Note

Metrics Selection

- **Intersection over Union (IoU)**

  For the selection of the evaluation metrics, firstly, I need to evaluate a single face. I choose **Intersection over Union (IoU)** because IoU is fundamental in object detection tasks, including face detection. It measures the overlap between predicted bounding boxes and ground truth, providing a quantitative measure of localization accuracy.

- **Precision, Recall, and F1-score**

  For face detection, Precision, Recall, and F1-score are widely used and intuitive metrics. They provide a clear understanding of how well the model identifies faces and handles false positives and false negatives.

  - **Precision**: (True Positives+False Positives)/True Positives. Indicates the accuracy of positive predictions, minimizing false positives.
  - **Recall**: (True Positives+False Negatives)/True Positives. Measures the model's ability to capture all relevant instances, reducing false negatives.
  - **F1-score**: 2*(precision*recall) / (precision+recall). Harmonic mean of precision and recall, offering a balanced evaluation when there's an imbalance between precision and recall.

- **Precision-Recall Curve**

  A precision-recall curve is a graph that shows the trade-off between precision (the number of true positives divided by the total number of predicted positives) and recall (the number of true positives divided by the total number of actual positives) at different thresholds.

- **Average Precision (AP)**

  AP summarizes the precision-recall curve into a single value. It considers how well the model ranks the detections and computes the area under the precision-recall curve.To compute AP, the precision-recall curve is divided into small intervals, and the average precision in each interval is calculated. These individual precisions are then averaged over all intervals, resulting in the average precision value.

  AP is essential when dealing with multiple faces in an image or when evaluating the model across various images with different numbers of faces, which is just the case in the task. It provide a more comprehensive view of the model's performance, especially when dealing

with varying complexities in face detection scenarios.