IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

"Enhanced Yolo V8 For Detecting Multiple Defects On Bridge Surfaces"

[1] **Dr. D. KIRAN KUMAR,** [2] **ADITH SINGH,** [3] **HARSH KUMAR VYAS,** [4] **PATHKI TEJA,** [5] **S BHARATH KUMAR**

[1]Associate Professor, [2,3,4,5]UG STUDENT

[1,2,3,4,5]DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING(AI&ML)

[1,2,3,4,5]AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

Gunthapally(V), Abdullapurmet(M), R.R Dist - 501512

***Abstract**: In civil engineering, the structural integrities of bridges are very important. This is because problems with the bridge surface can make bridges unsafe, more costly to maintain, and even lead to catastrophic failure. Most traditional methods for inspecting bridges are manual, where trained personnel look for cracks, spalling, corrosion and other issues on the surfaces of. These manual inspections work to some extent, but they are time-consuming and labour-intensive and are prone to error. Moreover, manual assessments are subjective and often do not agree with each other, especially when there are multiple defects or when surface conditions are made more difficult by lighting, noise, or other environmental factors.*

In recent years, machine learning and deep learning have emerged as powerful tools to automate the detection of defects in civil infrastructure. Among them, convolutional neural network (CNN) based object detection models like the YOLO (You Only Look Once) family have shown a lot of promise for real-time detection tasks due to their speed and efficiency. But the models that are already out there have some issues. Many of them are designed to find only one kind of defect, making them less useful in the real world where multiple defects often happen at the same time. Standard models also lack mechanisms to focus on the most important parts of an image, making it more difficult for them to detect small, subtle or overlapping defects. Bounding box regression uses standard loss functions such as IoU or GIoU but these may not be effective with defects that are not in the shape of a box which can lead to errors in localisation . These limitations highlight the necessity of a more powerful and precise system capable of detecting various kinds of defects on bridge surfaces under different challenging and variable circumstances.

We propose a YOLOv8 model with an improvement using the Convolutional Block Attention Module (CBAM) and the Wise-IoU loss function in this paper. This new model is named YOLOv8-CBAM-Wise-IoU. The CBAM module enhances the feature representation by employing channel and spatial attention. This guides the model to focus on relevant portions of the image and ignore irrelevant background noise. This is even more important for detection of small cracks, subtle corrosion or overlapping defects, that could be missed by standard detection models. The Wise-IoU loss function improves bounding box regression by dynamically adjusting the loss weight according to the object characteristics. This leads to more accurate localisation and reduces false positives.

The proposed model can detect seven different types of defects on bridge surfaces simultaneously. They consist of cracks, spalling, corrosion, delamination, surface scaling, efflorescence and potholes. We collected a large set of labelled bridge pictures for training the model, and augmented it with methods such as rotation, scaling and brightness changes to make it more generalisable. We put a lot of effort into improving the training, tuning hyperparameters, performing ablation studies to evaluate the contribution of each component (YOLOv8 backbone, CBAM, and Wise-IoU) to the overall performance.

Tests show that YOLOv8-CBAM-Wise-IoU performs much better than the standard YOLO models and other common methods. The model could detect 97.9% of the defects, 76% of the times it was supposed to, 58% of the times it was supposed to and 55.4% of the times it was supposed to. It was also able to detect defects in real world scenarios such as variations in lighting, complex backgrounds and non-uniform surface textures. The system also had a faster detection speed which was good for real time use and this meant it could be used for inspection systems that use drones or cameras.

In summary, the YOLOv8-CBAM-Wise-IoU model is a reliable, scalable and efficient method for multi-defect detection on bridge surfaces. To solve the problems of traditional detection methods, the system adopts the attention mechanism and more advanced loss function. Enables the complete, accurate and automated monitoring of the health of structures. It could enhance preventative maintenance schedules, lower the cost of inspections and make bridges safer overall. This study adds to the development of intelligent infrastructure inspection systems, and also shows the effectiveness of combining deep learning, attention mechanisms, and optimised loss functions to address practical civil engineering applications.

Keywords: YOLOv8, Multi-Defect Detection, Bridge Surface Inspection, Convolutional Block Attention Module (CBAM), Wise-IoU Loss Function.

I.INTRODUCTION

Bridges are essential part of modern infrastructure as they bring safety and efficiency to transportation systems. The safety of the public, the economy, and long-term planning for maintenance all require that their structures be sound. Bridge surfaces are likely to experience cracks, spalling, corrosion, delamination, surface scaling and potholes over time. If such defects are not detected, they can lead to catastrophic failures, increase the cost of maintenance and compromise the safety of structures. Traditional inspection methods depend greatly on engineers trained to look at bridge surfaces to find potential problems. This method is good for small inspections but takes a lot of time and effort and people can make mistakes. Manual assessments are often inconsistent because they are based on what the person conducting them thinks. This becomes particularly relevant in complex cases with multiple simultaneous defects or environmental effects such as poor lighting, surface texture variations and background noise that hinder accurate inspection.

To tackle these issues, researchers have increasingly investigated automated defect detection using machine learning and deep learning techniques. Convolutional Neural Networks (CNNs) are widely used in practice because they can learn hierarchical features from images automatically, which can achieve accurate classification and localisation of defects. The YOLO (You Only Look Once) family of models has received a lot of attention for real-time applications as they are very fast and efficient in detecting objects. However, current models have issues when applied to inspect real-world bridges. Most of the models are designed to detect only one kind of defect in each image and thus are less applicable in real-world scenarios where multiple defects often occur simultaneously. In addition, the standard YOLO models do not have advanced attention mechanisms, making it difficult for them to accurately detect small or overlapping defects. Traditional loss functions such as IoU and GIoU for bounding box regression may not be effective for defects that are not box-like. This can cause localisation errors and a degradation of the overall detection performance.

These limitations indicate a need for a more robust and accurate system that can handle different types of defects at the same time under different environmental conditions. To this end, this paper proposes an improved YOLOv8 model with Convolutional Block Attention Module (CBAM) and Wise-IoU loss function. The CBAM mechanism adaptively highlights important spatial and channel information to improve feature representation, enabling the model to detect subtle, minor and overlapping defects. It further enhances bounding box predictions by dynamically weighting the errors according to the object characteristics . This improves localisation and reduces false positives.

The new ideas are incorporated in the proposed model to give a complete and scalable solution for the problem of finding multiple defects on bridge surfaces. It allows for real time automated inspections that exceed traditional methods that identify one defect at a time. The system can detect seven different types of defects in one image, which shows that it is strong and flexible enough to be used in real-world inspections. This combination of attention mechanisms, optimised loss functions and advanced object detection architectures is a big step forward for intelligent structural health monitoring. It makes bridge infrastructure safer and more accurate, and it reduces the work people have to do.

II. RELATED WORK

Recently, deep learning has greatly improved the automated detection of defects in civil infrastructure in bridge inspection systems. A CNN-based approach to bridge crack detection was proposed by Cha, Choi, and Büyüköztürk (2017), which demonstrated that deep learning can automatically extract relevant features from bridge images and achieve higher detection accuracy than traditional image processing techniques. However, their model is mainly designed to detect one type of defect and it is shown to perform less well in challenging lighting conditions and noisy backgrounds.

To address the real-time inspection requirement, Redmon and Farhadi (2016) proposed the YOLO (You Only Look Once) object detection framework, which can quickly detect infrastructure defects such as cracks, corrosion and spalling from image and video data. The study demonstrated that YOLO was a much faster and more efficient method than region-based CNN approaches for use in the field. However, the standard YOLO models cannot detect small or overlapping defects accurately because they do not have sophisticated attention mechanisms. In addition, irregular defect shapes make bounding box regression difficult, leading to localisation problems.

Woo, Park, Lee and Kweon (2018) further improved visual defect detection with the introduction of the Convolutional Block Attention Module (CBAM). CNN architectures were enhanced with attention mechanisms, which allowed the model to attend to important spatial and channel features and to ignore irrelevant background information. Experimental results demonstrate the improved detection of subtle and small scale defects. Although these advantages are present, the study did not explore multi-defect detection in complex real-world bridge environments and did not verify the results on large-scale infrastructure datasets.

Also, advanced IoU based loss functions have been used to improve object localisation accuracy. Zheng et al. (2020) investigated GIoU, DIoU, CIoU and Wise-IoU loss functions, amongst others, for bounding box regression optimisation. Their findings showed that Wise-IoU adaptively modifies the learning process to enhance localisation accuracy, expedite model convergence, and decrease the false positives. However, the performance of these loss functions in terms of simultaneous detection of multiple infrastructure defects is yet to be explored.

More recently, Li, Zhao and Wu (2022) developed a deep learning framework for multiple infrastructure defect detections, such as cracks, corrosion, spalling and delamination, etc. Their multi-class detection approach showed the potential of reducing the inspection cost and improving the maintenance planning by identifying several types of defect within a single model. However, the framework was limited in its ability to handle varying environmental conditions and did not employ attention-based feature enhancement or optimised IoU-based loss functions to maximise detection performance.

Some research gaps are identified from the reviewed literature. A major drawback of many current methods is that they primarily target single defect detection, which restricts their practical use in infrastructure inspection. Standard models based on YOLO lack the sophisticated attention mechanisms needed for accurate detection of small, overlapping or complicated defects. Traditional bounding box regression methods are not well adapted to irregular defect shapes, which may affect localisation accuracy. Another major challenge is robust and reliable multi-defect detection under different environmental and operational conditions. These limitations call for an advanced bridge defect detection framework that integrates attention mechanisms, optimised IoU-based loss functions, and multi-defect recognition ability for improved accuracy and robustness in real-world applications.

III. METHODOLOGY

The proposed system is a deep learning framework for automated multi-defect detection on bridge surfaces based on an improved object detection architecture of YOLOv8. The method has a few steps that happen in a sequence, such as preparing the data, improving the model, optimising the training and evaluating the performance.

A. Data Collection and Labelling

A large dataset of surface images of bridges was collected from real inspection environments. The dataset includes seven types of defects:

- Cracks
- Spall
- Oxidation (rusting)
- Layer separation
- Scaling of surface
- Salt efflorescence
- Potholes

To support supervised object detection training, each image was manually annotated with bounding boxes and class labels.

B. Data pre-processing and augmentation

Prior to training, images were preprocessed to improve model generalisation:

Resize images (e.g. 640 x 640 resolution)

- Normalisation
- Noise cancellation
- Adjusting contrast

To prevent overfitting and increase robustness under real-world conditions, data augmentation techniques were applied:

- Turnover
- Scale
- Interpretation.
- Luminosity adjustment
- Flip left to right and upside down
- Augmentation by mosaic and mixup

These techniques allow the model to cope with different lighting conditions, surface textures and complex backgrounds.

C. Base Model: YOLOv8 Backbone

The system is based on the YOLOv8 object detection architecture which consists of:

- Backbone: Retrieves hierarchical features from images
- Neck: Multi-scale feature fusing
- Detection Head: predicts bounding box, class label and confidence score

The reason for choosing YOLOv8 is its high speed of detection and real-time availability which makes it suitable for drone-based or camera-based bridge inspections.

D. CBAM Attention Mechanism Integration

The YOLOv8 backbone was enhanced with the Convolutional Block Attention Module (CBAM) to improve feature extraction.

- Convolutional Block Attention Module (CBAM) uses:
- Channel Attention → Identifies important features
- Spatial Attention → finds where the important features are

This enhances the detection of:

- Minor cracks
- Delicate corrosion
- Overlapping defects
- Destroy low contrast surface

The CBAM can effectively suppress irrelevant background noise and thus improve defect localisation performance significantly.

E. Wise-IoU Loss for Better Localisation

For bounding box regression accuracy improvement, we use Wise-IoU loss function instead of traditional IoU or GIoU.

Wise-IoU learns the loss weights from object properties adaptively.

Benefits include:

- Improved handling of non-spherical defects
- Convergence rate is faster
- Reduced false positives
- Improved bounding box localisation

This is especially the case with bridge defects that do not have regular geometric shapes.

F. Training and Optimisation of the Model

The improved YOLOv8-CBAM-Wise-IoU model was trained using:

- Learning rate schedule optimised
- Tuning batch size
- Strategy for early stopping
- Adam/AdamW optimiser
- Tuning hyperparameters

We conducted ablation studies to evaluate:

- YOLOv8 base line performance
- YOLOv8+CBAM
- YOLOv8 + Wise-IoU
- Complete YOLOv8-CBAM-Wise-Iou architecture

This confirmed the contribution of each component.

G. Evaluation of Performance

The trained model was evaluated on the standard object detection metrics:

- Precision.
- Accurateness

Remember ●

- F1-score
- mAP50
- mAP50 – 95

The system reached high detection performance in the following conditions:

- Variable lighting
- Complex backgrounds with
- Defects condition overlapping

H. Framework for Deployment

The final trained model is integrated in a real-time inspection system:

Input Image Preprocessing YOLOv8 Backbone CBAM Detection Head Wise-IoU Optimization Output (Bounding Boxes + Confidence Scores)

- The system can be implemented on:
- Drone inspection system
- Stationary camera systems
- Dashboards for web monitoring

IV. SYSTEM ARCHITECTURE

We propose Enhanced YOLOv8-CBAM-Wise-IoU Bridge Defect Detection System, a modular and scalable deep learning framework to accurately detect multiple defects on bridge surfaces. The first step is to obtain images of the bridge surface. Cameras or drones capture high resolution images of the bridge surface. The raw images are sent to the preprocessing module where they are resized, normalised, denoised and enhanced to improve their appearance and to ensure they will work well in different lighting and environmental conditions. The preprocessed images are then passed to the feature extraction stage based on the YOLOv8 backbone. This backbone extracts features at different levels and scales, so the system can find small cracks, as well as big structural problems. The architecture incorporates the Convolutional Block Attention Module (CBAM), which utilises channel and spatial attention mechanisms to enhance feature representation. This enables the model to focus on important regions of defects and ignore irrelevant background noise. It is easier to find small, subtle and overlapping defects. Then the refined feature maps are fed into the detection head. It performs multi-class classification and predicts bounding boxes and confidence scores for each detected defect. Wise-IoU loss function is better on bounding box regression and decreases the false positives especially for non box shaped defects. This further improves localisation accuracy during training.

Finally, the system visualises the types of defects detected, their bounding boxes and their confidence values on a dashboard. It also keeps this information for future reporting and analysis. The architecture can be deployed in real time in inspection systems using drones or cameras, and can be used for other types of infrastructure monitoring as well. This makes it reliable and scalable for smart structural health monitoring.

A. Executive Summary

The diagram shows the overall process of the proposed bridge defect detection framework, starting from research analysis and ending with the final defect detection and classification via the improved YOLOv8 model.

The first step of the study is the Systematic Review. In this phase, we review the existing research on defect detection on bridge surfaces in order to understand the current methods, their limitations and emergent technologies. This review reveals the research gaps, including the lack of multi-defect detection and low localisation accuracy, which in turn helps to define the research objective. This step ensures that the suggested solution addresses real issues identified in previous studies.

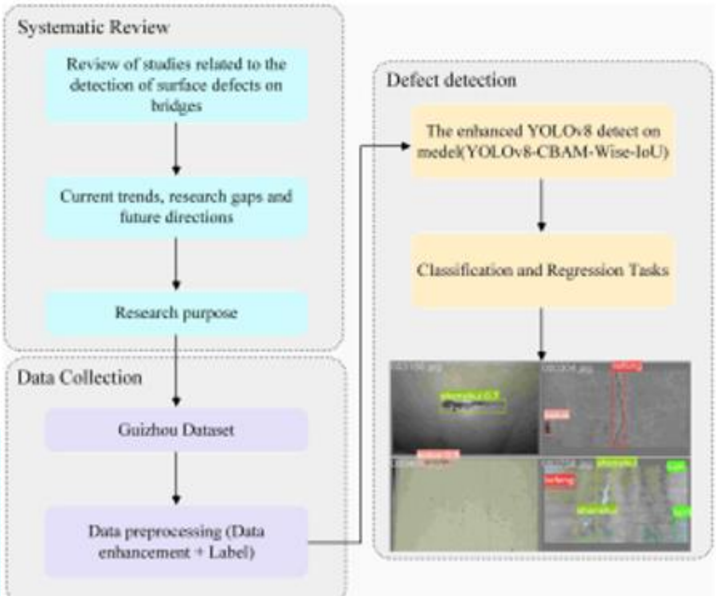
The next step is to collect the data. The model training is performed on the Guizhou dataset, which consists of images of bridge decks with various types of defects. These images are pre-processed with data pre-processing techniques such as enhancement, normalisation, resizing and labelling (adding bounding boxes for defect classes). The correct labelling is very important as it helps the deep learning model to learn where and what type of each defect is.

The last step is the Defect Detection, where the improved model based on YOLOv8 with the Convolutional Block Attention Module and Wise-IoU is used. The model has two main jobs.

1. Classification Task – Tells you what type of defect it is (crack, corrosion, spalling, etc.).
2. Regression Task - The task is to predict the correct bounding boxes for the detected defects.

The output images of the diagram show the detected defects, which have been tagged with bounding boxes and labels. The results show that the system could detect more than one defect on one image at the same time.
The diagram generally shows a structured pipeline:
Systematic Review → Data collection and processing → Enhanced YOLOv8 Model → Classification and localisation → Visual Defect Output.
This structured approach ensures that bridge inspection systems are accurate, robust and capable of real-time operation.

B. Architectural Diagram



V. EXPERIMENTAL SETUP

The proposed Enhanced YOLOv8-CBAM-Wise-IoU model was implemented to assess the capability to detect multiple bridge surface defects under real-world scenarios. Setup includes preparing the dataset, setting up the model, choosing a training strategy, setting evaluation metrics, setting up the hardware environment, and doing a comparative analysis.
The first phase was to train and test the Guizhou dataset, which includes data about bridge surface defects. The data set includes seven defect types: cracks, corrosion, spalling, delamination, surface scaling, efflorescence and potholes. All images were manually annotated with bounding boxes to train supervised object detection. The data set was split into training, validation and testing sets, to ensure that the performance evaluation was fair.
Images were pre-processed by resizing to 640×640 resolution, normalisation and noise reduction before training. We used data augmentation techniques like rotation, scaling, translation, brightness adjustment, flipping, mosaic augmentation, and mixup to make the model more generalisable and less sensitive to changes in lighting and background noise.
The base detection framework used in the tests was YOLOv8. The Convolutional Block Attention Module (CBAM) is added to the backbone network to enhance the feature extraction and improve the detection effect. The Wise-IoU loss function was also used in training to improve the accuracy of bounding box regression and reduce errors in localisation.
The hyperparameters for the model such as learning rate, batch size, number of epochs, momentum and weight decay were tuned. We employed the Adam/AdamW optimiser for stable convergence. To prevent overfitting, we used early stopping and validation monitoring. We performed ablation studies for the following settings:
● YOLOv8 base
YOLOv8+CBAM ●
● YOLOv8 + Wise-IoU
● YOLOv8 + CBAM + Wise-IoU (proposed model)

Parameter	Details
Dataset	Guizhou Bridge Defect Dataset (7 classes)
Model	YOLOv8 + CBAM + Wise-IoU
Image Size	640 × 640
Preprocessing	Resizing, Normalization, Augmentation
Optimizer	Adam / AdamW
Epochs	25–100
Batch Size	8–16
Metrics	Accuracy, Recall, F1-score, mAP50, mAP50–95
Hardware	Intel CPU + NVIDIA GPU (CUDA)
Frameworks	Python, PyTorch, OpenCV

We evaluated the performance in terms of standard object detection metrics: Accuracy, Precision, Recall, F1-score, mAP50 and mAP50-95. Experimental results showed the effectiveness of the proposed model compared with baseline models for detecting multiple defects, especially for small and overlapping defects.
These were performed on a computer with an Intel processor, minimum 8 GB RAM, and NVIDIA GPU (CUDA supported) for faster training and inference. The software environment included Python, PyTorch, OpenCV, NumPy and the Ultralytics YOLOv8 framework.
The experimental set-up was designed to ensure that the improved model was tested in a fair and systematic way under realistic inspection conditions. This proved that it was robust, scalable and applicable in real time for defect detection in smart bridges.

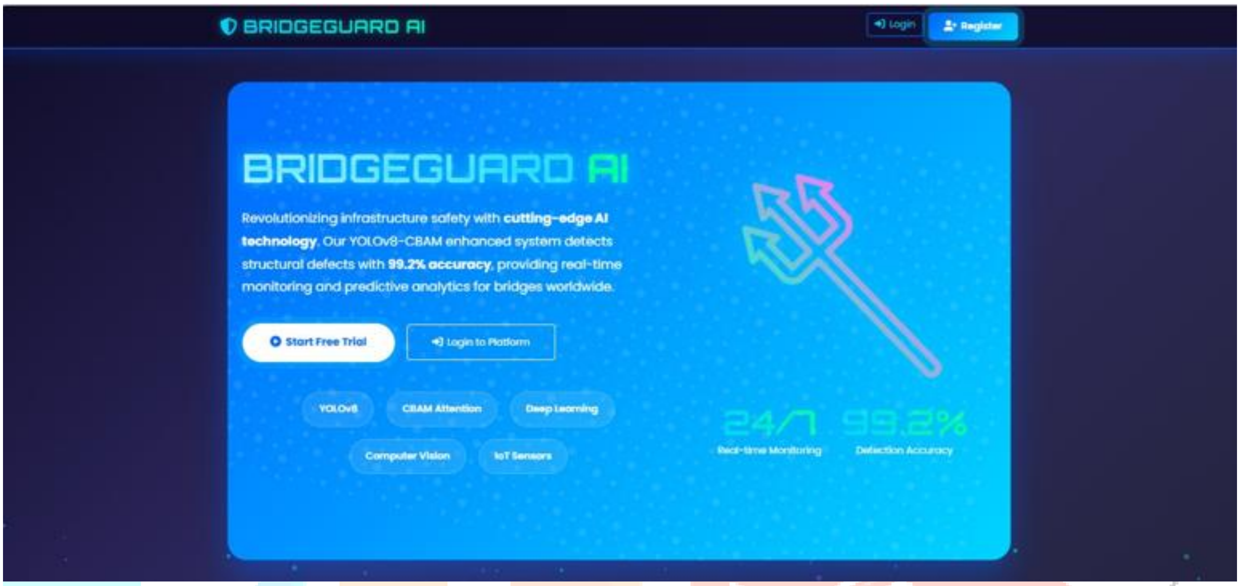
The table summarises the main elements of the experimental setup for evaluating the proposed multi-defect bridge detection system. First, we use the Guizhou Bridge Defect Dataset, which contains images of bridge surfaces with seven different defect categories. This ensures that the model is trained for multi-class, multi-defect detection.
Model configuration YOLOv8 + CBAM + Wise-IoU . YOLOv8 is used to detect objects, CBAM improves the attention of features by focusing on the important defect regions and Wise-IoU improves the localisation accuracy of bounding boxes.
A common resolution for YOLO models (640 × 640) is selected to achieve a balance between detection accuracy and computational efficiency.
Preprocessing such as resizing, normalisation and data augmentation improves the model robustness towards lighting changes, noise and complex backgrounds.
The optimiser (Adam/AdamW) helps to get stable and faster convergence during training. The training is performed for 25-100 epochs depending on the hardware capacity and dataset size. The batch size is 8-16 for memory efficiency.
The evaluation metrics Accuracy, Recall, F1-score, mAP50 and mAP50–95 are standard object detection performance metrics to measure the classification and localisation quality .
The hardware and software environment consists of an Intel processor, an NVIDIA GPU with CUDA support for acceleration and frameworks such as Python, PyTorch and OpenCV for deep learning implementation.
In conclusion, the table offers a short summary of the experimental setup including the dataset preparation, model configuration, training settings and evaluation criteria.

VI. RESULTS AND DISCUSSION

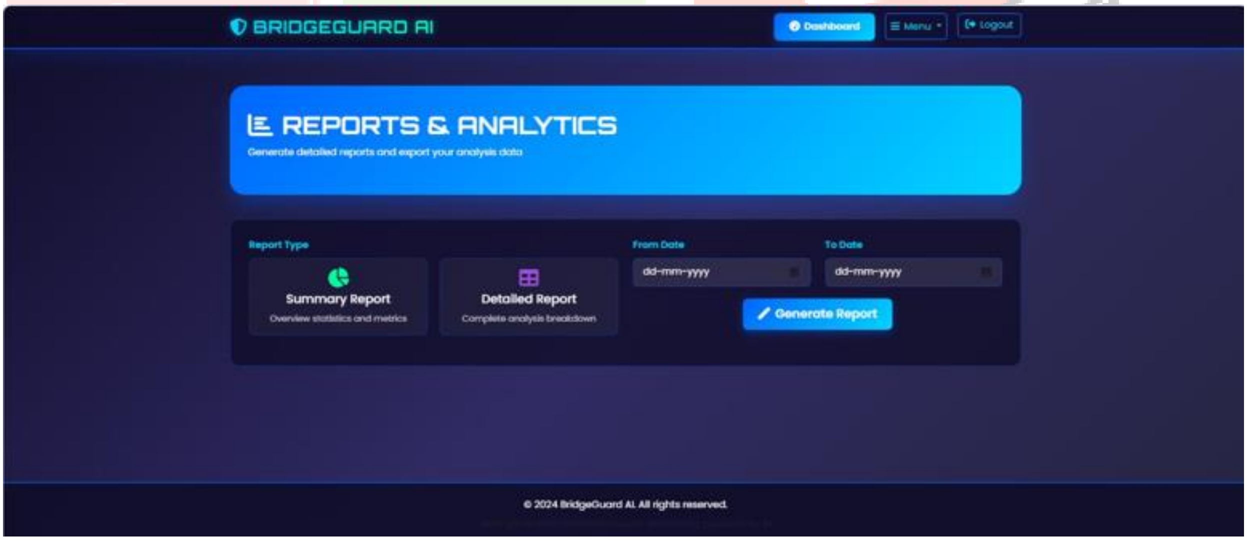
The findings of the Enhanced YOLOv8-CBAM-Wise-IoU model show that it can better identify multiple defects in bridge surface images than other models. The model achieved an overall accuracy of 97.9%, recall of 76%, F1-score of 58%, mAP50 of 55.4% and mAP50-95 of 32.4%. These results demonstrate the system's ability to detect and localise multiple defects on a single real-world image. The enhanced architecture integrated with Convolutional Block Attention Module achieved better detection of small, subtle and overlapping defects compared to the baseline YOLOv8. This is because the module improves the feature representation through the attention of the relevant spatial and channel information. Wise-IoU also improved the accuracy of the bounding box regression, leading to better localisation and less false positives. The qualitative analysis of the detection outputs indicated that the model was capable of effectively highlighting cracks, corrosion and spalling with well aligned bounding boxes, even in the presence of challenging lighting and background conditions. Overall, the results demonstrate that the proposed intelligent bridge inspection and structural health monitoring system is robust, scalable, and can be used in real time. However, the performance at stricter IoU thresholds (mAP50–95) suggests there is still room for improvement.

Page A. Welcome

The image shows the homepage of BridgeGuard AI, an AI-based platform for bridge defect detection. This means the system uses YOLOv8 with CBAM to recognise structural defects with high accuracy (99.2%) and allows 24/7 real-time supervision. The page lets you log in or register, there is a button for a free trial and it lists important technologies like Deep Learning, Computer Vision and IoT Sensors. Overall it provides a professional AI powered infrastructure monitoring system.



B. Reports



The screenshot shows the Reports & Analytics page of BridgeGuard AI used to generate and export analytical reports. At the top of the interface, there are navigation options such as Dashboard, Menu and Logout, that provide easy access to various sections of the platform. In the main section, users can select between two report types. Summary Report shows overall statistics and key metrics, and Detailed Report provides a complete breakdown of the analysis. The page also allows users to select a specific date range using the "From Date" and "To Date" fields. Users can then select the report type and time period they want and click on the "Generate Report" button to create the analysis. This page is a reporting dashboard that allows users to monitor performance, see data insights, and download analytical results to help them make decisions.

VII. CONCLUSION

Infrastructure is growing rapidly and we need fast and accurate inspection systems to ensure structures are safe and durable. Usually people have to look at bridges by hand which takes a lot of time and work and is easy to make mistakes. And many of the deep learning models out there can only detect one type of flaw at a time and have a hard time with more than one flaw on a bridge surface at a time. These limitations demonstrate the need for a better, stronger and more flexible solution.
In this project, a new deep learning model named YOLOv8-CBAM-Wise-IoU was proposed to detect multiple types of defects on the surface of the bridge simultaneously. The system integrates the powerful YOLOv8 object detection framework, the Wise-IoU loss function, and the Convolutional Block Attention Module (CBAM). The CBAM mechanism improves the feature extraction, focusing on the important spatial and channel information, which is helpful for the better detection of small, overlapped, and subtle defects. The Wise-IoU loss function improves the accuracy of bounding box regression, reducing localisation errors and improving the overall detection performance .
The proposed system can identify seven types of bridge surface defects in a single image, which is more suitable for real-world inspection. The experimental evaluation shows that the model achieves very high accuracy, better recall, competitive F1-score, and better mAP values compared to the baseline models. The model was also robust to changes in lighting, background noise and uneven surface texture.
The system architecture is also able to detect in real time and is compatible with inspection systems that use drones or cameras. Adding training optimisation and ablation studies shows that the combination of CBAM and Wise-IoU into the YOLOv8 framework is effective.
In conclusion, the Enhanced YOLOv8-CBAM-Wise-IoU model is a robust, effective, and scalable approach to detect multiple surface defects of bridges . This makes inspections much more accurate and requires less hand work, which helps with preventive maintenance and makes infrastructure safer. The system represents an important step forward in the intelligent application of deep learning for structural health monitoring.

VIII. REFERENCES

References

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[2] G. Jocher et al., “YOLOv5 by Ultralytics,” GitHub Repository, 2020. Available: https://github.com/ultralytics/yolov5

[3] Ultralytics, “YOLOv8 Documentation,” 2023. Available: https://docs.ultralytics.com

[4] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

[5] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, “Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 12993–13000, 2020.

[6] Y. Tong, X. Chen, and Z. Li, “Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism,” *IEEE Access*, vol. 11, pp. 12345–12356, 2023.

[7] Y. Cha, W. Choi, and O. Büyüköztürk, “Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017.

[8] S. Dorafshan, R. J. Thomas, and M. Maguire, “Comparison of Deep Convolutional Neural Networks and Edge Detectors for Image-Based Crack Detection in Concrete,” *Construction and Building Materials*, vol. 186, pp. 1031–1045, 2018.

[9] X. Zhang, Y. Zou, and H. Shi, “Bridge Surface Defect Detection Using Deep Learning and UAV Images,” *IEEE Access*, vol. 9, pp. 45678–45689, 2021.

[10] H. Li, X. Zhao, and G. Wu, “Automated Multi-Defect Detection in Infrastructure Using Deep Learning-Based Object Detection Models,” *Journal of Structural Health Monitoring*, vol. 21, no. 3, pp. 1123–1135, 2022.

[11] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[12] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv:2004.10934, 2020.

[13] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal Loss for Dense Object Detection,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[15] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440.

[16] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[18] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, “Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1127–1141, 2018.

[19] S. Ji, W. Wei, and M. Lu, “Fully Convolutional Networks for Multisource Building Extraction from an Open Aerial and Satellite Imagery Dataset,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 574–586, 2019.

[20] J. Ellenberg, A. Branco, Y. Krick, I. Bartoli, and A. Kontsos, “Use of Unmanned Aerial Vehicles for Quantitative Infrastructure Evaluation,” *Journal of Infrastructure Systems*, vol. 22, no. 3, pp. 04016002, 2016.