HiFace: High-Fidelity 3D Face Reconstruction by Learning Static and Dynamic Details

NUS National University of Singapore · 清華大學 · Microsoft · ICCV23 PARIS

webapge · paper

Zenghao Chai, Tianke Zhang, Tianyu He, Xu Tan, Tadas Baltrušaitis, HsiangTao Wu, Runnan Li, Sheng Zhao, Chun Yuan, Jiang Bian

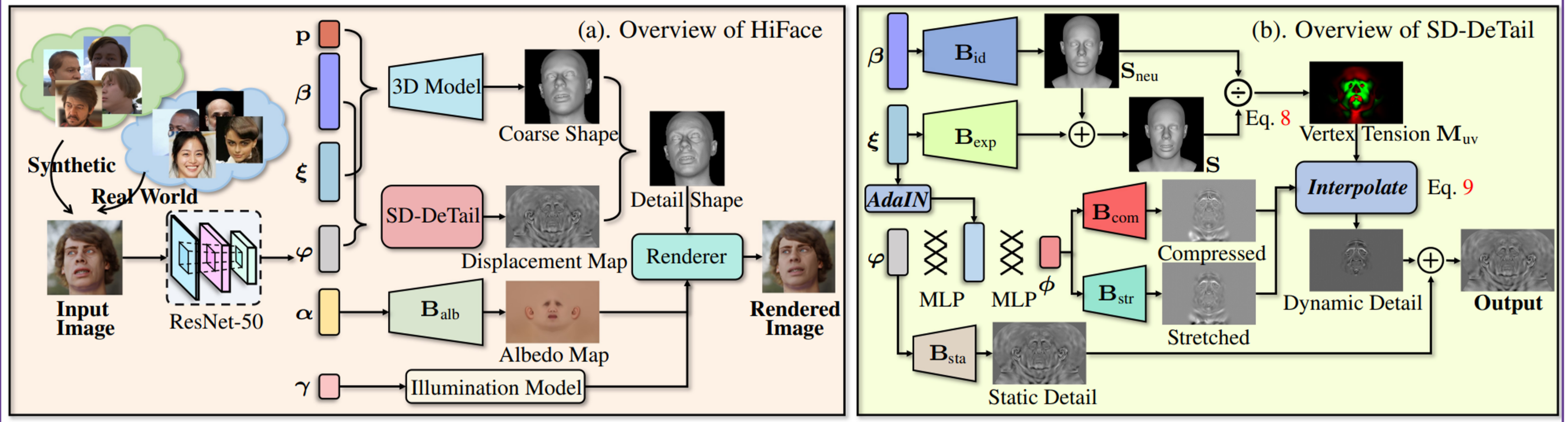
Contributions of HiFace & SD-DeTail

- Model the static and dynamic details explicitly, demonstrate the benefits of synthetic data in detailed 3D face reconstruction.
- Novel loss functions to learn 3D representations of coarse shape and fine details simultaneously from both the synthetic and real-world images.
- SOTA reconstruction quality both quantitatively and qualitatively, with over 15% performance gains in the region-aware benchmark
- Easy to plug-and-play into optimization-based methods and can transfer expressions and details from one to another for face animation.

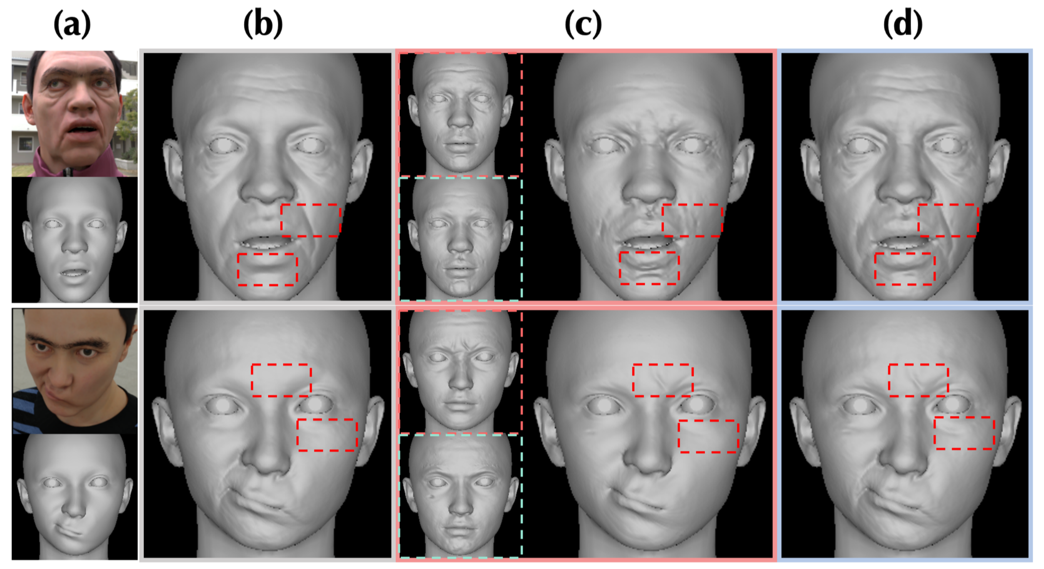
Overview of HiFace

- Reconstruction: given a single image, HiFace faithfully reconstructs a coarse shape (2nd-row) with vivid details (3rd-row).
- Animation: given a source image, HiFace can animate the static (4th-row), dynamic (5th-row), or both (6th-row) details of the driving images.

Model Architecture



p: pose coef β: identity coef ξ: expression coef φ: static coef α: albedo coef γ: SH coef ϕ: dynamic coef



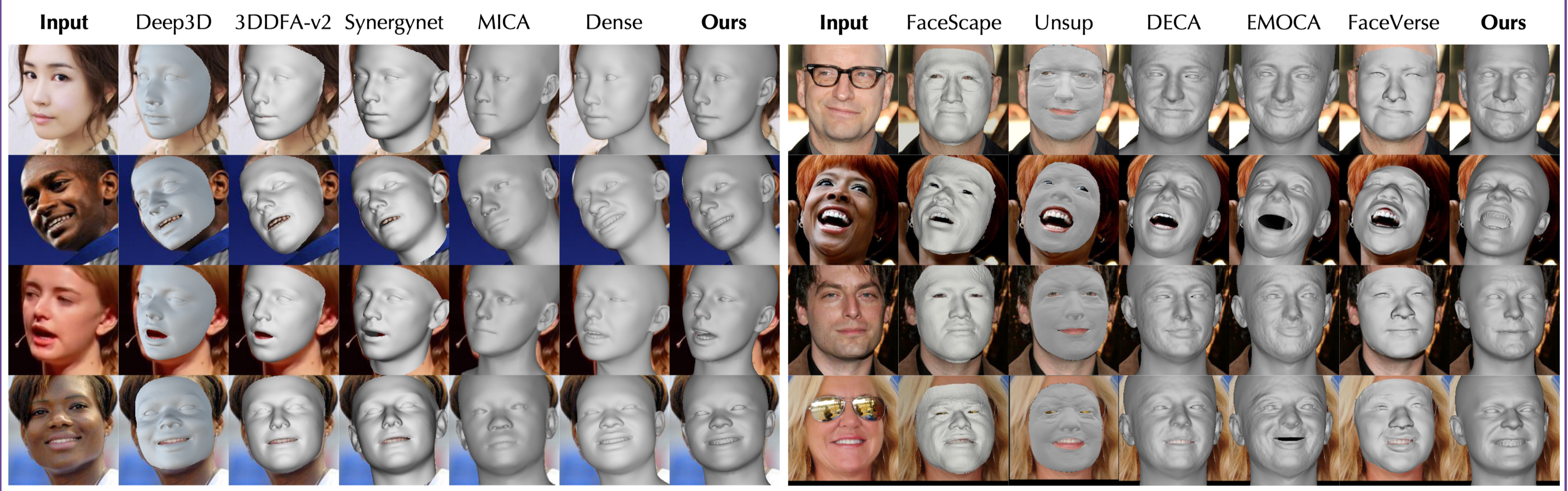
- Given (a). an image (top) to reconstruct its coarse shape (bottom), we formulate the detail as (b). a static factor and (c). a dynamic factor interpolated by polarized states w.r.t. compressed (top) and stretched (bottom). (d). the output displacement map is linearly combined by (b) and (c) to present vivid details.

- Train HiFace with Synthetic Images & Real-world Images

Static and Dynamic Detail Losses $\mathcal{L}_{\rm detail} = \mathcal{L}_{\rm sta} + \mathcal{L}_{\rm com} + \mathcal{L}_{\rm str}$ **Coarse Shape Losses** $\mathcal{L}_{\rm shp} = \mathcal{L}_{\rm ver} + \mathcal{L}_{\rm kl}$

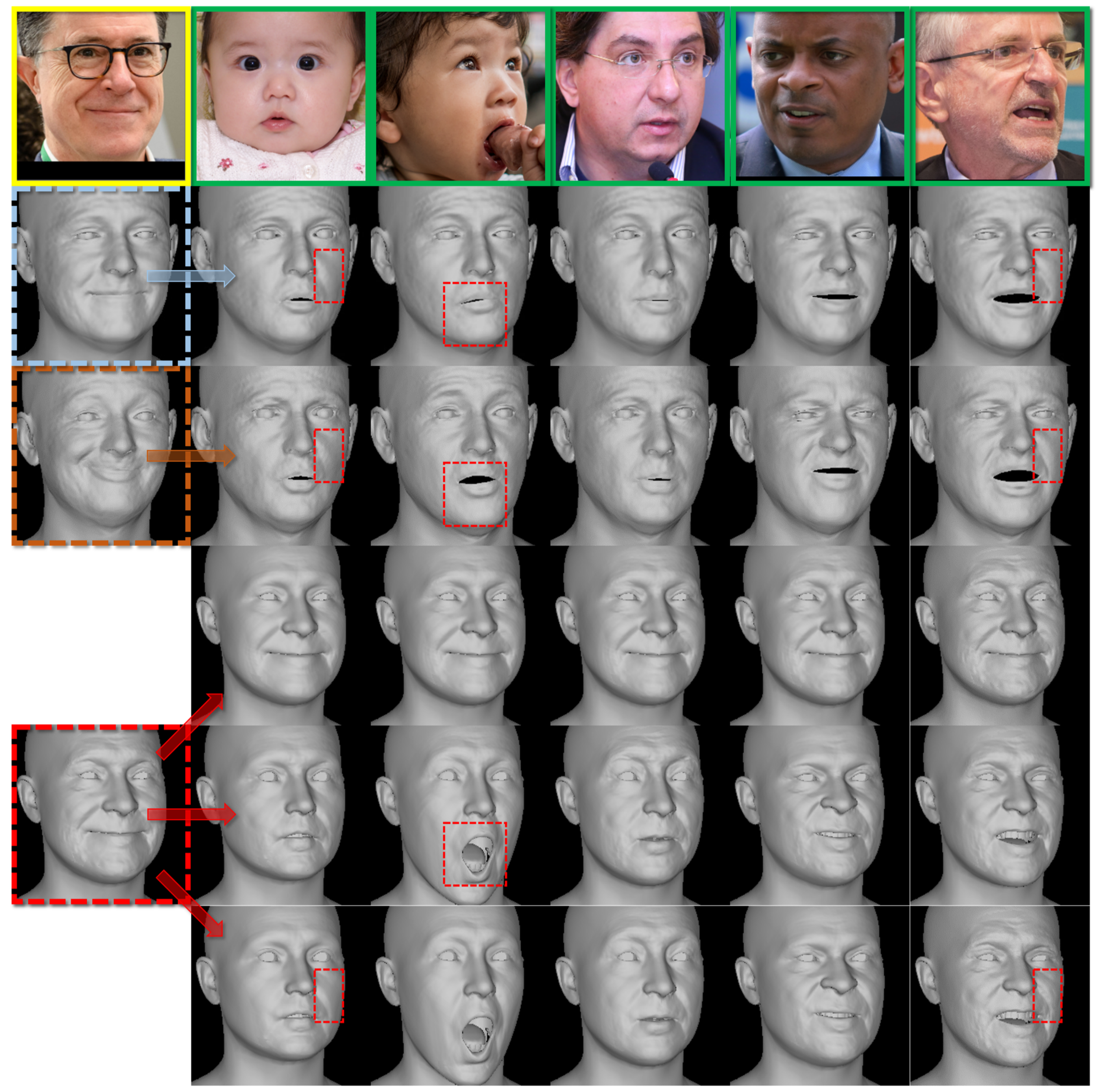
Self-supervised Losses $\mathcal{L}_{\rm self} = \mathcal{L}_{\rm pho} + \lambda_{\rm id}\mathcal{L}_{\rm id} + \lambda_{\rm lmk}\mathcal{L}_{\rm lmk}$ **Knowledge Distillation** $\mathcal{L}_{\rm kd}$ **Regularization** $\mathcal{L}_{\rm reg}$

3D Face Reconstruction Comparisons



Group	Methods / e (mm)	frontal-view @$\mathcal{R}_N$	frontal-view @$\mathcal{R}_M$	frontal-view @$\mathcal{R}_F$	frontal-view @$\mathcal{R}_C$	frontal-view all	side-view @$\mathcal{R}_N$	side-view @$\mathcal{R}_M$	side-view @$\mathcal{R}_F$	side-view @$\mathcal{R}_C$	side-view all
Coarse	Deep3D [21]	1.719±0.354	**1.368±0.439**	2.015±0.449	1.528±0.501	1.657	1.749±0.343	1.411±0.395	2.074±0.486	1.528±0.517	1.691
	MGCNet [64]	1.771±0.380	1.417±0.409	2.268±0.503	1.639±0.650	1.774	1.827±0.383	**1.409±0.418**	2.248±0.508	1.665±0.644	1.787
	3DDFA-v2 [29]	1.903±0.517	1.597±0.478	2.447±0.647	1.757±0.642	1.926	1.883±0.499	1.642±0.501	2.465±0.622	1.781±0.636	1.943
	DECA-c [24]	1.694±0.355	2.516±0.839	2.394±0.576	1.479±0.535	2.010	1.903±1.050	2.472±1.079	2.423±0.720	1.630±1.135	2.107
	SADRNet [61]	1.791±0.542	1.591±0.488	2.413±0.537	1.856±0.701	1.913	1.771±0.521	1.560±0.462	2.490±0.566	2.010±0.715	1.958
	EMOCA-c [19]	1.868±0.387	2.679±1.112	2.426±0.641	1.438±0.501	2.103	1.867±0.554	2.636±1.284	2.448±0.708	1.548±0.590	2.125
	MICA [81]	1.585±0.325	3.478±1.204	2.374±0.683	**1.099±0.324**	2.134	1.525±0.322	3.567±1.212	2.379±0.675	**1.109±0.325**	2.145
	Ours-c (w/o Syn. Data)†	1.227±0.407	1.787±0.439	1.454±0.382	1.762±0.436	1.558	1.187±0.379	1.826±0.490	1.470±0.426	1.653±0.450	1.534
	Ours-c	**1.054±0.317**	1.461±0.430	**1.331±0.347**	1.342±0.384	**1.297**	**0.992±0.246**	1.505±0.454	**1.427±0.400**	1.439±0.429	**1.341**
Detail	DECA-d [24]	2.138±0.461	2.802±0.868	2.457±0.559	1.443±0.498	2.210	2.286±1.103	2.684±1.041	2.519±0.718	1.555±0.822	2.261
	EMOCA-d [19]	2.532±0.539	2.929±1.106	2.595±0.631	1.495±0.469	2.388	2.455±0.636	2.948±1.292	2.606±0.686	1.599±0.563	2.402
	HRN [42]	1.722±0.330	**1.357±0.523**	1.995±0.476	**1.072±0.333**	1.537	1.642±0.310	**1.285±0.528**	1.906±0.479	**1.038±0.322**	1.468
	Ours-d (w/o Syn. Data)†	1.465±0.557	1.790±0.425	1.528±0.373	1.618±0.362	1.600	1.422±0.537	1.849±0.473	1.530±0.414	1.572±0.399	1.594
	Ours-d (w/o static)*	1.055±0.290	1.469±0.415	1.336±0.337	1.319±0.374	1.295	1.004±0.233	1.491±0.437	1.418±0.392	1.418±0.415	1.332
	Ours-d (w/o dynamic)*	1.069±0.318	1.469±0.414	1.358±0.336	1.270±0.344	1.292	0.991±0.239	1.496±0.437	1.411±0.393	1.375±0.402	1.318
	Ours-d	**1.036±0.280**	1.450±0.413	**1.324±0.334**	1.291±0.362	**1.275**	**0.985±0.237**	1.489±0.436	**1.399±0.388**	1.360±0.395	**1.308**

3D Face Animation Comparisons



- Given a source image, we use the driving images to drive its expressions. DECA and EMOCA can animate the expression-driven details but lack realistic.
- As a comparison, HiFace is flexible to animate details from static (4th-row), dynamic (5th-row), or both (6th-row) factors, and presents vivid animation quality with realistic shapes.

Flexibility of SD-DeTail