

# Motion Transfer

Hsuan Ouyang, Jiaming Liu, Sheng Xu, Zichen(Chen) Li

Carnegie Mellon University - Deep Learning 2019 Fall

## Abstract

### Motivation

Our purpose is to research on the task of motion transfer and to present a model combination to achieve better synthesis result, especially to address the blur issue and the distortion problem around face area in the synthesised video. Also, most importantly, to have fun with this project!

### Approach

- Extract pose from target video we shot using the latest OpenPose
- Train the full body GAN model to generate images of target subject from pose figures
- Apply the model to source video
- Add face GAN and deblur GAN to optimise result

### Model

- Full-body GAN & Face GAN : based on pix2pixHD [3]
- Deblur GAN: WGAN [2]

### Conclusion

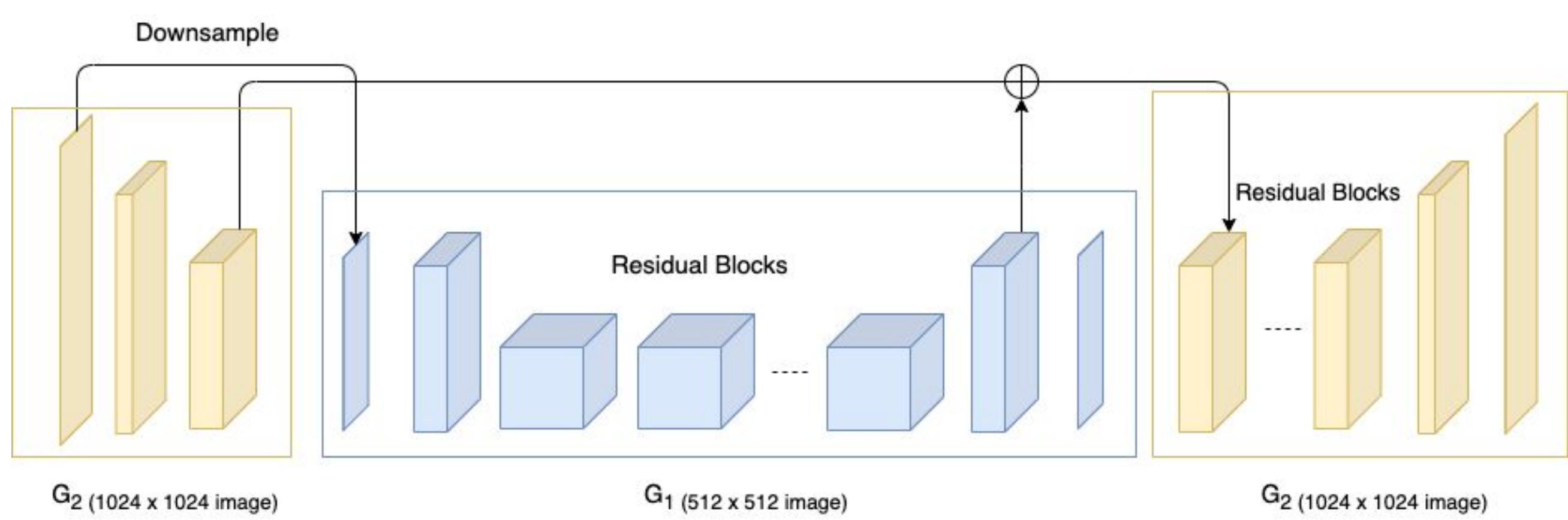
- With the restriction of computational resources, we can achieve a surprisingly satisfying video-to-video synthesis result with a simplified version of pix2pixHD [3]
- The addition of Face GAN and Deblur GAN does not significantly improve the result

### Limitation

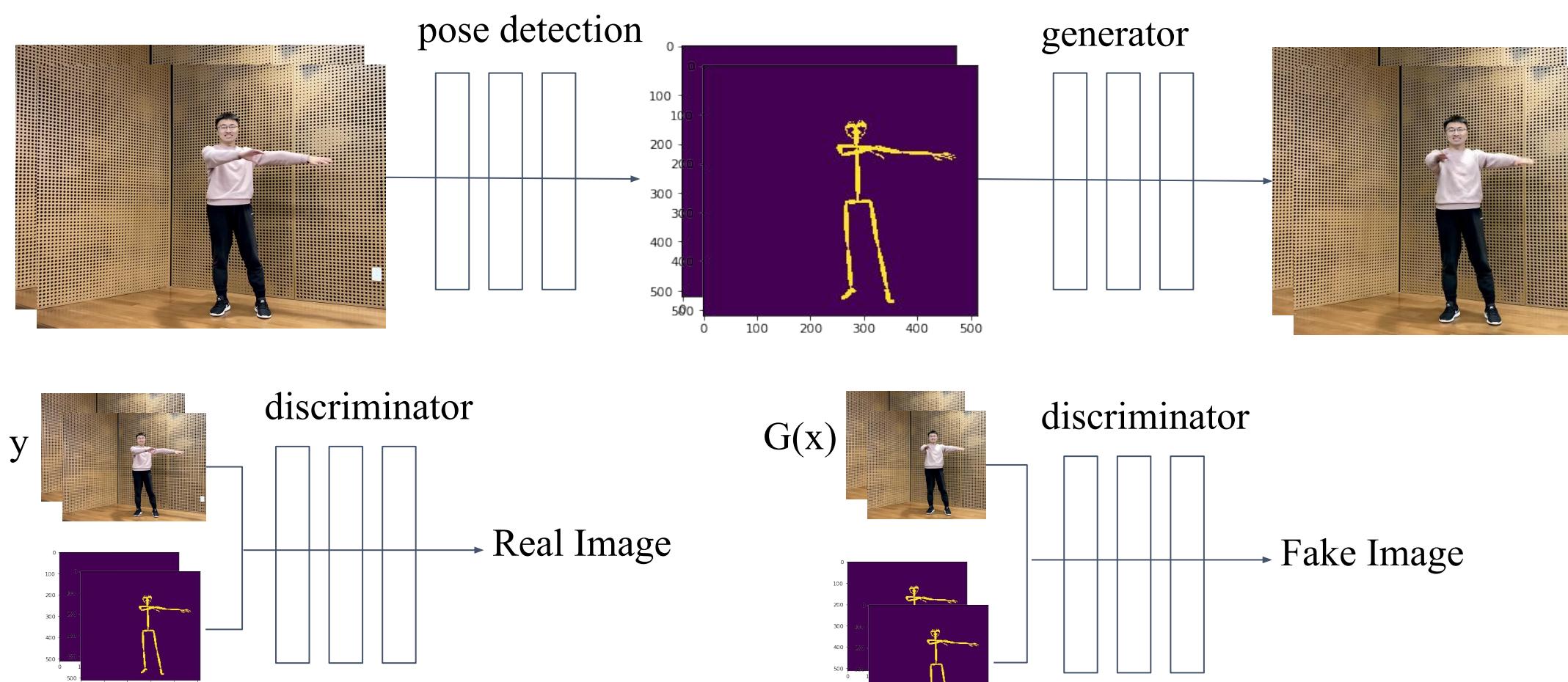
- We have limited time and computational resources to train our GAN model
- The gesture in the target video used for training will highly influence the transferring result

## Method

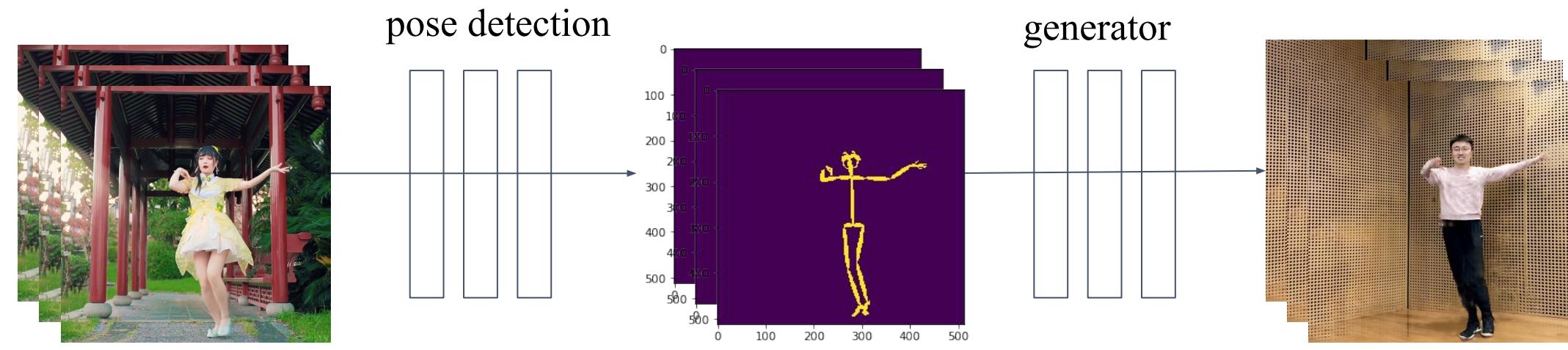
### Generator Network Structure (pix2pixHD [3])



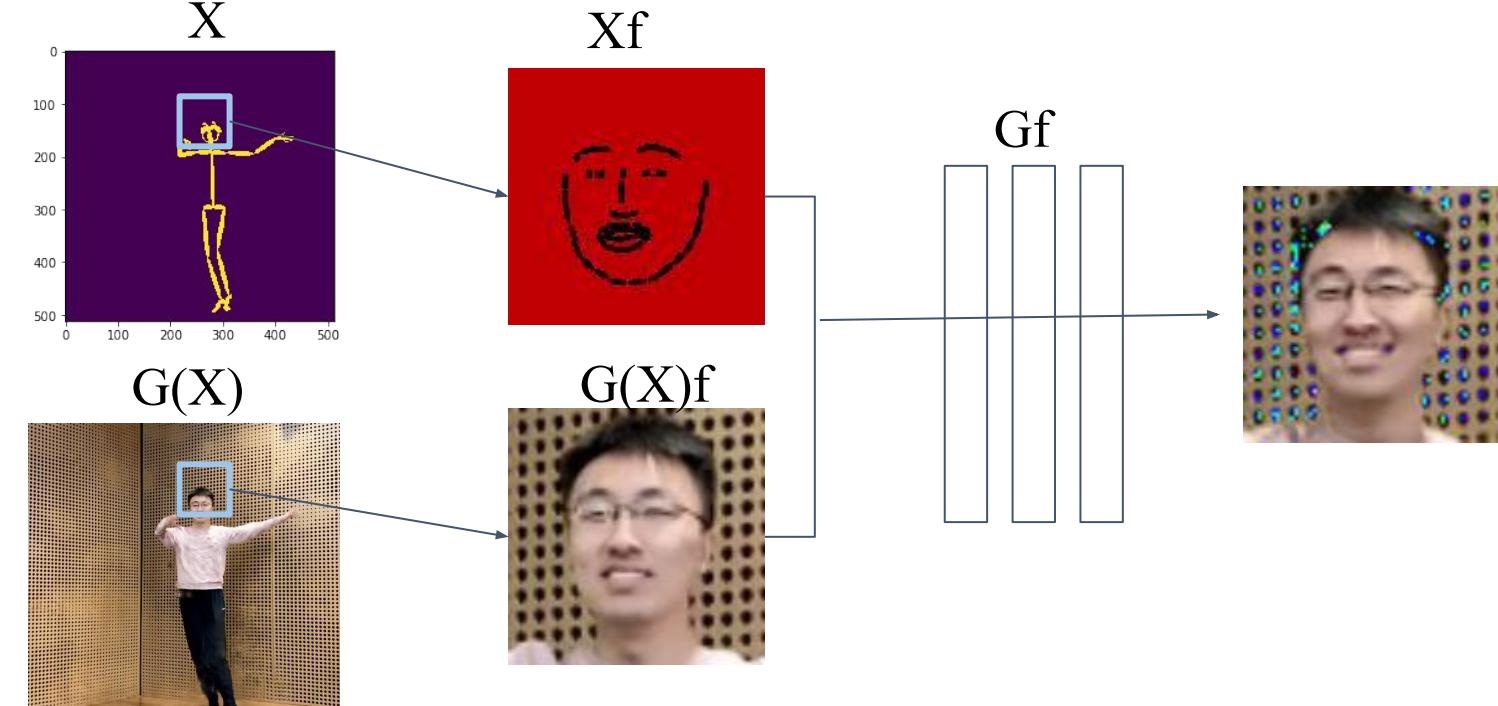
### Full-body GAN



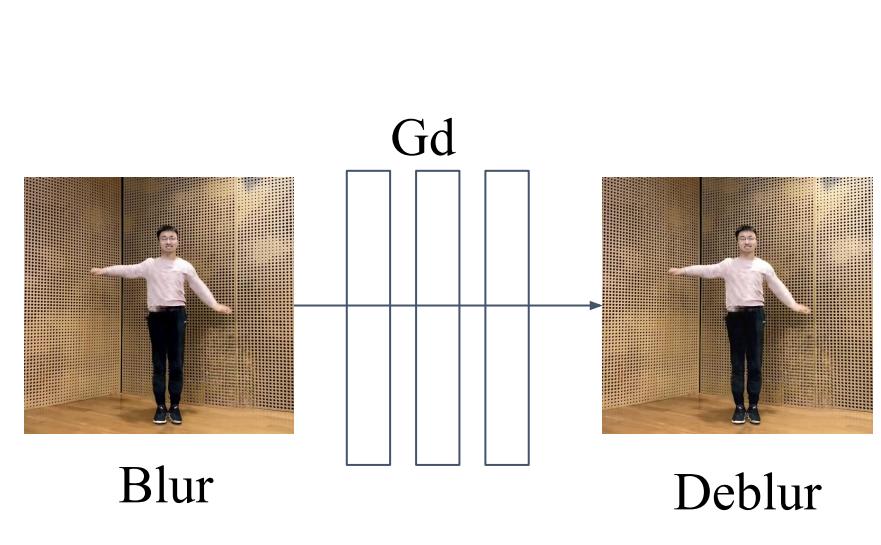
### Transfer



### Face GAN



### Deblur GAN

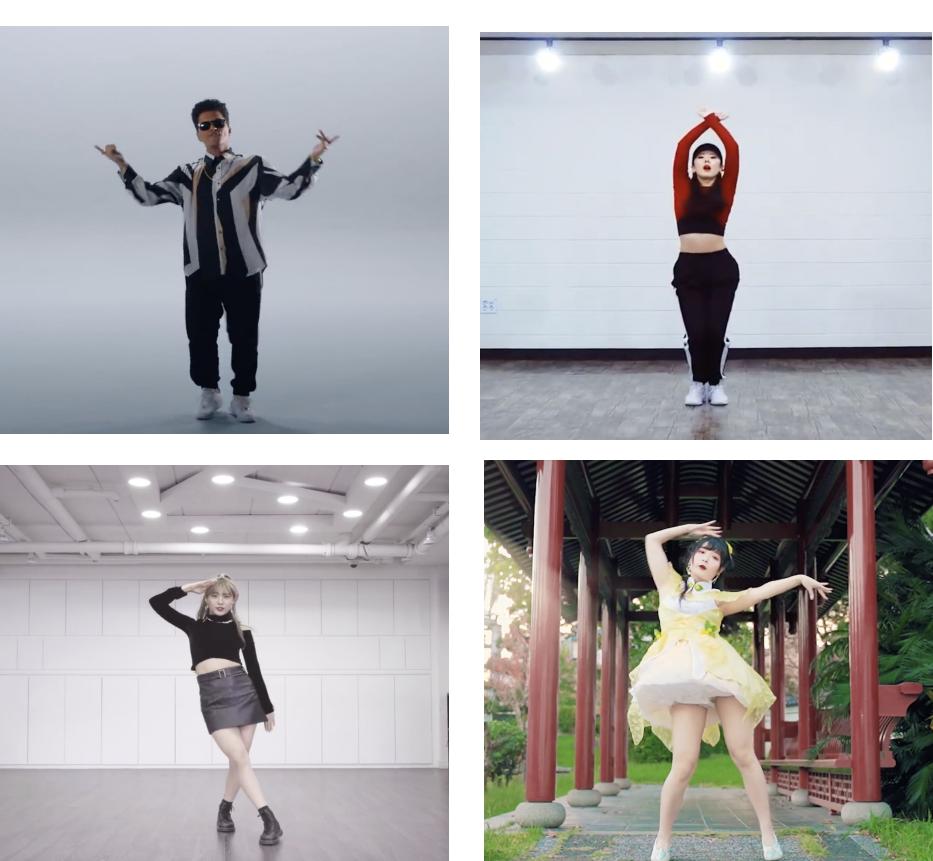


## Dataset

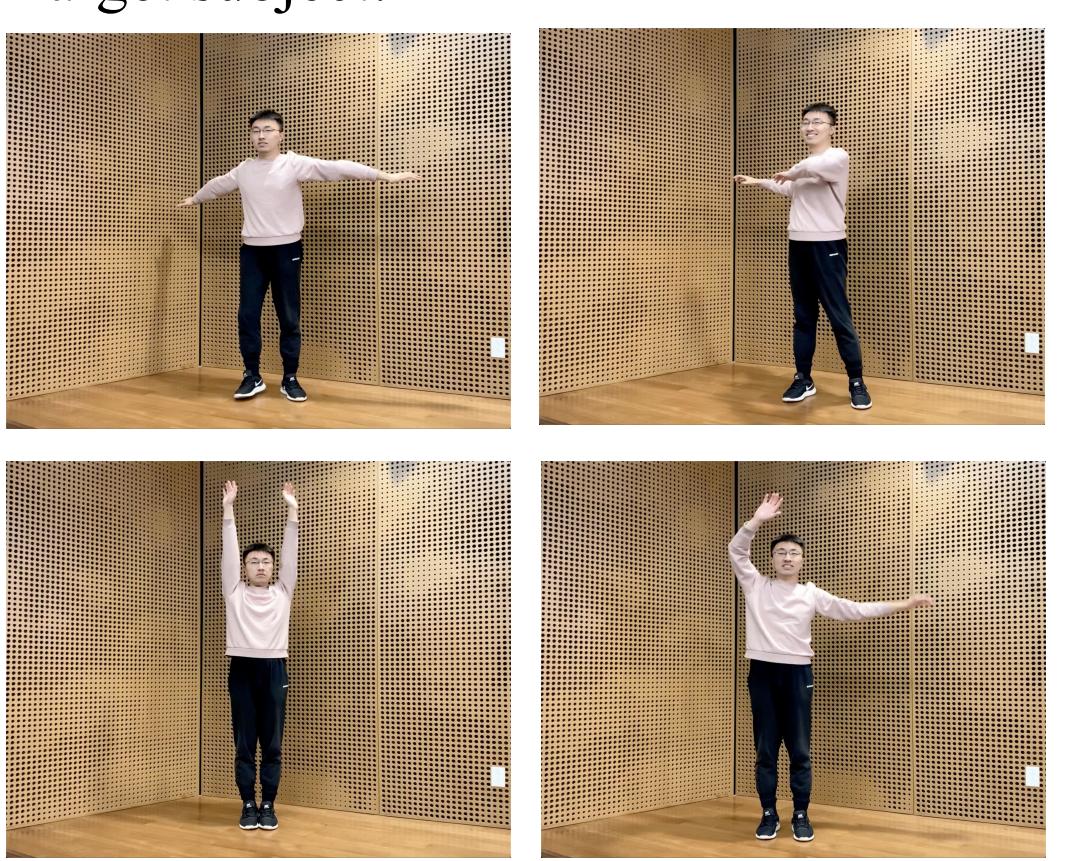
The dataset we used includes 2 parts: target person images for training and source dance videos for motion transfer. We chose one of our team members as the target subject. Detail information and dataset preview are listed below:

Name	Quantity/Size	Source
Source short dance videos	4	Bilibili, Youtube
Target basic movement video	1	Shot by team
Target subject images (512 * 512)	60 GB	Extracted from target video
Target subject images (1024 * 1024)	220 GB	Extracted from target video

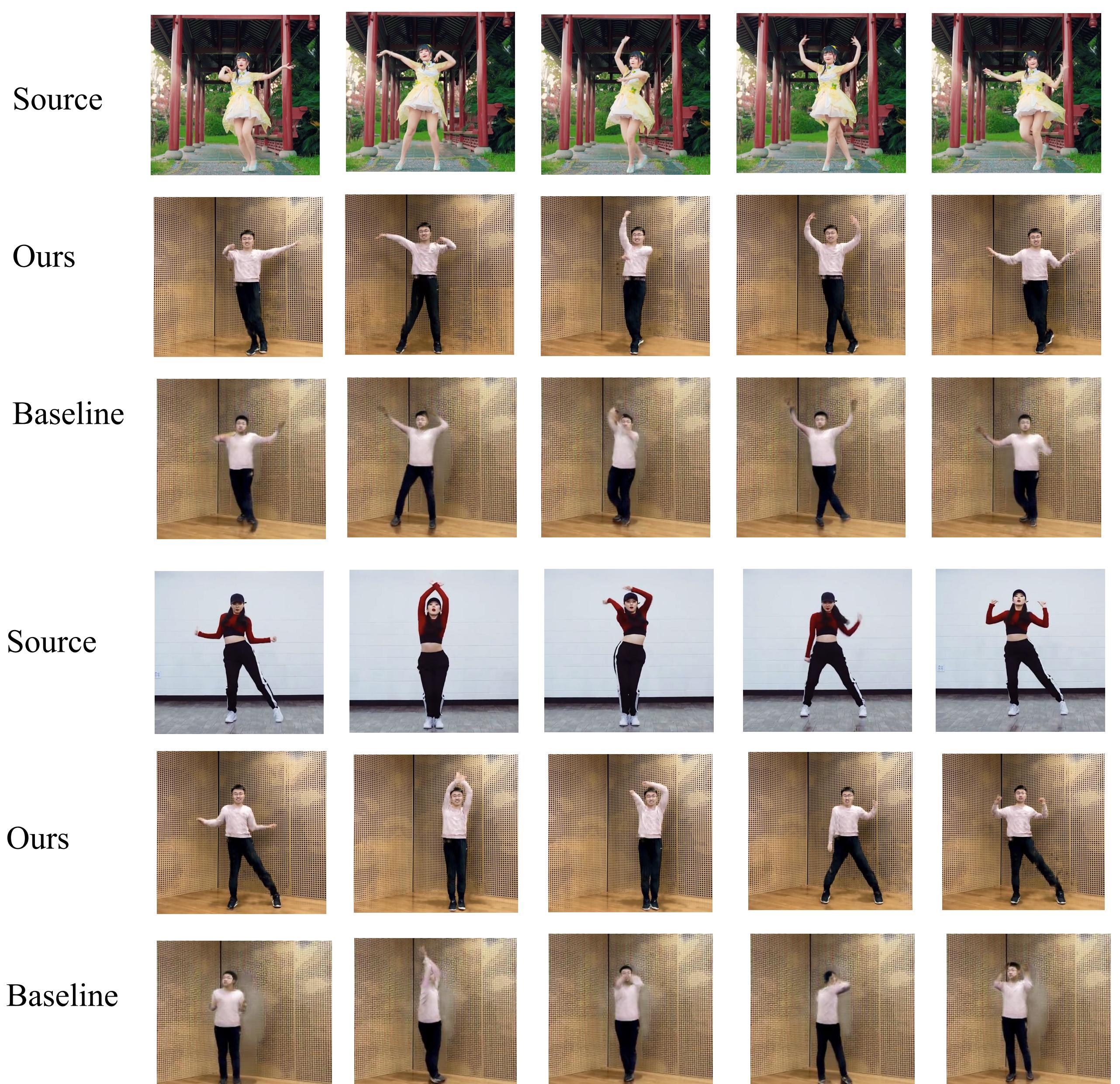
Source videos:



Target subject:



## Qualitative Comparison



## Quantitative Comparison

Metrix	Original paper	Baseline	Ours
SSIM	0.838 [1]	0.65	0.936

## Reference

- [1] Chan, C., Ginosar, S., Zhou, T. & Efros, A., (2019) Everybody dance now. ArXiv, abs/1808.07371
- [2] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin & J. Matas, (2018) Deblur-GAN: Blind Motion Deblurring Using Conditional Adversarial Networks. ArXiv, abs/1711.07064
- [3] Ting-Chun Wang, etc. High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.