Deep Reinforcement Learning for Real-world Anomaly Detection in Surveillance Videos

1st Sabrina Aberkane
Ecole Superieur en Informatique 08 Mai 1945 de Sidi Bel Abbes , Algeria
Mostaganem, Algeria
s.aberkane@esi-sba.dz

2nd Mohamed Elarbi Boudihir
Ecole Superieur en Informatique 08 Mai 1945 de Sidi Bel Abbes , Algeria
Sidi Bel Abess, Algeria
m.elarbiboudihir@esi-sba.dz

***Abstract*—Surveillance videos are considered as a very important mechanism in a smart city project. In recent years, a method called deep learning was combined with reinforcement learning techniques to learn useful representations for the problems with high dimensional raw data input. In this paper, we develop a Deep Q Learning Network (DQN) to localize anomalies in videos by enabling the agent to learn how to detect and recognize the abnormalities in videos. Our idea is inspired by multiple instance learning (MIL) techniques based on common share features with reinforcement learning. We consider normal and abnormal videos as bags and the selection of videos clips as actions. In our DQN architecture we will design a fully connected layer, which compute probability for each video segment in both positive(anomalous) and negative(normal) bags indicating how likely a clip is containing an anomaly. Our method is applied to a new large-scale dataset of 128 hours of videos called UCF-Anomaly-Detection-Dataset, it is about of 1900 long and untrimmed real-world surveillance videos, with 13 cases of realistic anomalies.**

***Index Terms*—Deep Learning; Reinforcement Learning; Action Recognition; Surveillance Videos.**

I. Introduction

To ensure human safety, there is an immediate need for surveillance cameras to increase safety in public places e.g. banks, intersections, stores, streets, parking, etc. However, mounting of security cameras is cheap, but there is a flagrant gap in the utilization of surveillance cameras and an elevated cost of finding available human resources to observe the output. The ultimate goal in video surveillance is to prevent dangerous situations by detecting anomalous events such as road accidents or fighting. Generally What is needed is continuous 24-hour monitoring of surveillance video to alert security officers. Thus, the computer vision community developed an intelligent agent for automatic action videos recognition in order to alleviate the waste of labor and time. Such an intelligent system would be highly beneficial as it detects and alerts timely an activity that deviates normal patterns.

The difficulty in action recognition task is to give a formal and general definition of all real-world anomalous events, for instance, the solution developed in [1], [2] for violence and aggression detection cannot be generalized to detect other abnormal events. In other words, the anomaly detection algorithms rely on prior information about anomalous activities occurring in a specific environment.

In recent years there have been many successes of using anomaly detection system in surveillance videos to resolve security issue due to several attempts to signal violent activities in videos. For example, Gao and al introduced violence detection framework using oriented violent flows in crowd scenes. In [2] they proposed a system for aggressive actions detection using video and audio data. More recently, Waqas and al. [3] proposed to learn anomaly through the deep multiple instance learning framework by leveraging weakly labeled training videos. Various approaches were introduced such as context-driven model [4], social force models [5], motion patterns [6], Hidden Markov Model on local spatio-temporal volumes [7], Histogram-based methods to learn global motion patterns [8], beyond these approaches that consider anomalies as being low probable patterns comparing to the distribution of normal motion patterns, a method based on trajectory model is introduced in [9], [10], the concept is to track the normal motion of people to detect diversion from the general frequent model as an abnormality.

The advent of deep learning concepts has had a significant impact on action recognition area, several methods have been used Deep neural Network [11]–[14]. However, the annotation for training remains a barrier to obtain good performance, specifically for videos.

II. Proposed Approach

As we mentioned above, it is very difficult or improbable to define general patterns which cover all possible normal/abnormal motions. In addition, in daily life, the same behavior could be interpreted differently depending on a defined context. For example, in our approach, we propose to learn anomalous events through a deep Q learning Network framework by taking advantage of the Action/Reward process

978-1-7281-3156-6/19/$31.00 ©2019 IEEE

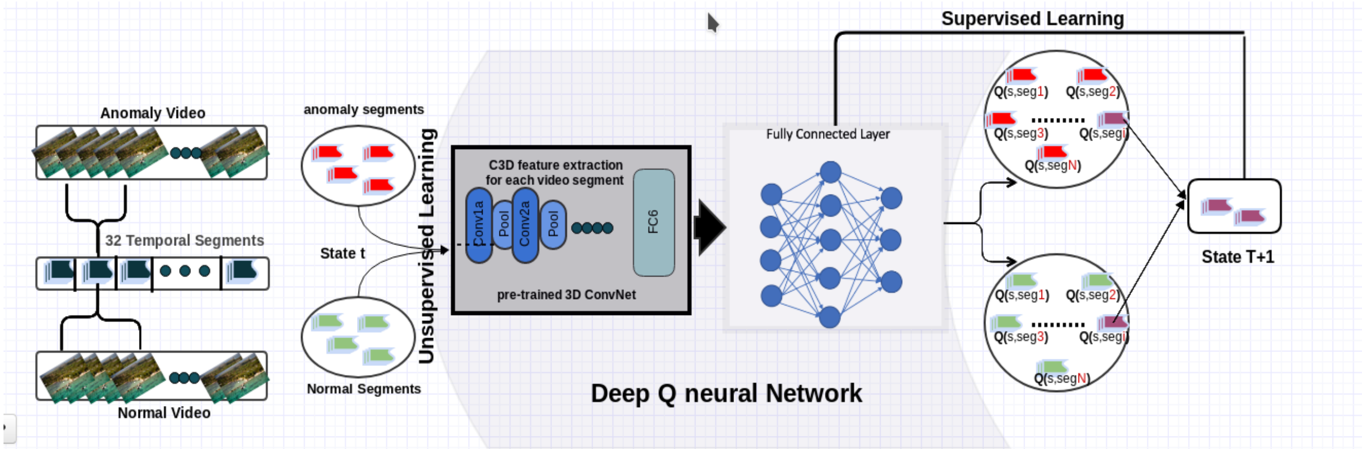

Fig. 1. The flow diagram of the anomaly detection method using deep reinforcement learning. States represent a set of video segments X_i, action a_i represents the decision of selecting a segment having the highest Q-value $Q(S_t, X_i)$. Action a_i will lead to specific state $t+1$ containing two segments: the first from a positive video and second from a negative video.

of MDP (Markov Decision Process). Fig. 1 illustrates the flow diagram of our proposed model. It consists of two main components: a C3D for features extraction from all segments in video and Deep Q Neural Network responsible for detecting segments, which contain anomalous behavior.

A. Space-time features extraction: C3D

In recent years, the 3D Convolutional is considered as one of the most successful techniques that learned action features by deep learning tools, 3D ConvNets is probably the most intuitive to capture temporal dynamics in a short period of time. It is mainly used to generate directly hierarchical representations of spatiotemporal data as shown in different works [15]–[17]. In our approach, we will use the implementation of a modern deep architecture named C3D discussed in [18] to extract features from our used video dataset. Still, C3D was implemented to support 3-Dimensional Convolutional Networks, the authors in [18] proved the C3D efficiency by making the experiments on different datasets e.g. Sports-1M, UCF10, HMDB51 dataset, it also provides C3D pre-trained model which were trained on Sports-1M dataset with necessary tools for video features extraction.

B. Deep Reinforcement Learning

Reinforcement Learning is an exciting field of machine learning, which has been proposed in recent years. The concept is based on the understanding of the humans decision-making process [19]. The algorithm aims to enable an agent to decide the right behavior by trial-and-error while receiving feedback form of reward signals. Deep reinforcement learning [20] is a combination of reinforcement learning and deep learning to settle more complex tasks. Solving such tasks involve dealing with high-dimensional data and environments, sparse reward signals, and uncertainties in the agents observations. DRL approaches have already been applied to a wide range of problems, such as the vision control in robotics [21], where control policies for robots can now be learned directly from camera inputs in the real-world problems. More recently, the authors in [22] proposed a framework named attention-aware deep reinforcement learning (ADRL) based on deep reinforcement learning to find the focus of attention in face videos for person recognition. Zhou *et al.* [23] proposed a new deep summarization network for video summarization process based on a new reward formula called Diversity-Representativeness Reward.

Generally, the idea behind using reinforcement learning to decision-making problems is that an agent will be able to learn from the environment by interacting with it and receiving rewards for performing actions, it comes from the natural humans' process to learn through their experiences. Usually, the reinforcement learning (Rl) model is built as a loop that works as follows:

- An agent, which receives State S_0 from the environment.
- Based on that state S_0, this agent takes an action (randomly).
- Environment transitions to a new state S_1.
- Environment gives some reward R_1 to the agent.

This loop outputs a sequence of state, action, and reward. The aim of the agent is to maximize future reward.

C. Deep Action Recognition Network

In real life, anomalous events rarely happen compared to normal actions, thats why mostly, the anomaly detection is handled as low likelihood pattern detection [13], [24] due to the unavailability of sufficient anomaly instances. Otherwise, the authors in [3] treat the anomaly detection as a regression problem by encouraging high scores for anomalous sequences in videos.

In our work, we handle the anomaly detection as Markov Decision Process (MDP) and introduce a deep reinforcement learning approach to train the evaluation network. We aim to create an agent possessing the ability to decide if some video contains any abnormality through a set of normal/abnormal videos. Based on the idea that in real-word anomaly activities happen only for a short moment and only a few segments may contain the anomaly event through the whole video, we consider the process of finding the abnormality in the video as the process of finding the most representative segments from the video. Hence, we present a deep model to evaluate the situation of video segment X_i and calculate its anomaly weight Q_i:

$$Q_i = C\big(S_t, X_i\big) \tag{1}$$

$$Q_i^* = C_1\big(S_t, X_i\big) \tag{2}$$

Where S_t represent the state at time t regrouping all segments of normal/abnormal videos, Since we only know the video-level label, i.e we only know that the video is abnormal, but we dont know exactly in which segment. Thus, we train our network C without additional supervision, we consider pre-trained network C_1 as an expert in action recognition and build an algorithm to supervise C by C_1 as the recognition execution of C_1 outputs the anomaly score of input video segments [3].

To select the most representative segment from the video, we can employ two different strategies. One is to immediately score each segment and then determine the most representative segment from each video. The other is to drop the worst segment progressively, and the last remaining segment is considered as the most representative one. It is evident that locate the worst segment from a video is less difficult and more effective than directly finding the most important sequence since we don't have the segment-level labels to obtain a good value for Q_i.

Therefore, we choose to take the second strategy to allow our agent to learn from each transition in a given video. We adopt Deep Q network architecture to determine the most representative clip in a set of a video segment. We designate a set of segments in a video as state S_t which is defined as:

$$\forall X \in V_{n,a}, \quad S_t = X_1, X_2, X_3 \ldots\ldots\ldots\ldots\ldots\ldots X_n \tag{3}$$

The action of choosing to drop the worst segments from both normal/abnormal video states leads to a state S_{t+1}. The termination state means that we have already found the best representative segment among normal/abnormal videos segments. The following formula describes the segment selection process within state S_t:

$$(X_i|S_{t+1}) = \begin{cases} 0 & \text{if} \quad X_i* = \text{argmin}_X Q(S_t, X_i) \\ 1 & \text{otherwise} \end{cases}$$

The Markov decision model uses some formula called the decision policy to select to best action at certain state. In accordance with the Q-Learning approach [20], [25], we fix Q_i as the expectation value of action a_i at S_t, where the action a_i represents the selection of a segment X_i. The policy is defined as π where:

$$\pi = \text{argmax}_{a_i} Q_i \tag{4}$$

Therefore, Q_i can be rewritten as:

$$Q(S_t, a_i) = R(s_t, a_i) + \gamma \max_a Q(S_{t+1}, \text{argmax}_a Q(S_{t+1}, a)) \tag{5}$$

$$R(S_t, a_i) = Q^*(S_t, a_i) \tag{6}$$

where γ is the discount factor, which takes a trade-off between the immediate reward and the prediction of feature reward. As shown in [20] we use a deep neural network $C(S_t, X)$ to evaluate $Q^*(S_t, X)$. If the evaluation is correct, we consider $Q^*(S_t, X)$ as $Q(S_t, X)$.

It exists two manners to build the architecture of the deep Q-network (DQN) C, one is taking the state S_t as the input and producing the Q-value of all possible actions, which is used in [20]. The other is taking both the state S_t and the action a_i as the input and producing single Q-value of the action.

By default, we consider a state S_t as a set of 32 segments per videos, our Q-network produces a Q-value for each segment in the state. But, In order to avoid a false positive anomaly segment, we proposed to add a new special state S_{ts}, which include only two segments having the higher Q-value in the normal and abnormal state.

$$S_{ts} = \{ \max_{X_i \in V_n} (X_i), \max_{X_i \in V_a} (X_i) \} \tag{7}$$

To avoid to correlation issue between consecutive states, we decide to use the experience replay (ER) technique [26], which it consists of storing agents experience $(S_t; X_t; R_t; S_{t+1})$ at each time step into a dataset D, whereS_{t+1} is the state at the next time step $t + 1$. For updating Q-values, it uses stochastic minibatch updates with uniformly random sampling from experience replay memory (previous transitions). More recently, Schaul [27] introduced Prioritized Experience Replay (PER) based on the concept that some experiences may be more important than others for our training, but might occur less frequently. But unlike the (ER), (PER) try to change the sampling distribution by using a criterion to define the priority of each experience.

In order to avoid as much as possible the false alert, we decide to take into priority experience with states containing only the two best anomaly segments among normal/abnormal video. It is also a way to employ the video-level labels instead of a totally rely on expert $C1$ predictions

for a better final policy quality, the network is trained with a target Q-network θ to obtain consistent Q-learning targets by fixing weight parameters used in Q-learning target and updating them periodically.

III. Experiments

A. Implementation details

To extract visual features from the video set, we use fully connected (FC) layer FC6 of the C3D network [12] for every

16-frame video clip followed by L2 normalization. Each state includes normal and abnormal video features (4096D), we input these states to start the training phase of our deep neural network. We use the 3-Layer FC neural network. The first FC layer has 512 units followed by 32 units and 1 unit FC layers. We fix ReLU [28] activation and Sigmoid activation for the first and the last FC layers respectively and take the Adam optimizer with the initial learning rate of 0.001, we compute loss by using a keras function called mse. We employ the generated model in[3] as an action recognition expert to evaluate our network and calculate the reward for each action. Finally, for the Prioritized Experience Replay (PER) we set the mini-batch size to 500 elements with a discount factor γ of 0.95.

B. UCF-Anomaly-Detection-Dataset

We evaluate our approach on a new large-scale dataset [3]. It's composed of long untrimmed surveillance videos which include 13 real-world anomalies, covering Abuse, Arrest, Arson, Assault, Accident, Burglary, Explosion, Fighting, Robbery, Shooting, Stealing, Shoplifting, and Vandalism. The dataset was divided into to parts: 800 normal videos and 810 anomalous videos for the training phase.

TABLE I
NUMBER OF VIDEOS OF EACH ANOMALY CLASS IN DATASET.

Classes	Number of Videos
Abuse	50
Arrest	50
Arson	50
Assault	50
Burglary	100
Explosion	50
Fighting	50
RoadAccidents	150
Robbery	150
Shooting	50
Shoplifting	50
Stealing	100
Vandalism	50
Normal events	950

C. Results

Based on [3], the authors demonstrate the invalidity of traditional action detection methods in real-world surveillance videos. Moreover, the experimental results of different techniques on UCF-Anomaly-Detection-Dataset are available in [3] as the dictionary method to learn anomalous pattern [24] and the proposed method of Hasan *et al* in [13]. Figure[2, 3, 4, 5] illustrate the obtained results by our method on many videos representing different anomalies. Our approach renders a precise and well-timed detection of anomalies by assigning high weights for the actions of choosing the anomaly segments, at the same time, it disqualifies the normal segments through a low anomaly weight.

Our approach gave better outputs than our expert network [3], as we predicted, the deep reinforcement architecture is more efficient than a classical deep learning network. We assume that our used architecture is more suitable for raw input data like videos. In the table below we compare the Deep Action Recognition Network to other methods:

D. False positive analysis

We notice that our system generates no alert for videos with bad quality(not clear, filmed at nighttime) and also a false alarm in the normal crowded scene. We concluded that it needs more powerful features extraction method to capture exclusively only the relevant features. In the other side, we will enhance the Experience Replay (ER) function by giving priority to specific segments, for example, the clips filmed at night.

TABLE II
COMPARISON (AUC) OF OUR METHOD TO OTHER APPROACHES ON UCF-ANOMALY-DETECTION-DATASET .

Approach	Accuracy
Binary classifier	50.0
Hasan *et al.* [13].	50.6
Lu *et a.* [24]	65.51
The proposed Deep Action Recognition Network	78.20

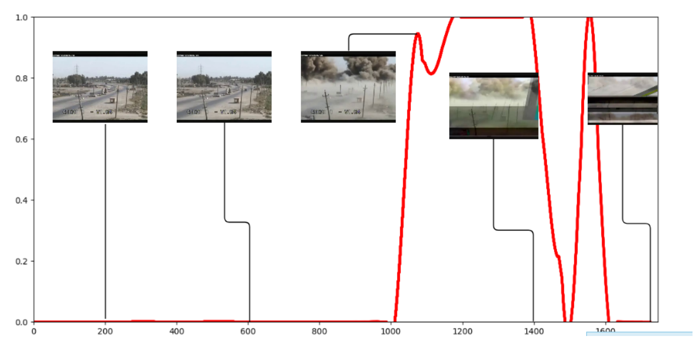

Fig. 2. Detection of anomaly (explosion) in a video.

Fig. 3. Detection of anomaly (shooting) in a video.

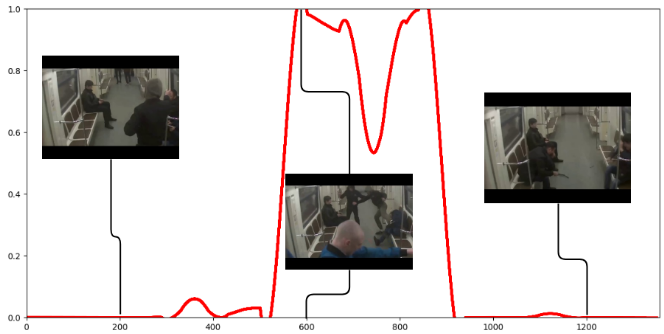

Fig. 4. Detection of anomaly (fighting) in a video.

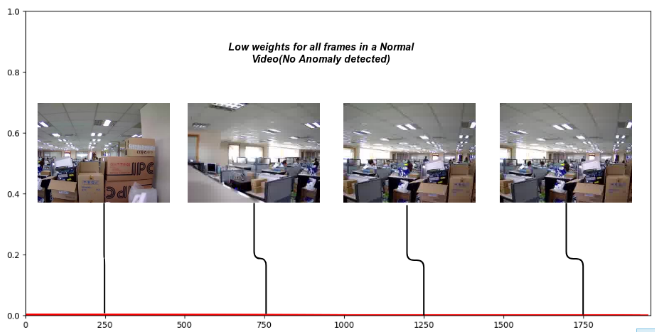

Fig. 5. Detection of anomaly in a normal video.

E. Conclusion

In this paper, we developed deep reinforcement learning architecture for anomalies detection in the surveillance video. Our proposed model tends to recognize the most abnormal segment in a video. We also attempt to resolve the false alarm problem in normal videos. We choose to use a new large-scale anomaly dataset consisting of a variety of real-world anomalies to validate the proposed methods. The experimental results on this dataset achieve a very competitive performance of video action recognition comparable to even superior than other methods.

In our work we compute the weight of each segment on the videos, it could be computationally expensive. In the future, we plan to improve the developed architecture in such a way that we could minimize the computational cost.

REFERENCES

[1] Y. Gao, H. Liu, X. Sun, C. Wang, and Y. Liu. Violence detection using oriented violent flows. Image and Vision Computing, 2016.
[2] J. Kooij, M. Liem, J. Krijnders, T. Andringa, and D. Gavrila. Multi-modal human aggression detection. Computer Vision and Image Understanding, 2016.
[3] S. Waqas, C. Chen, S. Mubarak. Real-world Anomaly Detection in Surveillance Videos, 2018.
[4] Y. Zhu, I. M. Nayak, and A. K. Roy-Chowdhury. Contextaware activity recognition and anomaly detection in video. In IEEE Journal of Selected Topics in Signal Processing, 2013
[5] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In CVPR, 2009.
[6] I. Saleemi, K. Shafique, and M. Shah. Probabilistic modeling of scene dynamics for applications in visual surveillance.TPAMI, 31(8):14721485, 2009.
[7] L. Kratz and K. Nishino. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In CVPR, 2009.
[8] X. Cui, Q. Liu, M. Gao, and D. N. Metaxas. Abnormal detection using interaction energy potentials. In CVPR, 2011.
[9] S. Wu, B. E. Moore, and M. Shah. Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In CVPR, 2010.
[10] A. Basharat, A. Gritai, and M. Shah. Learning object motion patterns for anomaly detection and improved object detection. In CVPR, 2008.
[11] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In CVPR, 2014.
[12] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3d convolutional networks. In ICCV, 2015.
[13] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis. Learning temporal regularity in video sequences. In CVPR, June 2016.
[14] D. Xu, E. Ricci, Y. Yan, J. Song, and N. Sebe. Learning deep representations of appearance and motion for anomalous event detection. In BMVC, 2015.
[15] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, Learning spatiotemporal features with 3d convolutional networks, in ICCV, 2015.
[16] S. Ji, W. Xu, M. Yang, and K. Yu, 3d convolutional neural networks for human action recognition, IEEE Trans. Pattern Analysis and Machine Intelligence, 2013.
[17] G. W. Taylor, R. Fergus, Y. LeCun, and C. Bregler, Convolutional learning of spatio-temporal features, in ECCV, 2010.
[18] Du Tran, H. Wang, L. Torresani, J. Ray, Y. leCun ,M. Paluri. A Closer Look at Spatiotemporal Convolutions for Action Recognition. arXiv:1711.11248v3 [cs.CV] 12 april 2018.
[19] M. L. Littman. Reinforcement learning improves behaviour from evaluative feedback. Nature, 521(7553):445451, 2015.
[20] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski. Human-level control through deep reinforcement learning. Nature, 518(7540):529533, 2015.
[21] F. Zhang, J. Leitner, M. Milford, B. Upcroft, and P. Corke. Towards vision-based deep reinforcement learning for robotic motion control. arXiv preprint arXiv:1511.03791, 2015.
[22] Y. Rao, J. Lu, J. Zhou. Attention-aware Deep Reinforcement Learning for Video Face Recognition. In IEEE International Conference on Computer Vision (ICCV)2017.
[23] K. Zhou, Y. Qiao, T. Xiang. Deep Reinforcement Learning for Unsupervised Video Summarization with Diversity-Representativeness Reward. arXiv:1801.00054v3 [csCV] 13 Feb 2018.
[24] 18 C. Lu, J. Shi, and J. Jia. "Abnormal event detection at 150 fps in matlab". In ICCV, 2013.
[25] C. J. C. H. Watkins and P. Dayan. "Q-learning. ML", 8(3):279292, 1992.
[26] Lin, L.-J., 1993. "Reinforcement learning for robots using neural networks". Ph.D. thesis, Fujitsu Laboratories Ltd.
[27] T. Schaul, J. Quan, I. Antonoglou and D. Silver," Prioritized Experience Replay".Google Deep Mind.conference paper at ICLR 2016.
[28] G. E. Hinton. Rectified linear units improve restricted boltzmann machines vinod nair. In ICML, 2010.