A Review of Intelligent Security Research: Progress in Integrated Edge-Cloud-Graph Approaches

Yunwei Wang
Shenzhen High School of Mathematics and Physics Shenzhen, Guangdong, China
**Corresponding Author*

Abstract: This paper provides a systematic review of the current development status and future trends in intelligent security systems. Against the backdrop of accelerating urbanization and growing public safety demands, traditional security systems struggle to meet requirements for real-time responsiveness, intelligence, and robustness. With the rapid advancement of artificial intelligence, edge computing, and the Internet of Things, security systems are progressively transitioning from "passive response" to "proactive prevention and control". This paper focuses on five key research areas: video surveillance and object detection, behavioral analysis and anomaly detection, facial recognition and identity verification, edge computing and resource optimization, and multimodal fusion with security knowledge graphs. Drawing on existing research, it identifies shortcomings in current methods regarding adaptability to complex environments, cross-modal semantic understanding, and edge-cloud collaborative scheduling [1-2]. To address these challenges, this paper proposes constructing an "integrated edge-cloud-graph" intelligent security framework, aiming to upgrade capabilities from mere "visibility" to "comprehension and rapid response". Finally, future research directions are explored, including lightweight edge deployment, unsupervised anomaly detection, multimodal semantic modeling, and dynamic resource scheduling.

Keywords: Intelligent Security; Edge Computing; Multimodal Fusion; Knowledge Graph; Anomaly Detection

1. Introduction

1.1 Background

As urban areas continue to expand and public safety incidents become increasingly frequent, traditional security systems exhibit significant shortcomings in rapidly identifying complex scenarios and responding promptly. China's 14th Five-Year Plan and the "Digital China Initiative" documents emphasize accelerating the digital and intelligent upgrading of public safety systems [2] to achieve a closed-loop system of "perception–understanding–response".

Technologically, artificial intelligence, the Internet of Things, and edge computing provide robust support for upgrading security systems.

1.2 Research Questions and Objectives

Most contemporary security systems still rely predominantly on centralized video storage and manual identification, suffering from issues such as bandwidth constraints, high latency, substantial energy consumption, and a lack of semantic understanding. Consequently, this study aims to review current research progress in intelligent security, analyze existing challenges, and propose a solution based on an integrated "edge-cloud-graph" architecture, thereby providing reference for future research and applications.

2. Literature Review

2.1 Video Surveillance and Object Detection

As the core component of intelligent security systems, video surveillance aims to achieve precise identification and tracking of pedestrians, vehicles, and suspicious objects. In recent years, deep learning models have demonstrated outstanding performance in this domain. [8]

Mainstream models include the YOLO series (YOLOv3, YOLOv5, YOLOv8), Faster R-CNN, and DETR (Detection Transformer). Among these, the YOLO series models are renowned for their end-to-end architecture and high-speed detection performance, making them suitable for surveillance tasks demanding high real-time requirements. Two-stage detection algorithms like Faster R-CNN excel in feature extraction and localization accuracy, frequently employed for static scene analysis; whereas the DETR series models, by incorporating the Transformer self-attention mechanism, transform object detection into a sequence matching problem, achieving true end-to-end detection. This approach demonstrates enhanced robustness and generalization capabilities in complex, multi-object environments.

In recent object detection research, the DETR family of algorithms has established itself as the new paradigm for end-to-end detection. Unlike traditional anchor-based detection methods, DETR employs self-attention mechanisms for global feature modeling, eliminating complex steps such as region candidate generation and non-maximum suppression (NMS). This simplifies the detection workflow while enhancing overall stability. However, the original DETR model suffers from slow convergence and reduced sensitivity to small objects, limiting its application in real-time monitoring scenarios.

To address these limitations, researchers have proposed several improvement strategies:

Deformable DETR: Incorporates a deformable attention module that aggregates features at only a few key points, significantly reducing computational complexity while improving detection accuracy for small objects and dense scenes;

Conditional DETR and Anchor-DETR: Combine anchor prior information with conditional attention mechanisms to enhance model convergence stability in complex backgrounds;

DETR3D and BEVFormer: Implement Transformer-based object detection in three-dimensional scenes, offering novel solutions for intelligent traffic systems and multi-camera video surveillance.

Experimental results on mainstream datasets (e.g., COCO, KITTI, VisDrone) demonstrate that Deformable DETR achieves approximately 4–6% higher small object detection accuracy (mAP-small) than the original DETR, alongside a 30–40% increase in inference speed. In high-resolution video surveillance tasks, its multi-scale feature aggregation capability significantly outperforms YOLOv5. [8]

Overall, the YOLO series exhibits notable advantages in speed and real-time performance, making it suitable for deployment on edge devices. Faster R-CNN and the DETR series demonstrate superior detection accuracy and generalization capabilities, with the latter particularly representing the transition trend from convolutional networks to Transformer architectures. Future research is advancing towards lightweight Transformer architectures, sparse attention mechanisms, and edge-cloud collaborative inference frameworks to better meet the performance demands of real-time security surveillance.

Research Progress: Scholars have substantially enhanced detection performance in complex environments by incorporating Transformer attention mechanisms, Feature Pyramid Networks (FPN), and cross-camera object tracking (ReID). For instance, addressing facial recognition challenges caused by mask occlusions, a fusion approach combining CNNs and Transformers has been proposed, achieving significant accuracy improvements over traditional CNN models.

Shortcomings: High-accuracy models often feature complex architectures and substantial computational demands, rendering them unsuitable for resource-constrained edge devices. Most research remains reliant on standardized datasets (e.g., COCO, LFW), exhibiting considerable divergence from real-world security environments. Furthermore, adaptability to specialized scenarios such as night-time surveillance, foggy/rainy conditions, and strong backlighting remains inadequate.

2.2 Behavior Analysis and Anomaly Detection

Behavior recognition and anomaly detection represent a critical transition from "post-event tracing" to "pre-event warning".

Methodological Advances: In recent years, models such as the skeleton-feature-based ST-GCN (Spatio-Temporal Graph Convolutional Network) and the frame-based TSM (Time Shift Module) have demonstrated promising performance in identifying abnormal behaviors like fighting, falling, tailgating, and running. Concurrently, unsupervised and self-supervised learning approaches have gained prominence, addressing the scarcity of abnormal behavior data through pseudo-label generation, domain adaptation transfer, and contrastive learning. References: [5-7]

Notable research: Scholars have proposed a weak object detection method combining machine vision with Gamma correction, demonstrating strong performance in low-light conditions; [8] Another study employs pseudo-supervision for detecting floating objects on water surfaces, offering insights for cross-scenario behavior detection. [6]

Challenges and Limitations: Current anomaly detection models predominantly rely on batch training, resulting in insufficient real-time performance; pseudo-label methods are prone to cumulative errors, leading to diminished detection accuracy; most research remains focused on static frames, lacking temporal dynamic modeling of actions, and thus struggles with handling continuous and highly complex behaviors.

2.3 Facial Recognition and Identity Verification

Facial recognition stands as one of the most widely deployed technologies within intelligent security systems, extensively utilized in access control, visitor management, and surveillance deployment.

Research Progress: During the pandemic, traditional methods suffered performance degradation due to mask occlusions. Research proposed a hybrid CNN + Transformer model to enhance robustness [4]. For long-range and complex background conditions, the combined YOLO and FaceNet framework demonstrated favorable performance [2]. Furthermore, infrared polarization and RGB fusion recognition methods extended application boundaries to night-time and low-light environments [1].

Practical Applications: Examples include facial recognition-based laboratory attendance systems, demonstrating the technology's potential for deployment on low-cost hardware [3].

Limitations: Most models exhibit high computational demands, hindering deployment on low-power devices such as access control terminals and cameras; lack cross-modal auxiliary mechanisms, resulting in insufficient robustness under varying weather and lighting conditions; certain datasets were collected under idealized environments, leading to weaker generalization capabilities.

2.4 Edge Computing and Resource Optimization

As surveillance scales expand, traditional "cloud-centric" processing faces challenges in latency and bandwidth. Edge computing emerges as a key solution to mitigate these constraints.

Research Progress: Existing studies have proposed a "edge-cloud collaboration" architecture, deploying lightweight models at the edge for preliminary detection. [2] Only feature vectors are transmitted to the cloud for further identification, thereby reducing network bandwidth consumption and latency. Lightweight models such as MobileNet and EfficientNet have become mainstream choices and have been validated on low-power hardware platforms like Jetson Nano and Raspberry Pi.

Typical Case: In edge-cloud collaborative facial recognition, the edge performs feature extraction while the cloud handles high-precision matching, significantly enhancing system responsiveness.

Limitations: Lightweight models exhibit significant accuracy degradation in scenarios involving occlusion or low-resolution images. Most research lacks dynamic scheduling and resource-aware mechanisms, preventing flexible allocation of computational power based on task priority in complex environments. Furthermore, edge inference predominantly remains confined to the "detection" level, with limited exploration of integration with higher-level semantic understanding.

2.5 Multimodal Fusion and Security Knowledge Graphs

Single-modal information is highly susceptible to environmental interference, making multimodal fusion and knowledge modeling crucial trends for enhancing system intelligence.

Research Progress: Investigators have attempted to fuse multi-modal images such as RGB, infrared, and polarized data to enhance recognition performance in complex environments. For instance, combining infrared and visible light imagery compensates for recognition deficiencies in low-illumination night-time scenarios. In image translation research, methods based on spatial correlation and latent encoding spaces have been proposed, offering novel approaches for multi-domain knowledge transfer.

Knowledge Graph Applications: Within security scenarios, constructing semantic knowledge graphs based on the "person–object–event–time–location" quintuple enables event-level causal reasoning and risk assessment. Graph neural networks (GNNs) integrated with rule-based inference engines facilitate systems' transition from mere visual perception to genuine comprehension.

Limitations: Existing research predominantly focuses on image-level fusion, lacking structured semantic alignment mechanisms; no comprehensive system-level multimodal fusion architecture has yet been established; causal reasoning capabilities remain limited, hindering dynamic risk decision-making in complex scenarios.

3. Research Methodology

Addressing the shortcomings of existing intelligent security systems in real-time performance, accuracy, and adaptability, this paper proposes an "edge-cloud-graph integrated" research framework. This framework aims to organically combine edge computing, cloud fusion, and knowledge graph reasoning to form a closed-loop system spanning data acquisition, real-time recognition, and semantic inference. The specific research methodology can be divided into the following four aspects:

3.1 Design of Edge-Optimized Detection Models

Deploying lightweight models at the front end of security systems effectively reduces latency and power consumption. To achieve this objective, the research primarily employs the following technical approaches:

Model Structure Optimization: Building upon lightweight architectures such as MobileNetV3 and ShuffleNet, inference speed is enhanced through structural pruning and separable convolutions.

Knowledge Distillation: Utilizing high-precision large models in the cloud to distill knowledge into edge-based compact models, enabling near-equivalent recognition performance while maintaining low computational overhead.

Hardware Adaptation: Selecting edge computing platforms such as NVIDIA Jetson Nano and Raspberry Pi to evaluate model latency, power consumption, and accuracy, ensuring feasibility in real-world security scenarios.

3.2 Cloud-Based Multimodal Fusion Recognition System

Limited computational capabilities at the edge preclude complex recognition tasks, necessitating reliance on the cloud for higher-level fusion and analysis.

Multimodal Feature Fusion: Unify modeling of RGB images, infrared images, and behavioral trajectory data uploaded from the edge. Employ Transformer or gated fusion architectures to enhance feature complementarity.

Cross-Camera Target Consistency Tracking: Enables pedestrian re-identification (ReID) across multiple cameras, maintaining consistent target identity under varying perspectives and environmental conditions.

Complex Event Modeling: Leveraging cloud-based high-computational resources to conduct in-depth analysis of group behavior and anomalous events, enhancing the system's adaptability in dynamic environments.

3.3 Security Knowledge Graph Construction and Inference Mechanism

Traditional security systems often remain confined to the detection layer, lacking semantic understanding and reasoning capabilities for events. To address this, this paper proposes the introduction of a knowledge graph:

Pentadic Component Modeling: Abstracting "person–object–event–time–location" into nodes and edges within a knowledge graph to form a semantic relationship network.

Graph Neural Network (GNN) Inference: Integrating deep learning with symbolic reasoning to model causal relationships between events and assess risk levels.

Dynamic Feedback Mechanism: Feeds knowledge graph inference results back to the edge in real time, enabling policy optimization and resource allocation adjustments to build the system's adaptive capabilities.

3.4 Edge–Cloud–Graph Collaborative Scheduling Architecture

The efficiency of security systems depends not only on the accuracy of recognition models but also on effective collaboration between modules.

Event-Driven Mechanism: An asynchronous communication mechanism based on MQTT or gRPC is designed to enable rapid information exchange between edge, cloud, and graph modules.

Task Priority Scheduling: Dynamically adjust resource allocation based on contextual scenarios and graph inference outcomes, such as prioritizing computational resources for high-risk events.

Simulation and Field Testing Optimization: Evaluate the collaborative mechanism's performance in latency, power consumption, and accuracy through comparative experiments and real-world deployment tests, progressively refining system capabilities.

4. Conclusions

Intelligent security, as a vital component of smart cities and public safety, is undergoing a transformation from traditional surveillance towards intelligent and proactive systems. Through systematic review and analysis of existing research, the following conclusions can be drawn:

4.1 Research Status Summary

Video surveillance and object detection have achieved significant advances in accuracy, yet suffer from insufficient generalization capabilities in complex environments;

Behavioral analysis and anomaly detection are progressively evolving from supervised to unsupervised and self-supervised methods, though real-time performance remains suboptimal;

Facial recognition performance has improved under occlusion and low-light conditions, though lightweight solutions and cross-modal fusion require further enhancement;

Edge computing alleviates cloud-based pressures, though collaborative scheduling and resource optimization mechanisms remain underdeveloped;

Multimodal fusion and knowledge graph research are emerging, demonstrating potential for further expansion into system-level applications.

4.2 Future Development Trends

Lightweighting and Edge Deployment: Achieving low-power, high-efficiency front-end detection to enable real-time responsiveness in security systems.

Unsupervised and Adaptive Learning: Enhancing model robustness against unknown anomalies and cross-domain environments, overcoming the limitation of "reliant on labeled data for recognition".

Multimodal semantic fusion: Integrating visual, audio, sensor and other multi-source data to enhance comprehension and judgment capabilities for complex events.

Knowledge Reasoning and Causal Analysis: Utilizing knowledge graphs and reasoning mechanisms to endow security systems with semantic comprehension and decision-making capabilities, achieving a leap from "perception" to "cognition".

Edge-Cloud-Graph Collaboration Mechanism: Establishes a dynamic resource scheduling and task prioritization management system, ensuring efficient system operation under large-scale deployment.

4.3 Application Value

The integrated edge-cloud-graph architecture holds not only theoretical research value but also broad practical application prospects. It offers reference solutions for scenarios such as smart cities, campus security, and transport hub management, elevating public safety standards and driving the intelligent upgrading of the security industry.

In summary, the trajectory of intelligent security has shifted from isolated recognition to system-level integration and intelligence. Future research should focus on deepening cross-modal fusion, knowledge reasoning, and collaborative scheduling mechanisms, striving to achieve security objectives characterized by "perceptual awareness, interpretative understanding, and rapid response".

References

[1] Wang, F., Jin, X., Zhu, D., Hu, Z., Tang, S., & Lei, J. (2022). Infrared polarised face recognition based on RGB colour space. Laser & Optoelectronics Progress, 59(12), 1233001.

[2] Wei, Q., Li, Y., Lou, P., Yan, J., & Hu, J. (2021). Face recognition method based on edge-cloud collaboration. Computer Science, 49(5), 27–47.

[3] Zhu, Y., Lai, Z., Li, F., & Huang, Y. (2021). Laboratory attendance system based on face recognition. Journal of Jiangxi University of Science and Technology, 42(3), 115–120.

[4] Jiang, S., Yao, K., Chen, L., Wang, Z., & Guo, F. (2023). Masked face recognition method based on CNN–Transformer hybrid model. Journal of Jiangsu University of Technology, 39(2), 55–62.

[5] Wu, Y., & Sang, N. (2023). Unsupervised domain adaptation for person re-identification based on consistency constraint and label optimisation. Journal of Huazhong University of Science and Technology (Natural Science Edition), 51(4), 70–78.

[6] Chen, R., Peng, Y., & Li, Z. (2023). Water-surface floating object detection method based on continuous unsupervised domain adaptation strategy. Journal of

Dalian University of Technology (Hydraulic Engineering Edition), 63(5), 98–105.
[7] Wang, Y., Sun, G., Zhao, G., & Xue, H. (2024). Unsupervised novel domain adaptation learning based on self-supervised knowledge. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 44(2), 120–129.
[8] Wu, D., Zhang, Z., & Li, Q. (2024). Light and weak target detection method combining machine vision and unsupervised domain adaptation. Journal of Heilongjiang University of Science & Technology, 34(2), 329–336.