

Peer Review

Review of: "Reconstructing In-the-Wild Open-Vocabulary Human-Object Interactions"

Matej Džijan¹

1. University of Osijek, Croatia

This paper is well-written and presents meaningful contributions to 3D Human-Object Interaction (HOI) research. The authors introduce a valuable dataset that may support future work in this area. They also propose an iterative method for reconstructing HOI from single images and demonstrate that existing general-purpose 3D models struggle with 3D HOI understanding.

I have only a few minor suggestions for improvement:

- 1) The following part of the introduction is awkwardly phrased: "In detail, we built our annotation pipeline on two bases: 1) Existing 2D HOI datasets provide rich 2D annotations, including bounding boxes, and a wide variety of objects and actions. This diversity creates the potential for reconstructing 3D assets from 2D HOI images. 2) The development of existing image-based 3D object/human reconstruction techniques."
- 2) Figure 4 is difficult to interpret since it provides some information about the dataset, but there are many categories that are not named and are just shown to exist. Some other way of visualizing this might be more appropriate. However, the full list of object classes is given in the appendix, which helps mitigate the issue.
- 3) In Subsection 3.3. Open3DHOI Dataset Construction, the sentence "It can be seen that the object and action categories in the current benchmarks are much fewer than those in our dataset." sounds awkward. A clearer alternative might be: "The current benchmarks contain significantly fewer object and action categories than our dataset."

- 4) In Subsection 4.1. 3D Gaussian Splatting, the sentence “3D Gaussian Splatting is advanced for rendering and reconstructing scenes by representing objects continuously, in a volumetric manner.” feels grammatically off. Consider rephrasing as: “3D Gaussian Splatting is an advanced method...” for clarity.
- 5) In Subsection 4.2. HOI-Gaussian Optimizer, “thus reducing reliance on prior”, should be corrected to “thus reducing reliance on priors”
- 6) In Subsection 4.3. Contact in Gaussian Model, when explaining Figure 6, you state the following: “In Fig. 6 (a) and (b), the blue area represents regions with high opacity, while the red area starts with relatively high opacity,” which seems confusing. If both red and blue indicate high opacity, the distinction isn’t clear. Perhaps clarify what distinguishes the red areas from the blue — or confirm if red actually denotes lower opacity? Furthermore, the rest of the explanation “As the optimization progresses, occluded parts of the human leg turn blue, designating them as potential contact regions. Meanwhile, the opacity of points behind the camera remains relatively stable.” is unclear. A more precise explanation would help here.
- 7) There is inconsistency in the spelling of “GauHuman” — it appears both as “Gauhuman” and “GauHuman”. This should be standardized.

Overall, this is a solid contribution to the field. These are primarily comments about clarity and presentation, and I believe they can be resolved with minor revisions.

Declarations

Potential competing interests: No potential competing interests to declare.