

Supplementary Document:

Ego2HandsPose: A Dataset for Egocentric Two-hand 3D Global Pose Estimation

1. ManoFit Annotation Tool

In this section, we provide additional details for the introduced hand pose annotation tool, which consists of three major panels.

1.1. Display Panel

As shown in Figure 1, we display three images for annotation. The center image displays the rendered MANO hand model with a resolution of 400×400 . The left image consists of the input image resized and padded to match the same resolution for fitting. The user has the option to crop a hand region for a magnified view (the right image) or automatic 2D/3D hand pose estimation using our pretrained models. 2D joint locations can be annotated by first selecting the joint using the top slider and then clicking on the left or the right image.

1.2. Control Panel

We provide sliders to control global orientation γ_1 and translation γ_2 . In addition, sliders for the control of each finger is provided. While the first joint of the selected finger has degrees of freedom to rotate in the directions of θ , ϕ and ψ with the predefined ranges, the second and third joint of each finger can only rotate in the direction of ψ . This panel allows the user to have full manual control of the MANO hand model.

1.3. Action Panel

With an arbitrary number of provided 2D joint locations, the user can apply our parametric fitting algorithm to automatically fit MANO parameters using the defined losses. To automatically obtain the 2D joint locations, we allow the user to estimate the 2D/3D canonical hand poses using our pretrained models. A toggle button is provided to assist the user with improved ability to visualize how well the rendered hand model matches the input image. Other basic functionalities such as navigation between input images, resetting the MANO pose and saving MANO parameters are also provided.

2. Qualitative Examples

We provide qualitative examples for our annotated hand poses in Ego2HandsPose in Figure 2. This is the first RGB-based 3D two-hand pose dataset collected in non-laboratory environment, which allows the training and evaluation for the task of two-hand 3D global pose estimation in the wild using monocular data.

3. Additional Comparison

We provide quantitative results for all sequences in the test set of Ego2HandsPose in Table 1. Training on the training set of Ego2HandsPose outperforms other datasets on all sequences for all stages. We find that models trained on H2O and FreiHAND struggle particularly on sequences with bright (seq-2) and dark illumination (seq-5). In addition, these methods generally have lower performance on sequences with fast motion (seq-7 and seq-8), which introduces lower-quality input images challenging for 2D hand pose estimation. Qualitative comparison is provided in Figure 3.

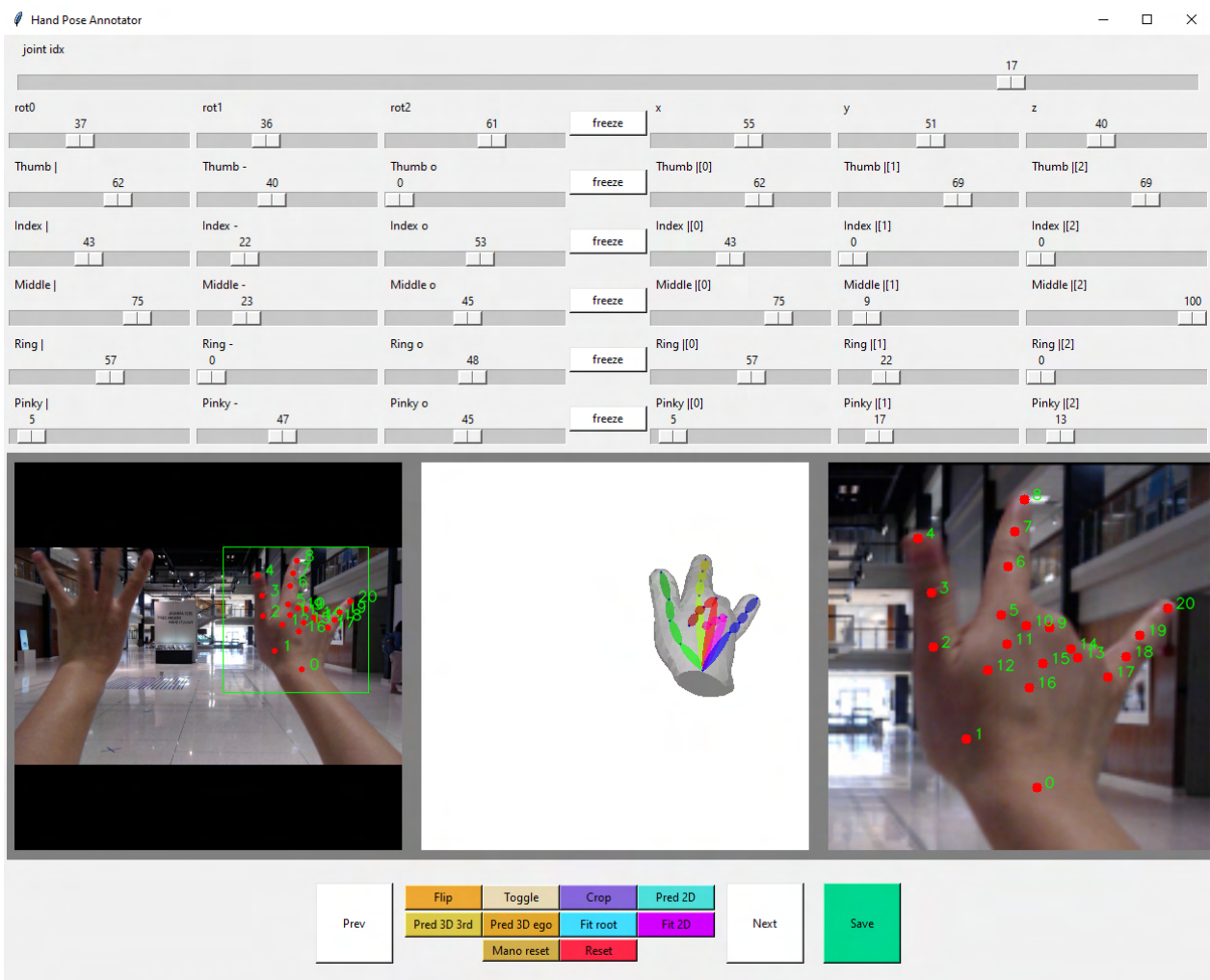


Figure 1. Hand pose annotation tool for ManoFit using a single image.

Sequence	AUC _{2D}			AUC _{3D}			AUC _{angle}			AUC _{radius}		
	Ours	H2O	Frei	Ours	H2O	Frei	Ours	H2O	Frei	Ours	H2O	Frei
seq-1	0.574	0.476	0.464	0.477	0.338	0.377	0.938	0.913	0.907	0.786	0.658	0.581
seq-2	0.475	0.095	0.136	0.378	0.149	0.203	0.917	0.770	0.762	0.741	0.192	0.078
seq-3	0.564	0.429	0.319	0.518	0.415	0.415	0.943	0.924	0.865	0.759	0.747	0.547
seq-4	0.581	0.480	0.432	0.499	0.438	0.474	0.936	0.925	0.907	0.846	0.765	0.731
seq-5	0.516	0.270	0.289	0.417	0.278	0.282	0.911	0.858	0.843	0.660	0.315	0.211
seq-6	0.511	0.374	0.278	0.499	0.425	0.411	0.926	0.899	0.856	0.825	0.612	0.472
seq-7	0.574	0.255	0.396	0.362	0.165	0.310	0.923	0.828	0.902	0.760	0.460	0.510
seq-8	0.375	0.155	0.246	0.303	0.137	0.232	0.899	0.782	0.840	0.693	0.347	0.336
Average	0.508	0.317	0.320	0.432	0.293	0.338	0.924	0.862	0.860	0.759	0.512	0.433

Table 1. Quantitative comparison on sequences in the test set of Ego2HandsPose.

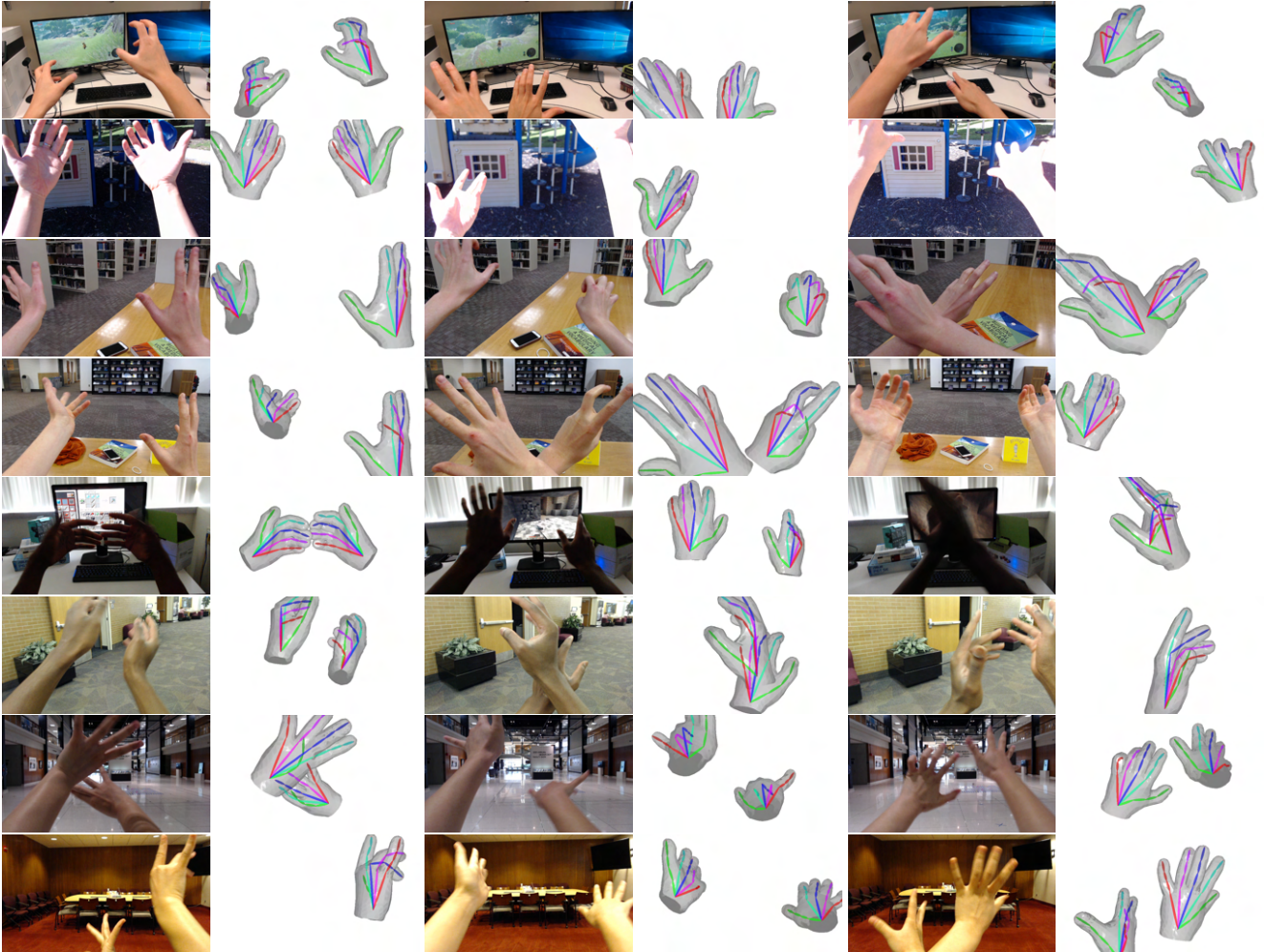
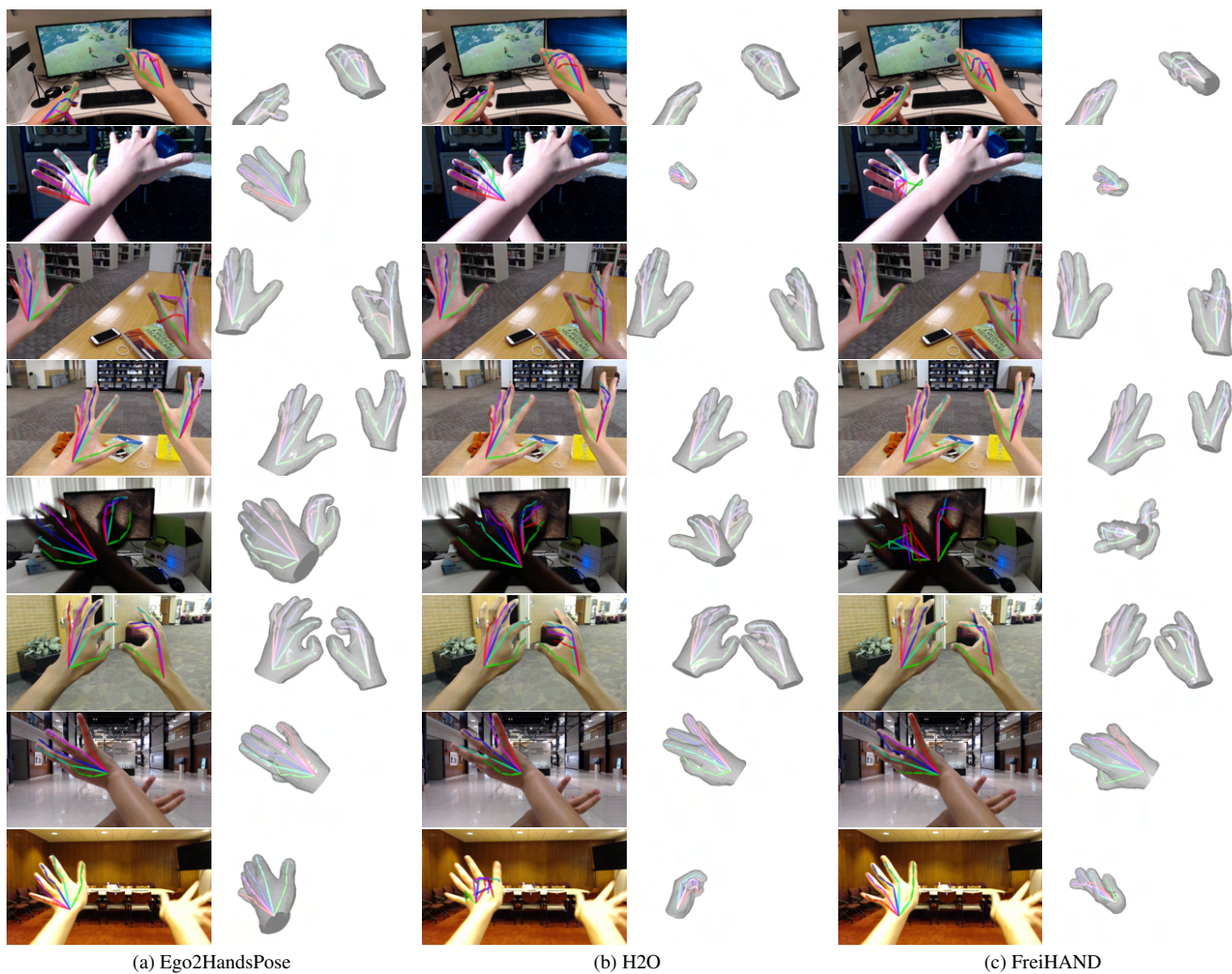


Figure 2. Qualitative examples of annotated instances in the evaluation set of Ego2HandsPose. Each row represents a test sequence. Hands with multiple joints outside of the camera view are skipped in annotation due to incomplete information.



(a) Ego2HandsPose

(b) H2O

(c) FreiHAND

Figure 3. Qualitative comparison between models trained on selected datasets on sequences of Ego2HandsPose. Results from each dataset are shown in two columns. The first column displays input images with visualized estimated 2D hand poses. The second column displays the final global two-hand poses. Hand poses that are partially outside of the view are not estimated.