Informatics Bulletin, Faculty of Computers and Artificial Intelligence, Helwan University
Published Online ***Vol 3 Issue 1, January 2021***
(https://fcihib.journals.ekb.eg)

A Review on Anomalous Events Detection and Recognition

Manar M. F. Donia
Computer Science Department,
Faculty of Computers and Artificial Intelligence, Helwan university, Cairo, Egypt

Wessam M. H. El-Behaidy
Computer Science Department,
Faculty of Computers and Artificial Intelligence, Helwan university, Cairo, Egypt

Aliaa A. A. Youssif
Computer Science Department,
Faculty of Computers and Artificial Intelligence, Helwan university, Cairo, Egypt

***Abstract*—Anomalous behaviour may indicate threats and dangers to others. We can define anomalous event as what drift from what is anticipated, common, or natural action. Which make video surveillance a key to increase public security. The main goal of event detection is to detect the occurrences of events and categorize them into normal or abnormal actions. Such detection requires detecting and tracking objects then recognize what is happening around those tracked objects. Recently, researches depend on one of two techniques: handcrafted features and deep learning models. The handcrafted features depend on extracting low-level features, its strength depend on choosing the best features, which gives the best results. After the successful of deep learning techniques for image classification, researchers have examined the deep learning techniques ability for detection, which skip the manual step of features extraction and deals directly with the images. This paper presents a survey on both handcrafted and deep learning models for abnormal events detection.**

Keywords: Abnormal events, Event analysis, Human computer interaction, Deep learning, Handcrafted features.

1. INTRODUCTION

The video surveillance is a key to increase public security. The main goal of event detection is to detect the occurrences of events and categorize them into normal or anomalous event. To do so we need to recognize what is happening around those tracked objects. In order to decide whether it's normal or abnormal event that requires actions. Abnormal behaviour may indicate threats and dangers to others. Finding such abnormalities in videos is crucial for applications ranging from automatic quality control to visual surveillance scenarios like in prisons, schools, blocking of inappropriate violence in movies watched by children. Most of the work in event analysis based on two main paths [1]: the first walking or jogging of video sequences whereas the other path is anomalous detection that focused on detecting the rare or unusual event such as aggressive behaviours. Anomalous detection can further classified into [2] local anomalous and global anomalous detection. Local anomalous is the behaviour of an individual such as driving a car in opposite direction. Whereas, global anomalous refers to the behaviour of more than one individual such as most of the people run in different directions. Different approaches proposed to detect human activity in a scene [3]. Those approaches can be fallen into two main types as shown in figure 1 that can detect normal and anomalous events in crowded or uncrowded scenes. First type is the handcrafted features based approach that depends essentially on extracting set of features like motion or texture, which make it more appropriate for crowded scenes. While the other type is deep learning based approach that depend on tracking objects of interest to produce a trajectory, any trajectory different from learnt trajectories treated as anomalous. Deep learning based approach is a black box we have no explanation about how a model works. If it counters a problem, it will be difficult to debug. The handcrafted feature based approaches are depending on choosing the best features which make it suitable for crowed or uncrowded scenes. But its main obstacle is the wide-ranging features of actions in traditional approaches. The rest of the paper organized as follows. Methodologies covered in section 2. The classification methods categorized in one is explicit event recognition, which focused mostly on recognizing human activity like clapping, section 3. Well-known dataset used presented in section 4. Finally conclusion presented in section 5.

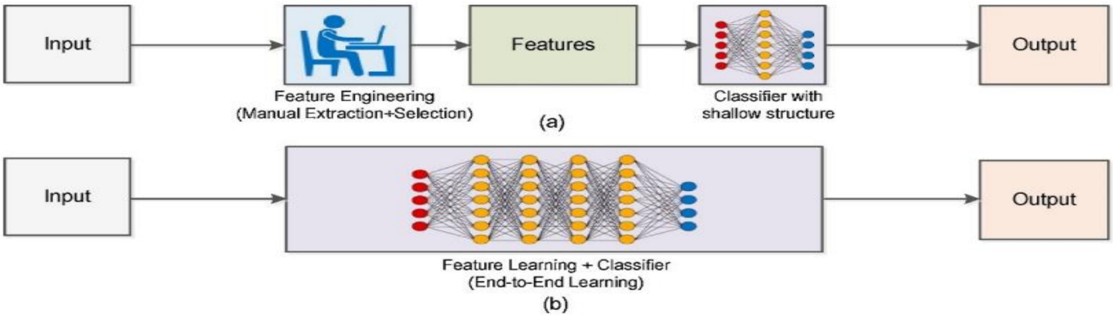

Fig. 1. a) Handcrafted features based model. b) Deep learning based model. From[43]

2. METHODOLOGIES

Currently there are various researches interesting in abnormal event detection, which can detect anomalous events also categorizing them into local or global behaviours. Recent researches fall into two main categories: handcrafted features based models, and deep learning models.

2.1. HANDCRAFTED BASED MODELS

The feature learning techniques used every day in how we teach a child to recognize different objects. Handcrafted features based approach depends essentially on the extraction of feature in the spatio-temporal domain. Where a set of low-level features combined in what called a bag-of-words that can used in supervised or unsupervised type learning. Its strength depends on choosing the features in a way that gives the best results. Then classification performed by training a machine learning classifier. Deepak et al [4] detect and recognize violent actions in crowded scenes. By adopting (STACOG) spatio-temporal autocorrelation of gradient as feature extraction technique to extract motion information from video frames. It extracts local relationships among space–time gradient then pass it as input to SVM classifier for violence and non-violence activities recognition. They evaluate the methodology on two benchmark datasets: The crowd violence dataset and The hockey fight dataset for violence detection. Their proposed methodology achieved 91.3% recognition rate over "the crowd hockey fight dataset". Yang et al [5] detect abnormal vehicle behavior in complex surveillance scenes. They recognize abnormal behaviour by decompose complex global behaviours according to the spatio-temporal characteristics of behaviour. The cascaded topic model methodology proposed, depend on learning behavioural spatio-temporal context. It identifies local behavioral pattern by modelling the relation between multiple trajectory segments of the same moving object using MRF-LDA (Markov random fields - latent dirichlet allocation) then cluster local behavioural pattern topics into categories using spectral clustering algorithm. After that, LDA topic model used to learn the temporal context of global behaviour. The abnormal behaviour recognition method developed based on the learned spatio-temporal context of behaviours but in a top-down strategy. It firstly detects abnormal video clips by training LDA topic models independently for each type of non-overlapping video clips that checked for abnormal behaviours. An abnormality scoring function (*abf*) calculated depending on likelihood value of anomalous behaviour in the video clip if abf is lower than a threshold value the video clip considered to contain an abnormal behaviour. Once video clip recognized as abnormal a local behaviour topic categories is determined then locate anomalous moving object trajectories. The cascaded topic model methodology evaluated on QMUL Street Intersection Dataset then compared its performance along with two different methods single layer" LDA violence dataset" while achieved 90.4% on "The

highest performance among them while single layer " LDA model " has the lowest performance. Das & Mishra [6] detect anomalous trajectory in crowded scenes by clustering trajectories based on several independent features (density, shape, mean position and standard deviation) in order to detect the overall anomalous as an anomalous trajectory may be similar with a normal pattern in one aspect, but differs significantly in other aspects. In classification, they depend on the shannon entropy where the anomalous trajectory would have higher levels of entropy when compared to normal trajectories. So The entropy of the probability distribution evaluated and if it exceeds a threshold, it classified as an anomalous. The method was tested on two datasets crowded scenes dataset and UCF crowd dataset showing an accuracy of 98% detection rate. Zhu et al. [7] detect local abnormal behaviour such as walking in the wrong direction in crowded scene. Based on the idea of optical flow energy of an abnormal behaviour is larger than of a normal one. They describe the behaviour feature of the whole video sequence based on a histogram of categories of optical flow features extracted from image. Then consider local behaviour as anomalous if the log-likelihood computed using mixed naive bayes model is less than a predefined threshold value. The methodology compared with three approaches: social force-based method, mixed dynamic texture based method and adaptive optical flow filtering based method. These methods are applied on two benchmark datasets: The UCSD dataset and the subway dataset where its results was superior to other three approaches in detecting fast abnormal behaviour but skip the slow one. Handcrafted features based model [8] [13] [14] is remarkable, as the amount of training data is not a factor compared to deep learning model. Also produce the opportunity to visualize and analyse features in order to choose the best features largely contribute to the results. But its main obstacle is the wide-ranging features of

model" and two level hierarchal "Cas-LDA model" as it achieved the

researchers to explore other techniques in order to have a better detection.

2.2. DEEP LEARNING BASED MODELS

After the successful of deep learning techniques in image classification, researchers have examined its ability for anomalous detection. Deep learning based approaches [15] [18]; need not human obtrusion as it skip the manual step of features extraction and deals directly with raw data. The amount of learning depends on the quality of the data, which affect result quality. If the data quality is not good enough its performance will affect. Pang et al. [24] proposed a self-training deep neural network in order to establish an end-to-end deep video anomaly detection approach. By performing initial detection using Sp + iForest to generate anomaly scores, which used to separate the frames of unlabelled videos into anomaly set or normal set. These sets then used to train a ResNet-50 model and a fully connected network in order to optimize the anomaly scores. Then a self-training used to improve the anomaly score through re-computing the anomaly scores of all frames, which yield to have better anomaly scores than the initial anomaly scores. The methodology was evaluated on three benchmark datasets UCSD, subway and UMN. It is compared its performance along with unsupervised methods based "Sp + iForest" where it achieves about 2%-15% improvement than others. There are two types of abnormalities in the surveillance systems. I) Appearance anomaly of objects that is visually different from normal ones which needs to identify unusual samples in data. II) Motion anomaly refers to an unusual motion of normal appearance object. Ruff et al [25] they proposed an end-to-end deep method for appearance anomaly detection in images by establishing a semi-supervised deep support vector data descriptor (SS-DSVDD). They extended boundaries around data in a way that enables detection of unusual samples. The proposed model trajectories in traditional approaches which made

both have ten classes. They set one of the ten classes to be the normal class and let the remaining nine classes represent anomalies. The proposed method performance compared along with state of art methods where it on par other methods in detection.

The end-to-end-deep learning is a black box due to their multilayer structure, their predictions not easy to traceable by human. If it counter a problem will be difficult to solve. Therefore, the researchers try to combine both deep leaning with traditional methods [26] [28] [30]. Anitha & Arun [36] proposed unsupervised deep learning method for video anomaly detection where both training and testing are unsupervised. Spatial feature consists of a combination of original frames and edge frames of unlabelled video given as input to the deep learning model. The deep learning model consists of convolutional auto encoder and convolutional LSTM for both spatial and temporal features learning then reconstructed frames given as output. Abnormal event detected based on the reconstruction error calculated from the euclidean distance between the original frame and the reconstructed frame. Methodology was evaluated on two benchmark datasets: avenue and UCSD. It achieves 90.7% accuracy in avenue dataset, while 98.4% accuracy in UCSD datasets. Kamoona et al [37] detect anomaly events in real world surveillance video by computing the sparsity information of C3D features (3D convolution features). Then using naïve bayes for classification by computing the likelihood of normal class only after modifying it based on sparsity estimation of normal video bag and consider those less than a user defined threshold as anomalous. They evaluate the methodology on UCF dataset, which is real-world surveillance video and compare its performance along with standard naïve bayes method, passion naïve bayes and BOVW+SVM. Their methodology along with BOVW+SVM achieved a high close performance while the passion naïve bayes has the lowest performance. Combining

trained on MNIST and CIFAR-10 datasets which

ones give strong baseline by extracting distinguished features help improve the results.

3. CLASSIFICATION

The discussed methodologies can be classified into two categories. I) supervised models as shown in table 1 that trained on labelled data in order to adjust its precision. II) Unsupervised models as shown in table 2, it doesn't need labelled data as the model tries to extract pattern by its own. Different classifiers have been used in either supervised or unsupervised. According to the reviewed papers, SVM was the most used supervised classifier [4] [11] [12]; while CNN was the most used unsupervised classifier [17] [24] [26]. In both categories different features have been used for anomalous representation. Features are classified into I) motion information [8] [13] [16]. II) Appearance information [3] [9] [36]. III) Both motion and appearance information [14] [20] [37].

4. DATASETS

In this section frequently public datasets used for anomalous detection are discussed. Those datasets can be categorized according to the videos content, into violence dataset [38] and interaction datasets [39-41]. Violence datasets containing videos of violent action. The most well-known violence dataset are hockey-fight dataset and movie-fight dataset. The first dataset contains 1000 clips of action that labelled as fight or non-fight. While the other dataset contains 200 clips from action movies that captured at variety of scenes and different resolutions. Interaction datasets containing videos of abnormal interaction between groups of people. The most well-known interaction datasets are I) UCSD dataset that consists of 78 video samples which split into 2 subsets. Video samples containing normal event of pedestrian walk and abnormal events used for training and testing. Abnormal event due to either anomalous pedestrian motion patterns or non-pedestrian entities in the walkways.

both deep learning based approaches with traditional

Table 1: supervised anomaly detection

Model	Method	Scene
Handcrafted based models	Violence detection based on autocorrelation of gradients [4]	Crowded
	Local abnormal behavior detection based on optical flow[7]	Crowded
	Global abnormal event detection based on covariance matrix for optical flow [8]	Crowded
	Abnormal event detection based on Contextual information using histogram of oriented gradient[9]	Crowded
	Abnormal event detection based on appearance and motion information [10]	Crowded
	Violence detection based on motion weber local descriptor (MoWLD)[11]	Crowded
	Violence detection based on spatio-temporal interest point [12]	Crowded and uncrowded
	Anomaly detection based on spatially localized histogram of optical flow (SL-HOF) descriptor [13]	Crowded
Deep learning based models	Abnormal event detection based on generative adversarial nets (GANs) [15]	Crowded
	Abnormal event detection based on deep one class convolutional neural network [16]	Crowded
	Anomaly detection based on gaussian mixture variational autoencoder[20]	Crowded
	Anomaly detection based on sparsity information of C3D deep features[37]	Uncrowded

Table 2: unsupervised anomaly detection

Model	Method	Scene
Handcrafted based Models	Local and global anomalies detection via hierarchical feature representation and gaussian process regression (GPR)[3]	Crowded and uncrowded
	Abnormal behaviors detection in complex scenes based on spatio-temporal context[5]	Street Intersection
	Anomalous trajectories detection using multi-object tracker and then cluster them based on multiple independent features[6]	Crowded
	Abnormal behaviors localization and detection based on optical flow and histogram based descriptor[14]	Crowded
Deep learning based Models	Abnormal events detection based on joint representation learning of appearance and motion[17]	Crowded
	Anomaly detection via deep predictive coding network[18]	Crowded
	Anomaly detection based on end-to-end neural network[24]	Crowded
	Local anomalies detection by combining semantic information with optical flow[26]	Crowded
	Anomaly detection based on hybrid deep learning framework[36]	Crowded

II) Avenue dataset contains 37 video clips for training and testing in total of 30652 frames. III) Subway dataset contains videos in a total of 64902 frames of different types of anomalous events.

5. CONCLUSION

In this review, we provide a survey of state-of-the-art anomalous detection approaches including both handcrafted and deep learning based approaches. Anomalous detection approaches consists of many steps (pre-processing, feature extraction, classification, decision making). Each step is charge of a specific action that will be responsible on the whole system results. opportunity to train model on different classifiers. Also, they choose and know the best feature to extract that will produce reliable results, but its main obstacle is the wide-ranging features of trajectories in traditional approaches. Although pure deep learning is more suitable for real time anomalous detection but it may not give effective results as compared to other techniques. Deep learning is a black box due to their multilayer structure. We cannot provide an explanation about how a model works and if it counters a problem, it will be difficult to debug. The combination of both handcrafted feature and feature learning give a strong baseline by extracting more distinguished

Handcrafted features based approaches gives the features not only help improve the results but also overcome the problem of no exact definition of abnormal events. The amount of learning depends on the quality of the data, if the amount and quality of data is not good enough the performance will be affected. Till now the researchers trying to enhance anomalous detection throughout trying different adjustments in each approach "handcrafted, pure deep learning or the combination of both approaches" in order to find the optimum result.

REFERENCES

[1] Aggarwal, J. K., & Ryoo, M. S. (2011). Human activity analysis: A review. ACM Computing Surveys (CSUR), 43(3), 1-43.

[2] Cheng, K. W., Chen, Y. T., & Fang, W. H. (2015). Gaussian process regression-based video anomaly detection and localization with hierarchical feature representation. IEEE Transactions on Image Processing, 24(12), 5288-5301.

[3] Sargano, A. B., Angelov, P., & Habib, Z. (2017). A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition. applied sciences, 7(1), 110.

[4] Deepak, K., Vignesh, L. K. P., & Chandrakala, S. (2020). Autocorrelation of gradients based violence detection in surveillance videos. ICT Express.

[5] Yang, Y., Li, L., Liu, Z., & Liu, G. (2020). Abnormal behavior recognition based on spatio-temporal context. Journal of Information Processing Systems, 16(3), 612-628.

[6] Das, D., & Mishra, D. (2018). Unsupervised Anomalous Trajectory Detection for Crowded Scenes. In 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS) (pp. 27-31). IEEE.

[7] Zhu, S., Hu, J., & Shi, Z. (2016). Local abnormal behavior detection based on optical flow and spatio-temporal gradient. Multimedia Tools and Applications, 75(15), 9445-9459.

[8] Wang, T., Qiao, M., Zhu, A., Niu, Y., Li, C., & Snoussi, H. (2018). Abnormal event detection via covariance matrix for optical flow based feature. Multimedia Tools and Applications, 77(13), 17375-17395.

[9] Hu, X., Huang, Y., Duan, Q., Ci, W., Dai, J., & Yang, H. (2018). Abnormal event detection in crowded scenes using histogram of oriented contextual gradient descriptor. EURASIP Journal on Advances in Signal Processing, 2018(1), 54.

[10] Amraee, S., Vafaei, A., Jamshidi, K., & Adibi, P. (2018). Abnormal event detection in crowded scenes using one-class SVM. Signal, Image and Video Processing, 12(6), 1115-1123

[11] Zhang, T., Jia, W., Yang, B., Yang, J., He, X., & Zheng, Z. (2017). MoWLD: a robust motion image descriptor for violence detection. Multimedia Tools andApplications, 76(1), 1419-1438

[12] Mabrouk, A. B., & Zagrouba, E. (2017). Spatio-temporal feature using optical flow based distribution for violence detection. Pattern Recognition Letters, 92, 62-67.

[13] Wang, S., Zhu, E., Yin, J., & Porikli, F. (2016). Anomaly detection in crowded scenes by SL-HOF descriptor and foreground classification. In 2016 23rd International Conference on Pattern Recognition (ICPR) (pp. 3398-3403). IEEE

[14] Zhang, X., Yang, S., Zhang, J., & Zhang, W. (2020). Video Anomaly Detection and Localization using Motion-field Shape Description and Homogeneity Testing. Pattern Recognition, 107394.

[15] Ravanbakhsh, M., Sangineto, E., Nabi, M., & Sebe, N. (2019, January). Training adversarial discriminators for cross-channel abnormal event detection in crowds. IEEE Winter Conference on Applications of Computer Vision (WACV), (pp. 1896-1904).

[16] Sun, J., Shao, J., & He, C. (2019). Abnormal event detection for video surveillance using deep one-class learning. Multimedia Tools and Applications, 78(3), 3633-3647.

[17] Yu, J., Yow, K. C., & Jeon, M. (2018). Joint representation learning of appearance and motion for abnormal event detection. Machine Vision and Applications, 29(7), 1157-1170.

[18] Ye, M., Peng, X., Gan, W., Wu, W., & Qiao, Y. (2019). Anopcn: Video anomaly detection via deep predictive coding network. In Proceedings of the 27th ACM International Conference on Multimedia (pp. 1805-1813).

[19] Sabokrou, M., Fathy, M., Zhao, G., & Adeli, E. (2020). Deep End-to-End One-Class Classifier. IEEE Transactions on Neural Networks and Learning Systems.

[20] Fan, Y., Wen, G., Li, D., Qiu, S., Levine, M. D., & Xiao, F. (2020). Video anomaly detection and localization via Gaussian mixture fully convolutional variational autoencoder. Computer Vision and Image Understanding, 102920.

[21] Medel, J. R., & Savakis, A. (2016). Anomaly detection in video using predictive convolutional long short-term memory networks. arXiv preprint arXiv:1612.00390.

[22] Sabokrou, M., Fayyaz, M., Fathy, M., Moayed, Z., & Klette, R. (2018). Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. Computer Vision and Image Understanding, 172, 88-97.

[23] Nawaratne, R., Alahakoon, D., De Silva, D., & Yu, X. (2019). Spatiotemporal anomaly detection using deep learning for real-time video surveillance. IEEE Transactions

on Industrial Informatics, 16(1), 393-402.

[24] Pang, G., Yan, C., Shen, C., Hengel, A. V. D., & Bai, X. (2020). Self-trained Deep Ordinal Regression for End-to-End Video Anomaly Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12173-12182).

[25] Ruff, L., Vandermeulen, R. A., Gornitz, N., Binder, A., Muller, E., & Kloft, M. (2019). Deep support vector data description for unsupervised and semi-supervised anomaly detection. In Proceedings of the ICML 2019 Workshop on Uncertainty and Robustness in Deep Learning, Long Beach, CA,USA

[26] Ravanbakhsh, M., Nabi, M., Mousavi, H., Sangineto, E., & Sebe, N. (2018). Plug-and-play cnn for crowd motion analysis: An application in abnormal event detection. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 1689-1698)..

[27] Smeureanu, S., Ionescu, R. T., Popescu, M., & Alexe, B. (2017). Deep appearance features for abnormal behavior detection in video. In International Conference on Image Analysis and Processing (pp. 779-789).

[28] Gkountakos, K., Ioannidis, K., Tsikrika, T., Vrochidis, S., & Kompatsiaris, I. (2020). A Crowd Analysis Framework for Detecting Violence Scenes. In Proceedings of the 2020 International Conference on Multimedia Retrieval (pp. 276-280).

[29] Zhao, Y., Deng, B., Shen, C., Liu, Y., Lu, H., & Hua, X. S. (2017). Spatio-temporal autoencoder for video anomaly detection. In Proceedings of the 25th ACM international conference on Multimedia (pp. 1933-1941).

[30] Xu, D., Yan, Y., Ricci, E., & Sebe, N. (2017). Detecting anomalous events in videos by learning deep representations of appearance and motion. Computer Vision and Image Understanding, 156, 117-127.

[31] Yang, M., Rajasegarar, S., Erfani, S. M., & Leckie, C. (2019). Deep Learning and One-class SVM based Anomalous Crowd Detection. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

[32] Hamdi, S., Bouindour, S., Loukil, K., Snoussi, H., & Abid, M. (2019). Hybrid deep learning and HOF for Anomaly Detection. In 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT) (pp. 575-580).

[33] Nguyen, T. N., & Meunier, J. (2019). Anomaly detection in video sequence with appearance-motion correspondence. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1273-1283).

[34] Huang, S., Huang, D., & Zhou, X. (2018). Learning multimodal deep representations for crowd anomaly event detection. Mathematical Problems in Engineering.

[35] Xia, H., Li, T., Liu, W., Zhong, X., & Yuan, J. (2019). Abnormal event detection method in surveillance video based on temporal cnn and sparse optical flow. In Proceedings of the 2019 5th International Conference on Computing and Data Engineering (pp. 90-94).

[36] Ramchandran, A., & Sangaiah, A. K. (2019). Unsupervised deep learning system for local anomaly event detection in crowded scenes. Multimedia Tools and Applications, 1-21.

[37] Kamoona, A. M., Gostar, A. K., Bab-Hadiashar, A., & Hoseinnezhad, R. (2019). Sparsity-Based Naive Bayes Approach for Anomaly Detection in Real Surveillance Videos. In 2019 International Conference on Control, Automation and Information Sciences (ICCAIS) (pp. 1-6).

[38] Nievas, E. B., Suarez, O. D., García, G. B., & Sukthankar, R. (2011). Violence detection in video using computer vision techniques. In International conference on Computer analysis of images and patterns (pp. 332-339). Springer, Berlin, Heidelberg.

[39] Lu, C., Shi, J., & Jia, J. (2013). Abnormal event detection at 150 fps in matlab. In Proceedings of the IEEE international conference on computer vision (pp. 2720-2727).

[40] Adam, A., Rivlin, E., Shimshoni, I., & Reinitz, D. (2008). Robust real-time unusual event detection using multiple fixed-location monitors. IEEE transactions on pattern analysis and machine intelligence, 30(3), 555-560.

[41] Mehran, R., Oyama, A., & Shah, M. (2009). Abnormal crowd behavior detection using social force model. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (pp. 935-942). IEEE.

[42] Lin, W., Hasenstab, K., Cunha, G. M., & Schwartzman, A. (2020). Comparison of handcrafted features and convolutional neural networks for liver MR image adequacy assessment. Scientific Reports, 10(1), 1-11.

[43] Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. Journal of Manufacturing Systems, 48, 144-156.