Detection of person talking with mobile in no mobile zone

Team: Forger House

Heritage Institute Of Technology, Kolkata

| Anidipta Pal	Ankana Dutta	Anjishnu Mukherjee	Ananyo Dasgupta
1st Year B.Tech in CSE – AIML	1st Year B.Tech in CSE – AIML	1st Year B.Tech in CSE – AIML	1st Year B.Tech in CSE - AIML

Introduction:

Dealing with people making calls in no-mobile zones or inappropriate places can be frustrating.Phone calls in quiet or designated no-mobile zones can disrupt the peace and quiet of the environment. Phone use in certain areas like gas stations, train and metro stations, and hospitals may pose safety risk.To cope with this problem, we propose the idea of a mobile detector model system, that would detect a person talking on a phone at a no-mobile zone, and set an alert and would eventually store the details of the person in the database.
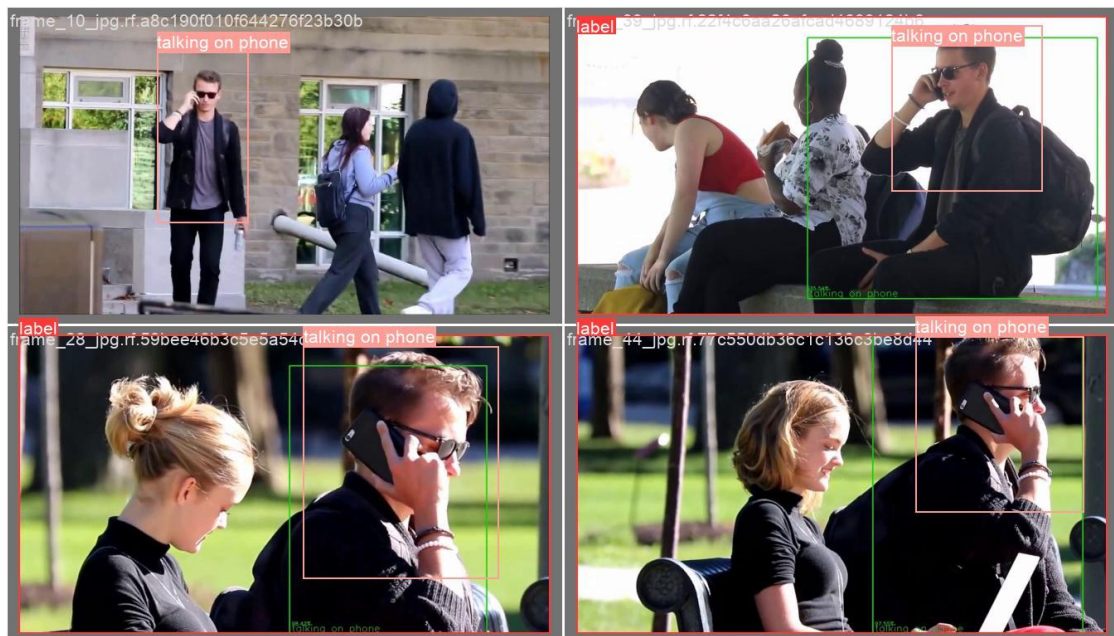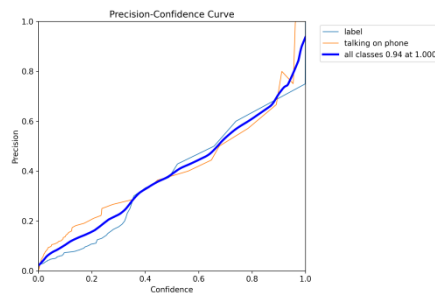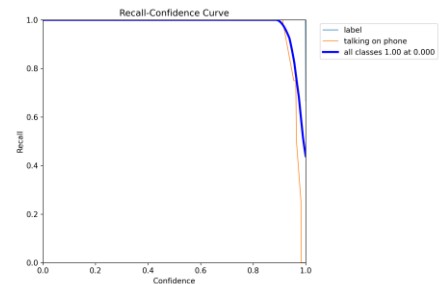

Idea/Approach Details:

- **Data Collection and Annotation:** We collected the CCTVs and Surveillance Camera Footages and a dataset of images are made from the video with *40FPS (Frames per Second)* . Annotation of these images with bounding/anchor boxes around the objects is made by using LabelImg.

- **Image Preprocessing:** The images are reshaped , scaled , normalized , filtering , and processed for training , validation and test sets under the YOLO's.yaml extension, which includes the class label and the bounding box coordinates relative to the image size , for the model training.

- **Model Building and Configuration:** Two Models , YOLOv8x and a Hybrid Architecture of YOLOv8 and ResNetv2 , are configured and made for the person talking in mobile and for the public place detection , respectively. The Models are made to detect and classify using the (hyper)parameters like img_size , batch , anchor boxes , activation function , etc.

- **Model Training:** The Imaginally Dataset is splitted in the ratio 70-20-10 and is trained for both the models.During training, the model learns to detect objects and adjust its weights to minimize the detection loss.

- **Evaluation:** The trained model is evaluated to assess its performance. Calculate metrics such as precision, recall, and Mean Average Precision (mAP) are used to measure the model's accuracy.
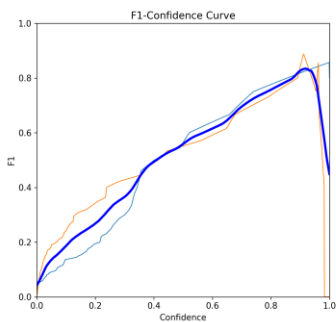
| F1-confidence curve | Precision-Confidence Curve | Recall-Confidence Curve |

- **DataBase Connection:** For the fine System , the model with connect with a SQL Database and the fine information will be stored there temporarily and will be shown in a chart format. (**To be developed and implemented into the project**)

- **Fine-Tuning:** The performance of the person detection is fine-tuned using an extra layer of CNNs(Convolutional Neural Networks) and by training more epochs for a lesser Log Loss and better Accuracy , ROC and IOC.

- **Model Testing:** Both the models are fused together and tested on unseen data footages to evaluate its generalization ability and real-world performance. Deployment: The Model after training , packed and deploy and will be monitored under the infrastructure of Azure ML.
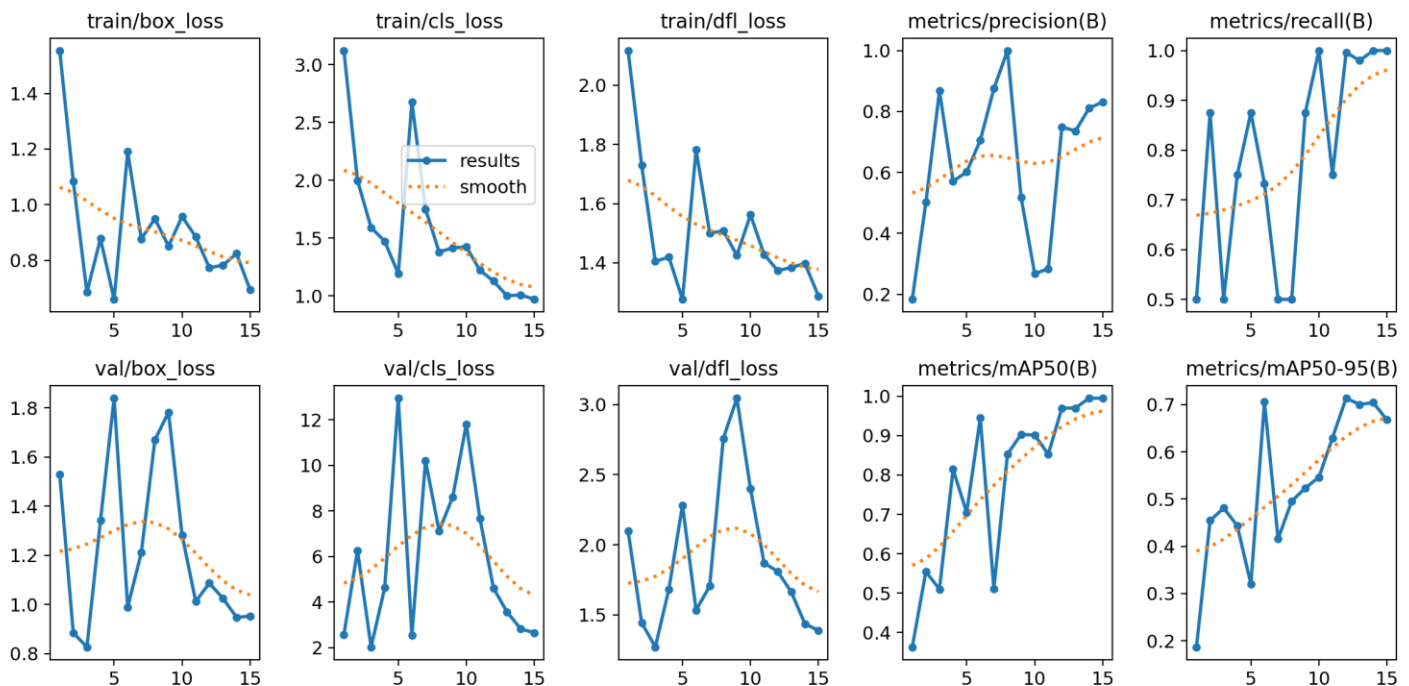
Practical Use Cases:

- **Hospitals:** People talking on their phones in hospital wards and cabins will be detected via the CCTV units and fined accordingly. This will also ensures a peaceful environment for patients and visitors.

- **Pedestrian Crossings:** People talking on their phones while crossing a road will be detected and enhancing people and road safety.

- **Gas Stations:** Using mobile phones in gas stations can be hazardous; hence people talking on their phones at gas stations will be detected via the CCTV units and fined, preserving fire safety.

- **Library:** Surveillance systems will detect phone calls to maintain a quiet environment. Patrons will be reminded to keep noise to a minimum for peaceful studying and reading. This ensures a conducive atmosphere for all library users.

Difficulty Faced:

- **Dataset Collection:** Obtaining a real-time CCTV footages for our model was difficult as this kind of datasets are not available publicly and are protected under the surveillance department.

- **Image Processing:** Standardizing and normalizing the images for training and testing was difficult as it contains a lot of noises and availability was also less as mentioned earlier.

Tech Stacks:

- Matplotlib
- YOLOv8x
- Scikit-Learn
- Pillow
- ResNetV2
- OpenCV
- Streamlit

GitHub Link :

https://github.com/Anidipta/Hack-Fusion-2k24

DEMO VIDEO LINK:

https://youtu.be/qmikzlzULWY?si=6qZTEpnY5xJG7MlN

Conclusion:

In conclusion, our work showcases the potential of image processing in enhancing security and enforcement measures in various settings, such as public places and mobile-free zones. Future research could focus on refining the models further, improving real-time processing capabilities, and exploring additional applications of image processing in security and surveillance.