**1.0 Evaluation**

**1.1 Metrics**

Below is the metric evaluation for model training with 50 and 100 epochs.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **epoch** | **train/box\_loss** | **train/cls\_loss** | **train/dfl\_loss** | **val/box\_loss** | **val/cls\_loss** | **val/dfl\_loss** |
| 50 | 1.1483 | 0.51484 | 0.8649 | 1.5142 | 0.63646 | 0.91213 |
| 100 | 0.89689 | 0.46219 | 0.83563 | 1.3244 | 0.56175 | 0.89177 |

1. **Training Loss**: The training losses decrease from the 50-epoch to the 100-epoch run, indicating that the model continues to learn and improve during the additional epochs. Lower loss values suggest better convergence.
2. **Validation Loss**: The validation losses for both runs shows a slight increase. This suggests that the model may begin to overfit, as it becomes less capable of generalizing to unseen data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **epoch** | **metrics/precision(B)** | **metrics/recall(B)** | **metrics/mAP50(B)** | **metrics/mAP50-95(B)** |
| 50 | 0.91164 | 0.9554 | 0.97305 | 0.57454 |
| 100 | 0.98075 | 0.94345 | 0.96823 | 0.61107 |

1. **Precision and Recall**: Precision(B) significantly increases from 0.91164 to 0.98075 in the 100-epoch run, showing that the model's positive predictions are more accurate. However, Recall(B) slightly decreases from 0.9554 to 0.94345, indicating a minor trade-off between precision and recall.
2. **mAP50**: mAP50(B) decreases from 0.97305 to 0.96823, signifying that the model's average precision at an IoU threshold of 0.5 for object detection has slightly declined, indicating a minor decrease in performance.
3. **mAP50-95**: mAP50-95(B) increases from 0.57454 to 0.61107 in the 100-epoch run, indicating that the model's performance over a broader range of IoU thresholds from 0.5 to 0.95 has slightly improved.

The 100-epoch training run generally results in better precision, improved mAP50-95, and a minor reduction in recall compared to the 50-epoch run. However, there are indications of potential overfitting in the 100-epoch run as seen in the increased validation losses.

**1.2 Graphs**

1. **Confusion Matrix**

A blue squares with white text

Description automatically generated

*(Confusion Matrix for 50 epoch)*

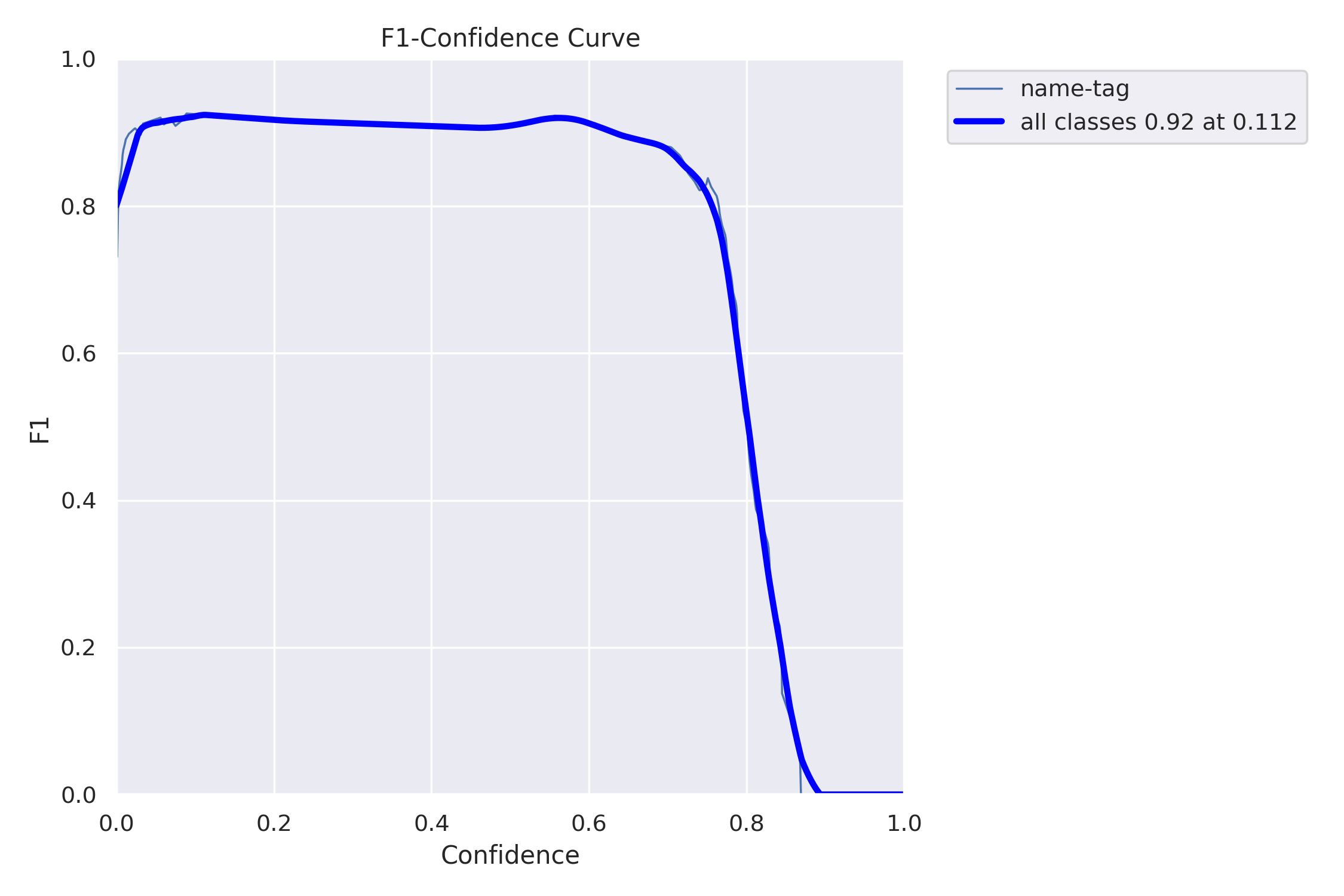
A blue squares with white squares

Description automatically generated

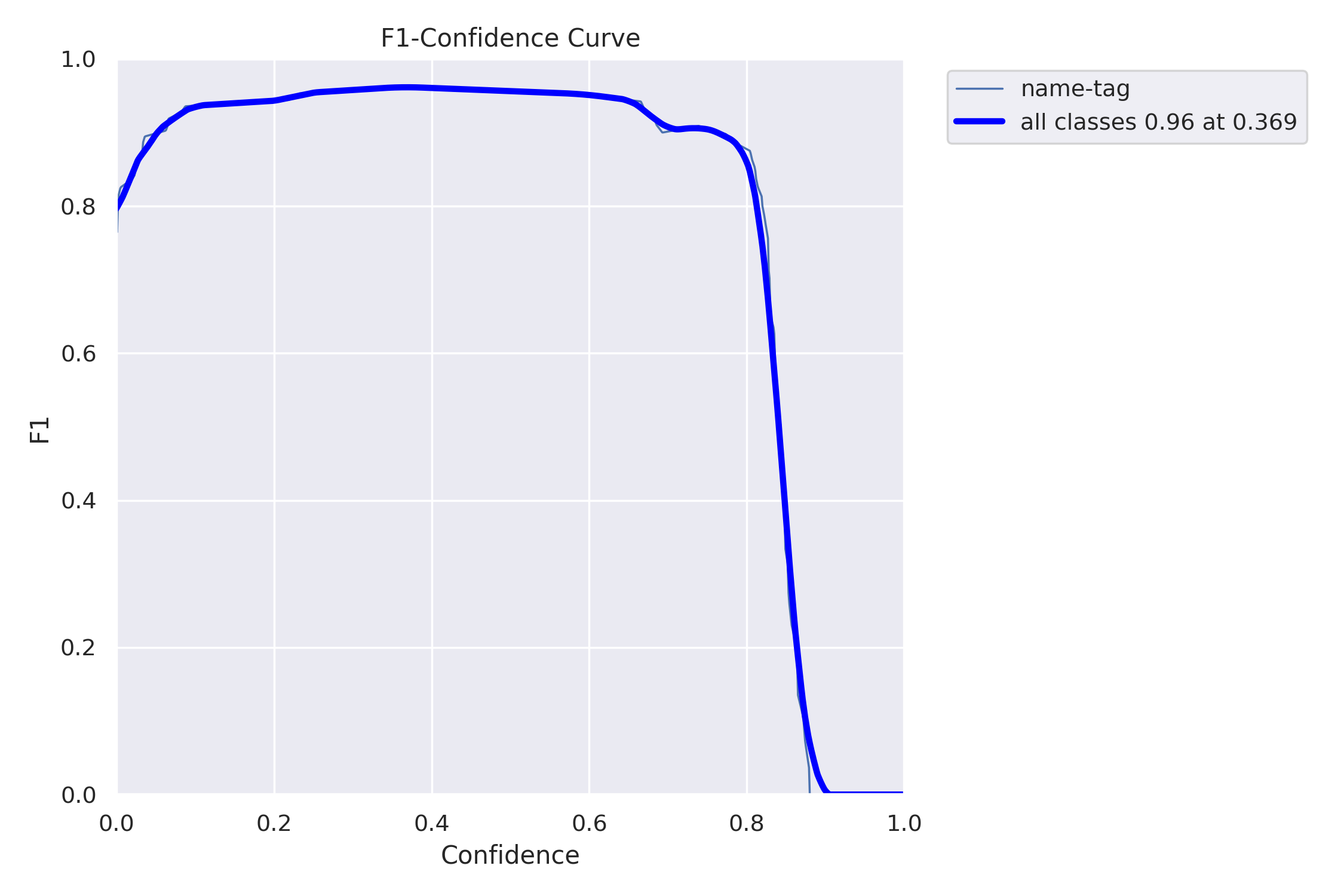
*(Confusion Matrix for 100 epoch)*

From the confusion matrix graph, the 50-epoch training run exhibited a True Positive rate of 89%, indicating its ability to correctly identify objects, while it had a 100% False Positive rate, implying some incorrect positive predictions. In contrast, the 100-epoch run showed an improved True Positive rate of 94% and a lower False Negative rate of 6%, indicating enhanced object detection accuracy. However, specific False Positive values for the 100-epoch run are not provided, making it essential to consider other metrics and aspects of the model's performance to comprehensively assess its capabilities, with the 100-epoch run showing better object detection performance.

1. **F1 Confidence Curve**



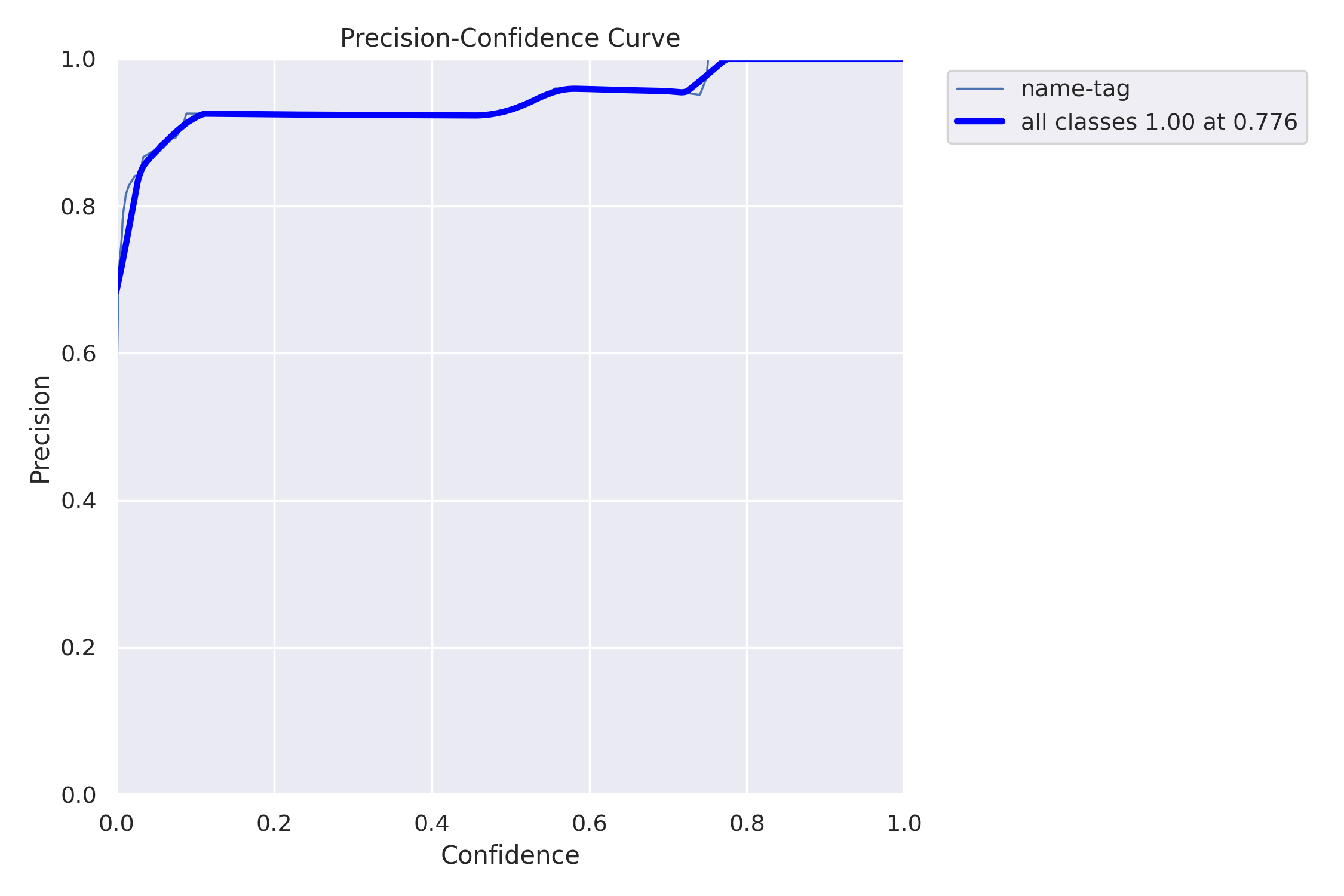
*(F1 Curve for 50 epoch)*



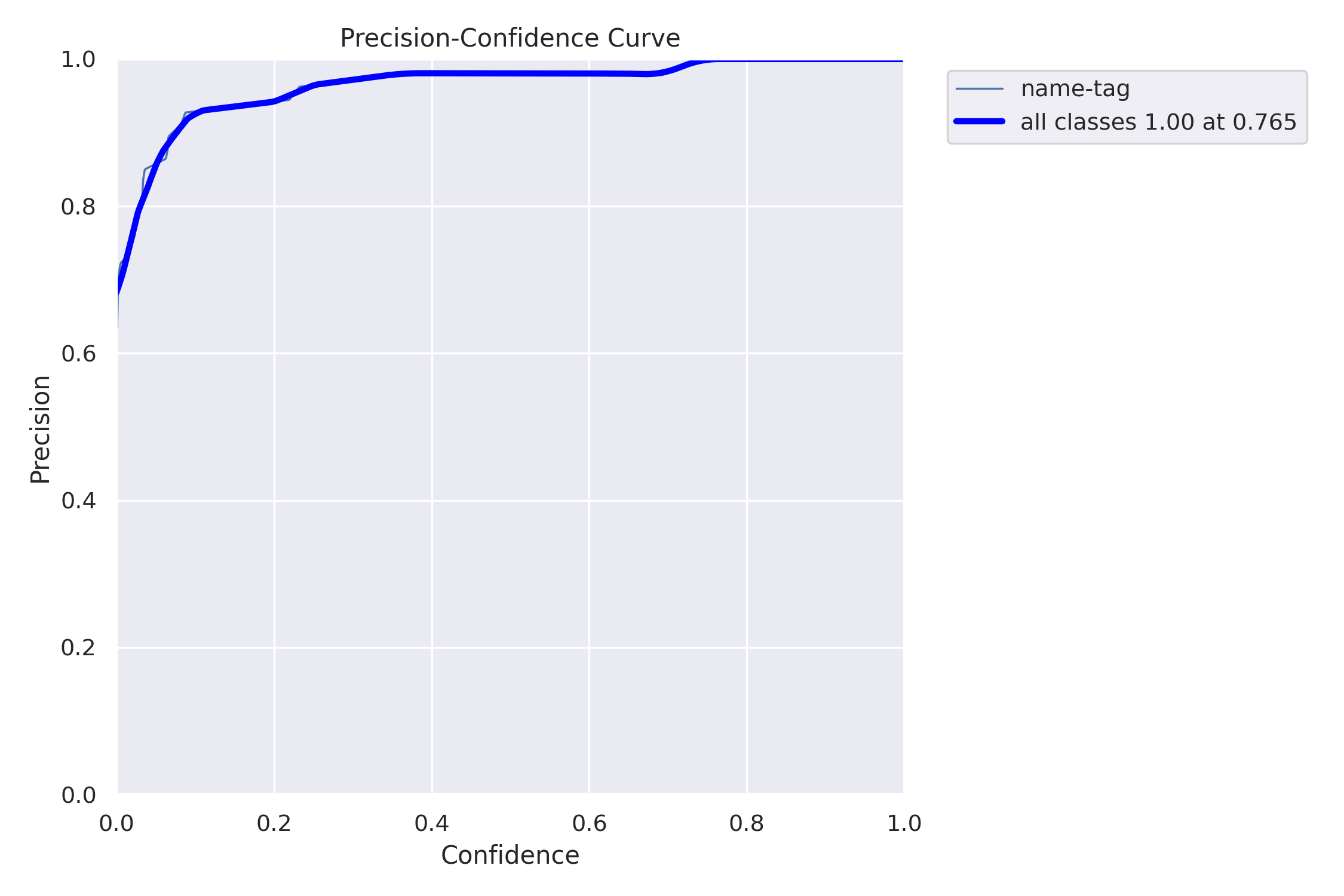
*(F1 Curve for 100 epoch)*

The 100-epoch model achieves an F1 score of 0.96 at a threshold of 0.369, outperforming the 50-epoch model, which attains a score of 0.92 at a threshold of 0.112. This suggests that the 100-epoch model excels in striking a balance between precision and recall at a higher threshold, indicating superior overall performance in object detection tasks.

1. **Precision- Confidence Curve**



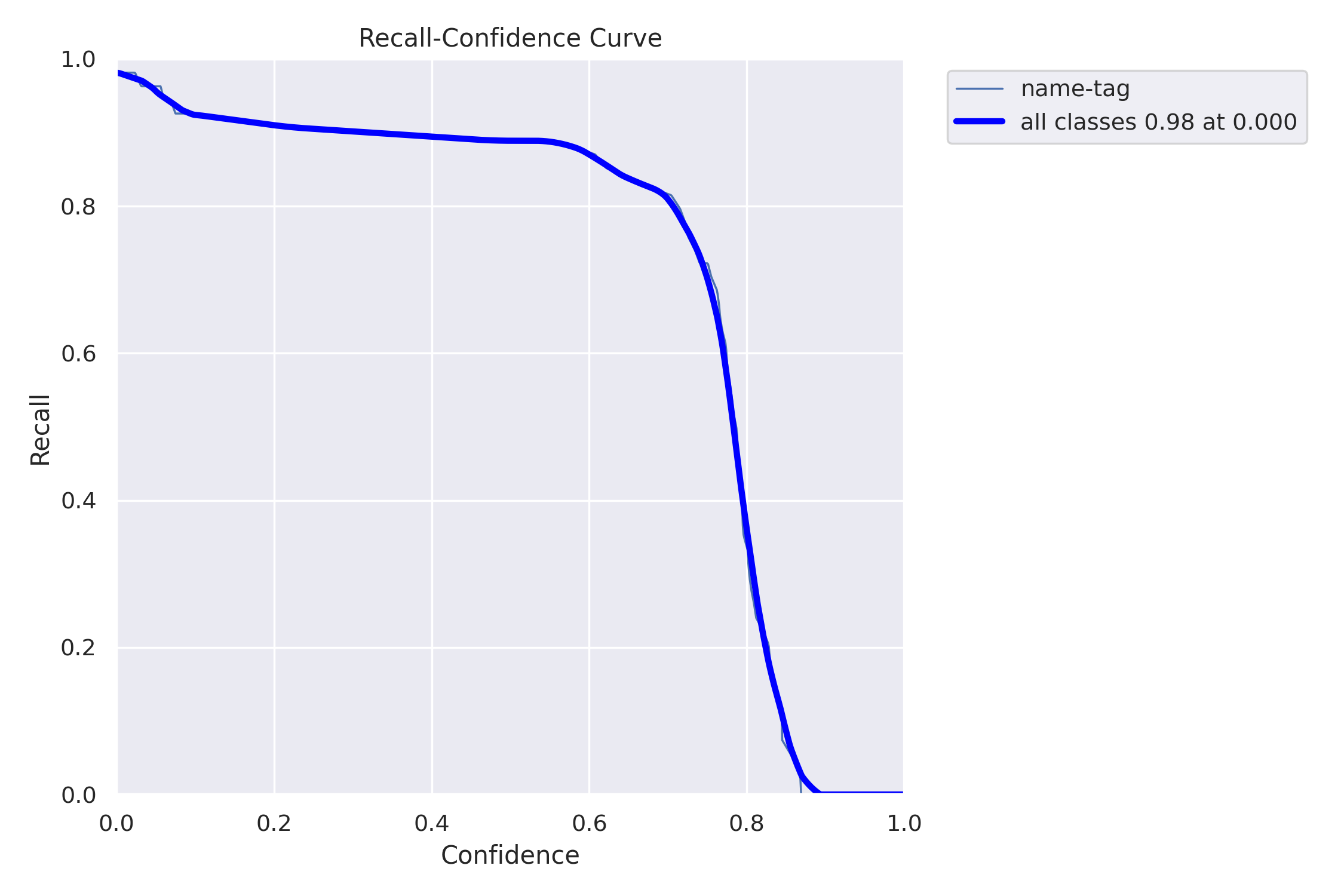
*(Precision Curve for 50 epoch)*



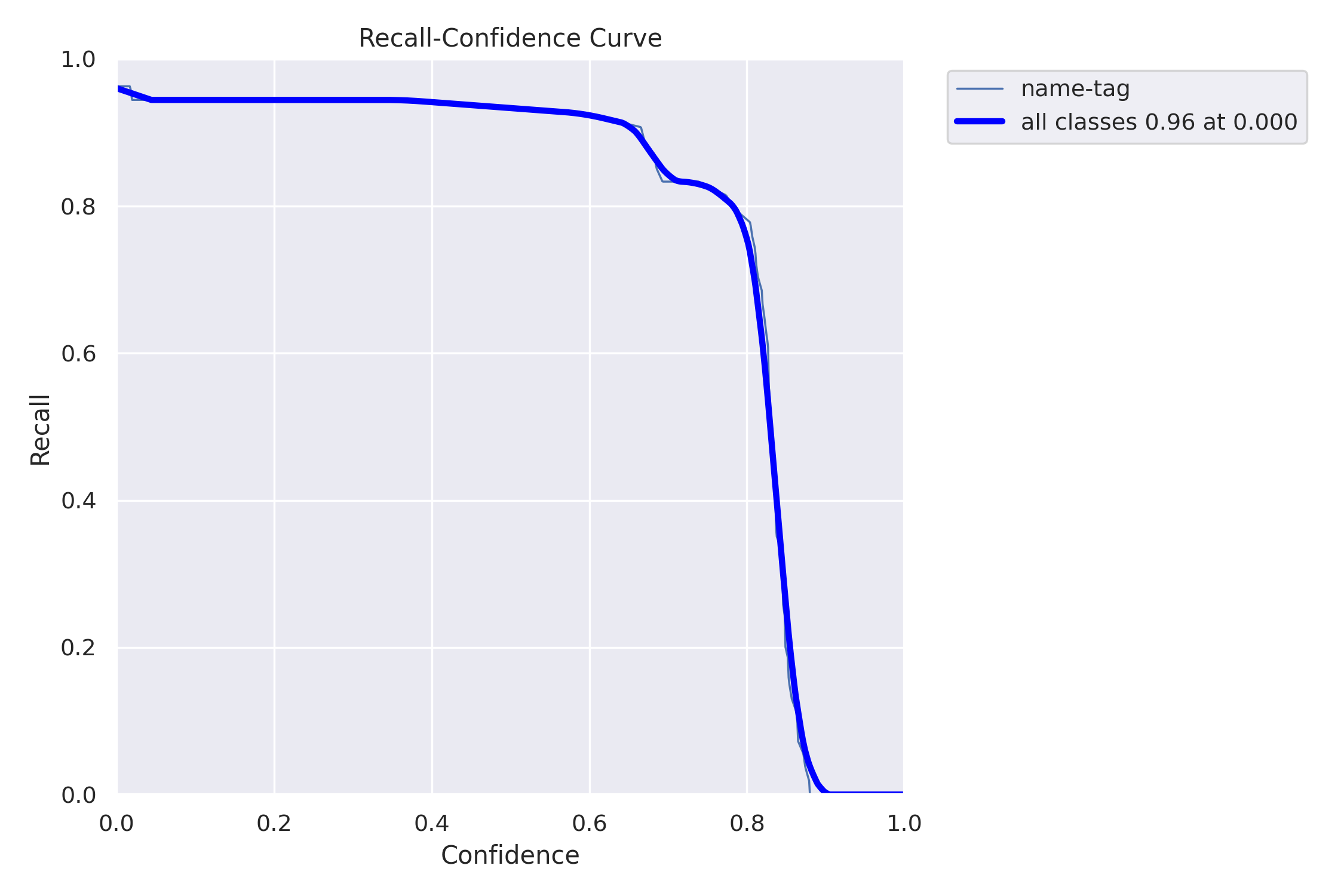
*(Precision Curve for 100 epoch)*

Precision denotes the accuracy of positive predictions, and is exceptionally high for both models, with a perfect precision score of 1.0 for all classes. This implies that both models excel in making accurate positive predictions, showcasing their reliability in object detection tasks.

1. **Recall-Confidence Curve**



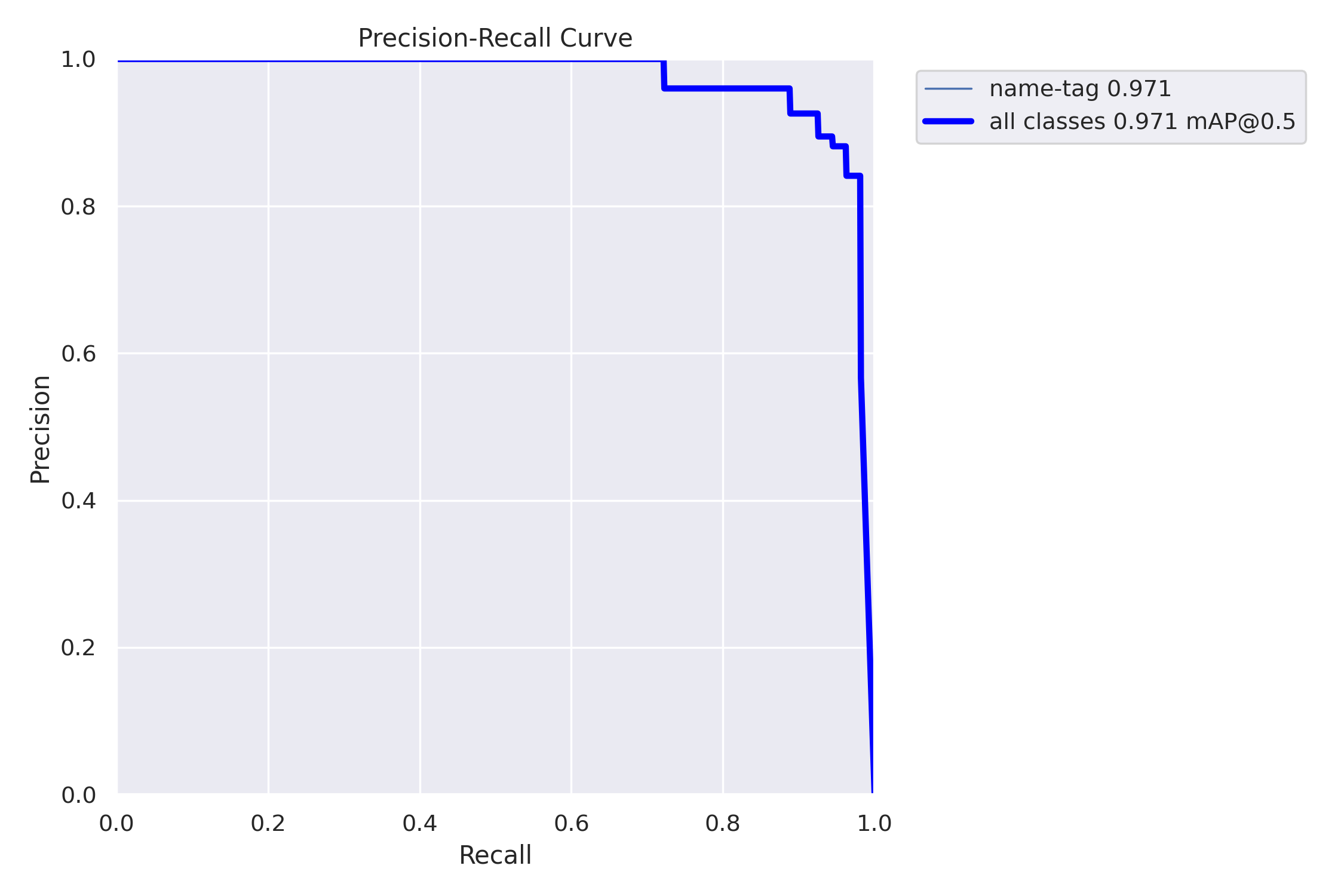
*(Recall Curve for 50 epoch)*



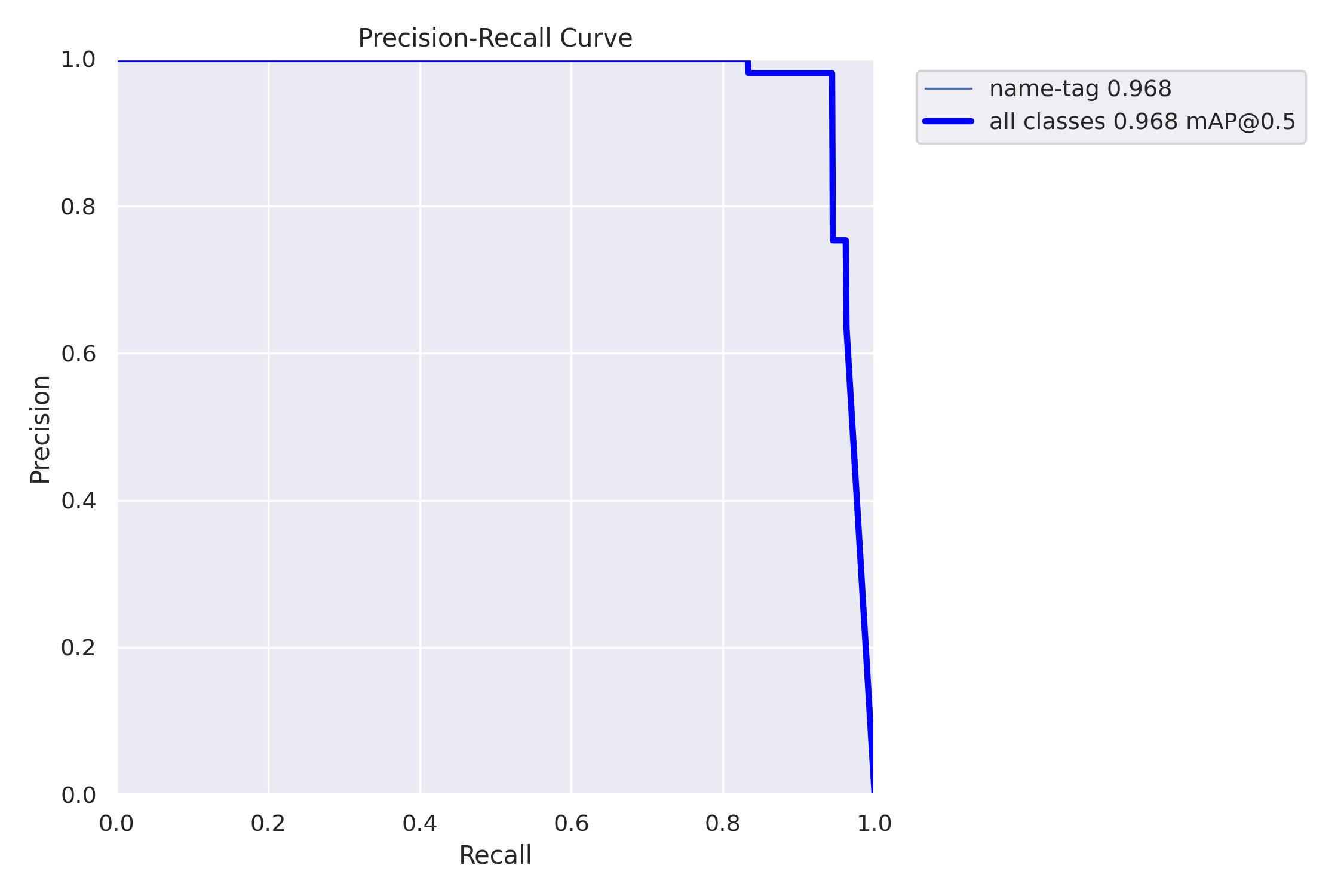
*(Recall Curve for 100 epoch)*

Recall is used for measuring a model's ability to identify all relevant instances, exhibits relatively lower values for both models. The 50-epoch model achieves a recall of 0.98, while the 100-epoch model attains a slightly lower recall of 0.96. However, the difference in recall between the two models is minimal, indicating comparable performance in this aspect.

1. **Precision-Recall Curve**



*(Precision Recall Curve for 50 epoch)*



*(Precision Recall Curve for 100 epoch)*

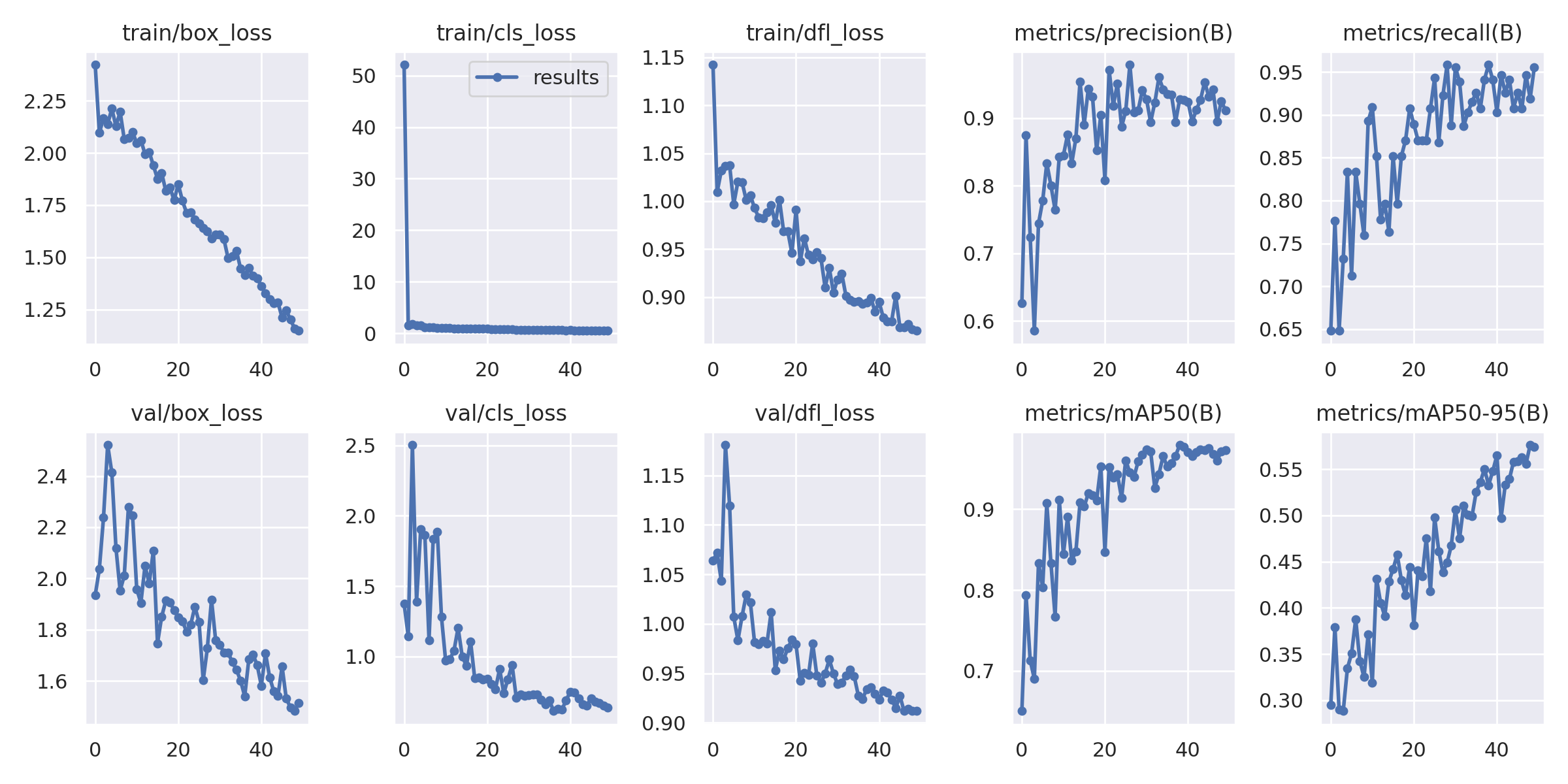
Both the 50-epoch and 100-epoch models yield remarkably similar mean average precision (mAP) scores at 0.5 which are 0.971 and 0.968. This demonstrates that the models perform on par in terms of object detection accuracy at a moderate IoU threshold.

**Conclusion**

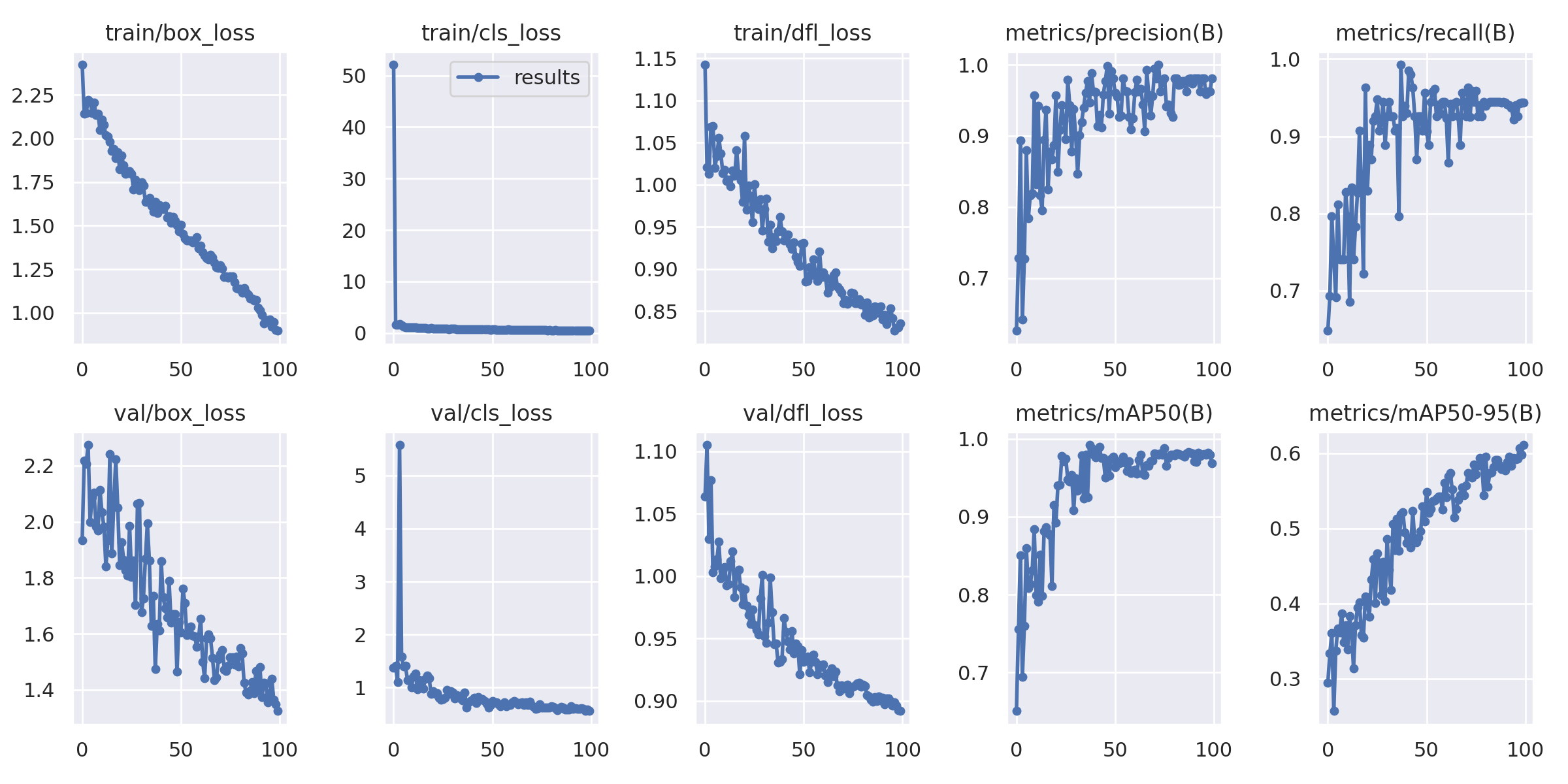
Considering the metrics and graphs, the 100-epoch training run stands out as the better choice for the name tag detection task. It demonstrates superior precision, achieving a precision score of 0.98075, indicating highly accurate positive predictions. While there is a slight trade-off in recall (0.94345), the overall balance between precision and recall remains excellent. The mAP50 value of 0.96823 showcases improved object detection accuracy, and the F1 Confidence Curve reinforces its superior performance, attaining an F1 score of 0.96 at a threshold of 0.369. Additionally, the confusion matrix illustrates an improved True Positive rate of 94%. Despite a minor reduction in mAP50-95, the 100-epoch run offers exceptional name tag detection capabilities. It outperforms the 50-epoch run in all relevant aspects, making it the preferable choice for this task.

However, it's important to note that the 100-epoch model exhibits signs of potential overfitting as evidenced by the increased validation losses. To mitigate the risk of overfitting, further regularization techniques or adjustments to the training process may be necessary to ensure the model's generalization ability while retaining the improvements in precision and mAP observed in the 100-epoch run.

**Other Results**



*(Results for 50 epoch)*



*(Results for 100 epoch)*

For both results from 50 and 100 epoch the trend of the graph is similar in which the higher number of epochs the lower the loss score.