A Review of AI-Powered Video Surveillance Systems for Automated Suspect Identification

1st Haroon Arshad
Faculty of Engineering and Technology
The Islamia University of Bahawalpur
Punjab, Pakistan
haroonarshad.cyber@gmail.com

2nd Muhammad Amaaz
Faculty of Engineering and Technology
The Islamia University of Bahawalpur
Punjab, Pakistan
muhammadamaaz.cyber@gmail.com

3rd Hamid Iqbal
Faculty of Engineering and Technology
The Islamia University of Bahawalpur
Punjab, Pakistan
hamidiqbal.cyber@gmail.com

4th Asjad Amin
Faculty of Engineering and Technology
The Islamia University of Bahawalpur
Punjab, Pakistan
asjad.amin@iub.edu.pk

5th Umar Fayyaz
Faculty of Engineering and Technology
The Islamia University of Bahawalpur
Punjab, Pakistan
umar.fayyaz@iub.edu.pk

***Abstract*—AI in video surveillance has been a game changer, elevating systems into a new industry, widely known as smart monitoring systems. This review summarizes recent developments in the domain of AI-based video surveillance systems, emphasizing the convergence of object detection, face detection, and face recognition. In Particular, it expounds the real-time human detection, explained the deep learning techniques including YOLO, the facial localization by MTCNN and RetinaFace, identity verification by ArcFace and FaceNet. This paper discusses the relative merits and drawbacks of these approaches, how they have been employed in practice, and the ethical and societal implications they conjure. The goal of this review is to summarize and analyze the state of research, describe a selection of performance benchmarks, spot notable gaps in the literature, and recommend paths for future development in automated suspect identification.**

***Index Terms*—Artificial Intelligence, Video Surveillance, Human Detection, Face Recognition, YOLO, MTCNN, ArcFace, Public Safety, Privacy, Deep Learning**

I. Introduction

Surveillance systems have evolved from simple manual monitoring to intelligent automated frameworks – transforming public safety and security enforcement [1] [2]. In the face of growing urban populations and increasingly sophisticated threats facing jurisdictions today, timely and accurate identification of suspects are of paramount importance [3] [4]. Traditional systems that depend on human operators reviewing static footage are no longer sufficient for timely threat mitigation. AI-enabled surveillance overcomes this problem through automation at several levels in the process—from detecting individuals in crowded areas to confirming they are who they claim through facial biometrics [5] [6] [7] [8]. This review focuses specifically on AI methods in the field of video surveillance with an emphasis on the last task in the video surveillance pipeline: Automated Suspect Identification. Among others, the scope covers object detection, face detection, face recognition, system integration strategies, ethical concerns, and contemporary deployment case studies.

The purpose of this review is to evaluate the usefulness of modern AI techniques utilized in automated video surveillance systems, compare model performance, and highlight current research gaps and future potential.

II. Evolution of Video Surveillance Technologies

Video surveillance systems have progressed long ago from analog to digital systems ranging from resolution and storage to the intelligent [6] [9]. As social interest in crime prevention and public safety continues to rise, the demand for image security and automated surveillance has increased significantly. The increasing demands are addressed by the deployment of low-cost imaging systems [10] [11] [12]. In the beginning, traditional systems depended on human observation, requiring around-the-clock manual surveillance to identify abnormal activity. Emerging AI methods can make surveillance systems capable of what can be termed as high-level features (i.e. detect, analyze, and track activities on real time), where the dependency on human is reduced, enabling a higher responsiveness [13] [14] [15] [16].

III. Video and Image Preprocessing in Surveillance

Before real-time face detection in video stream, raw image frames go through additional steps as preprocessing. Surveillance cameras generate streams of images consisting of nearly continuous streams with each image frame processed independently in real time [17] [18]. Preprocessing usually involves scaling, resizing, brightness enhancement, and sharpening to maintain consistent and accurate detection Resolving differences in frame resolution and frame size between the frames is vital to enhancing the model's performance [19]. Photos taken in low-light or night time settings may have brightness adjustments made to enhance visibility and bring focus to items of interest. These techniques are tailored to optimize the input frames for downstream detection and

DOI: https://doi.org/10.67535/tsp.000001.034
Published under the ICACNC 2025 Proceedings

recognition algorithms, maximizing their performance under optimal visual conditions [20] [21] [22].

Generalized pipeline of AI-powered video surveillance for automated suspect identification is shown in fig.1

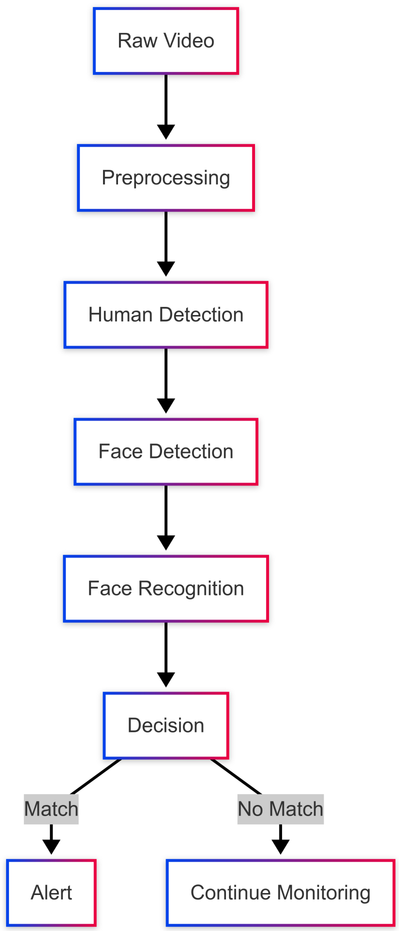

Fig. 1. Generalized pipeline of AI-powered video surveillance for automated suspect identification.

IV. Human Detection Techniques

Person detection in video frames is a basic but important task of automated surveillance. Many deep learning approaches have been proposed for this task, each with a different speed-accuracy trade-off [23] [24] [25].

A. *YOLO (You Only Look Once)*

The success of YOLO is attributed to the fact that it performs localization and classification on the whole image within a single pass, which allows it to be extremely fast and suitable for real-time applications [26] [27]. Versions such as YOLOv4 and YOLOv5 are commonly used in security monitoring systems, especially in places like transport hubs and crowded streets. Recent studies show that YOLO is able to achieve competitive accuracy at very high processing rates, outperforming both accuracy and speed of traditional region-based methods in real-time applications [28] [29].

B. *SSD and Faster R-CNN*

Single Shot MultiBox Detector (SSD) offers a compromise of speed and accuracy, often slower than YOLO [30]. Although Faster R-CNN is very accurate, it is usually employed in offline analysis, as it requires considerable computation. Previous comparative work indicates YOLO is the more fluid choice for live suspect identification, while Faster R-CNN is more appropriate for forensically reviewing videos [31] [32].

C. *Model Comparison*

Object detection algorithms in smart surveillance system are important for distinguishing human presence in video frames. This section provides an extensive comparison between three widely used models with high accuracy: YOLOv4, SSD (MobileNet), and Faster R-CNN (ResNet- 50), based upon the efficiency of implementation and effectiveness in real-time environments. Evaluation focuses on inference time, frame rate (fps) and GPU memory consumption-costing your models as little as possible on edge or resource-constrained devices.

TABLE I
Average Inference Time and Frames Per Second (FPS) of Object Detection Algorithms. [31]

Algorithm	**Inference Time (ms)**	**FPS**
YOLOv4	25	40
SSD (MobileNet)	45	22
Faster R-CNN (ResNet-50)	120	8

Discussion: YOLOv4 manages this impressive speed due to its single-pass detection mechanism. It's real-time capable for tasks like autonomous driving and live video surveillance. SSD is a good balance of both speed and accuracy that can be used in many different scenarios. Faster R-CNN Although it has a high level of accuracy, real-time applications are very troublesome for it.

TABLE II
GPU Memory Consumption of Object Detection Algorithms. [31]

Algorithm	***GPU Memory Usage (MB)***
YOLOv4	2800
SSD (MobileNet)	3200
Faster R-CNN (ResNet-50)	5400

Discussion: Faster R-CNN's architecture is more complex, and its two - stage detection-process requires a greater number of computational resources. Even at its faster processing speed, YOLOv4 manages to keep its memory usage moderate, which makes it feasible for application on devices with limited resources. SSD's memory usage lies somewhere in between that of YOLO and Faster R-CNN, making its performance and simplicity balance have some influence on how much of your secretary data you keep handy.

V. Face Detection Methods

Once human presence is confirmed, it must isolate faces within those frames. Face detection algorithms have become much more robust to variations in illumination, occlusion and pose [4].

A. MTCNN (Multi-task Cascaded Convolutional Networks)

MTCNN can detect faces and facial landmarks at the same time. Using a cascade of three neural networks to refine detections, this makes it a popular choice to use in real-time surveillance as the orientation of the face can vary greatly [33] [34].

B. RetinaFace and Haar Cascades

RetinaFace is a single-stage detector and shows high performance on benchmark datasets. This is especially adept in managing extreme poses and expressions [35]. Haar cascades once prevailed, but have since faded into irrelevance, with abysmal performance in environments with temporal variation. While more resource-intensive, RetinaFace has been shown to be more robust than MTCNN [36] [37] [38].

TABLE III
Dataset Scenarios Used for Face Detection Evaluation, Representing Different Lighting and Crowd Density Conditions. [39]

Scenario	Number of Students	Illuminance (Lux)
1	40	152
2	38	152
3	32	152
4	32	40

TABLE IV
Evaluation Metrics of Face Detection Models Across Four Scenarios. [39]

Scenario	Model	*Precision*	*Recall*	*F1-score*	*Inference Time (s)*
1	Haar Cascade	0.93	0.33	0.48	0.5409
	MTCNN	1.00	0.90	0.95	4.6908
	RetinaFace	1.00	1.00	1.00	9.1460
2	Haar Cascade	0.93	0.34	0.50	0.7535
	MTCNN	1.00	0.97	0.99	10.084
	RetinaFace	0.97	1.00	0.99	12.7296
3	Haar Cascade	0.92	0.34	0.50	0.8030
	MTCNN	0.97	0.97	0.97	4.2710
	RetinaFace	1.00	1.00	1.00	9.6616
4	Haar Cascade	0.89	0.25	0.39	0.5511
	MTCNN	1.00	0.84	0.92	5.5228
	RetinaFace	1.00	1.00	1.00	11.4697

Discussion: The results in Table IV show that in all kinds of lighting and crowd scenes MTCNN has good performance which is both strong and stable.In Scenario 3, the lighting is low and the crowds are high. The F1-score is 0.97 (meaning a trade-off between speed and accuracy is worthwhile), and MTCNN's inference time is moderate.In Scenario 4, although lighting was low and the crowd was smaller, its performance was still good with an F1-score of 0.92 and outperformed Haar Cascade significantly. Although RetinaFace achieved absolute perfect precision, recall, and F1-score in every scene, its inference time was in a class by itself–in all tests, over 9 seconds. Its practitioners must be aware of this limitation because it cannot be used for real-time surveillance. Haar Cascade, on the other hand, was computationally fast but in every metric, especially recall, statical defeat. And as a result of this poor performance it cannot be relied upon under complex conditions. These matching results confirm that MTCNN is a practical solution for face detection under surveillance systems in different situations. It represents the best balance of processing efficiency and detection quality over this wide range of uses.

VI. Face Recognition Models

Face recognition is the last and most decisive part of automated identification of suspects. This involves comparing the detected face against known entries in a database to establish identity [40].

A. ArcFace

ArcFace enhances the field of face embeddings using additive angular margin loss function. It outperforms benchmark datasets such as LFW and MegaFace by a substantial margin, which is why it is often used in verification tasks when it comes to surveillance-based systems [41] [42].

B. FaceNet and FaceNet512

FaceNet proposed the triplet loss that minimizes the distance between a pair of samples from the same class while maximizing the distance between samples of different classes [43].On the other hand, FaceNet512 is an upgradation of FaceNet. It uses a 512-dimensional embedding vector to raise recognition accuracy and lower false positives. The two models use the same underlying architecture and loss function. But by raising the dimensionality of the embedding above 512, in FaceNet512 it becomes possible to give more detailed feature representations. This makes it particularly effective for certain surveillance environments [44] [45] [46].
Though performances of all three methods are competitive, ArcFace has demonstrated superior robustness in noisy, real-world surveillance conditions [47] [48].

TABLE V
Classification Performance Comparison of ArcFace, FaceNet, and FaceNet512 Models. [41]

Metric	*ArcFace*	*FaceNet*	*FaceNet512*
Accuracy	0.878	0.921	0.974
Specificity	0.872	0.938	0.994
Sensitivity	1.000	0.600	0.600
Precision	0.303	0.353	0.857
F1-Score	0.465	0.444	0.706

Discussion: The results presented in Table V shows that FaceNet512 enjoys a good lead in accuracy (0.974), specificity (0.994), precision (0.857), and F1-score (0.706), while the perfect sensitivity is of ArcFace (1.000). This means that in the 190 verification tests carried out, ArcFace made no mistakes at all: it has been able to successfully identify every real attacker and was not found wanting by any given suspect (even if they would have caused a mistake for another system). In a system for identifying criminals, where letting a guilty person go free can have disastrous consequences, What's more, ArcFace still delivers sound overall accuracy (0.878) and specificity (0.872), making it desirable if you want to use it in real-world surveillance scenarios. It may be true that FaceNet512 has an excellent recall rate and therefore is best suited in cases where false positives must be kept to a minimum, but compared to this balance between detection capability and dependability, ArcFace really provides a more even trade-off-of the two. This makes it particularly appropriate for forensic use and security-critical deployments in general.

VII. Results and Comparative Findings

Comparing object detection models, we find that YOLOv4 is the best performing for real-time surveillance. Its high frame rate (40 FPS) and moderate memory usage (2.8 GB) make it particularly attractive. SSD provides a happy ratio between speed and accuracy, while Faster R-CNN not only achieves a higher accuracy but also demands more resources and has a slower speed (8FPS).

Four scenarios were provided to measure the performance of the face detection model under different lighting and crowd scenarios.. MTCNN was able to perform consistently no matter what the conditions, achieving high values of the F1-score (as high as 0.99) and from a machine learning point it has a relatively long serving time. RetinaFace got perfect detection results, however, these came at an increased cost of latency, which to a certain extent is not fit for real-time application. So In terms of reliability in practice, Haar Cascade had reported the fastest time to infer but it also exhibited correspondingly low accuracy.

In facial recognition, ArcFace had perfect sensitivity (1.000). All true positives were found to have been detected.FaceNet512 achieved the highest accuracy (0.974) and precision (0.857), making it an ideal algorithm to use when the emphasis is on minimizing false postivites. FaceNet scored about average across the board degree of excellence in any metric, therefore, can only be used with an area that uses some risk.

These results suggest that we need to pick models on a case-by-case basis according to specific requirements such as real-time detection, forensic accuracy and sufficiently balanced trade-offs.

VIII. Real-World Applications

A. Public Safety Monitoring

CCTV networks in cities such as London, Dubai and Singapore, using AI, actively patrol public areas looking for known suspects. They combine YOLO (You Only Look Once) for human detection systems with ArcFace for identity matching, providing a law enforcement solution that can be scaled up [49] [50] [51].

B. Border Control and Transportation Hubs

Facial recognition technology is applied for passenger identity verification at airports and border checkpoints [52]. They use MTCNN for face detection and ArcFace for recognition to fasten security checking and improve detection [53].

C. Post-Incident Investigation

In post-event situations, footage is analyzed retrospectively based on the AI models applied. These tools assist authorities in searching for relevant frames, identify persons and reconstruct timelines for the events, allowing much faster forensic analysis [54].

IX. Ethical, Legal, and Social Implications

A. Privacy and Consent

Surveillance systems are deployed in common areas, and individuals may not be aware that their data is being collected. This raises serious ethical concerns regarding consent, data usage and storage. Regulatory frameworks, such as GDPR, emphasize on transparency and purpose limitation [55].

B. Bias and Misidentification

These models have demonstrated variable performance across demographics, often performing less accurately in the case of women and ethnic minorities. Such differences can lead to wrongful identification, casting doubt on law enforcement technologies. To address the problem of algorithmic bias, the researchers recommend balanced datasets and the use of regular audits [56].

C. Transparency and Accountability

AI decisions need to be comprehensible, especially in law enforcement contexts.. Alongside human-in-the-loop initiatives to verify AI-created alerts, demands for algorithmic transparency are growing [57].

X. Challenges and Research Gaps

A. Environmental Constraints

Surveillance systems generally operate in oppressive environments — like low-light, crowd density, occlusion, etc. Improving model robustness in this kind of environments is a hot research topic [58].

B. Real-Time Performance on Edge Devices

Models like YOLOv5 and ArcFace are extremely optimized but still cannot run in real time on embedded hardware. There is ongoing work towards edge-compatible variants and hardware-accelerated solutions (such as NVIDIA Jetson) [59].

C. Lack of Standardized Benchmarks

Existing datasets related to face recognition (LFW, MegaFace) do not provide a unified benchmark that mimics real-life surveillance conditions. An exhaustive standard would save for a powerful base of measuring and comparing models in a uniform methodology [60].

XI. Future Directions

The future research lies towards lightweight models to be used for deployment on the edge devices, multi-modal biometric fusion (voice and gait) and privacy-preserving techniques (like federated learning). Policy adjustments are needed for a balance between civil liberties and technological capabilities [61] [62] [63].

XII. Conclusion

AI surveillance systems are used to detect and track the targets using data mining and machine learning–based search algorithms. Technologies like YOLO, MTCNN and ArcFace have made surveillance more dynamic, accurate and scalable. However, there are technical limitations along with ethical dilemmas and social resistance but we are with building sufficient safeguards before mass scale implementation. Maintaining close collaborations across academia, policy, and industry will be critical to the realization and wisdom of these systems.

References

[1] C. Sivakumar, T. K. Vali, P. S. B. Reddy, M. L. Meghana, and Y. Sukumar, "Ai-powered video surveillance for enhanced intrusion detection," in *2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*. IEEE, 2024, pp. 1630–1634.

[2] J. Sivapriya, D. R. Ramani, R. P. Srivastava, K. Kumar, and R. V. Nair, "Ai-powered anomaly and threat detection for surveillance footage analysis," in *2024 8th International Conference on Inventive Systems and Control (ICISC)*. IEEE, 2024, pp. 240–247.

[3] F. Nilsson *et al.*, *Intelligent network video: Understanding modern video surveillance systems*. crc Press, 2023.

[4] O. Yakovleva, A. Kovtunenko, V. Liubchenko, V. Honcharenko, and O. Kobylin, "Face detection for video surveillance-based security system." in *COLINS (3)*, 2023, pp. 69–86.

[5] B. R. Ardabili, A. D. Pazho, G. A. Noghre, C. Neff, S. D. Bhaskararayuni, A. Ravindran, S. Reid, and H. Tabkhi, "Understanding policy and technical aspects of ai-enabled smart video surveillance to address public safety," *Computational Urban Science*, vol. 3, no. 1, p. 21, 2023.

[6] A. Shabbir, N. Arshad, S. Rahman, M. A. Sayem, and F. Chowdhury, "Analyzing surveillance videos in real-time using ai-powered deep learning techniques," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 12, no. 2, pp. 950–960, 2024.

[7] P. Dhar and D. P. Chowdhury, "Source camera identification for securing ai-enabled surveillance cameras," in *2024 2nd International Conference on Advancements and Key Challenges in Green Energy and Computing (AKGEC)*. IEEE, 2024, pp. 1–6.

[8] D. Asuquo, I. J. Udo, M. Ekpenyong, and K. Attai, "Ai-enabled iot framework for smart traffic surveillance and communication," in *International Conference on Intelligent and Fuzzy Systems*. Springer, 2024, pp. 193–202.

[9] H.-B. Kim, N. Choi, H.-J. Kwon, and H. Kim, "Surveillance system for real-time high-precision recognition of criminal faces from wild videos," *IEEE Access*, vol. 11, pp. 56 066–56 082, 2023.

[10] Y. Manu and M. Hemchandru, "Surveillance system for real time high precision recognition of criminal," 2024.

[11] O. K. Pal, M. Shovon, M. Mridha, and J. Shin, "In-depth review of ai-enabled unmanned aerial vehicles: trends, vision, and challenges," *Discover Artificial Intelligence*, vol. 4, no. 1, pp. 1–24, 2024.

[12] P. A. D. S. Y. RAMU and K. B. REDDY, "Ai enabled system with real time monitoring of public surveillance videos for abnormality detection and notification," *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 9, 2024.

[13] V. Tsakanikas and T. Dagiuklas, "Video surveillance systems-current status and future trends," *Computers & Electrical Engineering*, vol. 70, pp. 736–753, 2018.

[14] M. Kashef, A. Visvizi, and O. Troisi, "Smart city as a smart service system: Human-computer interaction and smart city surveillance systems," *Computers in Human Behavior*, vol. 124, p. 106923, 2021.

[15] O. Elharrouss, N. Almaadeed, and S. Al-Maadeed, "A review of video surveillance systems," *Journal of Visual Communication and Image Representation*, vol. 77, p. 103116, 2021.

[16] C. Fontes, E. Hohma, C. C. Corrigan, and C. Lütge, "Ai-powered public surveillance systems: why we (might) need them and how we want them," *Technology in Society*, vol. 71, p. 102137, 2022.

[17] R. Sharma, A. Sungheetha *et al.*, "An efficient dimension reduction based fusion of cnn and svm model for detection of abnormal incident in video surveillance," *Journal of Soft Computing Paradigm (JSCP)*, vol. 3, no. 02, pp. 55–69, 2021.

[18] M. M. Mukto, M. Hasan, M. M. Al Mahmud, I. Haque, M. A. Ahmed, T. Jabid, M. S. Ali, M. R. A. Rashid, M. M. Islam, and M. Islam, "Design of a real-time crime monitoring system using deep learning techniques," *Intelligent Systems with Applications*, vol. 21, p. 200311, 2024.

[19] M. J. Iqbal, M. M. Iqbal, I. Ahmad, M. O. Alassafi, A. S. Alfakeeh, and A. Alhomoud, "Real-time surveillance using deep learning," *Security and Communication Networks*, vol. 2021, no. 1, p. 6184756, 2021.

[20] S. Ravindran and G. Aghila, "A fog-assisted framework for intelligent video preprocessing in cloud-based video surveillance as a service," *IEEE Transactions on Sustainable Computing*, vol. 7, no. 4, pp. 825–838, 2022.

[21] R. Archana and P. E. Jeevaraj, "Deep learning models for digital image processing: a review," *Artificial Intelligence Review*, vol. 57, no. 1, p. 11, 2024.

[22] M. Ju and X. Yu, "Semantic-aware low-light image enhancement network for recognizing semantics in intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, 2025.

[23] M. Nallasivam and V. Senniappan, "Moving human target detection and tracking in video frames," *Studies in informatics and control*, vol. 30, no. 1, pp. 119–129, 2021.

[24] Y. Guo, Z. Liu, H. Luo, H. Pu, and J. Tan, "Multi-person multi-camera tracking for live stream videos based on improved motion model and matching cascade," *Neurocomputing*, vol. 492, pp. 561–571, 2022.

[25] J. Wang, F. Hu, G. Abbas, M. Albekairi, and N. Rashid, "Enhancing image categorization with the quantized object recognition model in surveillance systems," *Expert Systems with Applications*, vol. 238, p. 122240, 2024.

[26] A. Sarda, S. Dixit, and A. Bhan, "Object detection for autonomous driving using yolo [you only look once] algorithm," in *2021 Third international conference on intelligent communication technologies and virtual mobile networks (ICICV)*. IEEE, 2021, pp. 1370–1374.

[27] S. Kukreti, R. Praveen, S. Bansal, H. Raju, N. Singh, and R. Begum, "Object detection in real-time surveillance using deep learning-based yolo framework," in *2025 International Conference on Computational, Communication and Information Technology (ICCCIT)*. IEEE, 2025, pp. 601–606.

[28] A. Nazir and M. A. Wani, "You only look once-object detection models: a review," in *2023 10th International conference on computing for sustainable global development (INDIACom)*. IEEE, 2023, pp. 1088–1095.

[29] Z. Dahirou and M. Zheng, "Motion detection and object detection: Yolo (you only look once)," in *2021 7th Annual International Conference on Network and Information Systems for Computers (ICNISC)*. IEEE, 2021, pp. 250–257.

[30] K. Akshatha, A. K. Karunakar, S. B. Shenoy, A. K. Pai, N. H. Nagaraj, and S. S. Rohatgi, "Human detection in aerial thermal images using faster r-cnn and ssd algorithms," *Electronics*, vol. 11, no. 7, p. 1151, 2022.

[31] D. D. Aboyomi and C. Daniel, "A comparative analysis of modern object detection algorithms: Yolo vs. ssd vs. faster r-cnn," *ITEJ (Information Technology Engineering Journals)*, vol. 8, no. 2, pp. 96–106, 2023.

[32] T. Palwankar and K. Kothari, "Real time object detection using ssd and mobilenet," *Int. J. Res. Appl. Sci. Eng. Technol*, vol. 10, pp. 831–834, 2022.

[33] F. R. Chandra, H. R. Ngemba, O. A. Hamid, N. T. Lapatta, S. Hendra, D. W. Nugraha, S. Syahrullah *et al.*, "Analysis of the use of mtcnn and landmark technology to improve the accuracy of facial recognition on official documents," *Journal of Applied Informatics and Computing*, vol. 9, no. 1, pp. 16–22, 2025.

[34] S. Singh *et al.*, “Incorporating mtcnn and deepface: A novel hybrid deep learning method for human face detection and recognition.”
[35] B. Liu and H. Yu, “A lightweight and accurate face detection algorithm based on retinaface,” *arXiv preprint arXiv:2308.04340*, 2023.
[36] A. Ramachandra, R. Tomer, Y. Swathi, P. Kadri, and H. Jagadeesh, “Face recognition and detection using cnn and cascade detection,” in *2024 International Conference on Recent Advances in Science and Engineering Technology (ICRASET)*. IEEE, 2024, pp. 1–7.
[37] S. Xue, “Facial recognition for surveillance videos based on retinaface and mobilefacenet,” in *The International Conference Optoelectronic Information and Optical Engineering (OIOE2024)*, vol. 13513. SPIE, 2025, pp. 878–883.
[38] E. Wahab, W. Shafique, H. Amir, S. Javed, and M. Marouf, “Robust face detection and identification under occlusion using mtcnn and resnet50,” *Sukkur IBA Journal of Emerging Technologies*, vol. 7, no. 2, pp. 100–112, 2024.
[39] G. F. Ananda, H. A. Nugroho, and I. Ardiyanto, “Comparative analysis of multi-face detection methods in classroom environments: Haar cascade, mtcnn, yoloface, and retinaface,” in *2024 Seventh International Conference on Vocational Education and Electrical Engineering (ICVEE)*. IEEE, 2024, pp. 268–273.
[40] Z. Xiao, X. Gao, C. Fu, Y. Dong, W. Gao, X. Zhang, J. Zhou, and J. Zhu, “Improving transferability of adversarial patches on face recognition with generative models,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 11 845–11 854.
[41] A. Firmansyah, T. F. Kusumasari, and E. N. Alam, “Comparison of face recognition accuracy of arcface, facenet and facenet512 models on deepface framework,” in *2023 International conference on computer science, information technology and engineering (ICCoSITE)*. IEEE, 2023, pp. 535–539.
[42] M. G. PELOSO, “Development of a deep learning model with a novel loss function for facial recognition based on arcface.”
[43] R. Jin, H. Li, J. Pan, W. Ma, and J. Lin, “Face recognition based on mtcnn and facenet,” in *AAAI Conference on Artificial Intelligence*, 2021.
[44] J. Diez-Tomillo, J. M. Alcaraz-Calero, and Q. Wang, “Face verification algorithms for uav applications: An empirical comparative analysis,” *Journal of Communications Software and Systems*, vol. 20, no. 1, pp. 1–12, 2024.
[45] A. Laoprakiang and S. Watanapa, “A comparative study of face recognition algorithms for investigating watchlist,” in *2024 8th International Conference on Information Technology (InCIT)*. IEEE, 2024, pp. 331–336.
[46] S. Park, H. Na, and D. Choi, “Verifiable facial de-identification in video surveillance,” *IEEE Access*, 2024.
[47] H. Deng, Z. Feng, G. Qian, X. Lv, H. Li, and G. Li, “Mfcosface: a masked-face recognition algorithm based on large margin cosine loss,” *Applied sciences*, vol. 11, no. 16, p. 7310, 2021.
[48] J. Chi, C. K. On, H. Zhang, and S. S. Chai, “A review of deep convolutional neural networks in mobile face recognition.” *International Journal of Interactive Mobile Technologies*, vol. 17, no. 23, 2023.
[49] M. F. B. A. Rahman, “Smart cctvs for secure cities: Potentials and challenges,” 2017.
[50] A. Mosaif and S. Rakrak, “A new system for real-time video surveillance in smart cities based on wireless visual sensor networks and fog computing.” *J. Commun.*, vol. 16, no. 5, pp. 175–184, 2021.
[51] Y. Myagmar-Ochir and W. Kim, “A survey of video surveillance systems in smart city,” *Electronics*, vol. 12, no. 17, p. 3567, 2023.
[52] N. Khan and M. Efthymiou, “The use of biometric technology at airports: The case of customs and border protection (cbp),” *International Journal of Information Management Data Insights*, vol. 1, no. 2, p. 100049, 2021.
[53] D. T. Viet, P. Van Thien, N. H. Tu, H. G. Minh, and N.-T. Bui, “Design of a face recognition technique based mtcnn and arcface,” in *International Conference on Material, Machines and Methods for Sustainable Development*. Springer, 2022, pp. 71–78.
[54] M. Tahir, Y. Qiao, N. Kanwal, B. Lee, and M. N. Asghar, “Real-time event-driven road traffic monitoring system using cctv video analytics,” *IEEE Access*, vol. 11, pp. 139 097–139 111, 2023.
[55] M. Wiesenthal, “The ethical implications of ai-based mass surveillance tools,” Master’s thesis, 2022.
[56] M. Shah and N. Sureja, “A comprehensive review of bias in deep learning models: Methods, impacts, and future directions,” *Archives of Computational Methods in Engineering*, vol. 32, no. 1, pp. 255–267, 2025.
[57] G. B. Mensah, “Artificial intelligence and ethics: a comprehensive review of bias mitigation, transparency, and accountability in ai systems,” *Preprint, November*, vol. 10, no. 1, 2023.
[58] H. Jeon, H. Kim, D. Kim, and J. Kim, “Pass-cctv: Proactive anomaly surveillance system for cctv footage analysis in adverse environmental conditions,” *Expert Systems with Applications*, p. 124391, 2024.
[59] E. Ozen, F. Alim, S. B. Okcu, E. Kavakli, and C. Cigla, “Real-time face recognition system at the edge,” in *Signal Processing, Sensor/Information Fusion, and Target Recognition XXXIII*, vol. 13057. SPIE, 2024, pp. 243–256.
[60] N. M. Karie, N. M. Sahri, W. Yang, C. Valli, and V. R. Kebande, “A review of security standards and frameworks for iot-based smart environments,” *IEEe Access*, vol. 9, pp. 121 975–121 995, 2021.
[61] H. Sharma and N. Kanwal, “Video surveillance in smart cities: current status, challenges & future directions,” *Multimedia Tools and Applications*, pp. 1–46, 2024.
[62] D. Bhagat, A. Hanwate, R. V. Salunkhe, T. Jagyasi, P. Sardare, and M. Sahu, “Future perspectives on surveillance systems,” in *Modern Advancements in Surveillance Systems and Technologies*. IGI Global Scientific Publishing, 2025, pp. 349–370.
[63] R. Vijeikis, V. Raudonis, and G. Dervinis, “Towards automated surveillance: a review of intelligent video surveillance,” in *Intelligent Computing: Proceedings of the 2021 Computing Conference, Volume 3*. Springer, 2021, pp. 784–803.