

Decoding the Blackbox of AI Video Surveillance: What Makes Anomaly Detection So Difficult?

Elena Ramlow
Georgetown University

ABSTRACT

Anomaly detection can be classified as both an object detection and action recognition problem, as violence often is recognized by interaction and objects present. Video analysis and action recognition are at the forefront of the field of artificial intelligence. However, they are the most difficult to find information on and to comprehend. This project explores the data collection, processing, and model building process that goes into video object detection and recognition and shed light on the mechanisms that enterprise “AI Video Surveillance” is built on.

BACKGROUND

- Applying AI models to video data is a hot topic
- The potential for video surveillance to use AI to recognize anomalous behavior could improve emergency response times and general safety
- Publicly available research on the subject does not provide detail or large model capability
- Enterprise models seem to offer significant utility, but methodology is unknown

DATA

UCF Crime Dataset:
1900 untrimmed surveillance videos depicting 13 forms of anomalies: abuse, arrest, arson, assault, burglary, explosion, fighting, robbery → 48 videos labeled Abuse

Annotated using CVAT in PVOOC format



MODELS

Transfer learning: final layers of pretrained models

Original models from Tensorflow Detection Zoo

Both models use Inception V2 Feature Extractor for consistency

Faster RCNN
Higher accuracy

Image → CNN
Feature Map → RCNN
Proposals → pooling
Softmax classification
and box regression

SSD “Single Shot”
Faster, “Real time”

Combines region proposals and region classifications to simultaneously predict bounding box and class as image is processed, considering every bounding box in every location

RESULTS

Faster RCNN

Average Precision: 0.095
Total Loss: 0.13



SSD

Average Precision: 0.068
Total Loss: 5.79



Neither models performed well on the task. Potentially due to lack of input data, poor annotations, or there being better suited models

DISCUSSION

- Improving quality of annotations—multiple annotators for each image, requiring many hours of work
- Machine learning applications in video are time and resource expensive, require large amounts of manual encoding and large complex models
- Ethics:
 - Distress caused by having to watch disturbing videos to annotate

REFERENCES

Sultani, W., Chen, C., Shah, M. (2018) Real-world anomaly detection in surveillance videos. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6479-6488

Ko, T. (2011). “A Survey on Behavior Analysis in Video Surveillance Applications” Ch. 16 of Video Surveillance.