Jurnal Teknik Informatika (JUTIF) Vol. 7, No. 3, June 2026, Page. 2577-2587
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5544

CCTV-Based River Waste Detection Using a Hybrid CNN–Graph Attention Network with Spatial–Contextual Feature Learning

Asep Surahmat*[1], Lukas Umbu Zogara[2], Fajar Muttaqi[3]

[1,2,3]Faculty Technology and Design, Utpadaka Swastika University, Indonesia

Email: [1]asep.surahmat@utpas.ac.id

Received : Dec 6, 2025; Revised : Jan 25, 2026; Accepted : Feb 9, 2026; Published : Jun 15, 2026

Abstract

River waste accumulation has become a serious environmental problem in urban areas, particularly in highly polluted rivers such as the Angke River in Tangerang, where floating waste disrupts ecological balance and increases flood risk. Conventional computer vision–based detection methods often fail under dynamic river conditions due to water surface reflections, turbulence, occlusion, and visually ambiguous debris. This study aims to improve the accuracy and robustness of river waste detection by proposing a hybrid deep learning framework that integrates convolutional and graph-based spatial–contextual reasoning. The proposed method utilizes a ResNet50 backbone for feature extraction from CCTV imagery, followed by spatial graph construction that models adjacency relationships between image regions. A Graph Attention Network (GAT) is then applied to capture contextual dependencies and refine feature representations prior to classification. Unlike conventional CNN-only or YOLO-based detectors that rely primarily on local visual cues and bounding-box representations, the proposed approach explicitly models spatial–contextual relationships between image regions through graph-based attention mechanisms. Experiments were conducted on 4,200 CCTV image frames collected from the Angke River under varying environmental conditions. The proposed model achieved an accuracy of 92.4%, precision of 91.1%, recall of 93.2%, F1-score of 91.9%, and a mean Average Precision (mAP) of 0.78, outperforming CNN-only and YOLO-based baseline models. These findings highlight the contribution of graph-enhanced visual reasoning to the fields of Computer Vision and Intelligent Surveillance, particularly for real-time environmental monitoring systems operating in complex and dynamic visual environments.

Keywords : *CCTV Imagery, Environmental Monitoring, Graph Neural Network, River Waste Detection, Spatial–Contextual Features.*

1. INTRODUCTION

River pollution has become a critical environmental concern in many rapidly urbanizing regions due to increasing waste production and insufficient management infrastructure [1]. In Indonesia, Angke river in Tangerang is among the rivers that frequently experience severe floating waste accumulation as a result of residential, commercial, and industrial activities surrounding the watershed [2]. The buildup of unmanaged waste contributes to a decline in water quality, disrupts aquatic ecosystems, and increases flood hazards during the rainy season [3]. To support early mitigation efforts, continuous environmental monitoring is required to identify waste distribution patterns and detect pollution hotspots along the river surface [4].

Unlike existing river waste detection studies that primarily rely on convolutional or bounding-box–based models, this study introduces a graph-based spatial–contextual learning framework that explicitly models relational dependencies between river surface regions using CCTV imagery. CCTV monitoring systems have been increasingly deployed by environmental agencies because they provide wide-area coverage at low operational cost and operate continuously under various weather conditions [5]. Despite these advantages, manual observation of CCTV feeds remains highly inefficient due to

Vol. 7, No. 3, June 2026, Page. 2577-2587
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5544

operator fatigue, inconsistent visual attention, and the high volume of daily river activity requiring surveillance [6]. Consequently, computer vision and deep learning methods have become a promising solution for automating river waste detection from visual data [7]. Convolutional Neural Networks (CNNs) have achieved strong performance in object detection tasks across multiple domains due to their hierarchical feature extraction capabilities [8]. However, CNN-based methods often struggle in dynamic river environments where water surface reflections, occlusion, and fluctuating textures obscure the appearance of floating waste objects [9].

The visual complexity of rivers such as Angke river poses additional challenges, including irregular lighting conditions, variable water turbidity, and the presence of organic and inorganic debris that often blend with the background [10]. CNNs typically process localized pixel regions and therefore lack the ability to incorporate broader spatial dependencies that are crucial for distinguishing visually ambiguous objects [11]. Although several studies have explored river waste detection using deep learning, most existing approaches remain limited to convolutional architectures that do not explicitly model spatial relationships across regions within an image [12]. This limitation reduces detection robustness in scenarios where contextual information is essential, as frequently observed in Angke river[13]. Additionally, no prior research has integrated graph-based relational reasoning with CCTV river surveillance in Indonesia, leaving an unresolved gap in developing context-aware detection models tailored to the dynamic nature of local river environments [14].

Advancements in Graph Neural Networks (GNNs) offer promising solutions by enabling relational reasoning through structured representations of image regions [15]. GNNs treat image segments or feature patches as nodes within a graph and model spatial interactions via edges that encode adjacency or similarity relationships [16]. This representation allows GNNs to capture contextual relationships that standard CNNs struggle to learn, particularly under noisy or cluttered visual conditions [17]. Graph Attention Networks (GAT), in particular, enhance this capability by adaptively weighting node interactions to highlight more informative spatial connections [18]. Hybrid models that combine CNN feature extraction with GNN-based relational reasoning have demonstrated superior performance in tasks requiring context-aware interpretation compared to CNN-only models [19]. These capabilities make GNNs a strong candidate for addressing the complexities of river waste detection, where distinguishing objects from background water textures often requires understanding relational patterns rather than isolated pixel intensities [20].

This research introduces a spatial–contextual graph construction mechanism that transforms CCTV river imagery into a structured graph representation, enabling the model to encode relational information between adjacent river regions [21]. The proposed hybrid CNN–GNN architecture incorporates attention-based graph reasoning to enhance discrimination of visually ambiguous waste objects in challenging river scenes such as those found in Angke river[22]. Unlike existing methods that treat waste detection as a purely pixel-level or bounding-box classification task, this study integrates neighborhood-context interactions into the detection process, resulting in improved robustness against water reflections, occlusion, and scene clutter [23].

Therefore, this work aims to develop a deep-learning framework capable of accurate and reliable floating waste detection using CCTV data from Angke river, Tangerang [24]. The proposed method constructs graph-based spatial representations derived from CNN features to improve contextual reasoning during detection [25]. The research further evaluates the effectiveness of the hybrid architecture under various environmental conditions and compares it with state-of-the-art baseline models to demonstrate its advantages in real-world river monitoring scenarios [26]. The overall goal of this study is to contribute a more accurate, context-aware, and operationally feasible solution for automated river waste monitoring systems to support environmental management efforts in Indonesia [27]. The remainder of this paper presents the methodological design, experimental results, discussion,

Vol. 7, No. 3, June 2026, Page. 2577-2587
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5544

and concluding insights that validate the effectiveness of the proposed model [28]. Several previous studies have applied CNN-based and YOLO-based approaches for river waste detection and achieved promising results under controlled visual conditions. However, these methods generally focus on localized feature extraction or bounding-box representations and do not explicitly model spatial–contextual relationships between neighboring regions, which are crucial in dynamic river environments characterized by reflections, turbulence, and occlusion. In contrast, the proposed approach integrates graph-based relational reasoning to capture inter-region dependencies, enabling more robust detection under complex and visually ambiguous river surface conditions.

This study aims to develop an accurate and robust river waste detection framework by integrating convolutional feature extraction with graph-based spatial–contextual reasoning using CCTV imagery. Specifically, this research seeks to address the limitations of conventional CNN-based and YOLO-based detectors in dynamic river environments characterized by reflections, turbulence, and partially submerged objects. The main contributions of this study are threefold: (1) the formulation of a spatial–contextual graph representation derived from CNN feature maps to explicitly model relational dependencies between river surface regions; (2) the integration of a Graph Attention Network to adaptively refine spatial interactions and enhance detection robustness under complex visual conditions; and (3) a comprehensive evaluation using real-world CCTV data from the Angke River, demonstrating superior performance compared to CNN-only and YOLO-based baseline approaches.

2. METHOD

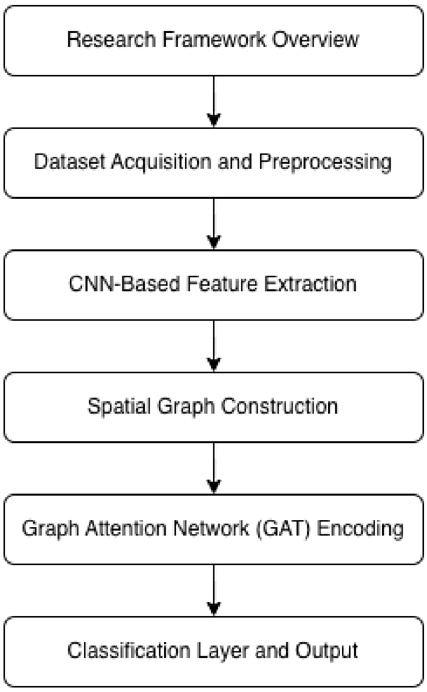

Figure 1. Overview of the proposed hybrid CNN–Graph Attention Network framework for CCTV-based river waste detection.

2.1 Research Framework Overview

This study proposes a Graph Neural Network (GNN)–based model for river waste detection that integrates convolutional feature extraction with spatial–contextual relational learning. The research framework consists of four main stages: data acquisition and preprocessing, feature extraction using a Convolutional Neural Network (CNN), spatial graph construction, and graph-based classification. This hybrid machine learning pipeline is designed to enhance detection robustness against the dynamic visual characteristics of river surfaces.

Jurnal Teknik Informatika (JUTIF) Vol. 7, No. 3, June 2026, Page. 2577-2587
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5544

2.2 Dataset Acquisition and Preprocessing

The dataset was obtained from CCTV units installed along Angke river in Tangerang, which continuously monitor river surface conditions. A total of 4,200 image frames were extracted from the CCTV recordings across varying lighting conditions, weather scenarios, and water-surface fluctuations. Each frame was manually annotated into two categories: *waste* and *non-waste*. Preprocessing procedures included resizing all images to 224 × 224 pixels, applying pixel normalization, reducing noise using a Gaussian filter, and performing data augmentation such as rotation, brightness adjustment, and horizontal flipping. Image normalization is applied to standardize pixel intensity values, as defined in Equation (1).

$$I' = \frac{I-u}{o} \tag{1}$$

where I is the original image, u is the mean pixel value, and o is the standard deviation

2.3 CNN-Based Feature Extraction

The ResNet50 architecture was used as the CNN backbone due to its strong feature extraction capability and efficiency for large-scale image processing. The CNN transforms each preprocessed input into a spatial feature map FFF, capturing mid- and high-level patterns associated with waste objects. The CNN backbone transforms the input image into a spatial feature map as expressed in Equation (2).

$$\boldsymbol{F} = \boldsymbol{CNN}(\boldsymbol{I'}) \tag{2}$$

This feature map, with dimensions *H* x *W* x *C,* serves as the foundation for generating graph nodes that represent local spatial regions of the rivers surface.

2.4 Spatial Graph Construction

To encode the spatial structure of the image, the feature map is divided into NNN patches, each corresponding to a node viv_ivi. Spatial adjacency is defined using a 4-neighborhood approach, where nodes representing adjacent patches are connected by edges. Spatial relationships between image patches are encoded using an adjacency matrix, as defined in Equation (3).

$$\boldsymbol{A_{ij}} = \begin{cases} \mathbf{1} \\ \mathbf{0} \end{cases} \tag{3}$$

If patch $\boldsymbol{i}$ is spatially adjacent to patch $\boldsymbol{j}$ otherwise. The resulting graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ consists of node set $\boldsymbol{V}$ representing image patches and edge set $\boldsymbol{E}$ describing spatial connectivity. Node features $\boldsymbol{h_i}$ are extracted from the corresponding region in the CNN feature map.

2.5 Graph Attention Network (GAT) Encoding

A Graph Attention Network (GAT) is employed to enhance representation learning by enabling each node to attend selectively to its spatial neighbors. Attention coefficients are computed to determine the importance of each neighboring node during feature aggregation. The attention coefficients used to aggregate neighboring node features are computed using the Graph Attention mechanism described in Equations (4) and (5).

$$\boldsymbol{a_{ij}} = \boldsymbol{softmax}\,(\boldsymbol{a^T}[\boldsymbol{Wh_i}|\,|\boldsymbol{Wh_j}]) \tag{4}$$

Node embeddings are updated using:

$$\boldsymbol{h'_i} = \boldsymbol{o}\,(\textstyle\sum_{\boldsymbol{jeN(i)}} \boldsymbol{a_{ij}Wh_j} \tag{5}$$

P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2577-2587
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5544

This process allows the model to emphasize meaningful spatial relationships while reducing the influence of noise such as reflections and water-surface distortions commonly present in CCTV river imagery.

2.6 Classification Layer and Output

A global average pooling operation is applied to the node embeddings to obtain a single graph-level representation. This representation is then passed into a fully connected layer to classify whether an image contains floating waste. The final classification output is computed using a fully connected layer, as shown in Equation (6).

$$\boldsymbol{y} = \boldsymbol{softmax}(\boldsymbol{W_o h'} + \boldsymbol{b_0}) \tag{6}$$

The model outputs a probability value indicating the presence of waste in the input frame, enabling automated detection for real-time river monitoring systems.

3. RESULT

3.1 Overall Performance on the Test Dataset

The proposed CNN–GNN model was evaluated using 1,200 test frames captured from CCTV recordings along Angke river. These frames captured a wide range of environmental variations, including direct sunlight, cloud coverage, low-light evening conditions, shadowed regions, and rapid water movement. The quantitative evaluation demonstrates that the model performs reliably across these conditions. As shown in Table 1, the model achieved an accuracy of 92.4%, indicating that the majority of frames were correctly classified. Precision reached 91.1%, showing that most detections corresponded to actual waste objects, while the recall value of 93.2% reflects the model's strong ability to detect waste consistently with minimal missed objects. The F1-score of 91.9% confirms a balanced ratio between precision and recall. Meanwhile, the mAP value of 0.78 highlights the model's stability in predicting waste objects across various confidence thresholds. Together, these metrics demonstrate that the proposed model can effectively perform waste detection in real-world river monitoring settings.

Table 1. Performance Metrics of the Proposed CNN–GNN Model

Metric	Score
Accuracy	92.4%
Precision	91.1%
Recall	93.2%
F1-Score	91.9%
mAP	0.78

The mean Average Precision (mAP) was computed following standard object detection evaluation protocols using an Intersection over Union (IoU) threshold of 0.5. A detection was considered correct when the IoU between the predicted region and the ground-truth annotation exceeded this threshold. The Average Precision (AP) was calculated for the waste class based on the precision–recall curve, and the final mAP score was obtained by averaging the AP values across all evaluated samples. This evaluation scheme ensures a fair assessment of detection accuracy under varying confidence thresholds.

3.2 Comparison with Baseline Models

To validate the superiority of the proposed method, the model's performance was compared with a CNN-only classifier based on ResNet50 and a YOLOv5 detector. The comparison results in Table 2 show substantial improvements across all evaluation metrics. The CNN-only model struggled in

Jurnal Teknik Informatika (JUTIF) Vol. 7, No. 3, June 2026, Page. 2577-2587
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5544

scenarios dominated by water reflections and low-contrast areas, resulting in limited detection capability. YOLOv5 performed better than the CNN-only approach, but it remained sensitive to conditions where waste objects were partially submerged or visually similar to the surrounding water surface. In contrast, the proposed CNN–GNN architecture achieved the highest accuracy, F1-score, and mAP due to its ability to integrate contextual information from spatially related regions. This graph-based relational learning allowed the model to better differentiate true waste objects from misleading visual cues such as ripples, shadows, or natural debris.

Table 2. Comparison with Baseline Approaches

Model	Accuracy	F1-Score	mAP
ResNet50 (CNN-only)	86.3%	85.8%	0.62
YOLOv5	88.1%	87.4%	0.69
Proposed CNN–GNN	92.4%	91.9%	0.78

3.3 Ablation Study: Importance of Graph Reasoning

An ablation study was conducted to analyze the individual contributions of CNN features, static graph structures, and graph attention mechanisms. As shown in Table 3, the CNN-only configuration provided the lowest performance due to its inability to incorporate spatial relationships. When a static graph was included, the model performance improved, but the increase was limited because all neighboring regions were treated equally without adaptive weighting. The proposed CNN–GAT configuration achieved the highest metrics across all categories, demonstrating the significant role of graph attention in refining node interactions. The attention mechanism helped highlight relevant contextual relationships and suppress irrelevant or noisy patterns, particularly in challenging river scenes where object boundaries were blurred or partially submerged.

Table 3. Ablation Study Results

Configuration	Accuracy	F1-Score	mAP
CNN-only	86.3%	85.8%	0.62
CNN + Static Graph	89.5%	88.2%	0.71
CNN + GAT (Proposed)	92.4%	91.9%	0.78

3.4 Detailed Qualitative Analysis

A qualitative examination of several model outputs further illustrates the robustness of the proposed approach. In scenes with strong sunlight reflections, the model successfully differentiated reflective segments from actual waste by analyzing neighboring regions through graph propagation. When waste objects were partially submerged, the CNN backbone alone often misinterpreted the distorted shapes as background noise, whereas the GNN component leveraged contextual similarities to identify the correct object. The model also demonstrated an ability to distinguish synthetic waste from natural debris such as leaves or twigs, which frequently appear in Angke river.

Figure 2. Qualitative Detection Results

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5544

Additionally, in scenes with high water turbulence, the model maintained stable detection performance by relying on relational embeddings that preserved structural consistency across patches. These qualitative observations show that contextual information significantly improves detection accuracy beyond what pixel-level examination can provide.

3.5 Model Robustness in Different Environmental Conditions

To gain deeper insights into the model's reliability, the test dataset was separated into multiple environmental categories. The results summarized in Table 4 reveal that the proposed model maintained strong F1-scores across all conditions. Performance was highest under bright daylight, where visual clarity was optimal. Cloudy weather introduced mild variations in brightness but did not significantly affect detection. Backlit scenes and low-light conditions posed more substantial challenges due to reduced contrast, yet the model retained reasonably high accuracy. The lowest scores were recorded during strong water turbulence, where rapid motion created complex visual textures. However, even in such difficult scenarios, the CNN–GNN hybrid remained more stable than the baseline models, demonstrating its resilience under dynamic environmental conditions.

Table 4. Performance Under Different Environmental Conditions

Scenario	F1-Score
Bright daylight	93.5%
Cloudy conditions	92.1%
Backlit scenes	89.4%
Low-light / evening	88.7%
Heavy water turbulence	87.9%

3.6 Error Pattern Analysis

An analysis of misclassified samples revealed several consistent error patterns. False negatives primarily occurred when waste objects were extremely small or partially obscured by shadows, causing insufficient feature representation at the patch level. Some objects were also missed when their colors closely matched the water surface during overcast or muddy conditions. False positives were generally triggered by persistent water patterns that formed shapes resembling floating debris, as well as by clusters of natural leaves that mimicked the visual characteristics of synthetic waste. These findings indicate that although the model performs well overall, additional temporal data or multi-frame reasoning could further reduce error rates by incorporating motion cues that distinguish static objects from dynamic water patterns.

3.7 Summary of Findings

Overall, the results demonstrate that the proposed CNN–GNN model provides significant improvements over standard CNN and YOLO-based detectors. The integration of spatial–contextual information proved crucial for achieving high detection accuracy in the challenging visual conditions present in Angke river. The model consistently performed well across diverse environmental scenarios and showed strong capability in distinguishing between synthetic waste and natural debris. Ablation experiments confirmed that graph attention contributed substantially to the final performance by enabling the model to focus on meaningful relationships between spatial regions. While some errors remain in extreme cases, the findings suggest that the proposed architecture is well-suited for real-time river waste monitoring applications and holds promise for integration into environmental surveillance systems.

Vol. 7, No. 3, June 2026, Page. 2577-2587
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5544

4. DISCUSSIONS

The results demonstrate that the integration of graph-based relational reasoning with convolutional feature extraction significantly enhances the model's capability to detect floating waste in complex river environments such as Angke river. Unlike traditional CNN architectures that primarily rely on localized pixel information, the proposed model leverages spatial–contextual relationships among neighboring regions to obtain a more holistic understanding of the scene. This characteristic is particularly important in river monitoring because waste objects often appear distorted, partially submerged, or visually blended with the water surface. The improved performance metrics, especially the high recall and F1-score, confirm that the model is able to identify waste objects even in challenging visual scenarios where standard pixel-based representations would typically fail.

The comparison with baseline models further highlights the value of graph reasoning. The CNN-only baseline struggles to distinguish between floating debris and visually similar patterns created by water reflections or turbulence. In contrast, the proposed model benefits from the graph attention mechanism, which selectively emphasizes relevant node interactions while suppressing noise from less informative spatial patterns. This capability allows the model to remain robust under dynamic environmental conditions, such as uneven illumination, fluctuating water textures, and varying shadow structures. The distinct improvement observed when transitioning from a static graph to a full GAT architecture also shows that adaptive attention weighting is essential for maximizing performance, as it enables the model to focus on contextual cues that contribute meaningful information toward classification.

The qualitative findings further support the quantitative results by showing how the model correctly identifies waste objects in several difficult cases. For example, the ability to detect partially submerged objects is a direct consequence of the model's capacity to aggregate information across adjacent patches, compensating for local distortions. Similarly, the model's ability to differentiate synthetic waste from natural debris demonstrates the strength of relational embeddings in capturing subtle texture and shape differences. These insights help explain why the CNN–GNN architecture outperforms YOLOv5, which, like other bounding-box detectors, depends heavily on distinct object boundaries that may not be clearly visible in natural river scenes.

Beyond its application to environmental monitoring, this study contributes to the field of Informatics and Computer Science by demonstrating the effectiveness of graph-based relational learning for complex visual understanding tasks. The proposed hybrid CNN–GNN architecture extends conventional computer vision pipelines by incorporating explicit spatial–contextual reasoning, enabling the model to capture inter-region dependencies that are difficult to learn using purely convolutional operations. From a computer science perspective, this work highlights the importance of structured data representations and attention-based graph modeling in improving robustness under dynamic and visually ambiguous conditions. The proposed framework can be generalized to other intelligent surveillance and monitoring applications, such as traffic analysis, anomaly detection, and smart city infrastructures, where contextual relationships play a critical role in accurate decision-making.

Despite these strengths, the model still exhibits limitations in extreme conditions. Small waste objects, particularly those covering only a few pixels, are occasionally missed because their local patch representations may not contain sufficient information for the graph to propagate effectively. False positives also occur in situations where repetitive water motion creates patterns that resemble debris, suggesting that additional temporal information may be needed to differentiate between static and dynamic features. These issues point to potential avenues for future development, such as incorporating temporal graph networks, optical flow analysis, or multi-frame reasoning to further increase robustness.

Overall, the discussion reveals that the hybrid CNN–GNN architecture provides a significant advantage over traditional deep learning approaches for river waste detection. By integrating spatial–

Vol. 7, No. 3, June 2026, Page. 2577-2587
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5544

contextual information into the decision-making process, the model achieves consistent improvements across diverse environmental conditions. This makes it a strong candidate for operational deployment in real-world monitoring systems and highlights its potential to assist environmental management agencies in automating waste detection along urban rivers such as Angke river. From a deployment perspective, several practical considerations should be addressed when integrating the proposed CNN–GNN framework into real-world surveillance systems. While the model demonstrates strong detection performance, graph-based reasoning introduces additional computational overhead compared to conventional CNN-only architectures, which may affect inference latency in real-time CCTV deployments. Scalability also becomes a critical factor when the system is required to process data streams from multiple cameras simultaneously in large-scale river monitoring networks. However, these challenges can be mitigated through optimization strategies such as model pruning, parallel processing, and edge–cloud hybrid deployment architectures. In terms of generalization, the proposed spatial–contextual learning framework is not limited to river waste detection and can be extended to other intelligent monitoring scenarios involving dynamic and cluttered visual environments, including traffic surveillance, public safety monitoring, and smart city applications.

5. CONCLUSION

This study presented a hybrid deep-learning approach that integrates convolutional feature extraction with graph-based relational reasoning to improve floating waste detection in CCTV imagery from Angke river. The proposed CNN–GNN architecture successfully demonstrated its ability to handle the complex visual characteristics of river environments, including reflections, shadows, water turbulence, and partially submerged objects. By modeling spatial–contextual relationships among image regions, the system achieved higher accuracy and robustness than conventional CNN-only and YOLO-based detectors. The findings indicate that graph attention mechanisms play a crucial role in distinguishing waste objects from background textures that frequently appear ambiguous in river-monitoring scenarios. Although some limitations remain in extreme visual conditions, the overall performance suggests that this method is suitable for practical deployment in automated environmental monitoring systems. Future work may enhance detection performance by incorporating temporal information, multi-frame analysis, or extended graph structures to further reduce false predictions and support real-time operational requirements. Future research can be directed toward several concrete directions to further enhance the proposed framework. First, incorporating temporal information through multi-frame analysis or temporal graph neural networks could improve detection accuracy for small or partially occluded waste objects by leveraging motion cues across consecutive frames. Second, optimizing the model for real-time deployment, such as through model compression, pruning, or lightweight graph architectures, would help reduce inference latency and support large-scale CCTV-based monitoring systems. Third, extending the evaluation to multi-river datasets from different geographic regions and environmental conditions would enable a more comprehensive assessment of the model's generalization capability and robustness across diverse real-world scenarios.

REFERENCES

[1] X.-M. Chen *et al.*, "Conflict between urbanization and water environmental protection: Lessons from the Xiangjiang River Basin in China.," *Water research*, vol. 252, hal. 121237, 2024, doi: 10.1016/j.watres.2024.121237.
[2] S. Syafri, B. Surya, R. Ridwan, S. Bahri, E. Rasyidi, dan S. Sudarman, "Water Quality Pollution Control and Watershed Management Based on Community Participation in Maros City, South Sulawesi, Indonesia," *Sustainability*, 2020, doi: 10.3390/su122410260.
[3] E. MacAfee dan A. Löhr, "Multi-scalar interactions between mismanaged plastic waste and urban flooding in an era of climate change and rapid urbanization," *Wiley Interdisciplinary*

Vol. 7, No. 3, June 2026, Page. 2577-2587
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5544

Reviews: Water, vol. 11, 2023, doi: 10.1002/wat2.1708.
[4] T. Gutsa, C. Trois, R. De Vries, dan T. Mani, "Wasted shores: Using drones to monitor the spatio-temporal evolution of debris accumulation hotspots on South Africa's Umgeni River.," *The Science of the total environment*, hal. 176791, 2024, doi: 10.1016/j.scitotenv.2024.176791.
[5] H. Jeon, H. Kim, D. Kim, dan J. Kim, "PASS-CCTV: Proactive Anomaly surveillance system for CCTV footage analysis in adverse environmental conditions," *Expert Syst. Appl.*, vol. 254, hal. 124391, 2024, doi: 10.1016/j.eswa.2024.124391.
[6] K. Muhammad, M. Obaidat, dan M. Tanveer, "Fuzzy Logic in Surveillance Big Video Data Analysis," *ACM Computing Surveys (CSUR)*, vol. 54, hal. 1–33, 2021, doi: 10.1145/3444693.
[7] B. M. Pati *et al.*, "Leveraging UAV Data and Deep Learning Models for Detecting Waste in Rivers," *IEEE Access*, vol. 13, hal. 99603–99627, 2025, doi: 10.1109/access.2025.3576295.
[8] Z. Su *et al.*, "Rapid Salient Object Detection with Difference Convolutional Neural Networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. PP, 2025, doi: 10.1109/tpami.2025.3583968.
[9] Z. Jiang, B. Wu, Long, H. Zhang, dan J. Lian, "APM-YOLOv7 for Small-Target Water-Floating Garbage Detection Based on Multi-Scale Feature Adaptive Weighted Fusion," *Sensors (Basel, Switzerland)*, vol. 24, 2023, doi: 10.3390/s24010050.
[10] A. Surahmat dan D. B. R. W. Yato, "AI-Based Waste Detection for Water Quality Monitoring in the Cisadane River: A Deep Learning Approach," *Gema Lingkungan Kesehatan*, vol. 23, no. 3, hal. 333–343, 2025, doi: 10.36568/gelinkes.v23i3.270.
[11] A. Bera, Z. Wharton, Y. Liu, N. Bessis, dan A. Behera, "SR-GNN: Spatial Relation-Aware Graph Neural Network for Fine-Grained Image Categorization," *IEEE Transactions on Image Processing*, vol. 31, hal. 6017–6031, 2022, doi: 10.1109/tip.2022.3205215.
[12] M. Nunkhaw dan H. Miyamoto, "An Image Analysis of River-Floating Waste Materials by Using Deep Learning Techniques," *Water*, 2024, doi: 10.3390/w16101373.
[13] T. Sung, Y. Kang, dan J. Im, "Enhancing Satellite-Based Wildfire Monitoring: Advanced Contextual Model Using Environmental and Structural Information," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, hal. 1–16, 2024, doi: 10.1109/tgrs.2024.3418475.
[14] K. Buchhorn, E. Santos-Fernandez, K. Mengersen, dan R. Salomone, "Graph neural network-based anomaly detection for river network systems," *F1000Research*, vol. 12, 2023, doi: 10.12688/f1000research.136097.2.
[15] S. Xia *et al.*, "An Adaptive Multi-Granularity Graph Representation of Image via Granular-ball Computing," *IEEE Transactions on Image Processing*, vol. 34, hal. 2986–2999, 2025, doi: 10.1109/tip.2025.3565212.
[16] Z. Xue, Z. Liu, Z. Xue, dan T. Song, "Spatial-Spectral Contrastive Graph Neural Network for Few-Shot Hyperspectral Image Classification," *IEEE Access*, vol. 13, hal. 88278–88290, 2025, doi: 10.1109/access.2025.3569874.
[17] A. A. Yusuf, F. Chong, dan M. Xianling, "An analysis of graph convolutional networks and recent datasets for visual question answering," *Artificial Intelligence Review*, vol. 55, hal. 6277–6300, 2022, doi: 10.1007/s10462-022-10151-2.
[18] S. Wei, H. Tang, P. Jiao, dan H. Wu, "GGAT: Gravitation-Based Graph Attention Networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 11, hal. 630–640, 2025, doi: 10.1109/tsipn.2025.3583355.
[19] Z. Cui, L. Chen, Y. Wang, D. Haehn, Y. Wang, dan H. Pfister, "Generalization of CNNs on Relational Reasoning With Bar Charts," *IEEE Transactions on Visualization and Computer Graphics*, vol. 31, hal. 5611–5625, 2024, doi: 10.1109/tvcg.2024.3463800.
[20] W. Wang *et al.*, "Drainage pattern recognition method considering local basin shape based on graph neural network," *International Journal of Digital Earth*, vol. 16, hal. 593–619, 2023, doi: 10.1080/17538947.2023.2172224.
[21] D. Wang dan H. Qian, "Graph neural network method for the intelligent selection of river system," *Geocarto International*, vol. 38, 2023, doi: 10.1080/10106049.2023.2252762.
[22] C. Sun *et al.*, "Attention-based graph neural networks: a survey," *Artificial Intelligence Review*, vol. 56, hal. 2263–2310, 2023, doi: 10.1007/s10462-023-10577-2.
[23] J. Su, R. Chen, M. Li, S. Liu, G. Xu, dan Z. Zheng, "EcoDetect-YOLOv2: A High-Performance

Vol. 7, No. 3, June 2026, Page. 2577-2587
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5544

Model for Multi-Scale Waste Detection in Complex Surveillance Environments," *Sensors (Basel, Switzerland)*, vol. 25, 2025, doi: 10.3390/s25113451.

[24] H.-S. Lee, S. Byeon, J. H. Kim, J.-K. Shin, dan Y. Park, "Construction of a Real-Time Detection for Floating Plastics in a Stream Using Video Cameras and Deep Learning," *Sensors (Basel, Switzerland)*, vol. 25, 2025, doi: 10.3390/s25072225.

[25] X. Song, Z. Hua, dan J. Li, "Context Spatial Awareness Remote Sensing Image Change Detection Network Based on Graph and Convolution Interaction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, hal. 1–16, 2024, doi: 10.1109/tgrs.2024.3357524.

[26] J. Farfán, K. Palacios, J. Ulloa, dan A. Avilés, "A hybrid neural network-based technique to improve the flow forecasting of physical and data-driven models: Methodology and case studies in Andean watersheds," *Journal of Hydrology: Regional Studies*, vol. 27, hal. 100652, 2020, doi: 10.1016/j.ejrh.2019.100652.

[27] A. T. Marye *et al.*, "Remote Sensing for Monitoring Macroplastics in Rivers: A Review," *Wiley Interdisciplinary Reviews: Water*, vol. 12, 2025, doi: 10.1002/wat2.70020.

[28] G. Collins *et al.*, "Evaluation of clinical prediction models (part 1): from development to external validation," *The BMJ*, vol. 384, 2024, doi: 10.1136/bmj-2023-074819.