BoIR: Box-Supervised Instance Representation for Multi-Person Pose Estimation

Uyoung Jeong[1], Seungryul Baek[1], Hyung Jin Chang[2], Kwang In Kim[3]

[1]Ulsan National Institute of Science and Technology [2]University of Birmingham [3]Pohang University of Science and Technology

Motivation

Existing single-stage MPPE methods

- Sparse instance representation learning
  - Supervise only on GT keypoint locations
  - No loss for single person
- Insufficient multi-task supervision
  - More task heads, more computational cost
  - Auxiliary tasks might dominate over the primary task

Contribution

- Spatially rich instance representation learning with box-level supervision
  - Bbox Mask Loss provides learning signal on the entire image region, even when only one person is present
- Auxiliary task heads without additional computational cost during inference
  - They are used only for training, and removed during inference
  - Share the bottleneck ASPP to prevent overtaking the primary task

Method (1)

Overall framework

- Backbone output feature f passed to task-specific heads
- Instance-wise keypoint head(kpt) is a primary task head

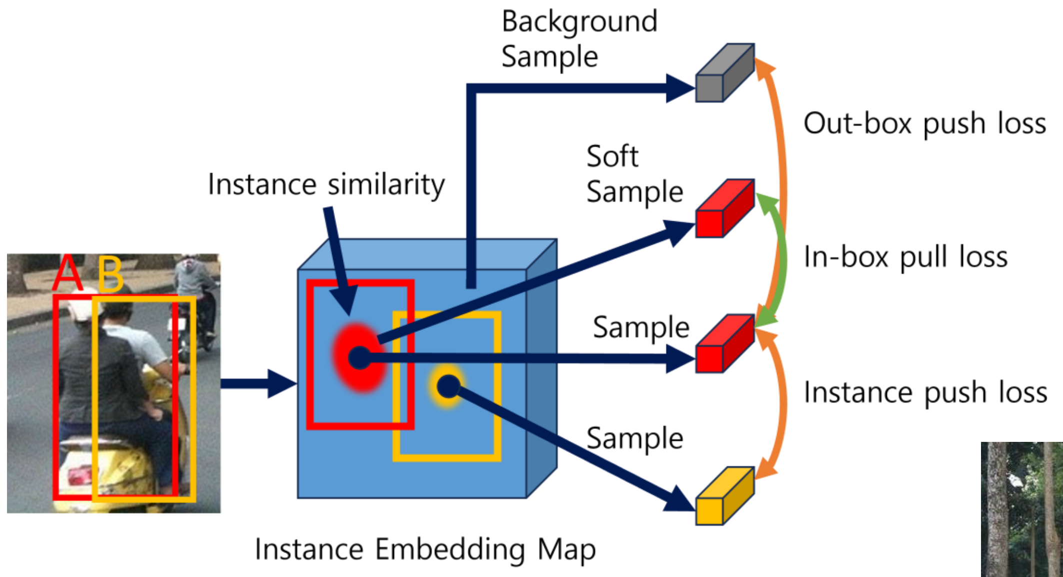
Auxiliary task heads

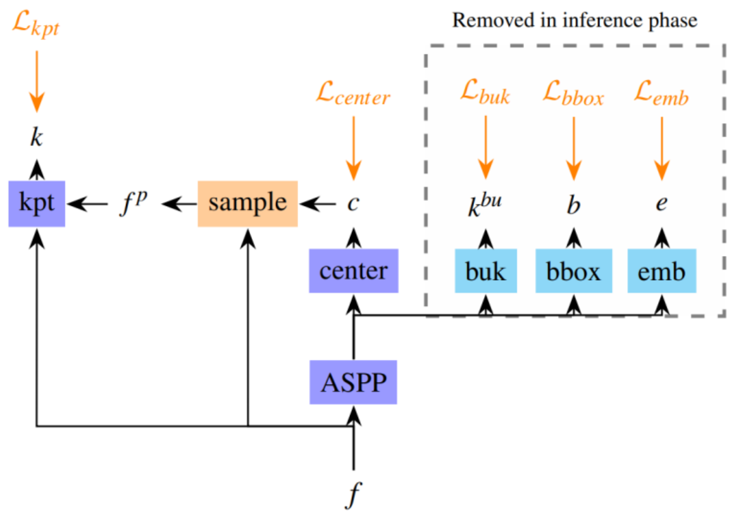
- Used only for training, and removed during inference
- Share the bottleneck ASPP module to prevent overtaking the primary task

Method (2)

Bbox Mask Loss

- Utilize box annotation which is far more abundant and easier to obtain than segmentation level annotation
- For each instance:
  - 3 embedding samples (center, positive, background)
  - 3 pull/push terms (in-box pull/push, out-box push)





Results

- Metric: mAP(mean Average Precision) (%)
- *: train on COCO and then apply finetuning

Method	COCO val	COCO test-dev	OCHuman val	OCHuman test	CrowdPose test
DEKR(W32)	68.0	67.3	37.9	36.5	65.7
DEKR(W48)	71.0	70.0	-	-	-
ED-Pose(R50)	71.7	69.8	-	-	69.9
CID(W32)	69.8	68.9	44.9	44.0	71.3 (74.9*)
CID(W48)	-	70.7	46.1	45.0	72.3
BoIR(W32)	**70.6**	**69.5**	**47.4**	**47.0**	70.6 (**75.8***)
BoIR(W48)	**72.5**	**71.2**	**49.4**	**48.5**	71.2 (**77.2***)