A Robust Approach to Real World Anomaly Detection in CCTV Surveillance

Presented by Thomas Scholtz
Supervisor: Dr M. Ngxande

Problem Statement

The task of detecting anomalies spans a wide variety of situations, each with differing opinions on **what is ‘normal’ and ‘abnormal’** activity.

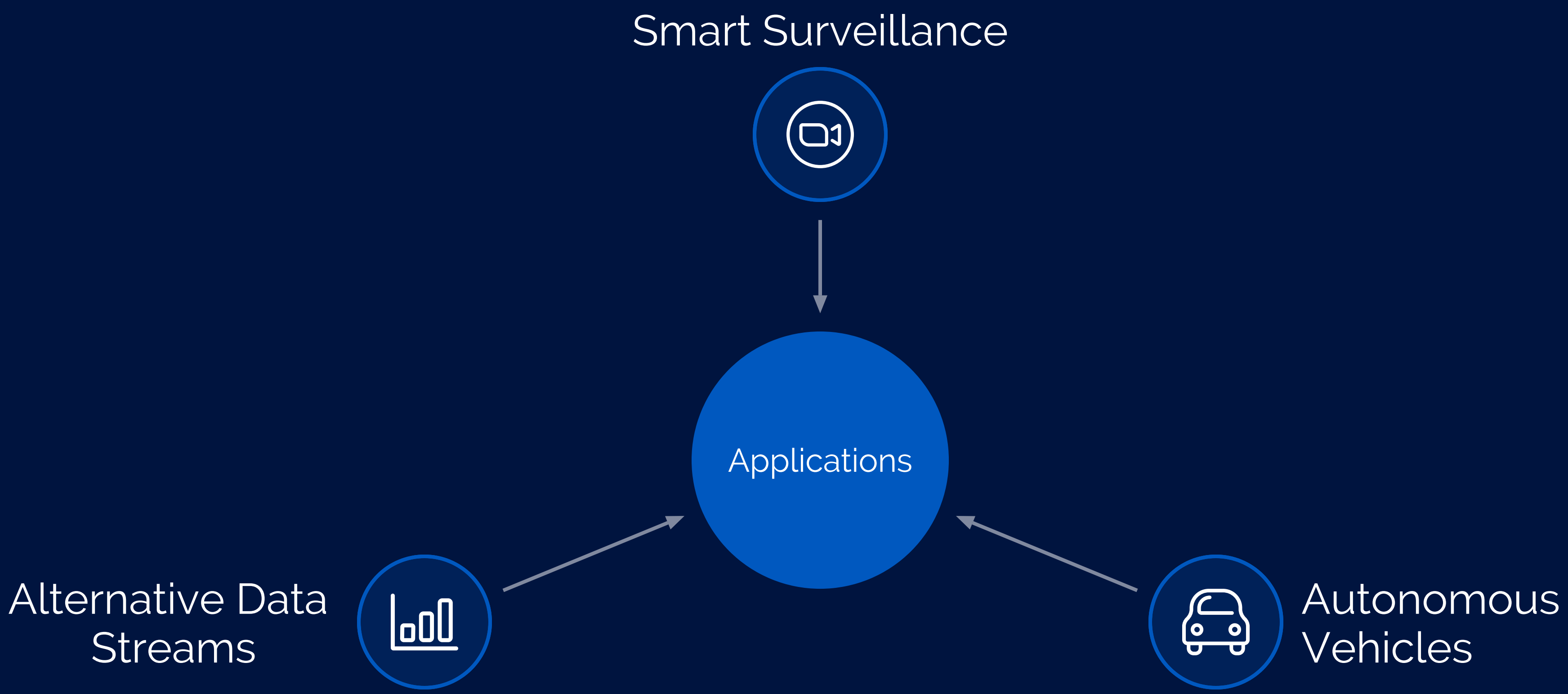
This research is concerned with investigating the **detection of anomalies in CCTV** surveillance, **irrespective of the context of the footage.**

Multiple existing implementations however, considering the vast domain of input, it is **difficult to create a robust framework.**

The **objective is to create a robust anomaly detection** framework by the following approach:

- Replicate a previously published state-of-the-art framework
- Introduce proposed extensions
- Experiment with different structural configurations, hyper-parameters and training methods
  Apply to opposing extremes of anomaly detection

Motivation


Smart Surveillance
Applications
Alternative Data Streams
Autonomous Vehicles

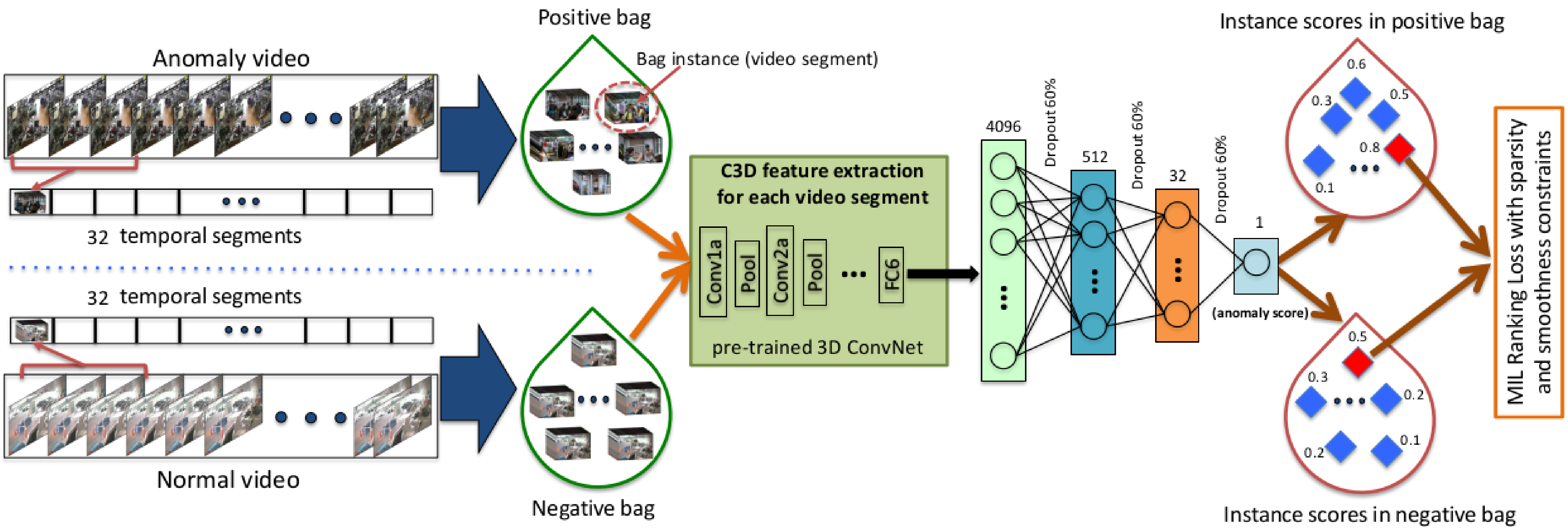
Background - A Review of Existing Approaches

- Many approaches: learn an idea of normal and recognize that (or not)
- Operates on the assumption that part of the definition of an anomaly is that it is rare, therefore not learned by a framework
- **Appearance-Motion Correspondence**
    - **common encoder, appearance decoder, motion decoder**
    - shared encoder: **appearance is associated with motion**
    - **optimized** according to the **reconstruction loss** and **optical flow loss**
- **Memory-Guided Normality**
    - relies on **robust view of normal** activity
    - record **prototypical patterns** of normal data on items **in memory**
    - **encoder, memory module, decoder**
    - optimized on **reconstruction loss** and **distance{query feature, nearest item in memory module}**
- **STT modelling on ST volumes**

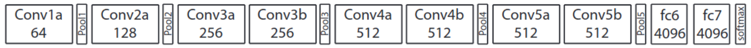
PROBLEM: dictionary of normal events doesn't adjust well to environment changes (day to night) - high false-alarm rate

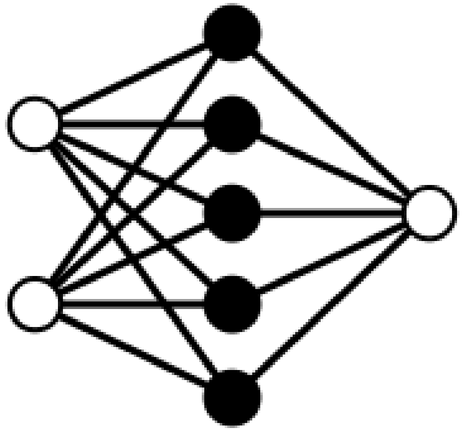
Proposed Approach

- Adopted from ***Real-world Anomaly Detection in Surveillance Videos*** by **W. Sultani [1]**
- Weakly-supervised learning approach
- **3D CNN** to **ANN**, wrapped in **MIL ranking model**
- **Regression problem** (maps **feature vector to a score between 0 and 1**)
- **Lower false alarms** compared to two state-of-the-art methods
- **Drawback:** weakly-labeled training videos



Proposed Approach (cont.)





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

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

$$\mathcal{L}(\mathcal{W}) = l(\mathcal{B}_a, \mathcal{B}_n) + \|\mathcal{W}\|_F$$

C3D Feature Extraction [2]

Spatio-temporal feature learning using pre-trained deep **3D ConvNets**

Features **extracted per 16 frames** of the input videos

Videos are divided into **32 segments, each represented by the average** of all feature vectors in the segment (4096D)

ANN

input: **4096**

hidden: **512**, **32**

output: **1** - anomaly score in [0, 1]

60% dropout between FC layers

ReLU and **Sigmoid** activation for **first** and **last** FC layers

Train with mini-batch SGD, **adagrad optimizer**

MIL

Learn **ranking model** which predicts high anomaly scores for anomalous segments

Enforce ranking **only on two instances having the highest anomaly score** in the anomalous and normal bags

Therefore learn to **deter false positives!** (something many frameworks struggle with)

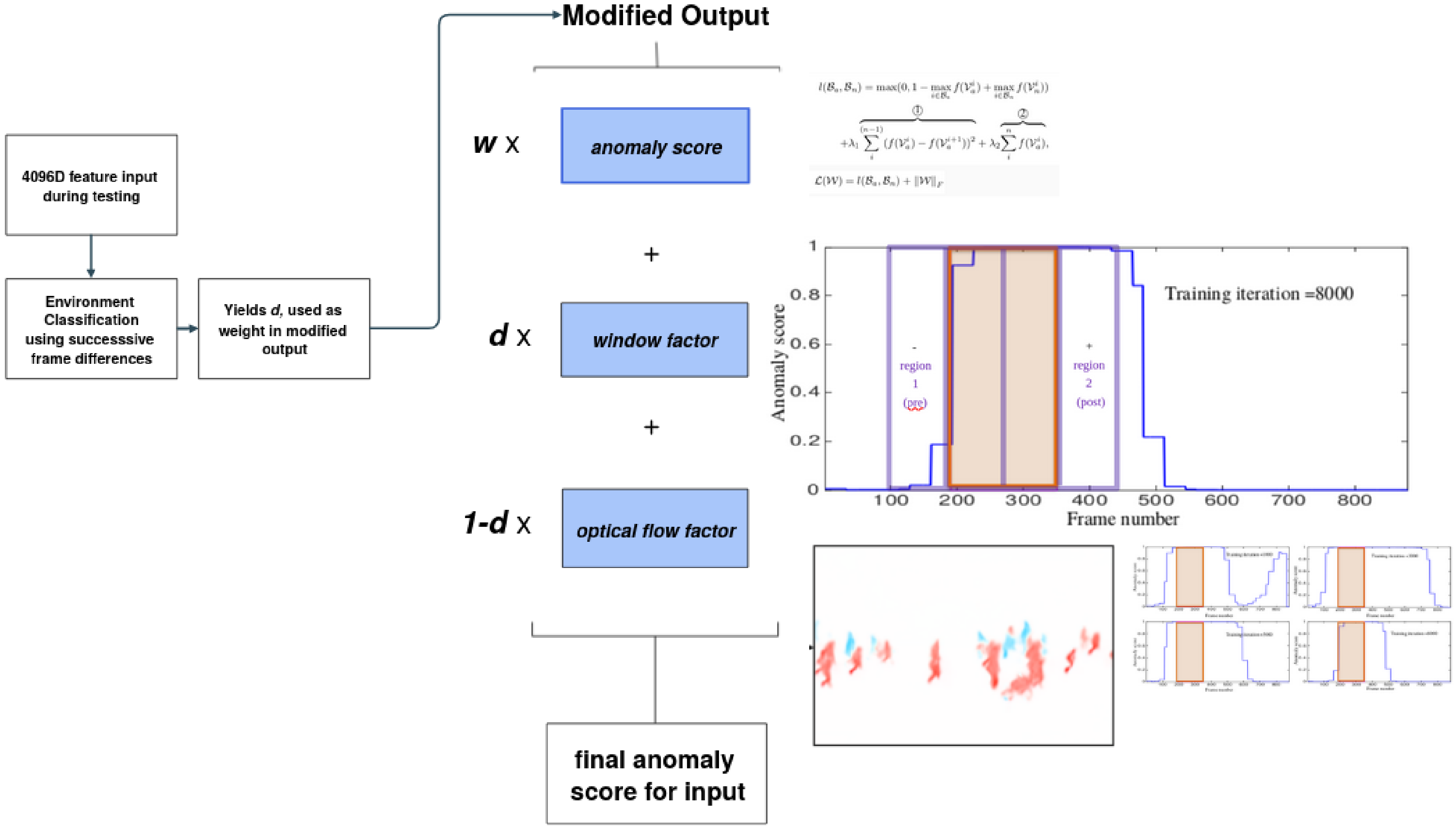
Proposed Extensions

Investigate Merit of Environment Classification

Static Environments	Dynamic Environments
Anomalies detected by subtle abnormal movement	Anomalies will invoke an unusual reaction from a dynamic environment
Employ Motion Prediction in Objective Function	Employ Peak-Window Monitoring

Proposed Extensions (cont.)

- Additional abstraction layer on anomaly score output

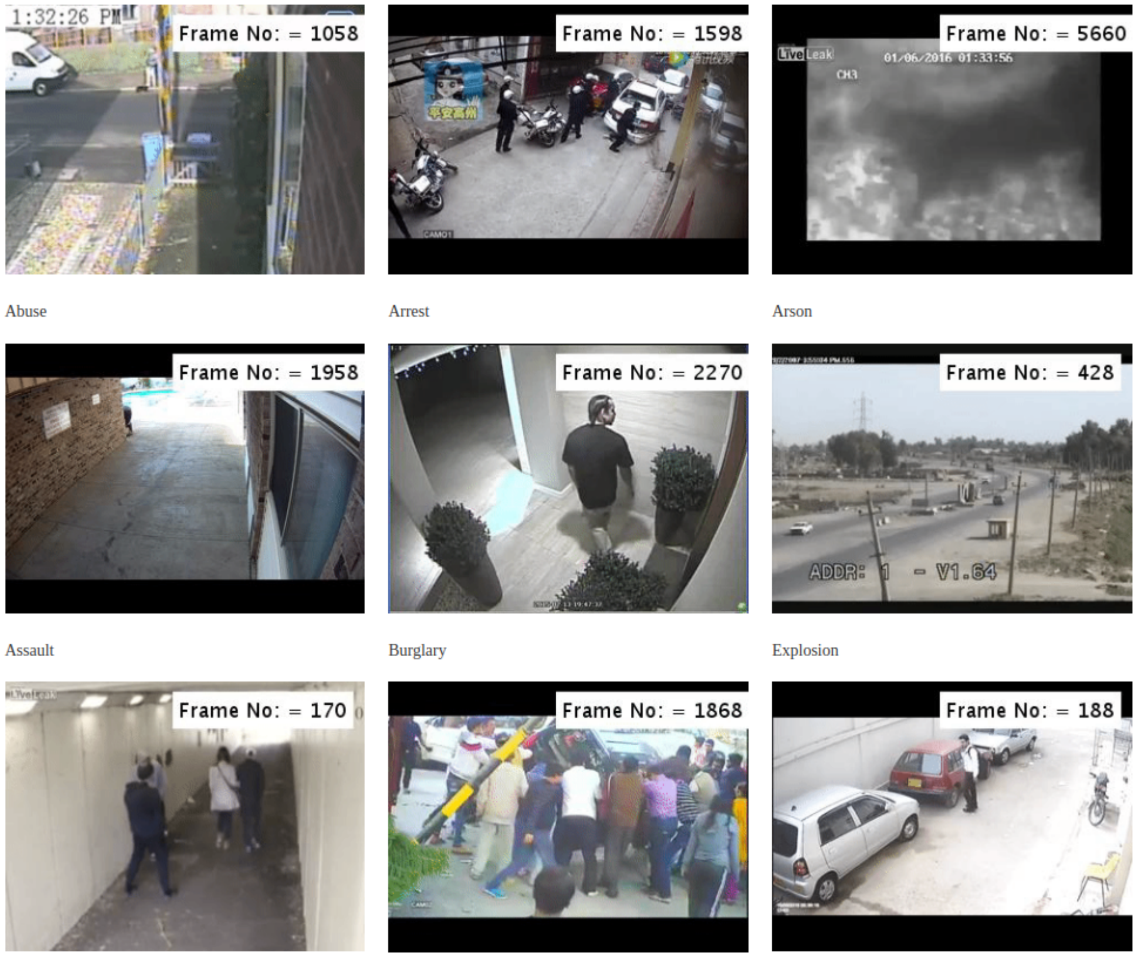


Investigate training method for ANN + MIL with DE

Data set

UCF-Crime data set [3]

- 950/950 - normal/anomalous, approximately 200 GB of videos
- long untrimmed videos - replicates real life scenarios
- 13 Anomalies considered namely, **Abuse, Arrest, Arson, Assault, Accident, Burglary, Explosion, Fighting, Robbery, Shooting, Stealing, Shoplifting, and Vandalism**
- Significant relevance in terms of public safety
- State-of-the-art methods show poor results - worthy challenge

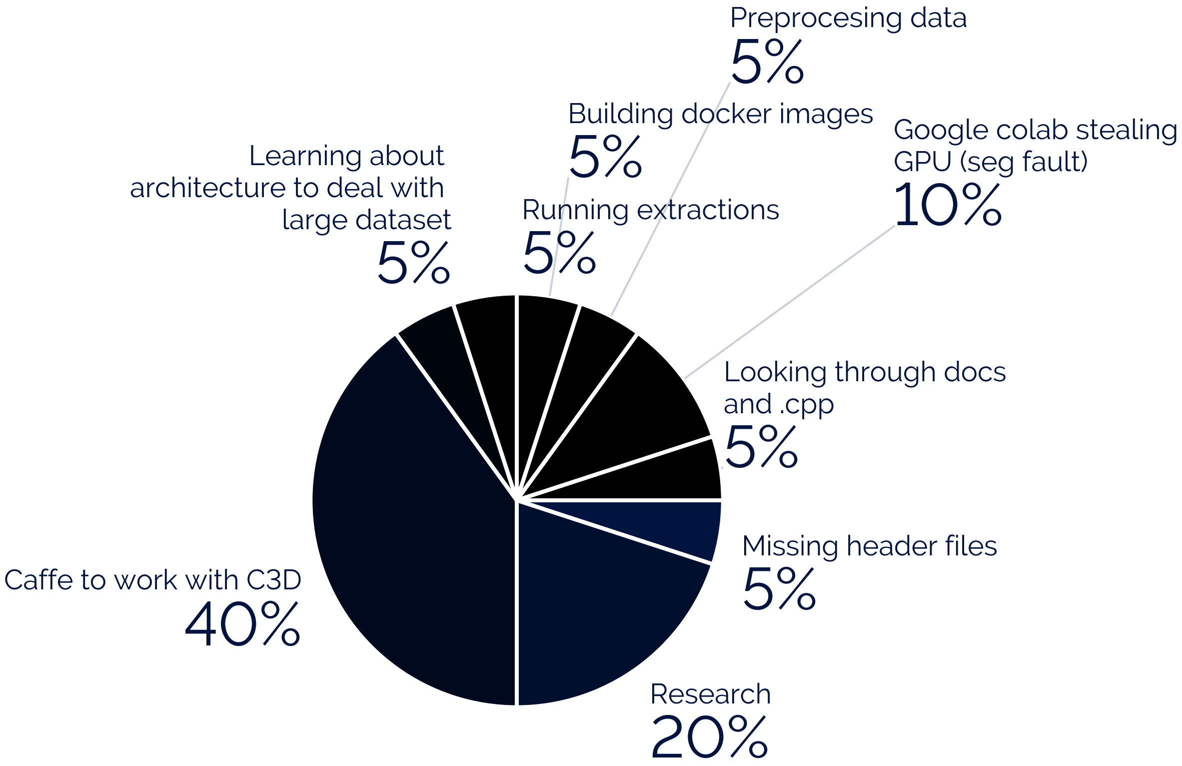


Progress

- **Convolutional Aspect Completed: Image to Vector**
  - C3D Features extracted per 16 frames of all 200GB of UCF-Crime data set
  - All videos preprocessed
    - segmentation
    - format conversion
    - average vectors computed
  - Data set represented by 32 x 4096D vectors per video

- ANN + MIL implementation underway in Keras + TensorFlow

- Visualization tool in Streamlit



current point

Complete ANN + MIL implementation

Develop environment classification method

Develop peak-window monitoring

Develop optical flow comparator

Repeat experiments for extended framework as well as baselines

DEMO 2

Investigate training method

Apply all frameworks to extremes of data set

Timeline of Future Work

References

[1] Waqas Sultani, Chen Chen, Mubarak Shah, **Real-World Anomaly Detection in Surveillance Videos,** Cornell University Library, [v1] Fri, 12 Jan 2018.

[2] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, **Learning Spatio-temporal features with 3D Convolutional Networks**, ICCV 2015.

[3] UCF-Crime Data set, https://www.crcv.ucf.edu/projects/real-world/

Questions?