**Monocular Human Pose Estimation: A Survey of Deep Learning-based Methods**

**Link-** [**https://arxiv.org/pdf/2006.01423.pdf**](https://arxiv.org/pdf/2006.01423.pdf)

**The 7 W's**

Q1. This survey extensively reviews the recent deep learning-based 2D and 3D human pose estimation methods published since 2014. This paper summarizes the challenges, main frameworks, benchmark datasets, evaluation metrics, performance comparison, and discusses some promising future research directions.

Q2. It is relevant because the recent developments of deep learning techniques have been brought significant progress and remarkable breakthroughs in the field of human pose estimation. Although there are some existing reviews for HPE, however, there still lacks a survey to summarize the most recent deep learning-based achievements. This paper extensively reviews deep learning-based 2D/3D human pose estimation methods from monocular images or video footage of humans.

Q3. The surveys (Aggarwal and Cai, 1999; Gavrila, 1999; Poppe, 2007; Ji and Liu) reviewed the early work of human motion analysis in many aspects (e.g., detection and tracking, pose estimation, recognition) and described the relation between human pose estimation and other related tasks. More recent surveys were mainly focusing on relatively narrow directions, such as RGB-D-based action recognition, 3D HPE, model-based HPE, body parts-based HPE, and monocular-based HPE.

Q4. There still lacks a survey to summarize the most recent deep learning-based achievements. Due to limitations in hardware device capability and the quantity and quality of training data, early networks are relatively shallow, used in a very straightforward way and can only handle small images or patches

Q5. In this paper, we have reviewed the recent deep learning-based research addressing the 2D/3D human pose estimation problem from monocular images or video footage and organize approaches into four categories based on specific tasks: (1) 2D single person pose estimation, (2) 2D multi-person pose estimation, (3) 3D single person pose estimation, and (4) 3D multi-person pose estimation. Further, we have summarized the popular human pose datasets and evaluation protocols.

Q6. Different from existing review papers, this survey extensively summarizes the recent milestone work of deep learning-based human pose estimation methods, which were mainly published from 2014. We further describe HPE approaches for both single-person pose estimation and multi-person pose estimation. Since data are a very important and fundamental element for deep learning-based methods, the recent HPE datasets and the evaluation metrics are summarized

Q7. Efficient networks and adequate training data are the most important requirements for deep learning-based approaches. Future networks should explore both global and local contexts for more discriminative features of the human body while exploiting human body structures into the network for prior constraints. Diversity data can improve the robustness of networks to handle complex scenes with irregular poses, occluded body limbs and crowded people.