

Dual-mode iterative denoiser: Tackling the weak label for anomaly detection

Shuheng Lin, Hua Yang[™]

Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University Shanghai key lab of digital media processing and transmission Shanghai, China

Introduction

Crowd anomaly detection suffers from limited training data under weak supervision. In this paper, we propose a dual-mode iterative denoiser to tackle the weak label challenge for anomaly detection. First, we use a convolution autoencoder (CAE) in image space to act as a cluster for grouping similar video clips, where the spatial-temporal similarity helps the cluster metric to represent the reconstruction error. Then we use the graph convolution neural network (GCN) to explore the temporal correlation and the feature similarity between video clips within different rough labels, where the classifier can be constantly updated in the label denoising process. Without specific image-level labels, our model can predict the clip-level anomaly probabilities for videos. Extensive experiment results on two public datasets show that our approach performs favorably against the state-of-the-art methods.

Proposed Method

• We propose a dual-mode iterative label denoising approach for the weak supervised anomaly detection task. With the only video-level label in the training process, our model can automatically generate clip-level detection results.

• We develop a pre-denoiser to guide the initial label denoising for the global optimization direction.

We propose the dual-mode label denoisers framework, where the cluster label denoiser works on the image space to group similar clips, and the GCN label denoiser works on the feature space to constrain the temporal correlations. As far as we know, it is the first work to leverage CAE to correct noised label in video analytics.

We conduct experiments on the largest anomaly detection datasets. The comparison results demonstrate the effectiveness of our model against state-of-the-art methods.



Experiments

Table 1: AUC comparison on UCF-Crime dataset.

Method	AUC(%)
Hasan et al. [10]	50.66
Lu et al. [3]	65.51
Sultani et al. [6]	74.89
Lin <i>et al.</i> [7]	78.28
Zhong et al. [22]	79.25
Our method	83.31
Table 2: AUC comparison on Sh	hanghaiTech dataset.
Method	AUC(%)
Sultani et al. [6]	73.51
Lin <i>et al.</i> [7]	74.83
Zhong et al. [22]	82.09
Our method	85.12



Conclusion

In this paper, to learn both effectively normal and abnormal patterns under weak supersion, i.e., as for the challenge of weak label in anomaly detection, we regard the video-level label as the noised label for its clips and propose a dual-mode iterative denoising model to train an end-to-end classifier. We use a pre-denoiser to guide the initial optimization direction, to avoid non convergence at later stage. To further clean the noised labels, we conduct the denoising both on the image space and feature space. We propose a cluster-based algorithm to gather the video clips which are similar in image space through two CAEs. After several iterations, this module learns both normal and abnormal patterns effectively. Under the guidance of the pre-denoiser, the cluster denoiser perform better at the initial training. On the feature level, we pay more attention on the temporal correction and the feature similarity between clips, GCN is leveraged to learn the special relations. With only video-level labels in training set, we finally conduct the anomaly detection on clip-level. Our method also achieves the state-of-the-art performance on two largest public datasets.