Supplemental Materials
Dynamic Local Aggregation Network with Adaptive Clusterer for Anomaly Detection

Zhiwei Yang[1], Peng Wu[2†], Jing Liu[1†], and Xiaotao Liu[1]

[1] Guangzhou Institute of Technology, Xidian University, Guangzhou, China
[2] School of Computer Science, Northwestern Polytechnical University, Xi'an, China
zwyang97@163.com, xdwupeng@gmail.com, neouma@163.com, xtliu@xidian.edu.cn

1 Inference Speed

Our method is implemented on a single Nvidia RTX-3090 GPU on a machine with CPU core of i7-10700K@3.80Ghz and 31.3 G memory. In the inference stage, our model takes about 0.009 seconds (111 FPS) to process a frame. This inference speed meets the real-time detection requirements in real applications.

2 Hyperparameter Selection

For hyperparameters such as the learning rate of AE, the initial neighborhood radius δ and the initial learning rate η of AC, we choose them based on empirical values. For the number of competing layer neurons L in AC and the loss function weights λ_{cp} and λ_{sp}, which are the two key hyperparameters, we use a grid search strategy to select them. We show the AUC of different settings of L, λ_{cp} and λ_{sp} on the Ped2 dataset as follows. We can observe that our method achieves optimal performance when L is set to 25 (M represents the number of prototypes) and λ_{cp} and λ_{sp} are set to 0.01. Even with some fluctuations in these two parameters, our method still maintains a relatively stable performance.

L / M	9/5	16/9	25/13	36/19	49/28
DLAN-AC	0.953	0.965	0.976	0.963	0.954

λ_{cp} or λ_{sp}	0.1	0.05	0.01	0.005	0.001
DLAN-AC	0.954	0.967	0.976	0.961	0.950

3 Additional Qualitative Results

To further demonstrate the effectiveness of our method, we show additional qualitative results here. Fig. 1 shows the prediction error map for some examples on Avenue and ShanghaiTech datasets, respectively. Obviously, for normal events, the future frame predicted by our model is almost close to the actual frame,

---

[†]Corresponding authors.

as shown by the darker error map. For abnormal events, the predicted future frame tends to be blurred and distorted compared to the real frame, and the location of the abnormality is very conspicuous in the error map. Fig. 2 shows more examples of anomaly score curves for the test video clips on three bench-mark datasets. It is easy to observe from Fig. 2 that the low abnormal score increases sharply with the occurrence of abnormal events, and then returns to the low level after the abnormality ends.
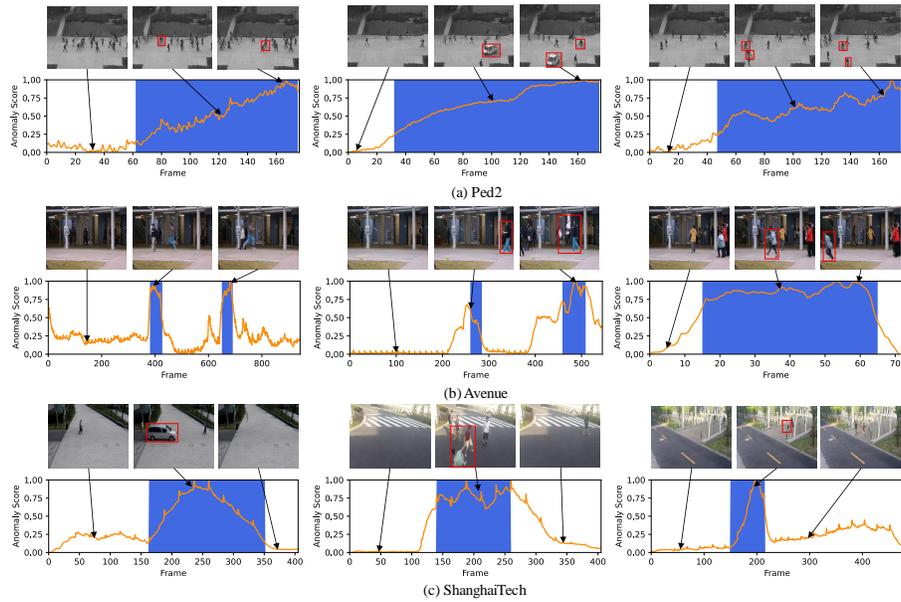

(a) Avenue

(b) ShanghaiTech

Fig. 1. Examples of frame prediction on Avenue and ShanghaiTech datasets. For each dataset, the first row is a prediction example of normal event, and the second row is a prediction example of an abnormal event. Left column: the real frame. Mid column: the prediction frame. Right column: the prediction error map. (Best viewed in color.)

Fig. 2. Anomaly score curves of several test video clips of our method on Ped2, Avenue, and ShanghaiTech datasets. Royalblue regions represent ground truth anomalous frames. It shows that the anomaly scores rise dramatically when anomalies occur and decrease when anomalies disappear. (Best viewed in color.)