Training and Inference Report on Person and PPE Detection Models

1. Introduction

In this report, it is detailed how the Person and Personal Protective Equipment (PPE) detection models were trained and how they may be used to make inferences. The steps followed in converting the annotations, preparing the dataset, training the models, and making inferences to determine if the trained model is efficient are described.

2. Dataset Preparation

a. Converting Annotations: Pascal VOC to YOLO Format

Script: pascalVOC_to_yolo.py

Importance: The backbone of training the models is the annotations from the images. Thus, the pascal VOC annotation format which saved the file as XML files had to be changed to YOLO format, which was saved in TXT files as this was suitable for YOLOv8 to train on.

How it works: The script opens the XML files that contained the objects' position in the images – persons, hardhats, and gloves – and then normalized all the boxes relative to the images such that they all hold on to a standard value irrespective of the image dimension. Then it saves the converted annotations to a new directory, one file for each image.

b. Splitting the Dataset

Objective: Ensure that the models train on a distinct dataset and validate and test them on distinct datasets as well, while the distribution between each class in the datasets is equal.

How it works:

Split the dataset randomly, where 80% is used to train, 10% to validate, and the remaining 10% to test.

The class distribution in each separate dataset is equal.

3.Model Training

Two models were trained, and each had a distinct objective to perform, as follows:

Person Detection Model

Objective: Detect a person exists in the images.

Demand: requiring to train on the whole images.

PPE Detection Model

Objective: after the person is detected, detect hardhats, gloves, and face masks on the person.

Model: This model was created by training image regions that contain people Training

Training Process:

Both models were trained on YOLOv8 We provided the images and their annotations as an input to models for them learn how they have to identify and classify objects at training time.

The weights of the best performing model were saved, in case they wanted to use this specific neural network for future predictions.

4. Inference

Script: `inference.py`

Aim: Test the trained models on new, unseen images to know how well is they able to detect among persons and PPE.

How it Works:

Person Detection Inference on Test Images.

Next, it zeroes down on the detected persons and again runs a secondary inference to inspect for PPE items within those locations

Results: Save the results.

5. Validation

Objective: This is utilized to evaluate the performance of models on unseen data.

How It Works: Models that are assessed with precision, recall and mAP (mean Average Precision) to understand their detection capabilities.

6. Execution Summary

1. Conversion of Annotation: Conversion from xml to txt for bounding boxes

2. Preprocessing: Training / Validation/ Test split of the dataset

3. Developing Model training: Develop special models for detecting person and PPE detections.

4. Inference Testing the models on new images to see how well they perform

5. Validation — testing the model using evaluation metrics.

7. Conclusion

The model building of the person and PPE detection were also a success. Inference results show that these models are good at person and PPE detections, making them ideal choices for safety surveillance in places where wearing PPE is mandatory.

This allows the process, which is described in this report to everyone who has a basic understanding of it even understandable and reproducible.