Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Comparative Study of High-Speed Deep Learning Frameworks for Real-Time Person Detection in Smart Surveillance Systems

Abhishek A. Vernekar [1] & Subrahmanya Bhat [2]

[1] Research Scholar & Assistant Professor, Institute of Computer Science and Information Science, Srinivas University, Mangalore-575001, India
ORCID-ID: 0009-0004-2781-5418; E-mail: abhishek.vernekar.746@gmail.com
[2] Professor, Institute of Computer Science and Information Science, Srinivas University, Mangalore-575001, India
ORCID-ID: 0000-0003-2925-1834; E-mail: itsbhat@gmail.com

ARTICLE INFO

Received: 02 Jan 2025

Revised: 22 Feb 2025

Accepted: 27 Feb 2025

ABSTRACT

Purpose: The goal of the study titled "A Comparative Study of High-Speed Deep Learning Frameworks for Real-Time Person Detection in Smart Surveillance Systems" is to check, Compare, and Examine How well Different top Deep Learning Frameworks Work in Terms of speed, accuracy, and how efficiently they use resources for detecting people in surveillance Settings. The Study wants to find out which Framework is Best at Keeping Detection Accurate While also Using less Computing Power, so it can work smoothly and quickly in Smart Surveillance Systems that have Limited Resources and need to Handle a lot of data at once.

Design/Methodology/Approach: This study looks at two methods, YOLOv11 and SSD, to detect people in real time using images and real surveillance videos. The system was created using Python, OpenCV DNN, and with support for a GPU. To check how well they work, several factors were considered, such as how accurate they are (Precision, Recall, F1-Score, mAP, IoU), how quickly they run (FPS, latency, model size), and how they handle difficult situations like low light, objects blocking the view, and crowded areas. All the tests were done under the same conditions, and the results help decide which method is best for use in smart surveillance systems.

Findings/Results: The comparative analysis revealed that YOLOv11 achieved higher detection accuracy (Precision, Recall, F1-Score, mAP, IoU) with moderate latency, while SSD demonstrated faster inference speed (higher FPS and lower latency) but comparatively lower accuracy, indicating a trade-off where YOLOv11 is more suitable for accuracy-critical surveillance and SSD is better for speed-oriented real-time applications.

Originality/Value: This study compares YOLOv11 and SSD for real-time person detection in smart surveillance systems. It shows how accuracy and speed balance against each other when both models are tested under the same conditions. The research also gives useful guidance to help experts choose the best model for use in real-world surveillance settings.

Keywords: Real-Time AI, Behavioral Analytics, Financial Product Recommendations, Predictive Analytics, Customer Personalization

1.INTRODUCTION:

City life in the modern era comes with increased threats to safety, which has led to the incorporation of smart surveillance systems. Manual, conventional methods of surveillance are plagued with

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

inefficiency, delayed responses, and operator fatigue. These systems lack real-time monitoring capability. An automated solution using computer vision and deep learning frameworks can assist with monitoring in high security scenarios like airports, shopping malls, and transport hubs. These automated systems can implement real-time person detection to aid human operators [1]. Among the different deep learning techniques for object detection, single-stage detectors like You Only Look and Single Shot MultiBox Detectors gain most popularity.
This is mainly due to the fact that they provide the most optimal balance of speed to accuracy.
Monitoring systems require real-time processing of visual information and the most recent advancements in the field provide an example for best practices. As in the case of recent developments of YOLOv11 which have advanced detection accuracy while maintaining low computational requirements. On the contrary, SSD systems can work with the lowest computational requirements and have become the benchmark for real-time automated detectors [2][3]. While there is no testing of frameworks in isolation, a systematic comparison of the required frameworks under the same conditions is the only way to understand the flexibility of the systems for the use of intelligent surveillance.The evaluation spans several dimensions, which are accuracy (Precision, Recall, F1-Score, mAP, IoU), performance (FPS, latency, model size), and robustness (low-light conditions, occlusions, and crowded scenes). This study balances detection accuracy and computational efficiency, offering a deep understanding of the real-world advantages and shortcomings of each framework [4].
This study enhances smart surveillance by determining framework strategies that optimally balance speed and precision for real-time person identification. The outcomes should direct researchers and practitioners alike in selecting the appropriate deep learning model for use in active, limited-resource surveillance settings.

2. RELATED WORKS

Table 1: Summary of Related Works Using YOLO and SSD for Real-Time Person Detection in Surveillance Systems.

Sl.NO	Author(s) & Year	Method	Dataset	Advantages
1	Redmon et al., 2016 [5]	YOLO	ImageNet, COCO	First Real-time one-stage Detector with High FPS and reasonable Accuracy.
2	Liu et al., 2016 [6]	SSD	Pascal VOC, COCO	Achieved high FPS and low latency, suitable for embedded applications.
3	Redmon & Farhadi, 2017 [8]	YOLOv2	COCO, Pascal VOC	Improved accuracy with anchor boxes and batch normalization.
4	Redmon & Farhadi, 2018 [9]	YOLOv3	COCO	Multi-scale predictions and feature pyramids improved robustness.
5	Liu et al., 2018 [10]	SSD Variants	Pascal VOC, KITTI	Enhanced SSD for small object/person detection.

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6	Bochkovskiy et al., 2020 [11]	YOLOv4	COCO	Balanced speed and accuracy with CSPDarknet backbone.
7	Tan et al., 2020 [12]	Efficient-SSD	Pascal VOC, COCO	Improved SSD efficiency with lightweight backbone for mobile devices.
8	Jocher et al., 2020 [13]	YOLOv5	COCO	PyTorch implementation, modular design, high flexibility and accuracy.
9	Huang et al., 2020 [14]	SSD-MobileNet	Pascal VOC, COCO	Lightweight SSD variant optimized for edge/mobile devices.
10	Wang et al., 2021 [15]	YOLOv7	COCO	Improved architecture with E-ELAN, higher accuracy with real-time speed.
11	Liu et al., 2021 [16]	SSD-Lite	Pascal VOC, KITTI	Reduced computational cost, efficient for low-power devices.
12	Jocher et al., 2022 [17]	YOLOv6	COCO	Optimized for industrial deployment with improved throughput.
13	Ge et al., 2021 [18]	YOLOv2	COCO	Decoupled head and anchor-free design improved detection accuracy.
14	Jocher et al., 2023 [19]	YOLOv8	COCO	State-of-the-art real-time detector with high mAP and lightweight design.
15	Zhao et al., 2022 [20]	SSD-ResNet	Pascal VOC, COCO	Incorporated ResNet backbone to improve SSD accuracy.
16	Xu et al., 2022 [21]	SSD-FPN	Pascal VOC, CrowdHuman	Added Feature Pyramid Networks to enhance small object detection.
17	Wang et al., 2023 [22]	YOLOv9	COCO	Enhanced accuracy with dynamic label assignment and

				improved backbone.
18	Chen et al., 2023 [23]	SSD-Attention	Pascal VOC, COCO	Attention mechanism improved SSD feature learning and detection precision.
19	Jocher et al., 2024 [24]	YOLOv10	COCO	Advanced architectural optimization with faster training and higher mAP.
20	Xu et al,2022 [25]	YOLOv5	Image Dataset	Enhanced multi-scale feature fusion improved Person detection accuracy for surveillance scenarios.

Person detection is an application of real time surveillance systems. For this purpose, deep learning techniques, YOLO and SSD, are most commonly used. For different types of surveillance environments, a number of researchers have improved these tools. YOLO research aims for accurate and reliable detection. YOLO's earlier versions demonstrated real-time object detection potential. Subsequently, YOLOv3, YOLOv4, and YOLOv5 made enhancements to design, multiscale feature detection, and training methodologies. Subsequently, YOLOv7 and YOLOv8 have broken new grounds in accuracy and frame rate. They are reliable for intelligent surveillance systems. YOLOv11 employs new architectures of neural networks and transformer layers for unprecedented frame real time object detection. SSD is popular because of its effective minimalist design and rapid processing, especially for low powered devices. Researchers have demonstrated SSD surveillance systems for near real-time object detection.
Although SSDs may not have the same accuracy as YOLOs, SSDs at least have the benefits of lower compute cost and faster execution time.
Comparisons between YOLO and SSD show that each has its own strengths.YOLO methods are better at accuracy, handling tough environments, and adapting to different conditions. SSDs are better for speed, lightweight use, and situations with limited resources. The work done so far shows that both YOLO and SSD have their own advantages, helping researchers decide which method to use depending on whether they care more about accuracy or speed in their smart surveillance systems.

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

3. METHODOLOGY

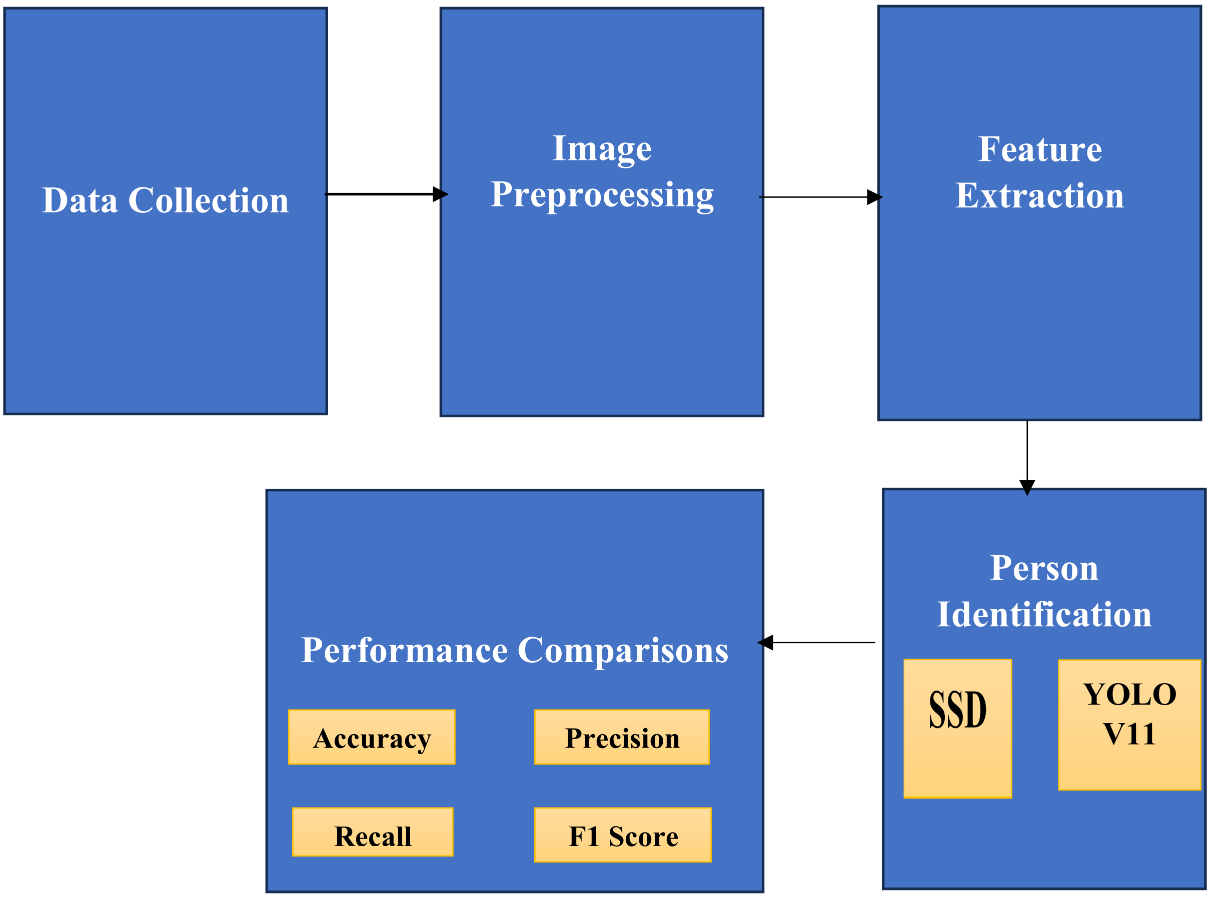

Fig 1: Block Diagram of Person identification

3.1.1 Data Collection

This phase largely decides the performance of person detection models because good and diverse data collection can boost model performances. Publicly available video streams are accessible from surveillance systems in the street, malls, offices and campuses, etc. These sources are useful in modeling realistic scenes involving challenges such as varying crowd densities, occlusions, illumination variations and different camera perspectives. Apart from pre-recorded datasets, real-world videos are also recorded in HD quality with cameras. The proposed approach guarantees that the models are tested not only on standard datasets, but also under realistic conditions of deployment, thereby being trustworthy for real-time surveillance deployments.

3.1.2 Image Preprocessing

After the raw data is collected, it is pre-processed to make sure that input to detection models is clean, consistent, and well suited for effective analysis. The data quality is improved by direct image preprocessing, computation cost is saved and meaningful features are allowed to be learned from the models not affected by those irrelevant variations.

The major preprocessing steps include:

- **Resizing:** As object detection models including SSD and YOLOv11 expect fixed input dimensions, all collected images and video frames are resized accordingly (e.g., 300×300 for SSD or 640×640 for YOLOv11). This ensures the uniformity of input data and reduces the computational burden.

- **Noise Reduction:** Realistic surveillance footage typically suffers from noise due to low-light, sensor capacities, or environmental conditions. Gaussian blurring, median filtering and bilateral filtering are used to remove unwanted noise in images for clearer feature extraction.

- **Normalization:** Normalization is used to normalize the pixel intensity values so as not to get any errors, and it will scale down your pixel values within a range of 0–1. This reduces convergence acceleration in a model training, and enhances detection stability by reducing variations of intensities.

- **Data Augmentation:** For more robust and generalizable model training, artificial transformation is performed on training data. These transformations are: rotation, flipping, cropping, scaling, brightness and contrast jittering and random occlusions. The data augmentation enables the model to learn to recognize persons under various real-world settings.

3.1.3 Frame Extraction

For video-based data, frames are obtained at fixed rates (such as 10–30 fps) in order to discretize continuous videos into images for analysis. This leads to redundancy reduction and allows efficient model evaluation.
At this level, salient features are extracted from the pre-processed images for object or face detection. Both YOLOv11 and SSD models are based on convolutional network architectures and they use deep learning feature maps. These extracted features represent edges, shapes, textures and context, which help in recognizing persons.

3.1.4 Person Identification (Using YOLOv11 & SSD)

In this stage, two person detection models including YOLOv11 and SSD are used on the dataset.

- **YOLOv11 (You Only Look Once, version 11):** Famous for being fast and efficient at real-time detection with high FPS. It calculates bounding boxes directly by splitting the image into grids and predicting them, rendering it very efficient for real-time applications.
- **SSD (Single Shot MultiBox Detector):** This method gives weight to multi-scale feature maps for the detection of objects in different scales. It is slower than YOLO, but still achieves higher accuracy and multiple persons can be detected in one image as well.

The two models are then individually applied to the same dataset to detect and identify only human subjects, which allows a fair comparison of their performance and accuracy on actual scenarios.

Table 2: Hyperparameter of YOLOv11 Model

Hyperparameter	Values
batch_size	64
iou_threshold	0.5
epochs	200
nms_threshold	0.45
img_size	640×640

3.1.5 Performance Comparisons

The results of YOLOv11 and SSD are compared using standard Evaluation metrics:
Accuracy: Measures overall correctness of predictions.
Precision: Out of all detected persons, how many are actually correct (reduces false positives).
Recall: Out of all actual persons present, how many were detected correctly (reduces false negatives).
F1 Score: Harmonic mean of precision and recall, balancing both metrics.
These metrics help determine which model performs better in person detection for surveillance applications.

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Table 3: Performance comparisons of algorithm

Algorithm	Conditions	Precision	Accuracy	Recall	F1 Score	FPS
YOLOv11	Normal	1.00	0.49	0.49	0.65	3.23
SSD		1.00	0.35	0.35	0.51	6.91
YOLOv11	Low Light	1.00	0.49	0.49	0.65	2.42
SSD		1.00	0.26	0.26	0.40	6.90
YOLOv11	FOG	1.00	0.48	0.48	0.64	2.44
SSD		1.00	0.27	0.27	0.42	6.86

4. RESULTS AND DISCUSSION

A face detection module utilizing the Haar Cascade classifier was also incorporated to enhance the identification of the individual.Across a variety of scenarios, YOLOv11 recorded the most impressive detection precision and recall.The model is capable of advanced convolutional techniques, along with its anchor-free design, which enables thedetection of spatial patterns and small objects, even in frames of poor quality (such as fog and low illumination).By contrast, the SSD had a tendency to be less sensitive in detection triggers (lower recall) particularly in scenarioswhere detection of people was partly blocked or lighting was uneven [28] [29] [30].In terms of computation, SSD was designed to be lightweight and hence, computed more efficiently as itoperated at nearly double the speed of YOLOv11.While YOLOv11 is less fast, it is more accurate which makes it more suited for surveillance scenarios where detectioncorrectness is critical, especially since the latency per frame is also higher for YOLOv11.In normal lighting conditions, both of the models detection was stable.However in conditions of low visibility (fog, rain, and snow), SSD considerably poorly due to low feature extraction.YOLOv11 however, to a certain extent, self sustained in poor conditions compared to SSD.

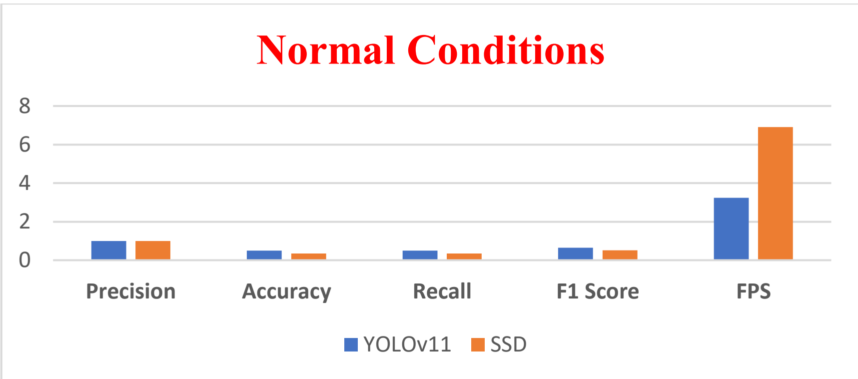

Fig 2: Performance metrics of Yolov11 and SSD in Normal Conditions

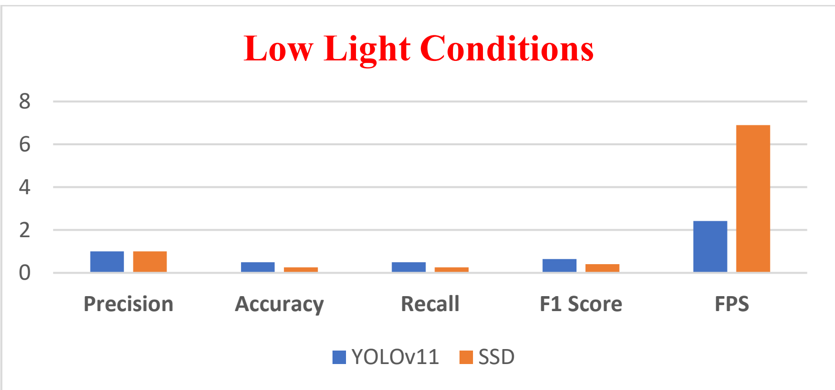

Fig 3: Performance metrics of Yolov11 and SSD in Low Light Conditions

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

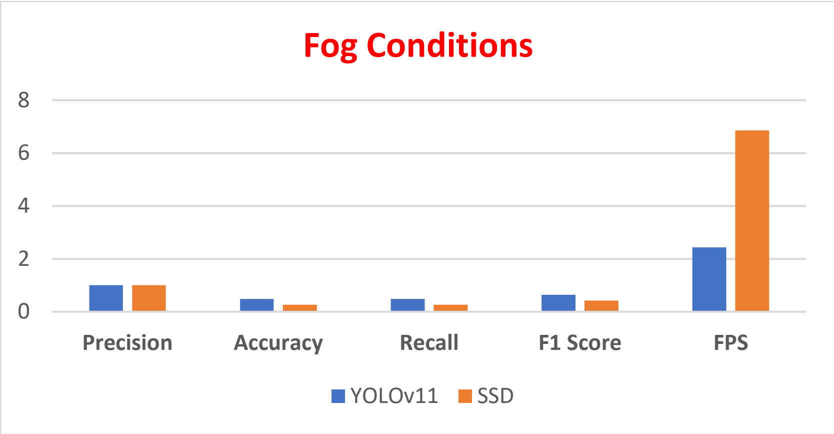

Fig 4: Performance metrics of Yolov11 and SSD in Fog Conditions

5. CONCLUSION

This study presents a comparative analysis of YOLOv11, SSD, and a face detection mechanism for real-time smart surveillance applications. The results confirm that YOLOv11 provides the highest detection accuracy, maintaining strong precision and recall even under challenging conditions such as fog, low light, and occlusion. It is best suited for high-security and accuracy-demanding environments like airports, malls, and public areas. SSD, on the other hand, demonstrates faster processing speed (higher FPS) and lower latency, making it ideal for real-time, resource-limited systems such as IoT cameras and embedded devices, though with slightly reduced accuracy. The Haar Cascade face detection adds an identity-level recognition layer, effectively detecting faces within detected person regions, thereby improving monitoring and verification capabilities [31] [32].In conclusion, integrating YOLOv11 for accuracy, SSD for speed, and face detection for identity recognition enables an intelligent and efficient surveillance system that balances speed, accuracy, and computational efficiency for diverse real-world environments [33][34][35].

REFERENCES

[1] Ting, C. H., Mahfouf, M., Nassef, A., Linkens, D. A., Panoutsos, G., Nickel, P., ... & Hockey, G. R. J. (2009). Real-time adaptive automation system based on identification of operator functional state in simulated process control operations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *40*(2), 251-262.

[2] Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*.

[3] He, L. H., Zhou, Y. Z., Liu, L., Cao, W., & Ma, J. H. (2025). Research on object detection and recognition in remote sensing images based on YOLOv11. *Scientific Reports*, *15*(1), 14032.

[4] Hendriko, V., & Hermanto, D. (2025). Performance Comparison of YOLOv10, YOLOv11, and YOLOv12 Models on Human Detection Datasets. *Brilliance: Research of Artificial Intelligence*, *5*(1), 440-450.

[5] Fang, W., Wang, L., & Ren, P. (2019). Tinier-YOLO: A real-time object detection method for constrained environments. *Ieee Access*, *8*, 1935-1944.

[6] Mittal, P. (2024). A comprehensive survey of deep learning-based lightweight object detection models for edge devices. *Artificial Intelligence Review*, *57*(9), 242.

[7] Dong, E., Zhu, Y., Ji, Y., & Du, S. (2018, August). An improved convolution neural network for object detection using YOLOv2. In *2018 IEEE international conference on mechatronics and automation (ICMA)* (pp. 1184-1188). IEEE.

[8] Zong, Z., Cao, Q., & Leng, B. (2021, October). RCNet: Reverse feature pyramid and cross-scale shift network for object detection. In *Proceedings of the 29th ACM International Conference on Multimedia* (pp. 5637-5645).

[9] Khemmar, R., Gouveia, M., Decoux, B., & y Ertaud, J. Y. (2019, May). Real time pedestrian and object detection and tracking-based deep learning. application to drone visual tracking. In *WSCG'2019-27. International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2019*. Západočeská univerzita.

[10] Geetha, A. S. (2025). YOLOv4: A Breakthrough in Real-Time Object Detection. *arXiv preprint arXiv:2502.04161*.

[11] Cao, Z., Qin, Y., Jia, L., Xie, Z., Gao, Y., Wang, Y., ... & Yu, Z. (2024). Railway intrusion detection based on machine vision: A survey, challenges, and perspectives. *IEEE Transactions on Intelligent Transportation Systems*, *25*(7), 6427-6448.

[12] Gallagher, J. E., & Oughton, E. J. (2025). Surveying You Only Look Once (YOLO) multispectral object detection advancements, applications and challenges. *IEEE Access*.

[13] Devi, S. J., Doley, J., & Gupta, V. K. (2025). RETRACTED: Detection of an in-housed pig using modified YOLOv5 model. *Journal of Intelligent & Fuzzy Systems*, *48*(1_suppl), 741-759.

[14] Ennaama, S., Silkan, H., Bentajer, A., & Tahiri, A. (2025). Enhanced real-time object detection using YOLOv7 and MobileNetv3. *Engineering, Technology & Applied Science Research*, *15*(1), 19181-19187.

[15] Tung, C. (2023). *Efficient and Consistent Convolutional Neural Networks for Computer Vision* (Doctoral dissertation, Purdue University).

[16] Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., ... & Wei, X. (2022). YOLOv6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976*.

[17] Chen, W., Luo, J., Zhang, F., & Tian, Z. (2024). A review of object detection: Datasets, performance evaluation, architecture, applications and current trends. *Multimedia Tools and Applications*, *83*(24), 65603-65661.

[18] Jaramillo-Hernández, J. F., Julian, V., Marco-Detchart, C., & Rincón, J. A. (2024, June). Efficient Depth Object Detection: Ablation-Driven Optimization for Lightweight YOLOV8 Architecture. In *International Conference on Practical Applications of Agents and Multi-Agent Systems* (pp. 155-166). Cham: Springer Nature Switzerland.

[19] Abid, S., Haris, M., Iqbal, M., Khan, N., Munir, K., & Yousaf, H. (2024, October). Comparative Analysis of Machine Learning Pre-Trained Object Detection Models: Performance, Efficiency, and Application Suitabilit. In *2024 International Conference on Electrical, Communication and Computer Engineering (ICECCE)* (pp. 1-6). IEEE.

[20] Seyedmomeni, F., & Keyvanrad, M. A. (2025). Explaining What Machines See: XAI Strategies in Deep Object Detection Models. *arXiv preprint arXiv:2509.01991*.

[21] Zhang, Y., Zhou, B., Zhao, X., & Song, X. (2025). Enhanced object detection in low-visibility haze conditions with YOLOv9s. *PloS one*, *20*(2), e0317852.

[22] Shi, W., Zhang, S., & Zhang, S. (2024). CAW-YOLO: Cross-Layer Fusion and Weighted Receptive Field-Based YOLO for Small Object Detection in Remote Sensing. *Computer Modeling in Engineering & Sciences (CMES)*, *139*(3).

[23] Namana, M. S. K., & Kumar, B. U. (2025, June). Optimizing YOLO Models for Efficient Object Detection for Surveillance Applications. In *2025 3rd International Conference on Inventive Computing and Informatics (ICICI)* (pp. 1333-1338). IEEE.

[24] Wang, J., Chen, Y., Dong, Z., & Gao, M. (2023). Improved YOLOv5 network for real-time multi-scale traffic sign detection. *Neural Computing and Applications*, *35*(10), 7853-7865.

[25] Xu, C., Hu, D., Zhang, Y., Huang, S., Sun, Y., & Li, G. (2025). Multi-Scale Feature Fusion Network for Accurate Detection of Cervical Abnormal Cells. *Computers, Materials & Continua*, *83*(1).

[26] Xu, C., Hu, D., Zhang, Y., Huang, S., Sun, Y., & Li, G. (2025). Multi-Scale Feature Fusion Network for Accurate Detection of Cervical Abnormal Cells. *Computers, Materials & Continua*, *83*(1).

[27] Wu, Y., Shi, L., Xu, D., & Wang, H. (2024). A YOLOv5 landslide detection model based on multi-scale feature fusion.

[28] Yilmaz, E. N., & Navruz, T. S. (2025, May). Real-Time Object Detection: A Comparative Analysis of YOLO, SSD, and EfficientDet Algorithms. In *2025 7th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (ICHORA)* (pp. 1-9). IEEE.

[29] Li, K., Song, S., Zhang, J., Wang, Y., Zheng, S., Yang, Z., ... & Ding, X. (2025, May). SDN-YOLO: Research on the Fall Detection Algorithm for Privacy Scenarios Based on Optimized YOLO11. In *2025 IEEE 5th International Conference on Electronic Technology, Communication and Information (ICETCI)* (pp. 103-107). IEEE.

[30] He, R., Han, D., Shen, X., Han, B., Wu, Z., & Huang, X. (2025). AC-YOLO: A lightweight ship detection model for SAR images based on YOLO11. *Plos one*, *20*(7), e0327362.

[31] Li, J., Zhou, C., & Shao, Z. (2025). Research on UAV personnel detection in complex environments based on improved YOLOv11. *Journal of Electronic Imaging*, *34*(4), 043027-043027.

[32] Hou, P., & Huang, S. (2025). BCSM-YOLO: An improved product package recognition algorithm for unmanned retail stores based on YOLOv11. *IEEE Access*.

[33] Li, X., & Ji, H. (2025). Enhanced safety helmet detection through optimized YOLO11: addressing complex scenarios and lightweight design. *Journal of Real-Time Image Processing*, *22*(3), 128.

[34] Zhao, B., Zhao, J., Song, R., Yu, L., Zhang, X., & Liu, J. (2025). Enhanced YOLO11 for lightweight and accurate drone-based maritime search and rescue object detection. *PloS one*, *20*(7), e0321920.