DriveSAM: Cognitive Perspective on Driving Maneuvers Based on Drivers' Attention Using Eye Gaze Data

Kelvin Kwakye and Younho Seong
Department of Industrial & Systems Engineering
North Carolina A&T State University
Greensboro, North Carolina, USA
kkkwakye@aggies.ncat.edu, yseong@ncat.edu

Sun Yi
Department of Mechanical Engineering
North Carolina A&T State University
Greensboro, North Carolina, USA
syi@ncat.edu

Armstrong Aboah
Department of Civil and Architectural Engineering and Mechanics
University of Arizona-Tucson
aaboah@arizona.edu

Abstract

Accurately predicting the intentions of drivers during their maneuvers and assessing their level of attentiveness hold paramount importance in comprehending driver behavior and enhancing decision-making processes. Given the intricate nature of driving environments, drivers are contended with a plethora of distractions that divert their focus away from the road. Thus, it is imperative to fathom the extent of a driver's attentiveness and its consequential impact on their cognitive decision-making abilities in the context of making driving maneuvers. Despite its pivotal significance, research in the following fundamental questions remains conspicuously absent: What garners the primary visual focus of drivers whilst operating a vehicle? Why do drivers occasionally veer into other vehicles or obstacles? The current study therefore alleviates the gaps in previous studies by providing answers to these questions. The overarching goal of this study is to employ a zero-shot learning technique to understand the complex patterns and relationships between driver's attention and their cognitive decision-making. To achieve this goal, the study leveraged on the synergistic fusion of both image features from the driving scene and eye gaze information from drivers. The data used from this study was gotten from DR(eye)VE dataset. We employ the Segment Anything Model (SAM) for our zero-shot learning to segment the attention of the drivers base on their gaze information. The result from our study indicates that drivers pay more attention to road when driving on a highway as compared to city drive. This insight provides valuable information about drivers' attention allocation in different driving contexts and can aid in developing tailored interventions to improve road safety and enhance driver awareness.

Keywords
Drivers' Attention, Eye-tracking, Human-vehicle Interaction, Segment Anything Model and Safety

1. Introduction
Driving is a complex cognitive task that requires the seamless integration of perceptual, motor, and decision-making processes (Lemonnier et al. 2014). The ability to understand how drivers allocate their attention during various driving maneuvers is crucial for improving road safety and optimizing driver assistance systems. In recent years, advancements in Artificial Intelligence (AI) and eye-tracking technology have provided researchers with valuable insights into drivers' visual attention patterns and image analysis (Gyimah et al. 2022; Gyimah et al. 2023). By

analyzing eye gaze data, we can gain a cognitive perspective on driving maneuvers, exploring how drivers process information from their visual environment to execute different driving tasks effectively.

In the realm of transportation safety and human-vehicle interaction, accurately predicting drivers' intentions and assessing their attentiveness play a pivotal role in enhancing road safety and optimizing decision-making processes. Driving, as a complex cognitive task, demands constant vigilance and prudent decision-making to navigate through diverse and dynamic environments (Büla et al. 2011). However, the driving experience is often challenged by a myriad of distractions that can divert drivers' focus away from the road, potentially leading to hazardous situations. Understanding the extent of drivers' attentiveness and its consequential impact on their cognitive decision-making abilities during driving maneuvers is of paramount importance in comprehending driver behavior and implementing effective safety measures (Kotseruba and Tsotsos 2021).

Despite the significance of driver attentiveness and decision-making, several fundamental questions remain unanswered. For instance, what specific elements capture drivers' primary visual focus while operating a vehicle, and why do drivers occasionally deviate into other vehicles or obstacles despite their driving experience? These questions underscore the need for comprehensive research in the domain of driver behavior and attentiveness during driving maneuvers.

In response to these research gaps, the current study presents a novel approach to address these critical questions by employing a zero-shot learning technique. Zero-shot learning leverages the synergy between image features from the driving scene and eye gaze information from drivers to gain a deeper understanding of the complex patterns and relationships between drivers' attention and their cognitive decision-making abilities.

The study utilizes data from the DR(eye)VE dataset, an extensive and valuable resource that captures real-world driving scenarios with precise eye gaze annotations. By leveraging the Segment Anything Model (SAM), the study successfully performs zero-shot learning to accurately segment drivers' attention based on their gaze information.

The overarching goal of this study is to contribute to a comprehensive understanding of driver behavior and attentiveness during maneuvers, shedding light on the factors influencing drivers' visual focus and decision-making. By illuminating drivers' attention allocation across different driving contexts, the study aims to provide valuable insights that can be harnessed to develop targeted interventions and optimize road safety measures.

In the following sections, we present a brief literature review, methodology and results of our study, highlighting the implications of our findings for driver safety and human-vehicle interaction. By bridging the gap in previous research and providing answers to these fundamental questions, this study contributes to the broader field of transportation safety and driver behavior analysis. The results hold promising applications in the design of advanced driver assistance systems and tailored interventions to enhance drivers' attentiveness and decision-making abilities on the road.

1.1 Objectives

The goal of this study is to employ a zero-shot learning technique to understand the complex patterns and relationships between driver's attention and their cognitive decision-making.

2. Literature Review

Drivers' primary visual focus while operating a vehicle is influenced by various factors such as the driving task and external stimuli. They need to attend to multiple concurrent tasks, including keeping the vehicle in the lane, observing other road users, reacting to hazards, and dealing with distractions. Research has shown that drivers tend to focus on the front area, but their attention can shift to different areas depending on the driving situation and cognitive workload (Nico et al, 2021). Occasionally, drivers may veer into other vehicles or obstacles due to factors like inattention, distraction, or cognitive overload. Failure to notice and respond to surrounding objects and events can lead to accidents (Amadori et al. 2021). Understanding drivers' visual attention patterns can help identify unsafe attention patterns and improve driver monitoring, driver assistance, and automated driving systems (Kotseruba and Tsotsos 2021).

This literature review explores relevant studies that utilize drivers' eye gaze data to investigate driving maneuvers from a cognitive perspective, incorporating theories such as selective attention, visual search, cognitive load, ecological perception, attentional tunneling, attentional capture, and central processing.

1. Selective Attention and Visual Search in Driving Maneuvers: Drivers' selective attention and visual search patterns during lane changing maneuvers were examined using eye gaze data. The study found that drivers primarily directed their gaze to the rear-view mirrors and adjacent lanes, while avoiding prolonged fixation on irrelevant areas. This supports the selective attention theory, suggesting that drivers allocate attention to critical information relevant to the maneuver (Sudkamp et al 2021).

2. Cognitive Load and Driving Performance: Bitkina et al. (2021) investigated the impact of cognitive load on drivers' attention allocation during merging maneuvers using eye gaze data. They found that under high cognitive load conditions, drivers displayed reduced gaze scanning patterns, focusing more on central areas of the road. The results align with cognitive load theory, indicating that drivers' attentional resources were constrained during demanding driving tasks. In a different study, Firoz and Seong (2023) delved into the realm of intuitive cognition during decision-making, utilizing the Artificial Grammar Learning paradigm. Meanwhile, Firoz et al. (2022) ventured into the classification of trust through EEG signals, employing machine learning techniques in a neurological context.
3. Ecological Perception and Decision-Making: Kaya et al. (2021) employed eye gaze data to study drivers' visual perception and decision-making during left turn maneuvers at intersections. The study revealed that drivers fixated on potential hazards, such as oncoming vehicles and pedestrians, before initiating the turn. This finding supports ecological perception theory, emphasizing the role of visual information in real-world decision-making scenarios.
4. Attentional Capture in Complex Driving Scenarios: Chan (2016) investigated attentional capture in driving maneuvers involving unexpected events using eye gaze data. They found that drivers' gaze was involuntarily captured by salient objects on the roadside, leading to momentary distractions. This supports attentional capture theory, demonstrating the influence of external stimuli on drivers' attention allocation.
5. Central Processing and Decision Points: Sun et al. (2016) analyzed eye gaze data to examine drivers' central processing activities during critical decision points in roundabout maneuvers. The study revealed specific gaze patterns associated with decision-making and route planning. This aligns with central processing theory, emphasizing the role of eye gaze in decision-related cognitive activities.

The utilization of eye gaze data renders significant revelations into the cognitive mechanisms that underscore driving maneuvers. The present literature review showcases how theories concerning selective attention, visual search, cognitive load, ecological perception, attentional tunneling, attentional capture, and central processing substantiate a more profound comprehension of drivers' allocation of attention and decision-making processes during driving maneuvers.

3. Methods

The architecture for capturing drivers' attention DriveSAM is designed to seamlessly integrate drivers' eye gaze data, and Segment Anything Model (deep learning, and attention segmentation).

3.1 Data Collection

The DR(eye)VE dataset comprises a total of 555,000 frames, organized into 74 sequences, with each sequence spanning a duration of 5 minutes. The driving experiment involved eight diverse drivers, aged between 20 and 40, consisting of 7 male participants and 1 female participant. The experiment extended over a period exceeding two months. The recorded videos capture various contexts, encompassing distinct landscapes such as downtown areas, countryside settings, and highways. The dataset encompasses a wide spectrum of traffic conditions, ranging from traffic-free scenarios to densely congested situations. Moreover, the videos were captured under diverse weather conditions, including sunny, rainy, and cloudy weather. The recording sessions were conducted throughout different times of the day, encompassing both daytime and nighttime periods (Palazzi et al. 2018). Table 1 presents some characteristics of the dataset. The data in question illuminates the extent of diversity inherent in the dataset as it pertains to the present study.

Table 1. Dataset characteristics

Gaze information	**Time of day**	**Weather conditions**
raw fixation	day	rainy
pupil dilation	evening	cloudy
gaze map	night	sunny

3.2 Segment Anything Model (SAM)

SAM functions based on the principle of zero-shot learning, which is a method that allows the model to detect and segment objects without the requirement of predetermined segmentation labels. Unlike conventional object detection models that necessitate substantial training data for specific objects, SAM is capable of generalizing its segmentation abilities to a broad array of objects, making it an ideal choice for intricate and ever-changing driving scenarios (Kirillov et al. 2023).

SAM's architecture comprises three integral components as depicted in Figure 1. These three components include an image encoder, a flexible prompt encoder, and a fast mask decoder.

Image Encoder. Motivated by considerations of scalability and the deployment of powerful pre-training methods, which utilizes a pre-trained Vision Transformer (ViT) that has been minimally adapted to process high-resolution inputs via mean absolute error (MAE) techniques. The image encoder operates through a single run per image and may be effectively applied prior to prompting the model.

Prompt Encoder. Consider two distinct sets of prompts, namely sparse and dense. The sparse comprises points, boxes, and text, while the dense encompasses masks. The positional encodings of points and boxes are augmented with learned embeddings specific to each type of prompt. Free-form text, on the other hand, is utilized in conjunction with a pre-existing text encoder sourced from CLIP. In the case of dense prompts, i.e., masks, the embedding process involves the use of convolutions, which are subsequently element-wise summed with the image embedding.

Mask Decoder. Effectively maps the image embedding, prompt embeddings, and an output token to a mask through the implementation of a Transformer decoder block that has been modified and is subsequently followed by a dynamic mask prediction head. This decoder block employs prompt self-attention and cross-attention in two directions, namely prompt-to-image embedding and vice versa, for the purpose of updating all embeddings.

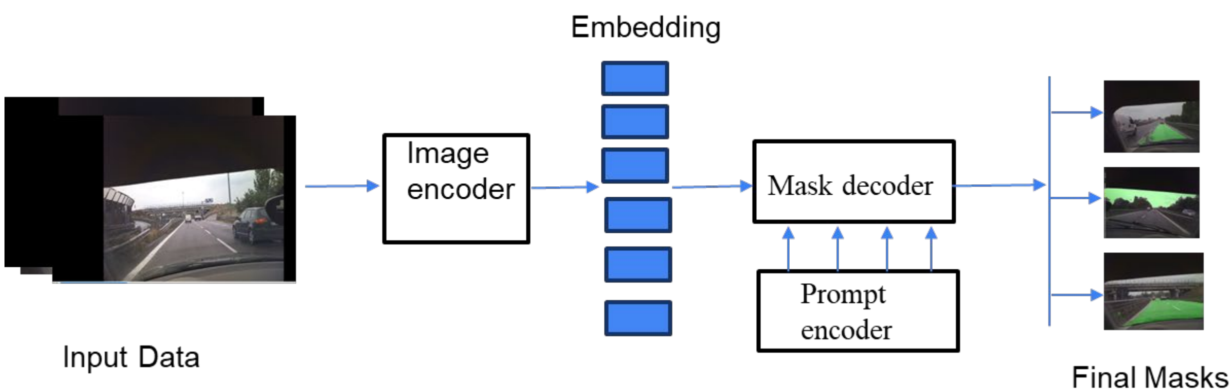

Figure 1. The architecture of DriveSAM system

3.3 Zero-Shot Learning and Attention Segmentation

SAM employs the technique of zero-shot learning (ZSL)to segment attention regions predicated on the input eye gaze data. Semantic embeddings are utilized in ZSL to capture the shared characteristics and relationships among different attention regions. These embeddings are learned from a semantic space that encodes attributes, features, or descriptions of the attention targets (Pourpanah et al. 2022). With attention segmentation, SAM identifies regions where drivers' gaze is concentrated, indicative of attention location. That is, it exhibits the capacity to independently identify regions of interest that encapsulate the focus of drivers' attention, thereby engendering attention heatmaps and density maps that visually depict the patterns of attention allocation.

4. Results and Discussion

Based on the data extracted from the DR(eye)VE dataset and the application of the Segment Anything Model (SAM) for zero-shot learning, a number of noteworthy observations were made. The study noted a distinct variance in the attention span of drivers when comparing highway drives to city drives. On highways as shown in Figures 2a and b, drivers predominantly allocated their attention to the road ahead, which could be attributed to the higher speeds, necessity to maintain a safe following distance, and the presence of fewer distractions than in city environments.

Moreover, by analyzing the eye gaze information in tandem with the driving scene, it was evident that drivers' gaze patterns are intrinsically linked to their cognitive decision-making processes. For instance, instances where drivers veered into other lanes or towards obstacles often coincided with their gaze being diverted away from the primary roadway.
In addition, the combination of image features from the driving scene and eye gaze information provided a holistic view of the driver's behavior. This comprehensive approach helped in unveiling the underlying patterns associated with driving maneuvers and their respective attention allocation. As such, the study underscores the importance of a synergistic fusion of data in understanding the complex interplay between drivers' attention, decision-making processes, and driving behavior.

Figure 2(a). Highway drive

Figure 2(b). Highway drive

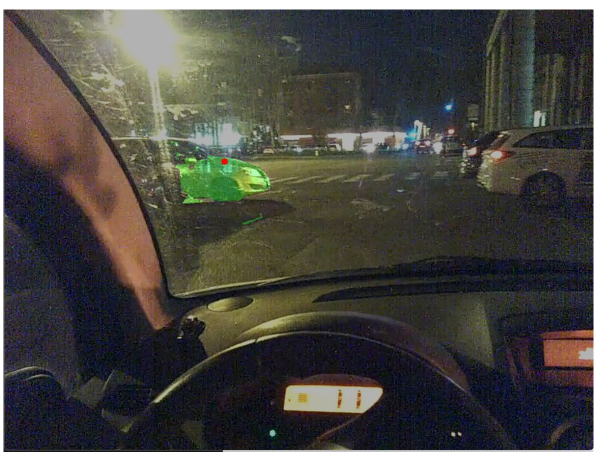

Figure 3(a). City drive

Figure 3(b). City drive

The results of this investigation offer insight into crucial aspects of driving behavior. An important discovery is that driving contexts play a significant role. The dissimilarities in attention allocation between highway and city drives emphasize the necessity of comprehending driving contexts. City drives as shown in Figures 3a and b, which are rife with distractions such as pedestrians, intersections, and frequent stops, may require more advanced interventions to ensure driver attentiveness. The information gathered has the potential to be pivotal in devising context-specific interventions to enhance road safety. For instance, adaptive driver assistance systems (ADAS) can be developed to provide more frequent alerts in city environments, taking into account the higher likelihood of distractions. The efficacy of the Segment Anything Model (SAM) in segmenting driver attention based on gaze information signifies the potential of zero-shot learning in behavioral research. Such models can be extended to other domains, providing insights without the need for extensive labeled data. Although this study addressed the primary visual focus of drivers and reasons behind some driving errors, additional research can delve deeper into specific distractions within city drives. Furthermore, expanding the dataset to encompass a broader spectrum of driving conditions, demographics, and vehicle types can enrich the understanding of driver behavior.

5. Conclusion

In conclusion, the correct prediction of driver intentions and the evaluation of attentiveness are critical in comprehending driver conduct and improving decision-making. This investigation employs zero-shot learning by fusing image features and eye gaze data to uncover intricate connections between attention and cognitive decision-making. By exploiting the DR(eye)VE dataset and the Segment Anything Model (SAM), the study reveals that drivers give priority to road attention when driving on highways as opposed to city drives. This finding has pragmatic implications for customized interventions aimed at enhancing road safety and driver awareness, propelling human-vehicle interaction research forward.

Acknowledgements

The authors would like to acknowledge supports from Dept. of Transportation – Center for Advanced Transportation Mobility: Expanding and Enhancing Autodrive Competition Experience for Aggie Autonomous Auto (Younho Seong), and Autodrive Project through SAE and General Motors (Sun Yi).

References

Amadori, P. V., Fischer, T., & Demiris, Y. HammerDrive: A task-aware driving visual attention model. *IEEE Transactions on Intelligent Transportation Systems*, vol. 23 no. 6, pp. 5573-5585, Febuary 9, 2021, https://doi.org/10.1109/tits.2021.3055120.

Bitkina, O. V., Park, J., & Kim, H. K. The ability of eye-tracking metrics to classify and predict the perceived driving workload. *International Journal of Industrial Ergonomics*, vol. 86, pp. 103193, 2021.

Büla, C., Eyer, S., von Gunten, A., Favrat, B., & Monod, S. Driving and cognitive impairment: how best to anticipate?. *Revue Medicale Suisse*, vol. 7, no. 316, pp. 2184-2189, November 1, 2011.

Chan, M. Brain and behavioural reflections of distracted driving associated with emotion processing and social factors. 2016.

Epple, N., Chopra, H., & Riener, A. How Do Drivers Observe Surrounding Vehicles in Real-World Traffic? Estimating the Drivers Primary Observed Traffic Objects. In *2021 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, pp. 594-601, July 11, 2021.

Firoz, K. F., & Seong, Y. A Neural study of intuitive mode of cognition while decision-making using Artificial Grammar Learning paradigm. In *IISE Annual Conference and Expo*. IISE, May 22, 2023.

Firoz, K. F., Seong, Y., & Oh, S. A neurological approach to classify trust through EEG signals using machine learning techniques. In *2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS)*, IEEE, pp. 1-6, November 17, 2022.

Gyimah, N. K., Girma, A., Mahmoud, M. N., Nateghi, S., Homaifar, A., & Opoku, D. A robust completed local binary pattern (rclbp) for surface defect detection. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, pp. 1927-1934, October 17, 2022.

Gyimah, N. K., Gupta, K. D., Nabil, M., Yan, X., Girma, A., Homaifar, A., & Opoku, D. (2023, March). A Discriminative DeepLab Model (DDLM) for Surface Anomaly Detection and Localization. In *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, pp. 1137-1144, March 8, 2023.

Kaya, N., Girgis, J., Hansma, B., & Donmez, B. Hey, watch where you're going! An on-road study of driver scanning failures towards pedestrians and cyclists. *Accident Analysis & Prevention*, vol. 162, pp. 106380, November 1, 2021.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A., Lo, W.-Y., Dollár, P., & Girshick, R. Segment anything. *arXiv preprint arXiv:2304.02643*, April 5, 2023.

Kotseruba, I., & Tsotsos, J. K. Behavioral research and practical models of drivers' attention. *arXiv preprint arXiv:2104.05677*, April 12, 2021.

Lemonnier, S., Brémond, R., & Baccino, T. Discriminating cognitive processes with eye movements in a decision-making driving task. *Journal of Eye Movement Research*, vol. 7, no. 4, July 25, 2014.

Palazzi, A., Abati, D., Solera, F., & Cucchiara, R. Predicting the Driver's Focus of Attention: the DR (eye) VE Project. *IEEE transactions on pattern analysis and machine intelligence*, vol.41 no.7, pp.1720-1733, June 8, 2018.

Pourpanah, F., Abdar, M., Luo, Y., Zhou, X., Wang, R., Lim, C. P., Wang, X.-Z., & Wu, Q. M. J. A review of generalized zero-shot learning methods. *IEEE transactions on pattern analysis and machine intelligence*. July 18, 2022.

Sudkamp, J., Bocian, M., & Souto, D. The role of eye movements in perceiving vehicle speed and time-to-arrival at the roadside. *Scientific reports*, vol. 11 no.1, p. 23312, December 2, 2021.

Sun, Q., Xia, J., Nadarajah, N., Falkmer, T., Foster, J., & Lee, H. (2016). Assessing drivers' visual-motor coordination using eye tracking, GNSS and GIS: a spatial turn in driving psychology. *Journal of spatial science*, vol. 61 no.2, pp. 299-316, July 2, 2016.

Biographies

Kelvin K. Kwakye is a Ph.D. candidate at North Carolina Agricultural and Technical State University (NCAT) with a strong research interest in human factors, machine learning applications and transportation safety. Kelvin, who was born and raised in Ghana, developed an early interest in science and technology. Kelvin's academic journey began with his undergraduate studies in Chemical Engineering, where he established a solid foundation. He decided to pursue further studies in Industrial and Systems Engineering after completing his undergraduate studies in 2017, recognizing the impact this field can have on a wide range of industries. Kelvin's desire to use his skills and knowledge to benefit society extends beyond his academic pursuits. He is also active in the community, having volunteered with several organizations in Ghana and the United States.

Younho Seong received the B.Sc. and M.Sc. degree in industrial engineering from Inha University, South Korea, in 1991 and 1993 respectively, and Ph.D. degree in industrial engineering from State University of New York at Buffalo in 2002. He is currently Professor of industrial and systems engineering with North Carolina A&T State University, Greensboro, NC, USA. His research interests include cognitive engineering, ecological approaches to human judgment & decision-making, machine learning, human trust in automation, social aspect of autonomous vehicles, human–robot interaction, designing interventions to support human cognition, virtual and augmented reality, neuroergonomics, and brain–computer interface.

Sun Yi received the B.Sc. degree in mechanical and aerospace engineering from Seoul National University, Seoul, Korea, in 2004, and the M.Sc. and Ph.D. degrees in mechanical engineering from the University of Michigan, Ann Arbor, in 2006 and 2009, respectively.
He is currently a Professor with the Department of Mechanical Engineering, North Carolina A&T State University, Greensboro. His current research interests include time-delay systems, and analysis and control of dynamic systems with applications to mobile robots and satellites.
Dr. Yi is the Publication Chair of the 2012 ASME DSCC/MOVIC Conference.

Armstrong Aboah was born in Kumasi, Ghana. He received a B.S. in civil engineering from
the Kwame Nkrumah University of Science and Technology in 2017 and MS in Transportation Engineering from Tennessee Technological University in 2019. After this, he obtained a Ph.D. degree in Transportation Engineering with a focus on Computer Vision and Machine learning from the University of Missouri, Columbia, MO, in 2022.
From 2018 to 2019, he was a Research Assistant at the Tennessee Technological University. Also, from 2020 to 2022, he was a Research Assistant at the University of Missouri-Columbia.
He authored about 20 articles. His research interests include computer vision, smart cities, the Internet of Things, ITS, autonomous vehicles, pavement, and structural health monitoring. Currently, Armstrong Aboah is an Assistant Research Professor at the University of Arizona.