An Innovative Approach for Detecting Bridge Defects Based on UAV Imagery in Low-Light Environments

LIANG Zhaolun [(a)], WU Hao [(a)], WANG Mingzhu [(b)], Jack C.P. Cheng [(a)]
[(a)] The Hong Kong University of Science and Technology, Hong Kong
[(b)] Loughborough University, United Kingdom
zliangaq@connect.ust.hk

Abstract. Bridge inspection plays a pivotal role in maintaining bridge safety and structural integrity, which is a critical task in the civil engineering industry. The emerging adoption of UAV (Unmanned Aerial Vehicle) image-based inspection, coupled with AI-driven defect detection models, offers a swift, secure, and cost-effective solution for maintaining bridges. However, when inspecting low-light environments, such as areas under bridges, the poor illumination and increased noise in the image often lead to issues with false detections or missed detection. Regarding the problem, this paper applies image enhancement techniques to low-light under-bridge images and evaluates the performance of commonly used deep learning-based detectors for defect detection. To improve detection performance, an attention mechanism CBAM is incorporated into YOLOv5. Results demonstrate that the proposed CBAM-YOLOv5 algorithm can improve detection accuracy by 2.2%-8.1% compared to other object detectors.

1. Introduction

Maintaining infrastructures is a common concern for civil engineers, as ensuring these assets' regular and safe operation is a persistent and complex issue. Continued monitoring and specific inspection strategies are necessary for maintaining infrastructures such as roads, tunnels, reservoirs, and bridges. Bridge inspection is one of the most difficult tasks, as it frequently entails working at elevated heights and examining the undersides of bridges in poorly lit environments. Inspecting bridges with traditional methods can pose significant risks to inspectors and induce high costs associated with their time-consuming and laborious implications. As bridges will inevitably deteriorate over time, it is essential to introduce an effective bridge inspection approach to prolong their lifespan and reduce the likelihood of disruptive failures. In recent years, there has been growing recognition of the potential of UAVs for bridge inspection with the rapid development of UAV technologies. UAV stands for Unmanned Aerial Vehicle, which is also commonly referred to as a drone. UAVs have the characteristic of high manoeuvrability and can capture high-quality data during their navigation, making them an exceptional tool for inspecting bridges. Another substantial benefit of using UAVs for bridge inspection is their ability to accommodate multiple sensors for various purposes, including high-resolution cameras, infrared cameras, and lidar.

Cameras are cheap and compatible in most cases, making them the competent candidate for visual inspections. The images captured by cameras offer a wealth of visual information, and the implementation of artificial intelligence (AI) algorithms has made it possible to process this visual data automatically. AI-driven detection models can employ sophisticated machine learning and deep learning algorithms to detect objects or anomalies in images or videos. Training these models on comprehensive bridge datasets and corresponding defect annotations can accurately pinpoint various defects, such as cracks, corrosion, and spalling, in bridge inspection works.

However, when dealing with real-world bridge inspections, most of the then-existing AI models are not capable to deliver satisfactory results. One particular challenge is the assessment of areas beneath bridges, as these locations are often subject to low-light conditions. Distractions include the spatiotemporal of sunrise and sunset, seasonal variations, weather, and the presence of obstructions that can affect the quality of under-bridge images. Detecting defects in these dimly lit environments is complicated for both traditional and drone-based inspection methods due to the lack of clarity and detail in the acquired images. The reduced contrast, uneven illumination, and increased noise in low-light images (LLIs) can adversely affect the performance of the detection models. Therefore, attempting to detect defects in these suboptimal images will lead to inaccurate results. To address these challenges, it is essential to employ pre-processing techniques, such as LLI enhancement and denoising methods, prior to applying object detection algorithms.

2. Methodology

2.1 Data Collection

The process of detecting defects in low-light environments initially starts by capturing under-bridge images using UAV. At the beginning, the site investigation determines the UAV flight path and viewpoints, and the UAV's camera must be titled upwards to capture the LLIs under the bridge during the inspection. Figure 1 illustrates the procedure of low-light environment defect detection, which encompasses a series of techniques and strategies designed to enhance the visibility and quality of images, facilitating better analysis and decision-making for maintenance purposes. It is worth noting that acquiring a flight permit from the related sectors is an essential prerequisite before undertaking any UAV inspection activity. The experiment was conducted on the T-shape bridges to validate the feasibility of the proposed method. The T-shape results from the combination of a vertical column (stem) and a horizontal beam (cap). This configuration forms a special box-like structure that acts as a barrier to sunlight and nature lights, casting additional shadows on the under-bridge areas. Real-time monitoring of UAV flight is imperative during the navigation process, as the region under the bridge constitutes a GPS-shielded environment. Except for the effects of the dark environment under the bridge, the lack of GPS signal can easily lead to pilot errors and improper actions, potentially resulting in safety incidents.

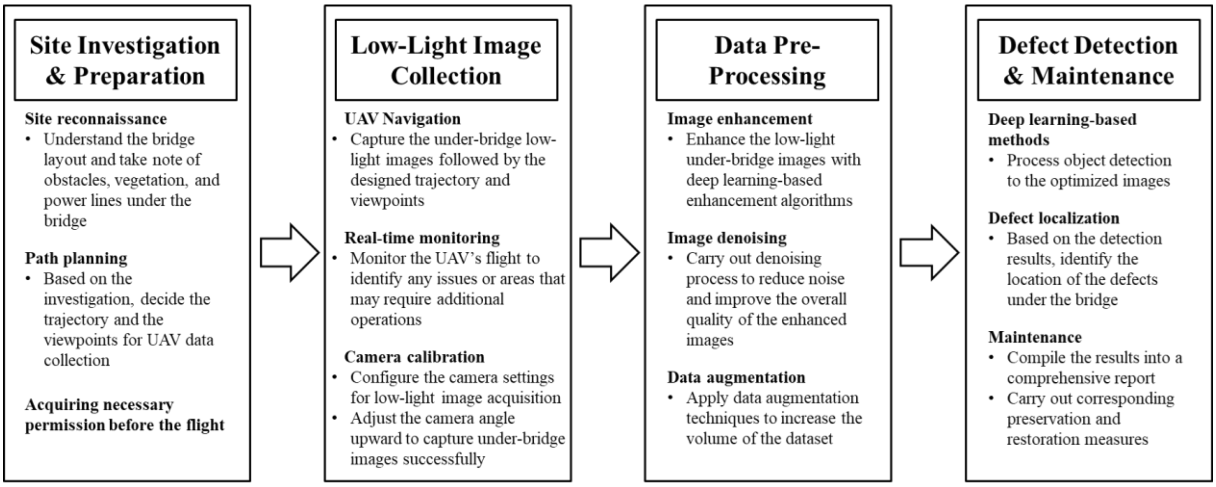
Site Investigation & Preparation
Site reconnaissance
• Understand the bridge layout and take note of obstacles, vegetation, and power lines under the bridge
Path planning
• Based on the investigation, decide the trajectory and the viewpoints for UAV data collection
Acquiring necessary permission before the flight
Low-Light Image Collection
UAV Navigation
• Capture the under-bridge low-light images followed by the designed trajectory and viewpoints
Real-time monitoring
• Monitor the UAV's flight to identify any issues or areas that may require additional operations
Camera calibration
• Configure the camera settings for low-light image acquisition
• Adjust the camera angle upward to capture under-bridge images successfully
Data Pre-Processing
Image enhancement
• Enhance the low-light under-bridge images with deep learning-based enhancement algorithms
Image denoising
• Carry out denoising process to reduce noise and improve the overall quality of the enhanced images
Data augmentation
• Apply data augmentation techniques to increase the volume of the dataset
Defect Detection & Maintenance
Deep learning-based methods
• Process object detection to the optimized images
Defect localization
• Based on the detection results, identify the location of the defects under the bridge
Maintenance
• Compile the results into a comprehensive report
• Carry out corresponding preservation and restoration measures

Figure 1: Flowchart of the low-light environment defect detection process

2.2 Image Enhancement Technologies

Upon obtaining the low-light images (LLIs) under the bridge, it is crucial to perform image pre-processing techniques to ensure the successful implementation of defect detection. LLIs often suffer from issues such as increased noise levels, low contrast, and reduced visibility, which can result in false detections and reduced confidence when conducting image detection. Moreover, many object detection models are trained with well-lit and high-quality images and may not be optimized for processing LLIs. Typical traditional machine learning enhancement models including Gamma correction, histogram equalization (HE) (Pisano et al., 1998), Retinex theory-based methods, and frequency-based methods (Huang et al., 2013). While traditional algorithms can be effective in certain scenarios, they often face considerable difficulties when performing specific image enhancement tasks. First, traditional methods are often built with fixed parameters and assumptions, tend to cause the problems of over-enhancement and under-enhancement. Second, the color balance of the images will be altered during the enhancement, resulting in unnatural color shifts in the targeted images. Third, traditional methods face difficulties in handling images with extremely dark or bright regions, as they lack the capacity to balance enhancement evenly across the entire image. Therefore, employing advanced deep learning-based enhancement techniques to deal with the under-bridge problem is more favorable.

In this paper, the enhancement model used is Zero-DCE (Zero-Reference Deep Curve Estimation) (Guo et al., 2020), a deep learning-based zero-reference enhancement technique. Zero-DCE leverages a deep network to estimate pixel-wise and high-order curves for dynamic range adjustment to enhance the input images. The dynamic range adjustment enables the model to adaptively adjust the contrast and brightness of different regions in the images, resulting in more balanced enhancements. Many image enhancement models, especially those based on supervised learning, require paired images for training. In these instances, the training dataset comprises low-light images and their counterpart high-quality images. Due to the absence of a public database containing dark images of under-bridge, the data utilized in this study were acquired by the authors. The well-known dark image datasets include the ExDark (Loh & Chan, 2019) and DICM (Lee et al., 2012) dataset; however, these datasets do not contain corresponding paired images. Moreover, many of these datasets are primarily used for object detection rather than defect detection. The Microsoft COCO (Lin et al., 2014) dataset also contains low-light images, but they comprise less than 0.2% of the overall data, posing difficulties for their effective utilization in low-light studies. Considering these factors, employing Zero-DCE to enhance the LLIs captured under bridges serves as an ideal solution. Compared to other image enhancement algorithms, Zero-DCE is an end-to-end model and does not require reference images for training, ensuring training efficiency and simplifying the enhancement pipeline. Additionally, the model can be easily fine-tuned or adjusted to suit different image modalities and datasets, making it a flexible and versatile solution for various enhancement applications.

2.3 Image Denoising and Augmentation Technologies

After completing the image enhancement process, a denoising procedure for the under-bridge images is necessary. In most cases, the step of image enhancement will amplify the noise present in the original image, which can lead to unwanted artifacts and disturbance. Image denoising is essential in alleviating or suppressing noise amplification, consequently enhancing the performance of downstream tasks, such as image classification, object detection, and segmentation. Block-Matching and 3D Filtering (BM3D) (Dabov et al., 2007) is a robust algorithm exploiting spatial and transform-domain redundancies to achieve effective image denoising. The BM3D consists of two stages: the block grouping stage and the collaborative filtering stage. In the grouping stage, the algorithm employs the block-matching technique to search for similar fragments; those fragments will be stacked to form 3D arrays. In the collaborative filtering stage, a shrinkage operation is applied to the transform coefficients of each 3D group, which can diminish noise while retaining the essential characteristics of the signal. The 3D group is later processed by an inversion of linear transform to form a set of 2D image fragments to the original image position, and the overlapping fragments are weight-averaged to further ensure the noises are effectively filtered. Moreover, as the number of acquired images is relatively small, it is essential to perform appropriate data augmentation techniques to prevent overfitting and instability problems of the detection model. In this study, several augmentation techniques are applied to the existing data, such as rotation, flipping, and cropping.

2.4 Object Detection Models

Object detection models are among the most frequently used techniques in computer vision tasks. Traditional object detection models, including Viola-Jones (Viola & Jones, 2001) detectors, Histogram of Oriented Gradients (HOG) (Dalal & Triggs, 2005), and Deformable Part Model (DPM) (Felzenszwalb et al., 2008), relied on hand-crafted features due to limited image representation available at the early age. The performance of traditional object detectors is suboptimal in many cases because they were designed on human intuition and sensitive to variations in object appearance, such as occlusion, pose, scale, and lighting. Bridge inspection images often encounter the problem of variant lighting and object occlusion, making it challenging to achieve good results using hand-crafted feature-based detectors. Compared to traditional methods, deep learning-based object detectors exhibit greater resilience to variations and provide notable enhancements in performance, adaptability, and effectiveness.

The Region-based Convolutional Neural Network (R-CNN) (Girshick et al., 2014) represents a significant breakthrough in the field of computer vision, leveraging deep convolutional networks to achieve exceptional object detection accuracy. In R-CNN, selective search is utilized to generate region proposals, and a pre-trained CNN is employed for feature extraction in each proposal. Object prediction is completed using the linear Support Vector Machine (SVM) (Cortes & Vapnik, 1995) classifiers in each region generated. The introduction of R-CNN has been followed by several iterations and improvements, culminating in the creation of Fast R-CNN (Girshick, 2015) and Faster R-CNN (Ren et al., 2017). Fast R-CNN introduces the Region of Interest (RoI) pooling layer that enables the detection network to extract fixed-size features from each RoI, resulting in faster detection speed than R-CNN. The subsequent release of Faster R-CNN introduced a significant improvement by replacing selective search with a Region Proposal Network (RPN) for generating region proposals. This enhancement further accelerated the object detection process.

In contrast to R-CNN methods, You Only Look Once (YOLO) (Redmon et al., 2016) is capable of detecting objects throughout the entire image rather than examining specific regions. For efficient detection and training, the YOLO network partitions the image into cells and estimates each cell's class probabilities and bounding box coordinates. Subsequently, it consolidates predictions from all cells to generate a final set of detections for the image. YOLO is categorized as a one-stage detector due to its ability to detect the entire image, leading to faster detection speeds compared to the two-stage detectors such as R-CNN, Faster R-CNN, and FPN (Feature Pyramid Networks) (Lin et al., 2017). Nonetheless, this speed comes at the cost of accuracy, particularly when dealing with small and closely clustered objects. In recent years, YOLO has spawned numerous derivative algorithms, such as YOLOv2, YOLOv3, up to YOLOv7, among others, with the aim of improving object detection accuracy and overall performance.

Other mainstream one-stage object detectors, such as SSD and RetinaNet, are proficient in detecting small objects and managing objects with diverse sizes. The Single Shot MultiBox Detector (SSD) (Liu et al., 2016) employs multi-reference and multi-resolution detection techniques to extract features, enabling it to effectively identify objects of varying sizes and aspect ratios, especially improving small objects' detection accuracy. RetinaNet (Lin et al., 2017) uses a focal loss function that down-weights the impact of easy negative examples, focusing instead on hard negative examples as they provide more informative training examples for the network to learn from.

2.5 YOLOv5 with Improved Attention Mechanism

Several object detectors were employed to detect the defects in the optimized under-bridge images, including Faster R-CNN, SSD, RetinaNet, YOLOv5, and CBAM-YOLOv5. Among the many detection models, YOLOv5 is distinguished by its high efficiency and real-time performance, demonstrating significant value in both industrial applications and academic research. In this paper, CBAM (Convolutional Block Attention Module) (Woo et al., 2018) is integrated with YOLOv5 to develop CBAM-YOLOv5 for achieving higher detection accuracy, where CBAM is an attention mechanism. The attention mechanism is a widely used data processing method in machine learning tasks across various fields. The main concept of the attention mechanism in computer vision is to identify correlations between raw data and then emphasize important features. These features can include channel attention, pixel attention, and multi-order attention. The CBAM is a lightweight module that performs attention operations in the channel and spatial dimensions. It consists of a channel attention module (CAM), which enables the network to focus more on the foreground and meaningful areas of the image, and a spatial attention module (SAM), which allows the network to prioritize locations that are rich in contextual information about the entire picture. The structure of CBAM mechanism is shown in Figure 2.

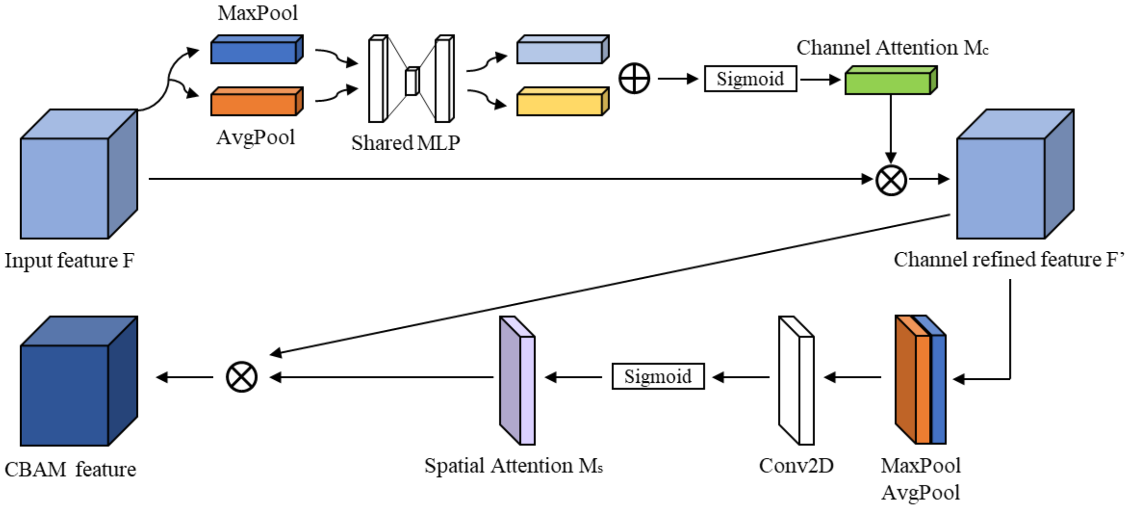

Figure 2: Structure of CBAM

In the channel attention module (CAM), the input feature map with a shape of H×W×C undergoes global max pooling (GMP) and global average pooling (GAP) operations, resulting in two feature maps of size 1×1×C. These two obtained feature maps are passed through a two-layer multilayer perceptron (MLP) with a hidden layer. The first layer of the MLP contains C/r neurons (where r is the reduction rate) with the ReLU activation function. The second layer has C neurons and shares weights with the first layer. The output features are then added element-wise and passed through a sigmoid activation function to produce the final channel attention features. These features are multiplied with the original input feature map to obtain the channel refined feature as the input for the spatial attention module (SAM). After applying global max pooling (GMP) and global average pooling (GAP) to the input features, two feature maps of size H×W×1 are obtained. These feature maps are concatenated, and the resulting tensor is passed through a series of convolutional layers to reduce its dimensionality and generate spatial attention feature using a sigmoid activation function. Finally, the spatial attention features are multiplied with the input feature map to obtain the CBAM feature map.

3. Validation of the Proposed Method

3.1 Application to the T-shape Bridge

Among all infrastructure inspection projects, bridge inspection is considered intricate because it requires high-altitude and low-light operations. While UAVs can serve as an effective solution for high-altitude inspection, they may not have the ability to address the low-light challenges directly. In most cases, the lighting conditions on a bridge's upper and lower structures can vary greatly. Due to the absence of obstacles, the components on the upper bridge are more easily observed, whereas under-bridge structures often remain in low-light environments for prolonged periods. To assess the under-bridge area, traditional bridge inspection methods often use additional lighting sources to improve visibility for inspections conducted in low-light conditions.

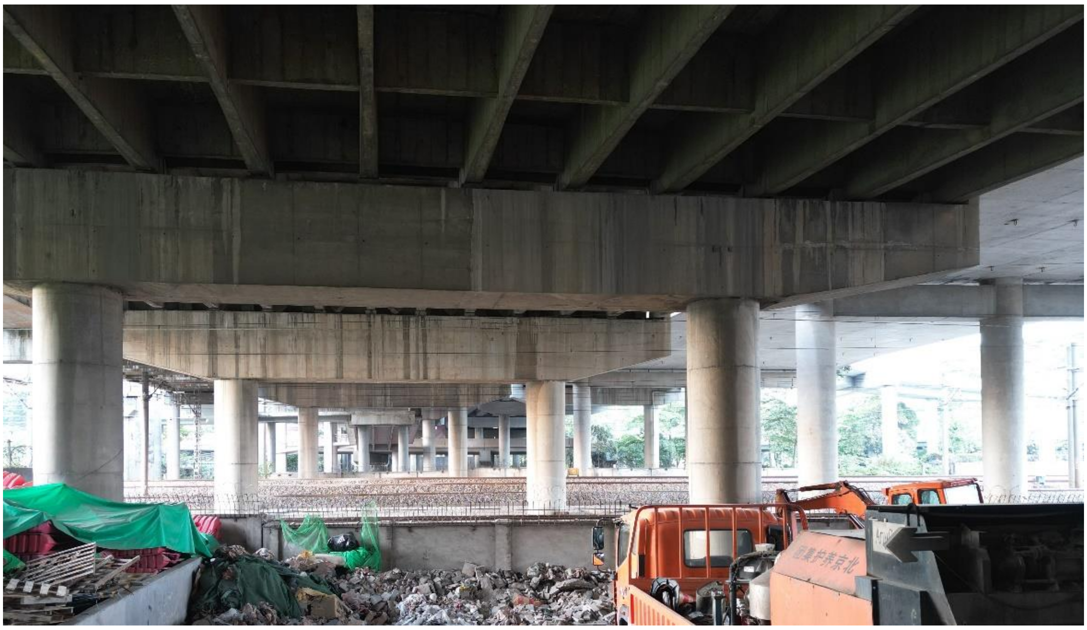

Figure 3: Complex under-bridge environment

The T-shape bridge selected for the experiment is located in Guangzhou, China. An example of the target bridge appearance and the environment of the under-bridge area captured by the UAV is shown in Figure 3. In comparison to other bridge types, the box-like under-bridge structure of a T-shape bridge can generate additional shadows, further exacerbating low-light challenges during inspections. It is important to note that the camera angle of the UAV must be facing upwards for image capturing during aerial inspections. The experiment was conducted under favorable weather conditions, with a total of 540 under-bridge LLIs captured by DJI Mini 3 Pro. The operating environment of the whole process was using Intel Core i7-6700 CPU @ 3.4 GHz×8 and an NVIDIA GeForce GTX 3070 GPU. The universality of the proposed approach is demonstrated using a generic operating system in this research, indicating that it can be replicated at an affordable cost.

3.2 Result and Discussion

Table 1 presents the detection results with different IoU thresholds in this study and demonstrates that the proposed CBAM-YOLOv5 model outperforms other classical deep learning algorithms in detecting enhanced LLIs. Compared to other algorithms, the mean average precision (mAP) improvement of CBAM-YOLOv5 is between 2.2% and 8.1% when the IoU threshold is set to 0.5. In Table 1, the metric mAP@0.5:0.95 denotes the average precision across a series of IoU thresholds, ranging from 0.5 to 0.95, with an increment of 0.05. Precision@0.5 evaluates the ratio of accurate detections to all detected objects when the IoU threshold is set to 0.5. Precision@0.5:0.95 computes the respective precision when the optimal IoU threshold is chosen between 0.5 and 0.95. Recall assesses the capability of the model to identify positive instances accurately. The F1-score reflects the model's performance by considering the harmonic mean of precision and recall.

Instead of adjusting the backbone of YOLOv5, the proposed method integrates CBAM into the enhanced feature extraction network, allowing it to be utilized without compromising the original features extracted by the network. The reason for avoiding backbone modification is that adding an attention mechanism may alter or reduce the original weights, resulting in undesirable prediction results. Figure 4 shows the image enhancement achieved by applying Zero-DCE and BM3D and the detection outcomes obtained with CBAM-YOLOv5. The results indicate that the proposed approach is more effective in identifying defects in the low-light environment of a T-shaped bridge.

Table 1: Detection results obtained by different object detectors

Models	**mAP**	**Precision**	**Recall**	**F1-Score**	**mAP**	**Precision**	**Recall**	**F1-Score**
	@0.5	**@0.5**	**@0.5**	**@0.5**	**@0.5:0.95**	**@0.5:0.95**	**@0.5:0.95**	**@0.5:0.95**
SSD	0.832	0.778	0.729	0.752	0.569	0.461	0.373	0.412
Faster R-CNN	0.878	0.844	0.729	0.781	0.630	0.512	0.435	0.470
RetinaNet	0.891	0.823	0.781	0.801	0.670	0.605	0.450	0.514
YOLOv5	0.902	0.853	0.776	0.812	0.708	0.625	0.420	0.502
YOLOv5+ CBAM	0.906	0.857	0.792	0.823	0.697	0.625	0.436	0.515

After comparing the results obtained from various detection models, it was found that SSD had the worst detection performance. One of the reasons for the inferior performance of SSD compared to other models is that it was developed earlier and makes a trade-off between detection speed and accuracy. Although SSD and Faster R-CNN were developed around the same period, Faster R-CNN can achieve higher detection accuracy due to its two-stage detection process. RetinaNet utilizes the focal loss function to achieve a better balance between recall and precision, which is important in crack detection where the target objects are relatively small or sparse. As a well-developed algorithm, YOLOv5 surpasses other models in detection capacity and speed due to its innovative neck structure, spatial attention modules, and anchor box design. Its fast processing speed also enables it to efficiently handle large-scale infrastructure inspections, such as bridge inspections. By integrating CBAM with YOLOv5, the feature extraction capability of YOLOv5 is further improved, resulting in enhanced detection accuracy of CBAM-YOLOv5.

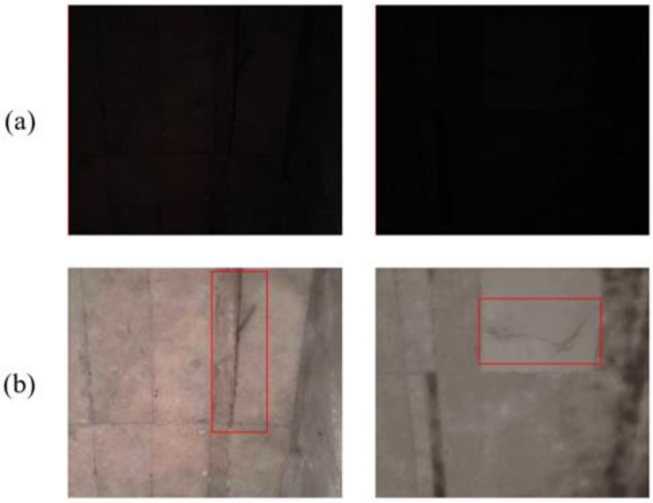

Figure 4: (a) Original low-light images (b) Detection results obtained by CBAM-YOLOv5 in the enhanced images

4. Conclusion

The paper introduces a novel approach for detecting defects in low-light environments, specifically in under-bridge regions. UAVs are employed for capturing low-light under-bridge images owing to their agility in the air and high-resolution image acquisition system. In general, AI algorithms cannot accurately detect low-light images (LLIs) captured under bridges due to factors such as high noise levels, low contrast, and reduced visibility. Therefore, the acquired LLIs must be enhanced and denoised prior to the defect detection process. The image enhancement algorithm utilized in this study is Zero-DCE, which can enhance low-light under-bridge images without requiring paired images. The denoising algorithm employed is BM3D, a classical and robust technique for reducing image noise. The introduction of CBAM, an attention mechanism, improves the feature extraction capabilities of YOLOv5. The article compares the performance of various detection models and finds that the CBAM-YOLOv5 detection algorithm outperforms other classical detection algorithms in detecting cracks under bridges by 2.2%-8.1%. This article offers valuable insights into the implications of defect detection in low-light environments, highlighting its potential for the future of civil infrastructure inspection.

References

Pisano, E.D. et al. (1998) “Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms,” Journal of Digital Imaging, 11(4), pp. 193–200. Available at: https://doi.org/10.1007/bf03178082.

Huang, S.-C., Cheng, F.-C. and Chiu, Y.-S. (2013) “Efficient contrast enhancement using adaptive gamma correction with weighting distribution,” IEEE Transactions on Image Processing, 22(3), pp. 1032–1041. Available at: https://doi.org/10.1109/tip.2012.2226047.

Guo, C. et al. (2020) “Zero-reference deep curve estimation for low-light image enhancement,” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) [Preprint]. Available at: https://doi.org/10.1109/cvpr42600.2020.00185.

Loh, Y.P. and Chan, C.S. (2019) “Getting to know low-light images with the exclusively Dark Dataset,” Computer Vision and Image Understanding, 178, pp. 30–42. Available at: https://doi.org/10.1016/j.cviu.2018.10.010.

Lee, C., Lee, C. and Kim, C.-S. (2012) “Contrast enhancement based on layered difference representation,” 2012 19th IEEE International Conference on Image Processing [Preprint]. Available at: https://doi.org/10.1109/icip.2012.6467022.

Lin, T.-Y. et al. (2014) “Microsoft Coco: Common Objects in Context,” Computer Vision – ECCV 2014, pp. 740–755. Available at: https://doi.org/10.1007/978-3-319-10602-1_48.

Dabov, K. et al. (2007) “Image denoising by sparse 3-D transform-domain collaborative filtering,” IEEE Transactions on Image Processing, 16(8), pp. 2080–2095. Available at: https://doi.org/10.1109/tip.2007.901238.

Viola, P. and Jones, M. (2001) “Rapid object detection using a boosted cascade of Simple features,” Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001 [Preprint]. Available at: https://doi.org/10.1109/cvpr.2001.990517.

Dalal, N. and Triggs, B. (2005) “Histograms of oriented gradients for human detection,” 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) [Preprint]. Available at: https://doi.org/10.1109/cvpr.2005.177.

Felzenszwalb, P., McAllester, D. and Ramanan, D. (2008) “A discriminatively trained, multiscale, deformable part model,” 2008 IEEE Conference on Computer Vision and Pattern Recognition [Preprint]. Available at: https://doi.org/10.1109/cvpr.2008.4587597.

Girshick, R. et al. (2014) “Rich feature hierarchies for accurate object detection and semantic segmentation,” 2014 IEEE Conference on Computer Vision and Pattern Recognition [Preprint]. Available at: https://doi.org/10.1109/cvpr.2014.81.

Cortes, C. and Vapnik, V. (1995) “Support-Vector Networks,” Machine Learning, 20(3), pp. 273–297. Available at: https://doi.org/10.1007/bf00994018.

Girshick, R. (2015) “Fast R-CNN,” 2015 IEEE International Conference on Computer Vision (ICCV) [Preprint]. Available at: https://doi.org/10.1109/iccv.2015.169.

Ren, S. et al. (2017) “Faster R-CNN: Towards real-time object detection with region proposal networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), pp. 1137–1149. Available at: https://doi.org/10.1109/tpami.2016.2577031.

Lin, T.-Y. et al. (2017) “Feature Pyramid Networks for Object Detection,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Preprint]. Available at: https://doi.org/10.1109/cvpr.2017.106.

Redmon, J. et al. (2016) “You only look once: Unified, real-time object detection,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Preprint]. Available at: https://doi.org/10.1109/cvpr.2016.91.

Liu, W. et al. (2016) “SSD: Single shot multibox detector,” Computer Vision – ECCV 2016, pp. 21–37. Available at: https://doi.org/10.1007/978-3-319-46448-0_2.

Lin, T.-Y. et al. (2017) “Focal loss for dense object detection,” 2017 IEEE International Conference on Computer Vision (ICCV) [Preprint]. Available at: https://doi.org/10.1109/iccv.2017.324.

Woo, S. et al. (2018) “CBAM: Convolutional Block Attention Module,” Computer Vision – ECCV 2018, pp. 3–19. Available at: https://doi.org/10.1007/978-3-030-01234-2_1.