A Hierarchical Representation Network for Accurate and Detailed Face Reconstruction From In-the-Wild Images

Biwen Lei, Jianqiang Ren, Mengyang Feng, Miaomiao Cui, Xuansong Xie

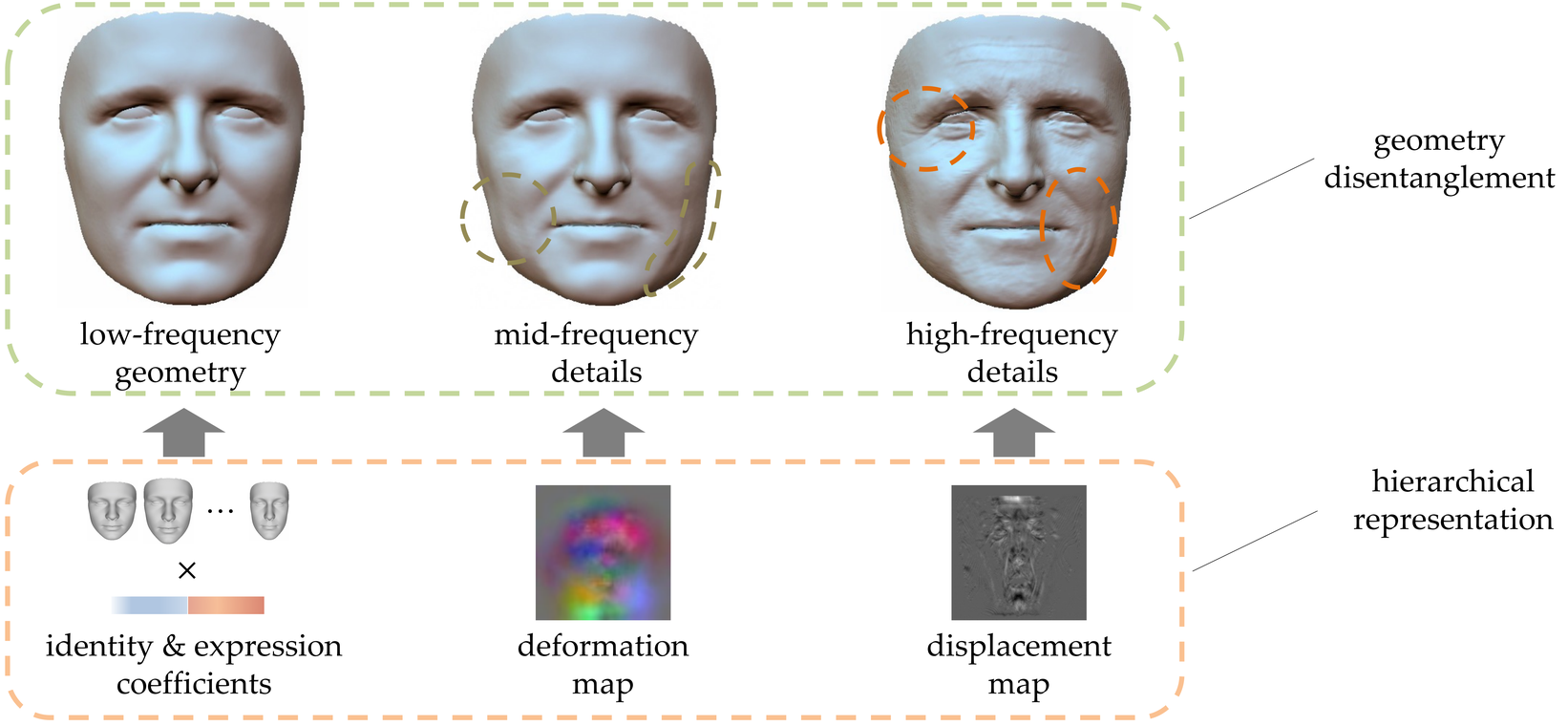
DAMO Academy, Alibaba Group

Paper tag: TUE-AM-038

Quick Preview

The result of traditional 3DMM-based face reconstruction method.

Quick Preview

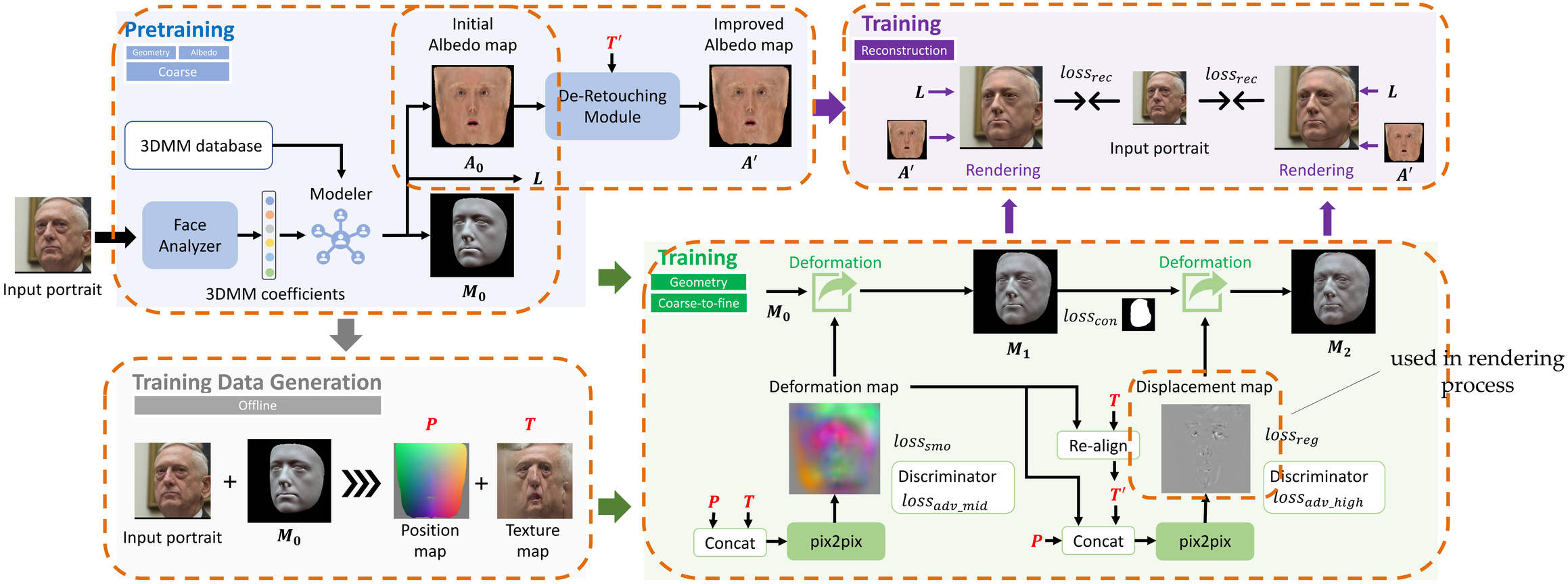


An illustration of the hierarchical modeling.

Quick Preview

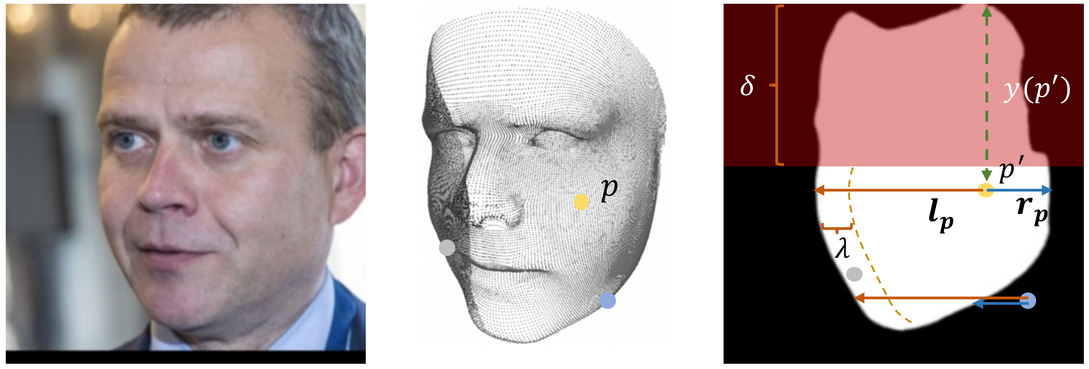
Some results of the proposed model HRN.

Framework



Overview of the proposed hierarchical representation network (HRN).

Contour-aware Loss

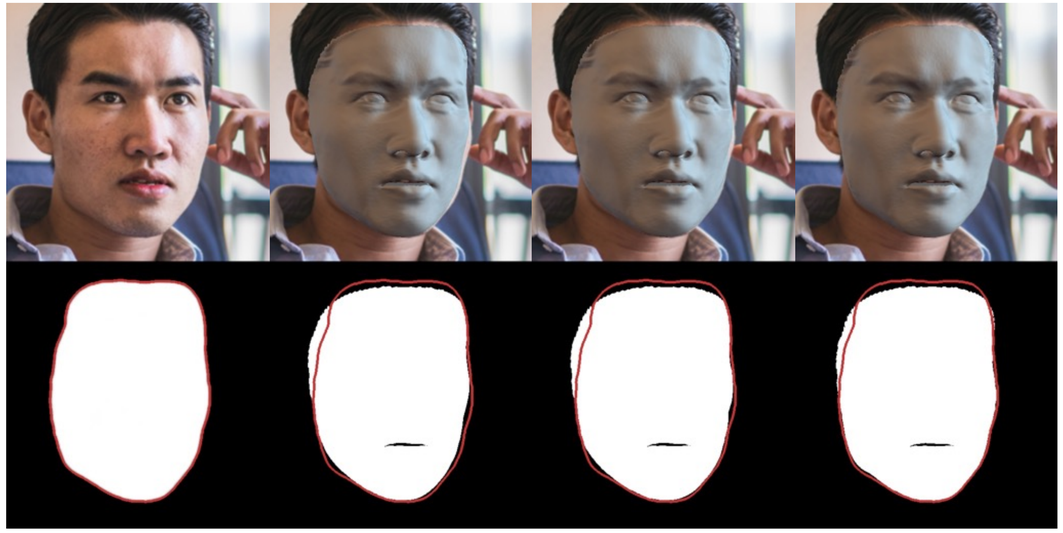


(a) (b) (c)

The details of the proposed contour loss.

$$L_{con} = \frac{1}{N_p} \sum_{p \in M_1} (f(p) \mathbb{1}[y(p') > \delta]),$$

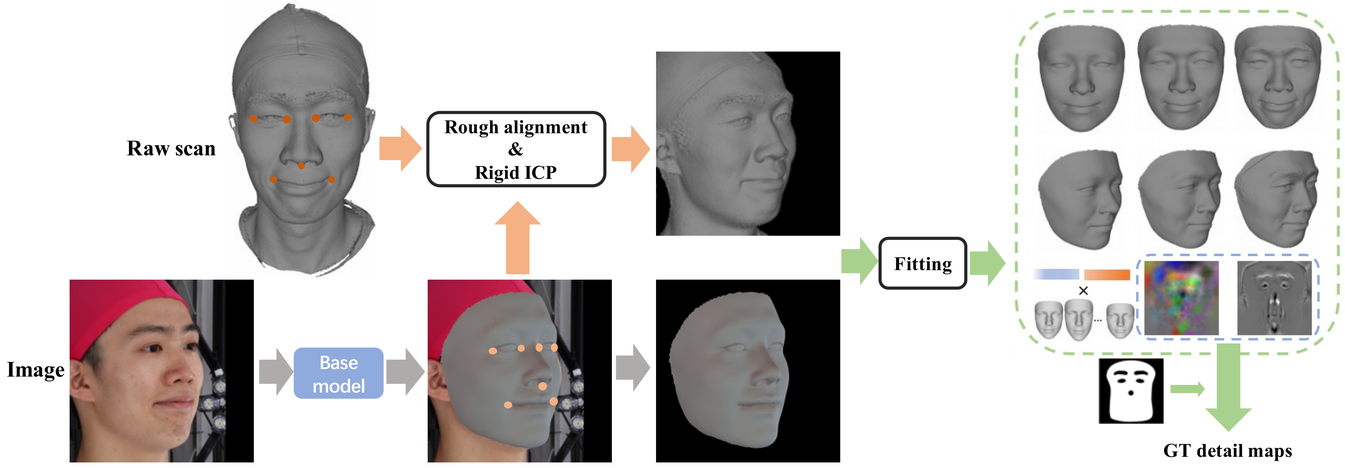
$$f(p) = |h(\frac{\boldsymbol{l_p} \cdot \boldsymbol{r_p}}{max(||\boldsymbol{l_p}||, ||\boldsymbol{r_p}||)} + \lambda) - \lambda|,$$



(a) Input (b) w/o Contour loss (c) BCE loss (d) Ours

Ablation study toward contour loss on FFHQ.

3D Priors of Facial Details

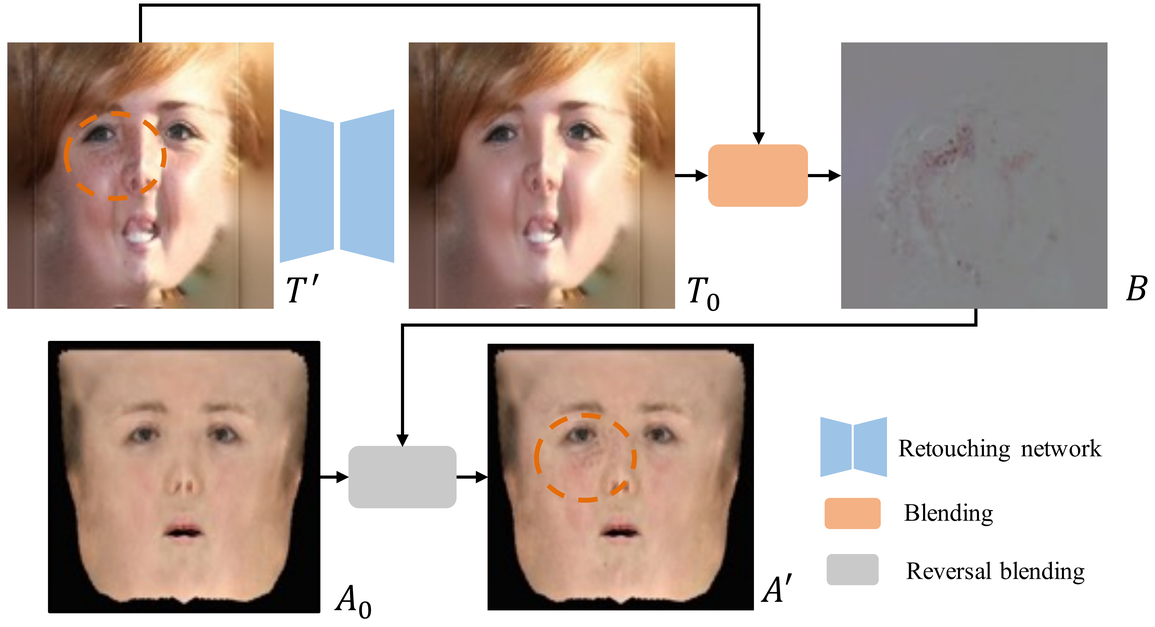


The pipeline of acquiring ground-truth deformation map and displacement map.

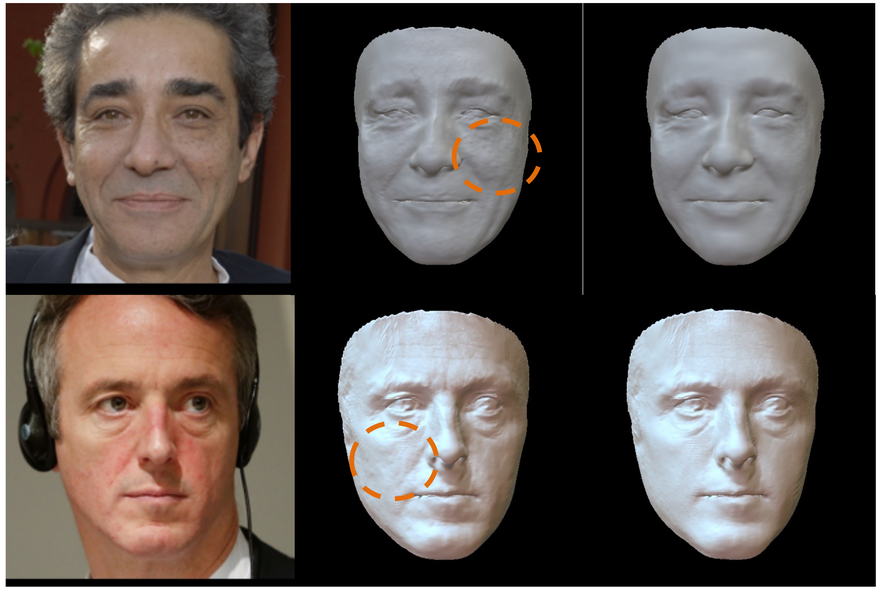
(a) Input (b) w/o 3D priors (c) Ours

Ablation study toward 3D priors on FFHQ.

De-retouching Module



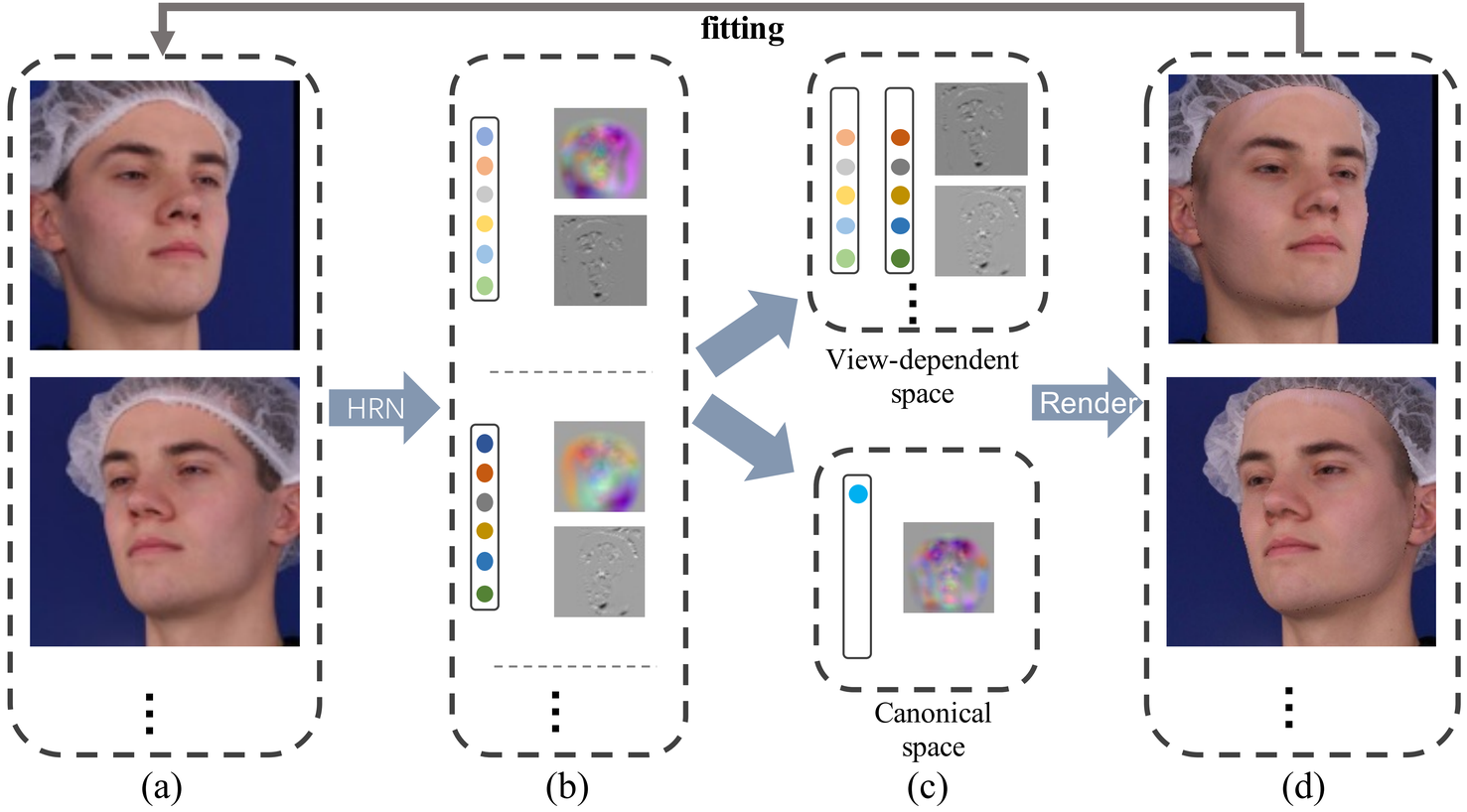
The details of the de-retouching module.



(a) Input (b) w/o De-retouching (c) Ours

Ablation study toward de-retouching module on FFHQ.

Multi-view HRN

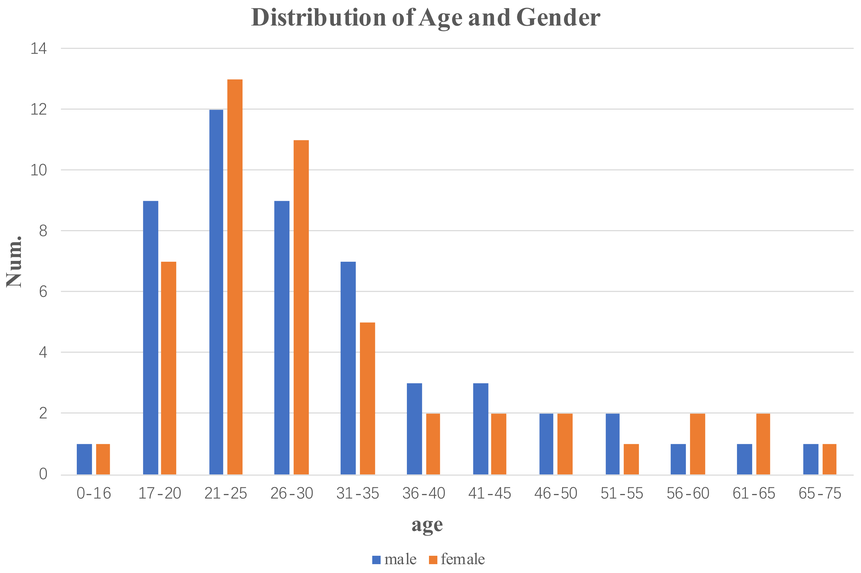


The pipeline of the proposed MV-HRN.

FaceHD-100 Dataset

(a) 9-view images (b) Mesh (c) Textured mesh (d) Texture details

An example from the FaceHD-100 dataset.



The age and gender distribution of the FaceHD-100 dataset.

Comparison with SOTAs (single-view)

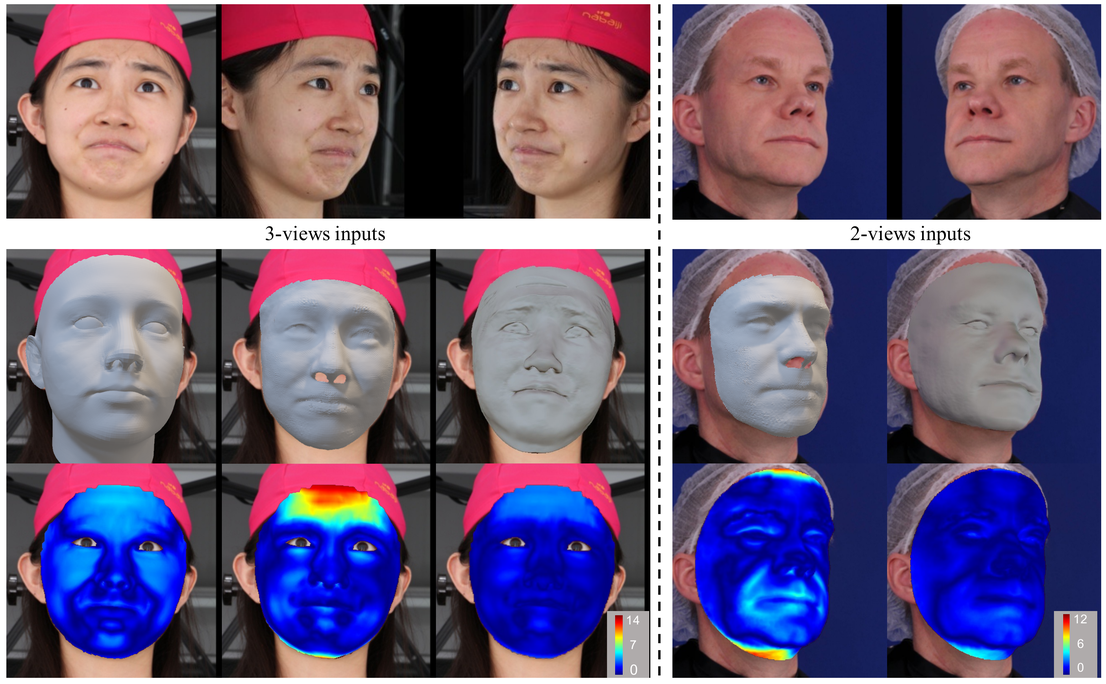
(a) Input (b) SADRNet (c) Deep3D (d) Pix2Vertex (e) DECA (f) Ours

The single-view qualitative comparison.

Table 1. Single-view quantitative comparison. REALY-F and REALY-S denote frontal-view and side-view reconstruction on REALY benchmark respectively.

Methods	FaceScape-wild CD (mm)	FaceScape-wild MNE (rad)	FaceScape-lab CD (mm)	FaceScape-lab MNE (rad)	REALY-F NMSE (mm)	REALY-S NMSE (mm)
Deep3D	3.8	0.092	5.28	0.118	1.657	1.691
MCGNet	3.22	0.077	4.00	0.093	1.774	1.787
PRNet	3.47	0.123	**3.56**	0.126	2.013	2.032
SADRNet	7.12	0.123	6.75	0.133	1.913	1.958
DECA	3.31	0.089	4.69	0.108	2.210	2.261
3DDFA-V2	3.00	0.080	3.60	0.096	1.926	1.943
Ours	**2.91**	**0.065**	3.67	**0.087**	**1.537**	**1.468**

Comparison with SOTAs (multi-view)



(a) MVFNet (b) DFNRMVS (c) Ours (d) DFNRMVS (e) Ours

The multi-view qualitative comparison on FaceScape and ESRC datasets .

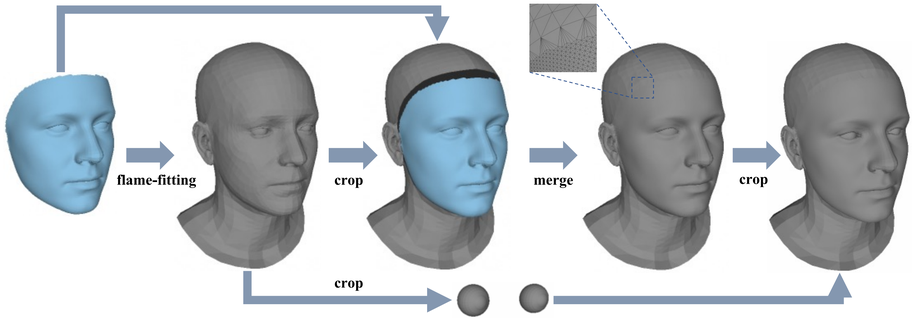
Table 2. Multi-view quantitative comparison. We only report MVFNet performance on FaceScape because its released model cannot process two-view inputs.

Methods	FaceScape (3 views) Median (mm)	Mean (mm)	Std (mm)	ESRC (2 views) Median (mm)	Mean (mm)	Std (mm)
MVFNet	1.76	2.12	**1.66**	N.A.	N.A.	N.A.
DFNRMVS	1.79	2.41	2.61	1.59	2.13	2.29
Ours	**1.13**	**1.51**	1.79	**1.29**	**1.69**	**1.72**

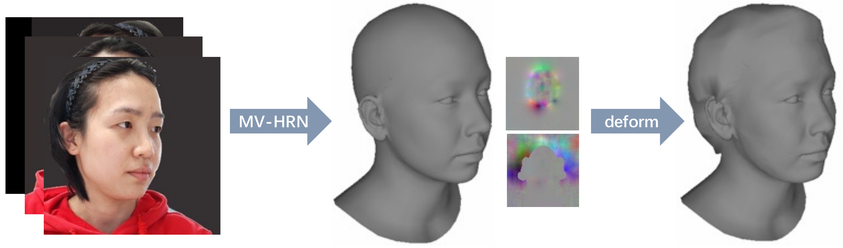
Methods	2 views	3 views	4 views	5 views	naive version (5 views)
Ours	1.17	1.13	1.11	**1.10**	1.23

Table 4. Quantitative ablation study toward sparse-view reconstruction on FaceScape. Only median distance(mm) is reported.

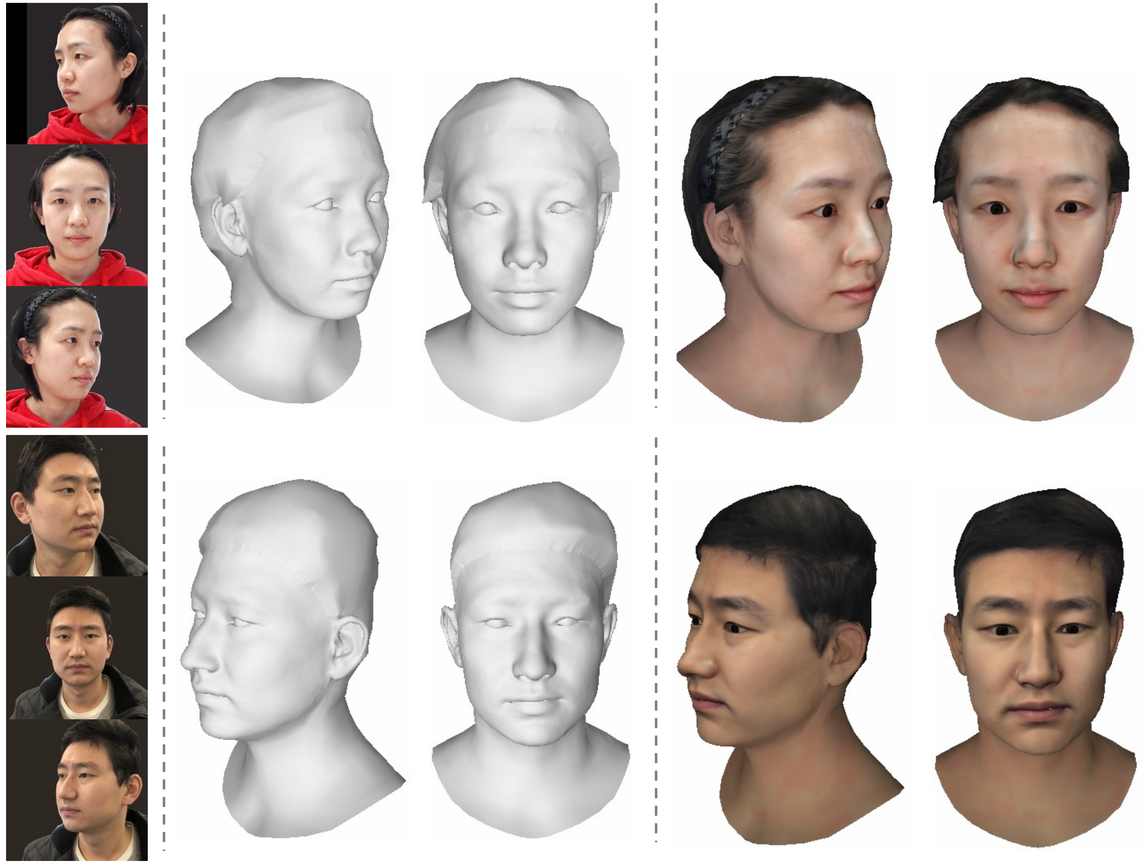
Extention work (head reconstruction)



The pipeline of generating a new head model from BFM and FLAME.



Simplified head reconstruction process.



(a) 3-view images (b) Predicted mesh (c) Textured mesh

Some head reconstruction results of our method on selfie data.

Alibaba Group 阿里巴巴集团 達摩院 ALIBABA DAMO ACADEMY

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Thanks!