A Hybrid Video Anomaly Detection Framework via Memory-Augmented Flow Reconstruction and Flow-Guided Frame Prediction

Zhian Liu[1] Yongwei Nie[1*] Chengjiang Long[2] Qing Zhang[3] Guiqing Li[1]
[1]South China University of Technology [2]JD Finance America Corporation [3]Sun Yat-sen University

2021 ICCV OCTOBER 11-17 VIRTUAL

Background & Motivation

- Surveillance system is widely used in many places.
- Video anomaly detection is an essential task to save human labor, with goal to identify unexpected events.
- Since the anomaly rarely happens and is free of forms, we treat it as an unsupervised learning problem with normal data only.

Contributions

- Design a novel hybrid framework, combining flow reconstruction and flow-guided frame prediction, named as HF^2-VAD.
- Design the Multi-Level Memory Autoencoder with Skip Connections (ML-MemAE-SC) for flow reconstruction.
- Propose to model the normal consistency between flows and frames by leveraging the conditional Variational Autoencoder (CVAE).

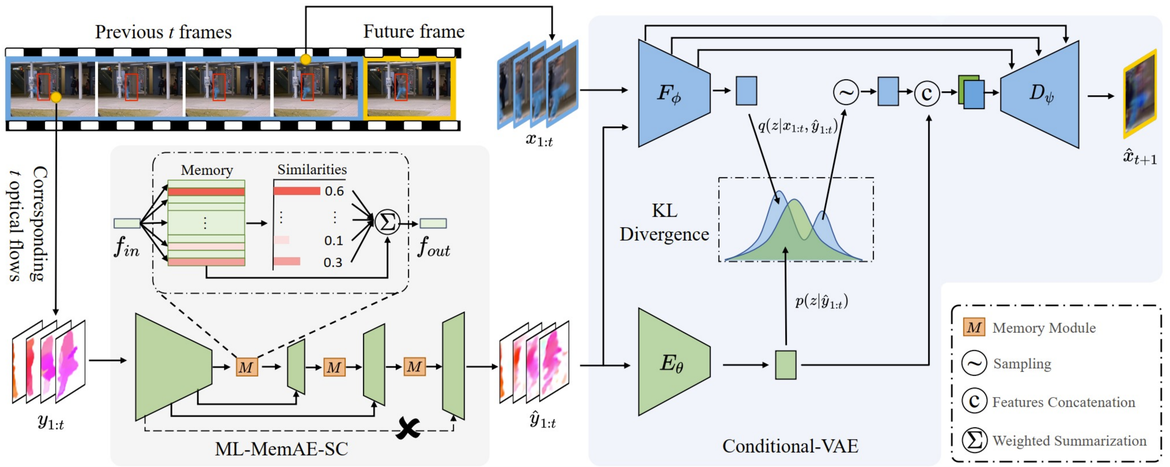
Datasets & Metric & Project

- Datasets: Ped2, Avenue, ShanghaiTech
- Metric: Area Under the Receiver Operation Characteristic (AUROC)
- Code: github.com/LiUzHiAn/hf2vad

Acknowledgment

This research is sponsored in part by the National Natural Science Foundation of China (62072191, 61802453, 61972160), in part by the Natural Science Foundation of Guangdong Province (2019A1515010860, 2021A1515012301), and in part by the Fundamental Research Funds for the Central Universities (D2190670).

Proposed HF^2-VAD



$$\mathcal{L}_{ML-\text{MemAE-SC}} = \lambda_{\text{recon}}\mathcal{L}_{\text{recon}} + \lambda_{\text{ent}}\mathcal{L}_{\text{ent}}$$

$$\mathcal{L}_{\text{recon}} = \|y_{1:t} - \hat{y}_{1:t}\|_2^2 \quad \mathcal{L}_{\text{ent}} = \sum_{i=1}^{M}\sum_{k=1}^{N} -\hat{w}_{i,k}\log(\hat{w}_{i,k})$$

$$\mathcal{L} = \lambda_{CVAE}L_{CVAE} + \lambda_{gd}||\Delta\hat{x}_{t+1} - \Delta x_{t+1}||_1$$

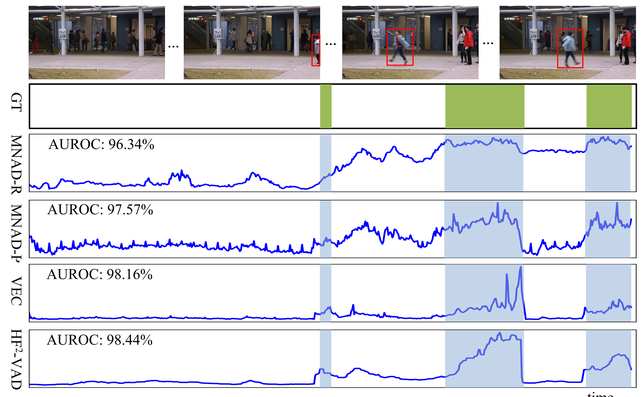
$$\mathcal{L}_{CVAE} = KL[q(z \mid x_{1:t}, y_{1:t})\|p(z \mid y_{1:t})] + \|x_{t+1} - \hat{x}_{t+1}\|_2^2$$

Anomaly score $S_r = \|\hat{y}_{1:t} - y_{1:t}\|_2^2 \quad S_p = \|\hat{x}_{t+1} - x_{t+1}\|_2^2 \quad S_{O_i} = w_r \cdot \frac{S_r - \mu_r}{\sigma_r} + w_p \cdot \frac{S_p - \mu_p}{\sigma_p}$

- Quantitative comparisons

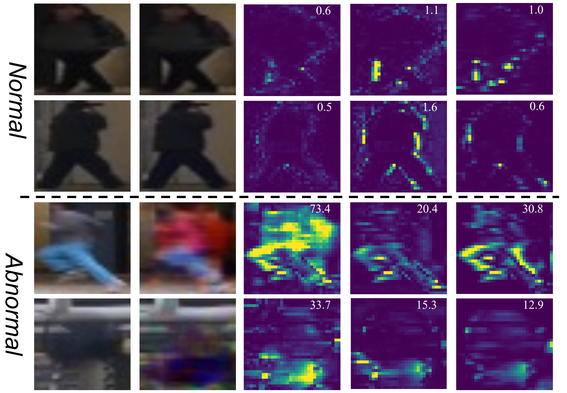
	Method	UCSD Ped2	CUHK Avenue	SHTech
Recon.	Conv-AE [11]	90.0	70.2	-
	ConvLSTM-AE [32]	88.1	77.0	-
	GMFC-VAE [7]	92.2	83.4	-
	MemAE [8]	94.1	83.3	71.2
	MNAD-R [39]	90.2	82.8	69.8
Pred.	Frame-Pred. [26]	95.4	85.1	72.8
	Conv-VRNN [31]	96.1	85.8	-
	MNAD-P [39]	97.0	88.5	70.5
	VEC [50]	97.3	90.2	74.8
Hybrid	ST-AE [53]	91.2	80.9	-
	AMC [37]	96.2	86.9	-
	AnoPCN [49]	96.8	86.2	73.6
	HF^2-VAD w/o FP	98.8	86.8	73.1
	HF^2-VAD w/o FR	94.5	90.2	76.0
	HF^2-VAD	**99.3**	**91.1**	**76.2**

- Qualitative comparisons



Abnormal event: kid running.

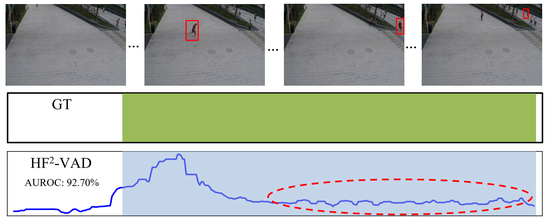
Detailed Visualization



Ablation study

	Memory-augmented Reconstruction Models	Prediction Models		AUROC
		VAE	CVAE	
Flow	✓			96.27
	✓			97.75
	✓			98.81
Frame		✓		89.96
			✓	94.48
Hybrid	✓		✓	96.91
	✓		✓	98.28
	✓		✓	**99.31**

Failure case



Fail when objects are far away from camera.