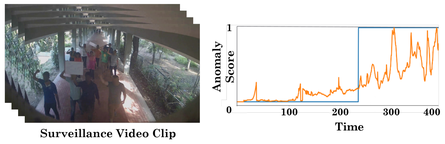
# Multi-timescale Trajectory Prediction for Abnormal Human Activity Detection

Royston Rodrigues, Neha Bhargava, Rajbabu Velmurugan and Subhasis Chaudhuri

WACV 2020

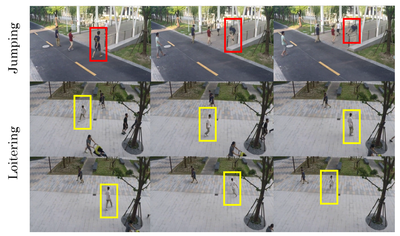
## Problem Definition and Contribution

**Goal:** To temporally localize the duration of abnormal human activities in surveillance streams.




### Motivation:

- Abnormal human activities occur at different timescales.




- A single and pre-defined timescale is not enough to capture the wide range of anomalies occurring with different time duration.

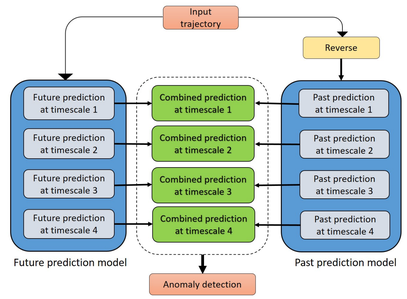
### Key Contributions:

- A bi-directional prediction framework that considers an input pose trajectory and predicts both future and past pose trajectories at different timescales.
- A large scale single-camera dataset that contains a diverse set of abnormal activities. This dataset contains a diverse range of human anomalies - single person, multiple persons or group.

## Method

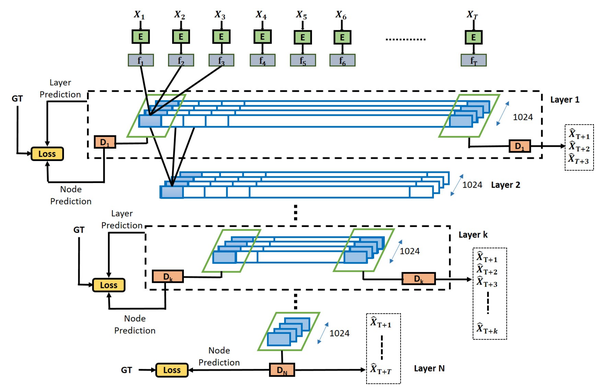
### Top-level block diagram:

- Our prediction model takes input pose trajectory sequence and generates predictions (past and future) at different timescales.




### Model Architecture:

- We make use of multi-layered 1D CNN model to predict pose trajectories at various time scales.




## Loss Function

- The total loss at the $j^{th}$ layer with $M_j$ number of nodes is given as,

$$L_j = \sum_{i=1}^{M_j} L_1^j(i) + \sum_{t=1}^{T} L_2^j(t), \tag{1}$$

- Let the $i^{th}$ node in $j^{th}$ layer make predictions for the duration $T(i) = [t_{si},\ t_{ei}]$ and generate prediction error $e(t,i)$ for a particular time instant $t \in [t_{si},\ t_{ei}]$. The $i^{th}$ node loss is computed as follows:

$$L_1^j(i) = \sum_{t=t_{si}}^{t_{ei}} e(t,i) \tag{2}$$

- For layer loss, we take the average of prediction errors generated by different nodes for a particular time instant.

$$L_2^j(t) = \frac{\sum_{i=1}^{M_j} e(t,i) 1(t \in T(i))}{\sum_{i=1}^{M_j} 1(t \in T(i))} \tag{3}$$

## Experiments & Results

### IIT-B Corridor Dataset:

- We created a large-scale single camera dataset for abnormal human activity detection.

| Camera Specific Datasets | | | |
|---|---|---|---|
| **Dataset** | **Training frames** | **Testing frames** | **Abnormal activities** |
| USCD Ped-1 | 6,800 | 7,200 | Bikers, small carts, walking across walkways |
| USCD Ped-2 | 2,550 | 2,010 | Bikers, small carts, walking across walkways |
| Subway | 20,000 | 116,524 | Climbing over fence, wrong direction |
| Avenue | 15,328 | 15,324 | Running, Throwing object, Wrong direction |
| ShanghaiTech | 2,74,515 | 42,883 | Throwing Object, Jumping, Pushing Riding a Bike, Loitering, Climbing |
| **IIT-B Corridor** | **3,01,999** | **1,81,567** | **Protest, Unattended Baggage, Cycling, Sudden Running, Fighting, Chasing, Loitering, Suspicious Object,Hiding, Playing with Ball** |
| Camera Independent Datasets | | | |
| UCF Anomaly Detection | 1610 Training videos | 290 Testing videos | Abuse, Arrest, Arson, Assault, Accident, Burglary, Explosion, Fighting, Robbery, Shooting, Stealing, Shoplifting, and Vandalism |

Samples from the IIT-B Corridor Dataset


### Ablation studies:

- We train our multi-time scale model and then test it with incremental combinations of timescales.
- We also compare performance of the model trained using only the layer loss, node loss, and the combined.

Effect of multiple timescales and past predictions

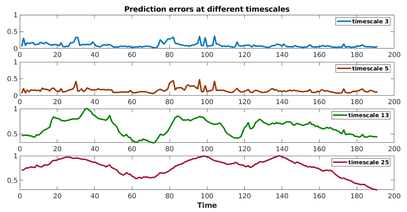
| Timescales | HR-ShanghaiTech | | | HR-Avenue | | |
|---|---|---|---|---|---|---|
| | Future | Past | Future+Past | Future | Past | Future+Past |
| 3 | 71.71 | 70.62 | 72.05 | 85.33 | 83.36 | 84.99 |
| 3, 5 | 72.89 | 71.69 | 73.39 | 86.96 | 84.70 | 86.82 |
| 3, 5, 13 | 74.51 | 73.39 | 75.65 | **88.29** | **86.20** | **88.43** |
| 3, 5, 13, 25 | **74.98** | **74.17** | **77.04** | 87.37 | 85.65 | 88.33 |

Effect of layer loss and node loss

| Loss | HR-ShanghaiTech | | | HR-Avenue | | |
|---|---|---|---|---|---|---|
| | Future | Past | Future+Past | Future | Past | Future+Past |
| Layer loss | 73.66 | 74.03 | 76.46 | 87.09 | 85.15 | 88.05 |
| Node loss | 74.26 | 74.03 | 76.67 | 87.31 | 85.19 | 88.12 |
| Layer and Node losses | **74.98** | **74.17** | **77.04** | **87.37** | **85.65** | **88.33** |

### Results:

Prediction Errors for *Loitering* example




We report *AUC* on existing single camera datasets and the proposed *IIT-B Corridor* dataset

| | HR-ShanghaiTech | ShanghaiTech | HR-Avenue | Avenue | *IIT-B Corridor* |
|---|---|---|---|---|---|
| Conv-AE (CVPR'16) | 69.80 | 70.40 | 84.80 | 70.20 | - |
| Liu *et. al* (CVPR'18) | 72.70 | 72.80 | 86.20 | **84.90** | 64.65 |
| Morais *et. al* (CVPR'19) | 75.40 | 73.40 | 86.30 | - | 64.27 |
| Ours | **77.04** | **76.03** | **88.33** | 82.85 | **67.12** |