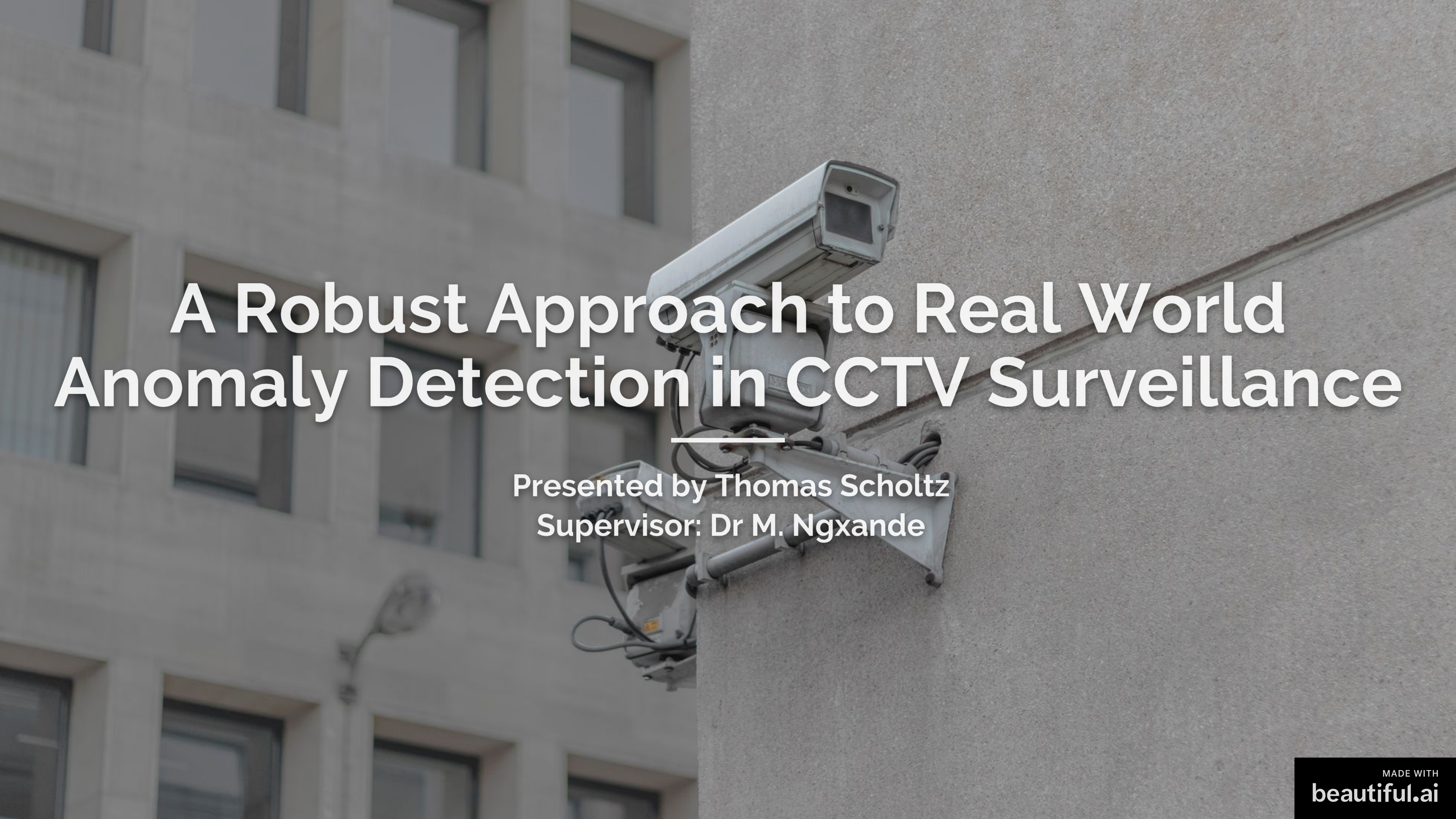


A Robust Approach to Real World Anomaly Detection in CCTV Surveillance

Presented by Thomas Scholtz
Supervisor: Dr M. Ngxande

Outline

Anomaly Detection Framework

- Combination of **deep learning techniques** and **heuristics**
- Emphasis on **generalization**
- Operates on **consensus**

Outline

Real-world Anomaly Detection in Surveillance Videos

Waqas Sultani[1], Chen Chen[2], Mubarak Shah[2]
[1]Department of Computer Science, Information Technology University, Pakistan
[2]Center for Research in Computer Vision (CRCV), University of Central Florida (UCF)
waqas5163@gmail.com, chenchen870713@gmail.com, shah@crcv.ucf.edu

Abstract

Surveillance videos are able to capture a variety of realistic anomalies. In this paper, we propose to learn anomalies by exploiting both normal and anomalous videos. To avoid annotating the anomalous segments or clips in training videos, which is very time consuming, we propose to learn anomaly through the deep multiple instance ranking framework by leveraging weakly labeled training videos, i.e. the training labels (anomalous or normal) are at video-level instead of clip-level. In our approach, we consider normal and anomalous videos as bags and video segments as instances in multiple instance learning (MIL), and automatically learn a deep anomaly ranking model that predicts high anomaly scores for anomalous video segments. Fur-ermore, we introduce sparsity and temporal smoothness -traints in the ranking loss function to better localize -ly during training.

1. Introduction

Surveillance cameras are increasingly being used in public places *e.g.* streets, intersections, banks, shopping malls, etc. to increase public safety. However, the monitoring capability of law enforcement agencies has not kept pace. The result is that there is a glaring deficiency in the utilization of surveillance cameras and an unworkable ratio of cameras to human monitors. One critical task in video surveillance is detecting anomalous events such as traffic accidents, crimes or illegal activities. Generally, anomalous events rarely occur as compared to normal activities. Therefore, to alleviate the waste of labor and time, developing intelligent computer vision algorithms for automatic video anomaly detection is a pressing need. The goal of a practical anomaly detection system is to timely signal an activity that deviates normal patterns and identify the time window of the occurring anomaly. Therefore, anomaly detection can be consid-ered as coarse level video understanding, which filters anomalies from normal patterns. Once an anomaly -tected, it can further be categorized into one of the

Base Model (1/3)

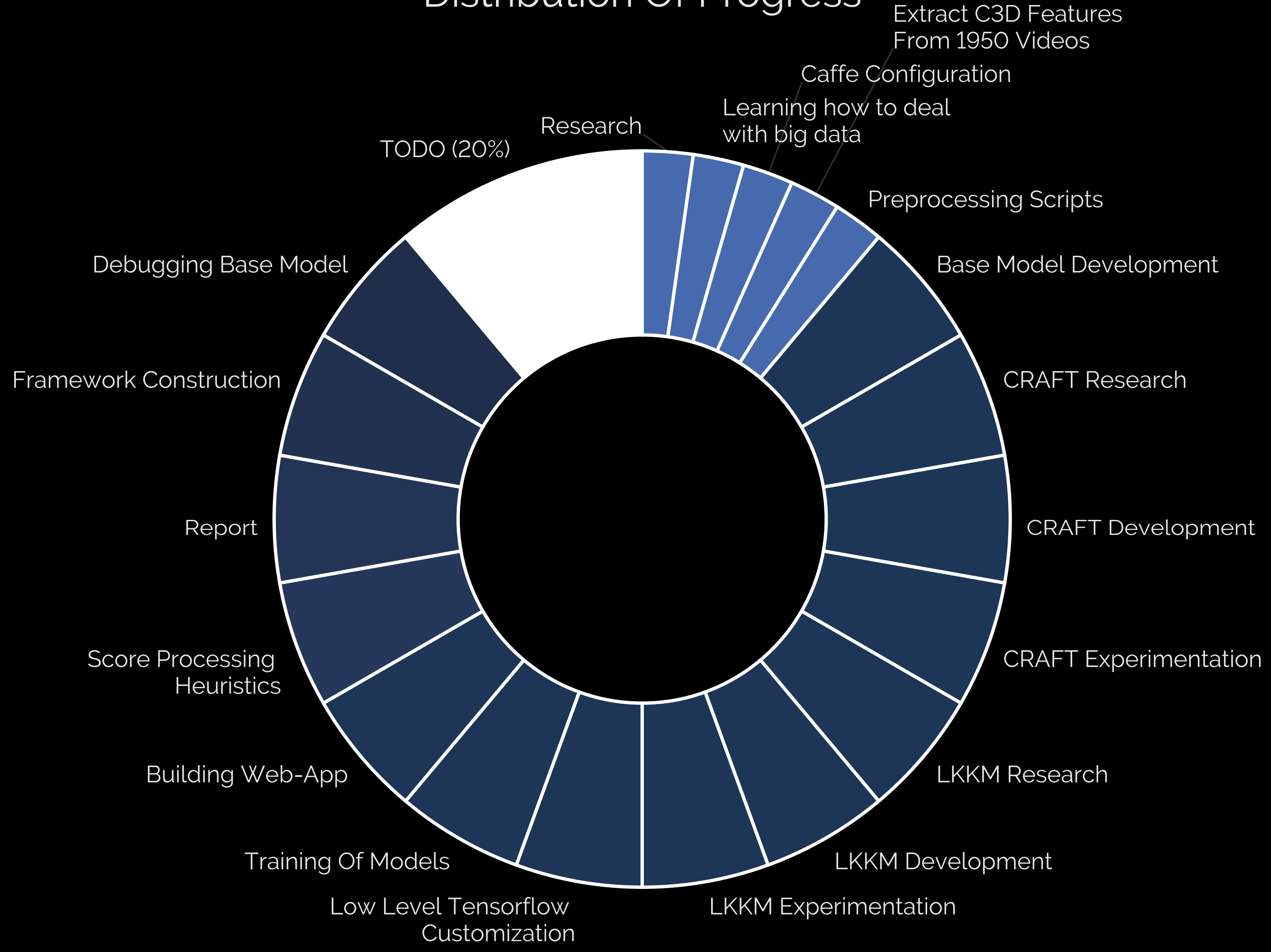
- **Higher-level, generalizable approach**
- **Focus on deterrence of false positives**
- Multiple **improvements** implemented

Outline

Novel Approaches

- **CRAFT** (2/3)
- **LKKM** (3/3)
- **Unsupervised**
- Makes use of recent research in emerging areas of Computer Vision and ML
- Results show **promising potential**

Distribution Of Progress



Problem Statement

Quantify the degree of abnormality

- assign anomaly scores along temporal axis
- minimize false positives
- problem statement is inherently vague

Problem Statement

An occasional bicycle on a sidewalk of pedestrians

Problem Statement

A rare accident on a highway

Problem Statement

Is one more anomalous than the other?
Can anomaly be defined in the general case?

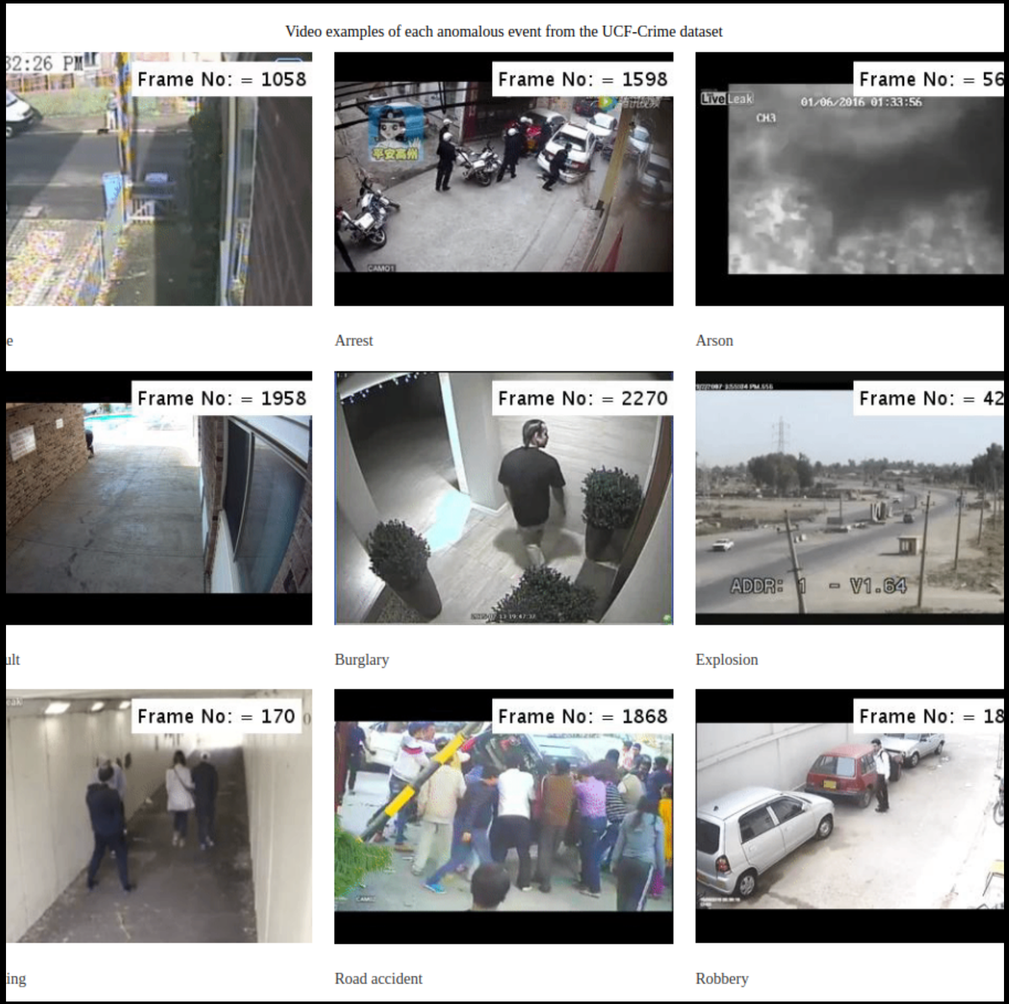
- Appearance
- Motion
- Context (difficult part)

Problem Statement

Focal Area

- The base model is **weakly supervised**
- Focus on **unwanted activity**
- Problem domain reduced but still under-specified (intentional)

Problem Statement



UCF-Crime Dataset [2]

- Abuse
- Arrest
- Arson
- Assault
- Burglary
- Explosion
- Fighting
- Road Accidents
- Robbery
- Shooting
- Shoplifting
- Stealing
- Vandalism

Background - A Review of Existing Approaches

- **Common approach:**
  - learn an idea of normal activity
  - does unseen footage conform to this idea?
- **Assumption:**
  - anomaly = rare event
  - therefore not learned by a framework
- **Examples:**
  - Appearance-Motion Correspondence
  - Memory-Guided Normality

Recurring Shortfalls Of Existing Solutions

Frameworks struggle to cover the spectrum of anomalies

Difficulty in dealing with false positives

Proposed Framework

Learn higher-level context.
Deter false positives.
Draw attention to suspicious footage.

BASE MODEL

CRAFT

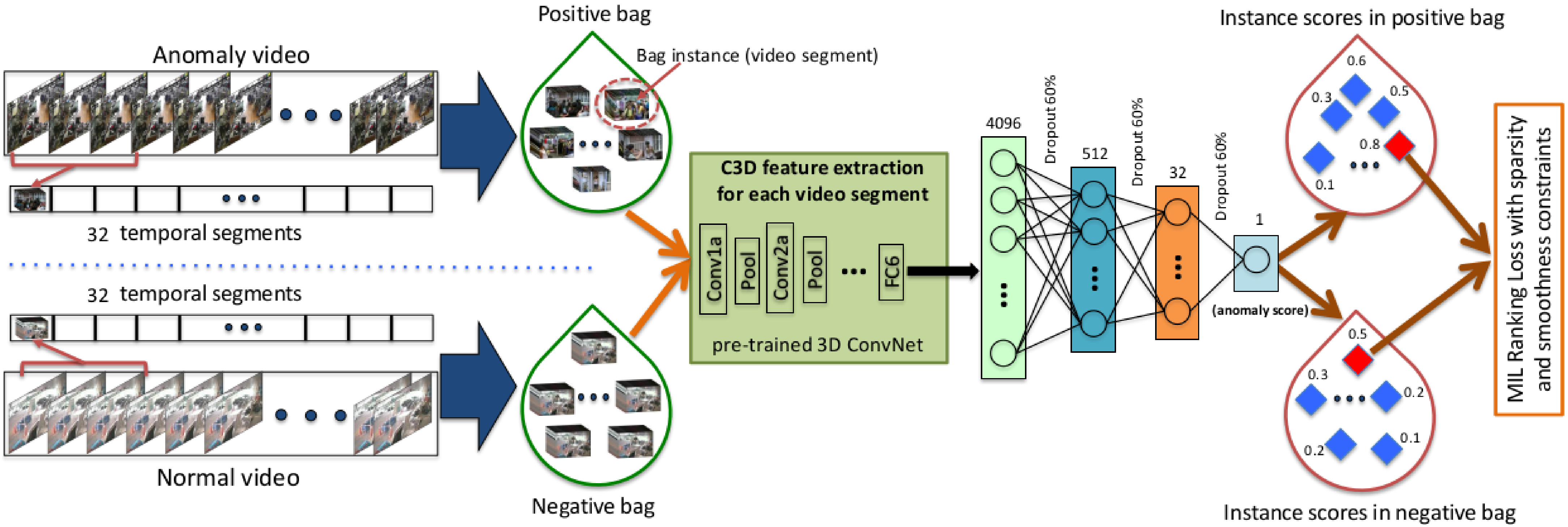
LKKM

Fine grained. Identify unpredictable changes in appearance/motion via frame construction with RAFT

Fine-grained.
Detect deviations from learned patterns.

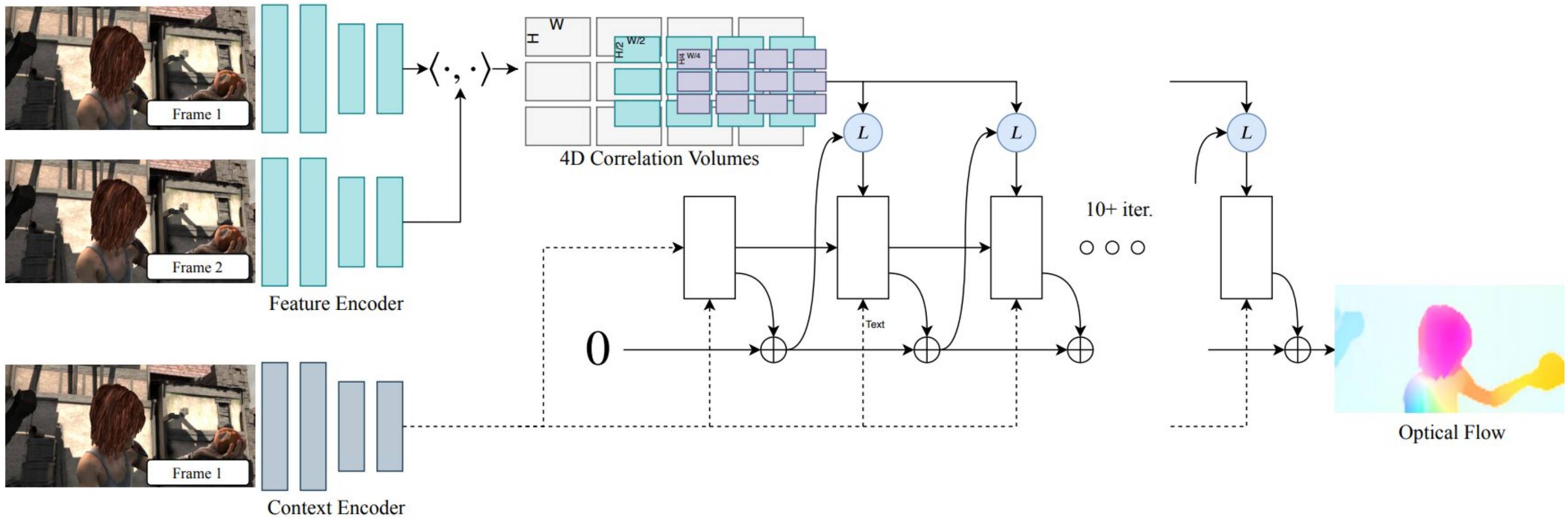
Implementation: Base model

- 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 alarm rate** compared to two state-of-the-art methods



Implementation: CRAFT

- Optical Flow Estimation: **displacement of pixels** from one frame to the next.
- CRAFT uses RAFT [3] to **CONSTRUCT future frames** and compare them to their ground truth.
- f_i is optical flow between $\mathbf{x_i}$ and $\mathbf{x_{i+1}}$
- $\mathbf{x_i + f_i = \hat{x}_{i+1}}$
- reconstruction error = $\| \mathbf{\hat{x}_{i+1} - x_{i+1}} \|$
- Exploit RAFT's **inaccurate predictions,** resulting from anomalous activity, to obtain **inaccurate constructions**

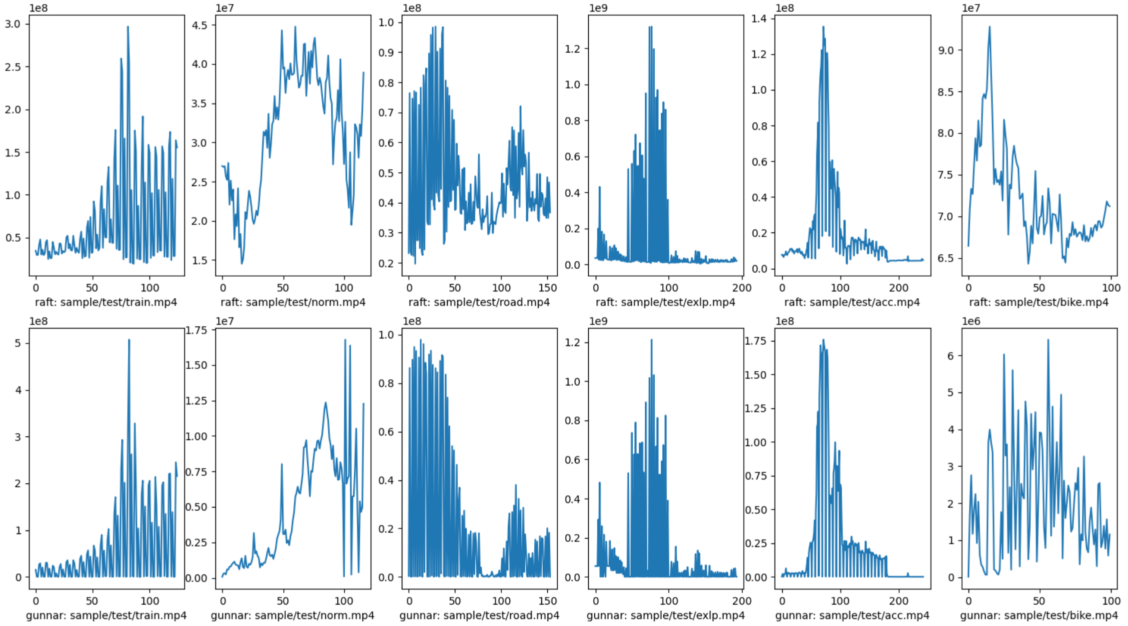


CRAFT

Below are reconstruction errors obtained from clips similar to the one on the right.

Top row: RAFT optical flow estimation.

Bottom row: Gunnar Farneback method for optical flow, which was used for comparison.



Video

Implementation: LKKM

- **Lukas-Kanade** [4]: classic method for sparse optical flow
- **K-Means** [5]**:** data clustering algorithm
- The two concepts are combined to form a heuristic

- **Quantify deviation from repetitive patterns**
- Extracts key features from each frame
- Compute sparse optical flow between frames' features
- Use K-Means to from cumulative clusters of flow vectors
- A score is proportional to the top 50 maximum distances of vectors to nearest cluster

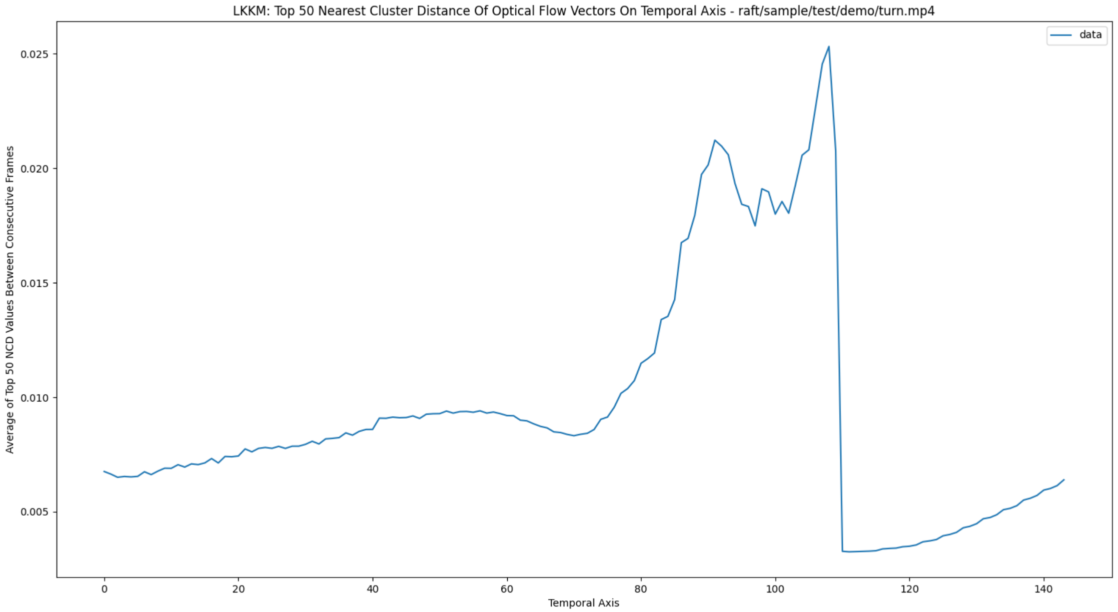
- Quality of scores proportional to video progress

  Time-series decomposition to remove trend

LKKM

The score profile below corresponds to the top fifty maximum nearest cluster distances at each frame.

Leading up to the anomaly, all traffic follows the same pattern of movement, therefore forming strong clusters of optical flow vectors.



Video

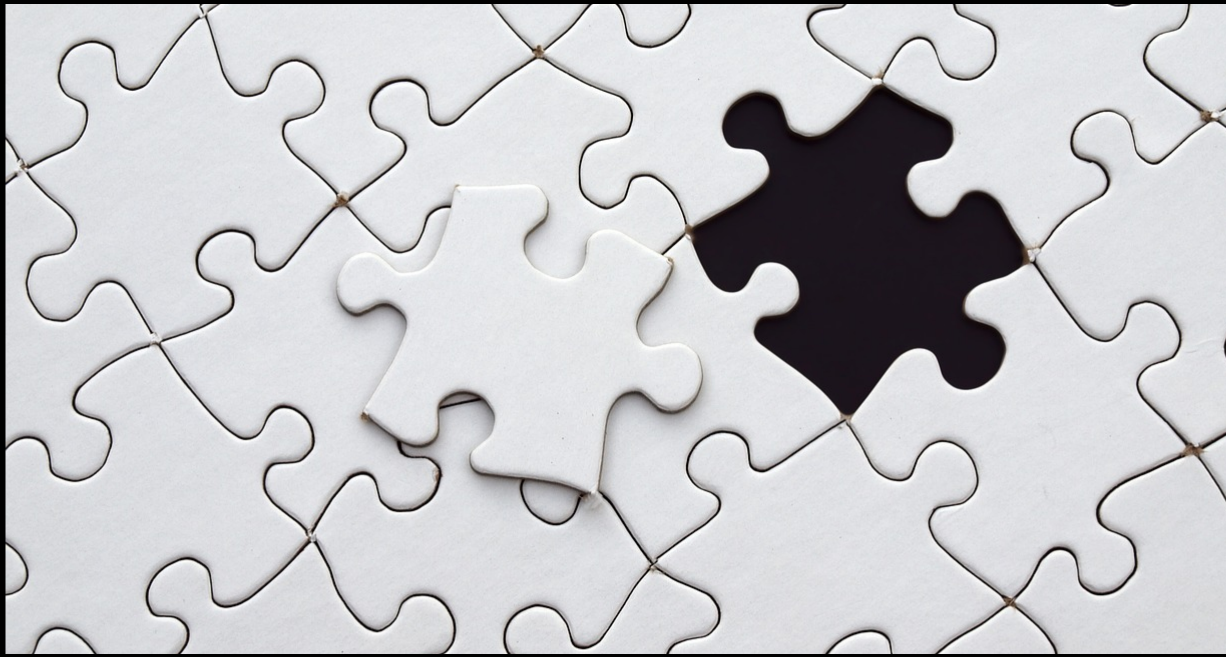
Results/DEMO

- **Competitive potential**
- Both LKKM and CRAFT locate most (>74%) anomalies with decent accuracy
- High false positive rates
- Framework **relies on the context learned by the base model**

- Base model is not performing
- Sultani claims:
  - AUC of 0.7541
  - 1 in 50 normal videos raising false alarm

- What could be wrong?
  - Logic issues: double checked
  - C3D features are incorrect?

    Author's results are perhaps unrealistic?



End Game

Today

- Report
- Clean dataset as much as possible
- Recompute all C3D features using Tensorflow+Keras
- Complete training of base model with new features
- 20 Sept
- Experiment with basic modifications to base model architecture
- Finalize the combination of scores to form consensus - optimize according to results
- Qualitative tests at extremes of problem domain
- 27 Sept

End

References

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

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

[3] RAFT: Teed, Z. and Deng, J., 2020, August. Raft: Recurrent all-pairs field transforms for optical flow. In *European conference on computer vision* (pp. 402-419). Springer, Cham.

[4] Lukas-Kanade: Bouguet, J.Y., 2001. Pyramidal implementation of the affine Lucas Kanade feature tracker description of the algorithm. *Intel corporation*, *5*(1-10), p.4.

[5] K-Means Clustering Burkardt, J., 2009. K-means clustering. *Virginia Tech, Advanced Research Computing, Interdisciplinary Center for Applied Mathematics*.

Questions?