The F1-Confidence Curve is a valuable visualisation for assessing the performance of our object detection model. It shows how model’s F1 score changes across different confidence thresholds taking into account both the precision as well as the recall of a model.

Understanding the graph

The X-axis represents the confidence threshold i.e the probability assigned by the model to each correct detection. It ranges from 0 to 1. Lower confidence means the model results in more detections while higher threshold results in fewer detections with high certainty. The Y-axis is the harmonic mean of precision and recall. High F1 score indicates good balance between precision and recall.

The light blue line represents the F1 score for detecting “crop” while the orange line represents the F1 score for detecting “weed”. F1 scores for both classes rise indicating model is more confident. After reaching its peak, the score drops likely due to its higher precision but decreased recall. The bold blue line represents the F1 score when considering both weed and crop classes collectively. The peak of this line indicates the optimal confidence for the model when detecting all classes. From the graph we can analyse that at a confidence of 0.305, the combined F1 score for all classes is 0.71.

Key Insights from he Curve  
The optimal confidence threshold for our model’s performance is 0.305, providing the best balance between detecting both classes with minimum false poitives or false negatives.

F1 score for crop is consistently higher than for weed at most confidence levels, suggesting that the model is generally better at detecting crops. A low F1 score for weed can also mean its harder for the model to distinguish the weed from the background noise.

The F1 score declines after the optimal confidence level likely because recall drops and precision increases, this trade-off results in lower false positives and more accurate results.