Letter

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

Dear Editor,

Traffic driving is a dynamic and complicated task in which drivers are required to pay close attention to the important targets or regions to maintain safe margins. Rainy weather conditions make it more challenging with factors such as low visibility, raindrops, pedestrian with umbrellas, wipers, etc. Studies showed that rainy condition affects driving safety significantly [1], [2]. It is reported that, in raining weather condition, the odds for a fatal accident are 3.340 times higher on highways than on streets [3]. An investigation of the relationship between rainfall and fatal crashes in Texas from 1994 to 2018 on fatality analysis reporting system (FARS) database illustrated that rain-related fatal crashes represented about 6.8% of the total fatal crashes on average, moreover, the proportion showed high variability at the annual, monthly, and hourly time scales [4]. Therefore, raining is a complex and critical factor for road safety planning and management. In fact, the traffic environment is a dynamic scene with multiple sources of information, including important targets that are highly relevant to the current driving task as well as irrelevant targets that may distract the driving task [5]. Driven by the visual selective attention mechanism, experienced drivers often focus their attention on the most important regions and only show concern for objects related to driving safety in those salient regions. This selective attention mechanism [6], [7] helps drivers reduce the interference of irrelevant scene information and guarantee the driving safety. Understanding the selective attention mechanism of experienced drivers and then simulating the efficient saliency detection process in rainy conditions may help driving a car in rainy conditions as well as on a sunny day.

Visual saliency prediction is a hot topic in the field of the driving assistance technologies. It applies the intelligent algorithms to simulate human visual search patterns and extract salient areas or regions of interest in the driving scenes. Its purpose is to find the regional targets that are highly relevant to the current driving task, such as cars, pedestrians, motorcycles, bicycles, traffic lights, traffic signs, etc., in order to give drivers supplementary tips or warnings and improve the driving safety. Many algorithms have been proposed to predict the traffic saliency or drivers' attention [8], [9]. Traditional models include Itti [10], image signature [11], and hyper complex Fourier transform (HFT) [12], etc. With the development of deep neural networks, convolutional neural network (CNN) and deep learning algorithms are adopted to predict video saliency. Typical models include spatiotemporal residual attentive networks [13], skip-layer visual attention network [14], multi-level network (MLNet) [15], expandable multilayer network (EML-NET [16]), TASED-Net [17], saliency exponential moving average (SalEMA) [18], convolution-deconvolution neural network (CDNN) [19], driving video fixation prediction with spatio-temporal networks and attention gates (DSTANet) [20], semantic context induced attentive fusion network (SCAFNet) [21], etc. However, it is worth noting that none of above models are based on driving datasets collected under rainy conditions.

To evaluate the robustness of the saliency models, there are many public datasets available in the field of visual saliency detection. Static datasets include MIT [22], PASCAL-S [23] and saliency in context (SALICON) [24]. Dynamic video datasets include Hollywood-2 [25], UCF-sports [25], dynamic human fixation (DHF1K) [26], etc. Most of these datasets are related to natural images/videos about sports and life scenes. There are also a few public datasets specifically for traffic scenes, e.g., Berkeley DeepDrive attention (BDD-A) [27], DR(eye)VE [28], DADA-2000 [21] and Deng *et al.* [19]. Some datasets with synthetic images from video game [29], [30] were presented for object detection [31]. The comparing statistics of the often-used saliency/attention datasets in natural and driving scenes are summarized in Supplementary Material[1].

Considering the lack of rainy driving datasets and the shortage of investigation on saliency prediction models in rainy conditions, we collected an eye tracking dataset from 30 experienced drivers, called DrFixD(rainy). Based on the multiple drivers' attention allocation dataset, we proposed a new model based on the theory of two cortical pathways to predict the salient regions of drivers in rainy weather conditions. A CNN-based module is adopted to simulate the function of ventral pathway to identify the image features of the traffic scenes, and a long short-term memory (LSTM)-based module is applied to simulate the function of dorsal pathway to process the dynamic information between the video frames. The results indicated that our proposed model showed competitive accuracy in predicting driver's attention areas in rainy weather. The dataset and source code of our method are available[1].

Material and data collection:

Participants: 30 participants took part in the eye movement experiment, including 17 females and 13 males aged 24 to 53 years old (M = 35.8; SD = 7.5). The participants are drivers who have at least 2 years of driving experience and drove frequently. As a result, their driving experience ranges from 2 to 17 years (M = 7.5; SD = 4.1). All participants had normal or corrected-to-normal vision and were provided with written informed consent prior to participation. The experimental paradigms were approved by the Ethics and Human Participants in Research Committee at the University of Electronic Sciences and Technology of China in Chengdu, China.

Stimuli and procedure: 16 traffic driving videos in rainy conditions were collected by the driving recorders. The driving scenes are urban roads including normal streets, crossroads, overpasses, etc. Each video lasts 100−180 seconds with a resolution of 1280×720 and a frame rate of 30 frames per second. Participants were seated 57 cm away from a 21-inch CRT monitor with a spatial resolution of 1280×1027 pixels and a refresh rate of 75 Hz. Their heads were stabilized with a chin and forehead rest. A steering wheel was placed in front of the participants by assuming that they were driving a car. In order to avoid fatigue and ensure the reliability of the experimental data, the subjects completed the whole experiment in two or three days according to their mental states. The videos were divided into eight blocks, and each block consisted of two trials. There is 20 seconds blank interval between the two trials. Eye movements were recorded using an eye-tracker (Eyelink 2000, SR Research, Ottawa, Canada) with a sampling rate of 1000 Hz and a nominal spatial resolution of 0.01 degree of visual angle.

Eye-movement analysis: The procedure of eye movement data analysis is same as Deng's research [19]. The subjects' eye fixations were recorded to construct the Drivers' Fixation Dataset in rainy weather conditions (DrFixD(rainy)). In the DrFixD(rainy), there were

Corresponding author: Tao Deng.

Citation: H. Tian, T. Deng, and H. M. Yan, "Driving as well as on a sunny day? Predicting driver's fixation in rainy weather conditions via a dual-branch visual model," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 7, pp. 1335–1338, Jul. 2022.

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

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

Digital Object Identifier 10.1109/JAS.2022.105716

[1] https://github.com/taodeng/DrFixD-rainy

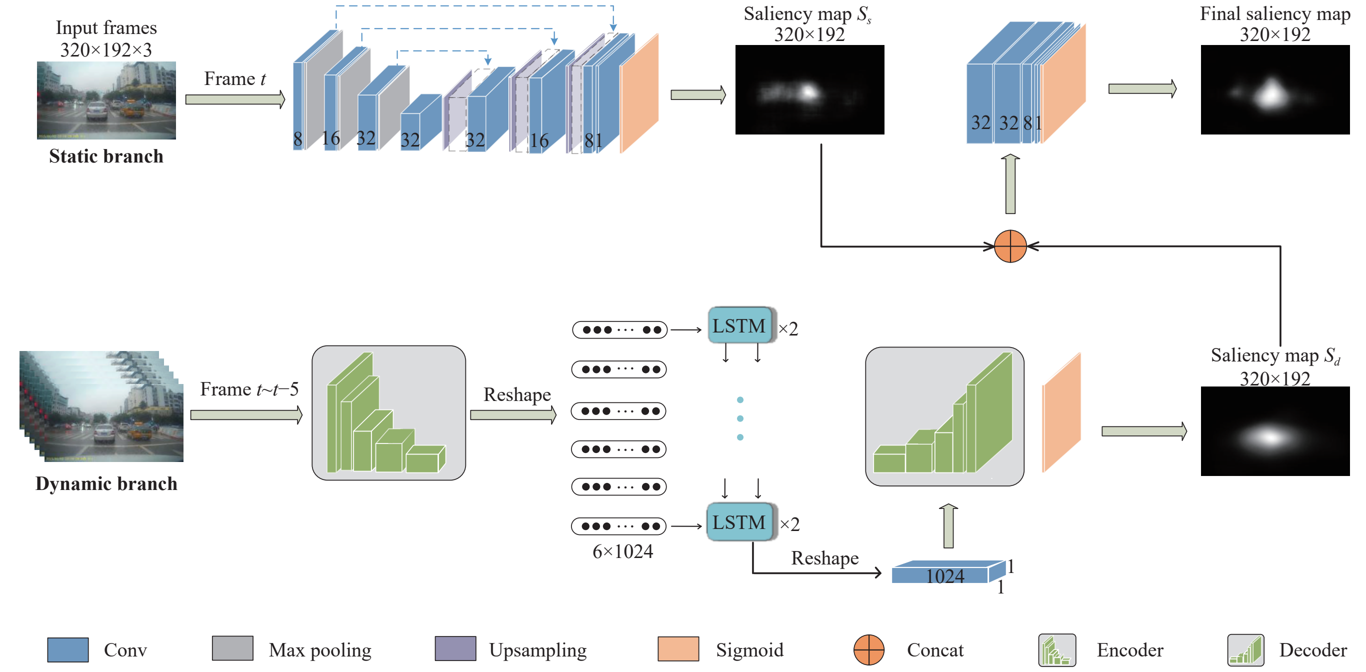

Fig. 1. The architecture of the dual-branch visual model.

30 fixation points per video frame, then all of the fixational points from the 30 participants were superimposed for each frame. The drivers' eye-tracking data fitted with a 2D Gaussian distribution were used as the ground truth in the following training and test. By taking advantage of multiple drivers' attentional experiences, the dataset may contain more than one salient region, that is, except for the primary salient region, a secondary even the tertiary region may be included if they exist.

Fixation prediction based on a dual-branch visual model: Human visual cortex is a remarkable visual information processing system. It is divided into functionally distinct areas, and each area is responsible for specific visual processing tasks. Two parallel processing pathways are commonly identified in visual cortex, one is the ventral (recognition) pathway and the other is the dorsal (motion) pathway [32]. The ventral pathway mainly identifies details such as the shape and size of objects, while the dorsal pathway is sensitive to the movement and location of objects. The two pathways work together to detect "what" and "where" the objects are quickly and accurately. Wolfe et al. also proposed a similar dual-path model from aspect of visual search in scenes: a "selective" path in which candidate objects must be individually selected for recognition and a "nonselective" path in which information can be extracted from global and/or statistical information [33].

As described above, when driving in rainy conditions, raindrops, low brightness and the movement of the wipers, etc. are the specific factors to influence the driving safety compared with good weather condition. Some factors, such as raindrops and brightness, can be treated as static image features, and some factors, such as the movement of wipers, have to be regarded as the temporal features. Therefore, considering the characteristics in rainy weather, inspired by the mechanism of the two cortical pathways [32], [33], we developed a new architecture that combines static and dynamic branches together to predict the saliency maps in the rainy traffic environment, as shown in Fig. 1 . On one hand, the CNN has excellent ability at extracting the rainy image features, so it is adopted to simulate the function of the ventral pathway to identify the traffic features in the static image saliency detection. On the other hand, the LSTM network has high capability of handling the dynamic changes of the time-dependence signals, so it is applied to simulate the function of the dorsal pathway to process the temporal information between the video frames in predicting the motion of the objects in a rainy weather condition. Thus, some missing or unnoticed information in the static branch can be made up by the dorsal (nonselective) pathway. Finally, the saliency maps of the two branches are fused together.

In the static branch, same with the model [19], a convolution-deconvolution structure is used to obtain the static features of the images. In order to reduce the amount of parameters and increase the computation speed, we resized the image to 320×192×3. The convolution path follows the typical architecture of a convolution network, namely, two 3×3 convolutions are included, and each convolution is followed by batch normalization (BN) and rectified linear unit (ReLU), then a 2×2 max-pooling operation with a stride of 2 for down-sampling. The deconvolution path consists of an up-sampling of the feature map, a concatenation with the corresponding feature map from the convolution path, and two 3×3 convolutions, each followed by a BN and ReLU. For each frame image, a static saliency map S_s with a size of 320×192×1 was obtained.

In the dynamic branch, we aim to obtain the temporal correlation between adjacent frames. Due to the bottom-up relationships of continuous video frames and the top-down pre-attention mechanism, the saliency of previous frames may have an influence on that of subsequent frames, so we expect to utilize the temporal correlation to improve the accuracy of the prediction. In the dynamic branch, the current frame and its previous five frames $F_{t\sim t-5}$ are packed into a continuous sequence as input, and the feature vector (6×1024) is obtained after passing through the Encoder module, corresponding to the six frames of the input sequence. After that, the feature vector is sent to the LSTM and the output (1×1×1024) is extracted as the motion information at time $t-5$. Then, going through the Decoder module, the dynamic saliency map S_d (320×192) is finally obtained. Specifically, the 1×1×1024 feature map is decoded as 5×3×512 feature map by a 2D transpose convolution operator firstly. Then, the 5×3×512 feature map is decoded as 320×192×1 by four upsampling and deconvolution operations. The size of feature map changes to (10×6×512), (20×12×256), (40×24×128), (80×48×16), (320×192×8) in turn. Finally, the 320×192×8 feature map is convolved as 320×192×1 by a 1×1 convolution and Sigmoid activation function.

To make full use of the temporal information as well as avoid attention shift within the period, 6 frames is chosen as the consecutive input of LSTM. According to the cognitive neuroscience researches, approximately 200 ms is required if attention shifts from one item to the next in visual search tasks [34], [35], which means that driver's attention usually remains stable within 200 ms. Therefore, we set the input length as 6 frames because the temporal course of 6 frames is 200 ms (the frame rate of our video is 30 frames per second).

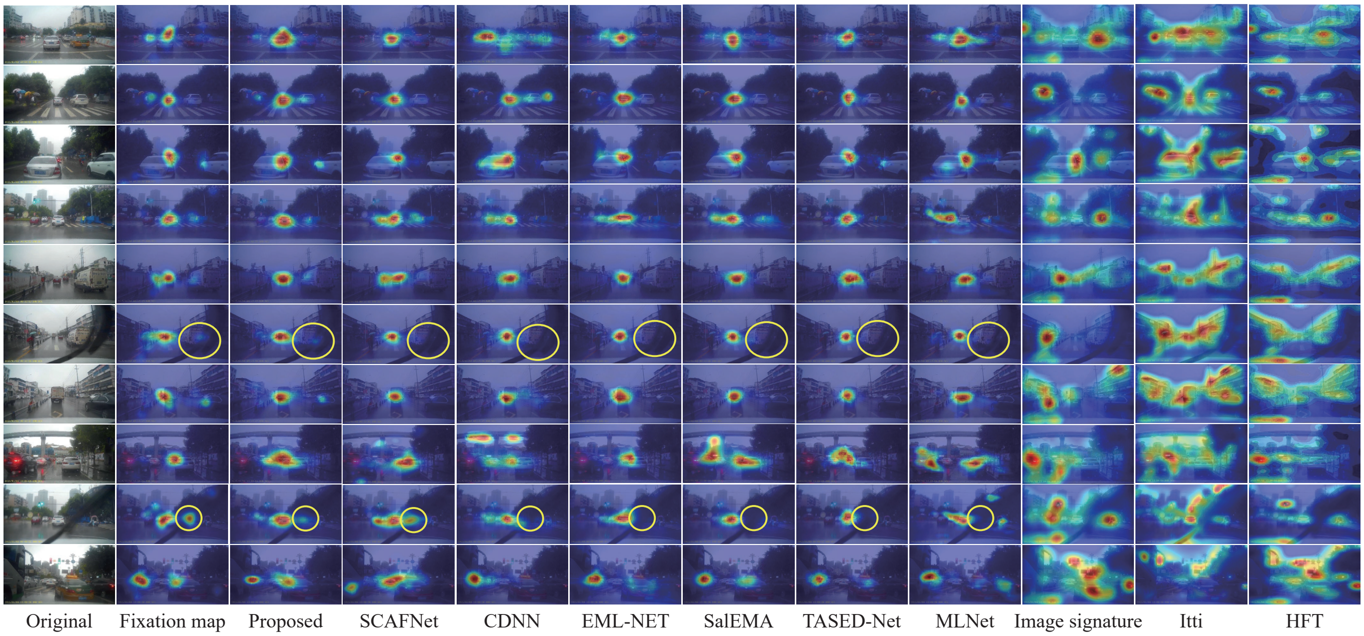

Fig. 2. Saliency maps selected from the four test videos randomly. The maps were yielded by our model and nine other methods. The yellow circles indicate the location of the truck (sixth row) and a crossing pedestrian (ninth row).

Finally, to integrate the static and dynamic feature information together, we concatenate the static saliency map S_s and the dynamic saliency map S_d, then a convolutional operation is followed to reduce the image dimension. The final saliency map is a 320×192 grayscale image. In the training process, the binary cross entropy (BCE) is used as the loss function to evaluate the difference between the label and the predicted result and minimize the difference.

Results: The sixteen videos in DrFixD(rainy) are randomly divided into three subsets. Ten videos are used as the training set, two videos are used as the validating set, and four videos are used as the testing set. All of these videos are untrimmed. In total, there are 52 291 frames in the training phase, 9816 frames in the validating phase, and 19 106 frames in the testing phase.

In order to illustrate the performance of the proposed model, we compared it with nine other models, including six deep learning and three non-deep learning methods. Among them, Itti [10], image signature [11], and HFT [12] are non-deep learning methods. MLNet [15], TASED-Net [17], SalEMA [18], EML-NET [16], CDNN [19] and SCAFNet [21] are deep learning-based methods. Note that, all the models have been retrained on our dataset, and the best parameters and results are saved.

Qualitative evaluation: Fig. 2 shows an intuitive result of the dual-branch visual model in predicting the drivers' salient regions in rainy weather conditions. From Fig. 2, we can see that the proposed model can predict the driver's gaze areas accurately. The last three columns are saliency maps by classical non-deep learning methods, which show a higher false detection results. On the contrary, our model combines bottom-up features and top-down dynamic prediction together to simulate the visual selective attention, therefore, it matches the driver's attention better and obtains the significant areas that are more relevant to the driving task.

In addition, the proposed model is also better than the other six deep learning-based methods. As shown in Fig. 2, e.g., for the third, sixth, eighth and tenth rows, when the scenes are complex or there are multiple targets, our model can predict the drivers' potential attention allocation accurately for the main target/region as well as the secondary one if it exists, which is consistent with the drivers' driving experience. Especially for the sixth and ninth rows, when the wiper occludes part of the truck (sixth row of Fig. 2) and a crossing pedestrian (ninth row of Fig. 2) labeled by the yellow circles, our model could still detect them while it is difficult for other deep learning-based state-of-the-art (SOTA) models to achieve this, indicating that our model shows excellent performance in an occlusion situation of wipers in rainy condition. Furthermore, the prediction results of the eighth and tenth rows in Fig. 2 present that our model can detect driver's fixational regions more efficiently than other models in the crowded driving scenes under rainy conditions.

Quantitative evaluation: For quantitative comparison, six classic metrics are used to evaluate the models, including the area under the ROC curve (AUC_Borji, AUC_Judd), normalized scanpath saliency (NSS), pearson's correlation coefficient (CC), similarity (SIM) and Kullback-Leibler divergence (KLD). Among them, AUC_Borji, AUC_Judd and NSS are location-based metrics, and CC, SIM and KLD are distribution-based metrics [19]. Table 1 shows the quantitative performance of our proposed model compared with other models. Note that, similar to the evaluation of baseline (infinite humans) on MIT300 [22], we convert the human drivers' fixation point maps into saliency maps with a 2D Gaussian distribution, which are used to calculate the standard of human driver. In this case, the human driver's AUC_Borji and AUC_Judd evaluation scores may not be 1, because the AUC calculates the similarity of fixation point and saliency map, as well as the differences of observers. The first row represents the drivers' fixations (ground truth). If the value of a model is closer to that of human, the model shows better performance. As expected, our proposed model (last row of Table 1) predicts the fixations more accurately than other models.

Ablation study: In order to validate the effectiveness of the dual-branch model, we compared its performance with that of each individual branch. The quantitative evaluation is shown in Table 2. The result indicates that the combination of the two branches shows better saliency prediction than individual one. Under a test computing condition of NVIDIA TITAN RTX 24GB GPU, the speed of saliency prediction achieves 289.39, 50.89 and 50.27 fps for static, dynamic and dual-branch model respectively, which can basically meet the real-time demand. Besides, we have done a further validation of our model on the public DR(eye)VE-rainy [28] and BDD-A-rainy [27] datasets in Supplementary Material[1].

Conclusions: In this work, we built up a rainy traffic video dataset DrFixD(rainy) containing eye tracking information from thirty drivers, which contributes to the researches on the traffic saliency in rainy weather condition. Further, we proposed a two-branch saliency model based on the theory of two cortical pathways to predict the driver's fixation in rainy weather conditions. The result shows that our model can predict the drivers' potential attention allocation of the

Table 1. Performance Comparison of Our Model With the State-of-the-Art Saliency Models Using Multiple Evaluation Metrics on DrFixD (rainy)

Models	AUC_ Borji↑	AUC_ Judd↑	NSS↑	CC↑	SIM↑	KLD↓
Human	0.9620	0.9822	5.4196	1	1	0
Itti	0.8426	0.857	1.5569	0.3695	0.2702	1.6868
ImageSig	0.6368	0.6912	0.5298	0.1404	0.1781	2.4159
HFT	0.7606	0.7867	1.0077	0.2311	0.224	2.0079
MLNet	0.8942	0.928	3.8999	0.7944	0.6282	3.6946
TASED-Net	0.8773	0.9471	4.2118	0.8393	0.5877	0.8451
SalEMA	0.8969	0.9536	4.1144	0.8465	0.6650	0.4734
EML-NET	0.8907	0.9462	**4.2951**	0.8512	0.5531	0.6968
CDNN	0.9001	0.9516	4.1069	0.8214	0.6339	0.5222
SCAFNet	0.8952	0.9416	4.1742	0.8351	0.6650	1.8686
Proposed	**0.9066**	**0.9555**	4.1902	**0.8534**	**0.6664**	**0.4670**

Table 2. Ablation Study on Dual-Branch Model on DrFixD (rainy)

Models	AUC_ Borji↑	AUC_ Judd↑	NSS↑	CC↑	SIM↑	KLD↓
Static	0.8925	0.9491	4.0049	0.8134	0.6393	0.5512
Dynamic	0.8980	0.9492	4.0308	0.8314	0.6414	0.5208
Static+ dynamic	**0.9066**	**0.9555**	**4.1902**	**0.8534**	**0.6664**	**0.4670**

main target or region as well as the secondary/tertiary ones if they exist.

Acknowledgment: This work was supported by the National Natural Science Foundation of China (NSFC) (61773094, 62076055, 62106208), the Science and Technology Project of Sichuan Province, (2020JDRC0031), and the China Postdoctoral Science Foundation (2021TQ0272, 2021M702715).

References

[1] S. Jung, K. Jang, Y. Yoon, and S. Kang, " Contributing factors to vehicle to vehicle crash frequency and severity under rainfall," *J. Safety Research*, vol. 50, pp. 1–10, 2014.

[2] S. Jung, X. Qin, and D. A. Noyce, "Rainfall effect on single-vehicle crash severities using polychotomous response models," *Accident Analysis & Prevention*, vol. 42, no. 1, pp. 213–224, 2010.

[3] A. Drosu, C. Cofaru, and M. V. Popescu, "Fatal injury risk model (firm) of the road accidents that occurred in rainy conditions—a probabilistic approach," *Intern. J. Auto. Techn.*, vol. 22, no. 5, pp. 1415–1426, 2021.

[4] 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. DOI: 10.3390/su12197976.

[5] A. Jain, H. S. Koppula, B. Raghavan, S. Soh, and A. Saxena, "Car that knows before you do: Anticipating maneuvers via learning temporal driving models," in *Proc. IEEE Intern. Conf. Comp. Vis.*, 2015.

[6] M. Corbetta, F. Miezin, S. Dobmeyer, G. Shulman, and S. Petersen, "Attentional modulation of neural processing of shape, color, and velocity in humans," *Science*, vol. 248, no. 4962, pp. 1556–1559, 1990.

[7] T. S. Lee, "Computations in the early visual cortex," *J. Physiol. Paris.*, vol. 97, no. 2–3, pp. 121–139, 2003.

[8] T. Deng, K. Yang, Y. Li, and H. Yan, "Where does the driver look? Top-down-based saliency detection in a traffic driving environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 7, pp. 2051–2062, 2016.

[9] T. Deng, H. Yan, and Y.-J. Li, "Learning to boost bottom-up fixation prediction in driving environments via random forest," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 9, pp. 3059–3067, 2018.

[10] L. Itti, C. Koch, and E. Niebur, " A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Patt. Analys. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[11] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Trans. Patt. Analysis Mach. Intell.*, vol. 34, no. 1, p. 194, 2012. DOI: 10.1109/TPAMI.2011.146.

[12] J. Li, M. D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Trans. Patte. Analys. Mach. Intell.*, vol. 35, no. 4, pp. 996–1010, 2013.

[13] Q. Lai, W. Wang, H. Sun, and J. Shen, "Video saliency prediction using spatiotemporal residual attentive networks," *IEEE Trans. Image Proc.*, vol. 29, pp. 1113–1126, 2020.

[14] W. Wang and J. Shen, "Deep visual attention prediction," *IEEE Trans. Image Proc.*, vol. 27, no. 5, pp. 2368–2378, 2018.

[15] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara, "A deep multi-level network for saliency prediction," in *Proc. IEEE Int. Conf. Patt. Rec.*, 2016, pp. 3488–3493.

[16] S. Jia and N. D. Bruce, "EML-Net: An expandable multi-layer network for saliency prediction," *Image Vis. Comp.*, vol. 95, p. 103887, 2020. DOI: 10.1016/j.imavis.2020.103887.

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

[18] 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 *Proc. British Machine Vision Conf.*, BMVA Press, 2019, p. 182.

[19] 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.

[20] T. Deng, F. Yan, and H. Yan, "Driving video fixation prediction model via spatio-temporal networks and attention gates," in *Proc. IEEE Intern. Conf. Multim. Expo*. 2021, pp. 1–6.

[21] 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. DOI: 10.1109/TITS.2020.3044678.

[22] Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba, " Mit saliency benchmark," 2015, Available: http://saliency.mit.edu/. Accessed on: Sept. 2021.

[23] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille, "The secrets of salient object segmentation," in *Proc. IEEE Conf. Comp. Vis. Patt. Rec.*, 2014, pp. 280–287.

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

[25] S. Mathe and C. Sminchisescu, "Dynamic eye movement datasets and learnt saliency models for visual action recognition," in *Proc. Eur. Conf. Comp. Vis.*, 2012.

[26] 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.

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

[28] 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, 2019.

[29] D. Liu, Y. Wang, K. E. Ho, Z. Chu, and E. Matson, " Virtual world bridges the real challenge: Automated data generation for autonomous driving," in *Proc. IEEE Intell. Veh. Symp.*, 2019, pp. 159–164.

[30] D. Liu, Y. Cui, Z. Cao, and Y. Chen, "A large-scale simulation dataset: Boost the detection accuracy for special weather conditions," in *Proc. IEEE Intern. Joint Conf. Neural Netw.*, 2020, pp. 1–8.

[31] D. Liu, Y. Cui, Y. Chen, J. Zhang, and B. Fan, "Video object detection for autonomous driving: Motion-aid feature calibration," *Neurocomputing*, vol. 409, pp. 1–11, 2020.

[32] M. Mishkin, L. G. Ungerleider, and K. A. Macko, "Object vision and spatial vision: Two cortical pathways," *Trends in Neur.*, vol. 6, no. 10, pp. 414–417, 1983.

[33] J. M. Wolfe, M. L.-H. Võ, K. K. Evans, and M. R. Greene, "Visual search in scenes involves selective and nonselective pathways," *Trends Cogn. Sci.*, vol. 15, no. 2, pp. 77–84, 2011.

[34] S. J. Luck and S. A, " Hillyard.Spatial filtering during visual search: Evidence from human electrophysiology," *Journ. Exper. Psych.: Human Perc. Perf.*, vol. 20, no. 5, pp. 1000–1014, 1994.

[35] G. F. Woodman and S. J. Luck, "Serial deployment of attention during visual search," *J. Exper. Psych.: Human Perc. Perf.*, vol. 29, no. 1, pp. 121–138, 2003.