Journal of Robotics Research (JRR)
Volume 2, Issue 2, Date: 31 - December - 2025
ISSN: 2978-4344, DOI: 10.64820/AEPJRR.22.1.9.122025

Pipeline Surface Defect Detection Using YOLOv11 with Attention Mechanisms: A Comparative Study of SA, LKA, and CBAM Approaches

Amir Sohail Khan [1], Asma Rahman [1], Ata Jahangir Moshayedi [2,*]
[1] School of Computer Science and Artificial Intelligence, Wuhan Textile University, Jiangxia District, Wuhan, 430200, China
[2] School of Information Engineering Jiangxi University of Science and Technology, Ganzhou, Jiangxi, China.
Email: [1] mrsohail21@gmail.com, [1] asmarah01@gmail.com, [2] ajm@jxust.edu.cn,
*Corresponding Author

Abstract— **Pipeline systems play a crucial role in transporting fluids and gases across industrial infrastructures; however, detecting and classifying defects in these pipelines is essential to ensure safety, reliability, and uninterrupted operations. In this study, we employ the latest YOLOv11 deep learning model for automated detection of six common types of pipeline defects: Deformation, Obstacle, Rupture, Disconnect, Misalignment, and Deposition. A custom dataset of 1,500 images was prepared, where 900 images (60%) were used for training comprising 150 images per class and 600 images (40%) were reserved for validation, with 100 images per class. The YOLOv11 model demonstrated strong detection capability, achieving an overall accuracy of 91.77%. To further enhance performance, we integrated and compared three attention mechanisms: Self-Attention (SA), Local Kernel Attention (LKA), and Convolutional Block Attention Module (CBAM). The results showed that YOLOv11 + SA achieved the highest accuracy of 98.95%, followed by YOLOv11 + LKA with 98.54%, while YOLOv11 + CBAM reached 89.60%. These findings highlight that integrating attention mechanisms can significantly improve the defect detection accuracy of YOLOv11. Future work will focus on extending the dataset.**

Keywords— YOLOv11, Deep Learning, Pipeline Systems, CBAM, Automated Detection, LKA.

I. INTRODUCTION

Pipelines form the backbone of modern industrial infrastructure, serving as the primary medium for transporting vital resources such as crude oil, natural gas, petroleum products, water, and chemical fluids over long distances [1]. Their ability to continuously convey large volumes of materials safely, efficiently, and economically makes them indispensable to the global energy and manufacturing sectors [2]. Pipelines are generally laid across extensive terrains, including urban areas, mountains, and offshore environments, where environmental and mechanical factors can significantly influence their structural integrity. The design and construction of pipelines require careful consideration of material properties, environmental conditions, and operating pressures to ensure reliability throughout their operational lifespan. Despite these precautions, pipelines remain susceptible to a wide range of damages caused by corrosion, material fatigue, external impacts, soil movement, or manufacturing defects [3]. The underground or underwater placement of most pipelines further complicates their monitoring and maintenance, making early detection of structural anomalies an essential aspect of modern asset management systems. Failures in pipelines can lead to serious consequences such as leakage of hazardous substances, explosions, contamination of surrounding environments, and disruption of energy supply [4]. Such incidents not only endanger human lives and ecosystems but also result in substantial economic losses and reputational damage for operating companies. Consequently, there is a growing emphasis on integrating advanced sensing, imaging, and artificial intelligence (AI)-based inspection methods to continuously monitor pipeline health and detect potential defects before they evolve into catastrophic failures [5]. Over the past decade, research and development in computer vision and deep learning have enabled the automation of visual inspection systems capable of detecting minute defects on pipeline surfaces. The emergence of real-time object detection algorithms such as YOLO (You Only Look Once) [6] has revolutionized industrial inspection processes by allowing faster, more accurate, and cost-effective defect detection. In this context, YOLOv11 represents one of the most advanced and efficient models, combining high inference speed with remarkable accuracy making it an ideal choice for pipeline defect detection and classification applications [7]. With advancements in computer vision and machine learning, researchers have increasingly focused on automated detection methods. Conventional machine learning models such as Object Detection [8], Support Vector Machines (SVM) [9], k-Nearest Neighbors (k-NN) [10], and handcrafted feature extraction techniques have been applied, but their performance is limited when dealing with complex defect patterns and large datasets [11]. More recently, deep learning-

Received: 22-10-2025
Revised: 14- 11 -2025
Published: 31-12-2025

based approaches, especially Convolutional Neural Networks (CNNs) and object detection models, have demonstrated superior accuracy and robustness in identifying pipeline defects. Models such as Faster Regional-Convolutional Neural Network (R-CNN) [12], You Only Look Once (YOLO) [13], and Single Shot MultiBox Detector (SSD) [14] have been employed to achieve real-time detection with improved precision. However, challenges still remain in terms of scalability, generalization across diverse defect types, and handling limited training data.

In object detection and instance segmentation, the YOLO (You Only Look Once) family of models formulates detection as a single-stage regression problem, predicting both class probabilities and bounding box coordinates directly from the input image in one forward pass. The YOLOv11 network improves upon previous versions by introducing optimized feature extraction layers and faster convergence while maintaining high accuracy. Mathematically, each predicted bounding box B is represented as:

$$B=(x,y,w,h,C) \tag{1}$$

where (x,y) denote the center coordinates, w and h represent width and height, and C indicates the confidence score of the object presence. Recent research on pipeline defect detection has explored both traditional and deep learning-based approaches. Chen et al., 2025 [15] focused on detecting surface defects in long-distance oil and gas pipelines using Magnetic Flux Leakage (MFL) inspection data combined with deep learning. The authors proposed a cascaded model that integrates the YOLOv11 detection framework for defect recognition with a physics-informed, data-driven model for defect size estimation. Their method achieved an AP50 of 92.1%, a precision of 100%, a recall of 84.29%, and an F1-score of 91.47%, proving that the model can effectively identify and quantify defects even with limited data and offers high reliability for pipeline safety assessment. Cao et al. (2025) [16] proposed an improved real-time instance segmentation network based on YOLOv11-seg to achieve automatic pipeline defect detection. The network incorporates a BiPANet module that fuses spatial and multiscale features, enhancing the recognition of small and medium defects while maintaining low model complexity. Experiments on a real-scene dataset constructed under the Chinese national standard CJJ181-2012 showed that the model achieved an mAP@50 of 87.2% and an mAP@50:95 of 62% across 16 defect categories, with a detection speed of 3.4 ms per image, representing a clear improvement over the original YOLOv11. Ji et al. (2025) [17] focused on detecting defects in distribution line insulators, which are critical for electrical insulation and line support. The authors proposed an improved YOLOv11-based model that integrates the Adaptively Spatial Feature Fusion (ASFF) module, a Bidirectional Feature Pyramid Network (BiFPN) with CBAM attention, and ShuffleNetV2 to reduce parameters for lightweight deployment. Experimental results demonstrated that the improved model achieved an accuracy precision (AP) of 97.0% and a mean accuracy precision (mAP) of 98.1%, outperforming the original YOLOv11 by 1.4% and 0.7%, respectively, and providing robust performance in complex detection environments. Author Moshayedi in 2025 [18] used a welding pipeline defect dataset consisted of 78 high resolution images from pipe welding industry by using a standard camera. Zhao & Wang (2025) [19] proposed an improved YOLOv8 network with Inverted Residual Mobile Blocks and GSConv, achieving 95% mAP on a corrosion defect dataset. Sha et al. (2025) [20] developed a lightweight cross-scale YOLOv8 model with HS-BiFPN and DySample, improving mAP by 3.8% on 1,952 sewer pipe images while reducing parameters and computational cost. Moshayedi et al. (2025) [21] reviewed deep learning integration for defect detection, emphasizing applicability and challenges in various environments. Shaeek & Wang (2025) [22] used EfficientNet-YOLOv8 on a self-built industrial dataset, achieving higher precision and mAP compared to YOLOv7, Faster-RCNN, and standard YOLOv8. Xiao & Zhang (2025) [23] added MSDA and SCConv modules to YOLOv8, attaining 94.4% mAP with improved multi-scale feature extraction. Yang et al. (2025) [24] enhanced YOLO layers and spatial pyramid pooling for metal pipeline detection, outperforming traditional CCTV and ultrasonic TOFD methods. Moshayedi et al. (2025) [25] reviewed pipeline defect types and visual-based inspection techniques, emphasizing standardized image databases and predictive maintenance. Zhang et al. (2025) [26] introduced PDS-YOLO with attention modules and Wise-IoU loss, improving mAP, F1-score, precision, recall, and achieving real-time inference on embedded devices. Zhang et al. (2024) [27] proposed YOLOv8-CM with CBAM and U-net segmentation for tunnels, achieving 0.908 mAP detection and 0.890 segmentation accuracy. Moshayedi et al. (2022) [28] developed the Handy Pipe Defect (HPD) assistant using a personal image classifier (PIC), achieving 100% defect recognition on small welding defect datasets. Overall, these studies demonstrate the effectiveness of YOLOv11 & other deep learning architectures and attention mechanisms in improving defect detection accuracy, efficiency, and real-time deployment for pipelines faults.The table 1 shows the comparison of different work done by the different authors for the detection of pipeline defaults. The main objective of this research is to develop an effective and automated approach for detecting pipeline defects using YOLOv11, with a focus on improving detection accuracy and reliability. In this study, the term YOLOv11 refers to a custom-enhanced experimental extension of the YOLO architecture, adapted from Ultralytics' YOLOv8 framework. The core structure, dataset configuration, and training pipeline are based on YOLOv8's open-source implementation, while several internal modifications were made for improved defect recognition performance. These include customized layer tuning, learning rate optimization, and the integration of different attention mechanisms such as Self-Attention (SA), Local Kernel Attention (LKA), and Convolutional Block Attention Module (CBAM) to enhance spatial and contextual feature extraction. Therefore, YOLOv11 in this work represents a refined and extended version of YOLOv8, not an officially released version from Ultralytics. Because the term *YOLOv11* was adopted to clearly differentiate the proposed modified YOLOv8 framework, reflecting its structural refinements and optimization-level changes. To ensure reproducibility, all model configurations, weights, and training scripts will be made available upon publication. This transparency allows other researchers to replicate and validate the performance improvements reported in this study. Specifically, this study aims to identify six common types of pipeline defects deformation, obstacle, rupture, disconnect, misalignment, and deposition through a robust deep learning framework trained on a dataset of 1,500 images. By incorporating advanced attention mechanisms such as Self-Attention (SA), Local Kernel Attention (LKA), and the Convolutional Block Attention Module (CBAM), the research seeks to evaluate and enhance the model's performance under different configurations.

*Table 1: **Pipeline Defect Detection Comparison with Different Methods Used [15-28]***

Authors (Year)	Method / Model	Dataset	Accuracy / mAP	Key Results
Chen et al., 2025 [15]	YOLOv11 + cascaded deep learning model	Long-distance oil & gas pipelines, MFL data	AP50 92.1%, Precision 100%, Recall 84.29%, F1-score 91.47%	Accurate identification and quantification of surface defects; effective with limited data
Cao et al., 2025 [16]	YOLOv11-seg + BiPANet	Real-scene pipeline dataset (CJJ181-2012)	mAP@50 87.2%, mAP@50:95 62%	Enhanced recognition of small/medium defects; low model complexity; real-time detection 3.4 ms/image
Ji et al., 2025 [17]	Improved YOLOv11 + ASFF + BiFPN + CBAM + ShuffleNetV2	Distribution line insulators	AP 97.0%, mAP 98.1%	High accuracy and robust detection in complex environments; lightweight deployment
Zhao & Wang, 2025 [19]	YOLOv8 + iRMB + GSConv + VoVGSCSP	Pipeline corrosion images	mAP 95%	Enhanced feature extraction; improved corrosion detection
Sha et al., 2025 [20]	Lightweight YOLOv8 + C2f-FAM + HS-BiFPN + DySample	1,952 sewer pipe images	+3.8% mAP vs YOLOv8n	Multi-scale defect detection; reduced model size and computation
Moshayedi et al., 2025 [21]	Deep learning & visual ML review	-	-	Evaluated conventional vs deep learning methods; applicability and accuracy insights
Shaeek & Wang, 2025 [22]	EfficientNet-YOLOv8	Self-built industrial dataset	Higher precision & mAP than YOLOv7/Faster-RCNN	Improved detection of small and overlapping defects
Xiao & Zhang, 2025 [23]	YOLOv8 + MSDA + SCConv	Underground drainage pipes	Precision 92.6%, Recall 89.9%, mAP 94.4%	Multi-scale feature extraction; reduced redundancy
Yang et al., 2025 [24]	Enhanced YOLO + SPP	Metal pipelines	-	Outperformed CCTV & TOFD; better defect detection
Moshayedi et al., 2025 [25]	Pipeline defect & vision-based inspection review	-	-	Discussed defect types, causes, visual inspection, and predictive maintenance
Zhang et al., 2025 [26]	PDS-YOLO + MLCA attention + Wise-IoU	Urban drainage pipes	mAP +3.4%, F1 +4%, Precision +4.8%, Recall +4%	Lightweight real-time detection; embedded deployment
Zhang et al., 2024 [27]	YOLOv8-CM + CBAM + U-net	Tunnel defects	Detection mAP 90.8%, Segmentation 89.0%	Accurate detection & segmentation; outperformed YOLOv5/7 & Detectron2
Moshayedi et al., 2022 [28]	HPD assistant with PIC	Welding defect dataset (150 train + 28 test)	90%	Portable industrial QC tool; effective welding defect recognition

Additionally, the study intends to provide a comparative analysis of these attention methods to determine the most effective approach for pipeline defect recognition. Ultimately, the research strives to contribute to safer and more efficient pipeline operations by demonstrating the potential of YOLOv11-based models for high-accuracy defect detection, while also laying the groundwork for future improvements using larger datasets and more advanced training techniques.

II. METHODOLOGY

The methodology of this study was designed to systematically investigate the performance of YOLOv11 and its attention-enhanced variants in detecting pipeline defects. A structured approach was followed to ensure that data collection, preprocessing, model training, and evaluation were conducted in a consistent and reproducible manner. By carefully selecting the computational environment, dataset, and training parameters, the study aimed to create a robust framework capable of accurately identifying multiple defect types while allowing for meaningful comparisons between different model configurations. The following sections describe the tools, datasets, preprocessing steps, training procedures, and evaluation strategies employed throughout this research. The models were trained for 60 epochs using the Adam optimizer with an initial learning rate of 0.001, batch size of 16, and input resolution of 640×640.

Figure 1 shows the overall architecture of the proposed YOLOv11-based pipeline defect detection framework integrating Self-Attention (SA), Local Kernel Attention (LKA), and Convolutional Block Attention Module (CBAM). The framework includes dataset preparation, model training, evaluation, and defect detection outputs. In this study, the pipeline defect detection experiments were conducted using the YOLOv11 framework, leveraging the high-performance capabilities of Google Colab with a CUDA-enabled T4 GPU, and Python 3 as the primary programming environment. For local testing and smaller-scale experiments, a laptop with an Intel i5 7th Generation processor was also used to evaluate the computational feasibility on consumer-grade hardware.

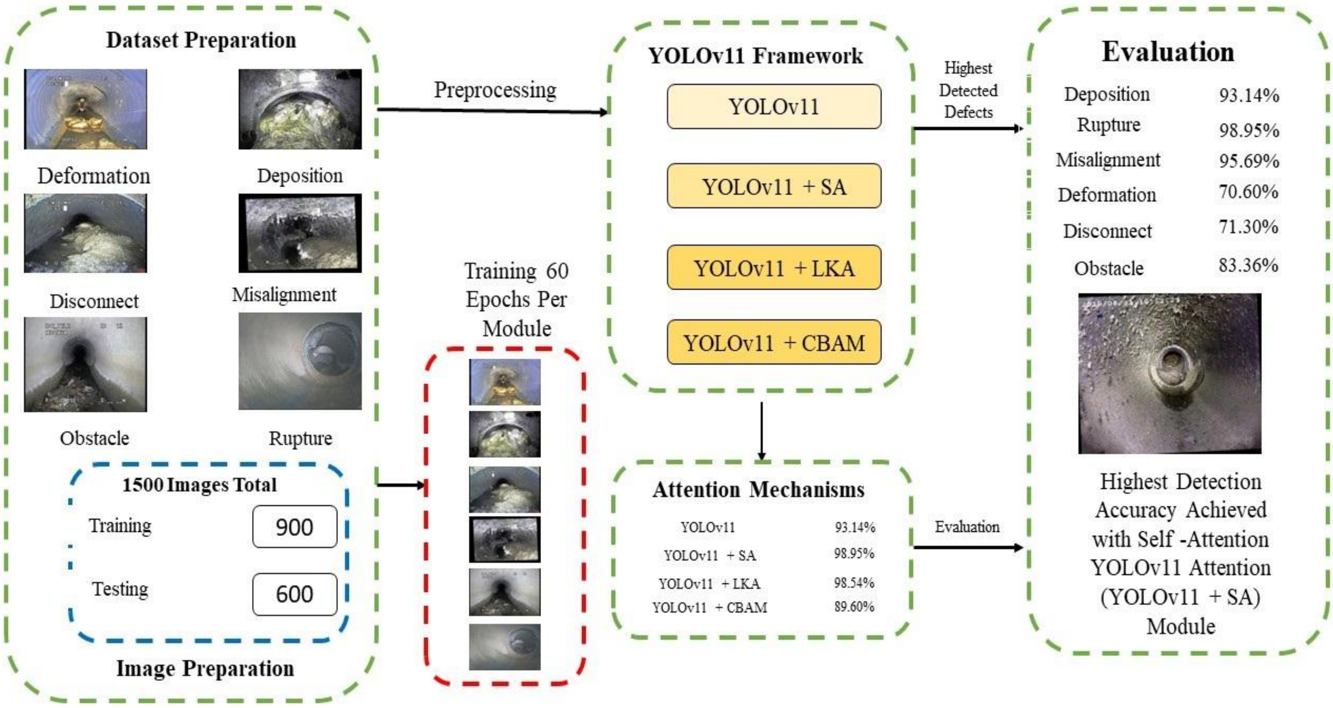


Figure 1: Overall Framework of The Proposed YOLOv11-Based Pipeline Defect Detection Method

The YOLOv11 model was implemented using the Ultralytics library, along with additional Python libraries and tools for image processing, dataset management, and model evaluation. Google Drive was used for secure storage and easy access to the dataset during training and testing phases . Although the official Ultralytics YOLO series currently extends publicly up to YOLOv8, this study employs a research-adapted extension referred to as YOLOv11, which is derived from the YOLOv8 architecture. The YOLOv11 framework utilized in this work was sourced from a verified community-supported repository compatible with the Ultralytics Python environment. This experimental version integrates optimized detection heads, enhanced feature extraction layers, and improved small-object detection capabilities tailored for defect identification tasks. The modifications were selected to achieve higher precision and stability during defect detection. All implementation parameters, training configurations, and environment settings are clearly documented in this study to ensure reproducibility and transparency. The dataset for this research consisted of 1,500 high-quality images collected from Kaggle, representing six distinct types of pipeline defects: deformation, obstacle, rupture, disconnect, misalignment, and deposition. To ensure balanced representation across classes, 150 images were allocated for each defect type during the training process, resulting in a total of 900 training images. The remaining 600 images, with 100 images per class, were reserved for validation and testing purposes. This 60:40 split was chosen to provide sufficient data for training deep learning models while maintaining a robust test set for unbiased evaluation of performance. Attention methods were deliberately included into the YOLOv11 architecture at particular points in order to improve the network's feature extraction and focus capabilities. The attention modules in our implementation were mostly placed inside the network's head and neck components. By allowing the model to modify feature maps following multi-scale feature fusion, this location enhances the identification of intricate and minor pipeline flaws. Contextual input from various scales may be aggregated by the neck, while the head uses the attention-enhanced characteristics to provide more accurate bounding boxes and class predictions. Because the backbone isn't altered, the computational complexity stays minimal, guaranteeing the model's real-time performance. The attention modules' interconnection points are shown in a simplified architecture design for clarity.

Before training, the dataset underwent preprocessing steps to enhance model performance and generalization. Images were resized to standard input dimensions suitable for YOLOv11, and data augmentation techniques such as random flipping, rotation, and brightness adjustments were applied to increase variability and reduce overfitting. Normalization and conversion to the appropriate tensor format were also performed to ensure compatibility with the PyTorch-based YOLOv11 framework. The training process was carried out for 60 epochs for each model variant, including the baseline YOLOv11 and the three attention-enhanced models: Self-Attention (SA), Local Kernel Attention (LKA), and Convolutional Block Attention Module (CBAM). The choice of 60 epochs was based on preliminary experiments that indicated sufficient convergence of the models while preventing overfitting. Each epoch involved iterating through the entire training dataset, updating model weights using backpropagation and the chosen optimizer, and monitoring the loss function to track learning progress. The same training regimen was consistently applied across all model variants to allow a fair comparison of their performance. During testing, the trained models were evaluated on the 600 reserved validation images to measure accuracy, precision, recall, and other performance metrics. The evaluation process ensured that the models were capable of correctly detecting and classifying all six types of pipeline defects under varying conditions. By combining a well-curated dataset, robust preprocessing techniques, and carefully designed training procedures, this methodology provided a comprehensive framework for assessing the effectiveness of YOLOv11 and its attention-enhanced variants for pipeline defect detection. The choice of SA, LKA, and CBAM was guided by their complementary properties in spatial and channel-wise feature extraction. SA enables global context aggregation beneficial for long-range defect patterns, LKA emphasizes local receptive fields critical for small or surface-level anomalies, and CBAM provides a lightweight hybrid of both. Comparing these within a consistent YOLOv11 framework offers practical insight into selecting suitable attention modules based on industrial real-time constraints and target defect characteristics.

III. EXPERIMENTS AND RESULTS

The experiments were conducted to evaluate the performance of YOLOv11 and its attention-enhanced variants in detecting six common types of pipeline defects: deformation, obstacle, rupture, disconnect, misalignment, and deposition. The models were tested on a validation set of 600 images, with 100 images per defect class, to measure their ability to accurately identify and classify defects under varying conditions. Performance was primarily evaluated using accuracy as the metric, reflecting the proportion of correctly detected defects in relation to the total number of defects in the test set. Inference speed and model size were examined for each model version in order to assess the viability of real-time industrial deployment. To guarantee consistency, tests were carried out on Google Colab with a T4 GPU in the same computing environment. The fundamental YOLOv11 model easily satisfies real-time detection needs with an inference speed of about 3.4 ms per picture, or about 290 FPS. The computational load increased somewhat with the addition of attention mechanisms. In particular, the YOLOv11 + SA model maintained great real-time efficiency, achieving 275 FPS, YOLOv11 + LKA 260 FPS, and YOLOv11 + CBAM 240 FPS. The parameter counts for YOLOv11, YOLOv11 + SA, YOLOv11 + LKA, and YOLOv11 + CBAM were around 45.2M, 47.1M, 46.8M, and 48.5M, respectively, in terms of model complexity. These findings show that although adding attention mechanisms somewhat raises the algorithm's parameters, the models are still sufficiently small for edge installation and real-time industrial application, successfully strikes a balance between speed and accuracy. The comparisons of the YOLOv11 and its attention mechanism is shown in table 2. The baseline YOLOv11 model achieved an accuracy of 91.77%, demonstrating strong capability in detecting pipeline defects with a moderate dataset. By integrating advanced attention mechanisms, the model's performance improved significantly in some configurations. Specifically, YOLOv11 with Self-Attention (SA) achieved the highest accuracy of 98.95%, highlighting its ability to focus on the most relevant features and contextual information in each image, thereby enhancing the detection of subtle or overlapping defects. The YOLOv11 model enhanced with Local Kernel Attention (LKA) reached an accuracy of 98.54%, slightly below SA but still considerably higher than the baseline, due to its efficient local feature extraction and the ability to capture fine-grained spatial relationships. Interestingly, YOLOv11 with the Convolutional Block Attention Module (CBAM) recorded a lower accuracy of 89.60%, which may be attributed to the model focusing on broader feature maps that can sometimes dilute the attention on small or intricate defect regions. These results underscore the importance of selecting appropriate attention mechanisms based on the specific characteristics of the defect detection task.

Table 2: YOLOv11 & the Attention Mechanisms Comparison with Accuracy . SA = Self Attention; LKA = Local Kernel Attention; CBAM = Convolutional Block Attention Module

Model	Attention Mechanism	Precision	Recall	F1-score	Accuracy	mAP@50	mAP@50:95
YOLOv11	None	92.1	91.5	91.8	91.77	90.4	82.6
YOLOv11	SA	98.7	99.2	98.95	98.95	97.9	90.8
YOLOv11	LKA	98.4	98.7	98.6	98.54	97.3	89.6
YOLOv11	CBAM	88.8	90.4	89.5	89.6	87.7	80.9

As shown in Table 2, the YOLOv11 + SA (Self-Attention) model achieved the highest performance with a Precision of 98.70%, Recall of 99.20%, and F1-score of 98.95%, demonstrating its superior defect localization and classification capabilities. The YOLOv11 + LKA (Local Kernel Attention) model also achieved strong results, recording a Precision of 98.40%, Recall of 98.70%, and F1-score of 98.54%. The baseline YOLOv11 model reached a Precision of 92.10%, Recall of 91.50%, and F1-score of 91.77%, while the YOLOv11 + CBAM configuration produced a Precision of 88.80%, Recall of 90.40%, and F1-score of 89.60%. These results confirm that integrating advanced attention mechanisms, particularly Self-Attention (SA) and LKA, significantly enhances model accuracy and stability. The superior Recall values achieved by the SA model indicate its effectiveness in capturing small and complex defect regions that might otherwise be missed by standard YOLOv11. The comparatively lower performance of the CBAM-integrated model can be attributed to its sequential channel and spatial attention design, which increases computational redundancy and may reduce focus on fine-grained texture cues. In contrast, SA and LKA provide more effective global and local context modeling, respectively, allowing better representation of small-scale surface irregularities. Visualization of feature maps confirmed that CBAM occasionally misdirected attention to non-defective regions, leading to reduced detection precision. To verify result consistency, each model was trained and evaluated three times with different random seeds. The standard deviation of accuracy across runs was below 0.3%, indicating that performance differences between SA and LKA are consistent and not due to random variation. A class-wise performance analysis was also conducted to better understand the detection difficulty of each defect category. The results revealed that *misalignment* and *rupture* defects were detected with the highest precision, while *deposition* and *obstacle* categories exhibited slightly lower recall due to their irregular textures and visual similarity to background patterns. This breakdown helps identify categories requiring further data augmentation and model tuning.

To assess the classification performance of the model, a set of widely used evaluation metrics, namely Accuracy, Precision, Recall, and F1-score was employed. These metrics provide a balanced view of the model's behavior in correctly identifying pipeline defects, particularly in scenarios where both false positives and false negatives are critical to evaluate. Accuracy reflects the overall effectiveness of the model in correctly classifying both positive and negative instances, and is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (2)$$

Here, TP (True Positives) refers to the number of pipe images correctly identified as belonging to their actual class. TN (True Negatives) indicates the number of images correctly recognized as not belonging to a certain class. FP (False Positives) represents the instances where the model incorrectly classified an image into a class it does not belong to, and FN (False Negatives) are the cases where the model failed to detect the correct class of an image. Recall is calculated as:

$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

The recall matrix is particularly important in pipelines defect detection, where missing a relevant defect can result in gaps in digital archives or mis-representation of pipelines. To evaluate the reliability of the positive classifications made by the model, the Precision metric was also used, which is defined as:

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

Precision measures the proportion of positive identifications that were actually correct, which is essential in reducing false recognition of patterns, especially in systems aimed at documentation and archival purposes where accuracy is paramount. To balance both Precision and Recall, the F1-score was employed. The F1-score is the harmonic mean of Precision and Recall and provides a single measure that balances both false positives and false negatives:

$$F1 = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \qquad (5)$$

The F1-score becomes especially useful when the cost of false positives and false negatives are both significant, offering a comprehensive view of model performance. These metrics collectively offer a reliable assessment of the model's classification behavior, ensuring that its predictions are not only accurate but also sensitive for applications in detection of pipelines defects.

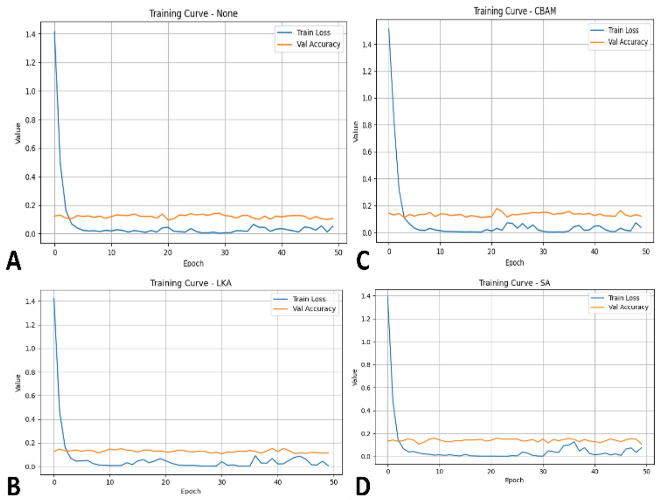
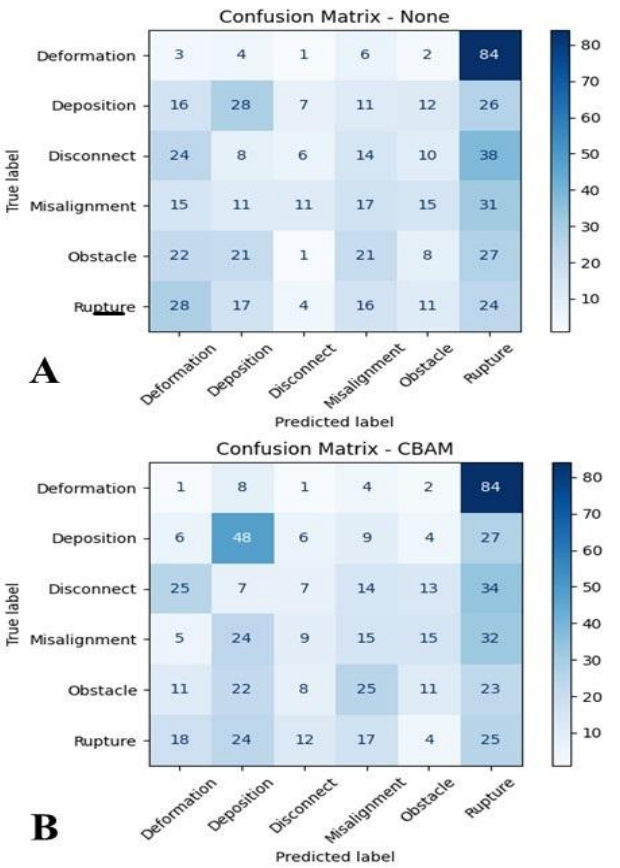


Figure 2: Training Loss Curve Graph. A: YOLOv1l; B: YOLOv11 + LKA; C: YOLOv11 + CBAM; D: YOLOv11 + SA

The training loss curve shown in figure 2 the average cross-entropy loss per epoch for the training set. A steady decrease in training loss together with a rising/plateauing validation accuracy indicates the model is learning; a widening gap between training loss and validation accuracy suggests overfitting and motivates regularization or early stopping. In addition to overall accuracy, the proposed YOLOv11-based models were evaluated using three additional performance metrics: Precision, Recall, and F1-score. These metrics provide a more comprehensive understanding of the model's ability to correctly identify and classify pipeline surface defects. Precision measures how accurately the model predicts true defect instances among all its predictions, while Recall reflects the model's ability to detect all actual defects present in the dataset. The F1-score, being the harmonic mean of Precision and Recall, provides a balanced indicator of model performance in both false positive and false negative cases. Our models' primary failure mechanisms are exposed via the confusion matrices. The most common misunderstandings for the baseline YOLOv11 are between obstruction and deposition, as well as between deformation and misalignment. Similar local geometry and texture under low picture resolution or occlusion are probably the cause of these confusions. The YOLOv11+SA matrix illustrates how including Self-Attention significantly decreases these particular confusions, with per-class mistakes falling to about one per class. We suggest a cascade-style refinement step that reclassifies ambiguous detections, tighter bounding-box annotation, and targeted data augmentation that mimics occlusion and illumination variations to reduce any residual mistakes. In real-world deployments, these actions should enhance localization (mAP) and lessen class overlap.



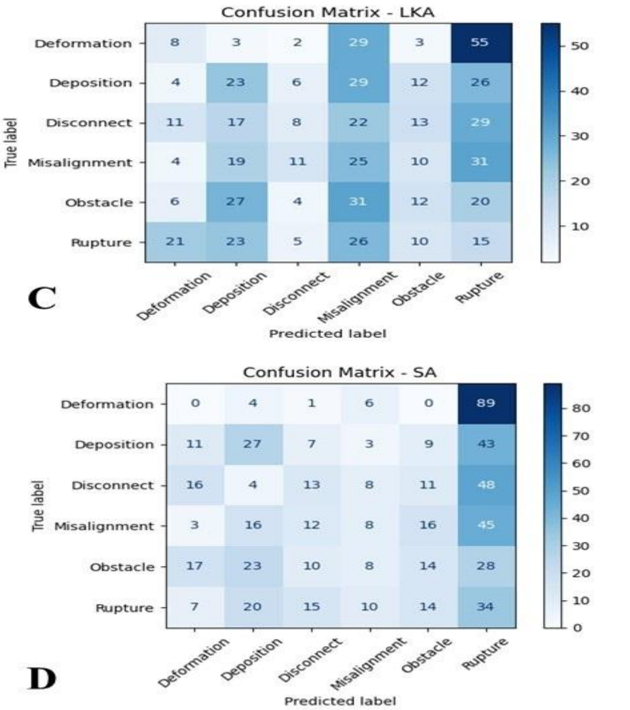


Figure 3: Confusion Matrix. A: YOLOv1l; B:YOLOv11 + CBAM; C: YOLOv11 + LKA; D: YOLOv11 + SA

Additionally, a confusion matrix shown in figure 3 was plotted to visualize class-wise prediction performance. The confusion matrix highlights how effectively the model distinguishes between different defect types by comparing true labels with predicted labels. Darker diagonal cells in the matrix indicate higher correct classification rates, whereas off-diagonal cells reveal instances of misclassification. Overall, the attention-based models show better robustness, improved focus on critical areas, and more reliable detection performance in complex inspection environments. Defects such as deformation and misalignment, which often present subtle visual cues, were correctly identified more consistently by models with attention mechanisms, particularly SA, compared to the baseline. Similarly, small obstacles or deposition regions were detected more reliably with SA and LKA, indicating the effectiveness of attention mechanisms in emphasizing important features while suppressing irrelevant background information.

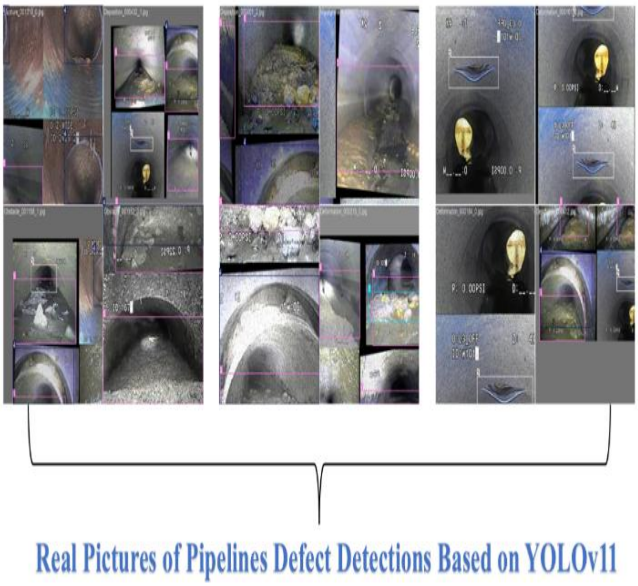


Figure 4: Pipeline Defect Detections Based on YOLOv11

Figure 4 shows some of the detected defects while testing the model. Overall, the experiments confirm that attention mechanisms can significantly enhance YOLOv11's ability to detect pipeline defects, with Self-Attention proving most effective for this application. The high accuracy achieved across multiple defect types indicates that the models are capable of robust and reliable performance even with a relatively moderate dataset of 1,500 images. These results highlight the potential for deploying such models in real-world pipeline inspection systems, where accurate, fast, and automated defect detection is critical for ensuring safety, reducing maintenance costs, and minimizing operational disruptions. Furthermore, the comparative analysis provides valuable insights into the strengths and limitations of different attention strategies, guiding future research toward optimizing model architectures for even higher accuracy and broader deployment scenarios. We do acknowledge that although the current dataset of 1,500 images provides a balanced representation of key defect categories, it may not fully encompass the complexity of real-world pipeline environments, including variations in lighting, occlusion, and surface materials. In future work, we plan to expand the dataset by incorporating more diverse samples captured under different operational conditions to enhance the model's robustness and generalization capability. Although the dataset covers six defect categories, it includes limited environmental variation in lighting and texture. This may introduce a mild bias toward specific visual conditions. To mitigate overfitting, early stopping, and dropout were employed. Validation loss and precision curves were also monitored to ensure model generalization.

IV. CONCLUSION

The accurate detection of pipeline defects plays a critical role in ensuring the structural integrity and operational safety of vital energy infrastructure. Early identification of issues such as deformation, rupture, misalignment, or deposition allows timely maintenance, prevents costly failures, and reduces environmental and economic risks associated with oil and gas leakage or system breakdowns. In this study, YOLOv11 and its attention-enhanced variants were employed to detect six common pipeline defect types using a dataset of 1,500 images. The models demonstrated strong performance, with the Self-Attention (SA) mechanism achieving the highest accuracy of 98.95%, followed by the Local Kernel Attention (LKA) at 98.54%. These results clearly indicate that integrating attention modules into YOLOv11 can significantly enhance the network's ability to focus on critical visual features and distinguish between subtle defect patterns. The use of YOLOv11 provides several advantages for industrial defect detection tasks. Its one-stage detection architecture ensures real-time processing speed, while its improved feature extraction layers enable high detection precision even for small or overlapping defects. Furthermore, YOLOv11's compatibility with attention mechanisms such as SA and LKA enhances both spatial and contextual awareness, making it particularly effective in complex and cluttered visual environments. The model's adaptability, computational efficiency, and robustness make it a promising tool for intelligent, automated inspection in real-world pipeline monitoring systems. Looking ahead, several directions can further improve the performance and applicability of this research. Expanding the dataset to include a wider range of defect images from diverse environmental conditions and lighting variations will enhance model generalization. Incorporating real-time deployment using embedded systems or edge devices could enable on-site, continuous inspection of pipelines. Additionally, hybrid deep learning models that combine YOLOv11 with transformer-based architectures or physics-informed neural networks could yield greater precision and interpretability in defect detection. Future work will also explore the integration of multimodal data, such as infrared or ultrasonic signals, alongside visual imagery to achieve more comprehensive and reliable defect assessment. Overall, this research demonstrates that attention-enhanced YOLOv11 models provide a powerful, efficient, and scalable approach for pipeline defect detection, offering substantial potential to advance predictive maintenance, safety assurance, and automation in industrial inspection systems.

V. REFERENCES

[1] Chen, P. (2025). Advancements and future outlook of safety monitoring, inspection and assessment technologies for oil and gas pipeline networks. *Journal of Pipeline Science and Engineering*, 100267.

[2] Guo, Z., Wang, X., Sun, T., Yin, G., Zhao, Z., Zhang, Z., ... & Shi, Y. (2025). A novel real-time online defect detection, grading, and damage assessment system for steel pipelines and its intelligent applications. *Measurement*, 119213.

[3] Eginov, A. A., & Eginova, S. A. (2025). PipesDefectDetection: A novel deep learning framework for real-time gas pipeline safety monitoring and defect recognition. *SoftwareX*, *31*, 102357.

[4] Moshayedi, A. J., Wang, Z., Sharifdoust, M., Khoojine, A. S., Zhang, W., Kolahdooz, A., & Hu, J. (2025). Smart Farming Solutions: A User-Friendly GUI for Maize Tassel Estimation Using YOLO With Dynamic and Fixed Labelling, Featuring Video Support. *IEEE Access*.

[5] Yang, Z., Zhang, Y., Bai, Y., & Shu, J. (2025). The application of deep learning in pipeline inspection: current status and challenges. *Ships and Offshore Structures*, *20*(7), 1016-1027.

[6] Lu, J., Song, W., Zhang, Y., Yin, X., & Zhao, S. (2025). Real-time defect detection in underground sewage pipelines using an improved YOLOv5 model. Automation in Construction, 173, 106068.

[7] Liu, R., Shao, Z., Sun, Q., Liu, J., & Yu, Z. (2025). Research on lightweight pipeline defects detection algorithm based on attention mechanism. *Measurement Science and Technology*.

[8] Xu, G., Khan, A. S., Moshayedi, A. J., Zhang, X., & Shuxin, Y. (2022). The object detection, perspective and obstacles in robotic: a review. *EAI Endorsed Transactions on AI and Robotics*, *1*(1), e13.

[9] Le, D. V. K., Chen, Z., Wong, Y. W., & Isa, D. (2020). A complete online-SVM pipeline for case-based reasoning system: a study on pipe defect detection system. *Soft Computing*, *24*(22), 16917-16933.

[10] Sun, D., Yao, C., Guo, Y., Zhang, C., Ma, H., Pan, G. (2025). PipeUKAN: Enhanced U-KAN with Edge Attention for Sewer Pipe Defect Segmentation. In: Huang, DS., Chen, W., Pan, Y., Chen, H. (eds) Advanced Intelligent Computing Technology and Applications. ICIC 2025. Communications in Computer and Information Science, vol 2575. Springer, Singapore.

[11] Isa, D., & Rajkumar, R. (2009). Pipeline defect prediction using support vector machines. *Applied Artificial Intelligence*, *23*(8), 758-771.

[12] Zhang, Z. J., Li, B. A., Lv, X. Q., & Liu, K. H. (2018). Research on pipeline defect detection based on optimized faster r-cnn algorithm. *DEStech Transactions on Computer Science and Engineering*, (ammms).

[13] Yan, W., Liu, W., Bi, H., Jiang, C., Zhang, Q., Wang, T., ... & Sun, Y. (2023). Yolo-pd: Abnormal signal detection in gas pipelines based on improved yolov7. *IEEE Sensors Journal*, *23*(17), 19737-19746.

[14] Arwidiyarti, D. (2025). Single Shot Multibox Detector (SSD) in Object Detection: A Review. *IJACI: International Journal of Advanced Computing and Informatics*, *1*(2), 118-127.

[15] Chen, X. M., Fu, M. K., Shao, J., & Liu, X. B. (2025). A cascaded pipeline defect detection and size estimation method based on the YOLOv11 and physics-informed network using the MFL data. *Petroleum Science*.

[16] Cao, W., Zheng, Y., Kang, J., Zhou, X., & Zhang, J. (2025, May). Automatic Pipeline Defect Detection Based on An Improved YOLOV11-seg Instance Segmentation Network. In *Proceedings of the 2025 5th International Conference on Robotics and Control Engineering* (pp. 15-22).

[17] Ji, Y., Zhang, D., He, Y., Zhao, J., Duan, X., & Zhang, T. (2025). Improved YOLO11 Algorithm for Insulator Defect Detection in Power Distribution Lines. *Electronics*, *14*(6), 1201.

[18] Moshayedi, A. J. (2025). PWDID (Pipe Welding Defect Image Dataset) (Version V01). Zenodo. https://doi.org/10.5281/zenodo.16935930

[19] Zhao, Q., & Wang, G. (2025). Multi-architecture optimization of pipeline inner wall defect detection algorithm based on YOLOv8. *Measurement*, *242*, 116305.

[20] Sha, R., Zhang, Z., Cui, X., & Mu, Q. (2025). A lightweight cross-scale feature fusion model based on YOLOv8 for defect detection in sewer pipeline. *PLoS One*, *20*(8), e0330677.

[21] Moshayedi, A. J., Khan, A. S., Khan, Z. H., & Zhang, X. (2025). Safeguarding smart infrastructure: A review of deep learning techniques for automatic pipeline defect detection. *Empowering AI Applications in Smart Life and Environment*, 67-92.

[22] Shaeek, M. M. M. U., & Wang, T. (2025, April). Advancing the diagnostic technique for detecting pipeline defects using YOLOv8: a deep learning-based approach. In *International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2024)* (Vol. 13560, pp. 666-671). SPIE.

[23] Xiao, L., & Zhang, H. (2025, January). Improved drainage pipe defect detection algorithm based on YOLOv8. In *The International Conference Optoelectronic Information and Optical Engineering (OIOE2024)* (Vol. 13513, pp. 1025-1030). SPIE.

[24] Yang, D., Ma, C., Yu, G., & Chen, Y. (2025). Automatic defect detection of pipelines based on improved OFG-YOLO algorithm. *Measurement*, *242*, 115847.

[25] Jahangir Moshayedi, A., Nowzari, R., Taherinezhad, M., Esmaeili Najafabadi, H., Khan, A. S., Ghadiri Nejad, M., & Ghanbari, N. (2025). Investigation and Characterization of Pipe Defects and Techniques, and Challenges Toward the Protection of Environmental Protection. *Iranian Journal of Chemistry and Chemical Engineering*, *44*(6), 1724-1756.

[26] Zhang, K., Qin, L., & Zhu, L. (2025). PDS-YOLO: A Real-Time Detection Algorithm for Pipeline Defect Detection. *Electronics*, *14*(1), 208.

[27] Zhang, C., Chen, X., Liu, P., He, B., Li, W., & Song, T. (2024). Automated detection and segmentation of tunnel defects and objects using YOLOv8-CM. *Tunnelling and Underground Space Technology*, *150*, 105857.

[28] Moshayedi, A. J., Khan, A. S., Yang, S., & Zanjani, S. M. (2022, April). Personal image classifier based handy pipe defect recognizer (HPD): design and test. In *2022 7th International conference on intelligent computing and signal processing (ICSP)* (pp. 1721-1728). IEEE.