Object detection of cyclists and helmets

Oskar Bohn Lassen (s1555827), Mads Thorstein Roar Henriksen (s153265),
Christoffer Riis (Project supervisor), Frederik Hüttel (Project supervisor), Ole Winther (Professor, course responsible)
Technical University of Denmark

Introduction

Using convolutional neural networks and object detection models we can enable traffic safety campaigns to be evaluated accurately.

Key technical objectives:

- Process raw video-file
- Create an efficient labelling algorithm
- Implement pretrained YOLO3 object detection model [3]
- Create baseline CNN classification model
- Improve pretrained YOLO3 using transfer learning

Model	Cyclist detection	Helmet detection
YOLOv3	Object detection for Person + Bike	None
YOLOv3 /w transfer learning	Object detection for Cyclist	Objection detection for Helmet + Hovding
Helm-Net	Classifier for Cyclist	Classifier for Helmet

Video processing

Video
- 86 RGB-videos
- 1920x2560 (res)
- 30 FPS

Preprocessing
- Upper ¼ is croped
- Downscaled to 128x128
- 10 FPS

Labelling algorithm

Current situation
- No labelled data
- More than 1 million frames in total
- Under 5% of the frames has cyclists

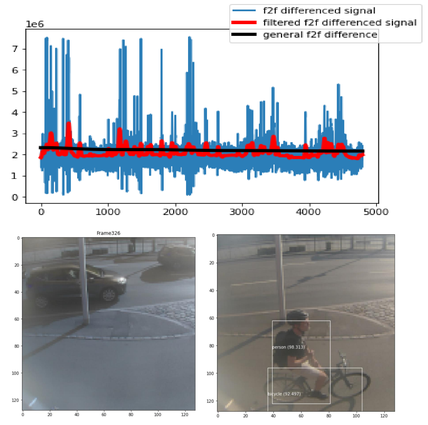
Algorithm

Frame-to-frame difference

Gaussian filter to reduce frames

YOLO3 object detection model

Images with bikes or persons

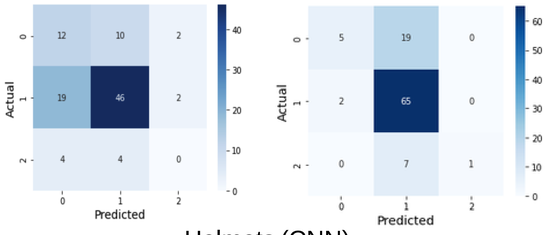


Baseline performance

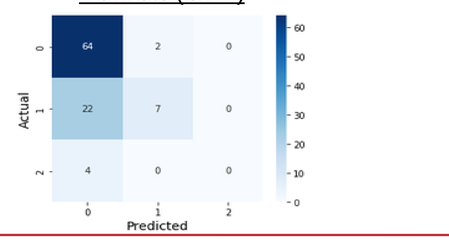
Baseline convolutional neural network (CNN) called “helm-net” as well as pretrained YOLO3 evaluated on labelled data.

Pretrained YOLO3 — CNN (“Helm-net”)

Cyclists



Helmets (CNN)

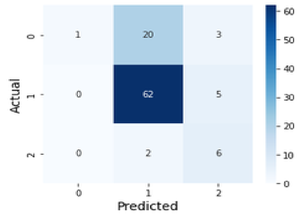


Improving performance

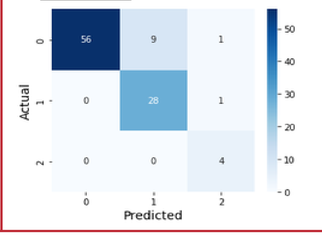
Methods for improving performance of baseline.

Transfer learning
- YOLO3 got tailored detection layers to our problem
- Model trained based on labelled data and new detection layers

Cyclists



Helmets



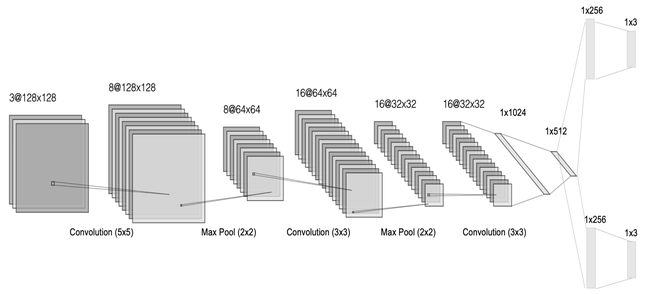
Next steps

Use active learning
- Investigate confidence of model predictions
- Label most uncertain frames
- Compare performance

Model

[4]

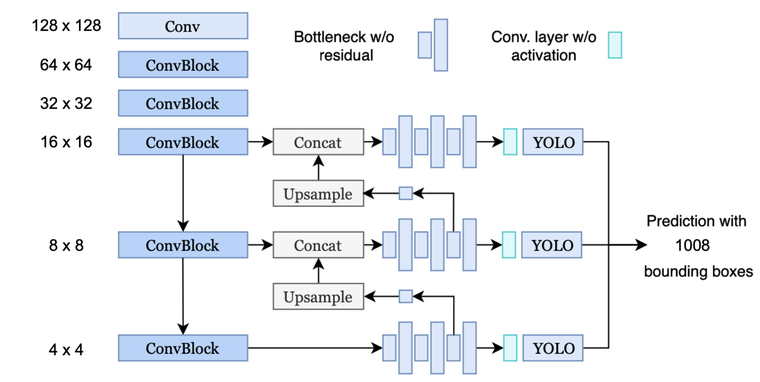
Convolutional Neural Network (“Helm-Net”)



- Reduce image to feature space using convolutions and max pool
- Two separate outputs: (0,1,2) cyclist, (0,1,2) helmets
- Binary encoded with softmax output

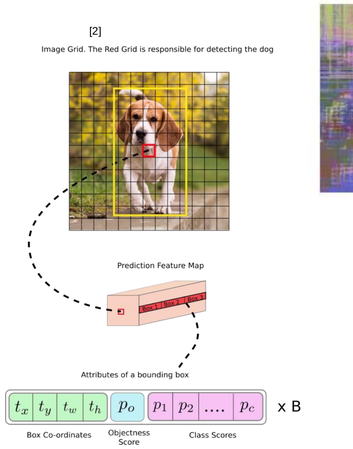
YOLO3 Object Detection

[1]



- Reduce image to feature space in three different scales
- For each scale object detection is made

[2]



- For each cell in each scale predict three bounding boxes

References

[1] Christoffer Riis, “Msc. thesis: Multi-modal engagement prediction on instagram,” 2020

[2] Ayoosh Kathuria, “How to implement a yolo (v3) object detector from scratch in pytorch: Part 1,” 2017.

[3] Joseph Redmon and Ali Farhadi, “Yolov3: An incremen-tal improvement,” 2018.

[4] http://neuralnetworksanddeeplearning.com/chap6.html

[5] Joseph Redmon and Ali Farhadi, “Yolo9000: Better,faster, stronger,” 2016.