**1.0 Introduction**

This documentation outlines the process of processing videos captured by 3D sensor to identify staff members wearing name tags using the tiny YOLOv7 object detection model. I used 2 separate models: one for person detection and another for name tag detection. First, the person detection model locates individuals in the image, drawing green bounding boxes around them. Next, the name tag detection model identifies name tags within these bounding boxes, marking them with red bounding boxes. Valid staff identifications are determined by checking if the name tag bounding box is contained within the person bounding box. Only detections meeting this criterion are considered valid. Finally, the detections are visualized by drawing bounding boxes on the image, enabling clear identification of staff members wearing name tags.

**2.0 Data Acquisition and Preprocessing**

**2.1 Frame Extraction**

Individual frames are extracted from the sample video (“sample.mp4”) using the OpenCV library. With the sample video having a duration of 53 seconds and a frame rate of 25 frames per second (fps), the total number of frames extracted amounts to 1325 frames. Each of these frames is saved as a separate image file in the JPEG format (.jpg).

**2.2 Annotation**

The images are annotated using LabelImg tool to mark the precise location of the name tag with class name on each frame. Annotations are saved in a label data file format named "~.txt", corresponding to the image name.

**2.3 Augmentation of the frames**

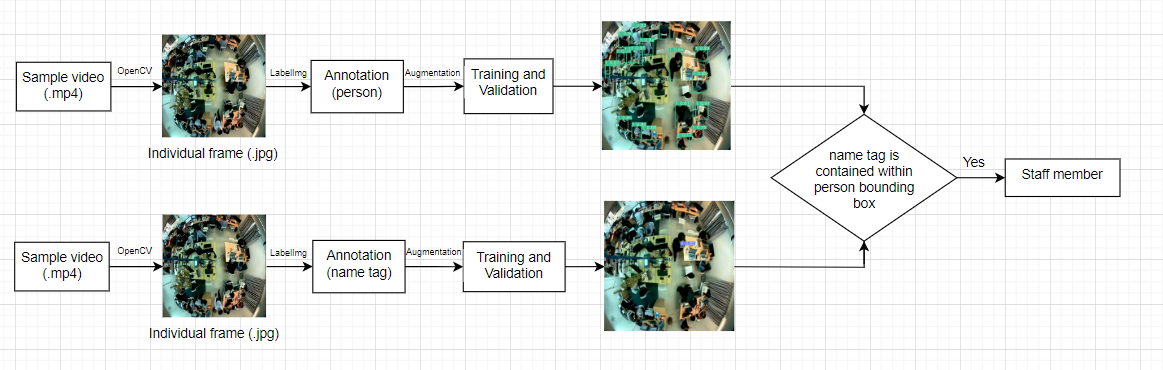
Two transformations are applied to enhance dataset diversity. Horizontal and vertical flips are implemented, along with rotations of 90° clockwise, 90° counter-clockwise, and complete upside-down flips. These augmentations contribute to enriching the dataset by providing mirrored and rotated versions of the original frames

**2.4 Dataset Splitting**

The total images of name tag dataset was 1759 images and was split into train (1357 images), valid (268 images) and test (134 images). The total images of person dataset was 5996 images and was split into train (2799 images), valid (266 images) and test (132 images).

**3.0 Model Selection and Training**

Tiny YOLOv7 architecture was selected due to its suitability for real-time applications and resource-constrained devices, offering rapid inference speeds and efficient object detection and classification. Transfer learning was employed using pretrained weights from Tiny YOLOv7 to enhance model convergence and performance. Training configurations included an image size of 640 pixels, 200 epochs for the name tag detection model and 50 epochs for the person detection model. The models were trained using NVIDIA GeForce RTX 3060 Ti.



**4.0 Evaluation Metrics**

Evaluation metrics such as precision, mean average precision (mAP), and recall are employed to assess model performance. Visualizing model performance is facilitated through various graphs, including the confusion matrix, F1 curve, precision-recall curve, precision curve, and recall curve.

**Person detection model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Total Testing Images** | **Precision** | **Recall** | **mAP** |
| Person | 132 | 0.978 | 0.959 | 0.977 |

**Name tag detection model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Total Testing Images** | **Precision** | **Recall** | **mAP** |
| Name Tag | 134 | 0.829 | 0.853 | 0.861 |

**XY coordinates of the staff**

YOLO models detect objects in an image and provide bounding box coordinates for each detected object. These coordinates are in the format (x\_min, y\_min, x\_max, y\_max), where (x\_min, y\_min) represent the top-left corner of the bounding box, and (x\_max, y\_max) represent the bottom-right corner.



**Github Repository:**

**Tasks:**

**1. Identify which frames in the clip have the staff present?**

The output video showed the identified staff with name tag using bounding boxes.

**2. [Bonus] Locate the staff xy coordinates when present in the clip.**

The xy coordinates of the identified staff was showed in the bottom left side of the video.