

Deep learning-based methods for anomaly detection in video surveillance: a review

Abdelhafid Berroukham¹, Khalid Housni¹, Mohammed Lahraichi^{1,2}, Idir Boulfrifi¹

¹Laboratory of research in informatics L@RI, Department of Computer Science, Faculty of Science, Ibn Tofail University, Kenitra, Morocco

²CRMEF Casablanca-Settat, Casablanca, Morocco

Article Info

Article history:

Received Apr 11, 2022

Revised Sep 24, 2022

Accepted Oct 8, 2022

Keywords:

Anomaly detection

Deep learning

Video processing

ABSTRACT

Detecting anomalous events in videos is one of the most popular computer vision topics. It is considered a challenging task in video analysis due to its definition, which is subjective or context-dependent. Various approaches have been proposed to address the anomaly detection problems. These approaches vary from hand-crafted to deep learning. Many researchers have gone into determining the best approach for effectively detecting anomalies in video streams while maintaining a low false alarm rate. The results proved that approaches based on deep learning offer very interesting results in this field. In this paper, we review a family of video anomaly detection approaches based on deep learning techniques, which are compared in terms of their algorithms and models. Moreover, we have grouped state-of-the-art methods into different categories based on the approach adopted to differentiate between normal and abnormal events, and the underlying assumptions. Furthermore, we also present publicly available datasets and evaluation metrics used in existing works. Finally, we provide a comparison and discussion on the results of various approaches according to different datasets. This paper can be a good starting point for such researchers to understand this field and review existing work related to this topic.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Abdelhafid Berroukham

Laboratory of research in informatics L@RI, Department of Computer Science, Faculty of Science

Ibn Tofail University

Kenitra, Morocco

Email: a.berroukham@gmail.com

1. INTRODUCTION

The huge deployment of surveillance camera systems in public areas in recent years has increased the demand of new systems that can automatically analyze video surveillance streams in real-time. Automatically detecting abnormal events in complicated and crowded scenes is a challenging task in intelligent video surveillance. This problem has attracted significant computer vision research interest in recent years. In this work, we aim to present and evaluate the anomaly detection approaches and deep learning-based methods, to automatically detect and localize anomalous events in which subject knowledge is continuously evolving. In this section, the research topic, background information, the research objectives are covered in order to introduce the study and finally the paper structure.

Anomaly detection in the video is the task of recognizing frames from a video sequence that reflect occurrences that differ significantly from the normal, identifying unusual incidents, such as fires, car accidents, escapes, stampedes, or fighting, and can be quite useful [1], [2]. The detection and localization of the anomaly are one of the most difficult tasks in video processing due to the definition of “anomaly” which

can have some degree of ambiguity within context. Visual behaviors are complicated and diverse in an unrestricted world, complicated backgrounds, moving cameras, occlusion, shadows, and lighting are challenges to overcome. In general, an occurrence is regarded as an "anomaly" if it occurs infrequently or unexpectedly [3], [4].

Anomaly detection is a growing field of research in and of itself. Although various methods have been put out to address this issue, they all have their limitations. Whereas, the inclusion of a labeled dataset with a collection of normal events is a requirement for the majority of approaches currently in use [1]. This presumption restricts their field of use because it prevents the system from being continuously retrained without human intervention.

Various approaches have been proposed, early literature relies on trajectory-based techniques [5], [6]. These techniques attempt to determine the target's trajectories by using visual tracking and a model is learned to describe normal actions. Then the anomaly is defined as an activity related to trajectories that differ significantly from the learned model. Though, these techniques are ineffective for complex and crowded scenes due to their high temporal complication and the occlusion issue caused by moving objects [7]. Therefore, more lately, non-object-centered unsupervised approaches have been more commonly used. These approaches tackle the problem of anomaly identification by learning representative activity patterns from the behavior-related characteristics of objects and humans in spatial and temporal contexts. Size, gradient, speed, and direction of the targets in the image are typically taken into account as behavioral attributes and are expressed with low-level representations like 3D spatio-temporal gradient, histogram of optical flow (HOF), histogram of oriented gradients (HOG) [8], and dense spatial-temporal interest points (dense STIPs). These methods have an advantage over trajectory-based methods in that they work at the pixel level, which makes them more robust in complicated scenes [7].

Dictionary learning is another proposed approach for anomaly event detection; this approach develops a dictionary of typical events and labels the events that the dictionary cannot adequately depict as abnormal. Low-level features like 3D gradient features and HOF or HOG features may also be subject to dictionary learning [1]. However, all of these methods depend on hand-crafted features that are difficult to describe a priori because there are so many different types of anomalous behaviors. In addition, they are unable to adapt to abnormalities that have never encountered before [7].

Recently, a variety of computer vision tasks have been successfully tackled using deep learning approaches, surpassing the state-of-the-art in a variety of difficult problems. Such as object classification [9]–[11], object detection [12], [13], and action recognition [8], [14], [15]. Deep learning is a subtype of machine learning that achieves high performance by learning to represent the information as a hierarchy of nested concepts within layers of the neural network [16]. As the volume of data increases, deep learning outperforms classical machine learning as illustrated in Figure 1.

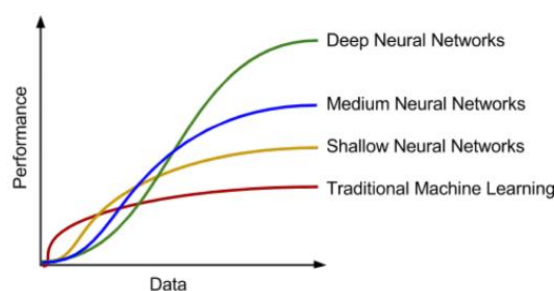


Figure 1. Deep learning-based algorithms's performance in comparison to traditional algorithms [16]

The deep learning-based methods for anomaly detection use one of these techniques: the reconstruction error to calculate the test data divergence from a series of normal training videos, the future frame prediction, the classifiers, or the scoring methods. Most of these techniques, specifically on "traditional" approaches, presuppose the existence of a labeled dataset that represents a collection of 'normal' events. In this work, we present a variety of contributions that tackle these issues. Especially we focus on deep learning-based methods to solve this issue. Today, these solving approaches based on deep learning are rapidly and constantly evolving, which makes it particularly difficult to master this area of expertise. Unlike the previous review papers which are general and tackle the anomaly detection problem in many fields, our paper is more specific for anomaly detection in video surveillance context using deep learning approaches and it covers this problem from different sides: techniques, used dataset, and metrics.

This paper is organized as follows: the first section serves as an introduction, in the second section, we review the deep learning-based methods for anomaly detection in surveillance video. In the third section, we provide the publicly available dataset and in the fourth section, we describe the most used evaluation metrics in order to evaluate and compare the methods. In the fifth section, we compare and discuss the results of different approaches according to several datasets. Finally, we terminate this paper with a conclusion.

2. DEEP LEARNING-BASED METHODS FOR ANOMALY DETECTION

Deep learning algorithms have proven effective in a variety of computer vision tasks, such as object classification [9], [17], object detection [12], [18], and action recognition [19], [20], including anomaly detection in video surveillance. As already introduced in the previous section, the approaches that have been proposed to tackle this challenge can be grouped into four categories: reconstruction error, future frame prediction, classifiers, and scoring.

2.1. Reconstruction error based methods

The reconstruction error is one of the most used approaches for solving the anomaly detection problem. The basic presumption of using the reconstruction error is that would be smaller for normal samples, because they are closer to the training data, and assumed to be higher for abnormal samples [21]. Deep learning-based methods typically train a deep neural network using an auto-encoder (AE) method and use it to reconstruct normal events with few reconstruction errors. But as it was claimed in [22], larger reconstruction errors for anomalous events don't necessarily happen. As a result, it can show that practically many methods based on the reconstruction of training data cannot guarantee the detection of abnormal events.

A method was proposed in [23] to learn normal patterns with minimal supervision using autoencoders; firstly, the authors use the conventional hand-crafted Spatio-temporal local features to train an autoencoder. The value of using this type of information for training is their capacity to work without or with minimal supervision. Then, they develop a fully convolutional AE to learn the classifiers and the local features in one framework.

Another method was proposed in [24] where the authors used generative adversarial networks (GANs) [25], which employ normal frames and associated optical-flow images as training data to learn the normal frame representation. The GANs cannot generate abnormal events because they have only been trained on normal data. Therefore, to detect abnormalities, a local differential between the actual and produced images is used during testing time. In future work, it could be possible to use dynamic images [26] to represent motion data.

Similarly, the work of [27] has also used GANs [28] and performs transfer learning algorithms on pre-trained CNN (VGG16). Transfer learning is a vital machine learning technique for addressing the fundamental issue of insufficient training data. Its goal is to transfer knowledge from one domain to another [16]. They also improve the model's effectiveness by processing the video's optical-flow information. The experiment of this work runs on University of California, San Diego (UCSD) datasets, and for the evaluation, they use various criteria such as area under the receiver operating characteristic (ROC) curve (AUC) and equal error rate (EER).

Vu *et al.* [28], propose an approach based on two-fold. They propose a customizable multi-channel framework for generating multi-type frame-level characteristics on one side and on other side; they investigate how supervised learning can be used to increase detection performance. The multi-channel framework that they propose is composed of four conditional GANs (CGANs) [29] that take various types of motion and appearance data as input and produce prediction data as output. Then peak signal-to-noise ratio (PSNR) is used to encode the difference between the generating and ground-truth information. For frame-level anomaly detection, the binary support vector machines (SVM) is used. Finally, they perform object-centric anomaly localization by using mask region-based convolutional neural networks (R-CNN) as a detector. They evaluate their solution on four different datasets: avenue, ShanghaiTech, and UCSD.

Sabokrou *et al.* [21], propose an approach for anomaly detection and localization based on two cubic patches, where one relies on the strength of an autoencoder to reconstruct an input video patch, while the other relies on the strength of sparse representation of an input video patch. These two stages are constructed based on the analysis of the reconstruction error of the AE and the sparsity value (SV). The main idea of their approach is that the anomaly patch in the testing phase has a more elevated reconstruction error than a normal patch if an AutoEncoder has been trained successfully on the normal patches.

2.2. Scoring based methods

There is another category of methods proposed by researchers based on score [6], [21], [22], [30]; the main idea of this approach is to generate an anomaly score that may be used to determine whether or not

a video segment or frame is abnormal. Sultani *et al.* [30] propose an approach to learn anomalies by utilizing both normal and abnormal videos; they postulated that the best way to detect anomalies might not be to use only normal data. Therefore, to save the time-consuming task of marking anomalous portions in training films, they suggest using weakly labeled training videos to learn anomalies using the deep multiple instance ranking system [31].

In their approach, the authors learn an anomaly ranking model that automatically predicts high anomaly scores for anomalous video segments by treating video segments as instances in multiple instances learning (MIL) and normal and abnormal videos as bags. MIL is a deep learning technique where training data is organized in bags, and each bag contains a collection of instances [32]. Research by Pang *et al.* [1], try to solve the problem by end-to-end anomaly scores learning on a collection of video frames without explicitly labeling any data as normal or abnormal. For that, they propose an end-to-end approach based on self-trained deep ordinal regression to detect the anomaly in the video. This approach overcomes some limitations of existing methods, the first one relies on manually labeled normal training data, and the second one is sub-optimal feature learning.

The framework that has been proposed receives a collection of videos without labels and then initially carries out initial detection to produce a set of pseudo anomalous and normal frames. Then, these collections are used to train a ResNet-50 model [33] and a fully connected network in an end-to-end fashion. ResNet50 is a pre-trained model that has the ability of take frame appearance characteristics. The network is composed of an output layer with one linear unit and a hidden layer with 100 units. Finally, the anomaly scores of all frames are then recalculated using the trained model. The abnormal and normal memberships are updated as needed, and the process is repeated.

Another method was proposed by Xu *et al.* [7] where they have proposed an unsupervised learning approach to learn feature representations automatically. They propose a new double fusion architecture to take advantage of the complementing information contained in both appearance and movement patterns, combining typical early fusion and late fusion advantages. In the early fusion, it is proposed to use stacked denoising auto-encoders (SDAE) to learn both the motion and appearance features of activities in a video separately. Then, they employ multiple one-class SVM models to predict the anomaly scores of each input using the learned features. Finally, the late fusion combines the obtained scores and detects anomalous events. As claimed by the author, this work is the first effort to tackle the challenge of abnormal event identification using deep learning. Despite the good results achieved, the approach still has a limit that is represented in the high computational for real-time processing. Therefore, in the future, it might be possible to research ways to cut the cost of computation.

2.3. Future frame-based methods

This approach is considered as another sight to address the anomaly detection challenge within a future frame prediction. The assumption of its use is that normal events are predictable whereas abnormal ones do not match expectations. The first work that introduces this approach is that of [22]. In which the authors propose a future frame prediction network. This approach is based on the generator-discriminator structure assimilated to that of a GAN network, and they use a U-net model as a prediction network to create a future frame while the discriminator at the end of the network determines whether or not the predicted frame is abnormal. Moreover, to predict a higher-quality future frame for normal events in addition to appearance constraints that are commonly used, they also use a motion constraint by forcing the optical flow between the ground truth and the anticipated frames.

Another method was proposed by Medel and Savakis [34], where they used a future frame prediction approach. Their approach is based on developing generative models that, with limited supervision, can detect anomalies in videos. They suggest a composite convolutional long short-term memory (ConvLSTM) network that is end-to-end trainable and can anticipate the development of a video sequence given a few input frames and predict future frames. The network learns to predict 'normal' activities that are comparable to those seen in the training videos. And with each succeeding timestep, the abnormality forecast deviates further from the ground truth. As a result, the regularity score produced can be used to identify when abnormalities occur in videos. At the evaluation level, the authors did not use the most used matrices for evaluating results and making comparisons with other methods like AUC and EER.

2.4. Classifier based methods

The work of Medel and Savakis [4] framed the anomaly detection problem as a classification problem. They proposed an approach for locating and detecting anomalies in videos by analyzing the output of deep layers, their approach uses fully convolutional neural networks (FCNNs) and information about time. The proposed FCN combines a pre-trained CNN using an AlexNet model [9] with a novel convolutional layer that trains kernels with regard to the training video. The network focuses on two key tasks: outlier detection and feature representation. This approach proved good results in terms of accuracy but it still has

some limitations, it occurs false positives in some cases like when people walk in different directions and when we have crowded scenes. Summary of past literature for anomaly detection techniques is shown in Table 1.

Table 1. Deep learning-based approaches for anomaly detection

Year	Learning type	Approach	Contribution	Used techniques	Ref
2020	Weakly supervised	Score	Self-trained deep ordinal regression for end-to-end video anomaly detection	ResNet50, end-to-end, self-training, ordinal regression	[1]
2018	Supervised weakly	Score	Using MIL to predict high anomaly scores for anomalous video segments	MIL sparsity, temporal smoothness	[30]
2016	limited supervision	Reconstruction error	learn normal patterns using autoencoders with limited supervision	Fully convolutional autoencoder	[23]
2017	Unsupervised	Reconstruction error	Train GAN to learn an internal representation of scene normality using normal frames and related optical-flow images	GAN, optical-flow	[24]
2021	Unsupervised	Reconstruction error	multi-channel framework based on 4 CGAN to generate multi-type frame-level features	CGAN, SVM, Full Flow; Mask R-CNN	[28]
2018	Unsupervised	Future frame	future frame prediction network for anomaly detection	GAN, U-Net, optical-flow	[22]
2016	Limited supervision	Future frame	end-to-end trainable composite Conv-LSTM networks	Conv-LSTM	[35]
2020	Unsupervised	Reconstruction error	Abnormal event detection using GAN and transfer learning	GAN, transfer learning pre-trained CNN (VGG16)	[27]
2018	Unsupervised	Classification	FCN: the combination of a pretrained CNN (AlexNet) and a novel convolutional layer	Optical flow FCNN, Alexnet, Gaussian classifier	[4]
2017	Unsupervised	Score	AMDN: unsupervised learning approach based on deep learning architectures	Sparse auto-encoder (SAE)	[7]
2016	Unsupervised	Reconstruction error	Two cubic patch approach based on AE and sparse representation	SDAE, multiple one-class SVM models, fine-tuning	[21]

3. BENCHMARK DATASETS

In this part, we describe the public datasets used for the anomaly detection tasks in the video. Many of the papers attempted to use at least one benchmark dataset to compare the performance of their suggested methods to previously published papers. Due to the variable crowd density and behavior patterns, all datasets exhibit dynamic scenarios. The datasets frequently used for activities involving anomaly detection are listed in Table 2.

Table 2. A comparison of anomaly datasets

Dataset	Number of videos	Number of frames	Average frames	Training video	Testing video	Anomalous events	Resolution	DATASET length	Number of scenes	Examples of Anomalies
Subway entrance	1	144,249	144,249	15 min		66	512×384	1,5 hours	1	No payment, loitering, Wrong way
Subway exit	1	64,900	64,900	15 min		19	512×384	43 min	1	Wrong direction, loitering
UMN	11	~7,700	1,290			11	320×240	5 min	3	Run
UCSD Ped1 [36]	50	14,000	200	34	16	40	238×158	5 min	1	Small cars, skaters, walking in the grass
UCSD Ped2 [37]	70	4,560	163	16	12	12	360×240	5 min	1	Skaters, small cars, bikers
CUHK avenue [38]	28	35,240	2,120	16	21	14	640×360	30 min	1	Running, throwing objects, and loitering
Street scene	15	203,257		46	35	205	1280×720		1	
Shanghai tech	437	317,398		330	107	130	856×480		13	
UCF-crime [30]	1900	~13.8M	4,052	1,610	290	13	320×240	128 hours	n	Burglary, fights, robbery, accidents on the road

3.1. UCSD pedestrian

The UCSD pedestrian dataset [37] contains 2 subsets: the UCSD Peds1 dataset and the UCSD Peds2 dataset, the size of the frame and the camera angle distinguish the two subsets. The dataset is divided into testing and training data. The training data is devoid of abnormalities, it is all normal activities and contains only pedestrians; however there is at least one anomaly in every testing clip, the anomalous events are either: object entities moving via pathways or anomalous people motion. Common anomalies contain small cars, skaters, bikes, and people walking in the grass, in certain frames, the anomalies appear in multiple locations.

UCSD pedestrian 1: this dataset has 34 video sequences for training, and 16 video sequences for testing in which one or more anomalies are present in some of the frames, pixel-level binary masks are given to a collection of ten clips in the testing set to identify regions having anomalous events, each clip contains about 200 frames. There are 5,500 normal and 3,400 abnormal frames, with a resolution 158×238 pixels. In this dataset, The camera is positioned at a considerable height.

UCSD pedestrian 2: this dataset contains around 1,652 anomalous and 346 normal frames across 12 testing and 16 training video sequences. The frame has a 360 by 240 pixel resolution. The camera here is placed at a lower altitude. Each testing clip in this dataset has only one anomalous event, which takes up the majority of the video segment.

Different works are usually evaluated independently on these two datasets. But due to the different camera viewpoints, Ped1 appears to be more challenging than Ped2. Figure 2 shows sample frames from the UCSD dataset for both normal and abnormal behavior in the scene and their ground truth.

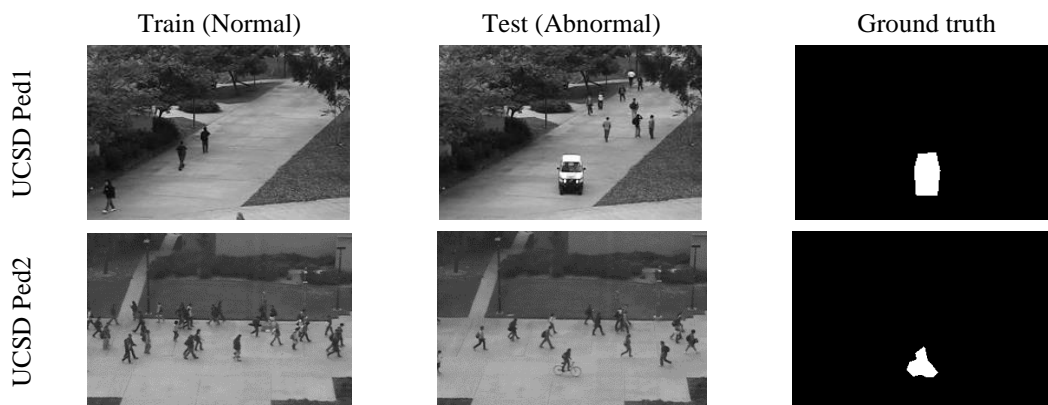


Figure 2. Samples from the UCSD dataset; left column illustrates normal pedestrian behavior, the middle shows the anomaly behavior in the scene and the right column shows their ground truth

3.2. CUHK avenue

Chinese University of Hong Kong (CUHK) avenue dataset [39] includes 16 video sequences for training and 21 video sequences for testing, each video around 2 minutes long. There are a total of 47 anomalous events, like throwing objects, running, loitering, and going the wrong way. Due to the camera position and viewpoint, people's sizes may vary. The training video contains generally normal events. However, there are a few abnormal situations. The number of normal samples in the test set is greater than the number of abnormal samples. Figure 3 illustrates sample frames of abnormal behavior from the CUHK avenue dataset.



Figure 3. Samples of abnormal behavior from the CUHK avenue dataset

3.3. Subway dataset

The subway dataset [40] comprises 2 video sequences recorded at the access point (144,249 frames, 1 hour 36 minutes long) and exit door (64,900 frames, 43 minutes long) of a subway station. The abnormal events mainly include individuals traveling in the opposite direction and no-payment events. The number of anomalies in this dataset are low. Figure 4 shows sample frames from the Subway dataset for both normal and abnormal events. Subway entrance: the surveillance video from the subway entrance shows a variety of anomalous events, such as people loitering, walking in the opposite way, and avoiding payment. Subway exit: similar anomalies to those seen in the subway entrance video can be seen in the surveillance video of the subway exit.



Figure 4. Samples from the subway dataset; the top row displays regular events, whereas the bottom row displays abnormal ones

3.4. UMN dataset

The University of Minnesota (UMN) dataset comprises 3 distinct sights of escape incidents, with a total number of frames 7740 (1,450 for scene 1, 4,415 for scene 2, and 2,145 for scene 3) and the resolution is 320×240. The abnormal activities are people spreads running at the same moment, while the normal events are pedestrians wandering aimlessly around the plaza or through the mall. There are 11 abnormal events in the entire video collection. Figure 5 illustrates example frames from the UMN dataset.



Figure 1. Samples from the UMN dataset; top row depicts normal crowd behaviour, while the bottom row depicts panicked crowd behavior

3.5. ShanghaiTech dataset

The ShanghaiTech dataset includes 330 videos for training and 107 videos for testing, with over 270,000 training frames. There are 130 abnormal events and numerous forms of anomalies with 13 scenarios that incorporate difficult lighting and camera positions. Furthermore, the ground truth of abnormal events is labeled. On the test set, normal samples outnumber abnormal samples, Figure 6 shows sample frames from this dataset for both abnormal and normal behavior.

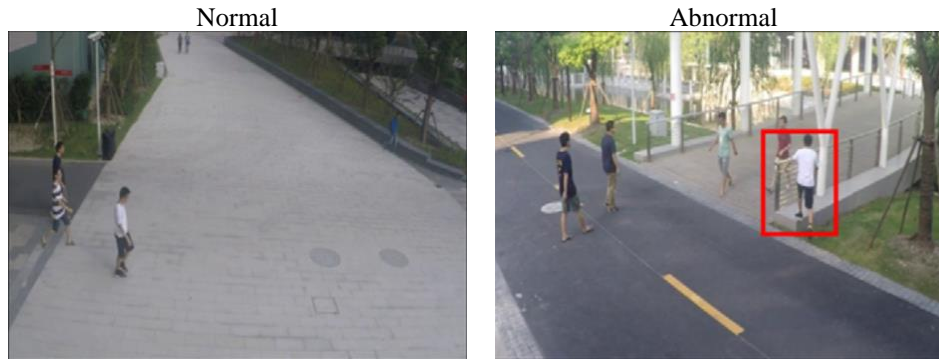


Figure 2. Normal and abnormal frames, the red box denotes an anomaly in an anomalous frame.

3.6. UCF dataset

The University of Central Florida (UCF) dataset is a sizeable dataset proposed by [30] to help solve the anomaly detection problem with about 128 hours of videos. It contains 1,900 lengthy actual surveillance movies, with 13 realistic abnormalities, including burglary, fights, robbery, accidents on the road, and also the normal activities. This dataset can be utilized for two different purposes. First, all anomalies are taken into account in one group, while all normal events are taken into account in another. Second, to identify each of the 13 anomalous activities. There are 15 times as many movies in this dataset as there are in other datasets. Figure 7 shows few examples of anomalies from the UCF dataset.

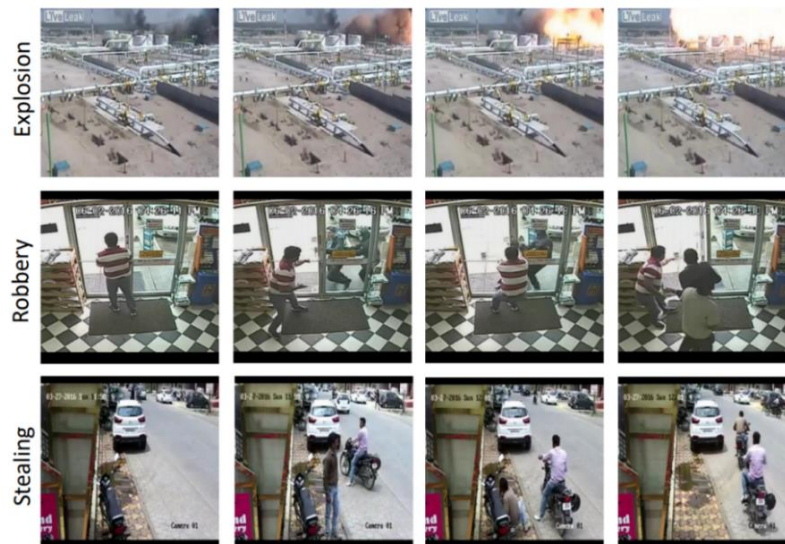


Figure 3. Samples of anomalies from the UCF dataset

4. EVALUATION METRICS

In this section, we will discuss the evaluation and comparison measures used in state-of-the-art methods.

- Frame level: a frame is deemed to have detection if it has at least one abnormal pixel. Each frame's ground truth annotation is compared to these detections. The process is carried out several times for different thresholds to create a ROC curve. This assessment does not confirm that the detection corresponds to the actual location of the anomaly. Therefore, some actual positive detections may be the result of "fortunate" co-occurrences of false positives and abnormal events [37].
- Pixel level: the accuracy of localization is evaluated by comparing detections to pixel-level ground truth masks, on a collection of ten clips. The process is comparable to what was previously stated. The frame is deemed as accurately detected if at least 40% of the actually anomalous pixels are found. otherwise, it is tallied as a false positive [37].
- ROC curve: to evaluate the accuracy for various threshold settings, the ROC curve is employed. The ROC is composed of false positive rate (FPR) and true positive rate (TPR), where FPR determines the proportion of false-positive findings that occur as compared to the total number of negative samples available through the test stage, and TPR defines a classifier test performance on accurately categorizing positive instances among all available positive samples throughout the test stage. These measurements are provided by (1) and (2):

$$\text{TPR} = \frac{\text{True positive}}{\text{False negative} + \text{True positive}} \quad (1)$$

$$\text{FPR} = \frac{\text{False positive}}{\text{True negative} + \text{False positive}} \quad (2)$$

where true positive (TP) denotes the anomalous events that have been properly identified; true negative (TN) denotes the normal events that have been properly identified, false positive (FP) denotes the anomalous events that have been improperly identified; and false negative (FN) denotes the normal events that have been improperly identified. We select several thresholds for both frame-level and pixel-level detection and compute the TPR and FPR in accordance to produce the ROC curve [39].

The AUC is employed as the evaluation metric. The ground truth and frame-level anomaly scores are used to calculate AUC. Figure 8 illustrates the area under the ROC curve.

The EER is the proportion of incorrectly categorized frames when the FPR and the miss rate are both equal. The lower the EER value, the higher the accuracy of the algorithm. The EER is a point in the ROC at the junction of the curve and a line going from (0.1) to (1.0). Figure 8 illustrates the EER. Time complexity is another important criterion. If an algorithm's overall execution time is sufficiently short, it is more appealing to be used in many applications.

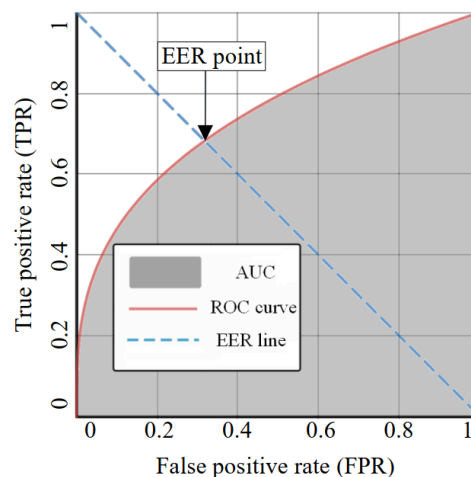


Figure 4. EER and ROC curve [27]

5. COMPARISON AND DISCUSSION

In this section, we will discuss and analyze the performance of anomaly detection methods in video sequences, exactly those based on deep learning approaches. Table 3 (in Appendix) lists the approaches discussed in the previous sections and other papers that tackle the anomaly detection problems in accordance

with the publicly available datasets. These approaches are grouped by the type of learning used and some evaluated metrics results obtained by applying some anomaly detection methods on different datasets. The comparison of accuracy between different methods is done by their frame and pixel-level scores.

We have classified papers based on deep learning into four categories of approaches: reconstruction errors, future frame prediction, scoring, and using classifiers. The accuracy of each one is tested on several datasets and evaluated using AUC and EER metrics for both frame and pixel-level. As shown in Table 3 (in Appendix), the deep learning-based methods are achieved good results for the most available dataset compared to hand-crafted based methods, except some methods for some specific dataset like [40] which has the lowest EER value (10%) compared to all the others methods for UCSD Ped2 dataset, and the method of [37] in subway entrance dataset, and also the method of [41] that achieved an accuracy of 99.70% in UMN dataset.

The analysis of the deep learning-based methods results demonstrates that the reconstruction errors are the most used approach and gives a superior accuracy in UCSD datasets for both frame and pixel-level, as shown in [23], [24], [28]. But in some situations, the larger reconstruction errors for anomalous events may not happen because of the higher capacity of the deep neural network. Whereas score approach has also achieved good results for some other datasets as in [42], especially in the subway exit (AUC=95,1%) and UMN(AUC=99,83%) datasets. In addition, the approach presented in [30], has also given a good accuracy (AUC=75,41%) in their dataset UCF compared to the results of other approaches, but it could not locate exactly the anomaly in some situations.

For the classifier approach, we can see that the approach presented in [43] has achieved good accuracy (AUC=97,80%) for the UCSD Ped2 dataset, but this approach generates a high rate of false-positives (AUC=68,4% in UCSD Ped1 dataset) in 2 situations: when people walk in the wrong way and in the crowded scenes. Despite the future frame prediction approach proves its effectiveness for anomaly detection on some datasets (AUC=95,4% in UCSD Ped2). In the avenue dataset, it fails to detect several anomalous events of jogging that occur in the background, because it could not differentiate jogging action from walking pedestrians. In general, using some datasets is more challenging than others. For example, all approaches give good results using the UMN dataset, due to its simplicity. But in UCF dataset, the higher result obtained is (AUC=75,41%).

Based on the reviewed literature papers and the results of Table 3 (in Appendix) [1], [4], [7], [21]–[24], [27], [28], [30], [35]–[37], [40]–[60], it appears clearly that several studies choose to tackle the anomaly detection problem using unsupervised learning methods, because do not require labeled video data and can be effectively employed for learning good representations. In addition, it is effective to the complexity and variety of visual behaviors of anomaly in an unconstrained environment. However, they still limited and did not achieve good results. Therefore, other researchers choose to surpass this limit by using the semi-supervised learning methods that use data only related to the "normal" class, thus these methods have greater specifications for anomaly detection problem as well as unsupervised methods, which only use the structure and configuration of the unlabeled data and do not use any other information.

Despite the very huge researchers in this topic, however, it still has some limits; many anomaly-detection algorithms work with very regular scenes, so it is necessary to evaluate how well these methods operate in less structured situations. Moreover, the real time application in unconstrained environment and the time complexity. Therefore, we propose to use the vision transformer model [61], which is a new deep learning technique that achieve good results in many problems and it could be a good approach to implement for anomaly detection problem.

6. CONCLUSION

This paper reviews deep learning-based methods for video anomaly detection, which cover a variety of approaches, techniques, datasets, and evaluation metrics. A thorough overview of anomaly detection should ideally enable readers to comprehend not just the rationale for using a specific technique, but also to compare different techniques and produce a comparative analysis, in addition to propose an approach. Firstly, we have classified the approaches into four types of categories: reconstruction errors, future frame prediction, scoring, and using classifiers. We also presented the strengths and weaknesses of each category according to several datasets. Each category can be applied in a supervised or unsupervised manner, but most researchers focused on tackling the anomaly detection problem by applying unsupervised learning.

Furthermore, we have presented the different publicly available datasets with their details such as the video resolution and example anomalies found within the respective datasets, and we found that many datasets are more challenging than the others. Finally, we have discussed the results of several categories applied to different datasets. Aiming to tackle some problems and achieve good results in both the accuracy and computational complexity, there are research opportunities to develop a new approach based on vision transformer to improve the detection of anomaly object in video sequences.

APPENDIX

Table 3. The results of different approaches according to several used dataset

Approach Categories	Dataset Ref	Dataset																				
		UCSD						Subway				CUHK Avenue				UMN		SHT	UCF			
		Ped 1		Pixel-level		Ped 2		Entrance		Exit		Frame-level		Pixel-level		All Scenes		Frame level	Frame level			
Hand-crafted based methods	[44]	50,30%						63,00%											76,50%			
	[40]					10%		17%														
	[45]	56,30%						67,50%				80,50%		91%					87,10%			
	[46]	40%	59,00%	81%	20,50%	30%	69,30%	71%														
	[47]	31%	67,50%	79%	19,70%	42%	55,60%	80%											96,00%			
	[37]	32%	68,80%	71%	21,30%	36%	61,30%	72%														
	[37]	25%	81,80%	58%	44,10%	25%	82,90%	54%	16,70%	90,80%	16,40%	89,7%										
	[36]	15%	91,80%	43%	63,80%	-	-													65,51%		
	[48]	19%		54%	45,30%	20%				24,40%	83,30%	26,40%	80,2%							97,80%		
	[41]		87,00%						91,00%											99,70%		
	[35]					19%			29,90%													
	supervised learning	Rec-error	[23]	27,9%	81,00%			21,7%	90%		26,20%	94,30%	9,90%	80,7%	25,1%	70,2%				61%	50,60%	
			[28]		85,00%				96%							92%	30,21%	74,43%			94%	
		score	[49]							92,20%												
			[50]		68,40%					82,20%		70,60%		85,7%		80,6%					95,10%	
Classification		[51]		69,00%					87,50%		71,60%		93,1%							95,20%		
		[4]					11%		15%	17,00%	90,40%	16%	90,2%									
Future frame		[52]					19%		24%									2,50%	99,60%			
		[22]		83,10%					95,40%						85,1%					72,8%		
Deep learning based methods		Unsupervised learning	[53]						94,10%													
			[54]		75,50%					88,10%		93,30%		87,7%		77%						
	[55]								92,20%						81,7%					68%		
	[27]		14%	93,00%	36%	73%	15%		17%													
	[21]							15%										2,50%	99,60%			
	Reconstruction error	[24]	8%	97,40%	35%	70,30%	14%	93,50%														
		[56]						96,20%							87%							
		[42]									93,50%		95,1%							99,30%		
		[1]		71,70%					83,20%		88,10%		92,7%							99,83%		
		[7]	16%	92,10%	40,10%	67,20%	17%	90,80%	42%													
Score	[57]		92,10%			20%	90,80%	42%														
	[58]		93,75%		65,11%		94,09%												99,65%			
	[59]													84,6%								
	[60]	8%	95,70%	40,80%	64,50%	18%	88,40%															
	Weakly sup	Classi- fi- cation	[43]						97,80%													
Score [30]																				75,41%		

REFERENCES

[1] G. Pang, C. Yan, C. Shen, A. van den Hengel, and X. Bai, "Self-Trained Deep Ordinal Regression for End-to-End Video Anomaly Detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 12170–12179, doi: 10.1109/CVPR42600.2020.01219.

[2] S. A. Mahmood, A. M. Abid, and S. H. Lafta, "Anomaly event detection and localization of video clips using global and local outliers," *IJECS*, vol. 24, no. 2, p. 1063, Nov. 2021, doi: 10.11591/ijeecs.v24.i2.pp1063-1073.




[3] O. P. Popoola and Kejun Wang, "Video-Based Abnormal Human Behavior Recognition—A Review," *IEEE Trans. Syst., Man, Cybern. C*, vol. 42, no. 6, pp. 865–878, Nov. 2012, doi: 10.1109/TSMCC.2011.2178594.

- [4] M. Sabokrou, M. Fayyaz, M. Fathy, Zahra. Moayed, and R. Klette, "Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes," *Computer Vision and Image Understanding*, vol. 172, pp. 88–97, Jul. 2018, doi: 10.1016/j.cviu.2018.02.006.
- [5] F. Jiang, Y. Wu, and A. K. Katsaggelos, "A Dynamic Hierarchical Clustering Method for Trajectory-Based Unusual Video Event Detection," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 907–913, Apr. 2009, doi: 10.1109/tip.2008.2012070.
- [6] C. Piciarelli and G. L. Foresti, "On-line trajectory clustering for anomalous events detection," *Pattern Recognition Letters*, vol. 27, no. 15, pp. 1835–1842, Nov. 2006, doi: 10.1016/j.patrec.2006.02.004.
- [7] D. Xu, Y. Yan, E. Ricci, and N. Sebe, "Detecting anomalous events in videos by learning deep representations of appearance and motion," *Computer Vision and Image Understanding*, vol. 156, pp. 117–127, Mar. 2017, doi: 10.1016/j.cviu.2016.10.010.
- [8] A. Jahagirdar and R. Phalnikar, "Comparison of feed forward and cascade forward neural networks for human action recognition," *IJECS*, vol. 25, no. 2, p. 892, Feb. 2022, doi: 10.11591/ijeecs.v25.i2.pp892-899.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
- [10] M. S. Naga Raju and B. S. Rao, "Colorectal multi-class image classification using deep learning models," *Bulletin EEI*, vol. 11, no. 1, pp. 195–200, Feb. 2022, doi: 10.11591/eei.v11i1.3299.
- [11] A. AL Smadi, A. Mehmood, A. Abugabah, E. Almekhlafi, and A. M. Al-smadi, "Deep convolutional neural network-based system for fish classification," *IJECE*, vol. 12, no. 2, p. 2026, Apr. 2022, doi: 10.11591/ijece.v12i2.pp2026-2039.
- [12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, Jun. 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.
- [13] S. M. Abas, A. M. Abdulazeez, and D. Q. Zeebaree, "A YOLO and convolutional neural network for the detection and classification of leukocytes in leukemia," *IJECS*, vol. 25, no. 1, p. 200, Jan. 2022, doi: 10.11591/ijeecs.v25.i1.pp200-213.
- [14] K. Simonyan and A. Zisserman, "Two-Stream Convolutional Networks for Action Recognition in Videos," arXiv, Nov. 12, 2014, doi: 10.48550/arXiv.1406.2199.
- [15] M. A. Alsaedi, A. S. Mohialdeen, and B. M. Albaker, "Development of 3D convolutional neural network to recognize human activities using moderate computation machine," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 6, pp. 3137–3146, Dec. 2021, doi: 10.11591/eei.v10i6.2802.
- [16] R. Chalapathy and S. Chawla, "Deep Learning for Anomaly Detection: A Survey," *arXiv:1901.03407 [cs, stat]*, Jan. 2019, Accessed: Aug. 03, 2021. [Online]. Available: <http://arxiv.org/abs/1901.03407>
- [17] A. Kherraki and R. El Ouazzani, "Deep convolutional neural networks architecture for an efficient emergency vehicle classification in real-time traffic monitoring," *IJ-AI*, vol. 11, no. 1, p. 110, Mar. 2022, doi: 10.11591/ijai.v11.i1.pp110-120.
- [18] R. Jain, A. Goyal, and K. Venkatesan, "Real-time eyeglass detection using transfer learning for non-standard facial data," *IJECE*, vol. 12, no. 4, p. 3709, Aug. 2022, doi: 10.11591/ijece.v12i4.pp3709-3720.
- [19] H. A. Razak, M. A. M. Saleh, and N. M. Tahir, "Review on anomalous gait behavior detection using machine learning algorithms," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 5, pp. 2090–2096, Oct. 2020, doi: 10.11591/eei.v9i5.2255.
- [20] S. Sharma, B. Sudharsan, S. Naraharisetti, V. Trehan, and K. Jayavel, "A fully integrated violence detection system using CNN and LSTM," *IJECE*, vol. 11, no. 4, p. 3374, Aug. 2021, doi: 10.11591/ijece.v11i4.pp3374-3380.
- [21] M. Sabokrou, M. Fathy, and M. Hoseini, "Video anomaly detection and localisation based on the sparsity and reconstruction error of auto-encoder," *Electron. lett.*, vol. 52, no. 13, pp. 1122–1124, Jun. 2016, doi: 10.1049/el.2016.0440.
- [22] W. Liu, W. Luo, D. Lian, and S. Gao, "Future Frame Prediction for Anomaly Detection - A New Baseline," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, Jun. 2018, pp. 6536–6545, doi: 10.1109/CVPR.2018.00684.
- [23] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning Temporal Regularity in Video Sequences," *arXiv:1604.04574 [cs]*, Apr. 2016, Accessed: Jul. 04, 2021. [Online]. Available: <http://arxiv.org/abs/1604.04574>
- [24] M. Ravanbakhsh, M. Nabi, E. Sanginetto, L. Marcenaro, C. Regazzoni, and N. Sebe, "Abnormal Event Detection in Videos using Generative Adversarial Nets," *arXiv:1708.09644 [cs]*, Aug. 2017, Accessed: Jul. 25, 2021. [Online]. Available: <http://arxiv.org/abs/1708.09644>
- [25] I. J. Goodfellow *et al.*, "Generative Adversarial Networks." arXiv, Jun. 10, 2014. Accessed: Oct. 12, 2022. [Online]. Available: <http://arxiv.org/abs/1406.2661>
- [26] H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, and S. Gould, "Dynamic Image Networks for Action Recognition," Jun. 2016, doi: 10.1109/cvpr.2016.331.
- [27] A. Atghaei, S. Ziaeinejad, and M. Rahmati, "Abnormal Event Detection in Urban Surveillance Videos Using GAN and Transfer Learning," *arXiv:2011.09619 [cs]*, Nov. 2020, Accessed: May 17, 2021. [Online]. Available: <http://arxiv.org/abs/2011.09619>
- [28] T.-H. Vu, J. Boonaert, S. Ambellouis, and A. Taleb-Ahmed, "Multi-Channel Generative Framework and Supervised Learning for Anomaly Detection in Surveillance Videos," *Sensors*, vol. 21, no. 9, p. 3179, May 2021, doi: 10.3390/s21093179.
- [29] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," *arXiv:1411.1784 [cs, stat]*, Nov. 2014, Accessed: Aug. 01, 2021. [Online]. Available: <http://arxiv.org/abs/1411.1784>
- [30] W. Sultani, C. Chen, and M. Shah, "Real-World Anomaly Detection in Surveillance Videos," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, Jun. 2018, pp. 6479–6488, doi: 10.1109/CVPR.2018.00678.
- [31] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artificial Intelligence*, vol. 89, no. 1–2, pp. 31–71, Jan. 1997, doi: 10.1016/s0004-3702(96)00034-3.
- [32] M. Combalia and V. Vilaplana, "Monte-Carlo Sampling Applied to Multiple Instance Learning for Histological Image Classification," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer International Publishing, 2018, pp. 274–281, doi: 10.1007/978-3-030-00889-5_31.
- [33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Jun. 2016, doi: 10.1109/CVPR.2016.90.
- [34] J. R. Medel and A. Savakis, "Anomaly Detection in Video Using Predictive Convolutional Long Short-Term Memory Networks," arXiv, Dec. 15, 2016. Accessed: Oct. 12, 2022. [Online]. Available: <http://arxiv.org/abs/1612.00390>
- [35] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly Detection and Localization in Crowded Scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 18–32, Jan. 2014, doi: 10.1109/tpami.2013.111.
- [36] C. Lu, J. Shi, and J. Jia, "Abnormal Event Detection at 150 FPS in MATLAB," *2013 IEEE International Conference on Computer Vision*, Dec. 2013, doi: 10.1109/iccv.2013.338.




- [37] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, Jun. 2010, pp. 1975–1981, doi: 10.1109/CVPR.2010.5539872.
- [38] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust Real-Time Unusual Event Detection using Multiple Fixed-Location Monitors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 555–560, Mar. 2008, doi: 10.1109/tpami.2007.70825.
- [39] Y. Cong, J. Yuan, and J. Liu, "Abnormal event detection in crowded scenes using sparse representation," *Pattern Recognition*, vol. 46, no. 7, pp. 1851–1864, Jul. 2013, doi: 10.1016/j.patcog.2012.11.021.
- [40] T. Xiao, C. Zhang, and H. Zha, "Learning to Detect Anomalies in Surveillance Video," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1477–1481, Sep. 2015, doi: 10.1109/LSP.2015.2410031.
- [41] Y. Zhang, H. Lu, L. Zhang, X. Ruan, and S. Sakai, "Video anomaly detection based on locality sensitive hashing filters," *Pattern Recognition*, vol. 59, pp. 302–311, Nov. 2016, doi: 10.1016/j.patcog.2015.11.018.
- [42] R. T. Ionescu, S. Smeureanu, M. Popescu, and B. Alexe, "Detecting abnormal events in video using Narrowed Normality Clusters," *arXiv:1801.05030 [cs]*, Nov. 2018, Accessed: Oct. 15, 2021. [Online]. Available: <http://arxiv.org/abs/1801.05030>
- [43] R. T. Ionescu, F. S. Khan, M.-I. Georgescu, and L. Shao, "Object-Centric Auto-Encoders and Dummy Anomalies for Abnormal Event Detection in Video," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 7834–7843, doi: 10.1109/CVPR.2019.00803.
- [44] A. D. Giorno, J. A. Bagnell, and M. Hebert, "A Discriminative Framework for Anomaly Detection in Large Videos," in *Computer Vision – ECCV 2016*, Springer International Publishing, 2016, pp. 334–349, doi: 10.1007/978-3-319-46454-1_21.
- [45] M. Sugiyama and K. Borgwardt, "Rapid Distance-Based Outlier Detection via Sampling," in *Advances in Neural Information Processing Systems*, 2013, vol. 26. Accessed: Oct. 12, 2022. [Online]. Available: <https://papers.nips.cc/paper/2013/hash/d296c101daa88a51f6ca8cfc1ac79b50-Abstract.html>
- [46] J. Kim and K. Grauman, "Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates," Jun. 2009, doi: 10.1109/cvpr.2009.5206569.
- [47] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," Jun. 2009, doi: 10.1109/cvpr.2009.5206641.
- [48] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," Jun. 2011, doi: 10.1109/cvpr.2011.5995434.
- [49] R. Hinami, T. Mei, and S. Satoh, "Joint Detection and Recounting of Abnormal Events by Learning Deep Generic Knowledge," Oct. 2017, doi: 10.1109/iccv.2017.391.
- [50] R. T. Ionescu, S. Smeureanu, B. Alexe, and M. Popescu, "Unmasking the Abnormal Events in Video," Oct. 2017, doi: 10.1109/iccv.2017.315.
- [51] Y. Liu, C.-L. Li, and B. Póczos, "Classifier Two-Sample Test for Video Anomaly Detections," *Machine Learning Department Carnegie Mellon University Pittsburgh, USA*, 2018.
- [52] M. Sabokrou, M. Fathy, M. Hoseini, and R. Klette, "Real-time anomaly detection and localization in crowded scenes," Jun. 2015, doi: 10.1109/cvprw.2015.7301284.
- [53] D. Gong *et al.*, "Memorizing Normality to Detect Anomaly: Memory-Augmented Deep Autoencoder for Unsupervised Anomaly Detection," Oct. 2019, doi: 10.1109/iccv.2019.00179.
- [54] W. Luo, W. Liu, and S. Gao, "Remembering history with convolutional LSTM for anomaly detection," Jul. 2017, doi: 10.1109/icme.2017.8019325.
- [55] W. Luo, W. Liu, and S. Gao, "A Revisit of Sparse Coding Based Anomaly Detection in Stacked RNN Framework," Oct. 2017, doi: 10.1109/iccv.2017.45.
- [56] T. N. Nguyen and J. Meunier, "Anomaly Detection in Video Sequence with Appearance-Motion Correspondence," Oct. 2019, doi: 10.1109/iccv.2019.00136.
- [57] D. Xu, E. Ricci, Y. Yan, J. Song, and N. Sebe, "Learning Deep Representations of Appearance and Motion for Anomalous Event Detection," 2015, doi: 10.5244/c.29.8.
- [58] Q. Sun, H. Liu, and T. Harada, "Online growing neural gas for anomaly detection in changing surveillance scenes," *Pattern Recognition*, vol. 64, pp. 187–201, Apr. 2017, doi: 10.1016/j.patcog.2016.09.016.
- [59] S. Smeureanu, R. T. Ionescu, M. Popescu, and B. Alexe, "Deep Appearance Features for Abnormal Behavior Detection in Video," in *Image Analysis and Processing - ICIAP 2017*, Springer International Publishing, 2017, pp. 779–789, doi: 10.1007/978-3-319-68548-9_70.
- [60] M. Ravanbakhsh, M. Nabi, H. Mousavi, E. Sangineto, and N. Sebe, "Plug-and-Play CNN for Crowd Motion Analysis: An Application in Abnormal Event Detection," Mar. 2018, doi: 10.1109/wacv.2018.00188.
- [61] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." *arXiv*, Jun. 03, 2021. Accessed: Oct. 12, 2022. [Online]. Available: <http://arxiv.org/abs/2010.11929>.

BIOGRAPHIES OF AUTHORS






Abdelhafid Berroukham    received the master degree in Distributed Systems Computing from Faculty of Science Agadir (FSA), University Ibne Zohr, Agadir, Morocco in 2013. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Faculty of Science, Ibn Tofail University, Kenitra, Morocco. His research interests are in deep learning, computer vision, video processing, and focusing on anomaly detection in video surveillance. He can be contacted at email: a.berroukham@gmail.com.






Khalid Housni    received the master of Advanced Study degree in applied mathematics and computer science, and the Ph.D. degree in computer science from the Ibn Zohr University of Agadir, Morocco, in 2008 and 2012, respectively. He joined the Department of Computer Science, University Ibn Tofail of Kenitra, Morocco, in 2014, where he has been involved in several projects in video analysis and network reliability. In 2019 he obtained his HDR degree (Habilitation à Diriger des Recherches: Qualification to supervise research) from Ibn Tofail University. He is a member of the Research in Informatics laboratory (L@RI) and head of the MISC team. His current research interests include image/video processing, computer vision, machine learning, artificial intelligence, pattern recognition, and network reliability. He can be contacted at email: housni.khalid@uit.ac.ma.



Mohammed Lahraichi    received the doctorate degree in video analysis from Faculty of Science Kenitra (FSK) university IbnTofail, Kenitra, Morocco in 2020. He is currently a researcher professor in CRMEF Casablanca-Settat. His research interests are: deep learning, computer vision, focusing on object detection, tracking in video sequence, and anomlay detection in video surveillance. He can be contacted at email: lahraichi.mohamed@gmail.com.



Idir Boulfrifi    was born in South of Morocco in 1983, after obtain his baccalaureate in Science of Mathematic in 2002 he joined the University Ibn Zohr to studying Science of Mathematics and Computer. In 2008 he had his master degree in system and networking, since 2009 he has occupied many jobs in computer science, in 2015 he joined the MISC Laboratory on the Ibn Tofail University to prepare his Ph.D in computer vision, specifically in risk detection by semantic analysis of video sequences. He can be contacted at email: iboulfrifi@gmail.com.