HandGCAT: Occlusion-Robust 3D Hand Mesh Reconstruction from Monocular Images

Shuaibing Wang, Shunli Wang, Dingkang Yang, Mingcheng Li,
Ziyun Qian, Liuzhen Su, Lihua Zhang

Academy for Engineering and Technology, Fudan University, Shanghai, China

Introduction

- **Background**: 3D hand mesh reconstruction from monocular images is a crucial yet challenging task, as hands are often severely occluded by objects.
- **Motivation**: Previous works often have disregarded essential 2D hand pose information, which contains hand prior knowledge that is strongly correlated with occluded regions.
- **Contributions**: We propose a novel 3D hand mesh reconstruction network HandGCAT, that can fully exploit hand prior as compensation information to enhance occluded region features. Extensive experimental results show that our method achieves state-of-the-art performance on 3D hand mesh benchmarks that contain severe occlusions.

Performance

Results on HO3D Datasets

TABLE I
COMPARISON WITH STATE-OF-THE-ART METHODS ON HO3D V2.

Method	PA-MPJPE ↓	PA-MPJPE AUC ↑	PA-MPVPE ↓	PA-MPVPE AUC ↑	F@5 ↑	F@15 ↑
I2L-MeshNet [32] (CVPR'20)	1.12	0.775	1.39	0.722	0.409	0.932
Hasson *et al.* [33] (CVPR'20)	1.10	0.780	1.12	0.777	0.464	0.939
Hampali *et al.* [34] (CVPR'20)	1.07	0.788	1.06	0.790	0.506	0.942
METRO [29] (CVPR'21)	1.04	0.792	1.11	0.779	0.484	0.946
Liu *et al.* [35] (CVPR'21)	0.99	0.803	0.95	0.810	0.528	0.956
I2UV-HandNet [2] (ICCV'21)	0.99	0.804	1.01	0.799	0.500	0.943
ArtiBoost [36] (CVPR'22)	1.14	0.773	1.09	0.782	0.488	0.944
Keypoint Trans. [14] (CVPR'22)	1.08	0.786	-	-	-	-
MobRecon [37] (CVPR'22)	0.92	-	0.94	-	0.538	0.957
HandOccNet [20] (CVPR'22)	0.91	0.819	0.88	0.819	0.564	**0.963**
HandGCAT (Ours)	**0.87**	**0.826**	**0.87**	**0.827**	**0.584**	**0.963**

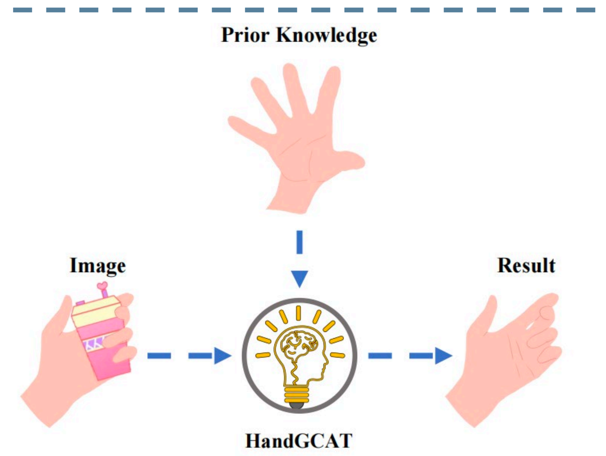
TABLE II
COMPARISON WITH STATE-OF-THE-ART METHODS ON HO3D V3.

Method	PA-MPJPE ↓	PA-MPJPE AUC ↑	PA-MPVPE ↓	PA-MPVPE AUC ↑	F@5 ↑	F@15 ↑
ArtiBoost [36] (CVPR'22)	1.08	0.785	1.04	0.792	0.507	0.946
Keypoint Trans. [14] (CVPR'22)	1.09	0.785	-	-	-	-
HandOccNet [20] (CVPR'22)	1.07	0.786	1.04	0.791	0.479	0.935
HandGCAT (Ours)	**0.93**	**0.814**	**0.91**	**0.818**	**0.552**	**0.956**

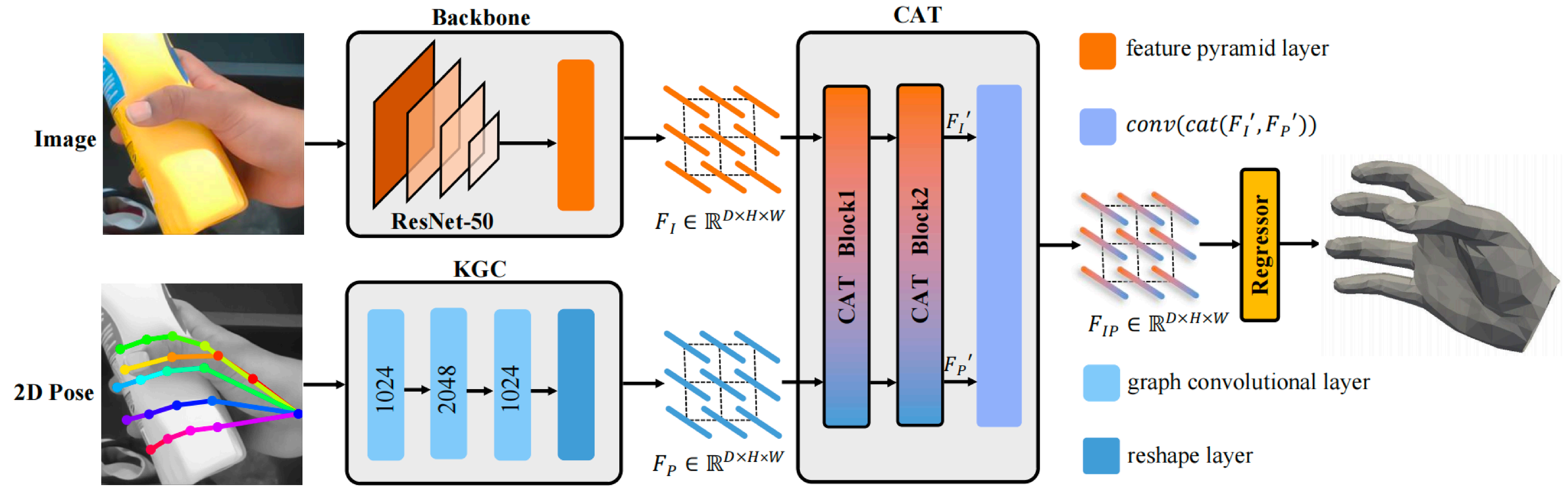
Results on DexYCB Dataset

TABLE III
COMPARISON WITH SOTA ON DEXYCB DATASET.

Method	MPJPE ↓	PA-MPJPE ↓
Spurr *et al.* [40] (ECCV'20)	17.34	6.83
METRO [29] (CVPR'21)	15.24	6.99
Liu *et al.* [35] (CVPR'21)	15.28	6.58
HandOccNet [20] (CVPR'22)	14.04	5.80
HandGCAT (Ours)	**13.76**	**5.60**



Method



Overview of the Proposed Method

- The proposed HandGCAT consists of backbone, KGC, CAT, and regressor. Resnet-50 with FPN extracts image feature F_I .
- KGC captures hand prior knowledge F_P using GCNs from the 2D pose. CAT fuses F_P into F_I and thus imagines occluded regions.
- Finally, the regressor reconstructs the 3D hand mesh.

Ablation Study

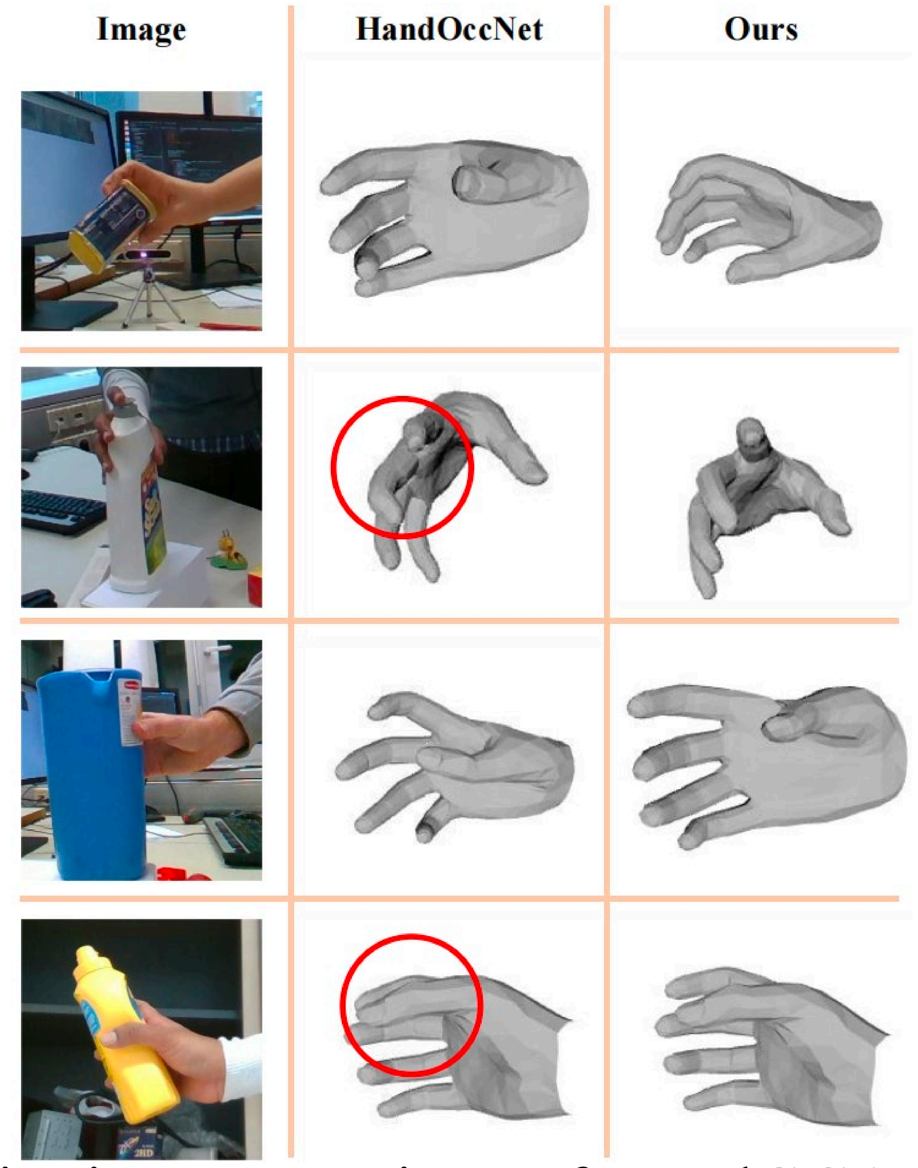
TABLE IV
COMPARISON OF MODELS WITH VARIOUS KGC ARCHITECTURES ON HO3D V2.

KGC architectures	PA-MPJPE ↓	PA-MPVPE ↓	F@5 ↑	F@15 ↓
MLP	0.93	0.93	0.547	0.959
1-layer GCN	0.92	0.92	0.546	0.961
2-layer GCNs	0.90	0.89	0.570	0.961
3-layer GCNs	0.89	0.88	0.573	**0.963**
4-layer GCNs	**0.87**	**0.87**	**0.584**	**0.963**
5-layer GCNs	0.89	0.88	0.579	0.962

TABLE V
COMPARISON OF MODELS WITH VARIOUS CAT ARCHITECTURES ON HO3D V2.

CAT architectures	PA-MPJPE ↓	PA-MPVPE ↓	F@5 ↑	F@15 ↑
Two Transformers	0.90	0.90	0.563	0.962
Single CAT block	0.89	0.88	0.574	0.962
Two CAT blocks	**0.87**	**0.87**	**0.584**	**0.963**
Three CAT blocks	0.88	**0.87**	0.583	**0.963**

Visualization



Qualitative comparison of HandGCAT and state-of-the-art method on HO3D v2.

Conclusions

- We propose a novel hand reconstruction method named HandGCAT, which is robust to occlusions.
- We propose a knowledge-guided graph convolution module to learn the hand prior.
- We design a cross-attention Transformer to fuse the hand prior into occluded regions.

Acknowledgements

This work is supported by National Key R&D Program of China (2021ZD0113502), Shanghai Municipal Science and Technology Major Project (2021SHZDZX0103).

Contact

Shuaibing Wang
sbwang21@m.fudan.edu.cn