TSC-Net: Prediction of Pedestrian Trajectories by Trajectory-Scene-Cell Classification

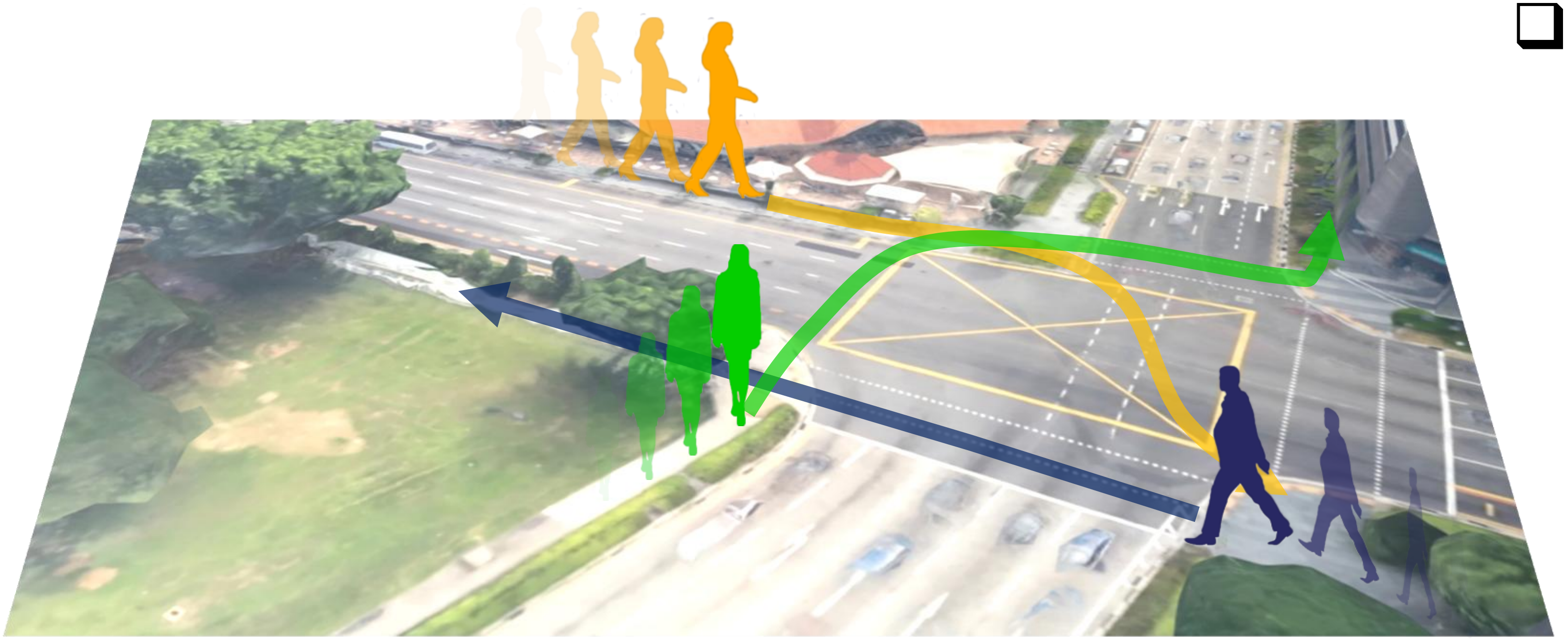

Bo Hu Tat-Jen Cham

College of Computing and Data Science

Nanyang Technological University

Trajectory Prediction

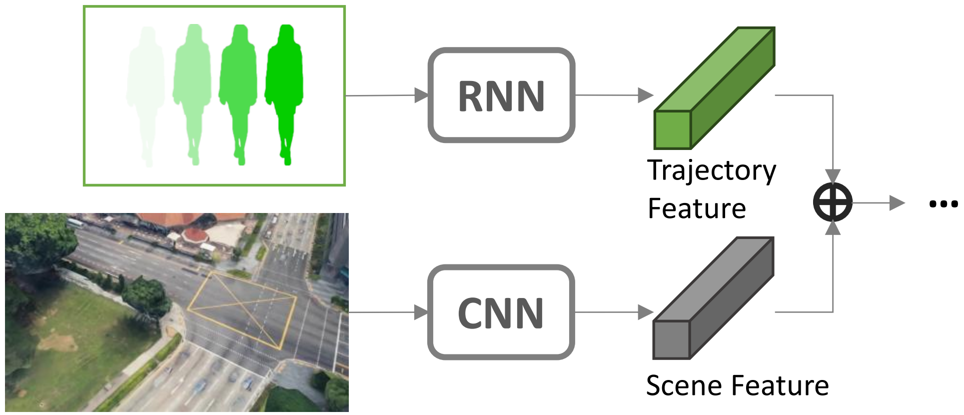
- Input: historical trajectories and scene map
- Output: future trajectories



- ❑ Temporal relation
- ❑ Human-human interaction
- ❑ Human-scene interaction

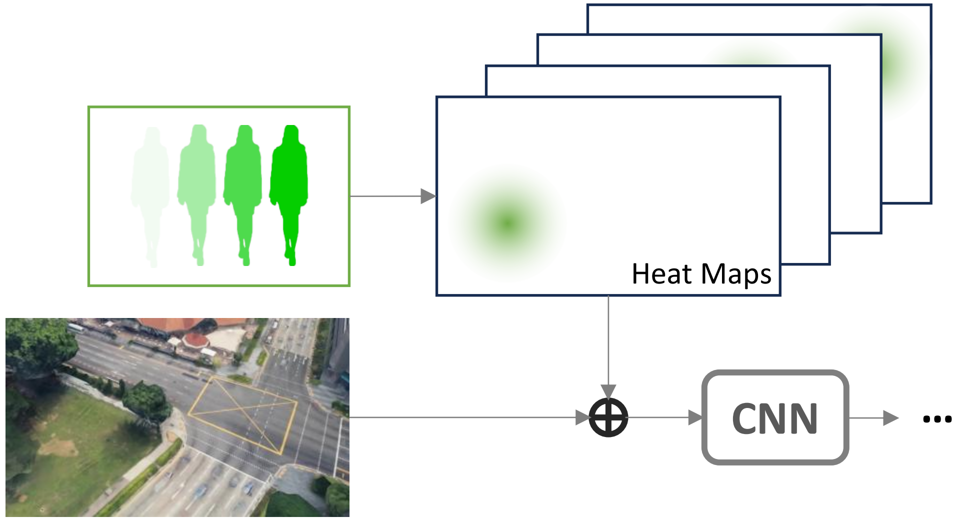
Related Works on Human-Scene Interaction

Scene Embedding



- Two feature types are not well aligned
- Difficult to align pedestrian coordinate and corresponding local scene

Trajectory Heat Map

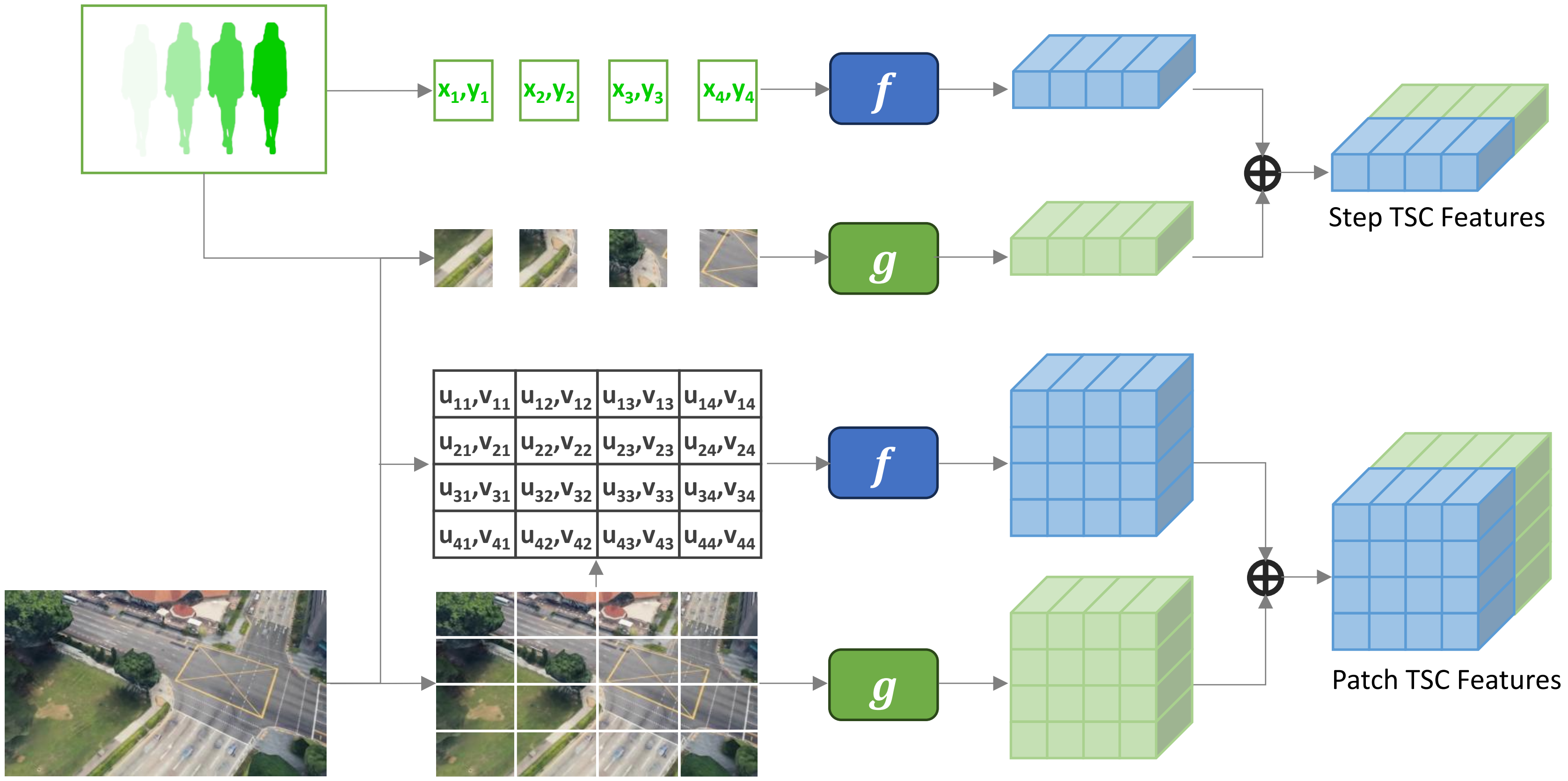


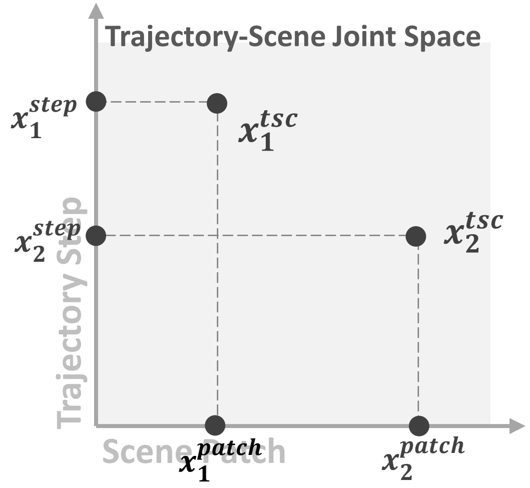
- Every pedestrian is represented by a 3D tensor
- Difficult to model human-human interactions

⊕: Concatenation

Trajectory-Scene-Cell Representation

1. **Decouple both trajectory and scene in to *cells***
2. **Build a *joint* feature space for trajectory and scene cells**



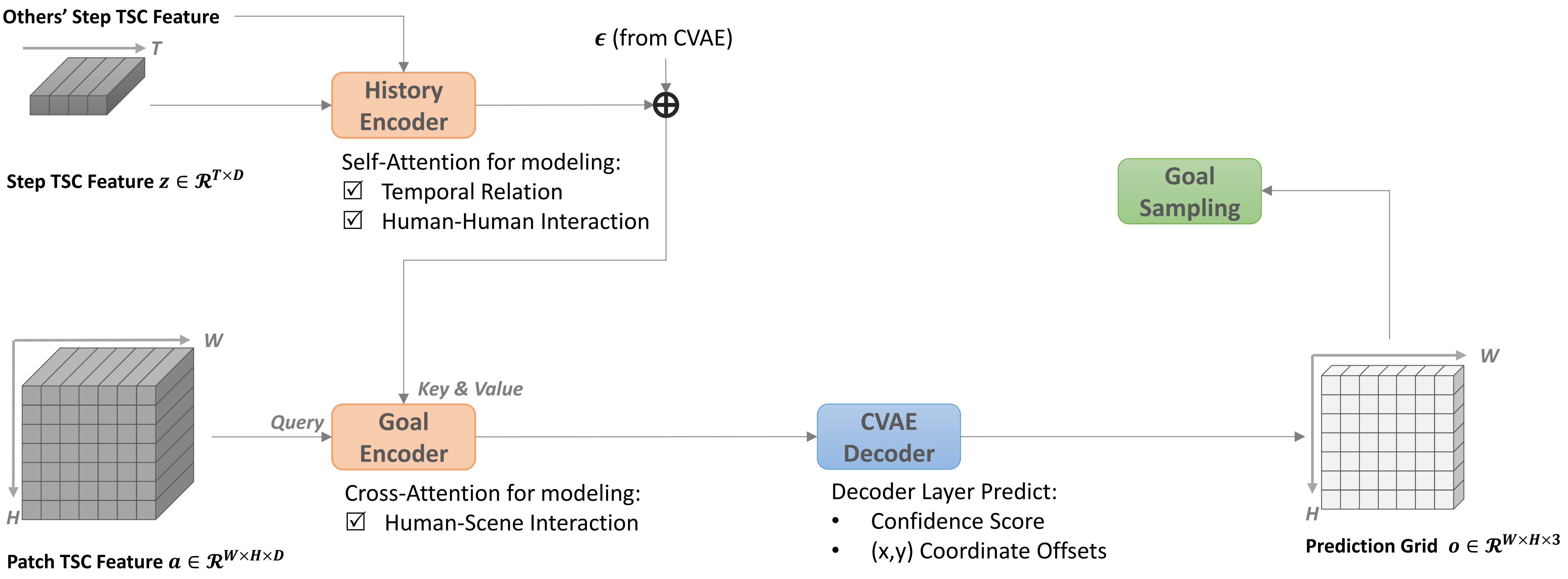


☑ ***Unify trajectory and scene features into one feature space***

f: Coordinate Embedding Network
g: Scene Embedding Network
$\oplus$: Concatenation

Prediction Framework – Goal Prediction

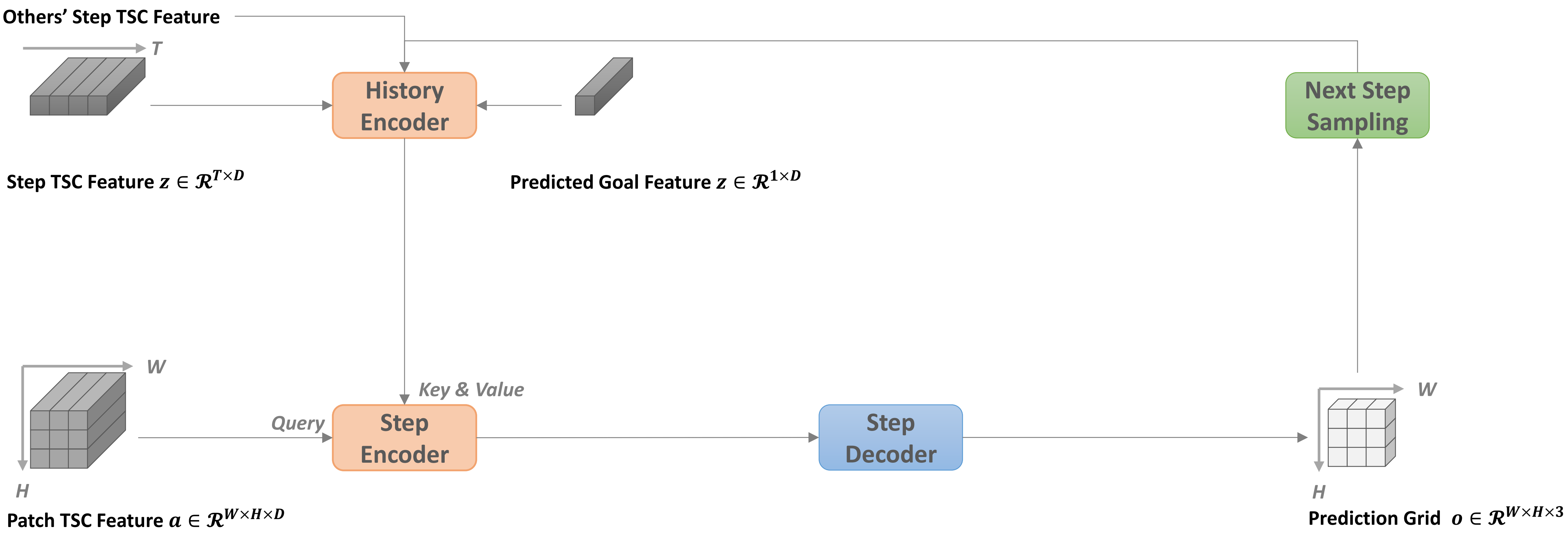
Predict goal by *cell classification*: which scene cell is the most probable location?



☑ ***Unify all types of interaction modeling with one operation***

Prediction Framework – Next Step Prediction

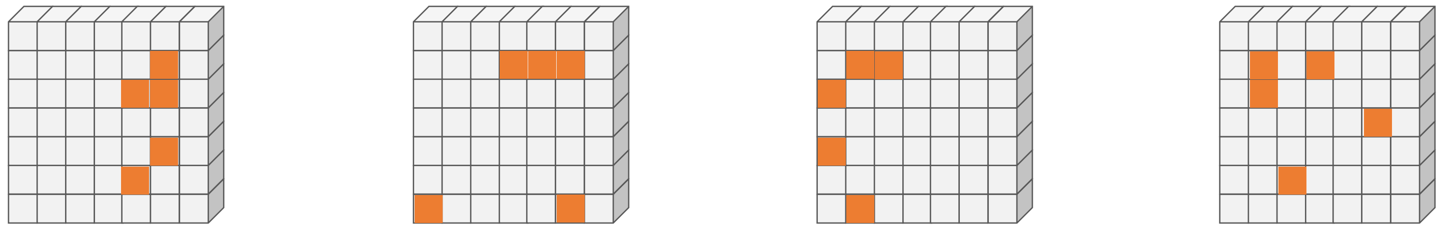
Predict <u>next step</u> by *cell classification*: which scene cell is the most probable?



☑ ***Unify all types of interaction modeling with one operation***

Prediction Framework – Hybrid Goal Sampling

- Two-Step Goal Sampling with CVAE:
  - Generate **m** goal output grid by CVAE
  - Pick **n** cells from each goal output grid with highest confidence score
- Hybrid Goal Sampling
  - Larger **m** performs better on short and smooth trajectories
  - Larger **n** performs better on long and non-smooth trajectories
  - Train a separate classifier, with historical trajectory as input, best **(m,n)** choice as output



$(m, n) = (4,5)$, results in $m \times n$ goals.

Prediction Results (ADE/FDE)

Results with Short-Term Prediction Setting (8 frames observation, 12 frames prediction, 2.5 fps)

Methods	ETH	HOTEL	UNIV	ZARA1	ZARA2	Average	SDD
Y-Net* Mangalam et al. (2021)	0.28/0.33	0.10/0.14	0.24/0.41	0.17/0.27	0.13/0.22	0.18/0.27	7.85 /11.85
NSP-SFM* Yue et al. (2022)	0.25/0.24	0.09/0.13	0.21/0.38	0.16/0.27	0.12/0.20	0.17/0.24	6.52 /10.61
MID Gu et al. (2022)	0.39/0.66	0.13/0.22	0.22/0.45	0.17/0.30	0.13/0.27	0.21/0.38	7.61 /14.30
PCCSNET Sun et al. (2021)	0.28/0.54	**0.11**/0.19	0.29/0.60	0.21/0.44	0.15/0.34	0.21/0.42	8.62 /16.16
MID Gu et al. (2022)	0.39/0.66	0.13/0.22	0.22/0.45	0.17/0.30	0.13/0.27	0.21/0.38	7.61 /14.30
FlowChain Maeda & Ukita (2023)	0.55/0.99	0.20/0.35	0.29/0.54	0.22/0.40	0.20/0.34	0.29/0.52	9.93/17.17
SICNet Dong et al. (2023)	0.27/0.45	**0.11**/0.16	0.26/0.46	0.19/0.33	0.13/0.26	0.19/0.33	8.44/13.65
TUTR Shi et al. (2023)	0.40/0.61	**0.11**/0.18	0.23/0.42	0.18/0.34	0.13/0.25	0.21/0.36	7.76/12.69
V^2-Net Wong et al. (2022)	**0.23/0.37**	**0.11**/0.16	0.21/0.35	0.19/0.30	0.14/0.24	0.18/0.28	7.12/11.39
LMTraj-SUP Bae et al. (2024)	0.41/0.51	0.12/0.16	0.22/0.34	0.20/0.32	0.17/0.27	0.22/0.32	7.80/10.10
E-V^2-Net-SC Wong et al. (2024)	0.25/0.38	0.12/**0.14**	**0.20/0.34**	0.18/**0.29**	0.13/**0.22**	**0.17/0.27**	6.54/10.36
TSC-Net (Ours)	0.32/0.39	0.12/0.19	0.25/0.46	0.17/0.30	0.15/0.26	0.20/0.32	**6.44/9.97**

Results with Long-Term Prediction Setting (5 frames observation, 30 frames prediction, 1 fps)

Methods	Social-GAN† Gupta et al. (2018)	PECNet† Mangalam et al. (2020)	R-PECNet†	Y-Net* Mangalam et al. (2021)	E-V^2-Net-SC ‡ Wong et al. (2024)	TSC-Net (Ours)
SDD	155.32/307.88	72.22/118.13	261.27/750.42	**47.94**/66.71	52.16/69.85	51.39/**63.97**
InD	38.57/84.61	20.25/32.95	341.8/1702.64	**14.99**/21.13	16.23/23.56	17.15/**19.53**

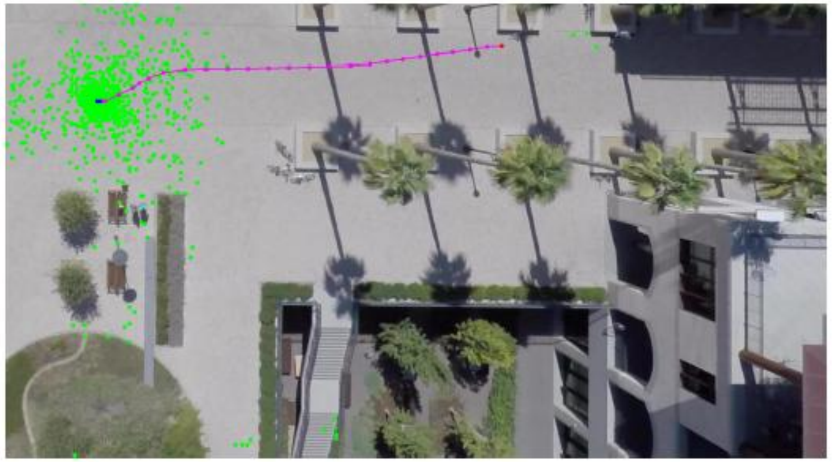
Prediction Results – Irregular Speed

- Compute velocity difference $V_{diff} = |\frac{\sum_{t=1}^{\tau} v_t}{\tau} - \frac{\sum_{t=\tau+1}^{T} v_t}{T-\tau}|$ for each trajectory sample
- Evaluate the trajectories with top x% largest velocity difference

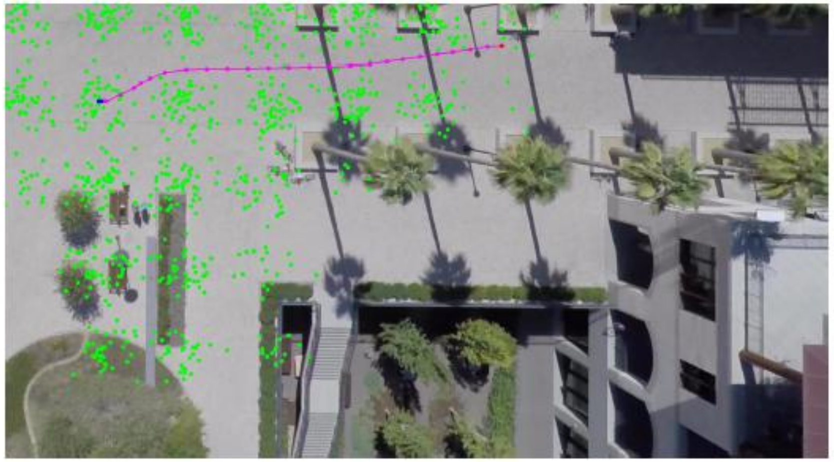
Results with Long-Term Prediction Setting (5 frames observation, 30 frames prediction, 1 fps)

Top (%) Samples with Largest V_{diff}	50%	40%	30%	20%	10%
Y-Net Mangalam et al. (2021)	**61.82** / 89.69	66.98 / 98.17	74.93 / 109.49	85.56 / 127.88	100.10 / 162.50
TSC-Net (Ours)	66.16 / **82.84**	**66.07 / 83.50**	**72.10 / 85.54**	**81.29 / 94.23**	**85.41 / 101.86**

Y-Net



Ours



History
Ground Truth Future
Sampled Goals

☑ ***Our method is good at predicting goals for sudden change of velocity.***

Thanks