EPFL
VITA

Visual Intelligence for Transportation

HONDA The Power of Dreams

Schindler

RICHEMONT

CarPostal La classe jaune.

HITACHI

SAMSUNG

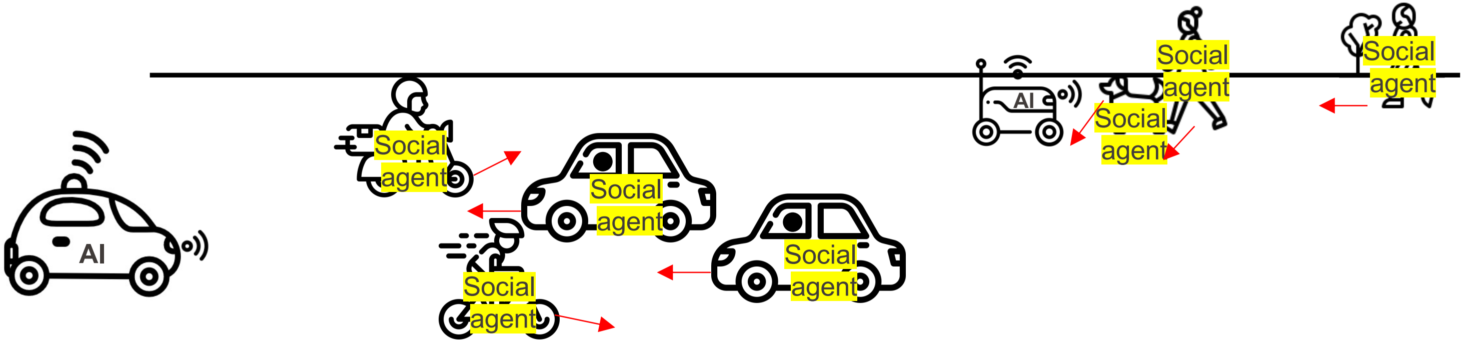
Valeo

Innosuisse

FNSNF

HASLERSTIFTUNG

Horizon 2020

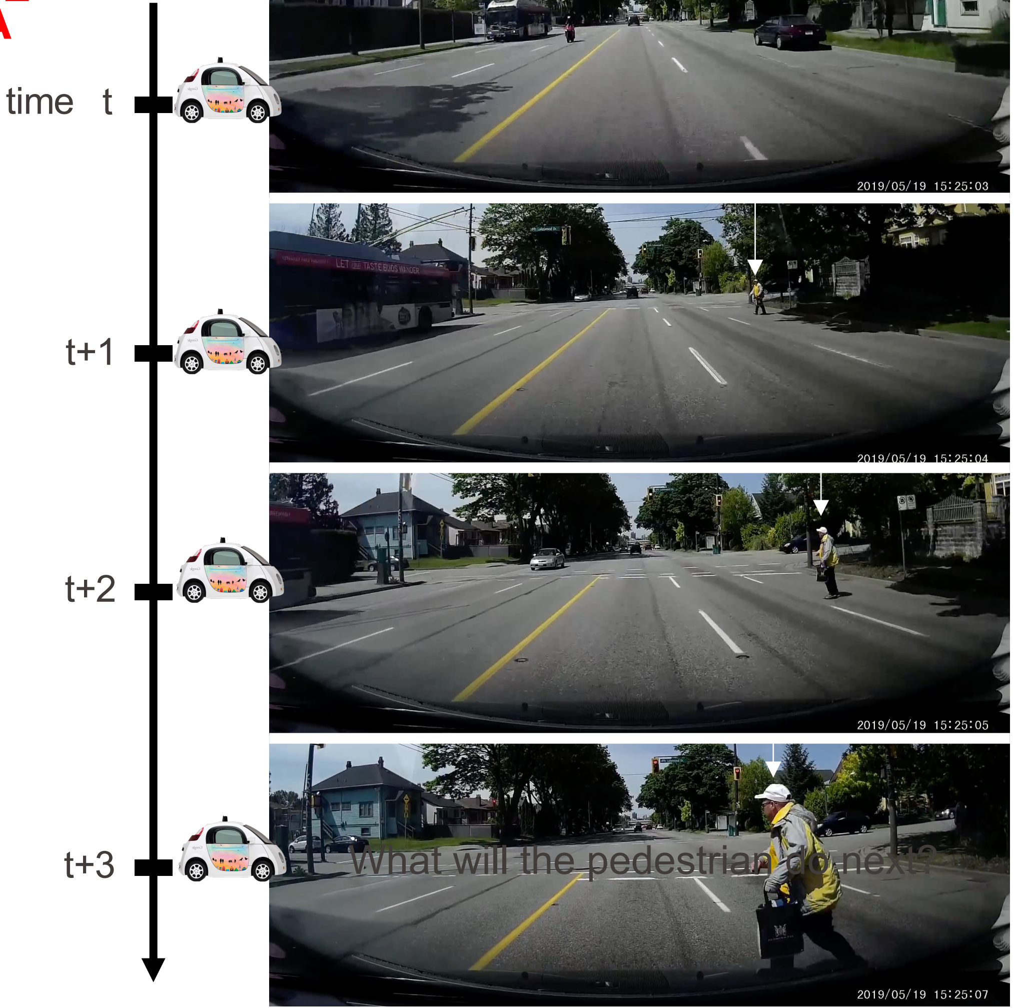


“Humans subconsciously **forecast the future**…
Autonomous Vehicles must have the same **forecasting** capability to **harmlessly** and **effectively co-exist**”,

Our lab goal.

Forecasting is essential…

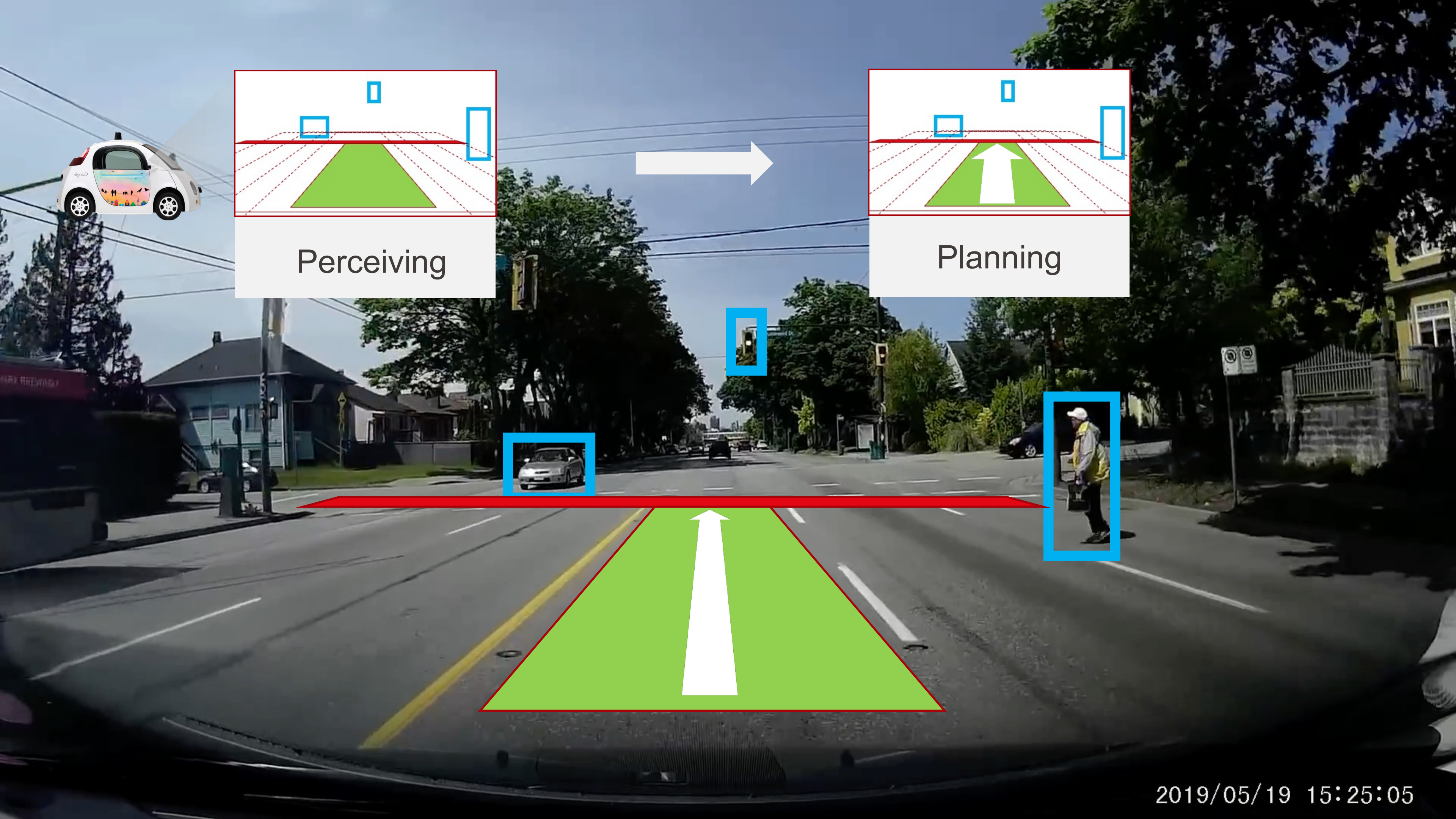
Mountain View, 2019/05/19



What will the pedestrian do next?

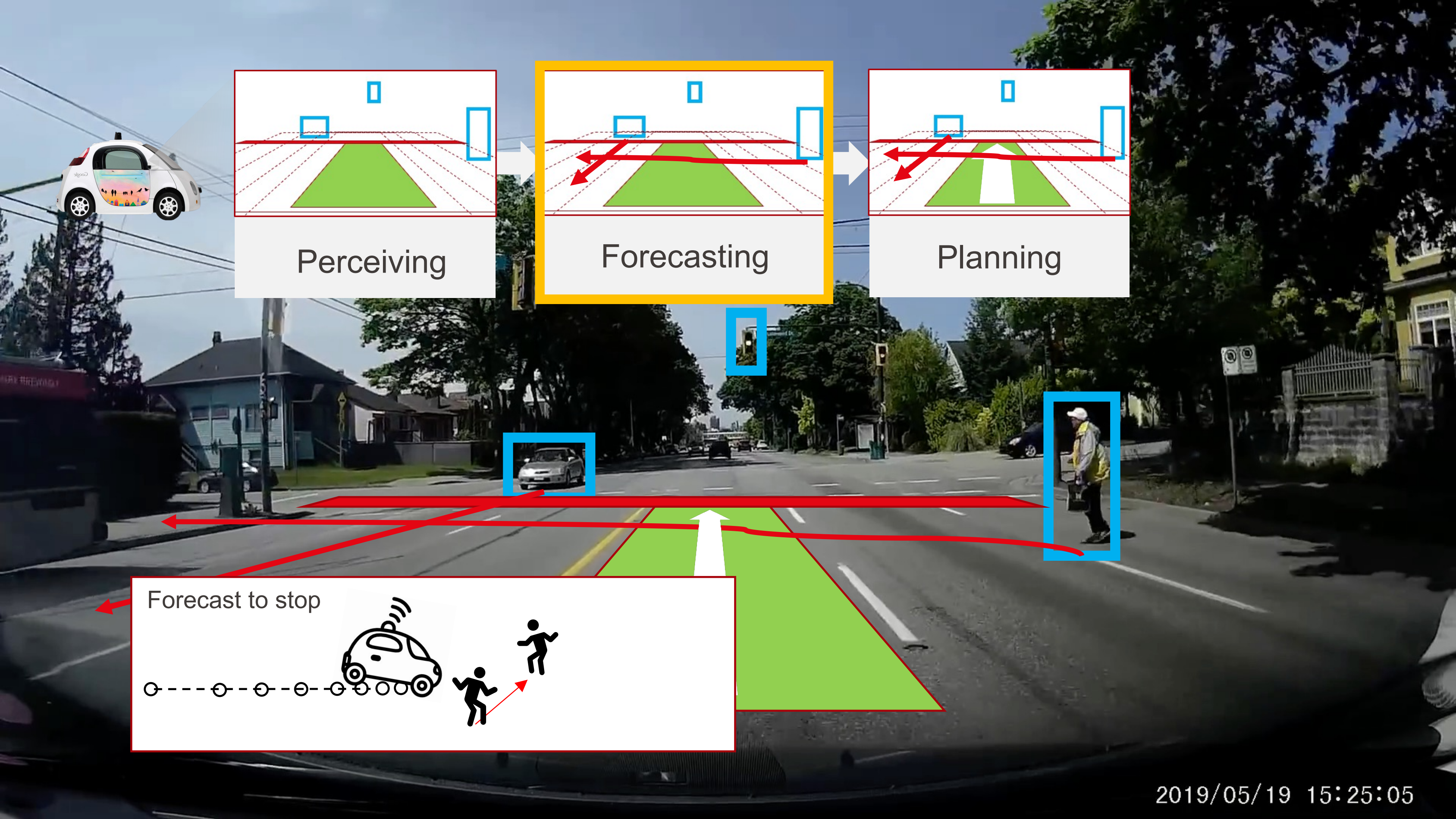
Autonomous ✓

Autonomous ✓



Perceiving

Planning



Perceiving

Forecasting

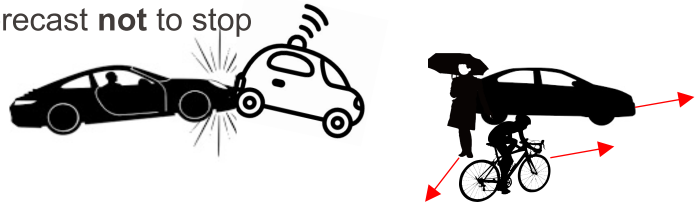
Planning

Forecast to stop

AI must forecast agent-agent* **interactions** = **Social Forecasting**

*agent = any moving entity in the world (driver, pedestrian, cyclist…)

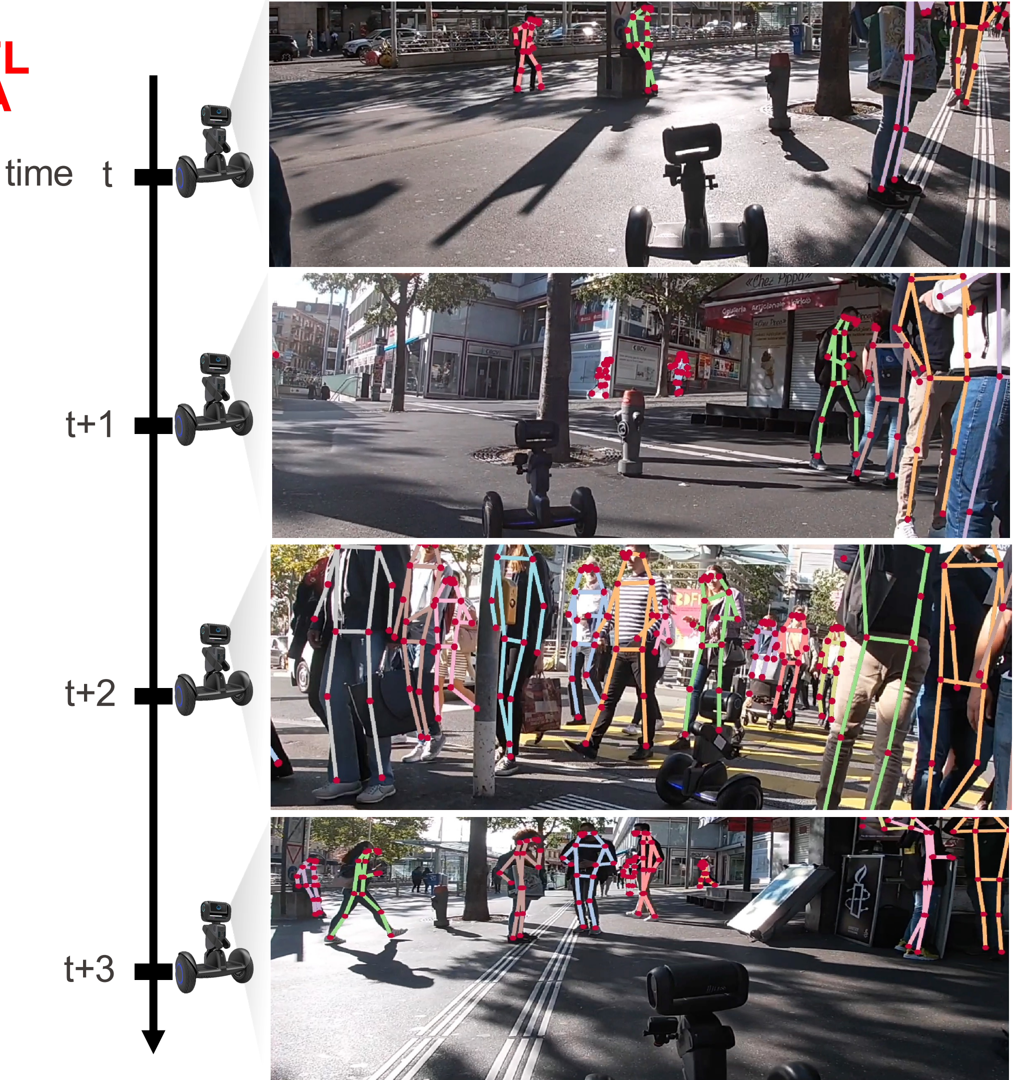
Forecast **not** to stop



98% of AV accidents are due to an unexpected stop*

* Favarò,F., et al., "Examining accident reports involving autonomous vehicles." PLoS one , ‘17

Lausanne, 2020/07/15



Autonomous ✓

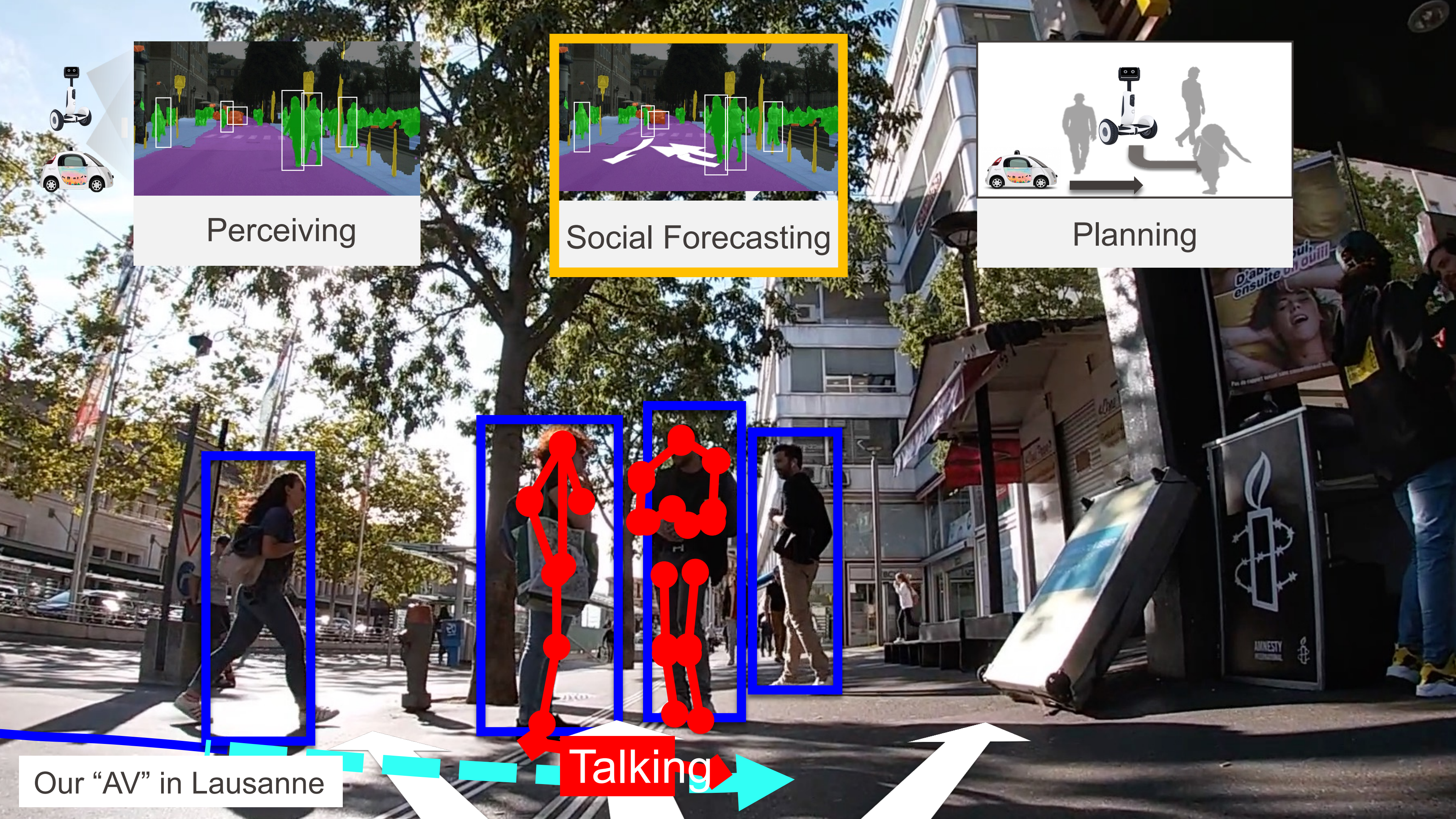
Autonomous ✓

Autonomous ✗

=> Our robot freezes in close human proximity

Autonomous ✗

=> Our robot does not comply with social norms

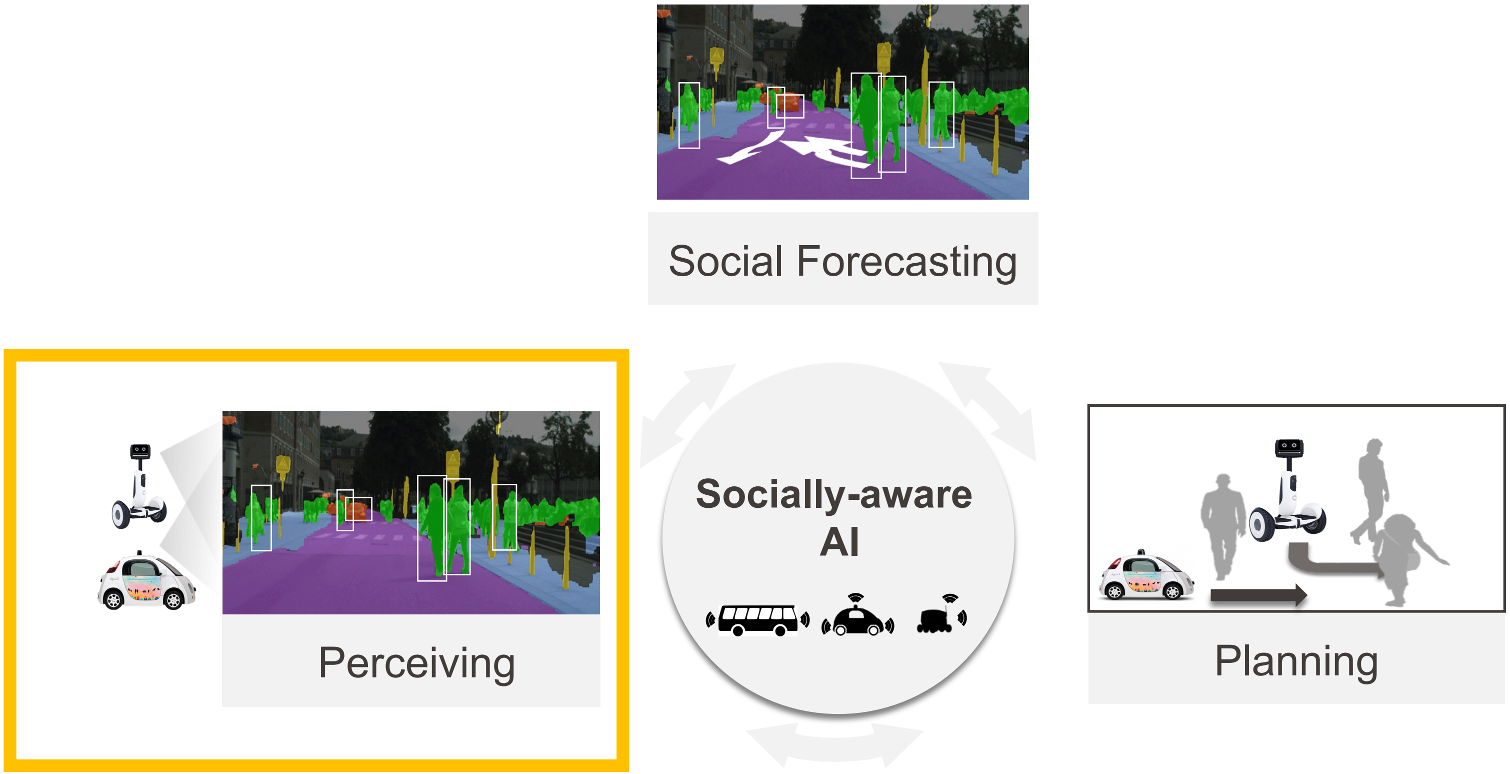


Perceiving

Social Forecasting

Planning

Our “AV” in Lausanne



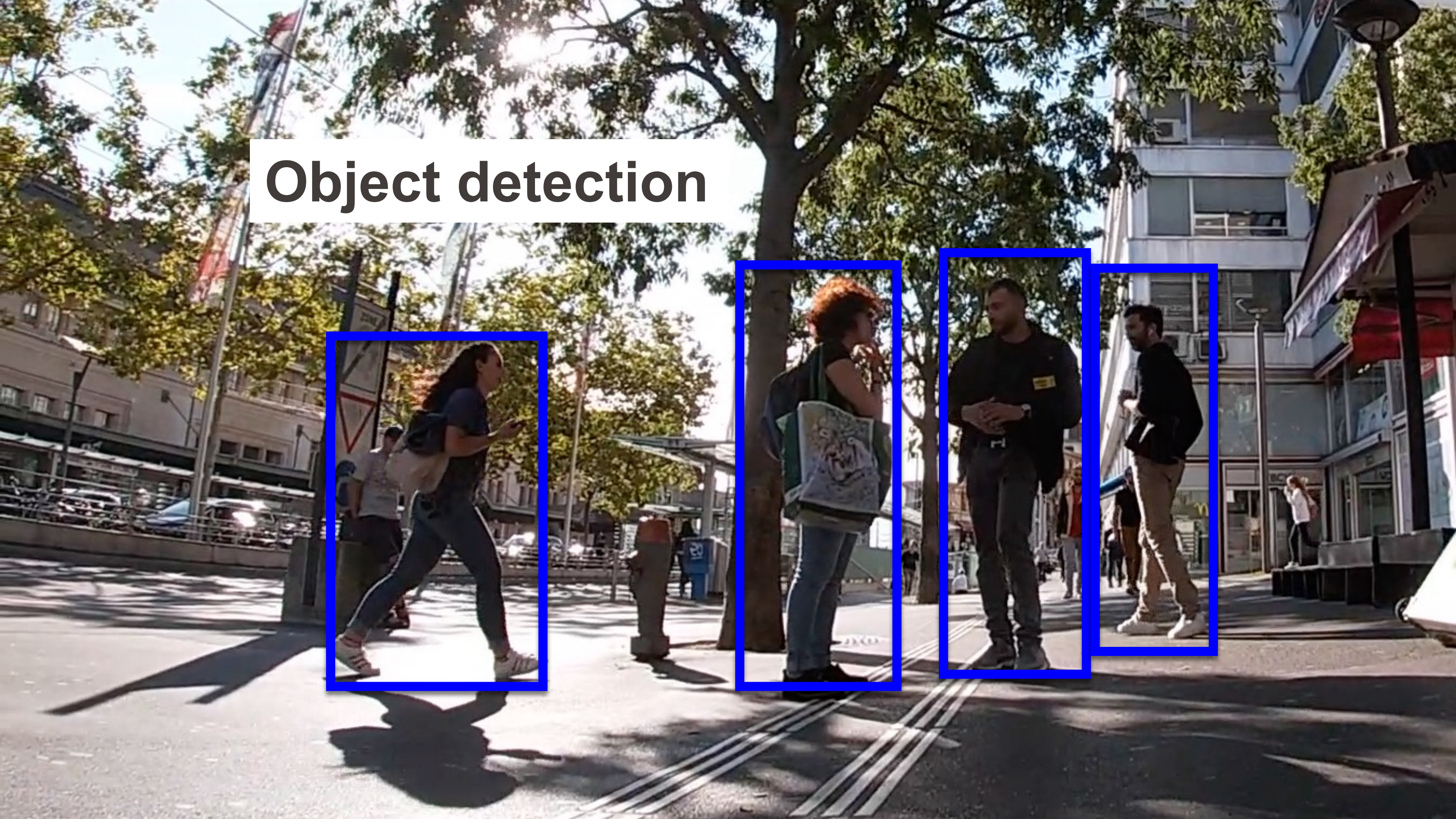
Social Forecasting

Perceiving

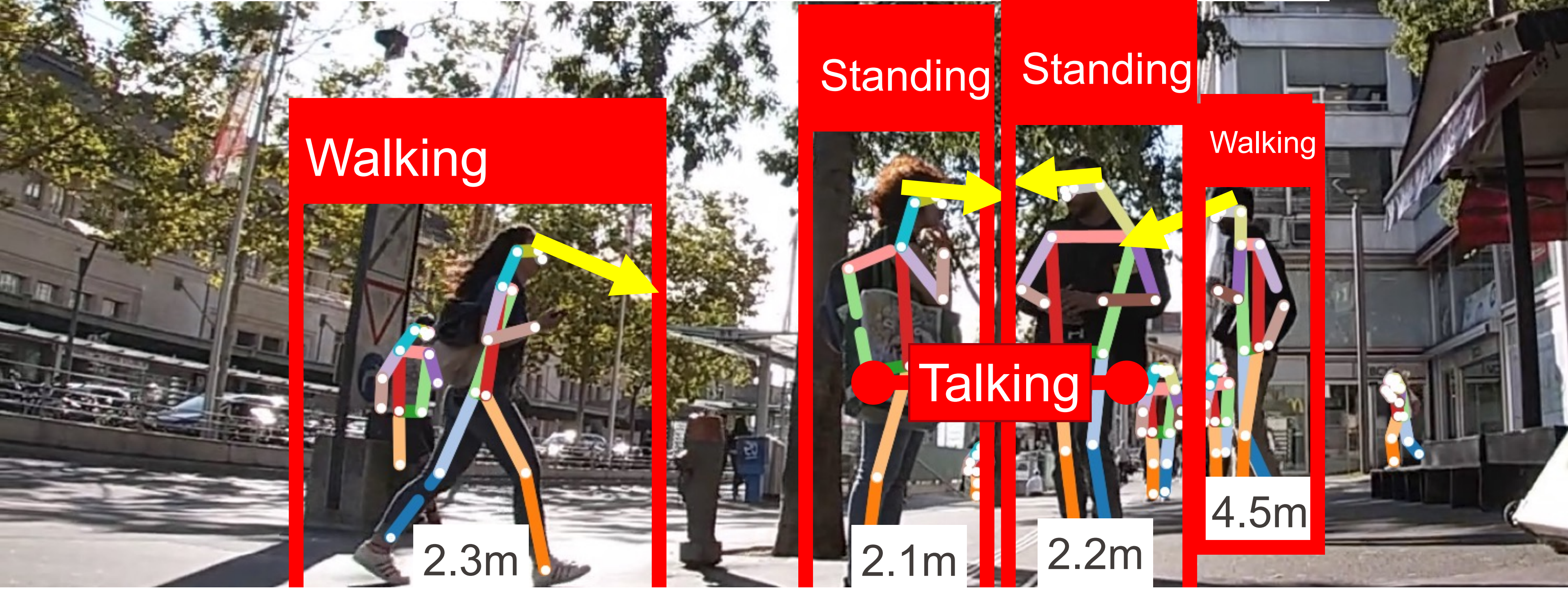
Socially-aware AI

Planning

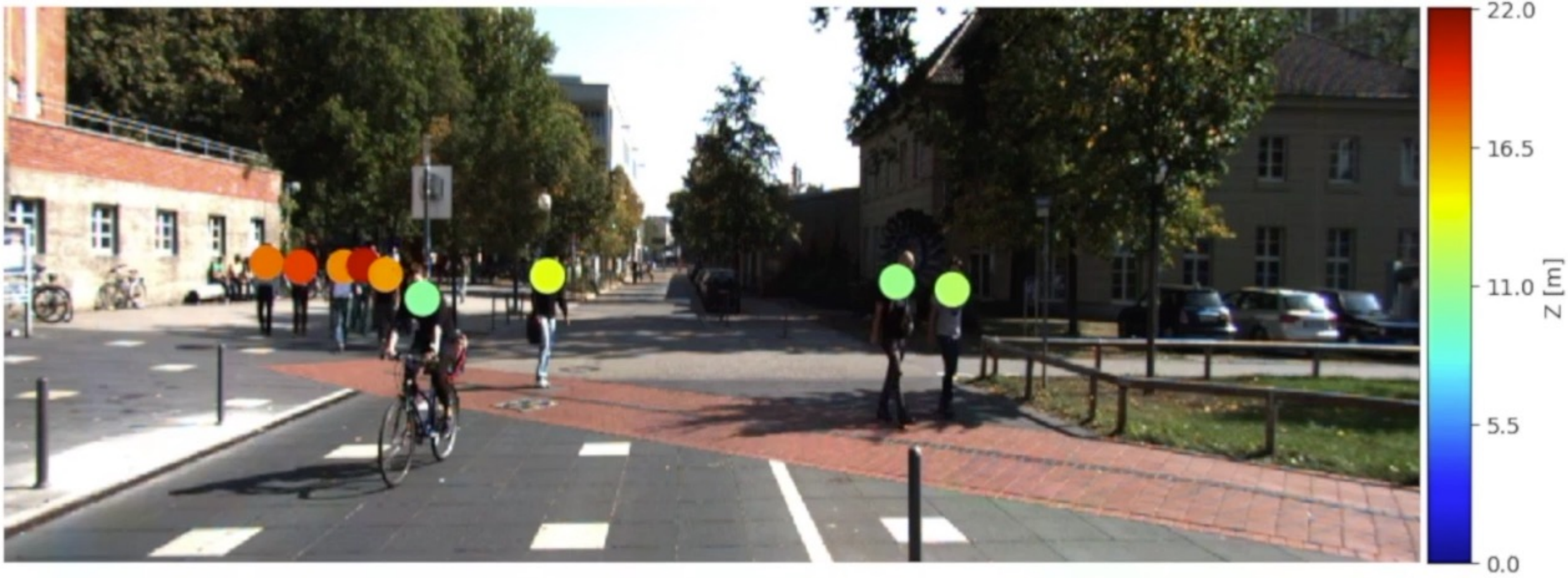
Object detection



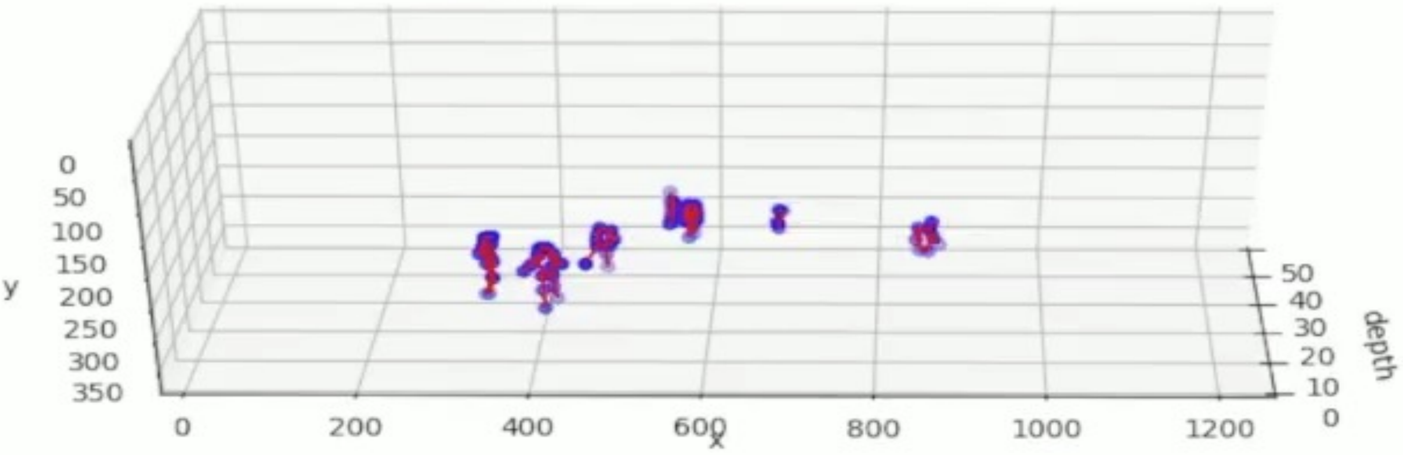
Body poses + Activities + Relationships

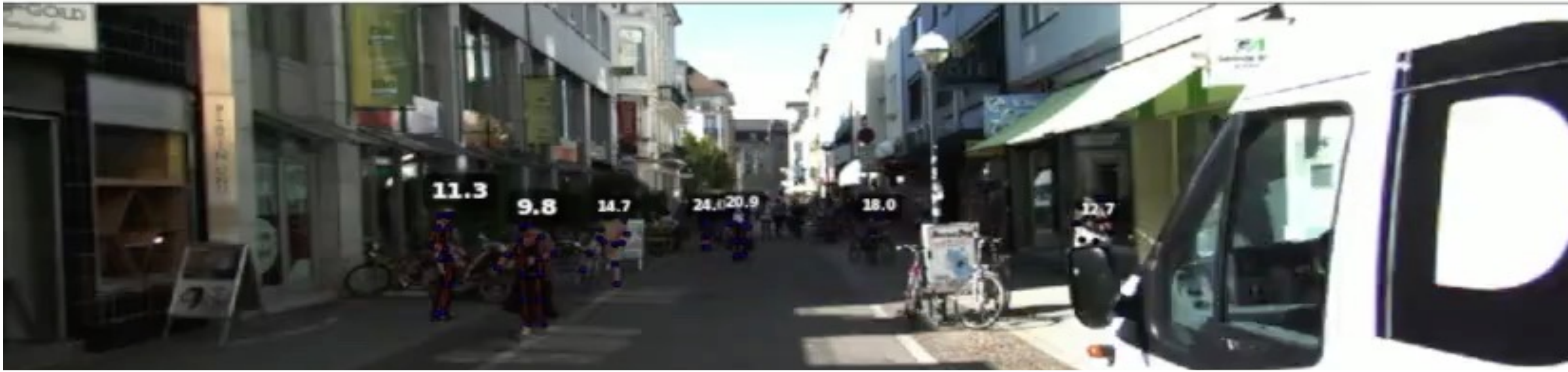


[1] PifPaf: Composite Fields for Human Pose Estimation, **CVPR'19; on-line demo:** vitademo.epfl.ch/movements
[2] Keypoints communities, **ICCV'21;** [3] OpenPifPaf, open-source library, **IEEE ITS'21**
[4] Convolutional Relational Machine, **CVPR'19;** [5] Detecting 32 Pedestrian Attributes, **IEEE ITS'21**
[6] Monoloco: Monocular 3D pedestrian localization and uncertainty estimation, **ICCV'19;** [7] MonStereo, **ICRA'21**

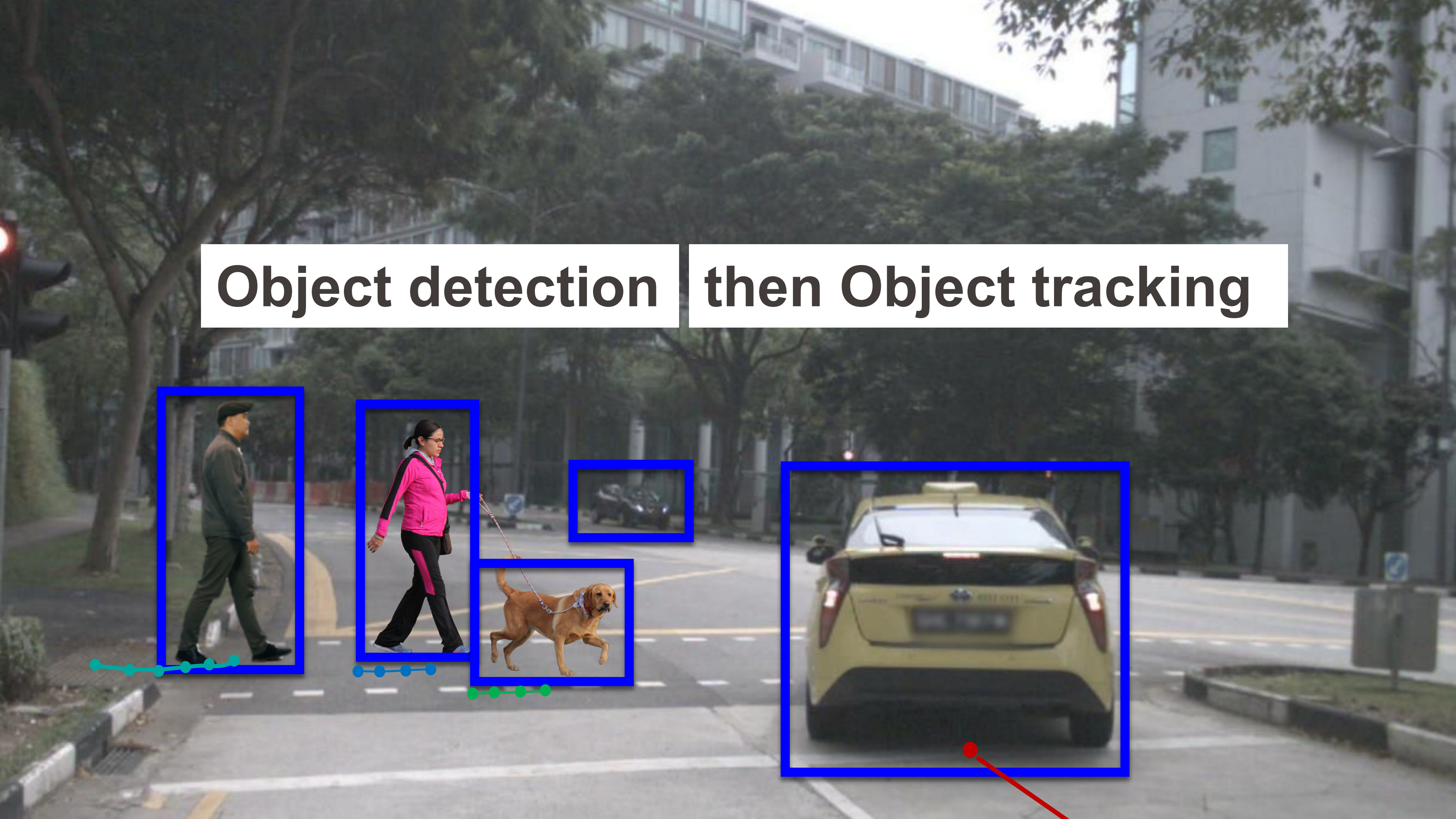


[1] Monoloco: Monocular 3D pedestrian localization and uncertainty estimation, **ICCV'19;** [2] MonStereo, **ICRA'21**



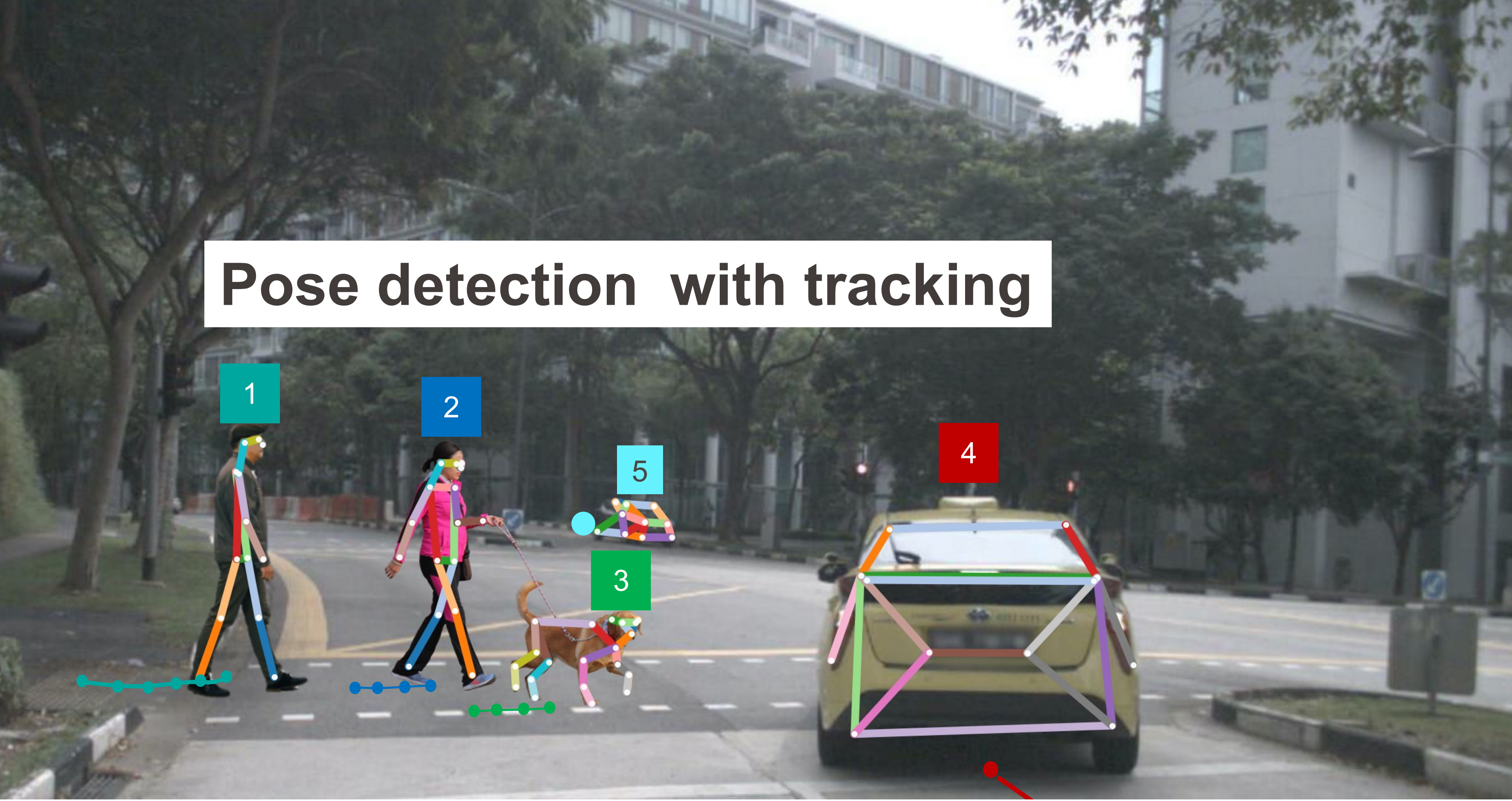


[1] Monoloco: Monocular 3D pedestrian localization and uncertainty estimation, **ICCV'19;** [2] MonStereo, **ICRA'21**

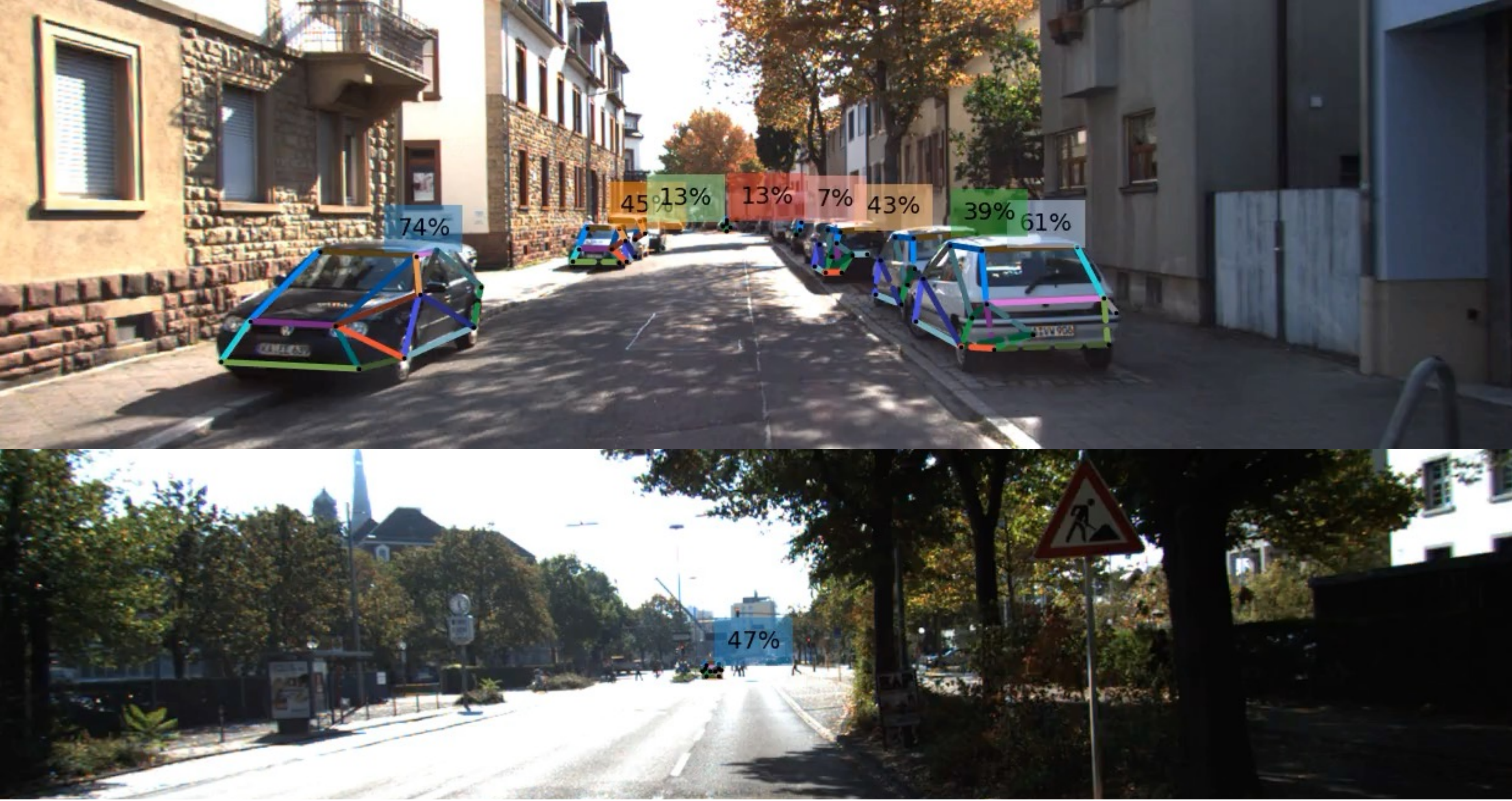


Object detection then Object tracking

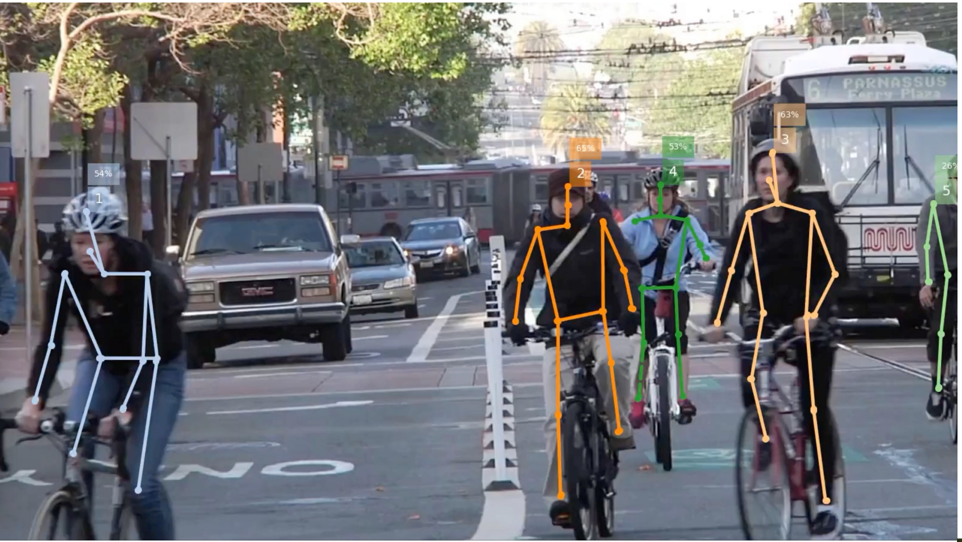
Pose detection with tracking

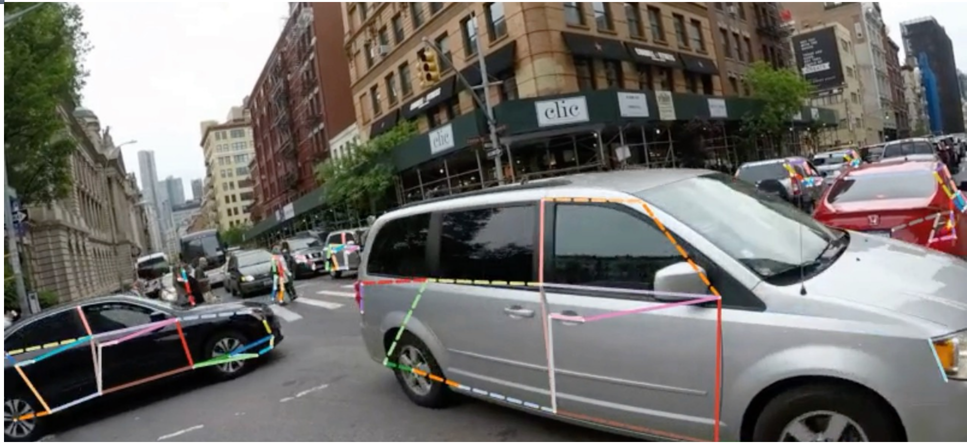


[1] OpenPifPaf, **IEEE transactions on ITS’21** (online demo: https://vitademo.epfl.ch/movements)



[1] OpenPifPaf, **IEEE transactions on ITS'21** (online demo: https://vitademo.epfl.ch/movements)

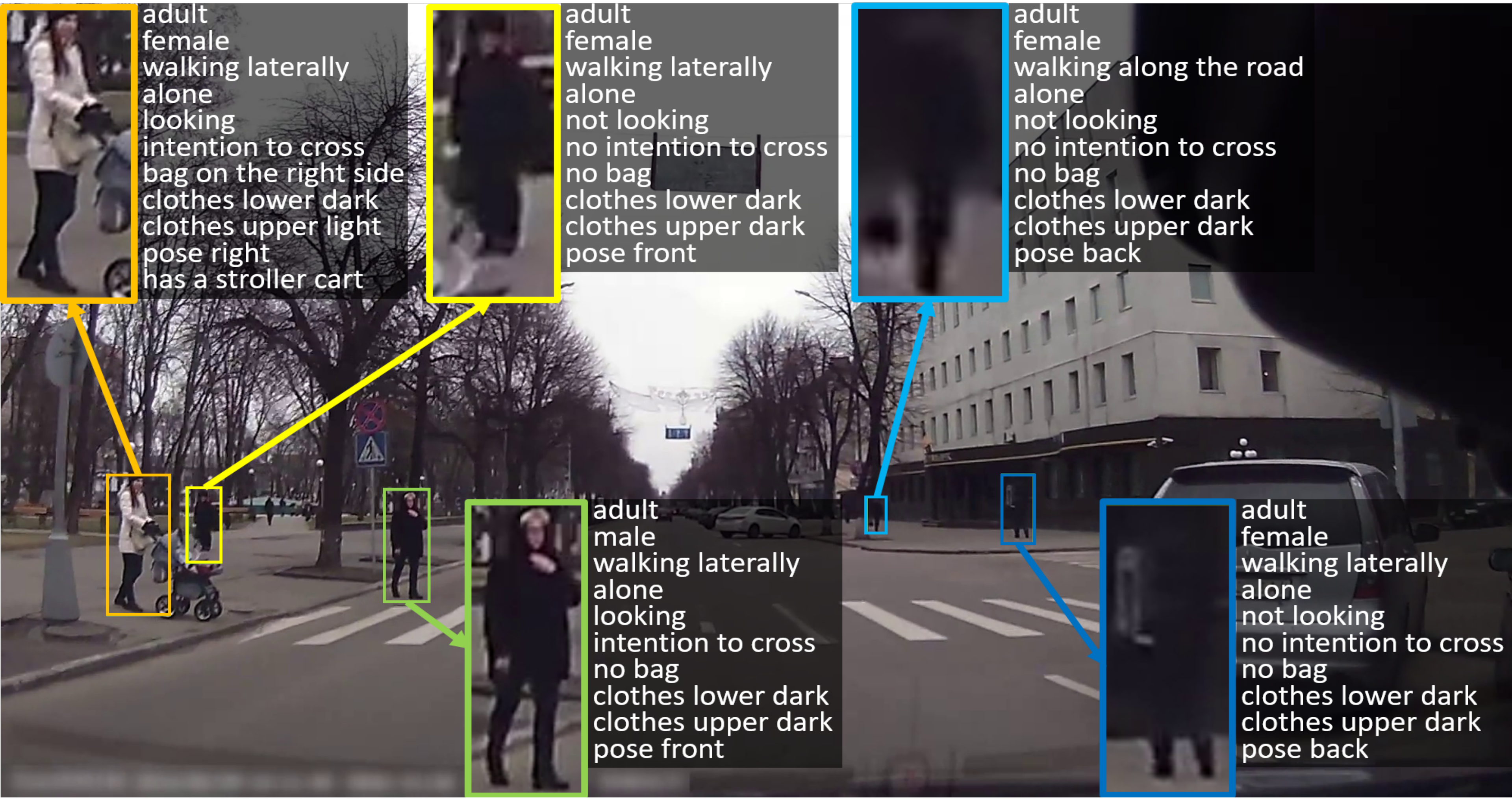




TA

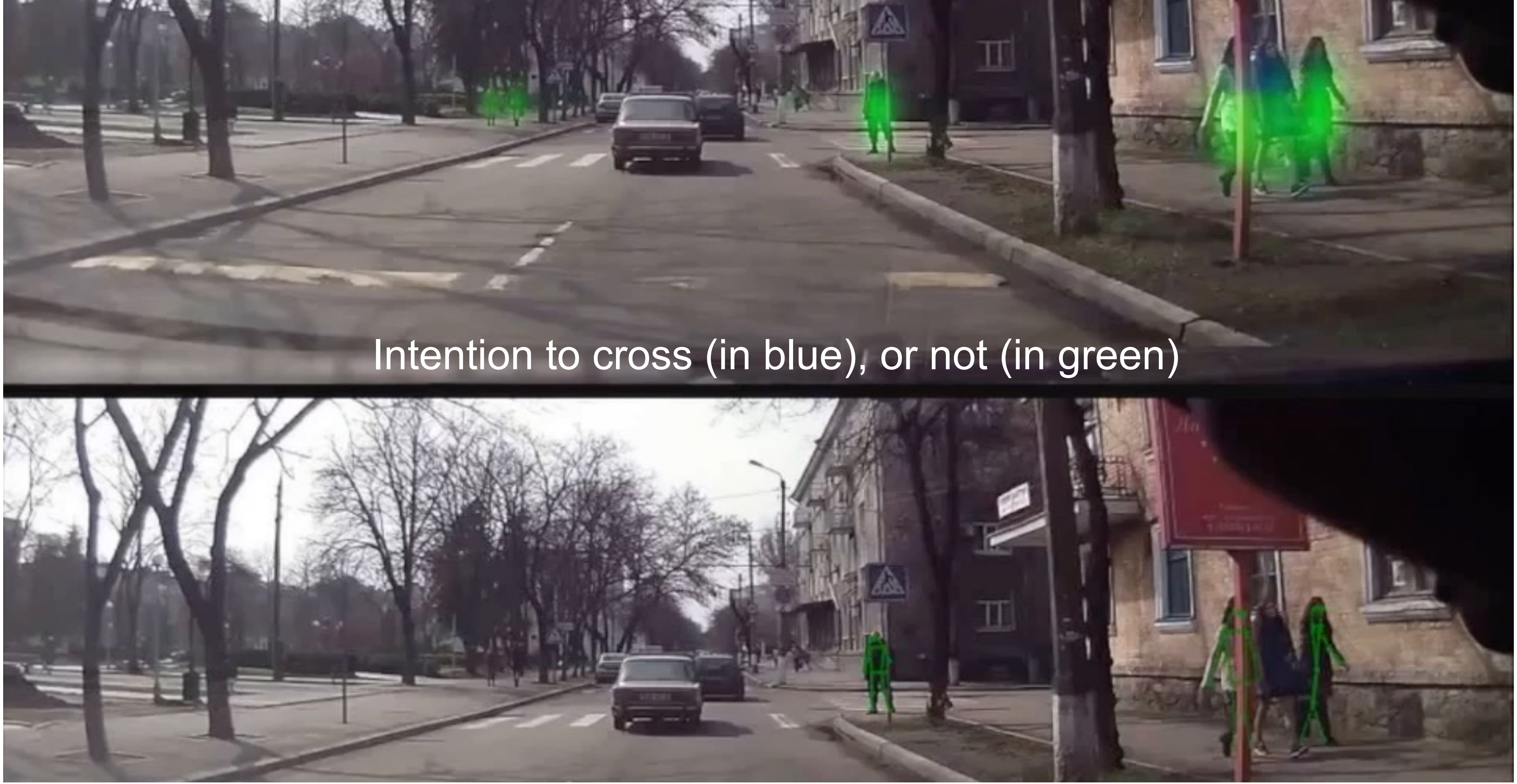
[1] **on-line demo:** vitademo.epfl.ch/movements

[2] Keypoints communities, **ICCV'21;** [3] OpenPifPaf, open-source library**,** **IEEE transactions on ITS'21**



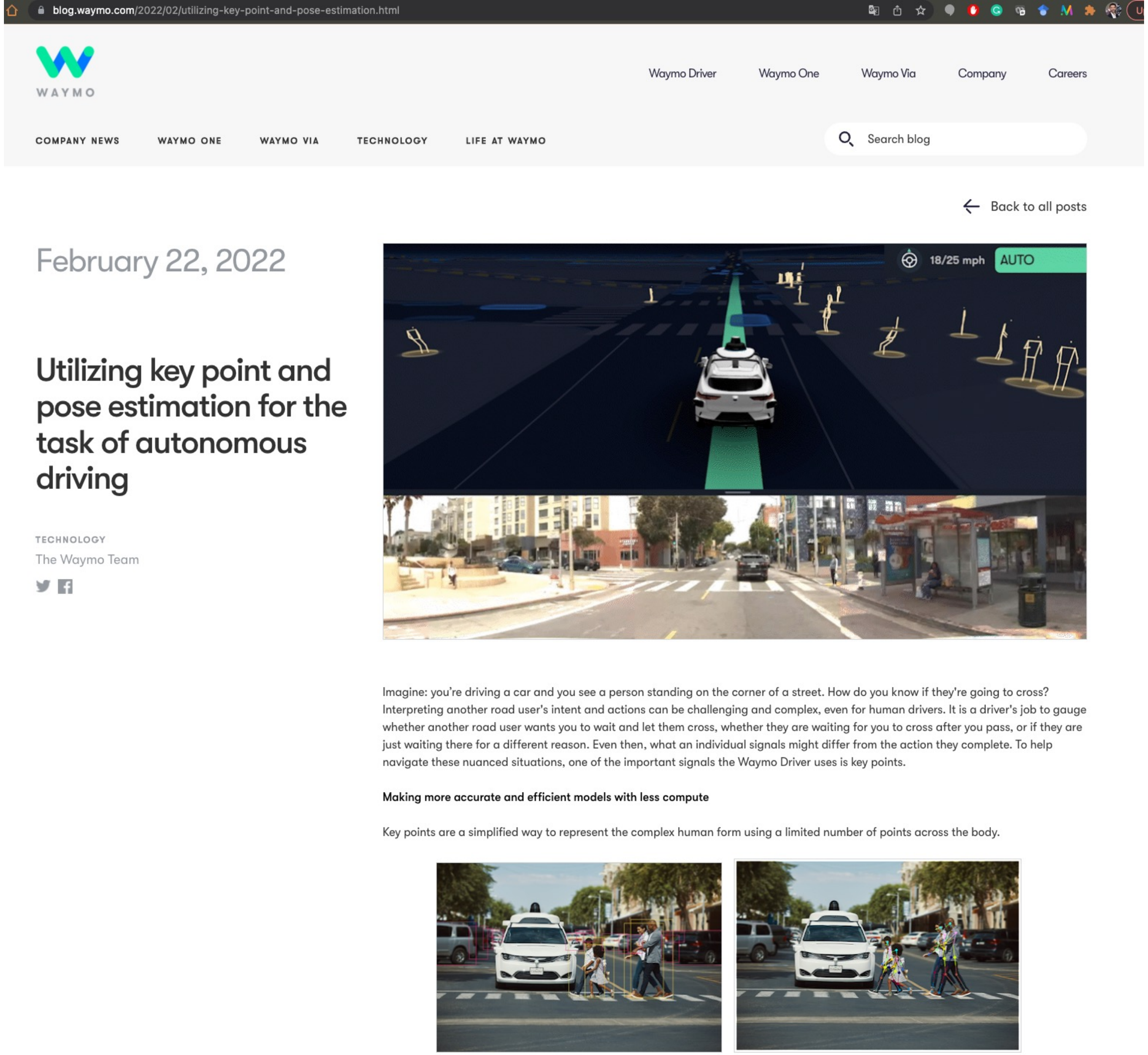
[1] Detecting 32 Pedestrian Attributes for Autonomous Vehicles, **IEEE transactions on ITS’21**

[1] Looking dataset: https://looking-vita-epfl.github.io/



Intention to cross (in blue), or not (in green)

[1] Pedestrian Intention Prediction: A Convolutional Bottom-Up Multi-task Approach, **TRC'21**



blog.waymo.com/2022/02/utilizing-key-point-and-pose-estimation.html

WAYMO

Waymo Driver Waymo One Waymo Via Company Careers

COMPANY NEWS WAYMO ONE WAYMO VIA TECHNOLOGY LIFE AT WAYMO

Search blog

Back to all posts

February 22, 2022

Utilizing key point and pose estimation for the task of autonomous driving

TECHNOLOGY

The Waymo Team

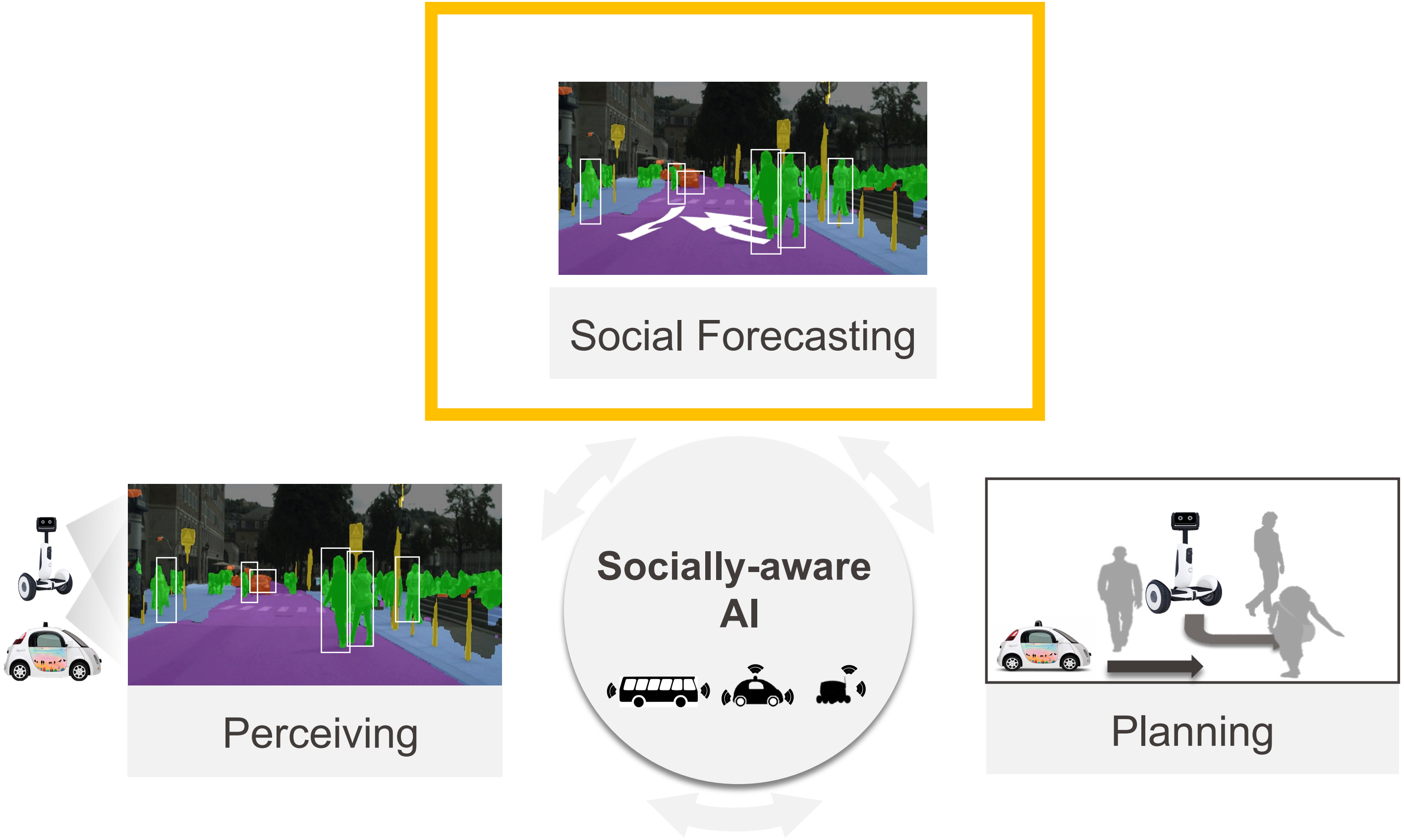
Imagine: you're driving a car and you see a person standing on the corner of a street. How do you know if they're going to cross? Interpreting another road user's intent and actions can be challenging and complex, even for human drivers. It is a driver's job to gauge whether another road user wants you to wait and let them cross, whether they are waiting for you to cross after you pass, or if they are just waiting there for a different reason. Even then, what an individual signals might differ from the action they complete. To help navigate these nuanced situations, one of the important signals the Waymo Driver uses is key points.

Making more accurate and efficient models with less compute

Key points are a simplified way to represent the complex human form using a limited number of points across the body.

Alexandre Alahi, EPFL

VITA



Social Forecasting

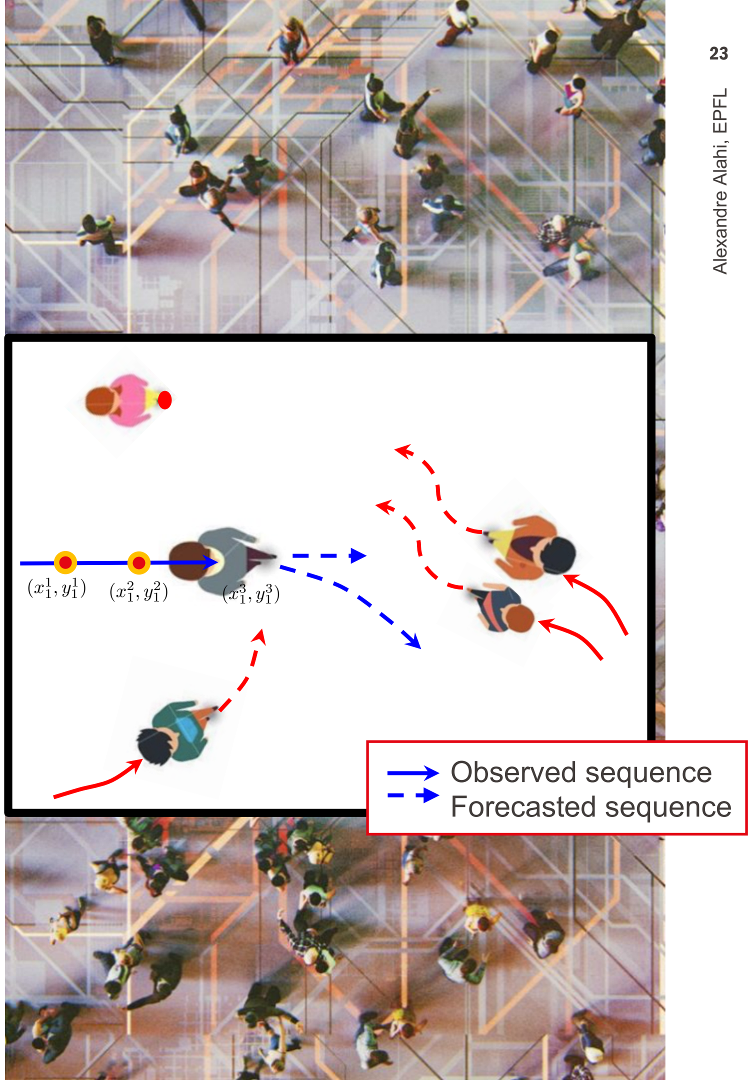
Perceiving

Socially-aware AI

Planning

Social Forecasting
(w/ pedestrians)

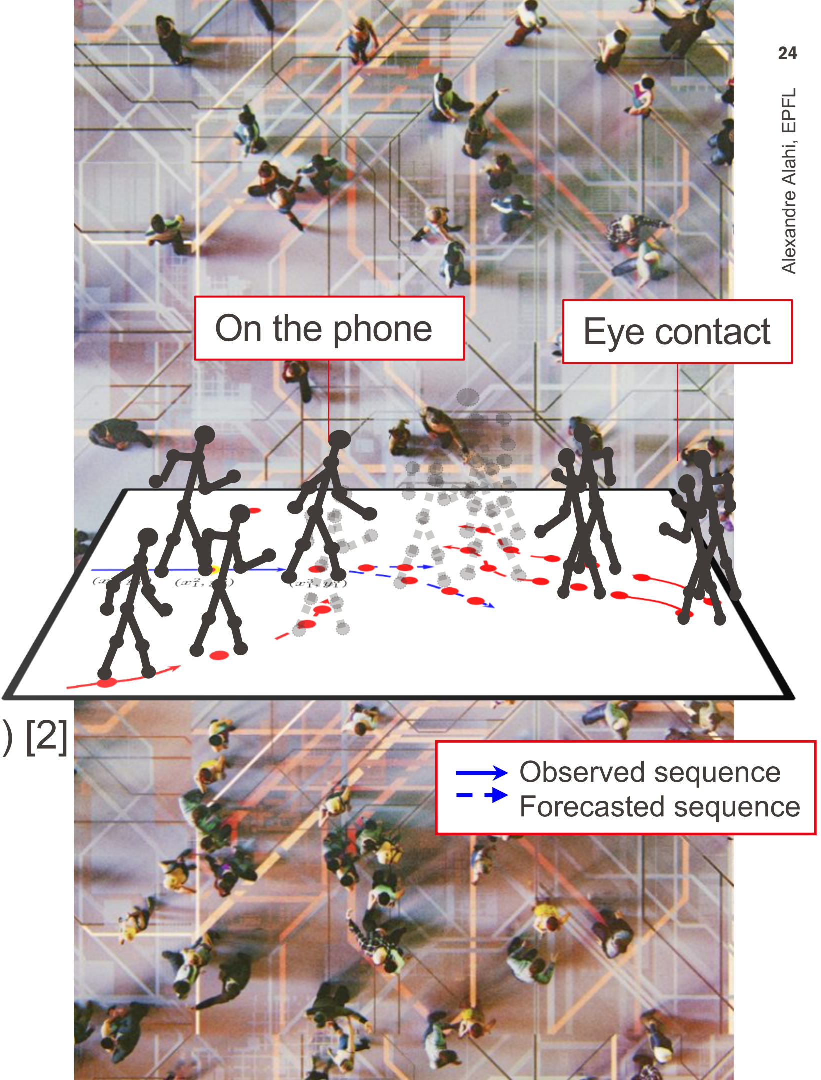
- **Input**: several sequences of states
- **Output**: forecast the future states, *e.g.*, next 5 seconds



Social Forecasting
(w/ pedestrians)

- **Input**: several sequences of states
- **Output**: forecast the future states, *e.g.*, next 5 seconds
- **State**:
  - (x^t,y^t) coordinates in time
  - Body pose [1]
  - Attributes (*e.g.*, on the phone, eye contact) [2]

- Challenge 1: agent-agent interactions
- Challenge 2: disentangle physics from social

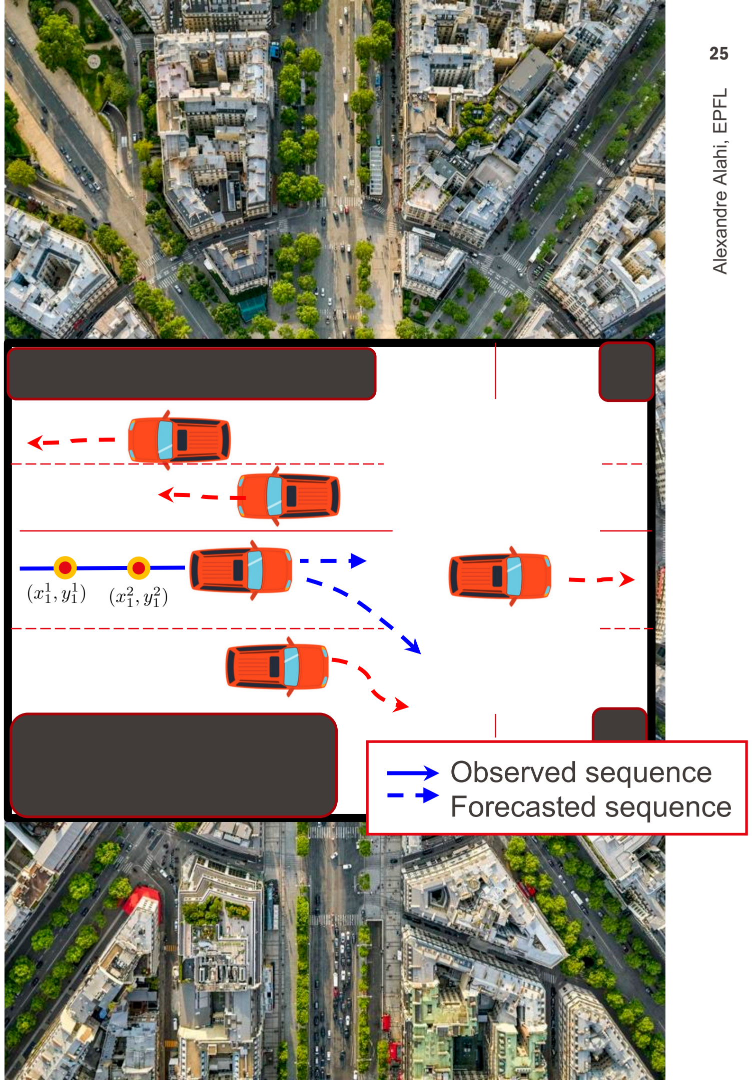


[1] PifPaf, CVPR’19
[2] 32 attributes detector, ITS transactions’21

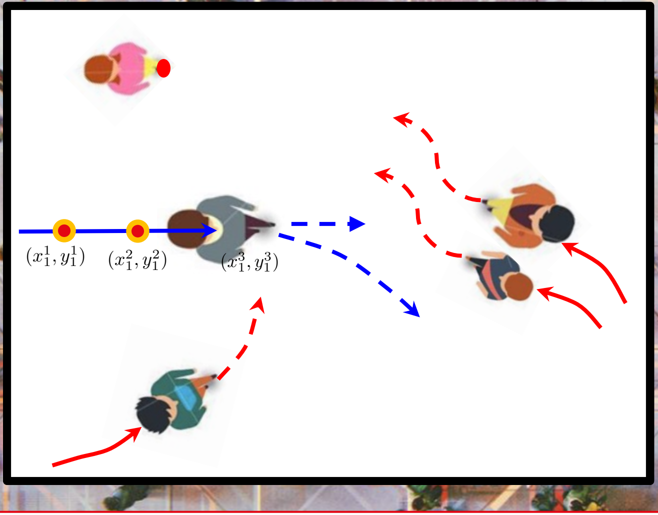
Social Forecasting
(w/ vehicles)

- **Input**: several sequences of states
  + scene infrastructure
- **Output**: forecast the future states, *e.g.*, next 5 seconds

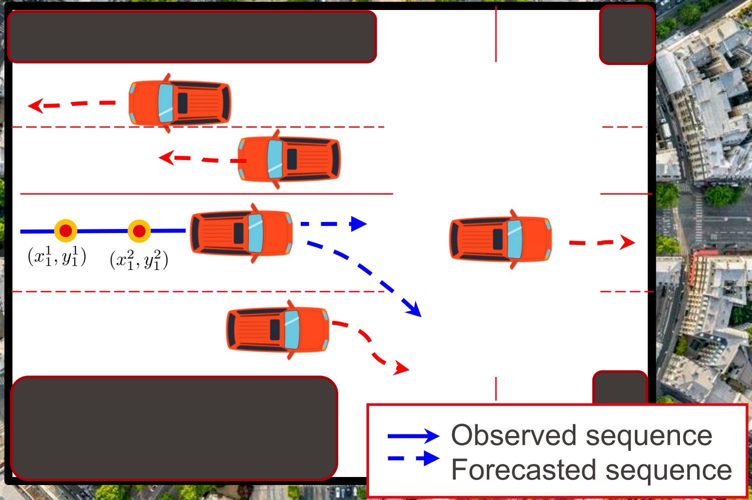
- Challenge 1: agent-agent interactions
- Challenge 2: agent-scene interactions
- Challenge 3: additional external constraints



Social Forecasting
(w/ pedestrians)

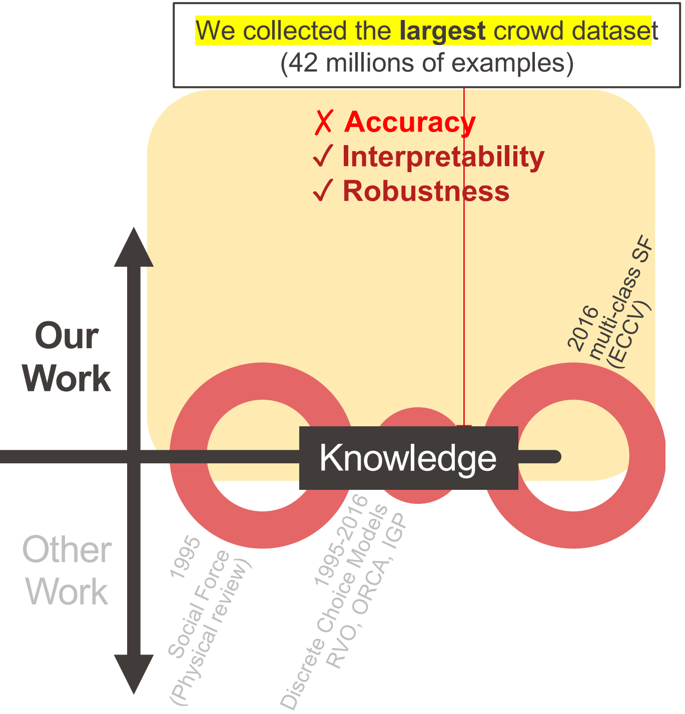


Social Forecasting
(w/ vehicles)



Robustness

Learning paradigms

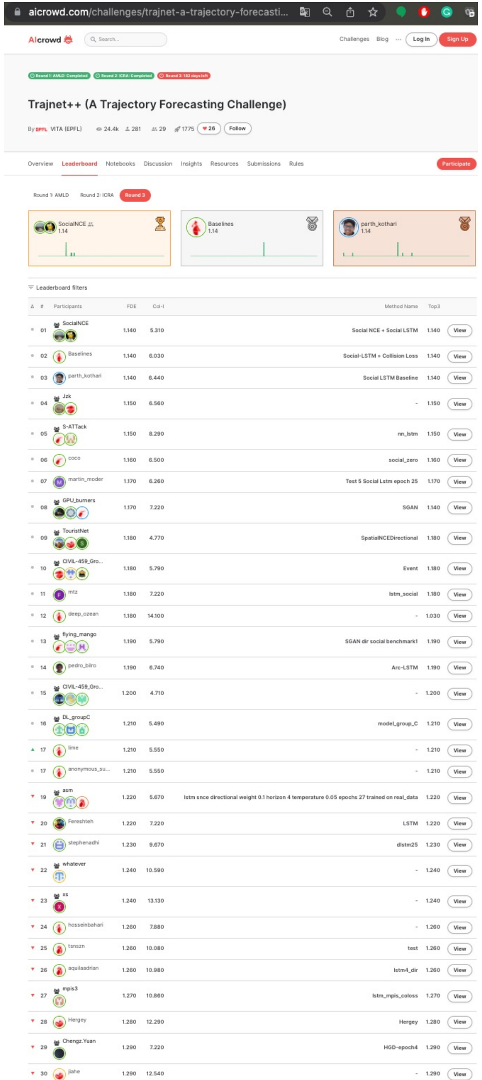


SF = Social Force
IGP = Interacting Gaussian Processes
RVO = Reciprocal Velocity Obstacle
ORCA = Optimal reciprocal collision-avoidance
LSTM = Long Short-Term Memory
GAN = Generative Adversarial Network
TTT = Test-Time Training

Trajnet++

- Open-source library (> 15 models)
  - https://github.com/vita-epfl/trajnetplusplusdata
- Data+evaluation protocols
- Challenge on Aicrowd
  - https://www.aicrowd.com/challenges/trajnet-a-trajectory-forecasting-challenge

GitHub


aicrowd.com/challenges/trajnet-a-trajectory-forecasti...

Trajnet++ (A Trajectory Forecasting Challenge)

By EPFL VITA (EPFL) 24.4k 281 29 1775 ♥ 26 Follow

SocialNCE 1.14

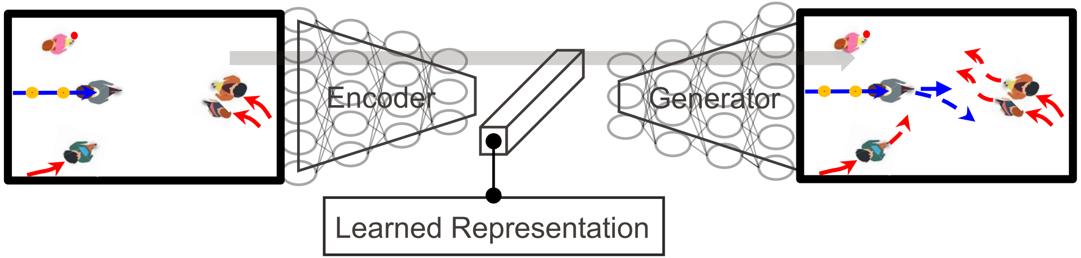
Baselines 1.14

parth_kothari 1.14

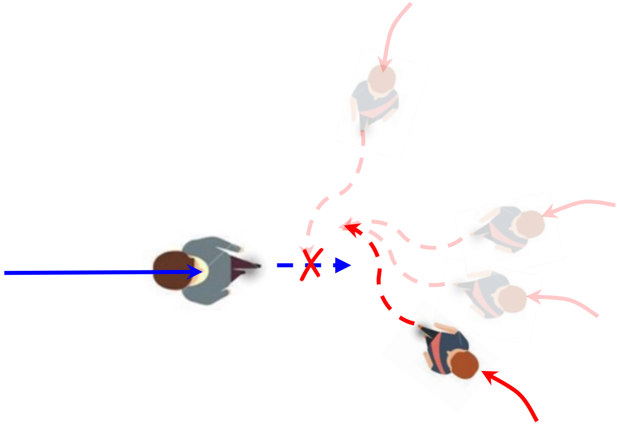
Leaderboard filters

Δ	#	Participants	FDE	Col-I	Method Name	Top3	
●	01	SocialNCE	1.140	5.310	Social NCE + Social LSTM	1.140	View
●	02	Baselines	1.140	6.030	Social-LSTM + Collision Loss	1.140	View
●	03	parth_kothari	1.140	6.440	Social LSTM Baseline	1.140	View
●	04	Jzk	1.150	6.560	-	1.150	View
●	05	S-ATTack	1.150	8.290	nn_lstm	1.150	View
●	06	coco	1.160	6.500	social_zero	1.160	View
●	07	martin_moder	1.170	6.260	Test 5 Social Lstm epoch 25	1.170	View
●	08	GPU_burners	1.170	7.220	SGAN	1.140	View
●	09	TouristNet	1.180	4.770	SpatialNCEDirectional	1.180	View
●	10	CIVIL-459_Gro...	1.180	5.790	Event	1.180	View
●	11	mtz	1.180	7.220	lstm_social	1.180	View
●	12	deep_ozean	1.180	14.100	-	1.030	View
●	13	flying_mango	1.190	5.790	SGAN dir social benchmark1	1.190	View
●	14	pedro_bilro	1.190	6.740	Arc-LSTM	1.190	View
●	15	CIVIL-459_Gro...	1.200	4.710	-	1.200	View
●	16	DL_groupC	1.210	5.490	model_group_C	1.210	View
▲	17	lime	1.210	5.550	-	1.210	View
●	17	anonymous_su...	1.210	5.550	-	1.210	View
▼	19	asm	1.220	5.670	lstm snce directional weight 0.1 horizon 4 temperature 0.05 epochs 27 trained on real_data	1.220	View
▼	20	Fereshteh	1.220	7.220	LSTM	1.220	View
▼	21	stephenadhi	1.230	9.670	dlstm25	1.230	View
▼	22	whatever	1.240	10.590	-	1.240	View
▼	23	xs	1.240	13.130	-	1.240	View
▼	24	hosseinbahari	1.260	7.880	-	1.260	View
▼	25	tsnszn	1.260	10.080	test	1.260	View
▼	26	aquilaadrian	1.260	10.980	lstm4_dir	1.260	View
▼	27	mpis3	1.270	10.860	lstm_mpis_coloss	1.270	View
▼	28	Hergey	1.280	12.290	Hergey	1.280	View
▼	29	Chengz.Yuan	1.290	7.220	HGD-epoch4	1.290	View
▼	30	jiahe	1.290	12.540	-	1.290	View

Current paradigm



✗ **Not Robust**



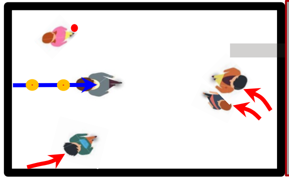
→ Observed sequence

- -→ Forecasted sequence by [1]

✗ **Collision**

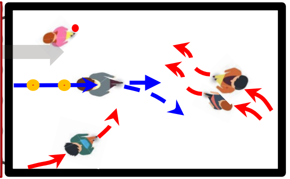
[1] Ynet, ICCV'21, Top ranked model in Trajnet++ public challenge

Current paradigm



Current paradigm:

Use **data to learn** the task of **navigation from scratch**



✗ **Not** Robust

Our proposed paradigm:

Use **data** to **only learn concepts we struggle to explain** (*e.g.*, our social norms)

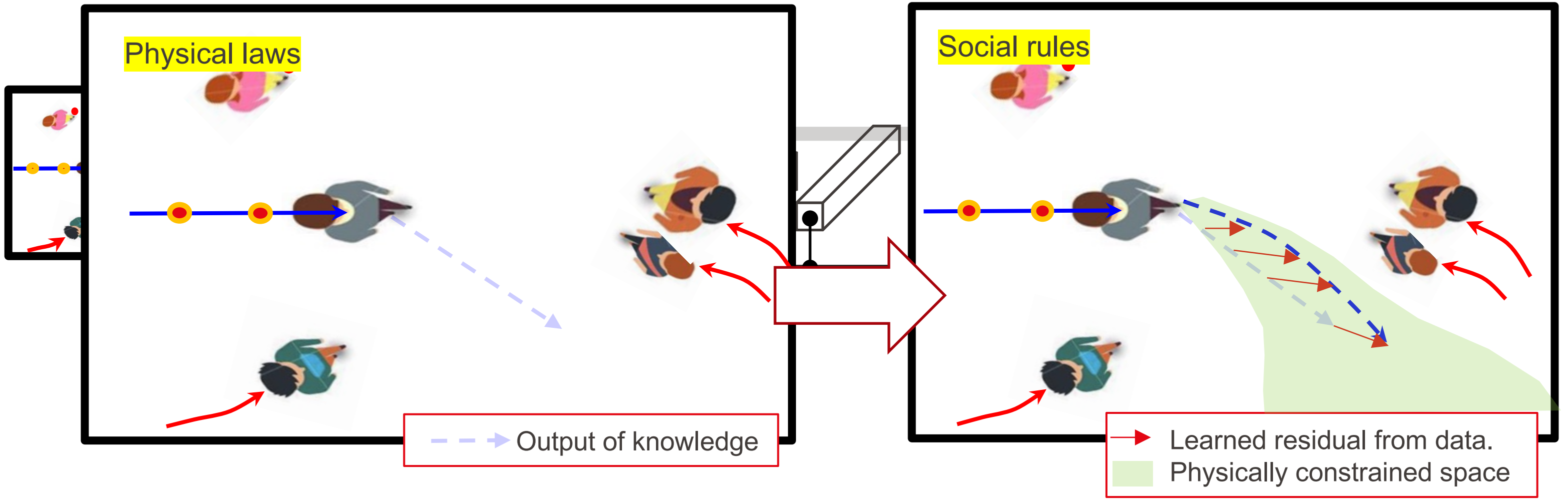
Because

1. Imbalanced/missing data

Solution

- Knowledge-Data

Proposed Knowledge-Data paradigm



Because

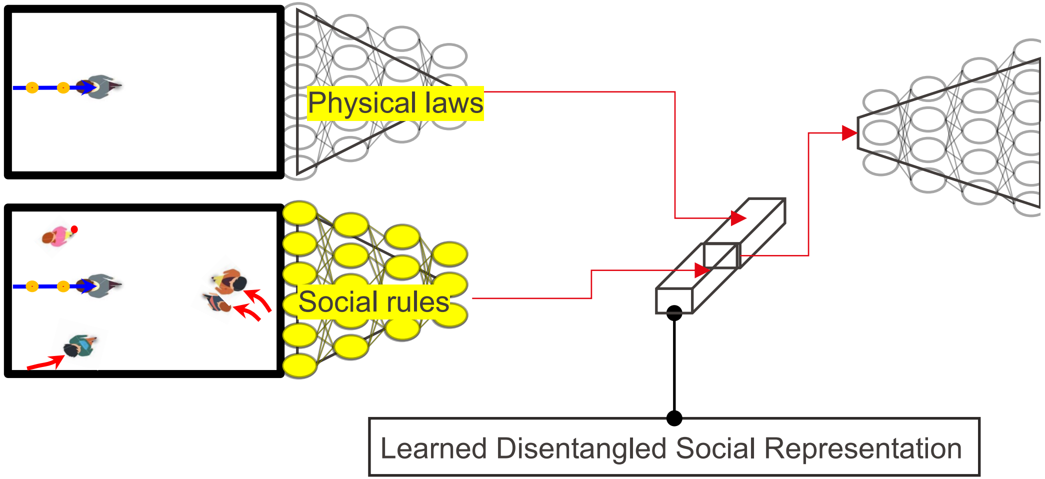
1. Imbalanced/missing data

Solution

- Knowledge-Data
  - Knowledge as input

[1] Injecting knowledge in data-driven vehicle trajectory predictors, **TRC'21**

Proposed Knowledge-Data paradigm



Outcome

✓ Generalizable (low-shot transfer)

Because

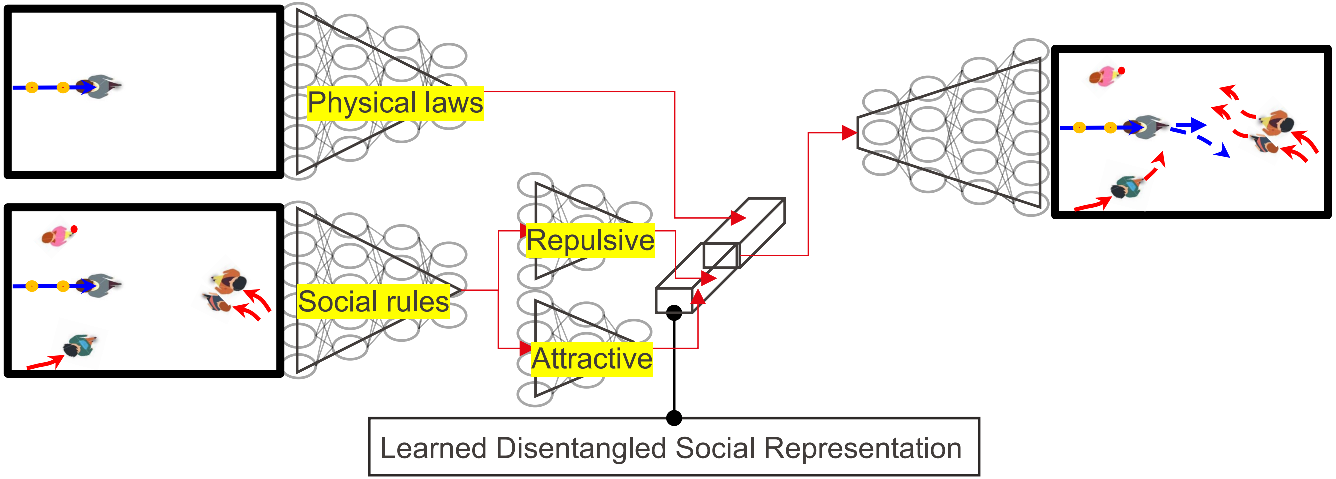
1. Imbalanced/missing data

Solution

- Knowledge-Data
  - Knowledge as input
  - Knowledge within

[1] Towards Robust and Adaptive Motion Forecasting: A Causal Representation Perspective, **CVPR’22**

Proposed Knowledge-Data paradigm



Because

1. Imbalanced/missing data

Solution

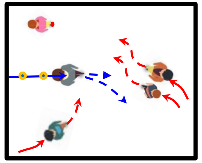
- Knowledge-Data
  - Knowledge as input
  - Knowledge within
  - Knowledge as supervision

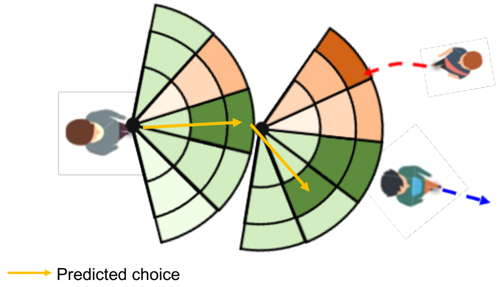
[1] Interpretable Social Anchors, **CVPR'21**

Knoweldge-data driven mathematical framework

Discrete Choice Models

$$\underbrace{U}_{\text{Utility}} = \underbrace{V}_{\text{Systematic}} + \underbrace{\varepsilon}_{\text{Random}}$$



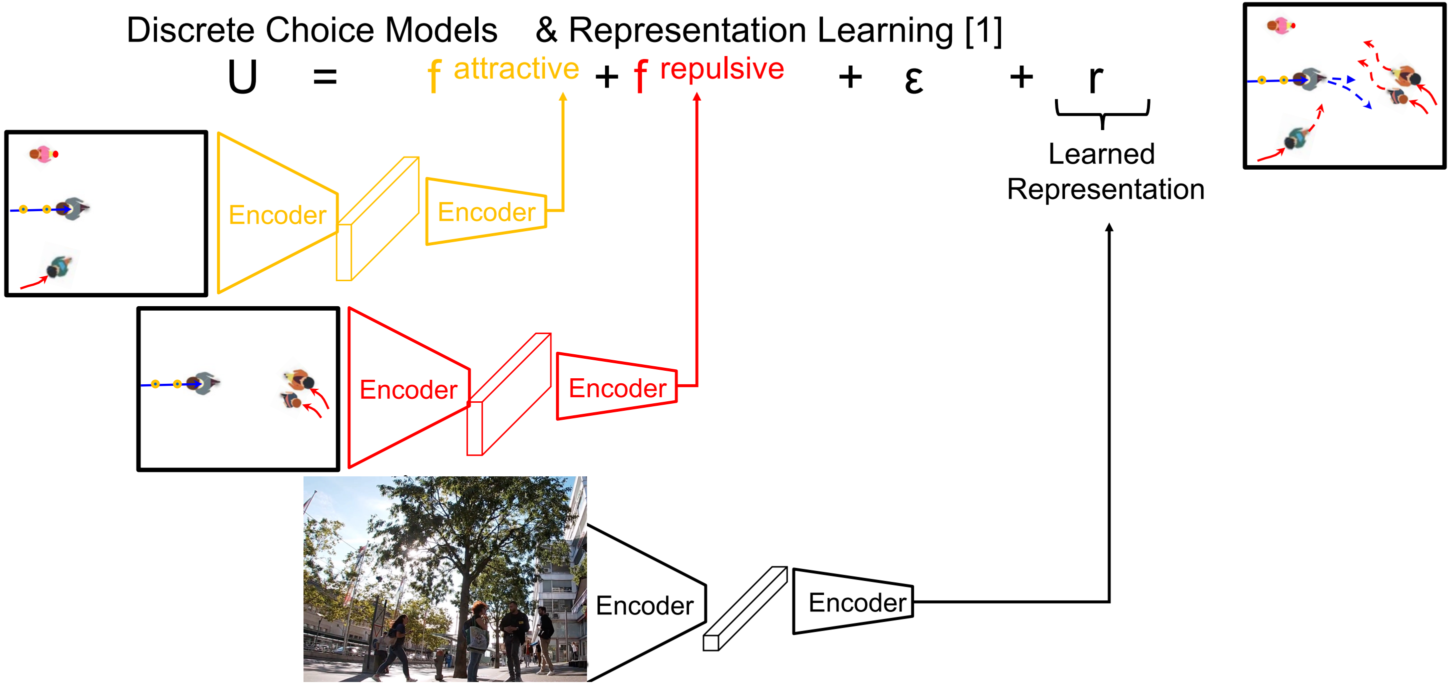


[1] Enhancing Discrete Choice Models with Representation Learning, **TRB’20**

Knoweldge-data driven mathematical framework

Discrete Choice Models & Representation Learning [1]

$$U = f^{\text{attractive}} + f^{\text{repulsive}} + \varepsilon + \underbrace{r}_{\text{Learned Representation}}$$

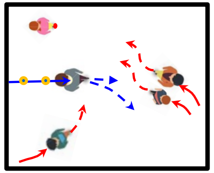


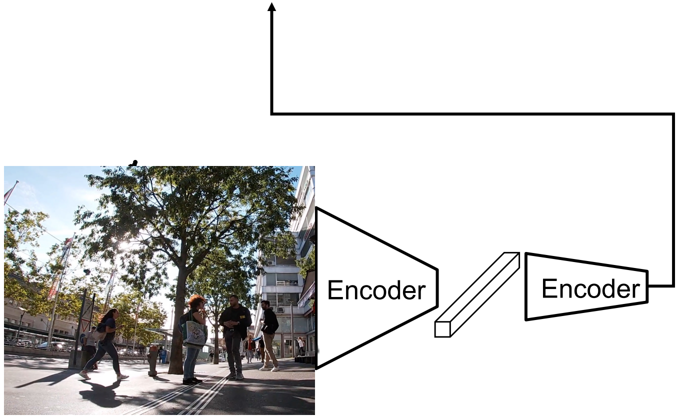
[1] Enhancing Discrete Choice Models with Representation Learning, **TRB’20**

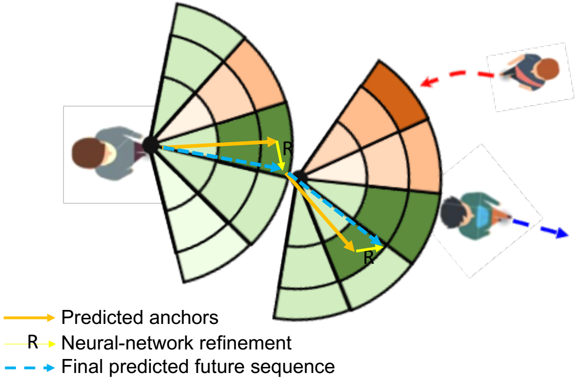
Knoweldge-data driven mathematical framework

Discrete Choice Models & Representation Learning

$$U{+}R = f^{\text{attractive}} + f^{\text{repulsive}} + \varepsilon + \underbrace{r}_{\text{Learned Representation}}$$

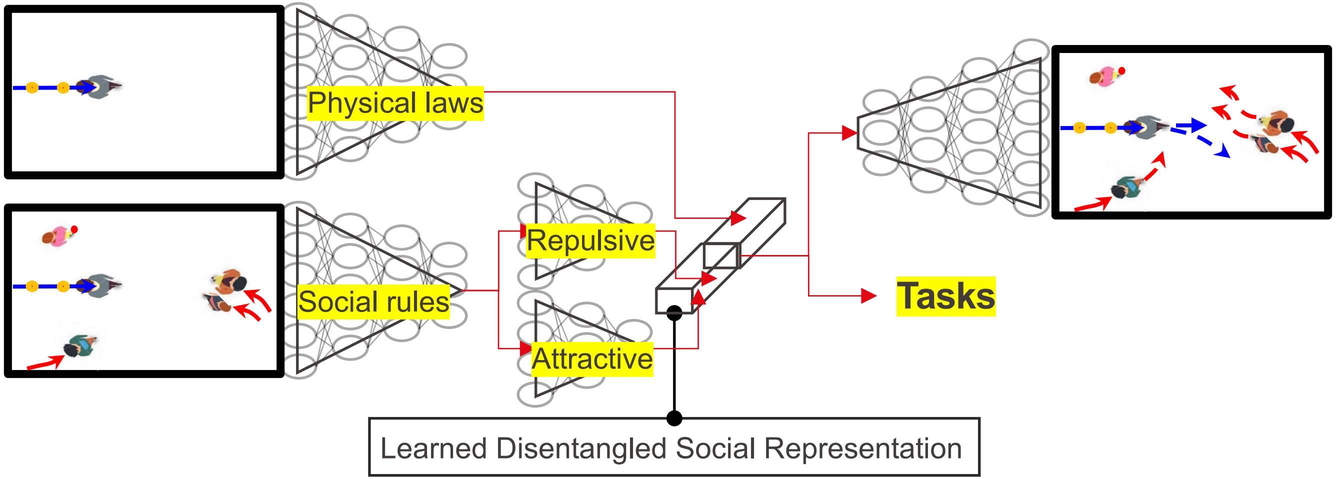






[1] Interpretable Social Anchors, **CVPR’21**

Proposed Knowledge-Data paradigm



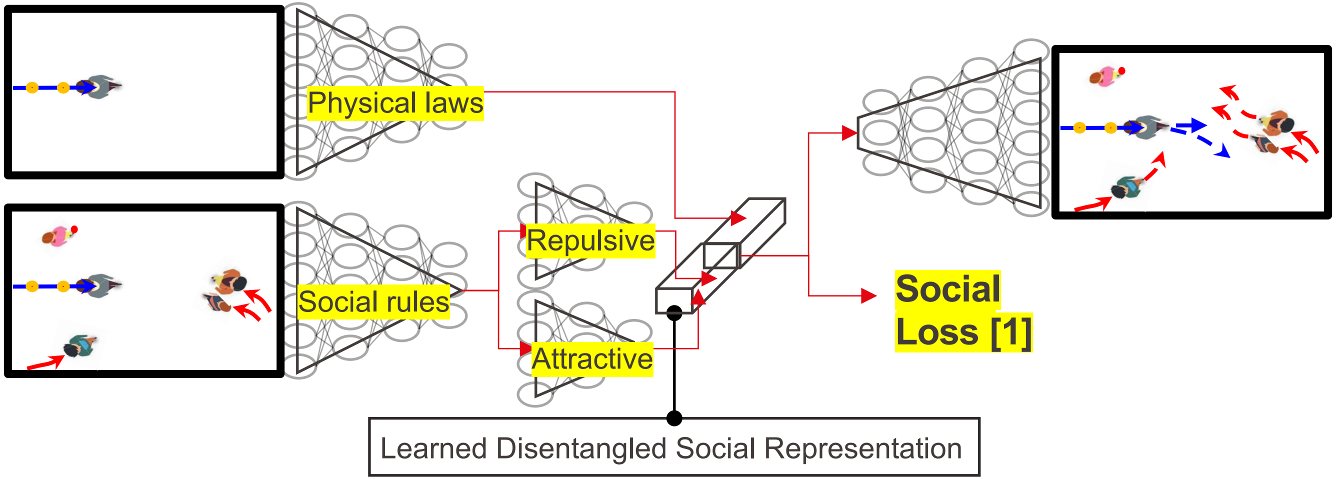
Because

1. Imbalanced/missing data

Solution

- Knowledge-Data
  - Knowledge as input
  - Knowledge within
  - Knowledge as supervision

Proposed Knowledge-Data paradigm



Because

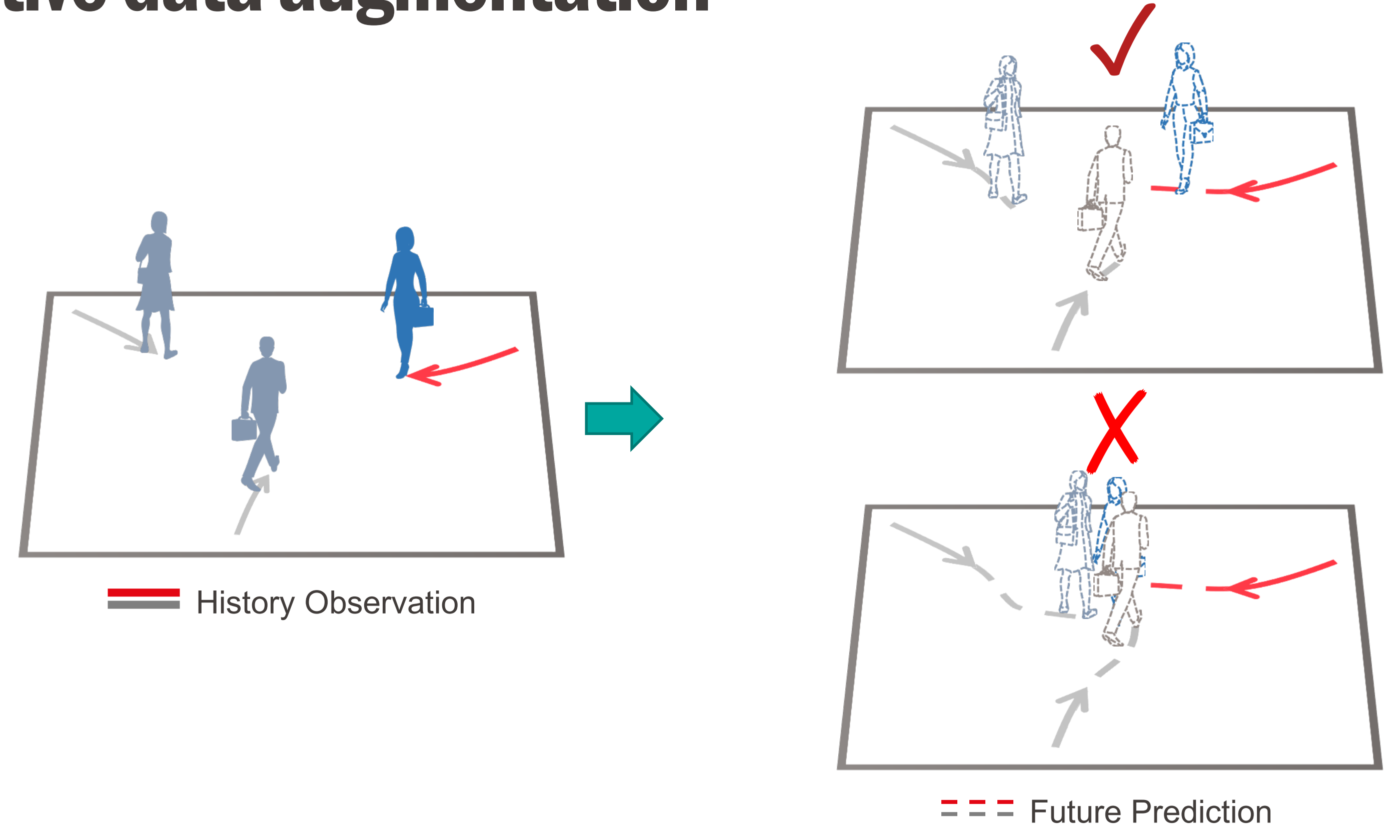
1. Imbalanced/missing data
2. Positive examples only

Solution

- Knowledge-Data
- w/ Opposite principle

[1] Social NCE, **ICCV'21**

Negative data augmentation

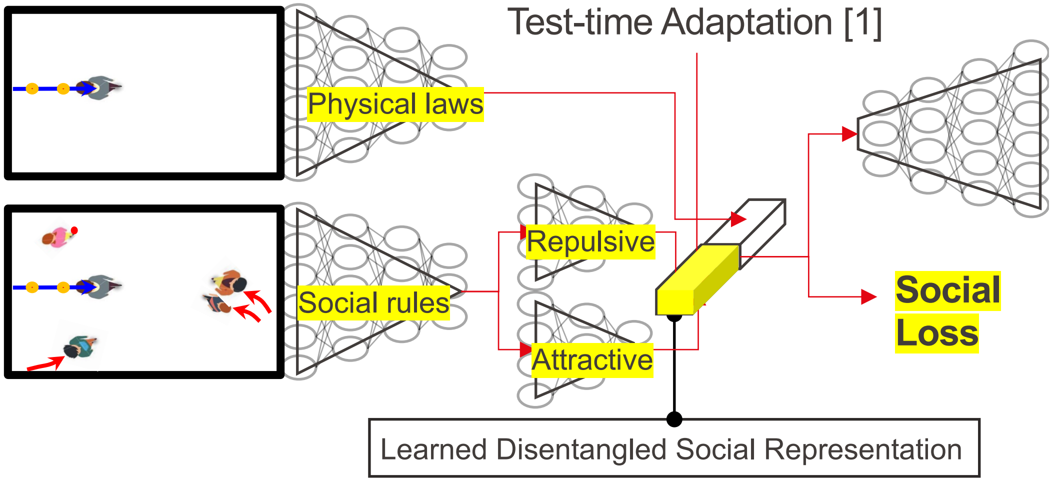


History Observation

Future Prediction

[1] Social NCE, **ICCV'21**

Proposed Knowledge-Data paradigm



Outcome

✓ Robust

✓ Generalizable

✓ Interpretable

Because

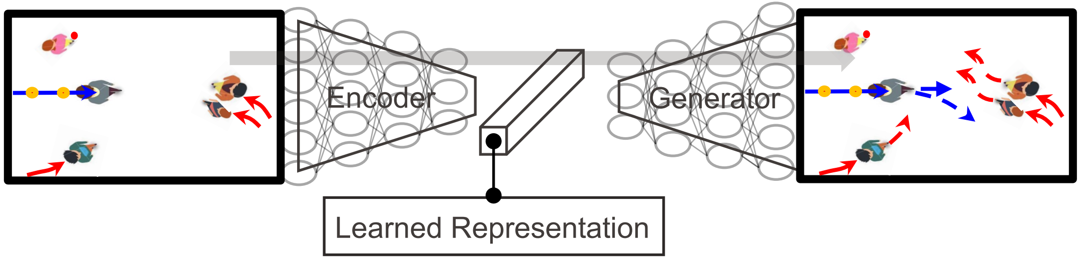
1. Imbalanced/missing data
2. Positive examples only
3. Distributional shifts

Solution

- Knowledge-Data
- w/ Opposite principle
- w/ Low-rank principle

[1] Test Time Training++, **NeurIPS'21**

New evaluation protocol



Encoder

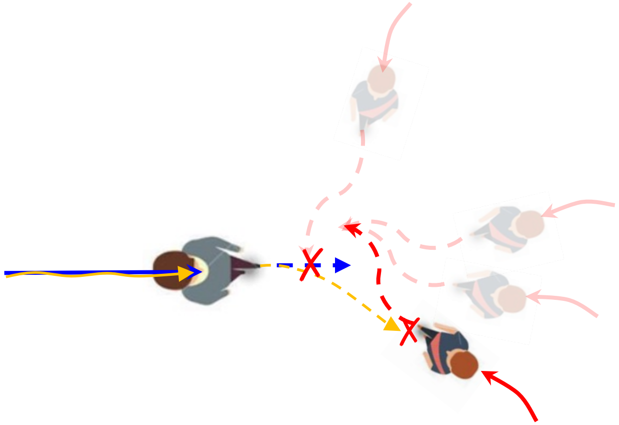
Generator

Learned Representation

Outcome

✓ **New evaluation** based on realistic adverserial examples [1]

✓ **Robust training**



→ Observed sequence

- - → Forecasted sequence by [2]

→ Perturbed observation by < 7 cm

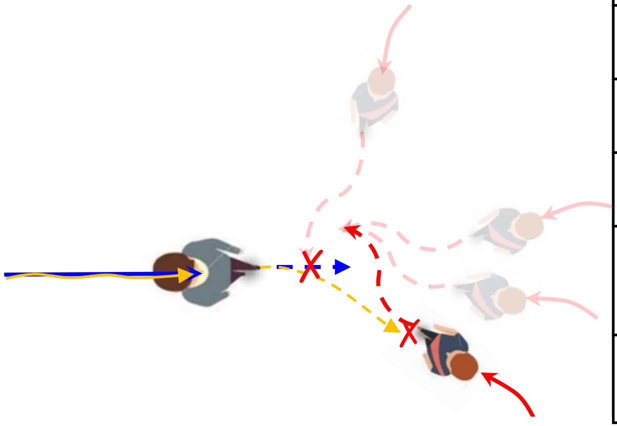
- - → Forecasted sequence leading to collision

✗ **Collision**

[1] S-attack library, https://s-attack.github.io/

[2] Ynet, ICCV'21, Top ranked model in Trajnet++ public challenge

Quantitative results



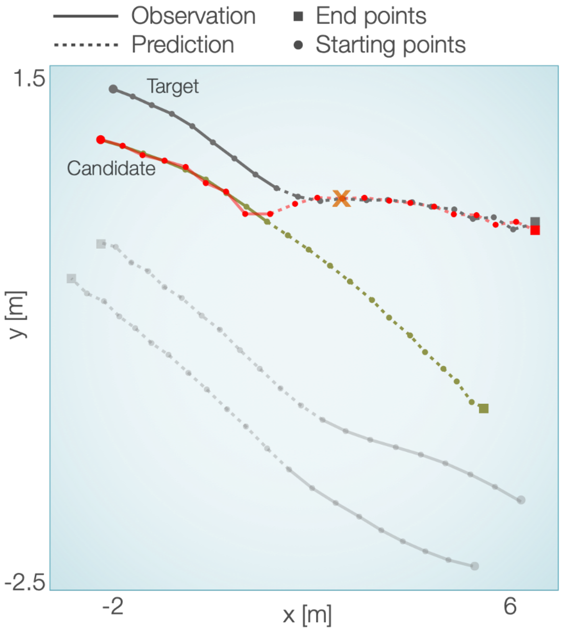
Baseline	Original collision rate
S-LSTM (CVPR'16)	7.8%
S-Att (ICRA'18)	9.4%
S-GAN (CVPR'18)	13.9%
D-Pool (ITS'2021)	7.3%
S-STGCNN (CVPR'20)	16.3%
PECNet (ECCV'20)	15.0%

=> 6.5% w/ aug

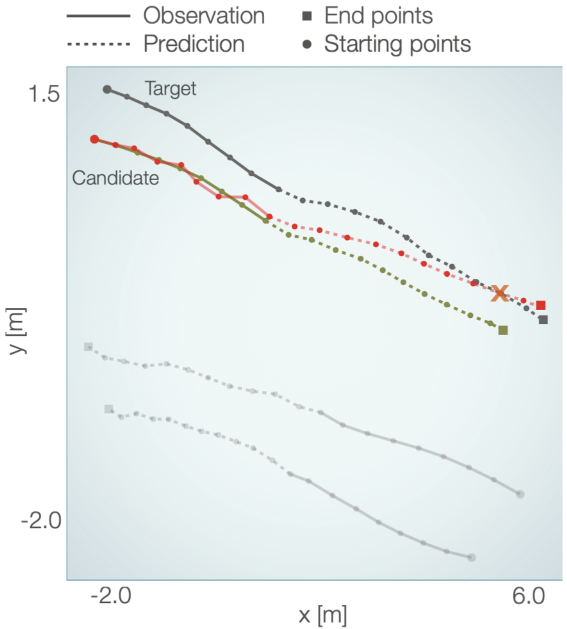
[1] S-attack library, https://s-attack.github.io/

Qualitative results

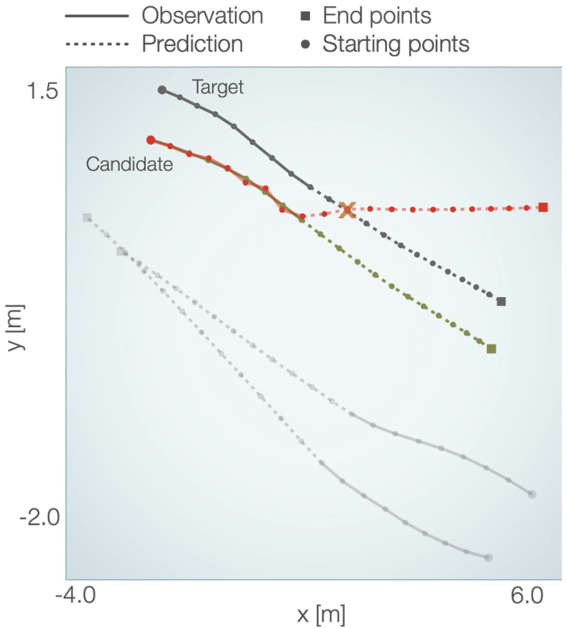
Alexandre Alahi, EPFL



D-Pool[ITS'2021]



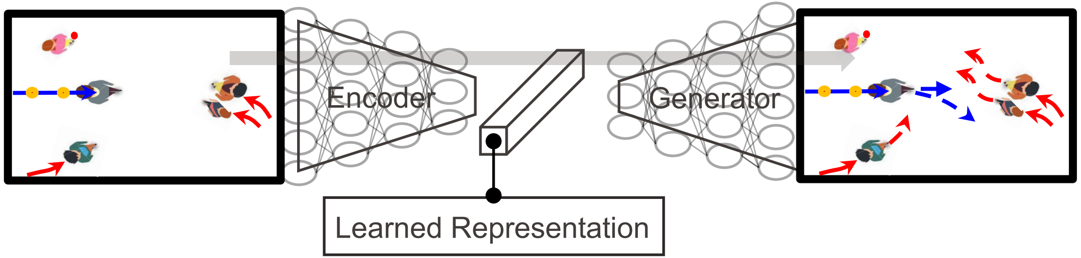
S-STGCNN[CVPR'20]



S-GAN[CVPR'18]

[1] S-attack library, https://s-attack.github.io/

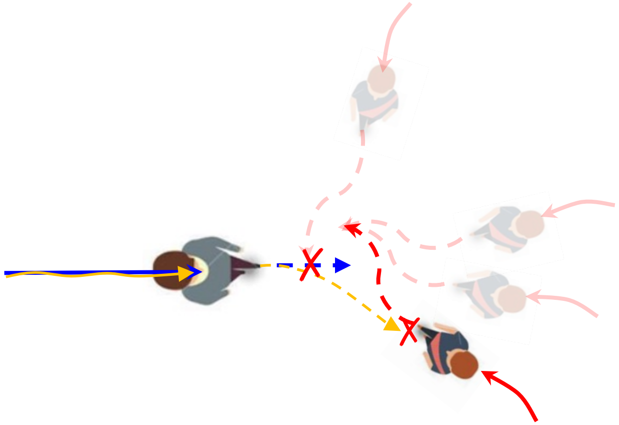
New evaluation protocol



Outcome

✓ **New evaluation** based on realistic adverserial examples [1]

✓ **Robust training**

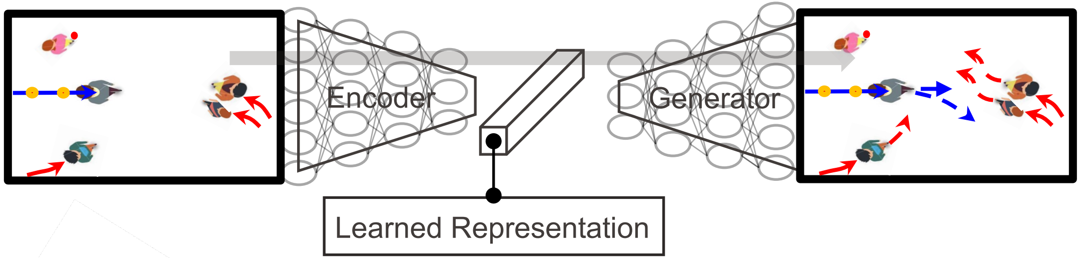


→ Observed sequence
- - → Forecasted sequence by [2]
→ Perturbed observation by < 7 cm
- - → Forecasted sequence leading to collision
✗ **Collision**

[1] S-attack library, https://s-attack.github.io/

[2] Ynet, ICCV'21, Top ranked model in Trajnet++ public challenge

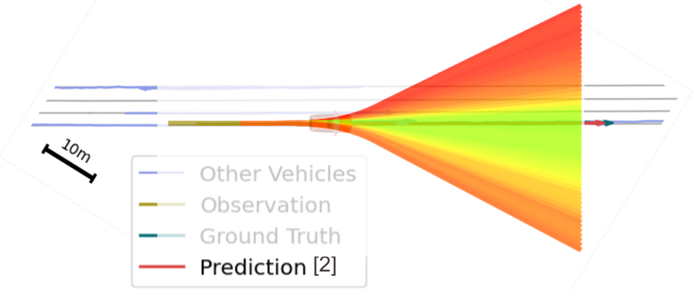
New evaluation protocol

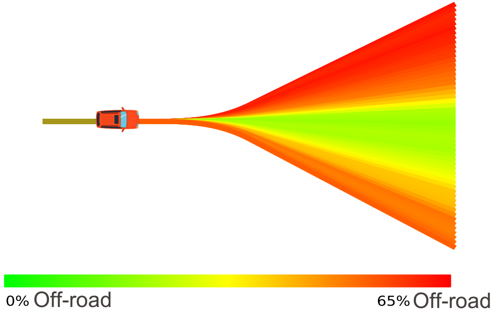


Outcome

✓ **New evaluation** based on realistic adverserial examples [1]

✓ **Robust training**



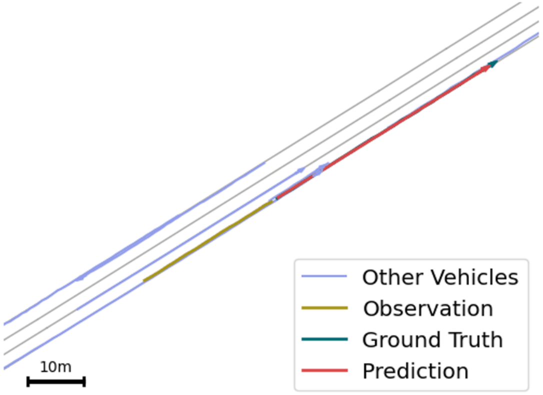


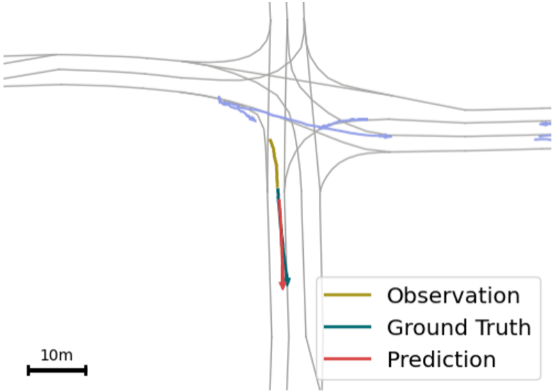
[1] Vehicle trajectory prediction works, but not everywhere, **CVPR'22**

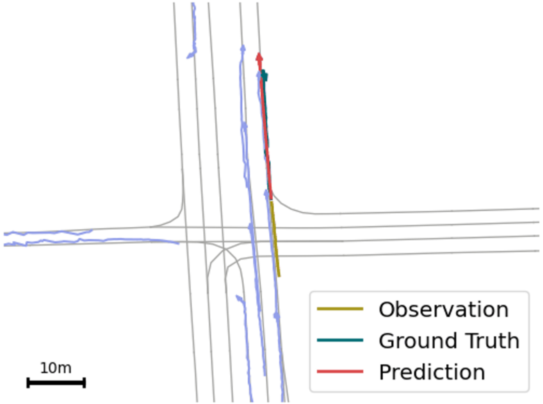
[2] LaneGCN, ECCV'20, Top ranked model in Argoverse public challenge

Scene generation

✓ Atomic scene generation functions







Simple turn

Double turn

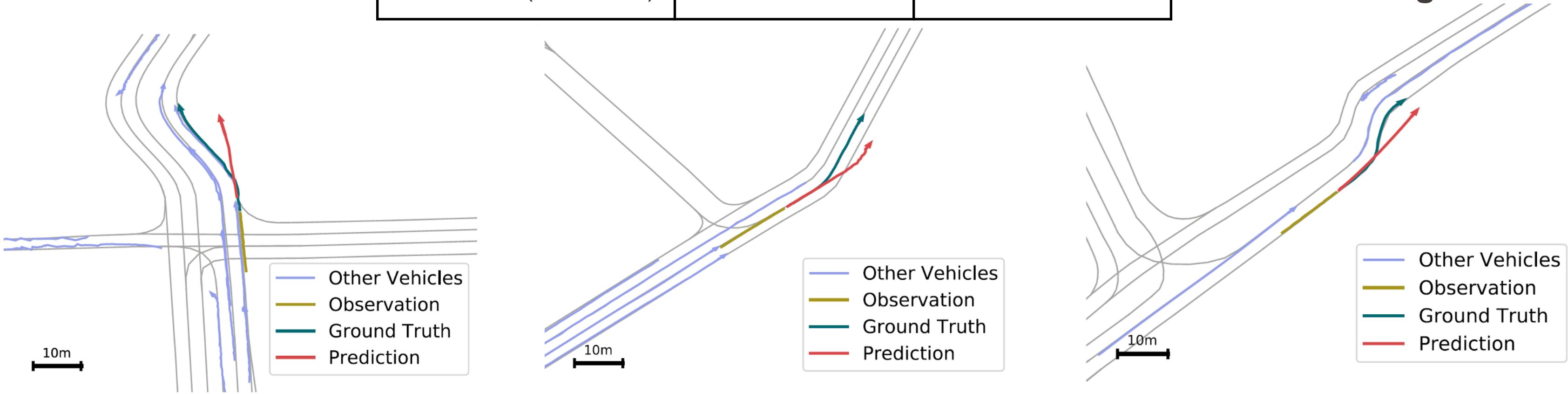
Ripple road

[1] Vehicle trajectory prediction works, but not everywhere, **CVPR'22**

Quantitative results

Baseline	Original off-road	Generated (ours) off-road
DATF (ECCV20)	2%	82%
WIMP (arXiv20)	1%	63%
LaneGCN (ECCV'20)	1%	66%

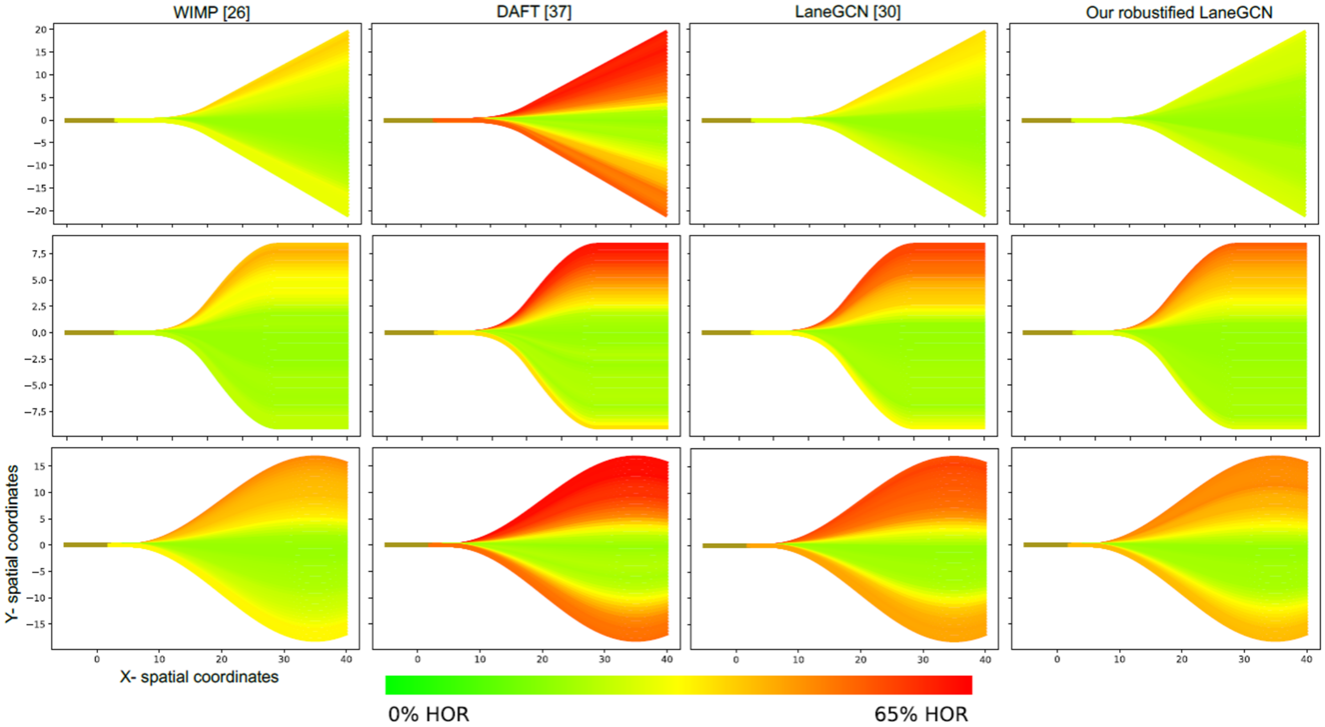
=> 46% w/ aug



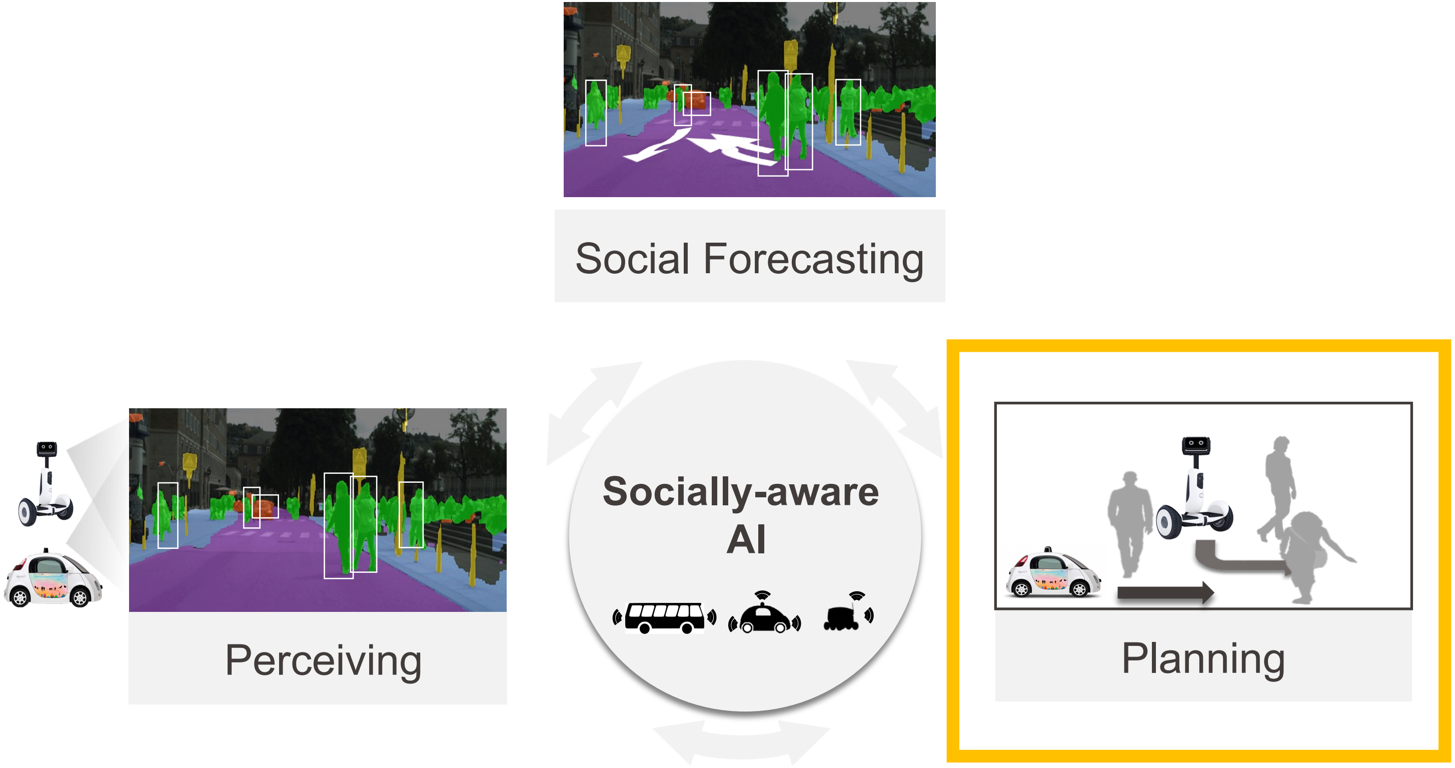
(a) DATF (b) WIMP (c) LaneGCN

[1] Vehicle trajectory prediction works, but not everywhere, **CVPR'22**

Discussions



- [1] Vehicle trajectory prediction works, but not everywhere, **CVPR’22**



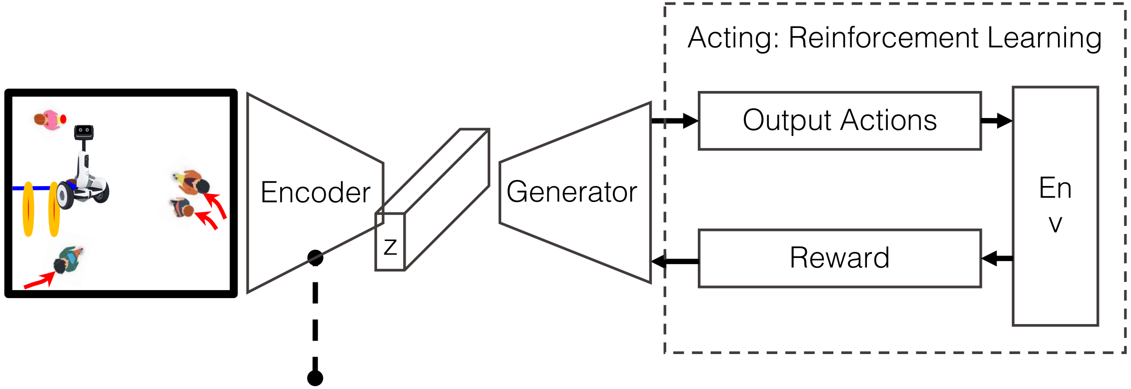
Social Forecasting

Perceiving

Socially-aware AI

Planning

Crowd-Robot Interaction [1]



Human-Robot Interaction [1,2]
&
Our Human-Human Interaction [3]

Previous works

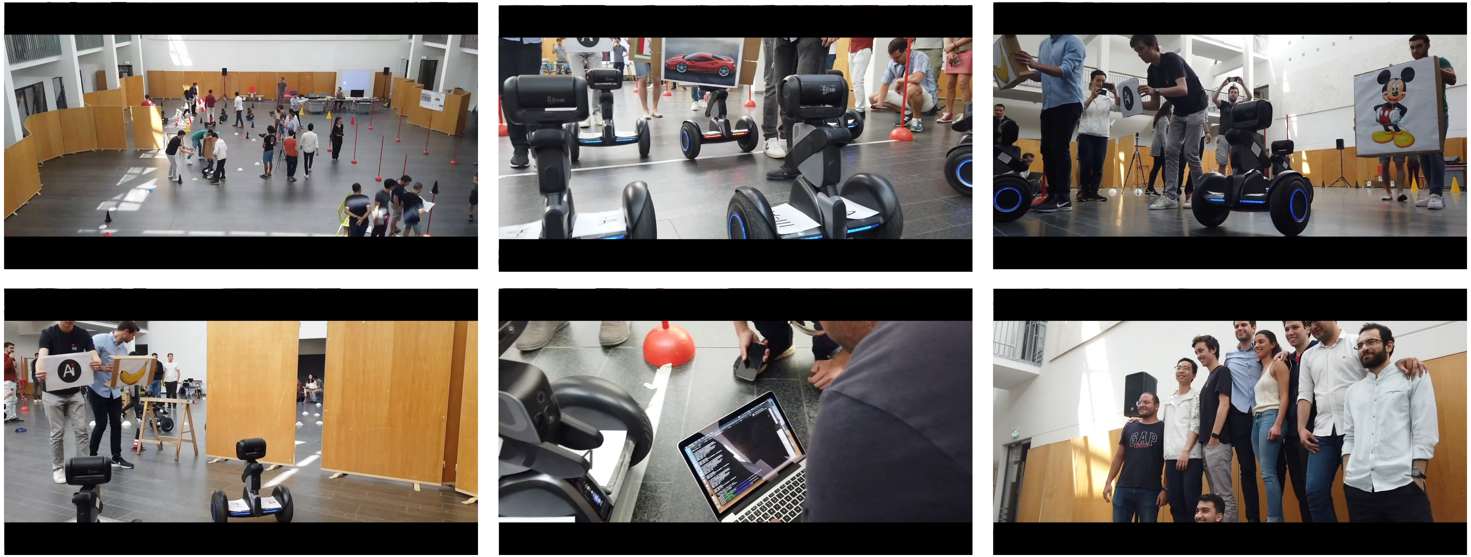
[1] HRI, Chen, C., *et al.,* IROS'17

[2] HRI, Everett, M., *et al.*, IROS'18

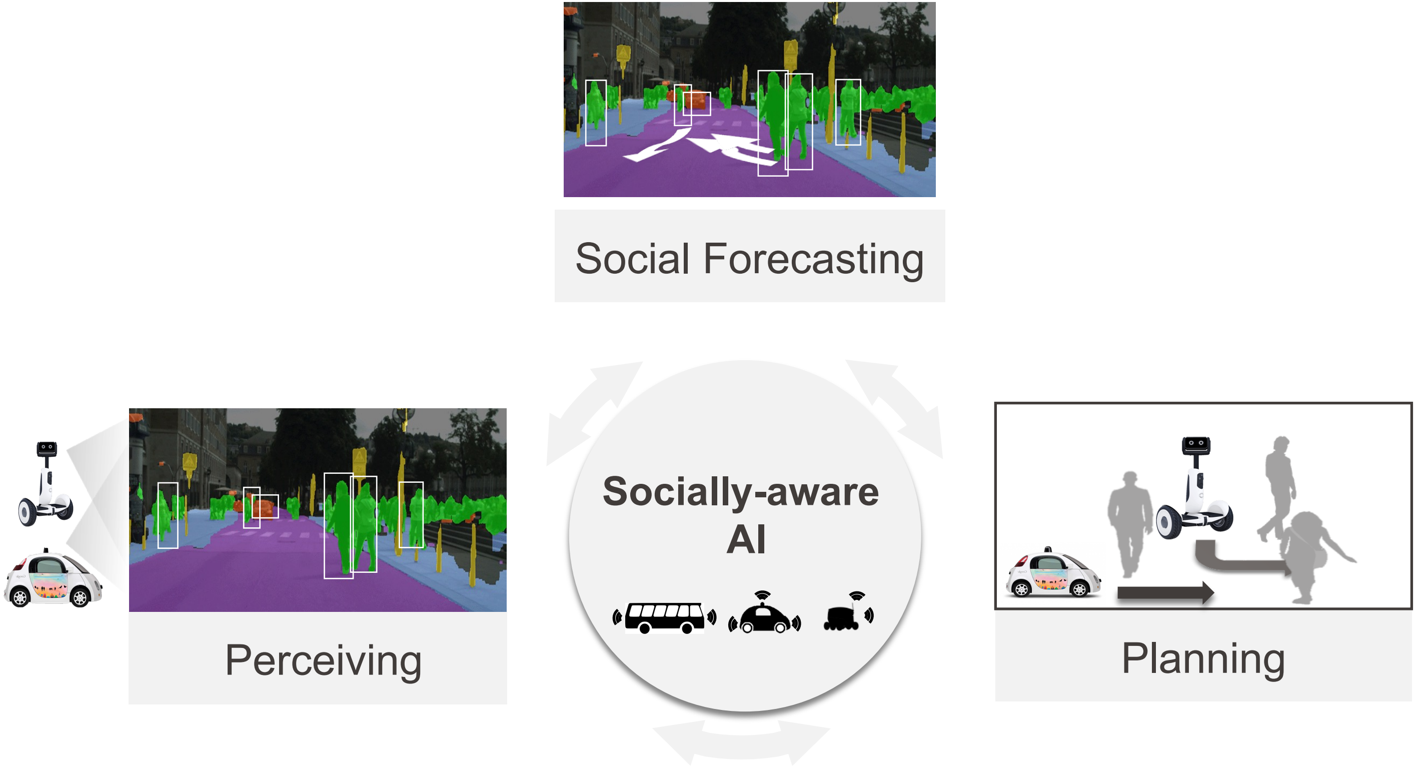
Our work

[3] Crowd-Robot Interaction, ICRA'19

Human-Robot Tandem Race



Ai
I·A
I·A



Social Forecasting

Socially-aware AI

Perceiving

Planning

Thank you!

EPFL
VITA

#Open Science

Alexandre Alahi, EPFL

Perception:
[1] S. Kreiss et al., OpenPifPaf **library** for pose estimation, **CVPR'19, ICCV'21 (licensed)**
[2] L. Bertoni et al., 3D perception **library**, **ICCV'19, ICRA'21**
[3] L. Bertoni et al., Perceiving Social Distancing, **ITS'20**
[4] G. Adaimi et al., Deep Visual Re-identification with Confidence, **TRC'21**
[5] T. Mordan et al., Detecting 32 human attributes, **ITS'21**
Prediction:
[6] Kothari et al., Trajnet++ **library** for spatio-temporal forecasting tasks (>15 implemented models)
[7] Kothari et al., Social Anchor, **ICCV'21**
[8] Liu et al., Social NCE, **ICCV'21**
Planning:
[9] C. Chen et al., Crowd-Robot Interaction, **ICRA'19**
Generative models:
[10] Y. Liu* et al., Collaborative Sampling in GAN, **AAAI'20**
[11] A. Carlier et al., Deep SVG, **NeurIPS'20**
DCM + NN
[12] B. Sifringer et al., L-MNL, **TRB'20**
Test-time training:
[13] Y. Liu* et al., TTT++, **NeurIPS'21**
Tools
[14] Video Ultimate labeling
[15] S-attack library**, CVPR'22**

GitHub

Code on-line: **vita.epfl**.ch/code

VITA