Numerical Methods in Civil Engineering, 10-3 (2026) 1-11

Numerical Methods in Civil Engineering

Journal Homepage: https://nmce.kntu.ac.ir/

A Computer Vision Approach to Detecting Shear Buckling in Thin Steel Plates Using Real and Synthetic Datasets

Mohammad Sabouri-Ghomi*, Anjan Bhowmick**

ARTICLE INFO

RESEARCH PAPER

Article history:
Received:
November 2025
Revised:
December 2025
Accepted:
February 2026

Keywords:
Computer vision; Shear buckling; Steel plate; Structural health monitoring; Finite element model; Synthetic Image Data

Abstract:

In recent decades, numerous innovative methods have been developed to improve the quality, speed, cost-effectiveness, and efficiency of structural damage detection. Among these, computer vision has emerged as a an auspicious approach, particularly due to recent advancements in machine learning. Many successful models have been developed to detect a wide range of structural damages such as cracks, spalling, corrosion, rusting, and bolt loosening. However, shear buckling damage has been almost entirely neglected in the literature. This type of damage commonly occurs in thin steel plates used in structural members such as steel plate shear walls, and its accurate detection and localization are essential for reliable post-event condition assessment, especially following seismic loading. This study investigates the application of computer vision for both detection and instance-level localization of shear buckling damage using an instance segmentation framework. A key challenge in this task is the limited availability of labeled real-world images. To address this issue, an innovative data augmentation strategy is proposed that combines synthetic images generated using finite element analysis (FEA) with visually enhanced synthetic images created using 3D modeling software. These synthetic datasets are then combined with real experimental images to form larger and more diverse training datasets. In total, five datasets were considered, including 208 real images, 343 synthetic images, and 551 combined real–synthetic images. All models were trained using the YOLO11 instance segmentation algorithm. Results demonstrate that a model trained solely on real images achieved strong segmentation performance, with a precision of 0.87 and a recall of 0.81. The best-performing model, trained using a combination of real and visually enhanced synthetic images, achieved a precision of 0.90 and a recall of 0.84, corresponding to an improvement of approximately 3% in mAP50 compared to the real-only model. These findings confirm that high-fidelity synthetic data can effectively mitigate data scarcity and significantly enhance shear buckling detection and localization performance.

1. Introduction

Steel structures are one of the most common types of structures constructed worldwide. Over the past few decades, many innovative solutions have been developed to improve different aspects of these structures, such as reducing construction time, enhancing ease of construction, minimizing weight and cost, and increasing reliability against lateral loads. Consequently, the use of thin-walled elements in steel structures has gained significant attention.

**Corresponding author: Assistant professor, SADRA Institute of Higher Education., Tehran, Iran. Email: Msabouri@sadra.ac.ir*
***Associate Professor, Dept. of Building, Civil and Environmental Engineering, Concordia University, Montréal, QC, Canada H3G 2W1.*
Email: anjan.bhowmick@concordia.ca

Thin steel plate shear walls are a recognized lateral load-resisting system known for their effectiveness against earthquakes due to their excellent properties, such as high shear strength, stiffness, and large energy dissipation capacity [1-7].

Although thin steel plates can exhibit large shear load-bearing capacity, they are prone to buckling even under relatively small lateral loads. Since lateral load-resisting systems are generally designed to act as a structural fuse, absorbing damage while critical members, such as columns, remain undamaged, buckling is often permitted in steel plate shear wall systems. Instead, the system is allowed to enter the plastic phase and dissipate large amounts of energy

https://doi.org/10.61882/NMCE.2511.1109

through the yielding of the thin steel infill plates. An advantage of this design approach is that the fuse (in this case, the thin plate) can be replaced after small or moderate earthquakes. Using this design method allows the structural frame to remain intact and avoids the high cost of rebuilding the entire structure. This, of course, requires post-earthquake inspections to assess damage to the structural members and determine whether replacement is necessary.

Inspecting a structure's condition after events such as earthquakes or for other purposes was traditionally performed manually by human inspectors. This process can be very time-consuming and requires a certain level of skill and knowledge for the inspector to make correct and accurate decisions. As a result, in recent years, various methods have been proposed and implemented to automate, accelerate, and enhance the quality and precision of structural health monitoring procedures. Examples of these innovative methods include the use of various types of sensors, such as wireless smart sensors and optical fiber sensors, along with signal processing techniques and laser-based monitoring systems [8-14].

Computer vision is another key method that has recently gained attention in the engineering community, particularly due to the exponential rise in the adoption of deep neural networks following the introduction of AlexNet [15]. By leveraging computer vision techniques, structural damage can be identified directly from photographs using a computer. This approach eliminates the need for skilled inspectors and significantly accelerates the inspection process. This technique allows a computer to process a large number of images in a short amount of time and detect structural damage with high accuracy.

Many studies in recent years have explored the use of computer vision to detect various types of structural damage. For example, crack detection in different civil engineering materials such as concrete, asphalt, and steel has been extensively studied and significantly improved [16-18]. Other types of concrete damage, such as crushing, spalling, and corrosion, have also been effectively identified using computer vision techniques [19-22]. Additionally, rust recognition and loose bolt detection are notable examples of damage and defect detection in steel structures [23, 24].

However, among the broad range of structural damage types, there are very few studies that focus on shear buckling damage in thin steel members. One of the rare examples in this area investigated the detection of buckling in corrugated panels using computer vision and reported high accuracy in identifying and localizing buckling damage in steel corrugated plates [25]. Given this significant gap in research, this study aims to advance the field by focusing on the detection of shear buckling damage in thin steel plates, an essential aspect of structural health monitoring that has largely been unexplored.

One of the key challenges in many machine learning projects is obtaining appropriate and sufficient data to train the desired model. This is particularly true for computer vision models, which require visual data, such as images or videos, to train object detection or segmentation algorithms. This challenge is especially noticeable in structural damage detection tasks due to the lack of sufficient images for certain types of damage. While there may be ample data available for some types of damage, such as crack damage, owing to the visibility of cracks, other types, such as shear buckling, are significantly more challenging to document.

Unlike damage types such as pavement distress, where databases can be created relatively easily by manually taking pictures or installing dashcams on vehicles [26], it is not straightforward to generate a high-quality database of shear buckling images. For this study, images from a recent experiment conducted by the authors [27] are used as the primary database for training the detection model. In addition, to enhance the model's accuracy and address the issue of limited data for buckling damage, an innovative approach is also implemented: the creation of synthetic images using finite element analysis (FEA).

Generating synthetic training data is a well-recognized technique in computer vision and broader machine learning applications, particularly when there is a lack of sufficient or high-quality real-world data [28-31]. This study demonstrates how synthetic data, derived from FEA, can be used to supplement real-world data and overcome limitations in datasets for structural damage detection tasks.

Based on the identified research gap, the primary objective of this study is to investigate the feasibility and effectiveness of computer vision–based methods for detecting and localizing shear buckling damage in thin steel plates. Specifically, this study aims to: (i) develop an instance segmentation framework capable of identifying shear buckling regions using limited real-world image data; (ii) address data scarcity by generating synthetic buckling images through finite element analysis and enhancing their visual realism using 3D modeling techniques; and (iii) quantitatively evaluate the impact of combining real and synthetic datasets on detection performance.

2. Methodology

This study adopts a deep learning–based instance segmentation framework for detecting and localizing shear buckling damage in thin steel plates. Given the need for accurate localization, efficient inference, and robustness to limited training data, the YOLO11 instance segmentation model was selected as the core detection architecture. YOLO11 is a single-stage, real-time detection framework which is capable of simultaneously predicting bounding

boxes and pixel-level segmentation masks, making it well-suited for identifying spatially complex buckling patterns.

The overall methodology consists of four main stages: (1) dataset generation and collection, including real and synthetic images; (2) data preprocessing and annotation; (3) model training using the YOLO11 instance segmentation framework; and (4) model validation and performance evaluation. More details of each stage of the proposed methodology are described in the following subsections:

1. Acquiring an appropriate image database:

This involved gathering images of real-world buckling instances, as well as synthetic images produced through finite element analysis (FEA) and post-processing techniques.

2. Preprocessing the database:

The images were prepared by annotating the buckling instances and applying augmentation techniques to improve the robustness of the model.

3. Training the model using the prepared database:

A deep learning framework was utilized to train an instance segmentation model specifically designed for buckling detection.

4. Validating and testing the trained model:

The model was evaluated on a separate test dataset to assess its accuracy, precision, recall, and overall performance.

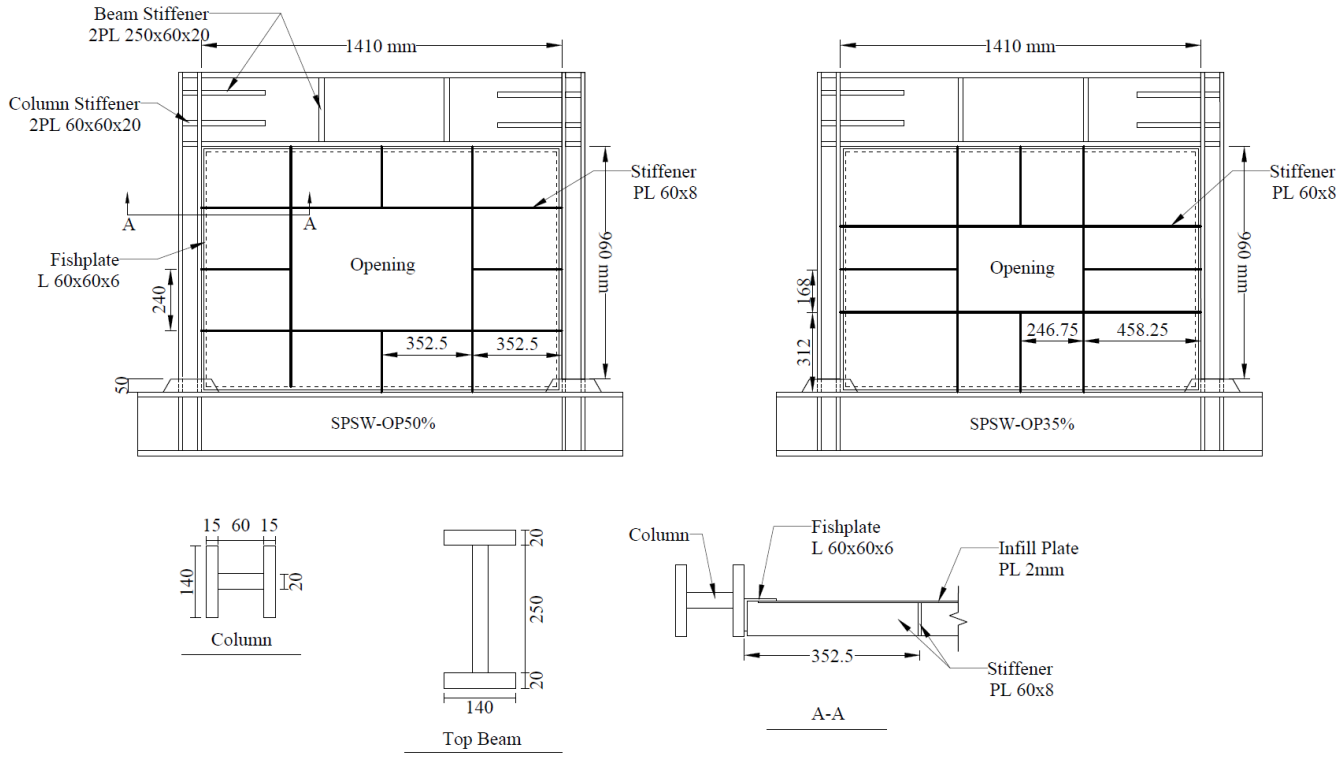

Fig. 1: schematics of the two test specimens [27]

2.1 Image Database Acquisition

The primary images for this study were obtained from an experiment conducted by the authors in 2023 [27]. The experiment was conducted on two one-third-scale steel plate shear wall (SPSW) specimens featuring a central rectangular opening. The specimens were stiffened with strategically placed stiffeners, dividing the plates into a number of subpanels. The schematics of the two specimens are presented in Fig. 1. The stiffener dimensions were selected to prevent global buckling of the plates. Thus, shear buckling occurred in the subpanels, which was at the early stages of cyclic loading. As a result, several instances of buckling were visible across the specimens. Fig. 2 shows a subpanel buckling as an example.

Hundreds of photographs were captured during all stages of the experiment from various angles, creating a considerable archive of images. These photos were taken using a Canon G7 X Mark II camera, which features a 20.1-megapixel resolution and provides high-quality images. However, not all the images were suitable for training the object detection model. For instance:

- Some photos were taken during the initial cycles of loading, where buckling on the subpanels had not yet developed.
- Other photos focused on features of the specimens that did not include visible buckling in the subpanels.

After a careful review of all the photographs, 188 images were selected for the task. While there is no strict minimum

number of images required for object detection tasks, it is generally recommended to use as many images as possible to improve model accuracy.

Fig. 2: Subpanel buckling during experiment [27]

A notable limitation of the selected dataset was that all images originated from two experiments with very similar setups and specimens. Despite the variation introduced by capturing photos from different angles, viewpoints, and lighting conditions, the images shared many similar parameters. This lack of diversity could result in a trained model that performs well only on images resembling those in the dataset, thereby limiting its ability to generalize to broader scenarios. To address these challenges, synthetic images were generated to:

- Increase the number of images in the training set.
- Introduce greater variation in image characteristics, thereby enhancing the model's ability to generalize across diverse conditions.

This approach aimed to overcome the dataset limitations and improve the model's performance for detecting buckling in real-world situations.

In order to use synthetic images of steel plate buckling for model training, it was crucial to ensure that the images accurately represented the shape and characteristics of shear buckling. To achieve this, finite element analysis (FEA) was chosen for image generation. FEA is widely recognized for its ability to predict the behavior of engineering models with high accuracy, making it an ideal method for generating realistic buckling shapes.

Several finite element (FE) models were developed for this study. These models were developed using ABAQUS [32] and had different specifications, such as the presence or absence of rectangular openings and different numbers of stiffeners, which resulted in subpanels of varying sizes in each model. Six examples of different variations for the considered models are shown in Fig. 3. The modeling approach employed here was similar to the previous approach used by the authors [27, 36].

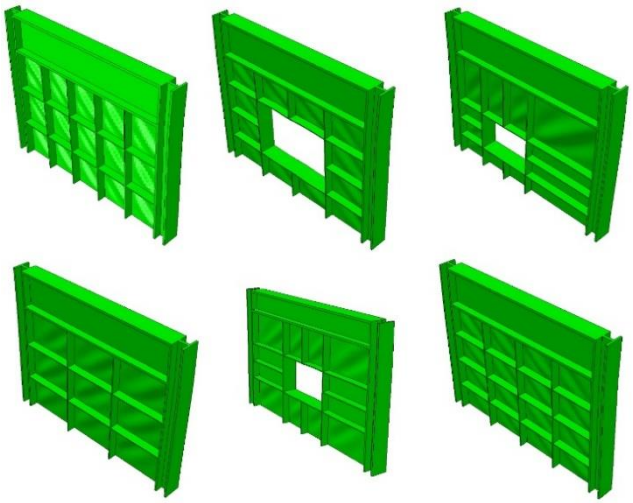

Fig. 3: Six examples of FE models variations considered for the study

In most FEA studies on steel plate shear walls (SPSWs), the structural members (plate, beams, and columns) are modeled using shell elements [33–37]. This is a rational approach since these studies typically focused on parameters such as shear strength or stiffness, and shell elements provide accurate results without unnecessarily increasing the computational costs. However, since the objective of this study was to generate visually realistic images that closely resemble real-world buckling, the models were created using 3D solid elements, which allow for a more faithful representation of local geometry, thickness, and buckling deformation patterns. The base of the model was fully fixed, while the out-of-plane displacement of the top beam was restrained. To ensure the initiation of shear buckling, a small initial geometric imperfection was introduced into the infill plate, consistent with common practice in buckling simulations.

Another critical consideration in creating FE models was the mesh size. In traditional FEA studies, the mesh size is usually chosen to be as large as possible while maintaining accuracy, which minimizes computational costs. However, for this study, larger mesh sizes would have reduced the resolution of the models, leading to polygonal shapes that appeared unnatural (as shown in Fig. 4). Conversely, opting for an excessively fine mesh size to achieve a smooth and high-resolution model would have resulted in unreasonable computational costs and analysis times, particularly given the complexity of 3D solid elements. Therefore, it was decided to adopt a balanced approach. A mesh size that maintained a realistic resolution without excessive computational burden was selected, and further enhancements were applied to the results outside the ABAQUS environment.

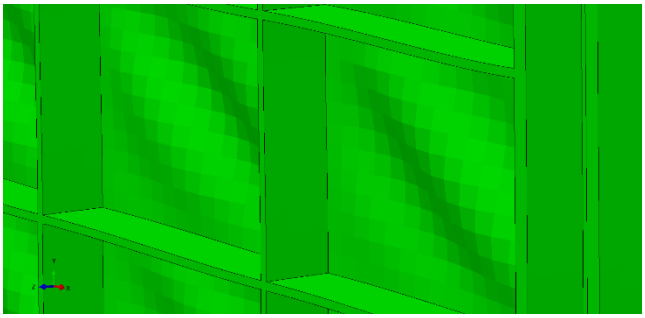

Fig. 4: An example of mesh sizes effect on buckling shape, making it appear low-resolution and unnatural

Additional domain adaptation steps were necessary due to the inherent limitations of ABAQUS as an engineering software. While ABAQUS excels at solving complex engineering problems, its capabilities for producing visually realistic images, such as natural lighting and textures, are very limited. The resulting images, although accurate from an engineering perspective, were clearly computer-generated and lacked the visual fidelity required for this study. Therefore, post-processing enhancements were applied to improve the realism of the images, ensuring they closely resembled real-world scenarios.

To enhance the FE results and adapt them for training purposes, the models were exported to Blender [38], a powerful 3D modeling software capable of producing highly realistic images. By using Blender, one of the primary issues, which was the low resolution of the raw FE models, was solved. This issue was mitigated by applying shade smoothing, a technique that interpolates surface normals to create a smooth appearance for low-resolution areas without changing the original geometry of the model.

Additionally, realistic textures were applied to the models to improve their visual representation further. These textures, combined with various lighting setups, were used to simulate different real-world conditions. Lights of varying intensities and orientations were applied from multiple angles, creating a diverse dataset that not only enhanced the realism of the images but also introduced variety, thereby addressing the model's ability to generalize to different conditions.

To further enhance the diversity of the synthetic dataset, Blender-rendered images were created with a blurry abstract background, unlike the FEA images, which had a simple white background. This approach was intended to mimic real-world conditions, where structural photographs often include complex and varied backgrounds. By introducing this variation, the model was better equipped to generalize to diverse scenarios.

The images were rendered using Blender's Cycles engine, which utilizes ray-tracing technology to achieve high visual fidelity. This process significantly improved the realism of the synthetic images compared to the raw FE outputs. The added diversity in textures and lighting ensured that the synthetic dataset captured a wide range of potential scenarios, enhancing the robustness of the trained model.

Despite the superior quality of the images rendered in Blender, the raw images from ABAQUS were also saved and included in the dataset. This allowed for testing the comparative effects of raw and enhanced images on the model's training performance. Fig. 5 illustrates this comparison, showing a raw FEA image alongside a rendered image from Blender. The images are from the same model and are captured from a similar angle and viewpoint.

Fig. 5: An FEA model (Right) compared to the enhanced version of the same model (Left)

2.2 Preprocessing the Database

Careful examination of the images revealed that annotating them with rectangles might not produce tight enough bounding boxes around the buckled areas for detection models to accurately learn the characteristics of buckling. To address this, all images were annotated using polygons to capture the irregular shapes of buckling instances more accurately. These annotations were applied carefully across all real-world, FEA, and Blender-generated images.

To improve the visual realism of FEA images, grayscaling was applied to them. While this adjustment did not fully bridge the gap between synthetic and real images, it enhanced their representation by a small but meaningful degree.

To evaluate the effects of combining real and synthetic images for model training, five distinct datasets were created:

1. A dataset containing only real-world images.
2. A dataset containing only raw FEA images.
3. A dataset containing only Blender-rendered images.
4. A dataset combining real-world and raw FEA images.
5. A dataset combining real-world and Blender-rendered images.

The first dataset only contains the images from the experiment. The second dataset is constructed exclusively from the images of various finite element models. The third

dataset contains only the Blender-rendered images. Since Blender provided the opportunity to introduce more variety to the FE models' visual appearances (e.g., different lighting conditions and various textures), more images were generated and included in the third database in comparison to the second database. The fourth dataset combines the experiment and FE model images (datasets 1 and 2) and the last dataset combines the experiment and Blender-rendered images (datasets 1 and 3). Each dataset was then split into training, validation, and test subsets. To ensure fair assessment and comparison of performance across all datasets, the test set was kept consistent for all configurations. Since the primary goal of the study was to detect real instances of buckling, only real-world images were included in the test set.

To further enhance the variety of the test set, 20 images were acquired from a different experimental study conducted in 2012 on stiffened SPSWs [39]. These photos, provided by the authors of the mentioned study, featured different specimens and testing setups from those in the primary dataset. This inclusion helped evaluate the models' generalization capabilities under diverse conditions.

As a result, a test dataset of 40 images was created, consisting of 20 real-world images from the current study and 20 images from the 2012 experiment. Careful measures were taken to ensure that none of the real-world images in the test set overlapped with the training or validation sets. Finally, approximately 15% of each dataset was allocated for validation, while the remaining images were used for training.

In addition to the annotated images, a small number of background images without any buckling instances were included in each dataset, including the test set. This was done to evaluate the model's ability to avoid false positives by correctly identifying non-damaged areas as negative cases. These background images introduce additional variety and contribute to testing the robustness and generalization capability of the trained models. Table 1 provides a summary of the number of images and the total number of buckling instances in each dataset.

Table 1: summary of the number of images and the total number of buckling instances in each dataset

Dataset Number	Dataset Name	Number of Images	Buckling Instances
1	Real	208	1294
2	FEA	141	1292
3	Blender	343	2598
4	Real-FEM	330	2586
5	Real-Blender	551	3892

2.3 Model Training

Recent advancements in object detection and instance segmentation have led to the development of numerous high-performing models. Among these, the groundbreaking algorithm You Only Look Once (YOLO), introduced in 2015 [40], has gained widespread recognition for its efficiency and robustness. Unlike traditional two-stage approaches, YOLO processes an entire image in a single pass, offering a well-balanced combination of speed and accuracy. Since its inception, several iterations of YOLO have been developed, each improving upon aspects such as speed, accuracy, generalization, and feature extraction capabilities.

The latest version, YOLO11, has demonstrated superior performance in key areas, including computational efficiency, precision, speed and compact model size [41, 42]. Thus, it was selected for this study, and instance segmentation was employed to train the YOLO11 model. Unlike traditional object detection, which identifies and classifies objects with bounding boxes, instance segmentation provides pixel-level segmentation masks for each detected object. All datasets were carefully formatted to meet YOLO11's input requirements. For each dataset, a separate model was trained over 300 epochs. Throughout the training process, validation results were closely monitored to ensure that overfitting did not occur. A batch size of 16 was used for training all models. YOLO11 automatically resizes images to 640 × 640 pixels, and this default setting was retained to reduce computational complexity. The training optimizer was set to auto-selection, which defaults to AdamW for iterations under 10,000. During training, metrics such as precision and recall were monitored, along with mAP (mean Average Precision). Although mAP's significance might appear less prominent in single-class scenarios like this study, it was included because it still provides valuable insights into model performance.

3. Results

Following the training phase, the models were validated using a test set that had remained unviewed throughout the training process. This approach was adopted to guarantee an unbiased evaluation. To comprehensively assess and compare the performance of each model, a variety of performance metrics were examined. These metrics provide different perspectives into the models' capabilities to detect buckling instances accurately and reliably. For this study, the following metrics were specifically considered:

1. Precision: Precision measures the ratio of correctly identified buckling instances out of all instances detected by the model. It reflects the model's ability to avoid false positives, ensuring that the detected buckling areas are indeed correct. High

precision is crucial in applications where false alarms could lead to unnecessary inspections or interventions.

2. Recall: Recall assesses the proportion of actual buckling instances in the images that were correctly identified by the model. It indicates the model's sensitivity, ensuring that critical damage instances are not overlooked. This metric is particularly important in structural health monitoring, as missed detections could pose significant safety risks.
3. F1 Score: The F1 score is the harmonic mean of precision and recall, effectively balancing the trade-off between false positives and false negatives. It serves as a useful summary metric in situations where both precision and recall are critical.
4. Mean Average Precision (mAP): Two variations of mAP were considered in this study to provide a more detailed evaluation of the model's performance:
    - mAP50: This metric calculates Mean Average Precision at a fixed Intersection over Union (IoU) threshold of 0.50. It evaluates the model's accuracy under relatively lenient conditions, where the overlap between the predicted bounding box and the ground truth only needs to exceed 50%. It is particularly useful for assessing the model's ability to detect "easy" instances of buckling. High mAP50 values indicate that the model is effective at identifying obvious or clear buckling areas, making it a valuable metric for initial assessments.
    - mAP50-95: This metric is the average of mAP values calculated at IoU thresholds ranging from 0.50 to 0.95 in increments of 0.05. It provides a more comprehensive evaluation of the model's performance by accounting for varying levels of detection difficulty. Unlike mAP50, mAP50-95 evaluates the model's ability to maintain accuracy even under stricter conditions, where the overlap between predicted and ground truth bounding boxes must be closer to perfect. This metric is particularly important for evaluating generalization and robustness, as it considers both straightforward and challenging detection scenarios.

By examining both mAP50 and mAP50-95, it is possible to gain valuable insights into the trade-offs between detecting easy instances and maintaining consistent performance across a range of detection challenges. The inclusion of these metrics ensures that the evaluation captures both high-level accuracy and nuanced detection capabilities.

The analysis of each metric was conducted to assess the strengths and weaknesses of the models trained on different datasets. This evaluation not only helps identify the most effective dataset for training but also provides insights into the model's ability to generalize to diverse scenarios. The results of the performance metrics are summarized in two tables Table 2 and Table 3). Table 2 presents the bounding box metric values, while Table 3 provides the mask (segmentation) metric values. The model numbers in both tables correspond directly to the dataset numbers listed in Table 1. This separation highlights the performance differences between object detection and instance segmentation tasks.

Table 2*:* bounding box metric values

Model Number	Precision	Recall	F1 Score	mAP50	mAP50-95
1	0.88	0.83	0.86	0.89	0.63
2	0.29	0.29	0.29	0.24	0.16
3	0.54	0.29	0.37	0.32	0.22
4	0.89	0.82	0.86	0.91	0.64
5	0.92	0.87	0.89	0.92	0.67

Table 3: mask (segmentation) metric values

Model Number	Precision	Recall	F1 Score	mAP50	mAP50-95
1	0.87	0.81	0.84	0.87	0.56
2	0.29	0.29	0.29	0.23	0.14
3	0.54	0.29	0.37	0.33	0.20
4	0.88	0.81	0.84	0.9	0.58
5	0.90	0.84	0.87	0.9	0.6

4. Discussion

4.1 Interpretation of Results

The results highlight the strengths and limitations of each model based on the datasets used for training. The first model, trained exclusively on real-world images, achieved high values across all metrics, including precision (0.88 for bounding box, 0.87 for mask), recall (0.83 for bounding box, 0.81 for mask), and mAP50 (0.89 for bounding box, 0.87 for mask). This indicates that real-world images alone can provide a strong baseline for buckling detection due to their direct relevance to the target domain. The high values suggest that the model effectively generalizes to unseen real-world test images.

For the second model, trained only on raw FEA images, the results were notably poor across all metrics, with precision,

recall, and F1 score hovering around 0.29 and mAP50 values of 0.24 (bounding box) and 0.23 (mask). These results are expected given the lack of real-world features in the dataset, which limits the model's ability to generalize. This highlights the gap between raw synthetic images and real-world applicability.

The third model, trained on Blender-rendered synthetic images, demonstrated better performance than the second model, with precision improving to 0.54 for the bounding box and 0.54 for the mask, and mAP50 increasing to 0.32 (bounding box) and 0.33 (mask). This improvement highlights the value of using enhanced synthetic images with higher visual fidelity. The rendering process in Blender, which introduced textures and lighting variations, appears to have narrowed the domain gap between synthetic and real-world images, making the dataset more effective for training.

The fourth model, trained on a combination of real-world and raw FEA images, achieved performance metrics similar to the first model, with only slight improvements in mAP50 (0.91 for bounding box, 0.9 for mask). While this indicates that raw FEA images do not significantly boost performance when combined with real-world images, their inclusion does not appear to detract from the model's ability to generalize either. This suggests that the raw FEA dataset provides little additional diversity or domain adaptation benefits.

The fifth model, trained on a combination of real-world and Blender-rendered synthetic images, achieved the best overall performance across all metrics. The bounding box mAP50 improved to 0.92, and the mask mAP50 increased to 0.9, with similar improvements in precision, recall, and F1 scores. These results confirm that incorporating enhanced synthetic images with real-world data provides meaningful gains in detection accuracy, likely due to the increased diversity and realism introduced by the Blender-generated dataset.

Comparing the bounding box metrics (Table 2) with the mask metrics (Table 3) reveals a general decrease in performance for mask predictions across all models. For instance, the first model's mAP50 decreased from 0.89 (bounding box) to 0.87 (mask). This trend suggests that detecting the precise outlines of buckling instances (mask prediction) is inherently more challenging than identifying the bounding boxes. However, the relatively smaller differences in mAP50 values between the bounding box and mask metrics for the fifth model indicate that the added realism and diversity in the Blender-enhanced dataset also benefited instance segmentation.

Fig. 6 illustrates the prediction results of each model for a sample image from the test set (explained in section 2.2), highlighting their respective strengths and weaknesses. This figure also serves as a visual reference for comparing the performance of the models. As it is observed, Model 5 has the best performance, missing only two of the buckling instances. In contrast, Model 2 performed very poorly, detecting a lot of false positives. A few false positives are also observed in the Model 3 prediction. However, Model 3 performed generally better than Model 2. The poor performance of these two models were expected since they are trained on datasets only constructed from synthetic images. The metric results in Table 2 and Table 3 also reflect this fact.

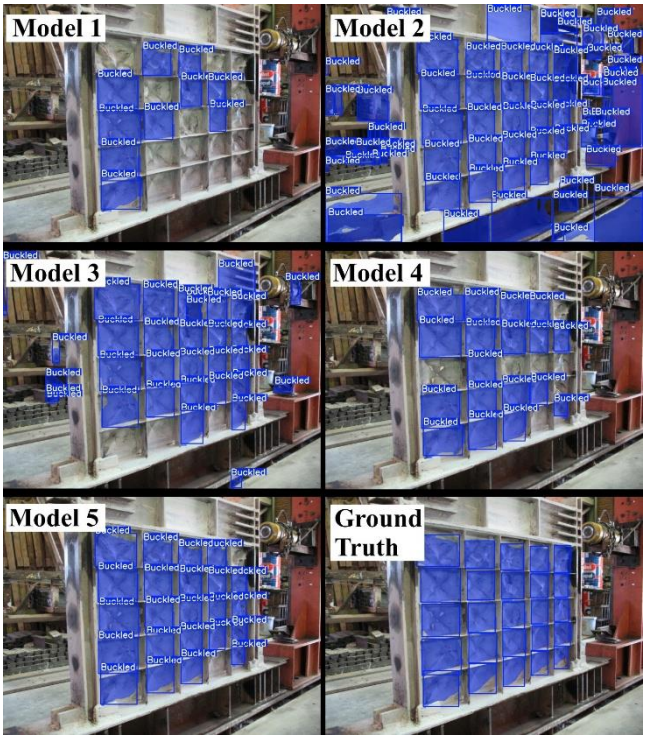

Fig. 6: Prediction results of each model for a sample image from the test set

4.2 Practical Implications

These results demonstrate the critical importance of dataset composition for training object detection and segmentation models. The combination of real-world and visually enhanced synthetic images (Model 5) outperformed all other datasets, highlighting the role of domain adaptation in improving model performance. The high precision (0.92 for bounding box, 0.90 for mask) and recall (0.87 for bounding box, 0.84 for mask) values achieved by Model 5 are particularly noteworthy. High precision indicates that the model effectively minimizes false positives, ensuring that detected buckling instances are highly likely to be accurate. Meanwhile, high recall reflects the model's ability to detect most of the true buckling instances in the test set, reducing the risk of missing critical damage. Together, these metrics highlight Model 5's potential for real-world applications, where both accurate detection and comprehensive coverage are crucial for structural health monitoring and safety assessments.

4.3 Limitations of the Study

Despite the promising results, several limitations should be acknowledged. First, although synthetic data were used to augment the dataset, the number of real-world images remained relatively limited and originated from laboratory experiments. As a result, the trained models may still face challenges when applied to field conditions with complex backgrounds, occlusions, and uncontrolled lighting.

Additionally, the synthetic images were generated based on a specific range of geometries and boundary conditions. While visual enhancement improved realism, the diversity of structural configurations remains constrained. Finally, the study employed a single deep learning architecture (YOLO11), and alternative architectures or hybrid approaches were not explored.

4.4 Future Research Directions

Future work could address these limitations by expanding the dataset to include field-acquired images from real post-earthquake inspections that would further improve model robustness and real-world applicability. Moreover, investigating domain adaptation techniques and alternative instance segmentation architectures could further enhance performance and generalization. These extensions would broaden the applicability of the proposed methodology to a wider range of structural health monitoring problems.

In addition to detecting the presence and location of shear buckling, future research could focus on quantifying the severity of buckling damage and linking it to performance-based decision criteria. By extracting geometric features from predicted segmentation masks (such as the affected area), it may be possible to estimate damage severity levels. These metrics could then be correlated with structural performance indicators or experimental damage thresholds to support automated decisions regarding repair, strengthening, or replacement of steel plate shear walls. Such an extension would move computer vision–based damage detection beyond qualitative assessment toward practical, decision-oriented structural health monitoring frameworks.

5. Conclusions

This research explored the application of computer vision for detecting shear buckling damage in thin steel plates. This is an area that has received minimal attention compared to other types of structural damage. Despite significant advancements in computer vision-based damage detection, the challenge of addressing buckling damage remains largely unexplored, partly due to the scarcity of image data, which is necessary for training robust models. To overcome this limitation, the study proposed a data augmentation framework that combines real experimental images with synthetic images generated through finite element analysis and enhanced using 3D modeling techniques. Instance segmentation models based on the YOLO11 architecture were trained and evaluated using multiple datasets to assess the effectiveness of the proposed approach.

The main findings of this study can be summarized as follows:

- A computer vision model trained solely on a limited set of real-world images was able to achieve high detection performance, demonstrating that reliable shear buckling detection is feasible even with relatively small experimental datasets.
- Models trained exclusively on synthetic data exhibited inferior performance, highlighting that raw synthetic image alone are insufficient for reliable generalization to real-world conditions.

- Enhancing synthetic images using 3D rendering significantly improved model performance compared to raw finite element images, confirming the importance of visual realism and domain adaptation when using synthetic data.
- The combined dataset of real-world and enhanced synthetic images produced the best overall performance, achieving the highest values across all evaluation metrics, including precision, recall, F1 score, mAP50, and mAP50–95.

- Compared to the real-only dataset, the combined dataset resulted in consistent quantitative improvements, with increases of approximately 4% across multiple segmentation-based performance metrics, demonstrating the effectiveness of the proposed data augmentation strategy.
- Segmentation metrics were consistently lower than bounding box metrics, indicating that accurately delineating buckling boundaries remains more challenging than localizing damaged regions, yet still benefited from the inclusion of enhanced synthetic data.

Overall, this research demonstrates the feasibility of automated shear buckling detection using computer vision and highlights the value of high-fidelity synthetic data for overcoming data scarcity in structural health monitoring applications. The proposed framework provides a foundation for future studies aimed at damage severity assessment, decision-oriented inspection tools, and broader applications of synthetic data generation in civil infrastructure monitoring.

Acknowledgements

The authors would like to express their gratitude to the authors of the 2012 experimental study on stiffened steel plate shear walls for generously providing the images used in the test set of this study. Their contribution was invaluable in evaluating the model's performance in diverse real-world scenarios and enhancing the quality of this research.

References

[1] L.J. Thorburn, G.L. Kulak, C.J. Montgomery, Analysis and Design of Steel Shear Wall Systems, Structural Engineering Report No. 107, Department of Civil Engineering, 1983.

[2] R.G. Driver, G.L. Kulak, D.L. Kennedy, A.E. Elwi, Cyclic test of four-story steel plate shear wall, Journal of Structural Engineering 124(2) (1998) 112-120.

[3] A.K. Bhowmick, R. Driver, G. Grondin, Seismic analysis of steel plate shear walls considering strain rate and p delta effects, Journal of Constructional Steel Research 65 (2009) 1149-1159.

[4] J.W. Berman, M. Bruneau, Experimental investigation of light-gauge steel plate shear walls, Journal of Structural Engineering 131(2) (2005) 259-267.

[5] A. Lubell, H. Prion, C. Ventura, M. Rezai, Unstiffened Steel Plate Shear Wall Performance under Cyclic Loading, Journal of Structural Engineering-asce 126 (2000) 453-460.

[6] R. Purba, M. Bruneau, Seismic performance of steel plate shear walls considering two different design philosophies of infill plates. II: Assessment of collapse potential, Journal of Structural Engineering 141(6) (2015) 04014161.

[7] A. Astaneh-Asl, Q. Zhao, Cyclic behavior of steel shear wall systems, Annual Stability Conference, Structural Stability Research Council Seattle, 2002, pp. 18015-3191.

[8] Qarib, H., & Adeli, H. (2014). Recent advances in health monitoring of civil structures. Scientia Iranica, 21(6), 1733-1742.

[9] Wang, T., Song, G., Liu, S., Li, Y., & Xiao, H. (2013). Review of bolted connection monitoring. International Journal of Distributed Sensor Networks, 9(12), 871213.

[10] Sofi, A., Regita, J. J., Rane, B., & Lau, H. H. (2022). Structural health monitoring using wireless smart sensor network–An overview. Mechanical Systems and Signal Processing, 163, 108113.

[11] Bado, M. F., & Casas, J. R. (2021). A review of recent distributed optical fiber sensors applications for civil engineering structural health monitoring. Sensors, 21(5), 1818.

[12] Hubbard, P. G., Xu, J., Zhang, S., Dejong, M., Luo, L., Soga, K., ... & Minto, C. (2021). Dynamic structural health monitoring of a model wind turbine tower using distributed acoustic sensing (DAS). Journal of civil structural health monitoring, 11(3), 833-849.

[13] Zhang, C., Mousavi, A. A., Masri, S. F., Gholipour, G., Yan, K., & Li, X. (2022). Vibration feature extraction using signal processing techniques for structural health monitoring: A review. Mechanical Systems and Signal Processing, 177, 109175.

[14] Luo, Y., Chen, Y., Wan, H. P., Yu, F., & Shen, Y. (2021). Development of laser-based displacement monitoring system and its application to large-scale spatial structures. Journal of Civil Structural Health Monitoring, 11, 381-395.

[15] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

[16] Deng, J., Singh, A., Zhou, Y., Lu, Y., & Lee, V. C. S. (2022). Review on computer vision-based crack detection and quantification methodologies for civil structures. Construction and Building Materials, 356, 129238.

[17] Ai, D., Jiang, G., Lam, S. K., He, P., & Li, C. (2023). Computer vision framework for crack detection of civil infrastructure—A review. Engineering Applications of Artificial Intelligence, 117, 105478.

[18] Hu, D., Yee, T., & Goff, D. (2024). Automated crack detection and mapping of bridge decks using deep learning and drones. Journal of Civil Structural Health Monitoring, 14(3), 729-743.

[19] Liu, J., Sun, H., Liu, H., Yue, Q., Xu, Z., Jia, Y., & Wang, S. (2025). Recognition and quantification of apparent damage to concrete structure based on computer vision. Measurement, 240, 115635.

[20] Nguyen, H., & Hoang, N. D. (2022). Computer vision-based classification of concrete spall severity using metaheuristic-optimized extreme gradient boosting machine and deep convolutional neural network. Automation in Construction, 140, 104371.

[21] Abdelkader, E. M., Moselhi, O., Marzouk, M., & Zayed, T. (2020). Evaluation of spalling in bridges using machine vision method. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 37, pp. 1136-1143). IAARC Publications.

[22] Nabizadeh, E., & Parghi, A. (2023). Automated corrosion detection using deep learning and computer vision. Asian Journal of Civil Engineering, 24(8), 2911-2923.

[23] Xu, J., Gui, C., & Han, Q. (2020). Recognition of rust grade and rust ratio of steel structures based on ensembled convolutional neural network. Computer-Aided Civil and Infrastructure Engineering, 35(10), 1160-1174.

[24] Lao, W., Cui, C., Zhang, D., Zhang, Q., & Bao, Y. (2023). Computer Vision-Based Autonomous Method for Quantitative Detection of Loose Bolts in Bolted Connections of Steel Structures. Structural Control and Health Monitoring, 2023(1), 8817058.

[25] Pan, X., Vaze, S., Xiao, Y., Tavasoli, S., & Yang, T. Y. (2022, May). Structural Damage Detection of Steel Corrugated Panels Using Computer Vision and Deep Learning. In Canadian Society of Civil Engineering Annual Conference (pp. 323-336). Cham: Springer Nature Switzerland.

[26] Malekloo, A., Liu, X. C., & Sacharny, D. (2024). AI-enabled airport runway pavement distress detection using dashcam imagery. Computer-Aided Civil and Infrastructure Engineering.

[27] Sabouri-Ghomi, M., Bhowmick, A., & Sabouri-Ghomi, S. (2023). Experimental and Numerical Study of Stiffened Steel Plate Shear Walls with Rectangular Openings. Journal of Structural Engineering, 149(12), 04023180.

[28] Dewi, C., Chen, R. C., Liu, Y. T., Jiang, X., & Hartomo, K. D. (2021). Yolo V4 for advanced traffic sign recognition with synthetic training data generated by various GAN. IEEE Access, 9, 97228-97242.

[29] Man, K., & Chahl, J. (2022). A review of synthetic image data and its use in computer vision. Journal of Imaging, 8(11), 310.

[30] Paulin, G., & Ivasic-Kos, M. (2023). Review and analysis of synthetic dataset generation methods and techniques for

application in computer vision. Artificial intelligence review, 56(9), 9221-9265.
[31] Zhai, G., Narazaki, Y., Wang, S., Shajihan, S. A. V., & Spencer, B. F. (2022). Synthetic data augmentation for pixel-wise steel fatigue crack identification using fully convolutional networks. Smart Struct Syst, 29(1), 237-250.
[32] ABAQUS 6.14. Dassault Systemes Simulia
[33] Barua, Kallol, and Anjan K. Bhowmick. "Nonlinear seismic performance of code designed perforated steel plate shear walls." Steel and Composite Structures, An International Journal 31, no. 1 (2019): 85-98.
[34] Bhowmick, A. K. (2014). Seismic behavior of steel plate shear walls with centrally placed circular perforations. Thin-Walled Structures, 75, 30-42.
[35] Purba, R., & Bruneau, M. (2009). Finite-element investigation and design recommendations for perforated steel plate shear walls. Journal of structural engineering, 135(11), 1367-1376.
[36] Sabouri-Ghomi, M., & Bhowmick, A. K. (2023). Analytical and numerical investigation of stiffened steel plate shear walls. Asian Journal of Civil Engineering, 24(6), 1841-1857.
[37] Sabouri-Ghomi, S., Ahouri, E., Sajadi, R., Alavi, M., Roufegarinejad, A., & Bradford, M. A. (2012). Stiffness and strength degradation of steel shear walls having an arbitrarily-located opening. Journal of Constructional Steel Research, 79, 91-100.
[38] Blender Development Team. (2024). Blender (Version 4.2.3) [Computer software]. https://www.blender.org
[39] Sabouri-Ghomi, S., & Sajjadi, S. R. A. (2012). Experimental and theoretical studies of steel shear walls with and without stiffeners. Journal of constructional steel research, 75, 152-159.
[40] Redmon, J. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition.
[41] Jegham, N., Koh, C. Y., Abdelatti, M., & Hendawi, A. (2024). Evaluating the evolution of yolo (you only look once) models: A comprehensive benchmark study of yolo11 and its predecessors. arXiv preprint arXiv:2411.00201.
[42] Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725.