# Overview

In the scenario given, I have a video sample ("sample.mp4") from a 3D sensor capturing a person wearing a staff name tag walking along a corridor. The tasks assigned to me are as follows:

1. Identify Frames with Staff Present:

My primary task is to analyse the video frames and identify the frames where a person wearing the staff name tag is present. This involves implementing a computer vision algorithm or using pre-trained models to recognize the staff name tag in each frame.

1. [Bonus] Locate Staff XY Coordinates:

As a bonus task, I am asked to locate the XY coordinates of the staff when present in the video clip. This involves determining the position of the staff member within each frame.

# Object Detection Workflow and usage

Here is my solution workflow and the usage of it:

Before start using the model, it is required to clone the repo.



Go the file location.



Create the virtual environment.



Activate it.



Install all the requirement.



## Extract frame

First, I have to extract the frame out of the video in order for me to do all the process afterwards.



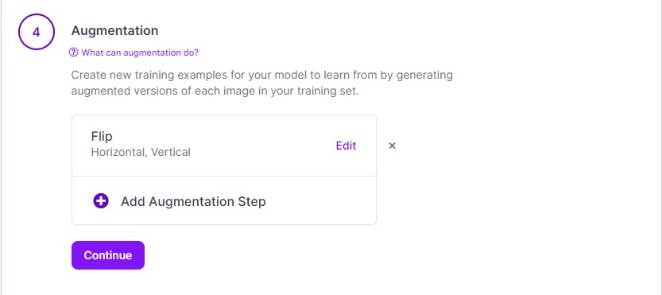
## Data annotation

Next step I am doing is the data annotation after extracting all the frame from the video. I have 53 images for me to annotate whether the person is staff or not and the tool that I am using is Roboflow which is an online platform suitable for data annotation task.



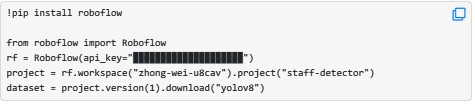
## Data augmentation

After I have manually annotated all of the data, I have to increase the size of my dataset in order for me to train my own model. Roboflow provided service such as data augmentation before we generate the dataset and I am using techniques such as: vertically flip and horizontal flip.



## Export dataset

Next after all of the data preprocessing step, the next step for me is to export the dataset and it is ready for the model training step.



## Train using custom dataset on yolov8

The model I have choose for this task is [yolov8](https://github.com/ultralytics/ultralytics) and the reason that I have choose this model instead of other is that this model is famous for the multiple object detection and is useful to detect the multiple classes. I am using the [google colab](https://colab.research.google.com/drive/1Bbctd2_IFzEevqcyd11mhvBrQXkWdq4P?usp=sharing) for the training purpose due to the fact that they are providing the GPU which may accelerate the training process. I have run with different epochs and also resizing the image to 800\*800 and I am also setting the plot=True so that It will save plots and images during training.

### Expand dataset and iterate the training.

The dataset size is still too small in order to achieve the task, so I decided to use active learning after I have trained my first yolov8 model. I am using the self-trained model and using the auto-Annotator from this [repo](https://github.com/naseemap47/autoAnnoter). After expanding the dataset size, I keep training the model using from the previous one.

## Inference the model on video.

After training the model, I have used the model to inference it on the “sample.mp4” video and I am able to achieve the bonus task by adding the coordinates in the frame only when the person is being detect as “Staff”.



# Results

After Interference with the video, the result video is being saved in your local device in .mp4 format. Here is some of the examples that the model successfully detects the staff with the input video being Interference.

