NATIONAL INSTITUTE OF TECHNOLOGY - ANDHRA PRADESH

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

ANOMALY DETECTION IN LIVE SURVEILLANCE VIDEOS

Jaya Shankar, Naveen and Deva

B.Tech. (CSE) 7Th Semester

Roll No. 29, 25 and 23

Under the Guidance of

Dr. K Hima Bindu

Asst.Prof.

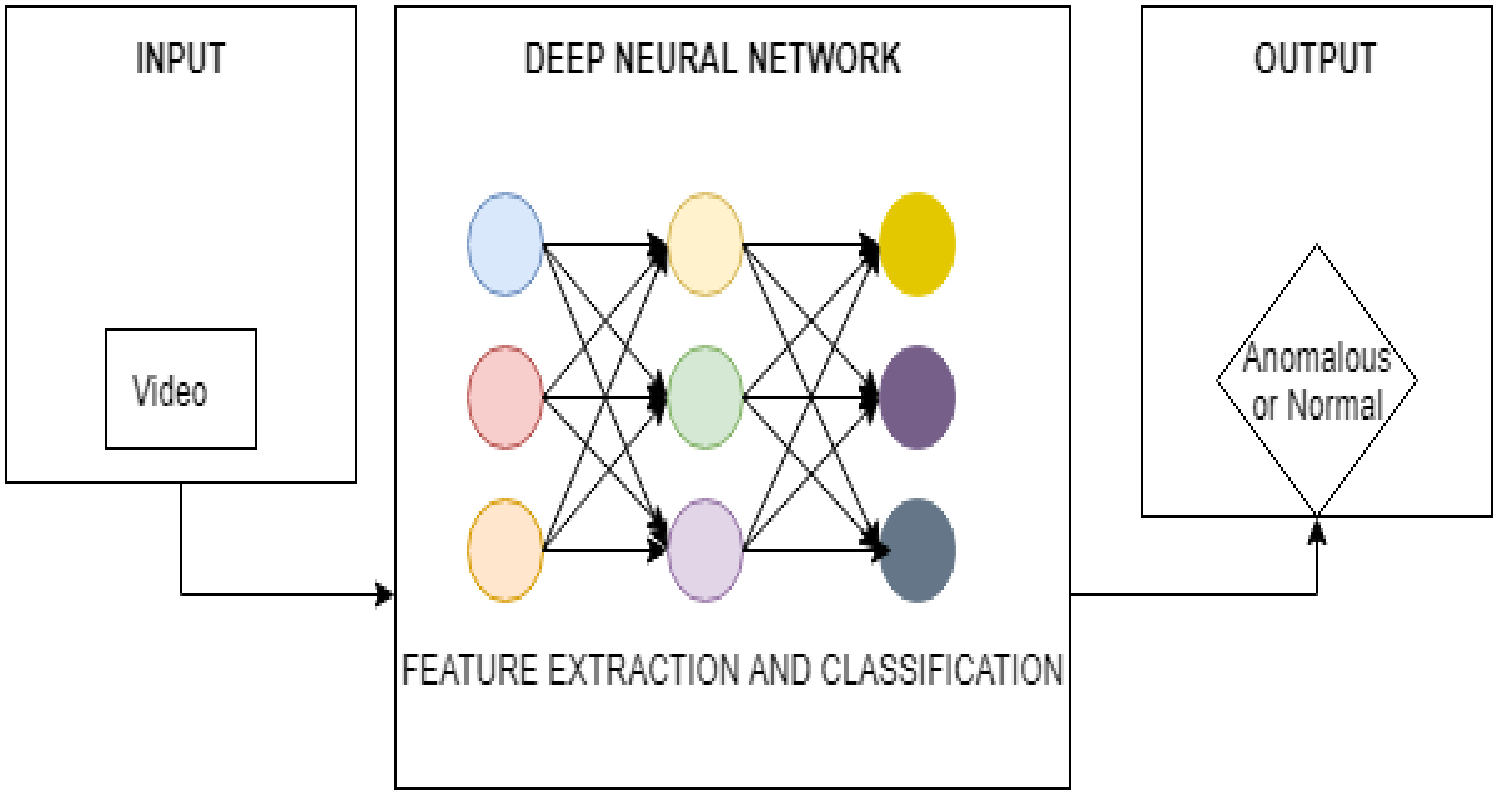
Dept. of CSE, NIT-AP

CONTENTS

PROBLEM STATEMENT

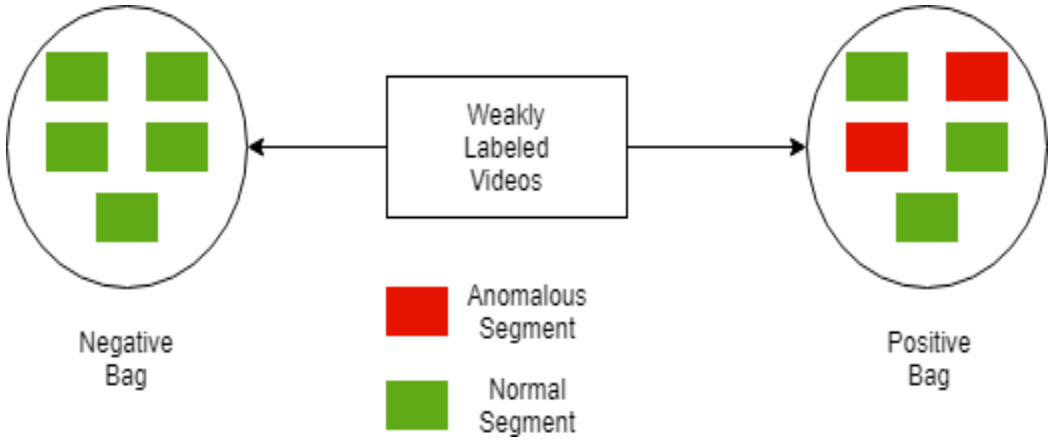
- There are many crimes occurring and there are around 250 million cameras in use today capturing them.
- Tiring process to manually detect and rise an alarm.
- Till date, specific anomaly detectors are available but not an overall detector.
- Ambiguity between normal and anomalous activity.
- Deep Learning helps in detecting such anomalous activities.

DEEP LEARNING MODEL



MULTIPLE INSTANCE LEARNING

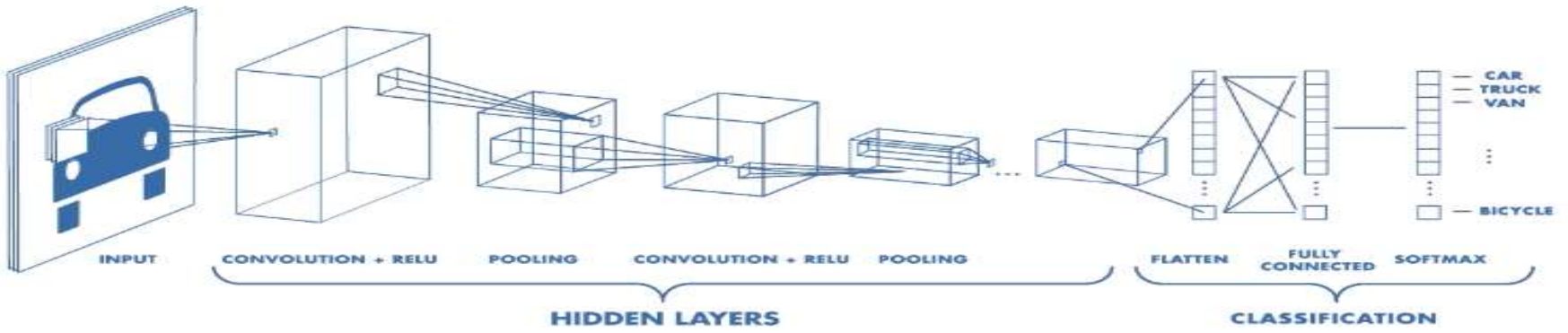
- Multiple Instance Learning



- Weakly labeled training data.
- Difference between traditional supervised learning and Multiple instance learning.

CONVOLUTIONAL NEURAL NETWORK

- Neural Network with Convolutional Layers.
- Convolutional layer has filter that does convolution operation.
- Filter is just a window of matrix that slides through the input features.



FEATURE EXTRACTION

Convolutional 3D Network

- Convolution Neural Network
- Pooling Layer

FEATURE EXTRACTION

Convolutional 3D Network

- 3D Convolution is better than 2D Convolution to model temporal information.
  - 2D CONV : performed only spatially, lose temporal information.
  - 3D CONV : performed spatio-temporally, preserve temporal information.
- Same phenomena is applicable for pooling.
- Generic, Compact, Simple and Efficient.

FEATURE EXTRACTION

Tube Convolutional Neural Network (TCNN)

- Video is divided into equal length clips
- Clips are fed to tube proposal network.
- Proposals are linked according to their scores.
- Spatio-temporal action detection is performed using these proposals.
- Tube of Interest pooling is applied to the proposal to generate fixed length feature vectors.

DEEP NEURAL NETWORK

- Input: Features.
- Output: Score.
- 3 Layer fully connected network.
- The first FC layer has 512 units followed by 32 units and at last 1 unit FC layer.
- Dropout regularization between the FC layers.
- Activation function to map score to either 0 or 1.

SCORING AND LOSS FUNCTION

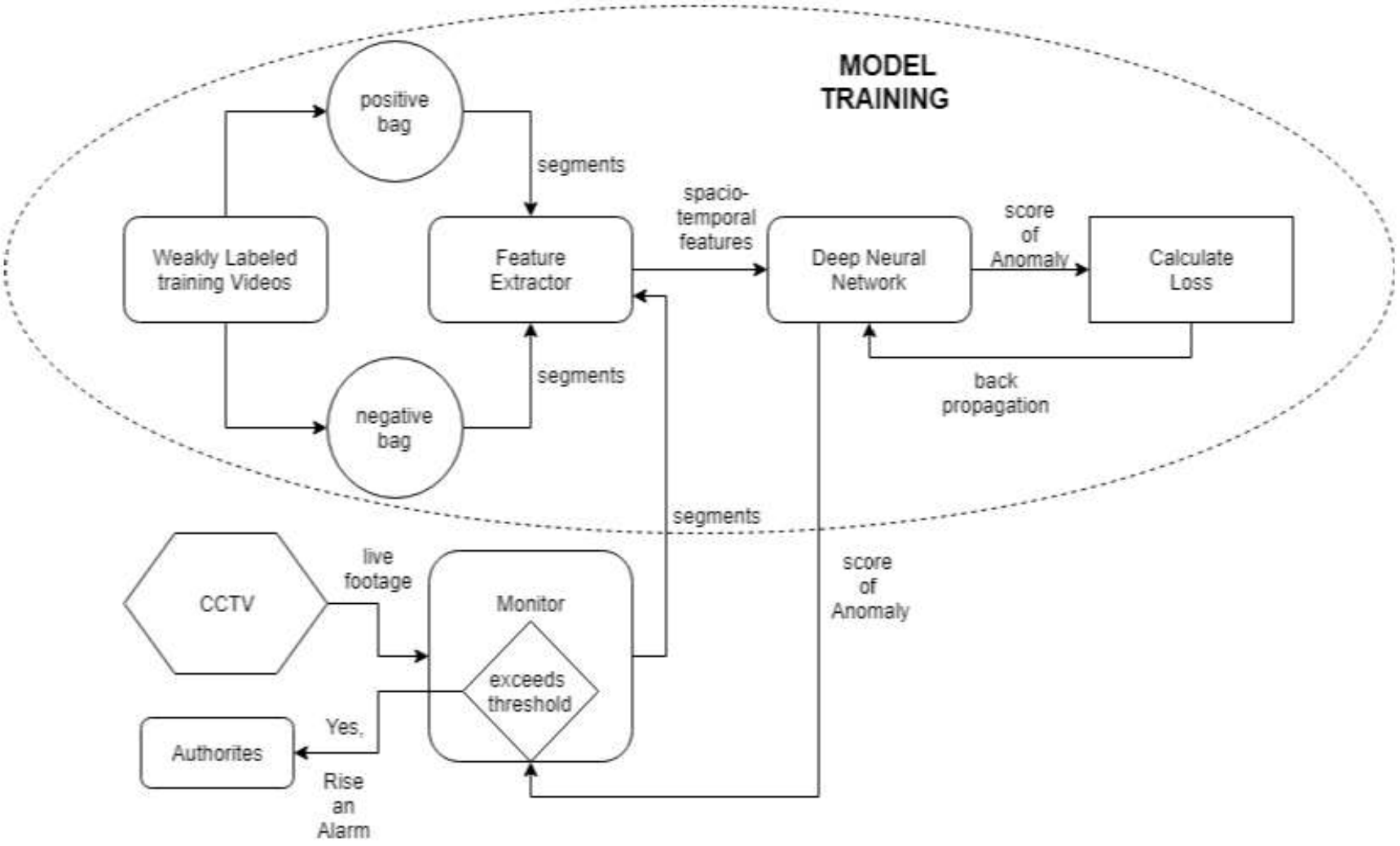
$$\min_{\mathbf{w}} \left[\frac{1}{k}\sum_{i=1}^{k} \overbrace{max(0, 1 - y_i(\mathbf{w}.\phi(x) - b))}^{①}\right] + \frac{1}{2}\|\mathbf{w}\|^2$$

$$\min_{\mathbf{w}} \left[\frac{1}{z}\sum_{j=1}^{z} max(0, 1 - Y_{\mathcal{B}_j}(\max_{i\in\mathcal{B}_j}(\mathbf{w}.\phi(x_i)) - b))\right] + \|\mathbf{w}\|^2$$

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

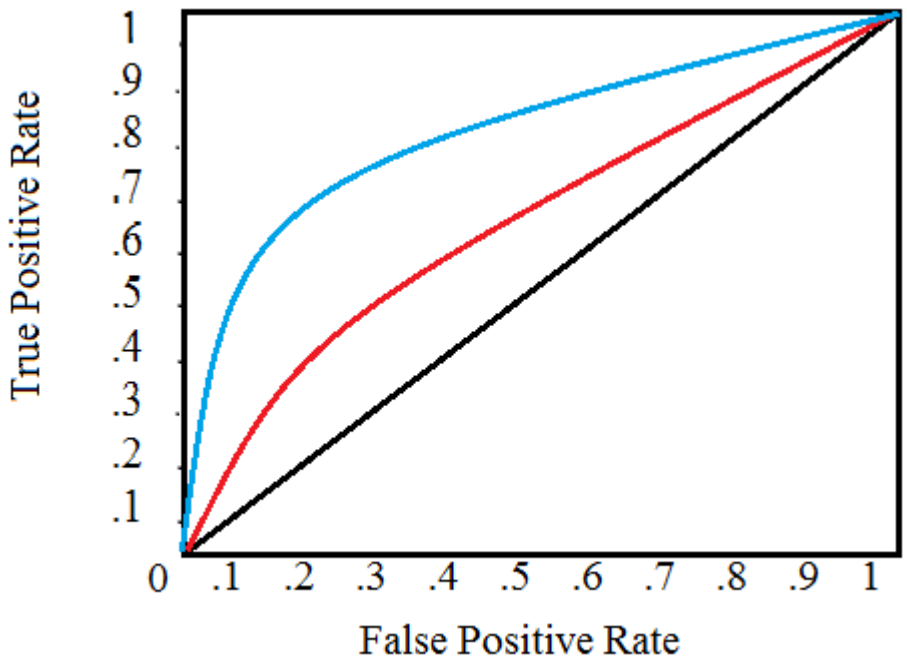
$$+\lambda_1 \overbrace{\sum_{i}^{(n-1)} (f(\mathcal{V}_a^i) - f(\mathcal{V}_a^{i+1}))^2}^{①} + \lambda_2 \overbrace{\sum_{i}^{n} f(\mathcal{V}_a^i)}^{②},$$

OVERALL MODEL



EVALUATION METRIC

- ROC and AUC to obtain the right threshold.



STATE OF THE ART METHODS

METHOD	AUC
Binary SVM Classifier	50
Dictionary Based Approach	65
Deep Neural Network w/o constraints	74
Deep Neural Network w constraints	75

• TCNN feature extraction method provides better accuracy than the C3D Feature Extraction method.

PROPOSED WORKPLAN

- Challenging Data set. (done)
- Environment Set up. (done)
- Feature Extraction.
- Training Model.
- Classifying a video into three classes
  - Criminal or violent activity
  - Potentially suspicious
  - Safe
- Testing

“Above to be completed in this semester”

PROPOSED WORKPLAN

- Challenges to be addressed:
  - Feature Extraction from a live running CCTV footage.
  - Less time delay between Anomaly occurrence and alarm.
  - Reducing the false alarms due to genuine normal activity.
  - Class Imbalance problem.
  - Extending the model to work even in night vision and in presence of external disturbances.

“Above to be done in next semester”

Implementation details

- Keras 1.1.0
- Theano 1.0.2
- Tensor Flow 1.1
- Python 3
- Ubuntu 18.10
- Nvidia Geforce GTX 1080 GPU.

REFERENCES

- 1. Sultani, Waqas and Chen, “Real-world Anomaly Detection in Surveillance Videos”, Center for Research in Computer Vision (CRCV), University of Central Florida (UCF), 2018.
- 2. Medel, Jefferson Ryan and Savakis, Andreas, “Anomaly Detection in Video Using Predictive Convolutional Long Short-Term Memory Networks”, arXiv preprint arXiv:1612.00390, 2016
- 3. http://crcv.ucf.edu/cchen/
- 4. https://dzone.com/articles/video-analysis-to-detect-suspicious-activity-based