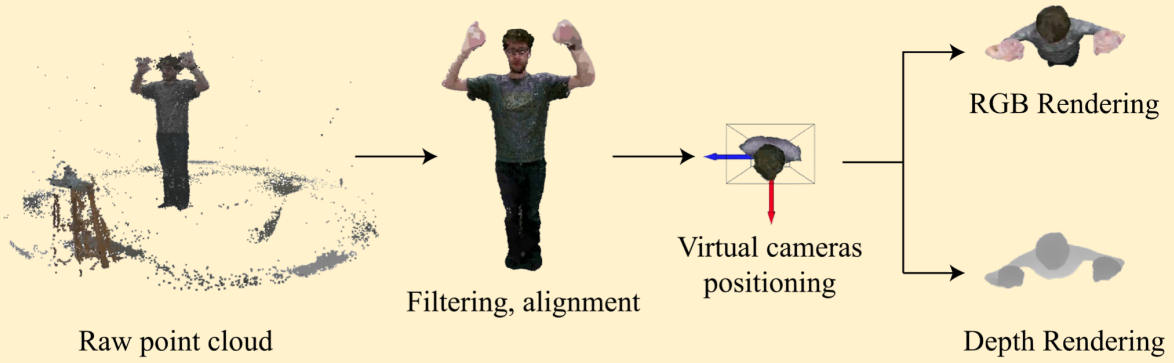
PanopTOP: a framework for generating viewpoint-invariant human pose estimation datasets

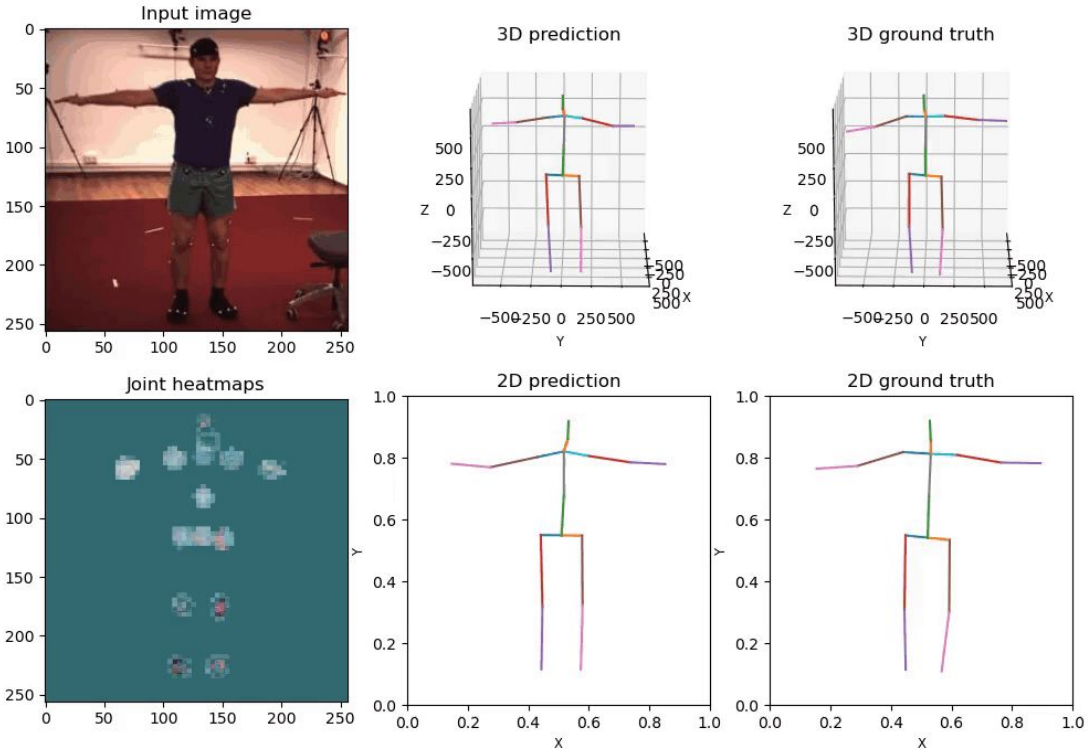
Nicola Garau, Giulia Martinelli, Niccolò Bisagno, Piotr Bródka, Nicola Conci
University of Trento, Italy



2021 ICCV OCTOBER 11-17 VIRTUAL

UNIVERSITY OF TRENTO

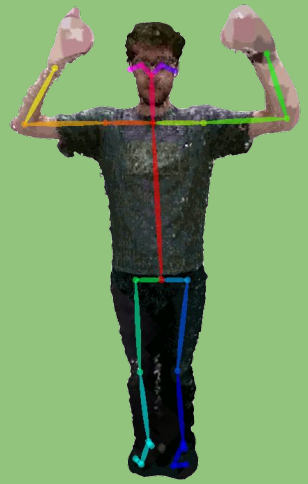
Human Pose Estimation 2D/3D



Issues and challenges

Viewpoint generalization

Front view



Top view

OpenPose

MaskRCNN

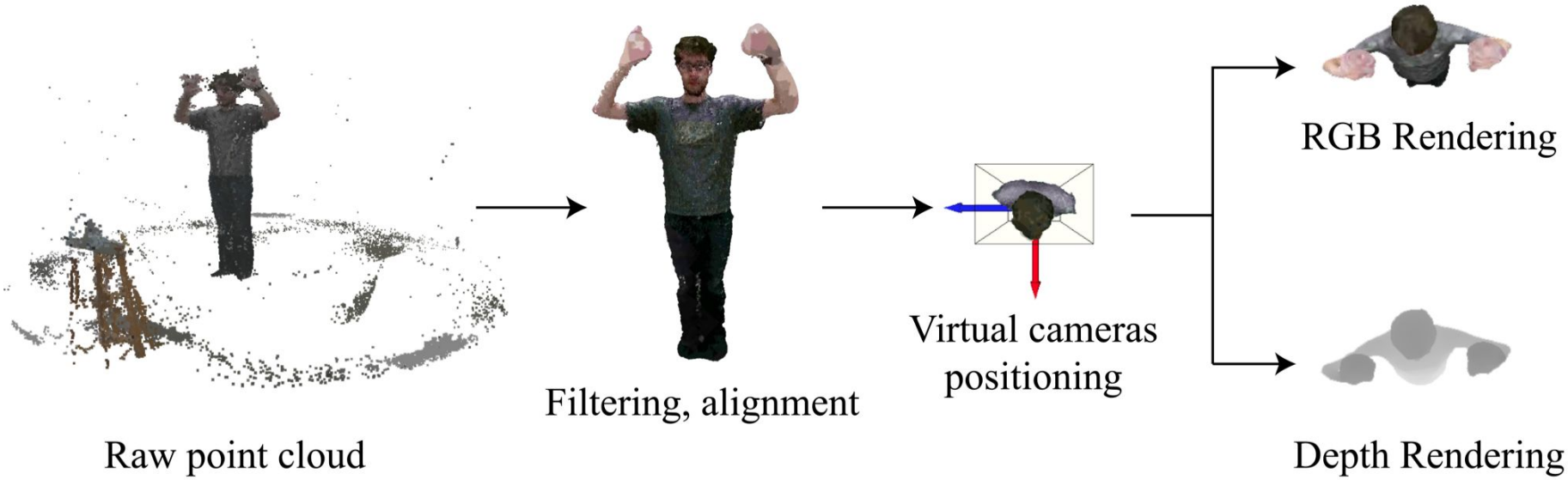
HMR

Lack of suitable dataset

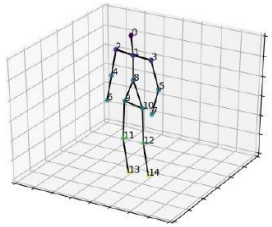
Dataset	RGB	Depth	Top-view	Multi-View	2D Pose GT	3D Pose GT	Camera parameters
PanopTOP31K	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
ITOP	N	**Y**	**Y**	**Y**	N	**Y**	**Y**
EVAL	N	**Y**	N	N	N	**Y**	N
TVPR	**Y**	**Y**	**Y**	N	N	N	N
TVPR 2	**Y**	**Y**	**Y**	N	N	N	N
K2HPD	N	**Y**	N	N	N	**Y**	N
UBC3V	N	**Y**	N	**Y**	N	**Y**	**Y**
Human3.6M	**Y**	N	N	**Y**	N	**Y**	**Y**

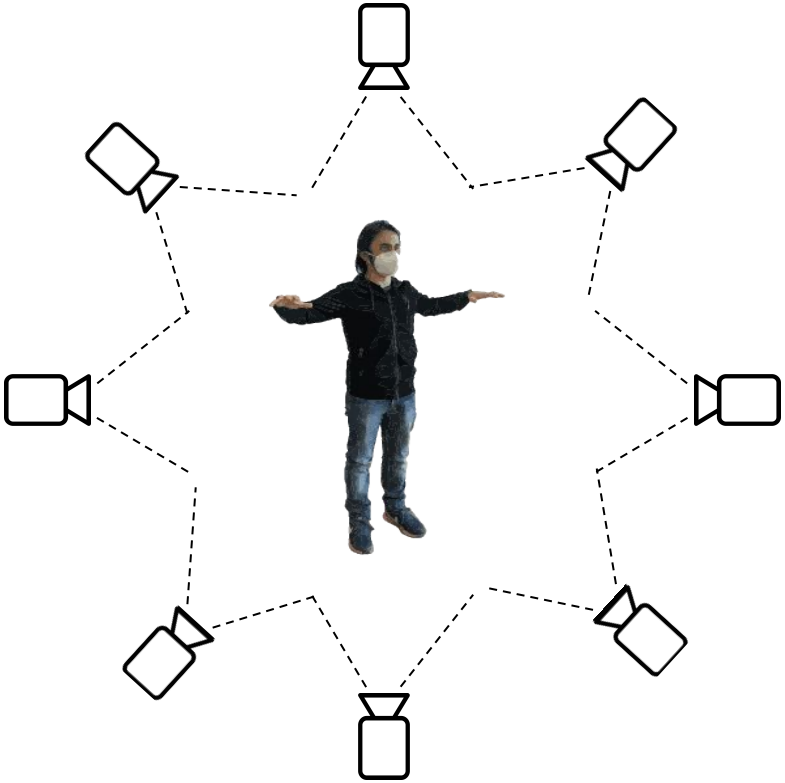
Proposed Solution

Proposed Solution



Advantages

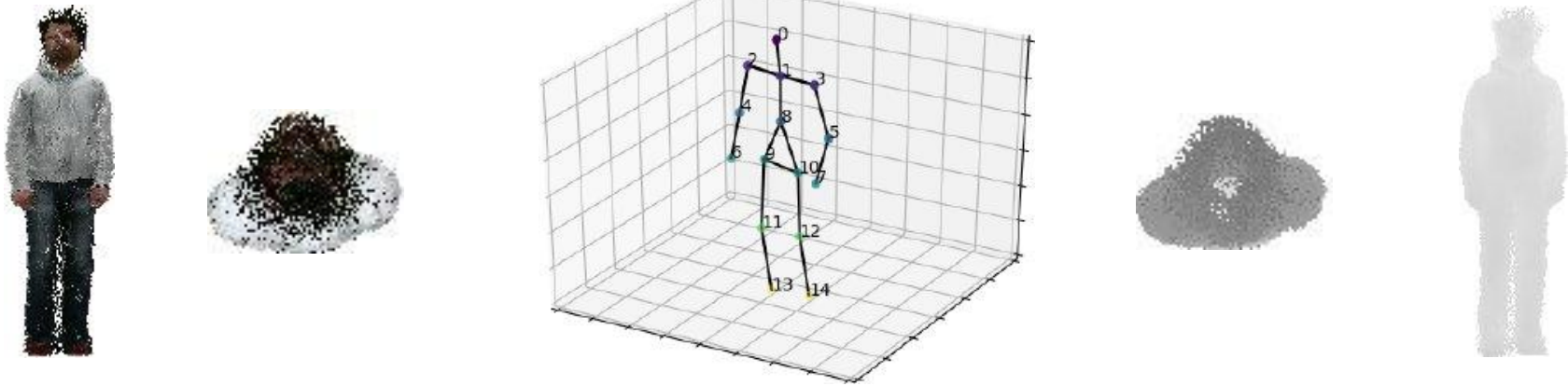




Experiments

PanopTOP31k

- 3 viewpoints: front, side, and top view
- 30K RGB images, 30K depth maps, 10K filtered point clouds, and 10K 3D meshes
- 23 different subjects
- 256x256 images
- 15 joints skeleton model

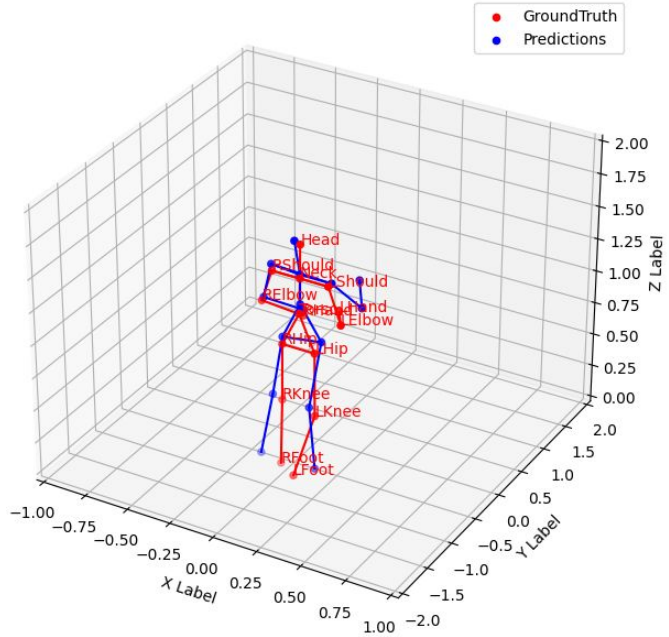


Benchmarking on depth: ITOP vs PanopTOP31k

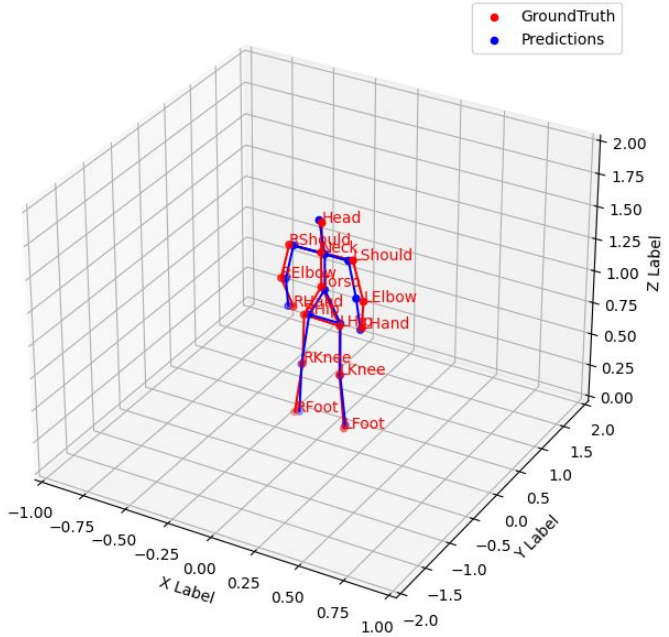
- Real vs semi-synthetic dataset
- Vanilla version of V2V network
- Same set, cross-validation, dataset transfer and combined experiments
- [I] = ITOP
- [P] = PanopTOP31k
- [Train][Validation][Test]

Benchmarking on depth

Same set



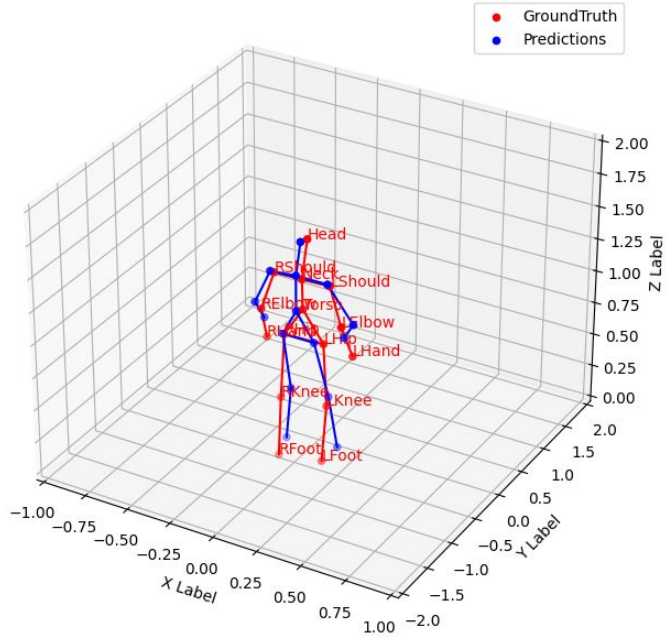
[I][I][I]



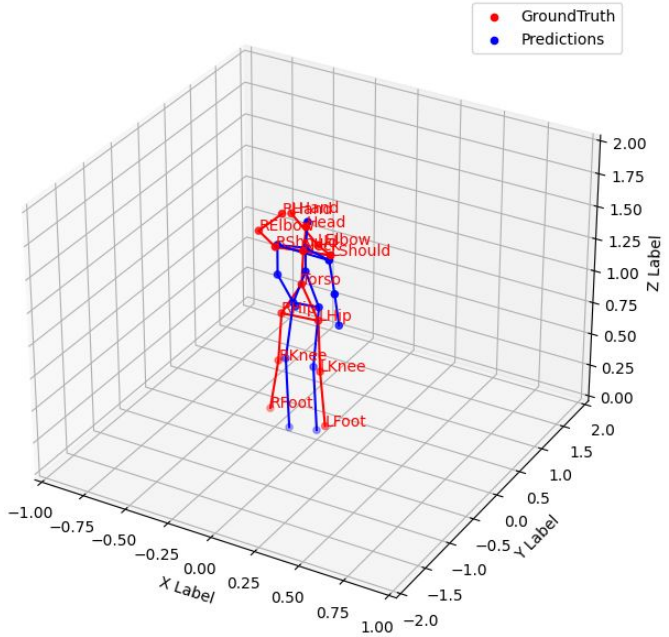
[P][P][P]

Benchmarking on depth

Dataset transfer



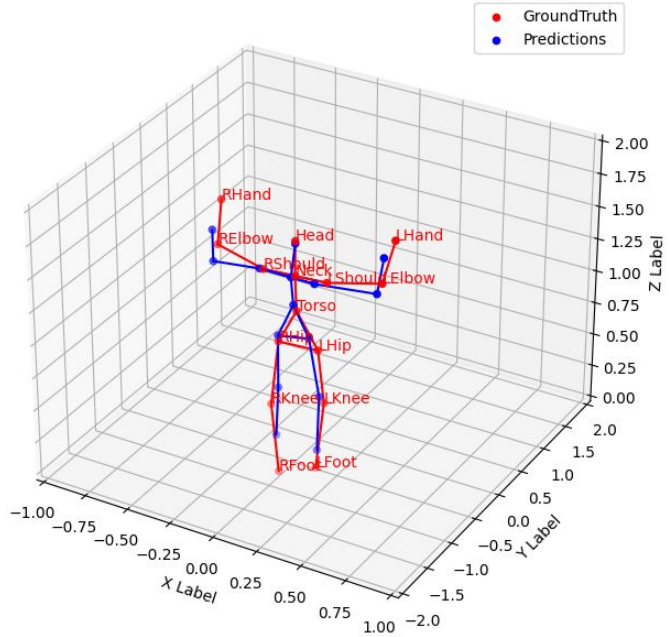
[P][P][I]



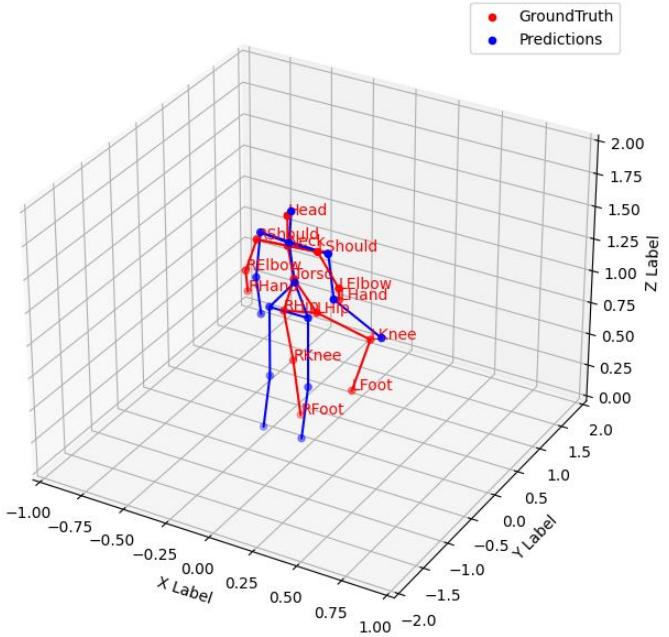
[I][I][P]

Benchmarking on depth

Cross-validation



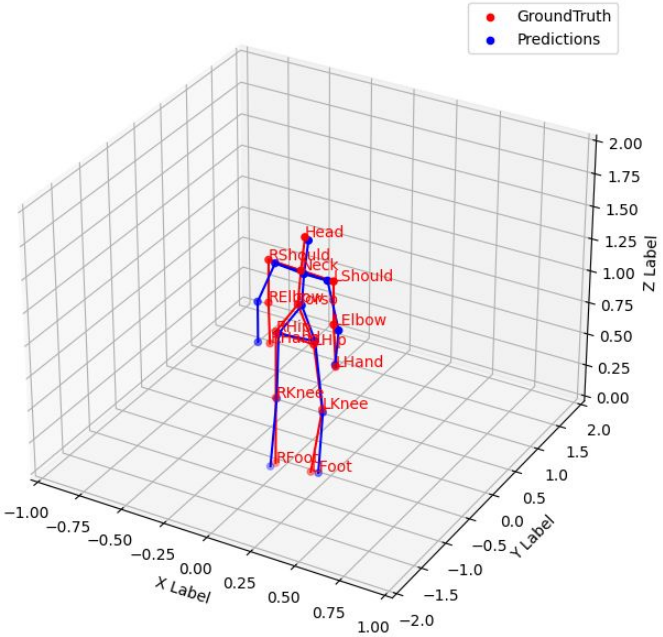
[P][I+P][I]



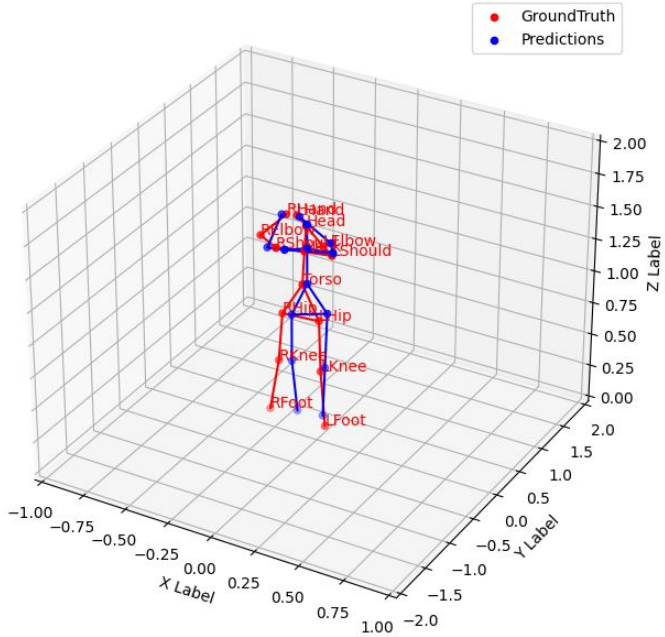
[I][I+P][P]

Benchmarking on depth

Combined



[I+P][I+P][I]



[I+P][I+P][P]

Benchmarking

	Experiment	Head	Neck	Shoulders	Elbows	Hands	Torso	Hips	Knees	Feet
(a)	**[I],[I],[I]**	99.50	99.60	99.05	**97.90**	**90.80**	***100.00***	98.55	95.20	87.15
(b)	**[I],[I],[P]**	96.60	97.90	93.80	76.10	63.60	97.80	89.90	84.60	46.50
(c)	**[I],[I+P],[P]**	97.20	98.10	95.45	77.15	59.10	98.00	90.25	70.20	35.80
(d)	**[I+P],[I+P],[P]**	**98.50**	**99.70**	**99.70**	***98.20***	***90.90***	**99.70**	**99.40**	95.80	***95.55***
(e)	**[P],[P],[P]**	**98.50**	**99.70**	**99.70**	97.80	90.85	99.60	99.35	***96.30***	95.45
(f)	**[P],[P],[I]**	99.50	99.50	98.10	93.90	61.45	99.30	94.85	75.45	26.80
(g)	**[P],[I+P],[I]**	99.60	99.80	97.95	94.00	66.60	99.50	94.45	83.55	59.20
(h)	**[I+P],[I+P],[I]**	***100.00***	***100.00***	***100.00***	97.80	90.35	***100.00***	***99.55***	***96.30***	**89.35**

Table 2: Percentages of correctly detected joints for the ITOP and PanopTOP31K datasets in our 8 conducted experiments. Each experiment is identified by a letter (**a-h**) and a data split **[train],[validation],[test]** (**P** = PanopTOP31K, **I** = ITOP). Each value represents the percentage of joints with L2 distance smaller than a threshold $T = 0.2m$ from the ground truth. The top scores for each joint regarding tests on the ITOP dataset are highlighted in **blue**, while the PanopTOP31K ones are highlighted in **green**. The top overall scores for each joint are in ***italic***.

Benchmarking

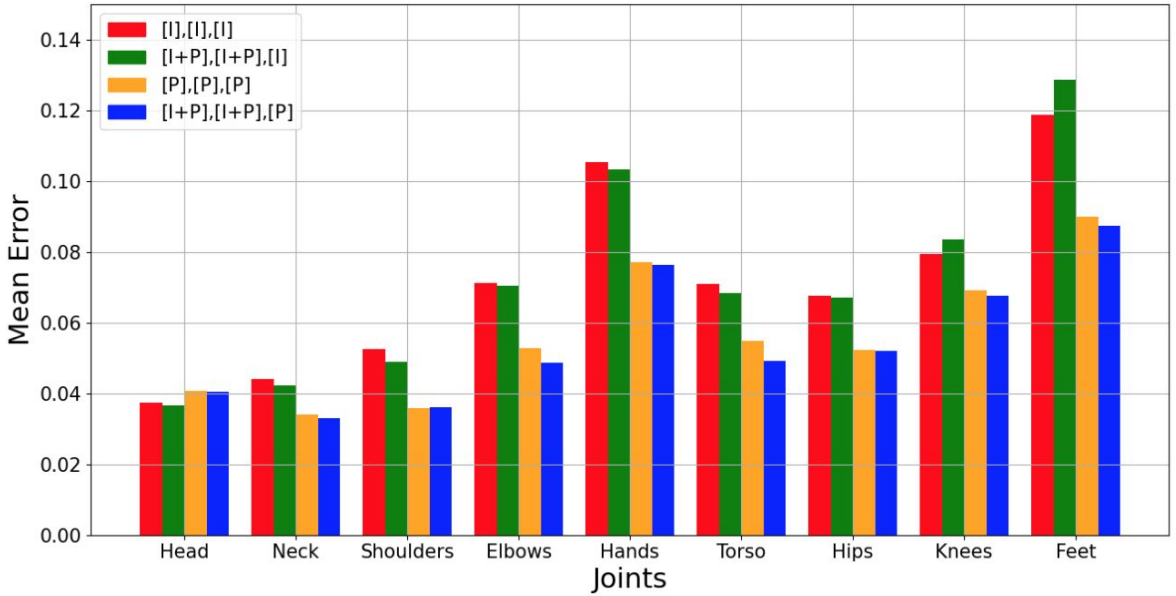
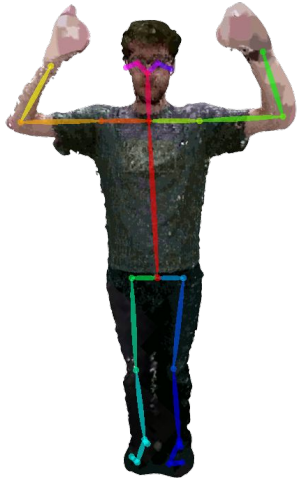


Figure 3: Mean per-joints errors in meters for ITOP and PanopTOP31K datasets, respectively, with (green, blue) and without (red, orange) training-wise augmentation. Red, green, yellow and blue bars correspond to experiments (a), (h), (e) and (d) respectively.

Results on RGB data

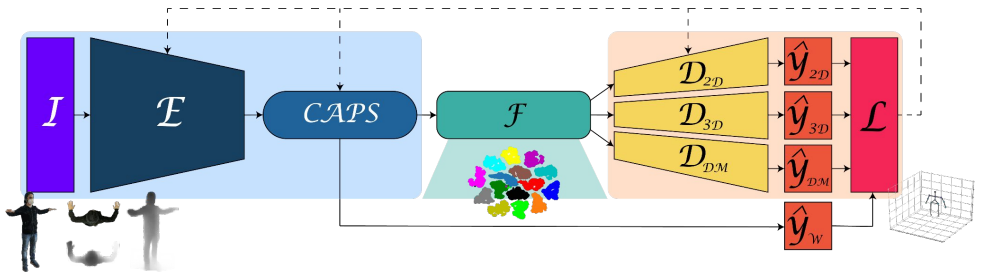


OpenPose

MaskRCNN

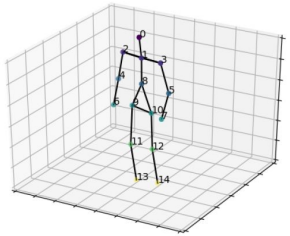
HMR

Enabling viewpoint equivariant approaches: DECA

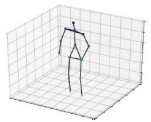


(a)

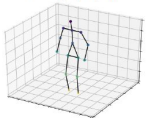
(b)



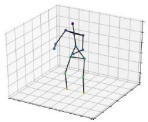
(c) GT



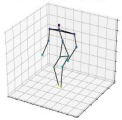
(d) {T}:{T}



(e) {F}:{F}



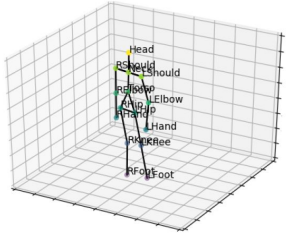
(f) {T}:{F}



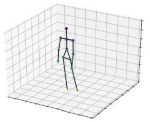
(g) {F}:{T}

(a)

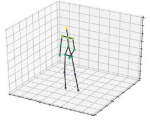
(b)



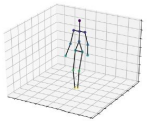
(c) GT



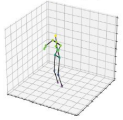
(d) {T}:{T}



(e) {F}:{F}



(f) {T}:{F}



(g) {F}:{T}

[5] N. Garau, N. Bisagno, P. Bródka, N. Conci, DECA: Deep viewpoint-Equivariant human pose estimation using Capsule Autoencoders (ICCV 2021- Oral)

Conclusions

We presented PanopTOP, a framework for generating viewpoint-invariant human pose estimation datasets

1. A new framework
2. PanopTOP31K dataset
3. Dataset benchmarking and validation

References

[5] N. Garau, N. Bisagno, P. Bródka, N. Conci, DECA: Deep viewpoint-Equivariant human pose estimation using Capsule Autoencoders (ICCV 2021 - Oral)