Qeios

Peer Review

Review of: "Optimizing Human Pose Estimation Through Focused Human and Joint Regions"

Israr Akhter[1]

1. Bahria University, Pakistan

Clarify the Scope of the Challenge

Specify whether the method targets 2D or 3D human pose estimation explicitly.

Mention any specific limitations in the datasets used for evaluation.

Improve the Problem Statement

Instead of stating that existing methods use "all pixels," quantify the issue with numerical metrics or examples.

Clearly define the impact of background distractions on accuracy and precision.

Enhance the Methodological Contributions

Provide more technical details on how the bilayer HumanKeypoint Mask module refines tokens. Does it employ multi-scale feature aggregation?

Explain how the deformable cross attention mechanism improves spatial and temporal reasoning. How does it compare computationally to standard attention mechanisms?

Describe the mathematical constraints applied in deformable cross attention—how do they differ from previous methods?

Benchmarking and Performance Analysis

Justify the choice of datasets—why were they selected, and how do they represent real-world scenarios?

Instead of just reporting mAP improvements, analyze improvements in specific scenarios (e.g., occlusions, complex backgrounds, multi-person interactions).

Mention computational efficiency: Does the proposed method improve inference speed or reduce model

doi.org/10.32388/48YMoK

complexity?

Improve the Writing Style

Replace vague phrases like "a remarkable highlight" with more technical phrasing, such as "our approach demonstrates superior performance, particularly on fine-grained joint localization, achieving X% improvement on wrist detection."

Ensure consistency: "methods has demonstrated" → "methods have demonstrated."

Declarations

Potential competing interests: No potential competing interests to declare.

doi.org/10.32388/48YMoK