Supplementary Materials For "SDPose: Tokenized Pose Estimation via Circulation-Guide Self-Distillation"

Sichen Chen[1*]   Yingyi Zhang[2*]   Siming Huang[2*]   Ran Yi[1]   Ke Fan[1]
Ruixin Zhang[2]   Peixian Chen[2]   Jun Wang[3]   Shouhong Ding[2]   Lizhuang Ma[1]

[1]Shanghai Jiao Tong University, [2]Tencent Youtu Lab, [3]Tencent WeChat Pay Lab33

1. Comparison with TokenPose-T

We further investigated the performance of the proposed method when based on the TokenPose-T structure. TokenPose-T is a pure transformer-based variant. We compared the SDPose setting with its alignment named SDPose-T-V2 and the results on the MSCOCO test-dev dataset are shown in Tab. 1. Our approach also achieves significant performance improvements compared to the baseline. This indicates that our approach is also applicable to a pure transformer-based model.

Methods	AP	AR
TokenPose-T*	60.5	67.3
SDPose-T-V2	**63.1**(↑ 2.6%)	**73.7**

Table 1. Results on MSCOCO test-dev dataset. * means we retrain and evaluate the models on MMPose.

2. More Results On CrowdPose

We further tested the results on Crowdpose based on SDPose-S-V1 with the version of the MCT module that uses three cycles, as shown in Tab. 2. Compared to SDPose-S-V1, SDPose-S-V1-3cycle gets better performance. This demonstrates the importance of the global information introduced by the multiple cycles of the MCT module in more difficult tasks like Crowdpose.

Methods	AP	AR
TokenPose-S-V1	55.7	65.2
SDPose-S-V1	**57.3** (↑ 1.6%)	**66.8**
SDPose-S-V1-3cycle	**57.5** (↑ 1.8%)	**67.0**

Table 2. Results on Crowdpose test dataset.

3. More Visualization Results

We provide more visualization cases of Human Pose Estimation by SDPose-S-V1, as shown in Fig. 1.

4. More Parameter Distributed Visualization Results

We provide more visualization cases of parameters distributed in SDPose-S-V1, as shown in Fig. 2. All parameter distributions of SDPose-S-V1 in Fig. 2 have a larger variance, proving that our method can fully optimize the whole transformer. Meanwhile, we note that layers at deeper locations are better optimized. This may represent richer global information learned by deeper layers.
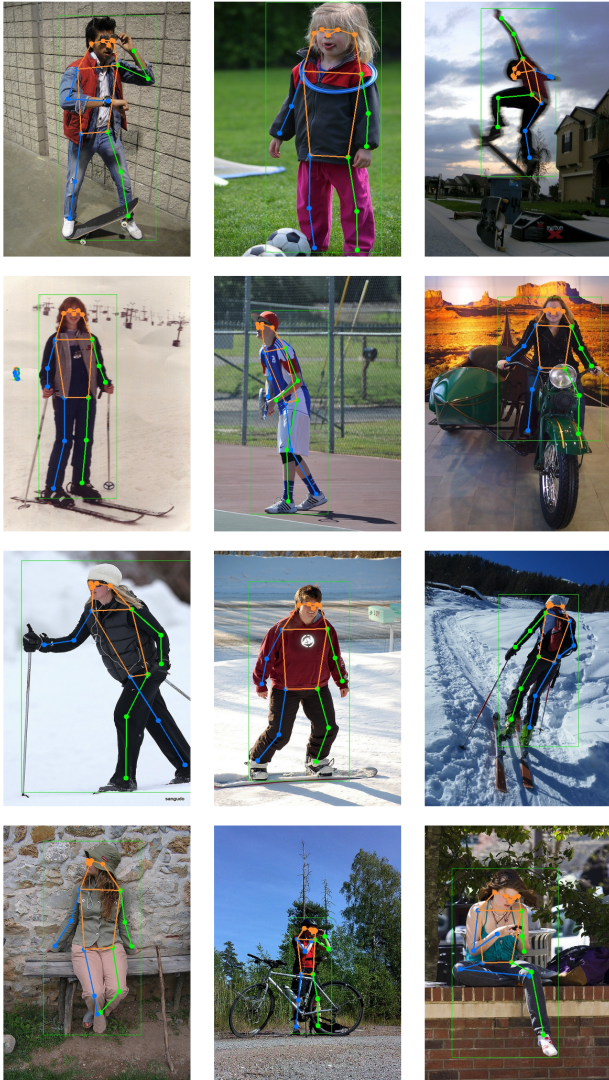
Figure 1. Visualization results of human pose estimation by SDPose-S-V1 on MSCOCO validation dataset.
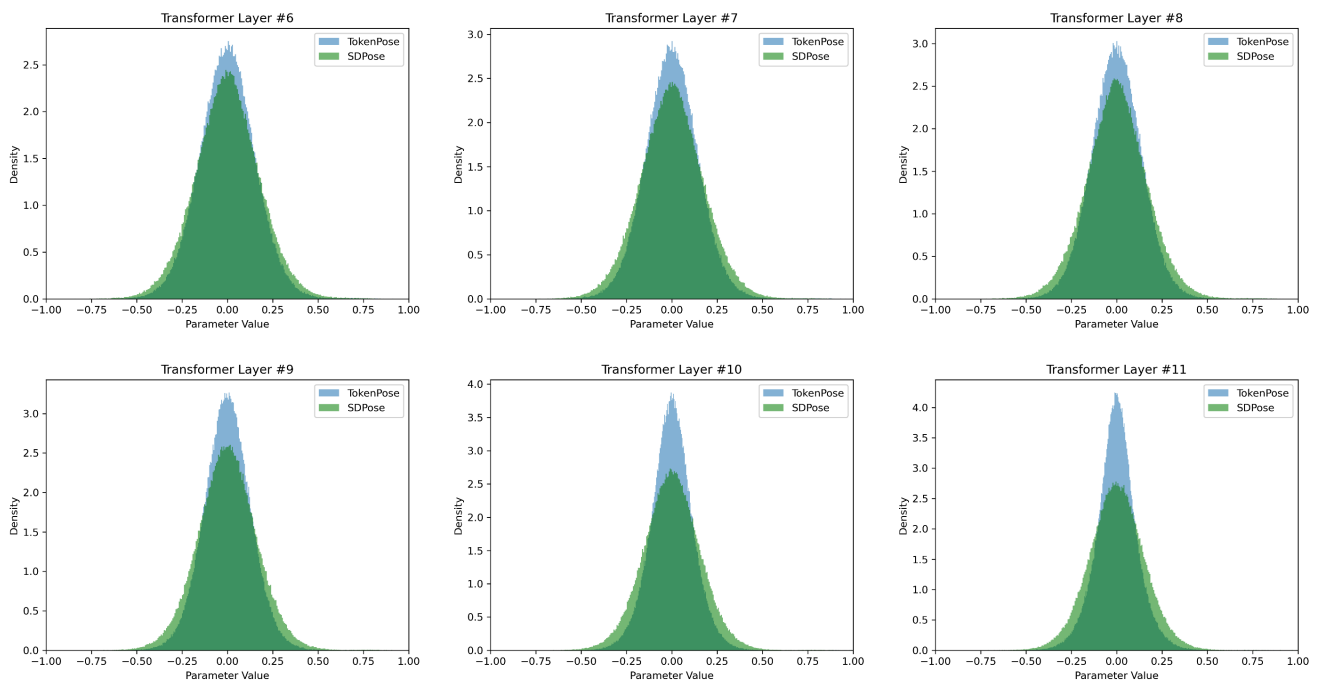
Figure 2. Visualization of parameter distributions for transformer layers # 6 - # 11. The blue represents TokenPose-S-V1 and the green represents SDPose-S-V1.