Application Research of Bridge Damage Detection Based on the Improved Lightweight Convolutional Neural Network Model

Fujun Du, Shuangjian Jiao * and Kaili Chu

Department of Civil Engineering, College of Engineering, Ocean University of China, Qingdao 266100, China; dufujun@stu.ouc.edu.cn (F.D.); ckl@stu.ouc.edu.cn (K.C.)

*Correspondence: oucjiaoshuangjian@163.com

Abstract: To ensure the safety and rational use of bridge traffic lines, the existing bridge structural damage detection models are not perfect for feature extraction and have difficulty meeting the practicability of detection equipment. Based on the YOLO (You Only Look Once) algorithm, this paper proposes a lightweight target detection algorithm with enhanced feature extraction of bridge structural damage. The BIFPN (Bidirectional Feature Pyramid Network) network structure is used for multi-scale feature fusion, which enhances the ability to extract damage features of bridge structures, and uses EFL (Equalized Focal Loss) to optimize the sample imbalance processing mechanism, which improves the accuracy of bridge structure damage target detection. The evaluation test of the model has been carried out in the constructed BDD (Bridge Damage Dataset) dataset. Compared with the YOLOv3-tiny, YOLOv5S, and B-YOLOv5S models, the mAP@.5 of the BE-YOLOv5S model increased by 45.1%, 2%, and 1.6% respectively. The analysis and comparison of the experimental results prove that the BE-YOLOv5S network model proposed in this paper has a better performance and a more reliable performance in the detection of bridge structural damage. It can meet the needs of bridge structure damage detection engineering with high requirements for real-time and flexibility.

Keywords: healthy structure detection; bridge structure; deep learning; lightweight convolutional neural network; YOLOv5S

1. Introduction

In recent years, with the deepening of the industrialization process in various countries in the world, the problem of aging infrastructure brought about by industrialization has become increasingly prominent, and the timely detection and improvement of aging infrastructure have become a focus of global attention. According to the American Society of Civil Engineers’ 2021 Infrastructure Transcript Report [1], there are more than 617,000 bridges in the United States, 42% of which are at least 50 years old, and about 7.5% of the bridges in the country are already defective to varying degrees. In China, according to official statistics, by the end of 2014, the number of highway bridges had reached 750,000 [2]. With the increase in the number, the problems caused by bridge structural damage have become more and more serious. From 2000 to 2014, 179 bridges collapsed in China [3], causing irreversible losses to the national economy and national life safety. Therefore, bridge safety is indispensable in the construction of our modern society and plays a very important role in the travel of human traffic and the connection of regional economies [4]. However, the aging problem of bridges caused by factors such as repetitive vehicle loading, temperature difference, corrosion, and human damage is becoming more and more prominent. If the bridge maintenance department can detect the potential problems of the bridge promptly and take measures to fix them, the loss of economy and life can be well reduced. For the normal operation of bridges, regional governments usually formulate strict laws to restrict the bridge inspection departments. For example, in North America, the inspection of bridges uses strict standards (The Ontario Structure Inspection Manual) to conduct routine

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
inspections every two years by manual means to ensure the daily quality of bridges [5]. In the United States, the government also conducts manual and periodic bridge damage inspections according to the strict National Bridge Inspection Standard (NBIS) [6]. In Europe, the quality of bridges is assessed by manual inspections using the standards set by TU1406 [7]. In order to meet the requirements of laws and regulations and ensure the safety of people’s transportation and the needs of the national economy, the bridge maintenance departments in various regions need to consume a lot of manpower and material resources to carry out routine inspections of bridge maintenance.

The inspection of bridge quality is not only required for daily bridge maintenance. During the life cycle of the bridge, it may suffer from various unpredictable hazards, and it is necessary to carry out a performance evaluation and prediction of the bridge in each period through regular inspection of the bridge [8]. The United States conducts visual inspections of the physical condition of bridge decks, superstructures, and substructures by using the National Bridge Inventory (NBI) Condition Rating System [9], quantifying the results of inspections to assess bridge life cycle reliability [10], life cycle cost [11], risk [12], sustainability [13], and utility [14,15]. Through the analysis of the above indicators, decision-makers can make optimal decisions on the design, construction, maintenance, repair, and management of the bridge’s life cycle under expected conditions to maximize the benefits of the bridge life cycle. Therefore, the detailed condition detection and evaluation of bridges is an important basis upon which to achieve optimal decision-making. However, the traditional inspection method mainly relies on manual visual inspection, which not only causes huge inspection cost, but also causes problems of low efficiency and inconsistent standards (affected by subjective factors) [16,17].

At present, the working methods for bridge defect detection can be roughly divided into four categories: manual inspection [18], hardware equipment participating in detection [19], traditional image processing methods [20], and deep learning-based computer vision technology [21]. The detection of bridge damage by manual inspection is currently the main method around the world. For example, mentioned in Section 1, bridge evaluations in the United States and China are still mainly conducted by manual inspection. The manual inspection usually first conducts professional and safety skills training for inspectors, and then searches and records the damage to the bridge through visual inspection on the bridge site, and then conducts a quantitative analysis after sorting. This method is currently used in most areas mainly because of its simplicity. However, there are high labor costs, which cause a large load on the inspection labor force, which is very time-consuming and inefficient. For bridges in bad weather and under dangerous conditions, the personal safety of the inspectors has been greatly challenged, and the detection accuracy is highly related to the professionalism of the inspectors. Therefore, at present, relevant researchers in defect detection are actively seeking alternative methods [20]. The non-destructive testing of bridge hardware equipment started relatively early. As early as 2007, American scientific researcher Sherif Yehia et al. [22] conducted a comparative analysis of three common bridge damage detection pieces of equipment: ground-penetrating radar, shock wave and infrared imaging. The results show that infrared thermal imaging is fast, easily affected by the environment, and has unreliable performance; the shock wave speed is slow and is greatly affected by the surface roughness. Ground penetrating radar is an ideal detection method, but the defect size must be greater than 0.5 inches to be detected, and even the lane needs to be closed for detection. The detection equipment mostly requires manual assistance, the equipment price is relatively high, and currently, hardware equipment is mostly used in the research of damage detection in the surface structure of bridges [23–25]; it is difficult to use on the side of bridges and under bridges overhead on land and sea, with considerable limitations, and it has difficulty meeting the needs of structural damage detection of viaducts, high-speed railway bridges and other land or sea overhead bridges in practical projects, so it is rarely used in practice. F. Huseynov et al. [26] proposed the use of the Weighing Motion (WIM) system for damage identification of bridge structures. The WIM system is a system including sensors, computers and cameras installed in bridges. The installation
often needs to be carried out through closed roads, and the WIN system is expensive, making it difficult to deploy large-scale applications for local bridges. At the same time, the biggest difficulty in bridge structural health inspection is how to accurately distinguish the impact of damage on the behavior of the structure from the impact of environmental and operational changes (EOV) [27]. Because of its dynamic characteristics, this method is sensitive to changes in the environment and operation, such as temperature, humidity, and sensor sensitivity. Therefore, it is more complicated in practical engineering applications and has difficulty meeting the requirements of low cost and high efficiency. Traditional image processing technology can use ordinary RGB images to conduct research on defect detection on a bridge structure’s surface. The traditional image processing technology generally needs to set the features manually, based on features such as color [28], shape [29], texture [30] and others. The classification of the classifier is carried out to complete the identification of bridge damage. N. T. Sy et al. [31] carried out detection research on the characterization of pavements through three traditional image processing techniques of double-layer threshold, morphological operation, and projection, and achieved ideal results under the condition of low complexity. However, the feature selection of this method is too simple, which makes it difficult to meet the complex background of practical engineering. Nhat-Duc Hoang [32] adjusted the gray level of the image, then used the Otsu method to preprocess the image, and then cooperated with the shaping algorithm to detect damage. The experimental results show that it needs to be combined with other shapes to perform accurate detection, and the detection method is complex and has difficulty meeting the requirements of real-time performance.

In recent years, with the accelerated development of computer hardware facilities, the computing power of GPU has been greatly improved, which has created a material basis for computer technology based on deep learning theory. On top of this, computer vision technology based on deep learning theory has been developed rapidly. Compared with the target detection of traditional image processing methods, target detection based on deep learning theory has a good performance in terms of generalization and robustness [33,34]. Generally, we divide the target detection algorithm based on deep learning theory into a two-stage detection algorithm and a single-stage detection algorithm. The two-stage detection algorithm is the pioneer of the target detection algorithm based on deep learning. The representative two-stage detection algorithms include Msk R-CNN [35], Fast RCNN [36] and Faster RCNN [37], and other network algorithms. The algorithm principle is generally to first generate a candidate frame and then use a convolutional neural network to extract features and classify them. Single-stage detection algorithms usually complete feature extraction, classification, and prediction in one step, with the advantages of small model size and fast speed. Typical single-stage detection algorithms include the SSD algorithm [38] and the YOLO series algorithm [39] etc. Their accuracy is lower than that of the usual two-stage algorithms, but with the introduction of the sample imbalance function, the performance of the single-stage object detection algorithm is also improving and even surpassing the two-stage detection algorithm [40]. However, at present, most of the bridge defect detection based on deep learning only focuses on the detection of damage to bridge pavements. For example, Jinsong Zhu et al. [41] realized defect detection of a concrete bridge pavement by improving the classical convolutional neural network and visual geometry group network-16 (VGG-16). The substructure of the bridge, as its load-bearing structure, is often more critical to the maintenance of the bridge. Licun Yu et al. [42] realized the detection of bridge damage by improving Faster RCNN and found that its accuracy is higher through experiments. However, the Faster RCNN network model is large, the flexibility is poor, and the detection speed is also slow, which makes it difficult to meet the real-time and flexibility requirements of practical engineering. Evan McLaughlin et al. [43] used deep learning theory combined with infrared and lidar to achieve end-to-end detection of regional defects in bridges, but the detection cost is relatively high. Ma, D et al. [44] improved the YOLOv3 network model to detect the number of cracks, and the detection speed has been greatly improved compared with the
two-stage detection algorithm. However, the model only performs single classification, and its generalization performance is difficult to determine for practical projects. Ping et al. [45] conducted a comparative analysis of YOLO, SSD, HOG with SVM and Faster R-CNN network models for the detection of concrete defects. The results show that the YOLOv3 model of the YOLO network algorithm series is the most robust for defect detection and has a reliable application prospect. Carlos Peña-Caballero et al. [46] proposed to deploy the YOLO algorithm into an embeddable device to detect pavement defects, but the network structure of the YOLOv3 network model is too large and has poor flexibility, which cannot meet the conditions of general hardware. Khaled R. Ahmed [47] conducted experimental analysis on three algorithms, YOLOv5, YOLOR and Faster R-CNN, for road surface defect detection. The results show that the YOLOv5 model is extremely flexible and suitable for real-time detection scenarios of embedded devices. However, in terms of accuracy (mAP@.5: 58.9%), further improvement is needed.

Although a large number of scientific researchers are currently conducting research in different defect detection fields, they mainly focus on the detection of pavement defects, largely cracks and potholes [21,48], and the computer vision technology for bridge load-bearing structures is less detected. There are many viaducts and bridges across waters in the world. The damage detection of load-bearing structures of such bridges is difficult to achieve through traditional methods. Therefore, there is an urgent need for a detection model with strong flexibility and reliable performance to meet the requirements of safety, low cost, high efficiency, and strong reliability for most bridge structure damage detection in practical engineering. Compared with other detection methods, the method based on computer vision can effectively reduce the detection cost while ensuring the safety of workers and improving efficiency. However, the current computer vision technology for bridge structural damage still lacks datasets, the target detection model is inflexible, and the detection speed is slow, which makes it difficult to meet the difficulty and practicability needs of actual bridge structural damage detection engineering. Given the above difficulties, in this paper, we propose the BE-YOLOv5S lightweight convolutional neural network model based on the YOLOv5S algorithm. The main contributions are as follows: (1) Build a bridge damage dataset and make it available to the public to provide practical data for the study of bridge structural damage detection; (2) Improve the backbone network of the lightweight convolutional neural network model to improve the performance of the lightweight model in damage detection of bridge structures; (3) Improve the sample imbalance processing mechanism in the lightweight convolutional neural network model and improve the training quality of samples for bridge damage models; (4) Establish a lightweight convolutional neural network model that is flexible, advanced, and practical for bridge damage detection to meet the needs of portable devices (such as drones) to achieve end-to-end rapid detection of bridge damage.

This paper is organized in the following sections as follows: Section 2 introduces the experimental method for bridge damage detection; Section 3 describes the results of the experiment, Section 4 discusses the analysis and the experiment, and Section 5 summarizes the paper.

2. Methods
2.1. Algorithm Introduction and BE-YOLOv5S Structure

Aiming at the specificity of the difficulty in detecting load-bearing structures of common ground bridges, viaducts, and bridges across waters [49], we urgently need a bridge damage detection model with high flexibility and reliable performance that can be easily embedded in small devices. The YOLOv5S target detection algorithm [50], a typical single-stage target detection algorithm, is the latest version of the YOLO series of algorithms and has shown good performance in defect detection with lightweight convolutional neural networks [51]. In this paper, we use the YOLOv5S network as the basic network and make improvements to establish the BE-YOLOv5S bridge damage detection model.
In the BE-YOLOv5S algorithm, we add the BiFPN network structure [52,53], which further strengthens the feature extraction ability for bridge damage images, which will be introduced in detail in the second part of this section. The network structure of BE-YOLOv5S is shown in Figure 1, and the detailed composition of each module is shown in Figure 2. We mainly divide the network structure of BE-YOLOv5S into four parts; the first part is the image input part for bridge damage detection, the second part is the backbone of the BE-YOLOv5S network, the third part is the enhanced feature extraction network part, and the last part is the detection layer. At the same time, in this paper, we improve the processing mechanism of the sample unbalance function of the network according to the specificity of the bridge damage structure sample image, which is introduced in detail in the third summary of this section.

Figure 1. The flow chart of BE-YOLOv5S network structure. (The data transmission process goes through Image Input, Backbone, Feature extraction Network’s and Detection Network from left to right respectively).

2.2. Feature Extraction Networks of the BE-YOLOv5S Model

Compared with the general neural network or traditional image processing technology, the target detection algorithm based on the deep learning network theory has a deeper network layer, and the feature extraction ability is gradually strengthened. However, with the continuous deepening of the number of convolutional neural network layers, the image feature information about bridge damage continues to be lost during transmission [54]. To enable the BE-YOLOv5S network model to obtain sufficient bridge damage characteristic information and improve the robustness of the model, and aiming at the characteristics of bridge damage with inconspicuous features, diverse shapes and complex backgrounds, we use the BiFPN network structure to enhance the feature extraction part of the YOLOv5S network.
which is defined as Equation (1) below. In Definition 1, the learning weight $w_i$ uses the ReLU activation function and takes the value $\varepsilon = 0.0001$ to enhance the numerical stability. Through the above operations, the characteristic information about the bridge damage is further fully integrated, and the performance of the BE-YOLOv5S bridge damage detection model is improved.

In the YOLOv5S network, the PANet network structure is used as the feature extraction network of the image. The network structure is shown in Figure 3a. PANet first proposed a top-down and bottom-up bidirectional backbone fusion network and built a “short-cut” between the bottom layer and the top layer to reduce the information transmission path at the top and bottom ends. Although PANet has a certain efficiency in the transmission of information when it transmits feature information, it adopts the same transmission priority for information of different scales. Deep features often contain richer semantic information, while shallow extracted features contain relatively more detailed information (local features and positions), but PANet treats features of different depths equally to process information. For example, both the feature information passed from $P_{5_{\text{out}}}$ through upsampling and the feature information from $P_{3_{\text{in}}}$ are found in the $P_{5_{\text{out}}}$ node. The feature information extracted by different feature layers is different. However, PANet processes both types of information with the same weight, ignoring the importance of deep feature information to some extent. Compared with the PANet structure of the YOLOv5S feature extraction network, in Figure 3b, the BE-YOLOv5S enhanced feature extraction network proposed in this paper adds residual connections, aiming to strengthen the feature representation ability of bridge damage through simple residual operations. As shown in Figure 3a, the nodes at the midpoints of both ends do not perform feature fusion, so the included bridge damage image feature information is less, and the final feature fusion effect is small. Therefore, it is pruned to further increase the speed of the BE-YOLOv5S model in bridge damage detection; different feature information is weighted. As mentioned above, different feature scales contain information of different importance. Through the fast normalized fusion method, the bridge damage images extracted from different scales are weighted and fused, which is defined as Equation (1) below. In Definition 1, the learning weight $w_i$ uses the ReLU activation function and takes the value $\varepsilon = 0.0001$ to enhance the numerical stability. Through the above operations, the characteristic information about the bridge damage is further fully integrated, and the performance of the BE-YOLOv5S bridge damage detection model is improved.

Figure 2. The detailed composition of each module in the BE-YOLOv5S network structure. (a) CONV module, (b) Bottleneck module (c) C3 $\times$ N module (d) Focus module, and (e) SPP module.
Through the interconnection and fusion between different layers, BiFPN’s bidirectional performance of the BE-YOLOv5S bridge damage detection model. To this end, in the work of this paper, we introduce Equalized Focal Loss [56] to improve the problem of unbalanced samples, and the problem increases with the degree of imbalance of positive and negative samples. Through the introduction of Fcoal Loss, its performance has improved. The main mechanism of action of Fcoal Loss is to control the weight of positive and negative samples and to control the weight of easy-to-classify and hard-to-classify samples. It is defined in the following Formula (3). Where $\alpha_i$ is the same as Definition 1. It is the learning weight. Through the interconnection and fusion between different layers, BiFPN’s bidirectional cross-scale connection and fast standardized fusion are finally realized. PANet and BiFPN feature extraction are shown in Figure 4.

$$O = \sum_i \frac{w_i}{\varepsilon + \sum_{j} w_j} \cdot I_i \quad (1)$$

$$P_{l,td}^{id} = \text{Conv}(\frac{w_1 \cdot P_{in} + w_2 \cdot \text{Resize}(P_{in})}{w_1 + w_2 + \varepsilon})$$

$$P_{l,out}^{out} = \text{Conv}(\frac{w_1 \cdot P_{in} + w_2 \cdot P_{td}^{id} + w_3 \cdot \text{Resize}(P_{out})}{w_1 + w_2 + w_3 + \varepsilon}) \quad (2)$$

The above Formula (2) is the definition formula of single-layer (not both ends) information transmission in the BiFPN network structure. Specifically, $P_{l,td}^{id}$ represents the intermediate feature of the top-to-bottom layer $l$, and $P_{l,out}^{out}$ is the output feature of the bottom-to-layer layer $l$, and $w_i$ is the same as Definition 1. It is the learning weight. Through the interconnection and fusion between different layers, BiFPN’s bidirectional cross-scale connection and fast standardized fusion are finally realized. PANet and BiFPN feature extraction are shown in Figure 4.

2.3. Improved Sample Imbalance Handling Mechanism for BE-YOLOv5S Bridge Damage Detection Model

The problem of sample imbalance has always restricted the development of single-stage object detection networks. With the introduction of Fcoal Loss, its performance has improved. The main mechanism of action of Fcoal Loss is to control the weight of positive and negative samples and to control the weight of easy-to-classify and hard-to-classify samples. It is defined in the following Formula (3). Where $pt \in [0,1]$ represents the prediction confidence score of a candidate frame target, and $\alpha_i$ is used as a balance parameter for balancing positive and negative samples in Fcoal Loss, $\gamma$ is the Fcoal parameter, which increases with the degree of imbalance of positive and negative samples. Through the use of the above definitions, the loss of simple samples can be reduced, and the focus of model learning can be biased towards complex and deep learning samples. However, for bridge damage targets, the traditional Fcoal Loss uses the same modulation factor for all damage target categories to balance the problem of unbalanced samples, and the problem of high similarity between bridge background and samples cannot be effectively solved. To this end, in the work of this paper, we introduce Equalized Focal Loss [56] to improve the performance of the BE-YOLOv5S bridge damage detection model.
2.3. Improved Sample Imbalance Handling Mechanism for BE-YOLOv5S Bridge Damage Detection Model

The problem of sample imbalance has always restricted the development of single-stage object detection networks. With the introduction of Fcoal Loss, its performance has improved. The main mechanism of action of Fcoal Loss is to control the weight of positive and negative samples and to control the weight of easy-to-classify and hard-to-classify samples. It is defined in the following Formula (3). Where \( p_t \in [0,1] \) represents the prediction confidence score of a candidate frame target, and \( \alpha_t \) is used as a balance parameter for balancing positive and negative samples in Fcoal Loss, \( \gamma \) is the Fcoal parameter, which increases with the degree of imbalance of positive and negative samples. Through the use of the above definitions, the loss of simple samples can be reduced, and the focus of model learning can be biased towards complex and deep learning samples. However, for bridge damage targets, the traditional Fcoal Loss uses the same modulation factor for all damage target categories to balance the problem of unbalanced samples, and the problem of high similarity between bridge background and samples cannot be effectively solved. To this end, in the work of this paper, we introduce Equalized Focal Loss \([56]\) to improve the performance of the BE-YOLOv5S bridge damage detection model.

This paper solves the problem of multi-class sample imbalance in the target of bridge damage by introducing the modulation factor, Focusing Factor, associated with the class. The definition of the Focusing Factor is as follows in Equation (4), where \( \alpha \) and \( p_t \) are the same as those in the above Equation (3), \( \gamma_j \) is the Focusing Factor of the \( j \)th category, and its function is similar to that of Fcoal Loss. Here, the value of \( \gamma_j \) is larger, which is used to alleviate the problem of serious sample imbalance. Of course, the small value of \( \gamma_j \) required for the samples with a small degree of sample imbalance in the bridge damage target is decoupled into two components, which are defined as Equation (5). Among them, \( \gamma^b \) represents the Focusing Factor that controls the basic behavior of the classifier, and the parameter \( \gamma_j^v \geq 0 \) represents the first variable with the unbalanced correlation degree of the \( j \)th class, It determines the attention distribution of BE-YOLOv5S in the process of training and learning the bridge damage model, and its value is determined according to the gradient guidance mechanism. The parameter \( g_j \) represents the cumulative gradient ratio of the \( j \)th class of positive and negative samples. When the value of \( g_j \) is large, it means that the training is balanced; otherwise, it is unbalanced. The hyperparameter \( \gamma \) is a scaling factor that controls the upper limit of \( \gamma_j \) in EFL. Through the application of the above definition, compared with Fcoal Loss, we can independently deal with different categories of damage targets encountered during the training process, thereby further improving the performance of the BE-YOLOv5S model in bridge damage detection.

\[
FL(p_t) = -\alpha (1 - p_t)^\gamma \log(p_t) \tag{3}
\]

\[
EFL = -\alpha (1 - p_t)^{\gamma^v} \log(p_t) \tag{4}
\]
\[
\gamma_j = \gamma_b + \gamma_j v
= \gamma_b + s(1 - g_j)
\] (5)

Currently, under the same \( x_t \) condition, the loss is negatively correlated with the value of \( \gamma \), which causes BE-YOLOv5S to learn a severely imbalanced bridge loss sample when learning a bridge loss sample. Having to sacrifice part of the loss contribution made in the learning process of the whole model makes our model lose performance on individual target classes to some extent. At the same time, under the condition that \( x_t \) is small, different bridge damage targets in different Focusing Factors will converge to a similar value, but the small number of individual categories cannot dominate the learning process of the BE-YOLOv5S model. Given the above situation, we introduce the Weighting Factor proposed by Bo Li et al. \[56\] to alleviate the above two types of situations by balancing the category loss of different bridge damage targets, which is defined as Equation (6). The Weighting Factor of the \( j \)-th bridge loss objective is defined as \( \gamma_b + \gamma_j v / \gamma_b \).

\[
EFL = -\sum_{j=1}^{C} a_i \left( \frac{\gamma_b + \gamma_j^v}{\gamma_b} \right) (1 - p_i) \gamma_b + \gamma_j^v \log(p_i)
\] (6)

Through the introduction of the above definitions, the BE-YOLOv5S model can improve the information of potential targets mined by the model and pay more attention to the learning difficulties encountered in training under the complex background of bridge damage, thereby improving the efficiency of model learning.

3. Experiment

At present, the YOLOv3-tiny algorithm has achieved remarkable results in the application research of many disciplines because of its excellent detection speed, flexibility and good robustness, such as in the field of public safety \[57\], the field of medical engineering \[58\], the field of agricultural disease prevention \[59\] and the field of industrial engineering \[60\], etc., which are representative in the field of lightweight target detection. To compare the contributions of this work through experiments, we selected the representative YOLOv3-Tiny network, YOLOv5S, B-YOLOv5S and BE-YOLOv5S in the field of lightweight target detection for comparative analysis experiments. Train–learn, test–comparison and validation analysis were performed in the same environment.

3.1. Development Environment and Evaluation Metrics

3.1.1. Development Environment

The work presented in this paper was established, validated, and tested in a laboratory environment, and the experimental results are analyzed and compared. The model development configuration for the lab server includes the hardware: CPU: Intel(R) Xeon(R) Gold 5218, GPU: GeForce RTX 2080 Ti/11GB, Software: System: ubuntu18.04, the accelerated training framework is CUDA10.2 and cuDNN7, we used pytorch1.7.0 as the training framework for the proposed and all compared models, and put all the experimental models in this paper in the same environment as above for testing, analysis, and comparison.

3.1.2. Evaluation Metrics

Model evaluation is crucial to an experiment and is an essential part of testing the performance of the model in the process of building the model. To fully verify the superiority of our model and comprehensively evaluate the performance of the model in all aspects, we introduced confusion matrix related indicators, Precision, Recall, F1-score, FPS and PR curve, and mAP@.5 model evaluation indicators, to conduct a comprehensive evaluation and analysis of the performance of the BE-YOLOv5S model in bridge defect detection, and further discuss it. In the following, we will introduce the functions and definitions of each index to evaluate the bridge damage detection model.
The Confusion Matrix

A confusion matrix is a standard format for accuracy evaluation in the field of target detection. In this paper, the confusion matrix was used as the most basic index to evaluate the performance of the above-mentioned models in terms of accuracy. The confusion matrix includes the definitions of four types of indicators, namely TN (predicting negative samples as negative samples), FN (predicting positive samples as negative samples), TP (predicting positive samples as positive samples), and FP (predicting negative samples as positive ones). The definitions of the above four types of confusion matrix indicators will also be applied to the following indicators as basic indicators.

Precision, Recall, F1-Score, and PR Curve

Accuracy is the evaluation index of how many real samples are predicted to be real samples in the bridge damage model assessment, that is, the proportion of real positive examples in the prediction results, which is defined as the following Formula (7). Recall (also known as recall) indicates how many samples of bridge damage are correctly predicted in the sample, that is, the proportion correctly predicted by the model in all positive examples, which is used to evaluate our experimental model on the sample. The detection coverage of the bridge damage target in the middle of the bridge is defined as the following Formula (8). In fact, in the actual experiment, the two indicators of Precision and Recall were a pair of contradictory measures. When Precision is high, Recall is low, and vice versa, to comprehensively consider the Precision and Recall performance of the experimental model in bridge damage target detection. We introduced the F1 score to perform weighted harmonic mean on the above two indicators, which is defined as the following Formula (9).

At the same time, in order positively show the performance of each model in this experiment in bridge damage detection, we introduced the PR curve evaluation method. The PR curve takes Recall as the abscissa and Precision as the ordinate to draw the curve. If the PR curve of model A can completely wrap up the PR curve of model B, it can be considered that the bridge damage detection performance of model A is better. If it cannot be judged directly, the comparison can be made according to the area under each model curve.

\[
\text{Precision}(P) = \frac{TP}{TP + FP} \quad (7)
\]

\[
\text{Recall}(R) = \frac{TP}{TP + FN} \quad (8)
\]

\[
\text{F1 score}(F1) = 2 \times \frac{P \times R}{P + R} \quad (9)
\]

Mean Average Precision IoU = 0.5 (mAP@.5)

Accuracy is a common evaluation index in the evaluation of target detection models. It evaluates the proportion of positive samples and negative samples accurately predicted in the bridge damage samples of this model. In the field of target detection, what is different from the traditional classification is the participation of Intersection over Union (IoU). The confidence level restricts the Accuracy index. To better demonstrate the performance of the model in bridge damage detection, we calculated Accuracy under the condition of IoU = 0.5, that is, under the condition of IoU = 0.5, we calculated the average precision (AP) of the detection results of the three types of bridge damage targets, which is defined as the following Formula (10) and then average the mean, which is defined as follows in Equation (11):

\[
AP = \frac{1}{11} \sum_{\gamma \in \{0,0.1,0.2,\ldots,1\}} \text{interp}(r) \quad (10)
\]

\[
mAP = \frac{\sum_{i=1}^{3} AP_i}{3} \quad (\text{IoU} = 0.5) \quad (11)
\]
Frames Per Second (FPS)

The detection speed is a key factor for bridge damage detection, especially in hazardous and harsh environments. Rapid detection allows our embedded mobile devices to quickly conclude damage detection to prevent potential damage from risk (for example, crushed by the wind and waves in the detection of a bridge across the sea, etc.). FPS is a commonly used definition of speed in the image field, which expresses the number of frames transmitted per second. Generally speaking, we believe that under the condition of \( \text{FPS} \geq 30 \), the fluency of the video can meet the general requirements. For example, in the bridge damage model evaluation, 60FPS means that 60 images or 60 video images are detected per second.

3.2. Creation of the Dataset

In the field of bridge damage detection, only a few researchers have published the datasets required for the experiments, such as the concrete defect dataset established by Martin Mundt et al. [61] and the bridge damage dataset created by Hüthwohl et al. [62]; however, the format of the above datasets is asymmetrical and cannot be directly used for our model training and testing. At the same time, we found some datasets with low quality and incorrect labeling through the exploration of other public datasets on the Internet. For data-driven learning models, such as deep learning, the current public datasets cannot meet our requirements for the establishment of the BE-YOLOv5S bridge damage detection model. Therefore, after in-depth research and discussion about online resources, we decided to build a bridge damage detection dataset based on the dataset created by Hüthwohl et al., which was named Bridge defect detection Dataset-D (BDD). We downloaded the public dataset and selected three types of bridge damage—including cracks, rust staining, and efflorescence—as the detection targets of this experiment to conduct bridge defect detection research. First, we downloaded the original dataset and checked it carefully. To be more suitable for the training of the YOLO model, we standardized the images and filled all images with white background to adjust the pixel size to 640 \( \times \) 640. At the same time, we carried out a small amount of field shooting to further expand the generalization performance of its data. The camera equipment used is the rear 4000W main camera and 800 W telephoto lens and the TOF deep-sensing camera equipped on Huawei Mate30 Pro to ingest bridge damage data. The parameters collected are ISO = 50, F:1.6, and S:1/182s. By introducing a few images, we can further improve the generalization performance of the model. The BDD part of the dataset is shown in Figure 5. After the acquisition of the image, we used the labeling image annotation tool to label the image in detail. First, we generated the labeled data into a text format and flipped the image at different angles under the condition of the text format, as well as the random adoption and free combination of changes in brightness, etc. Based on BDD, the amount of data preprocessing was expanded by eight times to establish the BDD-E dataset. The data volume requirement is very important for the deep learning model. BDD-E can provide important model-driven data resources for scholars who conduct research on bridge damage in the same field.

It is more suitable for the detection needs of lightweight models similar to the YOLO algorithm for the detection of bridges with difficult detection conditions such as cross-sea and viaducts. However, for data-driven deep learning networks, data are an indispensable part of model building and model validation, and play an irreplaceable role in model building. Because we found a dataset that can be directly used by the YOLO algorithm, we decided to be able to disclose the dataset we processed and the dataset on which we performed data enhancement. To train the trial in the experiment more effectively, we divided the BDD dataset into independent, train, tenant st, and Val datasets in the ratio of 8:1:1 for model training and comparative analysis research. A total of 1049 pieces of data were formed, and the detailed information is shown in Table 1 below.
3.3. Model Building Details

The learning process of the deep learning network model is crucial to the performance of the model. This paper mainly aims to explore the contribution of the lightweight convolutional neural network BE-YOLOv5S to bridge damage detection. To reduce the training time, we selected the BDD dataset to perform model training, validation, and a comparative analysis of YOLOv3-Tiny, YOLOv5S, B-YOLOv5S, and BV-YOLOv5S under the same experimental conditions.

The relevant parameters and settings of the model establishment are as follows: epoch = 1000, the learning method is Cosine annealing, the initial learning rate is lr0 = 0.01, and the cyclic learning rate is lrf = 0.2. The above settings make the model training more efficient. Under the environmental configuration conditions of the laboratory, the training of the entire model took about 3 h. The four types of model training and validation losses in the experiment are shown in Figure 6. The loss function mainly represents the prediction gap between the predicted value and the real value. As the training loss of the model gradually converges, the model performance in the experiment gradually approaches the upper limit that the dataset can provide.

Through the analysis of the above training and validation loss images, we found that different experimental models have large fluctuations in the curve of the loss function within a peri at the beginning of training. After continuous iteration, the loss function curve gradually converged, which also proved that our initial parameter settings were appropriate. In Figure 6a, showing the train box loss function graph and Figure 6d, showing the value box loss function graph, we can find that the above four types of models have little difference in box learning. In Figure 6b, showing the train classes loss function diagram and Figure 6e showing the value classes loss function diagram, it can be found that the YOLOv3-Tiny model has a high verification loss function, the learning ability of classes is relatively poor compared with other models, and there is serious overfitting.
The training effect of YOLOv5S and B-YOLOv5S is good, but there is a slight overfitting phenomenon. The BE-YOLOv5S model has the best effect on the learning of classes, and the convergence loss value is the smallest. In Figure 6c,f, we can see that, except for the poor learning ability of the YOLOv3-Tiny model, the learning process of the rest of the models has little difference; the BE-YOLOv5S model has the smallest verification loss and a better training effect.

Figure 6. Plots of training and loss functions for four types of models during the experiment. (a) Diagram of Train Box Loss function; (b) Diagram of Train Classes Loss function; (c) Diagram of Train Object Loss function; (d) Diagram of Value Box Loss function; (e) Diagram of Value Classes Loss function; (f) Diagram of Value Object Loss function.

4. Results and Analysis

In Section 4.1 of this paper, we evaluate the model by using an independent BDD test set and analyze and discuss the evaluation results. The experimental environment used is the same as the training environment for the four types of models. At the same time, in the Section 2, we use part of the data in the BDD-E dataset and network public data to conduct independent tests and conduct different complex environment tests. We aim to comprehensively evaluate the performance of various experimental models on bridge damage detection. The results are discussed in Section 3.

4.1. Evaluation Metrics Results and Discussion

The evaluation of the model is very important for the verification of the model performance. To comprehensively and quantitatively analyze the performance of the bridge damage detection model established in this paper, we draw the relevant data display diagrams according to the performance of the four types of models in the BDD test set. In Figure 7, according to the basic indicators of the four types of confusion matrices, we draw the confusion matrices of the relevant experimental models in this paper tested in the BDD test set. As shown in Figure 7, the performance of the YOLOv3-Tiny model in the classification of three types of bridge damage is poor, far lower than the other three types of bridge damage detection models, YOLOv5S, B-YOLOv5S, BE-YOLOv5S. The three models have the same ability to classify efflorescence, YOLOv5S is more sensitive to the
features of rust staining, and the B-YOLOv5S model with enhanced feature extraction greatly reduces the false detection rate of rust staining in background detection. Through feature enhancement, deep-level information can be mined, which is beneficial for the model to distinguish the background; the BE-YOLOv5S model after sample imbalance optimization is more sensitive to small target cracks. To more intuitively discover the specific performance of each model in the bridge damage classification prediction, we draw the PR curves of the relevant experimental models in this article, as shown in Figure 8.

<table>
<thead>
<tr>
<th>Model</th>
<th>ROC AUC (Cracks)</th>
<th>ROC AUC (Efflorescence)</th>
<th>ROC AUC (Rust Staining)</th>
<th>ROC AUC (Background FN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOLOv3-Tiny</td>
<td>0.20</td>
<td>0.31</td>
<td>0.36</td>
<td>0.80</td>
</tr>
<tr>
<td>YOLOv5S</td>
<td>0.80</td>
<td>0.97</td>
<td>0.80</td>
<td>0.36</td>
</tr>
<tr>
<td>B-YOLOv5S</td>
<td>0.78</td>
<td>0.97</td>
<td>0.75</td>
<td>0.36</td>
</tr>
<tr>
<td>BE-YOLOv5S</td>
<td>0.82</td>
<td>0.97</td>
<td>0.73</td>
<td>0.40</td>
</tr>
</tbody>
</table>

**Figure 7.** The confusion matrix plot of the relevant experimental models in this paper was tested on the BDD test set. (a) YOLOv3-Tiny; (b) YOLOv5S; (c) B-YOLOv5S; (d) BE-YOLOv5S.

As shown in Figure 8a, the PR curve of BE-YOLOv5S almost wraps the other three types of curves, which directly proves that the performance of BE-YOLOv5S in bridge damage detection is ahead of the other three types of models. For the PR curve analysis of cracks, efflorescence, and rust staining, it is intuitively found that the performance of the YOLOv3-Tiny lightweight detection network model in bridge damage detection is poor. By drawing the balance point F1 (the value when P = R, that is, when the slope is 1), we can...
find that the BE-YOLOv5S bridge damage detection model proposed in this paper is ahead of other detection models in all kinds of damage detection performance in the experiment.

Figure 8. The PR curve plot of the relevant experimental models in this paper was tested in the BDD test set. (a) PR curve of all classes, (b) PR curve of the cracks, (c) PR curve of the efflorescence, and (d) PR curve of the rust staining.

To evaluate the work in this paper in more detail, we tested the detection speed and mAP@.5, Precision, Recall, and F1-Score indicators of various models through experiments, as shown in Table 2. Given the special environment of bridge structures, there is a high demand for the detection speed of the model (for example, using UAVs to carry embedded devices for detection); a slow detection speed not only results in the loss of time and cost but also reduces the detection coverage, resulting in seriously missed detection. The detection speed of the four types of damage detection models for bridge damage is shown in Table 2. The YOLOv3-Tiny model has the fastest detection speed, and the BE-YOLOV5 and YOLOv5S models have similar detection speeds. The B-YOLOv5S model is improved by the BiFPN network structure. However, the BE-YOLOv5S network model after the sample imbalance treatment has significantly improved the detection speed of bridge damage. It can be seen that through the optimization of EFL sample imbalance, the established model can be more suitable for the feature learning of bridge damage, thereby improving the detection speed of the model. The same importance as the detection speed is the reliable performance
of the model. In Table 2, we analyze the performance of the four types of damage models in mAP@.5, Precision, Recall, and F1-Score evaluation metrics. The purpose is to evaluate the model proposed in this paper in a comprehensive and detailed manner. In this paper, mAP@.5 not only shows the accuracy of model classification but also evaluates its target localization ability, which is an important evaluation index in this work. The mAP@.5 of YOLOv3-Tiny is only 0.375, the mAP@.5 of YOLOv5S is 0.807, and the mAP@.5 of B-YOLOv5SY improved by enhanced feature extraction is 0.811, an increase of 0.4% compared with the unimproved. The enhanced feature extraction of the optimized sample imbalance processing function BE-YOLOv5S network model mAP@.5 is 0.827, which is 45.1% higher than YOLOv3-Tiny, 2% higher than YOLOv5S, and 1.6% higher than B-YOLOv5S. It can be seen that using the BiFPN network structure to enhance the feature extraction is more suitable for the features of the bridge damage target. With the optimization of the EFL sample unbalance function, the bridge damage detection BE-YOLOv5S model proposed in this paper achieves a significant improvement in mAP@.5 evaluation metrics. In the Precision evaluation metrics, YOLOv3-Tiny, YOLOv5S, B-YOLOv5S, and BE-YOLOv5S obtained scores of 0.293, 0.867, 0.841, 0.893, respectively. The BE-YOLOv5S proposed in this paper has the highest score. Compared with the first three models, the improvement was 57.4%, 2.6%, and 5.2% respectively. In the Recall evaluation indicators, YOLOv3-Tiny, YOLOv5S, B-YOLOv5S, and BE-YOLOv5S obtained scores of 0.701, 0.817, 0.803, and 0.821, respectively. The BE-YOLOv5S proposed in this paper has the highest score; compared with the previous three models, the improvement was 12%, 0.4%, and 1.8% respectively. In the F1 Score evaluation indicators, YOLOv3-Tiny, YOLOv5S, B-YOLOv5S, BE-YOLOv5S obtained the scores of 0.413, 0.841, 0.822, 0.855, respectively. The BE-YOLOv5S proposed in this paper has the highest score; compared with the previous three models, it obtained 44.2%, 1.4%, and 3.3% improvement. Through the evaluation of the above evaluation indicators, the BE-YOLOv5S bridge damage detection model proposed in this paper has shown good reliability and advancement. We strengthen the feature extraction capability of the bridge damage detection model BE-YOLOv5S in bridge damage targets by introducing the BiFPN network structure—it significantly improves the detection of small and indistinct features such as cracks, and improves the engineering applicability of the bridge damage detection model. At the same time, to establish a model with better performance, we improved Focal Loss and used EFL to improve the sample imbalance processing mechanism for the narrow and complex characteristics of bridge damage images. The experimental results show that, by improving the handling method of sample imbalance, the BE-YOLOv5S proposed in this paper can efficiently acquire image information in the process of training and learning. The optimization of sample imbalance processing plays an important role in the establishment of a practical bridge damage detection model. Through the comparative analysis of evaluation indicators, our established BE-YOLOv5S network model has advanced detection capabilities in the detection of bridge damage. Compared with the representative YOLOv3-Tiny and YOLOv5S networks in the field of lightweight target detection, they have achieved significant improvements in various evaluation indicators.

### Table 2. Evaluation Results of Four Types of Models.

<table>
<thead>
<tr>
<th>Model</th>
<th>mAP@.5</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
<th>FPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOLOv3-tiny</td>
<td>0.376</td>
<td>0.293</td>
<td>0.701</td>
<td>0.413</td>
<td>244</td>
</tr>
<tr>
<td>YOLOv5S</td>
<td>0.807</td>
<td>0.867</td>
<td>0.817</td>
<td>0.841</td>
<td>189</td>
</tr>
<tr>
<td>B-YOLOv5S</td>
<td>0.811</td>
<td>0.841</td>
<td>0.803</td>
<td>0.822</td>
<td>156</td>
</tr>
<tr>
<td>BV-YOLOv5S</td>
<td>0.827</td>
<td>0.893</td>
<td>0.821</td>
<td>0.855</td>
<td>185</td>
</tr>
</tbody>
</table>

### 4.2. Result and Discussion of Testing under Complex Conditions

In fact, the influence of EOV in actual engineering has a certain impact on the light and viewing angle of the inspection perspective. For the deep learning network model, the same object is different from different shooting angles, locations and lighting conditions. Therefore, in practical engineering applications, EOV is an important issue that needs
to be considered. We compare the detection effects of the models on the BDD-E dataset, aiming to comprehensively evaluate the reliability and advancement of our proposed BE-YOLOv5S model for bridge damage detection. We found that the BE-YOLOv5S bridge damage detection model has excellent detection performance and is more reliable in the detection of enhanced illumination and multiple small targets. As shown in Figure 9, the test images of the four types of models in the enhanced care and multi-small target environments are shown. Only BE-YOLOv5S can be successfully detected in the detection of enhanced illumination cracks and efflorescence, which indicates that the traditional Focal Loss using the same modulation factor balance training cannot meet the training requirements of the bridge damage detection model when faced with strong illumination effects. In this paper, FEL is used to optimize the imbalance problem of the training samples of bridge damage, and the learning and training using the modulation factor associated with the damage category achieve significant performance improvement. As shown in Figure 9c,d YOLOv5S and B-YOLOv5S are not sensitive to the small object detection in the upper right corner. However, the rust staining target detection confidence of B-YOLOv5S with enhanced feature extraction is significantly higher than that of YOLOv5S without enhanced feature extraction. It can be seen that after BiFPN feature extraction, deeper target information can be excavated. Therefore, the bridge damage target is detected more effectively, but the sensitivity is not high in the detection of small targets. As shown in Figure 9e, the optimized BE-YOLOv5S bridge damage detection model through enhanced feature extraction and sample imbalance processing will detect all damage targets in the image to be detected, the sensitivity to small targets is high, and it has a more reliable performance for practical engineering applications.

Figure 9. Comparison chart of detection performance of four types of bridge damage models under complex conditions. (a) Undetected original image; (b) Performance rendering of YOLO-V3-Tiny model detection; (c) rendering of YOLOv5S model detection performance; (d) rendering of B-YOLOv5S model detection performance; (e) BE-YOLOv5S model detection performance renderings (the images used are all bridge damage images from the BDD-E dataset).
4.3. Discussion

Through the analysis and comparison of the above results, we found that the YOLOv3-Tiny network model, which is representative of the lightweight convolutional neural network, performed poorly in the above evaluation and test results. YOLOv3-Tiny has excellent detection speed, but its network structure has fewer layers than YOLOv5S, B-YOLOv5S and B-V-YOLOv5S networks, so the detection speed is fast, but YOLOv3-Tiny has fewer layers of convolutional neural network. In the face of deep and complex feature information such as bridge defects, the ability to extract is obviously insufficient, resulting in poor performance of the model in bridge defect detection, and it has difficulty meeting the reliability requirements of bridge damage detection engineering. The original YOLOv5S network model has a low score in the evaluation indicators in detection, especially the key evaluation indicator mAP@.5, and also exposes that PANet has a weak feature extraction ability for bridge damage target images. The B-YOLOv5S bridge damage detection model after enhanced feature extraction has a significant improvement in the mAP@.5 index. The introduction of the BiFPN network structure has a slight adverse impact on the accuracy of bridge damage target detection and other indicators. However, overall performance still improved. The BE-YOLOv5S bridge damage detection model proposed in this paper obtains significant advantages in the evaluation index and shows good robustness in detection under complex conditions. Comparing the BE-YOLOv5S network model with the B-YOLOv5S network model, we can find that using EFL to deal with the sample imbalance problem is more suitable than the traditional Focal Loss; it still has good robustness for the detection of multiple small targets and complex conditions with strong lighting. In summary, the BE-YOLOv5S network model proposed in this paper has certain advanced and reliable properties in the field of bridge damage detection.

5. Conclusions

In the work in this paper, we propose a lightweight convolutional neural network model BE-YOLOv5S with enhanced feature extraction and an improved sample imbalance processing mechanism, we demonstrate the advances and reliability of our work in bridge damage detection by experimental analysis. First, we carried out the re-selection and labeling of the dataset established by the bridge damage detection model based on the public dataset, which reduced the error rate of mislabeling. At the same time, we obtained the data used in the experiments of this paper using the manual collection and web crawler and established the BDD dataset. To further test the generalization ability and reliability performance of the model in this experiment, we enhanced the dataset based on the BDD dataset. To a certain extent, the size and position of the original bridge damage targets are changed, and the brightness and clarity are processed to simulate the complex environmental conditions in actual engineering. We noticed that there are very few public datasets applicable to the YOLO algorithm. To better contribute our work to the field of bridge damage detection, we decided to disclose the BDD dataset and the enhanced BDD-E dataset we used. Through the disclosure of data, we aim to contribute to our work and facilitate scientific research for more researchers who are engaged in the research of the YOLO algorithm of lightweight convolutional neural networks. After the dataset is established, we introduce the representative YOLOv3-Tiny model in the field of lightweight target detection, the advanced YOLOv5S, and the B-YOLOv5S network model with enhanced feature extraction for experimental comparison and analysis. Under the same experimental environment, the BE-YOLOv5S model proposed in this paper has achieved significant advantages in various indicators and testing under complex conditions, especially in the field of target detection, which is an important indicator. In mAP@.5, the damage detection model on the BE-YOLOv5S bridge leads the YOLOv3-Tiny, YOLOv5S and B-YOLOv5S network models by 45.1%, 2%, and 1.6%, respectively, with a certain degree of advancement. Similarly, in the important index of F1-Score, the damage detection model on the BE-YOLOv5S bridge is ahead of the YOLOv3-Tiny, YOLOv5S, and B-YOLOv5S network models by 44.2%, 1.4%, 3.3%, respectively, and is more reliable in bridge damage detection.
BE-YOLOv5S shows significant advantages in detection under complex conditions, and the BE-YOLOv5S bridge damage detection model achieves significant advantages in the detection of illumination, sharpness, and multiple small targets. Through the experiments and results analysis in this paper, we find that the BE-YOLOv5S network model proposed in this paper has good robustness and advanced reliability in the detection of bridge damage targets.

The bridge damage detection model BE-YOLOv5S proposed in this paper has a practical engineering application type, good flexibility, and reliability, and can be suitable for the detection of embedded devices in complex environments where bridges are located. Compared with the current mainstream lightweight detection models YOLOv3-Tiny and YOLOv5S, the BE-YOLOv5S proposed in this paper is more advanced and applicable. However, for the deep learning model of bridge damage detection, the large-scale and high-quality data is the basis for the strong generalization ability of the model [63]. Therefore, in the next step, we will further improve the quality and scale of the data and improve the BE-YOLOv5S model, and improve the work of the lightweight convolutional neural network in bridge damage detection.

Author Contributions: Conceptualization, F.D. and S.J.; methodology, software, validation, formal analysis, investigation, F.D.; resources, F.D.; data curation, writing—original draft preparation, F.D. and K.C.; writing—review and editing F.D. and K.C.; visualization, F.D.; supervision, S.J.; project administration, F.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The BDD and BDD-E datasets used in this experiment will be publicly available, please contact the first author.

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

References
24. Montaggioli, G.; Puliti, M.; Sabato, A. Automated Damage Detection of Bridge’s Sub-Surface Defects from Infrared Images Using Machine Learning; SPIE: Bellingham, WA, USA, 2021; p. 11593. [CrossRef]
41. Zhu, J.; Song, J. An Intelligent Classification Model for Surface Defects on Cement Concrete Bridges. Appl. Sci. 2020, 10, 972. [CrossRef]
55. Du, F.-J.; Jiao, S.-J. Improvement of Lightweight Convolutional Neural Network Model Based on YOLO Algorithm and Its Research in Pavement Defect Detection. *Sensors* 2022, 22, 3537. [CrossRef]
58. Sha, G.; Wu, J.; Yu, B. The Shrank YoloV3-tiny for spinal fracture lesions detection. *IFS* 2022, 42, 2809–2828. [CrossRef]