



A Dataset of Relighted 3D Interacting Hands

NeurIPS 2023 (Datasets and Benchmarks Track)

Gyeongsik Moon, Shunsuke Saito, Weipeng Xu, Rohan Joshi, Julia Buffalini, Harley Bellan, Nicholas Matthew Rosen, Jesse Richardson, Mallorie Mize, Philippe De Bree, Tomas Simon, Bo Peng, Shubham Garg, Kevyn Alex Anthony McPhail, Takaaki Shiratori







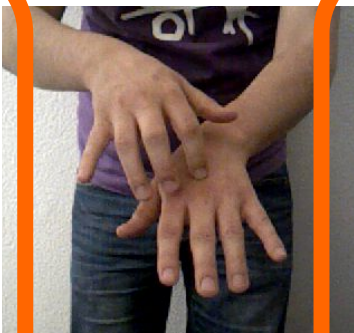
Table 1: Comparison of hand datasets that provide GT 3D poses. To count the number of images, we consider images from different viewpoints at the same time step as different ones.

Datasets	Image appearance	GT	# of subjects	# of images	Two-hand interactions
Dexter+Object [44]	Natural	3D fingertip coord.	1	3K	No
STB [59]	Natural	3D joint coord.	1	36K	No
EgoDexter [33]	Natural	3D fingertip coord.	4	3K	No
RHD [61]	Composite	3D joint coord.	20	44K	No
PanopticStudio [43]	Lab	3D joint coord.	N/A	15K	No
FPHA [9]	Natural	3D joint coord. + 3D obj.	6	105K	No
GANerated [31]	Composite	3D joint coord.	N/A	264K	No
FreiHAND [62]	Natural	MANO	32	134K	No
ObMan [15]	Composite	MANO + 3D obj.	20	150K	No
EHF [39]	Lab	SMPL-X	1	100	No
HO3D [13]	Natural	3D joint coord. + 3D obj.	10	78K	No
ContactPose [3]	Lab	3D joint coord. + 3D obj.	50	2.9M	No
HUMBI [56, 55]	Lab	MANO	453	24M	No
DexYCB [4]	Natural	MANO + 3D obj.	10	582K	No
AGORA [38]	Realistic	SMPL-X	350	19K	No
DART [8]	Composite	MANO	N/A	787K	No
BlurHand [34]	Lab	MANO	11	156K	No
H2O [21]	Natural	MANO + 3D obj.	4	571K	Weak
Assembly101 [42]	Natural	3D joint coord. + action labels	53	111M	Weak
AssemblyHands [35]	Natural	3D joint coord.	34	3M	Weak
ARCTIC [71]	Lab	MANO + SMPL-X + 3D obj.	10	2.1M	Weak
HIC [52]	Natural	MANO	1	36K	Strong
RGB2Hands [53]	Natural	3D joint coord. wo. fingertips	2	1K	Strong
InterHand2.6M [30]	Lab	MANO	27	2.6M	Strong
Ego3DHands [23]	Composited	3D joint coord. + masks	1	55K	Strong
Re:InterHand (Ours)	Realistic	MANO + masks	10	1.5M	Strong

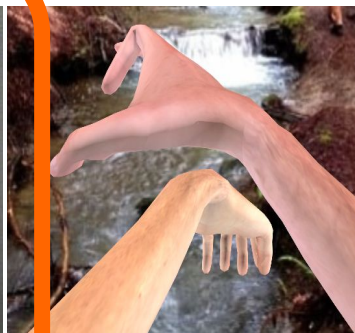
Datasets for 3D Interacting Hands Understanding



(a) InterHand2.6M
(Lab dataset)



(b) HIC
(Natural dataset)



(c) Ego3DHands
(Composited dataset)

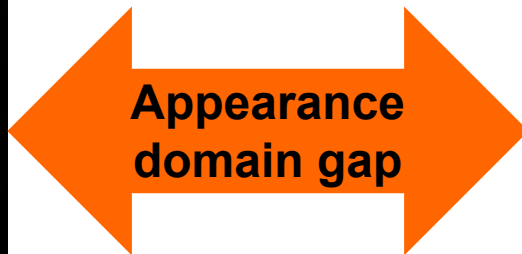
Appearance diverse?	No	No	Yes
Appearance realistic?	Yes	Yes	No
Appearance close to ITW?	No	Yes	No
GT 3D pose diverse?	Yes	No	Yes

Moon et al. "InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose Estimation from a Single RGB Image", ECCV. 2020.

Tzionas et al. "Capturing hands in action using discriminative salient points and physics simulation." IJCV. 2016.

Lin et al. "Two-hand global 3D pose estimation using monocular rgb." WACV. 2021.

Image appearances are monotonous...



Re:InterHand - A Dataset of Relighted 3D Interacting Hands



Relighting network



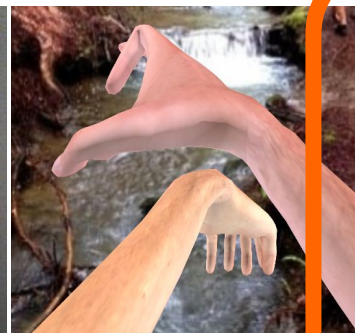
Datasets for 3D Interacting Hands Understanding



(a) InterHand2.6M
(Lab dataset)



(b) HIC
(Natural dataset)



(c) Ego3DHands
(Composited dataset)



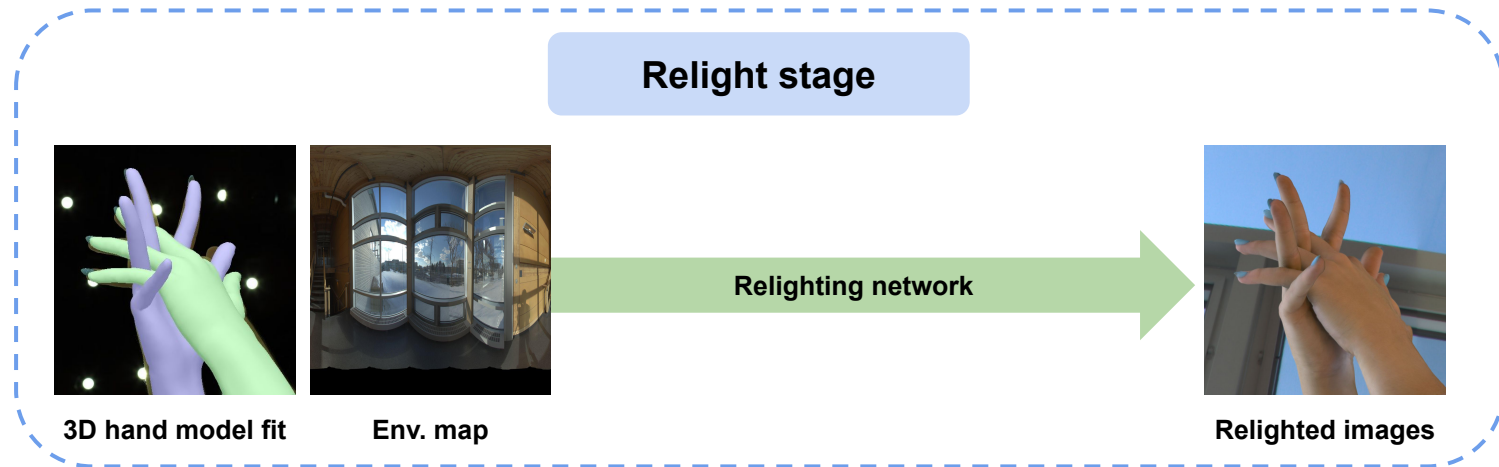
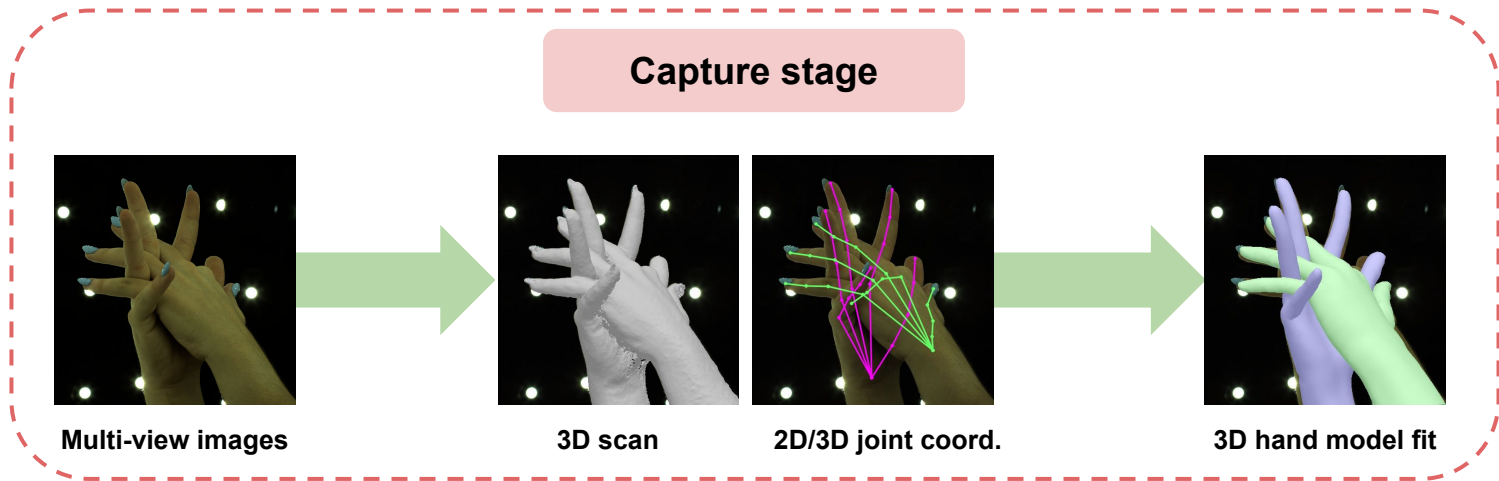
(d) Re:InterHand
(Ours)

Appearance diverse?	No	No	Yes	Yes
Appearance realistic?	Yes	Yes	No	Middle
Appearance close to ITW?	No	Yes	No	Middle
GT 3D pose diverse?	Yes	No	Yes	Yes

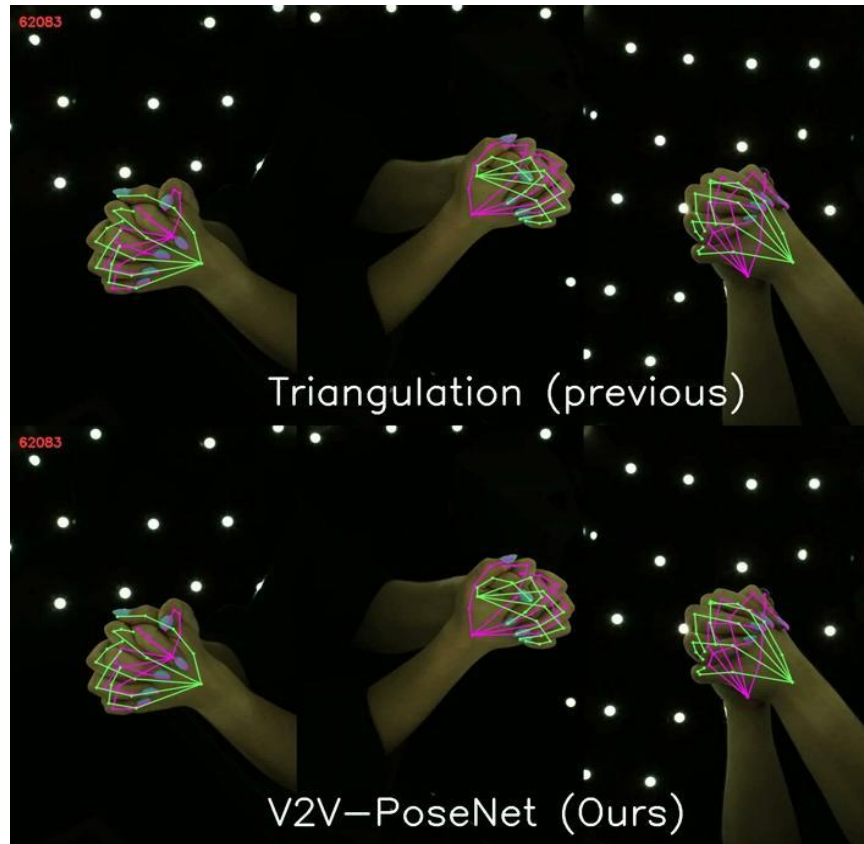
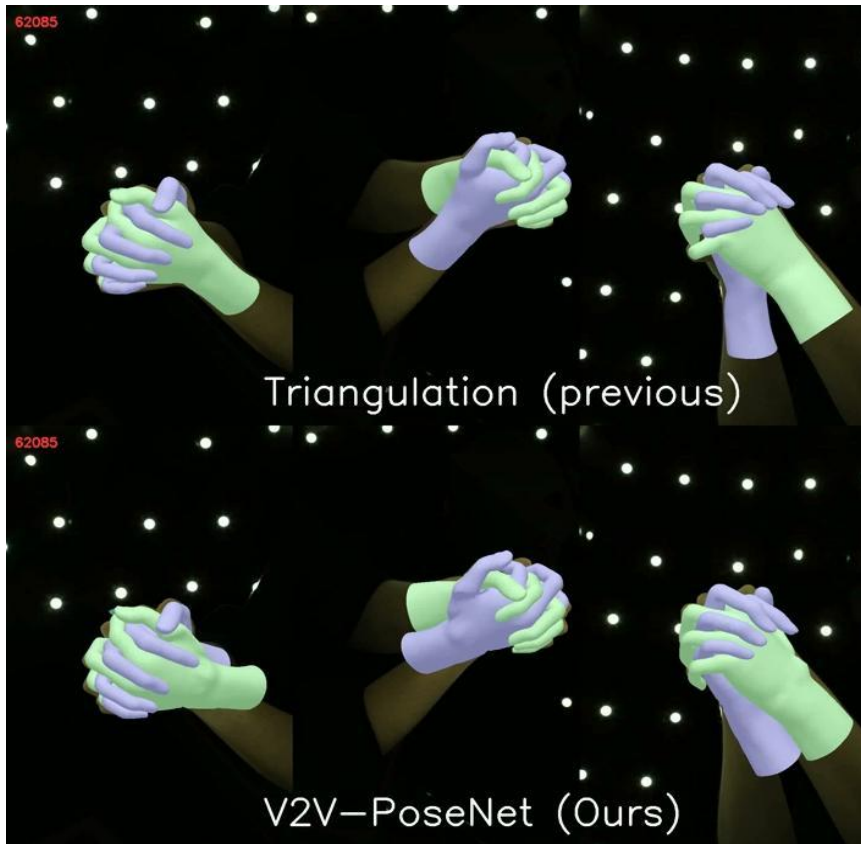
Moon et al. "InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose Estimation from a Single RGB Image", ECCV. 2020.

Tzionas et al. "Capturing hands in action using discriminative salient points and physics simulation." IJCV. 2016.

Lin et al. "Two-hand global 3D pose estimation using monocular rgb." WACV. 2021.

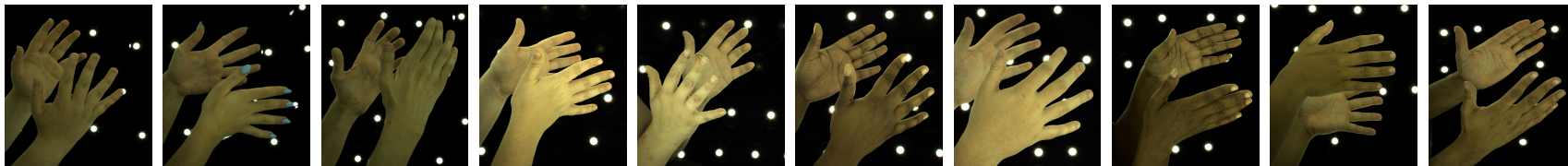


Better and stable capture stage than InterHand2.6M



10 Captures with both 3rd-person/egocentric viewpoints!

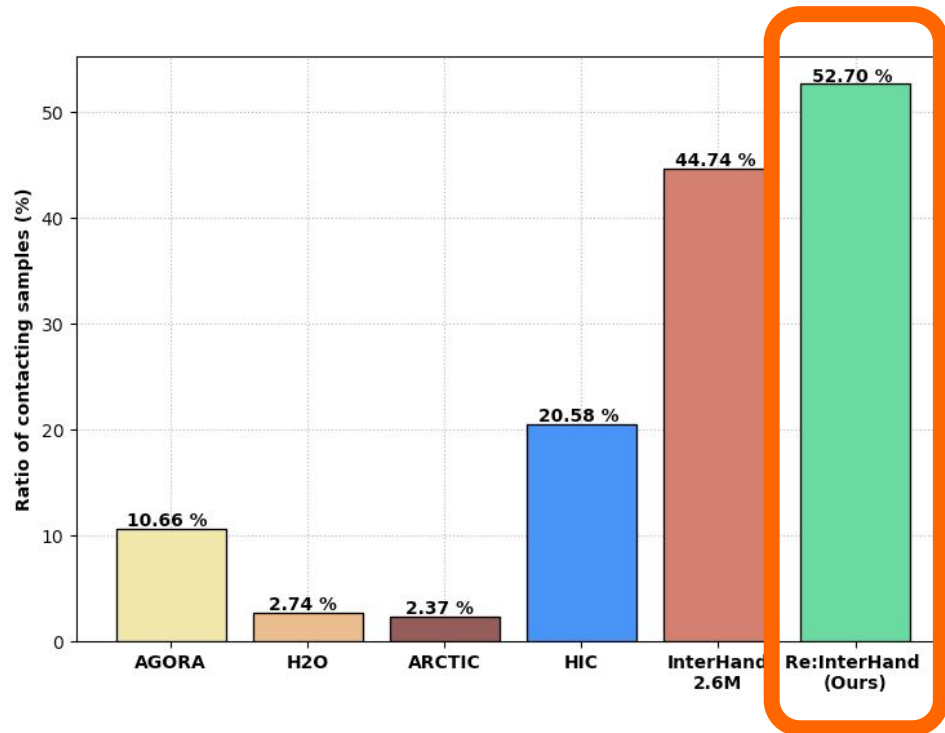
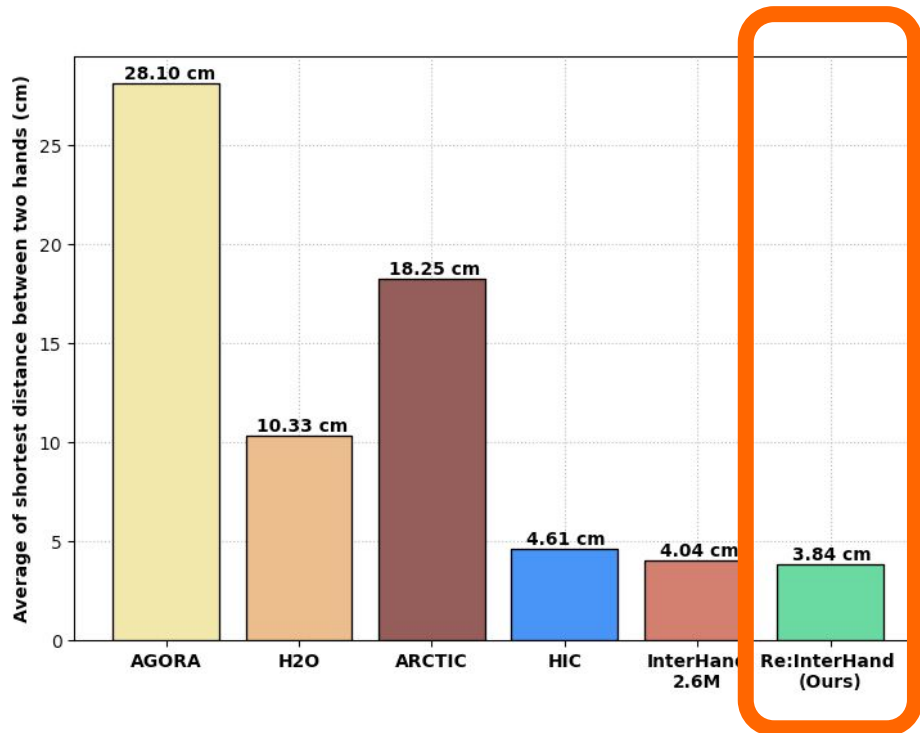
3rd-person viewpoints



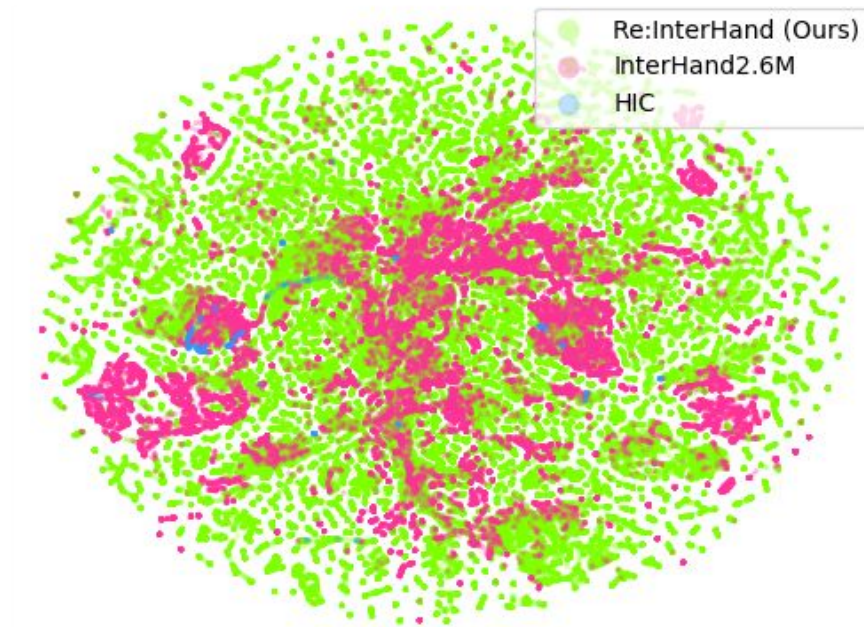
egocentric viewpoints



The strongest interactions between two hands



The most diverse 3D interacting hands



What to release?



What to release?



Dataset Splits

- **3rd-person viewpoint / frame-based version (5 fps. 493K images.)**
 - 20 cameras picked with the iterative farthest sampling
 - All frames have different environment maps
 - Same frames with different cameras share the same environment maps
- **3rd-person viewpoint / video version (30 fps. 739K images.)**
 - 5 cameras picked with the iterative farthest sampling
 - Frames that belong to the same segment have the same environment maps
 - Same frames with different cameras share the same environment maps
- **Egocentric viewpoint / frame-based version (30 fps. 148K images.)**
 - Simulated Oculus Quest camera
 - All frames have different environment maps and camera augmentations
- **Egocentric viewpoint / video version (30 fps. 148K images.)**
 - Simulated Oculus Quest camera
 - Frames that belong to the same segment have the same environment maps and camera augmentations

Table 2: RRVE comparison of InterWild [26] trained on different data including variants of our dataset. We use the 3rd-person viewpoint split of our Re:InterHand.

Training sets	Testing sets		
	InterHand2.6M	HIC	Re:InterHand
InterHand2.6M + MSCOCO	19.74	23.59	37.59
+ Capture stage	19.14	23.09	34.23
+ Capture stage and Composite	19.50	24.37	31.30
+ Capture stage and Composite (w. AdaIn [16])	19.44	24.83	27.96
+ Capture stage and Relight stage (Ours)	19.40	21.36	20.07

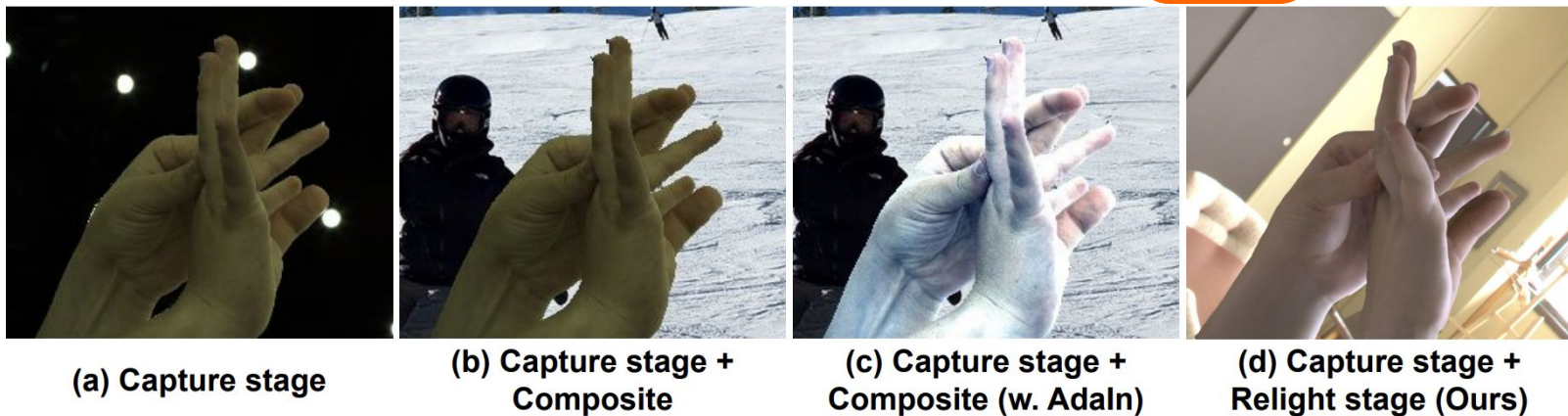


Figure 8: Image examples of datasets constructed in Tab. 2.



<https://mks0601.github.io/ReInterHand/>