How Robust is 3D Human Pose Estimation to Occlusion?

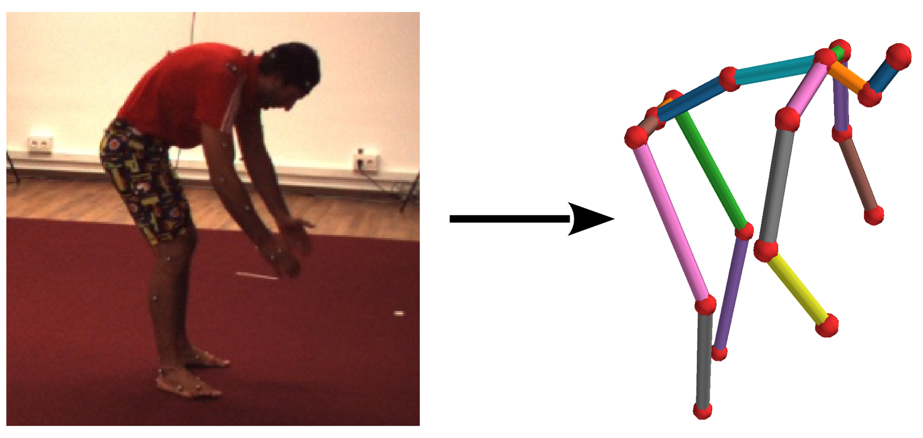
István Sárándi[1], Timm Linder[2], Kai O. Arras[2], Bastian Leibe[1]

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

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

Overview

3D Human Pose Estimation Task

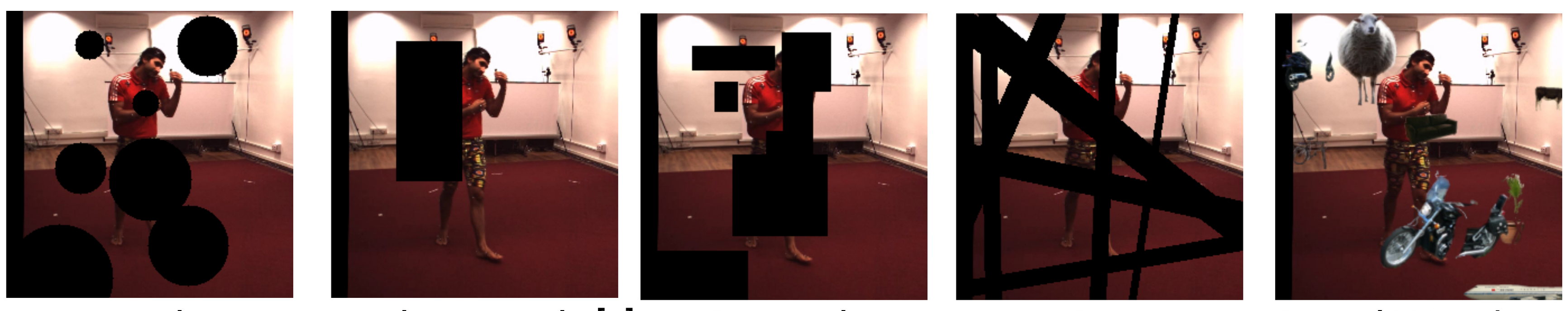


- Localize body joints in 3D camera space from an RGB image
- Useful for collaborative robotics
- Much progress in the last few years, as measured on current benchmarks, such as Human3.6M

Real Environments are More Challenging Than Benchmarks

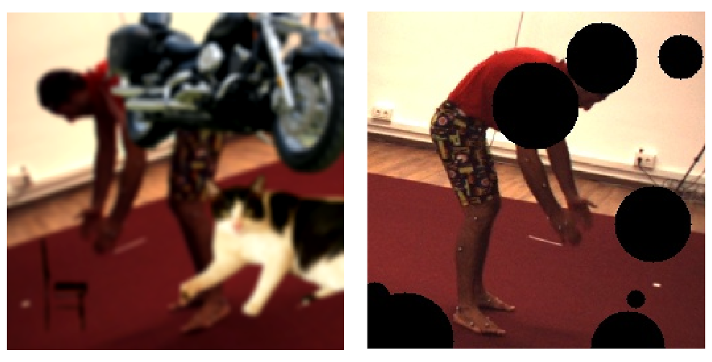
- Occlusion is common in shared human-robot environments
- Current benchmarks don't systematically model this
- How well do current methods work under occlusion?

Measure Robustness to Synthetic Occlusions



Circles | Single rectangle [1] | Rectangles | Bars | Pascal VOC objects

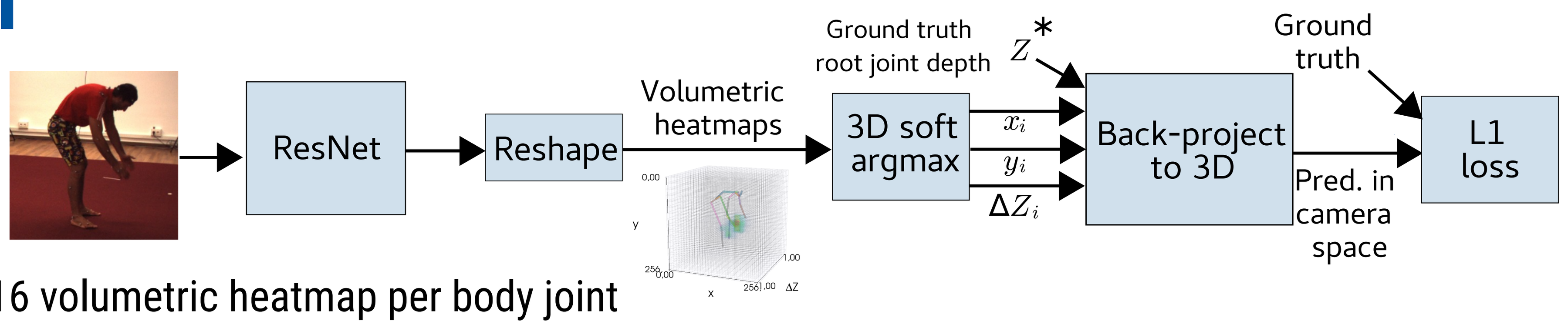
Apply Data Augmentation for Improved Robustness and Regularization



- Add synthetic occlusions to the training data as well
- In addition to usual augmentations
- Turns out to be a good regularizer even in non-occluded cases

Investigated Pose Estimation Model

- Based on two of the best recent methods [2][3]
- Performance on Human3.6M is at state-of-the-art level
- High frame rate inference (204 fps) on Titan X GPU
- Fully-convolutional backbone (ResNet-50) directly predicts a 16x16x16 volumetric heatmap per body joint
- Heatmaps are converted to coordinates with soft-argmax and are back-projected into metric 3D space, where the L1 loss is minimized

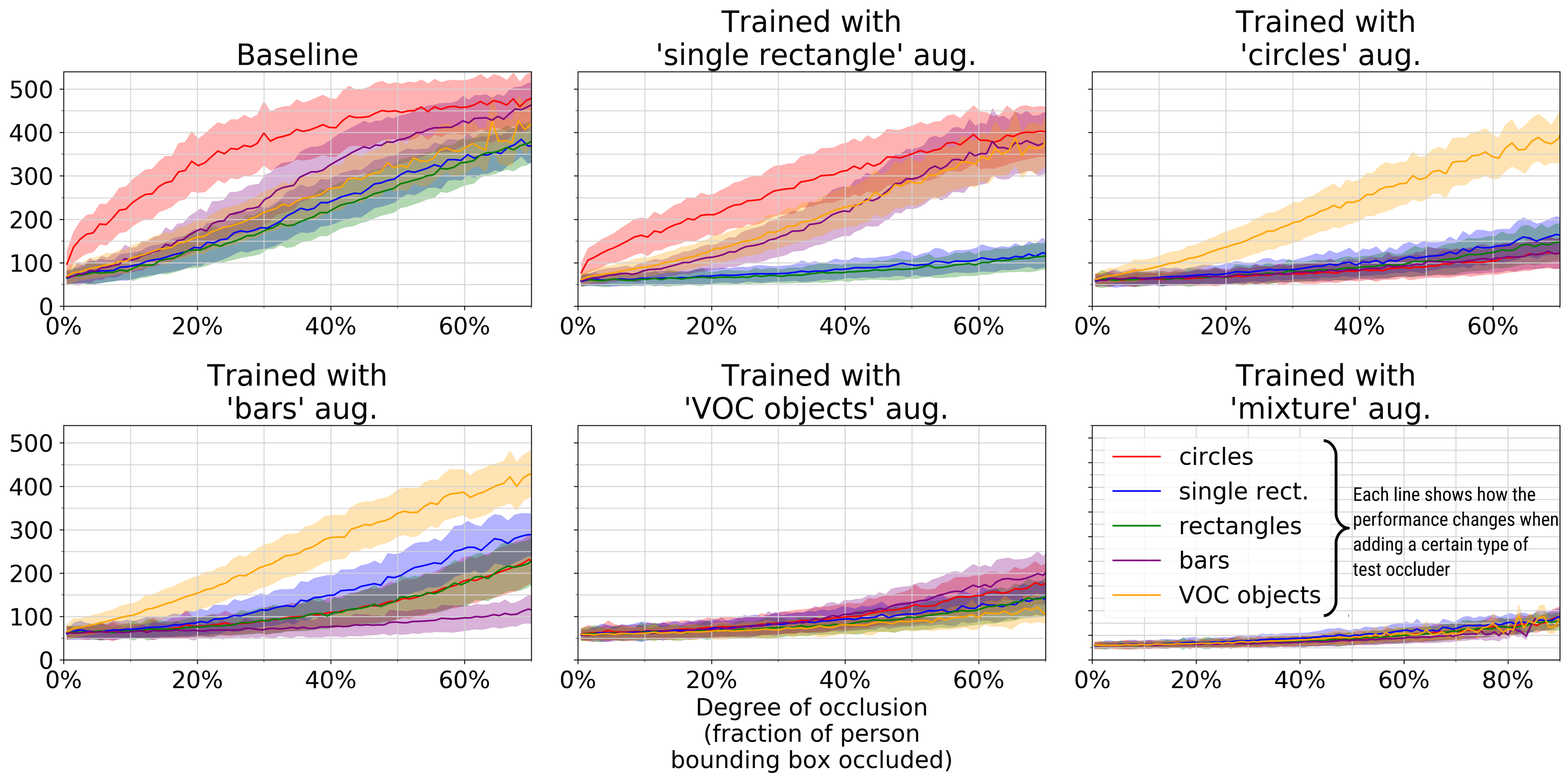


Evaluation

- Does the occluder shape matter? Which type of training augmentation improves robustness to which type of test-time occlusion?
- Evaluation measure: mean per joint position error after skeleton alignment at root joint (MPJPE)

Training-time augmentation (rows) vs. Test-time occlusion (columns)

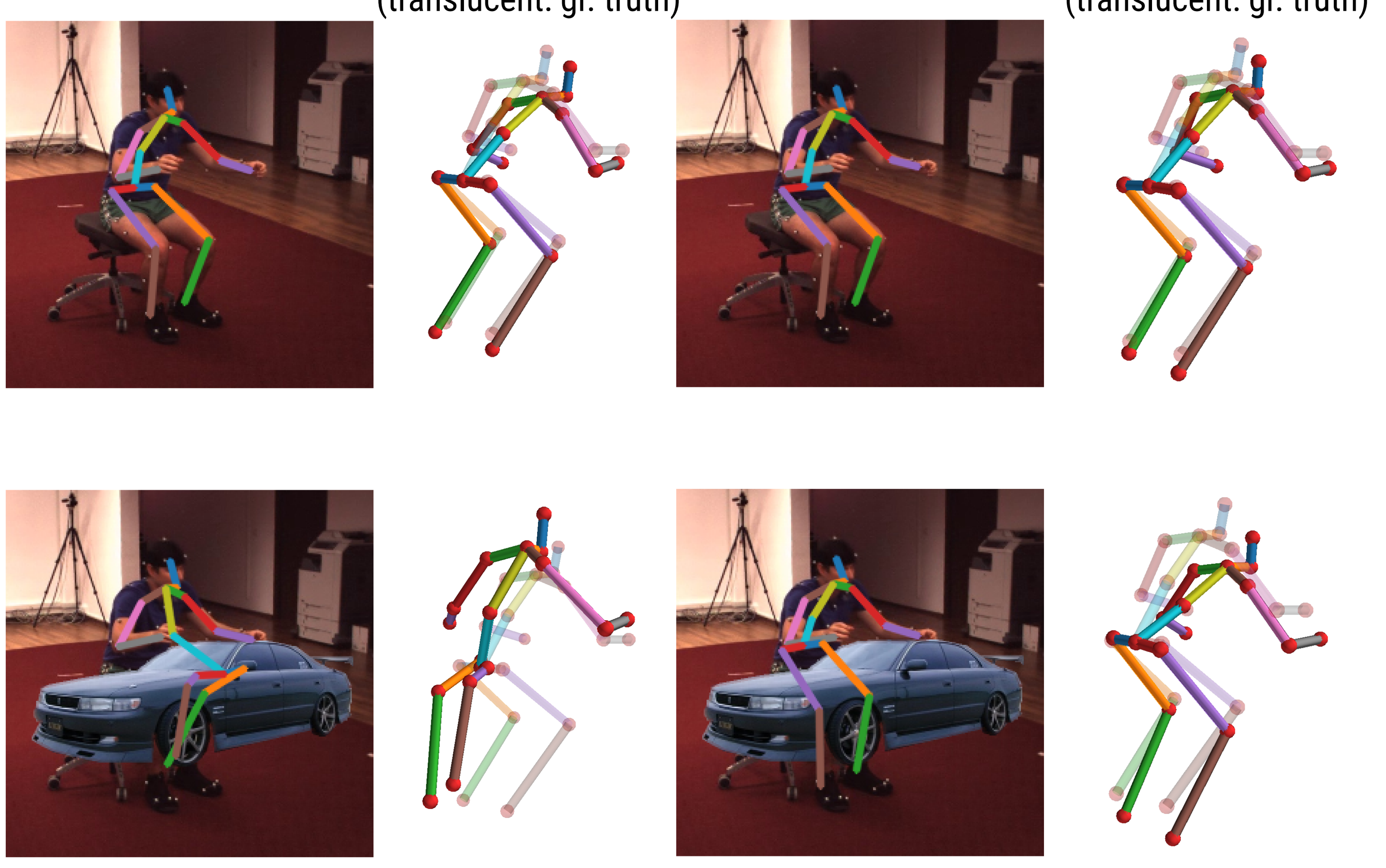
	none	single rect.	rectangles	circles	bars	VOC objects
none	67.3	190.4	176.8	370.9	249.8	216.5
single rect.	59.6	77.6	72.3	263.4	170.5	179.9
rectangles	60.0	81.4	71.9	217.4	154.6	194.4
circles	60.3	88.0	80.1	75.0	78.0	194.4
bars	63.3	120.6	93.8	94.7	72.7	218.3
VOC objects	59.3	83.3	75.9	88.0	88.6	72.9
mixture	59.6	79.7	71.8	72.5	68.1	74.4



Qualitative Example

Baseline model | Trained with VOC object aug.

Prediction | View from left (translucent: gr. truth) | Prediction | View from left (translucent: gr. truth)

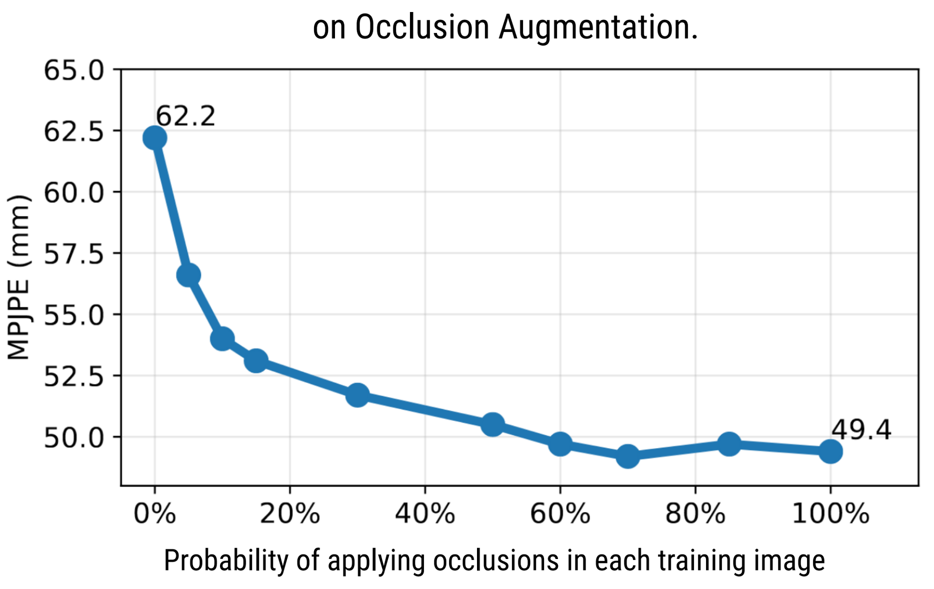


Key Findings

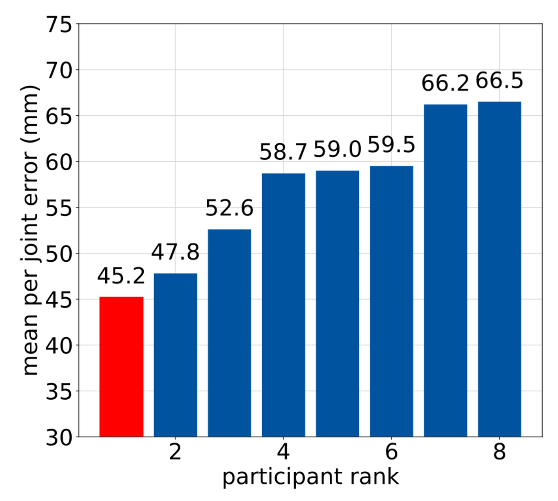
- Baseline 3D pose estimator is sensitive even to low degrees of occlusion
- Circular occluders are the most difficult
- Training with circles improves robustness to all simple shapes
- Robustness to Pascal VOC occluders not improved by augmenting with simple shapes
- Occlusion augmentation helps even for *unoccluded* test cases

➥Won the PoseTrack 3D Challenge at ECCV 2018, ahead of methods using external 2D datasets in training (details in [4])

PoseTrack'18 Validation Set Ablation on Occlusion Augmentation.



PoseTrack'18 Challenge Results



[1] Zhong, Z.; Zheng, L.; Kang, G., Li, S.; Yang, Y.: Random erasing data augmentation. arXiv:1708.04896, 2017

[2] Pavlakos, G.; Zhou, X.; Derpanis, K. G.; Daniilidis, K.: Coarse-to-fine volumetric prediction for single-image 3D human pose. CVPR 2017

[3] Sun, X.; Xiao, B.; Liang, S.; Wei, Y.: Integral human pose regression. ECCV 2018

[4] Sárándi, I.; Linder, T.; Arras, K. O.; Leibe, B.: Synthetic occlusion augmentation with volumetric heatmaps for the 2018 ECCV PoseTrack Challenge on 3D human pose estimation. ArXiv:1809.04987 (2018)

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).