WASP | WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

ÖREBRO UNIVERSITY

Learning to Understand and Predict Heterogeneous Trajectory Data

Tiago Rodrigues de Almeida[1]
Andrey Rudenko[2]
Johannes A. Stork[1]
Achim J. Lilienthal[1,2]
1. Örebro University; 2. Technical University of Munich

2025-09-26

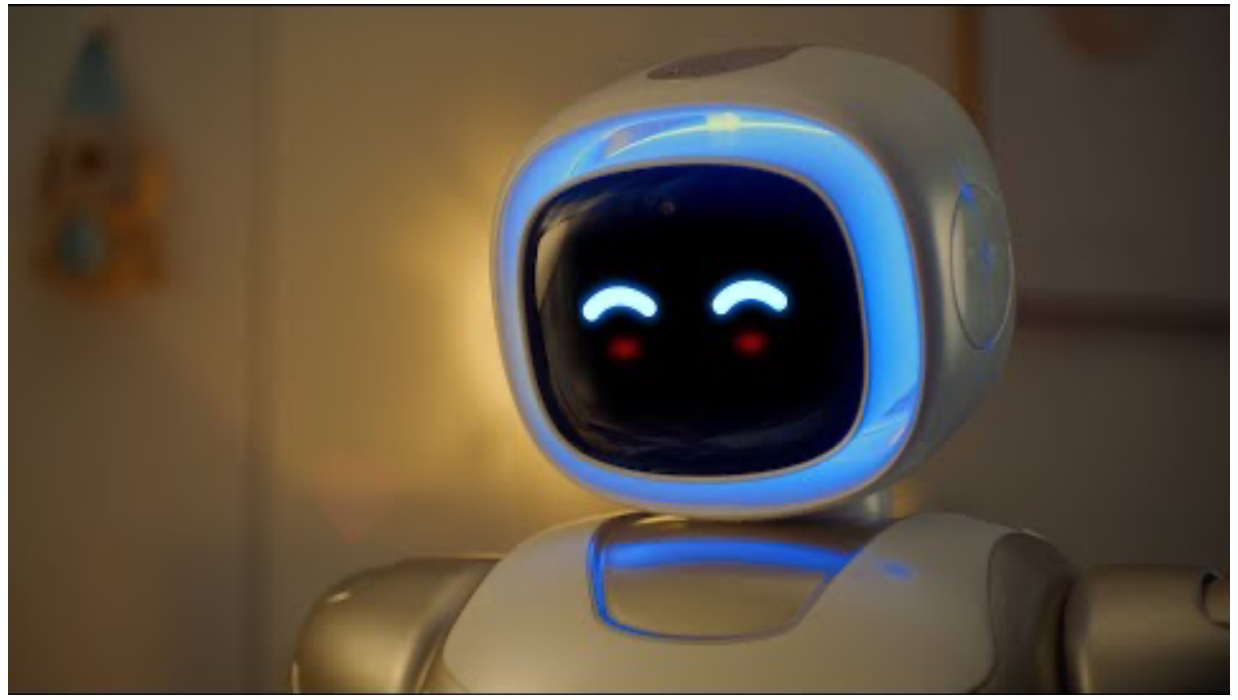
Outline

1. **Abstract.** Why is trajectory prediction important?
2. **Introduction.** What is this talk about?
3. **Gaps, Our Approaches & Results.** What are current methods lacking? How could this thesis bridge the gaps?
4. **Contributions Overview.** What could this thesis accomplish?
5. **Future Work.** Where should we go from here?

1. Abstract.

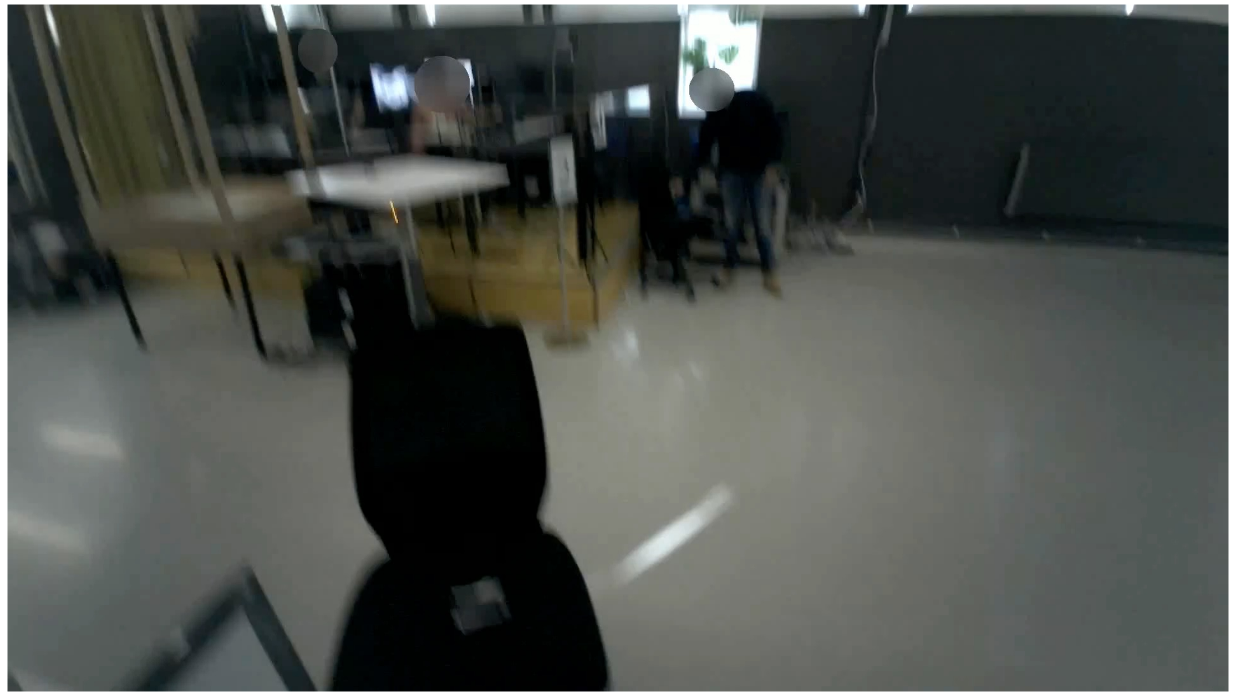
Why is trajectory prediction important?

Autonomous mobile robots must **anticipate** other agents' **intentions** in order to move efficiently and effectively (collision avoidance) in shared environments.



Courtesy of Ubtech.

Intentions can be encoded as **trajectories**, which must be predicted.



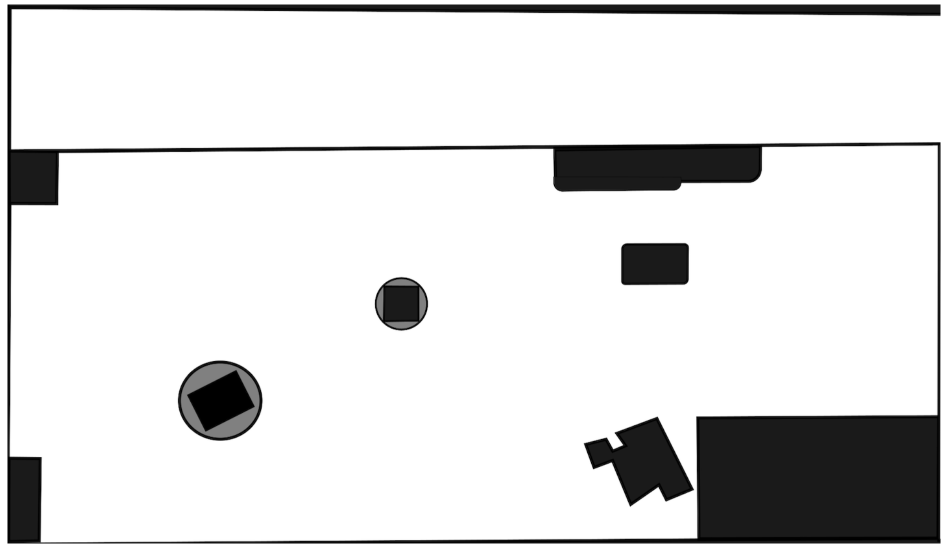
2. Introduction.

What is this talk about?

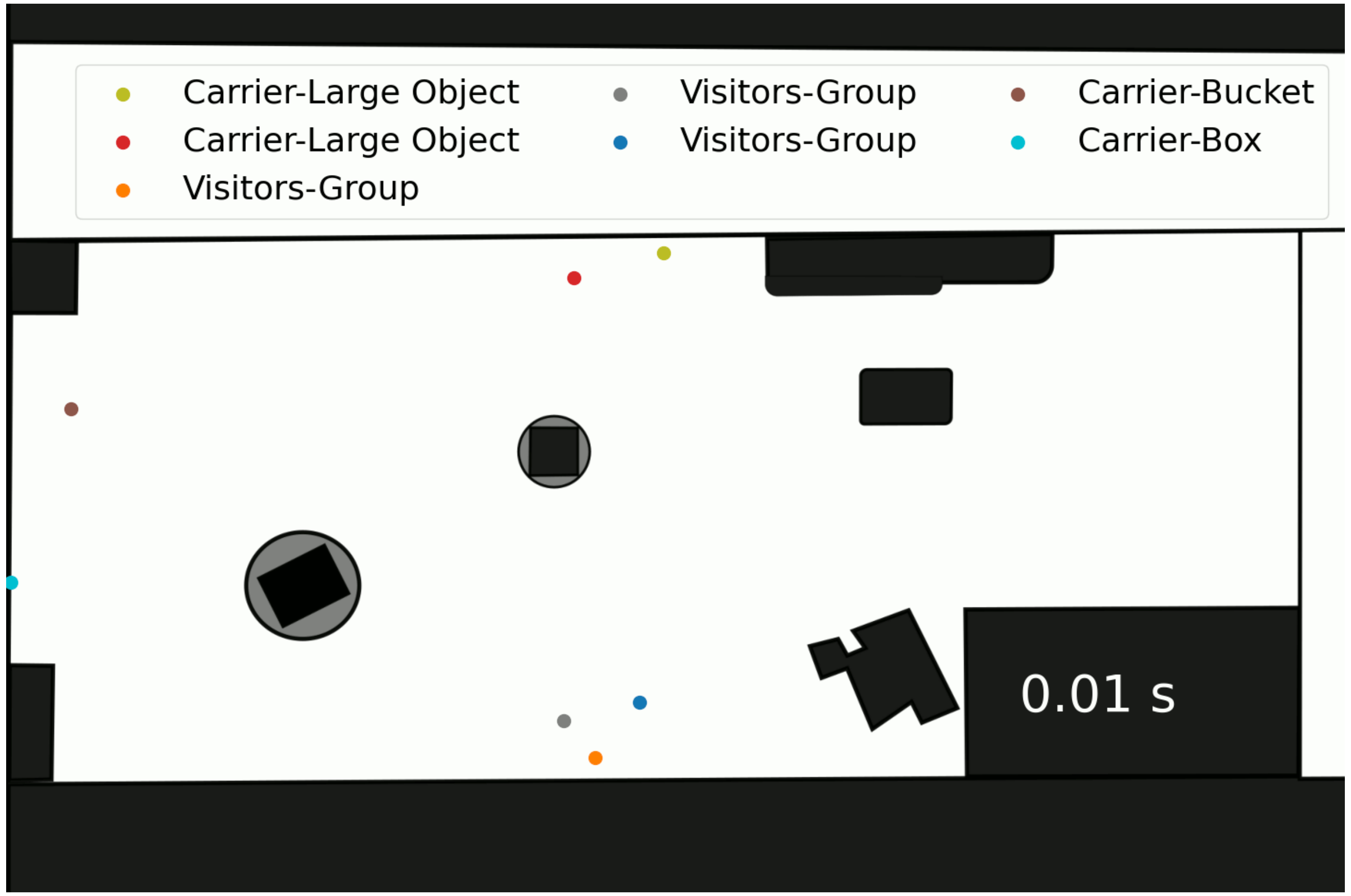
What is the usual form of trajectory data?

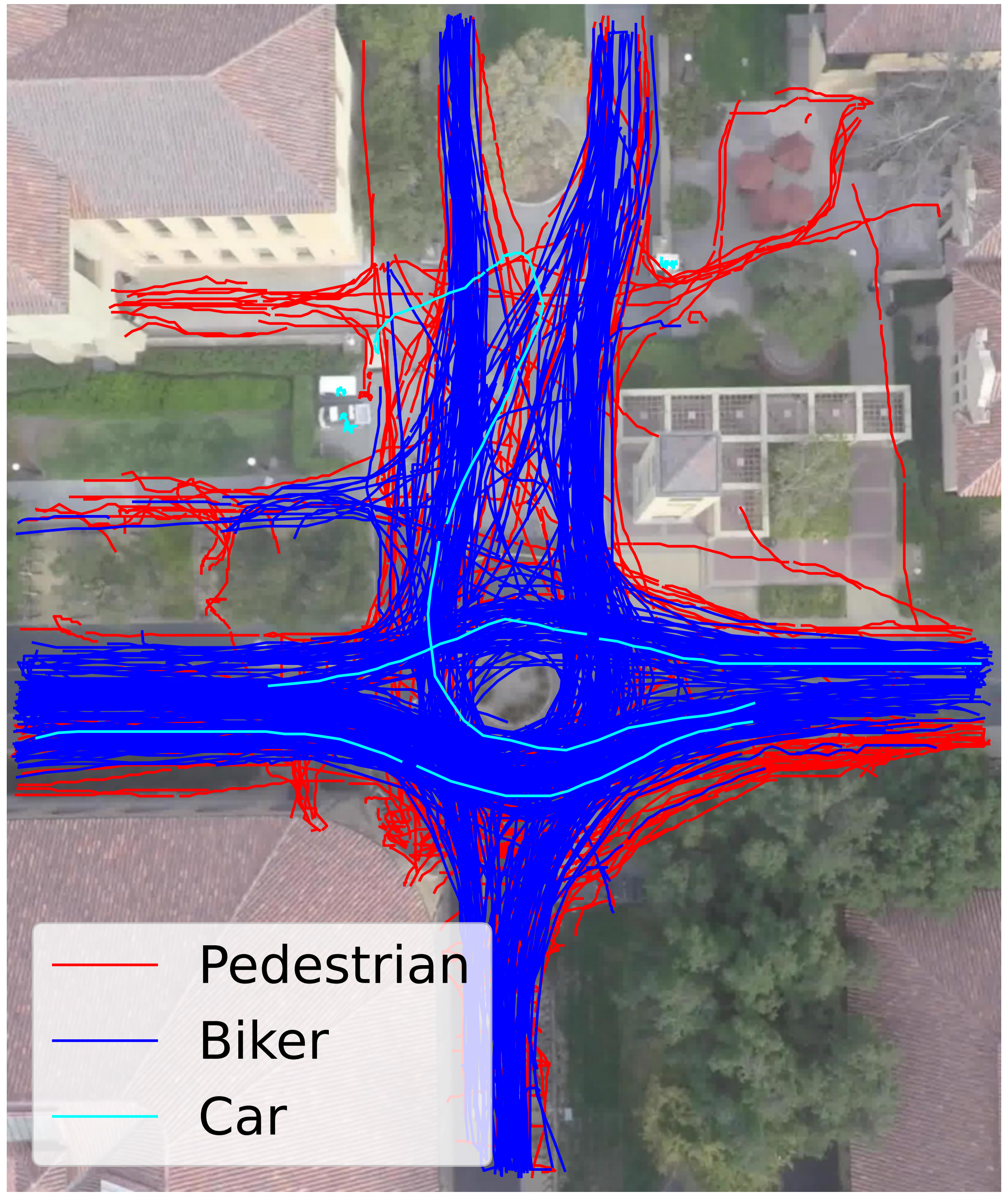
Moving agents and their respective navigational styles, roles, preferences, etc..

A **scenario / environment**, its affordances and contextual information.



What is the usual form of trajectory data?

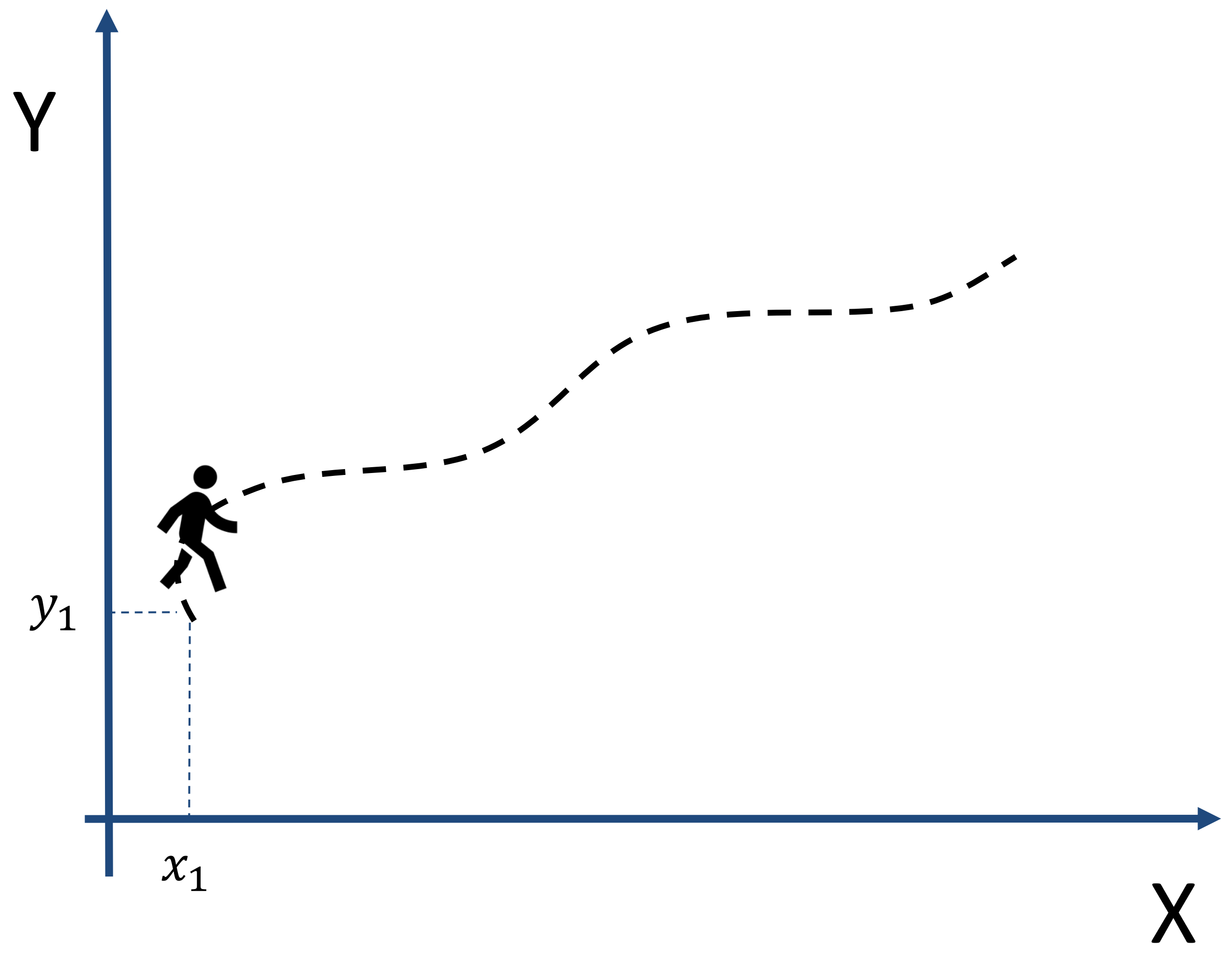




Temporal sequences of 2D positions with respect to a global reference frame.

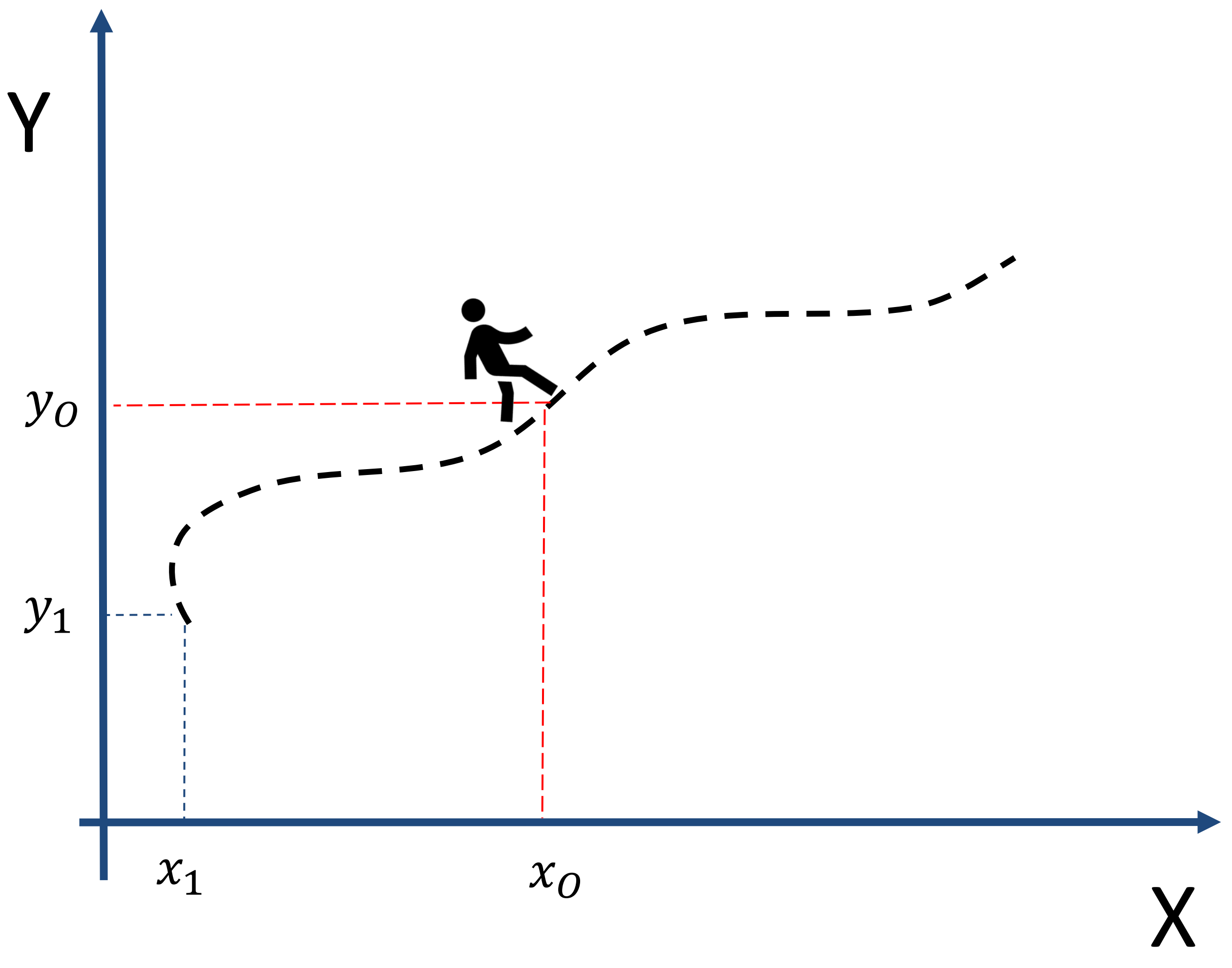
What is trajectory prediction?

We start by observing a moving agent…



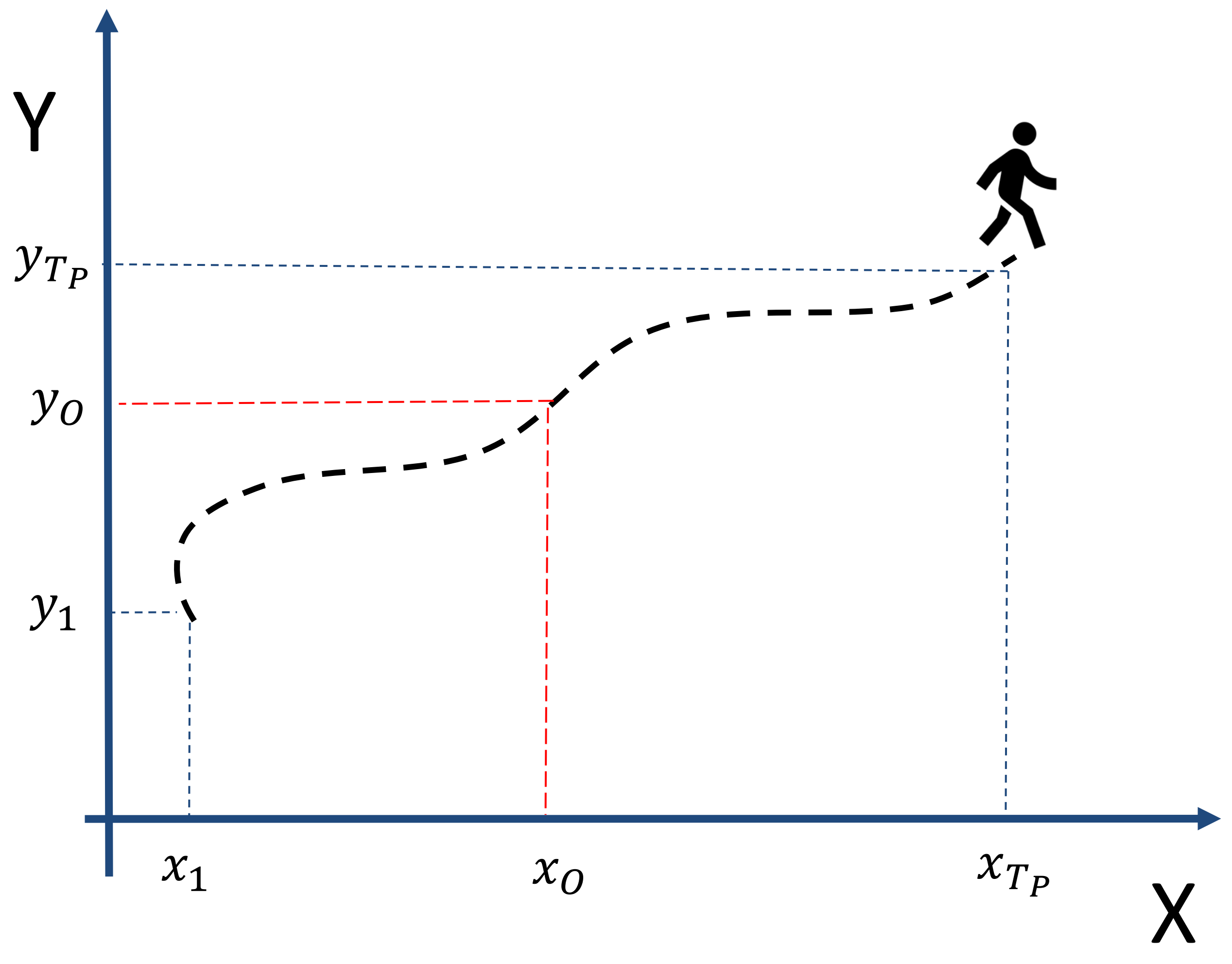
What is trajectory prediction?

We observe it for a given **observation horizon**, O…



What is trajectory prediction?

Based on *observed trajectory states* and other contextual factors, we predict future velocities, which are then converted into positions for a **given future horizon**, $T_p - O$.



<u>Trajectory Prediction</u>

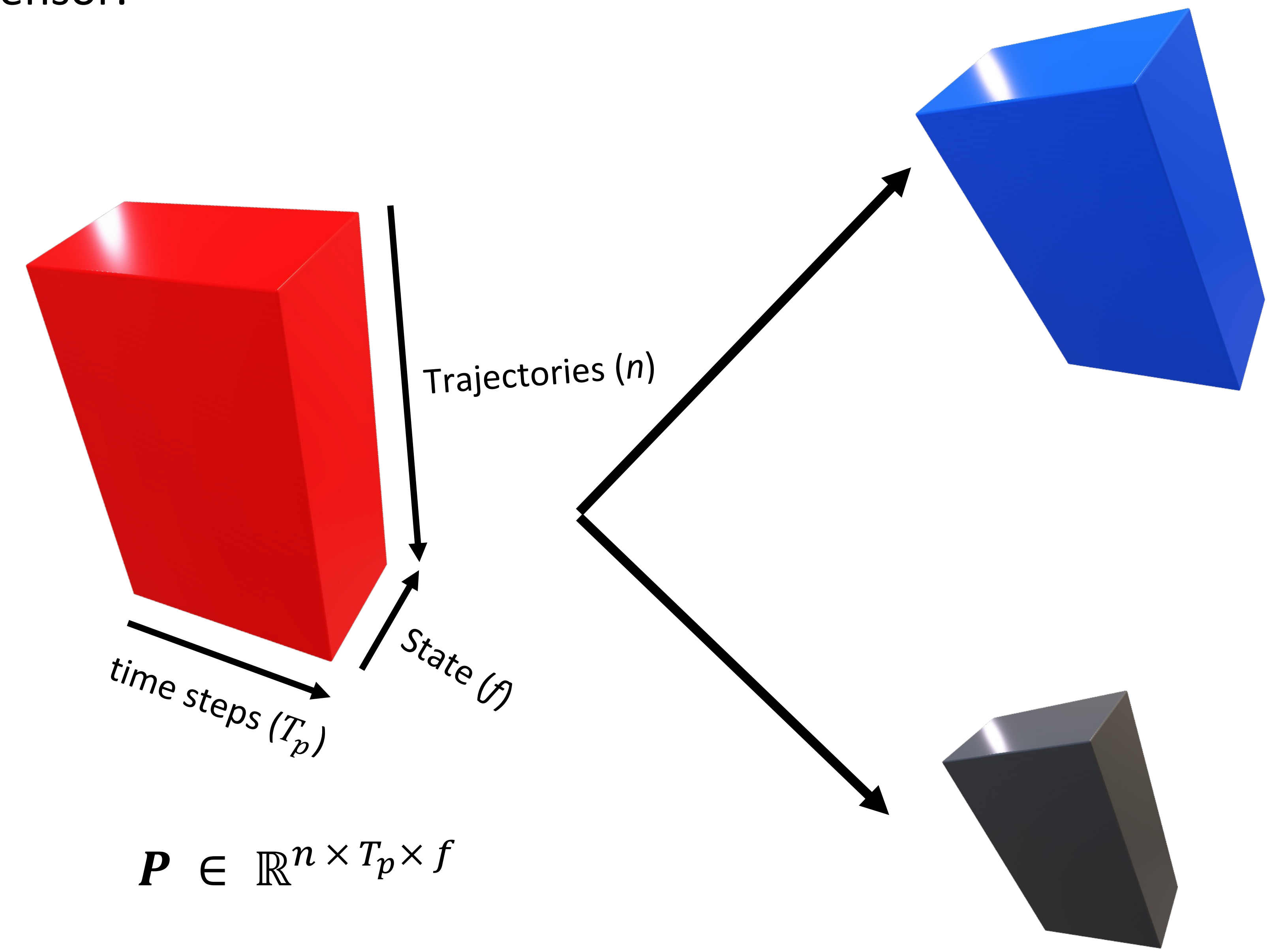
Observe $\boldsymbol{S} = (\boldsymbol{s_t})_{t=1}^{O}$

Predict $\boldsymbol{Y_S} = ((\dot{x}_t, \dot{y}_t))_{t\,=\,O+1}^{T_P}$

What are trajectory *states*?

- Trajectory cues that can characterize the moving agent trajectory, such as: 2D absolute positions, displacements, rotated trajectories, velocities, actions, head orientation, etc..

- Let's say that we cutoff all trajectories to a defined horizon and stack them. Then, a dataset of trajectories becomes a 3D tensor:



$\boldsymbol{P} \in \mathbb{R}^{n \times T_p \times f}$

Observation:

$\boldsymbol{S} \in \mathbb{R}^{n \times O \times f}$

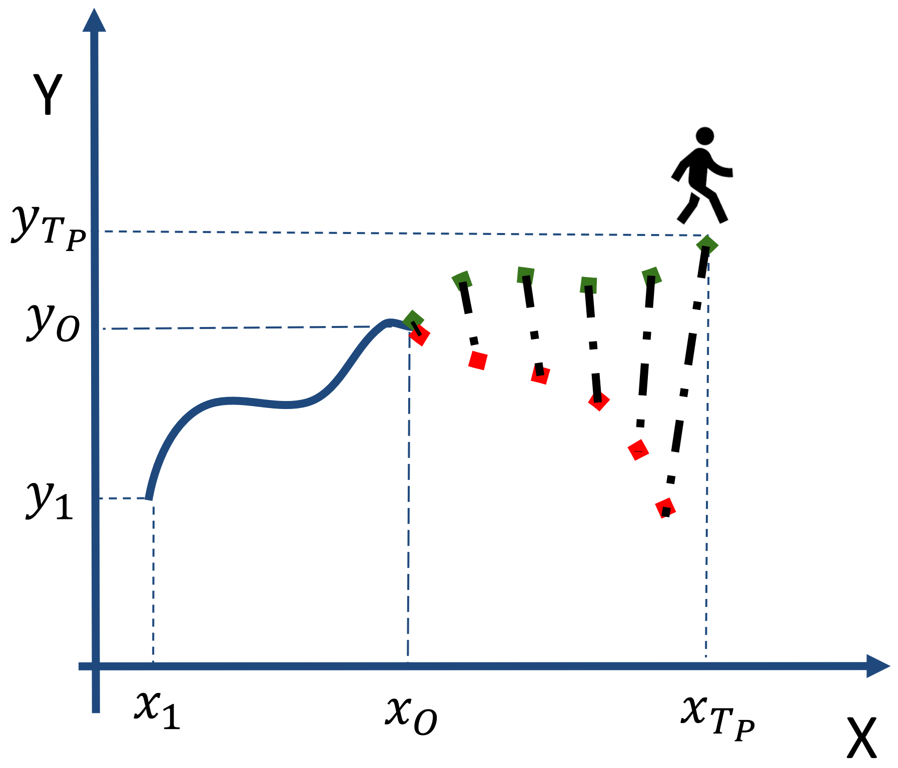
Future: **Only 2D velocities in our work!**

$\mathbf{Y_S} \in \mathbb{R}^{n \times L \times 2}, L = T_p - O$

$\boldsymbol{P}$ is the entire trajectory dataset.
$\boldsymbol{S}$ is the observation part of the trajectory dataset.
$\boldsymbol{Y}$ is the future part of the trajectory dataset.
n is the number of trajectories.
T_p is the total number of time steps.
f is the length of the state configuration (features).
O is the number of observed time steps.
L is the number of future time steps.

How can we evaluate trajectory prediction methods?

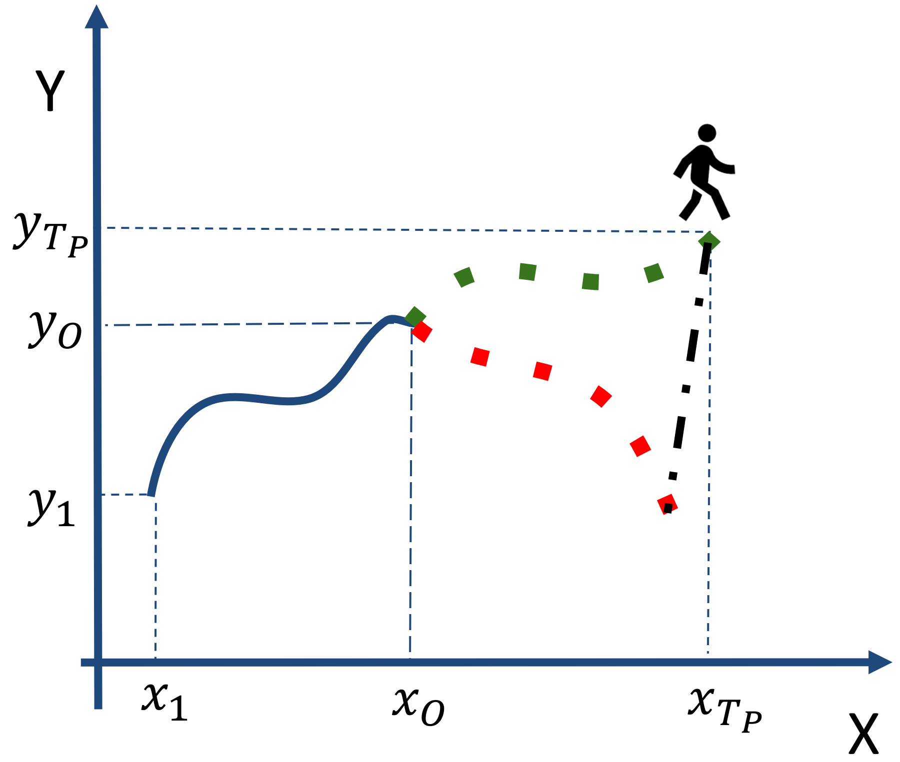
Top-K Average Displacement Error (ADE)



Ground truth
Trajectory prediction

T_p is the number of time steps.
O is the number of observed time steps.

Top-K Final Displacement Error (FDE)



3. Gap, Approach & Results.

What are current methods lacking?

How could this thesis bridge the gaps?

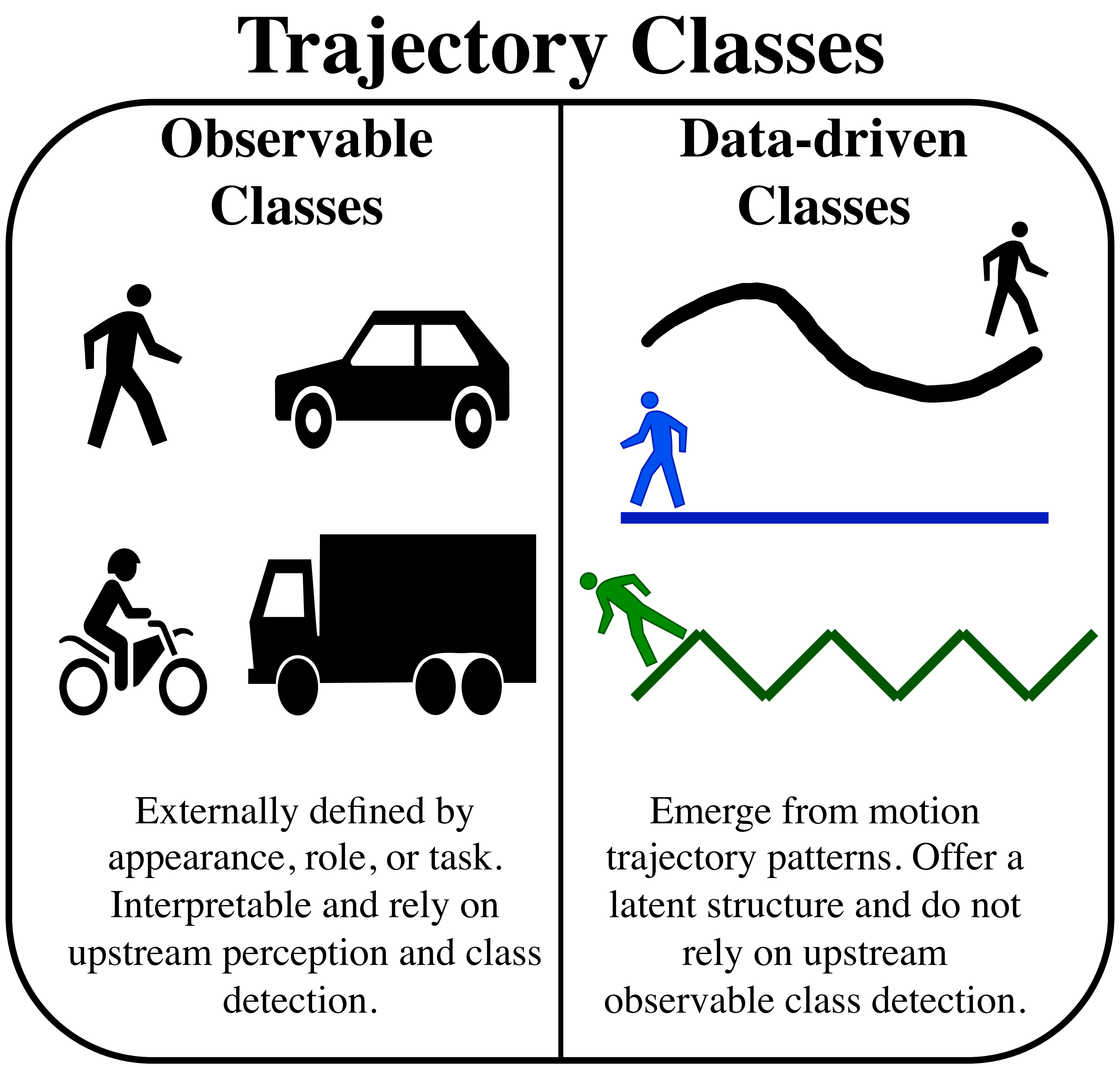
General claim

Robots share space with numerous dynamic agents in anthropocentric environments. The behaviors of dynamic agents are shaped by a complex interplay between **external** and **internal factors**.

External factors include environment, such as obstacles, and **internal factors** comprise activities, roles, intentions, and personal preferences.

This thesis addresses the challenges of discovering and modeling trajectory heterogeneity, a phenomenon arising from these factors, as ***trajectory classes***, which group trajectories based on perceived appearance or trajectory cues.

These classes can originate from two primary sources: *observable classes*, defined by human semantics and accessible via perception systems, and *data-driven classes*, which are automatically learned from the structure and dynamics of the trajectory data.

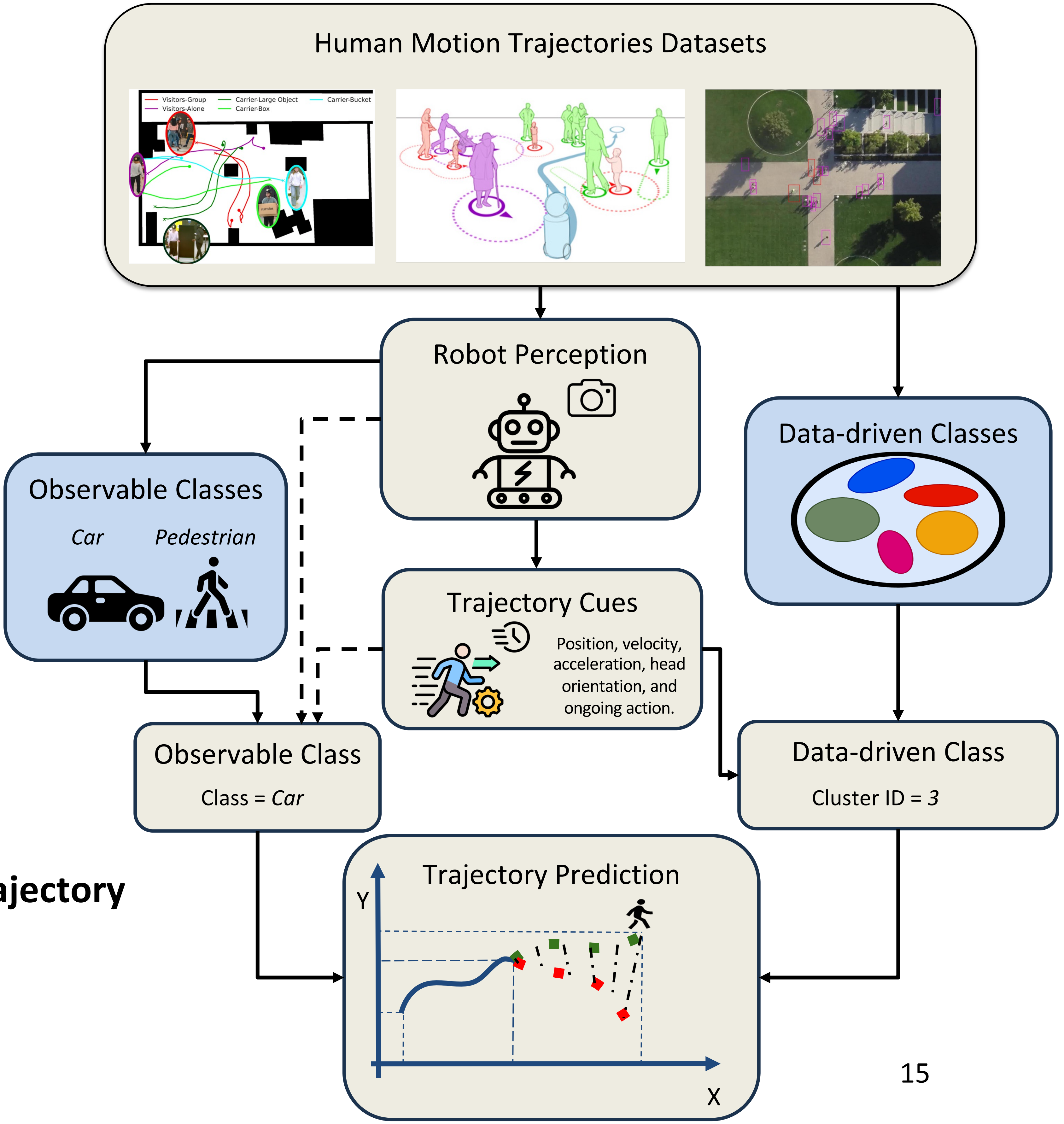


General claim

(1) We find these classes in human motion trajectories datasets, where observable classes can be detected through robot perception and data-driven classes are directly found from the data.

(1.1) External and internal factors affect measurable trajectory cues, such as velocity, acceleration, heading, ongoing action, etc., which can be used to detect/infer the trajectory classes.

(2) We incorporate trajectory classes both *observable* and *data-driven* into trajectory prediction methods to enhance trajectory predictions.



Gaps:

No systematic study addressing trajectory heterogeneity for trajectory prediction.

Existing methods lack appropriate datasets and mechanisms to leverage and cope with heterogeneity in trajectory data.

How can we bridge the gaps?

Research Goal: How can observable and data driven classes be effectively used to analyze and predict human trajectory data?

RQ1. What datasets are needed to study heterogeneity in human trajectories, and how to collect them?

RQ2. How can observable classes improve trajectory prediction?

RQ3. How can frame-based actions improve trajectory prediction?

RQ4. How to learn data-driven classes for trajectory prediction?

RQ5. How can data-driven classes improve trajectory prediction?

C1. Heterogeneous trajectory data: THÖR-MAGNI data collection.

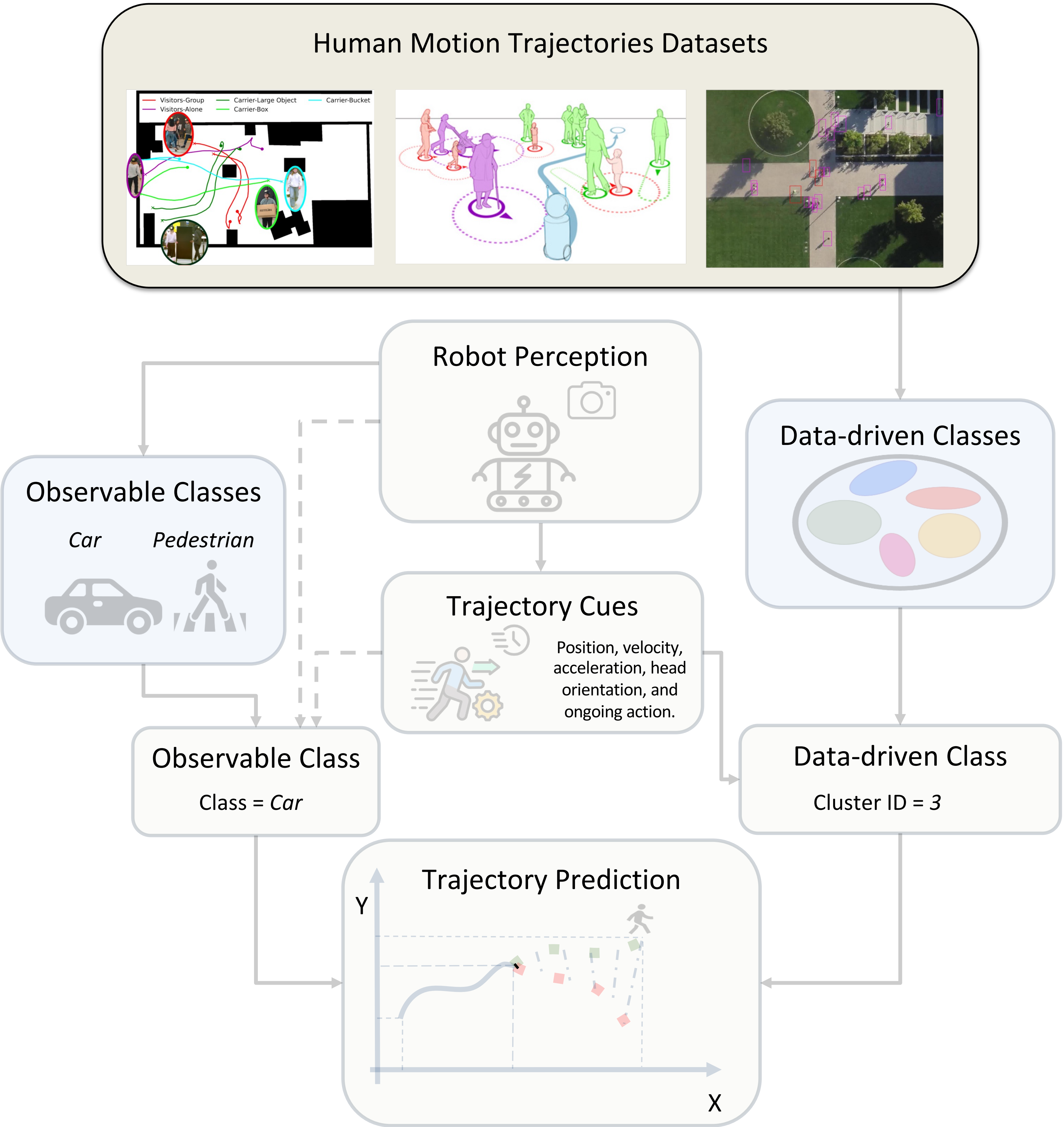
C2. Study of observable class-conditioned methods on various data settings (balanced vs. imbalanced datasets and low data regimes).

C3. Pitfalls of observable classes and THÖR-MAGNI *Act.*

C4. Data-driven classes (observation-driven, future-driven, and full-driven) and Self-conditioned GAN.

C5. Improved training settings for GAN-based forecasters and a multi-stage framework with novel predictions ranking methods.

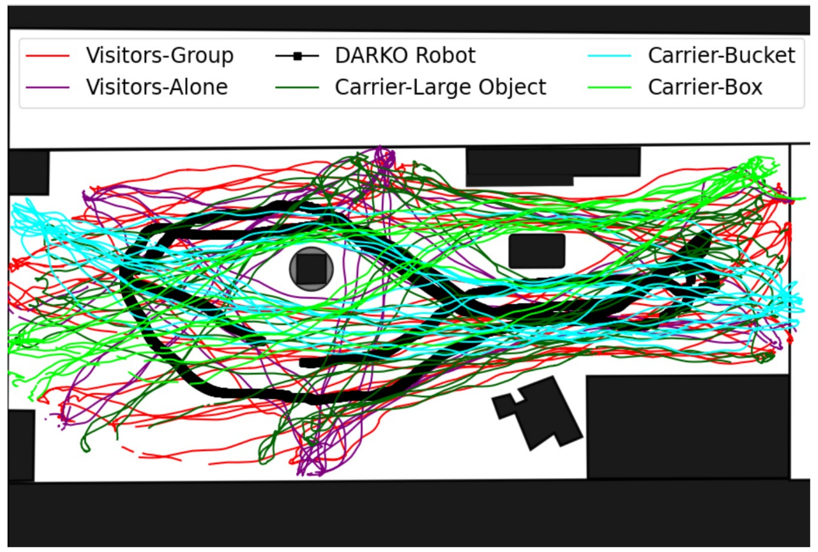
RQ1: What datasets are needed to study heterogeneity in human trajectories, and how to collect them?



What datasets are needed to study heterogeneity in human trajectories, and how to collect them?

THÖR-MAGNI data collection[1]

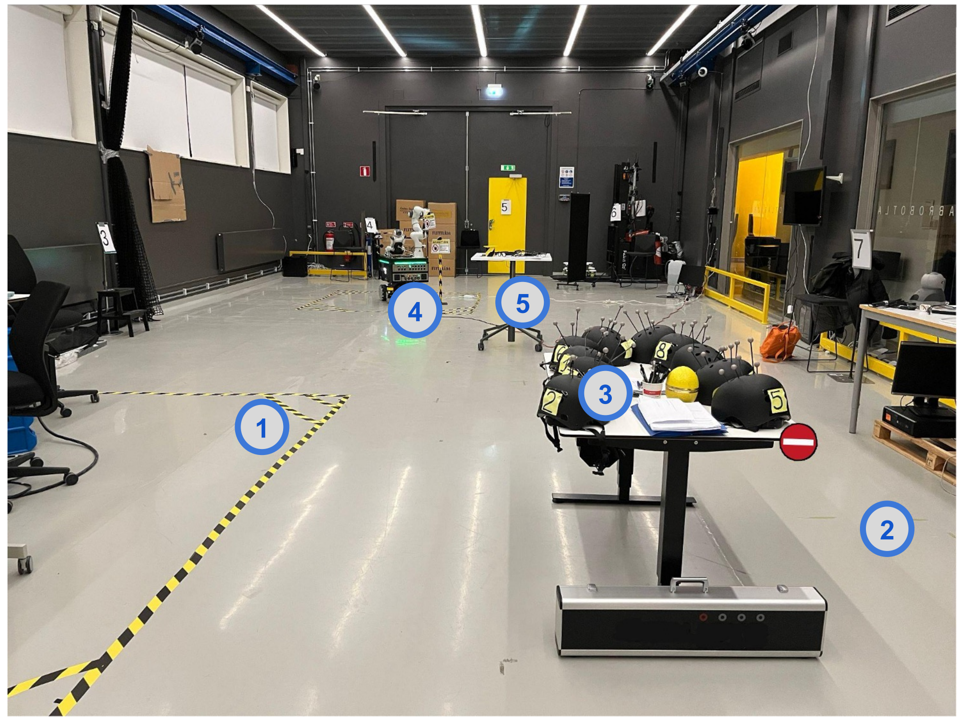
[1] T. Schreiter, **T.R. de Almeida**, et al. THÖR-MAGNI: A large-scale indoor motion capture recording of human movement and robot interaction. *IJRR '24*.



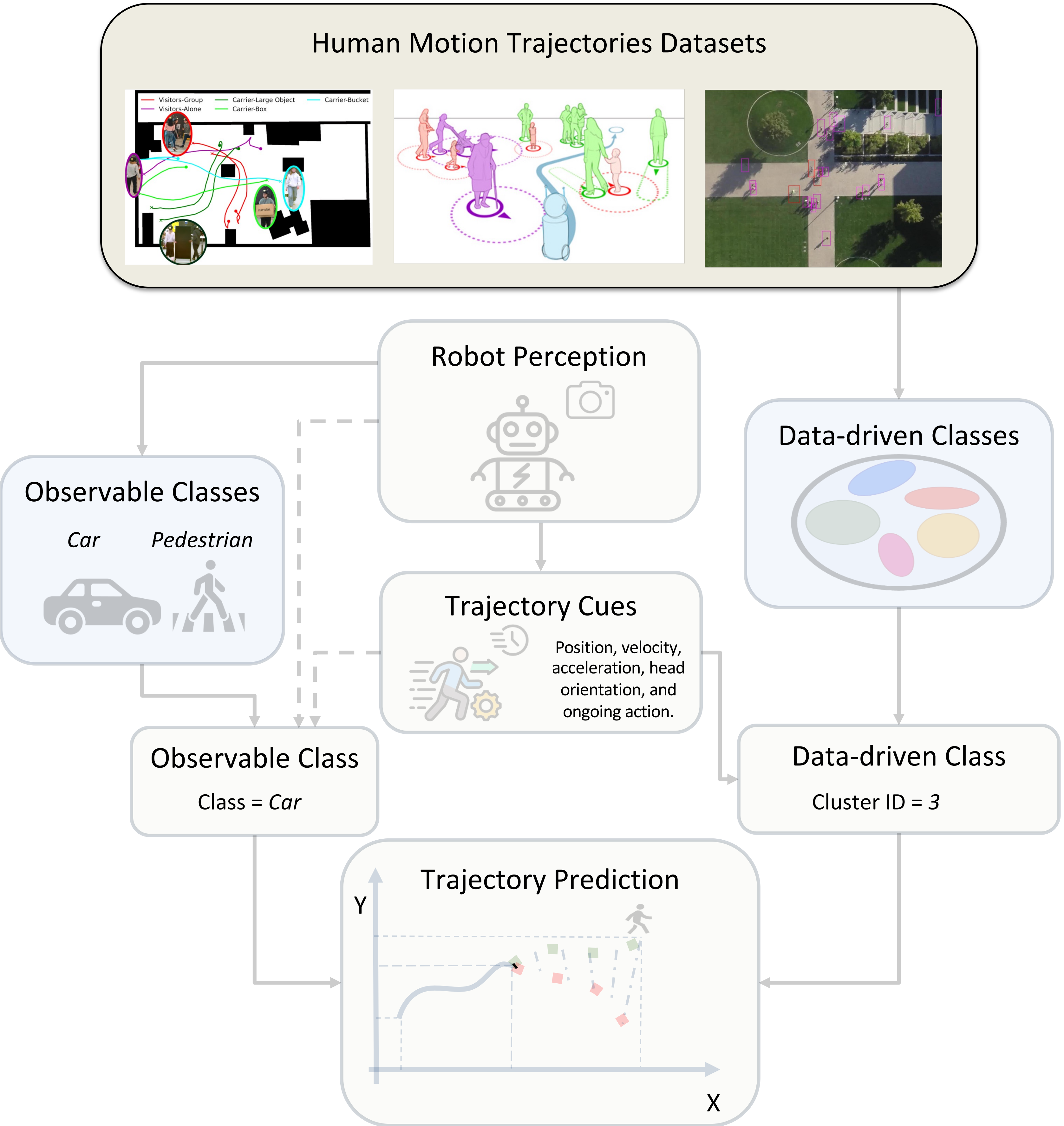
We need data, specially in robotics environments where heterogeneous trajectory annotated data is scarce.

Main features:

- Motion capture system to accurately track moving agents.
- Constrained **semantically-rich environment** promote different types of trajectories in 5 different scenarios.

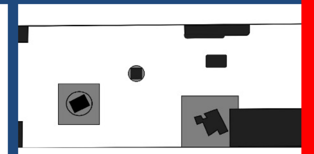
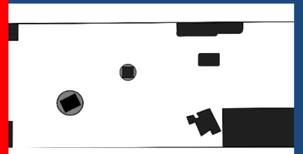
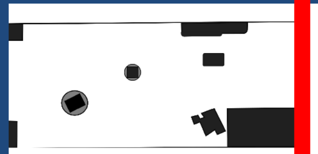


- People wearing tracking helmets 3 moving between 7 goal points.
- Mobile robot in the scene 4 moving and as static obstacle.
- Contextual semantics: lane markings 1, one-way corridors 2, static obstacles 5.



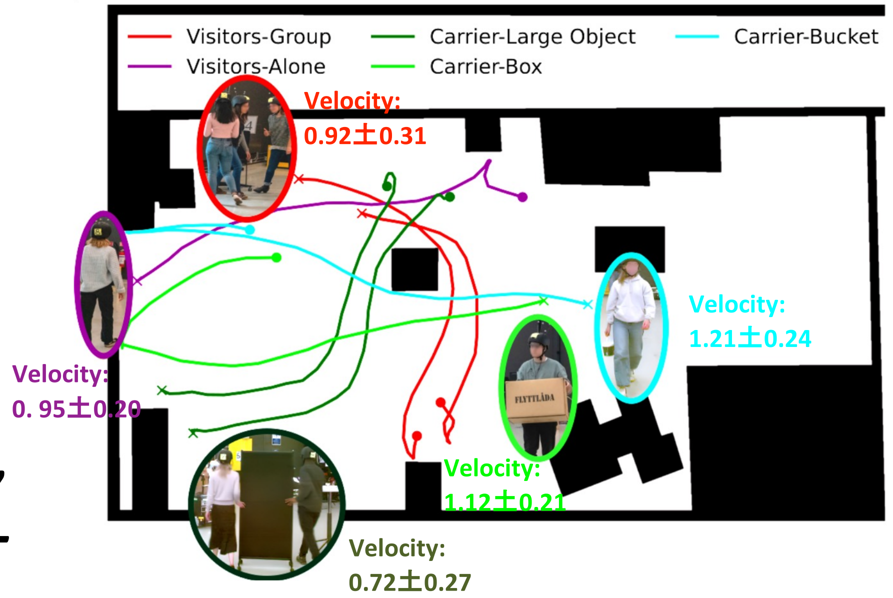
What datasets are needed to study heterogeneity in human trajectories, and how to collect them?

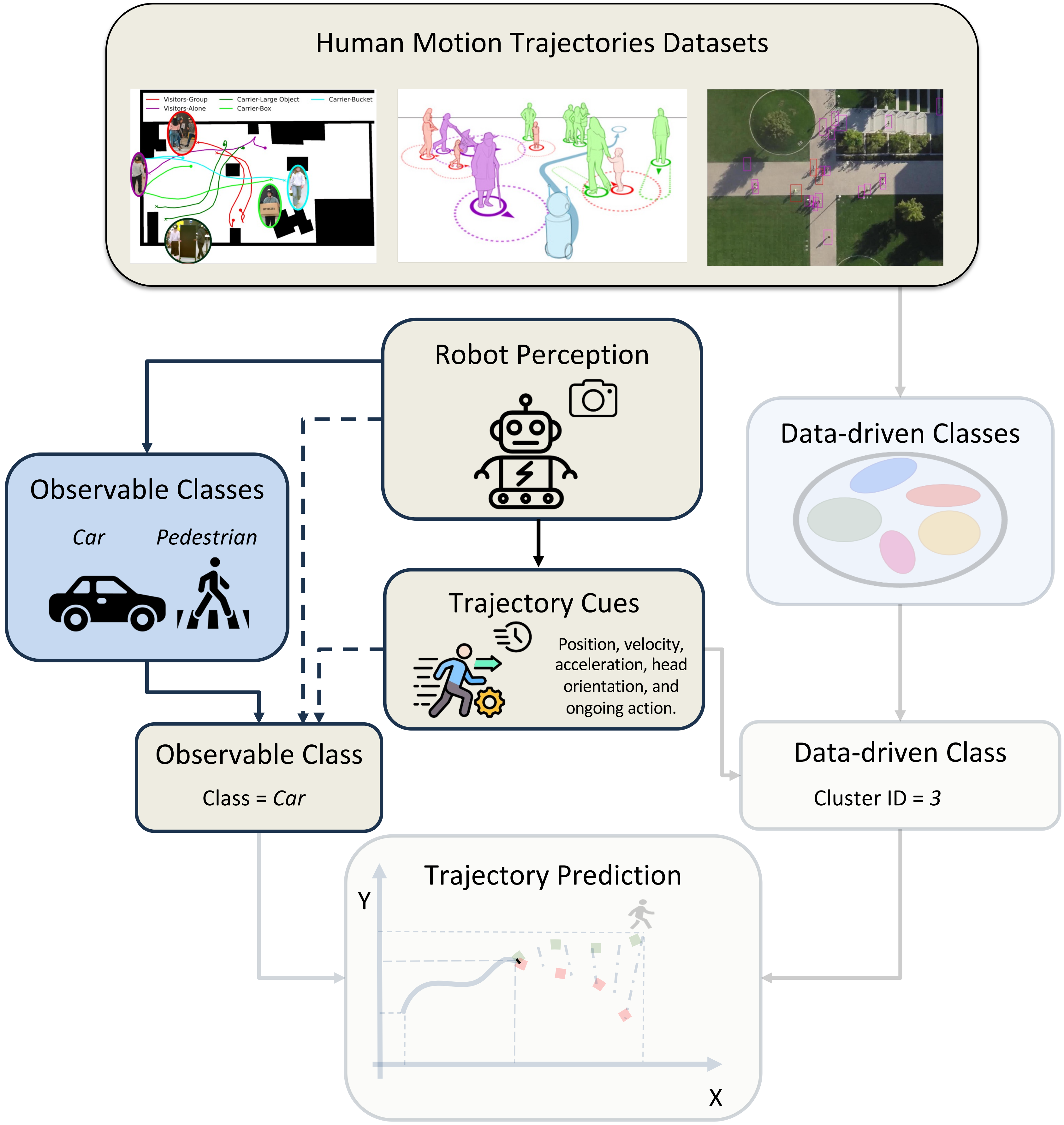
THÖR-MAGNI data collection

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Human Roles	Alone Groups	Alone, Groups, Objects Carriers	Alone, Groups, Objects Carriers	Alone, Groups, Human-Robot Interaction (HRI)	
Robot Motion	Stationary (obstacle)	Stationary (obstacle)	Moving with 2 driving styles	Directional (semi-autonomous)	
Environment Layout					
Scenario Conditions (A and B)	Semantics No semantics		2 robot driving styles	Verbal HRI Multimodal HRI	

Motion Heterogeneity:

- Scenarios 2 and 3: 90 min. of trajectory data.
- Human roles corresponding to various activities: moving in groups of 2 and 3 people (***Visitors-Group***), moving individually (***Visitors-Alone***), ***Carrier-Bucket***, ***Carrier-Box***, ***Carrier-Large Object***.



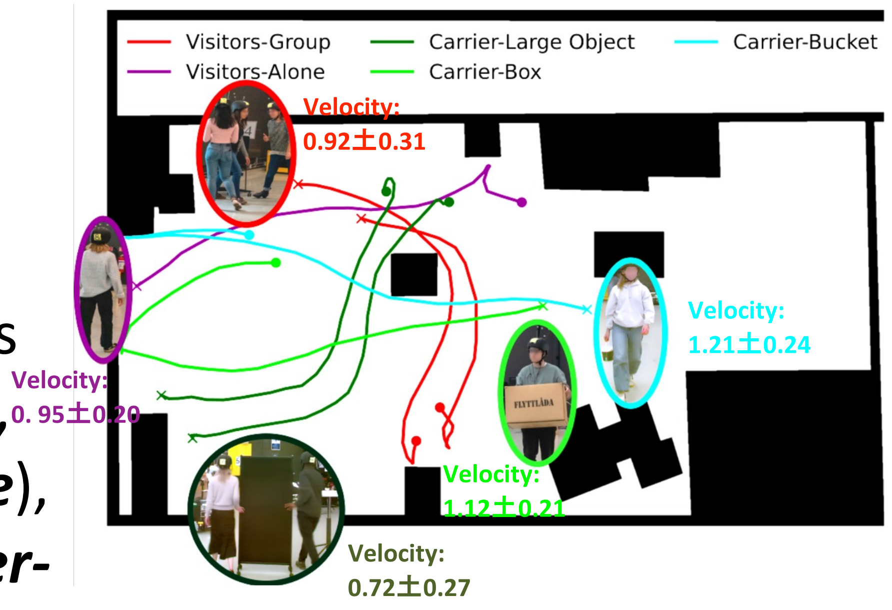


What datasets are needed to study heterogeneity in human trajectories, and how to collect them?

THÖR-MAGNI data collection

Motion Heterogeneity:

- Scenarios 2 and 3: 90 min. of trajectory data.
- Human roles corresponding to various activities: moving in groups of 2 and 3 people (***Visitors-Group***), moving individually (***Visitors-Alone***), ***Carrier-Bucket***, ***Carrier-Box***, ***Carrier-Large Object***.



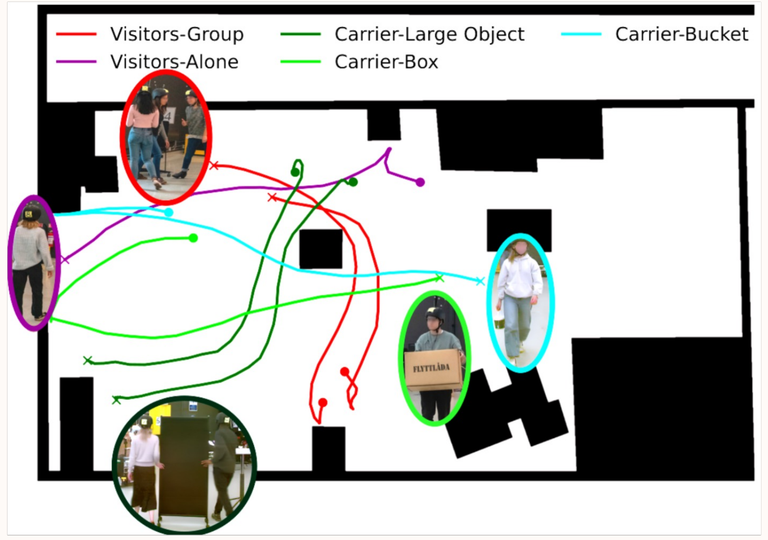
Velocities average and standard deviation per role in 20-time steps trajectories.

Role	MAGNI-S2	MAGNI S3A	MAGNI S3B
Carrier-Box	$1.12\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.21$	$1.15\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.27$	$1.08\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.26$
Carrier-Bucket	$1.21\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.24$	$1.21\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.20$	$1.13\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.18$
Carrier-Large Object	$\boldsymbol{0.72\ \frac{m}{s} \pm 0.27}$	$\boldsymbol{0.68\ \frac{m}{s} \pm 0.32}$	$\boldsymbol{0.76\ \frac{m}{s} \pm 0.36}$
Visitors-Alone	$0.95\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.20$	$0.92\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.29$	$0.87\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.32$
Visitors-Group	$0.92\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.31$	$0.87\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.26$	$0.84\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.31$
Total	$0.95\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.51$	$0.91\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.48$	$0.90\ \frac{\mathrm{m}}{\mathrm{s}} \pm 0.48$

Previously…

RQ1. What datasets are needed to study heterogeneity in human trajectories, and how to collect them?

C1. THÖR-MAGNI, a dataset of heterogeneous trajectory data, which was scarce for robotics environments.



Learning Outcome

The observable classes in THÖR-MAGNI demonstrate distinct motion patterns that could be important for **trajectory prediction** in robotics environments.

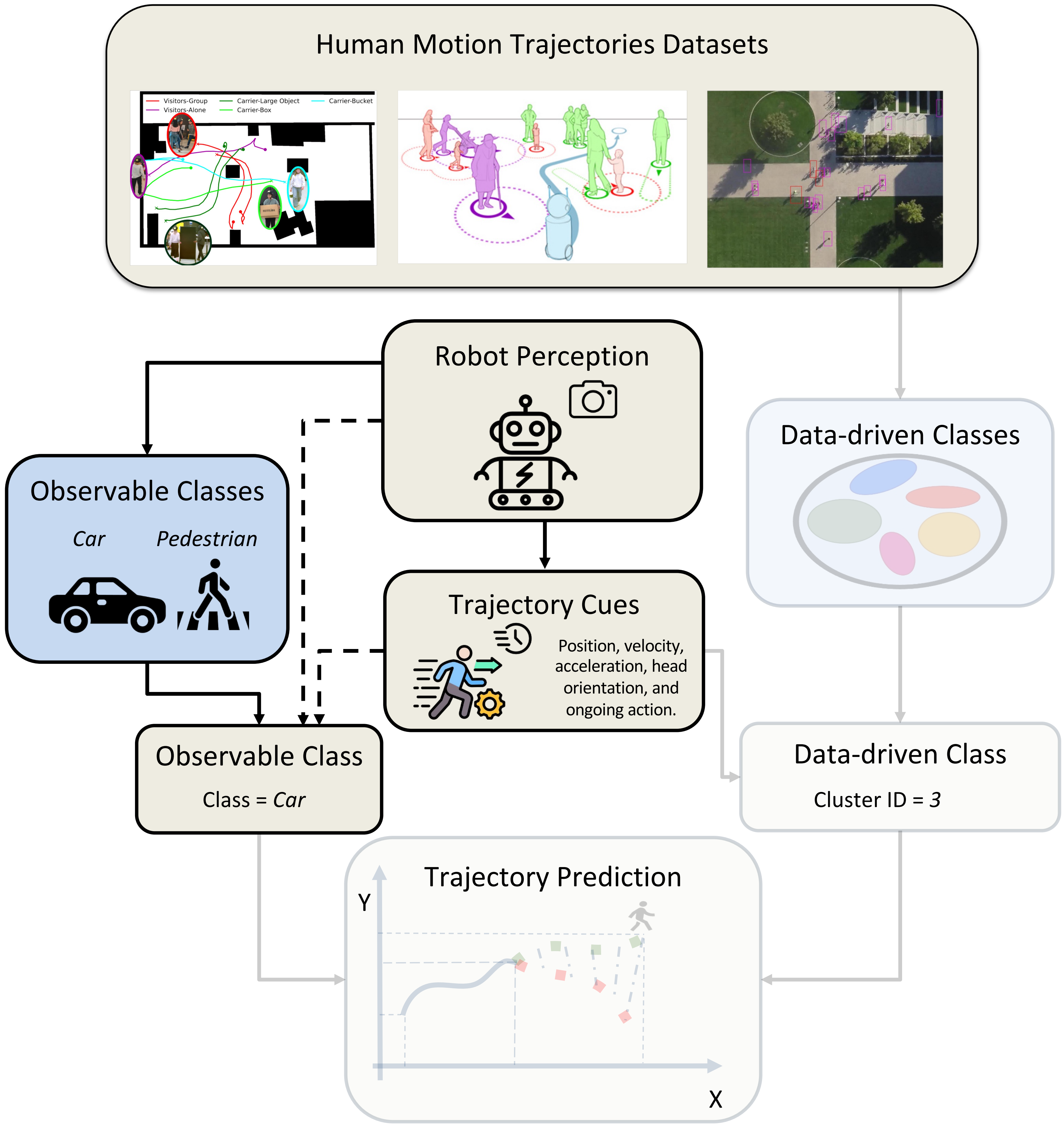
Next.

Limitations of prior art

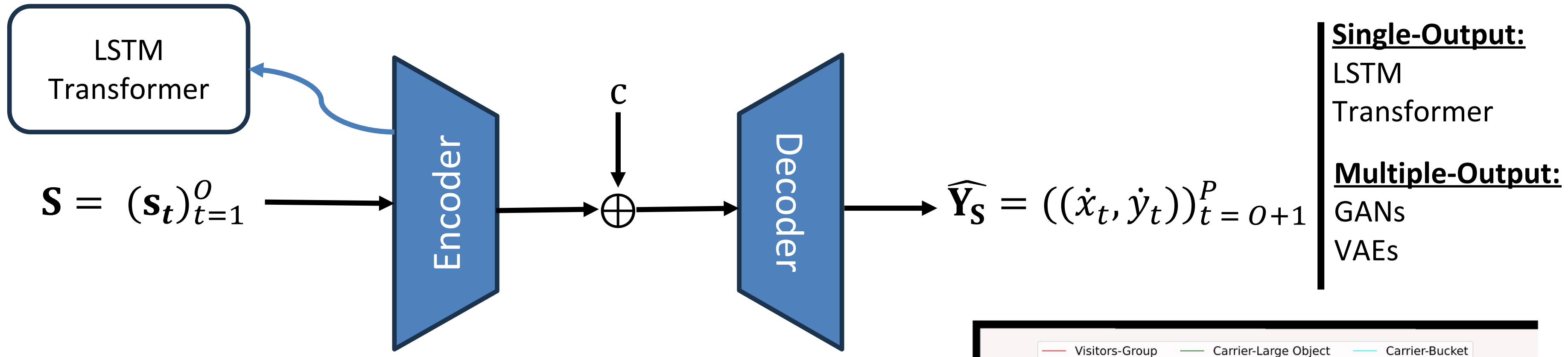
- Existing heterogeneous trajectory prediction methods tailored for autonomous driving depend on domain-specific contextual features (e.g., agent shape).
- Robotics applications present unique challenges (cold-start scenario).
- Robotics and autonomous driving domains may feature non-uniform class distributions (imbalanced data), leading to decreased performance of deep learning-based methods.

RQ2. How can observable classes improve trajectory prediction?

How can observable classes improve trajectory prediction?



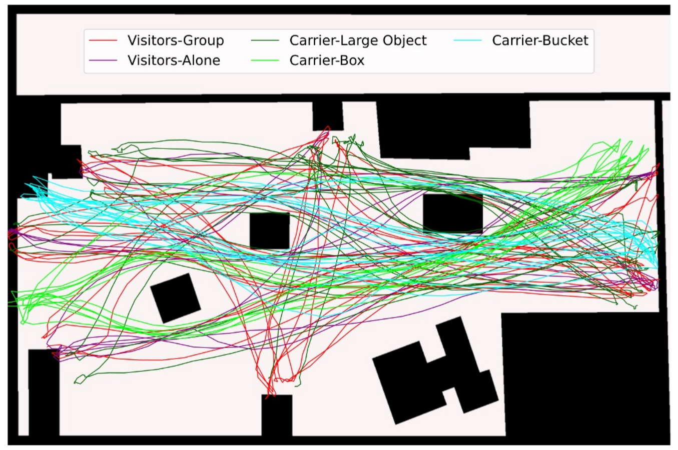
Study of class-conditioned trajectory prediction methods in various data settings[2]



Datasets & Class Proportions:

THÖR-MAGNI dataset

5 observable classes: *Carrier-Large Object* (25.7%), *Visitors-Group* (23.6%), *Visitors-Alone* (22.7%), *Carrier-Box* (14.1%), and *Carrier-Bucket* (13.9%)



Stanford Drone Dataset

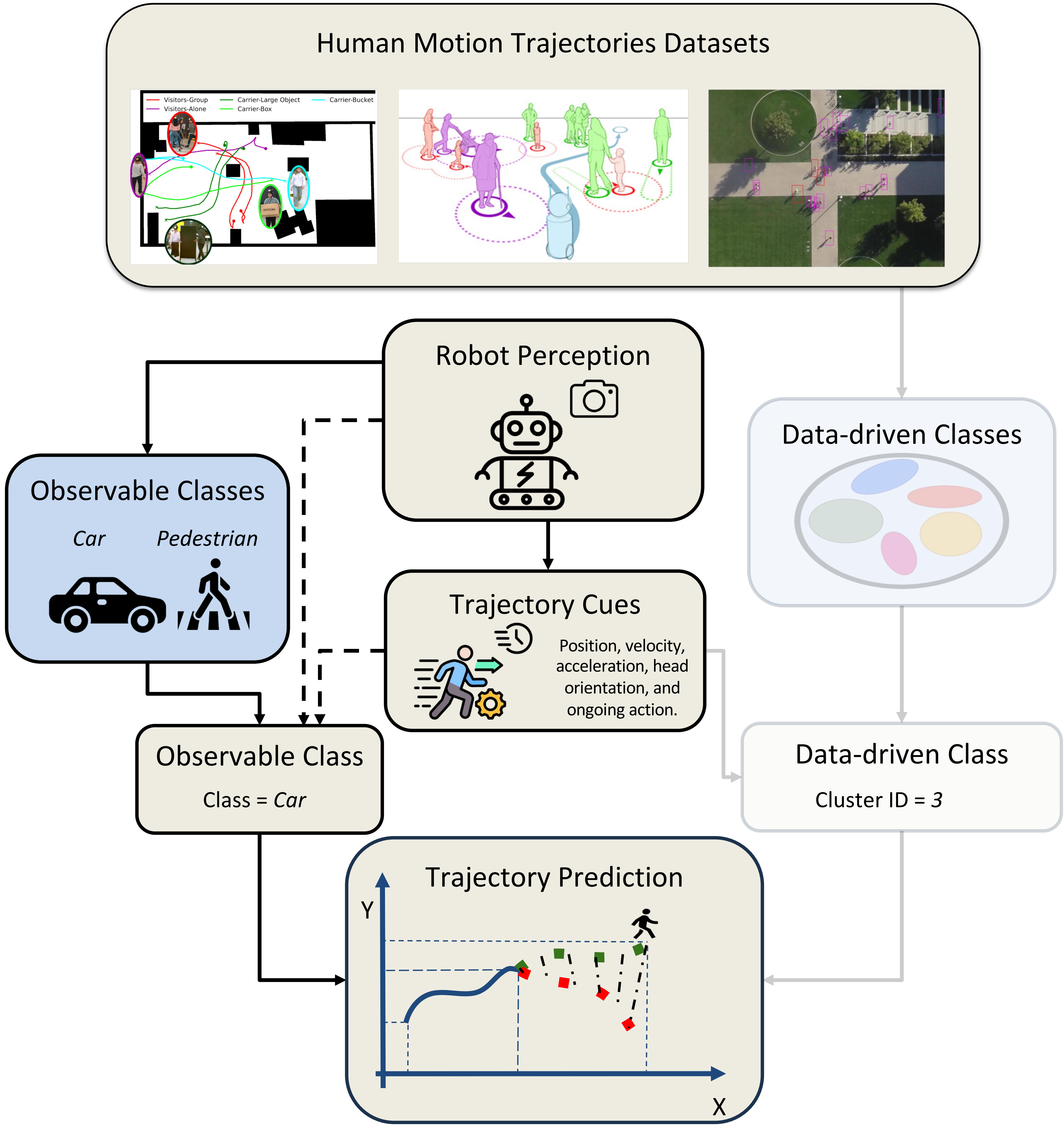
3 observable classes: *Pedestrians* (66.4%), *Bicyclists* (34.3%), and *Cars* (1.1%)

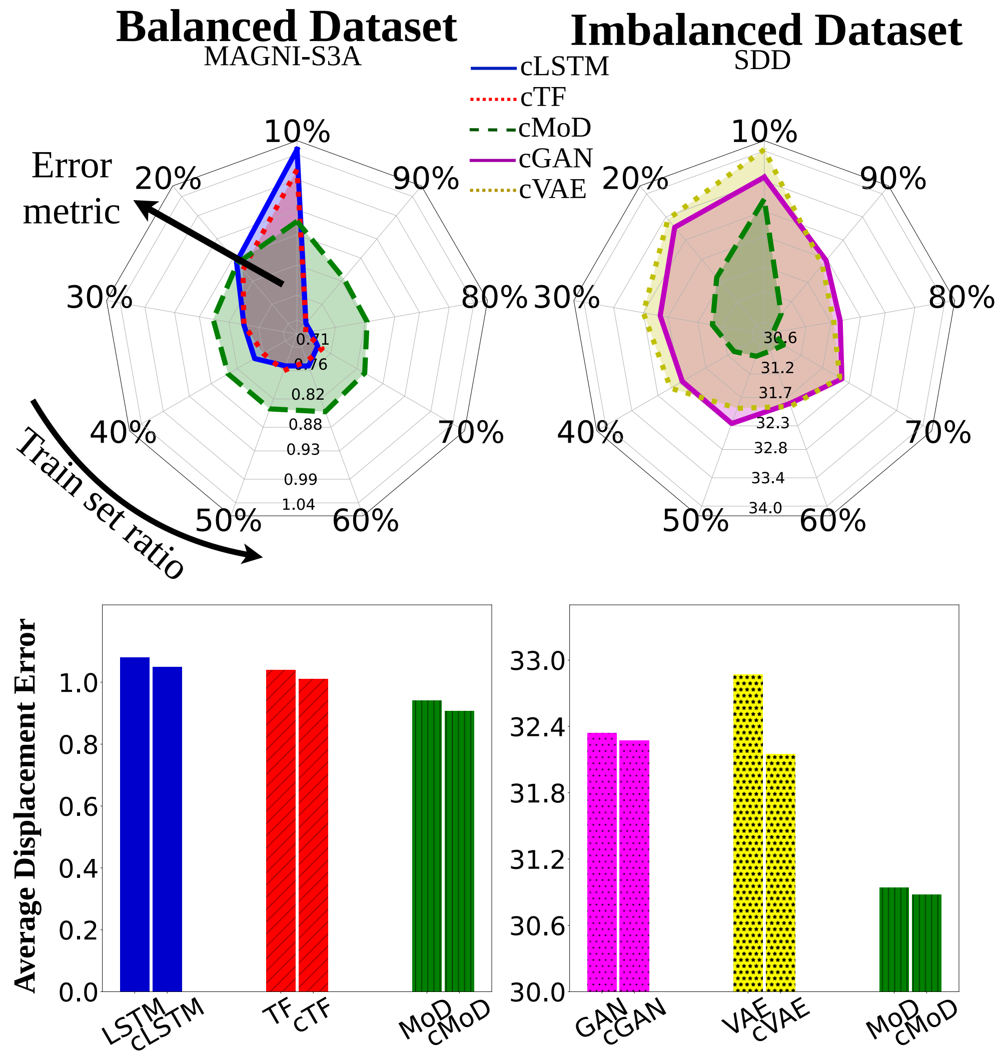
Imbalanced dataset! (non-uniform class proportions!)

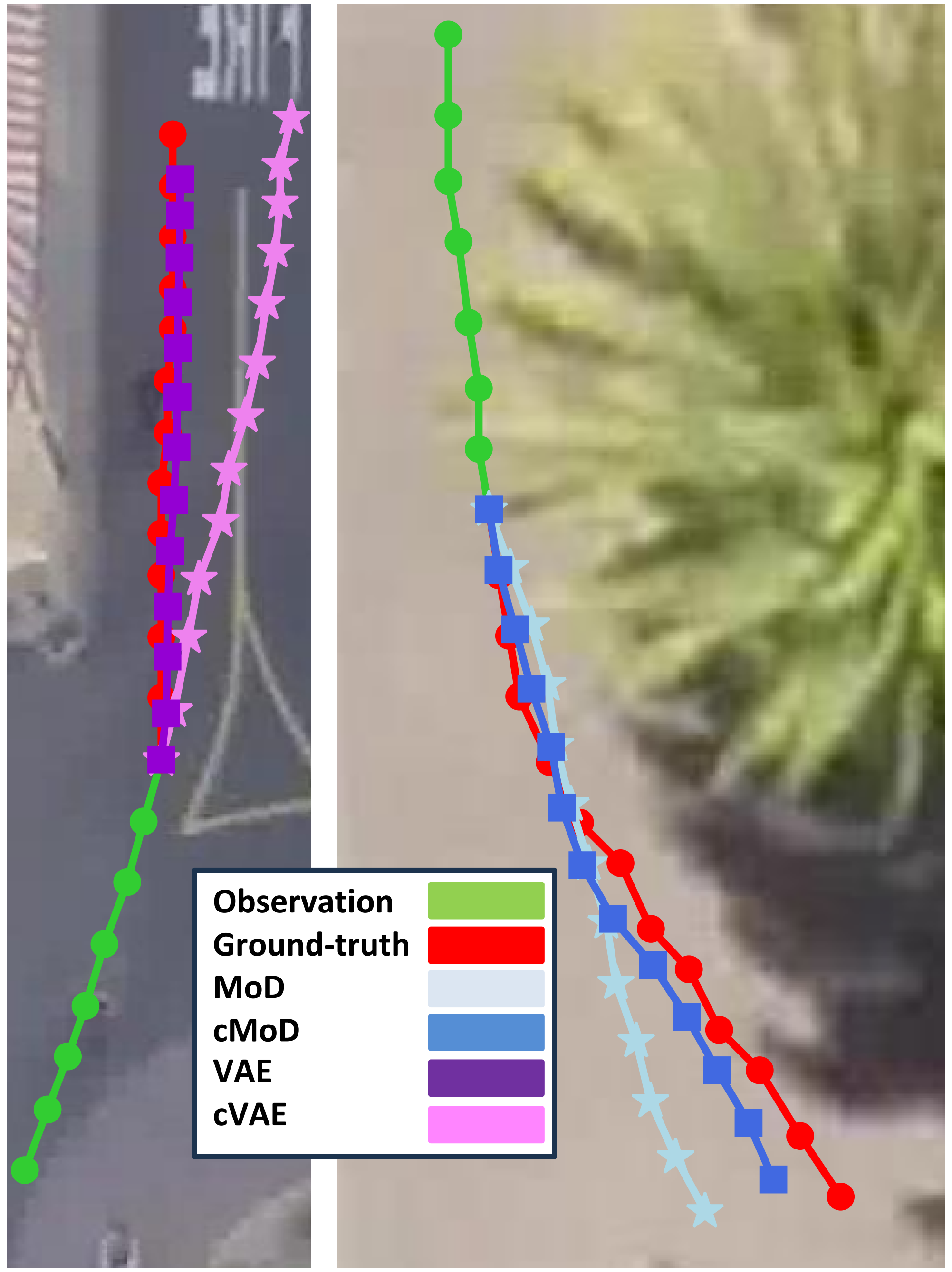
[2] **T. R. de Almeida**, Zhu Y, et al., "Trajectory Prediction for Heterogeneous Agents: A Performance Analysis on Small and Imbalanced Datasets," *RA-L '24*.

How can observable classes improve trajectory prediction?

Study of class-conditioned trajectory prediction methods

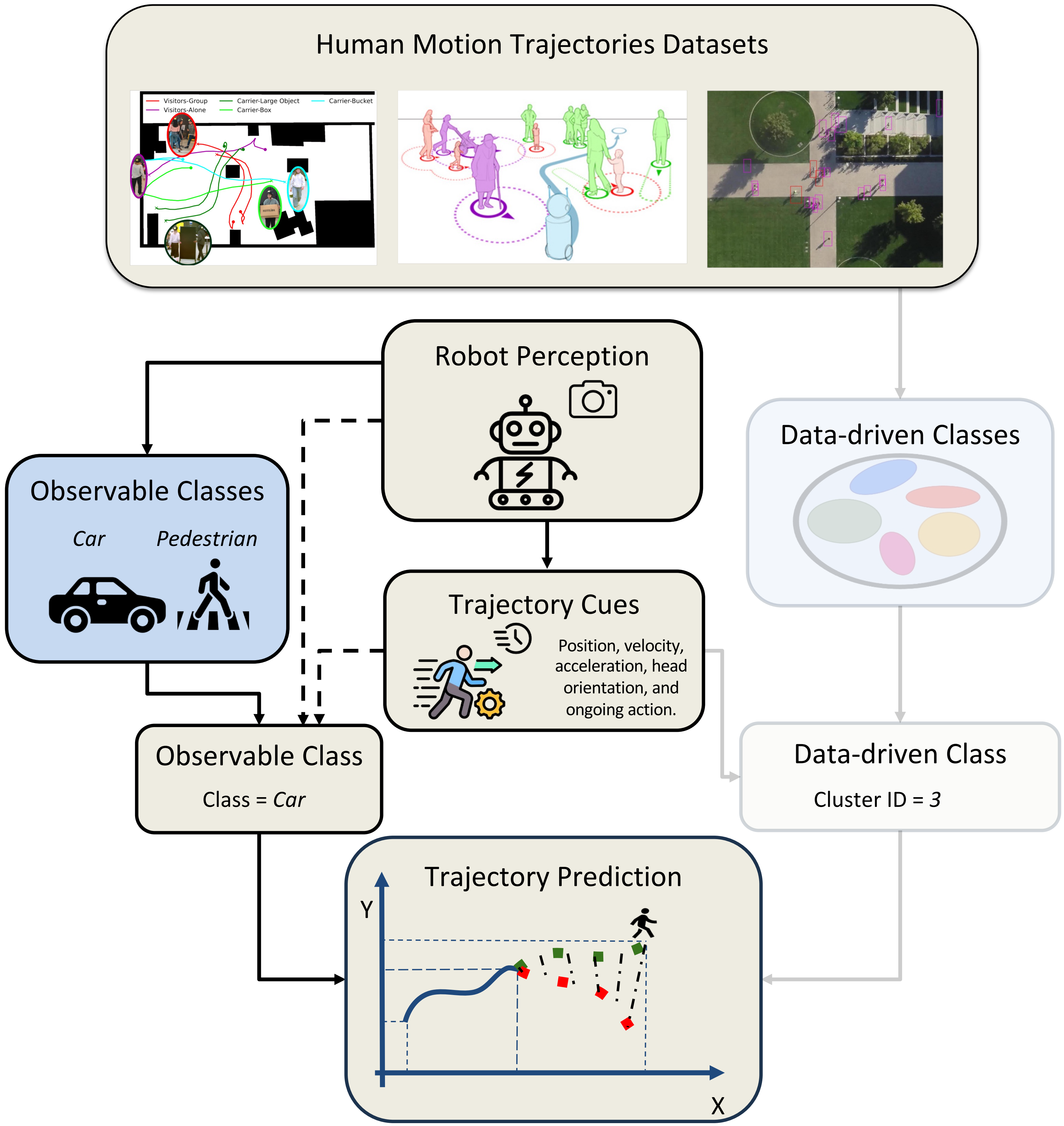


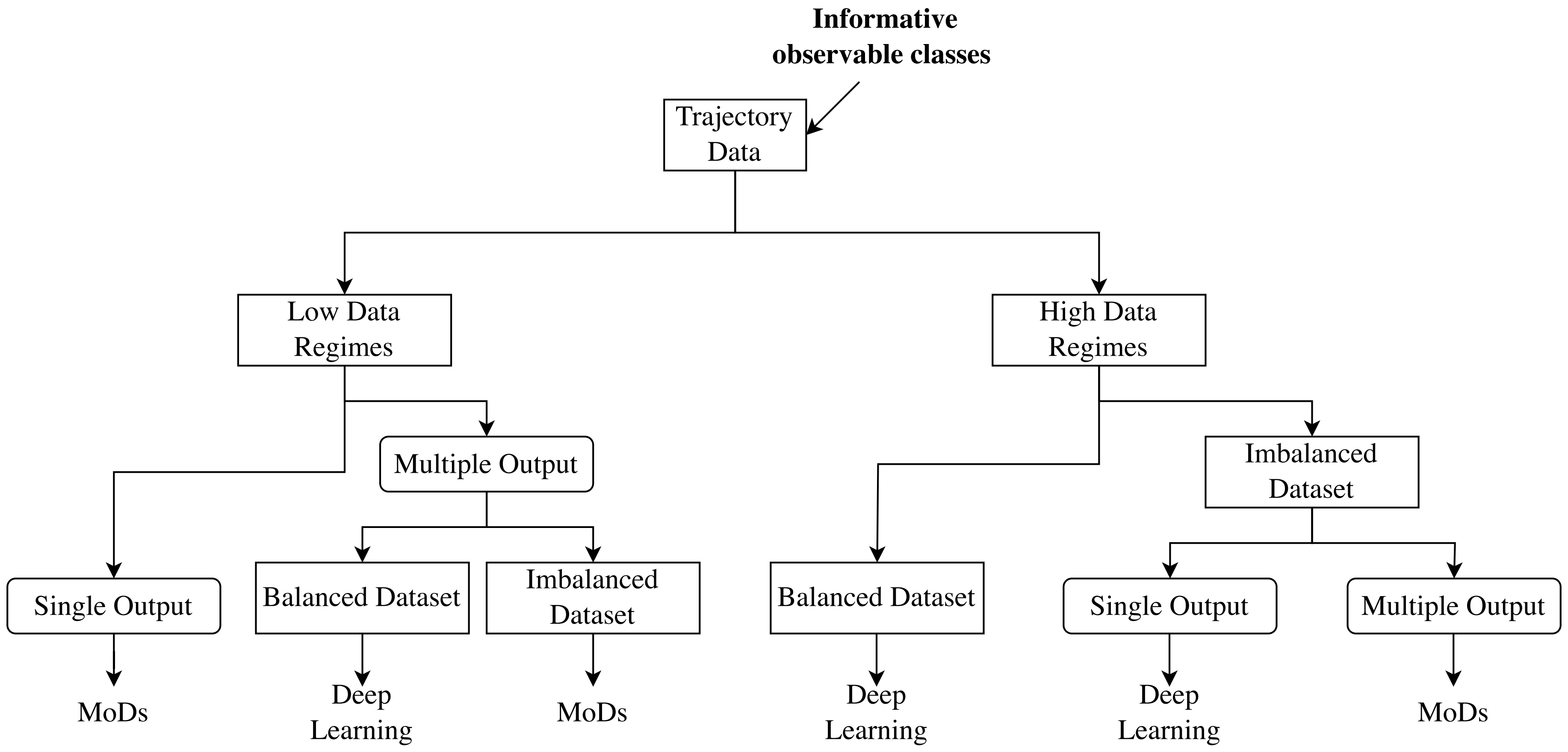




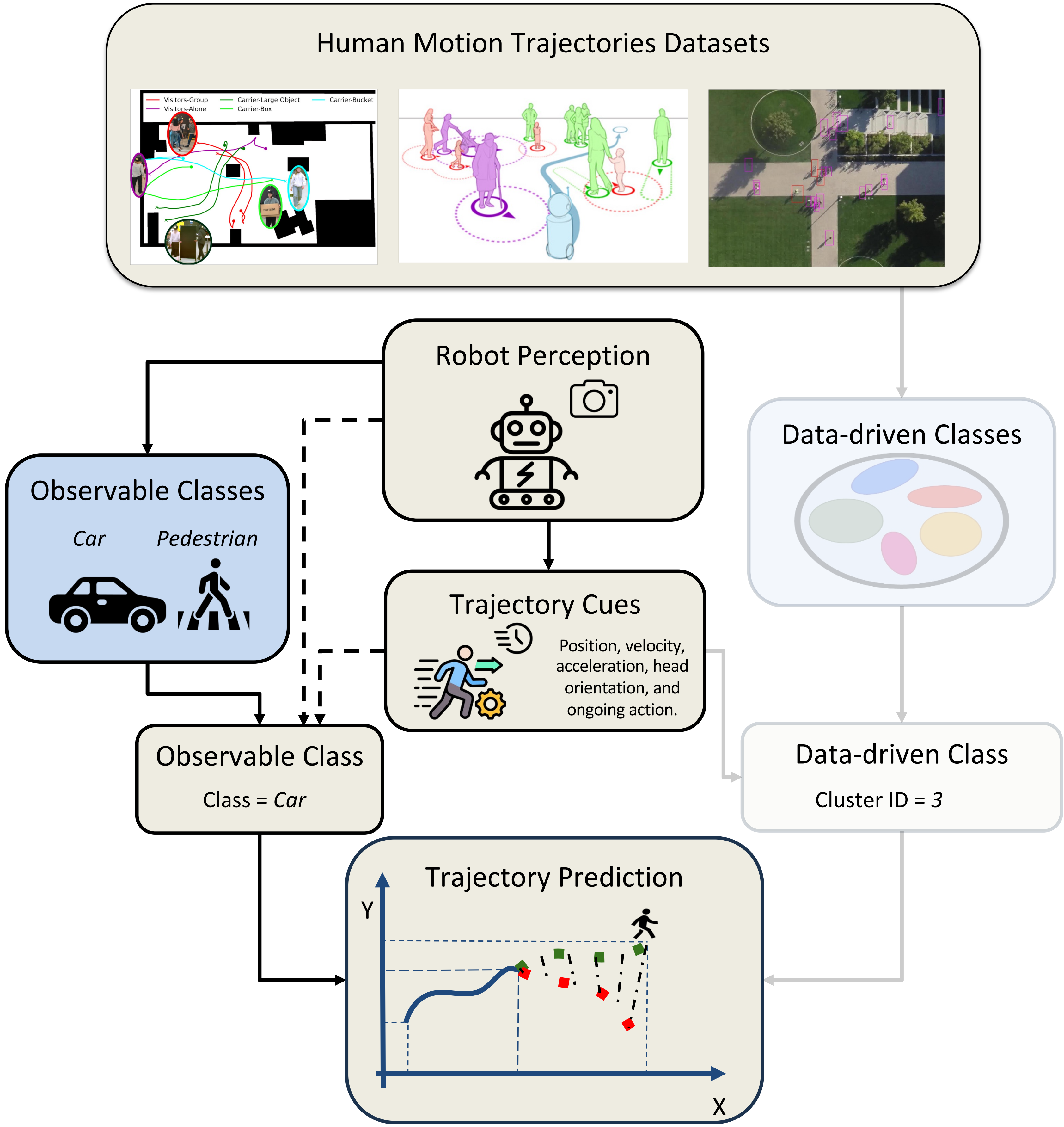
How can observable classes improve trajectory prediction?

Study of class-conditioned trajectory prediction methods



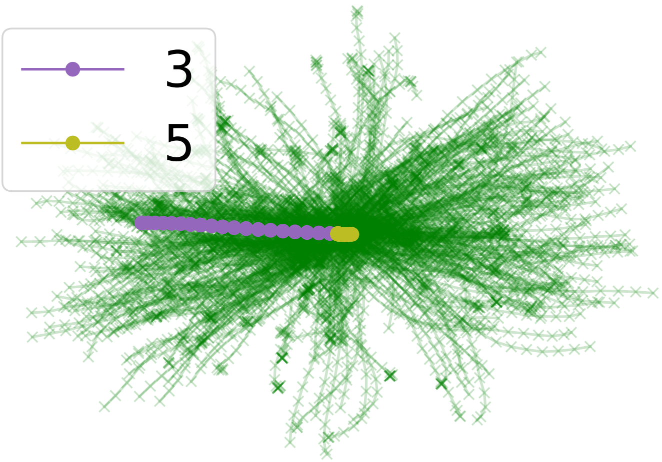


Observable Classes Pitfalls

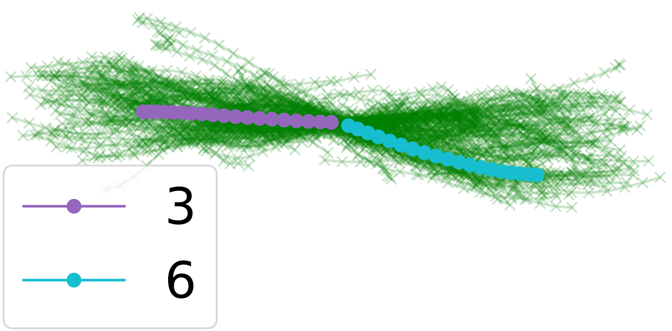


Ambiguity: different classes may share the same trajectory patterns, and a single class may encompass different trajectory patterns.

Visitors-Alone



Carrier-Bucket



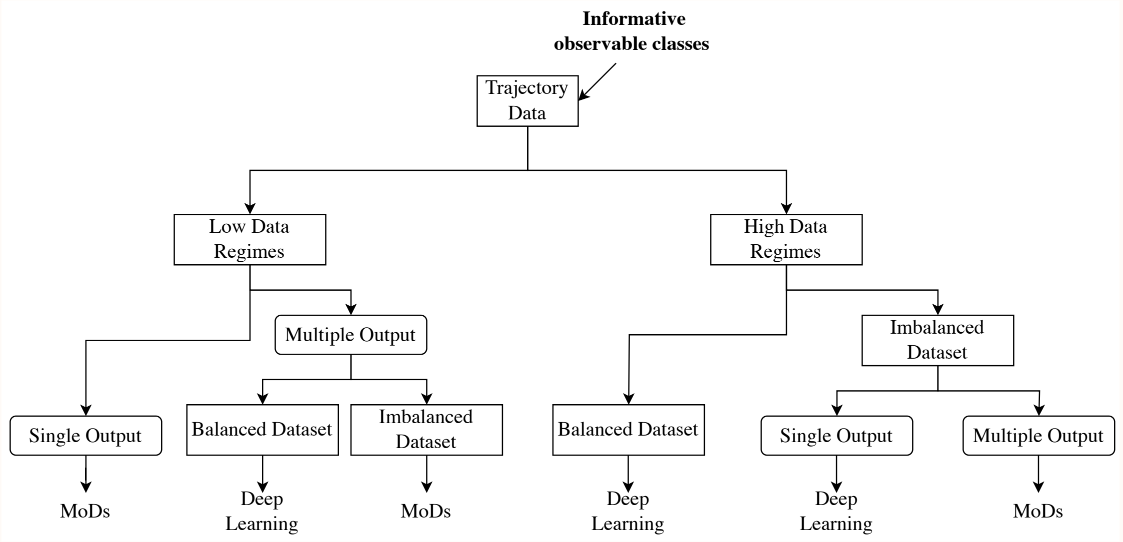
Latent Semantic Analysis

Cluster id	Carrier Box	Carrier Bucket	Carrier Lar. Obj.	Visitors Alone	Visitors Group 2	Visitors Group 3	Total
1	**62**	3	54	73	**43**	8	243
2	6	1	27	53	33	3	123
3	35	**91**	53	**87**	30	**21**	**317**
4	**77**	37	**64**	60	36	**17**	291
5	20	21	**132**	**106**	**37**	11	**327**
6	2	**54**	29	47	19	7	158
Total	202	207	359	426	198	67	1459

Previously…

RQ2. How can observable classes improve trajectory prediction?

C2. Study of class-conditioned trajectory prediction methods on class imbalanced and low-data regimes settings.



Learning Outcome

MoDs have an edge over deep generative methods in imbalanced data scenarios and over single output methods in low-data regimes.

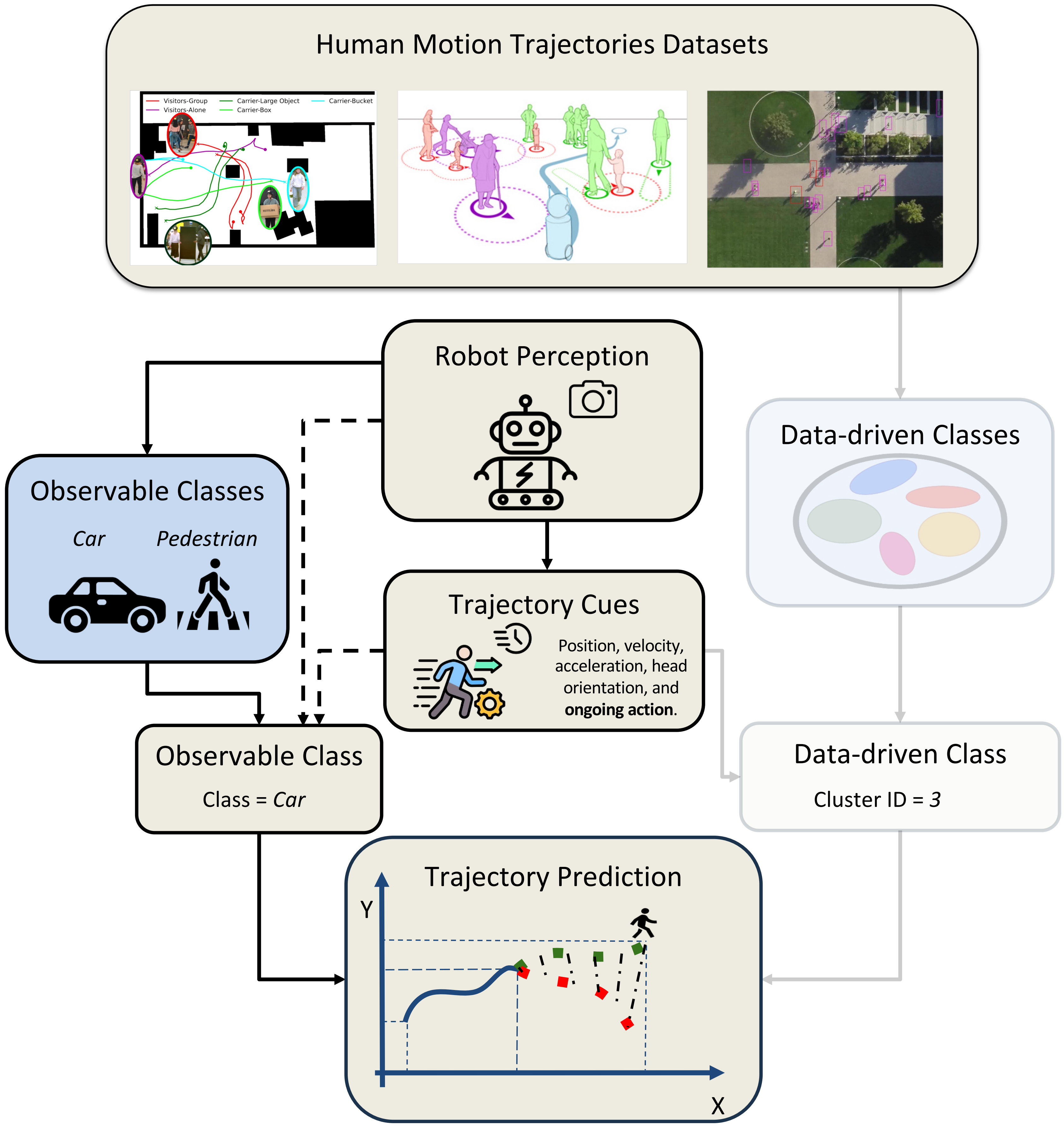
Next.

Limitations of observable classes

- **Ambiguity**: the same observable class may contain different motion patterns. For instance, a *Car* being parked or moving.
- Different observable classes may contain the same motion patterns. For instance, a *Car* moving slowly and a moving *Biker*.
- The **static nature** of observable classes can limit their representation power given the complex behavior of dynamic agents.

RQ3. How can frame-based actions improve trajectory prediction?

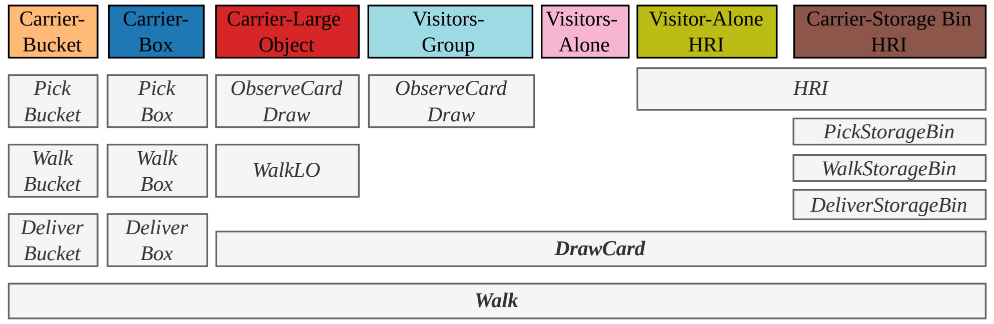
How can frame-based actions improve trajectory prediction?

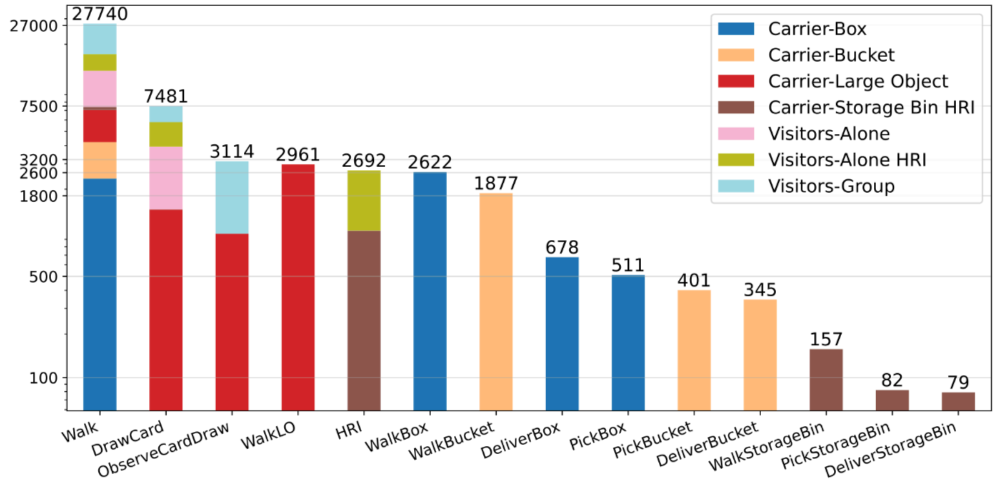


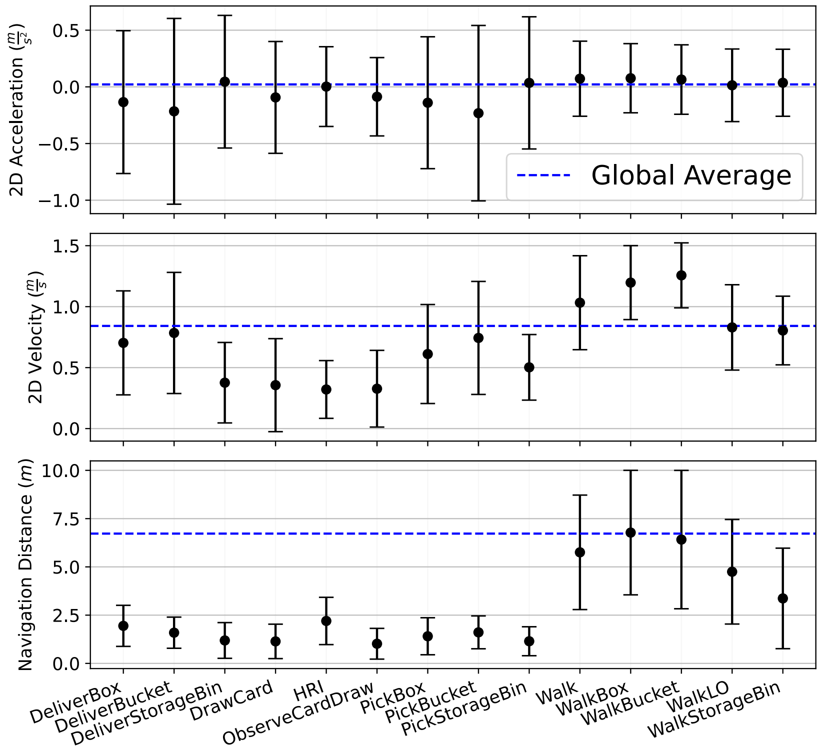
THÖR-MAGNI *Act* data collection[3]

8.3 hours of fine-grained actions are part of the sequence of input states and can **reduce ambiguity** as they decompose the observable class in a time-varying sequence of actions.

$\mathcal{A} = \{$ *Walk, DrawCard, ObserveCardDraw, WalkLO, PickBucket, WalkBucket, DeliverBucket, PickBox, WalkBox, DeliverBox, PickStorageBin, WalkStorageBin, DeliverStorageBin, HRI* $\}$

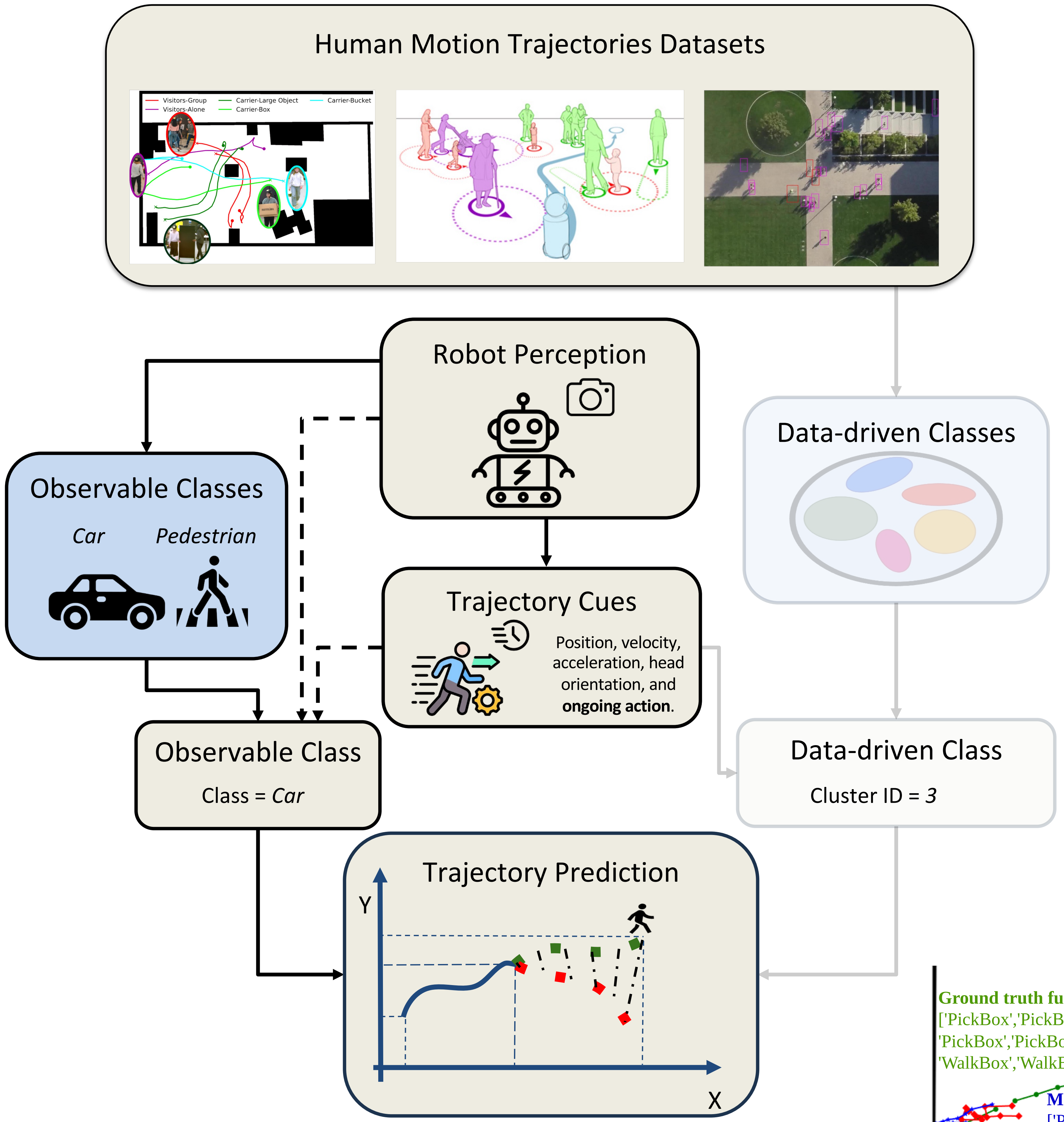




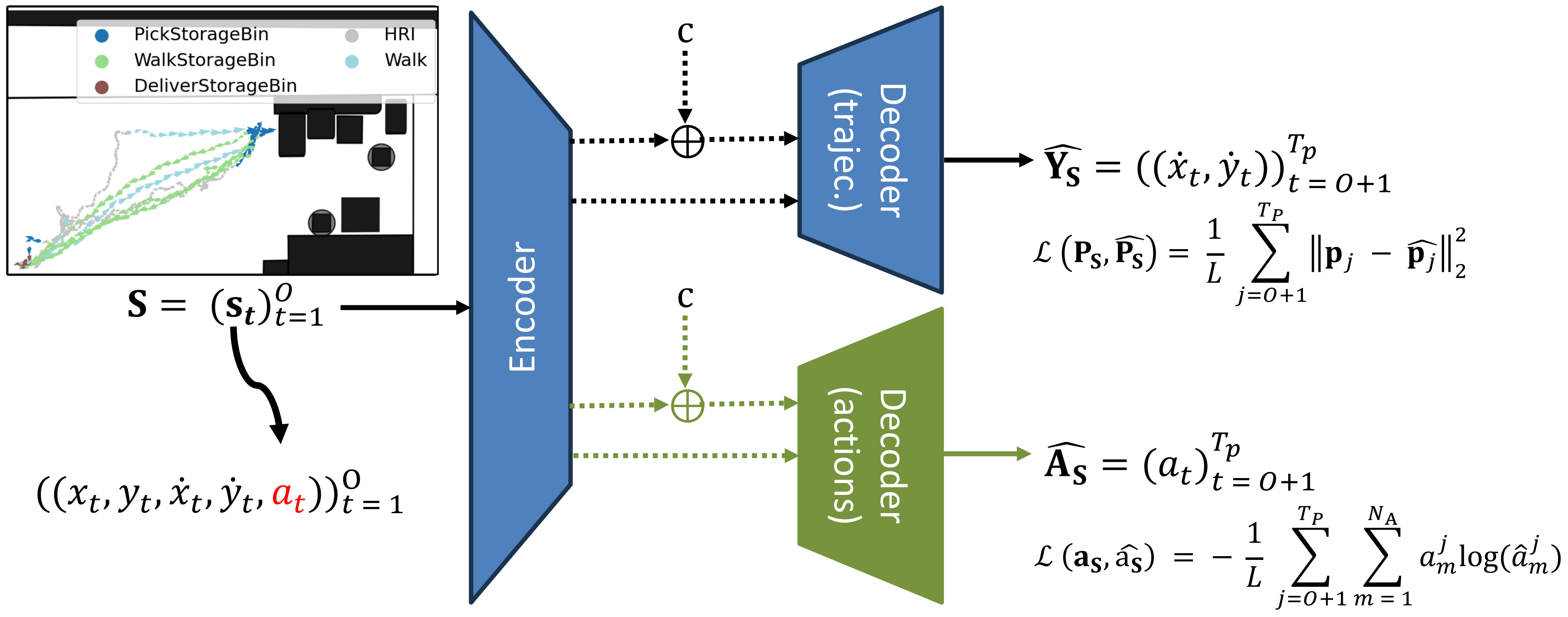


[3] **T. R. de Almeida**, et al. 2025. THÖR-MAGNI Act: Actions for Human Motion Modeling in Robot-Shared Industrial Spaces. *HRI '25*.

How can frame-based actions improve trajectory prediction?

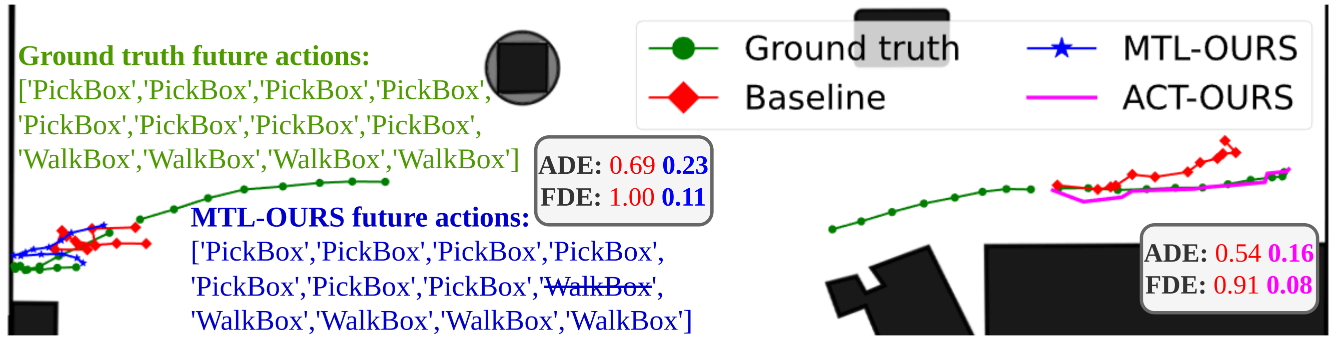


THÖR-MAGNI *Act* data collection



Model	Agent Class	Actions Class	ADE FDE	# Params (K)
BASELINE			0.71 ± 0.03 1.37 ± 0.05	36.7
OURS	✓		0.68 ± 0.03 1.30 ± 0.07	38.1
		✓	0.69 ± 0.03 1.31 ± 0.07	37.3
	✓	✓	$\mathbf{0.67 \pm 0.03}$ $\mathbf{1.28 \pm 0.07}$	38.7

Model	Agent Class	Actions Class	ADE FDE ACC F1	# Params (K)
BASELINE			0.71 ± 0.03 1.37 ± 0.05 $\mathbf{0.85 \pm 0.01}$ $\mathbf{0.85 \pm 0.01}$	36.7 + 42.6
OURS	✓		$\mathbf{0.68 \pm 0.04}$ $\mathbf{1.29 \pm 0.08}$ 0.62 ± 0.02 0.61 ± 0.02	46.3
		✓	0.70 ± 0.03 1.33 ± 0.07 0.83 ± 0.01 0.83 ± 0.01	43.3
	✓	✓	0.70 ± 0.04 1.32 ± 0.08 $\mathbf{0.85 \pm 0.01}$ $\mathbf{0.85 \pm 0.01}$	46.8



Previously...

RQ3. How can frame-based actions improve trajectory prediction?

C3. THÖR-MAGNI *Act*, an extension of THÖR-MAGNI, including frame-based actions to augment the state representation of prediction approaches. Extension of previous prediction methods to model sequences of actions.

Learning Outcome
Actions can enhance trajectory prediction by mitigating some of the ambiguity present in observable classes.

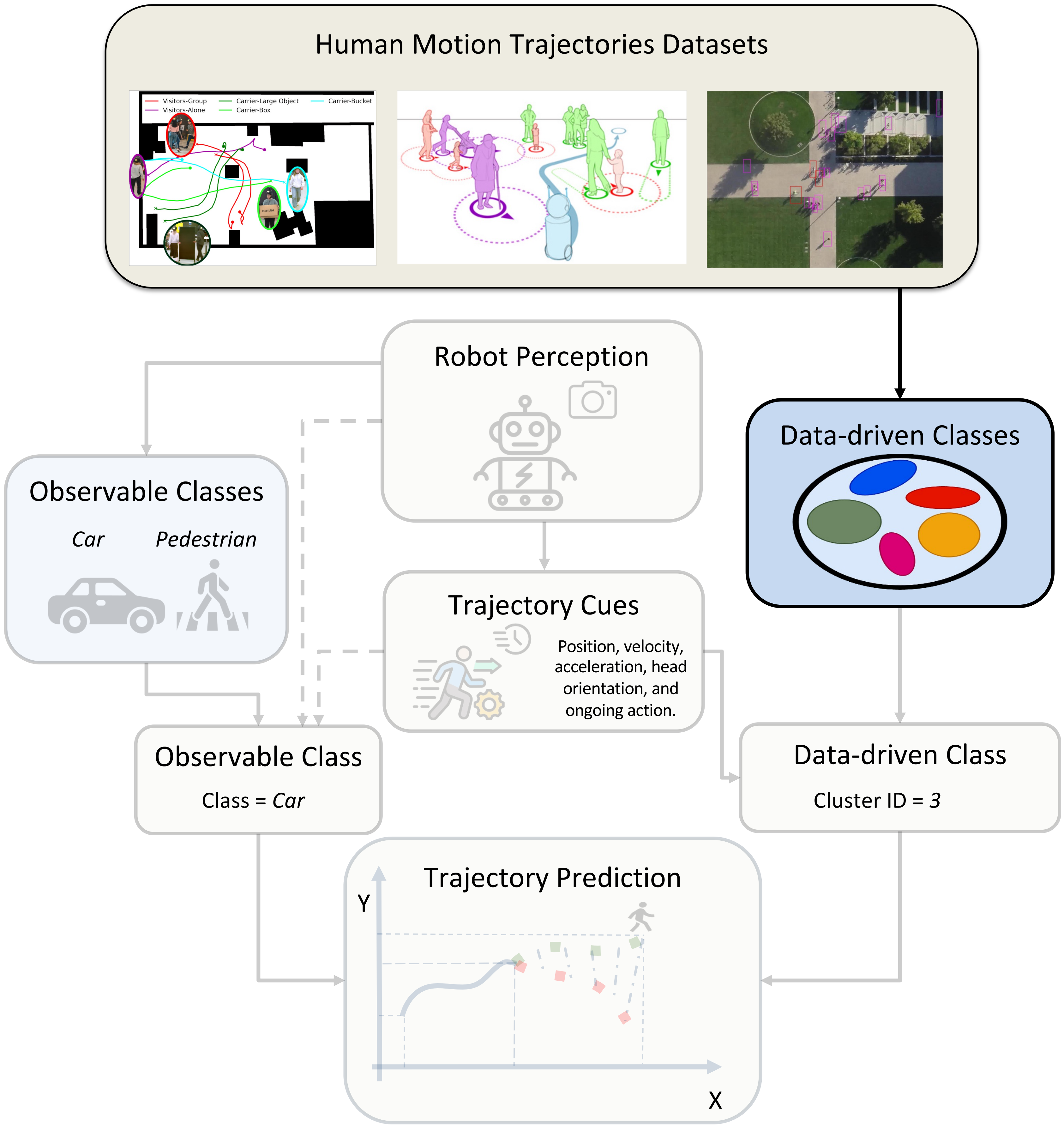
Next.

The observable classes hypothesis is simple: agents that belong to the same trajectory class within a certain predefined framework of abstraction *should* also move similarly. Frame-based actions help mitigating some of the ambiguity associated with observable classes.

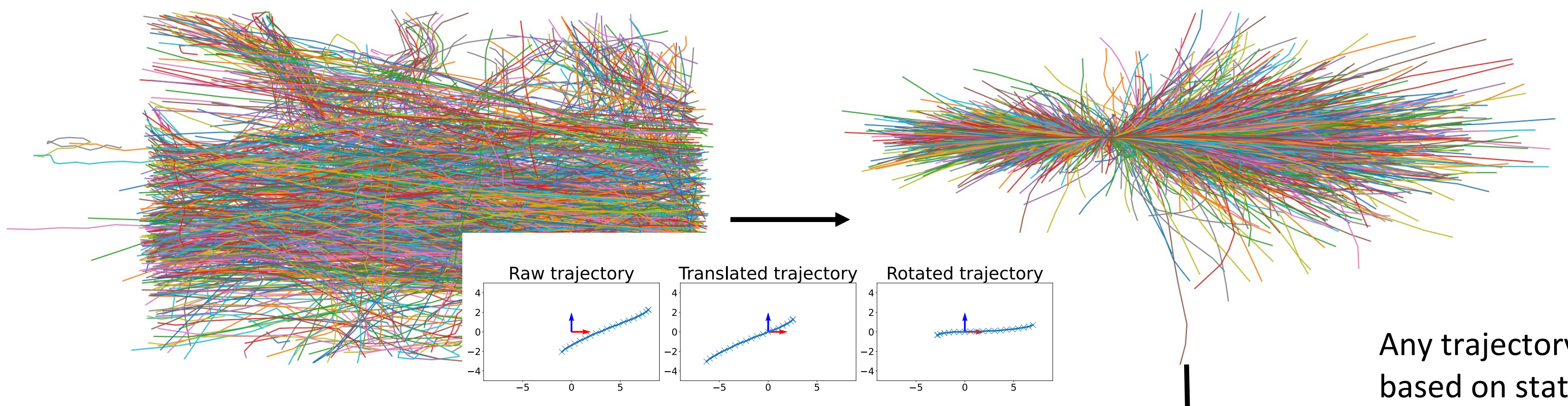
However, both rely on perception and downstream detection methods that can be erroneous and negatively impact the predictions.

RQ4. How to learn data-driven classes for trajectory prediction?

How to learn data-driven classes for trajectory prediction?

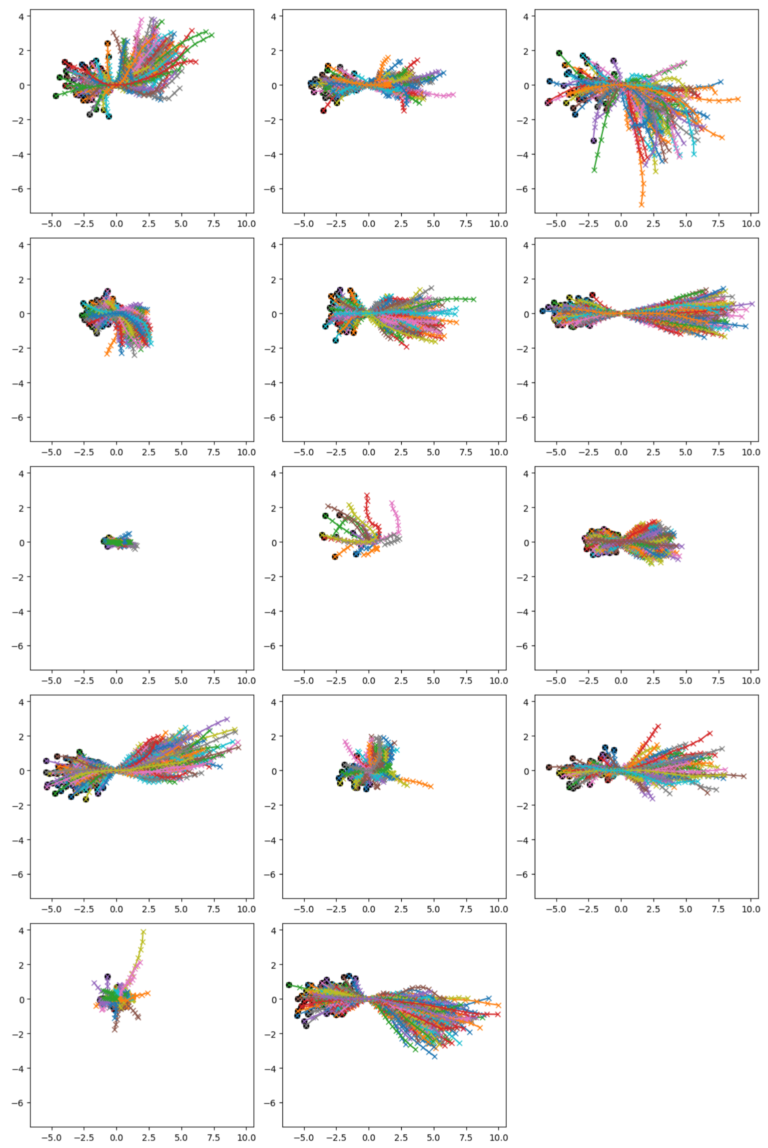


Data-driven classes



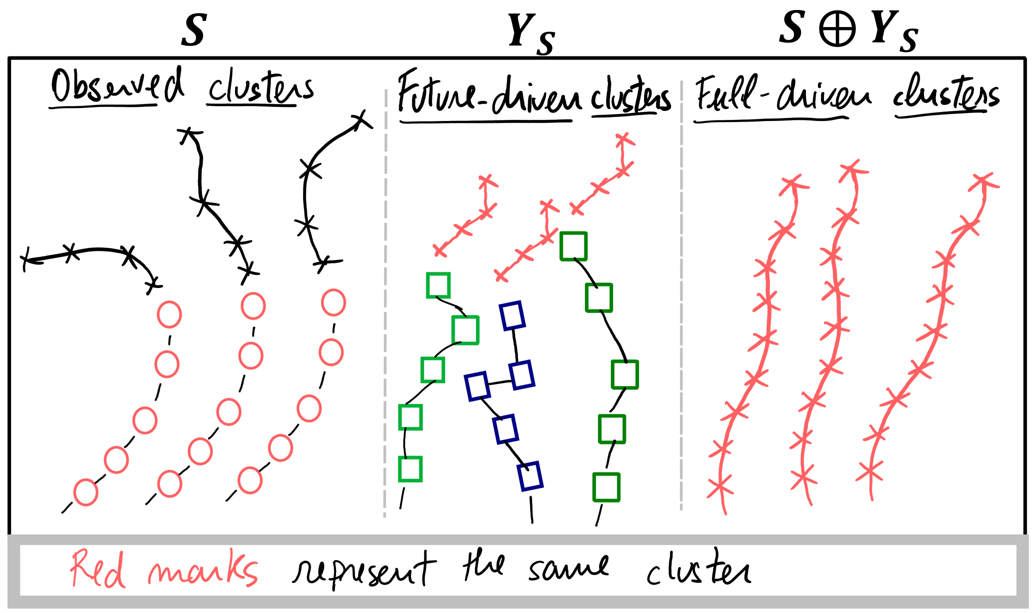
Any trajectory feature based on statistics:

- Velocity
- Acceleration
- Linearity
- Heading
- PCA



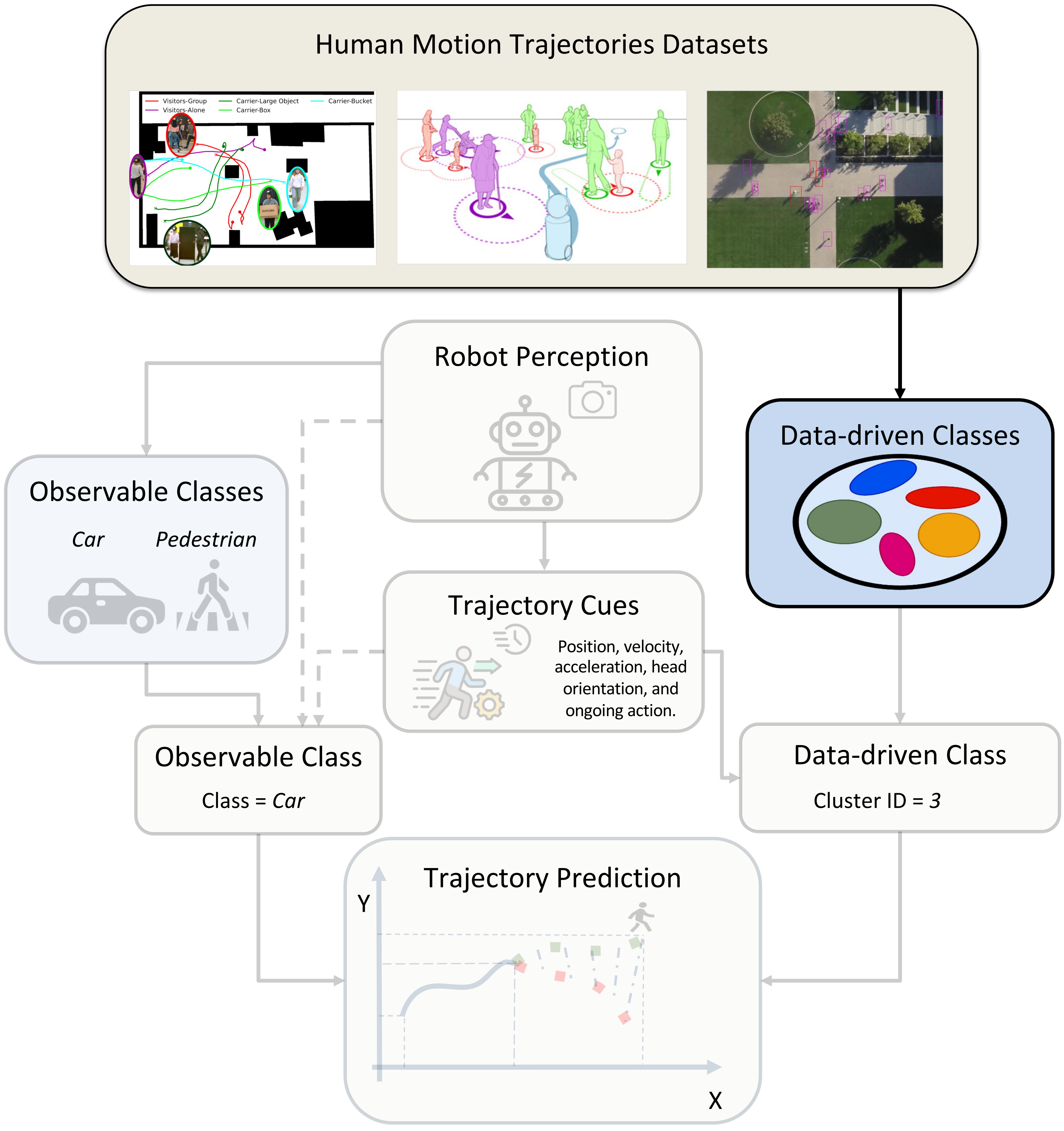
Feature Extraction

clustering



$\boldsymbol{S}$ is the observation part of the trajectory dataset
$\boldsymbol{Y}$ is the future part of the trajectory dataset

How to learn data-driven classes for trajectory prediction?

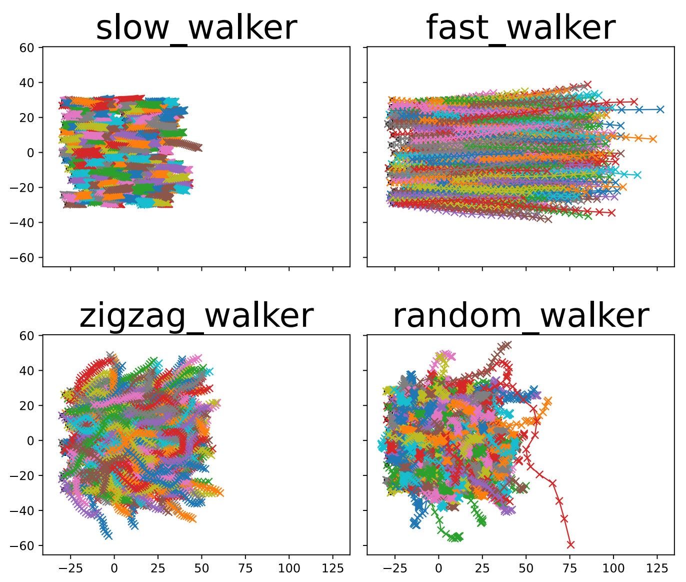


Datasets & Class Proportions:

THÖR-MAGNI dataset and…

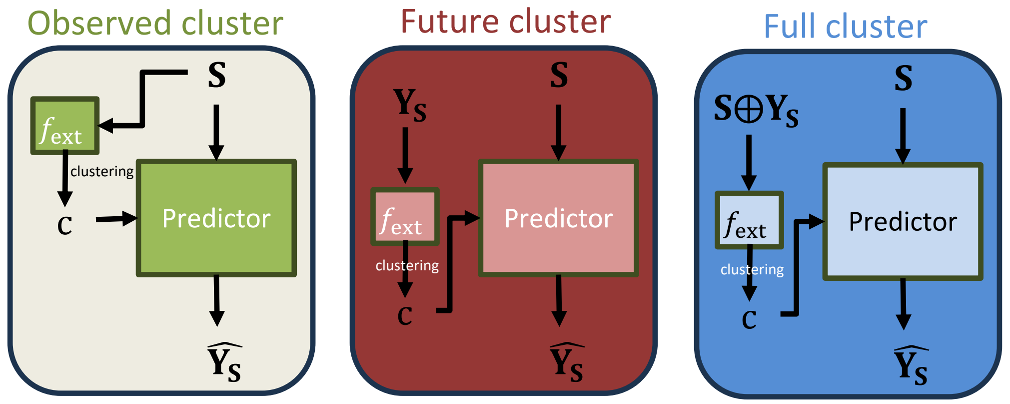
Synthetic Dataset

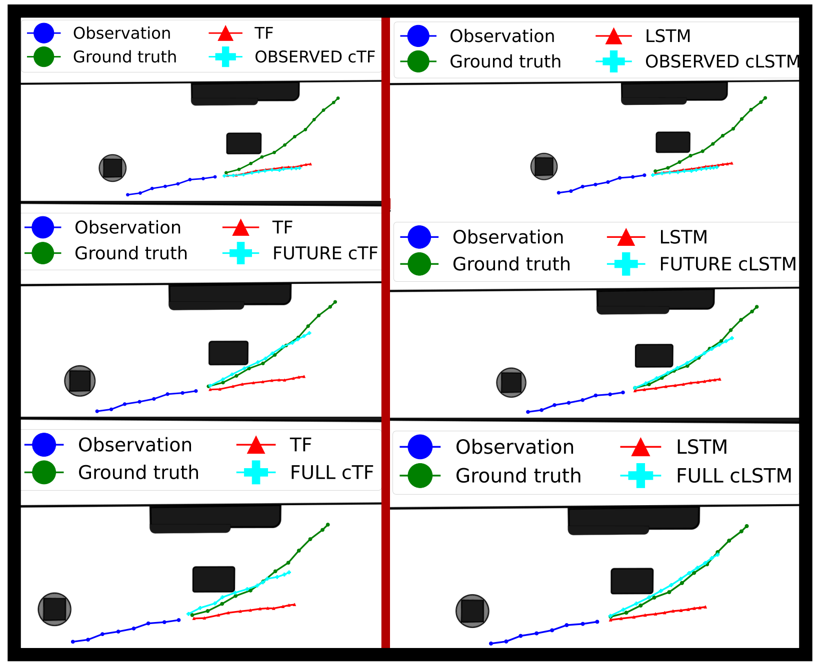
4 observable classes: *slow-walker* (25%), *fast-walker* (25%), zigzag-walker (25%), and *random-walker* (25 %)



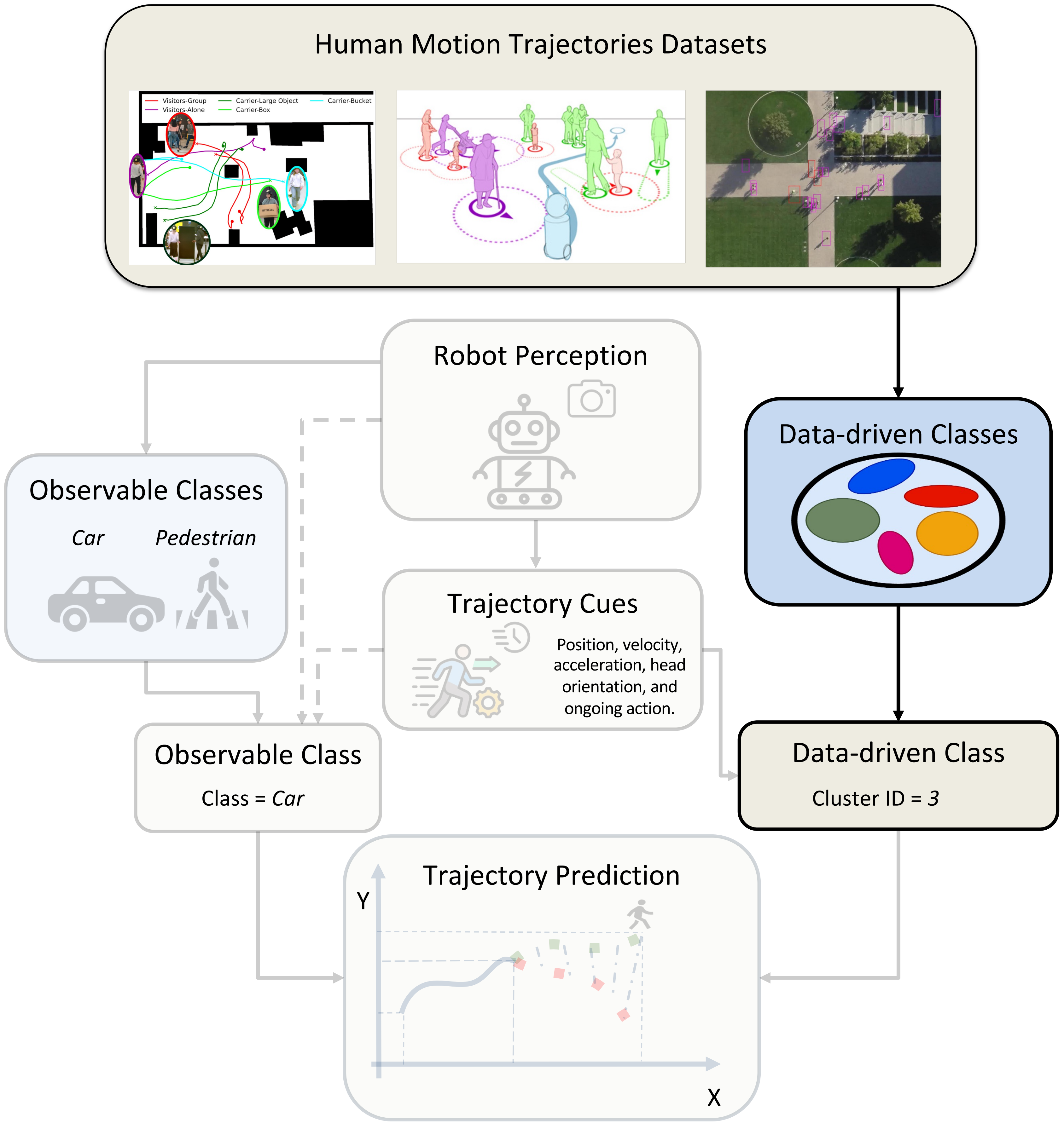
Predictors & Prediction Results:

Model	Model Type	Synthetic	MAGNI
TF	Baseline	2.03 ± 0.08 4.14 ± 0.12	0.89 ± 0.01 1.92 ± 0.03
	Observable class	1.98 ± 0.08 3.93 ± 0.09	0.87 ± 0.02 1.87 ± 0.04
	Observed cluster	2.04 ± 0.05 4.09 ± 0.10	0.89 ± 0.02 1.93 ± 0.04
	Future cluster	$\mathbf{1.80 \pm 0.05}$ $\mathbf{3.61 \pm 0.16}$	$\mathbf{0.65 \pm 0.01}$ $\mathbf{1.47 \pm 0.02}$
	Full cluster	$\mathit{1.89 \pm 0.09}$ $\mathit{3.80 \pm 0.15}$	$\mathit{0.78 \pm 0.01}$ $\mathit{1.71 \pm 0.02}$
LSTM	Baseline	2.04 ± 0.09 4.06 ± 0.12	0.89 ± 0.02 1.92 ± 0.04
	Observable class	1.96 ± 0.09 3.95 ± 0.12	0.87 ± 0.02 1.86 ± 0.04
	Observed cluster	2.03 ± 0.09 4.12 ± 0.13	0.89 ± 0.01 1.92 ± 0.03
	Future cluster	$\mathbf{1.77 \pm 0.07}$ $\mathbf{3.62 \pm 0.11}$	$\mathbf{0.64 \pm 0.01}$ $\mathbf{1.45 \pm 0.02}$
	Full cluster	$\mathit{1.84 \pm 0.06}$ $\mathit{3.76 \pm 0.13}$	$\mathit{0.77 \pm 0.01}$ $\mathit{1.69 \pm 0.02}$





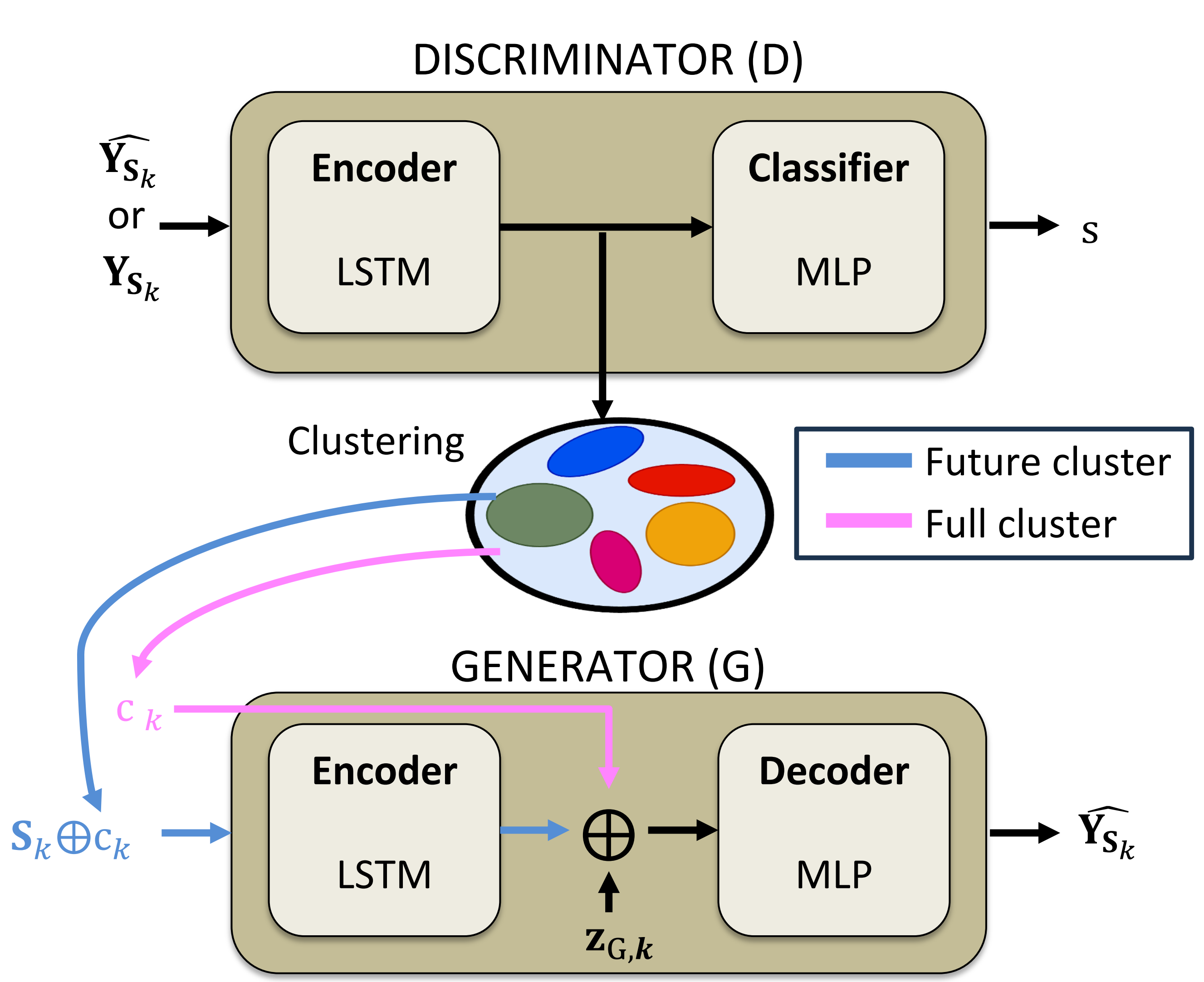
How to learn data-driven classes for trajectory prediction?



Traditional clustering techniques, such as K-means and TS K-means face from limitations:

- K-means consumes 2D data, disregarding sequential patterns that are similar in dynamics but misaligned in time.
- TS K-means uses DTW, which is computationally inefficient.

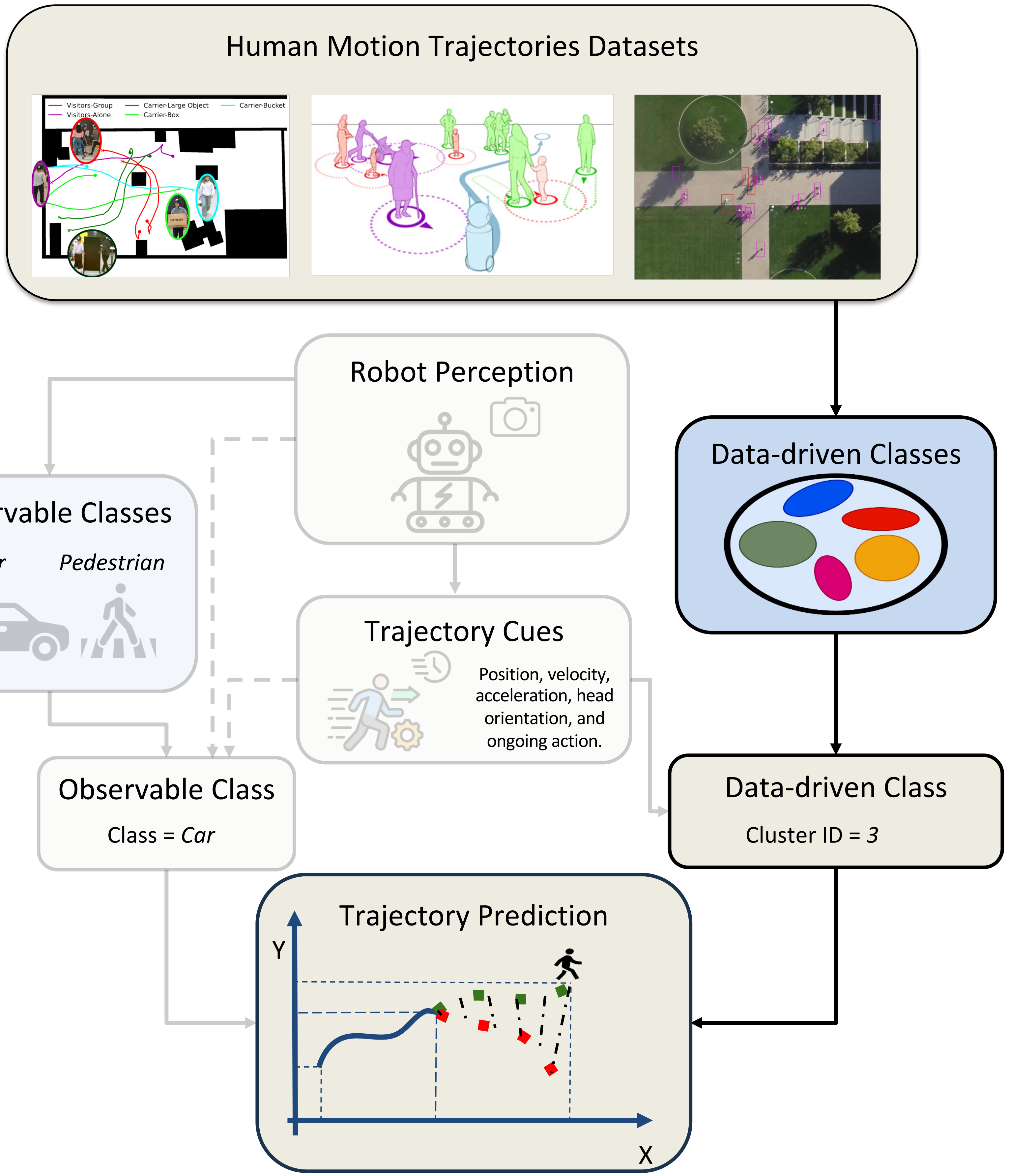
Self-Conditioned GAN (SC GAN)[4]

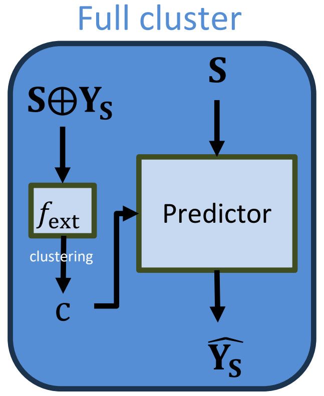


- Clusters in the discriminator's feature space (updated throughout the training).
- Self-learned classes (generator conditioned on clusters' ids).
- Clustering task and generation task linked.

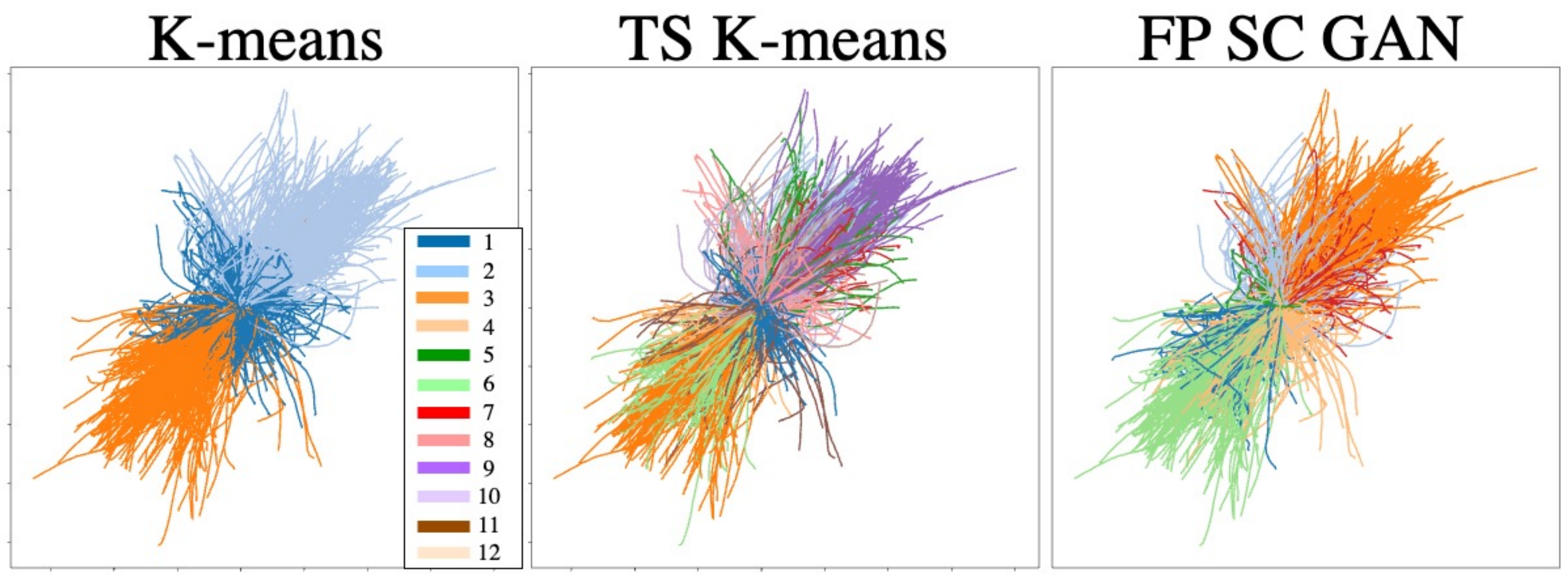
[4] T. Rodrigues de Almeida, et al., "Context-free **Self-Conditioned GAN** for Trajectory Forecasting," *ICMLA '22*.

How to learn data-driven classes for trajectory prediction?

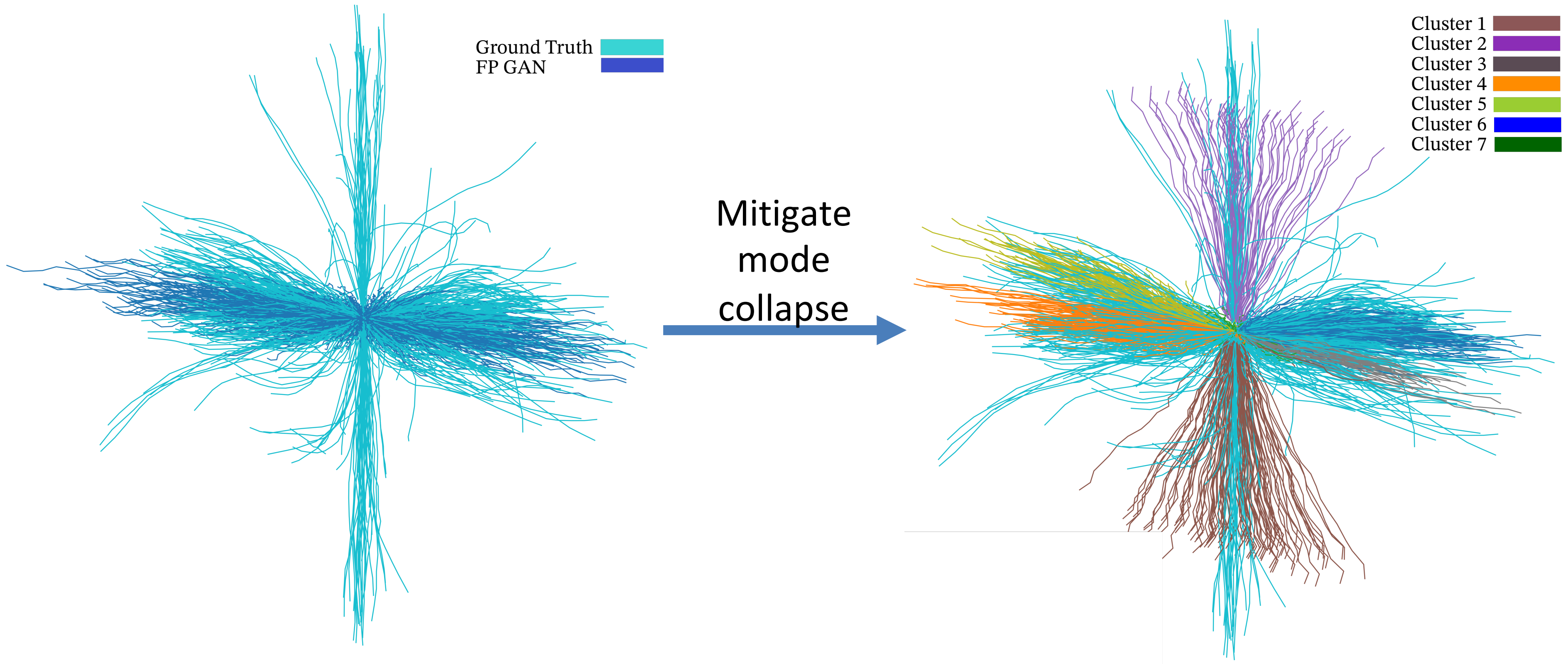




Full Cluster SC GAN - Experiments



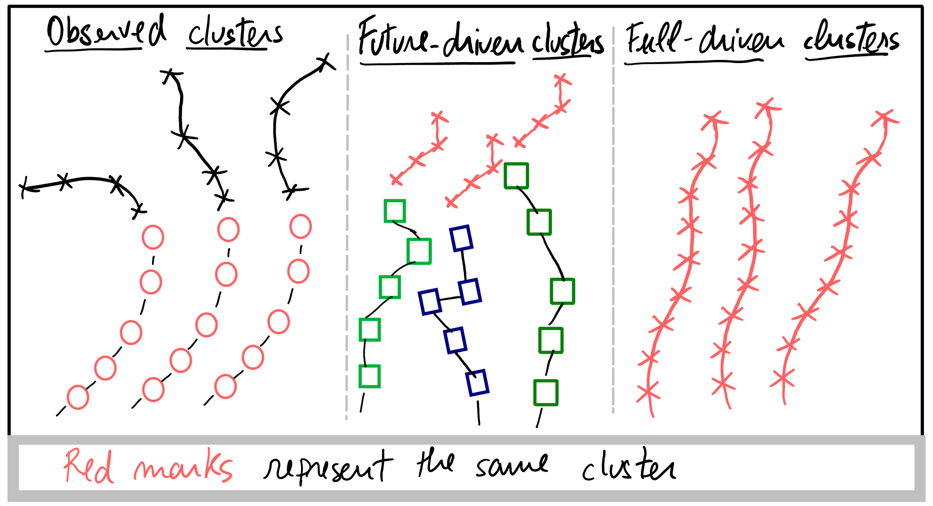
But still why full cluster SC GAN?



Previously…

RQ4. How to learn data-driven classes for trajectory prediction?

C4. A study on the most suitable trajectory segment for defining a data-driven class in trajectory prediction.
We also noted limitations of current clustering methods and proposed SC GAN to address them.

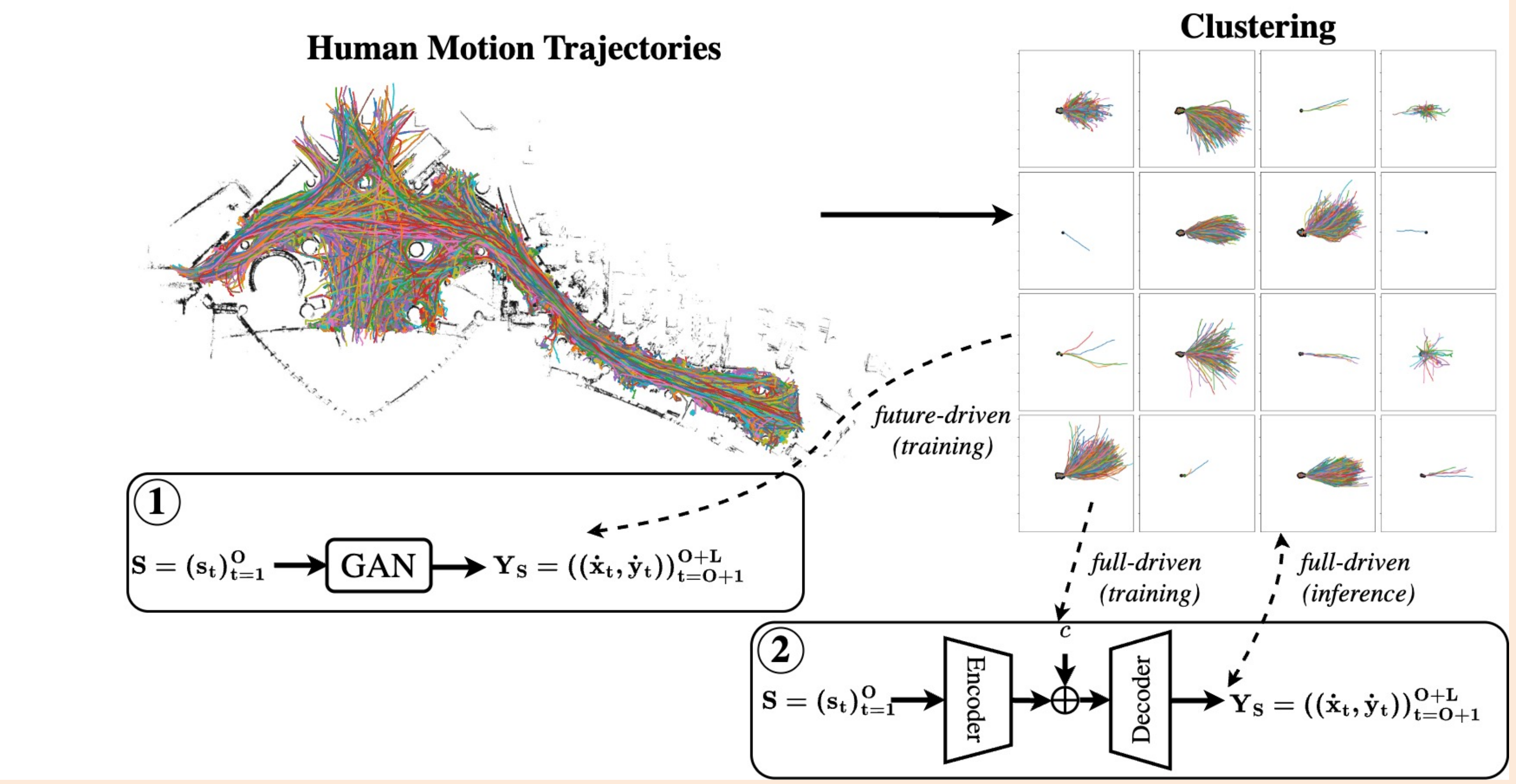


<u>Learning Outcomes</u>

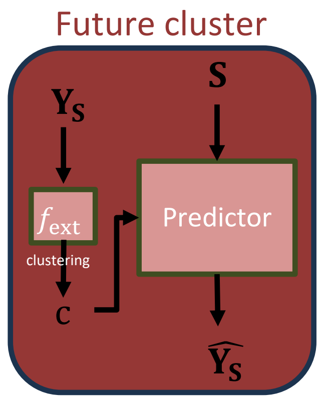
Clusters based on the **future and entire trajectory** are the most **informative** for trajectory prediction.
SC GAN can cluster trajectories in a deep feature space **efficiently**, **overcoming misaligned motion patterns** and **mitigating mode collapse**.

Next.

Now, let's leverage these outcomes to build more accurate trajectory predictors conditioned on data-driven classes.

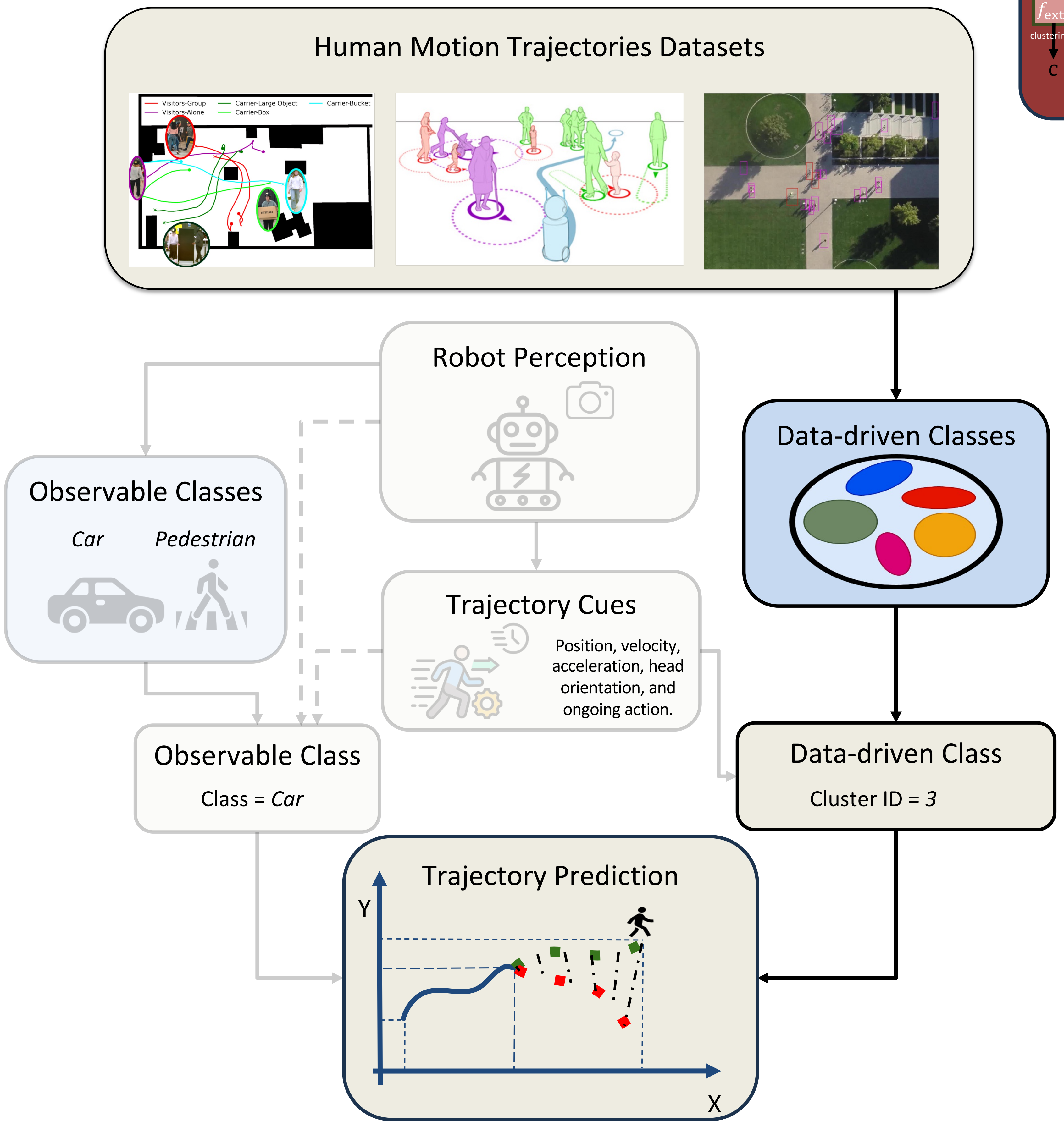


RQ5. How can data-driven classes improve the prediction of trajectories?



How can data-driven classes improve the prediction of trajectories?

Future Cluster SC GAN - Experiments



Training settings of a regular GAN-based forecaster[4]:

Penalize MSE (wL2)
Weighted batch sampler (wB)
Both (wL2 + wB)

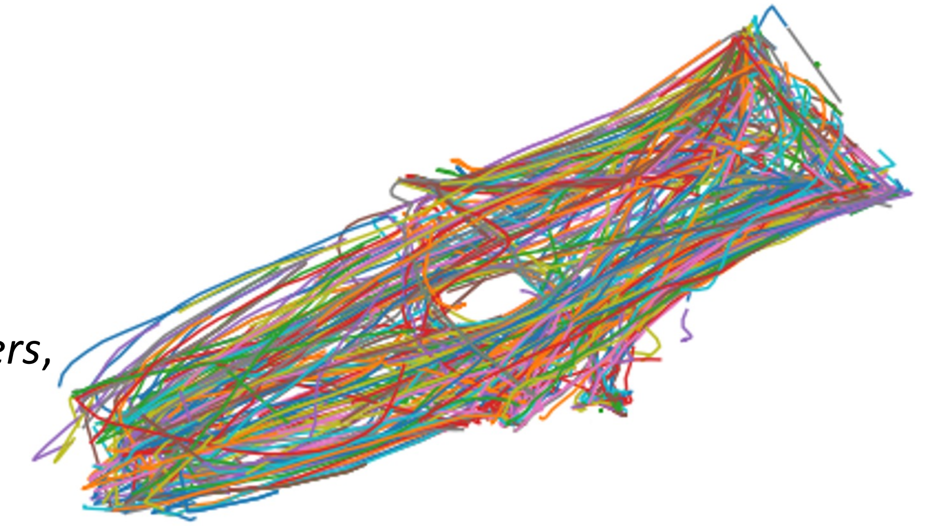
For each trajectory in cluster *i*:

$$\Lambda^i = \lambda_{\text{ADE}} \frac{\text{ADE}^i}{\text{ADE}^{\max}} + \lambda_{\text{FDE}} \frac{\text{FDE}^i}{\text{FDE}^{\max}} + \lambda_{\text{Dist.}} \frac{\#^i}{\#\text{Total}}$$

Datasets & Class Proportions

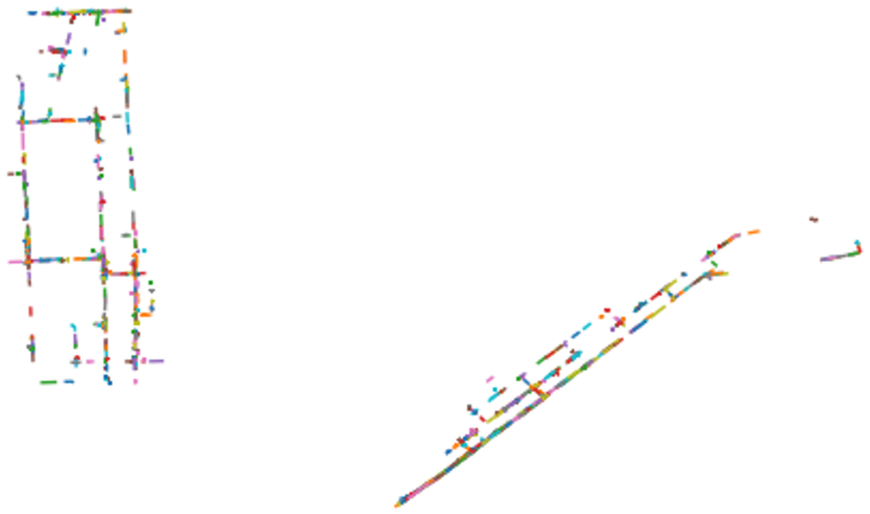
THÖR dataset[6]

3 observable classes: 5 or 6 *visitors*, 2 *workers*, and 1 *inspector*

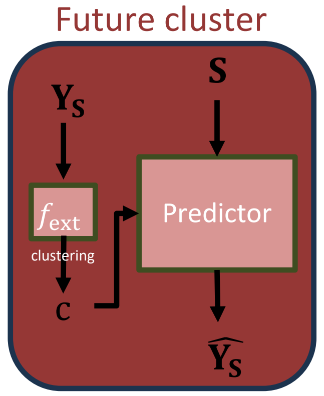


Argoverse dataset[7]

3 observable classes: autonomous vehicles (*av, 45.4%*), regular vehicles (*agents, 45.4%*), and other road agents (*others, 9.2%*)

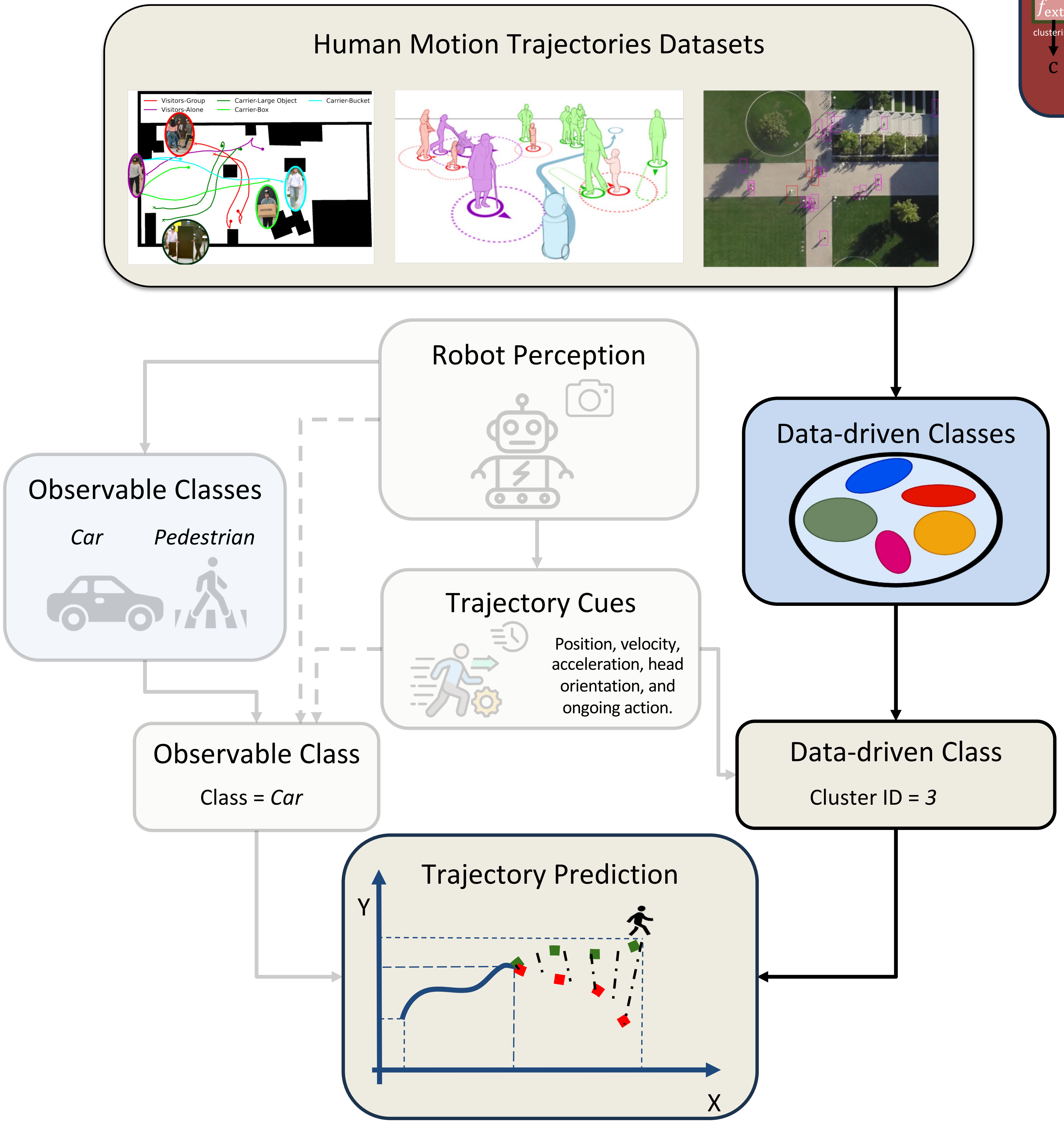


[4] **T. Rodrigues de Almeida**, et al., "Context-free Self-Conditioned GAN for Trajectory Forecasting," *ICMLA '22.*
[6] A. Rudenko, et al. "THÖR: Human-robot navigation data collection and accurate motion trajectories dataset," *RA-L '21.*
[7] M. -F. Chang, et al. , "Argoverse: 3d tracking and forecasting with rich maps". *CVPR '19.*



How can data-driven classes improve the prediction of trajectories?

Future Cluster SC GAN - Experiments



Training settings of a regular GAN-based forecaster:

Penalize MSE (wL2)
Weighted batch sampler (wB)
Both (wL2 + wB)

For each trajectory in cluster i:

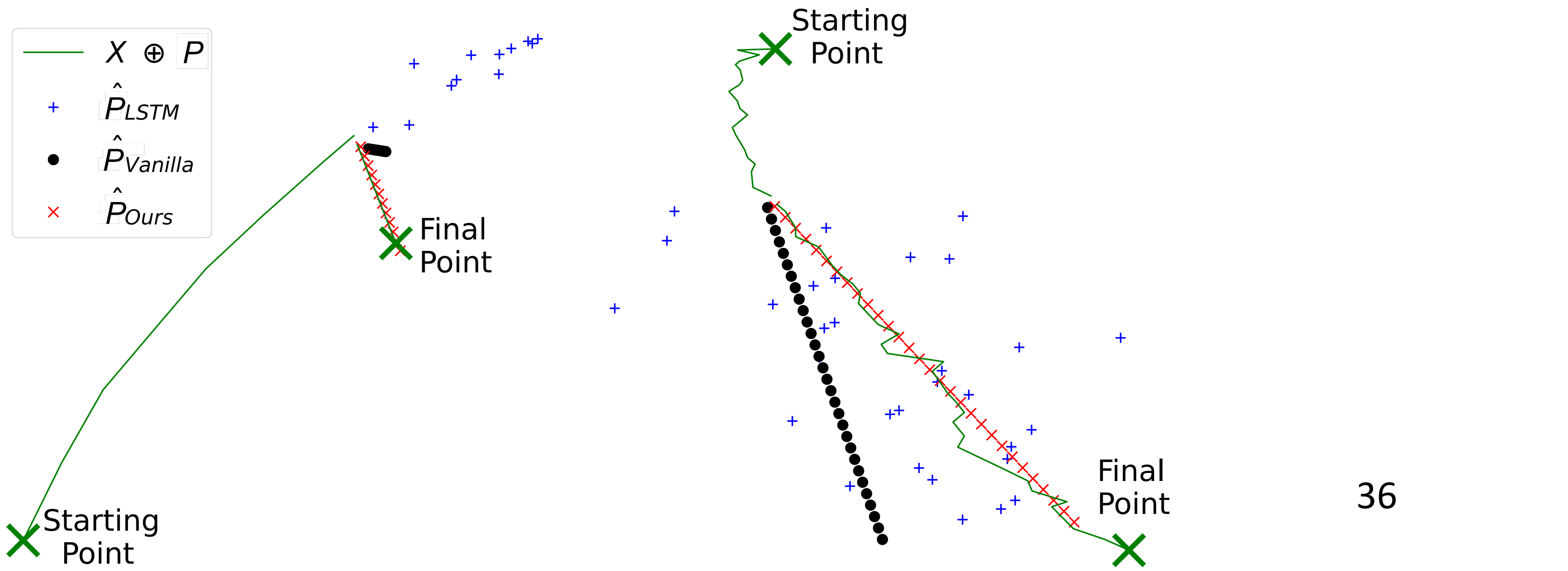
$$\Lambda^i = \lambda_{\text{ADE}} \frac{\text{ADE}^i}{\text{ADE}^{\max}} + \lambda_{\text{FDE}} \frac{\text{FDE}^i}{\text{FDE}^{\max}} + \lambda_{\text{Dist.}} \frac{\#^i}{\#\text{Total}}$$

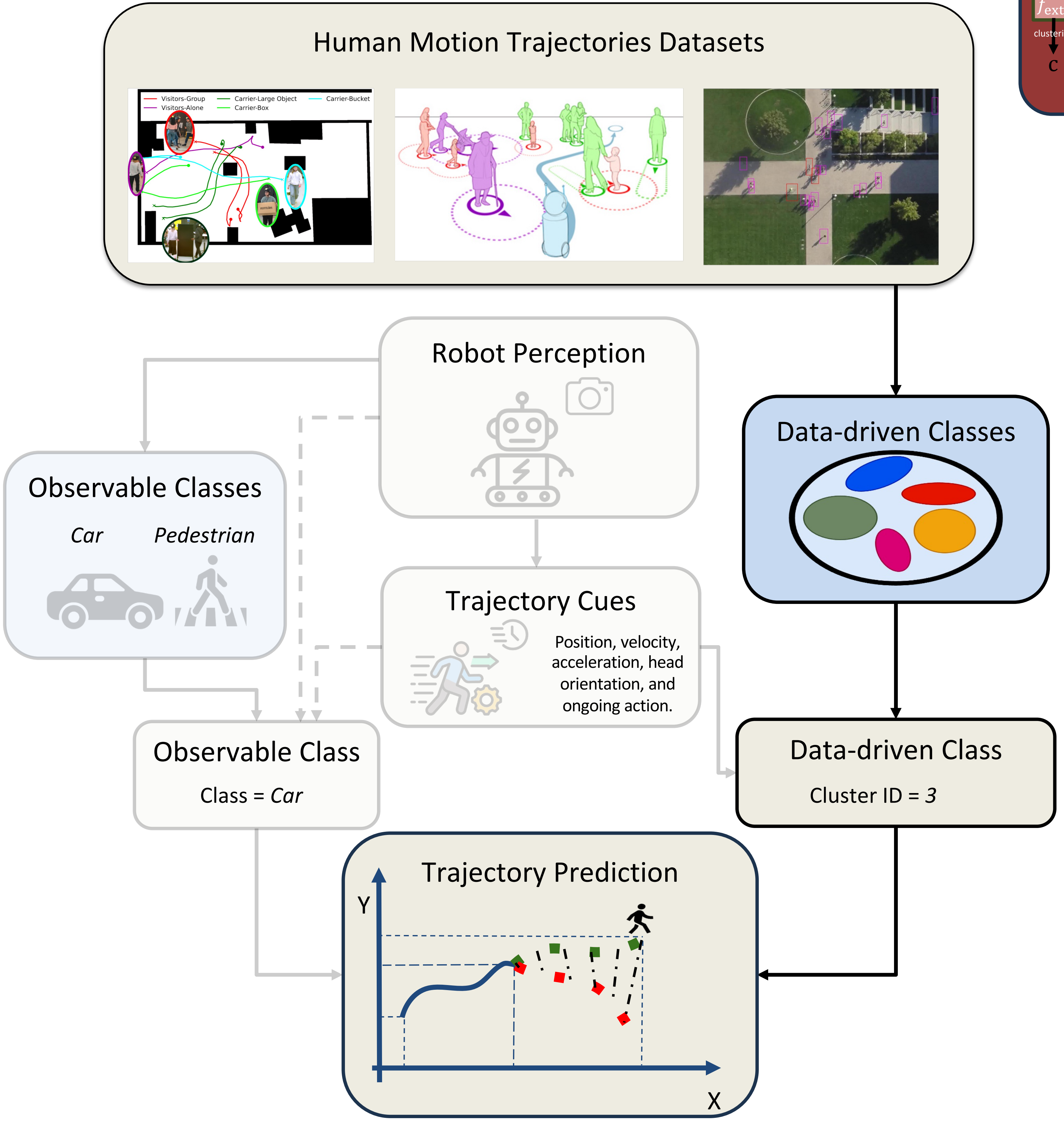
Results (Top-1 ADE/ Top-1 FDE)

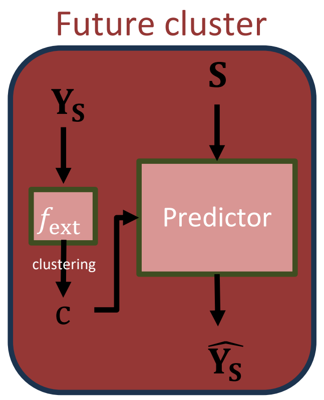
Dataset	Cluster ID (# samples)	Vanilla	Ours wL2	Ours wB	Ours wL2+wB
THÖR	9 (23)	1.12 ± 0.03 2.76 ± 0.08	$\mathbf{1.05 \pm 0.05}$ $\mathbf{2.51 \pm 0.13}$	1.12 ± 0.04 2.81 ± 0.14	$\mathbf{1.04 \pm 0.06}$ $\mathbf{2.51 \pm 0.11}$
	0 (1003)	0.31 ± 0.01 0.40 ± 0.01	0.32 ± 0.01 0.42 ± 0.02	0.32 ± 0.01 0.42 ± 0.02	0.32 ± 0.02 0.42 ± 0.02
Argoverse	10 (16)	7.18 ± 0.18 18.40 ± 0.42	7.11 ± 0.12 18.23 ± 0.30	7.12 ± 0.06 18.28 ± 0.11	$\mathbf{7.05 \pm 0.08}$ $\mathbf{18.13 \pm 0.19}$
	18 (1542)	0.81 ± 0.02 1.10 ± 0.02	0.81 ± 0.01 1.09 ± 0.03	0.81 ± 0.01 1.09 ± 0.03	$\mathbf{0.80 \pm 0.01}$ $\mathbf{1.06 \pm 0.03}$

Outperform on the underrepresented unsupervised classes.

Underrepresented are also the most complex. Therefore, the weight applied to the loss function is a good tradeoff.



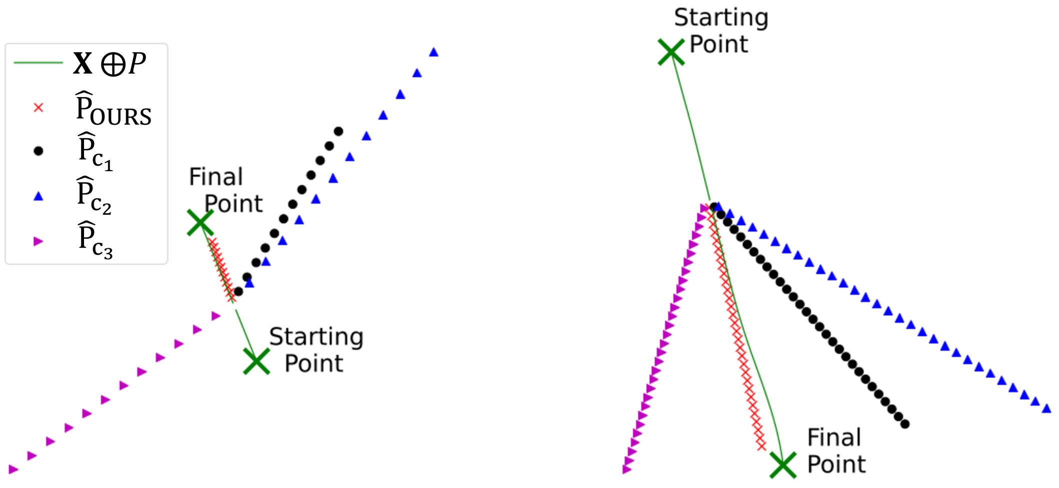




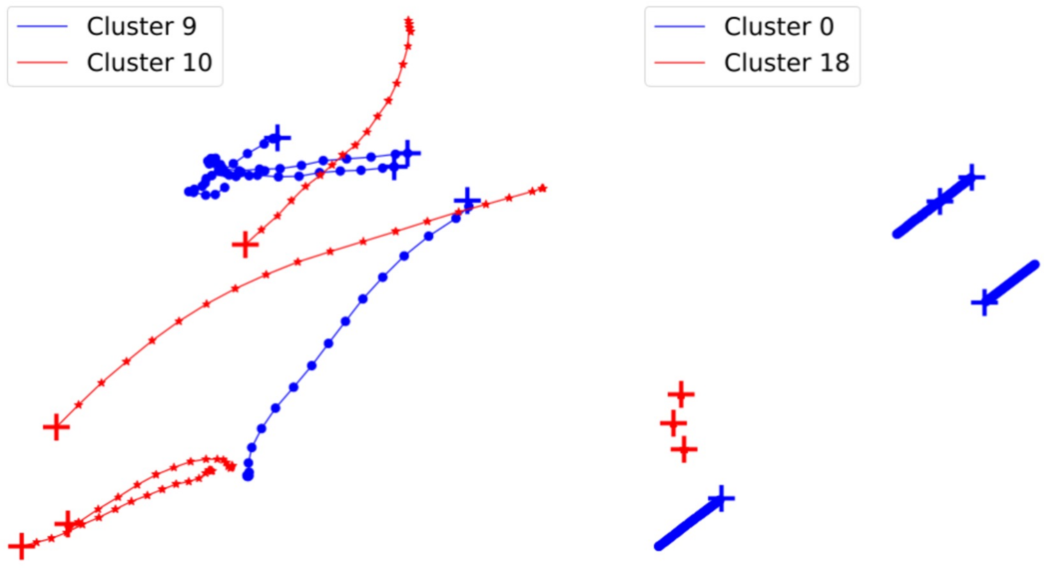
How can data-driven classes improve the prediction of trajectories?

Future Cluster SC GAN – Generation Results

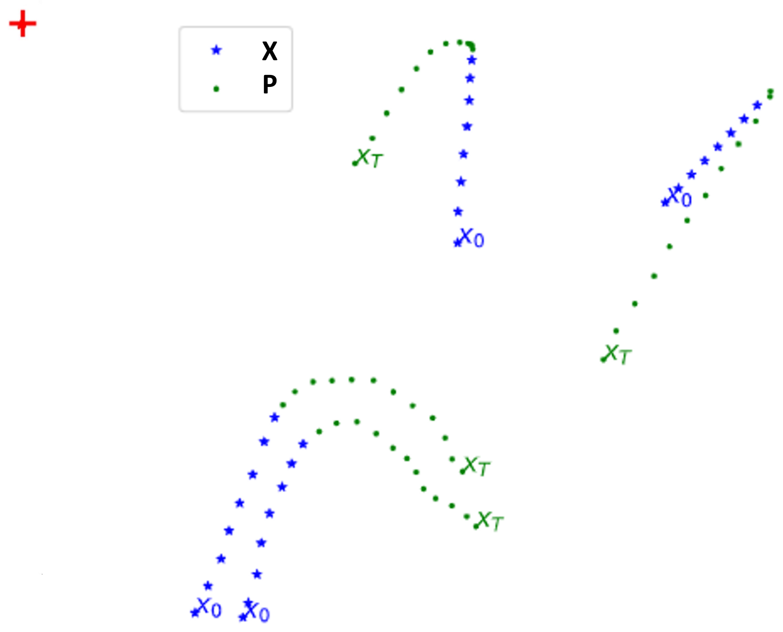
Dataset	Metric	SC GAN vs cGAN
THÖR	ADE	-10%
	FDE	-16%
Argoverse	ADE	-7%
	FDE	-13%



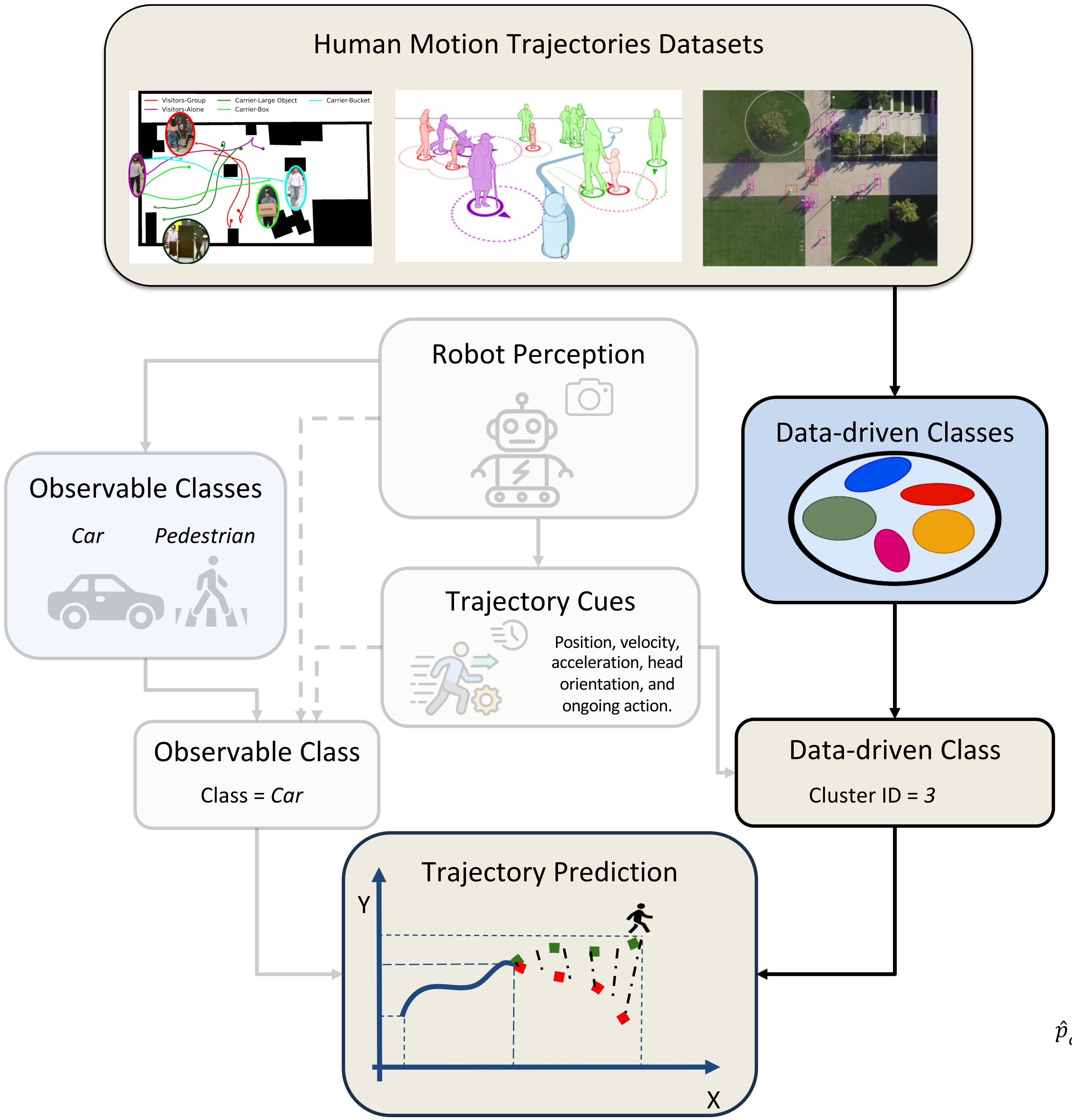
Four clusters from Argoverse



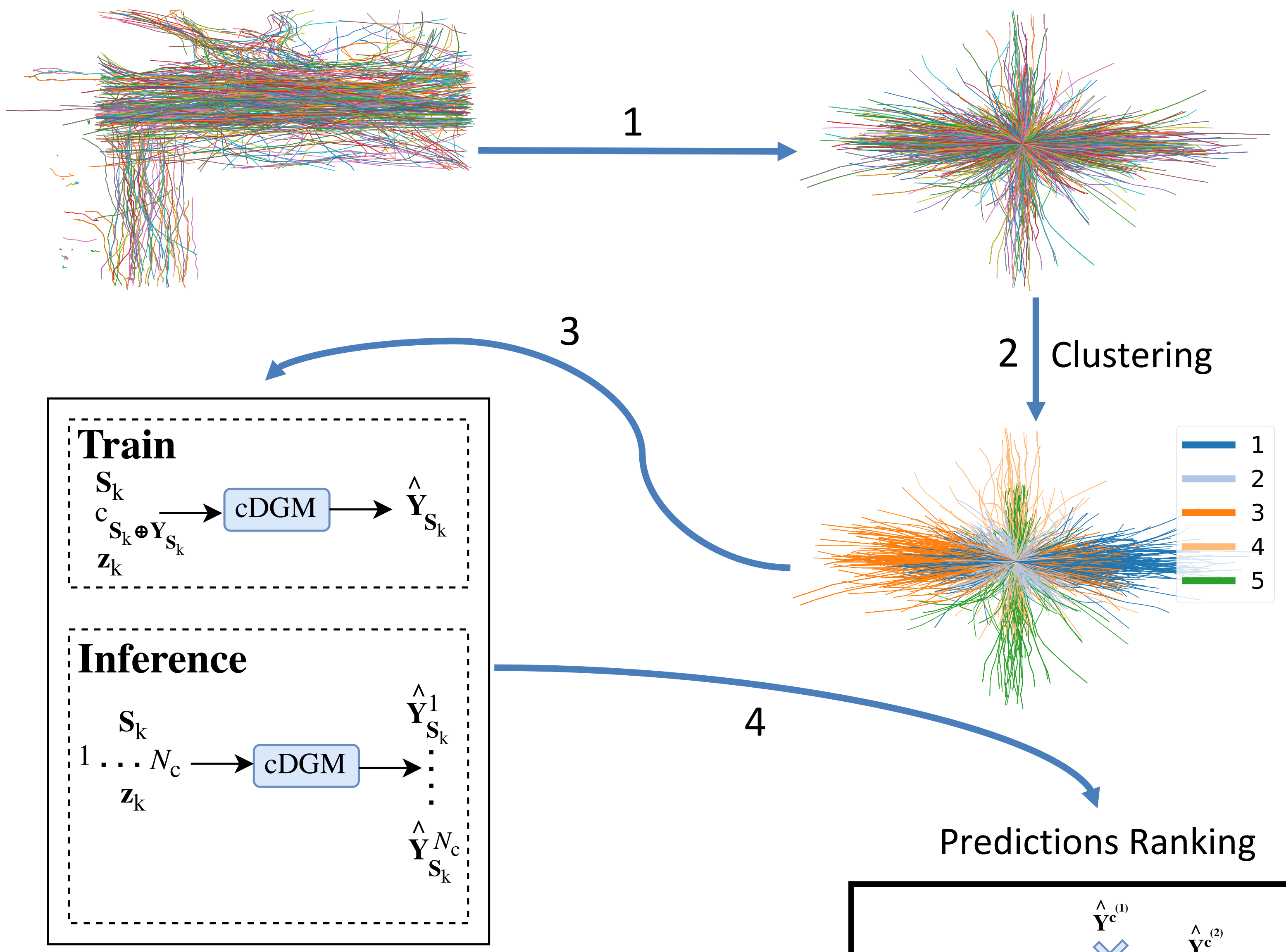
Most challenging cluster in THÖR



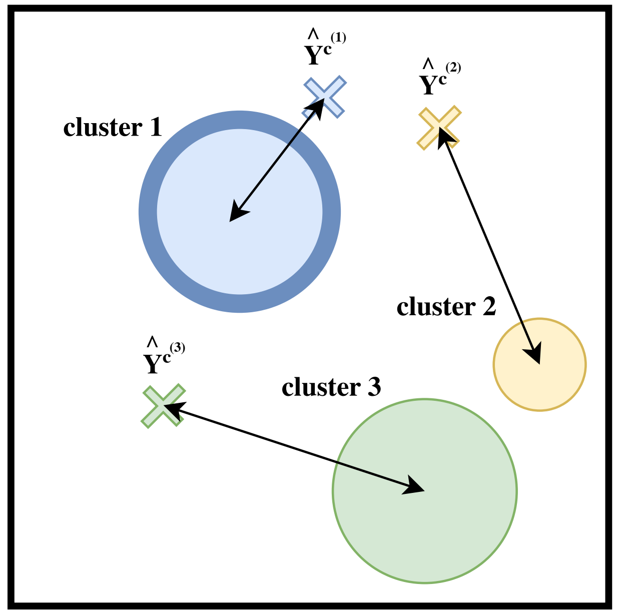
How can data-driven classes improve the prediction of trajectories?



Multi-stage probabilistic framework[5]



Predictions Ranking



$$\hat{p}_{c^{(i)}} = \frac{\exp\left(\frac{1}{m_{c^{(i)}}/\tau}\right)}{\sum_j^{N_c} \exp\left(\frac{1}{m_{c^{(j)}}/\tau}\right)}$$

Centroids (*cent*)

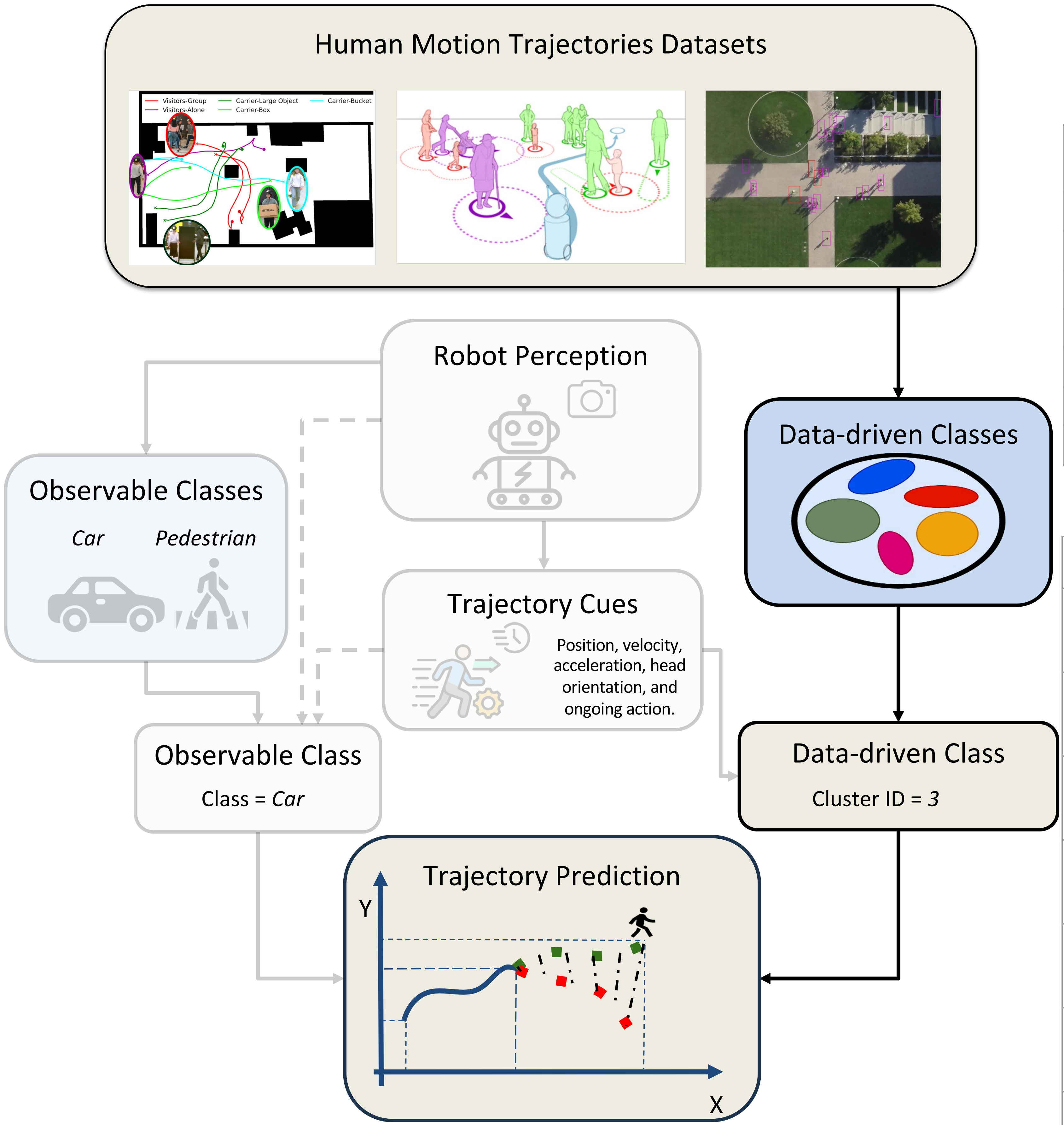
$m_{c^{(i)}}$ is the L2-distance between prediction's embeddings and the centroid of each cluster $c^{(i)}$.

Neighbors (*neigh*)

$m_{c^{(i)}}$ is the average L2-distance between the prediction (sample or embeddings) to the N_{neig} closest neighbors from $c^{(i)}$.

[5] **T. R. de Almeida** et al., "Likely, Light, and Accurate Context-Free Clusters-based Trajectory Prediction". *ITSC '23*.

How can data-driven classes improve the prediction of trajectories?



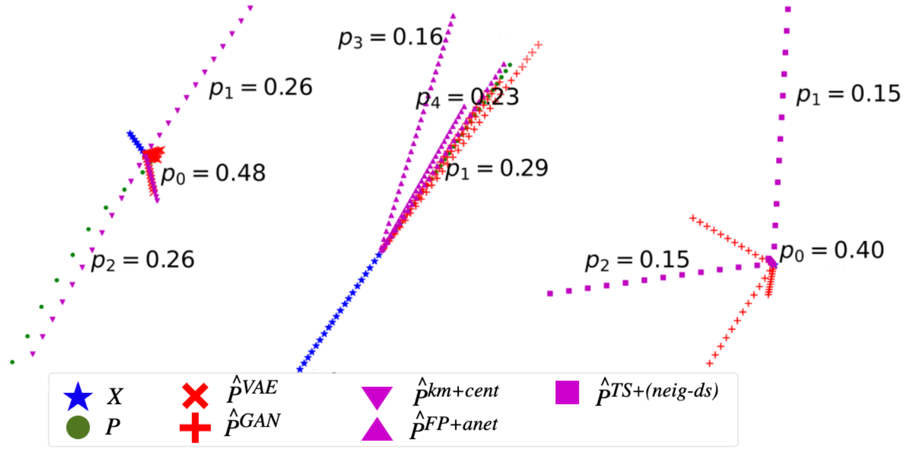
Experiments

Dataset	K-means	TS K-means	FP SC GAN
THÖR	3	12	7
Argoverse	5	5	5
ETH	5	5	6
HOTEL	5	5	5
UNIV	5	5	7
ZARA1	5	5	4
ZARA2	5	5	4

$$\mathrm{DBI} = \frac{1}{K}\sum_{i=1}^{K} \max_{j \neq i}\left(\frac{\sigma_i + \sigma_j}{d(\mu_i, \mu_j)}\right)$$

- σ_i is the average distance from points in cluster *i*.
- d is the Euclidean distance between cluster centroids.

THÖR Argoverse ZARA2



Top-3 ADE/FDE

Model	GAN	GAN-OURS	VAE	VAE-OURS
THÖR	0.57 ± 0.01 1.04 ± 0.03	$\mathbf{0.53 \pm 0.01}$ $\mathbf{0.84 \pm 0.03}$	0.62 ± 0.02 1.05 ± 0.05	$\mathbf{0.56 \pm 0.01}$ $\mathbf{0.89 \pm 0.02}$
Argoverse	1.62 ± 0.07 2.81 ± 0.14	$\mathbf{1.56 \pm 0.02}$ $\mathbf{2.69 \pm 0.02}$	1.96 ± 0.02 3.44 ± 0.06	$\mathbf{1.62 \pm 0.02}$ $\mathbf{2.82 \pm 0.04}$
ETH	0.84 ± 0.03 1.64 ± 0.06	$\mathbf{0.77 \pm 0.04}$ $\mathbf{1.60 \pm 0.11}$	0.94 ± 0.02 1.84 ± 0.04	$\mathbf{0.82 \pm 0.02}$ $\mathbf{1.60 \pm 0.04}$
HOTEL	0.87 ± 0.07 1.64 ± 0.12	$\mathbf{0.81 \pm 0.10}$ $\mathbf{1.46 \pm 0.11}$	1.08 ± 0.04 1.95 ± 0.06	$\mathbf{0.97 \pm 0.08}$ $\mathbf{1.72 \pm 0.14}$
UNIV	0.56 ± 0.01 1.08 ± 0.02	$\mathbf{0.51 \pm 0.01}$ $\mathbf{0.98 \pm 0.03}$	0.61 ± 0.01 1.16 ± 0.01	$\mathbf{0.56 \pm 0.01}$ $\mathbf{1.06 \pm 0.02}$
ZARA1	0.43 ± 0.02 0.82 ± 0.07	$\mathbf{0.37 \pm 0.01}$ $\mathbf{0.72 \pm 0.03}$	0.48 ± 0.01 0.98 ± 0.04	$\mathbf{0.44 \pm 0.01}$ $\mathbf{0.91 \pm 0.03}$
ZARA2	0.46 ± 0.01 0.81 ± 0.05	$\mathbf{0.40 \pm 0.01}$ $\mathbf{0.65 \pm 0.03}$	0.49 ± 0.01 0.86 ± 0.05	$\mathbf{0.43 \pm 0.01}$ $\mathbf{0.73 \pm 0.01}$

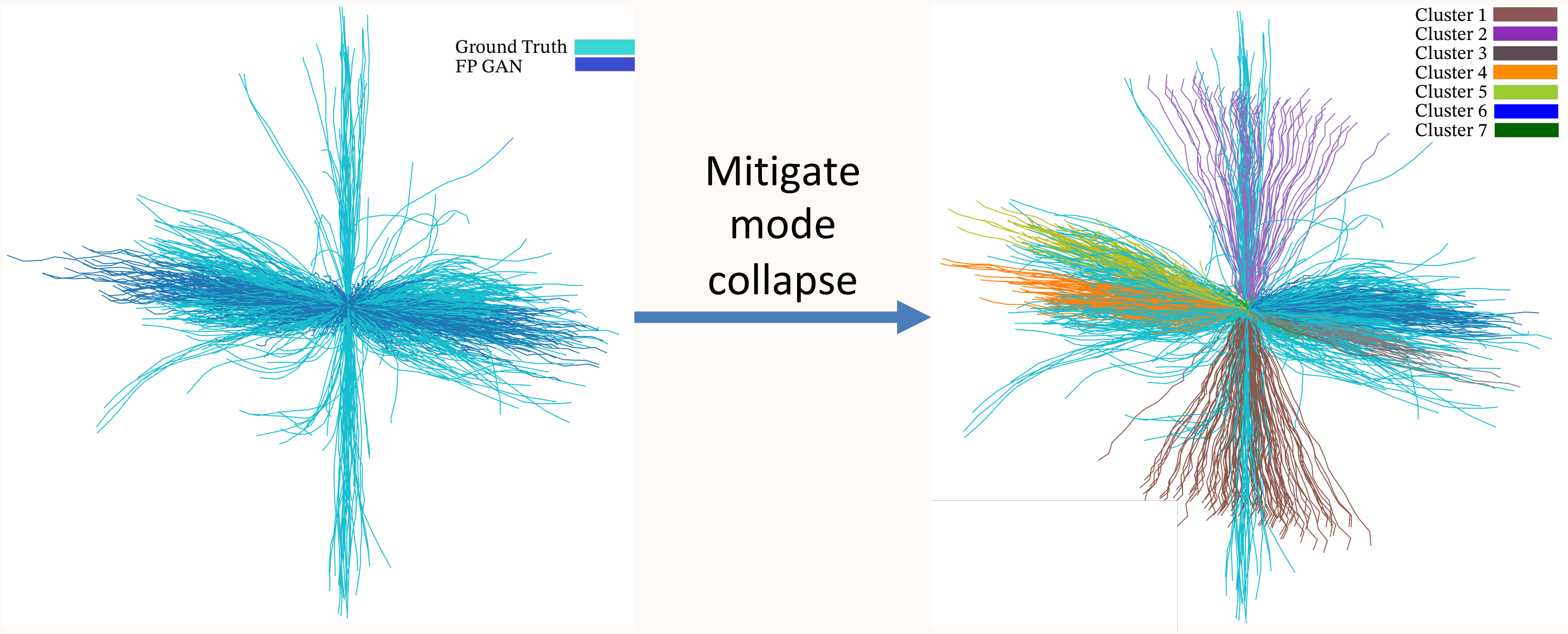
Accuracy (0-1)

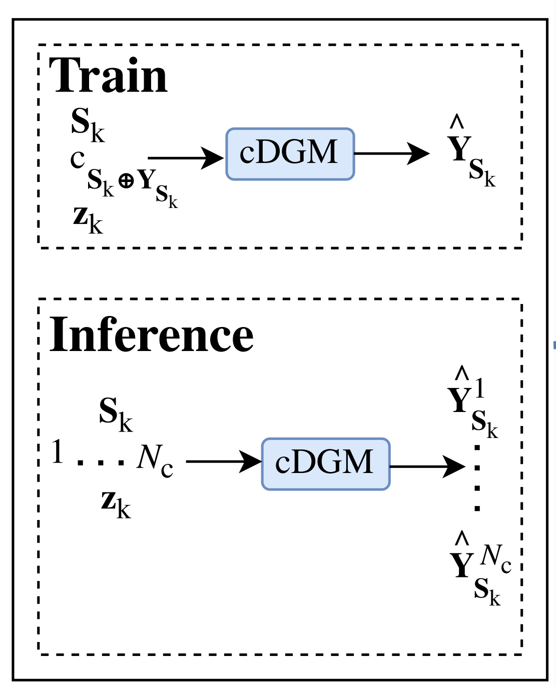
Clustering method	Ranking method	ETH	UNIV
K-means	*cent* *neigh-ds* *Neural Net*	0.95 ± 0.01 $\mathbf{0.98 \pm 0.01}$ 0.93 ± 0.02	$\mathbf{0.84 \pm 0.01}$ 0.81 ± 0.01 0.83 ± 0.01
TS K-means	*cent* *neigh-ds* *Neural Net*	0.95 ± 0.01 $\mathbf{0.98 \pm 0.01}$ 0.95 ± 0.01	$\mathbf{0.87 \pm 0.01}$ 0.82 ± 0.01 0.84 ± 0.01
FP SC GAN	*cent* *neigh-fs* *Neural Net*	0.70 ± 0.11 $\mathbf{0.75 \pm 0.05}$ 0.69 ± 0.08	$\mathbf{0.65 \pm 0.02}$ $\mathbf{0.65 \pm 0.03}$ 0.56 ± 0.03

Previously…

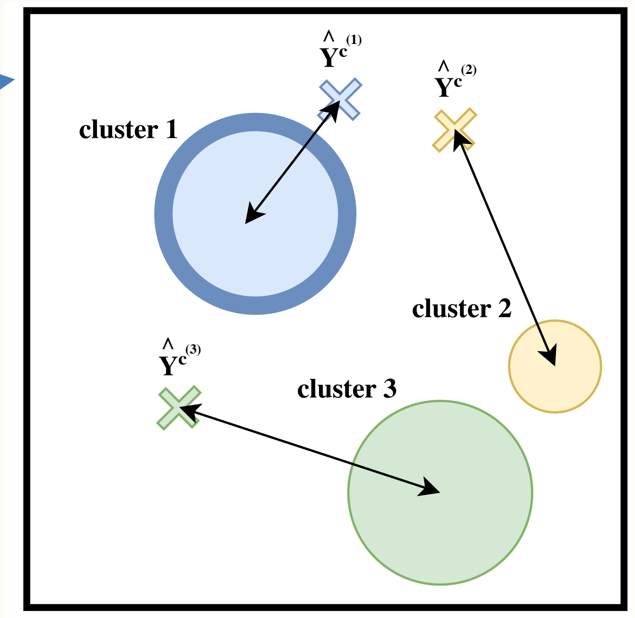
RQ5. How can data-driven classes improve the prediction of trajectories?

C5. SC GAN to mitigate mode collapse in GAN-based forecasters.
Data-driven classes incorporated in a multi-stage approach using novel predictions ranking methods.





Predictions Ranking



<u>Learning Outcomes</u>

SC GAN and the tree training settings effectively reduce mode collapsing by lowering the prediction errors in the clusters associated with the highest prediction errors.

The multi-stage framework is more accurate than the baselines. The corresponding predictions ranking methods are more efficient and sometimes more accurate than neural networks.

4. Contributions Overview.

What could this thesis accomplish?

Actions to augment the state!
Still observable…

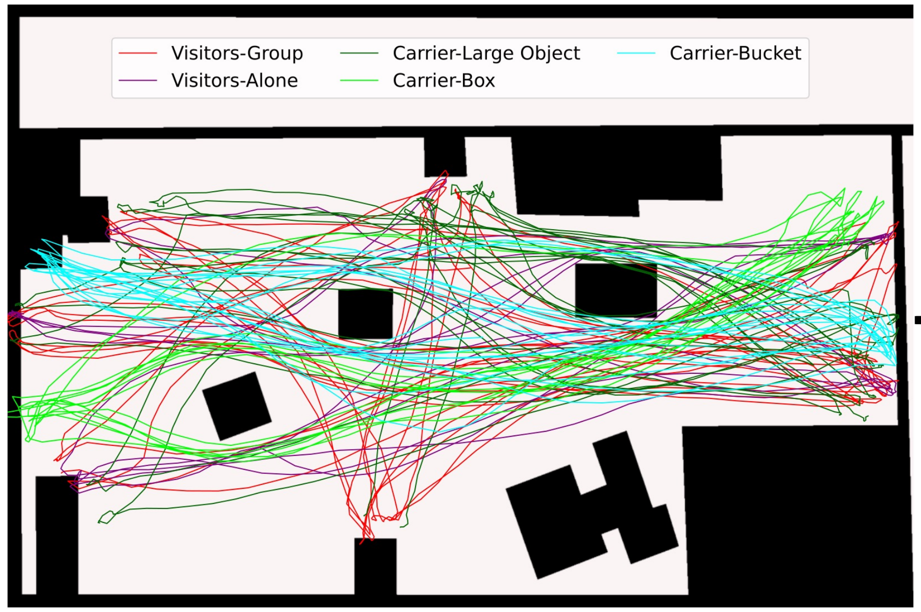
[3] T. R. de Almeida, et al., "**THÖR-MAGNI Act**: Actions for Human Motion Modeling in Robot-Shared Industrial Spaces". *HRI '25*.

Observable classes:

- Can be ambiguous.
- Depend on external perception ("observable").
- Are explainable and promote safe decision making based on human semantics.
- Deep learning methods struggle on class imbalanced and low data regimes.

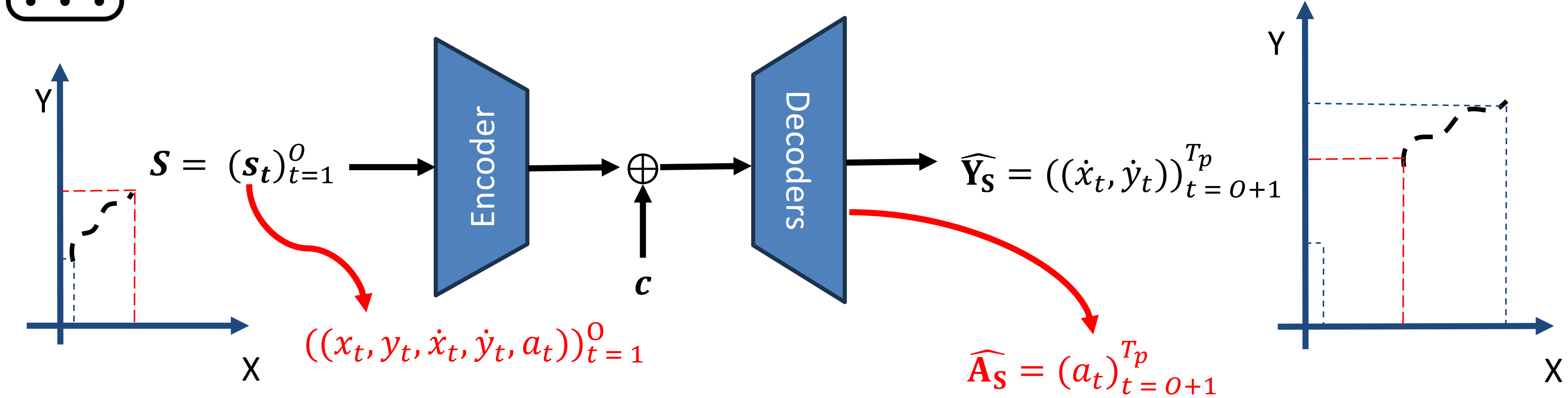
[2] T. R. de Almeida, Zhu Y, et al., "**Trajectory Prediction for Heterogeneous Agents**: A Performance Analysis on Small and Imbalanced Datasets". *RA-L '24*.

To model heterogenous trajectory data, we propose **Trajectory Classes**!

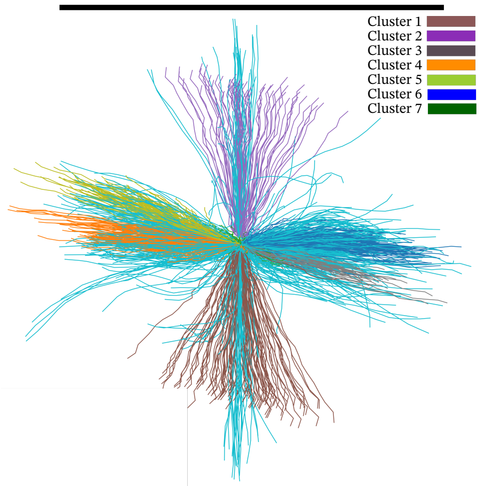


[1] T. Schreiter, T.R. de Almeida, et al., **THÖR-MAGNI**: A large-scale indoor motion capture recording of human movement and robot interaction. *IJRR '24*.

c , **Observable class**: agent type, activity, etc.

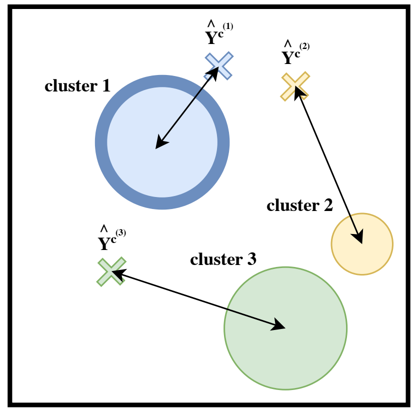


Data-driven classes



[4] T. Rodrigues de Almeida, et al., "Context-free **Self-Conditioned GAN** for Trajectory Forecasting," *ICMLA '22*.

Data-driven class detection



[5] T. R. de Almeida et al., "**Likely, Light, and Accurate Context-Free Clusters-based Trajectory Prediction**". *ITSC '23*.

Centroids (*cent*)

$m_{c^{(i)}}$ is the L2-distance between prediction's embeddings and the centroid of each cluster $c^{(i)}$.

Neighbors (*neigh*)

$m_{c^{(i)}}$ is the average L2-distance between the prediction (sample or embeddings) to the N_{neig} closest neighbors from $c^{(i)}$.

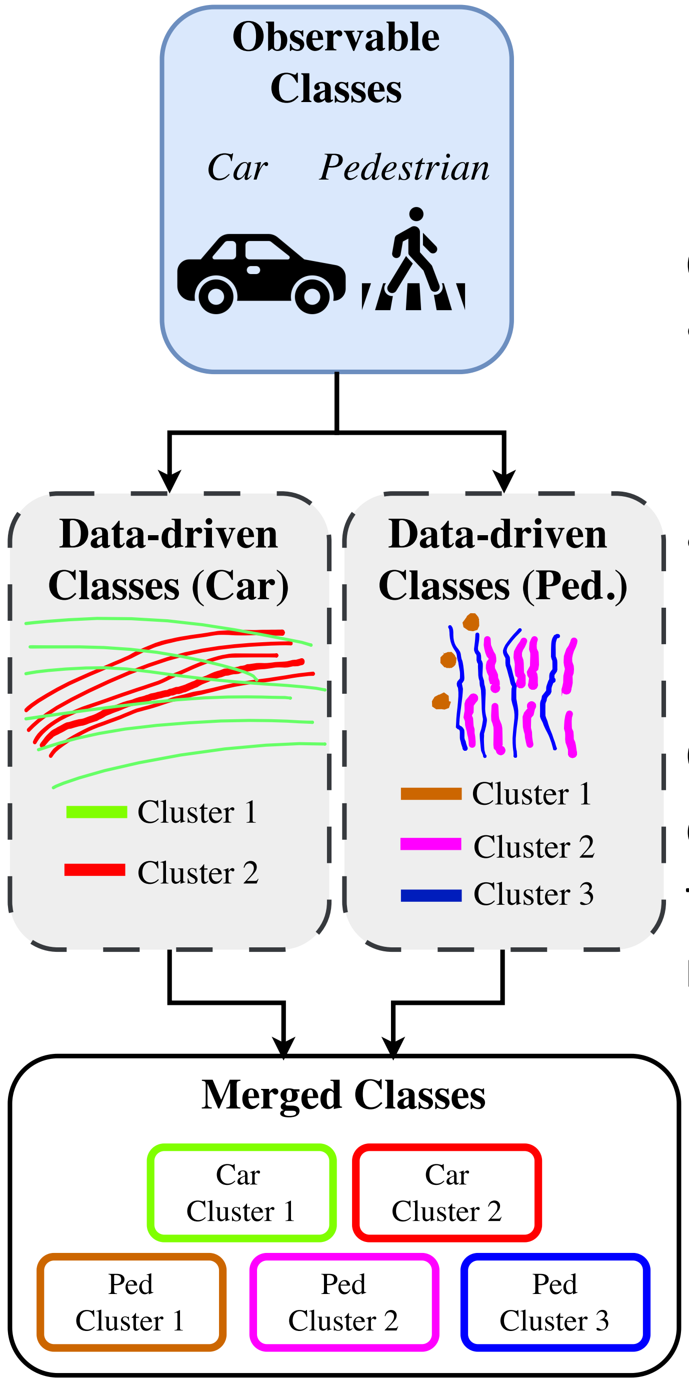
Data-driven class:

- Can overfit to the training data.
- Are less explainable.
- Depend on future trajectory to enhance trajectory prediction.
- Require detection mechanisms.
- Once detected accurately, are very powerful.

5. Future Work.

Where should we go from here?

Hybrid Observable and Data-driven Class Conditioning

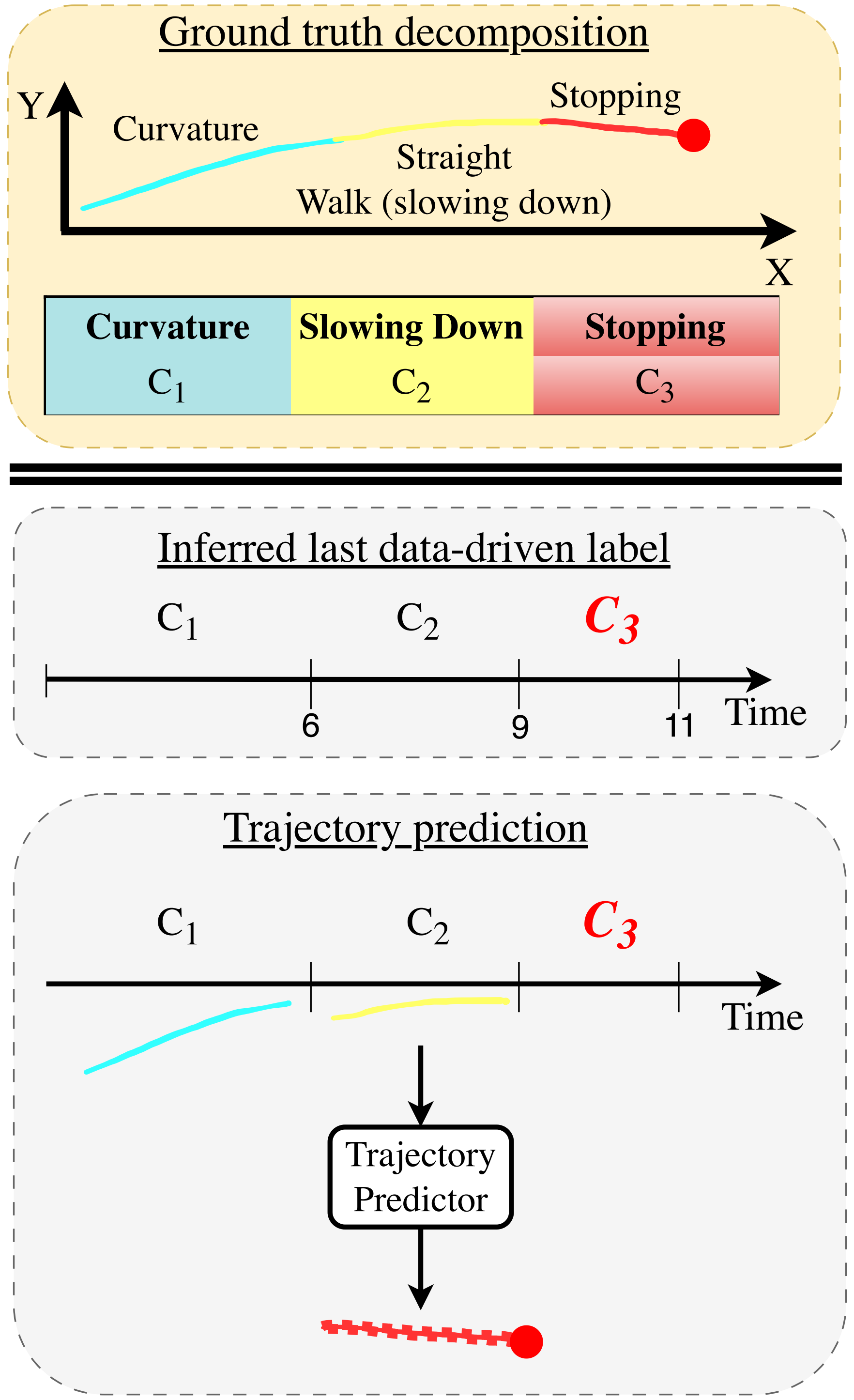


Complementary strengths:

- Observable classes provides explainability and human semantics.
- Data-driven classes captures motion-specific patterns.

Combining observable and data-driven classes presents an opportunity to unify **semantic interpretability** with **motion-based expressiveness**.

Time-Granularity in Data-driven Classes



Current formulation of data-driven classes overlooks temporal evolution of motion patterns within a single trajectory. Temporal decomposition of data-driven classes aligns conceptually to the notion of **fine-grained actions** enhancing:

- Predictors can identify and respond to **local behaviors** such as an upcoming stop or turning.
- **Improved generalization** as shorter segments are more generalizable across agents and contexts.
- **Realism and interpretability** by reflecting the inherently uneven structure of motion.

Learning to Understand and Predict Heterogeneous Trajectory Data

Tiago Rodrigues de Almeida[1]
Andrey Rudenko[2]
Johannes Andreas Stork[1]
Achim J. Lilienthal[1,2]
1. Örebro University; 2. Technical University of Munich

2025-09-26