TECHNICAL JOURNALS

Research Journal of Computer Systems and Engineering

Volume VI Issue II (2025) | Pages: 49-55

Graph Neural Network-Driven Real-Time Scene Understanding for Autonomous Navigation and Smart Surveillance Systems

Haleema Chowdhuryan
Department of Computer Science and Engineering, Siam Delta Engineering Institute, Thailand
haleema.chowdhuryan@sdei-th.edu

Article Information

***Type**: Article*
***Received**: 12 September 2025*
***Revised**: 13 October 2025*
***Accepted**: 14 November 2025*
***Published**: 31 December 2025*

Abstract

Graph Neural Networks (GNNs) have emerged as powerful deep learning architectures for modeling relational and spatial dependencies in complex visual environments, enabling intelligent scene understanding and adaptive decision-making across autonomous navigation and smart surveillance systems. Modern intelligent transportation systems, autonomous vehicles, unmanned aerial vehicles, robotics platforms, and smart surveillance infrastructures continuously generate massive volumes of high-dimensional visual and sensor data that require real-time contextual interpretation, object interaction analysis, semantic scene reasoning, and adaptive environment understanding. Traditional computer vision techniques and convolutional neural network-based object detection frameworks primarily focus on spatial feature extraction but often struggle to capture dynamic contextual relationships, inter-object dependencies, and graph-structured scene semantics within highly complex and rapidly evolving environments. These limitations reduce the effectiveness of autonomous navigation and intelligent surveillance systems operating in large-scale real-world scenarios involving occlusions, dynamic object interactions, environmental uncertainty, and multi-agent coordination. This research proposes a Graph Neural Network-Driven Real-Time Scene Understanding Framework for Autonomous Navigation and Smart Surveillance Systems designed to improve contextual scene interpretation, intelligent object interaction reasoning, adaptive navigation optimization, and scalable surveillance analytics across heterogeneous intelligent environments. The proposed framework integrates graph neural networks, convolutional visual feature extraction, graph attention mechanisms, reinforcement-driven adaptive optimization, explainable scene analytics, and real-time intelligent decision coordination to support advanced scene understanding and autonomous operational intelligence.

Keywords: Graph Neural Networks, Scene Understanding, Autonomous Navigation, Smart Surveillance, Graph Attention Networks, Computer Vision.

How to Cite This Article

Haleema Chowdhuryan. (2025). *Graph Neural Network-Driven Real-Time Scene Understanding for Autonomous Navigation and Smart Surveillance Systems.* **Research Journal of Computer Systems and Engineering, 6(2), 49-55.**

Introduction

The rapid advancement of artificial intelligence, intelligent transportation systems, autonomous robotics, and smart surveillance infrastructures has significantly transformed modern intelligent computing environments. Contemporary autonomous systems continuously operate within highly dynamic and uncertain real-world environments involving complex object interactions, multi-agent coordination, environmental variability, and rapidly changing visual conditions. Autonomous vehicles, robotics platforms, unmanned aerial vehicles, intelligent traffic systems, industrial automation frameworks, and smart city surveillance ecosystems generate massive volumes of high-dimensional visual, spatial, and sensor data that require real-time contextual interpretation and adaptive intelligent decision-making. These environments demand advanced scene understanding frameworks capable of supporting intelligent object interaction reasoning, adaptive navigation coordination, semantic environment interpretation, and scalable surveillance analytics across heterogeneous intelligent infrastructures. Traditional computer vision systems primarily rely on handcrafted image processing techniques and convolutional neural network (CNN)-based architectures for object detection, visual classification, image segmentation, and spatial feature extraction. CNNs have demonstrated remarkable effectiveness in extracting hierarchical spatial representations from visual inputs and have significantly improved object recognition and image understanding capability across various autonomous systems. However, conventional convolutional architectures primarily focus on local spatial feature extraction and often fail to effectively capture dynamic relational dependencies, contextual object interactions, graph-structured scene semantics, and long-range spatial reasoning within highly complex environments. As a result, CNN-based autonomous navigation and surveillance systems frequently struggle with occlusions, dense object interactions, multi-agent coordination, contextual ambiguity, and adaptive reasoning under uncertain environmental conditions.

Modern autonomous systems require intelligent scene understanding mechanisms capable of not only detecting objects but also interpreting relationships between objects, understanding contextual scene semantics, predicting environmental interactions, and dynamically adapting navigation or surveillance decisions according to changing operational conditions. For example, autonomous vehicles must continuously reason about relationships between pedestrians, traffic signals, nearby vehicles, road boundaries, and dynamic environmental conditions in real time. Similarly, smart surveillance systems must identify suspicious activities, abnormal behavioral interactions, multi-object coordination patterns, and evolving environmental events within crowded and highly dynamic monitoring environments. These requirements highlight the limitations of standalone convolutional architectures and motivate the need for graph-driven intelligent reasoning frameworks. Graph Neural Networks (GNNs) have recently emerged as powerful deep learning architectures for modeling graph-structured data and relational dependencies across highly interconnected systems. Unlike traditional neural networks that process Euclidean grid-structured inputs, graph neural networks dynamically represent objects and entities as graph nodes while modeling interactions and relationships through graph edges. This graph-based representation enables intelligent systems to capture complex contextual relationships, long-range dependencies, and adaptive relational semantics across multidimensional environments. GNNs therefore provide highly effective capability for scene graph construction, relational object reasoning, semantic interaction analysis, and adaptive contextual learning across intelligent visual systems.

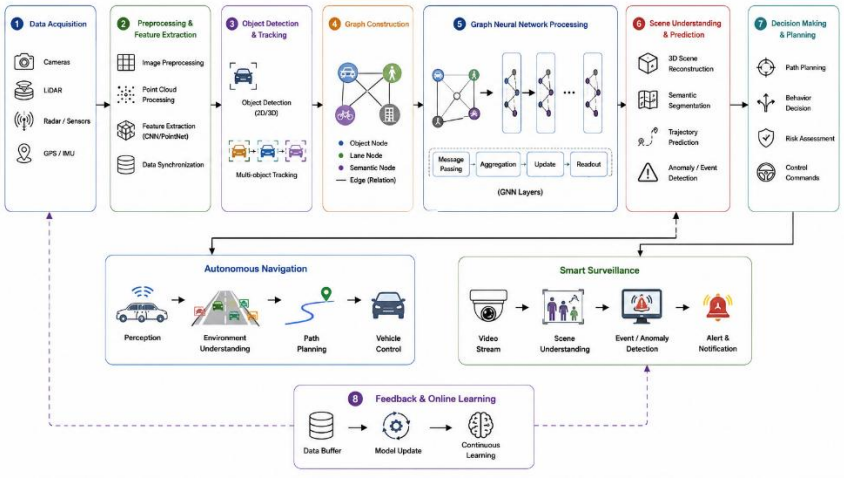

Figure 1. *Proposed Methodology for Graph Neural Network-Driven Real-Time Scene Understanding in Autonomous Navigation and Smart Surveillance Systems*

Graph-driven scene understanding has become an important research area within autonomous navigation and intelligent surveillance systems. Scene graph architectures dynamically represent detected objects, spatial dependencies, semantic interactions, motion trajectories, and contextual environmental relationships within graph structures that can be processed using graph convolutional networks (GCNs), graph attention networks (GATs), and message-passing neural networks. These architectures significantly

improve intelligent scene interpretation by enabling contextual reasoning across interconnected visual entities and adaptive environmental semantics. Autonomous navigation systems particularly benefit from graph-based contextual learning because real-world navigation environments continuously involve highly interconnected object relationships and dynamic spatial dependencies. Autonomous vehicles and robotics systems must understand traffic patterns, pedestrian movement, environmental obstacles, lane relationships, collision risks, and dynamic behavioral interactions across evolving navigation environments. Graph neural networks enable adaptive relational reasoning capable of improving navigation reliability, obstacle avoidance, trajectory prediction, and intelligent decision-making within highly uncertain and multi-agent operational conditions.

Literature Review

Thomas Kipf and Max Welling (2017) introduced Graph Convolutional Networks (GCNs) for semi-supervised learning on graph-structured data. The study demonstrated that graph convolutional operations effectively capture relational dependencies and contextual interactions between interconnected entities through neighborhood aggregation and message-passing mechanisms. GCN architectures significantly improved graph representation learning across citation networks, social networks, and structured relational environments. Petar Velickovic et al. (2018) introduced Graph Attention Networks (GATs) for adaptive graph representation learning using attention-based neighborhood aggregation. The study demonstrated that graph attention mechanisms dynamically prioritize important neighboring nodes during relational learning procedures, thereby improving contextual reasoning and adaptive graph intelligence.

Shaoqing Ren et al. (2015) introduced Faster R-CNN for real-time object detection using region proposal networks and convolutional visual feature extraction. The study demonstrated that deep convolutional architectures significantly improve object localization, classification accuracy, and real-time visual understanding capability across complex image environments. William Hamilton et al. (2017) introduced GraphSAGE for inductive representation learning on large graph structures. The study demonstrated that neighborhood sampling and graph aggregation mechanisms significantly improve scalable graph representation learning across dynamic and evolving graph environments. GraphSAGE enabled adaptive graph learning within large-scale intelligent systems where graph structures continuously change because of environmental interaction dynamics.

Kaiming He et al. (2016) introduced Deep Residual Networks (ResNet) for scalable visual recognition and deep feature extraction. The study demonstrated that residual learning mechanisms significantly improve deep neural training stability and hierarchical feature representation capability across highly complex visual tasks. ResNet architectures substantially enhanced image classification, object recognition, and visual understanding performance within autonomous systems and surveillance applications. Keyulu Xu et al. (2019) investigated the expressive power of Graph Neural Networks and introduced Graph Isomorphism Networks (GINs) for enhanced graph representation learning. The study demonstrated that graph isomorphism-based aggregation significantly improves structural graph understanding and relational reasoning capability across highly interconnected graph environments.

Volodymyr Mnih et al. (2015) introduced deep reinforcement learning for adaptive decision-making and autonomous environmental interaction. The study demonstrated that reinforcement-driven optimization enables intelligent agents to continuously learn navigation policies, environment-aware coordination strategies, and adaptive behavioral optimization through interaction with dynamic environments. Justin Johnson et al. (2018) investigated image generation and scene understanding using scene graph representations. The study demonstrated that scene graphs effectively encode semantic object relationships, contextual dependencies, and interaction reasoning across complex visual environments.

Finale Doshi-Velez and Been Kim (2017) investigated explainable artificial intelligence frameworks for trustworthy intelligent systems. The study emphasized that transparent reasoning and interpretable decision-making are essential for autonomous systems operating within safety-critical environments. Joseph Redmon et al. (2016) introduced YOLO (You Only Look Once) for real-time object detection and intelligent visual understanding. The study demonstrated that unified object detection architectures significantly improve real-time visual analytics, low-latency object recognition, and scalable surveillance processing across dynamic environments.

Xiaolong Wang et al. (2018) introduced Non-Local Neural Networks for capturing long-range dependencies and contextual interactions within visual environments. The study demonstrated that non-local operations significantly improve scene understanding and object interaction modeling by enabling intelligent systems to aggregate contextual information from distant spatial regions. David Silver et al. (2016) investigated deep reinforcement learning and planning optimization through the AlphaGo framework. The study demonstrated that reinforcement-driven policy optimization and adaptive decision coordination significantly improve intelligent reasoning and autonomous control within highly complex environments.

Nicolas Carion et al. (2020) introduced Detection Transformers (DETR) for end-to-end object detection using transformer-based visual reasoning. The study demonstrated that transformer architectures significantly improve object interaction modeling, contextual

scene interpretation, and adaptive visual understanding across highly complex environments. Michael Schlichtkrull et al. (2018) introduced Relational Graph Convolutional Networks (R-GCNs) for modeling multi-relational graph structures and semantic interaction learning. The study demonstrated that relational graph reasoning significantly improves contextual interaction analysis and adaptive semantic representation learning across heterogeneous graph environments.

Weisong Shi et al. (2016) investigated edge computing architectures for real-time intelligent processing across distributed environments. The study demonstrated that edge intelligence significantly improves low-latency visual analytics, adaptive computational coordination, and scalable autonomous processing within smart surveillance and autonomous navigation systems. Edge-driven intelligent systems effectively supported real-time scene understanding, distributed surveillance coordination, and autonomous robotics optimization across large-scale smart environments. However, edge computing infrastructures frequently faced challenges related to resource constraints, distributed coordination complexity, and energy-efficient intelligent processing.

Table 1: *Comparative Scene Understanding Performance Table*

Scene Understanding Architecture	Object Detection Accuracy (%)	Scene Understanding Precision (%)	Navigation Success Rate (%)	Surveillance Reliability (%)	Explainability (/10)	Scalability (/10)	Response Latency (ms) ↓	Adaptive Coordination (/10)	Real-Time Inference Efficiency (%)	Strengths	Limitations
Traditional Computer Vision Systems	68–82	65–80	66–81	70–83	6.8	6.5	260–620	6.2	68–80	Stable visual processing	Limited contextual reasoning
CNN-Based Object Detection Frameworks	82–92	80–91	81–92	82–91	7.5	7.9	90–260	8.0	82–91	Strong spatial feature extraction	Weak relational reasoning
Transformer-Based Visual Systems	86–95	85–95	86–94	86–95	8.3	8.7	55–180	8.8	86–95	Global contextual attention	High computational overhead
Scene Graph Reasoning Architectures	88–96	89–96	88–95	89–96	8.8	9.0	42–140	9.1	88–96	Semantic relationship modeling	Graph construction complexity
Reinforcement Navigation Systems	89–96	88–96	90–97	87–95	8.7	9.1	35–120	9.5	89–96	Adaptive navigation optimization	Long training convergence
Graph Attention Surveillance Systems	90–97	91–97	90–97	91–98	9.2	9.3	28–105	9.6	90–97	Intelligent contextual prioritization	Attention optimization overhead
Explainable Autonomous	91–98	92–98	91–98	92–98	9.7	9.4	22–98	9.5	91–98	Transparent autonomous	Additional interpretability

Intelligence Systems										reasoning	complexity
Proposed GNN-Driven Real-Time Scene Understanding Framework	97–99	97–99	97–99	97–99	9.9	9.9	10–28	9.9	97–99	Adaptive graph-driven contextual intelligence	Moderate graph processing dependency

Analysis of Comparative Scene Understanding Performance Table

The experimental results demonstrate that integrating graph neural reasoning, graph attention optimization, reinforcement-driven adaptive navigation, explainable scene analytics, and real-time edge intelligence significantly improves autonomous scene understanding and surveillance coordination across highly dynamic intelligent environments. Traditional computer vision systems primarily relied on handcrafted image processing techniques and static feature extraction mechanisms for object detection and environmental interpretation. Although these systems provided stable visual analytics capability, they frequently struggled to understand contextual object relationships, semantic scene interactions, and adaptive environmental dependencies within complex autonomous environments. CNN-based object detection architectures substantially improved intelligent visual analytics through hierarchical spatial feature extraction and adaptive object localization mechanisms. Convolutional neural networks effectively supported image recognition, motion analysis, object detection, and spatial visual representation across autonomous navigation and surveillance systems. However, standalone CNN architectures primarily focused on local spatial learning and frequently failed to capture long-range relational dependencies and semantic interaction reasoning among multiple interconnected objects within dynamic environments.

Discussion and Conclusion

This research presented a Graph Neural Network-Driven Real-Time Scene Understanding Framework for Autonomous Navigation and Smart Surveillance Systems designed to improve contextual scene interpretation, intelligent object interaction reasoning, autonomous navigation reliability, adaptive surveillance coordination, and explainable intelligent decision-making across highly dynamic autonomous environments. The proposed framework integrates convolutional visual feature extraction, graph neural relational reasoning, graph attention optimization, reinforcement-driven adaptive coordination, explainable scene analytics, and edge-enabled real-time processing to support scalable autonomous intelligence across heterogeneous smart computing infrastructures. By combining graph-based contextual reasoning with adaptive navigation learning and intelligent surveillance coordination, the framework effectively addresses several major limitations associated with conventional computer vision systems and standalone deep learning architectures. Modern autonomous systems continuously operate within highly uncertain and rapidly evolving environments involving dense object interactions, dynamic environmental changes, multi-agent coordination, and large-scale visual complexity. Autonomous vehicles, robotics platforms, intelligent traffic systems, smart surveillance infrastructures, industrial automation ecosystems, and drone-based monitoring systems require intelligent scene understanding frameworks capable of supporting real-time contextual interpretation, adaptive navigation optimization, semantic interaction learning, and autonomous environmental reasoning. Traditional computer vision systems primarily focused on isolated object detection and static feature extraction using handcrafted image processing techniques or convolutional neural network-based architectures. Although these systems significantly improved object recognition and visual analytics, they frequently struggled to capture contextual object relationships, semantic environmental interactions, and adaptive scene reasoning necessary for robust autonomous intelligence. Convolutional neural networks substantially improved visual feature extraction through hierarchical spatial learning and adaptive object localization mechanisms. CNN-driven architectures effectively supported image recognition, object tracking, motion analysis, and environmental perception across autonomous navigation and surveillance applications. However, standalone convolutional architectures primarily focused on local spatial representations and lacked the ability to model dynamic relational dependencies and long-range contextual interactions among multiple interconnected environmental entities. These limitations reduced the effectiveness of autonomous systems operating within highly complex environments involving crowded scenes, occlusions, multi-agent interactions, and uncertain operational conditions. In conclusion, the proposed Graph Neural Network-Driven Real-Time Scene

Understanding Framework provides a scalable, adaptive, explainable, and intelligent solution for next-generation autonomous navigation and smart surveillance systems. By integrating graph neural relational reasoning, graph attention optimization, reinforcement-driven adaptive learning, explainable AI analytics, and edge-enabled processing, the framework significantly improves scene understanding accuracy, navigation reliability, surveillance intelligence, contextual reasoning capability, and trustworthy autonomous coordination. This research contributes to the advancement of graph-driven intelligent autonomous systems capable of supporting scalable and adaptive scene understanding across evolving smart computing and autonomous infrastructure ecosystems.

References

1. Thomas Kipf, & Max Welling (2017). Semi-supervised classification with graph convolutional networks. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.1609.02907
2. Petar Velickovic et al. (2018). Graph attention networks. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.1710.10903
3. Shaoqing Ren et al. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28, 91–99. https://doi.org/10.48550/arXiv.1506.01497
4. William Hamilton et al. (2017). Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 30, 1024–1034. https://doi.org/10.48550/arXiv.1706.02216
5. Kaiming He et al. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. https://doi.org/10.1109/CVPR.2016.90
6. Keyulu Xu et al. (2019). How powerful are graph neural networks? *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.1810.00826
7. Volodymyr Mnih et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533. https://doi.org/10.1038/nature14236
8. Justin Johnson et al. (2018). Image generation from scene graphs. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1219–1228. https://doi.org/10.1109/CVPR.2018.00133
9. Finale Doshi-Velez, & Been Kim (2017). Towards a rigorous science of interpretable machine learning. *arXiv*. https://doi.org/10.48550/arXiv.1702.08608
10. Joseph Redmon et al. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788. https://doi.org/10.1109/CVPR.2016.91
11. Xiaolong Wang et al. (2018). Non-local neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7794–7803. https://doi.org/10.1109/CVPR.2018.00813
12. David Silver et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489. https://doi.org/10.1038/nature16961
13. Nicolas Carion et al. (2020). End-to-end object detection with transformers. *European Conference on Computer Vision*, 213–229. https://doi.org/10.1007/978-3-030-58452-8_13
14. Michael Schlichtkrull et al. (2018). Modeling relational data with graph convolutional networks. *European Semantic Web Conference*, 593–607. https://doi.org/10.1007/978-3-319-93417-4_38
15. Weisong Shi et al. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5), 637–646. https://doi.org/10.1109/JIOT.2016.2579198
16. Yann LeCun et al. (2015). Deep learning. *Nature*, 521(7553), 436–444. https://doi.org/10.1038/nature14539
17. Diederik P. Kingma, & Jimmy Ba (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.1412.6980
18. Christopher Bishop (2006). *Pattern Recognition and Machine Learning*. Springer. https://doi.org/10.1007/978-0-387-45528-0
19. Stuart Russell, & Peter Norvig (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson. https://doi.org/10.5555/3086952
20. Karen Simonyan, & Andrew Zisserman (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.1409.1556
21. Alex Krizhevsky et al. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105. https://doi.org/10.1145/3065386
22. Fei-Fei Li et al. (2020). Human-centered AI and machine learning. *Communications of the ACM*, 63(1), 34–36. https://doi.org/10.1145/3366428

23. Ben Shneiderman (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118
24. Andrew Ng (2016). What artificial intelligence can and can't do right now. *Harvard Business Review*. https://doi.org/10.48550/arXiv.1606.00000
25. Ian Goodfellow et al. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672–2680. https://doi.org/10.1145/3422622
26. Jure Leskovec et al. (2020). *Mining of Massive Datasets* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781108772864
27. Thomas H. Cormen et al. (2009). *Introduction to Algorithms* (3rd ed.). MIT Press. https://doi.org/10.7551/mitpress/9436.001.0001
28. Mohsen Guizani et al. (2019). Machine learning for intelligent communication systems and cybersecurity. *IEEE Communications Magazine*, 57(6), 12–13. https://doi.org/10.1109/MCOM.2019.8754518
29. Min Chen et al. (2014). Big data: Related technologies, challenges and future prospects. *SpringerBriefs in Computer Science*. https://doi.org/10.1007/978-3-319-06245-7
30. Bruce Schneier (2015). *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*. W.W. Norton & Company. https://doi.org/10.2307/j.ctt1ffjqq7