

P-YOLOv8: Efficient and Accurate Real-Time Detection of Distracted Driving

Mohamed R. Elshamy*, Heba M. Emara[†], Mohamed R. Shoab[‡], Abdel-Hameed A. Badawy*

*Klipsch School of ECE, New Mexico State University, Las Cruces, NM 88003, United States

[†]Dept. of Electronics and Communications Engineering, Pyramids Higher Inst. for Engineering & Technology, Egypt

[‡]College of Computing and Data Science, Nanyang Technological University, Singapore

*{elshamy, badawy}@nmsu.edu, [†]heba.emara@el-eng.menofia.edu.eg, [‡] Mohamedr003@e.ntu.edu.sg

Abstract—Distracted driving is a critical safety issue that leads to numerous fatalities and injuries worldwide. This study addresses the urgent need for efficient and real-time machine learning models to detect distracted driving behaviors. Leveraging the Pretrained-YOLOv8 (P-YOLOv8) model, a real-time object detection system is introduced, optimized for both speed and accuracy. This approach addresses the computational constraints and latency limitations commonly associated with conventional detection models. The study demonstrates P-YOLOv8’s versatility in both object detection and image classification tasks using the *Distracted Driver Detection* dataset from state farm, which includes 22,424 images across ten behavior categories. Our research explores the application of P-YOLOv8 for image classification, evaluating its performance compared to deep learning models such as VGG16, VGG19, and ResNet. Some traditional models often struggle with low accuracy, while others achieve high accuracy but come with high computational costs and slow detection speeds, making them unsuitable for real-time applications. P-YOLOv8 addresses these issues by achieving competitive accuracy with significant computational cost and efficiency advantages. In particular, P-YOLOv8 generates a lightweight model with a size of only 2.84 MB and a lower number of parameters, totaling 1,451,098, due to its innovative architecture. It achieves a high accuracy of 99.46% with this small model size, opening new directions for deployment on inexpensive and small embedded devices using Tiny Machine Learning (TinyML). The experimental results show robust performance, making P-YOLOv8 a cost-effective solution for real-time deployment. This study provides a detailed analysis of P-YOLOv8’s architecture, training, and performance benchmarks, highlighting its potential for real-time use in detecting distracted driving.

Index Terms—Distracted Driving, Machine Learning, Image Classification, Convolutional Neural Networks, YOLOv8.

I. INTRODUCTION

Distracted driving poses significant risks to road safety, resulting in approximately 3,142 deaths and 424,000 injuries in the United States in 2019, according to the Centers for Disease Control and Prevention (CDC) [5]. This averages nine deaths per day, and 20% of the deceased were pedestrians or cyclists, highlighting the widespread impact beyond vehicle occupants. The urgent need to mitigate these risks emphasizes the importance of effective methods to detect distracted driving behaviors.

Machine learning (ML) can facilitate the automatic identification of driver inattention, crucial to improving road safety

and transforming how insurance companies assess driving behaviors. By monitoring habits through dashboard-mounted cameras, insurers can adjust premiums based on driver attentiveness, offering lower rates to safer drivers. Although numerous studies have introduced efficient algorithms for this problem, most are computationally expensive and produce large model sizes, focusing mainly on accuracy without adequate consideration of detection speed [3], [4], [6], [12], [23], [27]. In contrast, the proposed P-YOLOv8 (You Only Look Once, version 8) algorithm optimizes accuracy and speed, resulting in a more efficient and scalable solution for real-time applications [24].

P-YOLOv8 achieves superior performance through several key innovations. Its streamlined architecture reduces parameters and computational overhead while maintaining accuracy, contrasting starkly with traditional deep learning models such as VGG16, VGG19, and ResNet [19], [26]. The effective use of anchor boxes by P-YOLOv8 and improved bounding box prediction strategies further boost detection precision and speed. Furthermore, real-time performance is achieved through batch normalization and efficient memory usage during inference [26]. The YOLO model, originally designed for object detection, has been adapted for image classification. P-YOLOv8 represents a significant evolution in the YOLO series, incorporating enhancements that improve performance, flexibility, and efficiency, making it suitable for applications requiring real-time processing [13], [15], [20], [24].

This study focuses on the P-YOLOv8 models, specifically the variant yolov8n-cls.pt, optimized for efficient image classification tasks. These models assign a single class label to an entire image, accompanied by a confidence score, which is advantageous for applications where determining the overall class is sufficient without identifying specific objects. Using the State Farm “Distracted Driver Detection” dataset [21], which consists of 22,424 images in ten behavior categories, the aim is to improve detection speed and classification of potentially dangerous activities. The dataset includes various forms of driver distractions, such as texting and talking on the phone. Experimental results demonstrate that P-YOLOv8 achieves competitive accuracy in image classification tasks while offering significant advantages in speed and computational efficiency, making it a viable alternative to traditional deep learning classification models.

The authors would like to thank the following funding agencies: NSF grant 2219680.

The rest of the paper is organized as follows: Section II reviews related work. Section III details the dataset and the proposed algorithm. Section IV presents experimental results and a discussion. Section V concludes the paper.

II. RELATED WORK

Distracted driving detection is crucial for road safety, with various models developed to identify driver distractions, each showing unique strengths and limitations. Hossain *et al.* [9] proposed an automatic driver distraction detection method using deep convolutional neural networks, achieving a maximum accuracy of 99.98% with the MobileNet v2 model on the State Farm dataset. Despite its high accuracy, the relatively large parameter count of the model (3.5 million) may hinder deployment on resource-constrained devices. In comparison, the VGG-16 model has 138.3 million parameters, while the ResNet50 model has 25,6 million parameters.

Sajid *et al.* [18] developed an efficient deep learning framework for distracted driver detection using the State Farm dataset, reporting that the EfficientDet-D3 model achieved a maximum mean average precision (MAP) of 99.16%. However, its complexity poses challenges for deployment on constrained devices, as the model requires extensive training epochs to achieve high accuracy. Bahari *et al.* [3] also reported high accuracy with the ResNet50 model (94%), but its large size and complexity can limit practical deployment. Aljasim *et al.* [2] created an ensemble model combining ResNet50 and VGG16, achieving a maximum accuracy of 92%. However, this model’s complexity and parameter count (ResNet50: 23 million and VGG16: 138 million) could affect deployment on resource-constrained devices.

Masood *et al.* [12] reported that the VGG16 model achieved a maximum accuracy of 99.57%, but its high parameter count (138 million) presents challenges for resource-constrained environments. Fang *et al.* [7] achieved a maximum accuracy of 99.57% using the Vision Transformer (ViT) model with transfer learning, but its complexity poses similar deployment challenges. Subbulakshmi *et al.* [22] presented an ensemble model with a maximum accuracy of 97.5%, but its large parameter sizes (*e.g.*, NasNet-A Large: 88 million, ResNeXt-101: 44 million) complicate deployment.

Li *et al.* [10] developed a framework achieving maximum accuracies of 99.92% (AlexNet), 100% (VGG16) and 99.99% (ResNet18), but all exhibited complexity and substantial parameter counts (*e.g.*, AlexNet: 217 million, VGG16: 491 million). Abbas *et al.* [1] reported that the optNet-50 model achieved a maximum accuracy of 98%, but its parameter count (317.4 million) complicates the deployment. Li *et al.* [11] presented a hybrid convolutional transformer model (MViTCNet) achieving an accuracy of 91.04% with a more manageable parameter count of 1.36 million, offering a favorable balance between accuracy and computational efficiency.

Detecting distracted driving behaviors is critical for road safety, as various models demonstrate high accuracy but often face deployment challenges on resource-constrained devices due to complexity and large parameter sizes. For example,

MobileNet v2 achieves 99.98% precision but has 3.5 million parameters [9]. VGG16 [12] and ResNet50 [3] have 138 million and 25.6 million parameters, respectively. Models such as EfficientDet-D3 [18], ResNet50 [3], and ensemble models [2] also report high accuracy but require significant computational resources. To address this, we propose leveraging P-YOLOv8, which balances accuracy and computational efficiency, making it suitable for real-time applications on devices with limited resources. Our approach aims to enhance real-world applicability by offering a superior trade-off.

III. MATERIAL AND METHOD

A. Dataset Description

To evaluate the proposed method (P-YOLOv8), we utilized the State Farm dataset [21], which contains 22,424 images classified into ten distinct classes. Each image in the dataset is presented as a 640×480 RGB image. The first category represents safe driving, while the remaining nine categories relate to various forms of distracted driving (*e.g.*, texting, talking on the phone) as illustrated in Figure 1. The distribution of the images across these classes is shown in Table I. The dataset was divided into training, validation, and test sets, with 70% of the images allocated for training, 15% for validation, and 15% for testing.

TABLE I
CLASS NAMES AND THEIR CORRESPONDING NUMBERS

Class	Name	Number of Images
c0	Safe driving	2489
c1	Texting - right hand	2267
c2	Talking on the phone - right hand	2317
c3	Texting - left hand	2346
c4	Talking on the phone - left hand	2326
c5	Operating the radio	2312
c6	Drinking a beverage	2325
c7	Reaching behind	2002
c8	Hair and makeup	1911
c9	Talking to passenger	2129

B. The Proposed Algorithm

P-YOLOv8, the latest iteration of the YOLO series by Ultralytics, represents a significant advance in computer vision. Building on the success of its predecessors, YOLOv8 introduces enhancements that improve performance, flexibility, and efficiency. It supports a wide range of vision AI tasks, including object detection, segmentation, pose estimation, tracking, and classification, making it a versatile tool for various applications [13]–[15], [20]. Table II summarizes the YOLOv8 pre-trained classification models. Detection, segmentation, and pose estimation models are trained on the COCO data set, while classification models use the ImageNet dataset [25]. These models, trained in a large number of labeled images, exhibit strong generalization capabilities for image recognition tasks. This study employs the P-YOLOv8 algorithm for classification on a NVIDIA RTX-A4000 GPU.

Table II summarizes the P-YOLOv8 models: YOLOv8n-cls, YOLOv8s-cls, YOLOv8m-cls, YOLOv8l-cls, and YOLOv8x-cls, which vary in size and complexity, affecting performance

TABLE II
YOLOV8 CLASSIFICATION MODELS PERFORMANCE

Model	size (pixels)	acc top1	acc top5	Speed CPU ONNX (ms)	Speed A100 TensorRT (ms)	params (M)	FLOPs (B) at 640
YOLOv8n-cls	224	69.0	88.3	12.9	0.31	2.7	4.3
YOLOv8s-cls	224	73.8	91.7	23.4	0.35	6.4	13.5
YOLOv8m-cls	224	76.8	93.5	85.4	0.62	17.0	42.7
YOLOv8l-cls	224	76.8	93.5	163.0	0.87	37.5	99.7
YOLOv8x-cls	224	79.0	94.6	232.0	1.01	57.4	154.8



Fig. 1. Examples of different driving behaviors captured in the dataset.

and speed. All models use a uniform input size of 224 pixels. The models' top-1 and top-5 accuracies on the ImageNet validation set improve with complexity, ranging from 69.0% and 88.3% for YOLOv8n-cls to 79.0% and 94.6% for YOLOv8x-cls. Inference times vary, with YOLOv8n-cls processing an image in 12.9 ms on a CPU and 0.31 ms on an NVIDIA A100 GPU, while YOLOv8x-cls requires up to 232 ms on a CPU. The complexity of the model, indicated by the number of parameters and FLOPs, ranges from 2.7 million parameters

and 4.3 billion FLOPs for YOLOv8n-cls to 57.4 million parameters and 154.8 billion FLOPs for YOLOv8x-cls. This study focuses on the variant YOLOv8n-cls.pt, optimized for efficient image classification, assigning a single class label with a confidence score, ideal for tasks that require general class determination [16], [17].

P-YOLOv8 is a state-of-the-art real-time object detection system that improves its predecessors with a powerful backbone (CSPDarknet53) [8], a robust neck (PANet), and an efficient prediction head. The backbone extracts features via convolutional layers with batch normalization and activation functions, while the neck aggregates features for the head to predict bounding boxes, objectness scores, and class probabilities, enabling high-speed single-pass processing.

The output of a convolutional layer is given by:

$$F_{l+1} = \sigma(W_l * F_l + b_l) \quad (1)$$

where F_l is the feature map, W_l are the weights, $*$ denotes convolution, b_l is the bias, and σ is the activation function.

Bounding box predictions are represented as:

$$\mathbf{b} = [x, y, w, h] \quad (2)$$

where x and y are the center coordinates, and w and h are the dimensions. Class probabilities are calculated using:

$$\hat{y}_j = \frac{\exp(z_j)}{\sum_{k=1}^K \exp(z_k)} \quad (3)$$

where \hat{y}_j is the predicted probability for class j . P-YOLOv8 can also be adapted for classification by utilizing its backbone and head to produce class probabilities through feature extraction, grouping, flattening, and fully connected layers. Its efficiency and real-time performance stem from a unified architecture and optimized inference, while accuracy is enhanced by advanced feature extraction and data augmentation techniques [22]. The detailed algorithm structure is provided in Algorithm 1. It describes a method for detecting distracted driving using a P-YOLOv8 model M . The algorithm includes data preprocessing (loading, resizing, normalizing, augmenting, and splitting the dataset), model training (loading the pre-trained model, fine-tuning, training, validating, and updating model parameters with backpropagation), and model evaluations. This approach leverages the robust YOLOv8 architecture to ensure the trade-off between accuracy and computational cost.

IV. RESULTS AND DISCUSSION

The results of the P-YOLOv8 models in the provided dataset elucidate the trade-offs between accuracy and computational

Algorithm 1 P-YOLOv8-Based Distracted Driving Detection

Require: Dataset D with images of driving behaviors, pre-trained YOLOv8 model M (yolov8n-cls.pt)

Ensure: Dataset Structure for YOLO Classification Tasks [24]

1: **Data Preprocessing:**

- 2: Load the dataset D
- 3: Resize images to 224×224 pixels
- 4: Normalize pixel values
- 5: Perform data augmentation (rotation, flipping, scaling)
- 6: Split the dataset into training, validation, and test sets

7: **Model Training:**

- 8: Load the pretrained YOLOv8 model M (e.g., yolov8n-cls.pt)
- 9: Fine-tune M on the training set
- 10: **for** each epoch do
- 11: Train the model M on the training set
- 12: Validate the model M on the validation set
- 13: Compute training and validation loss:

$$L = \sum_{i=1}^N y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (4)$$

- 14: Update model parameters using backpropagation:

$$\theta \leftarrow \theta - \eta \nabla_{\theta} L \quad (5)$$

15: **end for**

16: **Model Evaluations**

cost. The key metrics summarizing the performance include an impressive overall accuracy of 99.464% and detailed precision, recall, and F1 score for each class, as shown in Table III. Furthermore, the size of the model is remarkably tiny at 2.84MB, featuring a lower number of parameters, totaling 1,451,098, due to its innovative architecture. The classification metrics for the P-YOLOv8 model, as presented in Table III, demonstrate its high performance in multiple classes. The model achieves nearly perfect precision, recall, and F1 scores for most classes, with precision and recall values consistently close to or at 1.000000 for classes c1, c2, c5, c6, and c7. The average precision, recall, and F1 score are all above 0.99, specifically 0.994239, 0.994648, and 0.994422, respectively. These results highlight the model’s exceptional ability to accurately classify different behaviors, maintaining balanced performance and robust reliability. Additionally, the performance of class c8, while slightly lower than others, still remains high, indicating the overall effectiveness and precision of the model in classification tasks.

The confusion matrix was also used to gain a more granular view of our model performance. The confusion matrix provides a detailed breakdown, indicating the number of correct and incorrect predictions for each class. As shown in the confusion matrix visualization in Figure 2, the high diagonal values indicate that the model accurately predicts the majority of instances in all classes.

To further illustrate the effectiveness of our model, we include a set of predicted images showcasing various driving

TABLE III
CLASSIFICATION METRICS

Class	Precision	Recall	F1 Score
c0	0.991957	0.991957	0.991957
c1	1.000000	1.000000	1.000000
c2	1.000000	1.000000	1.000000
c3	0.997151	0.997151	0.997151
c4	0.997126	0.994269	0.995696
c5	1.000000	1.000000	1.000000
c6	1.000000	0.991453	0.995708
c7	0.996667	1.000000	0.998331
c8	0.972028	0.996416	0.984071
c9	0.987461	0.975232	0.981308
Average	0.994239	0.994648	0.994422

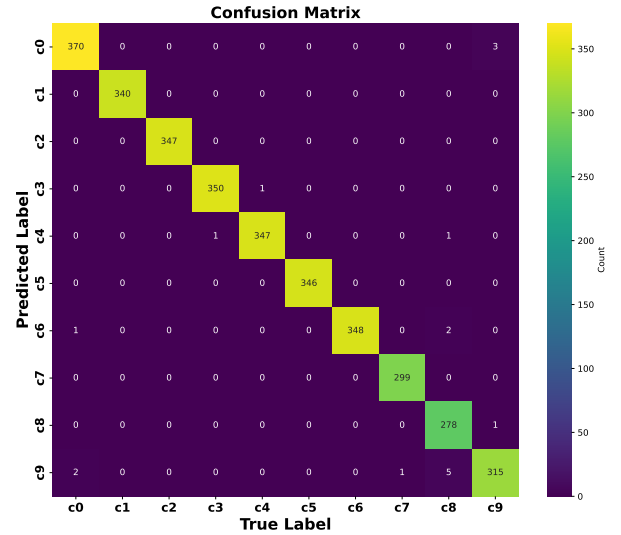


Fig. 2. Confusion Matrix Visualization. The matrix shows the number of correct and incorrect predictions for each class.

behaviors, such as “Normal driving,” “Texting - right” and “Talking on the phone - right,” with predictions that are consistent with the actual behaviors, demonstrating the applicability of the model in real-world scenarios as shown in Figure 3.

Table IV compares various models in the literature review with respect to the accuracy and the number of model parameters. The models cited include those by Hossain *et al.* [9], Li *et al.* [11], Abbas *et al.* [1], Li *et al.* [10], Subbulakshmi *et al.* [22], and Masood *et al.* [12]. These models demonstrate a range of accuracies, from 91.04% to 100%, and model sizes, from 1.36 million to 491 million parameters. The proposed algorithm, highlighted in the table, achieves an impressive accuracy of 99.46% with a significantly smaller model size of 1.45 million parameters. This represents an optimal trade-off between accuracy and model complexity. Although Li *et al.* [10] reports a perfect accuracy of 100%, it comes at the cost of a substantially larger model size of 491 million parameters, making it less practical for real-world applications with limited computational resources. In contrast, the proposed algorithm provides nearly equivalent accuracy with a fraction of the parameters, which improves its feasibility and efficiency. This analysis underscores the superiority of the proposed algorithm, which offers a balanced solution by maintaining high performance while being computationally efficient. Thus,



Fig. 3. Predicted images illustrating the model’s accuracy in identifying driving behaviors.

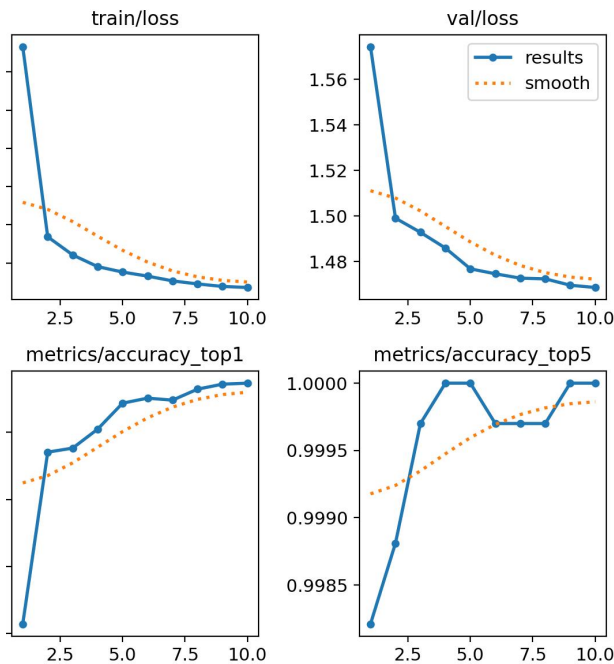


Fig. 4. Training and validation loss curves, along with Top-1 and Top-5 accuracy metrics over 10 epochs.

TABLE IV
MODEL COMPARISON IN TERMS OF ACCURACY AND MODEL SIZE

Model	Accuracy	Model Parameters (millions)
Hossain <i>et al.</i> [9]	99.98%	3.5
Li <i>et al.</i> [11]	91.04%	1.36
Abbas <i>et al.</i> [1]	98%	317.4
Li <i>et al.</i> [10]	100%	491
Subbulakshmi <i>et al.</i> [22]	97.5%	5.3
Masood <i>et al.</i> [12]	99.57%	138
Proposed algorithm	99.46%	1.45

it stands out as an optimal choice for applications where both accuracy and model size are critical considerations.

Our results highlight the proposed classification model accuracy and reliable performance in identifying different driving behaviors. The evaluation, supported by detailed metrics, confusion matrix analysis, and visual examples, confirms the robustness of the model and its potential for use in driver monitoring systems. We also present additional graphs that provide further insights into the model performance metrics and comparisons as depicted in Figure 4. These curves represent the training and validation performance of a YOLOv8 classification model across epochs. The training loss curve shows that the training loss decreases significantly as the number of epochs increases, indicating that the model is learning and fitting the training data better over time. Similarly, the validation loss curve shows a decrease in validation loss over epochs, signifying that the model performance on unseen validation data is improving. This decrease in both training and validation loss is a positive sign that the model is generalizing well to new data. The Top-1 accuracy curve represents the proportion of times that the model top prediction (the one with the highest probability) is correct. This plot shows that Top-1 accuracy improves and stabilizes as training progresses, indicating that the model is becoming more accurate in its primary predictions. The top-5 accuracy curve represents the proportion of times the correct label is within the Top-5 predictions of the model. This plot shows a high Top-5 accuracy, approaching 100%, which means that the model almost always includes the correct label among its top five predictions. The solid blue lines represent the actual values observed during training and validation, while the dotted orange lines represent smoothed versions of these metrics, helping to visualize the overall trends by reducing the noise in the data.

V. CONCLUSION

This study presents a real-time distracted driving detection system utilizing the Pretrained-YOLOv8 (P-YOLOv8) model, addressing both speed and accuracy challenges typically faced by traditional models. The results demonstrate the exceptional performance of P-YOLOv8 in image classification tasks, as evidenced by its application to the State Farm Driver Distraction Detection dataset. The model achieved an impressive accuracy of 99.46% while maintaining a compact parameter size of 2.84MB and 1,451,098 parameters. This demonstrates the model’s computational efficiency and suitability for

deployment on resource-constrained devices, such as those used in Tiny Machine Learning (TinyML). The P-YOLOv8 model, originally designed for object detection, has been adapted for efficient real-time processing in image classification. The P-YOLOv8 proves to be a viable alternative to traditional deep learning models, offering significant speed and computational cost advantages. The detailed analysis of the architecture, training, and performance benchmarks of P-YOLOv8 underscores its potential for practical applications to improve road safety through the detection of distracted driving behaviors. This approach, leveraging advances in the YOLO series, provides a balanced solution that maintains high accuracy while significantly reducing computational demands.

ACKNOWLEDGMENT

The authors express their deepest gratitude to Prof. Francesco Fabiano for his input. Authors Mohamed R. Elshamy and Abdel-Hameed A. Badawy were partially supported by NSF grant 2219680.

REFERENCES

- [1] Tahir Abbas, Syed Farooq Ali, Aadil Zia Khan, and Irfan Kareem. optnet-50: An optimized residual neural network architecture of deep learning for driver's distraction. In *2020 IEEE 23rd International Multitopic Conference (INMIC)*, pages 1–5, 2020.
- [2] Mustafa Aljasim and Rasha Kashef. E2dr: A deep learning ensemble-based driver distraction detection with recommendations model. *Sensors*, 22(5), 2022.
- [3] Munif Alotaibi and Bandar Alotaibi. Distracted driver classification using deep learning. *Signal, Image and Video Processing*, 14:617 – 624, 2019.
- [4] Muhammad Saiful Haqem Saiful Bahari and Lucyantie Mazalan. Distracted driver detection using deep learning. In *2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA)*, pages 198–203, 2022.
- [5] Centers for Disease Control and Prevention. Distracted Driving. <https://www.cdc.gov/distracted-driving/about/index.html>, 2020. Accessed: 2024-07-09.
- [6] Annu Dhiman, Anukrity Varshney, Faeza Hasani, and Bindu Verma. A comparative study on distracted driver detection using cnn and ml algorithms. In Mukesh Saraswat, Chandreyee Chowdhury, Chintan Kumar Mandal, and Amir H. Gandomi, editors, *Proceedings of International Conference on Data Science and Applications*, pages 663–676, Singapore, 2023. Springer Nature Singapore.
- [7] Zhenwu Fang, Jinxin Chen, Jinxiang Wang, Zejiang Wang, Neng Liu, and Guodong Yin. Driver distraction behavior detection using a vision transformer model based on transfer learning strategy. In *2022 6th CAA International Conference on Vehicular Control and Intelligence (CVCI)*, pages 1–6, 2022.
- [8] Chaima Gouider and Hassene Seddik. Yolov4 enhancement with efficient channel recalibration approach in cspdarknet53. In *2022 IEEE Information Technologies & Smart Industrial Systems (ITSIS)*, pages 1–6, 2022.
- [9] Md. Uzzol Hossain, Md. Ataur Rahman, Md. Manowarul Islam, Arnisha Akhter, Md. Ashraf Uddin, and Bikash Kumar Paul. Automatic driver distraction detection using deep convolutional neural networks. *Intelligent Systems with Applications*, 14:200075, 2022.
- [10] Guofa Li, Qing Liu, and Zizheng Guo. Driver distraction detection using advanced deep learning technologies based on images. *IEEE Journal of Radio Frequency Identification*, 6:825–831, 2022.
- [11] Penghua Li, Qiyun Mou, Jie Hou, and Yushan Tu. Hybrid convolutional-transformer neural network for driver distraction detection. In *2023 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*, pages 1–6, 2023.
- [12] Sarfaraz Masood, Abhinav Rai, Aakash Aggarwal, M.N. Doja, and Musheer Ahmad. Detecting distraction of drivers using convolutional neural network. *Pattern Recognition Letters*, 139:79–85, 2020.
- [13] U.M.M.P.K. Nawarathne, H.M.L.S. Kumari, and H.M.N.S. Kumari. Comparative analysis of jellyfish classification: A study using yolov8 and pre-trained models. In *2024 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, volume 7, pages 1–6, 2024.
- [14] Supriya Pulparambil, Basel Bani-Ismael, Hazem Migdady, Youcef Baghdadi, Sara Al-Ghafri, and Abdulaziz Al-Sakhbouri. Electronic waste recognition and classification using yolov8 for ict sector. In *2023 24th International Arab Conference on Information Technology (ACIT)*, pages 1–6, 2023.
- [15] Luyi-Da Quach, Khang Nguyen Quoc, Anh Nguyen Quynh, Hoang Tran Ngoc, and Nguyen Thai-Nghe. Tomato health monitoring system: Tomato classification, detection, and counting system based on yolov8 model with explainable mobilenet models using grad-cam++. *IEEE Access*, 12:9719–9737, 2024.
- [16] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016.
- [17] Joseph Redmon and Ali Farhadi. Yolo9000: Better, faster, stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6517–6525, 2017.
- [18] Faiqa Sajid, Abdul Rehman Javed, Asma Basharat, Natalia Kryvinska, Adil Afzal, and Muhammad Rizwan. An efficient deep learning framework for distracted driver detection. *IEEE Access*, 9:169270–169280, 2021.
- [19] Mohamed R. Shoaib, Mohamed R. Elshamy, Taha E. Taha, Adel S. El-Fishawy, and Fathi E. Abd El-Samie. Efficient brain tumor detection based on deep learning models. *Journal of Physics: Conference Series*, 2128(1):012012, dec 2021.
- [20] Somantri, Alun Sujjada, Yusuf Saeful Bayan, Viky Rodiatul Ulum, Fahmi Adrial Ilhami, and Gina Purnama Insany. Enhanced plastic detection and classification: Advancing recognition of plastic varieties using yolov8. In *2023 IEEE 9th International Conference on Computing, Engineering and Design (ICCED)*, pages 1–6, 2023.
- [21] State Farm. State Farm Distracted Drivers Dataset. <https://www.kaggle.com/c/state-farm-distracted-driver-detection/data>, 2017. Accessed: 2017-06-15.
- [22] S. Subbulakshmi, A.K Arathy, and Praseetha Madhu. Ensemble model for driver distraction detection. In *2023 4th IEEE Global Conference for Advancement in Technology (GCAT)*, pages 1–7, 2023.
- [23] Ikromjanov Kobiljon Komil Ugli, Ali Hussain, Beom Su Kim, Satyabrata Aich, and Hee-Cheol Kim. A transfer learning approach for identification of distracted driving. In *2022 24th International Conference on Advanced Communication Technology (ICACT)*, pages 420–423, 2022.
- [24] Ultralytics. Ultralytics documentation - classification datasets. <https://docs.ultralytics.com/datasets/classify/>, 2023. Accessed: 2024-07-09.
- [25] Ultralytics. Imagenet dataset configuration. <https://github.com/ultralytics/ultralytics/blob/main/ultralytics/cfg/datasets/ImageNet.yaml>, 2024. Accessed: 2024-07-09.
- [26] D. Umanandhini, M. Shyamala Devi, N Beulah Jabaseeli, and S. Sridevi. Batch normalization based convolutional block yolov3 real time object detection of moving images with backdrop adjustment. In *2023 9th International Conference on Smart Computing and Communications (ICSCC)*, pages 25–29, 2023.
- [27] Naveen Kumar Vaegae, Kranthi Kumar Pulluri, Kalapaveen Bagadi, and Olutayo O Oyerinde. Design of an efficient distracted driver detection system: Deep learning approaches. *IEEE Access*, 10:116087–116097, 2022.