Detection Anomaly in Video Based on Deep Support Vector Data Description

Bokun Wang,1 Caiqian Yang2, and Yaojing Chen3

1College of Civil Engineering and Mechanics, Xiangtan University, Xiangtan 411100, China
2School of Civil Engineering, Southeast University, Nanjing 210096, China
3Jiangsu Expressway Engineering Maintenance Co., Ltd., Huai'an 223005, China

Correspondence should be addressed to Caiqian Yang; ycqjxx@hotmail.com

Received 3 April 2022; Accepted 18 April 2022; Published 4 May 2022

Academic Editor: Baiyuan Ding

Copyright © 2022 Bokun Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Video surveillance systems have been widely deployed in public places such as shopping malls, hospitals, banks, and streets to improve the safety of public life and assets. In most cases, how to detect video abnormal events in a timely and accurate manner is the main goal of social public safety risk prevention and control. Due to the ambiguity of anomaly definition, the scarcity of anomalous data, as well as the complex environmental background and human behavior, video anomaly detection is a major problem in the field of computer vision. Existing anomaly detection methods based on deep learning often use trained networks to extract features. These methods are based on existing network structures, instead of designing networks for the goal of anomaly detection. This paper proposed a method based on Deep Support Vector Data Description (DSVDD). By learning a deep neural network, the input normal sample space can be mapped to the smallest hypersphere. Through DSVDD, not only can the smallest size data hypersphere be found to establish SVDD but also useful data feature representations and normal models can be learned. In the test, the samples mapped inside the hypersphere are judged as normal, while the samples mapped outside the hypersphere are judged as abnormal. The proposed method achieves 86.84% and 73.2% frame-level AUC on the CUHK Avenue and ShanghaiTech Campus datasets, respectively. By comparison, the detection results achieved by the proposed method are better than those achieved by the existing state-of-the-art methods.

1. Introduction

In order to improve the safety of public life and assets, video surveillance systems have been widely deployed in public places such as shopping malls, hospitals, banks, and streets. In most cases, how to detect video abnormal events in a timely and accurate manner is the main goal of social public safety risk prevention and control. Video abnormal events are defined as abnormal or irregular patterns in the video that do not conform to normal patterns. These incidents often include fights, riots, violations of traffic rules, trampling, holding arms, and abandoning luggage. However, due to the ambiguity of anomaly definitions, the scarcity of anomalous data, and the complex environmental background and human behavior, video anomaly detection is a major problem in the field of computer vision. In a nutshell, most of the current research work on video anomaly detection can be divided into two steps, such as feature extraction and normal model training [1]. Feature extraction can be achieved by manual technology or automatic feature extraction technology (representation learning or features based on deep learning). In normal model training, normal samples are used for learning, and then samples that do not conform to the learned model are judged as abnormal events. Then, the classification according to features can be divided into three different methods [2]. The first type is the trajectory-based methods [3]. This type of method obtains trajectory features by tracking the target. However, in dense scenes, the target tracking is a big problem. The second type of methods is based on global features [4, 5]. This type of method takes the video frame as a whole and extracts some low-level or middle-level features such as spatiotemporal

1. Introduction

In order to improve the safety of public life and assets, video surveillance systems have been widely deployed in public places such as shopping malls, hospitals, banks, and streets. In most cases, how to detect video abnormal events in a timely and accurate manner is the main goal of social public safety risk prevention and control. Video abnormal events are defined as abnormal or irregular patterns in the video that do not conform to normal patterns. These incidents often include fights, riots, violations of traffic rules, trampling, holding arms, and abandoning luggage. However, due to the ambiguity of anomaly definitions, the scarcity of anomalous data, and the complex environmental background and human behavior, video anomaly detection is a major problem in the field of computer vision. In a nutshell, most of the current research work on video anomaly detection can be divided into two steps, such as feature extraction and normal model training [1]. Feature extraction can be achieved by manual technology or automatic feature extraction technology (representation learning or features based on deep learning). In normal model training, normal samples are used for learning, and then samples that do not conform to the learned model are judged as abnormal events. Then, the classification according to features can be divided into three different methods [2]. The first type is the trajectory-based methods [3]. This type of method obtains trajectory features by tracking the target. However, in dense scenes, the target tracking is a big problem. The second type of methods is based on global features [4, 5]. This type of method takes the video frame as a whole and extracts some low-level or middle-level features such as spatiotemporal
gradients and optical flow. In a moderately crowded and dense environment, these methods can keep effective. The third type is the grid feature-based methods [6]. This type of method often divides the video frame into multiple small grids through dense sampling and then extracts the underlying features of a single grid because each grid can be individually evaluated. According to different normal model training methods, the present methods can also be divided into three different types. The first type is the cluster-based method [7]. This type of method is often based on an assumption that the normal sample belongs to a category or is relatively far from the cluster center. The abnormal samples do not belong to any category or are far away from the cluster center and then cluster the normal samples to build the model. The second is the method based on sparse reconstruction [8, 9]. This type of method assumes that the sparse linear combination of patterns can represent normal activities with the smallest reconstruction error. Because there is no abnormal activity in the training data set, it can represent abnormal patterns with a large reconstruction error. The third type is probabilistic model-based methods. This method considers that normal samples conform to a certain probability distribution, while abnormal samples do not conform to this distribution.

Recently, the latest progress of deep learning has proved the obvious advantages of deep learning-based methods in many computer vision applications [10]. As one of the tasks in computer vision, video anomaly detection is no exception. Different from traditional manual feature-based methods, deep learning methods often use pretrained networks to extract high-level features from videos or use existing network structures to establish end-to-end anomaly detection models based on normal models. For the former idea [11, 12], there is not much difference between the two steps of traditional abnormal event detection. For the latter idea [13–16], the two steps of feature extraction and model building are often jointly optimized in a deep network. These end-to-end deep networks include Auto-Encoder (AE), Deep Siamese Network (DSN), and Generative Adversarial Nets (GANs) [14, 17–20]. However, these network models are often designed for other tasks such as generative models and compression, rather than for anomaly detection tasks.

In the framework of deep learning, this paper proposes a new anomaly detection method based on Deep Support Vector Data Description (DSVDD) for anomaly detection tasks. Through DSVDD, not only can the smallest size data hypersphere be found to establish SVDD but also useful data feature representations and normal models can be learned. To this end, DSVDD uses a jointly trained deep neural network to map normal sample data to the smallest volume hypersphere. Then, in the test, the samples mapped inside the hypersphere are judged as normal, while the samples mapped outside the hypersphere are judged as abnormal. The RGB graph and the optical flow graph are composed of a 6-channel data and directly input into a DSVDD model; that is, it can detect the appearance abnormality and movement abnormality at the same time. The experimental results on the two public data sets of Avenue [9] and ShanghaiTec [17] show that the detection results of the method proposed in this paper are excellent, which exceed the state of the art.

## 2. Principle of Algorithm

The overall process of the method proposed in this paper is shown in Figure 1. In the training phase, the RGB images and optical flow diagrams of the training samples are intensively sampled and then merged into a 6-channel data to train the DSVDD model. In the testing phase, the RGB image and optical flow diagram composition of the video frame to be tested are also obtained after inputting the 6-channel data into the learned DSVDD model. It is determined whether the area is abnormal. In this section, the principle of SVDD is first briefly introduced, and then the training and testing process of video abnormal events based on DSVDD is described.

### 2.1. SVDD

SVDD is a description method based on boundary data (support vector). Its goal is to find a hypersphere that contains all or almost all training samples and has the smallest volume (the center is \( c \in \mathcal{F} \), and the radius is \( R > 0 \)). In fact, the SVDD optimization problem can be transformed as follows:

\[
\min_{R, \mathcal{W}} R^2 + \frac{1}{n} \sum_{i=1}^{n} \xi_i
\]

\[\text{s.t. } \left\| \phi_k(x_i) - c \right\|_{\mathcal{F}}^2 
\leq R^2 + \xi_i, \quad \xi_i \geq 0.\]

In (1), the slack variable allows a soft boundary; \( \xi_i \geq 0 \) and \( \nu \in (0, 1] \) are hyperparameters to control the balance between the penalty term and the volume edge of the hypersphere. Therefore, a point that falls outside the hypersphere, such as \( \left\| \phi_k(x_i) - c \right\|_{\mathcal{F}}^2 > R^2 \), is decided to be abnormal. SVDD has been widely used in fields such as anomaly detection, face recognition, speech recognition, image restoration, and medical imaging [21].

### 2.2. DSVDD

DSVDD learns a deep neural network \( \phi(\cdot; \mathcal{W}) \) with the weight \( \mathcal{W} \), so that the input normal sample space can be mapped to a hypersphere with the center and radius of the smallest. The normal sample is mapped in the hypersphere \( \mathcal{X} \subseteq \mathbb{R}^d \), and the abnormal sample is mapped on the hypersphere.

Specifically, for the sample area input space \( \mathcal{X} \subseteq \mathbb{R}^d \) and output space \( \mathcal{F} \subseteq \mathbb{R}^p \), a neural network with \( L \in \mathbb{N} \) hidden layers can project the input space to the output space \( \mathcal{X} \rightarrow \mathcal{F} \), where \( \mathcal{W} = \{W^1, W^2, \ldots, W^L\} \) are the weights of the hidden layers \( \ell \in \{1, 2, \ldots, L\} \) correspondingly. Then, \( \phi(x; \mathcal{W}) \in \mathcal{F} \) is the characteristic representation of the input sample \( x \in \mathcal{X} \). The goal of the DSVDD method is to jointly optimize the network weights \( \mathcal{W} \) and the output space to meet the minimum hyperspherical constraints of the center \( c \) and the radius \( R \). Then, given the training sample \( \mathcal{D}_n = \{x_1, x_2, \ldots, x_n\} \), the soft-boundary objective function of DSVDD is as follows:
The flow chart of video anomaly detection based on DSVDD.

\[
\begin{align*}
\min_{R, \mathcal{W}} & \quad R^2 + \frac{1}{n} \sum_{i=1}^{n} \max \left\{ 0, \| \phi(x_i; \mathcal{W}) - c \|_2^2 - R^2 \right\} \\
+ \lambda & \sum_{l=1}^{L} \| W_l \|_F^2.
\end{align*}
\] (2)

For (2), in the SVDD method, the minimization of \( R^2 \) means to minimize the volume of the hypersphere. The second item is the penalty items that are mapped out of the hypersphere through the neural network, such as those from the center of the hypersphere \( \| \phi(x_i; \mathcal{W}) - c \| > R \). The hyperparameter \( \nu \in (0, 1] \) controls the balance between the volume of the hypersphere and the deviation of the boundary, which allows certain points to be mapped to the outside of the sphere. The last item is the network parameter weight \( \mathcal{W} \) the attenuation regularization term, where \( \lambda > 0 \) and \( \| \cdot \|_F \) represents the Frobenius norm.

The optimization of (2) enables the network to learn weights \( \mathcal{W} \), so that the data points can be closely projected to the center of the hypersphere \( c \) nearby. For this reason, the deep network must extract the common factors of data changes. In fact, normal samples can often be mapped closer to the center of the hypersphere \( c \), while abnormal samples are mapped farther from the center or outside the hypersphere. In this way, a compact description of the normal model is obtained.

In actual tasks, it is often assumed that the training samples are all normal samples, so the objective function can be simplified to a single-class classification problem as follows:

\[
\min_{\mathcal{W}} \quad \frac{1}{n} \sum_{i=1}^{n} \| \phi(x_i; \mathcal{W}) - c \|_2^2 + \lambda \sum_{l=1}^{L} \| W_l \|_F^2.
\] (3)

DSVDD simply uses a secondary loss to punish the distance of each deep network representation \( \phi(x_i; \mathcal{W}) \) and \( c \). The second term is the regularization term of network parameter weight attenuation \( \mathcal{W} \), \( \lambda > 0 \). Equation (3) can also be regarded as a hypersphere with the smallest volume as the center. However, unlike Equation (2) using a soft boundary, Equation (3) shrinks the sphere by minimizing the average distance from the center of all data representations, instead of directly penalizing the radius and data representation that falls outside the sphere. Similarly, in order to map the samples as close to the center of the hypersphere as possible, the deep neural network must extract the changing common factors.

The weights \( \mathcal{W} \) of the neural network in DSVDD can be optimized by common back propagation methods (such as stochastic gradient descent). Because the network weight \( \mathcal{W} \) and hypersphere radius \( R \) are with different scales, it is impossible to optimize DSVDD with one learning rate. Therefore, it is necessary to alternately optimize the network weights \( \mathcal{W} \) and hypersphere radius \( R \) by the alternate minimization/block coordinate descent method.

2.3. Test Phase. Given test sample area \( x' \in \mathcal{X} \), the anomaly score can be calculated as follows:

\[
s(x') = \| \phi(x'; \mathcal{W}^*) - c \|_2^2.
\] (4)

where \( \mathcal{W}^* \) are the trained network model parameters. It is worth noting that network parameters can fully describe the DSVDD model. And predictions can be made without storing any data, so DSVDD has a very low storage complexity. Therefore, the computational complexity during testing is small.

In order to infer whether the test sample area is an abnormal sample, thresholds can be set on \( s(x') \) to make judgments as follows:

\[
s(x') \leq_{\text{normal}} \theta,
\] (5)

where \( \theta \) is the threshold that determines the sensitivity of the detection method in this paper.

3. Experiment

3.1. Dataset. This paper evaluates the performance of the DSVDD method on two publicly available data sets, i.e., the Avenue data set [9] and ShanghaiTech data set [17]. The Avenue data set is one of the most widely used benchmarks for video anomaly detection. It contains 16 training video clips and 21 test video clips, including 47 abnormal incidents that occurred on the streets of the Chinese University of Hong Kong. Each video is about 1 minute long and has a resolution of \( 640 \times 360 \). Normal events are walking on the street, and abnormal events include running, loitering, and throwing. ShanghaiTech data set [17] is one of the largest newly proposed datasets for video anomaly detection.
Unlike other data sets, the video clips in this data set come from 13 different cameras with different lighting conditions and camera angles. It has 330 training video clips and 107 test video clips containing 130 abnormal events. The resolution of the video frame is $856 \times 480$. Abnormal events in this data set include chasing and noise.

### 3.2. Evaluation Index

According to previous work [14], this paper calculates the frame-level receiver operating characteristic (ROC) curve and uses the area under the curve (AUC) score as an evaluation indicator. A higher AUC score indicates better anomaly detection performance. If an area in the video frame is judged to be abnormal, the frame is judged to be abnormal. We first obtain the anomaly scores of all video frames and then calculate the frame-level AUC scores.

### 3.3. Supplementary Details

For the two data sets, each frame is adjusted to a size of $320 \times 240$, and the optical flow image is calculated by the RAFT optical flow method provided in [22] through a network pretrained on the things data set. The original video frame and the calculated optical flow graph are combined into a 6-channel data, then cropped into $16 \times 12$ grid images according to the size of $20 \times 20$, and then input into DSVD for training and prediction. The deep neural network part of DSVD is in accordance with Conv (16, $3 \times 3$)-Leaky ReLU-ConvTran (32, $3 \times 3$)-BN-Leaky ReLU-ConvTran (64, $3 \times 3$)-BN-Leaky ReLU–FullyConnected64 structure. In the training phase, the batch size is set to 128, the initial learning rate is 0.0003, the weight decay is 0.0001, and the training is performed 1000 iterations.

### 3.4. Experimental Results

This section compares the proposed DSVD with the results obtained by several latest methods that only use positive sample training. These methods include Conv-AE [14], Stacked RNN [17], Unmasking [18], Davide et al. [19], Object-centric auto-encoder [20], MemAE [23], and New Baseline [24]. Table 1 lists the evaluation results of these methods on the frame-level anomaly detection on the two data sets [25–27].

On the Avenue dataset, the DSVD method proposed in this paper is superior to the results obtained by other methods, with an AUC score of 87.4%, which is 2.3% higher than the baseline method proposed in 2018 [24]. As far as we know, in terms of the frame-level AUC scores of all test videos in this data set, the DSVD proposed in this paper has achieved the best results. It is worth noting that the Object-centric auto-encoder [20] method achieved 89.3% of the frame-level AUC in their paper, but this is calculated through different indicators in their paper and the actual calculation of Object-centric. The frame-level AUC score obtained by the auto-encoder [20] method should be 86.5%, which is 0.9% lower than the method proposed in this paper.

On the ShanghaiTech dataset, the method DSVD proposed in this paper achieves a frame-level AUC score of 74.5%, which is 1.7% higher than the baseline method proposed in 2018 [24] and second only to Object-centric auto-encoder [20]. The method achieved 78.5%. The Object-
centric auto-encoder [20] method uses an object detection-based method for anomaly detection, and its performance largely depends on the output of its object detection algorithm. Therefore, detection-based methods cannot determine abnormal events that have not occurred before, and this often occurs in abnormal detection. Similarly, the MemAE method [23] requires the help of a pretrained pose estimator to achieve better results, so it is limited to detecting abnormal events related to people. In contrast, the DSVDD method proposed in this paper does not have this limitation and is very reliable when applied to various scenarios. Obviously, in addition to these two specially limited methods, the DSVDD method proposed in this paper is at least 1.7% ahead of other methods in frame-level AUC.

In Figure 2, some examples of abnormal score curves in the method proposed in this paper are shown, and some key frames with normal or abnormal events are given. Among them, the abscissa is the number of video frames, and the ordinate anomaly score has been normalized to 1. It can be seen that in the two data sets, the method proposed in this paper can correctly distinguish between normal and abnormal events. If an abnormal event occurs suddenly, such as running as shown in Figure 2(a), the abnormal score will increase sharply. If the abnormal event occurs slowly, as shown in Figure 2(b), the abnormal score will gradually increase. If the object that caused the abnormality disappears from the camera's field of view, the abnormality score will quickly decrease to close to 0.

4. Conclusion

In this paper, a video anomaly detection method based on DSVDD is proposed. DSVDD can be seen as a combination of deep learning and SVDD. It uses a jointly trained deep neural network to map normal sample data to the smallest volume hypersphere. Then, in the test, the samples mapped inside the hypersphere are judged as normal, while the samples mapped outside the hypersphere are judged as abnormal. A large number of experimental results on two public data sets show that the proposed method is significantly better than the existing methods, which proves the effectiveness of the anomaly detection method proposed in this paper. In the future, we will reduce the computational complexity on the basis of ensuring the accuracy of the algorithm and focus on improving the real-time performance of the algorithm to better apply it to actual scenarios. [25–27].

Data Availability

The datasets used in this paper can be accessed upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] W. Luo, W. Liu, D. Lian et al., “Video anomaly detection with sparse coding inspired deep neural networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, pp. 1–15, 2019.
[2] R. Nayak, U. C. Pati, and S. K. Das, “A comprehensive review on deep learning-based methods for video anomaly detection,” Image and Vision Computing, vol. 106, Article ID 104078, 2021.
[3] S. Cosar, G. Donatiello, V. Bogorn, C. Garate, L. O. Alvaress, and F. Bremond, “Toward abnormal trajectory and event detection in video surveillance,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, no. 3, pp. 683–695, 2017.
[4] K. Xu, X. Jiang, and T. Sun, “Anomaly detection based on stacked sparse coding with i classification strategy,” IEEE Transactions on Multimedia, vol. 20, no. 5, pp. 1062–1074, 2018.
[5] K. W. Cheng, Y. T. Chen, and W. H. Fang, "Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression[C]," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2909–2917, 2015.
[6] R. Leyva, V. Sanchez, and C.-T. Li, “Video anomaly detection with compact feature sets for online performance,” IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3463–3478, 2017.
[7] T. Zhao, F. Li, and P. Tian, "A deep-learning method for device activity detection in mMTC under imperfect CSI based on variational-autoencoder," IEEE Transactions on Vehicular Technology, vol. 69, no. 7, pp. 7981–7986, 2020.
[8] D. Xu, Y. Yan, E. Ricci, and N. Sebe, "Detecting anomalous events in videos by learning deep representations of appearance and motion," Computer Vision and Image Understanding, vol. 156, pp. 117–127, 2017.
[9] C. Lu, J. Shi, and J. Jia, “Abnormal event detection at 150 FPS in MATLAB,” in Proceedings of the IEEE International Conference on Computer Vision, pp. 2720–2727, Sydney, NSW, Australia, December 2013.
[10] Y. Zerrouki, F. Harrou, N. Zerrouki, A. Dairi, and Y. Sun, “Desertification detection using an improved variational autoencoder-based approach through ETM-landsat satellite data,” Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 202–213, 2021.
[11] X. Zhang, B. Hu, and Z. Pan, “Tensor representation based target detection for hyperspectral imagery,” Optical and Precision Engineering, vol. 27, no. 2, pp. 488–498, 2019.
[12] M. Ravanbakhsh, M. Nabi, H. Mousavi, E. Sangineto, and N. Sebe, "Plug-and-Play CNN for crowd motion analysis:an application in abnormal event detection," in Proceedings of the IEEE Winter Conference on Applications of Computer Vision, pp. 1698–1698, Lake Tahoe, NV, USA, March 2018.
[13] Y. Wang, B. Dai, G. Hua, J. Aston, and D. Wipf, "Recurrent variational autoencoders for learning nonlinear generative models in the presence of outliers," IEEE Journal of Selected Topics in Signal Processing, vol. 12, no. 6, pp. 1615–1627, 2018.
[14] M. Hasan, J. Choi, J. Neumanny, A. K. Roy-Chowdhury, and L. S. Davis, “Learning temporal regularity in video sequences,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, Las Vegas, Las Vegas, NV, June 2016.
[15] B. Ramachandra, M. J. Jones, and R. R. Vatsavai, “Learning a distance function with a Siamese network to localize anomalies in videos,” in Proceedings of the IEEE Winter Conference on Applications of Computer Vision, pp. 2598–2607, Snowmass Village, CO, March 2020.
[16] X. Wang, Z. Che, B. Jiang et al., "Robust unsupervised video anomaly detection by multi-path frame prediction," 2020, https://arxiv.org/abs/2011.02763.

[17] W. Luo, W. Liu, and S. Gao, "A revisit of sparse coding based anomaly detection in stacked rnn framework," in Proceedings of the IEEE International Conference on Computer Vision, pp. 341–349, Venice, Italy, October 2017.

[18] R. T. Ionescu, S. Smeureanu, B. Alexe, and M. Popescu, "Unmasking the abnormal events in video," in Proceedings of the IEEE International Conference on Computer Vision, pp. 2895–2903, Venice, Italy, October 2017.

[19] D. Abati, A. Porrello, S. Calderara, and R. Cucchiara, "Latent space autoregression for novelty detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 481–490, Long Beach, California, June 2019.

[20] R. T. Ionescu, F. S. Khan, M. I. Georgescu, and L. Shao, "Object-centric auto-encoders and dummy anomalies for abnormal event detection in video," in Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7842–7851, Long Beach, California, June 2019.

[21] R. Xie, N. M. Jan, K. Hao, L. Chen, and B. Huang, "Supervised variational autoencoders for soft sensor modeling with missing data," IEEE Transactions on Industrial Informatics, vol. 16, no. 4, pp. 2820–2828, April 2020.

[22] Z. Teed and J. Deng, "RAFT: recurrent all pairs field transforms for optical flow," in Proceedings of the European Conference on Computer Vision, pp. 402–419, Glasgow, Scotland, August 2020.

[23] R. Morais, V. Le, T. Tran, B. Saha, M. Mansour, and S. Venkatesh, "Learning regularity in skeleton trajectories for anomaly detection in videos," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 11996–12004, Long Beach, California, June 2019.

[24] W. Liu, W. Luo, D. Lian, and S. Gao, "Future frame prediction for anomaly detection—a new baseline," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6536–6545, Salt Lake City, UT, USA, June 2018.

[25] L. Weixin Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 1, pp. 18–32, 2014.

[26] Y. Chang, Z. Tu, W. Xie, and J. Yuan, "Clustering driven deep autoencoder for video anomaly detection," in Proceedings of the European Conference on Computer Vision, pp. 329–345, Springer, Glasgow, Scotland, August 2020.

[27] L. Ruff, R. A. Vandermulen, N. Gornitz et al., "Deep one-class classification," in Proceedings of the 35 th International Conference on Machine Learning, Stockholm, Sweden, July 2018.