2021 ICCV OCTOBER 11-17 VIRTUAL

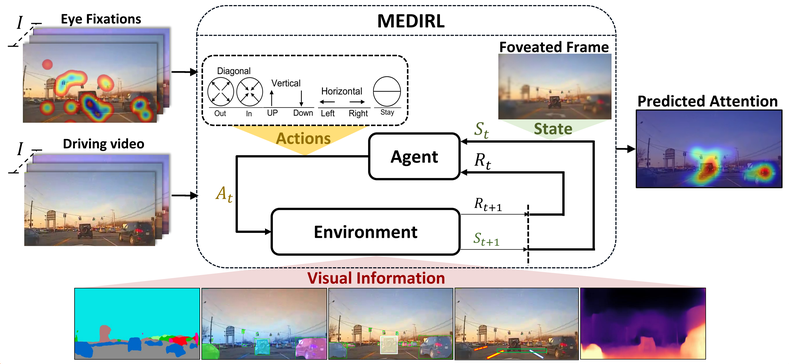
MEDIRL: Predicting the Visual Attention of Drivers via Maximum Entropy Deep Inverse Reinforcement Learning

Sonia Baee[1], Erfan Pakdamanian[1], Inki Kim[2], Lu Feng[1], Vicente Ordonez[3], Laura Barnes[1]

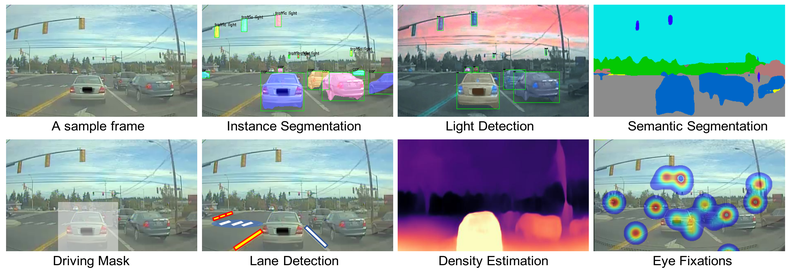
UNIVERSITY of VIRGINIA leidos NSF

Contribution at a Glance

- We propose **MEDIRL**, a novel IRL formulation for predicting driver visual attention in accident-prone situations.
- We introduce **EyeCar**, a new driver attention dataset comprised of rear-end collisions videos for the goal-directed attention problem in critical driving situations.
- Extensive experimental evaluations show that **MEDIRL** outperforms existing models for attention prediction and achieves state-of-the-art performance.

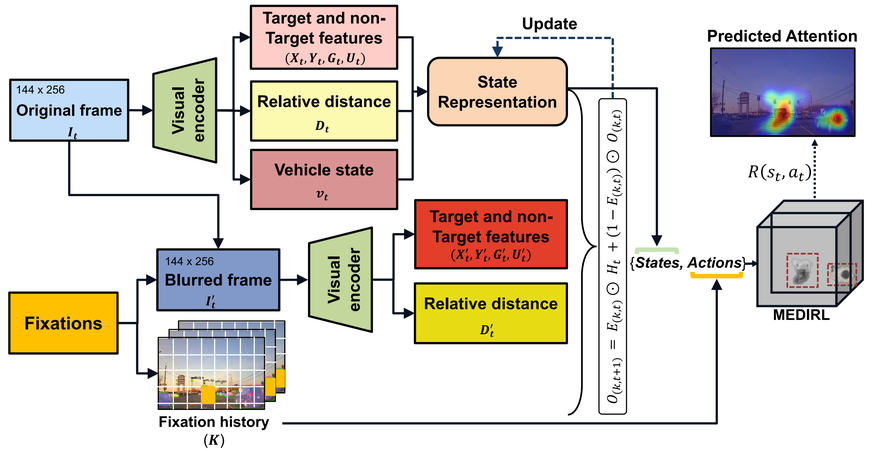


EyeCar Dataset



- **EyeCar** is the only dataset includes collisions from a driver point-of-view.
- **EyeCar** contains 3.5 hours of gaze behavior (aggregated and raw) captured from more than 315,000 rear-end collisions video frames.
- Each frame comprises 4.6 vehicles on average which makes **EyeCar** driving scenes more complex than other visual attention datasets.

MEDIRL

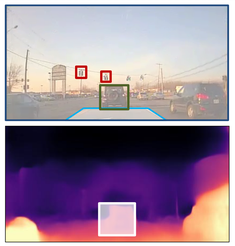


State Dynamics

Collecting the context of spatial cues

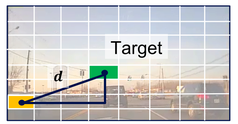
✓ Encode target and non-target visual features

Driving tasks and vehicle state



✓ lane-keeping, merging-in, braking
✓ lead vehicle speed
✓ relative distance

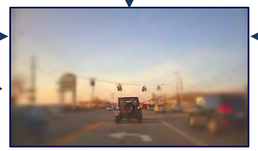
Low-resolution frame



Fixation point (k_i)

✓ Low-resolution frame for each fixation point

Foveated Frame



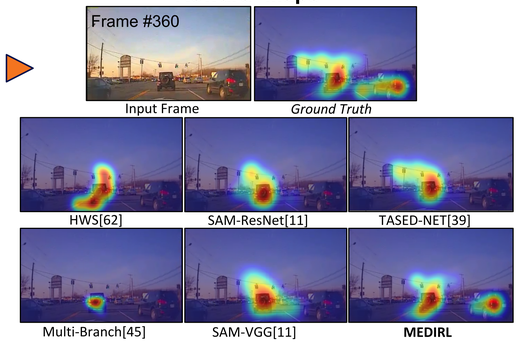
Results

Data	Method \ Task	Merging-in			Lane-keeping			Braking		
		CC	$s-AUC$	KLD	CC	$s-AUC$	KLD	CC	$s-AUC$	KLD
BDD-A	Multi-branch[45]	0.46	0.48	4.42	0.51	0.61	3.57	0.61	0.64	3.08
	HWS [62]	0.41	0.47	4.36	0.69	0.81	3.55	0.67	0.68	2.86
	SAM-ResNet[11]	0.55	0.48	3.85	0.85	0.72	3.29	0.79	0.74	2.46
	SAM-VGG[11]	0.53	**0.49**	3.92	0.84	0.70	3.22	0.77	0.70	2.49
	TASED-NET[39]	0.55	**0.49**	3.78	0.84	0.71	3.12	0.77	0.76	2.47
	MEDIRL	**0.58**	**0.49**	**2.81**	**0.86**	**0.73**	**2.43**	**0.79**	**0.81**	**2.30**

▲ Models trained on BDD-A train sets and tested on EyeCar.

MEDIRL reliably captures the important visual cues in a braking task of a complex frame.

Reward Maps



▼ **Ablation Studies**

Ablated version of *MEDIRL* \ Dataset	EyeCar		
	CC	KLD	f_β
global image + IRL	0.18	4.21	0.10
non-target + IRL	0.19	4.15	0.12
target + non-target + IRL	0.29	3.51	0.18
target + non-target + distance + IRL	0.30	3.62	0.19
lead vehicle + lane + IRL	0.30	3.57	0.23
target + non-target + lane + lead vehicle + IRL	0.36	3.53	0.21
target + non-target + distance + lane + lead vehicle + IRL	0.33	3.43	0.26
target + non-target + distance + lane + driving task + IRL	0.51	3.41	0.31
target + non-target + distance + lead vehicle + driving task + IRL	0.66	2.91	0.49
target + non-target + distance + lane + lead vehicle + driving task + IRL	0.70	2.78	0.60
***MEDIRL**: target+non-target+distance+lane+lead vehicle+driving task+speed+IRL*	**0.74**	**2.51**	**0.61**

Main Insights

- **MEDIRL** effectively learns to model the fixation selection as a sequence of states and actions.
- **MEDIRL** predicts a maximally-rewarding fixation location by perceptually parsing a scene and accumulating a sequence of visual cues

[1]University of Virginia, [2]University of Illinois at Urbana Champaign, [3]Rice University