thrive
SIGGRAPH2019
LOS ANGELES • 28 JULY – 1 AUGUST

Neural Rendering and Reenactment of Human Actor Videos

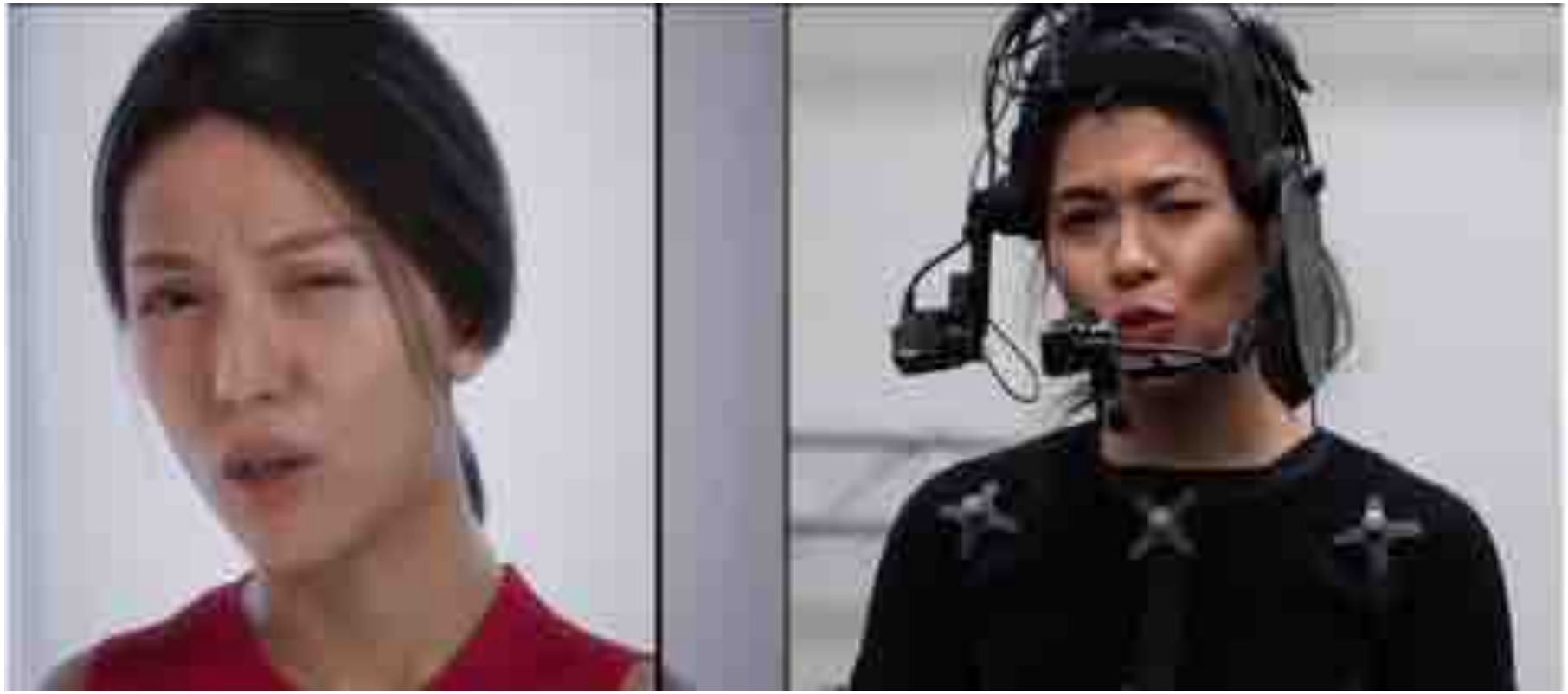
Lingjie Liu, Weipeng Xu, Michael Zollhoefer,
Hyeongwoo Kim, Florian Bernard, Marc Habermann,
Wenping Wang, Christian Theobalt

PHOTOGRAPHY &
RECORDING ENCOURAGED

Conventional Computer Graphics Modeling and Rendering

Modeling of a virtual character

Conventional Computer Graphics Modeling and Rendering



Motion Capture

Conventional Computer Graphics Modeling and Rendering

Rendering

Conventional Computer Graphics Modeling and Rendering

Modeling of a virtual character

Motion Capture

Rendering

Our Goal

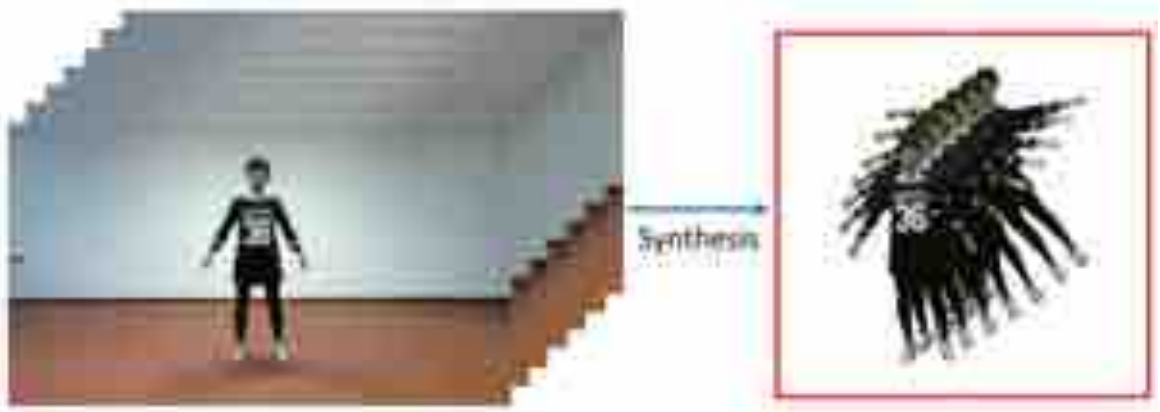
To design a more lightweight approach to capture and render video-realistic animations of real humans under user control.

Synthesized new motions

Related Work

Model-based Video Synthesis

[Xu et al. 2011]



[Li et al. 2017]

[Casas et al. 2014], [Volino et al. 2014], [Carranza et al. 2003], [Collet et al. 2015a], [Li et al. 2014], [Zitnick et al. 2004], …

Related Work

Learning-based Video Synthesis

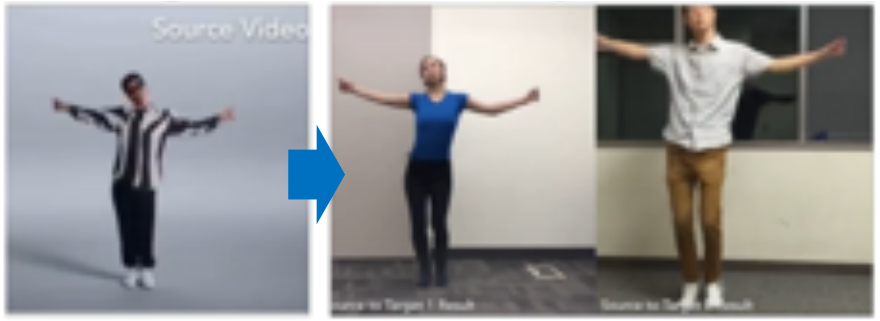
[Kim et al. 2018]

[Chan et al. 2018]

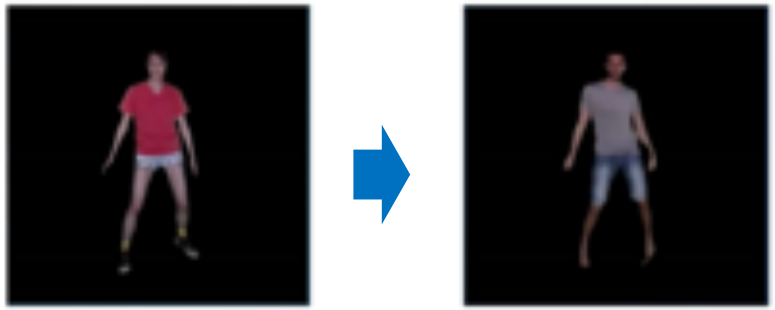
[Balakrishnan et al. 2018], [Ma et al. 2017], [Siarohin et al. 2018], [Wang et al. 2018], [Esser et al. 2018], [Zhou et al. 2019], [Gafni et al. 2019], [Aberman et al. 2018], [Lorenz et al. 2019], [Shysheya et al. 2019], [Justin et al. 2018], [Chan et al. 2018], …

Concurrent and Follow-up Works

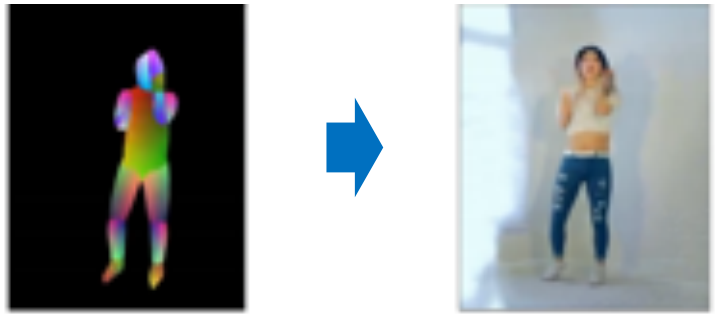
[Chan et al. 2018]



[Aberman et al. 2018]

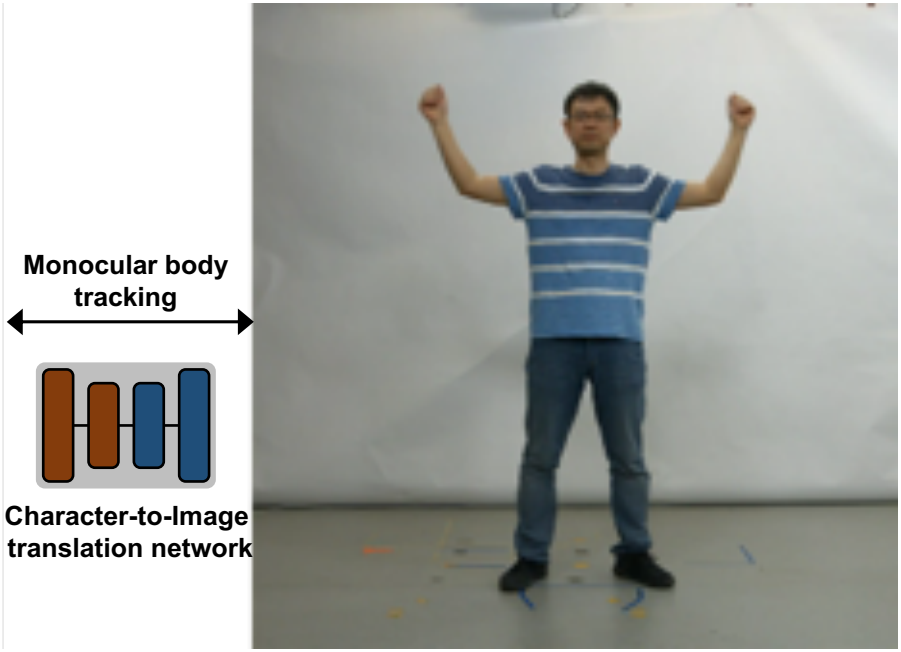


[Wang et al. 2018]



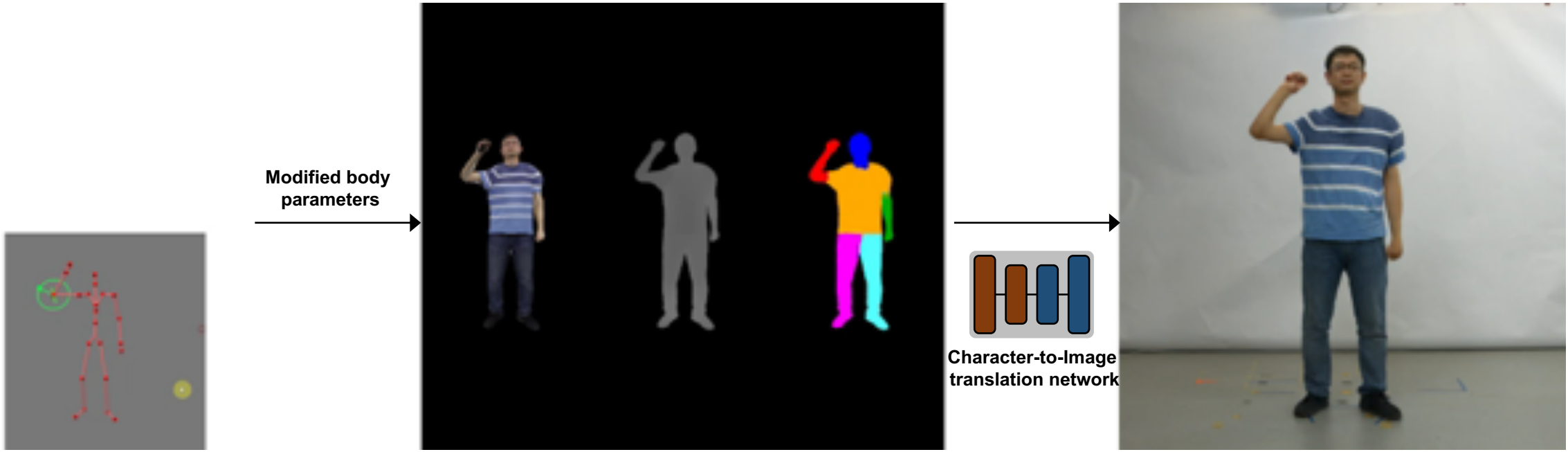
[Gafni et al. 2019], [Zhou et al. 2019], [Lorenz et al. 2019], …

Overview


Monocular body tracking
Character-to-Image translation network

Overview

Overview

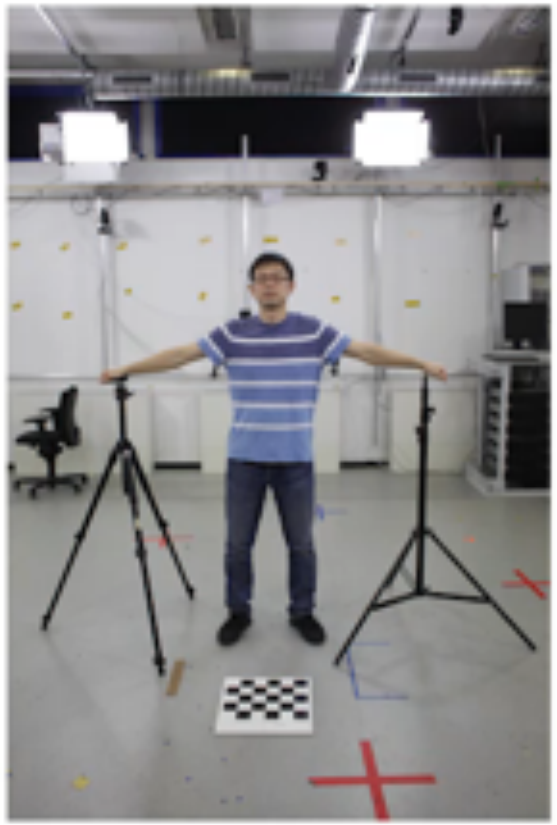

Modified body parameters
Character-to-Image translation network

Overview

Monocular body tracking

Training Data Acquisition

Template Acquisition



Template Acquisition

Template Mesh

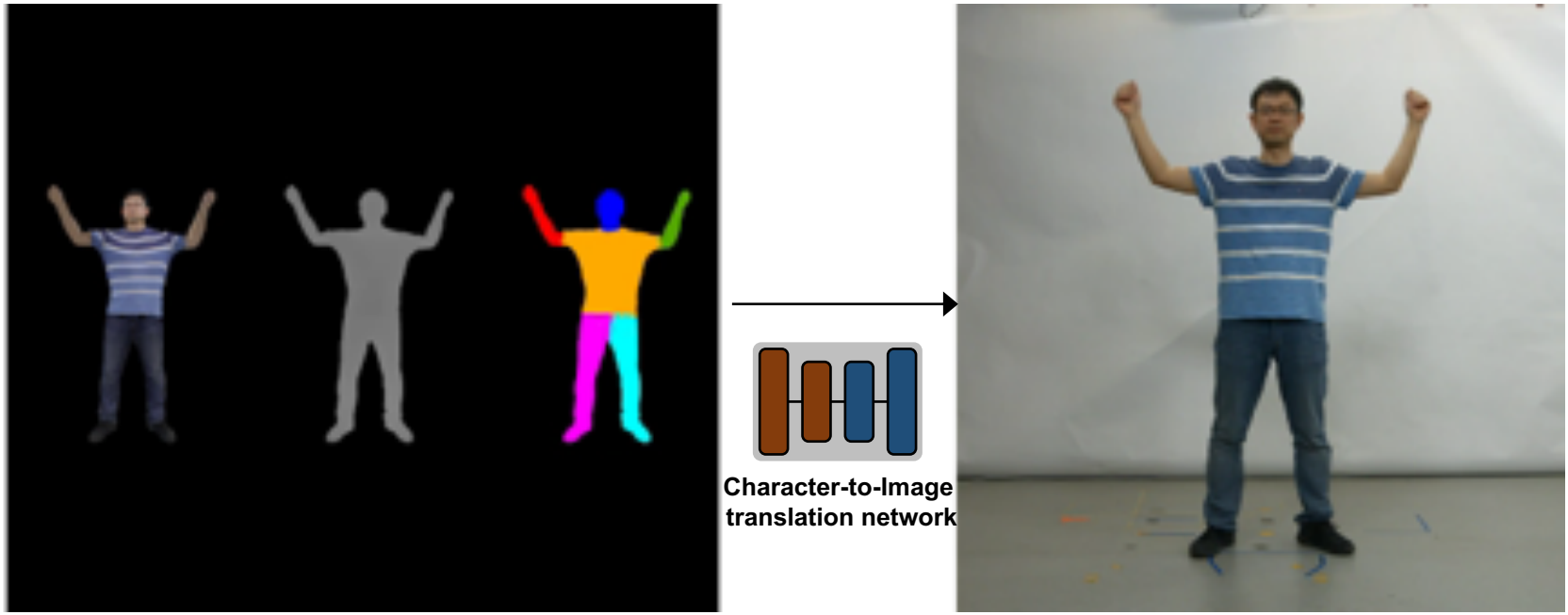
Rigged Template

Training Data Acquisition

Training Video

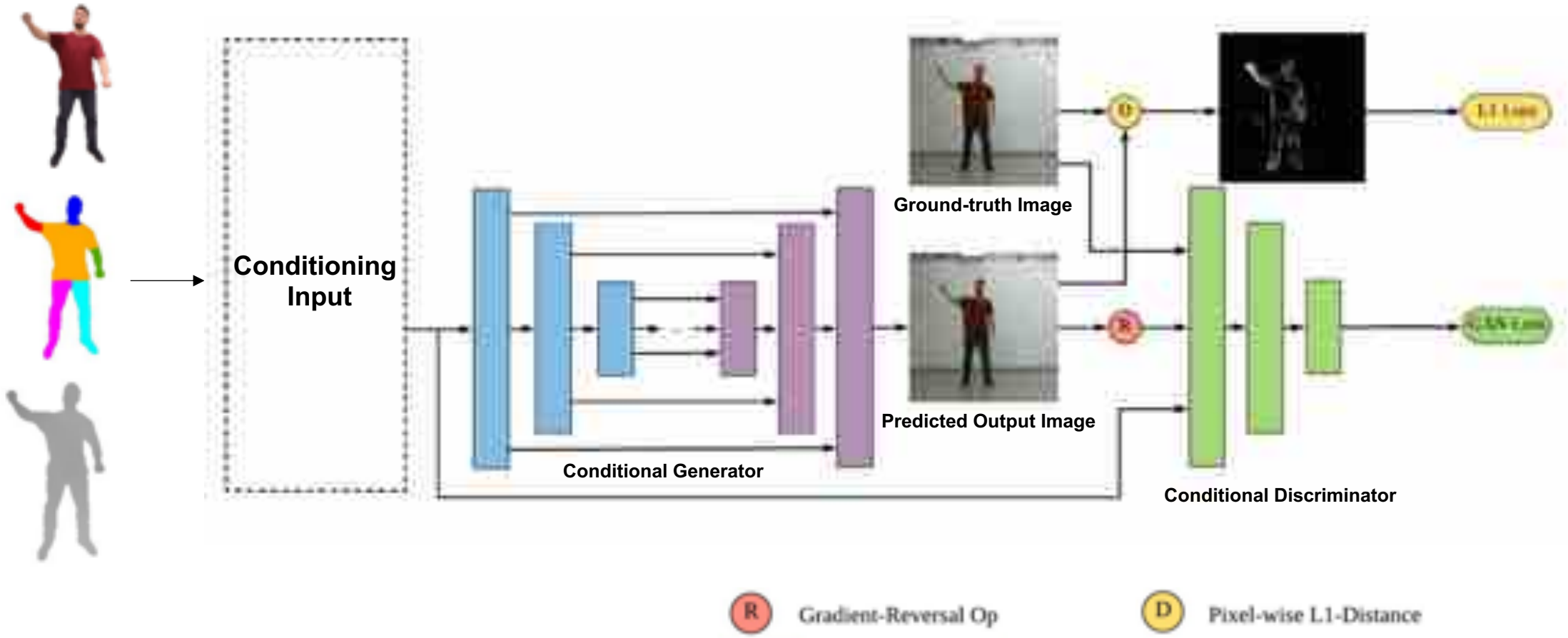
VNect [Mehta et al. 2017]

Overview



Character-to-Image Translation

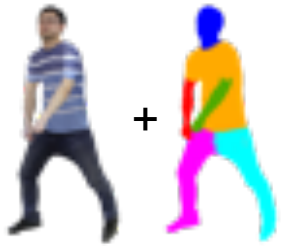
Character-to-Image Translation



Issue #1: How to choose conditioning inputs

Skeleton

RGB + mask

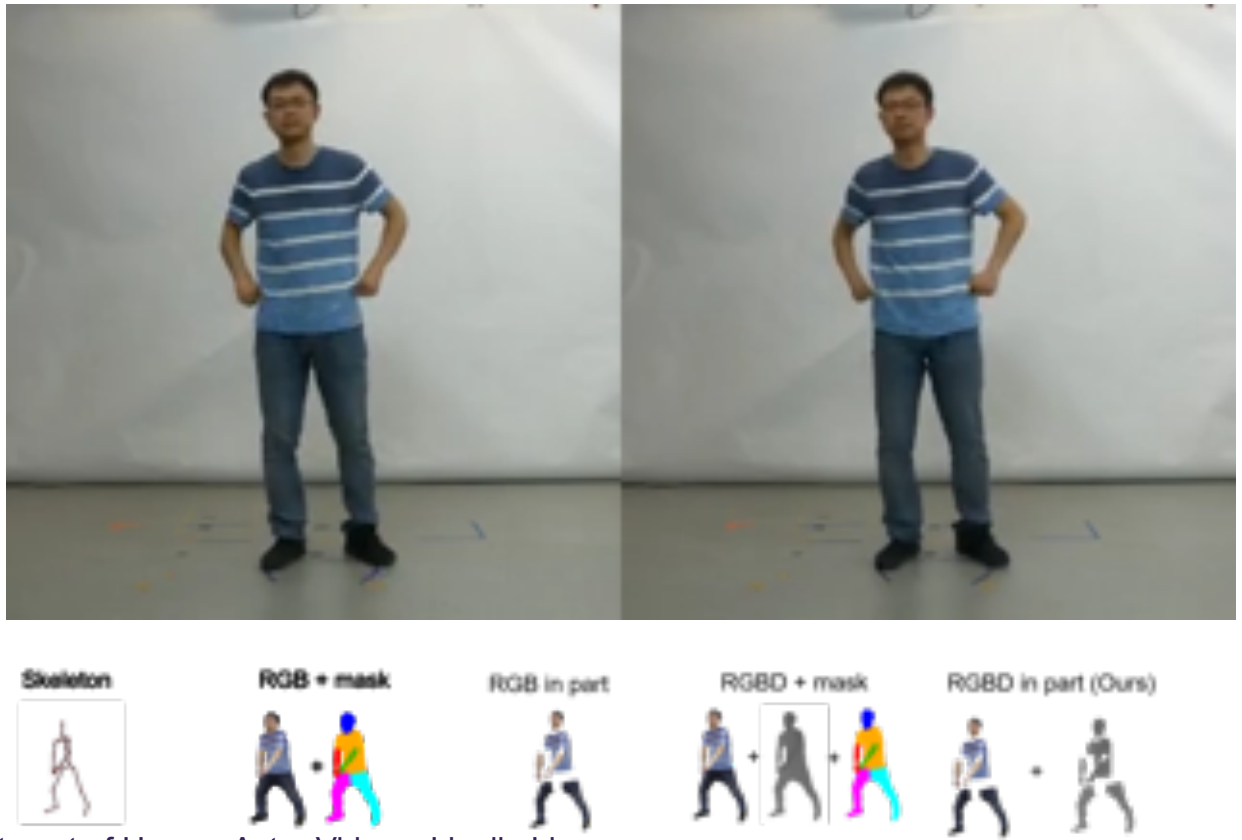


RGB in part

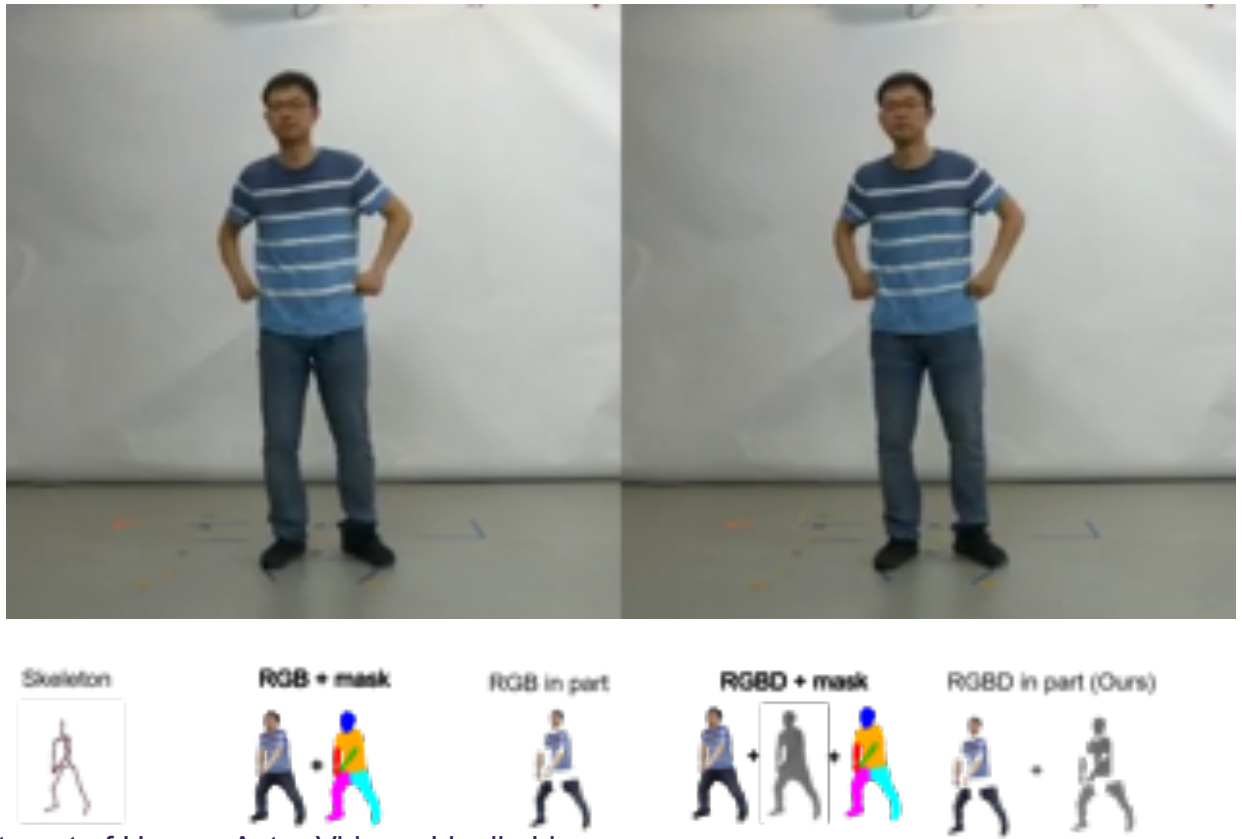
RGBD + mask

RGBD in part (Ours)

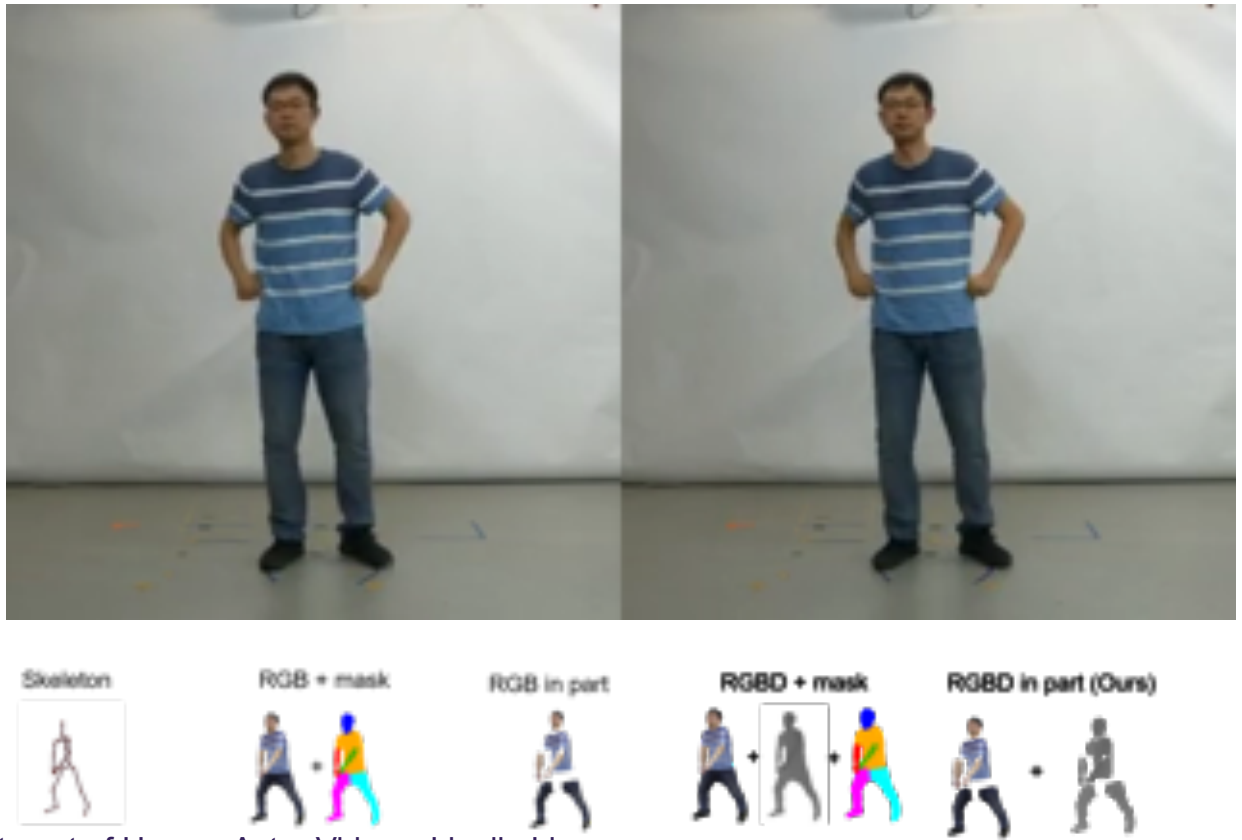
Issue #1: How to choose conditioning inputs



Issue #1: How to choose conditioning inputs



Issue #1: How to choose conditioning inputs

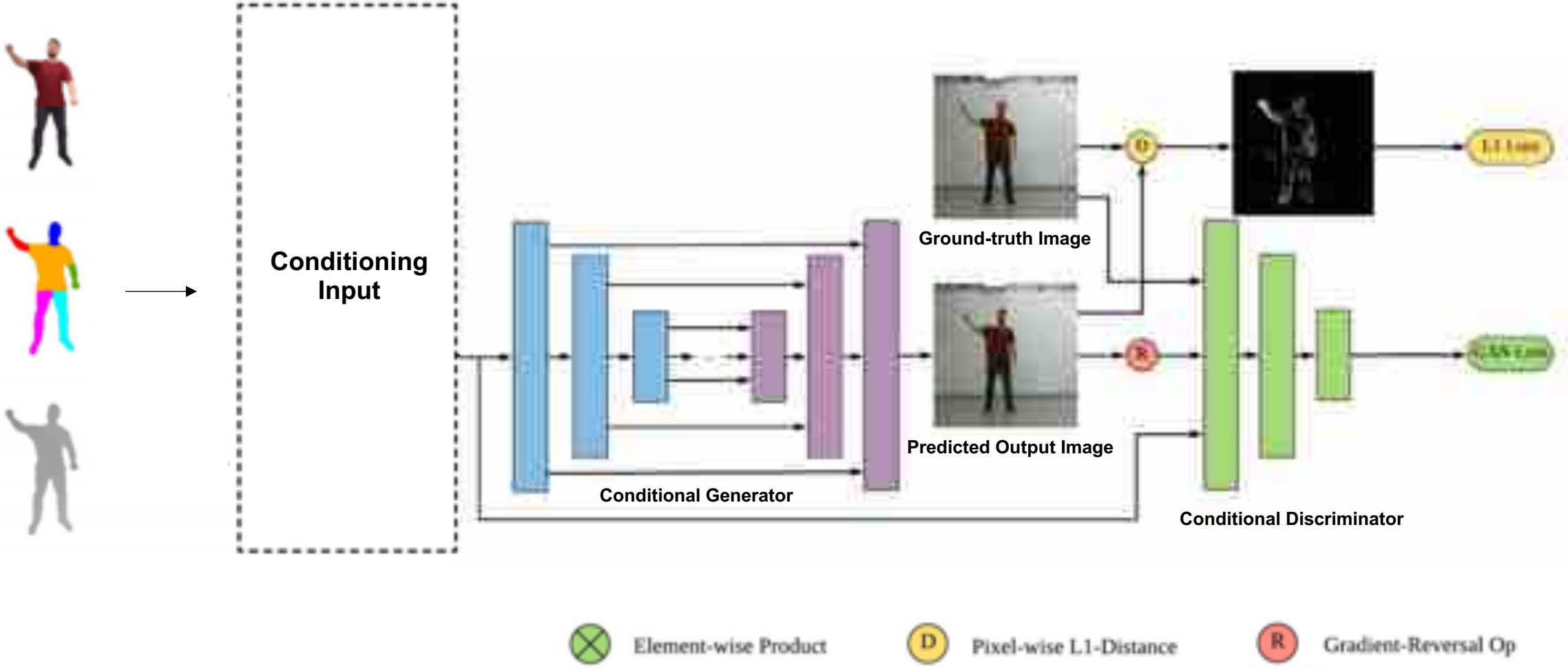


Issue #1: How to choose conditioning input

	L2 error (Lower is better)	SSIM (Higher is better)
Skeleton	17.64	0.60
RGB+mask	16.82	0.63
RGB part	16.12	0.64
RGBD+mask	16.25	0.64
Ours	**15.67**	**0.65**

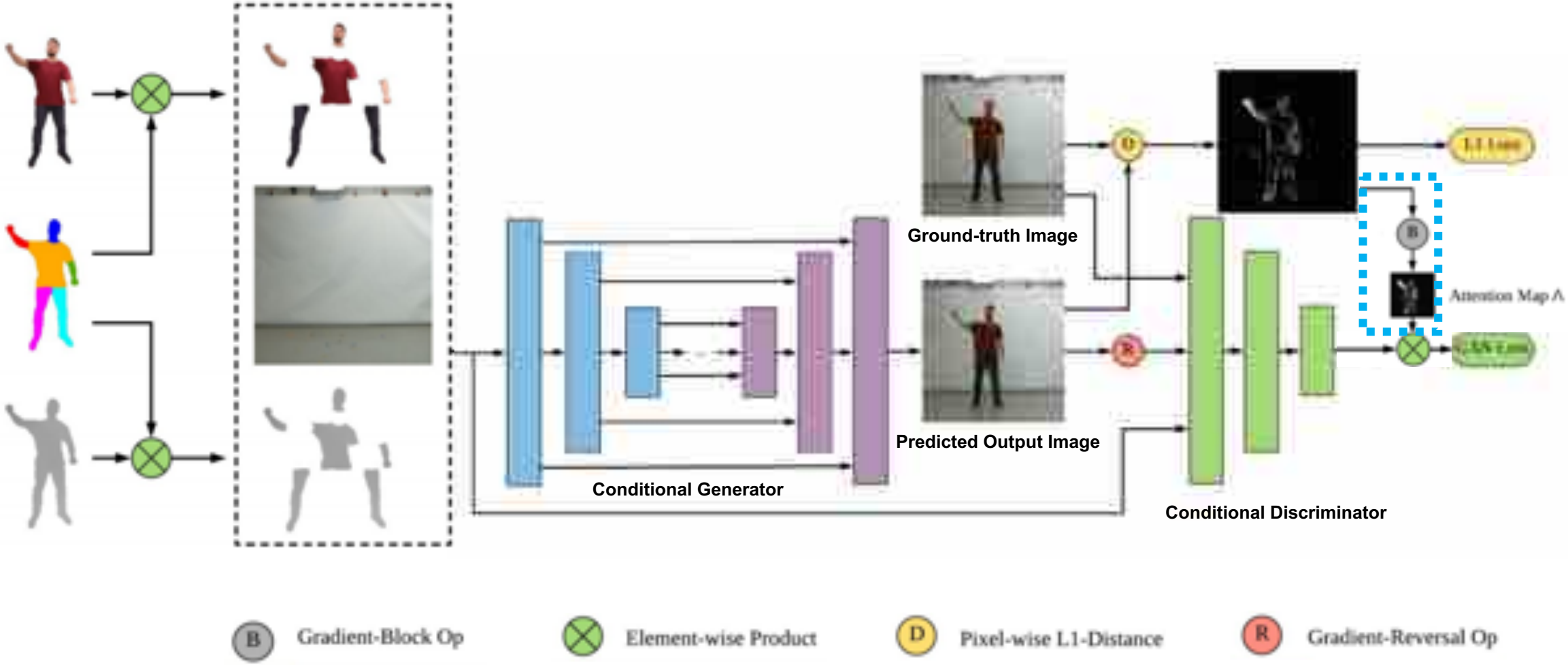
The L2 error and SSIM for the region of the person in the foreground in each image and report the mean value for the whole test sequence

Issue #1: How to choose conditioning inputs



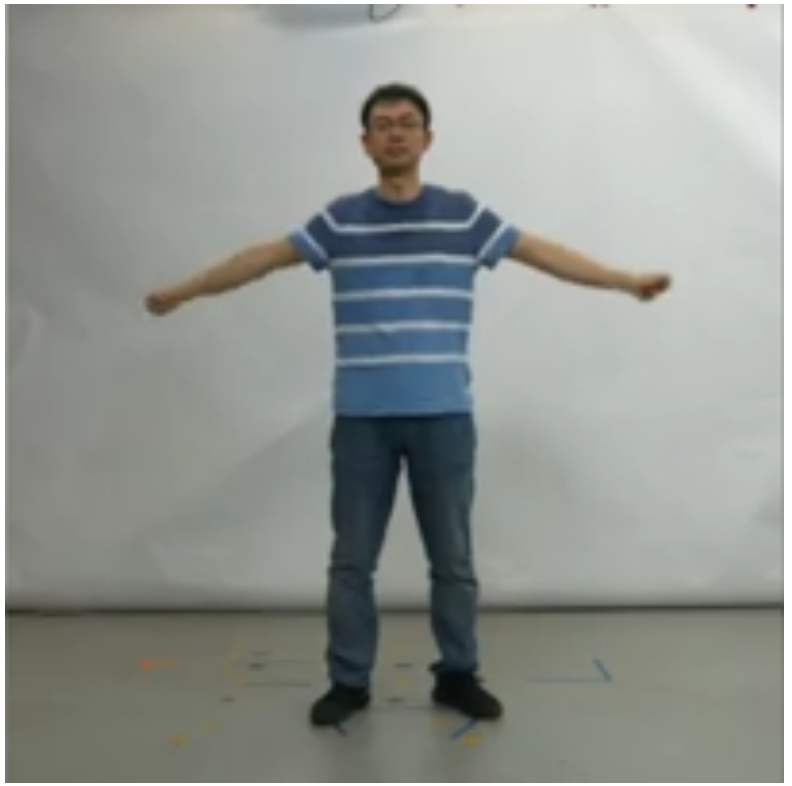
Issue #2: How to design loss function

Issue #2: How to design loss function



Issue #2: How to design loss function

W/o attentive mechanism



With attentive mechanism

Issue #2: How to design loss function

W/o attentive mechanism

With attentive mechanism

Issue #2: How to design loss function

	L2 error (Lower is better)	SSIM (Higher is better)
No attentive	16.39	0.64
Ours	**15.67**	**0.65**

The L2 error and SSIM for the region of the person in the foreground in each image and report the mean value for the whole test sequence

Results

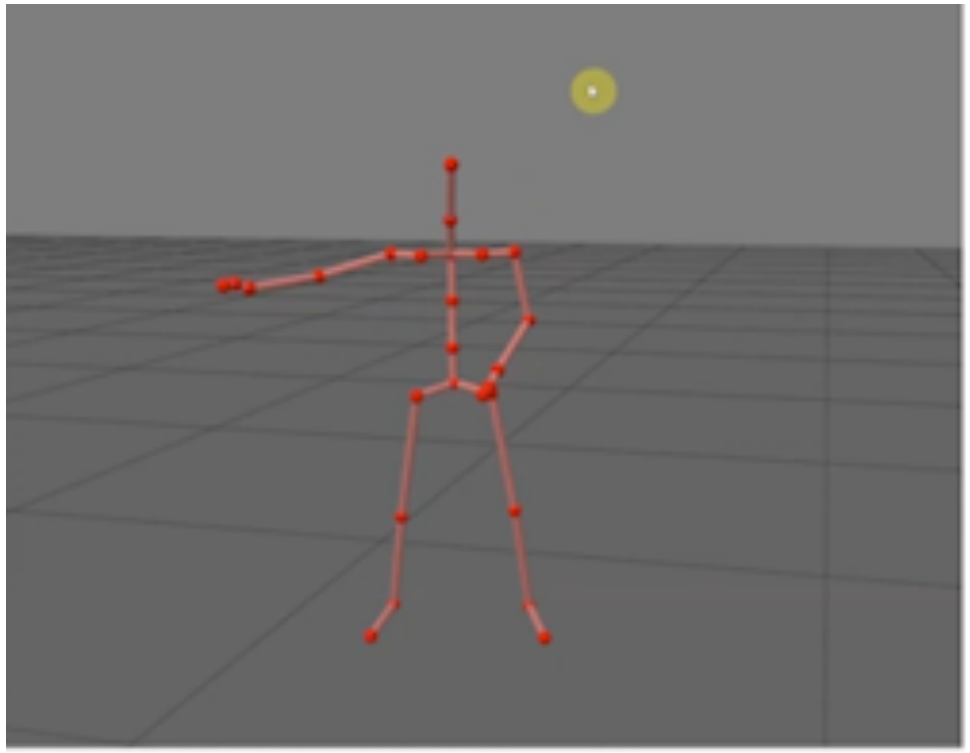
Challenging Motions

Driving Video

Ours (synthesized)

Results

Interactive Editing

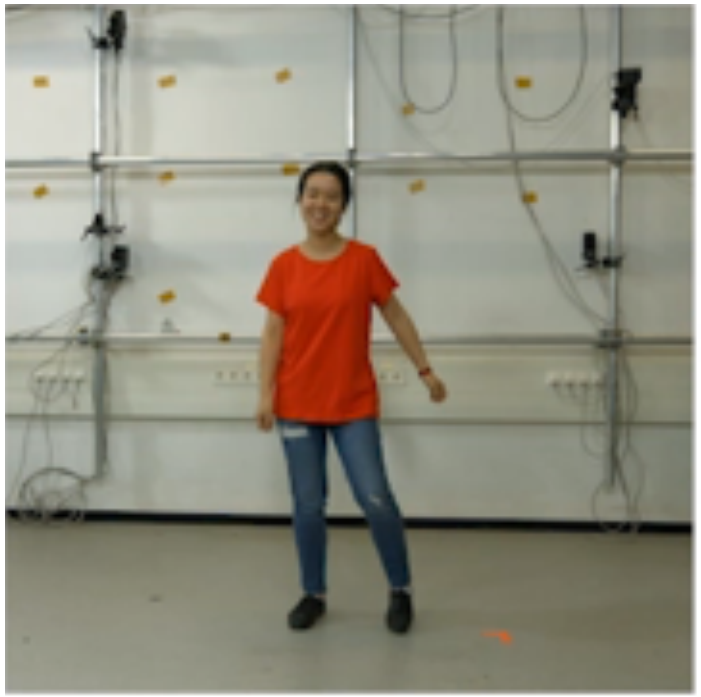


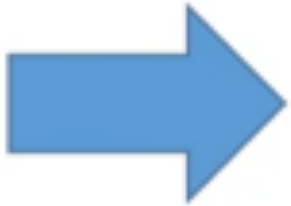
Artist-designed skeleton motion

Ours (synthesized)

Results

Reenactment



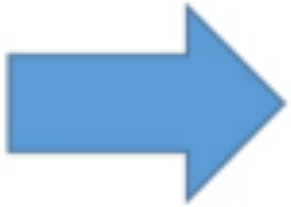


Driving Video

Ours (synthesized)

Results

Youtube Video as Driving Sequence

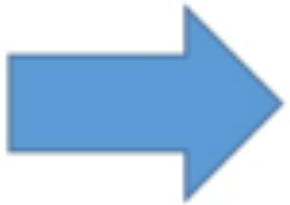


Driving Video

Ours (synthesized)

Results

Youtube Video as Driving Sequence



Driving Video

Ours (synthesized)

Comparison

Driving Video | Ours | [Ma et al. CVPR’18] | [Esser et al. CVPR’18]

Future Work and Limitations

- **Better synthesize highly articulated motions, unseen motions**
- **Handle the interaction with objects**
- **Design a person-agnostic network for generalization to other subjects.**
- **Incorporate a more complicated hand model and finger tracking components.**

Summary

A method for generating video-realistic animations of real humans under user control:

- **No need for a high-quality photorealistic 3D model of the human**
- **Near photo-realistic synthesis results for various motions**
- **Can be used for computer games, visual effects, telepresence, VR/AR.**

Thank you!

More information on the project website
http://gvv.mpi-inf.mpg.de/projects/wxu/HumanReenactment/