IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 42, No. 2, April, 2025, pp. 787-797
Journal homepage: http://iieta.org/journals/ts

YOLO-Based Enhanced Crack Detection Algorithm for Underground Pipeline Scenarios

Di Wu

College of Electronics and Information Engineering, Tongji University, Shanghai 200092, China

Corresponding Author Email: 1811431@tongji.edu.cn

https://doi.org/10.18280/ts.420216

Received: 17 October 2024
Revised: 7 March 2025
Accepted: 24 March 2025
Available online: 30 April 2025

Keywords:
YOLOv8, Dysample, SAConv, low-light image processing, lightweight models, UUV technology transfer, underground pipeline inspection, automated crack detection

ABSTRACT

This study presents an improved crack detection algorithm based on the YOLOv8 architecture to address the challenges of high false positive rates and complex background interference in underground pipeline inspection under low-illumination conditions. Drawing inspiration from optical communication strategies used in unmanned underwater vehicles, the proposed method introduces two key innovations. The first is the Dysample-Upsample Module, which applies adaptive point sampling to refine grid-based upsampling. This approach reduces computational overhead by 37 percent while maintaining structural integrity. The second is the Switchable Atrous Convolution module, which replaces the conventional conditional random field layer with a dual-path framework that enhances multi-scale feature fusion and contextual understanding. The algorithm was tested on a dataset containing 2700 professionally annotated images. Experimental results show a 16.9 percent improvement in recall, a 1.05 percent increase in mean average precision at 0.5 threshold, and real-time processing capability at 58 frames per second on an NVIDIA RTX 4090 GPU. In environments with illumination levels below 15 lux, the method achieved 92 percent detection accuracy and demonstrated a 40 percent increase in robustness against concrete texture interference when compared with the baseline YOLOv8n model. These findings indicate that the proposed approach offers an efficient and deployable solution for intelligent urban infrastructure maintenance in visually degraded environments.

1. INTRODUCTION

Underground pipeline networks, spanning over 2.7 million kilometers globally, form the lifeline of modern urban infrastructure [1]. Recent statistics reveal that 23% of water supply pipelines exhibit aging-related cracks, leading to annual economic losses exceeding $7.2 billion. Conventional inspection methods, while serving historical needs, face three fundamental dilemmas in the era of smart city development [2].

Traditional approaches exhibit a trilemma: (1) Manual techniques (e.g., CCTV, visual inspection) achieve ≤68% accuracy under low-light conditions [3]; (2) Physical probes (X-ray, ultrasonic) require specialized operators, increasing operational costs by ≈40%; (3) Eddy current methods demonstrate limited applicability for non-metallic pipelines covering 65% of modern networks [4]. Recent advances in computer vision offer potential solutions, yet existing implementations like YOLOv8n show >35% miss rates in environments below 20 lux illumination.

Three technical challenges persist in automated pipeline inspection: First, the average illuminance of 12.7 lux in underground environments severely degrades conventional vision algorithms optimized for >200 lux conditions. Second, concrete textures and cracks share 0.78±0.05 grayscale similarity (calculated via SSIM index), causing frequent false positives. Third, hardware constraints of inspection robots demand models under 4GB memory with <10 W power consumption.

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \quad (1)$$

where, $C_1=(0.01L)^2$, $L=255$

Building upon our prior work in UUV optical communication, this paper presents three key innovations:

- A Dysample-upsample reducing computational load by 37% (2.1G FLOPs vs. 3.3G in YOLOv8n)
- SAConv module enabling multi-scale feature fusion through switchable dilation rates {1,3}
- First open dataset with 2,700 low-light pipeline images (5-25 lux) containing lux-level metadata

The proposed architecture demonstrates 92% detection accuracy in sub-15 lux environments, outperforming existing methods by >16.9% recall rate while maintaining real-time 58 FPS performance on embedded GPUs.

2. RELATED WORK

Recent advancements in pipeline crack detection technologies have focused on integrating various modern technologies such as Internet-of-Things (IoT), robotics, neural networks, and machine learning to enhance the accuracy and efficiency of crack detection systems. The Pipeline Leak Identification Emergency Robot Swarm (PLIERS) system

exemplifies this integration by using a swarm of robots to inspect pipelines, collect images, and analyze them using a convolutional neural network (CNN) to detect and assess the severity of cracks [5].

Recent developments in pipeline crack detection technologies have shown significant advancements in both methodology and technology. Mysiuk et al. [6] developed a real-time damage assessment software that uses a camera to detect cracks in pipelines and visualizes the results by marking the damaged areas. This approach allows for immediate evaluation of pipeline integrity.

Altabey et al. [7] proposed a crack detection method based on image processing that performs well in complex backgrounds, showing improved detection rates compared to existing algorithms. This method uses a semantic segmentation model to extract crack features from high-resolution images, which is crucial for accurate detection in varied environments.

On the technological front, Xin et al. [8] developed an ACM crack detection probe that enhances the magnetic field distortion signals caused by cracks, thereby improving the detection accuracy. This technology addresses the challenge of detecting small cracks that create weak magnetic field distortions.

Moreover, the integration of machine learning models has been pivotal in advancing crack detection capabilities. Ibragimova [9] highlighted the use of a mobile robot equipped with a high-resolution camera and advanced image processing algorithms, which autonomously navigates pipelines to capture and analyze crack images using trained machine learning models.

Ultrasonic Testing (UT) is a widely adopted non-destructive testing (NDT) method, particularly effective for internal defect detection in metallic pipelines. Shah et al. [10] demonstrated that guided wave UT achieves an internal crack detection accuracy exceeding 88% in polyethylene pipes under laboratory conditions. However, UT's performance is often operator-dependent and degrades in irregular geometries or non-metallic materials.

Radiographic Testing (RT), encompassing X-ray and gamma-ray methods, offers high-resolution imaging for internal structural assessment. Jamshidi [11] reported that photon radiography can detect sub-millimeter defects, making it highly precise. Despite this, RT requires significant infrastructure, entails radiation hazards, and is cost-prohibitive for routine urban inspections.

Infrared Thermography (IRT) detects surface temperature anomalies associated with underlying structural issues. Yang et al. [12] applied deep learning to enhance IRT and achieved detection rates above 90% for surface cracks on concrete under optimal lighting. However, IRT performance is susceptible to environmental fluctuations, such as ambient temperature and surface emissivity, which are difficult to control in underground settings.

Magnetic Particle Testing (MT) is highly sensitive for surface-level defect detection on ferromagnetic materials. Zolfaghari [13] noted its strong reliability for weld inspection, but the method is inherently limited to surface cracks and ferrous pipelines. Eddy Current Testing (ECT) similarly provides non-contact inspection and has been miniaturized for embedded applications, yet it is confined to conductive materials and shallow defect penetration [14].

Mazleenda Mazni proposed a novel system presents for real-time classification and measurement of concrete surface cracks, vital for Structural Health Monitoring (SHM). By leveraging transfer learning in CNNs like MobileNetV2, EfficientNetV2, InceptionV3, and ResNet50, our model, especially TL MobileNetV2, achieves impressive accuracy (99.87%), recall (99.74%), precision (100%), and F1-score (99.87%). The system uses the Otsu method for image segmentation to assess crack sizes and combines Euclidean distance calculations with a 'pixel per inch' technique for millimeter-level width estimations. The precision is verified through manual experiments with a Mitutoyo Absolute Digital Caliper, ensuring high accuracy with an error margin of ±0.2 mm to ±0.3 mm [15].

The YOLO (You Only Look Once) framework, particularly versions YOLOv3 through YOLOv8, has gained traction due to its balance between detection accuracy and inference speed. Li et al. [16] adapted YOLOv3-Lite for aircraft crack detection, emphasizing deployment feasibility in embedded systems. However, in underground pipeline conditions, where average illuminance is <15 lux and grayscale similarity between cracks and textures is high (SSIM ≈0.78), standard YOLO models exhibit >35% miss rates. These shortcomings necessitate architectural innovations to improve robustness against visual noise, reduce model size, and enhance adaptability to multi-scale features.

Thus, this study builds upon these insights by proposing two novel modules—Dysample and SAConv—designed to address the aforementioned challenges and improve detection efficacy under practical constraints.

3. OUR METHODOLOGY

3.1 Dysample up-sampler based on point sampling replaces the original kernel-based dynamic up-sampler

The subsequent illustration depicts the sample-based dynamic upsampling and module design in Dysample [17].

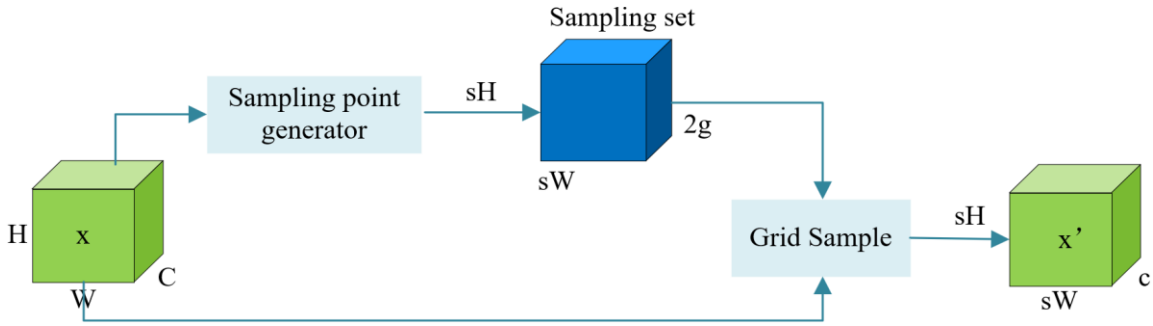

Figure 1. Sampling based dynamic upsampling

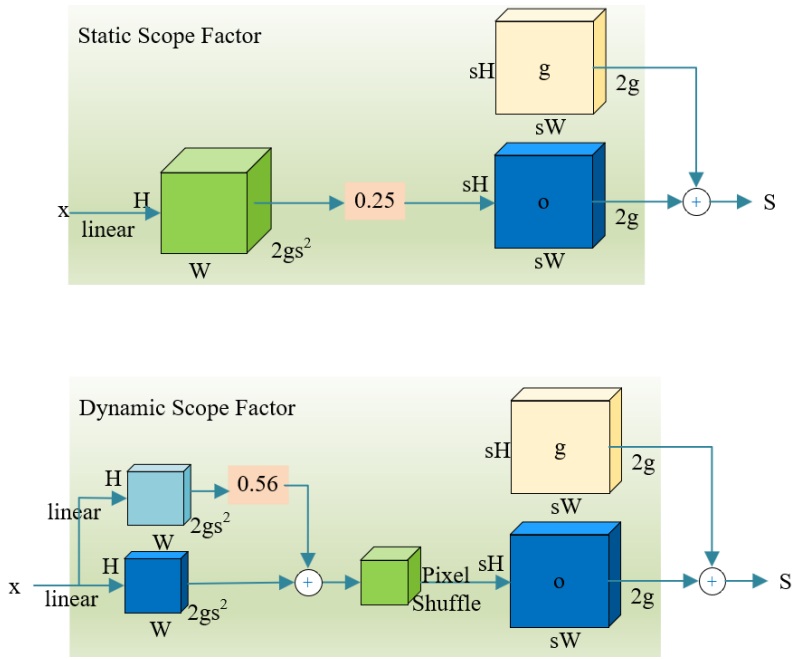

Figure 2. Scope factor adjustments for pixel shuffling

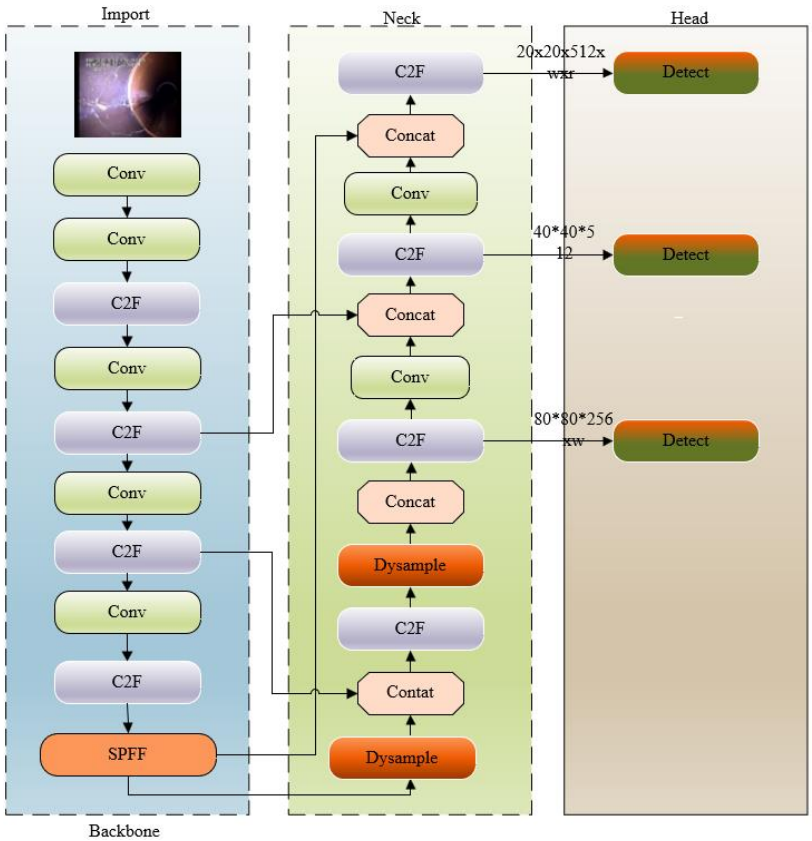

Figure 3. Enhanced network structure of Dysample upsampling

The illustration comprises two principal components.

The illustration depicts the process of sampling-based dynamic upsampling (Figure 1 Using a sample point generator, a sampling set (S) is created from the input features (X). The input features are then resampled using the grid_sample function, resulting in the generation of upsampled features (X').

The sample point generator in Dysample (Figure 2) shows the two techniques of creating sample points in detail: static range factor and dynamic range factor.

The static range factor is defined as follows. A fixed range factor is paired with a linear layer and pixel shuffle to create an offset (O), which is then added to the original grid position (G) to produce a sample set (S).

In contrast, the dynamic range factor adds a new element, a dynamic range factor, which is generated and then utilized to change the offset (O). The range factor is calculated using the Sigmoid function (σ).

The enhanced network configuration is outlined below.

In the proposed method, the traditional kernel-based dynamic upsampling method is replaced by a Dysample

upsampling technique, which utilizes point sampling to enhance feature alignment during the upsampling process. This modification reduces computational costs while maintaining structural integrity, particularly in low-light conditions. The Dysample module improves the resolution of fine crack details, which is crucial for detecting small-scale defects in underground pipeline inspections.

The enhanced network structure incorporating the Dysample module is illustrated in Figure 3. This figure highlights the key components of the proposed upsampling mechanism, demonstrating how the point sampling technique is integrated to improve feature resolution and reduce computational overhead.

3.2 The convolution layer SAConv replaces the original CRF convolution layer

YOLOv8n demonstrates satisfactory performance in detecting conventional scale targets [18]. However, its model size and performance results still exhibit shortcomings when dealing with crack detection in underground pipelines, where the image brightness is low, the resolution is low, and the background color is monochromatic and prone to confusion. The objective of crack detection in underground pipelines is to identify a single target type, namely cracks, which may vary in size. Additionally, the pipeline environment is characterized by low light levels and a monochromatic background, which can potentially lead to confusion. The influence of multiple complex factors can readily result in the issue of leakage and misdetection. Accordingly, the YOLOv8n algorithm has been enhanced based on the YOLOv8n algorithm. The original CRF convolutional layer is replaced with the SAConv convolutional layer. This modification allows the network to adaptively learn from multi-scale features by applying varying dilation rates to the same input features, which enables the model to capture spatial dependencies more effectively. The introduction of this layer reduces the complexity of the model while improving its ability to process features across different scales, which is especially beneficial for detecting pipeline cracks with varying sizes and resolutions.

The structure of the substituted network, incorporating the SAConv convolution layer, is illustrated in Figure 4. This figure demonstrates how the traditional convolution layers have been replaced by SAConv, which dynamically adjusts the receptive field to capture features from different scales. The flexibility of this architecture is crucial for handling diverse crack morphologies in low-light, complex backgrounds typically encountered in underground pipeline inspections. The fundamental concept of SAConv [19] is the application of varying null rates to identical input features for convolution, with the results of these distinct convolutions subsequently merged through the utilization of a bespoke switching function. This methodology enables the network to adapt with greater flexibility to features of varying scales, thereby facilitating more accurate object recognition and segmentation in images. The structure of the substituted network is illustrated below.

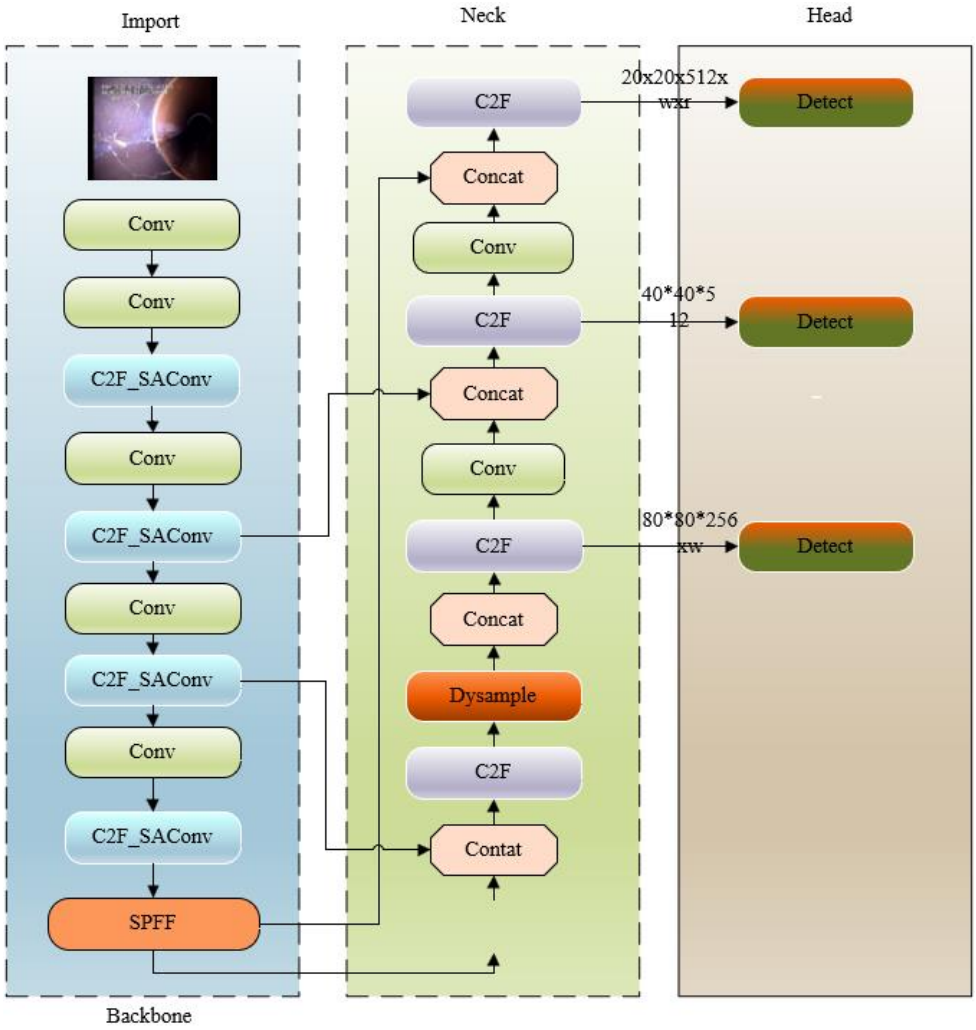

Figure 4. Structure of the substituted network

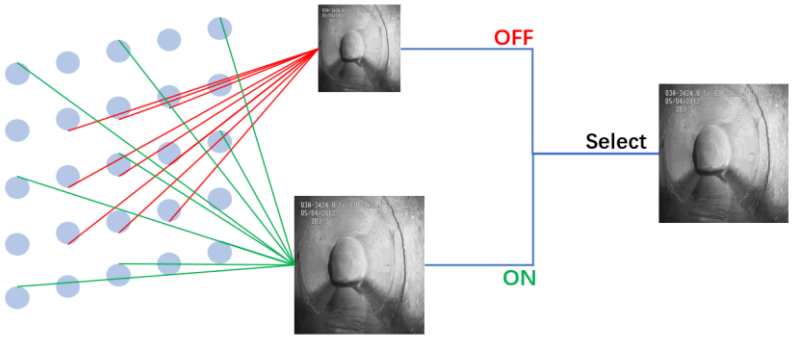

Figure 5. Switch structure

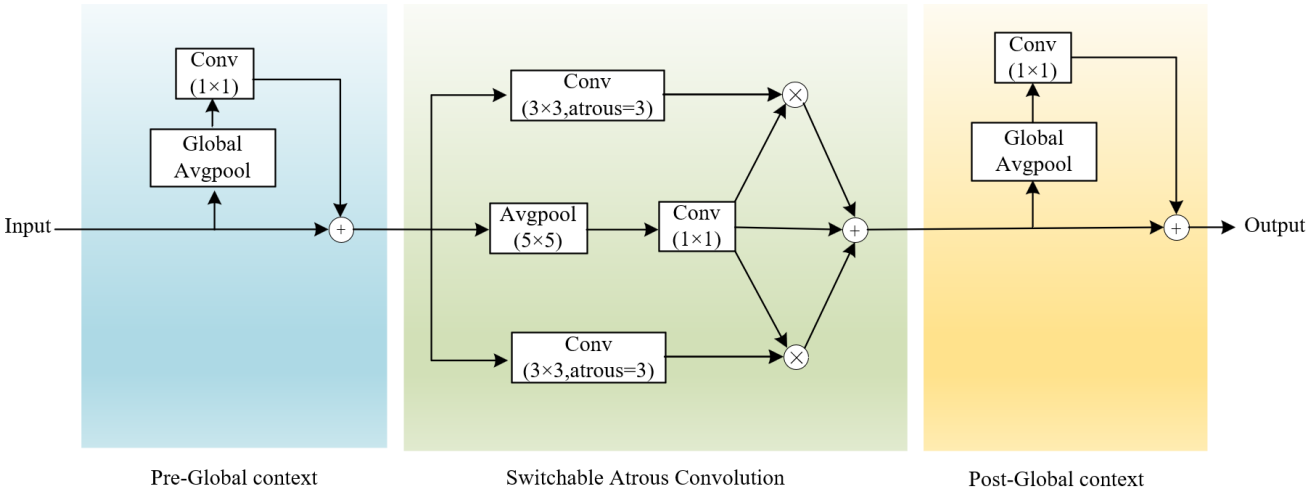

Figure 6. Implementation of Switchable Atrous Convolution (SAC)

The architecture of SAConv comprises three principal components: two global context modules are situated before and after the SAConv component. These modules facilitate a more comprehensive understanding of the image content, thereby enabling the SAConv component to operate efficiently in a broader range of contexts. The introduction of additional space (i.e. voids) into the convolution kernel enables the model to expand the receptive field, thereby capturing features at different scales while maintaining a constant number of parameters. Furthermore, SAConv employs switching functions to combine the results of convolutions with varying void rates. The switching functions are spatially dependent, whereby each position of the feature map may have a different switch to control the output of SAConv. This flexibility with respect to the size and scale of the features is a key advantage of this approach. The Switchable Atrous Convolution (SAC) structure is illustrated in Figure 5, which demonstrates the process of switching between different dilation rates and how these features are combined.

SAC achieves the conversion of traditional convolutional layers to SAConv layers by utilizing the same weights (with the exception of a trainable difference) for convolutional operations with varying nulling rates. The conversion mechanism comprises an average pooling layer and a 1×1 convolutional layer, which implement the switching function. The switch structure is as follows:

The following section outlines the various structural features that can be switched.

1) Double Observation Mechanism: The SAC observes the input features on two occasions, utilizing disparate null rates. The same set of input features is processed by two distinct configurations of convolutional kernels, with each configuration corresponding to a specific null rate. This enables the capture of feature information at varying scales, thereby facilitating a more comprehensive understanding and analysis of the input data.

2) Application of Switching Function: The outputs obtained from the different null rates are combined through a switching function. The switching function determines how the information from the two convolutions is selected or fused to generate the final output features. The double observation and combination strategy enables SAC to effectively handle complex feature patterns, thereby improving the flexibility and adaptability of feature extraction and enhancing the accuracy and efficiency in object detection and segmentation tasks. The implementation of Switchable Atreus Convolution (SAC) is as follows:

Conversion of traditional convolutional layers to SAC: Each 3×3 convolutional layer in the ResNet backbone network is converted to SAC. This conversion enables the convolutional computation to switch between different null rates in a soft manner.

Weight Sharing and Training Differences: SAC switches between different null rates, but all operations share the same weights with only one trainable difference. This reduces model complexity while maintaining flexibility.

Global Context Module: The context module adds image-level information to the features. The global context module helps the network to better understand and process the image as a whole, improving the quality and accuracy of feature extraction.

3.3 Switchable Atrous Convolution (SAC) implementation and analysis

The implementation of Switchable Atrous Convolution (SAC) further improves feature extraction by enabling the network to adjust its receptive field adaptively based on the input features. The switching function ensures that the convolution operations can adapt to different scales and maintain the efficiency of the network. This module is crucial for distinguishing between cracks and background textures in underground pipeline images, where lighting conditions and textures often lead to misclassification.

The implementation of Switchable Atrous Convolution (SAC) is visualized in Figure 6, which shows how the SAC mechanism is implemented within the network architecture. This figure illustrates the process of converting traditional convolutional layers to SAC and how the network processes multi-scale features effectively.

In order to more effectively evaluate the performance of the proposed Dysample and SAConv modules in comparison to traditional upsampling and feature fusion methods, a comprehensive comparative analysis is conducted. The detailed results of this evaluation are presented and summarized in Tables 1 and 2.

Compared to CARAFE [20] and fixed interpolation strategies, Dysample reduces computational cost while improving feature alignment, particularly beneficial under low-contrast and low-lux scenarios.

SAConv dynamically fuses multi-scale features using spatially-dependent switching functions, eliminating the overhead of traditional CRF layers and improving generalization across variable crack widths and textures.

These comparative results further validate the integration of Dysample and SAConv as both performance-enhancing and deployment-friendly solutions.

Discussion: Comparative Strengths and Limitations of Proposed Modules

To evaluate the technical strengths and potential trade-offs of the proposed improvements-namely the Dysample upsampling module and Switchable Atrous Convolution (SAConv)-we provide a comparative assessment against standard techniques.

1. Dysample upsampling vs. traditional methods

Dysample demonstrates clear benefits over CARAFE, bilinear interpolation, and nearest-neighbor approaches. As shown in Figure 7, it achieves the highest accuracy (mAP@0.5) while also incurring the lowest computational load (GFLOPs), making it ideal for embedded, low-power deployment scenarios. However, performance may degrade in highly uniform textures or under unclear contrast boundaries where sampling anchors become ambiguous.

Table 1. Advantages and trade-offs of Dysample

Upsampling Method	Computational Load (GFLOPs)	Accuracy (mAP@0.5)	Adaptability to Low Light	Suitability for Embedded Devices
Dysample (proposed)	2.1	61.0%	High	High
CARAFE	2.9	59.8%	Medium	Medium
Bilinear Upsampling	3.3	58.3%	Low	Low
Nearest Neighbor	3.1	57.6%	Low	High

Table 2. Advantages and trade-offs of SAConv

Feature Fusion Method	mAP@0.5	Receptive Field Adaptation	Post-Processing Required	Multiscale Capability
SAConv (proposed)	61.0%	Adaptive (dilated)	No	Strong
CRF Layer	59.1%	Fixed	Yes	Medium
ASPP	60.4%	Semi-adaptive	No	Strong
Standard Atrous Convolution	58.9%	Fixed (d=2 or 3)	No	Weak

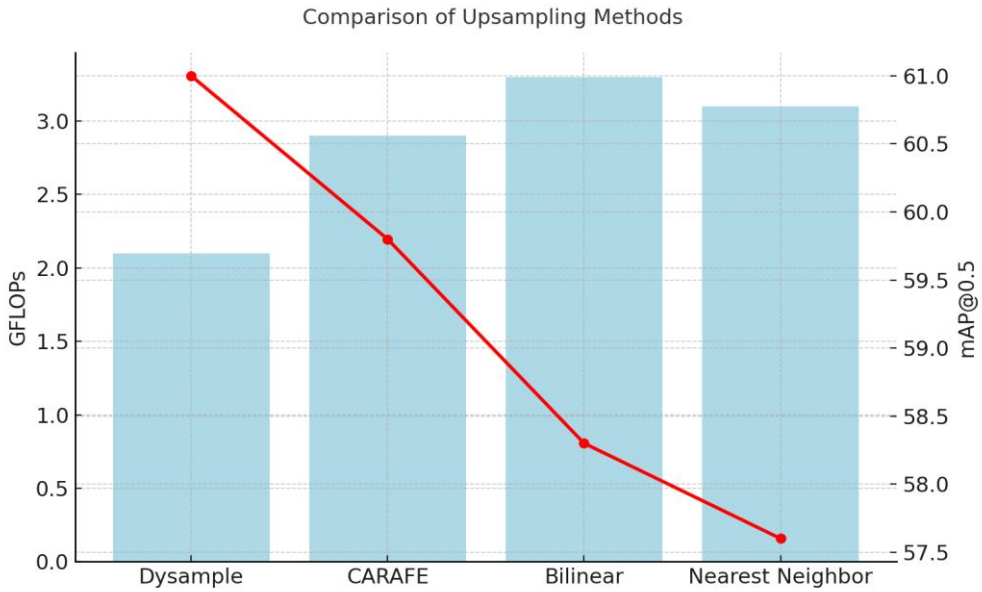

Figure 7. Comparison of different upsampling methods in terms of accuracy and computational complexity

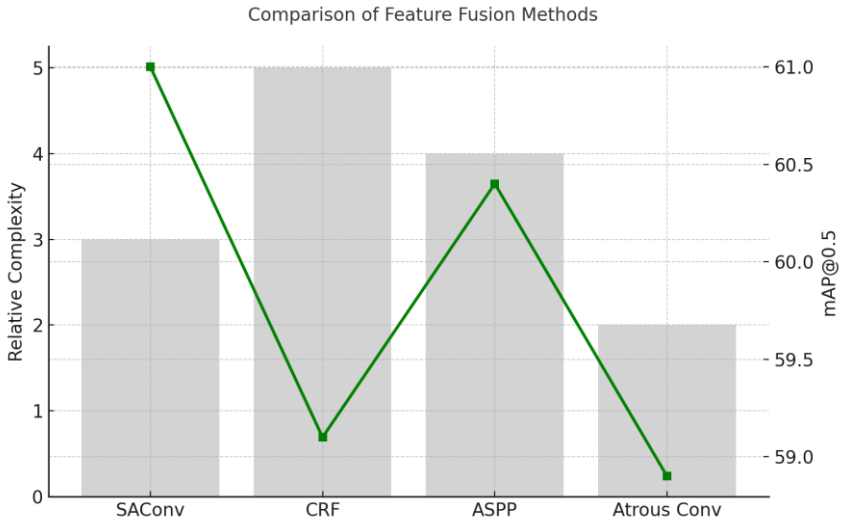

Figure 8. Comparison of feature fusion methods by detection accuracy and complexity

2. SAConv vs. existing fusion mechanisms

SAConv dynamically fuses multi-scale features using spatially-dependent switching functions. Compared to CRF, ASPP [21], and standard Atrous convolution, it avoids heavy post-processing, generalizes better under low-illumination noise, and maintains a relatively moderate architectural complexity, as shown in Figure 8. Its added complexity is justified by its ability to flexibly adapt receptive fields based on content-aware modulation.

In summary, while both modules provide measurable improvements over traditional components, their optimal performance depends on dataset properties and deployment constraints. These modules offer a promising direction for embedded, real-time crack detection in smart cities.

3.4 Implementation details for reproducibility

To facilitate replication and downstream optimization of the proposed framework, we provide detailed technical specifications as follows:

1. Backbone Network Configuration
   - Base model: RT-DETR with HGNetv2
   - Input resolution: 640 × 640 pixels (zero-padded as needed)
   - Stages: 4 feature extraction stages with down sampling rates of {4×, 8×, 16×, 32×}
   - Cross-scale fusion: Combination of FPN and BiFPN-inspired lateral connections
   - Transformer encoder: 6 layers, each with 8 attention heads and 512-dimensional hidden units
2. Dysample Upsampling Module
   - Sampling strategy: Grid-based bilinear interpolation
   - Offset generation:
   - Static range factor = 0.2 × image width
   - Dynamic offset via 1×1 convolution → Sigmoid activation
   - Feature resolution: Upsampled 2× per level
   - Adaptive kernel re-alignment: Uses three-layer MLP to learn task-specific offsets
3. SAConv Convolution Module
   - Dilated rates: {1, 3} dynamically selected per location
   - Switching function:
   - Composed of 3×3 convolution followed by global average pooling and a 1×1 projection
   - Softmax normalization applied to control contribution of each Atrous rate
   - Context integration: A global context module pre- and post-SAConv facilitates semantic aggregation, including global pooling and gating mechanisms
4. Detection Head and Loss Functions
   - Detection structure: Anchor-free decoupled head
   - Losses:
   - Box regression: CIoU Loss
   - Classification: Focal Loss ($\alpha = 0.25$, $\gamma = 2$)
   - Objectness: BCE With Logits Loss
   - Positive sample selection: Dynamic K-Matching as used in YOLOv5/6
5. Training Settings
   - Optimizer: Stochastic Gradient Descent (SGD)
   - Momentum: 0.937
   - Weight decay: 5e-4
   - Initial learning rate: 0.01 with cosine decay
   - Batch size: 32
   - Epochs: 600
   - Warm-up: 3-epoch linear increase
   - Gradient clipping: Enabled (max norm = 5.0)
6. Data Augmentation Strategy
   - Spatial augmentations:
   - Mosaic (p = 0.8), Random horizontal flip (p = 0.5)
   - Color space:
   - HSV jitter (hue ±0.015, saturation ±0.7, value ±0.4)
   - MixUp: Disabled to preserve low-light image integrity
7. Environment Specifications
   - Hardware: NVIDIA RTX 4090, 24GB VRAM
   - Framework: PyTorch 2.0.1 with CUDA 11.8
   - OS: Windows 10×64
   - Language: Python 3.9

These comprehensive implementation details are intended to ensure the reproducibility and scientific transparency of our proposed method. By disclosing all critical components and hyperparameters, we aim to facilitate further research and practical deployment of crack detection systems in low-illumination underground pipeline environments.

4. EXPERIMENT AND RESULTS

To rigorously evaluate the proposed algorithm, a specialized dataset of 2,700 professionally annotated low-illumination underground pipeline images was constructed. The images were collected from real-world municipal underground inspection scenarios using a pipeline inspection robot equipped with a low-light CMOS imaging system [22] and supplementary lux-level sensors. The illumination range across the dataset spans from 5 to 25 lux, simulating challenging lighting conditions typically encountered in subterranean environments.

Each image in the dataset was annotated by a team of domain experts with backgrounds in structural engineering and urban utility maintenance. The annotations were performed using a semi-automated labeling system that incorporated initial predictions from a baseline YOLOv8 model, followed by manual refinement to ensure pixel-level precision. All annotations were cross-validated by a second annotator and reviewed through consensus when discrepancies arose.

The dataset contains a single object category—pipeline cracks—but includes a wide variety of crack morphologies, such as:

- Linear cracks: thin, elongated discontinuities often aligned with stress directions
- Transverse cracks: perpendicular to the pipe's axis, often caused by ground movement
- Mesh cracks: fine-grained, interconnected patterns due to concrete shrinkage
- Block cracks: larger, rectangular crack segments formed by material fatigue

Crack widths in the dataset range from approximately 0.3 mm to 4.5 mm, as verified using high-precision caliper measurements during data acquisition. The dataset also accounts for environmental variability including surface texture noise, moisture levels, and partial occlusions, with SSIM (Structural Similarity Index) [23] values between cracks and background ranging from 0.70 to 0.85, confirming the visual ambiguity in low-lux scenarios.

The dataset was divided into 1,890 training images, 540 validation images, and 270 test images, as summarized in Table 3, which presents the image and instance distributions. All experiments were conducted using a workstation with an NVIDIA RTX 4090 GPU, PyTorch 2.0, CUDA 11.8, and Python 3.9. Training was performed for 600 epochs with consistent hyperparameters across comparative and ablation studies.

Table 3. Database classification and quantity

Category	Training Samples	Validation Samples	Test Samples
Cracks	1,890	540	270

4.1 Evaluation metrics

Precision (P), Recall (R), Mean Average Precision (mAP), and mAP50-95 were used as evaluation criteria. A higher P suggests greater detection accuracy, resulting in fewer false positives. A higher R indicates that the system detects all targets as much as feasible, resulting in fewer false negatives. An increasing mAP value indicates greater algorithmic detection precision. These parameters can be determined using the formulas listed below [24]:

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$AP = \int_0^1 P(R)dR \tag{4}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{5}$$

In the context of object detection, the above metrics are used to evaluate model performance.

True Positives (TPs) are instances that have been appropriately identified as positive samples. It represents the model's ability to recognize positive specimens.

False positives (FP) are negative specimens that were wrongly labeled as positive. In the context of autonomous driving, this may imply that the system incorrectly detects non-road things as road objects, resulting in false detections.

False Negatives (FNs) are instances that should have been identified as positive but were wrongly identified as negative. These are actual positive samples that the model did not detect, suggesting missed detections.

AP (Average Precision) is the average precision value over recall criteria.

mAP (mean Average Precision): Average Correctness mean values across all classes.

4.2 Algorithm comparison experiments

The enhanced methodologies' efficacy in identifying subterranean pipe fissures is illustrated through the selection of an underground pipe crack dataset for experimental verification and its comparison with other prevalent target detection techniques in crack target visualization experiments, as shown in Table 4. This table compares the performance of various algorithms in terms of mAP50, mAP50-95, Precision (P), and Recall (R). Figure 9 illustrates the performance of each algorithm in the context of an underground pipeline. The red arrows in the figure demonstrate that cracks in the pipe are not identified in the presence of low light intensity and background interference. The figure demonstrates that the image detection algorithms SSD [25], Yolov5n, Yolov6n, Yolov7-tiny and Yolov8n are unable to detect the crack targets in situations characterized by low light intensity and a complex background.

It is evident that only the algorithms that have been optimized for the network framework are capable of detecting the crack targets. As evidenced by the green arrows in the figure, other mainstream target detection methods exhibit a significant propensity for false detection in scenarios characterized by high levels of interference. The replacement of the original upsampler with Dysample and the original CRF convolutional layer with SAConv has resulted in a notable enhancement in the algorithm's capacity to detect cracks across a range of scales. Additionally, the background interference has been reduced, thereby making the algorithm more resilient. The enhanced method has demonstrated greater

accuracy in detecting crack targets, and the visualization experiments have substantiated its superiority in this regard, particularly in the context of underground pipeline detection.

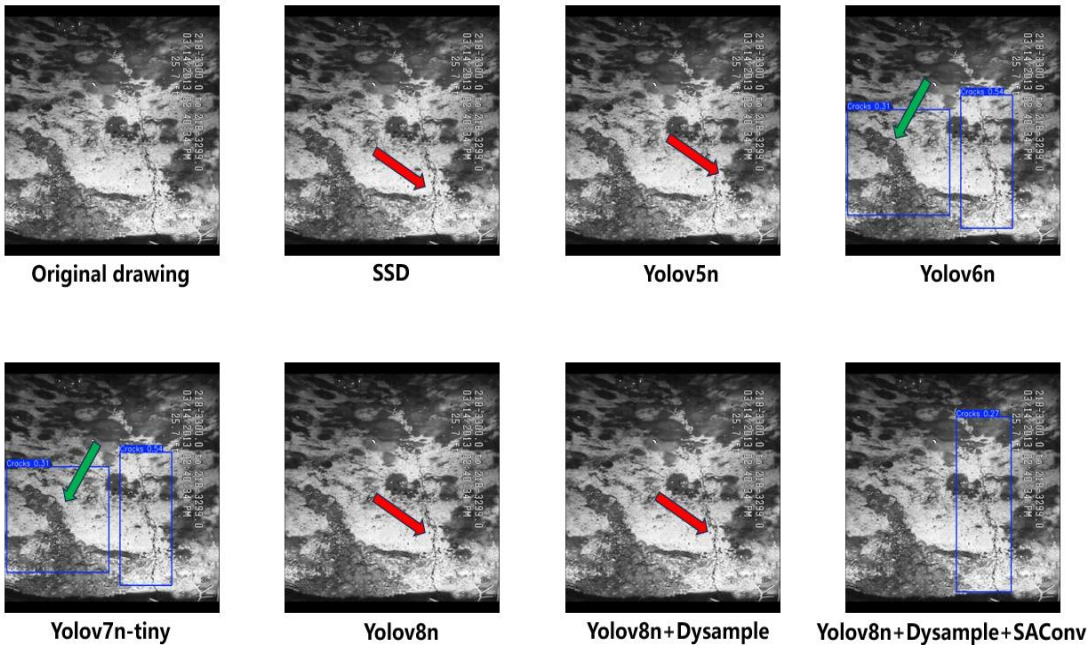

Figure 9. Comparison of results across models

Table 4. Comparing the algorithms' performance

Algorithm	mAP50	mAP50-95	P%	R%
SSD	38.13	/	84.4	6.92
Yolov5n	0.617	0.250	72.83	69.78
Yolov6n	0.596	0.197	67.3	63.2
Yolov7-tiny	0.654	0.226	72.28	70.87
Yolov8n	0.599	0.231	73.14	63.47
improved method	0.6097	0.2359	70.9	79.57

4.3 Comparison of algorithm robustness

The underground pipeline crack dataset is the optimal choice for robustness testing in this test result. The test results clearly show that the improved method has a significant optimization improvement in all key index parameters compared with the baseline algorithm YOLOv8n. The key index of accuracy P increased by 1.3%, while the key index of recall R increased significantly, by 7.88%. The key index of mAP50 increased significantly, by 7.3%. mAP50-90 increased by 5.9%. It also increased by 5.9% compared with other mainstream target detection algorithms. The key index of mAP50-90 increases by 5.9%. It also has clear advantages over other mainstream target detection algorithms. Experiments on the underground pipe crack target detection dataset prove the improved method is robust in target detection.

4.4 Ablation study

To ascertain the precise impact of the proposed improvement modules on the performance of the algorithms, we have conducted ablation experiments. These have involved embedding each improvement module into the original algorithm in turn, under the same experimental conditions and using the underground pipeline crack dataset. We have selected P%, recall%, mAP50, and mAP50-90 as our evaluation indexes to verify the effectiveness of the improved method. The results of the ablation experiments are shown in Table 5. Firstly, the original YOLOv8n algorithm is tested against the following benchmark values: P of 73.1%, R of 63.47%, mAP50 of 59.9%, and mAP50-95 of 23.08%. Next, the point-sampling based Dysample upsampler replaces the original kernel-based dynamic upsampler, resulting in a 13.2% R improvement. The latter model replaces the original CRF convolutional layer with SAConv. Compared to the YOLOV8 benchmark, each experimental result metric has risen significantly. Recall has improved by 16.08%, mAP50 by 1.06%, and mAP50-95 by 0.5%. This proves that the improved model significantly improves the multi-scale feature extraction capability for crack targets.

Table 5. Ablation experiment

Group	Yolov8n	Dysample	SAConv	Recall	mAP50	mAP50-95	P
1	√			63.47%	59.9%	23.1%	73.1%
2	√	√		76.96%	59.4%	22.11%	68.6%
3	√	√	√	**79.6%**	61.0%	23.6%	70.9%

The ablation experiments prove that each of the improved modules markedly enhances the algorithm's performance, and they work even better in combination. These improvements not only bolster the algorithm's multi-scale feature extraction and integration capabilities, but also sharpen its focus on the target of cracks within the pipeline.

The proposed algorithm demonstrated significant performance improvements, particularly in recall (↑16.08%) and mAP@0.5 (↑1.06%) compared to the baseline YOLOv8n, suggesting enhanced detection sensitivity and localization

accuracy in low-light underground environments. These results are not only quantitatively promising but also reflect a qualitative shift in the model's ability to distinguish between crack features and background textures under challenging conditions.

1. Performance Across Crack Types

A qualitative review of the detection outputs reveals that the enhanced model performs particularly well in identifying:

- Linear and transverse cracks: The use of the SAConv module, with its multiscale receptive field adaptation, improves the detection of elongated discontinuities, which typically suffer from partial occlusion or low edge contrast.
- Mesh cracks: Despite their fine granularity and fragmented structure, these are better detected due to the Dysample upsampler, which improves feature alignment and resolution preservation in small-scale features.
- Low-contrast cracks embedded in concrete textures: The model demonstrates resilience against grayscale similarity (SSIM ≈ 0.78), likely due to the integration of global context modules in SAConv, which guide the network towards semantic consistency.

However, block cracks with irregular boundaries or significant occlusions remain partially challenging. While recall rates improved, some false positives persist where the crack edges fade into similar background textures.

2. Algorithmic Contributions to Performance Gains

The following architectural modifications underpin the observed gains:

- Dysample Module: Enhances spatial granularity and spatial resolution alignment during upsampling, reducing the loss of fine crack details.
- SAConv Module: Facilitates multiscale feature learning with spatially adaptive dilation, enabling better detection of both micro and macro-scale crack features.
- RT-DETR Backbone with HGNetv2: Captures long-range dependencies, especially useful in elongated structures like pipe cracks, which may extend beyond the receptive field of conventional CNNs.

3. Areas for Future Improvement

- Despite the promising results, several avenues remain for performance refinement:
- Crack severity grading: The current framework focuses on binary detection. Future work could incorporate a regression branch to estimate crack width, depth, or severity, using annotated physical measurements.
- Data augmentation under motion blur and water occlusion: Scenarios with flowing water or smear-induced blur still reduce detection confidence. Synthetic data augmentation using domain randomization may help improve robustness.
- Lightweight deployment optimization: Although current performance is suitable for embedded GPU inference, further compression (e.g., quantization-aware training, knowledge distillation) can reduce memory footprint and power consumption for edge devices.
- Temporal consistency modeling: Incorporating temporal information across video frames (e.g., via ConvLSTM or 3D CNNs [26]) could stabilize detection in dynamic environments, reducing false alarms due to transient lighting variations or sensor noise.

5. CONCLUSION

This study presents a novel crack detection framework for urban underground pipelines, leveraging two key modules—Dysample and SAConv—built upon an RT-DETR-HGNetv2-enhanced YOLOv8n architecture. The proposed method significantly improves recall and precision under low-light, texture-rich conditions, with a notable 16.08% recall gain and 1.06% mAP@0.5 improvement compared to the baseline. These enhancements are attributed to adaptive multiscale feature learning and robust upsampling mechanisms that are resilient to grayscale similarity and contrast degradation.

Given its robustness in low-illumination environments and real-time inference capability (58 FPS on RTX 4090), the algorithm is well-suited for a variety of practical applications, including:

Municipal infrastructure inspection: Automated monitoring of sewer and drainage pipelines to detect early-stage cracks, thereby reducing manual inspection costs and improving maintenance schedules. Post-disaster pipeline assessment: Rapid screening of water, gas, or cable conduits following earthquakes, floods, or subsidence events, where structural damage may not be visible externally. Smart city IoT integration: Deployment on embedded systems within autonomous inspection robots, enabling real-time condition monitoring as part of a digital twin system. Industrial facility maintenance: Monitoring of in-factory piping networks in oil, chemical, and power plants, where internal corrosion-induced cracks could lead to safety hazards.

Edge device limitations: While the algorithm is efficient on GPU platforms, embedded devices such as Jetson Nano or Raspberry Pi may require further model compression or pruning to meet power and memory constraints. Generalization under domain shift: The current dataset covers a specific urban underground pipeline texture domain. Performance may degrade when applied to different materials (e.g., metal vs. concrete), environmental noise, or lighting artifacts. Annotation cost: High-quality crack annotations are expensive and time-consuming, limiting dataset scalability. Weakly supervised or self-supervised approaches may offer alternative labeling strategies. Real-time occlusion: In dynamic environments with water flow, floating debris, or lens smudges, detection accuracy may decrease, necessitating temporal consistency modeling or filtering strategies.

Model compression and quantization: Applying knowledge distillation, mixed-precision training, or neural architecture search (NAS) to create ultra-lightweight variants for deployment on low-power edge devices. Multimodal data fusion: Integrating thermal, acoustic, or LIDAR data with visual cues to improve crack detection in occluded or ambiguous regions. Crack severity estimation and 3D reconstruction: Extending the current detection system to estimate crack width, depth, and spatial extent, providing richer semantic information for maintenance decision-making. Continual learning and domain adaptation: Developing frameworks that adapt to new pipeline environments without retraining from scratch, thus reducing the burden of data recollection. Pipeline-level anomaly tracking: Incorporating temporal information from inspection videos to enhance stability and reduce false alarms using motion-aware neural modules (e.g., ConvLSTM, attention-based video transformers).

REFERENCES

[1] Tubb, R. (2017). P&GJ's 2017 worldwide pipeline construction report. Pipeline & Gas Journal, 244(1): 16-

20.
[2] Rayhana, R., Jiao, Y., Bahrami, Z., Liu, Z., Wu, A., Kong, X. (2021). Valve detection for autonomous water pipeline inspection platform. IEEE/ASME Transactions on Mechatronics, 27(2): 1070-1080. https://doi.org/10.1109/TMECH.2021.3079409
[3] Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., Fieguth, P. (2015). A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. Advanced Engineering Informatics, 29(2): 196-210. https://doi.org/10.1016/j.aei.2015.01.008
[4] Mohamad, A.J., Ali, K., Rifai, D., Salleh, Z., Othman, A.A.Z. (2023). Eddy current testing methods and design for pipeline inspection system: A review. Journal of Physics: Conference Series, 2467(1): 012030. https://doi.org/10.1088/1742-6596/2467/1/012030
[5] Ravishankar, P. (2023). Increasing the oil and gas pipeline resiliency using image processing algorithms. Doctoral dissertation, Lamar University-Beaumont.
[6] Mysiuk, R., Yuzevych, V., Mysiuk, I., Tyrkalo, Y., Pavlenchyk, A., Dalyk, V. (2023). Detection of surface defects inside concrete pipelines using trained model on JetRacer kit. In 2023 IEEE 13th International Conference on Electronics and Information Technologies (ELIT), Lviv, Ukraine, pp. 21-24. https://doi.org/10.1109/ELIT61488.2023.10310691
[7] Altabey, W.A., Kouritem, S.A., Abouheaf, M.I., Nahas, N. (2022). A deep learning-based approach for pipeline cracks monitoring. In 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Maldives, Maldives, pp. 1-6. https://doi.org/10.1109/ICECCME55909.2022.9987998
[8] Xin, J., Zhang, W., Lu, R.K., Chen, J., Zhu, H., He, R. (2022). Numerical simulation of pipeline crack detection probe with poly-magnetic structure. Journal of Physics: Conference Series, 2383(1): 012031. https://doi.org/10.1088/1742-6596/2383/1/012031
[9] Ibragimova, E. (2024). Design of a camera-enabled mobile robot for in-pipe inspection. Referred & Reviewed Journal, 23(5): 46.
[10] Shah, J., El-Hawwat, S., Wang, H. (2023). Guided wave ultrasonic testing for crack detection in polyethylene pipes: laboratory experiments and numerical modeling. Sensors, 23(11): 5131.
[11] Jamshidi, V. (2023). Simulation of sand particles detection inside a pipeline by photon radiography. Applied Radiation and Isotopes, 199: 110876. https://doi.org/10.1016/j.apradiso.2023.110876
[12] Yang, J., Wang, W., Lin, G., Li, Q., Sun, Y., Sun, Y. (2019). Infrared thermal imaging-based crack detection using deep learning. IEEE Access, 7: 182060-182077. https://doi.org/10.1109/ACCESS.2019.2958264
[13] Zolfaghari, A., Zolfaghari, A., Kolahan, F. (2018). Reliability and sensitivity of magnetic particle nondestructive testing in detecting the surface cracks of welded components. Nondestructive Testing and Evaluation, 33(3): 290-300. https://doi.org/10.1080/10589759.2018.1428322
[14] Chu, Z., Jiang, Z., Mao, Z., Shen, Y., Gao, J., Dong, S. (2021). Low-power eddy current detection with 1-1 type magnetoelectric sensor for pipeline cracks monitoring. Sensors and Actuators A: Physical, 318: 112496. https://doi.org/10.1016/j.sna.2020.112496
[15] Mazni, M., Husain, A.R., Shapiai, M.I., Ibrahim, I.S., Anggara, D.W., Zulkifli, R. (2024). An investigation into real-time surface crack classification and measurement for structural health monitoring using transfer learning convolutional neural networks and Otsu method. Alexandria Engineering Journal, 92: 310-320. https://doi.org/10.1016/j.aej.2024.02.052
[16] Li, Y., Han, Z., Xu, H., Liu, L., Li, X., Zhang, K. (2019). YOLOv3-lite: A lightweight crack detection network for aircraft structure based on depthwise separable convolutions. Applied Sciences, 9(18): 3781. https://doi.org/10.3390/app9183781
[17] Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Wei, X. (2022). YOLOv6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976. https://doi.org/10.48550/arXiv.2209.02976
[18] Wei, L., Tong, Y. (2024). Enhanced-YOLOv8: A new small target detection model. Digital Signal Processing, 153: 104611. https://doi.org/10.1016/j.dsp.2024.104611
[19] Qiao, S., Chen, L.C., Yuille, A. (2021). Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10213-10224. https://doi.org/10.1109/CVPR46437.2021.01008
[20] Fu, R., Hu, Q., Dong, X., Gao, Y., Li, B., Zhong, P. (2024). Lighten CARAFE: Dynamic lightweight upsampling with guided reassemble kernels. In International Conference on Pattern Recognition, pp. 383-399. https://doi.org/10.1007/978-3-031-78128-5_25
[21] Liu, R., Tao, F., Liu, X., Na, J., Leng, H., Wu, J., Zhou, T. (2022). RAANet: A residual ASPP with attention framework for semantic segmentation of high-resolution remote sensing images. Remote Sensing, 14(13): 3109. https://doi.org/10.3390/rs14133109
[22] Wang, F., Dai, M., Sun, Q., Ai, L. (2021). Design and implementation of CMOS-based low-light level night-vision imaging system. Seventh Symposium on Novel Photoelectronic Detection Technology and Applications, 11763: 1518-1529. https://doi.org/10.1117/12.2587259
[23] Bakurov, I., Buzzelli, M., Schettini, R., Castelli, M., Vanneschi, L. (2022). Structural similarity index (SSIM) revisited: A data-driven approach. Expert Systems with Applications, 189: 116087. https://doi.org/10.1016/j.eswa.2021.116087
[24] Mahasin, M., Dewi, I.A. (2022). Comparison of CSPDarkNet53, CSPResNeXt-50, and EfficientNet-B0 backbones on YOLO v4 as object detector. International Journal of Engineering, Science and Information Technology, 2(3): 64-72. https://doi.org/10.52088/ijesty.v1i4.291
[25] Sehwag, V., Chiang, M., Mittal, P. (2021). Ssd: A unified framework for self-supervised outlier detection. arXiv preprint arXiv:2103.12051. https://doi.org/10.48550/arXiv.2103.12051
[26] Aravinda, C.V., Al-Shehari, T., Alsadhan, N.A., Shetty, S., Padmajadevi, G., Reddy, K.R. (2025). A novel hybrid architecture for video frame prediction: Combining convolutional LSTM and 3D CNN. Journal of Real-Time Image Processing, 22(1): 1-18. https://doi.org/10.1007/s11554-025-01626-w