

Integrating Mixed Reality Technology, Deep Learning and Domain Knowledge for bridge inspection

Zheyu Liu¹, Zhen Sun¹, Patricia Vanova²

¹School of Civil Engineering, Southeast University, Nanjing 211189, China

²Center of Research and Innovation in Construction, Faculty of Civil Engineering, Technical University of Kosice,
Vysokoskolska 4, Kosice, 042 00, Slovakia

Email: 1318276241@qq.com; sunzhen@seu.edu.cn; patricia.vanova@tuke.sk

ABSTRACT: This study presents an integrated intelligent framework for bridge inspection that synergizes Mixed Reality (MR) technology, deep learning-based object detection, and domain-specific engineering knowledge. Utilizing Microsoft HoloLens 2 as the hardware platform, the system captures real-time 3D bridge surface imagery and deploys the optimized YOLOv11n-ZY model—enhanced with a ZZ convolutional module, YY attention mechanism, and SPPF-LSKA fusion module—to automatically detect and classify multi-category defects including cracks, corrosion, and spalling. Detection results are visualized within an MR interface and dynamically assessed through embedded expert knowledge. Validated on a custom dataset containing 4,176 images of 12 defect types under complex backgrounds, the proposed model achieves 40.3% mAP50 at 60 FPS with only 2.87 million parameters, outperforming existing YOLO variants. Implementation at the case study bridge demonstrates real-time defect localization, 3D model updating, and closed-loop maintenance functionality. The framework advances intelligent infrastructure management by establishing a scalable pipeline for accurate defect assessment and lifecycle-oriented bridge maintenance.

KEY WORDS: HoloLens 2; Object detection; YOLOv11; Visualization interface; Intelligent bridge operation and maintenance.

1 INTRODUCTION

As a critical component of modern transportation infrastructure, bridges are essential for maintaining socio-economic stability and ensuring public safety [1]. Consequently, their structural safety, stability, and health status are of paramount importance.

At present, traditional inspection methods, while capable of detecting visible defects such as cracks and corrosion, are constrained by inefficiency, heavy dependence on specialized knowledge, and challenges in handling complex environments, limiting their application in modern bridge inspection [2]. However, with the rapid development of computer vision technologies, deep learning-based object detection algorithms (e.g., YOLO, Faster R-CNN) have gradually been introduced into industrial defect detection, achieving significant improvements in efficiency and accuracy.

In the field of bridge defect detection [3], Mixed Reality (MR) technology has also provided novel solutions to traditional inspection approaches. Utilizing devices like HoloLens 2, 3D models can be projected into the real world, enabling engineering personnel to observe bridge surfaces from multiple angles and dimensions, thereby enhancing defect identification accuracy. For instance, certain Chinese bridge institutes have integrated 3D laser scanning with BIM technology to control modeling errors within $\pm 2\text{mm}$ [4]; however, algorithm robustness in complex environments still requires further improvement.

The integration of emerging technologies such as deep learning, augmented reality (AR), and mixed reality (MR) has rendered bridge defect detection more efficient, precise, and intelligent [5]. At the same time, machine learning approaches have been integrated for defect detection in concrete structures

in the past few years [6]. Nonetheless, adaptability to complex environments and capabilities for lifecycle management still need enhancement to better meet practical operational demands.

This study aims to develop an intelligent bridge defect detection system by integrating multiple technologies. The research focuses on two main objectives.

First, HoloLens 2 is leveraged to develop a mixed reality system that enables 3D visualization, virtual-physical integration, and multi-dimensional defect observation.

Second, the system combines real-time scanning data processed by YOLOv11 with pre-constructed 3D virtual models to create an innovative bridge inspection framework supporting long-term intelligent operation and maintenance.

2 UNITY-BASED 3D BRIDGE MODEL AND HOLOLENS 2 DEPLOYMENT

2.1 Case Study Bridge Information

This study selects the Wenxi Bridge (Figure 1) in Suijiang New County, Yunnan Province, China, as the exemplar bridge for the visualized intelligent bridge inspection platform. By leveraging the 3D scanning capabilities of HoloLens 2, comprehensive structural information of the bridge was collected. Through comparison with preliminary design drawings, discrepancies between the actual bridge structure and the original plans—caused by on-site construction adjustments or undocumented modifications—were effectively resolved, enabling the successful creation of a 3D model of the Wenxi Bridge in Unity.



Figure 1. Case study bridge - Wenxi Bridge.

The deck width of the case study Bridge is 9 m + 2×1.5 m (pedestrian walkways). The superstructure comprises 9×30 m prestressed concrete simply supported T-beams arranged in three-span continuous units. The substructure includes column piers, U-shaped abutments, and ribbed abutments, with foundation types consisting of spread foundations and pile foundations.

2.2 Characteristics of Mixed Reality (MR) Technology and HoloLens 2

Mixed Reality (MR) technology digitizes physical environments and integrates them with virtual objects to create a visualized interactive space where physical and virtual elements coexist. Compared to Virtual Reality (VR) and Augmented Reality (AR), MR not only superimposes virtual entities into real environments but also achieves precise spatial mapping and real-time interaction between virtual objects and physical spaces, forming a spatially consistent mixed reality environment [7].

Microsoft HoloLens 2 (Figure 2), the second-generation MR device released by Microsoft, demonstrates technical advantages in bridge defect detection [8].



Figure 2. HoloLens 2.

2.3 Unity Model Deployment on HoloLens 2

The mixed reality (MR) application development for HoloLens 2 is based on the following software and toolkits: Windows 10 SDK, Visual Studio 2023, HoloLens 2 Emulator, Unity 2022.3.53f1c1, Unity Hub, and MRTK 2.8. To ensure efficient development and deployment, the hardware configuration listed in Table 1 was adopted:

Table 1. Computer Hardware Configuration.

CPU	GPU	RAM	Storage	Display
i9-14900HX	RTX 4070	64GB	3TB SSD	2560x1600 / 240Hz / 18-inch

Deployment Workflow is described as follows. "Developer Mode" on both the host computer and HoloLens 2 within the Windows operating system is firstly enabled. And the following procedures are adopted.

(1) Project Creation:

Create a new project via the Unity Hub integrated development environment and access the Build Settings interface. Select Universal Windows Platform (UWP) and execute the "Switch Platform" operation. This process automatically performs platform compatibility checks and restructures project resource formats through the underlying engine to meet UWP-specific technical requirements.

(2) MR Toolkit (MRTK) Integration:

Import the Mixed Reality Toolkit using the Mixed Reality Feature Tool (MRFT).

The core principle involves modifying the manifestation configuration file to guide the Unity engine in correctly identifying and loading MRTK modules. Upon returning to the Unity environment, the system automatically initiates dependency detection and resource loading. Compared to traditional methods, MRFT integration effectively avoids dependency conflicts [7], significantly reducing the complexity of managing mixed reality toolchains.

(3) Unity Project Configuration:

As mixed reality applications fall under the extended reality (XR) domain, activate Unity's built-in XR framework. After configuring the Player module, navigate to the XR Plug-in Management section and install the plugin management component. Enable the "Initialize XR on Startup" parameter and activate "Windows Mixed Reality" to ensure precise hardware-software compatibility, laying the foundation for subsequent MR development.

(4) Unity Project Export:

Select the target scene file and configure parameters: (1) Set device compatibility to "Any Device" for universal platform support. (2) Select the x64 architecture for optimal runtime efficiency. (3) Optimize auxiliary parameters based on specific development requirements. Initiate the conversion of the Unity project to a Visual Studio solution via the Build button, establishing a standardized framework for application compilation and deployment.

(5) Deployment to HoloLens 2 Device:

Open the completed project in Visual Studio 2023 and proceed to the deployment parameter configuration phase. Simultaneously, ensure that the host development environment and the target device are connected to the same wired or wireless local area network (LAN), with sufficient network bandwidth to meet real-time transmission requirements for application image files, thereby guaranteeing the validity of the deployment process.

Within the Visual Studio integrated development environment (IDE), sequentially perform three critical configurations: (1) Program compilation using Debug mode; (2) Selection of the ARM64 instruction set architecture to align

with the target device's hardware specifications; (3) Configuration of remote computer deployment options.

Subsequently, input the target device's network IP address in the debugging parameters module and set the authentication protocol to Universal (Unencrypted) mode. After completing the above parameter configurations, initiate the compile-deploy automation process via the Start Debugging command.

This mechanism synchronously executes the generation of application binary files, their transmission, and device-side loading, establishing a complete end-to-end deployment pipeline. Through the above workflow, the Unity-based 3D model of the case study Bridge was successfully deployed on HoloLens 2, achieving 3D visualization as demonstrated in Figure 3.

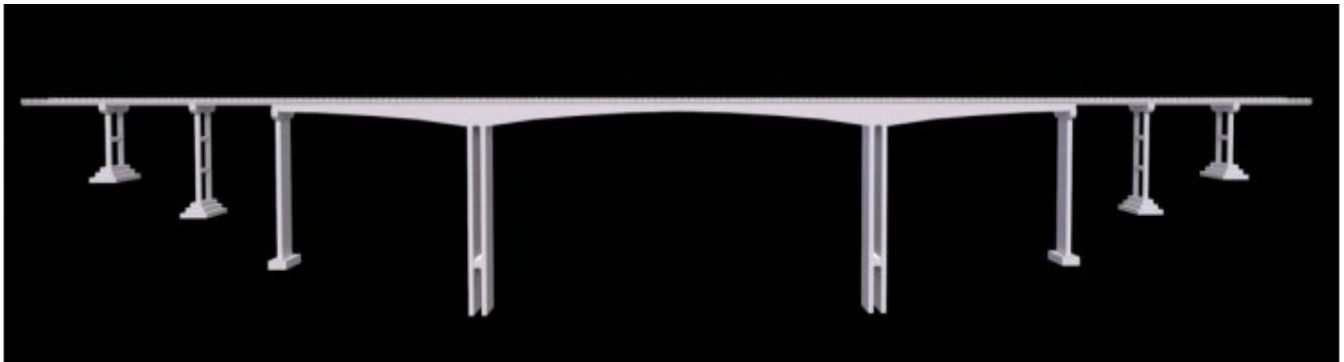


Figure 3. 3D Model layout diagram of case study Bridge.

To accurately simulate defect morphologies under real-world conditions and enable interactive analysis of damage data within the HoloLens 2 digital model, high-fidelity defect simulations were integrated into the case study bridge model. Leveraging HoloLens 2's augmented reality system, 3D defect visualization and spatial mapping were implemented [10]. This framework, combined with depth perception and multi-source data fusion algorithms, supports three critical technical requirements: on-site auxiliary defect diagnosis, remote collaborative structural assessment by experts, and dynamic coupling analysis of multidimensional human-machine interaction

3 DATASET CONFIGURATION

This work referenced and adapted the data collection standards of the VisDrone2021 [11] dataset, employing methods including web crawling, video frame extraction, and filtering of multiple publicly available bridge defect datasets [12]. Simultaneously, to enhance the model's generalization capability and prevent overfitting, data augmentation techniques—such as random rotation [13], random occlusion, color jittering, Gaussian blur, and noise addition—were applied to the dataset [14], as shown in Figure 4.



Figure 4. Defects on the bridge.

The experimental dataset for this study was ultimately compiled and generated, containing 4,176 images. Label files for the dataset were created using the labeling tool, with a total of 32,201 annotated bounding boxes. These annotations encompass bridge defect information across multiple categories, including exposed reinforcement, spalling, corrosion, water seepage in wet joints, mold growth, and cracks, as illustrated in Figure 5. The dataset aligns with the simulated defect conditions previously generated on the Unity bridge model.



Figure 5. Spalling & Exposed reinforcement.

Distinct from traditional bridge defect datasets that focus on one or several common types of surface-level defects, the self-constructed dataset employed in this study—"Small-Target Detection Dataset for Multi-Category Bridge Defects under Complex Background Interference"—simultaneously incorporates 12 bridge defect types, including those with low occurrence frequencies. To replicate real-world inspection scenarios and simulate the scanning and observational perspectives of inspectors wearing HoloLens 2 for subsequent comparative analysis and digital model superimposition tasks, no image or semantic segmentation [15] is performed on bridge defects in this dataset, preserving full panoramic small-target detection and recognition. Furthermore, unlike conventional small-target datasets, bridge defect images in real-world scenarios exhibit challenges such as blurred backgrounds, high inter-defect similarity, strong deceptive features, and extremely

small defect targets. These characteristics lead to core technical detection challenges, including weak multi-scale target sensitivity, significant complex background interference, insufficient fine-grained feature representation, and high inter-category similarity. Through the aforementioned diverse image augmentation methods, the dataset is further enriched and expanded to enhance robustness. The highly challenging "Small-Target Detection Dataset for Multi-Category Bridge Defects under Complex Background Interference" imposes greater demands on subsequent target detection and recognition tasks.

4 YOLOV11 ALGORITHM

4.1 YOLOv11 Model Introduction and Advantages

Ultralytics recently released YOLOv11, designed as a detection model achieving state-of-the-art (SOTA) performance across multiple tasks. The architecture of previous models has been optimized, enabling YOLOv11 to attain cutting-edge

performance in diverse tasks (object detection, segmentation, pose estimation). The overall network architecture is illustrated in Figure 6.

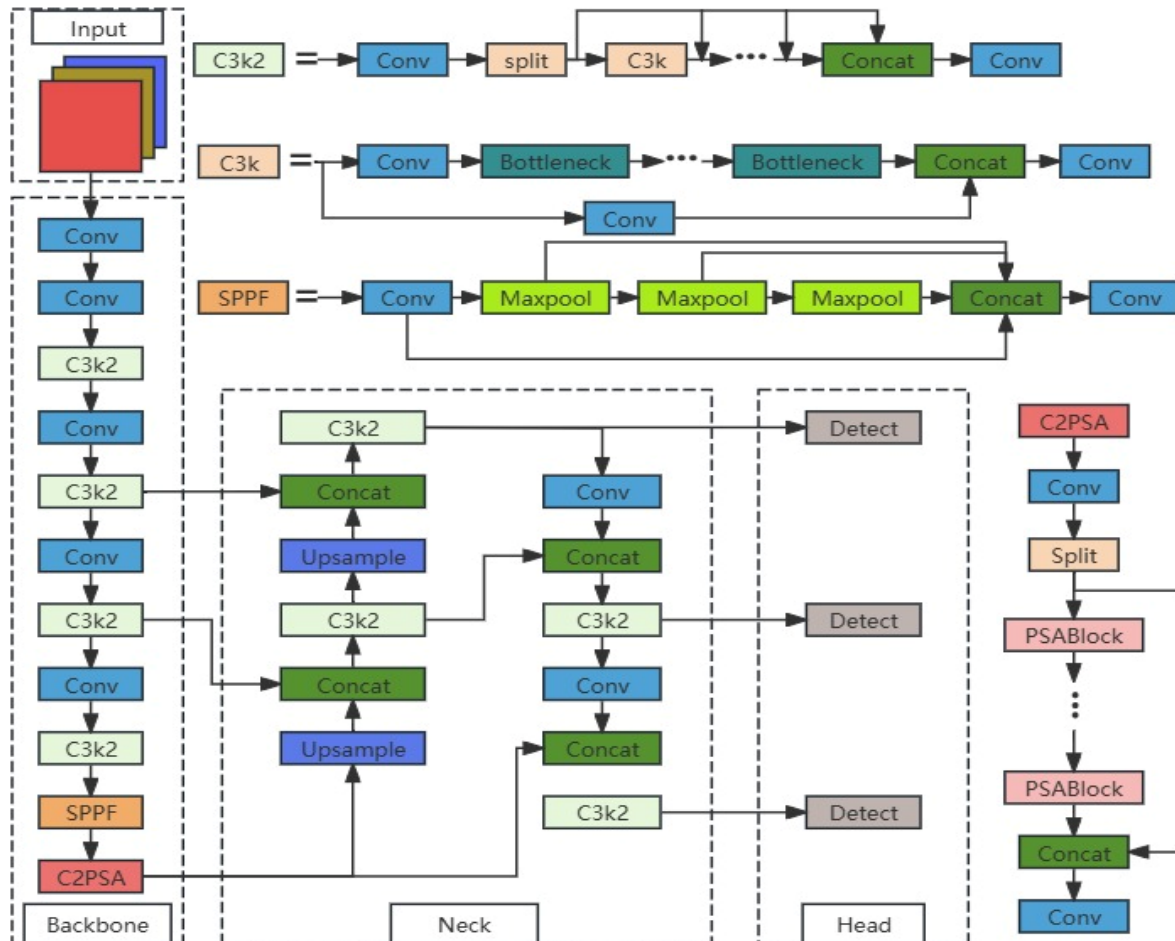


Figure 6. Overall network architecture diagram of YOLOv11.

Compared to YOLOv8, YOLOv11 reduces parameters by 22% on the COCO dataset while achieving higher mean Average Precision (mAP), as shown in Figure 7. Simultaneously, its inference speed is approximately 2% faster

than YOLOv10, reaching 60 frames per second (FPS), making it one of the fastest object detection models and providing enhanced support for real-time applications.

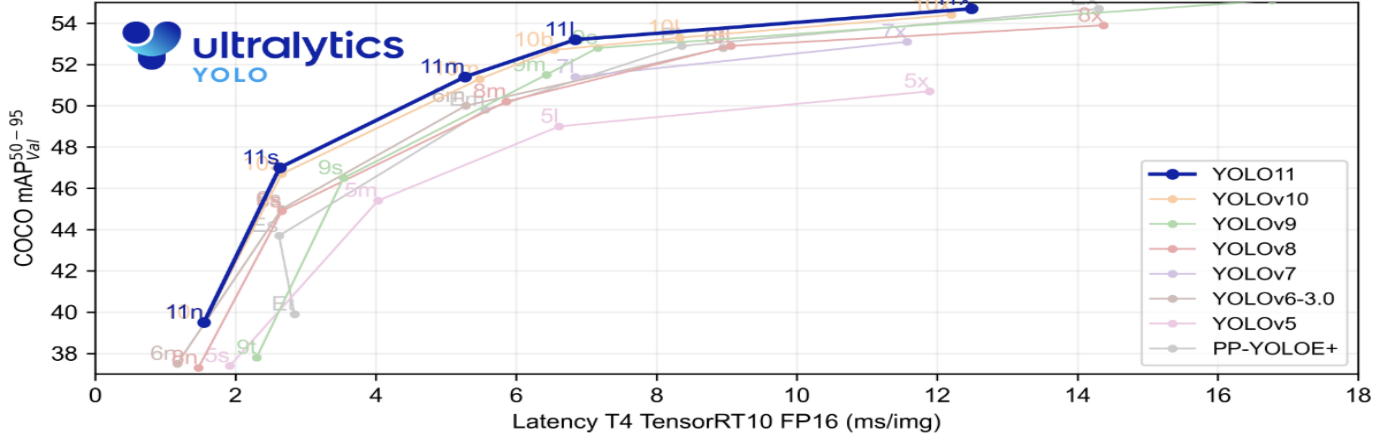


Figure 7. Performance comparison diagram between YOLOv11 and previous versions.

This study innovatively proposes the YOLOv11n-ZY model, an optimized and improved framework based on YOLOv11, to address core technical challenges in bridge defect detection tasks, including weak multi-scale target sensitivity, significant complex background interference, and insufficient fine-grained feature representation.

By deeply integrating the new ZZ convolutional module, new YY attention mechanism module, and SPPF-LSKA fusion module, a collaborative optimization system is constructed. The model adopts a hierarchical feature processing architecture, embedding three innovative modules into the feature extraction layer, attention enhancement layer, and multi-scale fusion layer, respectively, thereby establishing a complete technical chain from microscopic feature analysis to macroscopic semantic correlation. The three modules achieve compatible deep integration through granular allocation of computational resources and functional positioning: The ZZ module focuses on enhancing multi-granularity extraction efficiency of low-level features. The YY module implements domain-adaptive calibration during feature transmission. The SPPF-LSKA module accomplishes complementary fusion of multi-level semantic features. These three components collectively establish a progressive optimization pathway of “feature encoding → attention enhancement → pyramid fusion.”

At the parameter optimization level, the three modules respectively introduce learnable convolutional kernel scale ratios [17], dynamic attention weights [18], and large-kernel decoupled computation mechanisms. Through joint backpropagation, these components synergistically optimize the detection loss function. This hierarchically deployed collaborative paradigm provides a technical solution that combines theoretical innovation and engineering value for intelligent bridge defect detection.

4.2 Simulation Environment

The simulation was conducted on a Windows 11 operating system with 3T memory and an RTX 4070 GPU (64GB VRAM). Python 3.8 was utilized, with the PyTorch framework (Torch 1.12.0 version). The YOLOv11n-ZY model was trained

for 200 epochs with a batch size of 32 and a learning rate of 0.01.

To evaluate the comprehensive performance of the model, this experiment adopts the following quantitative evaluation metrics: parameter count, precision (P), recall (R), mean average precision (mAP), frame rate (FPS), and F1-score. These metrics collectively characterize the model’s robustness across detection accuracy, computational efficiency, and multiple confidence thresholds [16].

The mathematical definitions of precision and recall are given in Equations (1) and (2), where TP (True Positives) denotes the number of correctly detected positive samples, FP (False Positives) represents the number of positive samples incorrectly classified as negative, and FN (False Negatives) indicates the number of undetected positive samples. The F1-score is defined by Equation (3), which is essentially the harmonic mean of precision and recall. In the field of object detection, mAP serves as a core evaluation metric, quantified through the weighted average of precision values across confidence thresholds, with its computational methodology detailed in Equations (4) and (5).

$$P = \frac{TP}{TP+FP} \quad (1)$$

$$R = \frac{TP}{TP+FN} \quad (2)$$

$$F_1 = 2 \cdot \frac{P \cdot R}{P+R} \quad (3)$$

$$AP = \int_0^1 p(r) dr \quad (4)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP(i) \times 100\% \quad (5)$$

In the equations, the mAP reflects the balance between the detection precision and recall of the model across all categories (unit: %); mAP50 refers to the mAP value when the IoU is 0.5; AP_i is the average precision of the *i*-th category; and *n* is the total number of categories.

4.2.1 Comparison Experiments

To further validate the detection performance of the improved YOLOv11n-ZY model proposed in this paper, the algorithm is compared and analyzed with common algorithms in this field, and the results are shown in Table 2.

Table 2. Comparative Experimental Results.

Model	mAP50/%	P/%	R/%	Params/10 ⁶
YOLOv5s	33.8	37.0	39.7	7.82
YOLOv8n	33.4	40.5	36.7	3.01
YOLOv8s	35.1	40.7	39.1	11.13
YOLOv11n	34.9	44.6	37.8	2.58
YOLOv11s	36.3	43.0	41.1	9.41
Ours	40.3	49.3	41.1	2.87

According to the experimental results in Table 2, compared to the detection results of previous official Yolo series models (Yolov5s, Yolov8n, Yolov8s), the proposed Yolov11n-ZY algorithm in this study achieves significant improvements in mAP50 by 6.5%, 6.9%, and 5.2%, while reducing parameter sizes by 4.95 MB, 0.14 MB, and 8.26 MB, respectively. Compared to the baseline model YOLOv11n and its series counterpart YOLOv11s, the Yolov11n-ZY algorithm achieves accuracy improvements in mAP50 of 5.4% and 4%, respectively, despite a slight increase in parameter size. In summary, the proposed algorithm outperforms other methods in the accuracy of "multi-category bridge defect small target detection under complex background interference," including mAP50, precision (P), and recall (R), while maintaining real-time performance and achieving an optimal balance in model size.

5 VISUALIZATION INTERFACE DESIGN

The visual interactive interface of the proposed detection and recognition system integrates the YOLOv11n-ZY model for bridge defect detection and recognition, featuring multi-modal input source processing capabilities (including static images, video streams, real-time camera capture, and batch file processing). A multi-threaded parallel processing mechanism is adopted to ensure real-time responsiveness of the human-machine interaction interface. Detection results are visualized in real-time through the graphical interface, with dynamic parameter adjustment functions (confidence threshold, IoU threshold) and detection process control interfaces (start, pause, terminate detection, and result storage). The specific visual interface system is shown in Figure 8.

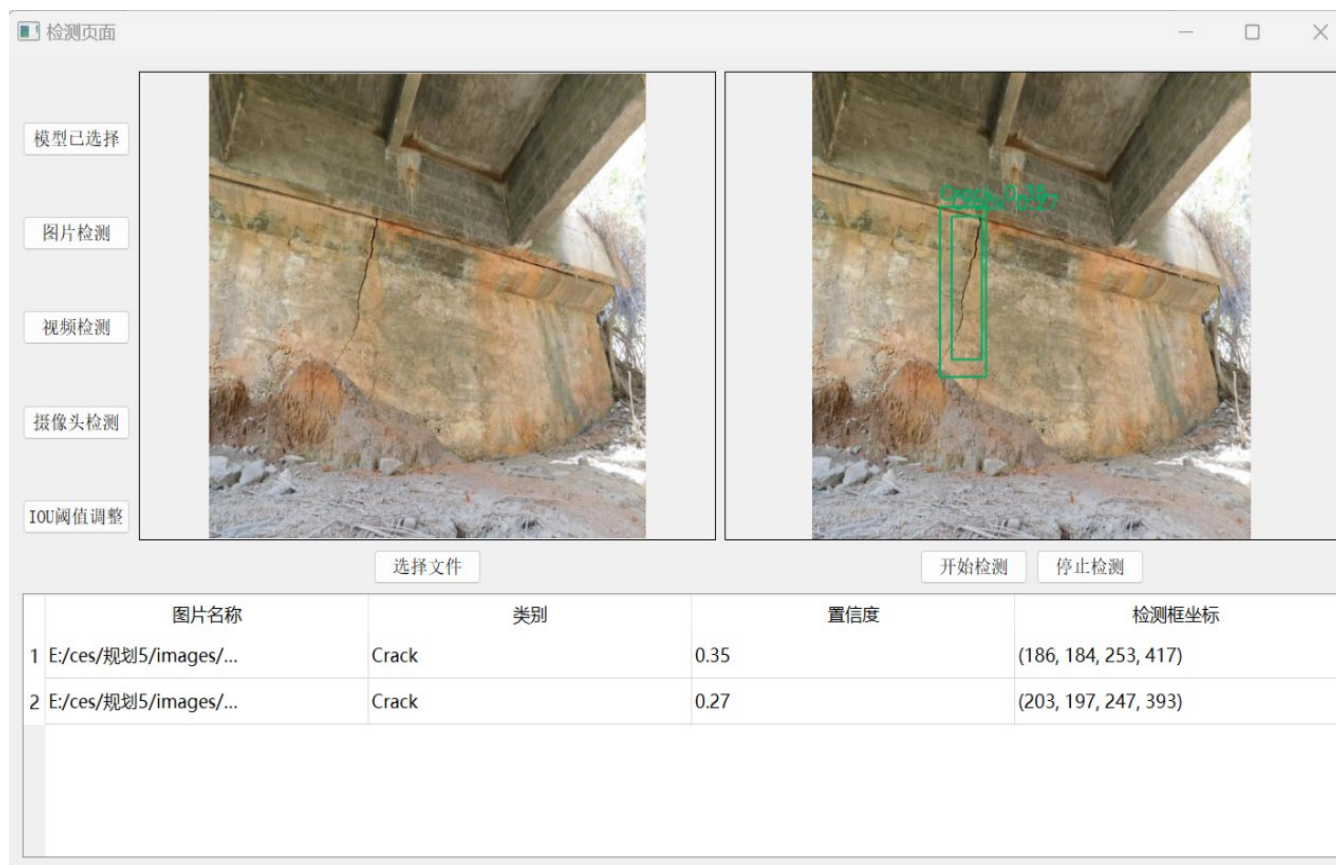


Figure 8. Visualization interface for bridge defect object detection.

In the input source selection unit, specify the camera capture device or local file path; dynamically configure the confidence threshold (Conf) and Intersection over Union threshold (IoU) of the target detection model through slider controls. The background processing thread is automatically initialized upon configuration completion.

The system routes the processing results to the main thread, transmitting bounding box-annotated detection images via

the (*send_detect_img*) signal and category statistics via the (*send_detect_info*) signal. Finally, the streaming inference approach governs the actual inference loop, as illustrated in Figure 9, which includes: reading input sources (images, videos, camera streams, etc.), preprocessing, executing model inference, processing post-inference results, and transmitting final recognition outcomes to the visualization main interface through signals.

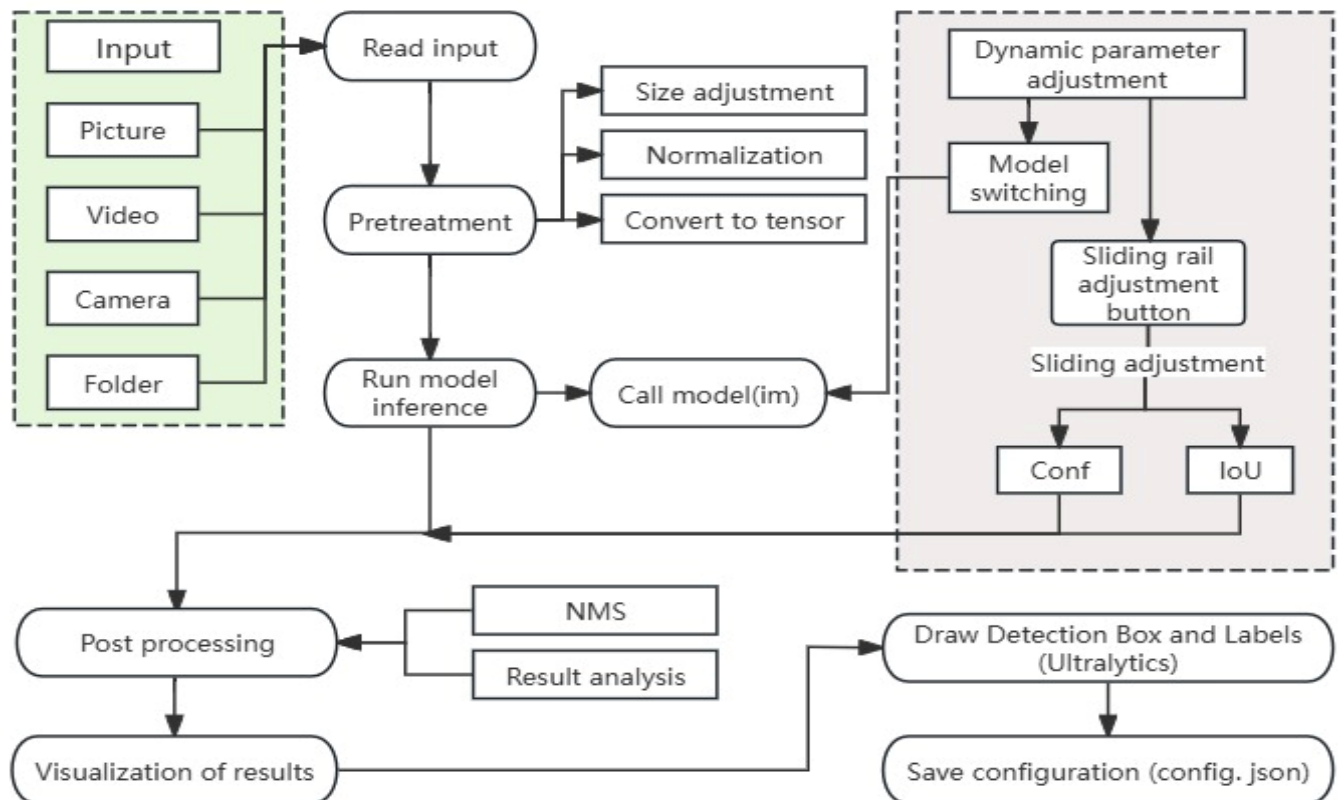


Figure 9. Flowchart of the visualization interface system operation.

6 CONCLUSION

Aiming at the bottleneck problems of low efficiency and insufficient accuracy in traditional bridge inspection methods, this paper proposes a "Comprehensive Visual Intelligent Bridge Inspection Platform" that integrates Mixed Reality (MR) technology, deep learning algorithms, and domain-specific bridge engineering knowledge. By leveraging the high-precision spatial perception capabilities of the HoloLens 2 device, the optimized architecture of the YOLOv11n-ZY model, and multimodal human-computer interaction technologies, a bridge defect detection system with real-time inspection, dynamic visualization, and full lifecycle management functions has been successfully established. Experimental results demonstrate that the improved YOLOv11n-ZY model significantly outperforms existing mainstream algorithms in detecting multi-category small-target defects under complex background interference, achieving a detection precision (mAP50 of 40.3%). Simultaneously, the MR technology facilitates closed-loop management of defect localization, remote collaboration, and dynamic 3D model updating. This study not only provides an efficient and reliable technical pathway for intelligent bridge operation and maintenance but also offers theoretical support and practical exemplars for the deep integration of mixed reality and deep learning in infrastructure inspection. Future work will focus on optimizing the lightweight deployment capability of the model, further enhancing the precision of the algorithm, expanding multi-source sensor data fusion mechanisms, and exploring a

digital twin-based predictive system for bridge performance throughout its lifecycle to advance the engineering application and standardized development of intelligent infrastructure inspection technologies.

ACKNOWLEDGMENTS

This study is supported by 'Data-Driven Intelligent Assessment of Bridge Infrastructure' granted by the Slovak Research and Development Agency under the contract No. SK-CN-23-0037.

REFERENCES

- [1] Spencer Jr, Billie F., Vedhus Hoskere, and Yasutaka Narazaki. "Advances in computer vision-based civil infrastructure inspection and monitoring." *Engineering* 5.2 (2019): 199-222.
- [2] Feng, D. C., Wang, W. J., Mangalathu, S., & Sun, Z. (2023). Condition Assessment of Highway Bridges Using Textual Data and Natural Language Processing-(NLP-) Based Machine Learning Models. *Structural Control and Health Monitoring*, 2023(1), 9761154.
- [3] Zhang, Yongjian , X. Chen , and W. Yan . "Automated intelligent detection system for bridge damages with Fractal-features-based improved YOLOv7." *Signal, Image and Video Processing* 19.3(2025).
- [4] Hsieh, Chang Cheng , H. M. Chen , and S. K. Wang . "On-site Visual Construction Management System Based on the Integration of SLAM-based AR and BIM on a Handheld Device." *KSCE Journal of Civil Engineering* 27.11(2023):20.
- [5] Malek, Kaveh, Ali Mohammadkhorasani, and Fernando Moreu. "Methodology to integrate augmented reality and pattern recognition for crack detection." *Computer-Aided Civil and Infrastructure Engineering* 38.8 (2023): 1000-1019.

- [6] Sun, Z., Caetano, E., Pereira, S., & Moutinho, C. (2023). Employing histogram of oriented gradient to enhance concrete crack detection performance with classification algorithm and Bayesian optimization. *Engineering Failure Analysis*, 150, 107351.
- [7] Kun, J., Zhenhai, Z., Jiale, Y., & Jianwu, D. . A deep learning-based method for pixel-level crack detection on concrete bridges. *IET Image Processing*.
- [8] Mascarenas, David DL, et al. "Augmented reality for next generation infrastructure inspections." *Structural Health Monitoring 20.4* (2021): 1957-1979.
- [9] Wang, Ying , et al. "Could I Have a Stack Trace to Examine the Dependency Conflict Issue?." *ACM* (2019).
- [10] Sahraie, A. , et al. "Psychophysical and pupillometric study of spatial channels of visual processing in blindsight." *Experimental Brain Research* 143.2(2002):249.
- [11] Zhu, Pengfei , et al. "Vision Meets Drones: A Challenge." *Springer, Cham* (2018).
- [12] Flotzinger, Johannes , P. J. Rsch , and T. Braml . "dac110k: Benchmark for Semantic Bridge Damage Segmentation." *IEEE* (2023).
- [13] Liu, Yu , et al. "Rotating Target Detection Method of Concrete Bridge Crack Based on YOLO v5." *Applied Sciences* (2076-3417) 13.20(2023).
- [14] LinXinrui, et al. "CCTSDDB dataset enhancement based on a cross-augmentation method for image datasets." *Intelligent Data Analysis* (2024).
- [15] Jin, T. , X. Ye , and Z. X. Li . "Establishment and evaluation of conditional GAN-based image
- [16] Peng, Shifeng , et al. "PS-YOLO: a small object detector based on efficient convolution and multi-scale feature fusion." *Multimedia Systems* 30.5(2024):1-16.
- [17] Wang, Hao , et al. "Multi-scale Location-aware Kernel Representation for Object Detection." *IEEE* (2018).
- [18] Jiang, Zhichao , D. Liu , and L. Cui . "Deep adaptively dynamic edge graph convolution network with attention weight and high-dimension affinity feature graph for rotating machinery fault diagnosis." *IOP Publishing Ltd* (2024).