A Supplementary Material

A.1 Qualitative Results

Figure 1: Qualitative results for sequences of the PoseTrack validation set [0].

A.2 Baseline Improvement

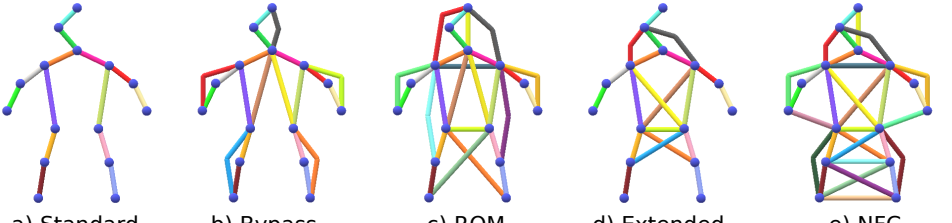

Figure 2: Different edge configurations used for the training of different spatial models.

We evaluate the robustness of different edge configurations as shown in Figure 2. This is motivated by the fact that edge configuration a) is prone to errors. If a single edge is not estimated correctly, the entire pose breaks. Similar to [4] we introduce skip connections to the standard model (Figure 2 b) *Bypass model*). Figure 2 c) illustrates a different idea to connect joints which we refer to as *Range of Motion (ROM) model* since pairs of joints are connected if both lie within the same ROM of a third joint. Further we train an edge configuration as proposed in [2] which we refer to as *Extended model*. For completeness, we introduce a *nearly-fully-connected (NFC) model* (Figure 2 e)) which connects most nearby

Model	VGG Layers	Trained on	Head	Shou	Elb	Wri	Hip	Knee	Ankl	Total mAP
Standard	12	MSCOCO + PoseTrack	82.9	80.3	69.9	**59.0**	67.8	59.2	51.4	68.3
Bypass	12	MSCOCO + PoseTrack	**83.0**	79.2	67.6	**59.0**	66.2	61.2	53.6	68.2
ROM	12	MSCOCO + PoseTrack	82.0	76.2	70.3	57.9	69.3	61.7	54.1	68.3
Extended	**12**	**MSCOCO + PoseTrack**	80.0	**80.8**	**71.3**	57.8	**72.5**	**63.3**	**53.9**	**69.3**
NFC	12	MSCOCO + PoseTrack	78.3	75.8	68.3	56.9	69.2	62.1	53.5	67.1

Table 1: The evaluation of different edge configurations reveals that the *Extended* edge configuration performs best compared to the *Standard* edge configuration.

joints. We rely on the metric proposed in [3] for the estimation of mean average precision (mAP) of all our pose estimation models. Table 1 shows the results, using $\tau_{NSM} = 0.1$. In all other experiments, we use the *Extended model*.

References

[1] M. Andriluka, U. Iqbal, E. Ensafutdinov, L. Pishchulin, A. Milan, J. Gall, and Schiele B. PoseTrack: A benchmark for human pose estimation and tracking. In *CVPR*, 2018.

[2] Eldar Insafutdinov, Mykhaylo Andriluka, Leonid Pishchulin, Siyu Tang, Evgeny Levinkov, Bjoern Andres, and Bernt Schiele. ArtTrack: Articulated Multi-person Tracking in the Wild. In *CVPR*, 2017.

[3] Leonid Pishchulin, Eldar Insafutdinov, Siyu Tang, Björn Andres, Mykhaylo Andriluka, Peter Gehler, and Bernt Schiele. Deepcut: Joint subset partition and labeling for multi person pose estimation. In *CVPR*, 2016.

[4] Xiangyu Zhu, Yingying Jiang, and Zhenbo Luo. Multi-person pose estimation for posetrack with enhanced part affinity fields. Technical report, Samsung Research Beijing, 2017.