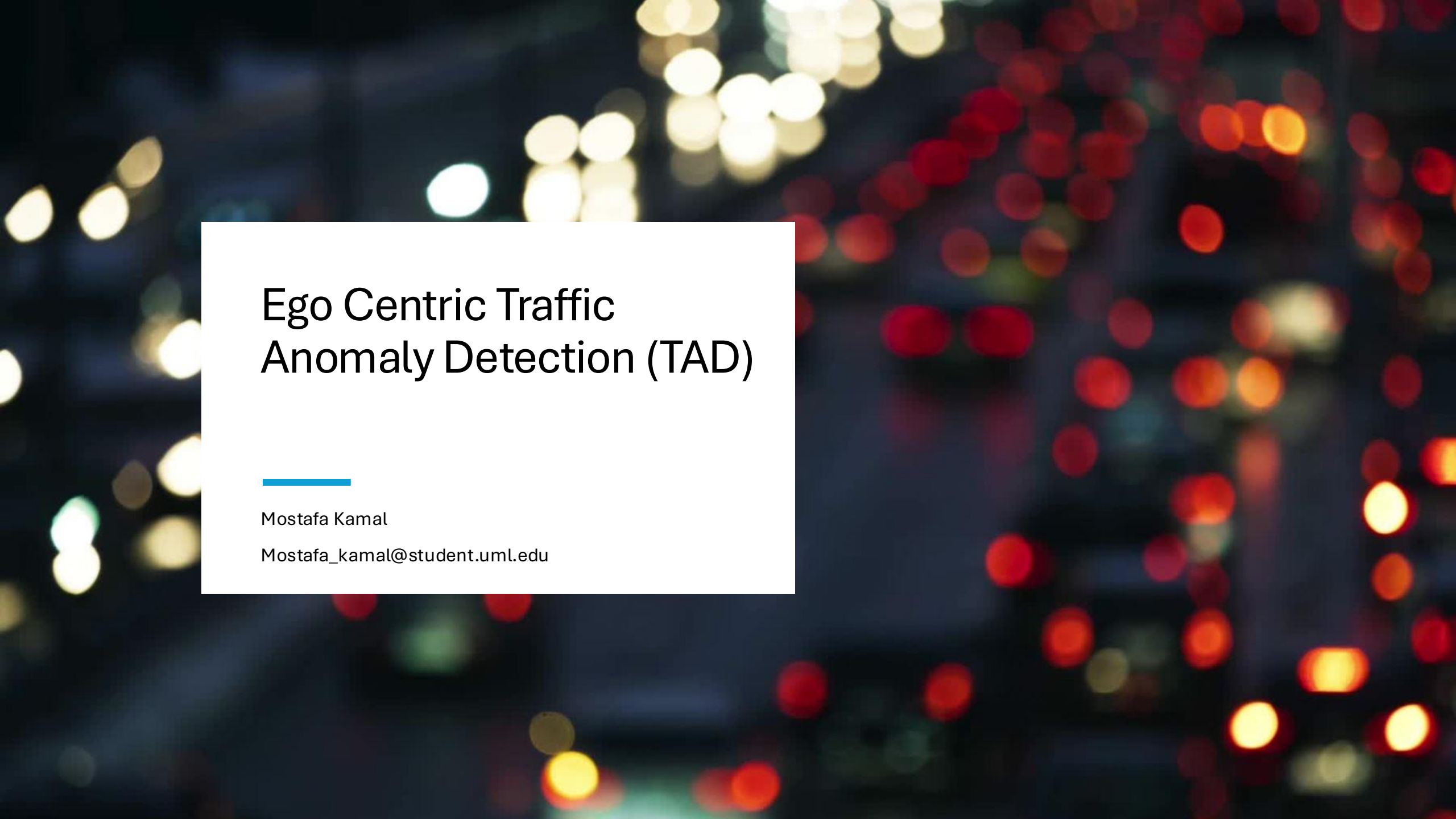


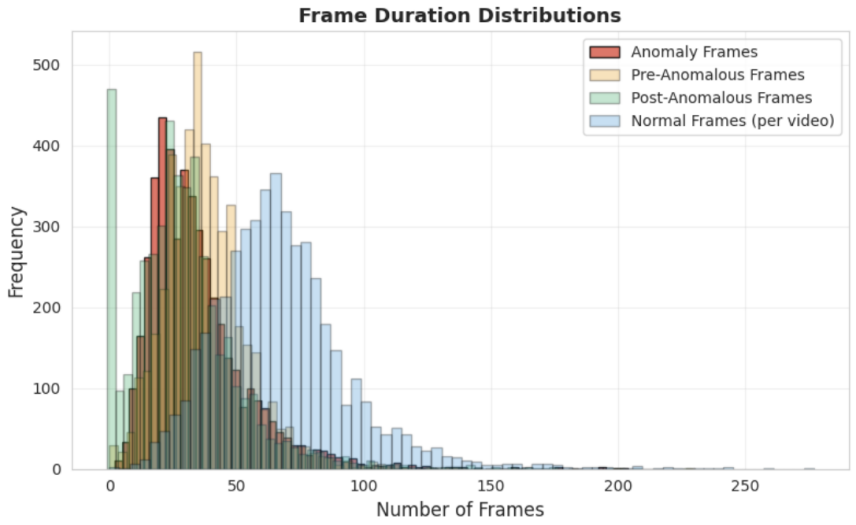
Ego Centric Traffic Anomaly Detection (TAD)

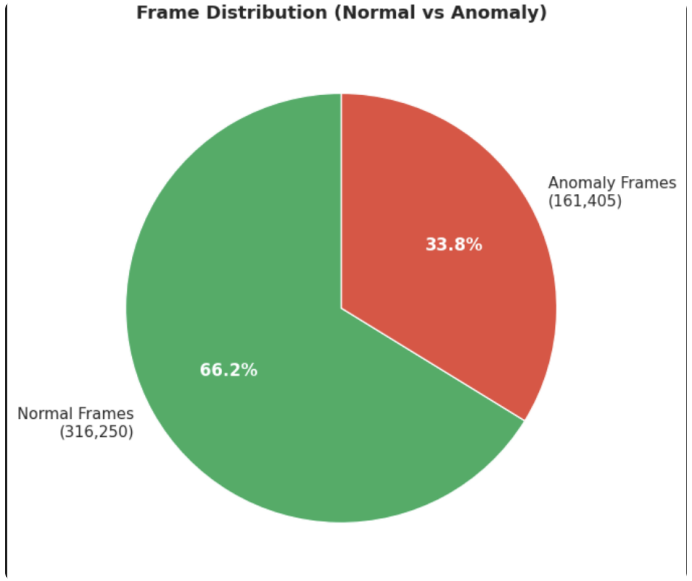
Mostafa Kamal

Mostafa_kamal@student.uml.edu

Introduction & Motivation

- Rising road traffic accidents cause over 1.19 million deaths annually worldwide
- Anomalies detections are critical in real-world applications
  - Monitoring/surveillance
  - Healthcare systems,
  - Intrusion detection
  - Autonomous driving.
- Failure to detect in autonomous vehicles causes life threatening injuries
- Task Definition: Traffic Anomaly Detection (TAD) using ego-centric dashcam videos
  - Is there an anomaly present?
  - When does the anomaly occur within the video?
  - Challenges: Dynamic moving videos, potential class imbalance, subtle differentiation





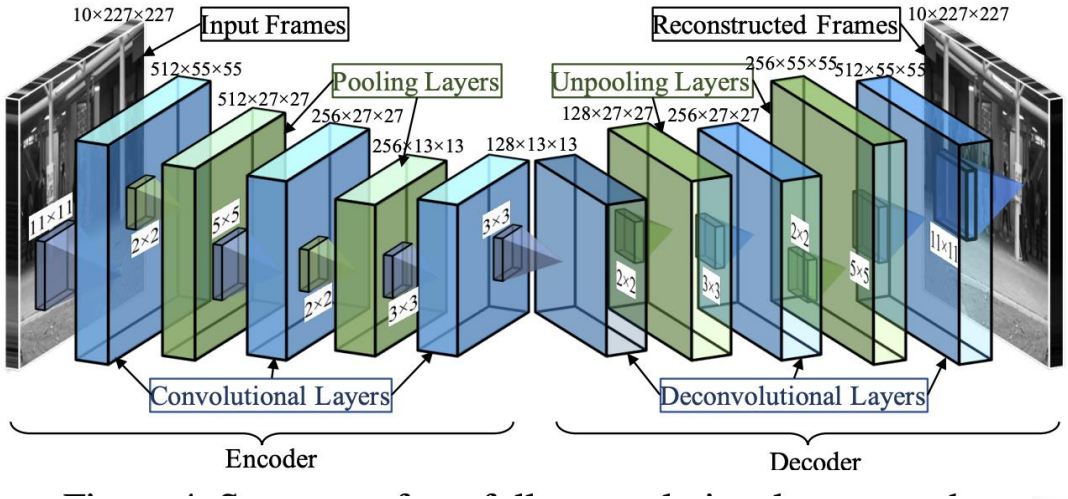
Dataset

- **DoTA Dataset Overview**
  - Detection of Traffic Anomaly (DoTA) dataset for dash-cam video analysis
  - 4,677 video clips from YouTube capturing real-world driving scenarios
  - 10 fps extraction with diverse weather and driving conditions
  - Frame-level annotations: Normal vs. Accidents
  - Class distribution: 66.2% Normal, 33.8% Anomalous frames
- **Data Partitioning**
  - 80/10/10 train/validation/test split
  - Random video-level partitioning prevents data leakage

Preprocessing

- Convolution Autoencoder:
  - Grayscale conversion with z-score normalization
  - 10-frame temporal stacks (10, 224, 224)
  - 3-frame overlap for data augmentation
  - Training on normal frames only
- VideoMAE + Classifier:
  - 16-frame sequences with sequence-level labels
  - Pretrained model preprocessing to 4D tensors of 16x3x224x224 shape
  - A video is labeled anomalous if it holds any anomaly frame; otherwise, it is normal clip

Method 1: Conv-AE (Architecture)



Source: Hasan et al. (2017)

- Core Features:
  - Symmetric encoder-decoder design for unsupervised anomaly detection
  - Trained on Normal Frames only
  - Stack of 10 consecutive grey scale frames for catching temporal regularities
- Architecture:
  - Encoder: 3 Convolution Block each containing:
    - Convolution Layers
    - batch norm
    - Rectified Linear Units (ReLU)
    - Maxpooling
  - Decoder: 3 Deconvolution Blocks
    - Transpose Convolution
    - Rectified Linear Units (ReLU)
    - Batch norm
  - Symmetric filter, padding, strides
  - Output padding to match input shape

Method 1: Conv-AE (Anomaly Detection)

Question: How do we detect anomalies/accidents?

Recall: Model is trained to learn normal driving patterns (spatial + temporal)

Reconstruction Error

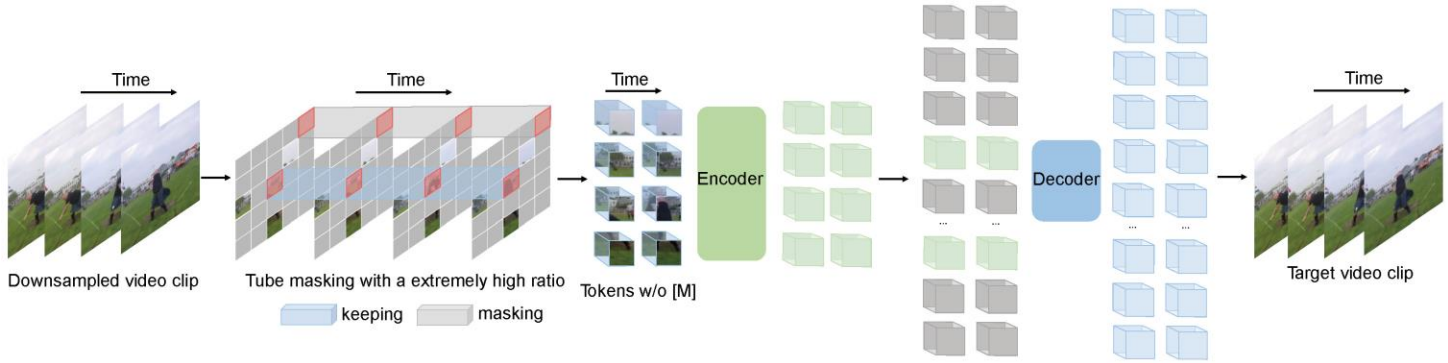
- High error identifies an anomalous frame
- Low error correlates to a normal frame
- Calculated using l2 norm of pixel intensity of reconstrued and original frame

Regularity Scoring

- High score -> normal frame
- Low score -> anomalous frame
- Captures regularity in normal frames
- Ranges: [0, 1]

Method 2: (VideoMAE + Classifier)

- Leveraged pre-trained VideoMAE for spatiotemporal feature extraction
  - Model name: *"MCG-NJU/videomae-base-finetuned-kinetics"*
  - MVM self-supervised pretraining approach
    - Works by learning masking out clips (spatial/temporal)
- Freeze first 9 layers and fine tune last 3 hidden layers
- Attached a classification head for binary output
  - Used binary cross entropy loss function
- Better results: Further pretraining on Domains Data



Source: Hugging face

Evaluation & Results

We evaluate our both our model on AUC-ROC metric

For supervised approach, we further evaluate it with, precision, recall, and F1 scores

Model Name	AUC-ROC Score
Conv-AE (Baseline)	0.556
VideoMAE + classifier	**0.828**

VideoMAE outperforms Conv-AE by 49%

Evaluation & Results

VideoMAE + Classifier
(further evaluation)

TABLE III: Per-Class Performance Metrics - VideoMAE Classifier

Class	Precision	Recall	F1-Score	Support
Normal	0.7437	0.8573	0.7965	1472
Anomaly	0.8094	0.6722	0.7345	1327
Weighted Avg	**0.775**	**0.769**	**0.767**	**2799**

TABLE II: Confusion Matrix - VideoMAE Classifier

	Predicted	
	Normal	Anomaly
Normal	TN: 1262	FP: 210
Anomaly	FN: 435	TP: 892

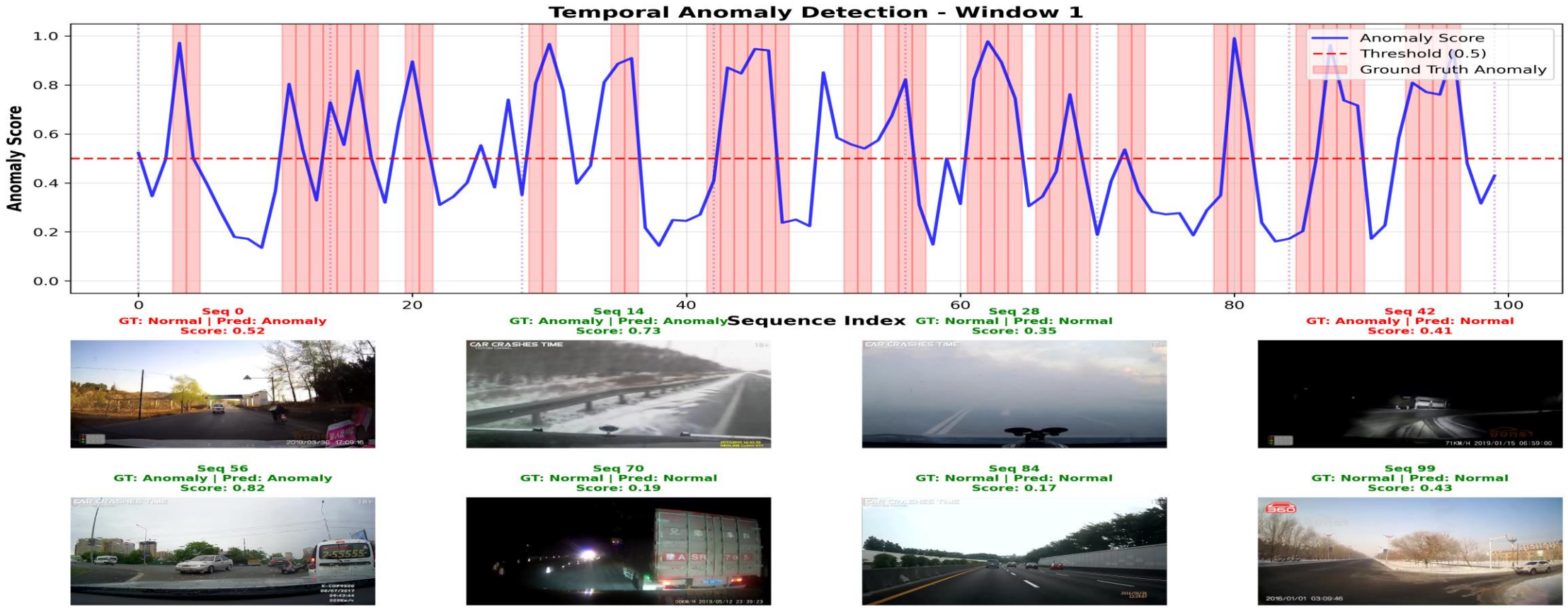
- VideoMAE classifier achieves better precision for anomaly detection (80.9%) but higher recall for normal events (85.7%)
- Useful for TAD

Visualization (Conv AE)

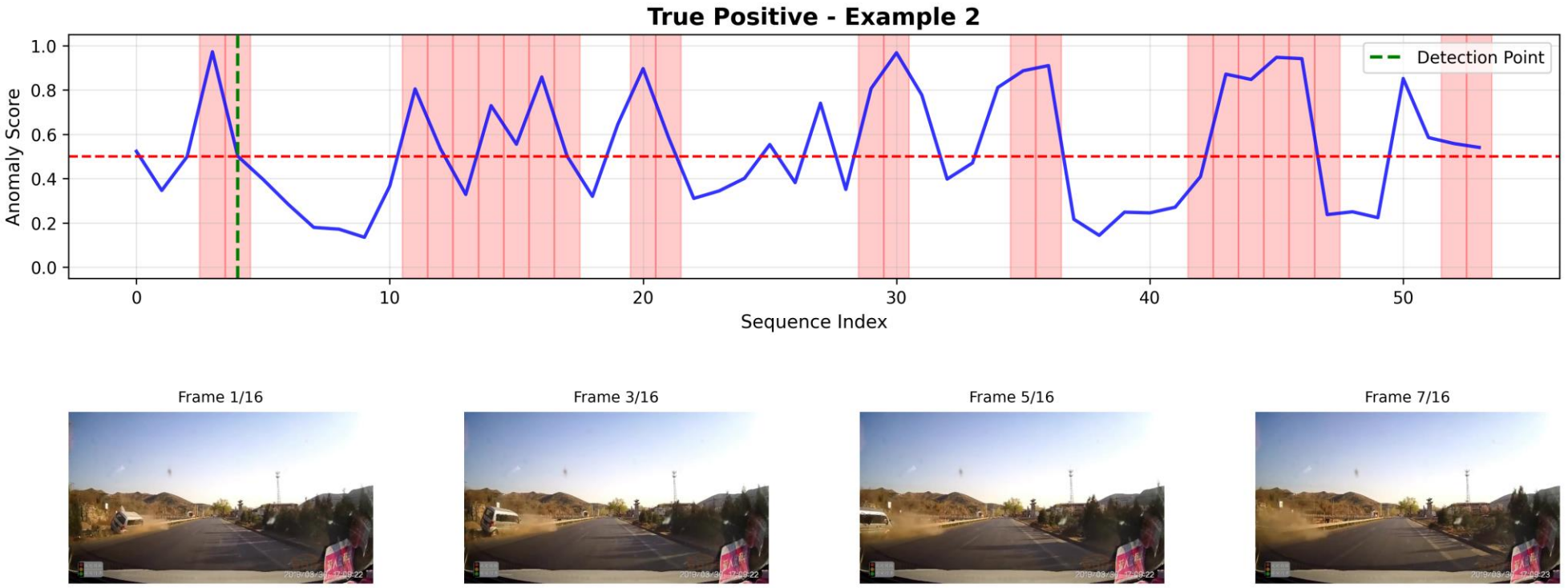
Reconstruction Error heat map

- Red hot shows high error region
- Can Implicitly find out anomaly location

Visualization (ViViT)

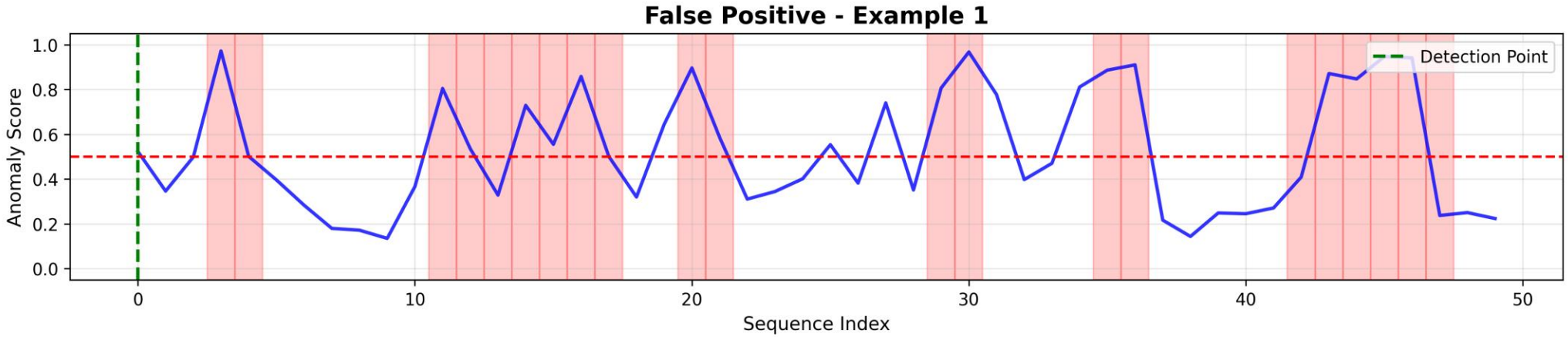


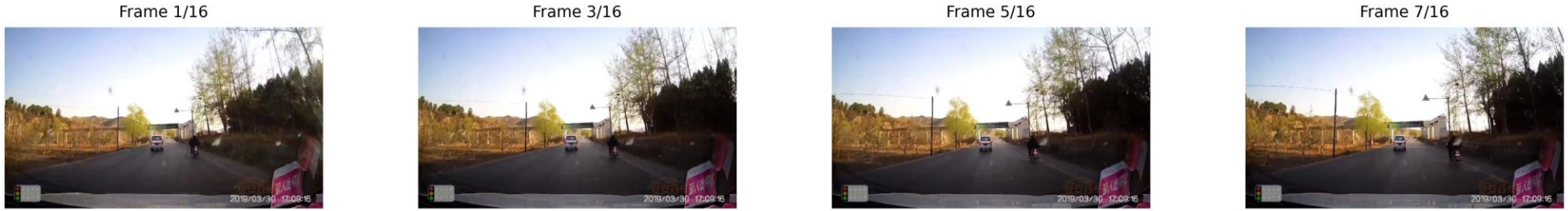
More Visualization (ViViT -- True Positive)



GT: Anomaly, Pred: Anomaly

More Visualization (ViViT – False Positive)





GT: Normal, Pred: Anomaly

Limitation & Future Work

- Complexity in ego centric motion
- Domain adaptability
- False positives in rare events

Future work can extend in:

- Detecting bounding box of anomalous objects
- Extends to accident level classification
- Heavy data augmentation (night/day settings, accident categories etc.)
- Further pretraining on VideoMAE using MVM strategy for domain adaptability in driving scenarios

REFERENCES

[1] Y. Yao et al., "DoTA: Unsupervised Detection of Traffic Anomaly in Driving Videos," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 1, pp. 444-459, 1 Jan. 2023, doi: 10.1109/TPAMI.2022.3150763

[2] Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A.K., Davis, L.S.: Learning temporal regularity in video sequences. In: CVPR (2016)

[3] Medel, J.R., Savakis, A.: Anomaly detection in video using predictive convolutional long short-term memory networks. arXiv:1612.00390 (2016)

[4] Y. S. Chong and Y. H. Tay, "Abnormal Event Detection in Videos Using Spatiotemporal Autoencoder," in Advances in Neural Networks—ISNN (Springer International Publishing, 2017): 189–196

[5] R. Muhammad, E. Amparore, E. Ferrari, and D. Verda, "Can I Trust My Anomaly Detection System? A Case Study Based on Explainable AI," in Proc. World Conf. Explainable Artif. Intell., Valletta, Malta, 2024, pp. 243-254, doi: 10.1007/978-3-031-63803-9_13

[6] Y. Yao, M. Xu, Y. Wang, D. J. Crandall and E. M. Atkins, "Unsupervised Traffic Accident Detection in First-Person Videos," 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 2019, pp. 273-280, doi: 10.1109/IROS40897.2019.8967556.

[7] Y. Yao et al., "DoTA: Unsupervised detection of traffic anomaly in driving videos," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 1, pp. 444–459, Jan. 2023.

[8] E. Orlova et al., "Simplifying Traffic Anomaly Detection with Video Foundation Models," in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW), (2025).

[9] Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Luci˘ c, and Cordelia Schmid. Vivit: A video ´ vision transformer. In Proceedings of the IEEE/CVF international conference on computer vision, pages 6836–6846, 2021.