

Regularity in Video Sequences

Lina Zhu

June 14 ,2018

Abstract

Abstract Sensing meaningful activities in long video sequences is a challenging issue of meaning and confusion in the scene due to ambiguous definitions. We solve this problem by learning a generated model. The use of multiple supervisory rules is very limited. Specifically, we propose two approaches based on automatic encoders because they have little or no supervision. We first use traditional manual spatio-temporal local functions and learn fully connected automatic encoders for them. Second, we have built a complete convolutional learning native function into an end-to-end learning framework. Our model can capture rules from multiple data sets. We evaluate our qualitative and quantitative methods from various aspects of displaying video learning patterns and displaying anomalous competitive performance detection datasets as applications.

1. Introduction

The availability of a large number of uncontrolled videos leads to long-time pointless scenes [8]. Automated segmentation makes sense with a very limited supervision of these video moments without supervision. It is a variety of basic problems in computer vision applications such as video annotation[10], abstract, indexing or time-segmentation[4], Anomaly detection and activity identification[3]. Instead of modeling supervised sparse irregularities or meaningful moments in a supervised attitude, we model the time to solve this problem with limited oversight. Learning meaningful time visual features or significant moments is a very challenging definition of these moments is not clear, that is, there is no visual limit. On the other hand, it is easier to learn time visual features, because they often show the regularity of time, such as periodic crowd movements. We focus on learning the characteristics of regular time patterns through very limited labeling. We assume that all events in the training video are part of the regular pattern. In particular, we use multiple video sources, such as different data sets, to learn the regular temporal appearance of changing the video mode in a single mode and then use

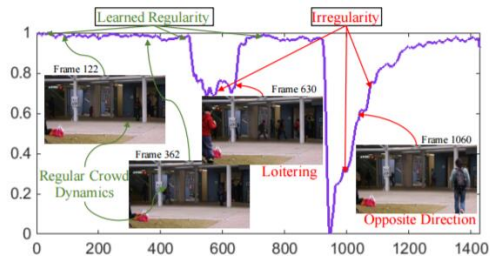


Figure 1. Learned regularity of a video sequence. Y-axis refers to regularity score and X-axis refers to frame number. When there are irregular motions, the regularity score drops significantly.

it for multiple videos. Given the training data of regular videos, the time dynamics of learning regular scenes is an unsupervised learning problem. This unsupervised state-of-the-art method modeling involves sparsely coded combination bags[6][12][1]. In addition, optimizing sparse coding training and testing involving both are computationally expensive, especially with large amounts of data, such as video. We propose a method based on an automatic encoder. Its objective function is more efficient than computational sparse coding and preserves spatiotemporal information while encoding dynamics.

The learned automatic encoder reconstructs regular motion with low error but results in higher error in recreating irregular motion. Figure 1 shows an example of a learning rule, which is a learning model calculated from reconstruction errors. We suggest that automatic encoders that learn the regularity of time are based on the following two types of functions. First, we use state-of-the-art hand-crafted motion functions and learn neural network-based deep automatic encoders to form fully connected layers. The most advanced sports functions, however, may not be optimal for learning time rules because they are not designed or optimized for this problem. Subsequently, we directly understand these two kinds of motion features and use a fully convoluted discriminative rule pattern based on a neural network-based automatic encoder.

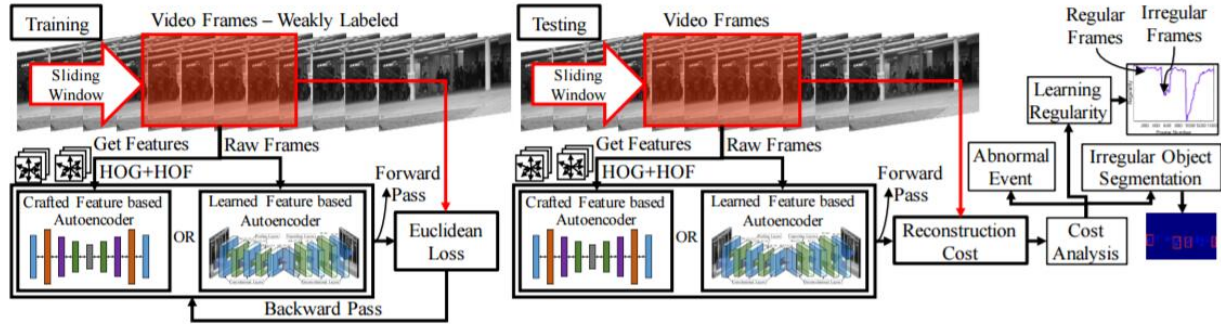


Figure 2. Overview of our approach. It utilizes either state-of-the-art motion features or learned features combined with autoencoder to reconstruct the scene. The reconstruction error is used to measure the regularity score that can be further analyzed for different applications.

2. Related Work

Learn unsupervised exercise patterns. The pattern of learning without supervision has received a lot of attention in recent years[5][2][9]. Learn previous signatures for the same time and use it for event retrieval, anomaly detection. One of the applications of our model is abnormal or abnormal event detection. The survey report contains comprehensive comments on this topic. The most video-based anomaly detection method involves a local feature extraction step and then learns the model in the training video. Any event that is anomalous to the learning model is considered abnormal. These models include the components of the mixed optical flow of probabilistic subjects, sparse dictionaries[6][12], probability frameworks based on Gaussian regression, spatiotemporal contexts, sparse automatic encoders, Based on the space-time volume analysis of the codebook[7], compared with the shape, our model is an end-to-end training generator that can be generalized over multiple data sets. Convolutional neural networks (CNN) perform regular systems in end-to-end learning video.

3. Approach

We use automatic encoders to learn regularity in video sequences. Intuition is an automatic encoder that learns to reconstruct a motion signature in a regular video with low error, but does not accurately reconstruct irregular motion video. In other words, the automatic encoder can model the dynamic complex distribution changes of the appearance rules. As an input to an automatic encoder, we initially used state-of-the-art techniques to manually create motion features with improved trajectory features by HOG and HOF[11]. Then we learn a network-based auto-encoder that automates signatures by neural rules. However, even out-of-the-box national movement characteristics may not be optimal for learning patterns because they are not designed specifically for this purpose. Therefore, we use video as input and learn two local motion functions based on a complete convolutional neural network through an end-to-end

learning model. We illustrate our overview of the method in Figure 2.

4. Conclusion

We propose a limited supervision of the use of automatic encoders to learn regular patterns. We first use traditional spatio-temporal local features and learn fully connected automatic encoders. Then we build a complete convolutional autoencoder to learn local features and classifiers in a single learning framework. Our model can be generalized across multiple data sets, even using potential dataset biases. We analyze our learning model in many ways, such as visualizing regular frames and pixels, and predicting the past and the future of conventional video will have only one image in the future. For quantitative analysis, we show that our method is competitive with the most advanced anomaly detection methods.

References

- [1] Y. Cong, J. Yuan, and J. Liu. Sparse reconstruction cost for abnormal event detection. In *CVPR*, 2011. 1
- [2] H. Jhuang, T. Serre, L. Wolf, and T. Poggio. A biologically inspired system for action recognition. In *ICCV*, 2007. 2
- [3] 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. 1
- [4] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. In *CVPR*, 2008. 1
- [5] Q. V. Le, W. Y. Zou, S. Y. Yeung, and A. Y. Ng. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In *CVPR*, 2011. 2
- [6] C. Lu, J. Shi, and J. Jia. Abnormal event detection at 150 fps in matlab. In *ICCV*, 2013. 1, 2
- [7] M. J. Roshtkhari and M. D. Levine. Online dominant and anomalous behavior detection in videos. In *CVPR*, 2010. 2
- [8] M. Sun, A. Farhadi, and S. Seitz. Ranking domain-specific highlights by analyzing edited videos. In *ECCV*, 2014. 1

- [9] G. W. Taylor, R. Fergus, Y. LeCun, and C. Bregler. Convolutional learning of spatiotemporal features. In *ECCV*, 2010. [2](#)
- [10] C. Vondrick, D. Patterson, and D. Ramanan. Efficiently scaling up crowdsourced video annotation. *IJCV*, 101(1):184–204, 2013. [1](#)
- [11] H. Wang and C. Schmid. Action recognition with improved trajectories. In *ICCV*, 2013. [2](#)
- [12] B. Zhao, L. Fei-Fei, and E. P. Xing. Online detection of unusual events in videos via dynamic sparse coding. In *CVPR*, 2011. [1](#), [2](#)