Exploring Techniques for Abnormal Event Detection in Video Surveillance

Varsha Anil Kshirsagar, Shou-Chih Lo, Guanling Lee*

Department of Computer Science & Information Engineering, National Dong Hwa University, Taiwan
varsha04kshirsagar@gmail.com, sclo@gms.ndhu.edu.tw, guanling@gms.ndhu.edu.tw

Abstract

This study explores and evaluates the effectiveness of various abnormal event detection techniques in video surveillance, addressing challenges such as intrusions, accidents, and suspicious activities. Through a systematic review of related papers, the study reveals the prevalence of traditional methods like background subtraction and motion detection despite their limitations in complex scenarios. It highlights the increasing use of deep learning techniques, particularly CNNs and RNNs, which show promise but require substantial labeled data. The findings underscore the importance of selecting proper detection techniques based on specific surveillance scenarios and emphasize the need for extensive labeled datasets for deep learning methods. The originality of this study lies in its comprehensive review and comparison of various abnormal event detection techniques, providing valuable insights and practical implications for advancing video surveillance systems.

Keywords: Abnormal event detection, Video surveillance, Object tracking, Machine learning

1 Introduction

Abnormal event detection (AED) in video surveillance is an important area of study that has big effects on public safety and security. People need to watch live feeds or playback recorded footage on traditional video surveillance systems to spot strange or suspicious activity. Modern surveillance systems, on the other hand, produce so much video data that monitoring by hand is both impractical and prone to mistakes. Because of this, automated techniques for finding strange events have been created [1]. These include the application of computer vision and machine learning to look at video streams and find strange events in real-time. One of the biggest challenges in detecting abnormal events is defining what constitutes an abnormal event. Strange things can happen in a lot of different ways depending on the situation and surroundings [2].

This paper addresses AED by providing a systematic review. Several approaches have been proposed for AED in video surveillance [3]. One common method is to use supervised machine learning algorithms, like Naive Bayes, K-nearest neighbors, support vector machines (SVMs), or deep neural networks, to train models with labeled data indicating normal and abnormal events [4]. A major challenge of supervised learning is the need for a vast amount of labeled data, which can be both difficult and expensive to acquire. In contrast, unsupervised learning trains algorithms on unlabeled data to identify patterns and detect anomalies.

Autoencoders are neural networks trained to reconstruct input data and are often used for AED in unsupervised learning. Anomalies are identified when the reconstruction error surpasses a predefined threshold [5]. Unsupervised learning methods are advantageous because they don’t require labeled data, but they may not be as effective at detecting complex or rare abnormal events as supervised methods.

Deep learning techniques, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs) are increasingly popular for AED. Deep learning models are good at finding unusual patterns in video data due to their ability to learn complex spatiotemporal patterns [6]. However, these models typically require substantial computing power and large amounts of training data.

Beyond algorithmic approaches, researchers are exploring the use of multi-modal sensor fusion in AED [7]. By integrating video data with information from other sensors, such as audio, thermal, or motion sensors, surveillance systems can more easily detect events and reduce false alarms. Combining advanced machine learning and computer vision techniques is essential to address the complex challenge of AED. Despite significant progress in recent years, many issues and research questions remain unresolved. Enhancing the reliability and robustness of AED techniques will improve the effectiveness and trustworthiness of video surveillance systems.

In the remainder of this paper, we will present the background knowledge in Section 2 and highlight the most effective event detection methods in Section 3. Section 4 will discuss the application domains, while Section 5 will suggest future research directions. Finally, the concluding remarks will be provided in Section 6.

2 Background Knowledge

Video surveillance technology is utilized in various fields, including healthcare, transportation, and security, primarily to monitor activities, detect anomalies, and ensure safety. It involves the automated detection of abnormal events in video footage.

*Corresponding Author: Shou-Chih Lo; E-mail: sclo@gms.ndhu.edu.tw
DOI: https://doi.org/10.70003/160792642024092505013

There are two primary types of abnormal detections in video data: single-scene and multi-scene. In single-scene detection, the environment remains consistent, meaning the background, lighting conditions, and typical activities do not change much. Conversely, in multi-scene detection, the environment changes significantly across different scenes. The multi-scene detection is more challenging than the single-scene detection.

In video surveillance, various kinds of anomalies must be considered based on the appearance, motion, trajectory, and interaction of objects within a video scene [8]. These anomalies include:

Appearance-only anomalies: These occur when an object appears unusual within a scene.

Motion-only Anomalies: These involve unusual motion of an object within a scene.

Trajectory anomalies: These involve unusual object trajectories within a scene.

Group Anomalies: These occur due to unusual interactions among objects within a scene.

Identifying anomalies based solely on appearance or motion necessitates the analysis of specific local regions within the video. On the other hand, detecting anomalies in trajectories involves examining video segments rather than individual frames. Lastly, identifying group anomalies is distinct from other types, as it requires an analysis of the relationships between different video regions. A graph-based technique, which models object relationships in video frames is commonly used for detecting group anomalies.

AED in video surveillance can be achieved through various methods [9]. A common approach involves using machine learning algorithms that train models to recognize patterns associated with normal and abnormal events. These algorithms analyze video frames to detect atypical changes in motion, appearance, or behavior [10]. The resulting models can then analyze real-time video feeds and issue alerts when anomalies are detected.

Figure 1 illustrates the anomaly detection process utilizing machine learning. During the training phase, raw visual data is first preprocessed, and features are extracted. A model representing normal activity is then developed from the extracted features across one or more video scenes, ensuring it contains no anomalies. In the testing phase, a new video (potentially containing anomalies) from the same scene is preprocessed, and the same features are extracted. These features, combined with the learned model and a trained classifier, are employed to detect anomalies and differentiate between normal and abnormal frames in the test video.

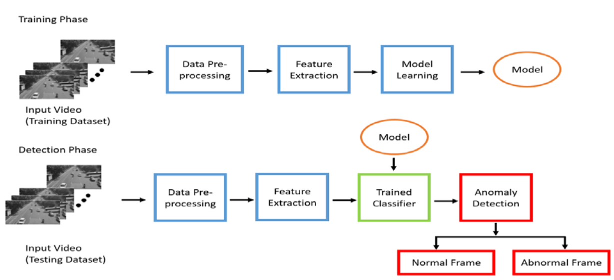

Figure 1. General block diagram of AED using machine learning

Another approach utilizes computer vision techniques, which extract features from video frames to identify anomalies. For example, optical flow algorithms can detect motion patterns, and background subtraction algorithms can identify objects that appear or disappear unexpectedly [11].

Real-life situations are complex and varied, making AED challenging. This requires robust algorithms capable of functioning in diverse settings [12]. Additionally, the quality of input data and the features extracted from video frames significantly impact the performance [13].

While research on AED techniques has advanced considerably, several gaps and challenges remain. A major issue is the lack of standard benchmarks or evaluation metrics for assessing the performance of these algorithms [14]. Another problem is the reliance on controlled environments and datasets in many studies, which may not be applicable in real-world scenarios involving dynamic lighting and obstructions [15].

Moreover, there has been insufficient research on the scalability and efficiency of AED algorithms. Further research is needed to develop scalable algorithms capable of quickly analyzing live video feeds from multiple cameras [16].

Another gap in the research is the limited focus on context-aware AED. Many studies analyze abnormal events in isolation, without considering the broader surveillance environment. Incorporating contextual information, such as the hour of day, geographical location, and lighting conditions, could enhance the accuracy of detection algorithms [17].

3 Abnormal Event Detection: Techniques and Evaluations

Detecting abnormal events in video surveillance employs various methods and algorithms, which can be categorized into traditional computer vision-based and deep learning-based approaches. Traditional computer-vision-based methods use image processing techniques and apply fundamental concepts such as motion detection, background subtraction, object tracking, and appearance-based detection [18]. Motion-based detection, akin to optical flow tracking, examines video motion patterns to identify sudden or unusual movements. Background subtraction identifies moving objects by removing the background from the current frame. Object tracking algorithms monitor objects over time, detecting anomalies when their paths or behaviors change. Appearance-based methods identify sudden changes in appearance, such as unexpected appearances or disappearances.

These traditional methods have their strengths and weaknesses, making them suitable for various scenarios. Motion-based detection is effective for detecting events involving noticeable motion changes and is particularly useful for identifying break-ins or unauthorized access to restricted areas [19].

Object tracking is valuable for monitoring the movement of specific objects over time, making it useful for scenarios where the primary goal is to observe the behavior of individuals or vehicles [20]. Appearance-based methods are versatile and can be applied in various scenarios where anomalies might not involve significant motion changes [21].

Traditional machine learning techniques, such as SVM classifiers, GMM (Gaussian Mixture Model) clustering, K-means clustering, were initially used in AED with hand-crafted features like Histogram of Oriented Gradients (HOG) and Histogram of Optical Flow (HOF) [22]. These methods, like traditional computer vision-based methods, suffer from high computational costs due to the extraction of manual features and are less adaptive to environmental changes.

Deep learning-based methods, utilize neural networks to both extract features and detect anomalies. Despite their advancements, these techniques still face limitations. Conventional approaches often face difficulties in handling intricate situations, such as densely populated settings or fluctuating lighting conditions. Deep learning techniques necessitate a significant amount of labeled data for effective training and often demand considerable computational resources.

CNNs are widely favored for various image-processing tasks because they can automatically detect and learn spatial features. A standard CNN is composed of multiple fundamental components, such as convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification. 3D CNNs are capable of extracting spatiotemporal features from video data, which makes them particularly useful for detecting anomalies [23]. The ability of deep learning to model intricate patterns in data through multiple layers of processing allows these algorithms to surpass traditional methods, delivering higher accuracy and faster processing speeds [51].

RNNs, which have cyclic connections within their neural network architecture, are designed to memorize previous inputs, allowing them to capture temporal information effectively in sequential data. Although RNNs can be trained to process images, they face challenges in distinguishing closely situated contrasting features. To overcome this problem, Long Short-Term Memory (LSTM) networks and bidirectional LSTM are implemented [24].

To reduce the overhead of labelling data, unsupervised learning models are trained using normal events, utilizing concepts such as reconstruction errors and generative models. The autoencoder architecture, which involves an encoder to compress the input data and a decoder to reconstruct the original input data, can be used to detect anomalies by identifying instances with high reconstruction errors.

The attention mechanism is an innovative technique that allows RNNs to concentrate on particular segments of the input sequence while making predictions. One attention-based autoencoder network is proposed in [25] to detect anomalies.

GANs have gained popularity in recent years and have successfully achieved effective results in anomaly detection. A standard GAN consists of two primary parts: a generator that produces synthetic data and a discriminator that distinguishes between genuine and synthetic data. The GAN can learn the data distribution to identify anomalies. Several variants of GANs are applied for AED by incorporating techniques such as self-attention [26], deep spatiotemporal translation [27], and dual discriminators [28].

Both traditional computer vision-based and deep learning-based approaches can generate false positives and may face challenges in real-time processing and scalability. Further research is needed on hybrid approaches that combine the strengths of both computer vision and deep learning methods [29]. Additionally, ensemble learning, which integrates multiple trained models, can significantly enhance the accuracy and robustness of the detection system.

To handle large video datasets, a scalable technique based on weakly supervised learning [30] and multiple-instance learning (MIL) [31] can be employed. Using MIL, video data is segmented into segments (bags). A training model is built by identifying a bag as normal (negative) if all video frames in the bag are normal, and as abnormal (positive) if at least one frame is abnormal.

Numerous publicly accessible datasets are available for training and testing anomaly detection. Researchers select specific datasets and evaluation parameters based on the application, as each dataset yields different results and outcomes. Commonly used evaluation measures include precision, recall, area under the curve (AUC), receiver operating characteristics (ROC), equal error rate (EER), and detection rate (RD).

Some popular datasets are briefly introduced as follows:

UCSD dataset [32]: This comprises various scenarios such as walking on the road and driving on the footpath. The videos capture events from various crowd scenes under different lighting conditions, weather conditions, and pedestrian densities, providing diverse scenarios for researchers. Additionally, the dataset includes both single-pedestrian and multiple-pedestrian scenarios, further increasing the complexity of the detection task.

Avenue dataset [33]: This contains various video sequences captured from a fixed camera overlooking a busy street scene. The videos are labeled with ground truth annotations indicating the location and duration of the anomalies.

ShanghaiTech dataset [34]: This is a crowd-counting dataset that contains images of crowded scenes captured from various locations in Shanghai, China. The crowd scenes in the dataset include various situations such as people walking on the street, waiting in line, or attending crowded events.

Subway dataset [35]: This contains footage of underground train stations from entry and exit points. It captures abnormal behavior such as individuals walking in the wrong direction, lingering, suspicious interactions, and evading turnstiles.

4 Applications and Case Studies

The ability to spot abnormal events in video surveillance has numerous practical applications across various fields and situations. In security, AED is utilized to identify unusual behavior or activities in public places such as hospitals, supermarkets, and subway stations [36]. For instance, AED algorithms can assist security personnel in identifying individuals loitering in restricted areas or acting strangely, potentially indicating a security threat.

In traffic management, AED monitors traffic flow and identifies accidents or traffic jams on highways and roads [37]. These algorithms can automatically notify authorities of accidents or road hazards, improving response times and enhancing traffic flow efficiency. For example, researchers utilized deep learning algorithms to detect abnormal events in traffic flow on a busy highway [38]. The algorithms analyzed video feeds from cameras along the highway to identify accidents, breakdowns, and other anomalies that could impede traffic.

In healthcare, abnormal event detection can monitor patient behavior and detect signs of distress or medical emergencies. Hospital room cameras equipped with detection algorithms can identify abnormal events. A study focuses on monitoring patients in the intensive care unit by analyzing video feeds from patient rooms to detect unusual behaviors or movements that may indicate a medical emergency [39]. The system alerted medical staff to potential issues, enabling prompt intervention and patient care.

In factories, AED monitors machines and equipment for signs of malfunction. Algorithms can analyze video feeds from cameras watching machinery to detect unusual movements or vibrations indicating potential issues [40].

The real-world impacts of AED in video surveillance are substantial, offering significant benefits across multiple fields. Improved security is a major benefit, as algorithms that detect abnormal events allow organizations to identify and address security threats in real-time. This capability helps prevent theft, vandalism, and other crimes, ensuring a safer environment for employees, customers, and the public. Additionally, AED can yield cost savings and improve operational efficiency [41].

Early detection of accidents or breakdowns in transportation can reduce traffic congestion and delays, lowering fuel costs and optimizing traffic flow. In healthcare, early detection of medical emergencies can accelerate response times and improve patient outcomes. Furthermore, AED supports better decision-making by providing organizations with real-time information about abnormal events, enabling more informed resource allocation, emergency management, and operational planning. This can enhance risk management and overall organizational efficiency [42].

The applications of AED are expected to expand as technology advances, making it increasingly valuable across diverse fields and situations.

5 Challenges and Future Directions

Even though current techniques for detecting abnormal events in video surveillance work, they have several problems and limitations. Noise and data variability are significant challenges. Real-world video surveillance data often contain noise from changing lighting, bad weather, and occlusions, making it difficult to accurately detect abnormal events. Techniques that depend on motion or appearance may struggle in these conditions, leading to false alarms or missed detections [43].

Abnormal events occur infrequently, while normal events are much more common. This class imbalance can lead to biased models that perform poorly in AED. To address this, data augmentation techniques like SMOTE, ENN, GANs, or diffusion are often required [44]. Additionally, transfer learning, which leverages knowledge learned from pre-trained models, can help mitigate the problem of limited training data.

Another major issue is accurately interpreting the scene. While algorithms can identify anomalies based on pre-established patterns, they may not fully understand the overall context of the video [45-46]. Interpretability is a common challenge for many AED techniques. Deep learning models can be hard to understand because their decision-making processes are often opaque. This lack of transparency can deter their use, especially in scenarios where openness and accountability are crucial. Addressing these challenges necessitates continuous research and development in the field of AED.

Emerging technologies such as computer vision, machine learning, and artificial intelligence (AI) are revolutionizing the detection of abnormal events in video surveillance. Deep learning techniques, especially CNNs and RNNs, are increasingly used to improve the accuracy and speed of abnormal event detection. These techniques enable computers to learn intricate patterns and relationships in video data, thereby improving their detection capabilities.

Integrating multiple data sources to improve the reliability of AED is another emerging trend. Enhancing multimodal data fusion is an opportunity [47]. Combining video data with sensor data and audio provides algorithms with a richer understanding of the environment, making it easier to detect and interpret anomalies. For example, audio data can add context to video footage, helping differentiate between normal and abnormal events.

Context-aware abnormal event detection [48], which considers the scene's context to improve detection accuracy, receives considerable attention. Elements like the hour of the day, weather conditions, and the presence of other individuals or objects can affect the detection process, thereby minimizing false alarms and improving overall system performance.

Edge computing is another area for future exploration in AED. By processing video data closer to the source, such as on cameras or edge devices, edge computing can reduce latency and bandwidth requirements, facilitating real-time anomaly detection [49]. This can enhance the performance of AED systems, especially when immediate processing is necessary.

Explainable AI, which aims to make AED algorithms more transparent, is another promising trend [50-51]. In addition, incorporating human feedback into AED systems can improve their performance. Allowing users to provide feedback on detected events enables the system to learn and improve over time, making detections more accurate and reliable.

6 Conclusion

This survey paper provides both a broad and insightful view of AED in video surveillance. Machine learning algorithms are frequently utilized to identify patterns in both normal and abnormal events, enabling the detection of unusual movements, appearances, or behaviors. Computer vision is another method used for event detection.

Detecting abnormal events in real-life scenarios is challenging due to constant changes and complexities. Surveillance cameras must differentiate between normal events and anomalies like emergencies or crimes, requiring robust algorithms applicable in various settings.

Future research should explore and develop several critical areas. Investigating the application of new technologies like computer vision, deep learning, and reinforcement learning in AED is essential. These technologies have the capability to greatly enhance the precision and effectiveness of surveillance systems.

Additionally, making detection systems more scalable and adaptable to various surveillance settings and situations is crucial. This might involve developing more flexible and customizable algorithms tailored to different surveillance needs. Moreover, addressing ethical and privacy concerns associated with abnormal event detection systems is vital for future research.

References

[1] R. J. Franklin, Mohana, V. Dabbagol, Anomaly Detection in Videos for Video Surveillance Applications Using Neural Networks, *Inventive Systems and Control (ICISC)*, Coimbatore, India, 2020, pp. 632-637.

[2] F. Meng, G. Yuan, S. Lv, Z. Wang, S. Xia, An Overview on Trajectory Outlier Detection, *Artificial Intelligence Review*, Vol. 52, No. 4, pp. 2437-2456, December, 2019.

[3] J. Ren, F. Xia, Y. Liu, I. Lee, Deep Video Anomaly Detection: Opportunities and Challenges, *Data Mining Workshops (ICDMW)*, Auckland, New Zealand, 2021, pp. 959-966.

[4] D. Xu, E. Ricci, Y. Yan, J. Song, N. Sebe, Learning Deep Representations of Appearance and Motion for Anomalous Event Detection, arXiv:1510.01553, October, 2015, https://arxiv.org/abs/1510.01553v1.

[5] C. Sun, Y. Jia, H. Song, Y. Wu, Adversarial 3D Convolutional Auto-Encoder for Abnormal Event Detection in Videos, *IEEE Transactions on Multimedia*, Vol. 23, pp. 3292-3305, September, 2020.

[6] K. K. Santhosh, D. P. Dogra, P. P. Roy, Anomaly Detection in Road Traffic Using Visual Surveillance: A Survey, *ACM Computing Surveys (CSUR)*, Vol. 53, No. 6, pp. 1-26, November, 2021.

[7] T. Ji, S. T. Vuppala, G. Chowdhary, K. Driggs-Campbell, Multi-Modal Anomaly Detection for Unstructured and Uncertain Environments, *arXiv:2012.08637*, December, 2020, https://arxiv.org/abs/2012.08637.

[8] B. Ramachandra, M. J. Jones, R. R. Vatsavai, A Survey of Single-Scene Video Anomaly Detection, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 44, No. 5, pp. 2293-2312, May, 2022.

[9] H.-T. Duong, V.-T. Le, V. T. Hoang, Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey, *Sensors*, Vol. 23, No. 11, Article No. 5024, June, 2023.

[10] Y. Liu, D. Yang, Y. Wang, J. Liu, J. Liu, A. Boukerche, P. Sun, L. Song, Generalized Video Anomaly Event Detection: Systematic Taxonomy and Comparison of Deep Models, *ACM Computing Surveys*, Vol. 56, No. 7, pp. 1-38, July, 2024.

[11] N. Cao, C. Lin, Q. Zhu, Y.-R. Lin, X. Teng, X. Wen, Voila: Visual Anomaly Detection and Monitoring with Streaming Spatiotemporal Data, *IEEE transactions on visualization and computer graphics*, Vol. 24, No. 1, pp. 23-33, January, 2018.

[12] J. T. Geiger, K. Helwani, Improving Event Detection for Audio Surveillance Using Gabor Filterbank Features, *European Signal Processing Conference (EUSIPCO)*, Nice, France, 2015, pp. 714-718.

[13] M. Z. Yazd, I. Taheri, A. Samimi, M. Sarvi, A Robust Machine Learning Structure for Driving Events Recognition Using Smartphone Motion Sensors, *Intelligent Transportation Systems*, Vol. 28, No. 1, pp. 54-68, January, 2024.

[14] N. Choudhry, J. Abawajy, S. Huda, I. Rao, A Comprehensive Survey of Machine Learning Methods for Surveillance Videos Anomaly Detection, *IEEE Access*, Vol. 11, pp. 114680-114713, October, 2023.

[15] M. Ramzan, A. Abid, H. U. Khan, S. M. Awan, A. Ismail, M. Ahmed, M. Ilyas, A. Mahmood, A Review on State-Of-The-Art Violence Detection Techniques, *IEEE Access*, Vol. 7, pp. 107560-107575, July, 2019.

[16] A. B. Mabrouk, E. Zagrouba, Abnormal Behavior Recognition for Intelligent Video Surveillance Systems: A Review, *Expert Systems with Applications*, Vol. 91, pp. 480-491, January, 2018.

[17] M. Gnouma, R. Ejbali, M. Zaied, Abnormal Events' Detection in Crowded Scenes, *Multimedia Tools and Applications*, Vol. 77, No. 19, pp. 24843-24864, October, 2018.

[18] C. B. Murthy, M. F. Hashmi, N. D. Bokde, Z. W. Geem, Investigations of Object Detection in Images/Videos Using Various Deep Learning Techniques and Embedded Platforms—A Comprehensive Review, *Applied Sciences*, Vol. 10, No. 9, Article No. 3280, May, 2020.

[19] Y. Zhu, S. Newsam, Motion-Aware Feature for Improved Video Anomaly Detection, arXiv:1907.10211, July, 2019, https://arxiv.org/abs/1907.10211v1

[20] B. Mirzaei, H. Nezamabadi-pour, A. Raoof, R. Derakhshani, Small Object Detection and Tracking: A Comprehensive Review, *Sensors*, Vol. 23, No. 15, Article No. 6887, August, 2023.

[21] M. I. Georgescu, R. T. Ionescu, F. S. Khan, M. Popescu, M. Shah, A Background-Agnostic Framework with Adversarial Training for Abnormal Event Detection in Video, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 9, pp. 4505-4523, September, 2022.

[22] S. Hamdi, S. Bouindour, K. Loukil, H. Snoussi, M.

Abid, Hybrid Deep Learning and HOF for Anomaly Detection, *Control, Decision and Information Technologies (CoDIT)*, Paris, France, 2019, pp. 575-580.

[23] I. C. Hwang, H. S. Kang, Anomaly Detection Based on a 3D Convolutional Neural Network Combining Convolutional Block Attention Module Using Merged Frames, *Sensors*, Vol. 23, No. 23, Article No. 9616, December, 2023.

[24] W. Ullah, A. Ullah, T. Hussain, Z. A. Khan, S. W. Baik, An Efficient Anomaly Recognition Framework Using an Attention Residual LSTM in Surveillance Videos, *Sensors*, Vol. 21, No. 8, Article No. 2811, April, 2021.

[25] H. Song, C. Sun, X. Wu, M. Chen, Y. Jia, Learning Normal Patterns Via Adversarial Attention-Based Autoencoder for Abnormal Event Detection in Videos, *IEEE Transactions on Multimedia*, Vol. 22, No. 8, pp. 2138-2148, August, 2020.

[26] W. Zhang, G. Wang, M. Huang, H. Wang, S. Wen, Generative Adversarial Networks for Abnormal Event Detection in Videos Based on Self-Attention Mechanism, *IEEE Access*, Vol. 9, pp. 124847-124860, September, 2021.

[27] Z. Li, Y. Li, Z. Gao, Spatiotemporal Representation Learning for Video Anomaly Detection, *IEEE Access*, Vol. 8, pp. 25531-25542, January, 2020.

[28] F. Dong, Y. Zhang, X. Nie, Dual Discriminator Generative Adversarial Network for Video Anomaly Detection, *IEEE Access*, Vol. 8, pp. 88170-88176, May, 2020.

[29] B. Sabzalian, H. Marvi, A. Ahmadyfard, Deep and Sparse Features for Anomaly Detection and Localization in Video, *Pattern Recognition and Image Analysis (IPRIA)*, Tehran, Iran, 2019, pp. 173-178.

[30] Y. Fan, Y. Yu, W. Lu, Y. Han, Weakly-Supervised Video Anomaly Detection with Snippet Anomalous Attention, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 34, No. 7, pp. 5480-5492, July, 2024.

[31] A. M. Kamoona, A. Khodadadian Gostar, A. Bab-Hadiashar, R. Hoseinnezhad, Multiple Instance-Based Video Anomaly Detection Using Deep Temporal Encoding–Decoding, *Expert Systems with Applications*, Vol. 214, Article No. 119079, March, 2023.

[32] W. Li, V. Mahadevan, N. Vasconcelos, Anomaly Detection and Localization in Crowded Scenes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 1, pp. 18-32, January, 2014.

[33] C. Lu, J. Shi, J. Jia, Abnormal Event Detection at 150 fps in Matlab, *Computer Vision*, Sydney, Australia, 2013, pp. 2720-2727.

[34] W. Liu, W. Luo, D. Lian, S. Gao, Future Frame Prediction for Anomaly Detection–A New Baseline, *Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, Utah, 2018, pp. 6536-6545.

[35] A. Adam, E. Rivlin, I. Shimshoni, D. Reinitz, Robust Real-Time Unusual Event Detection using Multiple Fixed-Location Monitors, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 3, pp. 555–560, March, 2008.

[36] Ruchika, R. K. Purwar, Abnormality Detection Using LBP Features and K-Means Labelling Based Feed-Forward Neural Network in Video Sequence, *Innovative Technology and Exploring Engineering*, Vol. 8, No. 9S, pp. 629-633, July, 2019.

[37] S. R. E. Datondji, Y. Dupuis, P. Subirats, P. Vasseur, A Survey of Vision-Based Traffic Monitoring of Road Intersections, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 10, pp. 2681-2698, October, 2016.

[38] H. Singh, E. M. Hand, K. Alexis, Anomalous Motion Detection on Highway Using Deep Learning, *Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates, 2020, pp. 1901-1905.

[39] J. Jorge, M. Villarroel, H. Tomlinson, O. Gibson, J. L. Darbyshire, J. Ede, M. Harford, J. D. Young, L. Tarassenko, P. Watkinson, Non-Contact Physiological Monitoring of Post-Operative Patients in the Intensive Care Unit, *NPJ Digital Medicine*, Vol. 5, No. 4, January, 2022.

[40] A. Choudhary, S. Chaudhury, Video Analytics Revisited, *IET Computer Vision*, Vol. 10, No. 4, pp. 237-249, June, 2016.

[41] S. Ramapatruni, S. N. Narayanan, S. Mittal, A. Joshi, K. Joshi, Anomaly Detection Models for Smart Home Security, *Intelligent Data and Security (IDS*), Washington, DC, USA, 2019, pp. 19-24.

[42] T. D'Orazio, C. Guaragnella, A Survey of Automatic Event Detection in Multi-Camera Third Generation Surveillance Systems, *Pattern Recognition and Artificial Intelligence*, Vol. 29, No. 1, Article No. 1555001, February, 2015.

[43] G. Bhola, D. K. Vishwakarma, A Review of Vision-Based Indoor HAR: State-Of-The-Art, Challenges, and Future Prospects, *Multimedia Tools and Applications*, Vol. 83, No. 1, pp. 1965-2005, January, 2024.

[44] M. Xu, S. Yoon, A. Fuentes, D. S. Park, A Comprehensive Survey of Image Augmentation Techniques for Deep Learning, *Pattern Recognition*, Vol. 137, Article No. 109347, May, 2023.

[45] K. Doshi, Y. Yilmaz, Towards Interpretable Video Anomaly Detection, *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, 2023, pp. 2654-2663.

[46] L. Caruccio, G. Polese, G. Tortora, D. Iannone, EDCAR: A Knowledge Representation Framework to Enhance Automatic Video Surveillance, *Expert Systems with Applications*, Vol. 131, pp. 190-207, October, 2019.

[47] X. Qu, Z. Liu, C. Q. Wu, A. Hou, X. Yin, Z. Chen, MFGAN: Multimodal Fusion for Industrial Anomaly Detection Using Attention-Based Autoencoder and Generative Adversarial Network, *Sensors*, Vol. 24, No. 2, Article No. 637, January, 2024.

[48] R. Kalshetty, A. Parveen, Abnormal Event Detection Model using an Improved ResNet101 in Context Aware Surveillance System, *Cognitive Computation and Systems*, Vol. 5, No. 2, pp. 153-167, June, 2023.

[49] S. Wan, S. Ding, C. Chen, Edge Computing Enabled Video Segmentation for Real-Time Traffic Monitoring in Internet of Vehicles, *Pattern Recognition*, Vol. 121, Article No. 108146, January, 2022.

[50] G. Balamurugan, J. Jayabharathy, An Integrated Framework for Abnormal Event Detection and Video Summarization Using Deep Learning, *International Journal of Advanced Technology and Engineering Exploration*, Vol. 9, No. 95, pp. 1494-1507, October, 2022.

[51] Y.-T. Chang, W. Gunarathne, T. K. Shih, Deep Learning Approaches for Dynamic Object Understanding and Defect Detection, *Journal of Internet Technology*, Vol. 21, No. 3, pp. 783-790, May, 2020.

Biographies

Varsha Anil Kshirsagar received a Master of Engineering in VLSI and Embedded Systems from Savitribai Phule Pune University, Pune, M.S., India. She is currently pursuing a Ph.D. in IoT and Machine Learning in the Department of Computer Science and Information Engineering at National Dong Hwa University in Taiwan.

Shou-Chih Lo received the Ph.D. degree in Computer Science from National Tsing Hua University, Taiwan, in 2000. Since 2004, he has been an Associate Professor in the Department of Computer Science and Information Engineering at National Dong Hwa University, Taiwan. His research interests focus on mobile and wireless networks.

Guanling Lee received the Ph.D. degree in computer science from National Tsing Hua University Taiwan in 2001. She joined National Dong Hwa University in the Department of CSIE and became an associate professor in 2005. Her research interests include resource management in the mobile environment and data mining.