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]

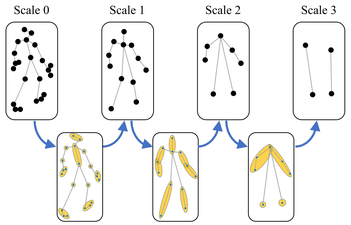
[1] South China University of Technology, [2] JD Finance America Corporation, [3] Sun Yat-sen University

csdanglw@mail.scut.edu.cn, nieyongwei@scut.edu.cn, cjfykx@gmail.com, zhangqing.whu.cs@gmail.com, ligq@scut.edu.cn

2021 ICCV OCTOBER 11-17 VIRTUAL

Motivation

- Human motion prediction is a challenging task due to the stochasticity and aperiodicity of future poses.
- Graph convolutional network has been proven to be very effective to learn dynamic relations among pose joints, which is helpful for pose prediction.
- One can abstract a human pose recursively to obtain a set of poses at multiple scales. With the increase of the abstraction level, the motion of the pose becomes more stable, which benefits pose prediction too.



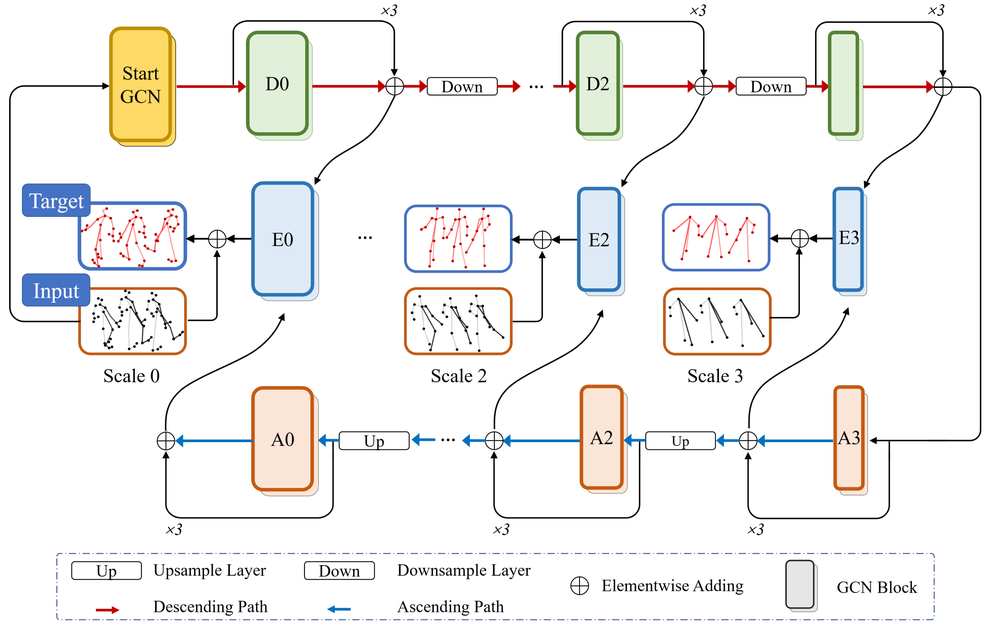
Contributions

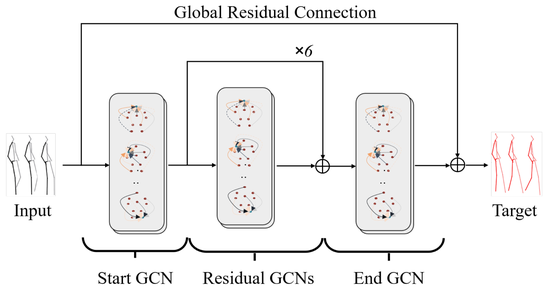
- We propose a novel multi-scale residual graph convolution network for human pose prediction in an end-to-end manner, which consists of multiple GCNs organized in a multi-scale architecture.
- The well-designed descending and ascending GCN blocks can extract features in both fine-to-coarse and coarse-to-fine manners.
- The intermediate supervision imposed at each scale enforces to learn more representative features, benefiting high-quality future prediction.

Datasets & Metric

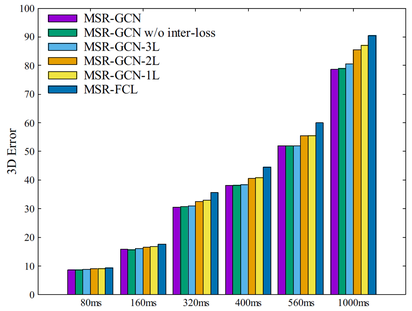
- Datasets: Human3.6M, CMU Mocap Dataset
- Metric: 3D Mean Per Joint Position Error (MPJPE)

Proposed Approach





Basic residual GCN model



Errors of different ablation variants on CMU Mocap

Quantitative Results

Short-term errors on H3.6M

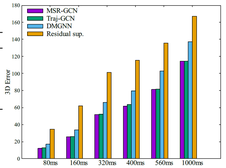
scenarios	walking				eating				smoking				discussion			
millisecond (ms)	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Residual sup. [34]	29.36	50.82	76.03	81.51	16.84	30.60	56.92	68.65	22.96	42.64	70.14	82.68	32.94	61.18	90.92	96.19
DMGNN [27]	17.32	30.67	54.56	65.20	10.96	21.39	36.18	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
scenarios	purchases				sitting				sittingdown				takingphoto			
millisecond (ms)	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Residual sup. [34]	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
DMGNN [27]	21.35	38.71	75.67	92.74	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 [33]	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
MSR-GCN	**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**
scenarios	waiting				walkingdog				walkingtogether				Average			
millisecond (ms)	80	160	320	400	80	160	320	400	80	160	320	400	80	160	320	400
Residual sup. [34]	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
DMGNN [27]	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 [33]	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
MSR-GCN	**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**

Long-term errors on H3.6M

scenarios	walking		Eating		Smoking		Discussion		Directions		average	
millisecond (ms)	560	1000	560	1000	560	1000	560	1000	560	1000	560	1000
Residual sup.[34]	81.73	100.68	79.87	100.20	94.83	137.44	121.30	161.70	110.05	152.48	97.56	130.50
DMGNN [27]	73.36	95.82	58.11	86.66	50.85	72.15	**81.90**	138.32	110.06	115.75	74.85	101.74
Traj-GCN [33]	54.05	**59.75**	53.39	77.75	50.74	72.62	91.61	121.53	**71.01**	101.79	64.16	86.69
MSR-GCN	**52.72**	63.04	**52.54**	**77.11**	**49.45**	**71.64**	88.59	**117.59**	71.18	**100.59**	**62.89**	**86.00**

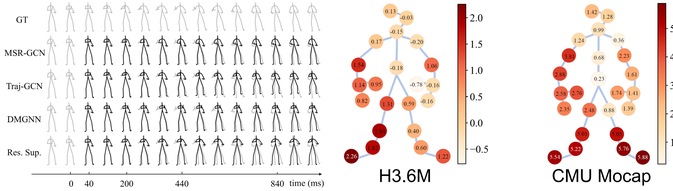
Long-term errors on CMU Mocap

scenarios	basket	bas_sig	dir_tra	jumping
Residual sup.[34]	**72.83**	60.57	153.12	162.84
DMGNN [27]	138.62	52.04	111.23	224.63
Traj-GCN [33]	97.99	54.00	114.16	127.41
MSR-GCN	86.96	**47.91**	**111.04**	**124.79**
scenarios	running	soccer	walking	washwin
Residual sup.[34]	158.19	107.37	194.33	202.73
DMGNN [27]	**46.40**	111.90	67.01	82.84
Traj-GCN [33]	51.73	108.26	**34.41**	**66.95**
MSR-GCN	48.03	**99.32**	39.70	71.30



Average error on H3.6M

Qualitative Results



Visualization on H3.6M

Performance gain over Traj-GCN

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).