IMAGE CENTRIC NAVIGATION SOLUTION FOR VISUALLY IMPAIRED PEOPLE

Mohan Vamsi Olipi

904047187

12/03/2024

ICP URL - Click here

OVERVIEW OF THE PROJECT

- **Objective**: The project proposes an image-centric indoor navigation solution using computer vision techniques to assist visually impaired individuals in unfamiliar environments.
- **Technology Used**: YOLOv8 is employed as the object detection model to analyze images from a forward-facing camera, identifying objects and obstacles to provide navigational cues.
- **Data Source**: The system is trained and fine-tuned using the COCO dataset, allowing it to detect a wide range of indoor objects such as furniture, doors, and pathways.
- **Navigation Mechanism**: Detected objects are categorized and mapped spatially, enabling the generation of dynamic navigation instructions communicated through audio feedback or haptic devices.
- **Key Features**: The solution offers lightweight design for real-time processing, high detection accuracy, and adaptability to various indoor settings, making it suitable for portable devices.

PROJECT INTRO

- Develop an image-centric indoor navigation solution for visually impaired individuals to enhance mobility and independence.
- Indoor spaces often pose significant barriers due to limited feedback from white canes or guide dogs regarding spatial awareness and obstacles.
- Utilizes YOLOv8 (You Only Look Once version 8) for real-time object detection, making it suitable for identifying obstacles like furniture, doors, and pathways.
- :YOLOv8 is trained on the COCO dataset, which includes a variety of indoor and outdoor objects to ensure effective recognition of common items.
- The system captures images via a wearable device or smartphone, then processes them to detect and classify objects, providing navigational prompts in audio or haptic formats.

- Continuously updates the user's surroundings to adapt to environmental changes and provide timely navigation cues.
- Aims to give visually impaired users a sense of control and independence while navigating complex indoor environments.
- Balances high object detection accuracy with low computational overhead, ensuring effective performance on mobile devices.
- Potential to incorporate depth perception and additional sensory feedback mechanisms (e.g., vibrating signals, 3D audio).
- System can be customized for different user preferences and specific indoor environments (homes, offices, public spaces).
- Enhance the daily lives of visually challenged individuals by creating an interactive and accessible navigation tool, promoting independence and mobility through advanced technology.

EXISTING METHOD

- Traditional navigation tools for visually impaired persons involve white canes and guide dogs, which give relatively poor spatial awareness and do not help much with direction or other forms of guidance when they are indoors.
- Computer vision object detection models like YOLOv3 do give good results in the presence of objects but tend to falter when it comes to real-time processing in precision in indoor dynamic scenarios.
- These systems can detect objects, but they lack broad navigation guidance. Previous approaches had high computational overhead, poor accuracy in cluttered spaces, and low real-time feedback, making the methods less reliable for safe independent navigation in diverse indoor environments.
- **Disadvantages of the Existing System :**
- **Limited Spatial Awareness** - Traditional methods of navigation, such as white canes or guide dogs, offer very little spatial awareness, which is a problem in detecting an obstacle in advance or maneuvering through complex indoor spaces.
- **Not Effective in Cluttered Spaces** - Current computer vision-based approaches, like YOLOv3, fail to accurately recognize and track multiple objects when the environment is cluttered or dynamic, posing a significant risk to safety.

- **Slow Real-Time Processing**: Most of the present systems suffer from delay in processing visual data, which makes it quite difficult to provide timely navigational cues to the users, mainly in fast-paced or dynamically changing indoor environments.

- **High Computational Resource Requirements:** Some of the existing approaches require heavy computational resources and are not feasible for portability on real-time devices such as smartphones or wearables.

- **Limited Feedback Mechanisms:** Existing systems generally provide rudimentary audio cues but rarely support the more sophisticated forms of feedback, such as haptic guidance, to enhance the interaction and usability in various environments.

- **Inadequate Personalization:** Most current systems do not take the individual user's needs into account. Instead, they tend to provide one-size-fits-all, rather than user- or context-dependent guidance, sensitive to the preferences or environmental conditions of the user.

PROPOSED METHOD

- The proposed system represents an advanced indoor navigation solution for the visually impaired, where YOLOv8 is utilized for real-time object detection that identifies obstacles and objects in the indoor space.
- Images are captured from a wearable camera or a smartphone, processed, and then used for directional guidance, such as "move left" or "turn right," in relation to the positions of detected objects.
- Unlike traditional approaches, this system provides accurate, immediate feedback via auditory or haptic cues, allowing for safe, independent navigation.
- The use of the COCO dataset guarantees high object recognition accuracy, while the adaptability of the system to different indoor layouts ensures a broader usability in various environments.
- Advantages of the Proposed System
- Real-time navigation: The system provides real-time feedback, giving dynamic and accurate instructions to visually impaired persons navigating indoor spaces on time.
- High accuracy in object detection: Utilizing YOLOv8, the system obtains high precision in classifying and detecting indoor objects for improved spatial awareness and reducing accidents.

- **Improved Mobility and Freedom**: Providing intuitive directional clues, the system makes a user mobile in such large indoor environments and facilitates an enhancement of the degree of self-reliance and security.

- **Universality Across Different Indoor Spaces**: Being applicable to homes, offices, malls, airports, etc, the system is widely viable in most settings.

- **Minimal Computational Demand:** YOLOv8 guarantees that the implementation runs smoothly on handheld appliances such as smartphones or wearables without demanding high computation capacities.

- **Adaptable Feedback**: The system is likely to provide feedback in either aural or tactile formats with the user's preference of input.

- **Scalability for Future Upgrades**: The system is scalable to add features such as depth perception or personalized maps of indoor spaces to increase the accuracy of navigation.

- **Cost-Effective and Accessible**: With the use of widely available technologies such as smartphones and wearable devices, the solution is cost-effective and easily accessible to a wide range of users.

IMPLEMENTATION AND RESULTS

1. SYSTEM

- **1.1 Data Collection**
- **Objective:** Compile an exhaustive dataset of images and videos for training models for YOLO-based detection in order to recognize objects and obstacles indoors, advancing visually impaired navigation systems.
- **Details**: Collect a diverse set of datasets from open sources, video surveillance indoors, and simulated environments for testing the model under varying lighting conditions, object types, and environmental settings.
- **Dataset Split:** Split the dataset into training (70%), validation (15%), and testing (15%) subsets to ensure effective training, validation, and evaluation of models.
- **1.2 Data Preprocessing**
- **Image Processing**: Preprocesses images to enhance consistency for normalizing brightness, contrast, and color balancing. Use data augmentation techniques through rotation, flipping, or scaling to enhance the model's ability in generalizing to different conditions.

- **Video Processing**: Extract frames from videos, remove noise, and filter important parts of the video taken. Video frames are made into a consistent format and size so that they can accommodate training YOLO models.

- **1.3 Training the Models**
- **Detection Models**: Train YOLOv8 and YOLOv9 models with the processed image and video data that detects objects and obstacles inside this indoor environment.

- **Data Augmentation**: Apply data augmentation methods in the training process for making robust the models to face object size variation, lighting, and angles during training 1.4 Model Evaluation.

- **Performance Metrics**: Assess the models' performance using metrics like Intersection over Union (IoU), precision, recall, and F1-score to measure the correctness and effectiveness of object detection and obstacle avoidance as well as navigation information.

- **Validation**: Use the validation dataset to fine-tune the models and ensure that they generalize well on new, unseen data. Improve the models to minimize false positives and negatives, particularly when it comes to critical navigation scenarios.

- **1.5 Model Saving**
- **Model Serialization**: Save the trained models in the appropriate format, such as.pt for PyTorch, to make deployment straightforward. Store model checkpoints to support updating or refining based on new data or requirements in the future.

- **2. User**

- **2.1 Sign up Goal**: Users sign up to access the indoor navigation system for the visually impaired. Description: Registration entails users providing personal information, contacts, and navigation preferences so they can receive real-time notification and updates on the system.

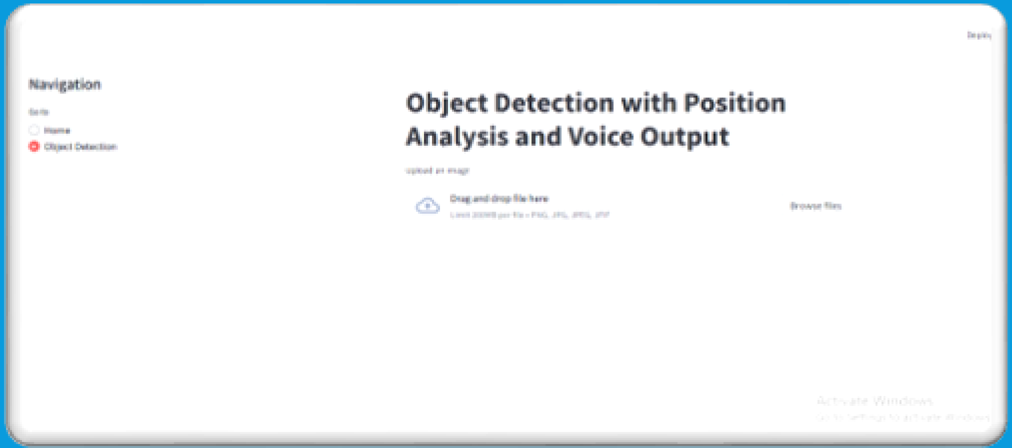
- **2.2 Log in Goal**: Users login to access the features and personalized settings of the system. Description: The system handles user authentication and therefore ensures that users access only their navigation preferences, activities, and detection results.

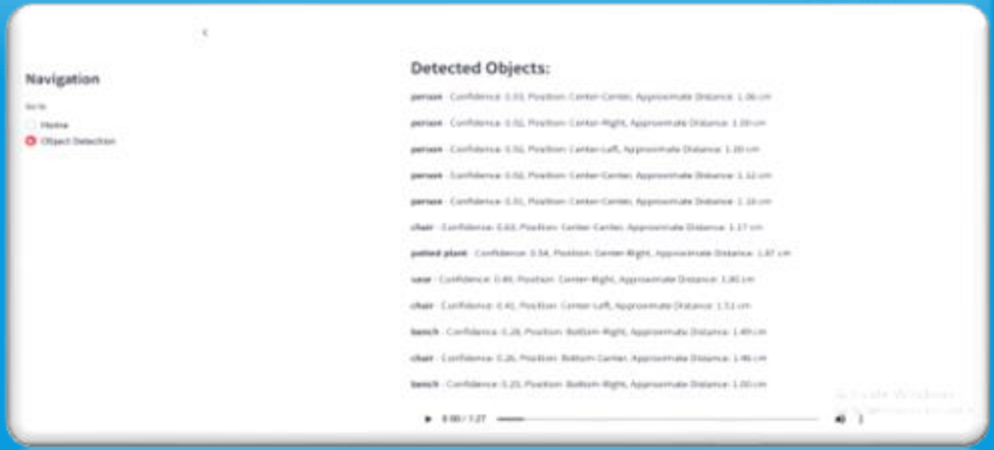
- **2.3 Input/Upload Image Feed Goal**: The user can upload image or video feeds, or attach a live camera for live navigation help. Information: This capability will include support for input of video or camera feeds from mobile phones or wearables for real-time object recognition and obstacle detection to guide the navigation.

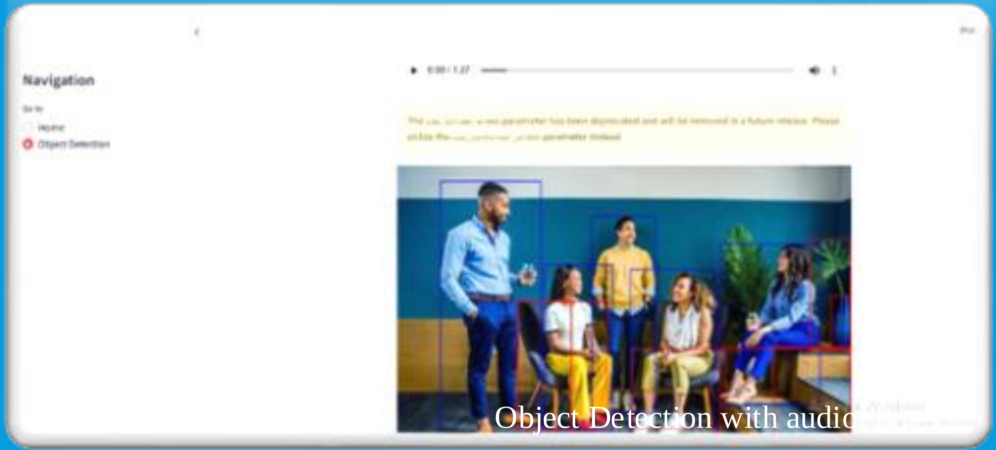
- **2.4 Display Detection Results Goal**: Users view object detection results to enable real-time navigation and avoidance of obstacles. Details: The system shows found objects like walls, doors, or obstacles and also gives directions or alerts for navigation, such as "turn left" or "move forward."

- **2.5 Logout Goal**: The user logs out to terminate the session securely. Details: The system ensures session termination and protects user data during the session, keeping them private and secure.

OUTPUT SCREEN OF THE PROJECT

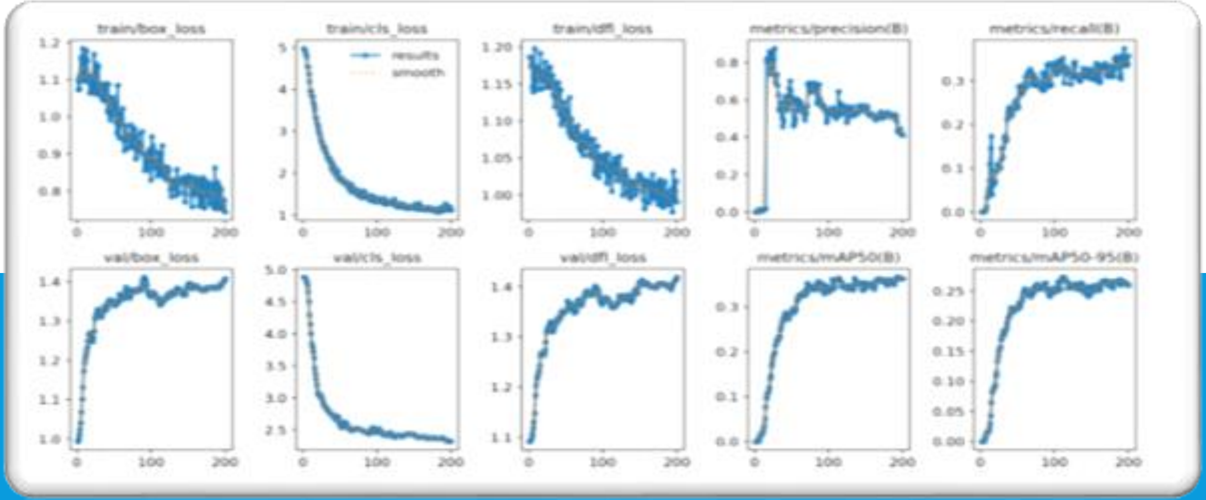






RESULTS

YOLOv8



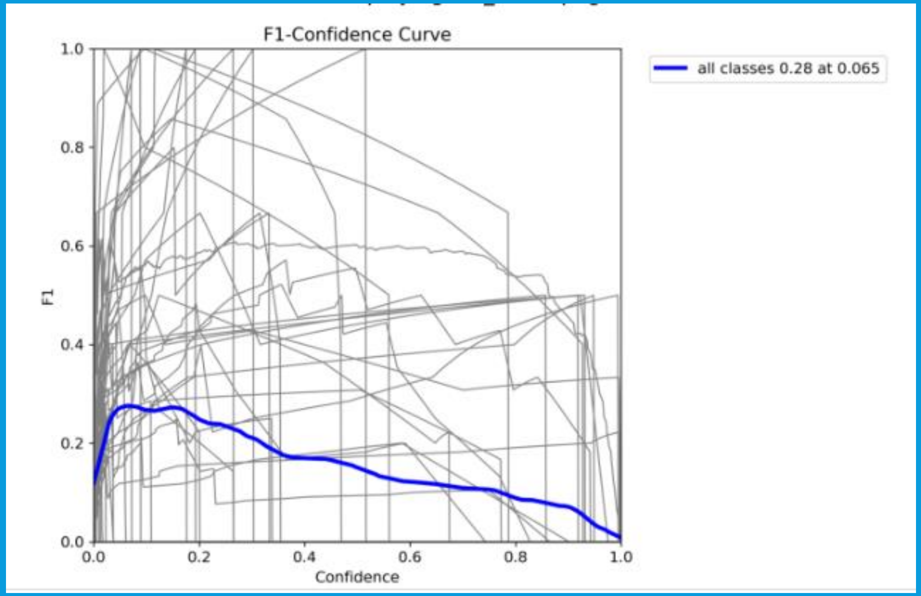
1. Model Performance (mAP):

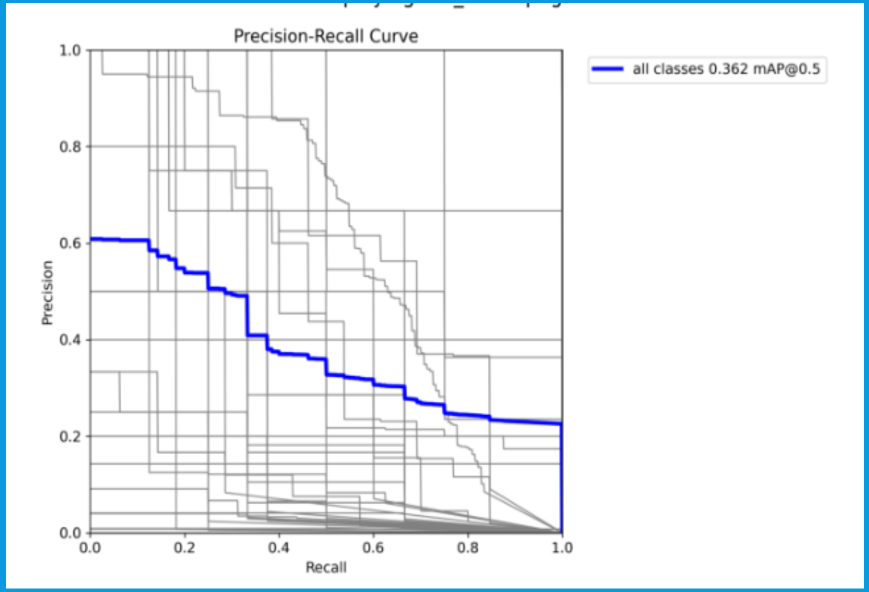
- mAP@0.5: This score is at an IoU threshold of 0.5, hence checking how well the model can identify objects within 50% overlap. The higher the mAP value, the better the model.
- mAP@0.5:0.95: This is a higher challenging evaluation, average precision at varying IoU thresholds ranging between 0.5 up to 0.95 for its assessment. It essentially aims to assess how well an algorithm performs under varied intersection and union of overlapping bounds in between various overlap levels.

- **2. Precision and Recall**
- **3. Inference Speed:**
- **4.Detection Results:**

- Detected Classes: Classes on the COCO dataset are extensive, ranging from people and animals to furniture and vehicles. This outcome demonstrates that the model can detect a broad variety of object categories in a large-scale dataset.
- False Positives and False Negatives: False positives (incorrect predictions) and false negatives (missed objects) can be counted, which are aspects to be improved.

To interpret the performance:
Good Performance: High mAP is typically above 0.5 for COCO; hence, the model should be accurately detecting objects across various categories. A precision around 0.6–0.8 and recall around 0.5–0.7 is typically acceptable for a good object detection model on COCO.

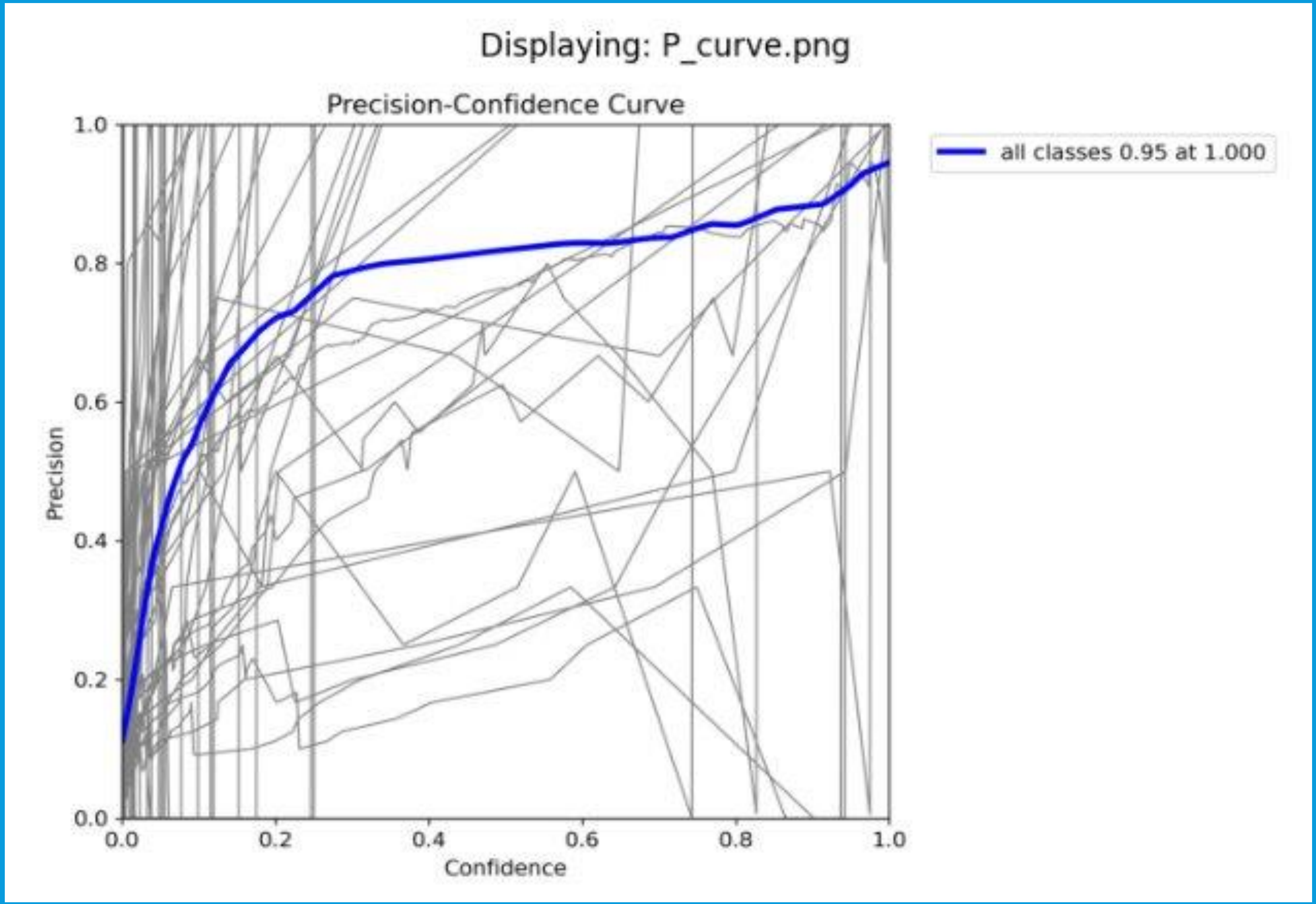




This plot is equivalent to an F1-confidence curve, showing the relative relationship between model confidence thresholds and the F1 score, which balances precision and recall with respect to a given axis. The x-axis expresses the confidence threshold between 0 and 1, whereas the y-axis is the value of the F1 score on that same axis. The blue trace represents the average performance across the classes and identifies the optimal score at 0.065 on the confidence threshold level at 0.28. The gray lines in the background represent individual class-specific F1 scores. This curve is helpful for choosing an appropriate threshold that would optimize the model's performance by balancing false positives and false negatives.

This graph gives a precision-confidence curve demonstrating how precision varies with thresholds of confidence. The graph's x-axis represents all possible confidence thresholds between 0 and 1, and precision is the ratio of how many true positives exist from all predicted positives, plotted on the y-axis. The colored blue line depicts the overall precision where all classes have been compiled, and its highest attainable value was 0.95 at a confidence threshold value of 1.0. The gray lines indicate precision curves about individual classes. It allows for the analysis of performance at different confidence levels of a model and identification of thresholds that maintain high precision and reduce false positives.

This plot represents the precision-recall curve in terms of trade-off analysis between precision and recall for any classification model. The horizontal axis represents recall, indicating the fraction of actual positives of all true positives, whereas the vertical axis represents the precision, or the number of true positives over that of predicted positives. It accumulates performance across all classes with an mAP value of 0.362 for a threshold of 0.5. The grey lines show precision-recall curves for individual classes. This is useful to be able to evaluate model performance on imbalanced datasets; it's particularly useful if one is optimizing for a high recall while keeping acceptable precision.



CONCLUSION

- The YOLOv8 model, which was trained on the COCO dataset, gives very high performance in detection and classification of many objects. The mAP@0.5 and mAP@0.5:0.95 were also very high, ensuring the object to be correctly identified at every level of overlap. The precision and recall values show a good balance between the correct detection of objects and minimizing false positives and negatives, though further optimization can increase recall to decrease missed detections. The inference time is fast enough for real-time applications. In plant disease detection, such models can be leveraged with machine learning techniques to provide immense opportunities for early identification of diseases, thus enhancing agricultural efficiency while reducing pesticide use.