applied sciences

MDPI

Review

An Analysis of Artificial Intelligence Techniques in Surveillance Video Anomaly Detection: A Comprehensive Survey

Erkan Şengönül [1], Refik Samet [1], Qasem Abu Al-Haija [2,*], Ali Alqahtani [3], Badraddin Alturki [4] and Abdulaziz A. Alsulami [5]

[1] Department of Computer Engineering, Faculty of Engineering, Ankara University, 06100 Ankara, Turkey
[2] Department of Cybersecurity, Princess Sumaya University for Technology (PSUT), Amman 11941, Jordan
[3] Department of Networks and Communications Engineering, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia
[4] Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[5] Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
* Correspondence: q.abualhaija@psut.edu.jo

Abstract: Surveillance cameras have recently been utilized to provide physical security services globally in diverse private and public spaces. The number of cameras has been increasing rapidly due to the need for monitoring and recording abnormal events. This process can be difficult and time-consuming when detecting anomalies using human power to monitor them for special security purposes. Abnormal events deviate from normal patterns and are considered rare. Furthermore, collecting or producing data on these rare events and modeling abnormal data are difficult. Therefore, there is a need to develop an intelligent approach to overcome this challenge. Many research studies have been conducted on detecting abnormal events using machine learning and deep learning techniques. This study focused on abnormal event detection, particularly for video surveillance applications, and included an up-to-date state-of-the-art that extends previous related works. The major objective of this survey was to examine the existing machine learning and deep learning techniques in the literature and the datasets used to detect abnormal events in surveillance videos to show their advantages and disadvantages and summarize the literature studies, highlighting the major challenges.

Keywords: video surveillance; abnormal events; anomaly detection; artificial intelligence

check for updates

Citation: Şengönül, E.; Samet, R.; Abu Al-Haija, Q.; Alqahtani, A.; Alturki, B.; Alsulami, A.A. An Analysis of Artificial Intelligence Techniques in Surveillance Video Anomaly Detection: A Comprehensive Survey. *Appl. Sci.* **2023**, *13*, 4956. https://doi.org/10.3390/app13084956

Academic Editor: Ke Gu

Received: 2 March 2023
Revised: 5 April 2023
Accepted: 12 April 2023
Published: 14 April 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The use of surveillance cameras in private and public spaces has become increasingly prevalent in recent years for various purposes, including tracking, monitoring, and preventing violations. An anomaly, as defined in the surveillance field, refers to a deviation from common rules, types, arrangements, or forms and can be characterized as an uncommon event that deviates from “normal” behavior.

Detecting anomalies in surveillance videos is crucial to maintaining security in various applications, such as crime detection, accident detection, abandoned object detection, illegal activity detection, and parking area monitoring. However, the manual detection of anomalies in surveillance videos is a tedious and labor-intensive task for humans. This is due to the large amount of data generated by critical systems in security applications, making manual analysis an impractical solution.

In recent years, there has been a significant increase in the demand for automated systems for detecting video anomalies. These systems include biometric identification of individuals, alarm-based monitoring of Closed-Circuit Television (CCTV) scenes, automatic detection of traffic violations, and video-based detection of abnormal behavior [1].

Automated systems significantly reduce human labor and time, making them more efficient and cost-effective for detecting anomalies in surveillance videos.

Identifying and tracking anomalies in the recorded video comprise a growing research problem in surveillance. Many methods have been proposed by researchers in academia, including the use of machine learning algorithms and image processing techniques.

In recent times, the utilization of surveillance cameras in public and private areas has risen significantly, serving multiple objectives. Identifying abnormalities in surveillance footage is vital to upholding safety measures in diverse scenarios. Nonetheless, the manual identification of anomalies in surveillance footage can be arduous and laborious for humans. Researchers have proposed automated systems for detecting video anomalies to address this issue, significantly reducing human labor and time. Despite the advancements that have been made, there is still room for improvement in the accuracy, reliability, and scalability in developing a flawless video surveillance system [2].

The Surveillance Video Anomaly Detection (SVAD) system is a sophisticated technology designed to detect unusual or suspicious behavior in video surveillance footage without human intervention. The system operates by analyzing the video frames and identifying deviations from normal patterns of movement or activity. This is achieved through advanced algorithms and machine learning techniques that can detect and analyze the position of pixels in the video frame at the time of an event.

Traditionally, anomaly detection methods have focused on identifying objects that deviate from normal trajectories. However, these methods need to be improved [3] for use in video surveillance due to the variety of objects that may be present in a video frame. As a result, two main approaches have been developed for video anomaly detection. The first approach involves measuring the magnitude of the error by calculating the reconstruction error of future frames. This is achieved by comparing the predicted future frames with the actual frames and identifying significant differences. The second approach involves predicting the future frames based on the previous frames and assigning a high anomaly score to any frame that deviates significantly from the predicted frame.

In recent years, with the advancement of hardware performance and the development of new models, smart learning techniques have become increasingly popular in video anomaly detection. However, the use of these techniques also brings several challenges. One of the major challenges is the production of big data, which requires a large amount of computational power for processing. High computational power also poses a significant challenge, requiring a significant resource investment [4].

Over the past two decades, a significant amount of research has been conducted on image and video processing to overcome these challenges. These studies have focused on developing new methods for anomaly detection that are more efficient and effective while also addressing the challenges associated with intelligent anomaly detection. Overall, understanding the issues of traditional anomaly detection methods and exploring new methods are crucial for the continued advancement of video surveillance.

This survey aimed to comprehensively examine the existing literature on Artificial Intelligence (AI) techniques for detecting abnormal events in surveillance videos. Specifically, the survey aimed to provide an overview of the most-commonly used datasets and evaluate their benefits and drawbacks. Additionally, the survey highlights key difficulties in the literature, providing insight into areas that require further research and development.

First, it is important to note that the use of AI in surveillance video analysis has gained significant attention in recent years due to its potential to improve the effectiveness and efficiency of surveillance systems. This is particularly relevant in security and surveillance, where detecting abnormal real-time events is crucial for public safety.

Various approaches and techniques, based on evolving artificial intelligence methods, enable the analysis of surveillance videos and the identification of abnormal events such as suspicious behavior, criminal activities, and other potential threats. The contributions of our study are as follows:

- A comprehensive survey of state-of-the-art AI approaches for SVAD was conducted. This analysis thoroughly examined the current research in the field, highlighting the most-popular techniques and methodologies used in SVAD.
- The commonly used datasets, needs, and issues of SVAD were explored in depth. This examination provides insight into the challenges faced by researchers in this field and the specific requirements of the datasets used for SVAD.
- The trade-offs in SVAD are discussed from the viewpoint of the performance of approaches that use AI techniques. This analysis provides a nuanced understanding of the trade-offs between performance and other factors, such as computational complexity and scalability.
- Areas of application, challenges, and possible future work in the field of AI for SVAD are presented. This examination provides a comprehensive overview of the potential applications of SVAD and the challenges that must be overcome to realize AI's potential in this field.

The remainder of this study is organized as follows: Section 2 presents a review of the related literature in the field of SVAD. Section 3 conducts an in-depth examination of AI techniques employed in SVAD. Section 4 provides an analysis of the commonly used datasets in this field. Section 5 evaluates the performance of existing SVAD applications. Section 6 engages in a critical discourse on the findings and implications, and finally, Section 7 presents our conclusions and recommendations for future research.

2. Related Works

There are many definitions of an anomaly. Frank E. Grubbs [5], in 1969, defined an outlier or an anomaly as "An outlying observation, or outlier, appears to deviate markedly from other members of the sample in which it occurs". Hawkins [6], in 1980, defined it as "an observation which deviates so much from other observations as to arouse suspicions that a different mechanism generated it". Barnett and Lewis [7], in 1994, defined it as "an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data".

In several situations, the same action can be interpreted as an abnormal or anomalous event because most anomaly detection techniques are based on the hypothesis that a pattern that deviates from previously acquired patterns is considered abnormal [8]. According to some studies [9–11], anomalies can be divided into three types:

- **Point anomalies:** These occur when only the entity's data behave somewhat irregularly compared to the rest of the data. Most research on anomaly identification focuses on this form of anomaly because it is the most basic. A car in the middle of the road can be termed a point anomaly.
- **Contextual anomalies:** These occur when a data value behaves irregularly compared to the rest of the data in a particular context. The context includes the observer's subjectivity and overall perception of the situation. Parking a passenger car in a bus-only car park can be considered a contextual anomaly.
- **Collective anomalies:** These occur when a collection of data samples is considered abnormal compared to the real data. A group of people gathered at the exit of a door can be called a collective anomaly.

Many approaches have been suggested within the scope of anomaly detection for both crowded and uncrowded environments. One study [12] highlighted the relationship between the number of moving objects in the clip and the complexity of the medium utilized to detect and identify abnormalities in the video: slightly crowded environment (10 square feet/per person), moderately crowded environment (4.5 square feet/per person), and crowded environment (2.5 square feet/per person).

Existing AI Techniques in SVAD

Because of a lack of knowledge sufficient to generalize their characteristics and correctly classify them as outliers, most machine learning (ML) and expert systems frequently need help detecting and classifying these anomalies. These rare events make identification challenging and contribute to an imbalanced data classification [13,14].

Once predictive models are built and sufficient data labels are provided, anomaly detection is challenging, considering the binary classification problem. However, the data available for training a model are restricted to containing few or no anomalous events, and such labels are frequently infrequent or cumbersome [15].

SVAD aims to find abnormal frames or pixel parts that contain various spatial and temporal data [16]. Spatial features can be collected from a single frame, whereas temporal features can be collected from the data on object movement and the order of frames. Generally, there are three methods for estimating abnormalities in SVAD [17]: (1) The characteristics of both regular and irregular events are reflected in a shared space, and the anomaly is identified based on the margin of the spatial distribution. (2) A dictionary was trained using the semantic properties of the event patterns. This dictionary is then used for anomaly calculation. (3) Anomalies are found through errors made during the prediction and reconstruction of prior or subsequent frames using various feature extractors trained to do so.

There are also various studies [18,19] on creating video synopses from surveillance cameras. These studies allow for a more compact view to select only active activities instead of whole frames and to achieve efficient video browsing. As a result, this contributes to focusing on the area of interest in SVAD and reducing false negatives more accurately.

The majority of methods in this domain have a variety of restrictions [20] such as: (1) The features used in many methods are handmade. (2) Most techniques demand a time-consuming stage for building models, having expertise that might not be useful for practical applications. (3) Perceiving deviations from normality as abnormal has been the subject of numerous earlier studies. In literature, hand-crafted descriptors for anomaly detection are particularly common. These traits continue to significantly contribute to research on anomaly identification [21].

Researchers have been inspired to use ML or DL techniques for abnormal event detection due to the success of similar techniques in computer vision and image processing [22]. Many challenging cognitive tasks, such as finding anomalies in surveillance video, have been solved using ML or DL [23].

Our review of the literature offers a comprehensive road map for SVAD. The published works in this field are grouped based on the learning method and algorithms.

3. Analysis of AI Techniques in SVAD

The taxonomy of SVAD, consisting of two main groups, is described in Table 1.

Table 1. Taxonomy for anomaly detection in video surveillance.

Learning	Algorithms
Supervised learning	Statistics-based algorithms
Unsupervised learning	Classification-based algorithms
Semi-supervised learning	Reconstruction-based algorithms
	Prediction-based algorithms
	Other algorithms

3.1. Learning

Several Artificial Intelligence (AI) subsets are based on various applications and use cases. This study mainly focused on Machine Learning (ML) and Deep Learning (DL). DL is a subset of machine learning methods. ML is a powerful technology that can be applied for anomaly detection. The process varies considerably depending on the problem.

The performance of an ML algorithm may vary depending on the features selected in the dataset or the weight assigned to each feature, even if the same model runs on two identical datasets [24]. A model may become overfit if it has fewer features that are only sometimes good. To better comprehend and construct a model using available ML techniques and data, reviewing and comparing the current solutions is worthwhile. Machine learning (ML) can be divided into three groups: Supervised Learning (SL), Unsupervised Learning (UL), and Semi-Supervised Learning (SSL).

3.1.1. Supervised Learning

SL acquires knowledge from pre-existing labeled datasets or "the training set", then compares the predicted output to the known labels. A high-level training set is always required to build a model that works effectively, but more is needed to ensure that the final product will be satisfactory; the training procedure is also a crucial element in creating a reliable predictor. A classifier model is first developed in SL through training, and after that, it can forecast either discrete or continuous outputs. The ASL model's performance, such as accuracy, is typically validated before prediction to demonstrate its dependability. Additionally, classification and regression techniques can be used to categorize SL tasks [25].

The training data are first divided into separate categories in the classification technique. It then calculates the probability of test samples falling into each category and chooses the category with the most votes [26]. This probability represents the likelihood that a sample is a class member. Credit scoring and medical imaging are examples of typical applications. The regression technique uses input factors such as temperature changes or variations in electricity demand to forecast continuous responses, often in quantity [27]. Forecasting power load and algorithmic trading are examples of typical applications. While the regression model can calculate the root-mean-squared error, the classification model can quantify the percentage of accurate predictions. Nevertheless, a discrepancy between the expected and actual values is acceptable since the output data are continuous.

Several works have been performed with SL. One of the suggestions in this area is presented by the study [28]. They proposed a unique way to identify fights or violent acts based on learning the temporal and spatial information from consecutive video frames that are evenly spaced. Using the proposed feature fusion approach, features with many levels for two sequential frames are retrieved from the first and last layers of the Convolutional Neural Network (CNN) and fused to consider the action knowledge. They also suggested a "Wide-Dense Residual Block" to learn the unified spatial data from the two input frames. These learned characteristics are subsequently consolidated and delivered to long-term memory components to store temporal dependencies. Using the domain adaptation strategy, the network may learn to efficiently merge features from the input frames, improving the results' accuracy. They evaluated their experiments by using four public datasets, namely HockeyFight, Movies, ViolentFlow, and BEHAVE, to show the performance of their model, which was compared with the existing models. There are several important learning techniques in SL, such the Hidden Markov Model (HMM) [29], Support Vector Machine (SVM) [30], Gaussian Regression (GR) [31], CNN [32], Multiple Instance Learning (MIL) [33], and Long Short-Term Memory (LSTM) [34]. It is clear that each technique has advantages and disadvantages in anomaly detection, and it is impossible to say that one technique can solve all problems efficiently.

3.1.2. Unsupervised Learning

UL groups data by identifying hidden patterns or intrinsic structures. Data input is necessary, but there are no predetermined output variables. There is neither labeled input data nor a training technique, in contrast to SL. As a result, it operates independently, and its performance could be more measurable. Although some researchers use the UL model's pre-existing labeled data to verify its results, this is only sometimes possible in practice. To conduct an external evaluation, specialists may need to analyze the results manually.

UL is mostly used for reducing dimensionality and clustering. UL is used in dimensionality reduction to find the dataset's linked features so that redundant data can be removed to reduce noise. Using clustering techniques, the clustering problem allows for the possibility of a sample belonging to more than one cluster or just one. Market research and object identification are common applications [35].

One proposed approach in UL is that of [36]. They provided a technique for detecting anomalies in surveillance missions, including UAV-acquired footage. They combined an unsupervised classification technique called One-Class Support Vector Machine (OCSVM) with a deep feature extraction technique utilizing a pre-trained CNN. Their quantitative findings demonstrated that their proposed strategy produces positive outcomes for the dataset studied. The authors in [37] extended their previous work by using mobile cameras to assist UAVs when acquiring videos. They added two feature extraction methods, the Histogram of Oriented Gradients (HOG) and HOG3D. They used the same UL method, which was OCSVM [38]. They obtained good results based on the used video-obtained datasets. There are many techniques under UL; PCA [39] and GANs [40] are examples of them.

3.1.3. Semi-Supervised Learning

SSL is a machine learning method that utilizes labeled and unlabeled data to create a classifier. This approach is particularly useful in situations with a limited amount of labeled data available. The SSL algorithm utilizes the training procedure described in Supervised Learning (SL) to create a predictor with a small amount of labeled data. The predictor then categorizes unlabeled samples and assigns each pseudo-labeled sample a confidence rating. This confidence rating informs the administrator of the prediction's certainty level. Once all data have been labeled, confident examples are added to the new training set to update the classifier.

Certain assumptions must be made before training unlabeled examples, such as smoothness and clustering. This is because unlabeled data are randomly labeled in the prediction process [41]. The anomaly detection (AE) model [42] is an important SSL model, as it utilizes labeled and unlabeled data to detect and identify anomalies in a given dataset. Overall, SSL is an effective method for creating a classifier with a limited amount of labeled data while leveraging the information present in unlabeled data to improve the accuracy of the classifier.

3.1.4. Supervised vs. Unsupervised vs. Semi-Supervised

Supervised learning techniques for SVAD offer several advantages, including the ability to accurately identify and classify anomalies using labeled data and the ability to identify specific types of anomalies. These techniques are also useful for detecting anomalies in surveillance and security applications. However, a significant amount of labeled data is required, and these techniques can be sensitive to environmental changes, affecting their accuracy.

Unsupervised learning techniques for SVAD offer advantages such as not requiring labeled data and the ability to detect anomalies in real-time. These techniques can also be used to identify patterns in the data that deviate from the norm and classify them as anomalies. However, unsupervised learning techniques are not able to identify specific types of anomalies and can also be sensitive to changes in the environment.

Semi-supervised learning techniques for SVAD can use labeled and unlabeled data, allowing for accurate identification and classification of anomalies. These techniques can also be used to identify specific types of anomalies and detect anomalies in real-time. However, semi-supervised learning techniques require significant labeled data and can also be sensitive to environmental changes.

In conclusion, supervised, unsupervised, and semi-supervised learning techniques each offer advantages and disadvantages when it comes to anomaly detection in SVAD. Each technique has its limitations, and the accuracy of the results can be affected by changes in the environment. Therefore, the choice of technique will depend on the specific needs of the application and the availability of labeled data.

3.2. Algorithms

We briefly outline the key classifications that result in a wide range of SVAD algorithms.

3.2.1. Statistics-Based Algorithms

Two main algorithms are used in video anomaly detection: parametric and non-parametric [43].

Parametric algorithms assume the data follow a specific probability distribution, such as a Gaussian distribution. These algorithms estimate the parameters of the distribution using the data and then use these parameters to calculate the likelihood of new data points. One popular parametric algorithm for video anomaly detection is the Gaussian Mixture Model (GMM). The GMM is a probabilistic model representing a dataset as a mixture of multiple Gaussian distributions. The algorithm estimates the parameters of the Gaussian distributions using the data and then uses these parameters to calculate the likelihood of new data points. If the likelihood of a new data point is below a certain threshold, it is considered an anomaly.

Non-parametric algorithms do not make any assumptions about the distribution of the data. Instead, these algorithms rely on the empirical distribution of the data, which is estimated using Kernel Density Estimation (KDE) [44]. One popular non-parametric algorithm for video anomaly detection is the Local Outlier Factor (LOF) [45]. The LOF is a density-based algorithm that calculates the local density of a data point by measuring the distance to its k-nearest neighbors. The algorithm then compares a data point's local density to its neighbors' density. The data point is considered an anomaly if the ratio is below a certain threshold. Several studies have been conducted on statistical-based algorithms, some of which are listed below: Gaussian Mixture Model (GMM), selective histogram of optical flow, Histogram of Magnitude and Momentum (HoMM), Histogram of the oriented Swarm (HoS), Histogram of Gradients (HoG), Bayesian, Fully-Convolutional-Network (FCNs)-based models, and Structural Context Descriptor (SCD). Some statistics-based studies are presented in Table 2.

3.2.2. Classification-Based Algorithms

One of the most-widely used methods for SVAD is classification-based methods, which involve training a classifier to distinguish between normal and anomalous video frames or segments.

The first step in using classification-based methods for video anomaly detection is to extract features from the video frames. These features can include spatial and temporal information, such as color, texture, motion, and object shape. Several feature extraction techniques have been proposed in the literature, including hand-crafted features, such as the Histogram of Oriented Gradients (HOG) and Scale-Invariant Feature Transform (SIFT), as well as in-depth learning-based features, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

Once the features have been extracted, the next step is to train a classifier to distinguish between normal and anomalous video frames or segments. Several classifiers have been proposed in the literature, including traditional machine learning classifiers, such as Support Vector Machines (SVMs), random forests, k-Nearest Neighbors (kNNs), and deep learning-based classifiers: CNNs and RNNs. The choice of the classifier will depend on the specific application and the type of features that have been extracted.

Table 2. Statistics-based methods.

Methods	**Summary**
HoMM [46]	The histogram of magnitudes is used to record the motion of objects. The anomalous motion of an object is represented by its momentum about the foreground region's occupancy. Feature descriptors for typical situations are learned in an unsupervised manner using K-means clustering. By measuring the distance between cluster centers and the test frame's feature vector, frame-level anomalies are discovered. The region and anything leaving that region are regarded as anomalous. **Datasets:** UCSD, UMN; **Techniques:** background subtraction, optical flow, K-means clustering.
Novel scheme based on SVDD [47]	Statistical histograms are used to model normal motion distributions. It combines motion detection and appearance detection criteria to find anomalous objects. They created a method based on Support Vector Data Description (SVDD), which creates a sphere-shaped boundary around the regular items to keep out anomalous ones. They took into account a fixed-dimension region, and anything left in that region is regarded as anomalous. **Datasets:** UCSD, UMN, Subway; **Techniques:** histogram model, optical flow, support vector data description.
Gaussian classifier [48]	Deep autoencoder networks and single-class image-level classification are proposed to detect event anomalies in surveillance videos. **Datasets:** UCSD, Subway; **Techniques:** Gaussian classifier, CNN, sparse autoencoder.
CoP [49]	The method called Consistency Pursuit (CoP) is based on the idea that normal samples have a very high correlation with each other, can span low-dimensional subspaces, and therefore, have strong mutual consistency with a large number of data points. **Datasets:** Hopkins155; **Techniques:** robust PCA, saliency map.

After the classifier has been trained, it can classify new video frames or segments as normal or anomalous. The classifier will output a score or probability for each frame or segment, indicating the likelihood that it is normal or anomalous. A threshold is usually set to make a final decision, and any frames or segments with a score below the threshold are considered anomalous.

One of the main advantages of classification-based methods for video anomaly detection is that they can be fine-tuned to a specific application by selecting appropriate features and classifiers. However, one of the main challenges is that these methods require a large amount of labeled training data to be effective. Additionally, they may be unable to detect anomalous events significantly different from the training data [43].

Several classification algorithms have been proposed in the literature on data science, which can be considered the most common in the field, and they were discussed in detail in [50]. Some commonly used algorithms are summarized as follows.

Support Vector Machine (SVM) is a widely used classification, regression, or other application method. An SVM generates a single hyperplane or a set of hyperplanes in a high or endless space. The goal is to separate the two classes using a hyperplane that reflects the greatest separation or margin. The larger the margin, the smaller the generalization error of the classifier is.

k-Nearest Neighbors (kNN) is a non-parametric supervised learning technique, also referred to as a "lazy learning" method. It maintains all occurrences that match the training set in an n-dimensional space, rather than focusing on building a large internal model. kNN uses data and employs similarity metrics to categorize new data points.

Decision Tree (DT) is a popular non-parametric SL approach. Both the classification and regression tasks are performed using DT learning techniques. The DT is a recursive operation; it starts with a single node and branches into a tree structure.

Some classification-based studies are shown in Table 3.

Table 3. Classification-based methods.

Methods	Summary
One-class classification [51]	A histogram of optical flow orientation is integrated with a one-class SVM to identify abnormal events. Modeling high-density scenes may be performed quickly and precisely using optical flow techniques. Pattern identification is performed after feature extraction to discriminate between regular and irregular activities. **Datasets:** UMN; **Techniques:** SVM, optical flow, histogram of optical flow orientation.
Asymptotic bounds [52]	The crowd escape anomaly is detected using statistical and deep learning algorithms that directly evaluate the pixel coordinates. **Datasets:** UCSD, Avenue, ShanghaiTech; **Techniques:** YOLOv3, GAN-based frame predictor, kNN.
Decision tree [53]	To detect abnormalities from video surveillance while precisely estimating the start and end times of the anomalous event, a decision-tree-enabled solution leveraging deep learning was created. **Datasets:** ImageNet, COCO; **Techniques:** decision trees, YOLOv5.
AE with kNN [54]	A new approach combines an AE-based method with single-class deep feature classification. An AE is trained using normal images; then, anomaly maps are embedded using a pre-trained CNN feature extractor. A one-class classifier with kNN is trained to calculate the anomaly score. **Datasets:** MVTec; **Techniques:** convolutional autoencoder, high-density embedding, one-class classification.
IGD [55]	There is a high probability of overfitting as abnormal datasets are insufficient. The Interpolated Gaussian Descriptor (IGD) method, an OCC model that learns a one-class Gaussian anomaly classifier trained with inversely interpolated training samples, is proposed to solve this problem. The IGD is used to learn more meaningful data descriptions from typical normal samples. The crowd escape anomaly is detected using statistical and deep learning algorithms that directly evaluate the pixel coordinates. **Datasets:** MNIST, Fashion MNIST, CIFAR10, MVTec AD; **Techniques:** Gaussian classifier.
Out of distribution [56]	A classifier that is simultaneously trained to give the GAN samples less confidence is used in conjunction with a GAN. Samples from each test distribution of anomalies are used to arrange the classifier and GAN. **Datasets:** CIFAR, tree-enabled, LSUN; **Techniques:** DNN, GAN, Kullback–Leibler, Gaussian distribution.

3.2.3. Reconstruction-Based Algorithms

Reconstruction-based methods operate under the presumption that normal data can be integrated into a lower-dimensional domain where normal samples and anomalies are represented in various ways [57].

An **Autoencoder (AE)** is a feed-forward neural network that includes an encoder and a decoder structure [58]. The objective is to train the network to capture the important parts of the input data and learn a lower-dimensional representation of the higher-dimensional data. The **Variational Autoencoder (VAE)** is a type of AE that includes an encoder network and a decoder network. The encoder network maps the input data to a low-dimensional latent space, while the decoder network maps the latent space back to the original data space. In this method, the VAE is trained on normal videos. The trained model is then used to reconstruct the input video, and the reconstruction error is calculated. Anomalies are detected by thresholding the reconstruction error. Any frame with a reconstruction error above a certain threshold is considered anomalous. The **Convolutional Autoencoder (CAE)** is also a type of AE consisting of convolution, deconvolution, pooling, and unpooling layers. The first two layer types may be found in the encoding step, whereas the others may be found in the decoding stage [59]. The **Variational Autoencoder (VAE)** is another type of AE that incorporates convolution, deconvolution, pooling, and unpooling layers. The first two layer types are used in the encoding step, while the others are used in the decoding stage [59].

Reconstruction-based methods are a variation of adversarial generative methods. **Generative-Adversarial-Network (GAN)**-based networks consist of two neural networks: a Generator (G) and a Discriminator (D) [58]. The generator network creates new examples in the target domain by mapping examples from the source domain to the target domain. The discriminator network then tries to distinguish between examples created by the generator and examples from the target domain. Through this process, the generator network learns to create examples indistinguishable from examples in the target domain.

In summary, reconstruction-based methods such as AEs and GANs have shown promising results in anomaly detection tasks by mapping normal data into a lower-dimensional domain and identifying anomalies based on the reconstruction error. Variants of AEs such as Conv AEs and variational AEs have also been utilized in this domain. These methods are part of a larger field of adversarial generative methods that include generative adversarial networks.

Some reconstruction-based studies are shown in Table 4.

Table 4. Reconstruction-based methods.

Methods	Summary
ST-AE [60]	The spatiotemporal AE comprises one encoder and two decoders of 3D convolutional layers. It employs parallel training of decoders with monochrome frames, which is noteworthy compared to the distillation process. **Datasets:** Traffic, UCSD, Avenue; **Techniques:** CNN, autoencoder.
AMDN [61]	The appearance and motion DeepNet model employs AEs and a modified two-stream network with an additional third stream to improve detection performance. The two-stream method has two major drawbacks: the requirement for a pre-processing technique, such as optical flow, which may be costly for real-world applications, and multiple networks for inference. **Datasets:** Train, UCSD; **Techniques:** one-class SVM, optical flow.
GMFC-VAE [62]	The Gaussian mixture fully convolutional-variational AE uses the conventional two-stream network technique and uses a variational AE to enhance its feature extraction capability. This method estimates the appearance and motion anomaly score before combining the two clues to provide the final detection results. **Datasets:** Avenue, UCSD; **Techniques:** convolutional autoencoder, Gaussian mixture model.
OF-ConvAE-LSTM [63]	This method uses the convolutional AE and long short-term memory to detect anomalies. The framework produces the error function and reconstructed dense optical flow maps. **Datasets:** Avenue, UCSD; **Techniques:** convolutional autoencoder, LSTM, optical flow.
Temporal cues [64]	A conditional GAN is trained to learn two renderers that map pixel data to motion and vice versa. As a result, normal frames will have little reconstruction loss, while anomalous frames will have significant reconstruction loss. **Datasets:** Avenue, ShanghaiTech; **Techniques:** GAN, LSTM, optical flow.
Ada-Net [65]	An attention-based autoencoder using contentious learning is proposed to detect video anomalies. **Datasets:** UCSD, Avenue, ShanghaiTech; **Techniques:** GAN, autoencoder.
Adversarial 3D CAE [66]	A 3D CAE-based competitor anomalous event detection method is proposed to obtain the maximum accuracy by simultaneously learning motion and appearance features. It was developed to explore spatiotemporal features that help detect anomalous events in video frames. **Datasets:** UCSD, Avenue, Subway, ShanghaiTech; **Techniques:** convolutional autoencoder.
Conv-AE + U-Net [67]	A two-stream model is created that learns the connection between common item appearances and their related motions. A single encoder is paired with a U-net decoder to predict motion and a deconvolution decoder that reconstructs the input frame under the control of the l_p reconstruction error loss terms using a single frame as the input. **Datasets:** UCSD, Avenue, Subway, Traffic; **Techniques:** convolutional autoencoder.

3.2.4. Prediction-Based Algorithms

Prediction-based techniques can identify anomalies by assessing the difference between the expected and actual spatiotemporal properties of a feature descriptor [57]. These models assume that normal activities are predictable, and any deviation from the prediction indicates an anomaly. They typically use a **Recurrent Neural Network (RNN)** to predict the next frame in the sequence, given the previous frames. The model minimizes the difference between the predicted frame and the ground truth during training. Here, are some commonly used algorithms:

Long Short-Term Memory (LSTM) is the most-widely used neural array model, combining the principles of the forget gate, entry gate, and exit gate and successfully avoiding back-propagation errors caused by vanishing/exploding gradients.

The **convolutional LSTM** is an LSTM variation that addresses the precipitation nowcasting problem. In contrast to LSTM, convolution operations are employed to calculate the feature maps instead of matrix operations, resulting in a significant decrease in the count of the training parameters of the model [59].

Another prediction-based approach is the **Vision Transformer (ViT)** [68–70]. The ViT model combines CNNs and transformers to extract spatiotemporal features from video data and model the temporal relationships between these features. This approach effectively captures long-term dependencies in the video data and is especially useful for detecting anomalies.

In summary, RNN-based prediction techniques are effective at detecting anomalies by comparing the expected and actual spatiotemporal properties of a feature descriptor. LSTM is the most-widely used and successful neural array model, while the convolutional LSTM and ViT are variations that address specific problems.

Some prediction-based studies are shown in Table 5.

Table 5. Prediction-based methods.

Methods	**Summary**
FFP [57]	Spatial and motion constraints are used to estimate the future frame for normal events in addition to density and gradient losses. **Datasets:** UCSD, ShanghaiTech, Avenue; **Techniques:** GAN, optical flow.
Deep BD-LSTM [71]	A model combining CNN and bidirectional LSTM is proposed to recognize human movement in video sequences. **Datasets:** YouTube 11 Actions, UCF-101, HMDB51; **Techniques:** LSTM, CNN.
LSTM [72]	By using the effective gradient and quadratic-programming-based training methods, the parameters of the LSTM architecture and the support vector data description algorithm are trained and optimized. **Datasets:** Avenue, Subway, ShanghaiTech, UCSD; **Techniques:** LSTM, one-class SVM.
SSPCAB [73]	A Self-Supervised Predictive Convolutional Attentive Block (SSPCAB) is proposed, which can be easily incorporated into various anomaly detection methods. The block acquires the ability to recreate the masked area utilizing contextual information for each site where the dilated convolutional filter is applied. **Datasets:** Avenue, MVTec AD, ShanghaiTech; **Techniques:** CNN, convolutional attentive block.
Spatiotemporal feature extraction [74]	A neural network built with transaction blocks, including dictionary learning, feature learning, and sparse representation, is proposed. A novel long short-term memory was also proposed and reformulated using an adaptive iterative hard-thresholding technique (LSTM). **Datasets:** UCSD, Avenue, UMN; **Techniques:** LSTM, RNN-based sparsity learning.
ISTL [75]	An Incremental Spatiotemporal Learner (ISTL) model is proposed to address the difficulties and limitations of anomaly detection and localization to keep track of the changing character of anomalies through active learning using fuzzy aggregation. **Datasets:** UCSD, Avenue; **Techniques:** convolutional LSTM, fuzzy aggregation.
Residual attention-based LSTM [76]	Using a light-weight CNN and an attention-based LSTM for anomaly detection reduces the time complexity with competitive accuracy. **Datasets:** Avenue, UCF-Crime, UMN; **Techniques:** residual attention-based LSTM.
CT-D2GAN [68]	A Conv-transformer is used to perform future frame prediction. Dual-discriminator adversarial training maintains local consistency and global coherence for future frame prediction. **Datasets:** UCSD Ped2, Avenue, ShanghaiTech; **Techniques:** GANs; transformer; CNN.
ViT-based framework [69]	Using a ViT model for anomaly detection involves processing a single frame as one patch. This approach yields good performance on the SVAD task while maintaining the advantages of the transformer architecture. **Datasets:** UCSD Ped2, Avenue, ShanghaiTech; **Techniques:** vision transformer.

3.2.5. Other Algorithms

Two clustering methods are available. Their argument is based on the idea that normal data are clustered, whereas anomalous data are not [77] connected to any cluster. The second type is predicated on the idea that, whereas anomalies belong to tiny clusters, typical data instances belong to massive or dense clusters. Fuzzy traffic density and flow are built using fuzzy theory to identify abnormalities in complicated traffic videos [78]. Heuristic techniques make decisions regarding anomalies based on feature values, geographical locations, and contextual data intuitively [79]. However, many real-world systems do not rely only on one technology. Using a light-weight CNN and an attention-based LSTM for anomaly detection reduces the time complexity with competitive accuracy.

3.2.6. Analysis of Algorithms

Statistics-based algorithms assume that normal behavior follows a certain statistical pattern, and any deviation from this pattern is considered an anomaly. They are simple and efficient and can detect anomalies in real-time without requiring a large amount of training data. However, they may not be effective at detecting novel anomalies or anomalies that do not follow a statistical pattern.

Classification-based algorithms use machine learning techniques to classify behavior or events as normal or abnormal based on labeled training data. They can detect novel anomalies and adapt to changing environments with high accuracy. However, they require a large amount of training data, and the labeling process can be time-consuming and costly.

Reconstruction-based algorithms reconstruct normal behavior or events and compare them to the actual behavior or events to detect subtle anomalies. They do not require labeled training data, but can be computationally expensive and may not be suitable for real-time anomaly detection.

Prediction-based algorithms use machine learning techniques to predict future behavior or events based on past behavior or events. Any deviation from the predicted behavior or events is considered an anomaly. They can detect anomalies before they occur, which can be useful in preventing security threats or safety issues. However, they require a large amount of training data, and the accuracy of the predictions may decrease over time as the environment changes.

In conclusion, the selection of the algorithm depends on the specific application and requirements. Statistics-based algorithms are simple and efficient, but may not detect novel anomalies. Classification-based algorithms have a high accuracy rate, but require a large amount of training data. Reconstruction-based algorithms can detect subtle anomalies, but can be computationally expensive. Prediction-based algorithms can detect anomalies before they occur, but require a large amount of training data, and the accuracy of predictions may decrease over time. Table 6 shows and overview of the algorithms.

Table 6. Overview of algorithms.

Algorithms	Strengths	Weaknesses
Statistics-based	Generally, they are suitable for real-time applications as they are simple and computationally efficient. Subtle or complex anomalies, such as those involving spatial or temporal relationship changes, cannot be detected.	They cannot detect subtle or complex anomalies, such as changes in spatial or temporal relationships. False alarms may also occur when the data distribution deviates from a Gaussian distribution.
Classification-based	High-dimensional datasets can be handled, and robustness to noise can be exhibited. Anomalies can be learned to be detected based on labeled training data, allowing them to adapt to changes in the data distribution over time. High-dimensional datasets can be handled, and global and local anomalies can be detected.	Generally, they could be improved for real-time applications, being more computationally expensive than statistics-based algorithms. A large amount of labeled training data is also required, which can be difficult and time-consuming. Only the anomalies encountered in the training data can be detected, and new unseen anomalies cannot be detected.

Table 6. *Cont.*

Algorithms	Strengths	Weaknesses
Reconstruction-based	A compact representation of normal data can be learned, allowing subtle or complex anomalies to be detected. High-dimensional datasets can be handled, and global and local anomalies can be detected.	They are generally less well-suited for real-time applications as they are more computationally expensive than statistics-based algorithms. A large amount of normal data for training is also required, which can be difficult to obtain in some scenarios. They are not robust to noise, and false alarms may occur when the data are noisy or corrupted.
Prediction-based	Temporal dependencies in the data can be leveraged to detect anomalies, making them well-suited for time series data. High-dimensional datasets can be handled, and global and local anomalies can be detected.	They are generally less well-suited for real-time applications as they are more computationally expensive than statistics-based algorithms. A large amount of normal data for training is also required, which can be difficult to obtain in some scenarios. Anomalies involving spatial or temporal relationship changes may be difficult to detect. They do not possess robustness to noise, and false alarms can be produced when the data are noisy or corrupted.

4. Analysis of the Existing Datasets

Due to the inherent rarity of anomalies, more real-world datasets need to have real anomalies. Natural or unnatural datasets have been created for researchers to use.

In some studies, video anomaly datasets have been categorized. For example, in a study [1], datasets were divided into three main categories: **heterogeneous, specific, and others**. By their nature, datasets containing various anomalies are categorized as heterogeneous, while datasets containing a certain anomaly are categorized as specific. Heterogeneous datasets consist of a greater variety of anomalies and scene variability, and specific datasets have mainly specific types of anomalies.

Table 7 shows the categories of video surveillance datasets examined in this study.

Table 7. Category of video surveillance datasets.

Heterogeneous	Specific
UCSD	Subway
Avenue	UMN
UCF-Crime	
ShanghaiTech	

4.1. CUHK Avenue Dataset

There are 21 test videos of anomalous events and 16 training videos of normal events in the Avenue dataset [80]. The Chinese University of Hong Kong (CUHK) is where the videos were shot, and they were released in 2013. A total of 47 abnormal events are included, including walking in the wrong direction, running, dancing, throwing objects, and similar anomalous actions. Clips recorded outdoors have a rate of 25 Frames Per Second (FPS) and a resolution of 640 × 360. The ground truth is available at both the pixel level and the frame level. Some examples of abnormal frames are shown in Figure 1.

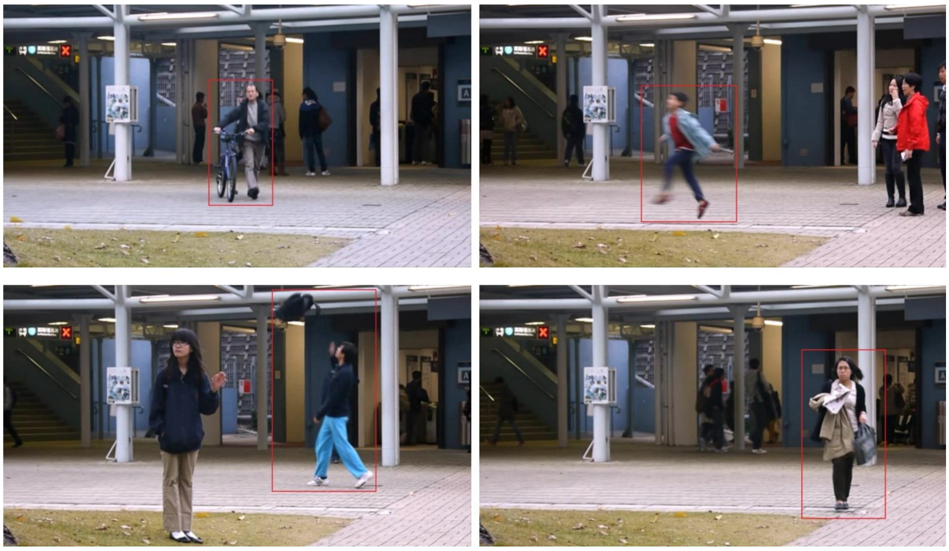

Figure 1. Examples of abnormal frames from the CUHK Avenue dataset are presented.

4.2. Subway Dataset

The Subway dataset [81], consisting of two sub-datasets, Subway Entrance and Subway Exit, was published in 2008. This dataset contains anomalous events such as going in the incorrect direction, avoiding payment, and similar actions. Some examples of abnormal frames are shown in Figure 2. The videos were produced at 25 FPS and a 512 × 384 resolution. The entrance subset contains 1 h 36 min of video (144,250 frames), and the exit subset contains 43 min (64,901 frames). Subway Entrance has 66 unusual events, and Subway Exit has 19 unusual events. Two restrictions apply to this dataset: the count of anomalies and predictable spatial localizations [48]. The ground truth is available only at the frame level.

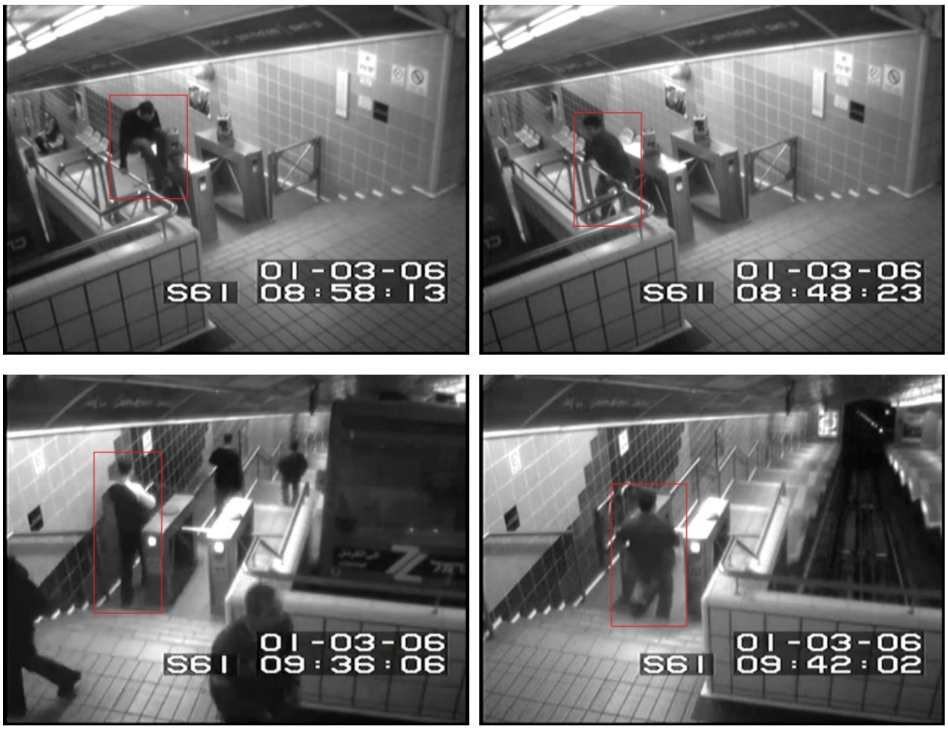

Figure 2. Examples of abnormal frames from the Subway dataset are presented. Abnormal frames are shown in the first row from the Subway Entrance, and abnormal frames are shown in the second row from the Subway Exit sub-datasets.

4.3. UCF-Crime Dataset

The UCF-Crime dataset [33] includes robbery, fighting, shooting, shoplifting, abuse, explosion, accident, arrest, burglary, arson, vandalism, stealing, and assault. Table 8 shows the number of videos by category in the UCF-Crime dataset. There are 950 normal and 950 abnormal videos taken from the actual world. Some videos are poor-quality as they are made from real-world footage. In addition, some videos may have anomalies that fall into multiple categories.

The ground truth is available at both the clip and frame levels. Some examples of normal and abnormal frames are shown in Figure 3.

Table 8. The number of videos in the UCF-Crime dataset.

Category	**UCF-Crime**	**HR-Crime**
Abuse	50	38
Arrest	50	42
Arson	50	48
Assault	50	47
Burglary	100	96
Explosion	50	26
Fighting	50	39
Road accident	150	68
Robbery	150	145
Shooting	50	46
Shoplifting	50	50
Stealing	100	98
Vandalism	50	46
Normal	950	782
Total	1900	1571

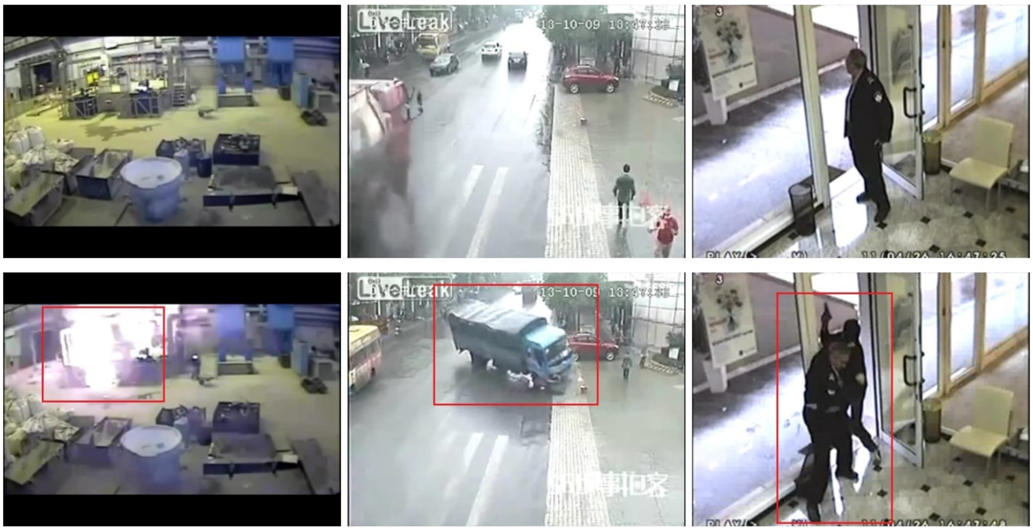

Figure 3. Examples of frames from the UCF-Crime dataset are presented. Normal frames are shown in the first row, and abnormal frames are in the second row.

4.4. UCSD Dataset

The videos in the UCSD dataset [82] contain events recorded in various crowd scenes. This was first published in 2010. The dataset includes anomalous actions, such as walking on grass, vehicles moving on the sidewalk and street, and unexpected behaviors such as skateboarding. Examples of abnormal frames are shown in Figure 4. In the Pedestrian (Ped) 1 sub-dataset, there were 34 training video samples and 36 test video samples, whereas the Ped2 sub-dataset has 16 training video samples and 12 test video samples. The Ped2 sub-dataset is generally used in studies. Ped1 has a low resolution and frame distortion [83]. The ground truth is available at both the pixel and frame levels.

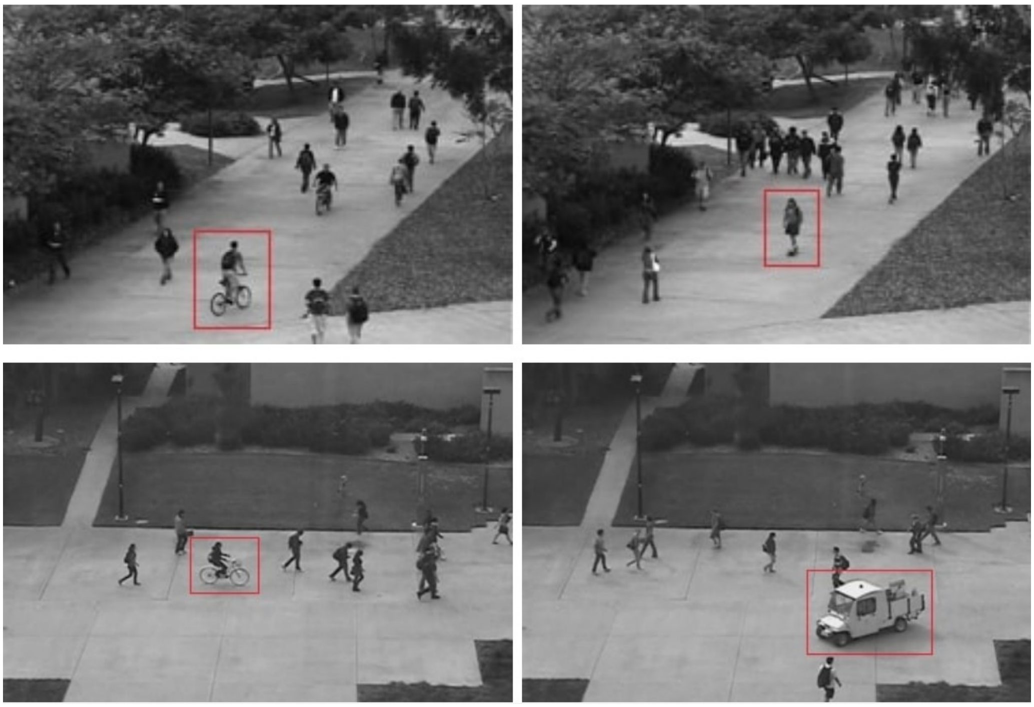

Figure 4. Examples of abnormal frames from the UCSD dataset are presented. Abnormal frames are shown in the first row from Ped1, and abnormal frames are shown in the second row from Ped2.

4.5. UMN Dataset

The UMN Dataset [84] is a collection of 11 short videos depicting a panicking crowd's abnormal movements. The dataset includes three distinct scenes, comprising two outdoor and one indoor environment. Each video in the dataset concludes with a sudden escape of individuals, depicted as walking in normal directions. It is important to note that the ground truth, or the labeled information, is only available at the frame level.

The UMN dataset is significant in abnormal behavior detection as it provides a unique opportunity to study the dynamics of panicking crowds. The dataset is particularly useful for developing anomalous actions, such as walking on grass and vehicles moving on the sidewalk. Items are essential for designing surveillance systems that identify and respond to potential safety hazards in crowded spaces.

In conclusion, the UMN dataset is a valuable resource for researchers and practitioners in abnormal behavior detection. The dataset provides a realistic representation of panicking crowds and can be used to test and improve the performance of abnormal behavior detection algorithms in crowded environments. Figure 5 shows examples of both normal and abnormal frames.

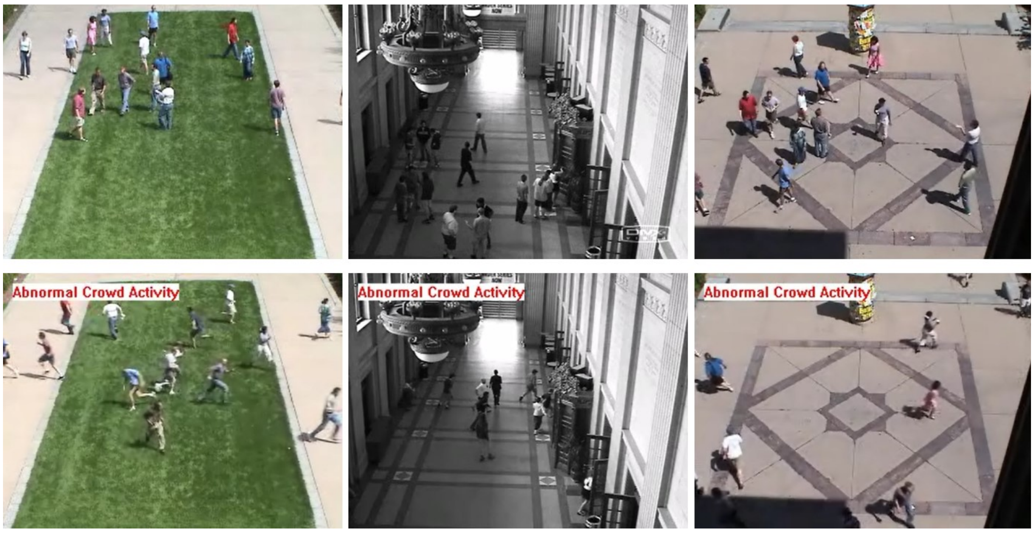

Figure 5. Examples of frames from the UMN dataset are presented. Normal frames are shown in the first row, and abnormal frames are in the second row.

4.6. ShanghaiTech Dataset

The ShanghaiTech dataset [34] contains 330 training videos and 107 test videos. Thirteen scenes in the dataset feature complex lighting and camera angles. This dataset has 42,883 test frames and 274,515 training frames, each with a resolution of 480 × 856 pixels. Examples of abnormal frames are shown in Figure 6. The ground truth is available at both the pixel and frame levels.

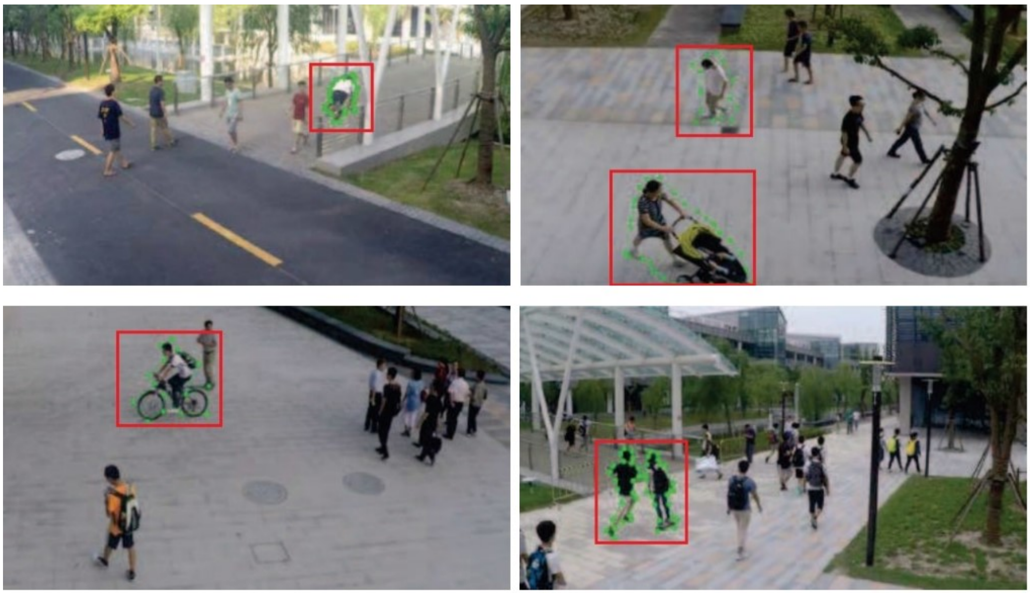

Figure 6. Examples of abnormal frames from the ShanghaiTech dataset.

4.7. HR-Crime Dataset

The HR-Crime dataset [85] contains 782 Human-Related (HR) normal videos and 789 HR abnormal videos. The ground truth is only available at the frame level, with 239 test videos annotated. HR-Crime is a subset of the UCF-Crime dataset containing the same categories. Table 8 shows the number of videos by category in the HR-Crime dataset. Since it is a new dataset, many studies have not used it. Some examples of normal and abnormal frames are shown in Figure 7.

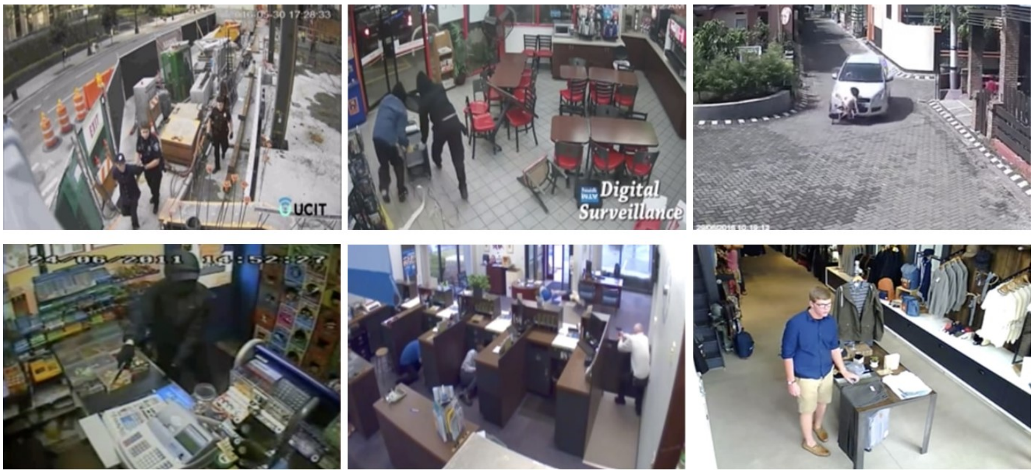

Figure 7. Examples of abnormal frames from the HR-Crime dataset are presented.

4.8. Comparison of Surveillance Datasets

The anomaly types in the datasets are listed in Table 9. It has been observed that humans mostly cause these anomalies, and the diversity of anomalies consisting of inanimate objects is low. In addition, when we examine Table 10, we see that the datasets are not recent and are generally low-resolution, and the ground truth data represent frame-level anomalies. When the structure and contents of the datasets are examined, we can categorize the datasets based on the recording environment and ground truth labeling.

Table 9. Anomalies of SVAD datasets.

Datasets	Anomalies
Avenue	Throwing objects, loitering, and running
UCF-Crime/HR-Crime	Abuse, arson, arrest, assault, burglary, explosion, road accident, fighting, shooting, robbery, shoplifting, stealing, and vandalism
UMN	Unusual behavior of the crowd and running
ShangaiTech	Chasing, brawling, running, and non-pedestrian assets, such as skaters and bicycles, on the pedestrian path
Subway Entrance/Exit	Wrong direction, no payment, unusual interactions, and loitering
UCSD Ped1/Ped2	Passage of non-pedestrian assets, such as vehicles and bicycles, from the pedestrian path

The CUHK Avenue, UCSD Ped2, and ShanghaiTech datasets are the most-suitable for evaluating anomaly detection performance in surveillance videos. Future studies should consider testing the proposed methods on these datasets to determine their effectiveness.

Table 10. Comparison of existing SVAD datasets.

Datasets	Release Year	Ground Truth	Resolution	FPS	Environment	Normal Frames	Abnormal Frames	Training Frames	Test Frames	Total Frames
Avenue	2013	frame/pixel	640 × 360	25	outdoor	26,832	3820	15,328	15,324	30,652
UCF-Crime	2018	frame/clip	320 × 240	30	indoor/outdoor	N/A	N/A	12,631,211	1,110,182	13,741,393
HR Crime	2021	frame/clip	320 × 240	30	indoor/outdoor	485,227	335,378	N/A	N/A	N/A
UMN	2009	frame	320 × 240	30	indoor/outdoor	6165	1576	N/A	N/A	7740
ShangaiTech	2016	frame/pixel	856 × 480	-	outdoor	300,308	17,090	274,515	42,883	317,398
Subway **Entrance** **Exit**	2008	frame	512 × 384	25	indoor	132,138 60,410	12,112 4491	76,453 22,500	67,797 42,401	144,250 64,901
UCSD **Ped1** **Ped2**	2010	frame/pixel	238 × 158 360 × 240	-	outdoor	9995 2924	4005 1636	6800 2550	7200 2010	14,000 4560

4.8.1. Based on Recording Environment

Various applications, such as public monitoring, are used indoors and outdoors. However, applications such as traffic monitoring are primarily used in outdoor environments. Examples of outdoor environments are train stations, parks, crosswalks, streets, etc.; examples of indoor environments are shopping malls, schools, factories, subways, etc.

Datasets produced in indoor environments are less affected by light changes and lens distortions than datasets produced in outdoor environments. Outdoor environments are heavily affected by weather events such as sun, clouds, rain, and snow.

We can separate the datasets produced in indoor and outdoor environments. Some datasets consist of videos recorded in both environments. Table 11 shows the distribution of the datasets by recording the environment. In selecting datasets, it is useful to consider the effects of external factors on the datasets.

Table 11. Based on the recording environment.

Indoor	Outdoor
UCF-Crime	Avenue
UMN	UCF-Crime
Subway Entrance	UMN
Subway Exit	ShanghaiTech
	UCSD Ped1
	UCSD Ped2

4.8.2. Based on Color Space

Previous research has demonstrated that, when constructing a three-channel model, it is crucial to select an appropriate dataset carefully. Specifically, it has been noted that only the UCSD and Subway datasets are represented in grayscale. In contrast, the remaining datasets (UMN, Avenue, UCF-Crime, and ShanghaiTech) are represented in the RGB color space. This is an important consideration when selecting a dataset for use in constructing a three-channel model, as the dataset's color space distribution can significantly impact the overall accuracy and performance of the model. Table 12 shows the color space distribution of the datasets.

Table 12. Based on the color space.

Grayscale	RGB
UCSD Ped1	Avenue
UCSD Ped2	ShanghaiTech
Subway Entrance	UMN
Subway Exit	UCF-Crime

4.8.3. Based on Ground Truth Labeling

The process of labeling datasets can take various forms, with different approaches being employed depending on the specific context and requirements of the study. However, it is generally acknowledged that there are better methods than labeling at the level of entire clips. This is due to the difficulty of explaining such labeling using real-world examples and the likelihood of inaccuracies in the resulting data.

A more widely accepted method is frame-based labeling, which provides a more accurate result by reducing the error rate. In certain cases, anomalies may occur at specific positions within the analyzed scene. In these instances, labeling at the pixel level may be employed to capture the anomaly in a particular spatial area of the scene. Table 13 illustrates the various labeling types commonly employed in the field.

Table 13. Based on ground truth labeling.

Clip-Level	Frame-Level	Pixel-Level
UCF-Crime	Avenue UCF-Crime UMN ShanghaiTech UCSD Ped1/Ped2 Subway Entrance/Exit	Avenue ShanghaiTech UCSD Ped1/Ped2

All these different forms of labeling are called the "ground truth", which refers to information that is assumed to be true or correct. The various types of datasets that are labeled in this way include clip-level (such as UCF-Crime), frame-level (such as UMN, Avenue, UCF-Crime, ShanghaiTech, UCSD, and Subway), and pixel-level (such as Avenue, ShanghaiTech, and UCSD).

5. Evaluation of the Performance of Existing Applications

Performance metrics are usually measured at the frame or pixel level concerning the ground truth data in datasets. If an abnormal event is identified in a frame, it is categorized as abnormal. The frame-level criterion considers only the entire frame. Instead of merely determining whether a frame contains abnormal events, the pixel-level analysis seeks to identify anomalous events within the frame. Therefore, the pixel-level criterion is better for assessing the quality of an algorithm [86]. In pixel-level evaluations, the locations of the detected objects are crucial. In rare cases, the detection can be performed at the clip level. However, this criterion is not preferred because it needs to provide sufficient evaluation details.

In anomaly detection studies, most models use the Receiver Operating Characteristic Curve (ROC) and its associated Area Under the Curve (AUC) as metrics. These metrics can be calculated by using a confusion matrix. Table 14 shows the general structure of the confusion matrix and related evaluation measures. In addition, as seen in Figure 8, the ROC curve shows the correlation between the false positive and true positive rates for different parameter cut-off values. Another metric used in models is the Equal Error Rate (EER). The EER selects the best threshold on the ROC curve to maximize the TPR and minimize the FPR. However, according to recent studies [87,88], EER evaluation criteria result in a severely unbalanced sample of normal and abnormal events. The performance comparison of framework-level applications using the AUC and EER is summarized in Table 15. The methods were published between 2016 and 2023.

Table 14. Confusion matrix and related evaluation measures.

		Actual Class	
		True	False
Predicted Class	True	True Positives (TPs)	False Positives (FPs)
	False	False Negatives (FNs)	True Negatives (TNs)

$$\textit{True Positive Rate (TPR)} = TP/(TP + FN)$$
$$\textit{False Positive Rate (FPR)} = FP/(FP + TN)$$

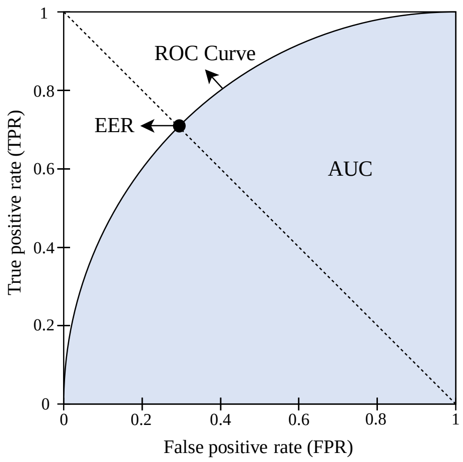

Figure 8. ROC curve—AUC.

Table 15. AUC/EER comparison of the accuracies of the various techniques with frame-level criterion.

		CUHK Avenue		Subway Entrance		Subway Exit		UCSD Ped1		UCSD Ped2		Shanghai Tech		UCF Crime	
Year	**Methods**	**AUC**	**EER**	**AUC**	**EER**	**AUC**	**EER**	**AUC**	**EER**	**AUC**	**EER**	**AUC**	**EER**	**AUC**	**EER**
2016	SL-HOF+FC [89]	-	-	-	-	-	-	87.4	18.0	95.07	9.0	-	-	-	-
	ConvAE [90]	70.2	25.1	**94.3**	26.0	80.7	9.9	81.0	27.9	90.0	21.7	-	-	-	-
2017	ConvLSTM-AE [91]	77.0	-	93.3	-	87.7	-	75.5	-	88.1	-	-	-	-	-
	S-RBM [92]	78.7	27.2	-	-	-	-	70.2	35.4	86.4	16.4	-	-	-	-
	ST-AE [93]	80.3	20.7	84.7	23.7	94.0	9.5	89.9	12.5	87.4	12.0	-	-	-	-
	3D gradients+conv5 [88]	80.6	-	70.6	-	85.7	-	68.4	-	82.2	-	-	-	-	-
2018	Baseline [57]	85.1	-	-	-	-	-	83.1	-	95.4	-	72.8	-	-	-
	WCAE-LSTM [94]	85.7	-	-	-	-		85.1	-	92.6	-	-	-	-	-
	NNC [95]	88.9	-	93.5	-	**95.1**	-	-	-	-	-	-	-	-	-
2019	TSN [96]	-	-	-	-	-	-	-	-	92.8	-	-	-	78.0	-
	MemAE [97]	83.3	-	-	-	-	-	-	-	94.1	-	71.2	-	-	-
	sRNN-AE [98]	83.4	-	85.3	-	89.7	-	-	-	92.2	-	69.6	-	-	-
	Attention [99]	86.0	-	-	-	-	-	83.9	-	96.0	-	-	-	-	-
	AnomalyNet [74]	86.1	22.0	-	-	-	-	83.5	25.2	94.9	10.3	-	-	-	-
	BMAN [100]	90.0	-	-	-	-	-	-	-	96.6	-	76.2	-	-	-
	3D ResNet [101]	-	-	-	-	-	-	-	-	-	-	-	-	76.6	-
2020	Dual D-b GAN [102]	84.9	-	-	-	-	-	-	-	95.6	-	73.7	32.2	-	-
	r-GAN [103]	85.8	-	-	-	-	-	86.3	-	96.2	-	77.9	-	-	-
	FFP+MS SSIM+FCN [104]	85.9	20.4	-	-	-	-	84.5	22.3	95.9	11.1	73.5	32.5	-	-
	Deep AE [105]	86.0	-	-	-	-	-	-	-	96.5	-	73.3	-	-	-
	Siamese CNN [20]	87.2	18.8	-	-	-	-	86.0	23.3	94.0	14.1	-	-	-	-
	P w/ Mem [106]	88.5	-	-	-	-	-	-	-	97.0	-	70.5	-	-	-
	Self-reasoning [107]	-	-	-	-	-	-	-	-	94.4	-	**84.1**	-	79.5	-
2021	Spatial+temporal [108]	80.3	-	87.3	-	90.8	-	-	-	84.5	-	-	-	-	-
	HMCF [109]	83.2	20.2	-	-	94.2	12.6	93.5	17.4	93.7	18.8	-	-	-	-
	Multi-task L. [110]	86.9	-	-	-	-	-	-	-	92.4	-	83.5	-	-	-
	Decoupled Arch. [111]	88.8	-	-	-	84.7	-	95.1	-	92.4	-	74.2	-	-	-
	GMM-DAE [112]	89.3	-	-	-	-	-	-	-	96.5	-	81.2	-	-	-
	DMRMs [113]	-	-	-	-	-	-	-	-	-	-	68.5	-	81.2	-
2022	Att-b residual AE [16]	86.7	-	-	-	-	-	-	-	97.4	-	73.6	-	-	-
	CR-AE [114]	-	-	-	-	-	-	-	-	95.6	-	73.1	-	-	-
	EADN [115]	**97.0**	-	-	-	-	-	93.0	-	97.0	-	-	-	**98.0**	-
	DR-STN [116]	90.8	11.0	-	-	-	-	**98.8**	2.9	97.6	6.9	-	-	-	-
	AMSRC [117]	93.8	-	-	-	-	-	-	-	99.3	-	76.6	-	-	-
2023	DMAD [118]	92.8	-	-	-	-	-	-	-	**99.7**	-	78.8	-	-	-
	Adjacent frames [119]	90.2	-	-	-	-	-	-	-	96.5	-	83.1	-	-	-

Higher AUC and lower EER are better.

For the CUHK Avenue dataset, the prediction-based EADN [115] method had the highest AUC of 97.0%. The AMSRC method [117] achieved the second-best result with an

AUC of 93.8%. However, the ConvAE method [90] had the worst result with an AUC of 70.2% and an EER of 25.1%.

For the Subway Exit dataset, the NNC [95] method demonstrated the best performance with an AUC of 95.1%. The second-best method was HMCF [109] with an AUC of 94.2%. The worst-performing method on this dataset was ConvAE [90] with an AUC of 80.7% and an EER of 9.9%. For the Subway Exit dataset, the NNC method [95] showed the best performance with an AUC of 95.1%. The HMCF method [109] achieved the second-best result with an AUC of 94.2%. However, the ConvAE method [90] had the worst result with an AUC of 80.7% and an EER of 9.9%.

For the UCSD Ped1 dataset, the reconstruction-based DR-STN method [116] exhibited the best performance with an AUC of 98.8%. The second-best method was decoupled Arch. [111] with an AUC of 95.1%. The worst-performing method on this dataset was S-RBM [92] with an AUC of 70.2%. For the UCSD Ped2 dataset, the reconstruction-based DMAD [118] method achieved the best performance with an AUC of 99.7%. The second-best method was AMSRC [117] with an AUC of 99.3%, and no EER value was reported. The worst-performing method on this dataset was 3D gradients+conv5 [88] with an AUC of 82.2%.

For the Shanghai Tech dataset, the self-reasoning method [107] showed the best performance with an AUC of 84.1%. The multi-task L. method [110] had the second-best result with an AUC of 83.5%, but the sRNN-AE method [98] had the worst result with an AUC of 69.6%.

Finally, for the UCF-Crime dataset, the EADN method [115] demonstrated the best performance with an AUC of 98.0%. However, the 3D ResNet method [101] had the worst result with an AUC of 76.6%.

The results in Table 15 demonstrate that there is no one-size-fits-all solution to anomaly detection, and the choice of the method and architecture depends on the specific dataset and the task at hand. The results also highlight the importance of benchmarking and comparing methods on multiple datasets to obtain a more complete understanding of their strengths and weaknesses.

6. Discussion

In this section, we discuss the current developments, existing limitations, and current challenges, as well as provide insights into future directions.

6.1. Current Developments

Significant advancements have been witnessed in the field of SVAD in recent years with the use of AI techniques. This sub-section discusses current developments in AI techniques in SVAD.

UL and SSL algorithms (AE, GAN, etc.) are more successful than SL algorithms (GMM, HMM, etc.) in SVAD. This is due to the lack of labeled data, the presence of unknown anomalies, the flexibility to adapt to different situations, and the ability to handle large datasets. SL algorithms require labeled data for training, making it difficult to apply them in scenarios where labeled data are scarce. UL and SSL algorithms do not require labeled data and can detect unknown anomalies, making them more flexible in adapting to different situations. Additionally, they can handle large amounts of data, making them more suitable for surveillance video feeds that generate massive amounts of data.

AI techniques, such as CNNs, LSTMs, and GANs, are extensively employed in the field of SVAD and have demonstrated highly promising outcomes. In particular, GANs can learn from large datasets and generate new data similar to the original, making them particularly well-suited to detecting anomalies in surveillance footage where unusual events may not have been previously observed. This approach, known as generative modeling, can be more effective than traditional supervised learning methods that rely on labeled data, which can be scarce and costly to obtain. By training a GAN on a large dataset of normal activities, it is possible to generate a model that can identify anomalous behavior in real-time. These techniques can learn complex patterns and relationships directly from the data. This allows them to detect subtle anomalies that may be missed by traditional statistical or classification-based algorithms.

Reconstruction-based (CAE, GAN, etc.) and prediction-based (LSTM, ViT, etc.) algorithms are more effective than statistics-based (GMM, etc.) and classification-based (kNN, etc.) algorithms in SVAD in terms of their robustness, flexibility, and efficiency. These algorithms are designed to learn and extract features from video data, making them less sensitive to irrelevant information and better able to adapt to changing environmental conditions. In addition, these algorithms are efficient in processing large amounts of data quickly, which makes them suitable for surveillance applications that require real-time anomaly detection. On the other hand, statistics-based and classification-based algorithms may require significant computational resources and may not be as effective at detecting anomalies in real-time. Overall, reconstruction-based and prediction-based algorithms offer better performance and reliability in SVAD.

Furthermore, there has been a growing interest in developing real-time anomaly detection systems that can detect anomalies in surveillance videos in real-time. Real-time anomaly detection is important for critical applications, such as security and public safety, where the timely detection of anomalies can prevent serious incidents.

SVAD's lack of available datasets can be attributed to several factors, including concerns around data privacy, the high cost and time requirements for collecting and labeling data, difficulties in detecting anomalies in real-world scenarios, and a general lack of standardization within the field. These factors pose significant challenges for the development and evaluation of algorithms in SVAD. However, there are initiatives underway to address these challenges. Some of these initiatives include the creation of privacy-preserving data collection techniques and standardized datasets for benchmarking.

Determining which dataset is better can be challenging due to the presence of global and local features that indicate abnormal events. Local features are emphasized in datasets such as UCSD, ShanghaiTech, and Subway, while global features are more prominent in UCF-Crime. The ground-truth values of these datasets are crucial for developing high-performance methods that require low computational complexity and memory requirements. It is important to determine the starting point of an abnormal situation, such as whether it begins with the subject's entrance or at the time specified by the dataset producer. For instance, it is worth considering whether the detection of a ball being thrown or the appearance of a ball on stage signifies the start of an abnormal situation.

Generally, in SVAD datasets, the training data typically contain only normal data, while the testing data contain both normal and aberrant data. This is due to the low frequency of abnormal events. Most methods for SVAD datasets are prediction- and reconstruction-based. However, the content of the training and testing data could be more clearly defined.

6.2. Limitations and Challenges

The field of SVAD poses several challenges, such as processing vast amounts of video data, performing real-time analysis, and distinguishing normal from abnormal behavior. Efficient algorithms that utilize parallel computing and GPU acceleration are necessary to meet the computational demands of processing large volumes of surveillance video data in real-time. Additionally, distinguishing normal from abnormal behavior across different contexts and types is another key challenge in SVAD. Machine learning models can be trained to identify patterns and features in the video data, enabling them to differentiate between normal and abnormal behavior.

False detection poses a major obstacle to the effective implementation of SVAD systems, as it can be triggered by changes in illumination, camera motion, occlusions, and scene clutter [10]. To mitigate this challenge, a range of techniques can be applied, including feature extraction, anomaly detection algorithms, and deep-learning-based methods. Additionally, strategies such as temporal modeling and background subtraction are employed to minimize the occurrence of false detections.

The behavior of people and objects can vary significantly, making it challenging to identify what constitutes an anomaly. For example, a person running in a park is normal behavior, but the same person running in a shopping mall may be considered abnormal. Therefore, the detection system must be able to learn and distinguish between different behaviors to identify anomalies accurately.

Tracking individuals and objects and identifying anomalies in their behavior is a complex task, particularly in challenging environments. Algorithms may struggle with accurately tracking and identifying objects, leading to errors and false positives. To achieve greater accuracy, sophisticated algorithms must be employed to analyze and interpret data.

Imbalanced datasets, where normal activities occur more frequently than anomalous events, can lead to model bias toward predicting normal activities. Deep learning models can automatically extract features from video frames and learn to detect anomalies despite differing lighting conditions, camera angles, and object sizes. Additionally, transfer learning techniques can adapt models trained on one dataset to another with similar characteristics, reducing the need for large amounts of labeled data.

Researchers face several challenges when working with SVAD datasets. One major challenge is the need for ground truth video sequences to analyze the data accurately. However, creating and annotating these sequences with the required level of detail is a very time-consuming task. Additionally, there are other obstacles, such as a shortage of training samples and annotations, a lack of diversity in terms of scenes and viewing angles, the exclusion of adverse weather conditions and varying illumination, insufficient coverage of anomaly events, and the limitations of camera devices [120]. These factors hinder the development of accurate and reliable SVAD analysis models.

Environmental noise and inadequate resilience make it more difficult to distinguish between normal and abnormal occurrences, resulting in false alarms or the failure to notice some events. However, deep learning techniques have shown potential in automatically learning important data features. Feature extraction and ensemble methods have also been utilized to develop effective anomaly detection systems that consider the challenging and noisy environments in which they will be utilized.

SVAD involves analyzing features within a video scene to identify patterns and detect anomalies. However, processing features that do not fit the expected patterns can lead to unnecessary computation, resulting in slower and less accurate detection. Therefore, selecting the appropriate features is crucial for effective anomaly detection in videos.

In order to tackle the challenges faced by SVAD systems in real-world scenarios, researchers need to come up with innovative methods that can enhance their accuracy and effectiveness. However, it is worth noting that there may be certain challenges and trade-offs associated with overcoming these challenges. For instance, Tsiktsiris et al. [121] highlighted that collecting more data or performing further fine-tuning may be necessary to achieve better performance. Similarly, He et al. [122] identified the primary challenge of spatially and temporally localizing anomalies in SVAD systems. These studies underscore the need for ongoing research and development to address the complex challenges involved in building effective anomaly detection systems.

6.3. Future Directions

Future research problems are selecting the proper video features, selecting the best classifiers for improved performance in anomaly detection, and creating an appropriate performance–cost balance.

SVAD has benefited from the successful application of UL techniques. However, to further enhance the performance of these methods, it may be beneficial to supplement them with other techniques, such as image processing or SL. By combining these different methods, we can potentially improve the accuracy and effectiveness of anomaly detection in videos. Moreover, UL techniques enable machine learning from unlabeled videos, eliminating the need for manually tagging significant amounts of data, as required in SL methods. As future studies focus on training DNNs, it is crucial to prioritize approaches that require attention from a human.

Research can focus on improving prediction models to capture complex spatiotemporal patterns in surveillance videos. This can involve integrating attention mechanisms, graph-based models, or other innovative methods to enhance accuracy and robustness. Another area of study is developing more resilient reconstruction models that can effectively handle common difficulties such as occlusions and lighting changes. This could entail incorporating techniques such as adversarial training or domain adaptation to improve the robustness of the reconstruction process.

In this field, researchers commonly use the "UCSD Ped2" and "CUHK Avenue" datasets, which were recorded outdoors. However, it is apparent that we need to improve the quantity and quality of datasets that include both synthetic abnormal events created within specific scenarios and real-world abnormal events captured by actual surveillance cameras. To improve video surveillance, we need higher-resolution datasets and greater diversity in anomaly detection techniques.

Integrating SVAD with other technologies can further enhance its capabilities. For instance, combining video analytics with sensor data from sources such as audio, temperature, or biometric data can provide a more comprehensive and precise understanding of anomalous events.

Achieving effective SVAD over edge networks is challenging due to the computational cost of deep learning models. Future directions should focus on optimizing online learning techniques for edge devices to enable real-time SVAD in resource-constrained environments.

To improve the flexibility and efficiency of SVAD models, they can be made adaptable to diverse data by incorporating parameters such as object/individual locations, distances, and motion trajectories into the model's design. This will enable real-time detection of anomaly patterns in video frames, including those that have not been seen before.

Incorporating human feedback and expertise is crucial for effective anomaly detection in human-in-the-loop systems. One possible avenue for future development is the use of human-in-the-loop surveillance video anomaly detection, which would allow operators to provide feedback, refine the model, and contribute their domain-specific knowledge. This approach has the potential to yield more precise and relevant anomaly detection results that are tailored to specific contexts.

As technology continues to advance, we can expect to see a proliferation of anomaly detection models designed for specific domains. These models will rely on domain-specific data to pinpoint unusual patterns and behaviors that may signal potential issues. By incorporating industry-specific characteristics, these models will offer even greater accuracy, providing more nuanced insights to support better decision-making and outcomes.

7. Conclusions

In conclusion, this survey article provided a comprehensive overview of the current state-of-the-art in the field of SVAD and the various AI techniques that have been applied to this problem. The review highlighted the methods, datasets, challenges, and future directions explored in previous studies. The need for automated systems for detecting abnormal events in real-time has been driven by the increasing use of CCTV and other video recording systems, leading to an overwhelming amount of video data being produced. The ability to learn from new observations and continuously improve anomaly detection capabilities is also paramount in video surveillance.

The field of SVAD is expected to see significant advancements in the coming years, thanks to the rapid progress in AI techniques and the availability of reasonably priced hardware. The survey has shown that prediction-based and reconstruction-based techniques are at the forefront of AI-based SVAD and are expected to provide improved anomaly detection capabilities and enable real-time monitoring of large-scale video surveillance systems.

It is also worth noting that the field of SVAD is still an active area of research, and there is still much room for improvement. Future research should focus on developing more robust and efficient algorithms and addressing existing methods' limitations. Additionally, more comprehensive and diverse datasets should be used to evaluate the performance of the proposed methods. Using large-scale datasets with various types of anomalies and scenarios will help improve the generalization capabilities of the proposed methods.

In summary, this survey article provided an up-to-date overview of the current state-of-the-art in the field of SVAD and highlighted the most-significant developments, techniques, and datasets used in this area. Soon, prediction-based and reconstruction-based techniques will be the trend in AI-based SVAD, providing improved anomaly detection capabilities and enabling real-time monitoring of large-scale video surveillance systems.

Author Contributions: Conceptualization, E.Ş.; Methodology, E.Ş. and R.S.; Software, E.Ş. and B.A.; Formal Analysis, E.Ş. and R.S.; Investigation, E.Ş. and Q.A.A.-H.; Data curation, R.S.; Visualization, Q.A.A.-H.; Resources, Q.A.A.-H.;Validation, B.A., A.A. and A.A.A.; Funding acquisition, B.A., A.A. and A.A.A.; Writing—original draft, E.Ş., R.S., Q.A.A.-H., A.A. and A.A.A.; Writing—review & editing, E.Ş., R.S., Q.A.A.-H., B.A., A.A. and A.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of education and the Deanship of Scientific Research at Najran University under the Research Groups Funding program grant (NU/RG/SERC/12/9).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kumari, P.; Bedi, A.K.; Saini, M. Multimedia Datasets for Anomaly Detection: A Survey. *arXiv* **2021**, arXiv:2112.05410.
2. Verma, K.K.; Singh, B.M.; Dixit, A. A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system. *Int. J. Inf. Technol.* **2019**, *14*, 397–410. [CrossRef]
3. Zhao, Y. Deep Learning in Video Anomaly Detection and Its Applications. Ph.D. Thesis, The University of Liverpool, Liverpool, UK, 2021.
4. Abu Al-Haija, Q.; Zein-Sabatto, S. An Efficient Deep-Learning-Based Detection and Classification System for Cyber-Attacks in IoT Communication Networks. *Electronics* **2020**, *9*, 2152. [CrossRef]
5. Grubbs, F.E. Procedures for detecting outlying observations in samples. *Technometrics* **1969**, *11*, 1–21. [CrossRef]
6. Hawkins, D.M. *Identification of Outliers*; Springer: Berlin/Heidelberg, Germany, 1980; Volume 11.
7. Barnett, V.; Lewis, T. *Outliers in Statistical Data*; Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics; Wiley: New York, NY, USA, 1984.
8. Wan, B.; Jiang, W.; Fang, Y.; Luo, Z.; Ding, G. Anomaly detection in video sequences: A benchmark and computational model. *IET Image Process.* **2021**, *15*, 3454–3465. [CrossRef]
9. Aldayri, A.; Albattah, W. Taxonomy of Anomaly Detection Techniques in Crowd Scenes. *Sensors* **2022**, *22*, 6080. [CrossRef]
10. Pannirselvam, P.M.; Geetha, M.K.; Kumaravelan, G. A Comprehensive Study on Automated Anomaly Detection Techniques in Video Surveillance. *Ann. Rom. Soc. Cell Biol.* **2021**, *25*, 4027–4037.
11. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv. (CSUR)* **2009**, *41*, 1–58. [CrossRef]
12. Abdelghafour, M.; ElBery, M.; Taha, Z. Comparative Study for Anomaly Detection in Crowded Scenes. *Int. J. Intell. Comput. Inf. Sci.* **2021**, *21*, 84–94. [CrossRef]
13. Wilmet, V.; Verma, S.; Redl, T.; Sandaker, H.; Li, Z. A Comparison of Supervised and Unsupervised Deep Learning Methods for Anomaly Detection in Images. *arXiv* **2021**, arXiv:2107.09204.
14. Abu Al-Haija, Q.; Alohaly, M.; Odeh, A.; A Lightweight Double-Stage Scheme to Identify Malicious DNS over HTTPS Traffic Using a Hybrid Learning Approach. *Sensors* **2023**, *23*, 3489. [CrossRef] [PubMed]
15. Medel, J.R. *Anomaly Detection Using Predictive Convolutional Long Short-Term Memory Units*; Rochester Institute of Technology: Rochester, NY, USA, 2016.
16. Le, V.T.; Kim, Y.G. Attention-based residual autoencoder for video anomaly detection. *Appl. Intell.* **2022**, *53*, 3240–3254. [CrossRef]
17. Liu, W.; Cao, J.; Zhu, Y.; Liu, B.; Zhu, X. Real-time anomaly detection on surveillance video with two-stream spatiotemporal generative model. *Multimed. Syst.* **2022**, *29*, 59–71. [CrossRef]

18. Başkurt, K.B.; Samet, R. Long-term multiobject tracking using alternative correlation filters. *Turk. J. Electr. Eng. Comput. Sci.* **2018**, *26*, 2246–2259. [CrossRef]
19. Baskurt, K.B.; Samet, R. Video synopsis: A survey. *Comput. Vis. Image Underst.* **2019**, *181*, 26–38. [CrossRef]
20. Ramachandra, B.; Jones, M.; Vatsavai, R. Learning a distance function with a Siamese network to localize anomalies in videos. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 2598–2607.
21. Colque, R.V.H.M. Robust Approaches for Anomaly Detection Applied to Video Surveillance. 2018. Available online: https://repositorio.ufmg.br/manutencao/ (accessed on 12 January 2023).
22. Chen, C.H. *Handbook of Pattern Recognition and Computer Vision*; World Scientific: Singapore, 2015.
23. Munyua, J.G.; Wambugu, G.M.; Njenga, S.T. A Survey of Deep Learning Solutions for Anomaly Detection in Surveillance Videos. *Int. J. Comput. Inf. Technol.* **2021**, *10*, 5. [CrossRef]
24. Alsulami, A.A.; Abu Al-Haija, Q.; Tayeb, A.; Alqahtani, A. An Intrusion Detection and Classification System for IoT Traffic with Improved Data Engineering. *Appl. Sci.* **2020**, *12*, 12336. [CrossRef]
25. Nasteski, V. An overview of the supervised machine learning methods. *Horizons B* **2017**, *4*, 51–62. [CrossRef]
26. Morente-Molinera, J.A.; Mezei, J.; Carlsson, C.; Herrera-Viedma, E. Improving supervised learning classification methods using multigranular linguistic modeling and fuzzy entropy. *IEEE Trans. Fuzzy Syst.* **2016**, *25*, 1078–1089. [CrossRef]
27. Angarita-Zapata, J.S.; Masegosa, A.D.; Triguero, I. A taxonomy of traffic forecasting regression problems from a supervised learning perspective. *IEEE Access* **2019**, *7*, 68185–68205. [CrossRef]
28. Asad, M.; Yang, J.; He, J.; Shamsolmoali, P.; He, X. Multi-frame feature-fusion-based model for violence detection. *Vis. Comput.* **2021**, *37*, 1415–1431. [CrossRef]
29. Wang, T.; Qiao, M.; Deng, Y.; Zhou, Y.; Wang, H.; Lyu, Q.; Snoussi, H. Abnormal event detection based on analysis of movement information of video sequence. *Optik* **2018**, *152*, 50–60. [CrossRef]
30. Patil, N.; Biswas, P.K. Global abnormal events detection in surveillance video—A hierarchical approach. In Proceedings of the 2016 Sixth International Symposium on Embedded Computing and System Design (ISED), Patna, India, 15–17 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 217–222.
31. Kaltsa, V.; Briassouli, A.; Kompatsiaris, I.; Strintzis, M.G. Multiple Hierarchical Dirichlet Processes for anomaly detection in traffic. *Comput. Vis. Image Underst.* **2018**, *169*, 28–39. [CrossRef]
32. Abu Al-Haija, Q., Al Badawi. A. High-performance intrusion detection system for networked UAVs via deep learning. *Neural Comput. Appl.* **2022**, *34*, 10885–10900.
33. Sultani, W.; Chen, C.; Shah, M. Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6479–6488.
34. Alsulami, A.A.; Abu Al-Haija, Q.; Alqahtani, A.; Alsini, R. Symmetrical Simulation Scheme for Anomaly Detection in Autonomous Vehicles Based on LSTM Model. *Symmetry* **2022**, *14*, 1450. [CrossRef]
35. Huang, G.; Song, S.; Gupta, J.N.; Wu, C. Semi-supervised and unsupervised extreme learning machines. *IEEE Trans. Cybern.* **2014**, *44*, 2405–2417. [CrossRef]
36. Chriki, A.; Touati, H.; Snoussi, H.; Kamoun, F. Uav-based surveillance system: An anomaly detection approach. In Proceedings of the 2020 IEEE Symposium on Computers and Communications (ISCC), Rennes, France, 7–10 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
37. Chriki, A.; Touati, H.; Snoussi, H.; Kamoun, F. Deep learning and handcrafted features for one-class anomaly detection in UAV video. *Multimed. Tools Appl.* **2021**, *80*, 2599–2620. [CrossRef]
38. Al-Qudah, M.; Ashi, Z.; Alnabhan, M.; Abu Al-Haija, Q. Effective One-Class Classifier Model for Memory Dump Malware Detection. *J. Sens. Actuator Netw.* **2022**, *12*, 5. [CrossRef]
39. Wang, L.L.; Ngan, H.Y.; Yung, N.H. Automatic incident classification for large-scale traffic data by adaptive boosting SVM. *Inf. Sci.* **2018**, *467*, 59–73. [CrossRef]
40. Ravanbakhsh, M.; Nabi, M.; Sangineto, E.; Marcenaro, L.; Regazzoni, C.; Sebe, N. Abnormal event detection in videos using generative adversarial nets. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1577–1581.
41. Van Engelen, J.E.; Hoos, H.H. A survey on semi-supervised learning. *Mach. Learn.* **2020**, *109*, 373–440. [CrossRef]
42. Bhakat, S.; Ramakrishnan, G. Anomaly detection in surveillance videos. In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, Kolkata, India, 3–5 January 2019; pp. 252–255.
43. Santhosh, K.K.; Dogra, D.P.; Roy, P.P. Anomaly detection in road traffic using visual surveillance: A survey. *ACM Comput. Surv. (CSUR)* **2020**, *53*, 1–26. [CrossRef]
44. Hu, W.; Gao, J.; Li, B.; Wu, O.; Du, J.; Maybank, S. Anomaly detection using local kernel density estimation and context-based regression. *IEEE Trans. Knowl. Data Eng.* **2018**, *32*, 218–233. [CrossRef]
45. Rüttgers, A.; Petrarolo, A. Local anomaly detection in hybrid rocket combustion tests. *Exp. Fluids* **2021**, *62*, 136. [CrossRef]
46. Bansod, S.D.; Nandedkar, A.V. Crowd anomaly detection and localization using histogram of magnitude and momentum. *Vis. Comput.* **2020**, *36*, 609–620. [CrossRef]
47. Zhang, Y.; Lu, H.; Zhang, L.; Ruan, X. Combining motion and appearance cues for anomaly detection. *Pattern Recognit.* **2016**, *51*, 443–452. [CrossRef]

48. Sabokrou, M.; Fayyaz, M.; Fathy, M.; Moayed, Z.; Klette, R. Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. *Comput. Vis. Image Underst.* **2018**, *172*, 88–97. [CrossRef]
49. Rahmani, M.; Atia, G.K. Coherence pursuit: Fast, simple, and robust principal component analysis. *IEEE Trans. Signal Process.* **2017**, *65*, 6260–6275. [CrossRef]
50. Sarker, I.H. Machine learning: Algorithms, real-world applications and research directions. *SN Comput. Sci.* **2021**, *2*, 160. [CrossRef]
51. Wang, T.; Snoussi, H. Detection of abnormal visual events via global optical flow orientation histogram. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 988–998. [CrossRef]
52. Doshi, K.; Yilmaz, Y. Online anomaly detection in surveillance videos with asymptotic bound on false alarm rate. *Pattern Recognit.* **2021**, *114*, 107865. [CrossRef]
53. Aboah, A. 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, Nashville, TN, USA, 20–25 June 2021; pp. 4207–4212.
54. Saeedi, J.; Giusti, A. Anomaly Detection for Industrial Inspection using Convolutional Autoencoder and Deep Feature-based One-class Classification. In Proceedings of the VISIGRAPP (5: VISAPP), Online, 6–8 February 2022; pp. 85–96.
55. Chen, Y.; Tian, Y.; Pang, G.; Carneiro, G. Deep one-class classification via interpolated gaussian descriptor. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2022; Volume 36, pp. 383–392.
56. Lee, K.; Lee, H.; Lee, K.; Shin, J. Training confidence-calibrated classifiers for detecting out-of-distribution samples. *arXiv* **2017**, arXiv:1711.09325.
57. Liu, W.; Luo, W.; Lian, D.; Gao, S. Future frame prediction for anomaly detection–a new baseline. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6536–6545.
58. Jiang, T.; Li, Y.; Xie, W.; Du, Q. Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 4666–4679. [CrossRef]
59. Cheng, H.; Liu, X.; Wang, H.; Fang, Y.; Wang, M.; Zhao, X. SecureAD: A secure video anomaly detection framework on convolutional neural network in edge computing environment. *IEEE Trans. Cloud Comput.* **2020**, *10*, 1413–1427. [CrossRef]
60. Zhao, Y.; Deng, B.; Shen, C.; Liu, Y.; Lu, H.; Hua, X.S. spatiotemporal autoencoder for video anomaly detection. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 1933–1941.
61. Xu, D.; Ricci, E.; Yan, Y.; Song, J.; Sebe, N. Learning deep representations of appearance and motion for anomalous event detection. *arXiv* **2015**, arXiv:1510.01553.
62. Fan, Y.; Wen, G.; Li, D.; Qiu, S.; Levine, M.D.; Xiao, F. Video anomaly detection and localization via gaussian mixture fully convolutional variational autoencoder. *Comput. Vis. Image Underst.* **2020**, *195*, 102920. [CrossRef]
63. Duman, E.; Erdem, O.A. Anomaly detection in videos using optical flow and convolutional autoencoder. *IEEE Access* **2019**, *7*, 183914–183923. [CrossRef]
64. Madan, N.; Farkhondeh, A.; Nasrollahi, K.; Escalera, S.; Moeslund, T.B. Temporal cues from socially unacceptable trajectories for anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 2150–2158.
65. Song, H.; Sun, C.; Wu, X.; Chen, M.; Jia, Y. Learning normal patterns via adversarial attention-based autoencoder for abnormal event detection in videos. *IEEE Trans. Multimed.* **2019**, *22*, 2138–2148. [CrossRef]
66. Sun, C.; Jia, Y.; Song, H.; Wu, Y. Adversarial 3d convolutional auto-encoder for abnormal event detection in videos. *IEEE Trans. Multimed.* **2020**, *23*, 3292–3305. [CrossRef]
67. Nguyen, T.N.; Meunier, J. Anomaly detection in video sequence with appearance-motion correspondence. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1273–1283.
68. Feng, X.; Song, D.; Chen, Y.; Chen, Z.; Ni, J.; Chen, H. Convolutional transformer based dual discriminator generative adversarial networks for video anomaly detection. In Proceedings of the 29th ACM International Conference on Multimedia, Nice, France, 21–25 October 2021; pp. 5546–5554.
69. Lee, J.; Nam, W.J.; Lee, S.W. Multi-Contextual Predictions with Vision Transformer for Video Anomaly Detection. In Proceedings of the 2022 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, 21–25 August 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1012–1018.
70. Yuan, H.; Cai, Z.; Zhou, H.; Wang, Y.; Chen, X. Transanomaly: Video anomaly detection using video vision transformer. *IEEE Access* **2021**, *9*, 123977–123986. [CrossRef]
71. Ullah, A.; Ahmad, J.; Muhammad, K.; Sajjad, M.; Baik, S.W. Action recognition in video sequences using deep bi-directional LSTM with CNN features. *IEEE Access* **2017**, *6*, 1155–1166. [CrossRef]
72. Ergen, T.; Kozat, S.S. Unsupervised anomaly detection with LSTM neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 3127–3141. [CrossRef] [PubMed]
73. Ristea, N.C.; Madan, N.; Ionescu, R.T.; Nasrollahi, K.; Khan, F.S.; Moeslund, T.B.; Shah, M. Self-supervised predictive convolutional attentive block for anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 13576–13586.
74. Zhou, J.T.; Du, J.; Zhu, H.; Peng, X.; Liu, Y.; Goh, R.S.M. Anomalynet: An anomaly detection network for video surveillance. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 2537–2550. [CrossRef]

75. Nawaratne, R.; Alahakoon, D.; De Silva, D.; Yu, X. Spatiotemporal anomaly detection using deep learning for real-time video surveillance. *IEEE Trans. Ind. Inform.* **2019**, *16*, 393–402. [CrossRef]
76. Ullah, W.; Ullah, A.; Hussain, T.; Khan, Z.A.; Baik, S.W. An efficient anomaly recognition framework using an attention residual LSTM in surveillance videos. *Sensors* **2021**, *21*, 2811. [CrossRef]
77. Ranjith, R.; Athanesious, J.J.; Vaidehi, V. Anomaly detection using DBSCAN clustering technique for traffic video surveillance. In Proceedings of the 2015 Seventh International Conference on Advanced Computing (ICoAC), Chennai, India, 15–17 December 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–6.
78. Li, Y.; Guo, T.; Xia, R.; Xie, W. Road traffic anomaly detection based on fuzzy theory. *IEEE Access* **2018**, *6*, 40281–40288. [CrossRef]
79. Chang, M.C.; Wei, Y.; Song, N.; Lyu, S. Video analytics in smart transportation for the AIC'18 challenge. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 61–68.
80. Lu, C.; Shi, J.; Jia, J. Abnormal event detection at 150 fps in matlab. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 2720–2727.
81. Adam, A.; Rivlin, E.; Shimshoni, I.; Reinitz, D. Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 555–560. [CrossRef]
82. Mahadevan, V.; Li, W.; Bhalodia, V.; Vasconcelos, N. Anomaly detection in crowded scenes. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 1975–1981.
83. Lee, S.; Kim, H.G.; Ro, Y.M. STAN: spatiotemporal adversarial networks for abnormal event detection. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1323–1327.
84. Mehran, R.; Oyama, A.; Shah, M. Abnormal crowd behavior detection using social force model. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 935–942.
85. Boekhoudt, K.; Matei, A.; Aghaei, M.; Talavera, E. HR-Crime: Human-Related Anomaly Detection in Surveillance Videos. In Proceedings of the International Conference on Computer Analysis of Images and Patterns, Virtual, 28–30 September 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 164–174.
86. Wang, T.; Qiao, M.; Lin, Z.; Li, C.; Snoussi, H.; Liu, Z.; Choi, C. Generative neural networks for anomaly detection in crowded scenes. *IEEE Trans. Inf. Forensics Secur.* **2018**, *14*, 1390–1399. [CrossRef]
87. Del Giorno, A.; Bagnell, J.A.; Hebert, M. A discriminative framework for anomaly detection in large videos. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 334–349.
88. Tudor Ionescu, R.; Smeureanu, S.; Alexe, B.; Popescu, M. Unmasking the abnormal events in video. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2895–2903.
89. Wang, S.; Zhu, E.; Yin, J.; Porikli, F. Anomaly detection in crowded scenes by SL-HOF descriptor and foreground classification. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 3398–3403.
90. Hasan, M.; Choi, J.; Neumann, J.; Roy-Chowdhury, A.K.; Davis, L.S. Learning temporal regularity in video sequences. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NA, USA, 27–30 June 2016; pp. 733–742.
91. Luo, W.; Liu, W.; Gao, S. Remembering history with convolutional lstm for anomaly detection. In Proceedings of the 2017 IEEE International Conference on Multimedia and Expo (ICME), San Diego, CA, USA, 10–14 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 439–444.
92. Vu, H.; Nguyen, T.D.; Travers, A.; Venkatesh, S.; Phung, D. Energy-based localized anomaly detection in video surveillance. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Nanjing, China, 22–25 May 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 641–653.
93. Chong, Y.S.; Tay, Y.H. Abnormal event detection in videos using spatiotemporal autoencoder. In Proceedings of the International Symposium on Neural Networks, Hokkaido, Japan, 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 189–196.
94. Yang, B.; Cao, J.; Ni, R.; Zou, L. Anomaly detection in moving crowds through spatiotemporal autoencoding and additional attention. *Adv. Multimed.* **2018**, *2018*, 2087574. [CrossRef]
95. Tudor Ionescu, R.; Smeureanu, S.; Popescu, M.; Alexe, B. Detecting abnormal events in video using Narrowed Normality Clusters. *arXiv* **2018**, arXiv-1801.
96. Zhong, J.X.; Li, N.; Kong, W.; Liu, S.; Li, T.H.; Li, G. Graph convolutional label noise cleaner: Train a plug-and-play action classifier for anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1237–1246.
97. Gong, D.; Liu, L.; Le, V.; Saha, B.; Mansour, M.R.; Venkatesh, S.; Hengel, A.V.D. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1705–1714.
98. Luo, W.; Liu, W.; Lian, D.; Tang, J.; Duan, L.; Peng, X.; Gao, S. Video anomaly detection with sparse coding inspired deep neural networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 1070–1084. [CrossRef]

99. Zhou, J.T.; Zhang, L.; Fang, Z.; Du, J.; Peng, X.; Xiao, Y. Attention-driven loss for anomaly detection in video surveillance. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 4639–4647. [CrossRef]
100. Lee, S.; Kim, H.G.; Ro, Y.M. BMAN: Bidirectional multi-scale aggregation networks for abnormal event detection. *IEEE Trans. Image Process.* **2019**, *29*, 2395–2408. [CrossRef] [PubMed]
101. Dubey, S.; Boragule, A.; Jeon, M. 3D resnet with ranking loss function for abnormal activity detection in videos. In Proceedings of the 2019 International Conference on Control, Automation and Information Sciences (ICCAIS), Chengdu, China, 23–26 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
102. Dong, F.; Zhang, Y.; Nie, X. Dual discriminator generative adversarial network for video anomaly detection. *IEEE Access* **2020**, *8*, 88170–88176. [CrossRef]
103. Lu, Y.; Yu, F.; Reddy, M.K.K.; Wang, Y. Few-shot scene-adaptive anomaly detection. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 125–141.
104. Yang, Y.; Zhan, D.; Yang, F.; Zhou, X.D.; Yan, Y.; Wang, Y. Improving video anomaly detection performance with patch-level loss and segmentation map. In Proceedings of the 2020 IEEE 6th International Conference on Computer and Communications (ICCC), Chengdu, China, 11–14 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1832–1839.
105. Chang, Y.; Tu, Z.; Xie, W.; Yuan, J. Clustering driven deep autoencoder for video anomaly detection. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 329–345.
106. Park, H.; Noh, J.; Ham, B. Learning memory-guided normality for anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 14372–14381.
107. Zaheer, M.Z.; Mahmood, A.; Shin, H.; Lee, S.I. A self-reasoning framework for anomaly detection using video-level labels. *IEEE Signal Process. Lett.* **2020**, *27*, 1705–1709. [CrossRef]
108. Feng, J.; Liang, Y.; Li, L. Anomaly detection in videos using two-stream autoencoder with post hoc interpretability. *Comput. Intell. Neurosci.* **2021**, *2021*, 7367870. [CrossRef]
109. Yang, F.; Yu, Z.; Chen, L.; Gu, J.; Li, Q.; Guo, B. Human-machine cooperative video anomaly detection. *Proc. ACM -Hum.-Comput. Interact.* **2021**, *4*, 1–18. [CrossRef]
110. Georgescu, M.I.; Barbalau, A.; Ionescu, R.T.; Khan, F.S.; Popescu, M.; Shah, M. Anomaly detection in video via self-supervised and multi-task learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 12742–12752.
111. Li, B.; Leroux, S.; Simoens, P. Decoupled appearance and motion learning for efficient anomaly detection in surveillance video. *Comput. Vis. Image Underst.* **2021**, *210*, 103249. [CrossRef]
112. Ouyang, Y.; Sanchez, V. Video anomaly detection by estimating likelihood of representations. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 8984–8991.
113. Dubey, S.; Boragule, A.; Gwak, J.; Jeon, M. Anomalous event recognition in videos based on joint learning of motion and appearance with multiple ranking measures. *Appl. Sci.* **2021**, *11*, 1344. [CrossRef]
114. Wang, B.; Yang, C. Video Anomaly Detection Based on Convolutional Recurrent AutoEncoder. *Sensors* **2022**, *22*, 4647. [CrossRef]
115. Ul Amin, S.; Ullah, M.; Sajjad, M.; Cheikh, F.A.; Hijji, M.; Hijji, A.; Muhammad, K. EADN: An Efficient Deep Learning Model for Anomaly Detection in Videos. *Mathematics* **2022**, *10*, 1555. [CrossRef]
116. Ganokratanaa, T.; Aramvith, S.; Sebe, N. Video anomaly detection using deep residual-spatiotemporal translation network. *Pattern Recognit. Lett.* **2022**, *155*, 143–150. [CrossRef]
117. Huang, X.; Zhao, C.; Wang, Y.; Wu, Z. A Video Anomaly Detection Framework based on Appearance-Motion Semantics Representation Consistency. *arXiv* **2022**, arXiv:2204.04151.
118. Liu, W.; Chang, H.; Ma, B.; Shan, S.; Chen, X. Diversity-Measurable Anomaly Detection. *arXiv* **2023**, arXiv:2303.05047.
119. Ouyang, Y.; Shen, G.; Sanchez, V. Look at adjacent frames: Video anomaly detection without offline training. In *Computer Vision–ECCV 2022 Workshops*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 642–658.
120. Sharif, M.; Jiao, L.; Omlin, C.W. Deep Crowd Anomaly Detection: State-of-the-Art, Challenges, and Future Research Directions. *arXiv* **2022**, arXiv:2210.13927.
121. Tsiktsiris, D.; Dimitriou, N.; Lalas, A.; Dasygenis, M.; Votis, K.; Tzovaras, D. Real-time abnormal event detection for enhanced security in autonomous shuttles mobility infrastructures. *Sensors* **2020**, *20*, 4943. [CrossRef]
122. He, C.; Shao, J.; Sun, J. An anomaly-introduced learning method for abnormal event detection. *Multimed. Tools Appl.* **2018**, *77*, 29573–29588. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.