WarpHE4D: Dense 4D Head Map toward Full Head Reconstruction

Supplementary Material

1. PCA visualization of Sapiens

As shown in Figure 1, we visualize Sapiens features across various head images. Although Sapiens is trained on a larger human-centric dataset compared to DINOv2, it fails to extract consistent features across different head images. In other words, while Sapiens can extract features that enhance downstream task performance during fine-tuning, it falls short in capturing semantic and 3D-aware features. Consequently, we conclude that Sapiens features cannot serve as a replacement for the template model prior, even without an ablation study.

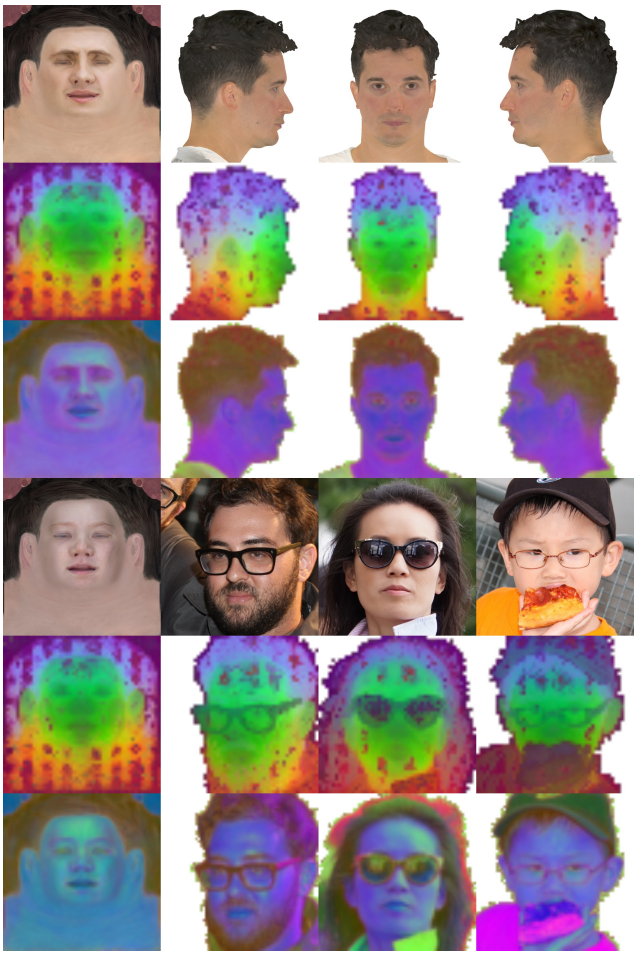

Figure 1. PCA visualization of DinoV2 and Sapiens features. The first, second, and third rows visualize the input images, DINOv2 features, and Sapiens features, respectively, while the fourth, fifth, and sixth rows shows the same results for other samples.

2. Qualitative Results of NoW Challenge

We evaluate our method on the NoW Challenge only qualitatively, as it excludes expressions. As shown in Figure 3, our approach successfully reconstructs the 4D head map even for NoW Challenge images, regardless of extreme head poses.

3. More Qualitative Results

We provide additional qualitative results on extreme images and FFHQ images in Figure 2 and Figure 4. Our method demonstrates consistent performance regardless of skin color, age, or gender. Interestingly, it also accurately distinguishes non-head regions despite the absence of such data in the training set. This further highlights the descriptive power of DINOv2 features, enabling generalization beyond the training data domain.

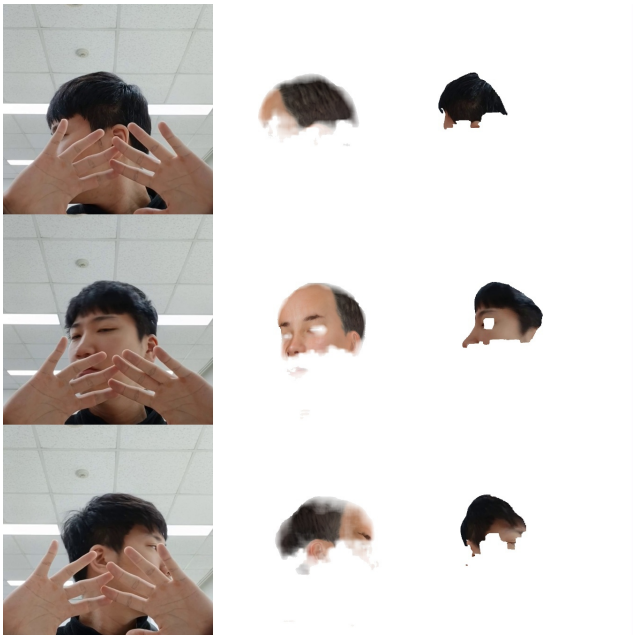

Figure 2. Qualitative results of extreme images. UV results are visualized by texture warping and masked by confidence.

4. More Quantitative Results

The results on the benchmarks can be evaluated across various categories. First, we present the FaceScape benchmark results categorized by pose angle and focal length in Tables 4, 5, and 6. Additionally, in Tables 7 and 3, we report

Figure 3. Qualitative results of NoW challenge. UV, depth and confidence are visualized as UV-colored meshes and mask respectively

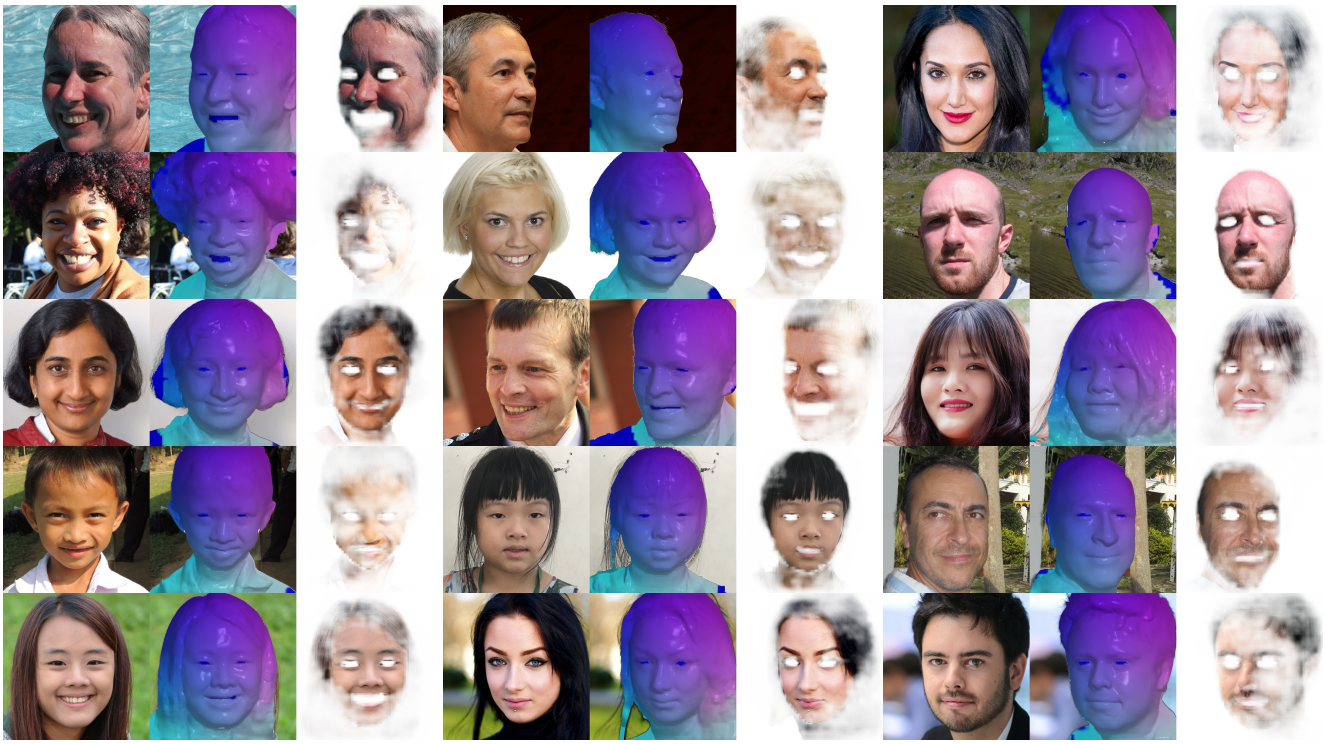

Figure 4. Qualitative results of in-the-wild images. UV, depth and confidence are visualized as UV-colored meshes and mask respectively

the FaceVerse benchmark results categorized by region and focal length, respectively.

5. Benchmark Details

Due to the large scale of the Multiface and FaceVerse datasets, using them entirely as benchmark data is not feasible. Therefore, we created the benchmark through subsampling. For the Multiface dataset, as shown in Table 1, we selected 10 subjects from the 13 available, along with 51 expressions and 38 cameras. For the FaceVerse dataset, as shown in Table 2, we used all subjects, but limited the selection to 15 expressions and 105 rendering cameras.

6. Mesh Registration Details

Our training dataset and benchmarks are rendered using registered meshes. For each scan in the dataset, we processed it through three steps: Procrustes alignment, parametric fitting, and non-rigid fitting. First, we aligned the mean shape of FLAME to each scan using 68 keypoints. If 3D keypoints are not available, we back-projected 2D keypoints from the rendered image to obtain them. During this stage, the scale, rotation, and translation between FLAME and the scan are optimized. Second, we optimized the FLAME parameters based on chamfer distance and 3D keypoint distance. Next, we removed the eyeballs and subdivided FLAME twice, increasing its density by a factor of 16. Finally, we assigned per-vertex displacement to the subdivided FLAME and apply non-rigid fitting. When matching the nearest vertex between the scan and the subdivided FLAME, we masked the distances using cosine similarity. Specifically, when the magnitude of cosine similarity between correspondences exceeds 0.3, we set the corresponding distance to infinity. This masking process ensures that the nearest point is found along the normal direction, which improves the non-rigid fitting performance, especially in areas with large variations, such as hair. Additionally, when vertex correspondences between the scan and FLAME are available, as in the case of the multiface tracked mesh, we optimized the distance based on the pre-built correspondence rather than the nearest point. An example of our registration on Multiface dataset is shown in Figure 5.

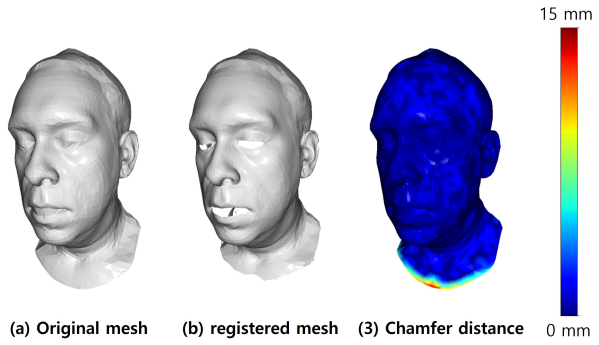

Figure 5. Retopology error of Multiface dataset. From left to right, original mesh, registered mesh into FLAME topology, and colorized retopology error in mm scale(max. 15mm) are visualized. Mean error over all meshes is 0.06mm.

Table 1. Lists of sampled categories from Multiface dataset

Multiface Benchmark							
Subject	Expression				Camera		
m–20171024–0000–002757580–GHS	E004	E018	E031	E044	400002	400023	400048
m–20180105–0000–002539136–GHS	E006	E019	E032	E045	400004	400024	400049
m–20180226–0000–6674443–GHS	E007	E020	E033	E046	400006	400026	400050
m–20180227–0000–6795937–GHS	E008	E021	E034	E056	400007	400027	400051
m–20180406–0000–8870559–GHS	E009	E022	E035	E057	400009	400028	400053
m–20180418–0000–2183941–GHS	E010	E023	E036	E058	400010	400029	400059
m–20180426–0000–002643814–GHS	E011	E024	E037	E059	400012	400030	400060
m–20180510–0000–5372021–GHS	E012	E025	E038	E060	400013	400031	400061
m–20180927–0000–7889059–GHS	E013	E026	E039	E061	400014	400035	400063
m–20181017–0000–002914589–GHS	E014	E027	E040	E064	400015	400037	400064
	E015	E028	E041	E065	400016	400039	400067
	E016	E029	E042	EXP_ROM07	400018	400041	400069
	E017	E030	E043		400019	400042	

Table 2. Lists of augmented categories from FaceVerse dataset

Faceverse Benchmark					
Subject	Expression	Focal length	Camera distance	Azimuth	Elevation)
001 - 110 (all)	01 15 02 16 04 17 06 18 07 19 08 20 10 21 13	Short 300 Mid 600 Long 1200	1 m 3 m 5 m	-90° -60° -30° 0° 30° 60° 90°	-30° -15° 0° 15° 30°

Table 3. Quantitative evaluation on Faceverse benchmark categorized by focal length.

Focal Length →	Short(300)		Mid(600)		Long(1200)	
Method ↓	UVAUC	3DAUC	UVAUC	3DAUC	UVAUC	3DAUC
PRNet	0.5772	0.6124	0.4034	0.4201	0.4608	0.4751
3DDFA v2	0.6380	0.6581	0.4137	0.4203	0.4784	0.4848
3DDFA v3	0.6562	0.7072	0.4039	0.4230	0.4636	0.4862
HRN	0.6424	0.6895	0.4031	0.4180	0.4641	0.4795
DECA	0.7288	0.7616	0.6604	0.6742	0.7043	0.7197
EMOCA	0.7153	0.7538	0.6530	0.6722	0.6946	0.7177
Ours	0.8774	0.8381	0.8546	0.8510	0.8697	0.8659

Table 4. Quantitative evaluation on FS-Wild dataset categorized by pose angle.

Pose Angle →	0° – 5°			5° – 30°			30° – 60°			60° – 90°			Success Rate (%)
Method ↓	CD	MNE	CR	CD	MNE	CR	CD	MNE	CR	CD	MNE	CR	
Ext3dFace	4.96	0.155	61.7	5.58	0.174	56.1	7.55	0.205	40.9	26.03	0.275	27.0	85.3
PRNet	2.64	0.116	83.3	3.07	0.113	82.9	4.28	0.118	78.6	3.88	0.141	75.2	100.0
Deep3DFaceRec	3.21	0.146	78.0	4.13	0.101	77.4	5.74	0.113	71.3	8.91	0.168	57.3	100.0
RingNet	2.41	0.082	99.8	2.98	0.090	98.4	4.86	0.100	94.1	10.79	0.170	96.8	100.0
DFDN	3.65	0.091	86.9	3.90	0.092	86.7	7.28	0.132	84.7	27.03	0.302	57.0	88.8
DF2Net	2.96	0.116	81.8	3.58	0.117	75.2	6.73	0.158	46.6	21.27	0.344	25.8	74.9
UDL	2.25	0.084	69.1	2.50	0.088	68.2	3.38	0.101	65.1	6.31	0.176	49.0	87.0
FaceScape (Opti.)	1.97	0.084	98.7	2.77	0.090	82.7	3.97	0.116	66.7	9.15	0.224	66.7	95.5
FaceScape (Learn)	2.63	0.085	90.2	3.22	0.099	86.5	4.11	0.090	85.3	9.09	0.151	70.8	100.0
MGCNet	3.11	0.072	84.6	2.92	0.072	84.6	2.85	0.070	81.6	4.20	0.091	74.4	100.0
3DDFA_v2	2.56	0.075	86.2	2.69	0.075	85.1	3.68	0.093	80.6	3.93	0.080	80.0	100.0
SADRNet	6.69	0.101	92.3	6.09	0.106	84.5	6.90	0.163	82.7	10.63	0.274	82.7	100.0
LAP	4.07	0.107	92.8	4.12	0.123	87.3	6.91	0.147	87.2	10.06	0.202	68.5	100.0
DECA	2.92	0.079	99.9	2.65	0.078	99.9	2.91	0.080	99.9	4.81	0.116	99.7	100.0
Ours	2.48	0.082	78.9	2.59	0.087	75.8	2.63	0.097	66.1	3.58	0.110	65.8	100.0
Ours(w/o kp loss)	2.60	0.082	78.4	2.79	0.088	75.4	2.73	0.095	65.4	3.72	0.110	41.6	100.0
Ours(w/ vgg)	5.54	0.127	70.1	5.86	0.147	67.1	6.73	0.168	62.2	9.89	0.264	59.6	100.0
Ours(w/ sapiens)	5.04	0.123	65.6	6.00	0.146	61.5	6.45	0.162	55.1	10.12	0.219	37.0	100.0

Table 5. Quantitative evaluation on FS-Lab dataset categorized by pose angle.

Pose Angle →	0°			30°			60°			Success Rate (%)
Method ↓	CD	MNE	CR	CD	MNE	CR	CD	MNE	CR	
Ext3dFace	4.49	0.131	86.2	7.42	0.170	69.1	8.51	0.175	55.2	85.9
PRNet	2.94	0.133	92.5	3.40	0.125	90.1	3.74	0.122	85.2	100.0
Deep3DFaceRec	3.99	0.106	87.6	5.90	0.120	81.3	5.55	0.137	75.3	98.9
RingNet	3.62	0.102	99.9	5.03	0.111	99.7	6.82	0.151	94.5	100.0
DFDN	4.28	0.111	98.4	6.71	0.132	95.2	23.63	0.280	81.0	94.7
DF2Net	4.48	0.152	64.1	7.64	0.200	52.2	-1.00	-1.00	-100.0	73.6
UDL	2.21	0.092	79.5	5.34	0.123	71.3	5.63	0.167	61.9	87.0
FaceScape (Opti.)	3.21	0.090	94.2	4.87	0.119	86.2	4.68	0.146	81.7	92.0
FaceScape (Learn)	2.40	0.086	96.7	7.28	0.124	87.7	3.87	0.108	90.5	100.0
MGCNet	3.45	0.085	92.7	3.91	0.093	90.1	3.65	0.090	83.2	100.0
3DDFA_v2	3.05	0.093	95.2	3.41	0.096	93.8	3.82	0.097	88.2	100.0
SADRNet	4.25	0.109	95.8	7.07	0.137	94.9	7.09	0.148	87.6	100.0
LAP	4.27	0.112	96.4	7.33	0.149	93.2	8.70	0.195	85.6	99.2
DECA	3.30	0.093	99.8	4.14	0.100	99.4	4.20	0.107	97.1	100.0
Ours	1.39	0.062	87.1	1.45	0.069	82.2	1.66	0.078	65.5	100.0

Table 6. Quantitative evaluation on FS-Lab dataset categorized by focal length.

Pose Angle → / Method ↓	Long(1200)			Mid(600)			Short(300)			Success Rate (%)
	CD	MNE	CR	CD	MNE	CR	CD	MNE	CR	
Ext3dFace	7.25	0.167	69.4.2	6.72	0.162	64.9	6.03	0.160	61.4	85.9
PRNet	3.42	0.125	89.4	3.48	0.124	89.0	3.79	0.128	90.2	100.0
Deep3DFaceRec	5.67	0.122	80.8	5.28	0.117	79.2	4.90	0.114	81.1	98.9
RingNet	5.23	0.117	98.8	5.25	0.117	99.4	5.37	0.119	99.8	100.0
DFDN	9.05	0.153	93.3	9.40	0.149	92.8	9.30	0.146	94.6	94.7
DF2Net	7.26	0.194	53.7	6.97	0.191	51.2	6.39	0.183	49.5	73.6
UDL	5.06	0.126	70.9	4.91	0.124	69.2	4.95	0.125	69.7	87.0
FaceScape (Opti.)	4.69	0.120	86.4	4.77	0.121	85.2	5.47	0.126	83.6	92.0
FaceScape (Learn)	6.21	0.118	89.0	6.19	0.118	88.8	6.43	0.125	86.4	100.0
MGCNet	3.82	0.091	89.1.7	4.01	0.091	89.4	4.18	0.098	91.3	100.0
3DDFA_v2	3.45	0.096	92.9	3.51	0.094	92.7	3.85	0.097	93.5	100.0
SADRNet	6.81	0.137	93.6	6.82	0.132	95.0	6.60	0.131	97.1	100.0
LAP	7.30	0.154	92.1	7.06	0.151	91.4	6.75	0.150	91.7	99.2
DECA	4.07	0.100	99.0	4.19	0.101	99.6	5.81	0.122	99.8	100.0
Ours	1.48	0.070	79.7	1.42	0.070	75.9	1.42	0.072	73.9	100.0

Table 7. Quantitative evaluation on FaceVerse benchmark categorized by region.

Eval. Region → / Method ↓	Face					Non-Facial				
	UVAUC	3DAUC	CR	MPUVE	MP3DE	UVAUC	3DAUC	CR	MPUVE	MP3DE
PRNet	0.7481	0.7759	0.9102	0.0228	0.0076	0.1201	0.1367	0.1783	0.0722	0.0142
3ddfav2	0.7779	0.7949	0.9257	0.0209	0.0071	0.1387	0.1505	0.1758	0.0637	0.0131
3ddfav3	0.7742	0.8182	0.9236	0.0199	0.0054	0.1357	0.1533	0.1632	0.0603	0.0113
HRN	0.7825	0.8190	0.9252	0.0206	0.0056	0.1169	0.1333	0.1435	0.0690	0.0118
DECA	0.8227	0.8569	0.9797	0.0213	0.0067	0.4976	0.5128	0.7910	0.0806	0.0187
EMOCA	0.8071	0.8491	0.9782	0.0232	0.0071	0.4951	0.5139	0.7892	0.0806	0.0185
Ours	**0.9119**	**0.905**	**1.0000**	**0.0126**	**0.0051**	**0.782**	**0.7699**	**1.0000**	**0.0554**	**0.0116**