

# An Anomaly Detection System via Moving Surveillance Robots with Human Collaboration

Muhammad Zaigham Zaheer<sup>1,2</sup> Arif Mahmood<sup>3</sup> M. Haris Khan<sup>4</sup> Marcella Astrid<sup>1,2</sup>  
Seung-Ik Lee<sup>1,2</sup>

<sup>1</sup>University of Science and Technology, Daejeon, South Korea.

<sup>2</sup>Electronics and Telecommunications Research Institute, Daejeon, South Korea.

<sup>3</sup>Information Technology University, Pakistan.

<sup>4</sup>Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi.

## Abstract

*Autonomous anomaly detection is a fundamental step in visual surveillance systems, and so we have witnessed great progress in the form of various promising algorithms. Nonetheless, majority of prior algorithms assume static surveillance cameras that severely restricts the coverage of the system unless the number of cameras is exponentially increased, consequently increasing both the installation and the monitoring costs. In the current work we propose an anomaly detection system based on mobile surveillance cameras, i.e., moving robots which continuously navigate a target area. We compare the newly acquired test images with a database of normal images using geo-tags. For anomaly detection, a Siamese network is trained which analyses two input images for anomalies while ignoring the viewpoint differences. Further, our system is capable of updating the normal images database with human collaboration. Finally, we propose a new tester dataset that is captured by repeated visits of the robot over a constrained outdoor industrial target area. Our experiments demonstrate the effectiveness of the proposed system for anomaly detection using mobile surveillance robots.*

## 1. Introduction

Surveillance systems are becoming indispensable towards making the modern cities safer. These systems not only aid in preventing crimes, but also enable the authorities to respond well in time in the face of unwanted events. In a typical surveillance system, cameras are installed at remote locations and human operators carefully monitor the live feed generated by these cameras to identify any malicious activity such as fire, stealing, shooting, accident, etc. Upon spotting an anomaly, the operator alerts the relevant

authorities.

A significant drawback of the manually monitored surveillance systems is the inability of human operators to continuously monitor video streams. Attention span of a human brain is often limited [17]. Moreover, the surveillance videos may not have any significant information most of the time as the anomalous activities do not occur often [18, 26]. Therefore, a human operator may lose focus which is not desirable in such systems. Given the possibility of human error in surveillance operations, designing an autonomous system is of paramount importance. Therefore, anomaly detection in surveillance videos has been gaining popularity in computer vision. The desideratum is to design a trainable system that can learn to detect anomalies in the streams of videos/images and generate an alarm if an anomaly appears in the scene.

A major limitation of recent anomaly detection solutions is that the corresponding systems assume stationary CCTV cameras [26, 18]. Therefore, the performance of these systems could deteriorate on video streams captured with moving cameras. To thoroughly monitor a relatively larger area, we require significantly large quantity of stationary cameras thereby resulting in increased initial deployment cost as well as recurring maintenance, energy and monitoring expenditures. Owing to current advancements in robotics, surveillance technologies are also adopting more sophisticated methods such as surveillance robots [2, 6]. However, the moving camera surveillance introduces several classical computer vision nuisances such as, moving backgrounds, motion blur, illumination variations, moving objects, etc. [19]. Existing approaches [4, 16, 26, 25, 24, 27], address anomaly detection problem only when video stream is captured from a stationary camera and could likely underperform under aforementioned challenges.

Anomaly detection is often seen as one-class classifi-

cation problem in which a model is trained only on normal data [7, 24]. It is because the normal data is abundantly available as compared to the anomaly examples. At test time, both normal and anomalous examples are used to simulate the real-world situations of anomalies appearing suddenly among the normal happenings. As the number of views increases for the case of moving camera with a surveillance robot, learning one class classifier becomes more and more challenging especially for the detection of anomalous objects. Illumination variations may cause the whole background to be predicted as anomalous.

Another limitation of existing approaches is their limited capability of continual learning after training. Unlike the machine learning architectures for usual object detection or other similar problems which can be trained and run in controlled environments, an anomaly detection system requires deployment in dynamic outdoor environments. This essentially requires that the definition of anomalies may need to be updated as and when necessary. For example, a car parked in the background, a new solar panel installed over a parking lot, or flower pots placed or removed from certain locations, etc. can be perceived as anomalous by a system and need to be manually changed thus requiring human intervention.

To this end, we aim to design an approach capable of performing well on the video data captured with moving camera surveillance robots as well as supporting the life-long adaptability through human interventions. The overall mechanism of our approach is illustrated in Figure 1. It takes live feed from a moving camera agent and passes it to a pre-trained anomaly detection algorithm. It is a network that takes a pair of images, one of which is the current image obtained via the camera agent and the other is its corresponding normalcy definition. This way, the system keeps looking for the major changes to be quantified as anomalies in the live feed. Once an anomaly is detected, the system notifies the human operator about the potential anomaly. The operator then decides whether the event/object flagged as anomalous is a false alarm or otherwise. If it is not a false alarm, which means that the anomaly is real, the operator may then alert the appropriate authorities to handle the matter. Otherwise, if it is a false alarm, the system receives the feedback, updates its normalcy definitions and continues testing the live feed.

The proposed system offers several benefits over the existing ones: 1) It can be updated based on human feedback, hence ensuring long-term learning. 2) It can detect anomalies in the imagery captured using moving camera agents. 3) It is a plug-and-play architecture, which means we can plug different change detection algorithms seamlessly without disintegration. 4) To handle a large number of views, each view is associated with a particular location point, viewing direction and angle. Such an arrangement avoids computationally expensive image search in a large-scale surveillance environment with excessive number of backgrounds.

tionally expensive image search in a large-scale surveillance environment with excessive number of backgrounds.

## 2. Related Work

Anomaly detection in images and videos has recently become an active area of research. In this section, we discuss several categories of existing anomaly detection approaches, their application contexts and their differences with our approach.

### 2.1. One-Class Classification

Most existing approaches utilize one-class classification (OCC) based training for anomaly detection, in which a normal class of data is used to carry out the training of one-class classifiers. Then at test time, the instances deviating from the learned representations are classified as anomalies.

Conventionally, to carry out the OCC training, some works proposed the idea of either utilizing hand-crafted features [8, 1, 21, 28, 11], or use deep features extracted using pre-trained convolution models [16, 12]. Recently, the idea of utilizing image regeneration based architectures has been explored by several researchers [3, 13, 22, 5, 9, 10, 23, 14, 15] to learn to reconstruct normal data in an unsupervised fashion. This way, at test time, it is expected to have more reconstruction error for the anomalous inputs. More recently, pseudo-supervised methods are also proposed in which pseudo anomalies are created using the normal training data and utilized to train binary classifiers [5, 24].

All these approaches generally follow the model of one-time training, in which the definitions of anomalies or normal may not be updated without a fresh training. Consequently, these methods also lack human intervention mechanisms, therefore any false alarm may never be corrected unless the system is re-trained completely. In contrast, our proposed architecture is designed by keeping in mind such limitations of the OCC methods and can overcome such drawbacks.

### 2.2. Weakly Supervised Anomaly Detection in Surveillance Videos

This is a relatively newer category of anomaly detection in which noisy annotations of videos are used to carry out the training in a weakly supervised setting [18, 29, 26, 27, 25, 20]. In typical settings, video-level labels are provided for training. Normal videos are noise free, which means no anomalies are present in normal labeled videos. However, in anomalous labeled videos, most of the scenes are still normal and the temporal location of the anomalies is unknown. While this setting has its own set of challenges, our approach is significantly different from this category. First, this category of approaches is only applicable to video-based training. Particularly, the videos are recorded using static cameras. Secondly, similar to the OCC methods, such

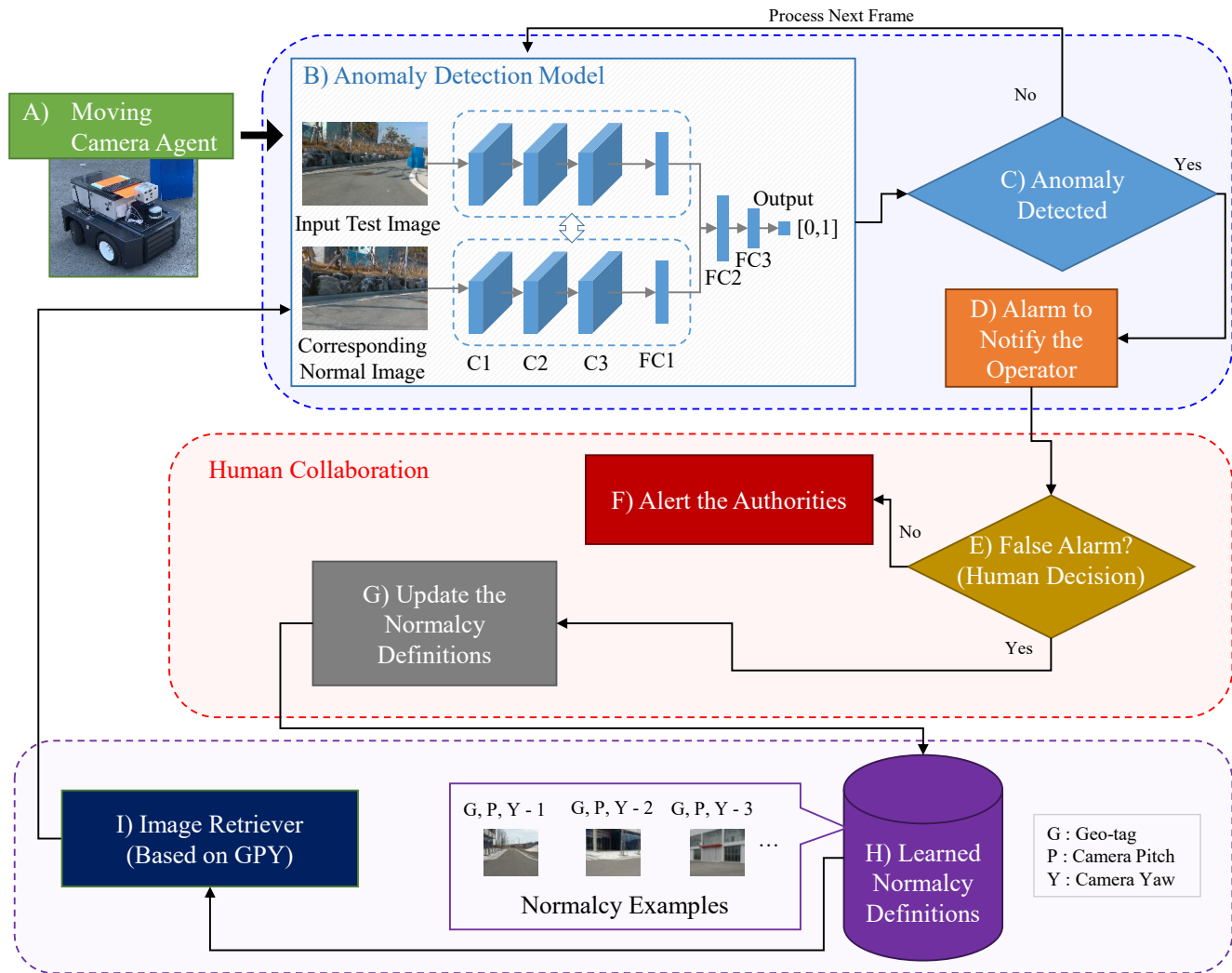


Figure 1. Overall architecture of the proposed system: A) Moving camera agent that performs surveillance over a target area. B) Change detection based anomaly detection model using Siamese network. C) & D) If anomaly is detected, human operator is notified. E) The operator decides whether the alarm is false or not. F) If the alarm is real, operator may take necessary actions such as alerting the authorities. G) & H) If false alarm, the corresponding normalcy definition is updated and added to the Learned Normalcy Definitions. (I) Image retriever is used to retrieve the image corresponding to the the input image from the Learned Normalcy Definitions using Geo-tag, camera Pitch and camera Yaw (GPY).

approaches are incapable of incorporating any new normal or anomalous into consideration without carrying out the re-training. Whereas our proposed approach considers moving camera imagery, while having the capability of incorporating human intervention based improvements.

### 3. Proposed Anomaly Detection System

The proposed system, as shown in Figure 1, has many parts, each of which is discussed next along with important implementation details.

#### 3.1. Camera Agent

Camera agent is a moving robot (as shown in Figure 1A), that continuously records images and transfers to the server. It also labels each image with the corresponding Geo-tag, Pitch and Yaw (GPY), while not utilizing camera roll. The geo-tags are the local coordinates of the zone under surveillance. Mostly, the yaw values change according to the surveillance path, while pitch remains almost constant.



Figure 2. Normal and anomalous geo-tagged image pairs taken from the proposed dataset. Each pair contains one normal image (reference) selected from the normalcy database while the second is captured by the robot (named as template) during test visits of the target area. Each pair of images is manually annotated for training and evaluation of the system.

### 3.2. Learned Normalcy Definitions

During training process, the camera agent (Figure 1A) explores the whole under-surveillance zone and, for each geo-tag point, captures normal images. These images are saved in a database (Figure 1H) and arranged according to the GPY values. This helps in retrieving the normal images previously saved at the similar location (Figure 1I) and the pair is passed to the anomaly detection model (Figure 1B).

### 3.3. Anomaly Detection Model

The desideratum of this research is to develop an algorithm that may not only support human feedback based life-long learning, but also be able to handle moving camera agents. To this end, a change detection based machine learning algorithm is developed that can be used to match the existing normalcy definitions with the new incoming test data and produce a score corresponding to the quantification of change happened in the new scene. This way, anything unusual will be detected by our system.

A Siamese based convolution network (SiamNet) is trained that can detect changes between two input images and is robust to view variations. An overview of this architecture is shown in Figure 1B. Based on this SiamNet, when the robot moves within a surveillance zone marked with geo-tags, each new test image is matched against the

normal images already stored in the database. To simplify, a new incoming image during the surveillance operation along with an already stored normal image in the database is forwarded to the SiamNet as a pair to be matched. If the SiamNet identifies noticeable changes, the prediction score is higher which translates directly to the presence of some anomaly in the scene.

#### 3.3.1 Architecture Details

The learning architecture used in our proposed framework is a conventional Siamese Network (SiamNet), designed to take a pair of images as input. In order to present the architecture details, we first define  $Conv(c_i, c_o, k, s)$  to represent a 2D convolution layer, where  $c_i, c_o, k,$  and  $s$  are input channels, output channels, kernel size and stride. We also define  $FC(c_i, c_o)$  to represent fully connected linear layers with  $c_i$  and  $c_o$  as input and output nodes. Each branch of the SiamNet is given as:  $Conv(3, 32, 3, 0)$ ,  $Conv(32, 64, 3, 2)$ ,  $Conv(64, 64, 3, 2)$ ,  $FC(50176, 1024)$ . The two branches are merged after difference using linear layers as:  $FC(1024, 256)$ ,  $FC(256, 1)$ . ReLU is used after each layer, except for the last layer where Sigmoid is used to regress scores between 0 and 1.

### 3.3.2 Training

Training of the network is carried out using supervised examples of paired images to minimize binary cross entropy loss between the labels and the network predictions. Each pair is formed by using a geo-tagged template image captured by the robot and selecting the corresponding reference image from the learned normalcy definitions. Each template image is manually labeled as normal or anomalous based on its comparison with the reference image. The focus of the training is to learn a change detection model that can quantify the change between the two paired images. It may be noted that this way of training is different from the fully supervised binary classification as, in a binary classifier, the network may learn to focus on identifying specific anomalous objects. However, using change detection to quantify anomalies can be a generic solution as the definition of normalcy can be updated by replacing the reference images.

### 3.4. Human Collaboration

Most of the existing machine learning algorithms for autonomous surveillance do not support long term learning. Which means if it is desired to change the normalcy definitions due to, for example, construction of a new building, a new installment in the under surveillance zone, etc. the whole model needs to be trained again with updated training examples. This is not only time consuming, but also cumbersome as optimizing the training model every time often requires expert supervision and several attempts. Therefore, we propose a model that does not require re-training every time and can sufficiently handle the new updates with minimal efforts. Hence, complementary to our change detection based machine learning algorithm, a human intervention mechanism is implemented that can update the normalcy definitions based on the feedback of the operator (Figure 1E). This way, we can ensure that the normalcy definitions are up-to-date and that a false alarm will not happen repeatedly for the same anomaly.

In case the incoming image is labeled as anomalous (Figure 1C) by the anomaly detection model, a human operator will verify it for the presence of a real anomaly (Figure 1E). In case the operator decides to mark the input image as normal, it will be added to the database of normal images at that particular GPY (Figure 1G). Thus the anomaly detector will not raise a false alarm on the same event next time.

## 4. Experiments

### 4.1. Proposed Dataset

Most of the currently popular anomaly detection datasets do not take moving camera robots into account. Therefore in the current research we have proposed a new dataset captured by a moving robot in a constrained outdoor industrial environment. The dataset is captured over an outdoor target

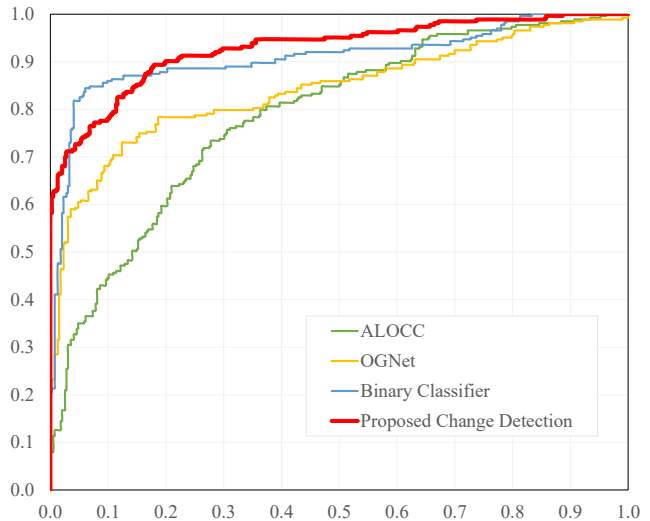


Figure 3. Receiver Operating Characteristic (ROC) curve of our proposed approach and other compared approaches including a convolutional network based binary classifier, OGNNet [24], and ALOCC [15].

area, with a local coordinate system dubbed as a test bed. The test bed consists of 480 geo-tag points. Among these points, 70% points (347) are used for training, while 133 points are used for testing. Average number of images captured at each geo-tag point is 5.7. Overall we have 1600 pairs for training, where each pair consists of one normal (reference image) and one template image which may be normal or anomalous. During training, SiamNet learns to produce 1.00 for the pairs having anomalous template images, and 0.00 for the pairs having normal template images. For the testing of our proposed system we have 659 pairs consisting of 263 pairs with anomalous template images and 396 with normal templates. Example images of the dataset are shown in Figure 2.

### 4.2. Experimental Settings

The trainable architecture used in our proposed approach is a Siamese network. Stochastic Gradient Descent is used for the training with the learning rate of the system set to  $10^{-3}$ , weight decay of  $5 \times 10^{-5}$  and momentum of 0.90. The network is trained for 50 epochs using a mini-batch size of 16 pairs. Training and testing are performed on a computer with Intel Core i7 processor, 128GB RAM, and NVIDIA TITAN Xp Graphics Card with 12GB memory.

### 4.3. Results and Comparisons

Area Under the Curve (AUC) % performance of our proposed anomaly detection approach is reported in Table 1 and the ROC curve plots are presented in Figure 3. Our approach demonstrates an AUC performance of 91.35%. In order to provide a vantage point to this perfor-



Table 1. AUC % performance comparison of the proposed algorithm with three existing approaches. OGNet and ALOCC are recently proposed one-class classification methods that utilize only normal data for training. Whereas, CNN is a convolution based binary classifier that takes one image input at a time and predicts corresponding anomaly score. All these methods do not support human collaboration.

Method	AUC %
ALOCC [15]	78.08
OGNet [24]	83.24
CNN - Binary Classifier	91.01
Ours	91.35

mance, we also compare our approach to two popular conventional anomaly detection approaches including ALOCC [15] which is essentially a unified autoencoder and discriminator network and OGNet [24], as well as a convolution neural network (CNN) trained as a binary classifier. The network architecture of the binary classifier is set similar to a branch of our SiamNet. OGNet and ALOCC are one-class classifiers and we train these models using only normal training examples within the training dataset. However, testing is carried out using both normal and anomalous examples from the test split. Only template images are used for training and testing while the reference images are completely discarded as these are only necessary/applicable in the case of change detection based anomaly detection algorithms. The performances of both OGNet and ALOCC are significantly lower which is understandable as these methods are not trained using actual anomaly examples. Binary classifier, on the other hand, demonstrates almost similar performance with our proposed approach. However, it may be noted that such a binary classifier is extremely dependent on the types of anomalies presented in the training set and cannot be generically used for different kinds of anomalies. Furthermore, it is also unable to support the proposed human collaboration mechanism. Whereas, our change detection based approach can not only support human intervention, but is also able to update the definitions of normalcy by updating the reference images.

## 5. Conclusions

In this work an anomaly detection system is proposed which is based on surveillance robots. The robots, while navigating a target area according to a local coordinate system, continuously capture images which are transmitted to a server. On the server, a normal images database is maintained which is arranged based on geo-tags, yaw and pitch of the robot camera. The test images and the corresponding normal images are matched using a Siamese deep neural network which is trained to find anomalies while ignoring the viewpoint differences. If the network detects the pres-

ence of anomaly, the test image is sent to a collaborating human which verifies if the image is anomalous. Samples which are found to be not-anomalous but declared anomalous by the Siamese network are added to the database of the normal samples at the corresponding geo-tag. Thus over the time, a geo-tagged database of normal images is developed which helps reduce false alarms. The system has consistently achieved excellent performance in the experiments. A new dataset captured by surveillance robots by multiple visits of the target area is also proposed. This dataset is a tester version, which is recorded with an intention of introducing the problem of anomaly detection using moving robots to the community. In future, it is being planned to extend this dataset by adding more types of anomalies, complex environments and evaluation protocols.

## 6. Acknowledgements

This work was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00306, Development of Multimodal Sensor-based Intelligent Systems for Outdoor Surveillance Robots, 50%) and the UST Young Scientist Research Program 2020 through the University of Science and Technology. (No. 2020YS21, 50%)

## References

- [1] Arslan Basharat, Alexei Gritai, and Mubarak Shah. Learning object motion patterns for anomaly detection and improved object detection. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.
- [2] Yifan Ding, Wenhao Luo, and Katia Sycara. Decentralized multiple mobile depots route planning for replenishing persistent surveillance robots. In *2019 International Symposium on Multi-Robot and Multi-Agent Systems (MRS)*, pages 23–29. IEEE, 2019.
- [3] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [4] Rui Hou, Chen Chen, and Mubarak Shah. Tube convolutional neural network (t-cnn) for action detection in videos. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5822–5831, 2017.
- [5] Radu Tudor Ionescu, Fahad Shahbaz Khan, Mariana-Iuliana Georgescu, and Ling Shao. Object-centric auto-encoders and dummy anomalies for abnormal event detection in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7842–7851, 2019.
- [6] Muhammad Umer Khan, Shuai Li, Qixin Wang, and Zili Shao. Cps oriented control design for networked surveillance robots with multiple physical constraints. *IEEE Transactions*

- on *Computer-Aided Design of Integrated Circuits and Systems*, 35(5):778–791, 2016.
- [7] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future frame prediction for anomaly detection—a new baseline. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6536–6545, 2018.
- [8] Gérard Medioni, Isaac Cohen, François Brémond, Somboon Hongeng, and Ramakant Nevatia. Event detection and analysis from video streams. *IEEE Transactions on pattern analysis and machine intelligence*, 23(8):873–889, 2001.
- [9] Trong-Nguyen Nguyen and Jean Meunier. Anomaly detection in video sequence with appearance-motion correspondence. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [10] Trong Nguyen Nguyen and Jean Meunier. Hybrid deep network for anomaly detection. *arXiv preprint arXiv:1908.06347*, 2019.
- [11] Claudio Piciarelli, Christian Micheloni, and Gian Luca Foresti. Trajectory-based anomalous event detection. *IEEE Transactions on Circuits and Systems for video Technology*, 18(11):1544–1554, 2008.
- [12] Mahdyar Ravanbakhsh, Moin Nabi, Enver Sangineto, Lucio Marcenaro, Carlo Regazzoni, and Nicu Sebe. Abnormal event detection in videos using generative adversarial nets. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 1577–1581. IEEE, 2017.
- [13] Huamin Ren, Weifeng Liu, Søren Ingvor Olsen, Sergio Escalera, and Thomas B Moeslund. Unsupervised behavior-specific dictionary learning for abnormal event detection. In *BMVC*, pages 28–1, 2015.
- [14] Mohammad Sabokrou, Mohsen Fayyaz, Mahmood Fathy, and Reinhard Klette. Deep-cascade: Cascading 3d deep neural networks for fast anomaly detection and localization in crowded scenes. *IEEE Transactions on Image Processing*, 26(4):1992–2004, 2017.
- [15] Mohammad Sabokrou, Mohammad Khalooei, Mahmood Fathy, and Ehsan Adeli. Adversarially learned one-class classifier for novelty detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3379–3388, 2018.
- [16] Sorina Smeureanu, Radu Tudor Ionescu, Marius Popescu, and Bogdan Alexe. Deep appearance features for abnormal behavior detection in video. In *International Conference on Image Analysis and Processing*, pages 779–789. Springer, 2017.
- [17] Noah Sulman, Thomas Sanocki, Dmitry Goldgof, and Rangachar Kasturi. How effective is human video surveillance performance? In *2008 19th International Conference on Pattern Recognition*, pages 1–3. IEEE, 2008.
- [18] Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6479–6488, 2018.
- [19] Keqiang Sun, Wayne Wu, Tinghao Liu, Shuo Yang, Quan Wang, Qiang Zhou, Zuochang Ye, and Chen Qian. Fab: A robust facial landmark detection framework for motion-blurred videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5462–5471, 2019.
- [20] Yu Tian, Guansong Pang, Yuanhong Chen, Rajvinder Singh, Johan W Verjans, and Gustavo Carneiro. Weakly-supervised video anomaly detection with robust temporal feature magnitude learning. *arXiv preprint arXiv:2101.10030*, 2021.
- [21] Jiang Wang, Yang Song, Thomas Leung, Chuck Rosenberg, Jingbin Wang, James Philbin, Bo Chen, and Ying Wu. Learning fine-grained image similarity with deep ranking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1386–1393, 2014.
- [22] Dan Xu, Elisa Ricci, Yan Yan, Jingkuan Song, and Nicu Sebe. Learning deep representations of appearance and motion for anomalous event detection. *arXiv preprint arXiv:1510.01553*, 2015.
- [23] Dan Xu, Yan Yan, Elisa Ricci, and Nicu Sebe. Detecting anomalous events in videos by learning deep representations of appearance and motion. *Computer Vision and Image Understanding*, 156:117–127, 2017.
- [24] Muhammad Zaigham Zaheer, Jin-ha Lee, Marcella Astrid, and Seung-Ik Lee. Old is gold: Redefining the adversarially learned one-class classifier training paradigm. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14183–14193, 2020.
- [25] Muhammad Zaigham Zaheer, Jin-ha Lee, Marcella Astrid, Arif Mahmood, and Seung-Ik Lee. Cleaning label noise with clusters for minimally supervised anomaly detection. In *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020.
- [26] Muhammad Zaigham Zaheer, Arif Mahmood, Marcella Astrid, and Seung-Ik Lee. Claws: Clustering assisted weakly supervised learning with normalcy suppression for anomalous event detection. In *European Conference on Computer Vision*, pages 358–376. Springer, 2020.
- [27] Muhammad Zaigham Zaheer, Arif Mahmood, Hochul Shin, and Seung-Ik Lee. A self-reasoning framework for anomaly detection using video-level labels. *IEEE Signal Processing Letters*, 27:1705–1709, 2020.
- [28] Tianzhu Zhang, Hanqing Lu, and Stan Z Li. Learning semantic scene models by object classification and trajectory clustering. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1940–1947. IEEE, 2009.
- [29] Jia-Xing Zhong, Nannan Li, Weijie Kong, Shan Liu, Thomas H Li, and Ge Li. Graph convolutional label noise cleaner: Train a plug-and-play action classifier for anomaly detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1237–1246, 2019.