Fine-Tuning a Pre-trained CAE for Deep One Class Anomaly Detection in Video Footage

Slim Hamdi[1,2(✉)], Hichem Snoussi[1], and Mohamed Abid[2]

[1] LM2S University of Technology of Troyes,
12, rue Marie Curie - CS 42060, 10004 Troyes Cedex, France
slim.hamdi@utt.fr
[2] CES Laboratory ENIS National Engineering School University of Sfax,
B.P. 3038, Sfax, Tunisia

Abstract. In recent years, abnormal event detection in video surveillance has become a very important task mainly treated by deep learning methods taken into account many challenges. However, these methods still not trained on an anomaly detection based objective which proves their ineffectiveness in such a problem. In this paper, we propose an unsupervised method based on a new architecture for deep one class of convolutional auto-encoders (CAEs) for representing a compact Spatio-temporal feature for anomaly detection. Our CAEs are constructed by added deconvolutions layers to the CNN VGG 16. Then, we train our CAEs for a one-class training objective by fine-tuning our model to properly exploit the richness of the dataset with which CNN was trained. The first CAE is trained on the original frames to extract a good descriptor of shapes and the second CAE is learned using optical flow representations to provide a strength description of motion between frames. For this purpose, we define two loss functions, compactness loss and representativeness loss for training our CAEs architectures not only to maximize the inter-classes distance and to minimize the intra-class distance but also to ensure the tightness and the representativeness of features of normal images. We reduce features dimensions by applying a PCA (Principal Component Analyser) to combine our two descriptors with a Gaussian classifier for abnormal Spatio-temporal events detection. Our method has a high performance in terms of reliability and accuracy. It achieved abnormal event detection with good efficiency in challenging datasets compared to state-of-the-art methods.

Keywords: Deep Learning · Anomaly detection · Convolutional Auto-Encoder

1 Introduction

Security is a founding value of any modern society, it contributes strongly to creating a climate of peace necessary for good social development. Currently,

C. Djeddi et al. (Eds.): MedPRAI 2020, CCIS 1322, pp. 3–17, 2021.
https://doi.org/10.1007/978-3-030-71804-6_1

the conditions and the various mechanisms for its implementation are major concerns, whether at the individual or collective level. In recent decades, cameras are used everywhere in public space for security purposes. Video surveillance is a system composed of cameras and signal transmission equipment. The use of video surveillance is an essential tool for fighting crime and strengthening security. It allows controlling the necessary conditions for security and the identification of the risked elements in the scene. In the current context, one operator is in charge of several scenes at the same time and may on the same screen. In [1], the author proves that an operator can miss 60% of target events when it is in charge of viewing 9 or more video streams. A possible solution to this problem would be the use of intelligent video surveillance systems. Theses systems will have to be able to learn the normal behavior of a monitored scene and detect any abnormal behavior that may represent a safety risk.

The AE auto-encoder is a fully connected and neural network widely used in unsupervised learning. It consists of an input layer, an output layer, and one or more hidden layers. The hidden layers are distributed between the encoder and the decoder, the encoder is used to encode the input data into a more compact representation, the decoder is used to reconstruct the data according to the representation generated by the encoder. To exploit its unsupervised learning capacity, the AE has been widely explored in the detection of abnormal events. The author in [2] proposes AMDN (Appearance and Motion DeepNet) a network consisting of three SDAEs (stacked denoising auto-encoders) a first trained to reconstruct patches extracted from normal images, a second trained with the optical flow representations corresponding to the patches and a third trained with the concatenation of the patches and their optical flow representations. Moreover, based on CAEs the author in [3] proposes to train a CAE for the reconstruction of 3D input volumes and the optical flux extracted from the image and the previous image. In, [4] compared two methods also based on CAEs. The first method suggests that a CAE should be trained to reconstruct low-level characteristics (HOG and HOF) extracted from samples in the normal class. In the second method, the authors propose to use a Spatio-temporal CAE trained on video volumes. In effect, in both approaches, the anomalies are captured using a regularity score calculated with the error of reconstruction. In recent years, many works exploit the progress that has been made in both areas of Deep Learning (DL) and Computer Vision (CV) to automate surveillance for abnormal events detection. Deep Learning automatizes the feature extraction from raw data to realize many purposes such as image classification [5], facial recognition [6], automatic generation of computer code [7], automatic natural language processing [8] and automatic speech recognition [9]. Unsupervised Deep Learning is often used in the field of anomaly detection not only due to the subjective aspect of the anomaly but also usually only normal data are available for training. The development of learning methods that do not require a labeled database has always been a primary objective in the field of automatic learning. In this perspective, many recent works have aspired to the development of deep one-class networks has have been proposed [35]. However, these methods

proposed to use an extra data set to ensure the compactness of normal features with a deep CNN. To remedy those drawbacks we propose in this paper, a new deep architecture for abnormal event detection. It consists of two convolutional auto-encodes, one formed on images and other on optical flow representations to obtain compact and descriptiveness features. This combination allows extracting high-level compact representation able to describe complex behaviors and dissociate between normal and abnormal events. In this paper, we propose new method based on a combination between auto-encoders to extract deep features contain both information about motion and shapes. The aim of this combination is to extract tight and representative spatio-temporal features of normal frames, and subsequently, these features are more easy to isolate it from abnormal frames. The originality of this work is to extract a deep spatio-temporal features of deep one class without using any external database.

2 Related

Anomaly detection in video footage is very import task in computer vision. Usually, state-of-the-art methods try to train a model to represent the normal events and labelled any new event at the testing phase that has small occurrence during the training as abnormal events. The earlier methods were proposed to extract low-level features to train a model, for example in [10], the author used the Histogram of Oriented Social Force (HOSF) to represent the events and in [11], the authors propose multiples features extraction such as size, color, and edges on small regions at any frame of input video obtained by foreground segmentation technique. Multiple classifier for each feature are exploited to decide if that region is contain anomaly or not. [12], use Histograms of Optical Flow (HOF) to represent the motion information of each frame enhanced by one class Support Vector Machine (SVM) classifier to pick up abnormal motion. In [13], the author propose to train a model from the available frames at the training using sparse coding and based on the assumption: "Usual events in a video footage are more reconstructible from a normal event dictionary compared to unusual events". The dictionary is obtain a model capable of computing normality score at each new event in order to dissociate normal and abnormal events. Moreover, other trajectory-based methods have been applied in order to recognize unusual trajectories in monitored scene. [14] propose to represent trajectories by Kanade Lucas-Tomasi Feature Tracker (KLT) and use Multi-Observation Hidden Markov Model (MOHMM) to determine if trajectory are normal or abnormal. [15] propose to train One-Class Support Vector Machine model to recognize the normal trajectories and pick up any abnormal events may occur. [16] combine two models; a vector quantization and a neural networks to extract robust representation. In last few years, several researchers based their works on deep learning. They have obtained greats results on various applications such as object detection [17], action recognition [18], face recognition [19]. This success come from to their capability to learn non-linear and complex representations from raw images, which is important because the real-world application contain many non-linear

relationships. These methods also have a good property of generalization: they can be applied on data unused during the learning process. The author of [20] propose to apply optical flow to extract spatial–temporal volumes of interest (SVOI) and use them to train a 3D - CNN to classify events into normal and abnormal. [21] combine pre-trained CNN completed with Binary Quantization Layer (BQL) and optical flow to detect local anomalies. [22] propose a method called AVID (Adversarial Visual Irregularity Detection) to detect and locate abnormalities in videos footage. A GAN composed of a generator trained to remove abnormalities in the input images and replace them with the dominant patterns of the same images and a discriminator in the form of an FCN that predicts the probability that the different regions (patches) of the input images will be abnormal. The two networks are trained in an adversarial manner and the abnormalities are simulated using Gaussian noise. After the training, each of the two networks is capable to detect abnormalities.

3 Proposed Method

3.1 Architecture

One-class classification is a machine learning problem that has received important attention by many researchers in different fields such as novelty detection, anomaly detection, and medical imaging. Nevertheless, the lack of data in the training phase reduces the depth of network architecture which in turn reduces the representativeness of features. To solve this weakness we propose to fine-tuning a pre-trained CAE for a one-class training objective constructed from VGG 16 CNN which is achieved 92.7% top-5 test accuracy. The database used to train VGG 16 CNN is ImageNet which is a dataset of over 14 million high-resolution images belonging to 1000 classes. The images were collected from the web and labeled by humans using Amazon's Mechanical Turk crowd-sourcing tool. We freeze the first layers of convolutions to properly exploit the richness of the database with which the CNN was trained (Fig. 1). The objective of the convolution operation is to extract the high-level features from the input image. Our architecture need not be limited to only one convolution layer. Conventionally, the first convolution layer is responsible for capturing the Low-Level features such as edges, color, gradient orientation, etc. With added layers, the architecture adapts to the High-Level features as well, giving us a network that has the wholesome understanding of images in the dataset, similar to how we would. So, we construct the encoder part of our CAE architecture based on convolutions layers of pre-trained CNN VGG16. We freeze the first convolutional block of VGG 16 and we keep the others convolutional blocks trainable (Fig. 2). In the hand, the decoder part is a plane network made up of four 2D-deconvolution layers to be able to reconstruct the original frames, Its hyper-parameters is given in (Table 1).

Similar to the traditional auto-encoder, the CAE is composed of two parts. The encoder part which is a sequence of convolutional layers aims to extract compressed data of input image at the bottleneck layer and the decoder part

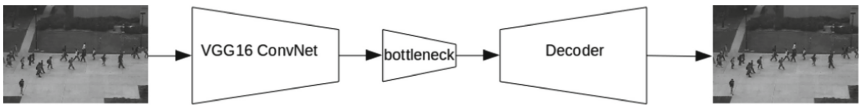

Fig. 1. 2D-CAE based on pre-trained CNN VGG16 ConvNet

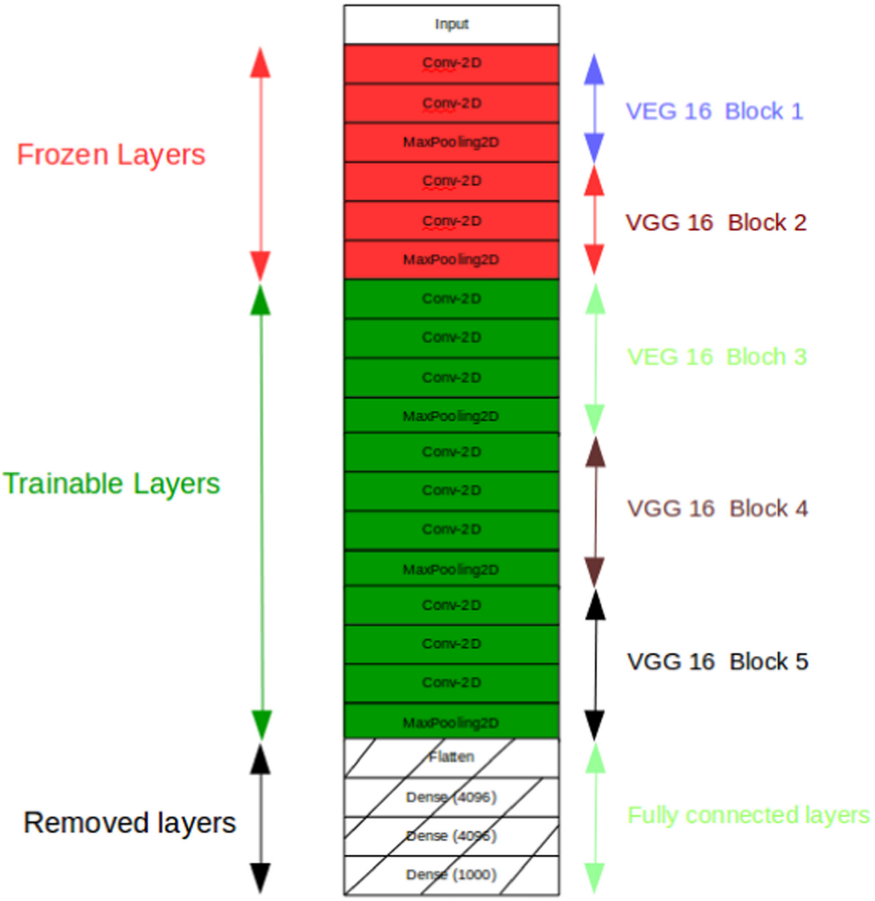

Fig. 2. VGG 16 architecture used for fine-tuning one class objective

which is successive of deconvolutional layers aims to reconstruct the input data from compressed data at bottleneck layer. The CAE can reconstruct better the data with was trained than the data that have ever seen, so the bottleneck layer must be reduced and representative as possible which in reality presents a compromise, many tests are done to select properly the bottleneck dimension (Table 1). A non-linear activation function is used at the convolutional and deconvolutional layers to obtain more useful and robust representations, except

Table 1. Hyper parameter of added layers

Input size	Layer type	Filter number	Kernel size	Strides	Activation	Output size
[7, 7]	2D-convolution	512	[3,3]	[2,2]	Relu	[3, 3]
[3, 3]	2D-deconvolution	256	[5,5]	[3,3]	Relu	[11,11]
[11, 11]	2D-deconvolution	128	[5,5]	[2,2]	Relu	[35,35]
[35, 35]	2D-deconvolution	96	[7,7]	[2,2]	Relu	[109,109]
[109, 109]	2D-deconvolution	1	[8,8]	[2,2]	linear	[224, 224]

for the last deconvolution layer we used linear activation function due to the range of our input data which is [−255, 255]. Our architecture consists of two parallel CAEs constructed as mentioned above. The first CAEs are trained on original images to be able to detect any abnormalities in shapes and the second CAEs are trained on optical flow representation aim to detect any abnormal motion relative to training (Fig. 3).

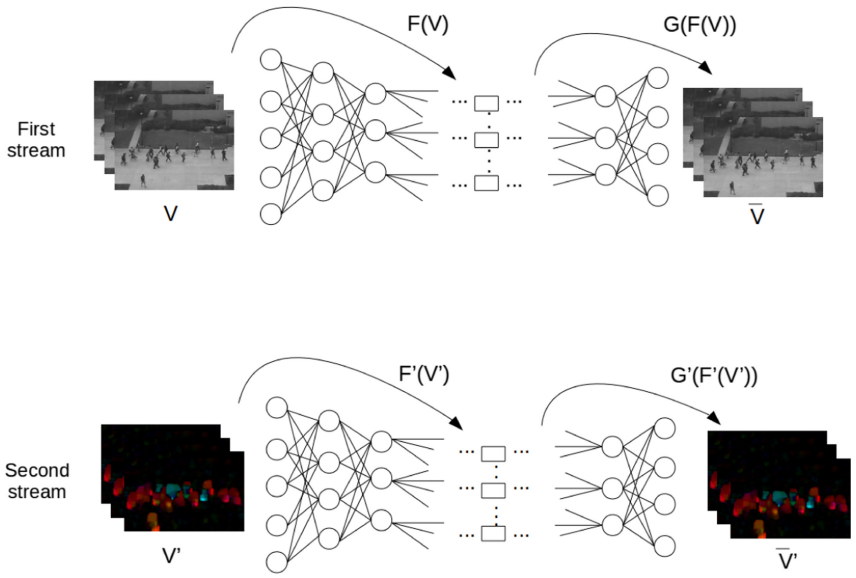

Fig. 3. Two stream learning

3.2 Training

The training phase aims to obtain a model capable to get representative and compact features of normal images for easy classification. We can ensure that in two methods; the first method (Fig. 4) is to do training in cascade objectives by training only at the beginning with the reconstruction objective and after a few

epochs we extract a representative point denoted "c" of features of the dataset which with our model is training at bottleneck layer as the mean of features. Then, we do training only with the compactness objective and we fix the point c as the target of our new features. The disadvantage of this training method is that the representativeness of the images is not robust but it gives very compacted features. To remedy this flaw, a second training method is proposed with pseudo-parallel objectives (Fig. 3), we start the training with only reconstruction objective then as we have done at the first method we extract a fixed point "c" as the target of features then we continue the training with both compactness and reconstruction objectives to get robust model.

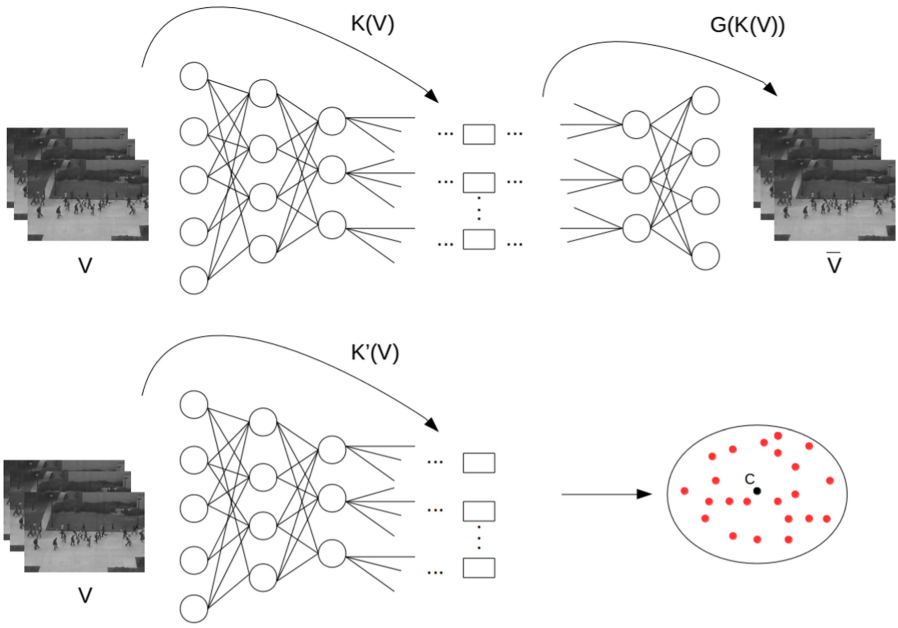

Fig. 4. The first training method: Cascade objectives

During the training phase (Fig. 5), both 2D-CAE are trained, one is trained with a stream of a sequence of original images and the other is trained with a stream of a sequence of optical flow representation. The optical flow is the pattern of apparent motion of image objects between two consecutive frames caused by the movements of the object. We have used a color code for better visualization. (Figure 6) shows some samples of images and optical flow images.

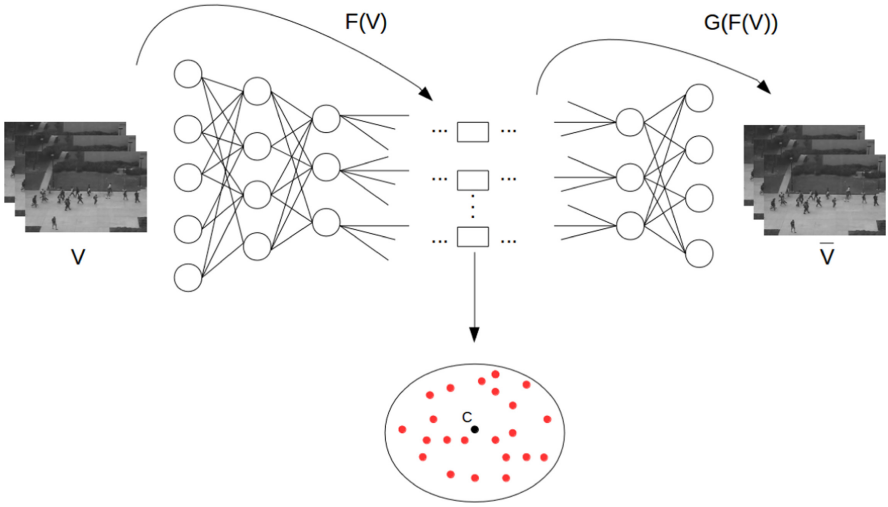

Fig. 5. The second training method: pseudo-parallel objectives

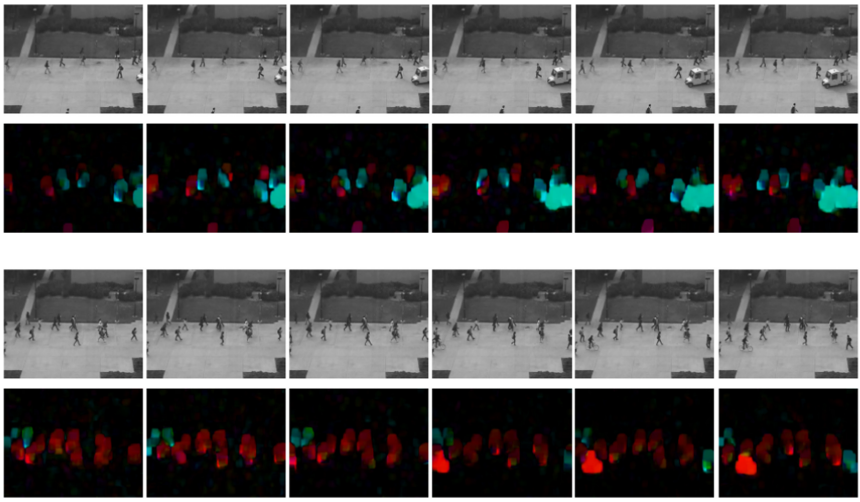

Fig. 6. Examples of optical flow representations and original images

Representativeness Loss: $\mathbf{L}_r$ The aim of representativeness loss is to evaluate the capacity of the learned feature to generalize normal class. The representativeness loss increases the capacity of our model to raise the distance inter-classes.

$$L_r = \frac{1}{n}\sum_{i=1}^{n}(V - \hat{V}) \tag{1}$$

Compactness Loss: $\mathbf{L_c}$ The objective of compactness loss is to tight all the features used during the training phase belonging to the normal class. Compactness loss evaluates the similarity between each feature vector and the fixed point 'C'. It is used to decrease the intra-class variance of the normal class.

$$L_c = \frac{1}{n}\sum_{i=1}^{n}(F(V) - M) \tag{2}$$

To perform back-propagation using this loss, it is necessary to assess the contribution each element of the input has on the final loss. For each ith sample F(V)= $\{Fv_{i1},Fv_{i2},...Fv_{ik}\} \in R^k$ and the fixed point as m_i= $\{m_{i1},m_{i2},...m_{ik}\}$, we define the gradient l_c with respect to the input Fv_{ij} is given as,

$$\frac{\partial L_c}{\partial Fv_{ij}} = \frac{2}{(n-1)n_k}[n \times (Fv_{ij}) - \sum_{k=1}^{n}(Fv_{ik} - m_{ik})] \tag{3}$$

3.3 Testing

The proposed testing procedure aims to classify features of testing images as normal or abnormal based on the Mahalanobis distance threshold. Both motion and shapes features vectors noted respectively F(v)= $\{Fv_{i1},Fv_{i2},...Fv_{ik}\} \in R^k$and F'(v')= $\{F'v'_{i1},F'v'_{i2},...F'v'_{ik}\} \in R^k$ are extracted from trained encoders parts to be concatenated into one vector. Then we apply PCA to this vector to reduce dimension and to extract important information noted X = PCA ([F(V); F'(V')] = $\{X_{i1},X_{i2},...X_{ik},X_{i1}\} \in R^p$ when p < 2 × k (Fig. 7). Using PCA is made the calculation of the covariance matrix Q faster and not complicate.

For each new feature vector X_{test} we calculate a Mahalanobis distance between each feature vector and $\bar{X}$ as given:

$$d = (X_{test} - \bar{X}) \times Q^{-1} \times (X_{test} - \bar{X})^t \tag{4}$$

When $\bar{X}$ as the mean of X $\in R^p$ and Q $\in R^{p\times p}$ as its covariance.

The classification process is carried out according to the following process: In the first step, we extract feature vectors X =$\{x_i\}, x_i \in R^{512}$ from the normal training examples, the mean M and the inverse of the covariance matrix Q of X are then calculated. In the second step, we evaluate each feature vector x_j of the testing frames with Mahalanobis distance d_j using M and Q. This is represented in the following equation:

$$d\_j = (x\_j - M) * Q * (x\_j - M)' \tag{5}$$

The outlier vectors, which actually represents abnormal frames, are then picked by thresholding the distance. If the distance exceeds a threshold α, the vector x_j is considered as outlier and the frame p_j is labeled as abnormal, Eq. (6).

$$p_j : \begin{cases} Normal \quad if \quad d_j \leq \alpha \\ Abnormal \; if \quad d_j > \alpha \end{cases} \tag{6}$$

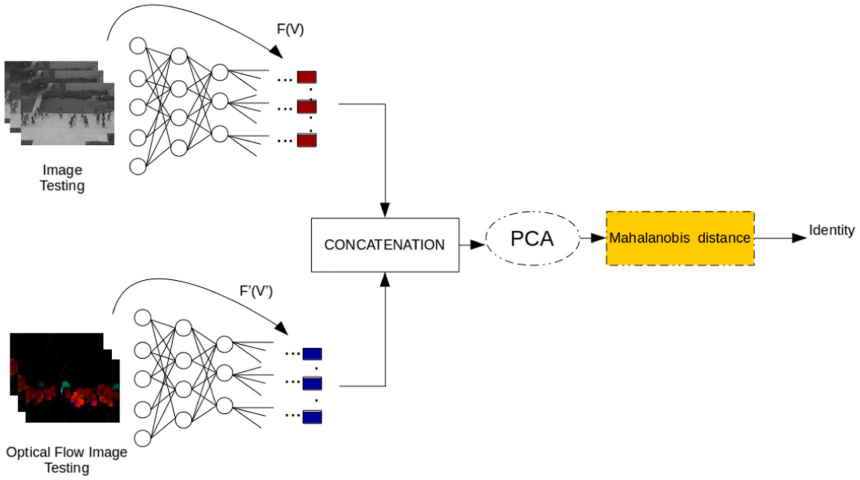

Fig. 7. Classification flowcharts

4 Experimental Results

UCSD Peds2 and UMN are challenging anomaly detection datasets. Both of them contain normal events like people are walking and abnormal events like the walking movement of bikers, skaters, cyclists, and small carts in the case of Ped2, and people are running in the case of UMN. Ped2 contains 16 training and 12 testing video samples and provides frame-level ground truth to evaluate the detection performance by comparing our method with others stat- of-the-art anomaly detection methods. In the other hand, The UMN dataset has consisted of 3 scenes: lawn (1450 frames), indoor (4415 frames) and plaza (2145 frames) and the ground truth is provided in the video frames that need to be extracted to evaluate the performance.

We evaluate our different methods using (Error Equal Rate) EER and (Area Under Curve ROC) AUC as evaluation criteria. A smaller EER corresponds with better performance. As for the AUC, a bigger value corresponds with better performance.

Our two methods have the same results nearly, with a little advantage for the pseudo-parallel objectives method.

It proves the robustness to occlusion and high performance in anomaly detection compared with state-of-the-art methods. To visualize the important effect of the compactness loss function we extract from each feature extracted by our architecture two components by applying the PCA. These components are named later features for visualization. Figure 8 illustrates the results, just to better understand its effects, we will categorize our database into three classes.

– Normal images contains only normal events as mentioned in ground truth, this class represented by green points in Fig. 8.

– Confused images when a portion of anomaly start to appear and not a whole of the anomaly enter in the scene, this class is presented by blue points in Fig. 8.
– Abnormal images when a more of the half of anomaly enter in the scene, this class represented by red points in Fig. 8.

The Fig. 8 1.a represents features for visualisations of our architecture trained with only representativeness loss, as we can see in this figures each of three classes reserved a region of space. Which is mean representativeness loss has increased the inter-classes distance between the three classes in an unsupervised way and using only the class of normal images (Class one). In order to decrease the intra-class distance for normal image we have used compactness loss. The Fig. 2 1.b represents features for visualisations of our architecture trained with both representativeness loss and compactness loss. In this case, the normal images not only are reserved region in space but also are very tight and easy to separate from abnormal images.

Combining the two CAEs have decreased the EER from 17% to 11% which make the importance of using of optical flow image to represent the motion in each frames. The Table 2 shows our results on Ped2 dataset and proves the robustness of our method compared to others state of the art methods.

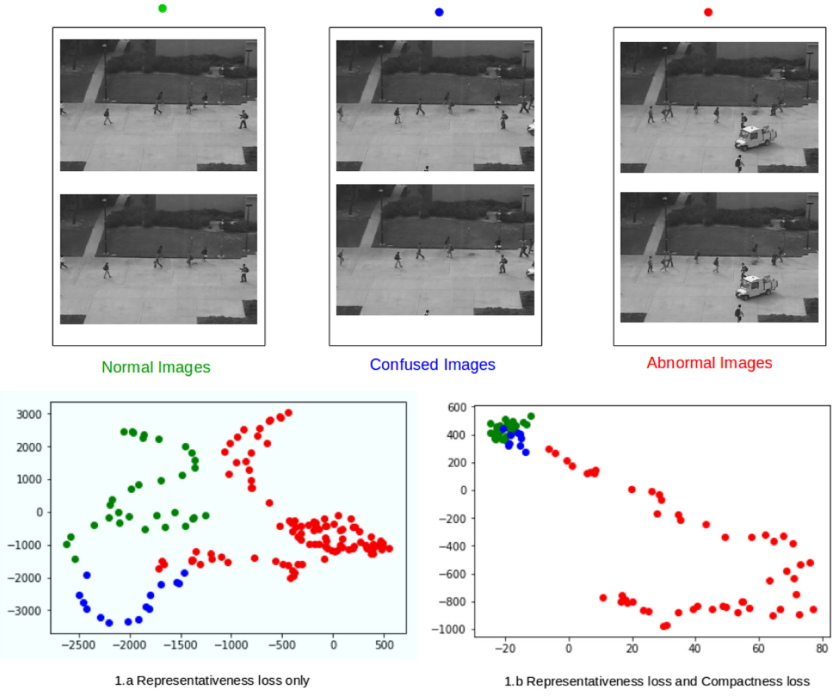

Fig. 8. Compactness loss importance

Table 2. EER comparison of UCSD Peds2

Method	EER
Mehran. [23]	42.00%
Kim (MPCCA). [24]	30.00%
Bertini. [25]	30.00%
Zhou. [26]	24.40%
Bouindour. [27]	24.20%
Hamdi. [28]	14.5%
Li. [29]	18.50%
Chong. [30]	12.00%
Tan Xiao. [31]	10.00%
Ours (Cascade)	**12%**
Ours (pseudo parallel)	**11%**

Table 3. Results in UMN dataset

Scene	EER	AUC
Lawn	3.17%	99.23%
Indoor	1.92%	99.37%
Plaza	1.11%	99.80%

Table 4. ERR comparison of UMN dataset

Method	EER
Mehran. [23]	12.60%
Chaotic invariants [32]	5.30%
Li. [29]	3.70%
Saligrama et al. [33]	3.40%
Sparse. [34]	2.80%
Ours	**2.28%**

Our results in scene of UMN is presented in the following table:

This table shows our results relatively at each scene. Despite that our model is trained on different scenes. It proves that our method have good efficiency for anomaly detection (Table 3).

This table shows our results for UMN dataset, in this case we use one threshold for whole the dataset and its independent to the scenes. It proves that our method have good efficiency and robust for variation of scenes (Table 4). This figure is plotted with tools from python library sklearn.metrics and roc_curve. It proves that our architecture achieve more then 99% of AUC.

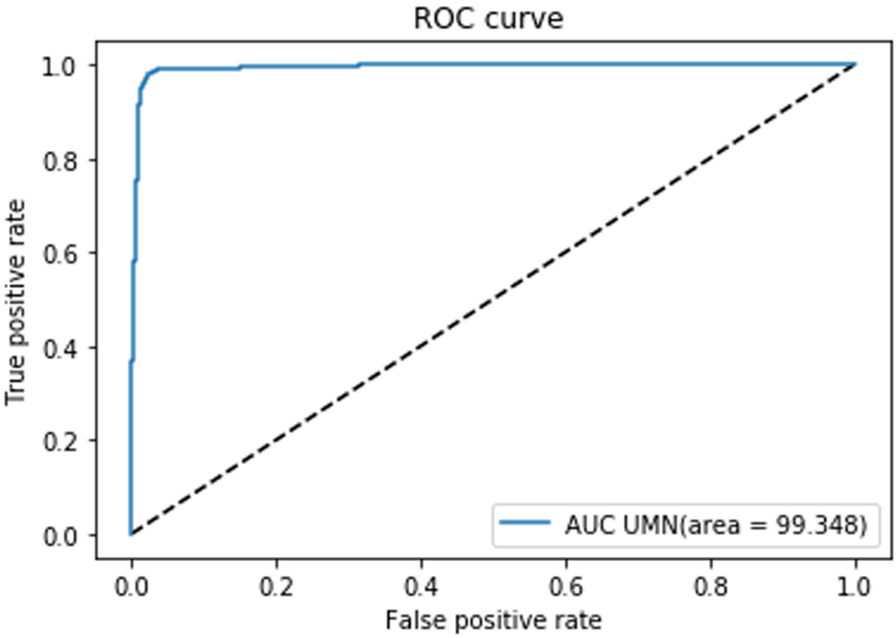

Fig. 9. ROC Curve of UMN dataset

5 Conclusion

In this paper, a new unsupervised methods were proposed to train CAEs for the Deep One-Class objective. We used these methods to learn a new architecture composed of two CAEs, one trained on video volumes and the second on optical flow representations. Our two networks allow extracting high-level Spatio-temporal features taking into account the movements and shapes present in each small region of the video. This robust representation makes possible, with a simple classifier, to differentiate between normal and abnormal events. We have tested our network on challenging datasets, containing crowded scenes (USCD Ped2 and UMN) Our method obtained high results competing with the best state-of-the-art methods in the detection of abnormal events (Fig. 9).
Our future works will investigate the strengthening of our learning process and apply our model on drone video for anomaly detection.

References

1. Beltramelli, T.: Generating code from a graphical user interface screenshot. In: Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems, pp. 329–343 (2018)
2. Xu, D., Ricci, E., Yan, Y., Song, J., Sebe, N.: Learning deep representations of appearance and motion for anomalous event detection (2015)

3. Gutoski, M., Aquino, N.M.R., Ribeiro, M., Lazzaretti, E., Lopes, S.: Detection of video anomalies using convolutional autoencoders and one-class support vector machines (2017)
4. Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A.K., Davis, L.S.: Learning Temporal Regularity in Video Sequences, pp. 733–742 (2016)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
6. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)
7. Beltramelli, T.: Generating code from a graphical user interface screenshot. In: Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing (2018)
8. Conneau, A., Schwenk, H., Barrault, L., Lecun, Y.: Very deep convolutional networks for natural language processing, vol. 2 (2016). arXiv preprint. arXiv:1606.01781
9. Amodei, D., et al.: Deep speech 2: end-to-end speech recognition in English and Mandarin. In: International Conference on Machine Learning, pp. 173–182 (2016)
10. Yen, S., Wang, C.: Abnormal Event Detection Using HOSF, pp. 1–4 (2013)
11. Reddy, V., Sanderson, C., Lovell, B.C.: Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture, pp. 55–61 (2011)
12. Wang, T., Snoussi, H.: Detection of abnormal visual events via global optical flow orientation histogram. IEEE Trans. Inf. Forensics Secur. **9**(6), 988–998 (2014)
13. Zhao, B., Fei-Fei, L., Xing, E.P.: Online detection of unusual events in videos via dynamic sparse coding. In: CVPR 2011, pp. 3313–3320 (2011)
14. Zhou, S., Shen, W., Zeng, D., Zhang, Z.: Unusual event detection in crowded scenes by trajectory analysis. In: 2015 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 1300–1304 (2015)
15. Piciarelli, C., Micheloni, C., Foresti, G.L.: Trajectory-based anomalous event detection. IEEE Trans. Circuits Syst. Video Technol. **18**, 1544–1554
16. Johnson, N., Hogg, D.: Learning the distribution of object trajectories for event recognition. Image Vis. Comput. **14**(8), 609–615 (1996). ISSN 0262–8856. https://doi.org/10.1016/0262-8856(96)01101-8
17. Redmon, J., Farhadi, A.: YOLOv3: An Incremental Improvement (2018)
18. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3D convolutional networks. In: The IEEE International Conference on Computer Vision (ICCV) (2015)
19. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
20. Zhou, S., Shen, W., Zeng, D., Fang, M., Wei, Y., Zhang, Z.: Spatial-temporal convolutional neural networks for anomaly detection and localization in crowded scenes. Sig. Process. Image Commun. **47**, 358–368 (2016)
21. Ravanbakhsh, M., Nabi, M., Mousavi, H., Sangineto, E., Sebe, N.: Plug-and-Play CNN for crowd motion analysis: an application in abnormal event detection. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (2018)
22. Sabokrou, M., et al.: Avid: adversarial visual irregularity detection. arXiv preprint arXiv:1805.09521 (2018)

23. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: Computer Vision and Pattern Recognition, pp. 935–942 (2009)
24. Kim, J., Grauma, K.: Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates. In: Computer Vision and Pattern Recognition, pp. 2921–2928 (2009)
25. Bertini, M., Del Bimbo, A., Seidenari, L.: Multi-scale and real-time non-parametric approach for anomaly detection and localization. Computer Vis. Image Underst. **116**(3), 320–329 (2012)
26. Zhou, S., Shen, W., Zeng, D., Fang, M., Wei, Y., Zhang, Z.: Spatial-temporal convolutional neural networks for anomaly detection and localization in crowded scenes. Signal Process. Image Commun. **47**, 358–368 (2016)
27. Bouindour, S., Hittawe, M.M., Mahfouz, S., Snoussi, H.: Abnormal event detection using convolutional neural networks and 1-class SVM classifier. In: 8th International Conference on Imaging for Crime Detection and Prevention (ICDP 2017) (2017)
28. Hamdi, S., Bouindour, S., Loukil, K., Snoussi, H., Abid, M.: Hybrid deep learning and HOF for Anomaly Detection. In: 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), pp. 575–580 (2019)
29. Li, W., Mahadevan, V., Vasconcelos, N.: Anomaly detection and localization in crowded scenes. IEEE Trans. Pattern Anal. Mach. Intell. **36**(1), 18–32 (2014)
30. Chong, Y.S., Tay, Y.H.: Abnormal event detection in videos using spatiotemporal autoencoder. In: Proceedings CVPRR in International Symposium on Neural Networks, pp. 189–196 (2017)
31. Xiao, T., Zhang, C., Zha, H.: Learning to detect anomalies in surveillance video. IEEE Sig. Process. Lett. I **22**(9), 1477–1481 (2015)
32. Wu, S., et al.: Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. In: IEEE Conference on Computer Vision Pattern Recognition, pp. 2054–2060 (2010)
33. Saligrama, V., Chen, Z.: Chaotic invariants based on local statistical aggregates. J. IEEE Conf. Comput. Vis. Pattern Recogn. 2112–2119 (2012)
34. Cong, Y., et al.: Sparse reconstruction cost for abnormal event detection. In: IEEE Conference on Computer Vision Pattern Recognition, pp. 3449–3456 (2011)
35. Perera, P., Patel, V.M.: Learning deep features for one-class classification. In: IEEE Conference on Computer Vision Pattern Recognition, pp. 3449–3456 (2011)