fails to fuse features well. In this regard, we use BiFPN [27] to replace PANet and find that it can achieve more efficient multi-scale feature fusion to improve detection accuracy. In the model training stage, since the original YOLOv5s uses the CIOU [28] loss function, it cannot solve the problem of sample imbalance. Therefore, we replace the CIOU with the Focal-EIOU [29] loss function. After experiments, it is found that it can solve the CIOU problem and also improve the quality of anchor boxes.

In Sect. 2, we organize the article as follows. In Sect. 2.2, we illustrate the problems with the current dataset and expand it, and show the types of road damage detected in this paper. After first describing the overall network structure of YOLO-LRDD in Sect. 2.3, we detail the backbone network and neck network of YOLO-LRDD in 2.3.1 and 2.3.2, and finally explain the Focal-EIOU loss function in detail in 2.3.3.

2.2 Data collection and processing

In this paper, we have constructed a road damage dataset called RDDC, which is based on the RDD2020 [30] dataset. RDDC is an extension of the RDD2020 dataset about Chinese road samples. As far as we know, there are various problems with the road damage

(a) Longitudinal Cracks   (b) Lateral Cracks   (c) Alligators   (d) Potholes
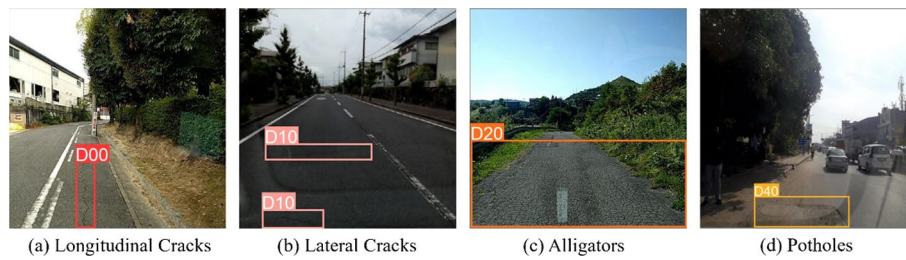
Fig. 1 Examples of damages types
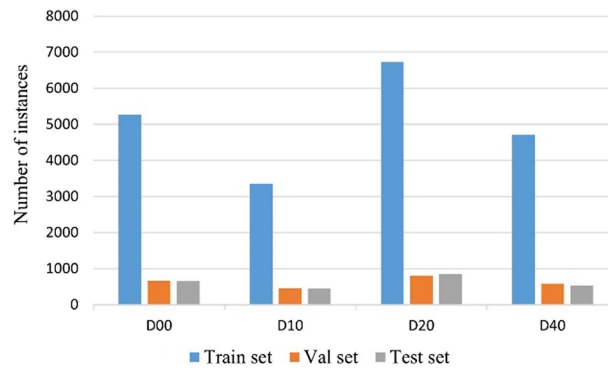
Fig. 2 Number of damage instances

images that exist today, such as inconsistent resolution, the equipment used to capture the image data, and a range of external factors such as lighting shadows; these seriously affect the quality of the dataset we use to train our models, so we decided to build our road damage dataset based on the RDD2020, and the quality of the RDDC dataset was improved by using field collection. Four common types of road damage are mainly studied in our dataset. Thus, for this study, the damage categories considered are D00 to represent longitudinal cracks, D10 for lateral cracks, D20 for alligators cracks, and D40 for potholes. Figure 1 shows four typical damage types in RDDC.

To make full use of the dataset and to improve the generalization ability of the model, the images containing multiple types of defects were preferentially used. After careful selection, the collected images were standardized before model training, and the images were reduced to 640×640 size so that the YOLO model could exert the best training performance. After standardization, according to different types of road damage, labeling was used to manually label data. The file format of the labeling was txt, and a total of 13,780 pieces of data were formed. According to the needs of the experiment, it was randomly divided into a training set, a validation set, and a test set, according to 8:1:1. The details of the damage distribution of the dataset are shown in Fig. 2.

2.3 The proposed YOLO-LRDD

The overall schematic of YOLO-LRDD is shown in Fig. 3. The network architecture of YOLO-LRDD is mainly composed of four parts: input module, backbone network, neck network, and prediction network. In the first part, the input module enhances features through mosaic data augmentation, adaptive anchor box computation, and adaptive

Wan *et al. EURASIP Journal on Advances in Signal Processing*   (2022) 2022:98

Page 6 of 18
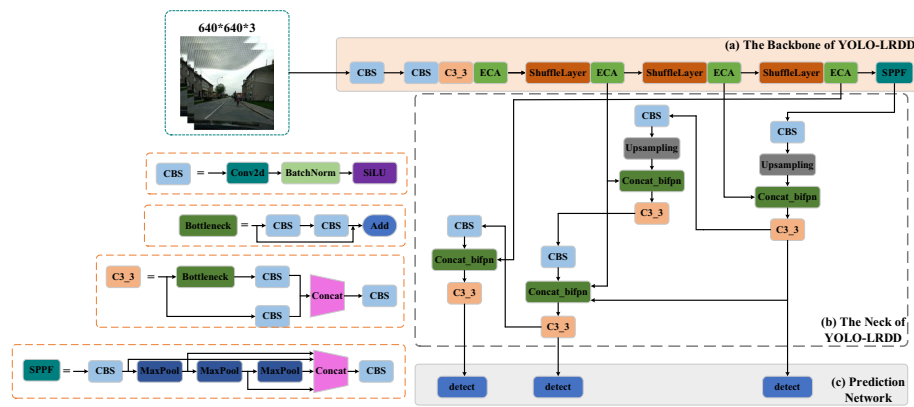
Fig. 3 Network architecture of the YOLO-LRDD method

image scaling. In the second part, we replace the backbone network of the original YOLOv5s with the proposed backbone network Shuffle-ECANet. In the third part, we use bidirectional feature pyramid network (BiFPN) instead of path aggregation network for instance segmentation (PANet). The last part is the prediction network, which performs object detection and classification tasks, which ultimately reveal the road damage type and predicted probability. In the model training phase, the focal and efficient IOU (Focal-EIOU) loss is used instead of the complete IOU loss to solve the problem of sample imbalance and improve the quality of the bounding box.

2.3.1 The backbone of YOLO-LRDD

The original YOLOv5s backbone feature extraction network adopts the C3 structure, which will bring a large number of parameters and cause the problem of slow detection speed. In addition, when the model faces the complex application scenario of road damage detection, there are often problems of insufficient memory and high detection delay in embedded devices. Therefore, it is crucial to study lightweight feature extraction networks. According to the ability of attention mechanism to obtain global information, we created a new feature extraction network ShuffleECA-Net, which combines Shuffle-NetV2 and ECA-Net attention mechanism, which not only makes the model lightweight, but also improves the detection speed.

ShuffleNet [31] is a convolutional neural network with high computational efficiency deployed on mobile devices. It uses pointwise group convolution and channel shuffle to provide better performance and faster-running speed for mobile devices. ShufflenetV2 is an improved version of ShuffleNet, which structure is shown in Fig. 4. It introduces the channel splitting operation, changes the element addition to concatenation, and then uses the channel shuffling operation to mix features. In this article, ShufflenetV2 is used as the basic backbone network.

ECA-Net is a network of attention mechanisms based on Squeeze and Excitation Networks (SENet) [32]. ECA-Net uses the one-dimensional sparse convolution filter which is used to generate channel weights to replace the full connection layer in the SENet. This solves the problem caused by the operation of dimension reduction, which significantly reduces the complexity of the network and can maintain the same performance as the original. ECA-Net is a novel, lightweight, and efficient attention mechanism module.
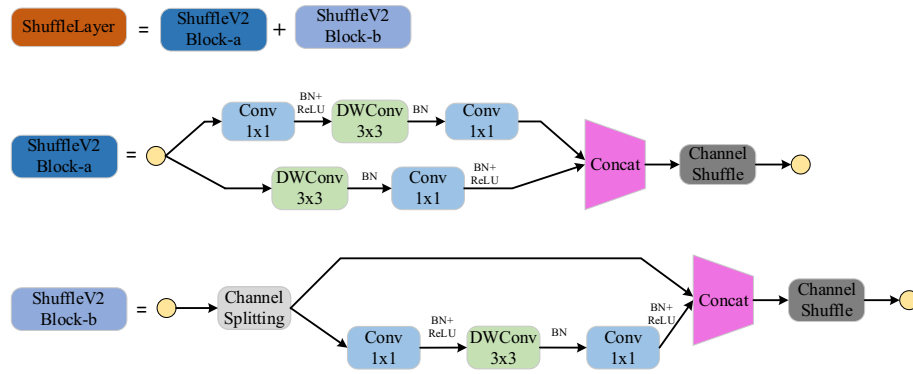
Fig. 4 Structure of the ShuffleLayer
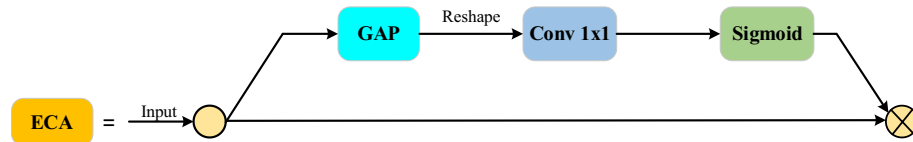
Fig. 5 Structure of the ECA module

The research shows that it can improve the prediction accuracy without increasing the computational complexity, and can be easily deployed to the mobile network. The structure of the ECA-Net module is shown in Fig. 5.

ECA-Net converts the input feature map $X \in R^{(W \times H \times C)}$ into a single real value through global average pooling and the obtained features are expressed as $X_{avg} \in R^{(1 \times 1 \times C)}$, where W, H, and C are, respectively, expressed as the width, height, and channel of the feature, as shown in formula (1). The one-dimensional convolution kernel with the size of K is used to extract the feature from X_{avg} by convolutional operation. K is shown as formula (2).

$$X_{avg} = \frac{1}{W \times H} \sum_{i=1, j=1}^{W, H} X_{ij} \tag{1}$$

$$K = \frac{\log_2 C + 1}{2} \tag{2}$$

Then, the sigmoid activation function is used to activate the output result after convolution to obtain the weight parameter $W \in R^{(1 \times 1 \times C)}$, which reflects the correlation and importance of each channel. Finally, the weight parameter W is multiplied by the original input feature map to complete the recording of each channel feature of the feature map. In this way, important features are enhanced by giving larger weights, while invalid features are suppressed by giving smaller weights.

The backbone network of YOLOv5s is replaced by the combination of ShufflenetV2 and ECA-Net. Meanwhile, the CBS module and SPPF module in the original YOLOv5s model are retained to reduce the pixel loss from the feature map in the initial stage, which can ensure the learning ability of the model and enhance the feature expression

ability of the feature map. We proposed that the backbone network of Shuffle-ECANet can not only ensure the road damage detection effect which is unchanged, but also greatly reduce the amount of model calculation and achieve real-time detection. However, we found that the detection accuracy of Shuffle-ECANet is not high for areas with small damage areas or unclear damage features, because such features carry little information and are easy to cause information loss in the process of forwarding calculation. To solve this problem, we replace the feature pyramid in YOLOv5s with BiFPN to enhance the ability to describe features.

2.3.2 The feature pyramid network of YOLO-LRDD

The purpose of the feature pyramid used by YOLOv5s is to extract features from different scales and further generate a feature pyramid network to detect targets from different scales by using the feature maps of different scales. YOLOv5 uses the architecture of PANet which adds a bottom-up channel based on the top-down (feature pyramid networks) FPN [33] structure, as the neck network. This makes the prediction layer have both high-level semantic information and bottom-level location information. However, in road damage detection, the characteristics of cracks are often long and discontinuous, and the damage is slender and tiny, which requires the network to have a strong ability of feature extraction. BiFPN was proposed by EfficientDet and is based on the structure of PANet. Through bidirectional connection and weighted feature fusion, it can enhance the feature extraction ability of the network, and introduce learnable weights to learn the importance of different input features. The network structure of PANet and BiFPN is shown in Fig. 6.

Bidirectional connection is consisted of three-part. To begin with, deleting the nodes with only one input, because the featureless fusion of such nodes has little contribution to the feature network, and will not have a great impact and simplify the network after deletion. Furthermore, an additional edge is added between the original input and output nodes to fuse more features. Finally, each top-down and bottom-up path is regarded as a repeated stack of feature network layers to achieve higher-level feature fusion.

2.3.3 The position loss of YOLO-LRDD

The loss function of YOLOv5 includes position loss, classification loss, and confidence loss. We keep the original binary cross-entropy loss [34] in YOLOv5 is used for
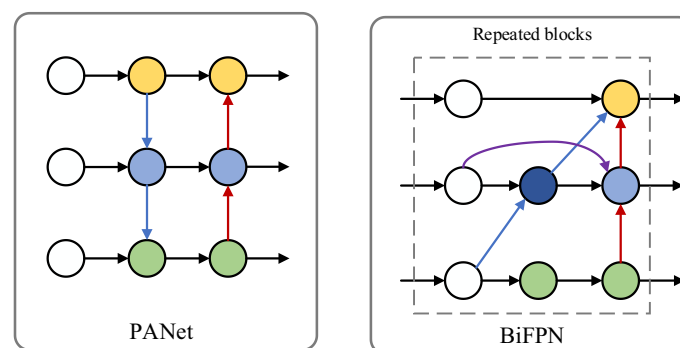
Fig. 6 Structure of PANet and BiFPN network

confidence loss and classification loss, while Focal-EIOU loss is used to replace the original CIOU loss for positioning loss. The most common position loss function is the IOU loss function, which calculates the intersection union ratio of the prediction boundary box and the ground truth box as shown in formula (3).

$$L_{IOU} = 1 - \left| \frac{B \cap B^{gt}}{B \cup B^{gt}} \right| \tag{3}$$

where the $\left| B \cap B^{gt} \right|$ is the intersection of the prediction boundary box and the ground truth box, and $\left| B \cup B^{gt} \right|$ is the union of the prediction boundary and the ground truth box. However, there are two problems with IOU loss. First, when the prediction boundary box and the ground truth box do not intersect, the IOU loss value is equal to 0, which makes the error unable to backpropagate. Furthermore, it cannot accurately reflect the area of overlap between the prediction boundary box and the ground truth box.

The current YOLOv5 model mainly uses the CIOU loss, which takes into account the distance between the center points of the predicted box and the center points of the ground truth bounding box and the aspect ratio of the predicted box and the ground truth box as shown in formula (4).

$$L_{CIOU} = 1 - IOU + \frac{\rho^2 \left(b, b^{gt} \right)}{c^2} + \alpha v \tag{4}$$

where $\rho^2 \left(b, b^{gt} \right)$ is the distance between the center points of the predicted box and the center points of the ground truth bounding box. c is the diagonal length of the smallest enclosing box covering the predicted box and the ground truth bounding box. αv takes into account the aspect ratio between the predicted box and the ground truth bounding box.

However, as the road damage types are diverse and the damaged area is not fixed, it is impossible to accurately predict the ground truth bounding box by CIOU loss. EIOU loss uses the calculation method of overlap loss and center distance loss in CIOU for reference, but the width and height loss use the minimum value of the difference between the width and height of the predicted box and the ground truth bounding box, which makes the model converge faster and obtains greater accuracy. The EIOU loss function is defined in formula (5).

$$L_{EIOU} = L_{IOU} + L_{dis} + L_{asp} = 1 - IOU + \frac{\rho^2 \left(b, b^{gt} \right)}{c^2} + \frac{\rho^2 \left(W, W^{gt} \right)}{C_w^2} + \frac{\rho^2 \left(h, h^{gt} \right)}{C_h^2} \tag{5}$$

C_w and C_h are the width and height of the smallest enclosing box covering the predicted box and the ground truth bounding box. However, there is a problem of unbalanced data samples in the road damage dataset, which will make the number of high-quality anchor boxes with small regression errors in the image far less than that of low-quality samples with large errors. The samples with poor quality will produce large gradients and affect the training process. Therefore, Focal-EIOU loss is used to improve loss of accuracy, which is shown in formula (6), where γ is the parameter that the degree of inhibition of outliers.

$$L_{Focal-EIOU} = IOU\gamma L_{EIOU} \tag{6}$$

3 Results and discussion

3.1 Experiment environment and metrics

The experiment environment is based on the Pytorch 11.0 framework, CUDA 11.3, and CUDNN 8.2, and the training model is based on an NVIDIA GeForce RTX 3060 (12 GB). An SGD optimizer was used in the training phase with an initial learning rate of 1E-5 and a weight decay of 5E-3, in addition to three warm-up periods of 0.8 momenta and a cosine annealing method to decay the learning rate, with 150 epochs per experiment and a batch size of 32, and the training process of the whole model took about 7 h. To better train the model, the mosaic method was used in the training phase to crop four images from the original dataset into one image after random scaling and stitching, and then an adaptive image scaling operation was performed to obtain a uniform 640×640 size image for training.

We use the RDDC dataset mentioned in Sect. 2.1 to validate the performance of the YOLO-LRDD method. Two commonly used metrics, Precision and Recall, are used to measure the performance of the model with an IOU threshold of 0.5 and a confidence threshold of 0.4 in order to objectively assess the experimental results. Precision is the probability of correctly predicting a positive sample out of all predicted positive samples, and recall is the probability of predicting a positive sample out of the actual positive samples. The formulas for precision and recall are shown in formulas (7), (8), TN (predict negative samples as negative samples), FN (predict positive samples as negative samples), TP (predict positive samples as positive samples), and FP (predict negative samples as positive samples).

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

In object detection, Precision and Recall interact with each other and cannot be used to evaluate the detection directly. Therefore, we introduce the AP to represent the detection precision and the comprehensive evaluation metric F1-Score to evaluate the model more comprehensively. Higher AP and F1-Score values imply higher network accuracy, and the mAP represents the average accuracy for n types of defects. The equations for AP, mAP, and F1-Score are shown in formulas (9), (10), and (11).

$$AP = \int_0^1 P(R)dR \tag{9}$$

$$mAP = \frac{1}{n} \sum_{i=1}^{m} AP^i \tag{10}$$

$$F1 \text{ - } Score = 2 \times \frac{P \times R}{P + R} \tag{11}$$

When evaluating the superiority of the algorithm, we need to define the detection speed, where we use the frame rate (FPS) to indicate the detection speed, which is an important indicator; if the FPS $\geq$ 30, it satisfies the requirements, and a video detection function with FPS $\geq$ 60 is superior.

### 3.2 Ablation experiments on YOLO-LRDD

To demonstrate the validity and necessity of each improved module in the YOLO-LRDD model, we use YOLOv5s as the baseline and gradually add improved modules for ablation experiments. Using, Precision, model size, F1-Score, and inference time per image as evaluation metrics, the experimental results are shown in Table 1.

We divided the ablation experiments into four steps to prove the superiority of the YOLO-LRDD model. (1) We first modified the original backbone network of YOLOv5s to a lightweight backbone ShuffleNetV2 network, and the parameter size and detection time of each image, respectively. The reductions are 45% and 34%, while the accuracy is reduced by 1.4%. The results show that using ShuffleNetV2 as the backbone network is more likely to be applied in practice and deployed on embedded devices. (2) Then, on the basis of the ShuffleNetV2 backbone network, the original PANet feature fusion network was replaced with BiFPN, the parameter size was reduced by 29% compared with the original YOLOv5s, and the accuracy was increased by 0.5% compared with the ShuffleNetV2 backbone network, which indicates that the BiFPN network. In this experiment, better fusion features can be used for road defect detection. (3) Secondly, the ECA-Net attention mechanism was integrated into the ShuffleNetV2 network, and a new backbone network named Shuffle-ECANet was created. Compared with the original YOLOv5s, the accuracy remained unchanged but the number of parameters was reduced, indicating that ECA-Net can be more focused on extracting useful information from features. (4) Finally, the Focal-EIOU loss function was used to replace the original CIOU localization loss function in the training phase. On the Shuffle-ECANet and BiFPN structures, the accuracy was improved by 0.3%, which proves that the Focal-EIOU loss function can make the model perform better regression and obtain higher-quality anchor boxes.

Table 1 Ablation study on YOLO-LRDD

Method			Precision (%)	#Param. (M)	F1-Score (%)	Inference time per image (ms)
Backbone	**Neck**	**Loss**				
CSPdarknet + SPPF	PANet	CIOU	56.8	27.6	57.5	4.4
ShuffleNetV2	PANet	CIOU	55.4	**15.2**	55.6	**2.9**
ShuffleNetV2	BiFPN	CIOU	55.9	18.5	56.5	3.1
Shuffle-ECANet	BiFPN	CIOU	56.8	19.6	57.1	3.2
Shuffle-ECANet	BiFPN	Focal-EIOU	**57.1**	19.8	**58.7**	3.4

The best results are marked in bold
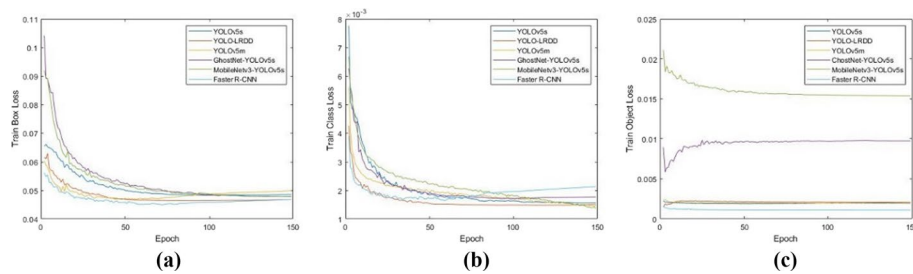
Fig. 7 Comparison of training losses of six types of methods

Table 2 Comparison of detection results between YOLO-LRDD and the other five methods

Model	mAP$_{0.5}$ (%)	mAP$_{0.95}$ (%)	Precision (%)	Recall (%)	F1-Score (%)	#Parama. (M)	FLOPs (G)	FPS
YOLOv5s	56.9	29.4	58.9	56.2	57.5	27.6	28.4	72
YOLOv5m	57.3	**38.9**	58.5	57.3	57.8	33.2	32.2	64
Mobile-Netv3-YOLOv5s	51.2	25.8	54.7	53.8	54.2	19.2	24.2	78
GhostNet-YOLOv5s	49.5	24.5	53.6	52.5	53.0	20.4	23.4	77
Faster R-CNN	**60.2**	31.8	**62.7**	**59.3**	**60.1**	38.3	34.6	51
YOLO-LRDD	57.6	30.6	59.2	58.2	58.7	**19.8**	**17.4**	**86**

The best results are marked in bold

3.3 Comparison with various methods

In this section, we compare the performance of the proposed YOLO-LRDD model with five other state-of-the-art models, including two one-stage lightweight models with YOLOv5s as the baseline, MobileNetv3-YOLOv5s, and GhostNet-YOLOv5s, respectively, two one-stage models including YOLOv5s and YOLOv5m.

3.3.1 Numerical analysis of the RDDC dataset

The loss represents the difference between the predicted value and the actual value. With the gradual narrowing and convergence of the gap, it means that the model is close to the upper limit of performance determined by the dataset. The comparison of the training loss function curves of the six methods is shown in Fig. 7.

As shown in Fig. 7, the loss values for each category fluctuate considerably at the beginning of the training, indicating that the initial hyperparameters were reasonable. After a certain number of iterations, the fluctuation of the loss curve gradually decreases. In Fig. 7a, it can be seen that the results compared with the other five methods can be seen that the Box curve converges faster and more stably when the Focal-EIOU loss is used in YOLO-LRDD than the CIOU loss in YOLOv5s. The trained loss curves are more convergent and stable compared to the other methods, providing higher position precision and greater stability and robustness in road damage detection.

To comprehensively verify the proposed YOLO-LRDD network's performance, the six object detection methods are quantitatively compared, and the comparison results are shown in Table 2.

By analyzing the experimental results, it can be seen that the model Precision of YOLOv5s and YOLOv5m is 58.9% and 58.5%, respectively, while the Precision of our proposed YOLO-LRDD model is 59.2%, an improvement of 0.7%, which is the result of the lightweighting of the model. (The size of the model is reduced from 32.2G to 17.4G.) This means that the YOLO-LRDD algorithm can still be highly accurate in embedded applications, and the frames per second transmitted (FPS) by the model has also improved significantly, from 64 to 86 FPS, an improvement of 25.6%, which will result in smoother and more consistent detection of road damage images.

3.3.2 Visualization results on the RDDC dataset

In this section, we compare YOLO-LRDD with another five detection methods in four distinct circumstances and display the predicted labels and predicted values in these samples in order to more intuitively observe the detection effect of YOLO-LRDD. The accuracy of the model prediction increases with the size of the projected value in these samples. Comparing the six prediction models to the actual results in the samples allows one to easily see how accurate the prediction anchor box is. These are the detecting instructions.

The general road environment is very complex, especially on rural roads. In order to comprehensively measure the performance of the YOLO-LRDD model, we used a single small target and unevenly exposed targets for visual experiments. As shown in Fig. 8, we found that the detection performance of the Yolov5m and GhostNet-Yolov5s models was significantly disturbed by rutting and lane boundaries and was less resistant to interference. YOLO-LRDD was highly resistant to interference in this environment.

As shown in Fig. 9, we have conducted a series of experiments on the performance of the six detection models in the case of multiple small targets, in which the image has the reflection effect. It can be seen from this that the detection performance of
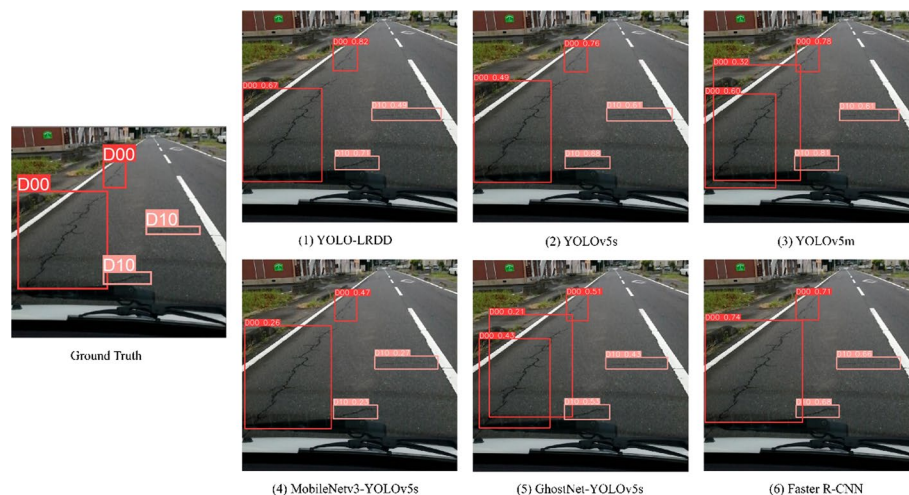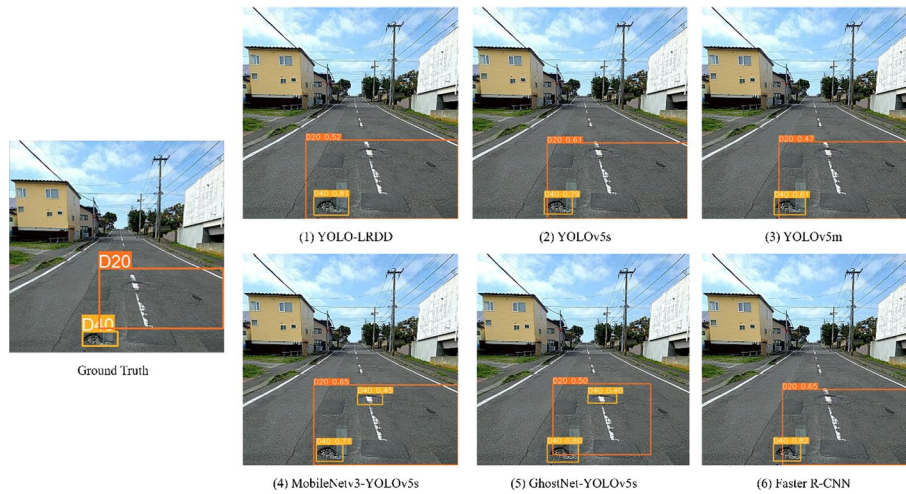

Fig. 8 Comparison of the detection of longitudinal cracks and lateral cracks
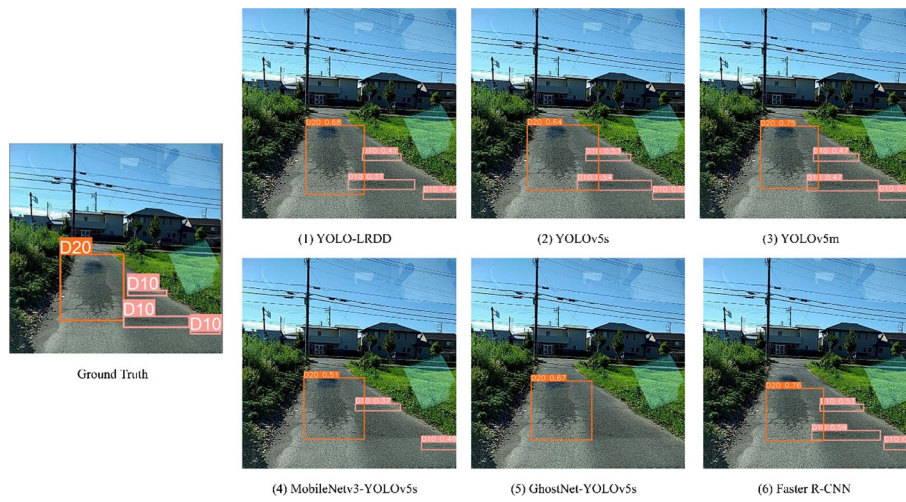
Fig. 9 Comparison of the detection in multiple small lateral cracks and alligators

Fig. 10 Comparison of the detection of alligators and potholes under strong light

GhostNet-YOLOv5s is general, and the small lateral cracks nearby are omitted, indicating that the model is insensitive to small targets, and the performance of the GhostNet-YOLOv5s model is the worst. Recently, the hole has been leaked, and the performance has serious defects. YOLO-LRDD model performs best in this test. It can detect potholes in this range and show an absolute advantage in performance.

As shown in Fig. 10, a series of experiments were carried out on the road damage detection performance of six methods under strong light. It can be seen that the YOLOv5s method can detect defects, but the confidence is very low. Faster R-CNN and GhostNet-YOLOv5s methods are greatly disturbed by light intensity, resulting in a significant decline in detection performance. Mobilenet3-YOLOv5s has error detection due to the interference of lane boundary. YOLO-LRDD can detect damages with high reliability and accuracy. It is not affected by uneven light intensity and has strong resistance to the external environment.

Fig. 11 Comparison of the detection in multi damaged road sections

As shown in Fig. 11, in the case of low exposure and partial shadow occlusion, these methods have similar detection ability, and a small part of shadow has little impact on the model. However, in the test, we found that GhostNet-YOLOv5s method has the lowest confidence, while YOLO-LRDD method has the highest detection confidence, and only this method with YOLOv5s can detect pits of small objects. This shows that Focal-EIOU loss's treatment of sample imbalance enhances the detection ability of the model for small objects to a certain extent. It can be seen that by using BiFPN to enhance feature extraction, the detection precision of the model can be improved to a certain extent. In addition, by optimizing the sample imbalance processing method, the sensitivity of small object recognition can be improved, the missed detection rate of targets with unclear features can be reduced, and better road defect detection performance can be achieved. By further testing the performance of our improved model, we confirmed its performance advantages. Compared with GhostNet-YOLOv5s, YOLOv5s, and Faster R-CNN methods, YOLO-LRDD shows stronger practical advantages. In the quantitative evaluation results and qualitative analysis, the YOLO-LRDD method proposed in this paper has strong anti-interference ability, high sensitivity to small targets, low missed detection rate of multiple targets, little influence by external environmental interference, good robustness, and strong versatility.

4 Conclusion

Firstly, in this paper, we have compiled a new dataset, RDDC, that is based on the RDD2020 and includes a notable sample of Chinese road damage photographs. This dataset contains four sorts of frequent damage situations, such as longitudinal cracks, lateral cracks, alligators, and potholes. The larger RDDC dataset permits the algorithm to be trained to have a higher capacity for generalization, resulting in a slight gain in algorithmic precision. In addition, the enhanced RDDC dataset can make the algorithm's strengths and flaws more apparent when comparing the algorithm's superiority.

Secondly, we present the most appropriate YOLO-LRDD deep network learning model for damage identification in real-world road scenarios, which has less parameters

Wan *et al. EURASIP Journal on Advances in Signal Processing*   (2022) 2022:98

Page 16 of 18

and less computation and outperforms previous object detection models based on test model performance. We propose a new backbone network called ShuffleECA-Net, which adds an ECA lightweight attention mechanism to the lightweight network Shuffle-NetV2. ShuffleECA-Net decreases model weights and increases the speed of detection.

To improve the accuracy of model detection, we employ BiFPN rather than PANet, which may enrich the description of features more efficiently. In the last phase of the model's training, Focal-EIOU loss is employed to correct the imbalance of the samples and create anchor boxes of greater quality. In the end, compared with YOLOv5s, our proposed YOLO-LRDD model reduces the model size by 28.8% and improves the accuracy by 0.3%, which is more suitable for lightweight requirements.

5 Future work

Our research is cutting-edge. Nowadays, many models pursue high accuracy and increase the number of layers of the network indefinitely. In this paper, we build a lightweight algorithm model without significantly reducing the accuracy of the algorithm, and it is clear that the improvement idea of our algorithm is a reference for future model lightweight. The YOLO-LRDD algorithm also expands for lightweight computer vision models, which are informative for innovations in road defect real-world detection techniques. The lightweight model allows it to be equipped on mobile devices and in-vehicle tools. It is important for the real-time detection of road defects and their timely maintenance.

For further research on road damage detection algorithms such as YOLO-LRDD, we propose beginning with the two aspects listed below:

A database of such samples will be collected and improved in the future in order to be able to include the added dataset in future model training. This will allow the model to be trained in order to produce a better and more complete road damage detection model. Improving the algorithm and the detection accuracy of transverse cracks in road damage will be accomplished by collecting and improving the database of such samples.

While maintaining a lightweight, or more lightweight, model to improve accuracy, there are still many variables that need to be debugged in the algorithm proposed in this paper, and changes in these variables can affect the accuracy of the algorithm. Future research should be directed toward reducing the number of uncontrollable variables in the algorithm so that its accuracy can be improved.

Abbreviations
YOLO   You only look once
YOLO-LRDDYOLOv5s-Lightweight road damage detection
RDDC   Road damage detection collection
IOU   Intersection over union
AP   Average precision
mAP   Mean average precision
FLOPs   Floating point operations per second

Author's information
Fang Wan, lecturer at the School of Computer Science, Hubei University of Technology, China, graduated from the School of Computer Science of Wuhan University with a PhD Founder of the Digital Twin Team of Hubei University of Technology. His main research interests are computer vision and computer graphics.

Chen Sun, graduate student at the School of Computer Science, Hubei University of Technology, China. Her current research fields include deep learning and object detection.

Hongyang He is studying for his MPhil in the School of Engineering and Physical Sciences at the University of Birmingham. He is also a member of the Institute of Data Science and Artificial Intelligence, where his recent research interests focus on deep learning-based signal processing and the use of machine learning algorithms to solve problems in hardware, as well as the optimization of IoT algorithms.

Guangbo Lei, lecturer at the School of Computer Science, Hubei University of Technology, graduate of Chongqing University, China, her current research field is deep learning and lightweight BIM model.

Li Xu, lecturer at the School of Computer Science, Hubei University of Technology, China, graduated from the School of Computer Science of Wuhan University with a master's degree, research directions are database technology, computer graphics, and computer vision.

Teng Xiao, received PhD degrees from the School of Geodesy and Geomatics of Wuhan University, China, in 2021. From 2018 to 2019, he is a visiting PhD student at the Institute of Photogrammetry and Geoinformation of Leibniz University Hannover, Germany. He is currently a lecturer of Hubei University of Technology, China. His main research interests are photogrammetry and computer vision.

Author contributions
All authors have contributed toward this work as well as in compilation of this manuscript. All authors read and approved the final manuscript.

Availability of data and materials
Please contact author for data requests.

Declarations

Ethics approval and consent to participate
These manuscripts do not involve human participants, human data, or human tissue and include a statement on ethics approval and consent. These manuscripts do not involve ethical approval and consent and include the name of the ethics committee that approved the study and the committee's reference number if appropriate. The studies not involving animals must include a statement on ethics approval.

Competing interests
The authors declare that they have no competing interests.

References

1. N. H. T. S. Administration. *National Highway Traffic Safety Administration Technical Report DOT HS* vol. 811 (2008). p. 059
2. M.E. Torbaghan, W. Li, N. Metje, M. Burrow, D.N. Chapman, C.D. Rogers, Automated detection of cracks in roads using ground penetrating radar. J. Appl. Geophys. **179**, 104118 (2020)
3. G.M. Hadjidemetriou, P.A. Vela, S.E. Christodoulou, Automated pavement patch detection and quantification using support vector machines. J. Comput. Civ. Eng. **32**, 04017073 (2018)
4. T. S. Nguyen, S. Begot, F. Duculty, M. Avila, in *2011 18th IEEE International Conference on Image Processing* (IEEE, 2011), p. 1069
5. H. Nguyen, L. Nguyen, D.N. Sidorov, A robust approach for road pavement defects detection and classification. J. Comput. Eng. Math. **3**, 40 (2016)
6. N. Safaei, O. Smadi, A. Masoud, B. Safaei, An automatic image processing algorithm based on crack pixel density for pavement crack detection and classification. Int. J. Pavement Res. Technol. **15**, 159 (2022)
7. A. Cubero-Fernandez, F. J. Rodriguez-Lozano, R. Villatoro, J. Olivares, J. M. Palomares, Efficient pavement crack detection and classification. *EURASIP J. Image Video Process.* **2017** (2017)
8. Y. Wang, K. Song, J. Liu, H. Dong, Y. Yan, P. Jiang, RENet: Rectangular convolution pyramid and edge enhancement network for salient object detection of pavement cracks. Measurement **170**, 108698 (2021)
9. K. Madasamy, V. Shanmuganathan, V. Kandasamy, M.Y. Lee, M. Thangadurai, OSDDY: embedded system-based object surveillance detection system with small drone using deep YOLO. EURASIP J. Image Video Process. **2021**, 1 (2021)
10. N. Aloysius, M. Geetha, A review on deep convolutional neural networks, in *2017 International Conference on Communication and Signal Processing (ICCSP)* (IEEE, 2017), p. 0588
11. F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang, Residual attention network for image classification, in *Proceedings of the IEEE Conference on Computer Vision and PATTERN recognition* (2017), p. 3156
12. Z.-Q. Zhao, P. Zheng, S.-T. Xu, X. Wu, Object detection with deep learning: a review. IEEE Trans. Neural Netw. Learn. Syst. **30**, 3212 (2019)
13. P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, G. Cottrell, Understanding convolution for semantic segmentation, in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* (IEEE, 2018), p. 1451

14. Y. Xu, D. Li, Q. Xie, Q. Wu, J. Wang, Road damage detection and classification with faster R-CNN. Automatic defect detection and segmentation of tunnel surface using modified Mask R-CNN. Measurement **178**, 109316 (2021)
15. W. Wang, B. Wu, S. Yang, Z. Wang, in *2018 IEEE International Conference on Big Data (Big Data)* (IEEE, 2018), p. 5220
16. V. Hegde, D. Trivedi, A. Alfarrarjeh, A. Deepak, S. H. Kim, and C. Shahabi, Yet another deep learning approach for road damage detection using ensemble learning, in *2020 IEEE International Conference on Big Data (Big Data)* (IEEE, 2020), p. 5553.
17. Q. Wang, J. Mao, X. Zhai, J. Gui, W. Shen, Y. Liu, Improvements of YoloV3 for road damage detection, in *Journal of Physics: Conference Series* (IOP Publishing, 2021), p. 012008.
18. S. Shim, J. Kim, S.-W. Lee, G.-C. Cho, Road surface damage detection based on hierarchical architecture using lightweight auto-encoder network. Autom. Constr. **130**, 103833 (2021)
19. A. Sheta, H. Turabieh, S. Aljahdali, A. Alangari, Pavement crack detection using a lightweight convolutional neural network, in *Proceedings of 35th International Conference*, vol. 69 (2020). p. 214
20. K. Guo, C. He, M. Yang, S. Wang, A pavement distresses identification method optimized for YOLOv5s. Sci. Rep. **12**, 1 (2022)
21. S. Vicente, J. Carreira, L. Agapito, J. Batista, Reconstructing Pascal voc, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), p. 41
22. S. Shim, J. Kim, S.-W. Lee, G.-C. Cho, Road damage detection using super-resolution and semi-supervised learning with generative adversarial network. Autom. Constr. **135**, 104139 (2022)
23. H. Maeda, T. Kashiyama, Y. Sekimoto, T. Seto, H. Omata, Generative adversarial network for road damage detection. Comput. Aided Civ. Infrastruct. Eng. **36**, 47 (2021)
24. N. Ma, X. Zhang, H.-T. Zheng, J. Sun, Shufflenet v2: Practical guidelines for efficient CNN architecture design, in *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), p. 116
25. Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) **2020**, 11531–11539 (2020). https://doi.org/10.1109/CVPR42600.2020.01155
26. M. Tan, R. Pang, Q. V. Le, Efficientdet: Scalable and efficient object detection, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), p. 107
27. S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, Path aggregation network for instance segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), p. 8759
28. Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, Focal and efficient IOU loss for accurate bounding box regression. arXiv:2101.08158 (2021)
29. Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, D. Ren, Distance-IoU loss: Faster and better learning for bounding box regression, in *Proceedings of the AAAI Conference on Artificial Intelligence* (2020), p. 12993
30. D. Arya, H. Maeda, S.K. Ghosh, D. Toshniwal, Y. Sekimoto, RDD2020: An annotated image dataset for automatic road damage detection using deep learning. Data Brief **36**, 107133 (2021)
31. X. Zhang, X. Zhou, M. Lin, J. Sun, Shufflenet: An extremely efficient convolutional neural network for mobile devices, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), p. 6848
32. J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), p. 7132
33. T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), p. 2117
34. U. Ruby, V. Yendapalli, Binary cross entropy with deep learning technique for image classification. Int. J. Adv. Trends Comput. Sci. Eng. **9** (2020)

Publisher's Note