BOSCH
Invented for life

BOSCH-FORSCHUNGSSTIFTUNG
IM STIFTERVERBAND

Visual Computing Institute | RWTH AACHEN UNIVERSITY

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

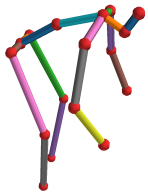
István Sárándi[1], Timm Linder[2], Kai O. Arras[2], Bastian Leibe[1]

[1]Visual Computing Institute, RWTH Aachen University – Aachen, Germany

[2]Robert Bosch GmbH, Corporate Research – Stuttgart, Germany

PoseTrack Challenge 2018 – 3D Task

Input image
Uncropped, static RGB image of a single person



```
[[   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]]
```

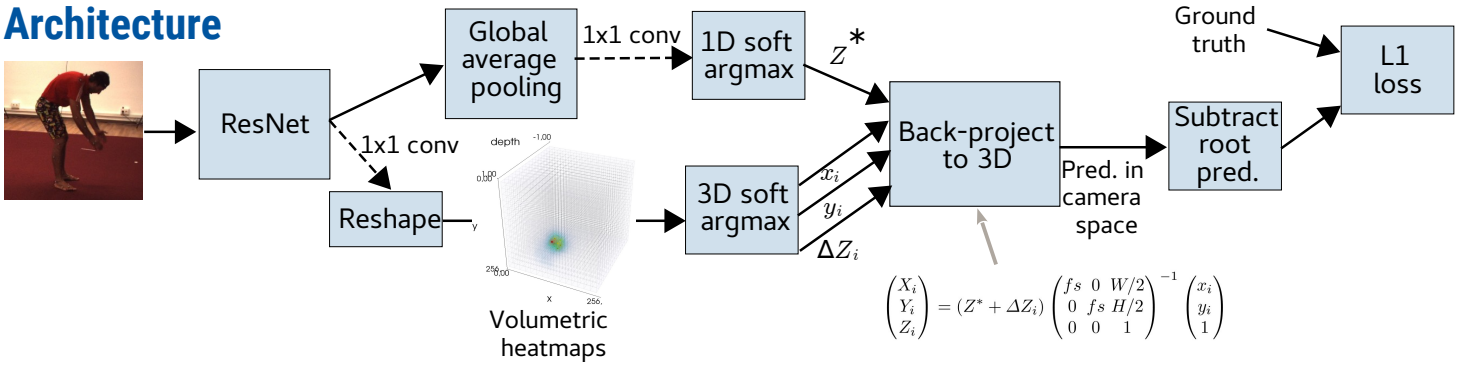
Output 3D skeleton
17 body joints in 3D camera space relative to the root (pelvis) joint

Our Approach

- Detect person with YOLOv3, then zoom & crop
- Predict **volumetric body joint heatmaps** directly, with a fully-convolutional backbone (ResNet-50v2)
- Predict person center depth with a **1D heatmap head**
- Obtain 3D points with **soft-argmax** and camera **back-projection**
- Minimize the **L1 loss** after subtracting root joint

- Achieved **first place** in the Challenge
- **No additional pose datasets** used for training
- **High frame rate** inference (204 fps, excl. detection) on Titan X GPU

Architecture



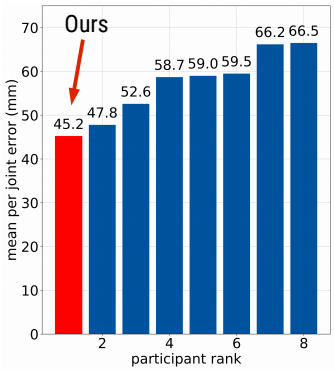
Occlusion Augmentation at Training

2638 occluder objects from Pascal VOC
Filter out 'person', 'truncated', 'difficult' and small object segments

Augmented inputs with pasted occluders
Applied with 50% probability, 1–8 objects, at random scale, at random position

Why not some simpler geometric shapes?
We found them less effective in our recent occlusion-robustness study[8]

Quantitative Results



Final challenge results

Method	Extra pose data in training? no	yes
Sun (ICCV'17) [3]	92.4	59.1
Martinez (ICCV'17) [4]	–	62.9
Zhou (ICCV'17) [5]	–	55.9
Pavlakos (CVPR'18) [6]	71.9	56.2
Sun (ArXiv) [7]	64.1	**49.6**
Ours (no occlusion augm.)	65.7	–
Ours (full)	**55.4**	–

Comparison on the full Human3.6M[1][2] benchmark
MPJPE, trained on subjects S1, S5, S6, S7, S8; tested on S9, S11

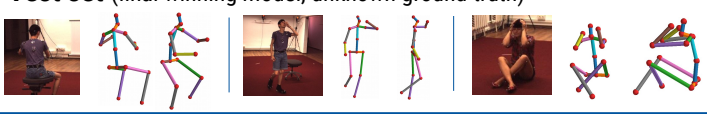
References

[1] Ionescu, C., Li, F., Sminchisescu, C.: Latent structured models for human pose estimation. ICCV (2011)
[2] Ionescu, C., Papava, D., Olaru, V., Sminchisescu, C.: Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments. PAMI (2014)
[3] Sun, X., Shang, J., Liang, S., Wei, Y.: Compositional human pose regression. ICCV (2017).
[4] Martinez, J., Hossain, R., Romero, J., Little, J.J.: A simple yet effective baseline for 3D human pose estimation. ICCV (2017)
[5] Zhou, X., Huang, Q., Sun, X., Xue, X., Wei, Y.: Towards 3D human pose estimation in the wild: a weakly-supervised approach. ICCV (2017)
[6] Pavlakos, G., Zhou, X., Daniilidis, K.: Ordinal depth supervision for 3D human pose estimation. CVPR (2018)
[7] Sun, X., Xiao, B., Liang, S., Wei, Y.: Integral human pose regression. ArXiv:1711.08229 (2017)
[8] Sárándi, I., Linder, T., Arras, K.O., Leibe, B.: How robust is 3D human pose estimation to occlusion? ArXiv:1808.09316 (2018)

Qualitative Results

Validation set (GT translucent)

Test set (final winning model, unknown ground truth)



Acknowledgment. This project has been funded by a grant from the Bosch Research Foundation, by ILIAD (H2020-ICT-2016-732737) and by ERC Consolidator Grant DeeViSe (ERC-2017-COG-773161).