Supplementary Materials

1. Comparison with latest regression SOTA Poseur

The recent SOTA method Poseur [15] has FPN feature and well-designed encoder/decoder, which brings a larger computational burden while achieving higher performance. Thus, as shown in Tabel A, 1) when with the same backbone (HRNet-W48), DistilPose achieves much higher FPS while Poseur achieves higher accuracy. 2) When with similar AP (DistilPose-S (71.6%) & Poseur-MobileNetV2 (71.9%)), the FPS of DistilPose-S is 4.73 times that of Poseur.

Method	Backbone	Resolution	Param	GFLOPs	mAP	FPS
Poseur	MobileNetV2	256×192	11.36M	0.5	71.9%	12.1
	ResNet-50	256×192	33.26M	4.6	75.4%	12.0
	HRNet-W32	256×192	38.19M	7.4	76.9%	5.5
	HRNet-W48	384×288	74.27M	33.6	78.8%	5.4
DistilPose	stemnet	256×192	5.36M	2.4	71.6%	40.2
	HRNet-W48-s3	256×192	21.27M	10.3	74.4%	13.7

Table A. Comparison between Poseur and DistilPose on MSCOCO ***val*** dataset.

2. Deviation for Basic Distribution Simulation

We conduct an ablation study to demonstrate that predicting deviations from different directions (σ_x/σ_y for horizontally/vertical respectively) can better help student model learn the distribution information from teacher heatmaps than predicting one deviation σ for all directions, as shown in Tab.B. We also provide a set of sample comparisons of predicted deviations based on the same input image as Fig.3.

Furthermore, we provide a set of visualization cases for comparison between teacher heatmaps and different kinds of basic distribution simulation, as shown in Fig.A. The three rows in Fig.A from top to bottom show the local distribution of teacher heatmaps, one-deviation Basic Distribution Simulation and two-deviations Basic Distribution Simulation, respectively. Since the deviation used for target generation during the training of teacher model is usually default to 2, the deviations we predict are also around 2. We can see that heatmaps generated by two-deviations are more similar to teacher heatmaps than heatmaps generated by one-deviation.

	mAP	Nose	Shoulder(l)	Shoulder(r)	Elbow(l)	Elbow(r)	Wrist(l)	Wrist(r)	Hip(l)	Hip(r)	Knee(l)	Knee(r)	Ankle(l)	Ankle(r)
σ	71.4%	1.99	2.04	2.02	2.00	2.04	2.10	2.03	2.00	2.01	2.04	2.05	2.07	2.08
(σ_x, σ_y)	71.6%	(2.03, 2.02)	(2.00,1.99)	(2.01,2.04)	(2.05,2.04)	(2.02,2.05)	(2.07,2.06)	(2.02,2.07)	(1.98,2.01)	(2.01,2.05)	(1.96,2.07)	(2.00,2.07)	(2.02,2.18)	(1.93,2.14)

Table B. Comparison between single deviation and horizontal/vertical deviations for Basic Distribution Simulation.

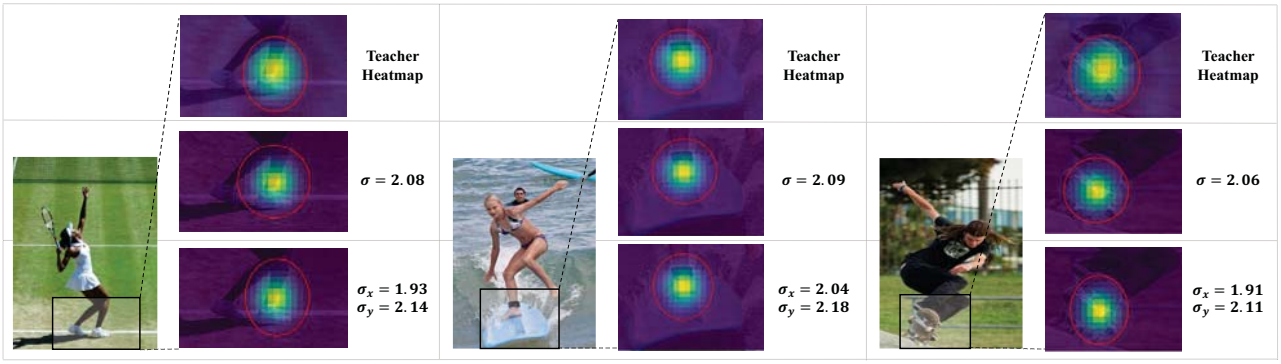

Figure A. Visualization of teacher heatmaps and different kinds of Basic Distribution Simulation. In the second row, σ represents the single deviation predicted for generating heatmaps. In the third row, σ_x and σ_y represent horizontal and vertical deviations for generating heatmaps, respectively.

3. More Visualization about Simulated Heatmaps

We provide more visualization cases of generated heatmaps for Basic Distribution Simulation in Simulated Heatmaps, as shown in Fig.B.

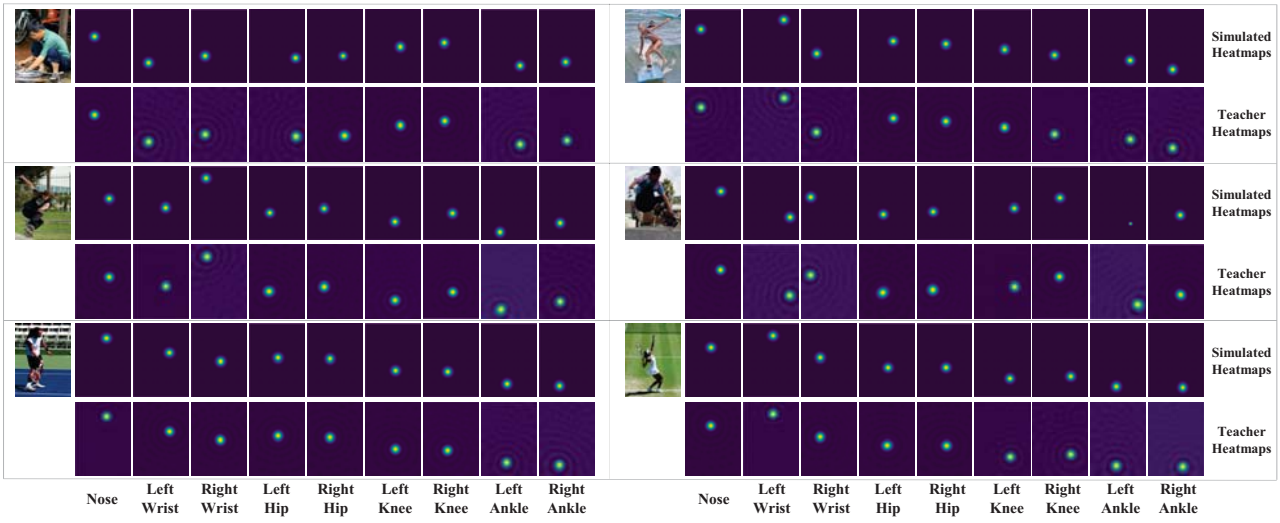

Figure B. Visualization of Simulated Heatmaps and teacher heatmaps.