electronics

MDPI

Review

Deep Learning for Abnormal Human Behavior Detection in Surveillance Videos—A Survey

Leonard Matheus Wastupranata [1], Seong G. Kong [1,*] and Lipo Wang [2,*]

[1] Department of Computer Engineering, Sejong University, Seoul 05006, Republic of Korea; leo.matt.547@sju.ac.kr
[2] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore
* Correspondence: skong@sejong.edu (S.G.K.); elpwang@ntu.edu.sg (L.W.)

Abstract: Detecting abnormal human behaviors in surveillance videos is crucial for various domains, including security and public safety. Many successful detection techniques based on deep learning models have been introduced. However, the scarcity of labeled abnormal behavior data poses significant challenges for developing effective detection systems. This paper presents a comprehensive survey of deep learning techniques for detecting abnormal human behaviors in surveillance video streams. We categorize the existing techniques into three approaches: unsupervised, partially supervised, and fully supervised. Each approach is examined in terms of its underlying conceptual framework, strengths, and drawbacks. Additionally, we provide an extensive comparison of these approaches using popular datasets frequently used in the prior research, highlighting their performance across different scenarios. We summarize the advantages and disadvantages of each approach for abnormal human behavior detection. We also discuss open research issues identified through our survey, including enhancing robustness to environmental variations through diverse datasets, formulating strategies for contextual abnormal behavior detection. Finally, we outline potential directions for future development to pave the way for more effective abnormal behavior detection systems.

Keywords: abnormal human behavior detection; video surveillance; deep learning; data scarcity; security

Citation: Wastupranata, L.M.; Kong, S.G.; Wang, L. Deep Learning for Abnormal Human Behavior Detection in Surveillance Videos—A Survey. *Electronics* **2024**, *13*, 2579. https://doi.org/10.3390/electronics13132579

Academic Editor: Yue Wu

Received: 4 June 2024
Revised: 25 June 2024
Accepted: 28 June 2024
Published: 30 June 2024

1. Introduction

Abnormal human behavior detection involves identifying unusual behavior or state transitions in a targeted subject. Behavior deviating from the norm is deemed abnormal [1]. In surveillance video monitoring, video footage from static cameras is analyzed for such behaviors [2–6]. The field of abnormal behavior detection primarily focuses on security and public safety, promoting the well-being of society [7,8]. Surveillance video provides valuable visual information within a defined field of view for detecting abnormal human behaviors [9–11].

Abnormal behavior detection can be categorized into short-term and long-term detection based on the period. Short-term abnormal behavior detection entails identifying abnormal actions from video frames of a relatively short time duration, facilitating instantaneous decision-making. This category covers various abnormal behavior detection scenarios such as fire detection [12], running [13–17], falling [18–22], crowding [23–28], throwing objects [29–31], fighting [32–35], trespassing [36–41], and moving in opposite directions [42]. On the other hand, long-term abnormal behavior detection requires longer video durations to reach a decision. Examples include suspicious loitering [43–48], leaving bags unattended [49–53], prolonged absence of movement in specific areas [54–56], or extended instances of erratic behavior.

Several survey papers have compared abnormal behavior detection techniques using various evaluation metrics, such as accuracy, equal error rate (EER), and area under

the curve (AUC), meaning the area under the receiver operating characteristic (ROC) curve [57–63]. These metrics are crucial for assessing the effectiveness of different approaches in identifying abnormal behaviors. In recent years, abnormal behavior prediction analysis utilizes a variety of deep learning algorithms [64]. Deep learning techniques for abnormal human behavior detection can be classified into three approaches: unsupervised, partially supervised, and fully unsupervised learning [65]. Specifically, the weakly supervised and semi-supervised learning paradigms are referred to as partially supervised learning [66]. The scarcity of abnormal human behavior data poses a significant challenge [67]. Hence, unsupervised and partially supervised detection approaches serve as alternatives to address the data scarcity issue [68–70]. In these approaches, the model learns normal behavior patterns from input layers during the training phase [59]. This model then detects abnormal behaviors by identifying deviations from the learned patterns or by comparing the new data against the identified normal behavior clusters. However, not all unsupervised learning schemes are equally effective in detecting abnormalities in images or videos. Two commonly used unsupervised learning approaches for abnormal human behavior detection are reconstruction-based [71] and generative detection [72] methods.

In summary, this study provides a comprehensive examination of methods across the three deep-learning-based detection approaches, addressing evident gaps such as the limited coverage of research in the past five years, the scarcity of abnormal human behavior data, and limited performance comparisons using popular datasets from the prior research.

1.1. Literature Review Methodology

The survey was conducted on the literature published in major journals. Their online platforms were explored to locate the latest articles regarding deep learning techniques for abnormal human behavior detection in surveillance videos. To ensure comprehensiveness, several reputable academic websites were consulted, including Web of Science, IEEE Xplore, Google Scholar, Science Direct, Scopus, ACM, and MDPI. Figure 1 illustrates the trend in the number of papers focusing on deep learning for abnormal human behavior detection over the past five years, from 2019 to 2023. With this upward trend, the topic of abnormal human behavior detection research using deep learning is becoming increasingly prominent, as evidenced by the substantial number of research publications in this field, particularly in 2023.

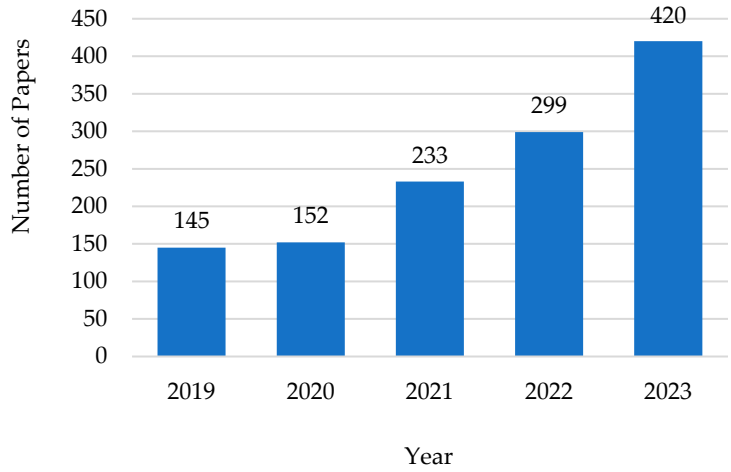

Figure 1. Trend in the number of publications on deep learning for abnormal human behavior detection over the past five years (2019–2023).

Several combinations of keywords were used, including “deep learning”, “unsupervised learning”, “weakly-supervised learning”, “semi-supervised learning”, “anomaly detection”, and “video surveillance”, to search for relevant articles. In total, 1284 results were obtained from several academic journal search engines. Among these, 416 results

were from Science Direct, 388 from Google Scholar, 93 from IEEE Xplore, and 63 from Scopus. The remaining results were obtained from the Web of Science, ACM, and MDPI search engines. Figure 2 illustrates the distribution of related published papers on abnormal human behavior detection by search engines. From the collected set of articles, all those published before 2019 were removed. Additionally, priority was given to journal articles. The remaining articles were screened to exclude those not related to abnormal behavior detection, particularly those that do not focus on video surveillance and static cameras. Then, the remaining papers were categorized into unsupervised, partially supervised, and fully supervised detection approaches. Finally, the screening process was completed with 97 papers, including 13 survey papers from related works. Approximately 90% of the selected papers are from the past four years, since 2020. Only 9% of the selected papers were published in 2019.

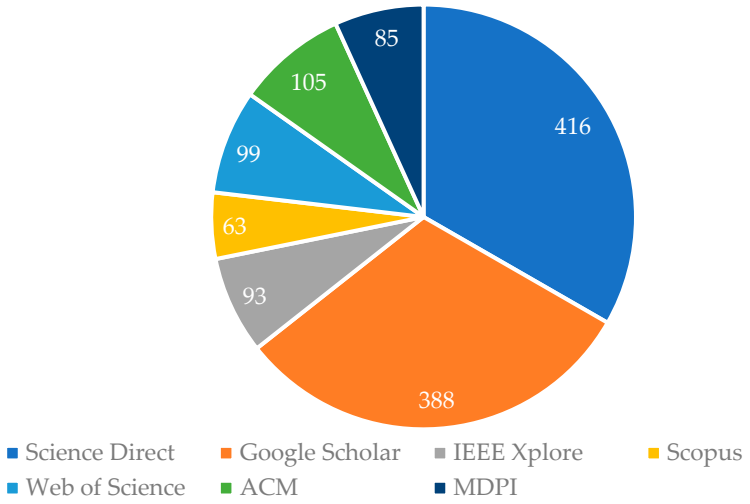

Figure 2. Distribution of related published papers on abnormal human behavior detection by search engines.

1.2. Contributions of the Paper

The objective of this paper is to assess the strengths and weaknesses of various abnormal behavior detection techniques within the domain of deep learning. Additionally, it provides an overview of recent breakthroughs from studies conducted over the past five years. Each piece of research is meticulously analyzed, including the abnormal behavior datasets utilized, performance evaluation results of the models in terms of AUC or accuracy, and an examination of the pros and cons associated with each prior research endeavor. This paper places particular emphasis on deep learning techniques, thereby narrowing down the focus compared to earlier review papers. The key contributions of this paper are as follows:

1. Categorizing deep learning techniques for abnormal human behavior detection into three main detection approaches: unsupervised, partially supervised, and fully supervised.
2. Discussing the strengths and drawbacks of each learning scheme for training a deep learning model for abnormal human behavior detection.
3. Conducting a comprehensive comparison of the performances of deep-learning-based abnormal human behavior detection techniques on popular benchmarking datasets.
4. Exploring open research issues in the field of abnormal human behavior detection in surveillance videos.

1.3. Organization of the Paper

This paper is organized as follows: Section 2 explains the types of abnormal human behavior detection and the prior research. Section 3 presents popular abnormal behavior datasets utilized in the prior research works. Section 4 surveys the deep learning techniques

for abnormal human behavior detection in surveillance videos, categorizing them into unsupervised, partially supervised, and fully supervised approaches. Section 5 discusses the open research issues of the current deep learning techniques. Finally, Section 6 presents the conclusion of this survey paper. Figure 3 illustrates the organizational structure of this survey to facilitate navigation through the paper.

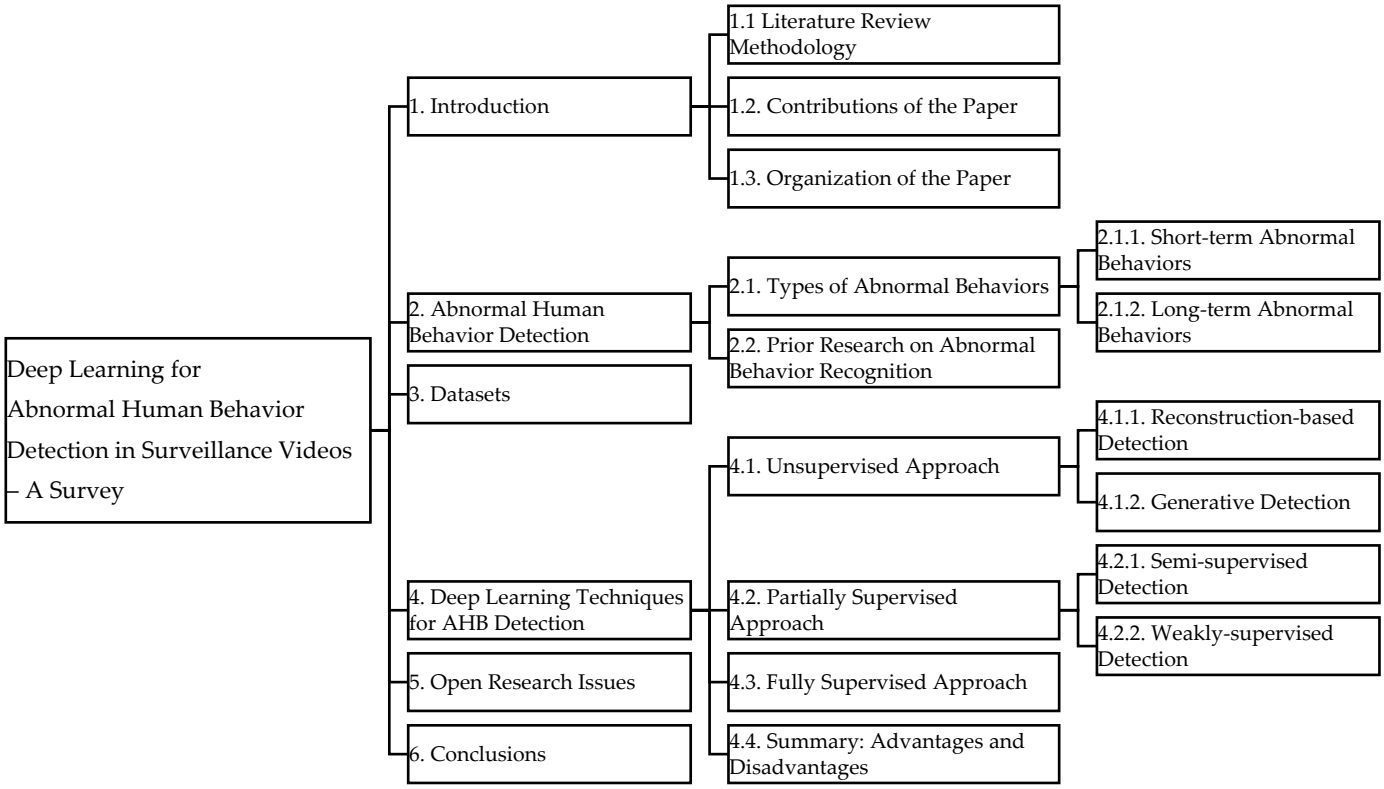

Figure 3. The organizational structure of the survey.

2. Abnormal Human Behavior Detection

Abnormal human behaviors (AHB) involve observing actions of human-like entities and identifying unusual patterns in behavior that deviate from the norm. These behaviors are labeled as 'abnormal' because they diverge from typical environmental contexts [73]. The rapid detection of abnormal behavior is crucial in real-time settings, particularly in environments where public safety is paramount [74]. AHB detection poses challenges due to the dynamic visual characteristics influenced by environmental conditions and the nature of abnormal actions [75].

2.1. Types of Abnormal Behaviors

In the process of abnormal human behavior detection, certain strategies prioritize detection time. Abnormal human behaviors are generally classified into two types: short-term and long-term abnormal behaviors. Further elaboration on these classifications is provided in the subsequent subsections.

2.1.1. Short-Term Abnormal Behaviors

Short-term abnormal human behavior refers to behaviors that deviate from the norm and can be identified by analyzing a relatively short duration of video frames. Decisions regarding such behaviors can be made immediately as their consequences become apparent in real-time [76]. Examples of short-term abnormal behaviors include fires, running, falling, crowding, throwing objects, fighting, trespassing, and moving in opposite directions.

Researchers have explored the early detection of burning fires in real-time scenarios [12,77,78]. Fires, whether resulting from accidents or intentional human acts, are classified as short-term abnormal behavior as they can be detected from a single-frame view of fire. The act of human running can be considered abnormal behavior in environments where most individuals typically walk. [17]. Numerous studies have explored the detection of running behavior as a significant aspect [13–16]. Visual differentiation between walking and running can be achieved using a single image [79]. Hence, running can be classified as short-term abnormal behavior. A fall condition occurs when a person loses body balance and ends up in an unstable position [80]. Research has been conducted on fall detection, establishing it as one of the classifications for short-term abnormal behavior detection [18–22]. A crowd is defined as a condition where two or more people are closely grouped within a single frame [23]. Detecting crowds is crucial as it often signifies an abnormal event [24–28]. Throwing objects involves the act of hurling potentially dangerous or harmful items [31]. Numerous studies on detecting this behavior emphasize the risk posed by the throwing of prohibited items [29,30]. Identifying suspicious objects in the air from a single frame classifies this behavior as short-term abnormal. Physical altercations involving two or more individuals with the potential for injury are categorized as fights [32]. Detection of these fights has been undertaken by several researchers to mitigate potential impacts [33–35]. Similar to the preceding category, fighting behavior falls under short-term abnormal behavior due to its detectability within a few frames. The high occurrence of accidents resulting from trespassing in restricted areas, such as railway lines, necessitates proactive measures [37]. Detecting human presence in restricted areas from a single frame allows for the identification of breaches in designated zones [36,38–41]. Thus, trespassing is classified as short-term abnormal behavior. Moving in the opposite direction, typically observed when an individual walks against the flow within a crowd, can be identified with only a small number of frames [42]. Therefore, this behavior is categorized as short-term abnormal.

2.1.2. Long-Term Abnormal Behaviors

Long-term abnormal human behavior refers to persistent patterns of unusual behavior observed over an extended period. Unlike short-term abnormal behavior, it requires prolonged observation to discern significant deviations from expected behavioral patterns. These behaviors may unfold gradually over time, necessitating continuous monitoring and analysis to understand their full impact. Examples of long-term abnormal behaviors include loitering, leaving bags unattended, and prolonged absence of movement in specific areas.

Loitering, where individuals aimlessly linger in crowded areas, can pose threats to public safety [48]. Detection of such behavior occurs when an individual follows others without any apparent purpose for an extended period [43–47]. Therefore, loitering is classified as long-term abnormal behavior. An unattended bag refers to a situation where the owner intentionally leaves it behind for a certain period [49]. If the bag contains potentially dangerous items, swift action must be taken [50–53]. The time required to detect the unattended bag places it in the category of long-term abnormal behavior. The lack of human movement detected by the camera can signal abnormal behavior. This phenomenon is also referred to as unusual inactivity or stationary movement detection. The absence of human movement in a specific area raises concerns, whether it involves a group of individuals [55,56] or an elderly person [54]. Detecting inactivity requires a longer time to reach a conclusive decision. Therefore, this category is classified as long-term abnormal behavior.

2.2. Prior Research on Abnormal Behavior Recognition

Deep learning offers advantages as it requires minimal hand engineering, especially with the increasing availability of computing power and data [81]. Techniques utilizing fully supervised learning frameworks in deep learning can achieve high accuracy but demand substantial amounts of data and computational resources [82,83]. Several studies

employ convolutional neural networks (CNN), long–short-term memory (LSTM), and gated recurrent unit (GRU) architectures to analyze abnormal human behavior spatially and temporally [84–94]. However, a significant obstacle in deep learning is the limited availability of data, often referred to as data scarcity [95].

The advancement of abnormal human behavior detection is hindered by issues related to data scarcity [96]. Detecting abnormal human behavior that has not been previously defined in the training data, especially given the wide variety of behaviors, poses significant challenges [97]. Furthermore, many instances of abnormal human behavior are context-dependent, with behaviors considered abnormal in one setting being normal in another [98]. This phenomenon arises from the model's incapacity to capture intrinsic uncertainty and often leads to decreased efficiency in the event recognition phase, particularly due to data scarcity [99,100]. To address these challenges, an important research question emerges: How can models be effectively trained to detect abnormal human behavior with limited labeled data, considering the diversity and context-dependency of behaviors? To mitigate the impact of data scarcity, the model is expected to learn from unlabeled data by identifying relationships and patterns through unsupervised and partially supervised learning paradigms [101,102].

In unsupervised learning, the primary focus lies on collecting unlabeled data including both normal and abnormal behaviors [60]. Within the unsupervised learning framework, a reconstruction-based approach is employed, where the model analyzes only normal event data and detects abnormalities by examining low reconstruction errors [58]. Initially, this approach utilized a basic auto-encoder to differentiate abnormal behavior from normal human behavior data exclusively [9,103–111]. Additionally, there is the use of variational auto-encoders, which incorporate a probability function during reconstruction through a neural network [112–120]. However, both auto-encoders and variational auto-encoders are not inherently designed for abnormal human behavior detection in two-dimensional data. To address this limitation, convolutional auto-encoders are employed, preserving the spatial locality of input data throughout the reconstruction stage [121–130].

Within the unsupervised learning framework, there exists an approach for detecting abnormal human behavior using generative detection [62]. In this approach, artificial images are generated from trained distribution patterns. Subsequently, the model discriminates whether the image is real or fake, effectively addressing the challenge of data scarcity [131–148]. However, both the reconstruction-based and the generative approaches pose difficulties in identifying specific abnormal behaviors and are highly sensitive to environmental changes [102].

There exists a scheme known as partially supervised learning that utilizes both labeled and unlabeled data [149]. Some research emphasizes maximizing the use of unlabeled data with labeled data serving as an anchor, a methodology termed semi-supervised learning [137,150–153]. Conversely, in contrast to semi-supervised learning, there are several studies that prioritize maximum model output with minimal reference label data, referred to as weakly supervised learning [145,154–169]. Consequently, partially supervised learning demonstrates capability in addressing data scarcity and alleviating the time-consuming nature inherent in deep-learning-based abnormal human behavior detection in surveillance videos [59,170].

There have been several previous works surveying abnormal human behavior detection systems. Patrikar and Parate [60] provide a survey on image-based detection systems for abnormal behaviors in video surveillance. They also survey edge-computing-based abnormal detection and divide the explanations into two main parts: learning and modeling algorithms. However, their focus is mainly on the application of edge computing, lacking a thorough exploration in the context of machine learning. Myagmar-Ochir and Kim [101] survey video surveillance systems (VSS) for smart city applications, but their explanation of the methods used in unsupervised learning methods is incomplete. Duong, Le, and Hoang [98] survey vision-based human activity recognition and describe popular databases commonly used. They also present data processing and feature engineering. However,

their survey does not include quantitative comparisons using metrics between research results for each prior study. Choudhry et al. [59] comprehensively explain the challenges of machine learning techniques for VSS and divide them into three categories: supervised, semi-supervised, and unsupervised. However, their scope does not focus on image-based detection. The previous surveys rarely cover research published in 2023. Moreover, there is a need for surveys to discuss the scarcity of abnormal human behavior data, as well as challenges and future applications. Table 1 summarizes the previous surveys.

Table 1. Summary of abnormal human behavior surveys in video surveillance.

Survey (Year)	Scope				Merits	Limitations
	Datasets	Deep Learning	Application	Metrics Comparison		
Patrikar and Parate [60] (2022)	✓	P	✓	✓	• Provides a survey on image-based detection systems • Divides explanations into learning and modeling algorithms • Emphasizes edge computing applications	Lacks thorough exploration of machine learning in the context of abnormal behavior detection
Myagmar-Ochir and Kim [101] (2023)	P	P	✓	P	Surveys VSS for smart city applications	The explanation of methods used in unsupervised learning methods is incomplete
Duong, Le, and Hoang [98] (2023)	✓	✓	-	-	• Describes popular databases used • Presents data processing and feature engineering	Does not include quantitative comparisons using metrics among research results
Choudhry et al. [59] (2023)	✓	✓	✓	P	• Comprehensively explains challenges of machine learning techniques for VSS • Divides techniques into supervised, semi-supervised, and unsupervised categories	The scope does not focus on image-based detection
Ours (2024)	✓	✓	✓	✓	• Categorizes existing AHB detection into unsupervised, partially supervised, and fully supervised approaches • Examines each approach's conceptual framework, strengths, and drawbacks • Provides extensive comparison using popular datasets	

(✓)—fully explained; (P)—partially explained; (-)—not explained.

3. Datasets

Several datasets have been widely used by researchers to benchmark related research. This section addresses the question, "What are the datasets used by prior research in abnormal human behavior detection?" The University of California, San Diego (UCSD) anomaly dataset consists of 70 pieces of video footage captured from an elevated perspective to monitor pedestrian walkways [171]. Abnormal events captured in this dataset include the presence of non-pedestrian entities in the walkways and anomalous pedestrian motion patterns. The UCSD anomaly dataset comprises two sets of videos, Ped1 and Ped2. Ped1 contains footage of people walking towards and away from the camera, with various perspective distortions, along with humans identified as abnormalities. Ped2 includes scenes with pedestrian movement parallel to the camera plane from a top angle, where non-human objects are considered abnormal.

The ShanghaiTech (ST) Campus dataset comprises 13 scenes with complex lighting conditions and camera angles from all sides [172]. It contains 130 abnormal events and over 270,000 training frames across a total of 437 videos for training and testing. The University of Central Florida (UCF)-Crime dataset consists of 1900 videos totaling 128 h, covering 13 anomalies in real-world environments, including fighting, vandalism, and robbery [173]. The Avenue dataset includes 37 videos, with 16 training video clips and 21 testing video clips. Filmed on the Chinese University of Hong Kong (CUHK) campus, it comprises 30,652 frames, evenly split between training and testing, and features 14 unusual incidents such as people running, loitering, and throwing objects [174]. The University of Minnesota (UMN) dataset encompasses 11 different abnormal scenarios with 3 scenes indoors and outdoors, totaling 22 videos for training and testing [175]. The Performance Evaluation of Tracking and Surveillance (PETS) dataset, recorded at the Whiteknights Campus, University of Reading, UK, captures abnormal behaviors including people counting, density estimation, person tracking, flow analysis, and event recognition [176]. The Subway dataset features two videos totaling two hours, containing 209 and 150 frames, comprising exit gate and entry gate videos. It includes 19 types of unusual events such as walking in the wrong direction, loitering, and wandering near exits [177]. The UBI-Fights dataset, generated by Universidade da Beira Interior in 2020, focuses specifically on fighting events. It consists of an 80-h video dataset labeled at the frame level, including 216 videos of fighting events and others depicting daily life [178].

The Live Videos (LV) dataset consists of 30 videos featuring various abnormal scenes, including 14 different abnormal events, with a total duration of 3.93 h [179]. The Surveillance Fight dataset consists of 300 videos, divided into fight and non-fight sequences taken from movies [180]. The Hockey Fight dataset consists of 1000 video clips from hockey games, manually labeled as fight or non-fight [181]. The Violent Flows dataset consists of 246 videos taken from YouTube and de-interlaced as audio video interleave (AVI) files [182]. The Traffic Anomaly Dataset (TAD) consists of 500 videos totaling 25 h, featuring abnormal actions such as vehicle accidents, illegal turns, illegal occupations, retrograde motion, pedestrian on road, road spills, and more [169]. The Universidad Panamericana Fall (UP-Fall) dataset includes 11 activities, as well as five different types of human falls, such as falling forward, backward, and sideways using hands or knees [183]. The Atomic Visual Actions (AVA) dataset annotates 80 atomic visuals for 437 video clips of human actions [184]. The Multiple Camera Fall (MCF) dataset was taken from eight cameras with different angles, capturing normal daily activities and simulated falls [185]. The University of Rzeszow-Fall (UR-Fall) dataset contains 70 videos, categorized into 30 fall videos and 40 daily living videos [186]. The VOC2007 dataset includes 9963 images, consisting of 24,640 annotated objects such as humans, animals, vehicles, and more [187]. The Penn-Fudan dataset contains 170 images with 345 labeled pedestrians, with 96 images from the University of Pennsylvania and 74 from Fudan University [188]. The UCF-50 dataset consists of 50 actions with a minimum of 100 videos for each category, taken from YouTube [189]. Extending from the UCF-50 dataset, the UCF-101 dataset contains 101 classes with a total of 13,320 clips [190].

The OpenImages dataset includes 600 object classes with a total of 3.68 million bounding boxes attached [191]. The Carnegie Mellon University (CMU) graphics lab dataset consists of 11 videos with a total of 2477 frames, 1268 of which depict abnormal actions [192]. The University of Texas (UT)-Interaction dataset contains videos of continuous human interactions, divided into six classes: shake-hands, point, hug, push, kick, and punch [193]. The Peliculas Movies (PEL) dataset includes 368 frames, of which 268 are fight frames taken from the movies [194]. The Web Dataset (WED) consists of 1280 frames comprising 12 sequences of normal crowd scenes such as walking and running and 8 scenes of abnormal scenes including escape panics, protesters clashing, and crowd fighting [195]. The Human Motion DataBase-51 (HMDB51) includes 51 action categories, which, in total, contains around 7000 manually annotated clips from YouTube [196]. The Kinetics-600 is a large human action dataset with 480,000 video clips and categorized into 600 action classes [197].

The YouTube Action dataset includes 1640 videos, categorized into 11 classes collected from YouTube [198].

The unsupervised methods, utilizing reconstruction-based techniques like AE, VAE, and CAE, demonstrate robustness in learning from unlabeled data, achieving notable AUC scores such as 0.984 on Ped2 by Wang et al. in 2023 [108] and 0.988 on Ped1 by Ganokratanaa et al. in 2022 [140]. These approaches capitalize on identifying patterns and anomalies without extensive labeled data, effectively addressing the challenge of data scarcity. Partially supervised methods, including semi-supervised and weakly supervised approaches, also show promising results with AUC scores like 0.945 on Ped1 by Sikdar and Chowdhury in 2020 [150], leveraging a combination of labeled and unlabeled data. Among these, state-of-the-art models continue to advance, exemplified by recent studies achieving high AUC scores through innovative techniques in abnormal behavior detection tasks. Table 2 shows the composition and features of some popular abnormal human behavior datasets. Figure 4 visually illustrates sample abnormal behavior images from each dataset.

Table 2. A summary of popular abnormal human behavior datasets.

Dataset	Characteristics	Merits	Challenges	Composition
Ped1 [171]	Anomalies include bikers, skaters, small carts, and people crossing	Ground truth annotations provided with binary flags per frame. Some clips include pixel-level masks for anomaly localization assessment	Perspectives include distortion, which might limit generalization	34 training videos, 36 test videos
Ped2 [171]	Pedestrian movement parallel to the camera plane from a top angle	Focuses on abnormal pedestrian motion patterns. Ground truth annotations provided with binary flags per frame. Some clips include pixel-level masks for anomaly localization assessment	Smaller dataset compared to Ped1	16 training videos, 12 testing videos
ST [172]	Includes abnormal behaviors caused by sudden motion, such as chasing and brawling	Pixel-level ground truth annotations of abnormal events	Complex lighting conditions and camera angles from all sides	270,000 training frames, 13 anomaly scenes
UCF-Crime [173]	Anomalies in real-world environments include abuse, arrest, arson, assault, accident, burglary, explosion, fighting, robbery, shooting, stealing, shoplifting, and vandalism	Extensive and diverse anomaly types relevant to public safety, with high-quality annotations by trained annotators, video-level labels for training, temporal annotations for testing, and a balanced set of 950 anomalous and 950 normal videos	Limited to surveillance footage, excluding other potential sources of anomalies	1900 videos with 13 classes
CUHK [174]	Contains unusual events such as running, throwing objects, and loitering	High frame rate detection (141.34 fps)	Unusual incidents with slight camera shake	30,652 frames, 14 abnormal classes
UMN [175]	Crowd behavior scenarios, where each video consists of a normal starting section and an abnormal ending section	Focuses on crowd behavior under panic conditions	Abnormal behavior typically appears at the end of the videos, which can lead to model overfitting	22 videos, 11 abnormal scenarios
UBI-Fights [178]	Various fighting scenarios in indoor and outdoor environments, with videos resized to 640 × 360 pixels and set to 30 fps	Provides a wide diversity of fighting scenarios with detailed frame-level annotations	Imbalance between fight and normal videos	1000 videos, where 216 videos contain a fight event, and 784 depict normal daily life situations

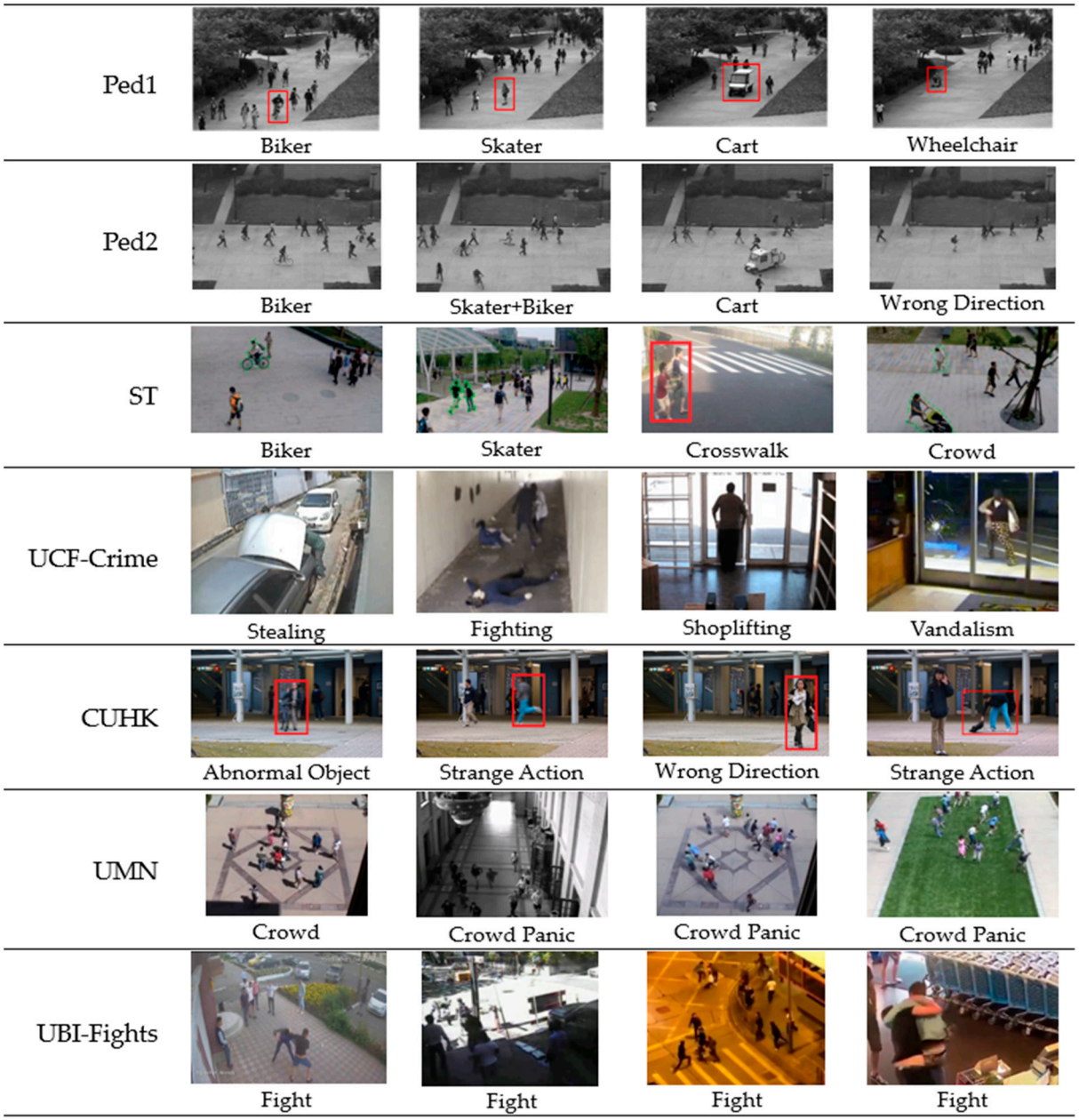

Figure 4. Sample abnormal behavior images from each dataset listed in Table 2.

4. Deep Learning Techniques for Abnormal Human Behavior Detection

4.1. Unsupervised Approach

Unsupervised detection refers to techniques that identify patterns, anomalies, or structures in data without the need for labeled examples. In the context of abnormal human behavior detection, this approach is particularly valuable because obtaining labels for various abnormal behaviors is often challenging and inefficient [199]. Two popular methods within this category are reconstruction-based and generative techniques. Reconstruction-based detection models learn patterns from input images, while generative detection models attempt to generate an artificial image that they have learned.

4.1.1. Reconstruction-Based Detection

Reconstruction-based methods model normal data distribution with the principle that the model is trained using only normal data. Anomalous data are then assigned high reconstruction errors by the model [200]. During the inference phase, if a test image is

abnormal, the model struggles to reconstruct the image. Reconstruction-based detection methods include auto-encoders (AE), variational auto-encoders (VAE), and convolutional auto-encoders (CAE). Auto-encoders are neural networks that learn input data and attempt to reconstruct new images based on previously learned patterns. Generally, AE consist of two structures: encoders and decoders. The objective is to minimize the reconstruction error, enabling the model to more accurately reconstruct images based on learned data. Table 3 summarizes the strengths and drawbacks of reconstruction-based methods for AHB detection with AUC scores on the Ped1, Ped2, and CUHK datasets.

Table 3. Strengths and Drawbacks of Reconstruction-based Methods.

	Research	Anomaly Datasets	Performance (AUC)			Strengths	Drawbacks
			Ped1	Ped2	CUHK		
Auto-encoders (AE)	Wang et al. [103] (2019)	Ped1, Ped2, UMN	0.897	0.913	N/A	Proposed a network for detecting abnormal events, which integrates a PCA network with kernel PCA	Reliance on hyperparameters and foreground detection may lead to false negatives by erroneously removing valid objects
	Hu et al. [104] (2019)	Ped1, Ped2, CUHK	0.809	0.959	0.842	Developed a three-stage framework for fast unsupervised anomaly detection in videos	May fail to detect instances such as a person walking with a bike
	Liu and Zhou [105] (2022)	Ped1, Ped2, CUHK	N/A	0.968	0.875	Proposed a memory-based connected network for video anomaly detection, utilizing an auto-encoder for reconstruction	The scoring threshold must be tuned for each environment
	Chang et al. [106] (2022)	Ped2, ST, CUHK	N/A	0.967	0.871	Proposed an auto-encoder for learning spatial and temporal regularity	Only detected abnormal events without classifying the object
	Wang et al. [107] (2022)	Ped1, Ped2, ST, CUHK	0.849	0.964	0.883	Proposed unsupervised video anomaly detection with frame prediction and noise tolerance loss	Require strategies for hyperparameter selection and model inference to ensure efficiency and accuracy
	Wang et al. [108] (2023)	Ped2, ST, CUHK	N/A	0.984	0.861	Proposed a pluggable spatio-temporal relationship attention module for indicating object relationships	Unable to fully utilize and understand the implicit video information
	Liu et al. [109] (2023)	Ped2, ST, CUHK	N/A	0.983	0.917	Proposed object-centric scene inference network for unsupervised video anomaly detection	Unable to identify the relationship between moving objects and background scenes
	Li et al. [110] (2023)	ST, CUHK	N/A	N/A	0.883	Proposed unsupervised algorithm based on skeleton features, eliminating manual specification of normal training data	May miss detecting some instances of abnormal pedestrian brawling but accurately identifies normal walking
	Yan et al. [9] (2023)	ST, UCF-Crime	N/A	N/A	N/A	Utilized auto-encoders and memory clustering to detect abnormal human actions	Challenges in crowd human pose prediction and conflicts in auto-encoders and clustering training
	Sampath and Kumar [111] (2023)	Ped1, Ped2, UMN	0.902	0.997	N/A	Proposed a spatiotemporal inter-fused auto-encoder for abnormal behavior detection	Reliant on a single modality, using only cameras for abnormal behavior detection

Table 3. *Cont.*

	Research	Anomaly Datasets	Performance (AUC)			Strengths	Drawbacks
			Ped1	Ped2	CUHK		
Variational Auto-encoders (VAE)	Wang et al. [112] (2019)	Ped1, CUHK, UMN, PETS	0.943	N/A	0.876	Used two VAEs for anomaly detection in crowded scenes	Very challenging due to the frame complexity
	Xu et al. [113] (2019)	Ped1, Ped2, ST	0.957	0.923	N/A	Introduced novel unsupervised VAE-based video anomaly detection approach	Dataset failure cases can hinder abnormal behavior detection performance
	Yan et al. [114] (2020)	Ped1, Ped2, CUHK, Subway	0.750	0.910	0.796	Proposed two-stream VAE structure: appearance and motion streams	Require additional resources for optical flow computation
	Wang et al. [115] (2021)	Ped2, ST	N/A	0.962	N/A	Proposed a cognitive memory-augmented network for decision-making based on past memory	Challenging to obtain normal sample distribution due to the dataset size
	Cho et al. [116] (2022)	Ped2, ST, UCF-Crime, CUHK, LV, UBI-Fights	N/A	0.992	0.880	Proposed implicit two-path auto-encoder with normal feature distribution modeling using normalizing flow	AE and normalizing flow model struggle to distinguish abnormal scenes due to visual similarity
	Huang et al. [117] (2022)	Ped2, CUHK, ST	N/A	0.981	0.888	Proposed temporal-aware contrastive network for unsupervised AHB detection	Require hyperparameter tuning to balance contrastive loss and task loss
	Wang et al. [118] (2022)	Ped1, Ped2, CUHK	0.884	0.888	0.872	Proposed double-flow convolutional LSTM with VAE probability calculation results	Challenging to detect small foreground target objects
	Slavic et al. [119] (2022)	Subway, CUHK	N/A	N/A	0.862	Proposed self-aware embodied agents for abnormal behavior detection, leveraging VAE regularization features	Challenging to detect camouflaged human objects in the background.
	Liu et al. [120] (2023)	Ped2, ST, CUHK	N/A	0.984	0.907	Proposed stochastic video normality network for unsupervised anomaly detection	Highly sensitive to hyperparameter settings
Convolutional Auto-encoders (CAE)	Chu et al. [121] (2019)	Ped1, Ped2, CUHK, Subway	0.909	0.902	0.937	Presented novel unsupervised spatiotemporal feature learning for video anomaly detection	Performance still unsatisfactory compared to fully supervised learning, which has made great progress
	Duman and Erdem [122] (2019)	Ped1, Ped2, CUHK	0.924	0.929	0.895	Detected AHB by generating reconstructed dense optical flow maps	Struggle to model distant activities
	Yan et al. [123] (2020)	Ped2, CUHK	N/A	0.892	N/A	Developed a 3D CAE for spatiotemporal irregularity detection in videos	Deeper layers in 3D convolutional auto-encoder may be unhelpful due to limited data
	Bahrami et al. [124] (2021)	Ped2, ST, CUHK,	N/A	0.975	0.801	Propose single-frame analysis and consideration of consecutive frames	Increased training time due to larger spatiotemporal architecture parameters
	Asad et al. [125] (2021)	Ped1, Ped2, CUHK, ST, Subway	0.898	0.958	0.892	Proposed two-staged CAE Framework for AHB detection	Takes a long time to train due to a large number of backpropagation iterations
	Li et al. [126] (2021)	Ped1, Ped2, CUHK, ST	0.850	0.951	0.888	Proposed CAE with extractor and latent code prediction for future frames	As training anomalies increase, the AUC score decreases

Table 3. *Cont.*

	Research	Anomaly Datasets	Performance (AUC)			Strengths	Drawbacks
			Ped1	Ped2	CUHK		
Convolutional Auto-encoders (CAE)	Wang et al. [127] (2022)	Ped2, CUHK	N/A	0.953	0.840	Combined criss-cross attention and bi-directional ConvLSTM in auto-encoder for AHB detection	AUC score improvement possible with added spatial and temporal features
	Kommanduri and Ghorai [128] (2023)	Ped1, Ped2, CUHK	0.847	0.977	0.867	Designed an end-to-end trainable bi-residual convolutional auto-encoder with long–short projection skip connections	Suffers from visual similarity and occlusions
	Taghinezhad and Yazdi [129] (2023)	Ped1, Ped2, CUHK	0.838	0.976	0.890	Introduced unsupervised video anomaly detection framework based on frame prediction	Significant improvements were not achieved in refined abnormality scores due to noise

Several of the recent studies have utilized auto-encoders for AHB detection. Wang et al. [108] and Sampath and Kumar [111] proposed a spatio-temporal AE, achieving an AUC value of over 0.98 for detecting abnormal behavior on the UCSD Ped1 and Ped2 datasets. However, the spatio-temporal AE is unable to fully utilize and understand the implicit video information, especially when using a single modality camera. To overcome these drawbacks, as illustrated in Figure 5, Liu et al. [109] developed AHB detection using an object-centric scene inference network (AUC 0.917 on CUHK), while Li et al. [110] utilized skeleton features to avoid the manual specification of normal data, achieving an AUC of 0.883 on the CUHK dataset. Both methods present better results compared with Wang et al. [108], who reported an AUC score of 0.861 on the CUHK dataset. Unfortunately, while the skeleton features accurately identify normal walking, they may miss detecting some instances of abnormal pedestrian brawling. Therefore, Yan et al. [9] introduced clustering and scoring system approaches using AE to better distinguish abnormal human behaviors. Wang et al. [103] also introduced a self-supervised framework known as the abnormal event detection network, comprising a principal component analysis (PCA) network and kernel principal component analysis. The framework achieved an outstanding AUC score of 0.997 using the UMN dataset. However, the method still depends on certain hyperparameters.

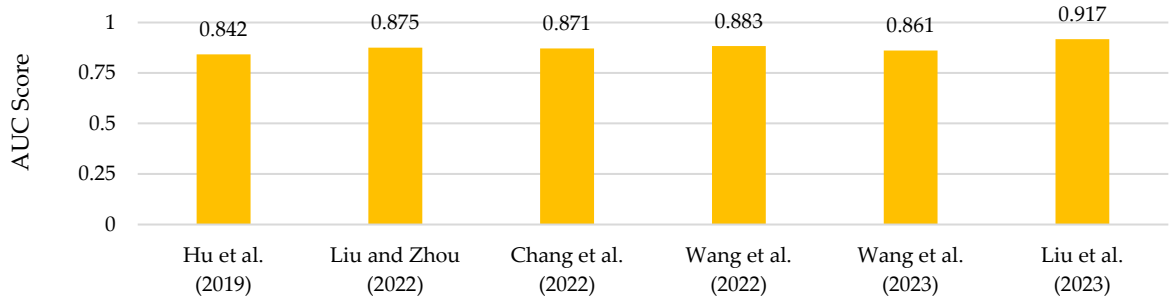

Figure 5. Reconstruction-based AHB detection results using AE on the CUHK dataset.

Additionally, foreground detection may inadvertently remove incorrect objects, resulting in false negative issues. VAEs are often conflated with traditional auto-encoders, despite being distinct entities. These models diverge in their mathematical formulations and objectives. VAE operates as a probabilistic generative model, requiring a neural network comprising an encoder and decoder. The encoder initially adjusts the parameters of the variational distribution, while the decoder maps from the latent space to the input space. VAEs are integral components of probabilistic graphical models and variational Bayesian methods [199,200].

Works on AHB detection using VAE were conducted by Wang et al. [112] to detect crowd scenes, which proved to be very challenging due to the complexity of frames. To address this issue, Yan et al. [114] and Wang et al. [118] generated a probability score using a double-flow VAE to differentiate abnormal behavior. However, the AUC score was not so high as when using two separate VAEs, as illustrated in Figure 6. Hence, several studies also employ temporal schemes to predict AHB scenes [115,117], resulting in significant improvements in AUC scores up to 0.961 on the Ped2 dataset. This indicates that the use of temporal schemes can substantially enhance the performance of AHB detection compared to methods relying solely on static features. The most recent research on AHB detection using the VAE method was conducted by Liu et al. [120]. They proposed stochastic video normality networks to learn various patterns of normal events in temporal, spatial, and spatiotemporal dimensions. The concept involves encoding past frames into a posterior distribution, from which latent variables are sampled using a VAE to predict future frames. The AUC results of this network reach 0.984 using the Ped2 dataset and 0.907 using the CUHK dataset. However, the performance of this network relies on hyperparameter settings for optimal AHB detection. Therefore, Cho et al. [116] introduced an implicit two-path auto-encoder and distribution modeling of normal features based on a normalizing flow model in an unsupervised manner for AHB detection. The achieved AUC score is impressive, reaching 0.992 using the Ped2 dataset and 0.880 using the CUHK dataset, indicating high performance. However, distinguishing between normal and abnormal scenes becomes challenging due to the similarity in appearance and motion between pedestrians and walking patterns. Consequently, the VAE and normalizing flow model struggle to differentiate between normal and abnormal behaviors.

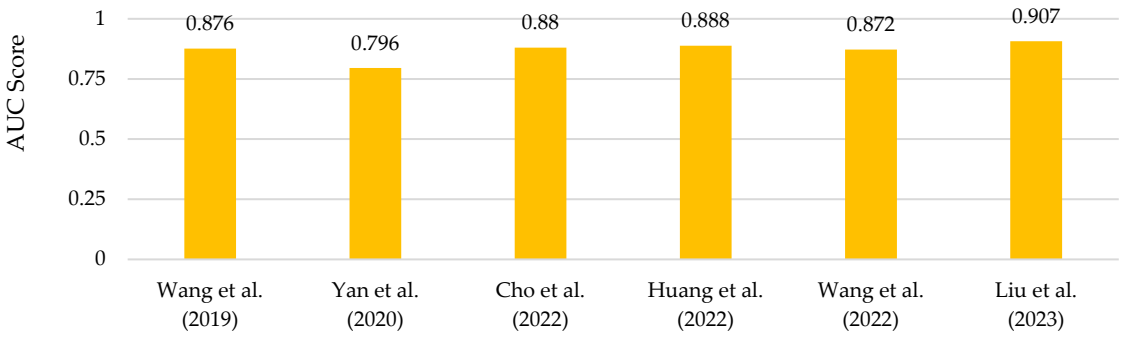

Figure 6. Reconstruction-based AHB detection results using VAE on the CUHK dataset.

Basic auto-encoders (AE), including VAE, do not consider the two-dimensional structure of an image. Therefore, a solution is required from an unsupervised learning paradigm that can evenly distribute weights for each area in the image. The convolutional auto-encoder is designed to preserve the spatial locality of the input image, which is then passed on to the reconstruction stage. Subsequently, reconstruction is carried out based on a linear combination of image patches using latent code [201]. Max-pooling is performed to ensure filter selectivity as an activation function across overlapping subregions. This prevents reliance on any single weight generated by multiple areas in the image. During the reconstruction phase, the sparse latent code further reduces the average filter contributing to the decoding phase of each pixel, resulting in filters with high generalization [202].

Bahrami et al. [124] achieved an AUC score of 0.975 on the Ped2 dataset for frame-level detection using a spatiotemporal approach. However, the training time increases due to complex larger spatiotemporal parameters. To overcome the challenge of preserving spatial information in the deep layers, Kommanduri and Ghorai [128] designed a bi-residual convolutional auto-encoder that is end-to-end trainable and introduces long–short projection skip connections. Additionally, Taghinezhad and Yazdi [129] proposed a novel multi-scale multi-path network architecture for AHB detection based on frame prediction. These two recent studies successfully achieved an AUC value above 0.976 using the Ped2

dataset, as illustrated in Figure 7. However, further research is needed on visual similarity, occlusions, and noise to achieve significant improvements in refined abnormality scores.

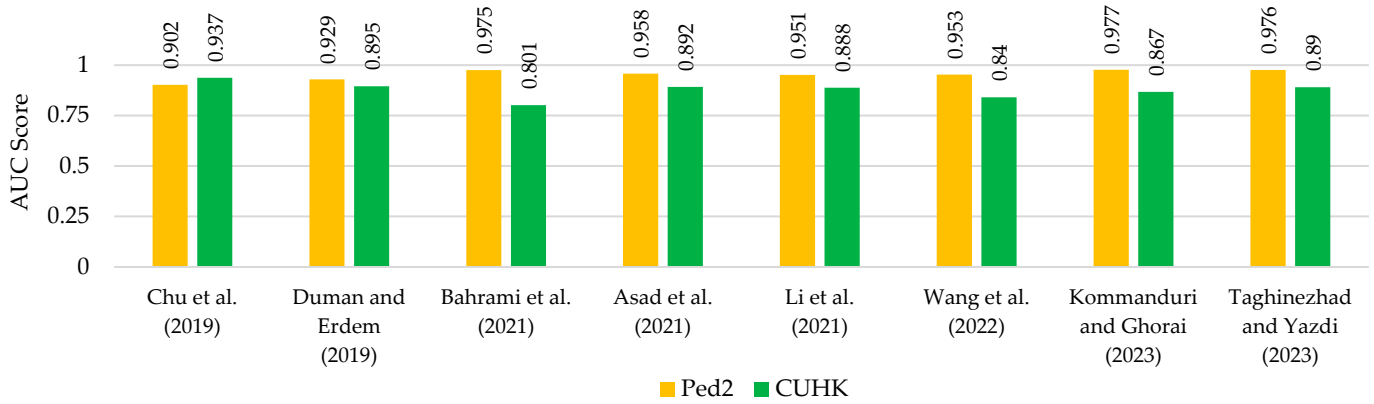

Figure 7. Reconstruction-based AHB detection using CAE on the Ped2 and CUHK datasets.

In summary, while AE-based methods, such as those by Wang et al. [103] and Hu et al. [104], demonstrate solid performance on Ped1 and Ped2 datasets, they often rely heavily on hyperparameters and struggle with false negatives and object removal. VAEs, such as those proposed by Cho et al. [116] and Huang et al. [117], show excellent results on Ped2 but are highly sensitive to hyperparameter settings and face challenges in distinguishing visually similar abnormal scenes. CAE models, exemplified by Chu et al. [121] and Duman and Erdem [122], effectively detect anomalies using spatiotemporal features but still fall short compared to fully supervised methods, especially in handling distant activities and complex scenes. Additionally, models across all types suffer from issues such as high computational costs, the need for extensive hyperparameter tuning, and difficulties in generalizing to different datasets. These findings underscore the necessity for further optimization to enhance model robustness, efficiency, and applicability to real-world scenarios.

4.1.2. Generative Detection

The artificial image is generated from a learned distribution pattern, and its similarity to the original image is assessed [203]. The difference between the original and fake images is used to detect whether abnormal human behavior is present in the captured frame. Since no labels are created, this approach remains within the unsupervised learning category. In generative adversarial networks (GANs), a random seed introduces some noise to the initial random image. Subsequently, the generator layers attempt to produce fake examples. The objective in this scenario is to generate the best normal image possible. Then, the real normal image serves as the second input to the discriminator layers. These layers aim to distinguish between the normal image and the fake image generated by the generator. The weights of both the discriminator and generator models are updated using the backpropagation method. This process iterates until the maximum number of training epochs, as previously specified.

The recent research on AHB detection using GANs was conducted by Li et al. [148] and Huang et al. [147,204], achieving AUC scores above 0.968 using the Ped2 dataset. However, the recent research requires computation of a large number of parameters. As the number of input frames increases, detection speed decreases, and there is difficulty in determining skip intervals for large foreground motion amplitudes in video anomaly detection. Ganokratanaa et al. [132,140] proposed a novel unsupervised spatiotemporal anomaly detection and localization for surveillance videos using GANs. The AUC scores reached 0.996 on the UMN dataset, demonstrating near-perfect performance. Additionally, the model may face difficulties in distinguishing similar abnormal events from normal patterns. Table 4 shows other works utilizing the generative detection approach with AUC scores on Ped1, Ped2, CUHK, and ST datasets.

Table 4. Strengths and Drawbacks of Generative Methods using GANs.

Research	Anomaly Datasets	Performance (AUC)				Strengths	Drawbacks
		Ped1	Ped2	CUHK	ST		
Li and Chang [131] (2019)	Ped1, Ped2, CUHK, UMN	0.850	0.916	0.842	N/A	Built on a two-stream framework for simultaneous appearance and motion anomaly detection	The lower AUC value is due to the noise removal of abnormal frames
Li et al. [133] (2019)	Ped1, Ped2, CUHK	0.838	0.966	0.845	N/A	Proposed novel spatiotemporal framework for video anomaly detection	Often fails to capture spatial characteristics due to camera angles
Ganokratanaa et al. [132] (2020)	Ped1, Ped2, UMN, CUHK	0.985	0.955	0.879	N/A	Proposed a spatiotemporal AHB detection and localization	Fails to detect abnormal events with similar object speeds
Wu et al. [137] (2021)	Ped1, Ped2, CUHK, ST	0.885	0.989	0.847	0.728	Used two independent GANs to predict optical flows or color frames	The model needs updates.
Yang et al. [138] (2021)	CUHK, Ped1, Ped2, ST	0.847	0.976	0.886	0.745	Proposed bidirectional prediction generator: forward and backward	The model struggles with small human objects in the presence of perspective distortion
Ganokratanaa et al. [140] (2022)	Ped1, Ped2, CUHK, UMN	0.988	0.976	0.908	N/A	Introduced unsupervised deep residual spatiotemporal translation network for video anomaly detection and localization	May struggle to distinguish similar abnormal events from normal patterns
Yu et al. [141] (2022)	Ped1, Ped2, CUHK, Subway, UCF-Crime	0.979	0.979	0.949	N/A	Proposed adversarial event prediction to detect rare pattern events in abnormal human behaviors	Absence of background detection preprocessing leads to slightly lower performance metrics in various scenarios
Zhong et al. [142] (2022)	Ped1, Ped2, CUHK, ST	0.826	0.977	0.889	0.707	Proposed cascade model: frame reconstruction and optical flow network with GAN	The average optical flow prediction error of normal frames increases due to perspective phenomena in datasets
Aslam et al. [143] (2022)	Ped1, Ped2, CUHK, ST	0.907	0.977	0.894	0.869	Proposed end-to-end trainable two-stream attention-based adversarial auto-encoder network	Struggles to learn typical features with small datasets
Hao et al. [144] (2022)	Ped1, Ped2, CUHK, ST	0.825	0.969	0.866	0.738	Proposed spatiotemporal consistency-enhanced network	Based on 3D CNN, struggles to converge if object size varies significantly
Yu et al. [205] (2022)	Ped1, Ped2, CUHK, UCF-Crime	0.975	0.971	0.947	N/A	Proposed adversarial predictive coding for abnormal event detection and localization	Requires large-scale dataset and motion data but increases the computational cost

Table 4. *Cont.*

Research	Anomaly Datasets	Performance (AUC)				Strengths	Drawbacks
		Ped1	Ped2	CUHK	ST		
Huang et al. [147] (2023)	Ped2, ST, CUHK	N/A	0.977	0.897	0.758	Predicted future frames using previous video frames and optical flow	Requires computing a large number of parameters
Huang et al. [206] (2023)	Ped1, Ped2, ST, CUHK	0.921	0.976	0.888	0.743	Proposed self-supervised attentive GAN for video anomaly detection	Detection speed decreases with increasing input frame numbers
Li et al. [148] (2023)	Ped2, CUHK, ST	N/A	0.968	0.887	0.767	Explored adversarial composite prediction for normal video dynamics learning feasibility	Difficulty in determining skip intervals for large foreground motion amplitudes in video anomaly detection

When conducting comparisons, it is essential to ensure fairness by comparing the prior research using the same dataset. As illustrated in Figure 8, results using the ST dataset tend to be lower than those from other datasets. Interestingly, Aslam et al. [143] proposed an end-to-end trainable two-stream attention-based approach that achieved an AUC score of 0.869 on the ST dataset and 0.894 on the CUHK dataset, which are the best results using these datasets compared to other studies. This is because during the inference stage, only the reconstruction branch is considered for computing the regularity score, while the prediction branch is utilized for better feature learning through GAN. These results highlight the need for further research to combine generative detection and reconstruction-based detection to achieve more optimal outcomes.

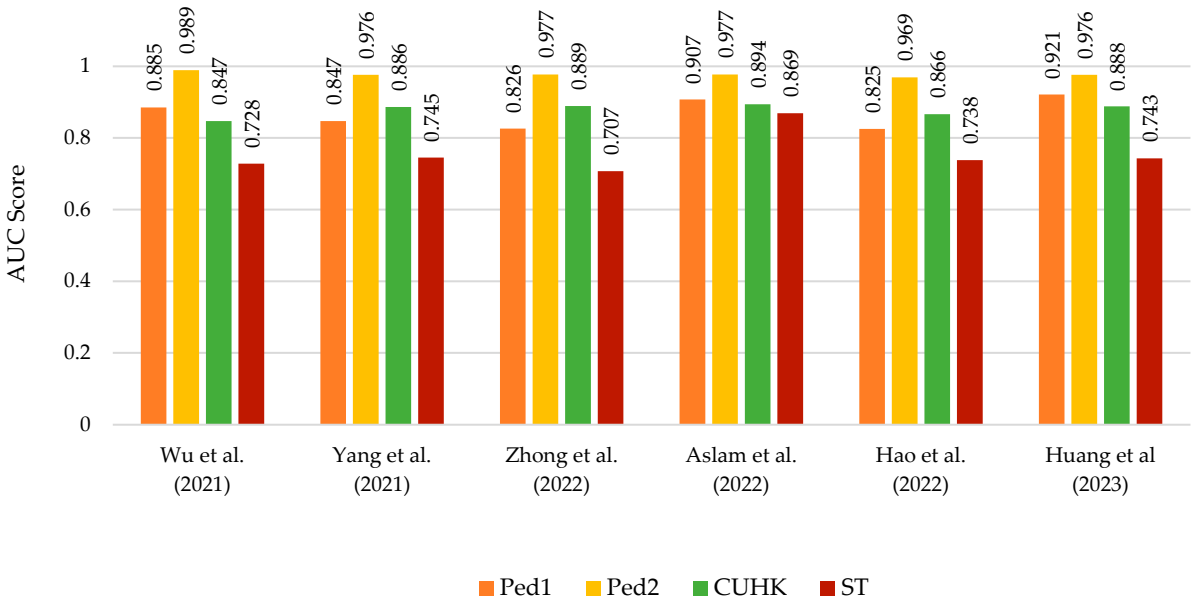

Figure 8. Generative AHB detection results on the Ped1, Ped2, CUHK, and ST datasets.

4.2. Partially Supervised Approach

Most reconstructive or generative approaches solely utilize normal samples, potentially resulting in a high false positive rate, particularly in real-world scenarios [154]. Many individuals commonly associate model training using labeled data with supervised learning, while training with unlabeled data is often termed unsupervised learning. However,

real-world situations often lack sufficient data for comprehensive training due to the high cost and time-consuming nature of full labeling [204].

Partially supervised learning occurs when both labeled and unlabeled data are available. Therefore, the question arises: How do partially supervised learning techniques leverage unlabeled data to improve model performance when trained on limited labeled data? Labeled data act as anchor points for training and prediction phases with the unlabeled data [149]. In partially supervised detection, two main schemes emerge: semi-supervised detection and weakly supervised detection.

Table 5 provides a summary of partially supervised detection research. Semi-supervised approaches, exemplified by Sikdar and Chowdhury [150] and Wu et al. [137], achieve high AUC scores (up to 0.989 for Ped2) through adaptive training and re-learning schemes but encounter challenges with sparse datasets, local descriptor construction, and dependencies on baseline models. In contrast, weakly supervised methods, such as those by Ullah et al. [162] and Chen et al. [164], leverage weakly labeled data, achieving high performance (up to 0.984 for Ped2) across diverse environments. However, they face challenges such as high false alarm rates, occlusion issues, and significant computational demands, particularly with transformer-based models. Therefore, while semi-supervised methods excel in data-specific performance, weakly supervised techniques offer broader applicability at the cost of increased complexity.

Table 5. Strengths and Drawbacks of Partially Supervised Approach.

	Research	Anomaly Datasets	Performance (AUC)				Strengths	Drawbacks
			Ped1	Ped2	UCF-Crime	ST		
Semi-supervised	Sikdar and Chowdhury [150] (2020)	Ped1, Ped2, CUHK, UMN, ST	0.945	0.979	N/A	N/A	Proposed adaptive training-less anomaly detection method	Performance lag due to sparse dataset and difficulty in constructing local descriptors
	Singh et al. [151] (2021)	CUHK, PETS, UMN	N/A	N/A	N/A	N/A	Proposed algorithm for suspicious event detection based on direction and magnitude	Not suitable for real-time application due to time-consuming optical flow calculation for each frame
	Wu et al. [137] (2021)	Ped1, CUHK	0.885	0.989	N/A	N/A	Implemented semi-supervised re-learning scheme to boost the baseline approach. Constructed new training selectively from the original testing set	Model performance is positively related to the baseline deep model, but occasional failure cases still occur

Table 5. *Cont.*

	Research	Anomaly Datasets	Performance (AUC)				Strengths	Drawbacks
			Ped1	Ped2	UCF-Crime	ST		
Weakly supervised	Li et al. [145] (2022)	Ped1, Ped2 UCF-Crime, ST	0.833	0.954	0.785	0.903	Proposed attention-based multiple instances learning using attention-based features and a stringent loss	Not robust to significant occlusion
	Hu et al. [154] (2020)	Ped1, Ped2, CUHK, UMN, Subway	N/A	N/A	N/A	N/A	Trained a discriminative classifier for anomaly detection with weakly labeled data	Unable to achieve end-to-end detection of abnormal behaviors
	Degardin and Proença [155] (2021)	Ped1, Ped2, UBI-Fights, UCF-Crime	0.819	0.819	0.769	N/A	Introduced an iterative learning framework, based on weakly and self-supervised paradigms	Performance gap between indoor and outdoor scenarios
	Ullah et al. [156] (2022)	UCF-Crime, Surveillance Fight, Hockey Fight, Violent Flows, ST	N/A	N/A	0.858	0.849	Introduced a dual-stream CNN framework for detecting anomalous events in surveillance and non-surveillance environments	Some highly complex video sequences are mispredicted, contributing to model failure cases
	Yi et al. [157] (2022)	UCF-Crime, ST	N/A	N/A	0.843	0.977	Presented a scheme to assess anomaly degree and used triplet loss to optimize the network	Limited discrimination for unseen normal events, leading to high false alarm rates
	Liu et al. [158] (2022)	Ped2, ST, UCF-Crime	N/A	0.914	0.831	0.882	Proposed a collaborative normality learning framework to address weakly supervised video anomaly detection	Some false detection cases due to image obscuration and low-resolution
	Ullah et al. [162] (2023)	Ped2, CUHK, ST	N/A	0.984	N/A	0.946	Proposed a weakly supervised hybrid CNN- and transformer-based framework to learn anomalous events using video-level labels	Transformer approach requires more computational resources due to model parameter variation
	Shao et al. [163] (2023)	UCF-Crime, ST	N/A	N/A	0.851	0.953	Enhanced temporal features for the entire video sequence, redefining integrity and coherence	Limited interpretability
	Chen et al. [164] (2023)	Ped2, ST, UCF-Crime, TAD	N/A	0.974	0.803	0.972	Proposed a spatial-temporal graph attention network to address video anomaly detection	Local discriminative representations may deteriorate in long videos with complex scenes, resulting in underfitting

Table 5. *Cont.*

	Research	Anomaly Datasets	Performance (AUC)				Strengths	Drawbacks
			Ped1	Ped2	UCF-Crime	ST		
Weakly supervised	Tang et al. [165] (2023)	UCF-Crime, ST	N/A	N/A	0.843	0.967	Prior knowledge guided pseudo label generator and improved self-guided attention encoder	High training time cost, and pseudo-label generator not robust enough
	Zhang and Xue [166] (2023)	UCF-Crime, Ped2	N/A	0.941	0.832	N/A	Proposed sub-Max method for anomaly detection	Pixel-level AUC result is suboptimal
	Wang et al. [167] (2023)	UCF-Crime, ST	N/A	N/A	0.815	0.940	Proposed attention mechanism-guided multi-instance learning weakly supervised video anomaly detection method	Difficulty in detecting anomalies in low-resolution video, challenging to evaluate confusing actions without additional context

4.2.1. Semi-Supervised Detection

Semi-supervised learning offers a method of learning the underlying structure of data using both labeled and unlabeled data [204]. It falls within the partially supervised detection approach, which is commonly encountered in real-world scenarios due to the limited availability of labeled data [207]. The semi-supervised detection scheme emphasizes augmenting limited labeled data with unlabeled data [208]. The model is initially trained with labeled data to understand underlying patterns. It then uses predictions on unlabeled data to create pseudo-labels. The model is subsequently retrained with this combination of labeled and pseudo-labeled data, improving generalization. Techniques such as consistency regularization and graph-based methods are also used to ensure the model produces consistent predictions and propagates label information from labeled data to unlabeled data [209].

Sikdar and Chowdhury [150] introduced an adaptive training-less method for anomaly detection. The model identifies abnormal behavior without pre-training, dynamically adjusting certain model parameters during runtime. Achieving an AUC of 0.992 on the UMN dataset, the method has shown promising results. However, a slight performance lag was observed, attributed to the sparse dataset nature and challenges in constructing local descriptors. Singh et al. [151] proposed an algorithm for suspicious event detection based on direction and magnitude using a semi-supervised scheme, achieving an impressive AUC score of 0.999 using the UMN dataset, indicating near-perfect performance. However, the method's suitability for real-time applications is limited due to the time-consuming optical flow calculation required for each frame. Wu et al. [137] further enhanced this baseline approach with a semi-supervised re-learning scheme. They constructed a new training set by selectively extracting training instances from the original testing set, resulting in an increased AUC score from 0.858 to 0.885 on UCSD Ped1 datasets. Nevertheless, the model's performance remains closely tied to that of the baseline deep model, and occasional failure cases may still occur.

The research gap in semi-supervised detection includes developing methodologies that effectively handle sparse datasets and improve the accuracy of local descriptors. Exploring innovative semi-supervised and adaptive learning strategies could enhance the adaptability and robustness of anomaly detection models across different datasets and environments. Closing these gaps is essential for advancing anomaly detection systems in practical applications.

4.2.2. Weakly Supervised Detection

Weakly supervised learning aims to generate predictions with high information content [204]. Unlike the semi-supervised scheme, the weakly supervised scheme focuses on enhancing detection results with limited labeled data [208]. In weakly supervised detection, the model uses labels that are less precise, coarser, or noisier than fully supervised labels. Techniques such as multiple instance learning, where groups of instances are labeled rather than individual instances [210], and expectation-maximization algorithms, which estimate the most likely labels and optimize model parameters, are used [211]. Regularization and adjustment of the loss function help the model deal with noise in the labels. This scheme addresses the challenge of missing training data without requiring extensive object annotation [212].

Recent works in weakly supervised AHB detection have utilized both video-level data [162–164,167], and image-level data [165,166]. These recent works have demonstrated promising results, achieving AUC scores above 0.940 using the ST dataset. Temporal features have been a primary focus in these recent works to better distinguish abnormal human behavior. However, drawbacks of video-level weakly supervised AHB detection include increased computational resources required by transformers due to model parameter variation, limited interpretability of results, potential degradation of local discriminative representations in lengthy videos, and challenges in detecting anomalies in low-resolution videos. Additionally, image-level weakly supervised AHB detection faces several drawbacks, including high training time costs and a pseudo-label generator that lacks robustness.

Therefore, Ullah et al. [156] introduced a dual-stream CNN framework for detecting anomalous events in surveillance and non-surveillance environments. The first scheme employs a two-dimensional CNN as an auto-encoder for visual feature extraction and further utilizes temporal relations. Subsequently, three-dimensional features are extracted and integrated into two-dimensional spatiotemporal features for accurate detection. The achieved AUC scores are high, nearly 0.990 on the Violent Flow and Hockey Fight dataset, which indicates excellent performance. As illustrated in Figure 9, this work also achieves the highest result compared to others, reaching 0.858 using UCF-Crime. However, some extremely complex video sequences were mispredicted, contributing to the failure cases of the proposed model. This finding suggests that combining reconstruction-based and weakly supervised approaches could be a promising avenue for future research to integrate temporal and spatiotemporal features.

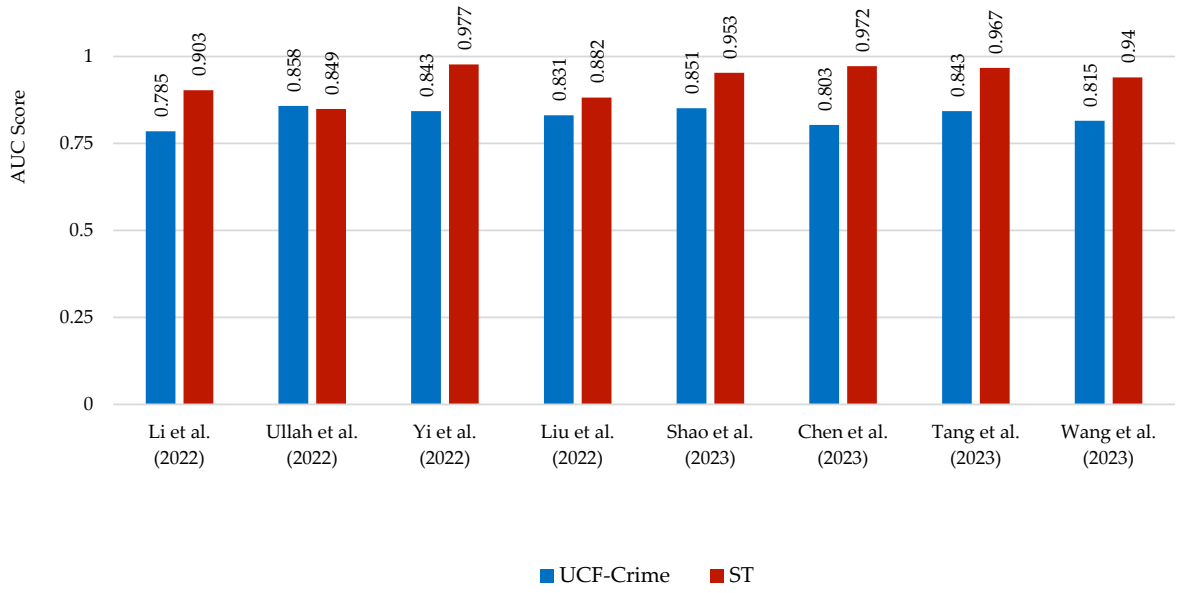

Figure 9. Weakly supervised AHB detection results on the UCF-Crime and ST datasets.

4.3. Fully Supervised Approach

Fully supervised learning is a paradigm where the model is trained to produce accurate detection outputs based on input data [59]. Therefore, most research employing fully supervised learning schemes uses accuracy as the primary metric to assess model performance [213]. Typically, the model captures local features using CNN layers and then incorporates them into the LSTM layer to learn temporal relationships between features [93].

The recent works in fully supervised AHB detection focus on specific tasks such as fall detection [84,85] and detecting suspicious activity at automated teller machines [94]. However, the effectiveness of the approach depends on factors such as image quality, camera position, and the presence of subjects. Moreover, fall detection encounters challenges in scenarios involving actions like crouching and sitting. The recent research addresses these challenges by combining CNN with LSTM and GRU to learn temporal features within video sequences [91,92]. Some frameworks also utilize Kalman Filter, dual-stream CNN, dual-attentional CNN (DA-CNN), and Bi-GRU. These studies use various datasets, including UP-Fall, AVA, UR-Fall, MCF, VOC2007, Penn-Fudan, OpenImages, CMU, UT-Interaction, PEL, Hockey Fight, WED, Ped1, Ped2, HMDB51, UCF-50, UCF-101, Kinetics-600, and YouTube Action. However, there are challenges in detecting motion on edge devices, and the model may generate non-zero probabilities for certain action classes. An outstanding result was achieved by Ahn et al. [89], who developed a vision-based factory safety monitoring system to detect human presence on assembly lines. They utilized YOLOv3 as a base model and employed the OpenImages dataset. The accuracy achieves a precision of 0.999 and a recall value of 0.964 across 24 detection classes. However, concerns were raised regarding the detection quality due to lens distortion issues.

The research gap in fully supervised detection encompasses several challenges: effectiveness depends on factors such as image quality, camera position, and subject presence. Additionally, optimizing hyperparameters for optimal results proves challenging. Lens distortion issues contribute to reduced detection accuracy compared to state-of-the-art methods. Furthermore, higher computational requirements increase costs, while motion detection on edge devices may result in non-zero probabilities for certain action classes. Table 6 summarizes the strengths and drawbacks of fully supervised AHB detection research.

Table 6. Strengths and Drawbacks of Fully Supervised Approach.

Research	Anomaly Datasets	Framework	Strengths	Drawbacks
Espinosa et al. [84] (2019)	UP-Fall	CNN	Presented multi-camera vision-based fall detection and classification system using CNN	The efficacy depends on image quality, camera position, and subject presence
Gomes et al. [85] (2022)	AVA, MCF, UR-Fall	CNN, Kalman Filter	Combined CNN and Kalman filter for fall tracking	Fall detection faces challenges in scenarios like crouching and sitting
Sivachandiran et al. [88] (2022)	VOC2007, Penn-Fudan	CNN	Enhanced model for person detection and tracking on surveillance videos	Difficulty in hyperparameter tuning for optimum results
Ahn et al. [89] (2023)	OpenImages	CNN	Designed vision-based factory safety monitoring system for detecting human presence on assembly lines	Low detection performance due to lens distortion issues
Michael Onyema et al. [90] (2023)	CMU, UT-Interaction, PEL, Hockey Fight, WED, Ped1, Ped2	CNN	Designed slow–fast CNN for abnormal behavior identification in surveillance videos	Consumes more computational time, increasing costs

Table 6. *Cont.*

Research	Anomaly Datasets	Framework	Strengths	Drawbacks
Hussain et al. [91] (2023)	HMDB51, UCF-50, YouTube Action	Dual-stream CNN	Proposed dual-stream network combining image enhancement, convolutional, and transformer techniques	Unable to detect motion in edge devices
Ullah and Munir [92] (2023)	HMDB51, UCF-50, UCF-101, YouTube Action, Kinetics-600	DA-CNN, Bi-GRU	Proposed cascaded spatial-temporal discriminative feature-learning framework for human activity recognition in video streams	May produce non-zero probabilities for some action classes
Kshirsagar and Azath [94] (2023)	YouTube Action	CNN	Used heuristic-assisted deep learning techniques for detecting suspicious human activities in the automated teller machines	Accuracy is slightly lower than state-of-the-art methods

4.4. Summary: Advantages and Disadvantages

Offering a definitive answer to the question "Which abnormal human behavior detection technique is most suitable for a specific application?" may not always be practical. Therefore, this section explains the advantages and disadvantages of each deep learning technique. In the reconstruction-based approach, the auto-encoder method is commonly employed for image dimension reduction. This learning process is subsequently utilized to compute the loss function within the network, which is used to identify abnormal behavior within the input data. However, the distribution of these data presents a significant challenge when implementing models in heterogeneous environments. Additionally, the effectiveness of the model using this auto-encoder method is also contingent upon data quality. Table 7 summarizes the advantages and disadvantages of each deep learning technique.

Methods have been devised to address generalization issues arising from data distribution using VAE. Techniques such as regularization and probabilistic formulation are incorporated into VAE methods, enabling their applications in detecting abnormal behavior across more heterogeneous environments. However, VAEs may struggle to identify abnormal behavior occurring within specific time frames. Moreover, probabilistic calculations throughout the image can lead to small object sizes and slightly disrupted pixel localization. A specialized strategy is required to fully leverage the potential of VAEs in detecting abnormal behavior in surveillance videos.

Within the reconstruction-based approaches, convolutional auto-encoders are employed to maximize detection for each pixel in the image. CAEs offer benefits due to their generalization capabilities, enabling weight distribution across all areas of the input image. This feature allows the model to localize and identify abnormal human behaviors within specific regions. However, convolutional networks often require augmentation with other methods to optimize output. Therefore, techniques like max-pooling and deconvolutional layers from a supervised learning framework are used to assist in distributing weight dependencies for each pixel in the image.

Generative detection approaches excel in recognizing environments unseen during the learning phase. This advantage can be leveraged by refining the generator module to improve the quality of training images, mitigate noise in images, and distinguish between foreground and background objects to streamline detection targets. However, careful consideration is required regarding the model's priorities, whether to emphasize the model's adaptability to objects or prioritize computational time and resources. This prioritization is crucial as it influences training paradigms and the balance between the generator and discriminator. By understanding the model's requirements, training objectives become more focused, enabling the mitigation of several drawbacks associated with generative detection approaches in identifying abnormal human behaviors.

Table 7. Advantages and disadvantages of deep-learning-based AHB detection approaches.

Approach	Methods	Advantages	Disadvantages
Unsupervised	Reconstruction-based Detection (AE, VAE, CAE)	• Can perform image dimensionality reduction • Capable of localizing AHB in frames • Can be trained solely with normal data • Handles data scarcity effectively	• Challenged with data distribution when implemented in heterogeneous environments • Difficult to accurately capture abnormal behavior occurring within certain timeframes • Disrupted by pixel localization
	Generative Detection (GAN)	• Capable of detecting AHB in new environments • Easily select AHB due to foreground object subtraction ability • Handles data scarcity effectively	• Experiences a gradient exploding problem • Challenged with high complexity • Requires a significant amount of computational resources
Partially Supervised	Semi-supervised Detection, Weak-supervised Detection	• Extends the amount of labeled data with minimal supervision • Rapidly modifies models based on recent data • Reduces time and minimizes human efforts for data labeling • Performs well in training with scarce data • Enable quick detection of video sequences • Handles data scarcity effectively	• Noisy labeled data as an anchor complicates the training phase • Limited interpretability • Suboptimal for pixel-level AHB detection
Fully Supervised	CNN, LSTM, GRU	• Useful for predefined AHB detection classes • Beneficial for short-term AHB detection	• Requires substantial human effort for data labeling • Not beneficial for long-term AHB detection • Consumes significant computing resources and time • Unable to handle data scarcity problems

In partially supervised detection, a semi-supervised detection scheme offers an alternative path to training detection models by concentrating on developing training data. By leveraging a small amount of labeled data, the focus shifts to instructing unlabeled data to serve as new references for subsequent model training. Through a pseudo-labeling scheme, additional labeled data can be generated with minimal supervision. This scheme finds diverse applications, particularly in adapting models to the latest data reflecting evolving real-world environments. Notably, it significantly reduces labeling time, minimizing human effort. However, it is crucial to acknowledge that if the reference data used for labeling are noisy or of poor quality, subsequent processes become more challenging. For endeavors aiming to augment unlabeled data and enhance dataset quality, the utilization of the semi-supervised detection scheme is recommended.

Weakly supervised detection is another scheme of partially supervised detection. This scheme focuses on training the model to yield higher-quality information. The emphasis lies in utilizing a small amount of data while still producing an accurate model. Much research has adapted layers and weighting strategies to enable models to learn patterns from sparsely labeled datasets and generalize from them. There are various advantages, such as quicker detection of video sequences and detecting abnormal human behavior. It is essential to note that hyperparameter tuning is crucial here. Recent advancements offer promising strategies to optimize model performance by tuning fewer hyperparameters, known as visual tuning [214]. Visual tuning holds significant potential for growth through the application of self-supervised learning, which enables models to learn from vast amounts of unlabeled data. This approach reduces the need for extensive labeled datasets and improves the model's ability to generalize across different scenarios of abnormal behavior. Additionally, the input data must be of high quality. Weakly supervised detection is utilized

when the focus and ultimate goal are on maximizing model output to make the model more informative and accurate.

Training a model to detect abnormal human behavior using fully supervised learning is highly beneficial if the behavior category to be detected has been determined. Detection using CNN is feasible only for short-term detection, i.e., detecting abnormal behavior at the frame level. CNN is less efficient for detecting behavior that requires a certain period to determine its abnormality, known as long-term abnormal behavior. Some research incorporates LSTM and GRU to assess behavior temporally [87,92,93]. Unfortunately, data on abnormal human behaviors are scarce and expensive. Therefore, the fully supervised scheme necessitates more human effort solely to determine normal and abnormal data in the model training phase. Additionally, this scheme consumes significant computing resources and time due to the diverse nature of abnormal human behaviors.

5. Open Research Issues

This section discusses open research issues with deep learning techniques for abnormal human behavior detection in surveillance videos. Table 8 outlines the open research issues for each deep learning approach.

For unsupervised detection, particularly within the reconstruction-based approach, managing the variability of data distribution is a significant challenge, as it changes according to the detection environment of the target object. Addressing this requires developing strategies to handle the high levels of environmental variability in the data, potentially leading to new research directions. Additionally, accurately detecting AHB within specific temporal frames poses a challenge due to the difficulty of detection within certain timeframes. Thus, developing coherent temporal models is crucial to prioritize these temporal AHB detections effectively. The VAE method faces issues with pixel localization accuracy, necessitating alternative mechanisms to improve AHB detection precision. This challenge warrants further exploration as one of many open research questions aimed at creating a more sophisticated AHB detector.

Table 8. Open research issues with deep learning techniques for AHB detection in surveillance videos.

Approach	Methods	Open Research Issues
Unsupervised	Reconstruction-based Detection (AE, VAE, CAE)	• Handling high environmental variation in data • Finding optimal combinations with other temporal models • Introducing additional mechanisms for a more fine-grained AHB detector to address pixel localization disruption
	Generative Detection (GAN)	• Addressing the gradient exploding issue through the development of model-learning techniques • Managing complexity for smoother detection • Developing solutions to reduce high computational resources requirements
Partially Supervised	Semi-supervised Detection, Weak-supervised Detection	• Formulating strategies for better understanding AHB in contextual settings • Implementing efficient preprocessing phases to clean noisy data • Developing strategies to integrate every pixel to gain comprehensive information
Fully Supervised	CNN, LSTM, GRU	• Designing lightweight AHB models • Integrating with LSTM and GRU for long-term AHB detection

Concerning generative detection, addressing issues such as gradient exploding is of significant importance. This phenomenon causes abrupt changes in weight values during calculations, disrupting the learning process, particularly as the model performance heavily relies on the training data [141]. This presents a significant challenge as the model needs to

efficiently and effectively generalize to new data. Establishing mechanisms for improving model learning while preserving the model's benefits is the core principle for overcoming these challenges. As the model processes more data, the demand for computer resources increases, raising questions about whether the current resources can meet the demand for AHB detection. This research problem arises from the need to model and allocate proper computing resources of sufficient quality. Concerning complexity reduction, several questions arise. Integration between the generator and discriminator is necessary, as their inability to work in unison undoubtedly leads to poor model performance [119]. Conversely, managing complexity is equally important to maintain the pace of the detection process.

The predominant challenge in partially supervised detection is limited interpretability. In this context, labels produced either automatically or manually may not always be accurate, casting doubt on the validity of the model interpretation. Therefore, special attention should be given to the development of new approaches to better predict abnormal human behavior. The noisy anchor data are also challenging, leading to a research question about implementing a robust preprocessing phase. Additionally, pixel-level AHB detection becomes inaccurate due to suboptimal performance, resulting in inaccurate results [166]. The extraction technique may fail to capture important information at the pixel level, leading to the omission of vital details necessary for AHB detection. Hence, the classifier should employ an adequate approach to integrate the detected object pixel information.

Another challenge associated with fully supervised detection is the requirement for high computational resources. Analyzing this situation, an effective lightweight model is needed. The research on lightweight AHB detection brings many benefits, especially in rapid inference, faster decision-making, and minimal computational consumption. However, there are still challenges in detecting long-term AHB using fully supervised learning. Therefore, the research issue opened in the use of LSTM and GRU for the feasibility of long-term AHB detection follows the path of fully supervised detection.

6. Conclusions

The detection of abnormal human behavior in video surveillance systems is a crucial task, yet datasets demonstrating such behavior are scarce and costly to collect. Hence, developing strategies that effectively utilize optimal abnormal data alongside accurate models becomes imperative. To address this challenge, we present a comprehensive survey of deep learning techniques for abnormal human behavior detection in surveillance videos. This survey begins by defining abnormal human behaviors and categorizing them into three detection approaches: unsupervised, partially supervised, and fully supervised. Each approach is extensively described, including its strengths and drawbacks. Additionally, we conduct a comparative analysis of the prior research findings on popular benchmarking datasets. In unsupervised detection, the reconstruction-based detection approach excels in reducing image dimensionality and localizing AHB in normal data. However, it struggles with environmental diversity and inaccurate timeframe detection, often due to pixel localization issues. Generative detection approaches are adept at identifying AHB in unfamiliar scenarios and addressing data shortages. Yet, they face challenges like exploding gradients, high complexity, and significant computational demands. Partially supervised detection mitigates data scarcity by enhancing limited labeled data with minimal supervision. However, it grapples with noisy labeled data, limited interpretability, and suboptimal AHB detection at the pixel level. Fully supervised detection, while suitable for defined AHB detection classes, is resource-intensive for labeling and less effective in long-term detection. Additionally, its data scarcity poses a trade-off between comprehensiveness and efficacy. Finally, we discuss several open research issues in AHB detection, including the issue of high environmental variation data, optimizing temporal AHB detection, tackling gradient exploding, and reducing computational resource usage. Through investigation of these potential research issues, we aim to drive progress in this field, ultimately bringing greater benefits to video surveillance systems in the future.

Author Contributions: Conceptualization, L.M.W. and S.G.K.; literature survey, L.M.W.; writing—original draft preparation, L.M.W.; writing—review and editing, S.G.K. and L.W.; supervision, S.G.K.; funding acquisition, S.G.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by MSIT of Korea under Grant 2019-0-00231 (Development of Artificial Intelligence Based Video Security Technology and Systems for Public Infrastructure Safety).

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: The authors are grateful to Jalil Piran for his comments on the paper's structure and survey.

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

References

1. Ito, R.; Tsukada, M.; Kondo, M.; Matsutani, H. An Adaptive Abnormal Behavior Detection using Online Sequential Learning. In Proceedings of the 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), New York, NY, USA, 1–3 August 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 436–440. [CrossRef]
2. Antonakaki, P.; Kosmopoulos, D.; Perantonis, S.J. Detecting abnormal human behaviour using multiple cameras. *Signal Process.* **2009**, *89*, 1723–1738. [CrossRef]
3. Kim, D.; Kim, H.; Mok, Y.; Paik, J. Real-Time Surveillance System for Analyzing Abnormal Behavior of Pedestrians. *Appl. Sci.* **2021**, *11*, 6153. [CrossRef]
4. Yoon, Y.-I.; Chun, J.-A. Tracking Model for Abnormal Behavior from Multiple Network CCTV Using the Kalman Filter. In *Computer Science and Its Applications: Ubiquitous Information Technologies*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 933–939. [CrossRef]
5. Park, H.-J. A Study on Monitoring System for an Abnormal Behaviors by Object's Tracking. *J. Digit. Contents Soc.* **2013**, *14*, 589–596. [CrossRef]
6. Patwal, A.; Diwakar, M.; Tripathi, V.; Singh, P. An investigation of videos for abnormal behavior detection. *Procedia Comput. Sci.* **2023**, *218*, 2264–2272. [CrossRef]
7. Tay, N.C.; Connie, T.; Ong, T.S.; Teoh, A.B.J.; Teh, P.S. A Review of Abnormal Behavior Detection in Activities of Daily Living. *IEEE Access* **2023**, *11*, 5069–5088. [CrossRef]
8. Wu, C.; Cheng, Z. A Novel Detection Framework for Detecting Abnormal Human Behavior. *Math. Probl. Eng.* **2020**, *2020*, 6625695. [CrossRef]
9. Yan, M.; Xiong, Y.; She, J. Memory Clustering Autoencoder Method for Human Action Anomaly Detection on Surveillance Camera Video. *IEEE Sens. J.* **2023**, *23*, 20715–20728. [CrossRef]
10. Sinulingga, H.R.; Kong, S.G. Key-Frame Extraction for Reducing Human Effort in Object Detection Training for Video Surveillance. *Electronics* **2023**, *12*, 2956. [CrossRef]
11. Wei, H.; Kehtarnavaz, N. Simultaneous Utilization of Inertial and Video Sensing for Action Detection and Recognition in Continuous Action Streams. *IEEE Sens. J.* **2020**, *20*, 6055–6063. [CrossRef]
12. Kim, B.; Lee, J. A Video-Based Fire Detection Using Deep Learning Models. *Appl. Sci.* **2019**, *9*, 2862. [CrossRef]
13. Wu, Q.; Zhou, Y.; Wu, X.; Liang, G.; Ou, Y.; Sun, T. Real-time running detection system for UAV imagery based on optical flow and deep convolutional networks. *IET Intell. Transp. Syst.* **2020**, *14*, 278–287. [CrossRef]
14. Zhao, Z.; Lan, S.; Zhang, S. Human Pose Estimation based Speed Detection System for Running on Treadmill. In Proceedings of the 2020 International Conference on Culture-Oriented Science & Technology (ICCST), Beijing, China, 28–31 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 524–528. [CrossRef]
15. Chen, K.-Y.; Shin, J.; Hasan, M.A.M.; Liaw, J.-J. Deep Transfer Learning Based Real Time Fitness Movement Identification. In Proceedings of the 2022 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Shah Alam, Malaysia, 25 June 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 102–106. [CrossRef]
16. Cao, Y.; Fan, S.; Cheng, W.; Zhao, Y.; Zheng, H.; Zhao, H. Human Body Movement Velocity Estimation Based on Binocular Video Streams. In Proceedings of the 2022 3rd International Conference on Computer Vision. Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA), Changchun, China, 20–22 May 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 977–985. [CrossRef]
17. Lao, S.; Wang, D.; Li, F.; Zhang, H. Human running detection: Benchmark; baseline. *Comput. Vis. Image Underst.* **2016**, *153*, 143–150. [CrossRef]
18. Ha, T.V.; Nguyen, H.M.; Thanh, S.H.; Nguyen, B.T. Fall detection using mixtures of convolutional neural networks. *Multimed. Tools Appl.* **2023**, *83*, 18091–18118. [CrossRef]

19. Yan, J.; Wang, X.; Shi, J.; Hu, S. Skeleton-Based Fall Detection with Multiple Inertial Sensors Using Spatial-Temporal Graph Convolutional Networks. *Sensors* **2023**, *23*, 2153. [CrossRef] [PubMed]
20. Zi, X.; Chaturvedi, K.; Braytee, A.; Li, J.; Prasad, M. Detecting Human Falls in Poor Lighting: Object Detection and Tracking Approach for Indoor Safety. *Electronics* **2023**, *12*, 1259. [CrossRef]
21. Zheng, K.; Li, B.; Li, Y.; Chang, P.; Sun, G.; Li, H.; Zhang, J. Fall detection based on dynamic key points incorporating preposed attention. *Math. Biosci. Eng.* **2023**, *20*, 11238–11259. [CrossRef]
22. Hoang, V.-H.; Lee, J.W.; Piran, M.J.; Park, C.-S. Advances in Skeleton-Based Fall Detection in RGB Videos: From Handcrafted to Deep Learning Approaches. *IEEE Access* **2023**, *11*, 92322–92352. [CrossRef]
23. Wastupranata, L.M.; Munir, R. Convolutional neural network-based crowd detection for COVID-19 social distancing protocol from unmanned aerial vehicles onboard camera. *J. Appl. Remote Sens.* **2023**, *17*, 44502. [CrossRef]
24. Kalshetty, R.; Parveen, A. Abnormal event detection model using an improved ResNet101 in context aware surveillance system. *Cogn. Comput. Syst.* **2023**, *5*, 153–167. [CrossRef]
25. Alafif, T.; Hadi, A.; Allahyani, M.; Alzahrani, B.; Alhothali, A.; Alotaibi, R.; Barnawi, A. Hybrid Classifiers for Spatio-Temporal Abnormal Behavior Detection, Tracking, and Recognition in Massive Hajj Crowds. *Electronics* **2023**, *12*, 1165. [CrossRef]
26. Bhuiyan, M.R.; Abdullah, J.; Hashim, N.; Al Farid, F.; Uddin, J. Hajj pilgrimage abnormal crowd movement monitoring using optical flow and FCNN. *J. Big Data* **2023**, *10*, 86. [CrossRef]
27. Hanif, M.S.; Bilal, M.; Balamash, A.S.; Al-Saggaf, U.M. Hypotheses Generation and Verification Based Framework for Crowd Anomaly Detection in Single-Scene Surveillance Videos. *Trait. Signal* **2023**, *40*, 115–122. [CrossRef]
28. Castellano, G.; Cotardo, E.; Mencar, C.; Vessio, G. Density-based clustering with fully-convolutional networks for crowd flow detection from drones. *Neurocomputing* **2023**, *526*, 169–179. [CrossRef]
29. Zubair, M.; Ali, A.; Naeem, S.; Anam, S. Video Streams for The Detection of Thrown Objects from Expressways. In Proceedings of the MOL2NET'22, Conference on Molecular, Biomedical & Computational Sciences and Engineering, 8th Ed.—MOL2NET: FROM MOLECULES TO NETWORKS, Paris, France, 1–15 January 2023; p. 13932. [CrossRef]
30. Ali, M.M. Real-time video anomaly detection for smart surveillance. *IET Image Process* **2023**, *17*, 1375–1388. [CrossRef]
31. Mahankali, S.; Kabbin, S.V.; Nidagundi, S.; Srinath, R. Identification of Illegal Garbage Dumping with Video Analytics. In Proceedings of the 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 19–22 September 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 2403–2407. [CrossRef]
32. Chaturvedi, K.; Dhiman, C.; Vishwakarma, D.K. Fight detection with spatial and channel wise attention-based ConvLSTM model. *Expert Syst.* **2024**, *41*, e13474. [CrossRef]
33. Pervaiz, M.; Shorfuzzaman, M.; Alsufyani, A.; Jalal, A.; Alsuhibany, S.A.; Park, J. Tracking and Analysis of Pedestrian's Behavior in Public Places. *Comput. Mater. Contin.* **2023**, *74*, 841–853. [CrossRef]
34. Alarfaj, M.; Pervaiz, M.; Ghadi, Y.Y.; al Shloul, T.; Alsuhibany, S.A.; Jalal, A.; Park, J. Automatic Anomaly Monitoring in Public Surveillance Areas. *Intell. Autom. Soft Comput.* **2023**, *35*, 2655–2671. [CrossRef]
35. Jebur, S.A.; Hussein, K.A.; Hoomod, H.K.; Alzubaidi, L. Novel Deep Feature Fusion Framework for Multi-Scenario Violence Detection. *Computers* **2023**, *12*, 175. [CrossRef]
36. Bashir, M.; Rundensteiner, E.A.; Ahsan, R. A deep learning approach to trespassing detection using video surveillance data. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 3535–3544. [CrossRef]
37. Zhang, Z.; Zaman, A.; Xu, J.; Liu, X. Artificial intelligence-aided railroad trespassing detection and data analytics: Methodology and a case study. *Accid. Anal. Prev.* **2022**, *168*, 106594. [CrossRef]
38. Grabušić, S.; Barić, D. A Systematic Review of Railway Trespassing: Problems and Prevention Measures. *Sustainability* **2023**, *15*, 13878. [CrossRef]
39. Zaman, A.; Ren, B.; Liu, X. Artificial Intelligence-Aided Automated Detection of Railroad Trespassing. *Transp. Res. Rec. J. Transp. Res. Board.* **2019**, *2673*, 25–37. [CrossRef]
40. Havârneanu, G.M. Behavioural and organisational interventions to prevent trespass and graffiti vandalism on railway property. *Proc. Inst. Mech. Eng. F J. Rail Rapid. Transit.* **2017**, *231*, 1078–1087. [CrossRef]
41. Zhang, T.; Aftab, W.; Mihaylova, L.; Langran-Wheeler, C.; Rigby, S.; Fletcher, D.; Maddock, S.; Bosworth, G. Recent Advances in Video Analytics for Rail Network Surveillance for Security, Trespass and Suicide Prevention—A Survey. *Sensors* **2022**, *22*, 4324. [CrossRef] [PubMed]
42. Bamaqa, A.; Sedky, M.; Bosakowski, T.; Bastaki, B.B.; Alshammari, N.O. SIMCD: SIMulated crowd data for anomaly detection and prediction. *Expert Syst. Appl.* **2022**, *203*, 117475. [CrossRef]
43. Mehmood, A. Abnormal Behavior Detection in Uncrowded Videos with Two-Stream 3D Convolutional Neural Networks. *Appl. Sci.* **2021**, *11*, 3523. [CrossRef]
44. Pouyan, S.; Charmi, M.; Azarpeyvand, A.; Hassanpoor, H. Propounding First Artificial Intelligence Approach for Predicting Robbery Behavior Potential in an Indoor Security Camera. *IEEE Access* **2023**, *11*, 60471–60489. [CrossRef]
45. Chen, H.; Bohush, R.; Kurnosov, I.; Ma, G.; Weichen, Y.; Ablameyko, S. Detection of Appearance and Behavior Anomalies in Stationary Camera Videos Using Convolutional Neural Networks. *Pattern Recognit. Image Anal.* **2022**, *32*, 254–265. [CrossRef]
46. Patel, A.S.; Vyas, R.; Vyas, O.P.; Ojha, M.; Tiwari, V. Motion-compensated online object tracking for activity detection and crowd behavior analysis. *Vis. Comput.* **2023**, *39*, 2127–2147. [CrossRef] [PubMed]

47. Wahyono; Harjoko, A.; Dharmawan, A.; Adhinata, F.D.; Kosala, G.; Jo, K.-H. Loitering Detection Using Spatial-Temporal Information for Intelligent Surveillance Systems on a Vision Sensor. *J. Sens. Actuator Netw.* **2023**, *12*, 9. [CrossRef]
48. Huang, T.; Han, Q.; Min, W.; Li, X.; Yu, Y.; Zhang, Y. Loitering Detection Based on Pedestrian Activity Area Classification. *Appl. Sci.* **2019**, *9*, 1866. [CrossRef]
49. Dwivedi, N.; Singh, D.K.; Kushwaha, D.S. An Approach for Unattended Object Detection through Contour Formation using Background Subtraction. *Procedia Comput. Sci.* **2020**, *171*, 1979–1988. [CrossRef]
50. Agarwal, H.; Singh, G.; Siddiqui, M.A. Classification of Abandoned and Unattended Objects, Identification of Their Owner with Threat Assessment for Visual Surveillance. In *Proceedings of 3rd International Conference on Computer Vision and Image Processing*; Chaudhuri, B., Nakagawa, M., Khanna, P., Kumar, S., Eds.; Springer: Singapore, 2020; pp. 221–232. [CrossRef]
51. Htun, B.; Sein, M.M. Observation of Unattended or Removed Object in Public Area for Security Monitoring System. In *Genetic and Evolutionary Computing*; Springer International Publishing: Cham, Switzerland, 2017; pp. 45–53. [CrossRef]
52. Park, H.; Park, S.; Joo, Y. Robust Real-time Detection of Abandoned Objects using a Dual Background Model. *KSII Trans. Internet Inf. Syst.* **2020**, *14*, 771–788. [CrossRef]
53. Bangare, P.S.; Bangare, S.L.; Yawle, R.U.; Patil, S.T. Detection of human feature in abandoned object with modern security alert system using Android Application. In Proceedings of the 2017 International Conference on Emerging Trends & Innovation in ICT (ICEI), Pune, India, 3–5 February 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 139–144. [CrossRef]
54. Planinc, R.; Kampel, M. Detecting Unusual Inactivity by Introducing Activity Histogram Comparisons. In Proceedings of the 9th International Conference on Computer Vision Theory and Applications, SCITEPRESS—Science and and Technology Publications, Lisbon, Portugal, 5–8 January 2014; pp. 313–320. [CrossRef]
55. Koehler, S.; Goldhammer, M.; Bauer, S.; Zecha, S.; Doll, K.; Brunsmann, U.; Dietmayer, K. Stationary Detection of the Pedestrian's Intention at Intersections. *IEEE Intell. Transp. Syst. Mag.* **2013**, *5*, 87–99. [CrossRef]
56. Yi, S.; Li, H.; Wang, X. Pedestrian Behavior Modeling From Stationary Crowds With Applications to Intelligent Surveillance. *IEEE Trans. Image Process.* **2016**, *25*, 4354–4368. [CrossRef] [PubMed]
57. Deep, S.; Zheng, X.; Karmakar, C.; Yu, D.; Hamey, L.G.C.; Jin, J. A Survey on Anomalous Behavior Detection for Elderly Care Using Dense-Sensing Networks. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 352–370. [CrossRef]
58. Nayak, R.; Pati, U.C.; Das, S.K. A comprehensive review on deep learning-based methods for video anomaly detection. *Image Vis. Comput.* **2021**, *106*, 104078. [CrossRef]
59. Choudhry, N.; Abawajy, J.; Huda, S.; Rao, I. A Comprehensive Survey of Machine Learning Methods for Surveillance Videos Anomaly Detection. *IEEE Access* **2023**, *11*, 114680–114713. [CrossRef]
60. Patrikar, D.R.; Parate, M.R. Anomaly detection using edge computing in video surveillance system: Review. *Int. J. Multimed. Inf. Retr.* **2022**, *11*, 85–110. [CrossRef]
61. Xefteris, V.-R.; Tsanousa, A.; Meditskos, G.; Vrochidis, S.; Kompatsiaris, I. Performance, Challenges, and Limitations in Multimodal Fall Detection Systems: A Review. *IEEE Sens. J.* **2021**, *21*, 18398–18409. [CrossRef]
62. Roka, S.; Diwakar, M.; Singh, P.; Singh, P. Anomaly behavior detection analysis in video surveillance: A critical review. *J. Electron. Imaging* **2023**, *32*, 42106. [CrossRef]
63. Newaz, N.T.; Hanada, E. The Methods of Fall Detection: A Literature Review. *Sensors* **2023**, *23*, 5212. [CrossRef] [PubMed]
64. Jenga, K.; Catal, C.; Kar, G. Machine learning in crime prediction. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 2887–2913. [CrossRef]
65. Pandiaraja, P.; Saarumathi, R.; Parashakthi, M.; Logapriya, R. An Analysis of Abnormal Event Detection and Person Identification from Surveillance Cameras using Motion Vectors with Deep Learning. In Proceedings of the 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2–4 March 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1225–1232. [CrossRef]
66. Zhou, Z.-H.; Schwenker, F. *Partially Supervised Learning*; Springer: Berlin/Heidelberg, Germany, 2013. [CrossRef]
67. Ren, J.; Xia, F.; Liu, Y.; Lee, I. Deep Video Anomaly Detection: Opportunities and Challenges. In Proceedings of the 2021 International Conference on Data Mining Workshops (ICDMW), Auckland, New Zealand, 7–10 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 959–966. [CrossRef]
68. Hao, Y.; Tang, Z.; Alzahrani, B.; Alotaibi, R.; Alharthi, R.; Zhao, M.; Mahmood, A. An End-to-End Human Abnormal Behavior Recognition Framework for Crowds With Mentally Disordered Individuals. *IEEE J. Biomed. Health Inf.* **2022**, *26*, 3618–3625. [CrossRef] [PubMed]
69. Zhang, C.; Li, G.; Xu, Q.; Zhang, X.; Su, L.; Huang, Q. Weakly Supervised Anomaly Detection in Videos Considering the Openness of Events. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 21687–21699. [CrossRef]
70. Zhu, S.; Chen, C.; Sultani, W. Video Anomaly Detection for Smart Surveillance. In *Computer Vision*; Springer International Publishing: Cham, Switzerland, 2020; pp. 1–8. [CrossRef]
71. Wang, Y.; Qin, C.; Bai, Y.; Xu, Y.; Ma, X.; Fu, Y. Making Reconstruction-based Method Great Again for Video Anomaly Detection. In Proceedings of the 2022 IEEE International Conference on Data Mining (ICDM), Orlando, FL, USA, 28 November–1 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1215–1220. [CrossRef]
72. Ganokratanaa, T.; Aramvith, S.; Sebe, N. Anomaly Event Detection Using Generative Adversarial Network for Surveillance Videos. In Proceedings of the 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Lanzhou, China, 18–21 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1395–1399. [CrossRef]

73. Popoola, O.P.; Wang, K. Video-Based Abnormal Human Behavior Recognition—A Review. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **2012**, *42*, 865–878. [CrossRef]
74. Wu, X.; Ou, Y.; Qian, H.; Xu, Y. A detection system for human abnormal behavior. In Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, AB, Canada, 2–6 August 2005; IEEE: Piscataway, NJ, USA, 2005; pp. 1204–1208. [CrossRef]
75. Fei, F.; Fang, Z.; Shu, L. A fast algorithm based on human visual system for abnormal event detection. In Proceedings of the 2017 International Conference on Computer, Information and Telecommunication Systems (CITS), Dalian, China, 21–23 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 185–189. [CrossRef]
76. Tran, C.H.; Kong, S.G. An Iterative Learning Scheme with Binary Classifier for Improved Event Detection in Surveillance Video. *Electronics* **2023**, *12*, 3275. [CrossRef]
77. Jin, C.; Wang, T.; Alhusaini, N.; Zhao, S.; Liu, H.; Xu, K.; Zhang, J. Video Fire Detection Methods Based on Deep Learning: Datasets, Methods, and Future Directions. *Fire* **2023**, *6*, 315. [CrossRef]
78. Cao, X.; Su, Y.; Geng, X.; Wang, Y. YOLO-SF: YOLO for Fire Segmentation Detection. *IEEE Access* **2023**, *11*, 111079–111092. [CrossRef]
79. Yam, C.; Nixon, M.S.; Carter, J.N. On the relationship of human walking and running: Automatic person identification by gait. In *Object Recognition Supported by User Interaction for Service Robots*; IEEE Computer Society: Washington, DC, USA, 2002; pp. 287–290. [CrossRef]
80. Gutiérrez, J.; Martin, S.; Rodriguez, V. Human stability assessment and fall detection based on dynamic descriptors. *IET Image Process* **2023**, *17*, 3177–3195. [CrossRef]
81. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
82. Shubber, M.S.M.; Al-Ta'i, Z.T.M. A review on video violence detection approaches. *Int. J. Nonlinear Anal. Appl. (IJNAA)* **2022**, *13*, 1117–1130. [CrossRef]
83. Zhao, X.; Wang, L.; Zhang, Y.; Han, X.; Deveci, M.; Parmar, M. A review of convolutional neural networks in computer vision. *Artif. Intell. Rev.* **2024**, *57*, 99. [CrossRef]
84. Espinosa, R.; Ponce, H.; Gutiérrez, S.; Martínez-Villaseñor, L.; Brieva, J.; Moya-Albor, E. A vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case study using the UP-Fall detection dataset. *Comput. Biol. Med.* **2019**, *115*, 103520. [CrossRef]
85. Gomes, M.E.N.; Macêdo, D.; Zanchettin, C.; de-Mattos-Neto, P.S.G.; Oliveira, A. Multi-human Fall Detection and Localization in Videos. *Comput. Vis. Image Underst.* **2022**, *220*, 103442. [CrossRef]
86. Chandrakala, S.; Vignesh, L.K.P. V2AnomalyVec: Deep Discriminative Embeddings for Detecting Anomalous Activities in Surveillance Videos. *IEEE Trans. Comput. Soc. Syst.* **2022**, *9*, 1307–1316. [CrossRef]
87. Gandapur, M.Q. E2E-VSDL: End-to-end video surveillance-based deep learning model to detect and prevent criminal activities. *Image Vis. Comput.* **2022**, *123*, 104467. [CrossRef]
88. Sivachandiran, S.; Mohan, K.J.; Nazer, G.M. Deep Learning driven automated person detection and tracking model on surveillance videos. *Meas. Sens.* **2022**, *24*, 100422. [CrossRef]
89. Ahn, J.; Park, J.; Lee, S.S.; Lee, K.-H.; Do, H.; Ko, J. SafeFac: Video-based smart safety monitoring for preventing industrial work accidents. *Expert. Syst. Appl.* **2023**, *215*, 119397. [CrossRef]
90. Onyema, E.M.; Balasubaramanian, S.; Suguna S, K.; Iwendi, C.; Prasad, B.V.V.S.; Edeh, C.D. Remote monitoring system using slow-fast deep convolution neural network model for identifying anti-social activities in surveillance applications. *Meas. Sens.* **2023**, *27*, 100718. [CrossRef]
91. Hussain, A.; Khan, S.U.; Khan, N.; Rida, I.; Alharbi, M.; Baik, S.W. Low-light aware framework for human activity recognition via optimized dual stream parallel network. *Alex. Eng. J.* **2023**, *74*, 569–583. [CrossRef]
92. Ullah, H.; Munir, A. Human Activity Recognition Using Cascaded Dual Attention CNN and Bi-Directional GRU Framework. *J. Imaging* **2023**, *9*, 130. [CrossRef] [PubMed]
93. Mao, J.; Zhou, P.; Wang, X.; Yao, H.; Liang, L.; Zhao, Y.; Zhang, J.; Ban, D.; Zheng, H. A health monitoring system based on flexible triboelectric sensors for intelligence medical internet of things and its applications in virtual reality. *Nano Energy* **2023**, *118*, 108984. [CrossRef]
94. Kshirsagar, A.P.; Azath, H. YOLOv3-based human detection and heuristically modified-LSTM for abnormal human activities detection in ATM machine. *J. Vis. Commun. Image Represent.* **2023**, *95*, 103901. [CrossRef]
95. Alzubaidi, L.; Bai, J.; Al-Sabaawi, A.; Santamaría, J.; Albahri, A.S.; Al-dabbagh, B.S.N.; Fadhel, M.A.; Manoufali, M.; Zhang, J.; Al-Timemy, A.H.; et al. A survey on deep learning tools dealing with data scarcity: Definitions, challenges, solutions, tips, and applications. *J. Big Data* **2023**, *10*, 46. [CrossRef]
96. Baxter, R.H.; Robertson, N.M.; Lane, D.M. Human behaviour recognition in data-scarce domains. *Pattern Recognit.* **2015**, *48*, 2377–2393. [CrossRef]
97. Tu, H.; Allanach, J.; Singh, S.; Pattipati, K.R.; Willett, P. Information integration via hierarchical and hybrid bayesian networks. *IEEE Trans. Syst. Man Cybern.—Part A Syst. Hum.* **2006**, *36*, 19–33. [CrossRef]
98. Duong, H.-T.; Le, V.-T.; Hoang, V.T. Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey. *Sensors* **2023**, *23*, 5024. [CrossRef] [PubMed]

99. Lavee, G.; Rivlin, E.; Rudzsky, M. Understanding Video Events: A Survey of Methods for Automatic Interpretation of Semantic Occurrences in Video. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **2009**, *39*, 489–504. [CrossRef]
100. Gawlikowski, J.; Tassi, C.R.N.; Ali, M.; Lee, J.; Humt, M.; Feng, J.; Kruspe, A.; Triebel, R.; Jung, P.; Roscher, R.; et al. A survey of uncertainty in deep neural networks. *Artif. Intell. Rev.* **2023**, *56*, 1513–1589. [CrossRef]
101. Myagmar-Ochir, Y.; Kim, W. A Survey of Video Surveillance Systems in Smart City. *Electronics* **2023**, *12*, 3567. [CrossRef]
102. Şengönül, E.; Samet, R.; Al-Haija, Q.A.; 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. [CrossRef]
103. Wang, T.; Miao, Z.; Chen, Y.; Zhou, Y.; Shan, G.; Snoussi, H. AED-Net: An Abnormal Event Detection Network. *Engineering* **2019**, *5*, 930–939. [CrossRef]
104. Hu, J.; Zhu, E.; Wang, S.; Liu, X.; Guo, X.; Yin, J. An Efficient and Robust Unsupervised Anomaly Detection Method Using Ensemble Random Projection in Surveillance Videos. *Sensors* **2019**, *19*, 4145. [CrossRef] [PubMed]
105. Liu, Q.; Zhou, X. A Fully Connected Network Based on Memory for Video Anomaly Detection. In Proceedings of the 2022 IEEE 8th International Conference on Cloud Computing and Intelligent Systems (CCIS), Chengdu, China, 26–28 November 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 221–226. [CrossRef]
106. Chang, Y.; Tu, Z.; Xie, W.; Luo, B.; Zhang, S.; Sui, H.; Yuan, J. Video anomaly detection with spatio-temporal dissociation. *Pattern Recognit.* **2022**, *122*, 108213. [CrossRef]
107. Wang, X.; Che, Z.; Jiang, B.; Xiao, N.; Yang, K.; Tang, J.; Ye, J.; Wang, J.; Qi, Q. Robust Unsupervised Video Anomaly Detection by Multipath Frame Prediction. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 2301–2312. [CrossRef] [PubMed]
108. Wang, Y.; Liu, T.; Zhou, J.; Guan, J. Video anomaly detection based on spatio-temporal relationships among objects. *Neurocomputing* **2023**, *532*, 141–151. [CrossRef]
109. Liu, Y.; Guo, Z.; Liu, J.; Li, C.; Song, L. OSIN: Object-Centric Scene Inference Network for Unsupervised Video Anomaly Detection. *IEEE Signal Process Lett.* **2023**, *30*, 359–363. [CrossRef]
110. Li, N.; Chang, F.; Liu, C. A Self-Trained Spatial Graph Convolutional Network for Unsupervised Human-Related Anomalous Event Detection in Complex Scenes. *IEEE Trans. Cogn. Dev. Syst.* **2023**, *15*, 737–750. [CrossRef]
111. Sampath, D.K.; Kumar, K. Abnormal Crowd Behaviour Detection in Surveillance Videos Using Spatiotemporal Inter-Fused Autoencoder. *Int. J. Intell. Eng. Syst.* **2023**, *16*, 470–481. [CrossRef]
112. 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.* **2019**, *14*, 1390–1399. [CrossRef]
113. Xu, M.; Yu, X.; Chen, D.; Wu, C.; Jiang, Y. An Efficient Anomaly Detection System for Crowded Scenes Using Variational Autoencoders. *Appl. Sci.* **2019**, *9*, 3337. [CrossRef]
114. Yan, S.; Smith, J.S.; Lu, W.; Zhang, B. Abnormal Event Detection From Videos Using a Two-Stream Recurrent Variational Autoencoder. *IEEE Trans. Cogn. Dev. Syst.* **2020**, *12*, 30–42. [CrossRef]
115. Wang, T.; Xu, X.; Shen, F.; Yang, Y. A Cognitive Memory-Augmented Network for Visual Anomaly Detection. *IEEE/CAA J. Autom. Sin.* **2021**, *8*, 1296–1307. [CrossRef]
116. Cho, M.; Kim, T.; Kim, W.J.; Cho, S.; Lee, S. Unsupervised video anomaly detection via normalizing flows with implicit latent features. *Pattern Recognit.* **2022**, *129*, 108703. [CrossRef]
117. Huang, C.; Wu, Z.; Wen, J.; Xu, Y.; Jiang, Q.; Wang, Y. Abnormal Event Detection Using Deep Contrastive Learning for Intelligent Video Surveillance System. *IEEE Trans. Ind. Inf.* **2022**, *18*, 5171–5179. [CrossRef]
118. Wang, L.; Tan, H.; Zhou, F.; Zuo, W.; Sun, P. Unsupervised Anomaly Video Detection via a Double-Flow ConvLSTM Variational Autoencoder. *IEEE Access* **2022**, *10*, 44278–44289. [CrossRef]
119. Slavic, G.; Baydoun, M.; Campo, D.; Marcenaro, L.; Regazzoni, C. Multilevel Anomaly Detection Through Variational Autoencoders and Bayesian Models for Self-Aware Embodied Agents. *IEEE Trans. Multimed.* **2022**, *24*, 1399–1414. [CrossRef]
120. Liu, Y.; Yang, D.; Fang, G.; Wang, Y.; Wei, D.; Zhao, M.; Cheng, K.; Liu, J.; Song, L. Stochastic video normality network for abnormal event detection in surveillance videos. *Knowl. Based Syst.* **2023**, *280*, 110986. [CrossRef]
121. Chu, W.; Xue, H.; Yao, C.; Cai, D. Sparse Coding Guided Spatiotemporal Feature Learning for Abnormal Event Detection in Large Videos. *IEEE Trans. Multimed.* **2019**, *21*, 246–255. [CrossRef]
122. Duman, E.; Erdem, O.A. Anomaly Detection in Videos Using Optical Flow and Convolutional Autoencoder. *IEEE Access* **2019**, *7*, 183914–183923. [CrossRef]
123. Yan, M.; Meng, J.; Zhou, C.; Tu, Z.; Tan, Y.-P.; Yuan, J. Detecting spatiotemporal irregularities in videos via a 3D convolutional autoencoder. *J. Vis. Commun. Image Represent.* **2020**, *67*, 102747. [CrossRef]
124. Bahrami, M.; Pourahmadi, M.; Vafaei, A.; Shayesteh, M.R. A comparative study between single and multi-frame anomaly detection and localization in recorded video streams. *J. Vis. Commun. Image Represent.* **2021**, *79*, 103232. [CrossRef]
125. Asad, M.; Yang, J.; Tu, E.; Chen, L.; He, X. Anomaly3D: Video anomaly detection based on 3D-normality clusters. *J. Vis. Commun. Image Represent.* **2021**, *75*, 103047. [CrossRef]
126. 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]
127. Wang, J.; Zhang, J.; Ji, G.; Sheng, B. Criss-Cross Attention Based Auto Encoder for Video Anomaly Event Detection. *Intell. Autom. Soft Comput.* **2022**, *34*, 1629–1642. [CrossRef]

128. Kommanduri, R.; Ghorai, M. Bi-READ: Bi-Residual AutoEncoder based feature enhancement for video anomaly detection. *J. Vis. Commun. Image Represent.* **2023**, *95*, 103860. [CrossRef]
129. Taghinezhad, N.; Yazdi, M. A New Unsupervised Video Anomaly Detection Using Multi-Scale Feature Memorization and Multipath Temporal Information Prediction. *IEEE Access* **2023**, *11*, 9295–9310. [CrossRef]
130. Jeong, J.; Jung, H.; Choi, Y.; Park, S.; Kim, M. Intelligent Complementary Multi-Modal Fusion for Anomaly Surveillance and Security System. *Sensors* **2023**, *23*, 9214. [CrossRef] [PubMed]
131. Li, N.; Chang, F. Video anomaly detection and localization via multivariate gaussian fully convolution adversarial autoencoder. *Neurocomputing* **2019**, *369*, 92–105. [CrossRef]
132. Ganokratanaa, T.; Aramvith, S.; Sebe, N. Unsupervised Anomaly Detection and Localization Based on Deep Spatiotemporal Translation Network. *IEEE Access* **2020**, *8*, 50312–50329. [CrossRef]
133. Li, Y.; Cai, Y.; Liu, J.; Lang, S.; Zhang, X. Spatio-Temporal Unity Networking for Video Anomaly Detection. *IEEE Access* **2019**, *7*, 172425–172432. [CrossRef]
134. Chen, D.; Wang, P.; Yue, L.; Zhang, Y.; Jia, T. Anomaly detection in surveillance video based on bidirectional prediction. *Image Vis. Comput.* **2020**, *98*, 103915. [CrossRef]
135. Patil, P.W.; Dudhane, A.; Murala, S. End-to-End Recurrent Generative Adversarial Network for Traffic and Surveillance Applications. *IEEE Trans. Veh. Technol.* **2020**, *69*, 14550–14562. [CrossRef]
136. Liu, S.; Yang, E.; Fang, K. Self-Learning pLSA Model for Abnormal Behavior Detection in Crowded Scenes. *IEICE Trans. Inf. Syst.* **2021**, *E104.D*, 473–476. [CrossRef]
137. Wu, R.; Li, S.; Chen, C.; Hao, A. Improving video anomaly detection performance by mining useful data from unseen video frames. *Neurocomputing* **2021**, *462*, 523–533. [CrossRef]
138. Yang, Z.; Liu, J.; Wu, P. Bidirectional Retrospective Generation Adversarial Network for Anomaly Detection in Videos. *IEEE Access* **2021**, *9*, 107842–107857. [CrossRef]
139. Chen, D.; Yue, L.; Chang, X.; Xu, M.; Jia, T. NM-GAN: Noise-modulated generative adversarial network for video anomaly detection. *Pattern Recognit.* **2021**, *116*, 107969. [CrossRef]
140. Ganokratanaa, T.; Aramvith, S.; Sebe, N. Video anomaly detection using deep residual-spatiotemporal translation network. *Pattern Recognit. Lett.* **2022**, *155*, 143–150. [CrossRef]
141. Yu, J.; Lee, Y.; Yow, K.C.; Jeon, M.; Pedrycz, W. Abnormal Event Detection and Localization via Adversarial Event Prediction. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 3572–3586. [CrossRef]
142. Zhong, Y.; Chen, X.; Jiang, J.; Ren, F. A cascade reconstruction model with generalization ability evaluation for anomaly detection in videos. *Pattern Recognit.* **2022**, *122*, 108336. [CrossRef]
143. Aslam, N.; Rai, P.K.; Kolekar, M.H. A3N: Attention-based adversarial autoencoder network for detecting anomalies in video sequence. *J. Vis. Commun. Image Represent.* **2022**, *87*, 103598. [CrossRef]
144. Hao, Y.; Li, J.; Wang, N.; Wang, X.; Gao, X. Spatiotemporal consistency-enhanced network for video anomaly detection. *Pattern Recognit.* **2022**, *121*, 108232. [CrossRef]
145. Li, Q.; Yang, R.; Xiao, F.; Bhanu, B.; Zhang, F. Attention-based anomaly detection in multi-view surveillance videos. *Knowl. Based Syst.* **2022**, *252*, 109348. [CrossRef]
146. Zhao, L.; Wang, S.; Wang, S.; Ye, Y.; Ma, S.; Gao, W. Enhanced Surveillance Video Compression With Dual Reference Frames Generation. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 1592–1606. [CrossRef]
147. Huang, H.; Zhao, B.; Gao, F.; Chen, P.; Wang, J.; Hussain, A. A Novel Unsupervised Video Anomaly Detection Framework Based on Optical Flow Reconstruction and Erased Frame Prediction. *Sensors* **2023**, *23*, 4828. [CrossRef]
148. Li, G.; He, P.; Li, H.; Zhang, F. Adversarial composite prediction of normal video dynamics for anomaly detection. *Comput. Vis. Image Underst.* **2023**, *232*, 103686. [CrossRef]
149. Pedrycz, W.; Waletzky, J. Fuzzy clustering with partial supervision. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **1997**, *27*, 787–795. [CrossRef]
150. Sikdar, A.; Chowdhury, A.S. An adaptive training-less framework for anomaly detection in crowd scenes. *Neurocomputing* **2020**, *415*, 317–331. [CrossRef]
151. Singh, G.; Kapoor, R.; Khosla, A. Optical Flow-Based Weighted Magnitude and Direction Histograms for the Detection of Abnormal Visual Events Using Combined Classifier. *Int. J. Cogn. Inform. Nat. Intell.* **2021**, *15*, 12–30. [CrossRef]
152. Khaire, P.; Kumar, P. A semi-supervised deep learning based video anomaly detection framework using RGB-D for surveillance of real-world critical environments. *Forensic Sci. Int. Digit. Investig.* **2022**, *40*, 301346. [CrossRef]
153. Pramanik, A.; Sarkar, S.; Pal, S.K. Video surveillance-based fall detection system using object-level feature thresholding. *Knowl. Based Syst.* **2023**, *280*, 110992. [CrossRef]
154. Hu, X.; Dai, J.; Huang, Y.; Yang, H.; Zhang, L.; Chen, W.; Yang, G.; Zhang, D. A weakly supervised framework for abnormal behavior detection and localization in crowded scenes. *Neurocomputing* **2020**, *383*, 270–281. [CrossRef]
155. Degardin, B.; Proença, H. Iterative weak/self-supervised classification framework for abnormal events detection. *Pattern Recognit. Lett.* **2021**, *145*, 50–57. [CrossRef]
156. Ullah, W.; Hussain, T.; Khan, Z.A.; Haroon, U.; Baik, S.W. Intelligent dual stream CNN and echo state network for anomaly detection. *Knowl. Based Syst.* **2022**, *253*, 109456. [CrossRef]

157. Yi, S.; Fan, Z.; Wu, D. Batch feature standardization network with triplet loss for weakly-supervised video anomaly detection. *Image Vis. Comput.* **2022**, *120*, 104397. [CrossRef]
158. Liu, Y.; Liu, J.; Zhao, M.; Li, S.; Song, L. Collaborative Normality Learning Framework for Weakly Supervised Video Anomaly Detection. *IEEE Trans. Circuits Syst. II Express Briefs* **2022**, *69*, 2508–2512. [CrossRef]
159. Kamoona, A.M.; Gostar, A.K.; Bab-Hadiashar, A.; Hoseinnezhad, R. Multiple instance-based video anomaly detection using deep temporal encoding–decoding. *Expert. Syst. Appl.* **2023**, *214*, 119079. [CrossRef]
160. Thakare, K.V.; Sharma, N.; Dogra, D.P.; Choi, H.; Kim, I.-J. A multi-stream deep neural network with late fuzzy fusion for real-world anomaly detection. *Expert. Syst. Appl.* **2022**, *201*, 117030. [CrossRef]
161. Krishna, N.S.; Bhattu, S.N.; Somayajulu, D.V.L.N.; Kumar, N.V.N.; Reddy, K.J.S. GssMILP for anomaly classification in surveillance videos. *Expert. Syst. Appl.* **2022**, *203*, 117451. [CrossRef]
162. Ullah, W.; Hussain, T.; Ullah, F.U.M.; Lee, M.Y.; Baik, S.W. TransCNN: Hybrid CNN and transformer mechanism for surveillance anomaly detection. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106173. [CrossRef]
163. Shao, W.; Xiao, R.; Rajapaksha, P.; Wang, M.; Crespi, N.; Luo, Z.; Minerva, R. Video anomaly detection with NTCN-ML: A novel TCN for multi-instance learning. *Pattern Recognit.* **2023**, *143*, 109765. [CrossRef]
164. Chen, H.; Mei, X.; Ma, Z.; Wu, X.; Wei, Y. Spatial–temporal graph attention network for video anomaly detection. *Image Vis. Comput.* **2023**, *131*, 104629. [CrossRef]
165. Tang, J.; Wang, Z.; Hao, G.; Wang, K.; Zhang, Y.; Wang, N.; Liang, D. SAE-PPL: Self-guided attention encoder with prior knowledge-guided pseudo labels for weakly supervised video anomaly detection. *J. Vis. Commun. Image Represent.* **2023**, *97*, 103967. [CrossRef]
166. Zhang, B.; Xue, J. Weakly-supervised anomaly detection with a Sub-Max strategy. *Neurocomputing* **2023**, *560*, 126770. [CrossRef]
167. Wang, L.; Wang, X.; Liu, F.; Li, M.; Hao, X.; Zhao, N. Attention-guided MIL weakly supervised visual anomaly detection. *Measurement* **2023**, *209*, 112500. [CrossRef]
168. Ullah, W.; Ullah, F.U.M.; Khan, Z.A.; Baik, S.W. Sequential attention mechanism for weakly supervised video anomaly detection. *Expert. Syst. Appl.* **2023**, *230*, 120599. [CrossRef]
169. Lv, H.; Zhou, C.; Cui, Z.; Xu, C.; Li, Y.; Yang, J. Localizing Anomalies From Weakly-Labeled Videos. *IEEE Trans. Image Process.* **2021**, *30*, 4505–4515. [CrossRef] [PubMed]
170. Jebur, S.A.; Hussein, K.A.; Hoomod, H.K.; Alzubaidi, L.; Santamaría, J. Review on Deep Learning Approaches for Anomaly Event Detection in Video Surveillance. *Electronics* **2022**, *12*, 29. [CrossRef]
171. 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. [CrossRef]
172. Luo, W.; Liu, W.; Gao, S. A Revisit of Sparse Coding Based Anomaly Detection in Stacked RNN Framework. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 341–349. [CrossRef]
173. Sultani, W.; Chen, C.; Shah, M. Real-World Anomaly Detection in Surveillance Videos. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 6479–6488. [CrossRef]
174. Lu, C.; Shi, J.; Jia, J. Abnormal Event Detection at 150 FPS in MATLAB. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 2720–2727. [CrossRef]
175. Detection of Unusual Crowd Activity Dataset. n.d. Available online: https://mha.cs.umn.edu/proj_events.shtml#crowd (accessed on 14 June 2024).
176. Ferryman, J.; Shahrokni, A. PETS2009: Dataset and challenge. In Proceedings of the 2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Snowbird, UT, USA, 7–9 December 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1–6. [CrossRef]
177. 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]
178. Degardin, B.; Proenca, H. Human Activity Analysis: Iterative Weak/Self-Supervised Learning Frameworks for Detecting Abnormal Events. In Proceedings of the 2020 IEEE International Joint Conference on Biometrics (IJCB), Houston, USA, 28 September–1 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–7. [CrossRef]
179. Leyva, R.; Sanchez, V.; Li, C.-T. The LV dataset: A realistic surveillance video dataset for abnormal event detection. In Proceedings of the 2017 5th International Workshop on Biometrics and Forensics (IWBF), Coventry, UK, 4–5 April 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6. [CrossRef]
180. Akti, S.; Tataroglu, G.A.; Ekenel, H.K. Vision-based Fight Detection from Surveillance Cameras. In Proceedings of the 2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA), Istanbul, Turkey, 6–9 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6. [CrossRef]
181. Nievas, E.B.; Suarez, O.D.; García, G.B.; Sukthankar, R. Violence Detection in Video Using Computer Vision Techniques. In *Computer Analysis of Images and Patterns*; Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 332–339. [CrossRef]

182. Hassner, T.; Itcher, Y.; Kliper-Gross, O. Violent flows: Real-time detection of violent crowd behavior. In Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, 16–21 June 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1–6. [CrossRef]
183. Martínez-Villaseñor, L.; Ponce, H.; Brieva, J.; Moya-Albor, E.; Núñez-Martínez, J.; Peñafort-Asturiano, C. UP-Fall Detection Dataset: A Multimodal Approach. *Sensors* **2019**, *19*, 1988. [CrossRef]
184. Gu, C.; Sun, C.; Ross, D.A.; Vondrick, C.; Pantofaru, C.; Li, Y.; Vijayanarasimhan, S.; Toderici, G.; Ricco, S.; Sukthankar, R.; et al. AVA: A Video Dataset of Spatio-Temporally Localized Atomic Visual Actions. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 6047–6056. [CrossRef]
185. Auvinet, E.; Rougier, C.; Meunier, J.; St-Arnaud, A.; Rousseau, J. *Multiple Cameras Fall Dataset*; Tech. Rep. 1350; DIRO-Université de Montréal: Montréal, QC, Canada, 2010; p. 24.
186. Kwolek, B.; Kepski, M. Human fall detection on embedded platform using depth maps and wireless accelerometer. *Comput. Methods Progr. Biomed.* **2014**, *117*, 489–501. [CrossRef]
187. Everingham, M.; Van, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
188. Wang, L.; Shi, J.; Song, G.; Shen, I. Object Detection Combining Recognition and Segmentation. In *Computer Vision—ACCV 2007*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 189–199. [CrossRef]
189. Reddy, K.K.; Shah, M. Recognizing 50 human action categories of web videos. *Mach. Vis. Appl.* **2013**, *24*, 971–981. [CrossRef]
190. Soomro, K.; Zamir, A.R.; Shah, M. UCF101: A dataset of 101 human actions classes from videos in the wild. *arXiv* **2012**, arXiv:1212.0402. [CrossRef]
191. Krasin, I.; Duerig, T.; Alldrin, N.; Ferrari, V.; Abu-El-Haija, S.; Kuznetsova, A.; Rom, H.; Uijlings, J.; Popov, S.; Veit, A.; et al. OpenImages: A Public Dataset for Large-Scale Multi-Label And Multi-Class Image Classification. 2017. Dataset. Available online: https://github.com/openimages (accessed on 12 June 2024).
192. CMU Graphics Lab Motion Capture Database. n.d. Available online: http://mocap.cs.cmu.edu/ (accessed on 3 June 2024).
193. Ryoo, M.S.; Aggarwal, J.K.; Dataset, U.T.-I. ICPR contest on Semantic Description of Human Activities (SDHA). 2010. Available online: https://cvrc.ece.utexas.edu/SDHA2010/Human_Interaction.html (accessed on 3 June 2024).
194. Peliculas Movies Fight Detection Dataset. n.d. Available online: http://academictorrents.com/details/70e0794e2292fc051a13f05ea6f5b6c16f3d3635/tech&h%20it=1&filelist=1 (accessed on 12 June 2024).
195. 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. [CrossRef]
196. Kuehne, H.; Jhuang, H.; Garrote, E.; Poggio, T.; Serre, T. HMDB: A large video database for human motion recognition. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 2556–2563. [CrossRef]
197. Carreira, J.; Noland, E.; Banki-Horvath, A.; Hillier, C.; Zisserman, A. A short note about kinetics-600. *arXiv* **2018**, arXiv:1808.01340. [CrossRef]
198. Liu, J.; Luo, J.; Shah, M. Recognizing realistic actions from videos "in the wild". 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. 1996–2003. [CrossRef]
199. Cinelli, L.P.; Marins, M.A.; da Silva, E.A.B.; Netto, S.L. *Variational Methods for Machine Learning with Applications to Deep Networks*; Springer International Publishing: Cham, Switzerland, 2021. [CrossRef]
200. Oliveira, E.E.; Rodrigues, M.; Pereira, J.P.; Lopes, A.M.; Mestric, I.I.; Bjelogrlic, S. Unlabeled learning algorithms and operations: Overview and future trends in defense sector. *Artif. Intell. Rev.* **2024**, *57*, 66. [CrossRef]
201. Ribeiro, M.; Lazzaretti, A.E.; Lopes, H.S. A study of deep convolutional auto-encoders for anomaly detection in videos. *Pattern Recognit. Lett.* **2018**, *105*, 13–22. [CrossRef]
202. Masci, J.; Meier, U.; Cireşan, D.; Schmidhuber, J. Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. In *Artificial Neural Networks and Machine Learning—ICANN 2011*; Honkela, T., Duch, W., Girolami, M., Kaski, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 52–59. [CrossRef]
203. Jovanovic, M.; Campbell, M. Generative Artificial Intelligence: Trends and Prospects. *Computer* **2022**, *55*, 107–112. [CrossRef]
204. Simmler, N.; Sager, P.; Andermatt, P.; Chavarriaga, R.; Schilling, F.-P.; Rosenthal, M.; Stadelmann, T. A Survey of Un-, Weakly-, and Semi-Supervised Learning Methods for Noisy, Missing and Partial Labels in Industrial Vision Applications. In Proceedings of the 2021 8th Swiss Conference on Data Science (SDS), Lucerne, Switzerland, 9 June 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 26–31. [CrossRef]
205. Yu, J.; Kim, J.-G.; Gwak, J.; Lee, B.-G.; Jeon, M. Abnormal event detection using adversarial predictive coding for motion and appearance. *Inf. Sci.* **2022**, *586*, 59–73. [CrossRef]
206. Huang, C.; Wen, J.; Xu, Y.; Jiang, Q.; Yang, J.; Wang, Y.; Zhang, D. Self-Supervised Attentive Generative Adversarial Networks for Video Anomaly Detection. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 9389–9403. [CrossRef]
207. Antoine, V.; Guerrero, J.A.; Romero, G. Possibilistic fuzzy c-means with partial supervision. *Fuzzy Sets Syst.* **2022**, *449*, 162–186. [CrossRef]

208. Oliver, A.; Odena, A.; Raffel, C.; Cubuk, E.D.; Goodfellow, I.J. Realistic evaluation of deep semi-supervised learning algorithms. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; Curran Associates Inc.: Red Hook, NY, USA, 2018; pp. 3239–3250.
209. Tian, Z.; Wang, W.; Zhou, K.; Song, X.; Shen, Y.; Liu, S. Weighted Pseudo-Labels and Bounding Boxes for Semisupervised SAR Target Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 5193–5203. [CrossRef]
210. Park, S.; Kim, H.; Kim, M.; Kim, D.; Sohn, K. Normality Guided Multiple Instance Learning for Weakly Supervised Video Anomaly Detection. In Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2–7 January 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 2664–2673. [CrossRef]
211. Xu, Z.; Zeng, X.; Ji, G.; Sheng, B. Improved Anomaly Detection in Surveillance Videos with Multiple Probabilistic Models Inference. *Intell. Autom. Soft Comput.* **2022**, *31*, 1703–1717. [CrossRef]
212. Peyre, J.; Laptev, I.; Schmid, C.; Sivic, J. Weakly-Supervised Learning of Visual Relations. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 5189–5198. [CrossRef]
213. Vu, T.-H.; Boonaert, J.; Ambellouis, S.; Taleb-Ahmed, A. Multi-Channel Generative Framework and Supervised Learning for Anomaly Detection in Surveillance Videos. *Sensors* **2021**, *21*, 3179. [CrossRef] [PubMed]
214. Yu, B.X.B.; Chang, J.; Wang, H.; Liu, L.; Wang, S.; Wang, Z.; Lin, J.; Xie, L.; Li, H.; Lin, Z.; et al. Visual Tuning. *ACM Comput. Surv.* **2024**. [CrossRef]