HACETTEPE UNIVERSITY

VADOR: Real World Video Anomaly Detection with Object Relations and Action

Halil İbrahim Öztürk, Ahmet Burak Can

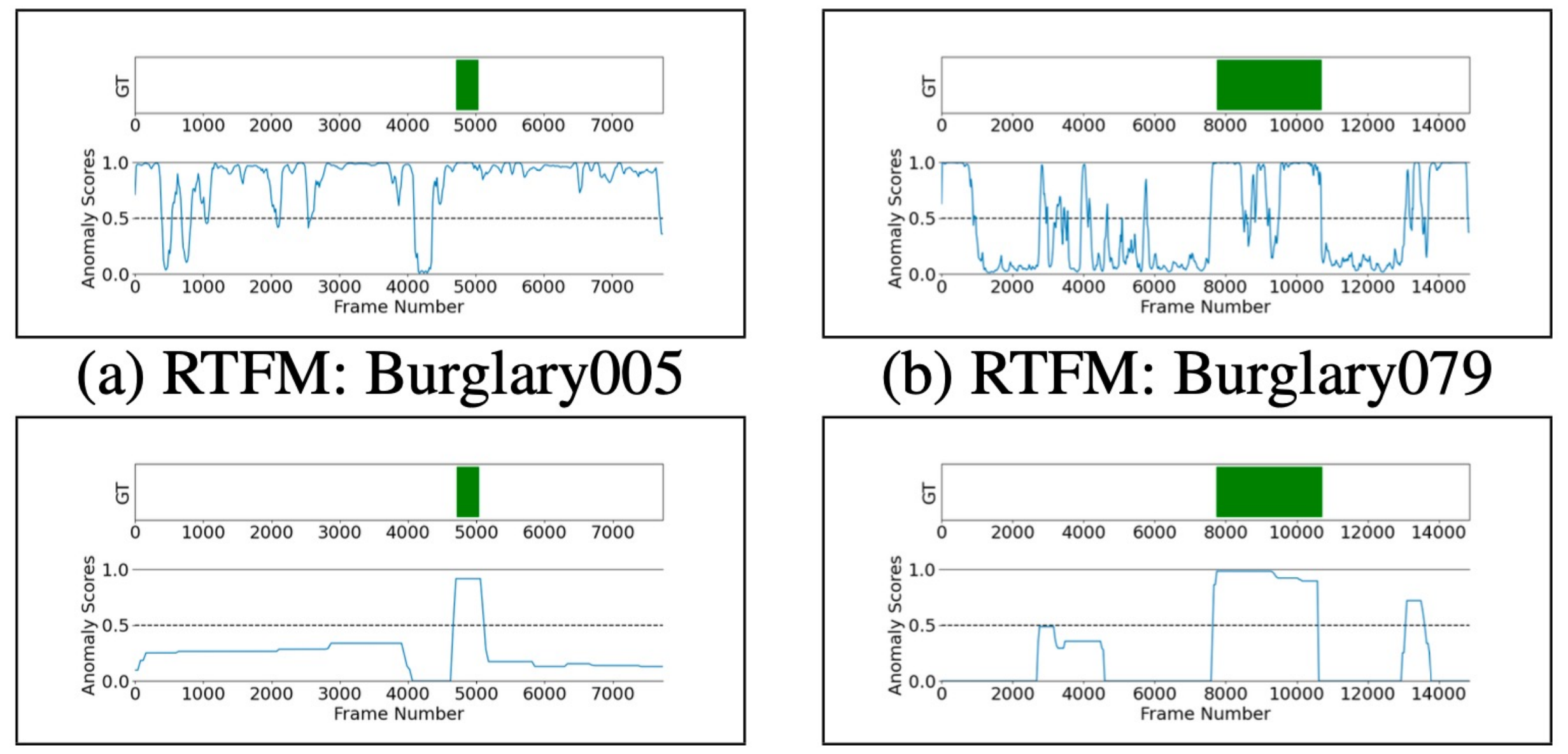
BMVC 2023

ABSTRACT

- Identifying anomalies in real-world scenarios is a challenging task that cannot solely rely on action-based knowledge
- To effectively recognize such complex actions, it becomes crucial to consider the objects involved and their interrelationships within the contextual scenes.
- We propose VADOR, a method understands complex scenes through the integration of action information and object relations

CONCLUSIONS

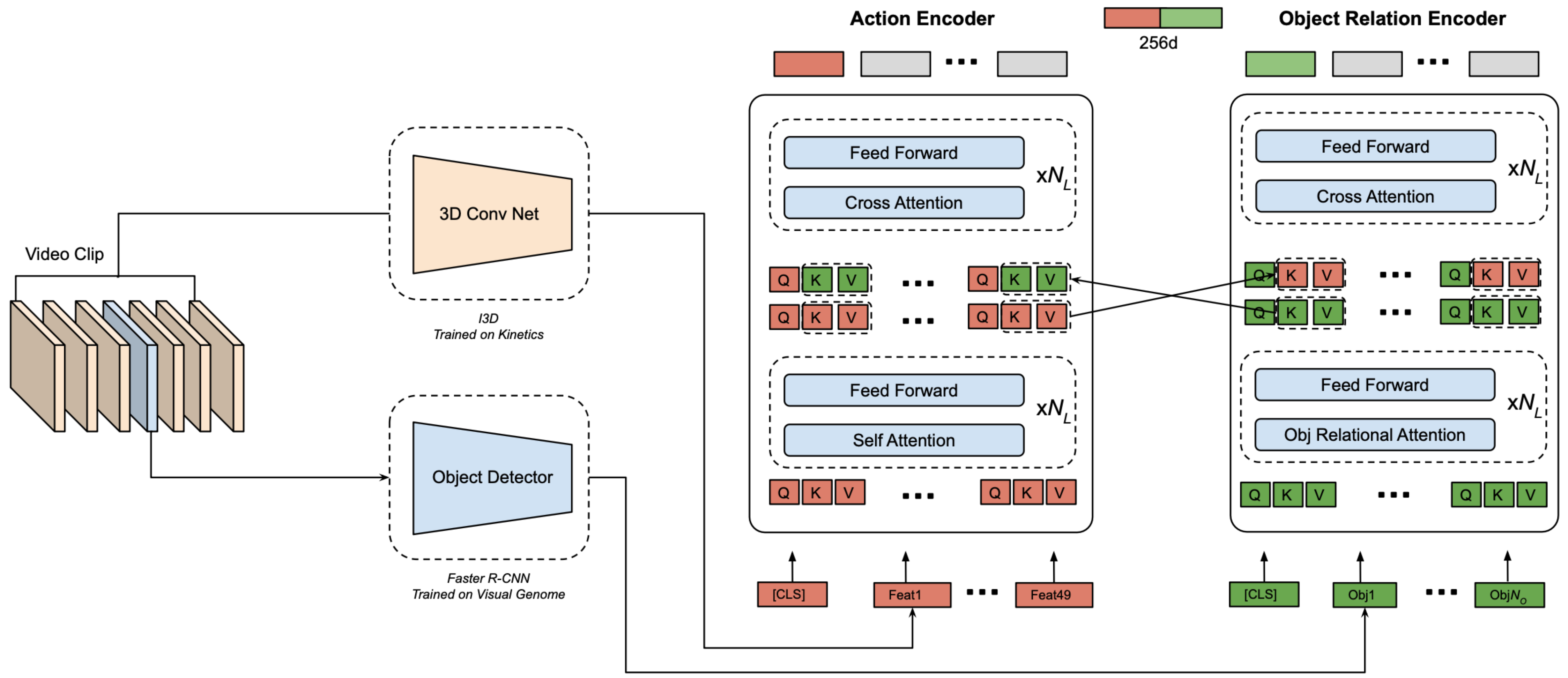
- ✓ Fusion of action and object relation information increases performance of VADOR
- ✓ Qualitative and quantitative results show that transformer encoders with cross attention layers provides better temporal anomaly segmentation performance



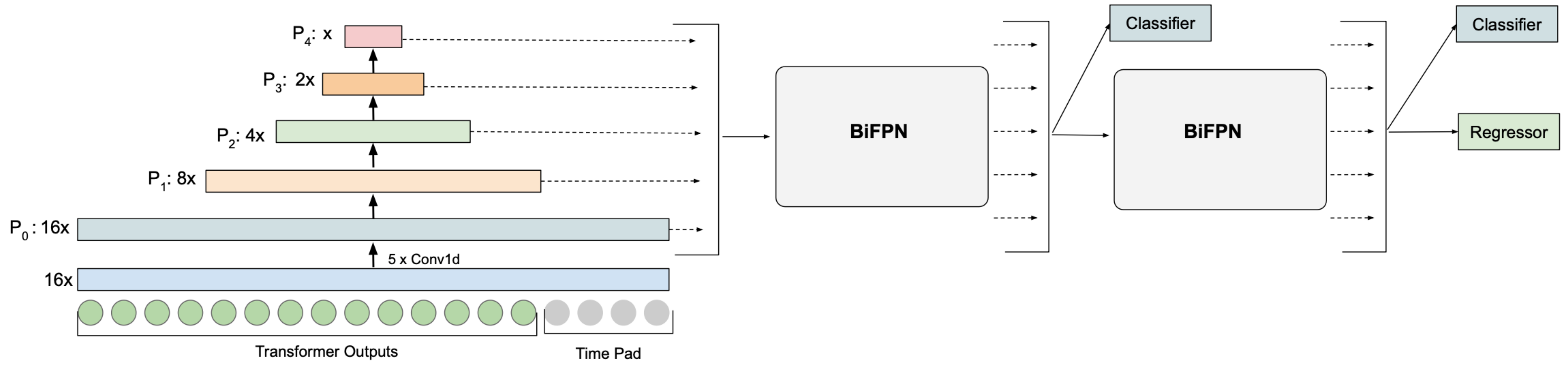
(a) RTFM: Burglary005 (b) RTFM: Burglary079

(a) Ours: Burglary005 (b) Ours: Burglary079

METHOD



- VADOR employs a two-stage approach, Video Clip Encoders (VCE) in first stage generates video clip features, TALNet in second stage localizes anomalies in time by using sequentially organized video clip features.
- VCE involves the object relation encoder and the action encoder.
  - The object relation encoder processes the object features and bounding boxes.
  - The action encoder handles the action features.
- Cross-attention layers between the encoders enable cross-relations between objects and actions within the same video clip.
- TALNet consists of 1D convolution layers, temporal BiFPN blocks and dense prediction heads. The model is similar to dense object detection methods.



RESULTS

	UCF Crime			
Methods	F1@10	F1@25	F1@50	AUC
Sultani *et al.*[15]	45.20	39.64	32.32	75.41
RFTM [17]	33.55	26.14	16.86	84.44
S3R [20]	43.30	33.43	21.76	**85.99**
ADNet [13]	58.16	51.85	41.29	70.57
TALNet w/o encoders	62.72	57.36	43.40	69.37
VADOR (ours)	**69.79**	**63.09**	**50.28**	83.62

- ✓ While VADOR's clip based AUC score of 83.62 is lower than S3R's score of 85.99, there is a significant difference in temporal F1 scores

	XD-Violance			
Methods	F1@10	F1@25	F1@50	AP
TALNet	36.65	26.43	12.67	51.30
RFTM [17]	41.23	31.05	15.28	58.35
S3R [20]	44.26	31.19	14.75	61.96
VADOR	**49.74**	**40.41**	**25.07**	**65.90**

- ✓ We evaluated UCF-Crime trained models on XD-Violance dataset. The results proof VADOR's generalization ability

	UCF Crime			
Methods	F1@10	F1@25	F1@50	AUC
VADOR only action	40.85	24.19	14.54	69.36
VADOR only object	65.78	57.78	42.75	74.50
VADOR cross-attention	**69.79**	**63.09**	**50.28**	**83.62**

- ✓ The results show that encoders with cross attention is important to get better performance. Furthermore, the results show that object relations are more useful than action to recognize anomalies in UCF Crime dataset.