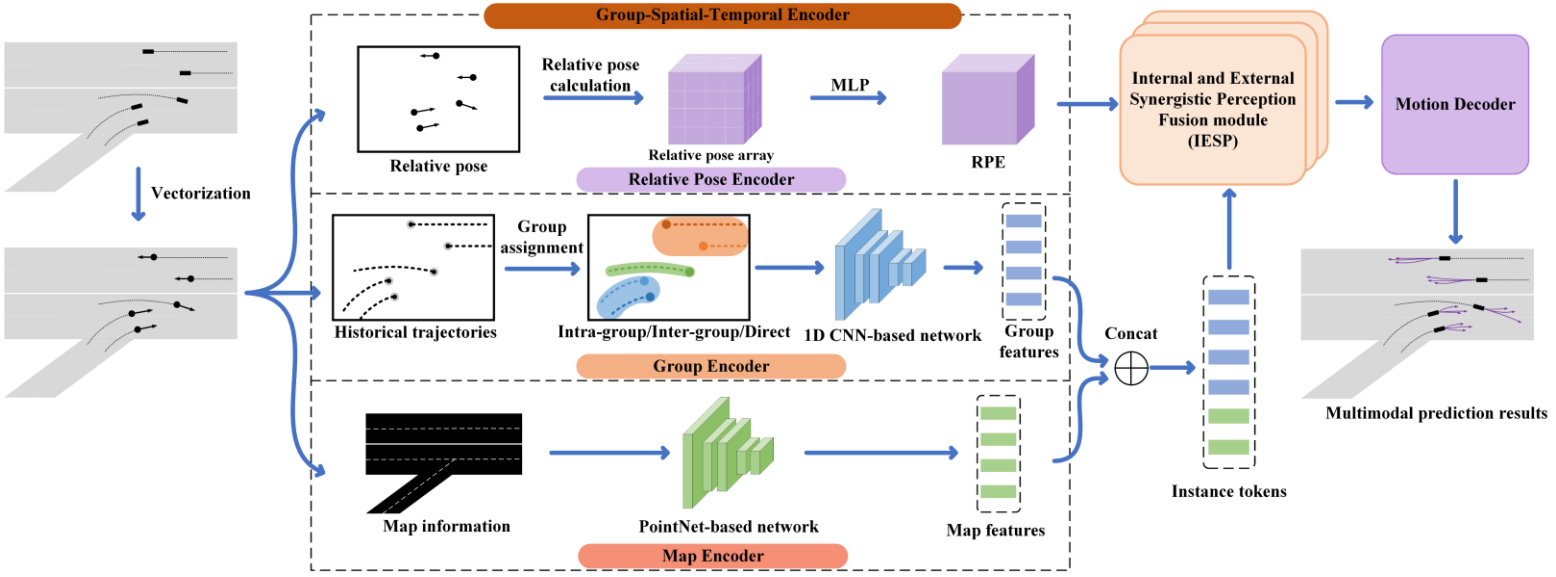
Trajectory Prediction Based on Grouped Spatial-Temporal Encoder

Di ZHOU, Ying GAO, Hui LI, Xiaoya LIU, Qinghua LIN

Frontiers of Computer Science, DOI: 10.1007/s11704-025-41156-z

Problems & Ideas

- Problems of conventional stereo matching approaches:
  - Traditional methods primarily focus on a single agent, overlooking the group representation of multiple agents.
  - Some methods rely excessively on unilateral perception mechanisms and fail to achieve comprehensive perception fusion.
- Ideas: Group-centered, grouping and pooling of all participants' characteristics and execute comprehensive information transfer and fusion through a dual perception mechanism.



The local features of the semantic instances are processed by the grouped spatial-temporal encoder, and the encoding results are input to the motion decoder after passing through the internal and external synergistic perception fusion module, which ultimately generates multimodal trajectory prediction results.

Main Contributions

- Contributions:
  - We propose a grouped spatial-temporal encoder to simultaneously encode group information and spatial-temporal features, capturing deep commonalities;
  - We propose an internal and external synergistic perception fusion module (**IESP**) for more comprehensive information fusion;
  - Experimental results show that despite its simplified design, **GSTEP** outperforms other state-of-the-art methods and baseline methods

Method	minADE_k	minFDE_k	MR_k	b-minFDE_k	Param
AutoBot	0.89	1.41	-	-	1.5M
LaneGCN	0.87	1.36	0.162	2.05	3.7M
mmTrans	0.84	1.34	0.154	2.03	2.6M
D-TNT	0.88	1.28	0.126	1.98	**1.1M**
THOMAS	0.94	1.44	0.100	1.97	-
TPCN	0.82	1.24	0.133	1.93	-
SceneTrans	0.80	1.23	1.126	1.89	15.3M
HiVT	0.77	1.17	0.127	1.84	2.5M
GANet	0.81	1.16	0.118	1.79	-
HPNet	<u>0.76</u>	<u>1.10</u>	<u>0.107</u>	<u>1.74</u>	-
LTP	0.83	1.30	0.155	1.86	**1.1M**
ADAPT	0.80	1.18	-	1.82	<u>1.4M</u>
SIMPL	0.79	1.18	0.123	1.81	1.8M
GSTEP(ours)	**0.65**	**0.97**	**0.084**	**1.60**	1.9M

Comparison of experimental results for different models on the test split on the Argoverse 1 motion forecasting dataset. (The best result is in **bold** while the second best result is <u>underlined</u>.)

Method	minADE_k	minFDE_k	MR_k	b-minFDE_k	Param
HDGT	0.84	1.60	0.214	2.24	12.1M
GoReal	0.76	1.48	0.220	2.01	-
QCNet	**0.65**	**1.29**	**0.160**	**1.91**	7.3M
SIMPL	0.72	1.43	0.192	2.05	**1.9M**
GSTEP(ours)	<u>0.67</u>	<u>1.35</u>	<u>0.180</u>	<u>1.95</u>	<u>2.0M</u>

Comparison of experimental results for different models on the test split on the Argoverse 2 motion forecasting dataset.