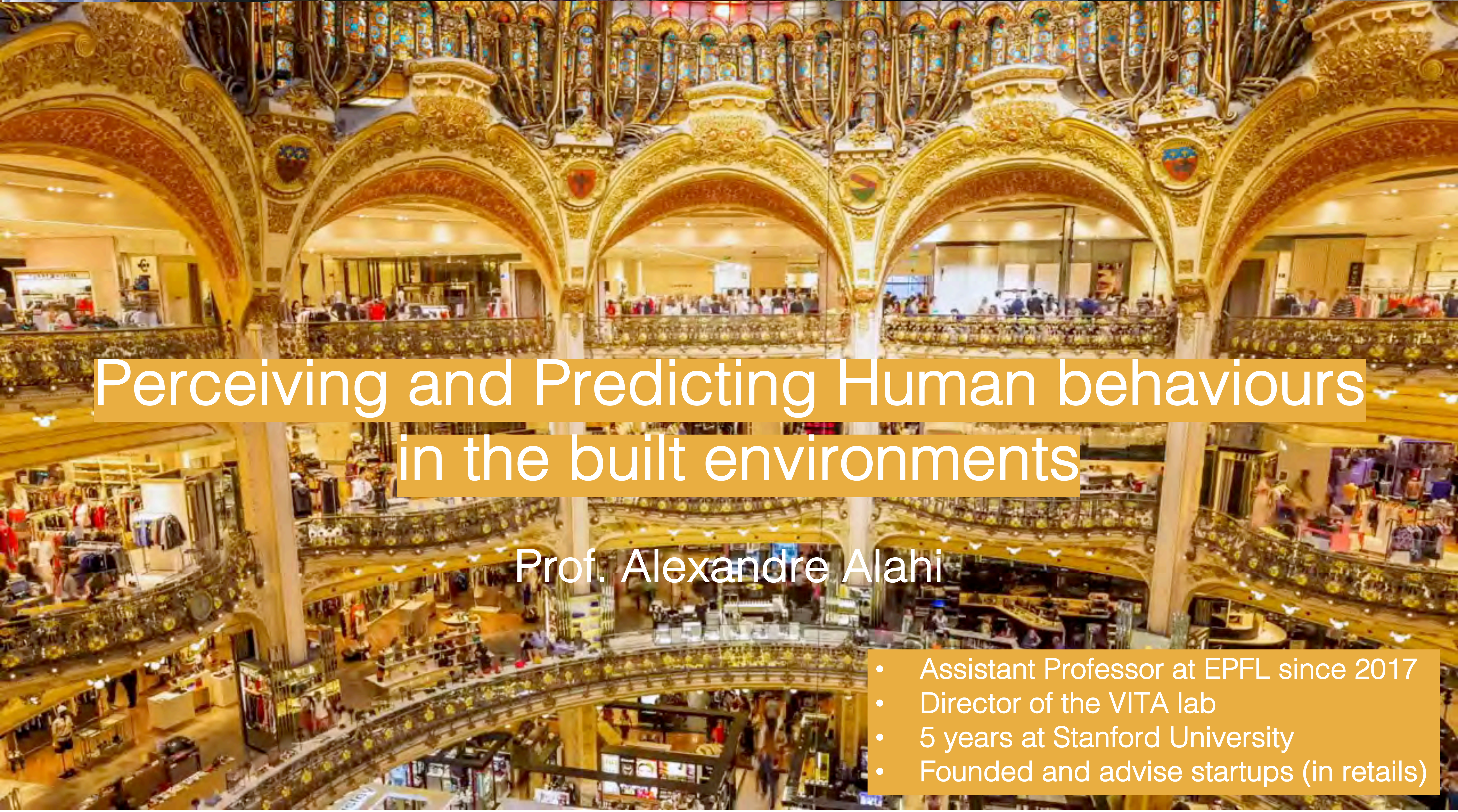


Perceiving and Predicting Human behaviours in the built environments

Prof. Alexandre Alahi

- Assistant Professor at EPFL since 2017
- Director of the VITA lab
- 5 years at Stanford University
- Founded and advise startups (in retails)

VITA

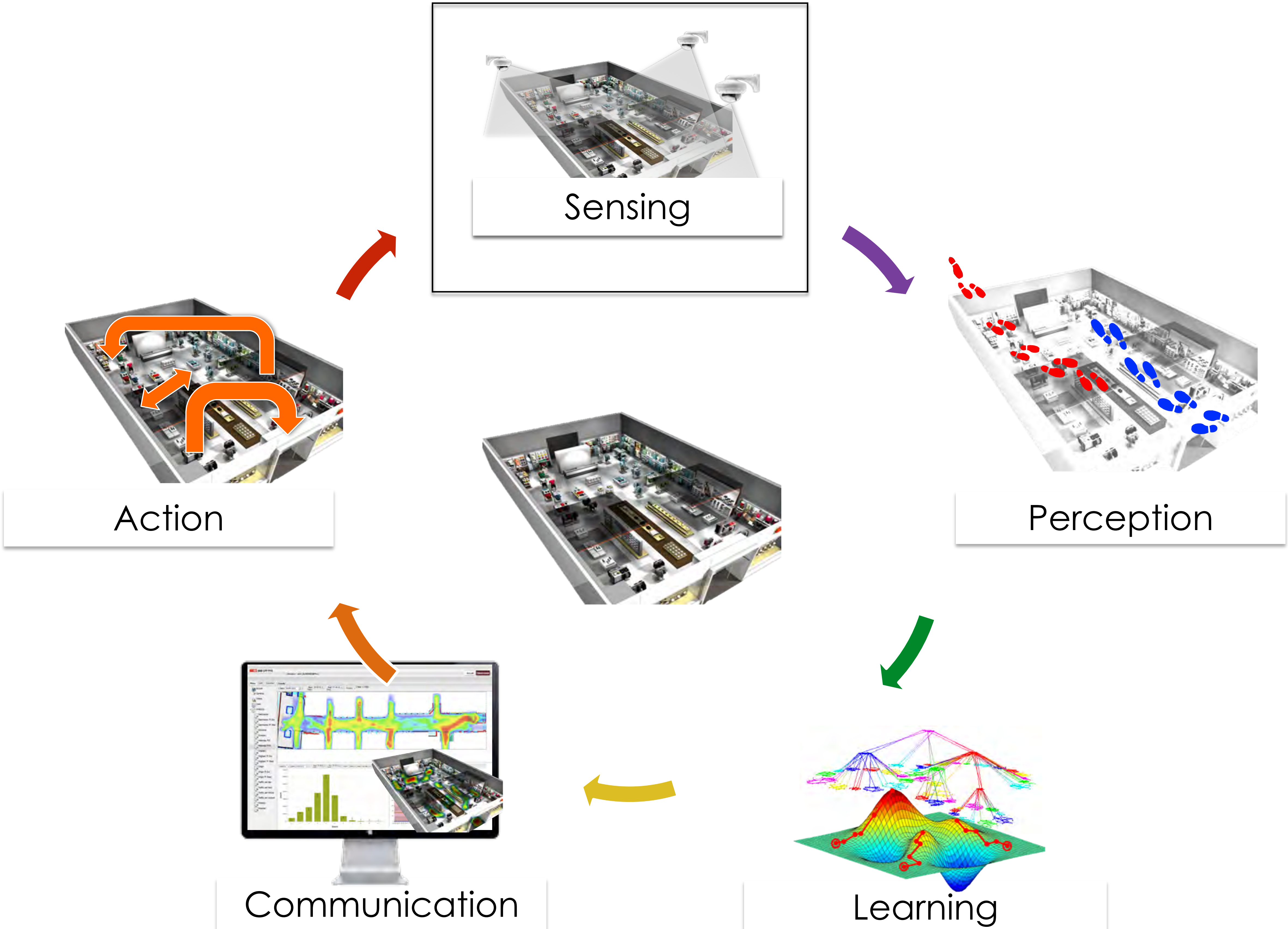
#Open Science

“Man is by nature a social animal” - Aristotle

How can machines learn **human-human** interactions?

How can machines learn **human-space** interactions?

AI for the built environments



2 Terminals

132 sensors

>20,000 m^2

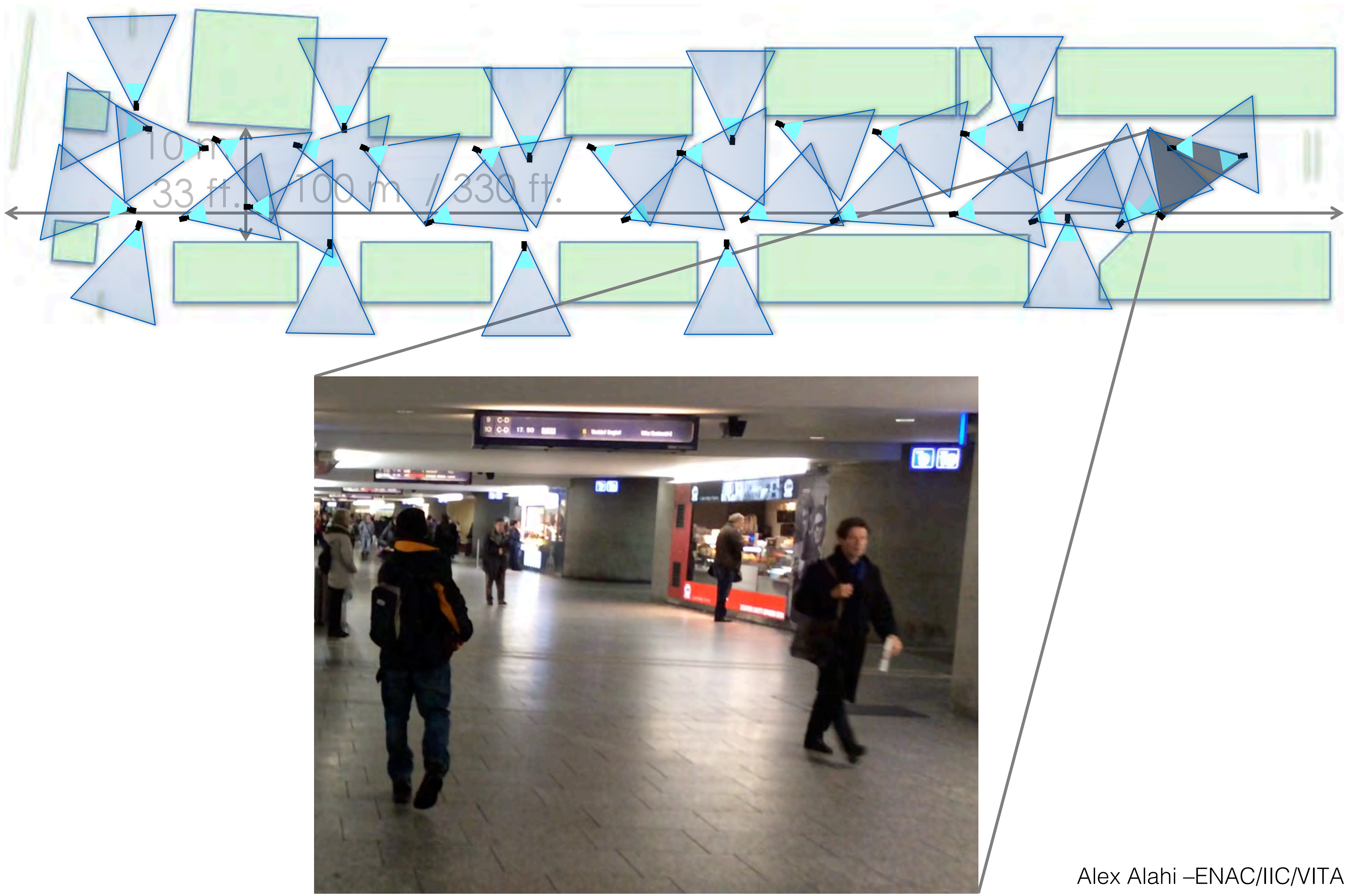
24/7 over a year

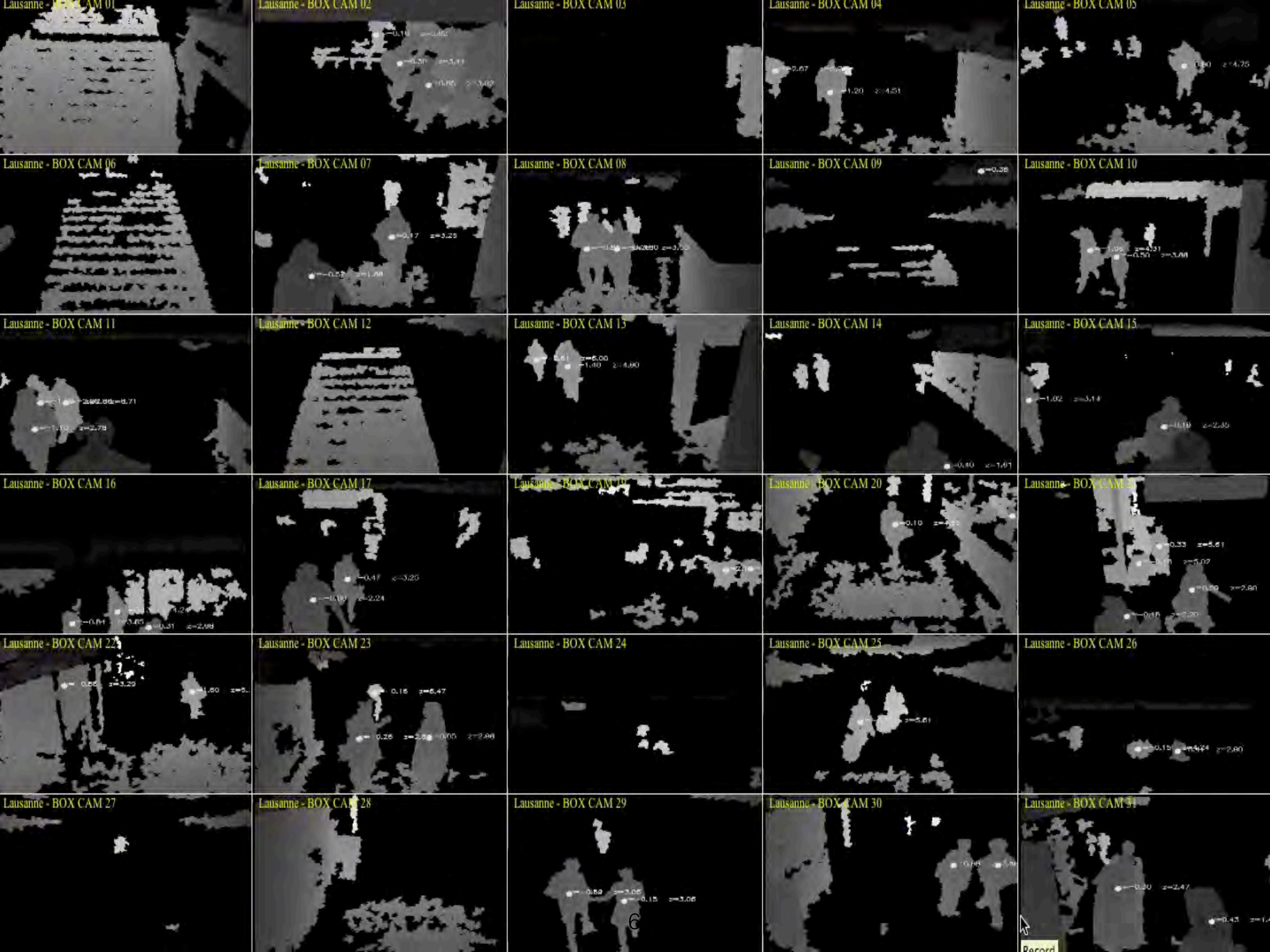
>42 million humans

(Lausanne, Basel)

A corridor with 32 sensors

Top View

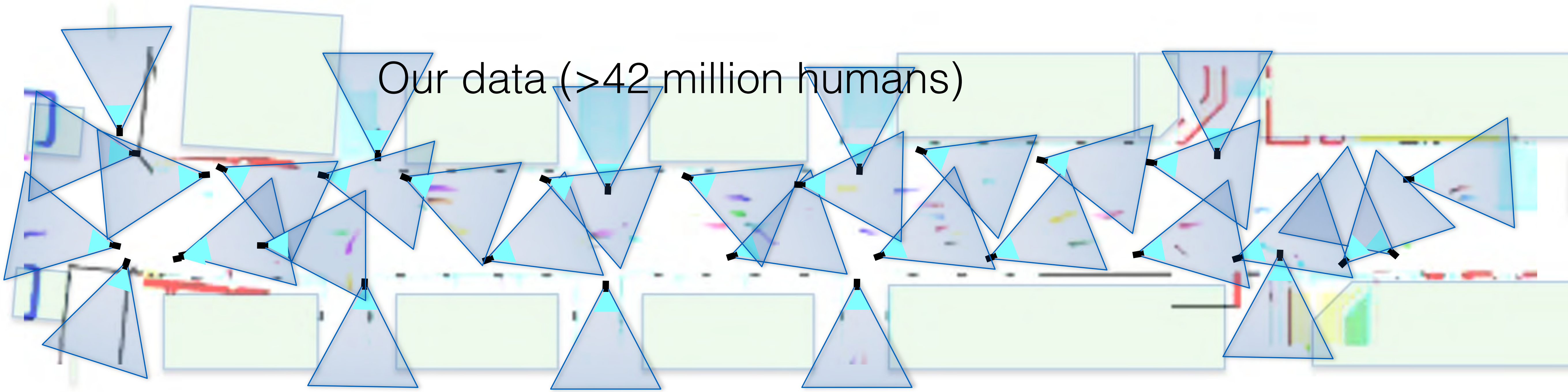



Lausanne - BOX CAM 02
Lausanne - BOX CAM 03
Lausanne - BOX CAM 04
Lausanne - BOX CAM 05
Lausanne - BOX CAM 06
Lausanne - BOX CAM 07
Lausanne - BOX CAM 08
Lausanne - BOX CAM 09
Lausanne - BOX CAM 10
Lausanne - BOX CAM 11
Lausanne - BOX CAM 12
Lausanne - BOX CAM 13
Lausanne - BOX CAM 14
Lausanne - BOX CAM 15
Lausanne - BOX CAM 16
Lausanne - BOX CAM 17
Lausanne - BOX CAM 20
Lausanne - BOX CAM 22
Lausanne - BOX CAM 23
Lausanne - BOX CAM 24
Lausanne - BOX CAM 25
Lausanne - BOX CAM 26
Lausanne - BOX CAM 27
Lausanne - BOX CAM 28
Lausanne - BOX CAM 29
Lausanne - BOX CAM 30
Lausanne - BOX CAM 31

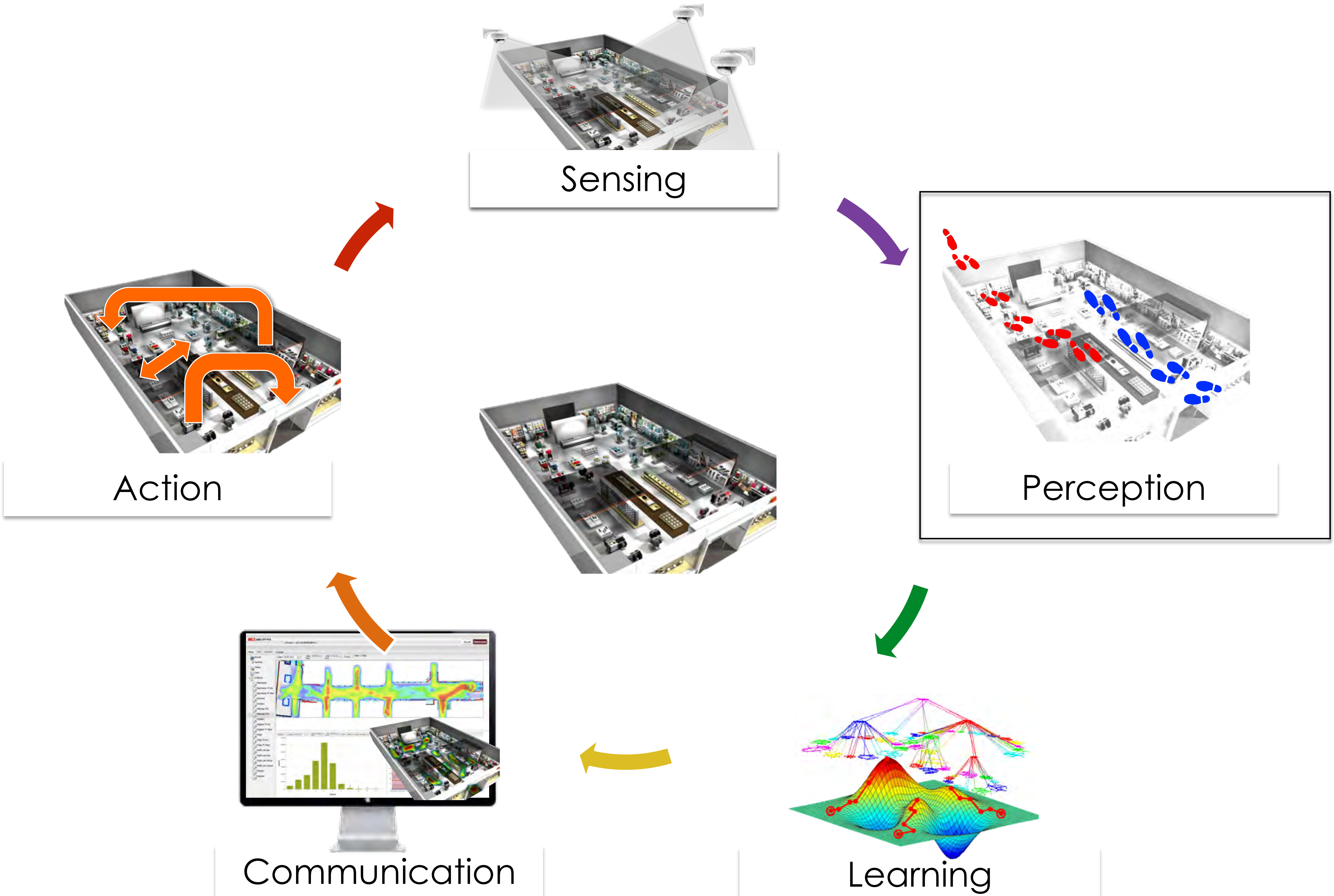

Basel - IP CAM 01
Basel - IP CAM 02
Basel - IP CAM 03
Basel - IP CAM 04
Basel - IP CAM 05
Basel - IP CAM 06
Basel - IP CAM 07
Basel - IP CAM 08
Basel - IP CAM 09
Basel - IP CAM 10
Basel - IP CAM 11
Basel - IP CAM 12
Basel - IP CAM 13
Basel - IP CAM 14
Basel - IP CAM 15

Collecting long-term trajectories

Our data (>42 million humans)

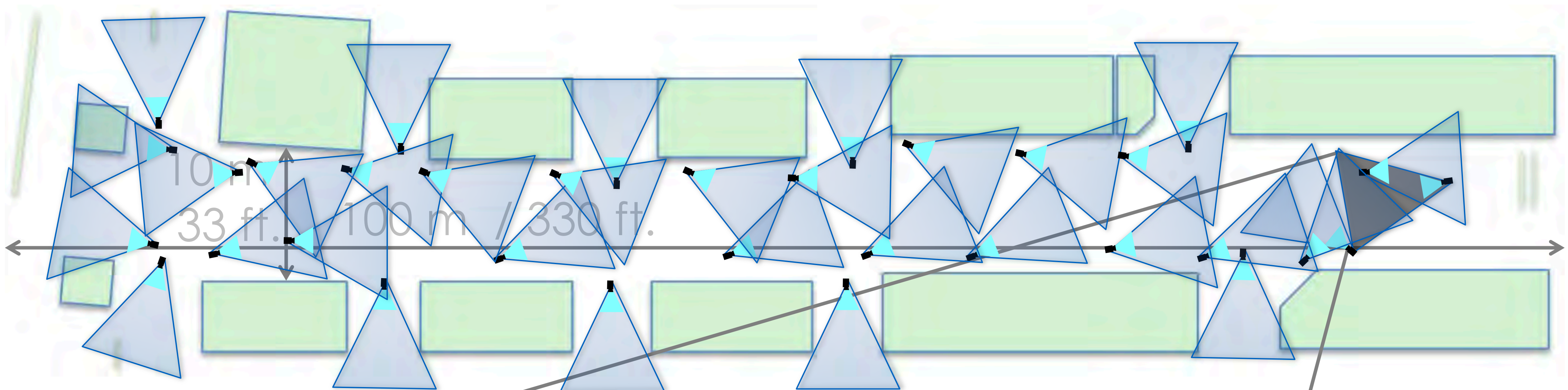


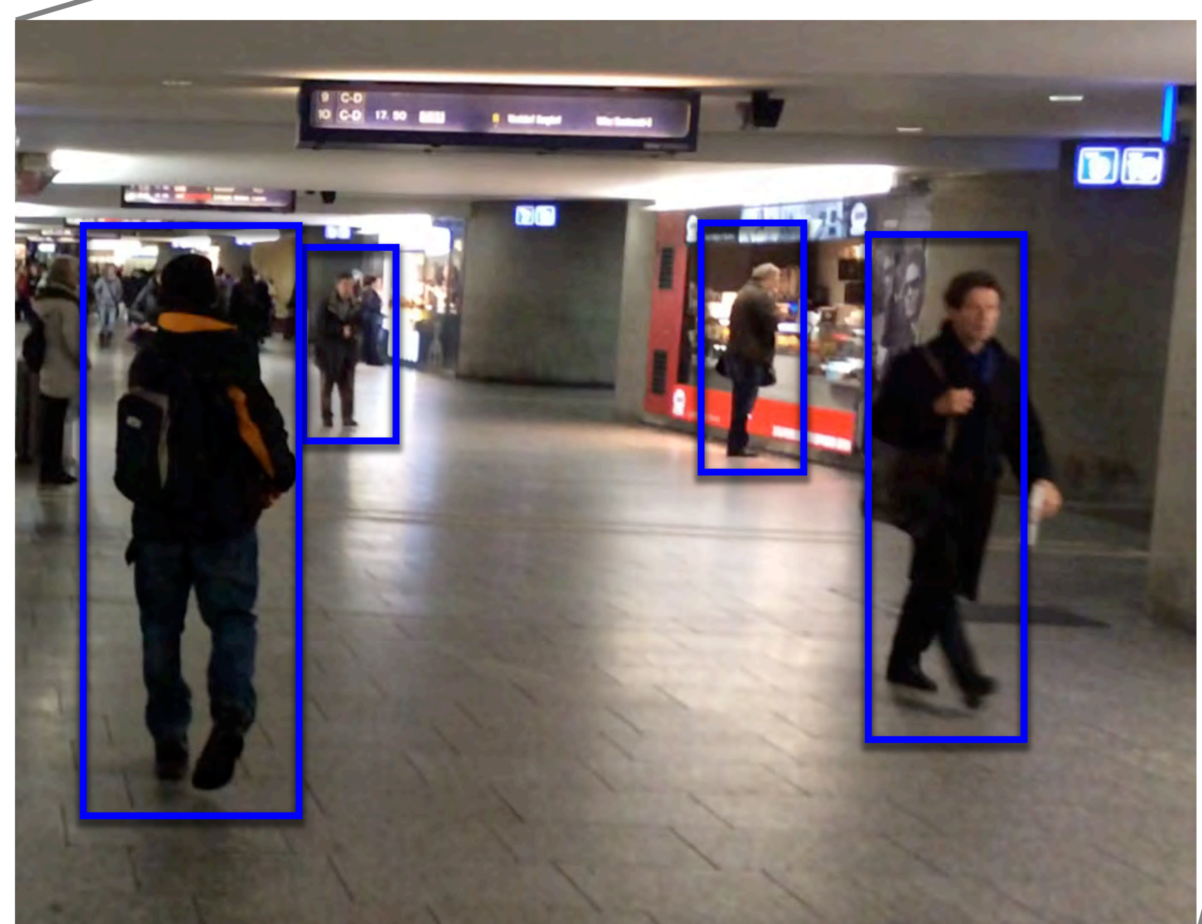
AI for the built environments



A corridor with 32 sensors

Top View

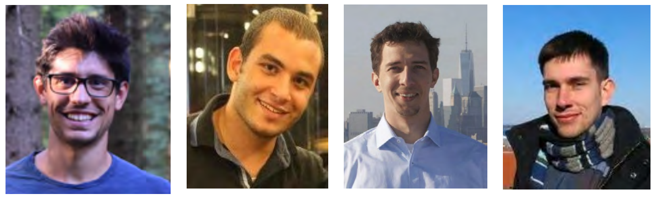


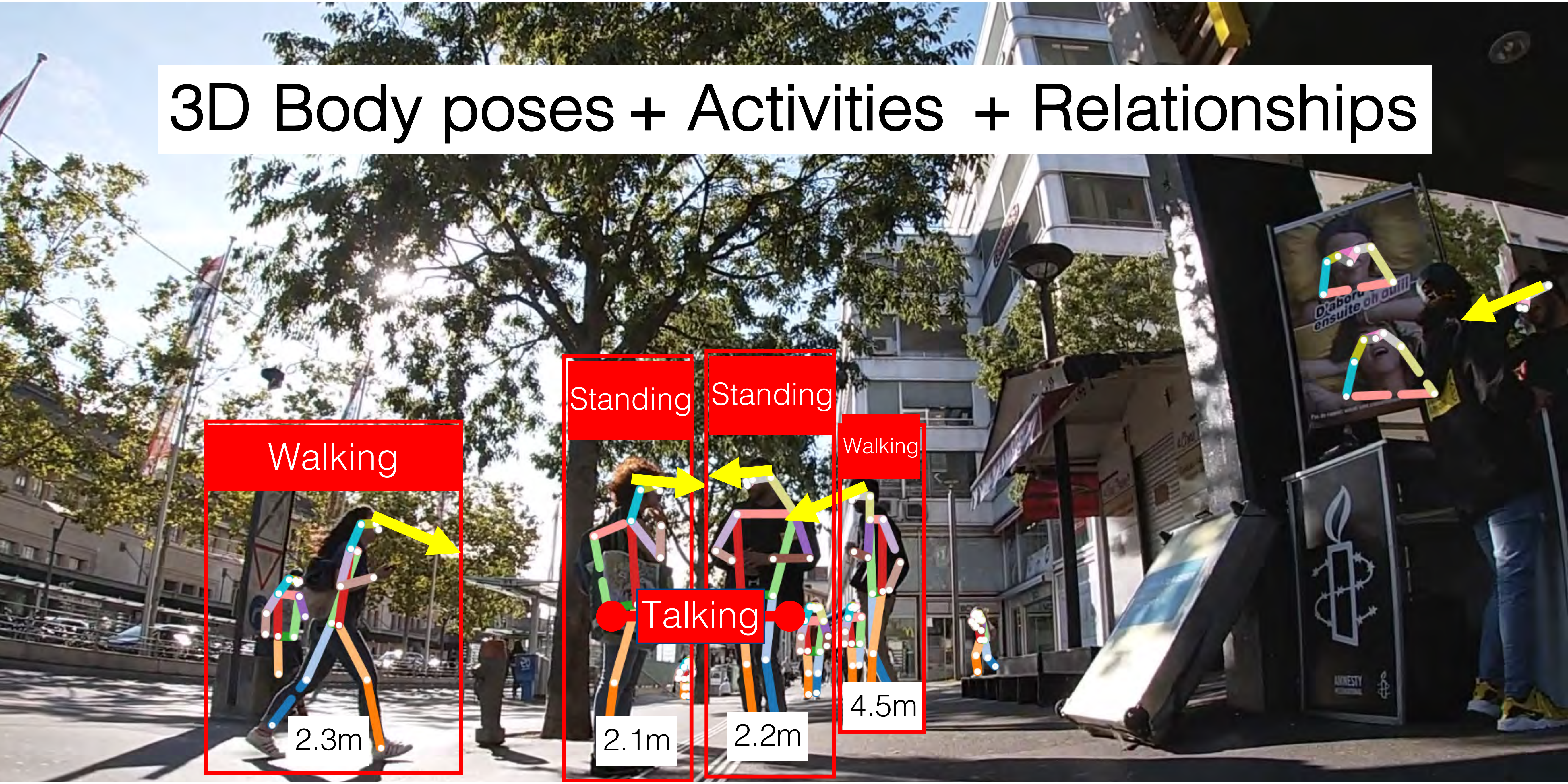


Perceiving

Previous works
[1] Yolo v1…Yolo v4

Perceiving
Socially-aware cues



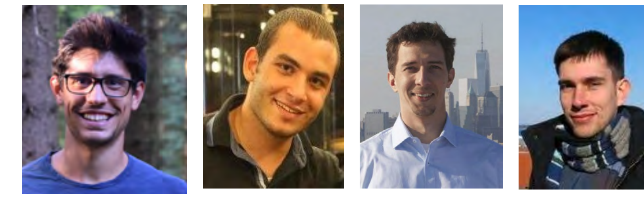


Our work

[1] PifPaf: Composite Fields for Human Pose Estimation, **CVPR'19** (Live Demo: https://vitademo.epfl.ch)

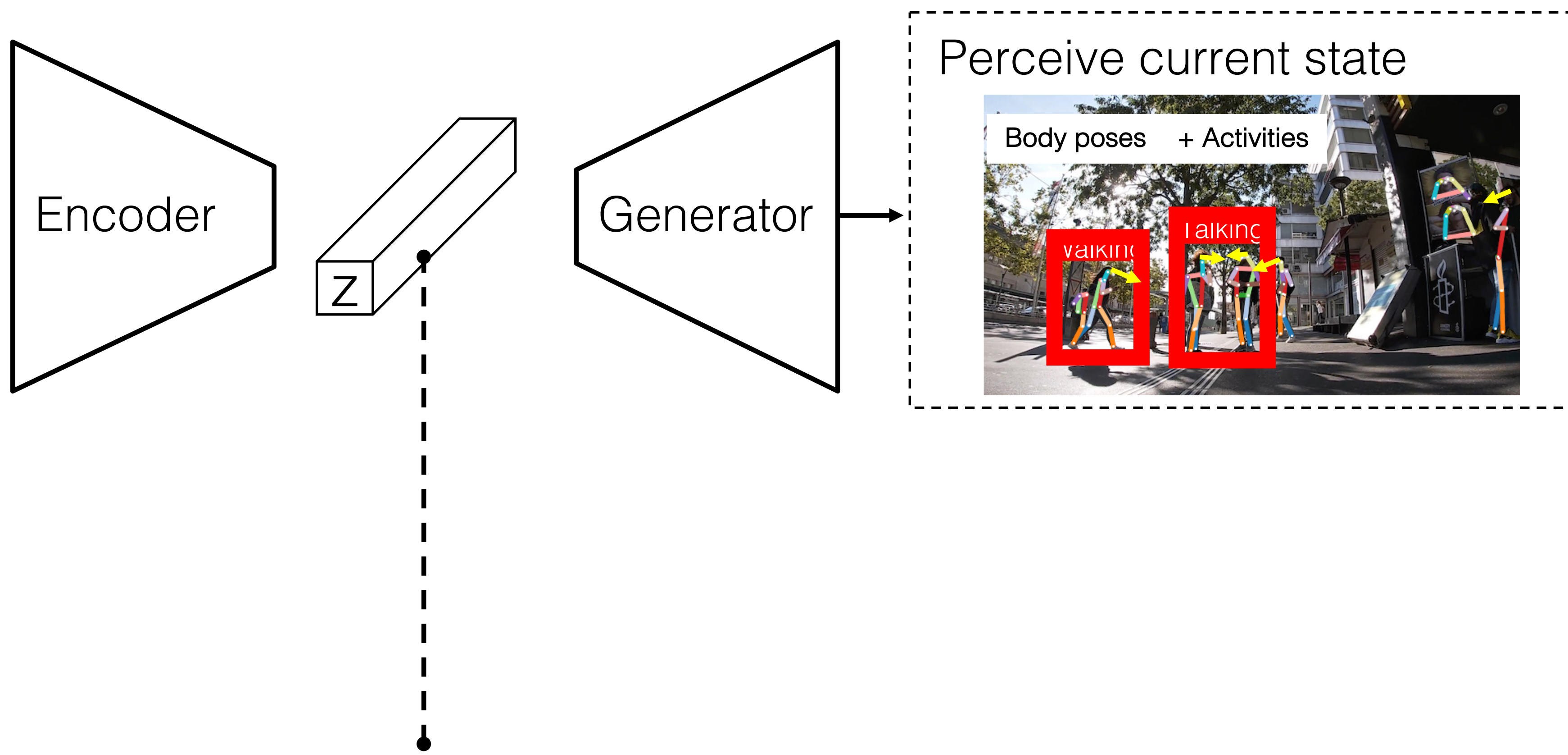
[2] Convolutional Relational Machine for Group Activity Recognition, **CVPR'19**

[3] Monoloco: Monocular 3D pedestrian localization and uncertainty estimation, **ICCV'19**



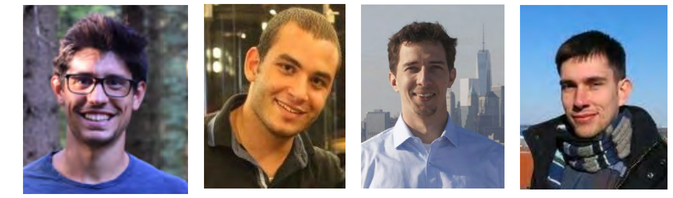
Computer vision for the built environments

3 Challenges :
1- Limited resolution with partial information
2- Efficiency (Real-time)
3- Many tasks with unbalanced labels



Shared representation for

Perception



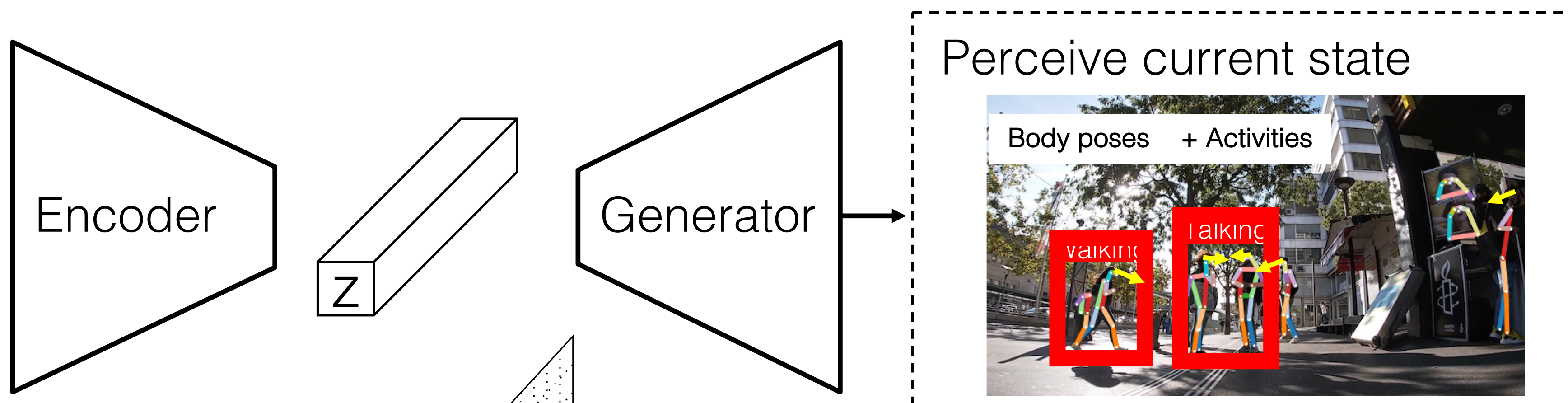
Computer vision for the built environments

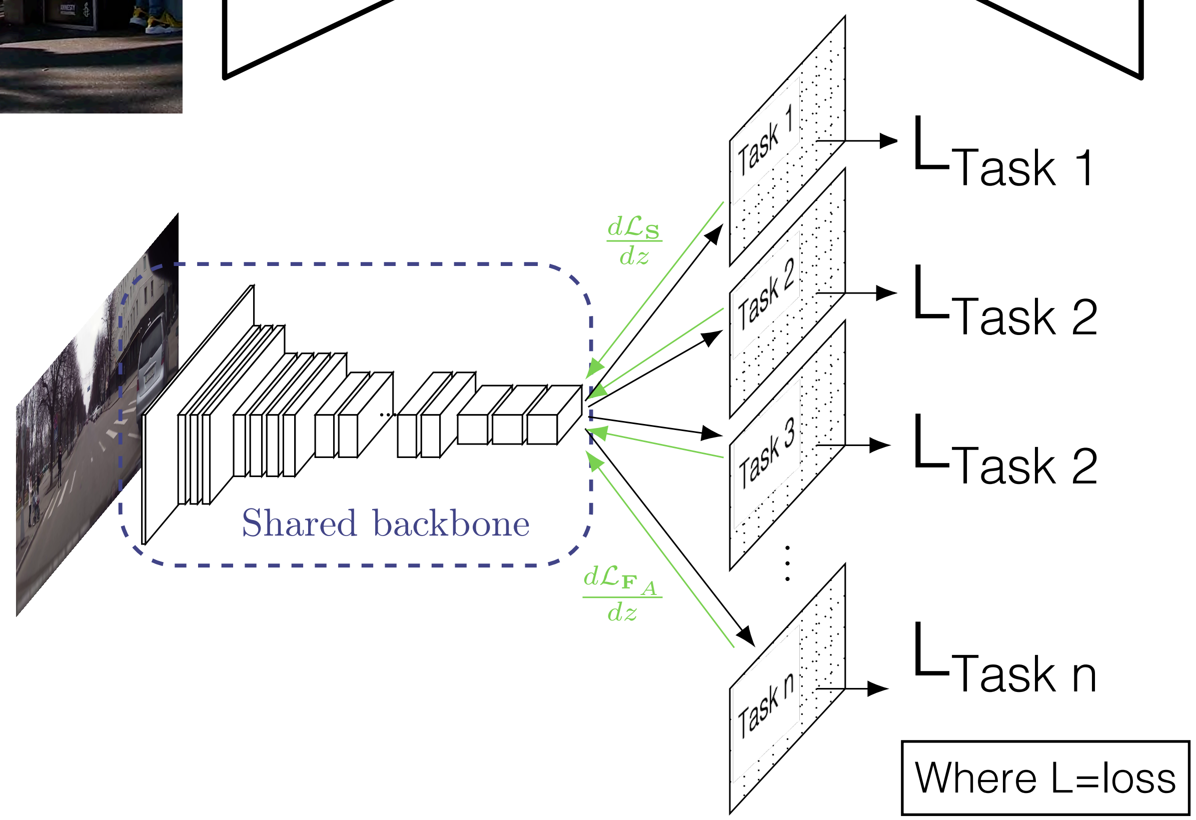
3 Challenges :

1- Limited resolution with partial information

2- Efficiency (Real-time)

3- Many tasks with unbalanced labels

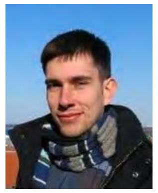


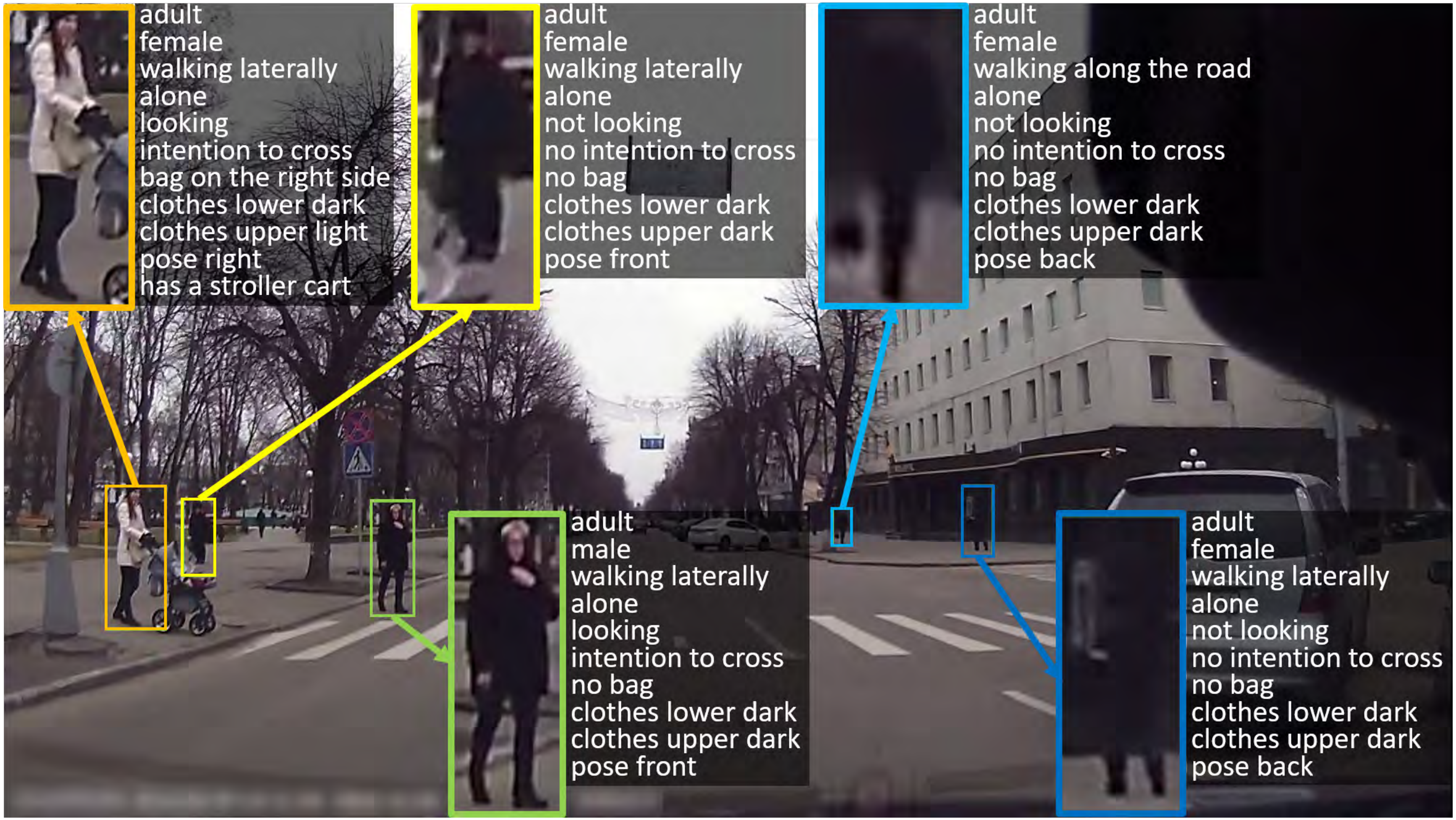


Composite fields formalism

A Feature map can jointly encode scalars and vectors

Jointly Perceiving 32 Attributes on Pedestrian's Appearance, Behaviour, Intention

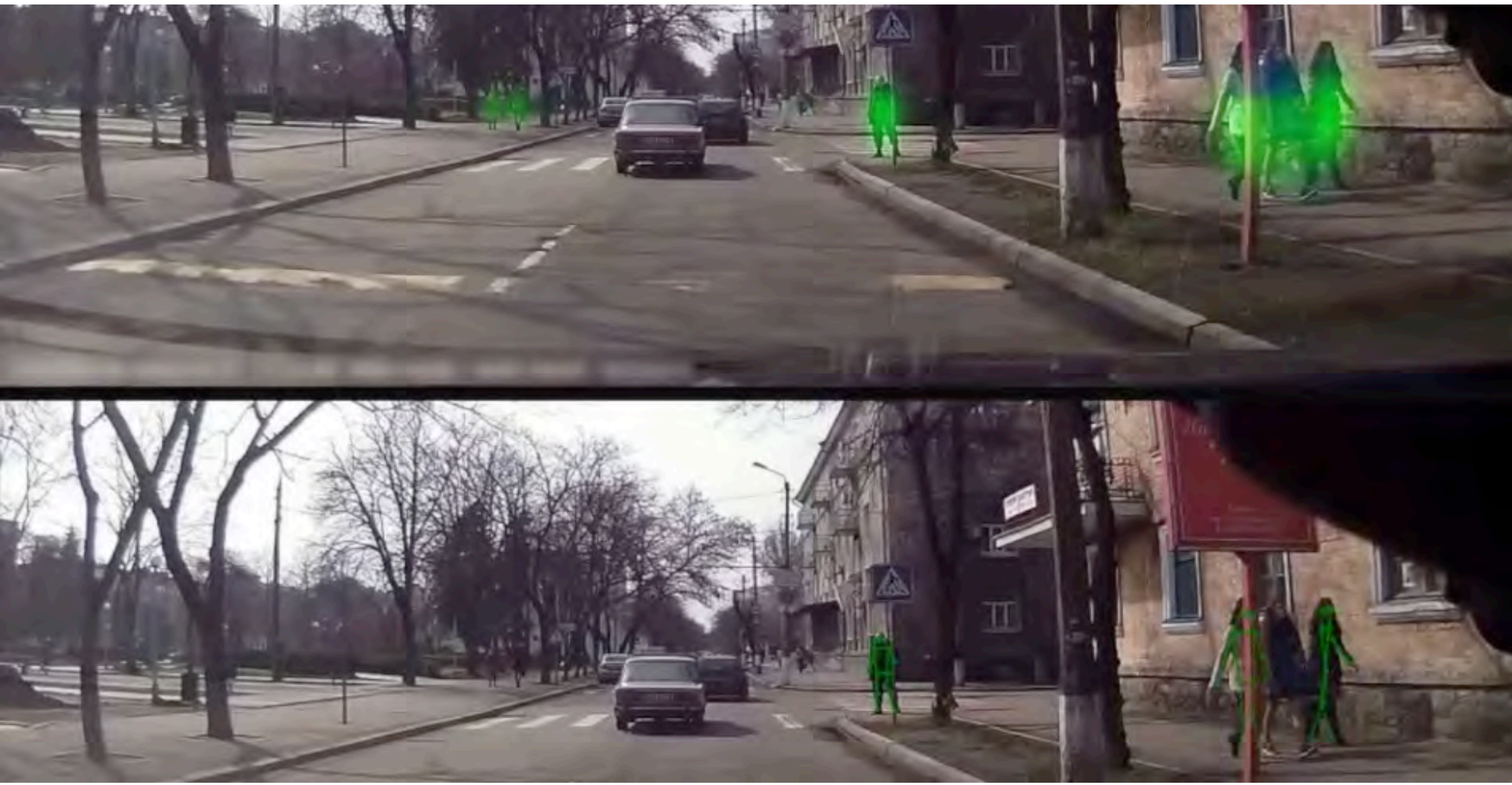




[1] Detecting 32 Pedestrian Attributes for Autonomous Vehicles, arxiv

Perceiving
Intentions

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

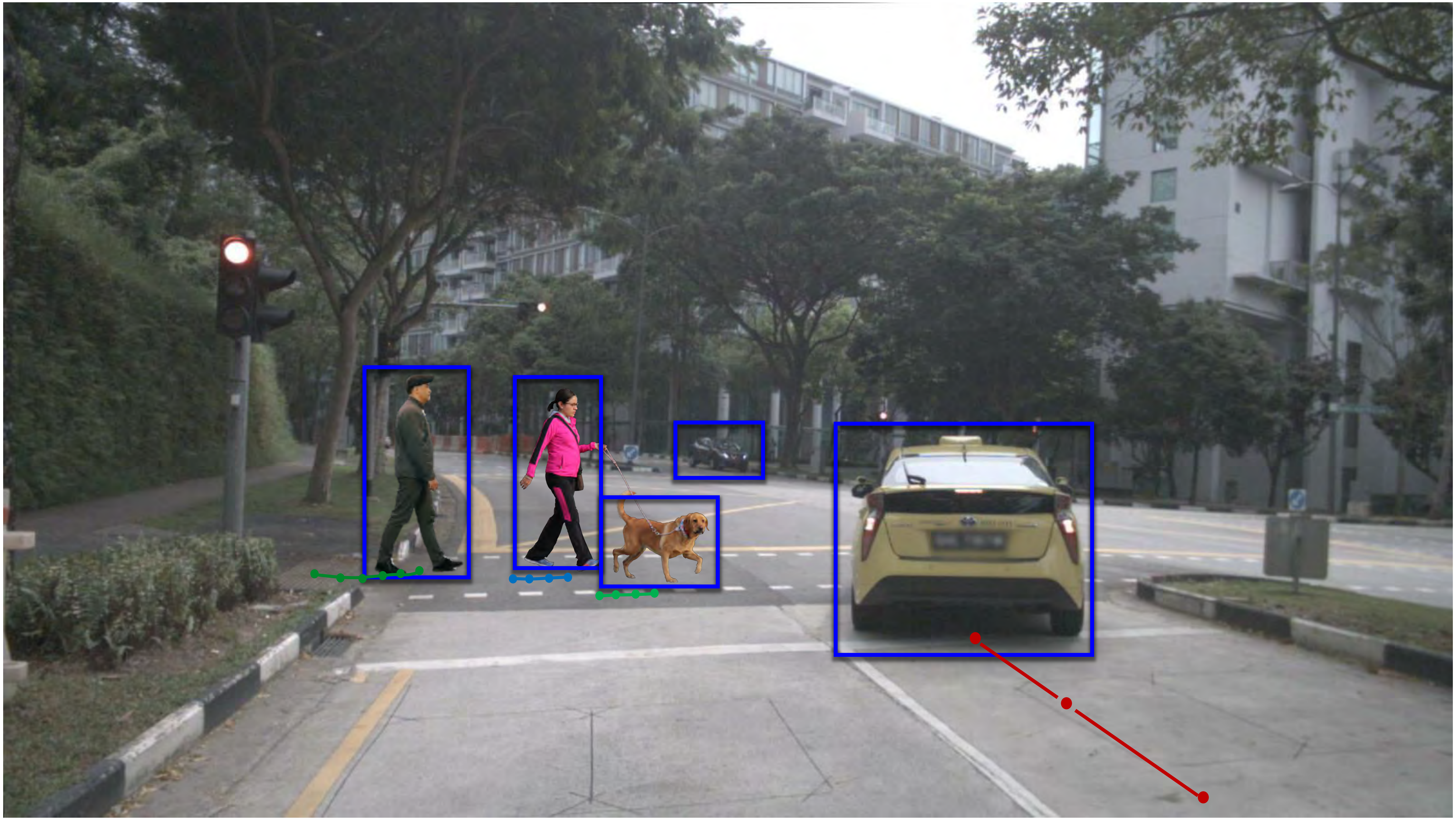


[1] Pedestrian Intention Prediction: A Convolutional Bottom-Up Approach, arxiv

Object

Detection & Tracking

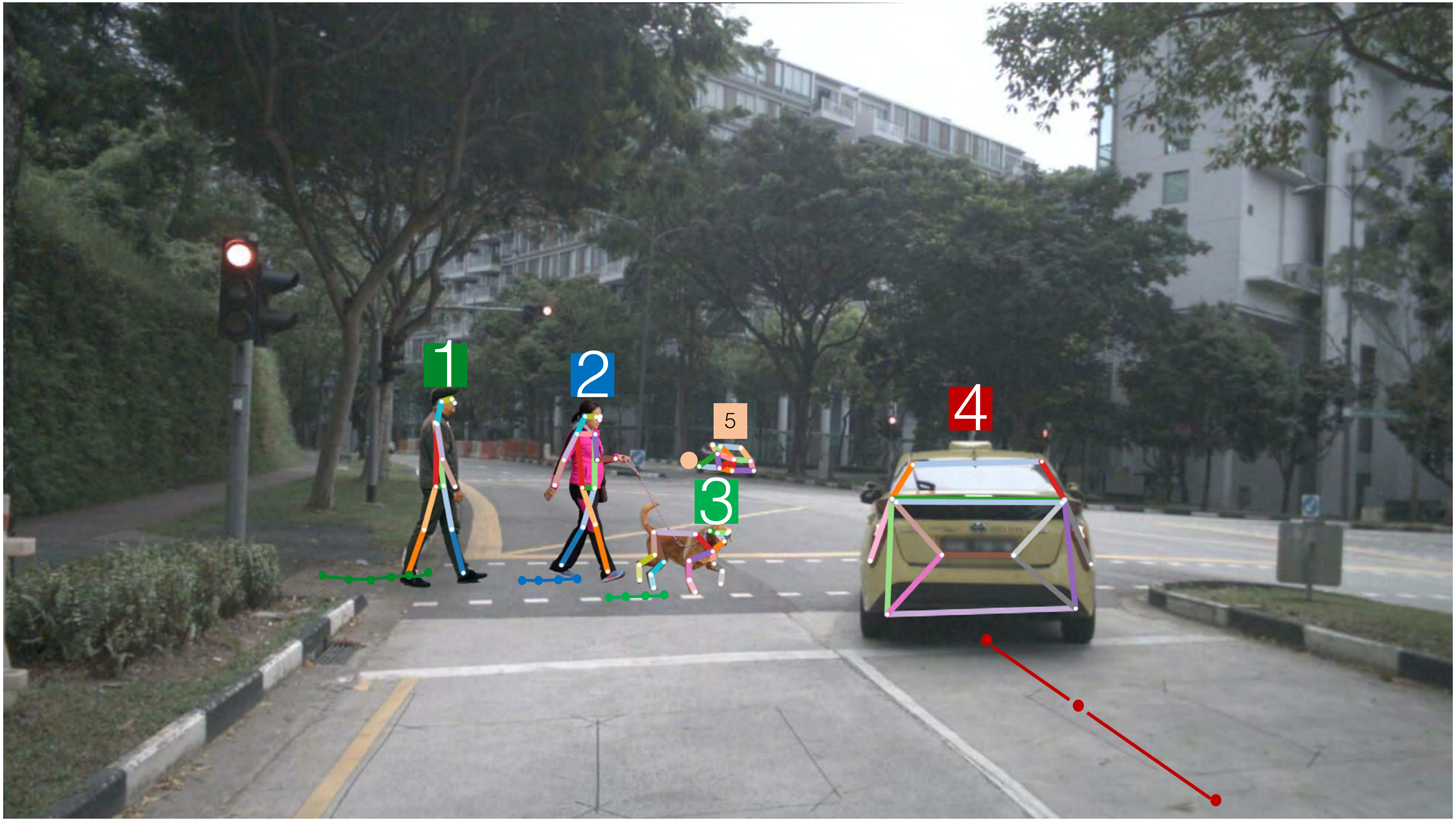
2 steps



Previous works
[1] Yolo v1…Yolo v5

Semantic keypoints

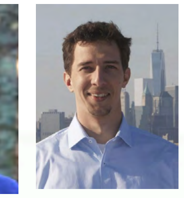
Joint Detection & Tracking

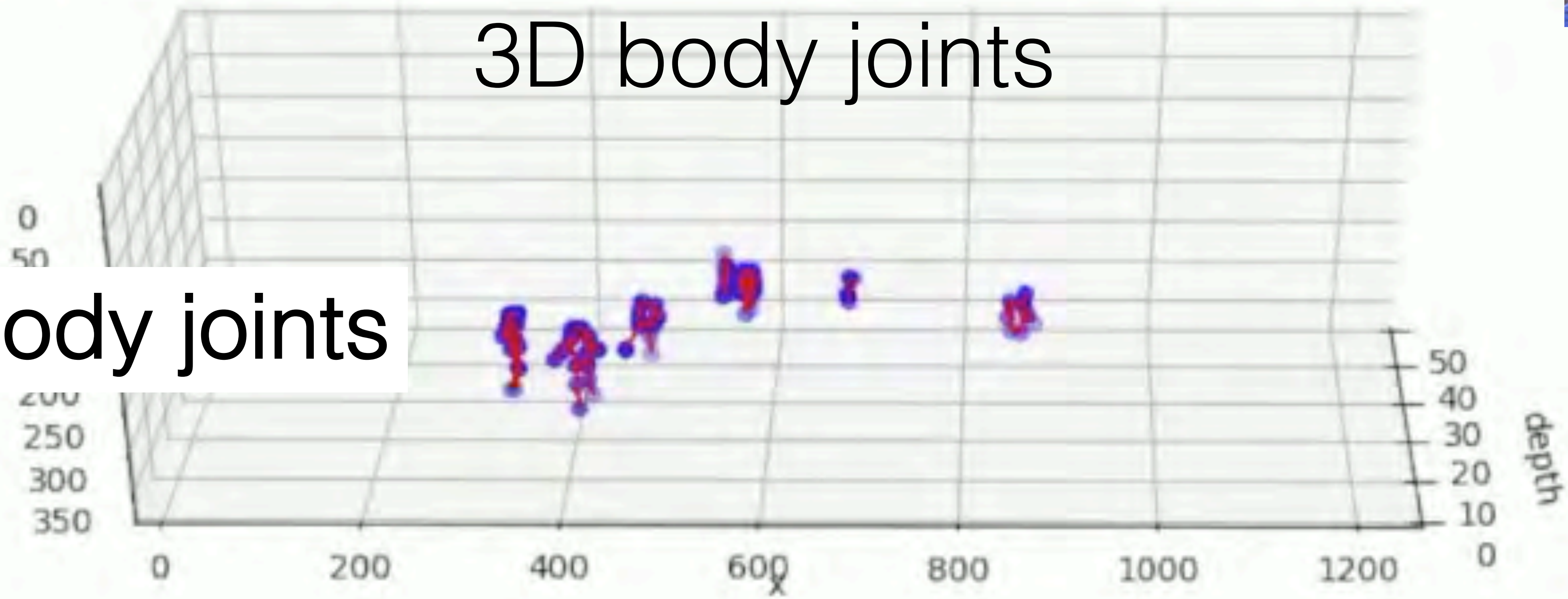


[1] PifPaf: Composite Fields for Human Pose Estimation, **CVPR'19**

(Live Demo: https://vitademo.epfl.ch)

Perceiving 3D body joints



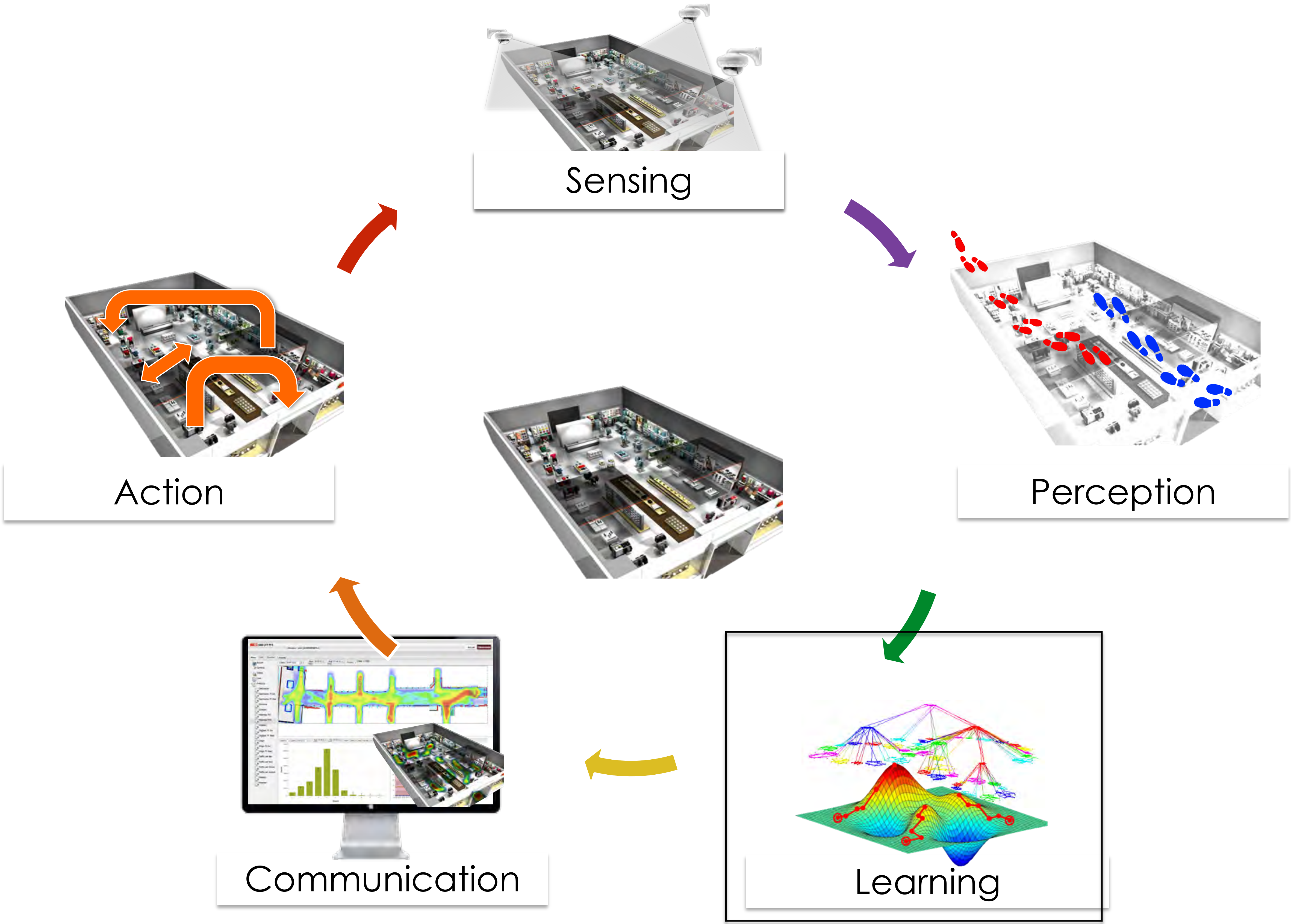


3D Body joints

[1] L. Bertoni *et al.*, MonoLoco, ICCV'19

[2] W. Deng et al., Joint Human Pose Estimation and Stereo Localization, ICRA'20

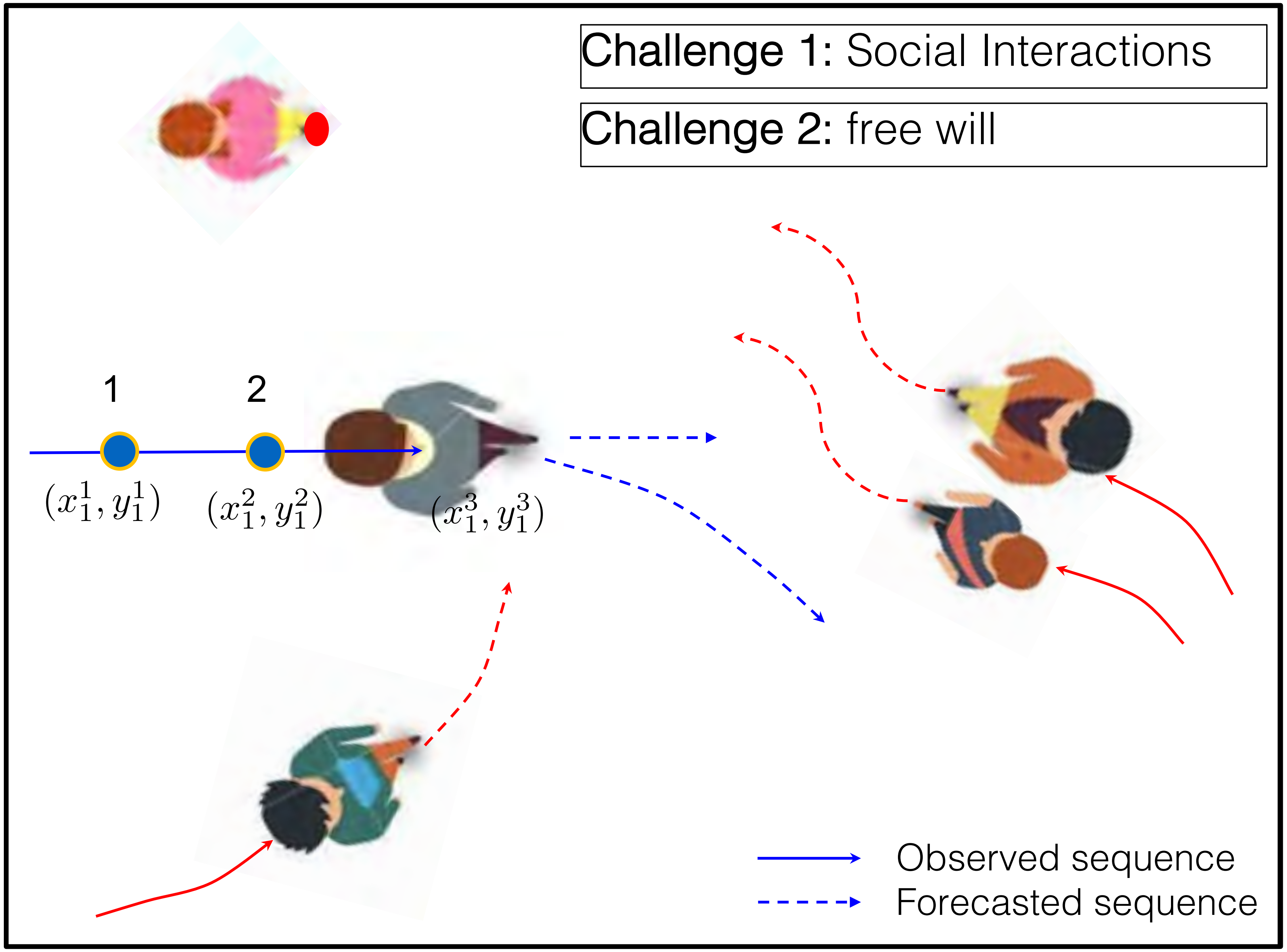
AI for the built environments



Social Forecasting

Input: Given a sequence of states, *e.g.*, (x^t,y^t) coordinates in time

Output: **Predict** the future states, *e.g.,* next 5 seconds



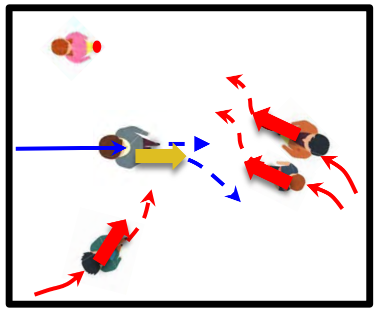
Social Forecasting

- Previous works

Knowledge-driven

- Social Forces Model [1],

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



- Discrete Choice Model [2]

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

✓ Interpretability
X Predictability

Previous works
[1] Helbing *et al.*, Physical review, '95
[2] Antonini *et al.*, Transportation Research, '06

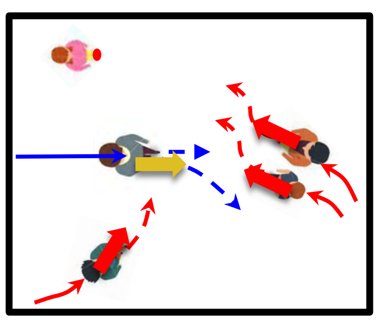
Social Forecasting

- Previous works

Knowledge-driven

- Social Forces Model [1],

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



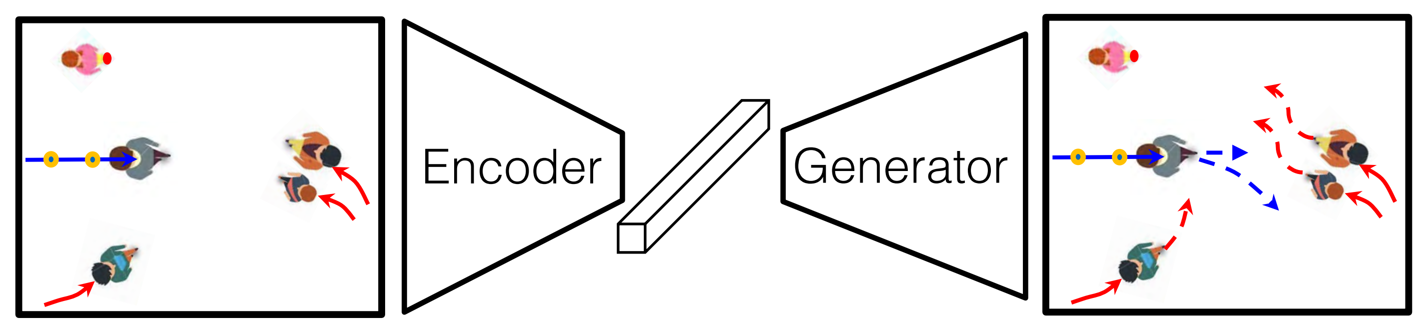
- Discrete Choice Model [2]

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

✓ Interpretability
X Predictability

- Our works

Data-driven methods



X Interpretability
✓ Predictability

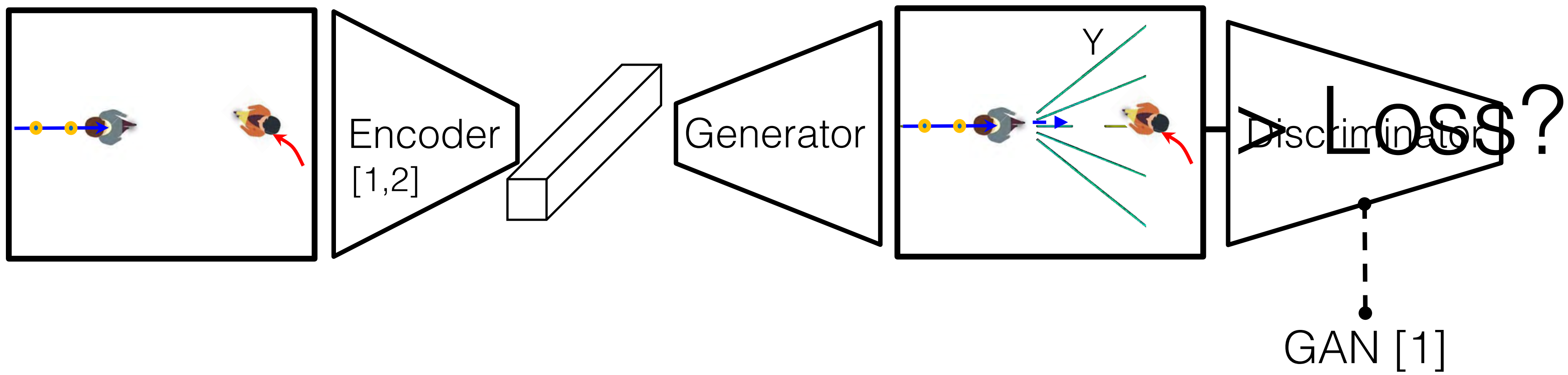
Previous works
[1] Helbing *et al.*, Physical review, ’95
[2] Antonini *et al.*, Transportation Research, ’06

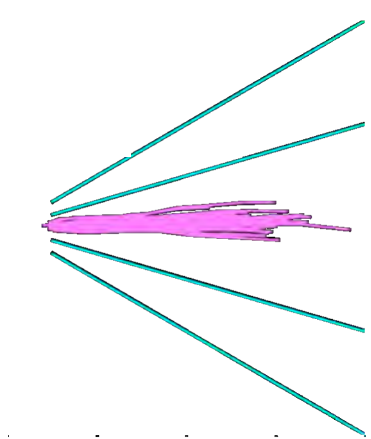


Social Forecasting

Learning Representation of Crowds

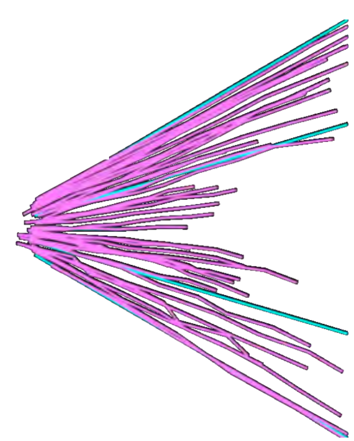
Challenge 2: Free will





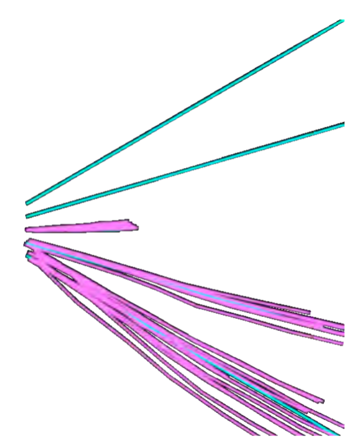
Average L2
Loss

$\ell_{Recon} = \frac{1}{K}\sum_k \|Y_i - \hat{Y}_i^{(k)}\|$

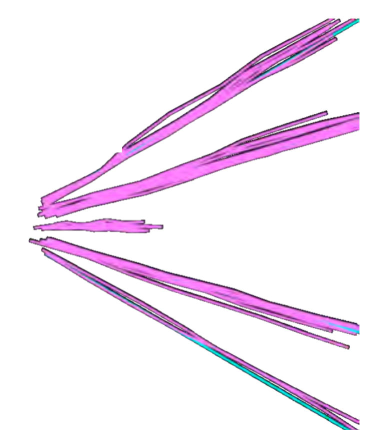


Winner-takes-all
Loss

$\mathcal{L}_{variety} = \min_k \|Y_i - \hat{Y}_i^{(k)}\|$



Adversarial
Loss [1]



Adversarial
Loss w/
Collab. Sampling [2]

Our work

[1] Social GAN (Generative Adversarial Network), CVPR'18

[2] Collaborative Sampling in Generative Adversarial Networks, AAAI'20



Performance evaluation

	Methods*	ADE/FDE
Knowledge-driven	Kalman Filter	0.87/1.69
	DCM '06	0.67/1.42
	Social Force '98	0.89/1.53
	ORCA '08	0.68/1.40
Our data-driven	LSTM '14	0.61/1.31
	S-GAN '18	0.57/1.24
	S-LSTM '16	0.54/1.17
	D-LSTM '20	0.57/1.24

0.68/1.40 0

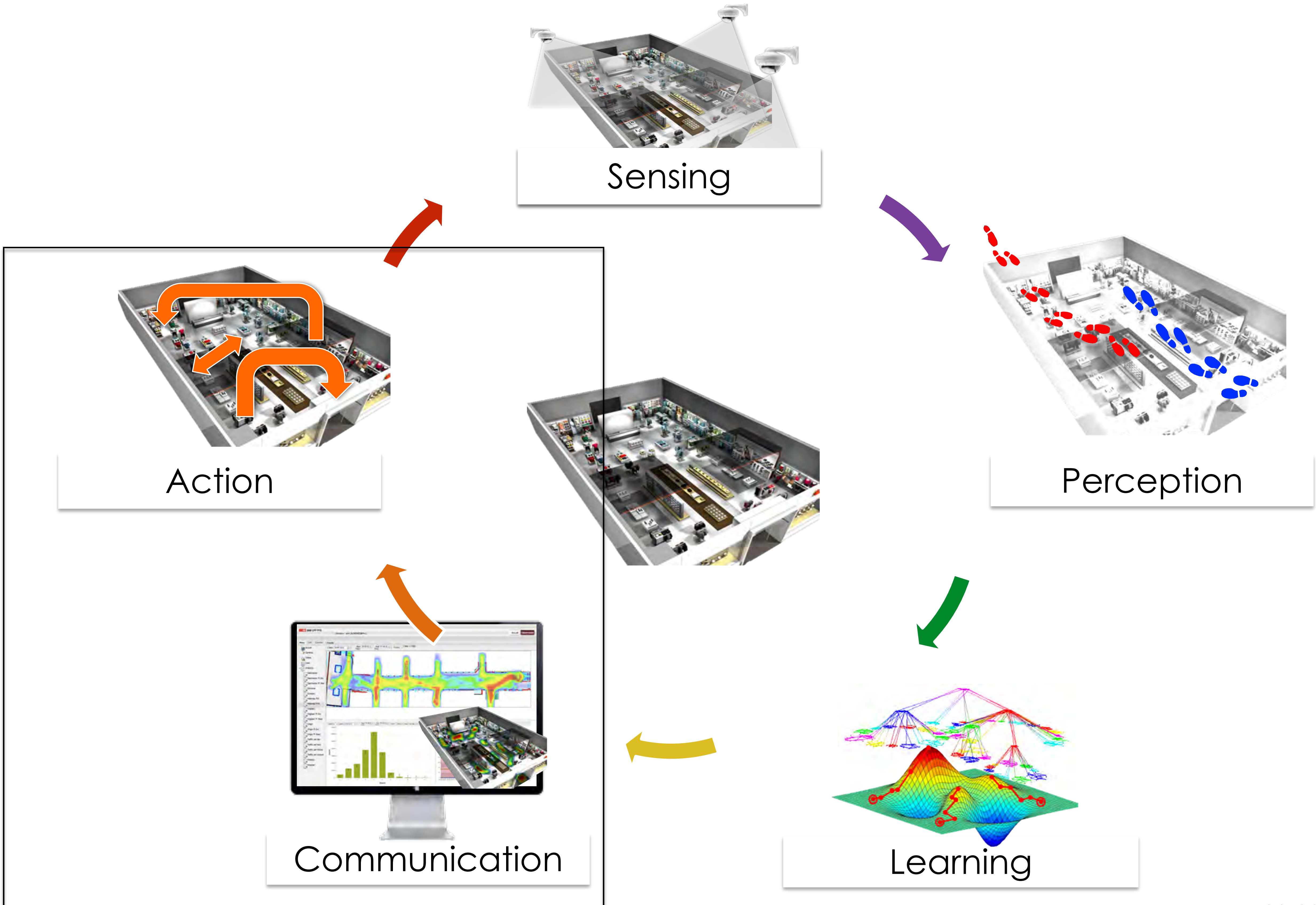
0.54/1.17 7

0.55/1.18 7.6

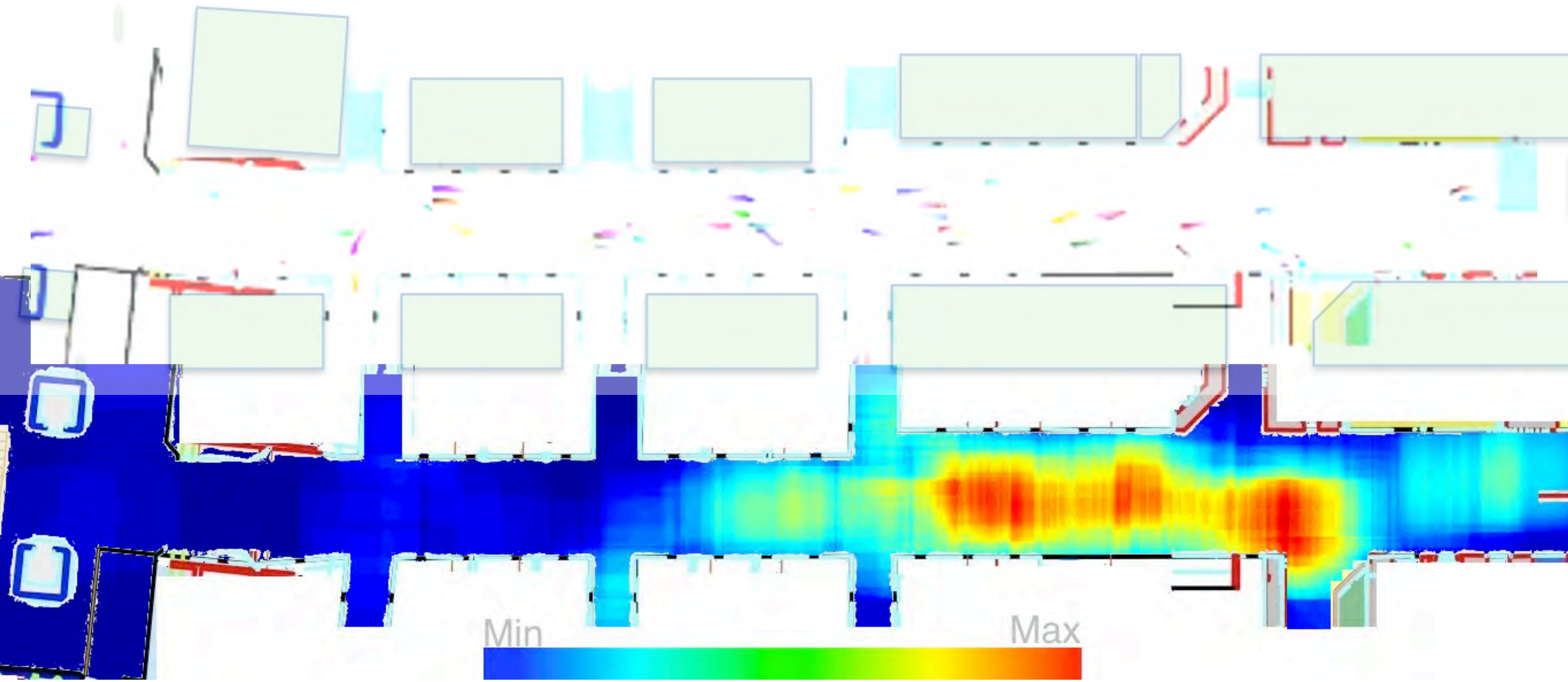
ADE: Average displacement error in m
FDE: Final displacement error in m
All references available in [1]

Our work
[1] Human trajectory forecasting: a deep learning perspective, arxiv

AI for the built environments



What can we learn from all these trajectories?



# People	Av. duration	Av distance	Density (up to)	# Paths (O/D)
42 million	1 minute	100m	1 pedestrian/m^2	196

Bertoni, L., Kreiss, S., Alahi,A,
Perceiving humans: from monocular 3D localization to social distancing,
arxiv

Thank you

#Open Science

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

Perception:
[1] S. Kreiss et al., OpenPifPaf library for pose estimation, CVPR'19 (licensed)
[2] L. Bertoni et al., Monocular 3D Pedestrian Localization and Uncertainty Estimation, ICCV'19
[3] L. Bertoni et al., MonStereo, Stereo 3D detection
[4] L. Bertoni et al., Perceiving Social Distancing, ITS'20
[5] G. Adaimi et al., Perceiving Traffic from Aerial Images
[6] G. Adaimi et al., Deep Visual Re-identification with Confidence

Prediction:
[7] Kothari et al., Trajnet++ library for spatio-temporal forecasting tasks (>15 implemented models)

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

Generative models:
[9] Y. Liu* et al., Collaborative Sampling in GAN, AAAI'20
[10] A. Carlier et al., Deep SVG, NeurIPS'20

DCM + NN
[11] B. Sifringer et al., L-MNL, TRB'20

Tools
[12] Video Ultimate labeling