Intelligent monitoring of steel bridge bearings using deep learning methods

COMPDYN 2025
10th ECCOMAS Thematic Conference on
Computational Methods in Structural Dynamics and Earthquake Engineering
M. Papadrakakis, M. Fragiadakis (eds.)
Rhodes Island, Greece, 15-18 June 2025

INTELLIGENT MONITORING OF STEEL BRIDGE BEARINGS USING DEEP LEARNING METHODS

Sina Roustaeikakaei[1], Emanuela De Luca[1,2], Gabriele Bertagnoli[2], Davide Masera[1]

[1] Masera Engineering Gourp S.R.L.
Corso Re Umberto, 8, Turin, Italy
e-mail: {sina.roustaeikakaei, emanuela.deluca, davide.masera}@masera-eg.com

[2] Politecnico di Torino - DISEG
Corso Duca degli Abruzzi, 24, Turin, Italy
{emanuela.deluca, gabriele.bertagnoli}@polito.it

Abstract

The advancement of mobile imaging technology has given rise to computer vision techniques, but there have been few attempts to use computer vision for automatic condition assessment of bridge bearings. In fact, human visual inspection still plays a major role in the assessment of bridge bearings conditions today. The objective of this research is to devise an intelligent Structural Health Monitoring (SHM) system for steel bearings of bridges, with a particular focus on aged steel bearings. These components are deeply related to the load-bearing capacity and overall structural integrity of bridges. The proposed SHM system leverages the capabilities of Deep Learning (DL), specifically Convolutional Neural Networks (CNNs), for image-based condition assessment. The model is trained to perform two sequential tasks: component detection and damage classification. Initially, the model identifies the presence of bearings within the collected images, subsequently, it classifies the detected bearings, then it identifies a possible damage, providing a quantitative measure of its extension. A dataset of images showing various conditions of steel bearings and connections is collected and annotated to support the development of the SHM system.

These images include a wide range of damage intensities to ensure the robustness of the model. The training process will involve finetuning a pre-trained model using transfer learning to optimize its performance for the specific task of component detection and damage classification.

Keywords: Structural Health Monitoring, Bridges, Steel Bearings, Deep Learning, Computer Vision, Damage Detection, Damage Classification.

1 INTRODUCTION

Structural health monitoring (SHM) technology was developed with the primary objective of ensuring the operational safety of engineering structures by deploying a variety of sensors and monitoring a wide range of physical parameters. Evaluating the structural condition and performance and providing guidance on routine inspection and maintenance [1-7]. Regarding a large-scale SHM system, the innovative sensing technologies from a variety of fields, such as mechanics, electricity, electromagnetism, optics, thermology, and chemistry, make a great contribution in accurately acquiring the huge amount of original data reflecting the real environmental and structural conditions. Over the previous thirty years, researchers from all around the world have made significant strides in creating innovative sensing technologies that can be used in the field of SHM research.

In the field of civil engineering, SHM system plays today a crucial rule in the continuous monitoring and maintenance of infrastructure to ensure safety and longevity. As vital parts of transportation systems, old bridges need to be monitored with great precision so that structural problems can be sensed early, before they develop into drastic failures. Bridge bearings are one of the most exposed structural elements to environmental aggression and are supposed to ensure load bearing capacity, thermal expansion, and resistance to seismic action. However, the traditional approach to their inspection relies heavily on manual visual assessments, which are time-consuming, expensive, subjective, and prone to human error.

Recent advancements in computer vision and deep learning present an innovative solution for monitoring bridges. The rise of high-resolution imaging devices and advanced AI models enables automated inspection systems to deliver real-time, precise, and scalable evaluations [8]. This study introduces an advanced framework that utilizes Convolutional Neural Networks (CNNs) for the automated detection and classification of damage in steel bridge bearings through image-based analysis.

By utilizing YOLOv8 and YOLOv8n-seg [9], two state-of-the-art object detection and segmentation models, this research develops a prototype of an highly efficient monitoring system capable of identifying bearing components and detecting their damage with high precision. These models are trained using a custom-annotated dataset of steel bridge components, encompassing various damage severities to enhance its robustness.

2 BACKGROUND AND RELATED WORK

2.1 Structural Health Monitoring and Computer Vision

The foundation of traditional SHM techniques is based on manual inspections, which are time-consuming, take place a fixed number of months or years, and could miss early-stage structural deterioration. Through image and video processing, **computer vision approach** allows for **automated real-time condition assessment**, which improves SHM.

The ability to recognize objects and segment damage inside structural components has been greatly enhanced by computer vision algorithms and deep learning models. They can extract information from images making **damage detection**, **classification**, and **predictive**

maintenance become achievable. These techniques have been extensively applied in **crack detection**, **corrosion analysis**, and material degradation assessment.
Along with that, advancements in this flied enables tracking changes in structures in real time and reducing the importance of manual interventions. Machine learning algorithms complement it even more through better accuracy in growing datasets. Such advances support ensuring long-term structural health monitoring system's safety, efficiency, and cost-effectiveness.

2.2 Deep learning algorithms for Structural monitoring

Deep learning (DL) algorithms and specifically CNNs revolutionized the processing of images by automatically extracting spatial features from image data. Unlike traditional machine learning algorithms that utilize handcrafted features, CNNs can learn **hierarchical representations** [10] and are thus most appropriate for **damage localization** and **classification**. The architecture of CNN is inspired by the feedforward networks that were designed by the behavior and functional principles of the visual cortex, a brain region that processes visual information [10]. More specifically, convolutional neural networks are feedforward neural networks that were developed for image recognition tasks. Nowadays they are applied to a variety of tasks, such as image classification, motion detection, voice recognition, and natural language processing. The properties of the convolutional networks include the following ones [11] [12]:

- Convolution Layer: used to reduce the dimension of data and computational complexity.
- Sparse Connection: refers to reducing the number of connections between neurons in the current layer and the next layer.
- Parameter Sharing: uses shared weights for a kernel at each level and shares these weights for other parts of the input features to extract features.
- Pooling: A strategy for reducing the dimensions of input data and learning useful features.

Convolution is an operation of two functions to produce the third function, it can be defined by the following relation:

$$(I * K)(t) = \sum_{a \in D_I} I(a)k(t-a) \tag{2.1}$$

Where I is the input and K is kernel. Usually the kernel's values are called kernel's weights. In Deep Learning, CNNs are usually displayed as rectangular arrays of real values, where the convolution is two-dimensional and discrete. Convolution is usually movable, meaning that kernels are shifted over the input. According to these definitions the convolution can be defined as:

$$(I * K)(i,j) = \sum_{1}^{m}\sum_{1}^{m} I(i-m, j-n)k(m,n) \tag{2.2}$$

The convolution process of an input image is illustrated in Figure 1:

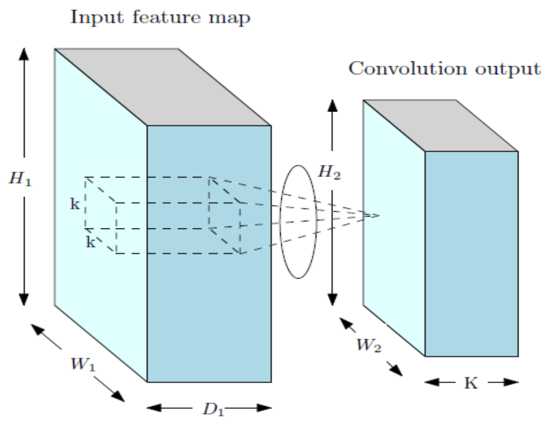

Figure 1. An example of input volume [W1 × H1 × D1], with operation of convolution with a kernel size k*k*k, resulting output size [W2 × H2 × D2] [13].

In this research, the two models **YOLOv8** and **YOLOv8n-seg** were trained using Roboflow platform [14] since they perform well in **detecting** and **segmenting** the captured images. The application of transfer learning is used also in order to enhance the model's adaptability to SHM-specific datasets that will be described in Chapter 3.2 with more detail.

3 METHODOLOGY

The proposed framework for intelligent monitoring of steel bridge bearings follows a structured methodology, which includes **dataset preparation, model training, and evaluation metrics**.

3.1 Dataset Collection and Annotation

A custom dataset comprising images of **steel bridge components** was used. The dataset was carefully labeled using the **Roboflow annotation tool** [14], where images were categorized on the base of:

- **Component types**: single pendulum, double pendulum, gerber connections, and other connections.
- **Damage conditions**: Presence of corrosion, oxidation, and debris accumulation.

To improve model generalization, **data augmentation techniques** were applied, producing multiple variations of each training image. The augmentation process included:

- **Outputs per training example:** 7
- **Flip:** Horizontal
- **90° Rotate:** Clockwise, Counter-Clockwise
- **Rotation:** Between -10° and +10°
- **Saturation:** Between -24% and +24%

3.1.1. Dataset Partitioning and Evaluation Strategy

In this section of the research, the dataset was split into three distinct subsets to enable a comprehensive assessment of the model and its performance.

- **Training Set:** This subset holds the majority of the dataset and is used for training the model. The model learns from this data to identify patterns and attributes concerning different states of steel bridge components.
- **Validation Set:** This is another distinct subset used during the training process to adjust model parameters and avoid the danger of overfitting. Its role is crucial in determining the optimal model configuration by giving an unbiased assessment during the training process.
- **Test Set:** This particular subset is used to evaluate the model's performance after training. It contains information that the model has not seen before, thus providing an unbiased measure of its generalization capability.

The following standard data partitioning strategy was employed to ensure robust model evaluation.

Model	Train Set	Validation Set	Test Set
YOLOv8 (Detection)	318 images (87%)	32 images (8%)	16 images (6%)
YOLOv8n-seg (Segmentation)	3016 images (84%)	124 images (14%)	62 images (2%)

Table 1: Dataset Distribution for Object Detection and Segmentation Models.

3.2 Model Training

Deep learning models were trained step by step grounding on the strategies established in this study. The **YOLOv8** and **YOLOv8n-seg** models were fine-tuned using a transfer learning strategy, so they could detect and segment steel bridge components more effectively.
Preprocessing of the dataset was performed (such as normalization and augmentation techniques defined in Section 3.1) in order to improve the model generalization. The training procedure involved:

- **Preprocessing:** Images were resized and normalized to maintain consistency and optimize input.
- **Transfer Learning:** The COCO [15] dataset pre-trained weights were used to improve convergence and accuracy in structural health monitoring tasks
- **Hyperparameter Optimization:** The model tuning process was repeated in multiple manners, testing different learning rates, batch sizes and weight decay coefficients to achieve the lowest possible losses and highest available precisions.
- **Evaluation Metrics: Mean average precision** (mAP) and **Segmentation score** (also called IOU) were used for measuring the efficiency of the model at object detection and segmentation tasks respectively.

Training was done over multiple epochs with convergence ensured to low loss while overfitting was avoided using techniques such as early stopping and validation supervision. Tested by using test set data to evaluate their strength in practical scenarios. The results of these training and validation processes are discussed in Section 4.

The annotated dataset is used for training YOLOv8 and YOLOv8n-seg models. The training process involved:

- **Image resizing and normalization** to ensure consistency in input data.
- **Fine-tuning of pre-trained models** using transfer learning from the COCO dataset [15].
- **Optimization of hyperparameters**, including batch size, learning rate, and number of epochs.

The loss functions and performance metrics used for evaluation included [16]:

- **Mean Average Precision (mAP@0.5)** for detection accuracy.
- **Intersection over Union (IoU)** for segmentation precision.
- **Training loss convergence graphs** to monitor model learning stability.

4 RESULTS AND DISCUSSION

4.1 Object Detection Performance

Object detection in steel bridge-beaings was performed using the YOLOv8 model. The model was trained and validated by Roboflow platform [14] using the dataset described in the corresponding section 3.1.1, thereby facilitating a comprehensive assessment of its generalization ability.

For evaluating the performance of our model, the standard object detection matrices were used, the **Mean Average Precision (mAP)**, to evaluate its performance in terms of how well it correctly identifies objects and minimizes false positives. Our model reached a mAP score of **96.5**%, which demonstrates the acceptable recognition performance of the steel components provided under various conditions.

The model's training progress over 300 epochs, is illustrated in Figure 2, showing the pattern of convergence and overall stability.

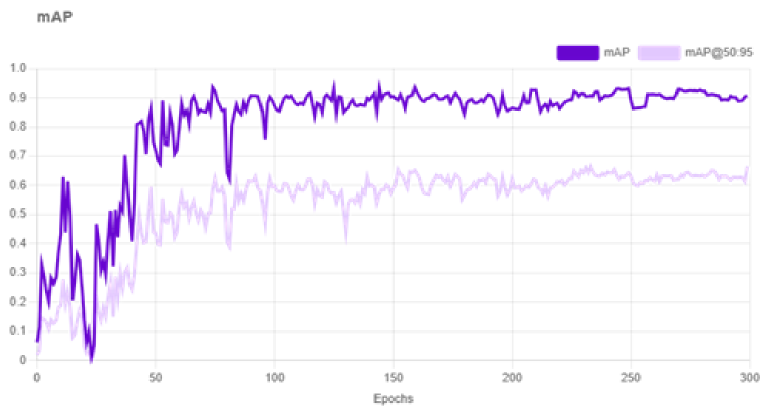

Figure 2. Training progress of object detection model, showing mAP convergence over 300 epochs.

The three Losses shown in Figure 3 are the key components in YOLO-based architecture. They indicate the measure of error or discrepancy between the model's predictions and the actual annotated data, briefly, Box Loss (Localization Loss) is a measure of how well the predicted bounding box matches the ground truth bounding boxes, Class Loss (Classification Loss)

evaluates the effectiveness of classified object and Object Loss (Confidence Loss) calculates the confidence of the model is in detecting actual objects versus background noise.
The goal in training a model is typically to minimize this loss: the decreasing value shown in the graph is a positive sign which means that the model is learning and improving over time.

Figure 4 illustrates outputs where the components are identified and labeled with confidence scores:

Figure 3. Training Loss Trends: Box Loss, Class Loss, and Object Loss over 300 epochs.

Figure 4. Example of object detection model's outputs. The detected component is highlighted with a bounding box and classification confidence score.

4.2 **Damage Classification and Segmentation**

The model's performance in damage classification and segmentation is evaluated using the Mean Average Precision (mAP) metric, which calculates the accuracy of the segmentation.

The maximum achieved mAP our model **is 77.8%**, which demonstrates its capability in recognizing corrosion, oxidation, and debris accumulation in general. The model is successful in the identification of the damaged regions and provides valuable information for our assessment.

Training performance and segmentation accuracy trends over 140 epochs are presented in Figure 5. Box, Class and object loss are pictured in Figure 6 whereas Figure 7 presents some output from the segmentation model, where steel components and their defects are accurately segmented and labeled with confidence scores.

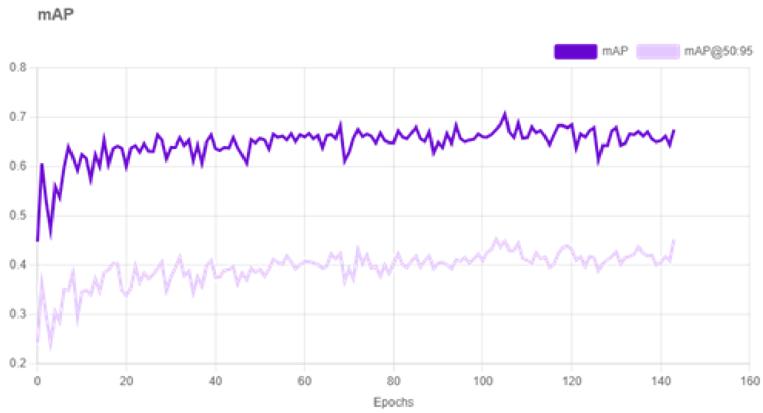

Figure 5. Training performance of YOLOv8n-seg segmentation model, depicting mAP trends for damage classification.

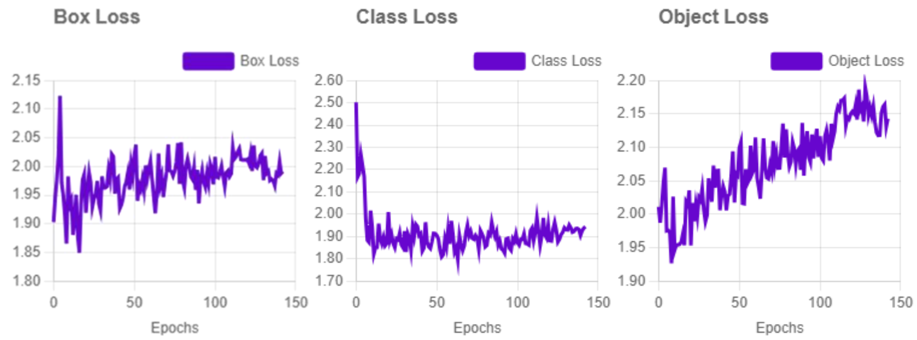

Figure 6. Training Loss Trends: Box Loss, Class Loss, and Object Loss over 140 epochs.

Figure 7. Example of segmentation model's outputs. The detected component is highlighted with a bounding box and classification confidence score.

4.3 Challenges and Limitations

Although the models showed acceptable accuracy in segmentation and detection of the provided steel bearings, numerous challenges arose upon application:

- **Dataset Imbalance:** most of the provided images were showing only oxidation damage type, which can cause a bias in model predictions. Moreover, the dataset featured many more photos of undamaged bearings than degraded ones.

- **Complex Environmental Factors:** Variations in lighting, occlusion, and angle of capture of the picture challenged the generalization of the model. Data augmentation strategies mitigated some of these issues, but further refinement is required.

- **Computational Resources:** Emphasizing the need of effective model optimization strategies, training deep learning models on high-resolution images needed significant computational capability.

Future research should focus on overcoming these challenges by expanding the dataset, enhancing augmentation techniques for improved generalization.

5 CONCLUSION AND FUTURE WORK

This study proposes an enhanced artificial intelligence-powered framework for structural health monitoring (SHM) of steel bridge bearings that embraces deep learning-based computer vision approaches. The use of YOLOv8 as an object detector and YOLOv8n-seg as a damage segmenter revealed great potential in automating the inspection procedure. Through the utilization of a properly organized dataset within at least 700 images and the inclusion of diverse augmentation techniques, the proposed system obtained high detection accuracy (mAP 96.5%) and stable damage segmentation (mAP 77.8%).

The results show that deep learning models can potentially substitute conventional time-consuming inspection techniques, providing real-time data-driven evaluations of structural health. Nevertheless, issues like dataset imbalance, environmental heterogeneity, and computational requirements still constitute areas needing further enhancement.

Future research directions may include**:**

- **Enhanced Data Collection:** expanding the dataset to include a wider range of damage types and environmental conditions to improve generalization.
- **Optimized Model Architectures:** exploring lightweight yet powerful models to reduce computational costs and enable real-time deployment.
- **Integration with UAV Technology:** implementing drone-based monitoring systems to facilitate large-scale, autonomous inspections.
- **Multimodal Sensor Fusion:** combining computer vision with thermal imaging, LiDAR, or acoustic sensing for more comprehensive structural assessments.

By addressing these challenges, AI-based SHM solutions can be further developed in follow-up research to enhance their practicality for real-world bridge inspection and safety assessments.
This article presents an AI-driven SHM framework for automatic bridge bearing assessment using computer vision. The system successfully utilizes YOLO-based models to detect and classify the state of damage with high accuracy.

REFERENCES

[1] Bertagnoli, G.; Luca, F.; Malavisi, M.; Melpignano, D.; Cigada, A. (2020) A large scale SHM system: A case study on pre-stressed bridge and cloud architecture. In: Dynamic of Civil Structures Vol. 2 – Proceedings of the 37th IMAC, A Conference and Exposition on Structural Dynamics (Conference Proceedings of the Society for Experimental Mechanics Series), Orlando, 28-31 January 2019, pp. 75-83.

[2] X. W. Ye, C. Z. Dong, and T. Liu; A Review of Machine Vision-Based Structural Health Monitoring: Methodologies and Applications.

[3] TUNG KHUC; Computer Vision Based Structural Identification Framework for Bridge Health Monitoring.

[4] Gianni Moor, Niculin Meng and Colm O'Suilleabhain; New York, USA (2013) Remote structural health monitoring systems for bridge expansion joints and bearings.

[5] Matthew Yarnold & Nathaniel C. Dubbs; Bearing Monitoring Case Study.

[6] Y. Q. Ni, M.ASCE; X. W. Ye; and J. M. Ko, F.ASCE3; Monitoring-Based Fatigue Reliability Assessment of Steel Bridges: Analytical Model and Application.

[7] X. W. Ye, Y. H. Su, and J. P. HanStructural Health Monitoring of Civil Infrastructure Using Optical Fiber Sensing Technology: A Comprehensive Review.

[8] Mahta Zakaria, Enes Karaaslan and F. Necati Catbas; Advanced bridge visual inspection using real-time machine learning in edge devices.

[9] M. A. P. P. NYOMAN KARNA, "Towards Accurate Fused Deposition Modeling 3D Printer Fault Detection using Improved YOLOv8 with Hyperparameter Optimization," 2016.

[10] Joshua Peterson, Paul Soulos, Aida Nematzadeh, & Thomas L. Griffiths; Learning Hierarchical Visual Representations in Deep Neural Networks Using Hierarchical Linguistic Labels.

[11] A. M. L. &. J. C. Osval Antonio Montesinos López, "Convolutional neural networks," in *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, 2022, pp. 533-577.

[12] F. V. Vincent Dumoulin, "A guide to convolution arithmetic for deep learning," 2018.

[13] H. Söderlund, "Real-time Detection and Tracking of Moving Objects Using Deep Learning," 2019.

[14] Dwyer, B., Nelson, J., Hansen, T., et. al. (2024). Roboflow (Version 1.0) [Software]. Available from https://roboflow.com. computer vision.

[15] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollar, P., & Zitnick, C. L. (2014). *Microsoft COCO: Common Objects in Context*. In *European Conference on Computer Vision (ECCV)* (pp. 740–755). Springer., https://cocodataset.org/.

[16] R. Padilla, S. L. Netto and E. A. B. da Silva, "A Survey on Performance Metrics for Object-Detection Algorithms," 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 2020, pp. 237-242, doi: 10.1109/IWSSIP48289.2020.9145130.