Synthetic Occlusion Augmentation with Volumetric Heatmaps for 3D Human Pose Estimation

István Sárándi[1], Timm Linder[2], Kai O. Arras[2], Bastian Leibe[1]

[1] Visual Computing Institute, RWTH Aachen University – Aachen, DE
[2] Robert Bosch GmbH, Corporate Research – Stuttgart, DE

September 8, 2018 – Munich, Germany

Updated: Sep 17, 2018

BOSCH
Invented for life

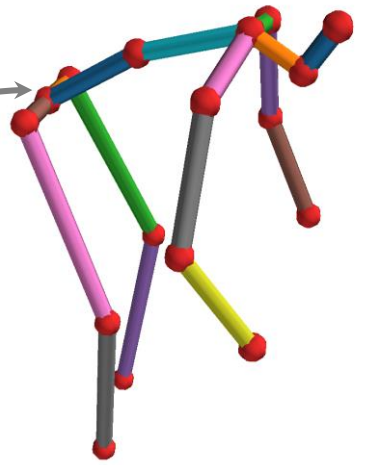
Visual Computing Institute

RWTH AACHEN UNIVERSITY

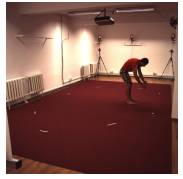
PoseTrack Challenge – 3D

```
[[   0.0    0.0    0.0]
 [ -96.9  -21.1  103.4]
 [ -43.0  456.4  150.3]
 [  42.6  902.6  249.0]
 [  96.0   22.1 -102.8]
 [  91.0  508.3 -100.2]
 [ 118.4  953.7   13.0]
 [   3.1 -262.6   13.2]
 [ -36.6 -502.0  -72.2]
 [ -96.7 -541.9 -162.8]
 [ -88.0 -651.5 -140.0]
 [  85.4 -439.5 -131.1]
 [ 278.0 -206.0 -121.3]
 [ 367.3   28.3 -184.0]
 [-139.0 -465.2   28.8]
 [-406.6 -322.4   42.9]
 [-414.1 -255.1 -202.1]]
```

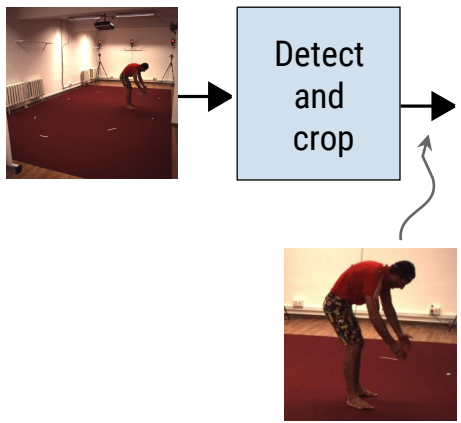
pelvis



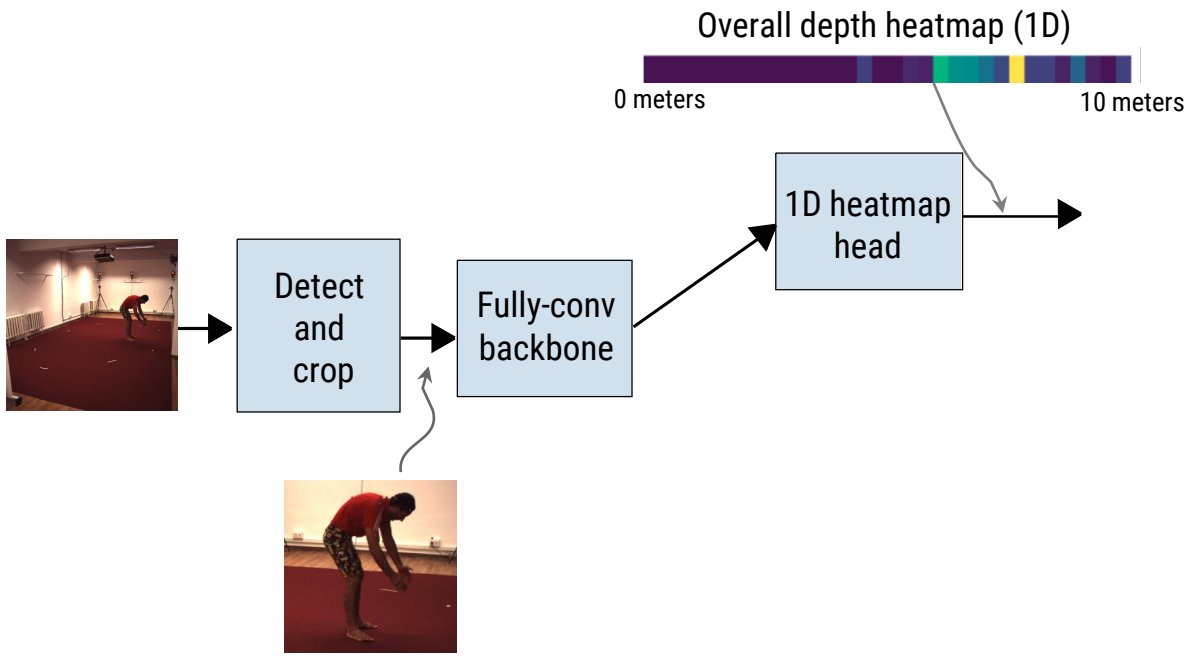
Approach



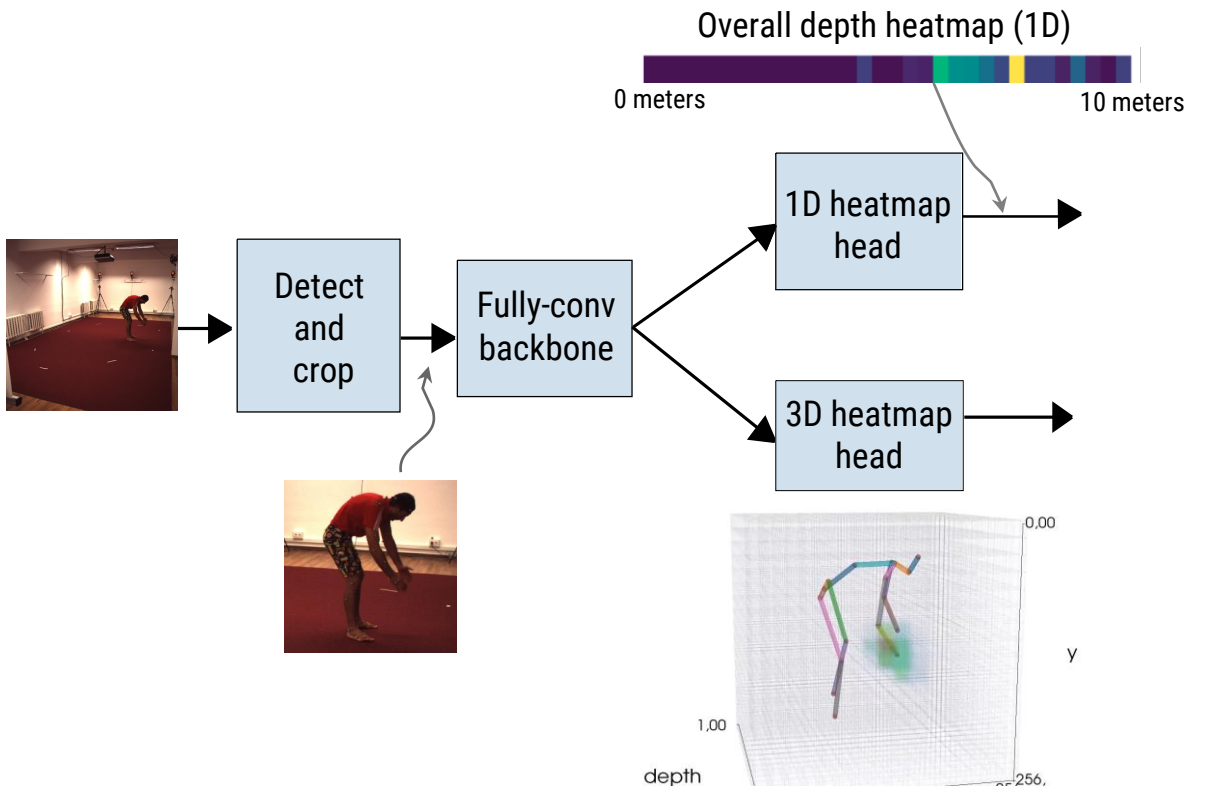
Approach



Approach

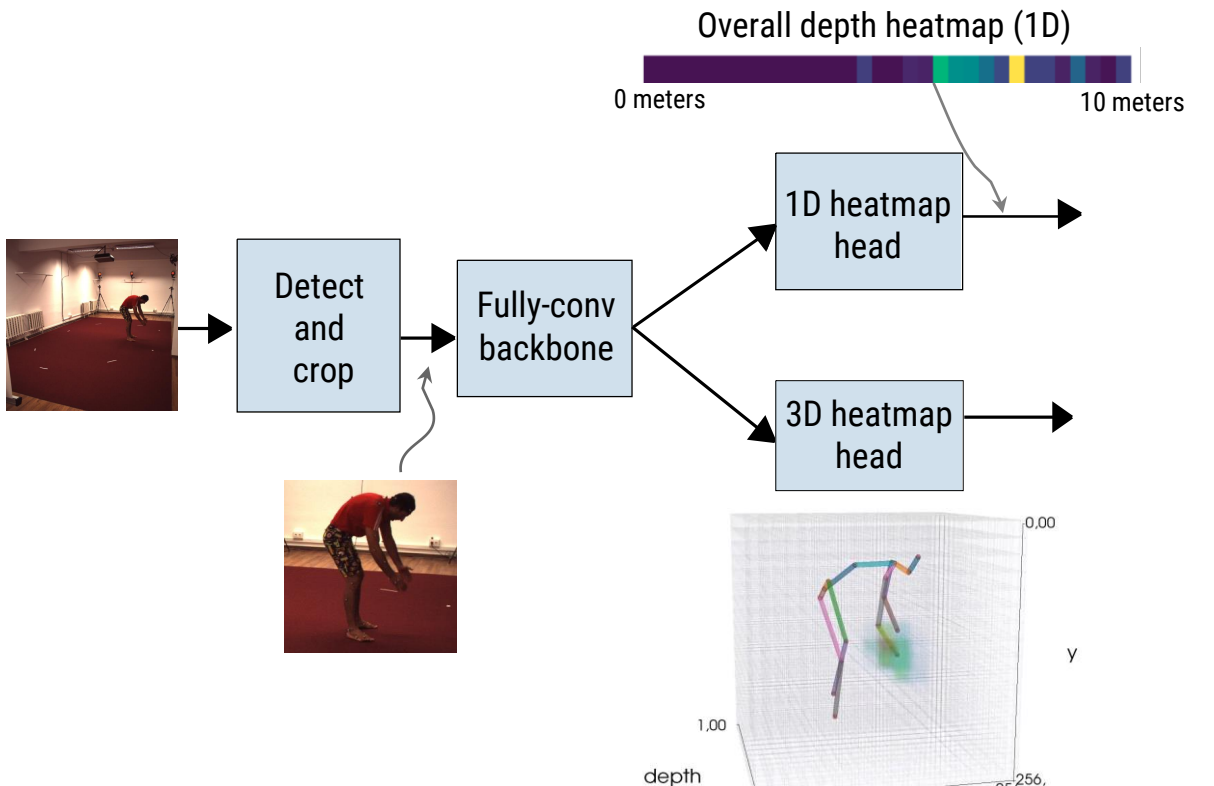

Overall depth heatmap (1D)
0 meters
10 meters
1D heatmap head
Detect and crop
Fully-conv backbone

Approach



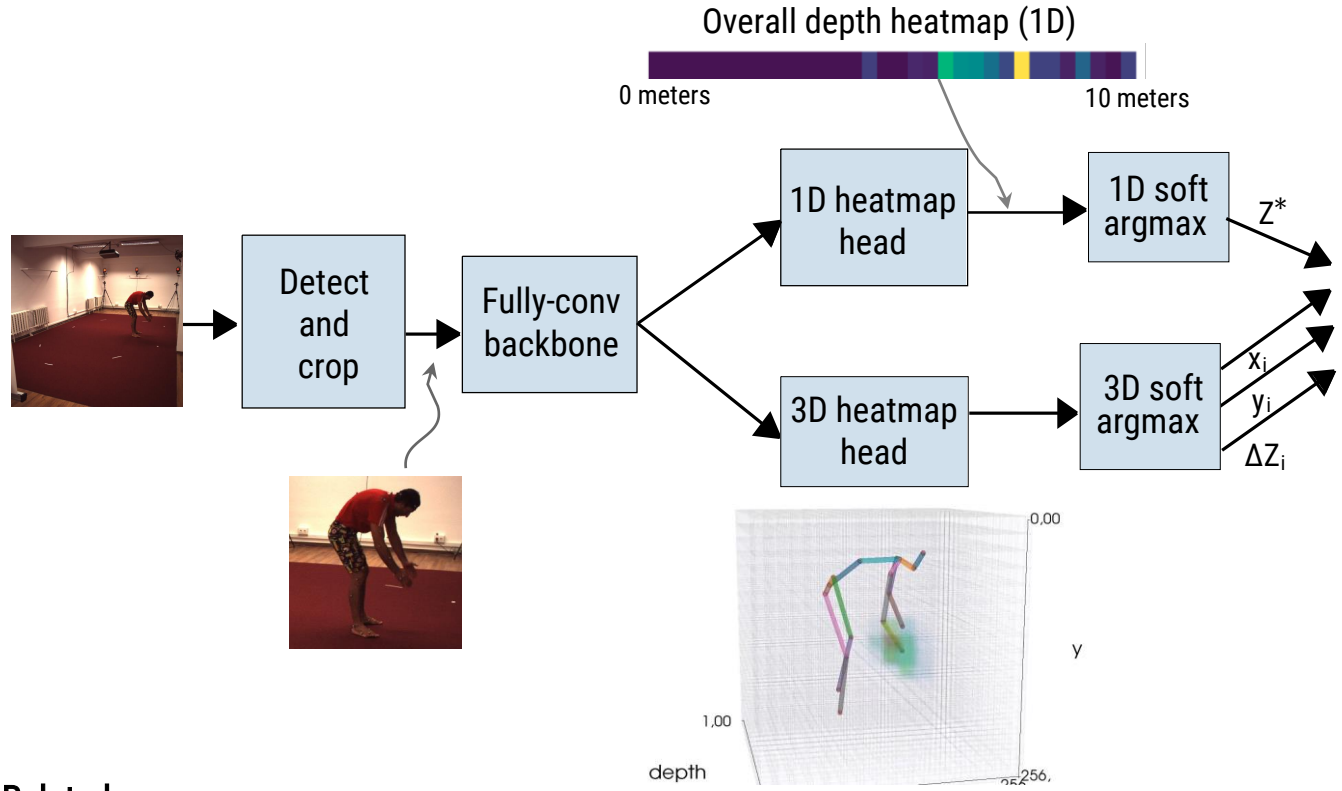
Related:
Pavlakos, CVPR'17
Sun, ECCV'18

Approach



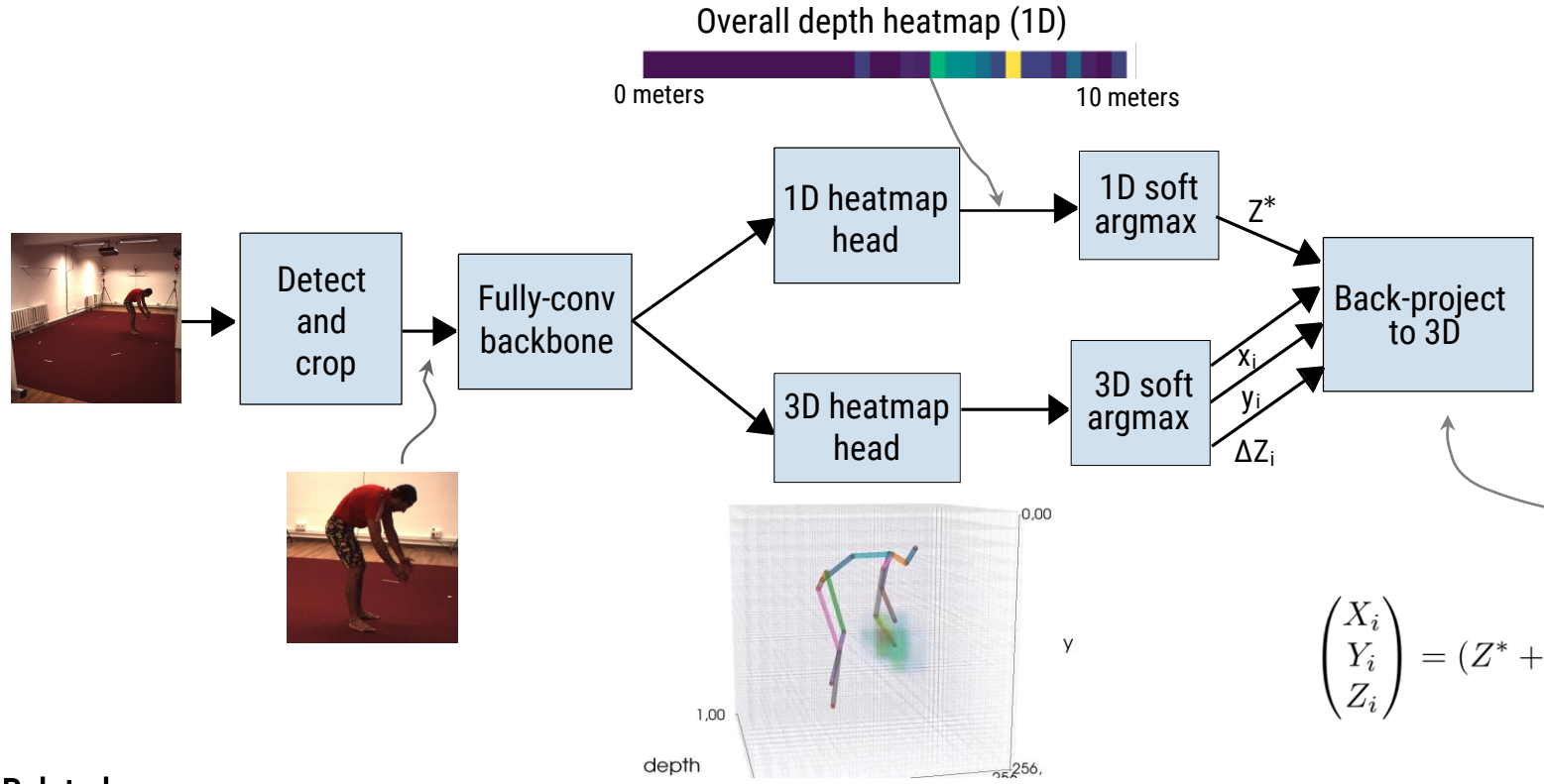
Related:
Pavlakos, CVPR'17
Sun, ECCV'18

Approach



Related:
Pavlakos, CVPR'17
Sun, ECCV'18

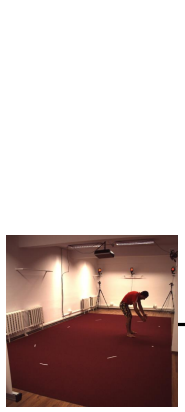
Approach



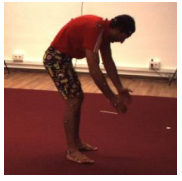
$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = (Z^* + \Delta Z_i) \begin{pmatrix} fs & 0 & W/2 \\ 0 & fs & H/2 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix}$$

Related:
Pavlakos, CVPR'17
Sun, ECCV'18

Approach



Detect and crop



Fully-conv backbone

1D heatmap head

3D heatmap head

Overall depth heatmap (1D)

1D soft argmax

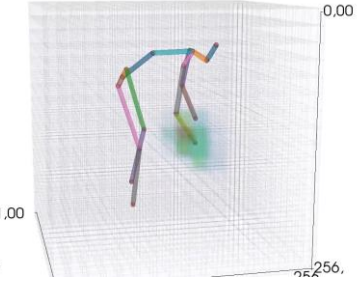
3D soft argmax

Z^*

x_i

y_i

ΔZ_i



Back-project to 3D

$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = (Z^* + \Delta Z_i) \begin{pmatrix} fs & 0 & W/2 \\ 0 & fs & H/2 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix}$$

Ground truth

Subtract root pred.

L1 loss

Related:
Pavlakos, CVPR'17
Sun, ECCV'18

Synthetic Occlusions

Pascal VOC objects

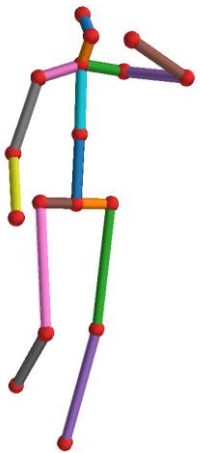
Sárándi *et al.*: How robust is 3D human pose estimation to occlusion? arXiv:1808.09316, IROS'18 Workshops

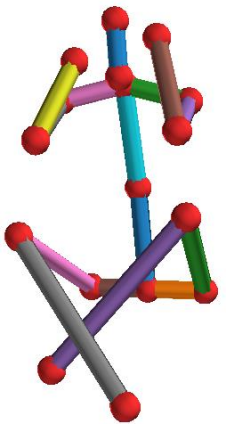
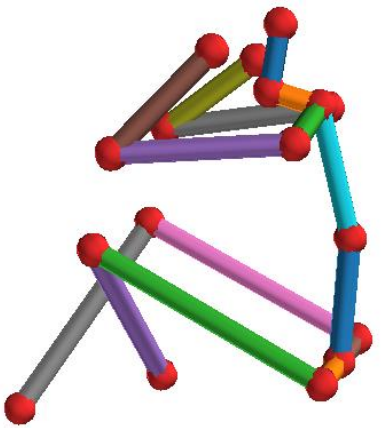
Synthetic Occlusions

\+ geometry and color =

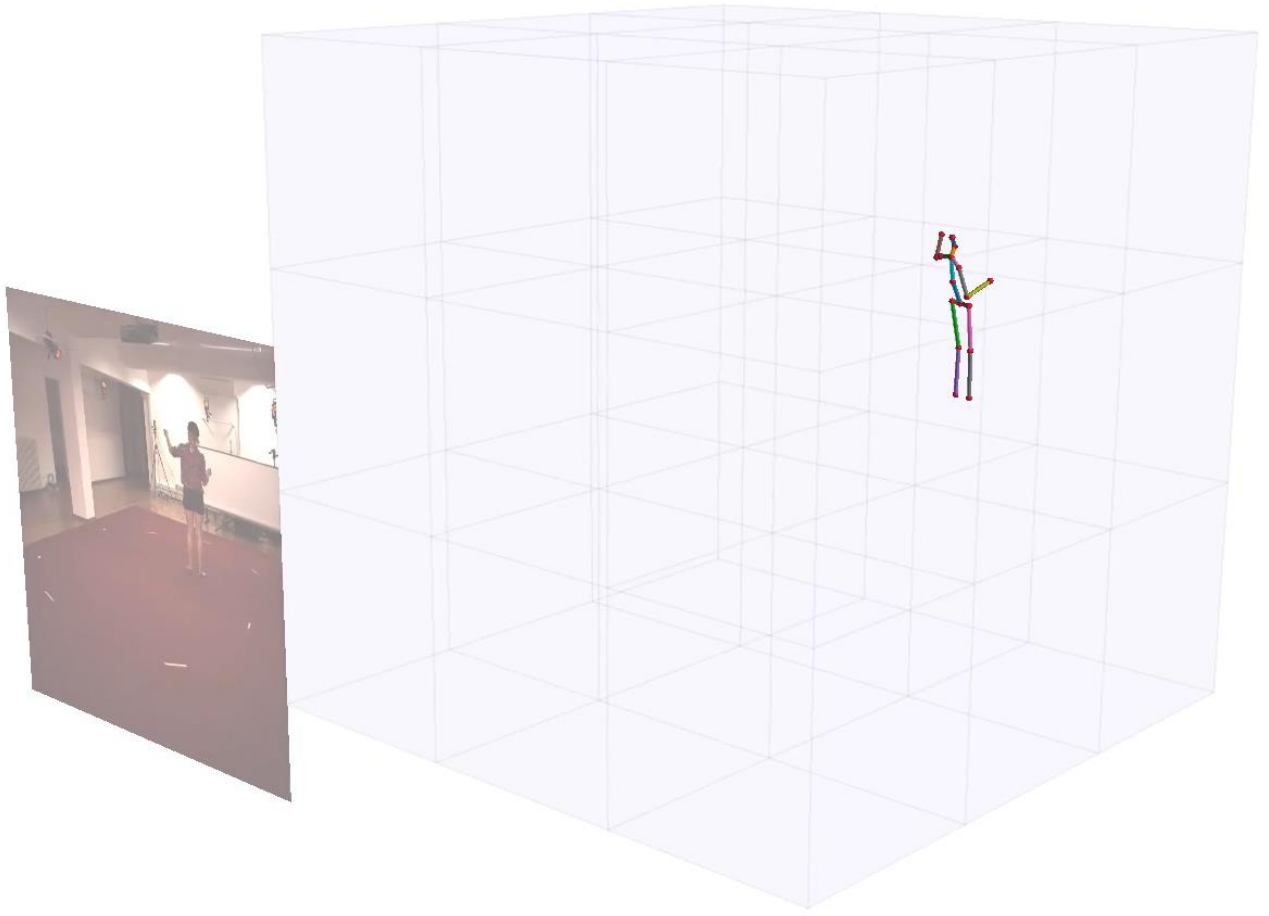
Sárándi *et al.*: How robust is 3D human pose estimation to occlusion? arXiv:1808.09316, IROS'18 Workshops

Results



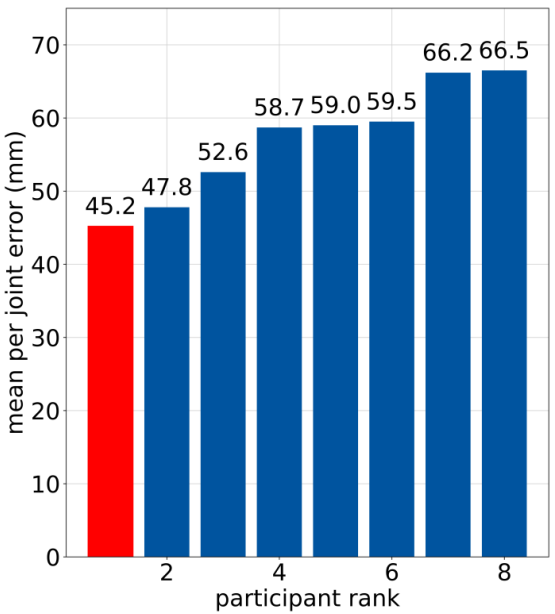



Results



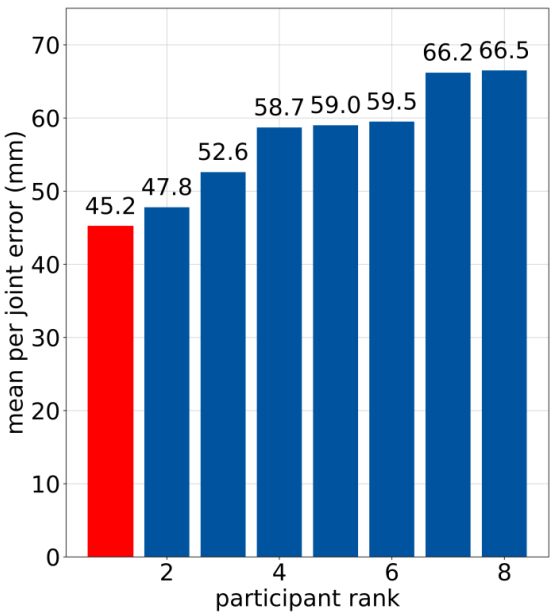
Results

1st place
in the Challenge



Results

1st place
in the Challenge

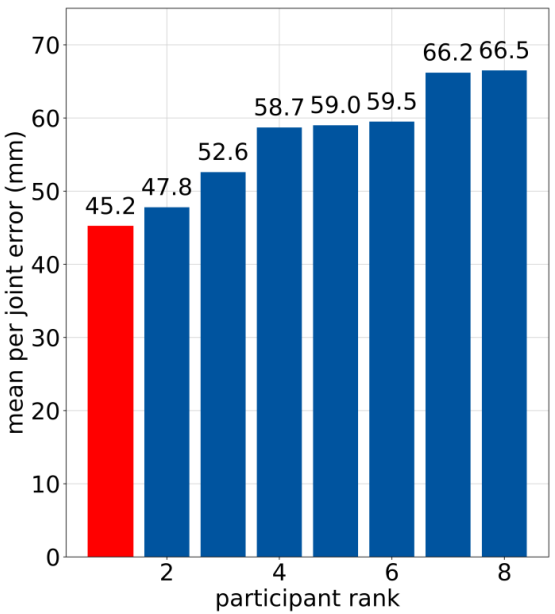


Best result on the full H3.6M
if no extra 2D pose datasets are used

Method	Extra pose data in training?	
	no	yes
Tekin (CVPR'16)	125.0	–
Zhou (CVPR'16)	113.0	–
Zhou (ECCV'16)	107.3	–
Sun (ICCV'17)	92.4	59.1
Martinez (ICCV'17)	–	62.9
Zhou (ICCV'17)	–	55.9
Pavlakos (CVPR'18)	71.9	56.2
Sun (ECCV'18)	64.1	**49.6**
Ours (no occl. augm., $p_{occ} = 0\%$)	65.7	–
Ours ($p_{occ} = 50\%$)	55.4	–
Ours ($p_{occ} = 100\%$)	**54.2**	–

Results

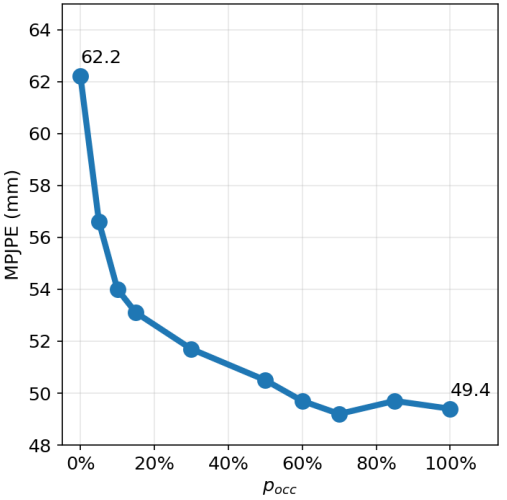
1st place in the Challenge



Best result on the full H3.6M if no extra 2D pose datasets are used

Method	Extra pose data in training? no	yes
Tekin (CVPR'16)	125.0	–
Zhou (CVPR'16)	113.0	–
Zhou (ECCV'16)	107.3	–
Sun (ICCV'17)	92.4	59.1
Martinez (ICCV'17)	–	62.9
Zhou (ICCV'17)	–	55.9
Pavlakos (CVPR'18)	71.9	56.2
Sun (ECCV'18)	64.1	**49.6**
Ours (no occl. augm., $p_{occ} = 0\%$)	65.7	–
Ours ($p_{occ} = 50\%$)	55.4	–
Ours ($p_{occ} = 100\%$)	**54.2**	–

Effect of occlusion augmentation



(evaluated on challenge validation set)

Conclusion

- Human3.6M has little appearance variation
  - Overfitting → data augmentation helps
- Simple, fast architecture, good performance
  - Heatmaps directly from backbone net
  - Soft-argmax on low-res heatmaps (16×16×16)
  - ~200 fps inference (Titan X GPU, excl. detection)
  - 1st place in 3D PoseTrack Challenge

Thank you!

sarandi@vision.rwth-aachen.de

BOSCH
Invented for life

BOSCH-FORSCHUNGSSTIFTUNG
IM STIFTERVERBAND

Visual Computing Institute

RWTH AACHEN UNIVERSITY