



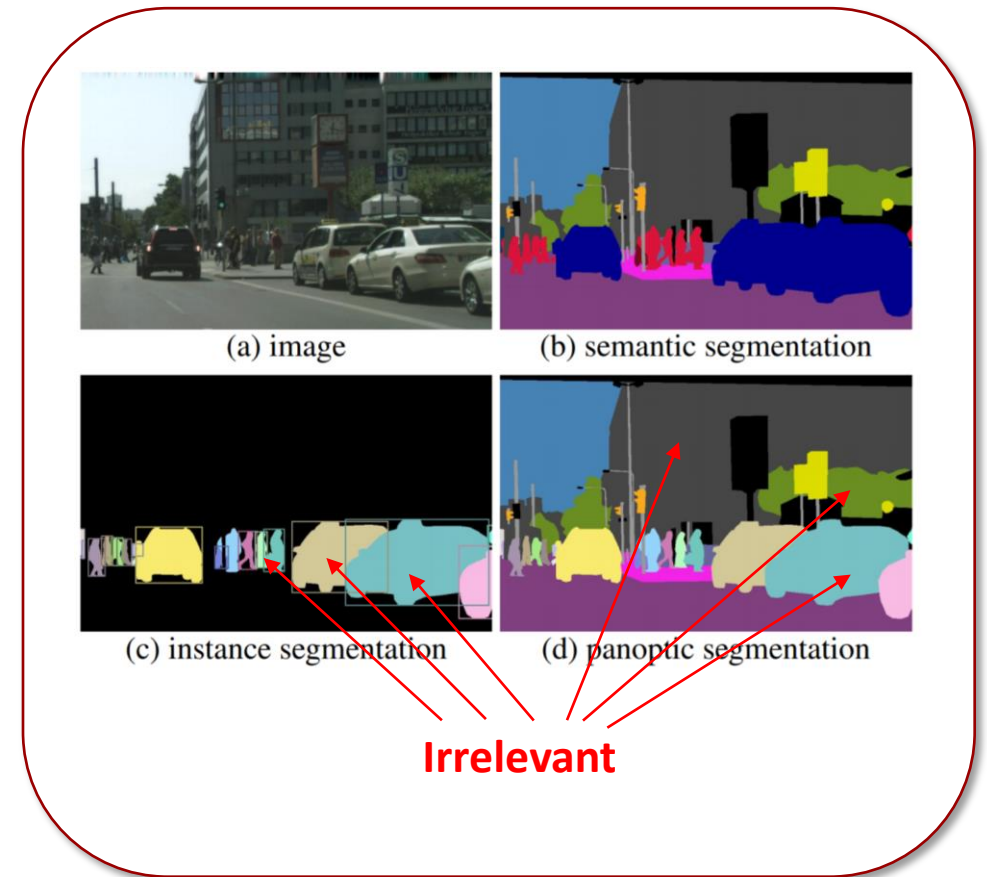
Action-Inducing Object Detection

NRI: FND: Towards Scalable and Self-Aware Robotic Perception
Award IIS-1924937

PI: Nuno Vasconcelos
UCSD

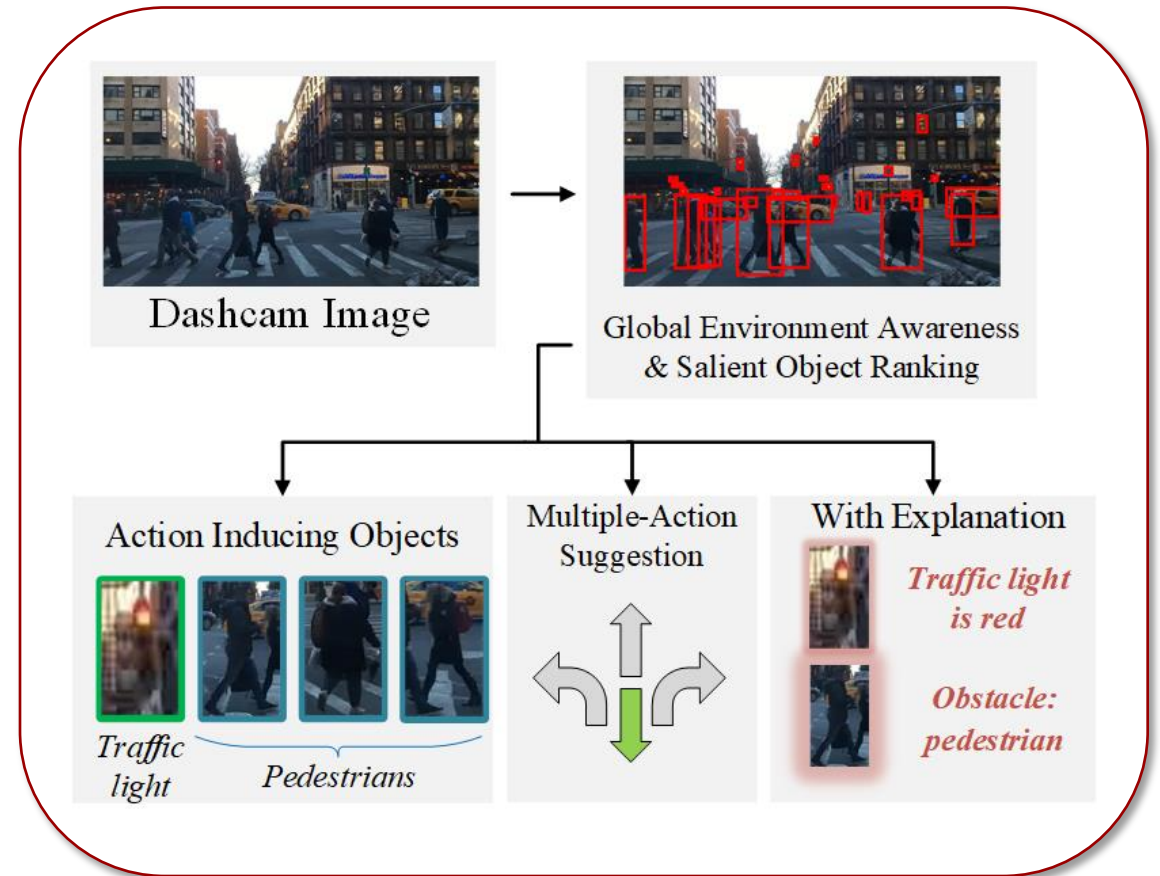
Broader Impact

- **current** paradigm in vision
 - segment **every possible object** in scene
 - computation **intensive and unnecessary**
- **human** drivers
 - **ignore** all objects that are **irrelevant for driving decisions**
 - buildings, trees, lamp posts, parked cars, ...
 - **perception is guided by action**
- **this work**
 - we define **action inducing object detection**
 - **simultaneous** prediction of allowable **actions** and detection of **objects that constrain those actions**



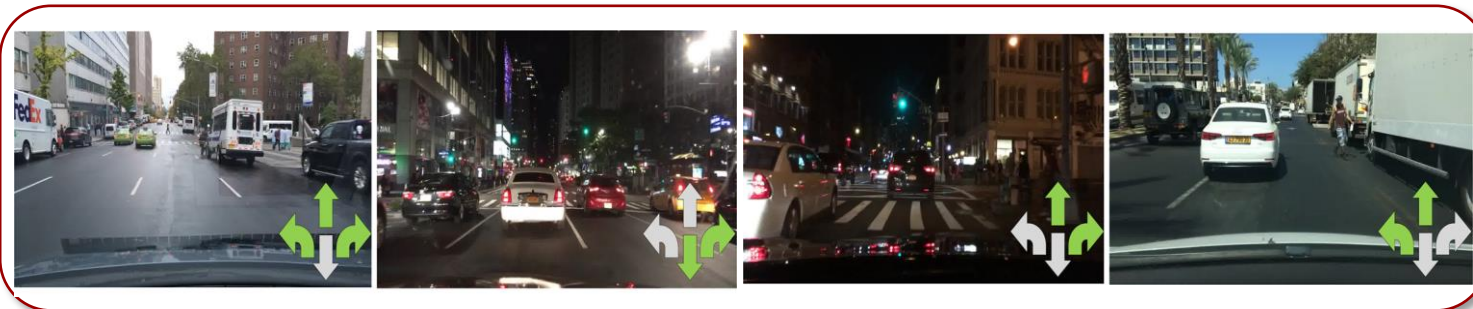
Action Induced Object detection

- given a **complex** scene
 - rather than all objects in the scene, **detect only those important to driving decisions**
 - these are **action inducing objects**: pedestrians, traffic lights, cars on the road, etc.
- goals:
 - Detect AIOs
 - **predict the actions allowable**
 - **combine the two into explanations**, e.g. “slow down because the light is red and there are pedestrians crossing”



BDD-OIA dataset

- new **dataset**
 - based on **BDD100K**
 - only **complex** scenes
 - 8 pedestrians, 12 vehicles per scene on avg
 - 4 **action categories**
 - 21 possible **explanations**
 - labelled for feasible actions, and their **explanations**

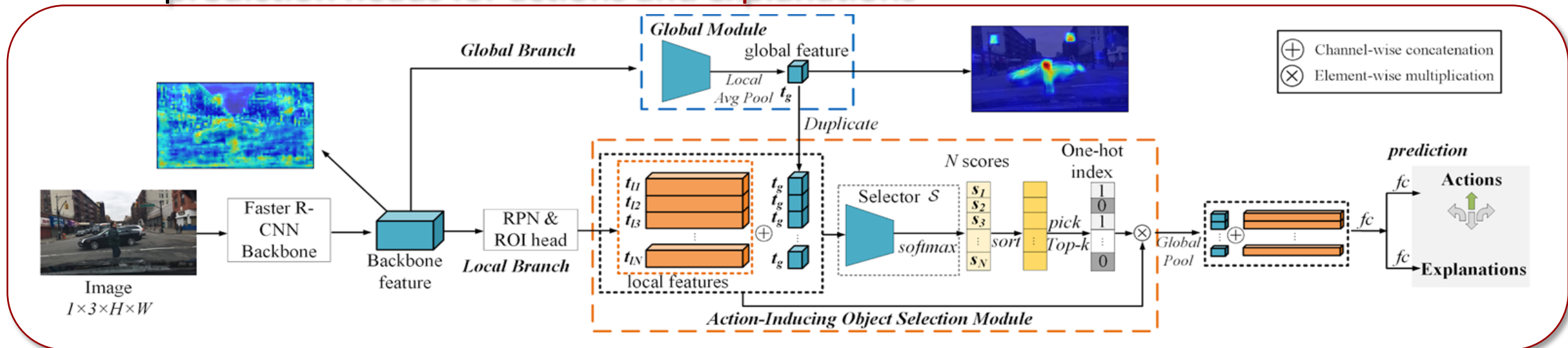


Dataset	# pedestrians	# vehicles
BDD100K [43]	1.2	9.7
KITTI [20]	0.8	4.1
Cityscapes [22]	7.0	11.8
BDD-OIA	8.0	11.8

Action Category	Number	Explanations	Number
Move forward	12491	Traffic light is green	7805
		Follow traffic	3489
		Road is clear	4838
Stop/Slow down	10432	Traffic light	5381
		Traffic sign	1539
		Obstacle: car	233
		Obstacle: person	163
		Obstacle: rider	5255
Turn left	838	No lane on the left	150
		Obstacles on the left lane	666
		Solid line on the left	316
Turn left	5064	On the left-turn lane	154
		Traffic light allows	885
		Front car turning left	365
Turn right	1071	No lane on the right	4503
		Obstacles on the right lane	4514
		Solid line on the right	3660
Turn right	5470	On the right-turn lane	6081
		Traffic light allows	4022
		Front car turning right	2161

Deep Learning Architecture

- new **architecture for AIO detection**
 - Faster R-CNN backbone
 - global processing branch to model **context**
 - localized processing branch to **detect objects**
 - selection module to **identify AIOs according to scene context**
 - prediction heads for **actions and explanations**





Questions?