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

István Sárándi¹, Timm Linder², Kai O. Arras², Bastian Leibe¹
¹Visual Computing Institute, RWTH Aachen University – Aachen, Germany
²Robert Bosch GmbH, Corporate Research – Stuttgart, Germany

PoseTrack Challenge 2018 – 3D Task

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]

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

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

Ours (no occlusion aug.) 65.7 –
Ours (full) 55.4 –

References