

A Review on Image and Video Processing with IoT-Enabled Supervised Learning for Intelligent Surveillance Systems

Don Joseph

Dept. of Computer Science & Engineering
Amal Jyothi College of Engineering
 Kanjirappally, Kottayam, India
 donjoseph722@gmail.com

Fiyona Ann Sojan

Dept. of Computer Science & Engineering
Amal Jyothi College of Engineering
 Kanjirappally, Kottayam, India
 fiyonaann@gmail.com

Jimmy Mathew

Dept. of Computer Science & Engineering
Amal Jyothi College of Engineering
 Kanjirappally, Kottayam, India
 jimmyathew765@gmail.com

Jobin Jomy Mathew

Dept. of Computer Science & Engineering
Amal Jyothi College of Engineering
 Kanjirappally, Kottayam, India
 jerryjoben@gmail.com

Bibin Varghese

Assistant Professor

Dept. of Computer Science & Engineering
Amal Jyothi College of Engineering
 Kanjirappally, Kottayam, India
 bibinvarghese@amaljyothi.ac.in

Abstract—Road traffic accidents are a major concern for global safety, where the delay in detection results in a rise in deaths and economic costs. Recent developments in Artificial Intelligence (AI), Deep Learning (DL), and Internet of Things (IoT) technology have made it possible to automatically detect accidents using visual and sensor data. The research works analyzed include image-based techniques using transfer learning and convolutional neural networks, video-based techniques using spatio-temporal and self-supervised anomaly detection, and IoT-based intelligent traffic monitoring systems. Additionally, resource-effective video compression and transmission techniques applicable to real-time surveillance systems are also analyzed. The comparison between the research works shows differences in input types, computational complexity, real-time capabilities, and applicability to smart city settings. The major issues that arise from the analysis of the research works include the lack of data for detecting rare accidents, the absence of temporal information in image-based techniques, high computational complexity of video-based techniques, and the lack of integration between detection and response systems. The paper concludes by providing recommendations for future research on efficient, scalable, and integrated accident detection systems for intelligent transportation systems.

I. INTRODUCTION

A. Background and Motivation

Road accidents are one of the major causes of death and serious injuries worldwide, creating significant social,

economic, and healthcare challenges [2], [3]. As urbanization increases, road networks become more congested, and factors such as overspeeding, driver distraction, and delayed emergency response further increase the severity of accidents [2]. Early detection of road accidents is crucial because faster emergency response can significantly reduce fatalities and property damage [1].

Traditional traffic monitoring systems mainly rely on human operators who continuously observe surveillance monitors. Such manual monitoring is time-consuming, prone to human error, and often incapable of responding quickly to sudden incidents [3]. These limitations have encouraged researchers to explore automated solutions for traffic surveillance and accident detection.

B. Research Problem

Recent advances in Artificial Intelligence (AI), Deep Learning (DL), and Internet of Things (IoT) technologies have enabled new approaches for automated accident detection [2], [6]. By analyzing visual data from CCTV cameras and roadside monitoring systems, intelligent algorithms can identify unusual traffic patterns and potential accidents in real time [1], [13]. As a result, the several research studies have proposed accident detection systems based on machine learning,

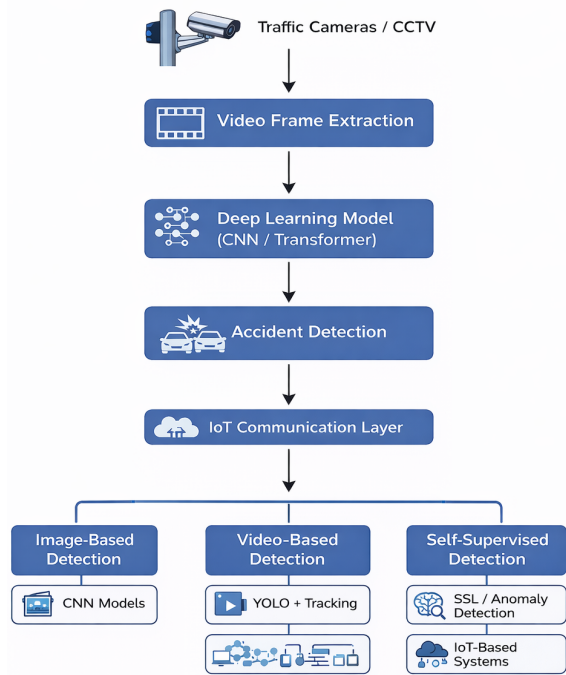


Fig. 1. Architecture of AI- and IoT-based accident detection system

computer vision, and intelligent traffic monitoring frameworks [2], [6].

However, existing studies often focus on specific components of the problem. Some approaches rely on static image analysis, while others use video-based motion analysis or IoT-enabled sensing systems. These approaches differ in terms of computational complexity, detection accuracy, and real-time feasibility. Furthermore, many systems treat accident detection, communication, and emergency response as separate components, which limits the overall effectiveness of intelligent traffic monitoring systems.

C. Scope and Contribution

This paper presents a review of recent research on image and video processing techniques combined with IoT-enabled supervised learning methods for intelligent traffic surveillance. The reviewed studies are grouped into three major categories: image-based accident detection, video-based accident detection, and IoT-based intelligent traffic monitoring systems.

The goal of this study is to analyze how these different approaches perform in terms of detection capability, computational requirements, and practical deployment in real-world environments. By examining the strengths and limitations of existing research, this paper highlights the major challenges in building scalable and efficient accident detection systems.

D. Paper Organization

The remainder of this paper is organized as follows. Section II presents the related work on accident detection techniques using image processing, video analysis, and IoT-based monitoring systems. Section III discusses the comparative analysis

of existing approaches and identifies key research challenges. Finally, Section IV concludes the paper and outlines possible directions for future research in intelligent traffic surveillance systems.

II. RELATED WORK

Road accident detection using AI has become an interdisciplinary area that has been fueled by advancements in AI, Deep Learning, Computer Vision, and IoT [2], [3]. The fifteen papers reviewed in this article discuss road accident detection from a variety of perspectives, including visual sensing, the dynamics of accidents in terms of time, learning strategies, construction of the system, and the efficiency of communication between systems [6], [13]. Instead of providing a single strategy for road accident detection, these papers provide a variety of techniques that can be adapted according to different constraints, such as the availability of data, processing capacity, and practical implementation [6], [11]. For ease of understanding, the papers have been categorized into four categories: image-based road accident detection, video-based road accident detection, self-supervised and anomaly-based road accident detection, and IoT-based intelligent traffic safety systems [3].

A. Image-Based Accident Detection Methods

Image-based accident spotting systems are based on static images of traffic taken by roadside cameras and CCTV systems [6], [17]. Usually, these systems consider accident spotting as a supervised learning classification task where images are classified as either accident or non-accident images [6]. The role of Convolutional Neural Networks (CNNs) in these systems cannot be overemphasized, especially considering their ability to extract spatial features such as vehicle damage, unusual positioning, road blockages, and the effects of accidents [6], [28].

The majority of image-based methods are heavily dependent on transfer learning, which involves the use of pre-trained CNN models like VGG, Inception, ResNet, DenseNet, MobileNet, and EfficientNet. These models, which were pre-trained on large-scale image datasets, have robust feature extraction capabilities that can be fine-tuned on a relatively small-scale traffic accident dataset [6]. Transfer learning is very useful in reducing the training time and improving convergence, especially when there is a lack of labeled accident images [6], [11].

In an attempt to further mitigate the issue of data scarcity, some studies make use of data augmentation and synthetic image generation of accidents [6]. Such methods attempt to create collision situations through the manipulation of vehicle orientation, occlusion, or scene composition [6]. Although image-based methods have low computational complexity and can be easily implemented on edge devices, they are limited in their ability to analyze motion cues and causal relationships between vehicles due to their reliance on single-frame analysis [6]. As such, such systems may find it difficult to differentiate

TABLE I
ANALYTICAL COMPARISON OF ACCIDENT DETECTION PARADIGMS

Approach	Understanding from Existing Studies	Practical Strength	Common Limitations Reported
Image-based detection	Most works treat accident detection as an image classification task using CNNs and transfer learning to capture visible signs like vehicle damage or road blockage.	Easy to deploy, requires less computation, and works well on edge devices.	Cannot understand how events develop over time; may confuse accidents with traffic jams or parked vehicles.
Video-based detection	Researchers analyze vehicle motion across frames, track trajectories, and study interaction patterns to detect abnormal events.	Provides better reliability in complicated and fast-changing traffic situations.	Needs high processing power, large storage, and strong communication support.
Self-supervised anomaly-based	Instead of learning accidents directly, systems learn normal traffic behavior and treat deviations as potential incidents.	Useful when labeled accident data is limited and can detect unseen situations.	Performance changes across locations; selecting proper thresholds is still challenging.
IoT-enabled systems	Combines visual intelligence with sensors, networking, and automatic alert mechanisms to support emergency response.	Helps in faster reporting and enables integration with smart city services.	Integration between modules, delay in communication, and coordination remain open problems.

between actual accidents and similar non-accident scenarios such as traffic congestion or roadside parking [6].

B. Video-Based Accident Detection Approaches

Video-based accident detection methods can be viewed as an extension of image-based analysis techniques, incorporating the benefits of temporal information available in continuous traffic video streams [6]. These methods are able to model the motion of vehicles and their interactions from frame to frame, allowing for the detection of dynamic events like sudden braking, accidents, abnormal lane changes, and post-crash behavior [6], [13].

Most video-based systems integrate object detection algorithms such as YOLO and region-based detectors with multi-object tracking algorithms to track each car over time [21], [24], [27]. By analyzing the trajectories and movements, they are able to identify abnormal behaviors that indicate accidents [6]. Multi-stream networks improve accuracy by processing spatial appearance and temporal motion information in separate pathways [6], [19].

Current studies are exploring transformer-based architectures and spatio-temporal attention mechanisms to capture long-range dependencies in video data [7]. Models such as Vision Transformers and Video Swin Transformers have demonstrated a strong capability to learn hierarchical spatio-temporal representations, making them suitable for complex traffic scenarios [8], [10]. However, these models are computationally expensive and require robust hardware and high-quality video input [7].

Video-based systems are more accurate than image-based systems, but they also introduce their own challenges [6]. Scaling up, storing data, and transmitting it over networks become more difficult when continuous video processing is involved [5]. Continuous computation and communication requirements make it less feasible to scale up or implement in resource-constrained environments [4], [5].

C. Self-Supervised and Anomaly-Based Detection Techniques

The rarity of traffic accidents and the high cost of manually annotating data have driven researchers towards self-supervised learning and anomaly detection [11], [13]. Rather than training on examples of accidents that are labeled, self-supervised learning algorithms learn what normal traffic behavior is and mark deviations from it as possible accidents [11]. This is particularly useful in a surveillance context where data about accidents is scarce but data about normal traffic is abundant [13].

Self-supervised learning leverages proxy tasks such as frame reconstruction, learning from temporal consistency, and predicting future frames to learn meaningful spatial and temporal representations without requiring direct labels for accidents [11]. By enforcing temporal consistency and views, self-supervised learning improves robustness [13]. This enables the system to detect accidents that are rare or unseen during the training process [13].

Anomaly-based systems do not pick up on all the same indicators in each environment [11]. They are very susceptible to changes in lighting, weather, camera angle, and the density of traffic in the scene [11]. What is considered “normal” traffic flow can vary greatly from one location to another and from season to season, making it difficult to establish a universal standard [11]. The problem of threshold levels still remains, and if it is set too low, it will be inundated with false positives [11].

D. IoT-Enabled Intelligent Traffic Safety Systems

IoT-based accident detection systems integrate sensing, communication, computation, and response to ensure that real-time traffic monitoring and emergency response are functioning seamlessly [2], [3]. These systems combine AI-powered visual analysis with sensor networks, cloud computing, and edge computing to automatically detect accidents and send immediate alerts [2].

Some IoT-based frameworks have end-to-end traffic safety systems that include vehicle detection, accident identification,

severity analysis, and notification services [2]. Automatic notifications are sent to hospitals, traffic police, or control rooms, reducing the time taken compared to manual reporting [2]. Research papers related to smart cities emphasize the need for interoperability and scalability to integrate accident detection with traffic management and smart city monitoring [3].

For IoT applications that involve the use of continuous video streams, efficient data transfer is critical [5]. To address the constraints imposed by bandwidth and energy, recent works make use of adaptive video compression and bitrate control learned from data [4], [5]. The aim is to preserve the necessary visual cues for accident detection while reducing the communication overhead [5]. However, many existing approaches consider perception, compression, and communication as isolated modules, which obstructs overall optimization and responsiveness [4].

Despite their advantages, IoT-enabled traffic safety systems face several practical challenges. One of the primary concerns is the reliability of communication between distributed sensors, cameras, and cloud platforms. Network delays or connectivity failures may reduce the effectiveness of real-time accident detection and emergency response. In addition, large-scale deployment of IoT devices requires efficient energy management, secure data transmission, and reliable system maintenance. Privacy and data security also become critical concerns when surveillance cameras continuously collect traffic data. Therefore, future research should focus on developing lightweight edge computing frameworks, secure communication protocols, and scalable architectures that can support large-scale intelligent traffic monitoring systems in smart city environments.

E. Comparative Analysis and Research Limitations

Among the existing solutions, there are obvious trade-offs between accuracy and cost, the amount of data required, and ease of deployment [3], [6]. Image-based solutions are very efficient but lack dynamics [6]. Video-based solutions improve accuracy but require more resources [5], [6]. Self-supervised solutions reduce the need for labeling but may be unstable [11]. IoT-based solutions are more effective in practical scenarios but introduce communication and integration complexities [2], [3].

The literature reveals that there isn't a framework that encompasses perception, learning, communication, and response [3]. Addressing these issues is essential for developing road accident detection systems that are scalable, efficient, and reliable in the context of intelligent transportation [3], [6].

III. DISCUSSION

This section summarizes what has been revealed through the comparative study of existing AI- and IoT-based road accident detection methods [3], [6]. It points out the trends in methodologies, the trade-offs in performance, and the implications of the system, as well as the challenges that still exist [3], [6]. These have been identified through the study

of image-based, video-based, self-supervised, and IoT-based methods [2], [6], [11].

A. Comparison of Detection Paradigms

From the studies analyzed, it is clear that the performance of accident detection systems is largely dependent on the input and learning strategy used [6]. Image-based systems, which rely on static images, perform well in identifying accidents that are quite obvious, such as those involving crashed vehicles or roads that are blocked by accidents [6]. These systems take advantage of convolutional neural networks and transfer learning, which makes them efficient to implement and fast in their operations [6], [28]. However, they fail when it comes to identifying accidents that are in their early stages or when the situation is visually ambiguous [6].

Video-based approaches are likely to outperform other approaches in detection by capitalizing on the motion of things over time [6], [13]. By analyzing how cars move, how their speed changes suddenly, and how they interact from one frame to the next, these approaches become even more accurate in complex traffic scenarios [6]. Approaches that break down the processing task into multiple streams or space and time learning, which deal with what things look like and how they move independently, can take this even further [19]. However, all this comes with a cost [5].

Self-learned and anomaly-driven methods represent a paradigm shift in incident detection, focusing on normal traffic patterns rather than attempting to define precise patterns for accidents [11], [13]. These methods are particularly valuable when there is little labeled data available for accidents and appear very promising for identifying unusual or novel accident patterns [13]. However, their effectiveness may depend on consistent environmental conditions and carefully selected anomaly levels when you apply them in different environments [11].

B. Impact of Model Architecture and Learning Strategy

Architectural choices play an important role in both the effectiveness of detection and the feasibility of a system to be executed. The trade-off between effectiveness and efficiency is well-balanced in the traditional CNN-based architecture, making it suitable for real-time applications and edge devices [7], [28]. In contrast, transformer-based architectures, such as Vision Transformers and Video Swin Transformers, have better global understanding and modeling of long-term dependencies [7], [8], [10]. These models can improve representation capacity in complex scenarios, but at the cost of significantly increased computational complexity and difficulty in training [7].

Overall, the results indicate that no single architecture is the best in all situations [6]. Lightweight CNNs perform best in scenarios with limited resources, while transformer architectures and multi-stream architectures are most likely to emerge as the best ones in situations where accuracy is the only criterion [7], [19]. The message is clear: select the architecture depending on the deployment scenario [6].

TABLE II
COMPARISON OF AI-BASED ACCIDENT DETECTION TECHNIQUES

Method	Input Type	Key Model	Detection Accuracy	Main Limitation
Image-based detection	Static images	CNN / Transfer Learning	Medium	Cannot capture temporal information
Video-based detection	Video streams	YOLO + Multi-object tracking	High	High computational cost
Self-supervised detection	Video data	Self-supervised learning models	Medium-High	Sensitive to environmental changes
IoT-based systems	Sensors + Cameras	Edge AI + Communication modules	High	Integration and scalability challenges

TABLE III
PERFORMANCE VS RESOURCE TRADE-OFF ACROSS DETECTION STRATEGIES

Approach	Detection Reliability	Data Requirement	Real-Time Feasibility	Scalability
Image-based methods	Works well when accident signs are clearly visible in a frame.	Needs a reasonable amount of labeled images.	Easy to achieve in most systems.	Can be expanded without major difficulty.
Video-based methods	More dependable because vehicle movement and interactions are analyzed over time.	Requires long and carefully annotated video data.	May need optimization for strict time limits.	Scaling requires higher infrastructure support.
Self-supervised methods	Capable of identifying unusual or previously unseen events.	Very little labeled accident data is needed.	Possible, but depends on environment stability.	Adaptation to new locations can be challenging.
IoT-enabled systems	Strong at enabling quick alerts and coordinated response.	Depends on inputs from cameras, sensors, and networks.	Suitable when edge or cloud support is available.	Implementation becomes complex at large scale.

C. Role of Data Availability and Dataset Limitations

Availability of data is still a bottleneck in accident detection research [6], [11]. The data is highly imbalanced, with actual accident events constituting a very small portion of the traffic data [13]. For image-based solutions, the problem is attempted to be remedied by data augmentation and simulated accident generation [6]. In video-based solutions, the usual strategy is to use short, labeled video clips extracted from longer surveillance videos [6], [17].

However, self-supervised learning reduces the requirement for labeled data on accidents, but the problem is transferred to the definition of what constitutes normal behavior in a given scenario [11], [13]. Variations due to camera angles, traffic density, weather, and lighting conditions make generalization difficult [11]. Additionally, there is no standardized benchmark for accident detection, which is a setback in comparing research work on this topic [3], [17].

D. System-Level Considerations in IoT-Enabled Deployments

Accident detection using IoT goes beyond mere sensing, as it relates to the connection between perception and communication, as well as automation [2], [3]. From the studies surveyed, it is clear that automated notification systems can significantly reduce the time it takes to respond to emergencies compared to manual notification systems [2]. Cloud-based designs are scalable, and edge computing can reduce latency [2], [3].

Most systems break down perception, communication, and response into multiple modules [3]. This modularization means that you have to make compromises: you get suboptimal detection performance, inefficient or slower transmission, and

reduced overall responsiveness [3]. Methods such as video compression and adaptive transmission can address bandwidth issues, but they may impair image quality and detection reliability if not jointly optimized with perception models [4], [5].

E. Real-Time Feasibility and Scalability

Real-time processing is a crucial requirement for practical applications [2], [6]. Image-centric methods are more likely to satisfy real-time requirements since they are less computationally intensive, whereas video-centric and transformer-based models may require additional optimization to achieve latency requirements [6], [7]. Scaling up such implementations in smart city infrastructure introduces additional challenges for scalability, maintenance, and resource allocation [3].

IV. CONCLUSION

AI and IoT-based approaches have significantly improved the process of automated road accident detection by leveraging the potential of visual information, intelligent learning models, and networked infrastructure [2], [6]. Image processing techniques are computationally efficient and easy to implement, making them suitable for edge devices, but they lack temporal information that can sometimes impair the detection process in dense traffic scenarios [6]. Video processing approaches improve detection accuracy by analyzing space-time relationships and vehicle interactions, but they are computationally expensive [5], [6].

Self-learning and anomaly-driven approaches assist in overcoming the challenge of limited labeled data related to accidents, making it feasible to identify unusual or novel incidents

[11], [13]. However, their effectiveness depends on their adaptability to environmental changes and the selection of threshold values [11]. When IoT is considered, the system becomes more viable by associating perception with communication and automatic response to emergency

REFERENCES

- [1] M. I. B. Ahmed, R. Zaghdoud, M. S. Ahmed, et al., "A Real-Time Computer Vision Based Approach to Detection and Classification of Traffic Incidents," *Big Data and Cognitive Computing*, 2023.
- [2] S. Bhattacharya, H. Jha, and R. P. Nanda, "Application of IoT and Artificial Intelligence in Road Safety," in *2022 Interdisciplinary Research in Technology and Management (IRTM)*, 2022.
- [3] C. Englund, E. E. Aksoy, F. Alonso-Fernandez, et al., "AI Perspectives in Smart Cities and Communities to Enable Road Vehicle Automation and Smart Traffic Control," *Smart Cities*, 2021.
- [4] X. Zhou, D. Huang, Z. Qi, et al., "A Multi-Scale Spatial-Temporal Network for Wireless Video Transmission," *arXiv preprint arXiv:2411.09936*, 2024.
- [5] A. Hojjat, J. Haberer, and O. Landsiedel, "MCUCoder: Adaptive Bivariate Learned Video Compression for IoT Devices," *arXiv preprint arXiv:2411.19442*, 2024.
- [6] S. T. Fu, L. B. Theng, B. L. C. Shiong, et al., "A Multi-Stream Approach to Mixed-Traffic Accident Recognition Using Deep Learning," *IEEE Access*, vol. 12, 2024.
- [7] K. Han, Y. Wang, H. Chen, et al., "A Survey on Vision Transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 87-110, Jan. 2023.
- [8] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021.
- [9] Vision Transformer (ViT). Described in [8] and surveyed in [7].
- [10] Z. Liu, J. Ning, Y. Cao, et al., "Video swin transformer," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [11] R. Ali, M. U. K. Khan, and C. M. Kyung, "Self-Supervised Representation Learning for Visual Anomaly Detection," *arXiv preprint arXiv:2006.09654*, 2020.
- [12] S. Bhattacharya, H. Jha, and R. P. Nanda, "Application of IoT and Artificial Intelligence in Road Safety," *Conference Paper*, March 2022.
- [13] J. Fang, J. Qiao, J. Bai, et al., "Traffic Accident Detection via Self-Supervised Consistency Learning in Driving Scenarios," *IEEE Transactions on Intelligent Transportation Systems*, 2022.
- [14] J. Fang, J. Qiao, J. Bai, et al., "Traffic Accident Detection via Self-Supervised Consistency Learning in Driving Scenarios," *IEEE Transactions on Intelligent Transportation Systems*, 2022.
- [15] J. Fang, J. Qiao, J. Bai, et al., "Traffic Accident Detection via Self-Supervised Consistency Learning in Driving Scenarios," *IEEE Transactions on Intelligent Transportation Systems*, 2022.
- [16] J. Fang, J. Qiao, J. Bai, et al., "Traffic Accident Detection via Self-Supervised Consistency Learning in Driving Scenarios," *IEEE Transactions on Intelligent Transportation Systems*, 2022.
- [17] A. P. Shah, J.-B. Lamare, T. Nguyen-Anh, and A. Hauptmann, "CADP: A novel dataset for CCTV traffic camera based accident analysis," in *Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Nov. 2018.
- [18] L. Wen, D. Du, Z. Cai, et al., "UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking," *Comput. Vis. Image Understand.*, vol. 193, 2020.
- [19] S. T. Fu, L. B. Theng, B. L. C. Shiong, et al., "A Multi-Stream Approach to Mixed-Traffic Accident Recognition Using Deep Learning," *IEEE Access*, vol. 12, 2024.
- [20] S. T. Fu, L. B. Theng, B. L. C. Shiong, et al., "A Multi-Stream Approach to Mixed-Traffic Accident Recognition Using Deep Learning," *IEEE Access*, vol. 12, 2024.
- [21] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv preprint arXiv:1804.02767*, 2018.
- [22] C. Szegedy, W. Liu, Y. Jia, et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015.
- [23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017.
- [24] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint arXiv:2004.10934*, 2020.
- [25] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv preprint arXiv:2207.02696*, 2022.
- [26] Y. Ge, H. Wang, F. Gao, et al., "Real-Time Accident Detection in Surveillance Videos Using Deep Learning," *IEEE Access*, vol. 9, 2021.
- [27] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2014.
- [28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. .