PREDICTING THE DRIVER'S FOCUS OF ATTENTION: THE DR(EYE)VE PROJECT

Andrea Palazzi, Davide Abati, Simone Calderara, Francesco Solera, Rita Cucchiara

(name.surname@unimore.it)

Department of Engineering, University of Modena and Reggio Emilia, Italy

AImage Lab

UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

MOTIVATION

Autonomous cars promise to revolutionize the whole transport system.

… yet, when this will happen is still unclear.

What happens in the meanwhile?

DRIVER DISTRACTION is an important safety problem. Between 13% and 50% of crashes are attributed to driver distraction, resulting in as many as 5000 fatalities and $40 billion in damages each year [1]–[3]. Increasing use of

prompt safe decisions about driving maneuvers. Every year, traffic accidents result in approximately 1.2 million fatalities worldwide; without novel prevention measures, this number could increase by 65% over the next two decades [2]. In the U.S. alone, more than 43 000 fatalities are projected this year due to traffic accidents, with up to 80% of them due to driver inattention [3], [4]. To counter the effect of inattention,

and growing problem with global dimensions. A recent study by World Health Organization mentions that annually, over 1.2 million fatalities and over 20 million serious injuries occur worldwide [1]. Enhancement of traffic safety is pursued

is due to drivers with a diminished vigilance level. In the trucking industry, 57% of fatal truck accidents are due to driver fatigue. It is the number one cause of heavy truck crashes. Seventy percent of American drivers report driving fatigued.

Human error is the main cause of more than 90 percent

MOTIVATION

Most of cars accidents are still caused by human factors (i.e. distraction)

Goal: investigating human focus of attention (FoA) during the driving task.

Not an easy task due to the lack of [public] datasets.

Dataset	Frames	Drivers	Scenarios	Annotations	Real-world	Public
Pugeault *et al.*	158,668	n.d.	Countryside, Highway Downtown	9 classes in Environment Road, Junction, Attributes	Yes	No
Simon *et al.*	40	30	Downtown	Gaze Maps	No	No
Underwood *et al.*	120	77	Urban Motorway	n.d.	No	No
Fridman *et al.*	1,860,761	50	Highway	6 Gaze Location Classes	Yes	No
Dr(eye)ve	555,000	8	Countryside, Highway Downtown	Gaze Maps	Yes	Yes

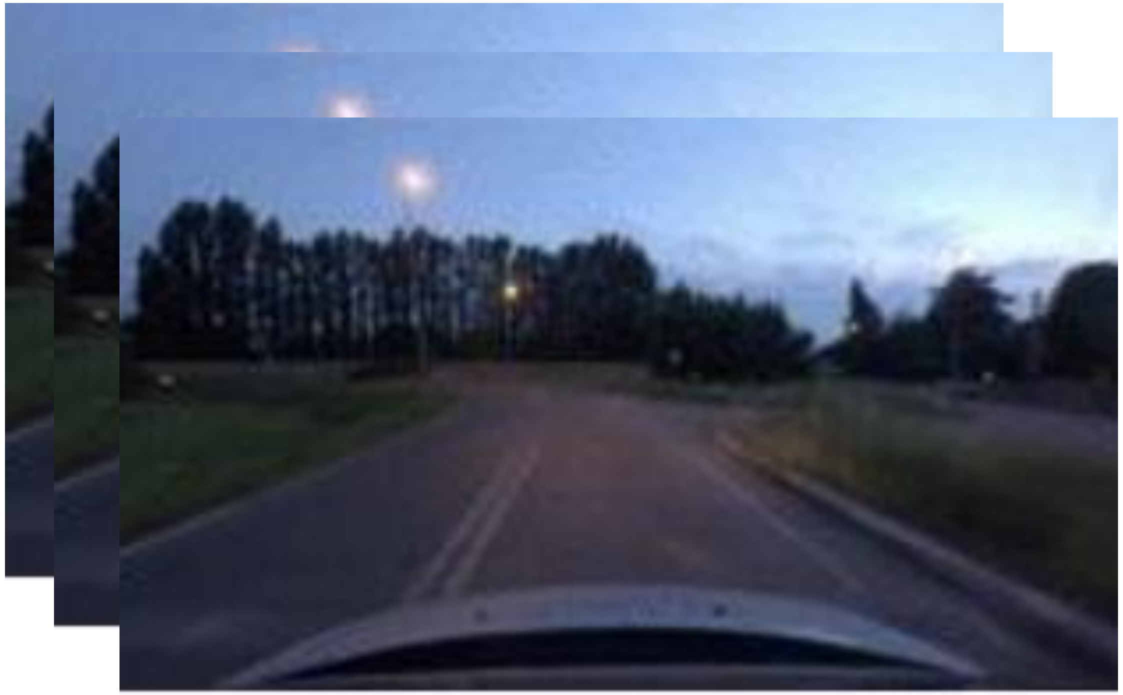
N. Pugeault and R. Bowden. How much of driving is preattentive? IEEE Transactions on Vehicular Technology, Dec 2015
L. Simon, J. P. Tarel, and R. Bremond. Alerting the drivers about road signs with poor visual saliency. In Intelligent Vehicles Symposium, June 2009.
G. Underwood, K. Humphrey, and E. van Loon. Decisions about objects in real-world scenes are influenced by visual saliency before and during their inspection. Vision Research, 2011
L. Fridman, P. Langhans, J. Lee, and B. Reimer. Driver gaze region estimation without use of eye movement. IEEE Intelligent Systems, 2016

DATA COLLECTION

DATA COLLECTION: ACQUISITION RIG

First-person camera: eye tracker SMI ETG HD camera 720p/30fps

Dashboard camera: Garmin VirbX 1080p/25fps, embedded GPS

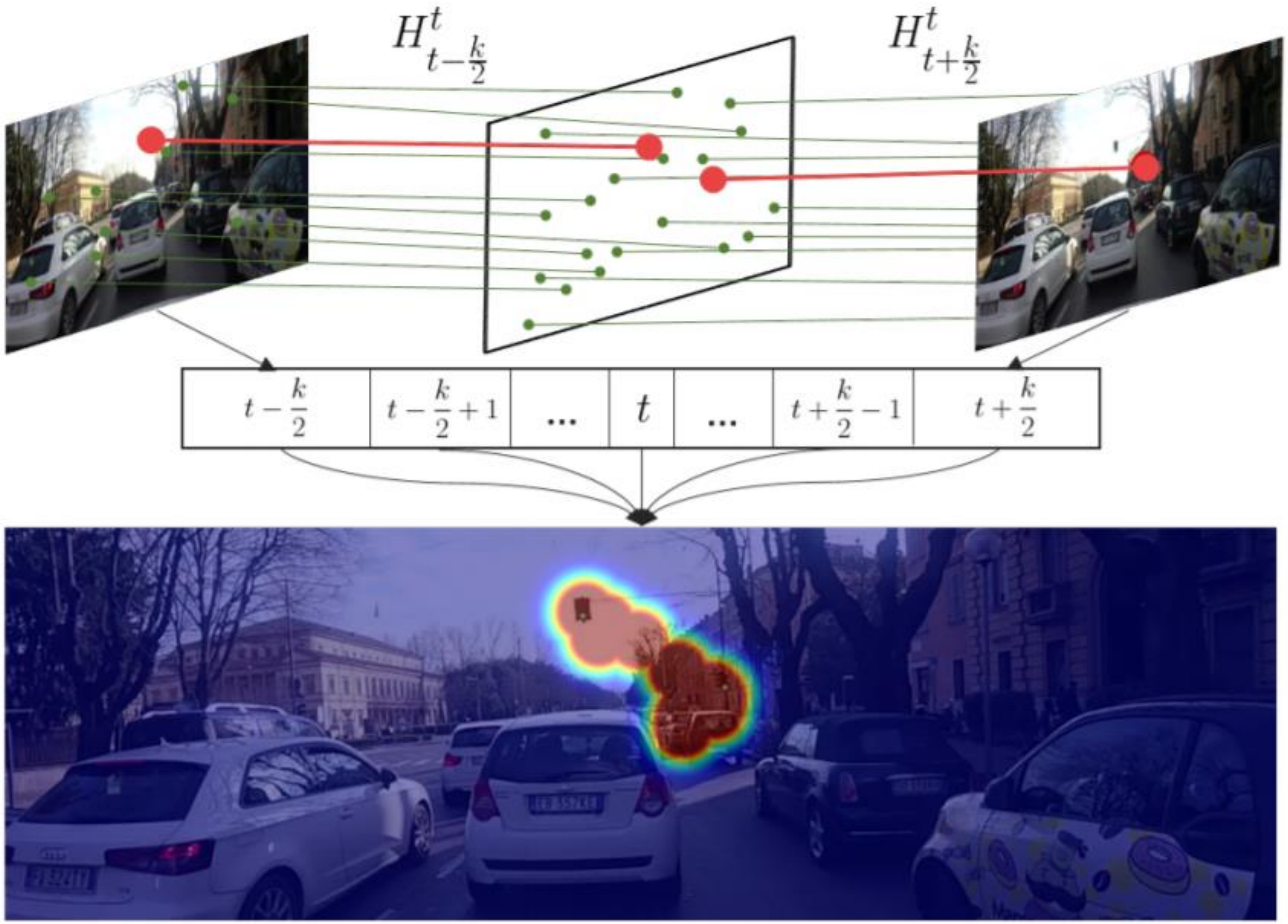


DATA COLLECTION: OVERVIEW

- 8 different drivers
- 3 different landscapes
  {Highway, Countryside, Downtown}
- 3 different weather conditions:
  {Sunny, Cloudy, Rainy}
- 3 different lighting conditions:
  {Morning, Evening, Night}

- 74 videos of 5 minutes each
- 555 000 annotated frames

DATA COLLECTION: TEMPORAL INTEGRATION (1/2)



Integrate over 25 consecutive frames (~ 1 second)

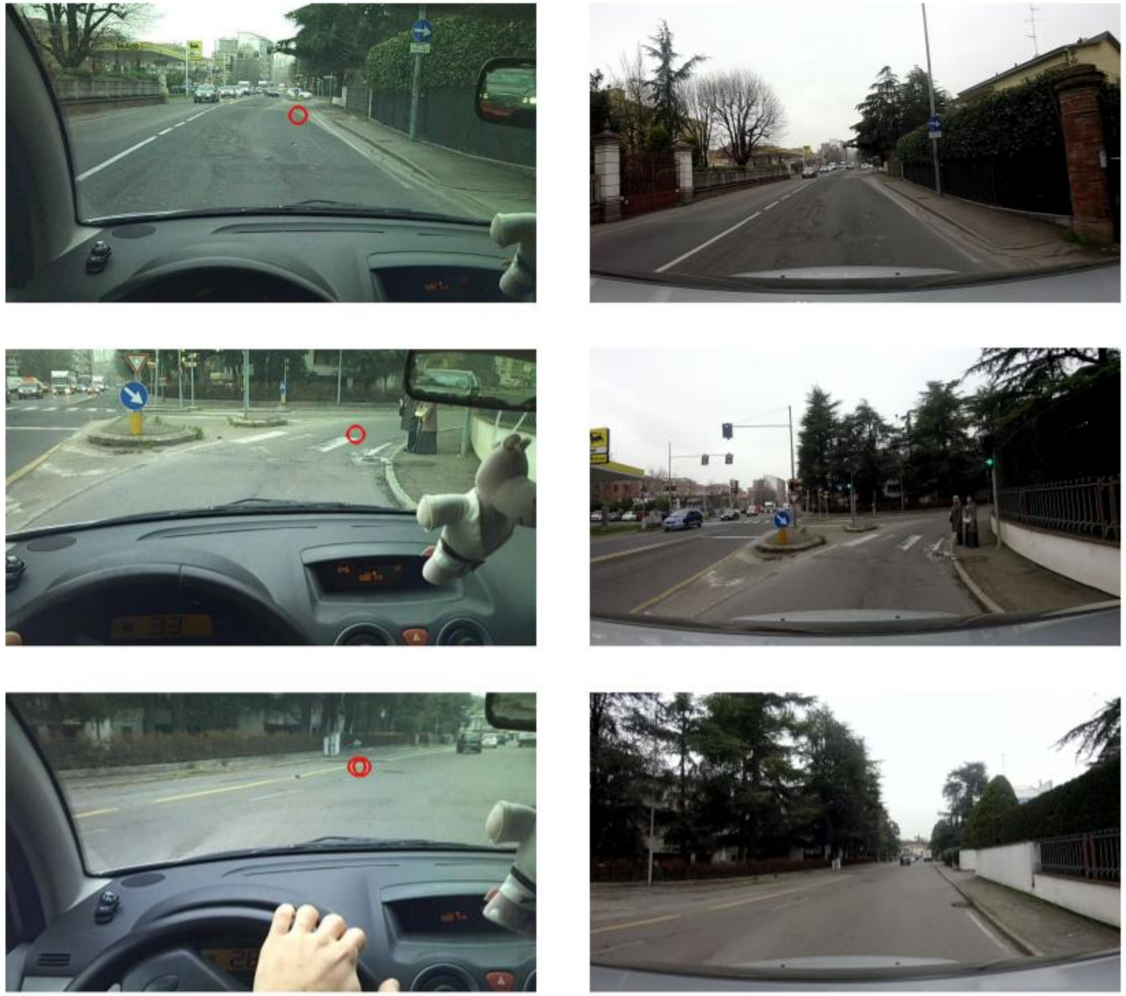
Eliminate scanpath subjectivity

DATA COLLECTION: TEMPORAL INTEGRATION (2/2)

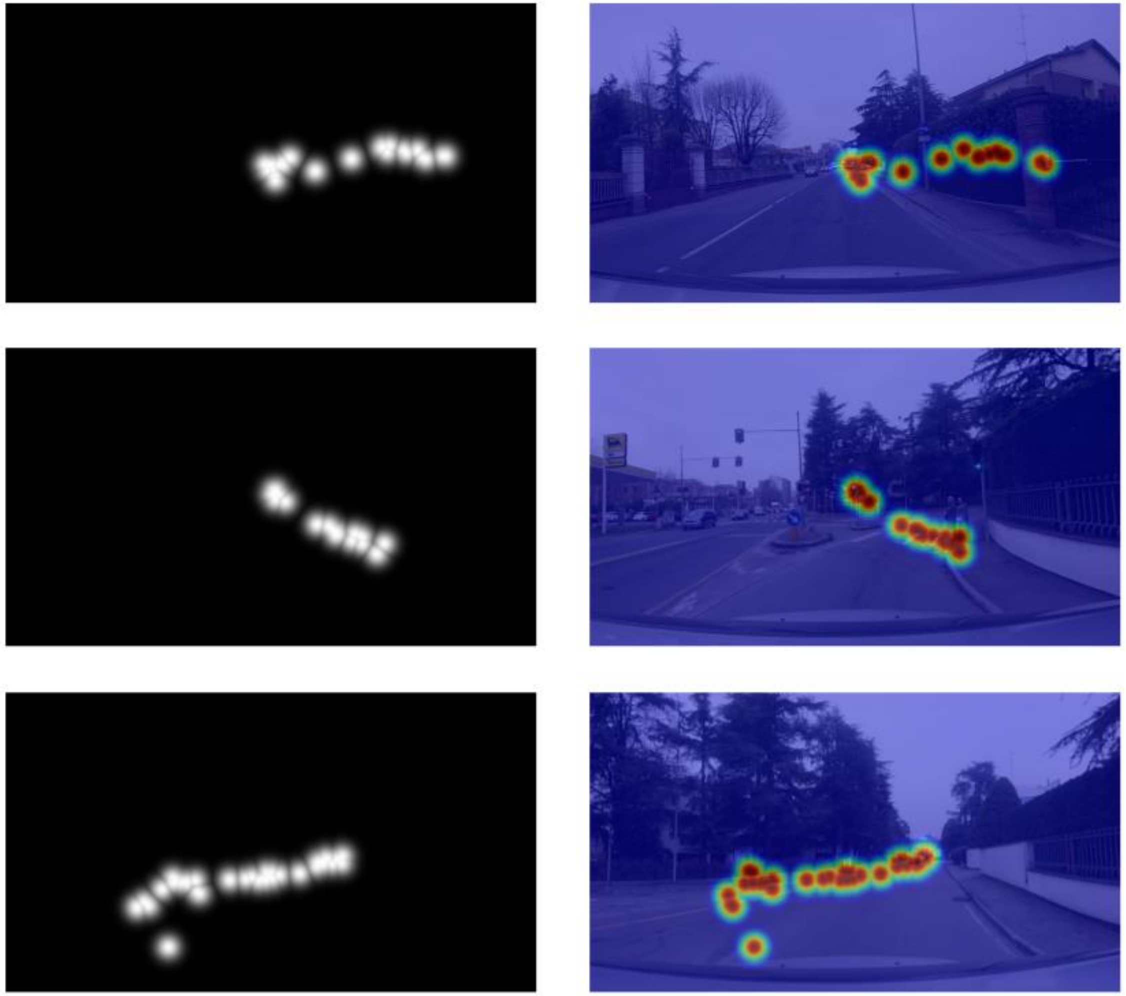
DATASET ANALYSIS

DATASET ANALYSIS: DATA OVERVIEW

X: frames from first-person and dashboard camera

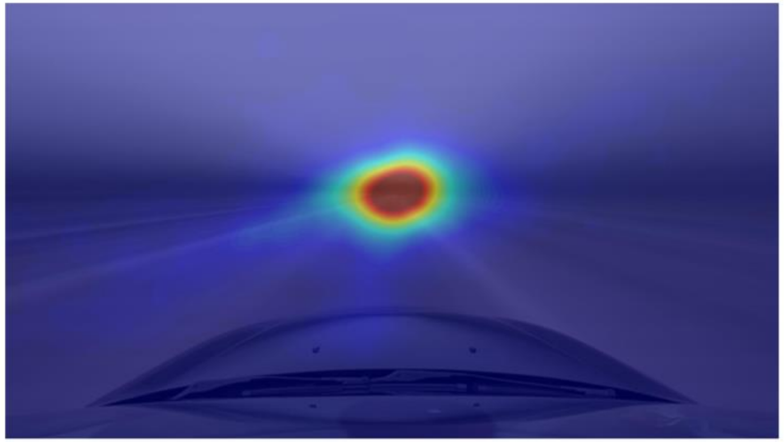


Y: driver’s fixation maps on the scene



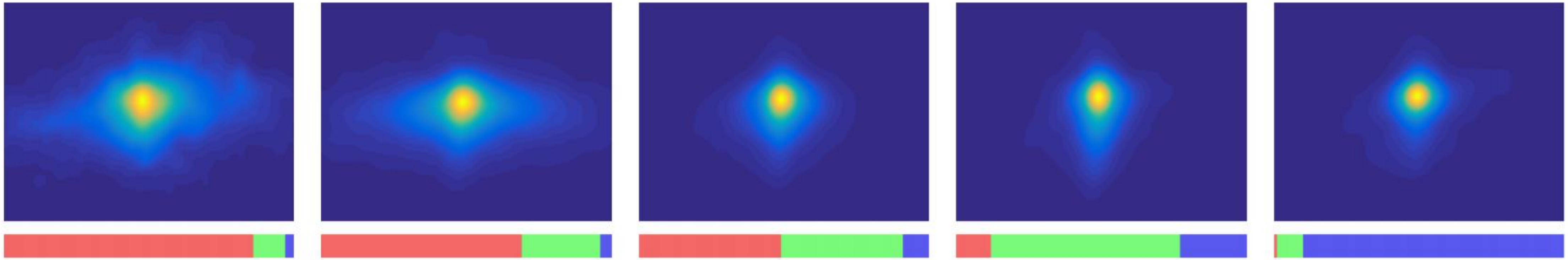
DATASET ANALYSIS: MOTION

5 minutes of highway driving *(mean frame)*



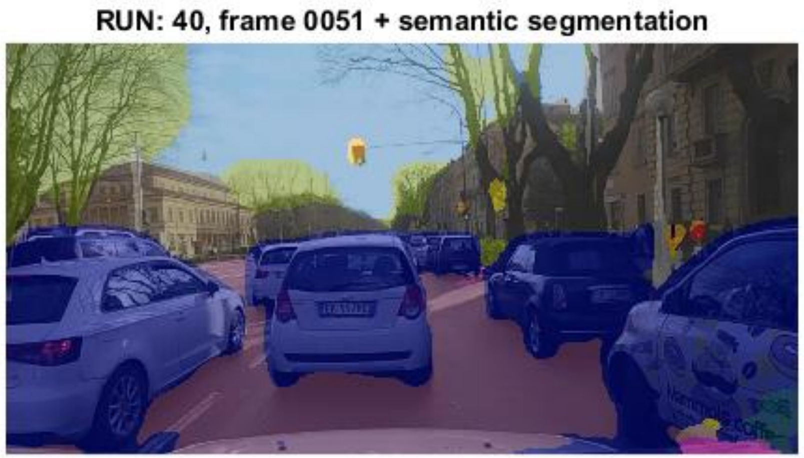
5 minutes of highway driving *(mean fixation)*

Strong bias towards the vanishing point of the road ($\propto$ speed)

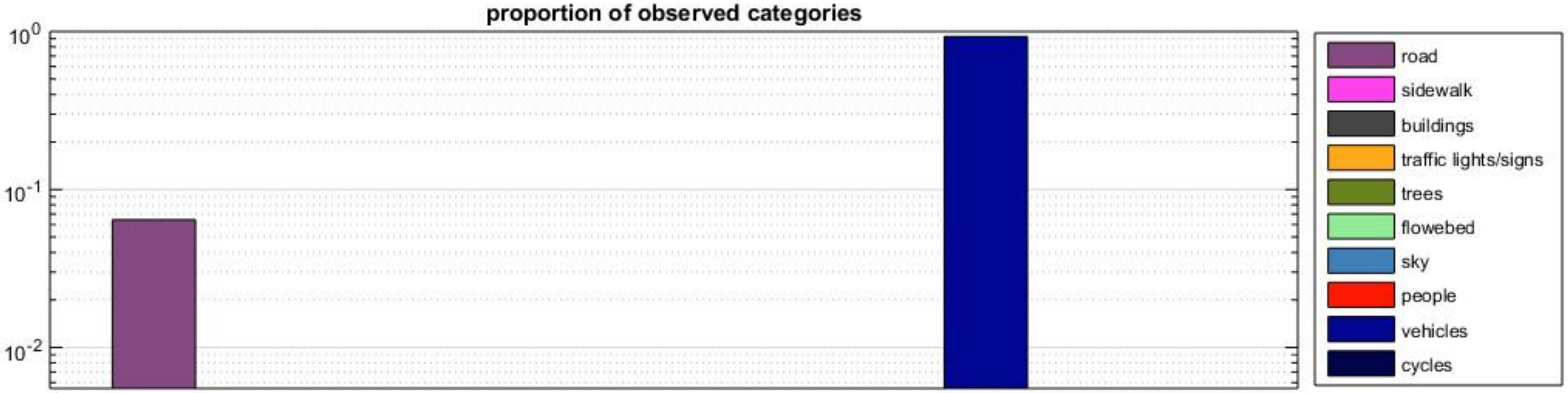


(a) $0 \leq \mathrm{km/h} \leq 10$ (b) $10 \leq \mathrm{km/h} \leq 30$ (c) $30 \leq \mathrm{km/h} \leq 50$ (d) $50 \leq \mathrm{km/h} \leq 70$ (e) $70 \leq \mathrm{km/h}$

DATASET ANALYSIS: SEMANTICS

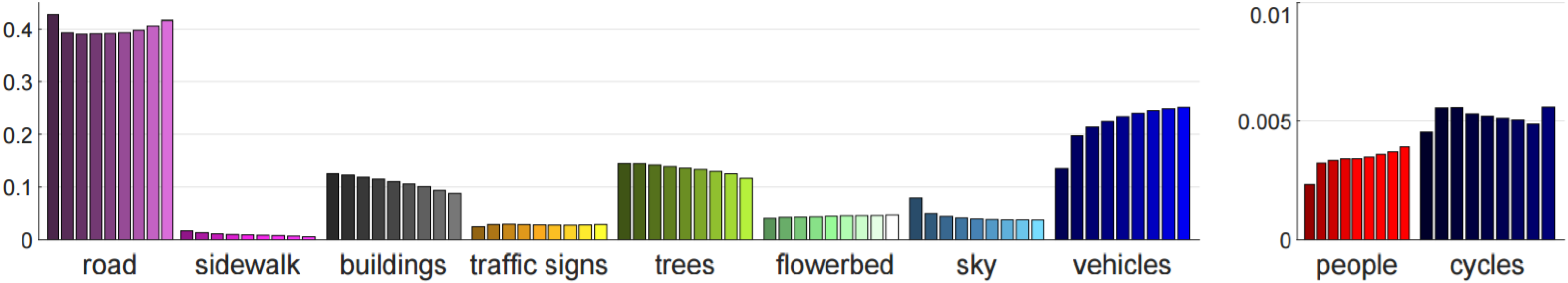


Attraction towards specific semantic categories



DATASET ANALYSIS: SEMANTICS

Not all categories “hit” by the gaze are the true focus of attention

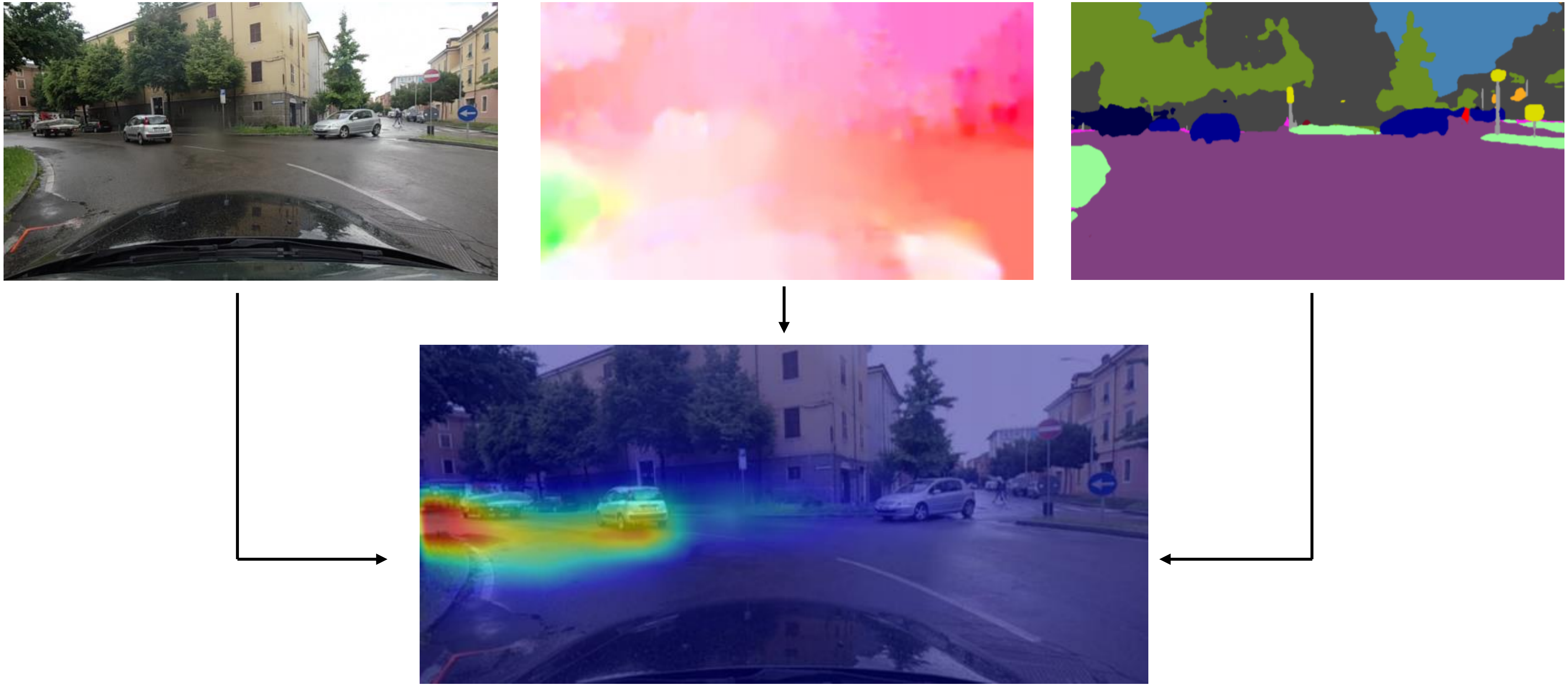


MODEL AND RESULTS

MIMICKING THE DRIVER: INTUITION

Exploiting multi-source information

Appearance | Motion | Semantics

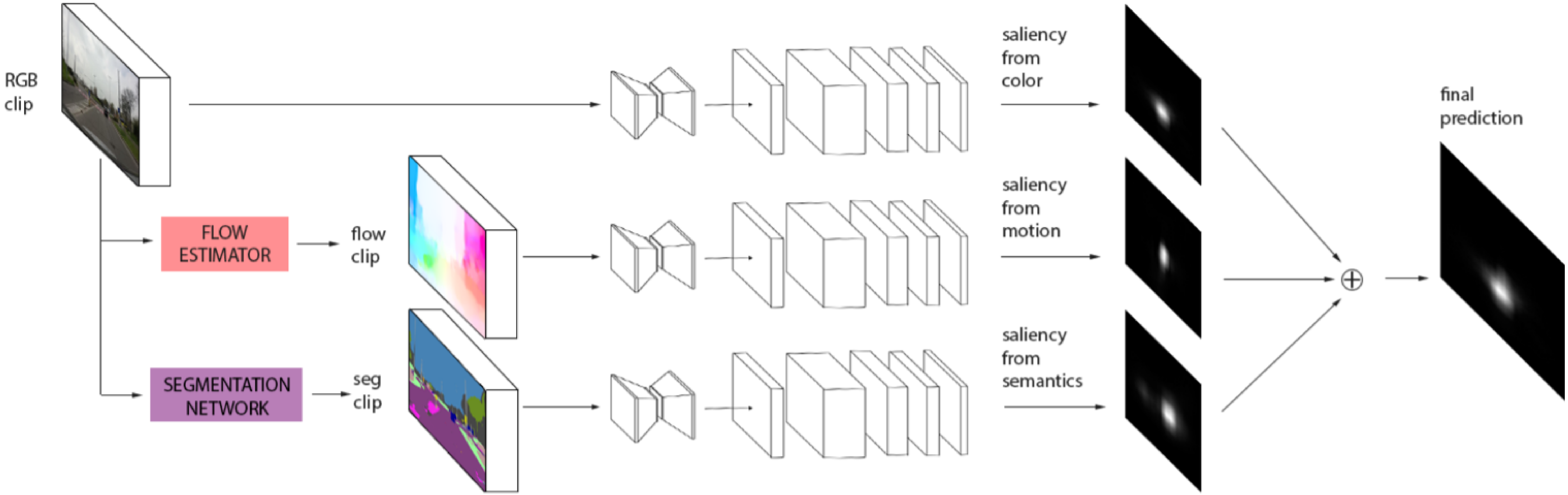


Automatically predicted FoA

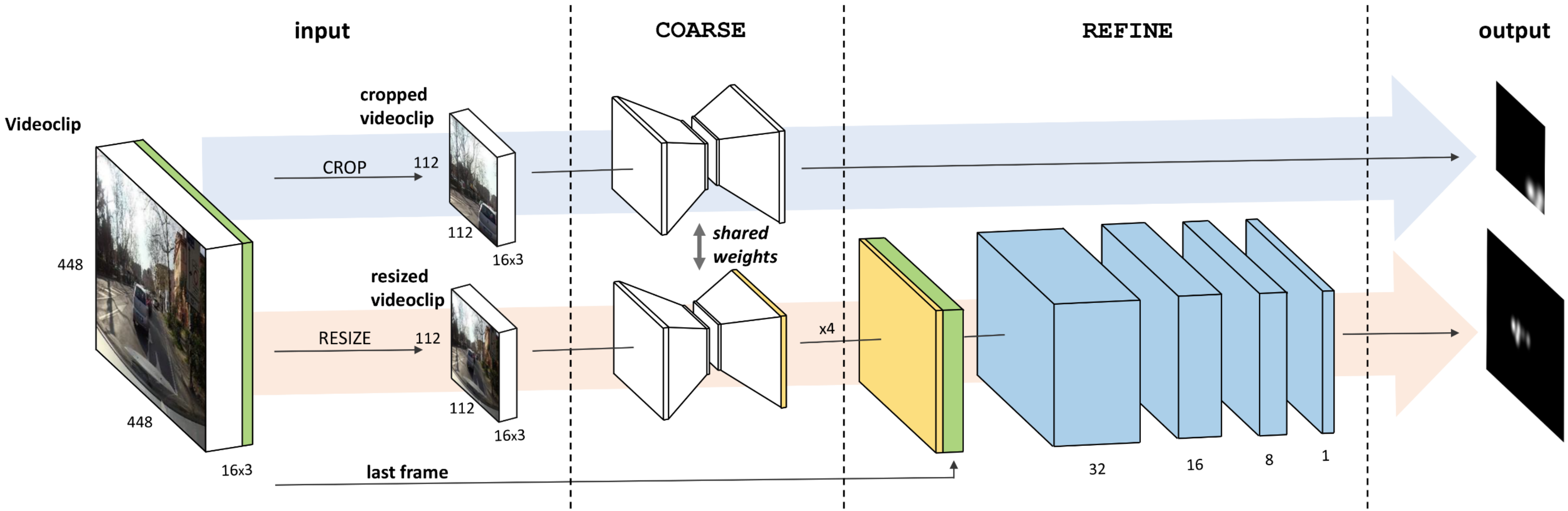
MULTI-BRANCH MODEL

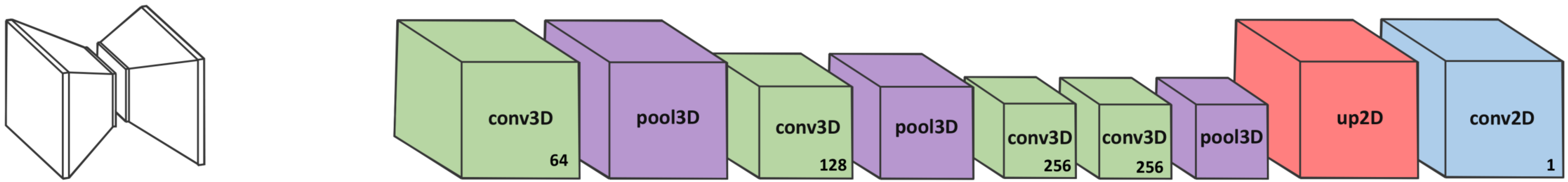
Deep learning model (approximately 43 M of learnable parameters)

Trained to predict human FoA given a driving scene



MODEL: SINGLE BRANCH

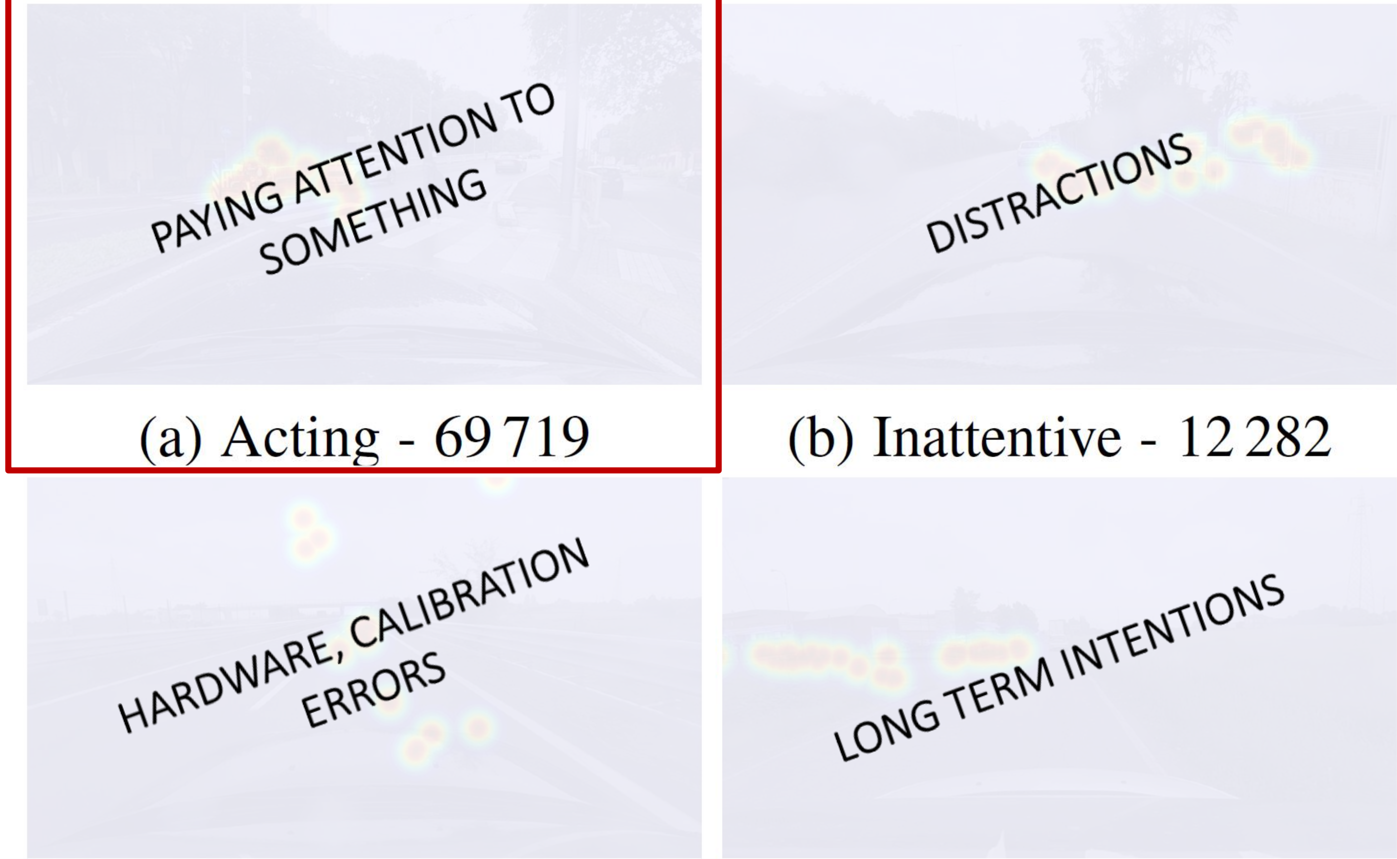




MIMICKING THE DRIVER: RESULTS

SUBSEQUENCES ANNOTATION

MOST IMPORTANT SUBSET!!



(a) Acting - 69 719

(b) Inattentive - 12 282

(c) Error - 22 893

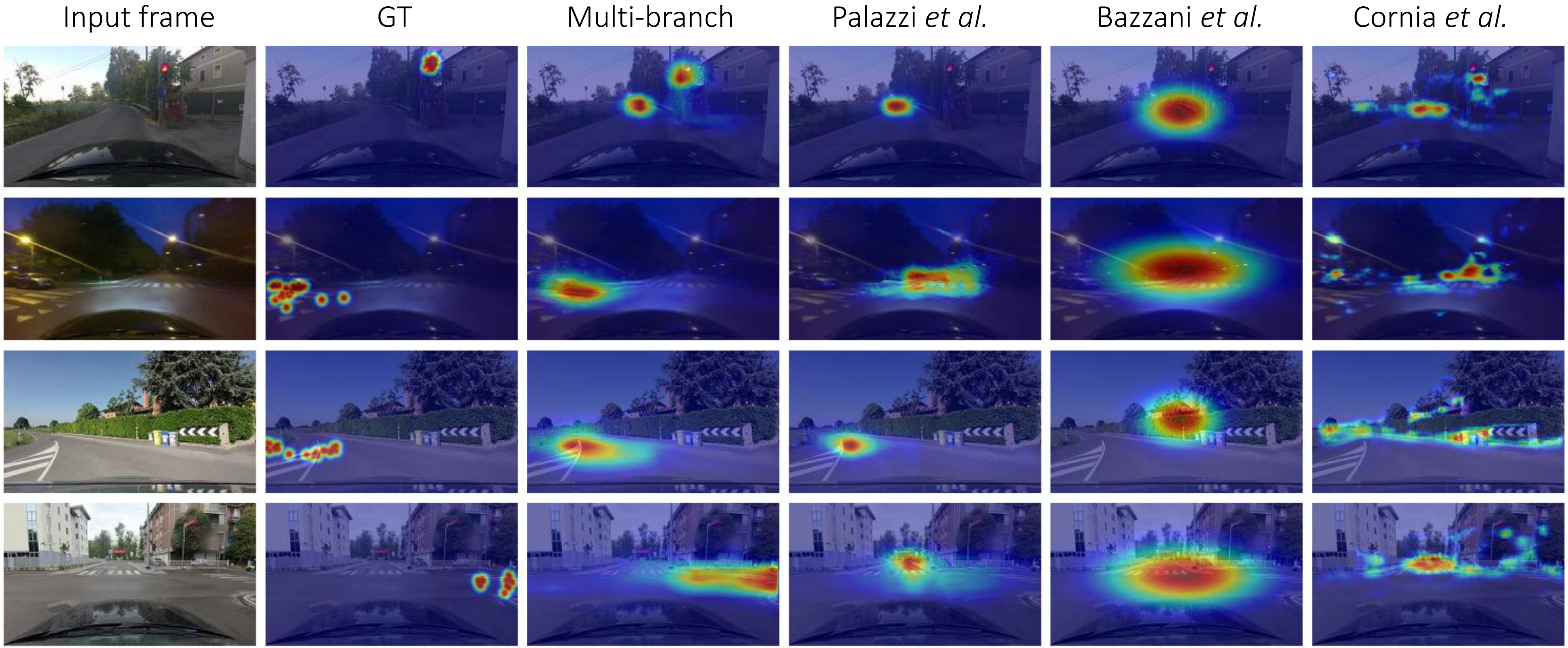
(d) Subjective - 3 166

RESULTS: QUANTITATIVE

	Test sequences			Acting subsequences		
	CC ↑	D_{KL} ↓	IG ↑	CC ↑	D_{KL} ↓	IG ↑
Baseline Gaussian	0.40	2.16	-0.49	0.26	2.41	0.03
Baseline Mean	0.51	1.60	0.00	0.22	2.35	0.00
Mathe *et al.*	0.04	3.30	-2.08	-	-	-
Wang *et al.*	0.04	3.40	-2.21	-	-	-
Wang *et al.*	0.11	3.06	-1.72	-	-	-
MLNet	0.44	2.00	-0.88	0.32	2.35	-0.36
RMDN	0.41	1.77	-0.06	0.31	2.13	0.31
Palazzi *et al.*	0.55	1.48	-0.21	0.37	2.00	0.20
`multi-branch`	**0.56**	**1.40**	**0.04**	**0.41**	**1.80**	**0.51**

S. Mathe et al, Actions in the eye: Dynamic gaze datasets and learnt saliency models for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.
W. Wang et. al, Saliency-aware geodesic video object segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
W. Wang et. al, Consistent video saliency using local gradient flow optimization and global refinement. IEEE Transactions on Image Processing, 2015.
M. Cornia et al, A Deep Multi-Level Network for Saliency Prediction. In International Conference on Pattern Recognition, 2016.
L. Bazzani et al, Recurrent mixture density network for spatiotemporal visual attention. In International Conference on Learning Representations, 2017.
A. Palazzi, et al, Learning where to attend like a human driver. Intelligent Vehicles Symposium, 2017.

RESULTS: QUALITATIVE



M. Cornia et al, A Deep Multi-Level Network for Saliency Prediction. In International Conference on Pattern Recognition, 2016.
L. Bazzani et al, Recurrent mixture density network for spatiotemporal visual attention. In International Conference on Learning Representations, 2017.
A. Palazzi, et al, Learning where to attend like a human driver. Intelligent Vehicles Symposium, 2017.

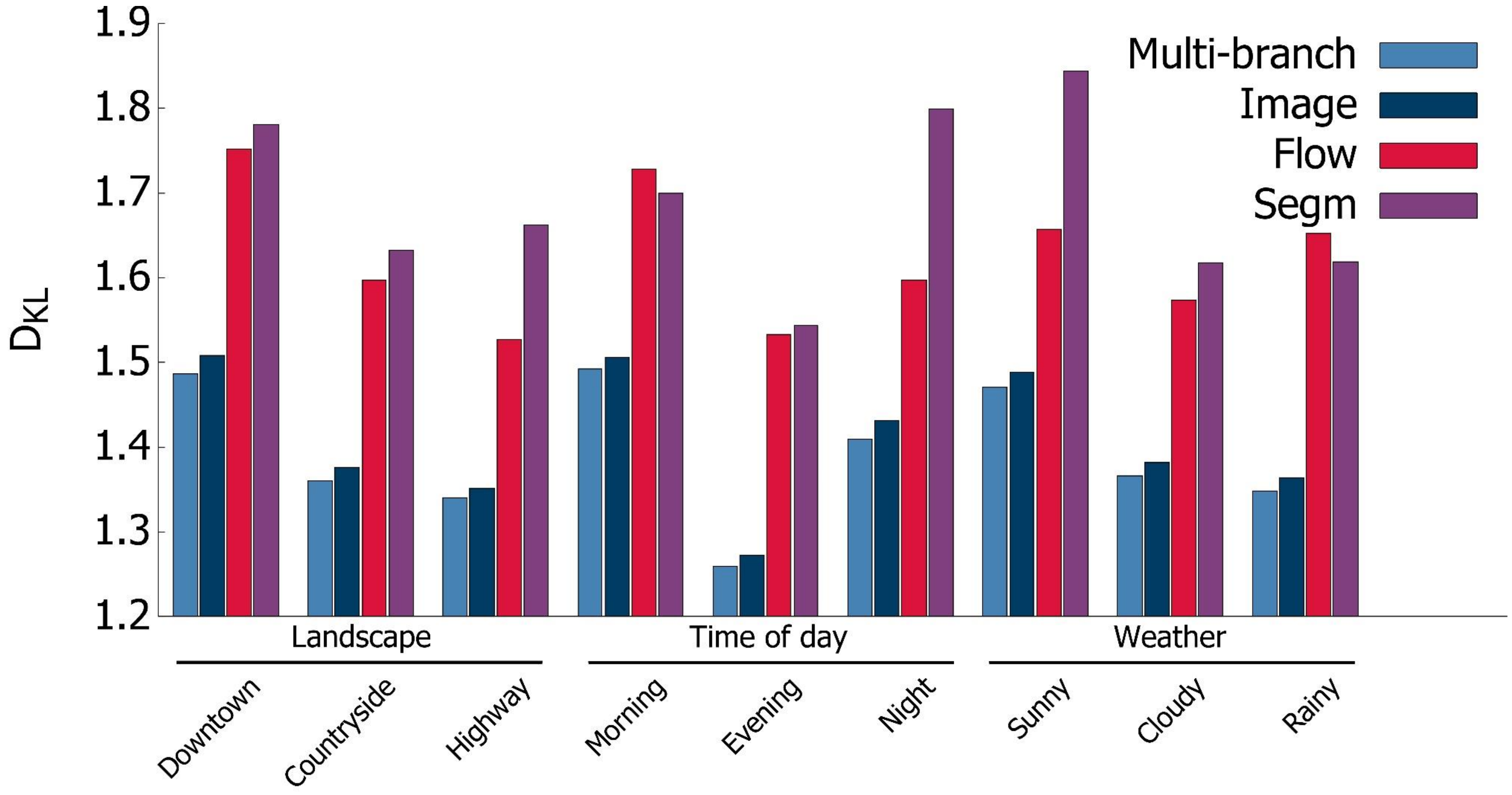
RESULTS: ABLATION STUDY

	Test sequences			Acting subsequences		
	CC $\uparrow$	D_{KL} $\downarrow$	IG $\uparrow$	CC $\uparrow$	D_{KL} $\downarrow$	IG $\uparrow$
I	0.554	1.415	-0.008	0.403	1.826	0.458
F	0.516	1.616	-0.137	0.368	2.010	0.349
S	0.479	1.699	-0.119	0.344	2.082	0.288
I+F	0.558	1.399	0.033	**0.410**	1.799	0.510
I+S	0.554	1.413	-0.001	0.404	1.823	0.466
F+S	0.528	1.571	-0.055	0.380	1.956	0.427
I+F+S	**0.559**	**1.398**	**0.038**	**0.410**	**1.797**	**0.515**

I = Image branch
F = Optical flow branch
S = Segmentation branch

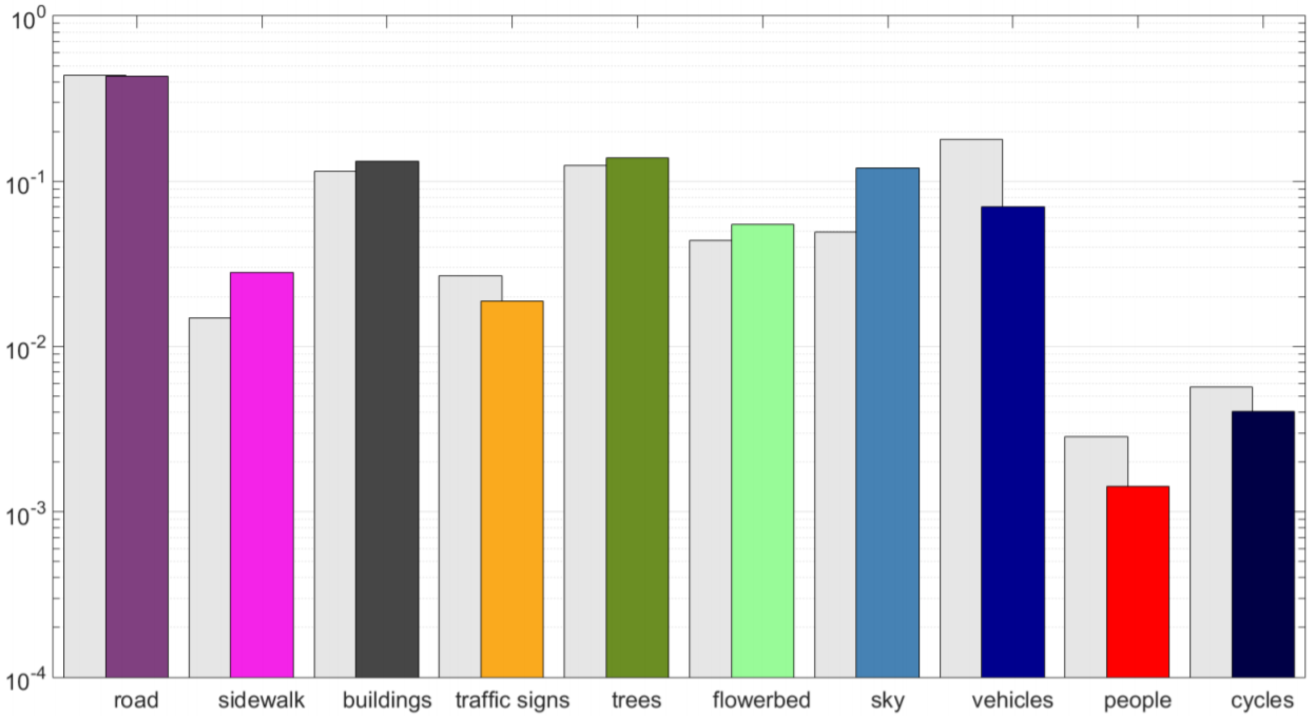
RESULTS: DRIVING ENVIRONMENT

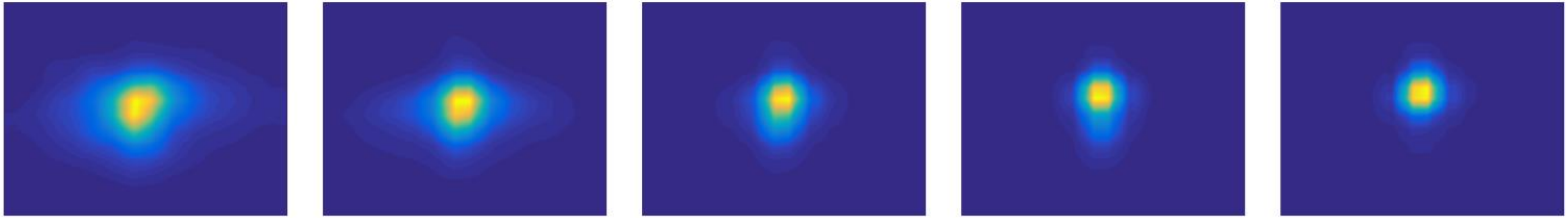
How does performance vary depending on the driving environment?



MODEL: ATTENTIONAL DYNAMICS

Model's attentional behaviors resemble human behaviors both in terms of semantics and speed.

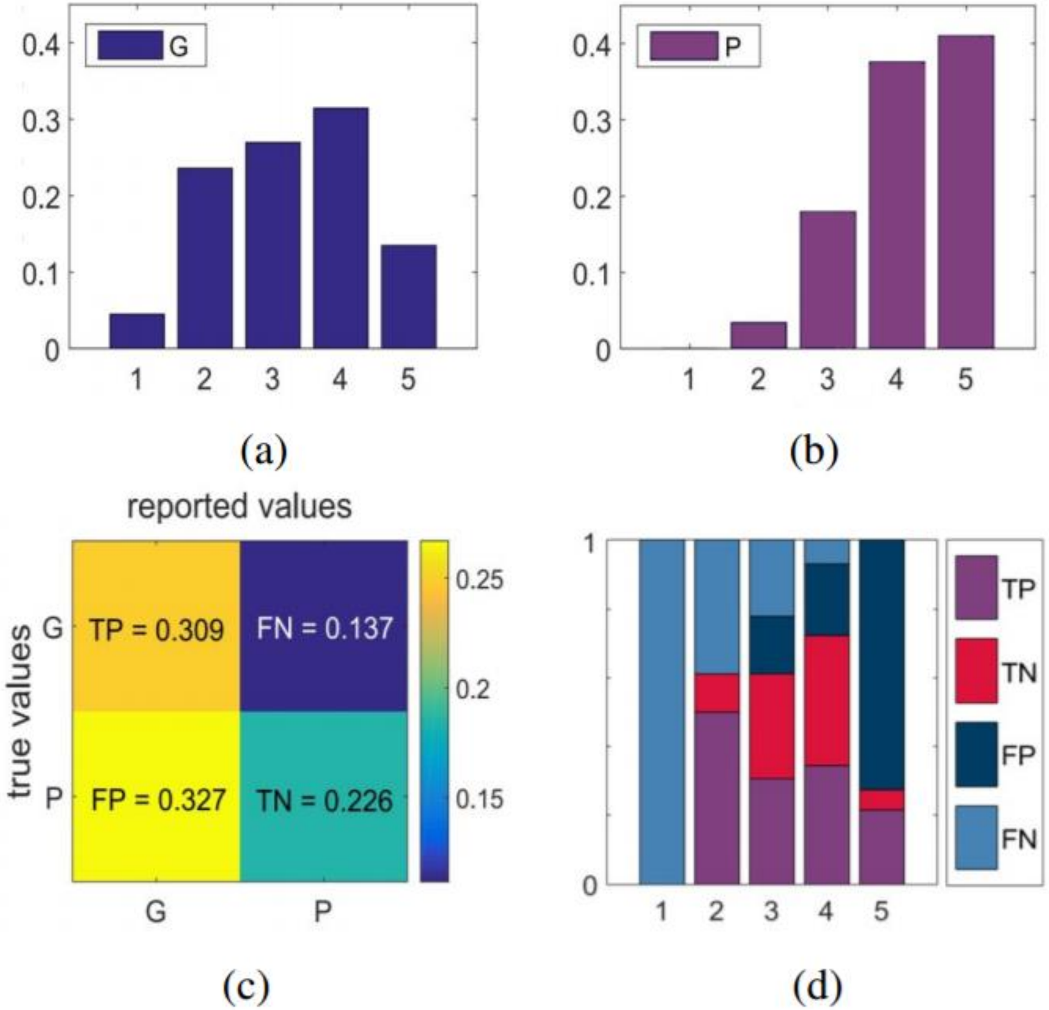




(a) $0 \leq \mathrm{km/h} \leq 10$ (b) $10 \leq \mathrm{km/h} \leq 30$ (c) $30 \leq \mathrm{km/h} \leq 50$ (d) $50 \leq \mathrm{km/h} \leq 70$ (e) $70 \leq \mathrm{km/h}$

VISUAL ASSESSMENT

Can you tell human behavior vs model behavior?



CONCLUSIONS

CONCLUSIONS

Goal: investigating and replicating human attentional dynamics during the driving task.

Contributions:

- Collection of a huge real-world, publicly available dataset of human fixations during the driving task
- Analysis of relationship between scene condition and driver's focus
- Development of a deep learning model that automatically infers where the driver should probably focus his attention, given a certain scene.

Thank you for your attention!
Questions?

PREDICTING THE DRIVER'S FOCUS OF ATTENTION:
THE DR(EYE)VE PROJECT

Andrea Palazzi, Davide Abati, Simone Calderara, Francesco Solera, Rita Cucchiara

(name.surname@unimore.it)