**3D Human Pose Estimation Using Convolutional Neural Networks with 2D Pose Information**

**Link-** [**https://arxiv.org/pdf/1608.03075.pdf**](https://arxiv.org/pdf/1608.03075.pdf)

**The 7 W's**

Q1. While there has been a success in 2D human pose estimation with convolutional neural networks (CNNs), 3D human pose estimation has not been thoroughly studied. Estimating a 3D human pose from a single image is more challenging than 2D cases due to the lack of depth information. In this paper, we tackle the 3D human pose estimation task with end-to-end learning using CNNs. Relative 3D positions between one joint and the other joints are learned via CNNs.

Q2. It is relevant because the proposed method improves the performance of CNN with two novel ideas. First, we added 2D pose information to estimate a 3D pose from an image by concatenating 2D pose estimation result with the features from an image. Second, we have found that more accurate 3D poses are obtained by combining information on relative positions with respect to multiple joints, instead of just one root joint.

Q3.

* Early works for 2D human pose estimation which are based on deformable parts model [9], pictorial structure [10,11,12], or poselets [13] train the relationship between body appearance and body joints using hand-crafted features.
* Joao et al. [2] proposed a self-correcting method by a top-down feedback. It iteratively learns a human pose using a self-correcting CNN model which gradually improves the initial result by feeding back error predictions.
* Xiao et al. [15] proposed an end-to-end learning system which captures the relationships among feature maps of joints. Geometrical transform kernels are introduced to learn features and their relationship jointly.

Q4 & Q5. The method proposed in this paper aims to provide an end-to-end learning framework to estimate 3D structure of a human body from a single image. Similar to [5], 3D and 2D pose information are jointly learned in a single CNN. Unlike the previous works, we directly propagate the 2D classification results to the 3D pose regressors inside the CNNs. Using additional information such as 2D classification results and the relative distance from multiple joints, we improve the performance of 3D human pose estimation over the baseline method.

Q6. By reusing 2D joint classification result, the relationship between 2D pose and 3D pose is implicitly learned during the training phase. Moreover, multiple regression results with different root nodes gives an effect of ensemble learning. When both strategies are combined, 3D pose estimation results are significantly improved and showed comparable performance to the state-of-the-art methods without exploiting any temporal information of video sequences.

Q7. We expect that the performance can be further improved by incorporating temporal information to the CNN by applying the concepts of recurrent neural network or 3D convolution [26]. Also, efficient aligning method for multiple regression results may boost the accuracy of pose estimation.