Lucerne University of
Applied Sciences and Arts
HOCHSCHULE
LUZERN

NN-based Person Detection using Transfer Learning with Synthetic Images

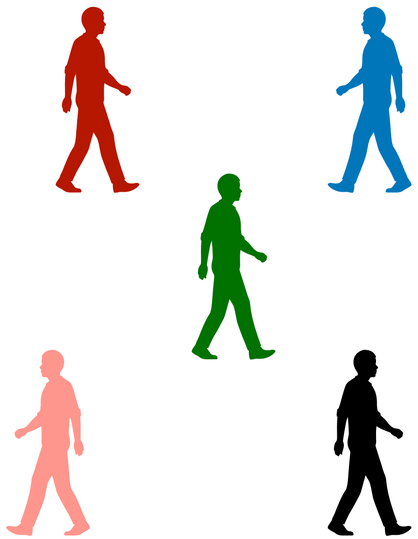
Master Thesis

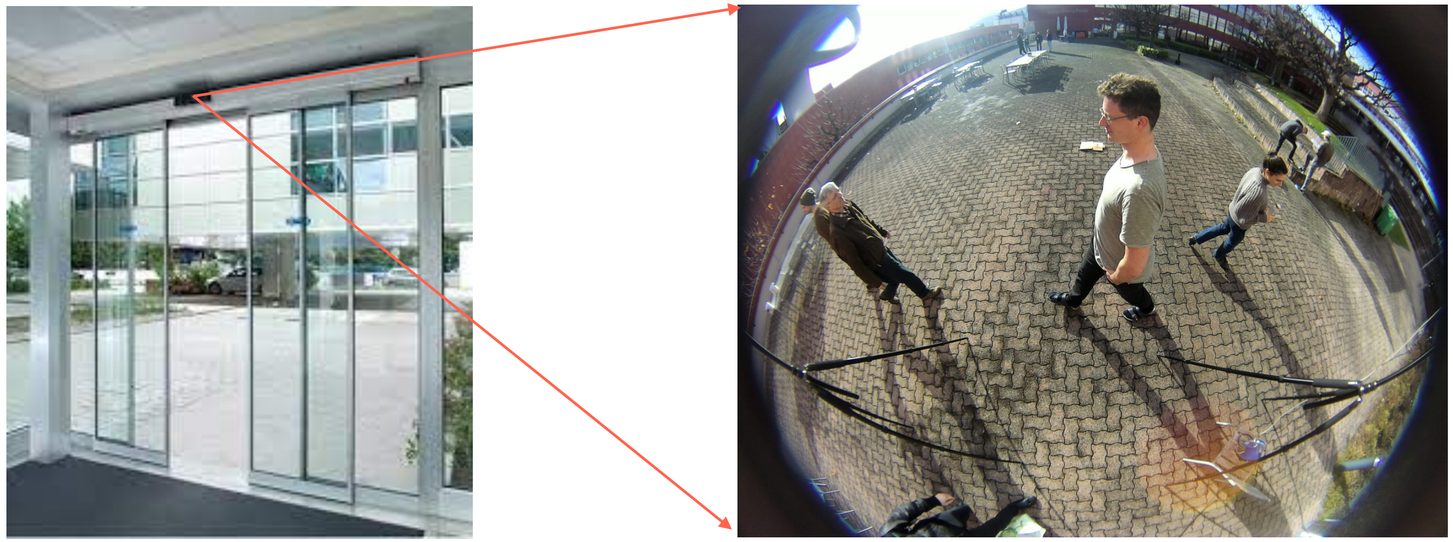
Odysseas Liagouris

odysseas.liagouris@stud.hslu.ch

Luzern, January 2023

Motivating Use Case



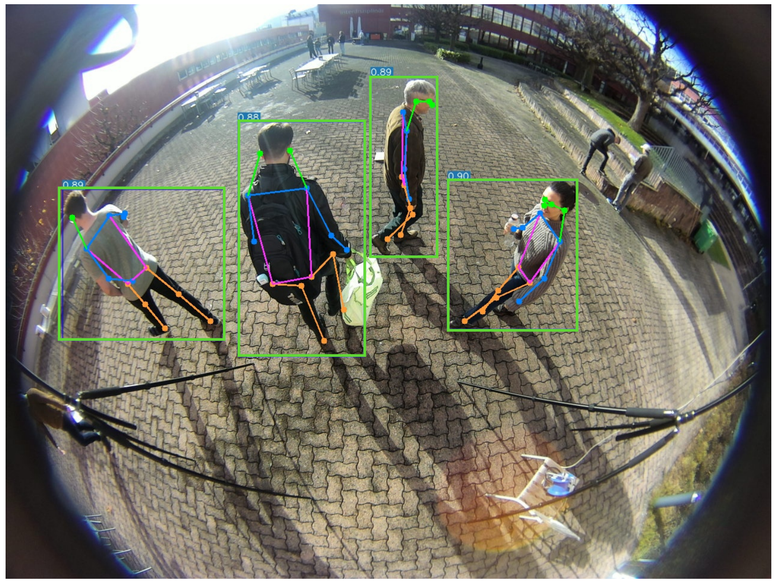


ICARUS: Intelligent Camera and Radar fUsion Sensor, CC Innovation in Intelligent Multimedia Sensor Networks, HSLU

Goal of Master Thesis

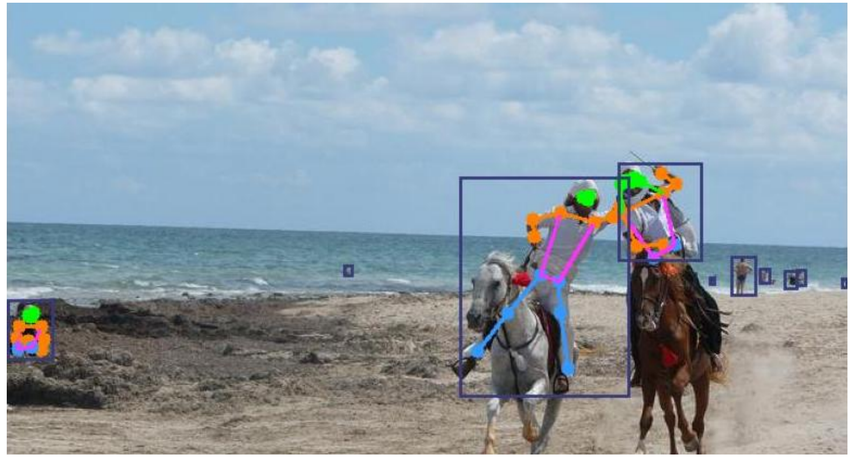
Person Detector

(CNN Model)

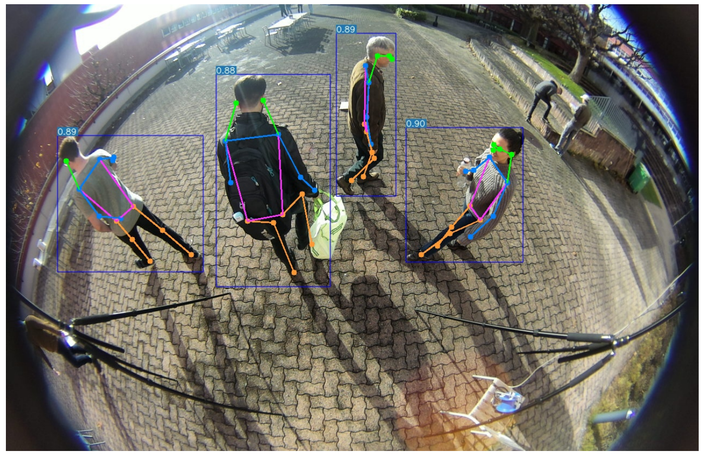


Detect persons in fisheye image frames with high accuracy

Technical Challenges



- Vast majority of person detection models are trained on front-view images
- Front-view images are not a good fit for top-view omnidirectional person detection



- Real fisheye datasets are scarce
- Manual labelling is time consuming

Overview of our approach

Transfer Lea

Real Images

- Use prior knowledge o
- Mitigate problem of Ca
  Interference

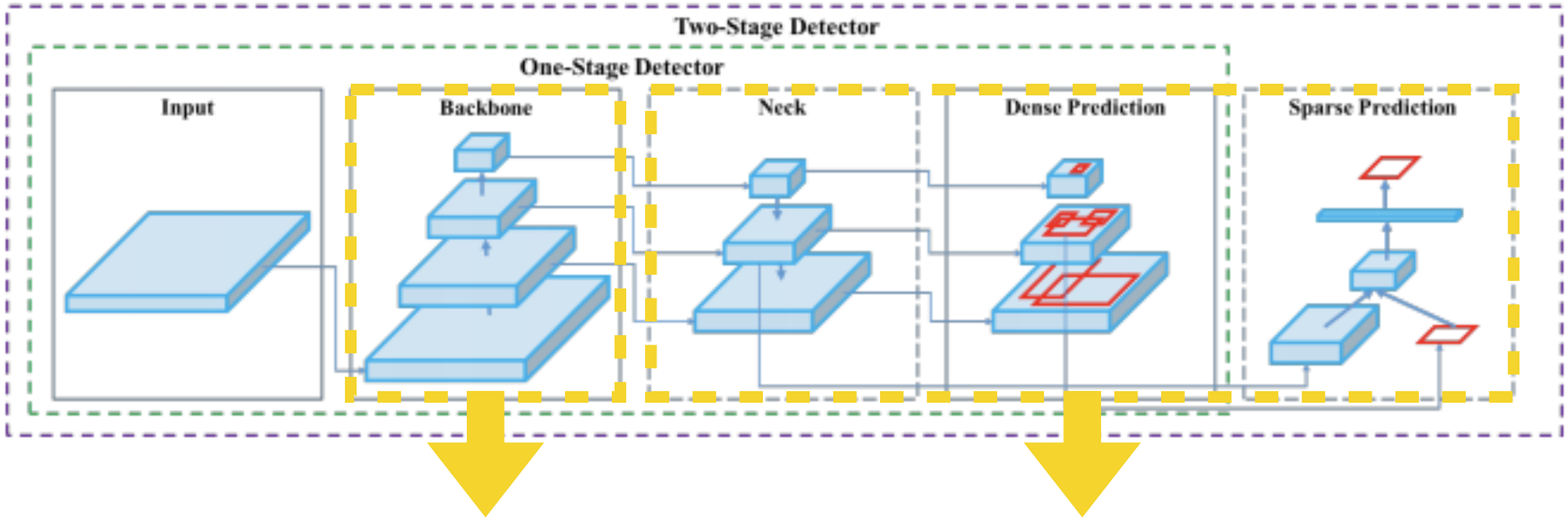
erated-images with
g
p between synthetic

Using our approach, we can improve:

- performance compared to the pre-trained model by 23%
- performance compared to prior work by 16%

Transfer Learning

YOLO V5 CNN model architecture



Backbone of the model remains untrained

Only the rest of the model layers are trained

YOLO V5 model repository: https://github.com/ultralytics/yolov5

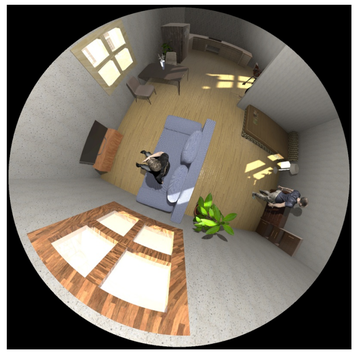
Training Datasets

Collected and annotated as part of the thesis

MS COCO[1]

- 64.000 images
- 220.000 persons
- 1-19 persons/image

THEODORE[2]



- 100.000 synthetic images
- 350.000 persons
- 0-4 persons/image

ICARUS

- 7.000 images
- 12.000 persons
- 1-7 persons/image

[1] Publicly available dataset created by Microsoft
[2] Provided by our academic collaborators at the University of Chemnitz

Test Datasets

MS COCO

- 2.300 images
- 9.500 persons
- 1-13 persons/image

FES

- 301 images
- 2.000 persons
- 1-8 persons/image

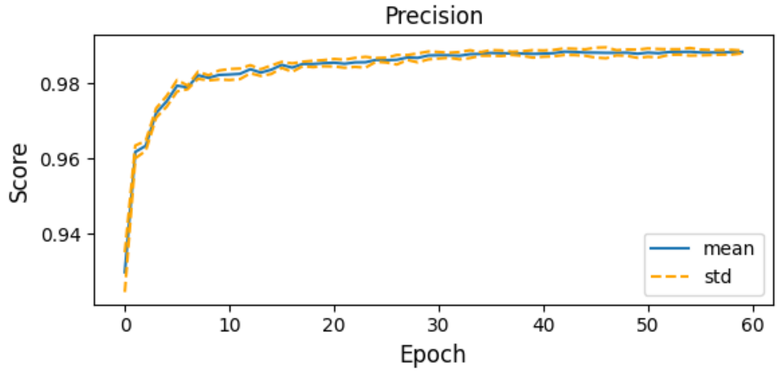
ICARUS

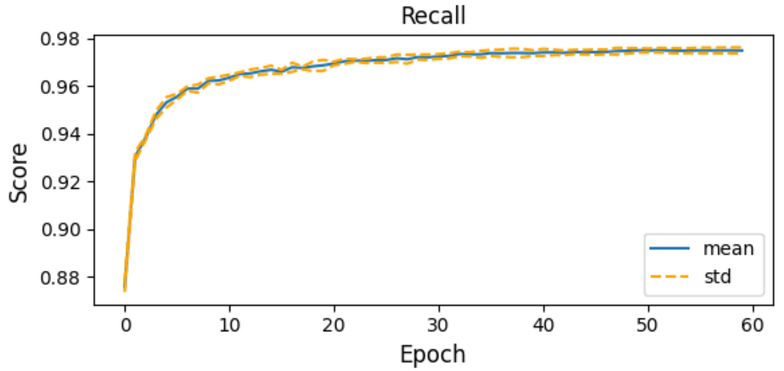
- 111 images
- 280 persons
- 1-4 persons/images

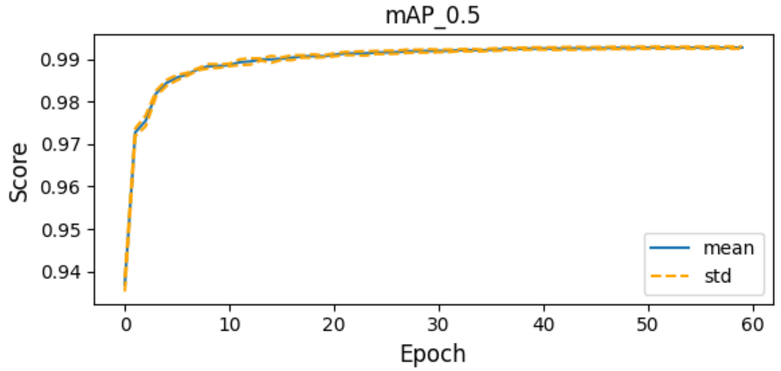
Which pre-trained layers to reuse?

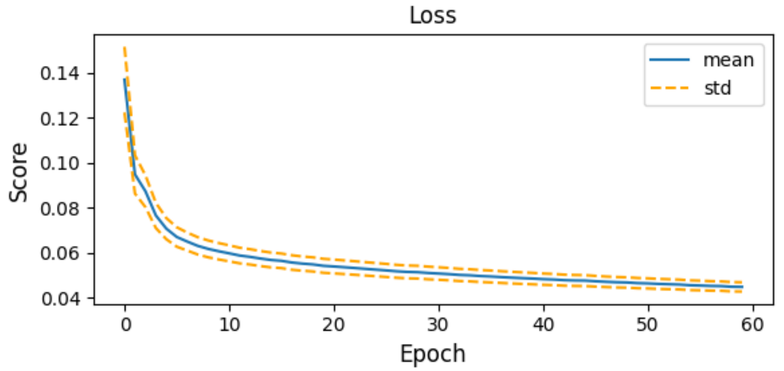
Model	Average Precision on FES
Yolo V5 from scratch (low augmentation)	0.672
Yolo V5 from scratch (medium augmentation)	0.813
Yolo V5 from scratch (high augmentation)	0.811
Yolo V5 Backbone frozen (low augmentation)	0.902
Yolo V5 Backbone frozen (medium augmentation)	**0.917**
Yolo V5 Backbone frozen (high augmentation)	**0.918**
Yolo V5 9 layers frozen (low augmentation)	0.901
Yolo V5 9 layers frozen (medium augmentation)	0.91
Yolo V5 9 layers frozen (high augmentation)	0.91
Yolo V5 8 layers frozen (low augmentation)	0.86
Yolo V5 8 layers frozen (medium augmentation)	0.905
Yolo V5 8 layers frozen (high augmentation)	0.905

K-fold cross validation









The resulting model does not deviate from the mean of each metric

Mixing synthetic and real images

Datasets	# Images		
	MS COCO training set	THEODORE training set	ICARUS training set
Dataset 1	64k	14k	7k
Dataset 2	64k	0	7k
Dataset 3	7k	7k	7k
Dataset 4	14k	7k	7k
Dataset 5	64k	7k	0
Dataset 6	64k	7k	7k
Dataset 7	0	25k	0
Dataset 8	0	7k	7k

Model accuracy

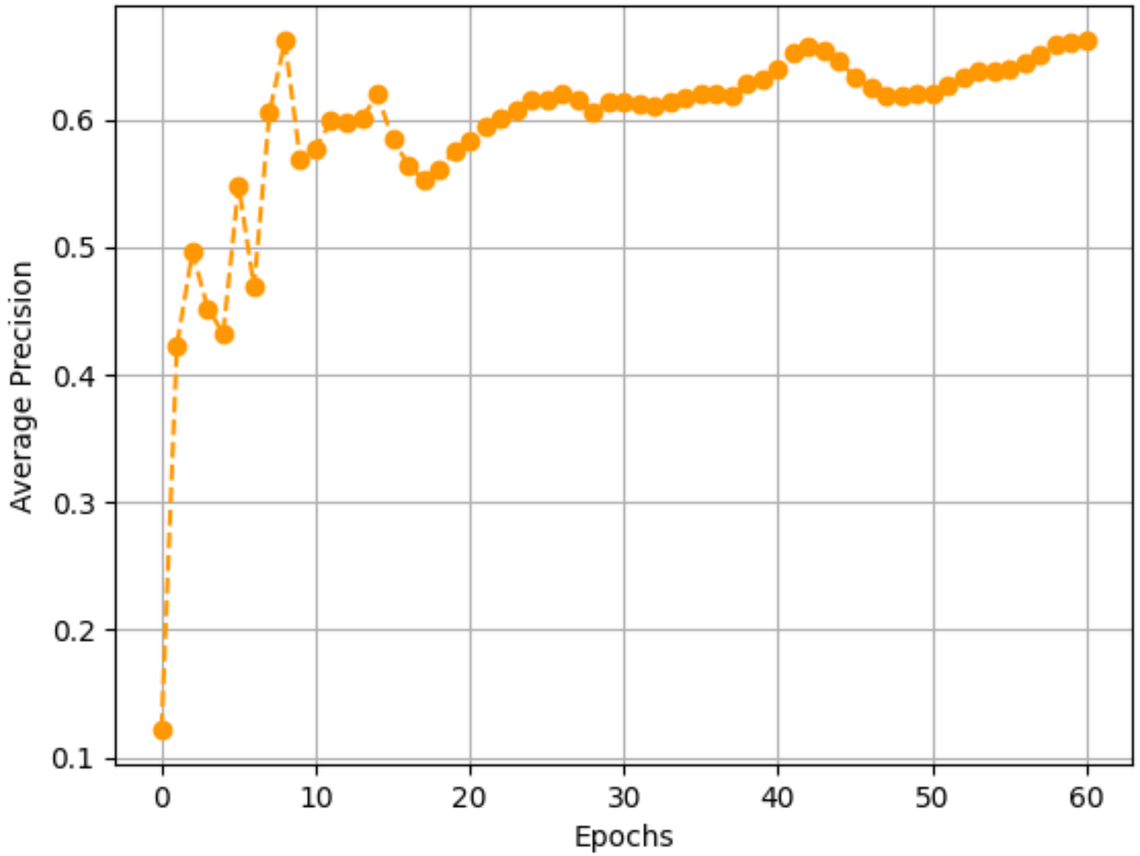
Datasets	Average Precision		
	MS COCO test set	FES test set	ICARUS test set
Dataset 1	0.731	0.936	0.959
Dataset 2	0.729	0.809	0.947
Dataset 3	0.688	0.941	0.963
Dataset 4	0.711	0.927	0.945
Dataset 5	0.728	0.928	0.829
Dataset 6	0.73	0.929	**0.966**
Dataset 7			0.684
Dataset 8			0.943

We achieve the best performance using 64k from MS COCO, 7k from THEODORE, and 7k from ICARUS

Comparison with prior work

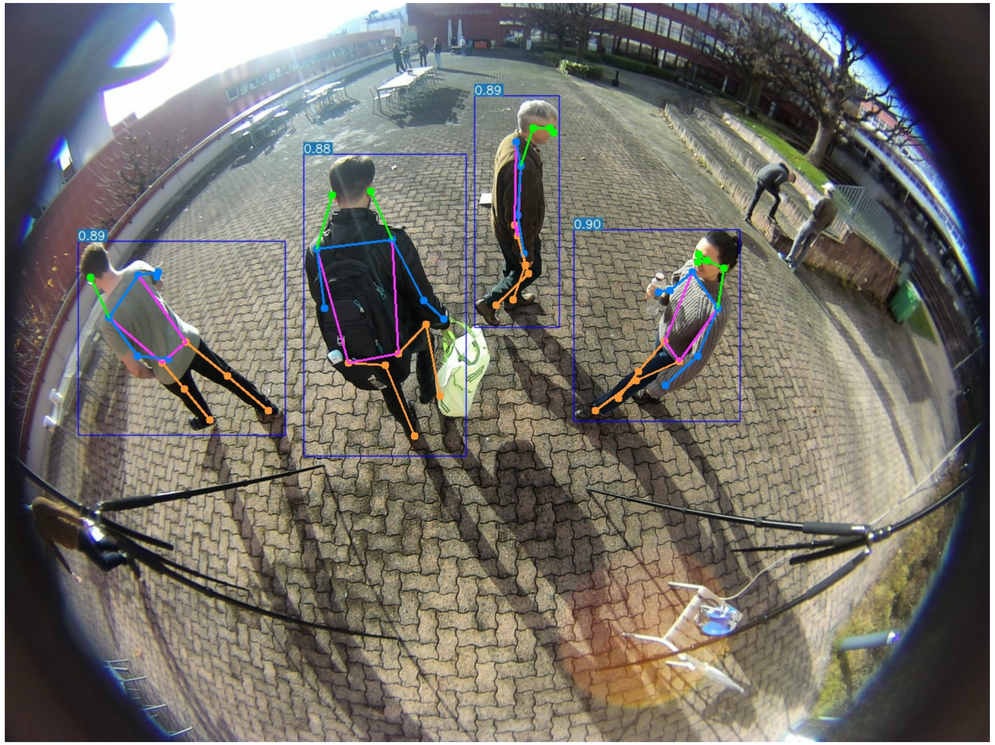
Models	Average precision on BOMNI test set
SSD model trained on MS COCO and THEODORE	0.579[1]
YOLO V5 (our model)	0.662

YOLO V5 performance on BOMNI



[1] T. Scheck , R. Seidel , G. Hirtz, *Learning from THEODORE: A Synthetic Omnidirectional Top-View Indoor Dataset for Deep Transfer Learning*. IEEE Winter Conference on Applications of Computer Vision (WACV), 2020.

Future work



- Improve pose estimation in omnidirectional images by training with a mixture of real and synthetic data
- Detect whether a person is approaching or just passing by in order to trigger the door accordingly