# Helmet Detection Using YOLO for Construction Site Safety
# Project findings

March 24, 2025

Zarin Karanikolov

---

## 1 Objective

This project aims to identify whether individuals in construction site images are wearing safety helmets. The system enhances worker safety by automating helmet compliance monitoring, reducing manual supervision, and preventing workplace injuries.

## 2 Methodology

### 2.1 Model Selection

Fine-tuning a pre-trained YOLOv5 model using transfer learning.
Using YOLOv5, because it has fast inference speed, suitable for real-time detection and is highly optimized for object detection with small and medium datasets.

### 2.2 Steps Involved

1. Images and Annotations split in train and test and then converting annotations to YOLO format
2. Model Training & Fine-Tuning
3. Model Evaluation
   Measuring mean Average Precision, Precision, and Recall.
   Testing performance on unseen images.

## 3 Results and evaluation

Every result(images and videos) and the evaluation(graphs of curves and confusion matrices) can be found in the folders "evaluation" and "results" inside the directory of the project.
After few tries, I concluded, that I cannot run the whole dataset, due to size, resources

and time. The solution I did, was to use only 701 images for the training. The model YOLOv5m, which is more stable and accurate had issues with my RAM, so I then changed it to the faster and lighter version - YOLOv5s. I tried two trainings, using 2 different image sizes, which are shown in the table below - 640x640 and 960x960 and compared the outcomes. Everything else is the same - 50 epochs, 8 batch size(any higher and RAM issue), device cpu(because otherwise CUDA/RAM issues).

### 3.1 yolov5s training

| | | |
|---|---|---|
| first run | 640x640 | duration: 1.55 hours |
| second run | 960x960 | duration: 4.133 hours |

According to the results of the first run, mAP@0.5 of 0.926 and 0.948 for respectively helmet and head is achieved, which is more than the initial target of 0.80. The model has issues with identifying people, even though it is almost perfect identifying human heads, which is why I focus only on heads and helmets. The F1 score is around 0.85 for these 2 classes.

The model was tried on unseen images and videos and it did very well, which could be seen in the results folder of my project.

According to the results of the second run, mAP@0.5 of 0.935 and 0.937 for respectively helmet and head is achieved, which is again more than the initial target of 0.80. The F1 score is a little bit higher for these 2 classes.

The model was tried on unseen images and videos and it did very well again.

It can be concluded from the results, that the model with higher resolution is a bit better of detecting helmets, which was the initial target, but is worse at detecting heads.

## 4 Reflection on challenges and possible improvements)

The main challenge I faced during this project was the lack of computing power. My limited RAM and not having access to a GPU acceleration meant I had to use a smaller dataset and lower model settings, which affected the performance of the model. I couldn't fully explore the best hyperparameters because of these hardware restrictions and the long time of training.

For future improvements, I would aim to use a larger dataset with more diverse images to make the model stronger. Having access to better hardware would allow me to train the model with higher image resolutions, editing the hyper parameters in the yaml files in the directory of the YOLO model and eventually use more recent version. I also want to test the model in real-time using a live camera feed. By tuning hyperparameters like the learning rate and confidence thresholds, I could improve the accuracy and reduce false positives.