**Learning to Refine Human Pose Estimation**

**Link-** [**https://arxiv.org/pdf/1804.07909.pdf**](https://arxiv.org/pdf/1804.07909.pdf)

**The 7 W's**

Q1. Despite the large improvements in human pose estimation enabled by the development of convolutional neural networks, there still exist a lot of difficult cases where even the state-of-the-art models fail to correctly localize all body joints.

In this work, we introduce a pose refinement network (PoseRefiner) which takes as input both the image and a given pose estimate and learns to directly predict a refined pose by jointly reasoning about the input-output space.

Q2. It is relevant because it aims to refine or motivates the need for an additional refinement step that addresses these challenging cases and can be easily applied on top of any existing method.

Q3.

* Single-person pose estimation: More recent methods rely on localizing body joints by employing convolutional neural networks (CNNs), which contributed to large improvement in human pose estimation. [43] directly predicts joint coordinates via a cascade of CNN pose regressors, while further improvement in the performance is achieved by predicting heatmaps of each body joint [42, 30] and using very deep CNNs with multi-stage architectures [45].
* Multi-person pose estimation: Top-down approaches [36, 21, 19, 12, 25, 16] employ a person detector and then perform single-person pose estimation for each detected person. These methods highly depend on the reliability of the person detector and are known to have trouble recovering poses of people in close proximity to each other and/or with overlapping body parts.

Q4. Although great progress has been made, the problem remains far from being solved. There still exist a lot of challenging cases, such as person-person occlusions, close proximity of similar looking people, rare body configurations, partial visibility of people and cluttered backgrounds.

Q5.

* We introduce an effective post-processing technique for body joint refinement in human pose estimation tasks, that works on top of any existing human body pose estimation approach. Our proposed pose refinement network is efficient due to its feed-forward architecture, simple and end-to-end trainable.
* We propose a training data augmentation scheme for error correction, which enables the network to identify the erroneous body joint predictions and to learn a way to refine them.
* We show that our refinement model allows to systematically improve over various state-of-the-art methods and achieve top performing results on four different benchmarks.

Q6. With our refinement network we improve the best reported results on MPII Human Pose and PoseTrack datasets for multi-person pose estimation and pose tracking tasks.