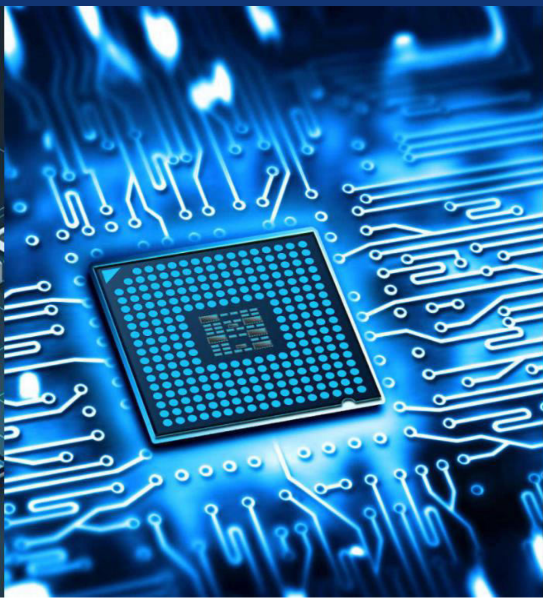
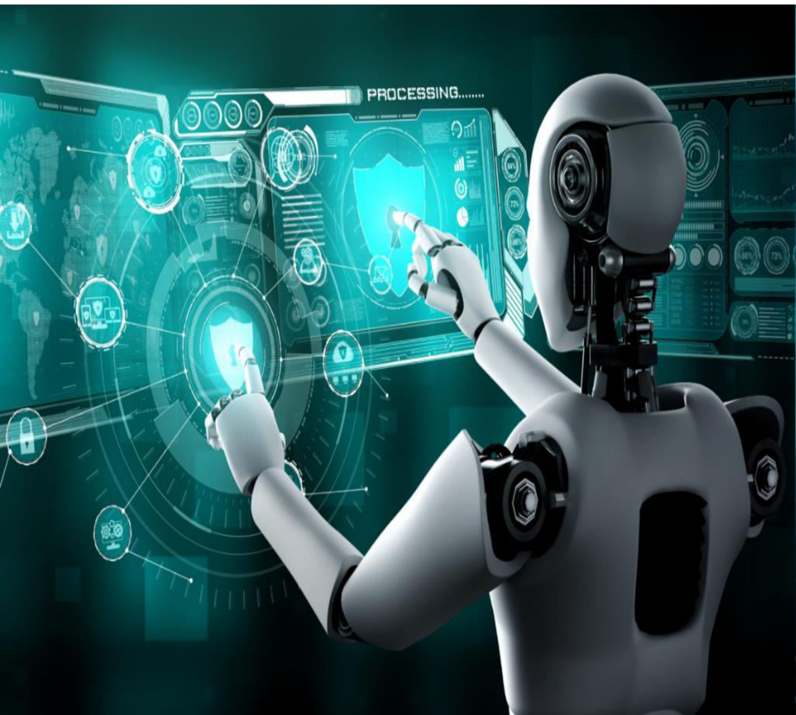
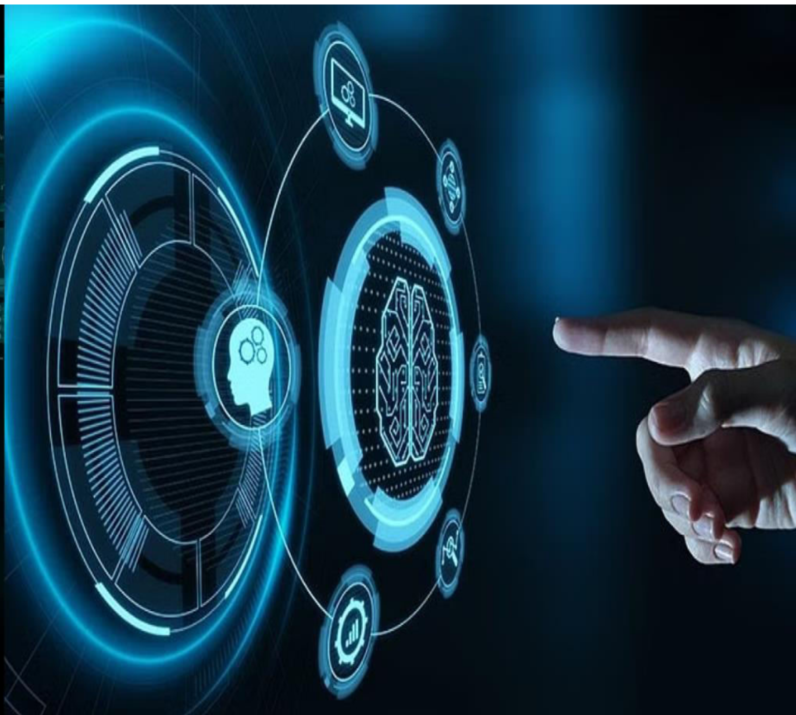
ISSN(O): 2320-9801
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 9.927

Volume 14, Issue 6, June 2026

www.ijircce.com ijircce@gmail.com +91-9940572462 +91 63819 07438

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406054

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

AI-Based Real-Time Multi-Class Abnormal Activity Detection Using Slow-Fast 3D CNN for Smart Surveillance Systems

Om Anerao[1], Sandeep Dharurkar[2], Ajinkya Jadhav[3], Vardhan Jadhav[4], Priyanka Jadhav[5]

Student, Department of Computer Engineering, RMD Sinhgad School of Engineering, Warje, Pune, Maharashtra, India[1-4]

Assistant Professor, Department of Computer Engineering, RMD Sinhgad School of Engineering, Warje, Pune, Maharashtra, India[5]

ABSTRACT: Modern surveillance systems are becoming increasingly important for maintaining safety in public, commercial, and institutional environments. Traditional monitoring methods depend heavily on human operators to continuously observe CCTV footage, which can often be inefficient, time-consuming, and prone to delayed responses. This work presents an AI-based real-time surveillance system designed to detect violence and abnormal activities from live video streams.

The proposed framework utilizes a Slow-Fast 3D Convolutional Neural Network (SF-3D CNN) to capture both spatial and temporal motion features from video sequences. By analyzing video frames at multiple temporal speeds, the model effectively distinguishes abnormal behavior from normal activities in environments such as colleges, hotels, public areas, and smart city infrastructures. Whenever suspicious or violent activity is identified, the system automatically generates alerts to support faster response from the concerned authorities. Experimental results demonstrate that the system can detect violence and abnormal events with good accuracy and efficient real-time performance, making it suitable for practical surveillance applications.

I. INTRODUCTION

The growing number of violent incidents and physical fights in public and private environments has increased the demand for intelligent surveillance systems capable of providing real-time monitoring and rapid response. Traditional CCTV surveillance mainly relies on human operators to continuously observe multiple video feeds, which can often result in fatigue, reduced attention, delayed reactions, and missed incidents. To address these challenges, this research proposes an AI-based smart surveillance system focused on detecting violence and fighting activities automatically from live video streams.

The proposed system combines Slow-Fast Convolutional Neural Networks and 3D CNN-based models to capture both spatial and temporal motion information from video sequences. By analyzing human movements, aggressive interactions, and sudden abnormal actions across multiple frames, the system can effectively differentiate violent behavior from normal day-to-day activities. The framework is suitable for deployment in locations such as colleges, hostels, hotels, public spaces, transportation hubs, and smart city environments where continuous monitoring is essential.

Whenever violent or suspicious fighting activity is detected, the system automatically generates alerts and notifies the concerned authorities for immediate action. For instance, alerts can be sent to college security staff during campus fights or to surveillance personnel in crowded public areas to prevent escalation of violence. The proposed framework aims to reduce dependency on manual monitoring, improve response time, enhance public safety, and provide an efficient real-time surveillance solution for modern security applications.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. LITERATURE REVIEW

The reviewed literature highlights the growing importance of intelligent surveillance systems for detecting violence and abnormal activities in real-time environments. Earlier surveillance approaches mainly relied on handcrafted features and manual monitoring techniques, which often suffered from limited accuracy, delayed response, and difficulty in handling complex environments. Recent studies have shown that deep learning-based approaches, especially 3D Convolutional Neural Networks (3D CNNs), are more effective because they can automatically learn spatial and temporal information directly from video sequences. Research on violence detection using 3D CNN architectures demonstrated improved performance in identifying aggressive human activities while reducing dependency on manual feature extraction methods.

Several researchers further explored Slow-Fast CNN architectures for abnormal activity detection in surveillance videos. These models process video frames at different temporal resolutions, where the slow branch captures long-term contextual information and the fast branch focuses on rapid motion patterns. This dual-path strategy improves the system’s ability to recognize violent interactions, sudden movements, and suspicious behavior in real time. Additional studies integrating attention mechanisms with Slow-Fast networks reported enhanced accuracy by emphasizing important motion regions while suppressing irrelevant background information. Comparative research involving CNN and Transformer-based models also showed that deep learning frameworks provide strong classification performance for surveillance applications, especially in crowded and dynamic environments.

Other research works focused on anomaly detection, object tracking, and behavioral analysis in surveillance systems. Techniques such as YOLO-based object detection, motion trajectory analysis, Dynamic Bayesian Networks, and rule-based activity recognition were used to identify unusual activities and suspicious movement patterns. Researchers also discussed practical challenges including occlusion, environmental noise, changing lighting conditions, and computational limitations in real-world deployments. These studies collectively indicate that combining object detection, tracking, temporal analysis, and spatio-temporal deep learning models can significantly improve the efficiency and reliability of modern smart surveillance systems.

III. PROPOSED METODOLOGY

The proposed system follows a structured real-time surveillance pipeline that integrates human detection, tracking, proximity analysis, temporal buffering, and deep learning-based violence classification into a unified framework. The system is designed to continuously monitor live video streams and automatically identify violent and abnormal human activities with high efficiency and accuracy. The framework supports real-time alert generation and intelligent monitoring for environments such as colleges, hostels, hotels, public places, and smart city infrastructures.

4.1 Video Input Acquisition

- The system accepts live or recorded video streams as input through multiple sources such as webcams, CCTV cameras, IP cameras, or stored video files.
- Frames are continuously captured from the video stream using a stream reader module. The captured frames act as the primary input for further surveillance analysis and real-time activity monitoring.

4.2 Frame Preprocessing

- The captured video frames undergo preprocessing to improve detection quality and system performance.
- The preprocessing stage includes frame resizing, normalization, color conversion, and noise handling. Frames are resized to a fixed resolution suitable for deep learning models, while normalization techniques are applied to maintain consistent input quality.
- These preprocessing steps help improve computational efficiency and ensure stable model performance during real-time execution.

4.3 Human Detection and Tracking

The proposed framework uses the YOLOv8 object detection model to identify human subjects present in each video frame. Bounding boxes are generated around detected individuals, and only person-class detections are considered for further processing. After detection, a tracking module assigns unique IDs to each person and maintains their movement

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

information across consecutive frames. This tracking mechanism helps the system analyze interactions between individuals and monitor suspicious movement patterns over time.

4.4 Proximity Gate Mechanism

To reduce unnecessary computation and improve processing speed, the system implements a proximity gate mechanism. The gate opens only when two or more individuals come sufficiently close to each other, indicating possible interaction or aggressive behavior. If no close interaction is detected, the classifier remains idle. This approach minimizes redundant processing and allows the framework to focus only on potentially suspicious activities.

4.5 Temporal Buffering

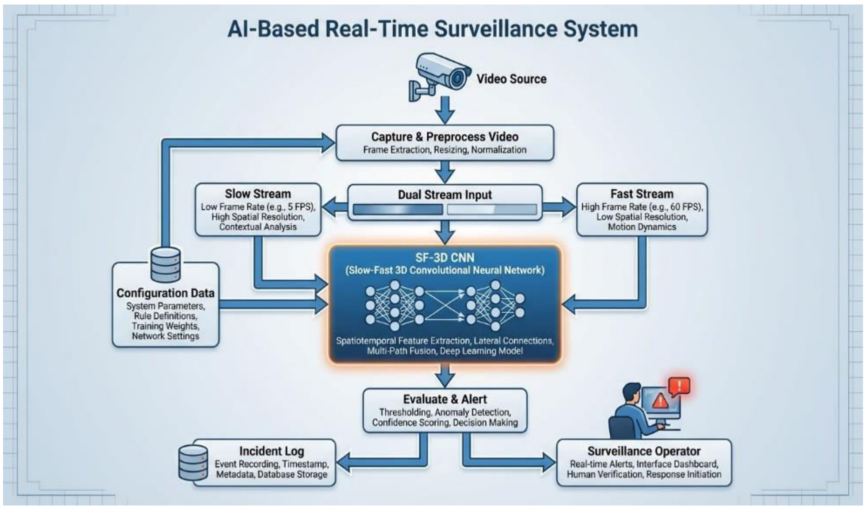
Once the proximity gate is activated, the system stores a sequence of consecutive frames in a temporal buffer. Unlike image-based approaches, violence detection requires understanding motion and interaction across time. Therefore, buffered video clips containing multiple frames are generated and passed to the classification module. This enables the model to capture both spatial appearance and temporal motion information effectively.

4.6 Deep Learning-Based Violence Classification

The proposed system integrates Slow-Fast CNN and 3D CNN-based architectures for violence detection. The default implementation utilizes a Kinetics-based heuristic model capable of recognizing violent actions without requiring additional training. The framework also supports advanced models such as R3D-18 and SlowFast networks for improved spatio-temporal feature extraction. These models analyze motion patterns, aggressive interactions, and abnormal activities from buffered video clips and classify them as violent or non-violent.

4.7 Alert Generation and Monitoring

- Whenever the predicted violence probability exceeds a predefined threshold, the system automatically triggers alerts and records the incident.
- Snapshots and short video clips of the detected event are stored for future analysis. The framework also supports API-based monitoring and dashboard visualization using FastAPI and Streamlit, enabling real-time surveillance monitoring and system status tracking.



4.8 System Architecture

- The proposed framework is implemented using a modular architecture to ensure scalability and flexibility.
- Python is used as the primary programming language for implementing the surveillance pipeline.
- YOLOv8 is used for object detection, while PyTorch-based 3D CNN models are used for activity recognition.
- FastAPI handles backend APIs and Streamlit provides the real-time monitoring dashboard.
- All modules communicate efficiently to maintain smooth real-time processing and automated alert generation.

4.9 Output Generation

- The system generates multiple outputs including
- Real-time violence alerts
- Annotated surveillance frames
- Tracked person IDs
- Confidence scores
- Stored incident clips.
- The dashboard visualizes detected interactions and classifier predictions, while the API server provides access to incident logs and system configuration details. This enables efficient surveillance management and rapid response to violent situations.

IV. EXPERIMENTAL SETUP & IMPLEMENTATION.

The proposed system is implemented as an AI-based real-time surveillance framework designed to automatically detect violence and abnormal human activities from live video streams. The system combines computer vision, object tracking, spatio-temporal deep learning, and automated alert generation into a unified surveillance platform. The implementation is carried out using Python along with libraries and frameworks such as PyTorch, OpenCV, Ultralytics YOLOv8, FastAPI, and Streamlit. The experimental setup includes testing the framework on live webcam feeds, CCTV surveillance videos, and recorded footage under varying conditions such as different lighting environments, camera angles, crowd density, and motion intensity. The system is designed to operate without requiring specialized surveillance hardware, making it suitable for practical deployment in colleges, hotels, public areas, and smart city environments.

The implementation consists of multiple modules including video acquisition, human detection, tracking, proximity analysis, temporal buffering, violence classification, and alert management. YOLOv8 is used to detect human subjects and generate bounding boxes, while the tracking module assigns unique IDs to maintain movement information across frames. A proximity gate mechanism activates the classifier only when multiple individuals come close enough to indicate possible interaction or aggressive behavior, thereby reducing unnecessary computation. For violence recognition, the framework utilizes Slow-Fast CNN and 3D CNN-based architectures capable of learning both spatial and temporal motion patterns from buffered video sequences. The default implementation uses a Kinetics-based heuristic model for zero-shot violence detection, while advanced models such as R3D-18 and SlowFast networks are also supported. Whenever the predicted violence confidence exceeds a predefined threshold, the system automatically generates alerts, stores snapshots and incident clips, and updates the Streamlit dashboard and FastAPI services for real-time monitoring and incident management.

4.1 Algorithms Used

- 1. YOLOv8 Human Detection Algorithm
- Capture video frames from live stream or recorded footage
- Perform image preprocessing and resizing
- Apply YOLOv8 object detection model on each frame
- Detect human subjects and generate bounding boxes
- Filter only person-class detections for surveillance analysis
- Output detected persons with confidence scores and coordinates

- 2. Object Tracking Algorithm
- Assign unique IDs to detected individuals
- Track movement of persons across consecutive frames
- Maintain temporal consistency of detected subjects
- Update track information based on position and motion
- Remove stale or lost tracks after inactivity
- Enable continuous monitoring of human interactions

- 3. Proximity Gate Algorithm
- Compute distance between detected individuals
- Identify groups of people based on spatial proximity
- Activate gate only when multiple persons come close together
- Reduce unnecessary classifier execution during normal activity
- Allow only suspicious interactions to proceed for classification

- 4. Temporal Buffering Algorithm
- Capture consecutive video frames into a temporal buffer
- Store fixed-length frame sequences for activity analysis
- Generate video clips representing motion over time
- Prepare buffered clips for deep learning classification
- Enable extraction of temporal motion features

- 5. Slow-Fast CNN and 3D CNN Violence Classification Algorithm
- Input buffered video clip into the classification model
- Extract spatial and temporal motion features using 3D convolutions
- Analyze aggressive actions and abnormal human interactions
- Use Slow-Fast architecture to process both slow and fast motion patterns
- Predict violence probability score for each video clip
- Classify activity as violent or non-violent based on threshold value

- 6. Kinetics-Based Heuristic Classification Algorithm
- Load pretrained Kinetics-400 action recognition model
- Run inference on buffered surveillance clips
- Extract probabilities of violence-related action classes
- Combine probabilities of actions such as fighting, boxing, and wrestling
- Scale confidence score to estimate violence probability
- Generate zero-shot violence predictions without additional training

- 7. Alert Generation Algorithm
- Compare predicted confidence score with predefined threshold
- Trigger alert if violence probability exceeds threshold
- Store incident snapshots and video clips
- Update monitoring dashboard and incident logs
- Send real-time notifications for detected violent activities

V. RESULTS

The proposed surveillance system was tested on live webcam feeds, CCTV-style videos, and recorded surveillance footage under different environmental conditions such as varying lighting, crowd density, and motion intensity. Experimental results show that the framework successfully detects violent and abnormal human activities in real time using YOLOv8-based human detection, tracking, and deep learning-based violence classification. The proximity gate mechanism effectively reduced unnecessary computations by activating the classifier only during close human interactions, thereby improving processing efficiency and reducing false detections. The system was also able to generate real-time alerts, incident snapshots, and video clips whenever violent activities such as fighting or aggressive behavior were identified.

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406054
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The integration of Slow-Fast CNN and 3D CNN-based models enabled the framework to capture both spatial and temporal motion features from video sequences, improving the accuracy of violence recognition. The Streamlit dashboard and FastAPI services provided real-time monitoring, visualization of tracked individuals, confidence score analysis, and incident management capabilities. The system demonstrated stable performance for practical surveillance applications without requiring specialized surveillance hardware. Overall, the proposed framework provides an efficient and scalable real-time violence detection solution that enhances automated monitoring, reduces manual surveillance effort, and improves public safety in environments such as colleges, hotels, and public spaces.

VI. CONCLUSION

The proposed system presents an AI-based real-time surveillance solution for detecting violence and abnormal human activities using computer vision and deep learning techniques. The framework integrates YOLOv8-based human detection, object tracking, proximity analysis, temporal buffering, and 3D CNN-based violence classification into a unified surveillance platform. By utilizing Slow-Fast CNN and spatio-temporal deep learning models, the system effectively analyzes both spatial and motion-based features from live video streams. The implementation also includes FastAPI services and a Streamlit dashboard for real-time monitoring, incident visualization, and automated alert generation. This integrated approach reduces dependency on manual surveillance and improves the efficiency of monitoring crowded public and institutional environments.

Experimental observations demonstrate that the system is capable of detecting violent and suspicious activities with reliable performance under different real-world conditions such as varying lighting, camera angles, and crowd density. The proximity gate mechanism helps optimize computational efficiency by activating the classifier only during close human interactions, thereby reducing unnecessary processing. The modular architecture ensures scalability, flexibility, and ease of deployment without requiring specialized surveillance hardware. Overall, the proposed framework provides a practical, cost-effective, and intelligent surveillance solution suitable for applications in colleges, hotels, public areas, transportation hubs, and smart city environments, contributing toward improved public safety and automated real-time security monitoring.

VII. FUTURE SCOPE

The proposed surveillance system can be further enhanced in multiple directions to improve its intelligence, scalability, and real-world applicability. Future improvements may include the integration of advanced deep learning architectures for achieving higher violence detection accuracy under crowded and complex environments. The system can also be extended to support multi-camera surveillance, allowing continuous tracking of suspicious individuals across different locations. Additional optimization techniques can be incorporated to improve real-time performance on low-resource devices and edge computing platforms. Furthermore, integrating advanced behavioral analysis and anomaly detection models can help the system identify a wider range of suspicious activities beyond physical violence and fighting.

The framework can also be expanded for smart city and traffic surveillance applications by integrating the system with traffic signal cameras and public monitoring infrastructure. In future implementations, the system can automatically detect road accidents, violent incidents, or emergency situations near traffic junctions and immediately send alerts to the nearest police station, hospital, or emergency response center. GPS-based location sharing, automated emergency notifications, and cloud-based incident management can significantly reduce response time and improve public safety. The integration of mobile applications, IoT devices, and centralized monitoring dashboards can further transform the framework into a fully intelligent real-time public security and emergency response system for smart city environments.

REFERENCES

[1] A. K. Tripathi, S. Mukherjee, and A. K. Dubey, "Toward Fast and Accurate Violence Detection for Automated Video Surveillance Applications," IEEE Access, vol. 11, pp. 38177–38191, 2023.
[2] R. K. Sharma, S. K. Singh, and A. Tiwari, "Remote Monitoring System Using Slow-Fast Deep CNN Model for Identifying Anti-Social Activities," Measurement: Sensors, vol. 25, pp. 100712–100721, 2023.

[3] T. Nguyen, H. Zhang, and M. Abdel-Mottaleb, "Enhancing Situational Awareness: Anomaly Detection Using Real-Time Video Across Multiple Domains," IEEE Access, vol. 13, pp. 50541–50556, 2025.
[4] M. N. Nahar and S. Paul, "Real-Time Violence Detection in Video Using Deep Learning," Multimedia Tools and Applications, vol. 81, no. 14, pp. 20321–20340, 2022.
[5] C. Feichtenhofer, H. Fan, J. Malik, and K. He, "SlowFast Networks for Video Recognition," IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6201–6210, 2019.
[6] C. Feichtenhofer, "X3D: Expanding Architectures for Efficient Video Recognition," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 203–213, 2020.
[7] A. Ullah, K. Muhammad, W. Ding, S. W. Baik, and M. Sajjad, "Violence Detection Using Spatio-Temporal Features with 3D Convolutional Neural Network," Sensors, vol. 19, no. 20, pp. 4470–4484, 2019.
[8] A. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent Flows: Real-Time Detection of Violent Crowd Behavior," IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1–6, 2012.
[9] Y. Tian, G. Lu, and D. Zhang, "Anomaly Detection in Surveillance Videos Using Deep Learning: A Review," IEEE Access, vol. 9, pp. 102363–102391, 2021.
[10] D. Wu, L. Shao, and Y. Fu, "Deep Learning for Video Classification and Captioning: A Review," IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 10, pp. 4159–4177, 2018.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details