The image features a dark blue background with several overlapping geometric shapes. On the left side, there is a large blue parallelogram and a light green parallelogram, both tilted at an angle. These shapes are layered over each other and the background. The text is positioned to the right of these shapes.

Abnormal Events Detection in CCTV Surveillance Systems

Introduction

- Increased use of cameras for surveillance
- Main task for these surveillance is to detect anomalies
 - Examples: Illegal Activities, Crimes, Fire etc.
- Anomalies occur for a short period of time compared to normal events

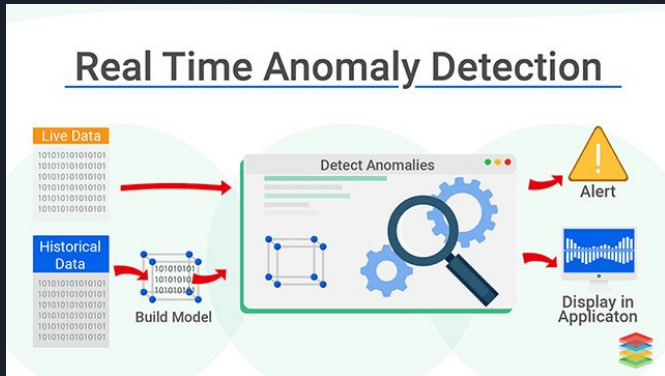


Problem Statement

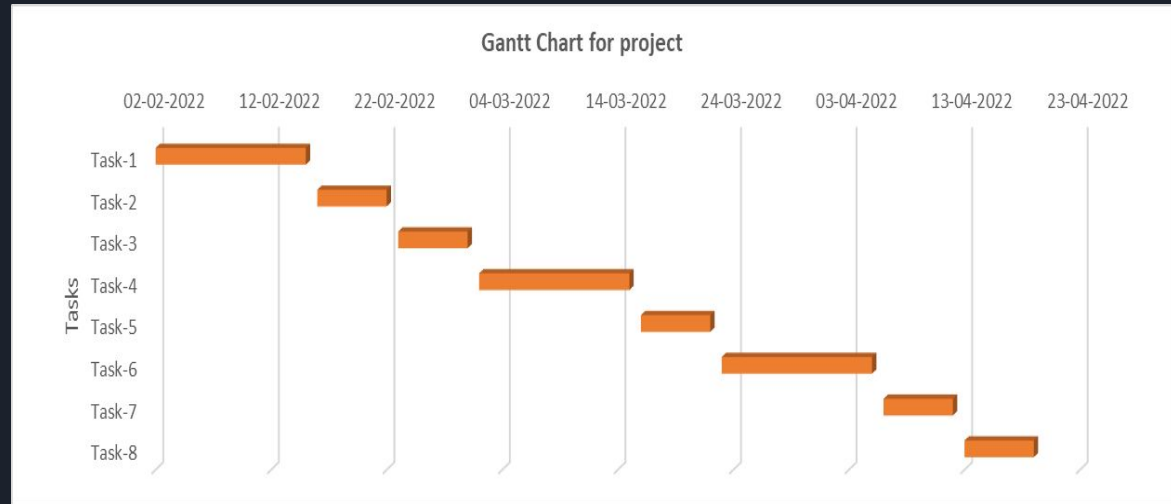
- Importance of abnormal events detection in video content analysis.

- Anomalous events v/s Normal events
 - The need for development of intelligent computer vision algorithms.

- The goal of this project is to develop a practical anomaly detection system is to timely signal an activity that deviates normal patterns and identify the time window of the occurring anomaly



Gantt Chart for project work timeline



Task-1	Task-2	Task-3	Task-4	Task-5	Task-6	Task-7	Task-8
Understanding the research paper	Started feature extraction	Completed feature extraction	performed sampling over all the videos	started feature extraction for a larger dataset	Training the model against larger dataset with hyperparameters tuned	Started research for another model which could be better than C3d	Trained the I3d model with smaller dataset to check its performance
Gaining more understanding about the dataset	Faced some errors in feature extraction	Understood how MIL and C3D works	started training and improving the results		Tested the model with this larger dataset and generated the results	Looked for more research papers which implemented this same	Tested the model and generated the results
Gaining more insights about the model						Found an I3d model which was quite better than C3d	



Existing body of work

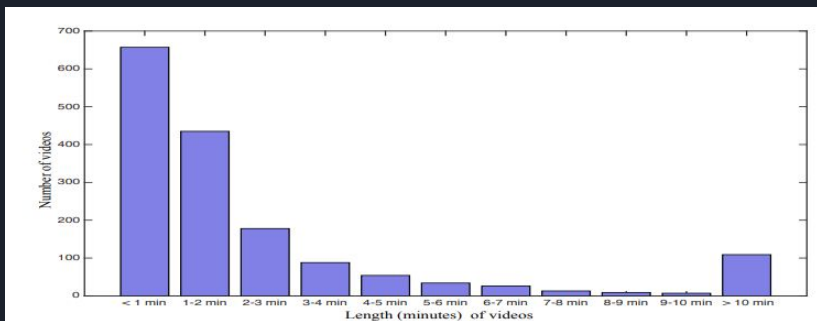
- Several different attempts in surveillance systems to detect anomalous behaviour.
- Yihao Zhanget proposes anomaly detection in traffic video with the information provided in the HEVC compressed domain.
- In High Efficient Video Coding Tian Wanga propose event detection based on moment feature descriptor and classification. The feature descriptor extracts the optical flow and computes the histogram of optical flow orientations(HOFO).

Dataset-1

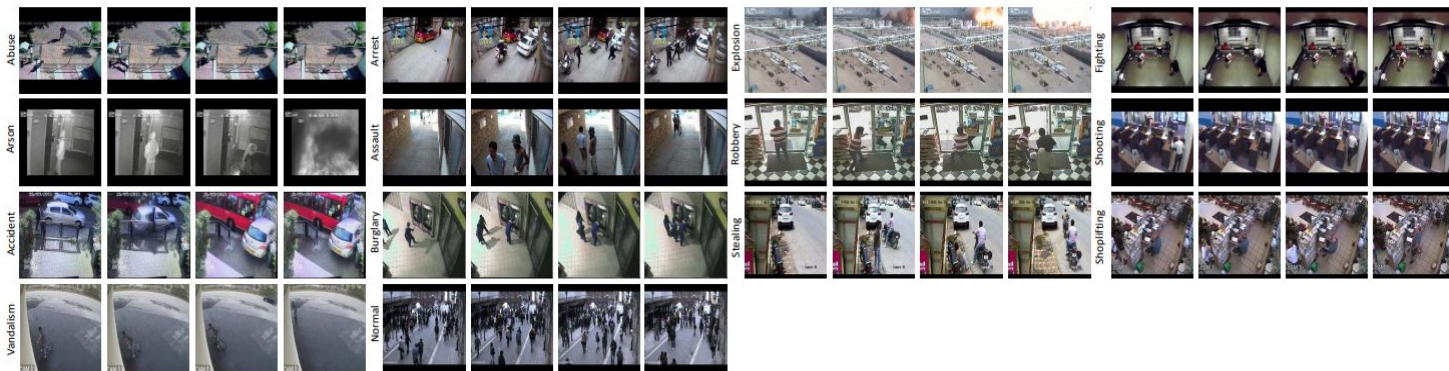
- The dataset used for this project is University of Central Florida(UCF) anomaly detection dataset.
- Consists of long untrimmed surveillance videos of Abuse, Arrest, Arson, Assault, Accident, Burglary, Explosion, Fighting, Robbery, Shooting, Stealing, Shoplifting, and Vandalism.
- A total of 1900 videos in which 950 are normal and anomaly videos.
- We have worked on two approaches for which are considering a sample of 130 videos, 65 videos from both normal and anomaly for each approach respectively. Also, 10 videos for testing part.

	# of videos	Average # of frames	Dataset length	Example anomalies
UCSD Ped1 [27]	70	201	5 min	Bikers, small carts, walking across walkways
UCSD Ped2 [27]	28	163	5 min	Bikers, small carts, walking across walkways
Subway Entrance [3]	1	121,749	1.5 hours	Wrong direction, No payment
Subwa Exit [3]	1	64,901	1.5 hours	Wrong direction, No payment
Avenue [28]	37	839	30 min	Run, throw, new object
UMN [2]	5	1290	5 min	Run
BOSS [1]	12	4052	27 min	Harass, disease, panic
Abnormal Crowd [31]	31	1408	24 min	Panic, fight, congestion, obstacle, neutral
Ours	1900	7247	128 hours	Abuse, arrest, arson, assault, accident, burglary, fighting, robbery

Dataset-2



<https://arxiv.org/abs/1801.04264>

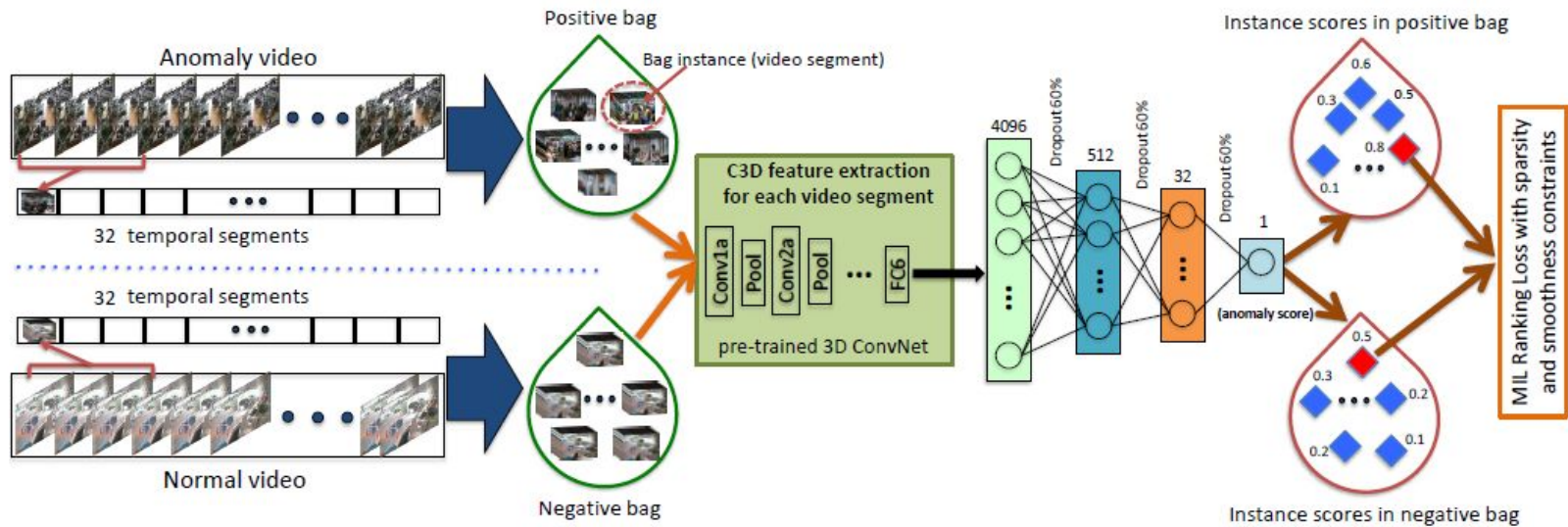


<https://arxiv.org/abs/1801.04264>

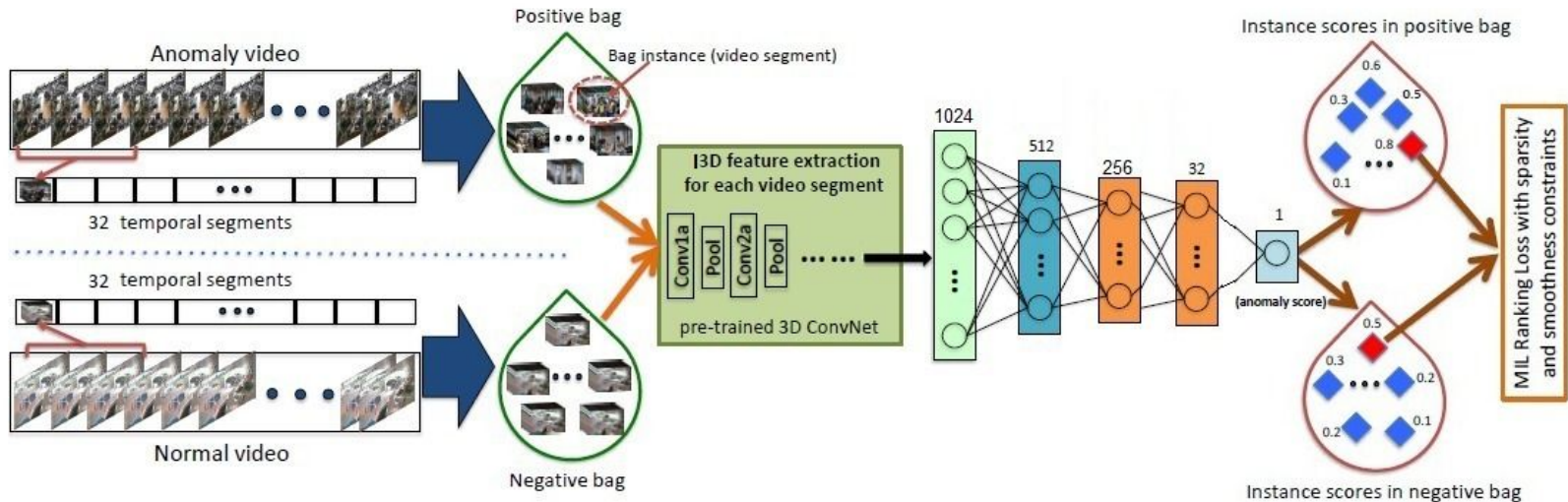
# of videos	Anomaly
50 (48)	Abuse
50 (45)	Arrest
50 (41)	Arson
50 (47)	Assault
100 (87)	Burglary
50 (29)	Explosion
50 (45)	Fighting
150 (127)	Road Accidents
150 (145)	Robbery
50 (27)	Shooting
50 (29)	Shoplifting
100 (95)	Stealing
50 (45)	Vandalism
950 (800)	Normal events

<https://arxiv.org/abs/1801.04264>

Our Approach(C3d)

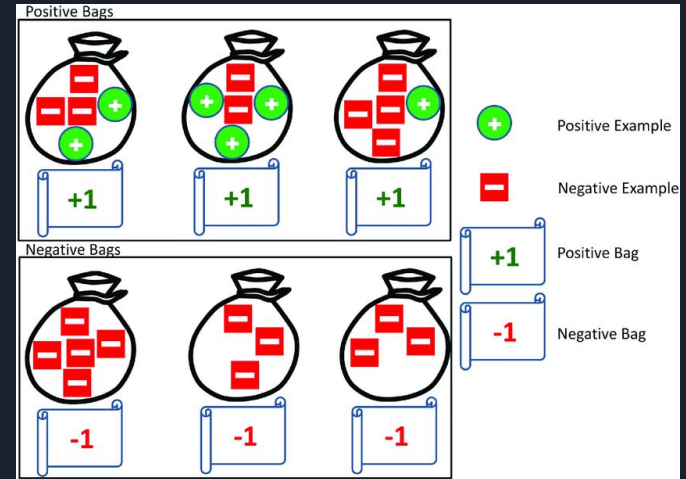


Our Approach(I3d)



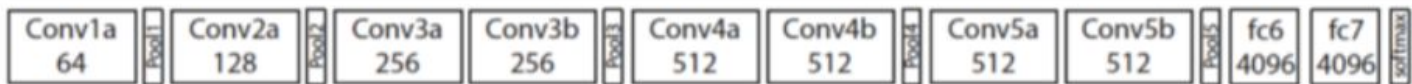
Multiple Instance learning

- We formulate anomaly detection as a regression problem in the ranking framework by utilizing normal and anomalous data.
- In MIL, precise temporal locations of anomalous events in videos are unknown. Video-level labels indicating the presence of an anomaly.
- Single video is a bag if the instance of video contains the anomaly we label it as a positive bag(anomalous video) else we consider it negative video(normal video).



https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.researchgate.net%2Ffigure%2Ffigure-illustration-of-the-concept-of-multiple-instance-learning-in-MIL-training-examples_fig1_315925709&psig=AOvVaw0cKoj8zHcr2o5xsVh5dJ6g&ust=1648551839569000&source=images&cd=vfe&ved=0CAsQjRxqFwoTCJDWvajU6PYCFQAAAAAdAAAAABAD

C3d Architecture



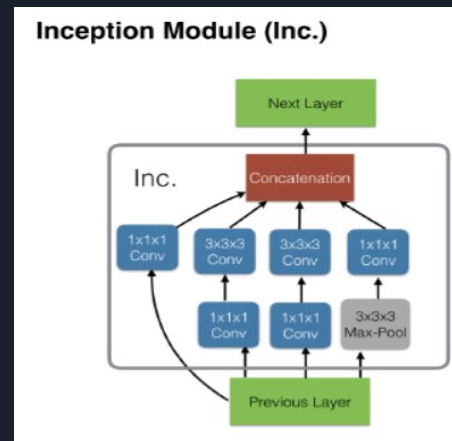
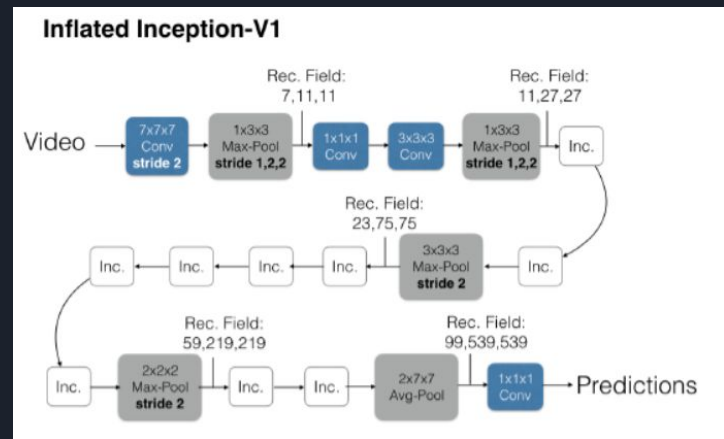
C3D Architecture [1]

<https://arxiv.org/abs/1801.04264>

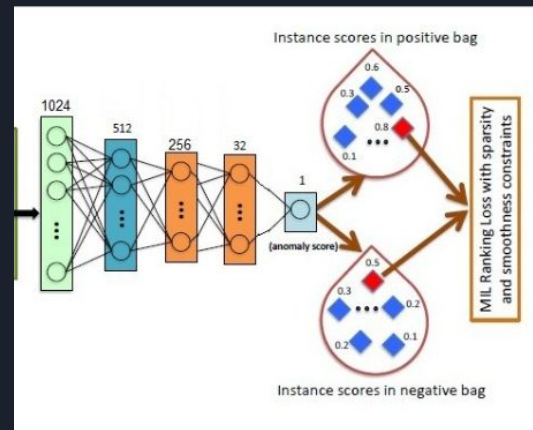
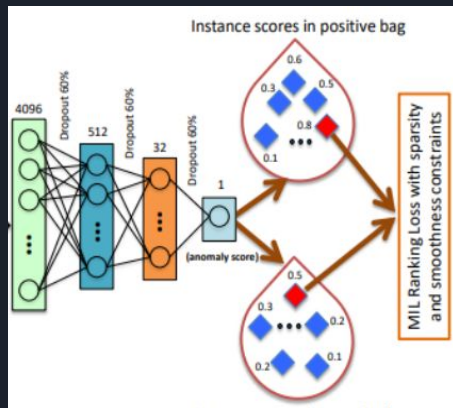
- The C3D model is given an input video segment of 16 frames (after downsampling to a fixed size which depends on dataset used) and the outputs a 4096-element vector.
- The fully connected layers have a size of 4096 dimensions which will be used in the DNN model for calculating the anomaly score
- There are total 61,214,464 params and 17 layers.
- This model would process a single clip at 16 frames per second

I3D Architecture

- The I3D model is given an input video segment of 16 frames (after downsampling to a fixed size which depends on dataset used) and the outputs a 1024-element vector.
- The global average pooling layer have a size of vector 1024 dimensions which will be used in the DNN model for calculating the anomaly score
- There are total 12,294,544 params and 195 layers.
- This model would process a single clip at 16 frames per second



DNN Model



- Fully connected layers have a size of 4096 and 1024 dimensions for each approach respectively using it as a DNN model for anomaly score
- Feature of 16 frames clip are represented in the form of (4096D and 1024D) were fed into a 3-layer feed forward neural network. This approach will use forward propagation and backward propagation using hinge loss formulation, sparsity and smoothness.



Loss Function

- The straightforward approach would be to use a ranking loss which encourages high scores for anomalous video segments as compared to normal segments, such as:

$$f(V_a) > f(V_n),$$

where V_a and V_n represent anomalous and normal video segments, $f(V_a)$ and $f(V_n)$ represent the corresponding predicted anomaly scores ranging from 0 to 1

$$l(\mathcal{B}_a, \mathcal{B}_n) = \max(0, 1 - \max_{i \in \mathcal{B}_a} f(\mathcal{V}_a^i) + \max_{i \in \mathcal{B}_n} f(\mathcal{V}_n^i)).$$

Loss Function with sparsity and smoothness

- First, in real-world scenarios, anomaly often occurs only for a short time. In this case, the scores of the instances (segments) in the anomalous bag should be sparse, indicating only a few segments may contain the anomaly.
- Second, since the video is a sequence of segments, the anomaly score should vary smoothly between video segments

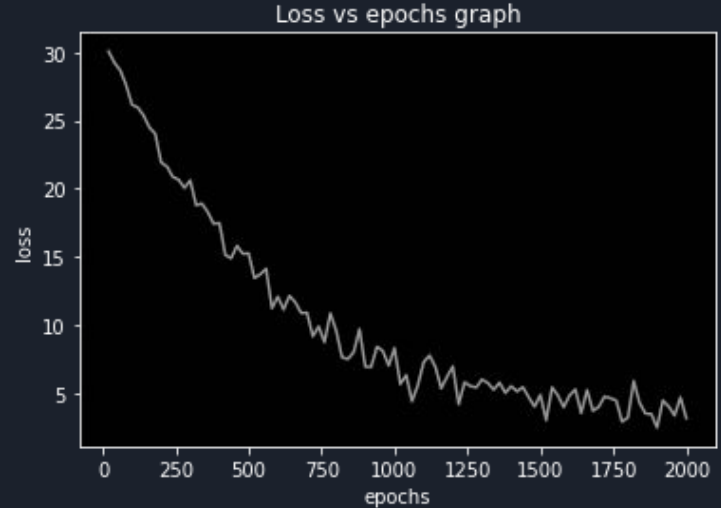
$$l(\mathcal{B}_a, \mathcal{B}_n) = \max(0, 1 - \underbrace{\max_{i \in \mathcal{B}_a} f(\mathcal{V}_a^i)}_{\textcircled{1}} + \underbrace{\max_{i \in \mathcal{B}_n} f(\mathcal{V}_n^i)}_{\textcircled{2}}) + \lambda_1 \sum_i^{(n-1)} (f(\mathcal{V}_a^i) - f(\mathcal{V}_a^{i+1}))^2 + \lambda_2 \sum_i^n f(\mathcal{V}_a^i),$$

<https://arxiv.org/abs/1801.04264>

- By incorporating the sparsity and smoothness constraints on the instance scores, the loss function becomes where 1 indicates the temporal smoothness term and 2 represents the sparsity term.

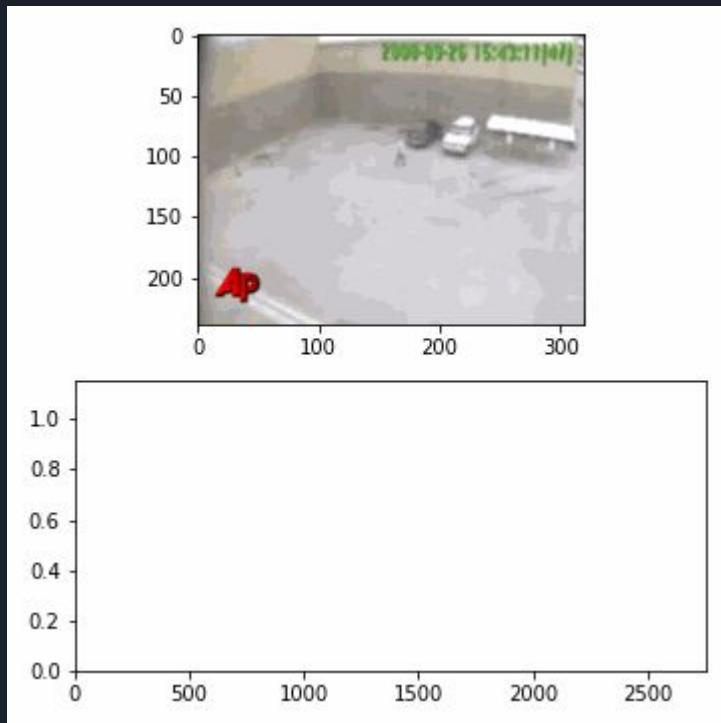
Results for 130 videos(C3d)

- We have trained our model for 2000 iterations, batch size is 60, learning rate is 0.01 and we have got the sum of hinge-loss, sparsity loss and smoothness loss which is 5.38.

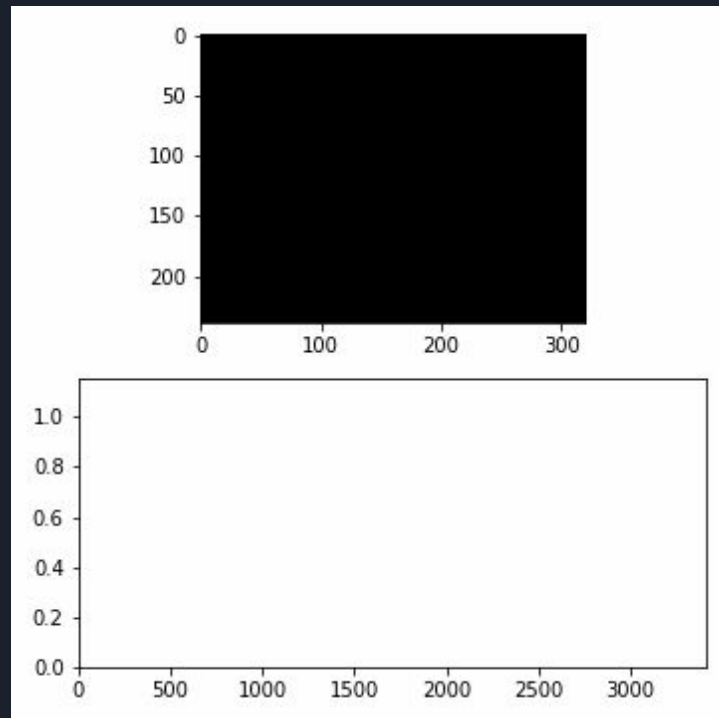


True Positive and False Negative Results

Explosion Anomaly

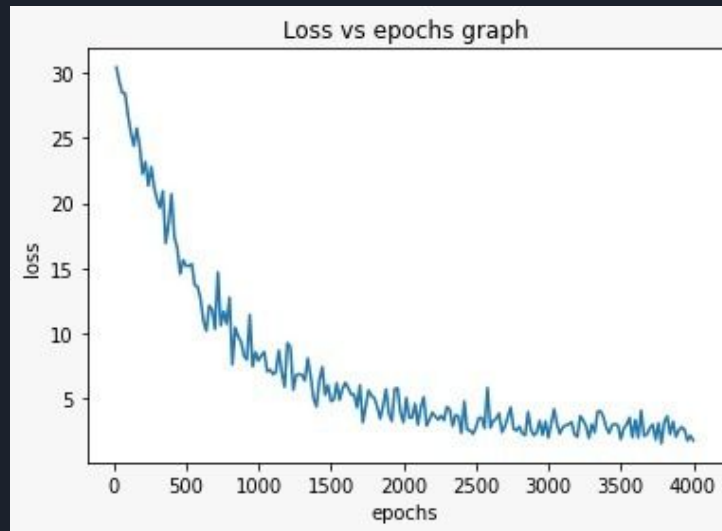


Abuse Anomaly



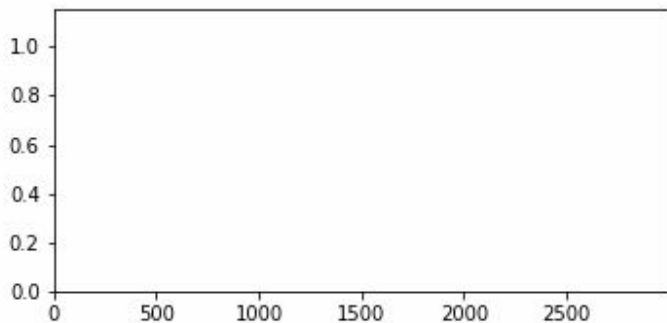
Results for C3d(260 video sample size)

- We have trained our model for **4000** iterations, batch size is **32**, learning rate is **0.01** and we have got the sum of hinge-loss, sparsity loss and smoothness loss which is **1.7413**.

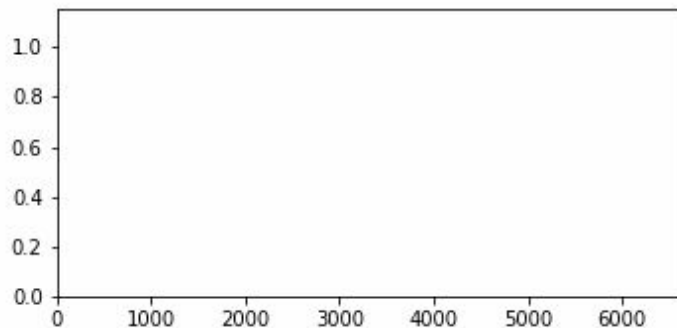
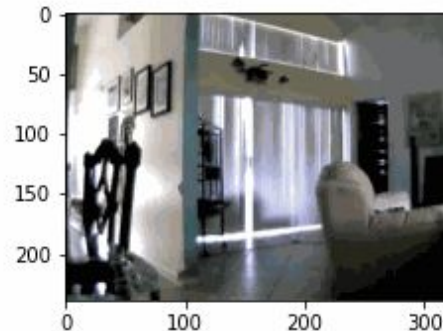


True Positive and False Negative Results

Fighting Anomaly

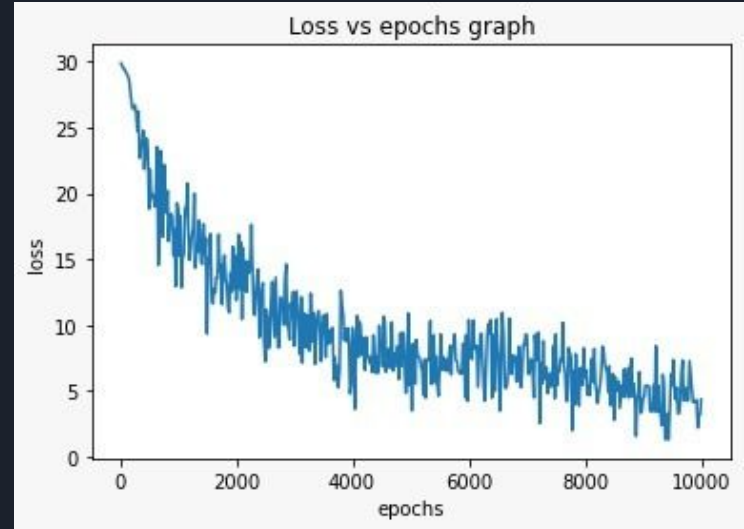


Burglary Anomaly



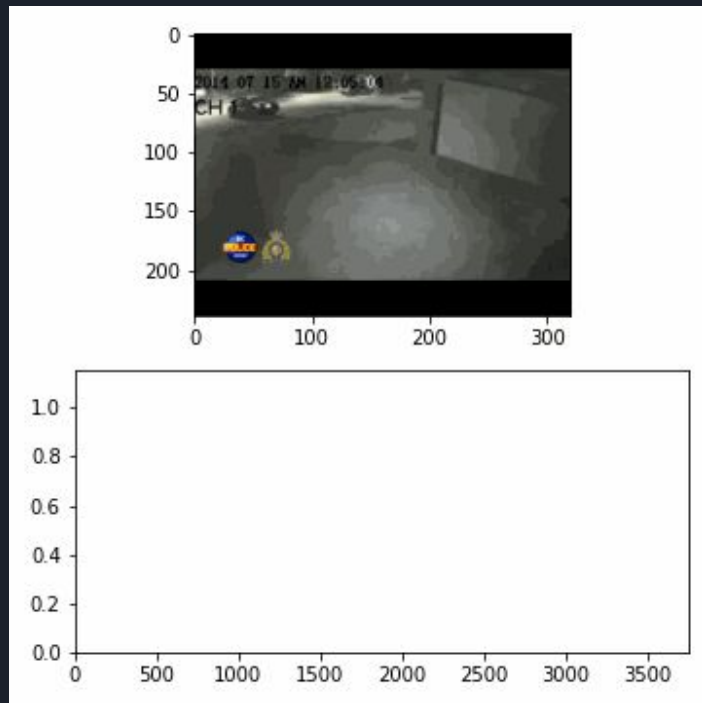
Results for I3d(130 video sample)

- We have trained our I3d model for **10000** iterations, batch size is **32**, learning rate is **0.01** and we have got the sum of hinge-loss, sparsity loss and smoothness loss which is **2.23**.

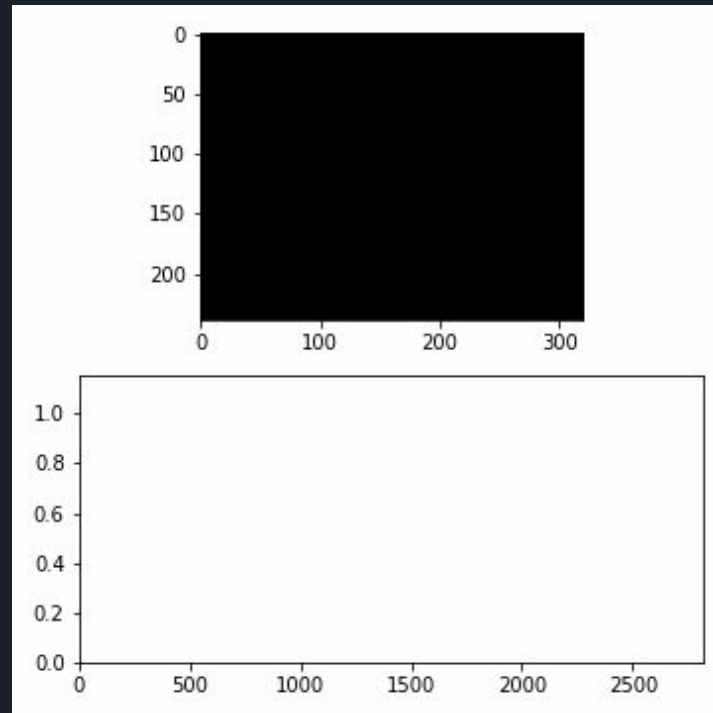


True Positive and False Negative Results

Arson Anomaly



Explosion Anomaly





Individual Roles:

Feature extraction using c3d and i3d for anomaly videos	Feature extraction using c3d and i3d for normal videos	Training on videos	Testing manually on anomalous videos	Documentation
Dinesh Nariani, Henil Shah	Devanshu Magiawala, Henil Shah	Dinesh Nariani	Devanshu Magiawala	All



Future Work

- Defining a testing mechanism for proper testing of this model.
- In the future works one can try on increasing the sample size and trying different features extractor to get best results.

References

C3d Model Feature extraction understanding:

B. M. Nair, "Deep Dive into Convolutional 3D features for action and activity recognition (C3D)," Medium, 23-Jul-2018. [Online]. Available: <https://medium.com/@nair.binum/quick-overview-of-convolutional-3d-features-for-action-and-activity-recognition-c3d-138f96d58d8f>. [Accessed: 28-Mar-2022].

I3d understanding:

J. Carreira and A. Zisserman, "Quo Vadis, action recognition? A new model and the kinetics dataset," [1705.07750v2] Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset, 20-Jul-2017. [Online]. Available: <https://arxiv-export-lb.library.cornell.edu/abs/1705.07750v2>. [Accessed: 27-Apr-2022].

S.-H. Tsang,

"<https://medium.com/nerd-for-tech/review-quo-vadis-action-recognition-a-new-model-and-the-kinetics-dataset-video-classification-a7535aa8bf48>," Medium, 29-Jun-2021. [Online]. Available: <https://medium.com/nerd-for-tech/review-quo-vadis-action-recognition-a-new-model-and-the-kinetics-dataset-video-classification-a7535aa8bf48>. [Accessed: 27-Apr-2022].

Dataset:

"Anomaly-detection-dataset," Dropbox. [Online]. Available: https://www.dropbox.com/sh/75v5ehq4cdg5g5g/AABvnJSwZI7zXb8_myBA0CLHa?dl=0. [Accessed: 28-Mar-2022].

Research Link and Literature Review:

W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," arXiv [cs.CV], pp. 6479–6488, 2018.

MIL Link:

P. Maia, "An introduction to multiple Instance Learning," NILG.AI, 18-May-2021. [Online]. Available: <https://nilg.ai/blog/202105/an-introduction-to-multiple-instance-learning/>. [Accessed: 28-Mar-2022].