Report Extract – Project Testing & Deployment Results
Pages 62-70
[Click Here](#) to access full document

(5.3.2) F1-Confidence Curve

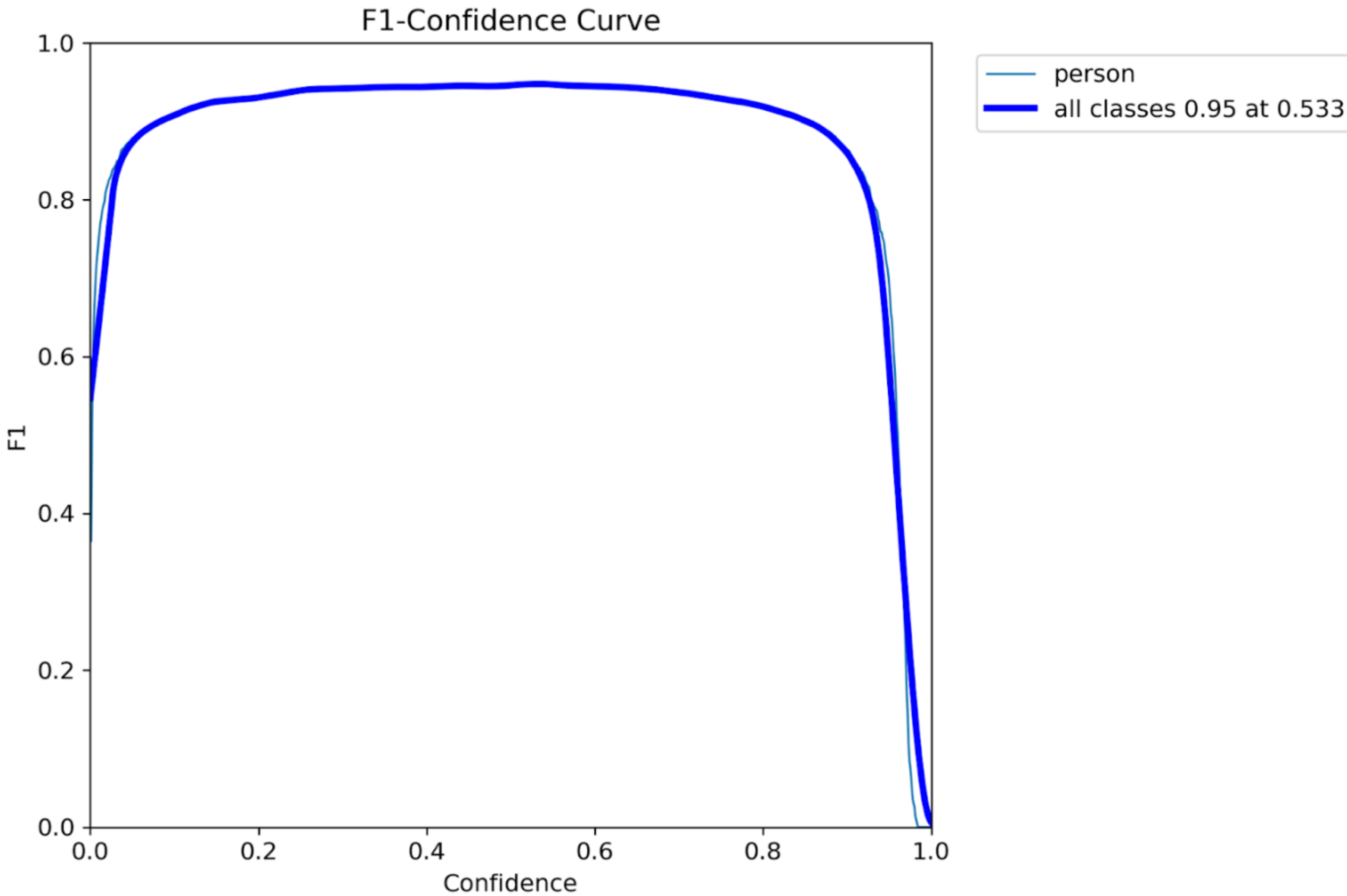


Figure 5:11 The confidence curve for the validation testing set results.

The F1 score is a metric to measure accuracy of a model in machine learning [8]. The above graph is made using the precision and recall values. The graph's shape comes from precision and recall. When the confidence is above 0.9, the F1 score plummets, as the precision remains high while the recall is significantly lower.

The model's F1 score being incredibly high for most of the above graph indicates the model performs very well, indicating a high level of accuracy, and that the object detection classifier works as intended. The model is 95% confident of its accuracy when F1 is around 0.533, however, this range extends from when the confidence is about 15% to 90%.

(5.3.3) Precision-Confidence-Recall Curves

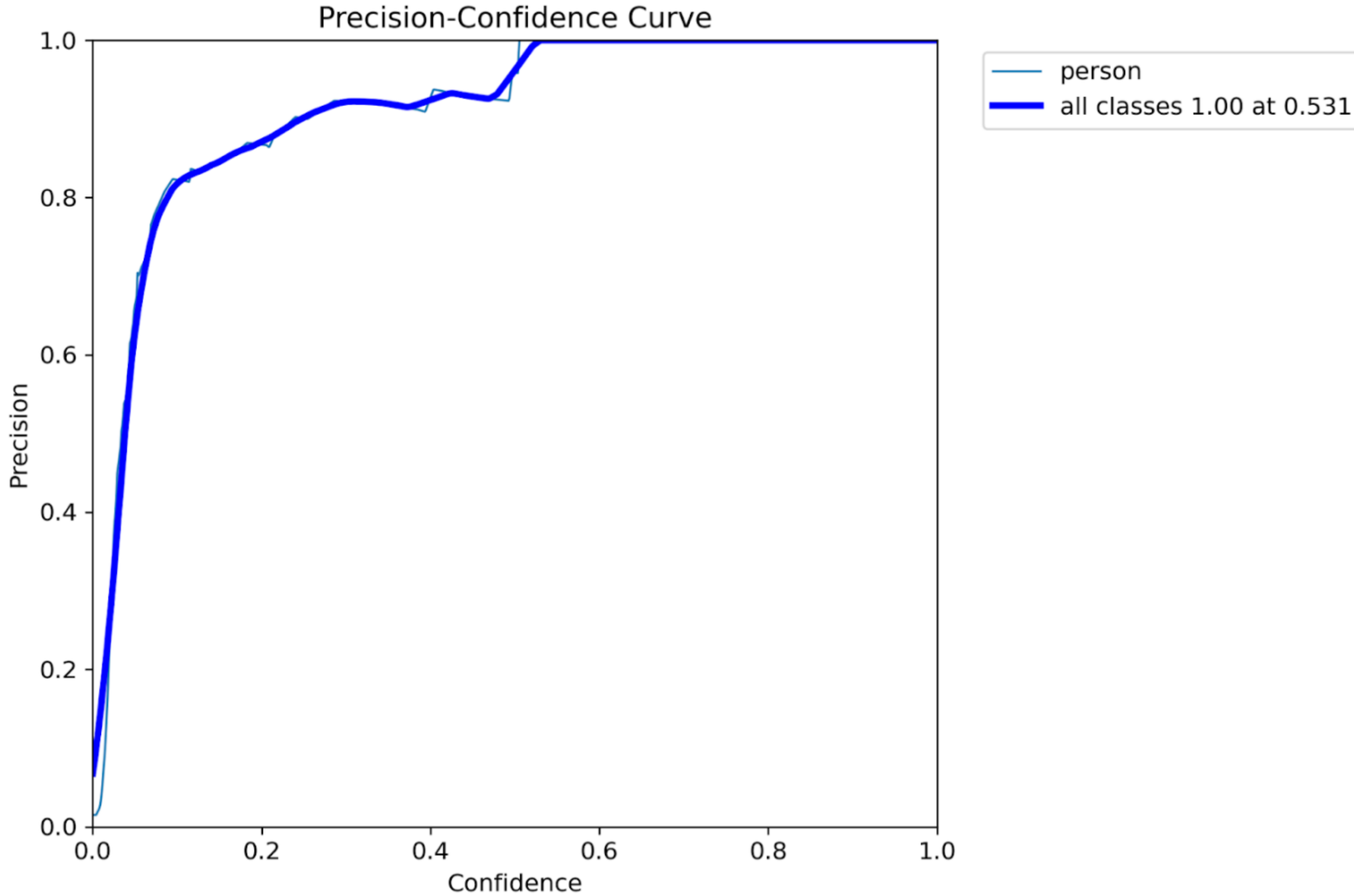


Figure 5.12: Precision-Confidence curve for the sample validation data.

This graph presents the comparison of the precision and confidence results. The graph shows that when the confidence is very low, at < 0.1, the precision is also very low. However, this rate grows incredibly quickly as the confidence rises, with precision reaching above 0.8 before confidence is even 0.2. What this means is that when the model is confident, it is precise. In layman's terms, the model is almost always correct when it believes it is correct. The precision reaches 1, the maximum, when the confidence is 0.531 or greater.

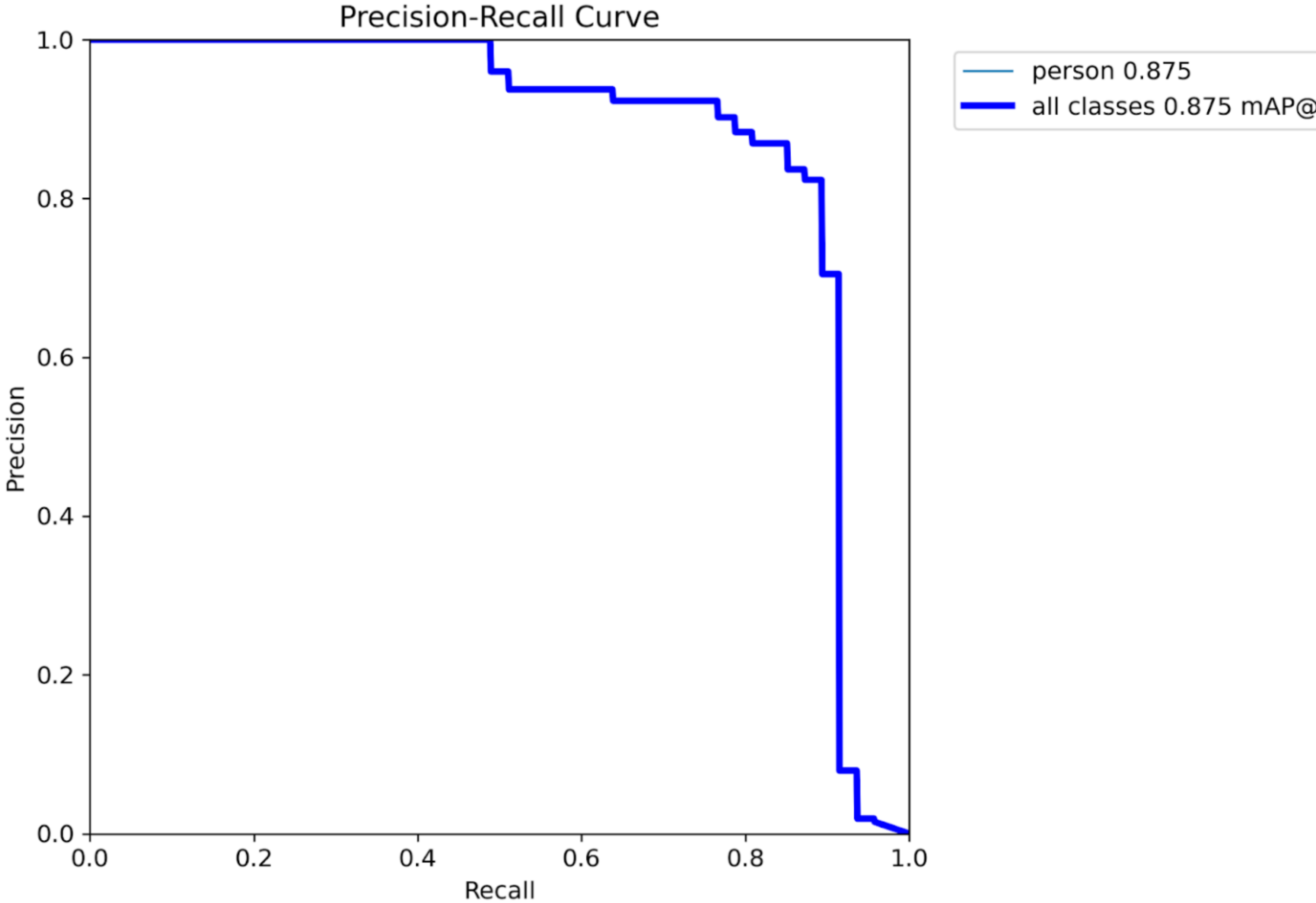


Figure 5.13: Precision-Recall Curve for the sample validation data.

The Precision-Recall curve, like the graphs before, have minimum values of 0 and maximum values of 1. As stated in Section 5.2, a high recall indicates that the false-negative rate is low, and a high precision indicates that a false-positive rate is low. The precision and recall in the results is overall very high, as shown by the large area under the curve of the graph. There is an increased rate of false positives when the recall reaches around 0.9. This is not unusual, as this number is being examined at the very edge of the model's accuracy. This graph compares the precision to recall balance point when mAP is 0.5.

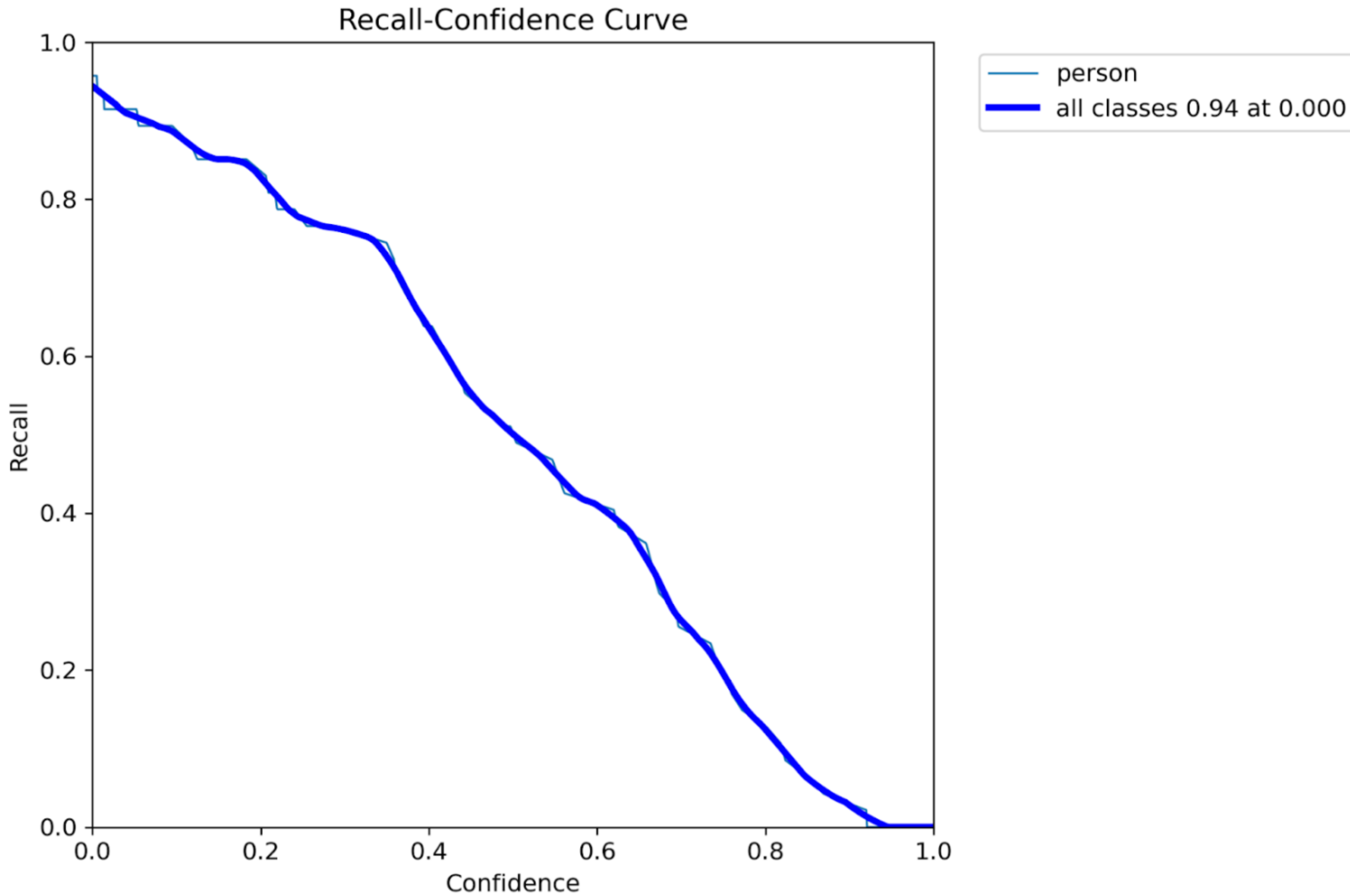


Figure 5.14: Recall-Confidence Curve of the validation data results.

The Confidence and Recall comparison compares the false-negatives rate to how confident the model is. Much like the Precision-Confidence graph, the more confident the model is, the more likely that the model is identifying true negative results, rather than false positive results. However, this result takes a much more linear form, showing that the model does struggle more with recall than precision. However, this is still a relatively good result. This means that the model is more likely to identify an object as being a person when it is not when the model does make an error.

(5.3.4) Instances

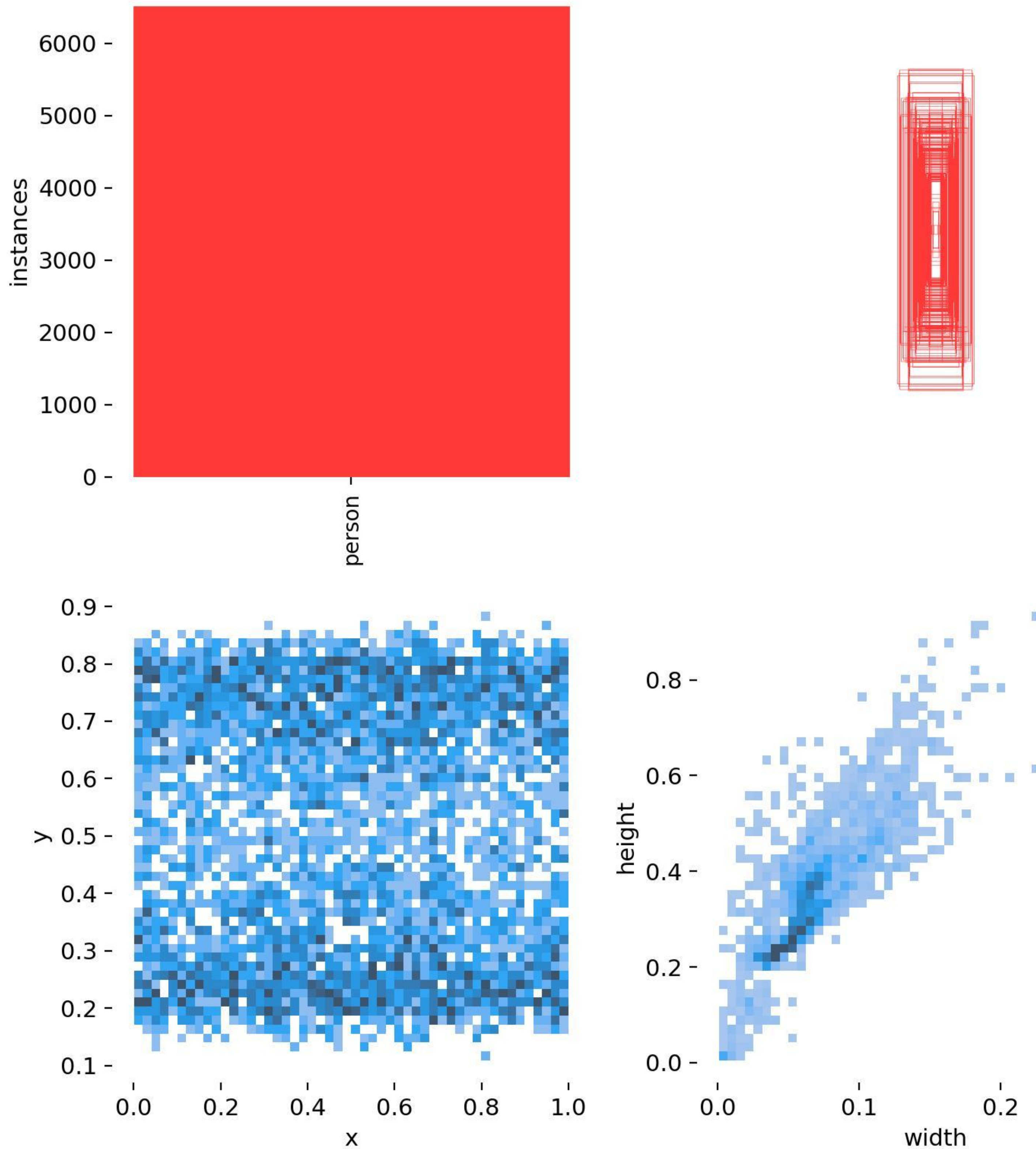
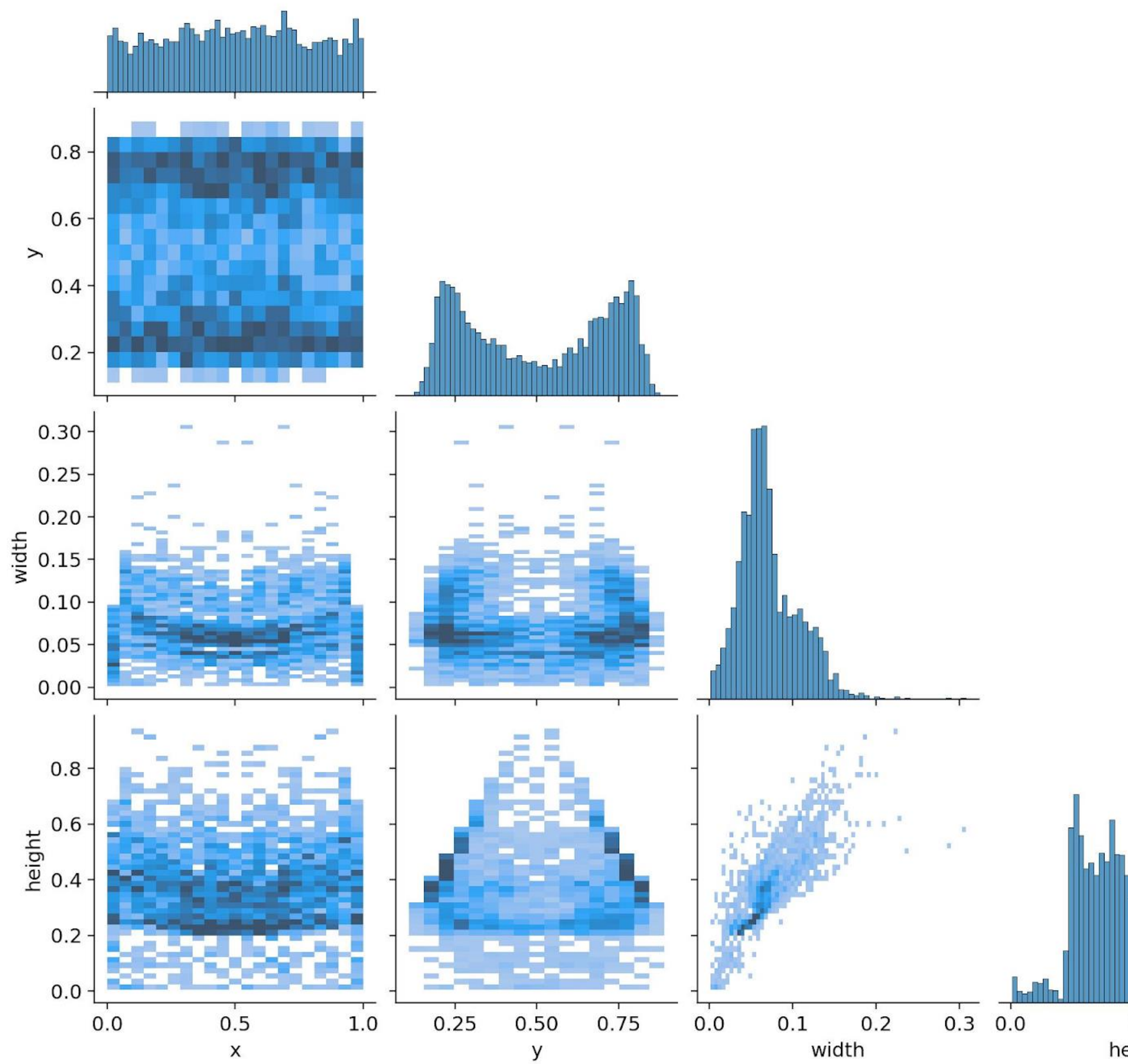


Figure 5.15: Chart of instances for the validation data.

The above chart shows the data relating to the bounding boxes used for object detection. It provides a summary of where in the images the bounding boxes were typically placed, how large they were, how frequent, and the x-y function. This is not directly related to the model accuracy results, however, it is important to consider any bias within the dataset, as this can be used as an indicator. Given the above results, the model trained on relatively balanced data. It also ended up typically creating bounding boxes that are taller than they are wide. However, given the typical human anatomy, this is an expected result.



0.8
0.6
0.4
0.2
y
0.30
0.25
0.20
0.15
0.10
0.05
0.00
width
0.8
0.6
0.4
0.2
0.0
height
0.0
0.5
1.0
x
0.25
0.50
0.75
y
0.0
0.1
0.2
0.3
width
0.0
he

Figure 5.16: Another group of graphs and charts, displaying the information from 5.15 in higher detail. This provides a clearer image towards the average bounding box dimensions.

(5.3.5) Confusion Matrix

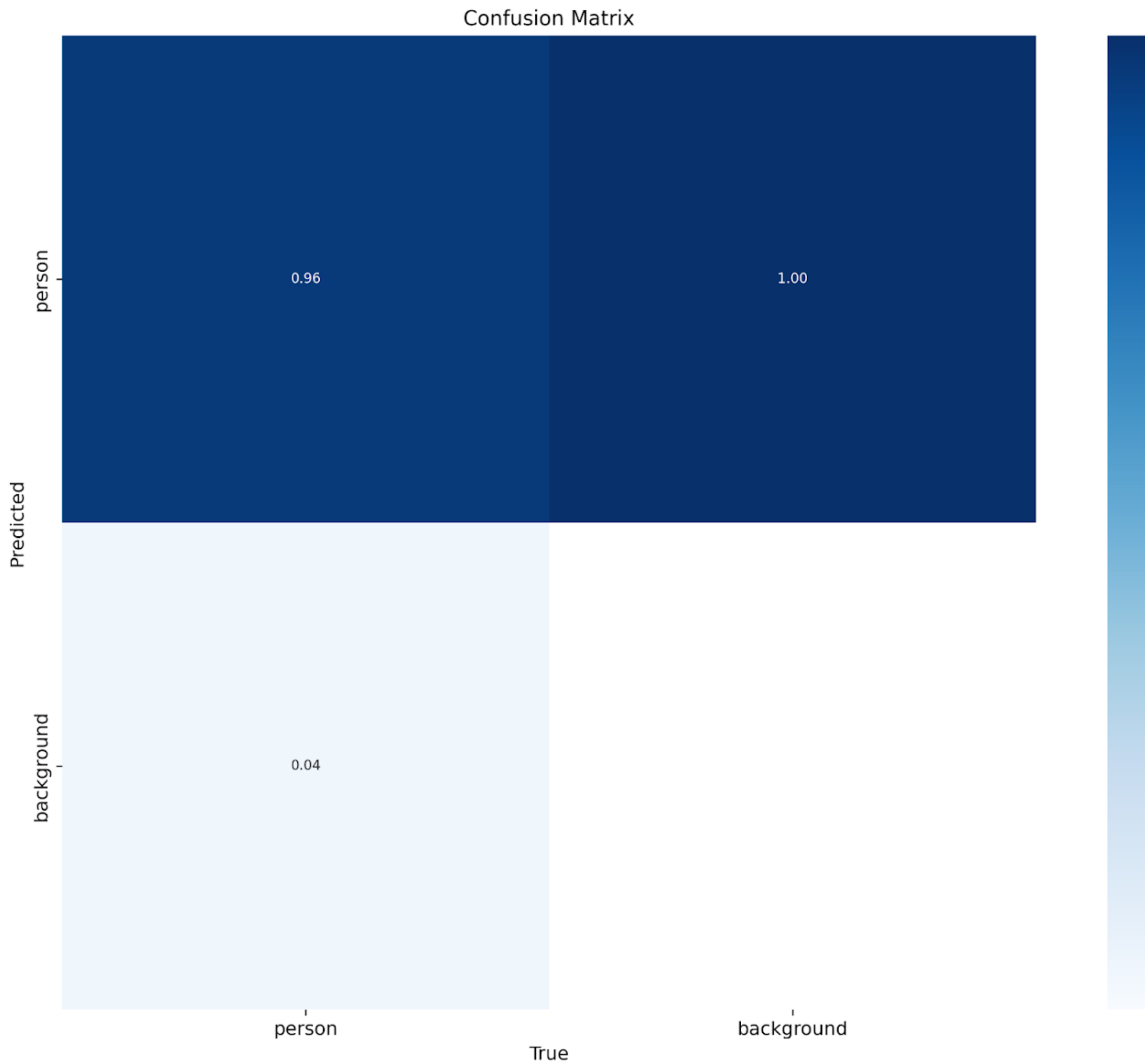


Figure 5.17: The confusion matrix for the validation data results.

The confusion matrix is made up of the true positives and negatives, as well as the false positives and negatives. In this instance, the confusion matrix compares the label the model predicted and what the true label was. The two classes being compared are the "person" class and the "background" class. Within the model, areas that are not occupied by a human in the image are considered background.

The above matrix indicates a similar result that the Confidence-Recall graph does, in that when the model makes mistakes, it typically misclassifies a part of the background as being a person. This is observed by the upper-right quadrant, the quadrant representing "predicted object was a person, was actually a background", and being valued at 1, on a scale of 0 to 1. The lower-left quadrant indicates that the opposite (model mistakes a person as being part of the background) rarely occurs.

The upper-left quadrant indicates that the model is fairly reliable in people being people when the model predicts this, however, can be inaccurate about this if the confidence of the label is low. The lower-right quadrant indicates that there is never any confusion when it predicts a background is a background, as previously indicated.

(5.3.6) Training, Loss, and Metrics

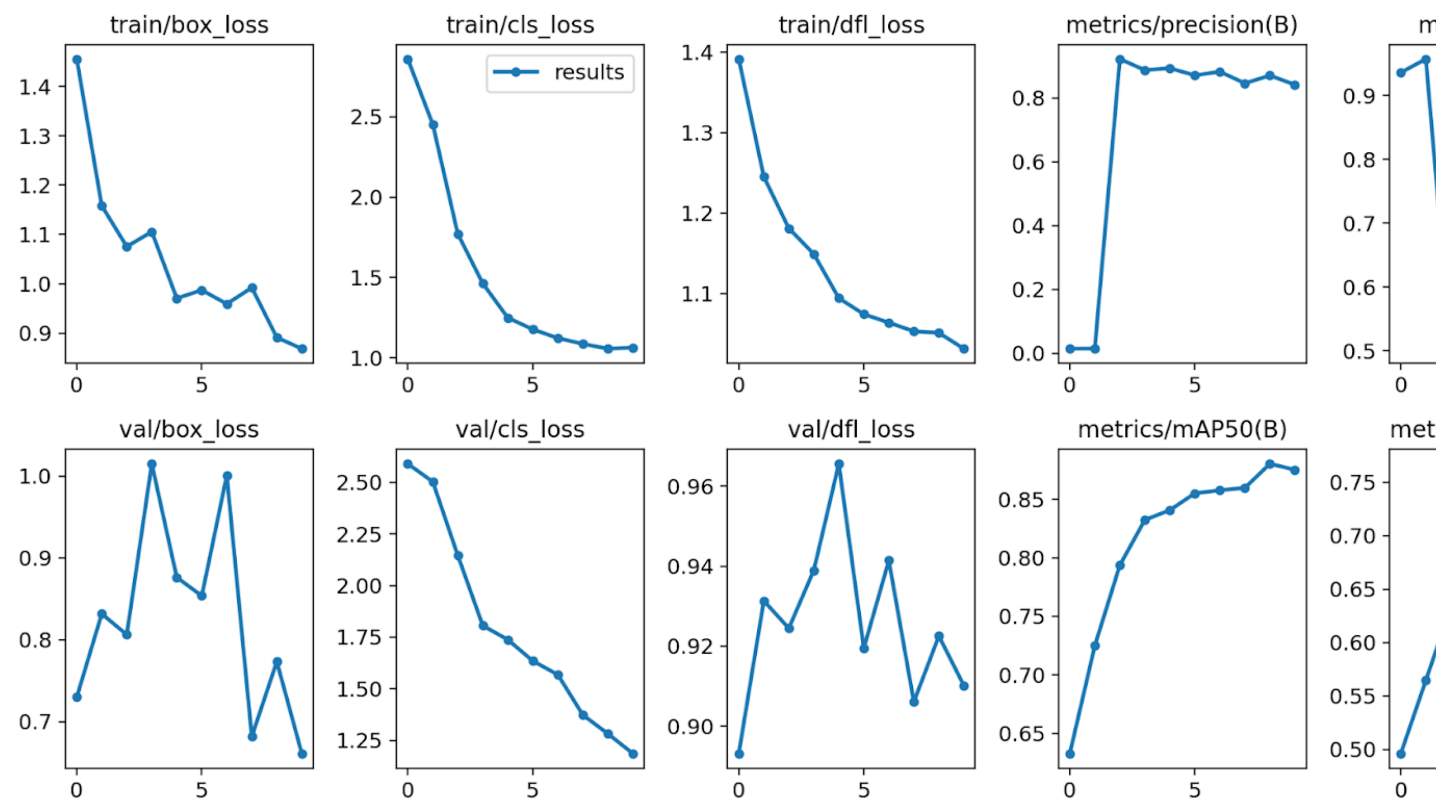


Figure 5.18: Graphs depicting the training, value, and metrics.

Figure 5.18 contains ten graphs. The results are overall very good, as they show the model is accurate with minimal costs for outlier detection. In order from top left to bottom right, going across, the graphs are as follows:

Train/Box Loss: The function represents how close the bounding box is to the object itself [8]. The graph shows that the model overall places the bounding box close to the person, rather than around the general area of the person.

Train/Class Loss: Also called entropy loss, this function shows how much of another class (in this case, the background class) is within the bounding box identifying a person (or vice versa) [8].

Train/Distribution Focal Loss: The DFL function helps balance an imbalance with the training data, in which there is an imbalance of examples between classes [8]. This graph being low indicates the imbalance is not large.

Metrics/Precision(B): This graph represents the precision across the epochs.

Metrics/Recall(B): This graph represents the recall across the epochs.

Value/Box Loss: This graph is another representation of how far the bounding box is from the centre of the person. In the model, this fluctuates significantly, meaning the bounding boxes aren't generally either close to or far away from the centre of the person. This makes sense, given the samples of crowds, people at differing distances, and obscured in the image, etc.

Value/Class Loss: The "Loss" function generally relates to the cost of making the model more accurate [8]. This represents the cost of an inaccurate classification. The graph shows a generally favourable class loss function for the model.

Value/Distribution Focal Loss: The DFL function, compared to value, is rather sporadic. This is not an ideal result, however, this does not have a significant impact on the model's accuracy or speed.

Metrics/mAP50(B): The metrics for mAP are shown. The resulting function is a great result.

Metrics/mAP50-95(B): The metrics for mAP are shown, ranging from 0.5 to 0.95. Though there is more fluctuation than just mAp at 0.50, this result is still relatively good.