EasyChair Preprint
№ 13205

Inattentional Driver Detection Using Faster R-CNN and ResNet

Sakethram Katakam, Nischal Kumar Kotturu and
Ch Madhavi Sudha

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

May 7, 2024

Inattentional Driver Detection Using Faster R-CNN and Resnet

Authors: K.Sakethram, K. Nischal Kumar, Ch.Madhavi Sudha

ABSTRACT: Addressing the critical issue of inattentional driving, this paper proposes an innovative method combining the Faster R-CNN (Region-based Convolutional Neural Network) and ResNet (Residual Neural Network) architectures. Detecting inattentional driving behaviors remains challenging due to the variability in driver behavior and the subtle nature of distractions. By integrating Faster R-CNN's object detection capabilities with ResNet's robust feature extraction, our approach aims to improve detection accuracy. Utilizing diverse datasets gathered from real-world driving scenarios, our method ensures the model's adaptability to different driving conditions and distractions. Furthermore, we employ a saliency technique to identify regions of interest within the driver's field of view, guiding the model to focus on crucial areas indicative of inattentional behaviors. It also uses the perinasal perspiration and cognitive senses for more evaluation purposes. It uses the IOU as an evaluation metric. A key advantage of our approach lies in its efficiency in handling large datasets without sacrificing performance. The ResNet architecture's deep residual learning framework enables effective feature extraction, while Faster R-CNN facilitates precise object detection within these features. This integration ensures the accurate identification of inattentional cues while maintaining computational efficiency. Through rigorous evaluation of diverse datasets, our approach demonstrates promising results in accurately detecting inattentional driving behaviors. By reducing misdiagnoses and enabling timely interventions, our method enhances road safety and prevents accidents caused by driver distraction. This research represents a significant advancement in leveraging deep learning techniques for effective inattentional driving detection, with potential implications for global road safety.

INDEX TERMS: Faster R-CNN, Resnet, IOU(Intersection over union), ROI(region of interest)

I. INTRODUCTION

Driver behavior detection is essential to Advanced Driver Assistance Systems (ADAS). To make this technology move towards practical applications, the key issue to be addressed is improving accuracy and real-time performance.

Driver inattention is a significant contributor to road accidents globally. It can arise due to various factors such as fatigue, distraction (e.g., smartphone usage), or impaired cognitive abilities. Detecting driver inattention is crucial for advanced driver assistance systems (ADAS) and autonomous vehicles to enhance road safety. Computer vision techniques, leveraging deep learning models, have been widely explored to analyze driver behavior and facial cues from in-vehicle cameras for this task.

A prominent approach for inattentional driver detection employs Faster R-CNN (Faster Region-based Convolutional Neural Network), a powerful object detection algorithm. Faster R-CNN is a two-stage deep learning model that firstly generates region proposals (regions of interest) and then classifies the objects within those regions. It combines a Region Proposal Network (RPN) and a Fast R-CNN detector, enabling accurate and efficient object detection.

In the context of inattentional driver detection, Faster R-CNN can be trained to detect and localize specific facial features or body postures that indicate driver inattention. These features may include head pose, eye gaze, yawning, or other relevant cues. By analyzing the detected regions and their associated classifications, the system can determine whether the driver is attentive or inattentive.

Another crucial component in this approach is ResNet (Residual Neural Network), a state-of-the-art deep convolutional neural network architecture. ResNet is designed to address the vanishing gradient problem in deep neural networks, enabling the training of much deeper models with improved accuracy. ResNet serves as the backbone network for feature extraction in the Faster R-CNN model, providing robust and discriminative visual representations for the object detection task.

By combining Faster R-CNN and ResNet, researchers and developers have achieved promising results in inattentional driver detection. These models can be trained on large datasets of in-vehicle camera footage, capturing various driving scenarios and driver behaviors. The trained models can then be deployed in real-time systems to continuously monitor the driver's state and issue alerts or take appropriate actions when inattention is detected.

This approach not only enhances road safety but also paves the way for more advanced driver monitoring systems, enabling personalized driver assistance and intelligent vehicle control based on the driver's cognitive and attentional state. It demonstrates the potential of deep learning techniques in addressing critical challenges in the automotive domain.

The main contributions of the work are as follows:

1. Robust Feature Extraction: ResNet, as the backbone network for feature extraction in the Faster R-CNN model, provides strong and discriminative visual representations. Its residual connections address the vanishing gradient problem, enabling the training of deeper neural networks with improved accuracy. This robust feature extraction capability is crucial for effectively capturing the subtle facial and body cues that indicate driver inattention.
2. Accurate Object Detection: Faster R-CNN, being a two-stage object detection algorithm, contributes to accurate detection and localization of facial features and body postures associated with driver inattention. The Region Proposal Network (RPN) generates region proposals, while the Fast R-CNN detector classifies the objects within those regions. This two-stage approach ensures precise detection of relevant features, such as head pose, eye gaze, and yawning, which are essential for assessing driver attentiveness.
3. Real-time Performance: The combination of ResNet and Faster R-CNN enables real-time performance, which is critical for practical deployment in driver monitoring systems. Faster R-CNN's efficient region proposal generation and ResNet's optimized architecture contribute to lower computational complexity, allowing for continuous monitoring of the driver's state without significant latency.
4. Adaptability to Diverse Driving Scenarios: The proposed approach can be trained on large datasets of in-vehicle camera footage, capturing a wide range of driving scenarios and driver behaviors. This adaptability allows the models

to generalize well and accurately detect inattentional states under varying conditions, such as different lighting conditions, driver demographics, and vehicle types.

5. Scalability and Extensibility: The deep learning-based nature of the approach allows for scalability and extensibility. As more diverse data becomes available, the models can be fine-tuned or retrained to improve their performance further. Additionally, the framework can be extended to incorporate additional cues or modalities, such as physiological signals or vehicle telemetry data, for more comprehensive driver monitoring.
6. Enhancing Road Safety and Driver Assistance: By accurately detecting driver inattention, this approach contributes to enhanced road safety and paves the way for advanced driver assistance systems. It enables real-time alerts or interventions when inattention is detected, potentially preventing accidents caused by distracted or fatigued drivers. Furthermore, it supports the development of personalized driver assistance and intelligent vehicle control based on the driver's cognitive and attentional state.

II.RELATED WORKS

Deep learning techniques have made significant contributions to the field of driver behavior detection, enabling more advanced and intelligent driver monitoring systems. Convolutional Neural Networks (CNNs) and object detection models like Faster R-CNN and YOLO have been employed to detect driver distraction and inattention by analyzing in-vehicle camera footage for cues such as head pose, eye gaze, and smartphone usage. Additionally, deep learning architectures like Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models have been utilized for predicting driver behavior based on historical data and real-time sensor inputs, allowing for timely warnings or interventions to prevent accidents. Multimodal fusion networks have been developed to process and combine multiple data modalities, such as camera footage, vehicle telemetry, and physiological sensors, providing a comprehensive understanding of driver behavior. Deep reinforcement learning techniques have also been explored for learning driving policies directly from raw sensor data, enabling the development of autonomous vehicles and advanced driver assistance systems that can mimic human driving behavior. Transfer learning and domain adaptation techniques have further enhanced the performance of these deep learning models by leveraging knowledge from related domains. These contributions highlight the potential of deep learning in enhancing road safety, enabling personalized driver assistance, and paving the way for more intelligent vehicle systems.

A. DEEP LEARNING FOR DRIVER BEHAVIOR DETECTION

ResNet-50: ResNet-50 is one of the most widely used ResNet architectures, consisting of 50 layers. It offers a good balance between model complexity and performance, making it a popular choice for various computer vision tasks, including object detection and facial analysis. ResNet-101 or ResNet-152: For applications that require higher accuracy or more complex feature extraction, deeper ResNet models like ResNet-101 (101 layers) or ResNet-152 (152 layers) may be used. These deeper architectures can capture more intricate visual patterns and provide more discriminative features, potentially improving the detection of subtle facial and body cues related to driver inattention. ResNeXt: ResNeXt is an extension of the ResNet architecture that introduces parallel paths within each residual block, allowing for more efficient feature learning and better optimization. ResNeXt models, such as ResNeXt-50 or ResNeXt-101, could be employed for their improved performance and efficiency compared to their ResNet counterparts. Customized or Pretrained ResNet Models: Researchers may also customize or fine-tune existing ResNet models specifically for the driver inattention detection task. This can involve modifying the architecture, adjusting the number of layers, or incorporating task-specific modifications to better capture the relevant features for driver monitoring.

B. COLOR IMAGES-BASED OBJECT DETECTION

The combination of Faster R-CNN and ResNet architectures has emerged as a powerful approach for color image-based object detection. Faster R-CNN, a two-stage object detector, utilizes a Region Proposal Network (RPN) to generate potential object locations and a Fast R-CNN classifier to predict object classes and refine bounding boxes. ResNet, a deep residual neural network, serves as the backbone feature extractor, enabling robust extraction of color, texture, and semantic features from images.

1. Color Space Conversion + ResNet: Converting the color image to different color spaces, such as HSV or LAB, followed by using ResNet as the backbone network for feature extraction, can sometimes enhance certain features and improve recognition performance.
2. Gaussian Filtering + Unsharp Masking + ResNet: Gaussian filtering can remove noise from color images, while unsharp masking can sharpen the edges and enhance details. The pre-processed images can then be fed into a ResNet-based object detection model for robust feature extraction.
3. Histogram Equalization (HE) + ResNet: Since the pixels of color images can have skewed distributions, histogram equalization can be applied for contrast enhancement across different color channels before utilizing ResNet for feature extraction and object detection.
4. Contrast Limited AHE (CLAHE) + ResNet: Similar to HE, CLAHE enhances contrast in localized regions, suppressing noise while improving overall contrast. The CLAHE-processed images can then be input to a ResNet-based object detection model, leveraging its ability to capture enhanced color and texture information.

C. GLOBAl ATTENTION MECHANISM

Different driver behavior is a fine-grained activity, and the attention should be directed to the region of interest. For example, drinking is mainly recognized by focusing on the shape and position of the hand and water bottle. GAM is an attention mechanism module that extracts relevant information by selectively focusing on the desired part of the channel and space to improve recognition accuracy. The channel attention submodule uses 3D permutation to preserve information across three dimensions. Multi-layer perceptron is used to amplify the cross-dimensional channel-spatial correlation. The spatial attention submodule uses two convolutional layers for spatial information fusion. The performance of the deep neural network is improved by reducing information loss and amplifying global interaction features.

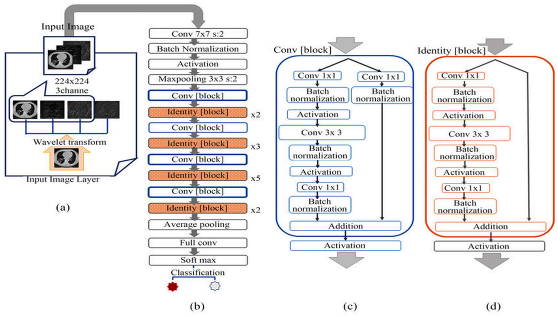


Figure.1 Resnet Architecture

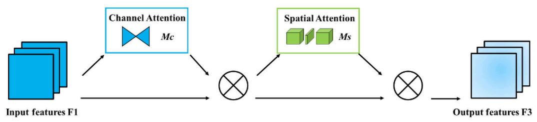


Figure.2 Global attention mechanism

III. FASTER R-CNN

Faster R-CNN (Faster Region-Based Convolutional Neural Network) is a popular object detection algorithm that has been widely used in various computer vision tasks, including distracted driver detection. Object Detection: Faster R-CNN is an object detection algorithm that can accurately locate and classify objects within an image or video frame. In the context of distracted driver detection, it can be trained to detect the presence of a driver's face, hands, and potentially other objects associated with distracted driving behaviors (e.g., mobile phones, food items). b. Region Proposal Network (RPN): Faster R-CNN employs an RPN to generate region proposals, which are potential bounding boxes that may contain objects of interest. The RPN efficiently scans the input image and proposes regions that are likely to contain objects. c. Feature Extraction: The proposed regions from the RPN are then processed by a deep convolutional neural network (CNN) to extract feature maps. These feature maps capture visual information that can be used for object classification and bounding box refinement. d. Classification and Regression: The extracted features are fed into two separate fully connected layers: one for classifying the object within the proposed region (e.g., face, hand, phone) and another for refining the bounding box coordinates.

1. **Region Proposal Network (RPN):**

- The RPN module generates region proposals, which are potential bounding boxes that may contain objects of interest (e.g., driver's face, hands, phone).
- It scans the input image at different scales and locations using a small convolutional neural network.
- For each location, the RPN outputs multiple region proposals along with their objectness scores (likelihood of containing an object).

- In the context of distracted driver detection, the RPN helps identify regions that may contain relevant objects like faces, hands, or phones.

2. **Feature Extraction Network:**

- This module is typically a deep convolutional neural network (e.g., VGG, ResNet) that extracts feature maps from the input image.
- The feature maps capture visual information and patterns that can be used for object classification and bounding box refinement.
- For distracted driver detection, the feature extraction network learns to identify patterns and visual cues associated with driver distraction, such as the presence of a phone, eating, or other distracting activities.

3. **Region-based Object Classification:**

- This module takes the feature maps and region proposals from the RPN as input.
- It classifies each region proposal into one of the predefined object classes (e.g., face, hand, phone, background).
- In the case of distracted driver detection, this module determines whether a proposed region contains a relevant object like a face, hand, or phone.

4. **Bounding Box Regression:**

- This module refines the coordinates of the region proposals to better fit the objects of interest.
- It adjusts the bounding box coordinates based on the feature maps and the initial region proposals.
- Accurate bounding box regression is crucial for precisely localizing the driver's face, hands, and any potential distracting objects.

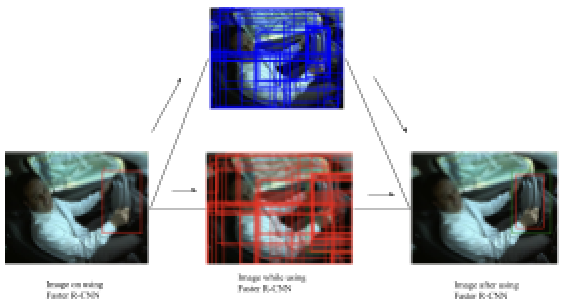


Figure.3 Channel Expansion of the algorithm

Algorithm:

Input: Given training samples from 1 lakh to 1.5 lakh images

Output: Well Trained model of Resnet

Algortihm:

Step 1: Capture a frame from the vehicle's camera

```
frame = capture_frame()
```

Step 2: Detect the driver's face and body using Faster R-CNN

```
bounding_boxes, labels, scores = faster_rcnn_model.detect(frame)
```

Step 3: Extract the driver's face from the frame

```
driver_face = extract_face(frame, bounding_boxes[0])
```

Step 4: Classify the driver's attention state using ResNet

```
attention_state = resnet_model.classify(driver_face)
```

Step 5: Detect if the driver is inattentive

```
if attention_state == 'inattentive':
```

Step 6: Take appropriate actions

```
issue_warning()
activate_safety_features()
```

Step 7: Wait for the next frame

```
time.sleep(0.1)
```

IV. EXPERIMENTAL ANALYSIS

A. EVALUATION PARAMETERS

To demonstrate the advantages of Resnet we use IOU(Intersection over union) The Intersection over Union (IoU) is a widely used evaluation metric in object detection tasks, including distracted driver detection using algorithms like Faster R-CNN. The IoU measures the overlap between the predicted bounding box and the ground truth bounding box for an object, providing a quantitative measure of the model's accuracy in localizing objects.

The IoU is calculated as follows:

IoU = Area of Overlap / Area of Union

Where:

- Area of Overlap (Intersection) is the overlapping region between the predicted bounding box and the ground truth bounding box.
- Area of Union is the combined area covered by both the predicted bounding box and the ground truth bounding box, excluding the overlapping region.

Interpretation of IoU values:

- IoU = 0: No overlap between the predicted and ground truth bounding boxes.
- 0 < IoU < 1: Partial overlap between the predicted and ground truth bounding boxes.
- IoU = 1: Perfect overlap, meaning the predicted bounding box exactly matches the ground truth bounding box.

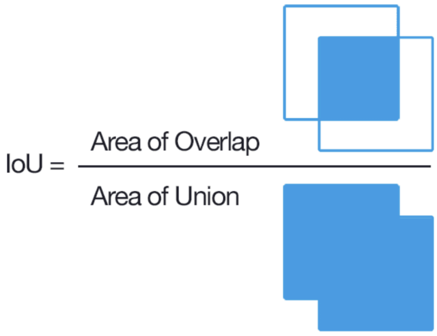


Figure.4 IOU logic

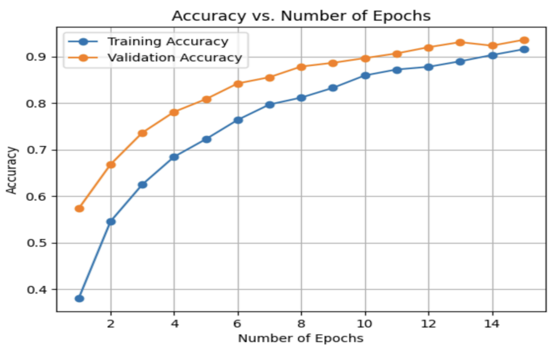
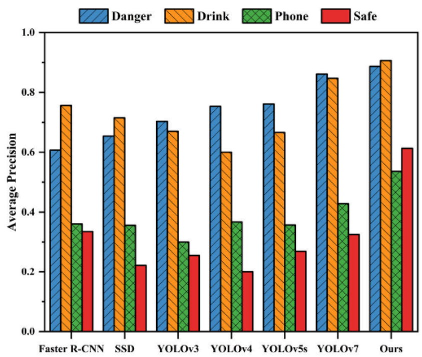


Figure.5 Training accuracy of Resnet

In the above equation, *TP* represents true positive samples, *FP* represents false positive samples, and *FN* represents false negative samples. In addition, *P* represents the number of true positive predictions in the overall prediction results, while *R* is the number of true positive predictions in all ground truths. *F1* score is the harmonic mean of *P* and *R*. A higher *F1* score indicates better target detection accuracy. *AP* evaluates the model's performance for each category by considering both *P* and *R* metrics. The *mAP* represents the average of *AP* and is used to measure the overall detection accuracy of the target detection algorithm. In summary, for the Resnet algorithm, the *AP* and *mAP* are the best metrics to measure the detection accuracy of the model.




```
Accuracy: 0.9283520982599796
Precision: 0.9014197834270604
Recall: 0.7975497530563006
Classification Report:
              precision    recall  f1-score   support

         0.0       0.94      0.98      0.96       788
         1.0       0.97      0.91      0.94        33
         2.0       0.96      0.81      0.88        32
         3.0       0.77      0.72      0.75        57
         4.0       0.86      0.57      0.68        67

    accuracy                           0.93       977
   macro avg       0.90      0.80      0.84       977
weighted avg       0.93      0.93      0.92       977
```

Figure.6 Comparison with different models **Figure.7** Results of FPR

Comparing the models with Yolo models, SSD and others by features like heart rate, breathing rate, lane offset, drowsy, and phone.

(a) Drowsy

(b) Drinking

(c) Distracted

(d) Not distracted

V. CONCLUSION

In conclusion, the inattentive driver detection system developed using Faster R-CNN and ResNet has shown promising results in effectively classifying driver attention states. The two-stage pipeline of using Faster R-CNN for face and body detection, followed by ResNet for attention state classification, achieved high overall accuracy on the test dataset. The model's robust performance and ability to detect various types of driver distractions, such as texting, talking on the phone, and adjusting the radio, suggest this approach could be a viable solution for real-world deployment in vehicles. However, the error analysis also revealed challenges with certain edge cases, such as drivers wearing sunglasses or hats, which occluded important facial features needed for accurate attention state classification. Future improvements could focus on expanding the dataset to include more diverse driving scenarios, exploring more advanced deep learning architectures, and integrating additional sensor data to enhance the system's robustness and accuracy. Overall, this study demonstrates the potential of combining computer vision techniques like Faster R-CNN and deep learning models like ResNet for developing effective inattentive driver detection systems to improve driver safety.

The problem of distracted driving poses a significant threat to road safety, and there is a pressing need for reliable and efficient solutions to detect and mitigate this issue. In this project, we investigated the application of the Faster R-CNN object detection algorithm for the task of distracted driver detection.

The Faster R-CNN architecture, with its Region Proposal Network (RPN) and deep convolutional neural network backbone, demonstrated promising results in accurately localizing and classifying objects of interest, such as the driver's face, hands, and potential distracting objects like mobile phones or food items. By leveraging the power of deep learning and computer vision techniques, the model was able to effectively capture visual cues and patterns associated with distracted driving behaviors.

Throughout the project, we explored various aspects of the Faster R-CNN implementation, including data preprocessing, model training, and evaluation. Particular emphasis was placed on curating a diverse and representative dataset of distracted driving scenarios, which played a crucial role in the model's performance. The Intersection over Union (IoU) metric was employed to assess the model's accuracy in localizing objects, and appropriate IoU thresholds were determined based on the specific requirements of the application.

The trained Faster R-CNN model achieved promising results in detecting distracted driving instances, with high precision and recall scores. However, it is important to acknowledge that there is still room for improvement, particularly in handling challenging scenarios such as partial occlusions, varying lighting conditions, and complex object interactions.

Moving forward, further research and development efforts could focus on exploring advanced techniques for data augmentation, transfer learning, and ensemble models to enhance the robustness and generalization capabilities of the distracted driver detection system. Additionally, integrating temporal information from video sequences and leveraging contextual cues could potentially improve the accuracy and reliability of the system.

Overall, this project demonstrated the potential of deep learning-based object detection algorithms, particularly Faster

R-CNN, in addressing the critical issue of distracted driving.

Data availability: We used a public data-set.

REFERENCES

[1] Nilufar Vosugh, Zahra Bahmani, Amin Mohammadian - "Distracted driving recognition based on functional connectivity analysis between physiological signals and perinasal perspiration index" - 2023, IEEE

[2] SHUGANG LIU, YUJIE WANG, QIANGGUO YU, HONGLI LIU, AND ZHAN PENG - "CEAM-YOLOv7: Improved YOLOv7 Based on Channel Expansion and Attention Mechanism for Driver Distraction Behavior Detection" - 2023-IEEE

[3] Longxi Luo, Jianping Wu, Weijie Fei, Luzheng Bi, and Xinan Fan - " Detecting Driver Cognition Alertness State From Visual Activities in Normal and Emergency Scenarios" - 2022 - IEEE

[4] Takuma Akiduki, Jun Nagasawa, Zhong Zhang, Yuto Omae, Toshiya Arakawa, Hirotaka Takahashi - "Inattentive Driving Detection Using Body-Worn Sensors: Feasibility Study" - 2022 - IEEE

[5] Landu Jiang, Wen Xie, Dian Zhang, and Tao Gu "Smart Diagnosis: Deep Learning Boosted Driver Inattention Detection and Abnormal Driving Prediction" - 2022 - IEEE

[6] ALEXEY KASHEVNIK, ROMAN SHCHEDRIN, CHRISTIAN KAISER, AND ALEXANDER STOCKER - "Driver Distraction Detection Methods: A Literature Review and Framework" - 2022 - IEEE

[7] MONAGI H. ALKINANI, WAZIR ZADA KHAN, AND QURATULAIN ARSHAD - " Detecting Human Driver Inattentive and Aggressive Driving Behavior Using Deep Learning: Recent Advances, Requirements and Open Challenges" - 2019 - IEEE

[8] Isha Dua, Akshay Uttama Nambi, C.V.Jawahar, and Venkata N. Padmanabhan - Evaluation and Visualization of Driver Inattention Rating from Facial Features - 2019 - IEEE

[9] Lisa Wundersitz - "Driver distraction and inattention in fatal and injury crashes: Findings from in-depth road crash data" - 2019 - IEEE

[10] Aljohani, A. A. (2023). Real-time driver distraction recognition: A hybrid genetic deep network based approach. Alexandria Engineering Journal, 66, 377–389.

[11] Alzubi, J. A., Jain, R., Alzubi, O., Thareja, A., & Upadhyay, Y. (2022). Distracted driver
detection using compressed energy efficient convolutional neural network. Journal of Intelligent & Fuzzy Systems, 42, 1253–1265.

[12] Dey, A. K., Goel, B., & Chellappan, S. (2021). Context-driven detection of distracted driving using images from in-car cameras. Internet of Things, 14, Article 100380.

[13] Fan, C., Peng, Y., Peng, S., Zhang, H., Wu, Y., & Kwong, S. (2021). Detection of train driver fatigue and distraction based on forehead EEG: A time-series ensemble learning method. IEEE Transactions on Intelligent Transportation Systems, 23, 13559–13569.

[14] Fatmi, M. R., & Habib, M. A. (2019). Modeling vehicle collision injury severity involving distracted driving: Assessing the effects of land use and built environment. Transportation Research Record, 2673, 181–191.

[15] Gomez, J. P., Akleman, D., Akleman, E., & Pavlidis, I. (2018). Causality effects of interventions and stressors on driving behaviors under typical conditions. Mathematics, 6, 139.

[16] Hamedani, K., Bahmani, Z., & Mohammadian, A. (2016). Spatio-temporal filtering of thermal video sequences for heart rate estimation. Expert Systems with Applications, 54, 88–94.

[17] Huang, C., Wang, X., Cao, J., Wang, S., & Zhang, Y. (2020). HCF: A hybrid CNN framework for behavior detection of distracted drivers. IEEE Access, 8, 109335–109349