CNNs for Surveillance Footage

Utkarsh Contractor, Chinmayi Dixit, Deepti Mahajan

utkarshc@stanford.edu, cdixit@stanford.edu, dmahaj@stanford.edu

Dataset

CAVIAR (Content Aware Vision using Image-based Active Recognition)
INRIA (2013 - 14)

- Contains various scenarios such as people walking alone, meeting with others, window shopping, entering and exiting shops, fighting and passing out and leaving a package in a public place
- The videos were captured with wide angle lenses in two different locations

i-LIDS (Imagery Library for Intelligent Detection Systems)
The Home Office Scientific Development Branch, UK

- Contains videos in a train station with multiple events of abandoned luggage
- The videos were captured by a surveillance camera

The datasets are augmented:

- Greyscale / color images
- Flipped horizontally and vertically

Size:

- 9.3 GB, 65000 video frames
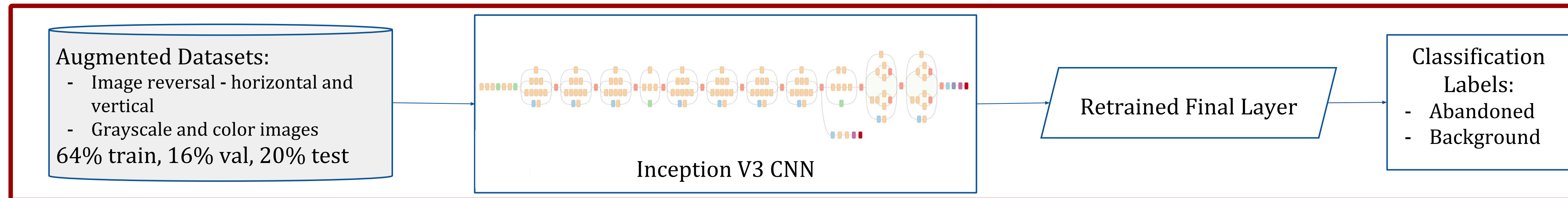- 30,000 abandoned luggage
- 35,000 background

Problem Statement

Using a mixed dataset with varying sources, camera angles, resolutions and backgrounds to identify abandoned baggage in surveillance videos.
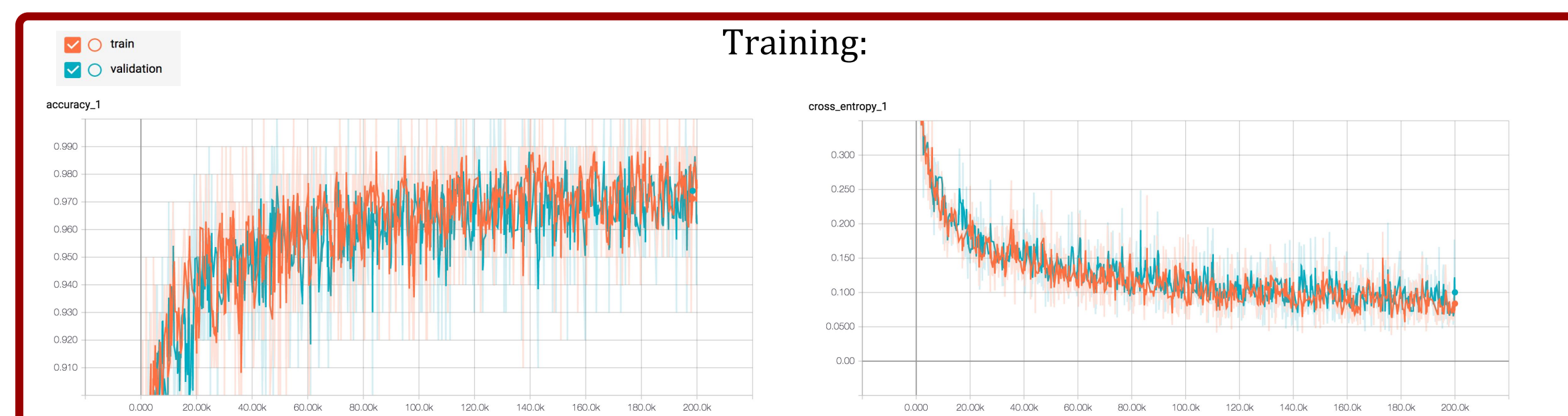
Challenges

- Varying resolution quality
- Different camera angles
- Variation in the background and activities that are considered 'normal'
- Low false negative as well as low false positive rates desired

Approach & Model

Augmented Datasets:
- Image reversal - horizontal and vertical
- Grayscale and color images

64% train, 16% val, 20% test

Inception V3 CNN

Retrained Final Layer

Classification Labels:
- Abandoned
- Background

Experiment Results

Training:

Step Count	200000	Validation 2 Accuracy	0.967
Learning Rate	0.01	Final Test Accuracy	0.98
Train accuracy	0.9587	False Positive Rate on Test	0.0065
Validation Accuracy	0.9459	False Negative Rate on Test	0.017

Background

Abandoned Luggage
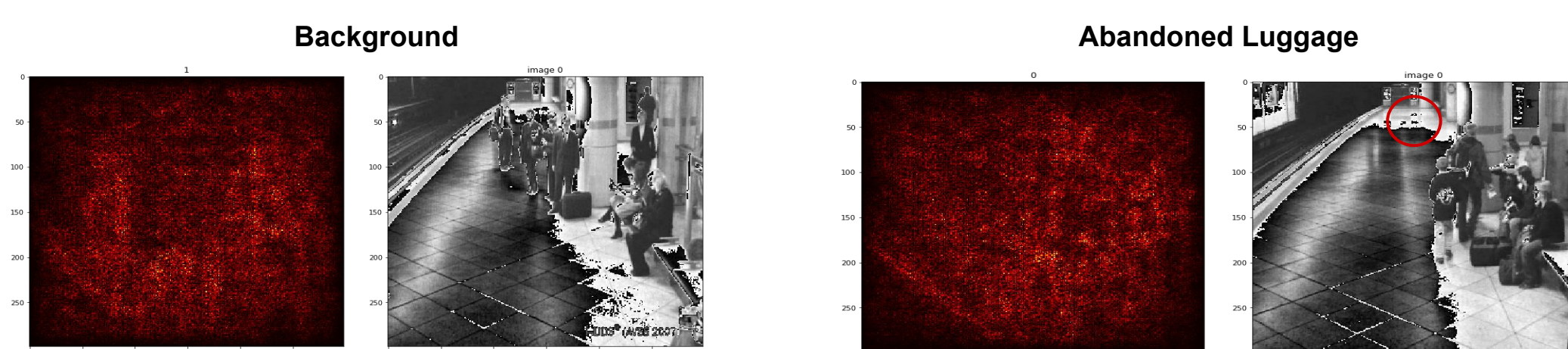
Analysis
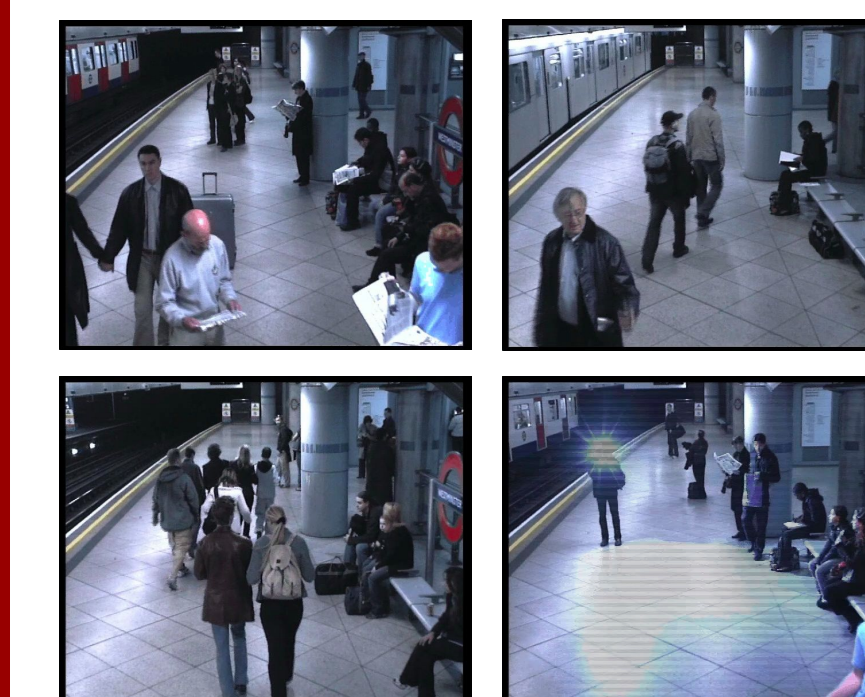
Misclassified as 'Background':
Bag partially/completely occluded
No temporal information

Misclassified as 'Abandoned':
Anomalous objects

- Model might have learnt the background!
- Generalizes well within the given type of data
- Might not work as well with a completely new setting

Future Work:
- Add temporal information using an RNN
- Augment dataset with different surveillance footage

REFERENCES:
Fisher, R., Santos-Victor, J., Crowley, J., 2005. "Caviar: Context aware vision using image-based active recognition."
i LIDS Team, 2006. "Imagery library for intelligent detection systems (i-lids); a standard for testing video based detection systems", *IEEE Carnahan Conf. Security Technology*, pp. 75–80.
"Train Your Own Image Classifier With Inception In Tensorflow". *Google Research Blog*. N.p., 2017. Web. 5 June 2017.
Karen Simonyan, Andrea Vedaldi, Andrew Zisserman., "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps", ICLR Workshop 2014