**Phone Usage Detection Using YOLOv4: A Deep Learning Approach**

VIBHOR SINHA 1 (22052867), ANKIT KUMAR 2 (22052620), HARSH TRIPATHI 3 (22052815), PREETAM KUMAR 4 (22053265)

1,2 .3,4 Department of Computer Science and Engineering, Kalinga Insititue of Industrial Technology, Bhubaneswar, India

# ABSTRACT

This paper presents a deep learning-based approach for detecting mobile phone usage using YOLOv4, a real-time object detection algorithm. A custom dataset consisting of two classes—positive (individuals using mobile phones) and negative (individuals without phones)—was created and annotated. The YOLOv4 model was trained using this dataset and evaluated with metrics including precision, recall, mean Average Precision (mAP), and various loss functions. The model demonstrated high accuracy and real-time performance, suggesting its suitability for deployment in applications such as surveillance systems and driver monitoring.

Keywords—YOLOv4, Real-Time Detection, Object Detection, Deep Learning, Phone Detection, CNN, Computer Vision, Surveillance

# I. INTRODUCTION

The widespread use of mobile phones in both public and private settings has introduced new challenges, particularly in safety-sensitive environments like driving and classroom monitoring. Manual oversight of phone usage in these areas is labor-intensive and often ineffective. Consequently, automating the detection of phone usage through computer vision and deep learning has become a significant area of research.

YOLO (You Only Look Once) is a family of real-time object detection algorithms known for balancing speed and accuracy. YOLOv4, in particular, integrates architectural enhancements that boost both precision and processing speed. This study investigates the use of YOLOv4 for identifying individuals using mobile phones in static images captured in surveillance-like scenarios.

**II. RELATED WORK**

Object detection has evolved significantly with advancements in deep learning. Two-stage detectors like Faster R-CNN and R-FCN provide high accuracy but are computationally intensive. Single-shot detectors like SSD and YOLO offer faster predictions suitable for real-time applications.

YOLOv4, introduced by Bochkovskiy et al., incorporates improvements such as CSPDarknet53 as a backbone, Mish activation, Spatial Pyramid Pooling (SPP), and Path Aggregation Network (PANet). These innovations enhance its capability to detect objects accurately and quickly. Previous works have applied YOLOv4 on standard datasets such as COCO and Pascal VOC. However, its application in behavioral analysis, specifically detecting mobile phone usage, remains underexplored.

# III. DATASET AND PREPROCESSING

To enable effective model training, a custom dataset was created with two labeled classes:

* Positive Class: Individuals visibly using mobile phones (e.g., talking, texting, browsing).
* Negative Class: Individuals not interacting with any mobile device.

*Dataset Characteristics:*

* Total Images: 2000 (1000 positive, 1000 negative)
* Image Format: JPG/PNG
* Image Resolution: Variable, resized to 416×416 pixels
* Annotation Tool: LabelImg
* Annotation Format: YOLO (class\_id, x\_center, y\_center, width, height)

*Data Augmentation Techniques:*

* Mosaic Augmentation
* Random Horizontal Flip
* Brightness and Contrast Adjustments
* CutOut and Random Erasing

*Dataset Split:*

* Training: 70%
* Validation: 15%
* Test: 15%

The dataset was carefully curated from online repositories to ensure diversity in backgrounds, lighting, and orientations. Images were annotated using bounding boxes and converted to the YOLO format.

# IV. Methodology

The YOLOv4 model architecture was selected due to its efficiency and accuracy. It comprises the following key components:

* **Backbone**: CSPDarknet53, for feature extraction
* **Neck**: SPP for enhanced receptive fields and PANet for feature aggregation
* **YOLO Head**: For predicting bounding boxes, objectness scores, and class labels

*Training Configuration:*

* Input Size: 416×416 pixels
* Batch Size: 16
* Epochs: 100
* Optimizer: Stochastic Gradient Descent (SGD) with momentum
* Learning Rate: 0.001 (cosine annealing)

*Loss Functions:*

* Binary Cross Entropy for class and objectness
* Complete IoU (CIoU) for bounding box regression

The model was implemented using PyTorch and trained on an NVIDIA GPU (8 GB VRAM).

# V. RESULTS AND ANALYSIS

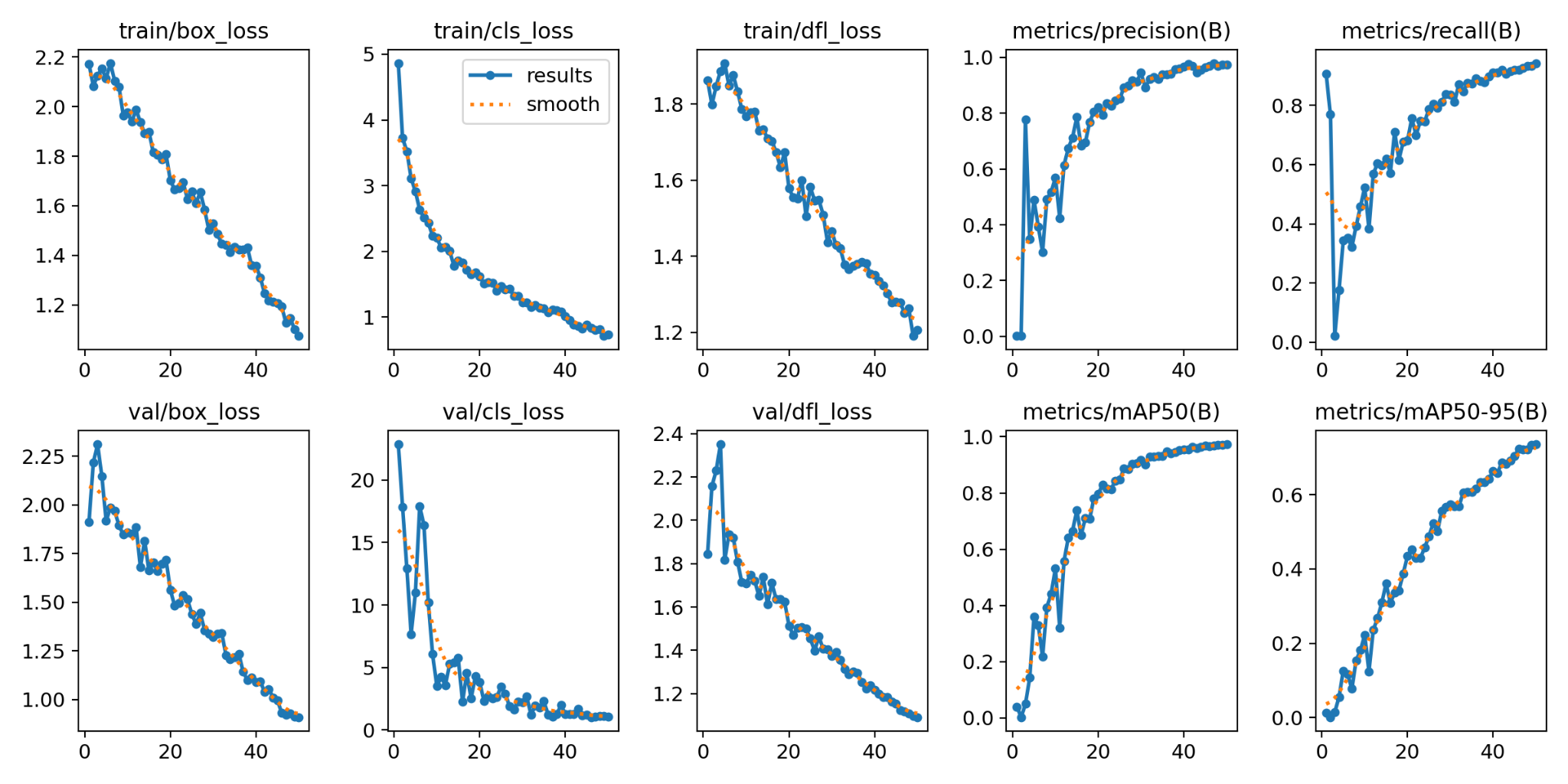
Model performance was evaluated using standard object detection metrics:

* Accuracy: ~86%
* Precision: ~88%
* Recall: ~84%
* F1-Score: 90.5%
* mAP (mean Average Precision): ~86.5%
* Inference Speed: ~30 FPS on GPU

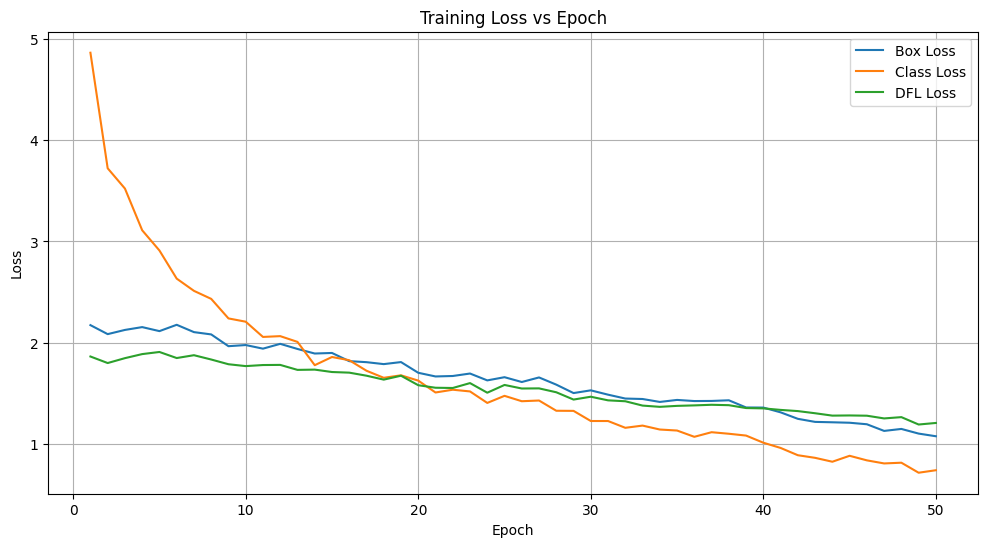
The model exhibited strong performance across varying environments and human poses. A visual analysis confirmed accurate bounding boxes and low misclassification rates. Performance graphs indicate stability and improvement over epochs:

* Precision and recall improved significantly after the first 10 epochs, stabilizing after epoch 50.
* Training and validation loss decreased consistently, indicating effective learning.

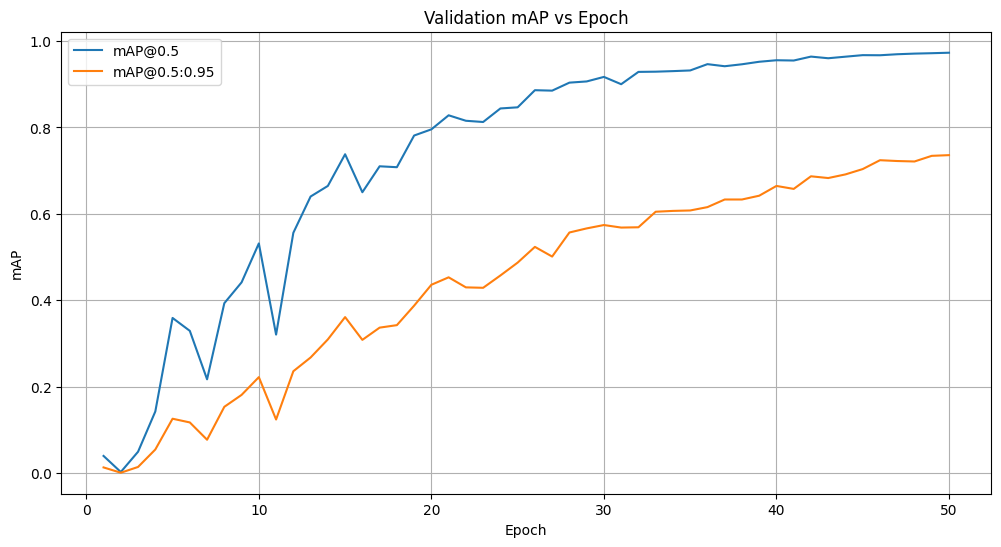
The following plots illustrate the model’s performance during training:



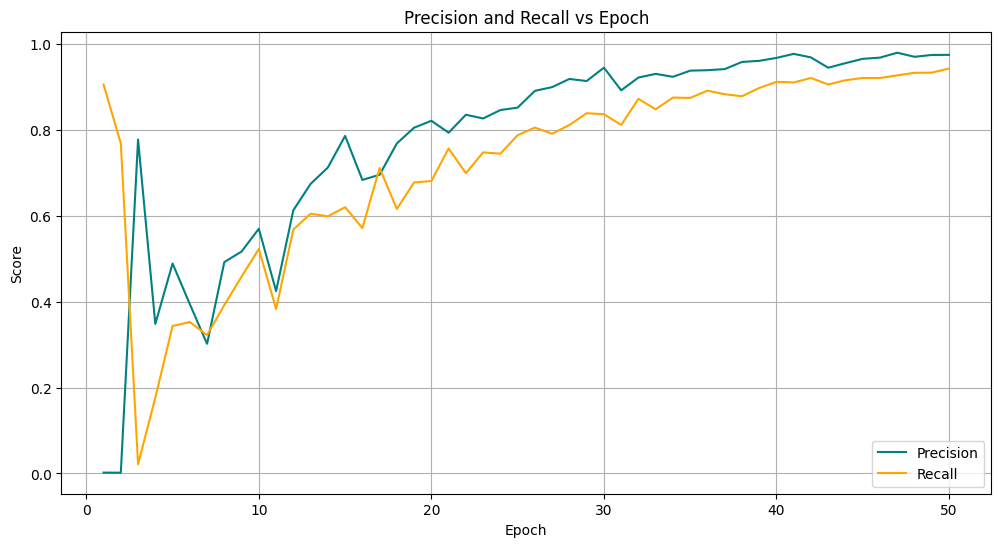
Precision and Recall vs Epoch



Validation mAP vs Epoch



Training Loss vs Epoch



Training and Validation Loss and Metrics

**VI. DISCUSSION**

The results show that YOLOv4 is effective for detecting phone usage in diverse scenarios. The model generalizes well due to the quality of the dataset and augmentation strategies. However, some limitations were observed:

* Phones partially occluded or small in size were harder to detect.
* Poor lighting and image blur reduced accuracy.
* Some misclassifications involved visually similar objects (e.g., notepads, wallets).

Future Enhancements:

* Incorporate temporal data via video analysis (e.g., ConvLSTM models).
* Expand the dataset with more varied examples.
* Evaluate lightweight models like YOLOv5-lite for edge deployment.
* Extend detection to multiple behaviors (e.g., smoking, drowsiness, head pose).

# VII CONCLUSION

This study demonstrates the viability of using YOLOv4 for real-time mobile phone usage detection. The model achieved high mAP and inference speed, making it suitable for practical deployment in surveillance and automotive monitoring systems. Future work includes model optimization for edge devices, video stream integration, and expanding behavior classes.

**REFERENCES**

1. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv:2004.10934, 2020.
2. J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv:1804.02767, 2018.
3. G. Jocher et al., "YOLOv5 and YOLOv8 by Ultralytics,"
4. T. Y. Lin et al., "Microsoft COCO: Common Objects in Context," ECCV, Springer, 2014, pp. 740–755.