Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 6932040, 16 pages
https://doi.org/10.1155/2022/6932040

Research Article

Deep Learning-Enabled Automatic Detection of Bridges for Promoting Transportation Surveillance under Different Imaging Conditions

Peng Han [1] **and Xiaoxia Yang** [2]

[1]*The Administrative Center for China's Agenda 21 (ACCA 21), Beijing 100038, China*
[2]*College of Earth Science, Chengdu University of Technology, Chengdu 610059, China*

Correspondence should be addressed to Xiaoxia Yang; 35923690@qq.com

Received 5 May 2022; Revised 29 July 2022; Accepted 2 August 2022; Published 7 September 2022

Academic Editor: Yi-Sheng Lv

The reliable and accurate detection of bridges plays an important role in imaging-driven transportation surveillance. It is capable of timely providing the traffic information, leading to safer and more convenient transportation. However, the visual quality of observed images is often inevitably reduced owing to the adverse weather conditions, e.g., haze and low lightness. It is still difficult to adopt the existing powerful deep learning methods to reliably and accurately detect the bridges under different imaging conditions. To achieve satisfactory bridge detection results, we first propose to exploit the data augmentation strategy and physical imaging method to generate the natural-looking experimental dataset, which contains latent high-quality images and their hazy and low-light versions. We then investigate how to further promote the deep learning-based bridge detection methods through the manually generated dataset. It is obvious that the generalization abilities of these deep neural networks are significantly improved using this data augmentation strategy. In this work, we constructed an original dataset consisting of 3500 images of size 900×600, collected under normal imaging condition. Extensive detection experiments will be performed based on the augmented dataset. Experimental results have demonstrated that our automatic bridge detection framework could generate more reliable and accurate results compared with existing detection methods.

1. Introduction

With the rapid developments of unmanned aerial vehicles (UAV) and remote sensing techniques [1, 2], it has become flexible to capture the visual data of bridges in a wide range [3]. The synthetic aperture radar (SAR) images [3] are able to expand the imaging range containing the bridges. However, the SAR images inevitably suffer from the random noise and low-contrast signal. In addition, it is not flexible to timely and accurately capture the visual images related to the traffic flow on bridges. In addition, the visual data collected from UAV-enabled imaging sensors are able to generate meaningful and important information in transportation surveillance. It is capable of timely providing the traffic information, leading to safer and more efficient transportation. Due to the poor weather conditions (e.g., haze [4] and low-light [5]), the visual quality of observed images is often inevitably impaired in practice. Under poor imaging conditions, it will be challenging to detect the bridges precisely. In recent years, considerable effort has been invested in the detection of bridges under normal imaging conditions. Traditional detection methods, deep learning–based methods, and combined detection methods comprise the three main categories of general detection techniques. Deep learning–based methods have significantly improved their detection precision and efficiency due to their powerful learning and representation capabilities. However, the degraded images provide few details in terms of geometrical structures and color appearance, making it difficult to detect bridges reliably and precisely under adverse imaging conditions.

The deep learning–based detection methods are highly dependent on the training dataset. To achieve satisfactory bridge detection results, we will construct an original dataset with 3500 images of size 900×600, collected under normal imaging conditions. To promote the volume and diversity of the training dataset, we will exploit the data augmentation method based on physical imaging modeling [6–8] to enlarge the original image dataset. In particular, the enlarged dataset will contain the latent sharp images and their hazy and low-light versions. Furthermore, four typical deep learning–based detection methods (i.e., Faster R-CNN [9], YOLOv3 [10], YOLOv4 [11], and YOLOv5 [12]) will be introduced to implement the detection of bridges of interest. In this work, we will investigate how to further promote the deep learning–based bridge detection methods through the enlarged dataset. Existing studies have demonstrated that the generalization abilities of these deep neural networks could be significantly promoted according to the data augmentation strategy. The proposed deep learning-enabled automatic computational framework could significantly improve the reliability and accuracy of bridge detection. Therefore, the main contributions of this work can be summarized as follows:

(1) We proposed a deep learning-enabled automatic computational framework for detection of bridges in transportation surveillance.

(2) We construct an original image dataset consisting of 3500 images of size 900×600 captured under standard imaging conditions. Owing to the physical imaging methods, the synthetically degraded images are generated to expand the image dataset and enhance the generalization capabilities of deep learning methods. The expanded dataset is used to train deep learning–based detection methods, resulting in more reliable and accurate detection of bridges under various imaging conditions.

(3) Experiments have shown that our automatic framework can produce more robust and accurate detection results than existing methods. The satisfactory detection performance is primarily attributable to the robust learning ability of neural networks and the physical imaging-based data augmentation strategy.

The remainder of this work is structured as follows. Section 2 briefly reviews the recent work related to bridge detection. Four typical deep learning-based bridge detection methods and data augmentation strategies are introduced in Section 3. Numerous experiments are implemented to illustrate the effectiveness of our method in Section 4. This work is finally ended by giving the main contributions in Section 5.

2. Related Work

Recent efforts in the literature have focused on the automatic detection of bridges of interest. The current detection methods can be roughly classified into three categories: traditional, deep learning–based, and combined versions. In this section, we will review these detection methods briefly.

2.1. Traditional Bridge Detection Methods. As early as 1989, Baker et al. [13, 14] studied the detection of concrete bridges in close-up side-shot images. It is assumed that the sides of the bridge had an approximate rectangular geometry. Therefore, there would be a pair of parallel lines on the upper and lower sides of the bridge. The rectangular side could be obtained by detecting the parallel lines. According to this assumption, Wang et al. [15] proposed a multistep bridge detection framework. It first extracted the river regions from the image and then selected the regions between two disparate water areas as candidate selections. The final detection results could be obtained by extracting parallel lines from these regions. To further enhance detection performance, more structural correlations between bridges and rivers have been exploited [16]. By considering both local radiometric and textural features, Loménie et al. [17] proposed a computational system for robustly detecting bridges from high-resolution satellite images. According to the image prior knowledge, an integrated method was developed [18] to automatically detect the bridges over rivers from satellite-borne visual images. In particular, this method first extracts the river areas and then detects the brides of interest through the knowledge-driven strategy. To take full advantage of the spatial relation between bridge and water, Liang et al. [19] developed an automatic computational method to detect bridges using the HJ-1 satellite remote sensing images. The multispectral and morphological information in images were combined to guarantee bridge detection.

The above computational methods mainly focus on the extraction and processing of line features rather than the position relationship between the bridge and river. However, methods only based on the parallel line features of the target will easily generate unsatisfactory detection results. To overcome this limitation, Sita [20] first proposed a method for template matching, moment feature matching, and transformation feature matching according to the geometric or regional features, moment features, and transformation features of the target. On this basis, Wang et al. [21] proposed a depict segmentation method for recognizing bridges, which used Hough transform to extract the longest straight lines from the coarse-resolution images. Though these methods perform well in detection efficiency and accuracy, the parameter selection of the Hough transform has a great influence on the result. It is not suitable for the situation where there are multiple similar targets in the image.

To effectively solve the above problems, Han et al. [18] first used the grey level cooccurrence matrix (GLCM) method to obtain the entropy, contrast, homogeneity, energy, and other characteristics from the remote sensing image. Then, they used these features to divide the image into river and land areas and finally detected bridges by using prior knowledge. As a consequence, Fu et al. [22] first established knowledge models of the bridges of interest.

They utilized concurrently the segmentation of waters, the extraction of regions of interest (ROI), the connectivity signal, and the detection of candidate regions for the approximate positioning of bridges. During the testing procedure, the grey characteristics were used to identify the potential bridge. To take full advantage of spatial information, both structural information and topological relations in bridges regions were considered to enhance bridge detection performance [23]. In addition, Liu et al. [24] proposed a bridge information extraction method based on multisource data fusion, which is mainly based on the representation of bridges, land, and rivers, by fusing panchromatic and near-infrared images to recognize and extract bridge information automatically.

The premise of these methods is that the road information in the image is available. For the positional relationship between roads, rivers, and bridges to be unknown, Shu-Kui and Nie [25] proposed a method for water bridge recognition which is object oriented, which used the region growth method to segment the image. According to the special reflection characteristics of the water body, the image objects generated after segmentation were used as the basic units to be classified. Then, based on the characteristics of the bridge, they extracted the bridges from images by using the shape feature of image objects and the contextual relationship between bridges and water bodies. However, due to the limitation of spatial resolution, the detection of bridges on some small tributaries was not perfect. Subsequently, through the research on the morphological filter and regional growing, Gu et al. [26] proposed a method of bridge extracting based on the difference between the above filtering methods. By combining the characteristics of bridges, the bridge information can be extracted from the difference in filtered DEM. The proposed method was capable of detecting different bridge designs. Based on the high spatial images, Yuan et al. [27] developed a target-oriented computational method to automatically extract bridges from original images. Firstly, they selected the optimal segmentation scale through multiscale segmentation experiments and the underlying surface features; secondly, they established a rule set by using water body index, threshold function, and other methods. They gradually obtained vectors of water bodies and potential areas of bridges. Finally, they successfully extracted the bridge through binarization, mathematical morphology processing, overlay analysis, and other methods.

2.2. Deep Learning-Based Bridge Detection Methods. The above-mentioned bridge detection methods are highly depended on traditional image processing techniques. The corresponding detection results often suffer from several limitations, such as unstable detection performance under complex imaging conditions. Besides, it becomes difficult to generate high accuracy and essentially fails to generate real-time detection. As the rapid development of artificial intelligence technique [28, 29], the deep learning–based detection methods have gained great success in the fields of target detection and recognition, due to its powerful learning capacity.

In the literature, the existing deep learning-based target detection methods can be broadly divided into two classes, i.e., the two-stage and one-stage methods. The representative two-stage detection methods mainly include the R-CNN [30] and its extensions, such as Fast R-CNN [31], Faster R-CNN [9], and mask R-CNN [32]. These methods first extract the candidate bounding boxes according to the position of the target of interest. The classification and regression are then implemented to produce the detection results. These methods are capable of accurately detecting the bridges from the original images. However, they inevitably suffer from high computational load, leading to unreal-time detection of bridges in practice. Fortunately, the introduction of one-stage detection method could deal with this limitation. The one-stage detection method mainly include the single shot multibox detector (SSD) [33], as well as you only look once (YOLO) [34] and its extensions. The popular YOLO [34] directly estimated the category probability, bounding box, and class confidence. It is able to significantly improve the detection efficiency. YOLOv2 [35] and YOLOv3 [10] have enhanced the capacities of feature extraction by introducing the Darknet-19 and Darknet-53, respectively. The target detection results were improved accordingly. More recently, the newly developed YOLOv4 [11] and YOLOv5 [12] have attracted considerable attention and could significantly improve the detection accuracy, efficiency, and robustness.

With the rapid developments of airborne and satellite-bone imaging systems, it becomes more flexible to detect the bridges to monitor traffic status. For example, based on the satellite-bone SAR images, Chen et al. [36] proposed an advanced bridge detection method, termed as single shot detection-adaptive effective feature fusion (i.e., SSD-AEFF), to perform accurate bridge detection from complex SAR images. The multiresolution attention and balance network (i.e., MABN) [3] were also proposed to identify bridges from SAR images. The proposed network is mainly composed of three parts, i.e., the attention and balanced feature pyramid module, the region proposal network (RPN), the regression, and classification. However, the SAR images often suffer from the random noise and low-contrast signal. It is thus difficult to detect small-scale bridges. In contrast, the optical aerial images become more popular and practical for bridge detection. By taking into consideration the auxiliary water body extraction task, an accurate bridge detection network [37] has been presented, which could enhance the feature representation through semantic context. However, the quality of optimal images is often degraded due to the poor weathers, such as hazy and low lightness, leading to unsatisfactory detection results. In this work, we will incorporate the data augmentation strategy into existing deep learning method to make bridge detection more robust and accurate under complex imaging conditions.

2.3. Combined Bridge Detection Methods. It is sometimes difficult for a single detection method to generate satisfactory detection results. To enhance the detection results, Mirmehdi et al. [38] proposed to introduce the specific

control strategies to enhance the traditional methods. Based on the joint features and knowledge rules, Sang et al. [39] proposed a new method for quickly detecting bridges on water from visible light remote sensing images. The whole detection framework could be mainly divided into four steps, i.e., water area segmentation, preliminary bridge edge detection, matching, and verification to remove false bridges. It is well known that traditional methods and deep learning methods have their own advantages and disadvantages. There is a potential to combine the advantages to further enhance bridge detection. Therefore, a hybrid detection framework, proposed by Loménie et al. [17], has been proposed to implement robust detection. This hybrid version exploited a voting strategy to classify the image pixels into the corresponding classes and introduced the bottom-up and top-down parts to accurately detect bridges. Tria-Sanz and Loménie [40] first introduced the radiation features to classify the pixels into different geographic types and then searched for bridges based on the above classification. By introducing the prior-information auxiliary module (PAM) [41], which is able to receive and integrate prior-information feature maps, four popular target detection networks, i.e., YOLOv3 [10], YOLOv4 [11], Faster R-CNN [9], and CenterNet [42], have been extended to obtain more meaningful features. By integrating the prior-information feature maps into the existing neural networks, the domain-specific knowledge will be more meaningful, beneficial for promoted bridge detection.

3. Deep Learning-Enabled Automatic Detection of Bridges

This paper mainly focuses on the automatic detection of bridges under hazy and low-light situations. To improve the generalization ability of our detection method, we first synthesize both hazy and low-light images to expand the original dataset. We then train the object detection methods and compare the computational accuracy and efficiency under the same imaging conditions. In particular, several typical networks, e.g., Faster R-CNN, YOLOv3, YOLOv4, and YOLOv5, will be exploited to detect the bridges of interest in this work. Experiments show that our automatic framework can produce more robust and accurate bridge detection results than existing methods.

3.1. Faster R-CNN. Faster R-CNN has recently become a typical two-stage target detection method. In particular, it first estimates the target candidate regions. These candidate regions are then classified and regressed, beneficial for accurate and robust target detection. This is due to the fact that Faster R-CNN exploits the RPN to generate candidate regions on the feature map. Compared with the selective search (SS) algorithm [43], the RPN reduces the number of candidate regions for each image to 300, which significantly reduces the calculations and greatly improves the efficiency during the network training and test. To make it easier to understand, the architecture of Faster R-CNN is visually shown in Figure 1. The widely used Faster R-CNN is mainly composed of four components, i.e., backbone network, feature pyramid network (FPN), RPN, and feature branches. The backbone network is capable of extracting the feature information from the original input image.

The RPN is a fully convolutional network, which greatly reduces the calculations because it does not contain a fully connected layer. The input of RPN is a feature map of any size. In addition, its output is a batch of candidate frame position information and the probability of whether there is a target. Because the size of feature map of the fully connected layer is fixed, the ROI pooling layer is required to perform feature transformation on the candidate region. This part is achieved by the local max pooling.

During the network training, the input image is firstly scaled to prevent distortion. The shared feature map is then extracted through the backbone network for subsequent RPN and ROI pooling. The RPN is composed of two layers, i.e., the convolutional layer and the activation function layer. In particular, it exploits the anchor mechanism to generate anchor boxes for three different scales and three aspect ratios. Let the size of the feature map of the input RPN be $w \times h$, and a total of $w \times h \times 9$ anchor boxes are generated accordingly. After screening the candidate frames, the feature extraction is performed through a convolution kernel of size 3×3. Two 1×1 convolutions are then exploited for target classification and frame regression. It becomes flexible to obtain the position information of the candidate frame and the probability of containing the targets.

To obtain the candidate area, the candidate frame obtained through the RPN is mapped to the last layer of the feature map. The feature size transformation is then performed through the ROI pooling layers. The object classification and bounding-box regression are finally performed in the detection network. During the training process, the multitask loss is utilized to jointly train the regression and classification networks in practical applications.

3.2. YOLOv3. The popular one-stage target detection network YOLOv3 essentially considers the target detection task as a regression problem. Compared with the Faster R-CNN, it avoids the generation of candidate regions. It directly performs classification and regression on the feature map, which greatly improves the detection efficiency and reduces the computational cost compared with traditional two-stage detection networks.

As shown in Figure 2, YOLOv3 can predict targets of interest at three different scales. By dividing the image into grid cells of the same size (i.e., 13×13, 26×26 or 52×52), each grid point is responsible for the prediction of a region. The final prediction result is represented using a 3-d tensor, which contains the position information of the predicted bounding box, the confidence score, and the class predictions. In particular, YOLOv3 utilizes the Darknet53 as a feature extractor, which applies the convolution kernels of size 3×3 and 1×1, alternately. It also employs the skip connections to enhance the ability of the network to extract meaningful features. In the literature [10], YOLOv3 can extract feature maps with three different scales, which are

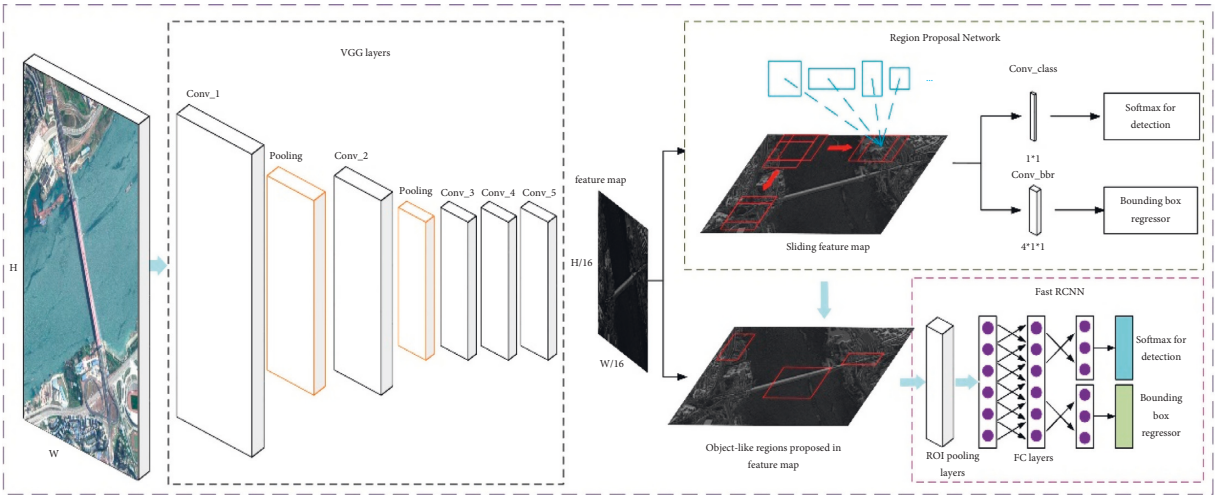


Figure 1: The architecture of Faster R-CNN.

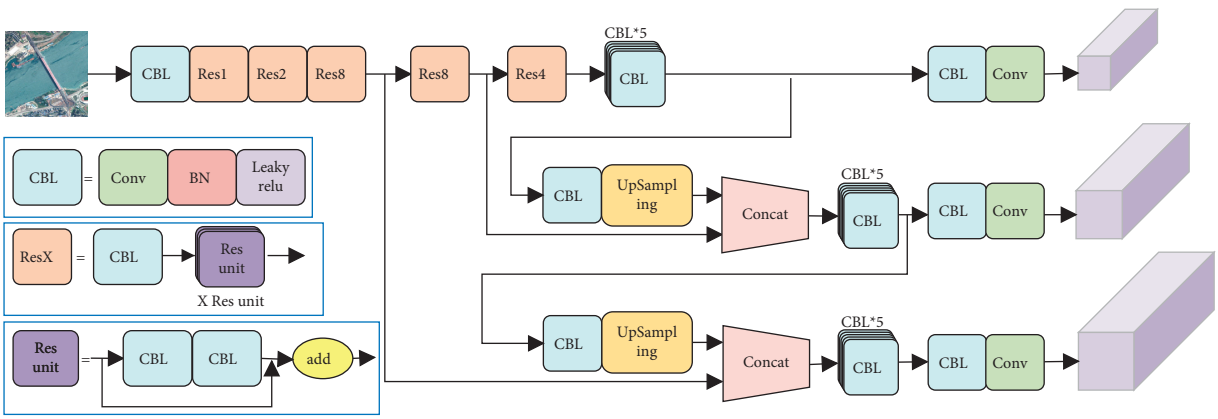


Figure 2: The architecture of YOLOv3 network.

$13 \times 13 \times 1024$, $26 \times 26 \times 512$, and $52 \times 52 \times 256$, respectively. By strengthening the capacity of feature extraction, the prediction heads could exploit the feature maps by FPN to robustly and accurately predict the bounding-box coordinates and class probabilities.

The YOLOv3 network firstly resizes the input image to a fixed size and then divides the image into grids of three scales. The k-means clustering is introduced to obtain a priori boxes of each different scale. The important feature extraction strategy is performed on the image through the backbone network. Through an existing multiscale prediction method, the feature maps with three different sizes extracted from the model are utilized to predict the final results. The feature fusion is finally performed on the prediction results of the three scales. The network filters out the bounding boxes whose confidence is lower than the predefined threshold. The final bounding boxes can be obtained through the widely used nonmaximum suppression (NMS).

3.3. YOLOv4. YOLOv4 is an extension of traditional YOLOv3, which is able to generate more robust and accurate detection results. In particular, YOLOv4 exploits the mosaic data augmentation method to enrich the existing dataset. This method randomly scales four training images and then randomly stitches them into a new image. This generation process of image dataset could make the detection network more robust and general under complex imaging conditions. As shown in Figure 3, YOLOv4 directly exploits the CSPDarknet53 as the backbone network to extract meaningful image features. Unlike the Darknet-53 by YOLOv3, YOLOv4 further modifies the residual block structure on Darknet-53. It also tends to replace the Leaky ReLU with the Mish activation function in the backbone network. Different from the FPN utilized as the parameter aggregation method in YOLOv3, the more powerful path aggregation network (PANet) is introduced in YOLOv4.

The neck is exploited to fuse image features and transfer these features to the prediction layer. In the neck part, the spatial pyramid pooling (SPP) [44] is introduced to increase the receptive field of the network. In addition, YOLOv4 employs the PANet to fuse and combine image features to promote the capacity of target detection with different scales, which is obviously different from the FPN utilized in YOLOv3. The head of YOLOv4 network is responsible for robustly and accurately predicting the bounding boxes and object classification.

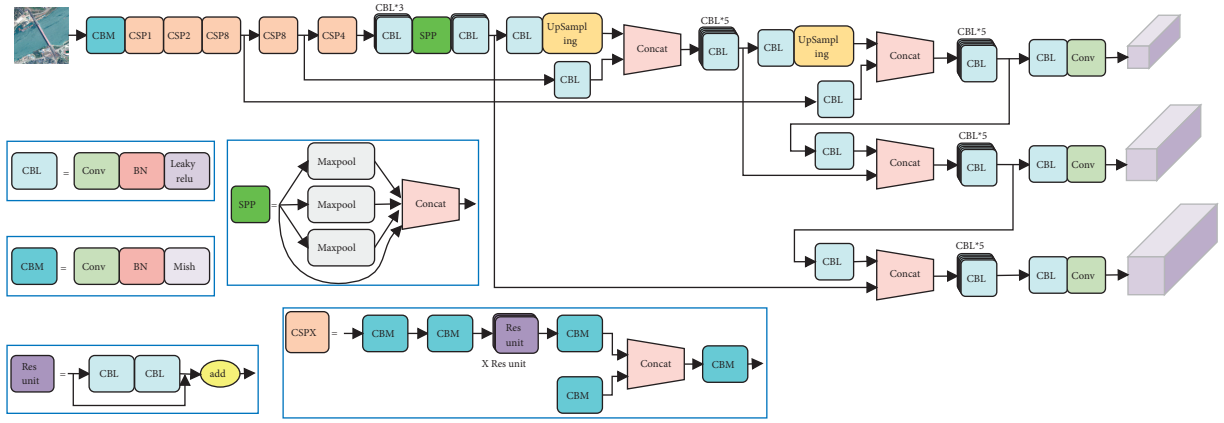


Figure 3: The architecture of YOLOv4 network.

In addition, YOLOv4 commonly exploits the complete intersection over union (CIoU) as the loss function, which considers not only the overlap area of the bounding box, but also the distance from the center point and the aspect ratio. As a consequence, the regression speed and accuracy, related to the bounding-box prediction, are improved accordingly.

3.4. YOLOv5. YOLOv5 also exploits the mosaic data augmentation method to enrich the image dataset. However, unlike the YOLOv4, it employs the mechanism of auto learning bounding-box anchors to learn the anchor boxes based on the enriched training data. The architecture of YOLOv5 network is visually illustrated in Figure 4. It can be observed that the backbone of YOLOv5 directly combines the focus structure and CSP structure. The focus layer slices the input image and warps $H \times W \times 3$ into $H/2 \times W/2 \times 12$. This method allows more detailed information to be retained during the downsampling process, which can prevent the information loss in practical applications.

The neck part contains FPN and PAN structures. In particular, the FPN transfers and integrates the high-level feature information through the upsampling from top to bottom. It is able to convey the robust semantic features accordingly. In addition, the PAN is a bottom-up feature pyramid which conveys the robust positioning features. Both FPN and PAN are simultaneously utilized to strengthen the network feature fusion capabilities. In the literature [12], YOLOv5 employs the generalized intersection over union (GIoU) as the loss function to constrain the prediction of bounding boxes. This strategy could effectively improve the performance for target classification and bounding-box regression. The nonmaximum suppression (NMS) is then exploited to eliminate the redundant bounding boxes and generate the optimal detection results.

3.5. Data Augmentation Method. To evaluate the detection performance of our learning method, it is necessary to enlarge the original image dataset under different imaging conditions. Both hazy and low-light imaging conditions will be considered to generate the degraded images.

3.5.1. Synthetic Generation of Hazy Images. According to the popular atmospheric scattering theory [6], the scattering of light by particulate impurities suspended in the air is the main reason for the degradation of hazy images. The incident light attenuation model refers to the attenuation of the reflected light on the surface of the object due to the scattering effect, which reduces the intensity of the light reaching the imaging system. Theoretically, as the propagation distance increases, the intensity of reflected light on the target surface will decay exponentially.

It is well known that the atmospheric scattering model has been widely exploited to describe the formation process of haze images, which can be expressed as follows:

$$I(x, y) = J(x, y)e^{-\beta\, d(x,y)} + \left[1 - e^{-\beta\, d(x,y)}\right]A. \tag{1}$$

(x, y) denotes the pixel location, I is the observed hazy image, J denotes the latent sharp image, $d(x, y)$ represents the depth at the pixel (x, y), β is a scattering coefficient which denotes the medium density, and A is the global airlight.

By manually setting different scattering coefficients and global airlights, we can synthesize different kinds of hazy images, contributing to an enlarged dataset which contains both original and hazy images. In our numerical experiments, we propose to randomly select different values of A and β, i.e., $A \in [0.7, 1.0]$ and $\beta \in [0.4, 0.8]$.

3.5.2. Synthetic Generation of Low-Light Images. The Retinex theory has been widely used to synthetically generate the low-light images. In particular, the Retinex theory could be exploited to decompose an image into two independent components, i.e., reflectance and illumination components. It is well known that the color of a target observed by the human eyes is only related to the reflection component on the target surface. The Retinex-based image decomposition is defined as follows:

$$S(x, y) = R(x, y) \cdot L(x, y), \tag{2}$$

where $S(x, y)$ is the observed image, $L(x, y)$ represents the light intensity (i.e., illumination component), and $R(x, y)$ represents the reflection component of the object image.

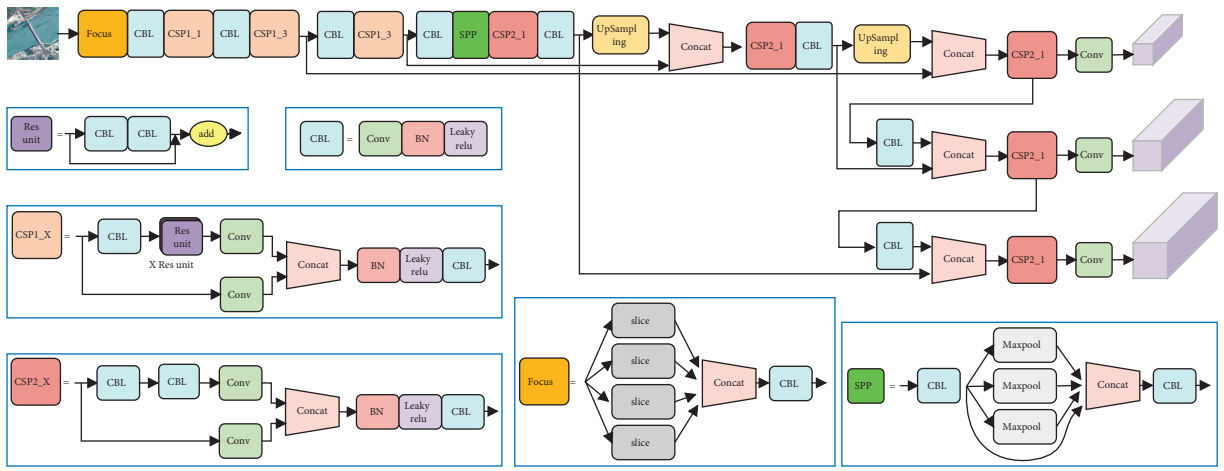


Figure 4: The architecture of YOLOv5 network.

Based on the Retinex theory, for each clean illumination image R, a low-light image can be synthesized by randomly setting a random illumination value L.

4. Experimental Results and Analysis

According to the representative target detection methods introduced in Section 3, we will implement extensive experiments using Faster R-CNN, YOLOv3, YOLOv4, and YOLOv5. In this study, all detection experiments are performed with Ubuntu 18.04 system and run in Intel® Core i9-9900X CPU @ 3.50 GHz, 128 GB ram, and a single NVIDIA GeForce RTX 2080 Ti computing environment. It is worth noting that the YOLOv5 mainly includes 5 different types, i.e., YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. We propose to select the YOLOv5s in this work, which is the lightweight version to efficiently and robustly detect the targets of interest.

4.1. Evaluation Criteria. In our experiments, two main evaluation metrics, i.e., frames per second (FPS) and average precision (AP), will be introduced to evaluate the detection performance. In particular, the FPS represents the execution efficiency for bridge detection. It is the number of FPS that the detection network is able to perform bridge detection. In addition, the AP denotes the area surrounded by the PR (precision-recall) curve. Instead of AP, we propose to use the average version (i.e., mAP) to evaluate the detection accuracy. The mAP represents the mean of the AP value of all classes. It is pointed out that there is only one class of bridge considered in our experiments. The value of mAP is thus equivalent to AP. Both FPS and mAP have become the most widely used standards for measuring target detection performance. The mAP (or AP) is defined as follows:

$$mAP = AP = \int_0^1 P(R)\mathrm{d}R. \tag{3}$$

4.2. Experimental Dataset and Settings. The original image dataset consists of 3500 images of size 900×600 collected under normal imaging condition. In our experiment, we take 67% of the bridge images for network training and the others for testing. Each image is annotated with correct bridge labels and ground truth boxes. In order to explore the influence of data augmentation on bridge detection, we enlarge the original image dataset according to the generation of synthetically degraded images shown in Section 3.5. In Figures 5 and 6, both hazy and low-light images include three different levels. We propose to exploit the TensorFlow 1.13.1 tool package as the framework and ResNet101 as the backbone of Faster R-CNN. We train YOLOv3 and YOLOv5s networks with the Pytorch1.8.1 package and the YOLOv4 network with the Darknet framework. During network training and testing, all detection networks are implemented with the optimal parameters.

4.3. Comparisons with Other Detection Methods. To evaluate the detection performance, we have implemented several experiments on bridge detection under normal imaging conditions. The bridge detection results can be visually illustrated in Figures 7–10. In particular, four typical methods, i.e., Faster R-CNN, YOLOv3, YOLOv4, and YOLOv5s, are simultaneously introduced to detect bridges of interest in our experiments. It can be found that it is easy to perform robust and accurate detection at the presence of large-scale bridges, as shown in Figure 7. These high-quality bridge detection results are beneficial for safer and more convenient transportation. From a theoretical point of view, the cooccurrences of water and land in Figure 8, single backgrounds in Figure 9, and complex backgrounds in Figure 10 could make bridge detection more difficult in practical applications. Due to the more powerful learning capacity, YOLOv5s is able to implement more robust and accurate detection results. The suboptimal detection results, yielded by Faster R-CNN, YOLOv3, and YOLOv4, could lead to poorer transportation surveillance.

Figure 5: The displays of original images and their hazy versions. From top to bottom: original sharp images and hazy images with $\beta = 0.3$, $\beta = 0.4$, and $\beta = 0.5$, respectively. In particular, β denotes the scattering coefficient. In addition, the global airlight A is set to 0.7 in these images.

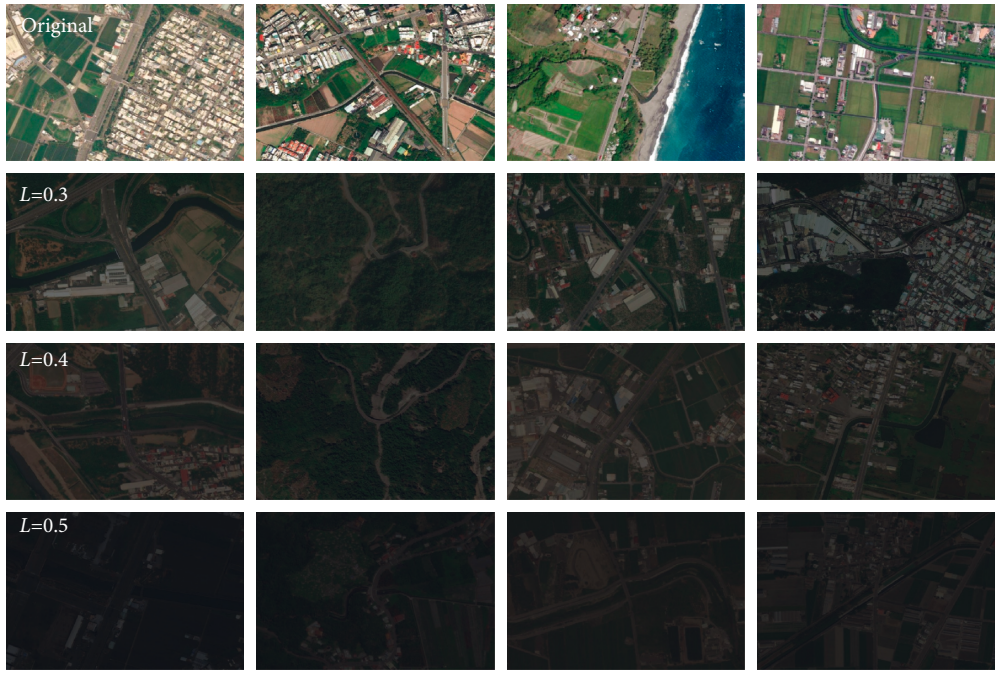
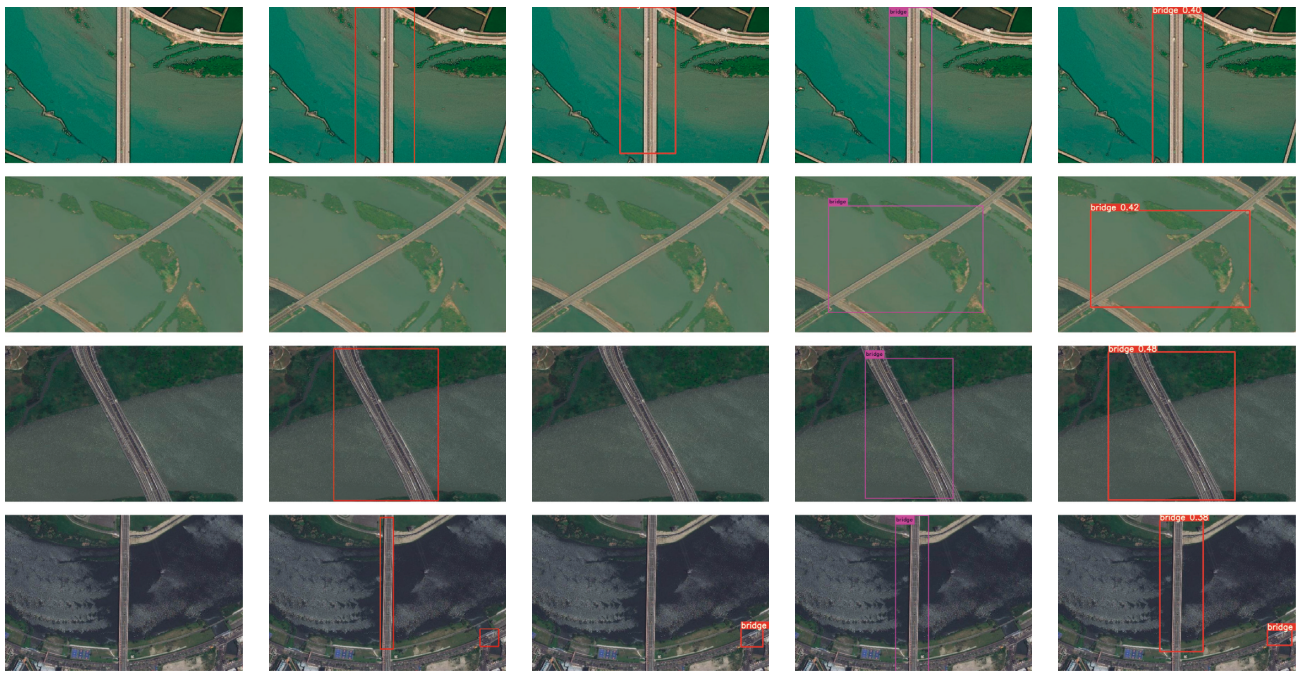


Figure 6: The displays of original images and their low-light versions. From top to bottom: original sharp images and low-light images with $L = 0.3$, $L = 0.4$, and $L = 0.5$, respectively. In particular, L denotes the illumination value.

4.4. Influences of Hazy and Low-Light Conditions on Bridge Detection

4.4.1. Detection Results on Synthetic Weather Conditions. There are two training sets and three testing sets in this experiment, which can be matched to six data groups. Data augmentation is used on both training and test sets. These datasets are composed of original, hazy, and low-light images. Visual comparisons with several detection methods under different conditions are shown in Figures 11–16. The composition of data sets is shown in Table 1.



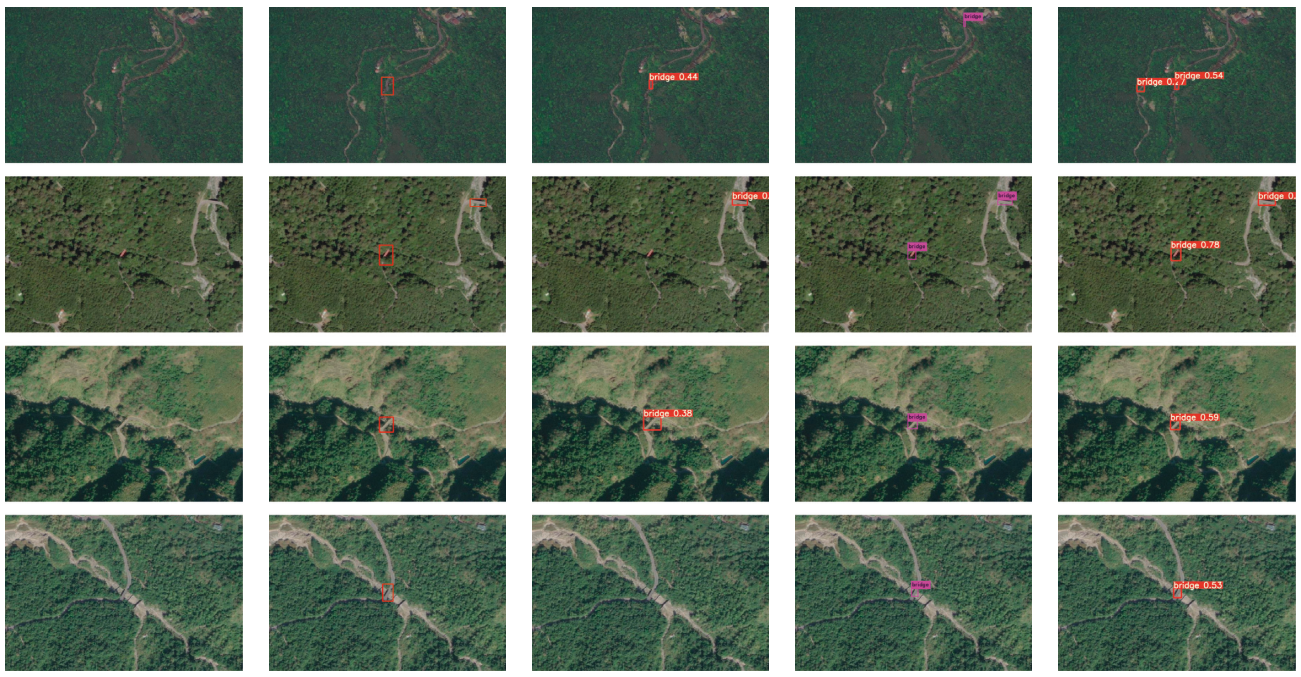
(a) (b) (c) (d) (e)

Figure 7: The comparisons of different competing methods for detection of large-scale bridges. From left to right: (a) original images and detected results by (b) Faster R-CNN, (c) YOLOv3, (d) YOLOv4, and (e) YOLOv5s, respectively.

(a) (b) (c) (d) (e)

Figure 8: The comparisons of different competing methods for bridge detection at the presence of cooccurrences of water and land. From left to right: (a) original images and detected results by (b) Faster R-CNN, (c) YOLOv3, (d) YOLOv4, and (e) YOLOv5s, respectively.

As shown in Table 1, the models are trained in original images and then used to test the bridge detection model in original, low-light, and hazy images, respectively. The mAP results from high to low is YOLOv4, YOLOv5s, Faster R-CNN, and YOLOv3. The FPS performance of the model from high to low is YOLOv5s, YOLOv4, YOLOv3, and Faster R-CNN. The models can meet the needs of real-time bridge detection, except for Faster R-CNN, of which FPS is



(a) (b) (c) (d) (e)

Figure 9: The comparisons of different competing methods for bridge detection under simple backgrounds. From left to right: (a) original images and detected results by (b) Faster R-CNN, (c) YOLOv3, (d) YOLOv4, and (e) YOLOv5s, respectively.

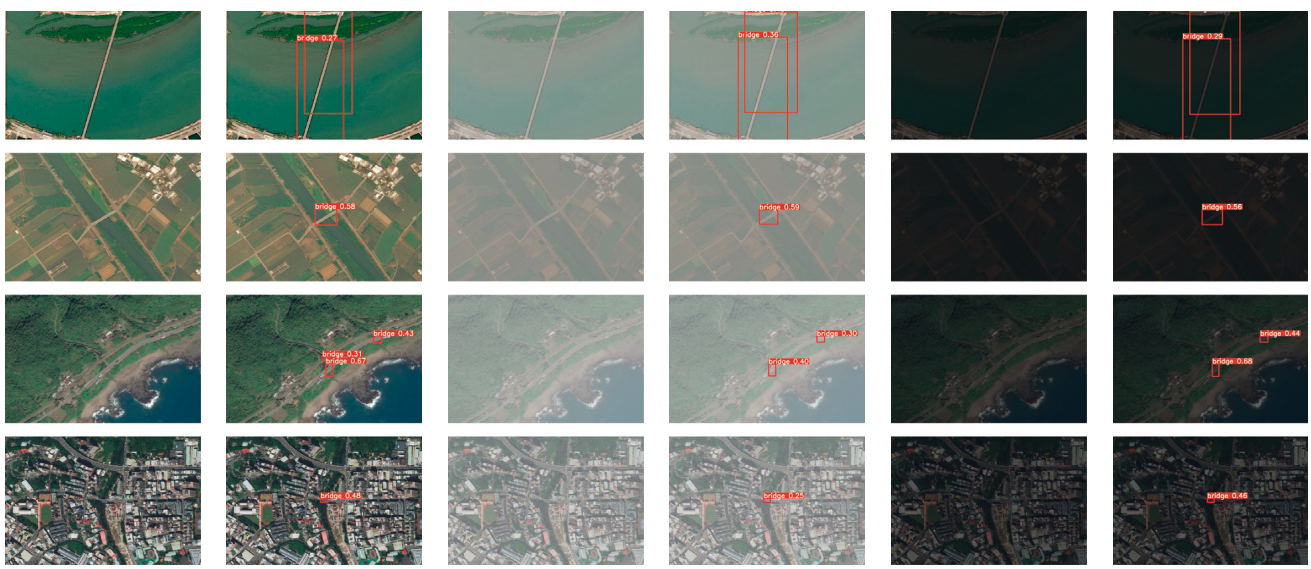
(a) (b) (c) (d) (e)

Figure 10: The comparisons of different competing methods for bridge detection under complex backgrounds. From left to right: (a) original images and detected results by (b) Faster R-CNN, (c) YOLOv3, (d) YOLOv4, and (e) YOLOv5s, respectively.

only 8. In Table 1, the precision of the hazy image test set is about 10% lower than the clear image test set. The precision of the low-light image test set is about 3% lower than the clear image test set. The Faster R-CNN model gets better precision than YOLOv3, but it cannot fulfill the need for real-time detection. In practice, even if Faster R-CNN gets better detection precision, it cannot effectively carry out the bridge detection task. YOLOv4 not only leads Faster R-CNN in precision but also surpasses Faster R-CNN in speed. The models are trained with augmented datasets and then used

(a) (b) (c) (d) (e) (f)

Figure 11: The comparisons of Faster R-CNN under normal, hazy, and low-light imaging conditions. F R-CNN represents the Faster R-CNN. (a) Original. (b) Original + F R-CNN. (c) Hazy. (d) Hay + F R-CNN. (e) Low-light. (f) Low-light + F R-CNN.



(a) (b) (c) (d) (e) (f)

Figure 12: The comparisons of YOLOv3 network under normal, hazy, and low-light imaging conditions. (a) Original. (b) Original + YOLOv3. (c) Hazy. (d) Hay + YOLOv3. (e) Low-light. (f) Low-light + YOLOv3.

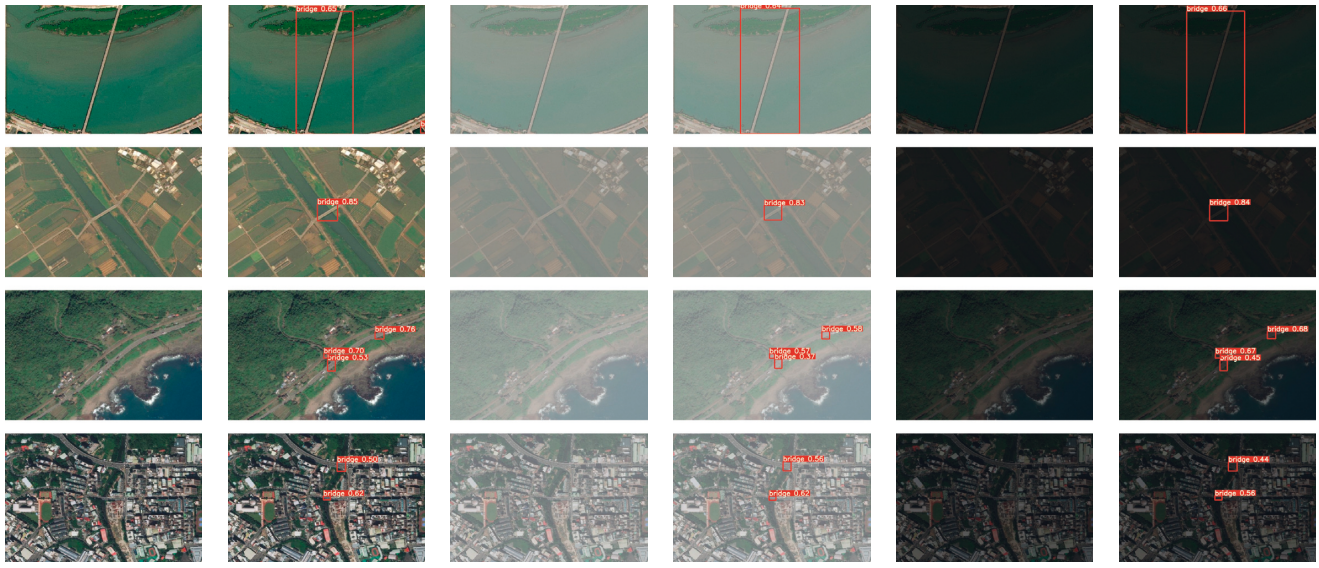
to test the bridge detection model in original, low-light, and hazy images, respectively. Except for YOLOv4, the precision of the hazy image test set is about 2% higher than the clear image test set. The precision of the low-light image test set is about 8% higher than the clear image test set. Surprisingly, the precision of the YOLOv5s model in the low-light image test set is improved by 16%. YOLOv4 gets insufficient improvement and relatively stable results. The YOLOv5s model performs very well in the low-light. It even surpasses YOLOv4. Meanwhile, the YOLOv5s model is more suitable for real-time detection, and FPS is also the best.

Therefore, Table 1 proves that data augmentation is a beneficial trick for training models. The additional data can improve the generalization ability and robustness of the bridge detection model. It enables the model to deal with the complex natural environment. It has momentous practical significance.

4.4.2. Detection Results under Real Weather Conditions. This subsection evaluates the performance of Faster R-CNN, YOLOv3, YOLOv4, and YOLOv5s under realistic weather

(a) (b) (c) (d) (e) (f)

FIGURE 13: The comparisons of YOLOv4 network under normal, hazy and low-light imaging conditions. (a) Original. (b) Original + YOLOv4. (c) Hazy. (d) Hay + YOLOv4. (e) Low-light. (f) Low-light + YOLOv4.

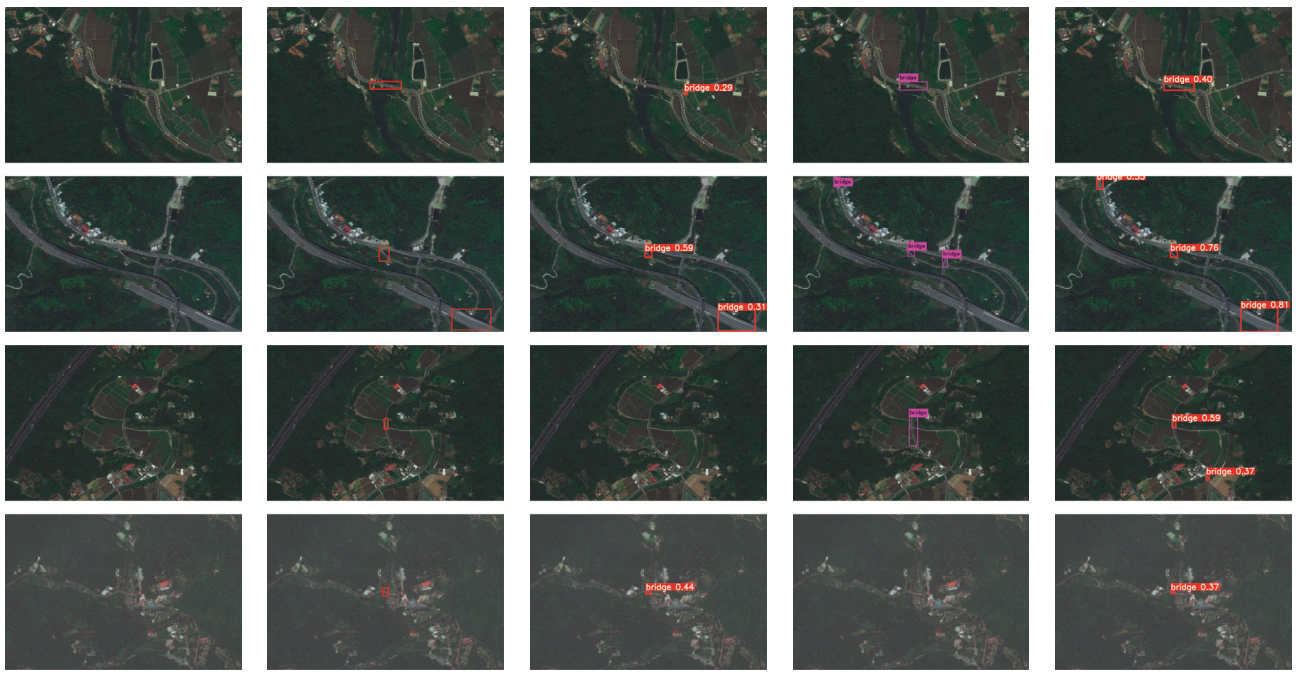


(a) (b) (c) (d) (e) (f)

FIGURE 14: The comparisons of YOLOv5s network under normal, hazy, and low-light imaging conditions. (a) Original. (b) Original + YOLOv5s. (c) Hazy. (d) Hay + YOLOv5s. (e) Low-light. (f) Low-light + YOLOv5s.

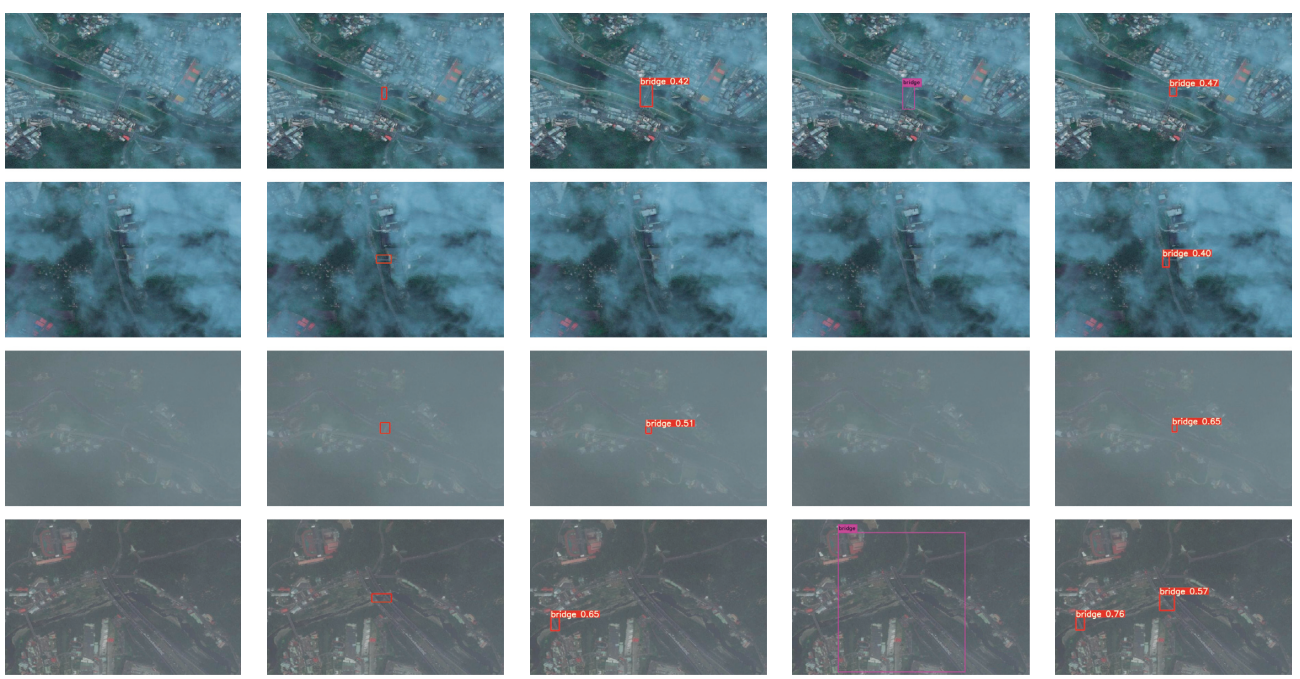
conditions. Figures 15 and 16 depict the detection results under low-light and hazy imaging conditions, respectively. The Faster R-CNN method has been found to have excellent detection accuracy. However, this method typically has a lengthy computational time. It is thus impossible to detect the bridges in real time. The detection capability of YOLOv3 and YOLOv4 may be impaired in low-light and hazy environments, whereas YOLOv5s have the best accuracy and speed performance.

4.5. Discussion. In practical applications, the purpose of data augmentation is to reduce the negative effects of bridge detection. We can conclude from the imaging experiments that hazy and low-light images have different effects on the detection of bridges. According to exhaustive experiments, both hazy and low-light images can easily have a negative impact on the representation of bridge features. In practice, the effect of low-light conditions is weaker than that of hazy conditions. The improvement of low-light conditions after



(a) (b) (c) (d) (e)

FIGURE 15: The comparisons of four different detection methods (i.e., (b) Faster R-CNN, (c) YOLOv3, (d) YOLOv4, and (e) YOLOv5s) for bridge detection under realistic low-light imaging scenarios. (a) Original. (b) Faster R-CNN, (c) YOLOv3, (d) YOLOv4. (e) YOLOv5s.



(a) (b) (c) (d) (e)

FIGURE 16: The comparisons of four different detection methods (i.e., (b) Faster R-CNN, (c) YOLOv3, (d) YOLOv4, and (e) YOLOv5s) for bridge detection under realistic hazy imaging scenarios. (a) Original. (b) Faster R-CNN, (c) YOLOv3, (d) YOLOv4. (e) YOLOv5s.

Table 1: The mean average precision (mAP) results of four bridge detection models trained with original dataset and enlarged dataset by data augmentation strategy.

Dataset	Methods	Clear	Hazy	Low-light	FPS
Original	Faster R-CNN [9]	55.25 ± 1.17	46.30 ± 1.15	51.11 ± 1.22	9
	YOLOv3 [10]	50.02 ± 0.51	42.24 ± 0.54	47.60 ± 0.47	77
	YOLOv4 [11]	60.05 ± 0.39	51.34 ± 0.42	57.70 ± 0.37	67
	YOLOv5 [12]	60.38 ± 0.35	49.50 ± 0.52	55.55 ± 0.44	140
Data augmentation	Faster R-CNN [9]	57.24 ± 0.33	59.45 ± 0.21	65.09 ± 0.41	8
	YOLOv3 [10]	55.76 ± 0.35	57.34 ± 0.23	63.44 ± 0.28	76
	YOLOv4 [11]	69.38 ± 0.09	68.06 ± 0.11	70.03 ± 0.07	65
	YOLOv5 [12]	60.49 ± 0.14	62.79 ± 0.17	76.47 ± 0.16	142

the implementation of a data augmentation strategy is superior to that of haze. Not only can data augmentation improve the detection accuracy, but it can also increase the robustness of bridge detection.

5. Conclusions

In this work, we propose a deep learning–enabled automatic bridge detection method for promoting transportation surveillance under different imaging conditions. It contributes to reliable and robust video surveillance to guarantee more reliable and effective intelligent transportation. The major contributions of this work are as follows: first, a deep learning-enabled automatic bridge detection framework is proposed for transportation surveillance. Second, we construct an original dataset consisting of 3500 images of size 900×600, collected under normal imaging conditions. Moreover, the synthetically degraded images obtained using physical imaging methods are generated to enlarge the imaging dataset to improve the generalization abilities of the deep learning methods. Last, the deep learning methods are effectively trained using the enlarged image dataset, contributing to more reliable and accurate bridge detection results under different conditions.

Although our proposed method has already provided a solution for high-accuracy bridge detection, to further improve the accuracy of bridge detection for better application in tasks such as traffic management, we will conduct the related work from the following three aspects.

(1) To improve the generalization ability for the detection of morphologically inconsistent bridges, we will further expand our dataset through field photography, web browsing, etc. In addition, we will perform the collection of bridges in really harsh imaging environments to further improve the robustness of bridge detection.

(2) To facilitate the subsequent work of bridge detection, we will introduce a rotated detection frame to match the length and width of the bridge as much as possible. Therefore, we intend to redesign the network to further refine the bridge features under high-altitude photography. It will be able to improve the accuracy of small-scale bridge detection.

(3) In the real-world harsh imaging environment, not only fog and low-light are included, but rainy days and sandstorms are also key factors affecting the imaging quality. To be able to adapt to bridge detection in various scenarios, we will develop a bridge detection model suitable for various harsh imaging scenarios in practical applications.

Data Availability

The image data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 41771444).

References

[1] G. Pajares, "Overview and current status of remote sensing applications based on unmanned aerial vehicles (UAVs)," *Photogrammetric Engineering & Remote Sensing*, vol. 81, no. 4, pp. 281–330, 2015.

[2] S. Feroz and S. Abu Dabous, "UAV-based remote sensing applications for bridge condition assessment," *Remote Sensing*, vol. 13, no. 9, p. 1809, 2021.

[3] L. Chen, T. Weng, J. Xing et al., "A new deep learning network for automatic bridge detection from SAR images based on balanced and attention mechanism," *Remote Sensing*, vol. 12, no. 3, p. 441, 2020.

[4] R. W. Liu, Y. Guo, Y. Lu, K. T. Chui, and B. B. Gupta, "Deep network-enabled haze visibility enhancement for visual IoT-driven intelligent transportation systems," *IEEE Transactions on Industrial Informatics*, p. 1, 2022.

[5] Y. Guo, Y. Lu, and R. W. Liu, "Lightweight deep network-enabled real-time low-visibility enhancement for promoting vessel detection in maritime video surveillance," *Journal of Navigation*, vol. 75, no. 1, pp. 230–250, 2022.

[6] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 598–605, 2000.

[7] E. H. Land and J. J. McCann, "Lightness and retinex theory," *Journal of the Optical Society of America*, vol. 61, no. 1, pp. 1–11, 1971.

[8] E. H. Land, "The retinex theory of color vision," *Scientific American*, vol. 237, no. 6, pp. 108–128, 1977.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[10] J. Redmon and A. Farhadi, "Yolov3: An Incremental Improvement," 2018, https://arxiv.org/abs/1804.02767.

[11] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "Yolov4: Optimal Speed and Accuracy of Object Detection," 2020, https://arxiv.org/abs/2004.10934.

[12] Z. Li, W. Xie, L. Zhang et al., "Toward efficient safety helmet detection based on YoloV5 with hierarchical positive sample selection and box density filtering," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022.

[13] D. C. Baker, J. K. Aggarwal, and S. S. Hwang, "Geometry guided segmentation of outdoor scenes. Applications of Artificial Intelligence VI," *SPIEL*, vol. 937, pp. 576–585, 1988.

[14] D. C. Baker, S. S. Hwang, and J. K. Aggarwal, "Detection and segmentation of man-made objects in outdoor scenes: concrete bridges," *Journal of the Optical Society of America*, vol. 6, no. 6, pp. 938–950, 1989.

[15] G. Wang, S. Huang, and L. Jiao, "An automatic bridge detection technique for high resolution SAR images," in *Proceedings of the Asian-Pacific Conference on Synthetic Aperture Radar*, pp. 498–501, Xi'an, China, October 2009.

[16] F. Gao, L. Hu, and Z. He, "Bridge extraction based on on constrained Delaunay triangulation for panchromatic image," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, pp. 1429–1432, Vancouver, BC, Canada, July 2011.

[17] N. Loménie, J. Barbeau, and R. Trias-Sanz, "Integrating textural and geometric information for an automatic bridge detection system," *IEEE International Geoscience and Remote Sensing Symposium*, vol. 6, pp. 3952–3954, 2003.

[18] Y. Han, H. Zheng, Q. Cao, and Y. Wang, "An effective method for bridge detection from satellite imagery," in *Proceedings of the IEEE Conference on Industrial Electronics and Applications*, pp. 2753–2757, Harbin, China, May 2007.

[19] F. Liang, Q. Tan, and Y. Liu, "Research on bridge extraction from HJ-1 remote sensing satellite images," *Engineering of Surveying and Mapping*, vol. 20, no. 5, pp. 13–17, 2011.

[20] D. Sita, "Target detection in SLAR images. Signal processing, sensor fusion, and target recognition III," *International Society for Optics and Photonics*, vol. 2232, pp. 291–299, 1994.

[21] W. Wang, S. Xu, and Q. Yao, "An image segmentation method using blackboard models and its application to bridge recognition," *Chinese Science Abstracts Series A*, vol. 4, no. 14, p. 51, 1995.

[22] Y. Fu, K. Xing, Y. Huang, and Y. Xiao, "Recognition of bridge over water in high-resolution remote sensing images," *World Congress on Computer Science and Information Engineering*, vol. 2, pp. 621–625, 2009.

[23] H. Lu, Y. Deng, Z. Huang, and J. Zhang, "Extraction of bridges from high resolution remote sensing image based on topology modeling," *IEEE International Conference on Consumer Electronics*, pp. 494–499, Boston, MA, USA, March 2014.

[24] W. Liu, Y. Jiang, L. Lei, and G. Kuang, "A method of bridge recognition based on multi-source remote sensing image fusion," *Signal Processing*, vol. 20, no. 4, pp. 427–430, 2004.

[25] B. O. Shu-Kui and R. Nie, "Extract bridge targets from remotely sensed imagery based on object-oriented method," *Computer Engineering and Applications*, vol. 44, no. 26, pp. 200–202, 2008.

[26] Y. Gu, Z. Fan, D. Fan, and J. Lin, "Extracting bridge from airborne LiDAR data based on the difference between diverse filtering methods," *Engineering of Surveying and Mapping*, vol. 23, no. 11, pp. 67–70, 2014.

[27] Z. Yuan, J. Ding, J. Wang, C. Wenqian, L. Xiang, and H. Shuai, "Object-oriented extracting bridges information based on China-made GF-1," *Chinese Journal of Sensors and Actuators*, vol. 28, no. 5, pp. 690–696, 2015.

[28] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[29] R. W. Liu, W. Yuan, X. Chen, and Y. Lu, "An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system," *Ocean Engineering*, vol. 235, Article ID 109435, 2021.

[30] R. Girshick, J. Donahue, T. Darrell, and M. Jitendra, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, June 2014.

[31] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, Montreal, BC, Canada, 2015.

[32] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, Montreal, BC, Canada, 2017.

[33] W. Liu, D. Anguelov, D. Erhan et al., *SSD: Single Shot Multibox Detector*, pp. 21–37, European Conference on Computer Vision, Cham, 2016.

[34] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of theIEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, Las Vegas, NV, USA, 2016.

[35] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263–7271, Honolulu, HI, USA, 2017.

[36] L. Chen, T. Weng, J. Xing et al., "Employing deep learning for automatic river bridge detection from SAR images based on adaptively effective feature fusion," *International Journal of Applied Earth Observation and Geoinformation*, vol. 102, Article ID 102425, 2021.

[37] H. Guo, R. Zhang, Y. Wang, W. Yang, H. C. Li, and G. S. Xia, "Accurate bridge detection in aerial images with an auxiliary waterbody extraction task," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 9651–9666, 2021.

[38] M. Mirmehdi, P. L. Palmer, J. Kittler, and H. Dabis, "Feedback control strategies for object recognition," *IEEE Transactions on Image Processing*, vol. 8, no. 8, pp. 1084–1101, 1999.

[39] L. Sang, Y. Zhang, and Y. Yan, "Joint feature and knowledge rule-based automatic recognition of bridge over water," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, pp. 457–460, Beijing, China, July 2016.

[40] R. Trias-Sanz and N. Loménie, "Automatic bridge detection in high-resolution satellite images," in *Proceedings of theInternational Conference on Computer Vision Systems*, pp. 172–181, Springer, Berlin, Heidelberg, March 2003.

[41] Z. Wang, Y. Zhang, Y. Yu, L. Zhang, J. Min, and G. Lai, "Prior-information auxiliary module: an injector to a deep learning bridge detection model," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 6270–6278, 2021.

[42] X. Zhou, V. Koltun, and P. Krähenbühl, *Tracking Objects as Points*, pp. 474–490, European Conference on Computer Vision, Glasgow, 2020.

[43] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.