**Evaluation and Justification of ONNX Model Outcomes**

This document highlights the points related to ONNX model inferencing/evaluation against 300 images (validation dataset) by comparing the predicted bounding boxes to the provided ground truth.

This evaluation has been done by varying the **iou\_threshold** and **conf\_threshold** parameters i.e., the test on the ONNX models has been carried out by choosing 6 pairs of the above parameters and observing and identifying which model among the provided 35 models is the best performing one.

The test scenario used to perform the evaluation is as given below,

**Test 1:**

iou\_threshold = 0.1

conf\_threshold = 0.4

**Result:** Best performing model on the validation dataset is **24.onnx** with an average accuracy of **0.5957** or **60%**

**Test 2:**

iou\_threshold = 0.3

conf\_threshold = 0.4

**Result**: Best performing model on the validation dataset is **24.onnx** with an average accuracy of **0.5890** or nearly **60%**

**Test 3:**

iou\_threshold = 0.4

conf\_threshold = 0.4

**Result**: Best performing model on the validation dataset is **24.onnx** with an average accuracy of **0.5824** or nearly **60%**

**Test 4:**

iou\_threshold = 0.4

conf\_threshold = 0.3

**Result:** Best performing model on the validation dataset is **24.onnx** with an average accuracy of **0.6624** or **66.24%**

**Test 5:**

iou\_threshold = 0.4

conf\_threshold = 0.35

**Result:** Best performing model on the validation dataset is **24.onnx** with an average accuracy of **0.6308** or **63.08%**

**Test 6:**

iou\_threshold = 0.4

conf\_threshold = 0.5

**Result:** Best performing model on the validation dataset is **24.onnx** with an average accuracy of **0.4394** or **43.94%**

As it is seen from above, in all the 6 tests, the best performing model was **24.onnx** with the highest average accuracy of **66.24%**. Other models outcome were also compared with this model’s result. It seems the other models were not able to detect some of the faces even if they were pretty clear and they were not detecting the faces with different orientations. Even if they were detecting the faces, the confidence score was pretty low i.e., around **20% - 30%**. Also, there were some false detections in all the models when **Test 4** was performed. But even then, **24.onnx** was the best performing model.

The evaluation has been implemented by calculating the IoU between predicted bounding boxes and ground truth data and comparing this IoU score with the IoU threshold to ensure the correct predictions and calculate the accuracy of all the ONNX models. And finally, based on the calculated best average accuracy, the best performing model will be finalized.