

Appendix of Skeleton-Parted Graph Scattering Networks for 3D Human Motion Prediction

Maosen Li¹, Siheng Chen^{1,2} ✉, Zijing Zhang³, Lingxi Xie⁴,
Qi Tian⁴, and Ya Zhang^{1,2} ✉

¹ Cooperative Medianet Innovation Center, Shanghai Jiao Tong University

² Shanghai AI laboratory

³ Zhejiang University

⁴ Huawei Cloud & AI.

{maosen_li, sihengc, ya_zhang}@sjtu.edu.cn, zijing_zhang@163.com,
198808xc@gmail.com, tian.qi1@huawei.com

I. Some details of Model

About SPGSN architectures. The entire SPGSN is constructed by cascaded MPGSBs, where we note that all the MPGSBs use distinct parameters without weight-sharing. The non-shared MPGSBs flexibly extract multi-level features to enrich the representation, just like ResNet uses distinct residual blocks to learn the hierarchical features.

Since the MPGSBs are formed in cascade, the input of the second MPGSB is the output of the first MPGSB, \mathbf{H}' , integrating the features from spatial and spectrum domains. We use IDCT to transform the output features to the predicted sequences.

About experimental scenarios. The coordinates of all the body-joints denote the relative positions of distinct body-joints to a central joint (e.g. chest). In this benchmark, all the studies target to predict the pure poses, regardless of the body positions in a global space.

II. Protocol Modification of STSGCN

Published in ICCV 2021, STSGCN [9] is one of the practical algorithms of human motion prediction, which leverages both spatial and temporal graph convolution on the different network stages. According to the official codes⁵, however, the experimental protocols have two main differences from most previous works [8, 4, 2, 3, 7, 5, 6], leading to their reported results having very different means:

- First, to evaluate the prediction performance at each timestamp, STSGCN learns a specific and unique model; that is, STSGCN uses different pre-trained models to generate poses at different timestamps such as 80ms and 160ms. However, most previous models use only one model to forecast the whole pose sequences directly and just calculate the errors at different timestamps based on

⁵ <https://github.com/FraLuca/STSGCN>

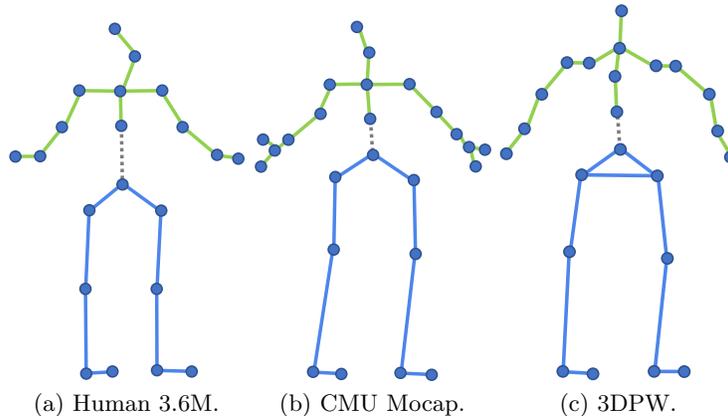


Fig. 1. The body separation strategies on various datasets, where we color the bones on upper bodies in green and the bones on lower bodies in blue.

the same model. The difference is that previous models generate one sequence only once to achieve low errors at any timestamps, while STSGCN builds specific models towards specific prediction timestamps, and the models are supervised by the target sequences within the corresponding timestamps.

- Second, to present the MPJPE at each timestamp, previous methods directly calculate the instant errors at the corresponding timestamp. However, according to the STSGCN’s test codes, at a certain timestamp, STSGCN presents the average MPJPE before this timestamp instead of the instant one; that is, the 1000ms-MPJPE reported by STSGCN is the average MPJPE across all the prediction frames from the 0ms to the 1000ms.

To achieve a fair comparison, we modify the STSGCN codes to directly generate the whole sequence using only one model and reported the instant MPJPE at each timestamp in our manuscript.

III. The Body Separation

Here we illustrate the body separation strategies on the three datasets: Human 3.6M, CMU Mocap and 3DPW. We show the decomposed upper bodies and lower bodies in Fig. 1. We color the bones on the separated upper bodies in green, and the bones on the lower bodies in blue. Different datasets show different predefined body structures, thus we should use specific body separation strategies to obtain valid body-parts.

IV. Effects of the spectrum aggregation

To study the proposed graph spectrum aggregation, we first compare our method to two common feature aggregation operations. First, directly averaging all

Table 1. SPGSNs with different spectrum aggregation methods.

Model	Average MAE					
	80	160	320	400	560	1000
SPGSN (average)	10.54	22.48	47.73	58.91	79.20	112.14
SPGSN (self-att.)	10.82	22.79	47.65	59.13	79.61	114.08
SPGSN	10.44	22.33	47.07	58.26	77.40	109.64

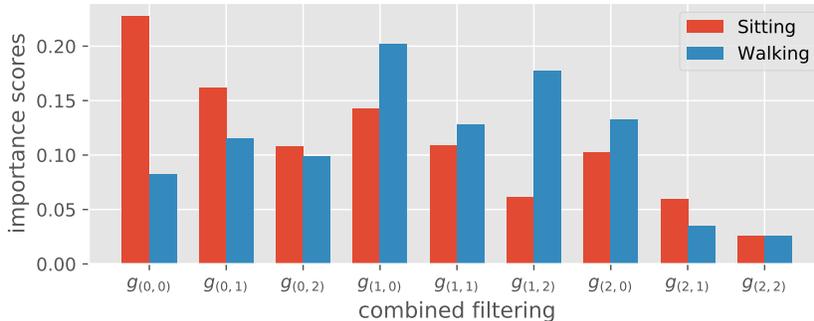


Fig. 2. Spectrum importance scores on different actions.

the channels; second, summing up channels with an inner-product-based inter-channel self-attention. Table 1 presents the prediction results of SPGSN with various aggregation methods on H3.6M. We see that, the proposed graph spectrum aggregation outperforms the two variants. Note that, the self-attention mixes up various channels before aggregation but loses the spectrum diversity, leading to larger errors.

Next, we visualize the learned spectrum importance of different actions. We show the importance scores calculated by the last MPGSB on actions ‘Sitting’ and ‘Walking’; see Fig. 2, where the x-axis denotes the graph scattering branch with different filter combinations. We see that, different actions lead to different importance distributions. For ‘Sitting’, the poses show slow movements, thus the low-pass features dominate the spectrum to stabilize pattern learning; for ‘Walking’, poses keep large movements, thus some high-frequency bands are emphasized.

V. Feature Responses of the Adaptive Graph Scattering Decomposition

To study the effects of the proposed adaptive graph scattering decomposition and provide a more interpretable analysis, we consider showing the feature responses of the adaptive graph scattering decomposition given the input motion sequence. We select a sample of the action ‘direction’, where the subject has a static body torso as well as performs hand raising and leg swaying. Since each body joint carries a DCT-formed multidimensional feature vector, we collect the second

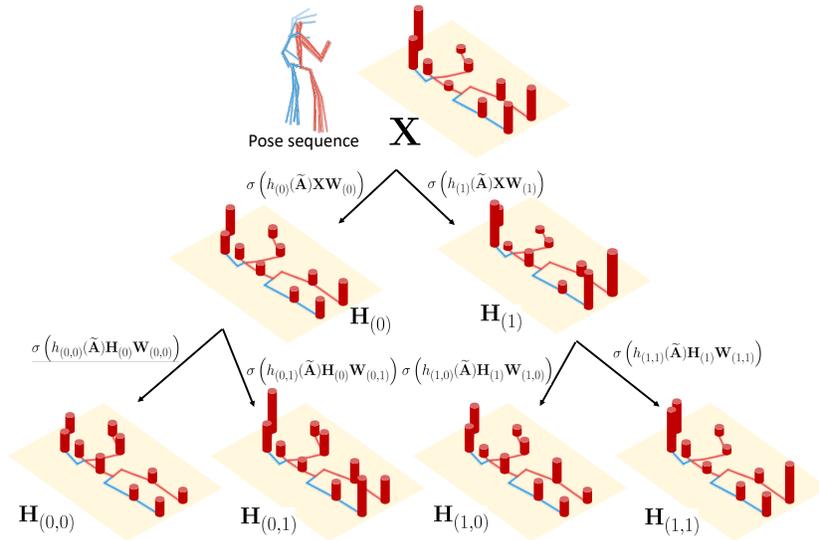


Fig. 3. Feature responses after the proposed adaptive graph scattering decomposition of a pose sequence of the action 'directions'.

DCT coefficients of all the joints as a graph signal, on which we perform the graph scattering. The second DCT coefficients carry large values at the highly dynamic joints but carry small values at relatively fixed joints, according to empirical tests. Therefore, it is convenient to reflect on how much information about the dynamic joints is preserved. Moreover, the selected graph signal also shows the difference between nodes, which can reflect graph scattering extract the feature at various graph spectral bands, and better demonstrate the performance of the model.

The feature responses at each channel in each layer of the graph scattering decomposition are visualized in Fig. 3. We crop a sub-tree of graph scattering decomposition from the pre-trained model, where each non-leaf channel has only two child branches instead of three child branches in the original model, for clearer presentation. We use two filters at each non-leaf channel; that is $h_{(0)}(\tilde{\mathbf{A}})$, a basic graph convolution, and $h_{(1)}(\tilde{\mathbf{A}})$, an adaptive graph wavelet; see Eq. (2) in our manuscript. We also apply the feature transform parameter \mathbf{W} and nonlinear activation $\sigma(\cdot)$ at each channel. In Fig. 3, the amplitude of the graph signals is plotted on the corresponding body-joints in the form of a histogram.

Fig. 3 reveals that, given the motion data with large signal values at moving joints and small values at static joints, the graph convolution tends to average their features, close their difference and derive a smoother signal; especially in $\mathbf{H}_{(0,0)}$, the differences between the high-speed limbs and the fixed torso are hardly reflected. This means that always using deep graph convolution will not be able to accurately describe the characteristics of different body joints, and

it is difficult to provide rich and detailed information for pattern recognition. As for graph wavelet filtering, the difference between joints could be effectively preserved; that is, in this sample, the graph wavelet filtering focuses on the high-frequency information to push joints farther and magnify their gaps. Note that, in the graph scattering, all the filtered channels covering both low-frequency and high-frequency bands are important, since the large spectrum contains significantly comprehensive information to promote pattern learning and motion prediction.

VI. Effects of Skip-connections across the Model

Table 2 compares SPGSN with and without the skip-connections on the H3.6M dataset. Skip-connections force to learn the feature displacements from observations to predictions, obtaining much smaller MPJPEs (especially in the short-term prediction) than the model without skip-connections that directly generates the poses.

VII. More Quantitative Results

Here we present the prediction results on all the contained actions in H3.6M and CMU Mocap for both short-term and long-term prediction. These results provide sufficient information for a detailed comparison of the algorithm development in future works.

First, we present the MPJPE of various models on H3.6M for short-term motion prediction, where the detailed results of any actions are shown in Table 3. We see that, the proposed SPGSN with spatial and spectrum feature decomposition could effectively outperform baselines at most actions, as well as achieve the best results in terms of the average prediction errors. Moreover, using the separated body parts could improve the prediction compared to the model variant using only the whole body.

Also, we show the prediction MPJPEs on H3.6M for long-term motion prediction, which are presented in Table 4. We see that, SPGSN obtains the effective performance in long-term motion prediction, since SPGSN shows lower MPJPEs on most actions as well as the lowest average MPJPE over all actions.

Finally, we present the MPJPEs of various methods on all the actions of CMU Mocap; see Table 5. The experiment results also verify the effectiveness of the proposed SPGSN.

Table 2. Prediction with/without using skip-connections.

	Average MPJPE					
millisecond	80	160	320	400	560	1000
No-skip	23.19	31.81	54.15	64.05	81.90	115.22
Skip	10.44	22.33	47.07	58.26	77.40	109.64

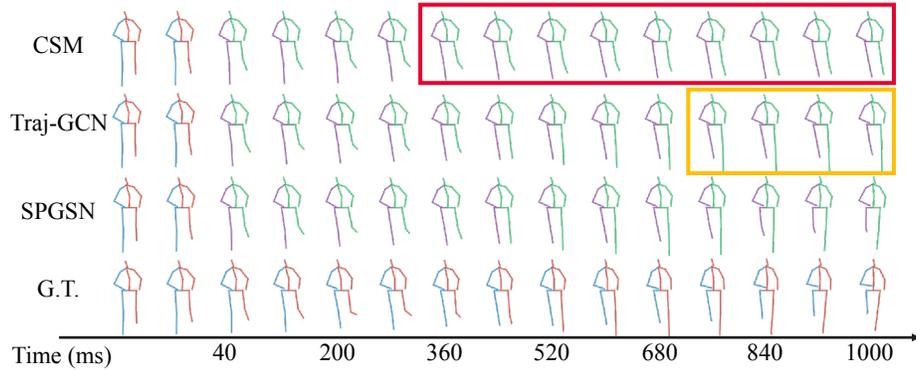


Fig. 4. The generated motion sequence of ‘walking dog’ in H3.6M

VIII. More Qualitative Results

Besides the Figure 5 in our manuscript, here we illustrate the predicted samples on the other action: walking dog, where we compare the proposed SPGSN to previous CSM [4] and Traj-GCN [7]; see Figure 4. We see that, compared to the baselines, SPGSN generates more precise and reasonable future poses that are close to the ground truth in both short-term and long-term. For CSM, the predicted motion shows large errors after the 360th ms; that is, the poses collapse to static where the right feet (green) are always lifted; see the red box. As for Traj-GCN, the generated poses keep static to lean to the left after the 720th ms; see the orange box.

References

1. Dang, L., Nie, Y., Long, C., Zhang, Q., Li, G.: Msr-gcn: Multi-scale residual graph convolution networks for human motion prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 11467–11476 (October 2021)
2. Gui, L., Wang, Y., Liang, X., Moura, J.: Adversarial geometry-aware human motion prediction. In: ECCV. pp. 786–803 (Sept 2018)
3. Guo, X., Choi, J.: Human motion prediction via learning local structure representations and temporal dependencies. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 2580–2587 (2019)
4. Li, C., Zhang, Z., Sun Lee, W., Hee Lee, G.: Convolutional sequence to sequence model for human dynamics. In: CVPR (June 2018)
5. Li, M., Chen, S., Zhao, Y., Zhang, Y., Wang, Y., Tian, Q.: Dynamic multiscale graph neural networks for 3d skeleton based human motion prediction. In: CVPR (June 2020)
6. Mao, W., Liu, M., Salzmann, M.: History repeats itself: Human motion prediction via motion attention. In: ECCV (Aug 2020)
7. Mao, W., Liu, M., Salzmann, M., Li, H.: Learning trajectory dependencies for human motion prediction. In: ICCV (Oct 2019)

8. Martinez, J., Black, M., Romero, J.: On human motion prediction using recurrent neural networks. In: CVPR. pp. 4674–4683 (July 2017)
9. Sofianos, T., Sampieri, A., Franco, L., Galasso, F.: Space-time-separable graph convolutional network for pose forecasting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 11209–11218 (October 2021)

Table 3. Prediction MPJPEs of various models for short-term motion prediction on H3.6M and the average MPJPEs across all the actions.

Motion millisecond	Walking				Eating				Smoking				Discussion			
	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Res-sup [8]	29.36	50.82	76.03	81.52	16.84	30.60	56.92	68.65	22.96	42.64	70.24	83.68	32.94	61.18	90.92	96.19
CSM [4]	21.70	43.56	66.29	75.48	14.50	26.13	47.47	55.63	19.42	37.70	62.49	68.55	26.35	53.41	79.12	83.01
SkelNet [3]	20.49	34.36	59.64	68.76	11.80	22.38	39.88	48.11	11.33	23.71	45.30	52.85	21.79	40.24	65.93	77.91
DMGNN [5]	17.32	30.67	54.56	65.20	10.96	21.39	36.18	43.88	8.97	17.62	32.05	40.30	17.33	34.78	61.03	69.80
Traj-GCN [7]	12.29	23.03	39.77	46.12	8.36	16.90	33.19	40.70	7.94	16.24	31.90	38.90	12.50	27.40	58.51	71.68
HisRep [6]	10.53	19.96	34.88	42.05	7.39	15.53	31.26	38.58	7.17	14.54	28.83	35.67	10.89	25.19	56.15	69.30
MSR-GCN [1]	12.16	22.65	38.64	45.24	8.39	17.05	33.03	40.43	8.02	16.27	31.32	38.15	11.98	26.76	57.08	69.74
STSGCN* [9]	16.26	24.63	40.06	45.94	14.32	22.14	37.91	45.03	13.10	20.20	37.71	44.65	14.33	24.28	52.62	68.53
SPGSN (Ibody)	10.13	19.51	35.52	44.67	7.13	15.02	31.87	41.18	6.83	13.94	28.77	36.78	10.42	23.90	54.13	69.99
SPGSN	10.14	19.39	34.80	41.47	7.07	14.85	30.48	37.91	6.72	13.79	27.97	34.61	10.37	23.79	53.61	67.12
Motion millisecond	Directions				Greeting				Phoning				Posing			
	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Res-sup [8]	35.36	57.27	76.30	87.67	34.46	63.36	124.60	142.50	37.96	69.32	115.00	126.73	36.10	69.12	130.46	157.08
CSM [4]	27.07	44.72	63.94	75.37	28.63	60.69	119.25	139.92	25.66	40.13	63.06	78.01	22.02	40.34	93.65	119.32
SkelNet [3]	16.06	27.12	62.97	72.75	24.71	56.90	111.74	134.25	18.91	34.69	59.34	72.09	18.51	34.67	80.83	106.39
DMGNN [5]	13.14	24.62	64.68	81.86	23.30	50.32	107.30	132.10	12.47	25.77	48.08	58.29	15.27	29.27	71.54	96.65
Traj-GCN [7]	8.97	19.87	43.35	53.74	18.65	38.68	77.74	93.39	10.24	21.02	42.54	52.30	13.66	29.89	66.62	84.05
HisRep [6]	7.77	18.23	41.34	51.61	15.47	34.04	73.77	88.90	9.78	20.98	39.81	50.87	13.23	27.70	63.68	81.82
MSR-GCN [1]	8.61	19.65	43.28	53.82	16.48	36.95	77.32	93.38	10.10	20.74	41.51	51.26	12.79	29.38	66.95	85.01
STSGCN* [9]	14.24	24.27	44.24	53.21	15.02	30.70	67.11	87.63	14.88	21.40	46.55	52.03	15.01	25.69	58.38	73.08
SPGSN (Ibody)	7.38	17.48	40.54	53.09	15.16	33.61	71.89	88.74	8.78	18.50	39.85	51.53	10.92	25.46	61.38	78.87
SPGSN	7.35	17.15	39.80	50.25	14.64	32.59	70.64	86.44	8.67	18.32	38.73	48.46	10.73	25.31	59.91	76.46
Motion millisecond	Purchases				Sitting				Sitting Down				Taking Photo			
	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Res-sup [8]	36.33	60.30	86.53	95.92	42.55	81.40	134.70	151.78	47.28	85.95	145.75	168.86	26.10	47.61	81.40	94.73
CSM [4]	25.69	47.85	82.49	93.90	22.25	34.67	58.72	75.80	23.67	51.76	102.93	119.47	20.29	38.92	61.14	77.40
SkelNet [3]	21.04	40.59	79.97	88.66	15.55	28.70	49.35	62.87	17.64	38.88	85.30	101.71	15.74	32.83	48.62	63.90
DMGNN [5]	21.35	38.71	75.67	82.74	11.92	25.11	44.59	50.20	14.95	32.88	77.06	93.00	13.61	28.95	45.99	58.76
Traj-GCN [7]	15.60	32.78	65.72	79.25	10.62	21.90	46.33	57.91	16.14	31.12	61.47	75.46	9.88	20.89	44.95	56.58
HisRep [6]	14.63	32.81	65.18	78.27	10.21	20.36	43.68	53.62	15.54	29.97	59.31	72.25	9.09	20.10	44.60	55.72
MSR-GCN [1]	14.75	32.39	66.13	79.64	10.53	21.99	46.26	57.80	16.10	31.63	62.45	76.84	9.89	21.01	44.56	56.30
STSGCN* [9]	15.26	26.26	63.45	74.25	15.19	22.95	46.82	58.34	16.70	28.05	56.15	72.03	16.61	24.84	45.98	61.79
SPGSN (Ibody)	12.78	28.86	62.59	77.01	9.25	19.58	43.47	56.32	14.34	28.10	58.23	74.44	8.72	18.95	42.62	55.22
SPGSN	12.75	28.58	61.01	74.38	9.28	19.40	42.25	53.56	14.18	27.72	56.75	70.74	8.79	18.90	41.49	52.66
Motion millisecond	Waiting				Walking Dog				Walking Together				Average			
	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Res-sup [8]	30.62	57.82	106.22	121.45	64.18	102.10	141.07	164.35	26.79	50.07	80.16	92.23	34.66	61.97	101.08	115.49
CSM [4]	19.14	33.11	69.72	95.21	58.67	97.36	129.74	158.57	22.60	38.51	71.13	84.37	25.17	45.92	78.08	93.33
SkelNet [3]	16.31	29.90	63.86	84.59	54.61	93.23	124.12	155.79	19.01	32.40	63.35	73.18	20.23	38.04	69.35	84.25
DMGNN [5]	12.20	24.17	59.62	77.54	47.09	93.33	160.13	171.20	14.34	26.67	50.08	63.22	16.95	33.62	65.90	79.65
Traj-GCN [7]	11.43	23.99	50.06	61.48	23.39	46.17	83.47	95.96	10.47	21.04	38.47	45.19	12.68	26.06	52.27	63.51
HisRep [6]	10.58	23.75	49.30	60.26	21.77	43.38	78.53	90.21	9.88	19.51	35.91	42.60	11.60	24.40	49.75	60.78
MSR-GCN [1]	10.68	23.06	48.25	59.23	20.65	42.88	80.35	93.31	10.56	20.92	37.40	43.85	12.11	25.56	51.64	62.93
STSGCN* [9]	16.30	27.33	48.12	59.79	16.48	37.63	70.60	86.33	11.38	22.39	39.90	47.48	15.34	25.52	50.64	60.61
SPGSN (Ibody)	9.24	20.02	43.80	56.80	18.31	38.12	73.63	86.74	8.91	18.46	34.88	42.98	10.55	22.63	48.21	60.96
SPGSN	9.21	19.79	43.10	54.14	17.83	37.15	71.74	84.91	8.94	18.19	33.84	40.88	10.44	22.33	47.07	58.26

Table 4. Prediction MPJPEs of methods for long-term prediction on H3.6M and the average MPJPEs across all the actions.

Motion millisecond	Walking		Eating		Smoking		Discussion		Directions		Greeting		Phoning		Posing	
	560	1k	560	1k	560	1k	560	1k	560	1k	560	1k	560	1k	560	1k
Res-sup [8]	81.73	100.68	79.87	100.20	94.83	137.44	121.30	161.70	110.05	152.48	156.32	184.29	143.92	186.79	165.41	236.79
CSM [4]	78.04	94.58	72.14	96.87	66.61	89.80	108.20	142.13	97.80	132.82	151.50	175.37	83.46	127.55	137.72	210.90
SkelNet [3]	73.58	91.84	63.58	90.88	58.96	80.53	98.28	135.68	93.77	124.89	148.38	168.06	75.42	113.34	131.90	196.21
Traj-GCN [7]	54.05	59.75	53.39	77.75	50.74	72.62	91.61	121.53	71.01	101.79	113.87	145.19	69.55	104.19	114.52	171.10
DMGNN [5]	71.36	85.82	58.11	86.66	50.85	72.15	81.90	106.32	102.06	135.75	144.51	170.54	71.33	108.37	125.45	188.18
MSR-GCN [1]	52.72	63.05	52.54	77.11	49.45	71.64	88.59	117.59	71.18	100.59	116.24	147.23	68.28	104.36	116.26	174.33
STSGCN* [9]	57.64	66.74	58.46	75.08	55.55	74.13	84.20	107.74	75.61	109.89	79.32	103.75	79.19	109.88	80.82	107.58
SPGSN	46.89	53.59	49.76	73.39	46.68	68.62	89.68	118.55	70.05	100.52	110.98	143.21	66.70	102.52	110.34	165.39
Motion millisecond	Purchases		Sitting		SittingDown		TakingPhoto		Waiting		WalkingDog		WalkingToge		Average	
	560	1k	560	1k	560	1k	560	1k	560	1k	560	1k	560	1k	560	1k
Res-sup [8]	119.36	176.92	166.20	185.16	197.09	223.58	107.03	162.38	126.70	153.14	173.61	202.31	94.51	110.48	129.19	164.96
CSM [4]	113.44	167.61	98.04	134.70	148.87	196.75	94.75	144.52	106.03	125.60	168.71	183.42	93.90	106.60	107.94	141.95
SkelNet [3]	109.51	155.72	84.76	127.11	125.89	184.24	86.43	130.41	90.49	112.02	166.65	176.79	79.07	99.25	99.11	132.46
Traj-GCN [7]	99.24	137.28	77.63	118.36	100.91	157.32	78.73	120.06	79.08	103.83	138.24	150.63	51.67	61.10	81.07	113.01
DMGNN [5]	104.86	146.09	75.51	115.44	118.04	174.05	78.38	123.65	85.54	113.68	183.20	210.17	70.46	86.93	93.57	127.62
MSR-GCN [1]	101.63	139.15	78.19	120.02	102.83	155.45	77.94	121.87	76.33	106.25	111.87	148.21	52.93	65.91	81.13	114.18
STSGCN* [9]	87.10	119.26	82.32	119.83	92.60	129.67	87.70	119.79	86.41	118.04	86.79	118.33	75.33	95.83	80.66	113.33
SPGSN	96.53	133.88	75.00	116.24	98.94	149.88	75.58	118.22	73.50	103.62	102.37	137.96	49.84	60.86	77.40	109.64

Table 5. Prediction MPJPEs of different methods on the 8 actions of CMU Mocap for both short-term and long-term prediction. We also present the average prediction results across all the actions.

Motion millisecond	Basketball					Basketball Signal					Directing Traffic					Jumping				
	80	160	320	400	1000	80	160	320	400	1000	80	160	320	400	1000	80	160	320	400	1000
Res-sup. [8]	15.45	26.88	43.51	49.23	88.73	20.17	32.98	42.75	44.65	60.57	20.52	40.58	75.38	90.36	153.12	26.85	48.07	93.50	108.90	162.84
DMGNN [5]	15.57	28.72	59.01	73.05	138.62	5.03	9.28	20.21	26.23	52.04	10.21	20.90	41.55	52.28	111.23	31.97	54.32	96.66	119.92	224.63
Traj-GCN [7]	11.68	21.26	40.99	50.78	97.99	3.33	6.25	13.58	17.98	54.00	6.92	13.69	30.30	39.97	114.16	17.18	32.37	60.12	72.55	127.41
MST-GCN [1]	10.28	18.94	37.68	47.03	86.96	3.03	5.68	12.35	16.26	47.91	5.92	12.09	28.36	38.04	111.04	14.99	28.66	55.86	69.05	124.79
STSGCN* [9]	12.56	23.04	41.92	50.33	94.17	4.72	6.69	14.53	17.88	49.52	6.41	12.38	29.05	38.86	109.42	17.52	31.48	58.74	72.06	127.40
SPGSN	10.24	18.54	38.22	48.68	89.58	2.91	5.25	11.31	15.01	47.31	5.52	11.16	25.48	37.06	108.14	14.93	28.16	56.72	71.16	125.20
Motion millisecond	Running					Soccer					Walking					Washing Window				
	80	160	320	400	1000	80	160	320	400	1000	80	160	320	400	1000	80	160	320	400	1000
Res-sup. [8]	25.76	48.91	88.19	100.80	158.19	17.75	31.30	52.55	61.40	107.37	44.35	76.66	126.83	151.43	194.33	22.84	44.71	86.78	104.68	202.73
DMGNN [5]	17.42	26.82	38.27	40.08	46.40	14.86	25.29	52.21	65.42	111.90	9.57	15.53	26.03	30.37	67.01	7.93	14.68	33.34	44.24	82.84
Traj-GCN [7]	14.53	24.20	37.44	41.10	51.73	13.33	24.00	43.77	53.20	108.26	6.62	10.74	17.40	20.35	34.41	5.96	11.62	24.77	31.63	66.85
MST-GCN [1]	12.84	20.42	30.58	34.42	48.03	10.92	19.50	37.05	46.38	99.32	6.31	10.30	17.64	21.12	39.70	5.49	11.07	25.05	32.51	71.30
STSGCN* [9]	16.70	27.58	36.15	36.42	55.34	13.49	25.24	39.87	51.58	109.63	7.18	10.99	17.84	22.61	44.12	6.79	12.10	24.92	36.66	69.48
SPGSN	10.75	16.67	26.07	30.08	52.92	10.86	18.99	35.05	45.16	99.51	6.32	10.21	16.34	20.19	34.83	4.86	9.44	21.50	28.37	65.08
Motion millisecond	Average																			
	80	160	320	400	1000															
Res-sup. [8]	24.21	43.75	76.19	88.93	139.00															
DMGNN [5]	14.07	24.44	45.90	55.45	104.33															
Traj-GCN [7]	9.94	18.02	33.55	40.95	81.85															
MST-GCN [1]	8.72	15.83	30.57	38.10	79.01															
STSGCN* [9]	10.80	18.19	31.18	41.05	81.76															
SPGSN	8.30	14.80	28.64	36.96	77.82															