Unsupervised Human Pose Estimation on Depth Images

Thibault Blanc-Beyne[1,2]
thibault.blanc-beyne@irit.fr

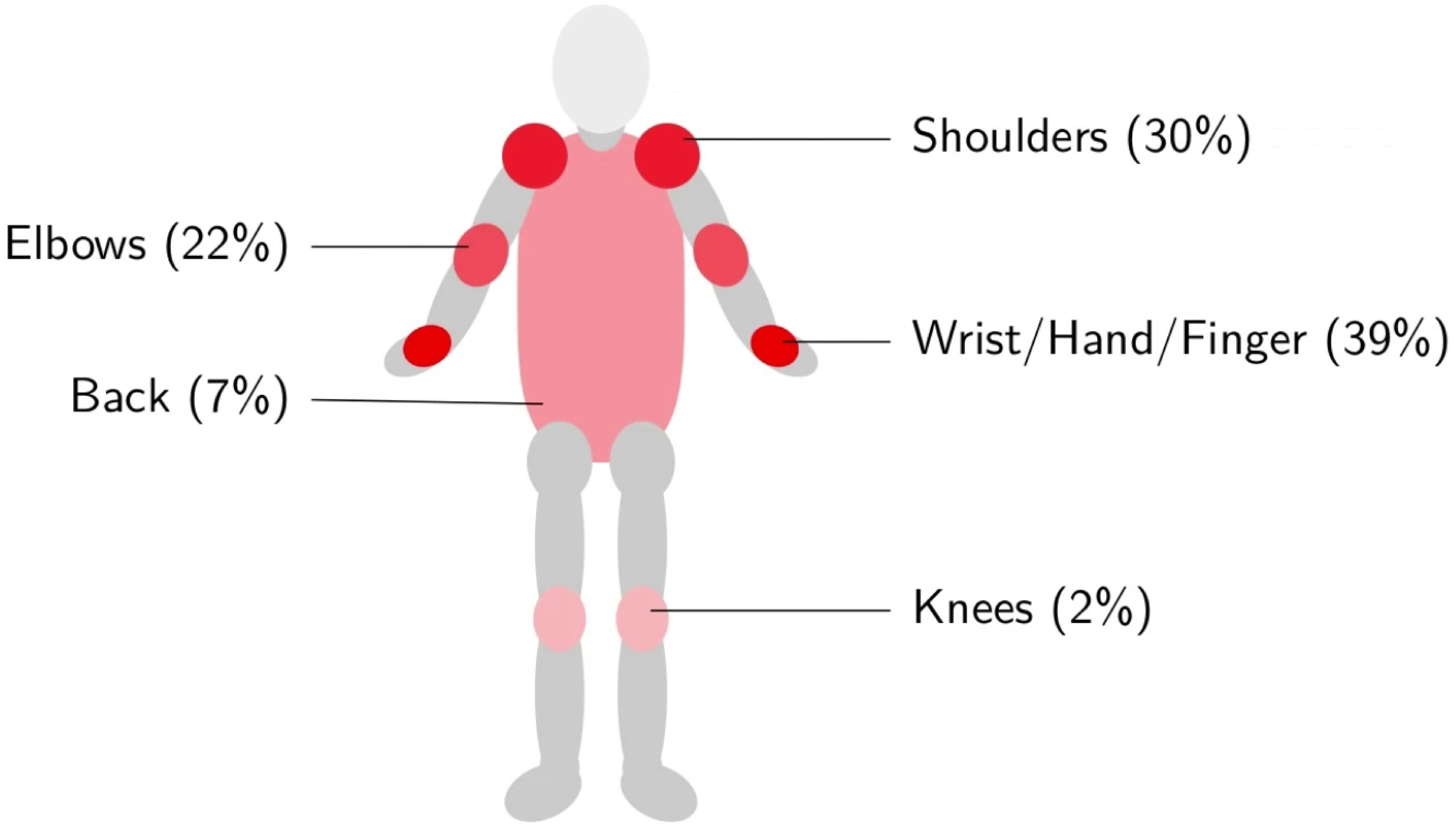
Axel Carlier[2]
axel.carlier@irit.fr

Sandrine Mouysset[2]
sandrine.mouysset@irit.fr

Vincent Charvillat[2]
vincent.charvillat@irit.fr

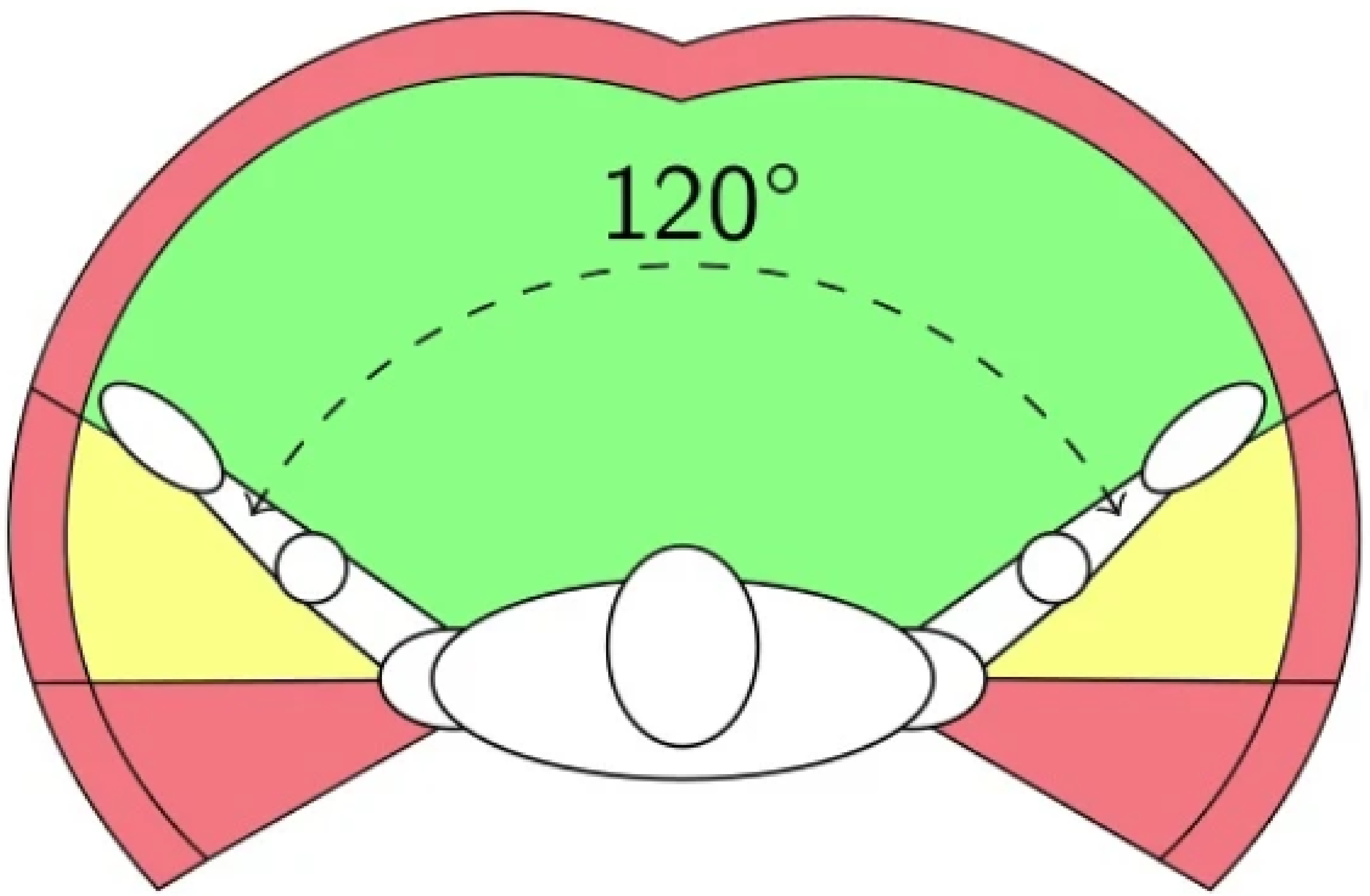
[1] Ebhys [2] University of Toulouse - IRIT

Musculoskeletal desorders (MSDs)



Distribution of MSDs on the human body

Indicators to measure

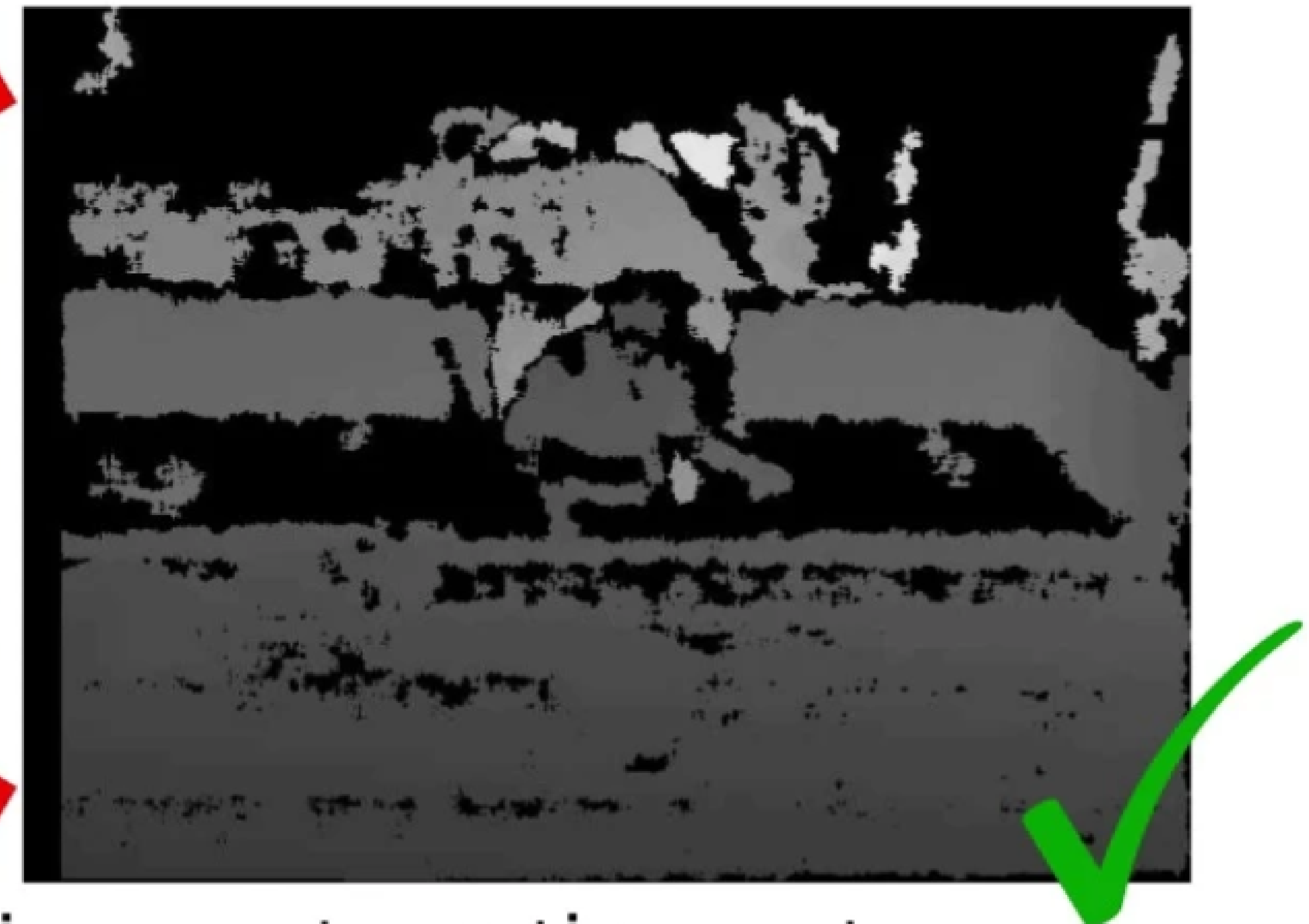


Stress on the operator's hands

Joint	Angulation
Neck	Flexion/Extension Lateral bending Rotation
Torso/Back	Flexion Lateral bending Rotation
Elbows	Flexion
Shoulders	Elevation Rotation

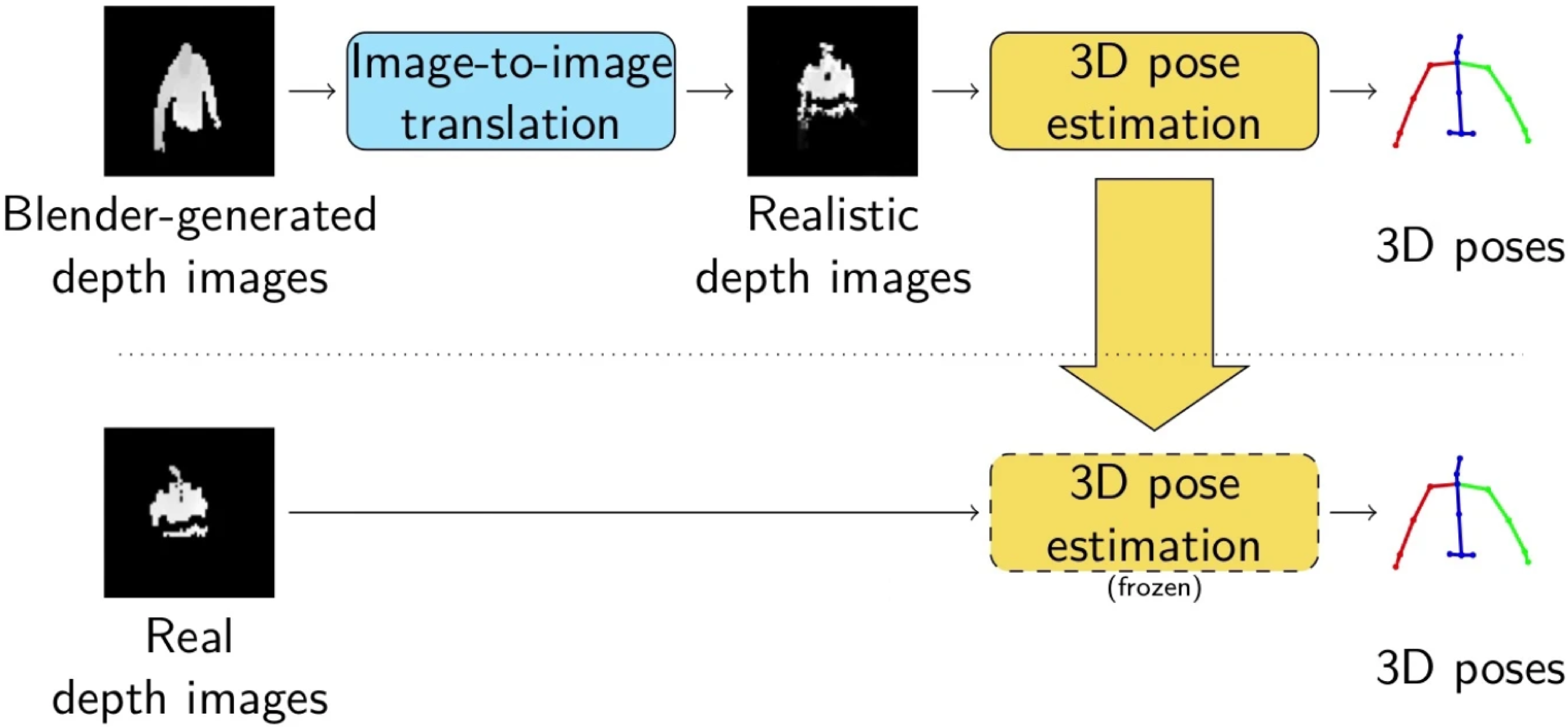
Stress on the operator's joints

Depth frames



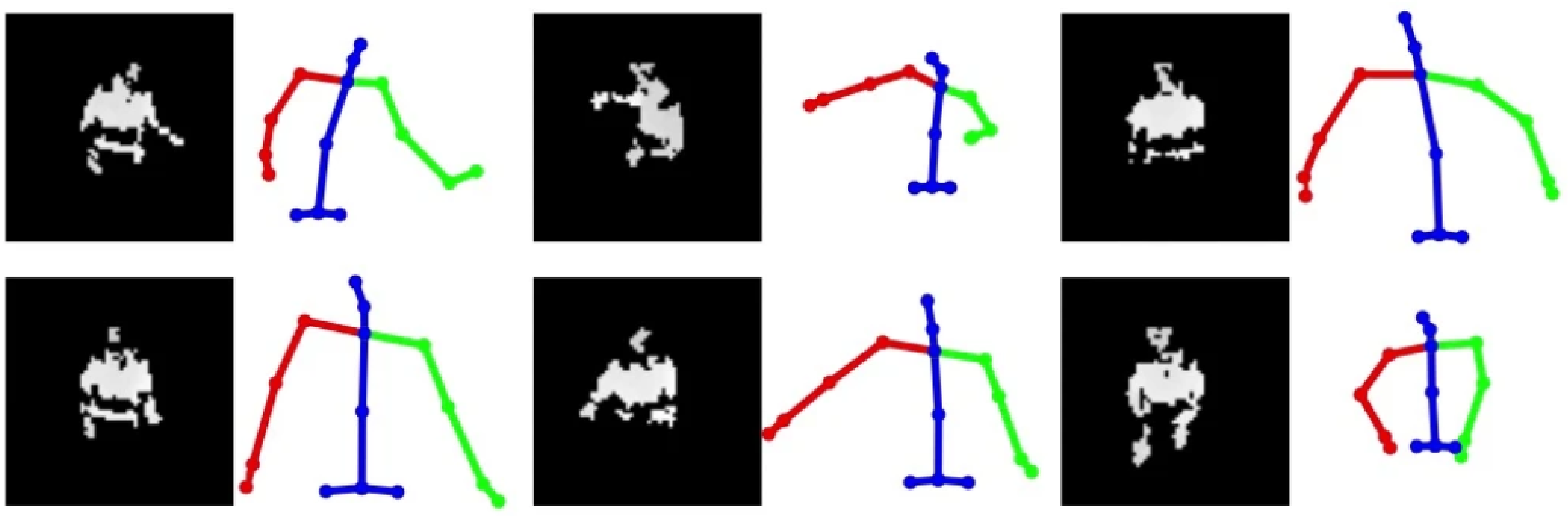
Images acquired in a waste sorting center,
where operators are especially prone to developing MSDs

Framework of our pose estimation process

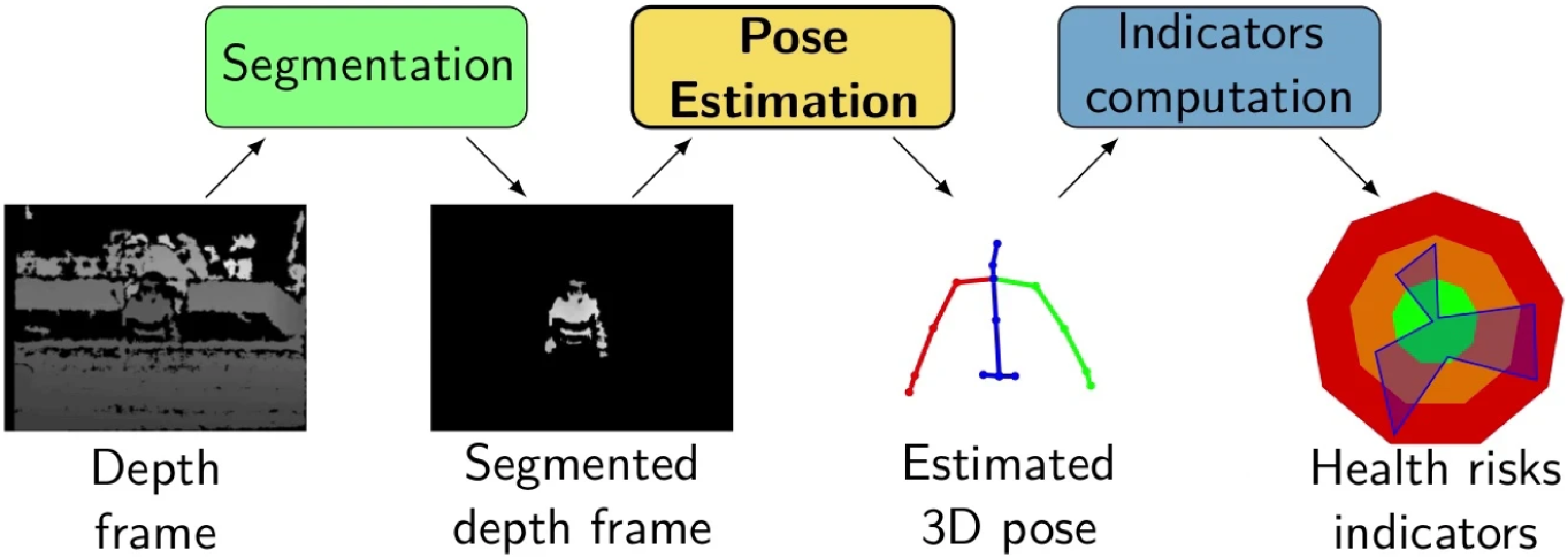


Numerical results and estimated posture from our proposed approach

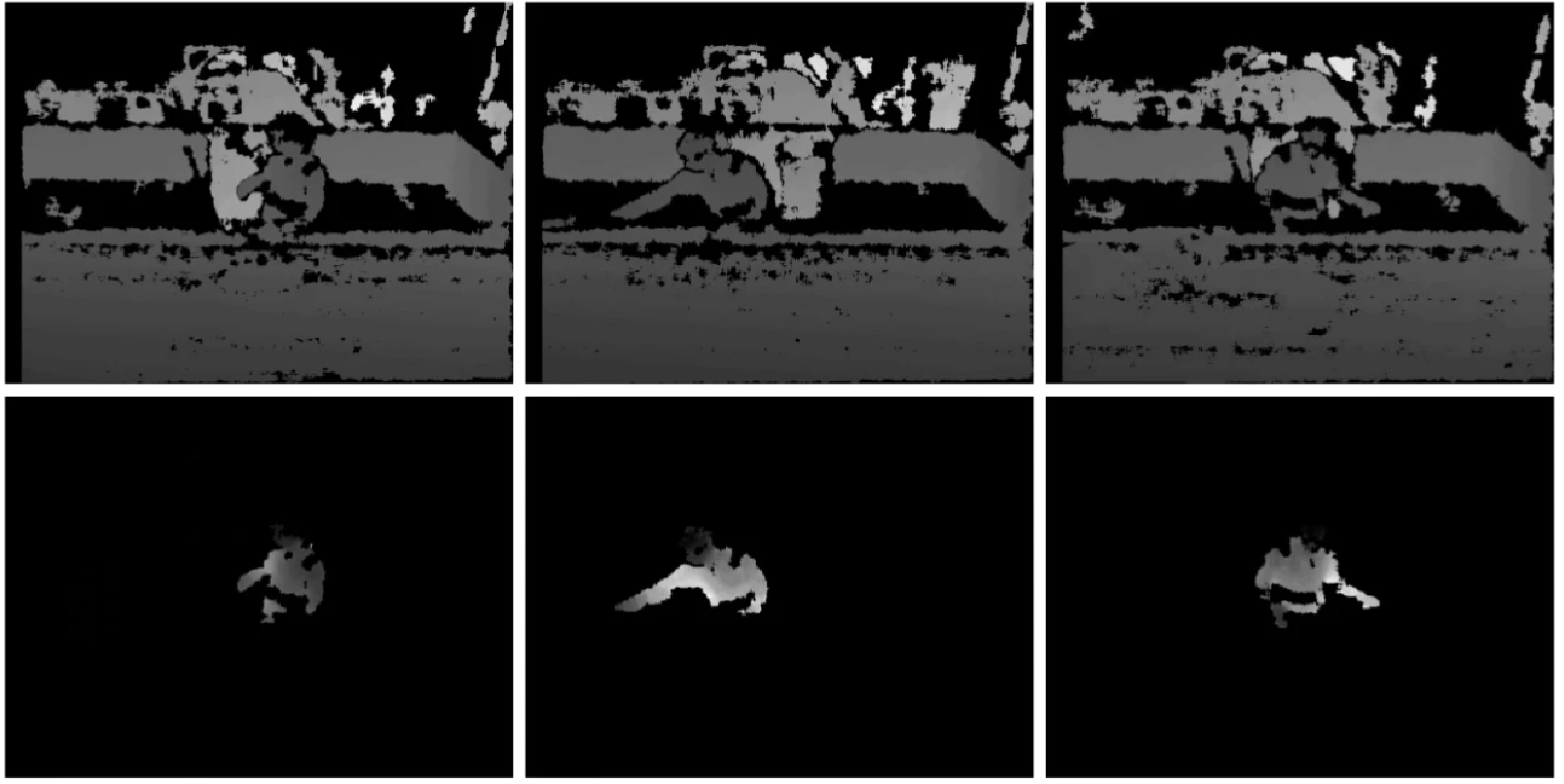
Kind of data	Training type	PCK 150mm	PCK 80mm	MPJPE 3D	PAMPJPE 3D
Noisy & Degraded	Supervised	100.0	100.0	21.4	18.8
	CycleGAN	**93.3**	**46.0**	**84.3**	**65.9**
	Unsupervised	55.0	0.2	148.0	87.6



Pipeline of our whole process

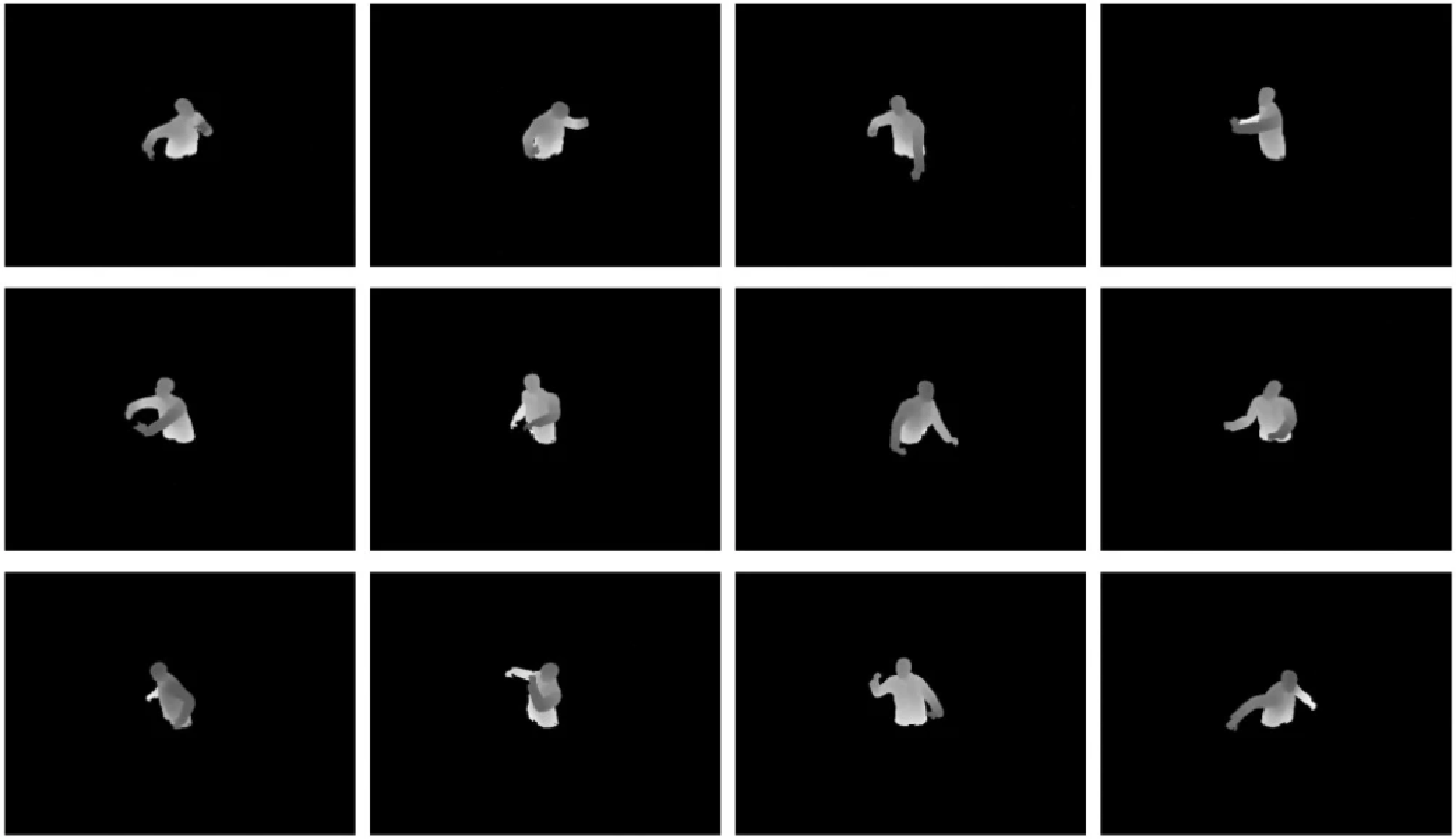


Our data



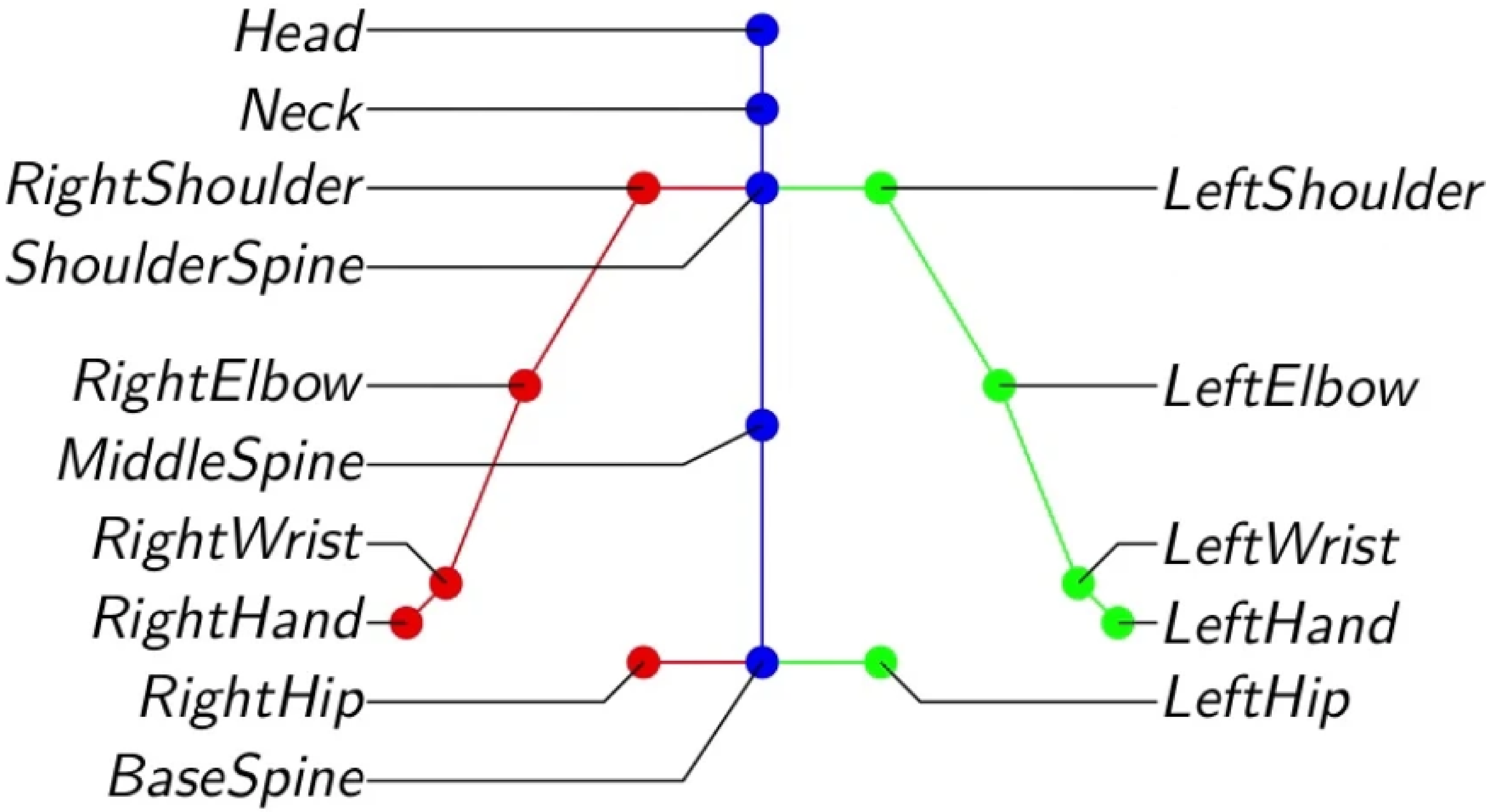
Raw depth frames (top) and segmented depth frames (bottom)

Data generation I



Images generated using MakeHuman and Blender

Data generation II

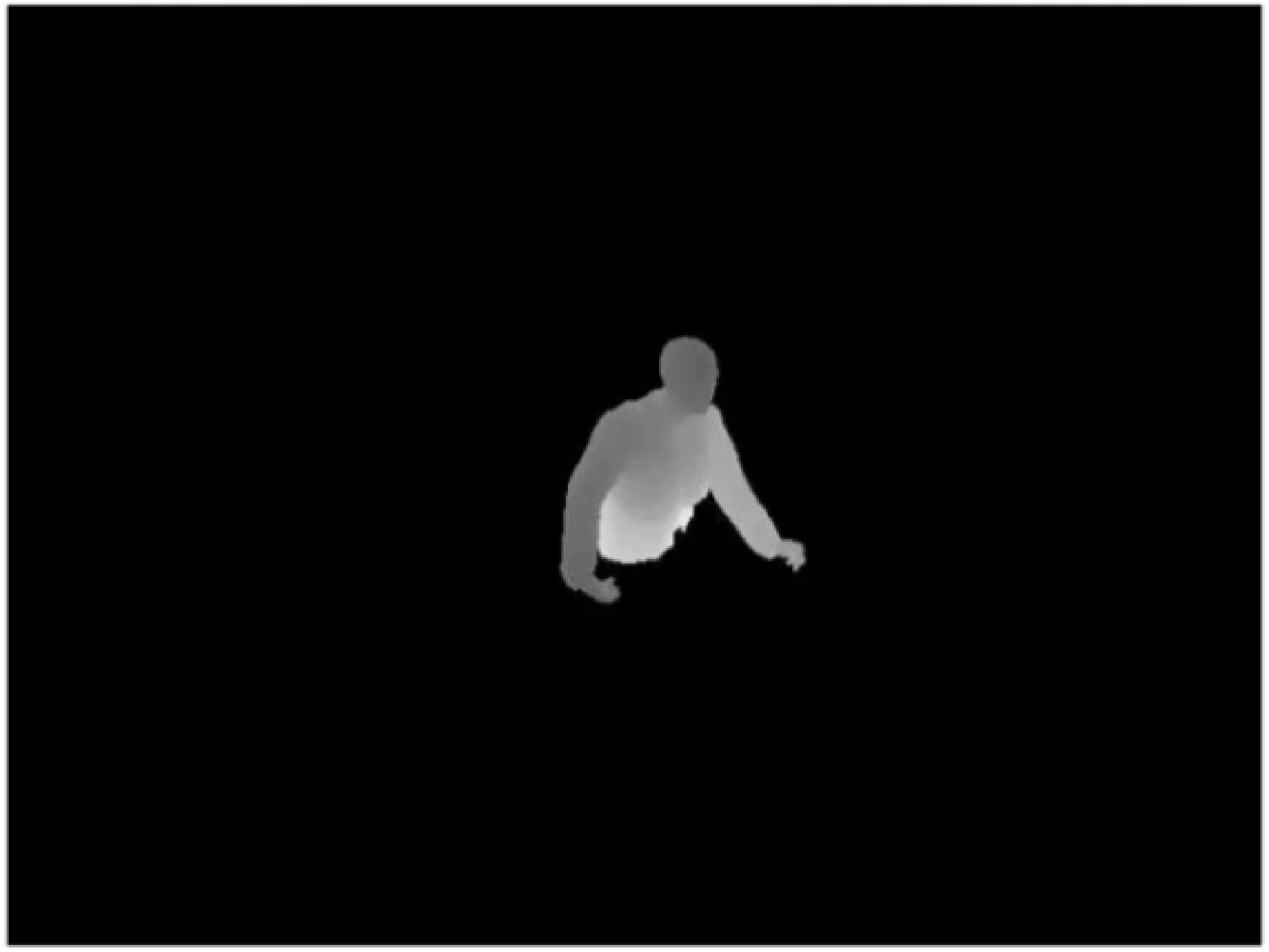


Base spine	0	0	0
Middle spine	-10	220	26
Shoulder spine	-35	459	24
Neck	-44	538	30
Head	-76	632	-12
Left shoulder	124	451	74
Left elbow	300	392	-82
Left wrist	363	446	-306
Left hand	315	477	-343
Right shoulder	-195	473	-46
Right elbow	-382	416	-189
Right wrist	-342	485	-413
Right hand	-327	533	-459
Left hip	100	-8	17
Right hip	-102	-8	-1

Joints for our pose estimation

Example of annotation

Blender-generated data against real data



$\neq$

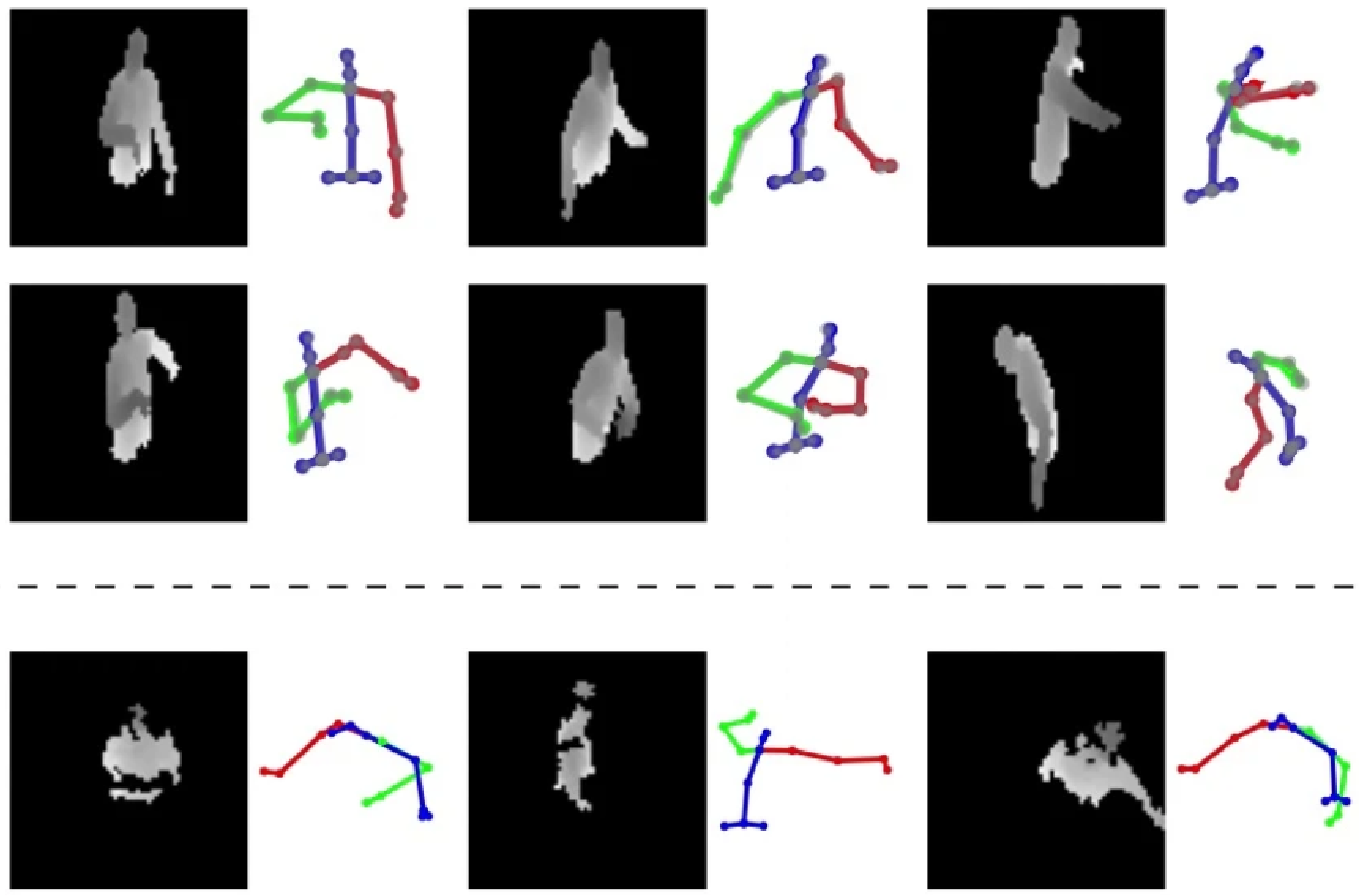
Blender-generated depth image

Real depth image

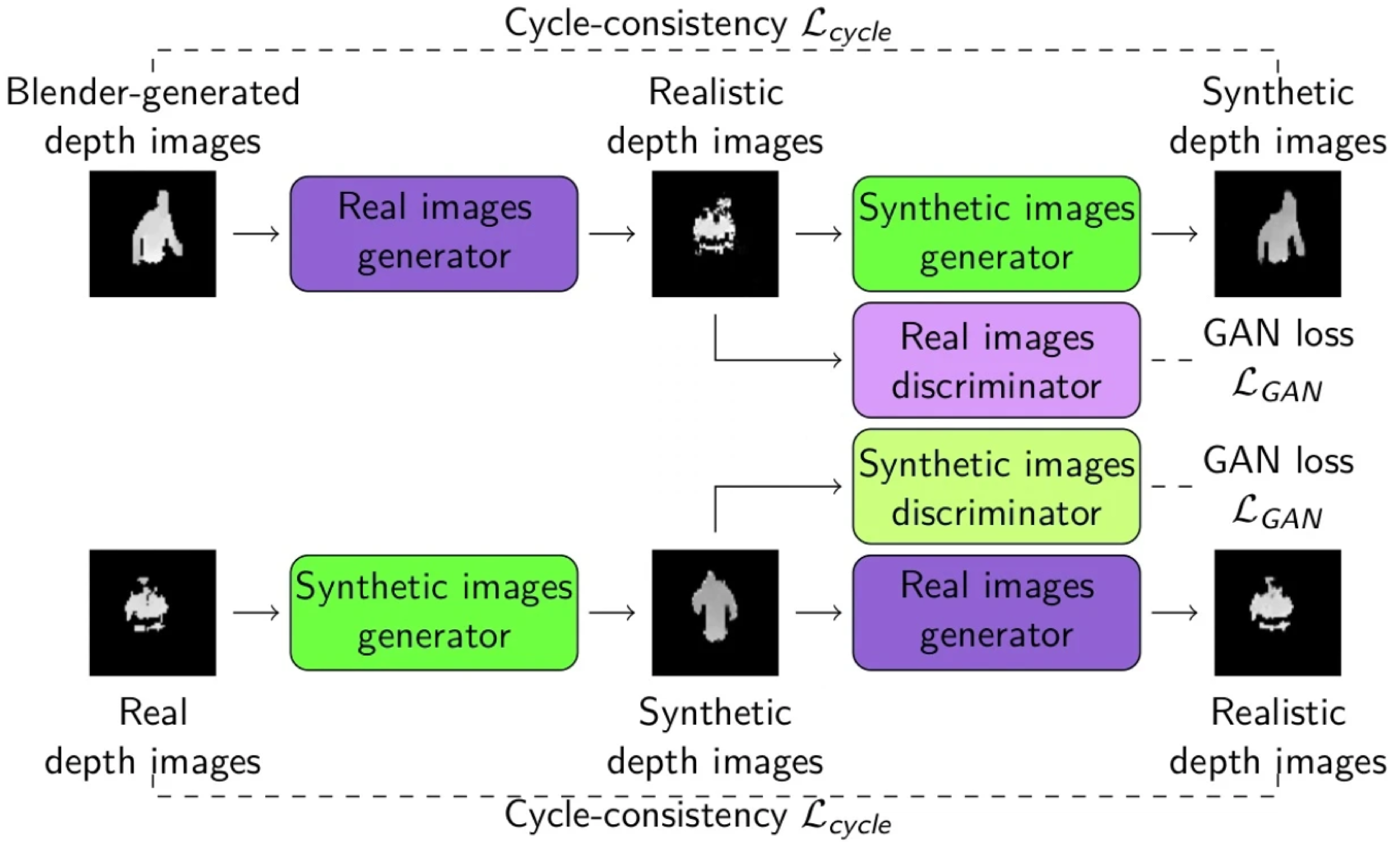
Pose estimator trained on blender-generated images

Blender generated data

Real data

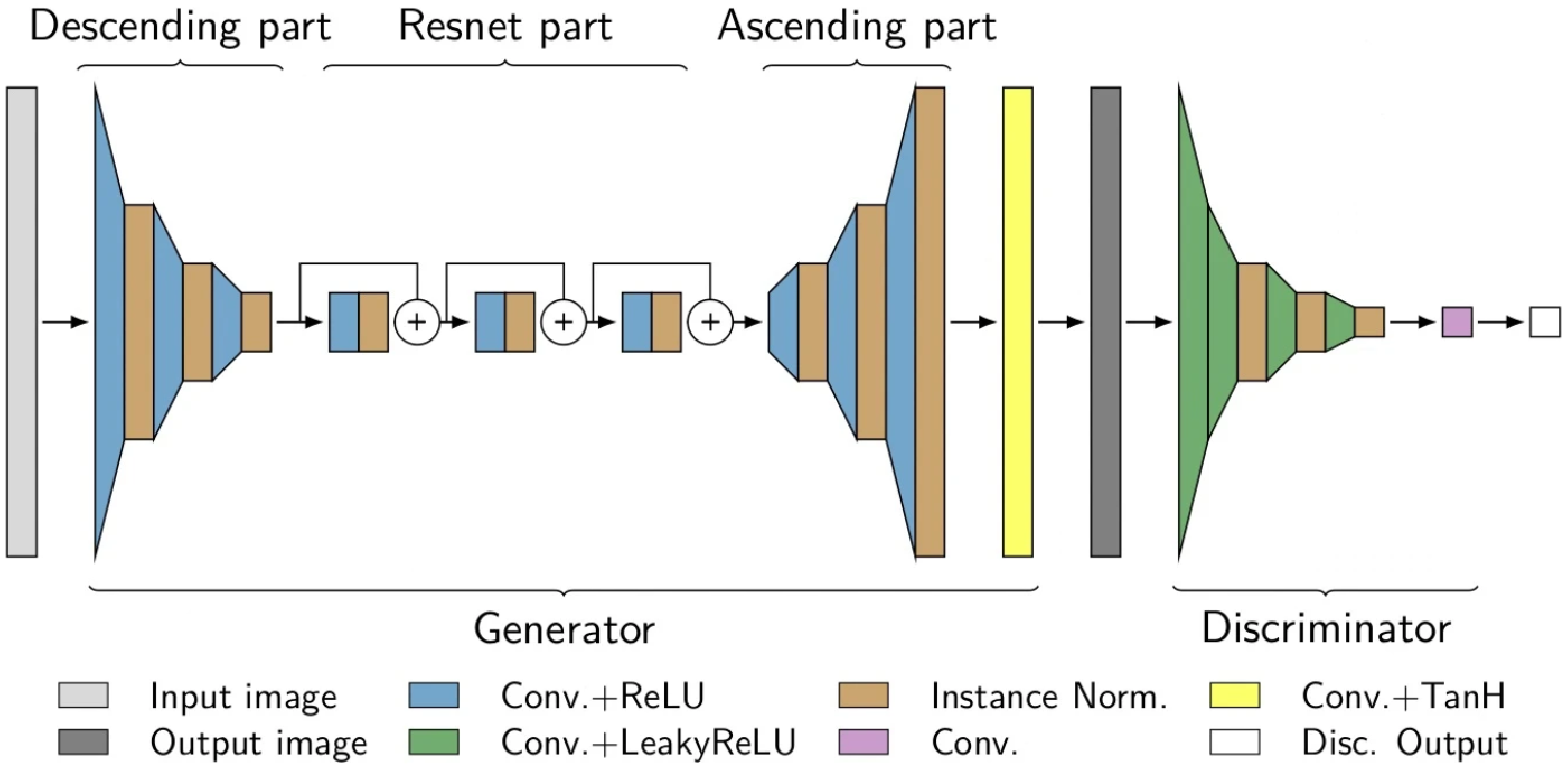


Model for image-to-image translation



CycleGAN model for image-to-image translation

Architecture of the CycleGAN model

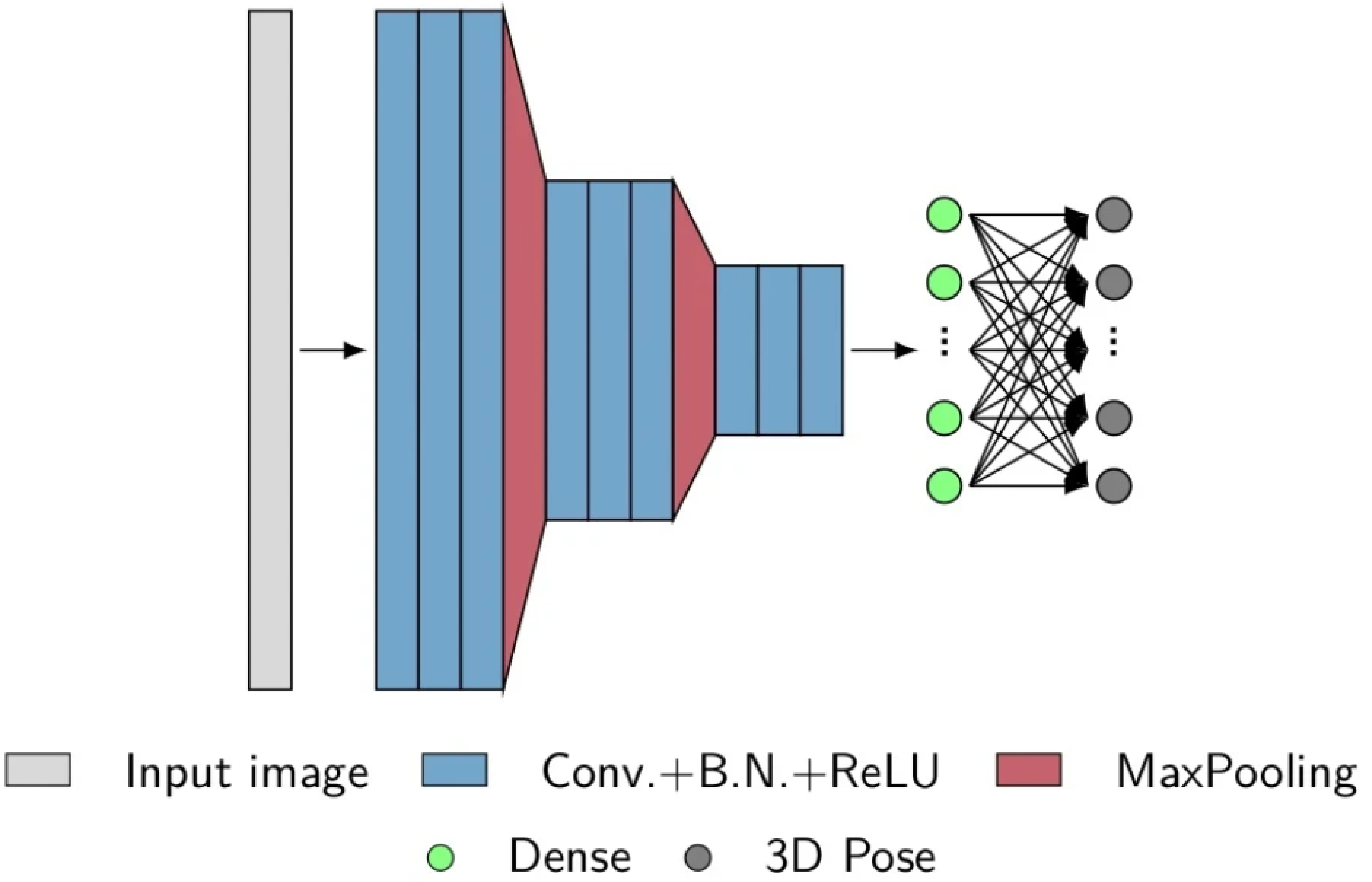


Results of our image-to-image translation model

Blender generated to realistic

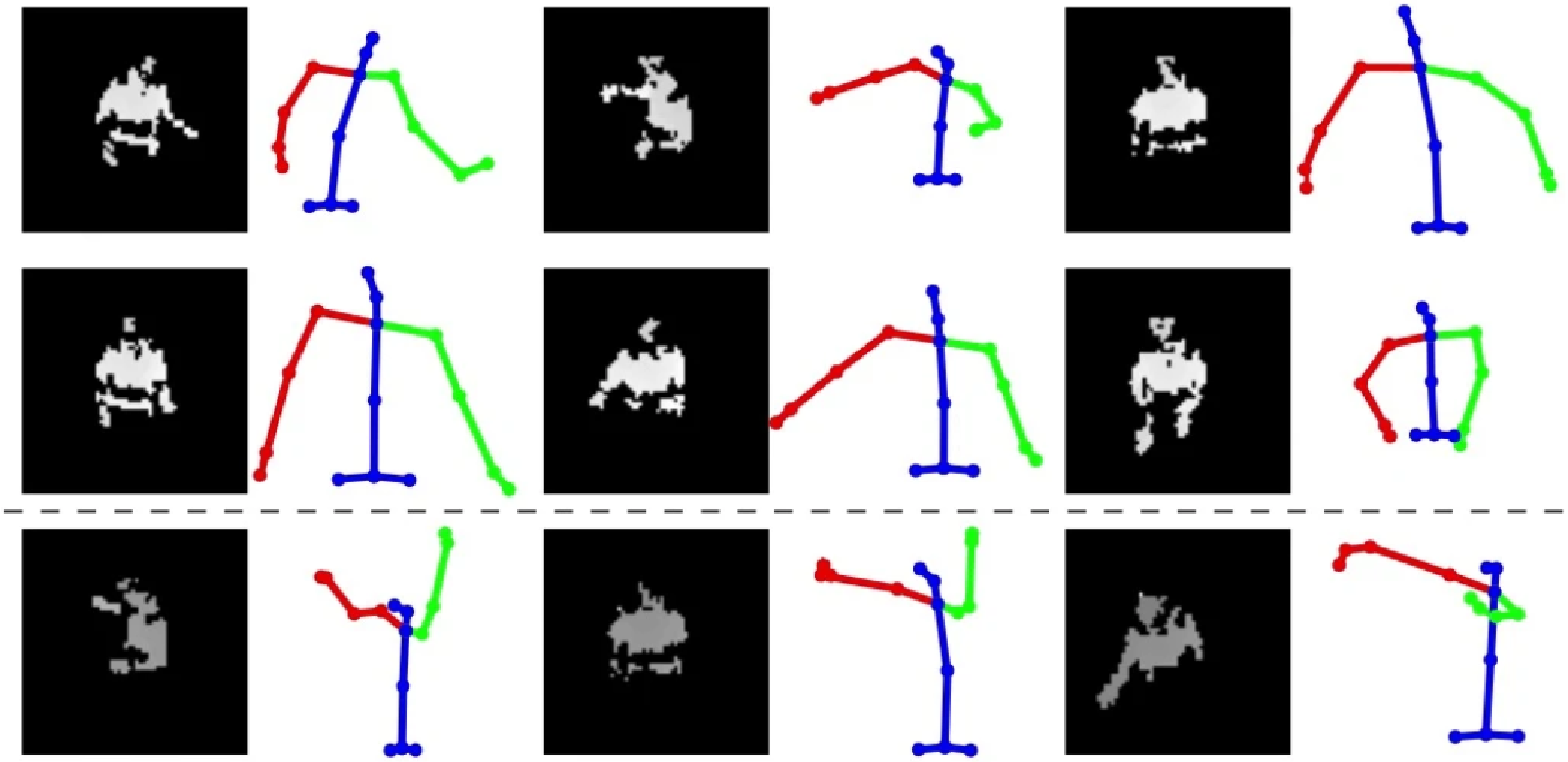
Real to synthetic

Training data



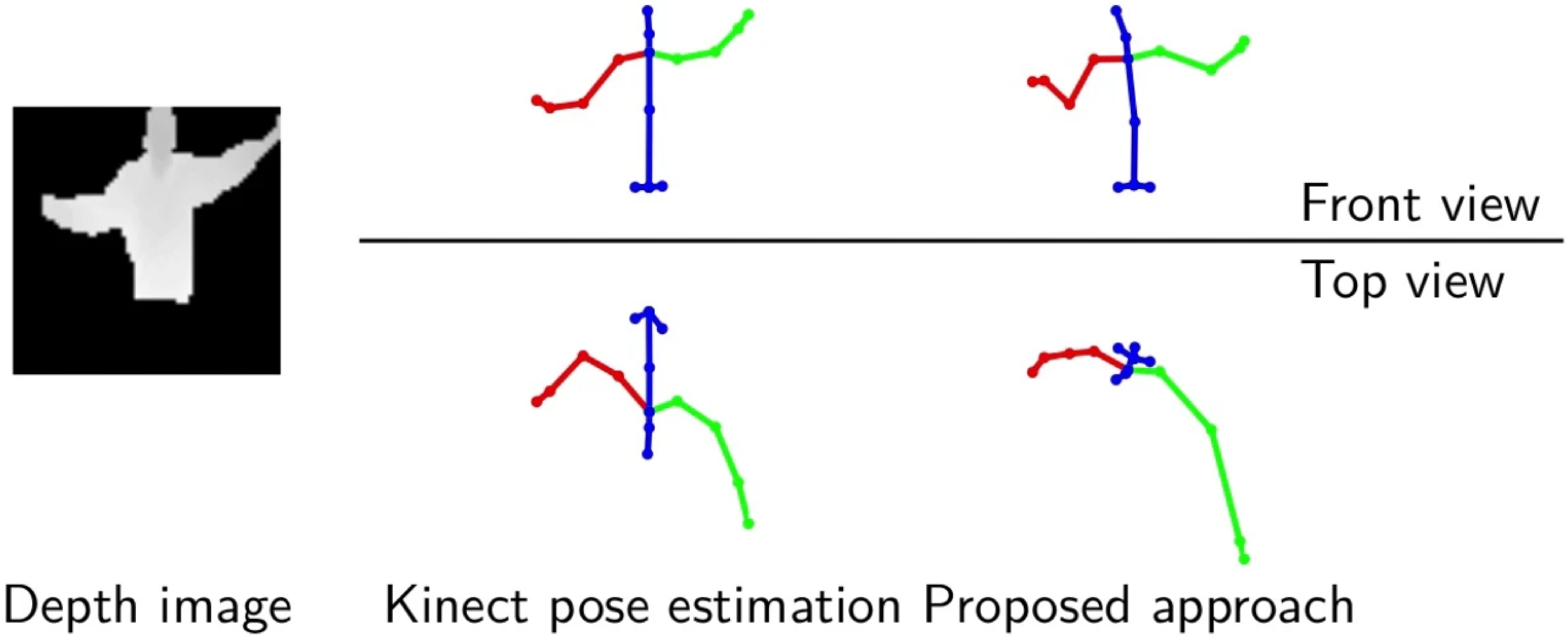
Architecture of the pose estimation network

Results of our pose estimation model



Estimated posture by our proposed approach

Comparison against the Kinect pose estimation



Training type	MPJPE 3D	PAMPJPE 3D
CycleGAN	**268.7**	**131.6**
Unsupervised	442.1	199.6

Usefulness of our approach

Example of noisy and degraded data

Kind of data	Training type	PCK 150mm	PCK 80mm	MPJPE 3D	PAMPJPE 3D
Synthetic	Supervised	100.0	100.0	10.9	9.9
Noisy & Degraded	Supervised	100.0	100.0	21.4	18.8
	CycleGAN	**93.3**	**46.0**	**84.3**	**65.9**
	Unsupervised	55.0	0.2	148.0	87.6

Numerical results on the noisy and degraded dataset

Conclusion

Our use case

- Preventing MSDs in waste sorting centers
- Computing indicators based on the 3D pose of the operator
- Working on depth frames

Our problem

- Lack of annotated data to train a pose estimator

Our proposed solution

- Use blender-generated depth frames
- Render realistic depth images using image-to-image translation
- Train the pose estimator on the realistic depth frames