中国科学院大学
University of Chinese Academy of Sciences

中国科学院自动化研究所
INSTITUTE OF AUTOMATION CHINESE ACADEMY OF SCIENCES

中国科学院香港创新研究院 人工智能与机器人创新中心
Centre for Artificial Intelligence and Robotics
Hong Kong Institute of Science & Innovation, Chinese Academy of Sciences

PEKING UNION MEDICAL COLLEGE HOSPITAL 1921

MICCAI2025 Daejeon REPUBLIC OF KOREA

Reconstructing 3D Hand-Instrument Interaction from a Single 2D Image in Medical Scenes

Miao Xu, Xiangyu Zhu, Jinlin Wu, Ming Feng, Zelin Zang, Hongbin Liu, Zhen Lei*
{miao.xu, jinlin.wu, hongbin.liu, dong.yi}@cair-cas.org.hk, fengming@pumch.cn, {xiangyu.zhu, zhen.lei}@ia.ac.cn,
*Corresponding Author

Background

Most HOI research focuses on daily object manipulation, where objects are textured and easy to track. In surgical scenes, instruments are textureless, reflective, and elongated, making hand reconstruction unstable and prone to penetration. Accurate modeling is essential for surgical skill assessment, yet current methods and datasets rarely address these challenges.

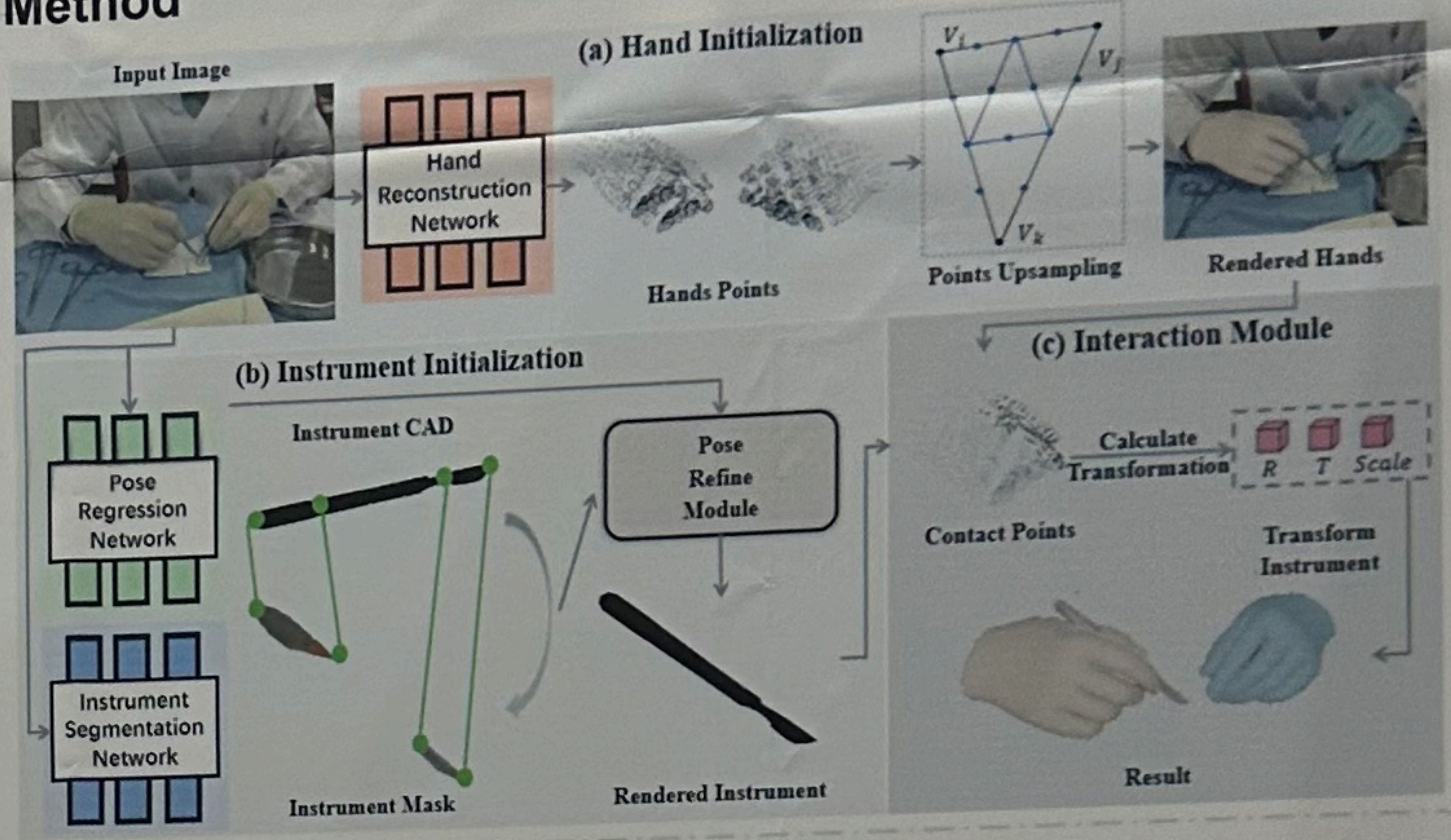
Motivation

Accurate modeling of hand–instrument interaction is a key foundation for behavioral analysis and surgical skill assessment. However, medical scenarios pose unique challenges that make existing HOI methods insufficient:

- Textureless & reflective instruments → vision-based pose estimation often fails.
- Thin and elongated shapes → combined with sparse hand reconstructions, this leads to unstable grasps and penetration.
- High-stakes medical context → even small errors in geometry can misrepresent surgical actions.

These challenges highlight the need for a framework specifically designed for robust hand–instrument interaction reconstruction in surgical environments.

Method



CPCI:

1. MedIns-3D: A 3D model library of real surgical instruments, used as geometric priors to refine instrument pose estimation.

2. Hand Initialization: Dual-hand reconstruction with MANO model, enhanced by fingertip surface upsampling to refine local geometry.

3. CPCI (Contact-Point-Centered Interaction): Aligns hands and instruments around contact points in a unified 3D space, ensuring stable grasping and penetration-free interaction.

Experiments

· **Our method achieves the best results on the POV-Surgery.**

Method	IV↓	Penetration Depth↓	Proximity Error↓	HO Motion Consistency↓
HandOCCNet [19]	2.31	2.60	3.08	21.35
HandOCCNet+TOCH [31]	3.04	2.19	3.14	4.42
Hasson *et al.* [9]	1.96	2.01	3.19	7.26
Hasson *et al.* [10]	1.78	1.85	2.98	4.13
HOISDF [22]	1.52	1.65	2.87	1.85
CPCI w/o UP (ours)	1.14	1.63	2.23	0.50
CPCI (ours)	**1.12**	**1.62**	**2.21**	**0.48**

· **We also achieve the lowest average error of hand reconstruction.**

Method	P_{2d}↓	MPJPE↓	PVE↓	PA-MPJPE↓	PA-PVE↓
METRO [15]	30.49	14.90	13.80	6.36	4.34
HandTailor [18]	25.42	13.20	12.48	5.89	4.19
Mesh Graphormer [16]	20.36	12.75	12.68	5.46	4.32
WiLoR [21]	18.48	13.72	12.91	4.33	4.20
SimpleHand [32]	16.52	13.45	12.61	4.32	4.19
SEMI [17]	13.42	15.14	14.69	4.29	4.23
HandOCCNet [19]	13.80	14.35	13.73	4.49	4.35
HaMeR [20]	13.05	13.15	12.55	4.41	**4.18**
CPCI (ours)	**12.08**	**12.21**	**12.25**	**4.21**	4.20

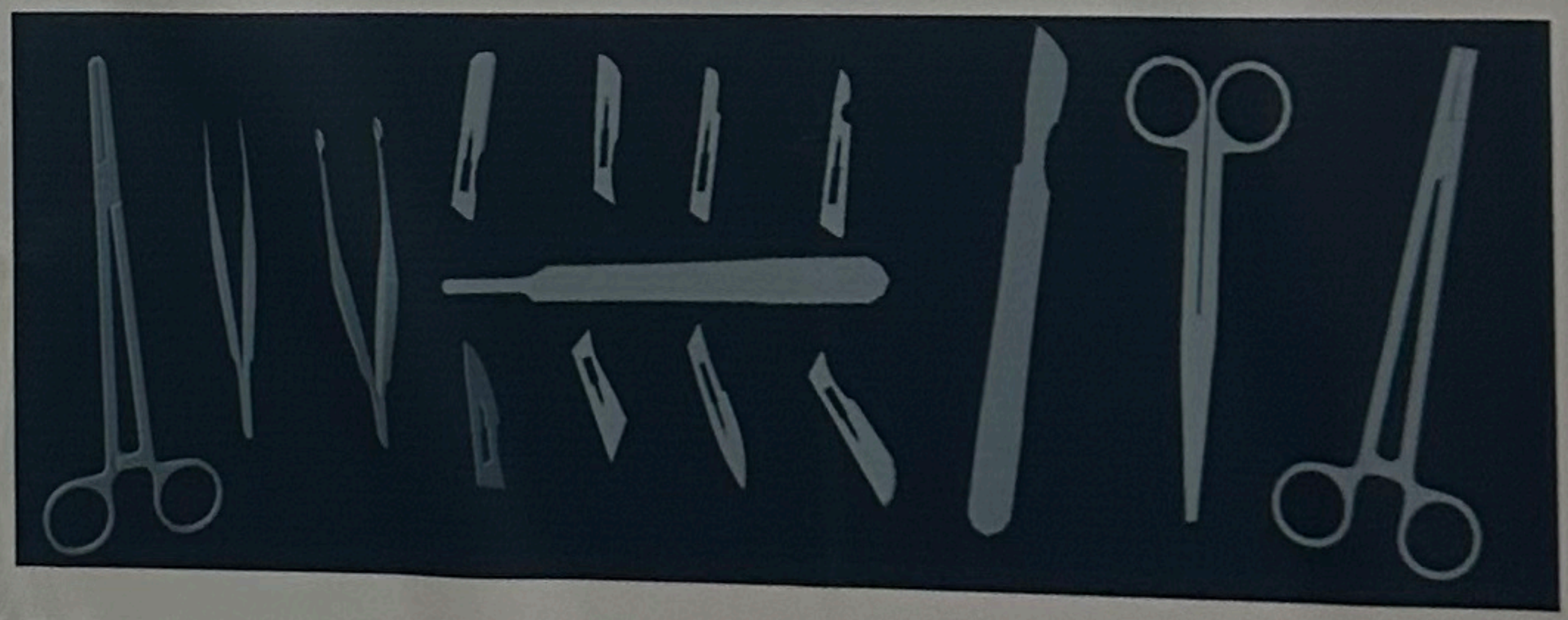
Metrics:

- **Hand rec[illegible] MPJPE, PVE, PA-MPJPE, PA-PVE.**
- **Intera[illegible]ality: IV, Penetration Depth, Proximity Error, HO Mo[illegible]stency.**

[illegible]

- **[illegible] (no upsampling): Using only original MANO 778 vertices [illegible]arse contact points, unstable grasping, larger errors.**
- **[illegible] + UP (ours): Fingertip geometry refinement → significantly [illegible]ed stability and accuracy in hand–instrument interactions.**

· MedIns-3D



- Examples of the 3D instrument models we have constructed, which are based on real medical instruments and include detailed texture.

· Qualitative Comparison on In-the-wild video