2022 International Conference on Computer Science and Software Engineering CSASE, Duhok, Kurdistan Region – Iraq

A Comprehensive Review for Video Anomaly Detection on Videos

Zainab K. Abbas
Dept of Information and Communication Engineering
College of Information Engineering
Al-Nahrain University
Baghdad, Iraq
zainabkudair@gmail.com

Ayad A. Al-Ani
Dept of Information and Communication Engineering
College of Information Engineering
Al-Nahrain University
Baghdad, Iraq
ayad.abdulaziz@coie-nahrain.edu.iq

***Abstract*—Video Surveillance Systems (VSS) are widely utilized in public and private areas to increase public safety, such as shopping malls, markets, banks, hospitals, educational institutions, streets, and smart cities. The accuracy and fast identification of video anomalies is usually the major goal of security applications. However, because of varying environmental factors, the complexities of human activity, the ambiguous nature of the anomaly, and the absence of appropriate datasets, detecting video anomalies is challenging. This paper surveys the last three years, a comprehensive study of detecting video anomalies, and the recently used dataset. Moreover, a comparison study on different approaches has been performed, which are used for anomalies detection. We have noticed that deep learning has outperformed other methods in this field.**

Keywords—Anomaly detection, deep learning, dataset, CNN, Video surveillance

I. Introduction

The capability of recognizing anomalous actions such as fighting, riots, crimes, road accidents, and stampedes, as well as anomalous things such as abandoned things and weapons in sensitive places, automatically and in real-time, is defined as Intelligent Video Surveillance Systems (IVSS) [1][2][3]. In other words, anomalous occurrences are outliers that deviate significantly from the training model [1-4]. The unusualness, abnormalities, deviants, and outliers are synonyms for literature anomalies [1-5].

A scene's abnormality is dependent on its context and time [1][2]. For example, whereas crowds in marketplaces are deemed normal on normal days, they are considered abnormal or anomalous if they are observed during curfews in the same market, implying that the abnormalities in the video are subject and context-dependent [1]. Unusual behavior can be defined as "an activity executed at an unusual location, at an unusual time" or "activities that are fundamentally different in appearance and motion" [1][2]. Anomaly detection in the video automatically detects abnormal video occurrences in spatiotemporal dimensions, such as anomalous actions or entities [2-6]. The video anomalies detection and video anomalies localization are inextricably linked. Whereas VAD seeks to determine whether a given video frame contains an anomaly by asking, "Does the given frame contain an anomaly or not?" Video anomaly localization focuses on the localization of anomalies by using the bounding box to determine the real location of anomalies in each video frame, which corresponds to the question: "Where is anomaly occurring in the given frame?" [1].

The research areas of human activity detection or recognition and VAD are closely related but not the same. On the other hand, human activity recognition could be defined as a supervised learning technique for classifying various human activities. While anomaly detection could be defined as an unsupervised learning technique that attempts to identify abnormal patterns [4]. Environmental conditions (variations, illumination, complex background, the objects' shadow effects, occlusions of objects, and so on), crowd density, the complexity of human conduct, noisy data, variations in spatial and temporal, setting of a recording camera, and difficulty in accessing good computational infrastructure are all domain challenges [1-8]. In addition to the intrinsic obstacles of Video Anomaly Detection (VAD) (data imbalance problem, ambiguous nature, significant variance within positive classes), for these difficulties, VAD is one of the most time-consuming computer vision jobs [9,10]. Consequently, anomaly event detection is still facing several challenges.

II. Literature Survey

A vital component is the VAD of the surveillance systems. This section reviews the essential related papers for the last four years, beginning from 2018.

Chaudhary *et al.* [6] suggested a method automatically detects several abnormal actions in videos. The feature extraction procedure identifies key features, speed, centroid, direction, and dimensions. They created a new dataset with 45 movies of various persons running, walking, and crawling without overlapping) for three activities.

Bhagyalakshmi *et al.* [11] proposed a weapon detection system. Live surveillance recordings are used in this study to monitor and detect unusual events using real-time image processing techniques. This project has three processing modules: the first one uses CNN to detect objects, the second one handles weapon identification, and the third one handles monitoring and alert operations.

Waqas *et al.* [12] propose to learn abnormalities by utilizing both anomalous and normal videos and present a system that can recognize anomalous attitudes and alert the

978-1-6654-2632-9/22/$31.00 ©2022 IEEE

user on the kind of abnormal behavior. This paper suggests using the deep Multiple Instance Learning (MIL) framework to learn anomalies from weakly labeled training videos. In this method, anomalous videos are treated as bags, segments are treated as instances in MIL, and a deep anomaly ranking model automatically predicts high anomaly scores for anomalous video segments. The UCF Crime dataset was used in this work.

Shine *et al.* [13] provided a real-time automated approach for detecting motorcyclists without helmets from traffic surveillance videos. Because there are no benchmarked datasets available, this work created a new dataset. The system's testing and development database comprise videos taken on public highways.

Shreyas *et al.* [14] offered a new implementation concept in which the videos are adaptively compressed before being passed via the activity recognition system. Anomaly event detection is achieved by integrating adaptive video compression and a C3D network.
Ramchandran et al. [10] In this paper, the benefits of both hand-crafted and hierarchical feature learning were combined, and an unsupervised Hybrid Deep Learning framework (HDLVAD) for VAD was proposed, which effectively detects video anomalies. Raw picture sequences are mixed with edge image sequences and fed into a convolutional Auto-encoder and ConvLSTM model to detect the abnormality. The datasets used in this research are Avenue, UCSD ped1, and UCSD ped2.

Anala *et al.* [7] described a system that can recognize abnormal actions and alert the user depending on the type of anomaly. In this work, three frames per second are used instead of extracting one frame per second, allowing capturing very subtle changes in the sequence of frames. This work treats anomaly detection as a regression problem. The evaluation of the performance of this approach is done on normal videos only. Using solely normal data may not be the best method for detecting anomalies. Extracting the features was done by the 3DCNN, and the experimentation was done on the UCF-Crime dataset.

For anomaly detection in Surveillance Videos (SV), a unique two-stream convolutional networks model was presented by Hao *et al.* [15]. They suggested a model consisting of RGB and Flow two-stream networks, with the combined score representing the final anomaly event detection score. This work treats the anomaly detection problem as a regression problem, a CNN ResNet was used for anomaly detection purposes, and the experimentation was done on the UCF-Crime dataset.

Venkatesh *et al.* [16] Discussed a practical approach to crime detection that may be utilized for on-device crime monitoring using Deep Learning. They can reduce latency, lower the expense of collecting data into a centralized unit, and mitigate the lack of privacy by making conclusions on-device. To provide low inference time, this work used the concept of Early-Stopping-MIL and LSTM for anomaly detection. The experimentation on UCF-Crime (videos crime includes eight classes for the anomaly and the Normal videos class).

Cheng *et al.* [17] presented the RWF-2000 database, containing 2,000 videos collected by security cameras in real-world scenarios and propose a new Flow Gated Network approach that combines the benefits of 3D-CNNs and optical flow.

Dubey *et al.* [5] proposed a framework, which is a deep network with Multiple Ranking Measures (DMRMs). This model addresses context-dependency using a joint learning technique for motion and appearance features. This work also considered the anomalous event detection problem as a regression problem, they used 3D ResNet-34 for anomaly detection purposes, and the implementation was done on the UCF-Crime dataset.

Ullah *et al.* [18] presented an efficient light-weight (CNN) based anomaly recognition framework functional in a surveillance environment with reduced time complexity. They used pre-trained light-weight CNN-multilayer BiLSTM for anomaly detection purposes, and the implementation was done on UCF-Crime (4 Class anomaly in addition to normal).

Ullah *et al.* [8] proposed an efficient deep features-based intelligent anomalous detection framework that can operate in surveillance networks with lesser time complexity. In this work, the feature vector used to discover the anomaly was generated from 15 successive video frames, and a CNN-ResNet-50 with Multilayer Bi-LSTM was used for anomaly detection purposes. The implementation was done on the UCF-Crime database.

Öztürk *et al.* [3] ADNet, an anomaly detection network that uses temporal convolutions to locate anomalies in videos, was proposed. The model works online by accepting multiple video clips in a row. AD-Net receives features taken from video clips in a window and uses them to localize anomalies in videos effectively. The AD Loss function is proposed to improve AD-anomalous Net's segment detection performance. Also, they propose that they employ the "F1@k" for temporal anomaly detection, the features extracted by I3D, and TSM is the second feature extractor. The implementation was done on UCF Crime. In addition, they added two different anomaly categories to the data set, namely "Molotov bomb" and "protest."

Wu *et al.* [19] provided a VAD approach based on CNN and MIL, in which the moving targets in the video are extracted using a Gaussian background model, and the features are extracted using a pre-trained VGG16 model. Finally, the MIL models are trained and forecasted at the pixel level using MISVM (Multiple-Instance Support Vector Machines) and NSK (Normalized Set Kernel) methods using the UCSD dataset.

Aziz *et al.* [20] described a method for detecting and locating anomalies in SV. A one-class support vector machine (OCSVM) is used to detect motion-based anomalous occurrences. The proposed approach detects anomalies at the frame level. It then localizes them at the pixel level in the classified anomalous frames using mask-rcnn, which provides pixel-level object masks, object class prediction, and bounding box regression. It was tested on the Avenue and UMN datasets.

Boekhoudt *et al*. [21] HR-Crime is a subset of the UCF-Crime dataset useful for human-related anomaly identification tasks, as described in this work. Build the feature extraction pipeline for human-related anomaly detection using cutting-edge approaches. In addition, they provide the HR-Crime baseline anomaly detection analysis.

Wan *et al.* [2] contributed a new Large-scale Anomaly Detection database (LAD) as a baseline for video sequence anomaly detection. It includes 2000 video sequences that include anomalous and normal video clips and 14 anomaly categories such as violence, crash, and fire. It also includes annotation data for anomaly detection, such as frame-level and video-level labels (abnormal/normal). A multi-task deep neural network is proposed to handle the anomaly detection problem, formulated as a fully supervised learning problem.

Zaheer *et al*. [22] developed a weakly supervised anomaly detection method that uses video-level labels to train. Furthermore, this research recommended the use of binary clustering, which aids in reducing noise in the labeling of anomalous films. The main network and clustering are encouraged to complement one other in reaching the goal in this task formulation. A pre-trained feature extractor model such as C3D, Fully Connected Network, and k-mean was used, and the implementation was done on UCF-crime.

Majhi *et al*. [23] proposed a method that jointly handles anomaly detection and classification in a single framework by adopting a weakly supervised learning paradigm. In weakly-supervised learning, only video-level labels are sufficient for learning instead of dense temporal annotations of weakly supervised training. A 3D CN (I3D) - many LSTM were used for anomaly detection purposes implementation on UCF-Crim.

Wu *et al.* [24] suggested a dual branch network that takes as input multi-granularity ideas in both the spatial and temporal dimensions. To record the correlation between tubes/videolets, each branch uses a relationship reasoning module, which can provide complex entity relationships and rich contextual information for the concept learning of abnormal behaviors, Weakly-Supervised Spatio-Temporal Anomaly Detection WSSTAD: aims to localize a spatiotemporal tube (i.e., a series of bounding boxes that appear at the exact moment), features was extracted by C3D and the implementation was done on a new dataset (denoted as ST-UCF-Crime) that annotates Spatio-temporal bounding boxes for abnormal events in UCF-Crime, STRA. Table I illustrates some methods recently used for features extraction purposes and the dataset used for experimentation with the AUC value they got.

TABLE I. METHODS WERE RECENTLY USED FOR FEATURE EXTRACTION PURPOSES, AND THE DATASET USED FOR EXPERIMENTATION WITH THE AUC VALUE THEY GOT.

Authors	Year	Features Extraction Method	Dataset	Performance by AUC %

[2]	2021	Pretrained I3D, LSTM	Avenue, UCSD Ped2, ShanghaiTech, UCF-crime, Proposes the LAD database	Avenue= 89.33
				UCSD Ped2= 95.12
				ShanghaiTech= 92.97
				UCF= 74.98
				LDA= 86.28

[3]	2021	I3D, TSM	UCF Crime added two different abnormal categories to the database, namely "Molotov bomb" and "protest."	-
[5]	2021	3D ResNet-34	UCF-Crime	81.91
[6]	2018	Gaussian Mixture Model	created a new dataset with 45 videos of three different activities	-
[7]	2019	CNN (VGG16) LSTM	UCF Crime (Explosion, Fighting, Road Accident and Normal)	85
[8]	2021	CNN-ResNet-50 with Multilayer Bi-LSTM	UCF-Crime	85.53

[9]	2021	C3D	New dataset" UBI-Fights", UCF-Crime, UCSD	UBI-Fights= 0.819
				UCF-Crime =74.4
				UCSD= 0.809

[10]	2019	convolutional autoencoder-ConvLSTM model	Datasets used are Avenue, UCSD ped1, and UCSD ped2	Avenue = 90.7
				ped1= 98.4
				ped2= 98.5

[11]	2019	CNN	live SVs	-
[12]	2019	Extract the 3D convolution features (3DCNN)	UCF-Crime	75.41
[13]	2020	Hand-crafted features	compiled a new dataset for this work from traffic	-
[14]	2020	3DConvNets	UCF crime	79.8
[15]	2020	CNN-ResNet	UCF-Crime	81.22
[16]	2020	LSTMs	UCF-Crime (videos crime Assault, Arson, Fighting, Burglary, Explosion, Arrest, Abuse, and	-------------

			Road Accidents are among the eight categories) +Normal	
[17]	2020	3DCNN	Proposes the RWF-2000 database, containing 2,000 movies taken by surveillance cameras in real-life situations.	85.75
[18]	2021	Pretrained lightweight CNN -multilayer Bi-LSTM	UCF-Crime (4 Class anomaly and normal)	78.43
[19]	2021	VGG16	UCSD dataset	62.66
[20]	2021	pre-trained mask-rcnn deep network	Avenue and UMN datasets	Avenue Frame level= 84.52, Pixel level 76.80 UMN Frame level= 97.53
[21]	2021	pre-trained MPED-RNN	UCF Crime	60
[22]	2021	pre-trained feature extractor C3D	UCF-crime	78.27
[23]	2021	I3D - many to many LSTM	UCF-Crim	82.12

[24]	2021	C3D	new dataset (ST-UCF-Crime) that annotates spatio-temporal bounding boxes for abnormal events in UCF-Crime, STRA	ST-UCF-Crime = 87.65
				STRA= 92.88

III. Video Anomalies Classification

The types of anomalies and the complexity of the environment are the primary elements in video anomaly identification and localization. Video anomalies can be classified as follows:

A. *Global and Local Anomalies*

A car going the wrong way is an example of a local anomalous or local anomaly activity that deviates significantly from nearby spatiotemporal activities [1]. In other words, local anomaly detection identifies local anomalous behavior in an individual or a group of people, such as a person riding a bicycle on a pathway where everyone is walking [10]. Global anomalous or global anomaly activity, on the other hand, is defined as activities that interact with one another globally in an abnormal, unusual, or suspicious manner, even if actions in isolation may be normal or abnormal. Examples of global anomalies include car accidents, bomb booms that disperse the crowd, and so on [1]. In other words, global anomaly detection is the technique of identifying people's global anomalous behavior in surveillance recordings. When an abnormal event occurs, such as a bomb blast, accident, or violence, the majority of people flee in various directions [10].

B. *Point and Interaction Anomalies*

The point anomaly is a random anomaly that can be further translated into an abnormal activity displayed by a person, such as loitering. [1].

C. *Conditional or Contextual Anomalies*

Behavioral features and contextual features identify the context. Behavioral features are qualities used to define normal activities, while time and space are considered contextual features. [1].

IV. Performance Evaluation Methodologies

The suggested algorithms' effectiveness in detecting and localizing video anomalies is assessed using computational infrastructure, datasets, and performance metrics from both quantitative and qualitative analysis.

A. *Database*

Because this is a comparatively new study subject, there are limited publicly available datasets for video anomaly identification. Furthermore, the scarcity of bench-marked datasets for video anomaly identification and location is related to the rarity of anomalous behaviors in real-life circumstances and the limitless variety of abnormal activities [1]. The most often used datasets for video anomaly detection and localization are UCSD Pedestrian dataset [25], Subway dataset [26], Avenue dataset [27], UMN dataset [28], BOSS dataset [29], Weizmann dataset [30]. The majority of publicly available datasets contain a small number of realistic anomalous behaviors, simulated abnormal behaviors, movies shot using predetermined scripts, training and test samples from various camera settings, and videos focusing mostly on ideal conditions [1]. Large datasets encompassing realistic anomalous behaviors are required for deep learning-based video anomaly detection systems. Several significant databases, such as the ShanghiTech dataset [31], UCF-Crime [12], RWF-2000 database [17], LAD database [2], UBI-Fights [9]. The datasets differ in terms of length, resolution, size, the monitoring environment, scenarios covered, challenges presented by the dataset, targeted applications, aberrant occurrences, and ground truth availability (GT). The GT is a concept used in statistics and machine learning to compare the accuracy of machine learning outcomes to the real world [32]. One of the most important aspects of deep learning-based advancements is selecting an appropriate combination of test data. Furthermore, there is a necessity for large-scale benchmarks to assess the algorithms employed for the detection and localization of video anomalies. Table II illustrates some recently used datasets with their specification. More details mentioned in [5,12,20].

B. Evaluation Criteria

In general, the performance of detection and localization of video anomaly is tested using three assessment criteria. The "Frame-Level (FL)" anomaly is identified for at least one pixel in a frame; the "Pixel-Level (PL)" 40% or more of the label of the pixel as anomalies; and the "Dual-Pixel-Level (DPL)" in which the frame must meet the PL evaluation requirements. At least 10% of the pixels detected as the anomaly and the "Object-level" result for a given threshold. This assessment criterion is utilized to detect anomalous frames with the "Detected Abnormality Area (DAA)" close to the "True Abnormality Area (TAA)" [1].

TABLE II. EXISTING VAD DATABASES WITH DETAILED INFORMATION

Database	No. of videos	Categorical	Supervision
UCSD Ped1	70	Not specified	Video-level
UCSD Ped2	28	Not specified	Video-level
Avenue	37	Not specified	Video-level
Subway	1	Not specified	Video-level
UMN	5	Not specified	Video-level
BOSS	12	Not specified	Video-level
ShanghiTech	437	Not specified	Video-level
UCF-Crime	1900	13	Video-level
RWF-2000	2000	Not specified	Video-level
LAD	2000	14	Video-level & Frame level
UBI-Fights	1000	Not specified	Frame level

C. Performance Metrics for Quantitative Analysis

Many performance measurements such as precision curves, Receiver Operating Characteristic (ROC), the Area Under the Curve (AUC), and others are used for quantitative analysis.

- *Error matrix (Confusion matrix):* A binary classification problem can be used to solve the challenge of video anomaly detection. Furthermore, each binary classifier decision can be represented by any of the performance analysis' essential parts, such as (FP), (TP), (TN), and (FN) [33]. Another advanced performance metric was mentioned in [1].
- *Receiver operating characteristic curve:* Plots TPR (on Y-axis) versus FPR (on X-axis) to assess detection performance at various FPRs. The AUC-ROC curve can be shown in Fig. 1. [34]. The anomaly detector's accuracy is measured using the (AU-ROC) for a particular test set. The AU-ROC value should be as high as possible [1].
- *Precision-Recall curve (PR):* is a graph that shows the relationship between recall (on X-axis) and precision(on Y-axis) [4]. AU-PR is more beneficial for anomaly identification than the AU-ROC. Within the range of zero to one, the AU-PR value should be as high as possible [1]. Another Performance metrics such as Equal Error Rate, Detection rate, Reconstruction error, Anomaly score, Regularity score, Peak Signal-to-Noise, and Computational complexities were described in [1, 35].

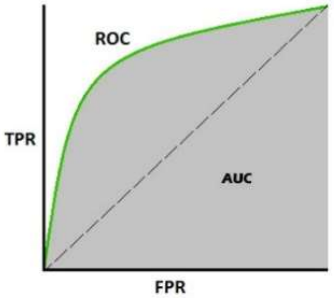

Fig. 1. AUC-ROC Curve [34]

V. CONCLUSIONS

Detection of video anomalies is an essential field for each surveillance system. Our article provides a comprehensive survey of the approaches and datasets recently employed for the VAD. Moreover, a comparative study on different approaches has been used for anomalies detection. Our comparative studies of the existing irregularity detection strategies provide a superior choice of a specific strategy that works best for a specific application. We have noticed that deep learning has outperformed other methods in this field. VAD can benefit from criminal activities, traffic violations, anomalous crowd behavior, weapons in critical places, abandoned objects, and other video surveillance application domains.

REFERENCES

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

[2] B. Wan, W. Jiang, Y. Fang, Z. Luo, and G. Ding, "Anomaly detection in video sequences: A benchmark and computational model," IET Image Process., pp. 1–10, 2021.

[3] H. İ. Öztürk and A. B. Can, "ADNet: Temporal Anomaly Detection in Surveillance Videos," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12664 LNCS, pp. 88–101, 2021.

[4] B. R. Kiran, D. M. Thomas, and R. Parakkal, "An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos," MDPI J. Imaging, pp. 1–15, 2018.

[5] S. Dubey, A. Boragule, J. Gwak, and M. Jeon, "Anomalous event recognition in videos based on joint learning of motion and appearance with multiple ranking measures," Appl. Sci., vol. 11, no. 3, pp. 1–21, 2021.

[6] S. Chaudhary, M. Aamir, and C. Bhatnagar, "ScienceDirect Multiple Anomalous Activity Detection in Videos," Procedia Comput. Sci., vol. 125, pp. 336–345, 2018.

[7] M. R. Anala, M. Makker, and A. Ashok, "Anomaly detection in surveillance videos," Proc. - 26th IEEE Int. Conf. High Perform. Comput. Work. HiPCW 2019, pp. 93–98, 2019.

[8] W. Ullah, A. Ullah, I. U. Haq, K. Muhammad, M. Sajjad, and S. W. Baik, "CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks," Multimed. Tools Appl., vol. 80, no. 11, pp. 16979–16995, 2021.

[9] B. Degardin and H. Proença, "Iterative weak/self-supervised classification framework for abnormal events detection," Pattern Recognit. Lett., vol. 145, pp. 50–57, 2021.

[10] A. Ramchandran and A. K. Sangaiah, "Unsupervised deep learning system for local anomaly event detection in crowded scenes," Multimed. Tools Appl., vol. 79, no. 47–48, pp. 35275–35295, 2019.

[11] P. Bhagyalakshmi, P. Indhumathi, and R. Lakshmi, "Real Time

Video Surveillance for Automated Weapon Detection," pp. 465–470, 2019.

[12] M. S. Waqas Sultani, Chen Chen, "Real-world Anomaly Detection in Surveillance Videos," 2019.

[13] L. Shine and C. V Jiji, "Automated detection of helmet on motorcyclists from traffic surveillance videos : a comparative analysis using hand-crafted features and CNN," 2020.

[14] D. G. Shreyas, S. Raksha, and B. G. Prasad, "Implementation of an Anomalous Human Activity Recognition System," SN Comput. Sci., vol. 1, no. 3, 2020.

[15] Hao, Wangli, Ruixian Zhang, Shancang Li, Junyu Li, Fuzhong Li, Shanshan Zhao, and Wuping Zhang. "Anomaly event detection in security surveillance using two-stream based model." Security and Communication Networks 2020 (2020).

[16] S. V. Venkatesh, A. P. Anand, G. S. Sahar, A. Ramakrishnan, and V. Vijayaraghavan, "Real-time surveillance based crime detection for edge devices," VISIGRAPP 2020 - Proc. 15th Int. Jt. Conf. Comput. Vision, Imaging Comput. Graph. Theory Appl., vol. 4, pp. 801–809, 2020.

[17] M. Cheng, K. Cai, and M. Li, "RWF-2000: An open large scale video database for violence detection," Proc. - Int. Conf. Pattern Recognit., pp. 4183–4190, 2020.

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

[19] G. Wu, Z. Guo, M. Wang, L. Li, and C. Wang, "Video abnormal event detection based on CNN and multiple instance learning," no. January, p. 30, 2021.

[20] Z. Aziz, N. Bhatti, H. Mahmood, and M. Zia, "Video anomaly detection and localization based on appearance and motion models," Multimed. Tools Appl., vol. 80, no. 17, pp. 25875–25895, 2021.

[21] K. Boekhoudt, A. Matei, M. Aghaei, and E. Talavera, "HR-Crime: Human-Related Anomaly Detection in Surveillance Videos," pp. 1–10, 2021.

[22] M. Z. Zaheer, J. Lee, M. Astrid, A. Mahmood, and S.-I. Lee, "Cleaning Label Noise with Clusters for Minimally Supervised Anomaly Detection," pp. 3–6, 2021, [Online]. Available: http://arxiv.org/abs/2104.14770.

[23] S. Majhi, S. Das, F. Bremond, R. Dash, and P. K. Sa, "Weakly-supervised Joint Anomaly Detection and Classification," 2021, [Online]. Available: http://arxiv.org/abs/2108.08996.

[24] J. Wu et al., "Weakly-Supervised Spatio-Temporal Anomaly Detection in Surveillance Video," pp. 1172–1178, 2021.

[25] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 1975–1981, 2010.

[26] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 3, pp. 555–560, 2008.

[27] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB," Proc. IEEE Int. Conf. Comput. Vis., pp. 2720–2727, 2013.

[28] N. Bird, S. Atev, N. Caramelli, R. Martin, O. Masoud, and N. Papainkolopoulos, "Real time, online detection of abandoned objects in public areas," Proc. - IEEE Int. Conf. Robot. Autom., vol. 2006, no. May, pp. 3775–3780, 2006.

[29] "BOSS Dataset," [Online]. Available: http://velastin.dynu.com/videodatasets/BOSSdata/whole_dataset.html.

[30] "Weizmann Dataset," [Online]. Available: https://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html.

[31] W. Liu, W. Luo, D. Lian, and S. Gao, "Future Frame Prediction for Anomaly Detection - A New Baseline," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 6536–6545, 2018.

[32] "Ground Truth definition," [Online]. Available: https://www.techopedia.com/definition/32514/ground-truth.

[33] A. Kulkarni, D. Chong, and F. A. Batarseh, "Foundations of data imbalance and solutions for a data democracy," Data Democr. Nexus Artif. Intell. Softw. Dev. Knowl. Eng., pp. 83–106, Jan. 2020.

[34] "ROC-AUC definition," [Online]. Available: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5.

[35] P. Wu, J. Liu, and F. Shen, "A Deep One-Class Neural Network for Anomalous Event Detection in Complex Scenes," IEEE Trans. Neural Networks Learn. Syst., vol. 31, no. 7, pp. 2609–2622, 2020