**Deep Learning-Based Human Pose Estimation: A Survey**

**Link-** <https://arxiv.org/pdf/2012.13392.pdf>

**The 7 W's**

Q1. Given the rapid progress in HPE research, this article attempts to track recent advances and summarize their achievements in order to provide a clear picture of current research on deep learning-based 2D and 3D HPE.

Q2. It is relevant because this survey paper provides a comprehensive review of recent deep learning-based solutions for both 2D and 3D pose estimation via a systematic analysis and comparison of these solutions based on their input data and inference procedures.

Q3.

* [4] [5] [6] [7] focus on the general field of visual-based human motion capture methods and their implementations including pose estimation, tracking, and action recognition. Therefore, pose estimation is only one of the topics covered in these surveys.
* A survey on both traditional and deep learning-based methods related to HPE is presented in [10]. However, only a handful of deep learning-based approaches are included.
* The survey in [13] only reviews 2D HPE methods and analyzes model interpretation. Monocular HPE from the classical to recent deep learning-based methods (till 2019) is summarized in [12]. However, it only covers 2D HPE and 3D single-view HPE from monocular images/videos. Also, no extensive performance comparison is given.

Q4. Although the recently developed deep learning-based solutions have achieved high performance in human pose estimation, there still remain challenges due to insufficient training data, depth ambiguities, and occlusion.

Q5&6.

* A comprehensive review of recent deep learning-based 2D and 3D HPE methods (up to 2020) is provided by categorizing them according to 2D or 3D scenario, single view or multi-view, from monocular images/videos or other sources, and learning paradigm.
* Extensive performance evaluation of 2D and 3D HPE methods. We summarize and compare reported performances of promising methods on common datasets based on their categories. The comparison of results provides cues for the strength and weakness of different methods, revealing the research trends and future directions of HPE.
* An insightful discussion of 2D and 3D HPE is presented in terms of key challenges in HPE pointing to potential future research towards improving performance.
* An overview of a wide range of HPE applications, such as gaming, surveillance, AR/VR, and healthcare.

These contributions make our survey more comprehensive, up-to-date, and in-depth than previous survey papers.

Q7.

* The recent trend to alleviate the domain gap is utilizing GAN-based learning approaches. Nonetheless, how to effectively transfer the human pose knowledge to bridge domain gaps remains unaddressed.
* Most existing methods ignore human interaction with 3D scenes. There are strong human-scene relationship constraints that can be explored such as a human subject cannot be simultaneously present in the locations of other objects in the scene. The physical constraints with semantic cues can provide reliable and realistic 3D HPE.
* Human body models such as SMPL, SMPLify, SMPLX, GHUM & GHUML, and Adam are used to model human mesh representation. However, these models have a huge number of parameters. How to reduce the number of parameters while preserving the reconstructed mesh quality is an intriguing problem.