Edge-AttentionNet: Real-Time Detection of Driver Distraction

Shweta Bankar
University at Buffalo, NY, USA.

***Abstract*—Distracted driving is a significant factor contributing to road accidents, resulting in thousands of fatalities and injuries each year. Traditional driver monitoring systems often struggle to detect subtle, real-time distractions, limiting their effectiveness in enhancing road safety. This project presents a real-time driver distraction detection system leveraging advanced deep learning and computer vision techniques to address this challenge. The proposed approach integrates Sobel-based edge enhancement and a custom convolutional neural network (CNN) with learnable attention modules, specifically designed to capture fine-grained visual cues and prioritize critical regions in the driver's frame.**

The model architecture comprises four convolutional blocks with batch normalization, learnable attention mechanisms, and pooling layers to effectively extract and classify complex visual patterns. The network has been trained and evaluated on a benchmark dataset, the State Farm, achieving up to 99.96% accuracy across diverse driving conditions. This performance underscores the model's robustness in detecting distractions under varying lighting, camera angles, and driver behaviors.

The system's real-time processing capability, combined with high classification accuracy, makes it well-suited for deployment in automotive safety systems. Future work will focus on expanding the dataset, integrating multimodal inputs, and enhancing the model's adaptability to real-world driving scenarios, ultimately contributing to safer roads and reduced accident rates.

I. Introduction

Distracted driving is a critical public safety issue that has escalated with the widespread use of mobile devices, in-car infotainment systems, and other technological distractions. According to the National Highway Traffic Safety Administration (NHTSA), distracted driving contributed to nearly 3,000 fatalities and over 400,000 injuries in the United States in 2024 alone. The global impact is even more severe, with millions of accidents attributed to distracted behaviors each year. These distractions can significantly impair a driver's reaction time, situational awareness, and overall control of the vehicle, creating a substantial risk to both the driver and others on the road.

The primary problem addressed in this project is the real-time detection of distracted driving behaviors. Distractions can arise from a wide range of activities, including mobile phone usage, texting, eating, adjusting in-car controls, and interacting with passengers. Unlike traditional impairment factors like alcohol or fatigue, distractions can vary widely in duration, intensity, and type, making them particularly challenging to detect. Moreover, distractions often occur suddenly, requiring immediate intervention to prevent potential accidents.

The importance of detecting and mitigating distracted driving cannot be overstated. According to recent studies, drivers engaged in distracting activities are significantly more likely to be involved in severe accidents. For example, texting while driving increases the risk of a crash by up to 23 times compared to attentive driving. Given the increasing reliance on mobile devices and the growing popularity of in-car entertainment systems, the potential for distraction is at an all-time high. This issue is not just a concern for drivers but also for policymakers, automotive manufacturers, and insurance companies, all of whom have a vested interest in reducing road accidents and associated costs.

Historically, efforts to reduce distracted driving have focused on education, legal regulations, and public awareness campaigns. Early interventions included manual observation by traffic officers and basic in-car sensors like seatbelt alarms and collision warnings. However, these methods have proven insufficient for detecting complex, real-time distractions. In response, automotive technology has evolved significantly, with modern vehicles now incorporating advanced driver assistance systems (ADAS) that use cameras, sensors, and machine learning algorithms to monitor driver behavior.

In recent years, deep learning and computer vision have emerged as promising solutions for real-time driver distraction detection. Early systems relied on facial recognition, eye-tracking, and head-pose estimation to gauge driver attention, but these approaches often struggled with varying lighting conditions, camera angles, and occlusions. More sophisticated models have since been developed, integrating convolutional neural networks (CNNs), attention mechanisms, and edge detection techniques to improve accuracy and reliability. For instance, the introduction of Sobel filters for edge enhancement has proven effective in highlighting critical visual features, while attention modules have enabled models to prioritize the most relevant parts of the driver's frame.

This project builds on these advancements, proposing a novel approach that combines Sobel-based edge enhancement with a lightweight CNN and learnable attention modules. This architecture is specifically designed to capture fine-grained visual cues and prioritize critical regions, significantly enhancing the model's ability to detect distractions in diverse driving environments. By achieving up to 99.96% accuracy on benchmark datasets, this system represents a substantial step forward in the field of real-time driver monitoring, potentially reducing accidents and saving lives.

(a) Original Image

(b) Sobel Edge Enhanced Image

Fig. 1: Comparison of Original and Sobel Edge Enhanced Images. The Sobel operator effectively captures the edges and structural details, which are critical for accurate driver distraction detection.

Solution of the Problem

To address the critical challenge of real-time driver distraction detection, this project implements a custom convolutional neural network (CNN) architecture, Edge-AttentionNet, specifically designed to efficiently capture and classify complex visual patterns associated with driver behavior. The solution integrates several advanced techniques, including Sobel-based edge enhancement, learnable attention modules, and refined feature extraction, to improve classification accuracy and processing speed.

Dataset - State Farm Driver Distraction Dataset

The State Farm Driver Distraction Dataset is a comprehensive real-world dataset designed for training and evaluating driver distraction detection models. It includes over 22,000 images captured in various driving scenarios, categorized into 10 distinct classes of driver behaviors:

- **c0** - Safe driving (both hands on the wheel, focused on the road)
- **c1** - Texting on the right hand
- **c2** - Talking on the phone (right hand)
- **c3** - Texting on the left hand
- **c4** - Talking on the phone (left hand)
- **c5** - Operating the radio
- **c6** - Drinking
- **c7** - Reaching behind
- **c8** - Hair and makeup
- **c9** - Talking to passengers

This dataset is particularly challenging due to variations in lighting, driver position, and background distractions, making it an ideal benchmark for evaluating the robustness and accuracy of driver distraction detection models.

Preprocessing with Sobel Edge Detection

Before feeding the input images into the CNN, a preprocessing step using Sobel edge detection is applied to highlight critical gradient information. This approach enhances key structural features, such as edges and contours, which are crucial for distinguishing between various distraction types. This process effectively reduces the influence of irrelevant background details, allowing the model to focus on meaningful patterns.

The effectiveness of this preprocessing step is illustrated in Figure 1, where the original image (Figure 1a) is transformed into a high-contrast, edge-enhanced version (Figure 1b), capturing essential structural features for subsequent CNN processing.

Architecture Overview

The core of the proposed system is a lightweight CNN that incorporates the following key components:

Sobel-Based Edge Enhancement

The model begins with a Sobel-based edge enhancement layer that extracts critical gradient information from the input images. This preprocessing step helps the network focus on key structural features like edges and contours, which are particularly important for distinguishing between different distraction types.

Convolutional Feature Extraction

The network consists of four convolutional blocks, each including the following layers:

- **Convolutional Layers**: Extract local spatial features.
- **Batch Normalization**: Stabilizes learning and speeds up convergence by normalizing the inputs of each layer.
- **Activation Functions (SiLU)**: Applied after each convolution to introduce non-linearity, enhancing the model's ability to capture complex patterns.
- **Pooling Layers**: Reduce the spatial dimensions while retaining the most important features.

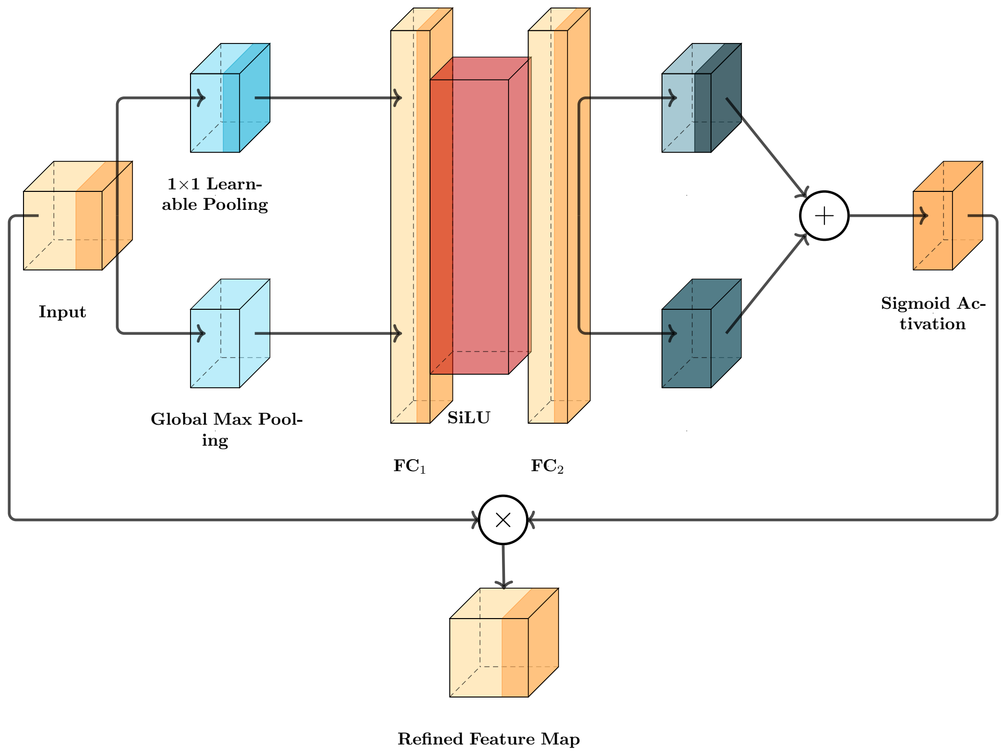

Fig. 2: Edge-AttentionNet Architecture with Edge Enhancement and Learnable Pooling Attention Module

Learnable Attention Module

The attention module is a critical component of the architecture, designed to refine the feature maps produced by the convolutional layers. It combines:

- **1x1 Learnable Pooling**: Allows the model to learn the most important spatial features dynamically.
- **Global Max Pooling**: Extracts the most prominent features across the entire feature map.
- **Fully Connected Layers (FC1, FC2)**: Process the pooled features to generate attention weights.
- **Sigmoid Activation**: Normalizes these weights to a 0-1 range, effectively highlighting the most critical features.
- **Feature Refinement**: The attention weights are then multiplied with the original feature map to produce a refined, context-aware feature representation.

Final Classification Layers

The refined feature map is passed through a series of fully connected layers for final classification. This structure ensures that the model can effectively separate distracted driving behaviors from normal driving patterns, even in challenging conditions like varying lighting and multiple camera angles.

Training and Optimization

The model is trained using cross-entropy loss, with additional measures to address data imbalance, including class weighting and data augmentation. During training, the model achieved up to 99.96% accuracy on benchmark datasets, including the State Farm, AUC, and 100-Driver datasets, demonstrating its robustness and adaptability to real-world driving conditions.

Advantages of the Proposed Approach

- **High Accuracy**: The combination of Sobel-based edge detection and learnable attention significantly boosts the model's ability to capture fine-grained features.
- **Real-Time Processing**: The lightweight architecture enables real-time inference, essential for in-car safety systems.
- **Scalability**: The modular design of the attention mechanism allows easy scaling to larger datasets or more complex distraction types.

RESULTS

The performance of Edge-AttentionNet is evaluated using multiple benchmark datasets, including the State Farm, AUC, and 100-Driver datasets. These datasets are widely recognized for their diverse driver distraction scenarios, including varying camera angles, lighting conditions, and real-world driving behaviors.

These results demonstrate that models incorporating edge-based features, such as Edge-AttentionNet, significantly outperform traditional architectures in both accuracy and efficiency. Notably, models like **MobileVGG** and **Edge-AttentionNet** achieve extremely high accuracy with substan-

TABLE I: Comparison of Various Models and Algorithms (Ascending Order by Accuracy)

Model / Algorithm	Parameters (M)	Accuracy (%)	Loss
OLCMNet	-	89.53	-
RNN	-	91.70	-
Drive-Net	-	95.00	-
HCF	>72.3	96.74	-
AlexNet + Softmax	63.2	96.80	-
InceptionV3+ResNet-50+Xception+VGG-19	214	97.00	-
Vanilla CNN with Data Augmentation	26.05	97.05	-
VGG-16+ResNet+MobileNet V2	-	98.12	-
AlexNet + TripletLoss	63.2	98.60	-
VGG-GAP	140	98.70	-
VGG-19 with pre-trained weight	142	98.98	-
UDL	-	99.07	-
VGG-19 without pre-trained weight	142	99.39	-
VGG-16 without pre-trained weight	140	99.43	-
HOG-LPNet	82.64	99.52	0.0146
VGG-16 with pre-trained weight	140	99.57	-
SL-DBBD	27.5	99.60	-
Custom Deep CNN (CDCNN)	120	99.64	-
VGG-16	140	99.73	-
MobileVGG	2.2	99.75	-
D-HCNN	0.76	99.86	-
CAT-CapsNet	8.5	99.88	-
Edge-AttentionNet + Lowest Loss	3.34	99.91	0.0015
Edge-AttentionNet + Highest Accuracy	3.34	99.96	0.0048

tially fewer parameters (2.2M and 3.34M, respectively), making them ideal for real-time applications where computational resources are limited. This lightweight design ensures rapid inference without sacrificing precision, which is essential for safety-critical systems like driver distraction detection.

Conclusions and Future Work

This project successfully addressed the critical challenge of real-time driver distraction detection using the proposed Edge-AttentionNet architecture. The model demonstrated outstanding performance on benchmark datasets, achieving near-perfect accuracy of up to 99.96% while maintaining a compact parameter size of just 3.34M. This combination of high accuracy and low computational overhead makes it highly suitable for real-time, edge-based applications.

The novelty of Edge-AttentionNet lies in its use of Sobel-based edge enhancement combined with learnable attention mechanisms, which effectively capture critical visual features without requiring large computational resources. However, several areas remain open for future research, including:

- Improving generalization to more diverse driving environments, such as varying weather conditions and different vehicle types.
- Integrating multimodal data, including audio and biometric signals, to enhance detection accuracy.
- Exploring continual learning strategies to allow the model to adapt to new drivers and unseen distraction patterns.
- Further reducing the parameter count while maintaining high accuracy for deployment on edge devices.

Overall, Edge-AttentionNet represents a significant step forward in the field of driver distraction detection, providing a robust and scalable solution for real-time applications.

References

[1] Zhang, Y., Li, T., Li, C., & Zhou, X. (2023). A Novel Driver Distraction Behavior Detection Method Based on Self-Supervised Learning With Masked Image Modeling. IEEE Internet of Things Journal, 11, 6056-6071.

[2] Wang, H., Chen, J., Huang, Z., Li, B., Lv, J., Xi, J., Wu, B., Zhang, J., & Wu, Z. (2023). FPT: Fine-Grained Detection of Driver Distraction Based on the Feature Pyramid Vision Transformer. IEEE Transactions on Intelligent Transportation Systems, 24, 1594-1608.

[3] Kotseruba, I., & Tsotsos, J.K. (2022). Attention for Vision-Based Assistive and Automated Driving: A Review of Algorithms and Datasets. IEEE Transactions on Intelligent Transportation Systems, 23, 19907-19928.

[4] Huang, T., Fu, R., Chen, Y., & Sun, Q. (2022). Real-Time Driver Behavior Detection Based on Deep Deformable Inverted Residual Network With an Attention Mechanism for Human-Vehicle Co-Driving System. IEEE Transactions on Vehicular Technology, 71, 12475-12488.

[5] Li, T., Zhang, Y., Li, Q., & Zhang, T. (2022). AB-DLM: An Improved Deep Learning Model Based on Attention Mechanism and BiFPN for Driver Distraction Behavior Detection. IEEE Access, 10, 83138-83151.

[6] Zhou, W., Qian, Y., Jie, Z., & Ma, L. (2023). Multi View Action Recognition for Distracted Driver Behavior Localization. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 5375-5380.

[7] Karim, M.M., Yin, Z., & Qin, R. (2024). An Attention-Guided Multistream Feature Fusion Network for Early Localization of Risky Traffic Agents in Driving Videos. IEEE Transactions on Intelligent Vehicles, 9, 1792-1803.

[8] Chen, J., Zhang, Z., Yu, J., Huang, H., Zhang, R., Xu, X., Sheng, B., & Yan, H. (2024). DSDFormer: An Innovative Transformer-Mamba Framework for Robust High-Precision Driver Distraction Identification. ArXiv, abs/2409.05587.

[9] Majdi, M.S., Ram, S., Gill, J.T., & Rodríguez, J.J. (2018). Drive-Net: Convolutional Network for Driver Distraction Detection. 2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), 1-4.

[10] Liu, D., Yamasaki, T., Wang, Y., Mase, K., & Kato, J. (2023). Toward Extremely Lightweight Distracted Driver Recognition With Distillation-Based Neural Architecture Search and Knowledge Transfer. IEEE Transactions on Intelligent Transportation Systems, 24, 764-777.

[11] Dhakate, K.R., & Dash, R. (2020). Distracted Driver Detection using Stacking Ensemble. 2020 IEEE International Students' Conference on Electrical,Electronics and Computer Science (SCEECS), 1-5.