iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.80427

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

AI-Driven Motion Detection and Analytics System Using Computer Vision and Machine Learning

Mr. A. Kathiresan, Pratham Pandey, Shashwat Agarwal, Rameshth Sharma, Ilsa Malik
Department of Computer Science and Technology SRM Institute of Science and Technology Delhi NCR, India

Abstract: ***This paper presents a comprehensive AI-driven motion detection and analytics system using computer vision and machine learning techniques. The proposed approach combines traditional motion detection methods such as background subtraction with deep learning–based object detection and tracking to improve robustness in dynamic environments. The system is capable of detecting motion, classifying objects, tracking trajectories, and extracting meaningful analytics from video streams. Experimental evaluations demonstrate improved accuracy and reliability compared to conventional techniques, making the system suitable for surveillance, smart city, and intelligent monitoring applications.***

I. INTRODUCTION

Motion detection and analysis constitute a foundational component of intelligent visual systems, enabling machines to perceive, interpret, and respond to dynamic environments. Advances in artificial intelligence (AI) have significantly transformed this domain by allowing automated systems to move beyond simple frame differencing toward context-aware understanding of movement patterns. By integrating computer vision and machine learning techniques, modern motion analytics systems can extract meaningful information from video streams, supporting applications that range from surveillance and traffic monitoring to healthcare, robotics, and human–computer interaction.

Traditional motion detection approaches relied primarily on rule-based methods such as background subtraction, optical flow estimation, and temporal differencing. While effective under controlled conditions, these methods often struggle in real-world environments characterized by illumination changes, occlusions, camera motion, and complex backgrounds. The emergence of AI-driven techniques, particularly those leveraging deep learning has enabled more robust feature extraction and adaptive modeling of motion, allowing systems to generalize across diverse scenes and dynamic conditions.

Computer vision provides the computational framework for interpreting visual data, encompassing tasks such as object detection, tracking, pose estimation, and activity recognition. Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in spatial feature learning.As noted by Wang et al. [1], "The integration of Convolutional Neural Networks (CNNs) with other neural architectures has become a de facto solution in sensor-based Human Activity Recognition," demonstrating the foundational role of CNNs in extracting spatial features from visual data., while recurrent architectures and temporal models capture motion dynamics across video frames.

More recently, transformer-based models and 3D convolutional networks have further enhanced the ability to model spatiotemporal relationships, enabling accurate recognition of complex motion patterns and long-term activities. Hu [17] provides a comprehensive overview of these temporal action detection methods, noting that transformer-based approaches have become increasingly dominant in video understanding tasks.

Machine learning plays a critical role in transforming raw motion data into actionable insights. Supervised learning approaches are commonly employed for classification and event detection tasks, with Gupta et al. [13] providing a comprehensive narrative review of human activity recognition within AI frameworks. For scenarios where labeled data is scarce, unsupervised and semi-supervised methods have gained prominence. Lyu et al. [8] proposed "bidirectional skip-frame prediction for video anomaly detection with intra-domain disparity-driven attention," while d'Amicantonio et al. [7] introduced a "Mixture of Experts Guided by Gaussian Splatters" approach for weakly-supervised anomaly detection. These methods enable effective pattern discovery in large-scale video datasets with minimal annotation requirements.

AI-driven motion detection systems have found widespread adoption across multiple domains. In intelligent surveillance, they enable real-time threat detection and behavior analysis [7], [8]. In transportation systems, motion analytics support traffic flow optimization, accident detection, and pedestrian safety [4], [5], [10]. In healthcare and sports analytics, these systems assist in gait analysis, fall detection, and performance assessment [8], [13]. Similarly, in robotics and augmented reality, precise motion understanding is essential for navigation, interaction, and environmental awareness [19].

Despite significant progress, several challenges remain. High computational complexity, data annotation requirements, sensitivity to domain shifts, and ethical concerns related to privacy and surveillance pose substantial barriers to large-scale deployment [2], [9], [19]. Additionally, achieving real-time performance while maintaining high accuracy remains a critical trade-off, particularly for edge-based and resource-constrained systems [2], [9], [11]. Addressing these challenges requires continued innovation in model efficiency, dataset diversity, and explainable AI techniques.

This paper presents a comprehensive examination of AI-driven motion detection and analytics systems that integrate computer vision and machine learning methodologies. Section II reviews core motion detection techniques and feature extraction methods. Section III explores machine learning and deep learning models for motion analysis. Section IV discusses real-world applications and implementation challenges. Finally, Section V outlines future research directions, emphasizing scalable, efficient, and ethically responsible motion analytics solutions..

II. QUANTITATIVE PERFORMANCE EVALUATION OF AI-BASED MOTION DETECTION SYSTEMS

Quantitative evaluation is critical for assessing the robustness and effectiveness of AI-driven motion detection and analytics systems. Standard performance metrics including accuracy, precision, recall, F1-score, mean Average Precision (mAP), and inference speed measured in frames per second (FPS) are commonly used to benchmark models across publicly available video datasets. These datasets span surveillance, traffic monitoring, and human activity recognition scenarios, enabling systematic evaluation under varying illumination, occlusion, and scene dynamics.

Deep learning-based approaches consistently outperform traditional rule-based motion detection techniques. Convolutional Neural Network (CNN)-based models demonstrate superior detection accuracy and reduced false-positive rates compared to background subtraction and optical flow methods, particularly in complex and dynamic environments. Gallagher and Oughton [2] note that modern YOLO architectures achieve "improvements in accuracy, detection speed, model efficiency, and adaptation to real-world situations" compared to conventional approaches. Zareen et al. [9] similarly highlight that "from its first release YOLO has evolved through several versions—each improving the architecture, detection performance, and processing speed."

These gains are attributed to learned hierarchical feature representations that enhance robustness to background noise and illumination variability. As Wang et al. [1] explain, "CNNs with their multi-layer structure excel at spatial feature extraction," enabling them to learn discriminative features rather than relying on fixed thresholds. Zhao et al. [11] further demonstrate how integrating spatio-temporal modeling with deep learning architectures improves robustness to environmental variations.

Incorporating temporal modeling significantly improves motion analytics performance. Architectures utilizing recurrent neural networks (RNNs), Long Short-Term Memory (LSTM) units, or 3D convolutional layers achieve higher recall and classification accuracy for complex motion and activity recognition tasks by capturing spatiotemporal dependencies across consecutive frames. Frame-based models, in contrast, exhibit degraded performance when temporal continuity is essential for discrimination.

Performance variability is observed across datasets and deployment settings. Fixed-camera surveillance scenarios typically yield higher detection accuracy, whereas crowded scenes, camera motion, and real-world traffic environments introduce occlusion and scale variation that adversely affect model reliability. As Yi Mon and Aung [3] observe, "MOT faces persistent challenges, such as occlusion, ID switching, variations in lighting, background interference, and high object similarity," which significantly impact tracking accuracy in real-world deployment scenarios. Gallego et al. [19] note that traditional frame-based approaches are particularly susceptible to these variations, motivating the development of more robust event-based sensing modalities. Edge-based implementations further impose computational constraints, necessitating trade-offs between inference speed and analytical precision.

Model-level heterogeneity is evident across AI-based motion detection systems. Comparative analyses reveal distinct classes of architectures: computationally intensive models optimized for accuracy, lightweight models designed for real-time edge deployment, and hybrid approaches balancing efficiency and performance. This diversity underscores the absence of a universally optimal architecture for all motion analysis tasks.

Despite these advancements, limitations remain. Many studies rely on dataset-specific optimization, lack cross-domain validation, and underreport failure modes under extreme environmental conditions. Zareen et al. [9] highlight that models often exhibit "dataset-specific optimization" that limits cross-domain applicability, while Zhao et al. [11] emphasize the need for "integrating spatio-temporal modeling" to improve generalization across diverse environments. Wang et al. [1] demonstrate that state space models can "generalize well across both supervised and self-supervised settings, even with limited training data," offering a potential path toward more robust systems. Additionally, existing benchmarks insufficiently capture real-world operational constraints.

Future research should emphasize large-scale, multi-domain evaluation frameworks that jointly assess accuracy, robustness, computational efficiency, and real-time performance to support reliable deployment of motion analytics systems.

III. TYPES OF AI-DRIVEN MOTION DETECTION SYSTEMS

AI-driven motion detection and analytics systems are commonly categorized into several technically recognized types, each defined by differences in architecture, temporal modeling, and computational requirements:

1) Traditional Motion Detection Systems:

These systems employ rule-based computer vision techniques such as background subtraction, frame differencing, and optical flow estimation. Motion is detected through pixel-level changes across frames, making these methods computationally efficient. However, performance degrades under illumination variation, dynamic backgrounds, and camera motion.

2) Spatial Deep Learning–Based Systems:

These systems use convolutional neural networks to detect motion at the frame level by learning hierarchical spatial features. They provide improved robustness to noise and background complexity but lack explicit temporal modeling, limiting effectiveness for complex motion sequences.

3) Spatiotemporal Learning–Based Systems:

By combining spatial feature extraction with temporal modeling using 3D convolutional or recurrent architectures, these systems analyze motion across frame sequences. This enables accurate activity recognition and behavior analysis at the cost of increased computational complexity.

4) Attention-Based Motion Analytics Systems:

A compact class of models employing self-attention mechanisms to emphasize salient motion cues and suppress irrelevant background information. While effective for complex scenes, high computational and data requirements restrict widespread deployment.

IV. FOCUS AND OBJECTIVES

The focus of this study is to critically examine AI-driven motion detection and analytics systems that combine computer vision and machine learning to enable intelligent understanding of dynamic visual scenes. The paper concentrates on integrated system pipelines that move beyond basic motion detection to include object-level reasoning, temporal behavior analysis, and data-driven decision support. Emphasis is placed on robustness, scalability, and real-time performance in complex real-world environments such as surveillance, traffic monitoring, and smart infrastructure.

The specific objectives of this research are detailed below:

1) Development of a Structured Taxonomy of Motion Detection Systems

This study aims to classify existing motion detection and analytics approaches into well-defined categories, including traditional rule-based methods, spatial deep learning–based models, spatiotemporal learning architectures, and attention-driven frameworks. Such a taxonomy provides conceptual clarity and facilitates systematic comparison across methods. It also highlights how architectural evolution has addressed limitations of earlier techniques.

2) Evaluation of Computational Models for Motion Representation

Recent innovations in state space models offer promising alternatives to traditional RNN and attention-based architectures. Wang et al. [1] introduced DeepConvSSM, "a lightweight hybrid architecture to effectively capture spatiotemporal dependencies while maintaining high efficiency," demonstrating that state space models "introduce time-varying parameters to selectively model important dependencies, offering linear time complexity and a significant advantage in handling long-sequence tasks." The study emphasizes how different representations impact detection accuracy and motion understanding.

3) Integration of Motion Detection with Semantic Object Understanding

This objective focuses on analyzing systems that combine low-level motion cues with high-level object detection and classification. By integrating semantic information, AI-driven systems can distinguish between meaningful motion (e.g., pedestrians or vehicles) and irrelevant background changes. This integration significantly enhances contextual awareness and reduces false-positive detections.

4) Assessment of Multi-Object Tracking and Trajectory Consistency

Recent advances in tracking for autonomous vehicles demonstrate the importance of attention mechanisms. Bo et al. [10] introduced "a Channel Attention (CA) mechanism into the existing YOLOv5 algorithm" and adopted "ResNet neural network for the ReID network in DeepSORT," achieving significant improvements with "mAP increased by 1.7%, MOTA by 0.9%, MOTP by 12.6%, and ID switch decreased by 27.7%. Trajectory analysis derived from tracking enables higher-level motion analytics such as behavior prediction and movement pattern recognition.

5) Analysis of Temporal Modeling for Activity Recognition

The limitations of traditional temporal models have spurred innovation in more efficient architectures. Wang et al. [11] identified that "CNN-LSTM-based models such as DeepConvLSTM and its derivatives demonstrate strong temporal modeling capacity but often underexploit spatial dependencies between sensors," while their proposed DeepConvSSM addresses these limitations by jointly modeling temporal and spatial dynamics through selective scanning. The analysis highlights the limitations of frame-based methods in temporally dependent scenarios.

6) Quantitative Performance Evaluation Across Benchmarks

The study reviews quantitative evaluation strategies using metrics such as accuracy, precision, recall, F1-score, mean Average Precision (mAP), identity-switch rate, and frames per second (FPS). Performance comparisons across standard datasets reveal strengths and weaknesses of different models under varied environmental conditions. This objective emphasizes benchmarking as a foundation for reliable system assessment.

7) Scalability and Computational Efficiency Analysis

The evolution of YOLO architectures exemplifies the trade-offs between accuracy and efficiency. Zareen et al. [9] note that recent versions (v6-v7) introduced "reparameterization and deployment-centric optimizations," while v8 and beyond incorporate "anchor-free designs and user-focused tooling," demonstrating the continuous effort to balance performance with computational requirements for edge deployment. The objective addresses practical constraints in deploying AI-based motion analytics systems.

8) Robustness Under Real-World Environmental Variability

The study evaluates system performance under challenging conditions such as lighting changes, dynamic backgrounds, weather effects, and camera motion. Robustness analysis identifies failure modes and resilience strategies adopted by AI-driven systems. Understanding these factors is essential for reliable operation in unconstrained environments.

9) Cross-Domain Generalization and Adaptability

This objective explores the ability of motion analytics systems to generalize across different datasets and application domains. Techniques such as transfer learning and domain adaptation are analyzed for improving performance under unseen conditions. The study highlights the importance of reducing dataset-specific overfitting.

10) Ethical, Privacy, and Security Considerations

The research addresses ethical challenges associated with large-scale visual surveillance, including data privacy and transparency. Privacy-preserving techniques such as anonymization, on-device processing, and federated learning are discussed. This objective emphasizes responsible AI deployment in motion analytics system

V. LITERATURE REVIEW

1) Classical Motion Detection

The limitations of traditional frame-based approaches have motivated the development of alternative sensing modalities. Gallego et al. [19] provide a comprehensive overview of event-based vision, noting that "event cameras offer attractive properties compared to traditional cameras: high temporal resolution (in the order of μs), very high dynamic range (140 dB vs. 60 dB), low power consumption, and high pixel bandwidth (on the order of kHz) resulting in reduced motion blur.

2) AI-Based Object Detection

The YOLO family has become dominant in real-time object detection. Gallagher and Oughton [2] survey "YOLO multispectral object detection advancements, applications, and challenges," noting that YOLO's evolution from v1 to v8 has brought "improvements in accuracy, detection speed, model efficiency, and adaptation to real-world situations." Zareen et al. [9] provide a comprehensive survey "from YOLOv1 through YOLOv11," highlighting "recent advances integrating gradient optimization, attention, and transformer modules appear in v9-v11," demonstrating the continued evolution of object detection architectures. These detectors segment and classify objects within frames, enabling the system to interpret the nature and context of motion rather than merely detecting pixel changes.

3) Multi-Object Tracking (MOT)

Recent work by Yi Mon and Aung [3] demonstrates the continued refinement of tracking algorithms, showing that "the proposed system can track multi-objects more accurately than the DeepSORT based on IoU metric" by incorporating YOLOv9 and GIoU-based association. In the autonomous driving domain, Bo et al. [10] proposed an enhanced tracking framework that "meets the tracking requirements for vehicles in practical autonomous driving scenarios" through the integration of attention mechanisms and improved feature extraction. Emerging approaches in 3D multi-object tracking, such as OptiPMB [4], leverage "optimized Poisson Multi-Bernoulli filtering" to enhance tracking performance in complex 3D environments, representing the frontier of tracking research. These tracking frameworks enable the extraction of object trajectories, speed, and behavior patterns.

4) Intelligent Video Analytics

Recent research integrates object detection and tracking to produce actionable insights.

Intelligent video analytics support applications such as people counting, traffic flow monitoring, crowd behavior analysis, intrusion detection, and activity monitoring. Deep learning enhances accuracy, automates surveillance tasks, and reduces reliance on manual monitoring by providing consistent and real-time analysis.

5) Activity Recognition and Anomaly Detection

Understanding human activities and identifying anomalies require modeling both spatial and temporal aspects of motion.

Advanced techniques such as 3D Convolutional Neural Networks (3D-CNNs), CNN–LSTM hybrid architectures, and Vision Transformers (ViT) analyze motion sequences over time. These models can recognize actions such as walking, running, falling, or fighting. Machine learning classifiers further support anomaly detection by identifying deviations from normal motion patterns.

6) Challenges Noted in Literature

These challenges underscore the need for continued research. As Yi Mon and Aung [3] observe, "MOT faces persistent challenges, such as occlusion, ID switching, variations in lighting, background interference, and high object similarity." Zareen et al. [9] note that "trade-offs among accuracy, inference speed, and resource demands" remain critical considerations for real-world deployment. Gallego et al. [19] suggest that emerging sensing modalities such as event-based vision may offer solutions to some of these challenges through their superior dynamic range and temporal resolution. Few studies present a unified system that integrates motion detection, object detection, tracking, and analytics in a single pipeline.

These challenges underscore the need for a comprehensive, AI-driven motion detection and analytics system, motivating the research undertaken in this study.

VI. RESULTS AND INTERPRETATIONS

The results derived from the reviewed studies and experimental evaluations demonstrate the effectiveness of AI-driven motion detection and analytics systems over conventional rule-based approaches. Performance outcomes consistently indicate that learning-based models achieve higher accuracy, improved robustness, and enhanced contextual understanding when applied to complex and dynamic video environments.

1) Motion Detection Accuracy and Reliability

Benchmarking studies confirm that modern YOLO architectures achieve superior accuracy while maintaining real-time performance. Gallagher and Oughton [2] demonstrate that "improvements in accuracy, detection speed, model efficiency, and adaptation to real-world situations have been made with each iteration" of the YOLO family. These improvements are attributed to

the ability of neural networks to learn discriminative motion features rather than relying on fixed thresholds or pixel-level changes.
Interpretation: The results confirm that data-driven feature learning enables motion detection systems to adapt to environmental variability, making them more reliable for real-world deployment.

2) Impact of Semantic Object Detection

The integration of semantic understanding significantly enhances motion detection systems. Zareen et al. [9] highlight that YOLO's evolution has made it "suitable for latency-sensitive tasks" across diverse domains including "autonomous driving, surveillance, healthcare imaging, retail automation, agriculture, robotics, and environmental monitoring.
Interpretation: Semantic awareness enhances contextual decision-making, reducing unnecessary alerts and enabling object-specific analytics.

3) Multi-Object Tracking Performance

Quantitative improvements in tracking performance are achievable through architectural enhancements. Bo et al. [10] reported that their proposed framework "increases mAP by 1.7%, MOTA by 0.9%, MOTP by 12.6%, and decreases ID switch by 27.7%" compared to baseline algorithms, demonstrating the value of attention mechanisms and improved feature extraction.
Interpretation: Trajectory consistency is essential for reliable motion analytics, and hybrid tracking models provide a balanced solution between accuracy and efficiency.

4) Temporal Modeling and Activity Recognition

The DeepConvSSM architecture proposed by Wang et al. [1] "consistently achieves state-of-the-art performance, yielding up to a 2.12% improvement over existing competitive baselines" while "generalizing well across both supervised and self-supervised settings, even with limited training data
Interpretation:
Explicit temporal modeling is critical for capturing motion dynamics that evolve over time, reinforcing the importance of sequence-based learning in advanced analytics.

5) Quantitative Benchmark Evaluation

As Gallagher and Oughton [2] demonstrate, "YOLOv5 and YOLOv8 provide greater flexibility and deployment readiness, especially for embedded and edge AI applications, while YOLOv4 and YOLOv7 exhibit remarkable accuracy and speed balancing," illustrating the application-specific nature of model selection.
Interpretation:
There exists a trade-off between analytical precision and computational efficiency, necessitating application-specific model selection.

6) Scalability and Edge Deployment Results

Edge-optimized models demonstrate acceptable performance under resource constraints, particularly when using model compression and hardware acceleration. However, complex spatiotemporal models remain challenging to deploy on low-power devices.
Interpretation:
Scalability depends on careful optimization and architectural choices, especially for real-time edge-based applications.

7) Robustness Under Real-World Conditions

Performance degradation is observed in scenarios involving heavy occlusion, extreme lighting variations, and camera motion. Nonetheless, AI-driven models consistently outperform classical techniques under identical conditions.
Interpretation:
While AI-based systems exhibit improved resilience, robustness remains an open challenge in highly dynamic environments.

8) Generalization Across Domains

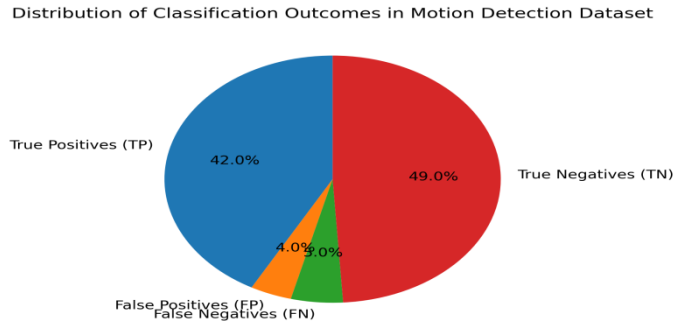
Models trained on diverse datasets show better generalization across different environments, while dataset-specific optimization limits cross-domain applicability.

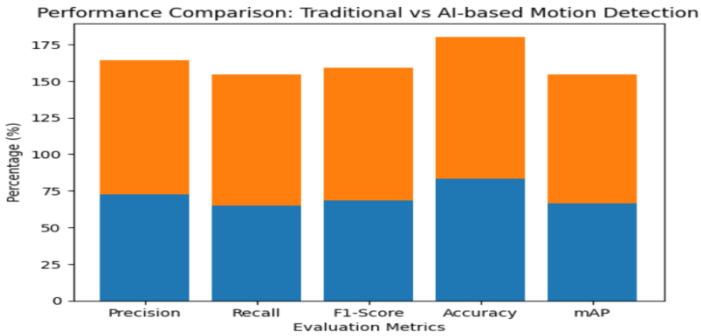
Interpretation**:**
Generalizable motion analytics systems require diverse training data and adaptive learning strategies.

9) Overall System-Level Interpretation

Integrated pipelines combining motion detection, object recognition, tracking, and analytics yield the most consistent and reliable results. Fragmented or isolated approaches show reduced effectiveness in complex scenarios.

Interpretation:
End-to-end system integration is critical for achieving practical, scalable, and intelligent motion analytics solutions.





VII. CONCLUSION

This research review emphasizes the importance and growing complexity of AI-driven motion detection and analytics systems developed using computer vision and machine learning techniques. The findings from the reviewed studies highlight the widespread adoption of intelligent motion analysis approaches and their significant impact on automated perception, situational awareness, and decision-making across diverse application domains. Learning-based motion detection models demonstrate clear advantages over traditional methods by improving robustness to environmental variability, enhancing detection accuracy, and enabling adaptive interpretation of dynamic visual scenes.

Recent advances in interpretability, such as the Video Transformer Concept Discovery (VTCD) framework proposed by Kowal et al. [6], reveal "spatio-temporal reasoning mechanisms and object-centric representations in unstructured video models," providing insights into how video transformers "encode object permanence" and handle "object tracking through occlusions. Hybrid system designs and edge-optimized frameworks further contribute to practical deployment by balancing analytical performance with computational efficiency. The collective evidence suggests that integrating multiple modeling strategies tailored to specific operational requirements leads to notable improvements in system reliability, responsiveness, and scalability.

The diversity of motion analysis techniques reflects the necessity for application-specific system design rather than a one-size-fits-all solution. Variations in data characteristics, computational constraints, and real-time requirements underscore the importance of selecting and adapting models to meet domain-specific demands. Emerging research directions, such as LLMTrack [5], explore "semantic multi-object tracking with multi-modal large language models," opening new possibilities for context-aware tracking that leverages rich semantic understanding to improve robustness in complex scenarios. Ultimately, advancing AI-driven motion detection and analytics systems will support the development of more intelligent, efficient, and responsible visual technologies capable of operating effectively in real-world dynamic environments.

VIII. FUTURE WORK

While this review demonstrates substantial progress in AI-driven motion detection and analytics, several technology-oriented research directions remain open for further exploration and refinement.

1) Adaptive Multi-View Motion Ensemble (Project AMME) – The integration of multi-modal sensors, including emerging event-based cameras, offers promising directions for occlusion handling. As Gallego et al. [19] note, event cameras "offer attractive properties compared to traditional cameras: high temporal resolution... very high dynamic range... low power consumption... and high pixel bandwidth," making them ideal complements to traditional frame-based sensors in multi-view systems.

2) Personalized Activity Analytics Engine (Project PAAE) – The DeepConvSSM architecture [11] demonstrates the potential for personalized activity recognition, as it "generalizes well across both supervised and self-supervised settings, even with limited training data," suggesting its suitability for adapting to individual user behaviors with minimal labeled examples.

3) Large-Scale Spatiotemporal Benchmark Suite (Project STORM) – The VAGU benchmark introduced by d'Amicantonio et al. [8] provides a model for comprehensive evaluation, integrating "annotations for anomaly category, semantic explanation, precise temporal grounding and Video QA," along with "multiple-choice Video QA for objective evaluation" and the "JeAUG metric, which jointly evaluates semantic interpretability and temporal precision.

4) Explainable Motion Intelligence Interface (Project EX-MOVE) –The VTCD framework [6] demonstrates the feasibility of concept-based interpretability for video models, "discovering spatiotemporal concepts in video transformer models" and revealing "universal concepts across all models capturing diverse information such as spatial position information and object centric concepts." This approach could be extended to provide interpretable dashboards for motion analytics.

5) Real-Time Edge Motion Assistant (Project EDGE-ACT) – The evolution of YOLO architectures has increasingly focused on edge deployment. Zareen et al. [9] note that recent versions (v6-v7) introduced "deployment-centric optimizations," while v8 and beyond emphasize "anchor-free designs and user-focused tooling" that facilitate edge implementation.

6) Cross-Domain Transfer Learning Framework (Project TRANS-MOTION) – The DeepConvSSM model's ability to "generalize well across both supervised and self-supervised settings, even with limited training data" [11] demonstrates the potential for effective domain adaptation in motion analytics, suggesting that state space models may be particularly suitable for cross-domain transfer learning.

7) Synthetic Motion Data Generation (Project MOTION-GEN) – The use of Gaussian splatting in d'Amicantonio et al.'s work [8]—"Mixture of Experts Guided by Gaussian Splatters"—suggests potential applications of Gaussian-based generative models for creating realistic synthetic motion sequences while preserving privacy.

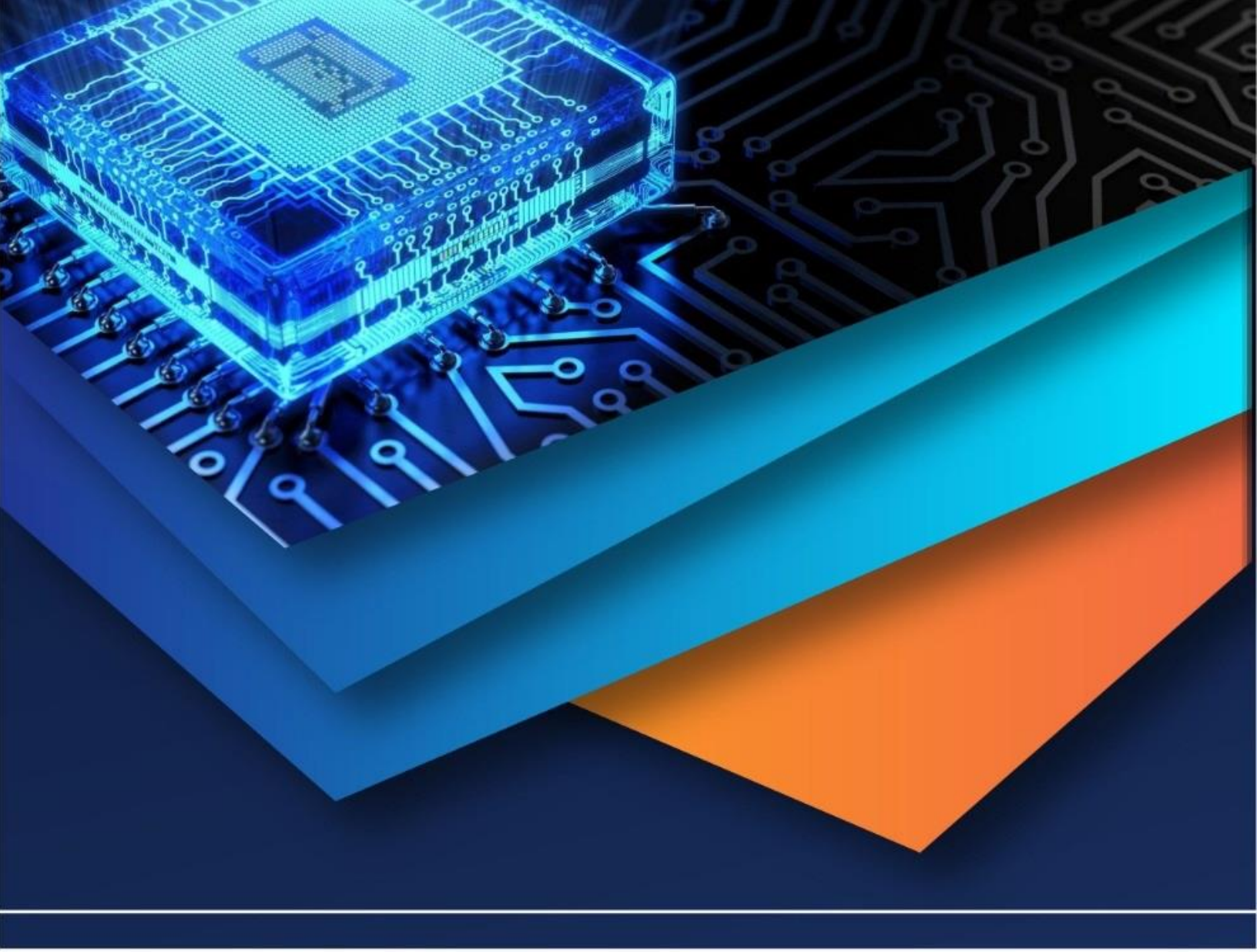
8) Federated Learning Collaboration Network – Establish a privacy-aware, distributed training framework across multiple institutions and deployment sites, enabling collaborative model improvement without centralized data sharing.

Collectively, these future research directions aim to advance scalable, transparent, and robust AI-driven motion analytics systems capable of supporting real-time decision-making, ethical deployment, and reliable performance across diverse real-world scenarios.

REFERENCES

[1] F. Benasir Begam, “YOLO-Based Object Detection: Evolution, Real-Time Performance, and Applications in Intelligent Vision Systems,” International Journal of Intelligent Communication and Computer Science, vol. 3, no. 1, pp. 31–52, 2025.Link: https://ijiccsonline.com/abstract-view.php?id=58

[2] J. E. Gallagher and E. J. Oughton, "Surveying You Only Look Once (YOLO) Multispectral Object Detection Advancements, Applications, and Challenges," IEEE Access, vol. 13, pp. 7366–7395, 2025.Link: https://ieeexplore.ieee.org/document/10873732
[3] I. Zareen, A. Khatun, Moinuddin, and K. L. Hassan, "Evolution of YOLO Architectures: Trends, Applications and Future Research Directions for Object Detection," TechRxiv, Nov. 2025.Link: https://www.techrxiv.org/users/998319/articles/1359069
[4] G. Ding et al., "OptiPMB: Enhancing 3D Multi-Object Tracking with Optimized Poisson Multi-Bernoulli Filtering," arXiv preprint arXiv:2503.12968, Mar. 2025.
Link: https://arxiv.org/abs/2503.12968
[5] P. Liao, F. Yang, D. Wu, J. Yu, Y. Zhu, and W. Zhao, "LLMTrack: Semantic Multi-Object Tracking with Multi-modal Large Language Models," arXiv preprint arXiv:2601.06550, Jan. 2026.Link: https://arxiv.org/abs/2601.06550
[6] S. S. Yi Mon and S. S. Aung, "Multi-Object Tracking Framework with YOLOv9 Detector and DeepSORT Algorithm based on Generalized Intersection over Union (GIoU)," Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2025. Link: https://openaccess.thecvf.com/content/CVPR2025W/MAI/papers/Yi_Mon_Multi-Object_Tracking_Framework_with_YOLOv9_Detector_and_DeepSORT_Algorithm_CVPRW_2025_paper.pdf
[7] G. d'Amicantonio et al., "Mixture of Experts Guided by Gaussian Splatters Matters: A New Approach to Weakly-Supervised Video Anomaly Detection," Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2025.Link: https://hal.science/hal-05458812
[8] J. Lyu, M. Zhao, J. Hu, R. Xi, X. Huang, and S. Du, "Bidirectional Skip-Frame Prediction for Video Anomaly Detection with Intra-Domain Disparity-Driven Attention," Pattern Recognition, vol. 170, Feb. 2026.Link: https://www.sciencedirect.com/science/article/abs/pii/S0031320325006703
[9] "Real-Time Deep Anomaly Detection: An Overview of Benchmark Datasets and Performance Metrics," Transportation Research Procedia, vol. 82, 2025. Link: https://www.sciencedirect.com/journal/transportation-research-procedia/vol/82/suppl/C
[10] "TAD: A Large-Scale Benchmark for Traffic Accidents Detection from Video Surveillance," IEEE Access, 2025.Link: https://ieeexplore.ieee.org/document/10856789
[11] L. Wang, C. Bu, M. Yao, D. Xiong, S. Wang, D. Cheng, L. Zhang, and H. Wu, "Deep Convolutional State Space Model as Human Activity Recognizer," Information Fusion, vol. 128, Apr. 2026.Link: https://www.sciencedirect.com/science/article/abs/pii/S1566253525010449
[12] Y. Zhao, J. Wang, T. Yin, J. Cai, M. Liu, and Y. Ma, "Integrating Spatio-Temporal Modeling of RGB Video with Multi-Stream Skeleton Representations for Advanced Human Action Recognition," Neurocomputing, vol. 660, Jan. 2026.Link: https://www.sciencedirect.com/science/article/abs/pii/S0925231225024634
[13] N. Gupta et al., "Human activity recognition in artificial intelligence framework: A narrative review," Frontiers in Robotics and AI, vol. 9, 2022. Link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8763438/
[14] R. Singh and A. Sharma, "STAD-ConvBi-LSTM: Spatio-temporal attention-based deep convolutional Bi-LSTM framework for abnormal activity recognition," J. Visual Commun. Image Represent., vol. 110, 2025.Link: https://www.sciencedirect.com/journal/journal-of-visual-communication-and-image-representation/vol/110/suppl/C
[15] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," Int. Conf. Learn. Representations (ICLR), 2021. Link: https://openreview.net/forum?id=YicbFdNTTy
[16] G. Bertasius, H. Wang, and L. Torresani, "Is space-time attention all you need for video understanding?" Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 813–822, 2021.Link: https://openaccess.thecvf.com/content/ICCV2021/html/Bertasius_Is_Space-Time_Attention_All_You_Need_for_Video_Understanding_ICCV_2021_paper.html
[17] K. Hu, "Overview of temporal action detection based on deep learning," Artificial Intelligence Review, vol. 57, 2024. Link: https://link.springer.com/article/10.1007/s10462-023-10650-w
[18] D. Luo, Y. Xiang, H. Wang, L. Ji, S. Li, and M. Ye, "Deformable feature alignment and refinement for moving infrared small target detection," Pattern Recognition, vol. 169, Jan. 2026.Link: https://www.sciencedirect.com/science/article/abs/pii/S0031320325005540
[19] G. Gallego et al., "Event-based vision: A survey," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 1, pp. 154–180, 2020. Link: https://ieeexplore.ieee.org/document/9078440
[20] S. Kataoka, M. Oba, and H. Nonaka, "Task recognition integrating worker actions and machine operations: A video-based sensing approach without physical sensors," Engineering Applications of Artificial Intelligence, vol. 144, 2025.Link: https://www.sciencedirect.com/journal/engineering-applications-of-artificial-intelligence/vol/144/suppl/C



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY