Deep Abnormality Detection in Video Data

Hung Vu
Centre for Pattern Recognition and Data Analytics (PRaDA), Deakin University, Australia
hungv@deakin.edu.au

Abstract

Automated detection of anomalous events plays an important role in video surveillance systems in practice. This task, however, requires to deal with three challenging problems of the lack of annotated training data, the inexact description of what to be "abnormal" and the expensive feature engineering procedure. Most anomaly detection systems are only able to satisfy some of these challenges. In this work, we propose a deep abnormality detection system to handle all of them simultaneously. Deep abnormality detection is a deep generative network that is an unsupervised probabilistic framework to model the normality and learn feature representation automatically. Furthermore, unlike other existing methods, our system can detect abnormality at multiple levels and be used as a powerful tool for video analysis and scene understanding.

1 Introduction

Nowadays, due to the rise of terrorism and crimes, there are more and more increasing concerns for intelligent systems to automatically discover unexpected behaviors or anomaly events in videos. Anomalous events are commonly assumed to be rare or significantly different from the others [Sodemann *et al.*, 2012]. Video anomaly detection is a non-trivial task because it faces both issues of anomaly detection in general and video processing. In particular, this dissertation specifies three key challenges. Firstly, anomaly objects cannot be explicitly defined, and then systems have to deal with the uncertainty. The second is the shortage of labeled training data, since video annotation (usually at the pixel-level) is expensive and labor-intensive, and hence annotated datasets are not large enough to learn effective anomaly detectors. Thirdly, video processing relies on hand-crafted features (e.g. histogram of orientation gradients or optical flow), which require exhaustive prior knowledge and expensive computation. As a result, feature engineering is a time-consuming phase in designing and developing anomaly detection systems.

This work was partially supported by the Australian Research Council under the Discovery Project DP150100031 and the DST Group.

Most existing methods are only able to deal with the first two issues of uncertainty and insufficient labeled data (via unsupervised probabilistic frameworks). Recently there have been several anomaly detection studies [Xu *et al.*, 2015; Hasan *et al.*, 2016] that have adopted deep learning techniques to automatically learn high-level representations, and then avoid the requirement of domain experts in designing features. However, they are non-probabilistic methods, and hence may fail to model the uncertainty.

This work introduces a solution to address three aforementioned issues simultaneously via deep generative networks. By capturing the distribution of regular events in unlabeled data, learned models can isolate abnormal events with low probabilities. Meanwhile, as deep learning networks, our models can automatically learn the data transformation at different levels and produce hierarchical feature representation. Moreover, unlike other systems that detect anomaly objects once after feature representation, by extending from deep generative networks, we propose a novel deep abnormality detection system which is able to produce anomaly detection results in every feature representation layer of the network.

2 Proposed Method

We discover three generative models, each of which is related to a project of the dissertation. In the first project, we start off with a shallow generative net which is known to be effective and efficient to do learning and inference. Next, we upgrade it to a deep generative model with a powerful capacity of higher level representations. The final project focuses on designing a deep network that specializes in anomaly detection.

2.1 Project 1: Detection Using Shallow Networks

Restricted Boltzmann machine (RBM) [Freund and Haussler, 1994] is an undirected generative network of a visible layer and a hidden layer. Due to its bipartite structure, the RBM can learn the data distribution efficiently in comparison to deep generative networks. RBMs are also extended for complicated data structures such as multiway tensor data [Nguyen *et al.*, 2015]. Furthermore, their pretraining roles contribute to the success in training deep networks and mark the rise of deep learning [Salakhutdinov and Hinton, 2009]. For this reason, we use RBMs in the first project to examine the possibility of applying generative networks to video anomaly detection.

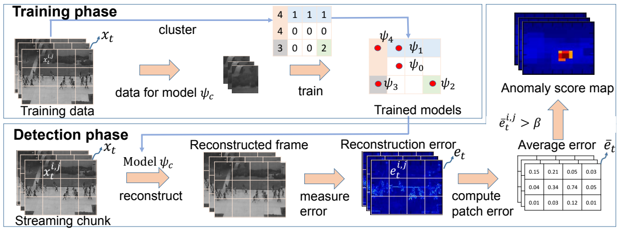

Figure 1: The framework overview.

To localize anomaly events in videos, we design a framework as demonstrated in Figure 1. In general, our system consists of two phases of training and detection.

Training phase: The purpose of this phase is to learn RBM models from a collection of unlabeled videos. We firstly split frames into 50% overlapping $h \times w$ patches, and then group similar patches into C clusters. All patches that are associated with the c^{th} cluster are used as training data to learn the c^{th} model.

Detection phase: We divide an input video of L frames into patches $\boldsymbol{x}_t^{i,j}$, where t is the frame index and (i, j) is the patch location. Each patch is fed into the corresponding learned model to yield the reconstructed patch $\tilde{\boldsymbol{x}}_t^{i,j}$. The patch error $\bar{e}_t^{i,j} = ||\tilde{\boldsymbol{x}}_t^{i,j} - \boldsymbol{x}_t^{i,j}||_2 / (h \times w)$ is then compared with a predefined threshold β. By discarding small objects spanning less than γ frames, we obtain the abnormalities in the video. One advantage of our proposed framework is that it is immediately amendable to a streaming and distributed implementation.

2.2 Project 2: Detection Using Deep Generative Nets

Deep generative nets (e.g. DBM) are multilayer networks with more than two layers. Multi-layer architecture enables the ability of better distribution modeling and hierarchical feature representation. Nevertheless, such architecture causes learning and inference to be more tricky. Hence, efficient learning and inference algorithms in deep networks should be intensively studied in this project. One direction should be investigated is the use of convolutional networks to improve the performance of anomalay detection in video data.

2.3 Project 3: Deep Abnormality Detection System

While the first two projects perform feature extraction and anomaly detection separately, this project aims to design a novel deep network that incorporates abnormality detection in every layer of the network. As a special generative network proposed for anomaly detection problem, our network is expected to produce better accuracy and hierarchical detection results. The coarse-to-fine detection ability allows our system to analyze and understand the video scenes accurately (e.g., in the scenario of pedestrian footpath, the bicycle and the rider can be considered as abnormality at a coarse level but only bicycle is abnormal at a fine level).

3 Current Results

At the current stage, we have developed a system using shallow generative nets. This section reports the experimental results of the first phase where we use RBMs to detect anomaly events in videos. Table 1 shows the detection results in AUC (Area Under Curve) and EER (Equal Error Rate) of our RBM and its streaming version (S-RBM) and some existing methods. Overall, our methods outperform PCA, OC-SVM and GMM while S-RBM obtains fairly comparable results with ConvAE [Hasan *et al.*, 2016]. This result is promising because ConvAE is a state-of-the-art deep networks with 12 layers, compared to only 2 layers in our RBMs.

	Ped1		Ped2		Avenue	
	AUC	EER	AUC	EER	AUC	EER
PCA	60.28	43.18	73.98	29.20	74.64	30.04
OC-SVM	59.06	42.97	61.01	44.43	71.66	33.87
GMM	60.33	38.88	75.20	30.95	67.27	35.84
ConvAE	**81.00**	**27.90**	**90.00**	21.70	70.20	**25.10**
RBM	64.83	37.94	76.70	28.56	74.88	32.49
S-RBM	70.25	35.40	86.43	**16.47**	**78.76**	27.21

Table 1: Anomaly detection results at frame-level. Higher AUC and lower EER indicate better performance. Best scores are in bold.

4 Contributions

Our work contributes a powerful and generalized tool with capacity of unsupervised learning and hierarchical representation in localizing irregular objects in videos. This research also provides a comprehensive understanding of the use of deep generative networks for anomaly detection. In addition, we introduce a deep abnormality detection model that is a deep generative network designed to localize abnormality effectively and hierarchically. Finally, many practical applications such as video analysis, traffic monitoring, fighting detection, will benefit from our proposed system.

Acknowledgments

I am deeply grateful to my supervisors, Professor Dinh Phung, Dr. Tu Dinh Nguyen, Dr. Anthony Trevors and Professor Svetha Venkatesh, for their instructions and advice.

References

[Freund and Haussler, 1994] Y. Freund and D. Haussler. Unsupervised learning of distributions on binary vectors using two layer networks. Technical report, CA, USA, 1994.

[Hasan *et al.*, 2016] M. Hasan, J. Choi, J. Neumann, AK. Roy-Chowdhury, and LS. Davis. Learning temporal regularity in video sequences. In *CVPR*, 2016.

[Nguyen *et al.*, 2015] TD. Nguyen, T. Tran, D. Phung, and S. Venkatesh. Tensor-variate restricted boltzmann machines. In *AAAI*, volume 4, pages 2887–2893, Austin Texas, USA, January 25–30 2015.

[Salakhutdinov and Hinton, 2009] R. Salakhutdinov and G. Hinton. Deep Boltzmann machines. In *AISTATS*, volume 5, pages 448–455, 2009.

[Sodemann *et al.*, 2012] AA. Sodemann, MP. Ross, and BJ. Borghetti. A review of anomaly detection in automated surveillance. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42:1257–1272, November 2012.

[Xu *et al.*, 2015] D. Xu, E. Ricci, Y. Yan, J. Song, and N. Sebe. Learning deep representations of appearance and motion for anomalous event detection. *BMVC*, 2015.