**Human Pose Estimation with Iterative Error Feedback**

**The 7 W's**

Q1. Our main contribution is in providing a generic frame- work for modeling rich structure in both input and output spaces by learning hierarchical feature extractors over their joint space. We achieve this by incorporating top-down feedback

Q2. It is relevant because it shows excellent performance on the task of articulated pose estimation in the challenging MPII and LSP benchmarks, matching the state-of-the-art without requiring ground truth scale annotation*.*

Q3. In concurrent work, Oberweger et al [27] proposed a feedback loop for hand pose estimation from kinect data that is closely related to our approach. The autocontext work of [41] is also related and iteratively computes label heatmaps by concatenating the image with the heatmaps previously predicted.

Q4. Previous works fails to cater to the needs of a simple full-fledged HPE model with an Iterative Error Feedback (IEF).

Q5. We propose a framework that expands the expressive power of hierarchical feature extractors to encompass both input and output spaces, by introducing top-down feedback. Instead of directly predicting the outputs in one go, we use a self-correcting model that progressively changes an initial solution by feeding back error predictions, in a process we call Iterative Error Feedback (IEF).

Q6. This solution is better because it provides high performance on MSPII and LSP benchmarks without ground truth scale annotation alongside matching state-of-the-art models.

Q7. We have only experimented so far feeding back "images” made up of Gaussian distributions. There may be more powerful ways to render top- down pose information using parametrized computational blocks (e.g., deconvolution) that can then be learned jointly with the rest of the model parameters using standard back- propagation.