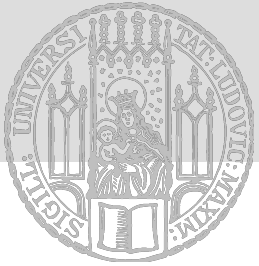
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN



Can you still see me?

The effect of image compression on state-of-the-art object detection neural networks

Diana Davletshina, Hitansh Singla
In cooperation with Harman and Dr. Matthias Schubert

17.02.2020

Ground truth

1% JPG

Motivation

Problem:

- High quality data generated by **front facing cameras** on cars (~9GB / minute)
- 10 cars x 8 hours x 60 mins x 9GB = **43 TB / day**
- ~1PB/month
- Storage of 1PB is costly for a company (~5000$ / month).

Can you still see me? - Motivation

Question: In the object detection framework, how much can image be **compressed** and **preserve accuracy** of object detection?

100% jpg

20% jpg

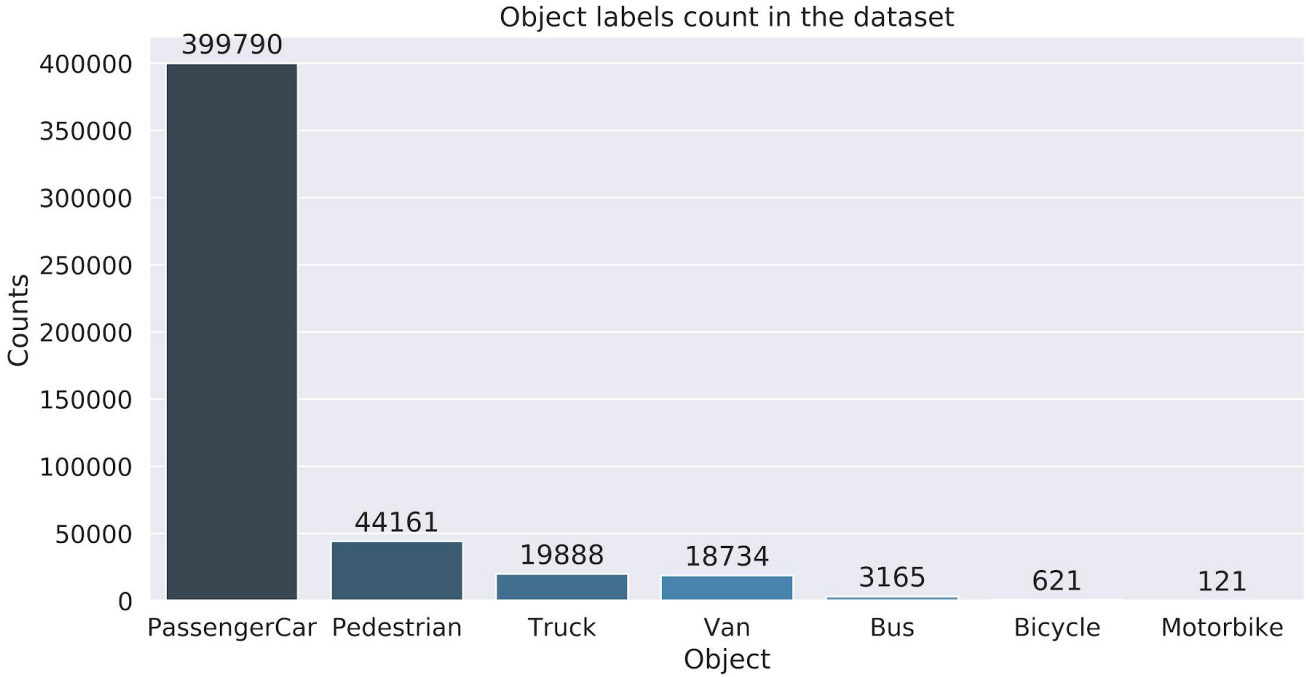
5% jpg

Data

- **57935 labelled images** in full resolution (.bmp format)
  - ~800 broken images among them
- **7 labelled classes**
- **Fisheye camera** (120°)
- **3 frames / minute**

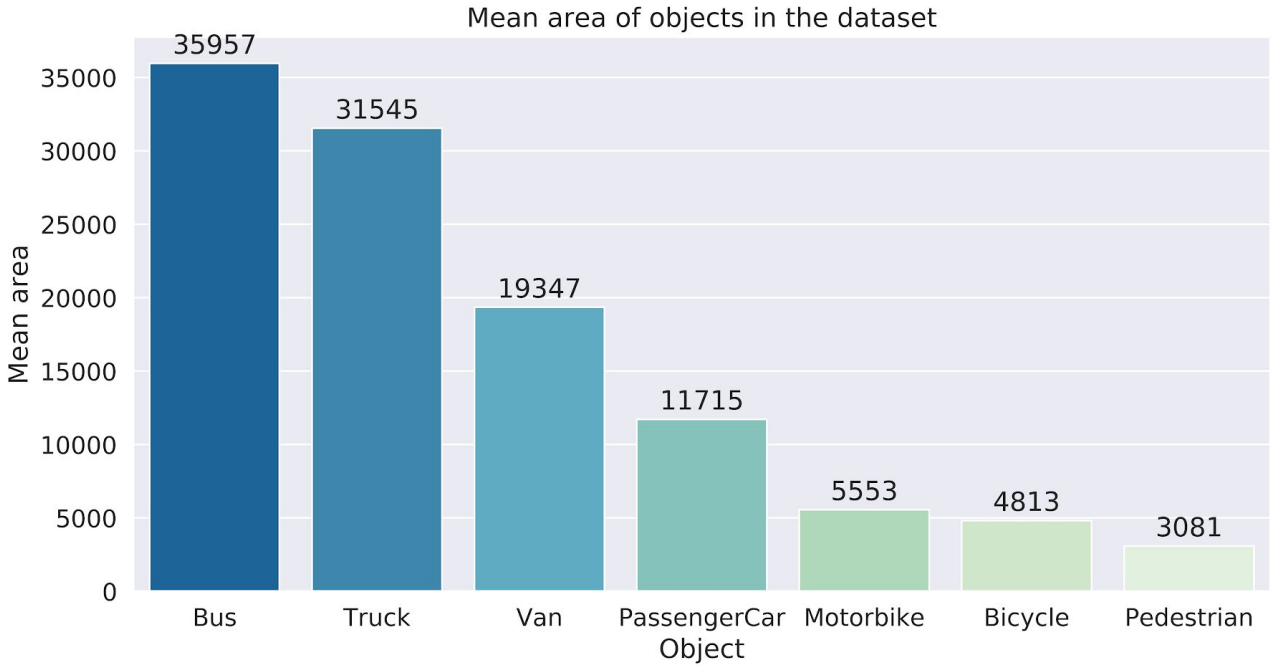
Data Labels

- PassengerCar
- Truck
- Van
- Bus
- Pedestrian
- Bicycle
- Motorbike



Object Localization

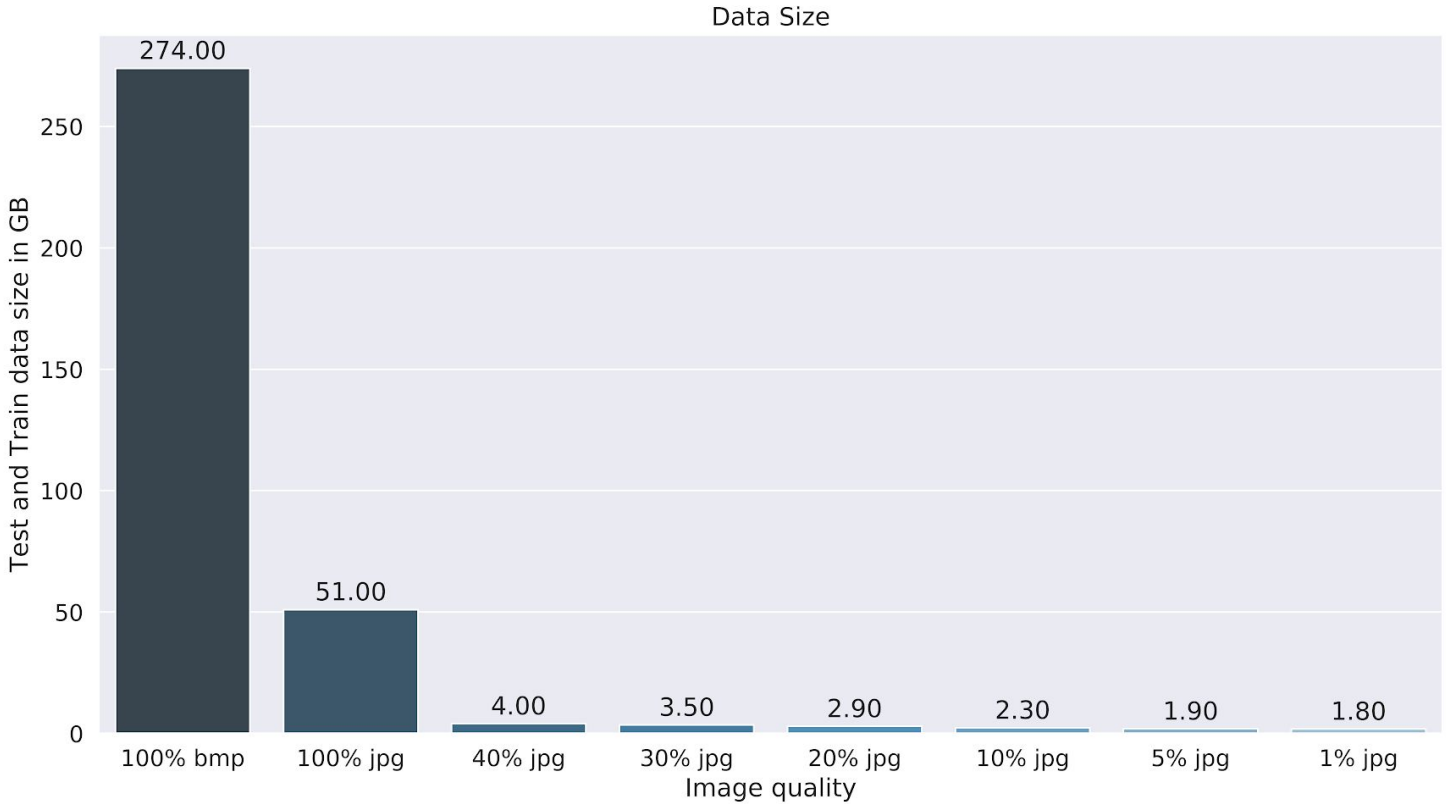
Bounding boxes



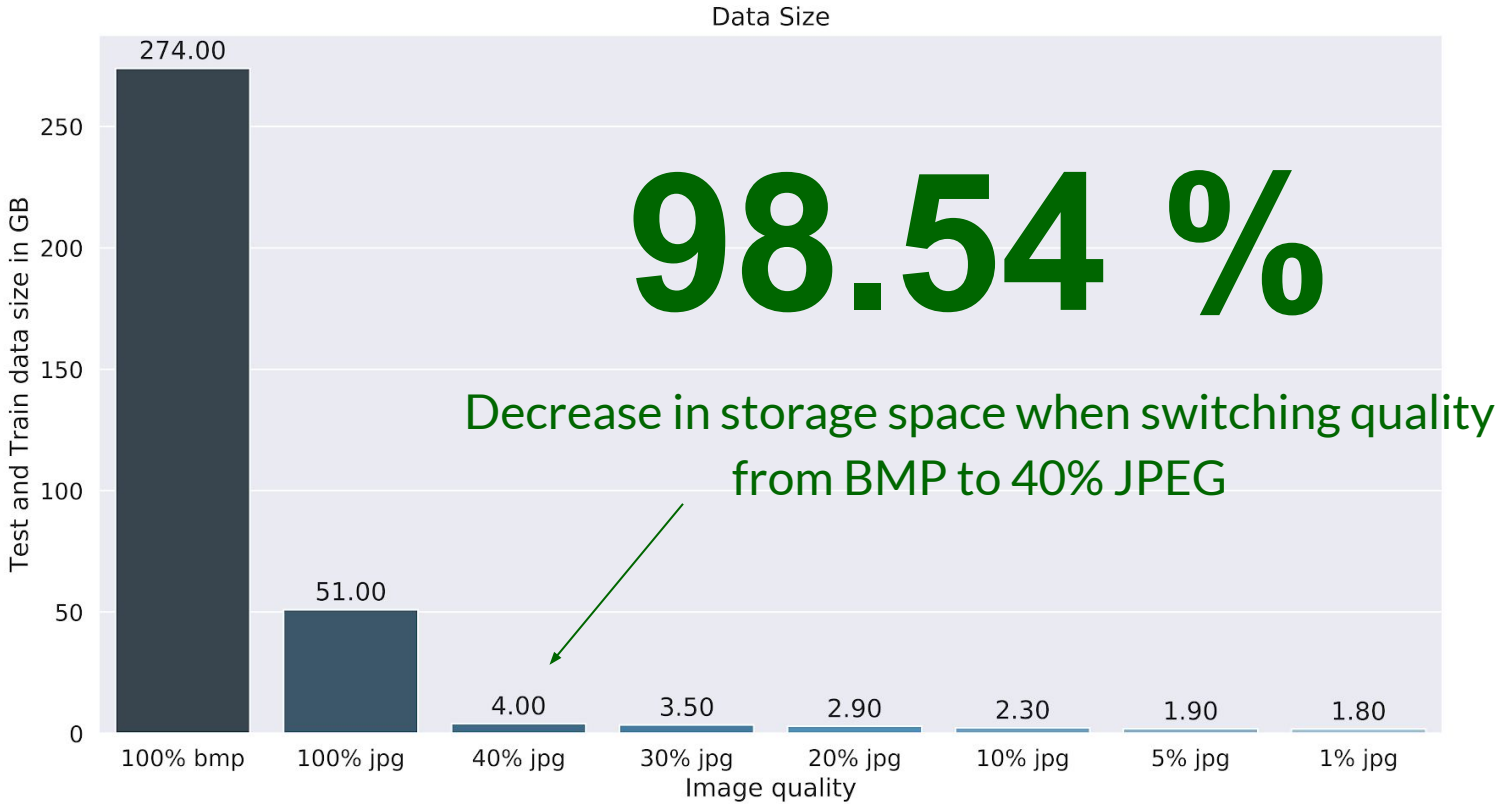
JPEG Compression

- **the most common** method of compression for digital images;
- uses a **lossy** form of compression based on the discrete cosine transform;
- "**lossy**" means that some original image information is lost and cannot be restored, affecting image quality.

Can you still see me? - Compression



Can you still see me? - Compression



Literature overview

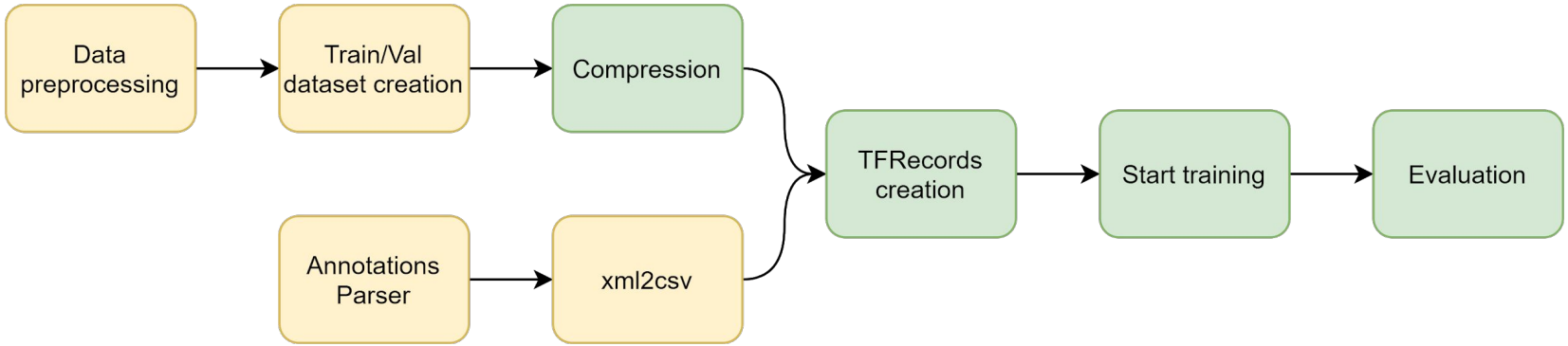
- **Low resolution [Koziarski and Cyganek, 2018]**
  - **Networks**: AlexNet, VGGNet, and ResNet
  - **Results**:
    1. Final accuracy decreases with picture size
    2. Applying super-resolution allows achieving classification accuracy close to that with original size

- **Image distortion [Dodge and Karam, 2016]**
  - **Image transformation**: blur, noise and contrast
  - **Results**:
    1. NN accuracy dropped dramatically
    2. NN rely heavily on textures in object classification, which get erased after transformations

Literature overview

- **Image compression**
  [Zanjani et al., 2019], [Dodge and Karam, 2016], [Dejean-Servieres et al., 2017]
  - **Compression:** JPEG
  - **Networks**: different variations of VGGNet, AlexNet
  - **Results**: pictures with similar-to-original quality have close values of target metrics
  - **[Zanjani et al., 2019]** experiments:
    - Training on uncompressed data, testing on compressed
    - Training and testing on compressed data -> better results

Solution Pipeline



Main tools: Python 3.7, TensorFlow 1.12, VCS Github

Transfer Learning

Used pre-trained tensorflow models:

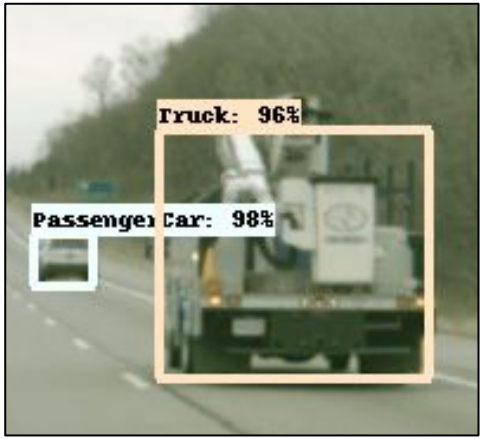
- SSD inception v2 coco
  - SSD refers to the architectures that use a single feed-forward CNN to directly predict classes and anchor offsets without requiring a second stage per-proposal classification operation
- Faster RCNN
  - Detection happens in 2 stages:
    - Region Proposal Network: Predicts class-agnostic box proposals
    - Predict a class and class-specific box refinement for each proposal
  - Resnet-101
  - Inception V2

Model settings

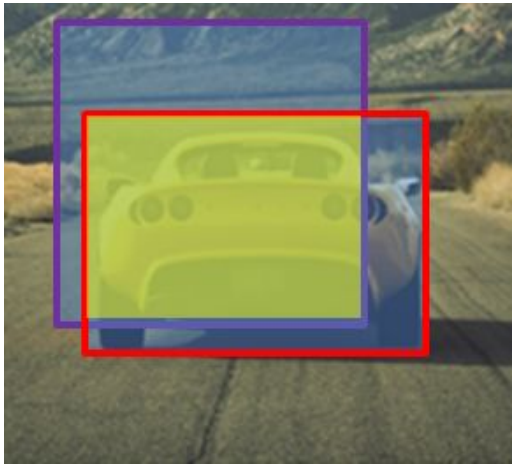
- **SSD inception v2 (COCO Dataset)**
  - 300x300, 600x600
  - 12-14 hours per training
  - 128 batch size
  - 15.000 steps
- **Faster RCNN ResNet101 (Kitti Dataset)**
  - 1820x940
  - 3 hours per training
  - 4 batch size
  - 20.000 steps
- **Faster RCNN Inception v2 (COCO Dataset)**
  - 1820x940
  - 6 hours per training
  - 16 batch size
  - 20.000 steps

Evaluation metrics

- **Classification**
  - Precision, Recall, F1
- **Object Detection**
  - Mean Average Precision (mAP)
  - Weighted mAP for imbalanced classes



Intersection over Union

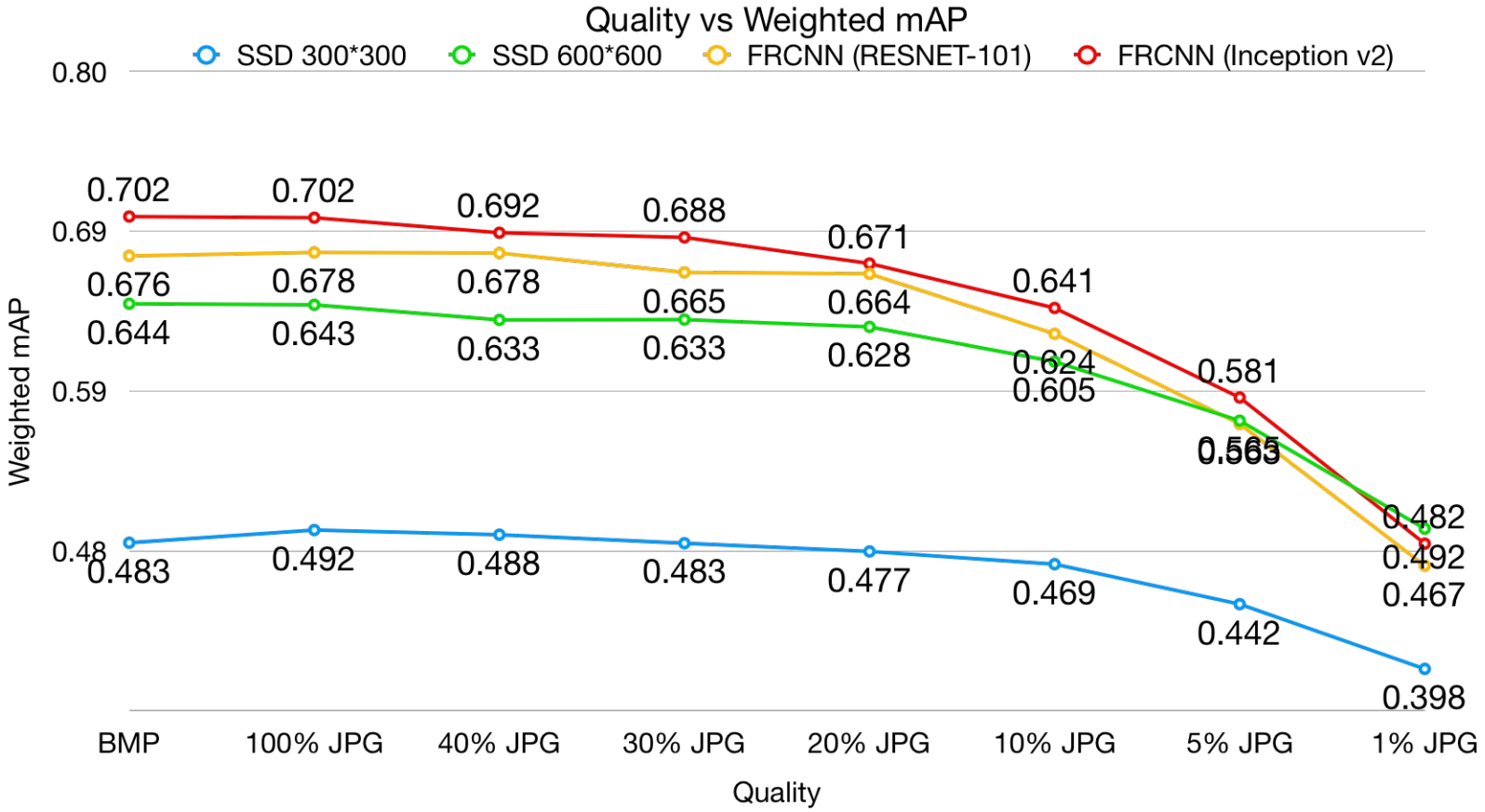


$$IoU = \frac{\text{Area of Intersection}}{\text{Area of Union}}$$

A prediction is correct if IoU ≥ 50%

Can you still see me? - Results

Can you still see me? - Results



Compression of grayscale images

- **Idea:** compression of grayscale image will better preserve some edge features for lower qualities

- Experiments on 40% and 5% quality data with SSD 600x600

Can you still see me? - Results

40%

Can you still see me? - Results

5%

Did grayscale help?

SSD at 600*600	Color image	Grayscale image
40%	0.632856 (4 GB)	0.635504 (3.1 GB)
5%	0.565084 (1.9 GB)	0.559135 (0.75 GB)

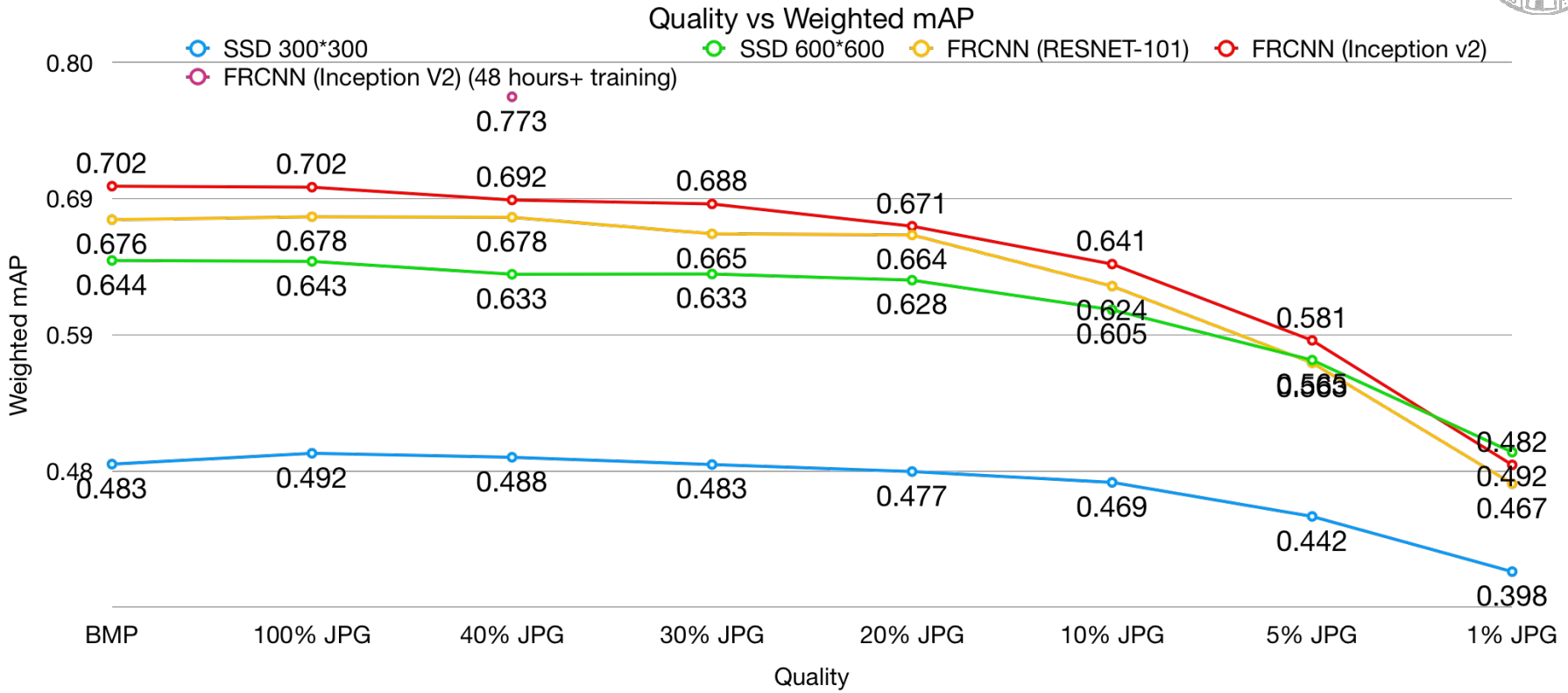
Did increasing the steps help?

FRCNN (Inception V2)	20.000	200.000
Time taken	ca 6 hours	ca 48 hours +
Weighted mAP	0.6915	0.772963

an increase of 8%

* at 40% JPEG quality

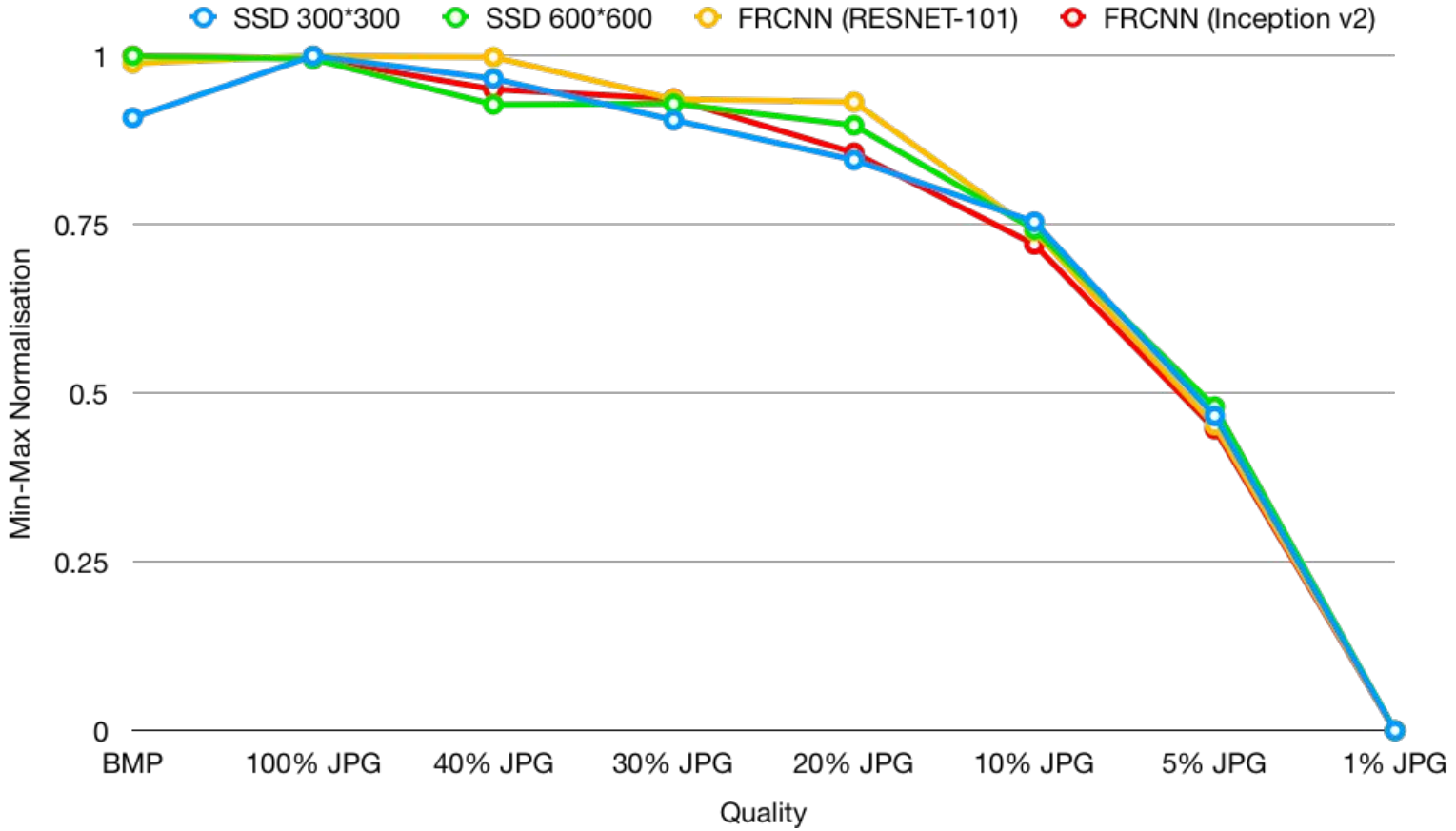
Can you still see me? - Results



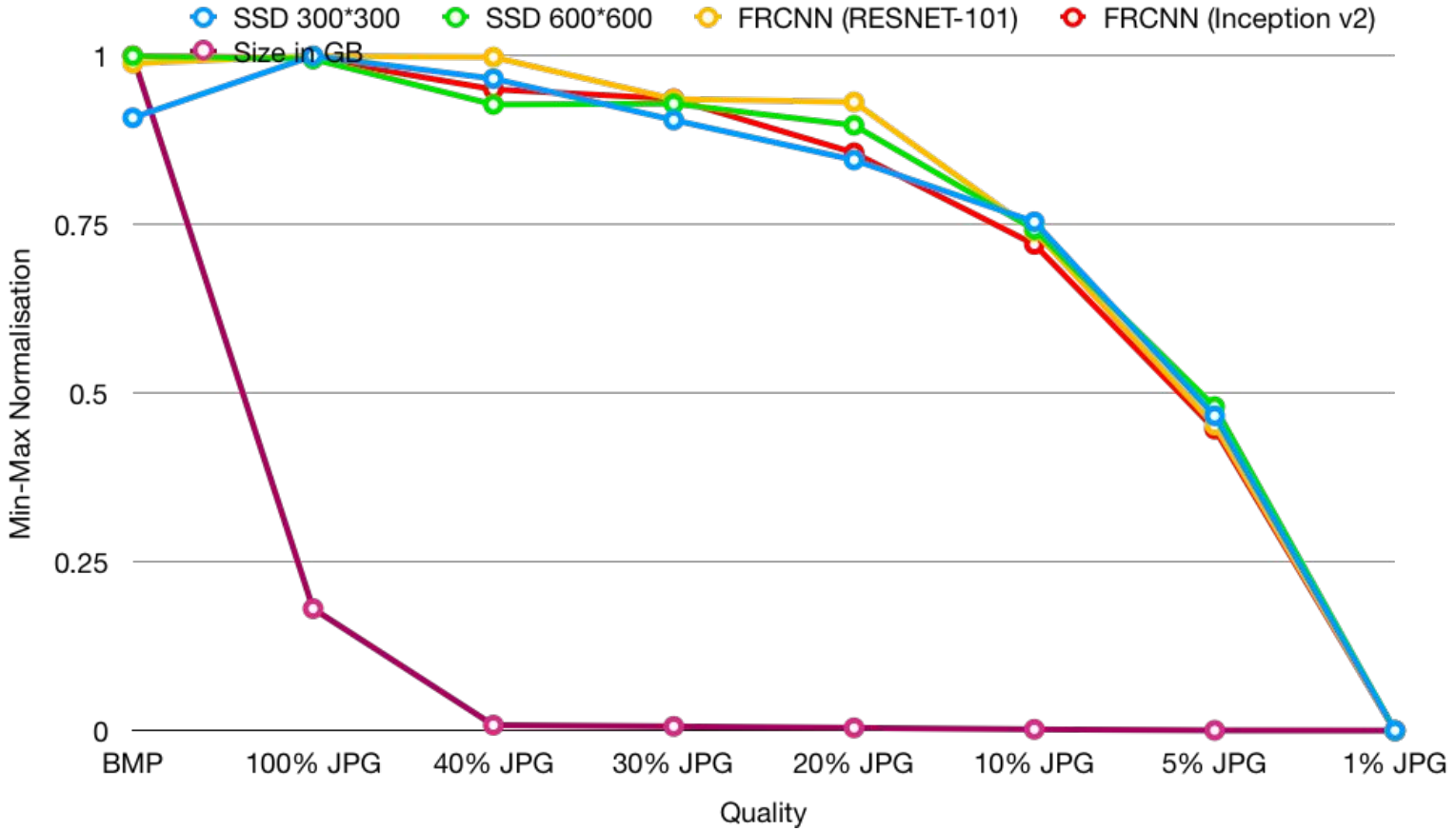
Comparing performance

We performed min-max scaling on each result, in order to normalise the data on 0-1 scale

Min-max Normalised Results



Min-max Normalised Results



Can you still see me? - Results

Ground truth

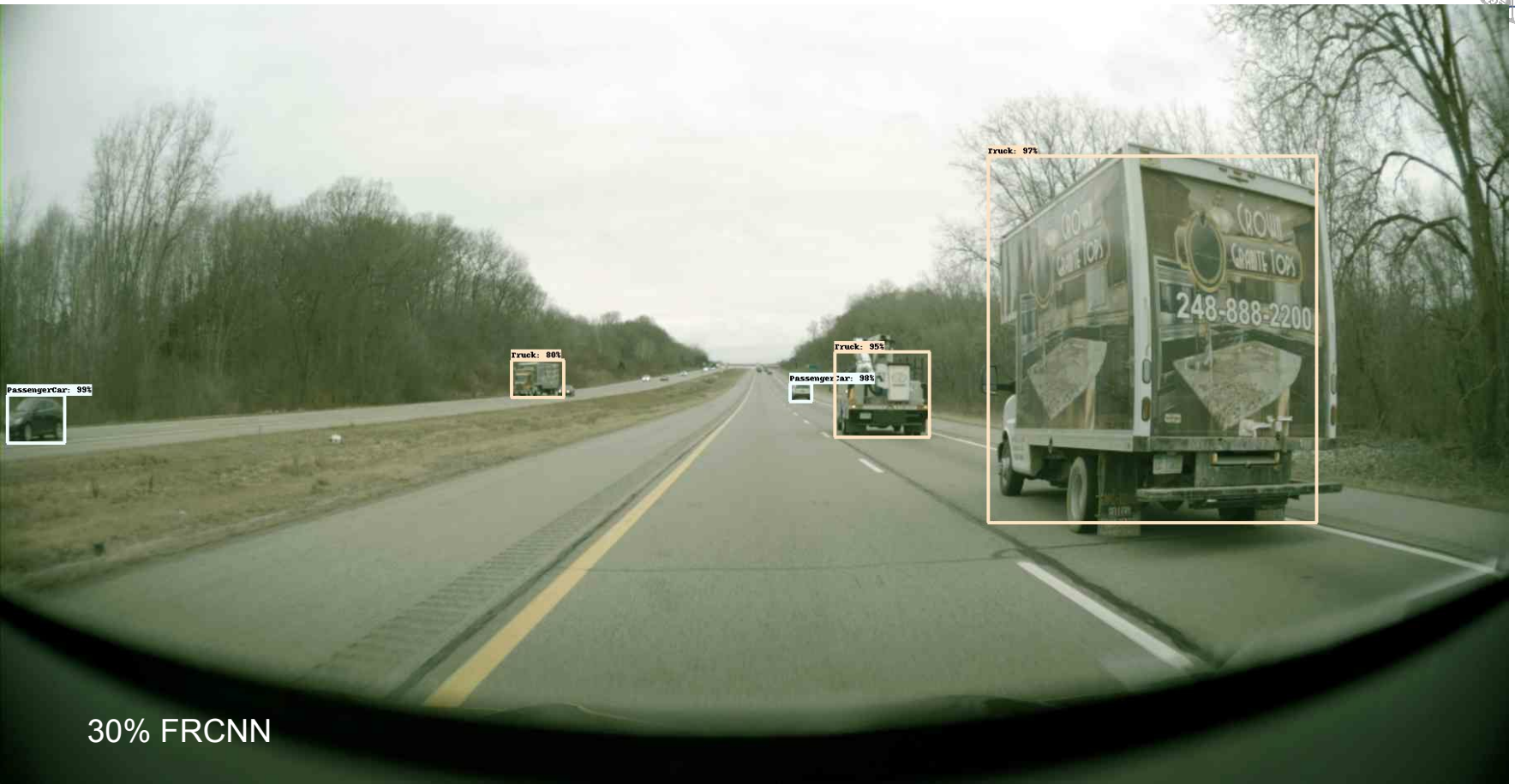
Can you still see me? - Results

100% FRCNN

Can you still see me? - Results

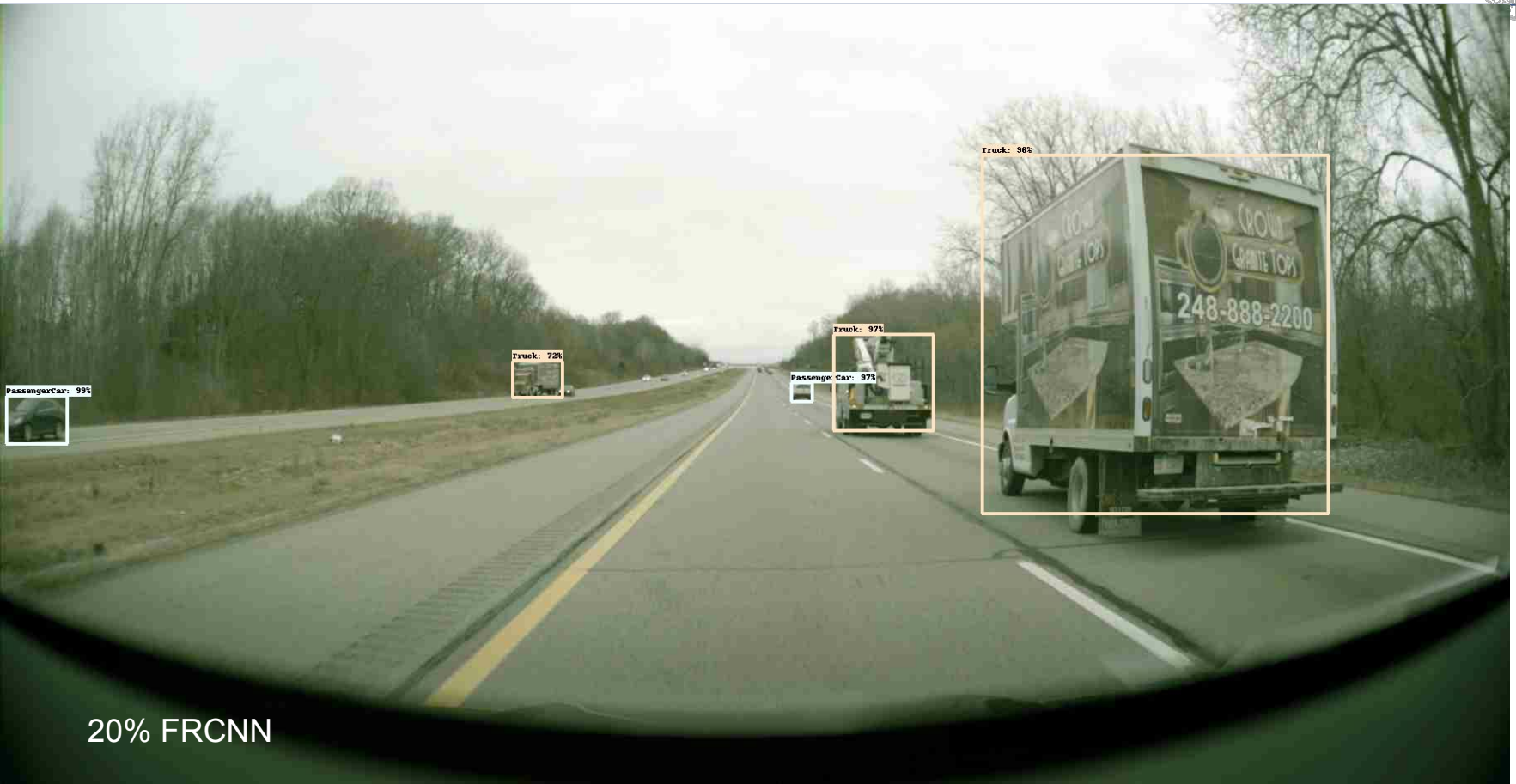
40% FRCNN

Can you still see me? - Results



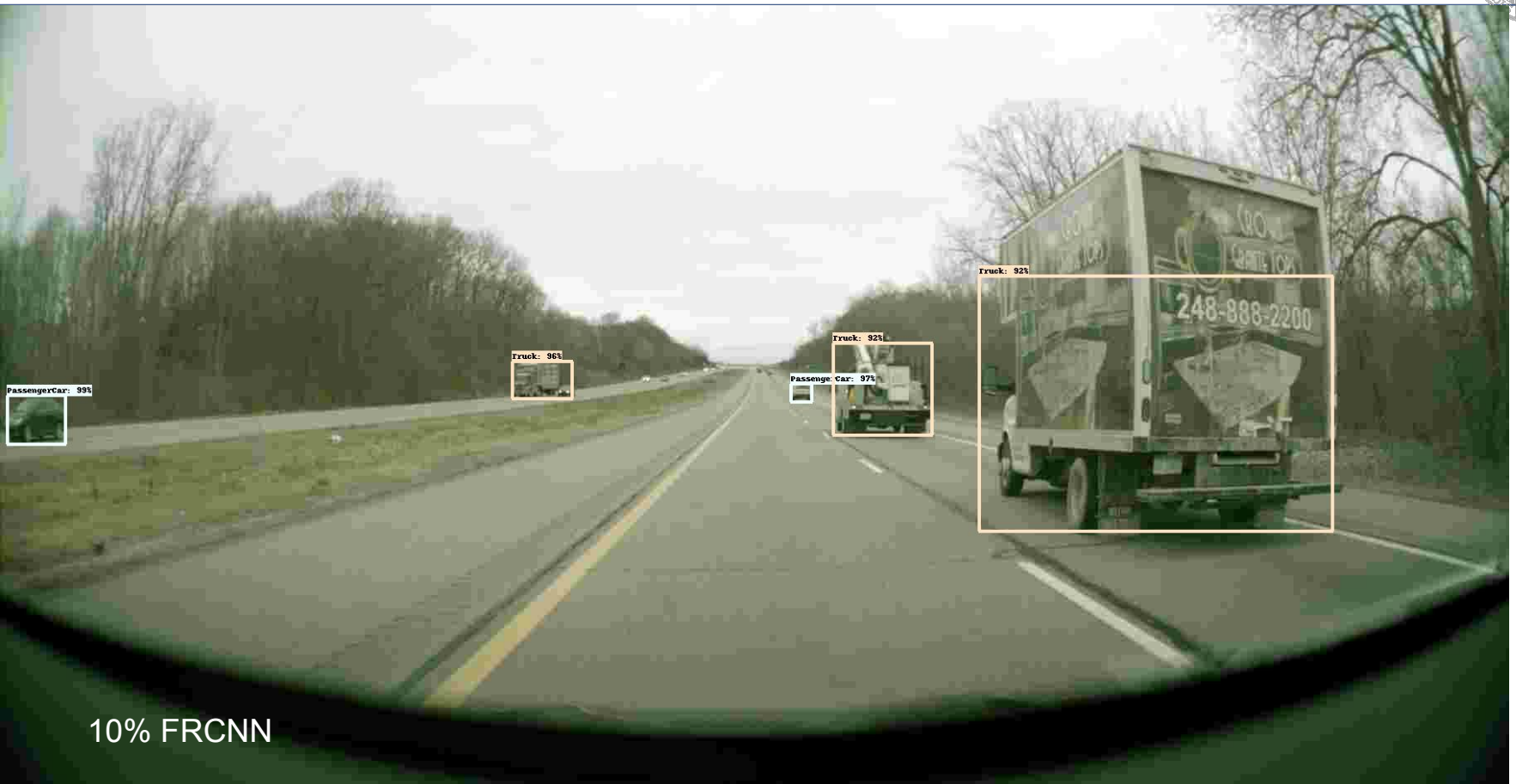
30% FRCNN

Can you still see me? - Results



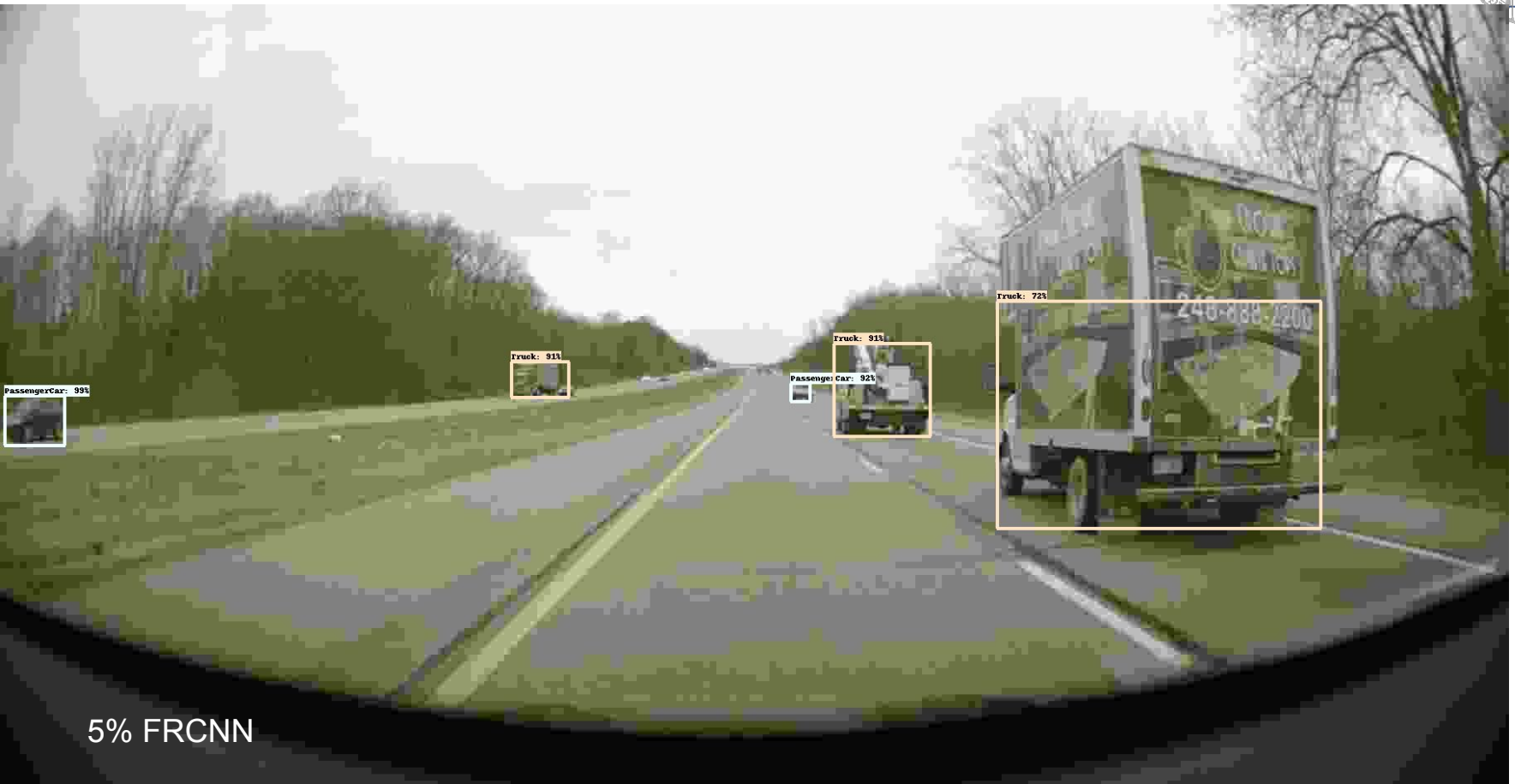
20% FRCNN

Can you still see me? - Results



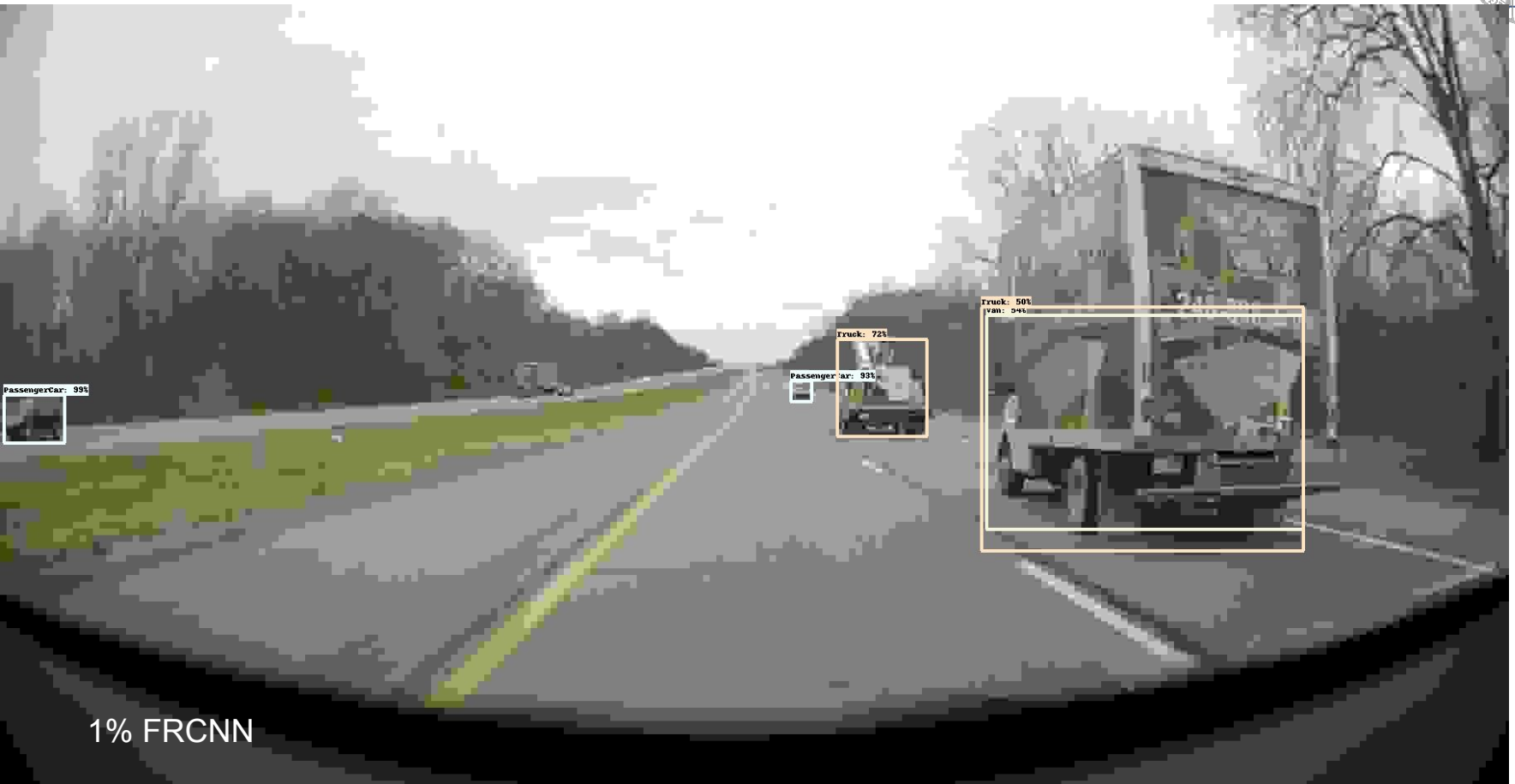
10% FRCNN

Can you still see me? - Results



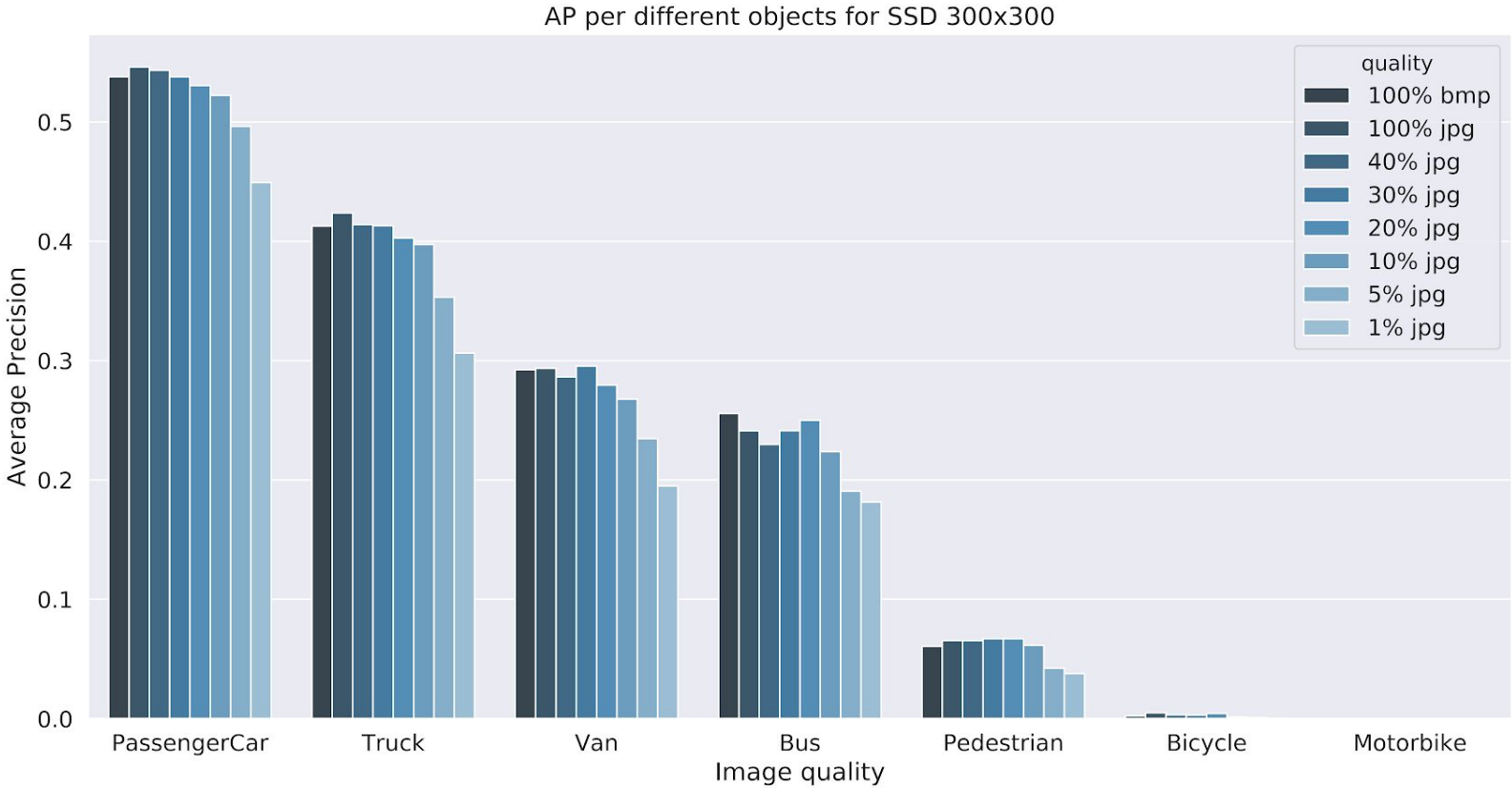
5% FRCNN

Can you still see me? - Results

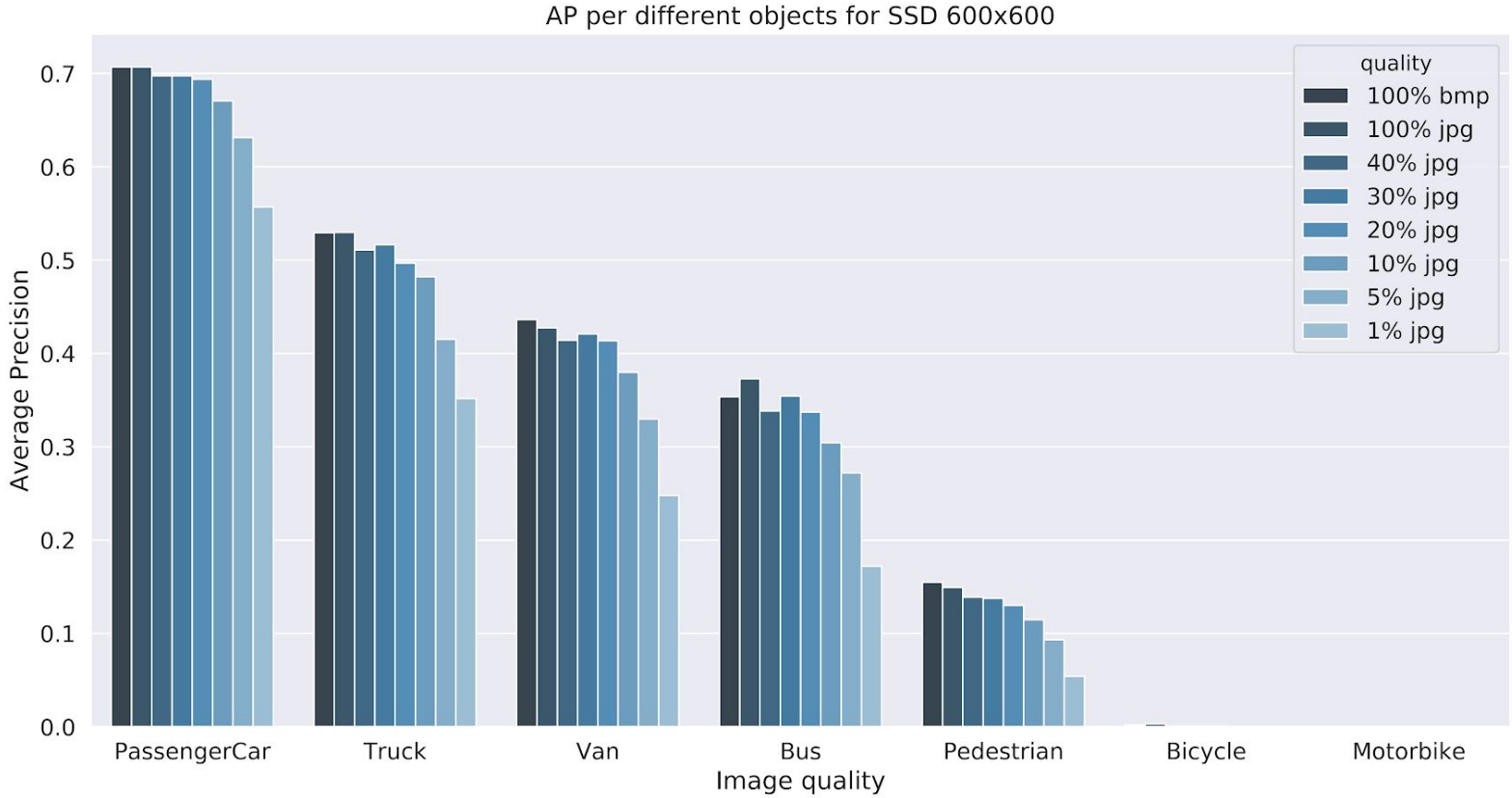


1% FRCNN

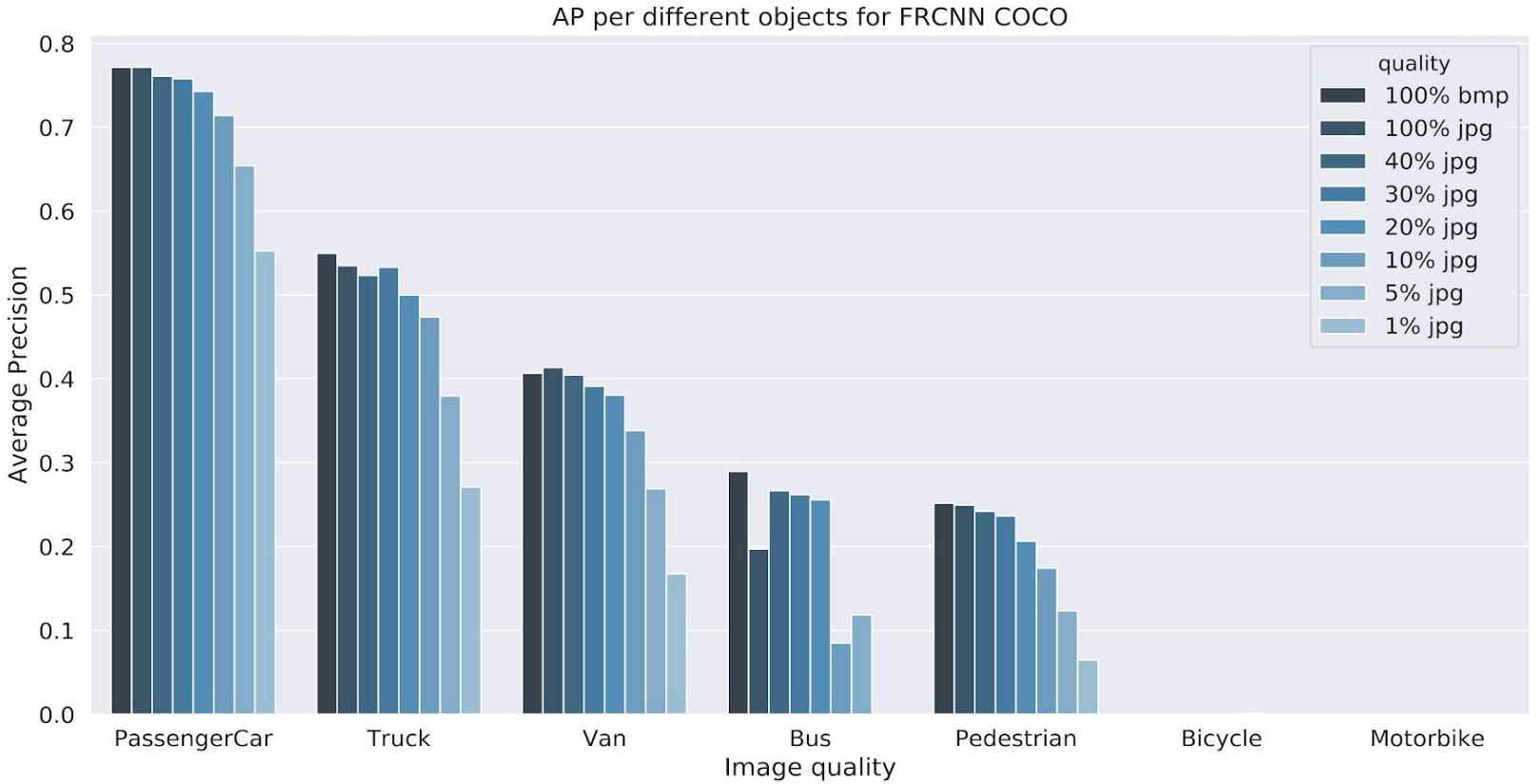
Can you still see me? - Results



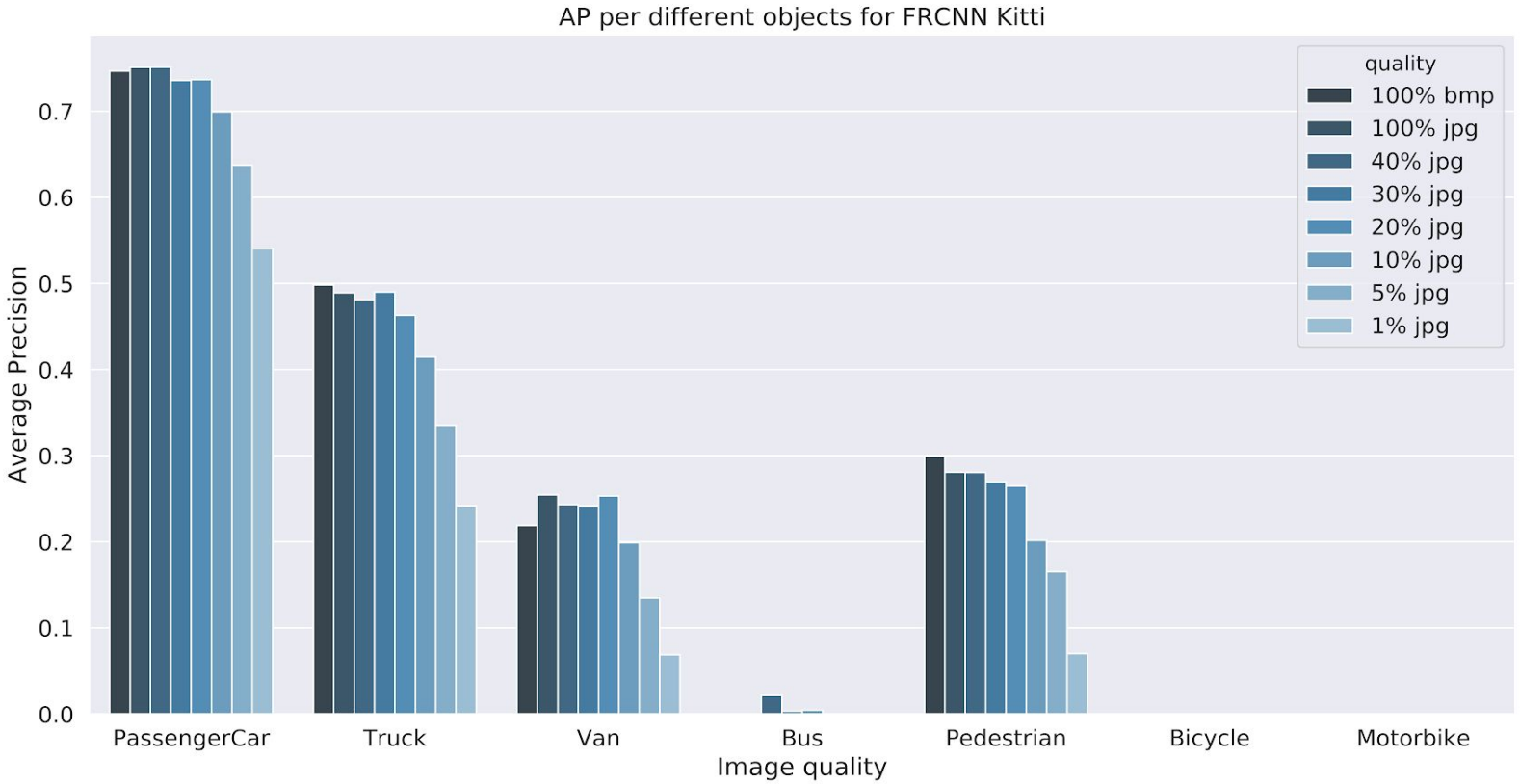
Can you still see me? - Results



Can you still see me? - Results

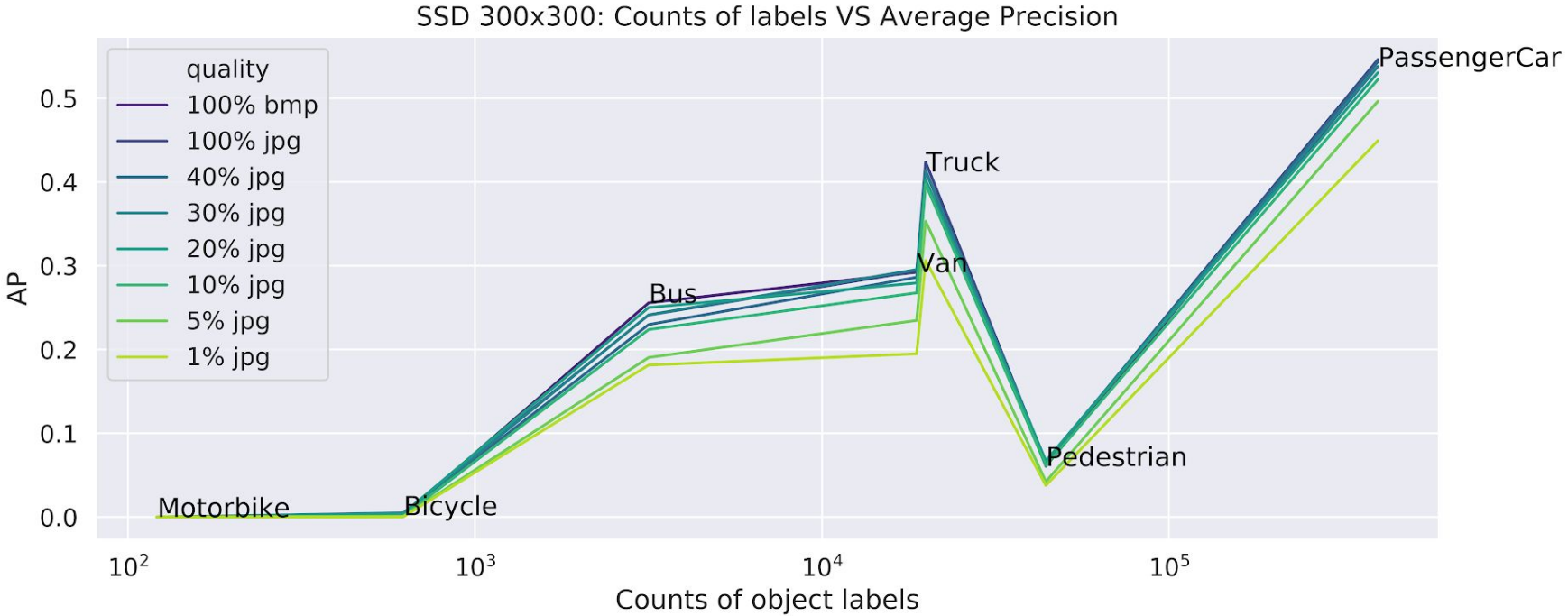


Can you still see me? - Results

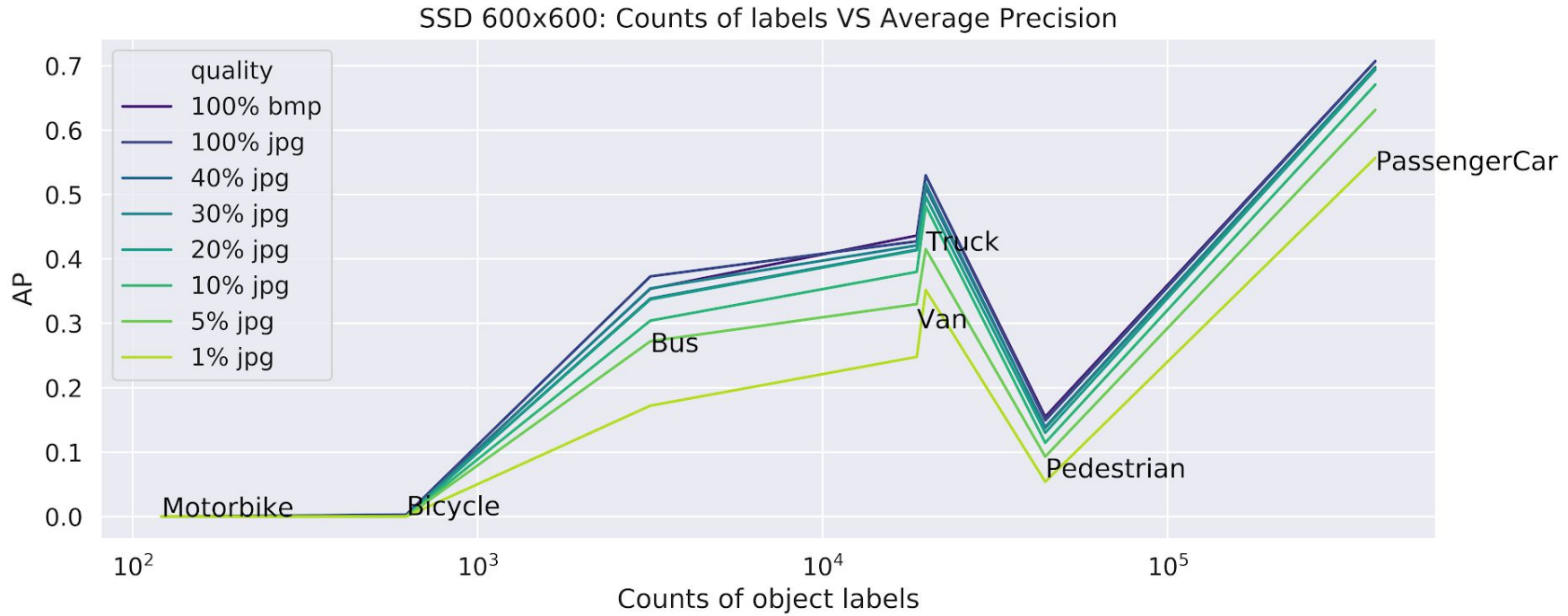


Performance insights

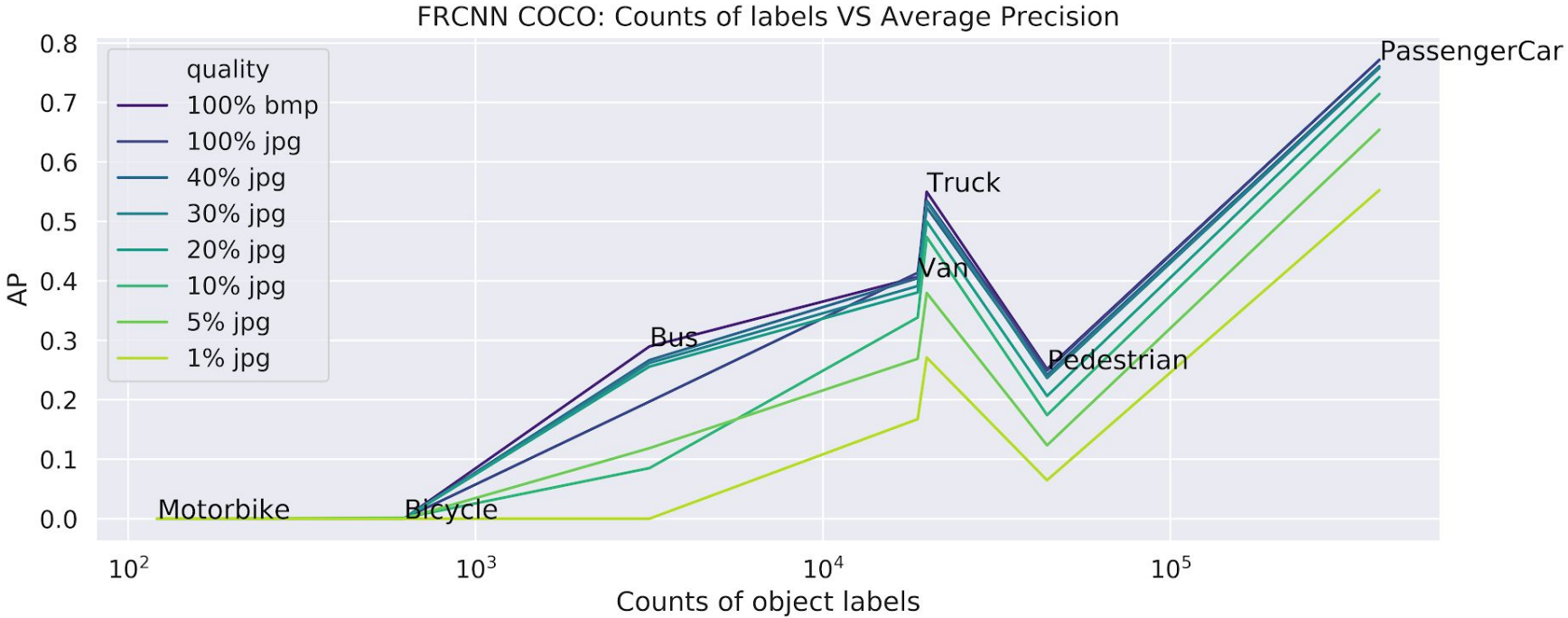
Can you still see me? - Results



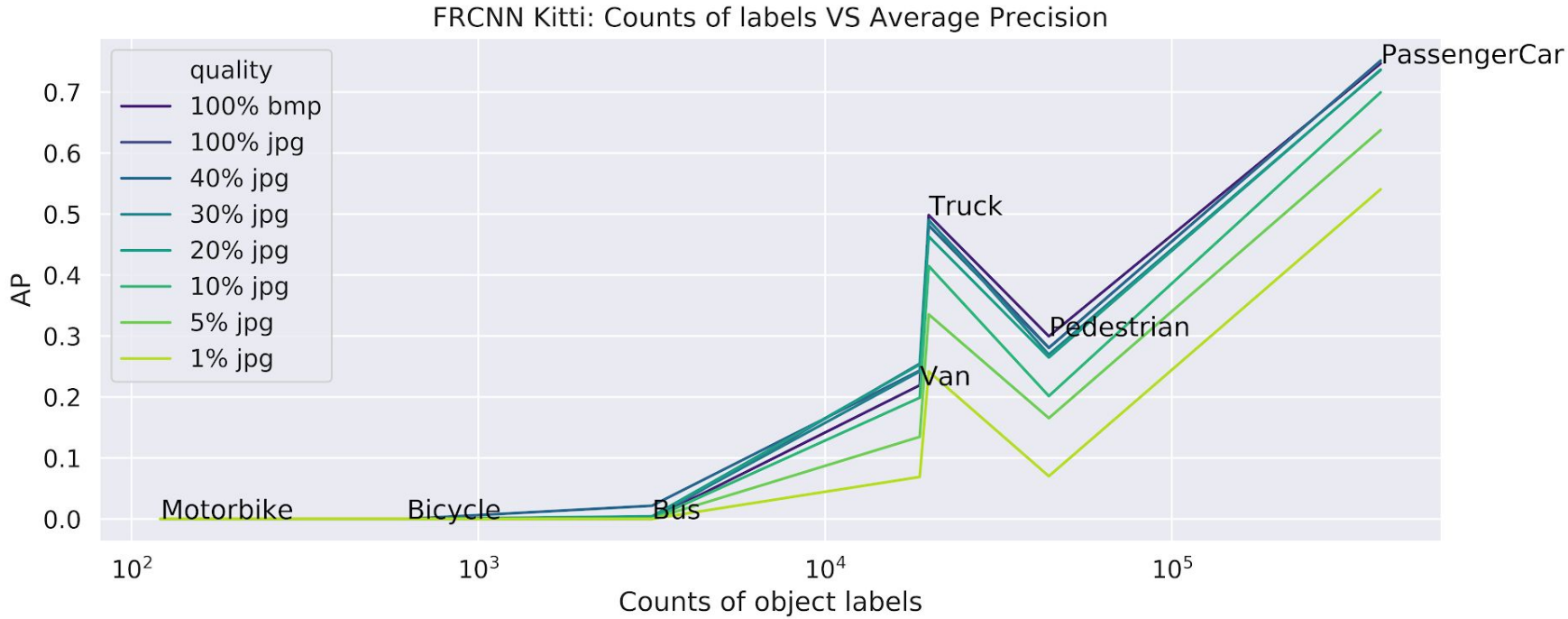
Can you still see me? - Results



Can you still see me? - Results

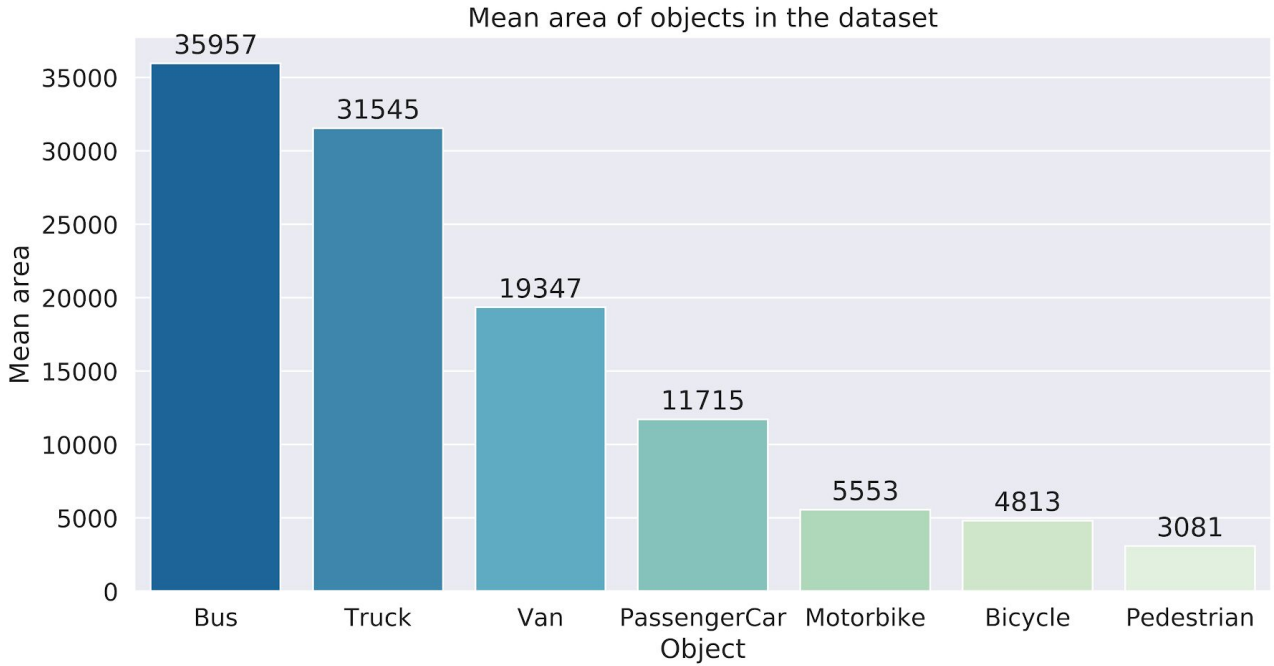


Can you still see me? - Results



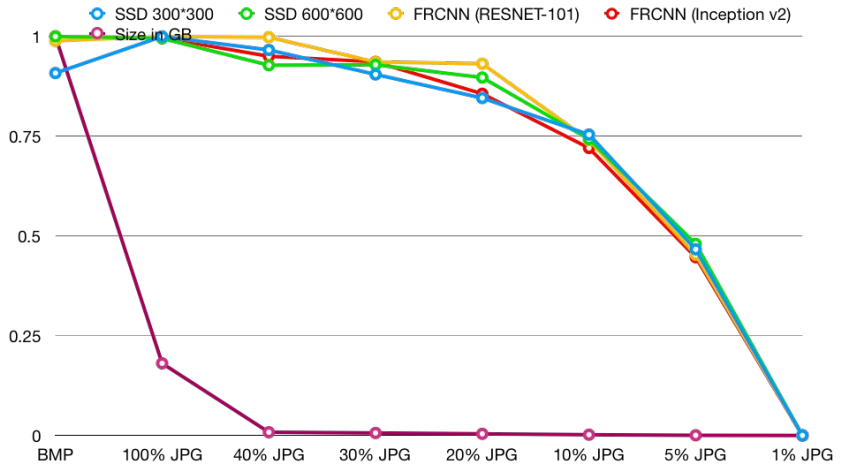
Even though pedestrians are more in number, they are small, so detected worse, than truck.

Can you still see me? - Results



Can you still see me? - Summary

While 40% and 30% jpg images show no significant difference in **precision** as compared to 100%, they bring huge drop in **storage size**.



Can you still see me? - Future work

- Research of other compression techniques, especially compression by neural nets, that preserve better features in the context of object detection;
- More experiments and more fine-tuning of the models
- Cross-validation
- Developing an NN architecture and training from scratch with more data
- Bigger batch size
- Uniform batch creation / Obtaining more data for classes like “Motorbike” and “Bus”

So that was it.
Thank you for coming.

d.davletshina@campus.lmu.de
hitansh.singla@campus.lmu.de

Can you still see me? - Reference

100% JPG


PassengerCar: 97%
PassengerCar: 73%
PassengerCar: 86%
PassengerCar: 92%
Van: 69%

40% JPG


PassengerCar: 96%
PassengerCar: 57%
PassengerCar: 92%
PassengerCar: 91%
Van: 63%

20% JPG


PassengerCar: 98%
PassengerCar: 74%
PassengerCar: 59%
PassengerCar: 63%

10% JPG

5% JPG