# Driving as well as on a Sunny Day? Predicting Driver's Fixation in Rainy Weather Conditions via a Dual-branch Visual Model

Han Tian, Tao Deng, *Member, IEEE*, and Hongmei Yan

## Supplementary Material

**Comparison of Our DriFixD(rainy) with Other SOTA Datasets:** In this work, we present a Drivers' Fixation Dataset in rainy weather conditions (DrFixD(rainy)). We focus on how to drive safely in rainy conditions, not all-weather conditions. As reported in literature [1], rain-related fatal crashes represented about 6.8% of the total fatal crashes on average. As shown in Table I, although the quantity of rainy videos in the DrFixD(rainy) we proposed is limited, the length of time for each video is long, and they are specific for rainy weather. Moreover, each video contains 30 drivers' eye-tracking data in DrFixD(rainy). By contrast, other driver attention benchmarks also include rainy videos, e.g., DADA-2000, DR(eye)VE, and BDD-A, but the number of viewers per video is low in quantity. There are about 5 viewers per video on DADA-2000, 1 viewer per video on DR(eye)VE, and 4 viewers per video on BDD-A, respectively. Therefore, the proposed DrFixD(rainy) dataset has more driver's attention samples than all other existing datasets. Besides, during the collection of the drivers' eye tracking data, the drivers were asked to view the rainy driving videos under the especial driving task, i.e., driving safely in rainy weather conditions. Thus, the eye fixation and attention pattern of subjects are specific for rainy conditions.

**Validation on Public Datasets:** In addition to verifying the performance of our model on other rainy weather conditions, we also evaluated it on DR(eye)VE-rainy [6] and BDD-A-rainy [5] dataset. The rainy condition videos are selected from the original datasets.

*DR(eye)VE-rainy:* In order to prove the robustness of the model, we made the qualitative and quantitative comparison of our model with other SOTA deep learning-based models, i.e., TASED-Net [9], SalEMA [10] and EML-NET [11], on the rainy weather conditions in DR (eye)VE dataset. There are totally 23 rainy videos in DR(eye)VE. We chose 11 rainy videos for training, 2 videos for validation and 10 testing videos. All of the models were retrained and tested on the DR(eye)VE-rainy dataset. Note that, DR(eye)VE recorded only one driver's fixation per frame, so the fixation map of the frame comes from one driver's fixations overlapped from previous 12 frames to afterward 12 frames.

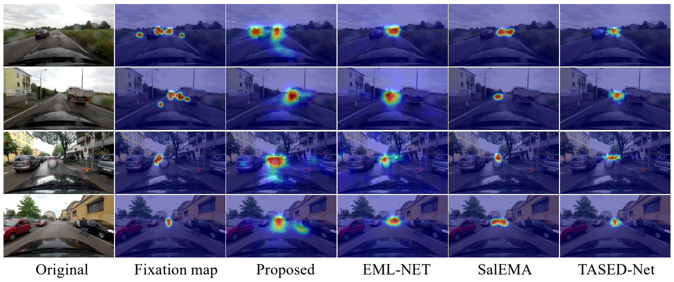

Fig. 1. Saliency maps selected from the DR(eye)VE-rainy randomly. From left to right: the original image, the fixation map (one driver's fixations overlapped from previous 12 frames to afterward 12 frames) of DR(eye)VE, the prediction map of current frame by our proposed model, EML-NET, SalEMA and TASED-Net.

The qualitative comparison between our model and the three SOTA deep learning-based models (TASED-Net, SalEMA and EML-NET) is shown in Fig. 1. For simple traffic scenes such as the first and the second rows in Fig. 1, all models can predict the salient regions (the road ahead) or targets (the car or truck ahead), but our model locates them more accurately compared with others. For complex traffic scenes such as the third and fourth rows, when multiple targets appear in the scene, our model can not only detect the main salient area (the road ahead), but also the secondary important target related to driving if they exist (the roadside traffic sign in the third row and the biker in the fourth row), even the third important target (the car at the left side in the third row). By comparison, other SOTA models may miss these salient targets or areas. All in all, our model has better performance in detecting multiple targets, mainly due to the facts that our model includes dual branches and the temporal correlation between adjacent frames is considered.

Furthermore, a quantitative comparison is presented in Table II. In this comparison, the metrics of CC, KLD and SIM are adopted to evaluate the performance of our model. As shown in Table II, our model shows a better performance than others on the KLD and SIM metrics, only worse than EML-NET on CC.

*BDD-A-rainy:* Berkeley DeepDrive Attention (BDD-A) [5] is a driver attention dataset in the critical driving situations. It consists of 1,232 videos in various weather and lighting conditions. However, it contains only 15 obvious and valid rainy videos (about 10 second per video, total 4,506 frames), so it is difficult to train the models on this dataset because the data is too small. Thus, we use the trained models to test on the BDD-A-rainy dataset. Each model is trained on the proposed DrFixD(rainy) dataset and tested on the BDD-A-rainy dataset. The quantitative evaluation results are reported in Table III, and our model achieves a promising performance.

TABLE I
COMPARISON OF SALIENCY/ATTENTION DATASETS IN NATURAL AND DRIVING SCENES.

| Dataset | Year | Specialty | Weather | Quantity of rainy | #Viewers | Dynamic |
|---|---|---|---|---|---|---|
| MIT [2] | 2012 | Natural | - | - | 39 | No |
| SLICON [3] | 2015 | Natural | - | - | 16 | No |
| DHF1K [4] | 2018 | Natural | - | - | 17 | Yes |
| BDD-A [5] | 2018 | Traffic Driving/brake events | Sunny, Cloudy, Rainy | 15 videos | 45 (4 per video) | Yes |
| DR(eye)VE [6] | 2018 | Traffic Driving | Sunny, Rainy, Snowy | 23 videos | 8 (1 per video) | Yes |
| DADA-2000 [7] | 2019 | Driving accidents | Sunny, Cloudy, Rainy | ≈ 140 videos | 20 (>5 per video) | Yes |
| Deng [8] | 2020 | Traffic Driving | Sunny, Cloudy | - | 28 per video | Yes |
| **DrFixD(rainy) (ours)** | 2022 | Traffic Driving | Rainy | 16 videos | 30 per video | Yes |

Han Tian and Hongmei Yan are with the MOE Key Lab for Neuroinformation, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, 610054, PR China. (E-mails: ivy_uestc@163.com (H. Tian); hmyan@uestc.edu.cn (H.M. Yan))

Tao Deng is with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, 611756, PR China. (E-mail: tdeng@swjtu.edu.cn)

TABLE II
PERFORMANCE COMPARISON OF MODELS TRAINED AND TESTED ON DR(EYE)VE-RAINY DATASET.

| Models | CC↑ | SIM↑ | KLD↓ |
|---|---|---|---|
| TASED-Net | 0.4483 | 0.3610 | 2.9156 |
| SalEMA | 0.3970 | 0.3214 | 2.3312 |
| EML-NET | **0.5113** | 0.2171 | 2.066 |
| **Proposed** | 0.4638 | **0.3709** | **1.9793** |

TABLE III
PERFORMANCE COMPARISON OF MODELS TRAINED ON DRFIXD(RAINY) AND TESTED ON BDD-A-RAINY DATASET.

| Models | CC↑ | SIM↑ | KLD↓ |
|---|---|---|---|
| TASED-Net | 0.4755 | 0.3920 | 2.1097 |
| SalEMA | 0.4811 | 0.3673 | 1.4570 |
| EML-NET | 0.5173 | 0.3264 | 1.5161 |
| **Proposed** | **0.5555** | **0.4020** | **1.3154** |

## REFERENCES

[1] Z. Han and H. O. Sharif, "Investigation of the relationship between rainfall and fatal crashes in texas, 1994–2018," *Sustainability*, vol. 12, no. 19, p. 7976, 2020.

[2] Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba, "Mit saliency benchmark," 2015, Available: http://saliency.mit.edu/.

[3] M. Jiang, S. Huang, J. Duan, and Q. Zhao, "Salicon: Saliency in context," in *Proc. IEEE Conf. Comp. Vis. Patt. Rec.*, 2015.

[4] W. Wang, J. Shen, J. Xie, M.-M. Cheng, H. Ling, and A. Borji, "Revisiting video saliency prediction in the deep learning era," *IEEE Trans. Patt. Analys. Mach. Intell.*, vol. 43, no. 1, pp. 220–237, 2021.

[5] Y. Xia, D. Zhang, J. Kim, K. Nakayama, K. Zipser, and D. Whitney, "Predicting driver attention in critical situations," in *ACCV*. Springer, May. 2018, Conference Proceedings, pp. 658–674.

[6] A. Palazzi, D. Abati, F. Solera, and R. Cucchiara, "Predicting the driver's focus of attention: the dr (eye) ve project," *IEEE TPAMI*, vol. 41, no. 7, pp. 1720–1733, Jul. 2019.

[7] J. Fang, D. Yan, J. Qiao, J. Xue, and H. Yu, "Dada: Driver attention prediction in driving accident scenarios," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–13, 2021.

[8] T. Deng, H. Yan, L. Qin, T. Ngo, and B. S. Manjunath, "How do drivers allocate their potential attention? driving fixation prediction via convolutional neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 5, pp. 2146–2154, 2020.

[9] K. Min and J. J. Corso, "Tased-net: Temporally-aggregating spatial encoder-decoder network for video saliency detection," in *Pro. IEEE Intern. Conf. Comp. Vis.*, 2019, pp. 2394–2403.

[10] P. Linardos, E. Mohedano, J. J. Nieto, N. E. O'Connor, X. Giró-i-Nieto, and K. McGuinness, "Simple vs complex temporal recurrences for video saliency prediction," in *British Machine Vision Conference*, 2019, pp. 1–12.

[11] S. Jia and N. D. Bruce, "Eml-net: An expandable multi-layer network for saliency prediction," *Image Vis. Comp.*, vol. 95, p. 103887, 2020.