Automated Structural Surface Damage Identification, Classification and Severity Estimation Using Deep Learning Approaches

MRINMOY KUMAR DAS, NIKHITA RAPOLU, DHATHRI MEDA, PRAFULLA KALAPATAPU and VENKATA DILIP KUMAR PASUPULETI

ABSTRACT

The maintenance of civil infrastructure has become a major challenge for modern engineers due to its continuous growth, especially buildings and bridges which are susceptible to structural damage caused by factors like aging, design flaws, and natural disasters, which can seriously jeopardize their safety, health, and structural soundness. The conventional inspection methods are not only costly and time-consuming, but also pose safety hazards, rendering them ineffective. The paper introduces an automated method for detecting and assessing the severity of damages in buildings and bridges, which aims to address and mitigate the limitations of traditional techniques. A collection of 5000 images, ranging in size from 416x416 to 640x640 pixels, were gathered from damaged sites and annotated in polygon annotation format, with damages categorized into three classes: spalling, corrosion, and crack. To expand the dataset and enhance the precision of the deep learning models, data augmentation techniques from the data Albumentation library were utilized for image processing. Several object-based instance segmentation deep-learning models, including Yolo V5, V7, V8 Instance Segmentation, and Mask-RCNN, were trained on 80% of the dataset to obtain the coordinates of the damaged area's outline and generate masks for the detected damages. The area of these masks and the percentage of damage severity in an image were computed. The trained models achieved an accuracy range of approximately 60% to 70%, indicating the potential effectiveness of instance segmentation deep learning models. The proposed approach offers a quick and efficient method for determining the severity of structural damages, resulting in improved safety and decreased maintenance costs for buildings and bridges. To enhance the accuracy and precision of the model, future research could include a larger dataset with additional types of structural failures.

Mrinmoy Kumar Das[1], Nikhita Rapolu[1], Dhathri Meda[1], Prafulla Kalapatapu[2] and Venkata Dilip Kumar Pasupuleti[2]
[1]UG Student, [2]Faculty
École Centrale School of Engineering, Mahindra University, Hyderabad, India
das20ucse101@mahindrauniversity.edu.in,
nikhita20ucse115@mahindrauniversity.edu.in,
dhathri20ucse040@mahindrauniversity.edu.in,
prafulla.kalapatapu@mahindrauniversity.edu.in,
venkata.pasupuleti@mahindrauniversity.edu.in

Furthermore, the automated approach can be integrated into mobile monitoring and inspection devices, such as drones, as well as street camera systems to identify damaged buildings following natural disasters. An alarm can be triggered based on a specific threshold value, and appropriate actions can be taken to reduce the response time for emergency management and improve the structures' resilience to natural disasters.

INTRODUCTION AND BACKGROUND

Many structures like buildings, bridges, and dams are susceptible to damages due to environmental conditions like aging, exposure to fire, chemicals, water, and other factors. These pose a threat to human life, since, they lose their structural integrity and are likely to collapse [1]. It is also a challenge for municipal governments to keep track of these efficiently. Hence, detection of these damages and calculation of their severities can help in determining the functionality and reliability of these structures. The most widely used methods are semi-automated approaches like collecting data manually and storing them in a database. But these approaches are time-consuming, costly, and rely on human intelligence. Advancements in computer-vision and deep learning have improved the evaluation of surface defects, particularly for detecting and assessing spalling in reinforced concrete bridges. Abdelkader et al. implemented a self-adaptive optimization-based method for automated detection and evaluation of spalling severities, which provided superior segmentation accuracies compared to conventional methods [1]. An important approach in segmentation is using Gabor filters, it is one of the most recent developments. It separates an object into distinctive regions using Principal Component Analysis (PCA) and K-Means Clustering [2]. To improve the speed and efficiency of damage detection, this paper utilizes deep learning models for instance segmentation like Mask R-CNN (Regional Convolutional Neural Networks), Yolo (You Only Look Once) V5, V7, V8 Instance Segmentation which segments individual objects in a scene. Instance segmentation is done based on similar attributes of instances such as texture, color, brightness, and distance index [3]. One promising approach is to use Convolutional Neural Network (CNN). For example, a study by Vundekode et al. used CNN to identify the types of damage in building structures [4]. The results showed that CNN was able to identify the different types of damage with a high degree of accuracy. Another study by Kandula et al. used a YOLO algorithm to detect, classify, and segment cracks in concrete structures [5]. Additionally, YOLO architecture was also able to detect structural damage in video footage with a high degree of accuracy [6]. Mask R-CNN is the first method to employ target detection and segmentation in one model [7]. Further, the Mask R-CNN model has demonstrated the feasibility of automatically locating and segmenting cracks and spalling in extreme events using 2D images with an accuracy between 67.6% and 81.1% [8]. YOLO V5, V7, V8 Instance Segmentation are real-time object detection systems that operate on a single CNN and have been designed for end-to-end training [9]. These models are able to segment various kinds of damages like spalling, corrosion, and cracks by generating outlines of the damages. These outlines are represented by contour points in Cartesian space, with a fairly reasonable accuracy. We have also calculated the severity of the damages by computing the area of damage and percentage of damage present in an image.

METHODOLOGY

Data Collection and Dataset Formation

The total number of images in the dataset is approximately around 5000 with the image sizes ranging from 416 x 416 to 768 x 768. Around 20% of the dataset consists of images from a project site and are annotated and labelled using a tool called LabelME, in which polygons are used to outline the boundaries of the cracks, spall, and corrosion. The remaining 80% of the dataset is the combination of existing crack dataset [10] and corrosion dataset [11]. The area represented by the closed region of these polygons are the damages in the images. Some examples of the annotated images are shown in Figure 1. In the images of Figure 1, cracks are represented by purple, and spall and corrosion are represented by yellow.

In order to increase the size of the dataset, the Albumentation library had been initially implemented, but Roboflow's data augmentation tool demonstrated ease of use and had been consequently implemented for data augmentation [12]. This technique uses pixel-level transformation and spatial transformation, using flipping, rotating, and shearing. Spatial level transformation was adopted in the method to pre-process the training data, as it can alter input images, masks and bounding boxes simultaneously [13]. This resultant final dataset consists of 13k images, and is split into train (80%) and test/validation (20%). In the dataset, there are 5240 cracks, 3310 corrosion, and 1041 spall images. The above numbers indicate that the dataset is biased towards crack and corrosion.

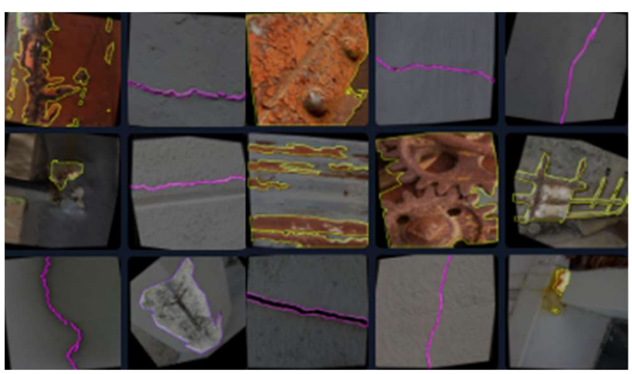


Figure 1: Examples of Annotated images in the Dataset [10] [11]

Model Training

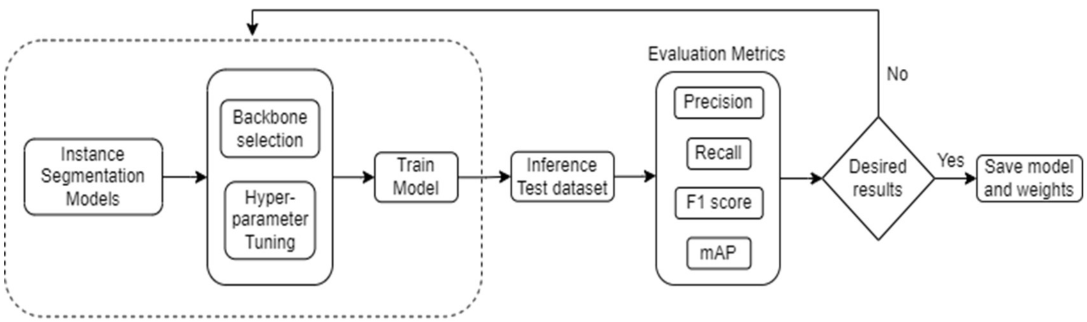


Figure 2: Model training process

There are numerous deep learning methods for instance segmentation. For our study, we used Mask R-CNN, Yolo V5, V7, and V8 Instance segmentation models. The Yolo models ranged from light-weight models to heavy models.

Mask R-CNN is an extension of Faster R-CNN, which is a region-based CNN, that returns bounding boxes for each detected object and its respective class label along with a confidence score [4]. Faster R-CNN model works in two stages: the first stage consists of two networks, backbone and Region Proposal Network (RPN) and the second stage consists of the network that predicts bounding boxes and object class for each of the proposed regions obtained from the first stage. VGG-16 has been used for the backbone of the model [14].

Yolo V5, V7, V8 Instance segmentation were the other models used for training and they use a single neural network to process an entire image [5]. The image is divided into regions and it detects the bounding boxes and outlines of objects for each region [15][16]. The above models were trained on a V100 Nvidia GPU card.

Damage and Severity Calculation

After image segmentation, we obtain the coordinates of the outline of the damage. The visual representation of the process used to calculate the area and severity is illustrated in Figure 3. The coordinates obtained after segmentation can be stored in a list. Masks are created based on these coordinates. Using pre-defined functions to calculate masks, the area of each mask is calculated**.** The total damage area is given calculated using the image dimensions, i.e., the width times the height of the image. Lastly, the damage percentage can be calculated.

$$Total\ damage\ area = area1 + area2 + \ldots + areaN \tag{1}$$

$$Image\ area = Size\ of\ image\ width \times Size\ of\ image\ height \tag{2}$$

$$Damage\ \% = (Total\ damage\ area\ /\ Image\ area) \times 100 \tag{3}$$

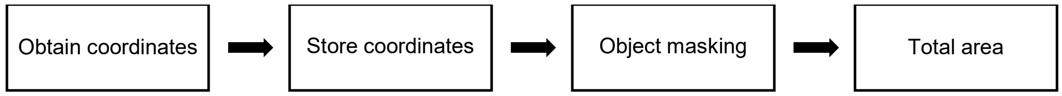


Figure 3: Process of obtaining damage severity

IMPLEMENTATION

Evaluation of the proposed models

The criteria used to evaluate the accuracy of the models are mainly Precision, Recall, F1 score, and predicted area. Precision is defined as true positive estimation over whole estimation, whereas Recall only provides positive estimation. The F1 score tells with what confidence the precision and recall values are scattered. The mAP (Mean Average Precision) summarizes the model. The following equations represent how the above parameters are computed.

$$Precision = \frac{TP}{(TP + FP)} \tag{4}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{5}$$

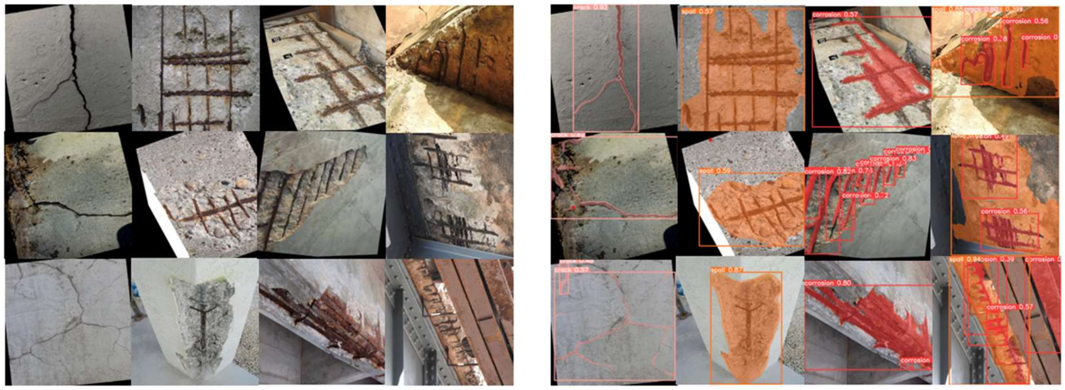
$$F1\ Score = \frac{2\ \times Recall \times Precision}{(Recall + Precision)} \tag{6}$$

Here, TP and FP stand for True Positive and False positive, while TN and FN stand for True negative and False negative, respectively [17].

Table I: COMPARISON OF PRECISION, RECALL, F1 SCORE, MAP-50, MAP-95, INFERENCE SPEED AND AREA PERCENTAGE OF DETECTED DAMAGE

Models		**Masks**						**Area (%)**
		P	**R**	**F1**	**mAP-50**	**mAP-95**	**Speed (ms)**	
Mask R-CNN	FPN	0.632	0.318	0.421	0.423	0.254	5.6	49.43
Yolo V5 Instance Segmentation	N	0.566	0.448	0.500	0.423	0.167	1.8	51.67
	M	0.670	0.538	0.596	0.518	0.232	5.9	54.09
	X	0.738	0.55	0.630	0.558	0.257	17.7	54.96
Yolo V7 Instance Segmentation	Seg	0.691	0.549	0.611	0.539	0.247	10.9	53.51
	SegX	0.687	0.555	0.616	0.550	0.254	17.4	53.48
Yolo V8 Instance Segmentation	N	0.614	0.498	0.549	0.506	0.213	1.9	52.83
	M	0.713	0.559	0.626	0.574	0.251	11.6	53.46
	X	0.717	0.568	0.633	0.584	0.264	20.6	54.03

Testing on test images



(a) (b)

Figure 4: (a) Input images for models [10][11], (b) Predicted damages with masks and bounding boxes drawn on the images

In the dataset, 20% of the images were used for testing. Some of the predictions made from the models are shown above in Figure 4. The following are depicted in Figure 4: the first column represents crack detection, the second column represents spall detection, the third column represents corrosion detection, and the fourth column represents a combination of spall and corrosion detections.

Failure cases

The trained models are unable to differentiate between cracks, spall, and corrosion with high confidence score during testing and validation. The major reason for the low confidence score is due to the insufficient training data as it does not cover diverse examples. This problem also leads to under-fitting results in real-world scenarios. For example, only one class of damage is detected, the damage is not completely detected, or there is a misclassification of the damage class. Some examples of wrong predictions from our dataset are show below in Figure 5.

Figure 5: Incorrect detections and classifications of damages

DISCUSSION

With the aim of providing a solution for automatically detecting and calculating severity of three types of structural damages, four deep-learning instance segmentation model architectures have been evaluated and validated on the dataset (test/validation dataset). The values for precision, recall, F1 score, and inference speed were also calculated for the models. The following are observed during analysis:

1. Data pre-processing challenges:
    - One of the important factors for the performance of instance segmentation models is the size of the dataset, and as the size of the dataset increases with more diverse examples, the validation results can be improved.
    - It is important to balance the ability of a model to generalize new data and training should be done in such a way that the model does not over fit to the training data. Moreover, the quantity and quality of images can drastically affect the performance of a model.
    - Collecting and annotating images in a dataset is both time-consuming and expensive.
    - Poor detection and segmentation of images containing spall is due to the dataset being unbalanced and biased towards crack and corrosion.

2. Inference from results:
    - Yolo V5x-seg has shown promising results with a precision of 73.8%, followed by Yolo V8x-seg and Yolo V8m-seg at 71.7% and 71.3% precision respectively.
    - Figure 4 demonstrates the outcomes produced from the 4 trained model architectures. The heavier models are able to segment better compared to the lighter models. However, the inference speed of lighter models is faster (around 2ms) than the heavier models (around 20ms).
    - The area of damage after prediction has a slight error which is due to factors like low confidence score, precision, F1 score, and overall accuracy. The difference between area calculated from the test dataset and Yolo V5x-seg is the lowest, indicating that despite the slight errors, the model has good performance in predicting area of damage.

CONCLUSIONS

In this paper, a dataset of approximately 5000 images was created. These images consisted of 3 classes of damages: Crack, Spall, and Corrosion. The dataset was split into 80% for training and 20% for testing and validation. Four different types of instance segmentation model architectures were trained and tested: Yolo V5, V7, V8 Instance Segmentation, and Mask R-CNN. The models ranged from lightweight models (like Yolov5n-seg, Yolov8n-seg) for mobile computing to heavier models (like Yolov5x-seg, Yolov7-segX, Yolov8x-seg) optimized for GPUs. The goal was to create an automated method for detecting and assessing the severity of damages.

Precision, recall, inference speeds, and area percentage of these models were compared. The trained models achieved precision in the range of approximately 60% to 70% and inference speeds ranging from 2ms to 21ms, indicating the potential effectiveness of instance segmentation deep learning models. Future research could include a larger dataset with additional types of structural failures, this will result in better training and also help in increasing the accuracy of detections and precision of outline of damages. Furthermore, the automated approach can be integrated into mobile monitoring and inspection devices, such as drones, as well as street camera systems to identify damaged buildings following natural disasters. An alarm can be triggered based on a specific threshold value, and appropriate actions can be taken to reduce the response time for emergency management and improve the structures' resilience to natural disasters.

REFERENCES

1. Mohammed Abdelkader, E., Moselhi, O., Marzouk, M., & Zayed, T. (2021). Entropy-based automated method for detection and assessment of spalling severities in reinforced concrete bridges. *Journal of Performance of Constructed Facilities*, *35*(1), 04020132.
2. Hoang, N. D. (2020). Image processing-based spall object detection using Gabor filter, texture analysis, and adaptive moment estimation (Adam) optimized logistic regression models. *Advances in Civil Engineering*, *2020*.
3. Hafiz, A. M., & Bhat, G. M. (2020). A survey on instance segmentation: state of the art. *International journal of multimedia information retrieval*, *9*(3), 171-189.
4. Vundekode, N. R., Kalapatapu, P., & Pasupuleti, V. D. K. (2021, January). A study on vision based method for damage detection in structures. In European Workshop on Structural Health Monitoring: Special Collection of 2020 Papers-Volume 1 (pp. 96-105). Cham: Springer International Publishing.
5. Kandula, H., Koduri, H. R., Kalapatapu, P., & Pasupuleti, V. D. K. (2022, June). Deep Convolutional Neural Network for Segmentation and Classification of Structural Multi-branch Cracks. In European Workshop on Structural Health Monitoring: EWSHM 2022-Volume 2 (pp. 177-185). Cham: Springer International Publishing.
6. Kadarla, S., Beeram, S. K., Kalapatapu, P., & Pasupuleti, V. D. K. (2021, January). Concrete crack detection from video footage for structural health monitoring. In European Workshop on Structural Health Monitoring: Special Collection of 2020 Papers-Volume 1 (pp. 79-88). Cham: Springer International Publishing.
7. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).
8. Bai, M., & Sezen, H. (2021, January). Detecting cracks and spalling automatically in extreme events by end-to-end deep learning frameworks. In *ISPRS Annals of Photogrammetry and Remote Sensing Spatial Information Science, XXIV ISPRS Congress, International Society for Photogrammetry and Remote Sensing*.
9. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 779-788)
10. University. (2022, December). Crack Dataset [Open Source Dataset]. Roboflow Universe. https://universe.roboflow.com/university-bswxt/crack-bphdr.
11. ComputerVision. (2023, April). Corrosion inspection Dataset [Open Source Dataset]. Roboflow Universe. https://universe.roboflow.com/computervision-a8zib/corrosion-inspection
12. Dwyer, B., Nelson, J. (2022), Solawetz, J., et. al. Roboflow (Version 1.0) [Software]. Available from https://roboflow.com. computer vision.
13. A. Buslaev, A. Parinov, E. Khvedchenya, V.I. Iglovikov and A.A. Kalinin, 2018. *Albumentations: fast and flexible image augmentations*, ArXiv, e-print = 1809.06839

14. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
15. Thuan, D. (2021). Evolution of Yolo algorithm and Yolov5: The State-of-the-Art object detention algorithm.
16. Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*.
17. Goutte, C., & Gaussier, E. (2005, March). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European conference on information retrieval* (pp. 345-359). Springer, Berlin, Heidelberg.