MDPose: Real-Time Multi-Person Pose Estimation via Mixture Density Model (Supplementary Material)

Seunghyeon Seo[1] **Jaeyoung Yoo**[2] **Jihye Hwang**[3] **Nojun Kwak**[1]

[1]Seoul National University
[2]NAVER WEBTOON AI
[3]LG Electronics

A ANALYSIS OF THE DISTRIBUTION OF MIXTURE COMPONENTS

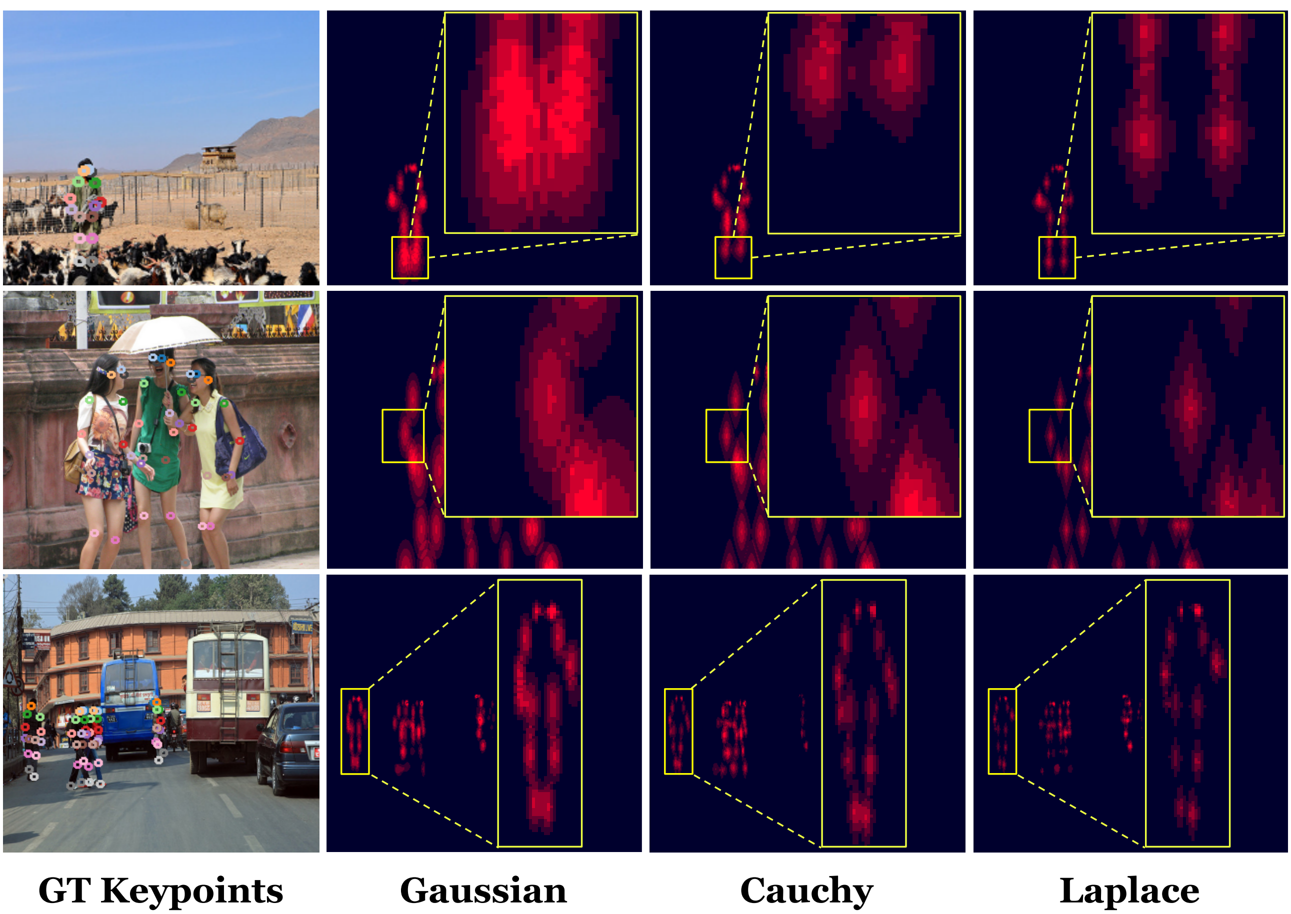

Figure A: **Visualization of estimation results with different mixture distributions of MDPose.**

Tab. A shows the accuracy and underflow ratio of different mixture distributions. The MDPose with Laplace mixture distribution outperforms the one with either the Gaussian or Cauchy with a noticeable gap of AP^{kp}. Since the tails of Laplace and Cauchy fall off **less sharply** than the Gaussian, they are relatively free from the underflow problem. Furthermore, as the tails of Laplace fall off **more rapidly** than the Cauchy and it has a **sharper peak**, it leads to more efficient weighting for

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

Table A: **Mixture model of different exponential distributions.** The Laplace is more suitable than the others for multi-person pose estimation.

Dist.	AP^{kp}	AP^{kp}_{50}	AP^{kp}_{75}	AP^{kp}_{M}	AP^{kp}_{L}	Underflow R.
Gaussian	50.5	79.7	54.0	41.1	63.8	0.184
Cauchy	50.6	79.6	54.1	41.4	63.5	**0.0**
Laplace	**51.5**	**80.4**	**55.1**	**42.0**	**64.7**	0.086

good and bad estimations during the training process. As demonstrated in Fig. A, the Laplace mixture distribution enables more accurate localization of human keypoints than the respective mixture distributions of the Gaussian and Cauchy.

B ANALYSIS OF GROUPING RANDOMNESS THROUGH VISUALIZATION

(a)

(b)

Figure B: **Visualization of our MDPose with (a) non-random grouping and (b) RKG.**

As mentioned in Sec. 4.2 in the main paper, our proposed RKG strategy enables learning of the overall joint distributions of all keypoints while the non-random grouping learns only the joint distributions of each pre-defined keypoint groups. Fig. B shows the qualitative results of our MDPose with (a) non-random grouping and (b) RKG. The results are obtained from the MDPose (ResNet-50 [He et al., 2016]) with $K_g = 3$, $N_g = 6$ and 320x320 input size on the COCO validation set [Lin et al., 2014]. As shown in Fig. B (a), the model trained by non-random grouping has a difficulty in differentiating the left and right of limbs, due to lack of learning the overall relationship between every keypoint. On the contrary, the MDPose trained by RKG (Fig. B (b)) shows superior performance with well-distinguished left and right of limbs.

C QUALITATIVE RESULTS

Figure C: **Qualitative results of MDPose (ResNet-101) on OCHuman validation set, with $K_g = 3$, $N_g = 6$ and 896x896 input size.**

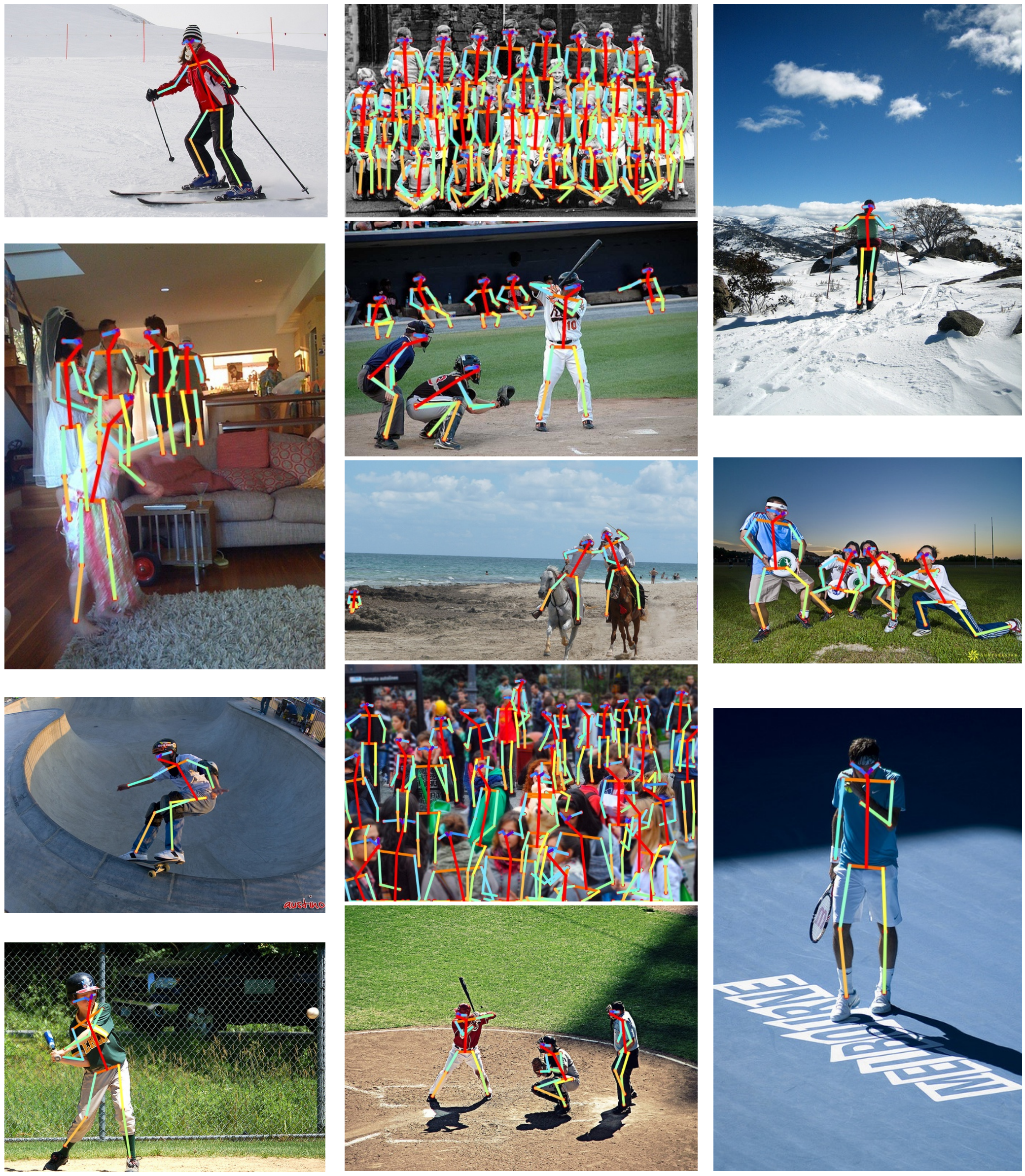

Figure D: **Qualitative results of MDPose (ResNet-50) on COCO validation set, with $K_g = 3$, $N_g = 6$ and 896x896 input size.**

References

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.