Supplemental Material: Accelerating Large-Kernel Convolution Using Summed-Area Tables

Table 1: Comparisons on the MPII Human Pose dataset using the validation set [2]. "Pretrain" indicates that the backbone network is pretrained on the ImageNet classification task.

Method	Pretain	#Params	FLOPs	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	PCKh@0.5	PCKh@0.1
SimpleBaseline [3]	N	34.0M	12.0G	96.3	94.9	87.7	81.7	87.5	82.1	77.8	87.5	32.1
SimpleBaseline [3]	Y	34.0M	12.0G	96.4	95.3	89.0	83.2	88.4	84.0	79.6	88.5	33.9
Dilated Convolution	N	1.88M	7.7G	96.1	94.1	87.1	81.4	86.2	81.3	75.7	86.7	33.4
3×3 Convolution	N	1.87M	7.6G	87.8	85.9	75.7	69.5	70.4	63.8	59.0	74.1	28.0
Burkov et al. [1]	N	1.85M	7.6G	96.4	94.5	86.6	79.9	87.1	81.1	75.7	86.6	30.8
Ours	N	1.85M	7.7G	96.7	95.3	89.4	83.8	88.1	83.3	77.7	88.4	36.4

Table 1 provides a comparison between our model, box convolution method by Burkov and Lempitsky [1], and other methods described in the main submission. To provide the comparisons, we use the widely accepted validation set by Tompson et al. [2]. The choice of the validation set over the official test set is caused by the strict rule on the number of evaluations permitted for the MPII dataset[1].

Notes on comparisons:

- SimpleBaseline [3] — we use the publicly available implementation and the pretrained model [2].
- Burkov and Lempitsky [1] — we replace our implementation of the box convolution layers with one provided by the prior work[3].

We also report PCKh@0.1, which uses a tighter threshold of 0.1. Our method prominently outperforms all other methods in terms of PCKh@0.1, indicating that our method is able to locate joints more precisely.

References

[1] Egor Burkov and Victor Lempitsky. Deep neural networks with box convolutions. In *Advances in Neural Information Processing Systems*, pages 6214–6224, 2018.

[2] Jonathan Tompson, Ross Goroshin, Arjun Jain, Yann LeCun, and Christoph Bregler. Efficient object localization using convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 648–656, 2015.

[3] Bin Xiao, Haiping Wu, and Yichen Wei. Simple baselines for human pose estimation and tracking. In *European Conference on Computer Vision (ECCV)*, pages 466–481, 2018.

---

[1] `http://human-pose.mpi-inf.mpg.de/#evaluation`

[2] `https://github.com/microsoft/human-pose-estimation.pytorch`

[3] `https://github.com/shrubb/box-convolutions`

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.