



EgoHandICL: Egocentric 3D Hand Reconstruction with In-Content Learning

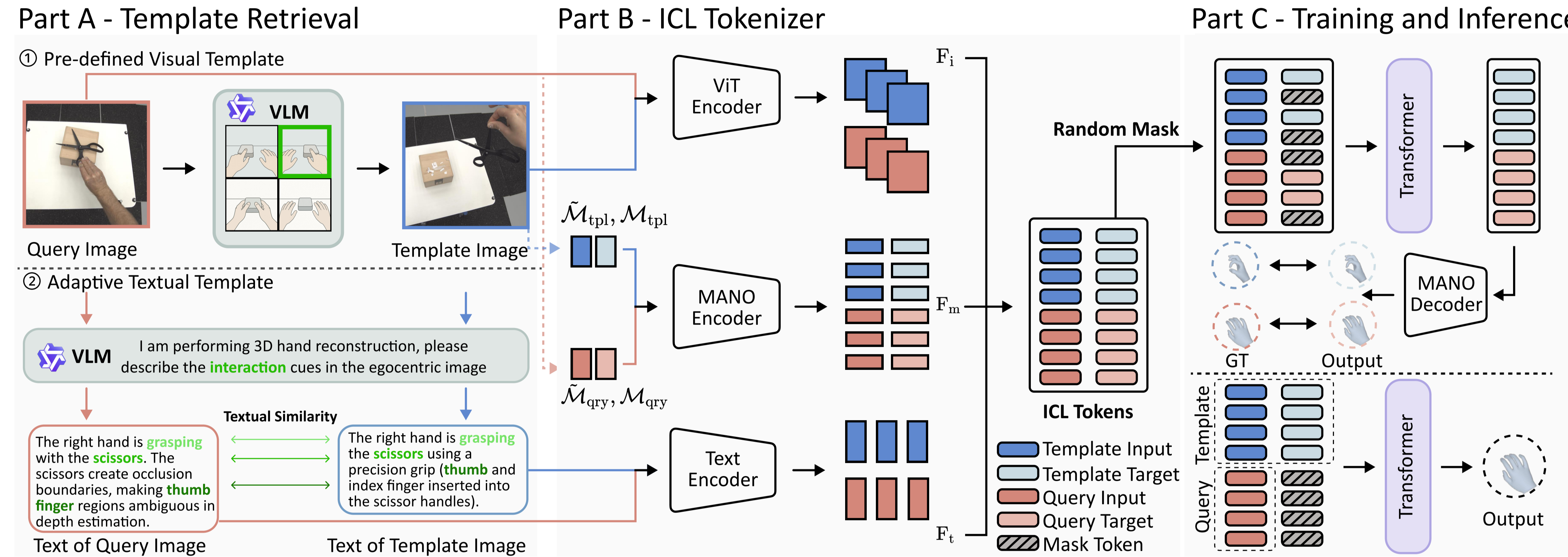
Binzhu Xie *, Shi Qiu *, Sicheng Zhang, Yinqiao Wang, Hao Xu, Muzammal Nasser, Chi-Wing Fu, Pheng-Ann Heng



TL;DR

- **The Bottleneck:** Egocentric 3D hand reconstruction suffers from severe self-occlusion, depth ambiguity, and complex hand-object interactions. Prior methods rely on auxiliary cues but still fail under challenging egocentric scenarios.
- **Core Insight:** In-context learning (ICL) enables example-guided reasoning by conditioning on relevant exemplars.
- **The Solution:** We propose **EgoHandICL**, the first ICL framework for 3D hand reconstruction.
- **Key Results:** Achieves state-of-the-art on ARCTIC and EgoExo4D benchmarks, improving P-MPVPE by **31.1%** (general) and **24.5%** (bimanual) on ARCTIC, while reducing MRRPE by **12%**. Also boosts EgoVLMs' hand-object interaction reasoning.

EgoHandICL Framework



Qualitative results

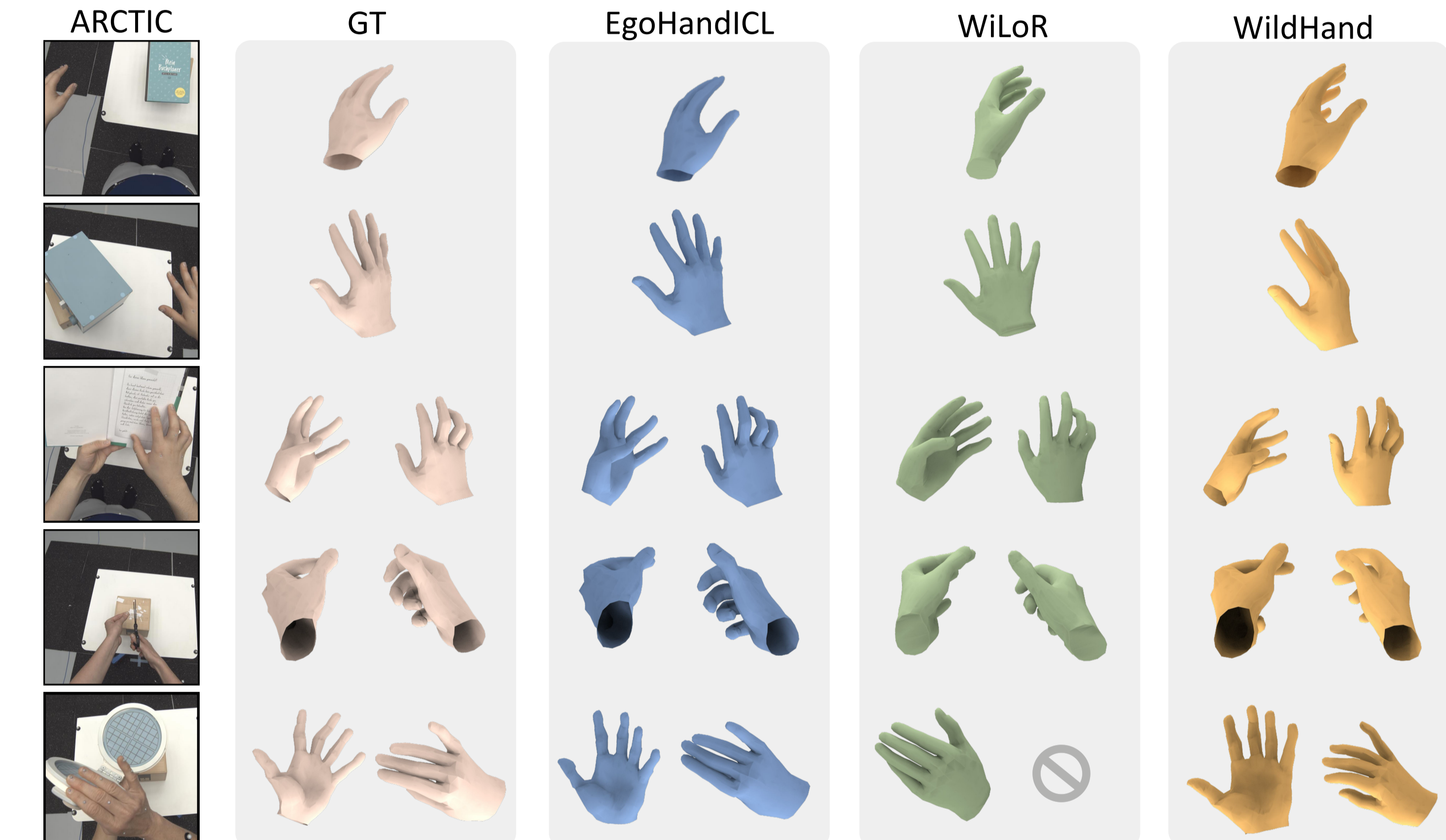
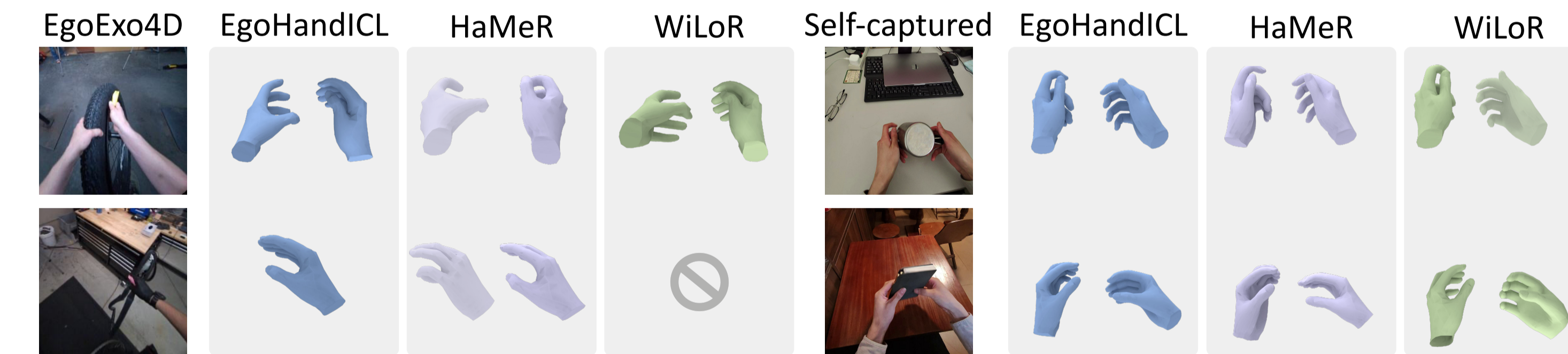


Table 1: **Quantitative results on the ARCTIC dataset.** We follow the standard evaluation protocol and report both the joint- and vertex-level metrics.

Method	General Setting				Bimanual Setting	
	P-MPJPE↓	P-MPVPE↓	F@5↑	F@15↑	P-MPVPE↓	MRRPE↓
HaMeR (Pavlakos et al., 2024)	9.9	9.6	0.046	0.911	9.9	10.1
WiLoR (Potamias et al., 2025)	5.5	5.5	0.524	0.994	5.7	9.8
WildHand (Prakash et al., 2024)	5.8	5.6	0.746	0.928	4.9	7.1
HaWoR (Zhang et al., 2025b)	6.2	6.1	0.474	0.896	6.0	8.6
EgoHandICL (ours)	4.0	3.8	0.801	0.996	3.7	6.2

Table 3: **In-context reasoning analysis across different hand-involvement types.** L, R, T, and N denote training on the left-hand, right-hand, two-hand, and non-hand involvement type sub-dataset, respectively. Results are tested on the ARCTIC under these four sub-dataset divisions.

Type	Left Hand		Right Hand		Two Hands		Non Hand	
	P-MPJPE↓	P-MPVPE↓	P-MPJPE↓	P-MPVPE↓	P-MPJPE↓	P-MPVPE↓	P-MPJPE↓	P-MPVPE↓
Proposed - L	4.6	4.4	4.6	4.5	5.1	4.8	5.2	5.1
Proposed - R	4.8	4.5	4.1	4.4	5.0	4.8	5.2	5.2
Proposed - T	4.7	4.6	4.9	4.4	4.3	4.2	5.0	5.1
Proposed - N	4.9	4.7	5.3	5.1	5.5	5.1	5.0	5.0
Proposed - Full	4.5	4.3	4.0	3.9	3.9	3.7	4.7	4.5



Q: Please understand the characters' hands and objects in the egocentric video.



... The right hand performs the primary task of **mixing the rice**. ... The right hand uses a utensil to interact with the rice, **moving it in circular motions to ensure thorough mixing**.

with visual prompts ... by EgoHandICL.

... Both hands appear to be working in tandem, ... The right hand is the primary actor, using its fingers to **press and shape** the rice within the bowl. ... **the right hand would likely have been moving rhythmically, pressing and shaping the rice.**