Joint CSCE Construction Specialty & CRC Conference 2025
Conférence conjointe spécialisée en construction de la SCGC et CRC-2025

Montreal, Quebec
July 28-31, 2025 / *28-31 juillet 2025*

Comparative Analysis of Deep Learning Models for Sewer Drainage Defects Detection: Faster-RCNN and YOLOv8

Rajbhandari, M.[1,3], Bouferguene, A.[2,3], Al-Hussein, M.[1,5]

1 Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada
2 Campus Saint-Jean, University of Alberta, Edmonton, Alberta, Canada
3 mrajbhan@ualberta.ca
4 ahmedb@ualberta.ca
5 malhussein@ualberta.ca

ABSTRACT: Preventive maintenance of municipal drainage infrastructure plays a vital role in ensuring its operational effectiveness, especially as aging systems require proactive solutions and municipalities emphasize preventive maintenance to avoid costly failures. Using object detection models to automate defect detection in drainage pipeline can increase efficiency and reduce the need for manual inspections. However, selecting the most suitable model remains a challenge. This study compares the performance of Faster R-CNN and YOLOv8 in detecting drainage pipe defects using a dataset of 1824 annotated images categorized according to Pipeline Assessment Certification Program (PACP) defect codes, covering seven defect types: broken, crack, deposits, fracture, tap, hole, and roots. Both Faster R-CNN and YOLOv8 models were trained and validated using 5-fold cross-validation to ensure robust evaluation. Their performance was measured using the metrics such as precision, recall, F1 score, mean Average Precision (mAP), and Intersection over Union (IoU). The results show that YOLOv8 outperforms Faster R-CNN in accuracy and speed, achieving a higher mean Average Precision (mAP@50) of 83.28% and (mAp@50-95) of 60.76% with faster inference times, making it more suitable for real-time applications. In contrast, Faster R-CNN showed lower detection accuracy and required significantly more computational resources. This study highlights the potential of deep learning models to automate drainage defect detection, reducing manual inspection efforts and improving efficiency in preventive maintenance programs. Future work will focus on hyperparameter tuning and real-time deployment to improve classification performance and adaptability to environmental conditions.

Keywords: Sewer pipe defect detection, Faster-RCNN, YOLOv8 (You Only Look Once), Cross-Validation.

1. INTRODUCTION

Maintaining sewage infrastructure is essential for urban drainage management, particularly considering aging systems and increasing maintenance costs. Preventive maintenance, such as routine inspections and quick repairs, are essential for avoiding expensive malfunctions, minimizing damage to the environment (like water body contamination from sewer overflows), and minimizing the social disruptions that can result from sewage backups or system failures. By establishing maintenance priorities based on accurate assessments of sewer conditions, municipalities can improve system performance, protect public health, and ensure compliance with regulatory standards (Fenner 2000). Detecting faults in sewer pipes, such as deposits, cracks, fractures, and root invasions, poses significant challenges. These defects often lack clear boundaries, requiring skilled engineers to interpret complex patterns. Additionally, since defects are rare in many cases, vast amounts of CCTV footage need to be reviewed, which is time-intensive and error-prone.

Manual inspection methods, even when standardized by systems like the PACP code, contribute to technician fatigue and inconsistencies in defect classification, further complicating the process (Yin et al. 2020).

To address these challenges, automated annotation using machine learning models can enhance the efficiency and accuracy of defect detection. By leveraging advanced computer vision techniques like Faster R-CNN (Wang et al. 2021), Single Shot Detector (SSD) (Shen et al. 2023), and YOLOv8 (Yuan and Wang 2024), these models can automatically identify and classify defects in CCTV footage, reducing the workload for inspectors and minimizing human error in the assessment process. Recent advancements in deep learning and computer vision offer promising solutions to these challenges. Automated defect detection systems can significantly enhance productivity by enabling technologists to focus on validation and decision-making, thereby reducing human error (Yin et al. 2020). However, selecting the most suitable model for sewer defect detection remains a challenge due to the complexity and diversity of these algorithms. Factors such as dataset size, training-validation partitioning, and hyperparameter tuning can significantly influence model performance (Yin et al. 2020).

This study aims to address this challenge by providing a comparative analysis of Faster R-CNN and YOLOv8 using cross-validation. The performance metrics such as precision, recall, F1 score, and Intersection over Union (IoU) will be evaluated. Additionally, the study will assess the performance of both deep learning models on a consistent drainage defect dataset using hyperparameter tuning, with especially focusing on epoch count. The findings of this study aim to provide actionable insights for implementing machine learning models in drainage defect detection, ultimately improving inspection efficiency and accuracy in real-world scenarios.

2. LITERATURE REVIEW

Recent advances in machine learning have significantly improved sewage problem identification, exceeding the constraints of manual inspection. Traditional approaches were based on human judgment, which is time-consuming and inconsistent. Early techniques, such as edge detection and handmade features, had difficulty identifying complicated problems in a variety of sewage settings. Automating feature extraction through deep learning, particularly with Convolutional Neural Networks (CNNs), has revolutionized the industry by increasing accuracy and efficiency. Through the application of attention mechanisms, detection is further improved by concentrating on important aspects of the image, which improves performance under difficult circumstances. In general, deep learning-based models perform better than conventional techniques and are more appropriate for sewer inspection in real time (Oh et al. 2022).

Several studies have investigated the use of Faster R-CNN for defect detection, emphasizing its superior performance in recognizing and localizing diverse defects with high mean average precision (mAP) scores. The ability of Faster R-CNN to share convolutional layers across the detection network and the region proposal network (RPN) lowers computational costs and increase efficiency, which is a significant advantage (Wang and Cheng 2018). However, despite its accuracy, the model’s detection speed is limited, especially when working with large datasets that need a lot of processing power. Furthermore, Faster R-CNN can sometimes generate false positives, misidentifying characteristics that mimic flaws (Wang and Cheng 2018). While the model is highly precise and ideal for thorough defect identification, its slower processing speed and high computational needs make it unsuitable for real-time application.

According to research, YOLOv8 offers an excellent balance between accuracy and real-time performance, making it very successful at identifying drainage pipeline defects. Its enhanced network design enables more accurate flaw diagnosis by incorporating sophisticated feature fusion and attention methods. The decoupled head design of YOLOv8 is a significant advantage since it isolates object location and categorization, resulting in a more computationally efficient model. Its improved receptive field modules also make it easier to identify minor defects like crack and intrusions of foreign objects. Even though YOLOv8 outperforms earlier iterations in terms of accuracy, it does so at the expense of a marginally lower frame rate, which makes it less suitable for real-time applications in the absence of additional optimization.

Despite this, YOLO v8 continues to be a powerful tool for accurately detecting defects in urban drainage system (Dong and Liao 2024).
Although machine learning models show great potential in detecting sewer defects, there are still a number of obstacles to overcome. These include managing large datasets, handling problems like false positives and negatives, and finding the right balance between accuracy and speed. Similarly, optimizing models for real-time application is difficult due to the high computational resources required. Despite having great accuracy, Faster R-CNN is less suitable for real-time applications due to its slow processing speed. On the other hand, YOLOv8 is faster but sometimes sacrifices accuracy for speed.

This study aims to address these challenges by performing a comprehensive comparison of Faster R-CNN and YOLOv8, focusing on their ability to detect defects in sewer systems. By using a diverse dataset with various defect types and employing 5-fold cross-validation, this research aims to provide deeper insights into the trade-offs between accuracy, processing speed, and resource consumption.

3. METHODOLOGY

This study aims to evaluate and compare the performance of two advanced object detection models- Faster R-CNN and YOLOv8 for automated sewer defect detection. This research methodology involves data preparation, model implementation, training, validation, testing, and performance evaluation using the key metrics such as precision, recall, F1 score, Intersection over Union (IoU), and training-validation losses.

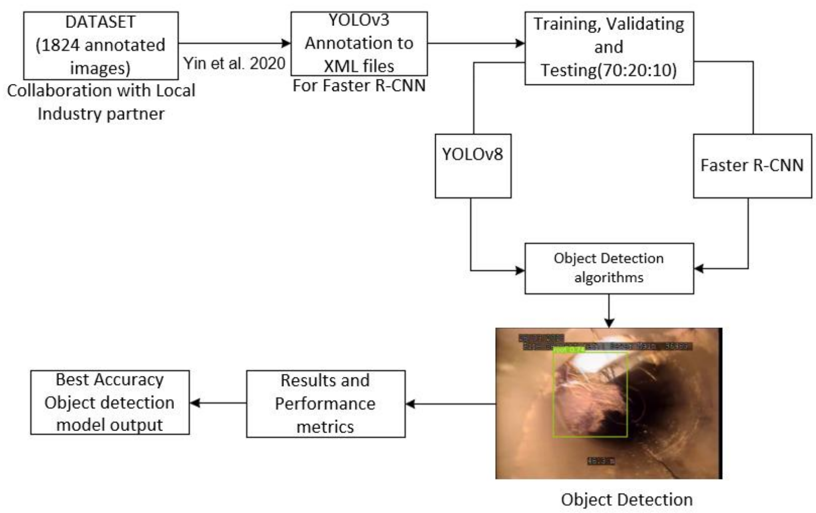

Figure 1: Work flow of the paper.

3.1 Data Preparation

The dataset used in this study consists of 1824 images labeled with seven defect types such as broken (405 instances), crack (473 instances), deposits (496 instances), fracture (318 instances), tap (275 instances), hole (312 instances) and root (317 instances). The images were collected in collaboration with a local industry partner and were annotated using the YOLOv3 format (Yin et al. 2020). These annotations were then converted into the XML format required by Faster R-CNN to ensure compatibility with the models. The dataset contains images with three different resolutions: 720×480 (74.45%), 320×240 (18.63%), and 320×238 (7.92%). The dataset was split into training, validation, and testing in a ratio of 70:20:10 percent. Additionally, a 5-fold cross-validation strategy was implemented to improve evaluation robustness and reduce overfitting.

3.1.1 Model Implementation

3.1.1.1 Faster R-CNN

Faster R-CNN is a two-stage object detection model that combines a Fast R-CNN module for classification and localization with a Region Proposal Network (RPN) to produce region proposals. By using anchor boxes to scan feature maps, the RPN generates possible object areas. For fixed-size feature maps, these regions are then further refined using ROI pooling layers. The model simultaneously optimizes bounding box regression and classification using a multi-task loss function, enabling precise and effective object detection (Liu 2018).

In this study the implementation of Faster R-CNN was tailored to optimize the model's performance on drainage defect detection through a structured training approach. The model was developed using PyTorch with a ResNet-50 backbone pretrained on the COCO dataset. The model was trained using two different configurations: one with 50 epochs and another with 100 epochs per fold. A batch size of eight and a learning rate of 0.005 were employed. To ensure robustness and minimize overfitting, the training and validating process utilized a 5-fold cross-validation strategy. The dataset for each fold consists of 1,167 training images and 292 validation images. A different dataset of 365 images was kept out of training and validation stages and used just for testing. The Adam optimizer was employed throughout the training process. The model's performance was measured during training using key metrics like precision, recall, F1 score and Intersection over Union (IoU). After the cross-validation process, the final model was evaluated on the reserved 365 test images. The validation results from each fold were averaged to provide a comprehensive assessment of the model's performance across the whole dataset.

3.1.1.2 YOLOv8

YOLOv8 is a one-stage object detection model that divides input images into N x N grids to predict bounding boxes and identify objects. To evaluate the existence of objects, each grid cell calculates confidence scores using the Intersection over Union (IoU) with ground truth. Non-Max Suppression (NMS) is used to refine forecasts, keeping the most confident ones and eliminating overlapping boxes. The model optimizes bounding box regression and classification at the same time, resulting in accurate detection of each object using bounding boxes, class labels, and confidence scores (Seth and Sivagami 2025).

For this study, YOLOv8 model was trained using a 5-fold cross-validation approach to evaluate its performance on the drainage pipe defect detection. Similar to Faster R-CNN, the model was trained under two configurations: 50 epochs and 100 epochs per fold using an NVIDIA GeForce RTX 4090 GPU. Key performance metrics, including precision, recall, mean average precision at Intersection over Union (IoU) 0.50 (mAP50), and mean average precision at Intersection over Union (IoU) 0.50-0.95 (mAP50-95) were recorded throughout the training. The model was fined-tuned from pretrained weights provided by the official Ultralytics YOLO implementation. After training, the best-performing weights from each fold were validated on their corresponding validation sets to accurately assess the model's performance. Finally, testing was conducted on a separate set of 365 reserved test images. The trained models were applied to these images to predict defects, and the results were analyzed to provide meaningful insights. This structured and methodical approach offered a comprehensive evaluation of YOLOv8's capability to identify various drainage defect types across the dataset.

3.1.2 Evaluation Metrics

3.1.2.1 Precision and Recall

Precision is defined as the ratio of true positive detections to the sum of true positives and false positives, indicating the correctness of the detected objects. It can be expressed mathematically as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

where TP and FP represent true and false positives respectively (Yin et al. 2020).

As for the recall, it assesses the completeness of detected objects among all labeled instances. It is mathematically defined as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

where FN denotes false negatives (Yin et al. 2020).

3.1.2.2 Performance Metrics

Intersection over Union (IoU) is a metric that quantifies the accuracy of an object detection model by calculating the ratio of the area of overlap between the predicted bounding box and the ground truth bounding box to the area of their union (Yin et al. 2020).

The mean average precision at an Intersection over Union (IoU) threshold of 0.50 (mAP50) evaluates the accuracy of a model by considering only those detections that meet or exceed this Intersection over Union (IoU) threshold, providing a focused assessment of object detection performance (Yin et al. 2020).

The mean average precision spanning Intersection over Union (IoU) thresholds from 0.50 to 0.95 (mAP50-95) offers a comprehensive evaluation of the model's performance, assessing its robustness across varying levels of detection specificity (Yin et al. 2020).

4. RESULT AND DISCUSSION

4.1 Performance of Faster R-CNN Model

The Faster R-CNN model was trained for 50 and 100 epochs on a sewer defect detection dataset, utilizing a 5-fold cross-validation approach. Table 1 and 2 presents the overall performance metrics of Faster R-CNN across five folds for 50 epochs and 100 epochs configurations, respectively.

Table 1: Performance of Faster R-CNN across 5 folds at 50 epochs

Fold	Intersection over Union (IoU)	Accuracy (%)	Precision (P)	Recall (R)	F1 Score
1	0.4013	53.19	0.7381	0.5319	0.6115
2	0.4608	58.15	0.7661	0.5815	0.6489
3	0.4238	54.37	0.7173	0.5437	0.6018
4	0.4204	52.18	0.7353	0.5218	0.6030
5	0.4381	57.28	0.7716	0.5728	0.6420

Table 2: Performance of Faster R-CNN across 5 folds at 100 epochs

Fold	Intersection over Union (IoU)	Accuracy (%)	Precision (P)	Recall (R)	F1 Score
1	0.3792	44.92	0.5377	0.4492	0.4639
2	0.4080	42.34	0.4824	0.4234	0.4290
3	0.3804	46.84	0.5226	0.4684	0.4779
4	0.4225	45.87	0.4950	0.4587	0.4506
5	0.3789	51.79	0.5646	0.5179	0.5332

From Table 1 and 2, we can clearly see that the performance of Faster R-CNN varies across the five folds at 50 and 100 epochs, respectively. The result shows that at 50 epochs, the model performed best in Fold 2, with an accuracy of 58.15%, precision of 76.61%, and F1 score 0.6489, indicating a good balance between precision and recall. However, Fold 5 achieved the highest precision of 77.16% at 50 epochs. As training continued for 100 epochs, there was a noticeable drop in accuracy, precision, and recall across all folds, which may indicate overfitting. The Intersection over Union (IoU) values remain low, ranging from 0.38 to 0.46, showing difficulties in precisely localizing defects. The performance decline after 100 epochs indicates that further training may result in poor generalization rather than better detection quality. It required 58.74 minutes and 117.48 minutes per fold for 50 and 100 epochs, respectively, to finish training. Figure 2, shows the training loss over epochs and validation accuracy per fold for both 50 and 100 epochs. We can see the spike in training loss over epochs due to the transitions between folds in 5-fold cross-validation. Each spike occurs when a new fold starts, because the model resets and trains on a new data split.

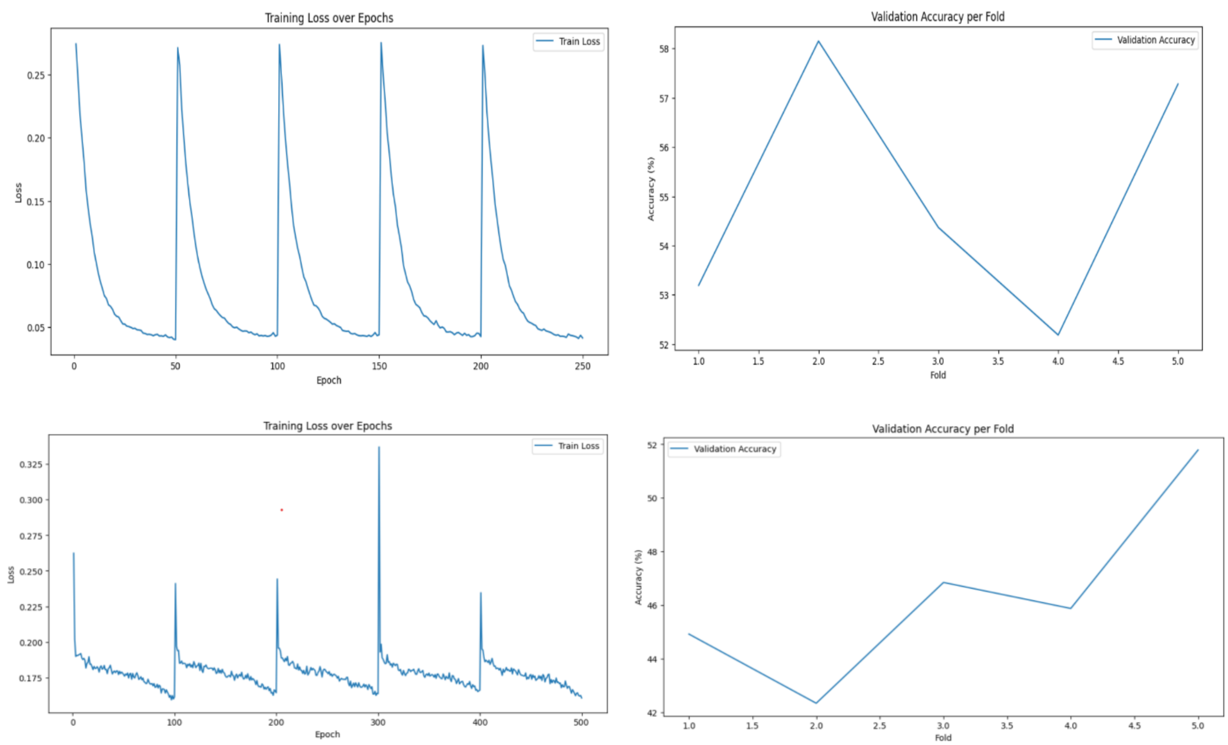

Figure 2: Training Loss and Validation Accuracy for (50 and 100) Epochs in 5- Fold Cross-Validation

4.2 Performance of YOLOv8 Model

The YOLOv8 model was trained for 50 and 100 epochs on a sewer defect detection dataset, utilizing a 5-fold cross-validation approach. Table 3 and 4 presents the overall performance metrics of YOLOv8 across five folds for 50 epochs and 100 epochs configurations, respectively.

Table 3: Performance of YOLOv8 across 5 folds at 50 epochs

Fold	mAP@50	mAP@50-95	Precision (P)	Recall (R)	F1 Score
1	0.450	0.219	0.502	0.443	0.471
2	0.825	0.559	0.815	0.742	0.777
3	0.816	0.564	0.843	0.733	0.785
4	0.820	0.545	0.813	0.730	0.769
5	0.801	0.556	0.808	0.696	0.746

Table 4: Performance of YOLOv8 across 5 folds at 100 epochs

Fold	mAP@50	mAP@50-95	Precision (P)	Recall (R)	F1 Score
1	0.445	0.225	0.501	0.452	0.475
2	0.945	0.719	0.869	0.901	0.885
3	0.909	0.669	0.897	0.822	0.859
4	0.929	0.708	0.847	0.885	0.866
5	0.936	0.717	0.893	0.858	0.875

From Table 3 and 4, we can clearly see that the performance of YOLOv8 improved significantly after Fold 1, with mAP@50 exceeding 90% in Folds 2-5, indicating high detection accuracy at 100 epochs. Among all the folds, Fold 2 achieved the best results, with a precision of 0.869, recall of 0.901, and mAP@50 of 0.945 at 100 epochs. Figure 3 illustrates the box loss and class loss for both the training and validation phases, along with the evaluation metrics used to assess the model's performance in Fold 2. Yolov8 took an average of 7.92 minutes and 15.69 minutes per fold for 50 and 100 epochs, respectively, to finish the training.

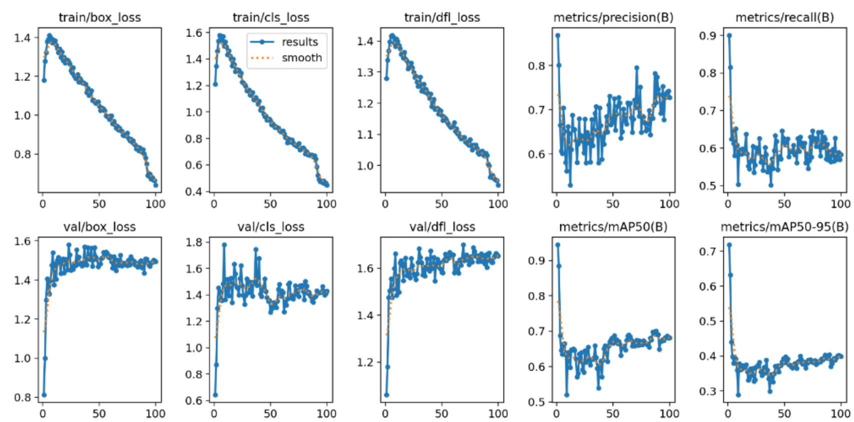

Figure 3: Training and Validation Box Loss, Class Loss, and Evaluation Metrics of YOLOv8 for 100 epochs

During testing, the YOLOv8 model was used on unseen sewer pipe images to evaluate its defect detection capabilities. Figure 4 shows sample results, highlighting different classes like broken, crack, deposits, fracture, tap, hole and root with bounding boxes and confidence scores. These images demonstrate the ability of model to identify defects under different conditions.

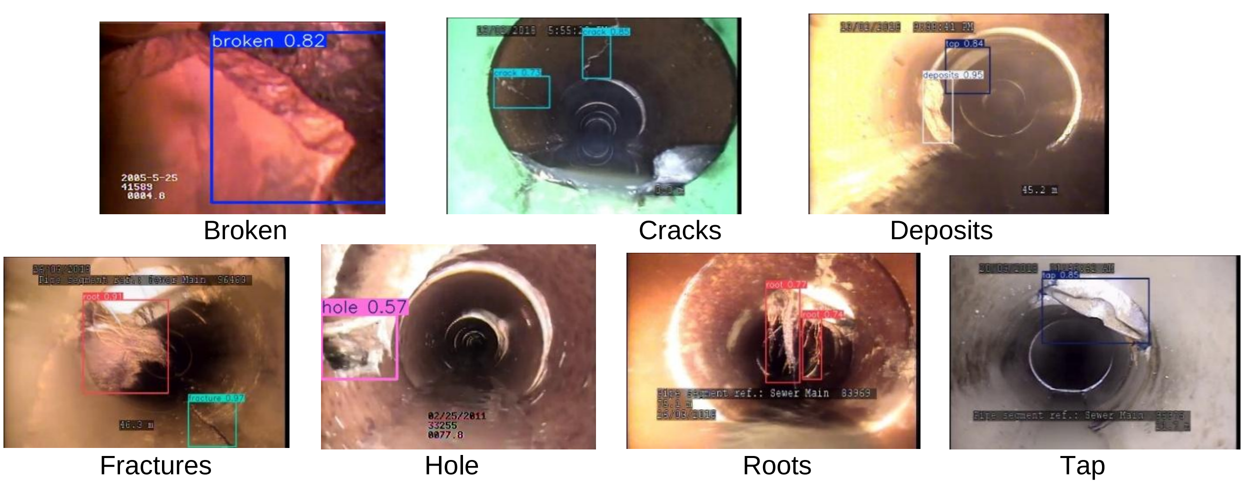

Figure 4: Sample Detection Results from YOLOv8 on Sewer Pipe Defect Images.

4.3 Comparison of Faster R-CNN and YOLOv8 Models

When comparing Faster R-CNN and YOLOv8 using the same dataset and hardware setup (NVIDIA GeForce RTX 4090), Tables 1,2,3 and 4 clearly show that YOLOv8 outperforms Faster R-CNN in both detection performance and efficiency. Faster R-CNN achieves an average accuracy of 55.03% with an Intersection over Union (IoU) of 0.4289 for 50 epochs, while YOLOv8 significantly outperforms it with an average mAP@50 of 83.28%, along with high precision (0.8014) and recall (0.7836). YOLOv8 also retains good overall detection performance, with a mAP@50-95 of 0.6076, indicating its durability at various Intersection over Union (IoU) thresholds. Beyond accuracy, one of YOLOv8's biggest advantage is its computational efficiency, completing training in just 15.69 minutes per fold for 100 epochs, whereas Faster R-CNN requires 117.48 minutes per fold, making YOLOv8 approximately eight times faster.

5. CONCLUSIONS

This study provides a comprehensive evaluation of Faster R-CNN and YOLOv8 for sewer defect detection, focusing on their accuracy, speed and efficiency. The results show that YOLOv8 outperforms Faster R-CNN, offering faster processing and higher accuracy, making it the preferred model for this dataset. It also achieves a higher mean average precision (mAP@50) of 83.28% and faster training times, making it more suitable for real-time applications. On the other hand, Faster R-CNN requires significantly more computational time and does not provide valuable improvement in detection accuracy. The finding show how deep learning-based models can improve sewage infrastructure monitoring by decreasing the need for manual inspection and increasing consistency and efficiency. Municipalities and service providers are able to ensure fast identification of defects and optimize maintenance operations through the integration of advanced object detection models. Future research will focus on fine tuning hyperparameter including dataset sizes, batch size, and learning rate as well as exploring real-time deployment efficiency and defect subclass classification to further refine the detection accuracy and model performance.

ACKNOWLEDGMENTS

The authors would like to acknowledge Alberta Innovates and the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support, which made this research possible. We are also grateful to the University of Alberta for providing the computational tools and resources, including computing time, that enabled this work.

REFERENCES

Dong, J., and M. Liao. 2024. “Defect Detection of Urban Drainage Pipeline Based on Improved YOLO-V8.” *2024 IEEE 7th International Conference on Information Systems and Computer Aided Education, ICISCAE 2024*, 284–289. Institute of Electrical and Electronics Engineers Inc.

Fenner, R. A. 2000. “*Approaches to sewer maintenance: a review*.”

Liu, Y. 2018. “An Improved Faster R-CNN for Object Detection.” *Proceedings - 2018 11th International Symposium on Computational Intelligence and Design, ISCID 2018*, 119–123. Institute of Electrical and Electronics Engineers Inc.

Oh, C., L. M. Dang, D. Han, and H. Moon. 2022. “Robust Sewer Defect Detection With Text Analysis Based on Deep Learning.” *IEEE Access*, 10: 46224–46237. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2022.3168660.

Seth, Y., and M. Sivagami. 2025. “Enhanced YOLOv8 Object Detection Model for Construction Worker Safety using Image Transformations.” *IEEE Access*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2025.3527511.

Shen, D., X. Liu, Y. Shang, and X. Tang. 2023. “Deep Learning-Based Automatic Defect Detection Method for Sewer Pipelines.” *Sustainability (Switzerland)*, 15 (12). MDPI. https://doi.org/10.3390/su15129164.

Wang, M., and J. C. P. Cheng. 2018. “Development and improvement of deep learning based automated defect detection for sewer pipe inspection using faster R-CNN.” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 171–192. Springer Verlag.

Wang, M., S. S. Kumar, and J. C. P. Cheng. 2021. “Automated sewer pipe defect tracking in CCTV videos based on defect detection and metric learning.” *Automation in Construction*, 121. Elsevier B.V. https://doi.org/10.1016/j.autcon.2020.103438.

Yin, X., Y. Chen, A. Bouferguene, H. Zaman, M. Al-Hussein, and L. Kurach. 2020. “A deep learning-based framework for an automated defect detection system for sewer pipes.” *Automation in Construction*, 109. Elsevier B.V. https://doi.org/10.1016/j.autcon.2019.102967.

Yuan, B., and Z. Wang. 2024. “Drainage Pipe Image Defect Detection Algorithm Based on SLE-YOLOv8.” *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems, ICPICS 2024*, 654–659. Institute of Electrical and Electronics Engineers Inc.