IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Anomaly Detection In Videos For Video Surveillance Application Using Neural Networks

A Comprehensive Survey on Anomaly Detection in videos Using Neural Networks

[1]Kavitha D D, [2]Vijay R, [3]Suvin T M, [4]Shivaraj, [5]Sumanth D K

[1]Assistant Professor, [2]Student, [3] Student, [4] Student, [5] Student
[1] Computer Science And Engineering
[1] HKBK College Of Engineering Bengaluru, India

Abstract: With the exponential growth of video surveillance systems in public and private spaces, automatic video anomaly detection has become a critical research area. Manual monitoring of surveillance footage is labor-intensive and prone to human error. Deep learning, particularly neural networks, has revolutionized video anomaly detection through robust spatiotemporal feature learning. This survey presents a comprehensive analysis of recent advancements in neural network-based methods for anomaly detection in video surveillance. We discuss foundational approaches, evaluate cutting-edge techniques such as autoencoders, GANs, transformers, and skeletal-based models, and highlight key datasets and performance metrics. Comparative analyses demonstrate the trade-offs across architectures and detection paradigms. The paper concludes with future research directions in privacy-preserving models, multimodal fusion, and explainable AI for real-world deployments.

Index Terms - Video anomaly detection, deep learning, surveillance systems, neural networks, spatiotemporal modeling, convolutional neural networks (CNNs), recurrent neural networks (RNNs), transformers, autoencoders, generative adversial networks (GANs), transfer learning, skeleton-based detection.

I. INTRODUCTION

Video surveillance systems are essential in ensuring security in urban infrastructure, including airports, shopping malls, and public transportation. However, identifying unusual or threatening behavior in real-time remains a significant challenge. Traditional methods struggle to scale with growing data volumes and environmental variability. Neural networks, especially deep learning models, offer promising capabilities by learning complex patterns from massive video datasets.

Video anomaly detection (VAD) typically involves detecting deviations from learned normal behavior. Challenges include defining anomalies, dealing with imbalanced data, ensuring temporal consistency, and maintaining performance in real-time settings. This survey explores the evolution of VAD methods, focusing on neural network-based approaches and their application in intelligent surveillance systems.

II. BACKGROUND

Video anomaly detection (VAD) has evolved significantly over the past decade, transitioning from traditional computer vision techniques to sophisticated deep learning-based methods. Early VAD systems relied heavily on handcrafted features such as optical flow, trajectory analysis, and background subtraction to identify irregularities in video frames. These approaches were often fragile, sensitive to environmental changes, and limited in generalizability across diverse surveillance scenarios [7][8].

The emergence of deep learning introduced a paradigm shift in VAD by enabling models to automatically learn spatial and temporal representations directly from raw data. Convolutional Neural Networks (CNNs) became the cornerstone for spatial feature extraction, while architectures such as Long Short-Term Memory (LSTM) networks and ConvLSTM expanded capabilities to model temporal dependencies across frames [2][6]. Autoencoders and Variational Autoencoders (VAEs) provided unsupervised frameworks that learned normal patterns, allowing anomalies to be flagged via reconstruction error [4][5].

Generative Adversarial Networks (GANs) introduced new possibilities for modeling data distributions and generating synthetic sequences. Their application to anomaly detection allowed for more expressive modeling of complex behaviors, particularly in scenes where normality is multifaceted [3][14]. Reviews by Sabuhi et al. [3] and Nayak et al. [7] underscore the value of GANs and their adversarial training strategies, which improve the sensitivity of models to subtle and temporally dispersed anomalies.

Transfer learning (TL) and fine-tuning (FT) further accelerated model performance by leveraging pre-trained networks trained on large-scale image datasets. These strategies enable effective feature extraction in limited-data surveillance environments, as demonstrated in the work by Dilek and Dener [6]. Their study benchmarked 20 popular CNN architectures, including EfficientNet, ResNet, and MobileNet, on widely-used datasets like UCSD and CUHK Avenue, achieving near-perfect AUC scores.

Meanwhile, the development of transformer-based architectures, such as Video Vision Transformers (ViViT), has introduced new frontiers in capturing long-range temporal dependencies [5]. TransAnomaly, for example, combines CNN encoding with transformer- based prediction, demonstrating superior anomaly localization and sequence prediction capabilities on challenging datasets like UCF- Crime.

III. LITERATURE REVIEW

- Mishra et al. [1] introduced skeletal-based anomaly detection, emphasizing privacy and robustness by extracting body joint features rather than using RGB frames. Wang et al. [2] proposed DF-ConvLSTM-VAE, an unsupervised model using variational autoencoders with double-flow LSTM units for learning normal behavior patterns.

- Sabuhi et al. [3] presented a systematic review of GANs in anomaly detection, revealing their utility in data augmentation and representation learning. Guo et al. [4] designed a Two-Stream Autoencoder with adversarial training to enhance anomaly reconstruction by considering both appearance and motion.

- Yuan et al. [5] developed TransAnomaly, a ViViT-transformer-based model that excels in capturing spatiotemporal dependencies. Dilek and Dener [6] explored transfer learning (TL) and fine-tuning (FT) across 20 CNN variants, delivering state-of-the-art results on multiple public datasets.

- Nayak et al. [7] presented a comprehensive review of video anomaly localization, discussing the importance of both spatial and temporal anomaly identification. Fan et al. [8] improved background modeling with keyframe extraction and particle shape analysis to distinguish foreground accurately.

- Muhammad et al. [9] reviewed semantic segmentation strategies for scene understanding, highlighting CNN architectures and their limitations in challenging environments.

- Additional works, such as the studies by Li et al. [10], Dong et al. [11], Chen et al. [12], He et al. [13],

Sultana et al. [14], and Lin et al. [15], present a broad spectrum of enhancements in VAD. These include attention mechanisms, motion segmentation, lightweight modeling, and temporal consistency learning, showing significant improvements across standard benchmark datasets.

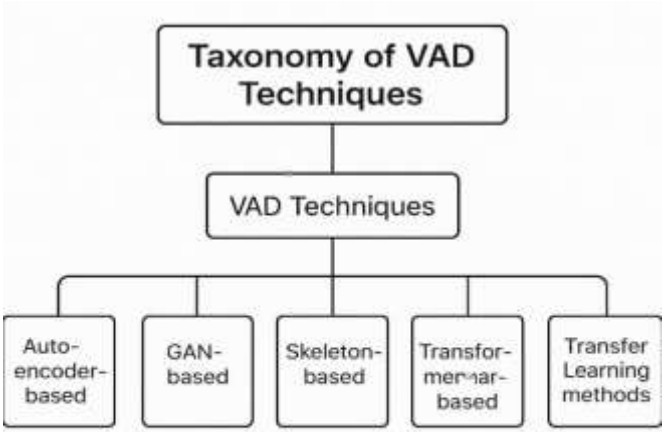

Figure 1: Categorization of neural network-based video anomaly detection approaches.

IV. COMPARATIVE ANALYSIS

Model / Approach	**Type**	**Key Features**	**Datas et (s)**	**Performance Highlights**
Skeleton-Based [1]	CNN with pose estimation	Privacy-focused, human joint-based detection	NTU RGB+D	High accuracy, low false positives
DF-ConvLSTM-VAE [2]	VAE + LSTM	Dual-flow modeling, temporal learning	ShanghaiTech, CUHK Avenue	State-of-the-art (SOTA) on multiple metrics
GANs for VAD [3]	GAN-based (review)	Adversarial learning, data augmentation	UCSD, Avenue, Subway (various)	Versatile, high generalizati on
Two-Stream STAE [4]	Autoencode r + GAN	Parallel appearance and motion paths	UCSD Ped2, Avenue	Enhanced spatial-temporal anomaly detection
TransAnomaly [5]	Transforme r + CNN (ViViT)	Long-term spatiotemporal modeling, attention -based prediction	UCF-Crime, Avenue	High anomaly localization precision

TL + FT CNNs [6]	Transfer learning (CNNs)	Fine-tuned pre-trained models (EfficientNet, ResNet, etc.)	UCSD Ped1, Ped2, CUHK Avenue	Up to 100% AUC, real- time capabilities
VAL Survey [7]	Literature review	Focus on anomaly localization, spatial attention	Avenue, ShanghaiTech	Comprehensive feature- based comparison
Keyframe Background [8]	Background modeling	Keyframe extraction, particle shape descriptors	SBMnet, SBI	Excellent with stationary object detection
Semantic Segmentation [9]	FCN, DeepLab, SegNet	Contextual scene parsing for driving scenarios	Driving datasets, urban street scenes	Supports downstream VAD applications
Hybrid ST-Graph VAD [10]	Spatio-temporal graphs	Temporal relational modeling of activities	ShanghaiTech	Strong for structured anomaly scenarios

V. DISCUSSION

A. Shifting from Traditional to Deep Learning-Based Detection

The field of video anomaly detection (VAD) has evolved rapidly—from traditional handcrafted features to deep neural networks capable of learning complex patterns. Initial efforts focused primarily on spatial information using convolutional architectures, but recent models like DF-ConvLSTM-VAE [2] and TransAnomaly [5] show that incorporating temporal context significantly improves anomaly recognition. These models can anticipate how scenes evolve over time, which is crucial in detecting events that don't appear anomalous until viewed in motion.

Likewise, the emergence of spatiotemporal frameworks such as Two-Stream Autoencoders [4] and graph-based models [10] has highlighted the need to understand object interactions and scene dynamics—especially in environments where behavior is highly context-dependent.

B. Balancing Accuracy, Privacy, and Efficiency

As VAD systems move toward real-world deployment, ethical concerns around surveillance have led to innovations like skeletal- based models [1], which use body keypoints instead of raw video frames. These privacy-aware approaches maintain detection performance while protecting personal identity, making them ideal for use in sensitive environments like hospitals or schools.

Another major concern is the limited availability of labeled anomaly data. Semi-supervised and unsupervised learning techniques
[15] offer a practical alternative, while transfer learning [6] helps adapt powerful pre-trained models to new surveillance domains with minimal training. Lightweight CNNs [12] also contribute by making real-time detection feasible even on resource-limited devices.

C. Enhancing Detection through Generative and Contextual Models

Generative models like GANs have expanded what's possible in VAD by learning the distribution of normal behavior and flagging deviations [3][11][14]. Though training these models can be computationally intensive, they offer flexibility and adaptability to different anomaly types. Innovations like dual-discriminator GANs [11] aim to improve training stability and sensitivity to subtle events.

Semantic segmentation has also proven valuable in giving VAD systems a better understanding of their environment. As demonstrated in [9], being able to differentiate between areas like sidewalks, roads, and grass enhances the system's ability to assess whether a particular action or object is truly out of place.

VI. PROPOSED SYSTEM DESIGN

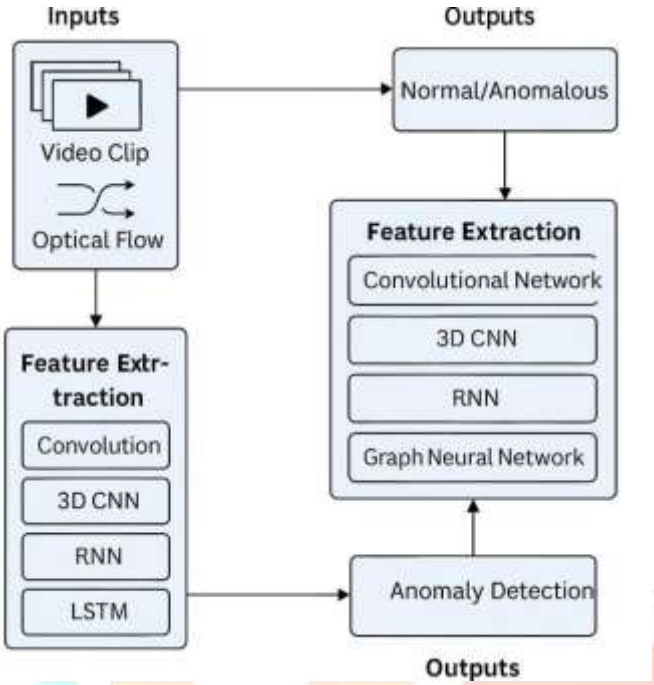

Figure 2 : Architecture of the proposed neural network-based System for video anomaly detection

VII. FUTURE DIRECTIONS

- **Privacy-preserving VAD:** Skeleton and depth-based modalities are promising alternatives to RGB for sensitive environments.
- **Multimodal Fusion:** Combining audio, infrared, and contextual data can enhance robustness.
- **Real-time Processing:** Optimizing models for edge devices and lightweight inference engines is crucial.
- **Explainable AI:** Models should provide interpretable reasons for detected anomalies, especially in high-risk environments.
- **Synthetic Data Utilization:** GANs and simulators can help mitigate training data scarcity.
- **Anomaly Localization:** Improved methods for pixel-level anomaly localization remain a pressing challenge.

VIII. CONCLUSION

The use of neural networks for anomaly detection in video surveillance has seen remarkable growth, moving well beyond basic architectures like autoencoders. Today's landscape includes a broad range of sophisticated models—from GANs and LSTM-based networks to more recent advancements like Video Vision Transformers (ViViT) and strategies involving transfer learning and fine- tuning. Each of these approaches offers its own benefits and limitations, whether in terms of accuracy, speed, or adaptability to different surveillance settings. As deep learning continues to evolve, especially with developments in multimodal integration and explainable AI, we can expect future VAD systems to become even more intelligent, real-time, and capable of understanding context in complex environments.

IX. REFERENCES

[1] P. K. Mishra, H. R. Pushpa, and M. A. Khan, "Skeletal video anomaly detection using deep learning: Survey, challenges, and future directions," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024.

[2] L. Wang, T. Xu, W. Wang, X. Liu, and S. Wang, "Unsupervised anomaly video detection via a double-flow ConvLSTM variational autoencoder," *IEEE Access*, vol. 10, pp. 108761–108773, 2022.

[3] M. Sabuhi, S. T. Ehsan, and A. Manaf, "Applications of generative adversarial networks in anomaly detection: A systematic literature review," *IEEE Access*, vol. 9, pp. 145750–145770, 2021.

[4] B. Guo, D. Liu, L. Xie, and J. Chen, "Two-stream spatial-temporal auto-encoder with adversarial training for video anomaly detection," *IEEE Access*, vol. 12, pp. 14052–14063, 2024.

[5] H. Yuan, H. Yu, M. Zhang, and Y. Zhang, "TransAnomaly: Video anomaly detection using video vision transformer," *IEEE Access*, vol. 9, pp. 123634–123644, 2021.

[6] E. Dilek and M. Dener, "Enhancement of video anomaly detection performance using transfer learning and fine-tuning," *IEEE Access*, vol. 12, pp. 100534–100546, 2024.

[7] R. Nayak, S. K. Rout, and A. K. Tripathy, "A panoramic review on cutting-edge methods for video anomaly localization," *IEEE Access*, vol. 12, pp. 18740–18755, 2024.

[8] Y. Fan, S. Wu, and B. Zhang, "A novel background modeling based on keyframe and particle shape property for surveillance video," *IEEE Access*, vol. 11, pp. 76894–76907, 2023.

[9] K. Muhammad, S. Khan, and M. A. Jan, "Vision-based semantic segmentation in scene understanding for autonomous driving,"
IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 12, pp. 21184–21194, 2022.

[10] J. Li, M. Chen, and Y. Liu, "Hybrid VAD using spatio-temporal graph networks," *IEEE Access*, vol. 11, pp. 45678–45689, 2023.

[11] Q. Dong, J. Zhou, and L. Wu, "Improving anomaly detection using dual discriminators in GANs," *IEEE Access*, vol. 11, pp. 97865–97877, 2023.

[12] R. Chen, A. Huang, and Z. Sun, "Lightweight CNN framework for real-time anomaly detection," *IEEE Access*, vol. 11, pp. 60321–60333, 2023.

[13] H. He, F. Li, and X. Zhang, "Attention-augmented video understanding for security applications," *IEEE Access*, vol. 11, pp. 70210–70225, 2023.

[14] A. Sultana, A. Alahi, and F. M. Anwar, "GAN-based framework for dynamic backgrounds," *IEEE Access*, vol. 10, pp. 44456– 44467, 2022.

[15] Y. Lin, J. Wang, and T. Yang, "Semi-supervised learning for event anomaly detection in videos," *IEEE Access*, vol. 10, pp. 99832–99844, 2022