

# Self-Supervised Robustifying Guidance for Monocular 3D Face Reconstruction



Hitika Tiwari<sup>1,4,\*</sup>, Min-Hung Chen<sup>2,\*\*</sup>, Yi-Min Tsai<sup>3</sup>, Hsien-Kai Kuo<sup>3</sup>, Hung-Jen Chen<sup>3</sup>, Kevin Jou<sup>3</sup>, K. S. Venkatesh<sup>1</sup>, Yong-Sheng Chen<sup>4</sup>

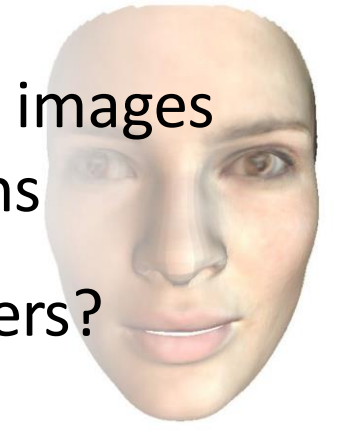
<sup>1</sup>Indian Institute of Technology Kanpur, India, <sup>2</sup>Microsoft AI R&D Center, Taiwan, <sup>3</sup>MediaTek Inc., Taiwan, <sup>4</sup>National Yang Ming Chiao Tung University, Taiwan

\*Work partially done during the research internship at MediaTek Inc., Taiwan, \*\*Work partially done during MediaTek Inc., Taiwan

## Motivation

### Monocular 3D Face Reconstruction

- Goal** Generate 3D faces from monocular face images without using 3D ground truth face scans
- Challenges** How to influence the output? Parameters?
- Applications** Face artifice, animation, etc.

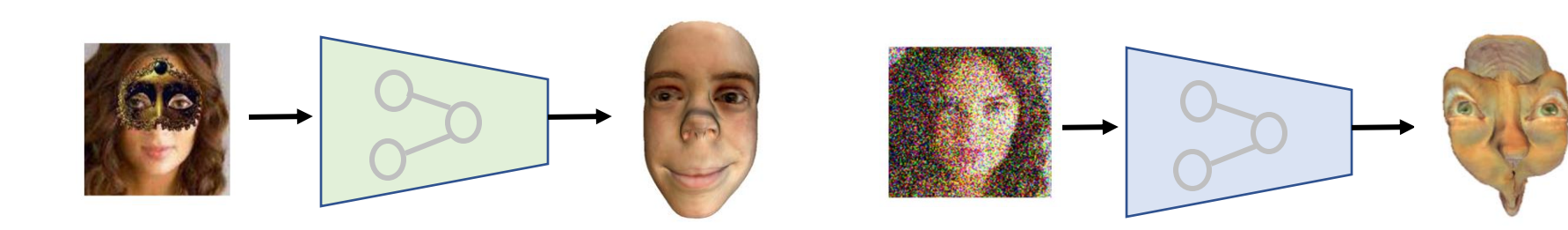


### Robust 3D Face Reconstruction

- Goal** Generate 3D faces from occluded and noisy monocular face images
- Challenges** How to attain robustness in a self-supervised manner?
- Applications** Face recognition, animation, etc.

#### Occlusion Robust Methods

#### Noise Robust Methods



- input+ weak label → output
- under-addressed issue

### Evaluation Dataset

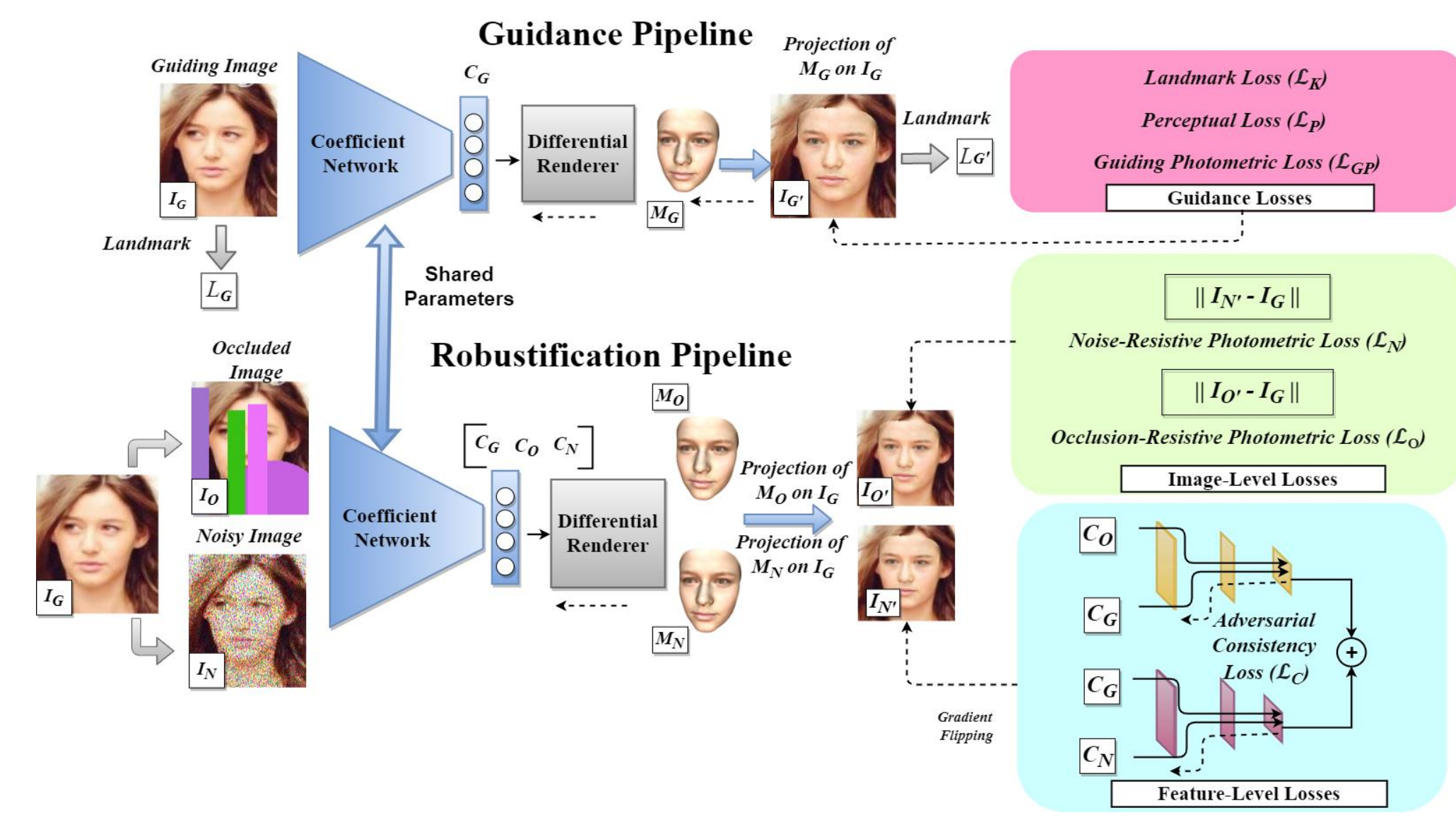
- Goal** Texture + shape evaluation
- Texture evaluation
- Challenges**
  - Open access
  - Occlusion dedicated dataset



## Method

Our *Self-Supervised ROBustifying GUIDance (ROGUE)* framework learns statistical facial coefficients for occluded, and noisy face images in a self-supervised manner using following pipelines

- Guidance Pipeline** Estimates coefficients for the clean target face
- Robustification Pipeline** Enforces the estimated coefficients of occluded and noisy faces to be consistent with clean images



#### Guidance Losses

$$\mathcal{L}_K = \|L_G - L_{G'}\|$$

$$\mathcal{L}_{GP} = \|I_G - I_{G'}\|$$

$$\mathcal{L}_P = 1 - \frac{\langle \theta, \theta' \rangle}{\|\theta\| \|\theta'\|}$$

$$\mathcal{L}_R = w_s \|s\| + w_t \|t\| + w_e \|e\|$$

$$\mathcal{L}_{guide} = \alpha_K \mathcal{L}_K + \alpha_{GP} \mathcal{L}_{GP} + \alpha_P \mathcal{L}_P + \alpha_R \mathcal{L}_R$$

#### Robustification Losses

$$\mathcal{L}_{CO} = \mathcal{L}_h(D(C_G, C_O), [d_G, d_O]),$$

$$\mathcal{L}_{CN} = \mathcal{L}_h(D(C_G, C_N), [d_G, d_N]),$$

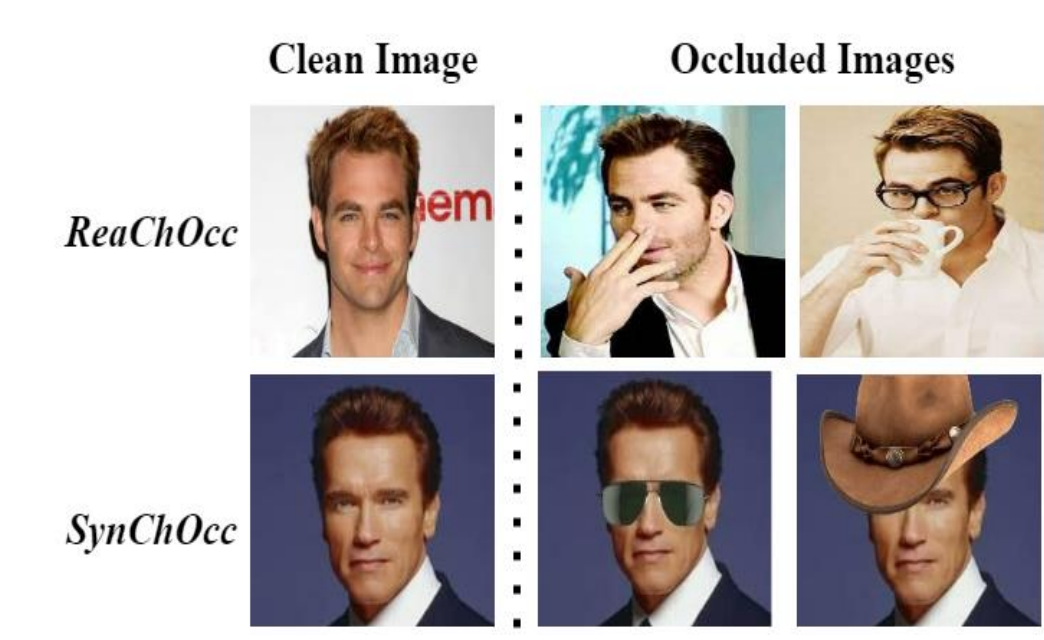
$$\mathcal{L}_C = \mathcal{L}_{CO} + \mathcal{L}_{CN}.$$

$$\mathcal{L}_O = \|I_{O'} - I_G\|, \quad \mathcal{L}_N = \|I_{N'} - I_G\|.$$

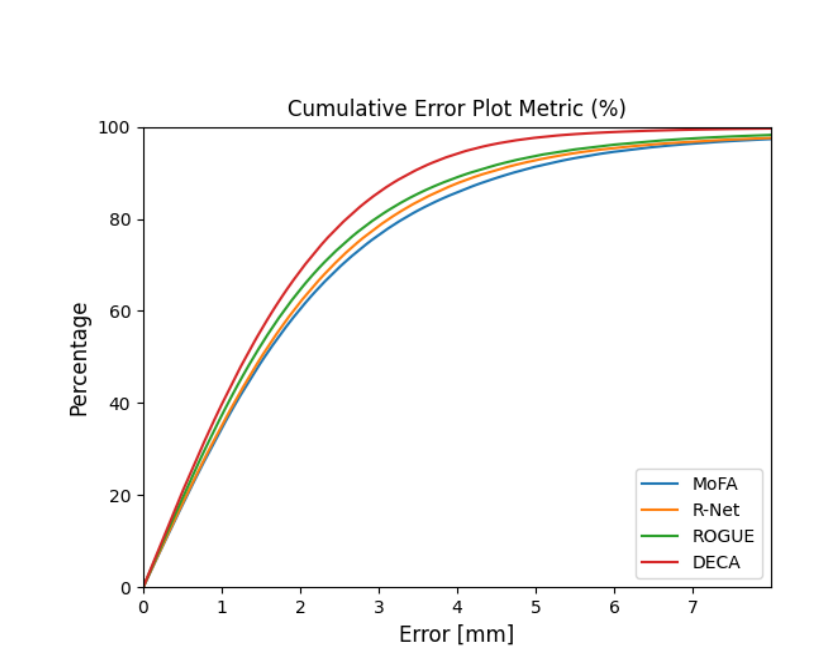
$$\mathcal{L}_{robust} = \beta_O \mathcal{L}_O + \beta_N \mathcal{L}_N - \beta_C \mathcal{L}_C$$

## Results

### Proposed Evaluation Datasets



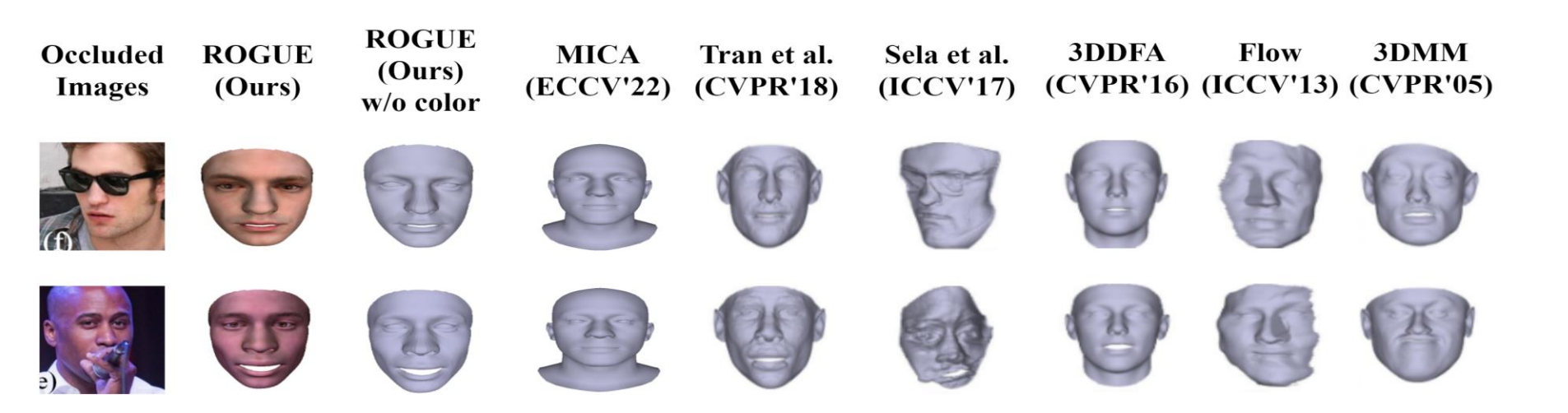
### NoW Challenge



### Perceptual Evaluation

Methods	ReaChOcc		SynChOcc		Noise	
	FaceNet-512	ArcFace	FaceNet-512	ArcFace	FaceNet-512	ArcFace
MoFA (TPAMI' 18)	1.237 ± 0.141	1.313 ± 0.114	1.195 ± 0.126	1.284 ± 0.150	1.245 ± 0.171	1.250 ± 0.274
R-Net (CVPRW' 19)	1.045 ± 0.173	1.188 ± 0.171	0.955 ± 0.187	1.131 ± 0.194	1.161 ± 0.253	1.221 ± 0.217
DECA (TOG' 21)	1.097 ± 0.176	1.196 ± 0.176	0.951 ± 0.184	1.061 ± 0.210	1.167 ± 0.295	1.170 ± 0.298
<b>ROGUE (Ours)</b>	<b>0.943 ± 0.187</b>	<b>1.025 ± 0.168</b>	<b>0.879 ± 0.174</b>	<b>0.983 ± 0.186</b>	<b>0.963 ± 0.185</b>	<b>1.017 ± 0.146</b>

### Shape Comparison



### Shape & Texture Comparison

