3D Interacting Hand Pose Estimation by Hand De-occlusion and Removal

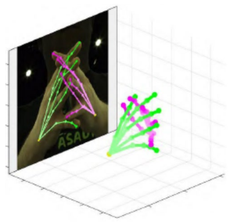
Hao Meng[1,3*] Sheng Jin[2,3*] Wentao Liu[3,4] Chen Qian[3]
Mengxiang Lin[1] Wanli Ouyang[4,5] Ping Luo[2]
[1] Beihang University [2] The University of Hong Kong
[3] SenseTime Research and Tetras.AI [4] Shanghai AI Lab [5] The University of Sydney

ECCV TEL AVIV 2022

Task

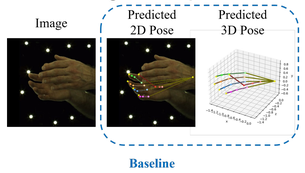
3D Interacting Hand Pose Estimation

Estimating 3D keypoint localization of two interacting hands.



Motivation

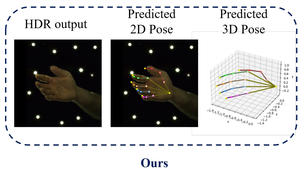
➢ Interacting hand pose estimation VS single hand pose estimation:



☹ Occlusion between two hands

☹ Ambiguity between two hands

➢ HDR (Hand De-occlusion and Removal):



☺ Single hand pose estimation for every hand

☺ De-occlusion: restore pixel values of occluded part of hand

☺ Removal: remove the other hand, i.e., predict pixel values of background behind the other hand

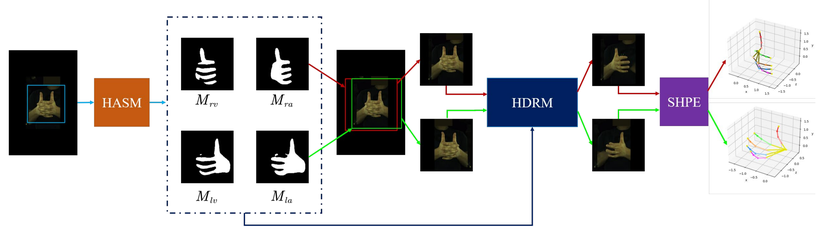
Contributions

➢ We propose a novel Hand De-occlusion and Removal (HDR) framework to tackle the challenging task of 3D interacting hand pose estimation.

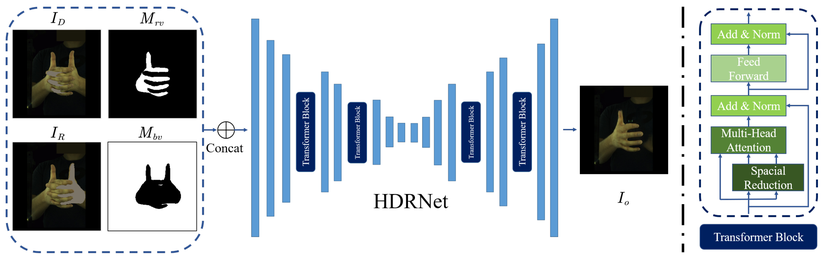
➢ We propose to explicitly handle the challenges of self-occlusion by hand de-occlusion and the homogeneous appearance ambiguity by distractor removal. To the best of our knowledge, we are the first to apply de-occlusion techniques to improve the downstream pose estimation accuracy.

➢ We propose the first large-scale synthetic Amodal InterHand Dataset (AIH) to settle the task of hand de-occlusion and removal. We envision that AIH will foster the development of the related research.

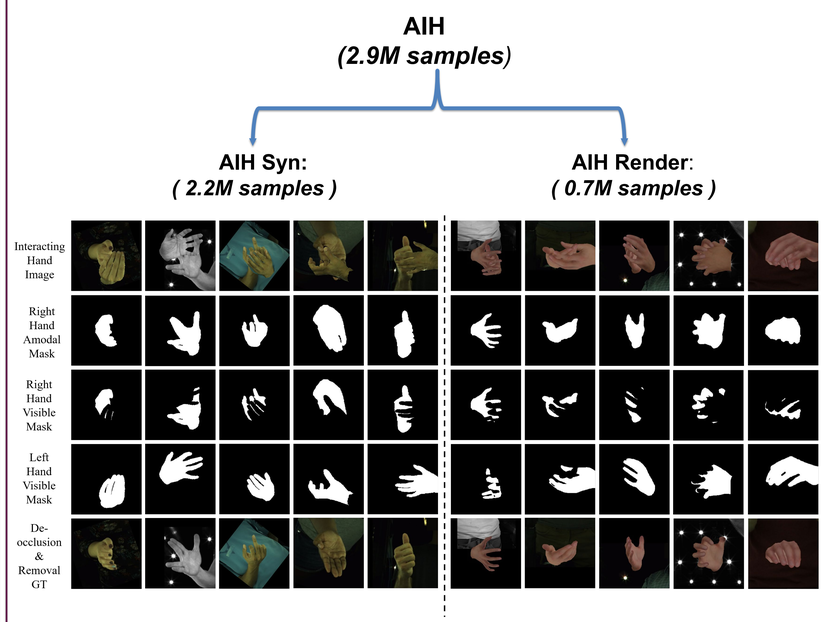
Overview



Hand De-occlusion and Removal Module (HDRM)



Amodal InterHand (AIH) Dataset



Experiment Results

➢ Quantitative results

Comparisons on InterHand2.6M V1.0 Dataset

Methods	InterHand2.6M - ALL branch				InterHand2.6M - M branch		
	IH26M-SH	IH26M-IH	IH26M-ALL	IH26M-Inter	IH26M-SH	IH26M-IH	IH26M-ALL
*Boukhayma *et al.* [4]	-	-	27.14	31.46	-	-	-
*Pose2Mesh [5]	-	-	27.10	32.11	-	-	-
*BiHand [35]	-	-	25.10	28.23	-	-	-
*Rong *et al.* [27]	-	-	17.12	20.66	-	-	-
DIGIT [7]	-	14.27	-	-	-	-	-
InterNet [21]	12.16	16.02	14.21	18.04	12.52	18.04	15.28
HDR (Ours)	**8.51**	**13.12**	**10.97**	**14.74**	**8.52**	**14.98**	**11.74**

Comparisons on Tzionas Dataset

Model	Boukhayma *et al.* [4]	Wang *et al.* [31]	InterNet [21]	Kim *et al.* [13]	SHPE	SHPE+HDR
EPE↓	12.91	13.31	17.61	12.42	14.88	**8.70**

Comparisons on InterHand2.6M V1.0 dataset

(ALL branch)

Methods	Train (ALL, IH26M-SH)		Train (ALL, IH26M-SH +AIH)	
	IH26M-IH	IH26M-ALL	IH26M-IH	IH26M-ALL
SHPE [39]	39.96	25.90	30.23	20.93
+HDR (Ours)	25.93	18.39	23.99	17.58

Comparisons on InterHand2.6M V1.0 dataset

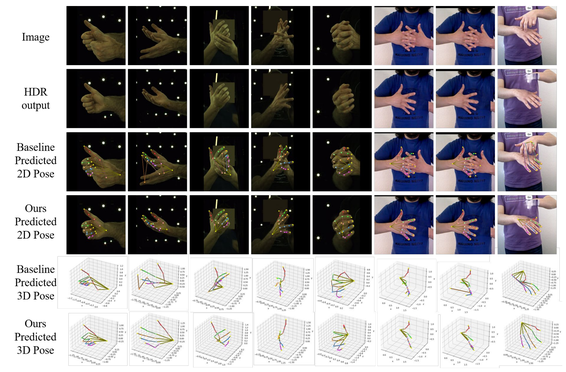
(Machine annot branch)

Methods	Train (M, IH26M-SH)		Train (M, IH26M-SH +AIH)	
	IH26M-IH	IH26M-ALL	IH26M-IH	IH26M-ALL
SHPE [39]	40.98	25.78	32.27	21.66
+HDR (Ours)	25.45	17.98	24.59	17.80

Ablation Studies

	Methods	MPJPE (mm)	Δ
#1	SHPE [39] only	25.78	+7.80 (43.4%)
#2	w/o Removal	24.16	+6.18 (34.4%)
#3	w/o De-occlusion	19.69	+1.71 (9.5%)
#4	w/o Discriminator	18.11	+0.13 (0.7%)
#5	w/o Transformer Block	18.85	+0.87 (4.8%)
#6	AIH_Render only	18.10	+0.12 (0.7%)
#7	AIH_Syn only	18.35	+0.37 (2.1%)
#8	Ours	**17.98**	-

➢ Qualitative results



(a) InterHand2.6M (b) Tzionas

➢ Time complexity analysis

Test on one Tesla P40 GPU in a single thread