Towards Robust and Socially-Adept Autonomous Vehicles Through Vehicle Trajectory Sensing for Safety Assessment

'YZ' Yezhou Yang, Assistant Professor, SCAI, ASU
Group Lead, Active Perception Group,
School of Computing and AI, Arizona State University
Projects Tech co-Lead, The Institute of Automated Mobility (IAM)

@Yezhou_Yang

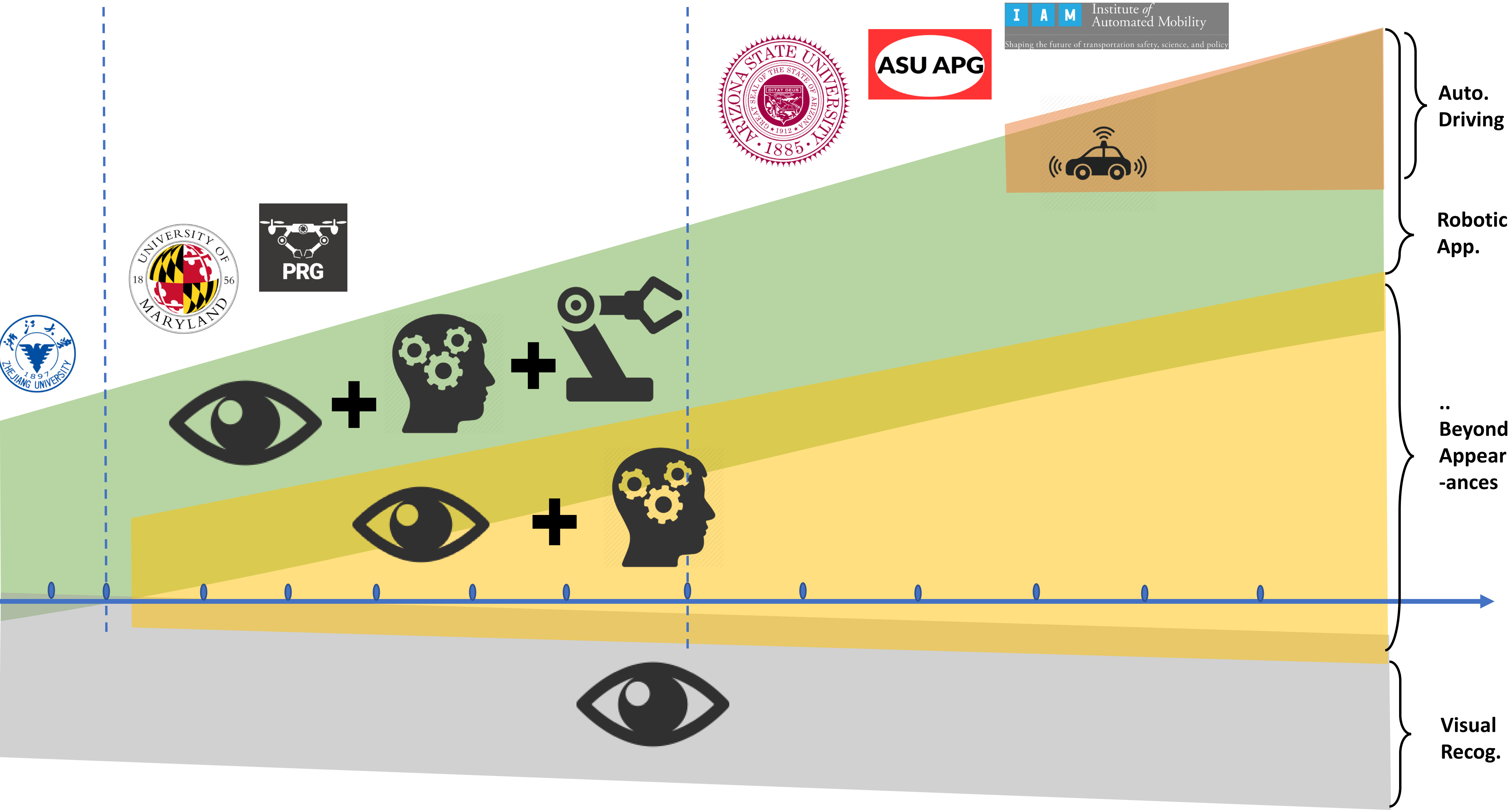
Arizona State University

ASU APG

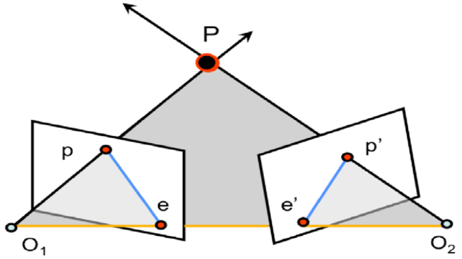
I A M Institute *of* Automated Mobility

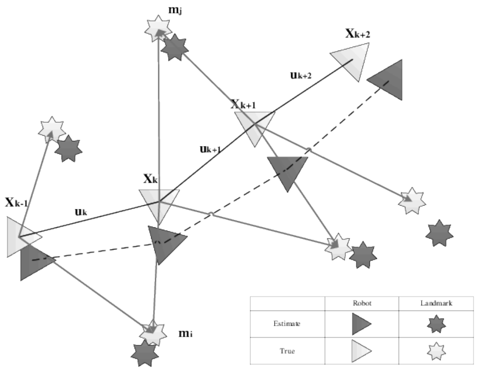
Shaping the future of transportation safety, science, and policy

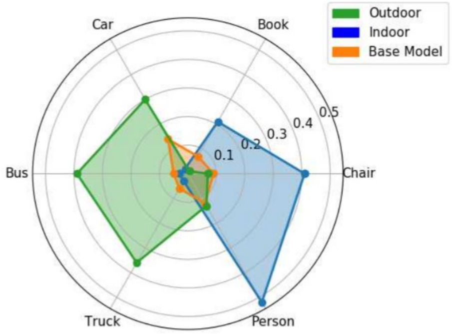
Sep 30th 2021 @ ITS AZ



Institute of Automated Mobility
Shaping the future of transportation safety, science, and policy
ASU APG
PRG
Auto. Driving
Robotic App.
.. Beyond Appear -ances
Visual Recog.







from Internet

CBS NEWS / August 31, 2019, 1:21 PM

"We're heading towards hell": Expert shares concerns with self-driving cars

CREDIT RSS OCTOBER 25, 2019 / 7:26 AM / UPDATED 2 HOURS AGO

INTERVIEW: Autonomous vehicles raise numerous regulatory issues

Jason Hsieh 13 MIN READ

4,615 views | Sep 26, 2019, 10:03am

What Happens When Self-Driving Cars Kill People?

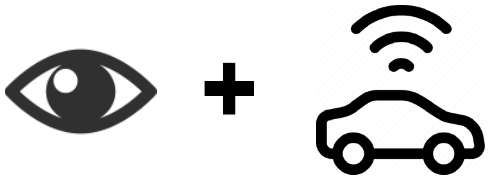
BUSINESS | LOGISTICS REPORT | WSJ LOGISTICS REPORT

Self-Driving Technology Threatens Nearly 300,000 Trucking Jobs, Report Says

Impact would come over 25 years, with projections for lighter job loss seen than other forecasts, but higher-paying trucking work could take a hit

56%: would not ride in an autonomous vehicle.
16%: feel safe to let an autonomous vehicle drive them without the option of taking control.
16%: feel autonomous cars will eventually eliminate the need for car insurance.

http://www.pewinternet.org/2017/10/04/automation-in-everyday-life/pi_2017-10-04_automation_0-02/
https://www.coxautoinc.com/news/evolution-of-mobility-study-autonomous-vehicles/
https://www.erieinsurance.com/blog/multi-gen-car-survey



So, what are these gaps?

And, how are we going to fill (or attempt to fill) these gaps (or a few of them) from AI/CV perspectives?

1) The signal to semantic gap ← Visual Recognition with Knowledge

2) From lab to the society gap ← Socially adept autonomous driving

3) The equipment gap ← AV Performance Evaluation with Existing Traffic Cameras;

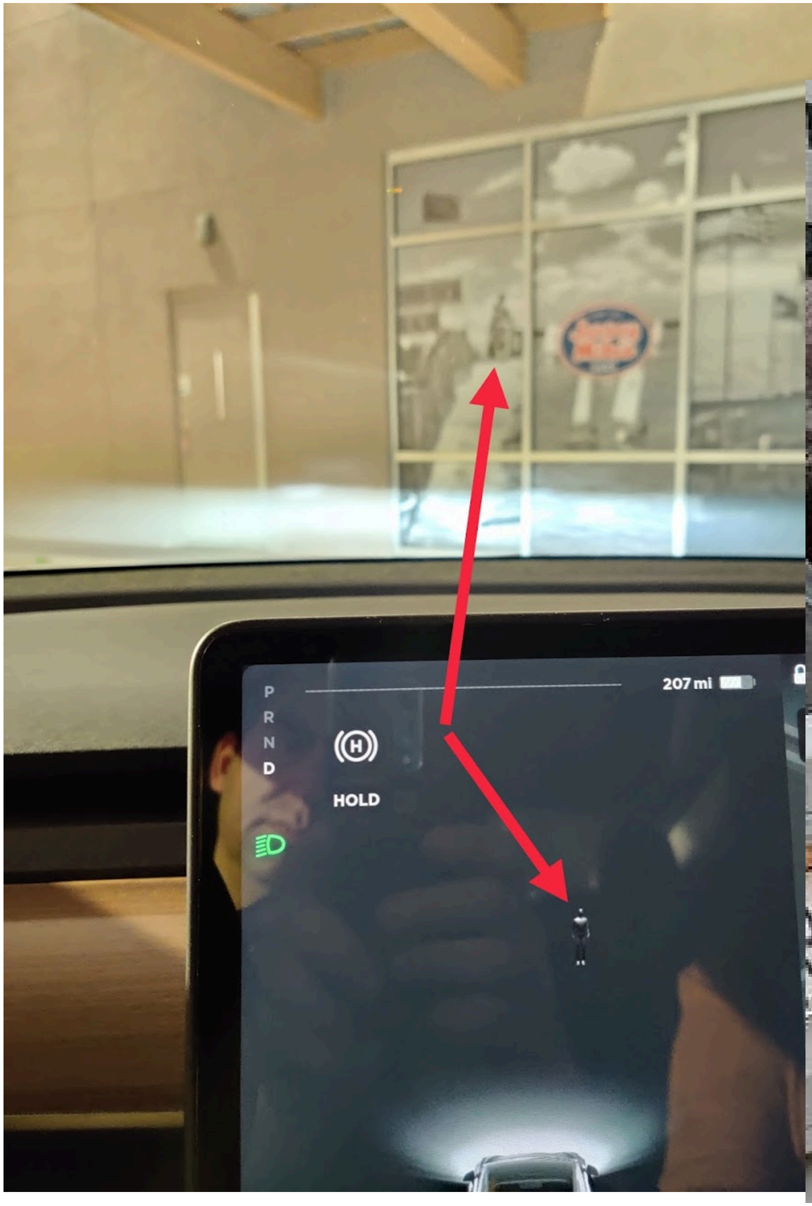
4) From tech to transportation practitioner gap ← ARGOS Vision.

1) The signal to semantic gap ← Visual Recognition with Knowledge

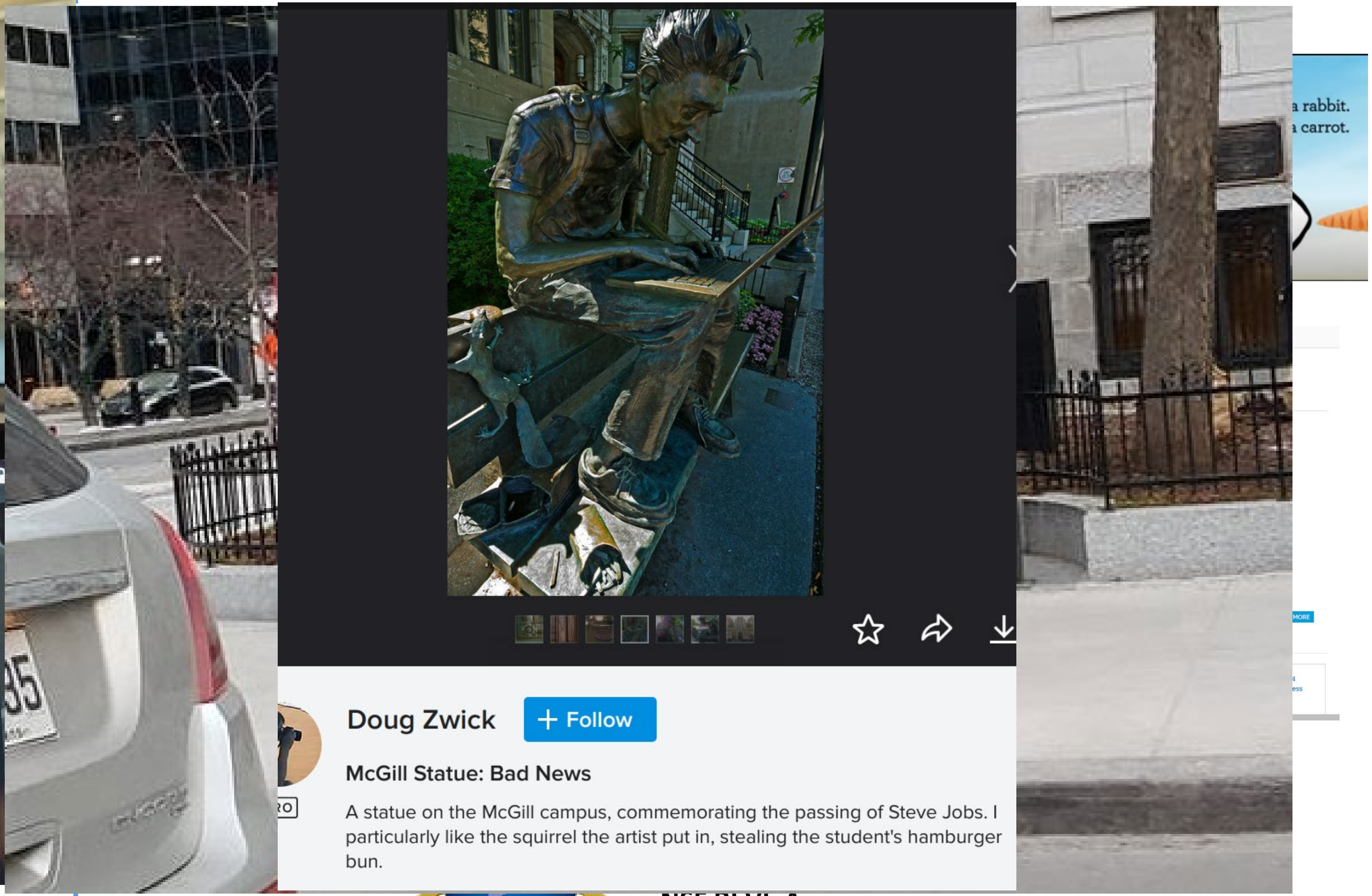
Visual Question Answering

Q: how many people are waiting for bus?

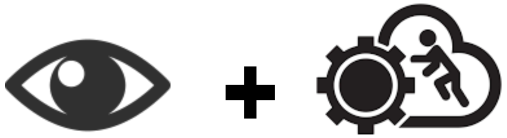
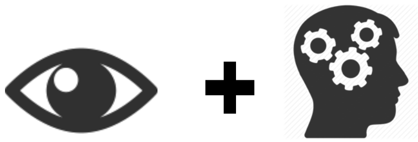
A: Two? or Three?



From Internet and a friend

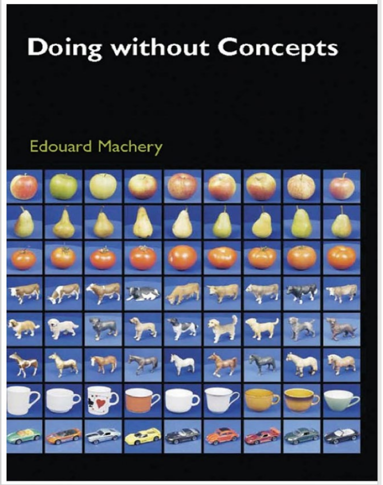


NSF RI VL-Aug

 ≠ 

Visual Recognition as Pattern Matching:

"Visual recognition is a cognitive process that involves identification of a visible CATEGORY from previous encounters "



Categories

Visual Recognition as it is:

"Visual recognition is a cognitive process that involves identification of a visible CONCEPT from previous encounters or KNOWLEDGE."

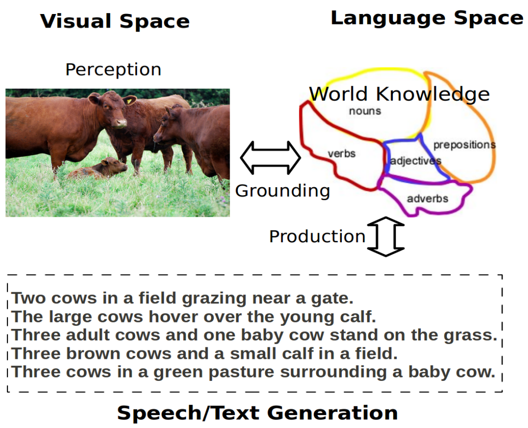
What is a concept?

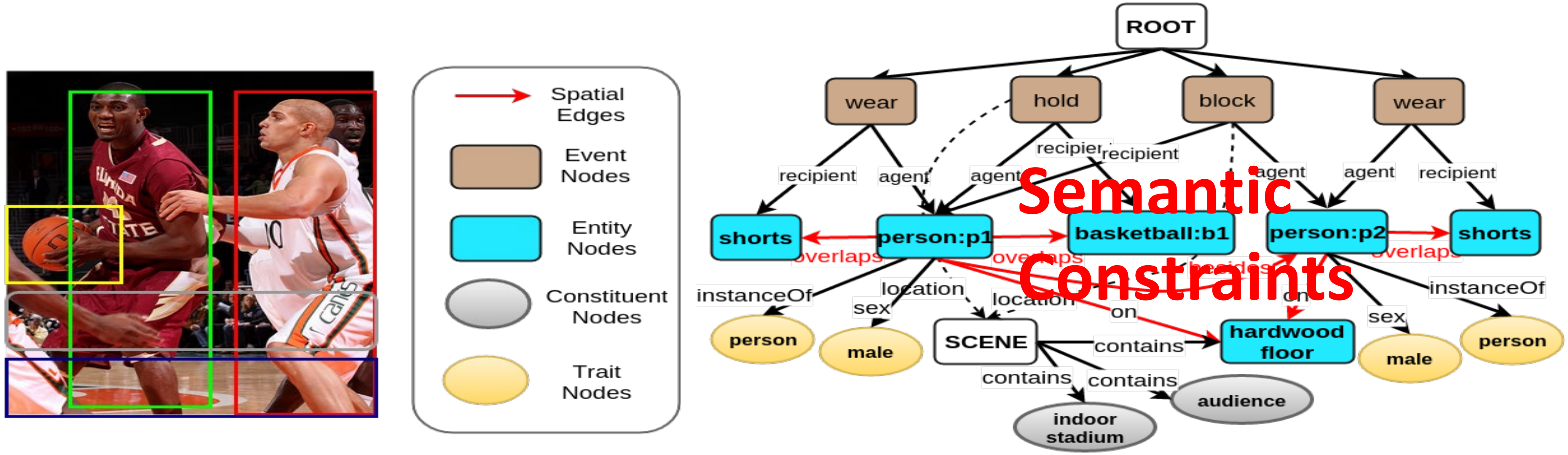
"… A theory of concepts should describe the kind of knowledge stored in concepts, the way they are used in agents' cognitive processes, their format, their acquisition, and their neural localization… "

≠

Concepts

1) The signal to semantic gap: the representation gap





EMNLP 11'
Sen. Gen. from Img, Captioning

ACS 16'
DeepIU Scene Description Graph (SDGs)

CVIU 17'
Image Under. w/ SDG

SDGs project webpage:
https://adityasomak.github.io/publication/sdg_cviu/

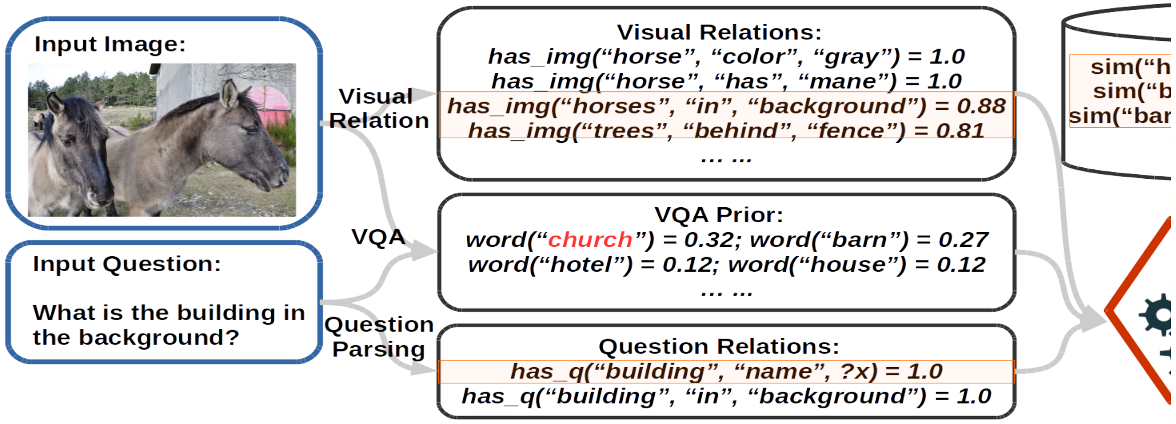
Experiment	BRNN-Karpathy	Our Method	Gold Standard
R ± D(8k)	2.08 ± 1.35	**2.82 ± 1.56**	4.69 ± 0.78
T ± D(8k)	2.24 ± 1.33	**2.62 ± 1.42**	4.32 ± 0.99
R ± D(30k)	1.93 ± 1.32	**2.43 ± 1.42**	4.78 ± 0.61
T ± D(30k)	2.17 ± 1.34	**2.49 ± 1.42**	4.52 ± 0.93
R±D(COCO)	**2.69 ± 1.49**	2.14 ± 1.29	4.71 ± 0.67
T±D(COCO)	**2.55 ± 1.41**	2.06 ± 1.24	4.37 ± 0.92

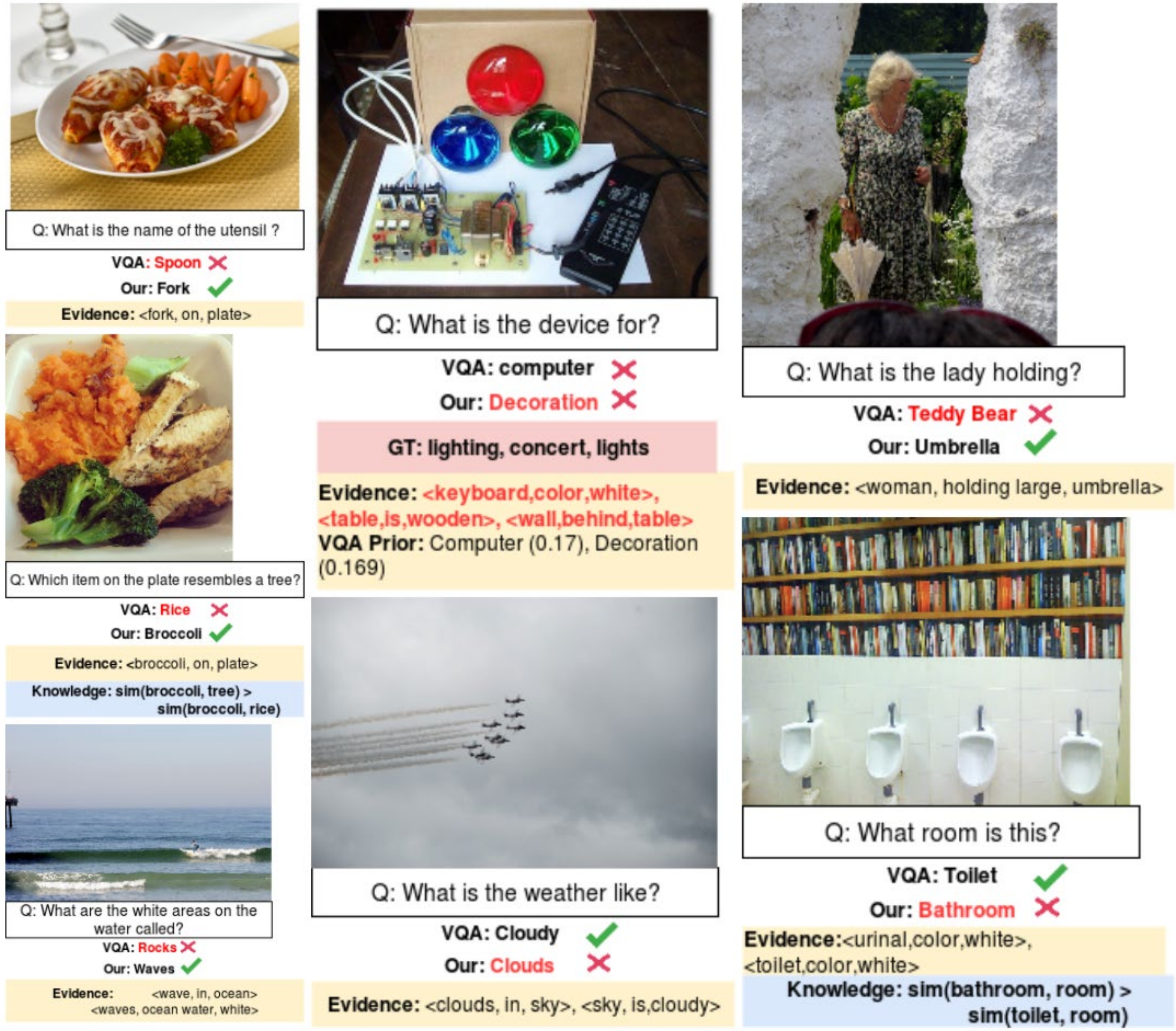
Table 1: Sentence generation relevance (R) and thoroughness (T) human evaluation results with gold standard and BRNN-Karpathy on Flickr 8k, 30k and MS-COCO datasets. D: Standard Deviation.

	Flickr8k			
Model	**R@1**	**R@5**	**R@10**	**Med r**
BRNN-Karpathy	11.8	32.1	44.7	12.4
Our Method-SDG	**18.1**	**39.0**	**50.0**	**10.5**
	Flickr30k			
BRNN-Karpathy	15.2	37.7	50.5	9.2
Our Method-SDG	**26.5**	**48.7**	**59.4**	**6.0**
	MS-COCO			
BRNN-Karpathy (1k)	**20.9**	**52.8**	**69.2**	**4.0**
Our Method-SDG (1k)	19.3	35.5	49.0	11.0
Our Method-SDG (2k)	15.4	32.5	42.2	17.0

Table 2: Image-Search Results: We report the recall@K (for $K = 1, 5$ and 10) and Med r (Median Rank) metric for Flickr8k, 30k and COCO datasets. For COCO, we experimented on first 1000 (1k) and random 2000 (2k) validation images.

1) The signal to semantic gap: the explicit reasoning for interpretation





AAAI 18’
Explicit Reasoning f/ VQA

UAI 18’
Knowledge & Reasoning for Image Puzzles

1) The signal to semantic gap: the fundamental logic-based reasoning

Is the plate green? Submit

Predicted top-5 answers with confidence:

yes	100.000%
no	0.000%
green	0.000%
broccoli	0.000%
unknown	0.000%

NEGATION

Is the plate not green? Submit

Predicted top-5 answers with confidence:

yes	94.785%
no	5.215%
green	0.000%
unknown	0.000%
y	0.000%

Tan & Bansal, LXMERT: Learning Cross-Modality Encoder Representationsfrom Transformers, EMNLP 2019

1) The signal to semantic gap: the fundamental logic-based reasoning

VQA-LOL: Visual Question Answering under the Lens of Logic

Question		Predicted Answer	Accuracy (%) SOTA	Accuracy (%) LOL
VQA				
Q_1:	Is there beer?	YES (0.96)	88.20	86.55
Q_2:	Is the man wearing shoes?	NO (0.90)		
VQA-Compose			50.69	82.39
$\neg Q_2$:	Is the man *not* wearing shoes?	NO (0.80)		
$\neg Q_2 \wedge Q_1$	Is the man *not* wearing shoes *and* is there beer?	NO (0.62)		
$Q_1 \wedge C$	Is there beer and does this seem like a man bending over to look inside of a fridge?	NO (1.00)		
VQA-Supplement			50.61	87.80
$\neg Q_2 \vee B$	Is the man not wearing shoes or is there a clock?	NO (1.00)		
$Q_1 \wedge anto(B)$	Is there beer and is there a wine glass?	YES (0.84)		

ECCV 20'
VQA-LOL

1) The signal to semantic gap: out-of-domain (OOD) generalization

Visual Question Answering
Q: Is it a fast vehicle?
A: Yes

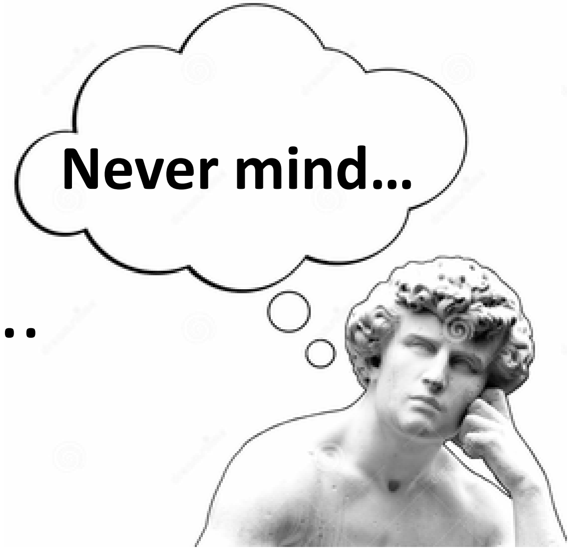
Visual Question Answering
Q: Is it a fast vehicle?
A: No

1) The signal to semantic gap: out-of-domain (OOD) generalization

A: Green

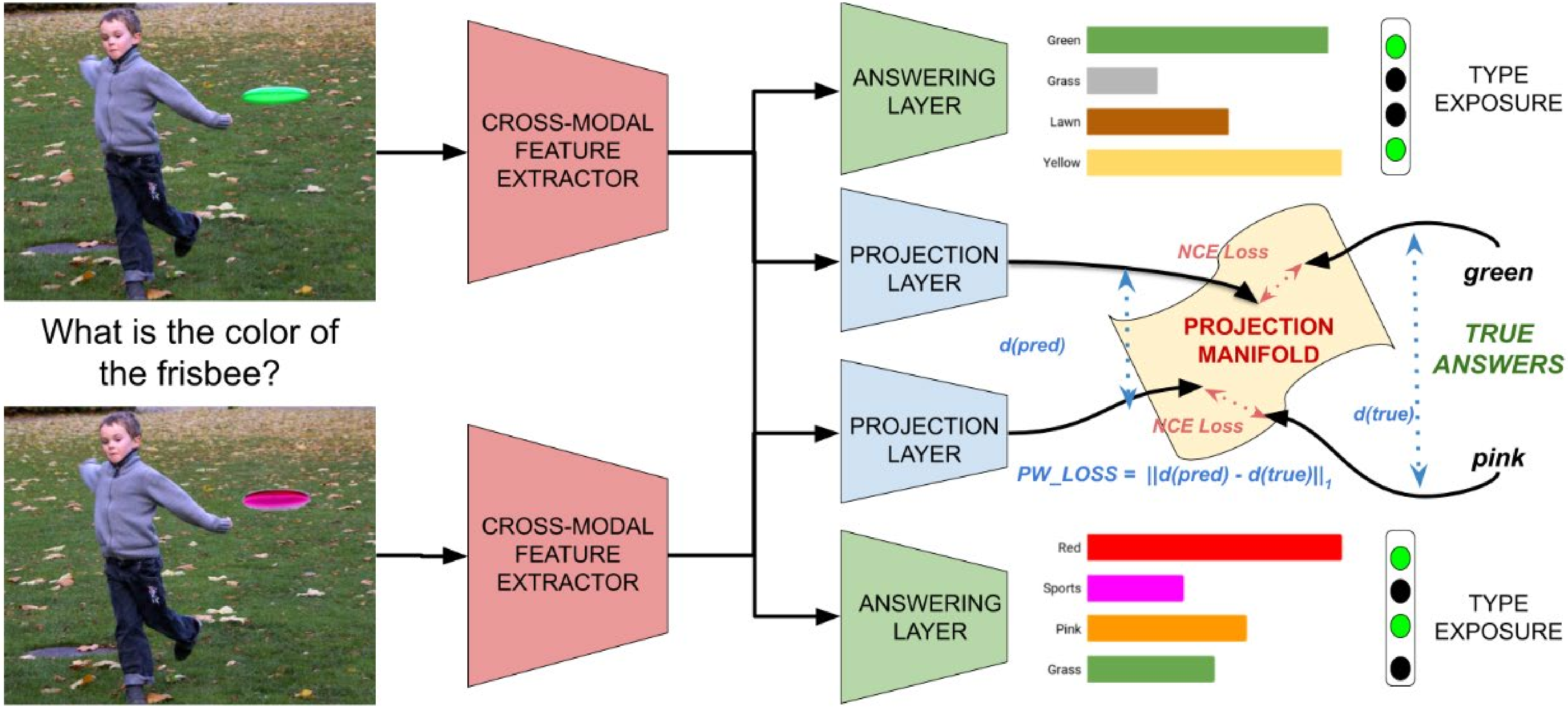
What is the color of the frisbee?

A: I think it is still green?...



1) The signal to semantic gap: out-of-domain (OOD) generalization

Mutant: A training paradigm for out-of-distribution generalization in visual question answering



EMNLP 20'
VQA-MUTANT

UpDn: Anderson et al. CVPR 2018
LXMERT: Tan et al. EMNLP 2019

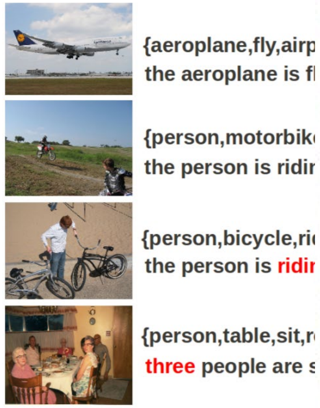
Analysis: Effect of Mutant Samples

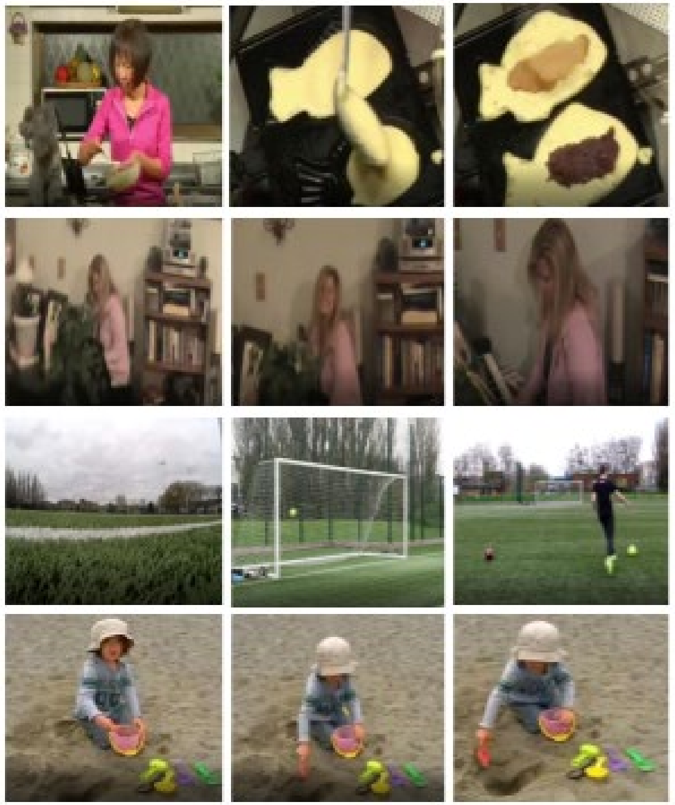
Model	Data	VQA-CP v2 test ↑ (%)			
		All	Yes/No	Num	Other
UpDn	VQA-CP	39.74	42.27	11.93	46.05
UpDn	VQA-CP + Mutant	50.16	61.45	35.87	50.14
Increase in Accuracy		*10.42*	*19.18*	*23.94*	*4.09*
LXMERT	VQA-CP	46.23	42.84	18.91	55.51
LXMERT	VQA-CP + Mutant	59.69	73.19	32.85	59.29
Increase in Accuracy		*13.46*	*30.35*	*13.94*	*3.78*
LXM + Ours	VQA-CP + Img. Mut.	64.85	85.68	66.44	53.80
LXM + Ours	VQA-CP + Que. Mut.	67.92	91.64	65.73	56.09
LXM + Ours	VQA-CP + Both Mut.	**69.52**	**93.15**	**67.17**	**57.78**

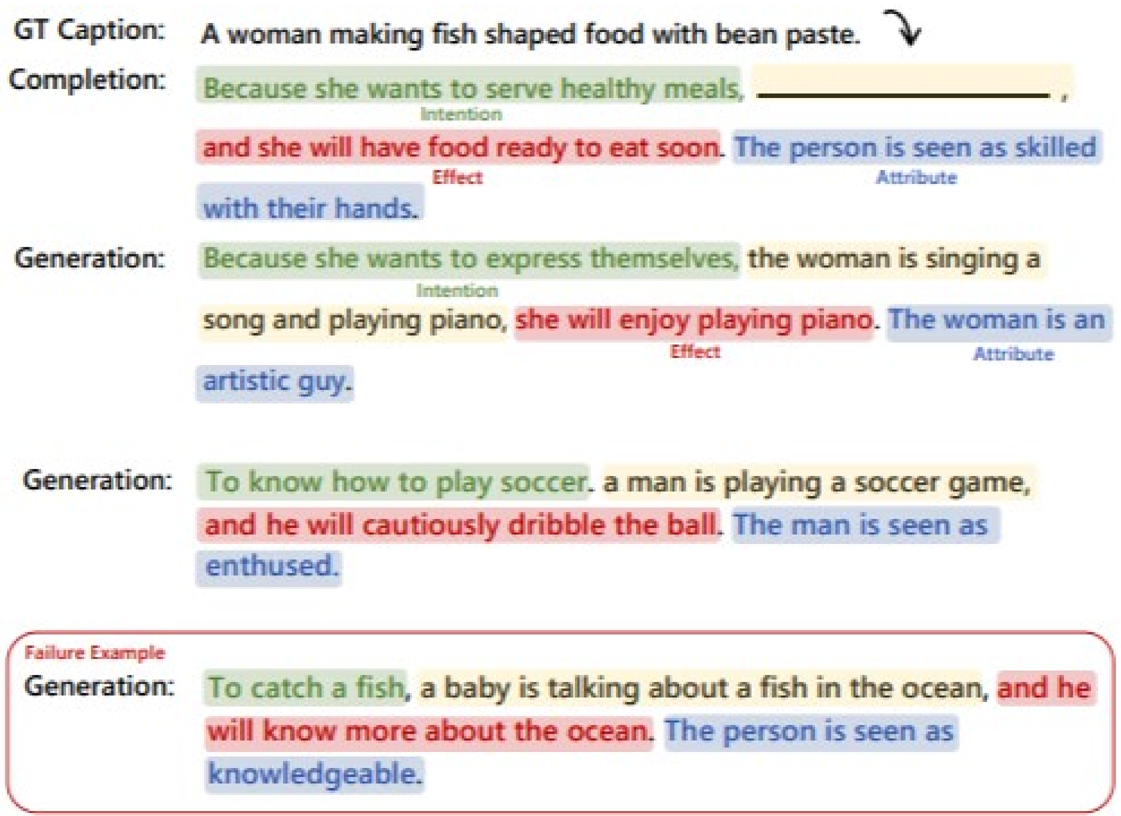
Comparison of Backbone models (UpDn, LXMERT) trained with VQA-CP data augmented with MUTANT samples.

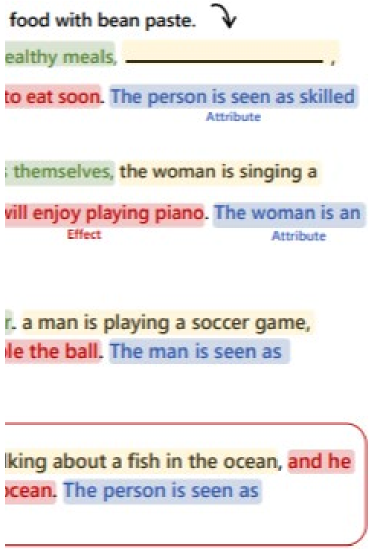
Comparison of our best model when trained with: image mutations, question mutations, and both types of mutations.

1) The signal to semantic gap: perceiving beyond appearance









(g) (h)

EMNLP 11'
Sen. Gen. from Img, Captioning

ACS 16'
DeepIU
Scene
Description
Graph (SDG)

CVIU 17'
Image Under.
w/ SDG

V2C: Video to
Commonsense

V2C VIDEO TO COMMONSENSE

https://asu-active-perception-group.github.io/Video2Commonsense/index.html

2) From lab to the society gap ← Socially adept autonomous driving

Irrationally courteous AV:

AV recognizes that its best action from the driver's perspective is to wait. Thus it waits...

How do we define a good driver?

https://jalopnik.com/how-to-recognize-a-good-driver-5947854

10. They move over after passing.

7. They are not overly polite at intersections.

6. They can park.

5. They use their turn signals.

3. They make confident lane changes.

1. They drive predictably

How to create a driver that is <u>naturally</u> good? How to evaluate whether a driver is <u>naturally</u> good?

NSF National Robotics Initiative 19-22
Socially-adept Autonomous Driving

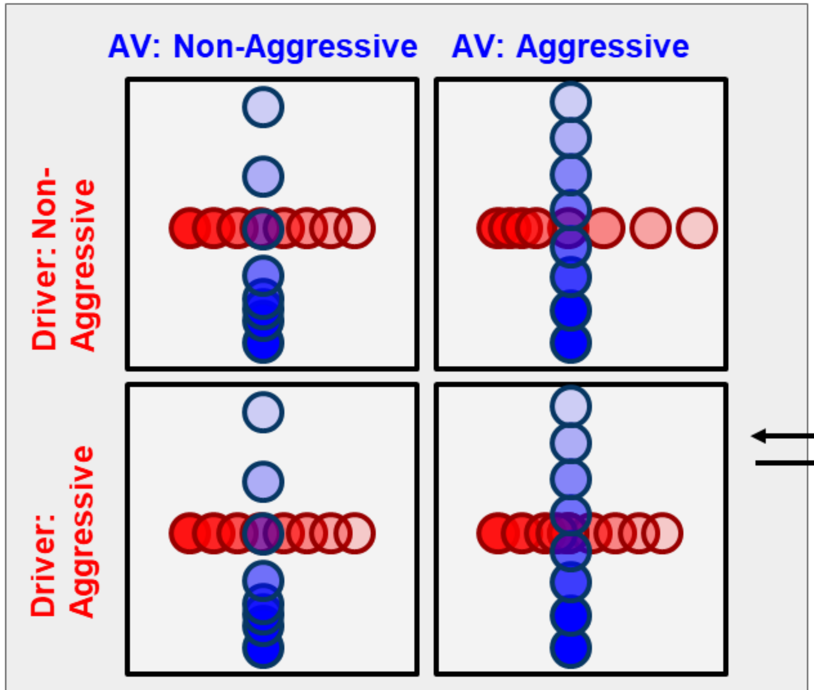
2) From lab to the society gap ← Socially adept autonomous driving

“Self-driving cars need to be nice, but not overly nice”:

Simply behave to satisfy others does not make a good driver.

Solution: **Rational courtesy** (through recognition of Nash Equilibria)

Brief explanation of our method



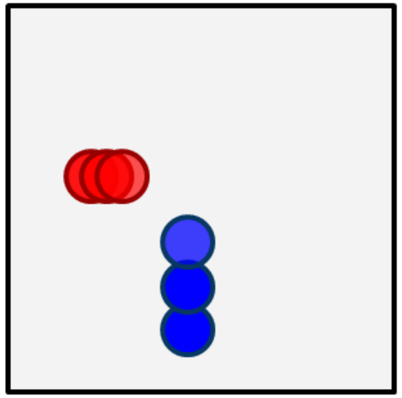
Hypothetical Nash Equilibria

Method: Find hypothesis that matches with the reality. Update belief using Bayesian update.

Challenge: Hypotheses change all the time. (Don’t have enough time to update them in reality).

Solution: Create intuition through experience (Offline fictitious self-play)

Find a match?



Reality

ICRA 19’
How shall I drive?

2) From lab to the society gap ← Socially adept autonomous driving

Irrationally courteous AV:

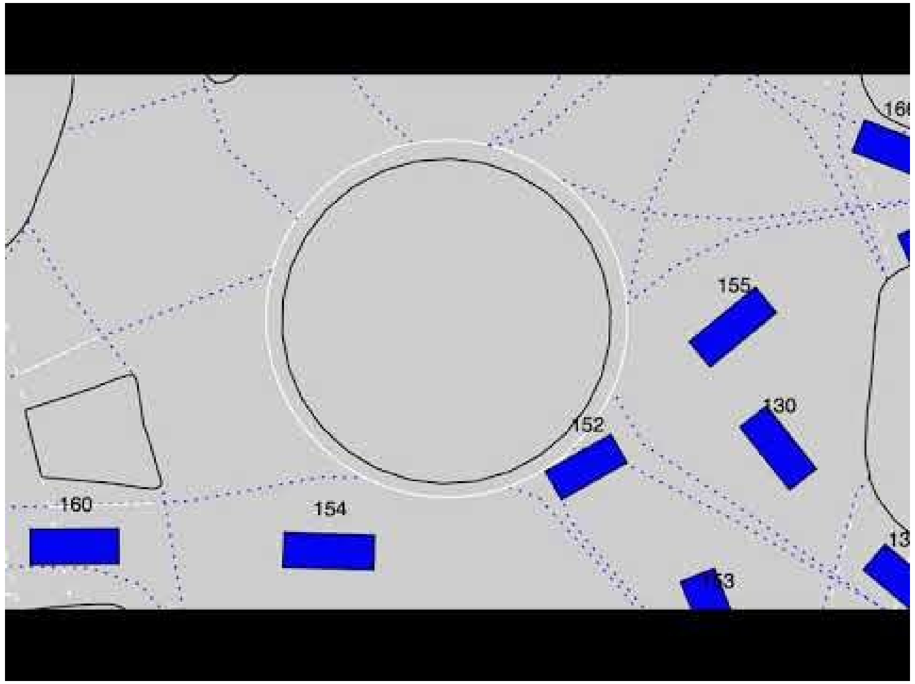
AV recognizes that its best action from the driver’s perspective is to wait. Thus it waits...

Rationally courteous AV:

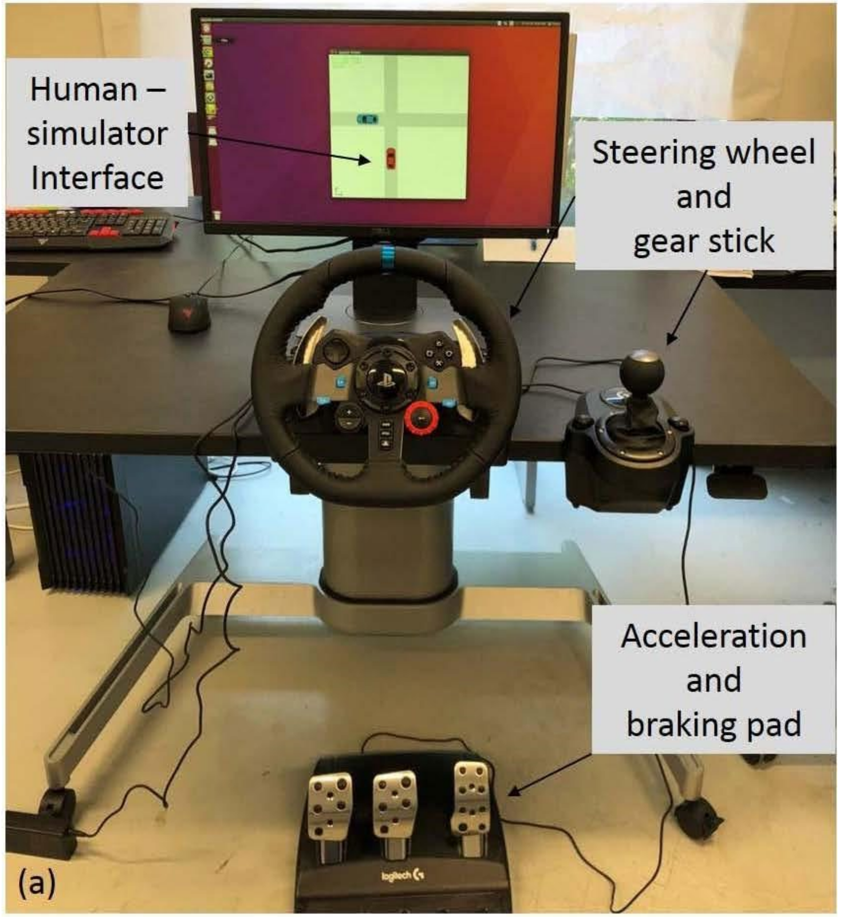
AV recognizes that from the driver’s perspective, its best action among all Nash Equilibria is to leave as soon as possible.

ICRA 19’
How shall I drive?

- **Scalability and other natural driving scenarios**
- **Human modeling and prediction**
- **Safety metrics & corner cases via human experiments**

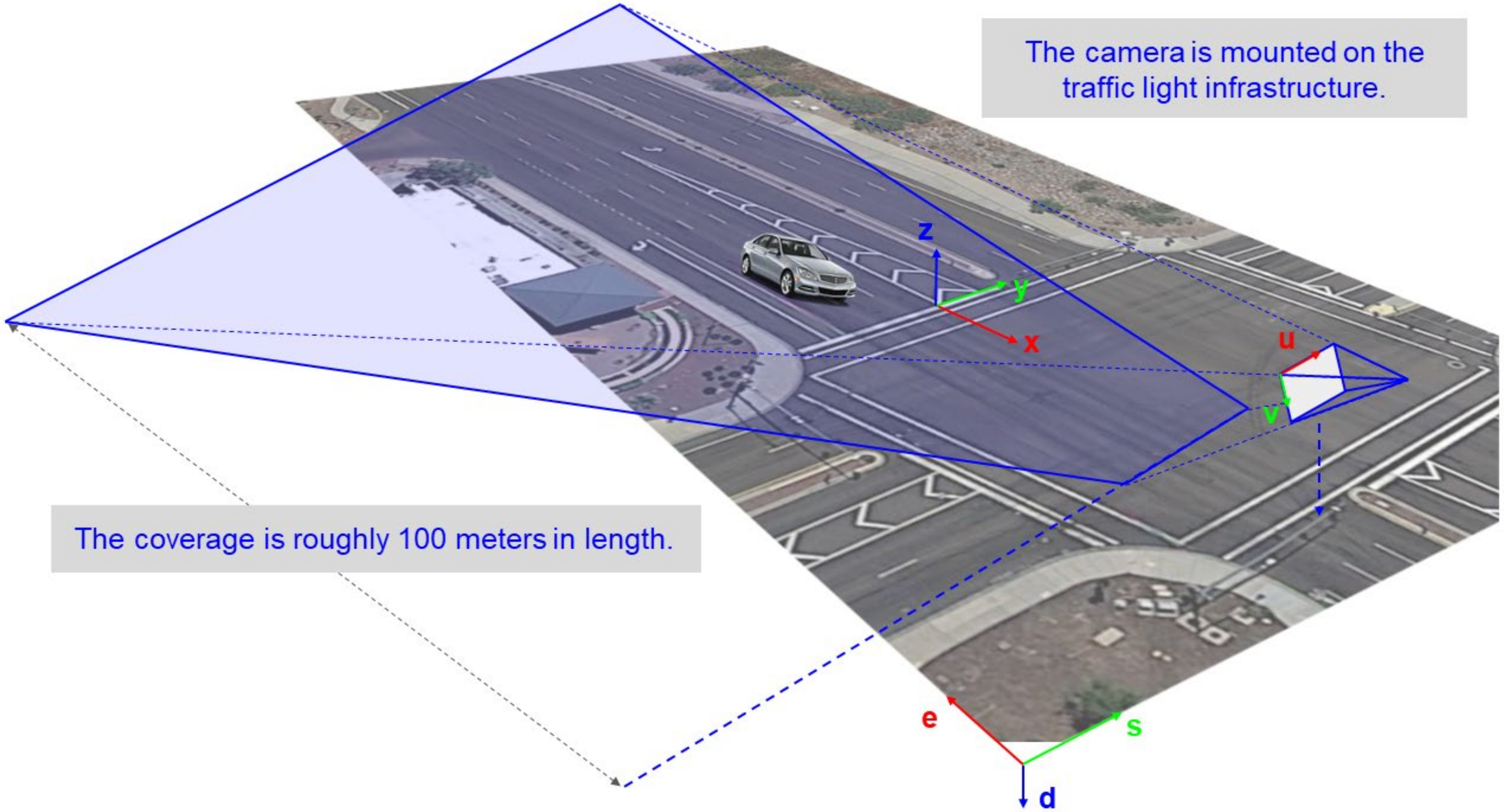


https://interaction-dataset.com/



3) The equipment gap ← AV Performance Monitoring with Existing Traffic Cameras;

Boss (CMU, DARPA Grand Challenge, 2002-2007)

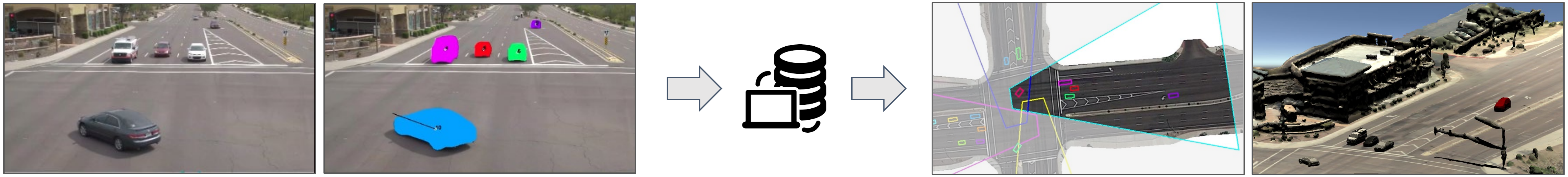


Institute of Automated Mobility
Shaping the future of transportation safety, science, and policy

AV Perf. Metrics I II III;
AVACAR: Avatar of Cars;

CAROM - CARs On the Map

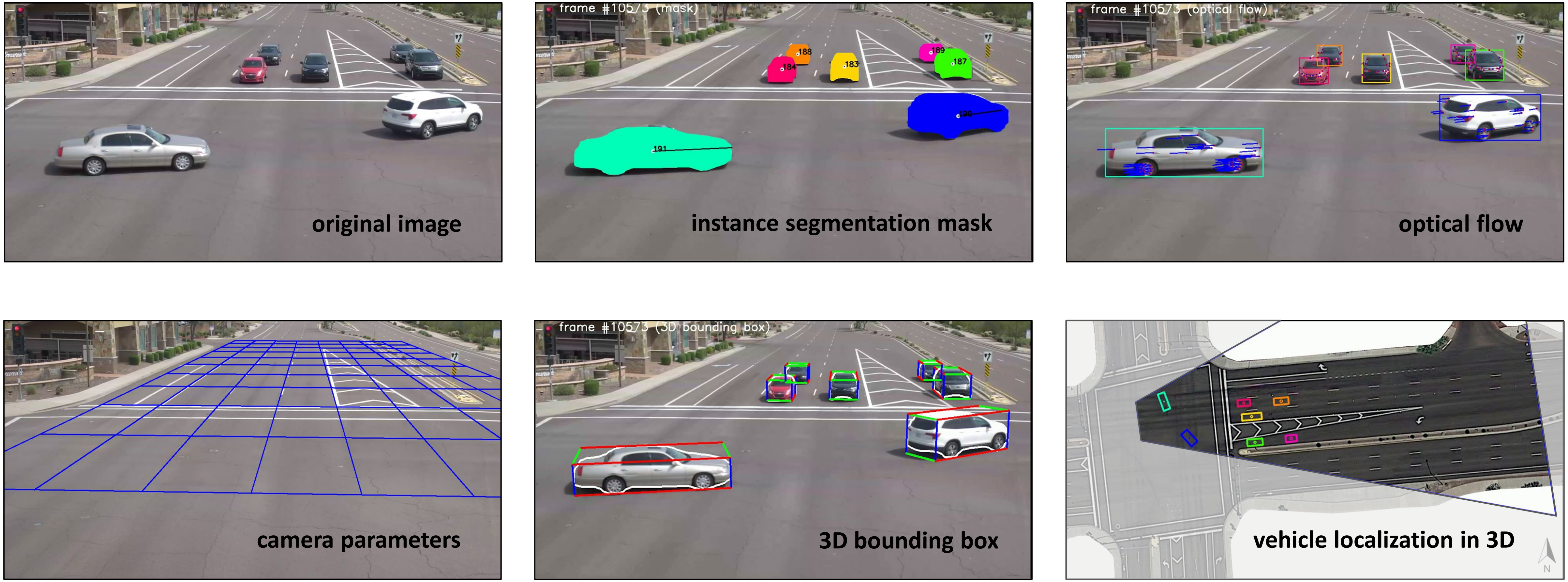
- **CAROM** is a framework to track and localize vehicles using monocular traffic monitoring cameras on road infrastructures.



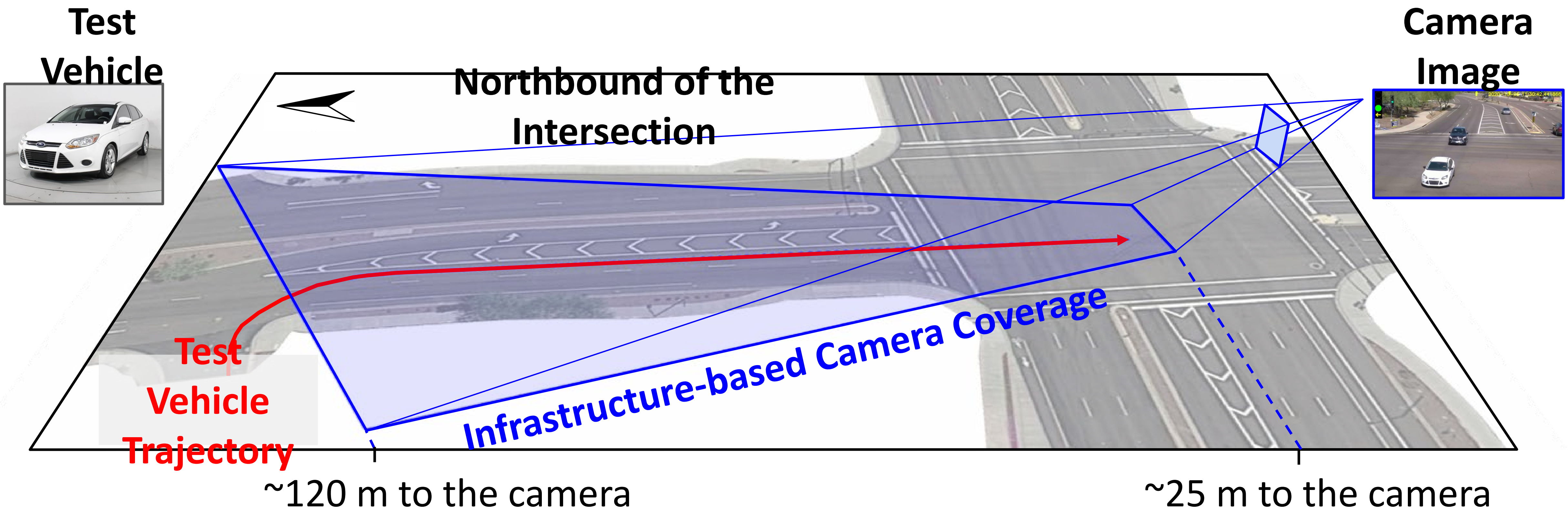
original video tracked vehicles data records replay in 2D replay in 3D

- The vehicle localization results are stored in files or in a database as records.
- Using the results, a traffic scene can be reconstructed and replayed on a map.

Vehicle Tracking

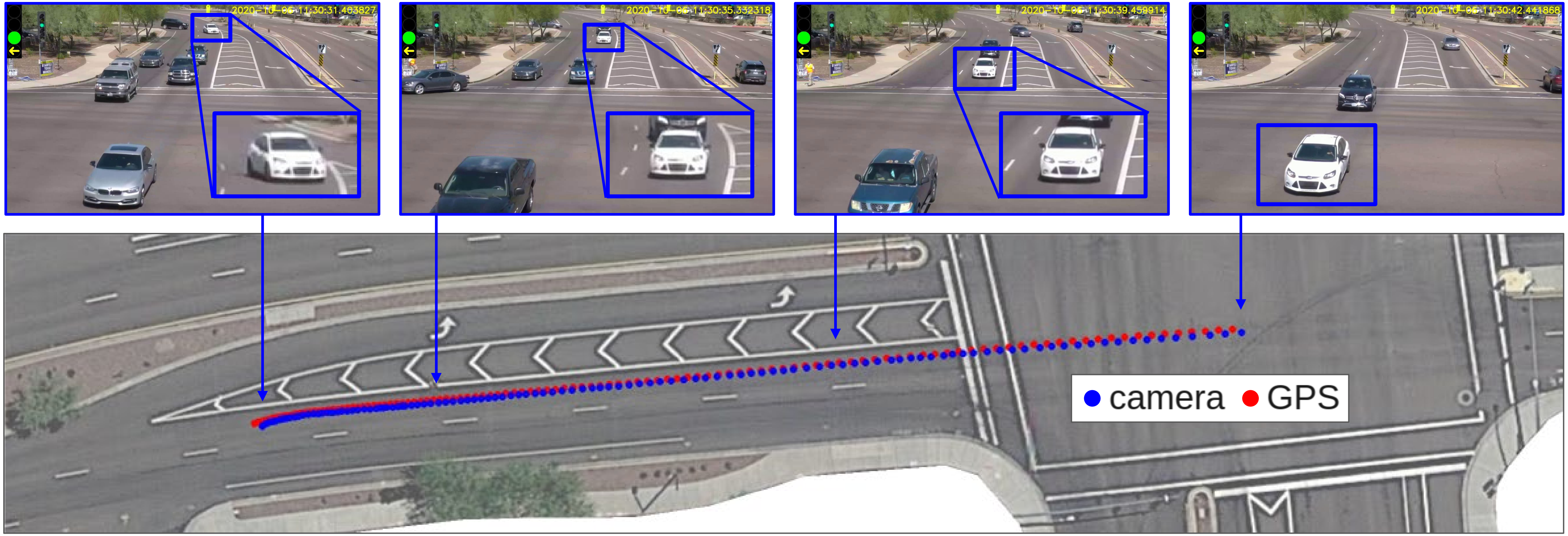


Evaluation - GPS

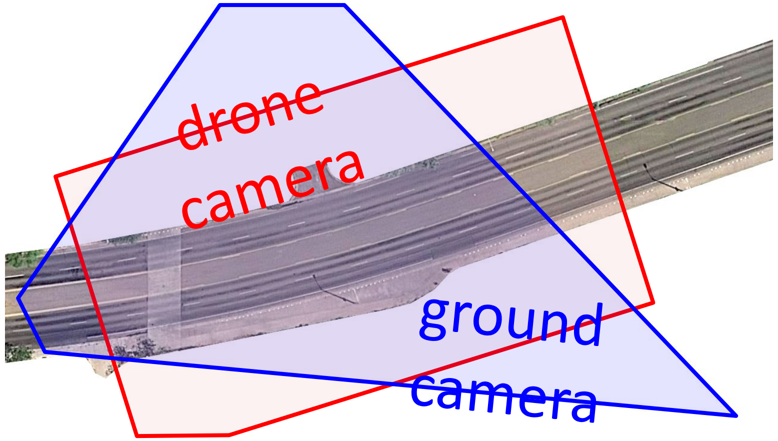


We drove a test vehicle with differential GPS in the first site for evaluation.

Evaluation - GPS

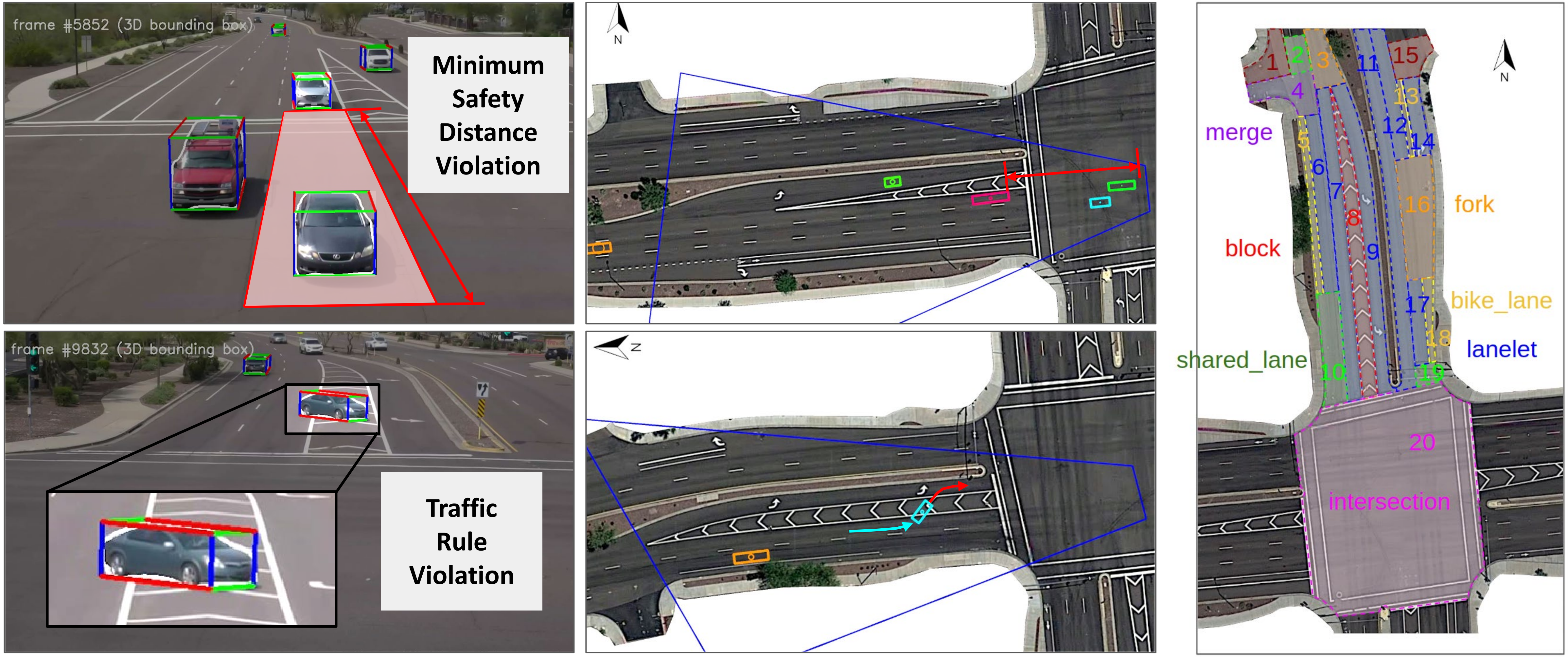


Evaluation - Drone

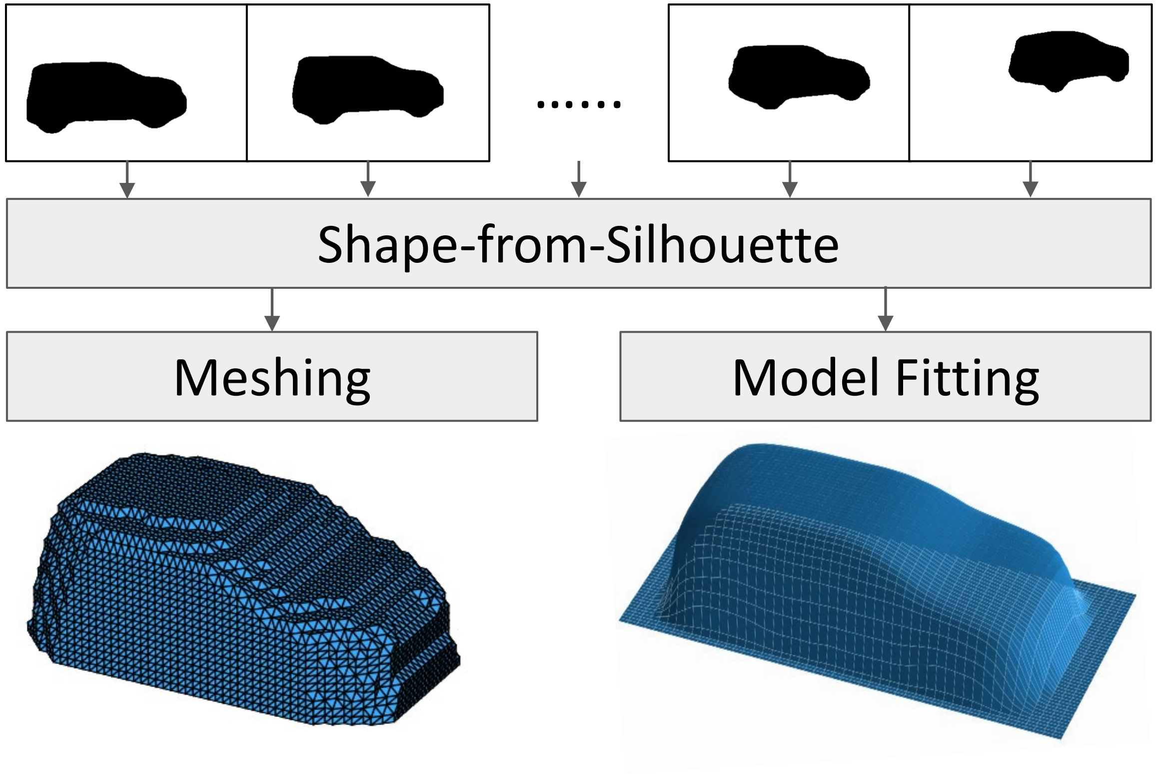


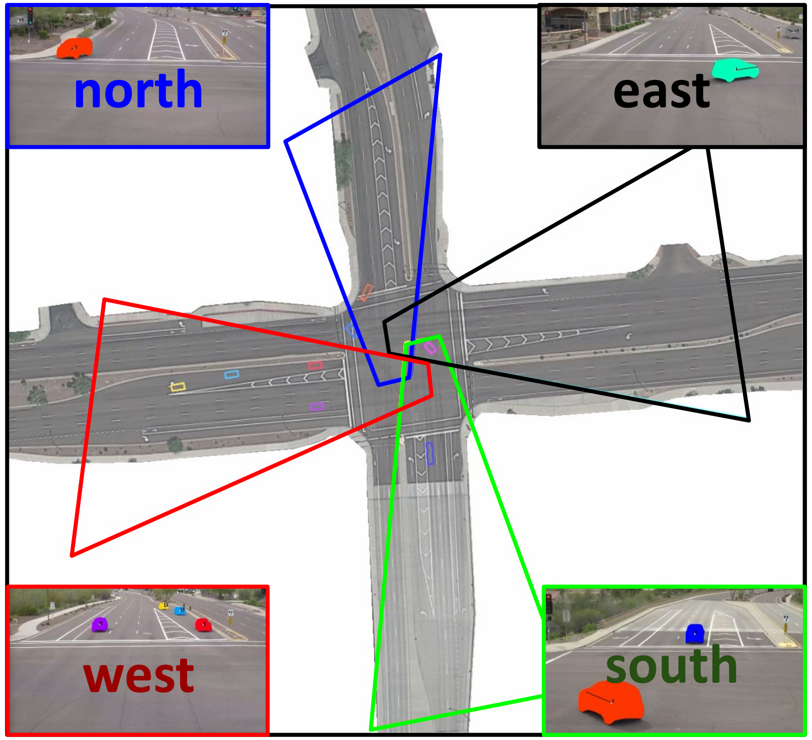
Video	L-Diff (m)	V-Diff (m/s)	#Vehicles (w/ Ref)	Coverage (m)	Ref Device
Track 1A	2.05	1.01	1	25 ~120	GPS
Track 1B	1.57	0.69	1	25 ~120	GPS
Track 2	1.68	1.47	69	15 ~110	Drone

Applications

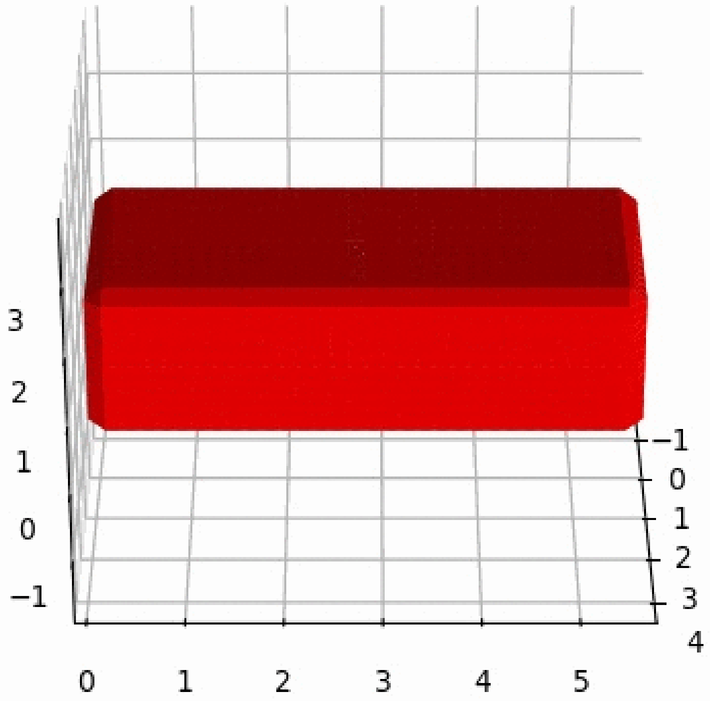


AvaCAR: avatar of vehicles

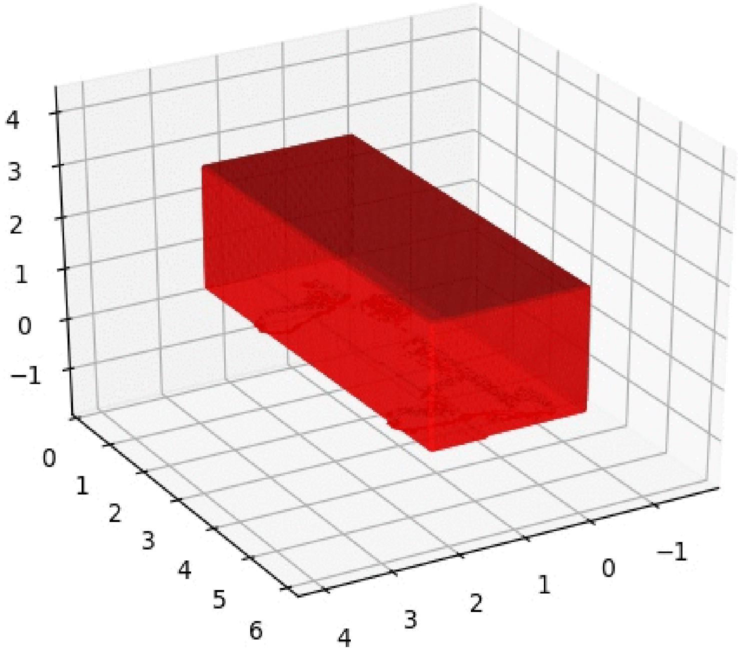




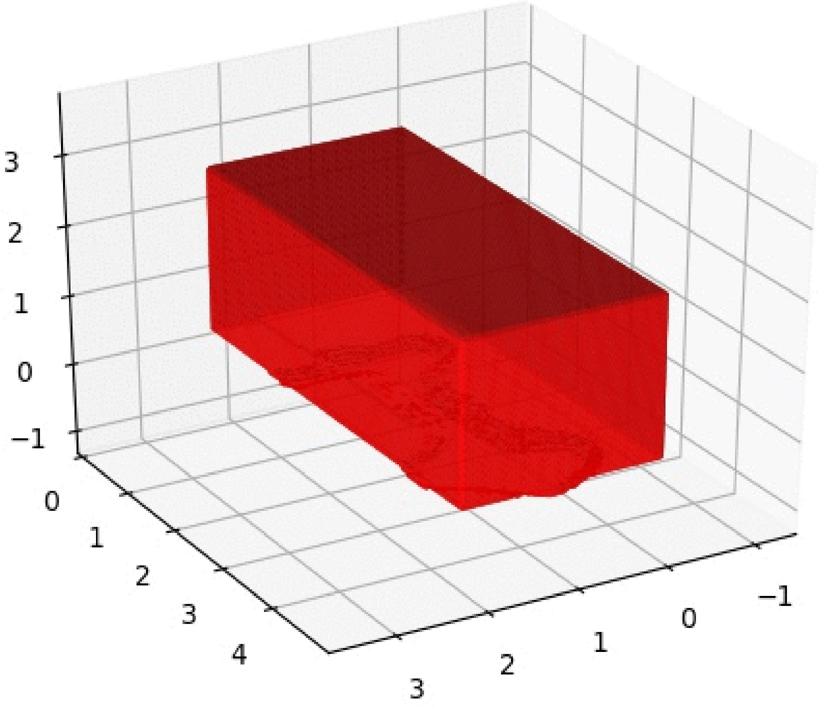
Chrysler Pacifica (Minivan)



Pick-up Truck



Nissan Rogue (SUV)



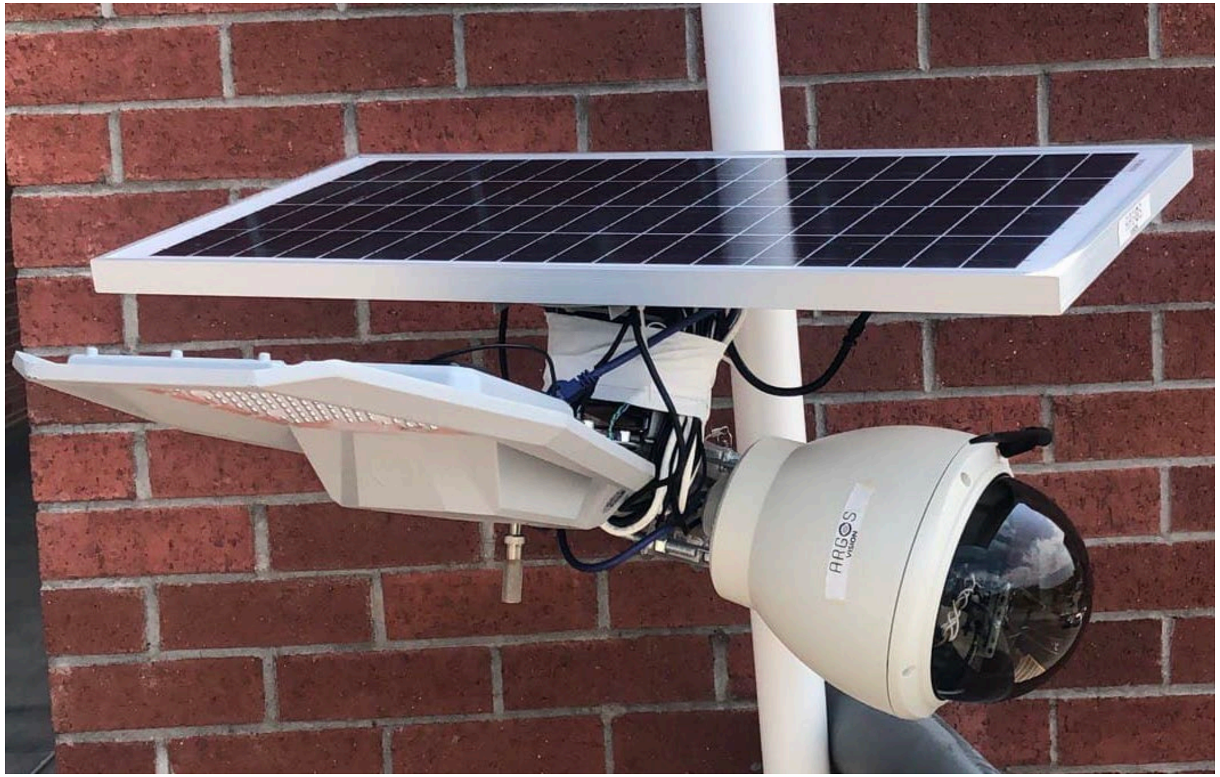
CAROM - Vehicle Localization and Traffic Scene Reconstruction from Monocular Cameras on Road Infrastructures

Demo Video Submission

Duo Lu[1], Varun C Jammula[1], Steven Como[1], Jeffrey Wishart[2],
Yan Chen[1], Yezhou Yang[1]

[1]{duolu, vjammula, scomo, yanchen, yz.yang}@asu.edu
[2]jwishart@exponent.com

4) From tech to transportation practitioner gap ← ARGOS Vision.



ARGOS project provides a full stack software + hardware intelligent camera solution with on-board CV/AI processing that performs semantic-level understanding of the environment and generate a vast amount of privacy-preserved, real-time, semantic **DATA.**

NSF I-CORPS NSF Innovation Corps

National Security Academic Accelerator (NSA2)

ARGOS VISION

ARGOS VISION

A data venture of visual analytics
with intelligent cameras



Mohammad Farhadi   Yezhou Yang   Ryan Kemmet

Thank you and Acknowledgements

NSF CAREER 18’ VR-K

2 NSF RI SMALLs

NSF NRI

NSF CPS

NSF SaTC

NSF CCRI (planning)

NSF I-Corps

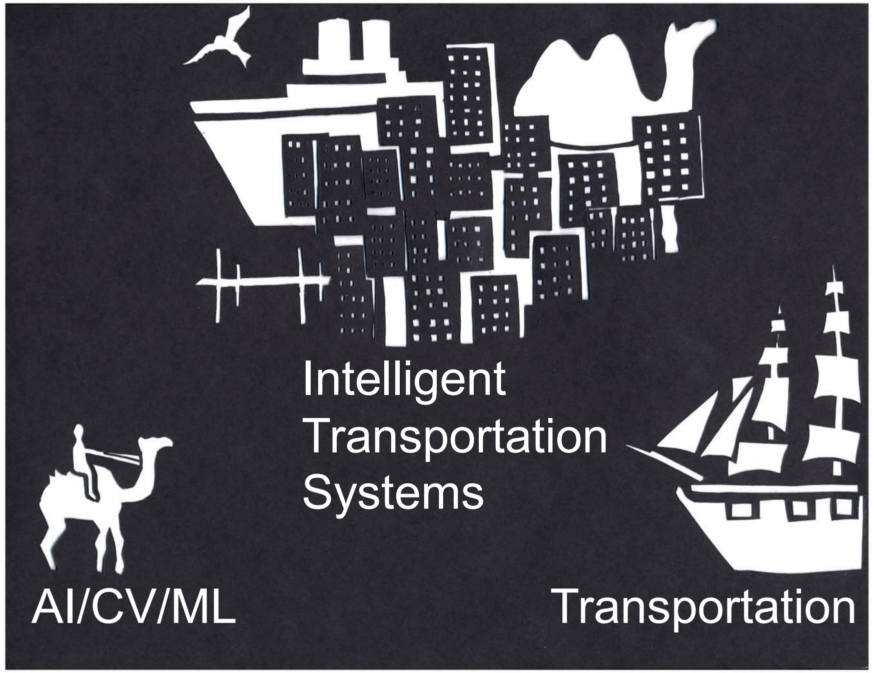
DARPA KAIROS
LESTAT project
And
GAILA ADAM-E

ONR Social
Interaction

Machine Learning
Research Award 19’

… …

and ASU close collaborating groups (C. Baral [KR & NLP], M. Ren/W. Zhang [Optimization & ML & Control], IAM collaborators: Jeff Wishart, Duo Lv, Mohammad Farhadi, , Maria Elli, Yan Chen, Larry Head, Greg Leeming, Prabal Dutta, Rahul Varma and many more).



ASU APG

Public/Business/Academia to APG: yz.yang@asu.edu

@Yezhou_Yang

ARGOS VISION

Public/Business/Academia to ARGOS: argos.vision.co@gmail.com
www.argos.vision

Check out our live demo @ ITS AZ!