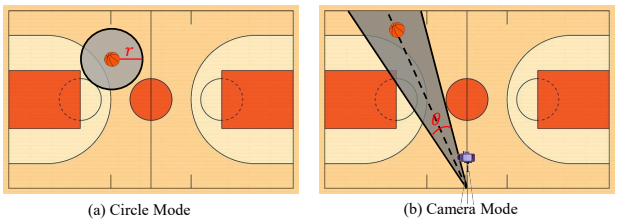
Supplementary Material for Uncovering the Missing Pattern: Unified Framework Towards Trajectory Imputation and Prediction

Yi Xu[1,2,*] Armin Bazarjani[2,3] Hyung-gun Chi[2,4] Chiho Choi[2,5] Yun Fu[1]
[1]Northeastern University [2]Honda Research Institute, USA
[3]University of Southern California [4]Purdue University [5]Samsung Semiconductor US
xu.yi@northeastern.edu, bazarjan@usc.edu, hgchi@purdue.edu
chiho1.choi@samsung.com, yunfu@ece.neu.edu

1. Details of -*TIP* Datasets

This section provides detailed information about the three curated datasets used in our experiments.

1.1. Basketball-TIP



(a) Circle Mode (b) Camera Mode

Figure 1. Two strategies, "circle mode" and "camera mode", to simulate the realistic appearance and disappearance of players.

Basketball-TIP is constructed using the NBA dataset [4], which contains 104,003 training sequences and 13,464 testing sequences. Each sequence includes 11 trajectories: the ball, 5 offensive players, and 5 defensive players, with 50 frames captured over 8 seconds. The basketball court size is 94 feet by 50 feet. We use two strategies, "circle mode" and "camera mode", to simulate the realistic appearance and disappearance of players during real sports game matches, as shown in Fig. 1. In the "circle mode" strategy, we define a circle centered on the ball with a radius of r. Players within this circle are visible while others are invisible. In the "camera mode" strategy, we place a simulated camera located at the sideline midpoint of the basketball court with a fixed field of view (FOV) angle of θ. The FOV bisector tracks the ball throughout the whole sequence, and by intersecting the FOV with the court plane, we obtain a projected in-frame polygon. Players within this polygon are observable while others are invisible. We curate Basketball-TIP with six scenarios by defining three different radii r (in feet) and three different angles θ (in degrees).

*Work done during Yi's internship at Honda Research Institute, under Chiho Choi's supervision.

1.2. Football-TIP

Football-TIP is constructed from NFL Football Dataset[1], which contains 10,780 training sequences and 2,492 testing sequences. Each sequence includes 21 trajectories: the ball, 10 offensive players, and 10 defensive players, with 50 frames captured over 10 seconds. The football field size is 120 yards by 53.3 yards. We use the same "circle mode" and "camera mode" strategies as in Basketball-TIP to curate different scenarios. We define three different radii r (in yards) and three different angles θ (in degrees) to curate six scenarios for evaluation.

1.3. Vehicle-TIP

We curated Vehicle-TIP using the Omni-MOT [3][2], which offers three levels of difficulty based on the camera's viewpoints: Easy, Ordinary, and Hard. With the Easy viewpoint, we have 29,239 training sequences and 6,419 testing sequences. With the Ordinary viewpoint, we have 33,831 training sequences and 7,427 testing sequences. And with the Hard viewpoint, we have 31,714 training sequences and 6,962 testing sequences. The scene size is 810 $\times$ 540 pixels. The dataset also includes an integrity value for each time step, which indicates how much of the vehicle is occluded. This allowed us to simulate real-world scenarios of occlusions while retaining the ground truth positional values of each vehicle. By thresholding the integrity value, we could determine if a vehicle was occluded or not.

1.4. Dateset Pre-Processing and Statistics

Pre-Processing. In our experiments, we normalized Basketball-TIP and Football-TIP data into $[-1, 1]$ based on

[1]https://github.com/nfl-football-ops/Big-Data-Bowl
[2]https://github.com/shijieS/OmniMOTDataset

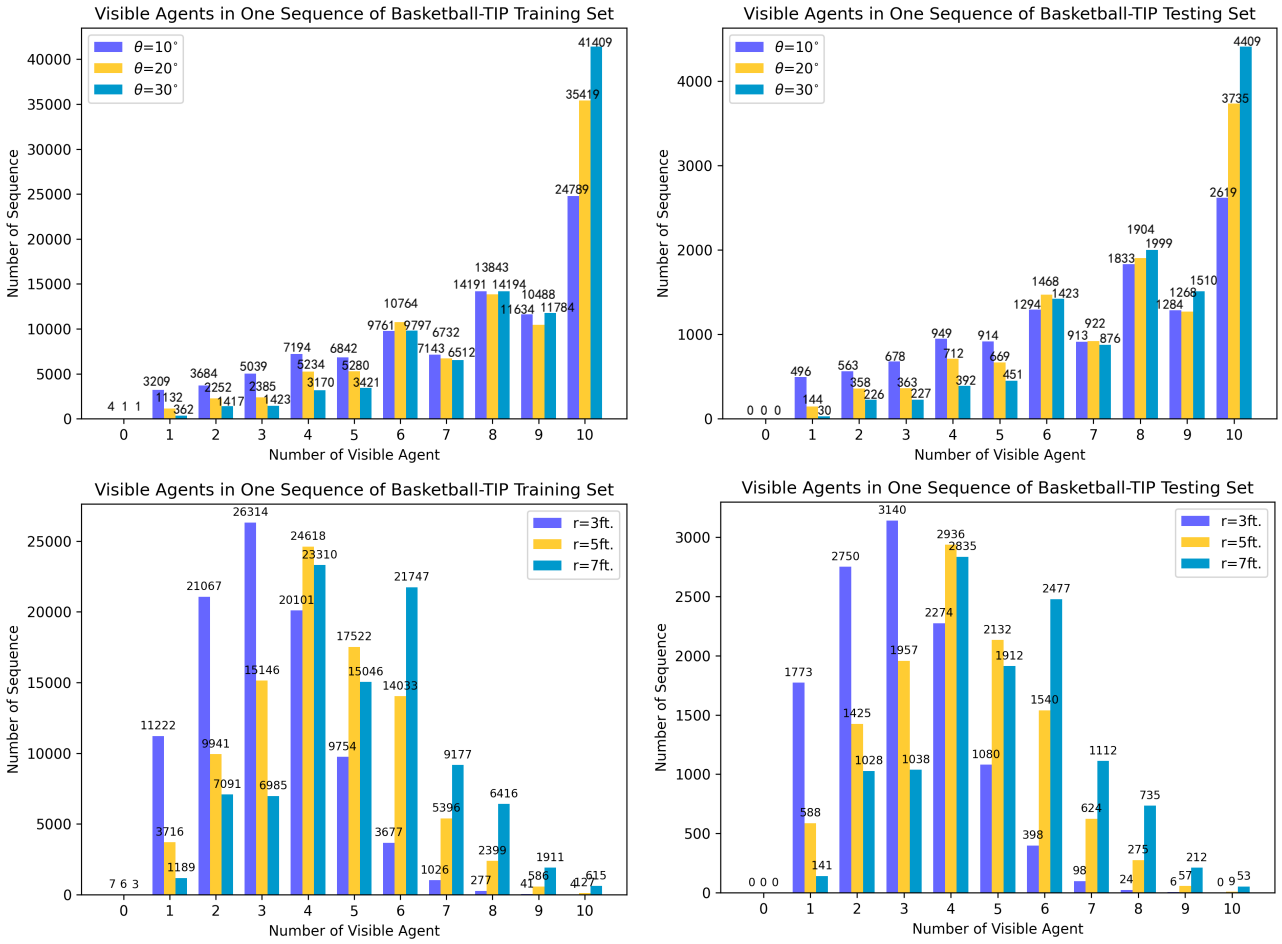


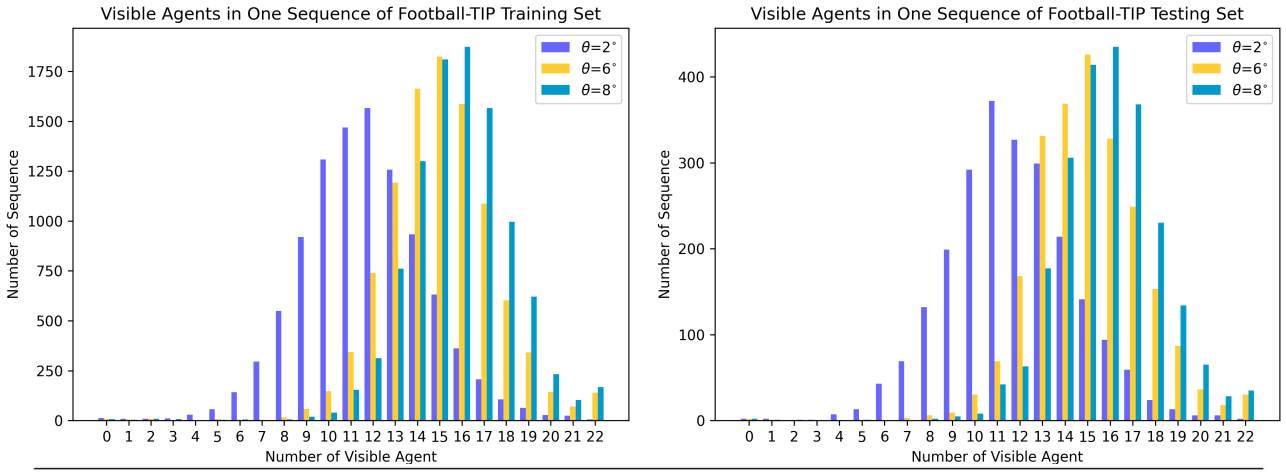
Figure 2. Histogram of visible agent numbers in one sequence of Basketball-TIP.

Datasets	Statistics On Average	$r=3$ ft. Train	$r=3$ ft. Test	$r=5$ ft. Train	$r=5$ ft. Test	$r=7$ ft. Train	$r=7$ ft. Test	$\theta=10°$ Train	$\theta=10°$ Test	$\theta=20°$ Train	$\theta=20°$ Test	$\theta=30°$ Train	$\theta=30°$ Test
Basketball-TIP	Visible Agent Number	3.14	2.99	4.31	4.17	5.07	4.95	7.10	6.81	7.82	7.57	8.29	8.09
	Visible Frame Length	3.93	3.93	5.22	5.21	6.92	6.91	7.53	7.35	12.35	12.06	16.60	16.30
	Missing Rate (%)	90.18	90.16	86.95	86.97	82.70	82.72	81.18	81.62	69.12	69.86	58.50	59.25
		$r=2$ yd. Train	$r=2$ yd. Test	$r=4$ yd. Train	$r=4$ yd. Test	$r=6$ yd. Train	$r=6$ yd. Test	$\theta=2°$ Train	$\theta=2°$ Test	$\theta=6°$ Train	$\theta=6°$ Test	$\theta=8°$ Train	$\theta=8°$ Test
Football-TIP	Visible Agent Number	11.67	11.69	14.97	14.99	15.83	15.85	6.20	6.31	10.87	10.92	13.45	13.55
	Visible Frame Length	3.56	3.61	10.59	10.62	17.43	17.45	4.48	4.47	11.30	11.36	14.28	14.31
	Missing Rate (%)	91.09	90.98	73.53	73.46	56.42	56.38	88.86	88.83	71.74	71.61	64.30	64.24

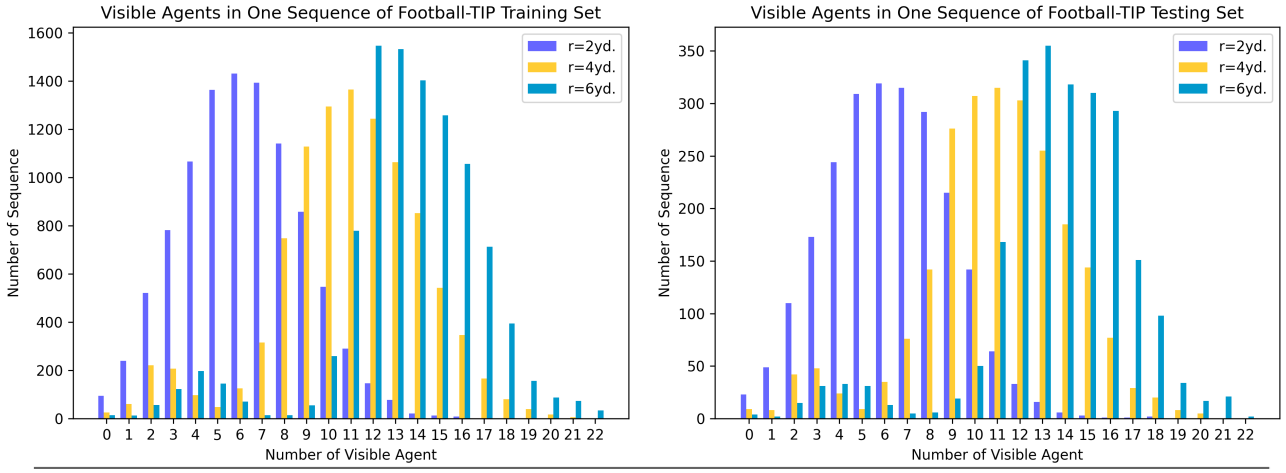
Table 1. Statistics of Basketball-TIP and Football-TIP.

the court/field size, while for Vehicle-TIP, we normalized the data into $[-1, 1]$ based on the scene size.

Statistics In sports games, not all players participate in the offensive or defensive rounds. Some players may be completely out of view for the entire observation duration. We calculate the number of visible agents (players) that appeared in at least one frame in each sequence during the observation. Tab. 1 shows the average results, where Visible Agent Number represents the average number of agents (out of 10 in Basketball-TIP and out of 22 in Football-TIP) that are visible in each sequence, Visible Frame Length represents the average length of visible observed frames (out of 40) of each agent, and Missing Rate represents the ratio of missing points number to total observed points of the



Visible Agent Number		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Training Set	$\theta = 2°$	12	8	8	11	29	56	142	296	550	919	1308	1469	1566	1258	934	631	362	207	106	64	28	24	6
	$\theta = 6°$	8	6	8	4	2	7	4	3	16	58	147	343	738	1193	1663	1824	1586	1087	602	341	143	71	140
	$\theta = 8°$	7	3	8	7	1	3	5	2	5	19	39	153	312	761	1300	1809	1873	1566	996	621	233	103	168
Testing Set	$\theta = 2°$	2	2	0	1	7	13	43	69	132	199	292	372	327	299	214	141	94	59	24	13	6	6	2
	$\theta = 6°$	2	1	1	0	0	1	0	3	6	9	30	69	168	331	369	426	328	249	153	87	36	18	30
	$\theta = 8°$	2	1	1	0	0	1	0	0	2	5	8	42	63	177	306	414	435	368	230	134	65	28	35



Visible Agent Number		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Training Set	$r = 2$ yd.	94	240	521	781	1066	1363	1431	1392	1141	858	546	290	147	78	21	13	8	2	2	0	0	0	0
	$r = 4$ yd.	26	61	221	207	98	48	126	315	748	1128	1294	1364	1243	1063	852	542	347	166	81	39	17	6	2
	$r = 6$ yd.	15	13	56	122	198	145	71	14	15	55	259	778	1546	1532	1402	1258	1056	713	394	157	88	73	34
Testing Set	$r = 2$ yd.	23	49	110	173	244	309	319	315	292	215	142	64	33	16	6	3	1	1	2	0	0	0	0
	$r = 4$ yd.	9	8	42	48	24	9	35	76	142	276	307	315	303	255	185	144	77	29	20	8	5	0	0
	$r = 6$ yd.	4	2	15	31	33	31	13	5	6	19	50	168	341	355	318	310	293	151	98	34	17	21	2

Figure 3. Histogram of visible agent numbers in one sequence of Football-TIP.

Dataset	Statistics	Easy						Ordinary						Hard					
		Training Set			Testing Set			Training Set			Testing Set			Training Set			Testing Set		
		#Ave	# Max	# Min	#Ave	# Max	# Min	#Ave	# Max	# Min	#Ave	# Max	# Min	#Ave	# Max	# Min	#Ave	# Max	# Min
Vehicle-TIP	Visible Agent Number	10.80	62	1	10.65	61	1	13.37	105	1	13.44	104	1	12.32	73	1	12.38	67	1
	Frame Length	69.55	90	1	69.06	90	1	71.28	90	1	71.16	90	1	66.88	90	1	67.14	90	1
	Observation Length	46.36	60	0	46.05	60	0	47.51	60	0	47.42	60	0	44.61	60	0	44.80	60	0
	Prediction Length	23.19	30	0	23.02	30	0	23.77	30	0	23.74	30	0	22.27	30	0	22.34	30	0
	Visible Frame Length	14.04	60	18	13.95	60	18	14.38	60	18	14.35	60	18	13.55	60	18	13.60	60	18
	Missing Rate (%)	69.71	–	–	69.70	–	–	69.74	–	–	69.74	–	–	69.63	–	–	69.64	–	–

Table 2. Statistics of Vehicle-TIP.

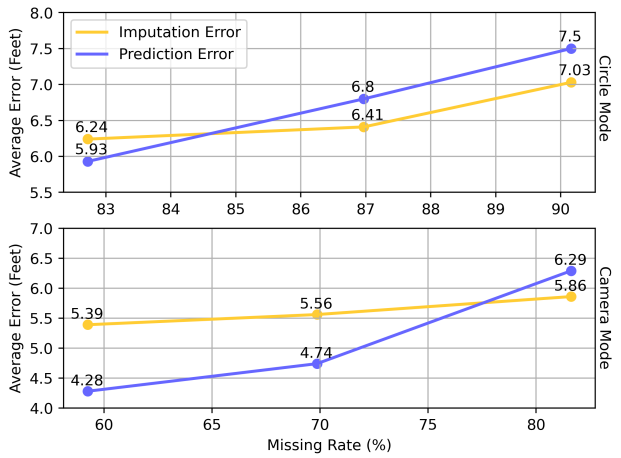


Figure 4. Relation between average error and missing rate on Basketball-TIP.

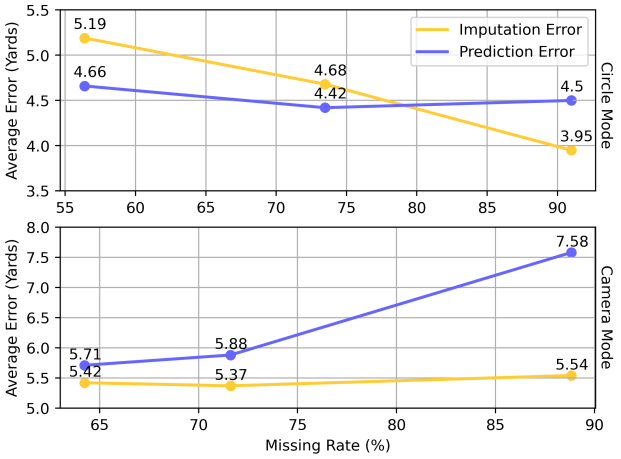


Figure 5. Relation between average error and missing rate on Football-TIP.

whole scenario. The average missing rates are over 50% of all 12 scenarios of Basketball-TIP and Football-TIP, posing a significant challenge for trajectory imputation. Fig. 2 and Fig. 3 show detailed results of Visible Agent Numbers on six scenarios.

In contrast, the number of agents (vehicles) in each sequence of Vehicle-TIP is not fixed, and the trajectory length of each agent varies. Tab. 2 provides statistics (mean, max, min) of Vehicle-TIP. In this table, Visible Agent Number represents the number of agents that are visible in each sequence, Frame Length represents the trajectory length of each agent, Observation Length represents the observed trajectory length (out of 60) of each agent, Prediction Length represents the to be predicted trajectory length (out of 30) of each agent, Visible Frame Length represents the visible trajectory length (out of the observed trajectory length) of each agent, and Missing Rate represents the ratio of missing points number to total observed points of the whole scenario. For all three scenarios, the average missing rates are over 69%, and the number of visible agents varies. Therefore, Vehicle-TIP is much more challenging than the sports datasets.

2. Experiments

In this section, we provide complete experiments including visualizations of our GC-VRNN.

2.1. Implementation Details of Baselines

For INAM [2], since there is no official implementation, we try our best to reproduce the method based on the implementation details reported in their paper.

For GMAT [4], we adopt their official implementation [3] in our Basketball-TIP and Football-TIP. For Basketball-TIP, the macro-intent labels are extracted as what they do in the paper, but we make imputations for all 10 players instead of only considering the offensive team. For Football-TIP, we segment the half-field into 12×11 grid of $5yd. \times 5yd.$ boxes and extract macro-intent labels accordingly. Other hyper-parameters such as feature dimensions, and the number of layers, are the same with [4].

For NAOMI [1], we adopt their official implementation [4] in our Basketball-TIP and Football-TIP. Instead of only imputing 5 players in their basketball dataset, we make imputations for a total of 10 players in Basketball-TIP and 22 players in Football-TIP. Other hyper-parameters such as feature dimensions, and the number of layers, are the same with their method.

We attempted to fine-tune the hyper-parameters during training for these baselines, but we did not observe any substantial improvements. As a result, we decided to use the hyper-parameters specified in the original papers for consistency.

2.2. Implementation details of Our Method

Within these three datasets, there are some sequences where some agents remain invisible throughout the entire observation period. In our implementation, we only conduct imputations and predictions for agents that have at least one observed frame.

2.3. Performance Analysis

We show the relation between the error and the missing rate on Basketball-TIP and Football-TIP in Fig. 4 and Fig. 5. For Basketball-TIP, the errors increase when the missing rate increases. However, it is interesting that the imputation error on Football-TIP (circle mode) decreases when the

[3] https://github.com/ezhan94/multiagent-programmatic-supervision.

[4] https://github.com/felixykliu/NAOMI.

Datasets	ID	GCL			$r = 3$ ft.		$r = 5$ ft.		$r = 7$ ft.		$\theta = 10°$		$\theta = 20°$		$\theta = 30°$	
		ST	DL	EC	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2
Basketball-TIP (In Feet)	1	✓			7.24	9.46	7.23	9.33	7.16	9.01	6.20	7.30	6.05	5.96	5.78	5.28
	2		✓		7.16	9.44	6.87	8.95	6.74	8.93	6.15	7.27	5.98	5.40	5.70	5.19
	3	✓	✓		7.11	9.33	6.84	7.89	6.55	8.54	5.98	7.18	5.60	5.27	5.51	5.08
	4	✓		✓	7.10	9.09	6.54	7.26	6.30	7.08	5.90	6.96	5.59	5.20	5.41	4.90
	5		✓	✓	7.07	8.10	6.49	6.91	6.26	6.04	5.87	6.32	5.59	5.11	5.88	6.09
	Ours	✓	✓	✓	**7.03**	**7.50**	**6.41**	**6.80**	**6.24**	**5.93**	**5.86**	**6.29**	**5.56**	**4.74**	**5.39**	**4.28**
	ID	GCL			$r = 2$ yd.		$r = 4$ yd.		$r = 6$ yd.		$\theta = 2°$		$\theta = 6°$		$\theta = 8°$	
		ST	DL	EC	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2
Football-TIP (In Yards)	1	✓			4.20	4.83	4.89	5.24	6.23	5.04	5.77	8.21	5.79	6.99	6.13	7.01
	2		✓		4.18	4.75	4.84	4.65	6.25	5.24	5.60	7.77	6.01	7.10	6.21	6.89
	3	✓	✓		4.14	4.71	4.76	4.63	6.18	5.23	5.67	7.75	5.58	7.12	5.71	6.30
	4	✓		✓	4.11	4.69	4.70	4.60	6.24	5.12	5.69	7.68	5.70	7.31	5.73	6.84
	5		✓	✓	4.10	4.59	4.76	4.58	6.24	4.83	5.65	7.66	6.11	7.33	6.09	6.78
	Ours	✓	✓	✓	**3.95**	**4.50**	**4.68**	**4.42**	**5.19**	**4.66**	**5.54**	**7.58**	**5.37**	**5.88**	**5.42**	**5.71**

Table 3. Component study of three GCLs in MS-GNN on Basketball-TIP and Football-TIP. ST denotes the Static Topology GCL, DL represents the Dynamic Learnable GCL, and EC represents the Edge Conditioned GCL.

Datasets	Variants	$r = 3$ ft.		$r = 5$ ft.		$r = 7$ ft.		$\theta = 10°$		$\theta = 20°$		$\theta = 30°$	
		I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2
Basketball-TIP (In Feet)	w/ IMP	7.21	28.74	6.74	29.68	6.58	30.06	5.94	31.84	5.77	31.67	5.56	31.57
	w/ PRE	29.15	7.78	29.10	7.78	29.07	6.48	30.61	6.83	30.98	5.46	32.35	4.76
	wo/ CON	7.11	16.38	6.57	14.25	6.40	13.32	5.96	15.05	5.86	13.14	5.48	13.96
	wo/ TD	7.25	7.70	6.70	7.30	6.37	6.26	5.98	6.61	5.69	5.10	5.51	4.52
	Ours	**7.03**	**7.50**	**6.41**	**6.80**	**6.24**	**5.93**	**5.86**	**6.29**	**5.56**	**4.74**	**5.39**	**4.28**
		$r = 2$ yd.		$r = 4$ yd.		$r = 6$ yd.		$\theta = 2°$		$\theta = 6°$		$\theta = 8°$	
		I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2
Football-TIP (In Yards)	w/ IMP	4.29	22.30	4.94	22.22	6.26	22.30	5.90	22.44	5.78	22.41	5.62	22.61
	w/ PRE	21.67	4.61	21.83	4.82	21.69	5.21	21.27	7.81	20.67	7.41	20.25	7.27
	wo/ CON	3.99	8.82	4.86	7.08	6.14	7.66	5.65	23.26	5.54	16.87	5.61	22.71
	wo/ TD	4.43	4.71	4.78	4.87	6.17	4.97	5.64	7.90	5.46	6.15	5.82	5.87
	Ours	**3.95**	**4.50**	**4.68**	**4.42**	**5.19**	**4.66**	**5.54**	**7.58**	**5.37**	**5.88**	**5.42**	**5.71**

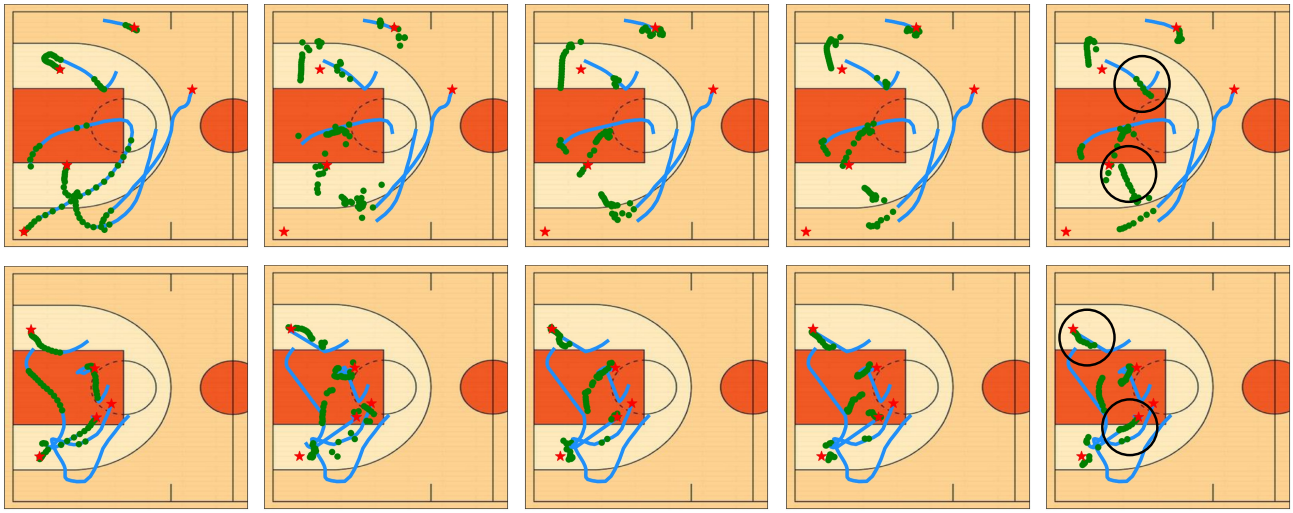
Table 4. Ablation study of the temporal decay module and the connection between the imputation and prediction stream on Basketball-TIP and Football-TIP.

Datasets	ID	GCL			Easy		Ordinary		Hard	
		ST	DL	EC	I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2
Vehicle-TIP (In Pixel)	1	✓			83.56	90.20	73.21	81.66	92.05	92.74
	2		✓		75.34	83.04	68.25	74.04	80.60	86.72
	3	✓	✓		73.56	80.04	66.98	69.34	79.22	86.83
	4	✓		✓	71.44	79.68	63.43	67.21	78.32	83.20
	5		✓	✓	69.32	75.34	60.16	66.79	76.51	81.40
	Ours	✓	✓	✓	**65.48**	**72.44**	**58.36**	**62.03**	**74.28**	**78.12**

Table 5. Component study of three GCLs in MS-GNN on Vehicle-TIP. ST denotes the Static Topology GCL, DL represents the Dynamic Learnable GCL, and EC represents the Edge Conditioned GCL.

Datasets	Variants	Easy		Ordinary		Hard	
		I-L_2	P-L_2	I-L_2	P-L_2	I-L_2	P-L_2
Vehicle-TIP (In Pixel)	w/ IMP	70.32	217.08	67.36	179.62	76.45	197.68
	w/ PRE	390.43	91.57	363.66	78.06	392.34	89.34
	wo/ CON	68.23	113.46	64.86	84.65	77.40	101.74
	wo/ TD	71.87	83.42	65.04	70.35	78.12	84.28
	Ours	**65.48**	**72.44**	**58.36**	**62.03**	**74.28**	**78.12**

Table 6. Ablation study of the temporal decay module and the connection between the imputation and prediction stream on Vehicle-TIP.



(a) GT (b) GMAT (c) NAOMI (d) INAM (e) Ours

Figure 6. Visualizations of imputed results. The red star denotes the starting point, the blue line represents the visible observation, and the green point represents the missing point.

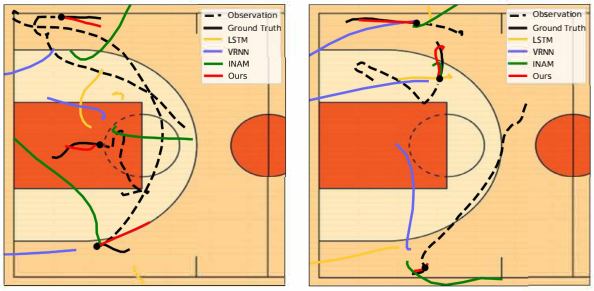


Figure 7. Visualizations of predicted results on Basketball-TIP ($\theta = 30^{\circ}$). The predicted results of ours are in the red line.

missing rate increases. This point is worth exploring to further understand and explain the missing patterns under different situations.

2.4. Ablation Study

Tab. 3 and Tab. 4 show the complete results of all scenarios on Basketball-TIP and Football-TIP. It can be seen from Tab. 3 that three GCLs all contribute to the final results, especially our designed Edge Conditioned GCL. This validates that our proposed EC GCL is able to uncover the spatial missing patterns and extract effective features for better spatial feature representations. In Tab. 4, "wo/ TD" means we remove the temporal decay module for comparison, "w/ IMP" means we only make imputations, and "w/ PRE" means we only make predictions. We also cut off the connection by introducing two different RNNs for imputation and prediction separately, which we refer to as "wo/ CON". It can be seen that when we make imputation or prediction separately, the performance drops compared with our complete model. In addition, when comparing with model "wo/ CON", the performance of both imputation and prediction drops, especially the prediction performance. One possible reason is that, as the conditions of the prediction task, the imputation information is relatively more important than the prediction information. When comparing with model "wo/ TD", it can be concluded that our designed temporal decay module is effective to learn the temporal missing patterns of incomplete observations.

Tab. 5 and Tab. 6 demonstrate the results on Vehicle-TIP. Although the statistics of Vehicle-TIP are different from those of sports datasets, our proposed method is still effective in all three different scenarios. Component study and ablation study also verify the effectiveness of our designed GCLs and TD module empirically. In addition, it also proves that considering the imputation task and the prediction task under a unified framework is a necessity for better performance on both tasks.

2.5. Qualitative Results

We provide visualized examples of our method and some baselines on Basketball-TIP ($\theta = 30^{\circ}$) in Fig. 6 and Fig. 7. It can be seen from Fig. 6 that our proposed method outperforms other baselines in imputation, some trajectory details are circled to highlight. However, there are still some missing values not well imputed, especially when the missing values are at the beginning of the trajectory. In Fig. 7, it is obvious that our predicted results are much better than others. Since our method handles the imputation and prediction simultaneously, better imputation results can be beneficial to future trajectory prediction.

	Path-L (ft.)	OOB (10^{-3})	Step (ft.)	Path-D(ft.)
GT	0.696	0	0.118	25.195
NAOMI [1]	2.118±0.020	2.137±0.085	0.624±0.038	68.900±5.144
INAM [2]	1.644	0.982	0.466	55.459
Ours	**1.121**±0.009	**0.138**±0.019	**0.327**±0.009	**47.031**±4.587

Table 7. Results using four additional metrics on Basketball-TIP ($r = 3$ ft.). The closer to GT (ground truth), the better.

2.6. Results on Additional Metrics

In [1] and [2], the evaluation includes four additional metrics: Path-L, OOB, Step, and Path-D. To be consistent, we also assess our method on Basketball-TIP ($r = 3$ ft.) and present the results in Tab. 7. As our datasets already exclude out-of-bound samples, OOB's ground truth is 0. Our method achieves better performance on these metrics, and we provide the standard deviation by inferring (sampling from predicted distribution) 20 times for further analysis. Note that the method NAOMI [1] may involve stochastic dynamics, while the method INAM [2] employs the deterministic loss (L2), resulting in a specific numerical result.

3. Limitations and Broader Impact

In our work, we point out a new direction in jointly learning trajectory imputation and prediction, we also curate and benchmark three practical datasets for further research in this domain. Naturally, due to our work being the pioneer of this joint problem, there are still some limitations.

Limitations. Our GC-VRNN has a limitation in that it is a single-modality method. A potential future direction of our work is to extend it to a multi-modality method, which could enhance the accuracy and robustness. Another limitation we have identified is that the imputation performance decreases when there is a long and continuous instance of missing data during the observed trajectory, particularly when it occurs at the beginning of the observation. We believe that exploring better solutions to address these situations would be a valuable direction for future research.

Broader Impact. The task of trajectory prediction plays a crucial role in many applications, ranging from autonomous driving to motion capture. We have demonstrated that our proposed method of combining the tasks of trajectory imputation and prediction mutually supports one another, enabling better performance for both tasks. We believe that our work, as well as our datasets, can serve as a useful baseline for follow-up works to explore this fertile area of research.

References

[1] Yukai Liu, Rose Yu, Stephan Zheng, Eric Zhan, and Yisong Yue. Naomi: Non-autoregressive multiresolution sequence imputation. 32, 2019. 4, 7

[2] Mengshi Qi, Jie Qin, Yu Wu, and Yi Yang. Imitative non-autoregressive modeling for trajectory forecasting and imputation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12736–12745, 2020. 4, 7

[3] ShiJie Sun, Naveed Akhtar, XiangYu Song, HuanSheng Song, Ajmal Mian, and Mubarak Shah. Simultaneous detection and tracking with motion modelling for multiple object tracking. In *Proceedings of the European Conference on Computer Vision*, pages 626–643, 2020. 1

[4] Eric Zhan, Stephan Zheng, Yisong Yue, Long Sha, and Patrick Lucey. Generating multi-agent trajectories using programmatic weak supervision. In *Proceedings of the International Conference on Learning Representations*, 2018. 1, 4