PORT SAID ENGINEERING RESEARCH JOURNAL
Faculty of Engineering - Port Said University
Volume (28) No. 3 pp: 88:102

Abnormal Human Activity Recognition in Video Surveillance: A Survey

Iman M. Yossef [1,*], *Kareem H. El-Safty* [2], *Marwa Gamal* [3], *Rehab F. Abdel-Kader* [4], *Khaled Abd Elsalam Ali* [5]

[1] *Electrical Engineering Department, Faculty of Engineering, Suez Canal University, email: iman.mostafa@eng.suez.edu.eg*

[2] *Electrical Engineering Department, Faculty of Engineering, Suez Canal University, email:* kareemessafty.pgs@eng.suez.edu.eg

[3] *Electrical Engineering Department, Faculty of Engineering, Suez Canal University, email:* marwa_gamal@eng.suez.edu.eg

[4] *Electrical Engineering Department, Faculty of Engineering, Port Said University, email:* rehabfarouk@eng.psu.edu.eg

[5] *Electrical Engineering Department, Faculty of Engineering, Suez Canal University, email:* khaled.abdelsalam@eng.suez.edu.eg

**Corresponding author,* *DOI: 10.21608/PSERJ.2024.275800.1328*

Received 11-3-2024
Revised 23-3-2024
Accepted 19-5-2024

ABSTRACT

Human Activity Recognition (HAR) is considered a multidisciplinary field that different branches of science contribute to its advancements. Vision-Based HAR is one of the means to use Computer Vision (CV) and its techniques to study and analyze the behavior of humans within the context of videos. Recently, Video Anomaly detection (VAD) has gained vast attention and becomes a popular research topic in recent years. This is due to their enormous potential in many fields such as healthcare monitoring, surveillance/crowd analysis, sports, Ambient Assistive Living (AAL), event analysis, and security. Manually detecting and analyzing inappropriate behavior was a very challenging task, especially in real-time scenarios which led to a great demand for smart surveillance systems. In recent work, deep learning approaches have been dominated in this field as they outperform the performance of other traditional methods. This literature provides the latest algorithms for anomalous human activities, the challenges facing this field, and a comprehensive review of the State-Of-The-Art (SOTA) approaches including the feature extractor, the method, and the loss function. In addition, we propose the effect of applying swarm optimization algorithms in the anomaly detection field in recent years. Moreover, it presents a chronological background to the subject with an emphasis on the recent advancements in the VAD field.

Keywords: Video Anomaly Detection, Video Surveillance, Video Transformer Networks, Swarm Optimization.

1 INTRODUCTION

The Human Activity Recognition (HAR) has become a trending research direction in Computer Vision (CV) because of sensors and accelerometer availability, low power consumption, live data streaming, and advancements in computer vision, Machine Learning (ML), and the Internet of Things (IoT). HAR is frequently linked to the process of identifying and naming human activities in real life through sensory perceptions such as walking, sleeping, running, sitting, standing, showering, cooking, driving, opening the door, abnormal activities, etc. [1,2]. It can be utilized in visual surveillance systems [3,4] to detect potentially dangerous human actions, as well as autonomous navigation systems [5] to detect human behaviors and ensure safe operations. It is also crucial for a variety of other applications, including video retrieval [6], home monitoring, human-robot interaction [7], Human-Computer Interfaces (HCI) [8], healthcare by tracking elderly people sitting alone [9,10], smart cities [11] and sports [12,13].

The advances in CV techniques and hardware accelerators made it possible to process the huge amount of data produced by live-stream cameras [14,15]. As a natural consequence, Abnormal Human Action Recognition (AbHAR) has become an interesting field in

CV due to the numerous applications that directly benefit from it such as public security, monitoring workers' safety during working hours, healthcare systems for the elder people, and the need for intelligent video surveillance systems (IVSS). In the past few years, intelligent video surveillance systems (IVSS) played a vital role in the computer vision field because of the increasing demand for security and the growing number of surveillance cameras outdoors and indoors. IVSS is capable of automatically detecting anomalous actions such as crimes, fights, traffic accidents, riots, kidnapping, and catastrophic events, as well as anomalous entities such as weapons in critical locations and abandoned objects. However, surveillance video analysis faces several challenges, one of which is detecting anomalous events, which demands extensive human effort and is time-consuming. As a result, relying on the human factor alone is insufficient, and IVSS is developed to assist in such scenarios.

In the context of Vision-Based Anomaly Detection (AD), anomalous events are considered rare, and their lifetime is relatively short compared to the complete live stream of the surveillance system. This is one of the main challenges in the AbHAR systems. Hence, different approaches to tackle this problem and to offer a robust framework for AbHAR have been presented in the last decade. In terms of the available surveys on HAR, numerous studies have been conducted [1,15–20]. On the other hand, only fewer surveys related to Deep Learning (DL) based VAD, are proposed.

Nayak et al. [21] presented an analysis of the performance evaluation approaches in terms of datasets and various evaluation criteria. Dhiman et al. [22] provided feature designs in videos of abnormal human activity recognition concerning the context or application. Besides, they introduced the drawbacks of each feature technique for 2D and 3D AbHAR categories. Mabrouk et al. [23] studied various levels of an intelligent video surveillance system and discussed some limitations of the recognition of abnormal behavior. Rezaee et al. [24] identified several automatic and real-time monitoring approaches for abnormal event detection in security applications to recognize dynamic crowd dynamics. Suarez et al. [25] discussed the effect of DL in the anomaly detection field and the classification of different DL methods relating to their objectives. Caetano et al. [26] presented a review of a large number of STOA methods and datasets related to the VAD field and they discussed the application-oriented issues related to deep anomaly detection for in-vehicle monitoring.

Optimization methods have previously been widely used in many fields, including Machine Learning (ML), data science, engineering, and many others. These methods seek the best values for parameters, weights, or configurations that result in the best solution to a given problem. They are useful in decision-making and problem-solving processes because they automate the search for the best solution in complex scenarios where an exhaustive search is not possible. Swarm optimization algorithms, which are inspired by the collective behavior of natural swarms, such as bird flocks, fish schools, and insect colonies, are one of these methods. There are many types such as Particle Swarm Optimization (PSO) [27], Artificial Bee Colony (ABC) [28], Ant Colony Optimization (ACO) [29], Firefly Algorithm (FA) [30], Bat Algorithm (BA) [31] and many others.

To our knowledge, this survey is the first to introduce swarm optimization in a VAD survey. Moreover, an extensive overview of the recent weakly-supervised SOTA models related to AbHAR will be explored with their availability codes.

This survey is structured in five sections as follows: section 2 will explore a discussion of Abnormal Human Activity Recognition (AbHAR). Section 3 provides a brief review of swarm optimization algorithms and their applications in VAD. Section 4 will be dedicated to proposing the challenges that face the AbHAR domain. The recent SOTA frameworks proposed in the field of VAD are introduced in Section 5. Finally, the survey will be concluded with a clear point of view of the current status of the field and the possible future directions in the last section.

2 ABNORMAL HUMAN ACTIVITY RECOGNITION (ABHAR)

Despite the popularity of the HAR topic and its various applications in many fields, AbHAR has become one of the trendiest topics in recent research, especially in security issues using video surveillance systems. The area of research in HAR seems to be close to that of AbHAR but they are not the same. Deep anomaly detection methods must be used to create new surveillance and monitoring systems that do not rely only on human supervision, lowering the risk of the aforementioned drawbacks. With the increasing number of crimes and the essential need for security in public areas like malls and banks, the demand for automotive surveillance systems becomes crucial. In this section, we will introduce more information related to this domain such as the anomaly definition, the framework of VAD, the learning mechanisms of anomalies, the types of anomalies, the detection learning approaches, and the feature learning methods.

2.1 Anomaly Definition

AbHAR and VAD are terms used interchangeably and are defined as the odd or irregular patterns found in videos that do not conform to the normal trained patterns. According to [32], VAD systems are either manually built by experts setting thresholds on data or constructed automatically by learning from the available data through Machine Learning (ML). VAD is widely used in many applications such as fraud detection [33,34], image processing [35,36], sensor networks [37,38], medical health [39,40], intrusion detection [41], IT security [42–44], and social media [45,46]. Fig. 1 shows an illustrative example of a normal and abnormal frame in a sample of the UCF-Crime dataset [47], where (a) is a normal frame of a woman withdrawing money from an ATM and (b) is the abnormal frame that shows a man steal the woman which has recorded as an anomaly activity there in the red window.

(a) (b)

Figure 1: A sample from the UCF-Crime dataset (Robbery part) [47]**; a) is a normal frame from the video, while b) is an abnormal frame of a woman being stolen**.

2.2 General Framework of Anomaly Detection

There are some sequential steps to form a complete surveillance system for VAD. As shown in Fig. 2, the video data are firstly captured or recorded by a surveillance camera, then segmented into several frames to determine any significant changes that occur in the content. After that, some pre-processing steps are performed according to our needs such as noise removal, resizing the frames, illumination adjustment, and others. The third step involves the extracting of features either by traditional or deep-based methods – which will be explained later in subsection 2.4 -. The next step is developing the model either for classification to determine if the presented video is normal or not or for detecting the anomaly type in the video such as fighting, a car moving in a wrong direction, robbery, etc. Lastly, depending on the model used, a score is generated to detect if it is a normal or abnormal video.

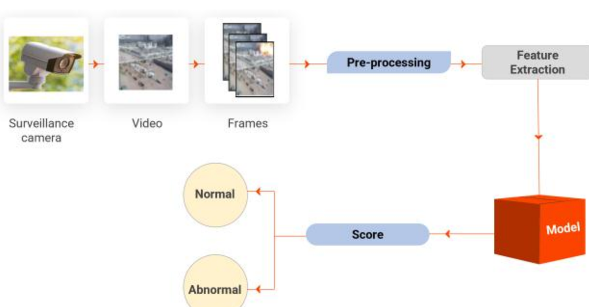

Figure 2: General framework of VAD. After collecting the data and performing some pre-processing on it, features can be extracted by different methods. Depending on the project objective, the model is chosen which generates a score to classify the data as normal or abnormal.

2.3 Anomaly Detection Learning Approaches

Based on the availability of annotated data during the training process, AD techniques can be categorized into three classes supervised, unsupervised, and semi-supervised learning. In the Supervised learning scheme, the normal and the abnormal data are associated with labels, which means all anomalies are known before [48,49]. Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Decision Trees (DT) are commonly used algorithms for this type. However, it is not suitable in most situations to label anomalies in videos due to many aspects.

In the unsupervised anomaly learning approach [50–57], the model learns the pattern of the normal data then it's used to predict if the new data point is an anomaly or not. The main aim is to detect previously unseen rare events without any previous information about these, which means the dataset points are not labeled as normal or abnormal. That is why unsupervised learning is more popular than supervised in the AD area. The main obstacle here is defining normal behavior in its context. Some of the popular algorithms used are K-means, Local Outlier Factor (LOF), and Auto-Encoders (AE).

Semi-supervised learning is widely used in AD as it grosses the benefits of both supervised and unsupervised learning methods [47,58–64]. In this type of learning, only the normal activity class is labeled while the abnormal class is not annotated with any labels. AE is one of the most popular approaches regarding the Semi-Supervised technique for AD.

2.4 Feature-Learning Based Methods

Feature extraction is one of the major primary steps in any CV pipeline such as image classification [65], VAD [66], and many others. Generally, the most frequent method for detecting visual anomalies is to extract features and model learning. Optimum choice of features plays a key role in detecting specific anomalies. Features can be extracted in AbHAR using two techniques: handcrafted (non-automatic) or deep-based (automatic feature extraction) approaches. Some surveys have discussed the difference between these two approaches [67,68].

2.4.1 Handcrafted Features

The handcrafted features method [69,70] is based on extracting low-level features (motion or texture) from the input data. Extracting features using well-defined feature descriptors - such as Scale Invariant Feature Transform (SIFT) [71], Speeded Up Robust Features (SURF) [72], and Binary Robust Independent Elementary Features (BRIEF) [73]- requires high expertise in the problem domain. Thus, different trials are needed to select and fine-tune the best technique. Moreover, feature selection and preprocessing are necessary to prepare the features before the modeling step. the main advantage of this approach is that it produces an understandable set of features that can be visualized. However, this method faces some challenges such as the time-consuming in extraction of features and the need for an expert. The main obstacle is that the dataset will probably contain occlusion issues and high complexity in crowded areas.

Generally, handcrafted feature approaches are classified into two main categories: object-oriented where detection of anomalies is by extracting objects or trajectories, and non-object-oriented where pixels or optical flow features are used in the AD. The most common methods used are trajectory-based approaches [74,75], spatial-temporal approaches, Silhouette-based approaches [76], appearance-based approaches [77], optical flow-based descriptors [78], Histograms of Oriented Gradients (HOG) [79], and Histogram of Oriented flow (HOF) [80].

2.4.2 Deep-Based Features

Deep-based feature extraction methods depend mainly on DL architectures. Within the context of transfer learning fine-tuning, a pre-trained model such as VGG16 or ResNet-50 can be used to extract the features and forward them to the modeling block directly. Adopting this technique can generally save time and reduce the overhead cost and complexity of the training pipeline compared to the aforementioned method, as it skips the manual feature selection. In addition, it can be used to train the neural network from scratch. This will allow the network to capture and learn complex representations and patterns. The quality of these patterns depends on several factors such as the network structure, the optimizer, and different hyperparameters such as the learning rate. On the other hand, it is not trivial to visualize and understand the learned patterns.

The most popular methods in DL are Convolution Neural Networks (CNN) [81,82], Recurrent Neural Networks (RNN), Long Short-term Memory (LSTM) [83], Auto-Encoders (AE), Variational Auto-Encoders (VAE) [84,85], Deep One-Class Classification [86,87], Generative Adversarial Networks (GAN) [88,89] and Transformers [90,91]. Some popular feature extractors are used with the previously mentioned methods to extract spatial features, temporal features, or spatial-temporal features in videos such as 3D Convolutional Network (C3D) [92], Inflated Convolutional Network (I3D) [93], Temporal 3D ConvNets (T3D) [94], Temporal Segment Networks (TSN) [95] and Action Vector of Locally Aggregated Descriptors (Action VLAD) [96].

Generally, DL methods are dominant now in the CV field and are superior to the traditional methods, but traditional methods are still effective in solving some problems. Traditional algorithms are well-known, transparent, and designed for performance and power economy, but Deep-Based approaches provide better accuracy and variety at the expense of a lot of computational power. In different contexts, it might be useful to combine the Handcrafted and the Deep-Based features. Hybrid approaches of both methods are performed such as in healthcare [97,98], image classification [99], and video analysis [100].

3 SWARM OPTIMIZATION IN ANOMALY DETECTION

Swarm optimization is a type of optimization algorithm that is inspired by the collective behavior of social organisms, specifically swarms in nature. These algorithms are intended to solve complex optimization problems by simulating natural system behaviors and interactions such as bird flocks, fish schools, ant colonies, and bee hives. In the 1980s, the concept of swarm intelligence was first proposed. Since then, it has piqued the interest of scientists in a wide range of disciplines, including engineering, economics, computer science, artificial intelligence, and many others. In recent years, swarm optimization algorithms have received a great deal of attention as modern optimization methods that achieved remarkable results in many fields, as traditional optimization methods rely on parameter selection and require the objective function to have high mathematical performance.

The fundamental concept behind swarm optimization is based on the emergence of intelligent global behaviour from the interactions and cooperation of simple individual agents, also known as "particles," "agents," or "individuals." These agents communicate, share information, and adjust their behaviour in response to local and global information, to collectively optimise a given objective function. There are many algorithms: one of the most popular is PSO algorithm. In PSO, a population of particles moves through the search space, adjusting their positions based on their own experience (personal best) and the overall swarm's best experience (global best). This constant movement and updating leads the swarm to optimal solutions.

Few surveys worked on swarm optimization in the AD field. Mishra et al. [101] reviewed different swarm-based anomaly detection methods in the cyber-security field. Iftikhar et al. [102] provided a survey of swarm applications in network security. In addition, a review was published on applying swarm in intrusion detection systems on various domains by Nasir et al. [103]. Unfortunately, there are no surveys done in VAD specifically may be due to the low number of papers in this field.

Swarm was first introduced in the VAD field by Vagia et al. [104] who combined swarm intelligence and histograms of oriented gradients (HOG) descriptor to form a new feature capable of determining normal regions using the SVM [105] framework. Some surveys introduced the methods of swarm optimization algorithms. Wei et al. [106] discussed seven of the recent optimization algorithms that have been introduced since 2010 such as Fireworks algorithm [107], Pigeons Algorithm (PA) [108], Dragonfly Algorithm (DA) [109], Moth-flame Optimization Algorithm (MFOA) [110], Butterfly Optimization Algorithm (BOA) [111], and crow search algorithm (CSA) [112] and Whale Optimization Algorithm (WOA) [113]. Rezvanian et al. [114] provided an overview of ACO which is one of the popular swarm algorithms that simulates the behavior of ants in searching for food.

4 CHALLENGES IN ANOMALY DETECTION

Real-world anomalous events are complex and varied, so many obstacles still face the VAD field. It is difficult to make a comprehensive list of all possible anomalous events. Thus, in this section, we will present some of the challenges that face the VAD field. It is difficult to define abnormal moments because there is no clear distinction between normal and abnormal events, which leads to more false alarms. In addition, anomalies in videos are irregular, and rare, and can be localized or distributed spatiotemporally in complex scenarios. Furthermore, under realistic circumstances, the same behavior could be normal or abnormal depending on the environment. For example: running in the middle of the road is unusual, whereas running in a park is not. As previously stated, anomalies are uncommon data instances, as opposed to normal instances, which frequently account for a significant portion of the data leading to an imbalance of the data. As a result, collecting a large amount of labeled abnormal instances is difficult. Moreover, noise is considered an abnormality, so it is a big challenge to distinguish between it and the real abnormal events in the videos. As a result, it will affect the actual accuracy of the model. In addition, real-time anomaly detection is limited by high computational and infrastructure costs. One of the main challenges is the availability of high-configuration hardware to deal with long and high-quality videos and to keep up with the latest deep-learning models. Also, there is still a scarcity of large-scale wide-ranging anomaly data for training and validation. Moreover, annotating large data is highly costly. Hence, there is a need for good benchmarks to evaluate the algorithms used for VAD and localization. Other Environmental issues affect the efficiency of algorithms such as low resolution, variations in background, environmental fluctuations, and occlusions, scaling of the moving target, light intensity changes, and the excessive cost of collecting data.

5 RECENT STATE-OF-THE-ARTS (SOTA)

In recent years, several papers used deep learning-based models to tackle the problem of VAD. They outperform performance by leveraging deep neural networks' powerful representation capabilities. Models related to AbHAR datasets tend to combine a feature extractor and a classification block incorporated with a custom loss function to mitigate the effect of having only video-level labeling. Experiments proved that weakly supervised methods are achieving the best performances for VAD as depicted in Table 1.

Sultani et al. [47] introduced the problem of VAD in the context of Multiple Instance Learning (MIL) [115] using only weakly supervised labels. They used a 3D Convolution Network (C3D) to extract the features from the dataset. In their approach, normal and anomalies videos are considered as bags designed for a network that processes video clips independently from each other with a novel hinge loss function. Bags that have at least one abnormal snippet is considered positive bag while the bag that has only normal snippets is a negative bag. The bag-level labels are used to learn the instance-level anomaly scores. Moreover, in their paper, they introduced a large-scale dataset called UCF-Crime for training and testing weakly supervised anomaly detection methods.

Several papers followed [47] by using the same framework but advocated some improvements. Zhong et al. [62] introduced a novel way to use Graph Convolutional Networks as noisy label cleaners along with an action classifier. They signified that during training MIL methods suffered from error propagation. This approach managed to overcome the issue of having video-level labeling in the UCF-Crime dataset and converted the problem into a direct classification task based on a cross-entropy function and a temporal-ensembling strategy. Although this method gives better performance, it is computationally expensive. Zhang et al. [63] adopted the approach in [47] as their baseline and introduced a new inner bag loss (IBL) to reduce the gap between the lowest and highest scores in the negative bag while increasing it in the positive one. They replaced the first fully connected layer (FCN) of [47] with a temporal convolution network (TCN) [116] to connect between the preceding and the current information of the instance followed by two fully connected layers.

Zhu et al. [117] modified the model [47] by adding an attention block [90] and making use of the PWC-Net [118] to extract the motion-aware features. Morais et al. [56] detected the human anomalies using the Spatio-Temporal patterns of skeleton features. However, the algorithm depends on the quality of the skeleton tracking and detection so it cannot be applied to low-quality videos. [119] improved the approach in Sultani et al. [47] by fusing both the weak and self-supervised schemes and adding a new term to the loss function to enhance the performance. Moreover, they used a Random Forest (RF) model to combine the outputs at the score level of the best top 3 performance models and developed a new dataset named UBI-Flight.

Lu et al. [55] addressed the problem of few-shot learning [120] in anomaly detection of large videos by using a meta-learning-based mechanism. They used GANs to spot the anomalies in a previously unseen scene with only a few frames instead of collecting a huge amount of data for each scene. Doshi et al [121] presented an online algorithm named Multi-Objective Neural Anomaly Detector (MONAD) to detect anomalies in streaming videos with minimal detection delays. This algorithm involves two modules: a deep learning-based feature extraction module and a statistical decision-making module. The feature extraction module is a combination between GAN-based frame prediction and YOLO object detector [122] to extract the important features, while the other module is a nonparametric statistical algorithm that uses the extracted features for online anomaly detection.

Wu et al. [123] introduced a novel large violence dataset called XD-Violence which contains both videos and audio signals. In addition, they proposed a neural network with three parallel branches which are the holistic branch, localized branch, and score branch to capture the different relations between video snippets. Moreover, online detection was also performed. Ullah et al. [124] reduced the time complexity using a pre-trained ResNet-50 [125] to extract the features followed by a multi-layer Bi-directional Long Short-term Memory (BDLSTM) model to classify the normal or the abnormal events in surveillance scenes.

Instead of processing video clips independently from each other as in [47], Kamoona et al. [60] treated the video instances (clips) as sequential visual data and they also introduced a new loss function that maximizes the mean distance between the normal and the abnormal instance predictions. This loss function is smoother than the one of [47]. Tian et al. [59] introduced a novel method named Robust Temporal Feature Magnitude learning (RTFM). RTFM learns a temporal feature magnitude mapping function that recognizes rare abnormal snippets from abnormal videos with many normal snippets while maintaining a wider margin between normal and abnormal snippets.

Some papers presented Transformers for anomaly instances in videos. Yuan et al. [91] combined the U-Net [126] and the Video Vision Transformer (ViViT) [127] to capture wider global contexts and deeper temporal information. They named their model TransAnomaly, which is a prediction-based VAD method. In addition, the model can execute anomaly localization. Feng et al. [128] proposed a model based on Convolution Transformer (CT) with dual discriminator GAN (D2GAN) and developed a new self-attention module that is focused on spatio-temporal modeling in video sequences. The CT is capable of encoding temporal information efficiently in a sequence of feature maps and the D2GAN was developed to enhance the prediction of future frames using the Wasserstein GAN with gradient penalty (WGAN-GP) [129]. Li et al. [130] proposed another method using a Transformer-based Multi-Sequence Learning (MSL) network to address the shortage in the other MIL-based methods. The extracted snippets features were encoded using a multilayer Convolution Transformer-Encoder. Rather than selecting the instance with the highest score, the method selects the sequence with the highest sum of anomaly scores to reduce the probability of incorrect selection. VideoSwin [131] is used as a feature extractor in this method gives a better performance than C3D and I3D traditional extractors.

Chen et al. [132] introduced a Magnitude-Contrastive Glance-and-Focus Network (MGFN) for anomaly detection to address the issue in [59] as it pushes the magnitude of abnormal features to be larger and the normal ones to be smaller without considering other video attributes. Unlike previous methods, it first scans the entire video sequence for long-term context information, and then addresses each specific portion for anomaly detection. In addition, they developed the Feature Amplification Mechanism (FAM) to improve feature learning and a Magnitude Contrastive (MC) Loss to encourage the separability of normal and abnormal features. The model is composed of two blocks: Glance block and Focus block respectively. In the Glance block, a video clip-level transformer (VCT) is used for global correlation learning among clips followed by 2 fully connected Feed-Forward Networks (FFN). The Focus block includes a self-attentional convolution (SAC) to improve the learning of features, followed also by FFN.

All previous methods concentrated on extracting anomaly data representations without taking the effect of normal data into their consideration. Zhou et al. [133] introduced an Uncertainty Regulated Dual Memory Units (UR-DMU) model to learn both the representation of normal and abnormal data. They designed a Global and Local Multi-Head Self Attention (GL-MHSA) model for learning the features, afterwards two memory banks for normal and abnormal data are used to differentiate between the normal and abnormal patterns. The model ends with Normal data Uncertainty Learning (NUL) for normality latent embedding learning using Gaussian distribution.

In Table 2, recent papers on applying swarm optimization algorithms in VAD are proposed. Qasim et al. [134] used a modified ACO clustering algorithm to identify prominent regions in video frames with high optical flow variations for abnormal event detection in

Table 1. Recent SOTA approaches applied in AbHAR with Area Under the Curve (AUC). The different colors in the table indicate different techniques with their results.

Paper	Year	Feature-extraction	Method	Loss function	Datasets	AUC (%)	Code
Sultani et al. [47]	2018	C3D- RGB	3 Fully Connected Layers	MIL+ hinge loss function	UCF- crime	75.41	Code
Zhong et al. [62]	2019	C3D TSN-RGB TSN-optical flow	GCN	Cross entropy + temporal-ensembling strategy [140]	UCF-Crime	81.08 82.12 78.08	Code
					UCSD Ped2 [141,142]	93.2 ± 2.3 (Greyscale) 92.8 ± 1.6	
					ShanghaiTech [143]	76.44 84.44 84.13	
Zhang et al. [63]	2019	C3D-RGB	TCN + 2 FCN	Inner bag loss (IBL)	UCF-Crime	78.66	-
Zhu et al. [117]	2019	PWC-Net Optical flow	Attention model	MIL	UCF-Crime	79.0	-
Morais et al. [56]	2019	Alpha Pose [57] + optical flow	Recurrent encoder-decoder	Perceptual loss + Global loss + Local loss	ShanghaiTech	73.4	Code
					CUHK Avenue [144]	86.3	
Degardin et al. [119]	2020	C3D- RGB	3 FC + 2 Bayesian classifiers	MIL + cross-entropy	UCF-Crime	74.4	Code
					UBI-Flight [119]	81.9	
					UCSD	80.9	
Lu et al. [55]	2020	RGB	U-Net+ConvLSTM +GAN	Meta-Learning for different tasks	ShanghaiTech	77.9	Code
					CUHK Avenue	85.8	
					UCSD Ped1	86.3	
					UCSD Ped2	96.2	
Doshi et al. [121]	2020	RGB	GAN + YOLOv3	Intensity+ gradient difference + optical flow + adversarial training	CUHK Avenue	86.4	Code
					UCSD Ped 2	97.2	
					ShanghaiTech	70.9	
Wu et al. [123]	2020	C3D-I3D + (RGB optical flow) VGGis [145]	Holistic Branch Network	Binary Cross-Entropy + Distillation loss within a MIL scheme	XD-Violence [123]	67.19 78.64	Code
Kamoona et al. [60]	2020	C3D-RGB	Temporal encoding-decoding network	MIL + mean between normal and abnormal instances score	ShanghaiTech	87.42	Code
					UCF-Crime	79.49	
Ullah et al. [124]	2021	Pre-trained ResNet-50	BD-LSTM	Cross-entropy	UCF-Crime	85.53	-
					UCFCrime2Local [49]	89.05	
Tian et al. [59]	2021	C3D-RGB I3D-RGB	Dilated convolutions + self-attention	MIL	ShanghaiTech	91.51 97.21	Code
					XD-Violence	75.89 77.81	
					UCF-Crime	83.28 84.03	
Yuan et al. [91]	2021	RGB	Transformer +GAN	Intensity + loss Gradient loss + Difference loss	UCSD Ped1	86.7	-
					UCSD Ped2	96.4	
					Avenue	87.0	
					UCSD Ped2	93.2 ± 2.3 (Greyscale) 92.8 ± 1.6	

					ShanghaiTech	76.44 84.44 84.13	
Feng et al. [128]	2021	Optical flow	CT +D2GAN	WGAN-GP	ShanghaiTech	77.7	-
					UCSD Ped2	97.2	
					Avenue	85.9	
Li et al. [130]	2022	I3D VideoSwin	Transformer + 1D convolution	MIL with sequences + Binary Cross Entropy	ShanghaiTech	96.08 97.32	-
					UCF-Crime	85.30 85.62	
Chen et al. [132]	2022	I3D VideoSwin	VCT+SAC+FNN	MC + Binary Cross Entropy	UCF-Crime	86.98 86.67	Code
					XD-Violence	79.19 80.11	
Zhou et al. [133]	2023	I3D	GL-MHSA+DMU	4 Binary Cross Entropy	UCF-Crime	86.97	Code
					XD-Violence	94.02	

Table 2. Recent Swarm Optimization methods applied in AbHAR with Area Under the Curve (AUC) or Accuracy (AC)

Paper	Year	Method	Swarm Optmization Algorithm	Datasets	AUC/AC (%)
Qasim et al.[134]	2019	SVM	ACO	UMN [146]	99.77 (AUC)
				UCF web	98.54 (AUC)
Priyadharsini et al [135]	2022	CNN+SVM	PSO	UCSD	97 (AC)
Alsolai et al. [136]	2023	EfficientNet	ICSO	UCSDPed1	87.87 (AUC)
				UCSDPed2	88.90 (AUC)
Kumar et al. [139]	2023	CNN	PSO	ADOC [147]	86 (AC)

crowded environments in surveillance videos. Priyadharsini et al. [135] built a hybrid DL system based on a pre-trained CNN and a One-class SVM where improved PSO is used to isolate the most salient regions in the video frames. Alsolai et al. [136] proposed a vision-based anomaly system based on the EfficientNet [137] with Improved Chicken Swarm Optimizer (ICSO) [138] to detect and classify anomalies to assist visually impaired people. Kumar et al. [139] Applied Multi-Feature Tensor Subspace Learning and Robust Principal Component Analysis for feature extraction while PSO-based CNN for anomaly detection.

6 CONCLUSIONS

The Vision-Based AbHAR is considered a challenging task despite the recent advancements. The lack of a generic dataset that contains numerous different scenarios, a general framework that can adapt to multiple environments, and dedicated edge devices that can handle

and scale with the intensive computations, is considered the main reasons behind its difficulties. Nevertheless, VAD is garnering a lot of attention because of its vital role in ensuring security and safety by detecting anomalous events like traffic accidents and crimes. This survey provides an in-depth look at the recently proposed models in terms of accuracy, datasets, and loss functions. One of the notable issues regarding AbHAR is the scarce number of frameworks that address real-time applications. Furthermore, novel datasets with varied forms of anomalies should be developed to cover all possible scenarios. In addition to that, the newly developed models should be able to adapt and generate new scenes to be robust enough if the dataset contains little to no anomalies at all. Moreover, swarm optimization algorithms can be used with MIL methods to save time by reaching optimal solutions faster which is very crucial in VAD field. Finally, End-To-End pipeline optimization with quantization techniques may be a powerful approach to combine the feature extraction and classification phases in one cycle. Experimentally, this can reduce the training pipeline complexity and enable us to efficiently deploy massive models onto edge devices.

REFERENCES

[1] Bhardwaj R, Singh PK. Analytical review on human activity recognition in video. 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence), IEEE; 2016, p. 531–6. https://doi.org/10.1109/CONFLUENCE.2016.7508177.

[2] Kong Y, Fu Y. Human Action Recognition and Prediction: A Survey 2018.

[3] Weiyao Lin, Ming-Ting Sun, Poovandran R, Zhengyou Zhang. Human activity recognition for video surveillance. 2008 IEEE International Symposium on Circuits and Systems, IEEE; 2008, p. 2737–40. https://doi.org/10.1109/ISCAS.2008.4542023.

[4] Vishwakarma S, Agrawal A. A survey on activity recognition and behavior understanding in video surveillance. Vis Comput 2013;29:983–1009. https://doi.org/10.1007/s00371-012-0752-6.

[5] Lu M, Hu Y, Lu X. Driver action recognition using deformable and dilated faster R-CNN with optimized region proposals. Applied Intelligence 2020;50:1100–11. https://doi.org/10.1007/s10489-019-01603-4.

[6] Aslam F, Hussain F, Yousaf MH. Human Activity Based Video Retrieval Using Optical Flow and Local Binary Patterns. NED University Journal of Research 2018;15:93–105.

[7] Rodomagoulakis I, Kardaris N, Pitsikalis V, Mavroudi E, Katsamanis A, Tsiami A, et al. Multimodal human action recognition in assistive human-robot interaction. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE; 2016, p. 2702–6. https://doi.org/10.1109/ICASSP.2016.7472168.

[8] Haria A, Subramanian A, Asokkumar N, Poddar S, Nayak JS. Hand Gesture Recognition for Human Computer Interaction. Procedia Comput Sci 2017;115:367–74. https://doi.org/https://doi.org/10.1016/j.procs.2017.09.092.

[9] Schrader L, Vargas Toro A, Konietzny S, Rüping S, Schäpers B, Steinböck M, et al. Advanced Sensing and Human Activity Recognition in Early Intervention and Rehabilitation of Elderly People. J Popul Ageing 2020;13:139–65. https://doi.org/10.1007/s12062-020-09260-z.

[10] Wang Z, Yang Z, Dong T. A Review of Wearable Technologies for Elderly Care that Can Accurately Track Indoor Position, Recognize Physical Activities and Monitor Vital Signs in Real Time. Sensors 2017;17:341. https://doi.org/10.3390/s17020341.

[11] Almeida A, Azkune G. Activity recognition approaches for smart cities: The City4Age use case. 2017 IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI), IEEE; 2017, p. 1–5. https://doi.org/10.1109/RTSI.2017.8065909.

[12] Stein M, Janetzko H, Lamprecht A, Seebacher D, Schreck T, Keim D, et al. From game events to team tactics: Visual analysis of dangerous situations in multi-match data. 2016 1st International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW), IEEE; 2016, p. 1–9. https://doi.org/10.1109/TISHW.2016.7847777.

[13] Schuldhaus D. Human Activity Recognition in Daily Life and Sports Using Inertial Sensors. doctoralthesis. FAU University Press, 2019. https://doi.org/10.25593/978-3-96147-226-0.

[14] Li S, Xu L Da, Zhao S. The internet of things: a survey. Information Systems Frontiers 2015;17:243–59. https://doi.org/10.1007/s10796-014-9492-7.

[15] Jobanputra C, Bavishi J, Doshi N. Human Activity Recognition: A Survey. Procedia Comput Sci 2019;155:698–703. https://doi.org/10.1016/j.procs.2019.08.100.

[16] Beddiar DR, Nini B, Sabokrou M, Hadid A. Vision-based human activity recognition: a survey. Multimed Tools Appl 2020;79:30509–55. https://doi.org/10.1007/s11042-020-09004-3.

[17] Vrigkas M, Nikou C, Kakadiaris IA. A Review of Human Activity Recognition Methods. Front Robot AI 2015;2:28. https://doi.org/10.3389/frobt.2015.00028.

[18] Subetha T, Chitrakala S. A survey on human activity recognition from videos. 2016 International Conference on Information Communication and Embedded Systems (ICICES), IEEE; 2016, p. 1–7. https://doi.org/10.1109/ICICES.2016.7518920.

[19] Singh T, Vishwakarma DK. Human Activity Recognition in Video Benchmarks: A Survey. Lecture Notes in Electrical Engineering, vol. 526, Springer; 2019, p. 247–59. https://doi.org/10.1007/978-981-13-2553-3_24.

[20] Raval RM, Prajapati HB, Dabhi VK. Survey and analysis of human activity recognition in surveillance videos. Intelligent Decision Technologies 2019;13:271–94. https://doi.org/10.3233/IDT-170035.

[21] Nayak R, Pati UC, Das SK. A comprehensive review on deep learning-based methods for video anomaly detection. Image Vis Comput 2021;106:104078. https://doi.org/10.1016/j.imavis.2020.104078.

[22] Dhiman C, Vishwakarma DK. A review of state-of-the-art techniques for abnormal human activity recognition. Eng Appl Artif Intell

2019;77:21–45. https://doi.org/10.1016/j.engappai.2018.08.014.
[23] Ben Mabrouk A, Zagrouba E. Abnormal behavior recognition for intelligent video surveillance systems: A review. Expert Syst Appl 2018;91:480–91. https://doi.org/10.1016/j.eswa.2017.09.029.
[24] Rezaee K, Rezakhani SM, Khosravi MR, Moghimi MK. A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance. Pers Ubiquitous Comput 2021. https://doi.org/10.1007/s00779-021-01586-5.
[25] Suarez JJP, Naval PC. A SURVEY ON DEEP LEARNING TECHNIQUES. Strad Research 2020;7. https://doi.org/10.37896/sr7.8/037.
[26] Caetano F, Carvalho P, Cardoso J. Deep Anomaly Detection for In-Vehicle Monitoring—An Application-Oriented Review. Applied Sciences (Switzerland) 2022;12. https://doi.org/10.3390/app121910011.
[27] Kennedy J, Eberhart R. Particle swarm optimization. Proceedings of ICNN'95 - International Conference on Neural Networks, vol. 4, 1995, p. 1942–8 vol.4. https://doi.org/10.1109/ICNN.1995.488968.
[28] Karaboga D. An idea based on honey bee swarm for numerical optimization. Technical report-tr06, Erciyes university, engineering faculty, computer …; 2005.
[29] Dorigo M, Birattari M, Stutzle T. Ant colony optimization. IEEE Comput Intell Mag 2006;1:28–39.
[30] Yang X-S. Nature-inspired metaheuristic algorithms. Luniver press; 2010.
[31] Yang X-S. A new metaheuristic bat-inspired algorithm. Nature inspired cooperative strategies for optimization (NICSO 2010), Springer; 2010, p. 65–74.
[32] Steenwinckel B. Adaptive Anomaly Detection and Root Cause Analysis by Fusing Semantics and Machine Learning, 2018, p. 272–82. https://doi.org/10.1007/978-3-319-98192-5_46.
[33] Awoyemi JO, Adetunmbi AO, Oluwadare SA. Credit card fraud detection using machine learning techniques: A comparative analysis. 2017 International Conference on Computing Networking and Informatics (ICCNI), IEEE; 2017, p. 1–9. https://doi.org/10.1109/ICCNI.2017.8123782.
[34] Pumsirirat A, Yan L. Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine. International Journal of Advanced Computer Science and Applications 2018;9:18–25. https://doi.org/10.14569/IJACSA.2018.090103.
[35] Minhas MS, Zelek J. Anomaly Detection in Images. ArXiv 2019.
[36] Beggel L, Pfeiffer M, Bischl B. Robust Anomaly Detection in Images Using Adversarial Autoencoders, 2020, p. 206–22. https://doi.org/10.1007/978-3-030-46150-8_13.
[37] Rajasegarar S, Leckie C, Bezdek JC, Palaniswami M. Centered Hyperspherical and Hyperellipsoidal One-Class Support Vector Machines for Anomaly Detection in Sensor Networks. IEEE Transactions on Information Forensics and Security 2010;5:518–33. https://doi.org/10.1109/TIFS.2010.2051543.
[38] Cauteruccio F, Fortino G, Guerrieri A, Liotta A, Mocanu D, Perra C, et al. Short-Long Term Anomaly Detection in Wireless Sensor Networks based on Machine Learning and Multi-Parameterized Edit Distance. Information Fusion 2018;52. https://doi.org/10.1016/j.inffus.2018.11.010.
[39] Wei Q, Shi B, Lo JY, Carin L, Ren Y, Hou R. Anomaly detection for medical images based on a one-class classification. In: Mori K, Petrick N, editors. Medical Imaging 2018: Computer-Aided Diagnosis, vol. 10575, SPIE; 2018, p. 57. https://doi.org/10.1117/12.2293408.
[40] Fernando T, Denman S, Ahmedt-Aristizabal D, Sridharan S, Laurens KR, Johnston P, et al. Neural memory plasticity for medical anomaly detection. Neural Networks 2020;127:67–81. https://doi.org/10.1016/j.neunet.2020.04.011.
[41] Jose S, Malathi D, Reddy B, Jayaseeli D. A Survey on Anomaly Based Host Intrusion Detection System. J Phys Conf Ser 2018;1000:012049. https://doi.org/10.1088/1742-6596/1000/1/012049.
[42] Bezerra VH, da Costa VGT, Barbon Junior S, Miani RS, Zarpelão BB. IoTDS: A One-Class Classification Approach to Detect Botnets in Internet of Things Devices. Sensors (Basel) 2019;19:3188. https://doi.org/10.3390/s19143188.
[43] Zalasiński M, Łapa K, Laskowska M. Intelligent Approach to the Prediction of Changes in Biometric Attributes. IEEE Transactions on Fuzzy Systems 2020;28:1073–83. https://doi.org/10.1109/TFUZZ.2019.2955043.
[44] Demertzis K, Iliadis L, Bougoudis I. Gryphon: a semi-supervised anomaly detection system based on one-class evolving spiking neural network. Neural Comput Appl 2020;32:4303–14. https://doi.org/10.1007/s00521-019-04363-x.
[45] Shah S, Goyal M. Anomaly Detection in Social Media Using Recurrent Neural Network. International Conference on Computational Science, Springer; 2019, p. 74–83. https://doi.org/10.1007/978-3-030-22747-0_6.
[46] HC M, R M. BMADSN: Big data multi-community anomaly detection in social

networks. The International Journal of Electrical Engineering & Education 2019. https://doi.org/10.1177/0020720919891065.

[47] Sultani W, Chen C, Shah M. Real-World Anomaly Detection in Surveillance Videos. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE; 2018, p. 6479–88. https://doi.org/10.1109/CVPR.2018.00678.

[48] Liu K, Ma H. Exploring Background-bias for Anomaly Detection in Surveillance Videos. Proceedings of the 27th ACM International Conference on Multimedia, New York, NY, USA: ACM; 2019, p. 1490–9. https://doi.org/10.1145/3343031.3350998.

[49] Landi F, Snoek CGM, Cucchiara R. Anomaly Locality in Video Surveillance. ArXiv 2019;abs/1901.1.

[50] Wang X, Che Z, Jiang B, Xiao N, Yang K, Tang J-B, et al. Robust Unsupervised Video Anomaly Detection by Multipath Frame Prediction. IEEE Trans Neural Netw Learn Syst 2021:1–12. https://doi.org/10.1109/TNNLS.2021.3083152.

[51] Yu J, Lee Y, Yow KC, Jeon M, Pedrycz W. Abnormal Event Detection and Localization via Adversarial Event Prediction. IEEE Trans Neural Netw Learn Syst 2021;PP:1–15. https://doi.org/10.1109/TNNLS.2021.3053563.

[52] Liu Z, Nie Y, Long C, Zhang Q, Li G. A Hybrid Video Anomaly Detection Framework via Memory-Augmented Flow Reconstruction and Flow-Guided Frame Prediction. ArXiv 2021;abs/2108.0.

[53] Markovitz A, Sharir G, Friedman I, Zelnik-Manor L, Avidan S. Graph Embedded Pose Clustering for Anomaly Detection. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, p. 10536–44. https://doi.org/10.1109/CVPR42600.2020.01055.

[54] Chang Y, Tu Z, Xie W, Yuan J. Clustering Driven Deep Autoencoder for Video Anomaly Detection, 2020, p. 329–45. https://doi.org/10.1007/978-3-030-58555-6_20.

[55] Lu Y, Yu F, Reddy MKK, Wang Y. Few-Shot Scene-Adaptive Anomaly Detection. European Conference on Computer Vision, Springer; 2020, p. 125–41. https://doi.org/10.1007/978-3-030-58558-7_8.

[56] Morais R, Le V, Tran T, Saha B, Mansour M, Venkatesh S. Learning Regularity in Skeleton Trajectories for Anomaly Detection in Videos. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE; 2019, p. 11988–96. https://doi.org/10.1109/CVPR.2019.01227.

[57] Nguyen TN, Meunier J. Anomaly Detection in Video Sequence With Appearance-Motion Correspondence. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE; 2019, p. 1273–83. https://doi.org/10.1109/ICCV.2019.00136.

[58] Lv H, Zhou C, Cui Z, Xu C, Li Y, Yang J. Localizing Anomalies From Weakly-Labeled Videos. IEEE Transactions on Image Processing 2021;30:4505–15. https://doi.org/10.1109/TIP.2021.3072863.

[59] Tian Y, Pang G, Chen Y, Singh R, Verjans JW, Carneiro G. Weakly-supervised Video Anomaly Detection with Robust Temporal Feature Magnitude Learning. ArXiv Preprint ArXiv:210110030 2021.

[60] Kamoona AM, Gosta AK, Bab-Hadiashar A, Hoseinnezhad R. Multiple Instance-Based Video Anomaly Detection using Deep Temporal Encoding-Decoding. ArXiv Preprint ArXiv:200701548 2020.

[61] Liu W, Luo W, Li Z, Zhao P, Gao S. Margin Learning Embedded Prediction for Video Anomaly Detection with A Few Anomalies. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, {IJCAI-19}, International Joint Conferences on Artificial Intelligence Organization; 2019, p. 3023–30. https://doi.org/10.24963/ijcai.2019/419.

[62] Zhong J-X, Li N, Kong W, Liu S, Li TH, Li G. Graph Convolutional Label Noise Cleaner: Train a Plug-And-Play Action Classifier for Anomaly Detection. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE; 2019, p. 1237–46. https://doi.org/10.1109/CVPR.2019.00133.

[63] Zhang J, Qing L, Miao J. Temporal Convolutional Network with Complementary Inner Bag Loss for Weakly Supervised Anomaly Detection. 2019 IEEE International Conference on Image Processing (ICIP), IEEE; 2019, p. 4030–4. https://doi.org/10.1109/ICIP.2019.8803657.

[64] Ramachandra B, Jones MJ, Raju Vatsavai R. Learning a distance function with a Siamese network to localize anomalies in videos. 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE; 2020, p. 2587–96. https://doi.org/10.1109/WACV45572.2020.9093417.

[65] Li J, Zhang B, Lu G, Zhang D. Generative multi-view and multi-feature learning for classification. Information Fusion 2019;45:215–26. https://doi.org/10.1016/j.inffus.2018.02.005.

[66] Chu W, Xue H, Yao C, Cai D. Sparse Coding Guided Spatiotemporal Feature Learning for Abnormal Event Detection in Large Videos. Trans Multi 2019;21:246–255. https://doi.org/10.1109/TMM.2018.2846411.

[67] O'Mahony N, Campbell S, Carvalho A, Harapanahalli S, Hernandez GV, Krpalkova L, et al. Deep Learning vs. Traditional Computer Vision, 2020, p. 128–44. https://doi.org/10.1007/978-3-030-17795-9_10.

[68] Chong YS, Tay YH. Modeling Representation of Videos for Anomaly Detection using Deep Learning: A Review. ArXiv 2015.

[69] Wang T, Qiao M, Zhu A, Niu Y, Li C, Snoussi H. Abnormal Event Detection via Covariance Matrix for Optical Flow Based Feature. Multimedia Tools Appl 2018;77:17375–17395. https://doi.org/10.1007/s11042-017-5309-2.

[70] Zhang X, Yang S, Zhang J, Zhang W. Video anomaly detection and localization using motion-field shape description and homogeneity testing. Pattern Recognit 2020;105:107394. https://doi.org/10.1016/j.patcog.2020.107394.

[71] Lowe D. Distinctive Image Features from Scale-Invariant Keypoints. Int J Comput Vis 2004;60:91-. https://doi.org/10.1023/B:VISI.0000029664.99615.94.

[72] Bay H, Tuytelaars T, Van Gool L. SURF: Speeded Up Robust Features. Comput Vis Image Underst, vol. 110, 2006, p. 404–17. https://doi.org/10.1007/11744023_32.

[73] Calonder M, Lepetit V, Strecha C, Fua P. BRIEF: Binary Robust Independent Elementary Features. In: Daniilidis K, Maragos P, Paragios N, editors. Computer Vision -- ECCV 2010, Berlin, Heidelberg: Springer Berlin Heidelberg; 2010, p. 778–92. https://doi.org/10.1007/978-3-642-15561-1_56.

[74] Xuan Mo, Monga V, Bala R, Zhigang Fan. Adaptive Sparse Representations for Video Anomaly Detection. IEEE Transactions on Circuits and Systems for Video Technology 2014;24:631–45. https://doi.org/10.1109/TCSVT.2013.2280061.

[75] Wu S, Moore BE, Shah M. Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE; 2010, p. 2054–60. https://doi.org/10.1109/CVPR.2010.5539882.

[76] Weinland D, Ronfard R, Boyer E. Motion history volumes for free viewpoint action recognition. Workshop on modeling people and human interaction (PHI'05), 2005.

[77] Vemulapalli R, Arrate F, Chellappa R. Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group. 2014 IEEE Conference on Computer Vision and Pattern Recognition, IEEE; 2014, p. 588–95. https://doi.org/10.1109/CVPR.2014.82.

[78] Yang Cong, Junsong Yuan, Yandong Tang. Video Anomaly Search in Crowded Scenes via Spatio-Temporal Motion Context. IEEE Transactions on Information Forensics and Security 2013;8:1590–9. https://doi.org/10.1109/TIFS.2013.2272243.

[79] Surasak T, Takahiro I, Cheng C, Wang C, Sheng P. Histogram of oriented gradients for human detection in video. 2018 5th International Conference on Business and Industrial Research (ICBIR), IEEE; 2018, p. 172–6. https://doi.org/10.1109/ICBIR.2018.8391187.

[80] Dalal N, Triggs B, Schmid C. Human Detection Using Oriented Histograms of Flow and Appearance. In: Leonardis A, Bischof H, Pinz A, editors. Computer Vision -- ECCV 2006, Berlin, Heidelberg: Springer Berlin Heidelberg; 2006, p. 428–41. https://doi.org/10.1007/11744047_33.

[81] Maqsood R, Bajwa UI, Saleem G, Raza RH, Anwar MW. Anomaly recognition from surveillance videos using 3D convolution neural network. Multimed Tools Appl 2021;80:18693–716.

[82] Sabokrou M, Fayyaz M, Fathy M, Moayed Zahra, Klette R. Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. Computer Vision and Image Understanding 2018;172:88–97. https://doi.org/https://doi.org/10.1016/j.cviu.2018.02.006.

[83] Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Comput 1997;9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

[84] Kingma DP, Welling M. An Introduction to Variational Autoencoders. Foundations and Trends® in Machine Learning 2019;12:307–92. https://doi.org/10.1561/2200000056.

[85] Kingma DP, Welling M. Stochastic gradient VB and the variational auto-encoder. Second International Conference on Learning Representations, ICLR, vol. 19, 2014, p. 121.

[86] Liznerski P, Ruff L, Vandermeulen R, Franks B, Kloft M, Müller K-R. Explainable Deep One-Class Classification 2020.

[87] Ruff L, Vandermeulen R, Goernitz N, Deecke L, Siddiqui SA, Binder A, et al. Deep one-class classification. International conference on machine learning, PMLR; 2018, p. 4393–402.

[88] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM 2020;63:139–44. https://doi.org/10.1145/3422622.

[89] Aggarwal A, Mittal M, Battineni G. Generative adversarial network: An overview of theory and applications. International Journal of Information Management Data Insights 2021;1:100004.

https://doi.org/https://doi.org/10.1016/j.jjimei.2020.100004.
[90] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is All You Need. Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA: Curran Associates Inc.; 2017, p. 6000–6010.
[91] Yuan H, Cai Z, Zhou H, Wang Y, Chen X. TransAnomaly: Video Anomaly Detection Using Video Vision Transformer. IEEE Access 2021;9:123977–86. https://doi.org/10.1109/ACCESS.2021.3109102.
[92] Tran D, Bourdev L, Fergus R, Torresani L, Paluri M. Learning Spatiotemporal Features with 3D Convolutional Networks. 2015 IEEE International Conference on Computer Vision (ICCV), IEEE; 2015, p. 4489–97. https://doi.org/10.1109/ICCV.2015.510.
[93] Carreira J, Zisserman A. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE; 2017, p. 4724–33. https://doi.org/10.1109/CVPR.2017.502.
[94] Diba A, Fayyaz M, Sharma V, Karami AH, Arzani MM, Yousefzadeh R, et al. Temporal 3D ConvNets: New Architecture and Transfer Learning for Video Classification 2017.
[95] Wang L, Xiong Y, Wang Z, Qiao Y, Lin D, Tang X, et al. Temporal Segment Networks: Towards Good Practices for Deep Action Recognition. European conference on computer vision, Springer; 2016, p. 20–36. https://doi.org/10.1007/978-3-319-46484-8_2.
[96] Girdhar R, Ramanan D, Gupta A, Sivic J, Russell B. ActionVLAD: Learning spatio-temporal aggregation for action classification 2017.
[97] AlMubarak HA, Stanley J, Guo P, Long R, Antani S, Thoma G, et al. A Hybrid Deep Learning and Handcrafted Feature Approach for Cervical Cancer Digital Histology Image Classification. International Journal of Healthcare Information Systems and Informatics 2019;14:66–87. https://doi.org/10.4018/IJHISI.2019040105.
[98] Pogorelov K, Ostroukhova O, Petlund A, Halvorsen P, de Lange T, Espeland HN, et al. Deep learning and handcrafted feature based approaches for automatic detection of angiectasia. 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), IEEE; 2018, p. 365–8. https://doi.org/10.1109/BHI.2018.8333444.
[99] Nanni L, Luca E, Facin ML, Maguolo G. Deep Learning and Handcrafted Features for Virus Image Classification. J Imaging 2020;6:143. https://doi.org/10.3390/jimaging6120143.
[100] Pogorelov K, Ostroukhova O, Jeppsson M, Espeland H, Griwodz C, de Lange T, et al. Deep Learning and Hand-Crafted Feature Based Approaches for Polyp Detection in Medical Videos. 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS), IEEE; 2018, p. 381–6. https://doi.org/10.1109/CBMS.2018.00073.
[101] Mishra S, Sagban R, Yakoob A, Gandhi N. Swarm intelligence in anomaly detection systems: an overview. International Journal of Computers and Applications 2021;43:109–18.
[102] Iftikhar MS, Fraz MR. A survey on application of swarm intelligence in network security. Trans Mach Learn Artif Intell 2013;1:1–15.
[103] Nasir MH, Khan SA, Khan MM, Fatima M. Swarm intelligence inspired intrusion detection systems—a systematic literature review. Computer Networks 2022;205:108708.
[104] Kaltsa V, Briassouli A, Kompatsiaris I, Strintzis MG. Swarm-based motion features for anomaly detection in crowds. 2014 IEEE international conference on image processing (ICIP), IEEE; 2014, p. 2353–7.
[105] Pisner DA, Schnyer DM. Support vector machine. Mach Learn, Elsevier; 2020, p. 101–21.
[106] Wei X, Huang H. A survey on several new popular swarm intelligence optimization algorithms 2023.
[107] Tan Y, Zhu Y. Fireworks algorithm for optimization. Advances in Swarm Intelligence: First International Conference, ICSI 2010, Beijing, China, June 12-15, 2010, Proceedings, Part I 1, Springer; 2010, p. 355–64.
[108] Duan H, Qiao P. Pigeon-inspired optimization: a new swarm intelligence optimizer for air robot path planning. International Journal of Intelligent Computing and Cybernetics 2014;7:24–37.
[109] Mirjalili S. Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. Neural Comput Appl 2016;27:1053–73.
[110] Mirjalili S. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. Knowl Based Syst 2015;89:228–49.
[111] Arora S, Singh S. Butterfly optimization algorithm: a novel approach for global optimization. Soft Comput 2019;23:715–34.
[112] Askarzadeh A. A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm. Comput Struct 2016;169:1–12.
[113] Mirjalili S, Lewis A. The whale optimization algorithm. Advances in Engineering Software 2016;95:51–67.

[114] Rezvanian A, Vahidipour SM, Sadollah A. An Overview of Ant Colony Optimization Algorithms for Dynamic Optimization Problems 2023.

[115] Wang X, Yan Y, Tang P, Bai X, Liu W. Revisiting multiple instance neural networks. Pattern Recognit 2018;74:15–24. https://doi.org/10.1016/j.patcog.2017.08.026.

[116] Lea C, Flynn MD, Vidal R, Reiter A, Hager GD. Temporal Convolutional Networks for Action Segmentation and Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, p. 1003–12. https://doi.org/10.1109/CVPR.2017.113.

[117] Zhu Y, Newsam S. Motion-Aware Feature for Improved Video Anomaly Detection. BMVC, 2019.

[118] Sun D, Yang X, Liu M-Y, Kautz J. PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE; 2018, p. 8934–43. https://doi.org/10.1109/CVPR.2018.00931.

[119] Degardin B, Proença H. Iterative weak/self-supervised classification framework for abnormal events detection. Pattern Recognit Lett 2021;145:50–7. https://doi.org/10.1016/j.patrec.2021.01.031.

[120] Wang Y, Yao Q, Kwok JT, Ni LM. Generalizing from a Few Examples. ACM Comput Surv 2021;53:1–34. https://doi.org/10.1145/3386252.

[121] Doshi K, Yilmaz Y. Online anomaly detection in surveillance videos with asymptotic bound on false alarm rate. Pattern Recognit 2021;114:107865. https://doi.org/10.1016/j.patcog.2021.107865.

[122] Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, p. 779–88. https://doi.org/10.1109/CVPR.2016.91.

[123] Wu P, Liu J, Shi Y, Sun Y, Shao F, Wu Z, et al. Not only Look, But Also Listen: Learning Multimodal Violence Detection Under Weak Supervision, 2020, p. 322–39. https://doi.org/10.1007/978-3-030-58577-8_20.

[124] Ullah W, Ullah A, Haq IU, Muhammad K, Sajjad M, Baik SW. CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. Multimed Tools Appl 2021;80:16979–95. https://doi.org/10.1007/s11042-020-09406-3.

[125] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE; 2016, p. 770–8. https://doi.org/10.1109/CVPR.2016.90.

[126] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI, 2015.

[127] Arnab A, Dehghani M, Heigold G, Sun C, Lučić M, Schmid C. ViViT: A Video Vision Transformer. ArXiv 2021.

[128] Feng X, Song D, Chen Y, Chen Z, Ni J, Chen H. Convolutional Transformer based Dual Discriminator Generative Adversarial Networks for Video Anomaly Detection. Proceedings of the 29th ACM International Conference on Multimedia, New York, NY, USA: ACM; 2021, p. 5546–54. https://doi.org/10.1145/3474085.3475693.

[129] Arjovsky M, Chintala S, Bottou L. Wasserstein Generative Adversarial Networks. In: Precup D, Teh YW, editors. Proceedings of the 34th International Conference on Machine Learning, vol. 70, PMLR; 2017, p. 214–23.

[130] Li S, Liu F, Jiao L. Self-training multi-sequence learning with Transformer for weakly supervised video anomaly detection. Proceedings of the AAAI, Virtual 2022;24.

[131] Liu Z, Ning J, Cao Y, Wei Y, Zhang Z, Lin S, et al. Video swin transformer. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, p. 3202–11.

[132] Chen Y, Liu Z, Zhang B, Fok W, Qi X, Wu Y-C. MGFN: Magnitude-Contrastive Glance-and-Focus Network for Weakly-Supervised Video Anomaly Detection 2022.

[133] Zhou H, Yu J, Yang W. Dual Memory Units with Uncertainty Regulation for Weakly Supervised Video Anomaly Detection 2023.

[134] Qasim T, Bhatti N. A hybrid swarm intelligence based approach for abnormal event detection in crowded environments. Pattern Recognit Lett 2019;128:220–5.

[135] PRIYADHARSINI NK, KAVITHA R, KALIAPPAN A, CHITRA DRD. Hybrid Deep Learning Technique with One Class Svm for Anomaly Detection in Crowded Environment n.d.

[136] Alsolai H, Al-Wesabi FN, Motwakel A, Drar S. Improved Chicken Swarm Optimizer with Vision-based Anomaly Detection on Surveillance Videos for Visually Challenged People. Journal of Disability Research 2023;2:71–8.

[137] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning, PMLR; 2019, p. 6105–14.

[138] Li Y, Lu Y, Li D, Zhou M, Xu C, Gao X, et al. Trajectory optimization of high-speed robotic positioning with suppressed motion jerk via improved chicken swarm algorithm. Applied Sciences 2023;13:4439.

[139] Kumar SN, Rani RS. Anomalous Human Action Monitoring in Video Images Using RPCA-MFTSL AND PSO-CNN. SN Comput Sci 2023;5:109.
[140] Laine S, Aila T. Temporal Ensembling for Semi-Supervised Learning 2016.
[141] Mahadevan V, Li W, Bhalodia V, Vasconcelos N. Anomaly detection in crowded scenes. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE; 2010, p. 1975–81. https://doi.org/10.1109/CVPR.2010.5539872.
[142] Mahadevan, Vijay. UCSD Dataset 2010. http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm.
[143] Luo W, Liu W, Gao S. A Revisit of Sparse Coding Based Anomaly Detection in Stacked RNN Framework. 2017 IEEE International Conference on Computer Vision (ICCV), IEEE; 2017, p. 341–9. https://doi.org/10.1109/ICCV.2017.45.
[144] Lu C, Shi J, Jia J. Abnormal Event Detection at 150 FPS in MATLAB. 2013 IEEE International Conference on Computer Vision, 2013, p. 2720–7. https://doi.org/10.1109/ICCV.2013.338.
[145] Hershey S, Chaudhuri S, Ellis DPW, Gemmeke JF, Jansen A, Moore RC, et al. CNN architectures for large-scale audio classification. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE; 2017, p. 131–5. https://doi.org/10.1109/ICASSP.2017.7952132.
[146] University M. UMN Dataset n.d.;16. http://mha.cs.umn.edu/proj_events.shtml.
[147] Pranav M, Zhenggang L. A day on campus-an anomaly detection dataset for events in a single camera. Proceedings of the Asian Conference on Computer Vision, 2020.