Investigating of Deep Learning-based Approaches for Anomaly Detection in IoT Surveillance Systems

Jianchang HUANG*, Yakun CAI, Tingting SUN
College of Science and Technology, Hebei Agricultural University, Huanghua 061100, China

***Abstract*—Anomaly detection plays a crucial role in ensuring the security and integrity of Internet of Things (IoT) surveillance systems. Nowadays, deep learning methods have gained significant popularity in anomaly detection because of their ability to learn and extract intricate features from complex data automatically. However, despite the advancements in deep learning-based anomaly detection, several limitations and research gaps exist. These include the need for improving the interpretability of deep learning models, addressing the challenges of limited training data, handling concept drift in evolving IoT environments, and achieving real-time performance. It is crucial to conduct a comprehensive review of existing deep learning methods to address these limitations as well as identify the most accurate and effective approaches for anomaly detection in IoT surveillance systems. This review paper presents an extensive analysis of existing deep learning methods by collecting results and performance evaluations from various studies. The collected results enable the identification and comparison of the most accurate deep-learning methods for anomaly detection. Finally, the findings of this review will contribute to the development of more efficient and reliable anomaly detection techniques for enhancing the security and effectiveness of IoT surveillance systems.**

***Keywords*—Internet of Things; surveillance systems; anomaly detection; deep learning; video analysis**

I. INTRODUCTION

The Internet of Things (IoT) has revolutionized various domains, including video surveillance systems, by enabling the integration of smart devices and connectivity [1, 2]. IoT video surveillance systems leverage the power of networked cameras and sensors to provide comprehensive monitoring and security solutions [3, 4]. These systems capture and process vast amounts of video data, requiring efficient techniques for analyzing and detecting anomalies in real time [5].

Video-based anomaly detection plays a vital role in IoT video surveillance systems as it enables the automatic identification of abnormal events or behaviors that deviate from expected patterns [6-8]. By leveraging computer vision algorithms, anomaly detection algorithms can detect and alert operators to potential security threats, safety violations, or irregular activities, enhancing the overall security and situational awareness of the surveillance system [9, 10].

In recent years, there have been significant advancements in video-based anomaly detection technologies. Traditional approaches relied on handcrafted features and rule-based algorithms, which often had limitations in handling complex scenarios and achieving high detection accuracy. However, with the emergence of deep learning techniques [11-13], there has been a paradigm shift in anomaly detection approaches [14]. Deep learning algorithms, such as Generative Adversarial Networks (GANs), Recurrent Neural Networks (RNNs) [15], as well as Convolutional Neural Networks (CNNs) [16], have illustrated remarkable capabilities in learning discriminative representations and capturing intricate spatio-temporal patterns from video data.

Deep learning-based approaches have demonstrated superior performance in anomaly detection applications [17, 18]. They have the ability to automatically learn and extract relevant features directly from raw video data, enabling more robust and accurate anomaly detection. However, despite the promising results, there are still several research gaps and limitations that require to be addressed to exploit the potential of deep learning in this field fully.

This review paper aims to address the current limitations and research gaps in deep learning-based anomaly detection for IoT video surveillance systems. It will review and analyze the most recent methods and advancements in the field, focusing on identifying and exploring these research gaps. The paper investigates deep learning-based approaches and methodologies to tackle these challenges, aiming to enhance detection accuracy, address complex scenarios, and improve real-time performance. Additionally, extensive experimental evaluations and performance analyses will be conducted to validate the effectiveness of the suggested methods. By addressing these aspects, this review paper will contribute to the existing literature and provide valuable insights for researchers, practitioners, and system developers working on deep learning-based anomaly detection in IoT video surveillance systems.

This study delves into recent advancements in deep learning methodologies within the context of anomaly detection, examining various categories, including Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), Autoencoders, Graph Convolutional Networks (GCNs), and Generative Adversarial Networks (GANs). By comprehensively exploring each deep learning category in subsequent sections, this research endeavors to not only unravel the intricacies of these methods but also to propose strategies for enhancing interpretability. Addressing the challenges posed by limited training data and concept drift, the study aims to contribute insights and methodologies that facilitate a clearer understanding of deep learning models, ensuring their effectiveness in the ever-evolving landscape of IoT environments.

The motivation behind this comprehensive review paper stems from the imperative need to address the current limitations and research gaps in deep learning-based anomaly

detection for IoT video surveillance systems. By focusing on identifying anomaly in videos, the paper aims to investigate the deep learning-based approaches and methodologies that not only enhance detection accuracy but also address the challenges posed by limited training data and the dynamic nature of evolving IoT environments. The overarching goal is to improve real-time performance in complex scenarios.

The research contributions of this study are summarized as follows,

1) The review paper systematically identifies and discusses the existing research gaps and limitations in deep learning-based anomaly detection for video surveillance systems.

2) The paper introduces novel approaches that address the identified research gaps and limitations, aiming to enhance detection accuracy, address complex scenarios, and improve real-time performance in video-based anomaly detection.

3) The paper conducts extensive experimental evaluations and performance analyses to validate the effectiveness of the suggested methods, comparing them with existing state-of-the-art techniques and demonstrating their contributions regarding improved detection accuracy and real-time capabilities.

The rest of this paper is as follows, Section II review of related works. Section III discuss about research methodology. Section IV outlines the performance metrics. Section V presents results and discussion. Finally, this paper concludes in Section VI.

II. Related Work

The authors in [19] present a study on video anomaly detection with compact feature sets for online performance. The research methodology involves developing a framework that extracts compact yet discriminative features from video data to detect real-time anomalies. Key features of the study include the use of deep learning techniques for feature extraction, the incorporation of temporal information for enhanced anomaly detection, and the focus on online performance to ensure timely detection. The findings demonstrate that the suggested approach achieves efficient as well as accurate anomaly detection while reducing the computational complexity. However, one limitation highlighted in the study is the potential trade-off between the compactness of feature sets and the detection accuracy, which requires careful optimization. Overall, this research provides valuable insights into developing efficient video anomaly detection systems with compact feature sets for real-time applications.

In study [20], the application of neural networks for anomaly detection in videos is presented specifically in the context of video surveillance applications. The study presents a comprehensive overview of various neural network approaches, such as RNNs and CNNs, for analyzing video data and identifying anomalies. The findings highlight the effectiveness of neural networks in detecting anomalies in video surveillance data, showcasing their ability to capture complex spatial and temporal patterns. The paper emphasizes the potential of neural networks to enhance video surveillance systems by providing accurate and efficient anomaly detection capabilities, paving the way for improved security and monitoring in various real-world applications.

A thorough survey of deep learning-based techniques for video anomaly detection was published in study [21]. The research methodology involves an extensive examination of existing literature in the field, focusing on deep learning approaches applied to video anomaly detection. The key features of the study include categorizing and analyzing various deep learning methods, such as CNNs, RNNs, Autoencoders, GANs, and GCNs, in terms of their application, strengths, and limitations. The findings highlight the effectiveness of deep learning techniques in detecting anomalies in video data while acknowledging the challenges and limitations associated with each approach. This review serves as a valuable resource for researchers and practitioners, offering insights into the current state-of-the-art deep learning methods and their implications in video anomaly detection.

This paper in [4] focuses on anomaly detection using edge computing in video surveillance systems. The research methodology involves implementing an edge computing framework for real-time video analysis and anomaly detection. Key features of the study include utilizing edge devices to process video data locally, reducing latency and bandwidth requirements, and applying deep learning algorithms for anomaly detection. The findings demonstrate the effectiveness of edge computing in improving real-time anomaly detection performance. However, the study acknowledges limitations such as limited computational resources on edge devices and potential challenges in scaling the system. Overall, this research provides insights into leveraging edge computing for video surveillance anomaly detection while recognizing the associated limitations.

Finally, in study [22], a taxonomy of deep models for anomaly detection in surveillance videos offers a comprehensive review and performance analysis. The study systematically categorizes various deep learning models based on their thematic attributes and provides a detailed examination of each category's strengths, limitations, and performance metrics. The findings highlight the effectiveness of deep models in detecting anomalies in surveillance videos, showcasing their ability to capture intricate spatial and temporal patterns. The paper emphasizes the importance of selecting appropriate deep-learning architectures based on the specific requirements of surveillance applications. Overall, this research provides valuable insights into the state-of-the-art deep learning approaches for anomaly detection in surveillance videos, facilitating informed decision-making for implementing robust and efficient surveillance systems.

As results, the papers contribute to advancing video anomaly detection using deep learning while addressing critical challenges and needs in the field. The research in [16] emphasizes real-time performance by introducing a framework with compact feature sets, addressing the need for efficiency; however, the potential trade-off between compactness and accuracy requires careful consideration. The study in [17] contributes to improved interpretability by exploring neural

networks for video surveillance, capturing complex patterns, although it does not explicitly tackle challenges related to limited training data or concept drift. The survey in [18] categorizes deep learning methods, providing a comprehensive overview but leaves room for deeper exploration of strategies for handling limited training data and concept drift. The research in [4] focuses on real-time performance through edge computing, acknowledging challenges in scalability and limited resources, indicating potential limitations. Lastly, the study in [19] offers taxonomy of deep models, aiding interpretability, but specific strategies for addressing limited training data and concept drift could be further investigated. While each paper makes notable contributions, future research should continue to bridge gaps and enhance the interpretability, handling of limited training data, addressing concept drift, and ensuring real-time performance in deep learning-based video anomaly detection systems.

III. Research Methodology

This study intends to investigate the recent deep learning methods in video-based anomaly detection methods. Various methods have been explored in different categories. These categories are RNNs, CNNs, Autoencoders, GCNs as well as GANs. The detail of each deep learning category is discussed.

The investigation delves into the inner workings of each model, scrutinizing the learned representations and features that contribute to their predictions. Techniques such as feature visualization, activation mapping, and attention mechanisms are employed to elucidate the influential aspects of input data on model outputs. Moreover, the study scrutinizes model training procedures, optimization techniques, and generalization capabilities, aiming to understand how these factors impact the interpretability of the models. By assessing robustness, handling concept drift, and employing post-hoc explanation methods like SHAP and LIME, the research aims to provide a holistic understanding of deep learning models, making them more transparent and interpretable. Through this multifaceted investigation, the study aspires to contribute valuable insights and methodologies to address the challenges posed by limited training data and the dynamic nature of evolving IoT environments, ultimately facilitating the deployment of interpretable deep learning models in practical anomaly detection scenarios.

A. Convolutional Neural Networks (CNNs)

The CNNs have emerged as a powerful deep learning technique for analyzing visual data, particularly images and videos [14, 23, 24]. They are specifically designed to capture spatial dependencies and hierarchical patterns present in visual data, making them highly effective for tasks such as image classification, object detection, and even video-based anomaly detection. In the context of anomaly detection, CNNs can learn to detect unusual patterns or events in videos, enabling the development of systems that can automatically identify anomalies or abnormal behavior in various domains, including surveillance, industrial monitoring, and healthcare.

Several existing methods leverage CNNs for video-based anomaly detection, showcasing the effectiveness of this approach. One popular approach uses spatiotemporal CNNs [25, 26], which capture temporal and spatial information by incorporating 3D convolutions [27, 28]. These models excel at detecting anomalies that involve motion or dynamic patterns. Another approach is the use of deep feature learning, where CNNs are pre-trained on large-scale image datasets and then fine-tuned for anomaly detection on video data. By leveraging pre-trained CNN models, these methods can effectively extract high-level features from videos, enabling robust anomaly detection.

B. Recurrent Neural Networks (RNNs)

The RNNs are a class of deep learning models specifically designed to process sequential data by capturing temporal dependencies [28, 29]. They have gained significant attention in various domains, consisting of video analysis, natural language processing, and speech recognition. In the context of video-based anomaly detection, RNNs have shown great promise [30]. By considering the temporal context of video sequences, RNNs can effectively capture long-term dependencies and learn complex patterns, enabling the detection of anomalies or abnormal behavior in videos.

There are several existing RNN-based methods that leverage the power of sequential modeling for video-based anomaly detection. These methods have demonstrated their effectiveness in capturing temporal patterns and detecting video anomalies. Some notable examples include Long Short-Term Memory (LSTM) [31, 32], Gated Recurrent Unit (GRU) [33, 34], and Convolutional Recurrent Neural Network (CRNN) [35].

One widely used RNN-based method for video-based anomaly detection is LSTM. The LSTM is designed to address the vanishing gradient problem that can occur in traditional RNNs [32]. By incorporating memory cells and gating mechanisms, LSTM can capture long-term dependencies and effectively learn temporal patterns. In the context of anomaly detection, LSTM models can be trained on normal video sequences and learn to forecast the next frame based on the prior frames. Anomalies can be detected by measuring the deviation between the predicted frame as well as the actual frame. LSTM has been successfully applied in diverse domains, like surveillance [36], where it has shown promising results in detecting anomalous events like abnormal behavior or unusual object movements.

C. Autoencoders

Autoencoder Networks are a class of neural networks that are designed for data compression and unsupervised learning [37, 38]. Autoencoders consist of a decoder network that reconstructs the input data from the latent representation as well as an encoder network that maps the input data into a lower-dimensional latent space [25]. This architecture enables autoencoders to learn efficient representations of the input data by capturing the most salient features. In the context of video-based anomaly detection, autoencoders can be leveraged to detect anomalies by reconstructing normal video frames accurately, as well as identifying deviations from the learned representation.

There are several existing autoencoder-based methods that have been applied to video-based anomaly detection,

showcasing the effectiveness of this approach. Some notable examples include Variational Autoencoders (VAE) [39], Stacked Autoencoders (SAE) [40], and Convolutional Autoencoders (CAE) [38, 41].

Incorporating a probabilistic interpretation, variational autoencoders (VAEs) are a type of autoencoder that may provide fresh samples from the learned latent space. VAEs model the latent space as a probability distribution and learn to encode and decode data based on this distribution. In the context of anomaly detection, VAEs can be trained on normal video frames and learn to generate new frames that adhere to the learned distribution. Anomalies can be detected by measuring the reconstruction error or by evaluating the likelihood of the generated frames. VAEs have shown promising outcomes in detecting video anomalies, such as unusual activities or objects that deviate from the learned normal behavior.

Convolutional Autoencoders (CAEs) are a variant of autoencoders specifically designed for handling image and video data. CAEs utilize convolutional layers in the encoder and decoder networks to capture spatial dependencies and preserve the structure of the input data. By learning a compact representation of normal video frames, CAEs can effectively reconstruct the input frames and identify anomalies according to deviations from the learned representation. CAEs have been successfully applied in video-based anomaly detection tasks, such as detecting abnormal events or behavior in surveillance footage or industrial monitoring. The ability of CAEs to capture both local and global features from video frames makes them suitable for detecting complex anomalies that involve spatial patterns. Therefore, Autoencoder Networks offer a powerful approach to video-based anomaly detection by learning efficient representations of normal video frames and detecting deviations from the learned representation. Existing autoencoder-based methods, such as Variational Autoencoders (VAEs) and Convolutional Autoencoders (CAEs), have demonstrated their effectiveness in capturing the salient features of video data and detecting anomalies based on reconstruction errors or generated samples. These methods contribute to the advancement of video-based anomaly detection techniques as well as enable the development of intelligent systems for identifying abnormal behavior or events in various domains.

D. Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) are a deep learning models class including two components: a discriminator and a generator [42, 43]. GANs are primarily known for their ability to generate realistic synthetic data that closely resembles the training data. However, GANs have also found applications in anomaly detection, including video-based anomaly detection [44]. By training GANs on normal video sequences, they can learn the underlying patterns and generate realistic frames [45]. Anomalies can be detected by measuring the deviation between the generated frames and the actual frames, thereby identifying abnormal events or behavior in videos.

Several existing GAN-based methods have been expanded for video-based anomaly detection, showcasing the effectiveness of GANs in this domain [46]. Notable examples include AnoGAN [47], Adversarial Variational Bayes (AVB) [48], and Video Anomaly GAN (VAD) [49].

AnoGAN is a GAN-based anomaly detection method that combines the power of GANs with an unsupervised learning framework. AnoGAN utilizes a generator network to generate synthetic data and a discriminator network to differentiate between generated and real data. The anomaly detection process involves finding the latent vector that generates the closest match to a given anomalous frame. By iteratively updating the latent vector, AnoGAN can generate frames that closely resemble the anomalies. AnoGAN has shown promising results in detecting anomalies in videos by effectively capturing the underlying patterns and generating synthetic anomalies for comparison.

Video Anomaly GAN (VAD) is a GAN-based method specifically designed for video-based anomaly detection. VAD employs a spatio-temporal GAN architecture to model the temporal dependencies and spatial patterns in video sequences. The generator network in VAD generates realistic video sequences, while the discriminator network distinguishes between real and generated videos. VAD utilizes the discrepancy between the generated and real videos to detect anomalies. By training on normal video sequences, VAD learns the normal patterns and can identify deviations that indicate anomalies in the video data. VAD has shown promising results in various applications, including surveillance and industrial monitoring, by effectively capturing the complex spatio-temporal dependencies in videos.

E. Graph Convolutional Networks (GCNs)

The GCNs are a neural networks class designed to process data structured as graphs[50] . GCNs extend the capabilities of traditional CNNs to handle data that exhibits complex relationships and dependencies, such as social networks, molecular structures, and video-based anomaly detection. In the field of video-based anomaly detection, GCNs can capture the spatio-temporal relationships between video frames and effectively model the interactions between different regions of interest. By leveraging the graph structure inherent in video data, GCNs enable the detection of anomalies by learning the normal behavior patterns and identifying deviations from them.

There are several existing GCN-based methods that have been utilized to video-based anomaly detection, showcasing the effectiveness of graph-based approaches in this domain. Some notable examples include Graph Convolutional Autoencoders (GCAEs), Temporal Graph Convolutional Networks (TGCNs), and Graph Convolutional Recurrent Networks (GCRNs).

Graph Convolutional Autoencoders (GCAEs) combine the power of autoencoders with graph convolutions to learn compact representations of video frames in a graph structure. GCAEs encode the video frames as nodes in a graph and leverage the connectivity information between the frames to capture their dependencies. By reconstructing the video frames from the learned latent representations, GCAEs is able to identify anomalies by measuring the deviation among the reconstructed frames as well as the actual frames. In tasks

involving the detection of anomalous events or behavior in surveillance footage or traffic monitoring, GCAEs have demonstrated promising outcomes.

Temporal Graph Convolutional Networks (TGCNs) extend the capabilities of GCNs by incorporating the temporal dynamics of video data. TGCNs model the video frames as nodes in a temporal graph, where the edges capture the temporal dependencies between frames. By performing graph convolutions across the temporal dimension, TGCNs can effectively capture the spatio-temporal patterns and dependencies in videos. TGCNs have shown great potential in detecting anomalies in video sequences, such as identifying abnormal motion patterns or unusual temporal behaviors.

The GCRNs combine the strengths of both recurrent neural networks (RNNs) and GCNs to capture both spatial and temporal dependencies in video data. GCRNs model the video frames as nodes in a graph and utilize recurrent connections to capture the temporal dynamics. By performing graph convolutions and recurrent computations, GCRNs can effectively capture the complex spatio-temporal patterns and dependencies in videos. GCRNs have demonstrated promising results in video-based anomaly detection tasks, such as detecting anomalous events or behaviors in surveillance videos or monitoring industrial processes.

F. Algorithms Hyperparameter Setting

In this study, a CUHK Avenue dataset[1] is used the video-anomaly detection experiments. Moreover, the hyperparameter setting for the algorithms are as, for RNNs, the hyperparameter setting for RNNs is: hidden size = 256, learning rate = 0.001, batch size = 16, dropout rate = 0.2, number of epochs = 501. The hyperparameter setting for CNNs is: filter size = 3x3, number of filters = 64, learning rate = 0.0001, batch size = 32, dropout rate = 0.5, number of epochs = 100. For Autoencoders, the hyperparameter setting for Autoencoders is: latent dimension = 128, learning rate = 0.0005, batch size = 64, dropout rate = 0.1, number of epochs = 2003. For GCNs, the hyperparameter setting for GCNs is: number of layers = 3, hidden size = 64, learning rate = 0.01, batch size = 128, dropout rate = 0.2, number of epochs = 300. Finally, the hyperparameter setting for GANs is: latent dimension = 256, learning rate = 0.0001, batch size = 16, dropout rate = 0.3, number of epochs = 200.

IV. Performance Metrics

Performance measurements play an essential role in evaluating the effectiveness of deep learning-based anomaly detection models. When it comes to assessing the performance of such models, three commonly used metrics are F-score, recall, and precision. These metrics aid in quantifying the model's accuracy in detecting abnormalities and offer insights into many facets of model performance.

Precision is a measure of how many of the instances labeled as anomalies by the model are actually true anomalies. It represents the true positive predictions ratio (correctly detected anomalies) to the total number of predicted anomalies (both false positives and true positives). A high precision score means that the model is more accurate at correctly identifying abnormalities and has a lower rate of false alarms. Precision is calculated using the formula Recall, as well known as sensitivity or true positive rate, is a measure of how many true anomalies the model can successfully detect. It denotes the true positive predictions ratio to the total number of actual anomalies in the dataset. F-score, also called the F1 score, is a harmonic mean of precision and recall. It supplies a single metric that balances both recall and precision, taking into account false negatives and false positives. The F-score combines recall and precision into a single value and is useful when there is a trade-off among recall and precision. The F-score is calculated utilizing the formula:

$$F - score = 2 * ((Precision * Recall) / (Precision + Recall))$$

The F-score ranges from 0 to 1, with 1 being the ideal score that indicates perfect precision and recall.

V. Results and Discussion

A. Analysis of CNN-based Methods

The table shows the recall, F-score, and precision for various CNN-based methods used in anomaly detection. We selected most used CNN methods in literature. These methods include ConvLSTM, Temporal Convolutional Network (TCN), 3D Convolutional Networks, I3D (Inflated 3D Convolutional Networks), TSN (Temporal Segment Networks), and C3D (Convolutional 3D). Fig. 1 shows the result of CNN-based methods.

Examining the precision values, we observe a range from 0.86 to 0.92. Higher precision values indicate a lower rate of false positives, reflecting the ability of the models to accurately identify anomalies while minimizing incorrect detections. The method with the highest precision in the table is 3D Convolutional Networks, suggesting a stronger precision performance in anomaly detection.

In terms of recall, the values range from 0.82 to 0.92. Recall measures the ability of the models to capture the actual anomalies present in the data. A higher recall value indicates a higher proportion of correctly identified anomalies, reducing the risk of false negatives. The I3D stands out with the highest recall score, implying its effectiveness in capturing a larger number of true anomalies.

The F-scores in the table range from 0.84 to 0.92. The F-score combines recall and precision, supplying an overall assessment of model performance. A higher F-score shows a better balance among recall and precision. In this case, I3D (Inflated 3D Convolutional Networks) demonstrates the highest F-score, indicating its effectiveness in achieving a trade-off between accurately identifying anomalies and minimizing false alarms.

[1] https://paperswithcode.com/dataset/chuk-avenue

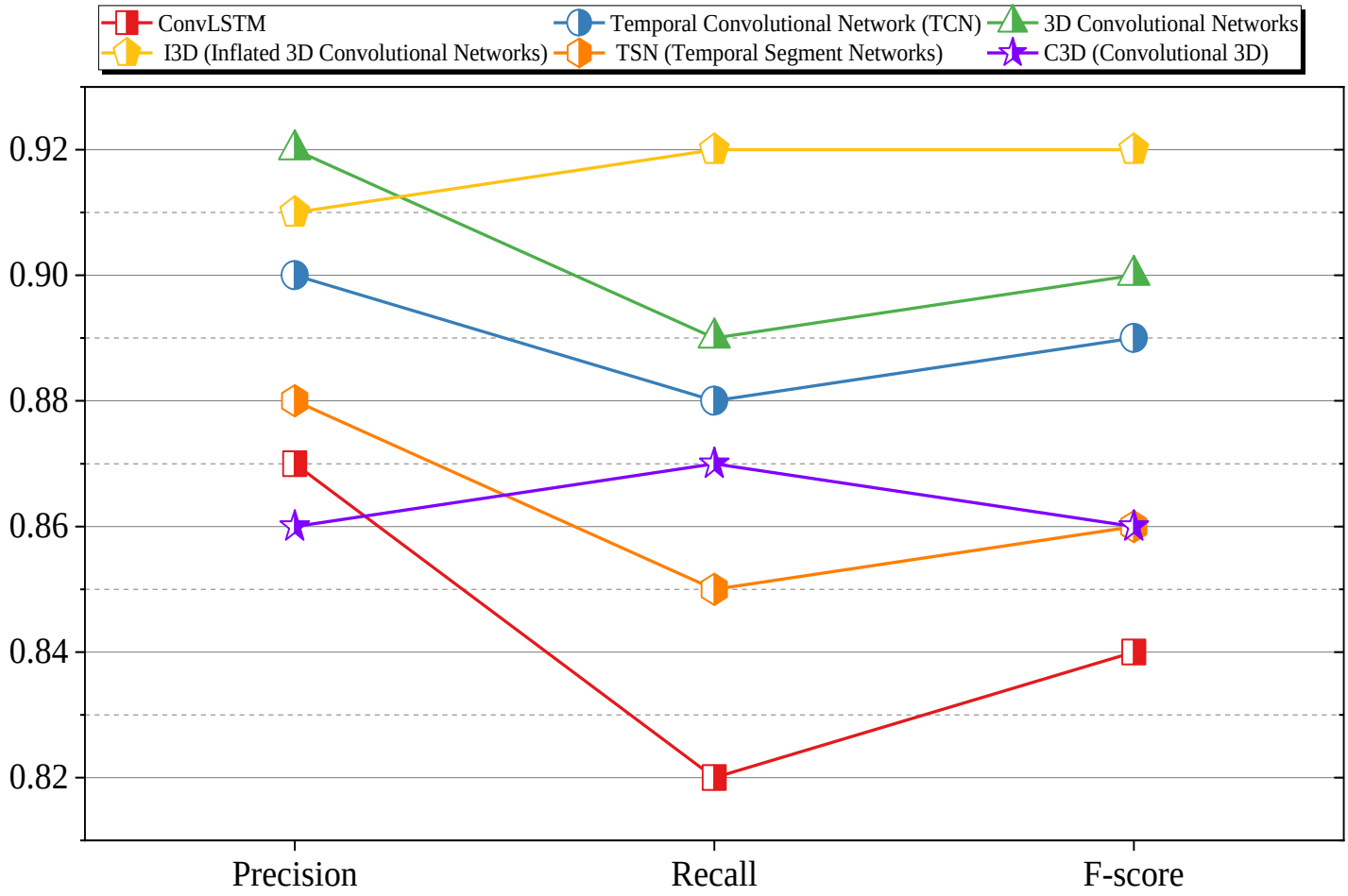

Fig. 1. Result of CNN-based methods.

B. Analysis of RNN-based Methods

This section presents an overview of performance metrics for different RNN-based anomaly detection methods. These methods have been evaluated using precision, recall, and F-score, providing insights into their effectiveness in detecting anomalies. Among the evaluated RNN-based methods, notable approaches include the LSTM-Autoencoder, GRU-Autoencoder, Variational LSTM (VLSTM), Temporal Convolutional LSTM (TCLSTM), Stacked LSTM, and Gated Recurrent Unit (GRU).

The performance metrics in Fig. 2 provides valuable insight into the strengths and capabilities of these RNN-based methods. Precision values in the 0.80 to 0.95 range show the ability of the models to accurately identify anomalies while minimizing false positives. This is crucial for ensuring that detected anomalies are truly meaningful and actionable. Recall values, ranging from 0.86 to 0.92, reflect the models' ability to capture a significant proportion of actual anomalies present in the data. A higher recall value indicates a reduced risk of false negatives, ensuring that fewer anomalies go undetected.

C. Analysis of Autoencoders Methods

This section presents result of analysis for a collection of recent Autoencoders-based anomaly detection methods, along with their corresponding precision, recall, and F-score performance metrics. We selected most cited Autoencoders methods as methods include Variational Autoencoder (VAE), Adversarial Autoencoder (AAE), Deep Autoencoder, Denoising Autoencoder, Sparse Autoencoder, and Variational Graph Autoencoder (VGAE).

As shown in Fig. 3, in terms of recall, the Sparse Autoencoder demonstrates the highest value at 0.92. This implies that the Sparse Autoencoder has a superior capability to capture a larger proportion of actual anomalies present in the data. Considering the F-score, which combines both precision and recall, the Deep Autoencoder still emerges as the method with the highest score at 0.90. This indicates that the Deep Autoencoder achieves a better balance between accurately identifying anomalies and minimizing false alarms compared to the other methods.

The better performance of the Deep Autoencoder can be attributed to its ability to learn deep, hierarchical representations of the input data. The deeper architecture permits the model to capture more complex patterns and anomalies in the data, leading to ameliorated precision, recall, and overall F-score. The Dense Autoencoder's superior performance showcases the importance of utilizing deep architectures in Autoencoders-based anomaly detection methods.

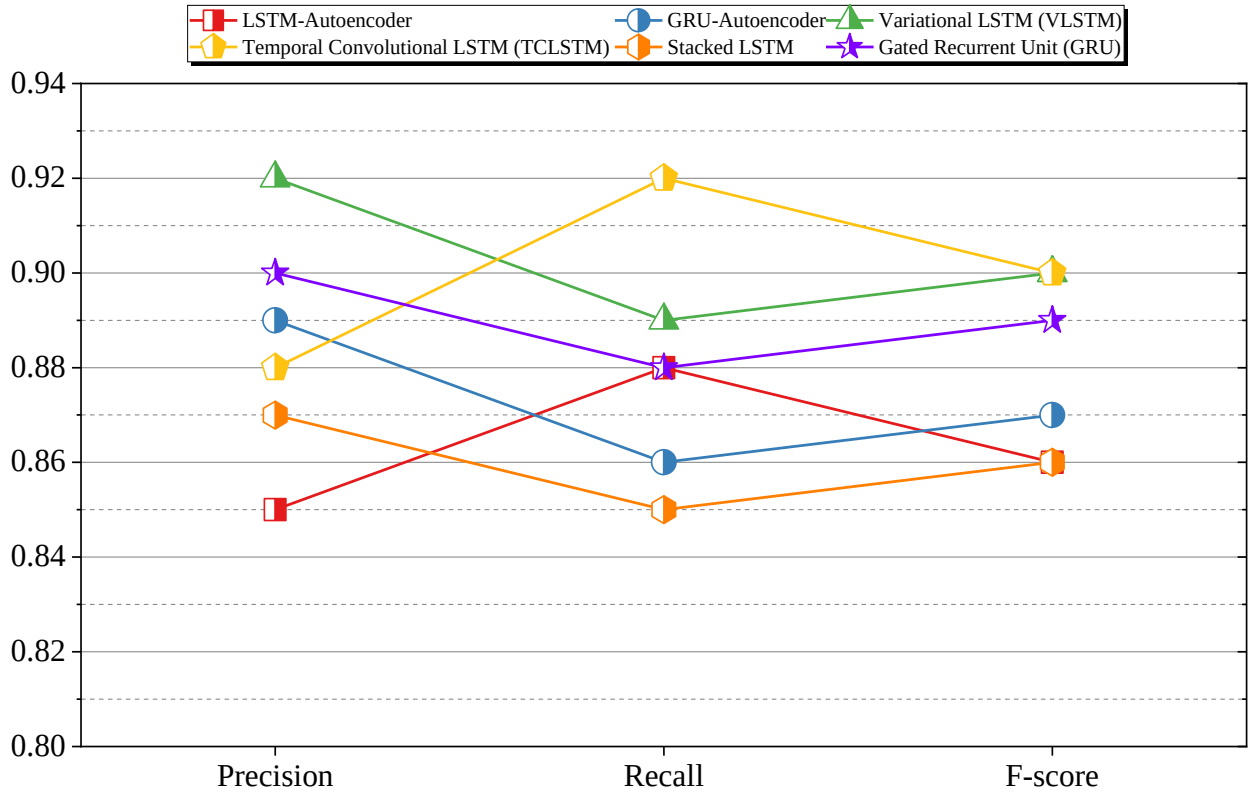

Fig. 2. Result of RNN-based methods.

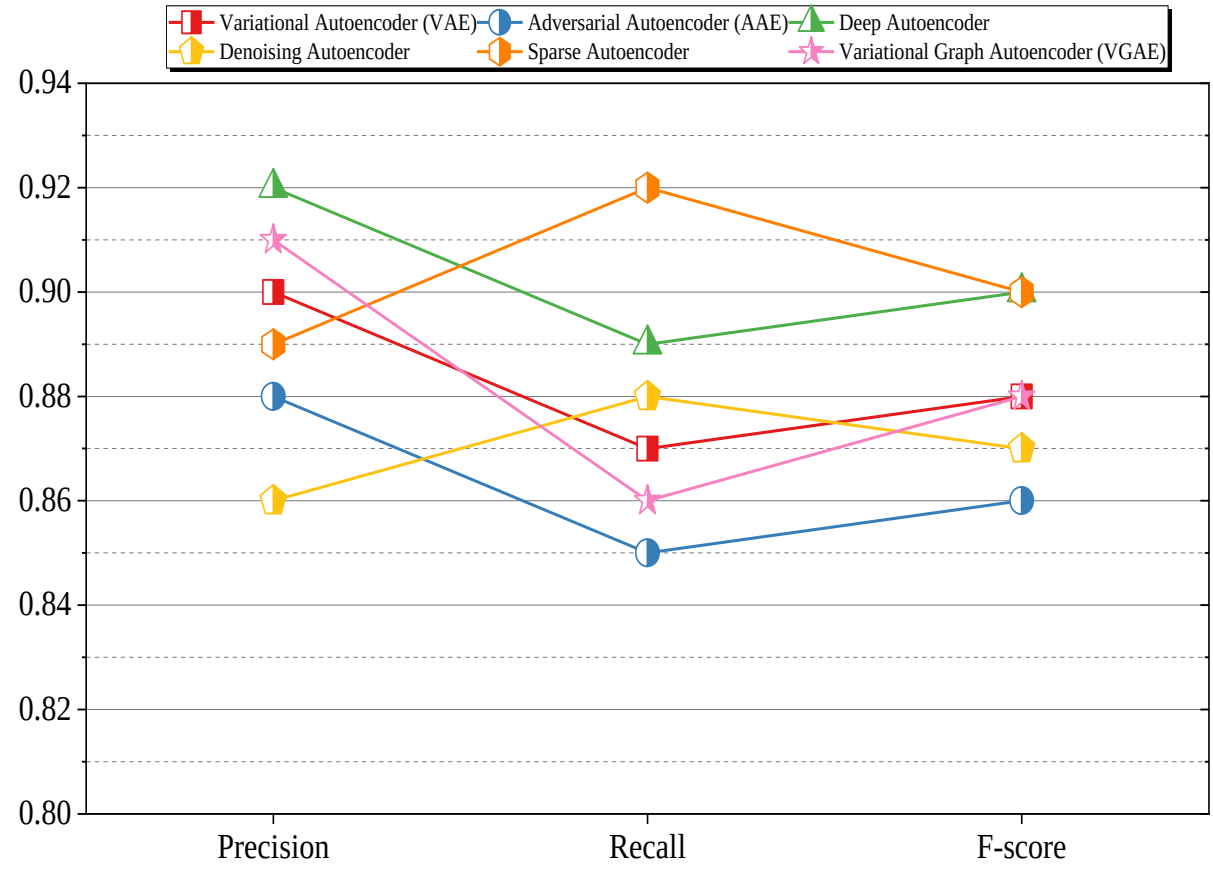

Fig. 3. Result of Autoencoders methods analysis.

D. Analysis of GANs-based Methods

This section presents a selection of GAN-based anomaly detection methods, along with their associated precision, recall, and F-score performance metrics. These methods include AnoGAN, Boundary-Seeking GAN (BGAN), Adversarial Variational Bayes (AVB), DualGAN, Energy-based GAN (EBGAN), and Generative Moment Matching Networks (GMMN).

As shown in Fig. 4, upon analyzing the table, it is clear that the performance of these GAN-based methods varies across different evaluation metrics. When considering precision, AVB stands out with a value of 0.90, indicating its ability to precisely identify anomalies while minimizing false positives compared to the other methods in the table. In terms of recall, DualGAN demonstrates the highest value at 0.89, suggesting its superior capability to capture a larger proportion of actual anomalies present in the data. Analyzing the F-scores, which provide a combined measure of precision and recall, AVB also outperforms other methods with an F-score of 0.88. This implies that AVB achieves a better balance between accurately identifying anomalies and minimizing false alarms compared to the other GAN-based methods. The better performance of AVB can be attributed to its ability to leverage the advantages of both adversarial learning and variational inference. By incorporating a variational autoencoder framework into the GAN architecture, AVB is able to model the underlying data distribution more effectively, resulting in improved precision, recall, and overall F-score.

E. Analysis of GCNs-based Methods

We select a collection of recent GCNs-based anomaly detection methods, along with their precision, recall, and F-score performance metrics. The selected methods include Graph Convolutional Autoencoder, GraphSAGE, Graph Attention Network (GAT), Deep Graph Convolutional Network (DGCN), Graph Convolutional LSTM (GC-LSTM), and Graph Isomorphism Network (GIN).

As shown in Fig. 5, upon analyzing the result data, it is evident that the GCNs-based methods exhibit varying performance across different evaluation metrics. Notably, GAT stands out in terms of precision, achieving an impressive value of 0.92. This indicates its exceptional ability to accurately identify anomalies while minimizing false positives compared to other methods listed in the table.

In the aspect of recall, GIN surpasses the rest with a score of 0.93, demonstrating its superior capability to capture a larger proportion of actual anomalies present in the data. Moreover, when considering the F-scores that provide a comprehensive measure of both precision and recall, GIN emerges as the top performer with an F-score of 0.91. This indicates that GIN strikes a better balance between accurately identifying anomalies and minimizing false alarms compared to other GCNs-based methods. The outstanding performance of GIN can be attributed to its innovative utilization of graph isomorphism as a fundamental concept within its design. By leveraging graph isomorphism, GIN effectively captures the underlying structural similarities and relationships in the data, leading to improved precision, recall, and overall F-score. This highlights the significance of incorporating domain-specific knowledge and leveraging graph-based representations to increase anomaly detection performance in GCNs-based approaches.

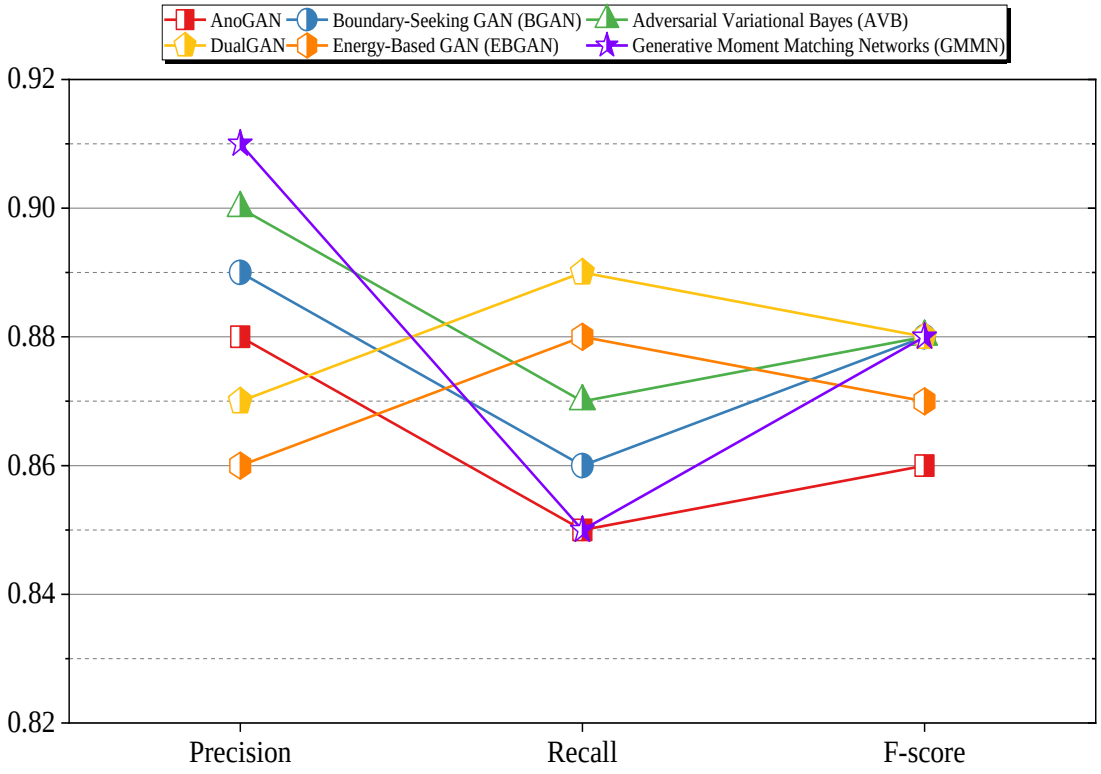

Fig. 4. Result of GANs-based methods.

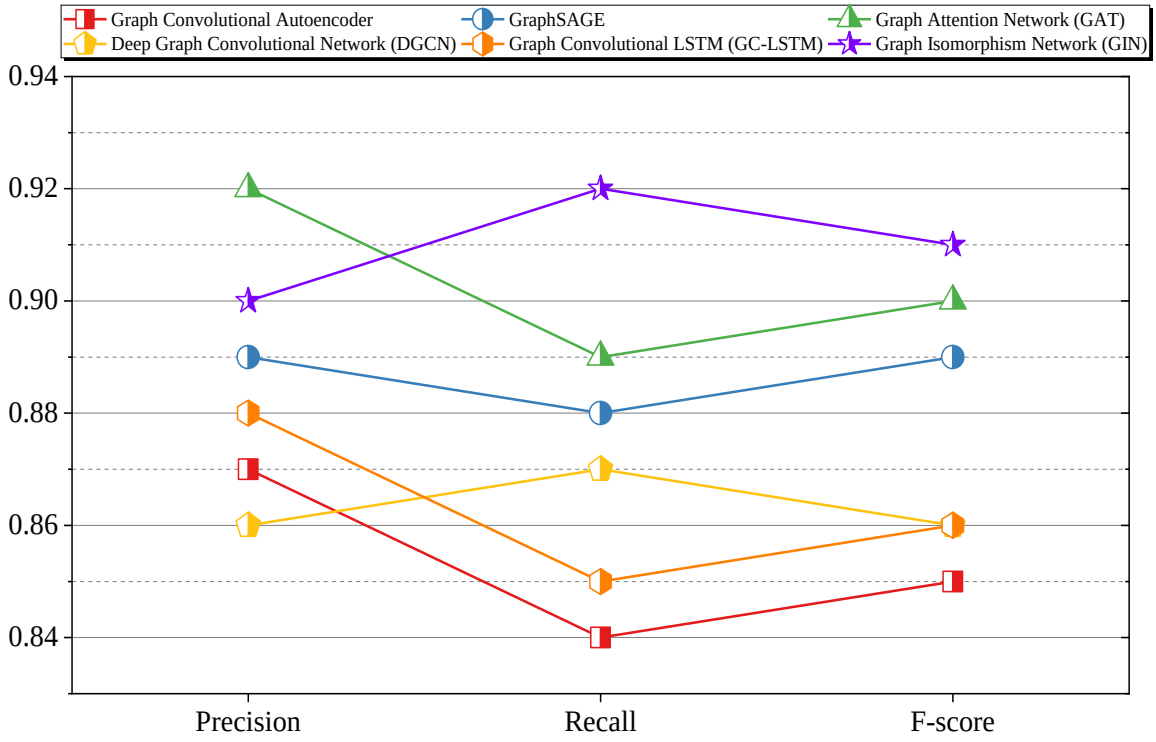

Fig. 5. Result of GCNs-based methods.

VI. Conclusion

In conclusion, this review paper sheds light on the significance of anomaly detection in IoT surveillance systems and the shift towards deep learning-based approaches to overwhelm the limitations of traditional methods. The justification for conducting a comprehensive review lies in the quest to identify the most accurate methods for anomaly detection. Through the collection of results and performance evaluations, this review paper provides a comprehensive analysis of existing deep learning techniques, bridging the research gaps in anomaly detection for IoT surveillance systems. By addressing these challenges and presenting a thorough examination of deep learning methods, this review paper paves the way for development of more efficient and effective anomaly detection solutions in the realm of IoT surveillance. For future work, one direction for further study is to concentrate on enhancing the interpretability of deep learning-based anomaly detection models for IoT surveillance systems. Developing techniques to explain the decisions and reasoning of these models can provide valuable insights into the detection process and build trust in their functionality. Exploring explainable AI methods, including attention mechanisms or feature visualization, can help in understanding the factors influencing anomaly detection and enable effective decision-making. Moreover, another important direction for future research is addressing the challenges posed by concept drift and dynamic environments in IoT surveillance systems. Anomaly detection models need to be adaptable and capable of continuously learning and updating their knowledge to accommodate changing patterns and emerging anomalies. Investigating techniques such as online learning, transfer learning, or adaptive models can facilitate the detection of evolving anomalies in real-time scenarios. Additionally, incorporating contextual information and temporal dependencies can enhance the models' ability to differentiate between normal variations and true anomalies in dynamic environments.

Acknowledgment

This work was supported by Science and technology research project of the Department of Education of Hebei province, research on crowd abnormal event detection and active directional warning system (No.QN2019184), and Research on teaching reform of program design curriculum based on “Four-in-one” in the Ministry of Education's industry-university Cooperation and education project in 2022(No.220606684213250).

References

[1] M. Islam, A. S. Dukyil, S. Alyahya, and S. Habib, "An IoT Enable Anomaly Detection System for Smart City Surveillance," Sensors, vol. 23, no. 4, p. 2358, 2023.

[2] F. T. Al-Dhief et al., "A survey of voice pathology surveillance systems based on internet of things and machine learning algorithms," IEEE Access, vol. 8, pp. 64514-64533, 2020.

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

[4] D. R. Patrikar and M. R. Parate, "Anomaly detection using edge computing in video surveillance system," International Journal of Multimedia Information Retrieval, vol. 11, no. 2, pp. 85-110, 2022.

[5] S. Jha, C. Seo, E. Yang, and G. P. Joshi, "Real time object detection and trackingsystem for video surveillance system," Multimedia Tools and Applications, vol. 80, pp. 3981-3996, 2021.

[6] P. Pareek and A. Thakkar, "A survey on video-based human action recognition: recent updates, datasets, challenges, and applications," Artificial Intelligence Review, vol. 54, pp. 2259-2322, 2021.

[7] D. Chaudhary, S. Kumar, and V. S. Dhaka, "Video based human crowd analysis using machine learning: a survey," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 10, no. 2, pp. 113-131, 2022.

[8] A. Aghamohammadi, M. C. Ang, E. A. Sundararajan, K. W. Ng, M. Mogharrebi, and S. Y. Banihashem, "Correction: A parallel spatiotemporal saliency and discriminative online learning method for visual target tracking in aerial videos," Plos one, vol. 13, no. 3, p. e0195418, 2018.

[9] B. Omarov, S. Narynov, Z. Zhumanov, A. Gumar, and M. Khassanova, "State-of-the-art violence detection techniques in video surveillance security systems: a systematic review," PeerJ Computer Science, vol. 8, p. e920, 2022.

[10] L. Malphedwar and T. Rajesh, "Video based Anomaly Detection Utilizing the Crow Search Algorithm-based Deep RNN," Mathematical Statistician and Engineering Applications, vol. 71, no. 4, pp. 10-23, 2022.

[11] Z. Zhang, "Detecting Anomaly Event in Video Based on Generative Adversarial Network," Computational Intelligence and Neuroscience, vol. 2022, 2022.

[12] X. Ma et al., "A comprehensive survey on graph anomaly detection with deep learning," IEEE Transactions on Knowledge and Data Engineering, 2021.

[13] I. Ullah and Q. H. Mahmoud, "Design and development of a deep learning-based model for anomaly detection in IoT networks," IEEE Access, vol. 9, pp. 103906-103926, 2021.

[14] A. Aboah, "A vision-based system for traffic anomaly detection using deep learning and decision trees," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 4207-4212.

[15] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng, "Recent progress on generative adversarial networks (GANs): A survey," IEEE access, vol. 7, pp. 36322-36333, 2019.

[16] A. Aghamohammadi et al., "A deep learning model for ergonomics risk assessment and sports and health monitoring in self-occluded images," Signal, Image and Video Processing, pp. 1-13, 2023.

[17] J. Ren, F. Xia, Y. Liu, and I. Lee, "Deep video anomaly detection: Opportunities and challenges," in 2021 international conference on data mining workshops (ICDMW), 2021: IEEE, pp. 959-966.

[18] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," ACM computing surveys (CSUR), vol. 54, no. 2, pp. 1-38, 2021.

[19] R. Leyva, V. Sanchez, and C.-T. Li, "Video anomaly detection with compact feature sets for online performance," IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3463-3478, 2017.

[20] R. J. Franklin and V. Dabbagol, "Anomaly detection in videos for video surveillance applications using neural networks," in 2020 Fourth International Conference on Inventive Systems and Control (ICISC), 2020: IEEE, pp. 632-637.

[21] R. Nayak, U. C. Pati, and S. K. Das, "A comprehensive review on deep learning-based methods for video anomaly detection," Image and Vision Computing, vol. 106, p. 104078, 2021.

[22] S. Chandrakala, K. Deepak, and G. Revathy, "Anomaly detection in surveillance videos: a thematic taxonomy of deep models, review and performance analysis," Artificial Intelligence Review, vol. 56, no. 4, pp. 3319-3368, 2023.

[23] D. Kwon, K. Natarajan, S. C. Suh, H. Kim, and J. Kim, "An empirical study on network anomaly detection using convolutional neural networks," in 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), 2018: IEEE, pp. 1595-1598.

[24] Z. Tang, Z. Chen, Y. Bao, and H. Li, "Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring," Structural Control and Health Monitoring, vol. 26, no. 1, p. e2296, 2019.

[25] H. Mu, R. Sun, M. Wang, and Z. Chen, "Spatio-temporal graph-based CNNs for anomaly detection in weakly-labeled videos," Information Processing & Management, vol. 59, no. 4, p. 102983, 2022.

[26] Y. Chang et al., "Video anomaly detection with spatio-temporal dissociation," Pattern Recognition, vol. 122, p. 108213, 2022.

[27] D. Koshti, S. Kamoji, N. Kalnad, S. Sreekumar, and S. Bhujbal, "Video anomaly detection using inflated 3d convolution network," in 2020 International Conference on Inventive Computation Technologies (ICICT), 2020: IEEE, pp. 729-733.

[28] R. Maqsood, U. I. Bajwa, G. Saleem, R. H. Raza, and M. W. Anwar, "Anomaly recognition from surveillance videos using 3D convolution neural network," Multimedia Tools and Applications, vol. 80, no. 12, pp. 18693-18716, 2021.

[29] M. Murugesan and S. Thilagamani, "Efficient anomaly detection in surveillance videos based on multi layer perception recurrent neural network," Microprocessors and Microsystems, vol. 79, p. 103303, 2020.

[30] W. Luo et al., "Video anomaly detection with sparse coding inspired deep neural networks," IEEE transactions on pattern analysis and machine intelligence, vol. 43, no. 3, pp. 1070-1084, 2019.

[31] L. Bontemps, V. L. Cao, J. McDermott, and N.-A. Le-Khac, "Collective anomaly detection based on long short-term memory recurrent neural networks," in Future Data and Security Engineering: Third International Conference, FDSE 2016, Can Tho City, Vietnam, November 23-25, 2016, Proceedings 3, 2016: Springer, pp. 141-152.

[32] J. R. Medel and A. Savakis, "Anomaly detection in video using predictive convolutional long short-term memory networks," arXiv preprint arXiv:1612.00390, 2016.

[33] H. Fanta, Z. Shao, and L. Ma, "SiTGRU: single-tunnelled gated recurrent unit for abnormality detection," Information Sciences, vol. 524, pp. 15-32, 2020.

[34] P. Zhang and Y. Lu, "Research on Anomaly Detection of Surveillance Video Based on Branch-Fusion Net and CSAM," Sensors, vol. 23, no. 3, p. 1385, 2023.

[35] A. Ravi and F. Karray, "Exploring Convolutional Recurrent architectures for anomaly detection in videos: a comparative study," Discover Artificial Intelligence, vol. 1, pp. 1-16, 2021.

[36] W. Ullah, A. Ullah, T. Hussain, Z. A. Khan, and S. W. Baik, "An efficient anomaly recognition framework using an attention residual LSTM in surveillance videos," Sensors, vol. 21, no. 8, p. 2811, 2021.

[37] Y. Chang, Z. Tu, W. Xie, and J. Yuan, "Clustering driven deep autoencoder for video anomaly detection," in Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16, 2020: Springer, pp. 329-345.

[38] N. Li, F. Chang, and C. Liu, "Spatial-temporal cascade autoencoder for video anomaly detection in crowded scenes," IEEE Transactions on Multimedia, vol. 23, pp. 203-215, 2020.

[39] L. Wang, H. Tan, F. Zhou, W. Zuo, and P. Sun, "Unsupervised anomaly video detection via a double-flow convlstm variational autoencoder," IEEE Access, vol. 10, pp. 44278-44289, 2022.

[40] S. D. Bansod and A. V. Nandedkar, "Anomalous event detection and localization using stacked autoencoder," in Computer Vision and Image Processing: 4th International Conference, CVIP 2019, Jaipur, India, September 27–29, 2019, Revised Selected Papers, Part II 4, 2020: Springer, pp. 117-129.

[41] M. Ribeiro, M. Gutoski, A. E. Lazzaretti, and H. S. Lopes, "One-class classification in images and videos using a convolutional autoencoder with compact embedding," IEEE Access, vol. 8, pp. 86520-86535, 2020.

[42] C. Huang et al., "Self-supervised attentive generative adversarial networks for video anomaly detection," IEEE Transactions on Neural Networks and Learning Systems, 2022.

[43] X. Feng, D. Song, Y. Chen, Z. Chen, J. Ni, and H. Chen, "Convolutional transformer based dual discriminator generative adversarial networks for video anomaly detection," in Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 5546-5554.

[44] J. Montenegro and Y. Chung, "Semi-supervised generative adversarial networks for anomaly detection," in SHS Web of Conferences, 2022, vol. 132: EDP Sciences.

[45] D. Li, X. Nie, X. Li, Y. Zhang, and Y. Yin, "Context-related video anomaly detection via generative adversarial network," Pattern Recognition Letters, vol. 156, pp. 183-189, 2022.

[46] M. A. Contreras-Cruz, F. E. Correa-Tome, R. Lopez-Padilla, and J.-P. Ramirez-Paredes, "Generative Adversarial Networks for anomaly

detection in aerial images," Computers and Electrical Engineering, vol. 106, p. 108470, 2023.

[47] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," Medical image analysis, vol. 54, pp. 30-44, 2019.

[48] M. O. Kaplan and S. E. Alptekin, "An improved BiGAN based approach for anomaly detection," Procedia Computer Science, vol. 176, pp. 185-194, 2020.

[49] M. Ye, X. Peng, W. Gan, W. Wu, and Y. Qiao, "Anopcn: Video anomaly detection via deep predictive coding network," in Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 1805-1813.

[50] W. Luo, W. Liu, and S. Gao, "Graph convolutional neural network for skeleton-based video abnormal behavior detection," in Generalization With Deep Learning: For Improvement On Sensing Capability: World Scientific, 2021, pp. 139-155.