HOKKAIDO UNIVERSITY

Title	Active Semi-Supervised Learning for Scalable Corrosion Detection in Steel Structures
Author(s)	Taher, Saad Al Deen Sal; Dang, Ji
Relation	Proceedings of the 10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026), June 28 - July 1, 2026, Sapporo, Japan
Citation	Proceedings of the 10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026)
Issue Date	2026-07-01
Doc URL	https://hdl.handle.net/2115/100450
Type	conference paper
File Information	C000090_1.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026),
28 June-01 July 2026, Sapporo, Japan
Kitahara, T., Miyamori, Y. & Beer, M. (editors)

Active Semi-Supervised Learning for Scalable Corrosion Detection in Steel Structures

Sal Saad Al Deen Taher[1]*, Ji Dang[2]

[1]*Department of Civil & Environmental Engineering, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama-shi, Japan.*
E-mail: sal.saad.a.612@ms.saitama-u.ac.jp
[2]*Department of Civil & Environmental Engineering, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama-shi, Japan.*
E-mail: dangji@mail.saitama-u.ac.jp
**Corresponding Author*

Detection of corrosion in steel structures is essential for reliable monitoring of structural health, but the creation of large, high quality labelled data sets remains costly and labor-intensive. This study introduces an active semi-supervised learning framework to efficiently scale corrosion detection by combining uncertainty-based active sample selection and semi-supervised learning strategies using large-scale unlabeled field images. The proposed pipeline will iteratively query only the most informative samples for expert annotations, refining the pseudo-labels by regularizing consistency and adaptive confidence filtering. The experimental results for steel corrosion image datasets show that the ASSL approach achieves performance comparable to fully supervised models with significantly fewer labelled samples, demonstrating improved efficiency, generalization capability and robustness to the variability of the real world. This work highlights ASSL as a practical route to cost-effective and scalable corrosion detection in the monitoring of civil infrastructures.

Keywords: Corrosion detection; Steel structures; Active learning; Semi-supervised learning; Structural health monitoring.

1 Introduction

The durability and safety of civil infrastructure systems have become critical global concerns as a significant portion of existing assets approach or exceed their design life. In Japan, nearly 57% of bridges are expected to be over 50 years old by 2034, (Ministry of Land, Infrastructure, Transport and Tourism, 2024) reflecting a rapidly aging infrastructure network and an increasing demand for efficient maintenance and inspection strategies. Among various deterioration mechanisms, corrosion of steel components is one of the most critical factors affecting structural performance, leading to section loss, stiffness degradation, and eventual structural failure. Recent failures, including the Morandi Bridge (2018) (Morgese, et al., 2020), Nanfangao Bridge (2019) (Taiwan Transportation Safety Board, 2020), and Fern Hollow Bridge (2022) (National Transportation Safety Board, 2024), demonstrate the severe consequences of inadequate monitoring and delayed maintenance, emphasizing the need for reliable and scalable structural health monitoring (SHM) systems. Traditional inspection relies on manual visual assessment, which is labor-intensive, subjective, and constrained by safety risks and a shortage of skilled inspectors. Consequently, computer vision and deep learning have been increasingly adopted for automated defect detection. Deep learning methods have evolved from classification and detection to pixel-level segmentation using architectures such as U-Net (Ronneberger, et al., 2015), DeepLabV3+ (Chen, et al., 2018), Mask R-CNN (He, et al., 2018), and Mask2Former (Cheng, et al., 2021). However, fully supervised approaches remain limited by the high cost of pixel-level annotation up to 90 minutes per image (Cordts, et al., 2016) and the variability of real-world data, which restricts model generalization.

To address these challenges, semi-supervised learning (SSL) has emerged as a promising approach by leveraging large volumes of unlabeled data. In SHM, early works introduced teacher–student frameworks with exponential moving average (EMA) updating and consistency regularization for crack segmentation (Wang & Su, 2021), while GAN-based SSL methods improved performance under low-resolution conditions (Shim, et al., 2022). More recent studies have demonstrated that high segmentation performance can be achieved with minimal labeled data through pseudo-labeling and consistency constraints (Asadi Shamsabadi, et al., 2024).

Further advancements incorporate foundation models and hybrid learning strategies. For instance, Li et al. leveraged the Segment Anything Model (SAM) (Kirillov, et al., 2023) with transfer learning to reduce annotation effort (Li, et al., 2025), while An et al. proposed a corrosion-focused SSL framework with weak-to-strong consistency and attention mechanisms (An, et al., 2025). Similarly, Gauthami et al. introduced a hybrid detection–segmentation approach combining YOLO (Redmon, et al., 2016) and SAM, eliminating the need for pixel-level annotations and improving scalability across defect types (Vijayakumar Kuttuva & Prawin, 2026).

Despite these advances, SSL in SHM remains less mature than in medical imaging and general computer vision. Unlike medical datasets with standardized acquisition protocols, SHM data exhibit complex backgrounds, environmental variability, heterogeneous materials, and severe class imbalance, often leading to noisy pseudo-labels and unstable training. Moreover, most existing approaches rely on passive pseudo-labeling and do not explicitly address efficient sample selection under limited annotation budgets.

Active learning (AL) (Rangnekar, et al., 2022) has been explored to improve data efficiency by prioritizing informative samples. However, unified frameworks integrating AL with SSL for corrosion detection remain limited.

In this context, this study proposes an Active Semi-Supervised Learning (ASSL) framework that combines uncertainty-driven sample selection with transformer-based segmentation. By iteratively selecting informative samples and refining pseudo-labels through adaptive confidence filtering and consistency regularization, the proposed method achieves near-supervised performance with significantly fewer labeled samples, offering a scalable and practical solution for corrosion detection in SHM.

2 Proposed ASSL Framework

This study proposes an Active Semi-Supervised Learning (ASSL) framework for corrosion segmentation that integrates transformer-based semantic segmentation, semi-supervised learning, and uncertainty-driven active learning. The objective is to achieve high segmentation performance while minimizing the requirement for large-scale pixel-level annotations. The overall pipeline of the proposed framework is illustrated in Figure 1, which shows the interaction between labeled data, unlabeled data, the teacher–student models, and the active learning loop.

2.1 *Mask2Former Architecture*

At the core of the framework is Mask2Former, a transformer-based mask classification architecture that formulates segmentation as a set prediction problem. Instead of directly assigning labels at the pixel level, Mask2Former utilizes a fixed set of learnable queries that interact with multi-scale feature representations through a transformer decoder. Each query predicts both a class probability and a corresponding segmentation mask, and the final semantic map is obtained by aggregating these predictions.

The architecture consists of a backbone network (Swin Transformer in this study) for feature extraction, a pixel decoder for multi-scale feature fusion, and a transformer decoder for query-based mask prediction. This design enables effective global context modeling and robust handling of complex corrosion patterns commonly observed in field inspection imagery.

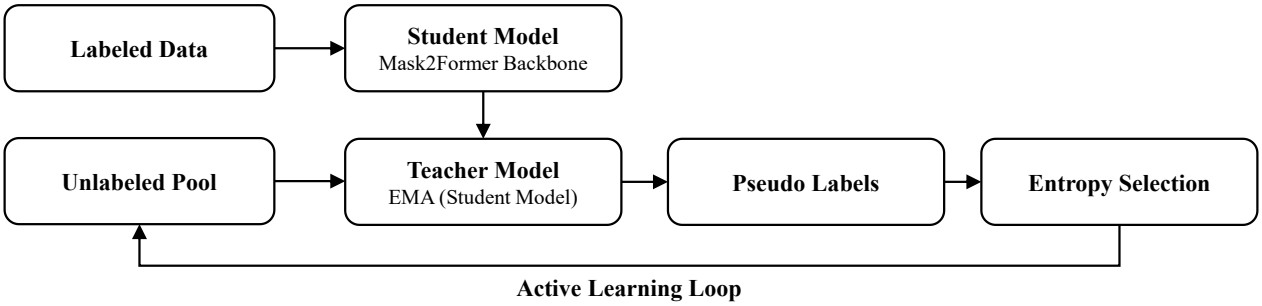

Figure 1. Proposed Active Semi-Supervised Learning (ASSL) Framework.

2.2 *Semi-Supervised Learning with EMA Teacher–Student*

To leverage unlabeled data, a semi-supervised learning strategy based on a teacher–student framework is adopted. The student model is trained using both labeled and pseudo-labeled data, while the teacher model provides stable supervision by generating pseudo-labels. The teacher parameters are updated using an exponential moving average (EMA) of the student parameters:

$$\theta_{teacher} \leftarrow m \cdot \theta_{teacher} + (1 - m) \cdot \theta_{teacher} \tag{1}$$

where m is the momentum coefficient. This EMA update stabilizes the teacher model and improves the reliability of pseudo-labels. During training, the teacher model produces pixel-wise class probability maps by combining class predictions and mask outputs. These probabilities are converted into pseudo-labels using a confidence threshold, τ:

$$\hat{y} = \begin{cases} 1, if\ p \geq \tau \\ 0, otherwise \end{cases} \tag{2}$$

A threshold ramping strategy is applied, where is gradually reduced across training generations. This allows the model to initially rely on high-confidence pseudo-labels and progressively incorporate more uncertain samples as training stabilizes. The overall training objective combines supervised and unsupervised losses:

$$\mathcal{L} = \mathcal{L}_{sup} + \lambda \mathcal{L}_{unsup} \tag{3}$$

where $\mathcal{L}_{sup}$ is computed on labeled data, $\mathcal{L}_{unsup}$ is computed on pseudo-labeled data, and λ balances their contributions.

2.3 *Active Learning via Uncertainty Sampling*

To improve annotation efficiency, the framework incorporates an active learning strategy that prioritizes the selection of informative samples from the unlabeled pool. As illustrated in Fig. 1, after each training cycle, the teacher model evaluates the unlabeled data and estimates prediction uncertainty using entropy:

$$H = \sum_{c} p_c log p_c \tag{3}$$

Where, p_c is the probability of class. The entropy values are aggregated at the image level to compute an overall uncertainty score. Images with higher entropy correspond to more uncertain predictions and are therefore selected for annotation. A fixed number of top-ranked samples are transferred from the unlabeled pool to the labeled dataset through oracle annotation. This targeted selection strategy ensures that labeling efforts focus on the most informative samples, improving learning efficiency.

2.4 *Iterative Training Strategy*

The proposed framework follows an iterative, generation-based training strategy that integrates semi-supervised learning and active learning into a unified loop. Initially, a small subset of the dataset (e.g., 10%, 20%, or 30%) is designated as labeled data, while the remaining samples form an unlabeled pool. During each generation, the student model is trained using both labeled data and pseudo-labeled data generated by the teacher model. The teacher is simultaneously updated via EMA to ensure stable pseudo-label generation. After training, the model evaluates the unlabeled pool using entropy-based uncertainty and selects the most informative samples. These samples are then added to the labeled dataset, and the process is repeated for multiple generations, as illustrated in Figure 1. Through this iterative refinement process, the labeled dataset is progressively expanded, leading to improved generalization while significantly reducing annotation cost.

3 Dataset And Experimental Setuo

3.1 *Dataset Description*

The dataset used in this study was constructed from inspection reports prepared by the General Incorporated Foundation Bridge Survey Committee (Japan). These reports contain detailed visual documentation of bridge conditions, including corrosion damage types and their spatial locations, providing realistic and diverse field scenarios. A total of 1,966 steel bridge corrosion images were collected, covering a wide range of environmental conditions, surface textures, and corrosion patterns. The dataset was divided into 1,572 training images (80%) and 394 validation images (20%), with the validation set strictly held out for model selection and performance evaluation. The original images have varying resolutions and aspect ratios; therefore, all images and corresponding segmentation masks were resized to a fixed resolution of 512 × 512 pixels during preprocessing to ensure consistency in training. The segmentation task is formulated as a binary classification problem, where each pixel is labeled as either corrosion or background.

.

3.2 *Details of Training*

The proposed ASSL framework is implemented using the Mask2Former architecture with a Swin-Small backbone, initialized from weights pre-trained on the ADE20K dataset. All input images are resized to 512 × 512 pixels during training to ensure consistency. The model is optimized using the AdamW optimizer with a learning rate of 5×10^{-5} and a weight decay of 5×10^{-2}. The batch size is set to 2 for labeled data and 2 for unlabeled data, allowing simultaneous optimization of supervised and unsupervised learning objectives.

To ensure stable pseudo-label generation, the teacher model is updated using an exponential moving average (EMA) of the student parameters with a momentum of 0.999. The training process is organized into five generations, each consisting of three epochs, enabling progressive refinement of the model and incremental expansion of the labeled dataset. A confidence thresholding strategy is applied to pseudo-labels, where the threshold is linearly decreased from 0.90 to 0.70 across generations. This approach allows the model to initially rely on high-confidence predictions and gradually incorporate more uncertain samples as training stabilizes.

The active learning component plays a central role in improving data efficiency. At the end of each generation, the model evaluates the unlabeled pool using entropy-based uncertainty and selects the most informative samples for annotation. Specifically, the top 100 high-uncertainty images are selected and transferred to the labeled dataset through oracle annotation. This targeted selection ensures that labeling efforts are focused on samples that provide the greatest contribution to model improvement. The annotation budget is explicitly controlled by combining limited initial supervision with iterative sample acquisition. The training process begins with a small labeled subset corresponding to 10%, 20%, or 30% of the total training dataset, while the remaining images are treated as unlabeled. Over successive

generations, additional samples are incrementally labeled through active learning. Assuming five generations, up to 500 additional images may be annotated, depending on the initial supervision level and the size of the unlabeled pool. Compared to fully supervised training, which would require labeling all 1,572 training images, the proposed approach significantly reduces annotation effort while maintaining competitive performance. This is particularly important for corrosion segmentation, where pixel-level annotation is highly time-consuming due to irregular and diffuse boundaries. By selectively labeling only the most informative samples and leveraging pseudo-labels for the remaining data, the framework ensures efficient utilization of annotation resources.

4 Performance Evaluation

4.1 *Effect of Active Semi-Supervised Learning*

The integration of semi-supervised learning and active learning plays a crucial role in improving performance under limited annotation budgets. As shown in Table 1 the labeled dataset grows progressively through active learning, reaching 35.43%, 26.33%, and 49.11% effective labeled data for the 10%, 20%, and 30% initial settings, respectively.

Table 1. Final Supervised Dataset Fraction

Initial Supervised Dataset Fraction	Final Labelled Images	Final Unlabeled Images	Final Supervised Dataset Fraction
10%	557	1015	35.43%
20%	414	1158	26.33%
30%	772	800	49.11%

A major advantage of the proposed ASSL framework is its ability to achieve competitive performance with reduced annotation effort. While fully supervised training requires labeling all 1,572 training images, the ASSL framework limits annotation through controlled acquisition. For instance, in the 10% setting, the model starts with only 158 labeled images and incrementally acquires additional samples, reaching approximately 557 labeled images after active learning. This corresponds to a reduction of more than 60% in annotation requirements compared to full supervision, while maintaining comparable segmentation performance

4.2 *Quantitative and Qualitative Evaluation*

The performance of the proposed ASSL framework is evaluated under different initial supervision levels (10%, 20%, and 30%) and compared with a fully supervised baseline. The quantitative performance of the proposed ASSL framework under different supervision levels is illustrated in Figure 2, which compares key evaluation metrics across fully supervised and ASSL settings. The metrics include mIoU, pixel accuracy (pACC), mean accuracy (mACC), corrosion IoU, precision, recall, F1-score, and boundary F1-score.

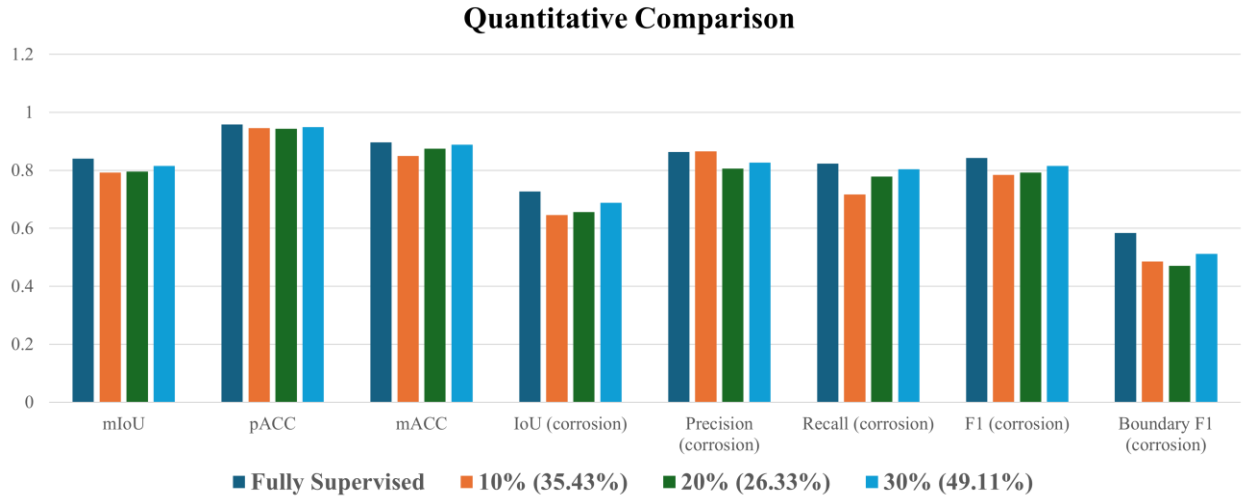

Figure 2. Quantitative performance comparison of the proposed ASSL framework and fully supervised baseline.

The fully supervised model achieves the highest performance with mIoU = 0.8401 and F1-score = 0.8423. However, the ASSL framework demonstrates competitive performance even with significantly reduced labeled data. Notably, the 10% supervision setting achieves mIoU = 0.8156 and F1 = 0.8151, while using only 35.43% of labeled data after active learning, indicating strong data efficiency. Similarly, the 20% and 30% settings achieve mIoU values of 0.7962 and 0.7924, respectively, with corresponding improvements in recall and F1-score.

A key observation is that the performance gap between ASSL and fully supervised learning remains relatively small despite substantial reductions in annotation effort. For example, the corrosion IoU decreases from 0.7275 (fully supervised) to 0.6879 (10% ASSL), while maintaining comparable recall (0.8233 vs 0.8034), indicating that the model preserves strong detection capability even under limited supervision. The results also show that the Boundary F1-score (BF1) is lower compared to other metrics across all settings (e.g., 0.5839 for fully supervised and 0.5122 for 10% ASSL), reflecting the inherent difficulty of accurately capturing irregular corrosion boundaries. This highlights the challenge of fine-grained segmentation in real-world inspection images.

Qualitative comparisons further illustrate the effectiveness of the proposed method. As shown in the visualization results (Figure 3), the ASSL models can capture the overall corrosion regions with reasonable accuracy across different supervision levels. The predictions generated using ASSL exhibit: (a) good localization of corrosion regions (b) consistent detection across varying textures and lighting conditions and (c) robustness to complex backgrounds.

However, some limitations are observed in boundary delineation, where predicted masks tend to be smoother and less precise compared to ground truth annotations. This is consistent with the lower BF1 scores and highlights the challenge of accurately segmenting irregular corrosion edges.

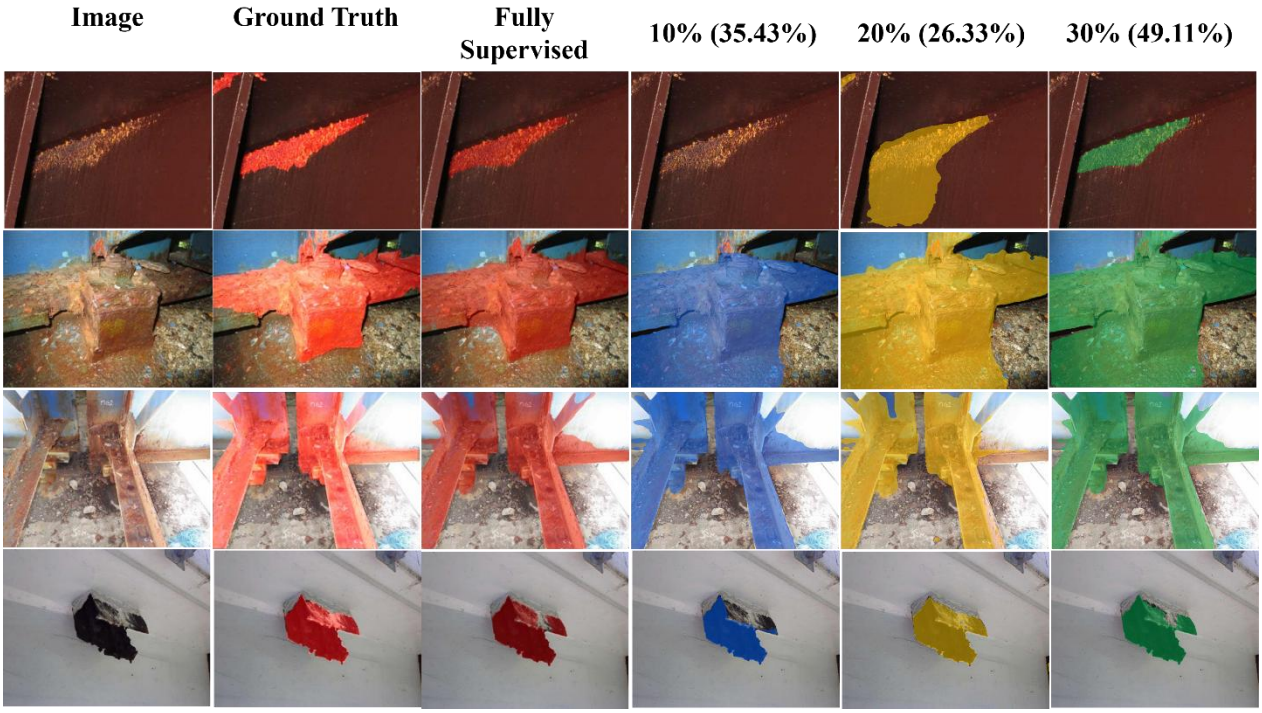

Figure 3. Detected Results for Different Supervised Fractions.

4.3 *Discussion*

The experimental results demonstrate that the proposed ASSL framework effectively addresses the limitations of fully supervised learning in SHM applications. By combining semi-supervised learning with uncertainty-driven active learning, the framework achieves strong performance with significantly reduced labeled data. The findings highlight three key insights:

- Label efficiency is significantly improved through targeted sample selection and pseudo-label utilization.
- Active learning plays a critical role in selecting informative samples, often outperforming random or larger initial labeling strategies.
- Boundary accuracy remains a challenge, indicating potential for future improvements through boundary-aware loss functions or refinement modules.

Overall, the proposed approach provides a scalable and practical solution for corrosion detection, particularly in scenarios where annotation resources are limited.

5 Conclusion

This study proposed an Active Semi-Supervised Learning (ASSL) framework for corrosion detection, integrating transformer-based segmentation with uncertainty-driven active learning to address the limitations of data scarcity in structural health monitoring (SHM). The results demonstrate that the proposed approach achieves near-supervised performance with significantly reduced labeled data, effectively mitigating the high cost and effort associated with pixel-level annotation.

By combining pseudo-labeling with entropy-based sample selection, the framework efficiently utilizes large volumes of unlabeled data while prioritizing the most informative samples for annotation. This targeted learning strategy enables the model to improve progressively with minimal supervision. The findings clearly indicate that annotation efficiency can be substantially enhanced without compromising segmentation performance, making the approach both practical and scalable for real-world infrastructure inspection scenarios.

Future work will focus on extending the framework to multi-class damage segmentation, incorporating boundary-aware loss functions to improve fine-grained edge delineation, and developing adaptive annotation strategies for further optimization of labeling efficiency. In addition, integrating the framework with component-level recognition and hierarchical damage assessment could broaden its applicability in comprehensive and intelligent infrastructure monitoring systems.

References

An, Y., Kong, L., Hou, C. & Ou, J., 2025. Semi-supervised method for automated detection and quantitative assessment of corrosion states in structural members. *Automation in Construction,* June, Volume 174, p. 106155.

Asadi Shamsabadi, E., Erfani, S. M. H., Xu, C. & Dias-da-Costa, D., 2024. Efficient semi-supervised surface crack segmentation with small datasets based on consistency regularisation and pseudo-labelling. *Automation in Construction,* February, Volume 158, p. 105181.

Cheng, B. et al., 2021. Masked-attention Mask Transformer for Universal Image Segmentation. *CoRR,* Volume abs/2112.01527.

Chen, L.-C.et al., 2018. *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation.* s.l.:s.n.

Cordts, M. et al., 2016. *The Cityscapes Dataset for Semantic Urban Scene Understanding.* s.l.:s.n.

He, K., Gkioxari, G., Dollár, P. & Girshick, R., 2018. *Mask R-CNN.* s.l.:s.n.

Kirillov, A. et al., 2023. *Segment Anything.* s.l.:s.n.

Li, J. et al., 2025. Semi-supervised crack detection using segment anything model and deep transfer learning. *Automation in Construction,* February, Volume 170, p. 105899.

Ministry of Land, Infrastructure, Transport and Tourism, 2024. *Roads in Japan 2024,* Tokyo: s.n.

Morgese, M., Ansari, F., Domaneschi, M. & Cimellaro, G. P., 2020. Post-collapse analysis of Morandi's Polcevera viaduct in Genoa Italy. *Journal of Civil Structural Health Monitoring,* Volume 10, p. 69–85.

National Transportation Safety Board, 2024. *Collapse of the Fern Hollow Bridge,* Washington: s.n.

Rangnekar, A., Kanan, C. & Hoffman, M., 2022. *Semantic Segmentation with Active Semi-Supervised Learning.* s.l.:s.n.

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A., 2016. *You Only Look Once: Unified, Real-Time Object Detection.* s.l.:s.n.

Ronneberger, O., Fischer, P. & Brox, T., 2015. *U-Net: Convolutional Networks for Biomedical Image Segmentation.* s.l.:s.n.

Shim, S., Kim, J., Lee, S.-W. & Cho, G.-C., 2022. Road damage detection using super-resolution and semi-supervised learning with generative adversarial network. *Automation in Construction,* March, Volume 135, p. 104139.

Taiwan Transportation Safety Board, 2020. *Final Report on Nanfangao Sea-Crossing Bridge Collapse,* New Taipei City, Taiwan: s.n.

Vijayakumar Kuttuva, G. & Prawin, J., 2026. Structural defect segmentation using a semi-supervised algorithm integrating YOLO and the segment anything model. *Automation in Construction,* February, Volume 182, p. 106709.

Wang, W. & Su, C., 2021. Semi-supervised semantic segmentation network for surface crack detection. *Automation in Construction,* August, Volume 128, p. 103786.