

Anomalous Activity Detection in Videos Using Incremental Learning: A Survey

Prerna Didwania and Vandana Jagtap

Abstract—Nowadays, there is a rapid growth in the number of video cameras at public and private sector because of the monitoring and security purposes. As video surveillance using Closed Circuit Television (CCTV) is in boom nowadays, it has got more research attention due to increased global security concerns. This rapidly growing data can be used to automatically detect the anomalous activities which are going around in our surrounding. Anomalous activity is something that deviates from its normal nature or something that opposes the normal events. This research mainly focuses on detecting anomalous activities in crowded scenes by using video data. Automatically detecting the anomalous activity without using the handcrafted feature has become the need of the hour. This paper contains a survey of different approaches used for anomaly detection in the past. Different incremental and transfer learning approaches are discussed in this paper and it was found that incremental learning has not been used for video-based anomalous activity detection.

Index Terms—Incremental Learning, Transfer Learning, Convolutional Neural Network, Machine Learning, Computer Vision.

I. INTRODUCTION

Recently, a rapid growth in the number of data generated by the videos placed at indoor and outdoor places for the monitoring and security purpose has been observed. According to a report IHS Markit, around 566PBs of data was generated by video surveillance cameras in 2015 and the report stated that by the end of 2019, it will generate 2500PBs of data on regular basis.

In the past few years, anomaly detection has become an area of interest for the research due to which various approaches has been proposed for detecting anomalies in crowded as well as non-crowded scenes. Things, actions or events that deviates or changes from their standard, normal or expected nature are considered as an anomaly.

Anomalous activity detection is the process whose goal is to find whether the given video frames is having an anomaly or not. There is another term called anomaly localization which tries to find the exact location of anomaly occurring in the video frame.

Depending on the number of entities present in the anomalous activity, anomaly can be of two different types, single entity based anomaly or interaction based anomaly. Single based anomaly considers an event as an anomaly if this behavior is different as compared to its neighbors. One of the examples of single based anomaly can be a person driving a vehicle in a wrong direction of the road. An

anomaly is interaction based anomaly if multiple events which are normal when performed individually, interact with each other in a different manner. Some of the examples of interaction based anomaly can be car accidents; group of people standing together during a riot.

Performance of anomaly detection can be evaluated using different levels of detection such as frame, pixel and dual pixel level detection. In pixel-level anomaly detection, if even a one pixel in the frame is detected as an anomaly then entire frame is considered as an anomaly. In frame-level anomaly, if 40 percent of pixels in a video frame are detected as an anomaly then the frame is considered as an anomaly. In dual pixel level, there are two conditions that needs to be satisfied. First condition is that it should satisfy frame-level anomaly detection and second condition is that, if β percent of pixels are detected as an anomaly then the frame is considered as an anomaly. This β parameter is defined by the user.

As humans have ability to incrementally learn new pieces of information, similarly incremental learning is a machine learning approach that learns new information as they arrive without forgetting the previous information learned. Incremental learning has become area of interest for the researchers as they require very small computational power and memory as compared to other existing approaches.

This research belongs to the domain of intelligent video surveillance where the aim is to detect anomalous activity from video of crowded scenes. This research aims to use incremental learning and transfer learning concepts along with convolutional neural network (CNNs) to detect anomalous activities in videos in crowded and non-crowded scenes.

This survey is divided into three sections. Section II has some background knowledge required to understand incremental and transfer learning. Section III contains a literature review of the approaches used for anomaly detection and incremental and transfer learning. The paper concludes in Section IV.

II. BACKGROUND

A. Incremental Learning

Incremental learning is a machine learning approach in which models learns from the new examples as they arrive without forgetting the knowledge previously learned. In incremental learning, learning process happens when input data gradually becomes available. The main aim of incremental learning is to preserve previously learned data while learning from new examples. Learning process is incremental learning if it satisfies following criteria:

- New information should be learned when new

Published on March 17, 2020.

Authors are with the School of Computer Engineering & Technology, MIT World Peace University, Pune, India.

examples are gradually coming or as new data arrives.

- When new data arrives, it should not need original data that was used to train the model.
- It should not suffer from catastrophic forgetting i.e. forgetting previously learned data when new information is learned.
- New classes learned from the new examples should be accommodated.

B. Transfer Learning

In transfer learning the knowledge from one or more sources is used or transferred to improve learning in related tasks. Transfer learning is a two-step process:

- first we train our base network from base dataset, then
- we use the learned features to train our target network from our target dataset.

This process works until all the features are general, which means features are not just specific to base tasks but also suitable to both target and base tasks.

The approach of using transfer learning will be different according to the size and similarity of base dataset and network to the target dataset and network. There are 4 approaches in general:

1. The target dataset is similar and small as compared to the base dataset.

As we have small target dataset, fine-tuning the network will lead to over fitting. Following steps are carried out in this process:

- Remove fully connected layers which are at the near end of the previously trained base network.
 - According to the number of classes in the target dataset, add new fully connected layers.
 - Now randomly assign the weights of new fully connected layers.
 - Train the network to update weights.
2. The target dataset is similar and large as compared to the base dataset.

Following steps are carried out in this process:

- According to the number of classes in the target dataset, add new fully connected layers.
 - Now randomly assign the weights of new fully connected layers.
 - Initialize the remaining of weights from pre-trained network
 - Retrain the entire network.
3. The target dataset is different and small as compared to the base dataset.

Over fitting will be a concern because the target dataset is small and as the target dataset is different the higher layers of the base network are not that useful. Therefore, we only use the lower layers of the base dataset.

4. The target dataset is different and large as compare to the base dataset.

As the target dataset is different and large as compared to base dataset, therefore we need to build the entire target network from the scratch.

III. LITERATURE REVIEW

The area of anomaly detection is in boom from past few years. This section has been classified into 2 categories as traditional approaches and deep learning based approaches.

A. Traditional Approaches

Traditional approaches include learning of normalcy model from the videos to detect the anomaly. Various distance or likelihood measures are used to detect the anomaly. If any event deviates from the normalcy model, it is considered as an anomaly.

In 2014, Wu et al. proposed a holistic approach. In this approach crowd behavior detection using Bayesian framework is proposed to directly model non-escape and escape crowd motions [2]. These crowd motions are characterized using optical flow field. These approaches can perform global anomaly detection. The major limitation of these approaches is their less accuracy.

In 2016, Blair et al. proposed an object-based approach which treats the scene to be detected as a collection of individuals. It uses object detection and approach based on Histogram of Gradients (HoG) and Mixture of Gaussian (MoG)[3]. These approaches can perform local anomaly detection but they can't handle densely crowded scenes.

In 2014, Li et al. proposed a dynamic texture model that uses mixture of dynamic texture. In this normal pattern is learned through mixture of dynamic texture. Patches of low probability under associated mixture of dynamic texture are anomalies [4]. They can perform local anomaly detection. This approach does not consider the relationship among local observations and works only for batch processing of video.

In 2013, Roshtkhari et al. proposed a method that extracts low-level visual features by implementing a pixel-level background model and employing spatio-temporal video volume. Observations are considered as anomalous if it can't be reconstructed with previous observation [5]. It can perform both local and global anomaly detection. This approach needs advanced threshold method for inference of suspicious activity.

In 2015 Cheng et al. proposed a method which is uses Gaussian process regression to learn spatio-temporal interest points (STIPS). In this method bottom-up greedy approach is used [6].

B. Deep learning based approaches

Various deep learning approaches are used for anomaly detection in video data. Some of them are described below.

In 2016, S. Zhou et al., proposed the networks that could capture information and motion of spatio temporal dimensions that were encoded in an adjacent frame. They model spatio-temporal relation of the patches using spatio-temporal CNN and infer the combined notion of motion and appearance of entities in the video. Given a sequence of images as an input, the recognition module generates a belief state about the objects that produced those images. Using this, the belief state generated, the control module produces actions that will affect the images observed in the future. It can detect both local and global anomalies [7].

In 2016, Hasan et al., proposed a deep generative model that uses reconstruction error for calculating anomaly score.

The general assumption done in this model is that the features are likely to fail if change in feature distribution is observed because all the features belongs to predefined type of distribution. This model can detect global anomalies but has poor anomaly localization [8].

In 2016, Jefferson et al. proposed a multiple stream network in which two parallel networks in which first network was spatial stream network and other network was temporal stream network. In this approach spatial stream network accepted raw video frames whereas temporal stream network used optical flow fields as an input. First this model uses global and context-aware features and then merges results with local and action-aware ones. This model can detect global anomalies but has poor anomaly localization [9].

In 2016, Mohammad Sabokrou et al. proposed a cascaded deep neural network in which classifier has two stages combined together, a very small stack of auto-encoders, and a CNN. In this approach first in the shallow layers of the network simple normal patches are detected using a weak classifier and then more complex patches are learned in the deeper layer using strong classifier [1].

C. Incremental and transfer learning approaches

This section consists of survey of different approaches of transfer and incremental learning.

RobiPolikar et al., 2001[10], introduced an incremental learning algorithm called Learn++ that can accommodate newly arrived data which includes the examples that are previously unseen. This algorithm doesn't require original dataset to train the model when new examples arrives. Multiple hypotheses are generated by using training datasets to utilize ensemble learning. Other incremental learning algorithm such as RBF networks, Fuzzy ARTMAP, TopoART, and IGNG can also be used and their accuracy and performance can be assessed.

Amir Rosenfeld et al., 2018[11], proposed a method that could learn new example by using linear combination of previously learned examples. This method is called deep adaption networks. Switching between various learned examples can be controlled by the new architecture that is learned which will enable a single network to process tasks of multiple domains. In this paper if the tasks are sufficiently related to each other then only the model works efficiently. The model will fail if tasks are not similar. For example, if a task requires horizontal counting and other requires vertical counting, in this case the model will not work efficiently.

Yang Yang et al., 2019 [12], developed an approach which uses incremental deep model which is adaptive in nature. This model has two main goals, one is to make model more flexible and faster by scaling the streaming data. Second, to provide capacity sustainability to the system because stream data changes continuously.

Ling Shao et al., 2015 [13], performed a survey on visual categorization which uses transfer learning approach. Visual categorization application generally consists of image classification, object recognition, activity recognition etc. In this paper two aspects are not taken into consideration. First, how to mine the information from the source domain which is noisy which would be helpful to target domain data and

second, how to incorporate the existing transfer learning methods which can deal with large data.

Yu Sun et al., 2015 [14], this paper deals with the scenario where emergence and disappearance of classes happens gradually. This paper proposes a class-based ensemble approach in which for each class a base learner is maintained and the base learners are dynamically updated with new data. The approach used in this paper emphasis more on evolved classes, so it may happen that its performance gets degraded when performed on non-evolved classes.

M Arfan Ikram et al., 2015 [15], this paper proposed the image segmentation approach using transfer learning. This paper comparison of supervised learning classifier and four transfer learning classifier is performed. In this paper, they have only used voxelwise classification as a comparison parameter. So voxelwise classification can be replaced in any of the four transfer learning classifier.

IV. CONCLUSION

As there is a rapid growth in data generated by the videos, detecting of anomaly in videos is a need of the hour. In crowded scenes abnormal event detection is a major issue. Many approaches have been used to detect anomaly in the video in the past. But many of the approaches suffers from catastrophic forgetting. This issue can be solved using incremental learning. Incremental learning allows a model to learn new example adaptively i.e. it allows to learn new examples as they arrive without losing the previously learned information, thus resolves the issue. Video-based data always suffers from the imbalanced training data for detecting anomaly as the video data available is huge and the videos actually containing the anomaly is less. A system that deal with the imbalance between training and testing dataset is required that can be resolved by using transfer learning.

REFERENCES

- [1] Mohammad Sabokrou, Mohsen Fayyaz, MahmoodFathy, and ReinhardKlette," Deep-cascade: Cascading 3D Deep Neural Networks for Fast Anomaly Detection and Localization in Crowded Scenes"IEEE Transaction on Image Processing,2017
- [2] S. Wu, H. S. Wong, and Z. Yu,"A Bayesian model for crowd escape behavior detection", IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 1, pp. 8598, 2014A. Agrawal and U. Gupta,"Clustering based extractive text summarization system", 2014
- [3] C. G. Blair and N. M. Robertson,"Video Anomaly Detection in Real Time on a Power-Aware Heterogeneous Platform", IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 11, pp. 21092122, Nov. 2016.
- [4] Li, W., Mahadevan, V. and Vasconcelos, "Anomaly detection and localization in crowded scenes", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no.1, pp.18-32, 2014
- [5] M. JavanRoshkhari and M. D. Levine,"An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions",Elsevier, Computer Vision and Image Understanding 117 (2013)14361452
- [6] K. W. Cheng, Y. T. Chen, and W. H. Fang,"Gaussian Process Regression- Based Video Anomaly Detection and Localization with Hierarchical Feature Representation",IEEE Transactions on Image Processing, vol. 24, no. 12, pp.52885301, 2015.
- [7] S. Zhou, W. Shen, D. Zeng, M. Fang, Y. Wei, Z. Zhang,"Spatialtemporal convolutional neural networks for anomaly detection and localization in crowded scenes",Elsevier,Signal Processing: Image Communication, Vol. 47, pp. 358-368, 2016.

- [8] Mahmudul Hasan, Jonghyun Choi, Jan Neumann, Amit K. Roy-Chowdhury and Larry S. Davis, "Learning Temporal Regularity in Video Sequences", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. DOI:10.1109/cvpr.2016.86
- [9] Jefferson Ryan Medel and Andreas Savakis, "Anomaly Detection in Video Using Predictive Convolutional Long Short-Term Memory Networks", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016
- [10] Robi Polikar, Lalita Udpa and Vasant Honawar, "Learn++: An Incremental Learning Algorithm for Supervised Neural Networks", IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews), March 2002
- [11] Amir Rosenfeld and John K. Tsotsos, "Incremental Learning Through Deep Adaptation", IEEE Transactions on Pattern Analysis and Machine Intelligence (Early Access), 2018. DOI: 10.1109/TPAMI.2018.2884462
- [12] Yang Yang, Da-Wei Zhou, De-Chuan Zhan, Hui Xiong "Adaptive Deep Models for Incremental Learning: Considering Capacity Scalability and Sustainability", ACM ISBN 978-1-4503-9999-9/18/06. DOI: 10.1145/1122445.1122456
- [13] Ling Shao, Fan Zhu, Xuelong Li, "Transfer Learning for Visual Categorization: A Survey", IEEE Transactions on Neural Networks and Learning Systems (Volume: 26, Issue: 5, May 2015) DOI: 10.1109/TNNLS.2014.2330900
- [14] Yu Sun; Ke Tang; Leandro L. Minku; Shuo Wang; Xin Yao, "Online Ensemble Learning of Data Streams with Gradually Evolved Classes", IEEE Transactions on Knowledge and Data Engineering (Volume: 28, Issue: 6, June 1 2016) DOI: 10.1109/TKDE.2016.252667
- [15] M Arfan Ikram, Meike W Vernooij, Marleen de Bruijne, "Transfer Learning Improves Supervised Image Segmentation Across Imaging Protocols", IEEE Transactions on Medical Imaging, 2014. DOI: 10.1109/TMI.2014.2366792.



Prerna Didwania is currently pursuing M. Tech in Data science and Analytics from School of CET, MIT WPU, Pune. She has completed her Bachelor of Engineering from Pune University, Maharashtra.



Prof. Vandana Jagtap presently working as Assistant professor in School of Computer Engineering & Technology, MIT World Peace University, Pune, India. She is Perceiving PhD from Pune University, Maharashtra. She did Master of engineering from Pune University, Maharashtra and Bachelor of Engineering from Amravati university Maharashtra. She has published about 20 papers in national and international Journals. She has presented about 20 papers in national and international conferences. Her current area of research includes Artificial Intelligence, Machine learning, and cloud computing and high performance computing. She is member of ISTE and ACM.