# Group Graph Convolutional Networks for 3D Human Pose Estimation

Zijian Zhang (zhangzj2015@bupt.edu.cn)

Beijing University of Posts and Telecommunications, Beijing, China

## Introduction

In skeleton-based 3D human pose estimation (HPE), graph convolutional networks (GCNs) have recently achieved encouraging performance. However, most previous GCNs are limited by coupling aggregation mechanism.

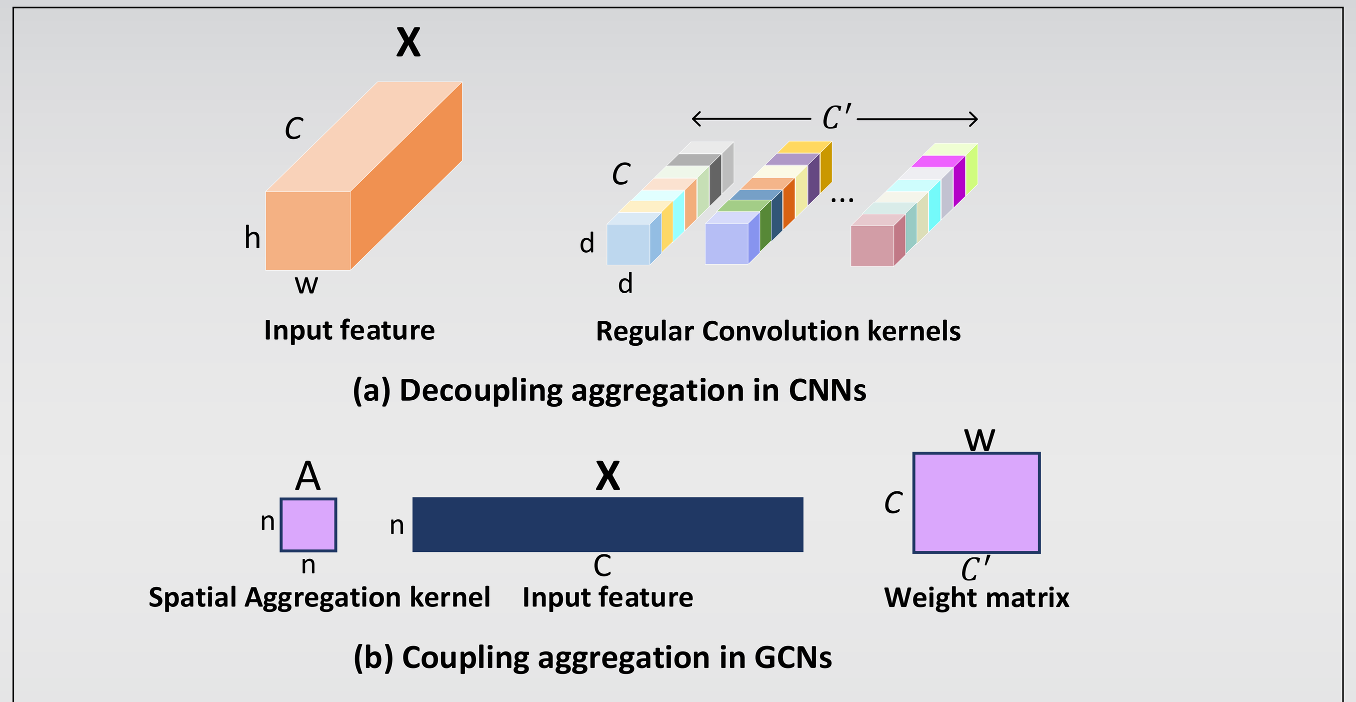


Figure 1. Aggregation mechanisms in CNNs and GCNs.

## Objectives

**Problem:**

How to dynamically learn different topologies and effectively aggregate joint features in GCNs for 3D human pose estimation?

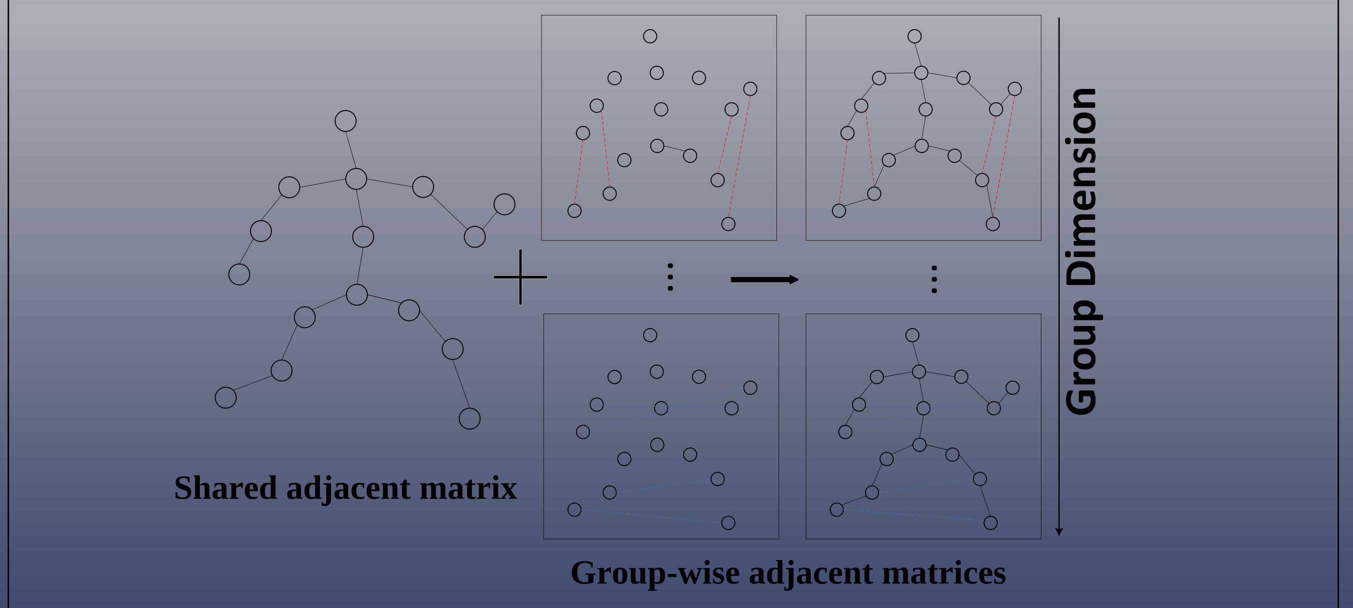


Figure 2. Group-wise adjacent matrices refinement. Lines of different colors correspond to adjacent matrices in different group.

## Methods

In this paper, similar to group convolution in CNNs, we propose group graph convolutional networks (GroupGCN), a novel decoupling GCN for 3D HPE. It consists of group convolution and group interaction. Group convolution ensures that every group has its own spatial aggregation kernel and weight matrix. A drawback of group convolution is that the status of the other groups is completely unknown because they are independent. As a result, the set of independent group convolution may not be globally coherent, leading to poor performance. So we propose group interaction to account for global information by making the features interact between groups.

### Group convolution

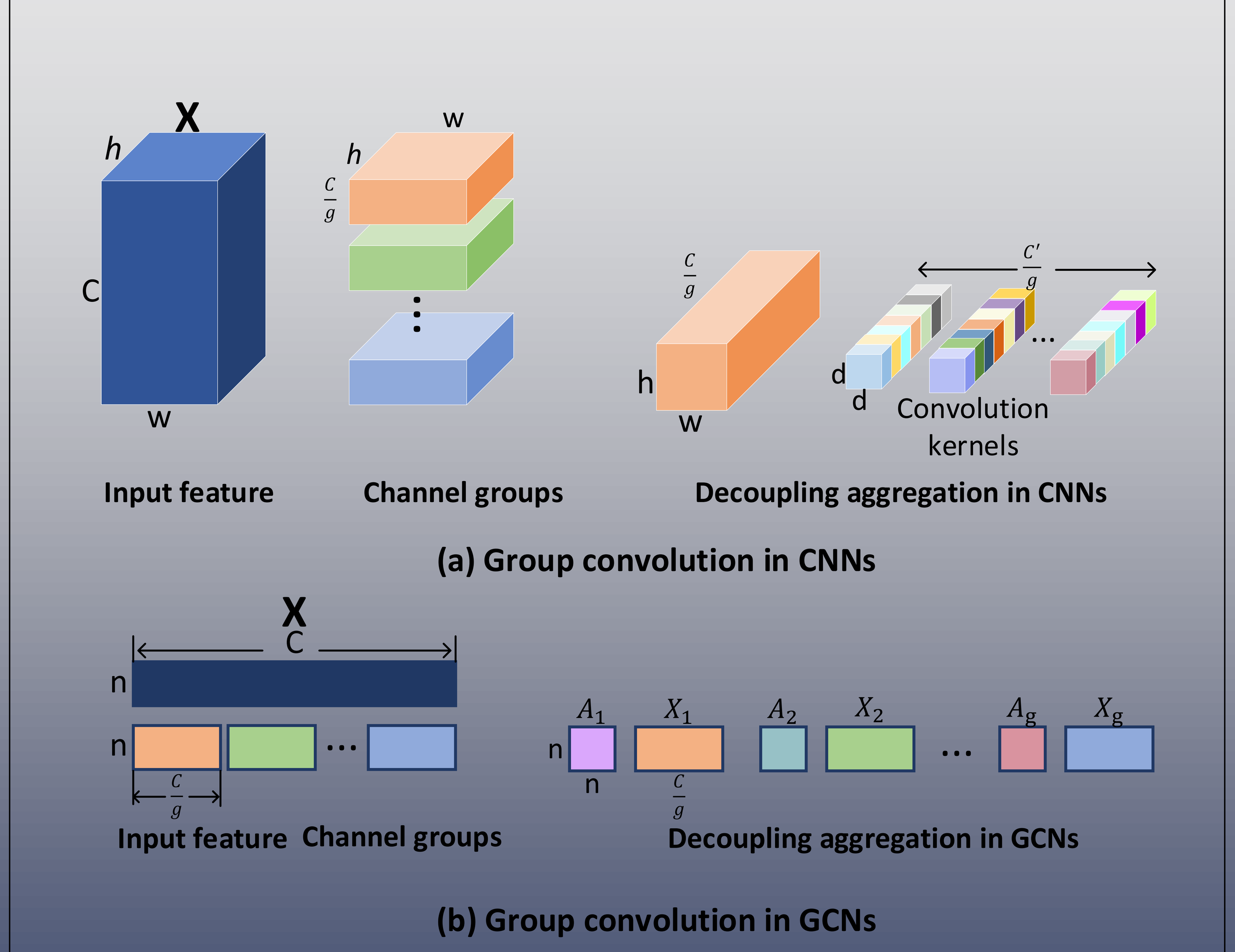


Figure 3. Illustration of group convolution. We introduce the group convolution into GCNs and propose GroupGCN.

## Methods

### Group interaction

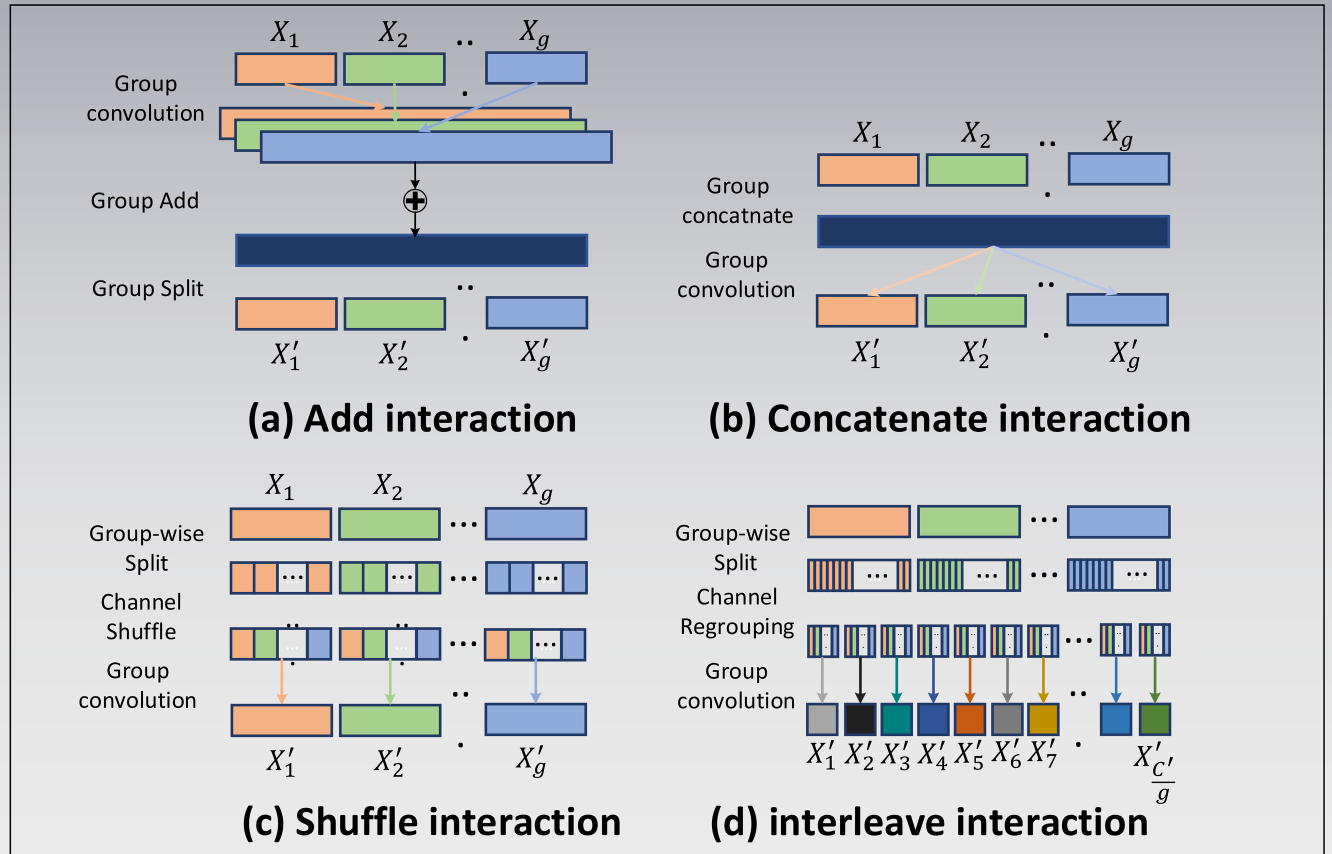


Fig 4. Illustration of four group interaction approaches

## Results

Experiments indicate that our approach achieve the state-of-the-art performance has strong generalization capabilities to unseen datasets.

### Results on H3.6M dataset

| Method | | Direct | Discuss | Eat | Greet | Phone | Photo | Pose | Purcha. | Sit | SitD | Smoke | Wait | WalkD | Walk | WalkT | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Martinez et al. [ ] | ICCV'17 | 51.8 | 56.2 | 58.1 | 59.0 | 69.5 | 78.4 | 55.2 | 58.1 | 74.0 | 94.6 | 62.3 | 59.1 | 65.1 | 49.5 | 52.4 | 62.9 |
| Park et al. [ ] | BMVC'18 | 49.4 | 54.3 | 51.6 | 55.0 | 61.0 | 73.3 | 53.7 | 50.0 | 68.5 | 88.7 | 58.6 | 56.8 | 57.8 | 46.2 | 48.6 | 58.6 |
| Zhao et al. [ ]† | CVPR'19 | 47.3 | 60.7 | 51.4 | 60.5 | 61.1 | 49.9 | **47.3** | 68.1 | 86.2 | **55.0** | 67.8 | 61.0 | **42.1** | 60.6 | 45.3 | 57.6 |
| liu et al. [ ]† | ECCV'20 | 46.3 | 52.2 | 47.3 | 50.7 | 55.5 | 67.1 | 49.2 | **46.0** | 60.4 | 71.1 | 51.5 | 50.1 | 54.5 | 40.3 | 43.7 | 52.4 |
| Xu et al. [ ]† | CVPR'21 | 45.2 | 49.9 | 47.5 | 50.9 | 54.9 | 66.1 | 48.5 | 46.3 | **59.7** | 71.5 | 51.4 | 48.6 | 53.9 | 39.9 | 44.1 | 51.9 |
| *Ours-AI* † | | 45.4 | 51.4 | 49.8 | 50.3 | 55.0 | **60.8** | 47.9 | 48.4 | 61.0 | 70.7 | 52.7 | 48.9 | 55.2 | 40.1 | 41.9 | 52.0 |
| *Ours-CI* † | | **45.0** | **50.9** | 49.0 | **49.8** | **52.2** | 60.9 | 49.1 | 46.8 | 61.2 | 70.2 | 51.8 | **48.6** | 54.6 | **39.6** | **41.2** | **51.6** |
| Martinez et al. [ ] | ICCV'17 | 37.7 | 44.4 | 40.3 | 42.1 | 48.2 | 54.9 | 44.4 | 42.1 | 54.6 | 58.0 | 45.1 | 46.4 | 47.6 | 36.4 | 40.4 | 45.5 |
| Zhao et al. [ ]† | CVPR'19 | 37.8 | 49.4 | 37.6 | 40.9 | 45.1 | 41.4 | 40.1 | 48.3 | 50.1 | **42.2** | 53.5 | 44.3 | 40.5 | 47.3 | 39.0 | 43.8 |
| liu et al. [ ]† | ECCV'20 | 36.8 | 40.3 | 33.0 | 36.3 | 37.5 | 45.0 | 39.7 | 34.9 | 40.3 | 47.7 | 37.4 | 38.5 | 38.6 | 29.6 | 32.0 | 37.8 |
| Zeng et al. [ ] | ECCV'20 | 35.9 | 36.7 | 29.3 | 34.5 | 36.0 | 42.8 | 37.7 | **31.7** | 40.1 | 44.3 | 35.8 | 37.2 | 36.2 | 33.7 | 34.0 | 36.4 |
| Ci et al. [ ]† | ICCV'19 | 36.3 | 38.8 | 29.7 | 37.8 | **34.6** | 42.5 | 39.8 | 32.5 | 36.2 | 39.5 | 34.4 | 38.4 | 38.2 | 31.3 | 34.2 | 36.3 |
| Xu et al. [ ]† | CVPR'21 | 35.8 | 38.1 | 31.0 | 35.3 | 35.8 | 43.2 | **37.3** | 31.7 | 38.4 | 45.5 | 35.4 | 36.7 | 36.8 | 27.9 | 30.7 | 35.8 |
| *Ours-AI* † | | **31.1** | 37.3 | **29.9** | **32.7** | 35.0 | 40.5 | 38.3 | 32.7 | 39.4 | 48.4 | 33.6 | 37.0 | 35.7 | **27.8** | 29.5 | 35.3 |
| *Ours-CI* † | | 32.5 | **36.4** | 30.7 | 33.2 | 34.9 | **40.0** | 37.8 | 33.1 | **38.3** | 47.8 | **34.4** | **36.2** | **35.1** | 28.4 | **29.2** | **35.2** |

### Results on 3DHP dataset

| | | GS | noGS | Outdoor | All (PCK) | All (AUC) |
|---|---|---|---|---|---|---|
| Martinez et al. [ ] | ICCV'17 | 49.8 | 42.5 | 31.2 | 42.5 | 17.0 |
| Ci et al. [ ] | ICCV'19 | 74.8 | 70.8 | 77.3 | 74.0 | 36.7 |
| Zeng et al. [ ] | ECCV'20 | - | - | 80.3 | 77.6 | 43.8 |
| liu et al. [ ] | ECCV'20 | 77.6 | 80.5 | 80.1 | 79.3 | 47.6 |
| Xu et al. [ ] | CVPR'21 | **81.5** | 81.7 | 75.2 | 80.1 | 45.8 |
| Zeng [ ] | ICCV'21 | - | - | **84.6** | **82.1** | 46.2 |
| *Ours-AI* † | | 81.1 | 84.0 | 77.6 | 81.3 | 49.7 |
| *Ours-CI* † | | 80.4 | **84.5** | 77.2 | 81.1 | **49.9** |

## Results

Fig 5 indicates the effectiveness of our proposed approach in tackling the 2D-to-3D pose estimation problem.

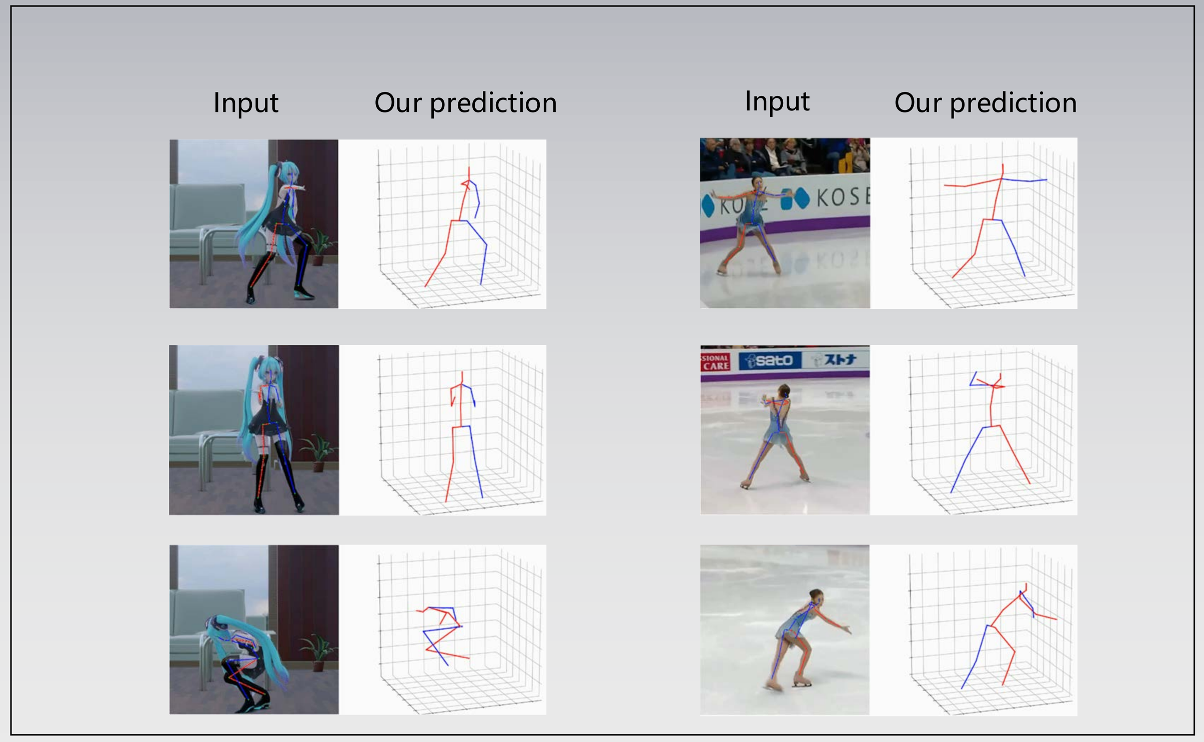


Figure 5. Qualitative results of our method on im-the-wild images.

## Conclusions

- We make new conclusions that (1) decoupling aggregation can effectively improve the performance of graph convolution, (2) different group interaction strategies and spatial aggregation kernels have a significant impact on the performance of 3D HPE.
- The experimental results prove that GroupGCN can achieve state-of-the-art performance with fewer parameters.

## References

1. Kenkun Liu, Rongqi Ding, Zhiming Zou, Le Wang, and Wei Tang. A comprehensive study of weight sharing in graph networks for 3d human pose estimation. In European Conference on Computer Vision, pages 318–334. Springer, 2020.
2. Julieta Martinez, Rayat Hossain, Javier Romero, and James J Little. A simple yet effective baseline for 3d human pose estimation. In Proceedings of the IEEE International Conference on Computer Vision, pages 2640–2649, 2017.
3. Long Zhao, Xi Peng, Yu Tian, Mubbasir Kapadia, and Dimitris N Metaxas. Semantic graph convolutional networks for 3d human pose regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3425– 3435, 2019.
4. Junhao Zhang, Yali Wang, Zhipeng Zhou, Tianyu Luan, Zhe Wang, and Yu Qiao. Learning dynamical human-joint affinity for 3d pose estimation in videos. IEEE Transactions on Image Processing, 30:7914–7925, 20.
5. Tianhan Xu and Wataru Takano. Graph stacked hourglass networks for 3d human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 16105–16114, 2021.