SRA: Spatial Reasoning Adapter via Evolving Social Interaction Graphs for Trajectory Prediction

Jaewoo Jeong[1*], Seonkyu Song[1*], Hyeonwoo Park[2], Jegyeong Cho[1], Yujin Bae[1], Giwon Lee[1], Daehee Park[2,3], and Kuk-Jin Yoon[1]

{jeong207, sok9854, kjyoon}@kaist.ac.kr ; [1]Visual Intelligence Lab., KAIST, Korea ; [2]Intelligent Systems and Learning Lab., DGIST, Korea ; [3]UNIV * Equal contribution

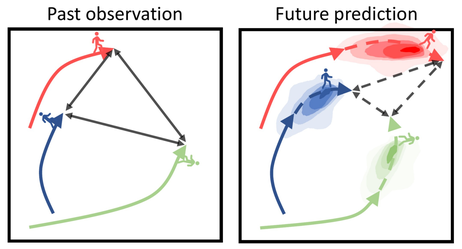
Code

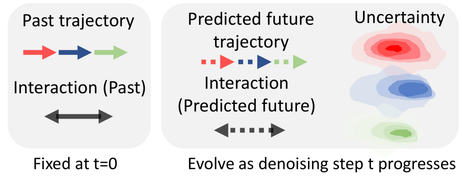
github.com/Jaewoo97/SRA

NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

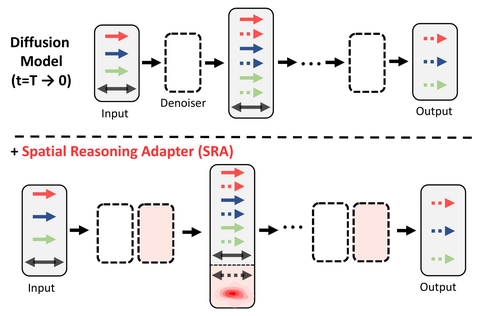
Problem: Existing predictors model interactions **only among past trajectories** (gray arrows) and keep them fixed during denoising → the predicted futures are **never aligned**.





Solution: SRA models interactions **among the predicted futures** (gray dashed arrows), rebuilt at every reverse step → aligned, **coherent** futures.

Proposed Method



Gated residual, injected into the host's update

$$\tilde{y}_i^\tau = \hat{y}_i^\tau + g_i\,\delta_i^\tau, \qquad g_i = \mathrm{sigmoid}\big(W_g[\,h_i,\ \bar{m}_i\,]\big)$$

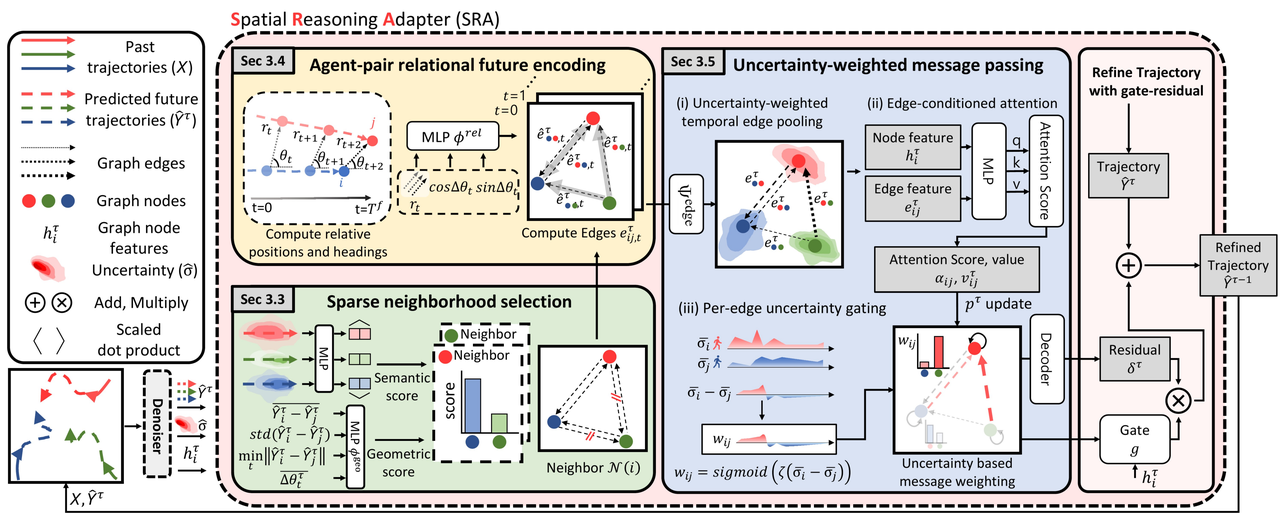
Message gate: confident source → uncertain target

$$w_{ij}^\tau = \mathrm{sigmoid}\big(\zeta\,(\bar{\sigma}_i^\tau - \bar{\sigma}_j^\tau) + b_w\big)$$

Training: host objective unchanged + NLL on σ

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{gen}} + \lambda_\sigma\,\mathcal{L}_\sigma$$

$$\mathcal{L}_\sigma = \tfrac{1}{2}\,\mathbb{E}\big[\log\sigma_\theta^2 + \|r_\theta\|^2/\sigma_\theta^2\big]$$



- **Sparse neighborhood selection:** top-K_{nbr} (= 5) neighbors per agent, scored by a semantic query–key match + a geometric bias on the predicted futures.
- **Agent-pair relational future encoding:** ego-centric relative position & heading difference of each selected pair, at every future timestep.
- **Uncertainty-weighted message passing:** predicted σ biases temporal edge pooling and gates each message by $\bar{\sigma}_i - \bar{\sigma}_j$ → confident agents anchor uncertain ones.

Quantitative Results

Main results — minADE / minFDE (K = 20, lower is better); (%) = relative gain over the host

Method	NBA ADE ↓	NBA FDE ↓	Soccer ADE ↓	Soccer FDE ↓	Football ADE ↓	Football FDE ↓	SDD ADE ↓	SDD FDE ↓
MID [15]	1.037	1.472	0.451	0.783	0.274	0.459	8.85	**15.92**
MID + SRA	**0.957** (+7.7%)	**1.346** (+8.6%)	**0.370** (+18.0%)	**0.595** (+24.0%)	**0.265** (+3.3%)	**0.438** (+4.6%)	**8.53** (+3.6%)	16.61 (−4.3%)
LED [37]	0.839	1.204	0.372	0.821	0.289	0.580	8.14	13.95
LED + SRA	**0.778** (+7.3%)	**1.161** (+3.6%)	**0.364** (+2.2%)	**0.710** (+13.5%)	**0.253** (+12.5%)	**0.505** (+12.9%)	**7.91** (+2.8%)	**13.31** (+4.6%)
MoFlow [14]	0.703	0.884	0.406	0.593	0.251	0.421	5.44	8.30
MoFlow + SRA	**0.695** (+1.1%)	**0.870** (+1.6%)	**0.374** (+7.9%)	**0.522** (+12.0%)	**0.245** (+2.4%)	**0.412** (+2.1%)	**5.37** (+1.3%)	**8.21** (+1.1%)

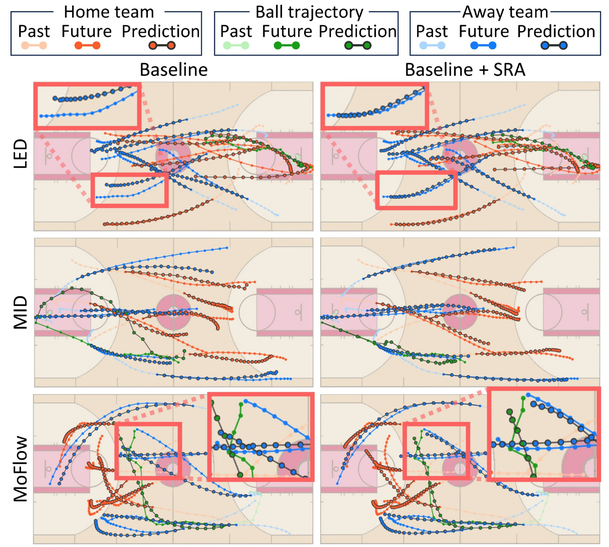
Ablation: cumulative component analysis (ADE)

Configuration	Enc.	Sel.	Unc.	NBA MID	NBA LED	NBA MoFlow	Soccer MID	Soccer LED	Soccer MoFlow	Football MID	Football LED	Football MoFlow	Improved
Host baseline				1.037	0.839	0.703	0.451	0.372	0.406	0.274	0.289	0.251	—
+ encoding (all neighbors)	✓			1.005	**0.776**	0.698	0.438	0.384	0.386	0.269	0.305	0.251	**6 / 9**
+ sparse selection	✓	✓		0.963	0.886	0.705	0.455	0.383	0.382	0.296	0.306	0.247	**3 / 9**
+ uncertainty (full SRA)	✓	✓	✓	**0.957**	0.778	**0.695**	**0.370**	**0.364**	**0.374**	**0.265**	**0.253**	**0.245**	**9 / 9**

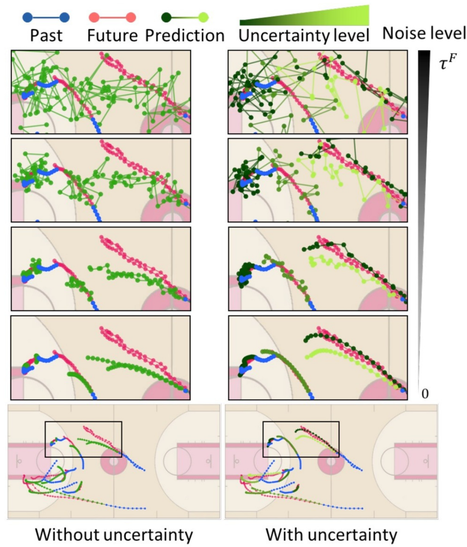
Bold: best per column · orange: worse than the host · Improved: pairs better than the host

- ✓ ADE improved on **all 12** host–dataset pairs; FDE on 11 / 12
- ✓ Average gains **5.8% ADE / 7.0% FDE**; up to 18.0% / 24.0%
- ✓ Only Enc. + Sel. + Unc. **together** improve all 9 sports pairs
- ✓ **Lightweight**: 0.63 M params, +1.5–8.6 ms per scene

Qualitative Results



NBA, host (left) vs. host + SRA (right): defender's return to defense (LED), ball kept with its carrier (MID), clean passing (MoFlow).



Without vs. with uncertainty (LED, NBA): messages from confident players recover the pass to the corner.