Dynamic Local Aggregation Network with Adaptive Clusterer for Anomaly Detection

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

1 Guangzhou Institute of Technology, Xidian University, Guangzhou, China

2 School of Computer Science, Northwestern Polytechnical University, Xi'an, China

MOTIVATION

Limitation:

Existing methods for anomaly detection based on memory-augmented autoencoder have the following drawbacks:

- Establishing a memory bank requires additional memory space.
- The fixed number of prototypes from subjective assumptions ignores the data feature differences and diversity.

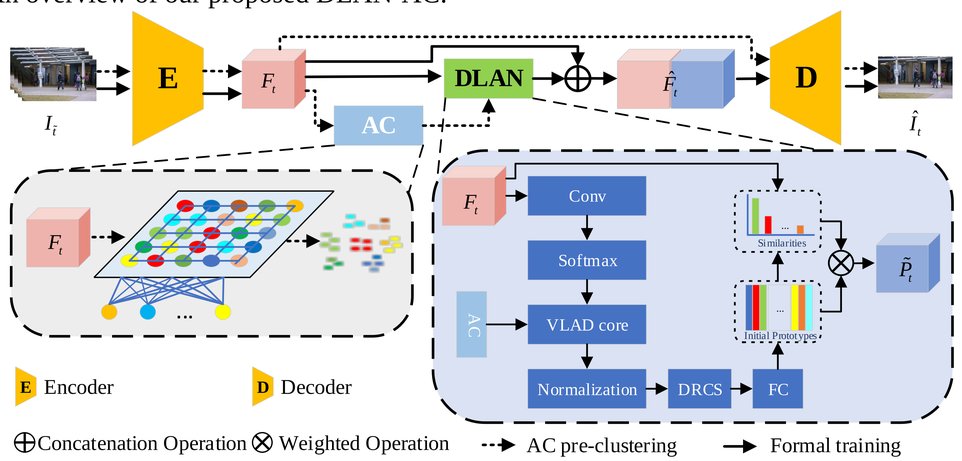
Motivation:

✓ How to automatically learn and aggregate high-level features from the AE to obtain more representative prototypes, while freeing up extra memory space?

✓ Whether it is possible to adaptively cluster video data to derive initial prototypes with prior information, rather than based on subjective assumptions?

METHODOLOGY

- An overview of our proposed DLAN-AC:



Our method mainly consists of four parts: an encoder, an AC, a DLAN, and a decoder. The encoder extracts a high-level feature map F_t of size $H \times W \times D$ from an input video frame sequence $I_{\tilde{t}}$. Then, AC adaptively clusters the high-level features to obtain the prior cluster number M and cluster center vector c_l^* which are used to initialize DLAN. Next, the F_t are further locally aggregated to obtain the prototypes $\tilde{P}_t$ by DLAN. Finally, the $\tilde{P}_t$ and the F_t are fused and fed to the decoder to predict the fifth frame $\tilde{I}_t$.

EXPERIMENTS

- Quantitative comparison with the state of the art for anomaly detection. We measure the average AUC (%) on Ped2, Avenue, and ShanghaiTech. ('R.' and 'P.' indicate the reconstruction and prediction tasks, respectively.)

	Methods	Ped2	Avenue	ShanghaiTech
	MPPCA [17]	69.3%	N/A	N/A
	MPPC+SFA [24]	61.3%	N/A	N/A
	Unmasking [37]	82.2%	80.6%	N/A
	AMC [29]	96.2%	**86.9%**	N/A
	AnomalyNet [53]	94.9%	86.1%	N/A
	DeepOC [43]	**96.9%**	86.6%	N/A
R.	Conv-AE [13]	90.0%	70.2%	60.9%
R.	ConvLSTM-AE [25]	88.1%	77.0%	N/A
R.	Stacked RNN [26]	92.2%	81.7%	68.0%
R.	CDDA [53]	96.5%	86.0%	73.3%
R.	MemAE [11]	94.1%	83.3%	71.2%
R.	MNAD [31]	90.2%	82.8%	69.8%
R.	AMCM [6]	**96.6%**	**86.6%**	73.7%
R.	LNRA-P [3]	94.77%	84.91%	72.46%
R.	LNRA-SF [3]	96.50%	84.67%	**75.97%**
P.	FFP [21]	95.4%	84.9%	72.8%
P.	AnoPCN [48]	96.8%	86.2%	73.6%
P.	IPRAD [36]	96.3%	85.1%	73.0%
P.	MNAD [31]	97.0%	88.5%	70.5%
P.	ROADMAP [39]	96.3%	88.3%	**76.6%**
P.	MPN [27]	96.9%	89.5%	73.8%
P.	**DLAN-AC**	**97.6%**	**89.9%**	74.7%

CONTRIBUTIONS

◆ We propose a novel dynamic local aggregation network with adaptive clusterer (DLAN-AC), where DLAN automatically learns and aggregates high-level features to obtain more representative prototypes of the normal video data; AC adaptively clusters the high-level features of video frames according to the scene, which can provide the prior information for prototype setting.

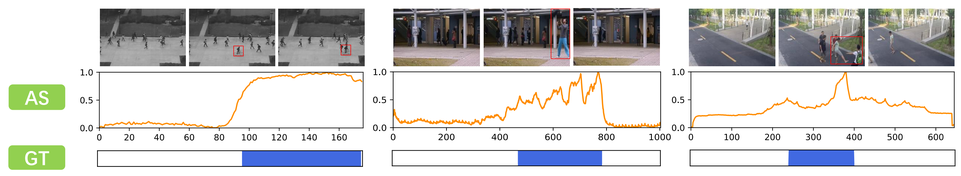
◆ We propose a dynamic redundant clustering strategy (DRCS) to eliminate the unimportant feature clusters and retain the feature clusters that has important contributions to the establishment of a prototype of normal video.

◆ DLAN-AC is on par with or outperforms other existing methods on three benchmarks, and extensive ablation experiments demonstrate the effectiveness of DLAN-AC.

RESULTS

- Example of frame prediction on Ped2 dataset. The first row is a prediction example of normal event, and the second row is a prediction example of an abnormal event.

- Anomaly score curves of several test video clips of our method on three benchmark datasets. (AS: the anomaly score, GT: the ground truth anomalous frame.)



RESULTS

- AC pre-clustering results on the Ped2, Avenue, and ShanghaiTech. The first row is the original image, and the second row is the corresponding clustering result. It can be seen that elements with similar attributes are grouped into the same category.