

# Simple Baselines for Human Pose Estimation and Tracking

Bin Xiao<sup>1</sup>, Haiping Wu<sup>2 †</sup>, and Yichen Wei<sup>1</sup>

<sup>1</sup> Microsoft Research Asia,

<sup>2</sup> University of Electronic Science and Technology of China  
{Bin.Xiao, v-haiwu, yichenw}@microsoft.com

**Abstract.** There has been significant progress on pose estimation and increasing interests on pose tracking in recent years. At the same time, the overall algorithm and system complexity increases as well, making the algorithm analysis and comparison more difficult. This work provides simple and effective baseline methods. They are helpful for inspiring and evaluating new ideas for the field. State-of-the-art results are achieved on challenging benchmarks. The code will be available at [https://github.com/leoxiaobin/pose\\_pytorch](https://github.com/leoxiaobin/pose_pytorch).

**Keywords:** Human Pose Estimation, Human Pose Tracking

## 1 Introduction

Similar as many vision tasks, the progress on human pose estimation problem is significantly advanced by deep learning. Since the pioneer work in [31, 30], the performance on the MPII benchmark [3] has become saturated in three years, starting from about 80% PCKH@0.5 [30] to more than 90% [22, 8, 7, 33]. The progress on the more recent and challenging COCO human pose benchmark [20] is even faster. The mAP metric is increased from 60.5 (COCO 2016 Challenge winner [9, 5]) to 72.1 (COCO 2017 Challenge winner [6, 9]) in one year. With the quick maturity of pose estimation, a more challenging task of “simultaneous pose detection and tracking in the wild” has been introduced recently [2].

At the same time, the network architecture and experiment practice have steadily become more complex. This makes the algorithm analysis and comparison more difficult. For example, the leading methods [22, 8, 7, 33] on MPII benchmark [3] have considerable difference in many details but minor difference in accuracy. It is hard to tell which details are crucial. Also, the representative works [21, 24, 12, 6, 5] on COCO benchmark are also complex but differ significantly. Comparison between such works is mostly on system level and less informative. About pose tracking, although there has not been much work [2], the system complexity can be expected to further increase due to the increased problem dimension and solution space.

---

Equal contribution.

<sup>†</sup>This work is done when Haiping Wu is an intern at Microsoft Research Asia.































