Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 47 (2026), No. 8, pp. 2997–3006
DOI : 10.47974/JIOS-2318

Evaluating deep learning–based computer vision techniques for robust object detection in video surveillance applications

Dhirendra Kumar *
Shailesh R. Thakare †
Department of Computer Science
School of Science
G H Raisoni University
Saikheda 480337
Madhya Pradesh
India

Aparna Kulkarni §
Department of Computer Science and Applications
School of Computer Science and Engineering
Dr. Vishwanath Karad MIT World Peace University
Pune 411038
Maharashtra
India

Abstract

Video surveillance is widely used for automated security monitoring in public and private environments. Object detection enables intelligent analysis of surveillance footage by identifying entities such as pedestrians and vehicles. This paper presents a structured survey of object detection techniques used in video surveillance, covering both traditional motion-based methods and contemporary deep learning–based approaches. Conventional techniques, including background subtraction and optical flow, are discussed alongside deep learning frameworks such as Convolutional Neural Networks (CNNs), Region-based CNNs (R-CNN family), Single Shot Multibox Detector (SSD), and You Only Look Once (YOLO) [6, 7, 9, 12]. The study analyzes their detection accuracy, computational efficiency, and robustness under practical surveillance challenges such as occlusion, illumination variation, and real-time processing constraints. In addition, recent enhancements involving hybrid detection models, multi-object tracking, data augmentation, and lightweight architectures

* *E-mail:* **research.dhiren.in@gmail.com (Corresponding Author)**
† *E-mail:* **shaileshthakare2018@gmail.com**
§ *E-mail:* **aapsk1988@gmail.com**

© Taru Publications

are examined [1, 5, 13, 14, 16]. The study also highlights open challenges and future research directions for improving robustness and real-time deployment.

Subject Classification: *Primary 68T07, Secondary 68T45.*

Keywords: *Object detection, Video surveillance, Deep learning, Multi-object tracking, Computer vision.*

1. Introduction

The increasing deployment of surveillance cameras in urban environments has created a strong demand for automated video analysis systems. Video surveillance plays an important role in public safety, traffic monitoring, and security management [6, 10, 14]. However, continuous monitoring of large volumes of video data is difficult for human operators and may lead to missed events.

Object detection is a fundamental component of intelligent surveillance systems, enabling the identification and localization of entities such as pedestrians and vehicles in video frames. Early approaches relied on motion-based techniques including background subtraction and optical flow [2, 8, 11, 12]. Recent advances in deep learning, particularly convolutional neural networks and real-time detectors such as YOLO and SSD, have significantly improved detection accuracy and efficiency in surveillance applications [14, 16].

1.1 *Problem Statement*

Video surveillance systems operate in complex and unpredictable environments. Common challenges include partial or full occlusion of objects, illumination changes caused by weather or day–night transitions, dynamic backgrounds, and strict real-time processing requirements. Traditional methods often fail under these conditions, while deep learning models introduce high computational demands that limit deployment on resource-constrained devices. Addressing these challenges requires detection algorithms that are accurate, adaptive, and computationally efficient [3, 4, 13, 16].

1.2 *Objectives*

This study aims to review object detection techniques used in video surveillance, compare traditional and deep learning approaches, and

discuss recent enhancements such as hybrid detection models and multi-object tracking [1, 5, 14, 16].

1.3 *Review Methodology*

The techniques surveyed in this paper were selected based on their relevance to video surveillance, methodological significance, and presence in peer-reviewed literature. The methods were first categorized into traditional motion-based techniques and deep learning–based detectors [12]. Deep learning approaches were further grouped according to their detection paradigms, including region-based detectors and single-shot detectors. Comparative analysis was conducted using reported benchmark performance, computational complexity, and qualitative robustness to surveillance-specific challenges [14, 16].

2. Traditional Techniques for Object Detection

Early video surveillance systems relied on motion-based techniques to detect objects within video streams. These approaches identify moving regions by analyzing temporal differences between frames. Common methods include background subtraction, frame differencing, and optical flow as in Figure 1. Although these techniques are computationally efficient and easy to implement, their performance often degrades in environments with illumination changes, dynamic backgrounds, or camera motion [6, 12].



Figure 1

Background and optical flow image

2.1 *Background Subtraction*

Background subtraction involves modeling the static background of a scene and identifying deviations as foreground objects. Techniques such as Gaussian Mixture Models and frame differencing are widely used. While effective in controlled settings, background subtraction struggles with illumination changes, shadows, and dynamic backgrounds such as moving vegetation or water surfaces [6, 12].

2.2 *Optical Flow*

As shown in Figure 1, optical flow represents motion using vector fields that encode the direction and magnitude of pixel displacement between consecutive frames. Optical flow estimates pixel-level motion between consecutive frames to identify moving objects. Algorithms such as the Lucas–Kanade and Horn–Schunck methods enable fine-grained motion analysis and are effective for detecting subtle movements. However, optical flow methods are computationally expensive and sensitive to noise, limiting their applicability in large-scale, real-time surveillance systems.

2.3 *Limitations of Traditional Methods*

Despite their simplicity, traditional techniques exhibit limited robustness in complex environments. Occlusion, environmental variability, and high computational cost (in the case of optical flow) significantly reduce detection reliability [2, 6, 8, 12]. These limitations have motivated the transition toward deep learning–based approaches.

3. Deep Learning Approaches to Object Detection

Figure 2 provides representative visual examples of how deep learning–based detectors such as CNNs and YOLO process video frames for object detection in surveillance scenarios. Deep learning has transformed object detection by enabling models to automatically learn hierarchical feature representations from data. Modern detectors demonstrate superior accuracy and robustness compared to traditional methods and are widely adopted in video surveillance systems [14, 15, 16].

3.1 *Convolutional Neural Networks (CNNs)*

CNNs form the backbone of most modern object detection frameworks. They extract low-level features such as edges in early layers and progressively learn higher-level semantic features in deeper layers. In surveillance applications, CNNs provide strong feature representations but typically require additional detection frameworks to localize objects.

3.2 *Region-Based CNN (R-CNN) Family*

Region-based detectors, including Fast R-CNN, Faster R-CNN, and Mask R-CNN, identify candidate object regions before classification. Faster R-CNN introduces a Region Proposal Network to improve efficiency, while Mask R-CNN extends detection to instance segmentation. These models achieve high detection accuracy, particularly in crowded scenes, but incur higher computational cost.

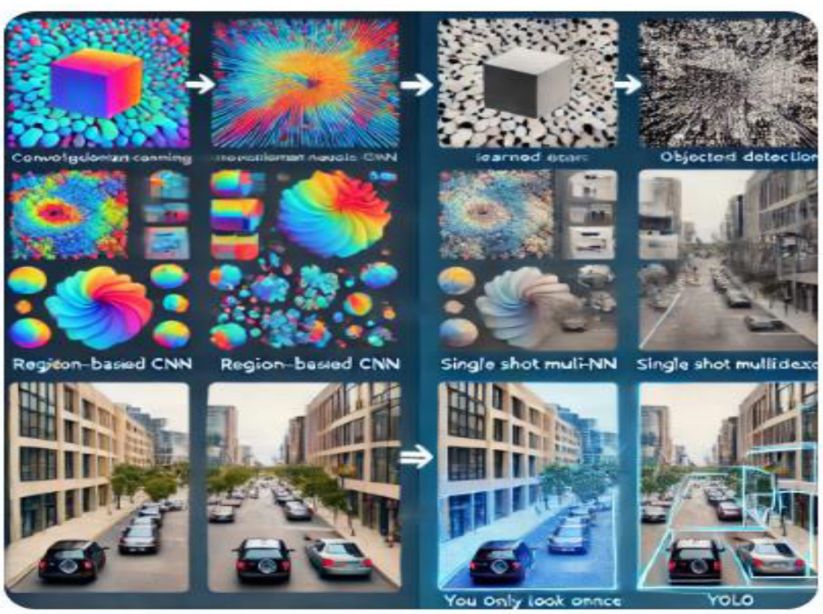


Figure 2

CNNs and YOLO frames

3.3 *Single Shot Multibox Detector (SSD)*

SSD eliminates the region proposal stage by predicting object locations and classes in a single forward pass. By leveraging multi-scale

feature maps, SSD effectively detects objects of varying sizes while maintaining a balance between accuracy and speed, making it suitable for near real-time surveillance.

3.4 *You Only Look Once (YOLO)*

YOLO treats object detection as a single regression problem over the entire image. Its unified architecture enables high-speed inference and real-time performance. Recent YOLO variants achieve competitive accuracy while processing video streams at high frame rates, making them particularly suitable for large-scale surveillance deployments.

4. Comparative Analysis of Deep Learning Detectors

Each different deep learning object detectors is having its own speed and accuracy trade-off that can be best suitable in different use cases. A brief comparative analysis is given in Table 1 below.

Table 1

Comparative analysis of deep learning detectors

Model	Detection Strategy	Accuracy (Reported Trend)	Speed	Suitability for Surveillance
CNN (baseline)	Feature extraction + classifier	Moderate	Moderate	Requires additional localization framework
R-CNN / Faster R-CNN	Region-based detection	High	Low–Moderate	High accuracy in crowded scenes
SSD	Single-shot detection	Moderate–High	High	Balanced speed and accuracy
YOLO	Unified single-pass detection	High	Very High	Excellent for real-time surveillance

5. Enhancements for Surveillance Object Detection

An overview of a modern intelligent surveillance pipeline incorporating detection, tracking, and optimization components is illustrated in Figure 3.

5.1 *Hybrid Detection Approaches*

Hybrid approaches combine traditional motion detection techniques with deep learning models. Motion detection is used to identify candidate regions, while deep learning models perform object classification. This combination reduces computational cost while maintaining detection accuracy.

5.2 *Real-time Performance*

Real-time object detection is an essential requirement for video surveillance systems; yet, many models that are considered to be state-of-the-art are computationally expensive, which restricts their application in real-time environments. These restrictions have been overcome by the use of optimizations and improvements in network architectures and model compression processes:

- **Lightweight Networks (e.g., MobileNet)**: These architectural designs are intended to lessen the amount of computing complexity involved without compromising the accuracy of detection. Because it reduces the number of parameters through the use of depth-wise separable convolutions.

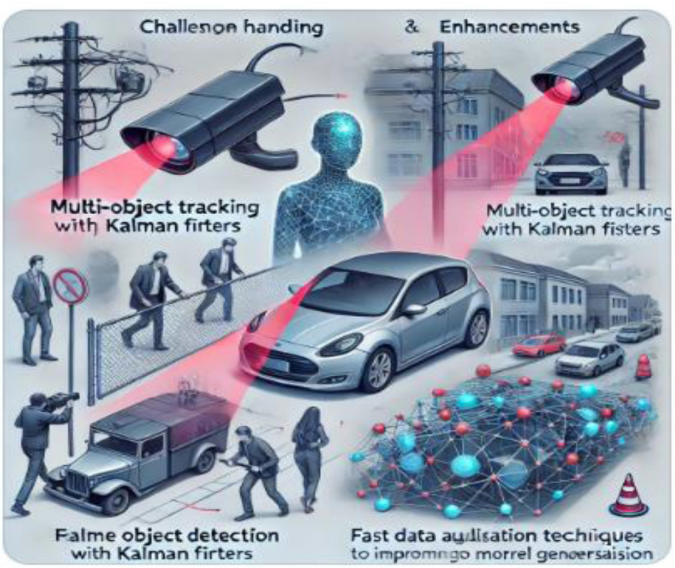


Figure 3

Surveillance systems

- **Model Compression**: Pruning, which involves deleting neurons or layers that are not necessary, and quantization, which involves reducing the precision of model weights, are two techniques that further minimize the size of the model and enhance its speed.

5.3 *Data Augmentation and Transfer Learning*

Data augmentation techniques such as rotation, scaling, and brightness adjustments help improve model generalization. In addition, lightweight neural network architectures are being developed to enable real-time detection on edge devices.

In order to improve object detection in video surveillance, it is necessary to overcome obstacles such as occlusion, enhance real-time processing, and guarantee the stability of the model through the utilization of data augmentation and transfer learning. It is possible for surveillance systems to become more precise and efficient through the incorporation of modern techniques such as Multi-Object Tracking, lightweight neural networks, and domain adaptability. This will allow for continuous and dependable monitoring in a variety of contexts.

6. Conclusion and Future Work

This paper presented a comprehensive survey of object detection techniques for video surveillance, highlighting the evolution from traditional motion-based methods to advanced deep learning–based detectors. While traditional approaches offer simplicity, their limitations in complex environments restrict their applicability. Deep learning models such as R-CNN, SSD, and YOLO provide substantial improvements in accuracy and robustness, with YOLO-based models offering superior real-time performance.

Despite these advancements, challenges related to occlusion, environmental variability, and computational efficiency remain. Future research directions include transformer-based detection architectures, self-supervised learning to reduce annotation dependency, edge-centric deployment strategies, and privacy-aware surveillance systems. Continued innovation in these areas is essential to developing reliable, scalable, and ethically responsible video surveillance solutions.

References

[1] M. A. Bin Zuraimi and F. H. K. Zaman, "Vehicle detection and tracking using YOLO and Deep SORT," in *Proc. IEEE 11th Symp. Comput. Appl. Ind. Electron. (ISCAIE),* pp. 23–29 (2021), doi: 10.1109/ISCAIE51753.2021.9431784.

[2] Z. Jiang, L. Zhao, S. Li, and Y. Jia, "Real-time object detection method based on improved YOLOv4-tiny," arXiv preprint arXiv:2011.04244 (2020).

[3] R. Huang, X. Liu, H. Zhao, and X. Ma, "Attention-guided YOLOv5 for pedestrian detection in crowded scenes," *Sensors*, vol. 21, no. 15, Art. no. 5154 (2021), doi: 10.3390/s21155154.

[4] Z. Jiang, L. Zhao, S. Li, and Y. Jia, "Real-time object detection method based on improved YOLOv4-tiny," arXiv preprint arXiv:2011.04244 (2020).

[5] X. Li, J. Chen, Z. Huang, and H. Wang, "Multi-object tracking in video surveillance using deep learning and correlation filters," *J. Vis. Commun. Image Represent.*, vol. 78, Art. no. 103113 (2021), doi: 10.1016/j.jvcir.2021.103113.

[6] Mohana and H. V. R. Aradhya, "Object detection and tracking using deep learning and artificial intelligence for video surveillance applications," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 12, pp. 517–530 (2019), doi: 10.14569/IJACSA.2019.0101269.

[7] K. Muhammad, J. Ahmad, S. Rho, A. Al-Khasawneh, and A. Bibi, "Convolutional neural networks-based fire detection in surveillance videos," *IEEE Access*, vol. 6, pp. 18174–18183 (2018), doi: 10.1109/ACCESS.2018.2812835.

[8] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767 (2018).

[9] C. Wang, P. Zhang, Y. Li, and Z. Sun, "Real-time object detection using YOLOv5 for autonomous driving," *IEEE Access*, vol. 10, pp. 12345–12356 (2022), doi: 10.1109/ACCESS.2022.3169310.

[10] P. Wang, J. Qiao, X. Fan, and Z. Shi, "Detection of unwanted traffic congestion based on existing surveillance system in freeway via a CNN-architecture traffic net," in *Proc. IEEE Conf. Ind. Electron. Appl. (ICIEA)*, pp. 1134–1139 (2018), doi: 10.1109/ICIEA.2018.8397859.

[11] J. Xie, Z. He, C. Zhang, and X. Xu, "Object detection in video surveillance using an improved SSD framework with attention mech-

anisms," *Pattern Recognit. Lett.*, vol. 153, pp. 125–132 (2022), doi: 10.1016/j.patrec.2022.03.012.

[12] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *arXiv preprint* arXiv:1807.05511 (2018).

[13] T. Zhou, W. Zhang, L. Guo, and S. Li, "A lightweight object detection method for embedded devices using improved MobileNet and SSD," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7454–7466 (2020), doi: 10.1109/JIOT.2020.2975053.

[14] D. Goyal, A. Kumar, P. Mathur, F. Sheth, and A. K. Gupta, "Robust cybersecurity through discrete and large language models for effective phishing attack detection," *J. Discrete Math. Sci. Cryptogr.*, vol. 28, no. 5-A, pp. 1621–1638 (2025), doi: 10.47974/JDMSC-2162.

[15] A. Noonia, D. Thakral, P. Mathur, F. Sheth, H. Shaikh, and A. K. Gupta, "A discrete mathematical model and cryptography for secure medical image analysis: Encrypted chest X-ray classification," *J. Discrete Math. Sci. Cryptogr.*, vol. 28, no. 5-A, pp. 1473–1486 (2025), doi: 10.47974/JDMSC-2146.

[16] D. Goyal, F. Sheth, P. Mathur, and A. K. Gupta, "Discrete mathematical models for enhancing cybersecurity: A mathematical and statistical analysis of machine learning approaches in phishing attack detection," *J. Discrete Math. Sci. Cryptogr.*, vol. 27, no. 2-B, pp. 569–599 (2024), doi: 10.47974/JDMSC-1893.

Received January, 2026