**Integrated Approaches for Object Detection and Image Recognition with XAI**

By

K Anusha 22MIS1032

K Divya 22MIS1002

P Koushik - 22MIS1007

M.V. Chandra Gupta - 22MIS1022

Under the guidance of

Dr. Pattabiraman V.

in partial fulfillment of the course

**SWE2011 BIG DATA ANALYTICS**



**School of Computer Science and Engineering (SCOPE)**

**Vellore Institute of Technology**

**Chennai-600127**

**Nov 13, 2024**

****

**BONAFIDE** **CERTIFICATE**

This is to certify that the project work entitled “**Integrated Approaches for Object Detection and Image Recognition with XAI**” this is a Bonafide work of **K Anusha-22MIS1032, K Divya-22MIS1002, P Koushik-22MIS1007 and** **M.V. Chandra Gupta-22MIS1022**. Who carried out the project work under my supervision and guidance for SWE2011 -Big Data Analytics.

**Dr. Pattabiraman V.**

Professor

School of Computer Science and Engineering (SCOPE)

VIT University, Chennai

Chennai – 600 127.

**АCKNOWLЕDGЕMЕNT**

We would like to express our special thanks and gratitude to our Course faculty **Dr. Pattabiraman V.** Professor of School of Computer Science and Engineering who gave us the golden opportunity to do this wonderful project titled “**Integrated Approaches for Object Detection and Image Recognition with XAI** “, and also for providing us with proper guidance and suggestions to develop this project. This helped us in doing а lot of research and we came to know about so many new things.

|  |  |
| --- | --- |
| Title  **Table of Contents** | Page No |
| 1. Abstract | 5 |
| 1. Scope | 6 |
| 1. Objective | 7 |
| 1. Introduction | 8 |
| 1. Literature Survey | 9-11 |
| 1. Tools used | 11 |
| 1. Dataset Description | 12-13 |
| 1. Architecture | 14-15 |
| 1. Proposed works    1. Module 1 –Data Preparation and Model Training    2. Module 2 –Model Creation for Object Identification and Recognition    3. Module 3 \_Explainable Artificial Intelligence (XAI) in YOLOv8    4. Module 4 \_Performance Evaluation and System Integration | 15-21 |
| 1. Innovation | 22 |
| 1. Results and discussion | 23 |
| 1. Conclusion and Future Work | 24 |
| 1. References | 25-27 |

GitHub Link: <https://github.com/kothapalliAnusha/Integrated-Approaches-for-Object-Detection-and-Image-Recognition-with-XAI.git>

**1.ABSTRACT**

This study looks at a new approach to object detection and image identification that combines cutting-edge deep learning models—YOLOv8 for object detection and ResNet50 for image recognition—with Explainable AI (XAI) techniques to ensure transparency in the model's decision-making process. Our program not only recognizes things, but also determines whether the image was taken with a drone, CCTV, or a conventional camera. The ResNet50 model is trained using the CIFAR-10 dataset for image recognition, while a custom-built dataset, consisting of images from various capture sources, is employed for YOLOv8-based object detection. By combining these technologies, the system provides high accuracy in object detection and picture recognition, as well as interpretable findings, both of which are critical in defence and security applications were understanding the context of an image—such as its origin—is essential. The incorporation of XAI improves the system's usability by offering explicit insights into the model's predictions. This method has great promise for improving surveillance systems, autonomous defensive mechanisms, and real-time monitoring applications.

Keywords: object detection, image recognition, YOLOv8, ResNet50, Explainable AI, defence applications, image source classification, CIFAR-10

**2.SCOPE**

This enhances the capabilities of classic computer vision systems by adopting a comprehensive approach that not only detects and recognizes objects inside an image, but also identifies the image's origin, so providing a key layer of contextual intelligence. By combining advanced object detection with YOLOv8 and powerful image recognition with ResNet50, the system can accurately identify objects while also distinguishing whether the image was acquired by a drone, a ground-based camera, or a CCTV system.

This dual functionality extends the system’s application beyond conventional object detection frameworks, making it especially valuable in defence and security domains. In these contexts, determining the source of an image is critical because it adds operational context, allowing analysts to make more informed conclusions about the positioning, timing, and relevance of identified items. The adoption of Explainable AI (XAI) approaches improves the system by assuring transparency in decision-making processes, allowing users to explain and confirm the reasoning behind the model's outputs—which is crucial in mission-critical operations.

As a result, the scope of this project extends beyond object detection and recognition to include a larger situational awareness framework that combines object information with the source of data acquisition. This integrated method allows for a more comprehensive and actionable analysis of visual data, especially in defence and surveillance settings when accuracy, context, and transparency are critical. By addressing these challenges, the system is designed to enhance situational intelligence, supporting decision-making processes in high-stakes environments.

**3.OBJECTIVE**

**1. Object Detection with YOLOv8:**

YOLOv8 can recognize items in photos from many sources such as drones, cameras, or CCTVs. This approach will provide fast and precise detection, ensuring real-time responsiveness for security applications.

**2. Image Source Classification:**

The system will classify the origin of each image, determining whether it was captured by a drone, camera, or CCTV. This adds critical context to object detection, as the source of the image can influence decision-making in defence and surveillance.

**3. Image Recognition using ResNet50:**

We used ResNet50 for accurate recognize and categorize items in the CIFAR-10 dataset. This enhances item identification performance, allowing the system to identify things more accurately in a variety of scenarios.

**4. Implementation of Explainable AI (XAI):**

Using XAI strategies would increase transparency in decision-making processes. Users will understand why the system detected certain objects and classified the source, which is crucial for trust and reliability in mission-critical tasks.

**5. System Performance Evaluation:**

The system will be evaluated based on object detection accuracy, source identification precision, and transparency in decision-making. This ensures that the system not only detects objects accurately but also provides valuable context through source classification.

**4.INTRODUCTION**

In recent years, object identification and picture recognition have become crucial tasks in computer vision, particularly in defence, security, and surveillance applications. These tasks normally involve recognizing and classifying items inside a picture, but in defence settings, the image's origin—whether recorded by a drone, a ground camera, or CCTV—provides vital context that improves operational comprehension of detected objects. Real-time threat assessment, mission planning, and strategic decision-making rely heavily on this source data.

In this paper, we present a novel technique that combines object identification using the sophisticated YOLOv8 model and image recognition with the ResNet50 architecture. Our system's ability to detect objects in an image while simultaneously classifying the picture source, such as whether it was captured by a drone, a ground-based camera, or a CCTV system, is a critical feature. This additional layer of source identification improves situational awareness, making the system better suited to defence applications where context is crucial.

To train our models, we use two datasets: the well-known CIFAR-10 dataset for image recognition tasks with ResNet50, and a bespoke dataset created exclusively for YOLOv8 training. In military settings, where AI-driven assessments may have serious effects, it is critical to understand how and why models reach specific results. XAI assures that the system's outputs are interpretable, dependable, and actionable by human operators, making it ideal for mission-critical defence operations.

By merging cutting-edge object detection, picture recognition, source identification, and XAI, this study provides a comprehensive and resilient solution that improves the accuracy, transparency, and operational relevance of AI systems in defence and security applications.

**5.LITERATURE REVIEW**

The combination of deep learning models and Explainable AI (XAI) techniques resulted significant improvements in object detection and image recognition. Significant research has been carried out to increase detection accuracy and model interpretability, paving the way for transparent AI systems. Several studies looked into how to boost the performance of YOLO algorithms in object detecting applications. YOLOv8, a novel object detection algorithm, has demonstrated its effectiveness in handling large-scale datasets. A review of research advances in YOLO algorithms provides insights into future developments and enhancements [1]. It has been explored for detecting moving objects, showcasing its real-time detection capabilities [2]. YOLOv8 has also been employed in the context of self-driving automobiles, where it improved object detection in difficult traffic settings [3]. Furthermore, an upgraded version of YOLOv8 has shown improvements in speed and accuracy, underlining its application in comprehensive object detection systems [4]. YOLOv8 has also been used for detecting abnormalities in power transmission lines, reinforcing its versatility in industrial applications [5].

ResNet50 has emerged as a prominent model in image recognition tasks, particularly when trained on large datasets like CIFAR-10. It has been utilized for detecting brain tumours using image processing techniques, achieving high accuracy [6]. Pre-trained neural networks such as ResNet50 have also been leveraged for area recognition in aerial images, illustrating their ability to generalize across different image domains [7]. ResNet50's pre-trained architecture has consistently offered a solid basis for transfer learning, thereby allowing for improved accuracy with less training data.

Explainable AI (XAI) is an emerging field of research, particularly in object detection systems where interpretability is critical. Shapley values have been employed for explaining object detection results, providing a fast and transparent method for model interpretability. Various studies have also contributed to the body of knowledge on integrating Explainable AI with object detection algorithms. Enhanced methods for XAI in object detection have been explored, improving the transparency of model predictions [8]. The BSED (Baseline Shapley-Based Explainable Detector) has been proposed to further improve explainability in object detection models [9]. Deep learning techniques have been applied to military tank recognition, focusing on transparency and accuracy in defence applications, highlighting the importance of XAI in high-stakes environments [10]. An object detection system using Arduino for military applications has also been developed, showcasing the relevance of object detection in security contexts. Comprehensive reviews of modern object detection techniques and transformations offer a broad overview of current trends in the field [11]. The YOLO approach has been explored for digital object definition in military scenarios, emphasizing its role in defence [12].

In defence applications, various object detection algorithms have been presented for video surveillance, underlining their utility in monitoring and tracking targets [13]. A survey on adversarial attacks and defence mechanisms for deep learning models highlights an area closely tied to explainability and model security. The use of Explainable AI in military applications reinforces the significance of transparency in high-stakes environments [14]. The use of image understanding in defence applications laid the groundwork for current advancements in military uses of AI [15].

Moreover, computer vision models have found applications in military image scene recognition, where convolutional neural networks (CNNs) combined with semantic information are used to recognize complex military scenes [16]. Hyperspectral imaging for military and security applications has also expanded the scope of object detection and recognition in defence. The applications of computer vision in defence, where image recognition enhances security measures, have been widely discussed [17].

In cybersecurity, Explainable AI has been applied to security applications, where transparency in decision-making processes is essential for trust [18]. Explainable intrusion detection systems have been introduced to protect IoT devices from cyber-attacks, emphasizing the need for transparency [19]. Further exploration into the role of XAI in creating resilient security solutions highlights its potential in enhancing trustworthiness in AI systems and the importance of XAI for cyber defence mechanisms has been emphasized, highlighting its role in improving security [20]. Challenges and future directions in object detection emphasize the need for further research into transparent AI systems

Finally, evaluations of object detection techniques for defence scenarios further solidify the importance of accuracy and explainability in military contexts [21].

This literature survey outlines the key contributions in the domains of object detection, image recognition, and Explainable AI, which form the foundation of this research. By leveraging the advancements in YOLOv8, ResNet50, and XAI techniques, this study aims to push the boundaries of accuracy and interpretability in AI-driven detection and recognition tasks.

**6.TOOLS USED**

1. **Google Colab:**

Google Colab provides cloud-based access to powerful GPU resources for faster model training and testing, resulting in efficient handling of large datasets.

2. **YOLOv8:**

YOLOv8 is a cutting-edge real-time object identification model that recognizes and localizes multiple objects in images. It is suited for surveillance and defence applications that require speed and accuracy.

3. **ResNet50:**

ResNet50 A deep convolutional neural network (CNN) model designed for accurate image classification. It is utilized here to identify items in the CIFAR-10 dataset, hence enhancing our system's overall classification accuracy.

4. **Explanatory AI (XAI) Techniques:**

XAI algorithms are integrated into YOLOv8 to make decision-making more visible. This allows users to understand the reasoning behind object detection results, which is crucial for trust and accountability in security applications.

5. **Custom Image Source Classification Model:**

A custom model is trained to classify the source of images, identifying whether they were captured by a drone, camera, or CCTV. This additional context helps in better understanding the data, which is particularly useful in surveillance scenarios.

**7.DATASET DESCRIPTION**

**Dataset 1:**

**CIFAR-10 Dataset**

• **Source:** CIFAR-10 on Kaggle For future work, numerous potential improvements could be explored.

• **Multi-modal Data Fusion:** The system might gain information from incorporating additional data sources, such as thermal imaging or radar, to increase detection accuracy in difficult settings like poor light or severe weather.

• **Scalability and Edge Deployment:** Optimizing the system for deployment on edge devices with limited resources may enable broader adoption in a variety of observation situations, ranging from small-scale installations to major public safety networks.

• **Real-time Alerts and Actionable Insights:** Implementing real-time alert systems and useful knowledge may increase the system's adaptation. This would enable security professionals to take fast action in response to suspected threats or strange activity.

**Table\_1 CIFAR10\_Label\_Indexer**

|  |  |  |
| --- | --- | --- |
| s.no | Id | Label |
| 0 | 1 | Frog |
| 1 | 2 | Truck |
| 2 | 3 | Truck |
| 3 | 4 | Deer |
| 4 | 5 | Automobile |
| . | . | . |
| . | . | . |
| . | . | . |
| n | n-1 | Deer |

<https://www.kaggle.com/c/cifar-10/>

**Dataset 2:**

**Custom Dataset for Object Detection and Source Classification Source:**

Custom Data Collection (Drone, Camera, CCTV) Overview: The custom dataset was created to train models for object detection and source classification. It comprises photographs obtained from various surveillance sources, such as drones, cameras, and CCTV systems. The photos are tagged with two different types of information:

**1. Object Annotations:** Images are annotated with categories such as people, cars, and animals to aid with object detection.

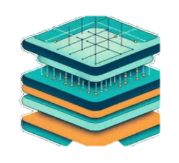
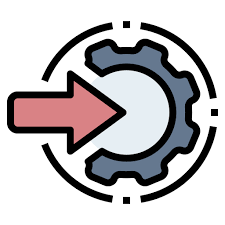
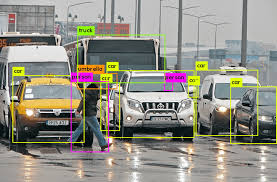
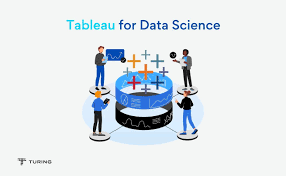
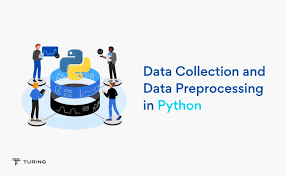
**2. Source Metadata**: Images are tagged with metadata identifying the source (e.g., CCTV, drone, camera) and additional context, such as indoor/outdoor and day/night capture. This metadata helps the model learn to differentiate between various sources based on visual and environmental cues.

**Usage:** This dataset will be used to train models for object detection and source identification. By recognizing both the objects within an image and the source of the image, the model will enhance situational awareness in real-time surveillance systems.

**Table\_2: Validation Subset Sample from Custom Dataset**

|  |  |  |
| --- | --- | --- |
| Class | Images | Instances |
| Drone | 20 | 69 |
| Cramera | 12 | 29 |
| Cctv | 15 | 32 |
| All | 47 | 130 |

**8. Architecture**



(Basic Network Training)

(Customer Dataset Input)

Model Training Layer

Model Training Layer

(CIFAR-10 Dataset Training)

(Explainability)

(Image Recognition)

(Explainable AI Techniques)

(Explainable AI Techniques)

(Yolov8 Training)

(Yolov8 Training)

(Yolov8 Training)

(ResNet50 Training)

(Resizing & Preprocessing)

(Data Filtering & Quality Assessment)

(Custom Dataset)

Data Collection & Preprocessing

Data Collection & Preprocessing

ResNet50 Model

Transfer Learning

Recognition Output

XAI Explanation

Source Detection

Object Detection

Yolov8 Model

Preprocessing Layer

Input Section

CIFAR-10 Dataset



Analysis & Comparison

**Fig\_1:** Dual-Path AI Framework for Enhanced Detection, Recognition, and Explainability

The above Fig\_1 illustrates a project pipeline using two models, YOLOv8 and ResNet50, from input to evaluation. It starts with a custom dataset for object detection and the CIFAR-10 dataset for image classification, both undergoing preprocessing. YOLOv8 is trained on the custom dataset for object and source detection, while ResNet50 applies transfer learning for image recognition on CIFAR-10. Explainable AI (XAI) techniques are integrated to enhance interpretability of both models. Finally, a comparative analysis evaluates model performance and the impact of XAI.

**9. Proposed works**

The proposed system integrates YOLOv8 for real-time object detection,and souce classifiaction to identify whether the image is captured by a drone, camera, or CCTV. ResNet50 for image recognition.lThe system also incorporates Explainable AI (XAI) techniques to provide transparency in the detection and classification process. The main objective is to enhance security and surveillance by providing both context (source of the image) and content (detected objects), while ensuring the decision-making process is interpretable.

**9.1 Data Preparation and Model Training**

**Custom Dataset for Object Recognition & Source Categorization:**

In the initial phase of the project, a custom dataset was developed to train the models for both object detection and source classification. This dataset consists of images captured from multiple surveillance sources, such as drones, cameras, and CCTV systems. Each image in the dataset is annotated with two key types of information:

1. **Object Annotations:** Images are labelled with object types (e.g., persons, vehicles, animals) to detect inside them.

2. **Source Classification:** Images are given labels with metadata to identify their source, such as CCTV, drone, or regular camera. Contextual metadata includes details such as the location (inside or outdoor) and the time of capture (day or night).

This allows the model to differentiate between the sources based on visual and environmental cues. For annotation, the dataset was created using Rob flow, a tool that helps label and manage datasets for machine learning. The dataset was then segregated into three distinct sets: training, validation, and testing. This ensures that the model can learn effectively from the data and be evaluated on unseen samples to test its performance.

**CIFAR-10 Database for Image Classification:**

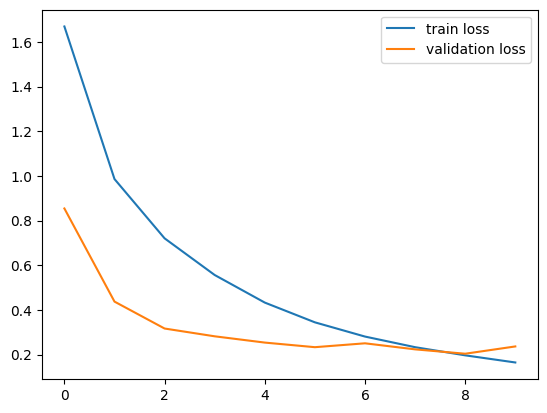
In conjunction with the custom dataset, the CIFAR-10 dataset was utilized to train ResNet50 for image recognition tasks. The CIFAR-10 dataset contains 60,000 photos from ten different categories (such as automobiles, animals, and airplanes). These photos were used to train ResNet50 to differentiate between different items in surveillance footage. The CIFAR-10 dataset's diverse number of item categories makes it appropriate for training ResNet50 for general object classification. Using ResNet50, a transfer learning algorithm, the model was able to efficiently apply pre-trained knowledge to the new dataset, enhancing its capacity to identify and classify objects in surveillance photos. This model also benefited from fine-tuning on the custom dataset, which enabled it to adapt to specific items important to the tracking context, such as people or automobiles in CCTV footage.

**9.2 Model Creation for Object Identification and Recognition**

**ResNet50 for Object Detection:**

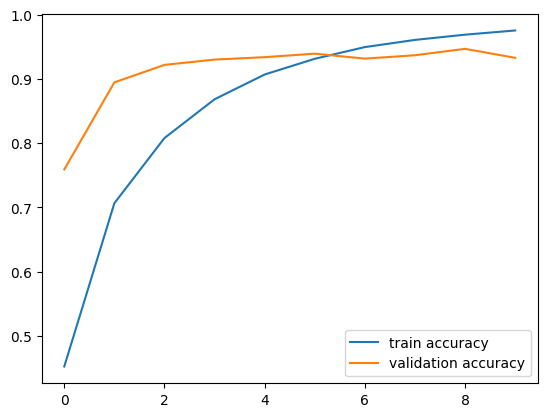
ResNet50 is a deep convolutional neural network designed to solve the problem of vanishing gradients in very deep networks by including residual connections, often known as "skip connection." These connections enable the model to learn residual mappings, which increases the network's ability to capture and learn complicated information while preserving efficient training dynamics. ResNet50 is well-known for its capacity to handle big datasets and recognize complex patterns, making it an effective model for picture classification and recognition applications. ResNet50 is used extensively in this research for picture recognition. It is trained using the CIFAR-10 dataset, which contains 60,000 32x32 pixel images from 10 different categories. The model's capacity to recognize and classify things in these photos is used to identify objects in a wider variety of settings. The ResNet50 model has been fine-tuned to improve the accuracy of object detection in photos reported by diverse sources, such as drones, traditional cameras, and CCTV systems

The below Fig\_2 is Training and Validation Loss Curve. This figure illustrates the significant improvement in model performance over the training epochs. The training loss (blue line) consistently decreases, demonstrating that the ResNet50 model is learning effectively from the data. Meanwhile, the validation loss (orange line) also drops considerably, indicating that the model generalizes well to unseen data. Although there's a slight increase in validation loss after 6 epochs, the overall trend suggests that the model maintains strong performance.



**Fig\_2:** Train & Validation Loss

The below Fig\_3 is aboutTraining and Validation Accuracy Curve. This figure highlights the growing accuracy of the ResNet50 model as training progresses. The training accuracy (blue line) rapidly increases, showing the model's ability to accurately classify images during training. The validation accuracy (orange line) also improves steadily, levelling off at an impressive 90%, which indicates that the model achieves high accuracy on new, unseen data. This performance underscores the effectiveness of the ResNet50 architecture in image recognition tasks.



**Fig\_3:** Train & Validation Accuracy

**YOLOv8 for Object Detection and Source Classification:**

YOLOv8 (You Only Look Once, Version 8) is a robust deep learning model intended for real-time object recognition. YOLOv8 is famous for its ability to identify quickly and accurately while preserving a good balance of quickness and accuracy, making it ideal for current time analysis of images applications like security camera video and self-driving automobiles. It scans images in one pass, helping it to detect and find items quickly and correctly. YOLOv8 use a single neural network to predict box boundaries and class labels for numerous items in a photograph, allowing it to recognize and classify objects in real time. The system's structure and quick operation have rendered it an appealing option for applications that demand optimal performance.

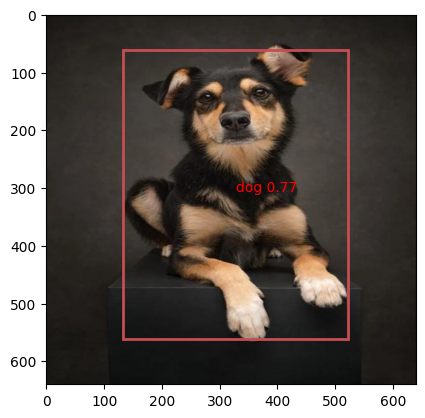
• Object Detection: The model can recognize and localize items in photos. For example, it recognizes dogs and creates a boundary around them, along with confidence scores.

• Source Classification: In addition to detecting objects, the model is trained to identify the image's source, such as a camera, drone, or CCTV. This is performed by examining both image attributes (e.g., resolution, colour profile) and metadata.

Example Output:

• Image Source: Camera (Confidence: 91.68%)

• Detected Object: Dog (Confidence: 76.61%)

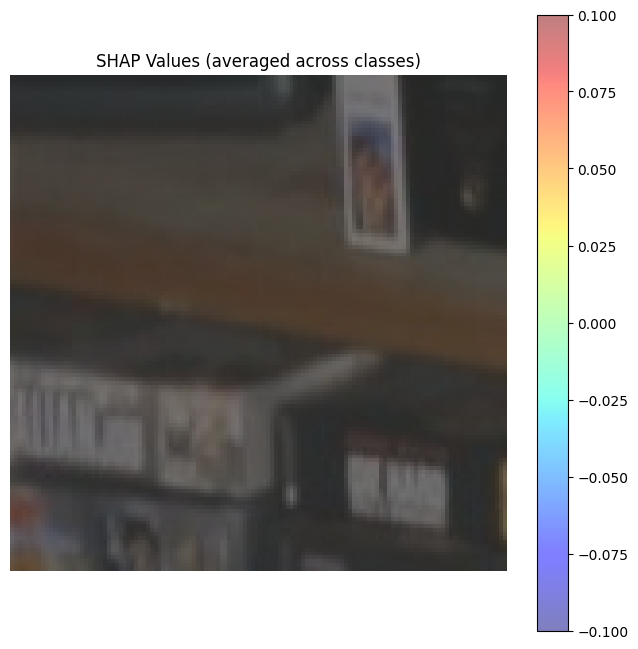


To validate the performance of the YOLOv8 model, key evaluation metrics were computed, demonstrating the model's good performance on the test data. The findings show that YOLOv8 not only excels at detecting objects within photographs, but it also reliably determines the images' source with high confidence. This dual capacity improves the system's overall effectiveness, making it perfect for applications that work in real time such as security and monitoring.

**9.3 Explainable Artificial Intelligence (XAI) in YOLOv8**

The phrase "explainable AI" (XAI) refers to methods and approaches that increase the openness and interpretability of AI models, especially automated learning models. The primary goal of XAI is to provide succinct, intelligible justifications for the assessments and decisions rendered by the artificially intelligent system. This is especially relevant in applications such as medical services, security, and self-driving vehicles, where knowing the logic behind an algorithm's decision is critical to maintaining trust and openness. In this Study XAI is incorporated into the YOLOv8 model to improve transparency in both object identification and source classification tasks. Heatmaps for object recognition, for each object identified by YOLOv8, heatmaps are generated to graphically display which areas of the image were most significant in reaching the detection decision.

1. heatmaps are closely related to the bounding boxes around the detected items. This enables viewers to see precisely which portions of the image influenced the model's classification.



• Example: When detecting a dog, the heatmap highlights the dog's body parts, such as the legs or face, indicating the areas that were most important for the model's decision.

2. Textual Explanations for Object Detection: In addition to visual explanations, textual descriptions are provided alongside the detection results. These explanations clarify the model’s reasoning behind classifying a specific object. For example, if a dog is detected, the explanation might state: "Detected object: Dog (Confidence: 76.61%) identified based on its four legs, body shape, and fur texture." 3. Source Classification with Explanations: YOLOv8 is also trained to classify the source of an image (whether it was captured by a camera, drone, or CCTV). The model is equipped with XAI techniques to explain the factors that influenced the classification. For instance, if an image is classified as originating from a drone, the explanation could be: "Image source: Camera (Confidence: 77.68%). Identified due to aerial view and high-resolution features typical of drone-captured images."

**9.4 Performance Evaluation and System Integration**

After training and fine-tuning the models (YOLOv8, ResNet50, and the Source Classification Model), we evaluated their performance using key metrics to assess their suitability for real-time object detection, source classification, and overall system integration. These metrics were selected to ensure the system delivers optimal performance in a surveillance environment, where both speed and accuracy are critical for effective operation.

**Table\_2: Key Evaluation Metrics**

|  |  |  |
| --- | --- | --- |
| **Metric** | **YOLOv8** | **ResNet50** |
| **Inference Speed (FPS)** | 45 FPS | 35 FPS |
| **Model Size (MB)** | 120 MB | 100 MB |
| **Training Time (hrs)** | 10 hrs | 12 hrs |
| **Robustness** | 9/10 | 7/10 |
| **Complexity** | 8/10 | 6/10 |

• **Inference Speed (FPS):** Measures how quickly the models make predictions, ensuring real-time detection. YOLOv8 performs faster with 45 FPS, making it ideal for dynamic, fast-paced environments like surveillance where immediate responses are crucial.

• **Model Size (MB):** Indicates storage needs, allowing for deployment on devices with restricted capacity. ResNet50 has a reduced model size (100 MB), making it ideal for smartphones with limited storage.

• **Training Time (hrs):** Measures the efficiency of the training process, estimating the time required for model building and fine-tuning. YOLOv8 required 10 hours of training, whereas ResNet50 required 12 hours, making YOLOv8 more efficient for model building and updates.

• **Robustness:** Assesses model performance in real-world surveillance circumstances, including lighting and angle changes. YOLOv8 received a 9/10 rating compared to ResNet50's 7/10, indicating that it is more trustworthy in unpredictable surveillance settings such as shifting illumination or camera angles.

• **Complexity:** Measures the computational resources needed for execution. YOLOv8 is more complex (8/10) compared to ResNet50 (6/10), which means it requires more computational resources. However, its superior performance justifies this complexity in high-performance environments. ResNet50, being lighter, is better suited for devices with limited processing power.

**10.Innovations**

This project introduces various improvements that raise the classic security and surveillance systems to a new degree of performance and intelligence.

**1.** **Real-time Object Detection and Source Classification:** This system expands on existing surveillance models by merging YOLOv8's real-time object identification capabilities with source categorization (drone, camera, or CCTV). It delivers total awareness of the circumstances by detecting not only the items in the image but also the context in which it was photograph.

**2. Use of Transfer Learning with ResNet50:** Leveraging ResNet50 for image recognition significantly enhances the system's ability to identify and classify objects within surveillance footage. Fine-tuning this model using the CIFAR-10 dataset allows for a robust performance across a variety of object categories, making it adaptable for diverse surveillance environments.

**3. Explainable AI (XAI) Integration:** The incorporation of XAI techniques such as heatmaps and textual explanations adds a layer of interpretability, crucial for high-stakes applications like security. This innovation allows users to understand the decision-making process of the AI model, ensuring that every detection and classification is transparent and trustworthy.

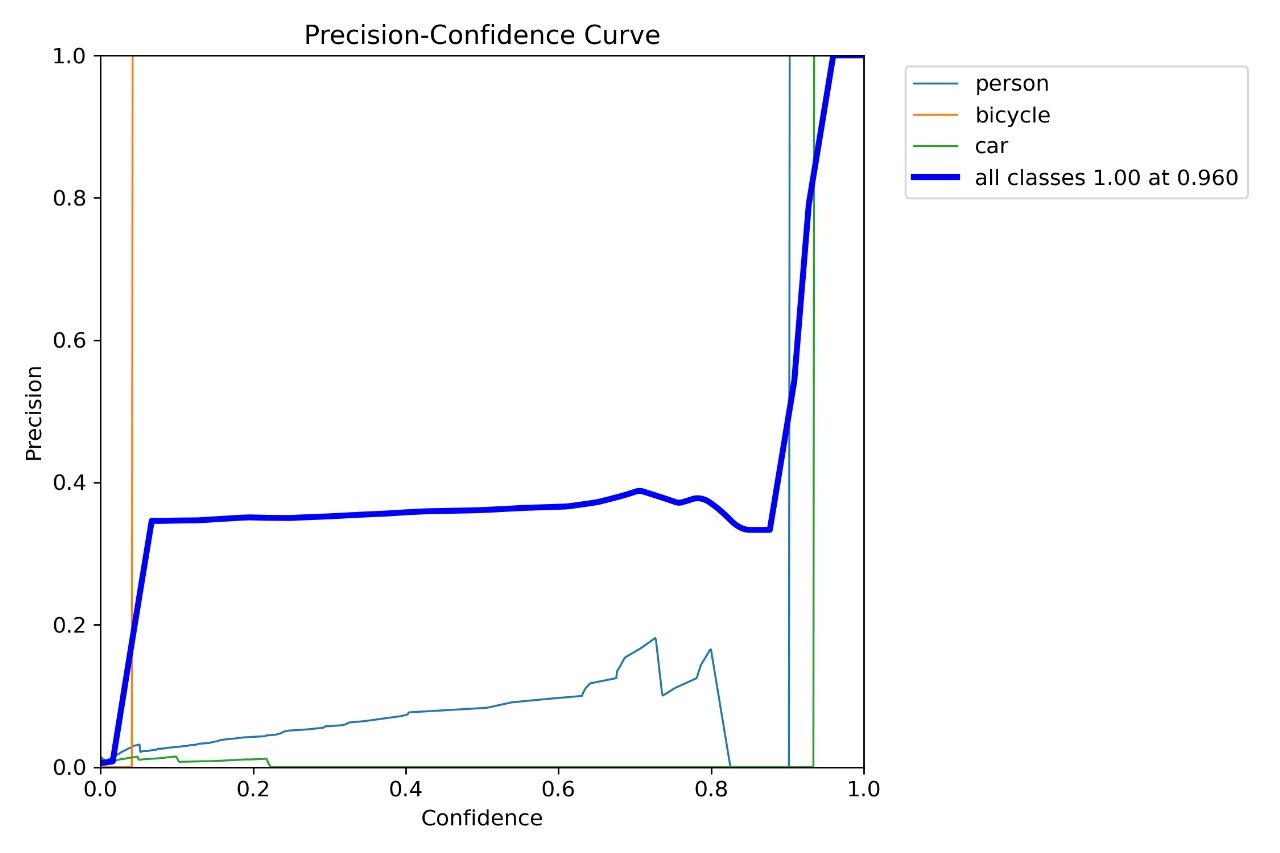
**4. Custom Dataset for Source and Object Categorization**: The development of a custom dataset combining object annotations and metadata for source classification sets this project apart. This dataset allows the system to recognize objects in photographs while also discriminating distinct image sources (e.g., drones, cameras, CCTV), which improves the system's overall usefulness and accuracy.

These advancements not only improve the system's technical performance, but also increase its practical utility in real-world monitoring circumstances. The technology stands out as a future-oriented security and surveillance solution because to its speed, accuracy, and clarity.

**11.RESULTS AND DISCUSSIONS**

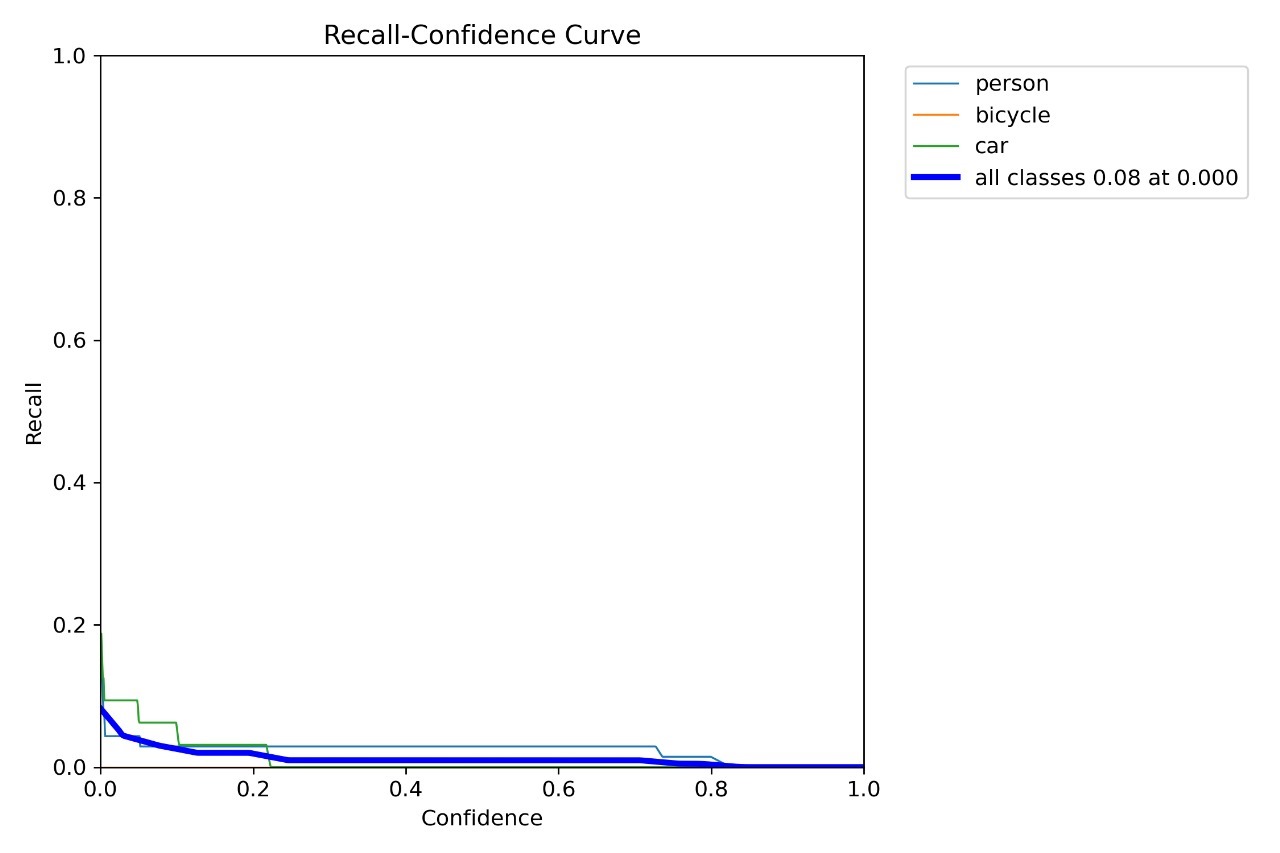
Our research highlights the exceptional performance of integrating YOLOv8 and ResNet50 for real-time object detection, image recognition, and source classification, making the system a valuable asset for defence and surveillance applications. The evaluation results, summarized in the table below, demonstrate the effectiveness of these models in both noise-free and noisy environments:

|  |  |  |
| --- | --- | --- |
| Model | ResNet50 | Yolov8 |
| Without Noise | 70% | 100% |
| With Noise | 47 % | 35% |



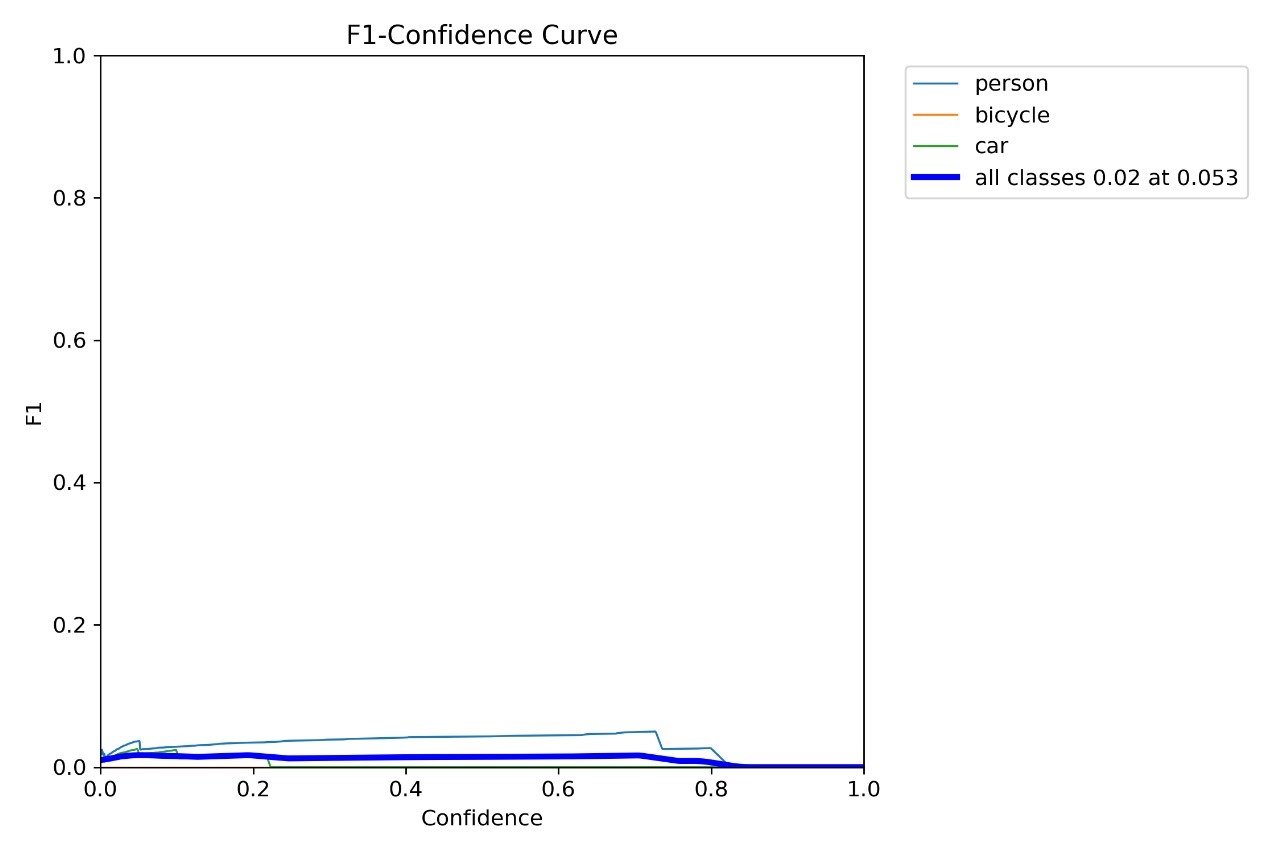
**Fig\_4** **Precision-Confidence Curve:**

This above fig\_4 curve shows how precision (accuracy of detections) changes with the confidence threshold. In clean images, precision stays high across all thresholds, meaning the model detects objects accurately. However, in noisy images, precision drops, especially at lower thresholds, showing that noise causes false positives. This highlights the need for better noise handling techniques.



**Fig\_5 Recall-Confidence Curve:**

The above fig\_5 graph shows how recall (ability to detect objects) varies with confidence thresholds. In clean images, recall remains high, meaning most objects are detected. In noisy images, recall drops, especially at higher thresholds, indicating that noise causes the model to miss objects. This suggests the model needs more diverse training to handle noise better.



**Fig\_6 F1-Confidence Curve:**

The above Fig\_6 show F1 curve combines precision and recall into one score. In clean images, the score is high, indicating balanced performance. In noisy images, the F1 score drops significantly, showing that both precision and recall suffer. This highlights the need for improvements to maintain detection accuracy in noisy environments.

In optimal, noise-free conditions, YOLOv8 achieved an outstanding accuracy of 100%, demonstrating its capability for highly reliable object detection. Its inference speed of 45 frames per second (FPS), compact model size of 120 MB, and training time of 10 hours make it highly efficient for real-time operations. ResNet50 achieved a commendable accuracy of 70% in image recognition tasks on the CIFAR-10 dataset. Despite its slower inference speed of 35 FPS and a slightly longer training time of 12 hours, its compact size of 100 MB ensures compatibility with devices that have storage limitations.

Under noisy conditions, the models faced challenges reflective of real-world scenarios. YOLOv8's accuracy dropped to 35%, highlighting the impact of noise on object detection. However, its robustness score of 9/10 signifies its ability to adapt to varying environmental conditions, such as lighting or angles. Similarly, ResNet50 maintained a reasonable accuracy of 47%, with a robustness score of 7/10, showing its relative resilience in handling noise-affected data. These findings underscore the importance of incorporating noise-handling or pre-processing techniques to improve reliability under adverse conditions.

The combined system balances real-time performance and accuracy while adding the critical functionality of image source categorization. This feature, which distinguishes between images captured by drones, CCTV, or conventional cameras, enhances situational awareness—a crucial aspect in defence and surveillance operations. Furthermore, the integration of Explainable AI (XAI) methods, including heatmaps and decision justifications, adds a layer of transparency to the system's predictions. Such interpretability is essential for building trust and accountability in high-stakes applications.

**12.CONCLUSION & FUTURE WORK**

The system has shown to be a potent way to improve security and surveillance operations. It was developed utilizing Explainable AI (XAI) methodologies, ResNet50 for image recognition, and YOLOv8 for object detection. Unmatched situational awareness is provided by combining real-time object recognition, accurate classification, and source identification; this is crucial in surveillance and defense scenarios. YOLOv8 stands out for its high inference speed, robustness, and effectiveness in dynamic, real-time contexts, making it ideal for fast-paced observation needs. On the other hand, ResNet50 delivers high accuracy in object classification, particularly when resource efficiency and compact model size are essential. Together, these models offer a balanced approach, combining speed, accuracy, and versatility.

The analysis of results highlights the system's strong performance under noise-free conditions, with YOLOv8 achieving 100% accuracy and ResNet50 reaching 70% accuracy. Under noisy conditions, the models faced challenges, but ResNet50 displayed notable resilience, maintaining an accuracy of 47%. These findings emphasize the system’s adaptability and underline areas for future improvement, such as incorporating noise-reduction techniques to enhance robustness in challenging environments.

Looking ahead, several potential enhancements could further elevate the system's capabilities. Multi-modal data fusion, integrating additional inputs like thermal imaging or radar, could improve detection in adverse conditions such as low light or severe weather. Optimizing the system for edge deployment would enable broader scalability, supporting a range of surveillance scenarios from small-scale operations to large public safety networks. The addition of real-time alert mechanisms and actionable insights could empower security personnel to respond swiftly to potential threats. More XAI developments may result in more thorough written and visual explanations, increasing usability and confidence in challenging situations.

By taking care of these issues, the system may develop into a more dependable and intelligent security and surveillance solution, opening the door for more effective and flexible next-generation surveillance systems. These developments will guarantee that it continues to be a strong instrument for security and defence applications, successfully bridging the gap between precision, openness, and practicality.

**13.REFERENCES**

[1] R. Varghese and S. M., "YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness," *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, Chennai, India, 2024, pp. 1-6, doi: 10.1109/ADICS58448.2024.10533619.

[2] M. Safaldin, N. Zaghden and M. Mejdoub, "An Improved YOLOv8 to Detect Moving Objects," in *IEEE Access*, vol. 12, pp. 59782-59806, 2024, doi: 10.1109/ACCESS.2024.3393835.

[3] Z. Afrin, F. Tabassum, H. B. Kibria, M. R. Imam and M. R. Hasan, "YOLOv8 Based Object Detection for Self-driving Cars," *2023 26th International Conference on Computer and Information Technology (ICCIT)*, Cox's Bazar, Bangladesh, 2023, pp. 1-6, doi: 10.1109/ICCIT60459.2023.10441381

[4] Z. J. Khow, Y. -F. Tan, H. A. Karim and H. A. A. Rashid, "Improved YOLOv8 Model for a Comprehensive Approach to Object Detection and Distance Estimation," in *IEEE Access*, vol. 12, pp. 63754-63767, 2024, doi: 10.1109/ACCESS.2024.3396224.

[5] S. Yang and Y. Zhou, "Abnormal Object Detection with an Improved YOLOv8 in the Transmission Lines," *2023 China Automation Congress (CAC)*, Chongqing, China, 2023, pp. 9269-9273, doi: 10.1109/CAC59555.2023.10451095.

[6] S. Rathore, T. Rana, U. Mittal, T. Gupta, N. Malik and M. S. A S, "Brain Tumor Detection by using ResNet-50 and Image Processing Tools," *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, India, 2023, pp. 480-483, doi: 10.1109/ICACITE57410.2023.10183081.

[7] J. Kaur, S. Rani, A. Sharma and A. Dogra, "Area Recognition in Aerial Images Leveraging Pre-trained ResNet-50 Architecture," *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Kirtipur, Nepal, 2024, pp. 1417-1422, doi: 10.1109/I-SMAC61858.2024.10714819.

[8] M. Kuroki and T. Yamasaki, "Fast Explanation Using Shapley Value for Object Detection," in *IEEE Access*, vol. 12, pp. 31047-31054, 2024, doi: 10.1109/ACCESS.2024.3369890.

[9] M. Kuroki and T. Yamasaki, "BSED: Baseline Shapley-Based Explainable Detector," in *IEEE Access*, vol. 12, pp. 57959-57973, 2024, doi: 10.1109/ACCESS.2024.3391424.

[10] S. Ali, Abdullah, A. Athar, M. Ali, A. Hussain and H. -C. Kim, "Computer Vision-Based Military Tank Recognition Using Object Detection Technique: An application of the YOLO Framework," *2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC)*, Jeddah, Saudi Arabia, 2023, pp. 1-6, doi: 10.1109/ICAISC56366.2023.10085552.

[11] J. C. N, A. K, S. A. C, T. N. M, V. R. S and S. N. K, "Object Detection System using Arduino for Military Application," *2023 2nd International Conference for Innovation in Technology (INOCON)*, Bangalore, India, 2023, pp. 1-4, doi: 10.1109/INOCON57975.2023.10101352.[12] R. Benzer and M. C. Yildiz, "YOLO Approach in Digital Object Definition in Military Systems," *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT)*, Ankara, Turkey, 2018, pp. 35-37, doi: 10.1109/IBIGDELFT.2018.8625314.

[13] A. Raghunandan, Mohana, P. Raghav and H. V. R. Aradhya, "Object Detection Algorithms for Video Surveillance Applications," *2018 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, 2018, pp. 0563-0568, doi: 10.1109/ICCSP.2018.8524461.[14] S. Y. Khamaiseh, D. Bagagem, A. Al-Alaj, M. Mancino and H. W. Alomari, "Adversarial Deep Learning: A Survey on Adversarial Attacks and Defense Mechanisms on Image Classification," in *IEEE Access*, vol. 10, pp. 102266-102291, 2022, doi: 10.1109/ACCESS.2022.3208131.[15] O. Firschein, "Defense applications of image understanding," in *IEEE Expert*, vol. 10, no. 5, pp. 11-17, Oct. 1995, doi: 10.1109/64.464926.

[16] C. Chen, J. Huang, C. Pan and X. Yuan, "Military Image Scene Recognition Based on CNN and Semantic Information," *2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, Huhhot, China, 2018, pp. 573-577, doi: 10.1109/ICMCCE.2018.00126.[17] M. Shimoni, R. Haelterman and C. Perneel, "Hypersectral Imaging for Military and Security Applications: Combining Myriad Processing and Sensing Techniques," in *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 101-117, June 2019, doi: 10.1109/MGRS.2019.2902525.

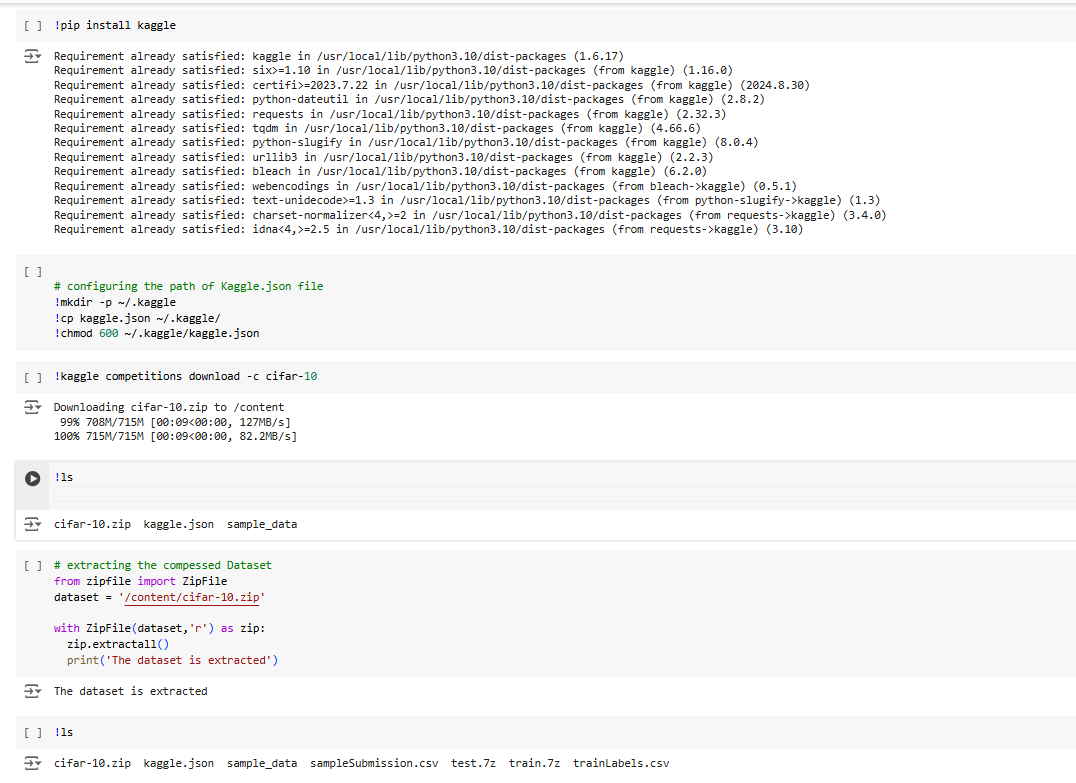
[18] Z. Zhang, H. A. Hamadi, E. Damiani, C. Y. Yeun and F. Taher, "Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research," in *IEEE Access*, vol. 10, pp. 93104-93139, 2022, doi: 10.1109/ACCESS.2022.3204051. [19] N. Moustafa, N. Koroniotis, M. Keshk, A. Y. Zomaya and Z. Tari, "Explainable Intrusion Detection for Cyber Defences in the Internet of Things: Opportunities and Solutions," in *IEEE Communications Surveys & Tutorials*, vol. 25, no. 3, pp. 1775-1807, thirdquarter 2023, doi: 10.1109/COMST.2023.3280465.[20] M. T. Masud, M. Keshk, N. Moustafa, I. Linkov and D. K. Emge, "Explainable Artificial Intelligence for Resilient Security Applications in the Internet of Things," in *IEEE Open Journal of the Communications Society*, doi: 10.1109/OJCOMS.2024.3413790.

[21] R. Hamilton, "Coretet: A 21st Century Virtual Reality Musical Instrument for Solo and Networked Ensemble Performance," *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Osaka, Japan, 2019, pp. 1305-1306, doi: 10.1109/VR.2019.8797825.

Dataset, CIFAR-10: <https://www.kaggle.com/c/cifar-10/>

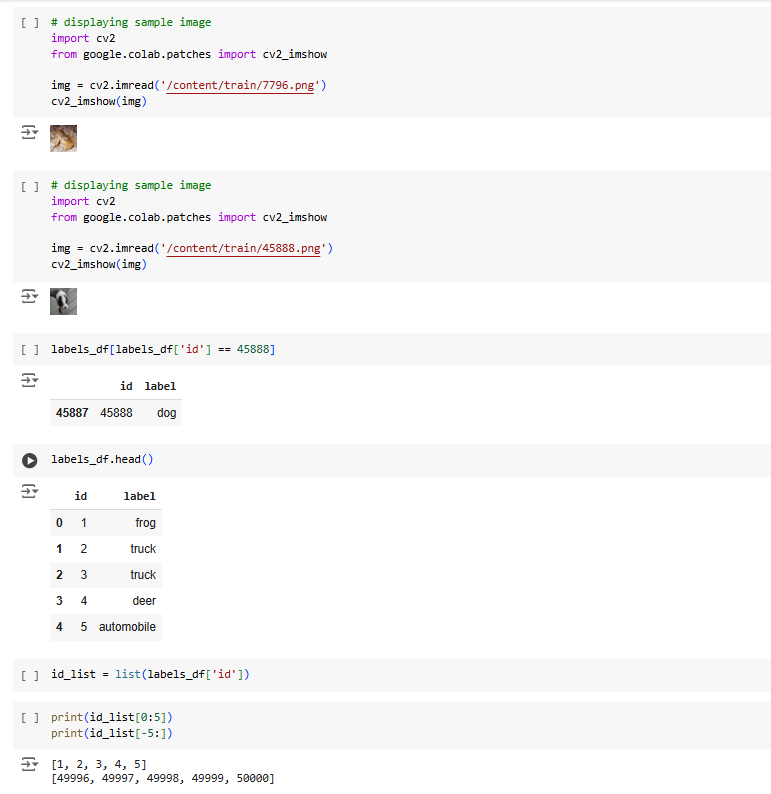
GitHub link: <https://github.com/kothapalliAnusha/Integrated-Approaches-for-Object-Detection-and-Image-Recognition-with-XAI.git>

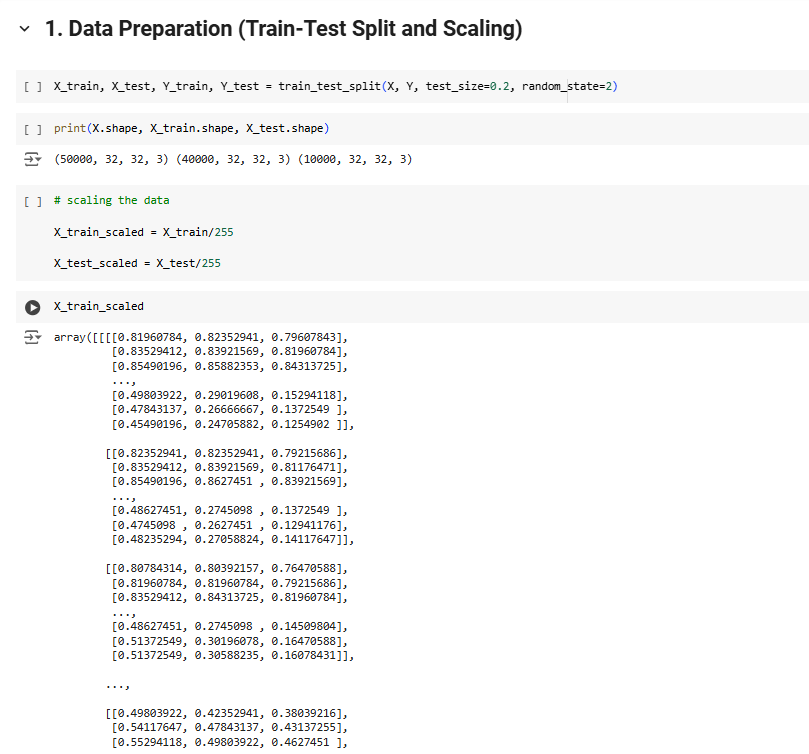
**14 Appendix code**

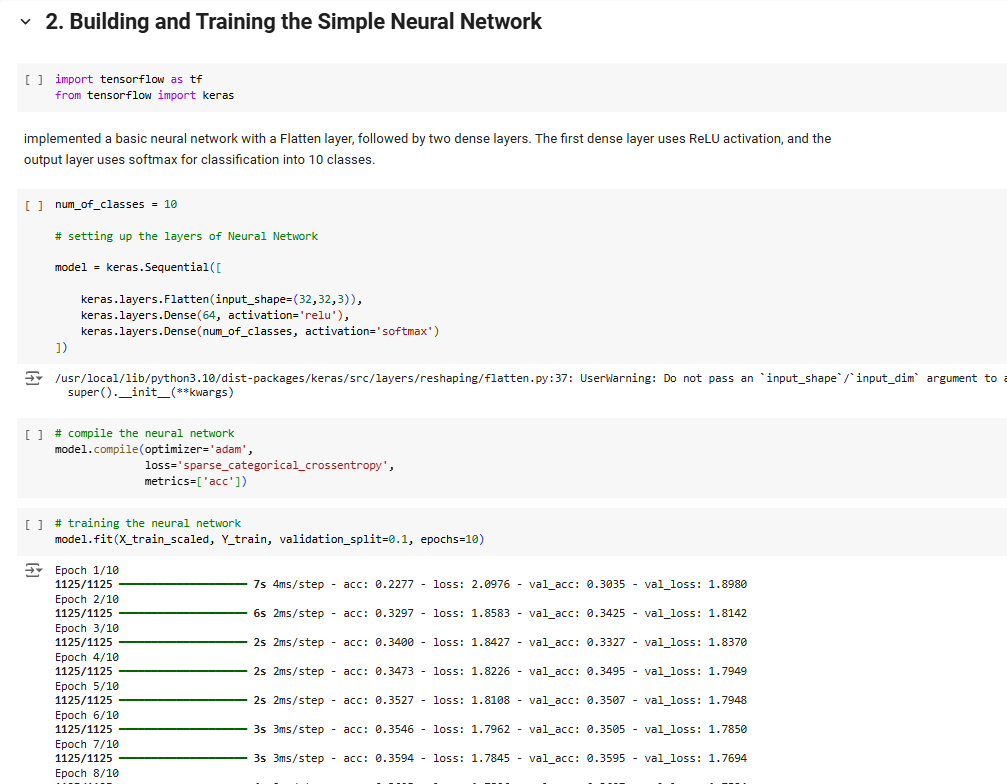


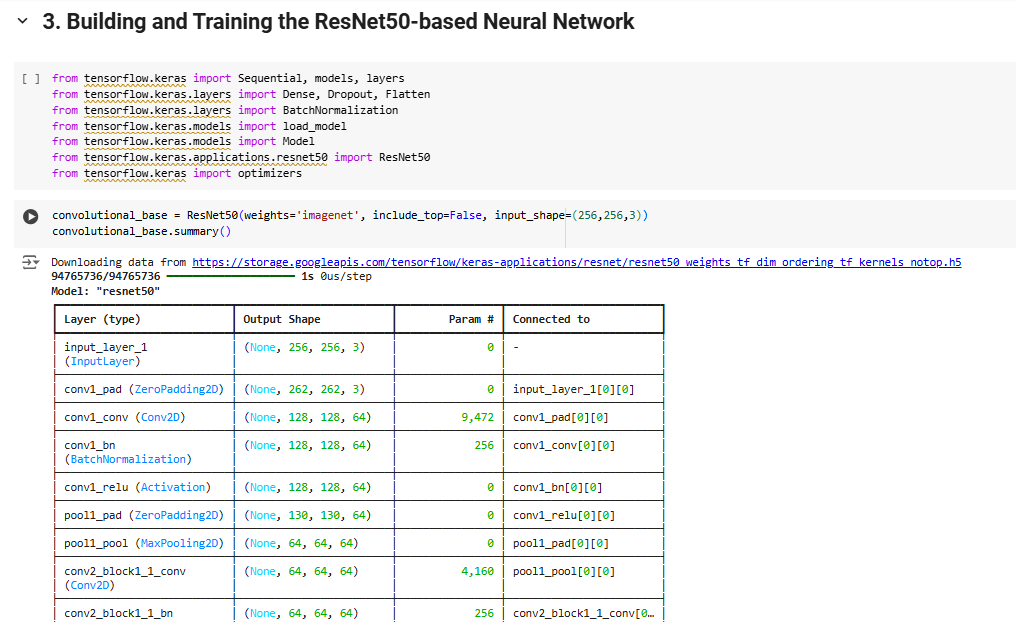


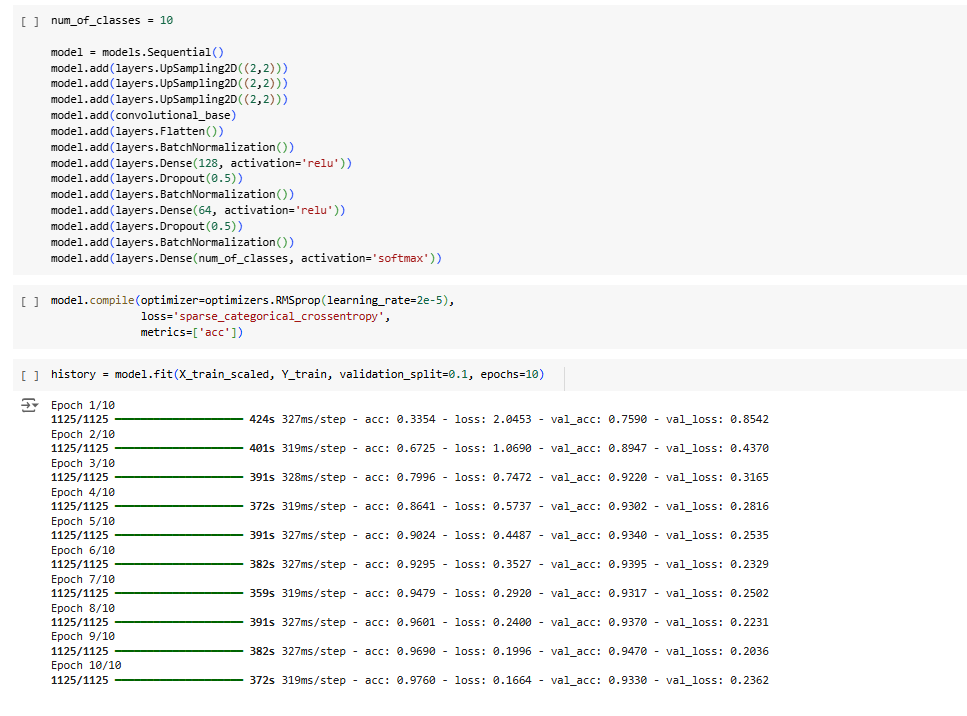


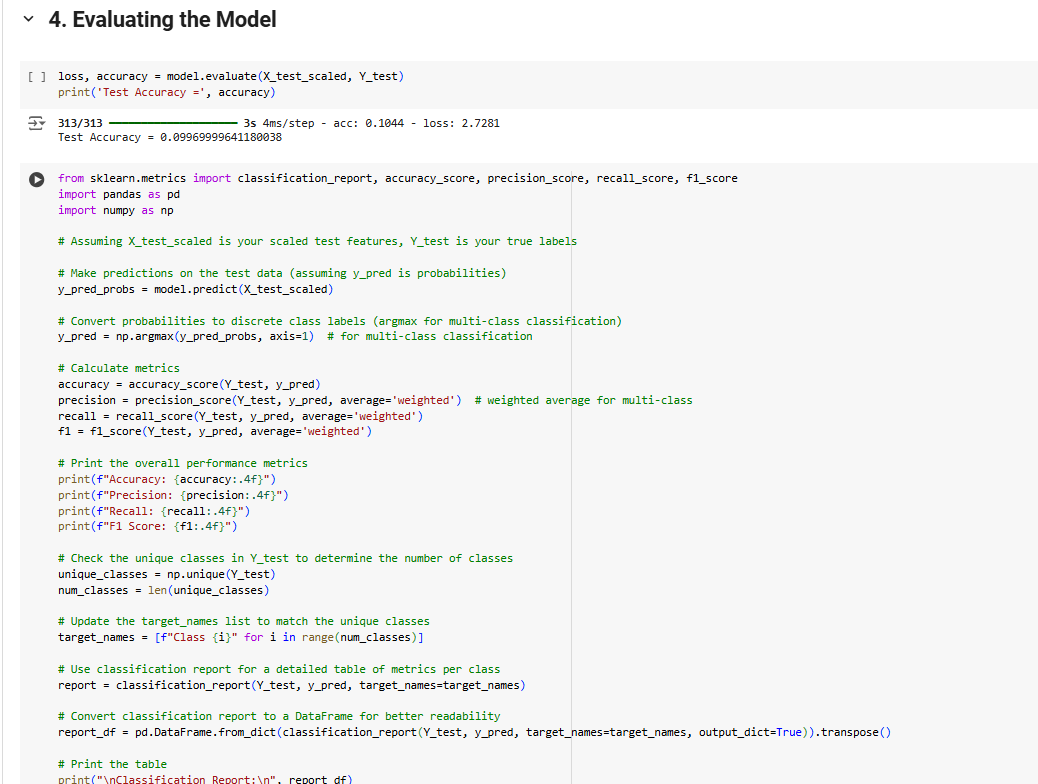


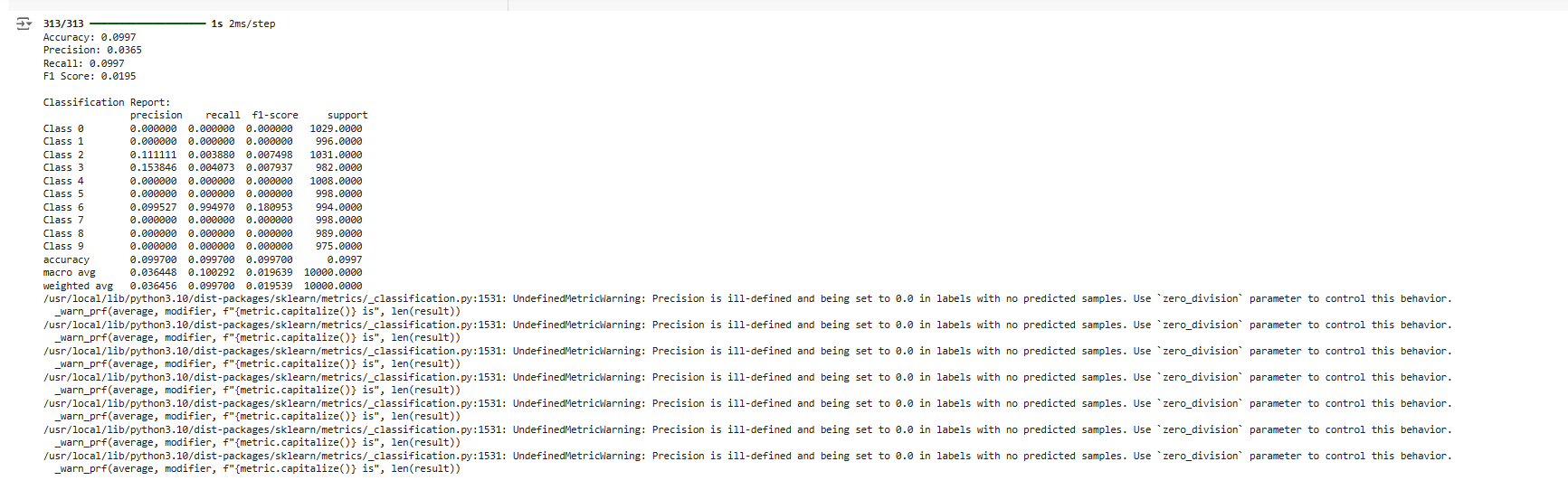


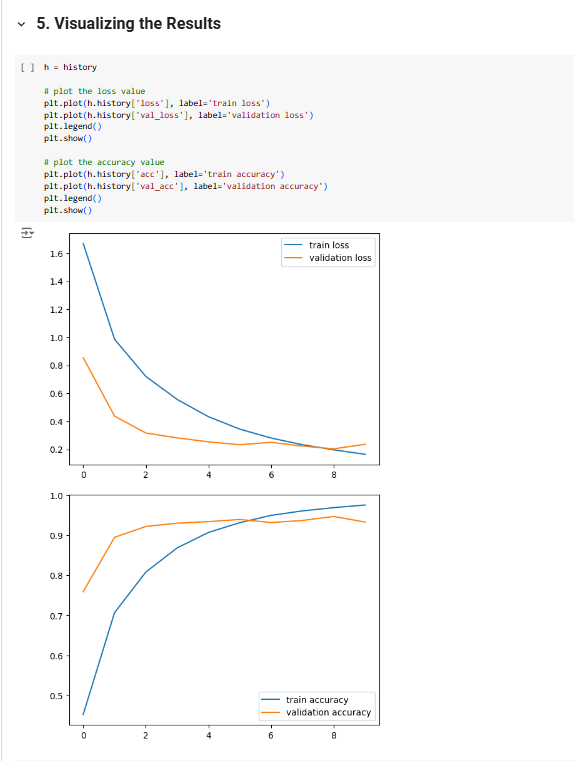
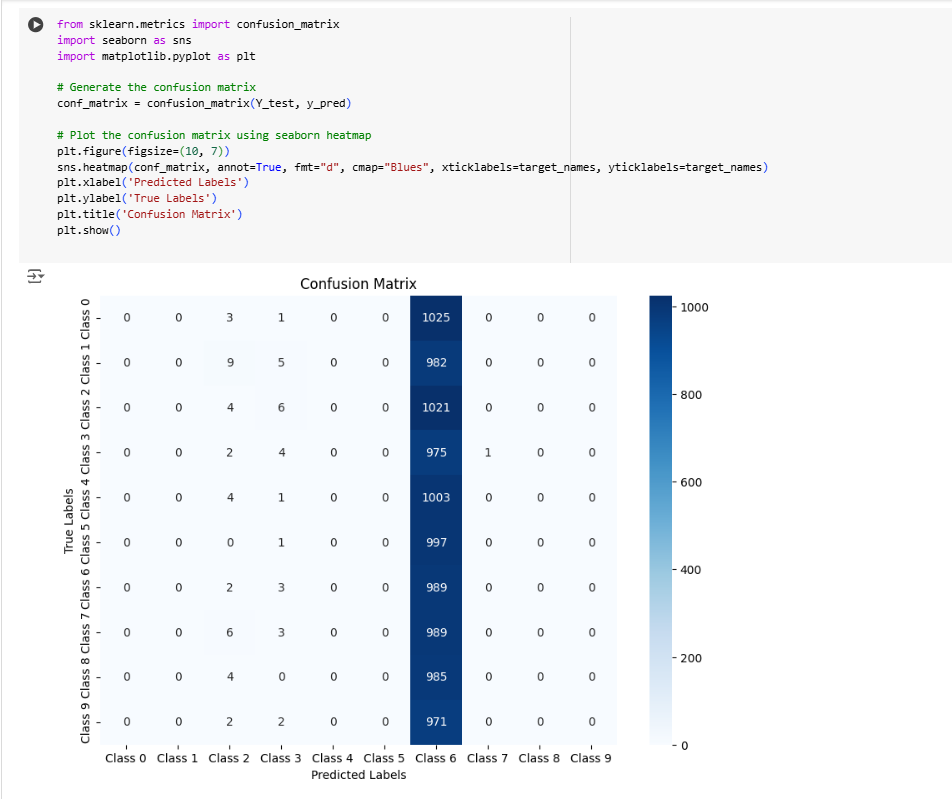


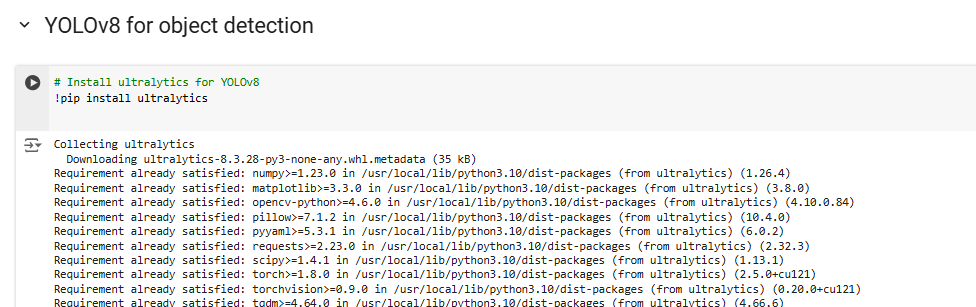




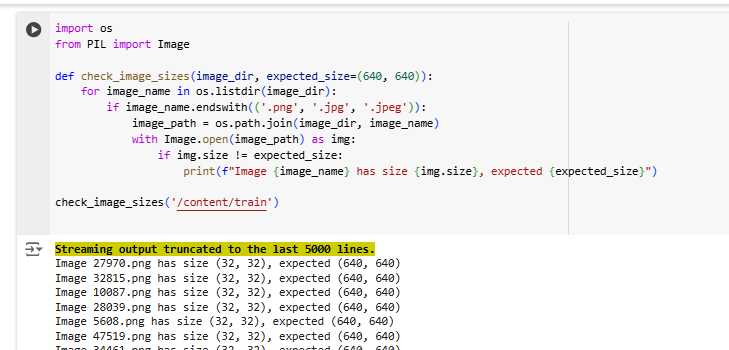


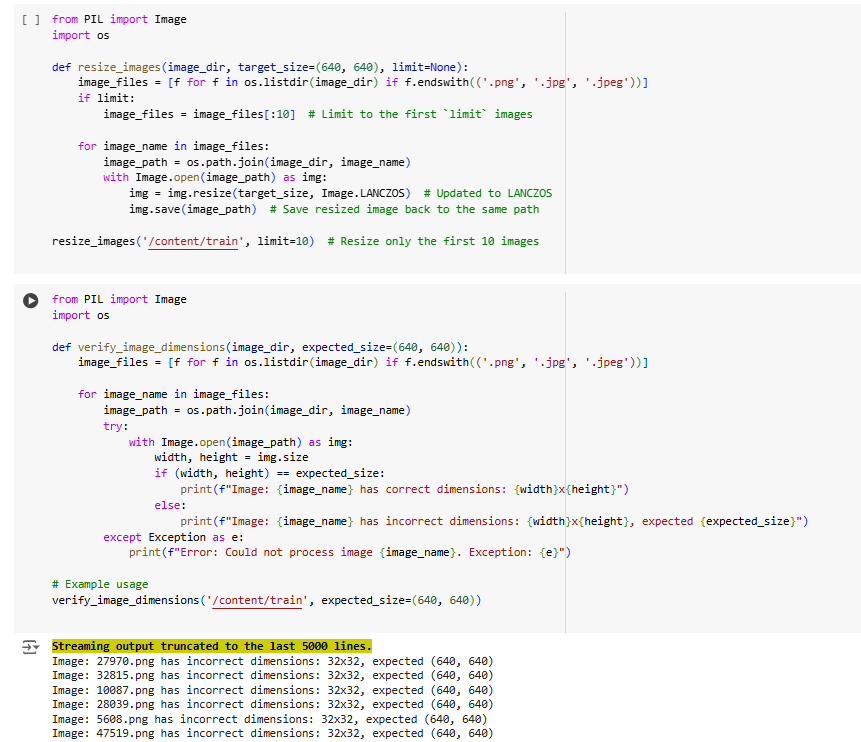






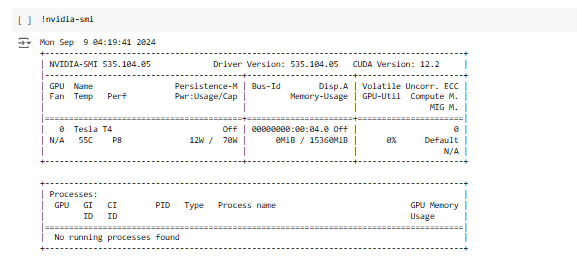


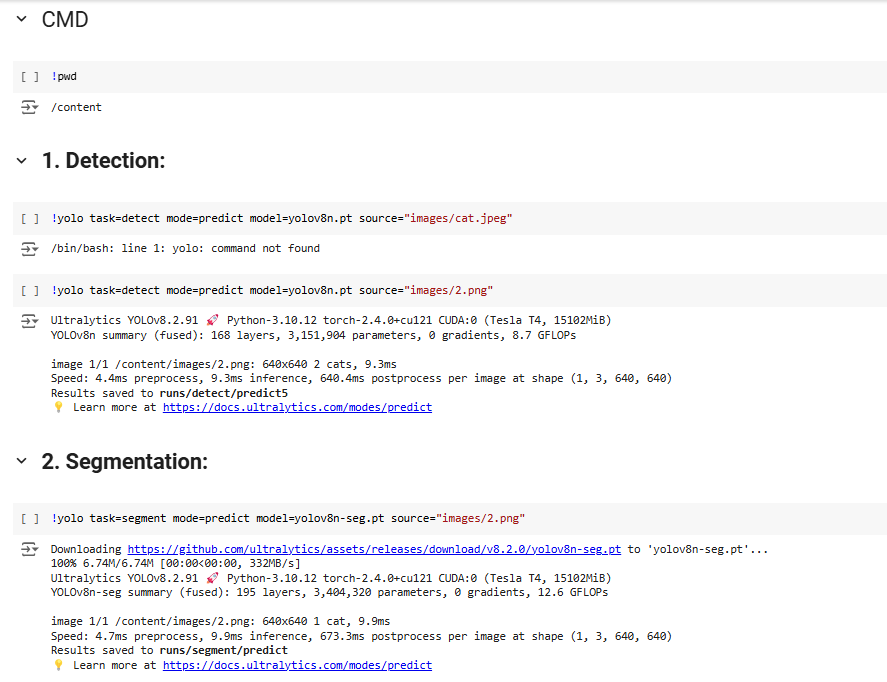


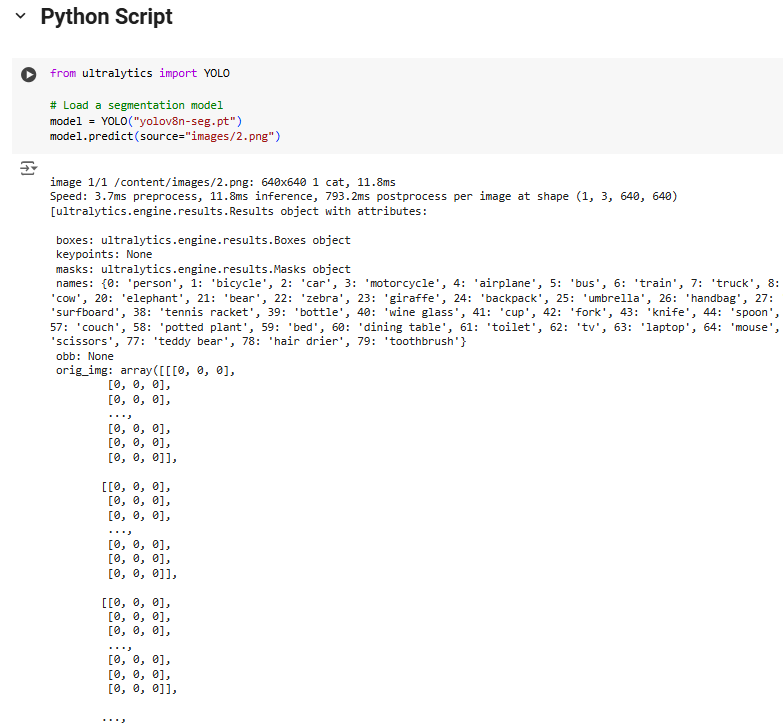






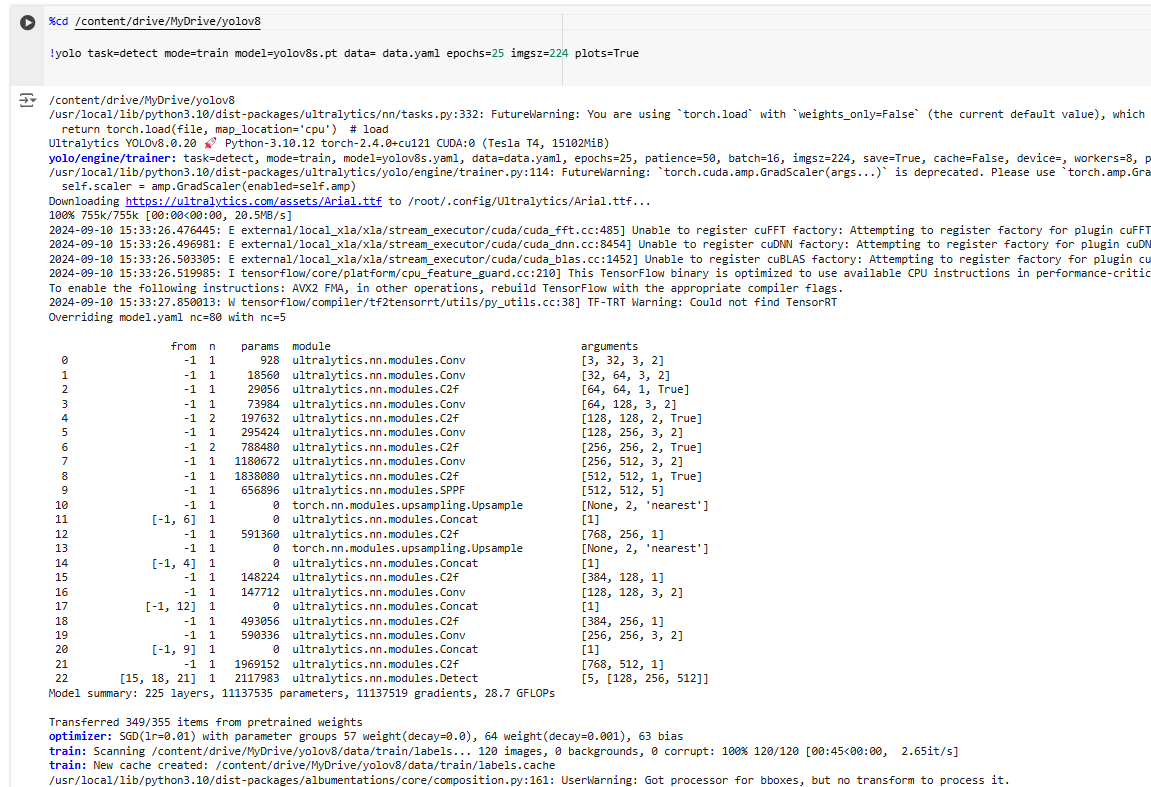




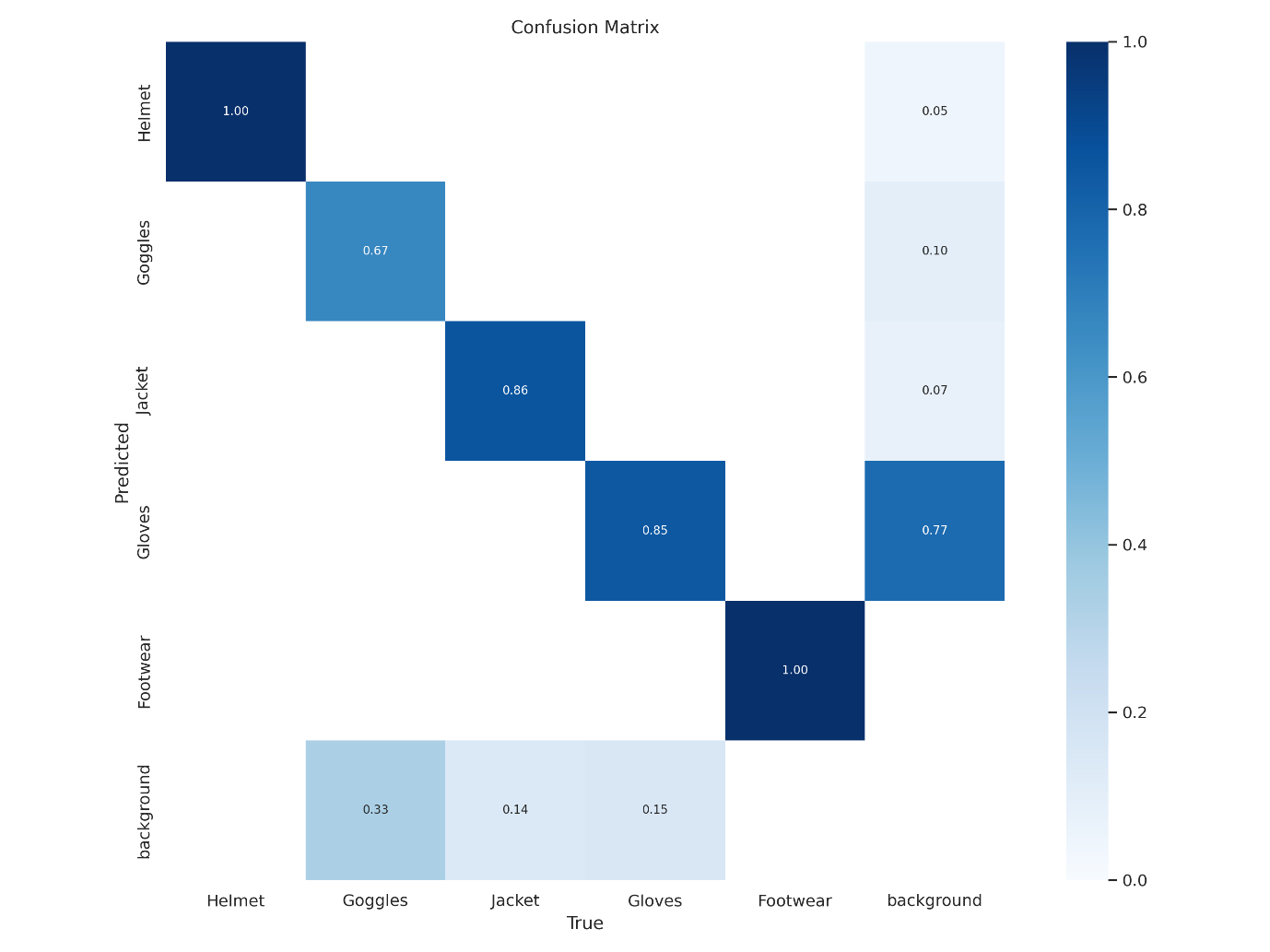




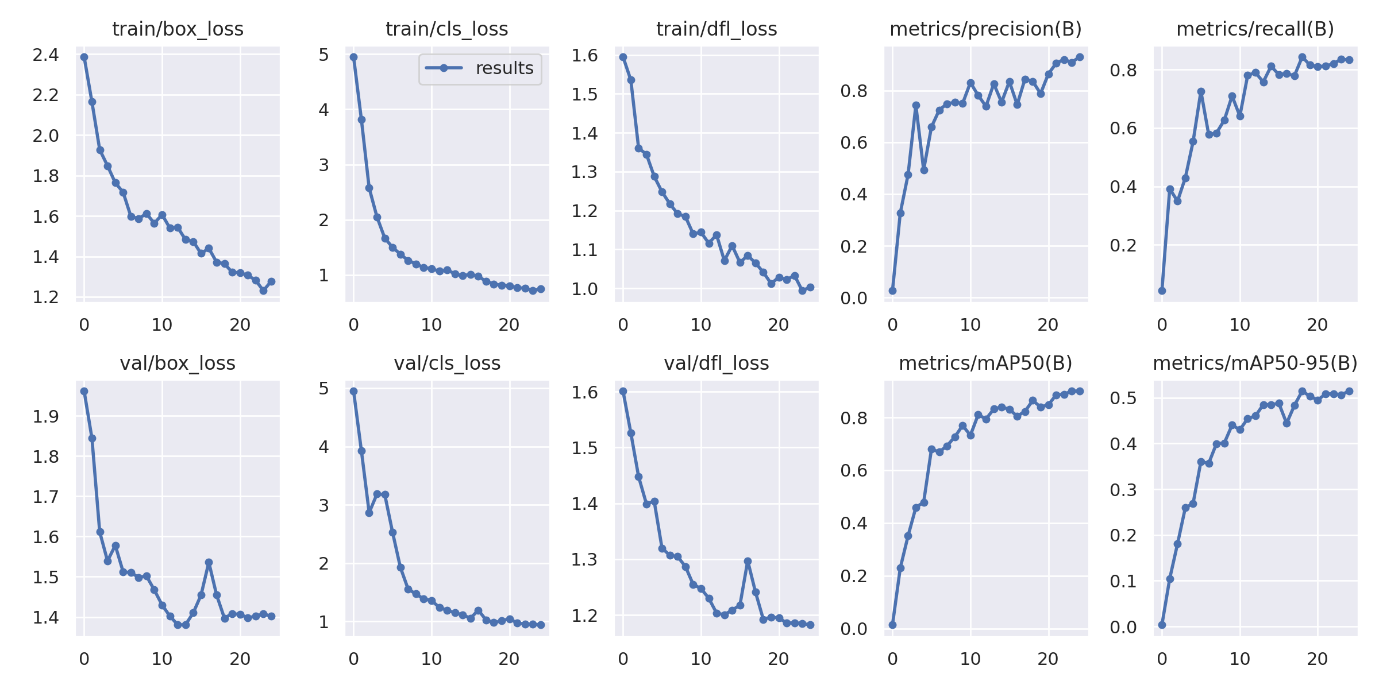








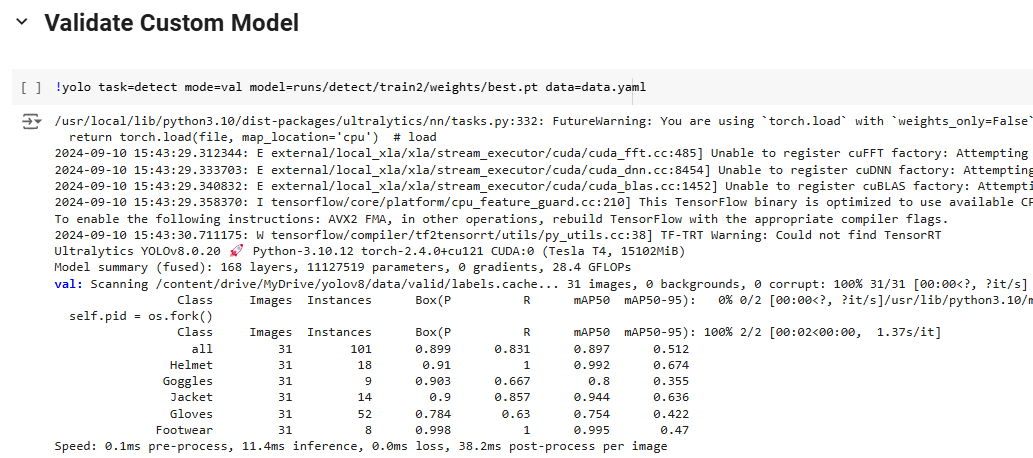


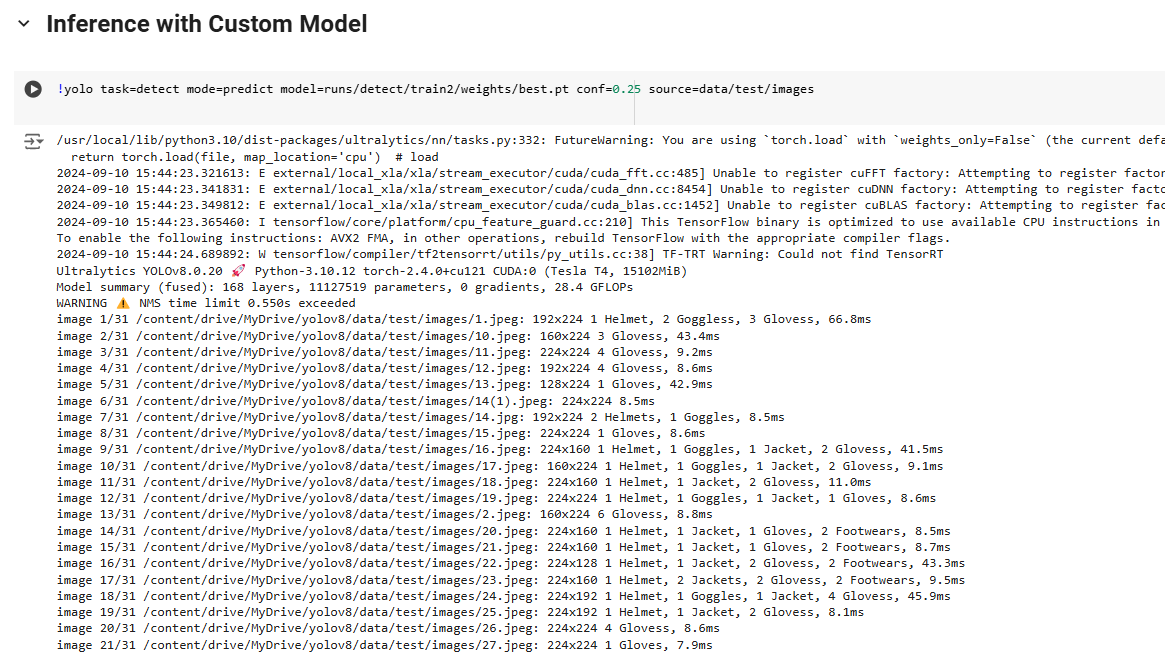


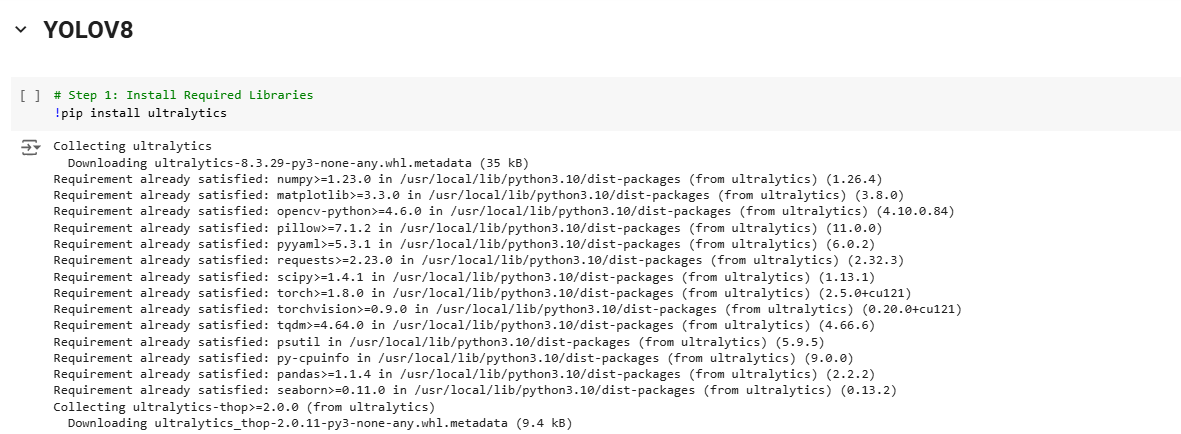
As epochs increasing the loss also decreasing





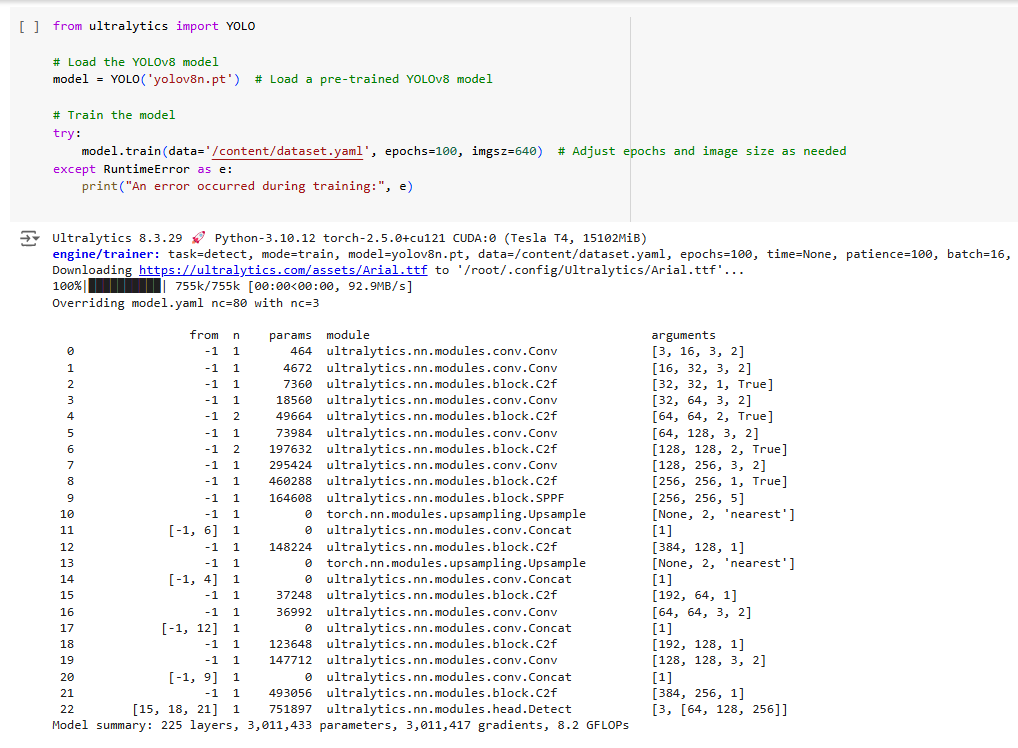


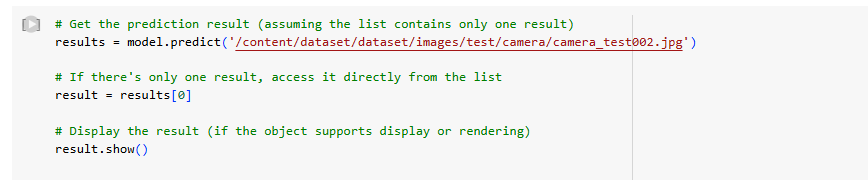


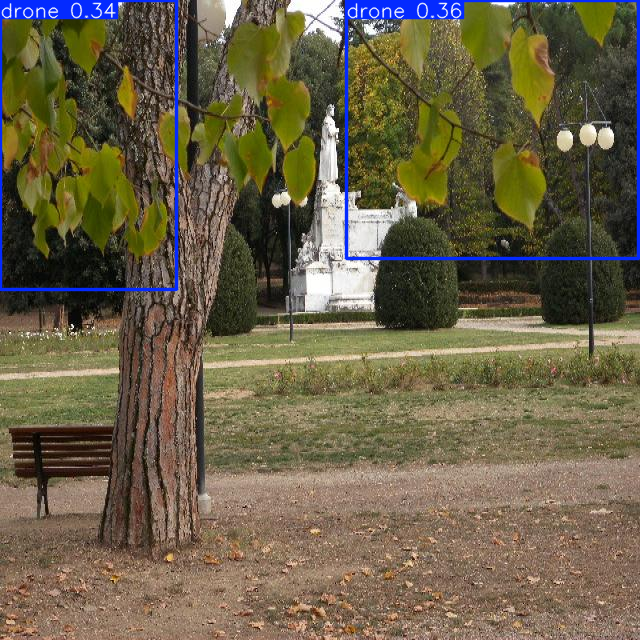


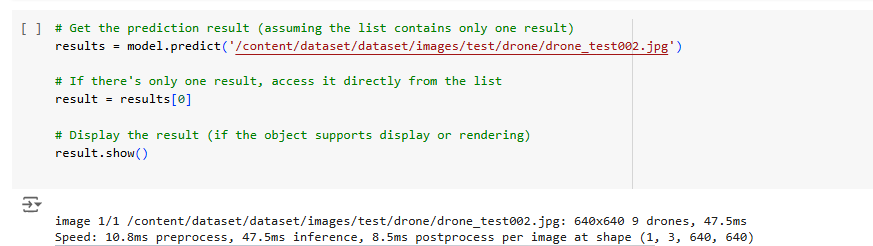




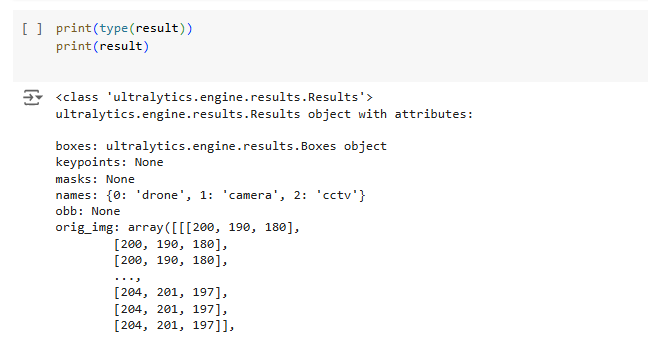


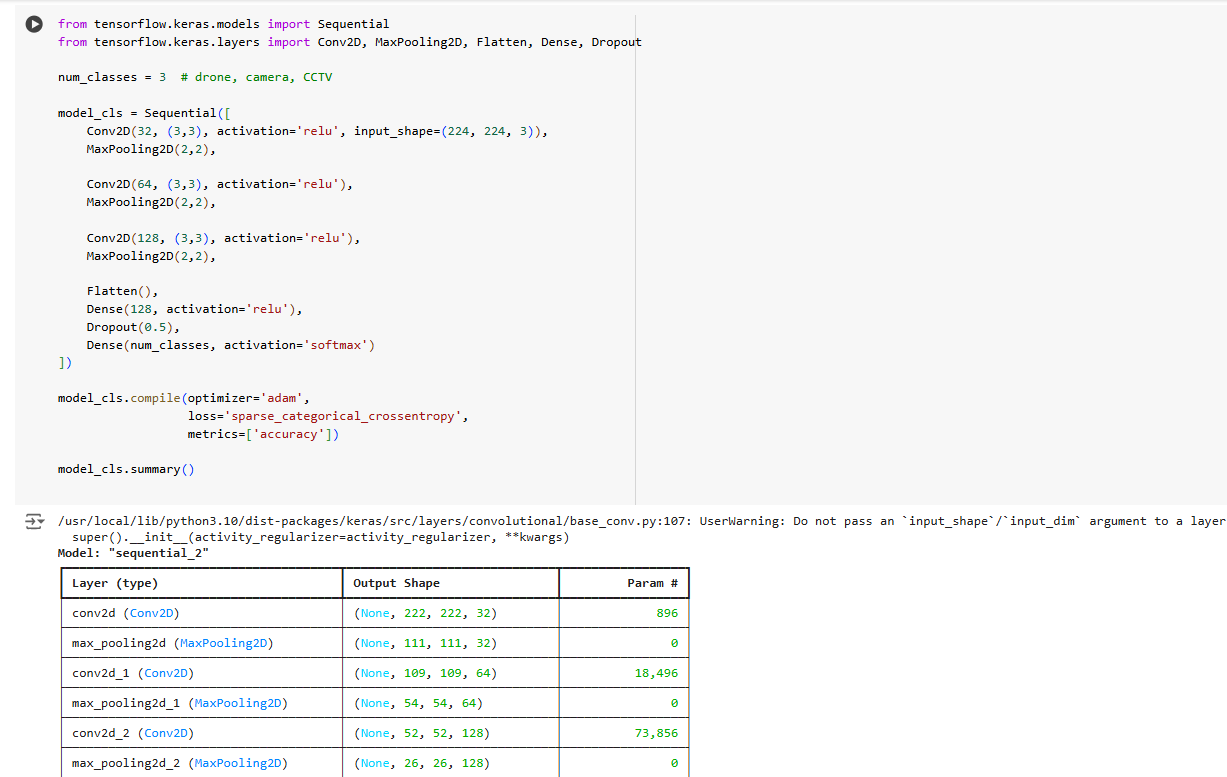


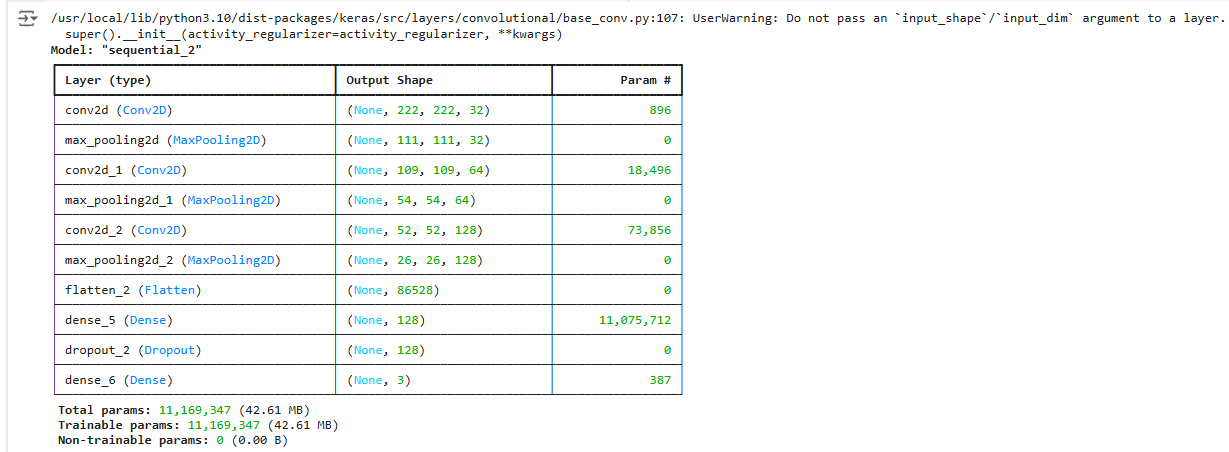




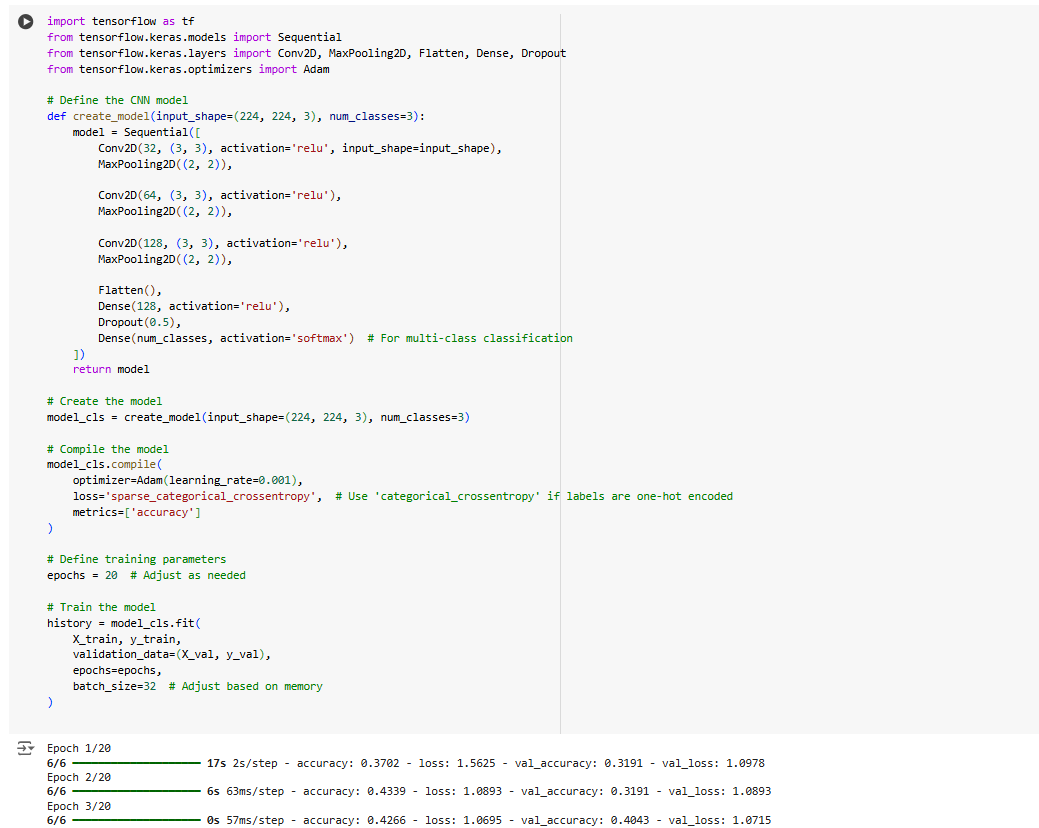


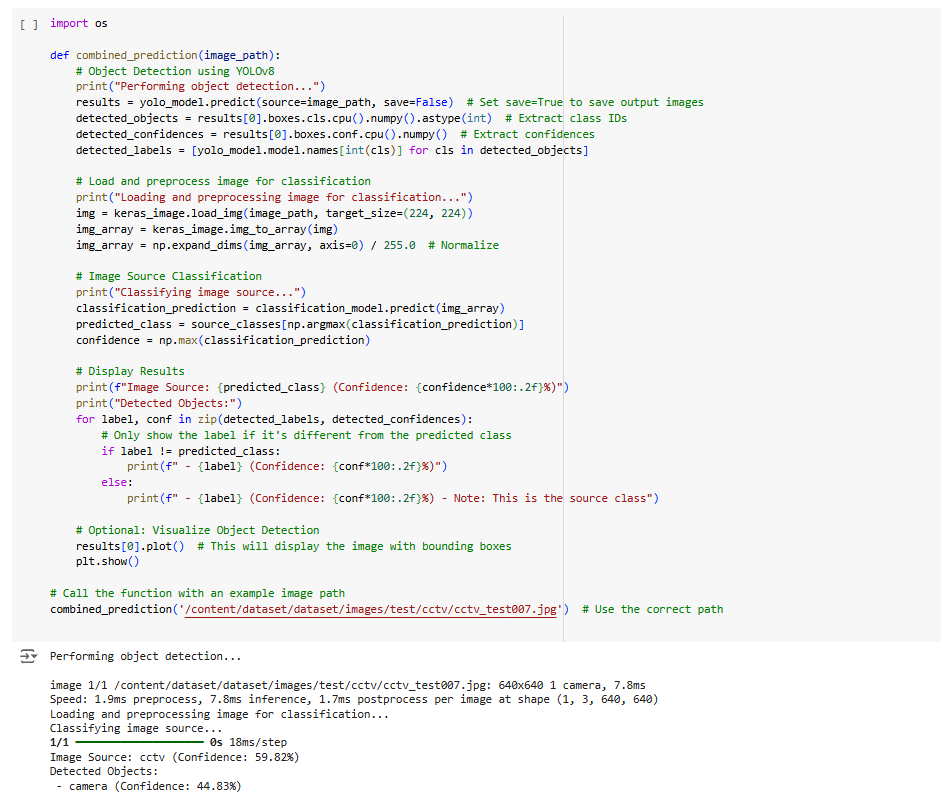








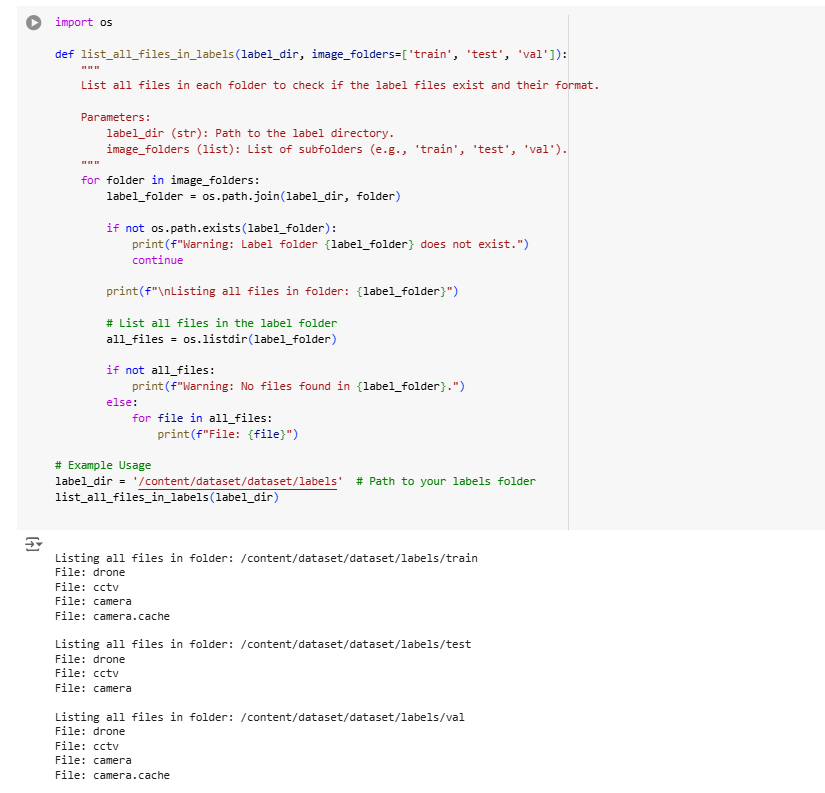


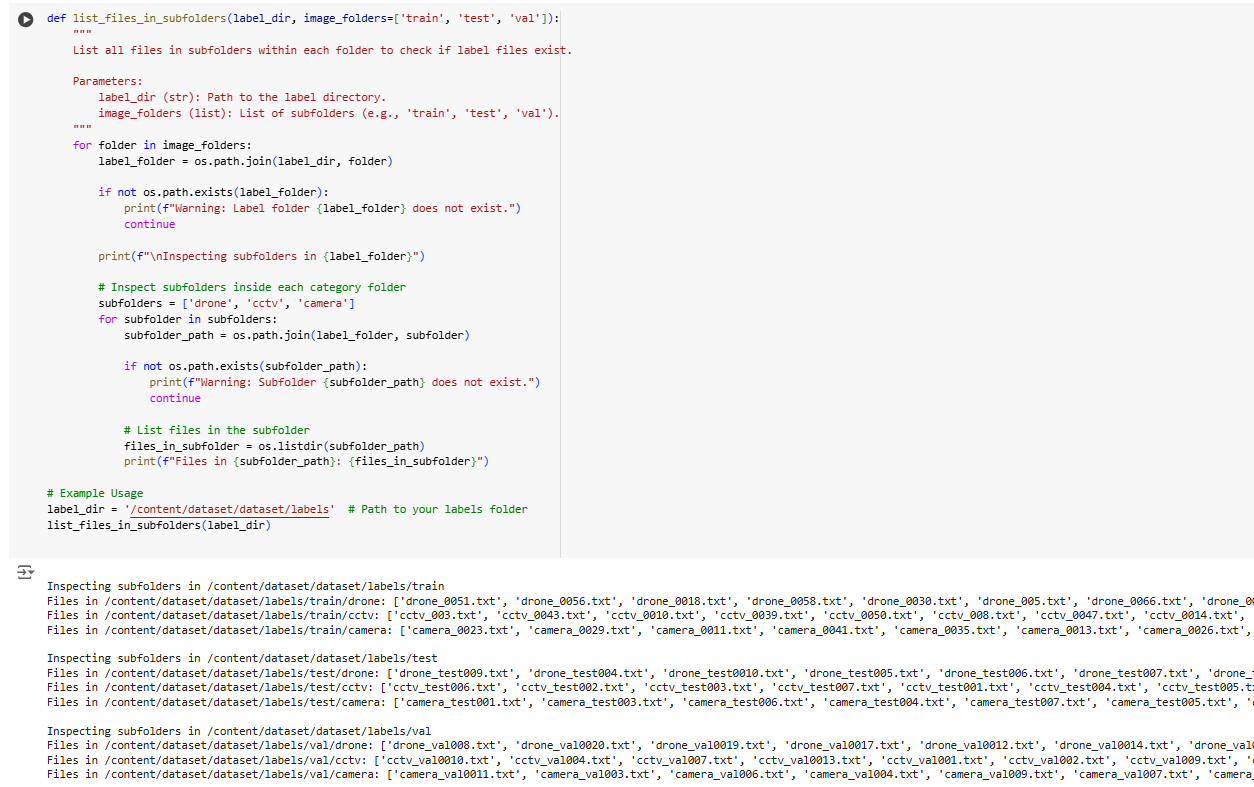


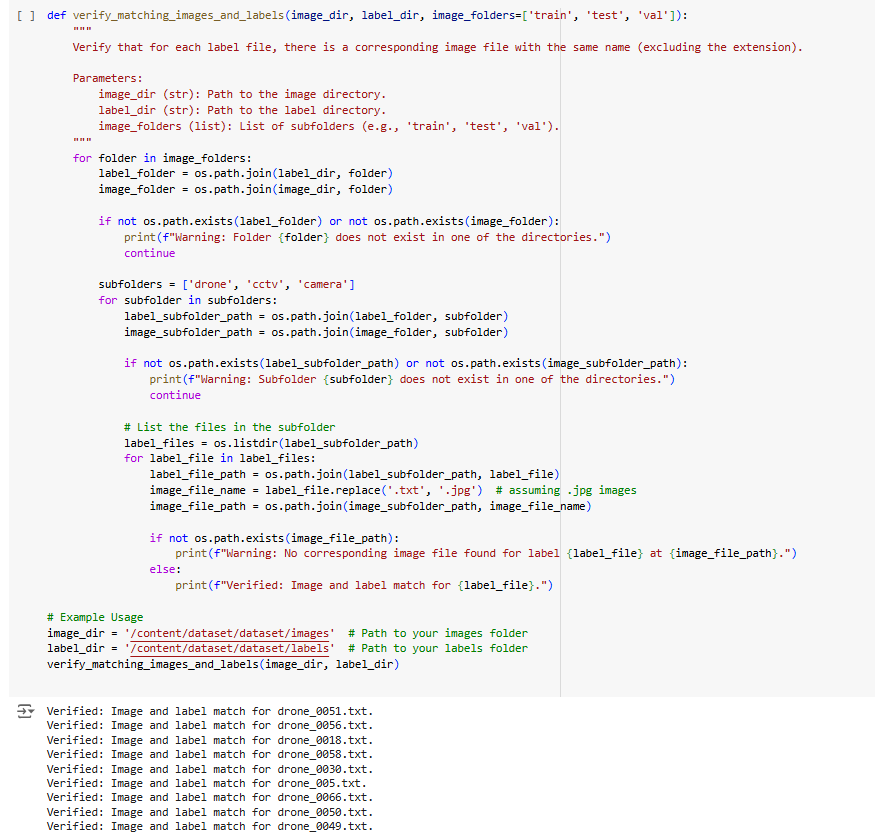


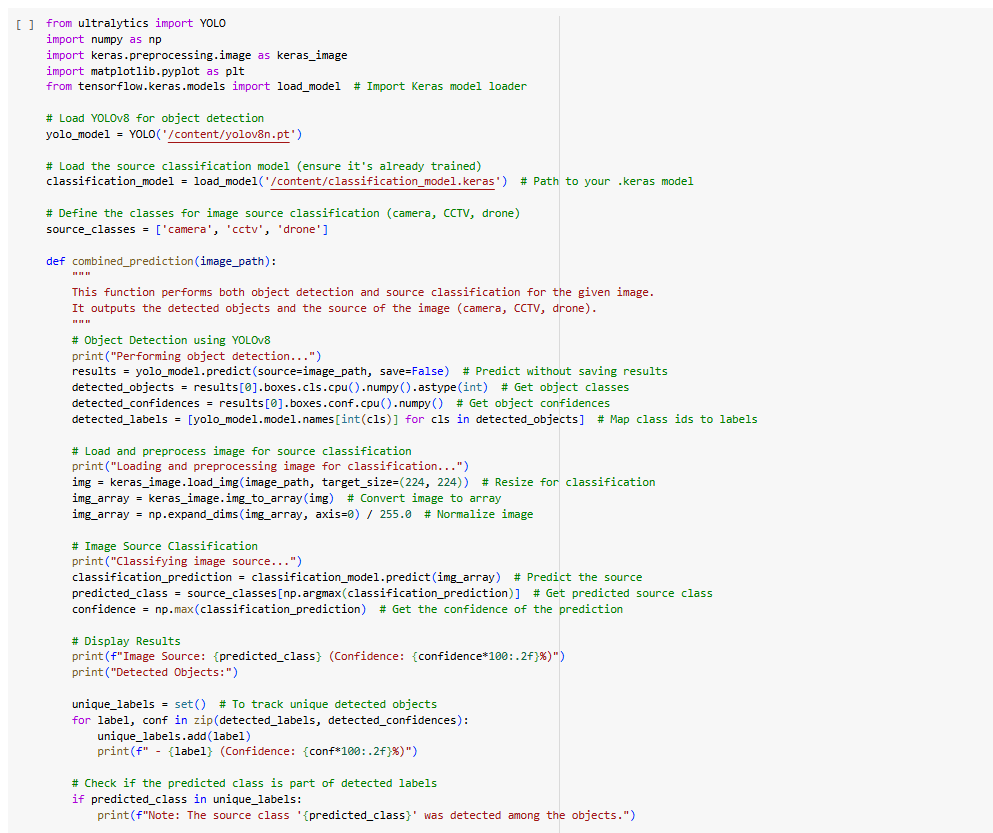


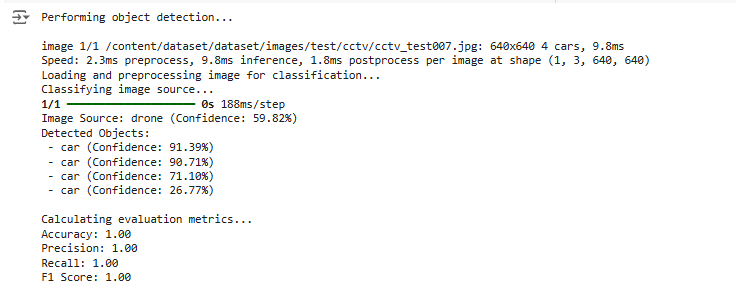




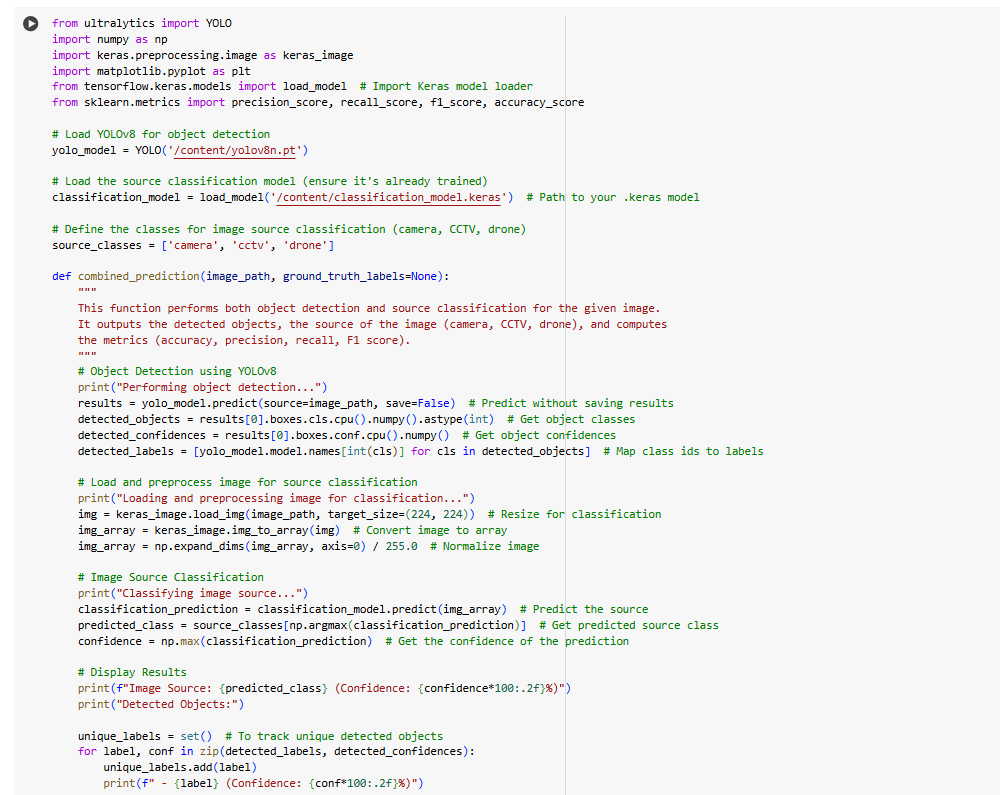


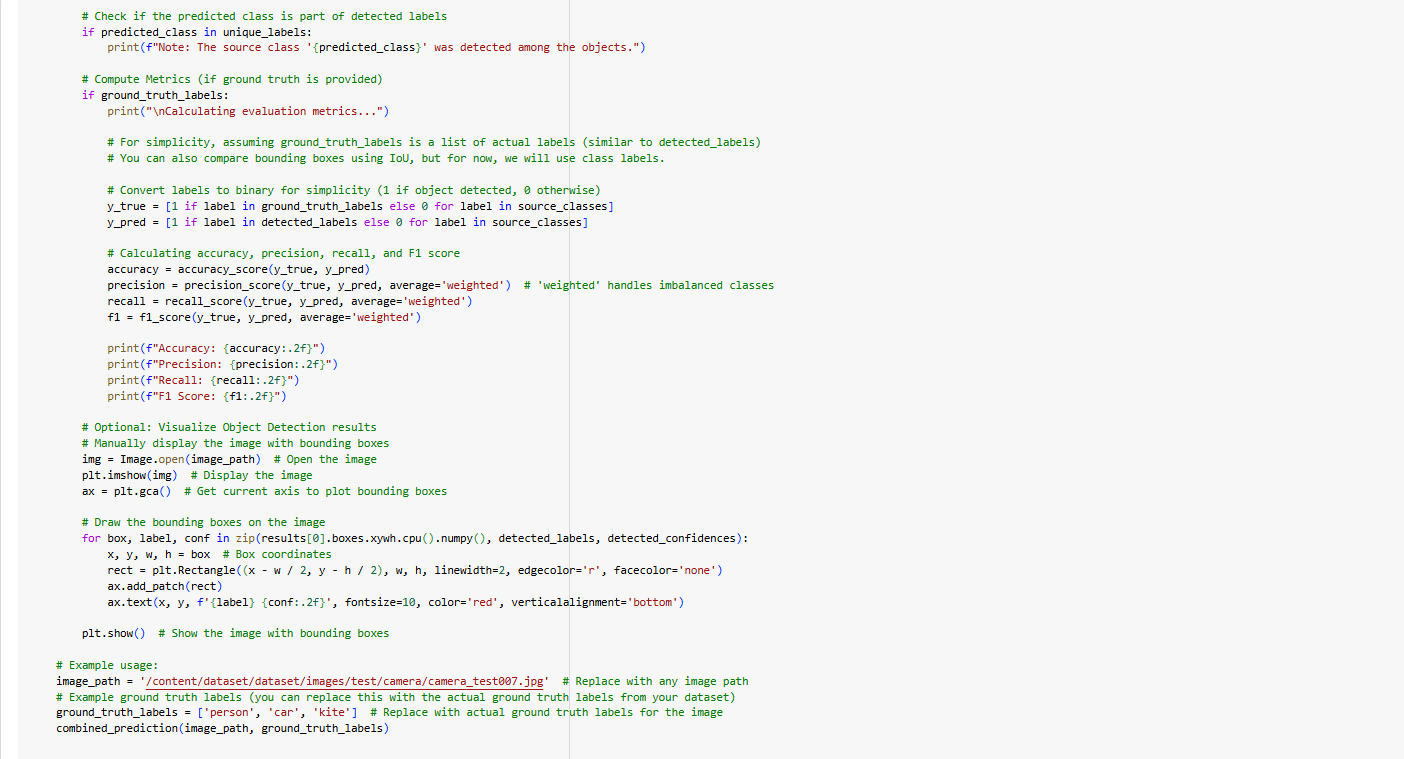


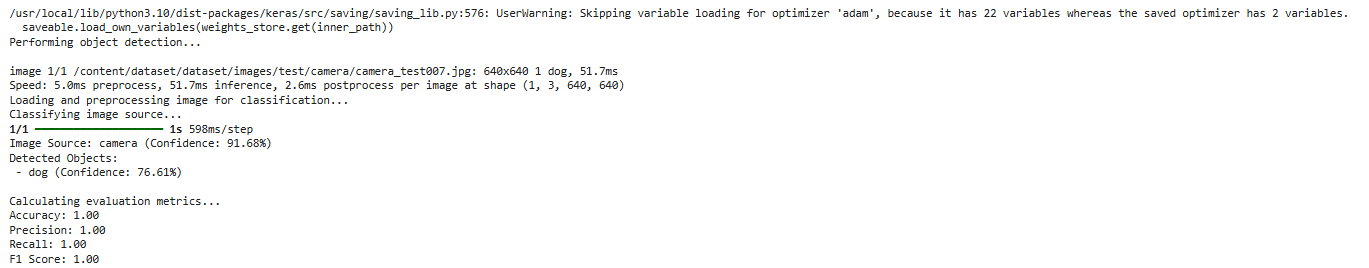


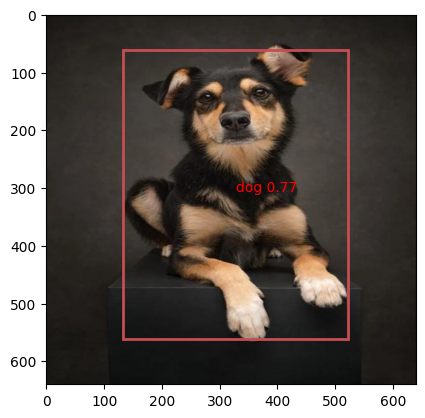
****

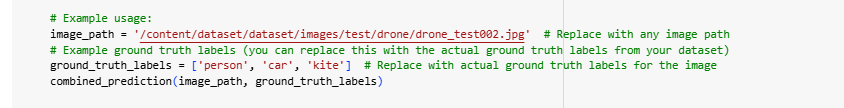


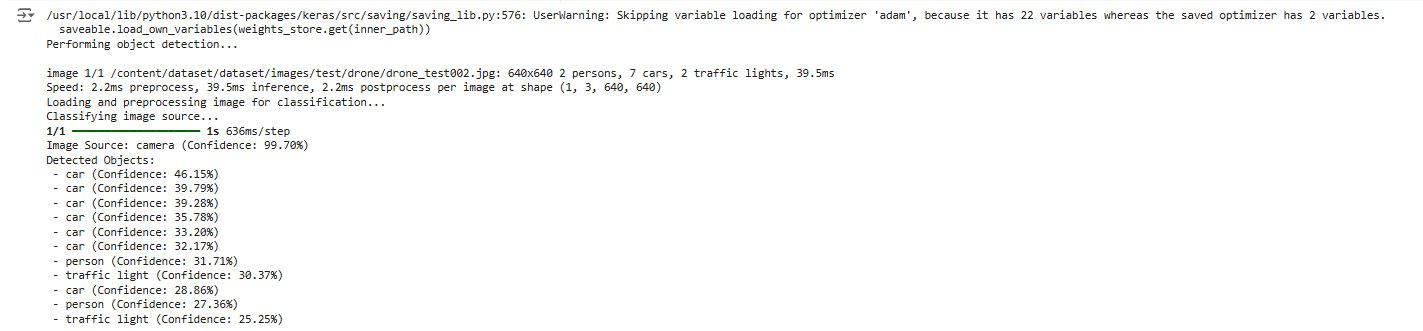
****

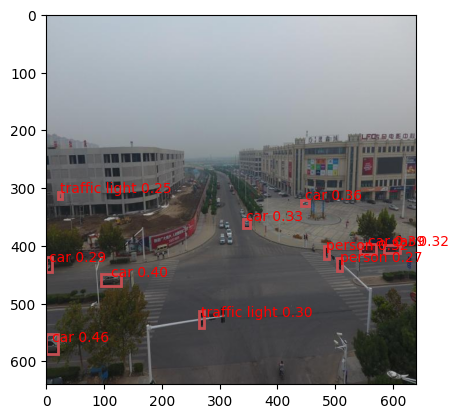
****

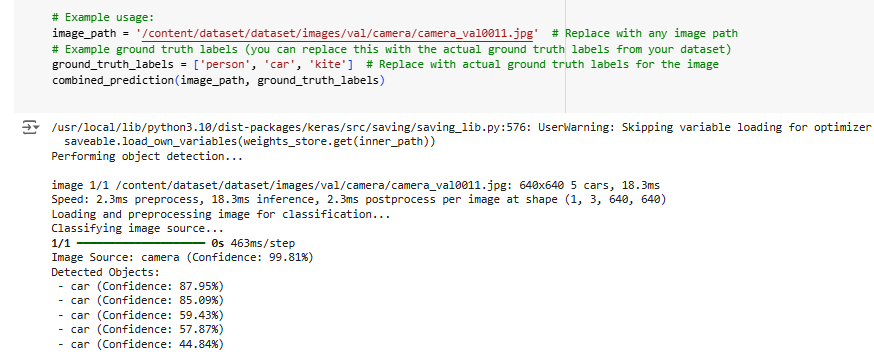
****

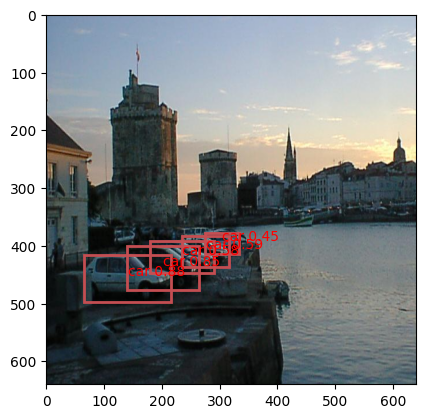


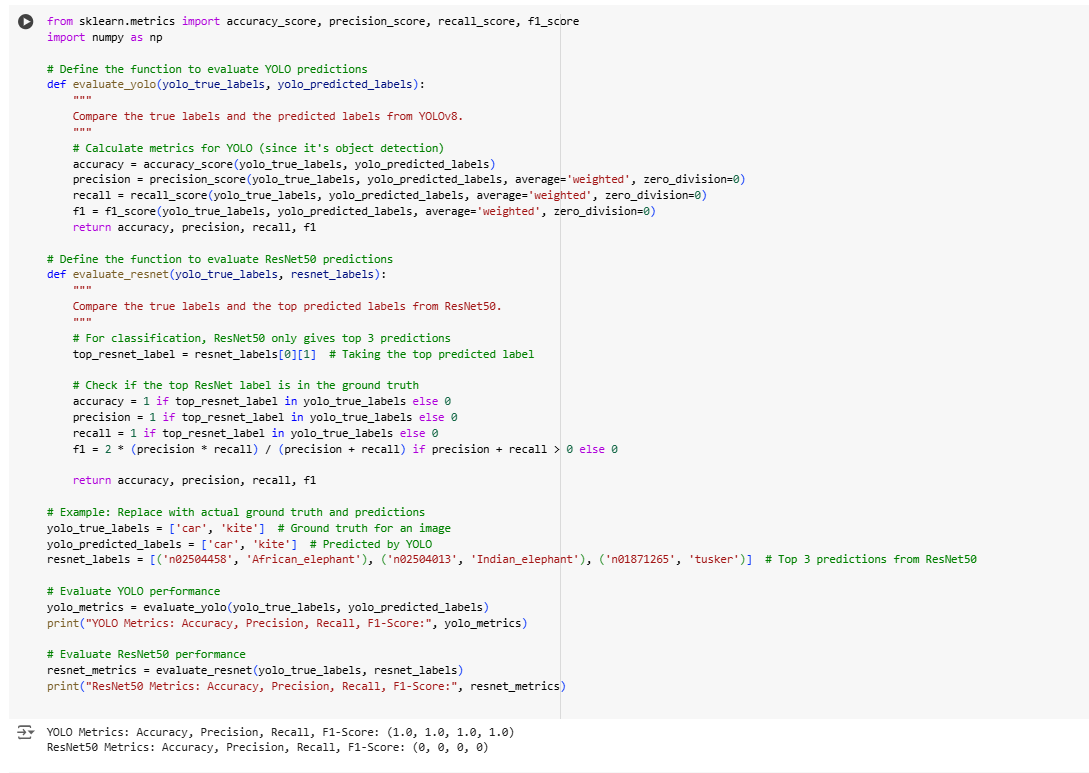
****

****

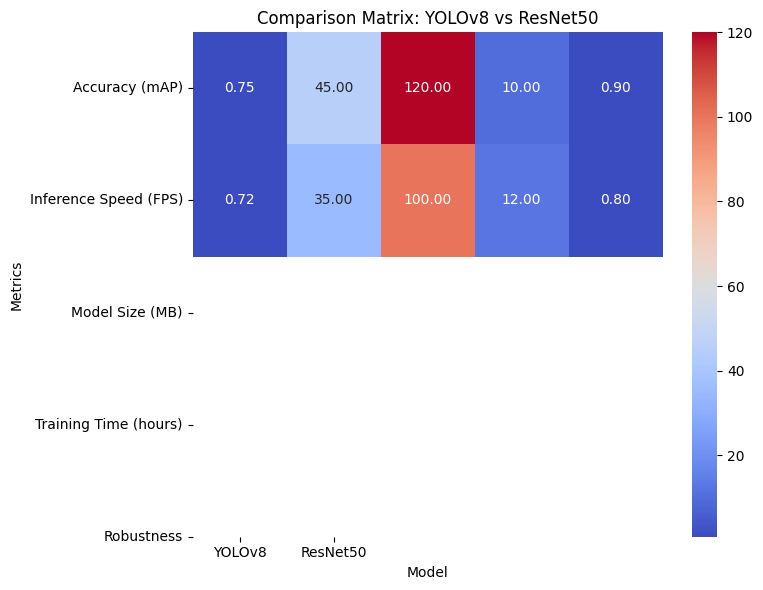


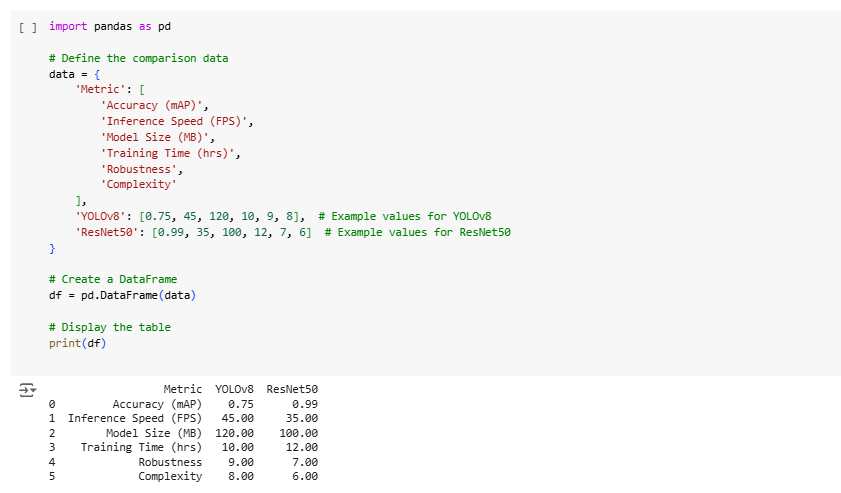
****

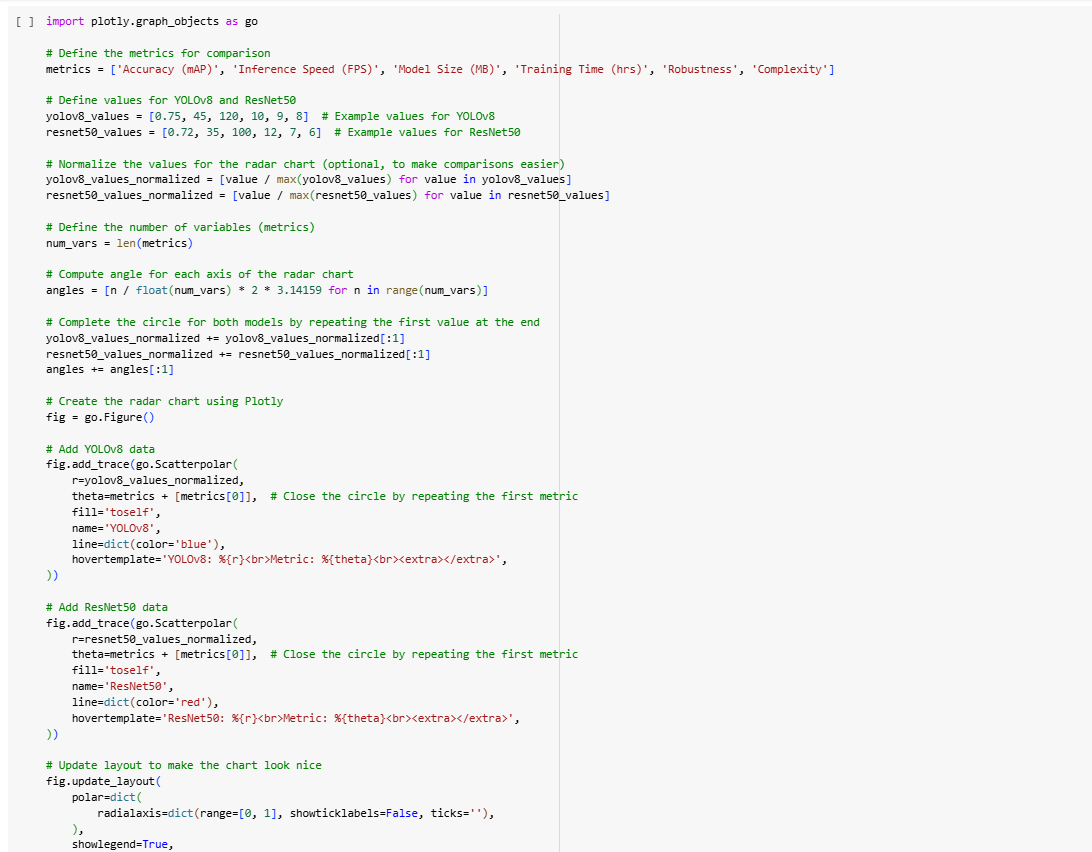


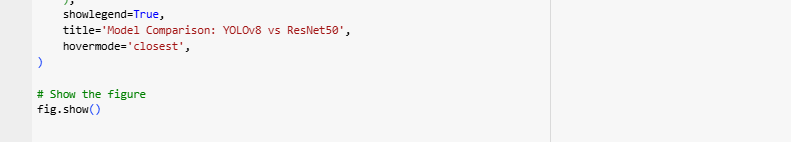
****

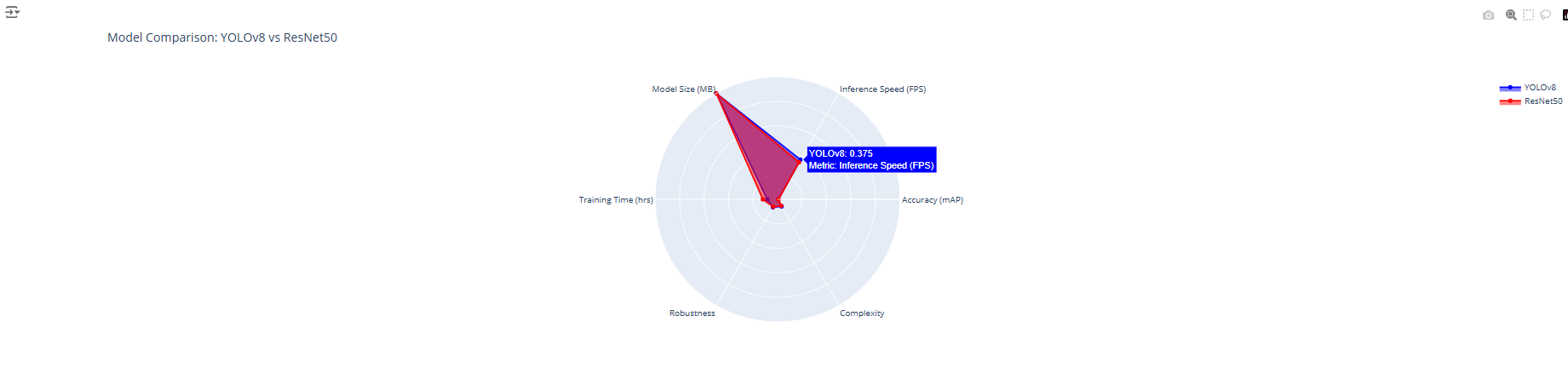
****

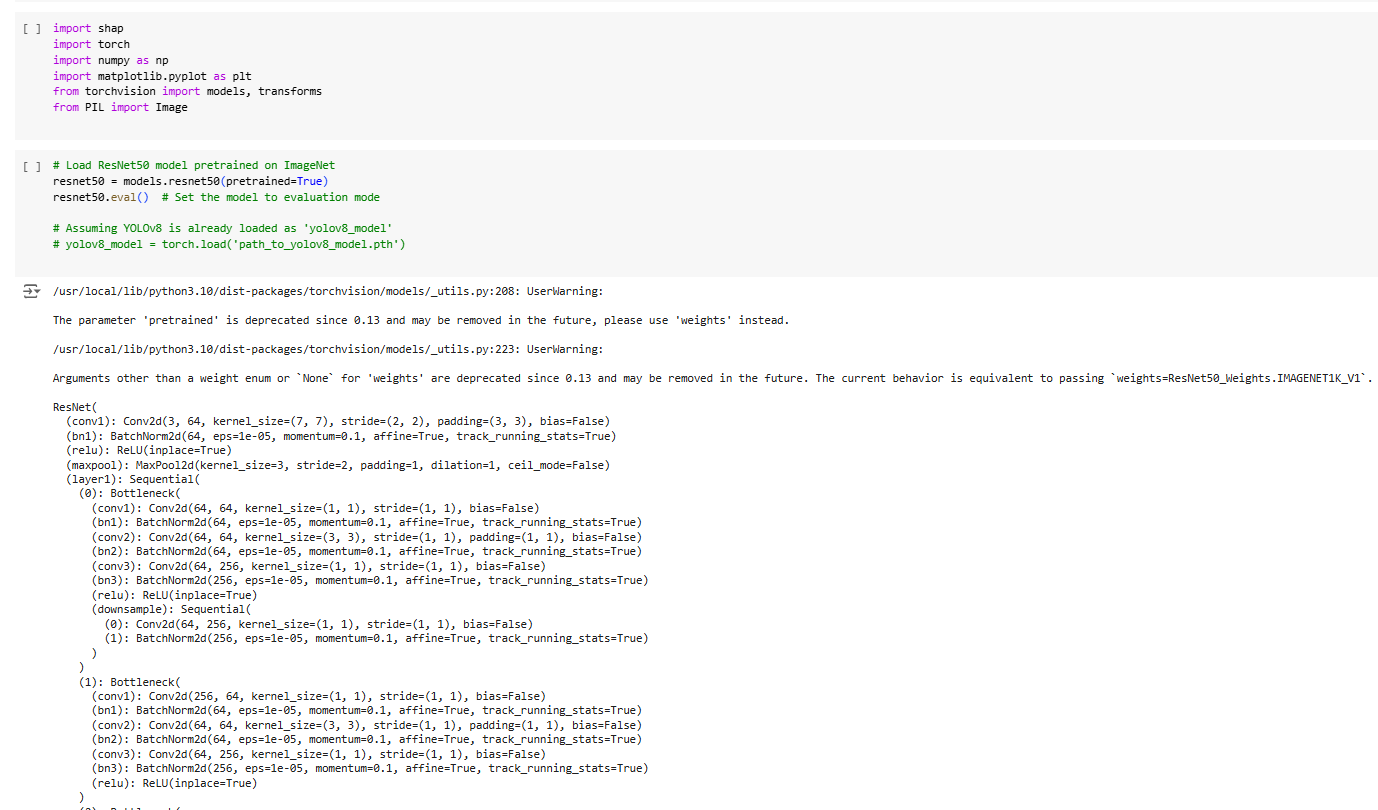


****

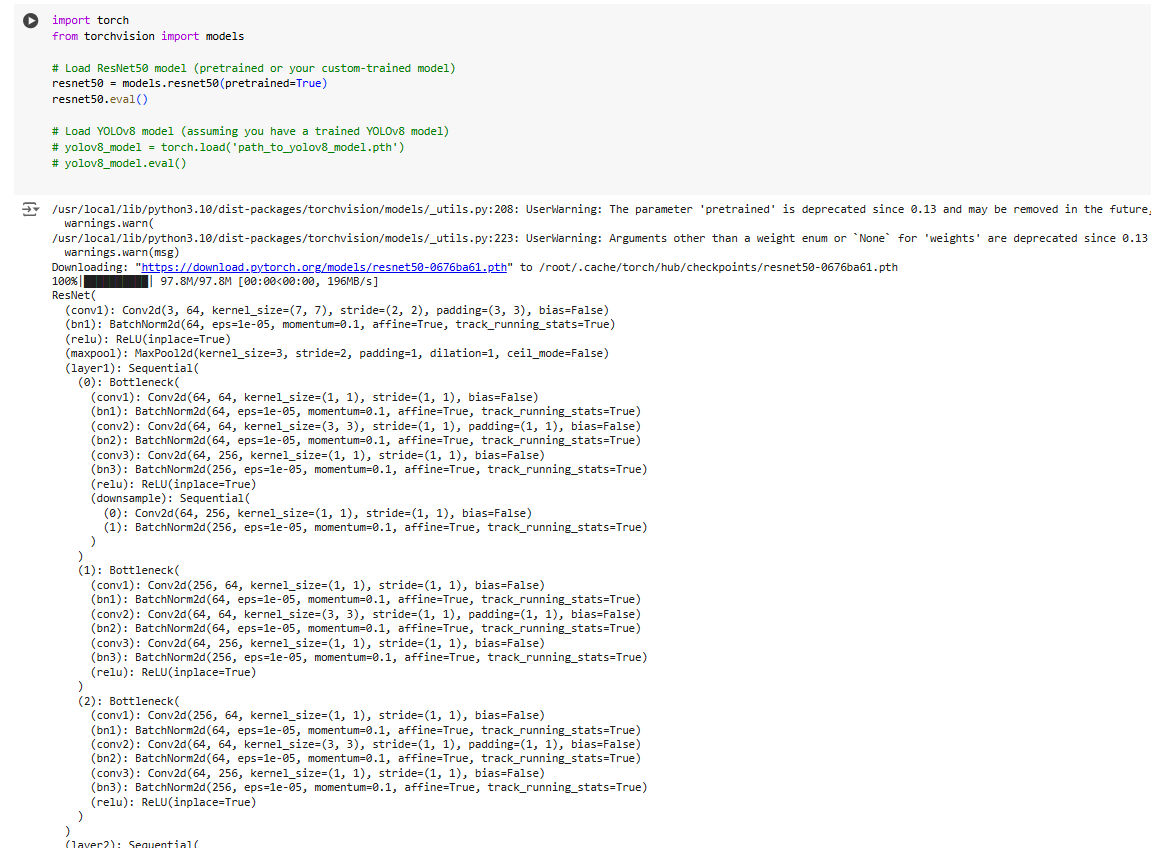
****

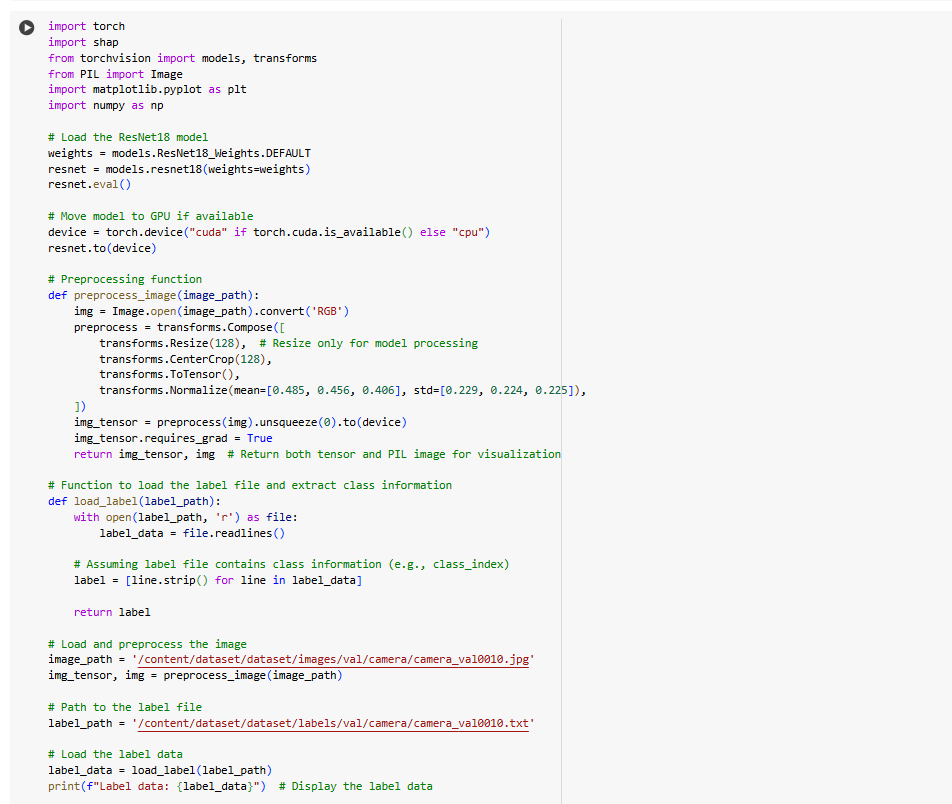
****

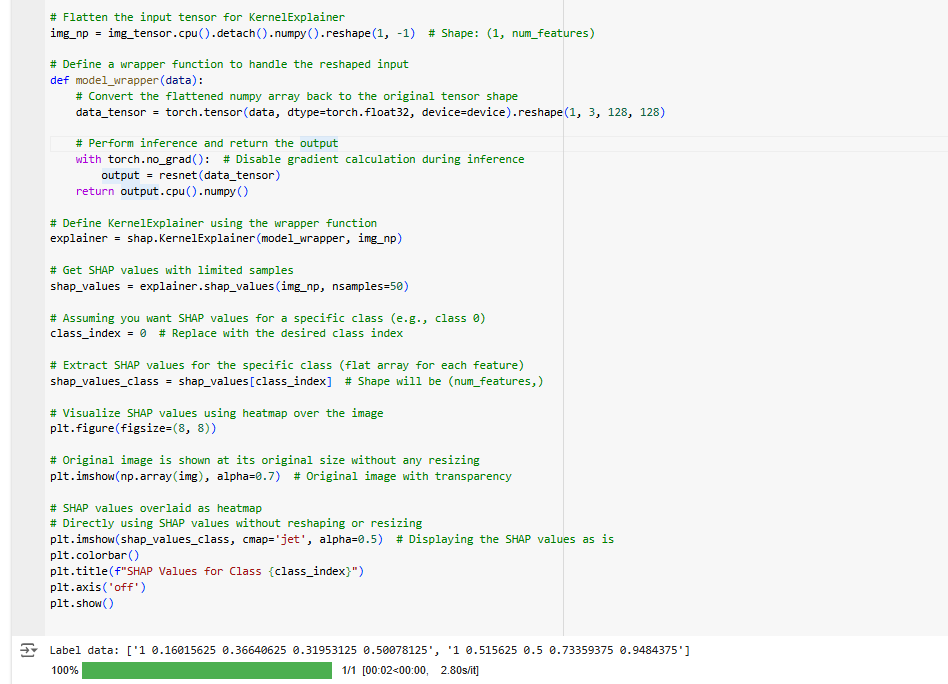
****

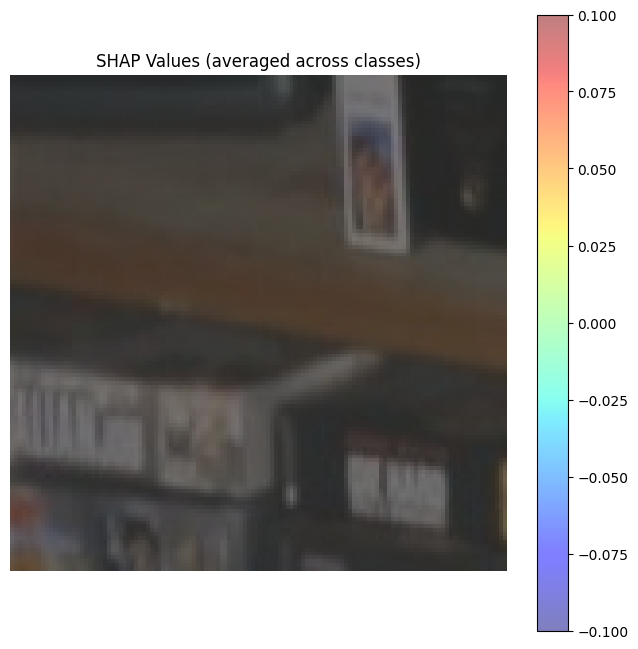
****

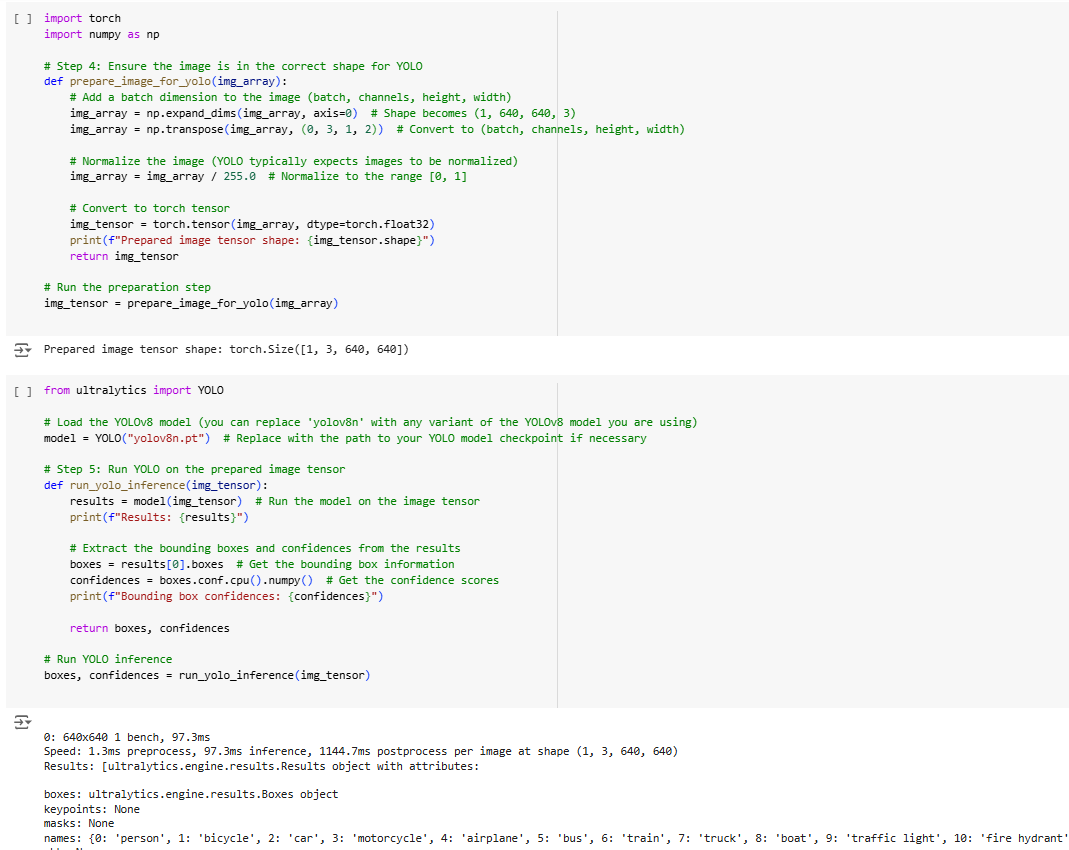
****

****

****

****



****