Our dataset was trained on YOLOv5 object detection model and the results of summary of performance metrics is shown below in the Table 6.1. As mentioned in previous chapter, our final dataset contains 1550 images.

6.1 YOLOv5 RESULTS

Table 6.1 : Result of YOLOv5 object detection model

CLASS	IMAGES	LABELS	P	R	MAP@0.5	MAP@0.5:0.95
all	310	480	0.37	0.246	0.0803	0.0407
Axe	310	29	1	0	0.0261	0.012
BillGates	310	32	0.107	0.469	0.113	0.0531
Bottle	310	179	0.0359	0.00559	0.0124	0.00273
ElonMusk	310	20	0.0708	0.7	0.192	0.118
Hammer	310	26	1	0	0.00839	0.00252
Handgun	310	40	0.132	0.05	0.0252	0.00768
Knife	310	54	1	0	0.0191	0.0043
LeonardoCaprio	310	29	0.116	0.724	0.167	0.0938
MarilynMonroe	310	29	0.0296	0.0345	0.0424	0.016
WillSmith	310	42	0.21	0.476	0.198	0.0968

Class Interpretations:

1. Axe:

The results of the "Axe" class shows perfect precision (P = 1.0) is achieved alongside zero recall (R = 0.0), which highlights the critical insights of model's behaviour and its inferences for practical applications.

A precision of 1.0 indicates that when the model predicts the presence of an "Axe" in an image, it is always right. This high value of precision implies a complete absence of false positive predictions for Axe class, suggesting that when the model identifies an "Axe," it is definitely present in the image. However, this remote study of precision alone does not provide a complete understanding of the model's efficacy.

Recall measures the capability of the model to detect all actual cases of a class within the dataset. Here, recall of 0.0 shows that the model fails to identify any true instances of "Axe" present in the images. This shortfall in recall implies a serious restraint in the model's ability to broadly separate and detects "Axe" objects, which results in a high number of false negatives instances of "Axe" that are not detected by the model.

Here, precision of 1.0 with 0 recall shows a disparity in the performance of the model. It can correctly detect images that do contain axes but it supervises every true instance of an "Axe" present in the dataset which highlights a crucial trade-off between accuracy and recall that needs to be addressed for the model to be almost viable.

2. BillGates:

The precision (P) is 0.107 and a recall (R) is 0.469 for "BillGates" class. It indicates that the model predicts "BillGates" class correctly only about 10.7% of the time.

So, there is a high rate of false positive calculations where the model incorrectly detects some objects or person as "BillGates." This aspect of the model's performance presents challenges in terms of accuracy and consistency for applications relying on precise object detection.

The relatively low precision value suggests room for upgrading in the model's ability to determine true instances of "BillGates" from other visual elements within the dataset. The significance of this is potential imprecisions and deceptive results in downstream applications that utilize the model's predictions.

On the other hand, the recall value of 0.469 signifies that the model can detect nearly 46.9% of all actual occurrences of Bill Gates present in the images. This moderate recall rate indicates a capability to identify a substantial proportion of relevant instances of "BillGates," indicating a reasonable level of efficacy in capturing instances of this specific individual within the dataset.

3. Bottle:

The precision (P) of 0.0359 and a recall (R) of 0.00559 for “Bottle” class indicates that the model detects with accuracy of only 3.59%. This low precision rate indicates that the predictions made by the model are mostly false positive. The values indicate serious shortfalls in the model’s ability to detect and identify the instances of the bottle within the dataset. The precision value of 0.0359 and accuracy of 3.59% shows large number of false positive results, indicating in practical challenges in the application of the model in the real-world scenario. Ironically the recall value of 0.00559 for the object of the bottle indicates that the model can detect all instances present in the dataset with accuracy of 0.559% only.

In contract with the precision value of 0.0359 and accuracy of 3.59%, indicates high chances of false positive results. On the other hand, the recall value of 0.00559 “Bottle” class signifies that the model can detect only around 0.559 % of all accurate instances of bottles present in the dataset. This significantly low recall rate indicates the models incompetency to detect the bottles accurately sabotages the performance of the model and its usability to detect the bottles in real world scenario.

4. ElonMusk:

The precision (P) of 0.0708 and a recall (R) of 0.7 for “ElonMusk” class indicates that the model detects with accuracy of only 7.08%. This low precision rate indicates that the predictions made by the model are mostly false positive. The results indicate serious challenges in the model’s ability to detect and identify the instances of the bottle with in the dataset. This precision value of 0.0708 indicates high chances of false positive results indicating the short falls of the model.

On the other hand, with the recall value of 0.7 for the class "ElonMusk", the model detects around 70% of all true instances within the dataset. This high recall value gives 70% accuracy in detecting all instances of “Elon Musk” within dataset. This shows the models stronger capability to detect this specific individual compared to other classes with lower recall rates.

This comparison of low precision with high recall for the "ElonMusk" indicates certain strength and weakness in the model’s ability to detect this specified individual. Model proves a strong ability to capture true instances of Elon Musk (high recall) while suffering false positive predictions due to the low precession rate leading to inaccuracies and certain short falls in predictions of this class.

5. Hammer:

The precision (P) of 1.0 and a recall (R) of 0.0 for "Hammer" class indicates models' strengths and weakness in accurately identifying instances of hammers. Giving a precision value of 1.0 the model achieves a prefect precision each time it detects hammer. This perfect precision score signifies no false positive predictions for hammers, indicating in great confidence in the model's ability to identify this specific object. On the other hand, due to the perfect precision we also need to critically examine recall metrics, to assess the model's overall efficiency.

In contract the recall of 0 signifies that the model misses all instances of hammers present in the images, resulting in complete inability to identify this object. Despite achieving perfect precision, the lack of any recall for hammers highlights a severe flaw in in the model's detection capability for this very class.

This understanding of precision and recall metrics for the "Hammer" class reveals unevenness in the performance of the model. While the model is superior in giving perfect precision its inability to detect any true instances of hammer with a zero recall questions the model's practicality and effectiveness in real-world applications needing an accurate object detection.

6. Handgun:

The precision (P) of 0.132 and a recall (R) of 0.05 for "Handgun" class shows that the model successfully detects Handgun 13.2% of the time. This low precision rate indicates that the predictions made by the model are mostly false positive. They do not match the actual handgun in the image. The model frequently misclassifies other objects as handguns, leading to potential inaccuracies and inadequacies in applications dependency on these predictions.

On the other hand, the recall value for "Handgun" is 0.05 resulting in only 5% detection of all true instances. This extremely low recall rate indicates that the model misses a huge number of actual handguns present in the images. This inability of the model to capture most actual instances of handguns shows a critical limitation in its competence to detect object efficiently.

The blend of low precision and very low recall suggests that the model struggles both in precisely identifying true instances of handguns and in generously capturing and recognizing this object class within the dataset. The models reveal specific challenges and shortcomings of the model in detecting handguns precisely to the values of precision and recall metrics.

7. Knife:

The precision (P) of 1.0 and a recall (R) of 0.0 for “Knife” class indicates models’ strengths and weakness in accurately identifying instances of Knife. Giving a precision value of 1.0 the model achieves a prefect precision each time it detects Knife. This perfect precision score signifies no false positive predictions for Knife, indicating in great confidence in the model’s ability to identify this specific object. On the other hand, due to the perfect precision we also need to critically examine recall metrics, to assess the model’s overall efficiency.

In contract the recall of 0 signifies that the model misses all instances of Knife present in the images, resulting in complete inability to identify this object. Despite achieving perfect precision, the lack of any recall for Knife highlights a severe flaw in in the model's detection capability for this very class.

This understanding of precision and recall metrics for the " Knife " class reveals unevenness in the performance of the model. While the model excels in giving perfect precision), its inability to detect any true instances of Knife (zero recall) raises significant questions in its practical utility and effectiveness in real-world applications needing an accurate object detection.

8. LeonardoCaprio:

The precision (P) of 0.116 and a recall (R) of 0.724 for “LeonardoCaprio” class indicates that the model successfully detects this class object 72.4% of the time. This low precision rate indicates that the predictions made by the model are mostly false positive. They do not match the actual Leonardo

DiCaprio in the image. Many of the instances that the model identifies as Leonardo DiCaprio are incorrect, which suggests that the model is not very sharp in its detection of this particular class. This high recall suggests that the model is quite active in capturing most of the real incidences of Leonardo DiCaprio, preventing it from not missing many true instances. However, while the model is good at finding instances of Leonardo DiCaprio, it also includes a lot of improper findings. The high recall paired with low precision suggests that the model prioritizes identifying as many true instances as possible, even if it means accepting a good number of false positives.

9. Marilyn Monroe :

The precision of 0.0296 for the "MarilynMonroe" class suggests that the model's predictions for Marilyn Monroe highly inaccurate supporting it with a very high rate of false positive predictions. The results indicate that only 2.96% of the instances are correctly identified. There is a significant drawback in the model's ability to precisely differentiate Marilyn Monroe from other objects or people in the dataset

In contrast, the recall of 0.0345 indicates that the model successfully detects only a small fraction, specifically 3.45%, of all true instances of Marilyn Monroe within the dataset. Having such a low recall rate Indicates that the model missies a vast majority of the real occurrences of MarilynMonroe. This signifies a serious shortage in the models training and capability to identify and classify Marilyn Monroe accurately.

10. WillSmith :

The precision (P) of 0.21 and a recall (R) of 0.476 for “WillSmith” class indicates that the model successfully detects this class object 21% of the time. This indicates a moderate rate of false positive predictions. Consequently, the model often wrongly identifies other people or objects for Will Smith, showing its limited accuracy in differentiating.

In contract to the precision, the recall of 0.476 indicates that the model successfully detects approximately 47.6% of all true instances of Will Smith within the dataset. This shows a moderate recall indicating that the model has brief ability to recognize and identify “WillSmith” even though it fails to detect more than half of the real instances present in the dataset. Therefore, the model displays a partial ability to detect Will Smith, but ironically its overall performance is tampered both the significant rate of false positives and its inability to dependably detect true instances, indicating a strong need in its training and accuracy.

Overall Interpretation:

The overall assessment of an object detection model indicates significant challenges across various classes, showing a gap in precision and recall metrics underlining the requirement of considerable improvements in performance. The model displays commendable precision (1.0) for specific classes such as "Axe," "Hammer," and "Knife," with complete absence of false positive results. On the contrary this precision comes at a cost of zero recall, indicating a critical failure to detect any real instance of these classes. This result is alarming as it signifies that the model is confident in its detection, however it misses the real presence of these objects in the images, highlighting a fundamental error in its ability to detect these objects.

On the other hand, for the classes like "BillGates," "ElonMusk," and "WillSmith," the model’s precision remain low leading to a significant number of false positive results. This variable inconsistency in precision highlights the hurdles faced by the model, resulting in misclassifications that can impact the applications reliable on accurate object detection. Besides, recall rates across all classes are usually poor, representative that the model scuffles to detect an important portion of accurate instances for most classes, further showing shortages in its ability to largely capture the objects of interest within images.

The Mean Average Precision (MAP) scores, particularly at IoU (Intersection over Union) thresholds of 0.5 (0.0407) and across thresholds from 0.5 to 0.95 (0.012), further reinforce the overall poor performance of the model in object detection tasks. The low MAP values suggest that the model's capability to accurately localize and detect objects across various classes is inadequate, especially under more stringent IoU thresholds, which are crucial for ensuring precise localization of objects within images.

The results of the YOLOv5 object detection are displayed in figure 6.1. The results show that the YOLOv5 model gives false positive results with a lower accuracy. From the image below, we can clearly see that the model wrongly detected Bill Gates and Leonardo DiCaprio as Elon Musk. The model also failed to identify other objects such as Knife, gun and hammer.

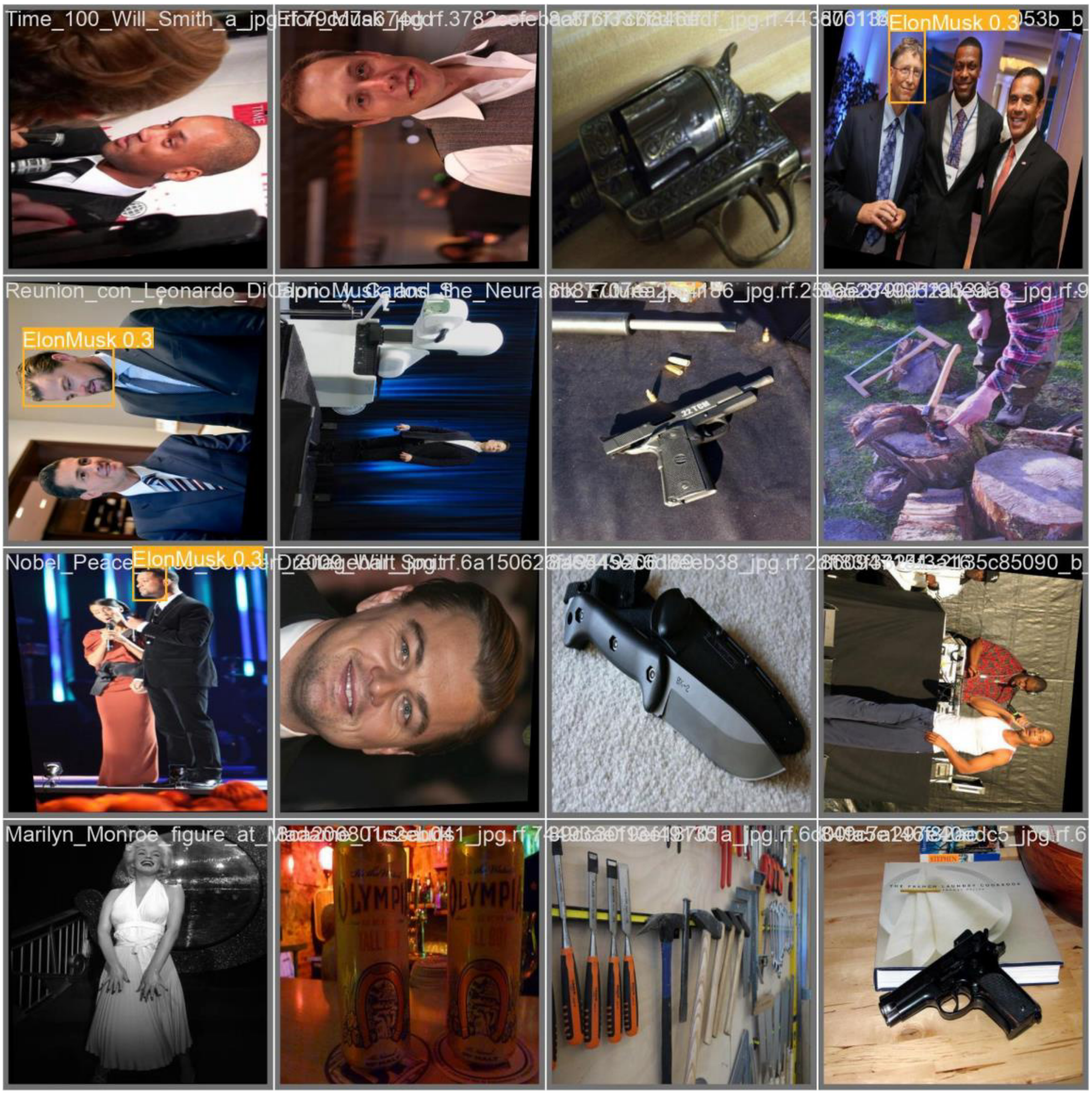

Fig. 6.1 : Result of YOLOv5 model on Image

6.2 YOLOv7 RESULTS

Table 6.2 : Result of YOLOv7 object detection model

CLASS	IMAGES	LABELS	P	R	MAP@0.5	MAP@0.5:0.95
all	310	480	0.435	0.474	0.432	0.238
Axe	310	29	0.307	0.0345	0.175	0.0769
BillGates	310	32	0.357	0.831	0.576	0.319
Bottle	310	179	0.463	0.145	0.144	0.0503
ElonMusk	310	20	0.287	0.95	0.712	0.425
Hammer	310	26	0.287	0.269	0.158	0.0637
Handgun	310	40	0.461	0.342	0.325	0.121
Knife	310	54	0.456	0.296	0.296	0.152
LeonardoCaprio	310	29	0.367	0.724	0.523	0.357
MarilynMonroe	310	29	0.75	0.722	0.826	0.442
WillSmith	310	42	0.62	0.429	0.588	0.371

The results of summary of YOLOv7 model's performance across different classes is displayed in Table 6.2. It reveals insights into its efficiency in detecting specific objects within the dataset. Analyzing the precision, recall, and Mean Average Precision (MAP) scores for each class provides an inclusive understanding of the model's strengths and weaknesses.

Class-Specific Interpretation:

1. **Axe:**

   The performance of YOLOv7 model in detecting the object “Axe” was below average due to a low precision rate of 0.307, indicating successful detection of this class object 30.7%. of the time. Such a low precision indicates a significantly high rate of false positive predictions, where the model falsely identifies other objects as axes as well.

   On the other hand, the recall for axes is significantly low at 0.0345, indicating that the model only captures 3.45% of all true instances of the object of axes within the dataset. The model certainly misses vast majority of real axes present in the dataset. This proves incompetency of the model to effectively detect axes, compromising its usefulness in tasks requiring precise identification of this object class.

Overall, the YOLOv7 model's performance comprises of high rate of false positive predictions and inability to detect all true instances of the object “axe”. Suggesting a substantial limitation and need for targeted improvements in the data quality, model architecture and training strategies to enhance precision and reliability in detecting “axe” objects class.

2. Bill Gates:

The YOLOv7 model displays a balanced performance in detecting instances of Bill Gates with a moderate precision score of 0.357 and successfully detects the images of Bill Gates 35.7% of the time. This precision value suggests in some false positive predictions with reasonable accuracy in detecting all true instances of Bill Gates.

The model shows a high recall rate of 0.831 for Bill Gates. This recall value indicates that the model successfully detects 83.1% of all instances of Bill Gates with in the dataset. This high recall shows that the model is competent to detect all majority of real instances of Bill gates, with a low false negative detection.

Overall, the YOLOv7 model's performance for Bill Gates shows efficiency of the model to detect the object. The moderate precision score indicates some presence of false positives. In contract to the high recall rate indicating the accuracy and efficiency of the model to detect bill gates, making this suitable for practical application of the model in real world.

3. Bottle:

The YOLOv7 model demonstrates a mixed performance in detecting instances of bottles with a high precision score of 0.463 and successfully detects the images of this class object 46.3% of the time. This indicates a low rate of false positives for the class bottle.

The model has a low recall rate of 0.145 indicating that the model only captures 14.5% of all true instances. On the other hand, the model misses a significant proportion of real bottle instances, leading to incomplete identification.

The high precision score indicates that the predictions of the model are accurate, with a low rate of false positives. However, the low recall rate indicates that the model misses a considerable number of instances.

However, the YOLOv7 model demonstrates both accuracy and efficacy in detecting the bottles with a high precision rate. In contract its performance is limited due to a low recall rate. Improvements in the model's recall to detect the object would enhance the efficacy in detecting this object class, ensuring reliability of the model in real world application.

4. ElonMusk:

The YOLOv7 model's performance in detecting instances of Elon Musk gives mixed results with a precision score of 0.287 and successfully detects the images of this class object 28.7% of the time, indicating a high rate of false positive predictions. The model incorrectly detects Elon Musk.

Although the model has low precision rate it displays a high recall score of 0.95 for Elon Musk, indicating that the model successfully detects 95% of all true instances of Elon Musk. The high recall suggests the efficacy of the model in detecting all instances of Elon Musk. This high recall rate signifies model's ability to detect Elon Musk. However, the low precision indicates a significant number of false positives leading to inaccuracies and inadequacies in real world application to detect Elon Musk.

Overall the YOLOv7 model demonstrates strong efficacy to detect Elon Musk due to a high recall rate on the other hand its performance is limited by low precision rate. The model needs improvement in reducing false positive predictions. Enhancement in precision would increase the model's accuracy and reliability in identifying Elon Musk, making it more able for practical usage in real world scenarios.

5. Hammer:

The YOLOv7 model's performance shows a balance between precision and recall to detect hammer. The precision score is 0.287 which indicates this model successfully detects the images of this class object 28.7% of the time, suggesting a moderate false positive rate for this class, some instances are wrongly identified as hammers.

The low recall score for hammers which is 0.269 indicates that it detects with an efficacy of only 26.9% of hammer within the dataset. This low recall suggest that the model misses a large number of real images of hammer. Indicating incomplete detection.

However, in spite of balanced performance of the model to detect hammer, it struggles to correctly identify and detect instances of this class. While achieving a moderate precision the model tries hard to minimize false positive predictions. Even though the low recall highlights the model's capability to recognize and detect hammers within images.

The YOLOv7 model's performance for hammers highlights challenges in efficient object detection. The model shows that the precision-recall, data quality, model design and optimization strategies is vital to improve overall models' efficacy in detecting hammers and other objects in diverse image datasets.

6. Handgun:

The YOLOv7 model's performance for the class "Handguns" gives a precision score of 0.461 which indicates that this model successfully detects the images of this class object 46.1% of the time. This data indicates a moderate rate of false positives. On the other hand, the low recall score of 0.342 suggests that the model detects only 34.2% of all true instances of handguns. This indicates that the model misses a Significant portion of handguns. This indicates incomplete detection.

Although the model shows a balanced performance to detect handguns it still struggles to precisely identify the instances of this class. While achieving a moderate precision the model shows great effort to minimize false positives.

On the contrary a low recall rate highlights the model's capability to recognize and detect handguns.

The YOLOv7 model's performance for Handguns encounters difficulty in efficient object detection. The model shows that the precision-recall, data quality, model design and optimization strategies is essential to improve overall models' efficacy in detecting Handguns and other objects in diverse image datasets.

7. Knife:

The YOLOv7 model's performance for the class “Knife” gives a precision score of 0.456 which indicates that this model successfully detects the images of this class object 45.6% of the time. This data indicates a moderate rate of false positives. On the other hand, the low recall score of 0.296 suggests that the model detects only 29.6% of all true instances of knives. This indicates that the model misses a Significant portion of Knife. This indicates incomplete detection.

However, in spite of balanced performance of the model to detect Knife, it struggles to correctly identify and detect instances of this class. While achieving a moderate precision the model tries hard to minimize false positive predictions. Even though the low recall highlights the model's capability to recognize and detect Knife within images.

The YOLOv7 model's performance for Knife highlights challenges in efficient object detection. The model shows that the precision-recall, data quality, model design and optimization strategies is vital to improve overall models' efficacy in detecting Knife and other objects in diverse image datasets.

8. LeonardoCaprio:

The YOLOv7 model's performance concerning the "LeonardoCaprio" class displays the precision score as 0.367 which indicates that this model successfully detects the images of this class object 36.7% of the time. This indicates a moderate rate of false positive predictions where "LeonardoCaprio" is correctly detected.

On the other hand, the recall score for Leonardo DiCaprio is 0.724 which indicates that this model successfully detects the images of this class object 72.4% of the time, for all true instances of dataset. This high recall value indicates the efficacy of the model to identify instances of Leonardo DiCaprio. This moderate precision and high recall performance indicate minimal false positives. Overall the YOLOv7 model's performance for Leonardo DiCaprio effectively detects specific individuals within Image dataset. By achieving a certain balance, we see the model has achieved certain efficacy in detection of the tasks.

9. MarilynMonroe :

In detection of the class "MarilynMonroe" using the YOLOv7 model indicates that the precision score is at 0.75, which indicates that this model successfully detects the images of this class object 75% of the time. Suggesting a low false positive rate where the Marilyn Monroe is identified incorrectly.

On the other hand, the recall rate of Marilyn Monroe is rather impressive standing at 0.722. This high recall rate suggests that the model has high efficacy rate in detecting all true instances of Marilyn Monroe. This combination of the precision with impressive recall indicates the efficacy and competency of the YOLOv7 model. The high precision gives high accuracy and the minimizes the rate of false positives. The YOLOv7 model shows high accuracy in detecting all instances of Marylin Monroe contributing in overall proficiency in object detection within the dataset.

10. WillSmith:

The YOLOv7 model's performance for the class “WillSmith” gives a precision score of 0.62. This data indicates a moderate rate of false positives. On the other hand, the low recall score of 0.429 suggests that the model detects only 42.9% of all true instances of WillSmith. This indicates that the model misses a Significant portion of WillSmith. This indicates incomplete detection.

However, in spite of moderate precision of the model to detect WillSmith, it struggles to correctly identify and detect instances of this class. While achieving a moderate precision the model tries hard to minimize false positive predictions. Even though the low recall highlights the model's capability to recognize and detect WillSmith within images.

The model suggests a room for improvement in recall, while achieving a fairly high precision is necessary, a higher recall would confirm more inclusive detection of Will Smith instances within the dataset.

The YOLOv7 model's performance for Will Smith highlights its ability to detect individuals with moderate precisions but also indicates limitations in its ability to inclusively capture all instances of this class. By improving the model's recall, we can work on improving the overall performances and efficacy of the model.

Overall Interpretation:

The results show that the YOLOv7 model shows a mixed performance across all classes some showing strong precision and recall, while others showing weak performance and limitation. Classes like "MarilynMonre" and "BillGates" display high precision and recall. Signifying accurate detection of all occurrences with in the dataset. The classes like "Axe" and bottle" displays weak performance displaying lower precision and recall. Signifying incorrect detection of all occurrences with in the dataset.

The mean average precision (MAP) values at IoU thresholds of 0.5 and across thresholds from 0.5 to 0.95 ranging from 0.0407 to 0.012 shows weaker performance moreover in rigorous IoU. These low MAP values highlight the model's restrictions in accurately localizing and detecting objects across various classes. The MAP score highlights the model's inclusive efficacy in object detection.

Overall, while the YOLOv7 model exhibits moderate efficiency across diverse classes. To improve its overall performance, targeted improvements in areas such as data quality, model architecture, and training strategies are vital. By addressing these factors and iteratively refining model, its efficacy, accuracy and reliability in object detection can be greatly enhanced, leading to more effective real-world usability.

The results of the YOLOv7 object detection are displayed in figure 6.2. The model has comparatively performed better than YOLOv5. It is able to detect Bill Gates but has poor accuracy with false positive results. It is able to detect Handgun, Knife and other objects but still it is not sufficient enough for our research because of low accuracy.

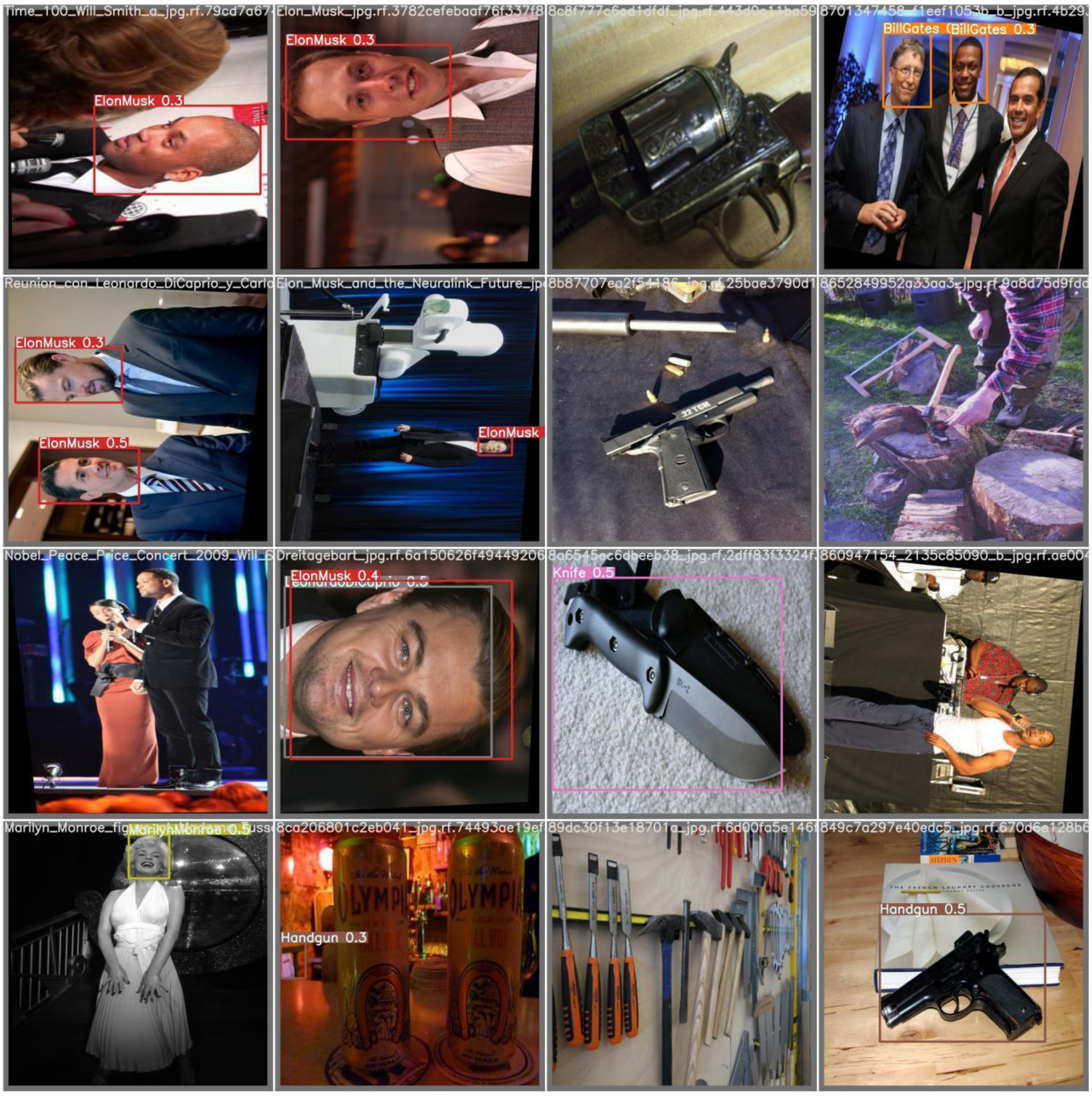

Fig. 6.2 : Result of YOLOv7 model on Image

6.3 YOLOv8 RESULTS

The table 6.3 summarizes the performance metrics of the YOLOv8 object detection model across different classes and overall, for our dataset.

Table 6.3 : Result of YOLOV8 object detection model

CLASS	IMAGES	LABELS	P	R	MAP@0.5	MAP@0.5:0.95
all	310	480	0.829	0.75	0.8	0.596
Axe	310	29	0.622	0.586	0.568	0.451

CLASS	IMAGES	LABELS	P	R	MAP@0.5	MAP@0.5:0.95
BillGates	310	32	0.951	1	0.992	0.61
Bottle	310	179	0.731	0.257	0.379	0.234
ElonMusk	310	20	0.861	1	0.99	0.696
Hammer	310	26	0.71	0.538	0.626	0.548
Handgun	310	40	0.901	0.683	0.804	0.613
Knife	310	54	0.586	0.629	0.674	0.56
LeonardoCaprio	310	29	1	0.94	0.992	0.783
MarilynMonroe	310	29	0.979	0.966	0.993	0.698
WillSmith	310	42	0.952	0.905	0.98	0.765

Class-Specific Interpretation:

1. **Axe:**

There is significant improvement in YOLOv8 with a precision value of 0.622. The model predicts all instances of the axe correctly with a significant reduction in false positive predictions as compared to the previous versions of YOLO. This indicates higher rate of accuracy in detecting axe objects.

The recall value of 0.586 increases the efficacy of the model to detect accurately 58.6% of all true instances of axes within the dataset. There is still some possibility for improvement in order to achieve a higher recall rate. The model indicates a significant transformation in enhanced recall and successful object detection. The Mean Average Precision (MAP) scores at IoU thresholds of 0.5 (0.568) and across thresholds from 0.5 to 0.95 (0.451) indicates enhanced performance for the "Axe" class making the model reliable in detecting axe across various images.

Overall, the YOLOv8 model's improved precision and recall metrics for the "Axe" class suggest higher accuracy and with minimal false positive predictions. YOLOv8 gives improved detection coverage compared to previous versions. More optimization and fine-tuning could possibly lead to higher efficacy, but the current results indicate significant development in object detection for this class.

2. BillGates:

Having a precision value of 0.951 this YOLOv8 model shows extra ordinary development. There is extraordinary development in YOLOv8 with the model predicts all occurrences of the Bill Gates correctly with a minimal false positive prediction as compared to the previous versions of YOLO. This indicates notable accuracy in detecting Bill Gates in images correctly.

The perfect recall of 1.0 indicates that model detects all instances correctly without missing on any true positives. This highlights the model's efficacy to detect Bill Gates across the dataset.

The Mean Average Precision (MAP) scores highlights the model's significant performance. With a MAP of 0.992 at IoU threshold 0.5 and 0.61 across thresholds from 0.5 to 0.95, the model achieves near-perfect identification. This model stands superior with astonishing precision and recall metrics for the "BillGates" class as compared to previous versions. The model highlights accuracy and dependability in detecting Bill Gates without losing precision and recall. This makes a treasured tool for the real-world people detection applications.

3. Bottle:

The YOLOv8 model detects bottle with a moderate precision and low recall with a precision value of 0.731 and successful identification 73.1% of the time. The results indicate some false positive and scope of improvement.

The recall value is 0.257, this low recall rate indicates that the model fails to detect many instances of the bottle, resulting in incomplete detection.

The MAP is 0.379 and MAP 0.5 to 0.95is 0.234 which indicates poor efficacy. This displays the incompetency of the model to accurately detect bottles in images. Overall the YOLOv8 model displays decent precision for the "Bottle" class. However, the low recall suggests major challenges in effectively detecting the bottles.

Addressing factors such as data quality, model architecture, and training strategies could help improve the model's performance and enhance its ability to accurately identify bottles in diverse image datasets.

4. ElonMusk:

The YOLOv8 model performs exceptionally in identifying instances of Elon Musk in images as it achieves a high precision of 0.861 for the "ElonMusk" class with the successful identification 86.1% of the time, suggestive off rare false positives. The model also displays a perfect recall of 1.0 signifying successful detection of all true instances of ElonMusk Class. It often does not miss on the ElonMusk Image detection with a MAP of 0.99 at IoU threshold 0.5 and 0.696 across thresholds from 0.5 to 0.95.

The high level of performance of YOLOv8 model efficacy in precisely detecting people and it's recall for the "ElonMusk" class, indicates its strength and dependability in identifying all instances of Elon Musk within varied image datasets. This is vital for various applications.

5. Hammer:

The YOLOv8 model shows dependability to detect all instances of hammer class with precision of 0.71. The model correctly detects approximately 71% of the time to achieve a precision of 0.71 for the "Hammer" class, indicating that when it predicts the presence of a hammer in an image, it is correct. The precision indicates low failure rate with reliability in correctly recognizing this class However, the recall is 0.538, which shows that the model only identifies approximately 53.8% of all true instances of this class. This comparatively low recall rate shows that the model fails to identify hammers in a large number of actual instances which leads to incomplete detection.

The MAP scores of 0.626 at IoU threshold 0.5 and 0.548 across thresholds from 0.5 to 0.95, the model shows moderate accuracy.

This model's high precision and a low recall rate for hammers, indicates scope of improvement. Enhancements in data quality, model architecture, and training strategies may help in improving the model's performance for the "Hammer" class.

6. Handgun:

The yolov8 model shows proficient object detection for the class Handguns, with a Precision score of 0.901 and successful identification 90.1% of the time. This high precision score indicates high proficiency and minimal false positive predictions.

The moderate recall score 0.683 suggest that the model successfully identifies approximately 68.3% of all actual instances of handguns indicating that the model may miss some objects. However, it demonstrates tremendous ability to detect Handguns across dataset.

The high MAP score 0.804, validates this model's efficacy and accuracy in identifying handguns. The MAP score and precision-recall balance, showcases the model's reliability and consistency. The satisfactory recall rate highlights minimal false positives.

7. Knife:

The precision score of 0.586 is improved compared to earlier versions, which indicates that it correctly identifies knife 58.6% of the time. This suggest that the model is more efficient in minimizing false positive predictions.

The Recall score of 0.629 indicates that the model successfully detects 62.9% of all true instances. This high recall signifies the high sensitivity of the model to detect knifes while missing few true positives.

The moderate MAP score 0.674, validates this model's efficacy and accuracy in identifying knifes. The MAP score and precision-recall balance, showcases the model's reliability and consistency. The high recall rate highlights minimal false positives.

8. LeonardoCaprio:

The YOLOv8 model achieves a flawless precision score of 1.0 for detecting the Leonardo DiCaprio class object. This model does not miss any instance of Leonardo DiCaprio. The model has no false positives which makes it extremely precise and consistent.

The model's high sensitivity is achieved with a recall score of 0.94. The model misses minimum instances while detecting 94% images of Leonardo DiCaprio correctly.

The excellent MAP score 0.992, and outstanding precision-recall, showcases the model’s high performance and consistency. The model’s excellent performance makes it flawless for real world usage.

9. MarilynMonroe:

The YOLOv8 model shows a high Precision score of 0.979 for this class MarilynMonroe by identifying it correctly 97.9% of the time. This high precision score indicates minimal false positive predictions.

With a recall score of 0.966, the Marilyn Monroe class item can be identified with high accuracy 96.6% of the time. This high recall indicates that the model has a high sensitivity to identify Marilyn Monroe with the fewest possible missed occurrences.

The exceptional MAP score 0.993and outstanding precision-recall, makes the model perfect for real world usage.

10. WillSmith:

The YOLOv8 model shows proficient object detection for the class WillSmith, with a Precision score of 0.952 and successful identification 95.2% of the time. This high precision score indicates high proficiency and minimal false positive predictions.

The recall score of 0.905 shows a high accuracy of detecting Leonardo DiCaprio class object 90.5% of the time. This high recall signifies the high sensitivity of the model to detect Leonardo DiCaprio while minimizing the number of missed instances.

The high MAP score 0.98, validates this model's accuracy and reliability in identifying Marilyn Monroe. The high MAP score and outstanding precision-recall, showcases the model’s reliability and consistency. The model’s high performance makes it perfect for real world usage.

Overall Interpretation:

The overall interpretation of the YOLOv8 model's performance indicates its significant developments in object detection accuracy and efficiency compared to previous versions. The model demonstrates exceptional precision and recall values across all classes, which indicates that it is able to accurately detect objects in the dataset.

The impressive MAP score further validates the model’s reliability and consistency. However, some classes with lower recall values such as Bottle need improvements.

In conclusion, the YOLOv8 model's remarkable performance metrics, including high precision, recall, and MAP scores, highlight its effectiveness in accurately detecting and localizing objects of interest. While there are areas for improvement, the model's overall advancements signify a promising direction for the future of object detection technology.

The result of YOLOv8 model is displayed on the below image in figure 6.3. This image clearly shows that YOLOv8 performance is better than the yolov5 and yolov7. It is able to detect the objects correctly with more accuracy compared to previous models.

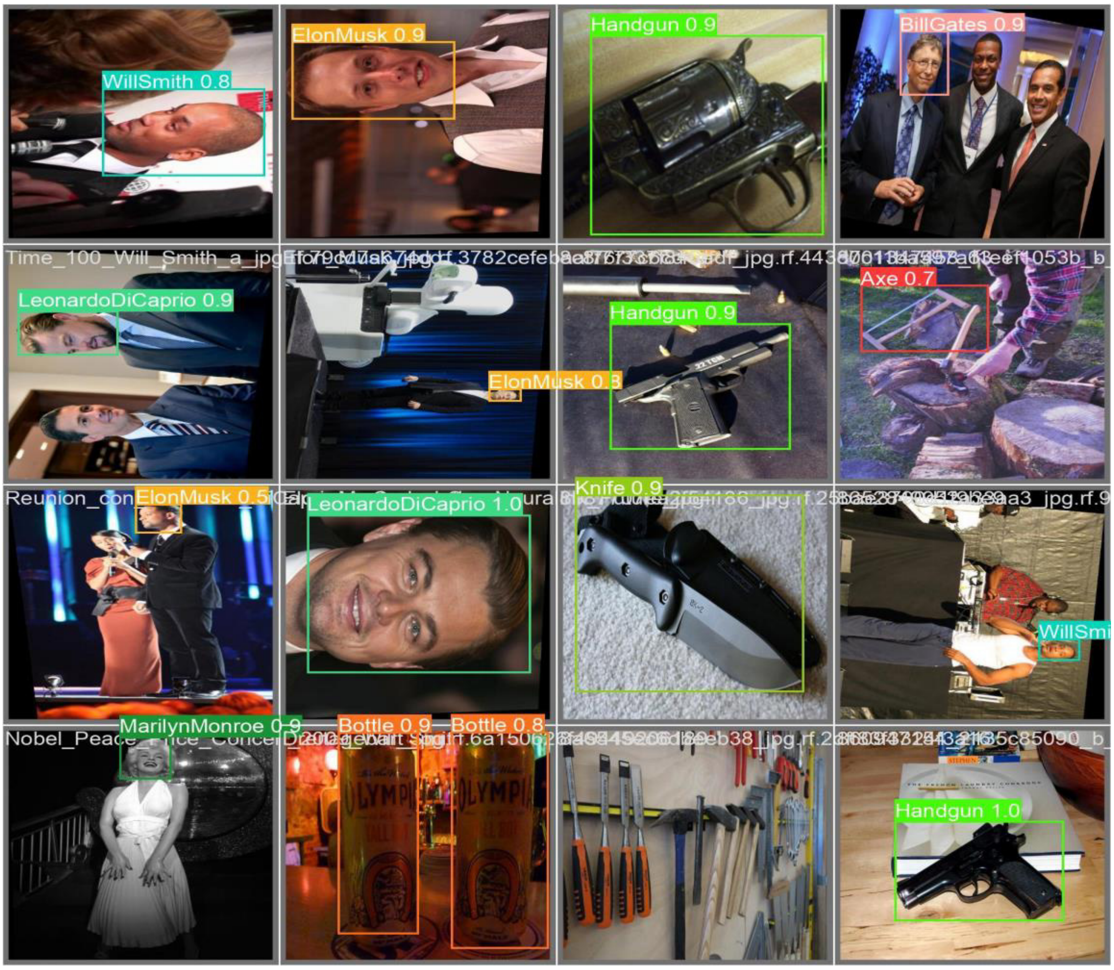

Fig. 6.3 : Result of YOLOv8 model on Image

6.4 COMPARATIVE ANALYSIS OF YOLOv5, YOLOv7 AND YOLOv8 RESULTS

Here is the comparative analysis of YOLOv5, YOLOv7, and YOLOv8 models trained on our custom dataset of 1550 images after pre-processing and augmentation.

The models are evaluated based on different evaluation metrics: Precision, Recall, F1 Score, mAP@0.5, and mAP@0.5:0.95. This comparison of 3 different versions of YOLO provides valuable insights and helps us take decision in real world scenarios. This class specific comparison also brings to light strengths and weaknesses of each versions

The YOLO (You Only Look Once) series have revolutionized object detection in the real-world scenario. Each succeeding version is aimed to improve upon previous versions.

Table 6.4 : Comparison of the various YOLO versions

	YOLOv5	YOLOv7	YOLOv8
Precision	0.37	0.435	0.829
Recall	0.246	0.474	0.75
F1 Score	0.296	0.454	0.788
mAP@0.5	0.0803	0.432	0.8
mAP@0.5:.95	0.0407	0.238	0.596
Completion time((in hours)	2.683	2.803	3.128
Weight(MB)	173.2	74.9	136.7
No. of Parameters	86278375	37245102	68133198

Based on Table 6.4, various metrics, characteristics and performance are compared and evaluated of YOLOv5, YOLOv7, and YOLOv8 models. Amongst the three versions, YOLOv8 is most advanced with its performance as compared to other versions. High Precision and minimal false positives make this version stand out. In contrast, YOLOv5 shows worst performance with its low precision scores. YOLOv7 gives moderate performance which is better than YOLOv5 but not proficient than YOLOv8. YOLOv8 demonstrates highest recall, F1 Score making it best model for capturing more true positives. This makes the model more dependable.

The mAP score of YOLOv8 outperforms both YOLOv5 and YOLOv7 enabling the model to detect objects with higher precision at a lower IoU threshold.

However, YOLOv8 has longest completion time than other versions which is indicative of more complex computations. YOLOv7 has the smallest model size 74.9MB preceding by YOLOv8 i.e. 136.7MB. The largest model size is of 173.2MB which is of YOLOv5.

The YOLOv7 model achieves high efficacy with a more compact model architecture. In terms of number of parameters, YOLOv8 falls between the two.

Overall YOLOv8 outperforms the other two models in terms of all the evaluation metrics.

Table 6.5 : mAP50 Performance of Individual class using YOLOv5, YOLOv7 and YOLOv8 models.

CLASS	YOLOV5L	YOLOV7L	YOLOV8L
all	0.0803	0.432	0.8
Axe	0.0261	0.175	0.568
BillGates	0.113	0.576	0.992
Bottle	0.0124	0.144	0.379
ElonMusk	0.192	0.712	0.99
Hammer	0.00839	0.158	0.626
Handgun	0.0252	0.325	0.804
Knife	0.0191	0.296	0.674
LeonardoCaprio	0.167	0.523	0.992
MarilynMonroe	0.0424	0.826	0.993
WillSmith	0.198	0.588	0.98

The Mean Average Precision (mAP) at IoU threshold 0.5 (mAP@0.5) provides a complete measure of the accuracy and localization capabilities of the object detection model. The mAP@0.5 scores of YOLOv5L, YOLOv7L, and YOLOv8L are compared in Table 6.5 to evaluate the performance of each model version.

The mAP@0.5 score of YOLOv5L suggest that the model faces challenges in localizing and detecting objects across classes at specified IoU threshold. However, YOLOv5L gives a higher mAP@0.5 scores for "BillGates" and "MarilynMonroe" class, as compared to the other classes, indicative of better performance.

As compared to YOLOv5L the YOLOv7L shows significant improvements, across all classes, with an mAP@0.5 of 0.432. classes like “BillGates”, “ElonMusk” and “MarilynMonroe” exhibit mAP@0.5 scores indicating of enhanced performance. This brings to light the development of YOLOv7L’s over YOLOv5L.

The highest mAP@0.5 score of 0.8 is achieved by YOLOv8L. This signifies substantial developments in terms of detection, accuracy and localization as compared to the previous versions. There is a consistent high map score across all classes, whereas some classes like "BillGates," "ElonMusk," and "MarilynMonroe" achieve exceptional mAP@0.5 scores. This validates the model’s ability to accurately detect objects with very few missed instances.

In terms of mAP@0.5 score each model starting from YOLOv5L shows improvement but YOLOv8L outperforms all models across classes. The consistent high mAP@0.5 scores across various classes in YOLOv8L makes it a robust and preferred choice for applications demanding high accuracy.

Table 6.6 : mAP50-95: Performance of Individual class using YOLOv5, YOLOv7 and YOLOv8 models.

CLASS	YOLOV5L	YOLOV7L	YOLOV8L
all	0.0407	0.238	0.596
Axe	0.012	0.0769	0.451
BillGates	0.0531	0.319	0.61
Bottle	0.00273	0.0503	0.234
ElonMusk	0.118	0.425	0.696
Hammer	0.00252	0.0637	0.548
Handgun	0.00768	0.121	0.613
Knife	0.0043	0.152	0.56
LeonardoCaprio	0.0938	0.357	0.783
MarilynMonroe	0.016	0.442	0.698
WillSmith	0.0968	0.371	0.765

The Mean Average Precision (mAP) at IoU threshold 0.5 to 0.95 (mAP50-95) provides a more rigorous evaluation of the object detection model's performance, considering a wider range of IoU thresholds. Comparing the mAP50-95 scores of YOLOv5L, YOLOv7L, and YOLOv8L in Table no. 6.6 reveals the efficiency of each versions in precisely localizing and detecting objects across different classes.

The lowest mAP50-95 score standing at 0.0407 achieved by YOLOv5L indicates the shortfalls of the model in accurately detecting objects with in a wider range of IoU thresholds. Despite some instabilities, classes like "BillGates" and "MarilynMonroe" exhibit higher mAP50-95 scores compared to others. Indicating relatively better performance across multiple IoU.

The YOLOv7L shows improvement with a mAP50-95 of 0.238 across all classes. This indicates better performance in detecting objects across different classes, when considering a wider range of IoU. Certain classes like "ElonMusk" and "WillSmith" demonstrate high mAP50-95 score, suggesting a superior performance detecting certain objects.

YOLOv8L demonstrates the highest mAP50-95 scores with the value of 0.596, signifying major improvements in detecting objects across various classes considering a wider range of IoU thresholds. The YOLOv8L with a consistent high mAP50-95, indicates major developments. YOLOv8L outperforms object detection with minimal missed instances and false positives.

The comparison of the three models from YOLOv5L to YOLOv8L highlights the improvements in comparison of mAP50-95 scores. The substantial developments in model architecture and training methodologies, lead to superior detection tasks.

Table 6.7 : Precision Performance of Individual class using YOLOv5, YOLOv7 and YOLOv8 models.

CLASS	YOLOV5L	YOLOV7L	YOLOV8L
all	0.37	0.435	0.829
Axe	1	0.307	0.622
BillGates	0.107	0.357	0.951
Bottle	0.0359	0.463	0.731

CLASS	YOLOV5L	YOLOV7L	YOLOV8L
ElonMusk	0.0708	0.287	0.861
Hammer	1	0.287	0.71
Handgun	0.132	0.461	0.901
Knife	1	0.456	0.586
LeonardoCaprio	0.116	0.367	1
MarilynMonroe	0.0296	0.75	0.979
WillSmith	0.21	0.62	0.952

Precision measures the number of true positive predictions among all positive predictions made by the model. Comparing the precision values across different classes for YOLOv5L, YOLOv7L, and YOLOv8L shown in Table 6.7 provides insights into the models' accuracy for each class.

YOLOv5L demonstrates a precision score of 0.37, indicating that he model predicts 37% of all instances accurately. In case of certain such as "Axe," "Hammer," "Knife," and "LeonardoCaprio" achieve perfect precision scores of 1. This means that the YOLOv5L model accurately detects all instances without any false positive. The other classes exhibit relatively lower precision values, indicating high amount of false positives.

The precision score for YOLOv7L has seen significant improvement as compared to the previous version. The precision score of 0.435 reflects improved accuracy. Certain classes such as "Axe," "Hammer," "Handgun," and "Knife" maintain high precision scores, indicating minimum false positives. On the other hand classes like "BillGates," "Bottle," "ElonMusk," "LeonardoCaprio," "MarilynMonroe," and "WillSmith" show significant improvement and higher accuracy, as compared to the precision score of YOLOv5L.

The highest and consistent precision among all predecessors is displayed by YOLOv8L with a value 0.829. The values indicate very few false positives and enhanced accuracy as compared to both the previous models. Classes like “BillGates," "ElonMusk," "LeonardoCaprio," "MarilynMonroe," and "WillSmith" have shown significant improvement in performance.

In all YOLOv8L out performs YOLOv7L, YOLOv5L in terms of precision. The consistent improvement in values high light the reliability and effectiveness in accurate predictions making it preferred model for object detection.

Table 6.8 : Recall Performance of Individual class using YOLOv5, YOLOv7 and YOLOv8 models.

CLASS	YOLOV5L	YOLOV7L	YOLOV8L
all	0.246	0.474	0.75
Axe	0	0.0345	0.586
BillGates	0.469	0.831	1
Bottle	0.00559	0.145	0.257
ElonMusk	0.7	0.95	1
Hammer	0	0.269	0.538
Handgun	0.05	0.342	0.683
Knife	0	0.296	0.629
LeonardoCaprio	0.724	0.724	0.94
MarilynMonroe	0.0345	0.722	0.966
WillSmith	0.476	0.429	0.905

Recall measures the proportion of true positive instances that the model correctly identifies among all actual positive instances in the dataset.

From Table 6.8 we can understand that, YOLOv5L has an overall recall of 0.246, giving it an accuracy of 24.6%. This model shows differed recall values across different classes. Certain classes like "BillGates," "ElonMusk," and "MarilynMonroe" achieve relatively high recall values. The high recall value is directly propionate to the efficiency and true positive instances. However, indicating the model effectively captures most true positive instances for these classes. On the other hand, classes such as "Axe," "Hammer," "Handgun," and "Knife" exhibit very low or zero recall values. YOLOv5L misses a significant portion of positive instances for these categories.

In Table 6.8, YOLOv7L shows advancement in overall recall compared to the previous version, with a value of 0.474, indicating better performance in detecting true positive instances. Certain classes like “BillGates," "ElonMusk," "MarilynMonroe," and "WillSmith" achieve relatively high recall values and showing true positive instance. Also, the remainder of the classes showed significant improvement.

The highest recall is seen with the YOLOv8L with a value of 0.75, indicating superior performance detecting all true positive instances across all classes as shown in Table 6.8. This version gives a higher recall value consistently as compared to the previous versions. The classes like "BillGates," "ElonMusk," "MarilynMonroe," and "WillSmith" show enhanced ability of the model to capture most true positive instances.

YOLOv8L out performs the previous versions in terms of precision and object detection. While the YOLOv7L shows improvements in recall compared to the other two, YOLOv8L still make it the most a advanced and accurate for object detection.

Table 6.9 : F1 Score Performance of Individual class using YOLOv5, YOLOv7 and YOLOv8 models.

CLASS	YOLOV5L	YOLOV7L	YOLOV8L
all	0.2955	0.4537	0.7875
Axe	0.0000	0.0620	0.6035
BillGates	0.1742	0.4994	0.9749
Bottle	0.0097	0.2208	0.3803
ElonMusk	0.1286	0.4408	0.9253
Hammer	0.0000	0.2777	0.6121
Handgun	0.0725	0.3927	0.7770
Knife	0.0000	0.3590	0.6067
LeonardoCaprio	0.2000	0.4871	0.9691
MarilynMonroe	0.0319	0.7357	0.9725
WillSmith	0.2914	0.5071	0.9279

The F1 score combines both precision and recall into a single value, providing a balanced insight of a model's performance.

YOLOv5L achieves an overall F1 score of 0.2955, indicates a moderate balance between precision and recall. The model displays varied F1 score for different classes. The classes like "Axe," "Bottle," "Hammer," "Knife," and "LeonardoCaprio" exhibit zero or significantly low F1 scores as shown in Table 6.9. This indicates limitation in the ability of the model to maintain a balance between recall and precision.

When it comes to YOLOv7L we see an overall improvement in the F1 score in Table 6.9, with a value of 0.4537, highlighting an enhanced performance in maintaining a balance between precision and recall. Just as YOLOv5L, certain classes like "BillGates" and "WillSmith" achieve improved F1 scores. On the other hand, classes such as "Axe," "Bottle," "Hammer," "Handgun," "Knife," "LeonardoCaprio," and "MarilynMonroe" display improved F1 scores.

Coming to the latest YOLOv8L we witness highest overall F1 score as compared to all previous versions. The value of 0.7875 indicates superior performance of the model in balancing precision and recall. YOLOv8L consistently achieves higher F1 scores across classes. Remarkably, classes such as "BillGates," "ElonMusk," "MarilynMonroe," and "WillSmith" show substantial improvements in F1 scores compared to previous versions as shown in Table 6.9.

The consistent development in improvements in F1 scores across various classes for YOLOv8L emphasize its dependability and effectiveness in achieving a balanced combination of precision and recall. Making this the most dependable and advanced model for accurate object detection tasks.

Precision Confidence Curve

A Precision-Confidence Curve is a graphical representation showing relationship amongst precision of the model's predictions and confidence scores.

- Precision: Precision measures the proportion of true positive detections (correctly identified objects) out of all positive detections (both true positives and false positives).
- Confidence Score: This is the score assigned by the model to indicate the likelihood that a predicted bounding box contains an object of interest. Higher confidence scores indicate greater certainty in the prediction.

Precision Confidence Curve of Yolov5, Yolov7 and Yolov8 is shown in figure 6.4,6.5, and 6.6 respectively

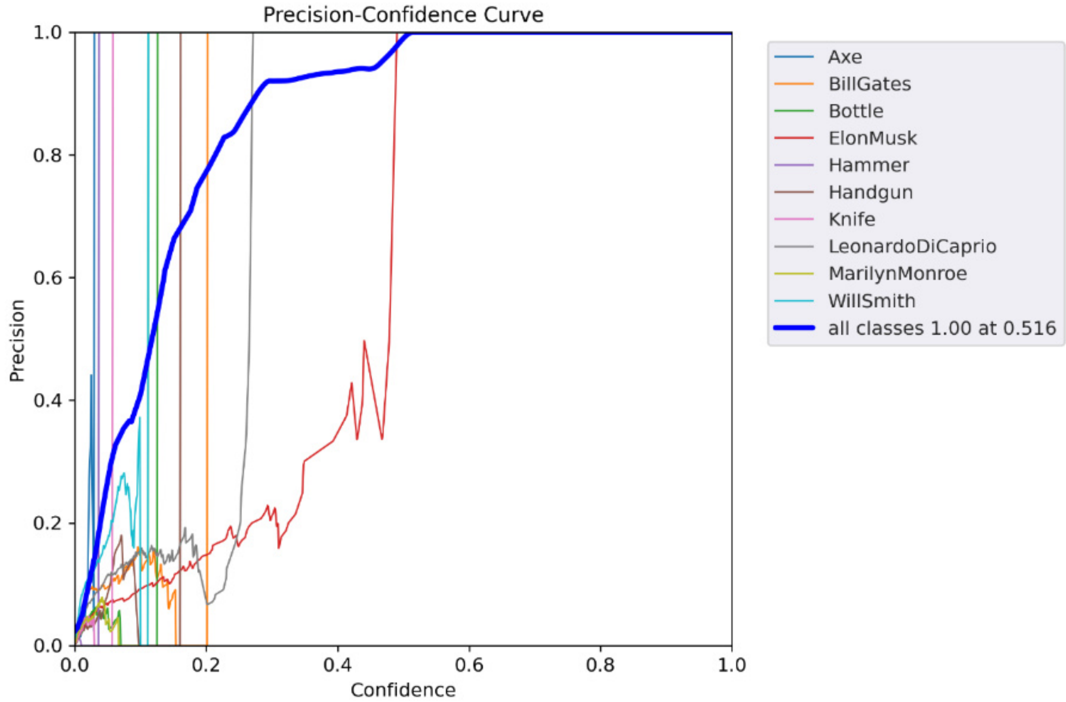

Fig. 6.4 : Precision -Confidence Curve of YOLOv5

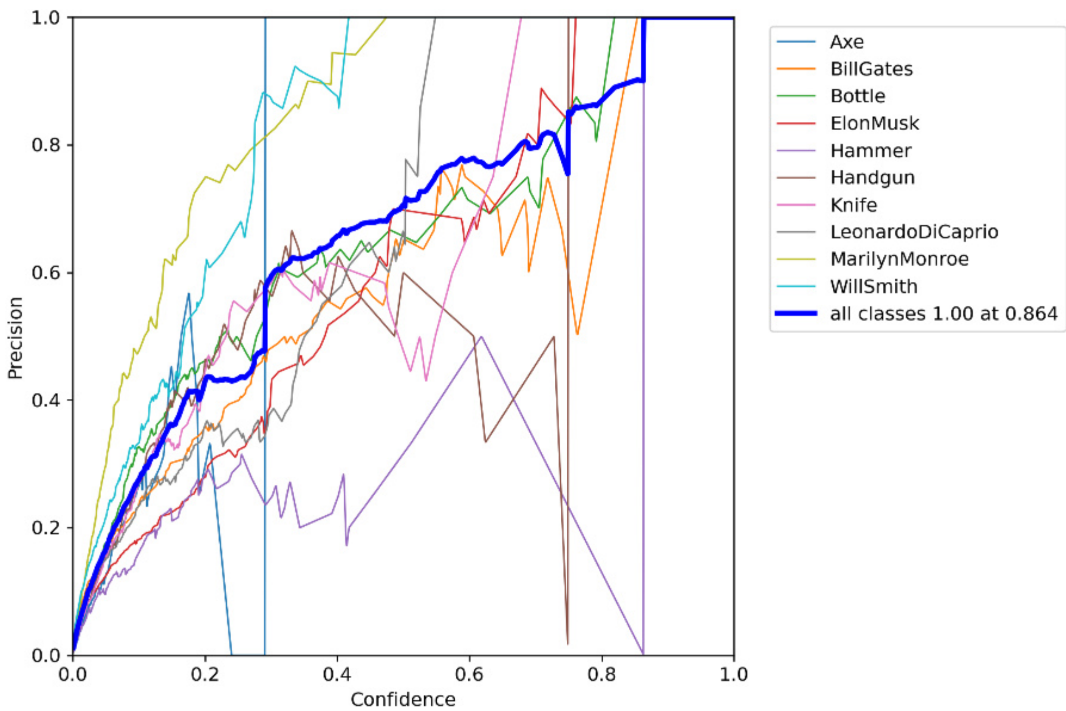

Fig. 6.5 : Precision -Confidence Curve of YOLOv7

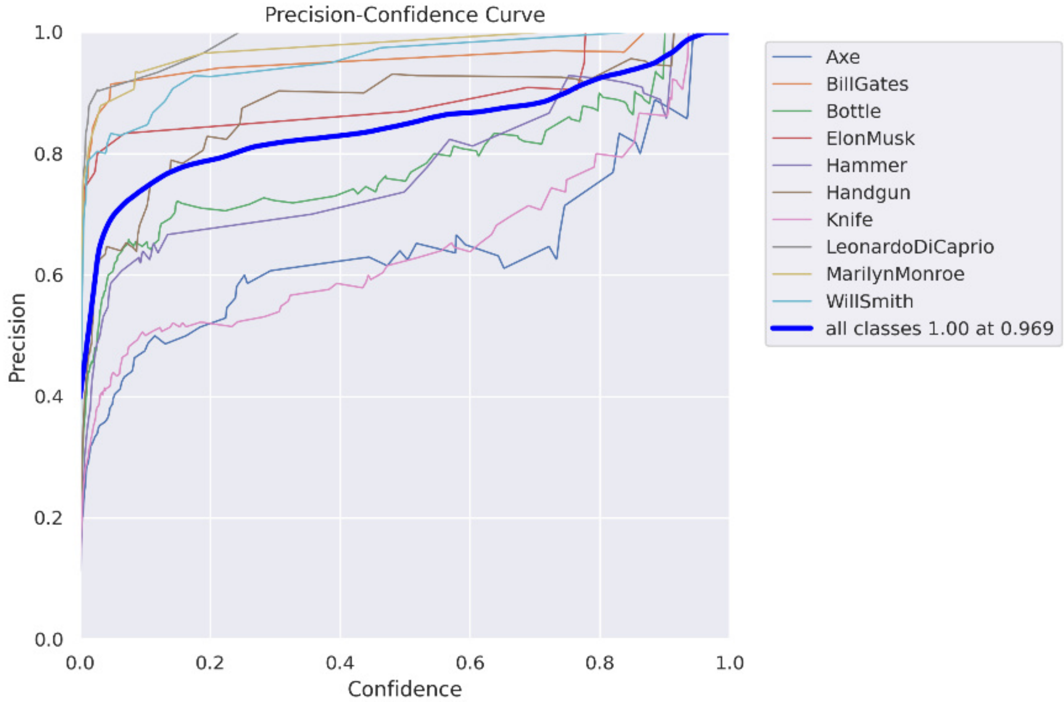

Fig. 6.6 : Precision -Confidence Curve of YOLOv8

When analyzing the precision-confidence curves for YOLOv8, YOLOv7, and YOLOv5, considering that all models achieve a precision of 1.00 at different confidence thresholds, the interpretation is as follows:

Precision-Confidence Curve Overview

1. **Precision**:
   - Precision measures the accuracy of positive predictions made by the model (i.e., the proportion of true positives among all positive predictions).
   - A precision of 1.00 means there are no false positives within the predictions classified as positive above a certain confidence threshold.
2. **Confidence Threshold**:
   - The confidence threshold is the minimum score at which the model considers a prediction to be positive.
   - Higher confidence thresholds mean the model is more selective about which predictions it classifies as positive.

YOLOv8, YOLOv7, and YOLOv5 Precision-Confidence Analysis

YOLOv8 achieves perfect precision (1.00) at a notably high confidence threshold of 0.969 indicating a confident model ensuring minimum false positive above this threshold as shown in Figure 6.6.

Figure 6.5 shows that the YOLOv7 achieves a perfect precision (1.00) with a slightly lower confidence score. Despite lower threshold YOLOv7 maintains high accuracy for positive predictions.

Figure 6.4 displays that the YOLOv5 with a lower confidence threshold of 0.516 with a precision (1.00). The YOLOv5 demonstrates significantly higher tolerance for positive detections while detecting a larger range of prediction confidence.

Model Reliability and Stringency:

In the processes of comparison of YOLOv8, YOLOv7, and YOLOv5, it is discovered that YOLOv8 displays higher threshold for perfect precision. This ensures predictions only when the model is confident. In contrast YOLOv7 operates at a moderate threshold, striking a balance between stringency and reliability, avoiding false positives within broader true positives.

Recall-Confidence Curve

Recall-Confidence Curve in YOLO Object Detection demonstrates how recall varies with different confidence thresholds for an object detection model. Recall measures the proportion of actual positives correctly identified by the model. At a confidence threshold of 0.000, every detected instance, regardless of confidence level, is counted as positive.

In this context, analyzing recall at a confidence threshold of 0.000 for YOLOv5, YOLOv7, and YOLOv8 reveals how well each model captures all possible positives without considering the confidence score. Recall-Confidence Curve of Yolov5, Yolov7 and Yolov8 is shown in figure 6.7,6.8, and 6.9 respectively

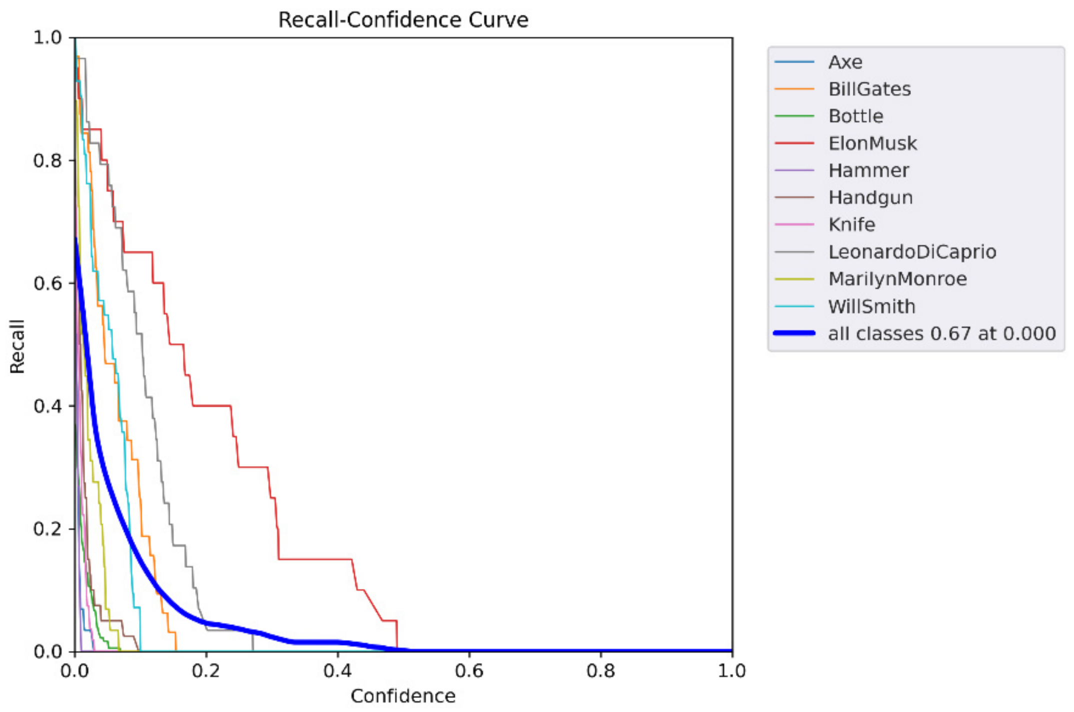

Fig. 6.7 : Recall -Confidence Curve of YOLOv5

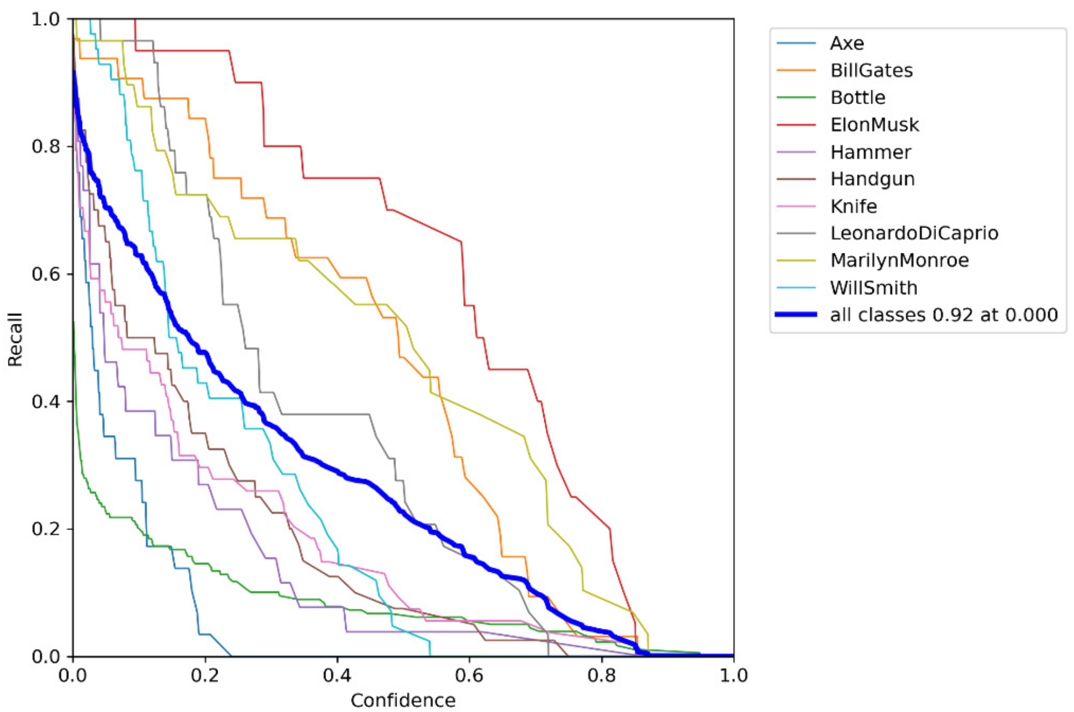

Fig. 6.8 : Recall -Confidence Curve of YOLOv7

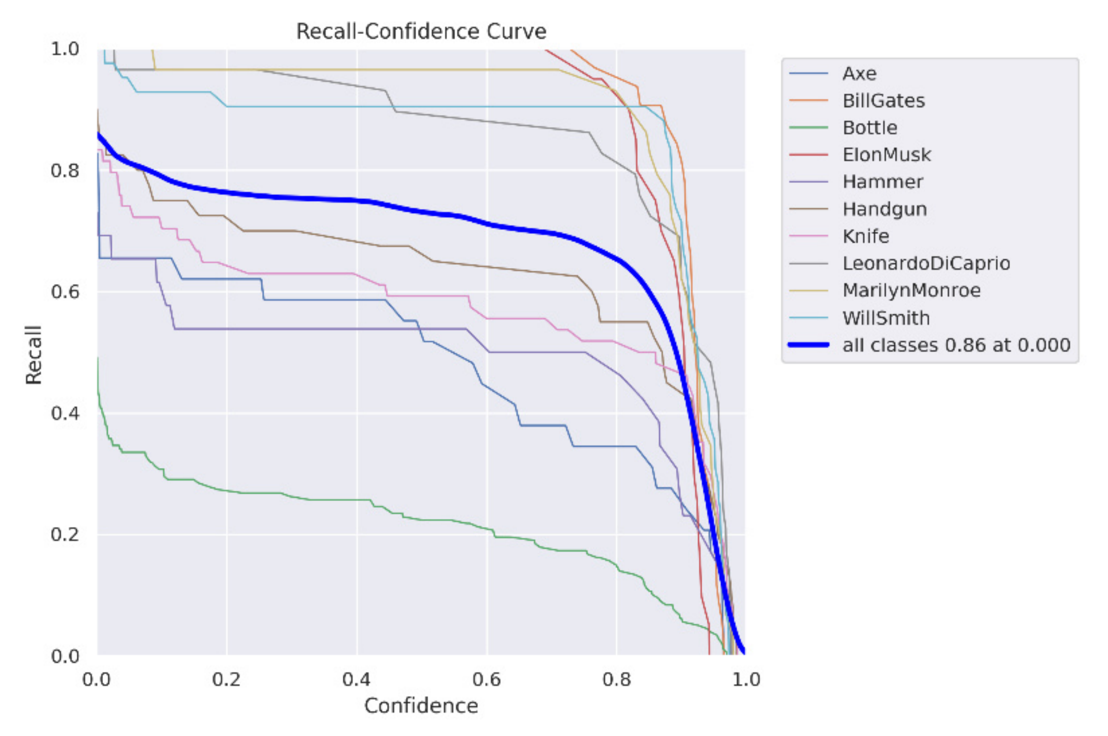

Fig. 6.9 : Recall -Confidence Curve of YOLOv8

In evaluating the recall performance of YOLOv5, YOLOv7, and YOLOv8 across varying confidence thresholds, distinct patterns emerge that influence their suitability for different applications. YOLOv7 stands out with the highest recall rate of 0.92 at a confidence threshold of 0.000 as shown in Figure 6.8, indicating its exceptional ability to capture 92% of all true positive instances without filtering based on confidence. This makes YOLOv7 particularly well-suited for tasks where comprehensive detection of every possible positive instance is essential. Following closely, YOLOv8 achieves a recall of 0.86 under similar conditions as shown in Figure 6.9, showcasing strong performance in identifying a significant portion of true positives while potentially offering more selective detections at higher confidence levels. In contrast, Figure 6.7 shows that the YOLOv5 exhibits a lower recall of 0.67 at the lowest confidence threshold, suggesting it captures a moderate proportion of true positives compared to YOLOv7 and YOLOv8.

Precision-Recall Curve in YOLO Object Detection

The precision-recall curve is a critical tool in evaluating object detection models, illustrating the trade-off between precision (the proportion of true positive detections among all positive detections) and recall (the proportion of true positive detections among all actual positives) for different confidence thresholds. In the context of YOLO models, these metrics reveal the performance of each version in accurately detecting and classifying objects. The Precision- Recall Curve of YOLOv5, YOLOv7 and YOLOv8 is shown in figure 6.10,6.11, and 6.12 respectively.

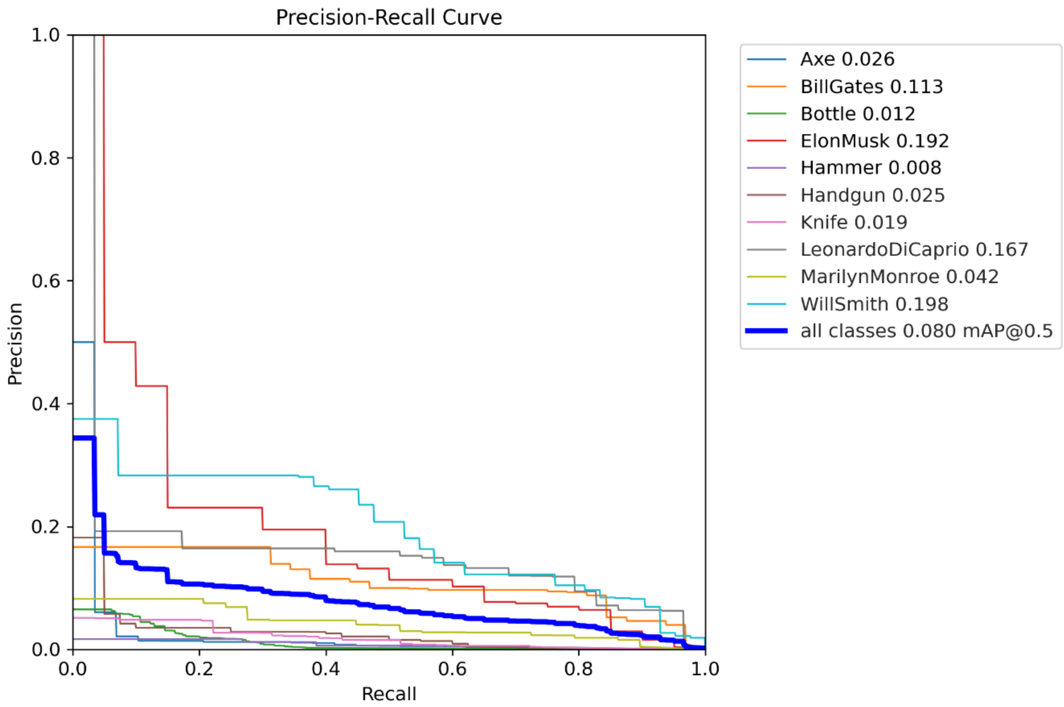

Fig. 6.10 : Precision- Recall Curve of YOLOv5

For YOLOv5, the precision-recall analysis reveals a mAP@0.5 (mean Average Precision at a 0.5 Intersection over Union threshold) of 0.080 as shown in figure 6.10. This low score indicates that YOLOv5 struggles significantly with both precision and recall, resulting in a high number of false positives and missed detections. Essentially, YOLOv5's object detection capabilities are limited, making it less reliable for applications requiring high accuracy.

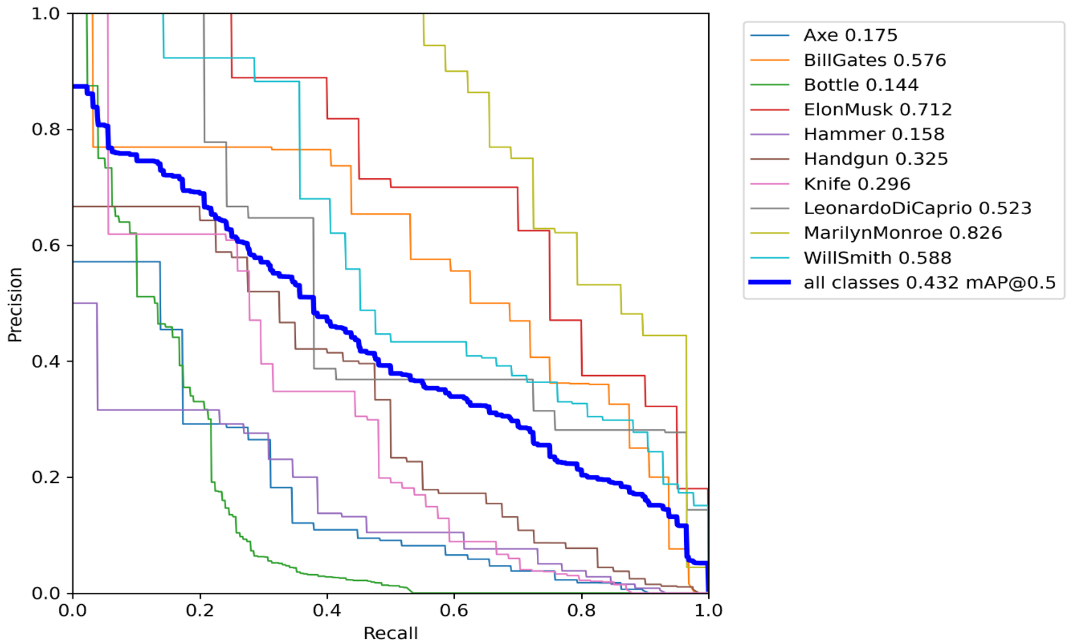

Fig. 6.11 : Precision- Recall Curve of YOLOv7

In contrast, YOLOv7 shows a marked improvement with a mAP@0.5 of 0.432 as shown in figure 6.11. This indicates a moderate level of performance where precision and recall are better balanced compared to YOLOv5. YOLOv7 can detect and classify objects more accurately, leading to fewer false positives and false negatives. This moderate performance suggests that YOLOv7 is suitable for applications where a reasonable level of accuracy is acceptable but not critical.

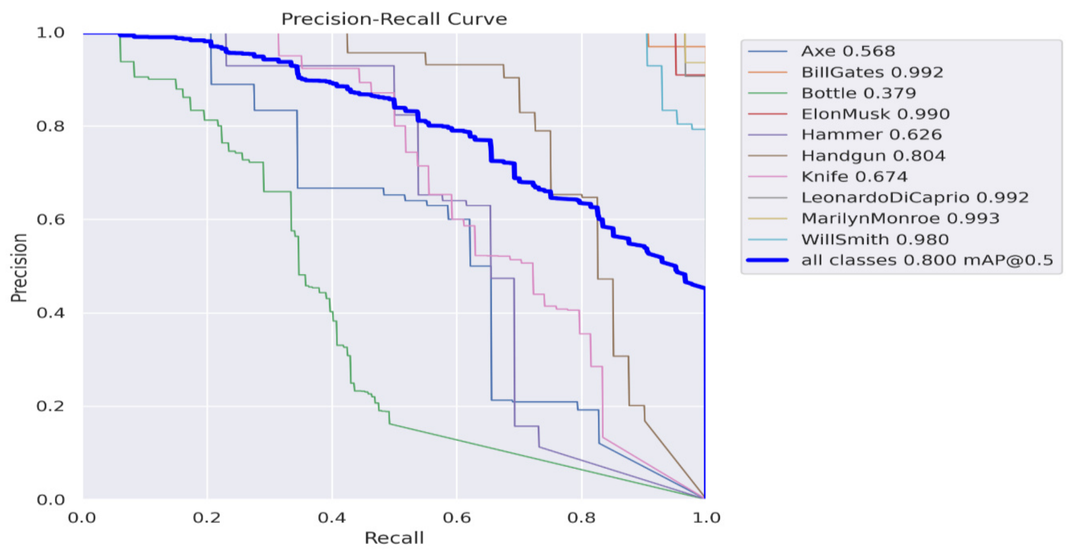

Fig. 6.12 : Precision- Recall Curve of YOLOv8

YOLOv8, however, outperforms both YOLOv5 and YOLOv7, with a high mAP@0.5 of 0.800 as shown in figure 6.12. This high score signifies excellent performance in terms of both precision and recall. YOLOv8 effectively detects and classifies objects with minimal errors, making it the most reliable model among the three. The high precision and recall indicate that YOLOv8 can accurately identify objects with few false positives and false negatives.

In summary, the precision-recall analysis clearly distinguishes the varying strengths of YOLOv5, YOLOv7, and YOLOv8. YOLOv5 shows the lowest performance with significant room for improvement, YOLOv7 offers a better balance and moderate performance, and YOLOv8 excels with the highest accuracy and reliability. These insights are crucial for selecting the appropriate model based on the specific needs of real-world applications.

F1 Confidence Curve

The F1 confidence curve is a vital evaluation tool in object detection models, combining both precision and recall into a single metric. The F1 score is the harmonic mean of precision and recall, providing a balance between the two. The confidence curve shows how the F1 score varies with different confidence thresholds, revealing the model's performance across a range of detection confidences. The F1 -Confidence Curve of Yolov5, Yolov7 and Yolov8 is shown in figure 6.13,6.14, and 6.15 respectively

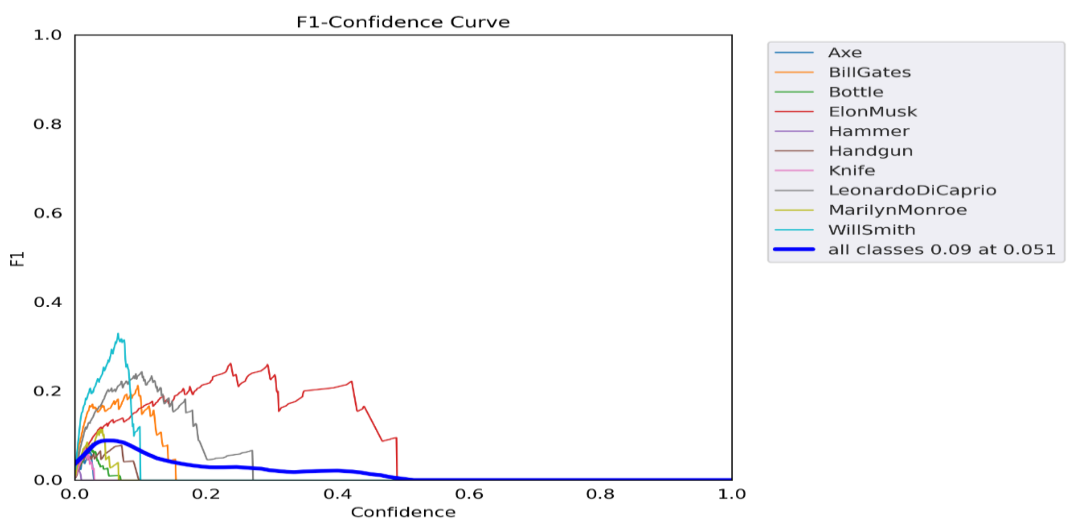

Fig. 6.13 : F1 Confidence Curve of YOLOV5

For YOLOv5, the F1 score at its best is 0.09 at a confidence threshold of 0.051 as shown in figure 6.13. This indicates that YOLOv5 performs poorly in balancing precision and recall, leading to a low F1 score. The low threshold at which this F1 score is achieved suggests that even at minimal confidence levels, the model struggles to identify and classify objects accurately, resulting in a high number of false positives and false negatives. This makes YOLOv5 less suitable for applications where accuracy is crucial.

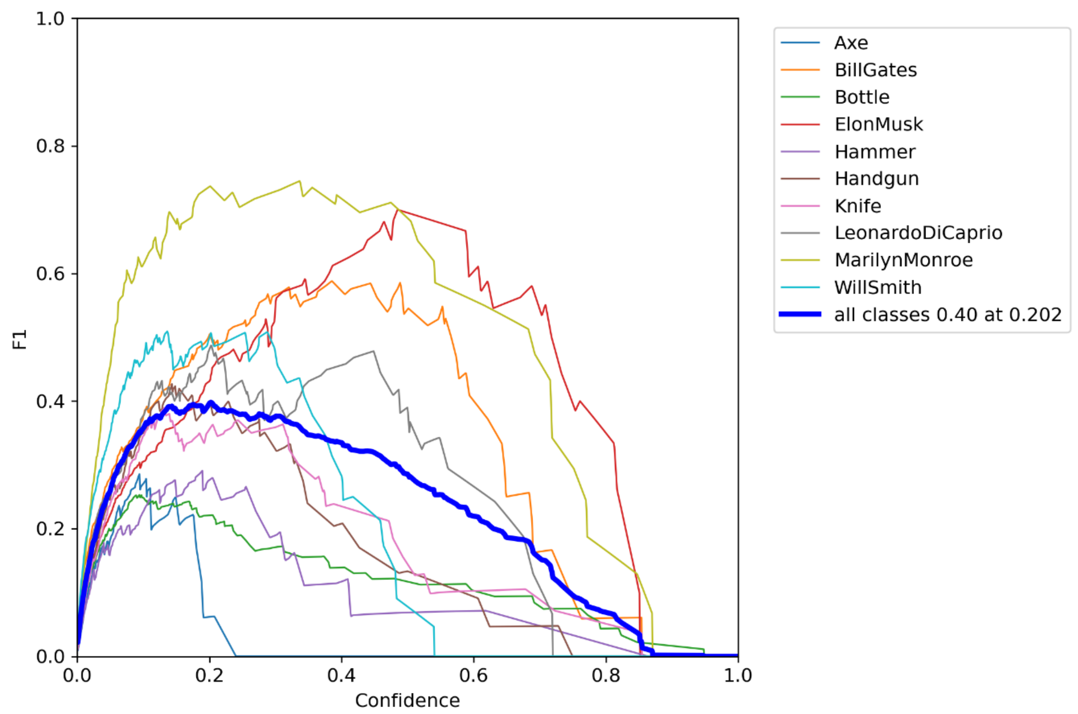

Fig. 6.14 : F1 Confidence Curve of YOLOv7

Figure 6.14 shows a significant improvement in YOLOv7, with an F1 score of 0.40 at a confidence threshold of 0.202. This higher F1 score reflects a better balance between precision and recall compared to YOLOv5. The model performs more reliably across various confidence levels, reducing the number of misclassifications. YOLOv7's performance indicates its suitability for applications where a moderate level of accuracy is acceptable and beneficial.

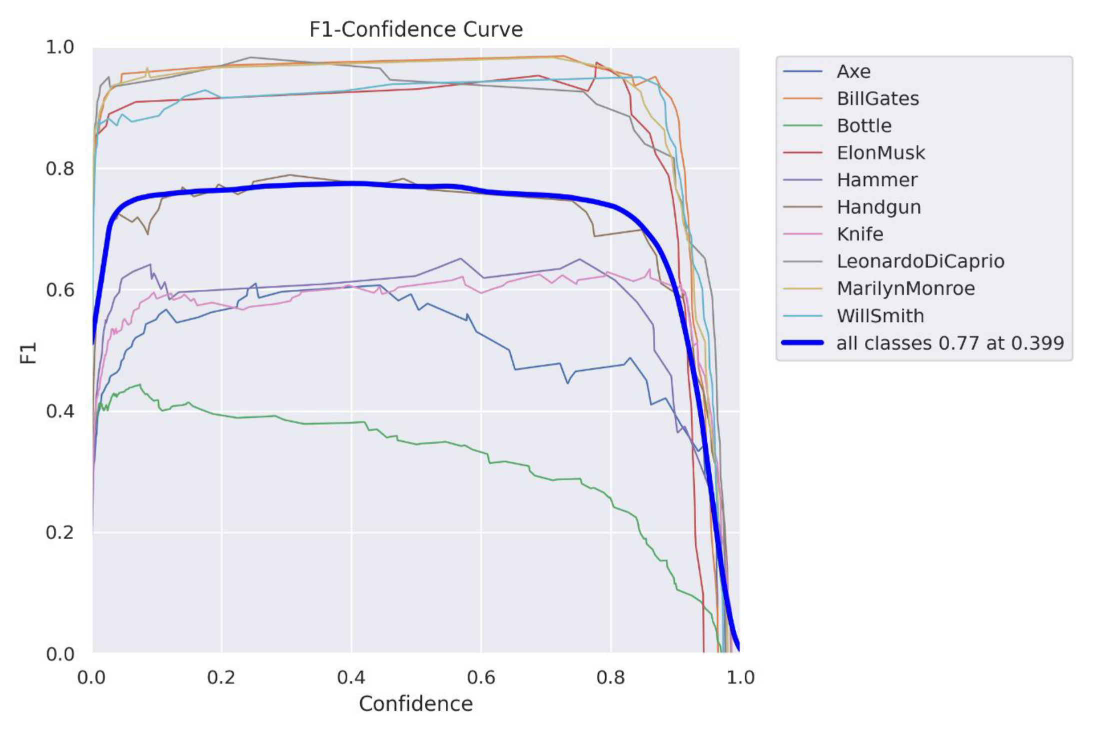

Fig. 6.15 : F1 Confidence Curve of YOLOv8

Through figure 6.15 we can understand that, YOLOv8 demonstrates the best performance, with an F1 score of 0.77 at a confidence threshold of 0.399 . This high F1 score indicates that YOLOv8 excels in achieving a balance between precision and recall, making accurate detections with fewer errors. The higher confidence threshold shows that the model maintains its reliability even as the confidence level increases, making it ideal for applications requiring high accuracy and minimal false detections. The superior F1 score of YOLOv8 underscores its effectiveness in detecting and classifying objects accurately, making it the most reliable choice among the three models.

In summary, the F1 confidence curve analysis highlights the varying capabilities of YOLOv5, YOLOv7, and YOLOv8 in object detection. YOLOv5 shows the weakest performance, YOLOv7 offers moderate improvement, and YOLOv8 provides the best balance of precision and recall, making it the most accurate and reliable model for real-world applications demanding high accuracy and minimal errors.

CONFUSION MATRIX

A confusion matrix for multiple classes provides a detailed breakdown of prediction results. Each row represents the actual class, while each column represents the predicted class. The diagonal elements represent the true positives (correct predictions), and off-diagonal elements represent false positives (misclassifications).

The Confusion Matrix of Yolov5, Yolov7 and Yolov8 is shown in figure 6.16,6.17, and 6.18 respectively

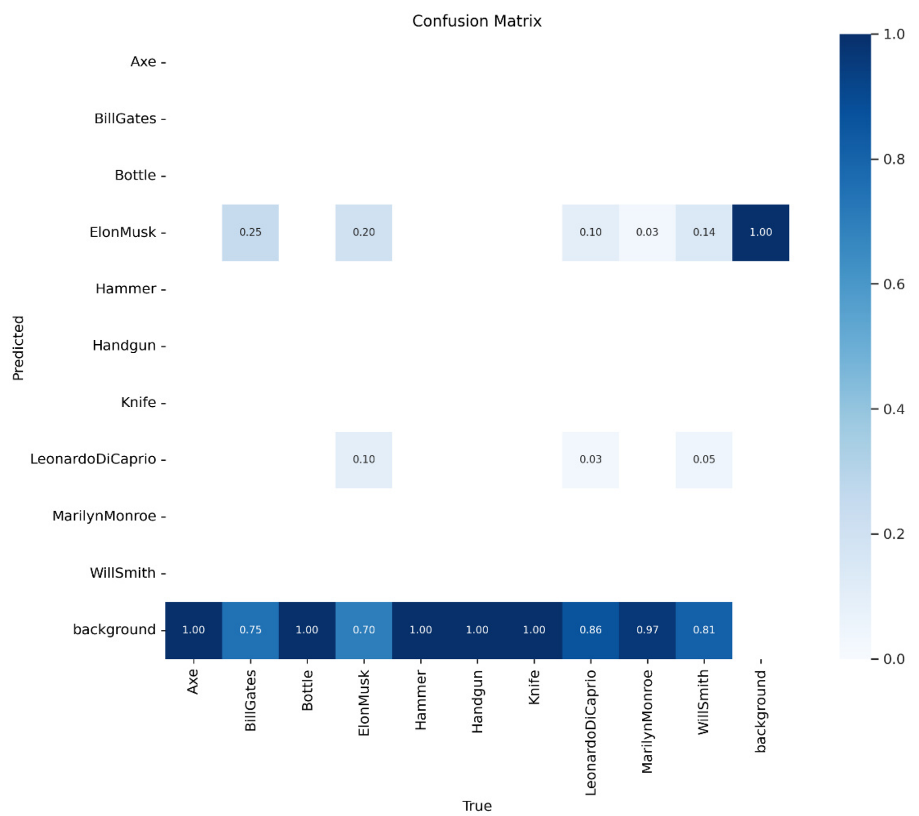

Fig. 6.16 : Confusion Matrix of YOLOv5

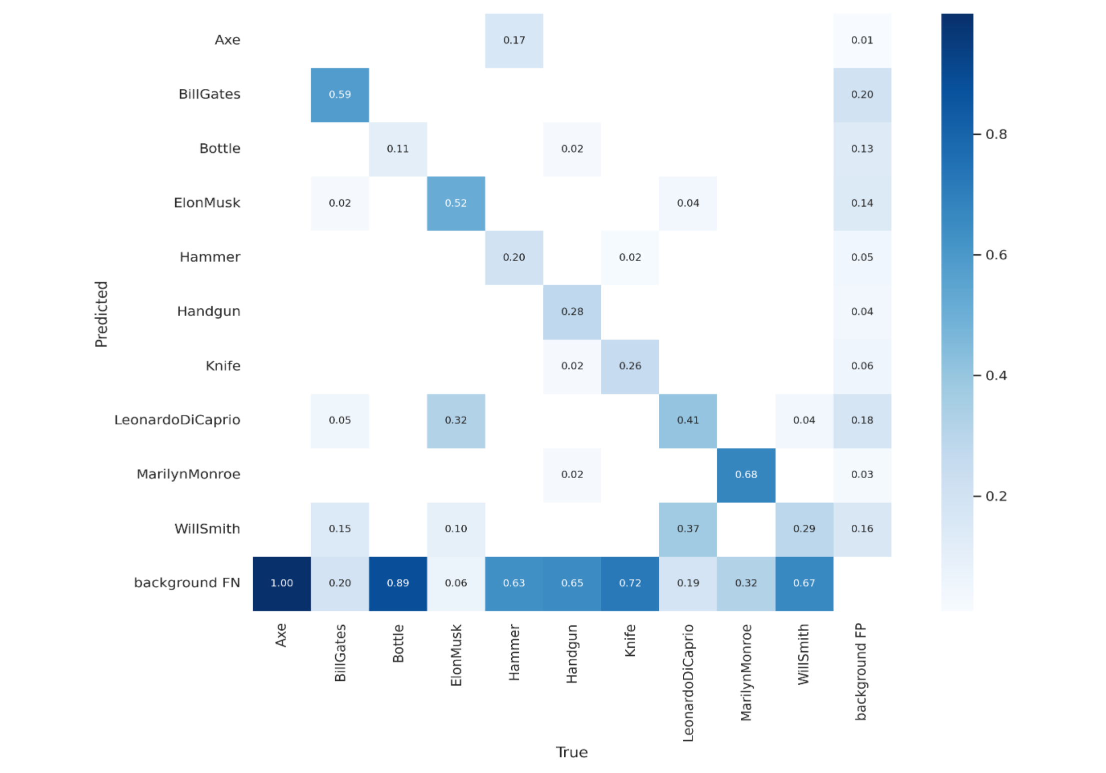

Fig. 6.17 : Confusion Matrix of YOLOv7

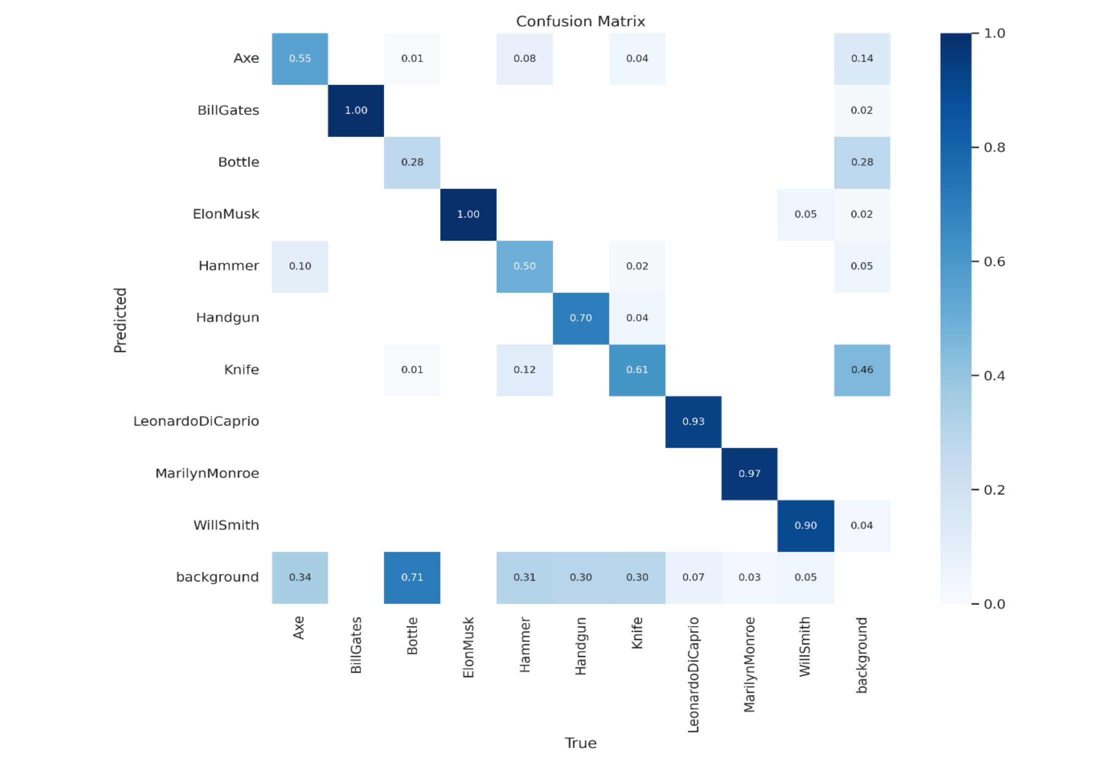

Fig. 6.18 : Confusion Matrix of YOLOv8

From the above figures of confusion matrices of yolov5, yolov7 and yolov8 we can interpret that yolov8 performs the best amongst other two. There are more diagonal elements in YOLOv8 confusion matrix compared to that of other two models.

Conclusion

The comparative analysis of YOLOv5, YOLOv7, and YOLOv8 reveals a clear progression in object detection capabilities, with YOLOv8 demonstrating the highest precision, recall, and MAP scores among the three. YOLOv5 shows moderate performance with balanced but subpar metrics, while YOLOv7 offers improvements over YOLOv5, especially in certain classes, yet falls short of YOLOv8's overall performance. YOLOv8 excels across most classes, though minor inconsistencies remain. The findings highlight YOLOv8 as the most effective version.