The 38th Annual AAAI Conference on Artificial Intelligence

Learning Explicit Contact for Implicit Reconstruction of Hand-Held Objects from Monocular Images

Junxing Hu[1,2], Hongwen Zhang[3], Zerui Chen[4], Mengcheng Li[5], Yunlong Wang[2], Yebin Liu[5], Zhenan Sun[1,2]

[1]University of Chinese Academy of Sciences, [2]CRIPAC, MAIS, CASIA, [3]Beijing Normal University,

[4]Inria, DI ENS, CNRS, PSL Research University, [5]Tsinghua University

Our code, data, and video results can be found here!

Motivation

- There are few methods to predict contact between the hand and objects. Some approaches require the 3D object model during estimation, while others are designed for body-scene interaction without the hand and object.
- For implicit reconstruction of hand-held objects, current model-free methods ignore formulating contacts in their frameworks or only use contact loss like attraction and repulsion loss to model the hand-object interaction.

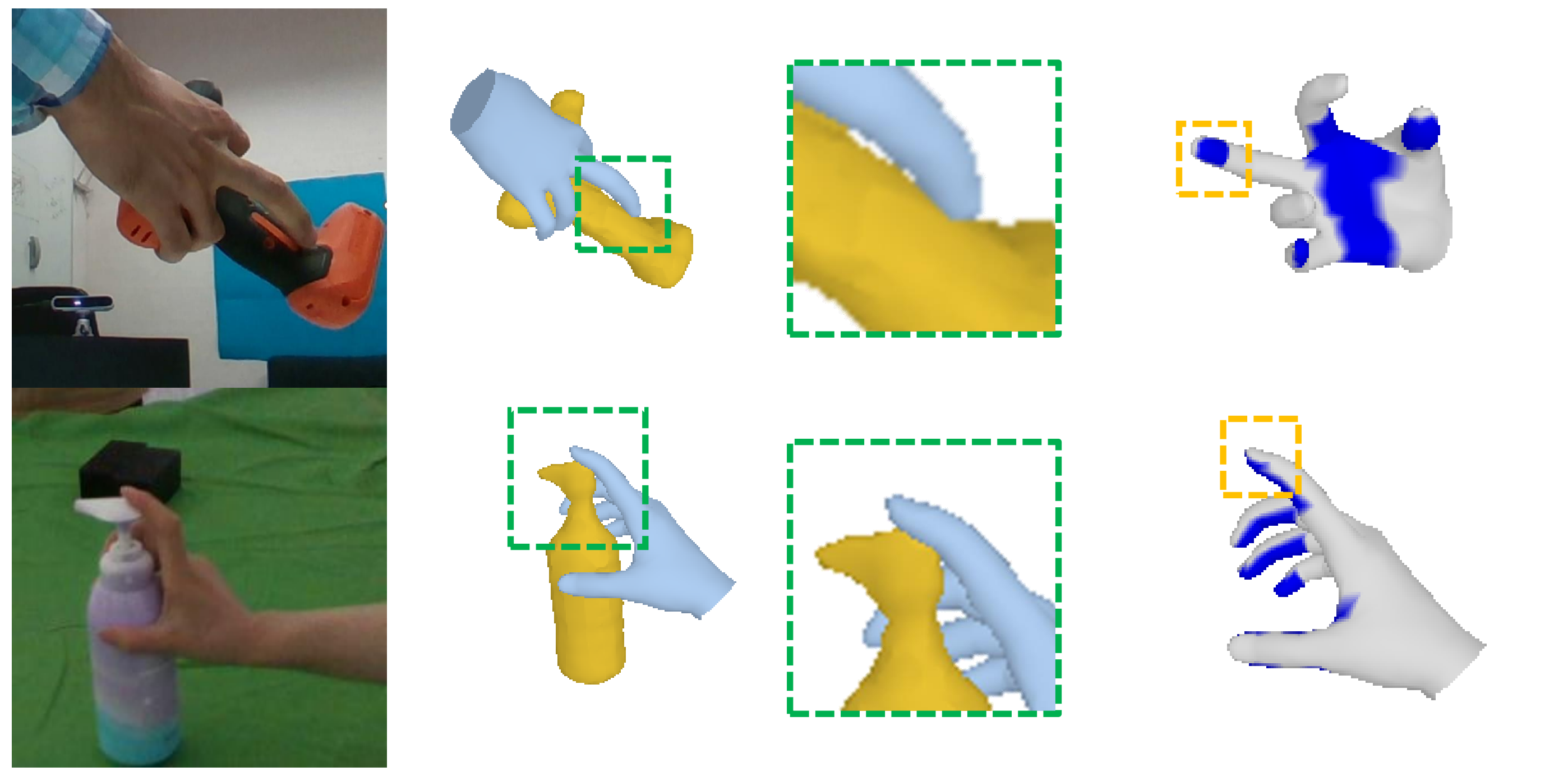


Image　　Object Reconstruction　　Contact Prediction

Contribution

- We propose to leverage contact priors for better reconstruction of hand-held objects. To estimate contact states more accurately, we introduce a novel framework that jointly improves part-level and vertex-level contact states in a coarse-to-fine manner.
- To make discrete contact states compatible with continuous implicit shape functions, we propose to diffuse contact features from the hand mesh surface to the whole 3D volume, which enables the continuous query of contact features for implicit object reconstruction.

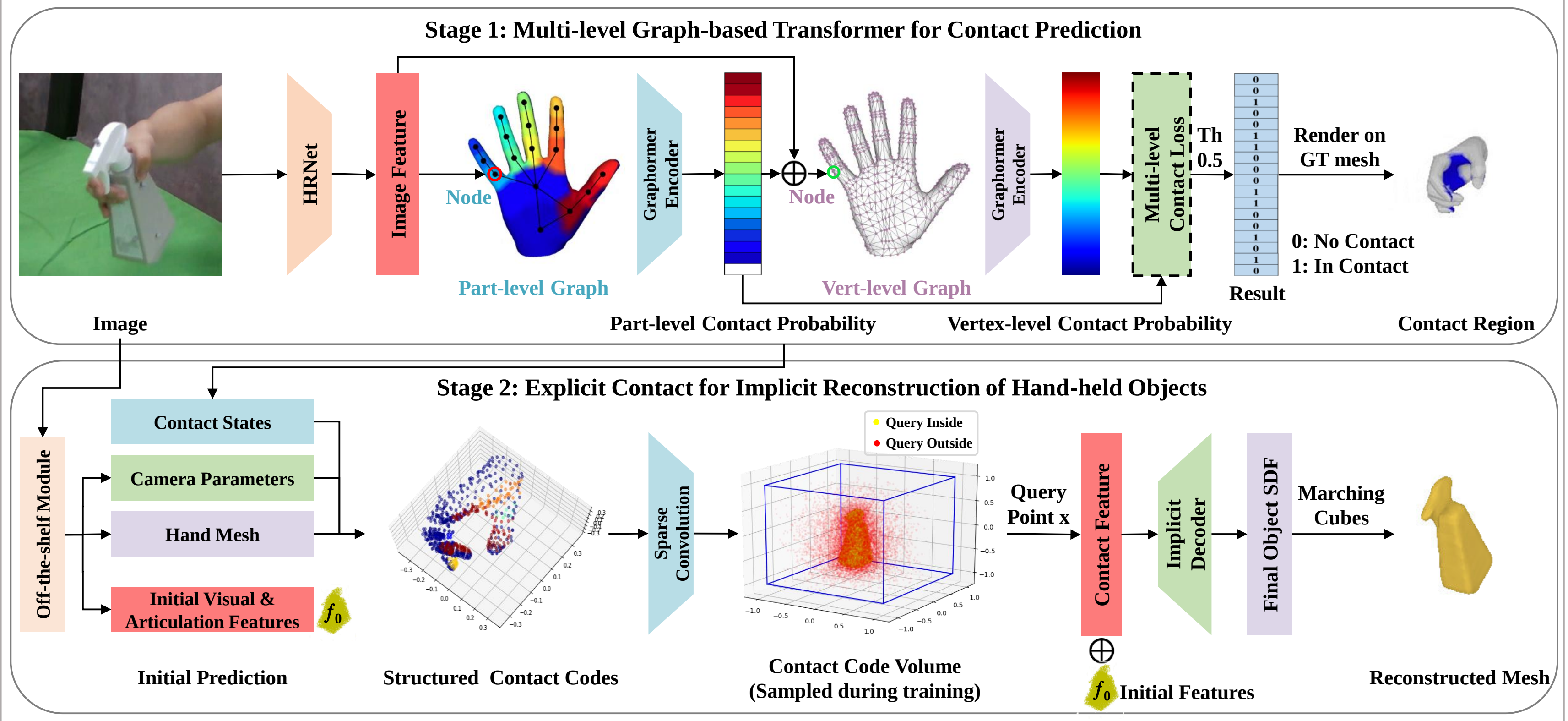
Related Work

[1] Hasson et al. Learning joint reconstruction of hands and manipulated objects. In Proceedings of CVPR, 2019.

[2] Karunratanakul et al. Grasping Field: Learning Implicit Representations for Human Grasps. In Proceedings of 3DV, 2020.

[3] Ye et al. What's in your hands? 3D Reconstruction of Generic Objects in Hands. In Proceedings of CVPR, 2022.

Method

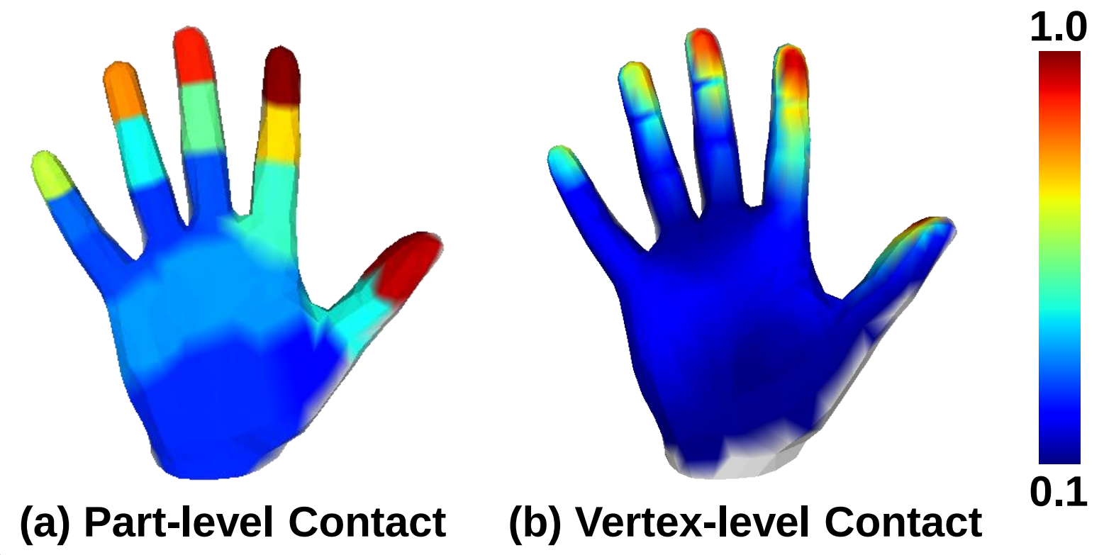


Contact Estimation

- **Multi-level Contact Graphs.** The part-level graph has 18 nodes relying on a coarse division of the hand regions, while the vertex-level graph is generated based on the template MANO mesh with 778 nodes.
- **Graph-based Transformer.** They are utilized to combine local interactions and global relationships of graph nodes. The contact probabilities are normalized to [0, 1], and the points greater than 0.5 are extracted.

Explicit Contact for Implicit Object Reconstruction

- **Structured Contact Codes.** The structured contact codes are generated by anchoring predicted contact states to the hand mesh surface. In the context of implicit reconstruction, the trilinear interpolation is performed on the estimated contact probabilities according to the contact point's position.
- **Contact Code Volume.** Since the implicit functions have continuous values in the 3D volume, the sparse convolutions are utilized to diffuse the discrete contact states to the continuous space as multiple contact code volumes. After that, the contacts on the hand surface are diffused to its nearby 3D space, facilitating the perception and reconstruction of the hand-held object.



(a) Part-level Contact　(b) Vertex-level Contact

- **Implicit Decoding.** The SDF value can be computed by the decoder conditioned by the contact feature.
- **Contact Frequency.** They are used as the weighted priors.

Results

- **Datasets and Metrics.** The methods are validated on HO3D and OakInk benchmarks. The metrics are precision, recall, and F1-score for contact prediction, while F@5/10mm, chamfer distance (CD, mm), penetration depth (PD, cm), and intersection volume (IV, cm^3) for object reconstruction.
- **The Contact Prediction is Positively Correlated with the Object Reconstruction.** The models are trained and evaluated multiple times with different seeds. The mean and standard deviation are reported.

Contact Prediction and Object Reconstruction

Method	Precision↑	Recall↑	F1↑	F@5mm↑	F@10mm↑	Chamfer Distance (mm)↓
BSTRO	0.467	0.416	0.400	0.363/0.007	0.607/0.010	0.764/0.035
Single Vertex-level	0.476 (1.9%↑)	0.422 (1.4%↑)	0.416 (4.0%↑)	0.371 (2.2% ↑)/0.005	0.615 (1.3%↑)/0.007	0.739 (3.3%↓)/0.013
Multi-level (Vertex output)	**0.510** (7.1% ↑)	**0.441** (4.5%↑)	**0.436** (4.8% ↑)	**0.374** (0.8% ↑)/0.004	**0.620** (0.8%↑)/0.006	**0.701** (5.1%↓)/0.059

Quantitative Comparisons on HO3D and OakInk

Method	F@5mm	F@10mm	CD	PD↓	IV↓
HO	0.110	0.220	4.190	-	-
GF	0.120	0.240	4.960	-	-
IHOI	0.280	0.500	1.530	-	-
Ours	**0.313**	**0.542**	**1.081**	**1.02**	**5.11**
IHOI*	0.351	0.600	0.656	0.90	4.10
Ours*	**0.393**	**0.633**	**0.646**	**0.67**	**2.91**

Method	F@5mm	F@10mm	CD	PD	IV
IHOI	0.432	0.658	0.491	0.75	4.36
Ours w/o $\mathcal{L}_{hoi}$	0.447	0.716	0.274	0.66	3.03
Ours	**0.459**	**0.718**	**0.260**	**0.62**	**2.67**

- The left table shows the results on HO3D, while the right one for OakInk.

Qualitative Comparisons on HO3D and OakInk

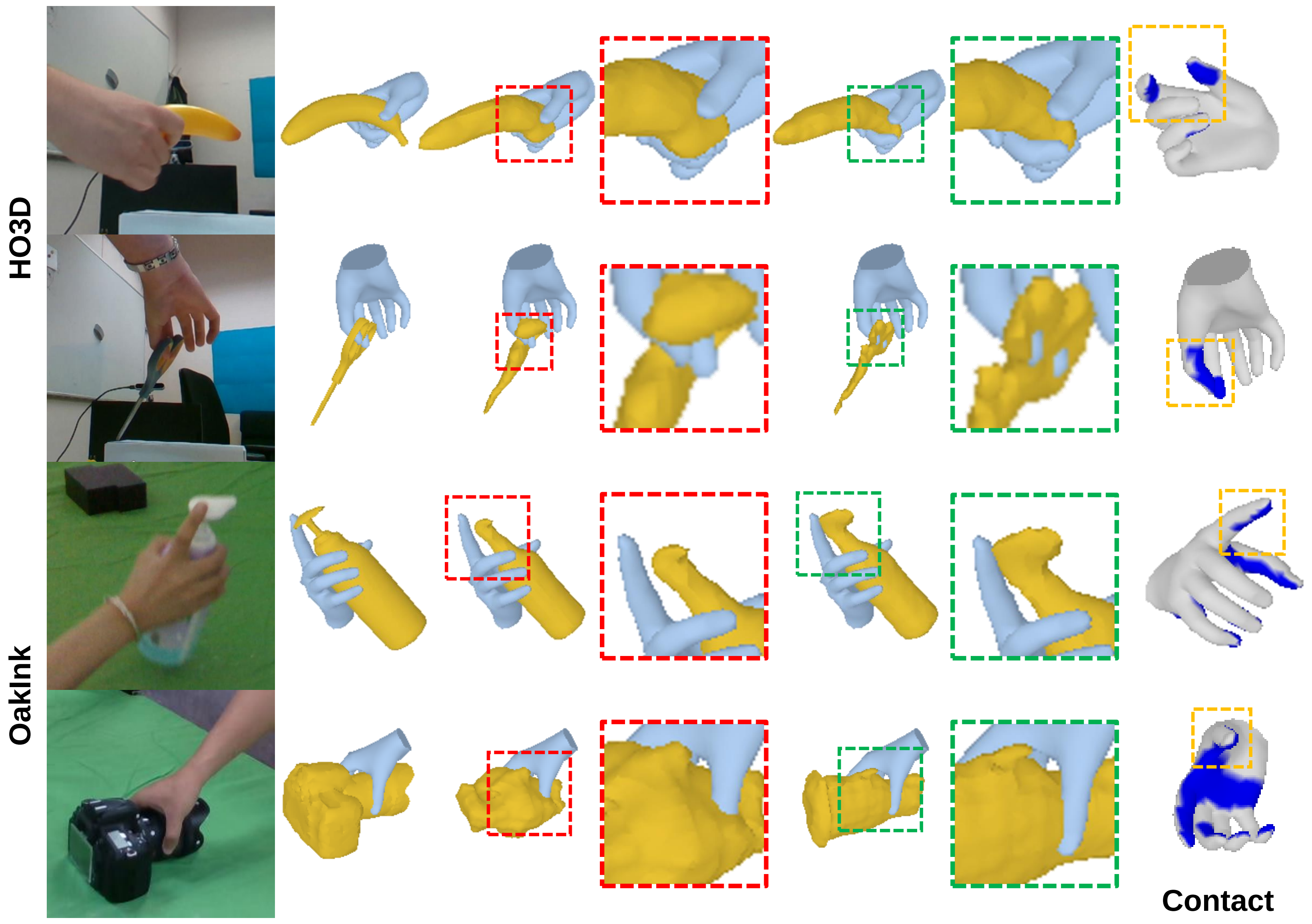


Image　GT　IHOI　Ours　Contact Prediction