Efficient Multimodal Pedestrian Trajectory Prediction with Vectorized Representations

Marwan Mashra – M2 student at Université Paris-Saclay

Supervisors: Julien Moreau, Lina Achaji

End of studies Internship (13 March 2023 – 12 september 2023)

Introduction

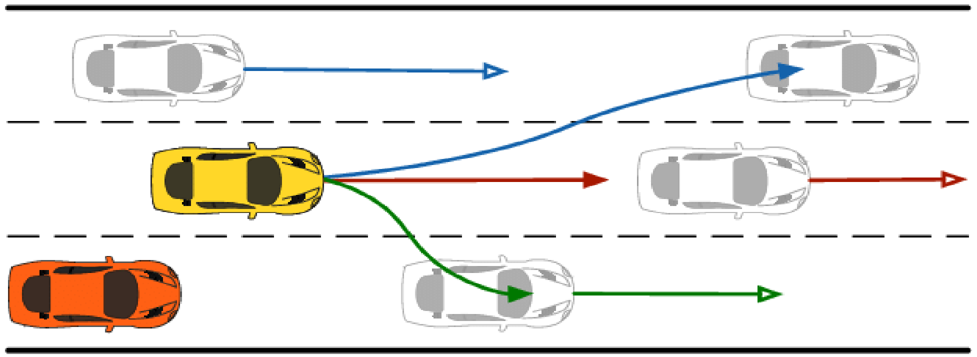
- STELLANTIS
- 6 months internship (March to September)
- Supervised by: **Julien Moreau** & **Lina Achaji**
- Worked on **Efficient Multimodal Pedestrian Trajectory Prediction**

STELLANTIS

Context & Background

Pedestrians Trajectory Prediction

- Given **past observations**, predict **future positions**
- Important for **Autonomous vehicles**
- Predict future **=** anticipate and avoid ~~**Collisions**~~
- **Multimodal** = multiple trajectories



Motivation

- Vehicle trajectory prediction focuses on **Efficiency**
- Pedestrian trajectory prediction, **not as much…**

STELLANTIS

Literature review

Goal-Based Method

- First predict **goals**, then predict the **trajectory**

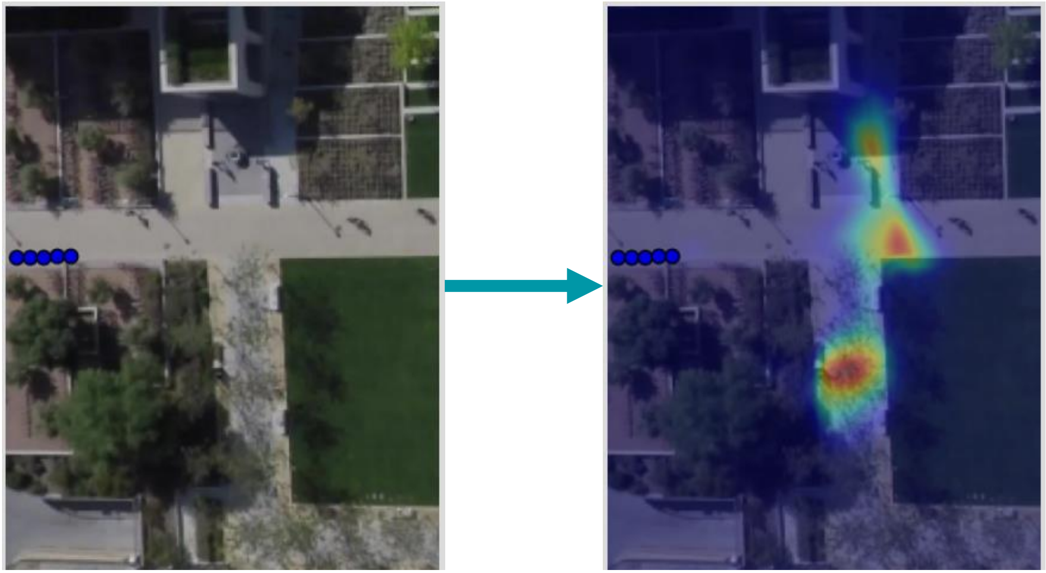
Goal-Based Method

- First predict **goals**, then predict the **trajectory**

input (observations)

Goal-Based Method

- First predict **goals**, then predict the **trajectory**



input (observations)   Generate Heatmap

Goal-Based Method

- First predict **goals**, then predict the **trajectory**

input (observations) Generate Heatmap Predict Goals

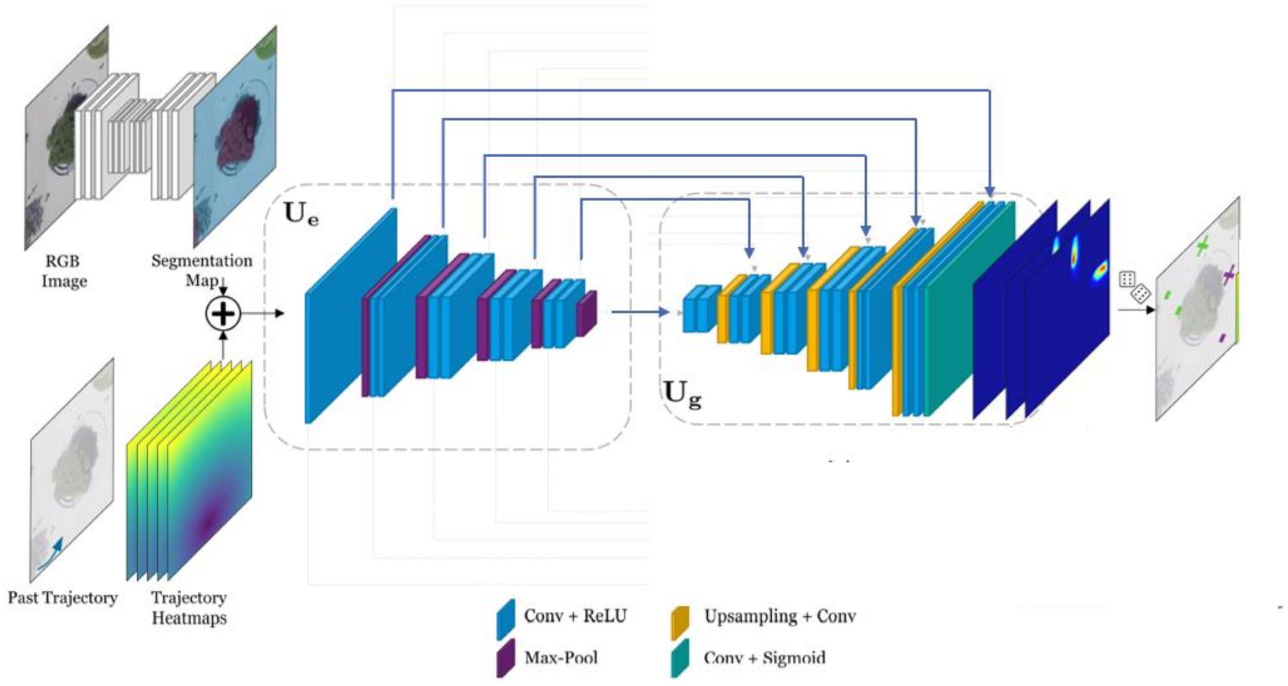
Goal-Based Method

- First predict **goals**, then predict the **trajectory**

input (observations) | Generate Heatmap | Predict Goals | Complete Trajectory

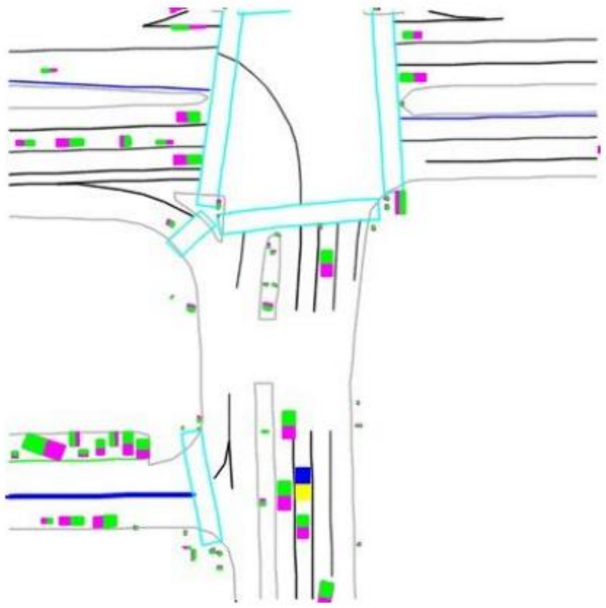
Goal-Based Method

- **Y-net** (UC Berkeley - 2020)
- **CNN** to predict the goals
- Lacks **Efficiency**
- **All SOTA** goal-based methods used it

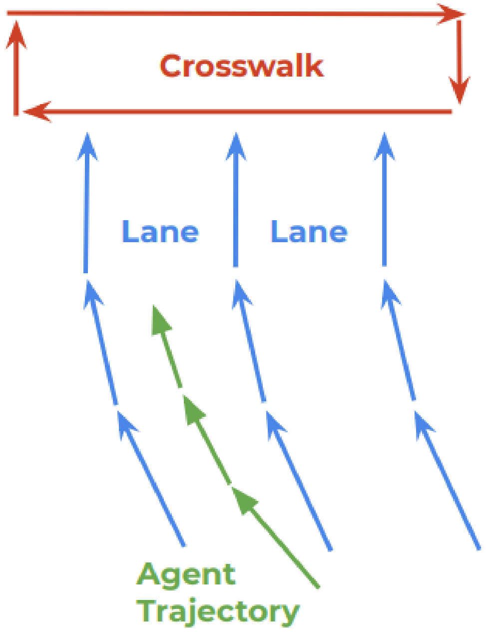


Goal prediction in Y-net

VectorNet: vectorized representations



vectorize



HD map

vectorized representations

DenseTNT

- Goal-based method, developed for vehicle trajectory predictions
- **VectorNet** backbone
- 1st place on the 2021 Waymo Challenge

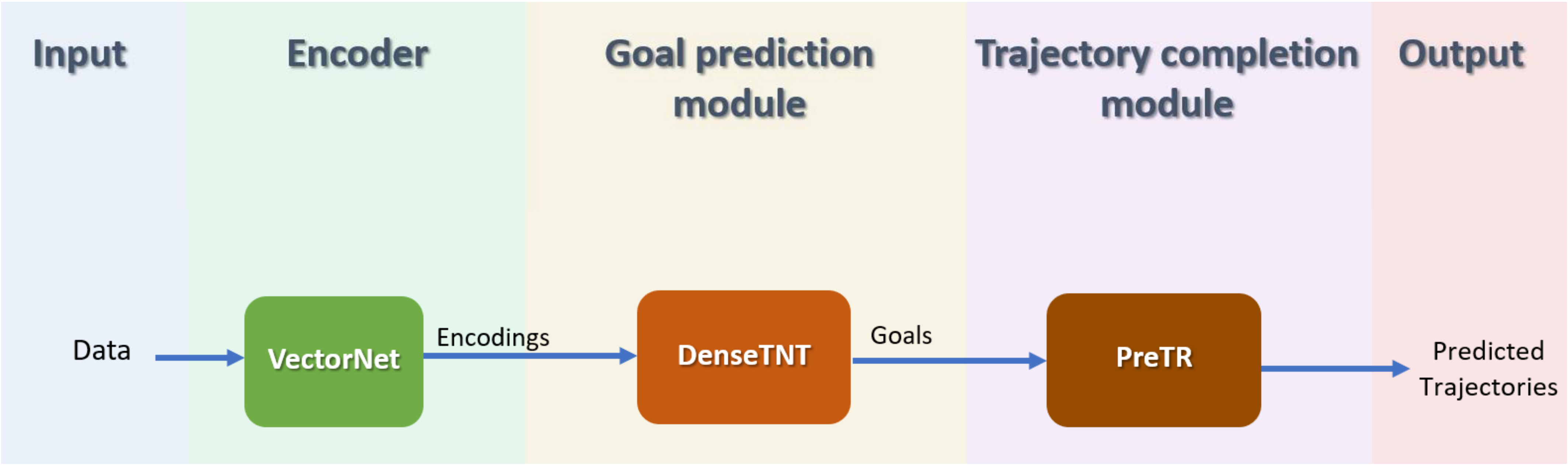
PreTR: Prediction Transformer

- Developed in **Stellantis,** for pedestrian trajectory prediction
- Focuses on modeling pedestrians social interactions
- Deterministic, **not multimodal**

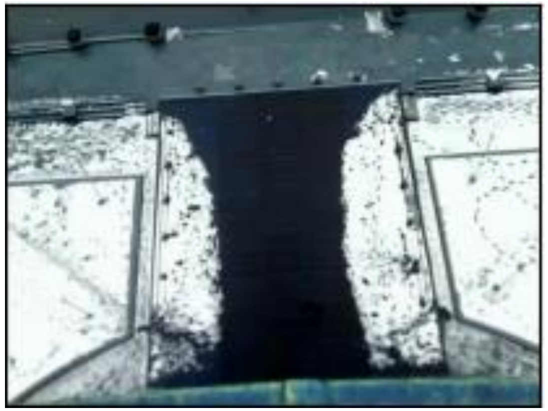
STELLANTIS

Methodology

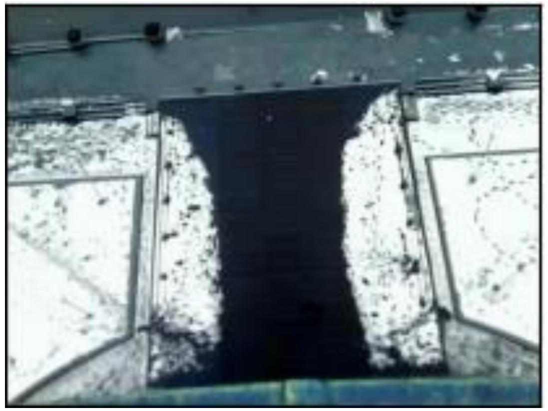
General architecture

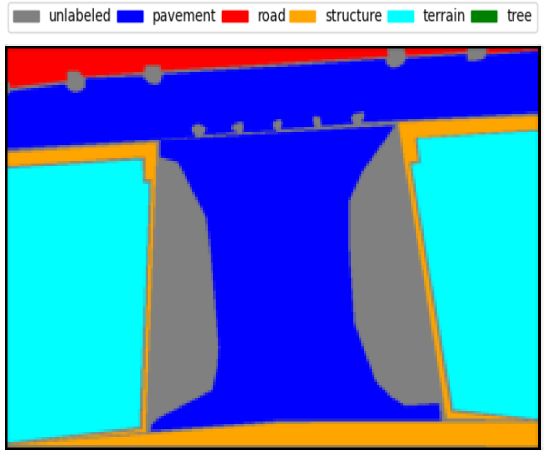


Vectorize the Map

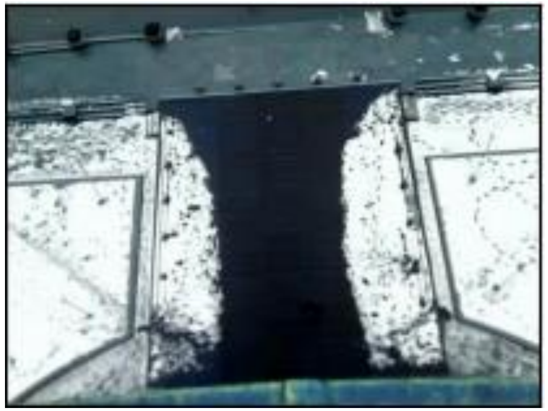


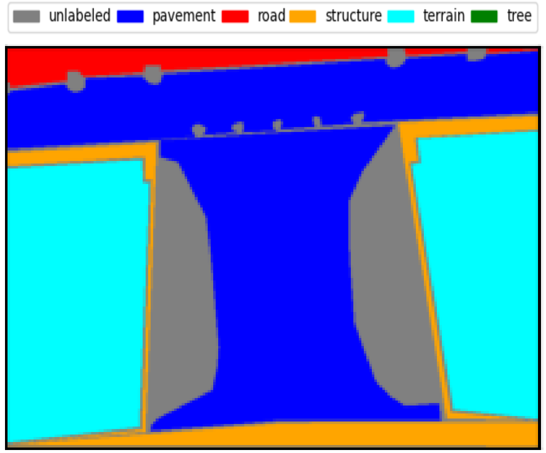
Vectorize the Map

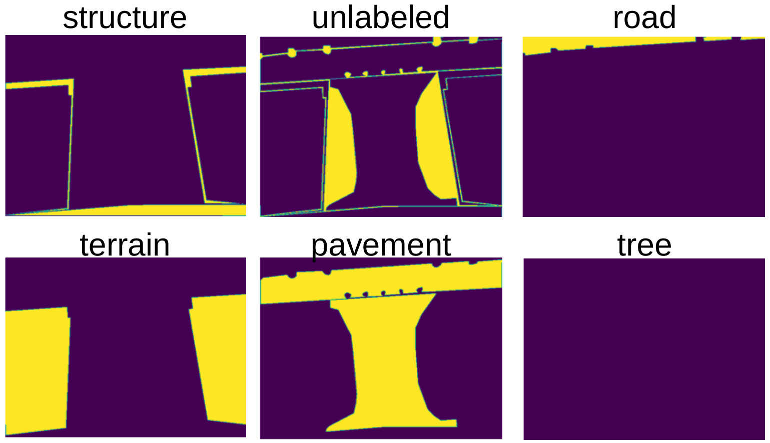




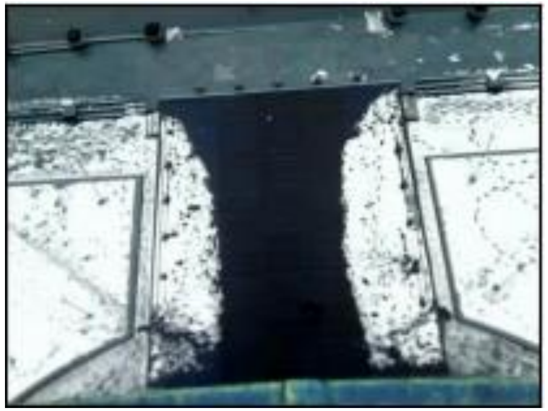
Vectorize the Map

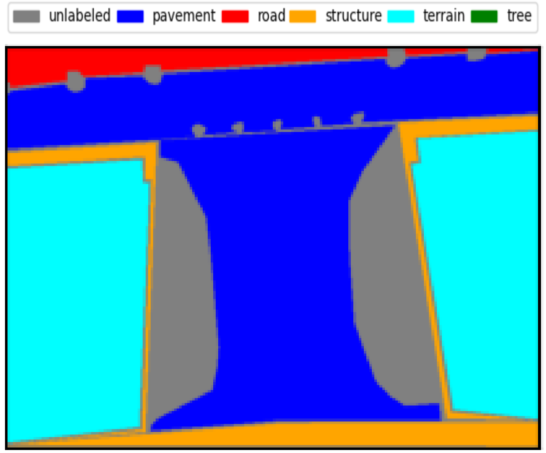


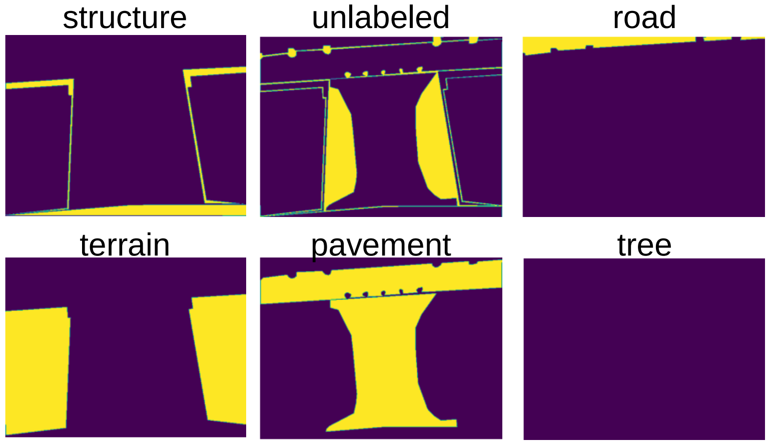


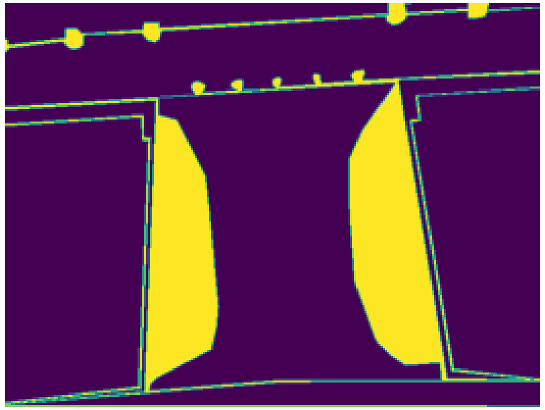


Vectorize the Map

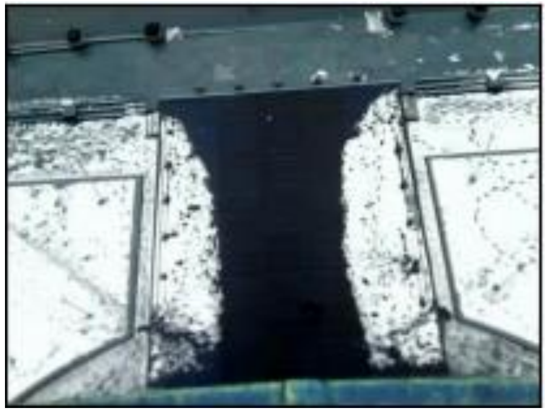


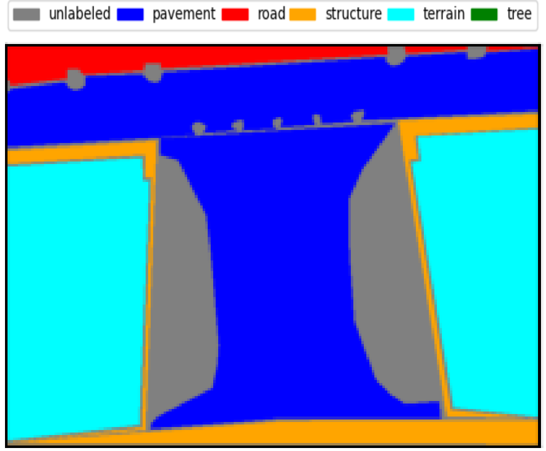


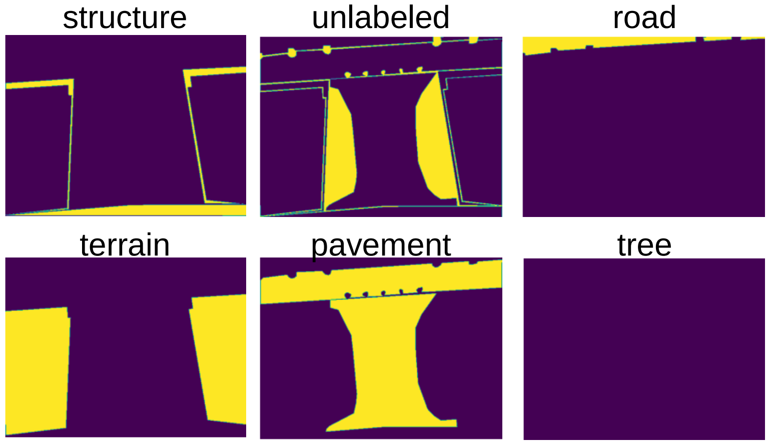


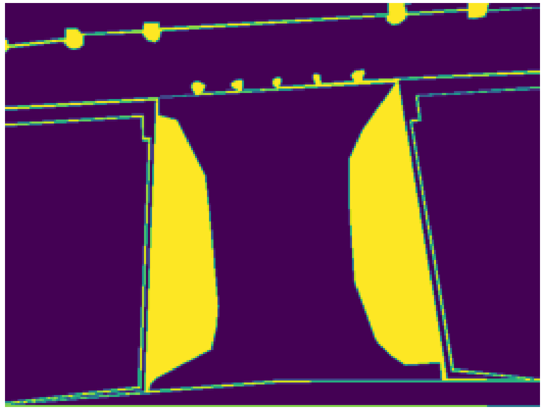


Vectorize the Map

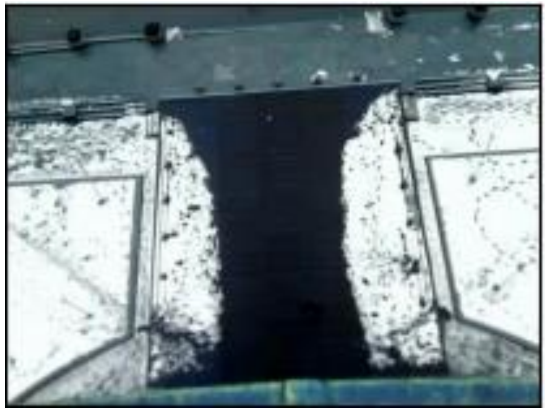


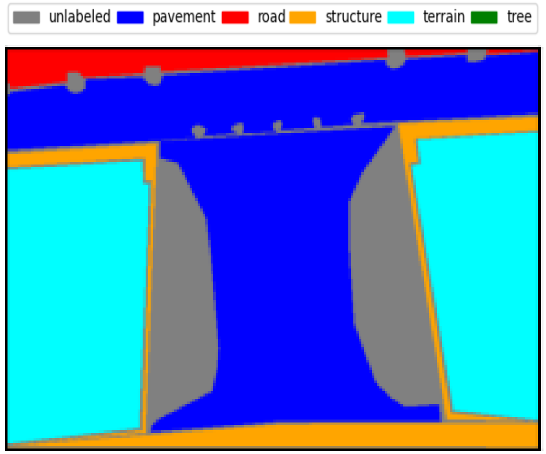


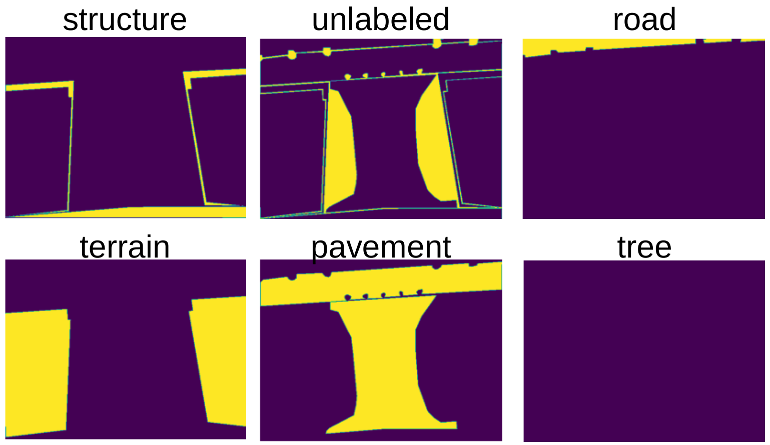


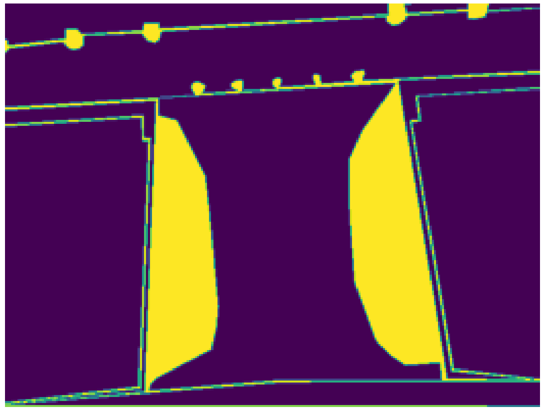


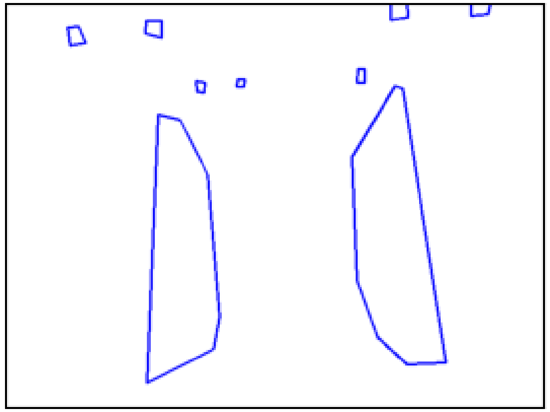
Vectorize the Map



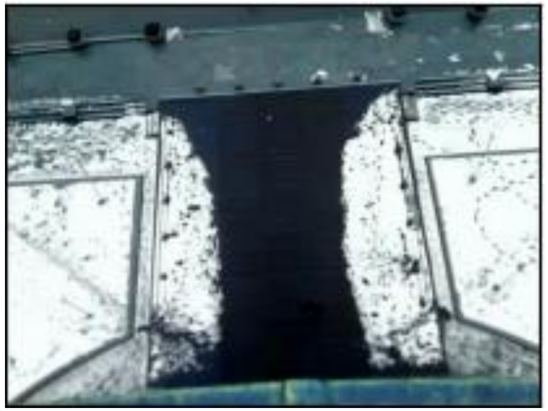


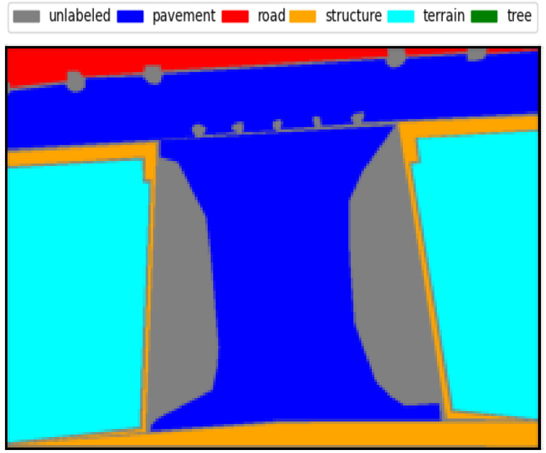


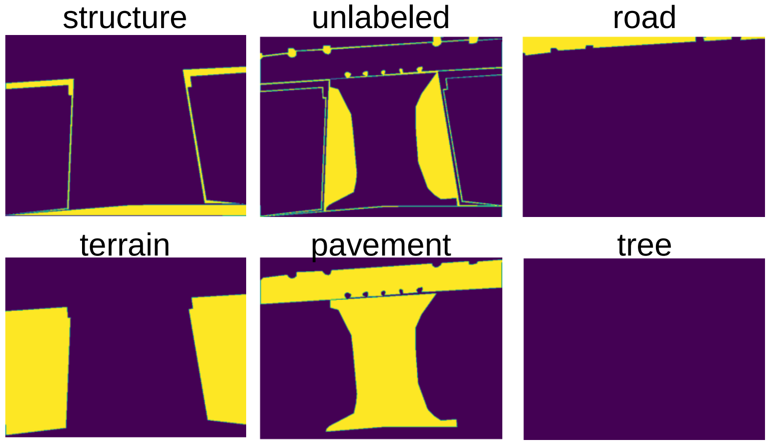


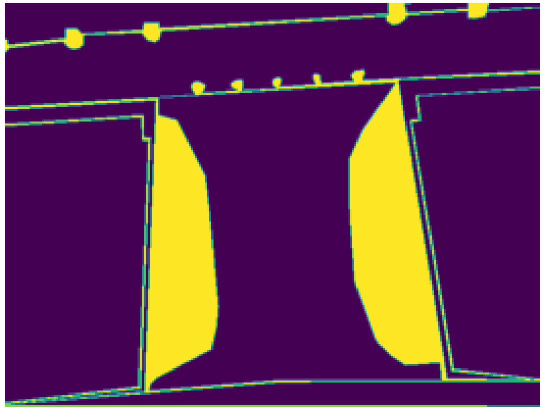


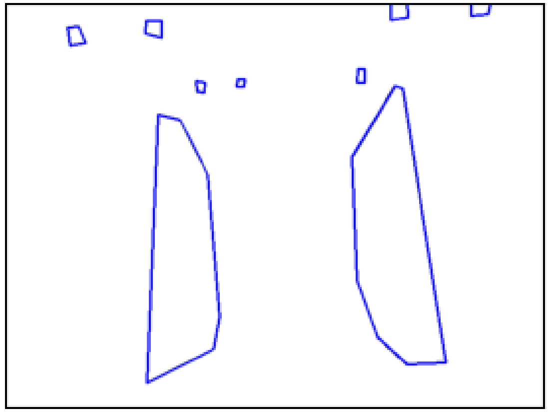
Vectorize the Map

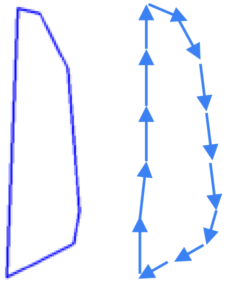








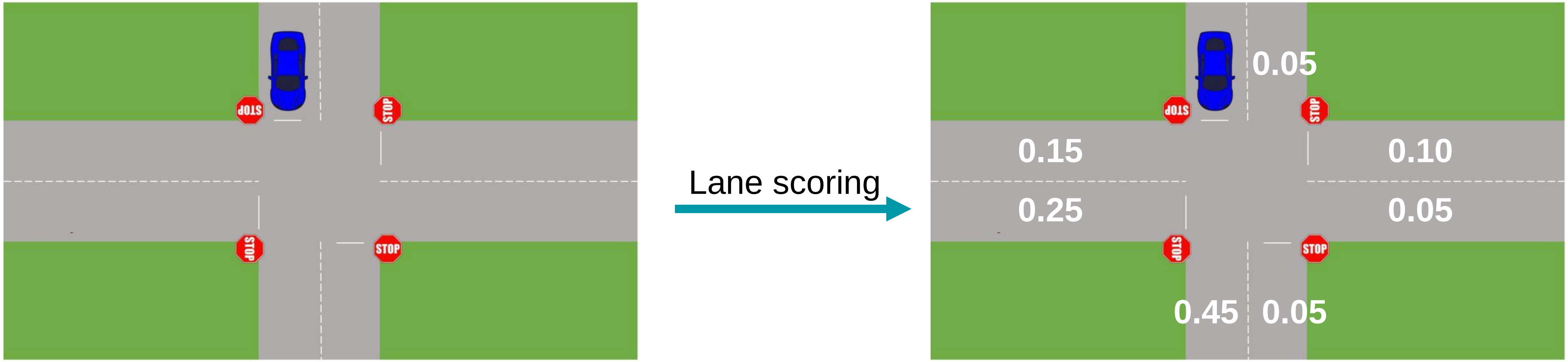




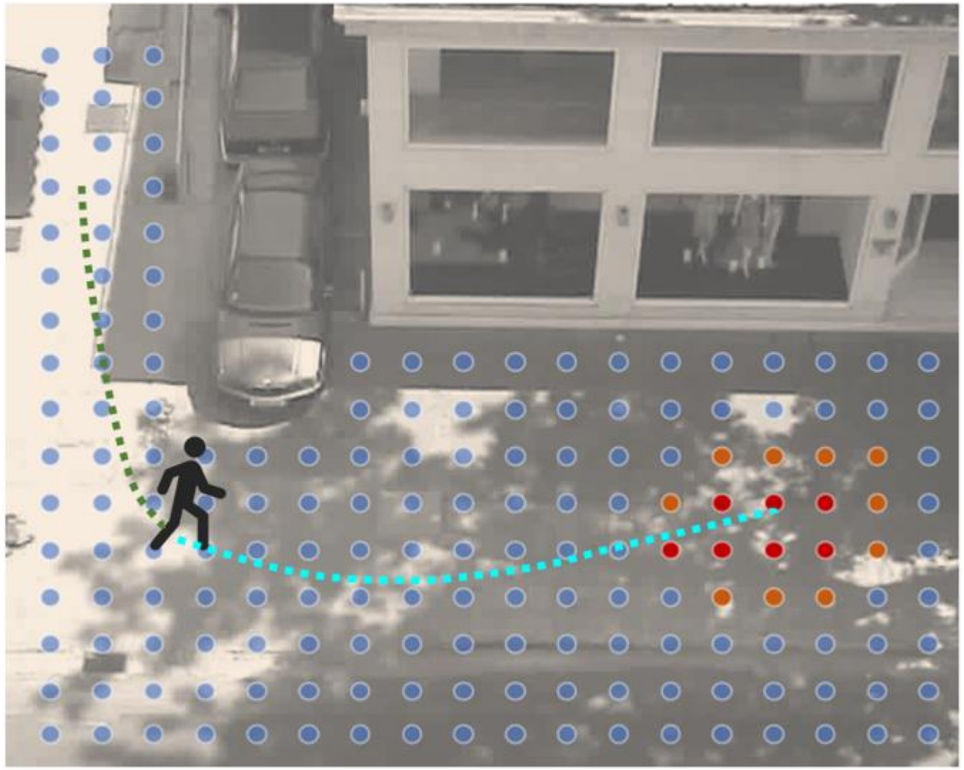
Goal prediction module

- Two stages training:
  - **First stage** learn to generate the **Heatmap**
  - **Second stage** learn to predict the **Goals**

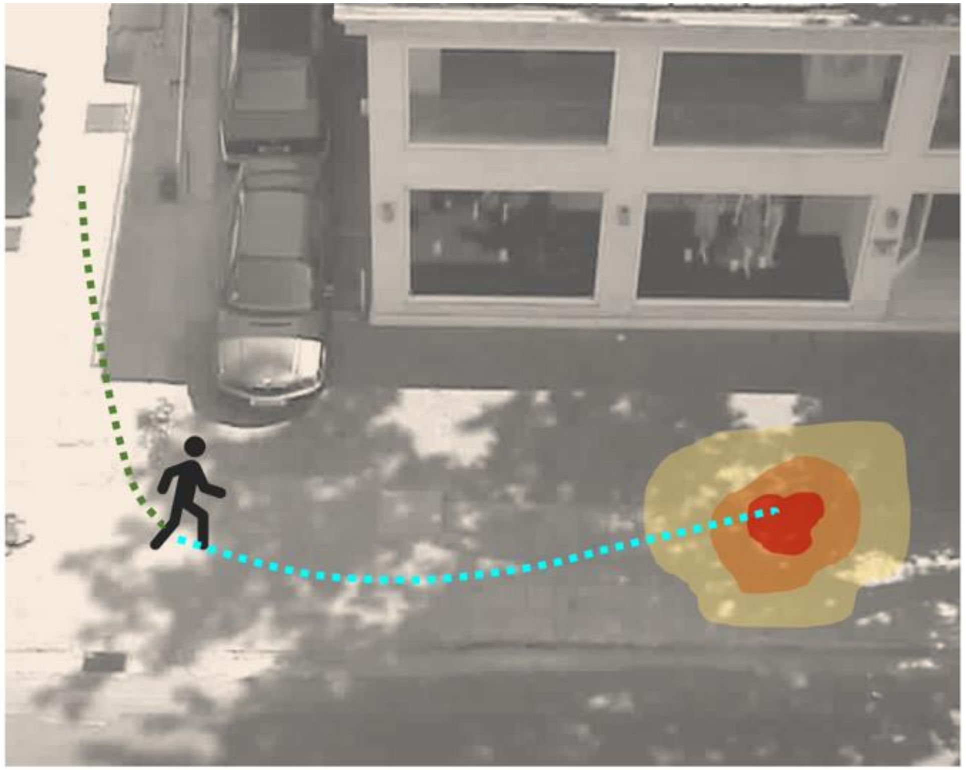
Heatmap generation



Heatmap generation



Sparse heatmap



Dense heatmap

Goal prediction

Offline optimization-based algorithm for pseudo-labels loss

→

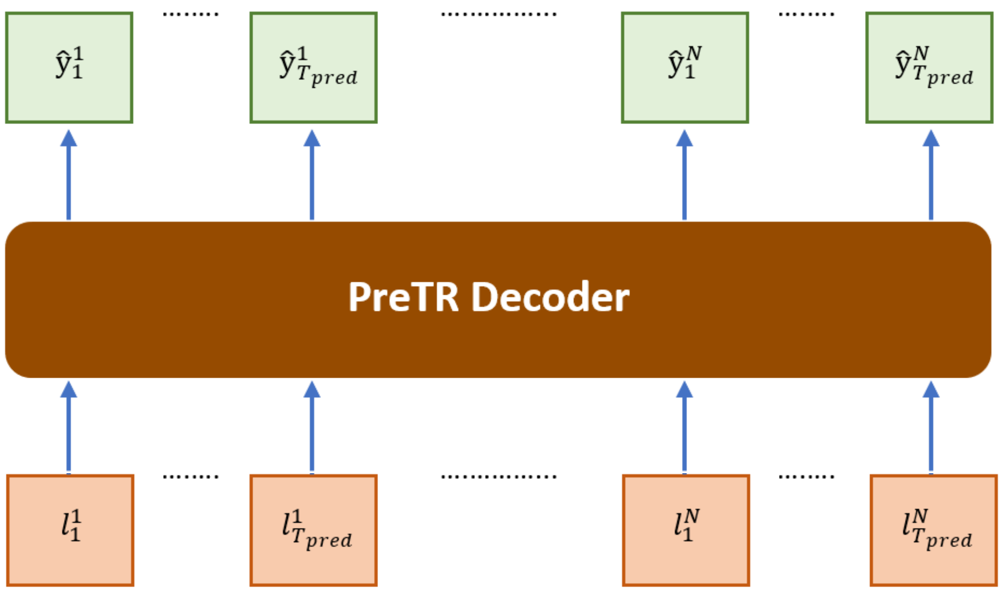
Winner-Takes-All loss

$$\mathcal{L}_{\text{wta}} = \min_{c_i \in \hat{C}} \left\| c_i - y_{T_{pred}} \right\|$$

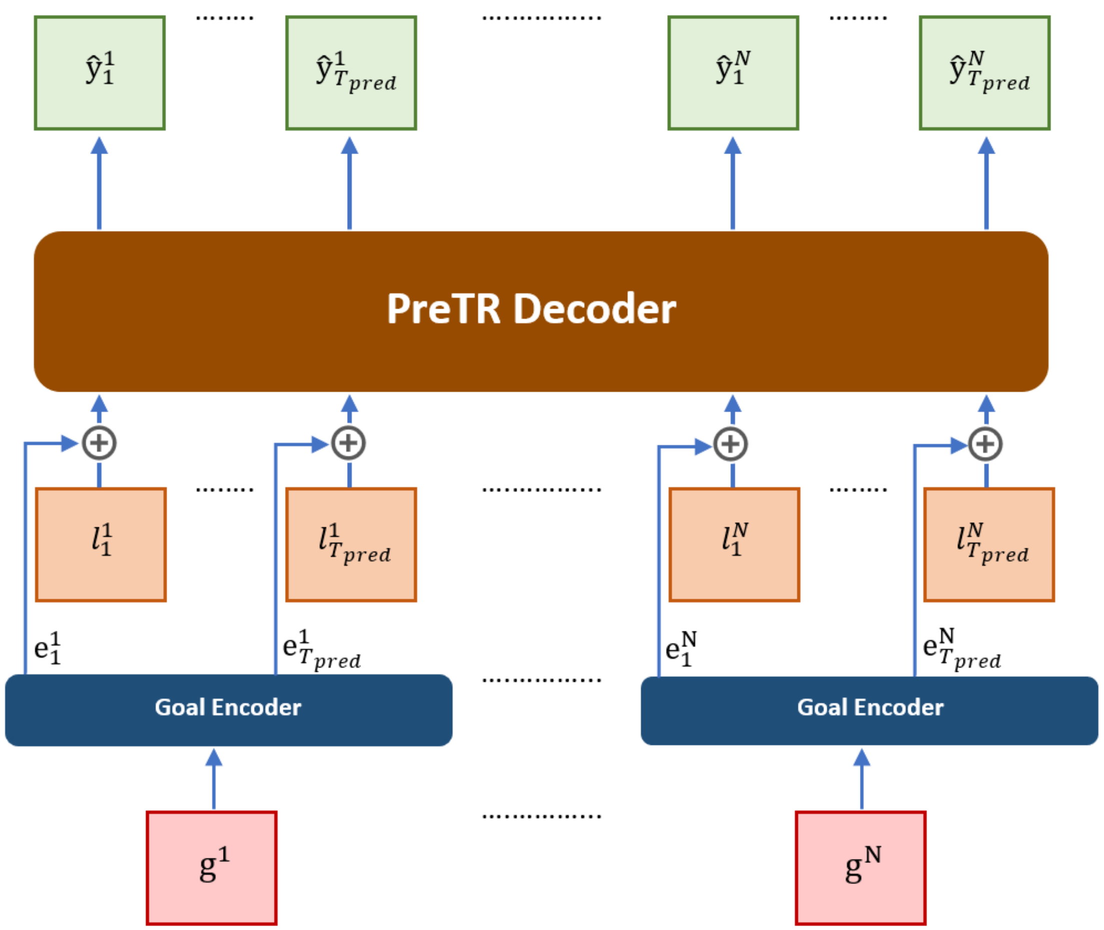
Trajectory completion module

- Feed the goals to **PreTR**
- Given a **goal**, predict a **trajectory**
- **K** predicted goals = **K** different trajectories

Trajectory completion module



Trajectory completion module



STELLANTIS

Results

Evaluation

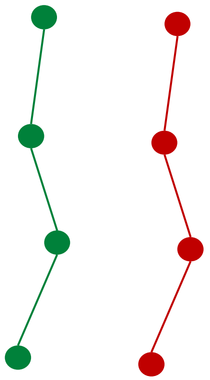
- **Datasets :** ETH/UCY datasets (ETH, HOTEL, UNIV, ZARA1, ZARA2)

Evaluation

- **Datasets :** ETH/UCY datasets (ETH, HOTEL, UNIV, ZARA1, ZARA2)
- **Evaluation protocol :** leave-one-out cross-validation strategy

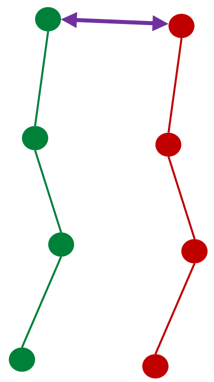
Evaluation

- **Datasets :** ETH/UCY datasets (ETH, HOTEL, UNIV, ZARA1, ZARA2)
- **Evaluation protocol :** leave-one-out cross-validation strategy
- **Metrics :**



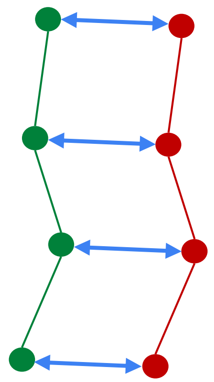
Evaluation

- **Datasets :** ETH/UCY datasets (ETH, HOTEL, UNIV, ZARA1, ZARA2)
- **Evaluation protocol :** leave-one-out cross-validation strategy
- **Metrics :**
  - minFDE (minimum final displacement error)



Evaluation

- **Datasets :** ETH/UCY datasets (ETH, HOTEL, UNIV, ZARA1, ZARA2)
- **Evaluation protocol :** leave-one-out cross-validation strategy
- **Metrics :**
  - minFDE (minimum final displacement error)
  - minADE (minimum average displacement error)



Results

Evaluation Metrics: $\text{minADE}_{20} \downarrow$ / $\text{minFDE}_{20} \downarrow$ [meters]

Method	ETH	HOTEL	UNIV	ZARA1	ZARA2	AVG
Social-LSTM (Alahi et al., 2016)	1.09/2.35	0.79/1.76	0.67/1.40	0.47/1.00	0.56/1.17	0.72/1.54
Social-GAN (Gupta et al., 2018)	0.81/1.52	0.72/1.61	0.60/1.26	0.34/0.69	0.42/0.84	0.58/1.18
Goal-GAN (Dendorfer et al., 2020)	0.59/1.18	0.19/0.35	0.60/1.19	0.43/0.87	0.32/0.65	0.43/0.85
ST-GAT (Huang et al., 2019)	0.65/1.12	0.35/0.66	0.52/1.10	0.34/0.69	0.29/0.60	0.43/0.83
MG-GAN (Dendorfer et al., 2021)	0.47/0.91	0.14/0.24	0.54/1.07	0.36/0.73	0.29/0.60	0.36/0.71
Transformer-TF (Giuliari et al., 2020)	0.61/1.12	0.18/0.30	0.35/0.65	0.22/0.38	0.17/0.32	0.31/0.55
STAR (Yu et al., 2020)	0.36/0.65	0.17/0.36	0.31/0.62	0.26/0.55	0.22/0.46	0.26/0.53
PECNet (Mangalam et al., 2020b)	0.54/0.87	0.18/0.24	0.35/0.60	0.22/0.39	0.17/0.30	0.29/0.48
Trajectron++ (Salzmann et al., 2021)	0.39/0.83	0.12/0.21	0.20/0.44	0.15/0.33	0.11/0.25	0.19/0.41
AgentFormer (Yuan et al., 2021)	0.45/0.75	0.14/0.22	0.25/0.45	0.18/0.30	0.14/0.24	0.23/0.39
Goal-SAR (Chiara et al., 2022)	0.28/0.39	0.12/0.17	0.25/0.43	0.17/0.26	0.15/0.22	0.19/0.29
Y-net (Mangalam et al., 2020a)	0.28/0.33	0.10/0.14	0.24/0.41	0.17/0.27	0.13/0.22	0.18/0.27
NSP-SFM (Yue et al., 2023)	0.25/0.24	0.09/0.13	0.21/0.38	0.16/0.27	0.12/0.20	0.17/0.24
Ours	0.52/0.65	0.17/0.24	0.32/0.50	0.24/0.37	0.21/0.31	0.30/0.42
$\pm$ *std*	$\pm$ *0.01/0.03*	$\pm$ *0.00/0.01*	$\pm$ *0.01/0.01*	$\pm$ *0.00/0.01*	$\pm$ *0.00/0.01*	$\pm$ *0.00/0.01*
Ours (offline)	0.54/0.69	0.17/0.27	0.32/0.49	0.26/0.39	0.22/0.34	0.30/0.44
$\pm$ *std*	$\pm$ *0.01/0.04*	$\pm$ *0.00/0.00*	$\pm$ *0.00/0.00*	$\pm$ *0.01/0.01*	$\pm$ *0.02/0.01*	$\pm$ *0.00/0.01*

CNN

Efficiency

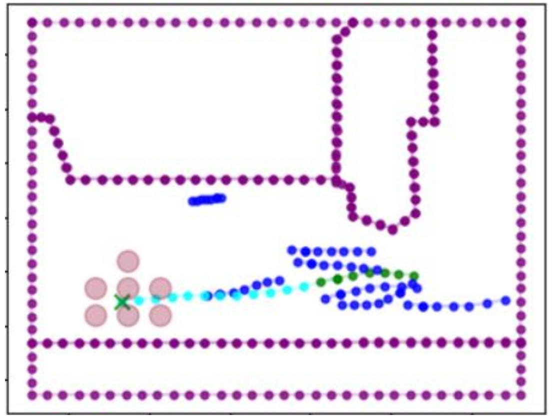
Model	**Training**		**Inference**	
	UNIV	ZARA1	UNIV	ZARA1
Goal-SAR	16:54:45	15:33:48	02:12:16	00:06:10
Ours (offline)	04:36:58	08:38:04	00:04:55	00:00:28
Ours	02:31:15	04:49:40	00:04:52	00:00:26

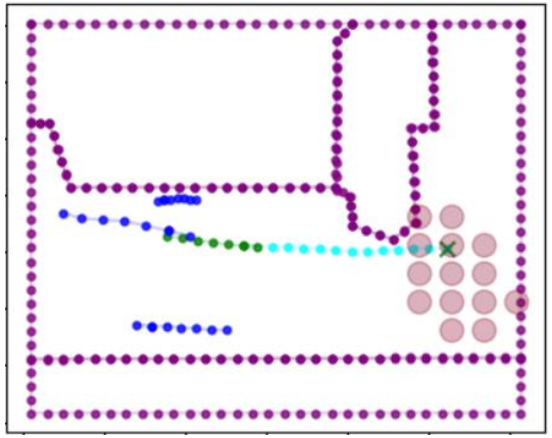
Visualizations

Other agents Past Future Predicted Map elements True goal Predicted goals

Visualizations

Other agents Past Future Predicted Map elements True goal Predicted goals

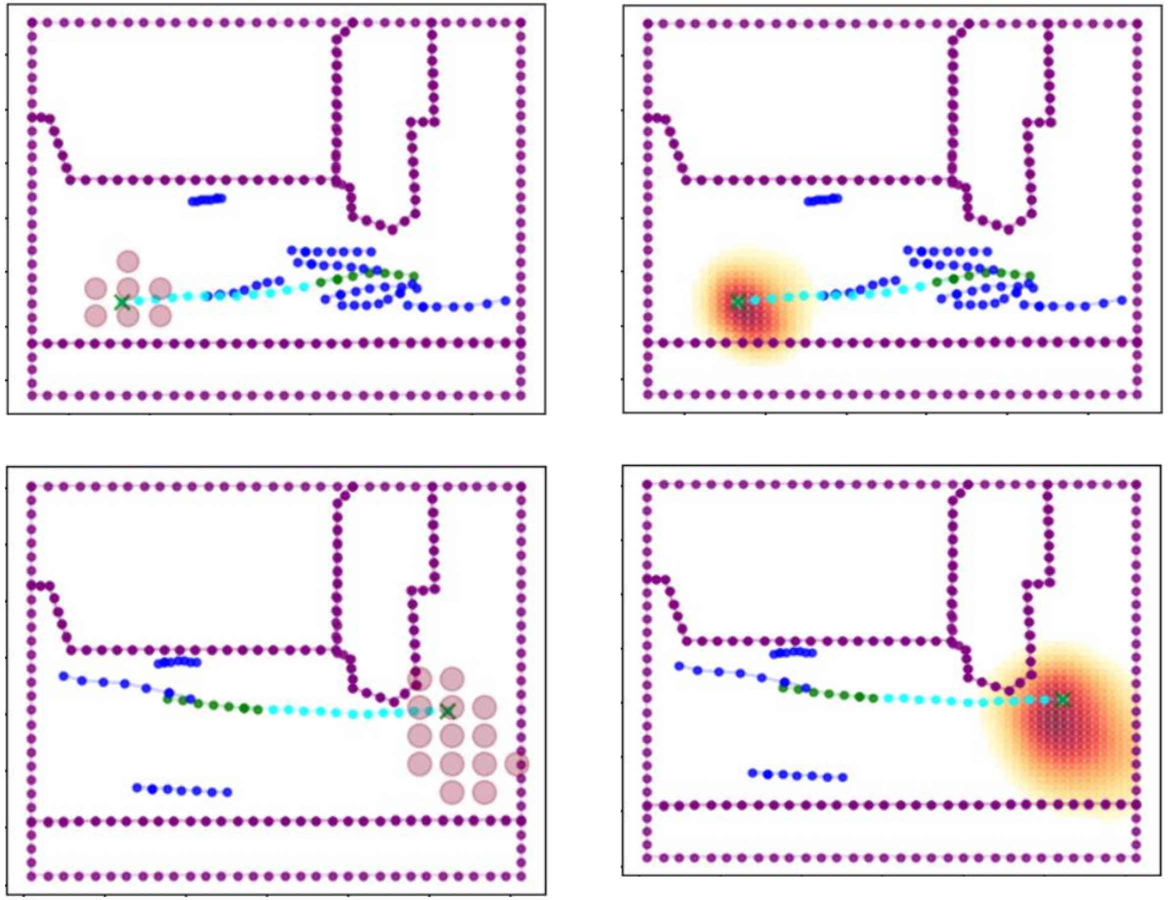




Predict General Area

Visualizations

Other agents Past Future Predicted Map elements True goal Predicted goals

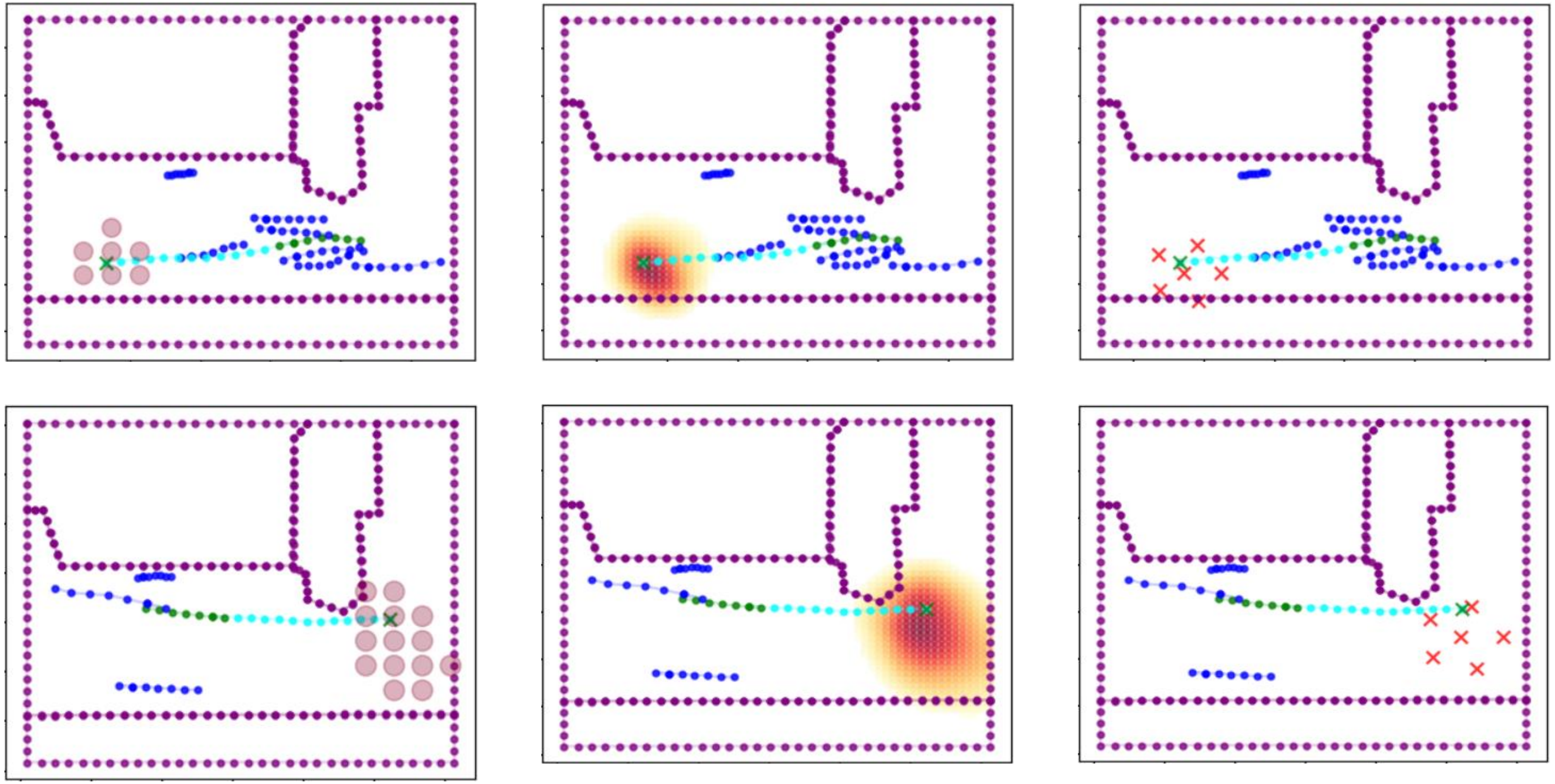


Predict General Area

Generate Heatmap

Visualizations

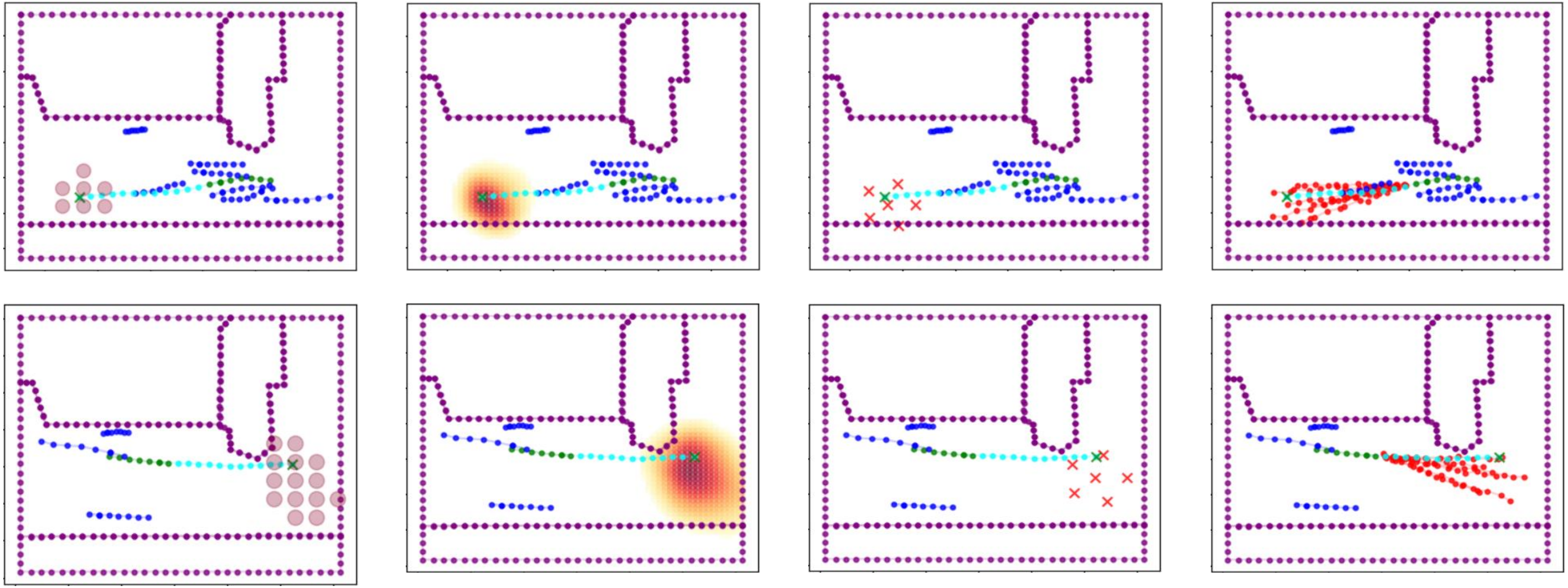
Other agents   Past   Future   Predicted   Map elements   True goal   Predicted goals



Predict General Area   Generate Heatmap   Predict Goals

Visualizations

Other agents · Past · Future · Predicted · Map elements · True goal · Predicted goals



Predict General Area | Generate Heatmap | Predict Goals | Predict Trajectories

STELLANTIS

Conclusion

Conclusion & Perspective

- Important gap with CNN-based methods, but more **efficient**
- Adapt vectorized representations to the pedestrian’s case

Questions