



# Learning to Disambiguate Strongly Interacting Hands via Probabilistic Per-pixel Part Segmentation

Zicong Fan<sup>1,2</sup>, Adrian Spurr<sup>1</sup>, Muhammed Kocabas<sup>1,2</sup>,  
Siyu Tang<sup>1</sup>, Michael Black<sup>2</sup>, Otmar Hilliges<sup>1</sup>

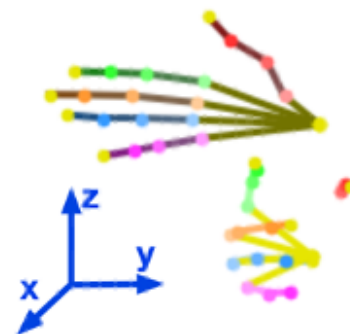
**MAX PLANCK INSTITUTE**  
FOR INTELLIGENT SYSTEMS

<sup>1</sup>ETH Zürich, Switzerland <sup>2</sup>Max Planck Institute for Intelligent Systems, Tübingen, Germany

**ETH** zürich

## Goal

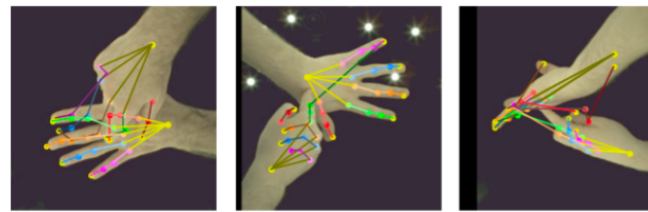
Estimate 3D poses of two interacting hands from a single RGB image.



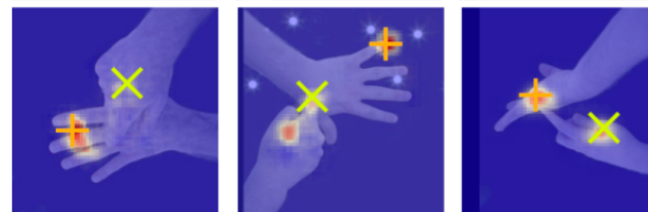
## Problem

Appearance ambiguity when two hands interact.

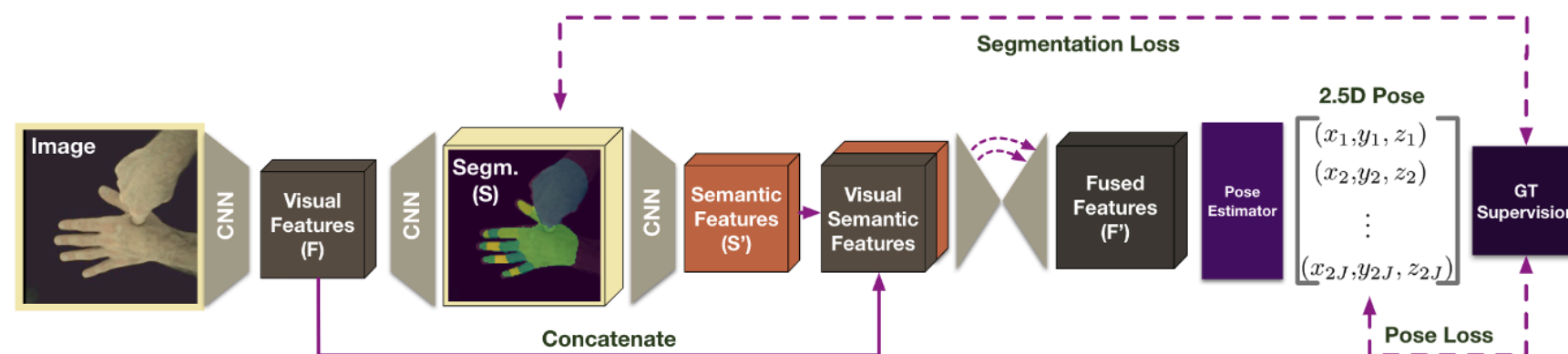
Keypoints



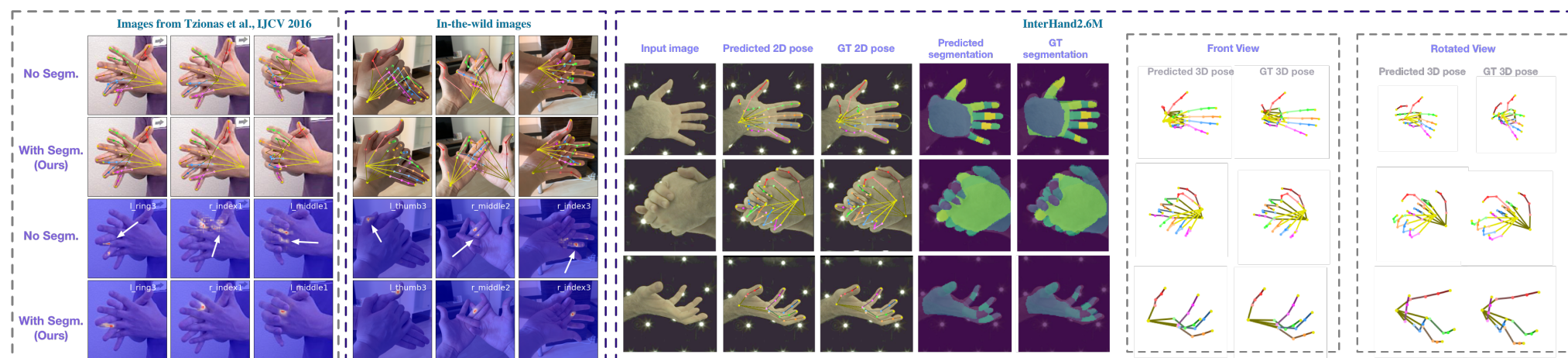
2D  
Heatmap



## DIGIT: DisambiGuating hands in InTeraction



## Qualitative Results

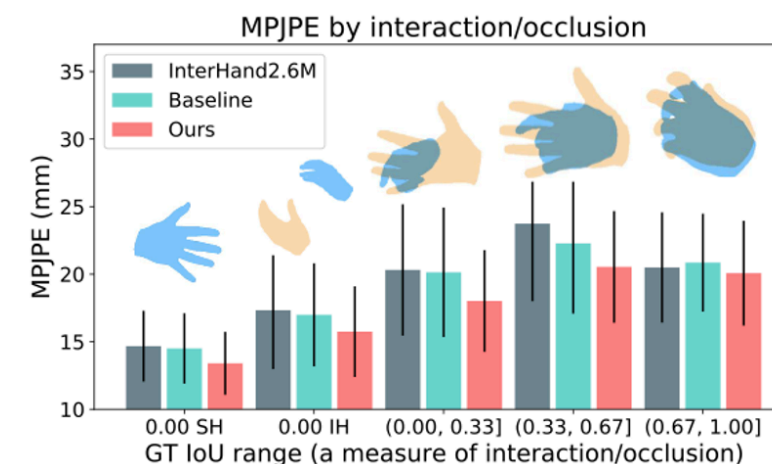


## Quantitative Results

SOTA performance

Methods	MPJPE Val	MRRPE Val	MPJPE Test	MRRPE Test
InterHand2.6M [33]	18.58	35.64	16.02	32.57
Baseline	17.79	33.90	15.06	31.36
Ours	<b>16.72</b>	<b>31.53</b>	<b>14.27</b>	<b>29.22</b>
% in improvement over [33]	10.01	11.53	10.92	10.29

## Ablation Study



Allows end-to-end training

Ablation Study	MPJPE Val	MRRPE Val	MPJPE Test	MRRPE Test
Part segm. <sup>§</sup>	16.68/23.52	41.99	14.35/20.57	38.87
Part segm. (ours)	<b>14.06/20.01</b>	<b>35.13</b>	<b>12.30/17.22</b>	<b>32.88</b>

Preserves segmentation uncertainty

Ablation Study	MPJPE Val	MRRPE Val	MPJPE Test	MRRPE Test
LR segm. <sup>†</sup>	28.72/36.05	50.85	25.75/31.46	46.98
Part segm. <sup>†</sup>	17.69/25.49	46.00	15.16/22.08	41.46
LR segm. (ours)	14.87/21.19	<b>34.70</b>	12.92/18.40	<b>32.13</b>
Part segm. (ours)	<b>14.03/20.01</b>	35.26	<b>12.29/17.23</b>	32.88



Project Page  
[zc-alexfan.github.io/digit](https://zc-alexfan.github.io/digit)