Author : Kena Hemnani
Date : 8th July 2024

Anomaly Detection in Traffic Videos:

Objective:
Develop a computer vision model for anomaly detection in traffic videos. The system should be capable of identifying unusual or unsafe activities within the footage. Examples of this can be eg, “near misses” (eg car that nearly hit a pedestrian) or unusual (heavy) vehicles.

Solution:

Dataset:

DoTA : Detection of Traffic Anomaly Dataset contains 4077 videos captured from egocentric vehicle view cameras as well as traffic surveillance cameras. It is the largest traffic video anomaly dataset and the first containing detailed temporal, spatial, and categorical annotations.

DoTA provides rich annotation for each anomaly:

1. temporal annotation [**When**?]
   per-frame anomaly scores to answer the When question
2. anomalous object bounding box tracklets [**Where**?]
   per-object anomaly scores as intermediate step to implicitly answer the Where question
3. type (category) [**What**?]
   classifies video type to answer the What question

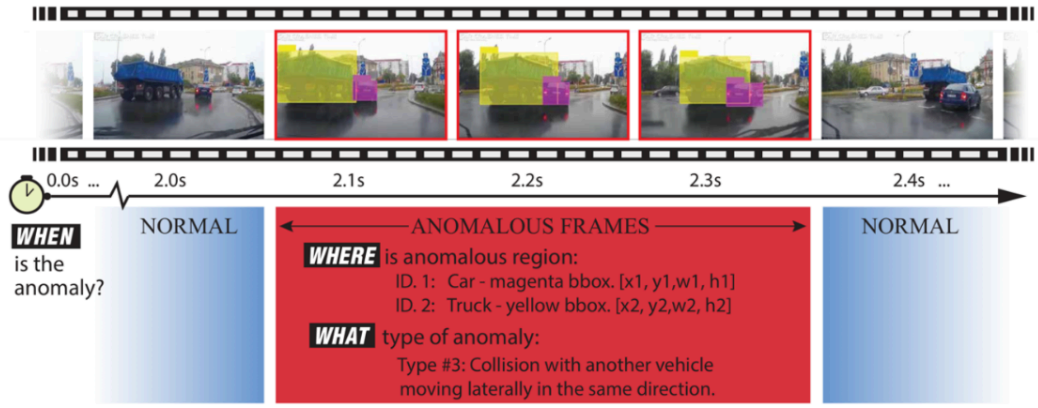

Fig. 1: Overview of DoTA dataset. Annotations are provided to answer the **When**, **Where** and **What** questions for driving video anomalies.

The anomaly categories annotated in the dataset are as follows:

Table 2: Traffic anomaly categories in the DoTA dataset

ID	Short	Anomaly Categories
1	ST	Collision with another vehicle which starts, stops, or is stationary
2	AH	Collision with another vehicle moving ahead or waiting
3	LA	Collision with another vehicle moving laterally in the same direction
4	OC	Collision with another oncoming vehicle
5	TC	Collision with another vehicle which turns into or crosses a road
6	VP	Collision between vehicle and pedestrian
7	VO	Collision with an obstacle in the roadway
8	OO	Out-of-control and leaving the roadway to the left or right
9	UK	Unknown

Model:

MOVAD : **Memory-Augmented Online Video Anomaly Detection** Model, a novel architecture for the frame-level VAD task. MOVAD operates in an online fashion, effectively managing the most restrictive VAD scenarios through end-to-end training, utilizing only RGB frames.

MOVAD is built on two primary modules: a Short-Term Memory Module, which extracts information about ongoing actions using a Video Swin Transformer (VST), and a Long-Term Memory Module integrated into the classifier, leveraging a Long-Short Term Memory (LSTM) network to consider past information and action context. MOVAD's strengths lie not only in its outstanding performance but also in its simple and modular design. It is trained end-to-end with only RGB frames and minimal assumptions, making

it easy to implement and experiment with. It is trained and evaluated on the Detection of Traffic Anomaly (DoTA) dataset.

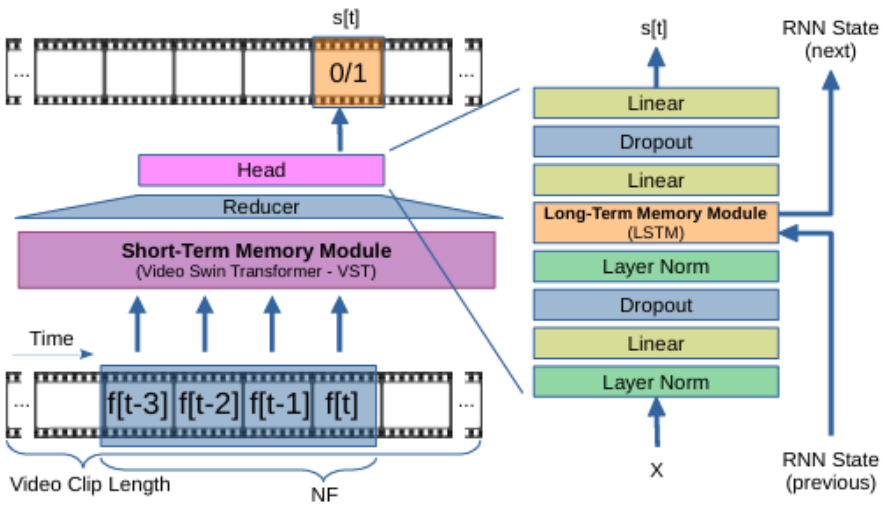

Fig. 1: The online frame-level VAD architecture. $f[t]$ is the frame at time t, x the output of the Reducer, NF the number of frames in input to the VST, $s[t]$ the anomaly classification score of the frame $f[t]$.

We will use the pretrained MOVAD model and evaluate the performance on 20 video frames and also generate a detection video. Hence we will benchmark the performance of MOVAD model on DoTA dataset.

Implementation Steps Followed :

1) Downloaded Dataset:

   Video Frames: First downloaded the DoTA dataset from link which is the split zip file of 54 GB containing only Video Frames.
   Annotations: Downloaded the annotations from link .
   meta_data: meta_data.json for validation set i.e., meta data of the dataset is downloaded from link .

   Since the dataset is very large it takes very long time almost 10-15 hours to download. Hence, I have created a mini DoTA Dataset that contains 20 video frames along with corresponding annotations and meta_data.json after writing this script : save_min_dota_dataset.ipynb

Hence, to run the model google colab file you can directory use the mini DoTA dataset [mini-dataset]

2) Run the model and evaluate on mini DoTA Dataset:

   The model file is given here : [movad_benchmarking.ipynb]

   The evaluation metrics on the 20 video frames are as follows:

```
[Correctness] f-AUC = 0.82233
              PR-AUC = 0.81332
            F1-Score = 0.69684
            F1-Mean  = 0.73005
            Accuracy = 0.73625

                precision    recall  f1-score   support

      normal       0.81      0.73      0.77      1106
     anomaly       0.65      0.75      0.70       748

    accuracy                           0.74      1854
   macro avg       0.73      0.74      0.73      1854
weighted avg       0.75      0.74      0.74      1854
```