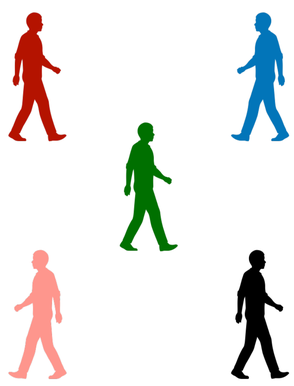
HSLU Hochschule Luzern

Technik & Architektur
Master of Science in Engineering Electrical Engineering

Master Thesis

NN-Based Person Detection using Transfer Learning with Synthetic Images

Motivating Use Case



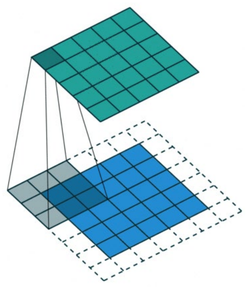
Goal Of Master Thesis

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 omnidirtectional person detection
- Real fisheye datasets are scarce
- Manual labelling is time consuming

Convolution Neural Networks

- Can extract features from images automatically
- Ideal for object detection, classification, image processing tasks
- The receptieve field is used to create a feature map
- In the thesis we use state-of-the-art YOLO V5 CNN model



Overview of our Approach

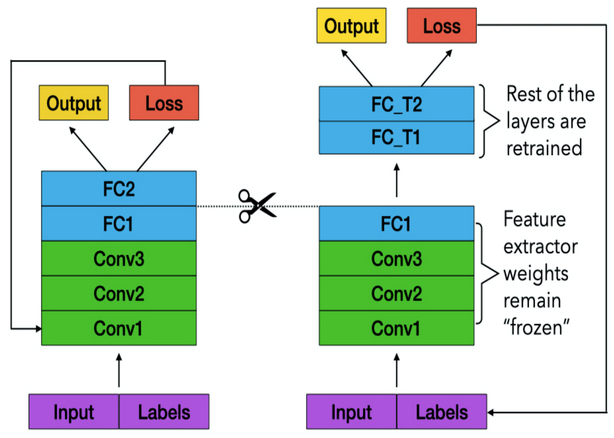
Transfer Learning

- Use prior knowledge of pre-trained model
- Mitigate problem of Catastrophic Interference

Synthetic + Real Images

- Use software-generated images with automatic labelling
- Bridge domain gap between synthetic and real images

Transfer Learning



Training Dataset

MS COCO

- 64.000 images

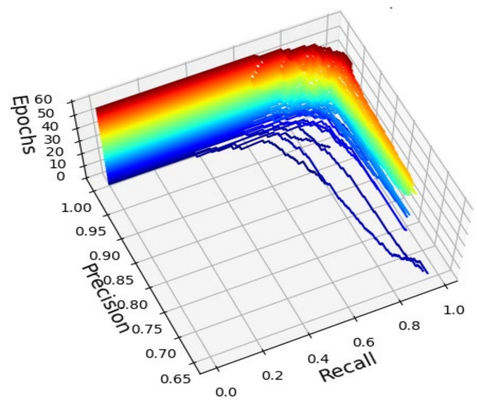
THEODORE

- 7.000 images

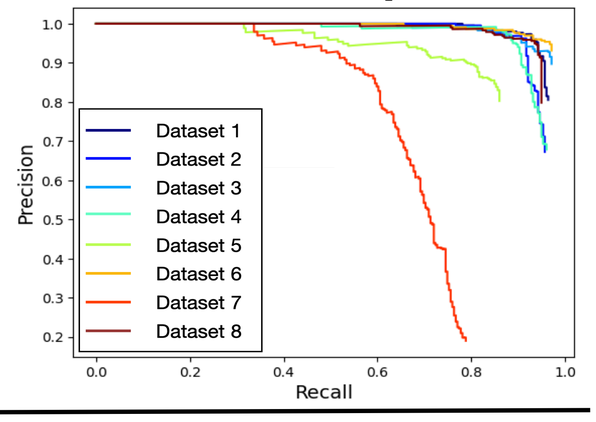
ICARUS

- 7.000 images

Precision-Recall Curves



Models Comparison



Person Detection in Fisheye Images

The majority of person detection models are trained on front-view images. However, when it comes to person detection in omnidirectional images, front-view images are not a good fit for improving the model's accuracy.

Synthetic 3D images of indoor scenes have been an increasing interest of research for computer vision related tasks. Thousands of images can be artificially generated and rendered to resemble the interior of an indoor room. The advantage of this method is that the labels are automatically generated and can, thus, eliminate the need for human labelling.

However, training only with synthetic data does not yield good results due to the domain gap between real and synthetic images. To tackle this problem, one can leverage prior knowledge gained by training the model on real and synthetic data using *Transfer Learning*. However, this technique requires extra care, as it may lead to *Catastrophic Interference.*

Core Idea of our Solution

We train a state-of-the-art YOLO V5 model to detect person in omnidirectional images. To improve model accuracy in real use cases, we use *Transfer Learning* and train certain layers of the model on a carefully selected mixture of synthetic and real images. We show that this approach mitigates the problem of *Catastrophic Interference* and bridges the domain gap between real and synthetic data.

Thesis Contributions

- We use a fisheye camera module to collect 7k real images for human pose estimation, person detection, and segmentation.
- We manually label the collected images with bounding box annotations that can be used for model training.
- We experimentally show that our approach improves model accuracy by 23% compared to baseline solutions and 14% compared to prior work on person detection in real omnidirectional images.

Real Use Cases

Our model could be used in an intelligent door control system for person detection and pose estimation. In this setting, the system can use the real-time model predictions on the fisheye camera frames to automatically control the door functions.

Odysseas Liagouris
odyliagouris@gmail.com

Adviser
Prof. Dr. Klaus Zahn

Expert
Marco Hauri

Cooperating Partner
Agtatec AG

FH Zentralschweiz

MSE | MASTER OF SCIENCE IN ENGINEERING