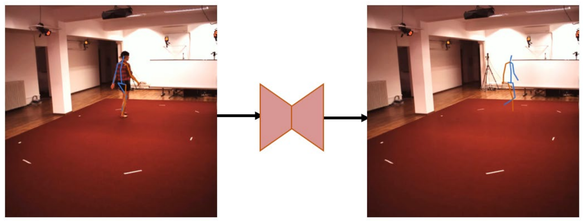
MSR-GCN: Multi-Scale Residual Graph Convolution Networks for Human Motion Prediction

Lingwei Dang [1], Yongwei Nie [1*], Chengjiang Long [2], Qing Zhang [3], Guiqing Li [1]
[1] South China University of Technology, [2] JD Finance America Corporation, [3] Sun Yat-sen University

2021 ICCV OCTOBER 11-17 VIRTUAL

Background & Motivation

- Human motion prediction is challenging for the stochasticity and aperiodicity of future poses.
- Graph Convolution Networks (GCNs) is suitable for non-grid graph-structured data like skeleton-based pose sequences.
- One can stabilize the motion pattern by gradually abstracting body parts in a fine-to-coarse manner.



Observed Seq Future Seq

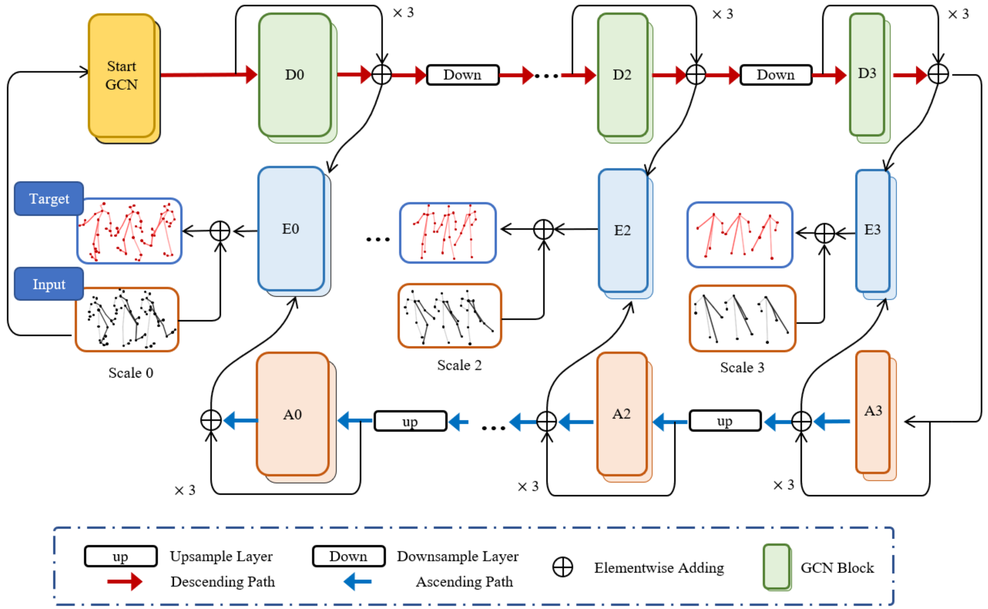
Contributions

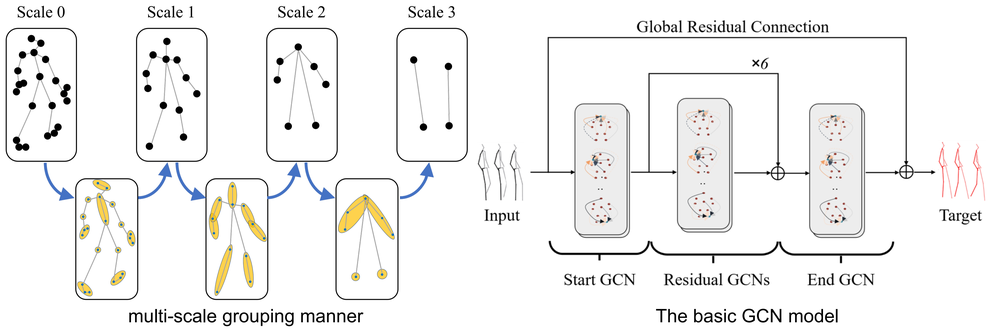
- A novel multi-scale residual graph convolution network.
- The descending and ascending GCN blocks can extract features in both fine-to-coarse and coarse-to-fine manners.
- The intermediate supervision benefiting high-quality future prediction.

Datasets & Metric & Project

- Datasets: Human3.6M, CMU Mocap Dataset
- Metric: Mean Per Joint Position Error (MPJPE)
- Project: https://github.com/Droliven/MSRGCN

Proposed Approach





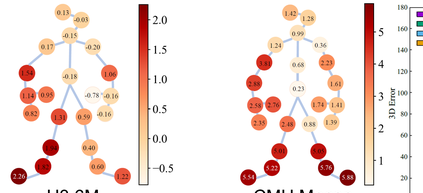
Quantitative Results

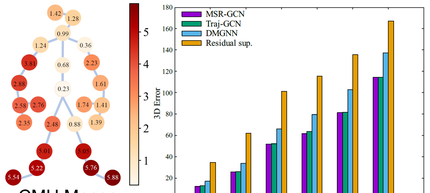
Short-term errors on H3.6M

scenarios	walking				eating				smoking				discussion			
millisecond (ms)	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Residual sup. [34]	29.36	50.82	76.03	81.51	16.84	30.60	56.92	68.65	22.96	42.64	70.14	82.68	32.94	61.18	90.92	96.19
DMGNN [27]	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 [33]	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
MSR-GCN	**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**
scenarios	directions				greeting				phoning				posing			
millisecond (ms)	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Residual sup. [34]	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
DMGNN [27]	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 [33]	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**
MSR-GCN	**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

Short-term errors on CMU Mocap

scenarios	basketball				basketball signal				directing traffic				jumping			
millisecond (ms)	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Residual sup. [34]	15.45	26.88	43.51	49.23	20.17	32.98	42.75	44.65	20.52	40.58	75.38	90.36	26.85	48.07	93.50	108.90
DMGNN [27]	15.57	28.72	59.01	73.05	5.03	9.28	20.21	26.23	10.21	20.90	41.55	52.28	31.97	54.32	96.66	119.92
Traj-GCN [33]	11.68	21.26	40.99	50.78	3.33	6.25	13.58	17.98	6.92	13.69	30.30	39.97	17.18	32.37	60.12	72.55
MSR-GCN	**10.28**	**18.94**	**37.68**	**47.03**	**3.03**	**5.68**	**12.35**	**16.26**	**5.92**	**12.09**	**28.36**	**38.04**	**14.99**	**28.66**	**55.86**	**69.05**
scenarios	running				soccer				walking				washwindow			
millisecond (ms)	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Residual sup. [34]	25.76	48.91	88.19	100.80	17.75	31.30	52.55	61.40	44.35	76.66	126.83	151.43	22.84	44.71	86.78	104.68
DMGNN [27]	17.42	26.82	38.27	40.08	14.86	25.29	52.21	65.42	9.57	15.53	26.03	30.37	7.93	14.68	33.34	44.24
Traj-GCN [33]	14.53	24.20	37.44	41.10	13.33	24.00	43.77	53.20	6.62	10.74	**17.40**	**20.35**	5.96	11.62	**24.77**	**31.63**
MSR-GCN	**12.84**	**20.42**	**30.58**	**34.42**	**10.92**	**19.50**	**37.05**	**46.38**	**6.31**	**10.30**	17.64	21.12	**5.49**	**11.07**	25.05	32.51

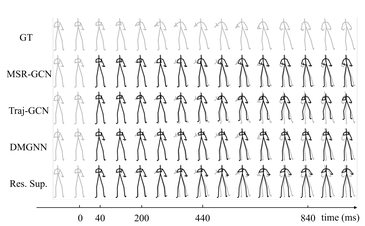




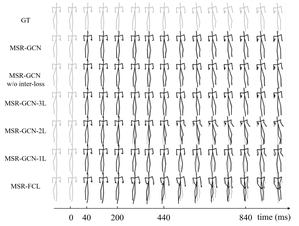
Performance gain over Traj-GCN

Average error on H3.6M

Qualitative Results



Visualization on H3.6M



Visualization of different ablation variants

Key References

[Traj-GCN] Mao W, Liu M, Salzmann M, *et al.* : Learning trajectory dependencies for human motion prediction. ICCV, 2019.
[DMGNN] Li M, Chen S, Zhao Y, *et al.* : Dynamic multiscale graph neural networks for 3d skeleton based human motion prediction. CVPR, 2020.
[Residual sup.] Martinez J, Black M J, Romero J. : On human motion prediction using recurrent neural networks. CVPR, 2017.

Acknowledgement

This research is sponsored in part by the National Natural Science Foundation of China (62072191, 61802453, 61972160), in part by the Natural Science Foundation of Guangdong Province (2019A1515010860, 2021A1515012301), and in part by the Fundamental Research Funds for the Central Universities (D2190670).