A Survey of Deep Learning Solutions for Anomaly Detection in Surveillance Videos

John Gatara Munyua
School of Computing and Information Technology
Murang'a University of Technology
Murang'a, Kenya
Email: munyuajohn89 [AT] gmail.com

Geoffrey Mariga Wambugu
School of Computing and Information Technology
Murang'a University of Technology
Murang'a, Kenya
Email: gmariga [AT] mut.ac.ke

Stephen Thiiru Njenga
School of Computing and Information Technology
Murang'a University of Technology
Murang'a, Kenya
Email: snjenga [AT] mut.ac.ke

Abstract— **Deep learning has proven to be a landmark computing approach to the computer vision domain. Hence, it has been widely applied to solve complex cognitive tasks like the detection of anomalies in surveillance videos. Anomaly detection in this case is the identification of abnormal events in the surveillance videos which can be deemed as security incidents or threats. Deep learning solutions for anomaly detection has outperformed other traditional machine learning solutions. This review attempts to provide holistic benchmarking of the published deep learning solutions for videos anomaly detection since 2016. The paper identifies, the learning technique, datasets used and the overall model accuracy. Reviewed papers were organised into five deep learning methods namely; autoencoders, continual learning, transfer learning, reinforcement learning and ensemble learning. Current and emerging trends are discussed as well.**

Keywords- Deep Learning; Anomaly Detection; Anomaly Detection in Videos; Intelligence Video Surveillance; Deep Anomaly Detection.

I. INTRODUCTION

In the recent past, the use of surveillance cameras has rapidly increased to enhance public safety. Unfortunately, the ability of security forces to monitor these surveillance footages has not kept up with the speed of generation and the volume of the surveillance data. [1].

This scenario has resulted in a critical problem in the utilization of the surveillance footage since more human monitors are required as the surveillance cameras increase. The monitoring task requires dedicated attention since anomalous events are very rare. Hence, human monitors might miss out to signal security incidents.

Anomaly detection refers to the act of identifying improper behaviors in surveillance videos. Security surveillance considers anomalies as security incidents or threats. For instance, some of the most popular anomalies include violence, abuse, theft, traffic accidents, explosions, fighting, abuse, shooting, weapons, stealing, vandalism and shoplifting [2]. Anomaly detection in videos is complex at several levels: due to the subjectivity of anomaly definition, the rarity of anomalies in videos, big video data and high computational power required. This research problem has spiked research interest with many scholars making contributions on how to achieve intelligent surveillance. This paper brings together such inputs for betterment of the video anomaly detection.

Deep learning is a subset of machine learning that is based on variants of artificial neural networks [3]. To address more complex cognitive intensive problems many layers of neural networks are stack together. The stacking of many layers to create a deep network of layers is referred to as deep learning. Like the name 'deep' implies deep learning is all about the scale where larger synthetic neural networks are trained with a huge amount of data, while their performance and accuracy continue to increase [4]. This is notably different from other machine learning techniques that reach a plateau in their performance. Like machine learning, deep learning algorithms can be categorized as supervised, semi-supervised and unsupervised algorithms.

Researchers have utilized supervised learning methods to develop models to detect specific anomalous events. For instance; traffic accident detectors [5],[6], violence detectors, home intrusion detectors [7] and shoplifting detectors [8]. Unfortunately, these early solutions cannot be generalized to detect other abnormal events/actions since they have limited use.

To address such shortcomings, from the supervised models, other researchers proposed to use unsupervised learning algorithms. For instance, Waqas Sultani [2] proposes Multiple Instance Learning that can be generalized across a variety of anomalies. Chong [9] also proposes to use the Conv2DLSTM Autoencoder as a solution for the generalization shortcomings.

The purpose of this article is to provide a comprehensive benchmarking of the deep learning solutions implemented for video anomaly detection, discover the underlying learning techniques, identify trends in deep learning models design and gaps in the existing solutions.

This paper examined the deep learning models published since 2016. Only the deep learning solutions in anomaly detection in surveillance videos were considered. *30* papers were reviewed and summarized for better analysis. Datasets used for training and testing, model accuracy, learning techniques and the underlying deep learning algorithms are identified and compared for ranking.

The structure of this paper is as follows; Section II describes other review papers in this field of video anomaly detection. This work is put into context with other surveys, gaps present and contributions are discussed. Section III describes the findings of the systematic review. Deep learning models are analyzed and their different approaches to solving the anomaly detection problem. The models are clustered into renowned deep learning design architectures and learning techniques. A roadmap of the researches and the rapid improvements are detailed as well. Section IV introduces the noted trends in deep video anomaly detection, datasets, evaluation procedures. Finally, Section V wraps up with a conclusion and future expectations.

II. RELATED WORKS

This paper is concerned with a review of deep learning solutions for video anomaly detection problems. Other publications outside that scope are not considered. Several reviews were found to be close to this delimitation.

A survey of video anomaly detection by Lomte and others investigated all solutions made for video anomaly detection: both machine learning and deep learning. The review considered very few papers, hence the need for a comprehensive review [10].

A survey by Mohammadi, Fathy and Sabokrou [11] on deep video anomaly detection concentrates on the technologies and learning techniques in deep learning. Their survey considered algorithms and models from the traditional models like Histogram of Gradient (HoG) and the deep learning models like autoencoders. The work is drawn more on the internal working of the models and not the publications done in that area.

Other related work includes the publication made by Monika Singh on video anomaly detection [12]. This paper investigates the image-based techniques used to detect anomalies. Image-based techniques span from image recognition or what is referred to sparse representations. The paper goes further to assemble the techniques based on a single object and multiple trajectories and motion. This paper is limited to sparse based solutions and it focuses more on the inner working mechanisms of this technology.

Other works with similar interests include Ramchandran [13], which concentrates on the video scenes and anomalies definition, another paper that cannot be ignored is the survey on traffic anomalies by Santhosh [14]. This paper analyses the computer vision-based techniques that sort to understand the traffic violations and on-road anomalies. Their work analyses the techniques, frameworks, datasets and gaps. This work is limited to traffic anomalies only.

Since deep learning is the backbone of this study, it is important to understand what deep learning models entail and the different design architectures. Deep learning models are composed of neural network variants that are multi-layered. The architecture of deep learning models is extremely flexible since the models can differ in the number of layers, filter size and dimension, as well as the basic constructs [4]. Most of the models used in the detection of anomalies in videos have ConvNets, ConvLSTM and 3DCNN as the basic building blocks.

A layer within a deep learning model is composed of interconnected nodes(neurons). A node may be connected to all other nodes in the adjacent layers or not. Data fed through the model goes through each layer and is transformed to an abstract representation also known as extracted features. The training process sets the weights across different transformation functions. Then the model modifies the weights using backpropagation where the output is traced back to input modifying the weights [3].

Researchers have utilized this knowledge to design models using deep learning frameworks like PyTorch, Keras, TensorFlow and others. A model can be composed of different deep learning algorithms, where several deep learning algorithms are stacked together to produce a model. For instance, a model may have Conv2D, ConvLSTM, layers combine to get a model that works for sequential problems using high dimensional data like videos. It can be noted that the nature of the problem inspires the design of the model.

Our literature survey provides a comprehensive road map of deep anomaly detection in videos. The works published in this area are clustered according to the learning technique. The methodologies used to detect anomalies are analyzed and compared to uncover the rapid changes and their motivations. This paper explores more publications than its peers and aims at establishing the trends and some ranking of the best model.

.

III. DEEP LEARNING MODELS

This chapter discusses the landmark publications in deep learning anomaly detection in surveillance videos that have received much attention. This section was arranged according to the learning techniques present in the reviewed papers. Autoencoders, transfer learning, ensemble learning, reinforcement learning and continual learning are the large thematic areas.

A. Transfer Learning

Transfer Learning describes the handover of the knowledge from one model to another. This approach uses an already pre-trained model to solve a different task. Transfer learning is very useful when there is a scarcity of data or computational resources since it allows the models to use less data by re-using the learned weights from the pre-trained model. Other advantages of transfer learning include improvement of performance accuracy of the base model and reduction of the training time.

Transfer Learning was found as a growing trend in video anomaly detection and deep learning. One strategy of implementing transfer learning is through feature extraction. Pretrained models were used to extract features from labelled video and imagery data. The pre-trained models used mostly in the reviewed models include deep 3-dimensional convolutional networks (C3D) Model [2], Inception V3 Module [15], I3D, You Look Only Once Version 3 (YOLOV3).

C3D borrows from BVLC Caffee which was modified to support 3D Convolution and pooling [16]. C3D model was trained on UCF-101 and sports 1M videos to extract features from videos. Which can be very useful in down sampling the dataset for effective processing. The 3D Convolution extracts, both temporal and spatial features of the motion of objects, human scenes detection and their interactions [16]. The pre-trained model has been used in various models. For instance, Sultani and others [2] utilize C3D for feature extraction in their paper. The model is set to input a video and then it extracts a tensor of 4096 features.

The use of pre-trained models to extract features from videos has been widely used in anomalies detection research. Other feature extractor models that were found include Inflated 3D (I3D), which was trained on the Kinetics-400 dataset. I3D is composed of two streams of inflated 3D ConvNets [17]. 2D kernels were inflated by adding a time dimension to filters and kernels from N x N to N x N x N. It extracts features from videos and gives an output of shape 1024. By default, frames fed should be of size 224x224 and video to be recorded at 25 frames per second (fps) [17].

Another important feature extractor used by researchers is You Look Only Once Version3 (YOLOv3) which is a deep convolutional neural network that identifies specific objects in videos or images [18]. YOLOv3 is an improved version of YOLOv2 that borrows heavily from the DarkNet model that was trained on Imagenet. YOLOv3 combines two 53 layers of Darknets to form a deep 106-layer network. Object detection in the model happens within three different locations. The first Detection happens at the 82nd layer that uses the 1x1 kernel, the second detection happens at the 94th layer that uses the 2x2 kernel while the third detection occurs at the 106th layer that uses the 2x2 kernel. The model also predicts bounding boxes on the objects and draws them around the objects and labels the objects [18]. This detector was used to extract objects from videos which were used to define anomalies and normal scenes, which anomaly detection was based on.

Transfer learning was identified in the following papers, Sultani [2], that used C3D pre-trained model combined with light classifier i.e. Support Vector Machine (SVM) to assign a ranking score for the normal and abnormal instances. The C3D model was used for feature extraction. Motion and trajectory features were extracted from the real-world UCF crime dataset. The model was trained on both normal and abnormal videos, which were used to generate the ranking bags of normal and abnormal instances. A novel ranking loss function is used to estimate the anomaly level for every video [2].

Nazare and others [19] also explored the use of Pretrained CNNs in anomaly detection. Nazare explored several CNN networks including VGG-16, ResNet-50, Xception and DenseNet-121.Their paper investigated the role of pre-trained image classifiers in feature extraction to solve the problem of anomaly detection. The paper found that the Xception model outperforms its counterparts and it can be used for features extraction even though the whole idea performs poorly compared to other anomaly detection methods. Other examples of transfer learning found in the review include [20], [21], [22], [23], [15], [24].

In this review, 30% out of the 30 papers reviewed had adopted the transfer learning paradigm by using pre-trained models to improve the performance of the new models. Pre-trained models design architecture is also borrowed to create other models by borrowing the design knowledge.

B. Autoencoders

Autoencoders are a substantial part of the survey, out of 30 papers reviewed, 11 papers were found to have used the autoencoder model design paradigm. Which is around 36% which is very significant statistically. Thus, autoencoders can be considered as a growing trend in video anomaly detection. Autoencoders are widely used due to their unsupervised nature, ability to learn without human supervision and unlabeled data. The golden idea behind autoencoders is the reconstruction error that arises after when reconstructing the abnormal frames. The reconstruction error of the irregular videos is larger than regular videos. This idea is applied in designing models that detect anomalies in videos.

The autoencoders found in the review have different architectures and deep learning algorithms. For instance, Nguyen and Meunier [25] integrate a Conv-AE and Inception Module to form a deep autoencoder that detects the appearance and motion features from the videos. The decoder part of the model has two units that are dedicated motion and appearance.

Duman and Erdem [26] autoencoder is composed of Convolutional Autoencoder and Convolutional LSTM. This framework uses Optical Flow to extract features of speed and trajectory from the videos. The optical flow output is fed to the autoencoder which returns the reconstructed optical flow map. The reconstructed output is subtracted from the input to acquire the mean squared error that is used to calculate the regularity score that indicates the abnormality level of every frame.

Ramchandran and Sangaiah [27] unsupervised solution for anomaly detection in crowded scenes was based on autoencoder. The model was constituted of Conv-LSTM. Raw images sequences and edge image sequences were used to train the model.

Spatial-Temporal autoencoder is another variation of the autoencoders encountered in the review. This model was made by Zhao and others [28] in their paper named Spatial-Temporal Autoencoder for Video Anomaly Detection. Their model is composed of 3D convolutional layers. The architecture of the network is made up of an encoder and two decoder branches. The decoder branches, consist of the prediction branch and reconstruction branch. The two branches are used to create prediction loss function and reconstruction loss function that are used to estimate regularity score for anomalies locating.

Pawar and Attar autoencoder is a hybrid of a convolutional autoencoder and LSTM autoencoder [29]. This presents another design paradigm of combining two different autoencoders to create a seamless model. The convolutional part takes care of the image part while the LSTM preserves the sequence. Reconstruction error is used to model the regularity score [29].

Variational Autoencoder is an improvement of autoencoders that employs the use of probabilistic modelling to select the best reconstruction from the latent space. Unlike the normal autoencoders that encode the latent space as a single point, variational autoencoders generate the latent space as a distribution. Wu and others [30] exploited this architecture to create a two-stream variational autoencoder to detect anomalies in both local and streaming videos.

Other unique autoencoders found include Bhakat and Ramakrishnan [31], Mahmudul Hasan and others [32], Sabokrou and Fathy [33] and another case of Spatial-temporal autoencoder by Chong and Tay [9]. The Spatial-temporal autoencoder is different due to its building constructs. It employs time-distributed layers wrapped conv2d layers for the spatial part and convlstm2d for the temporal part.

C. Ensemble Learning

Other reviewed models are random cases of ensemble learning that combines multiple learning algorithms to get better predictive performance than the constituent learning algorithms alone. For instance, Zahid and others [15] is a typical case of ensemble and transfer learning. The model combines both a 3D convolutional network and a Fully Connected (FC) Network. Vu and others [34] is another case of ensemble learning that combines Conditional Generative Adversarial Networks, R-CNN and Support Vector Machines (SVM).

D. Continual Learning

Continual Learning describes a non-stop learning mechanism, step by step maintaining the previously learnt knowledge. Other deep learning models incline to terribly forget the existing knowledge when they learn new observations. The model forgets when the new data to be learnt differs significantly from the previous observations. This causes the new information to overwrite the previous knowledge in the common internal representation of the neural network. To curb the catastrophic forgetting problem, a solution to regularize the whole network to preserve the trained knowledge was proposed. This type of learning has been used for anomaly detection by Doshi [35].

The continual learning was just a part of the model that enabled the newly learnt anomalies to be added to the knowledge without losing the previous knowledge. The anomaly detection part utilized Euclidian distance k in the nearest neighbors (KNN) to identify anomalies that lie away from the nominal manifold [35].

E. Reinforcement Learning

Reinforcement learning is another rare technique found in a single paper during the review. This technique describes a sequential decision-making system that uses an agent to make choices and receives a reward when it makes the right choice. This enables the agent to acquire new behavior and skills incrementally [36]. The learning process is cyclic since it involves repeating series of steps. The initial step entails an agent perceiving the environment and it acquires a new state and a reward, the second step involves the agent choosing the next cause of action. The third step involves the agent sending the action to the environment and finally, it modifies its internal state as inspired by the previous state and agents' actions [36].

This learning technique has been applied by Aberkane [37] to detect anomalies in videos. Aberkane and Elarbi used a Deep Q Learning Network (DQN) to locate anomalies in videos. The model design borrows heavily from the Multiple Instance Learning by Sultani [2]. The DQN enables the agent to learn how anomalies are detected and recognized in the videos. DQN is composed of a fully connected layer, that calculates the probability of every video clip in the anomalous and normal bags demonstrating the likelihood of a clip containing an anomaly [37].

F. A Summary of Deep Learning Models

Table 1: Summary Table Showing Deep Learning Models used in video anomaly detection

Publication	Learning Technique	Deep Learning Algorithm/Models	Datasets	Overall Accuracy
Chong & Tay 2017 [9]	Auto-encoder	ConvLSTMAE	UCSD Ped1, UCSD Ped2	87%
Sabokrou, Fathy & Hoseini 2017 [33]	Auto-encoder	Sparse AE & Non-Sparse AE	UCSD Ped2 & UMN	90.8%
Mahmudul Hasan et al 2016, [32]	Fully Conv Feed Forward Auto-encoder	FC ConvNet AE	UCSD Ped1 UCSD Ped2 CHUK Avenue	83.18%
Chalapathy, Menon & Chawla 2017 [38]	Robust PCA	PCA	Cifar10	89%
Sultani, Chen & Shah, 2017 [2]	Multiple Instance Learning (MIL)	C3D, FC Convnet SVM Classifier	UCF Crime Dataset	75.41%
Nguyen, 2020 [6]	Generative Adversarial Network	GAN	AI City Challenge	91%
Xu et al, 2017 [30]	Variation Auto-encoder	2 stream VAE/ GAN	-	-
Doshi & Yilmaz [35]	Continual Learning	YOLOv3 *KNN*	UCSD, Avenue, Shangai Tech	85%
Kavikuil & Amudha [39]	Feature Learning	CNN	-	-
Liu et al. [40]	Transfer Learning	Binary Networks, 3DCNN	citySCENE	94.6%
Ullah and others [24]	Residual Learning	CNN, Residual LSTM	UCF,UMN,Avenue	98.3%
Vu and others 2021 [34]	Ensemble Learning	R-CNN, SVM, CGAN	Avenue,UCSD Ped1, Ped2, Shangai Tech	91.7%
Cinelli and others 2017 [41]	Residual Network	ConvNet	CDNET2014	84.9%
Bhakat and Ramakrishnan 2019 [31]	Auto-encoder	ConvLSTM	Avenue, Surveillance Office, Police	73.6%
Ullah and others 2021 [42]	Transfer Learning	Pre-trained CNN, BD-LSTM	UCF Crime	89.05%
Zahid and others [15]	Transfer & Ensemble Learning	Fully Connected Network, Inception V3,	UCF Crime	-
Murugesan and Thilagamani [43]	Ensemble Learning	MLP-RNN	-	-
Nazare, Mello & Ponti [19]	Transfer Learning	Pre-trained CNNs	UCSD Ped2	76%
Aberkane and Elarbi [37]	Reinforcement Learning	Deep Q Learning Network (DQN),	UCF Crime	n/a
Bansod & Nandedkar [20]	Transfer Learning	Pre-trained CNN (VGG16)	UCSD, UMN	
Cinelli [21]	Transfer Learning	Pre-trained CNN ResNet	CDNET2014	85%
Pawar & Attar 2021 [44]	Auto-Encoder	ConvAE, LSTM AE	-	-
Zhao and others 2017 [45]	Spatial Temporal Auto-encoder (STAE)	ConvLSTM	UCSD Ped1 & Ped2, CUHK Avenue	86.8%
Ramchandran & Sangaiah 2020 [13]	Auto-Encoder	ConvLSTM	UCSD Ped1 & Ped2	-
Duman & Erdem 2019 [26]	Auto-Encoder	OF-ConvAE-ConvLSTM	Avenue, UCSD Ped1, Ped2	91.53%
Doshi & Yilmaz 2021 [23]	Transfer Learning	Pre-trained Convnet (YOLOV3) & Least Square Generative Adversarial Network LS-GAN	, UCSD Ped2 & CUHK Avenue	84.83%
Nasaruddin 2020 [46]	Transfer Learning	3D-CNN	UCF Crime	95.4%
[47] Khaleghi and Moin 2018	Transfer Learning	CNN	UCSD	-
Doshi & Yilmaz 2020[22]	Transfer Learning	Pre-trained Convnet (YOLOV3) & GAN	UCSD PED2, CUHK, Shanghai Tech	84.87%
Nguyen 2019 [25]	Auto-encoder Hybrid	Conv-Net, GAN	UCSD Ped2, CHUK Avenue, Subway Entrance, Exit	91%

A variety of models were discovered from the empirical review. All the models selected for the empirical review had some deep learning components in them. Some models combine both deep learning and traditional machine learning algorithms like Support Vector Machine (SVM) [2]. The most popular underlying deep learning algorithms are 3DCNN, ConvLSTM and ConvNets.

IV. DATASETS, EVALUATION METRICS AND TRENDS

A. Datasets

Publicly available, anomaly detection video datasets were used for the most of experiments in the reviewed works. UCF Crime dataset, UCSD Ped1 & Ped2, and Avenue Dataset. The UCF Crime dataset is 1900 hours long videos dataset that was introduced by Sultani [2], it is composed of real-life anomalies like Arrest, Arson, Abuse and many others. The training set has both abnormal and normal videos as well as the testing set. Although usage of both classes is dependent upon the nature of the model to be trained. For instance, in the auto-encoder model, only the normal videos are used for training while in his model was trained by both normal and abnormal videos [9].
University of California San Diego (UCSD) Ped1 & Ped2 datasets are used for training and testing [48]. UCSD Ped1 is composed of 70 videos with 34 as the training set and 36 as the testing set. The videos scenery is a group of people walking in a park. Anomalies include non-pedestrian entities like bikers, skaters, carts, Wheelchairs and people walking in the grass area.

Avenue dataset [49] contains 16 training and 21 testing video clips. A total of 30652 frames are available in the dataset. These videos are captured on a campus street using a still camera. Strange actions like the running of persons, riding a bike in the walkway are the abnormal events presented.
Other popular datasets found in the papers include UMN and ShangaiTech video datasets. The ShanghaiTech [50] Campus dataset is composed of 130 abnormal events within 13 different. In total, the ShanghaiTech contains over 270,000 frames. University of Minnesota (UMN) dataset was introduced by researchers developing anomaly detection models for crowded scenes. Therefore, it contains crowd escape and panic of 11 videos with 3 different scenes [51].

B. Evaluation Criteria

This subsection highlights some of the most popular model evaluation methods encountered during the review. Most models have employed the Receiver Operating Characteristic Curve (ROC) and its resultant Areas Under the Curve (AUC). The ROC curve is the plot of the successful cases of True Positives versus the False Positive [52]. This measures the specificity and the precision of the model.

The equal Error rate (ERR) metric was found in many of the papers as a measure that quantified the number of errors present in the model predictions. It establishes the threshold value for equalizing False acceptance and False Rejection [52]. The model accuracy is high if the ERR is low.

F1 Score was used in some papers. This metric is used to measure the accuracy of binary classification. It has been optimized by researchers to measure the accuracy of anomaly detection while anomaly detection is optimized as binary classification at the frame level. It is calculated as a weighted ratio of the product of precision and recall and the sum of the precision and recall. Unlike the ROC curve, it takes false positive and false negative into account [52]. Usually, the F1 score is between 0 and 1, and the higher the F1 value is the better the model is

C. Trends

It can be noted that transfer learning and autoencoders deep learning techniques have taken the biggest part of the reviewed papers. The ability to transfer knowledge has made it easier to develop and deploy models faster and the ability to improve the performance of the base model has made transfer learning more lucrative.
Autoencoders ability to learn with minimal or no supervision at all has made the researchers use its design technique to implement different deep learning models that have shown comparable performance to others models in the same area.
The emerging trends will be based on the ability of the model to continually learn new knowledge since the nature of anomalies is subjective. Detection of anomalies in real-time and the ability to progressively learn novel observations are important aspects of video surveillance.
Hence, the positioning of continual and reinforcement learning is the new direction. Currently, only a little work has been done and due to the nature of anomaly detection, more interest is expected to span that way to enable online anomaly detection and sequential improvements of the models.

V. CONCLUSION

In this study, we have analyzed the deep learning solutions, implemented to solve the anomaly detection in videos. Deep learning models were classified according to the learning techniques and design architecture. The mechanism of anomaly detection within the papers was discussed as well. The most popular datasets used for training and testing the models were discussed, as well as the evaluation criteria used in the reviewed works.
It can be noted that hybrid models work better as seen in both ensembles and transfer learning. There are emerging trends in a model’s ability to constantly improve its behavior in novel observations hence the reinforcement and continual learning are expected to receive much attention. Future work in this research domain should focus on online anomaly detection problems, for real-time detections and continued learning to optimize the model behavior.

VI. References

[1] R. Yadav and M. Rai, "Advanced Intelligent Video Surveillance System (AIVSS): A Future Aspect," *Research Gate,* 2018.

[2] W. Sultani, C. Chen and M. Shah, "Real-World Anomaly Detection in Surveillance Videos," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 6479-6488, 2018.

[3] A. Borner, "What is Deep Learning and How Does it Work? | Content Simplicity," 2019. [Online]. Available: https://contentsimplicity.com/what-is-deep-learning-and-how-does-it-work/.

[4] J. Brownlee, "What is Deep Learning?," 16 August 2019. [Online]. Available: https://machinelearningmastery.com/what-is-deep-learning/.

[5] M. U. Farooq, N. A. Khan and M. S. Ali, "Unsupervised Video Surveillance for Anomaly Detection of Street Traffic," *(IJACSA) International Journal of Advanced Computer Science and Applications,* pp. 270-275, 2017.

[6] K. T. Nguyen, D. T. Dinh, M. N. Do and M. T. Tran, "Anomaly Detection in Traffic Surveillance Videos with GAN-based Future Frame Prediction," *Proceedings of the 2020 International Conference on Multimedia,* pp. 457-463, 2020.

[7] A. Kushwaha, A. Mishra, K. Kamble, R. Janbhare and A. Pokhare, "Theft Detection using Machine Learning," *IOSR Journal of Engineering (IOSRJEN),* pp. 67-71, 2018.

[8] K. Wiggers, "AI Guardsman uses computer vision to spot shoplifters," 26 June 2018. [Online]. Available: https://venturebeat.com/2018/06/26/ai-guardsman-uses-computer-vision-to-spot-shoplifters/.

[9] Y. S. Chong and Y. H. Tay, "Abnormal Event Detection in Videos Using Spatiotemporal Autoencoder," *arxiv,* vol. 1701, no. 01546v1, 2017.

[10] V. Lomte, S. Singh, S. Patil, S. Patil and D. Pahurkar, "A Survey on Real World Anomaly Detection in Live Video Surveillance Techniques," *International Journal of Research in Engineering, Science and Management,* vol. 2, no. 2, pp. 2581-5792, 2019.

[11] B. Mohammadi, M. Fathy and M. Sabokrou, "Image/Video Deep Anomaly Detection: A Survey," *Computing Research Repository (CoRR),* vol. abs/2103.01739, 2021.

[12] M. Singh, "A Survey on Video Anomaly Detection," *International Journal of Engineering Research & Technology (IJERT),* vol. 5, no. 10, 2017.

[13] A. Ramchandran and A. K. Sangaiah, "Unsupervised deep learning system for local anomaly event detection in crowded scenes," *Multimedia Tools and Applications,* vol. 79, no. 47/48, p. 35275–35295, 2020.

[14] K. K. Santhosh, D. P. Dogra and P. P. Roy, "Anomaly Detection in Road Traffic Using Visual Surveillance: A Survey," *ACM Computing Surveys,* vol. 53, no. 6, pp. 1-26, 2019.

[15] Y. Zahid, M. A. Tahir and M. N. Durrani, "Ensemble Learning Using Bagging And Inception-V3 For Anomaly Detection In Surveillance Videos," in *2020 IEEE International Conference on Image Processing (ICIP),* Abu Dhabi, United Arab Emirates, 2020.

[16] D. Tran, L. Bourdev, R. Fergus, L. Torresani and M. Paluri, "Learning Spatiotemporal Features with 3D Convolutional Networks," *IEEE International Conference on Computer Vision (ICCV),* p. 4489–4497, 2015.

[17] J. Carreira and A. Zisserman, "Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset," *Computing Research Repository,* vol. abs/1705.07750, 2018.

[18] V. Meel, "YOLOv3: Real-Time Object Detection Algorithm (What's New?)," viso.ai, 25 Feb 2021. [Online]. Available: https://viso.ai/deep-learning/yolov3-overview/. [Accessed 17 August 2021].

[19] T. S. Nazare, R. F. de Mello and M. A. Ponti, "Are pre-trained CNNs good feature extractors for anomaly detection in surveillance videos?," *eprint arXiv,* no. 1811.08495v1, 2018.

[20] S. Bansod and A. Nandedkar, "Transfer learning for video anomaly detection," *Journal of Intelligent & Fuzzy Systems,* vol. 36, no. 3, pp. 1967-1975, 2019.

[21] L. P. Cinelli, "Anomaly Detection in Surveillance Videos Using Deep Resdiual Networks," Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2017.

[22] K. Doshi and Y. Yilmaz, "Any-Shot Sequential Anomaly Detection in Surveillance Videos," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops,* pp. 934-935, 2020.

[23] K. Doshi and Y. Yilmaz, "Online anomaly detection in surveillance videos with asymptotic bound on false alarm rate," *Pattern Recognition,* vol. 114, p. 107865, 2021.

[24] 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," *AI-Enabled Advanced Sensing for Human Action and Activity Recognition,* vol. 21, 2021.

[25] T.-N. Nguyen and J. Meunier, "Anomaly Detection in Video Sequence with Appearance-Motion Correspondence," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

[26] E. Duman and O. A. Erdem, "Anomaly Detection in Videos Using Optical Flow and Convolutional Autoencoder," *IEEE Access,* vol. 7, pp. 183914 - 183923, 2019.

[27] B. Ramachandra, M. J. Jones and R. R. Vatsavai, "A Survey of Single-Scene," *IEEE Transactions on Pattern Analysis and Machine Intelligence ,* 2020.

[28] Y. Zhao, B. Deng, C. Shen, Y. Liu, H. Lu and X.-S. Hua, "Spatio-Temporal AutoEncoder for Video Anomaly Detection," *Proceedings of the 25th ACM international conference on Multimedia,* pp. 1933-1941, 2017.

[29] K. V. Pawar and V. Attar, "Deep learning approaches for video-based anomalous activity detection," *World Wide Web,* p. 22, 27 May 2018.

[30] H. Wu, J. Shao, X. Xu, F. Shen and H. Shen, "A System for Spatiotemporal Anomaly Localization in Surveillance Videos," *Proceedings of the 25th ACM international conference on Multimedia,* pp. 1225-1226, 2017.

[31] S. Bhakat and G. Ramakrishnan, "Anomaly Detection in Surveillance Videos," *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, p. 252–255, 2019.

[32] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury and L. S. Davis, "Learning Temporal Regularity in Video Sequences," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 733-742, 2016.

[33] M. F. M. M. Sabokrou, "Video anomaly detection and localisation based on the sparsity and reconstruction error of auto-encoder," *Electronic Letters,* vol. 52, no. 13, pp. 1122-1124, 2016.

[34] T.-H. Vu, J. Boonaert, S. Ambellouis and A. Taleb-Ahmed, "Multi-Channel Generative Framework and Supervised Learning for Anomaly Detection in Surveillance Videos," *Human Activity Recognition Based on Image Sensors and Deep Learning*, vol. 21, no. 9, p. 3179, 2021.

[35] K. Doshi and Y. Yilmaz, "Continual Learning for Anomaly Detection in Surveillance Videos," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Seattle, WA, USA, 2020.

[36] J. Torres, "A gentle introduction to Deep Reinforcement Learning," Towards Data Science, 15 May 2020. [Online]. Available: https://towardsdatascience.com/drl-01-a-gentle-introduction-to-deep-reinforcement-learning-405b79866bf4. [Accessed 18 August 2021].

[37] S. Aberkane and M. Elarbi, "Deep Reinforcement Learning for Real-world Anomaly Detection in Surveillance Videos," in *2019 6th International Conference on Image and Signal Processing and their Applications (ISPA)*, Mostaganem, Algeria, 2019.

[38] R. Chalapathy, A. K. Menon and S. Chawla, "Robust, Deep and Inductive Anomaly Detection," *Machine Learning and Knowledge Discovery in Databases,* vol. 10534, 2017.

[39] K. Kavikuil and J. Amudha, "Leveraging Deep Learning for Anomaly

Detection in Video Surveillance," *First International Conference on Artificial Intelligence and Cognitive Computing,* vol. 815, no. I, pp. 239-247, 2018.

[40] K. Liu, M. Zhu, H. Fu, H. Ma and T.-S. Chua, "Enhancing Anomaly Detection in Surveillance Videos with Transfer Learning from Action Recognition," *Proceedings of the 28th ACM International Conference on Multimedia,* pp. 4664-4668, 2020.

[41] L. P. Cinelli, L. A. Thomaz, A. F. d. Silva, E. A. B. d. Silva and S. L. Netto, "Foreground Segmentation for Anomaly Detection in Surveillance Videos Using Deep Residual Networks," *XXXV Simposio Brasileiro de Telecomuniac, Oes e processamento de Sinais,* pp. 3-6, 2017.

[42] 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," *Multimedia Tools and Applications,* p. 16979–16995, 2021.

[43] M.Murugesan and S.Thilagamani, "Efficient anomaly detection in surveillance videos based on multi-layer perception recurrent neural network," in *Microprocessors and Microsystems*, 2020.

[44] V. A. Karishma Pawar, "Application of Deep Learning for Crowd Anomaly Detection from Surveillance Videos," in *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 2021.

[45] Y. Zhao, B. Deng, C. Shen, Y. Liu, H. Lu and X.-S. Hua, "Spatio-Temporal AutoEncoder for Video Anomaly Detection," *Proceedings of the 25th ACM international conference on Multimedia,* pp. 1933-1941, 2017.

[46] N. Nasaruddin, K. Muchtar, A. Afdhal and A. P. J. Dwiyantoro, "Deep anomaly detection through visual attention in surveillance videos," *Journal of Big Data,* vol. 7, no. 87, 2020.

[47] A. Khaleghi and M. S. Moin, "Improved anomaly detection in surveillance videos based on a deep learning method," in *2018 8th Conference of AI & Robotics and 10th RoboCup Iranopen International Symposium (IRANOPEN)*, Qazvin, Iran, 2018.

[48] UCSD, "UCSD Anomaly Detection Dataset," UCSD, 2014. [Online]. Available: http://www.svcl.ucsd.edu/projects/anomaly/dataset.html. [Accessed 10 May 2021].

[49] C. Lu, J. Shi and J. Jia, "Avenue Dataset for Abnormal Event Detection," The Chinese Univeristy of Hong Kong, 2013. [Online]. Available: http://www.cse.cuhk.edu.hk/leojia/projects/detectabnormal/dataset.html. [Accessed 10 May 2021].

[50] W. Liu, W. Luo, D. Lian and S. Gao, "Future Frame Prediction for Anomaly Detection -- A New Baseline," in *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA , 2018.

[51] R. Mehran, A. Oyama and M. Shah, "Abnormal Crowd Behavior Detection using Social Force Model," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Miami, 2009.

[52] T. Kanstren, "A Look at Precision, Recall, and F1-Score," Towards Data Science, 09 September 2012. [Online]. Available: https://towardsdatascience.com/a-look-at-precision-recall-and-f1-score-36b5fd0dd3ec. [Accessed 18 August 2021].

[53] H. M. Kun Liu, "Exploring Background-bias for Anomaly Detection in Surveillance Videos," *Proceedings of the 27th ACM International Conference on Multimedia,* pp. 1490-1499, 2019.

[54] R. V. H. M. Colque, C. Caetano and M. T. L. d. Andrade, "Histograms of Optical Flow Orientation and Magnitude and Entropy to Detect Anomalous Events in Videos," *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY,* vol. 27, no. 3, pp. 673-682, 2017.

[55] A. Sarkar, "Human Activity and Behavior Recognition in Videos. A Brief Review," 2014. [Online]. Available: https://www.grin.com/document/276054.

[56] A. Kushwaha, A. Mishra, K. Kamble and R. Janbhare, "Theft-Detection using Motion Sensing Camera," *International Journal of Innovative Science and Research Technology,* pp. 90-97, 2017.

[57] M. Sabokrou, M. Fayyaz, M. Klette and R. Fathy, "Deep-Cascade: Cascading 3D Deep Neural Networks for Fast Anomaly Detection and Localizaton in Crowded Scenes," *IEEE Transactions on Image Processing,* pp. 1992-2004, 2017.

[58] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayed and R. Klette, "Deep-Anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes," *Computer Vision and Image Understanding,* pp. 1-25, 2018.

[59] W. Badr, "Auto-Encoder: What Is It? And What Is It Used For? (Part 1)," towards data science, 22 April 2019. [Online]. Available: https://towardsdatascience.com/auto-encoder-what-is-it-and-what-is-it-used-for-part-1-3e5c6f017726. [Accessed 10 May 2021].