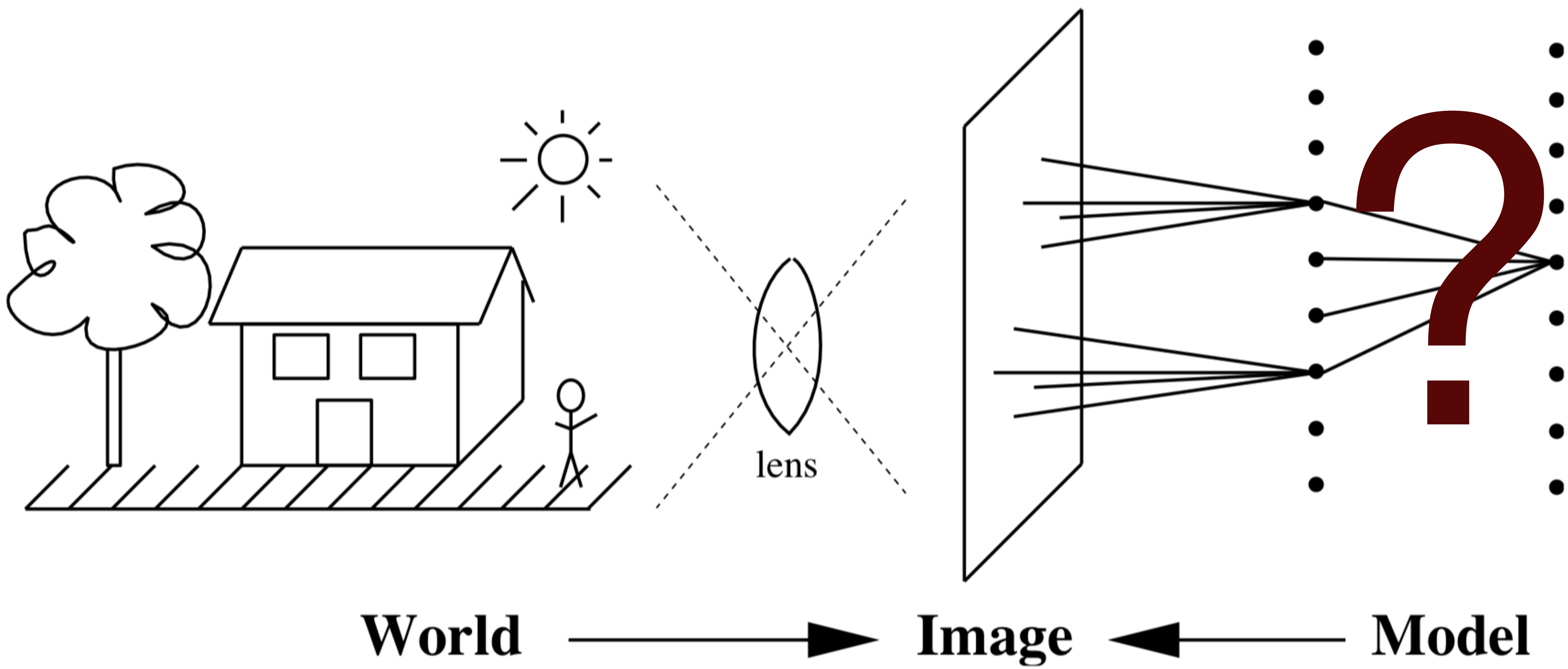
Adversarial Inverse Graphics Networks

Katerina Fragkiadaki

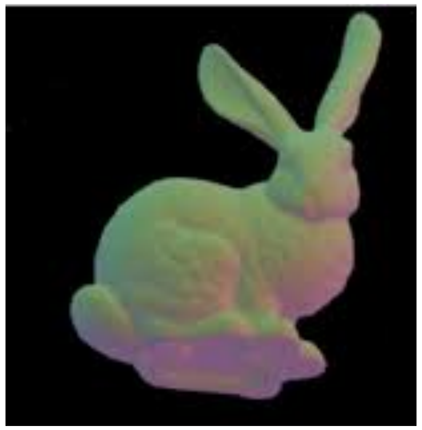
Carnegie Mellon University


lens
World
Image
Model
?

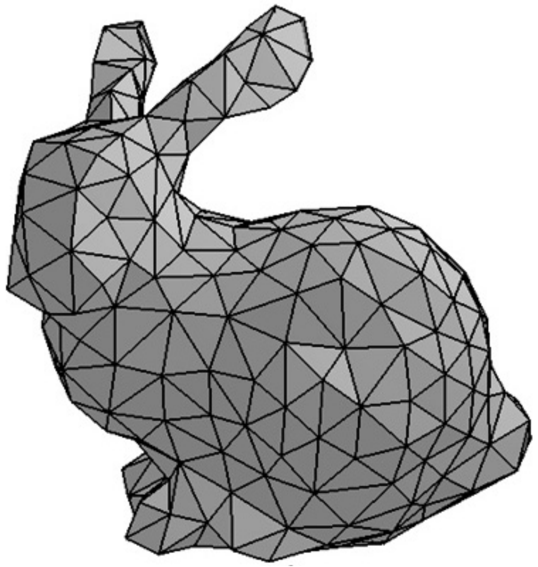
3D representations

depth map

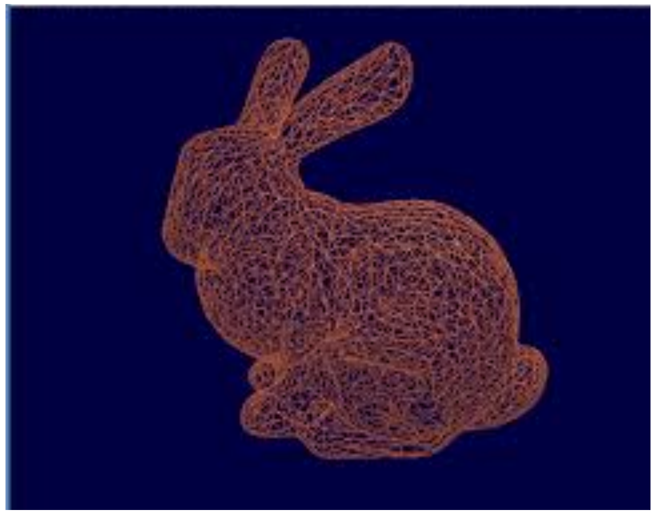
surface normals



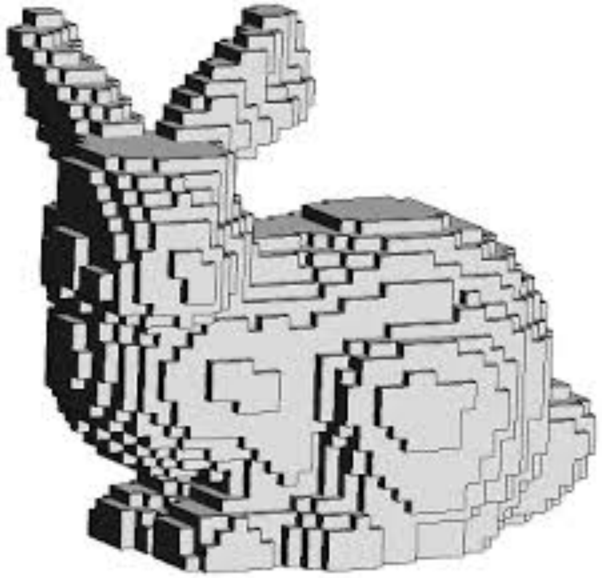
3D mesh



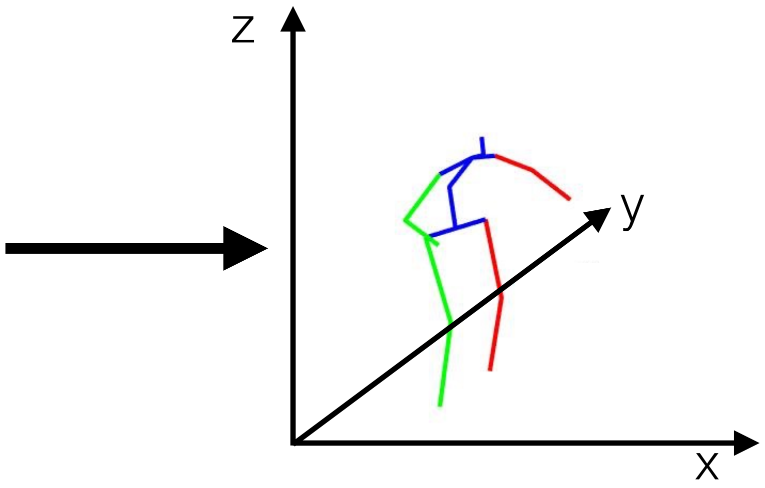
3D point cloud



3D voxel occupancy

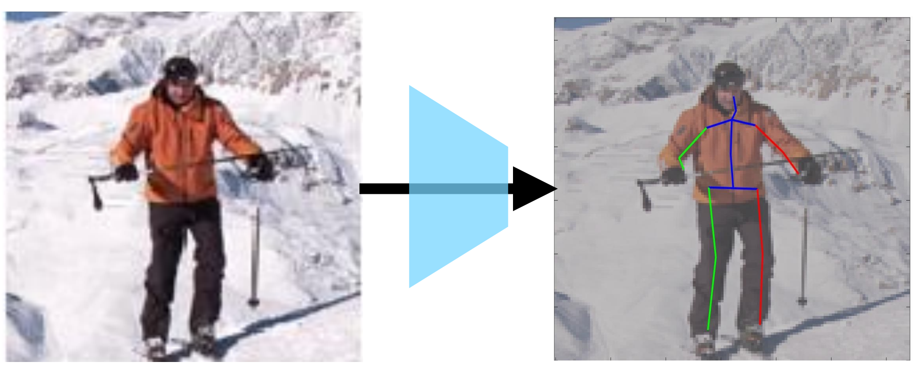


2D-to-3D synthesis



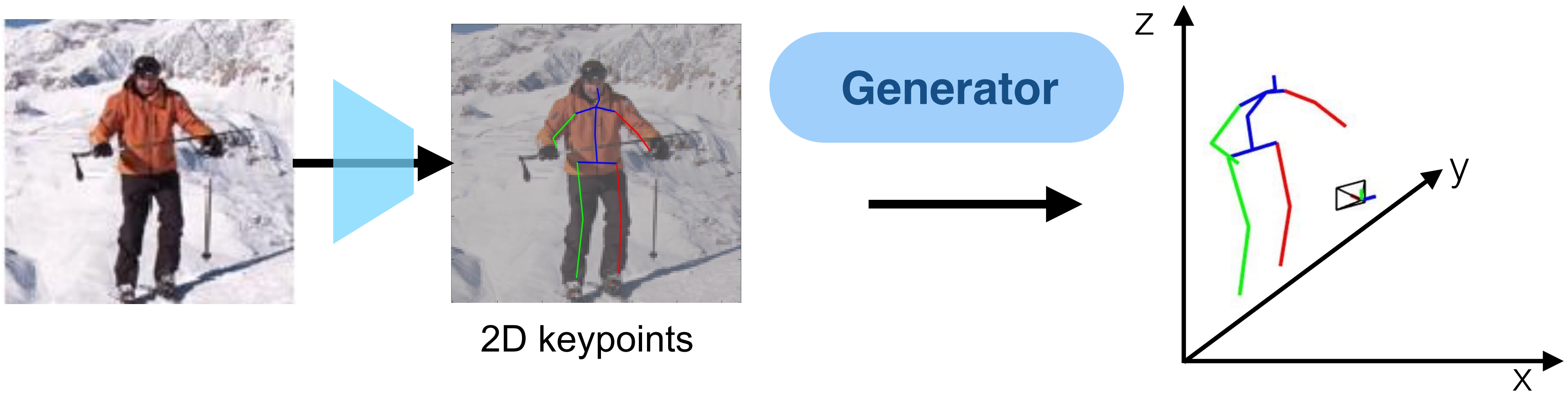
Hard to collect 3D annotations on real images/videos

Can we improve 2D-to-3D synthesis with unlabelled data?

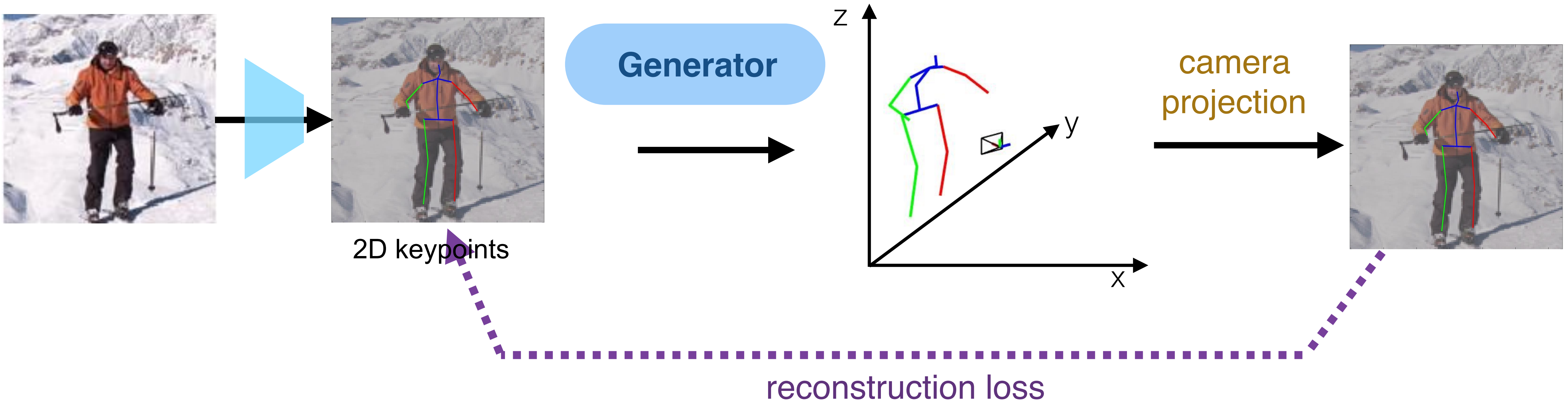


2D keypoints

Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, Cao et al.

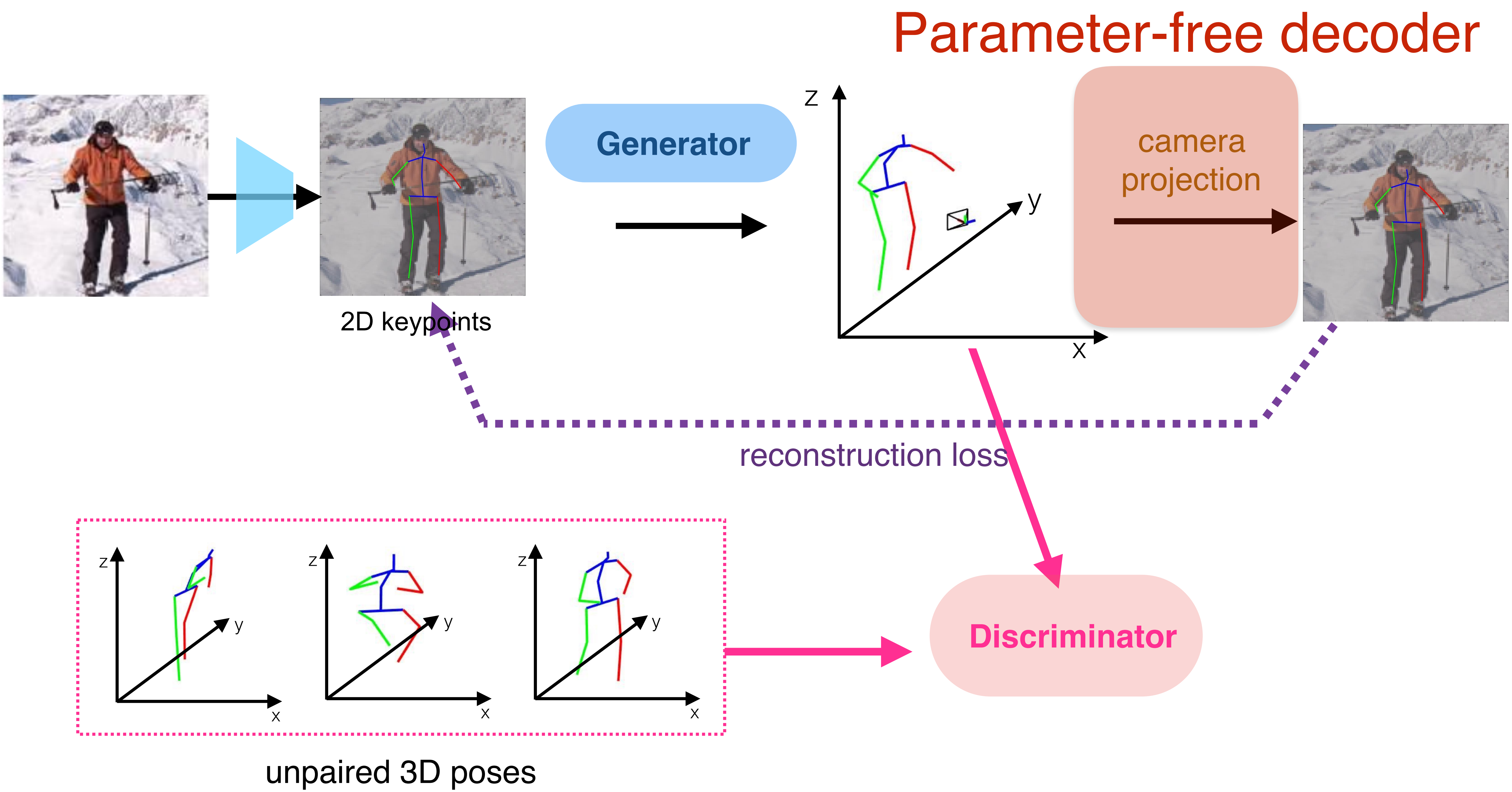


2D keypoints
Generator
z
y
x



Single Image 3D Interpreter Network, Wu et al., 2016

Adversarial Inverse Graphics Networks (AIGNs)



Tung at al. 2017

	Direct	Discuss	Eat	Greet	Phone	Photo	Pose	Purchase	Sit	SitDown	Smoke	Wait	Walk	Average
Forward2Dto3D	75.2	118.4	165.7	95.9	149.1	154.1	77.7	176.9	186.5	193.7	142.7	99.8	74.7	128.9
3Dinterpr [33]	56.3	77.5	96.2	71.6	96.3	106.7	59.1	109.2	111.9	**111.9**	124.2	93.3	58.0	88.6
Monocap [39]	78.0	78.9	88.1	93.9	102.1	115.7	71.0	**90.6**	121.0	118.2	102.5	82.6	75.62	92.3
AIGN (ours)	**53.7**	**71.5**	**82.3**	**58.6**	**86.9**	**98.4**	**57.6**	**104.2**	**100.0**	112.5	**83.3**	**68.9**	**57.0**	**79.0**

Table 1. **3D reconstruction error** in H3.6M using ground-truth 2D keypoints as input.

	Direct	Discuss	Eat	Greet	Phone	Photo	Pose	Purchase	Sit	SitDown	Smoke	Wait	Walk	Average
Forward2Dto3D	80.2	92.4	102.8	92.5	115.5	79.9	119.5	136.7	136.7	144.4	109.3	94.2	80.2	104.6
3Dinterpr [33]	78.6	**90.8**	92.5	89.4	108.9	112.4	77.1	**106.7**	127.4	139.0	103.4	91.4	79.1	98.4
AIGN (ours)	**77.6**	91.4	**89.9**	**88**	**107.3**	**110.1**	**75.9**	107.5	**124.2**	**137.8**	**102.2**	**90.3**	**78.6**	**97.2**

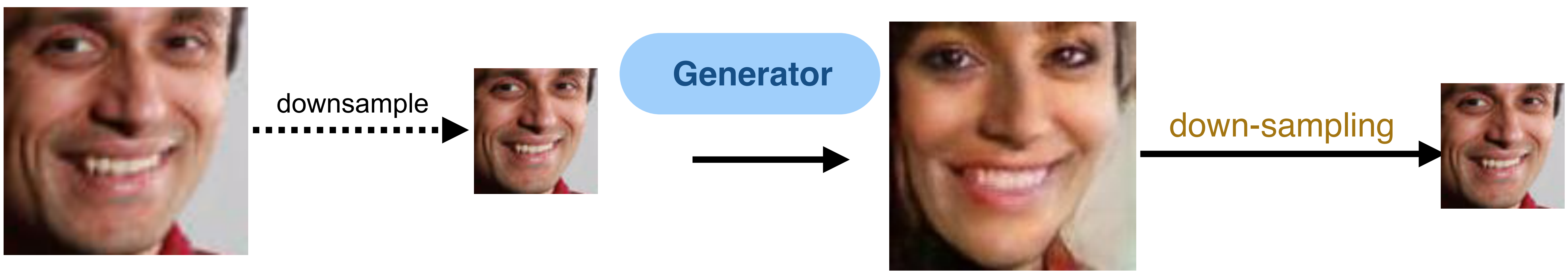
Table 2. **3D reconstruction error** in H3.6M using detected 2D keypoints as input.

AIGNs for Image-to-Image translation

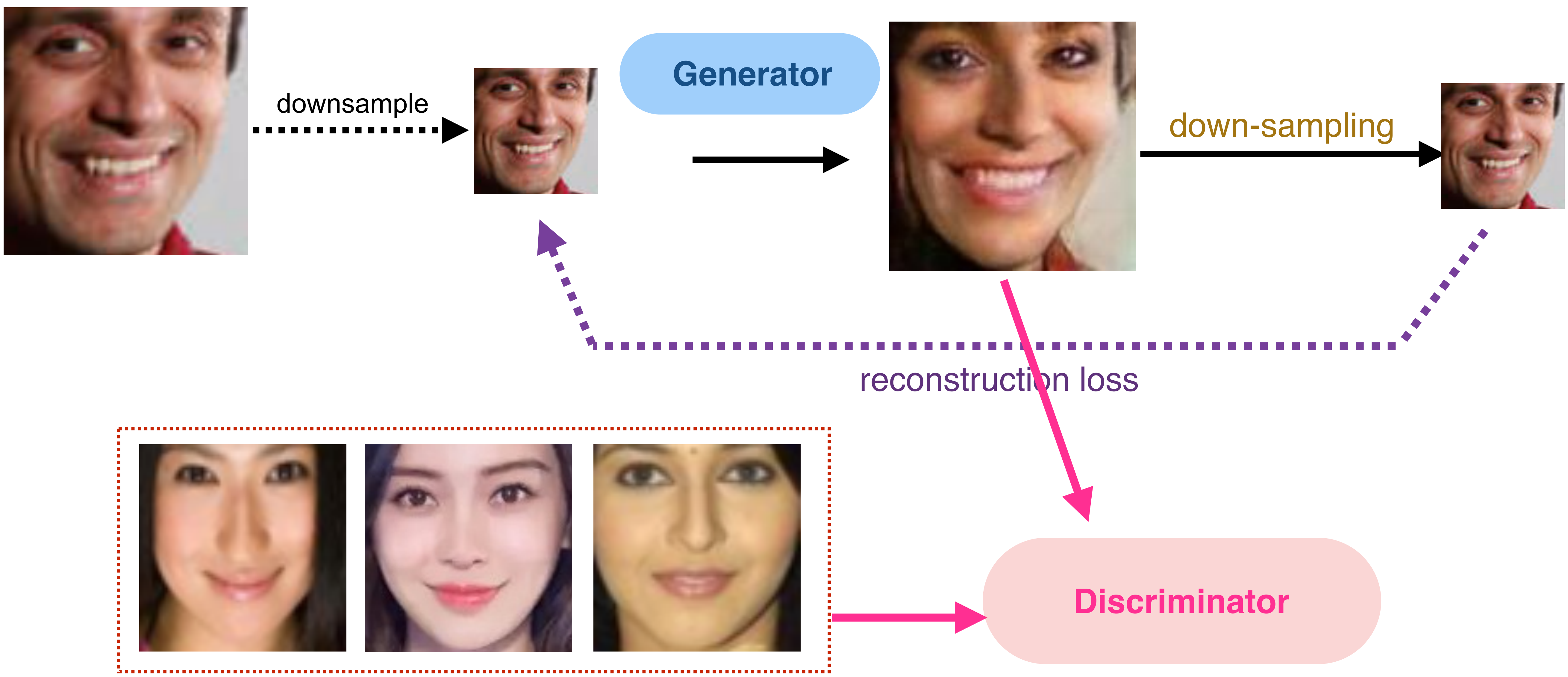
AIGNs for Image-to-Image translation

AIGNs for Image-to-Image translation

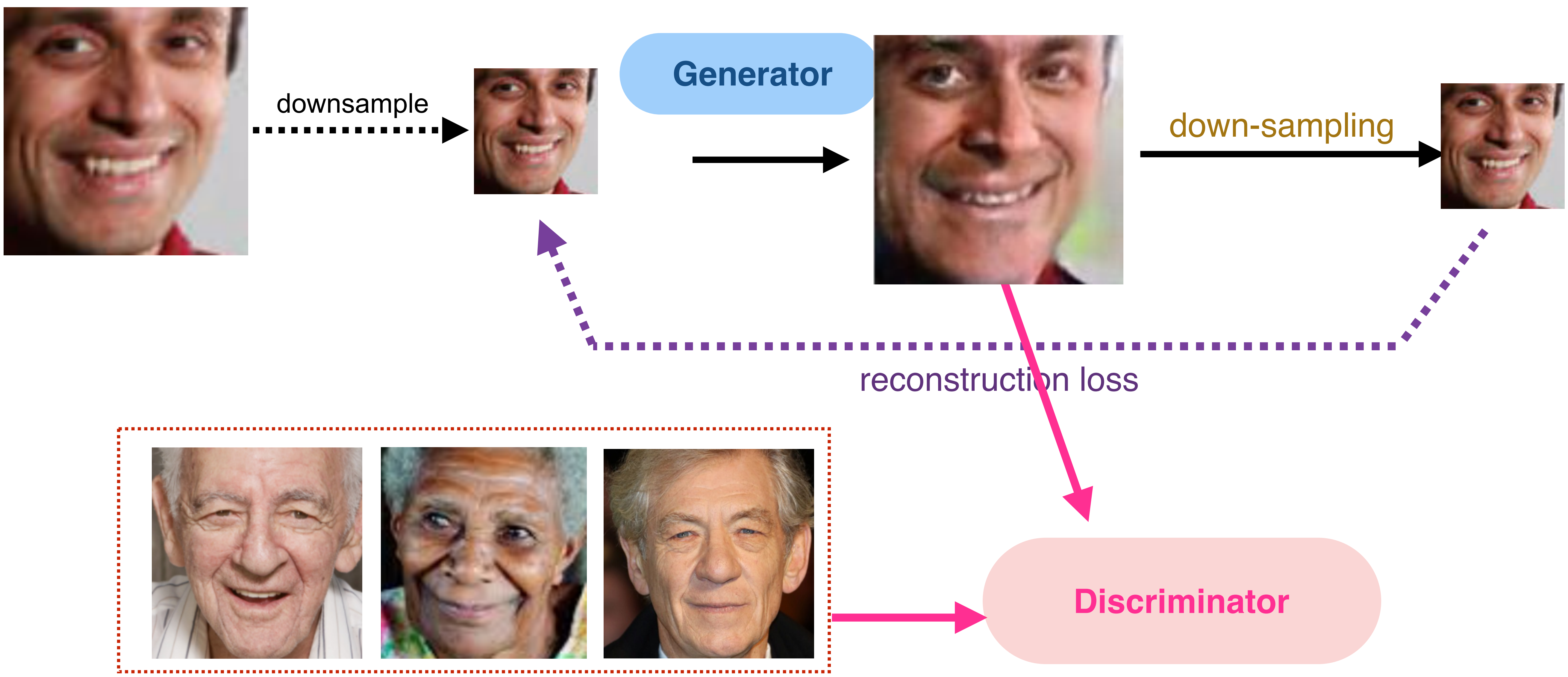
AIGNs for Image-to-Image translation



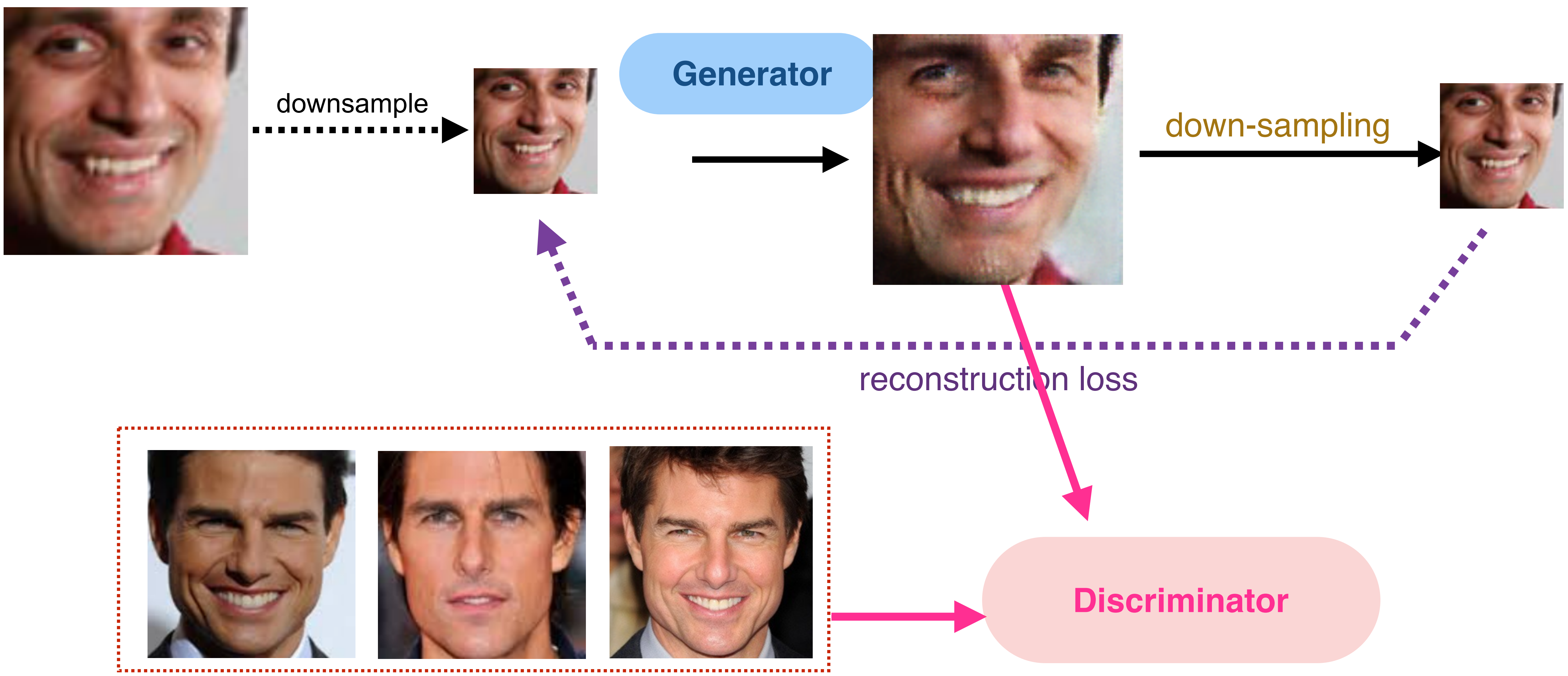
AIGNs for Image-to-Image translation



AIGNs for Image-to-Image translation



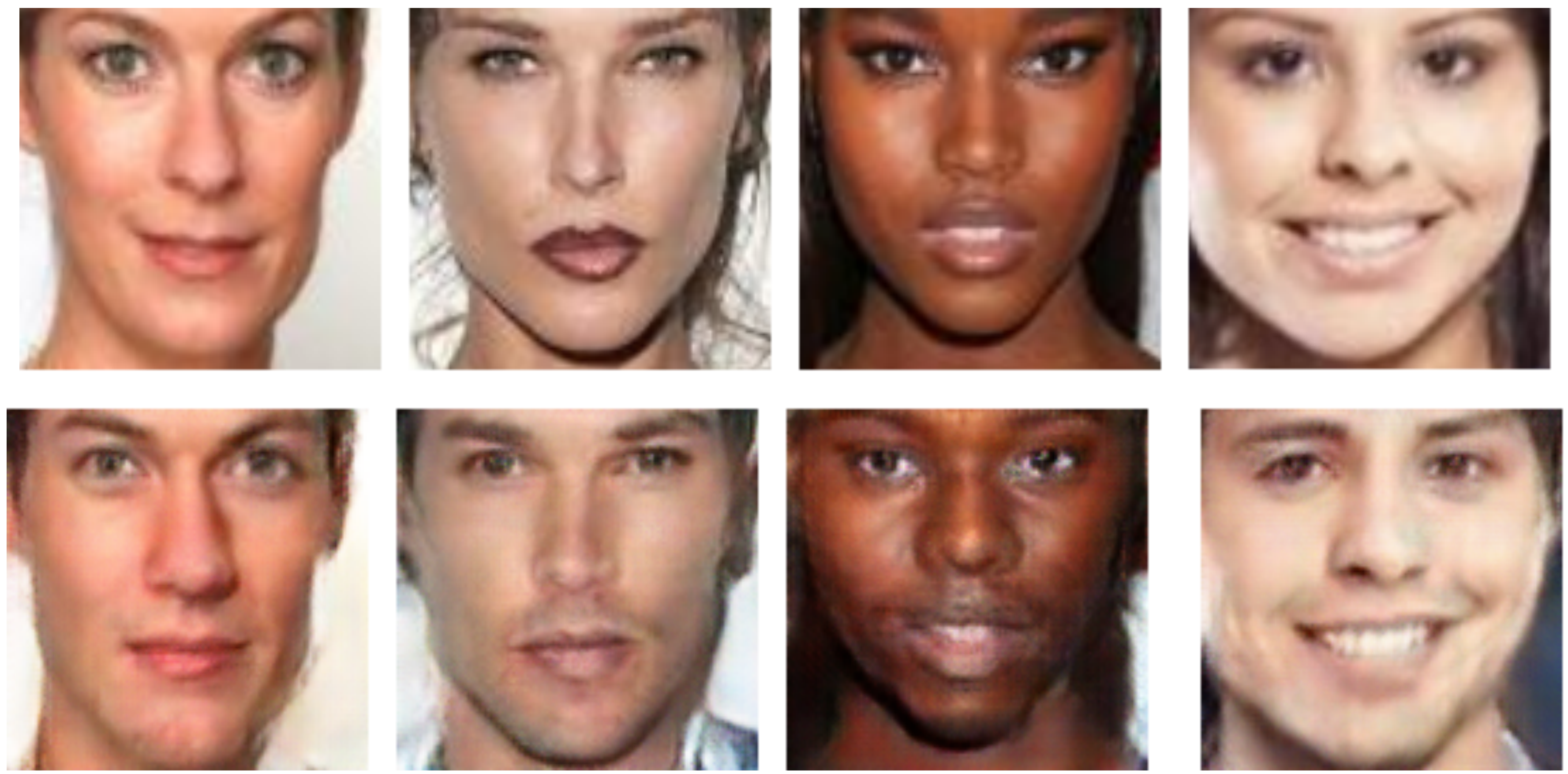
AIGNs for Image-to-Image translation



Male -> Female

CV researcher -> Tom Cruise

Female to male



Female to male

To older

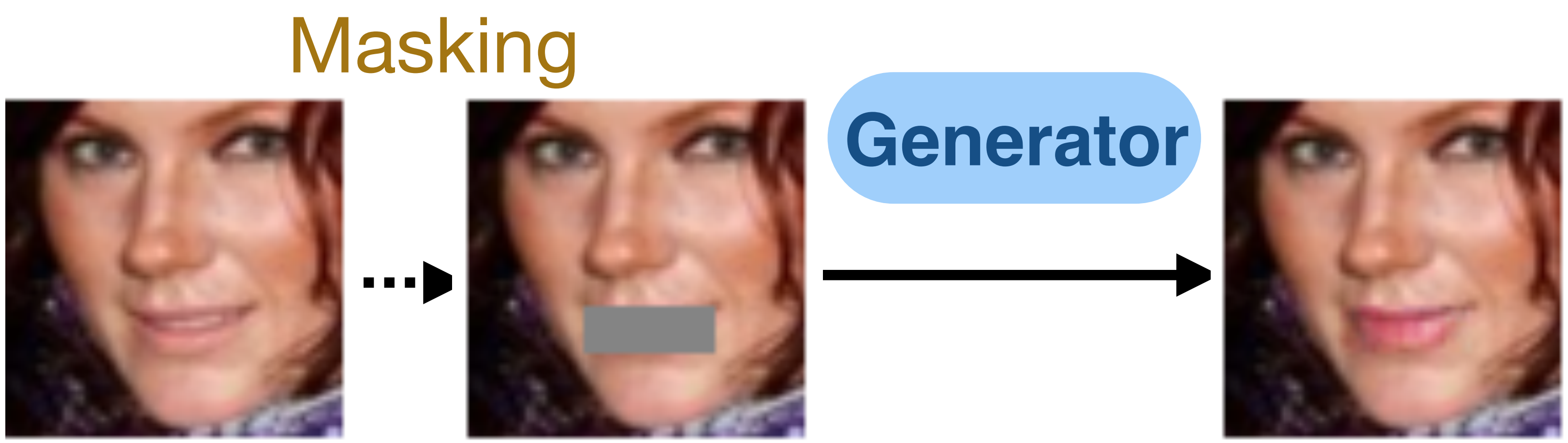
To younger

To younger

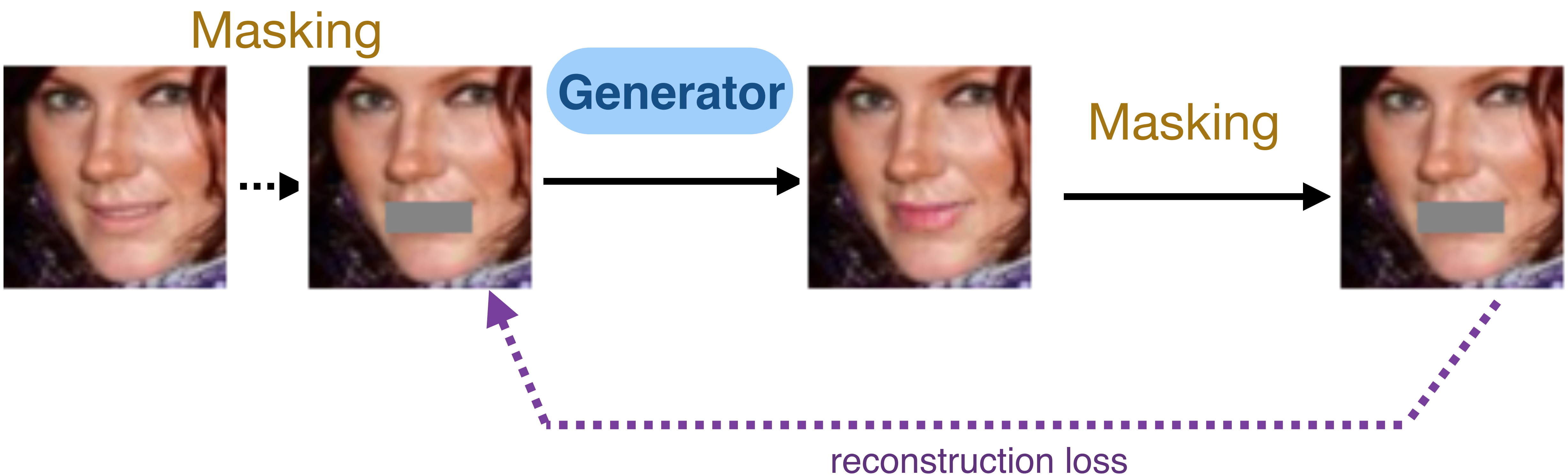
AIGNs for plastic surgary

Masking

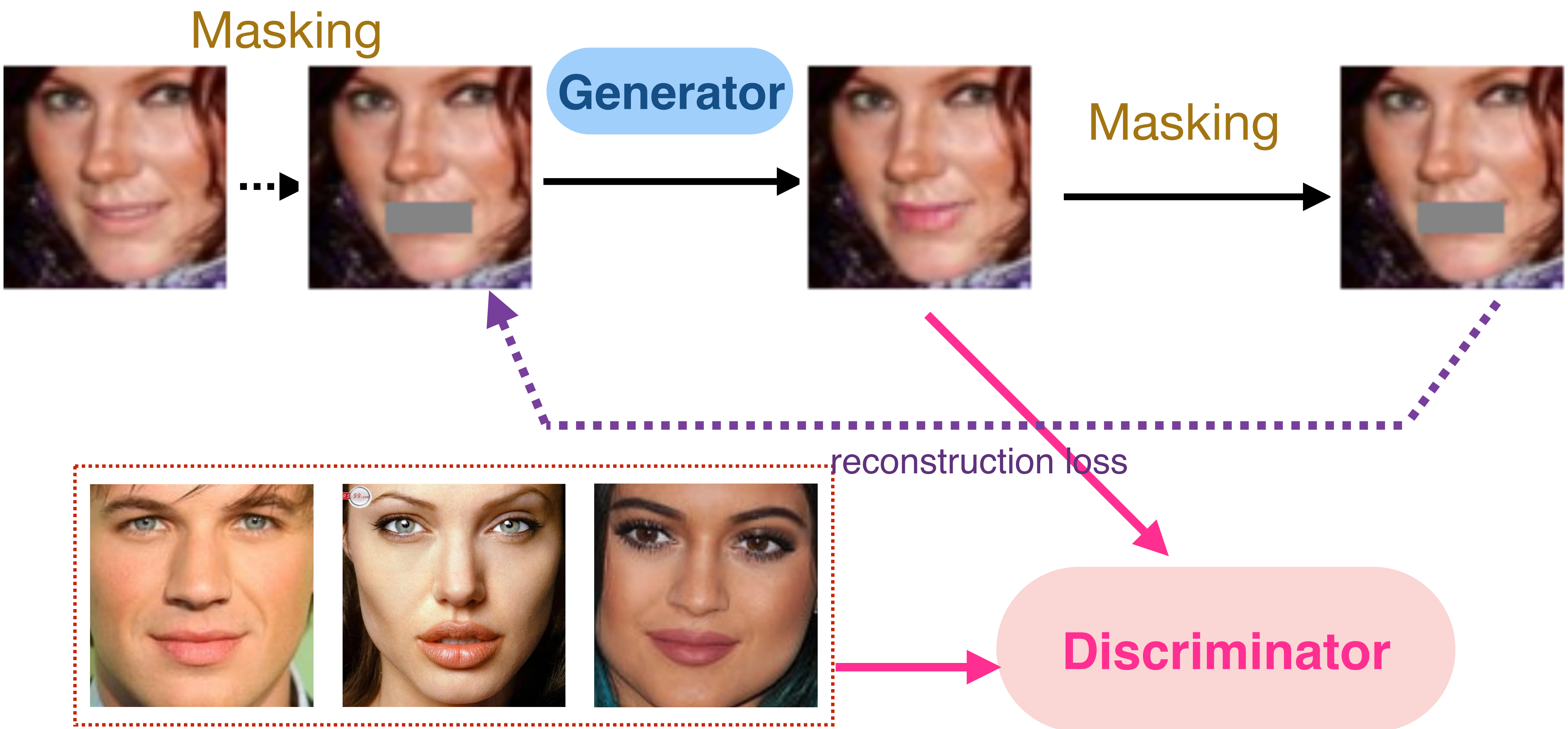
AIGNs for plastic surgary



AIGNs for plastic surgary



AIGNs for plastic surgary

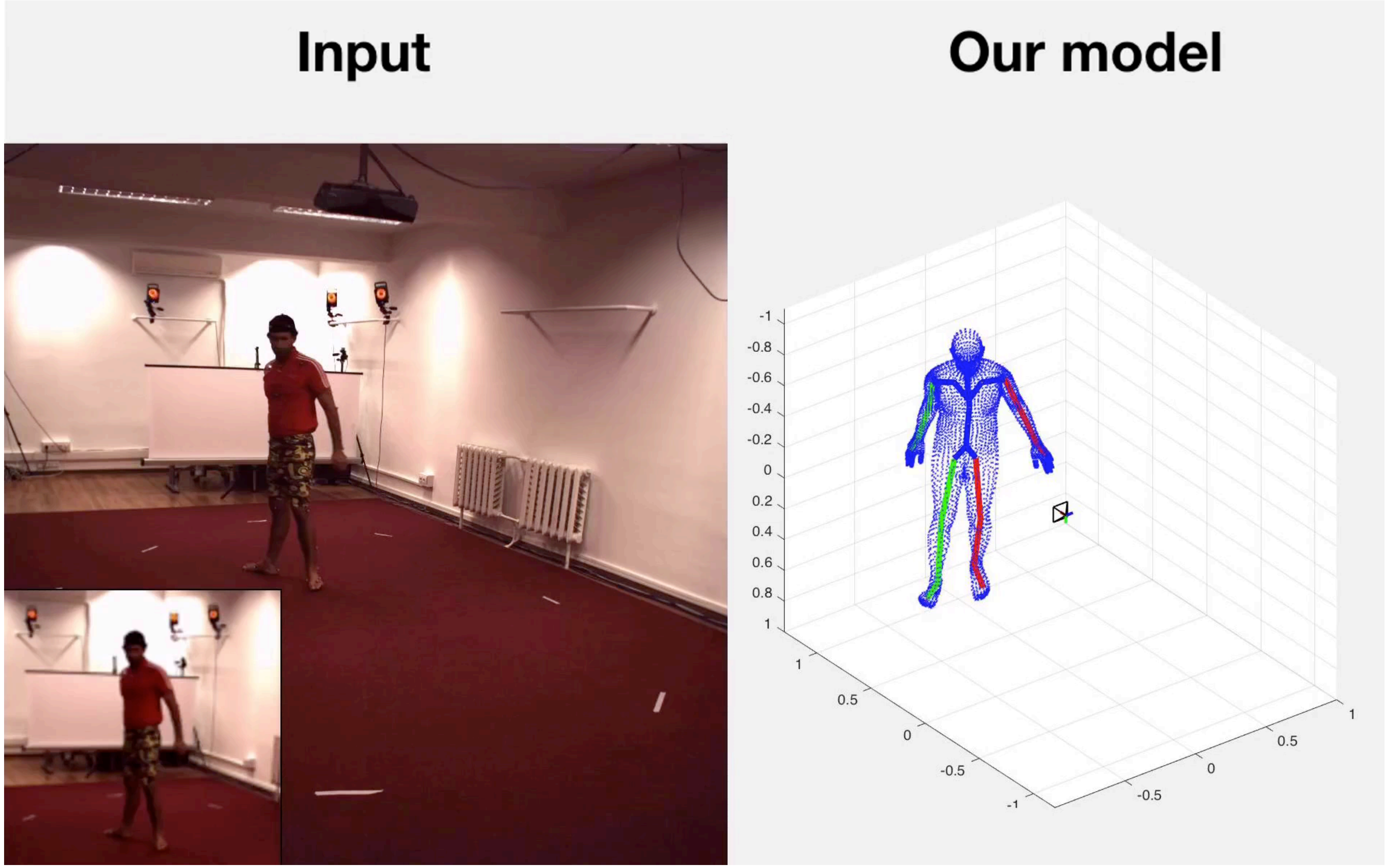


To bigger lips

To bigger lips

2D-to-3D synthesis

Recover a human 3D mesh from 2D videos



Can we improve with unlabelled data?

Self-supervised learning of motion capture, Tung et al. NIPS 2017

3D human shape model

SMPL [M. Loper et al.]: a low-parametric model learned from aligning high-resolution 3D scans.

3D mesh

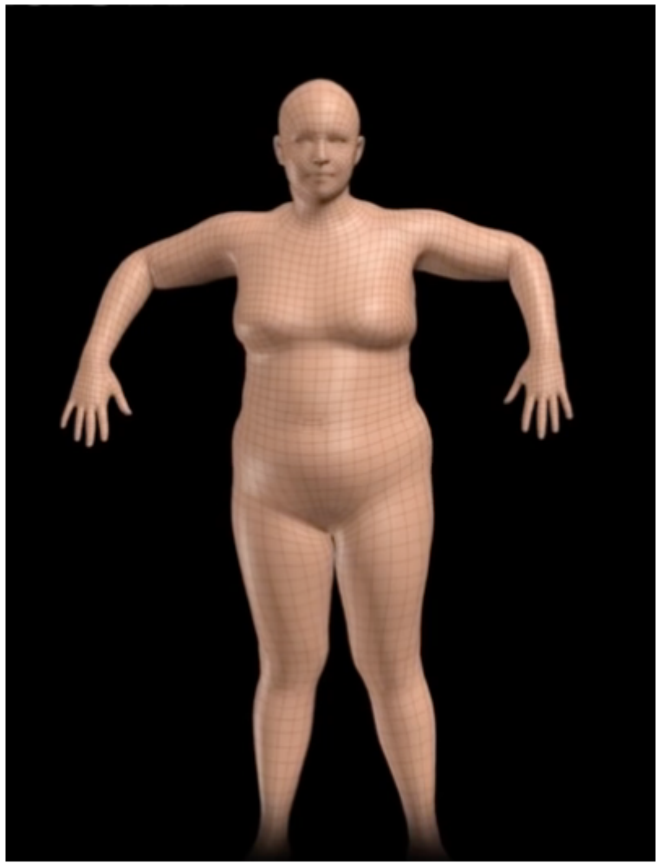
SMPL(θ , β)

Pose

θ

Shape

β



SMPL: A Skinned Multi-Person Linear Model Loper et al. SIGGRAPH Asia 2015

3D human shape model

SMPL [M. Loper et al.]: a low-parametric model learned from aligning high-resolution 3D scans.

3D mesh

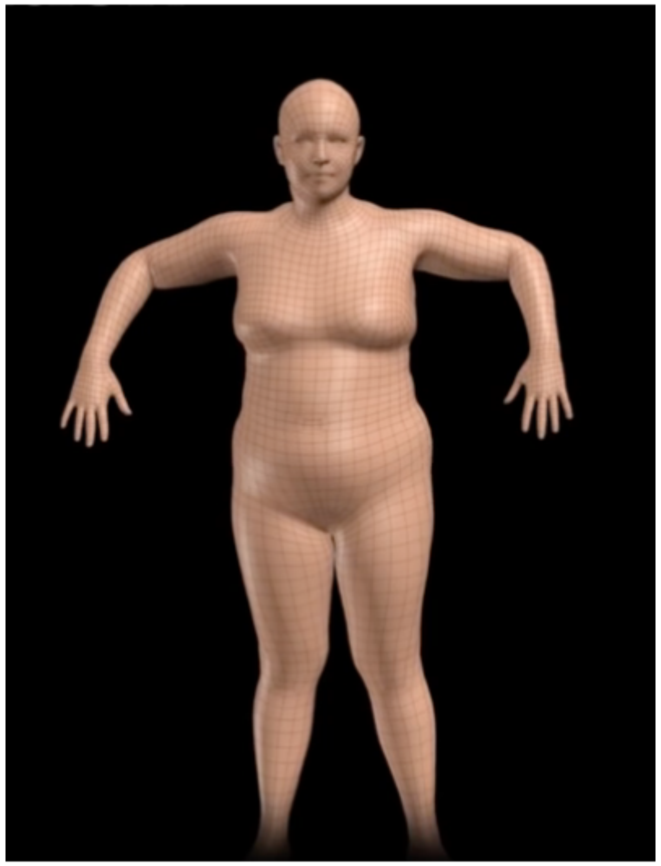
SMPL(θ , β)

Pose

θ

Shape

β



SMPL: A Skinned Multi-Person Linear Model Loper et al. SIGGRAPH Asia 2015

3D human shape model

SMPL [M. Loper et al.]: a low-parametric model learned from aligning high-resolution 3D scans.

3D mesh

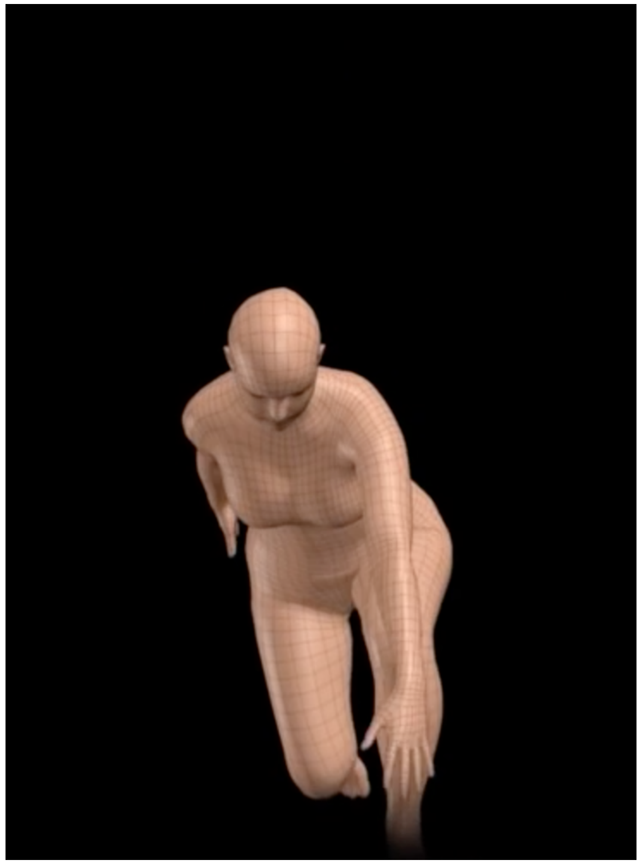
SMPL(θ , β)

Pose

θ

Shape

β



SMPL: A Skinned Multi-Person Linear Model Loper et al. SIGGRAPH Asia 2015

3D human shape model

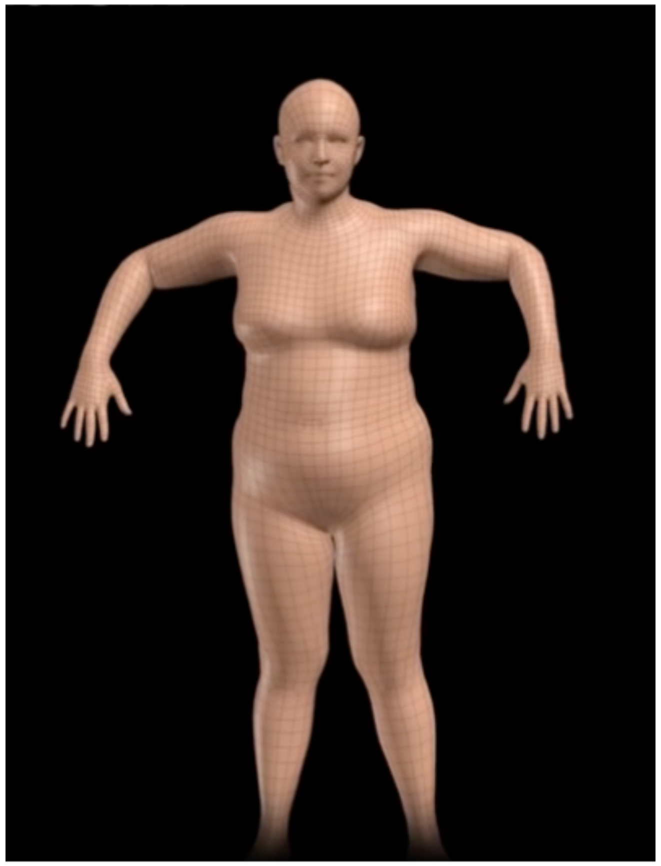
SMPL [M. Loper et al.]: a low-parametric model learned from aligning high-resolution 3D scans.

3D mesh

SMPL(θ , β)

Pose θ

Shape β $\longrightarrow$



SMPL: A Skinned Multi-Person Linear Model Loper et al. SIGGRAPH Asia 2015

3D human shape model

SMPL [M. Loper et al.]: a low-parametric model learned from aligning high-resolution 3D scans.

3D mesh

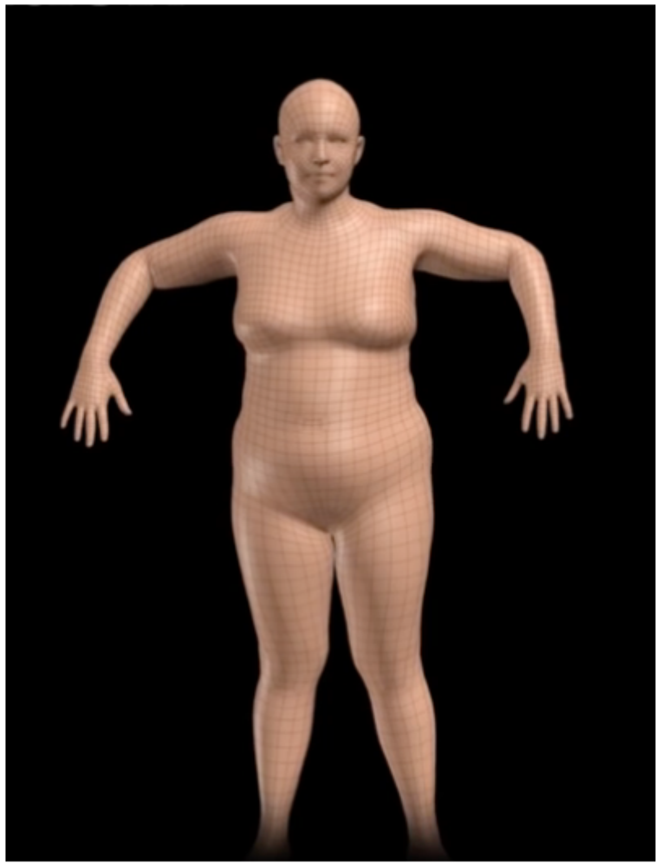
SMPL(θ , β)

Pose

θ

Shape

β



SMPL: A Skinned Multi-Person Linear Model Loper et al. SIGGRAPH Asia 2015

3D human shape model

SMPL [M. Loper et al.]: a low-parametric model learned from aligning high-resolution 3D scans.

3D mesh

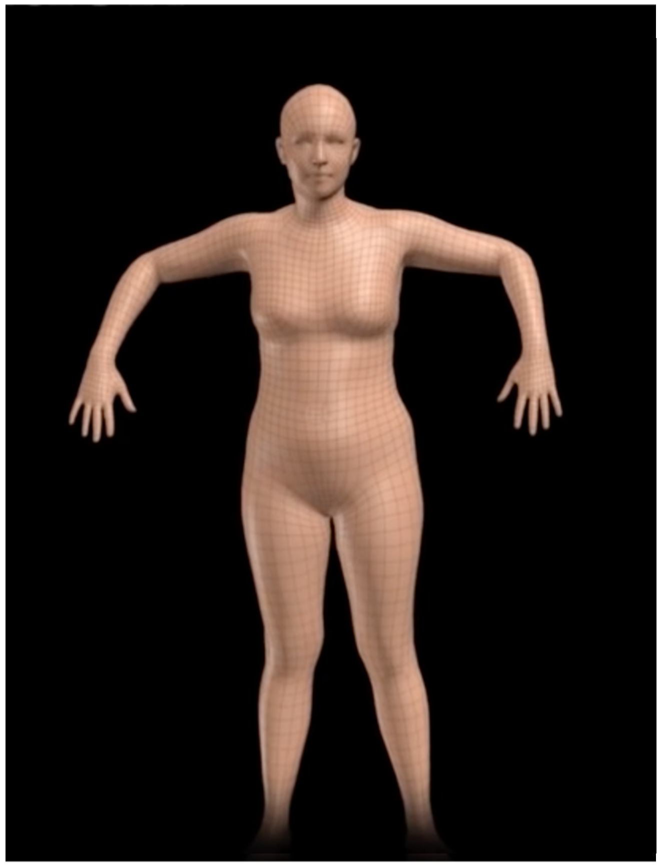
SMPL(θ , β)

Pose

θ

Shape

β



SMPL: A Skinned Multi-Person Linear Model Loper et al. SIGGRAPH Asia 2015

3D human shape model

SMPL [M. Loper et al.]: a low-parametric model learned from aligning high-resolution 3D scans.

Pose

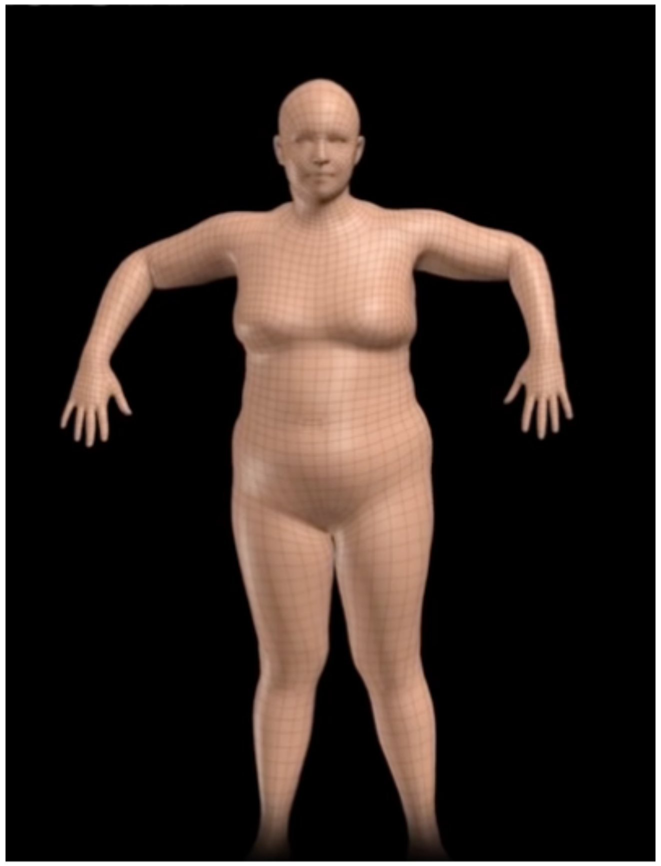
θ

Shape

β

3D mesh

SMPL(θ , β)



SMPL: A Skinned Multi-Person Linear Model Loper et al. SIGGRAPH Asia 2015

3D human shape model

SMPL [M. Loper et al.]: a low-parametric model learned from aligning high-resolution 3D scans.

3D mesh

SMPL(θ , β)

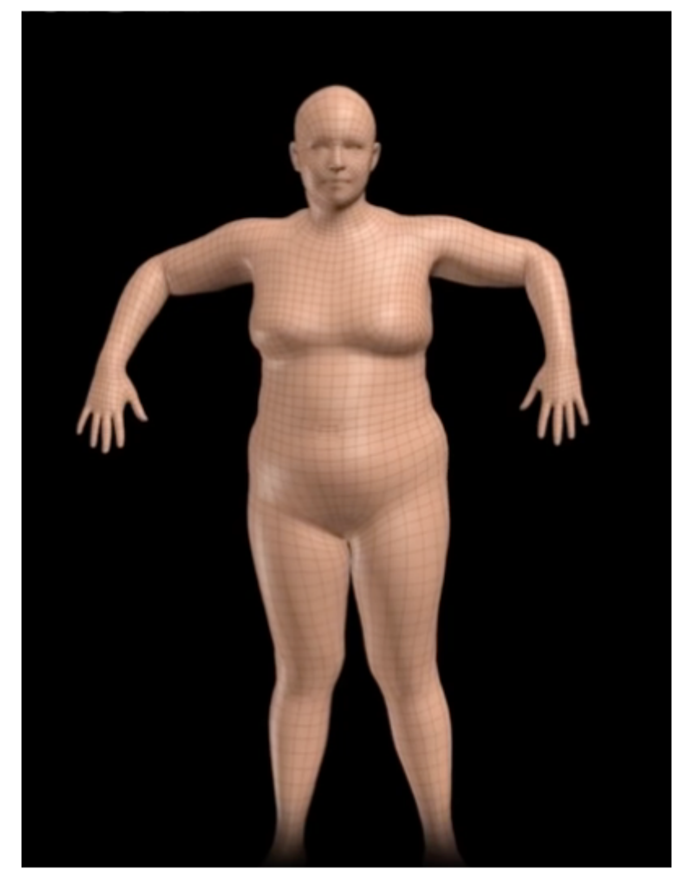
Pose

θ

Shape

β

Differentiable mapping

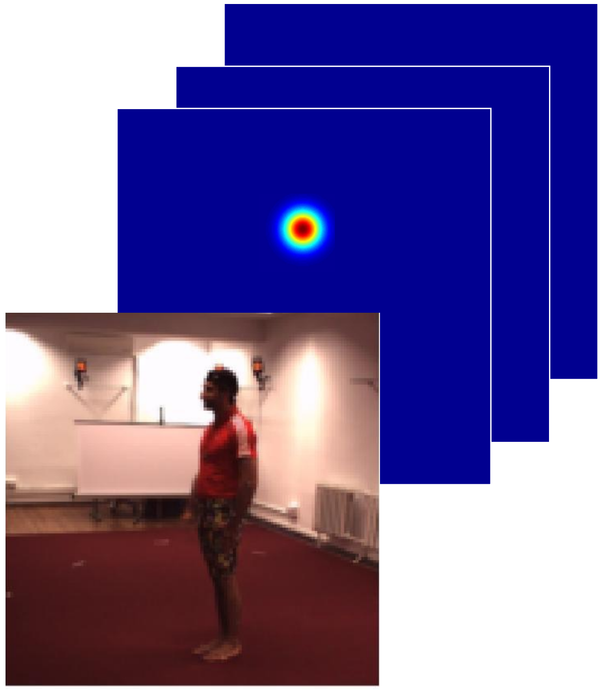


SMPL: A Skinned Multi-Person Linear Model Loper et al. SIGGRAPH Asia 2015

RGB - to - 3D mesh

Inputs:

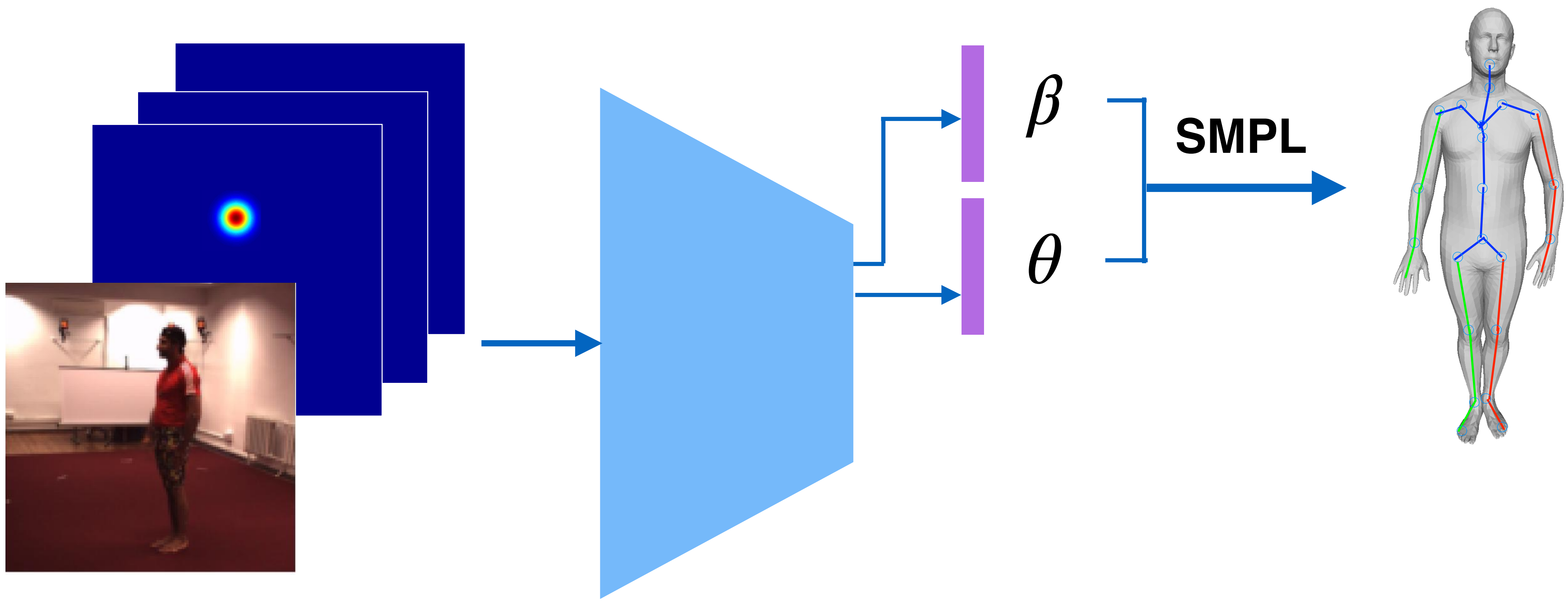
RGB frame
2D keypoint heatmaps



RGB - to - 3D mesh

Inputs:
RGB frame
2D keypoint heatmaps

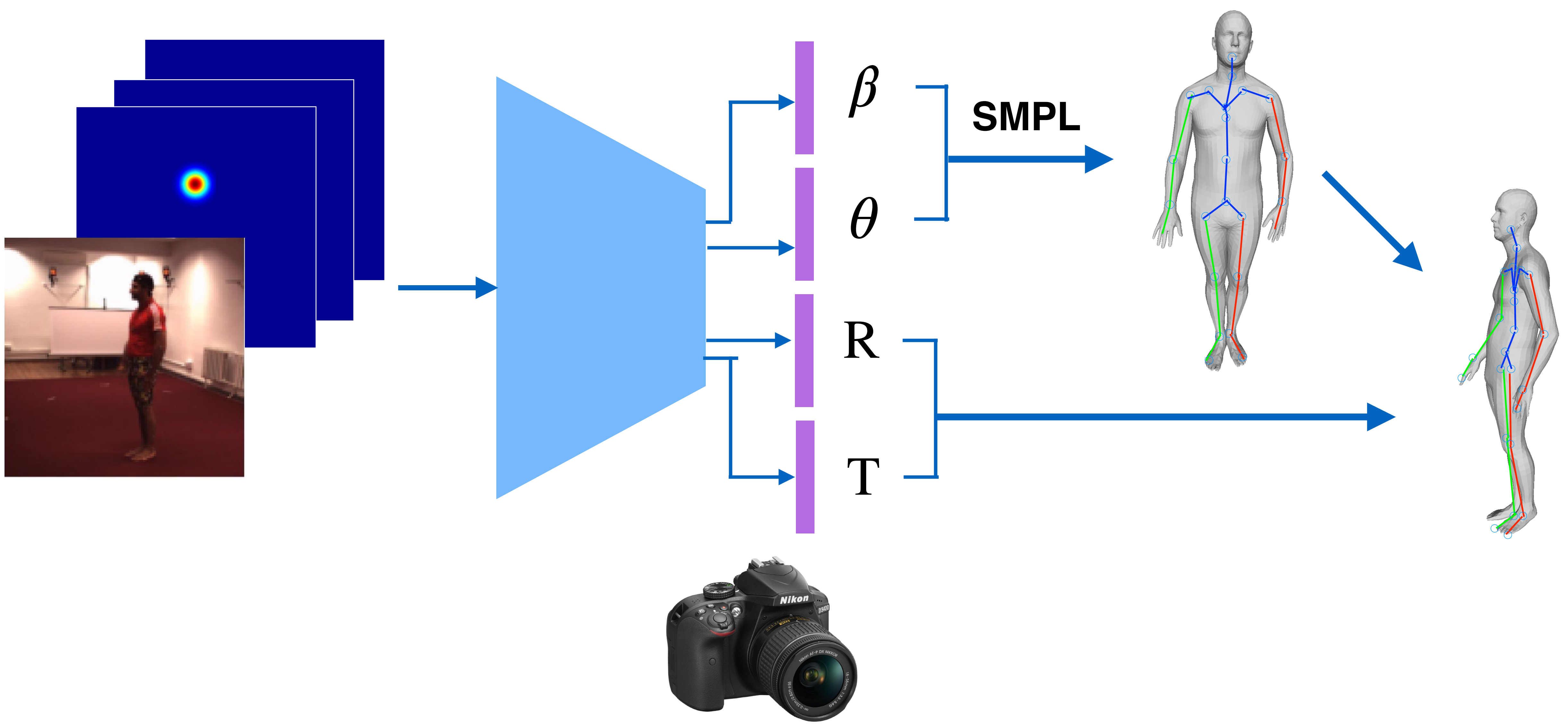
Outputs:
SMPL parameters (β , θ)



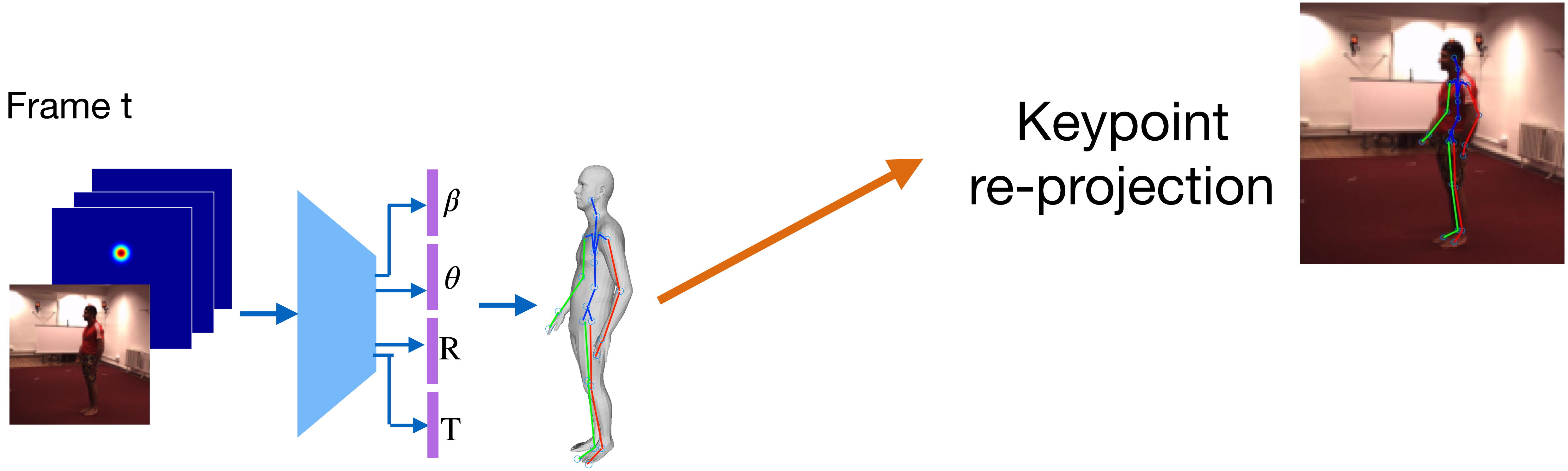
Our model

Inputs:
RGB frame
2D keypoint heatmaps

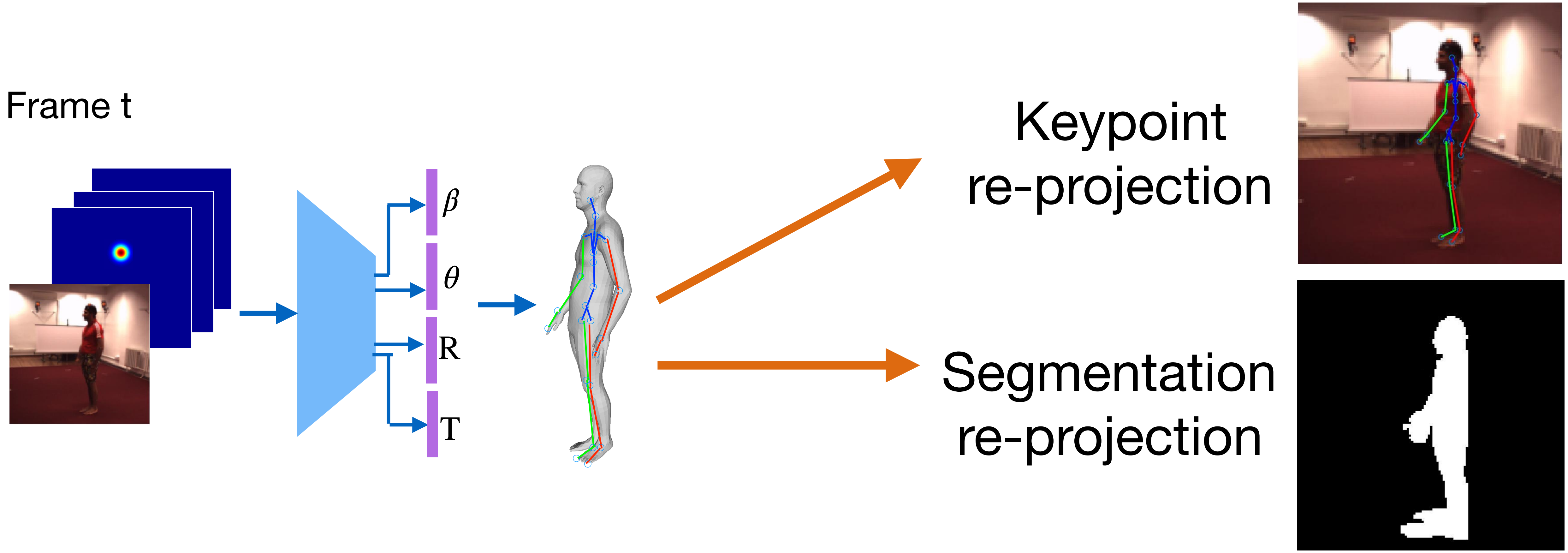
Outputs:
SMPL parameters (β , θ)
camera parameters(R, T)



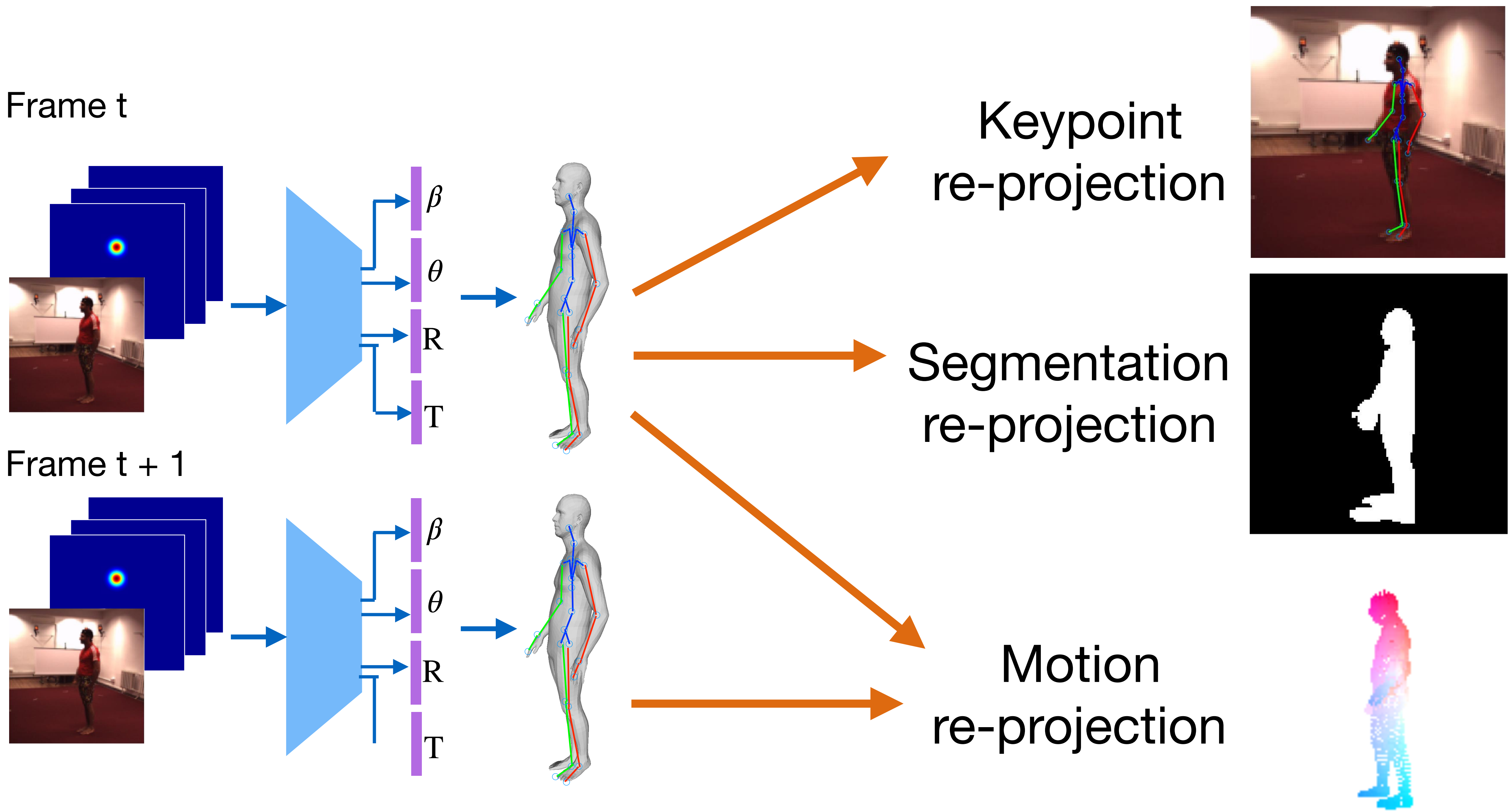
Self-supervised reprojection losses



Self-supervised reprojection losses

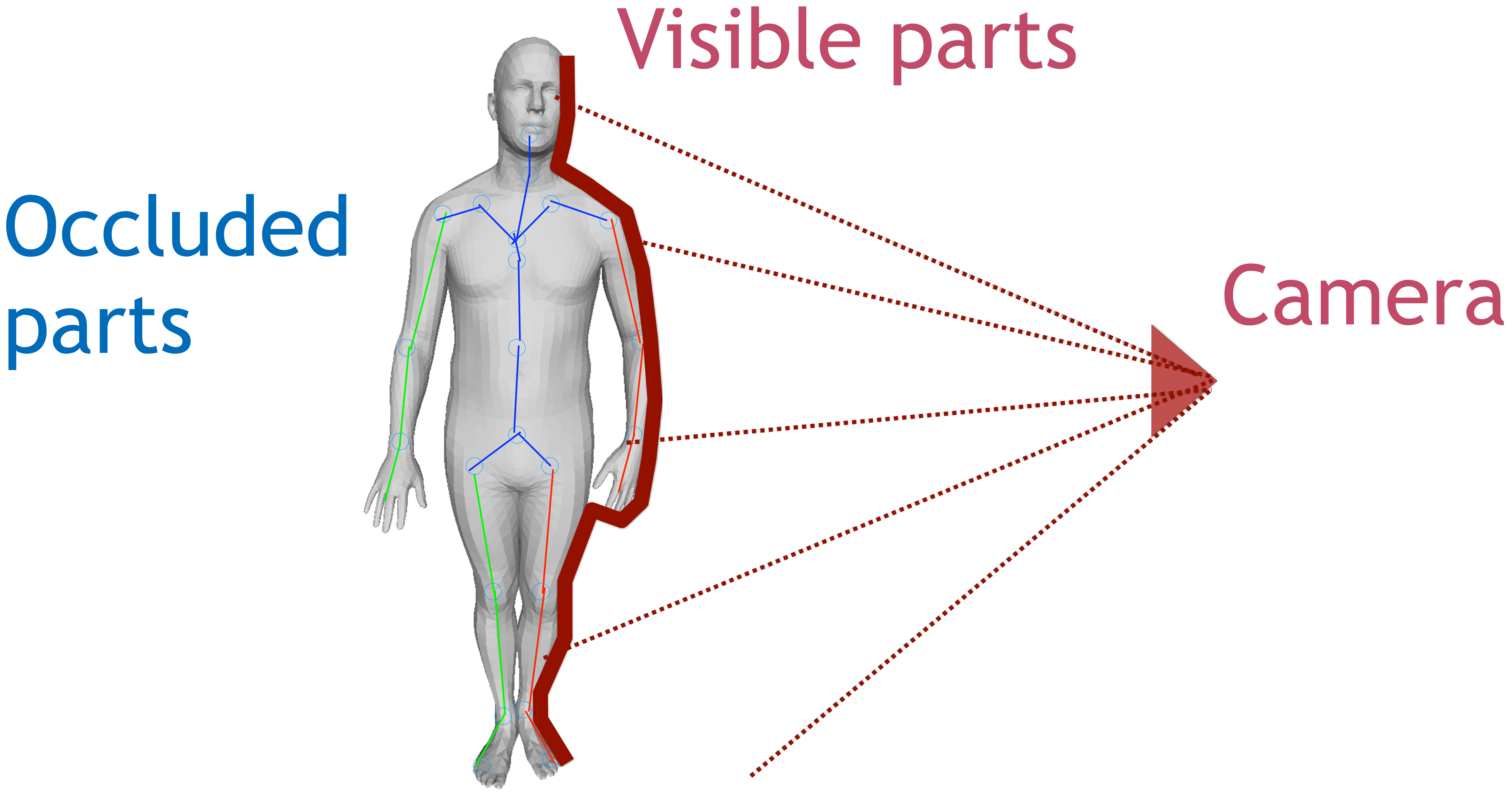


Self-supervised reprojection losses



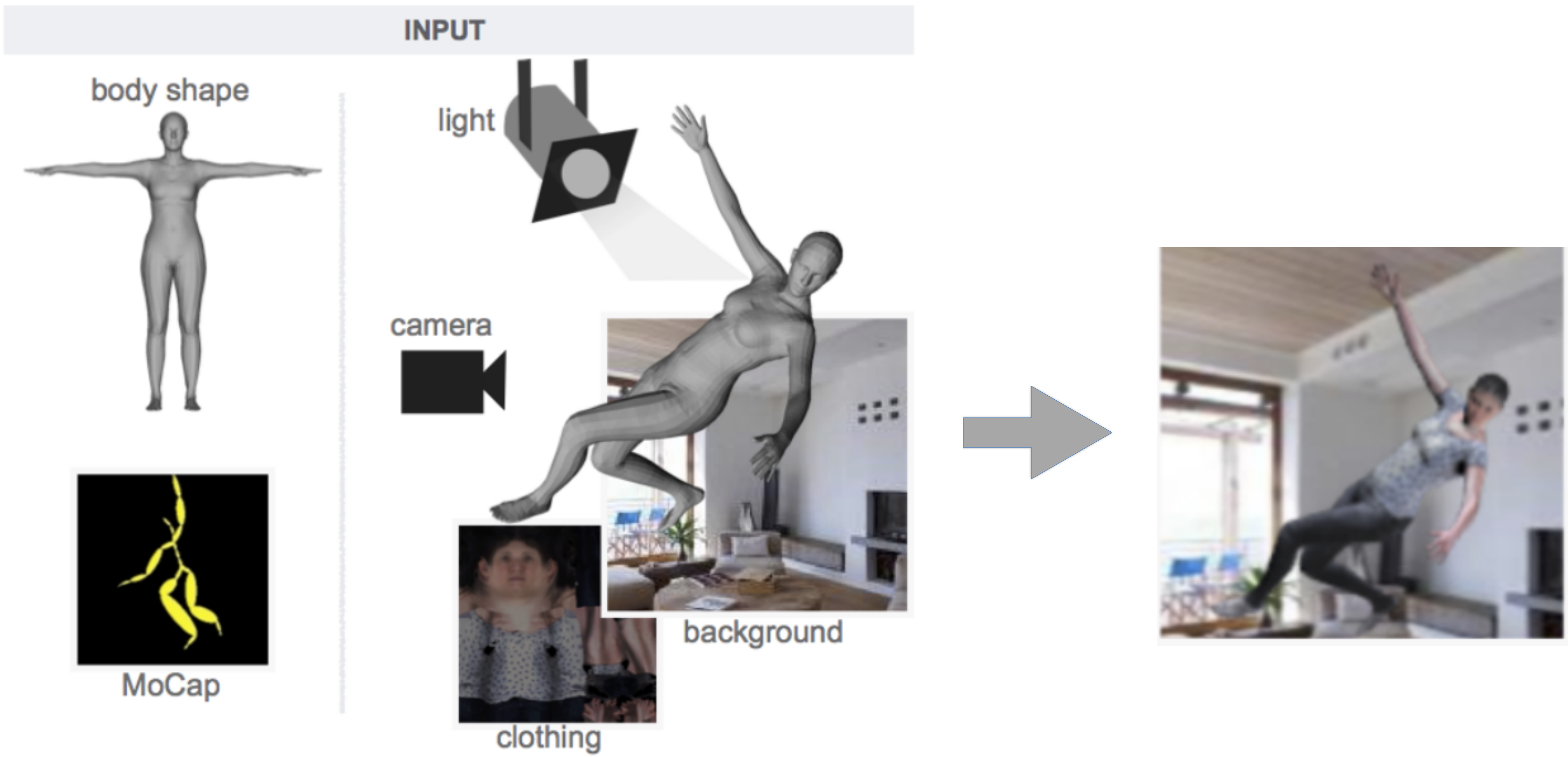
Flownet 2.0: Evolution of optical flow estimation with deep networks. Ilg at al., 2016

Visibility-aware reprojection



Supervised training

Synthetic data: SURREAL dataset



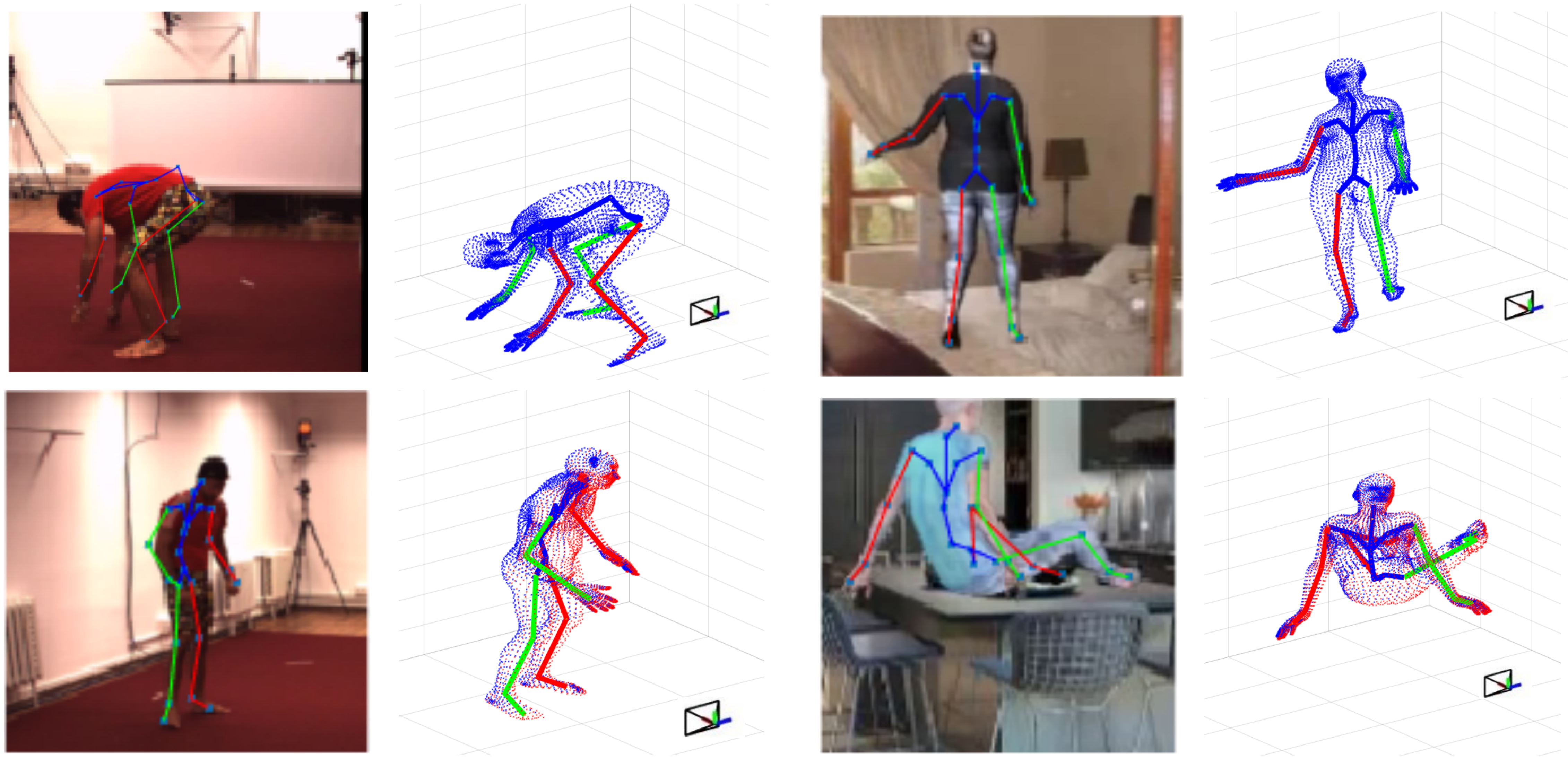
Learning from Synthetic Humans, Varol et al. CVPR 2017

Results

Per-Joint Error

	Per-Joint Error (mm)
optimization	562.4
supervised pretrained	125.6
Supervised+self-supervised	**98.4**

Results

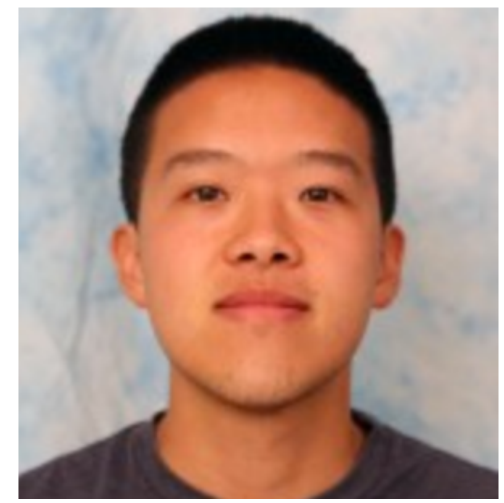


Thank you!

Fish Tung

Adam Harley

Hsiao-Wei Tung



William Seto

Ersin Yumer

- Adversarial Inverse Graphics Networks, Tung et al., ICCV 2017
- Self-supervised learning of motion capture, Tung et al. NIPS 2017