Cost-Efficient Driver Distraction Detection System: Transformer-Based Classification on Bayer Image

Chuheng Wei[1]; Ziye Qin[2]; Guoyuan Wu[3]; Matthew J. Barth[4]; and Zhanbo Sun[5]

[1]Center for Environmental Research and Technology, Univ. of California, Riverside, CA (corresponding author). Email: chuheng.wei@email.ucr.edu
[2]Center for Environmental Research and Technology, Univ. of California, Riverside, CA; School of Transportation and Logistics, Southwest Jiaotong Univ., Chengdu, China.
Email: ziye.qin@my.swjtu.edu.cn
[3]Center for Environmental Research and Technology, Univ. of California, Riverside, CA.
Email: gywu@cert.ucr.edu
[4]Center for Environmental Research and Technology, Univ. of California, Riverside, CA.
Email: barth@ece.ucr.edu
[5]School of Transportation and Logistics, Southwest Jiaotong Univ., Chengdu, China.
Email: zhanbo.sun@home.swjtu.edu.cn

ABSTRACT

Driver State Monitoring (DSM) is paramount for improving driving safety for both drivers of ego-vehicles and their surrounding road users, increasing public trust, and supporting the transition to autonomous driving. This paper introduces a Transformer-based classifier for DSM using an in-vehicle camera capturing raw Bayer images. Compared to traditional RGB images, we opt for the original Bayer data, further employing a Transformer-based classification algorithm. Experimental results prove that the accuracy of the Bayer Color-filled type images is only 0.61% lower than that of RGB images. Additionally, the performance of Bayer data is closely comparable to RGB images for DSM purposes. However, utilizing Bayer data can offer potential advantages, including reduced camera costs, lower energy consumption, and shortened Image Signal Processing (ISP) time. These benefits will enhance the efficacy of DSM systems and promote their widespread adoption.

INTRODUCTION

The safety of drivers is a paramount concern in the modern transportation industry. With the ever-increasing integration of technology into our vehicles, Driver State Monitoring (DSM) systems have emerged as a critical tool to enhance driving safety (Guettas et al. 2019). In addition to benefiting the drivers of ego-vehicles, these systems also contribute to the building of public trust in the era of autonomous driving (Wang 2022). Furthermore, they play a vital role in safeguarding the well-being of other road users (Gonçalves and Bengler 2015).

A unique approach to our research is motivated by the urgent need to improve road safety and reduce accidents caused by driver distraction. Our motivation is to explore the use of Bayer data as an alternative to traditional DSM systems that rely on RGB images. The objective of this project is to address a number of critical challenges directly by harnessing the original sensor data directly. These include reducing camera costs, lowering energy consumption, and shortening Image Signal Processing (ISP) time (Wei et al. 2023), ultimately enhancing the efficiency and cost-effectiveness of DSM systems.

Backgrounds

(1) Driver state monitoring and driver distraction detection systems

The DSM systems are integral components in contemporary vehicles aiming to detect, in real-time, the physiological and cognitive states of a driver (Baek et al. 2018). These states encompass fatigue, distraction, drunkenness, and a myriad of other conditions that could potentially impede one's ability to operate a vehicle safely.

The detection of driver distraction is a significant component of DSM. Distraction is any activity or factor that diverts the driver's attention away from the primary task of operating the vehicle in the context of driving. A person may experience this from internal stimuli, such as cognitive preoccupation or drowsiness, as well as external stimuli, such as using his/her mobile phone, adjusting the in-car entertainment or navigation system, or engaging in conversation with other passengers (Abouelnaga et al. 2017). A variety of road accidents are preceded by these distractions, which is why timely detection is of utmost importance (Eraqi et al. 2019). In order to identify signs of driver distraction and take preventive measures, modern DSM systems use a variety of techniques, including visual analytics from cameras.

Employing a combination of cameras, sensors, and sophisticated algorithms, these systems can issue warnings or even take corrective actions to prevent potential accidents (Wei et al. 2022). As the bridge between traditional vehicles and fully autonomous driving, DSM enhances safety and fosters public trust in next-generation vehicular technologies.

(2) Raw (bayer) image data

Raw image data pertains to a particular configuration of the color filter array (CFA) widely adopted in numerous digital cameras to record color imagery. Named after its inventor, Bryce Bayer, this unique filter mosaic consists of a strategic alignment of red, green, and blue filters organized on a square grid of photosensors (Bayer 1976). As illustrated in Figure 1, the elementary 2x2 unit of the Bayer CFA comprises one red pixel, two green pixels, and one blue pixel. Notably, the green filter is emphasized due to its alignment with the human visual system's spectral response band, making it particularly sensitive to the nuances of green light (Wang and Jeon 2015). During image acquisition, each pixel registers information corresponding to just a single color, be it red, green, or blue. Subsequently, the unrecorded color values for each pixel are deduced through interpolation of adjacent pixels, resulting in a comprehensive color image. This intricate procedure is termed "demosaicing".

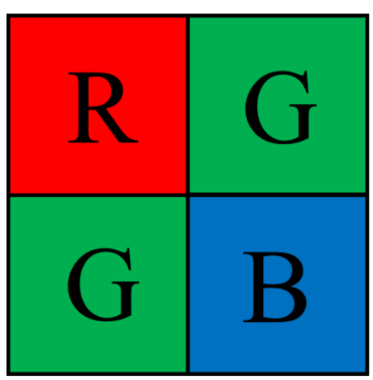

Figure 1. The minimum unit of Bayer CFA.

Opting for Bayer data is advantageous over RGB images in certain applications. Not only does it represent the rawest form of imaging data, but it also allows for greater flexibility in post-processing (Chan et al. 2023). By working directly with Bayer data, researchers can access unadulterated imaging information, potentially unlocking further nuances that may be pivotal in certain applications like DSM.

Contributions

This paper advances DSM research in two pivotal ways:

Transformer-based classification

We employ a Transformer-based classification algorithm, an architecture that has recently gained prominence in various tasks but is yet to be fully explored for DSM. Experimental evaluations manifest that the Transformer-based classifier offers competitive, if not superior, performance in distraction detection tasks.

Innovative input data

Instead of conventionally processed images, this work introduces the innovative use of raw Bayer images as the primary input for in-vehicle driver distraction detection. This offers the potential for significant cost and energy savings.

Structure of the Paper

Following the introduction, Section 2 delves into related work that contextualizes our study within the broader driver distraction detection research landscape. Section 3 elucidates our methodology, presenting a detailed overview of the Bayer data acquisition and the Transformer-based classification mechanism. Section 4 combines both experimental setups and the consequential findings. Finally, Section 5 consolidates our results, drawing conclusions and paving the way for future research endeavors.

RELATED WORK

Driving Distraction Detection

Distracted driving encompasses any activity that diverts a driver's attention from the primary responsibility of driving. Distractions can be categorized as visual, manual, or cognitive. Activities such as texting, talking on a cellphone, adjusting the radio, engaging in conversations with passengers, or even simply daydreaming are all examples of such distractions. The phenomenon of distracted driving has ascended as a grave concern in road safety, prompting widespread research in both academic and industrial sectors (Liang and Lee 2014; Wang et al. 2022). The imperative of timely detection of driver distraction, followed by either alerting the driver or an automatic vehicular control takeover, cannot be overemphasized. Traditional classifiers, such as the Support-vector machine (SVM), have also been employed for distraction recognition using camera data (Kutila et al. 2007). Furthermore, Liu et al. (2016) tapped into semi-supervised machine learning methodologies to classify drivers' states by assessing their eye and head movements.

To this end, a multitude of studies have delved into methods to accurately identify driver distractions. Wang et al. (2022) employed a Long Short-Term Memory (LSTM) model paired with vehicle dynamics sensor data to pinpoint cell phone usage as a driver distraction. Concurrently, bioelectric signals have been explored in conjunction with naturalistic driving data. Li et al. (2022) harnessed temporal data along with electroencephalography (EEG) signals to delve into the nuances of driver distraction. In recent times, in-vehicle cameras, offering practicality in their application, have been deployed to extract detailed insights from the facial expressions and movements of drivers. For instance, Grahn and Taipalus (2021) innovated a technique to discern driver distraction by analyzing varying glancing behaviors. In a similar vein, Craye and Karray (2015) utilized AdaBoost classifiers and the Hidden Markov Model to detect myriad distractions, such as making phone calls, drinking, texting, and general lack of focus on driving. Eye-specific metrics, including eyelid closure rate and distance changes, have been leveraged by Sigari et al. (2013) to determine a driver's state of distraction. Machine learning models and techniques have been particularly instrumental in this realm. Li et al. (2022) used a variety of driver actions as training data to craft models adept at recognizing distractions. Dong and Lin (2021) utilized Convolutional Neural Networks (CNN) to classify an array of distracted driving behaviors. Pushing the boundaries, Huang and Fu (2022) introduced a deep 3D residual network embedded with an attention mechanism and encoder-decoder (D3DRN-AMED) structure, specifically designed to detect driver distractions by examining their focal points.

Bayer Image Data Applications

Bayer image data is frequently transformed to an RGB-per-pixel format to facilitate image recognition tasks. For instance, Huang et al. (2008) pioneered a driver monitoring system using this methodology. Similarly, the Oxford RobotCar perceived its surroundings by converting Bayer data into more recognizable formats (Maddern et al. 2017). Horak (2011) employed this approach in driver eye tracking to identify signs of fatigue. Given these precedents, a pertinent question arises: is it possible to leverage Bayer data directly for image recognition? If so, this can significantly expedite image processing and result in substantial hardware cost savings. Recent research suggests that this is feasible. For example, Wei (2022) successfully employed raw Bayer data in vehicle detection and tracking tasks. Likewise, Chan et al. (2023) harnessed Bayer data for vehicle detection on the renowned KITTI dataset. Considering the critical nature of tasks like distraction detection, which necessitates real-time monitoring of a driver's status, the direct utilization of Bayer data is a promising avenue worthy of further exploration. In addition, there have been efforts to develop specialized models tailored to Bayer data. For instance, Wei et al. (2023) attempted to adapt the Faster R-CNN neural network model specifically for Bayer data and introduced camera parameters as supplementary input information for object detection. Their findings demonstrated that customizing deep neural network models to accommodate the inherent characteristics of Bayer data can yield beneficial improvements in performance.

METHODOLOGY

Bayer Data Generation

In the Bayer data generation section of this study, we adopted the widely used RGGB pattern for generating Bayer images. In this section, we outline the methodology employed to generate two types of Bayer data from RGB images, namely the Bayer 0-filled image and the Bayer color-filled image (Chan et al. 2023). It is important to note that the process of converting raw sensor data into RGB images involves the Image Signal Processing (ISP) stage, making the methods employed here inherently approximate.

(1) Bayer 0-filled data

In the Bayer 0-filled image generation method, a new image is created while maintaining its color format. Specifically, within each 2x2 pixel grid, only one channel's color value is

retained, and all other positions within the grid are set to zero. For example, for the R channel, only the top-left pixel within each 2x2 grid retains the R value, while all other positions within the grid are assigned zero. Similarly, for the B channel, only the bottom-right pixel within each 2x2 grid retains the B value, while the remaining positions are set to zero. In the case of the G channel, only the top-right and bottom-left positions within each grid retain their original G values, while the other two positions in the grid are set to zero.
The mathematical representation of this method is:

For the R channel

$$Bayer_R(i,j) = \{I_R(i,j)\ if\ (x \% 2 = 0)\ and\ (y \% 2 = 0)\ 0\ otherwise$$

For the G channel

$$Bayer_G(i,j) = \{I_G(i,j)\ if\ (x \% 2\ \neq y \% 2)\ 0\ otherwise$$

For the B channel

$$Bayer_B(i,j) = \{I_B(i,j)\ if\ (x \% 2 = 1)\ and\ (y \% 2 = 1)\ 0\ otherwise$$

where:

$I_R(i,j)$, IG(i,j), IB(i,j) represent the Red, Green, and Blue color channels of the raw sensor data at pixel *(i,j)*, respectively.

BayerR, BayerG, BayerR represent the Red, Green, and Blue value of the generated Bayer image at pixel *(i,j)*

i and *j* represent the row and column indices of the pixels in the sensor data or Bayer image.

% denotes the modulo operation.

(2) Bayer color-filled data

In the Bayer color-filled image generation method, a new image is created while preserving its color format. Within each 2x2 pixel grid, colors are replaced with specific values. To elaborate, for the R channel, all four pixels within the grid are assigned the value of the top-left R channel pixel. Similarly, for the B channel, all four pixels in the grid are assigned the value of the bottom-right B channel pixel. Regarding the G channel, the top-right and bottom-left pixels within each grid retain their original G values, while the other two pixels in the grid are set to the average of the adjacent G channel values.
The mathematical representation of this method is:

For the R channel

$$Bayer_R(i,j) = I_R(i-(i\%2), j-(j\%2))$$

For the G channel

$$Bayer_R(i,j) = \{I_G(i,j)\ if\ (i\%2) \neq (j\%2)\ \frac{1}{2}(I_G(i-(i\%2),j) + I_G(i,j-(j\%2)))\ otherwise$$

For the B channel

$$Bayer_B(i,j) = I_B(i - (i\%2) + 1, j - (j\%2) + 1)$$

where:

$I_R(i,j)$, $I_G(i,j)$, $I_B(i,j)$ represent the Red, Green, and Blue color channels of the raw sensor data at pixel (i,j), respectively.

$Bayer_R$, $Bayer_G$, $Bayer_R$ represent the Red, Green, and Blue value of the generated Bayer image at pixel (i,j)

i and j represent the row and column indices of the pixels in the sensor data or Bayer image.

% denotes the modulo operation.

Transformer Model

To harness the potential of Bayer data for driver distraction detection, the Transformer model, renowned for its proficiency in handling sequential data, is incorporated to discern and predict the driver's state.

The Transformer architecture, introduced by Vaswani et al. (2017), stands apart due to its distinctive self-attention mechanism. The multi-head attention layer in Figure 2 stands as a cornerstone in the Transformer model, epitomizing its novelty and efficacy in handling sequential data. Instead of utilizing a singular set of Query (Q), Key (K), and Value (V) weight matrices, this mechanism deploys multiple sets, referred to as "heads", to simultaneously process the input data from different representational spaces. This multiplicity ensures that for each input word or token, diverse relationships and contextual dependencies with other tokens can be captured (Han et al., 2023). Unlike conventional models that process input tokens sequentially or in fixed patterns, the Transformer's self-attention mechanism assesses and assigns varying weights to different tokens based on their contextual significance. This ensures that during the production of an output, the model can dynamically shift its focus across different portions of the input, allowing for a richer and more contextually-aware representation of data. This inherent ability of the Transformer to recognize patterns and relationships within the data makes it particularly suitable for tasks like driver distraction detection, where the nuanced understanding of a driver's state based on raw Bayer data becomes paramount.

Figure 3 illustrates the architecture of our proposed model, which is designed specifically to analyze Bayer image data in the context of Driver State Monitoring (DSM). The purpose of this model is to capture and discern the multifaceted states of the driver from raw Bayer images.

Convolutional Neural Network (CNN): The CNN is used as the first feature extractor in our model. This network is designed to process Bayer image data, exploiting the spatial hierarchy of images through convolutional filters that detect edges, textures, and other visual patterns at various scales. In the case of high-resolution Bayer images, the CNN architecture has the advantage of reducing the dimensionality of data while preserving the essential information.

Transformer-based Encoder: The Transformer-based encoder receives the feature maps produced by the CNN. It consists of a series of self-attention mechanisms that enable the model to weigh and integrate information across an entire image. By performing parallel computations, the Transformer-based encoder is able to capture complex relationships within data. This is different from traditional approaches, which process data sequentially. It is especially useful for determining the temporal dynamics of driver behavior, which may not be immediately adjacent to one another in a sequence of images.

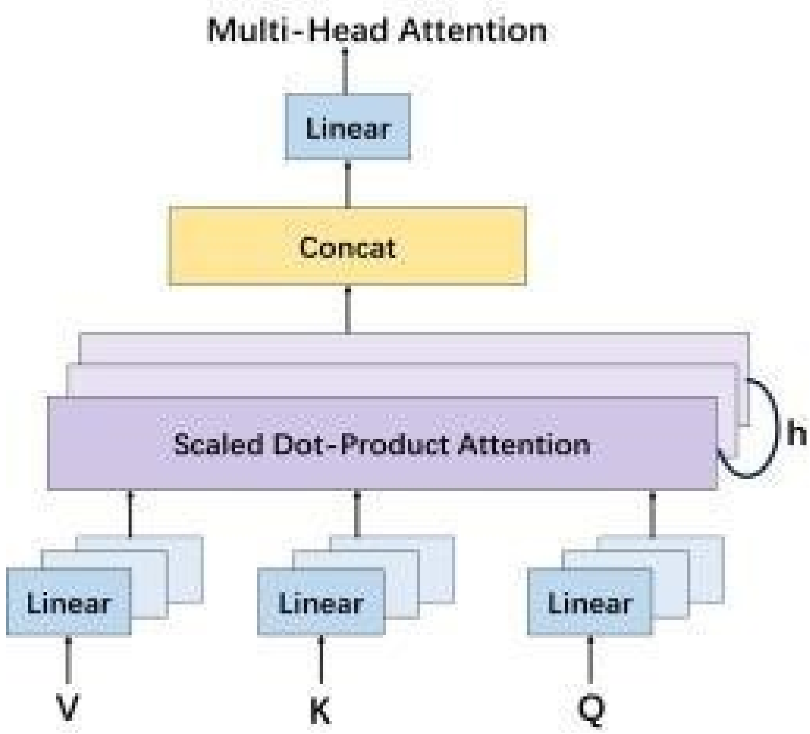

Figure 2. Multi-head Attention.

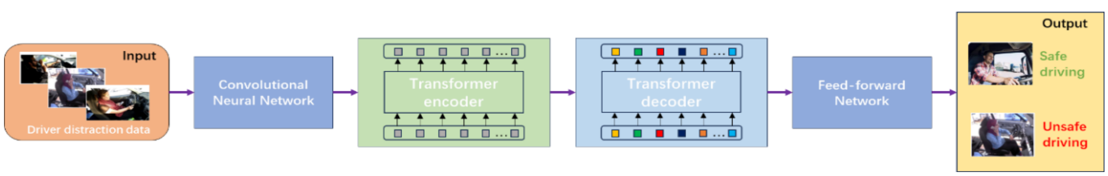

Figure 3. The Transformer architecture for driver distraction detection.

Transformer-based Decoder: Following the encoder, the Transformer-based decoder synthesizes the encoded features into a comprehensive representation. As a mirror of the encoder structure, it incorporates additional cross-attention layers that allow it to focus on specific segments of the input sequence in accordance with the encoder context. With this component, the sequence of driver states can be reconstructed and the current state can be inferred with a high level of accuracy.

Feed-Forward Network: The final component is the feed-forward network, which interprets the output from the Transformer-based decoder. The network is composed of dense layers that perform classification based on contextualized features. Feature representations are mapped to probabilities of safe or unsafe driving states based on a high-dimensional feature representation. During training, the network minimizes prediction errors, ensuring that the final output accurately reflects the likelihood that a driver is distracted.

EXPERIMENTS AND RESULTS

Dataset Selection

The dataset used in this study is the Distracted Driver Dataset (Abouelnaga et al. 2017; Eraqi et al. 2019) from the Machine Intelligence Group at the American University in Cairo (MI-AUC). It consists of individual images with a resolution of 1920×1080 pixels and captures real-world driving activities.

The study comprises a diverse cohort of 31 participants hailing from seven distinct countries, which consisted of 22 males and 9 females. Furthermore, the dataset encompasses imagery captured within four distinct car models. Additionally, the dataset is thoughtfully categorized, comprising a total of 17,308 frames distributed across ten distinct categories: Drive Safe (3,686 frames), Talk Passenger (2,570 frames), Text Right (1,974 frames), Drink (1,612 frames), Talk Left (1,361 frames), Text Left (1,301 frames), Talk Right (1,223 frames), Adjust Radio (1,220 frames), Hair & Makeup (1,202 frames), and Reach Behind (1,159 frames).

(a) RGB image (b) Bayer 0-filled image (c) Bayer color-filled image

Figure 4. **Example of different types of images used in this study.**

To streamline the experiment, we collapsed all categories into two groups: safe driving and unsafe driving. To prevent an unwieldy dataset size, we randomly selected 3,600 images from safe driving and an equal number of images from the various unsafe driving categories, resulting in a total experimental dataset of 7,200 images. Within this dataset, 80% of the images were allocated for training purposes, while the remaining 20% were designated for validation.

The RGB images utilized in the experiment were directly sourced from the Distracted Driver Dataset. Color-filled Bayer images, on the other hand, were generated using the previously described Bayer image generation method. Figure 4 illustrates a comparative display of the RGB and Bayer images employed in the experiment.

Results

(1) Quantitative Analysis (Detection Results)

In our investigation, we conducted training and evaluation experiments using three distinct Transformer-based classifier models. These models were trained on the same set of images in RGB format, Bayer 0-filled format, and Bayer color-filled format, each subjected to 10,000 training epochs. The achieved classification accuracies, as presented in Table 1, reveal noteworthy insights into the performance of these formats.

The RGB format exhibited the highest accuracy at 70.65%, highlighting its effectiveness as a baseline for classification tasks. Surprisingly, the Bayer color-filled format closely followed with an accuracy of 70.04%, outperforming the Bayer 0-filled format, which achieved an accuracy of 65.36%. This unexpected outcome suggests that the Bayer color-filled format, despite introducing some interpolation and color substitution effects, offers competitive performance, possibly due to its ability to retain more color information. These results underscore the potential utility of the Bayer color-filled format as an efficient alternative for driver distraction detection, but further investigation is warranted to explore its underlying advantages and limitations.

Table 1. Accuracy of models trained on images of different formats.

Format Number	Format	Accuracy
1	Original RGB	70.65%
2	Bayer Color -filled	70.04%
3	Bayer 0-filled	65.36%

However, it is crucial to note that achieving optimal performance with Transformer-based models requires extended training epochs. With increasing training epochs, the models in all three formats demonstrated improved accuracy (Carion et al. 2020). Although Transformer- based models require longer training times, they could capture complex patterns and features in the data with remarkable accuracy. In spite of the promising results demonstrated by the Bayer color-filled format, further investigation is recommended to fully unlock its potential and determine the optimal training duration in order to achieve peak accuracy.

(2) Qualitative Analysis (System Cost)

In evaluating the economic feasibility of the proposed driver distraction detection system, three key aspects were considered: hardware costs, processing costs, and energy consumption.

Hardware Costs: According to a comparative analysis, Bayer color cameras offer a significant price advantage over their three-CCD or three-CMOS counterparts. For instance, a high-quality 5-megapixel Bayer area scan camera is available for less than half the cost of an equivalent 3.2-megapixel prism camera. It is important to note that this notable price disparity plays a crucial role in large-scale deployments where initial capital expenditures are an important constraint (JAI, 2019).

Processing Costs: Bayer cameras utilize predefined patterns of color filters atop the pixels, requiring an interpolation process in order to estimate RGB values for each pixel (Adimec, 2013). Color accuracy may be marginally compromised by this approach, however, processing power requirements are significantly reduced compared to 3-CCD or 3-CMOS systems, which capture precise RGB values without interpolation, thus requiring increased computational resources and, consequently, higher processing costs (Adimec, 2013).

Energy Consumption: As compared to polymer filters used in conventional Bayer sensors, prism glass used in 3-CMOS cameras exhibits enhanced light transmission capabilities. This results in improved light sensitivity and a reduction in lighting requirements. However, this heightened sensitivity comes with an increased energy demand. In contrast, Bayer images' relatively low sensitivity could contribute to reduced energy consumption, especially in well-lit environments (Chouinard, 2018).

CONCLUSION

In this study, we introduce a novel Transformer-based classifier tailored for driver state monitoring (DSM) using two distinct raw Bayer images. Our aim is to explore its potential as a cost-effective, energy-saving alternative to traditional RGB imaging techniques. The experimental results reveal that models using Bayer data differ in accuracy by a mere 0.61% compared to their RGB counterparts, with a promising trend towards further convergence observed. Such findings underscore the feasibility of Bayer data as a suitable replacement for conventional RGB imaging. Moreover, our experiments demonstrate that Bayer data can closely match the performance of RGB images in DSM applications. Through this research, we not only highlight the benefits of employing Bayer data but also contribute to the broader adoption of DSM systems. This enhances road safety by making these systems more accessible and cost- effective, ultimately benefiting everyone on the road.

Firstly, we have demonstrated unequivocally the efficacy of Transformer-based methods for detecting driver distraction through a meticulous evaluation of classifiers trained on three distinct image categories. The results obtained from our experiment validate the robustness and adaptability of these models post-training, emphasizing their ability to distinguish and interpret complex visual cues indicative of driver distraction. As the field of deep learning continues to evolve, these models stand as a formidable tool in the pursuit of safer and more efficient driving experiences.

Secondly, the comparative analysis between RGB and Bayer color-filled images sheds light on a significant finding. The study illustrates that Bayer data has comparable potential to RGB images when it comes to detecting driver distraction states. This study suggests that even low-cost, energy-efficient cameras lacking sophisticated Image Signal Processing (ISP) units can be harnessed for driver state monitoring. Such an approach not only holds promise for reducing the hardware costs associated with driver monitoring systems but also paves the way for wider accessibility and adoption of this critical technology in enhancing road safety.

FUTURE WORK

To obtain more precise conclusions and enhance the performance of our model, our future work can be directed towards the following three aspects:

(1) Collection of Genuine Bayer Image Data: A critical direction for future research will be the collection of a real-world Bayer image dataset that is specifically tailored for monitoring the state of a driver. Rather than converting RGB images into Bayer data, which requires an irreversible Image Signal Processing (ISP) step, it is imperative to capture Bayer images directly from the vehicle's camera. It is anticipated that such an authentic dataset will not only enhance credibility of our findings, but will also serve as a valuable resource for training and evaluating DSM models. For the model to be robust and generalizable, it should include a variety of driving conditions, lighting scenarios, and driver behaviors.

(2) Introduction of Region of Interest (ROI) Mechanism: Through the implementation of a Region of Interest (ROI) mechanism within the model, key focal points in identifying driver behavior can be highlighted. As a result of this strategic integration, irrelevant information may be reduced, model precision could be enhanced, and training time could be reduced. In order for the model to be able to discern and categorize driver distractions effectively, we anticipate focusing its attention on critical areas of interest.

(3) Model Adaptation for Bayer Data: Adaptations tailored specifically for Bayer data should be explored in more detail. While we have demonstrated the effectiveness of Transformer-based models on this type of data, further refinement and optimization remain to be done. Customized architectures or modifications designed to harness the unique characteristics of Bayer data could potentially yield even better results. It may be possible to unlock untapped potential for driver state monitoring by fine-tuning or retraining existing Transformer-based models with a focus on Bayer data. By doing so, the model will be able to extract more nuanced and informative features from this unique image format.

(4) Robustness Assessment in Diverse Driving Environments: Our future research will explore the adaptability of our Transformer-based classification system in a variety of driving environments, including low-light and strobe scenarios. These conditions present distinct challenges to systems designed to monitor the driver's state of mind. We intend to conduct comprehensive testing and simulations in order to assess the efficiency of our system in these challenging lighting environments. During this assessment, data collected from real-world driving situations under a variety of lighting conditions will be collected and analyzed, a critical step in confirming the system's resilience against adverse lighting.

REFERENCES

Abouelnaga, Y., Eraqi, H. M., and Moustafa, M. N. (2017). "Real-time distracted driver posture classification." *arXiv preprint arXiv:1706.09498*.

Adimec. 2013. "Bayer Area Scan Color Cameras Compared to 3-CCD Color Cameras, Part 1." Adimec. Accessed [2023, 12 20]. https://www.adimec.com/bayer-area-scan-color-cameras-compared-to-3-ccd-color-cameras-part-1/.

Baek, J. W., Han, B. G., Kim, K. J., Chung, Y. S., and Lee, S. I. (2018). "Real-Time Drowsiness Detection Algorithm for Driver State Monitoring Systems." *Proc., 2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, 73-75.

Bayer, B. (1976). "Color imaging array." *United States Patent, no. 3971065*.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. (2020). "End-to-end object detection with transformers." *Proc., European conference on computer vision*, Springer, 213-229.

Chan, P. H., Wei, C., Huggett, A., and Donzella, V. (2023). "Raw camera data object detectors: an optimisation for automotive processing and transmission."

Chouinard, J. (2018) "3-CMOS machine vision cameras bring color fidelity to the market at half the price as previous models." Accessed 2023,12 21, https://www.linkedin.com/pulse/3- cmos-machine-vision-cameras-bring-color-fidelity-market-chouinard/.

Craye, C., and Karray, F. (2015). "Driver distraction detection and recognition using RGB-D sensor." *arXiv preprint arXiv:1502.00250*.

Dong, B. T., and Lin, H. Y. (2021). "An On-board Monitoring System for Driving Fatigue and Distraction Detection." *Proc., 2021 22nd IEEE International Conference on Industrial Technology (ICIT)*, 850-855.

Eraqi, H. M., Abouelnaga, Y., Saad, M. H., and Moustafa, M. N. (2019). "Driver distraction identification with an ensemble of convolutional neural networks." *Journal of Advanced Transportation*, 2019.

Gonçalves, J., and Bengler, K. (2015). "Driver state monitoring systems–transferable knowledge manual driving to HAD." *Procedia Manufacturing*, 3, 3011-3016.

Grahn, H., and Taipalus, T. (2021). "Refining distraction potential testing guidelines by considering differences in glancing behavior." *Transportation Research Part F: Traffic Psychology and Behaviour*, 79, 23-34.

Guettas, A., Ayad, S., and Kazar, O. (2011). "Driver state monitoring system: A review." *Proc., Proceedings of the 4th International Conference on Big Data and Internet of Things* , 1-7.

Horak, K. (2011)"Fatigue features based on eye tracking for driver inattention system." *Proc., 2011 34th International Conference on Telecommunications and Signal Processing (TSP)*, 593-597.

Huang, S., Giacalone, J.-P., and Fan, W. (2008). "A RGBI imager based background suppression method." *Proc., 2008 IEEE Intelligent Vehicles Symposium*, IEEE, 643-647.

Huang, T., and Fu, R. (2022). "Driver Distraction Detection Based on the True Driver's Focus of Attention." *IEEE Transactions on Intelligent Transportation Systems*, 23(10), 19374- 19386.

JAI. (2019). "When to Choose a Bayer Color Camera." JAI. Accessed 2023, 12 20. https://news.jai.com/blog/bayer-color-camera.

Kutila, M., Jokela, M., Markkula, G., and Rue, M. R. (2007). "Driver Distraction Detection with a Camera Vision System." *Proc., 2007 IEEE International Conference on Image Processing*, VI - 201-VI - 204.

Li, B., Chen, J., Huang, Z., Wang, H., Lv, J., Xi, J., Zhang, J., and Wu, Z. (2022). "A New Unsupervised Deep Learning Algorithm for Fine-Grained Detection of Driver Distraction." *IEEE Transactions on Intelligent Transportation Systems*, 23(10), 19272-19284.

Li, G., Yan, W., Li, S., Qu, X., Chu, W., and Cao, D. (2022). "A Temporal–Spatial Deep Learning Approach for Driver Distraction Detection Based on EEG Signals." *IEEE Transactions on Automation Science and Engineering*, 19(4), 2665-2677.

Liang, Y., and Lee, J. D. (2014). "A hybrid Bayesian Network approach to detect driver cognitive distraction." *Transportation research part C: emerging technologies*, 38, 146- 155.

Liu, T., Yang, Y., Huang, G. B., Yeo, Y. K., and Lin, Z. (2016). "Driver Distraction Detection Using Semi-Supervised Machine Learning." *IEEE Transactions on Intelligent Transportation Systems*, 17(4), 1108-1120.

Maddern, W., Pascoe, G., Linegar, C., and Newman, P. (2017). "1 year, 1000 km: The oxford robotcar dataset." *The International Journal of Robotics Research*, 36(1), 3-15.

Sigari, M.-H., Fathy, M., and Soryani, M. (2013). "A driver face monitoring system for fatigue and distraction detection." *International journal of vehicular technology*, 2013, 1-11.

Wang, L., and Jeon, G. (2015). "Bayer pattern CFA demosaicking based on multi-directional weighted interpolation and guided filter." *IEEE Signal Processing Letters*, 22(11), 2083- 2087.

Wang, Y., Chan, P. H., & Donzella, V. (2022). "A two-stage h. 264 based video compression method for automotive cameras." *IEEE 5th International Conference on Industrial Cyber-Physical Systems (ICPS)*, 01-06.

Wang, X., Xu, R., Zhang, S., Zhuang, Y., and Wang, Y. (2022). "Driver distraction detection based on vehicle dynamics using naturalistic driving data." *Transportation research part C: emerging technologies*, 136, 103561.

Wei, C. (2022). "Vehicle Detecting and Tracking Application Based on YOLOv5 and DeepSort for Bayer Data." *Proc., 2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 843-849.

Wei, C., Liu, C., and Chi, S. (2022). "A Driver Distraction Detection Method Based on Convolutional Neural Network." *Proc., 2022 7th International Conference on Computer and Communication Systems (ICCCS)*, IEEE, 272-277.

Wei, C., Wu, G., Barth, M., Chan, P. H., Donzella, V., and Huggett, A. (2023). "Enhanced object detection by integrating camera parameters into raw image-based faster R-CNN." *2023 IEEE 26th international conference on intelligent transportation systems (ITSC)*, IEEE, Spain.