UMassAmherst

Manning College of Information & Computer Sciences

EVAL: Explainable Video Anomaly Localization

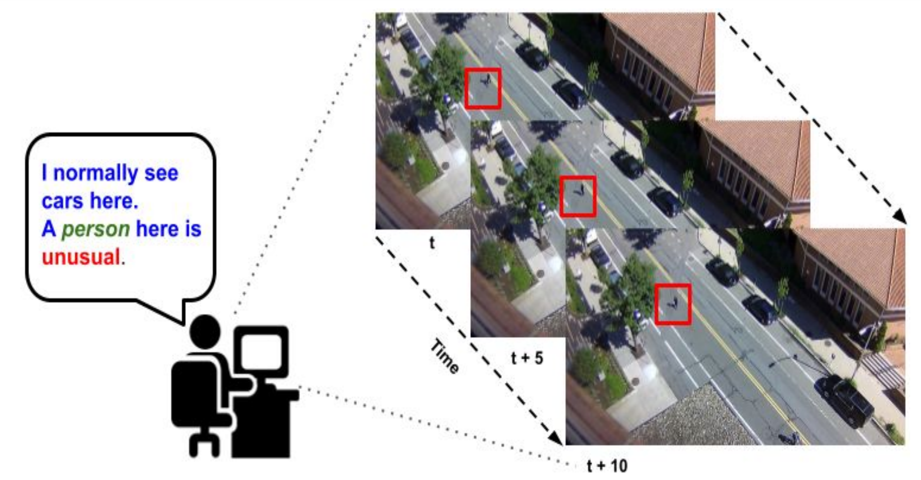
Ashish Singh, Michael Jones, Erik G. Learned-Miller

University of Massachusetts Amherst, Mitsubishi Electric Research Laboratories (MERL)

MITSUBISHI ELECTRIC Changes for the Better

MITSUBISHI ELECTRIC RESEARCH LABORATORIES, INC

Motivation

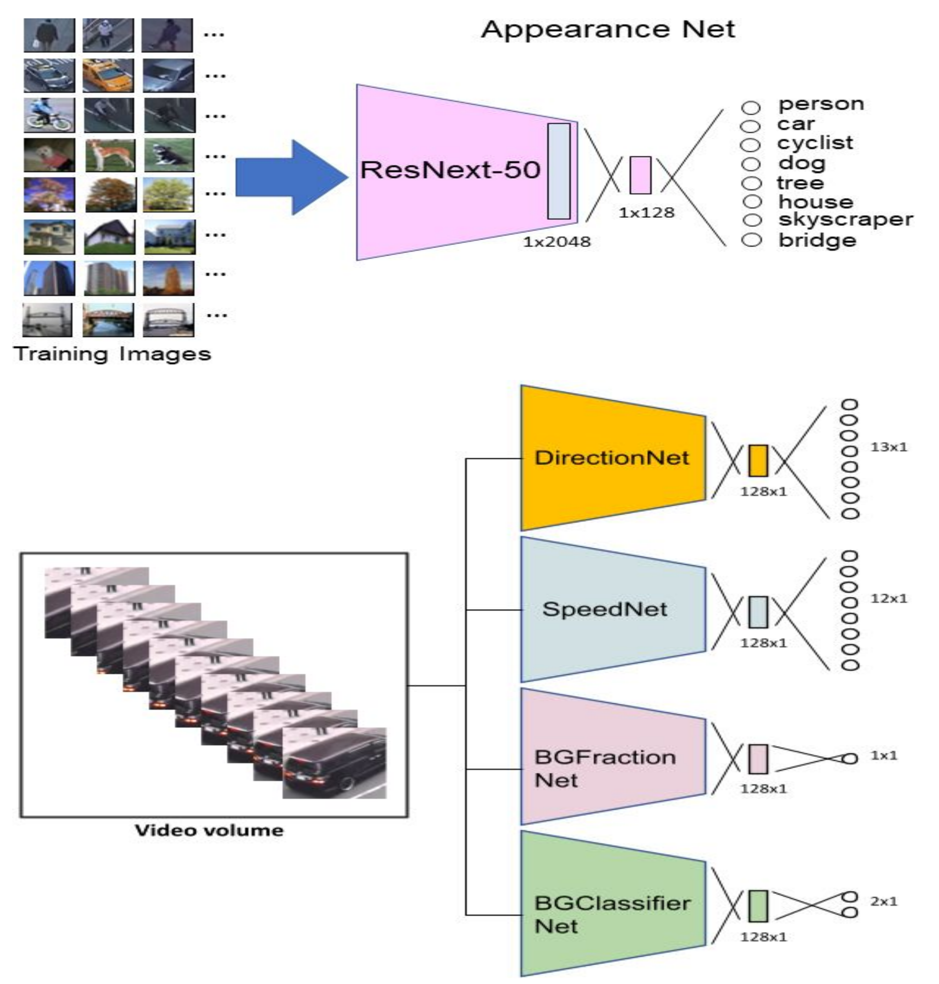


- We are interested in the problem of spatio-temporal localization of anomalous activities in videos of a given scene.
- Our aim is to design an interpretable, robust and accurate anomaly detection system.
- We are motivated by how humans are able to detect changes in a given scene after exposure to it by decomposing it into specific objects and their corresponding motion patterns.

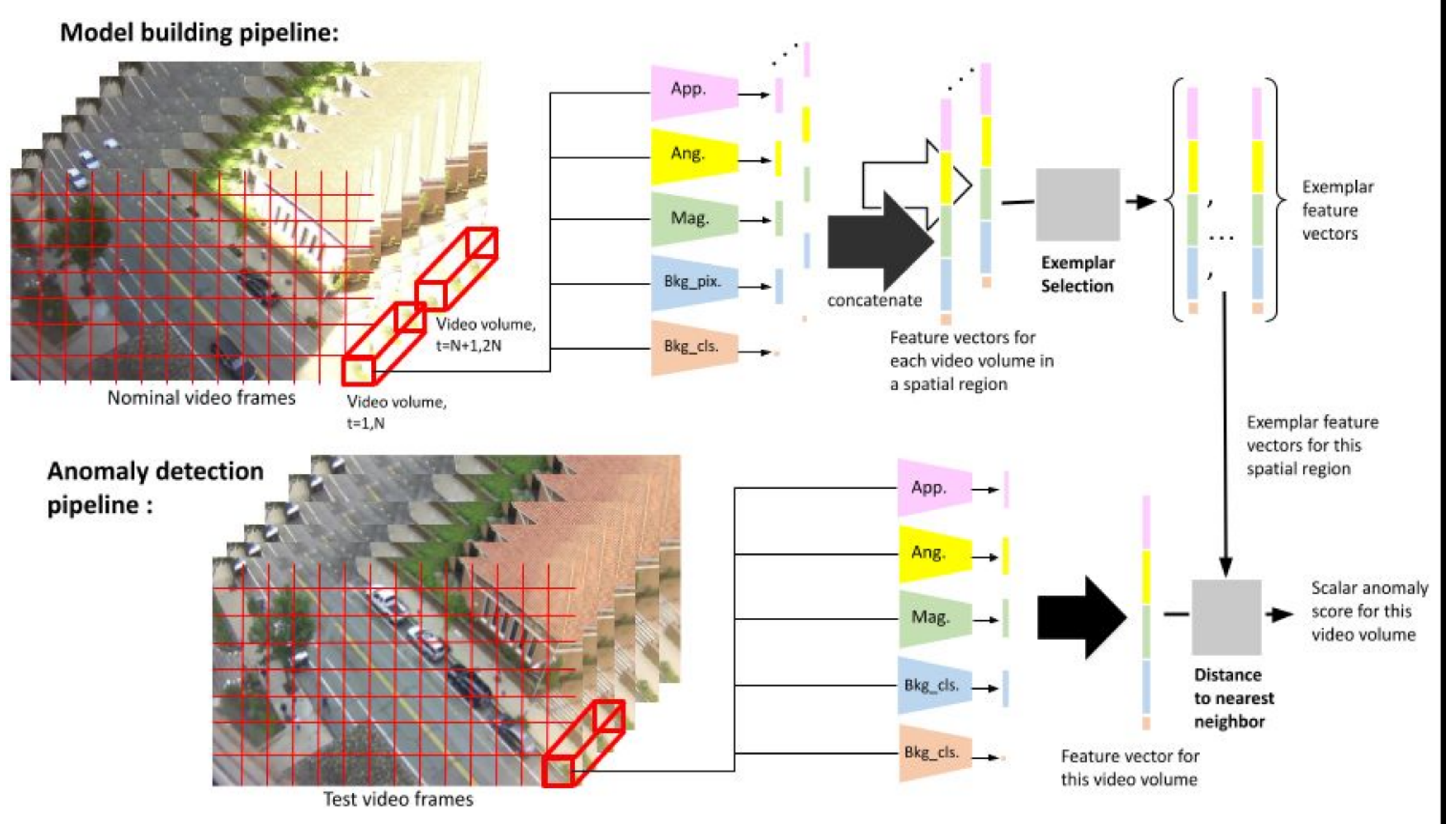
Method

Our method consists of three stages: Attribute Learning, Model Building, and Anomaly Detection.

- **Attribute Learning:** Attributes are human-understandable features of the objects present in a video volume. They give the system a high-level understanding of what is going on in a scene. In this work, the attributes we use are (a) Object Classes and (b) Motion directions, speeds and fraction of stationary pixels. We train neural networks to estimate these attributes given a video volume. The networks are only trained once (not trained for each different scene).
- **Model Building:** Given nominal video of a scene, learn a set of exemplars (attribute embeddings) for each different spatial region of the scene. Exemplar selection uses a simple greedy algorithm.
- **Anomaly Detection:** Given test video from the same scene, compute attribute embeddings for each video volume and find nearest neighbor exemplar from the nominal model. Distance to nearest neighbor is the anomaly score.



Stage 1: Attribute Learning



Stage 2: Model Building | Stage3: Anomaly Detection

Experimental Results

Threshold	UCSD Ped1 RBDC	UCSD Ped1 TBDC	UCSD Ped1 NUM EX	UCSD Ped2 RBDC	UCSD Ped2 TBDC	UCSD Ped2 NUM EX
3	36.9	77.8	288	64.8	89.1	350
2.5	49.4	89.4	424	78.8	93.7	761
2	57.5	89.6	944	84.7	96.0	1339
1.5	**61.7**	**88.9**	**4201**	**87.4**	**95.1**	**4470**
1.0	61.5	87.5	19926	87.4	95.8	19138
0.5	61.4	87.7	49113	87.2	95.1	34862
0.25	61.5	87.8	57636	87.2	95.1	45795

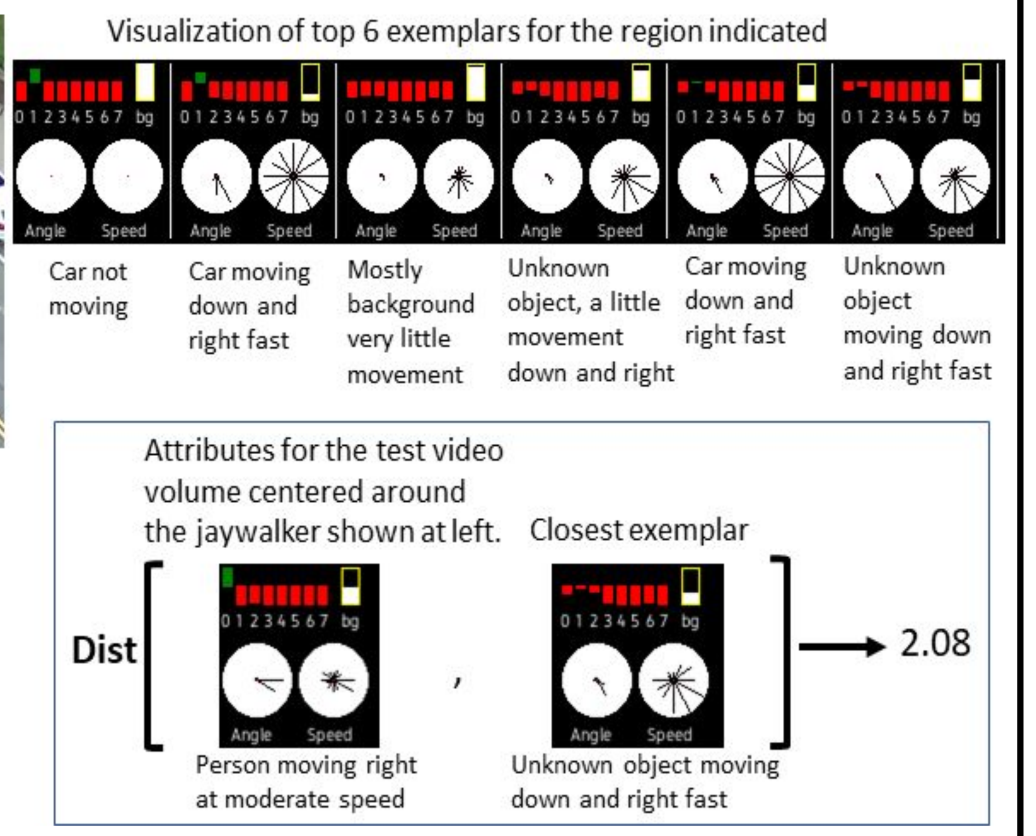
Methods	Avenue RBDC	Avenue TBDC	Avenue Frame	ShanghaiTech RBDC	ShanghaiTech TBDC	ShanghaiTech Frame
Ionescu (2019)	15.77	27.07	87.4	20.65	44.54	78.7
Ramachandra (2020)	35.80	80.90	72.0	-	-	-
Ramachandra (2020)	41.20	78.60	87.2	-	-	-
Georgescu (2021)	57.00	58.30	91.5	42.80	83.90	90.02
Liu (2018)	19.59	56.01	85.1	17.03	54.23	72.8
Liu (2021)	41.05	86.18	89.9	44.41	83.86	74.2
Georgescu (2021)	65.05	66.85	92.3	41.34	78.79	82.7
Liu + Ristea (2021)	20.13	62.30	87.3	18.51	60.22	74.5
Liu + Ristea (2021)	62.27	89.28	90.9	45.45	84.50	75.5
Georgescu + Ristea (2021)	65.99	64.91	92.9	40.55	83.46	83.6
Our Method	68.2	87.56	86.02	59.21	89.44	76.63

Street Scene

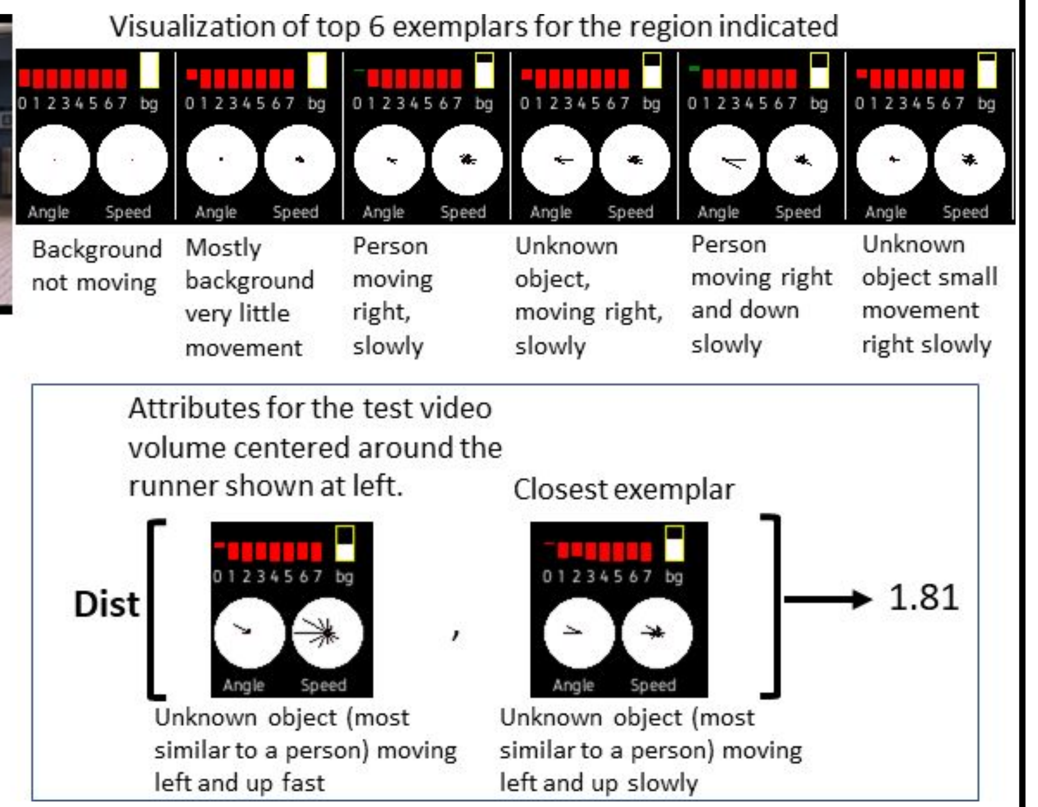
Methods	RBDC	TBDC
Auto-Encoder	0.29	2.0
Dictionary Methods	1.6	10.0
Flow Baseline	11.0	52.0
FG Baseline	21.0	53.0
Our Method	24.26	64.5

Explanation Visualization

Frame 1719 of Street Scene Test005. The bounding box indicates a particular spatial region containing a ground truth anomaly (jaywalking).



Frame 400 of Avenue Test004. The bounding box indicates a particular spatial region containing a ground truth anomaly (running).



Frame 119 of UCSD Ped2 Test001. The bounding box indicates a particular spatial region containing a ground truth anomaly (cyclist on sidewalk).