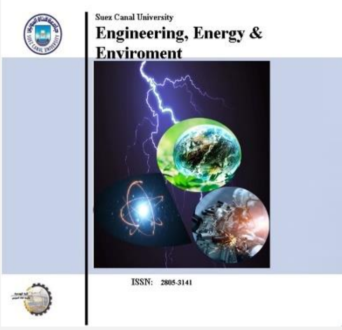
Suez Canal Engineering, Energy
and Environmental Science Journal
https://sceee.journals.ekb.eg/

Vol 1, No. 2, 1-9



A Review on Video Anomaly Detection Datasets

Iman M. Yossef [1], Marwa Gamal [1], Rehab F. Abdel-Kader [2], Khaled Abd Elsalam Ali [1]

[1] *Electrical Engineering Department, Faculty of Engineering, Suez Canal University, Ismailia, Egypt.*

[2] *Electrical Engineering Department, Faculty of Engineering, Port Said University, Port Said, Egypt.*

Abstract – In recent years, Video Anomaly Detection (VAD) has received a lot of attention and has become a popular research topic. This is due to their immense potential in a variety of fields, including healthcare monitoring, surveillance/crowd analysis, sports, Ambient Assistive Living (AAL), event analysis, and security. Manually detecting and analysing improper behavior was a hard process, particularly in real-time scenarios, resulting in a high demand for smart surveillance systems. Moreover, the availability of data plays a vital role in training and evaluating models. Datasets in VAD are typically composed of sequences of frames or videos, some of which depict normal activities and others that depict anomalous or unusual events. These datasets provide a rich resource that encapsulates everyday routine actions alongside irregular or unusual events, fostering the development and assessment of robust anomaly detection models. This paper provides an extensive review of the most popular and recent datasets in VAD including an extensive comparison between them.

Keywords:
Video Anomaly Detection, Video Surveillance, Video Anomaly Datasets.

I. INTRODUCTION

Human activity recognition (HAR) is a field of study within computer vision and machine learning that involves the development of algorithms and systems capable of identifying and classifying human actions and behaviors from input data, often in the form of sensor data or video streams. HAR finds applications in a wide range of domains, including autonomous navigation systems [1] to detect human behaviors and ensure safe operations. It is also crucial for a variety of other applications, including video retrieval [2], home monitoring, human-robot interaction [3], Human-Computer Interfaces (HCI) [4], healthcare by tracking elderly people sitting alone [5], [6], smart cities [7] and sports [8], [9]. HAR is frequently associated with the process of identifying and naming real-world human activities such as walking, sleeping, running, sitting, standing, showering, cooking, driving, opening the door, abnormal activities, and so on. [10], [11]. HAR and VAD are related fields that share commonalities in terms of analyzing and understanding activities within video data. However, they have distinct objectives and applications. HAR focuses on identifying, classifying, and understanding specific human activities or actions within a video stream or sensor data while VAD is concerned with identifying abnormal or unusual events or behaviors within a video stream.

The advancements in Computer Vision (CV) techniques and hardware accelerators enabled the processing of the massive amounts of data produced by live-stream cameras [12], [13]. As a result of the numerous applications that directly benefit from it, such as public security, monitoring workers' safety during working hours, healthcare systems for the elderly, and the need for Intelligent Video Surveillance Systems (IVSS), VAD has become an interesting field in CV. Because of the increasing demand for security and the growing number of surveillance cameras outdoors and indoors, IVSS have played an important role in the computer vision field in recent years. IVSS can detect anomalous actions like crimes, fights, traffic accidents, riots, kidnappings, and catastrophic events, as well as anomalous entities like weapons in critical locations and abandoned objects. However, there are several challenges to surveillance video analysis, one of which is detecting anomalous events, which requires extensive human effort and is time-consuming. As a result, relying solely on the human factor is insufficient, and IVSS was created to help in such situations.

Anomalies in video data are detected through the intricate analysis of sequential frames to identify unusual, unexpected, or abnormal occurrences. Datasets are critical in this domain because they serve as the foundation for developing, training, and evaluating algorithms designed for VAD. Several datasets are commonly used in this field (see section IV) to evaluate and benchmark the performance of anomaly detection algorithms. These datasets often contain video sequences capturing both normal and anomalous activities in various scenarios. Moreover, they cover a wide range of anomalies providing video anomaly detection algorithms with a diverse set of challenges. These datasets are frequently used by researchers and practitioners to assess the effectiveness and generalisation capabilities of anomaly detection models.

To our knowledge, this survey is the first to discuss a detailed overview of the most popular and latest VAD datasets, as well as a broad comparison between them such as the no. of videos, no. of frames, no. of anomalies, clip duration, and the resolution of each dataset. Moreover, an extensive overview of the challenges facing them.

This survey is structured in five sections as follows: section II will explore a brief overview of VAD. Section III will be dedicated to proposing the challenges that face the VAD domain. Section IV outlines the popular and recent datasets utilized in the research of VAD and their properties. 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.

II. Video Anomaly Detection

The word anomaly is defined as the odd or irregular patterns found in videos that do not conform to the normal trained patterns. According to [14], 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 [15], [16], image processing [17], [18], sensor networks [19], [20], medical health [21], [22], intrusion detection [23], IT security [24], [25], [26], and social media [27], [28]. Fig. 1 shows samples of abnormal frames in the UCF-Crime dataset [29], where (a) is an explosion, and (b) shows a man abusing a woman.

Fig. 1 Sample from the UCF-Crime dataset. (a) Explosion, (b) Abusing.

As shown in Fig. 2, the anomaly detection in videos passes through several steps. Firstly, a surveillance camera captures or records the video data, which is then segmented into several frames to determine any significant changes in the content. Following that, some pre-processing steps are carried out based on our requirements, such as noise removal, frame resizing, illumination adjustment, and so on. The next step entails feature extraction. The model is then developed, either for classification to determine whether the presented video is normal or for detecting the type of anomaly in the video, such as fighting, or robbery for example. Finally, a score is generated based on the model used to determine whether the video is normal or abnormal.

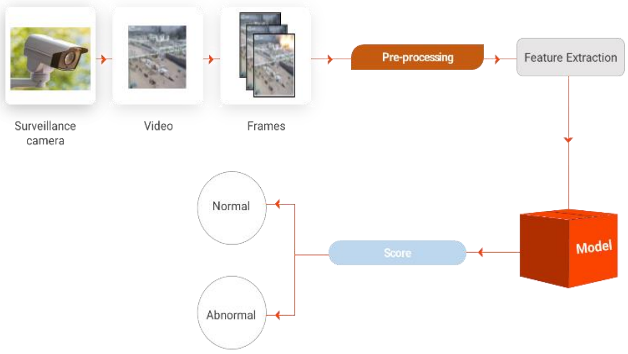


Fig. 2 Video Anomaly Detection Framework.

From the previous figure, data is the basic unit that will affect the performance and the accuracy of the model. As a result, the availability of datasets related to this domain is very important because it is so sensitive in some applications to get accurate results such as in detecting crimes or security issues.

III. VAD Datasets Challenges

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 datasets in VAD field:

- Exploring abnormality

  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, 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.

- Data Imbalance:
  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.

- Noise:
  Noise is considered an abnormality, so it is a big challenge to distinguish between it and the real abnormal events in the
  videos. Moreover, it will affect the actual accuracy of the model.
- Hardware requirements:
  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.

- Shortage of datasets:
  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:
  The efficiency of algorithms is affected by other external challenges 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.

IV. VAD Datasets.

Datasets for VAD detection vastly differ from each other and each one focuses on specific anomalies. They also differ in resolution and complexity. The main purpose of this paper is to introduce the most popular datasets and their features such as the number of videos, the number of frames, and the number of anomalies. Table 1 summarizes the datasets and their attributes. In addition to that, it presents how the dataset is labeled. The labeling procedure of these datasets affects how the loss functions are developed and the way we compare the efficiency of each architecture. Here some of the most popular datasets:

1. UMN Dataset:

UMN dataset is a dataset of panic scenarios taken from the web. It involves three different crowd scenes of 11 videos: one indoor and two outdoor. Normal crowd behavior is monitored until a predetermined moment in time, at which the behavior swiftly morphs into an escape scenario, in which each individual rushes out of view of the camera to mimic a panic moment. For example, people are walking and then suddenly run. The training set consists of the first 600 frames of each scene with a *320 x 240* frame resolution. Available at http://mha.cs.umn.edu/

2. UCSD Pedestrian Dataset:

The UCSD Pedestrian dataset is a small dataset of 98 videos [30], [31]. The dataset is acquired with a stationary camera that is mounted at an elevation that overlooks pedestrian walking with no changes in the illumination settings. The abnormal events are characterized by anything other than humans or sudden motion from pedestrians. UCSD is divided into two subsets according to different scenes: Ped 1 and Ped 2. Ped 1 has 80 videos, 34 for training and 36 for testing. Ped 2 has 26 videos, 16 for training and 14 for testing. Each video is 200 frames with a resolution of *158 X 238*. Many models are applied to it for anomaly detection [29], [32], [33], [34]. The crowd density in the scenes is one of the main challenges in this database. Available at http://www.svcl.ucsd.edu/projects/anomaly/dataset.html.

3. Avenue Dataset:

The Avenue dataset contains 37 videos in total which are taken from CUHK campus avenue [35]. It is split up into 16 training and 21 testing clips. The normal videos are of individuals walking between stairs and an entrance of a subway, while the abnormal clips are of people walking in an opposite direction, running, loitering, throwing, etc. All these videos were captured in one place. Some challenges are included in this dataset such as a few anomalies are contained in the training data and a slight camera shake. Available at http://www.cse.cuhk.edu.hk/leojia/projects/detectabnormal/dataset.html

4. ShanghaiTech Dataset:

W. Luo et al. [36] proposed a new dataset in their paper which was taken from 13 scenes in the Shanghai university campus (ShanghaiTech Dataset). It is a medium-scale database of 437 videos divided into 330 training videos and 107 testing videos having 130 abnormal events. The abnormal events include the occurrence of sudden movements in the data such as chasing and brawling. The challenges included are using different camera angles and complex light conditions. Available at https://svip-lab.github.io/dataset/campus_dataset.html.

5. Live Videos Dataset (LV dataset):

The LV dataset was introduced by [37] and includes 30 videos of real scenes taken by surveillance cameras. The captured scenes are from indoor and outdoor environments with camera motion and changes in the illumination which are difficult tasks. The anomalies are composed of 14 different events labeled as people fighting, people clashing, arm robberies, thefts, car accidents, hit and runs, fires, panic, vandalism, kidnapping, homicide, cars in the wrong way, people falling, loitering, prohibited U-turns and trespassing. All video sequences in the LV contain both normal and abnormal events and are not divided into training and testing sets as in the other datasets mentioned in this survey. The number of frames in the anomalous scenes is 68989 with variable resolutions varying between *176 x 144* and *1280 x 720*, but they are resized to a fixed size of *160 x 240*. Available at https://cvrleyva.wordpress.com/

6. UCF-Crime Dataset:

UCF-Crime is a large-scale dataset that has 1900 long real-world surveillance videos of 128 recorded hours [29]. Thirteen anomalies are included in it, which are explosion, fighting, robbery, shooting, abuse, arrest, arson, assault, road Accident, stealing, shoplifting, and vandalism. It can be used for two tasks:1) general AD by considering normal videos in a group and abnormal videos in another group, 2) recognizing only the type of the 13 different anomalies mentioned above. UCF-Crime is split into 1610 videos for training, where 800 are normal and 810 are abnormal videos, and 290 for testing, where 150 normal and 140 abnormal with a resolution of *240 x 320*. Video-level labels are only provided for the training videos and the temporal annotation is for the testing set. Many researchers prefer to make use of this dataset because the videos are from real-world scenes and the variety of anomalies it has [29], [38], [39], [40], [41]. The problem with this dataset is the varying duration of the clips and some clips have been repeated. Available at https://www.dropbox.com/sh/75v5ehq4cdg5g5g/AABvnJSwZI7zXb8_myBA0CLHa?dl=0

7. XD -Violence Dataset:

The XD-Violence is a large-scale and multi-scene audio-visual dataset of 4754 uncut videos collected from YouTube (in-the-wild scenes), sports streaming, surveillance cameras, and movies [42]. It is a dataset for violence detection. The dataset includes 2405 aggressive videos (abnormal) and 2349 non-aggressive videos (normal). The violent clips involve 6 violent types such as car accidents, abuse, explosion, riot, fighting, and shooting. The training set includes 3954 videos, and the testing set contains 800 videos. Video-level labels are obtained in the training set while frame-level labels are in the testing set. The good trait of this dataset is the multimodality information of both video and audio signals. Available at https://roc-ng.github.io/XD-Violence/

8. Extended UCF Crime Dataset:

The Extended UCF Crime dataset was proposed by [43]. They extended the UCF-Crime dataset by adding two extra anomaly classes to it, which are the protest and Molotov bomb classes. Furthermore, they added 33 videos to the fighting class, naming it UCF-Crime v2. The training set became 1826 videos after adding 216 videos and the testing set becomes 370 videos after adding 17 videos. In addition, the anomalies in the training videos are annotated in the temporal domain. Available at https://drive.google.com/file/d/1TnzMzk3TiHJHVsJmqQhzJXvNqml4MijB/view

9. Large-scale Anomaly Detection (LAD):

LAD was recently introduced and it contains 2000 video sequences which are considered the largest dataset available now of violence [44]. It consists of 14 different anomalies such as crash, crowd, destroy, drop, falling, fighting, fire, fall into the water, hurt, loitering, panic, thieving, trampling, and violence. For each anomaly class, there are more than 100 sequences are collected. Both the video-level labels and the frame-level labels are offered to assist in the detection of anomalies. The testing set is 560 videos with a resolution of *320 x 240* and the rest is for the training set. Available at https://drive.google.com/drive/folders/1WU2dld1rt5ajWaZqY3YLwLp-6USeQiVG

10. UBnormal:

This is a recently supervised dataset [45] of 268 training videos, 211 for testing, and 64 validation videos. These videos are annotated at both frame and pixel levels. It is synthetic data that was generated using the Cinema4D software. It introduces 22 abnormal events that were organized in a way that what in the testing data are different from those in the training and validation ones. Available at https://github.com/lilygeorgescu/UBnormal

11. Other Datasets:

Many other datasets were used in the VAD area. To overcome the lack of labeled data in the UCF-Crime dataset, [46] enriched a portion of the UCF-Crime with spatiotemporal annotations. They started by selecting six among the 13 anomalous labels presented in UCF-Crime, with a particular interest in human-based anomalies. After that, they selected 100 videos that belong to the designated categories, resulting in more than an hour of video sequences. That resulted in the creation of the UCFCrime2Local dataset. UBI-Fights was proposed by [47] which is a large-scale data of 1000 videos including 216 videos of fight events with frame-level annotation. Another novel dataset called Human Behavior Dataset 2021 (HBD21) [48] in which 456 videos are available in 4 categories: Assault violence, Gun violence, Sabotage violence, and Normal events. Street scene [49] is a dataset of 81 videos taken from a USB camera looking across a two-lane street, where the abnormal events include illegal activities such as prohibited U-turns and jaywalking. The Subway dataset [50] is collected from two scenes: an entrance and an exit of a subway. The Time-of-flight Indoor Monitoring (TIMO) dataset [51] was recently developed. It contains indoor spaces taken using a time-of-flight (ToF) camera [52] from two different views: top-down or tilted perspective. Hockey fight detection is a sports dataset of 1000 videos collected from hockey games that detect fighting that occurs during playing the game [53]. Moreover, some data sets are not of real scenes and

Table 1: Comparison of various datasets for human anomaly detection in videos

Dataset		Year	# Videos	# Frames	Resolution	Supervision	Scenes	# Of Anomaly types	Clip duration	Frame per sec (fps)
UMN [56]		2006	11	7700	320×240	Video-level	3	1	-	30
Subway [50]	Entrance	2008	1	72,401	512 x 384	Video-level	1	5	2 hr.	-
	Exit		1	136,524			1	3		
UCSD [30], [31]	Ped 1	2010	80	14,000	158 x 238	Video-level	1	5		
	Ped 2		26	4,560	240 x 360		1	5	-	10
Avenue [35]		2013	37	30652	640 x 360	Video-level	1	3	1-2 min	-
ShanghaiTech [36]		2017	437	317,398	846 x480	Video-level	13	130	-	-
LV [37]		2017	30	-	176 x 144 1280 x 720	Video-level	30	17	3.93 hrs.	7.5-30
UCF-Crime [29]		2018	1900	13M	240 x 320	Video-level	20	13	128 hrs. (total)	30
UCFCrime2Local [46]		2019	300	-	240 x 320	Video-level and Frame level	-	6	>1 hour	
XD-Violence [42]		2020	4754	-	160 x 120	Video-level and Frame level	Multiple scenes	6	217 hr. (total)	24
Street scene [49]		2020	81	203,257	1280 x 720	Video-level and Frame level	1	17	-	15
LAD [44]		2021	2000	-	320 x240	Video-level and Frame level	1895	14	-	25
Extended UCF Crime [43]		2021	2133	-	240 x 320	Video-level and Frame level	15	15	-	30
TIMO [51]		2021	1588	612,000	512 x 512 288 x320	Frame- level	2	-	-	30
UBI-Fights [47]		2021	1000	-	640 x 360	Frame level	Multiple scenes	1	80 hrs. (total)	30
UBnormal [45]		2022	543	236,902	224 x 224	Video-level and Frame level	29	22	2.2 hrs (total)	30

are performed by volunteers or actors such as the CASIA dataset [54] and IXMAS dataset [55].

It's obvious that these datasets require a specific evaluation criterion to benchmark different models. Most of them suffer from sparsity and they have different labeling techniques which would affect the choice of the loss function. They require dedicated resources to process the large amounts of the recorded hours.

V. CONCLUSION

Datasets in VAD are essential resources for the development, training, and evaluation of algorithms capable of distinguishing between normal and abnormal activity in video streams. These datasets drive innovation, benchmarking, and advancements in the field, making a significant contribution to the development of reliable and efficient anomaly detection systems in applications such as security, surveillance, and safety monitoring. This survey provides an in-depth examination of the most recent and widely used datasets for detecting abnormal human behaviour. The datasets clearly suffer from a lack of annotations, and each one focuses on a different type of anomaly. Since the UCF-Crime and LAD datasets are the only ones with large-scale and different types of anomalies, novel datasets with different types of anomalies should be created to cover all possible scenarios. Moreover, more real-world datasets should be proposed to produce real-world applications.

REFERENCES

[1] M. Lu, Y. Hu, and X. Lu, 'Driver action recognition using deformable and dilated faster R-CNN with optimized region proposals', *Applied Intelligence*, vol. 50, no. 4, pp. 1100–1111, Apr. 2020, doi: 10.1007/s10489-019-01603-4.

[2] F. Aslam, F. Hussain, and M. H. Yousaf, 'Human Activity Based Video Retrieval Using Optical Flow and Local Binary Patterns.', *NED University Journal of Research*, vol. 15, no. 4, pp. 93–105, 2018, [Online]. Available: https://search.ebscohost.com/login.aspx?direct=true&db=asn&AN=136432390&site=ehost-live

[3] I. Rodomagoulakis *et al.*, 'Multimodal human action recognition in assistive human-robot interaction', in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Mar. 2016, pp. 2702–2706. doi: 10.1109/ICASSP.2016.7472168.

[4] A. Haria, A. Subramanian, N. Asokkumar, S. Poddar, and J. S. Nayak, 'Hand Gesture Recognition for Human Computer Interaction', *Procedia Comput Sci*, vol. 115, pp. 367–374, 2017, doi: https://doi.org/10.1016/j.procs.2017.09.092.

[5] L. Schrader *et al.*, 'Advanced Sensing and Human Activity Recognition in Early Intervention and Rehabilitation of Elderly People', *J Popul Ageing*, vol. 13, no. 2, pp. 139–165, Jun. 2020, doi: 10.1007/s12062-020-09260-z.

[6] Z. Wang, Z. Yang, and T. Dong, '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*, vol. 17, no. 2, p. 341, Feb. 2017, doi: 10.3390/s17020341.

[7] A. Almeida and G. Azkune, 'Activity recognition approaches for smart cities: The City4Age use case', in *2017 IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI)*, IEEE, Sep. 2017, pp. 1–5. doi: 10.1109/RTSI.2017.8065909.

[8] M. Stein *et al.*, 'From game events to team tactics: Visual analysis of dangerous situations in multi-match data', in *2016 1st International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW)*, IEEE, Dec. 2016, pp. 1–9. doi: 10.1109/TISHW.2016.7847777.

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

[10] R. Bhardwaj and P. K. Singh, 'Analytical review on human activity recognition in video', in *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*, IEEE, Jan. 2016, pp. 531–536. doi: 10.1109/CONFLUENCE.2016.7508177.

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

[12] S. Li, L. Da Xu, and S. Zhao, 'The internet of things: a survey', *Information Systems Frontiers*, vol. 17, no. 2, pp. 243–259, Apr. 2015, doi: 10.1007/s10796-014-9492-7.

[13] C. Jobanputra, J. Bavishi, and N. Doshi, 'Human Activity Recognition: A Survey', *Procedia Comput Sci*, vol. 155, pp. 698–703, 2019, doi: 10.1016/j.procs.2019.08.100.

[14] B. Steenwinckel, 'Adaptive Anomaly Detection and Root Cause Analysis by Fusing Semantics and Machine Learning', 2018, pp. 272–282. doi: 10.1007/978-3-319-98192-5_46.

[15] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, 'Credit card fraud detection using machine learning techniques: A comparative analysis', in *2017 International Conference on Computing Networking and Informatics (ICCNI)*, IEEE, Oct. 2017, pp. 1–9. doi: 10.1109/ICCNI.2017.8123782.

[16] A. Pumsirirat and L. Yan, 'Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine', *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 1, pp. 18–25, 2018, doi: 10.14569/IJACSA.2018.090103.

[17] M. S. Minhas and J. Zelek, ‘Anomaly Detection in Images’, *ArXiv*, May 2019.

[18] L. Beggel, M. Pfeiffer, and B. Bischl, ‘Robust Anomaly Detection in Images Using Adversarial Autoencoders’, 2020, pp. 206–222. doi: 10.1007/978-3-030-46150-8_13.

[19] S. Rajasegarar, C. Leckie, J. C. Bezdek, and M. Palaniswami, ‘Centered Hyperspherical and Hyperellipsoidal One-Class Support Vector Machines for Anomaly Detection in Sensor Networks’, *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 518–533, Sep. 2010, doi: 10.1109/TIFS.2010.2051543.

[20] F. Cauteruccio *et al.*, ‘Short-Long Term Anomaly Detection in Wireless Sensor Networks based on Machine Learning and Multi-Parameterized Edit Distance’, *Information Fusion*, vol. 52, 2018, doi: 10.1016/j.inffus.2018.11.010.

[21] Q. Wei, B. Shi, J. Y. Lo, L. Carin, Y. Ren, and R. Hou, ‘Anomaly detection for medical images based on a one-class classification’, in *Medical Imaging 2018: Computer-Aided Diagnosis*, K. Mori and N. Petrick, Eds., SPIE, Feb. 2018, p. 57. doi: 10.1117/12.2293408.

[22] T. Fernando *et al.*, ‘Neural memory plasticity for medical anomaly detection’, *Neural Networks*, vol. 127, pp. 67–81, Jul. 2020, doi: 10.1016/j.neunet.2020.04.011.

[23] S. Jose, D. Malathi, B. Reddy, and D. Jayaseeli, ‘A Survey on Anomaly Based Host Intrusion Detection System’, *J Phys Conf Ser*, vol. 1000, no. 1, p. 012049, Apr. 2018, doi: 10.1088/1742-6596/1000/1/012049.

[24] V. H. Bezerra, V. G. T. da Costa, S. Barbon Junior, R. S. Miani, and B. B. Zarpelão, ‘IoTDS: A One-Class Classification Approach to Detect Botnets in Internet of Things Devices’, *Sensors (Basel)*, vol. 19, no. 14, p. 3188, Jul. 2019, doi: 10.3390/s19143188.

[25] M. Zalasiński, K. Łapa, and M. Laskowska, ‘Intelligent Approach to the Prediction of Changes in Biometric Attributes’, *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 6, pp. 1073–1083, 2020, doi: 10.1109/TFUZZ.2019.2955043.

[26] K. Demertzis, L. Iliadis, and I. Bougoudis, ‘Gryphon: a semi-supervised anomaly detection system based on one-class evolving spiking neural network’, *Neural Comput Appl*, vol. 32, no. 9, pp. 4303–4314, May 2020, doi: 10.1007/s00521-019-04363-x.

[27] S. Shah and M. Goyal, ‘Anomaly Detection in Social Media Using Recurrent Neural Network’, in *International Conference on Computational Science*, Springer, 2019, pp. 74–83. doi: 10.1007/978-3-030-22747-0_6.

[28] M. HC and M. R, ‘BMADSN: Big data multi-community anomaly detection in social networks’, *The International Journal of Electrical Engineering & Education*, Dec. 2019, doi: 10.1177/0020720919891065.

[29] 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*, IEEE, Jun. 2018, pp. 6479–6488. doi: 10.1109/CVPR.2018.00678.

[30] 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*, IEEE, Jun. 2010, pp. 1975–1981. doi: 10.1109/CVPR.2010.5539872.

[31] Mahadevan, Vijay, ‘UCSD Dataset’. [Online]. Available: http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm

[32] Y. Lu, F. Yu, M. K. K. Reddy, and Y. Wang, ‘Few-Shot Scene-Adaptive Anomaly Detection’, in *European Conference on Computer Vision*, Springer, 2020, pp. 125–141. doi: 10.1007/978-3-030-58558-7_8.

[33] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni, and N. Sebe, ‘Abnormal event detection in videos using generative adversarial nets’, in *2017 IEEE International Conference on Image Processing (ICIP)*, IEEE, Sep. 2017, pp. 1577–1581. doi: 10.1109/ICIP.2017.8296547.

[34] M. Sabokrou, M. Khalooei, M. Fathy, and E. Adeli, ‘Adversarially Learned One-Class Classifier for Novelty Detection’, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3379–3388, 2018.

[35] C. Lu, J. Shi, and J. Jia, ‘Abnormal Event Detection at 150 FPS in MATLAB’, in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 2720–2727. doi: 10.1109/ICCV.2013.338.

[36] W. Luo, W. Liu, and S. Gao, ‘A Revisit of Sparse Coding Based Anomaly Detection in Stacked RNN Framework’, in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2017, pp. 341–349. doi: 10.1109/ICCV.2017.45.

[37] R. Leyva, V. Sanchez, and C.-T. Li, ‘The LV dataset: A realistic surveillance video dataset for abnormal event detection’, in *2017 5th International Workshop on Biometrics and Forensics (IWBF)*, IEEE, Apr. 2017, pp. 1–6. doi: 10.1109/IWBF.2017.7935096.

[38] Y. Tian, G. Pang, Y. Chen, R. Singh, J. W. Verjans, and G. Carneiro, ‘Weakly-supervised Video Anomaly Detection with Robust Temporal Feature Magnitude Learning’, *arXiv preprint arXiv:2101.10030*, 2021.

[39] A. M. Kamoona, A. K. Gosta, A. Bab-Hadiashar, and R. Hoseinnezhad, ‘Multiple Instance-Based Video Anomaly Detection using Deep Temporal Encoding-Decoding’, *arXiv preprint arXiv:2007.01548*, 2020.

[40] J. Zhang, L. Qing, and J. Miao, ‘Temporal Convolutional Network with Complementary Inner Bag Loss for Weakly Supervised Anomaly Detection’, in *2019 IEEE International Conference on Image Processing (ICIP)*, IEEE, Sep. 2019, pp. 4030–4034. doi: 10.1109/ICIP.2019.8803657.

[41] R. Xue, J. Chen, and Y. Fang, ‘Real-Time Anomaly Detection and Feature Analysis Based on Time Series for Surveillance Video’, in *2020 5th International Conference on Universal Village (UV)*, 2020, pp. 1–7. doi: 10.1109/UV50937.2020.9426191.

[42] P. Wu *et al.*, ‘Not only Look, But Also Listen: Learning Multimodal Violence Detection Under Weak Supervision’, 2020, pp. 322–339. doi: 10.1007/978-3-030-58577-8_20.

[43] H. İ. Öztürk and A. B. Can, ‘ADNet: Temporal Anomaly Detection in Surveillance Videos’, *ArXiv*, Apr. 2021.

[44] B. Wan, W. Jiang, Y. Fang, Z. Luo, and G. Ding, ‘Anomaly Detection in Video Sequences: A Benchmark and Computational Model’. 2021.

[45] A. Acsintoae *et al.*, ‘UBnormal: New Benchmark for Supervised Open-Set Video Anomaly Detection’, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 20143–20153.

[46] F. Landi, C. G. M. Snoek, and R. Cucchiara, ‘Anomaly Locality in Video Surveillance’, *ArXiv*, vol. abs/1901.1, 2019.

[47] B. Degardin and H. Proença, ‘Iterative weak/self-supervised classification framework for abnormal events detection’, *Pattern Recognit Lett*, vol. 145, pp. 50–57, May 2021, doi: 10.1016/j.patrec.2021.01.031.

[48] M. Jayaswal, R., Dixit, ‘A Framework for Anomaly Classification Using Deep Transfer Learning Approach’, *Revue d'Intelligence Artificielle*, vol. 35, no. 3, pp. 255–263, 2021, doi: doi.org/10.18280/ria.350309.

[49] B. Ramachandra and M. J. Jones, ‘Street Scene: A new dataset and evaluation protocol for video anomaly detection’, in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, Mar. 2020, pp. 2558–2567. doi: 10.1109/WACV45572.2020.9093457.

[50] 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, Mar. 2008, doi: 10.1109/TPAMI.2007.70825.

[51] P. Schneider *et al.*, ‘TIMo - A Dataset for Indoor Building Monitoring with a Time-of-Flight Camera’, *ArXiv*, vol. abs/2108.1, 2021.

[52] R. Horaud, M. Hansard, G. Evangelidis, and M. Clément, ‘An Overview of Depth Cameras and Range Scanners Based on Time-of-Flight Technologies’, *Mach Vis Appl*, vol. 27, 2016, doi: 10.1007/s00138-016-0784-4.

[53] E. B. Nievas, O. Déniz-Suárez, G. B. García, and R. Sukthankar, ‘Violence Detection in Video Using Computer Vision Techniques’, in *CAIP*, 2011.

[54] Y. Wang, K. Huang, and T. Tan, ‘Human Activity Recognition Based on R Transform’, in *2007 IEEE*

Conference on Computer Vision and Pattern Recognition, IEEE, Jun. 2007, pp. 1–8. doi: 10.1109/CVPR.2007.383505.

[55] D. Weinland, R. Ronfard, and E. Boyer, 'Free viewpoint action recognition using motion history volumes', *Comput. Vis. Image Underst.*, vol. 104. 2006.

[56] M. University, 'UMN Dataset'. [Online]. Available: http://mha.cs.umn.edu/proj_events.shtml