EPFL ViTA

Forecasting Human Mobility with Deep Learning: Challenges and Recipes - Alex Alahi

EPFL

Stanford

- Assistant Professor at EPFL (since Sept'17)
- Director of the VITA lab
- Research Scientist at Stanford University (before Sept'17)
- Co-founded startups & collaborate w/ industry leaders

Thank our sponsors

HONDA
The Power of Dreams

SAMSUNG

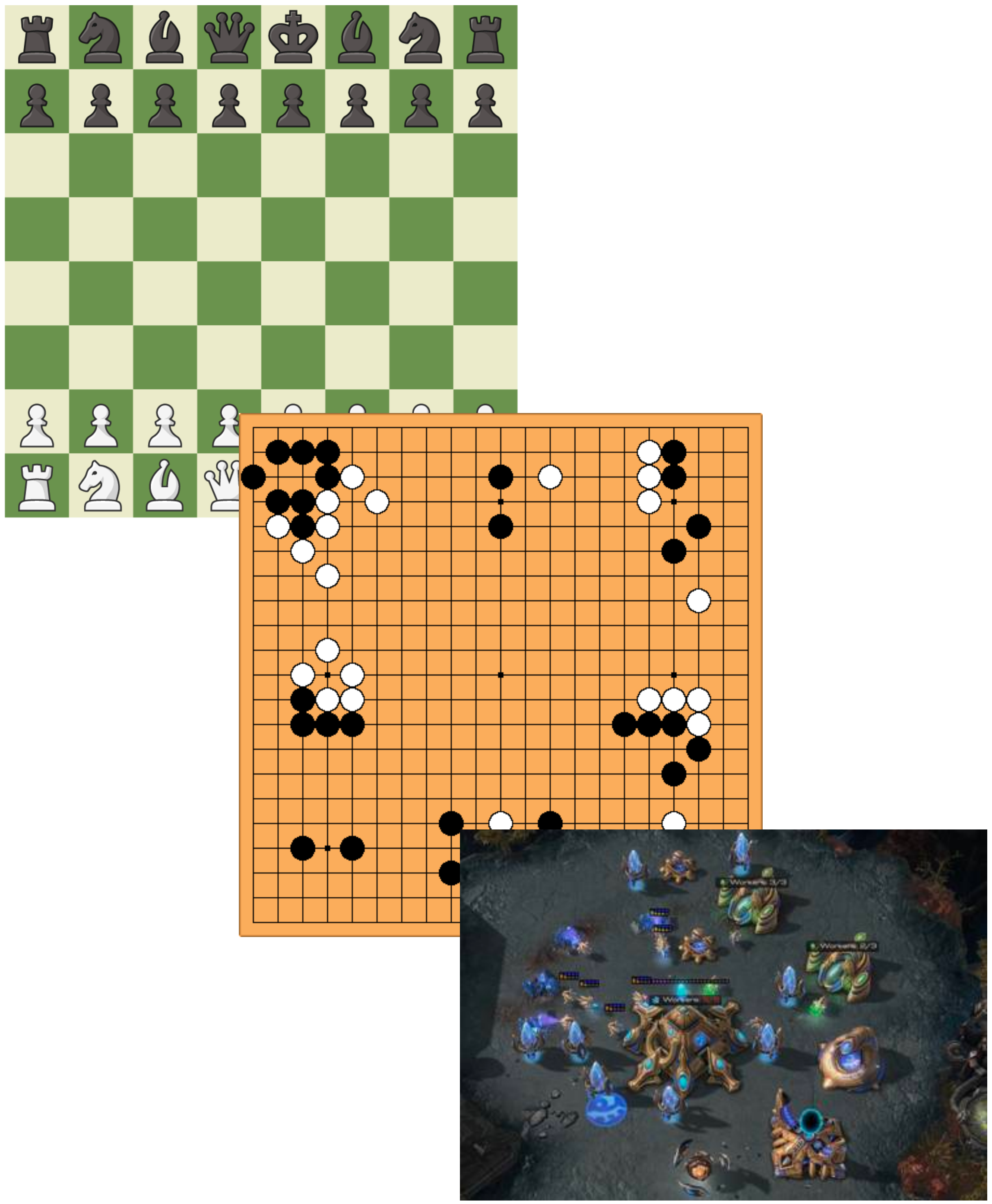
Schindler

RICHEMONT

HITACHI

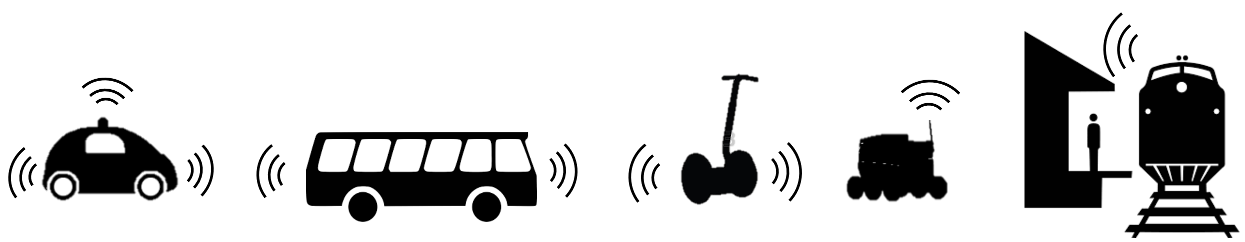
⋮

From *in vitro* To *in vivo*



AI in vivo

Intelligent **Agents**



Save thousands of lives every year

Release driving task / delivery task

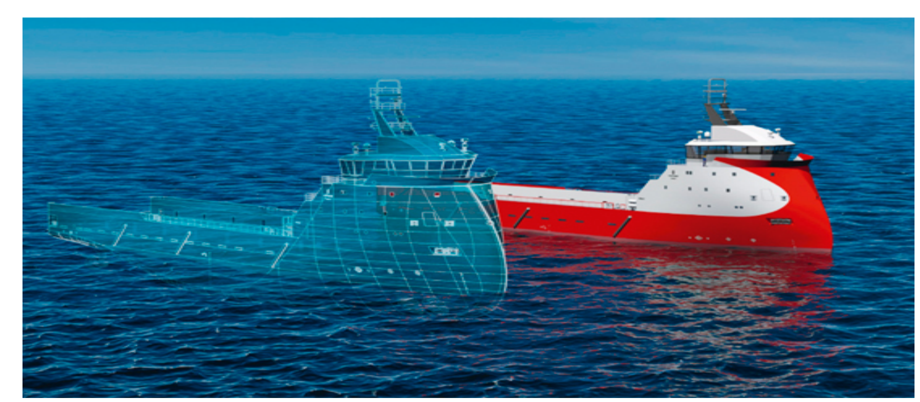
Assist / Guide / Help

Intelligent **Spaces**

Save energy / cost

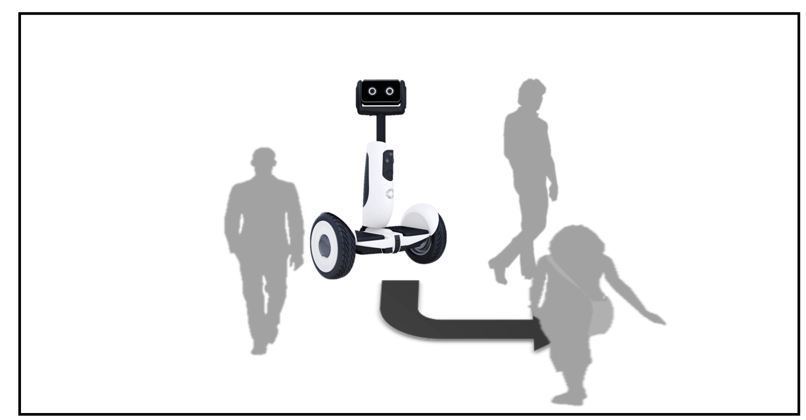
Reduce maintenance cost

Increase productivity / comfort / Safety



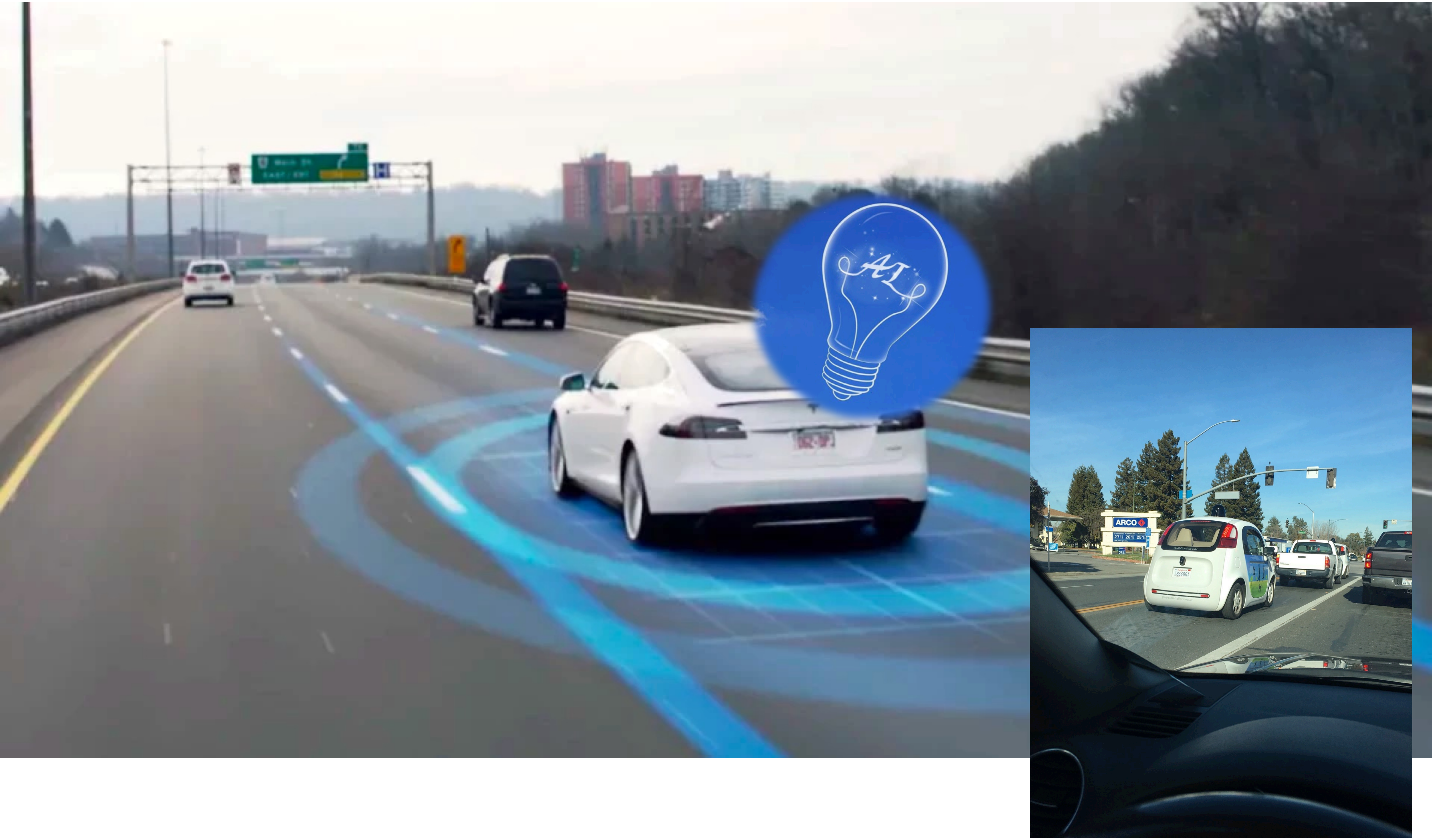
"AI = any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals"

Perceiving



Planning

Exciting existing results

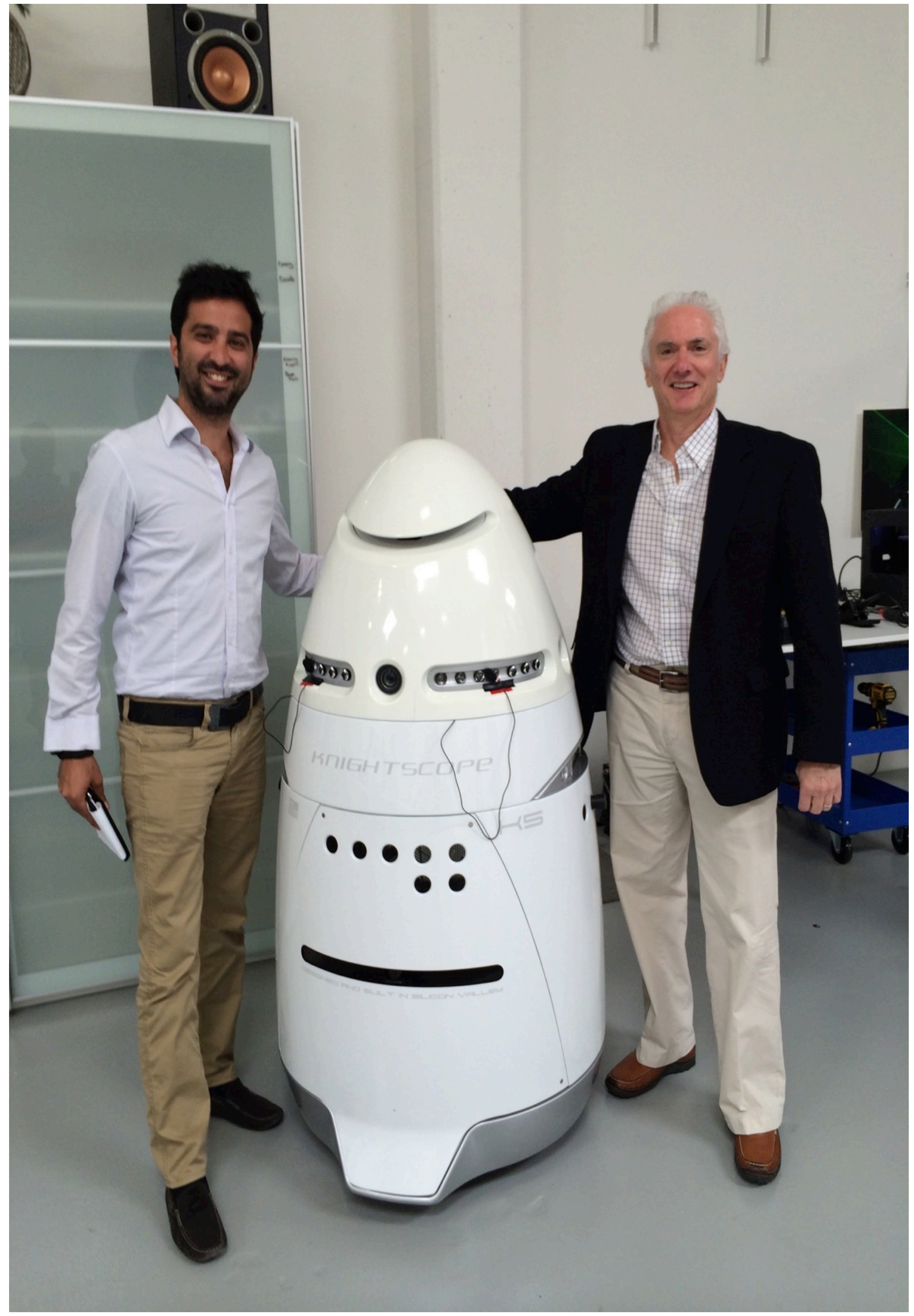


Yet…

Tips on how to drive in Paris…

AI must understand **social etiquettes**

We have …



Security guard robot ends it all by throwing itself into a watery grave

Knightscope K5 security bot shows your job is probably safe from automation. For now.

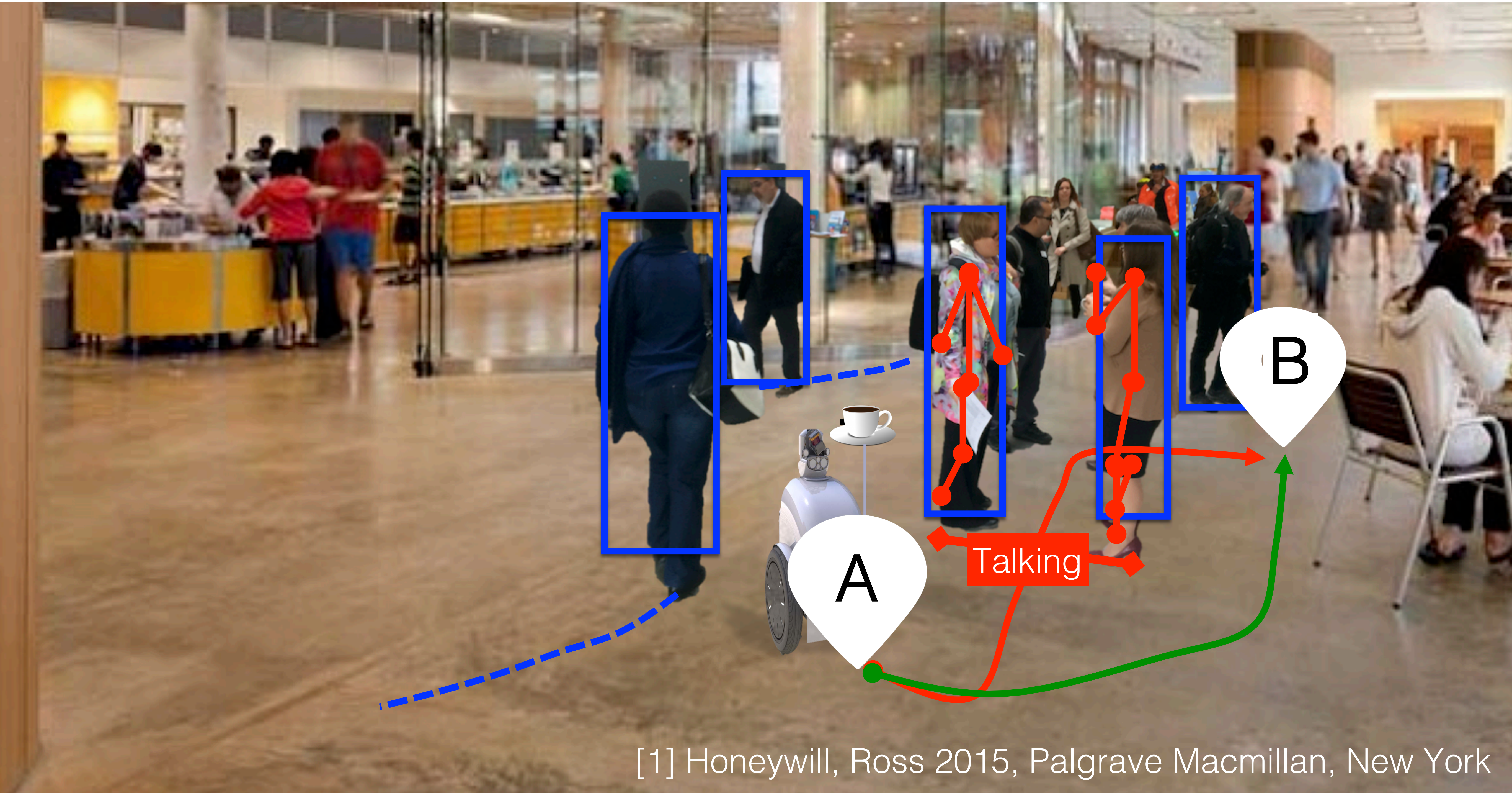
SEBASTIAN ANTHONY - 7/18/2017, 2:58 PM

Enlarge

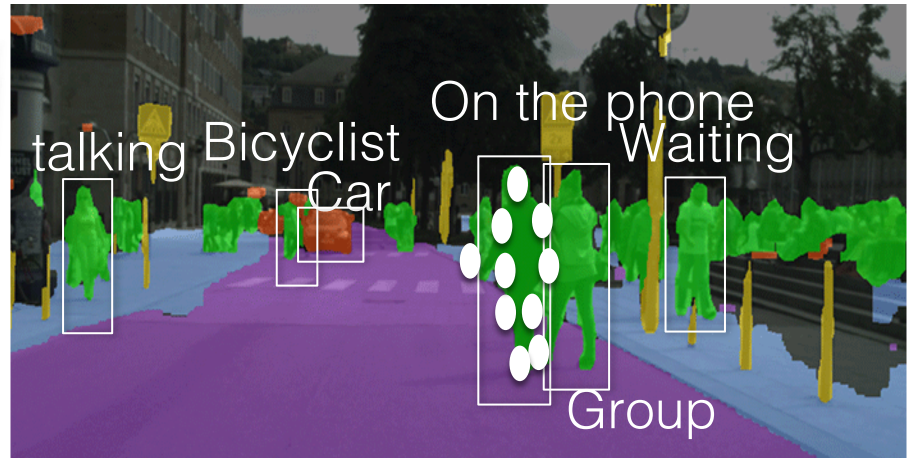
Robots must predict social/ethical interactions to
co-exist & gain society's trust

Socially-aware AI = Perception + Social Intelligence

Social Intelligence = "… the capacity to **effectively** navigate & negotiate complex social **relationships**" [1]

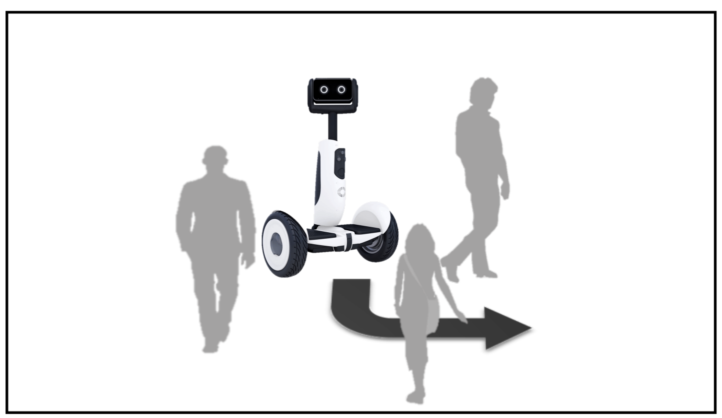


[1] Honeywill, Ross 2015, Palgrave Macmillan, New York

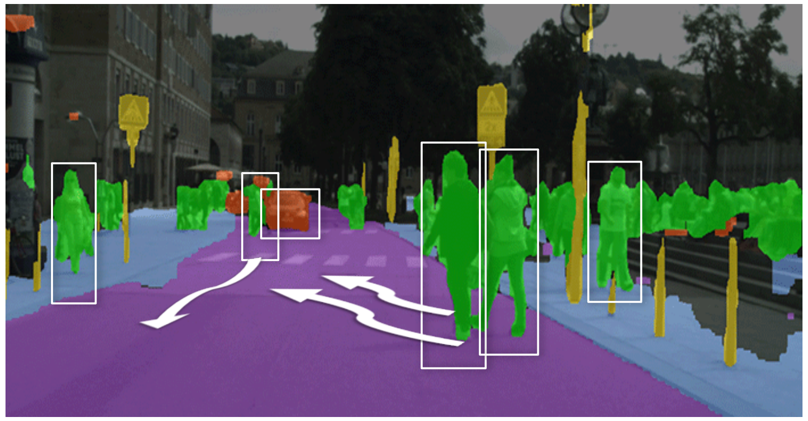


Perceiving

Socially-aware AI

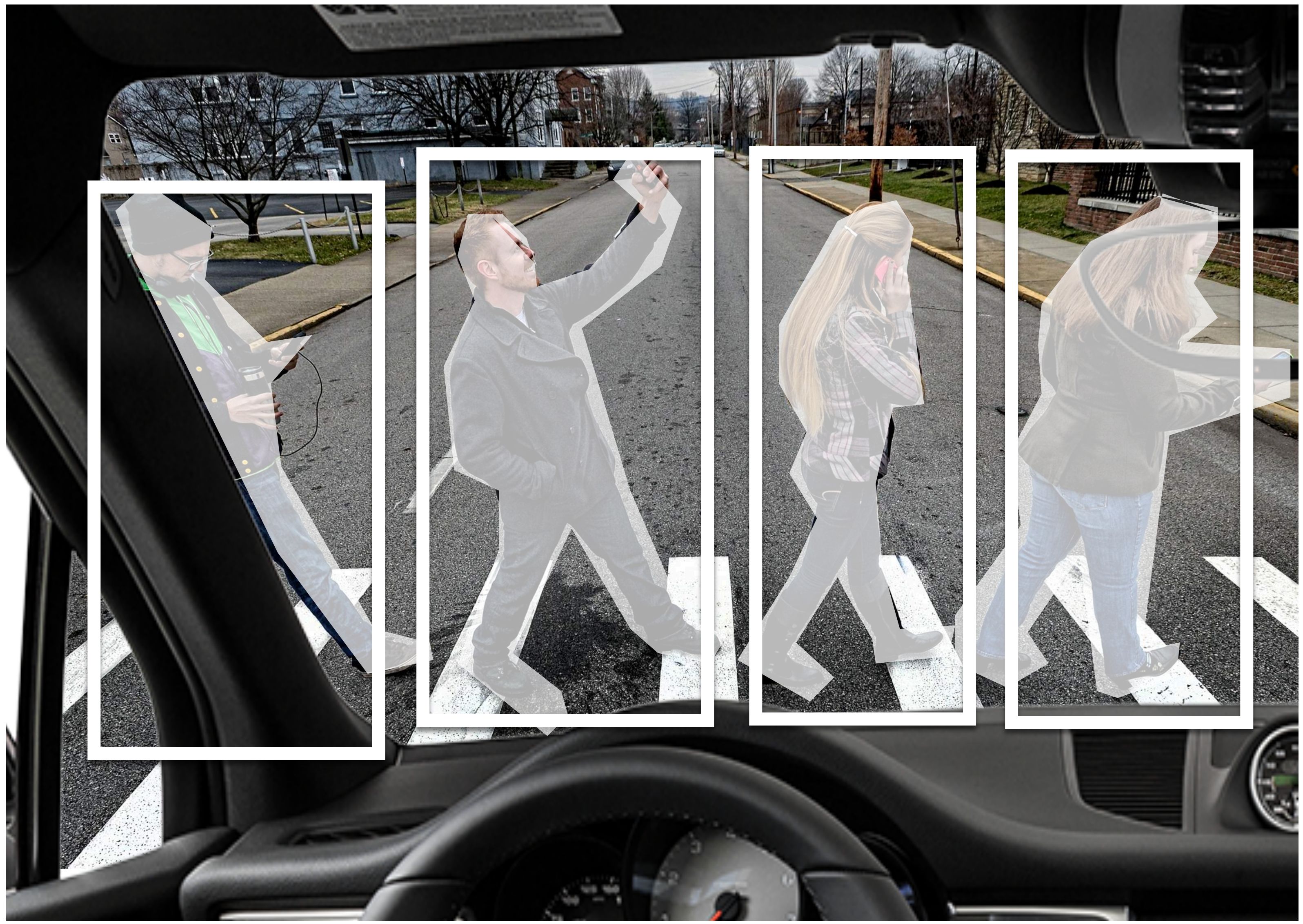


Planning



Predicting

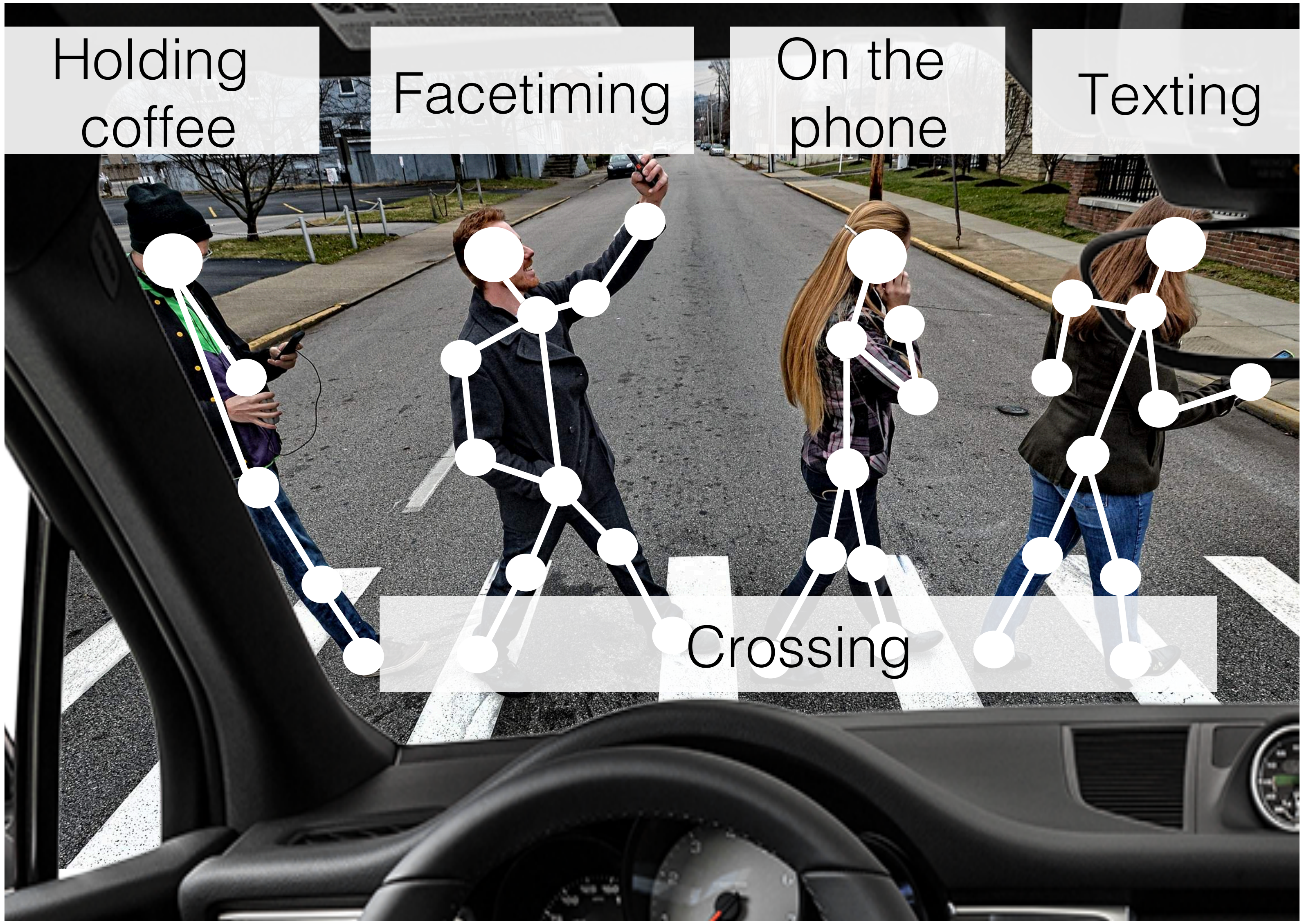
Perceiving



- Detection: Yolo v3 [1], RetinaNet [2]…
- Segmentation: Mask RCNN [3] …

[1] J. Redmon & A. Faradi, Yolov3: An incremental improvement, arxiv '18
[2] T.-Y Lin et al., Focal loss for dense object detection, ICCV'17
[3] K. He et al., Mask R-CNN, ICCV'17

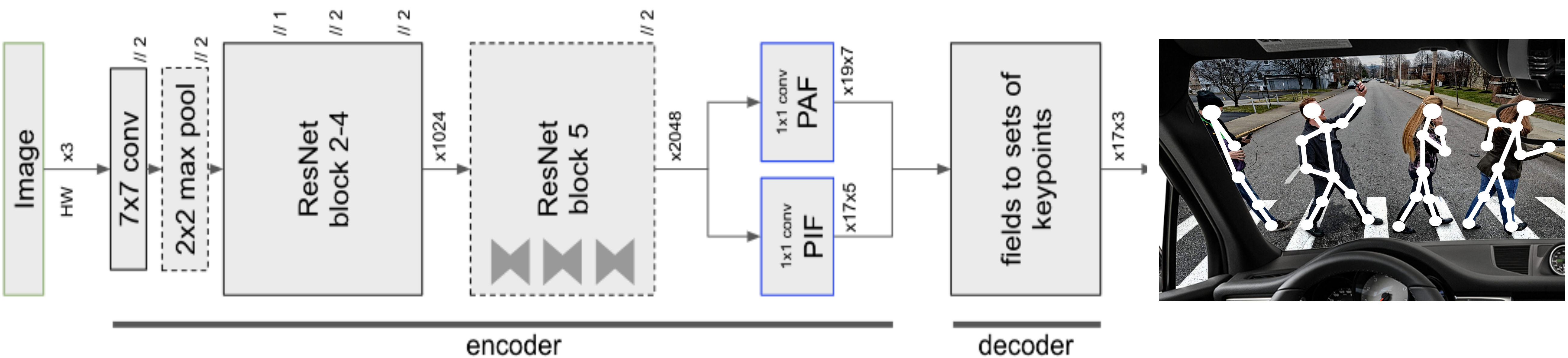
Perceiving
Social cues



How can we learn a representation that jointly solves perception tasks given limited labels?

Unified framework

e.g., Human Pose [1]



[1] S. Kreiss et al., Composite Fields for Human Pose Estimation, **CVPR'19**

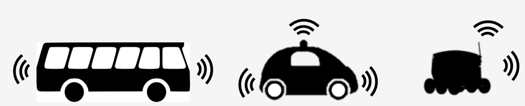
Self-Driving Car

- People occluding people
- Small instance size in wide field of view

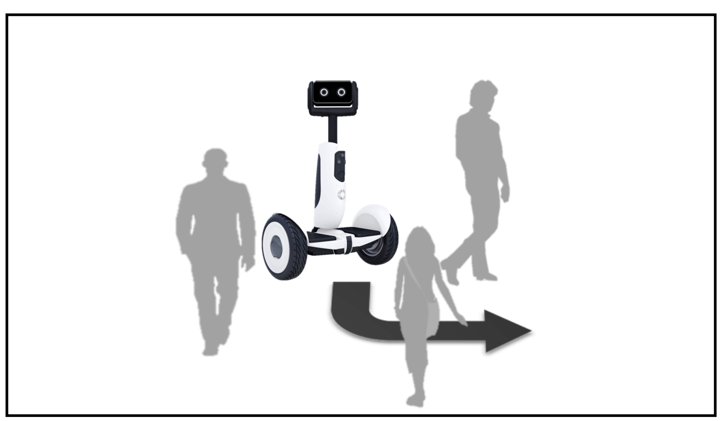
[1] S. Kreiss, L. Bertoni, A. Alahi, Composite Fields for Human Pose Estimation, CVPR'19

Perceiving

Socially-aware AI



Predicting



Planning

Problem Statement

Jointly reason and predict the future trajectories of **all** the agents in a scene **conditioned on the observed trajectories.**

Input: $X = X_1, X_2, \ldots, X_n$

Target: $Y = Y_1, Y_2, \ldots, Y_n$

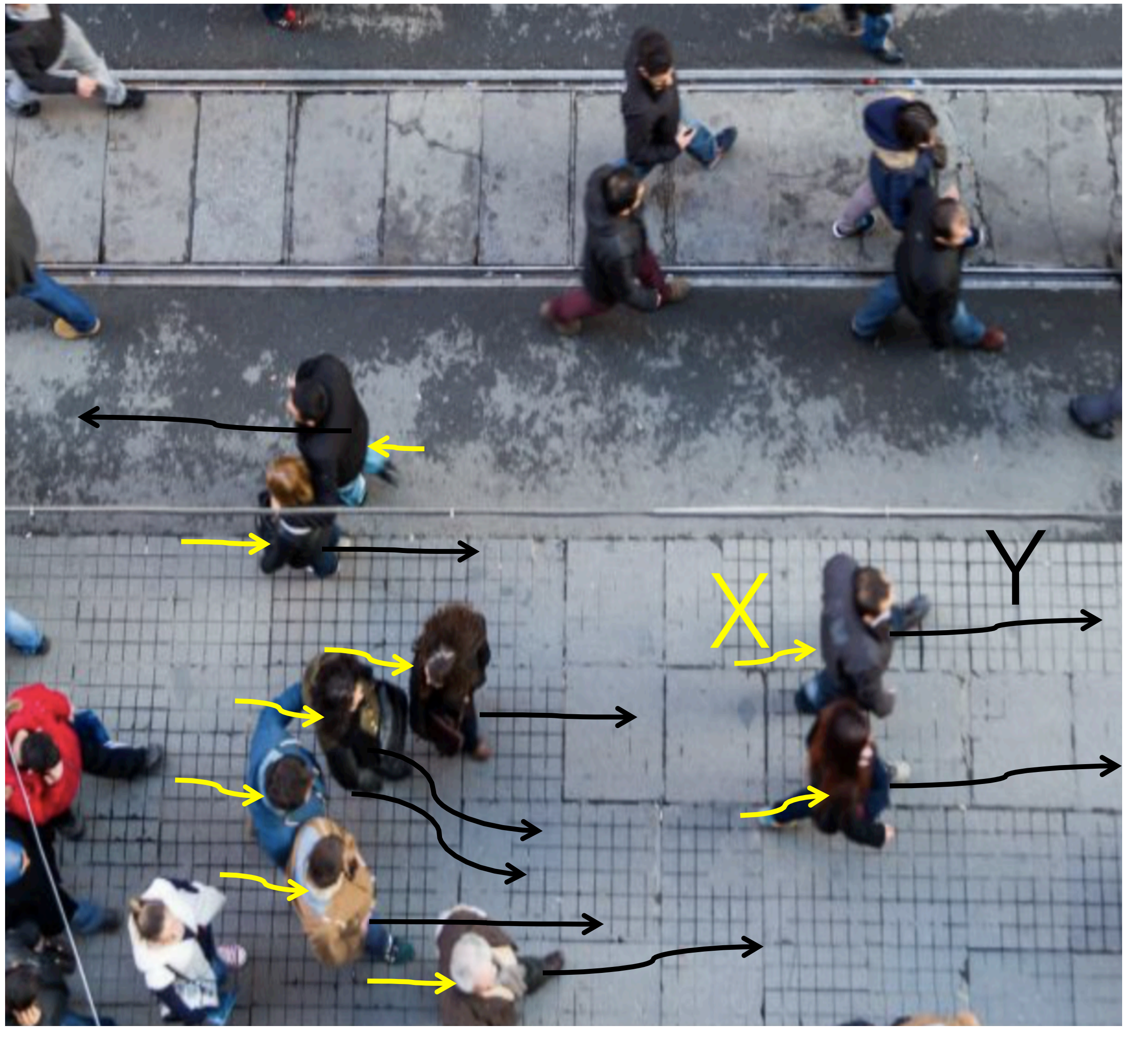
Output: $\hat{Y} = \hat{Y_1}, \hat{Y_2}, \ldots, \hat{Y_n}$

$$X_i = \{(x_t^i, y_t^i),\ 1 \leq t \leq T_{obs}\}$$

$$Y_i = \{(x_t^i, y_t^i),\ T_{obs}+1 \leq t \leq T_{pred}\}$$

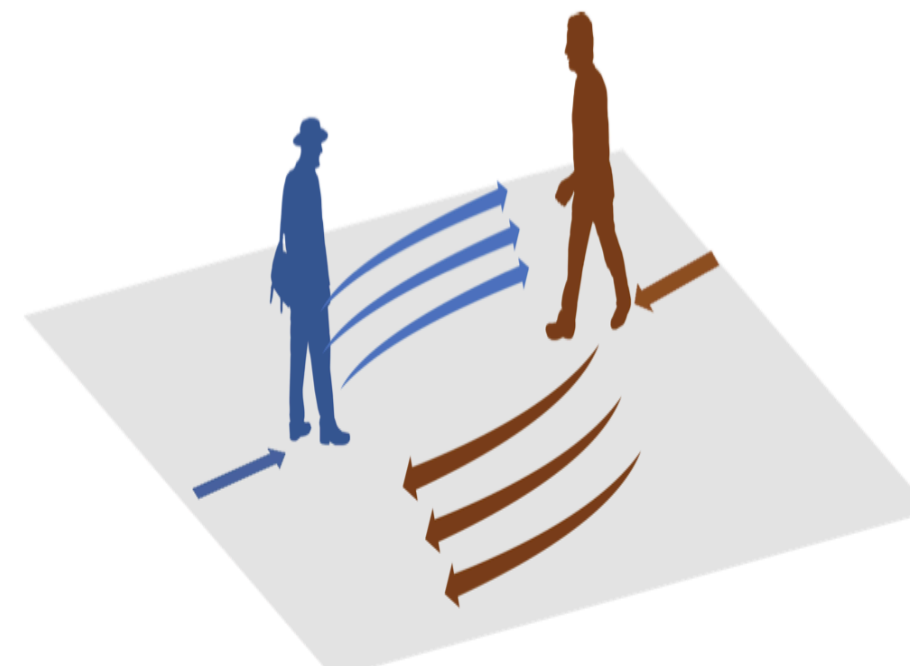
$$X_i^t = (x_t^i, y_t^i)$$

$$\hat{Y}_i^t = (\hat{x}_t^i, \hat{y}_t^i)$$



Challenges

1. Presence of Social Interactions
2. Socially Acceptable Trajectories
3. Multimodality



Previous works

Hand-crafted methods

e.g., Social Forces Model [1,2,3]

$F = F^{\text{attractive}} + F^{\text{repulsive}} \ldots$

Model Social sensitivity (reckless...) [3]

Fail to model:

- long-term dependencies
- Broad set of social interactions

[1] Helbing *et al.*, Physical review '95, [2] Leal-Taixé *et al.*,SVAC'13
[3] A. Robicquet et al., ECCV'18

Predicting

Previous works

Hand-crafted methods

e.g., Social Forces Model [1,2,3]

$F = F^{\text{attractive}} + F^{\text{repulsive}} \ldots$

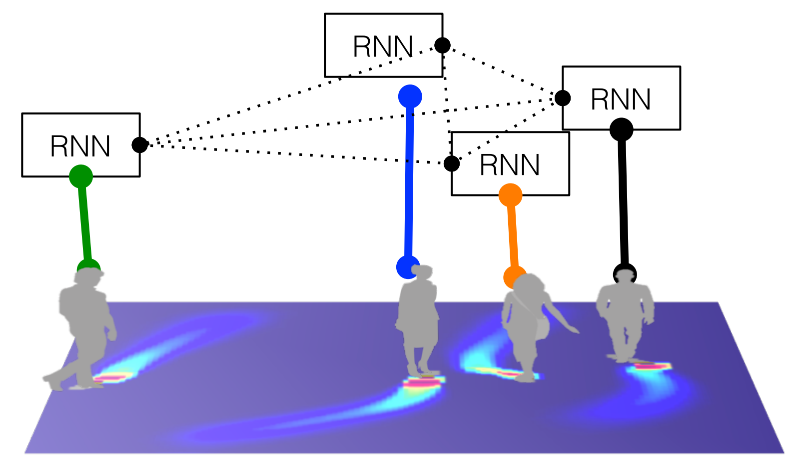
Model Social sensitivity

Fail to model:

- long-term dependencies
- Broad set of social interactions

Proposed work

DATA DRIVEN



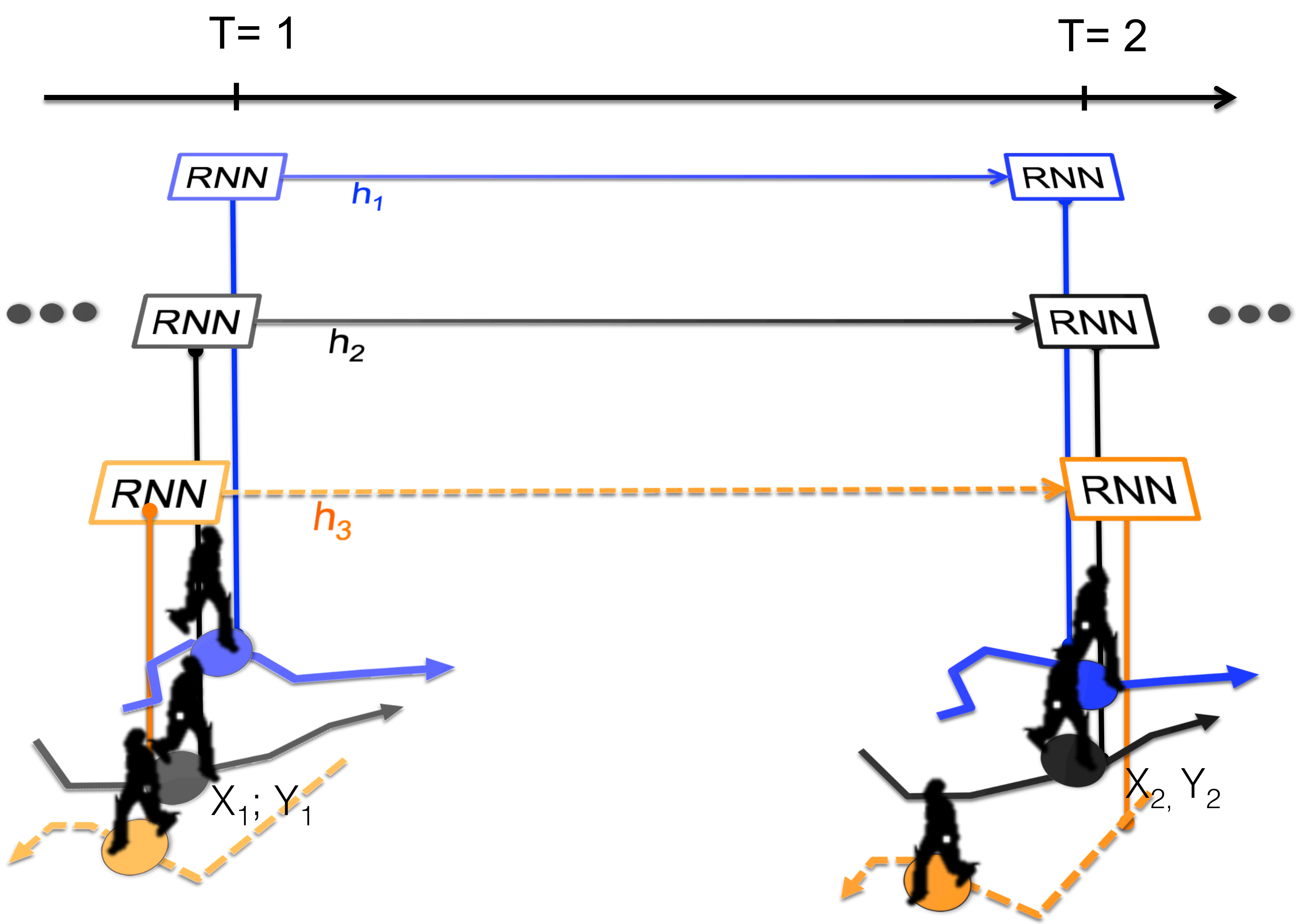
Recurrent Neural Network (RNN)

- Store dependencies in their hidden state
- Capacity to learn diverse behaviors

[1] Helbing *et al.*, Physical review '95, [2] Leal-Taixé *et al.*,SVAC'13
[3] A. Robicquet et al., ECCV'18

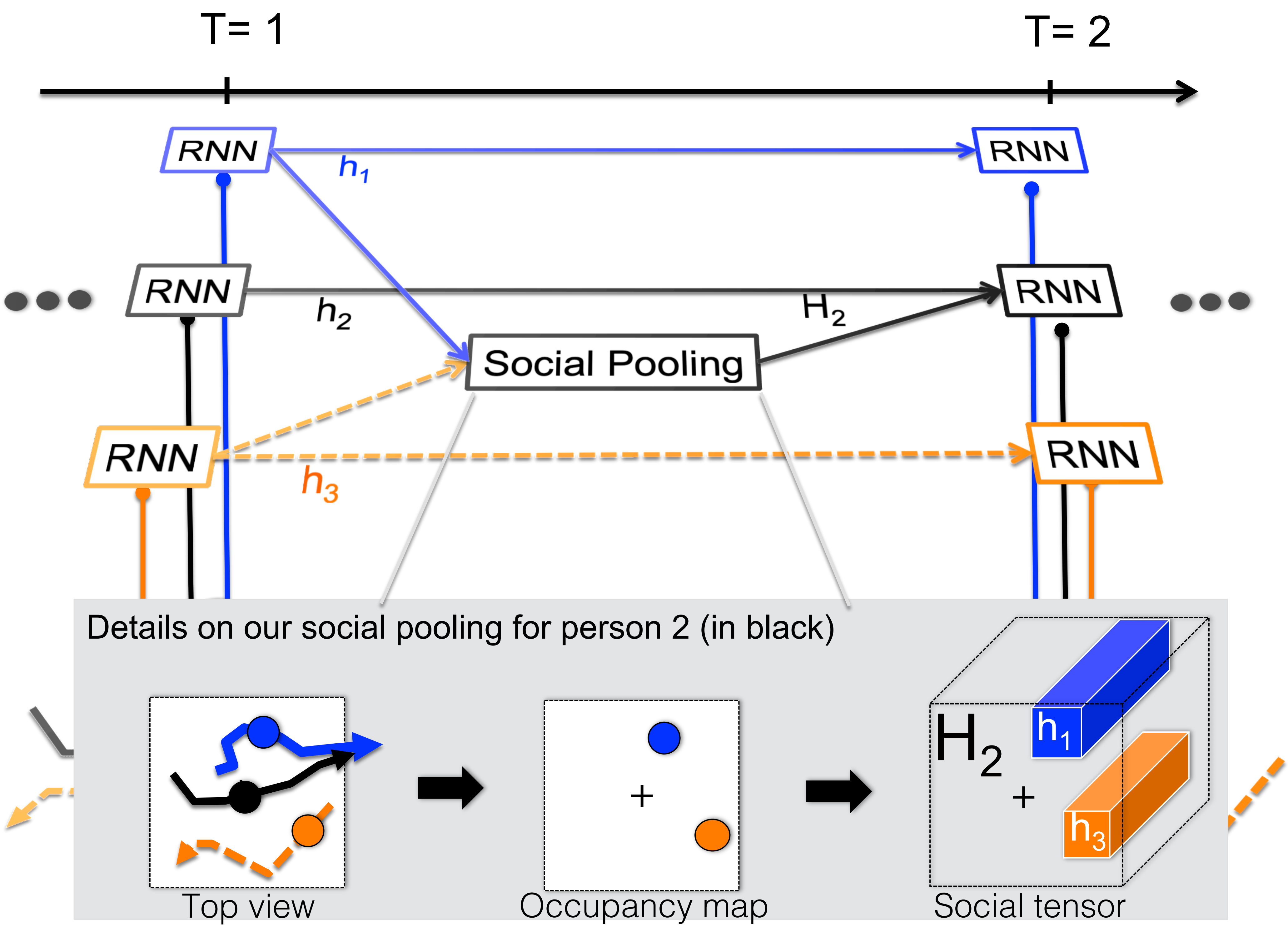
Forecasting human trajectories

With Recurrent Neural Network (RNN)



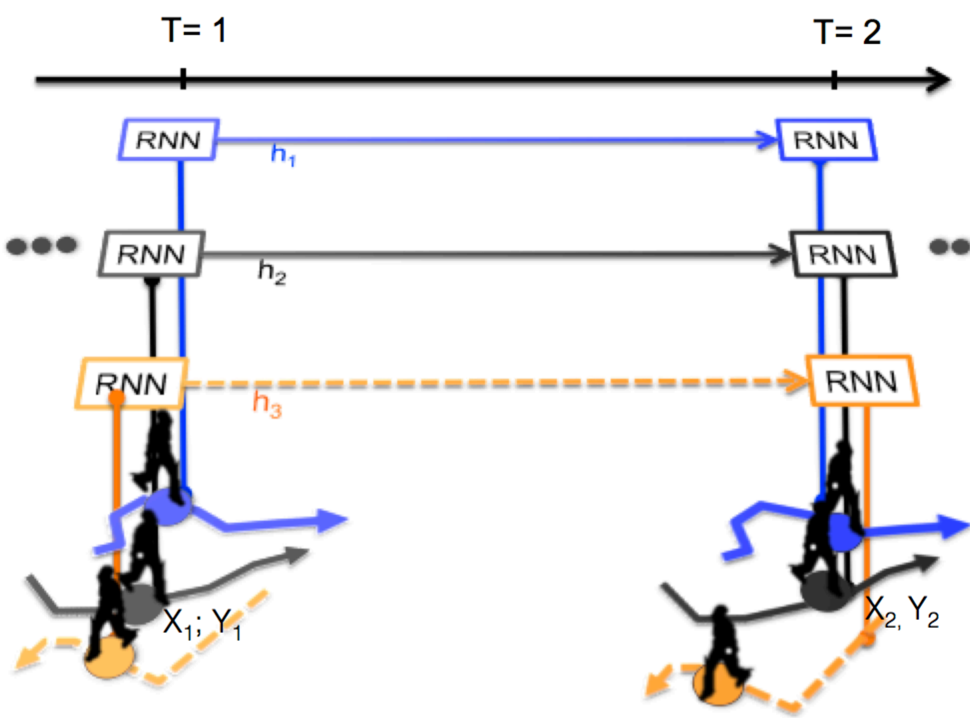
RNN “as is” will fail to model social interactions

Forecasting human trajectories



A. Alahi *et al.*, CVPR'16

Forecasting human trajectories

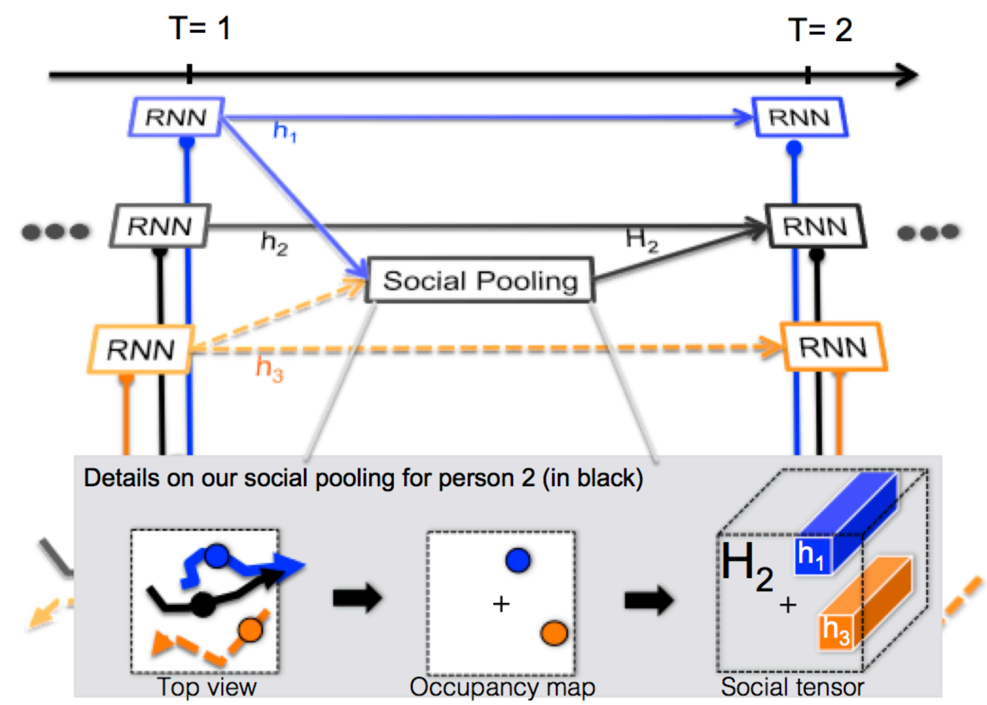


Single Sequence

$$h_t = f_{RNN}(h_{t-1},(x_t,y_t))$$

$$y = W_3 h_t$$

RNN "as is"



Multi-Sequences

$$H_t^i = \sum_{j \in N_i} 1_{mn}\left[x_t^j - x_t^i, y_t^j - y_t^i\right] h_{t-1}^j$$

$$a_t^i = \phi(H_t^i; W_a)$$

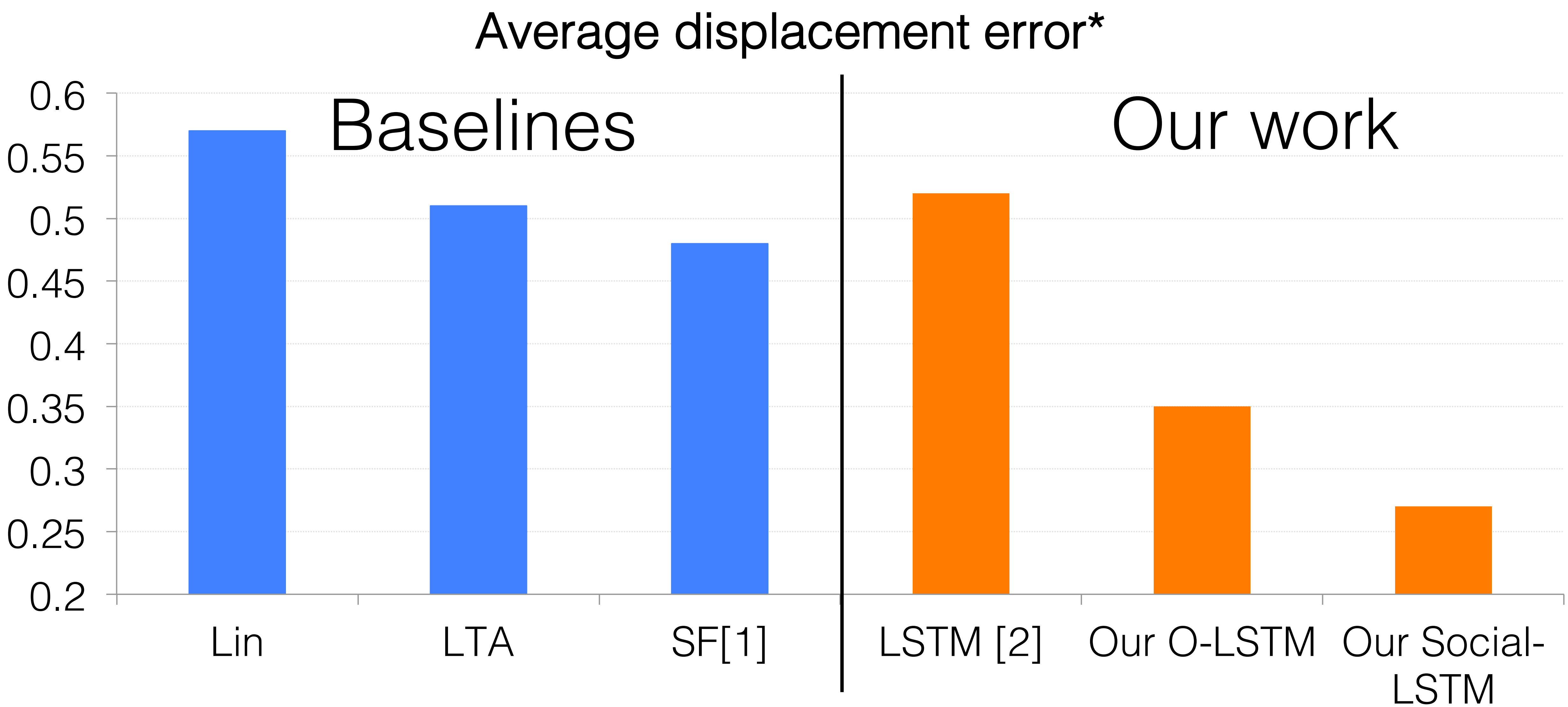
$$e_t^i = \phi(x_t^i, y_t^i; W_e)$$

$$h_t^i = f_{RNN}(h_{t-1}^i, e_t^i, a_t^i; W_l)$$

$$y = W_3 h_t$$

A. Alahi *et al.*, CVPR'16

Quantitative results



[1] Yamaguchi & Berg, CVPR'11
[2] Graves, '14
[3] Lerner & Lischinski, wiley '07

* On UCY [3] dataset

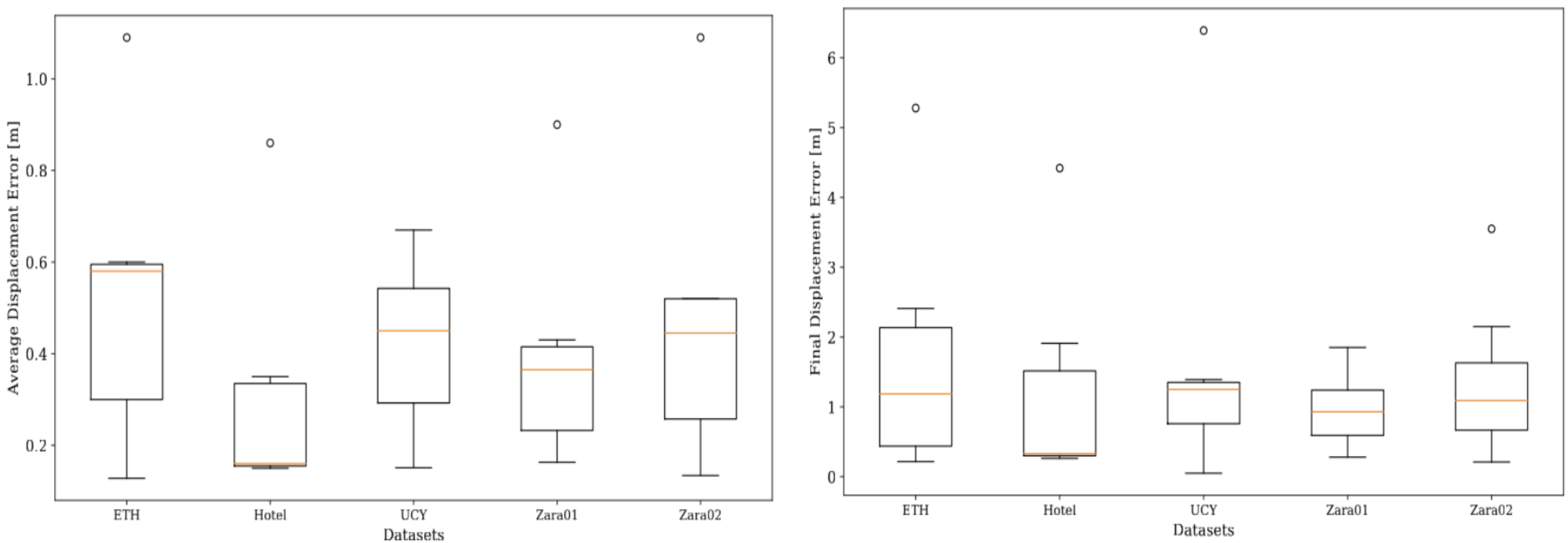
A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, S. Savarese, CVPR'16

DATA ?

Discrepancy in Published Results*

Published papers report up to 87% discrepancy for the ADE (figure a) and 99% for the FDE (figure b) for **simple Vanilla LSTM** results!

The results in both metrics are somewhat contradictory: eg. High ADE but smaller FDE for ETH.



(a) Average displacement error

(b) Final displacement error

Reasons.

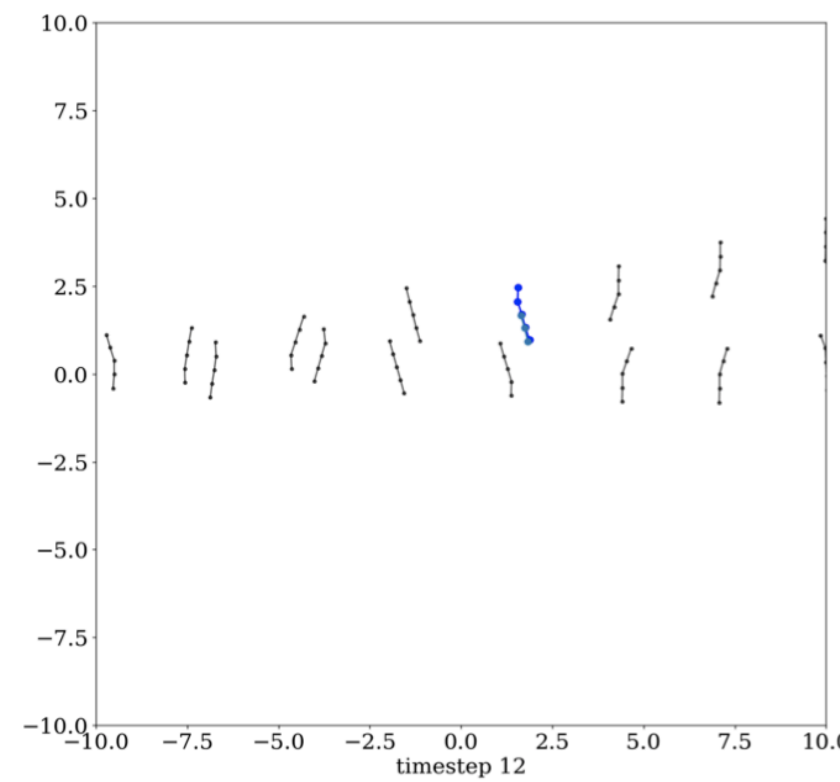
1. Indexing of Trajectories is different
2. Preprocessing of trajectories is *non-uniform* across methods.
3. *Defined categorization* of trajectories into linear and non-linear is missing.
4. Absence of a good Train-Test Split

Solutions!

1. Defined Indexing of Trajectories
2. Defined Preprocessing
3. *Defined categorization* of trajectories into linear and non-linear.
4. Carefully designed Train-Test Split

We propose TrajNet++

1. Trajnet++ : Carefully indexed real-world trajectories involving social interactions. Trajectories are divided into 4 types for fair evaluation. Static (Type I), Linear (Type II), Socially Interacting (Type III), Others (Type IV).



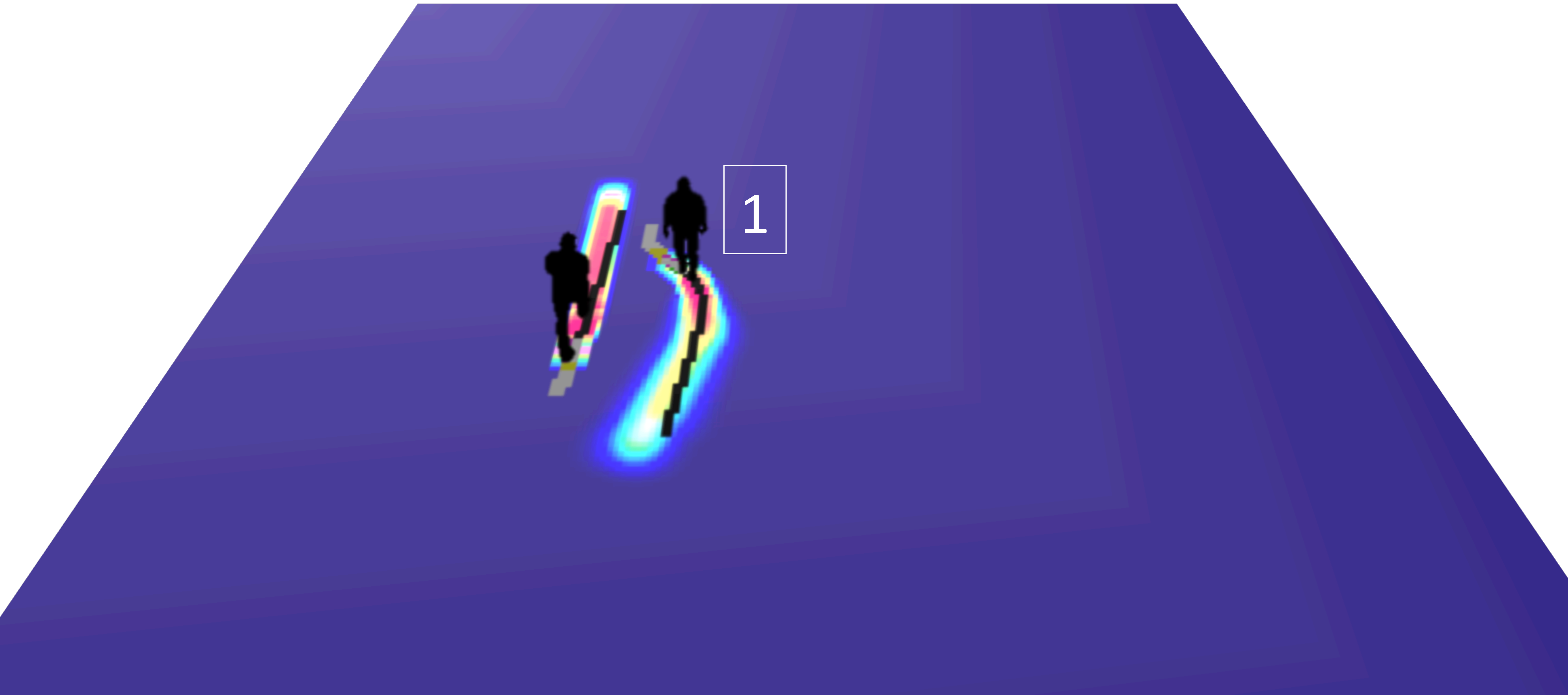
2. Synthetic Data : Clean trajectories affected only by neighbourhood interactions. Ideal sanity checks to evaluate the performance of different components.

Email me alexandre.alahi@epfl.ch

Our method learned to avoid a pedestrian

- Black line is the ground truth trajectory
- Gray line is the past
- Heatmap is the predicted distribution of our method

Our method learned to avoid a pedestrian

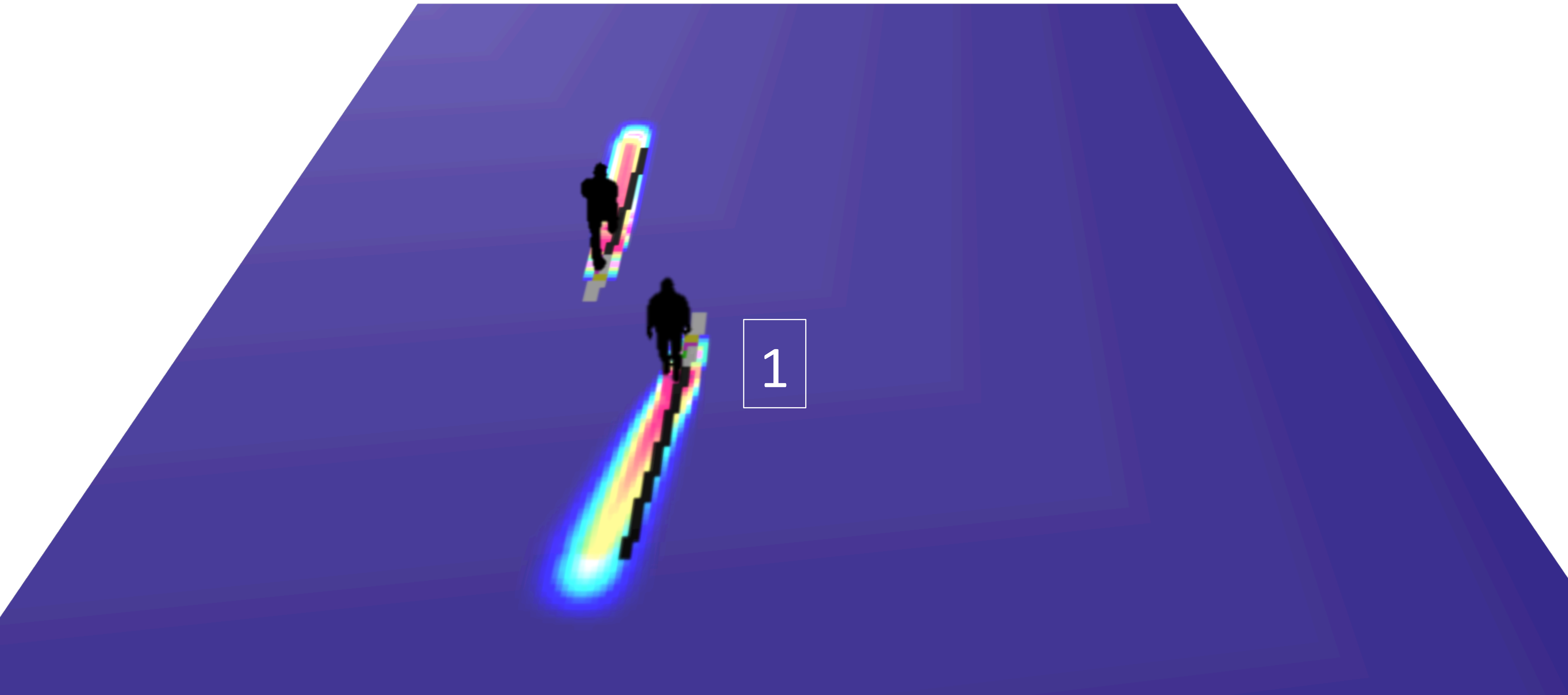


- Black line is the ground truth trajectory
- Gray line is the past
- Heatmap is the predicted distribution of our method

Our method learned to avoid a pedestrian

- Black line is the ground truth trajectory
- Gray line is the past
- Heatmap is the predicted distribution of our method

Our method learned to avoid a pedestrian



- Black line is the ground truth trajectory
- Gray line is the past
- Heatmap is the predicted distribution of our method

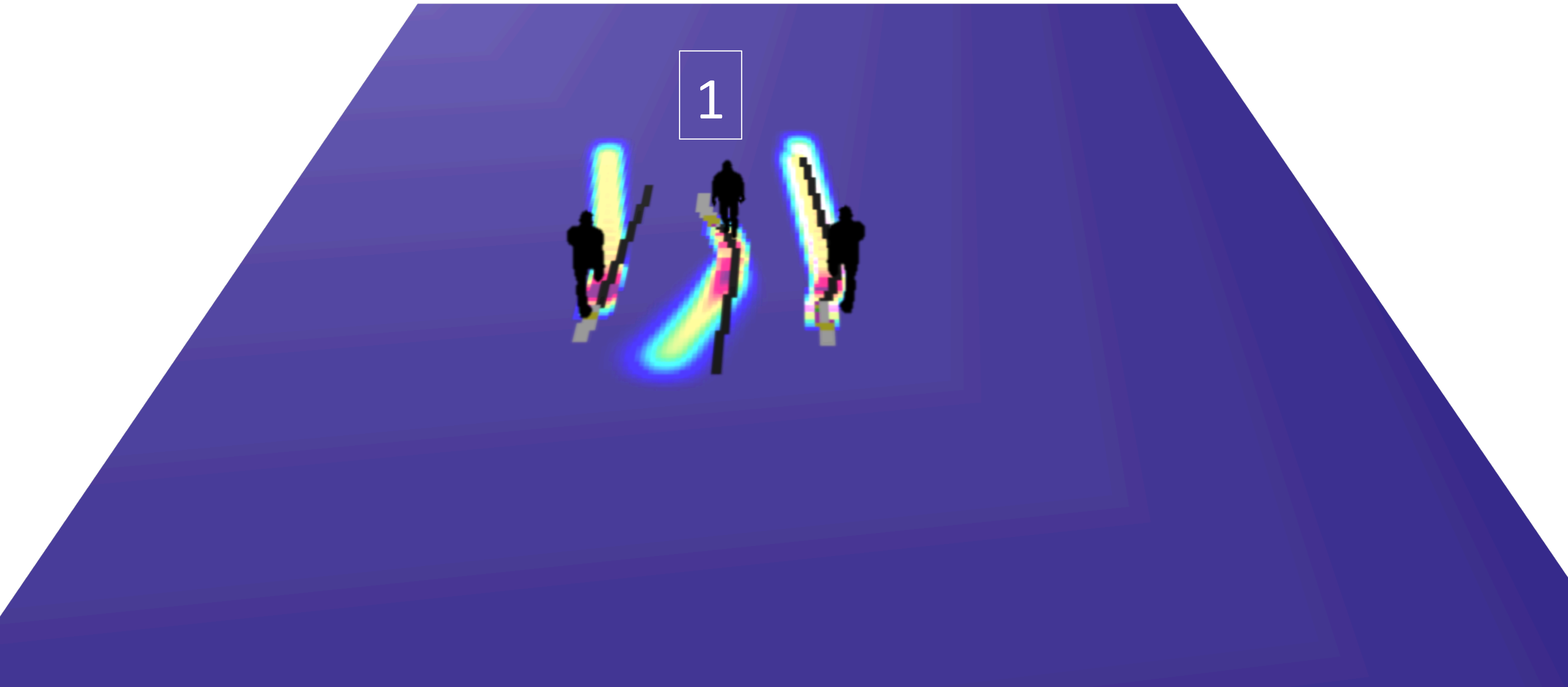
Our method learned to turn around a group

- Black line is the ground truth trajectory
- Gray line is the past
- Heatmap is the predicted distribution of our method

Our method learned to turn around a group

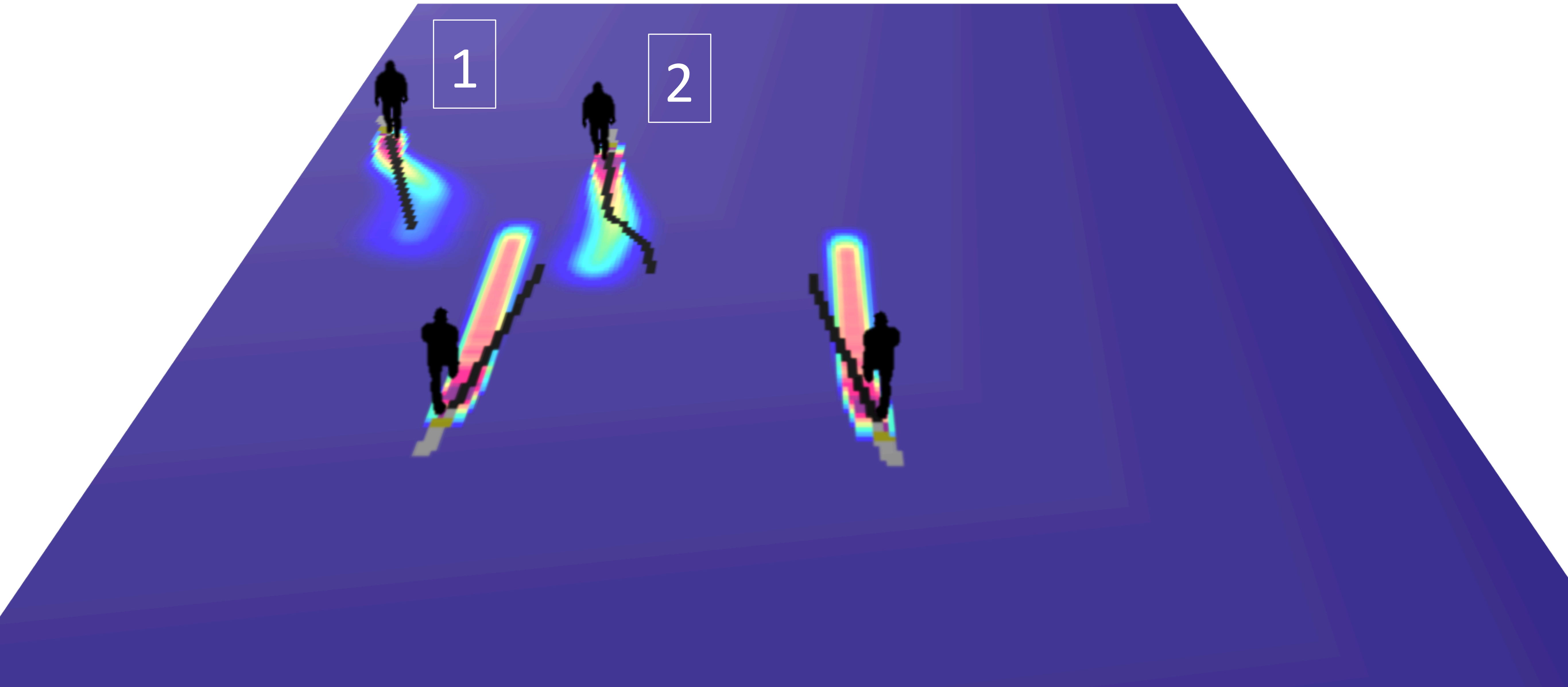
- Black line is the ground truth trajectory
- Gray line is the past
- Heatmap is the predicted distribution of our method

Walk in-between pedestrians



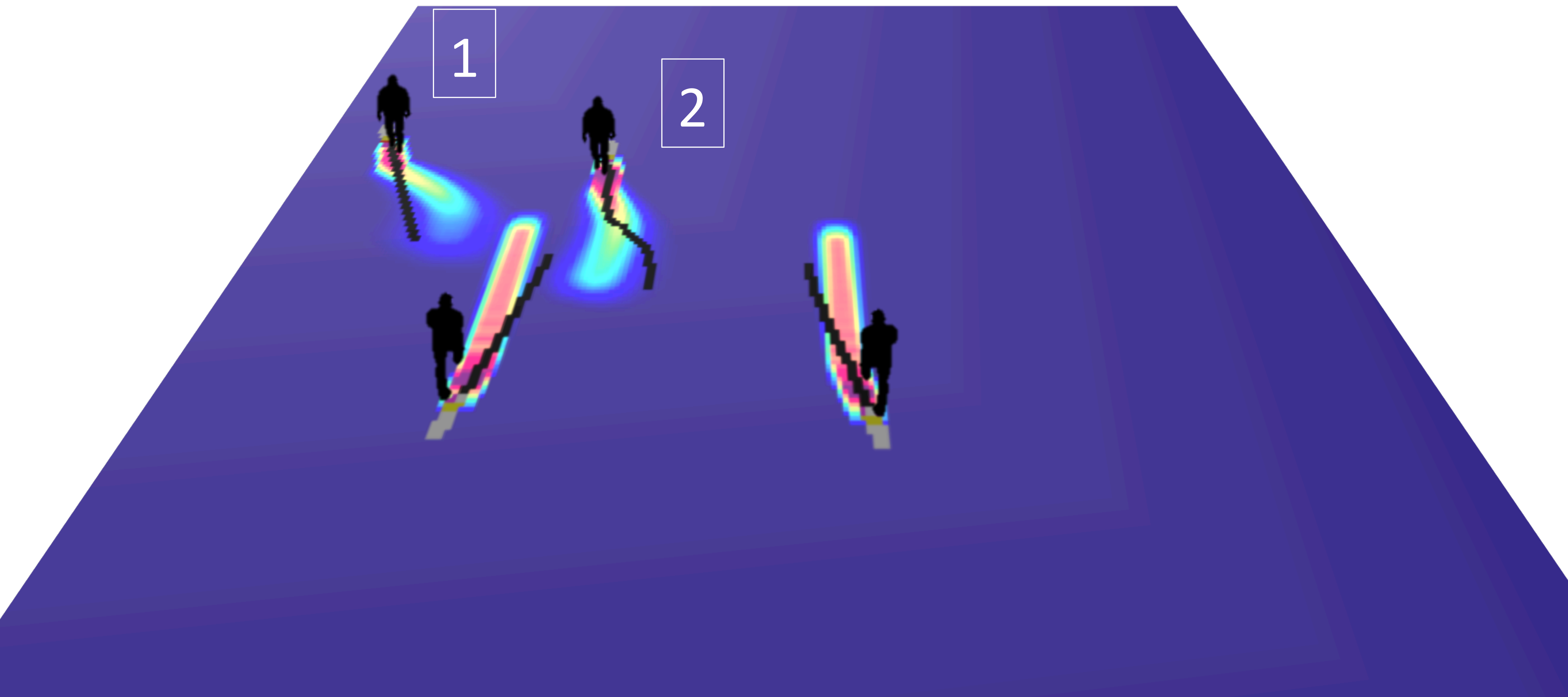
- Black line is the ground truth trajectory
- Gray line is the past
- Heatmap is the predicted distribution of our method

Walk in-between pedestrians



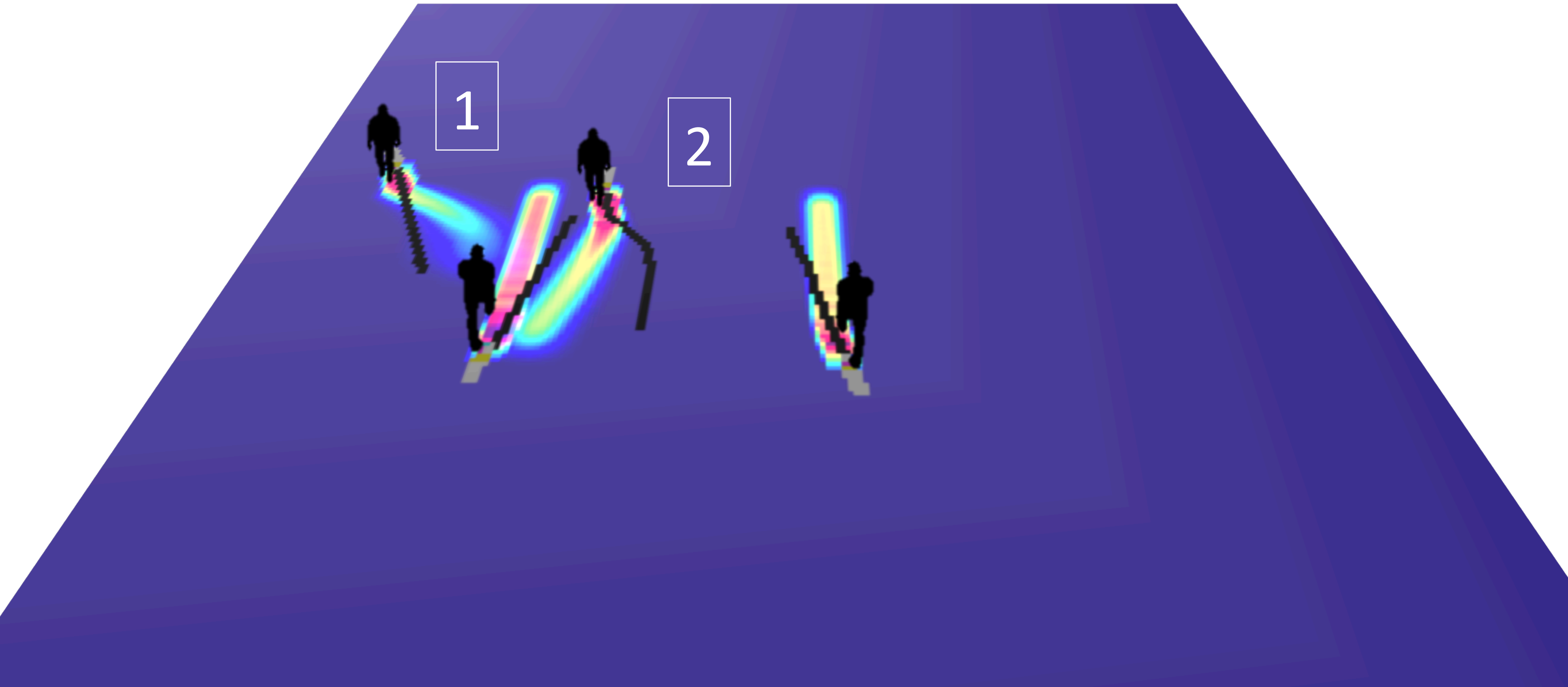
- Black line is the ground truth trajectory
- Gray line is the past
- Heatmap is the predicted distribution of our method

Walk in-between pedestrians



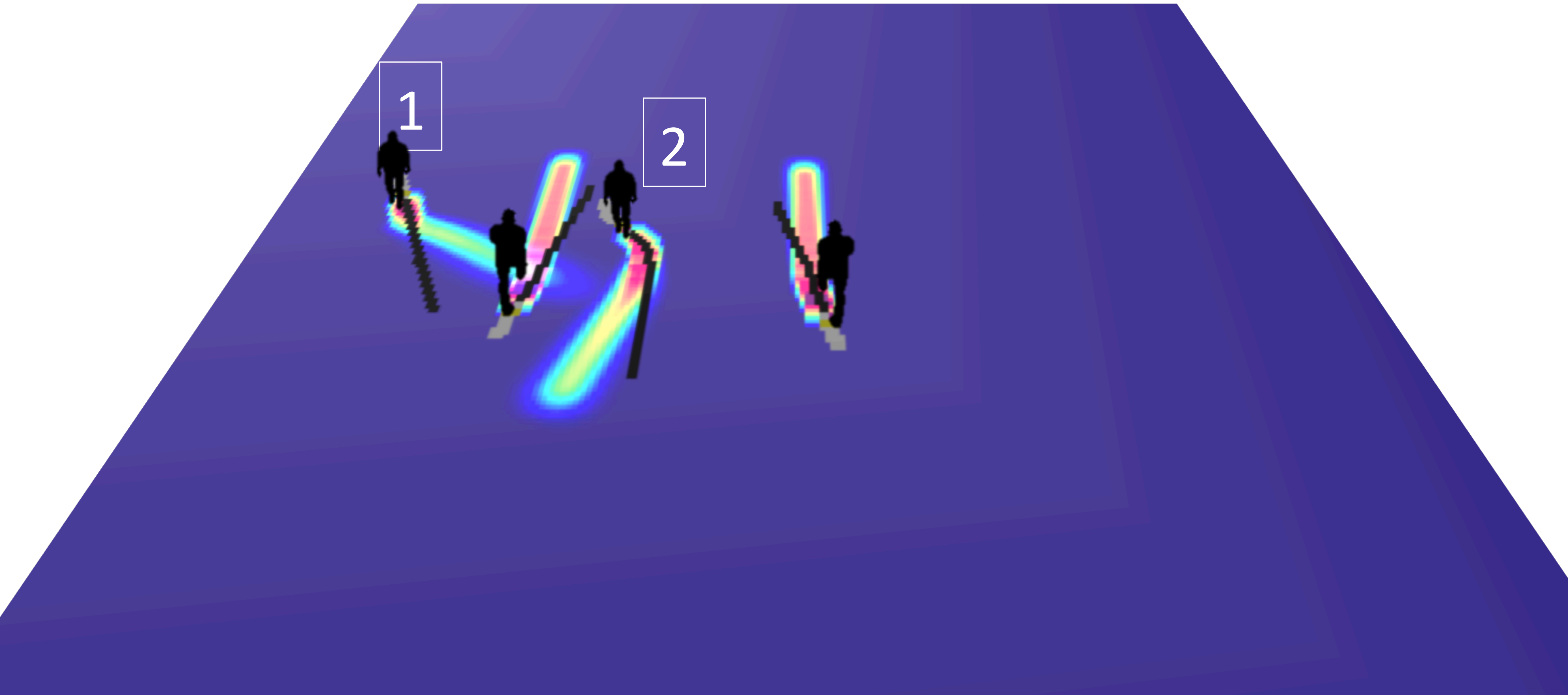
- Black line is the ground truth trajectory
- Gray line is the past
- Heatmap is the predicted distribution of our method

Walk in-between pedestrians



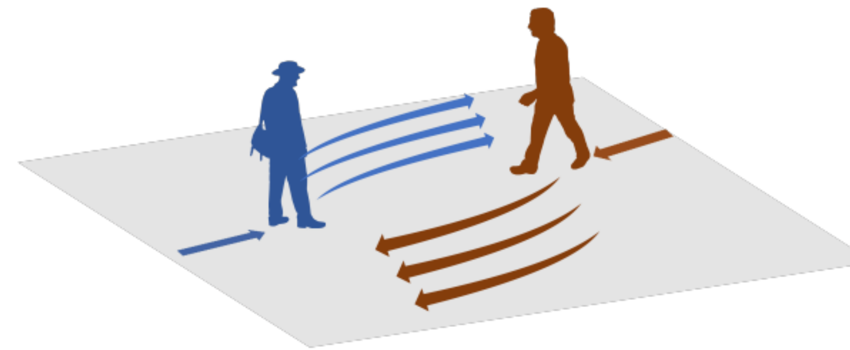
- Black line is the ground truth trajectory
- Gray line is the past
- Heatmap is the predicted distribution of our method

Walk in-between pedestrians



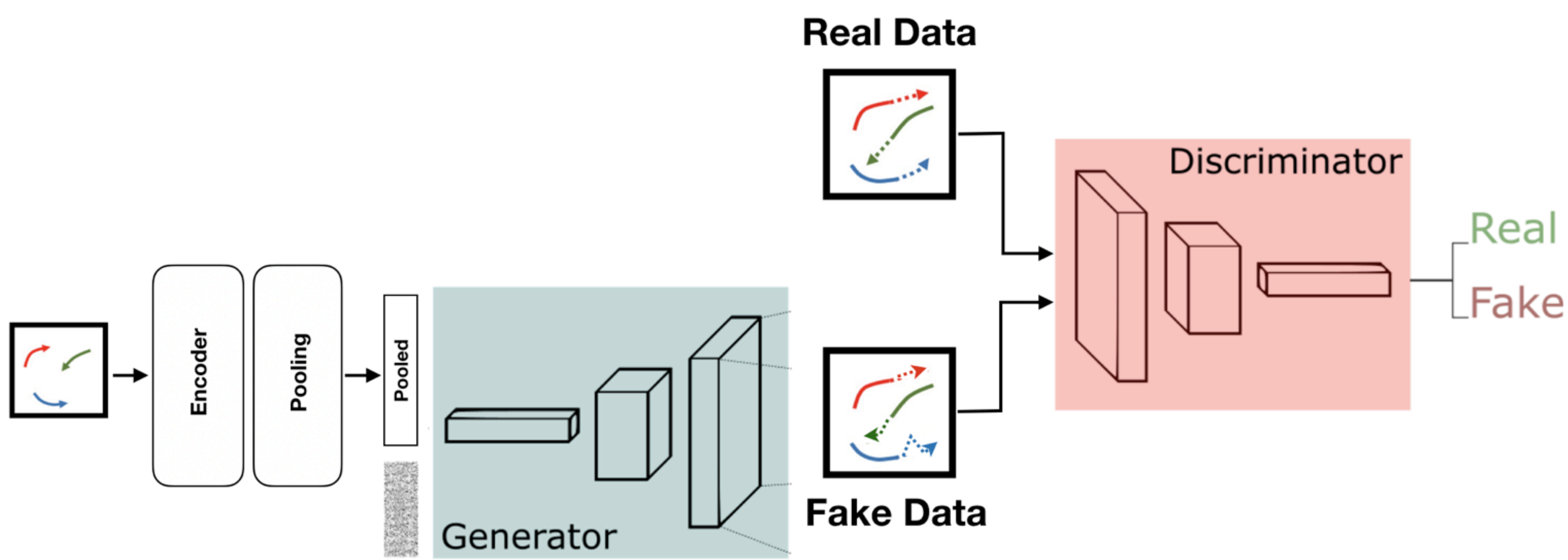
- Black line is the ground truth trajectory
- Gray line is the past
- Heatmap is the predicted distribution of our method

Generative models



- GAN: Minmax game between :

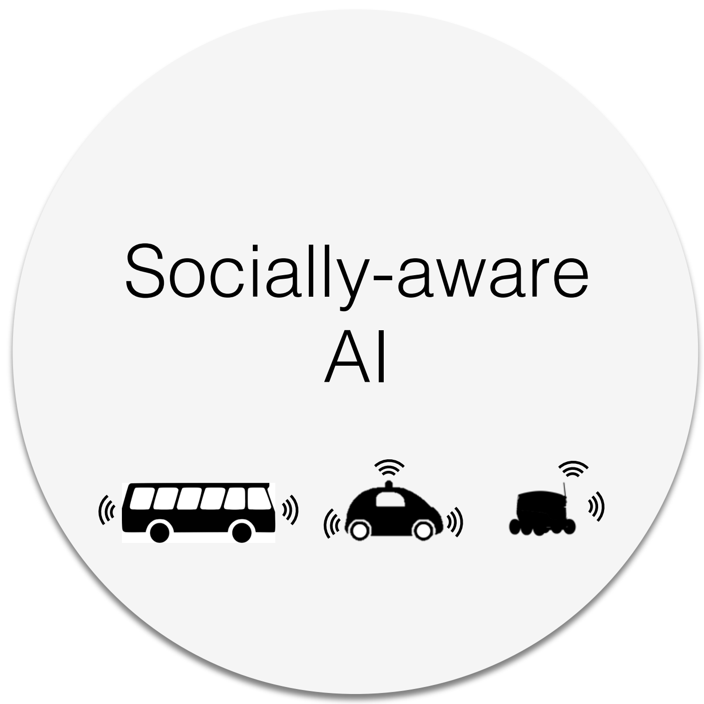
G: Generator of social plausible interactions
D: Discriminator classifying real social interactions

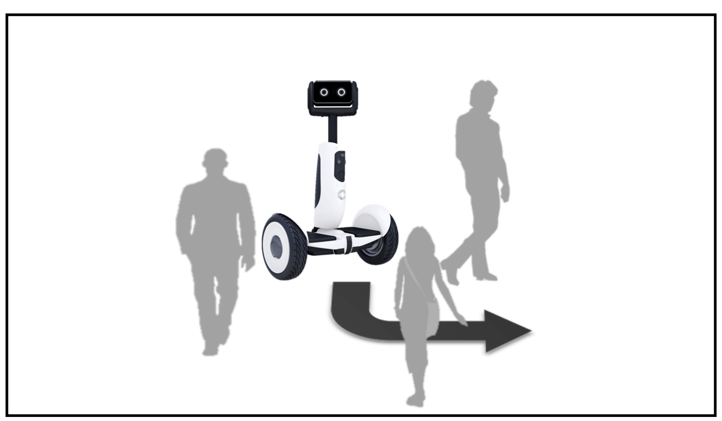


⇒ Model Social constraints in G & D [1]
⇒ Tackle mode collapse G, catastrophic forgetting for G & D [2]

[1] Gupta *et al.*, CVPR'18
[2] Yuejiang Liu & Parth Kothari, Collaborative Sampling in GAN, arxiv'19

Perceiving

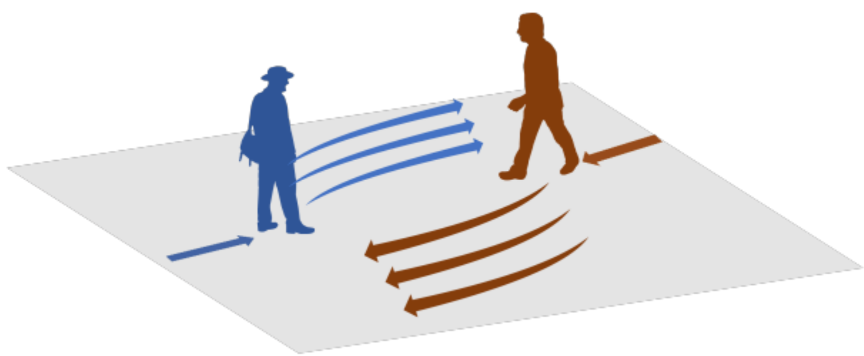


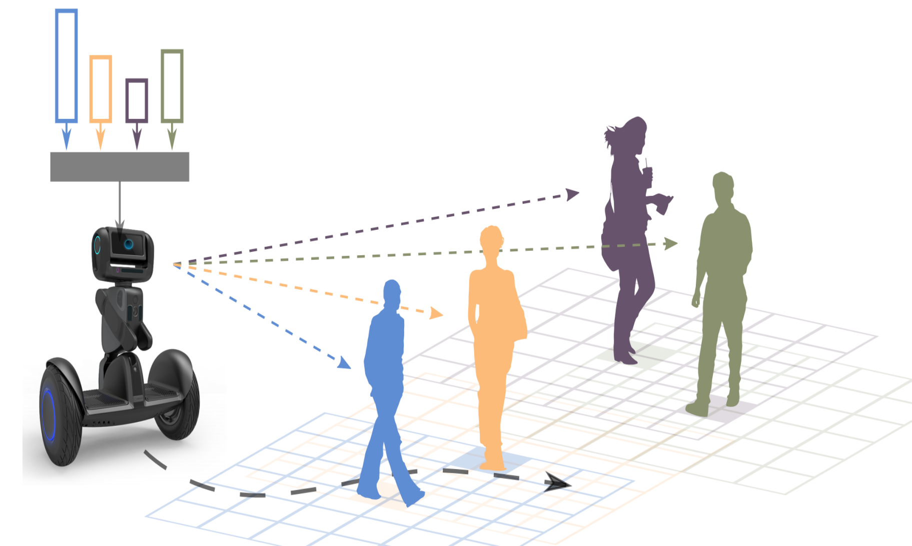


Planning

Predicting

From Prediction to Navigation





Socially-aware Trajectory Prediction

Social Force [1]
Discrete Choice Model [2]
Social-LSTM [3]

Socially-aware Robot Navigation

Sequential approach [4]
Interacting GPs [5]
CADRL [6,7]

Crowd-aware Robot Navigation

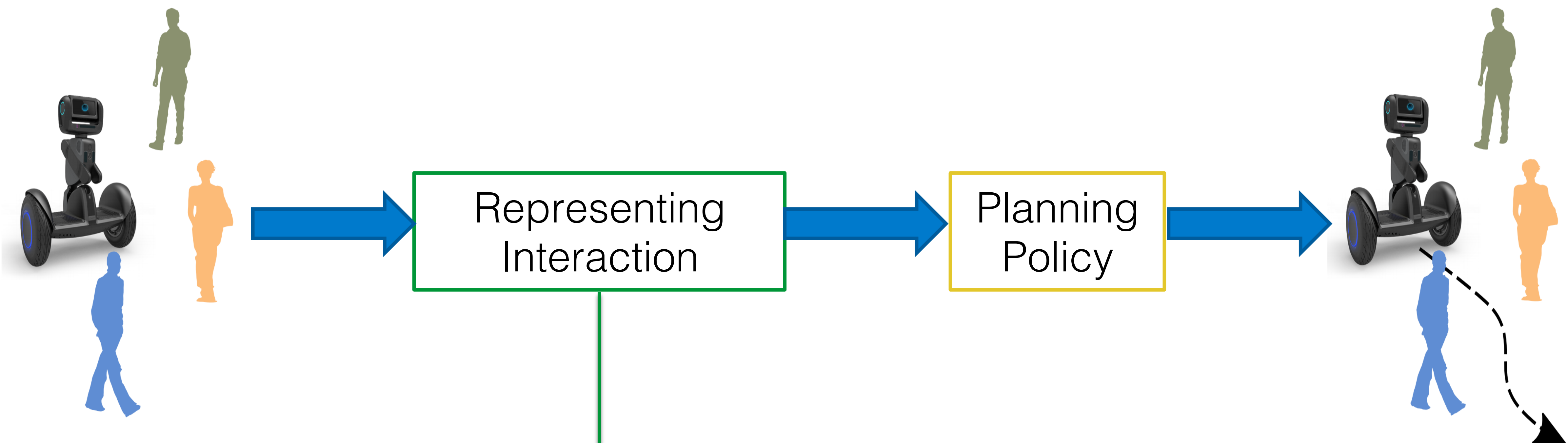
LSTM-CADRL [8]
Our recent work [9]

[1] Helbing, D. , et al. 1995
[2] Antonini, G., et al. 2006
[3] Alahi, A., et al. 2016

[4] Aoude, G. S., et al. 2013
[5] Trautman, P., et al. 2010
[6,7] Chen, Y., et al. 2017

[8] Everett, M., et al. 2018
[9] Chen, C., et al. 2019

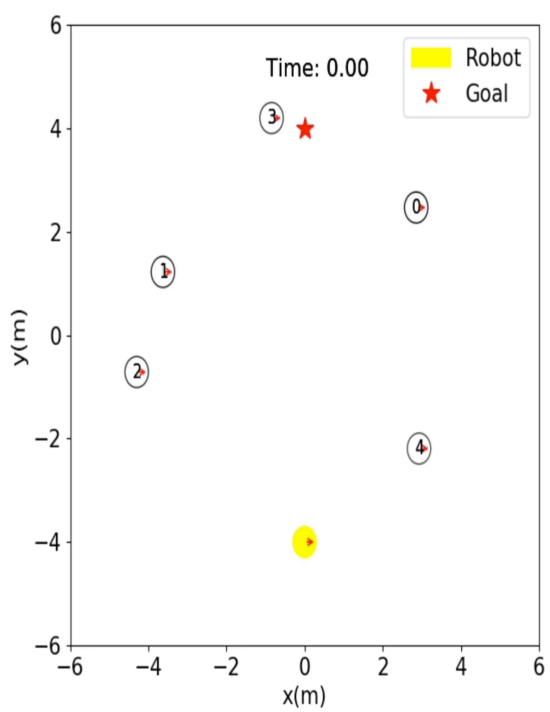
Crowd-Robot Interaction



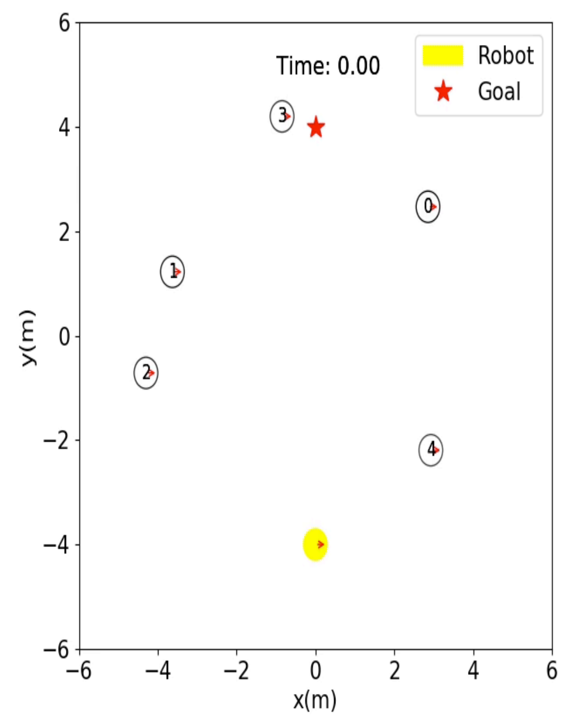
- Jointly model Human-Robot as well as Human-Human interactions
- Aggregate interactions with a self-attention mechanism

Crowd-Robot Interaction

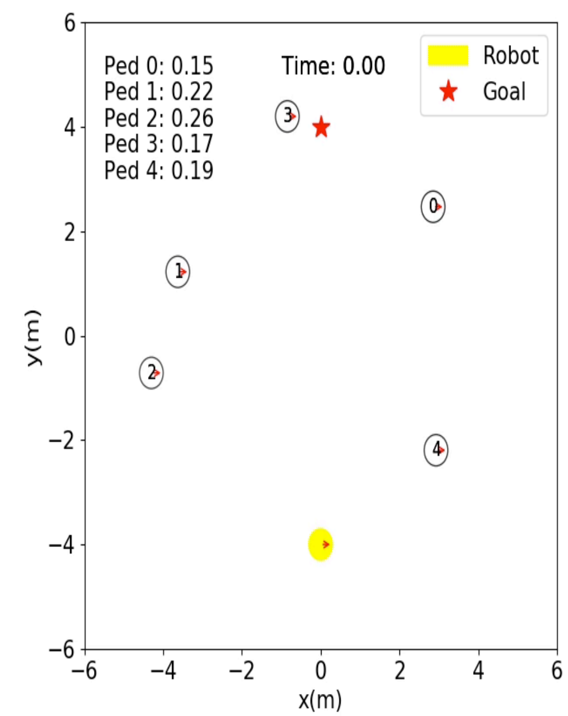
Experiments



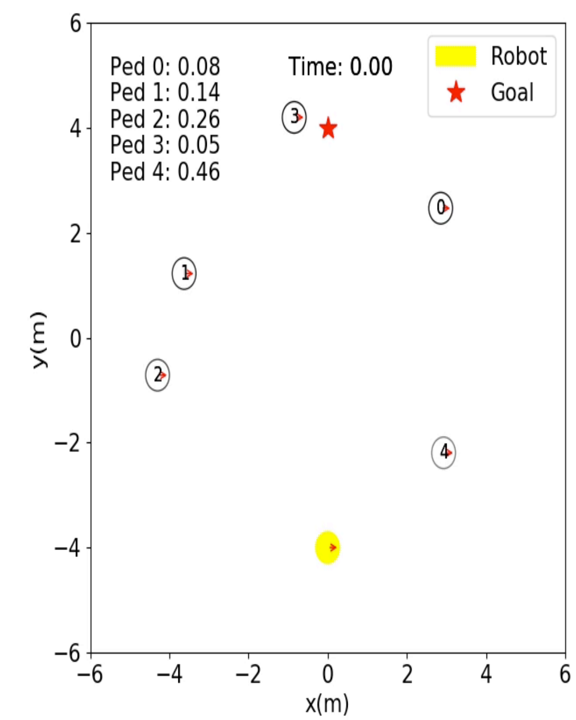
CADRL
(maximin)



LSTM-RL
(sequence)



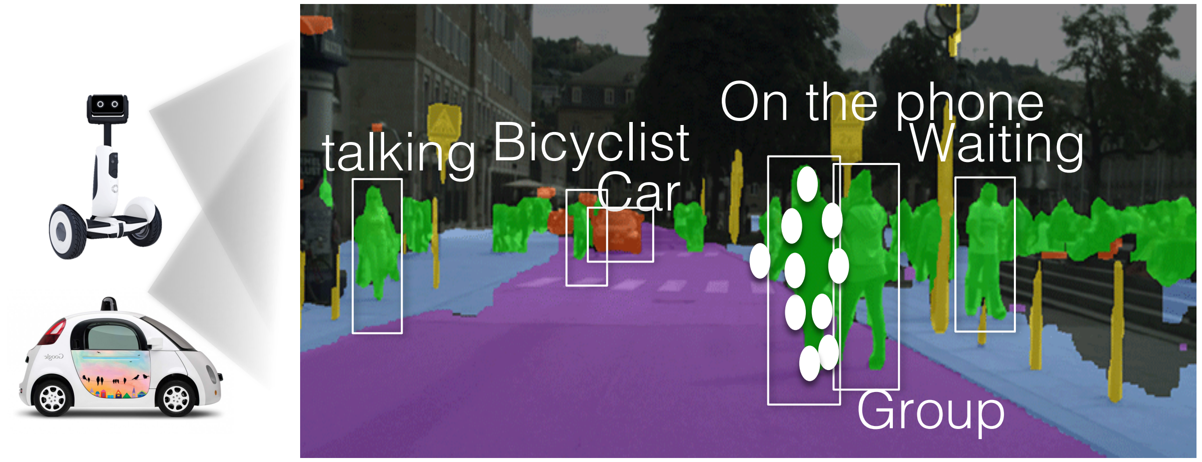
Our SARL [1]
(Human-Robot)



Our LM-SARL [1]
(Crowd-Robot)

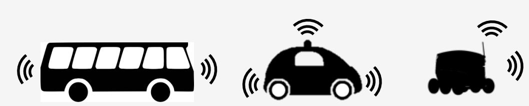
[1] C. Chen et al., Crowd-Robot Interaction: Crowd-aware Robot Navigation with Attention-based Deep Reinforcement Learning, ICRA'19


agepoly

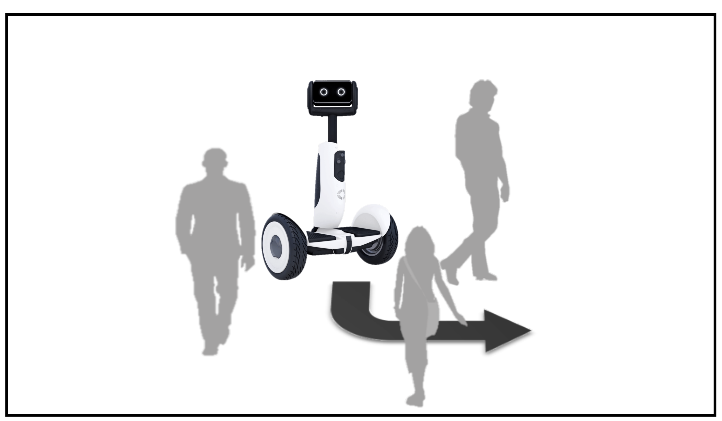


Perceiving

Socially-aware
AI

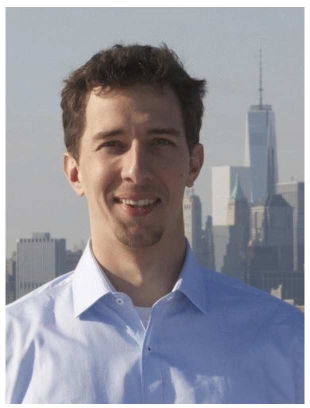


Predicting



Planning

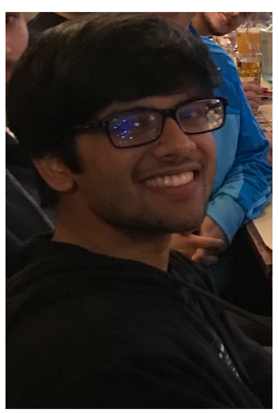
Thank you!
Our lab members & sponsors:


Dr Sven Kreiss


George Adaimi,


Hossein Bahari,


Lorenzo Bertoni


Parth Kothari


Yuejiang Liu


Saeed Saadatnejad


Brian Sifringer

HITACHI RICHEMONT HONDA The Power of Dreams Schindler SAMSUNG

Prof. Alex Alahi – VITA lab – EPFL

#Open Science

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

Perception:
[1] S. Kreiss et al., Composite Fields for Human Pose Estimation, CVPR'19

Prediction:
[2] A. Gupta et al., Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks, CVPR'18
[3] Y. Liu, et al., Collaborative Sampling in Generative Adversarial Networks, arxiv'19

Planning:
[4] C. Chen et al., Crowd-Robot Interaction: Crowd-aware Robot Navigation with Attention-based Deep Reinforcement Learning, ICRA'19