

Pose-aware Multi-feature Fusion Network for Driver Distraction Recognition

Mingyan Wu^{1, 2}, Xi Zhang^{1, 2}, Linlin Shen^{1, 2*}, Hang Yu³

¹Computer Vision Institute, College of Computer Science and Software Engineering ²Guangdong Key Laboratory of Intelligent Information Processing ³Shenzhen Laboratory of IC Design for Internet of Things

E Abstract

We propose a novel multi-feature fusion network based on pose estimation, for image based distracted driving detection. Since hand is the most important part of driver to infer the distracted actions, our proposed method firstly detects hands using the human body posture information. In addition to the features extracted from the whole image, our network also include the important information of hand and body posture. The global feature, hand and pose features are finally fused by concatenation of feature maps. The experimental results show that our method achieves state-of-the-art performance on SZ Bus Driver dataset and AUC dataset.

Motivation



- · The whole original image contains global information.
- · The actions of hand are important cues in driver distraction recognition.
- The pose information is robust against the interference of backgrounds.

Datasets



Fig: Examples of SZ Bus Driver





(a) C6: Drink



(f) C5: Adjust radio

(b) C7: Reaching behind Fig: Examples of AUC Distracted Driver dataset.

(i) C8: Hair and makeun

(i) C9: Talk to passang



Fig: Pose-aware multi-feature fusion network.

Experimental Results

Method

۲

e

Table: Result on SZ Bus Driver dataset Feature Backbone C0 C1 C2 C3 Total VGG-16 74.04 83.01 41.78 85.80 73.35 Global ResNet-50 87.11 47.89 48.85 83.48 75.70 InceptionV3 83.19 82.14 22.48 90.53 77.28 VGG-16 90.84 84.55 72.25 92.45 88.78 Late ResNet-50 93.12 92.78 75.36 91.14 90.87 Fusion InceptionV3 95.85 90.63 80.87 92.74 92.93 VGG-16 94.13 95.38 82.35 66.16 91.09 Early ResNet-50 95.43 99.46 64.13 94.76 92.58 Fusion InceptionV3 96.46 97.66 89.40 95.27 95.75

Table: Result on AUC V1 and V2 dataset

	Dataset	Method	Accuracy
	AUC V1	GA-Weighted Ensemble (2017)	95.98
		DenseNet+Latent Pose (2018)	94.20
		VGG with Regularization (2018)	96.31
		I3D-two stream (2019)	77.10
		AlexNet+HOG features (2019)	93.19
		Our method	96.28
	AUC	GA-Weighted Ensemble (2019)	90.07
	V2	Our method	90.38



Fig: The confusion matrix of three datasets.

Ablation Study



Visualization



Fig: The t-SNE visualization of global feature for SZ dataset (a) and AUC V2 dataset (c) and that of fused feature for SZ dataset (b) and AUC V2 dataset (d).



(a) C0: Playing the phone (b) C1: Talking on the phone (c) C2: Smoking





Fig: The saliency maps on AUC Distracted Driver dataset.