Analyzing Traffic Streams Using CNNs

A method comparison

Niels Broekmans Lennart Faber
Finn Gaida Rohit Malhotra

University of Groningen

22/01/2019

Introduction

- Problem description
- Methods
- Data
- Training
- Results
- Conclusion

Problem Description

- Gemeente Groningen: looking for ways to control traffic in city centre
- Live information on cyclist/pedestrian flows needed
- Solution: CNNs?

---

screenshot: https://www.rtvnoord.nl/nieuws/201369/Stad-roept-hulp-bedrijven-in-bij-oplossen-drukte-fietsers-en-voetgangers

Convolutional Neural Networks

- Used with success in various object detection tasks
- Fast: real-time detection on video is possible

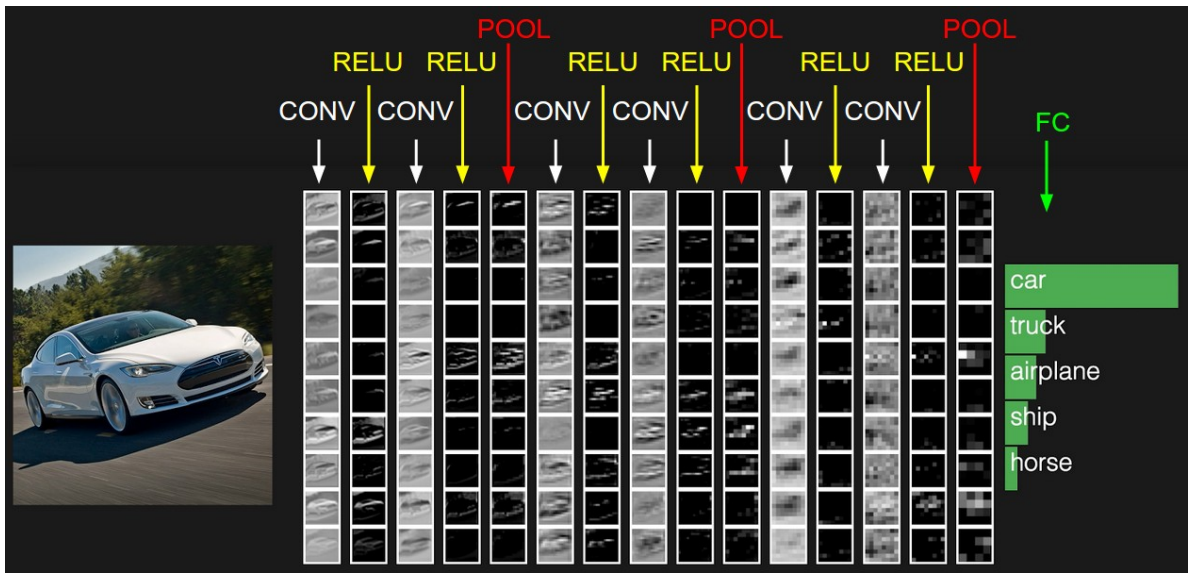


Image source + CNNs explained:
http://cs231n.github.io/convolutional-networks/

Methods

- Single Shot Detectors
  - YOLOv3 (original DarkNet & Keras implementation)
    - By Joseph Redmon (PJReddie) 2018
  - RetinaNet (aka Focal Loss)
    - By Lin et al. 2017
- Two-Stage Approach Detectors
  - Faster R-CNN
    - By Ren et al. 2015

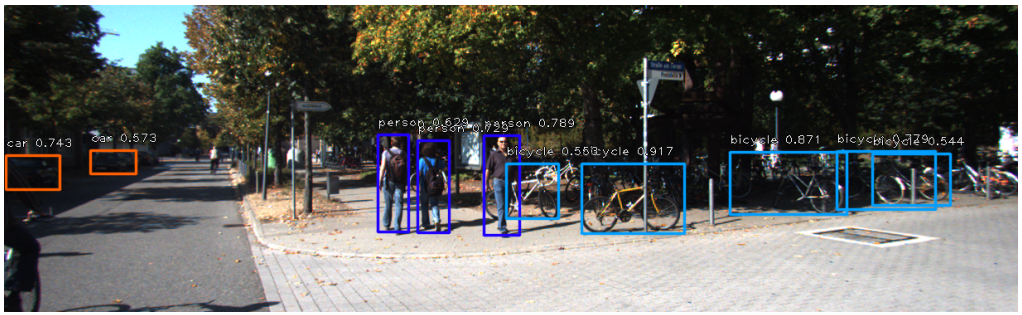
Data

- Training: MIO-TCD-Localization
  - 110,000 images gathered from traffic cams in USA/Canada
  - Released in 2018
  - 4452 images containing 8466 pedestrians + cyclists

Dataset reference: http://podoce.dinf.usherbrooke.ca/challenge/dataset/

Data

- Testing: KITTI Vision Benchmark
  - 15,000 images recorded with Google StreetView-like vehicle
  - testing on 2486 images of 4487 pedestrians and 1627 cyclists



  - including position in 3D space, but we're only using 2D bounding boxes from images with cyclists or pedestrians

---

Dataset reference: http://www.cvlibs.net/datasets/kitti/

Training

- Detectors
  - Faster-RCNN (Vgg16 backbone)
  - Faster-RCNN (ResNet50 backbone)
  - RetinaNet (Vgg16 backbone)
  - RetinaNet (ResNet50 backbone)
  - RetinaNet (DenseNet121 backbone)
- Pretrained weights for feature map detector (backbone)
  - Trained on COCO dataset
- Device
  - 1 Intel Xeon 2.5GHz CPU node and
  - 1 NVIDIA K40 GPU node available at peregrine cluster

Training Details

Detector	**Backbone**	**Loss Function**	**Optimizer**	**Learning Rate**
Faster RCNN	Vgg16	Classifier: Cross Entropy Regression: Smooth L1	Adam	0.00005
Faster RCNN	ResNet50	Classifier: Cross Entropy Regression: Smooth L1	Adam	0.00005
RetinaNet	Vgg16	Regression: Smooth L1 Classifier: Focal	Adam	0.00005
RetinaNet	ResNet50	Regression: Smooth L1 Classifier: Focal	Adam	0.00005
RetinaNet	DenseNet121	Regression: Smooth L1 Classifier: Focal	Adam	0.00005
YOLOv3 (Keras)(pretrained)	DarkNet53	Custom YOLO	Adam	0.001

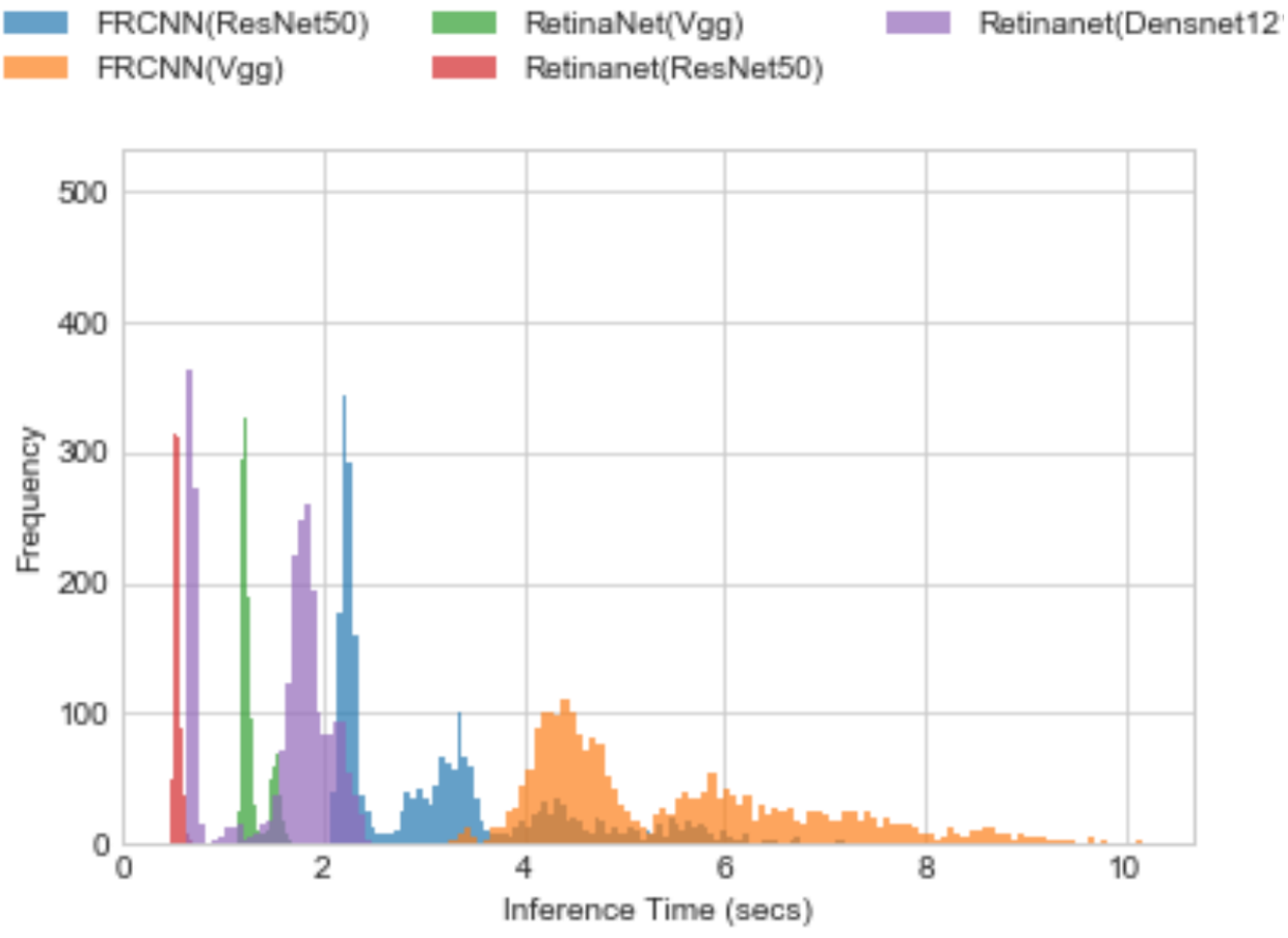
Results (Average Precision and mAP)

Detector	**AP (Bicycle)**	**AP(Pedestrian)**	**mAP**	**IoU Threshold**
Faster-RCNN (Vgg16)	0.24260	0.17862	0.21058	0.5
Faster-RCNN (Resnet50)	**0.76961**	**0.65782**	**0.7135**	**0.5**
RetinaNet (Vgg16)	0.1493	0.3516	0.2505	0.5
YOLOv3 (Keras)(pretrained)	0.0026	0.1918	0.00972	0.5
Retinanet (ResNet50)*	0.0000	0.0077	0.0057	0.5
RetinaNet (DenseNet121)*	0.0000	0.0003	0.0002	0.5

- Testing environment
  - 1 Intel Xeon 2.5GHz CPU node.

* : Still training

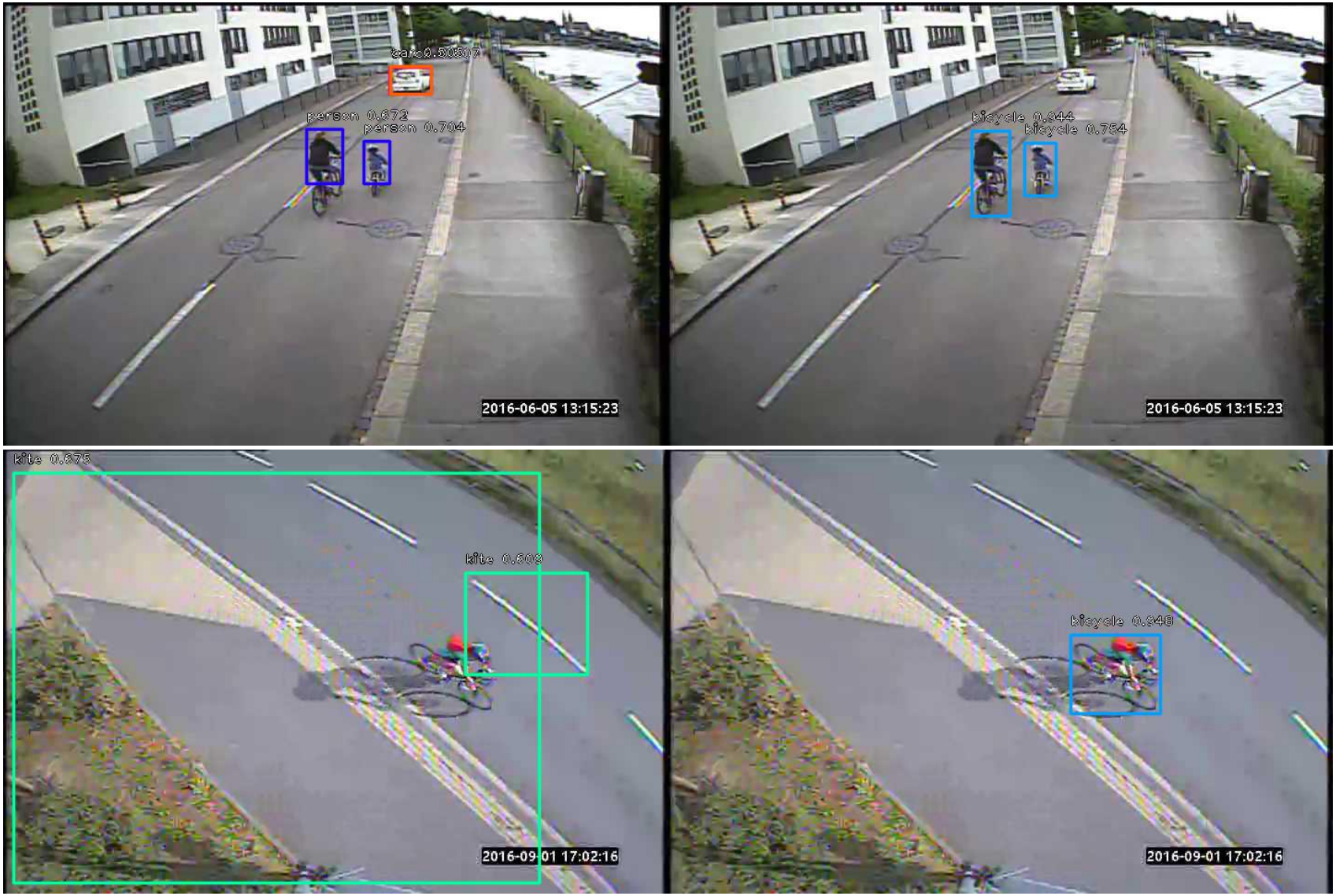
Results (Inference timing)



Results (Retinanet - ResNet50)

DEMO

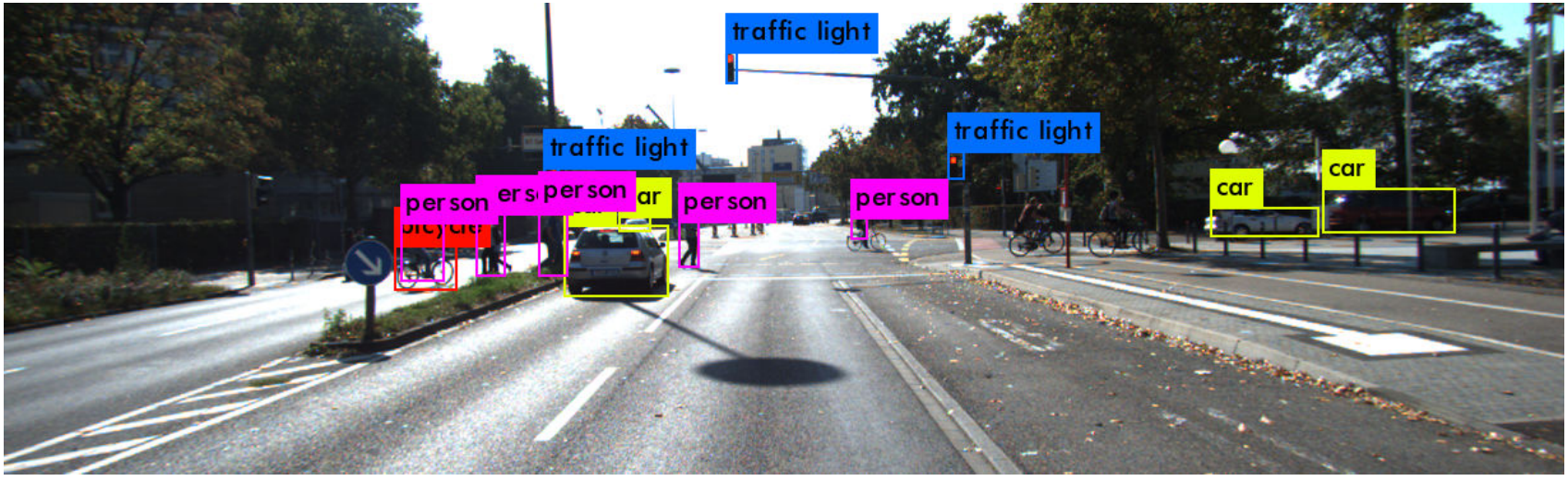
Results (Retinanet - ResNet50)



Results (Retinanet - ResNet50)

Expectation

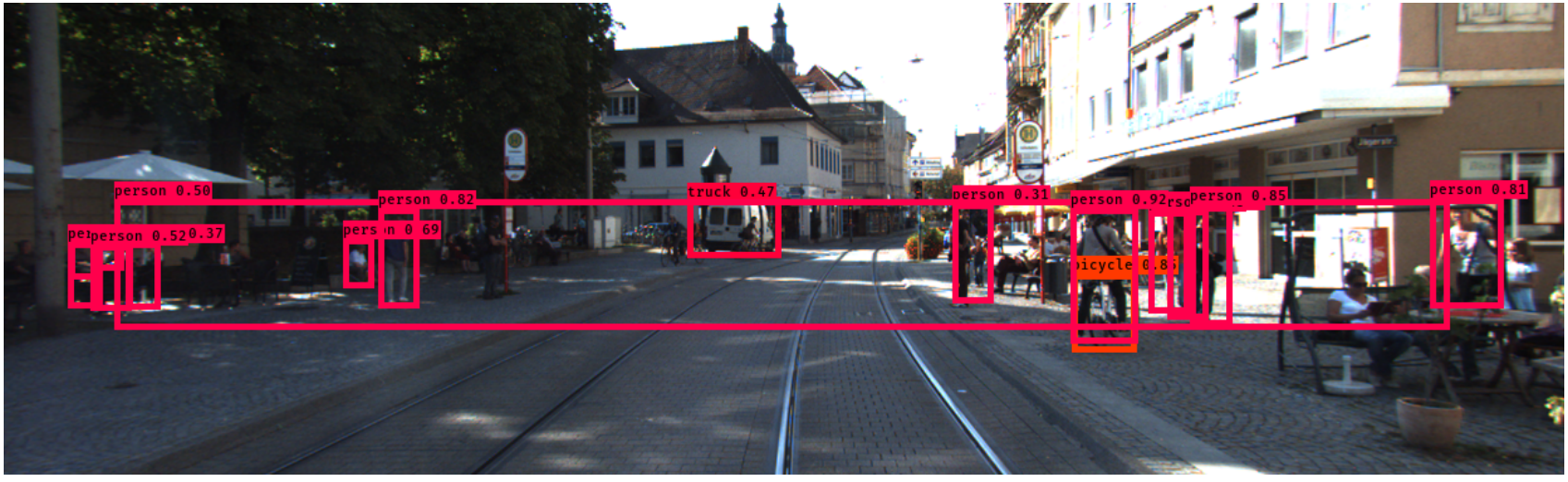
Results (YOLO Darknet)



Results (YOLO Darknet)

- Evaluated first 29% of output on default model:
  - Actual: 1238 pedestrians and 489 cyclists
  - Predictions: 1202 pedestrians (97.1%) and 498 cyclists (101.8%)
  - Conversion from bikes/persons still naive
- Slow: 29 seconds per image
- To do: train and calulate mAP

Results (YOLO Keras (pretrained)



Results (YOLO Keras (pretrained))

- Fast: < 0.5 seconds on CPU
- Training also problematic
- Problem: YOLO predicts bicycles & persons separately

Conclusion

- We have seen two stage detectors give higher accuracy but there inference time is also high.
- It's hard to account for differences in data quality
- Need to run on GPU to get close to real time processing

- **But:** Notable increase in cyclist and pedestrian detection after training
  $\Longrightarrow$ More data + more training might increase performance more

Thanks for your attention!

Are there any questions?