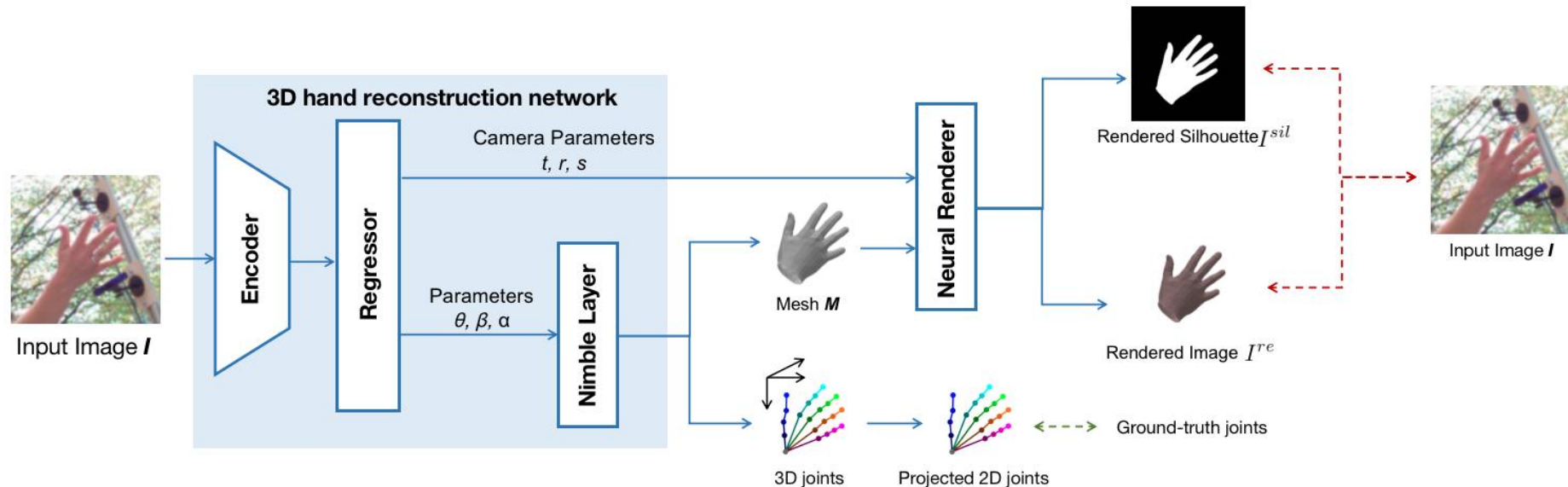# Problem & Approach

➢ **3D hand recontruction from monocular images**

- Applications:
  - ➢ AR, VR, HCI, AIGC
- Existing frameworks
  - ➢ Coarse hand mesh
  - ➢ Coarse texture
  - ➢ Insufficient high-quality ground-truth poses
  - ➢ Current Render-and-Compare scheme is less effective

- My attempts
  - ➢ Pipeline (similar to S2HAND [1])
  - ➢ Mano -> Nimble
  - ➢ Augmented training from synthetic data with Nimble (Future)
  - ➢ Improve Render-and-Compare scheme (Future)

[1] Chen, Yujin, et al. "Model-based 3d hand reconstruction via self-supervised learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

Interesting Observations

➢ Self-supervised scheme performs better than weakly-supervised scheme with confidence scores. Confidence scores are useful additional corrections to OpenPose noisy labels

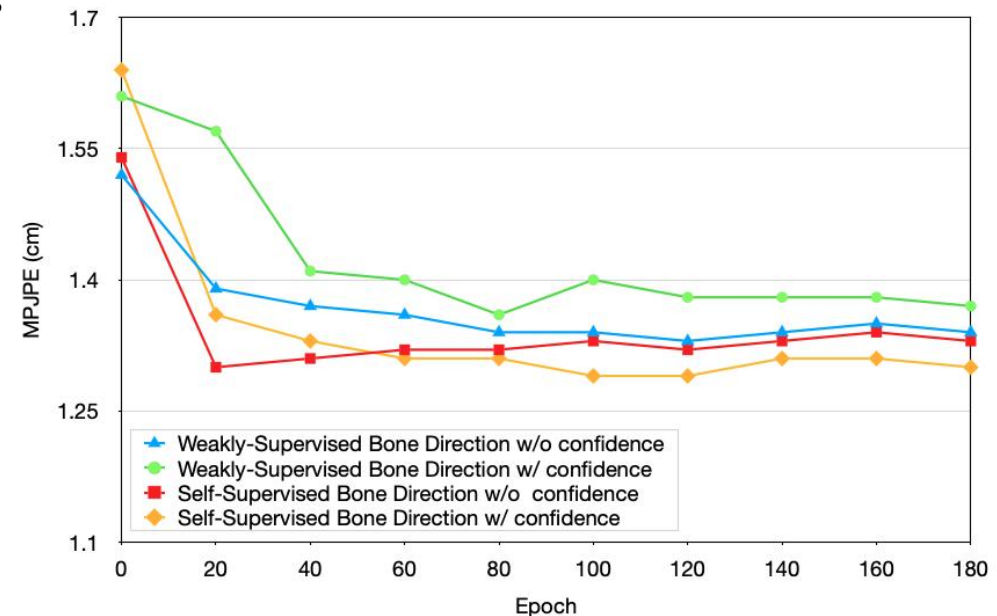- Comparison with S2HAND on FreiHAND testset:

Supervision	Method	AUC$_J$↑	MPJPE↓	AUC$_V$↑	MPVPE↓	F5↑	F15↑
OpenPose	S2HAND	0.77	1.18	0.77	1.19	0.48	0.92
GT-2D	Ours	0.74	1.33	0.74	1.33	0.42	0.89

- Similar experiment in S2HAND:

Dataset	Method	AUC$_J$↑	AUC$_V$↑	F$_5$↑	F$_{15}$↑
FreiHAND	WSL	0.730	0.725	0.42	0.89
	SSL	**0.766**	**0.765**	**0.48**	**0.92**
HO-3D	WSL	0.765	0.769	0.44	**0.93**
	SSL	**0.773**	**0.777**	**0.45**	**0.93**

- Comparison of Different Supervised Bone Orientation Loss

Method	MPJPE↓	MPVPE↓
WS Bone Orientation w/o Confidence	1.33	1.33
WS Bone Orientation w Confidence	1.38	1.40
SS Bone Orientation w/o Confidence	1.33	1.33
SS Bone Orientation w Confidence	**1.29**	**1.30**

Interesting Observations

➢ Adding a ground-truth 3D joint loss does not improve aligned results, but rather unaligned results. Adding a ground-truth 3D joint norm loss can improve both aligned and unaligned results
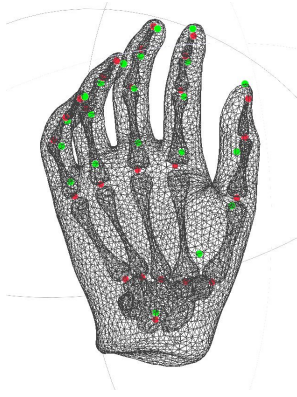
- Align results:

Supervision	Method	AUC$_J$↑	MPJPE↓	AUC$_V$↑	MPVPE↓	F5↑	F15↑
OpenPose	S2HAND	0.77	1.18	0.77	1.19	0.48	0.92
2D	Ours	0.74	1.33	0.74	1.33	0.42	0.89
2D + 3D	Ours	0.73	1.39	0.73	1.38	0.41	0.89
2D + 3D Norm	Ours	**0.76**	**1.24**	**0.76**	**1.22**	**0.46**	**0.91**

- Unaligned results:

Supervision	Method	Losses	MPJPE↓	MPVPE↓
FS	S2HAND	2D + 3D	8.66	8.77
SS	S2HAND	–	10.57	10.60
WS	Ours	2D	11.65	11.70
FS	Ours	2D + 3D	8.31	8.34
FS	Ours	2D + 3D Norm	**7.72**	**7.77**

Challenges

➢ Regression: Nimble joint rig is different from Mano joint rig

➢ Texture: Transform texture image to TextureUV for hand mesh rendering to perform photorealistic consistency

➢ Collision problem: Proper regularizers on pose, shape and texture parameters are required
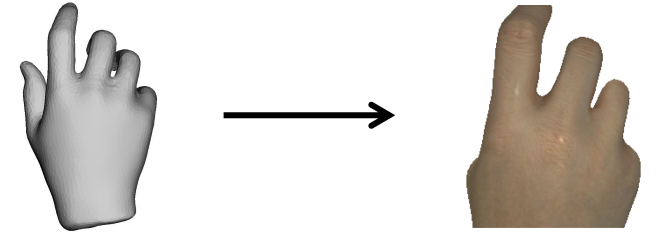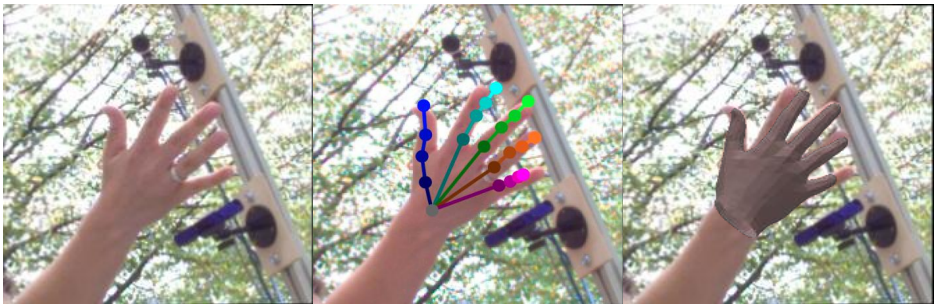

➢ Render: Camera intrinsic matrix K required by Pytorch3D is of shape 4 x 4 instead of commonly used 3 x 3

`K: (optional) A calibration matrix of shape (N, 4, 4)`

$$K = \begin{bmatrix} f_x & 0 & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

Visualizations & Results
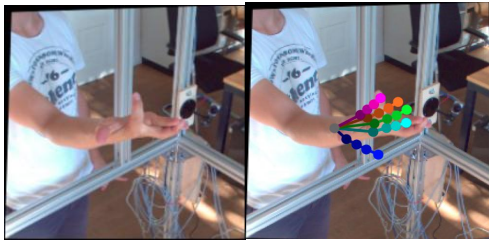
➢ Qualitative results with MANO
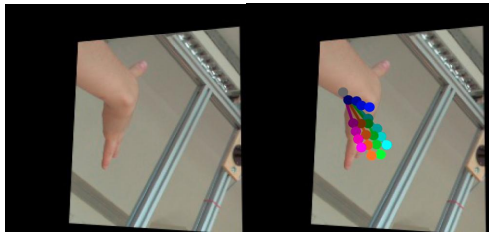

➢ Qualitative results with NIMBLE

➢ Comparison of using NIMBLE layer

Hand Model	Method	PA MPJPE↓
NIMBLE	I2L-MeshNet [1]	9.4
NIMBLE	Ours	3.5

➢ Failure Cases

Extreme Pose

Extreme Viewpoint

[1] Li, Yuwei, et al. "NIMBLE: a non-rigid hand model with bones and muscles." ACM Transactions on Graphics (TOG) 41.4 (2022)