

Summary

- Top-down approach to person detection and pose estimation:
 - Faster-RCNN based person box detection.
 - Pose estimation by Hough voting on heatmaps + offsets.
 - Keypoint-based rescoring and non max suppression.
- COCO 2017 keypoints competition results (*single model*):
 - 0.710 mAP on COCO test-dev.
 - 0.691 mAP on COCO test-challenge.



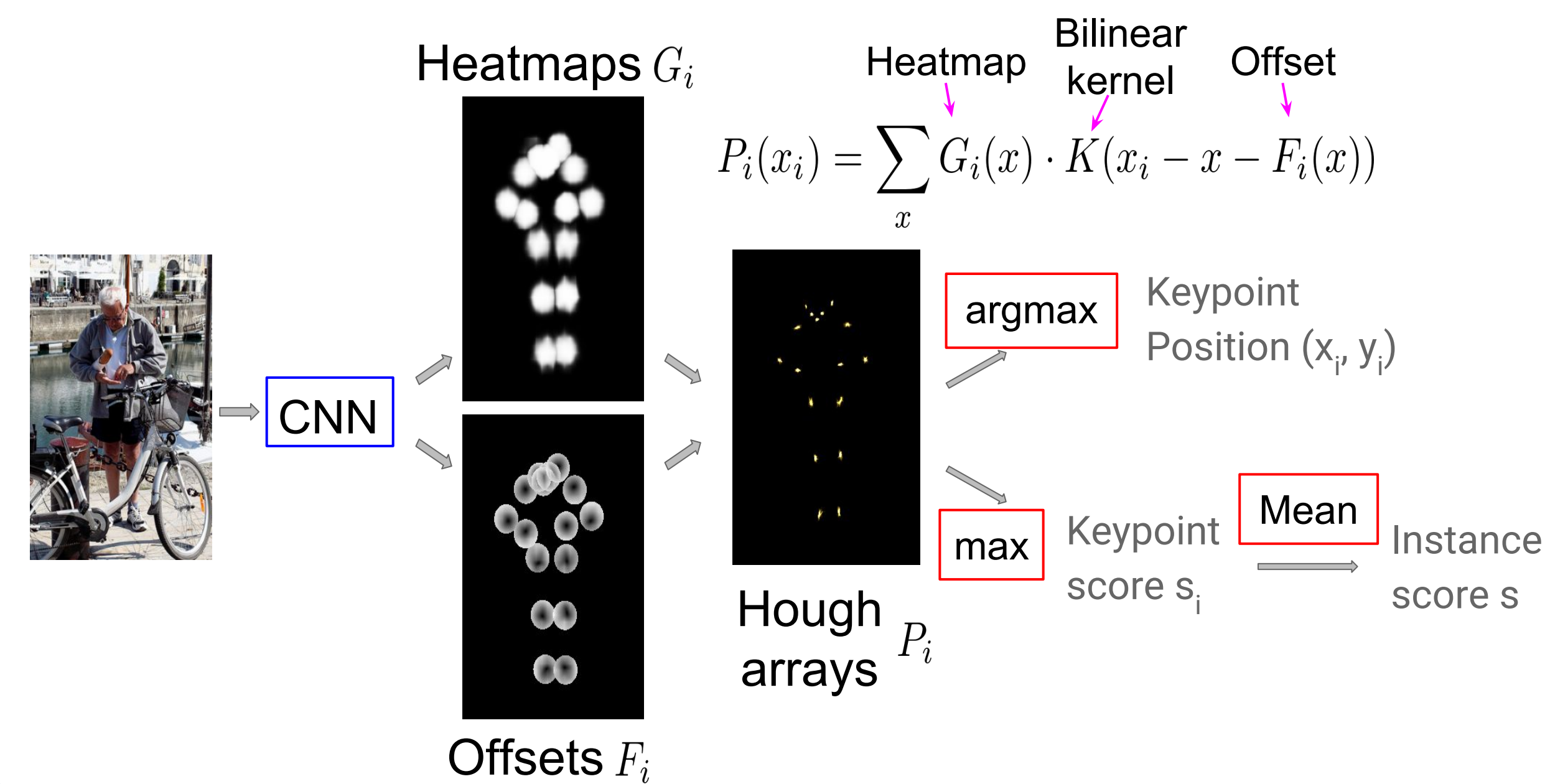
(1) Person detection + Crop

(2) Pose estimation

1. Person detection

- Faster-RCNN model, using ResNet-101 features, trained for person bbox detection on COCO using Tensorflow Object Detection API [2].
- Fixed-size image crop extraction, preserving aspect ratio.
- Pass each crop separately to pose estimation model.

2. Pose Estimation



Evaluation on COCO Keypoints 2017 (test-dev)

	AP	AP@.5	AP@.75	AP (M)	AP (L)
CMU-Pose [3]	0.618	0.849	0.675	0.571	0.682
Mask-RCNN [4]	0.631	0.873	0.687	0.578	0.714
Assoc. Embed [5]	0.655	0.868	0.723	0.606	0.726
Ours (COCO data)	0.669	0.864	0.736	0.640	0.720
Ours (COCO+internal)	0.696	0.872	0.766	0.670	0.742
Ours (COCO+internal) w. ResNet-152	0.710	0.879	0.777	0.690	0.752

- Our internal dataset is roughly 2x the size of COCO. We used it to augment the training set for the second stage pose estimator (first stage was COCO only).

Example Results



References

1. G. Papandreou et al, "Towards Accurate Multi-person Pose Estimation in the Wild ", CVPR 2017.
2. J. Huang et al., "Speed/accuracy trade-offs for modern convolutional object detectors", CVPR 2017
3. Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", CVPR 2017
4. K. He, G. Gkioxari, P. Dollár, R. Girshick, "Mask-RCNN", ICCV 2017
5. A. Newell, Z. Huang, J. Deng, "Associative Embedding", ICCV 2017