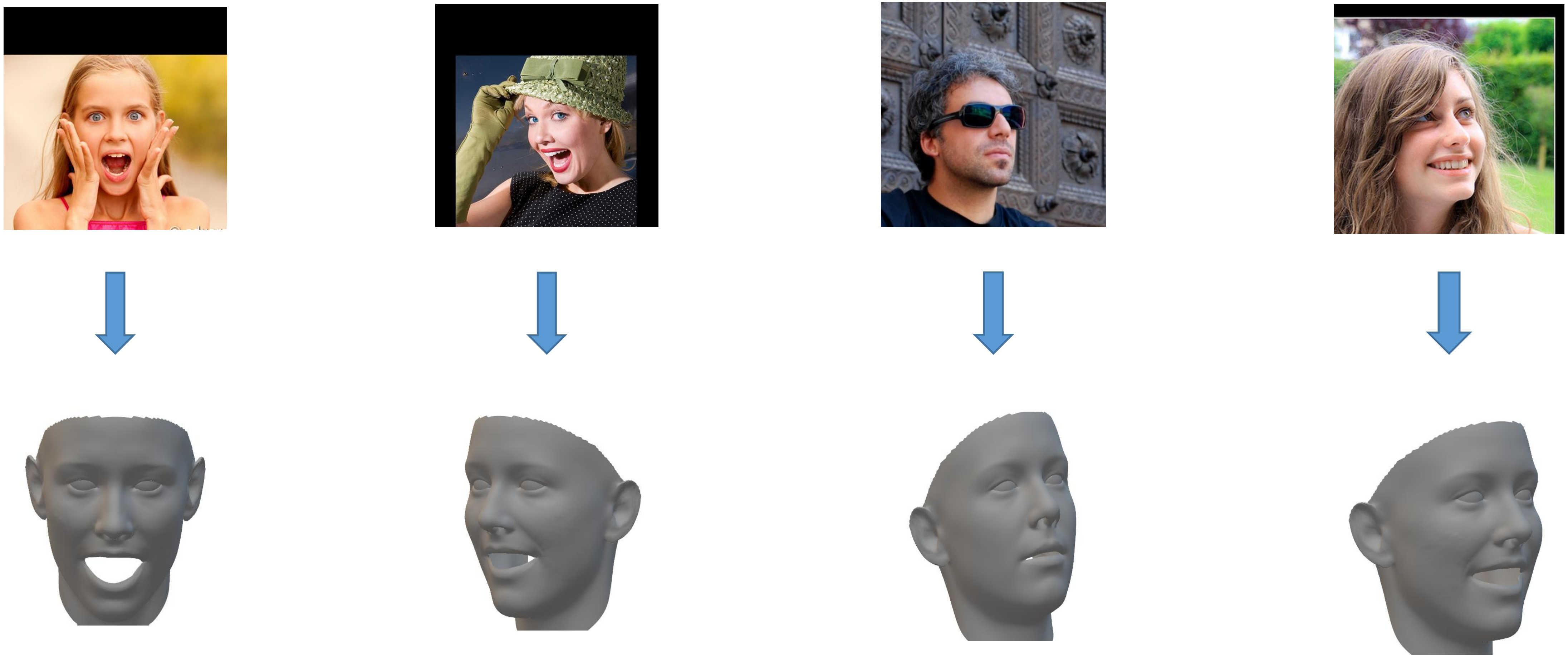
Multi-Attribute Regression Network for Face Reconstruction

XiangZhengLi, Suping*Wu

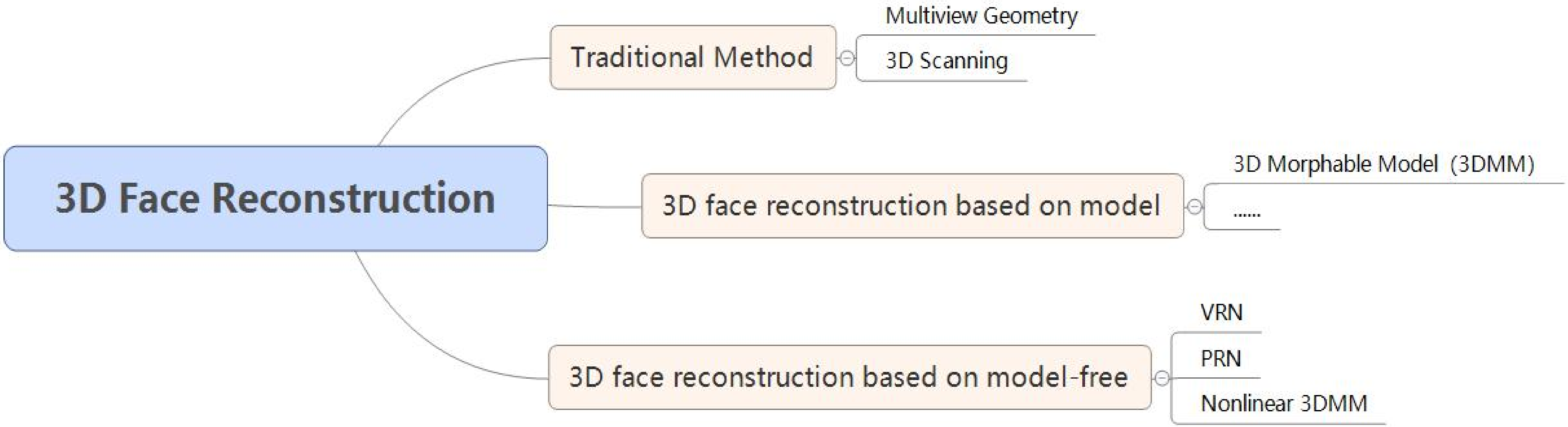
School of Information Engineering, Ningxia University

lxz_nxu@163.com wspg123@163.com

Main task



Related Work



Issue

1. The traditional 3DMM parametric regression method does not distinguish the learning of identity, expression, and attitude attributes, resulting in lacking geometric details in the reconstructed face.

2. The face reconstruction of a single unconstrained image is still challenging, especially in those cases when faces undergo large variations including severe poses, extreme expressions and partial occlusions in unconstrained environments.

Main Contribution of Our Method

(1) We use the multi-attribute regression method to fully capture feature information of pose, expression, and identity attribute to improve the networks ability to learn each attribute. At the same time, we introduce three loss functions to constrain the above three face attributes respectively.

(2) In order to enhance the network's learning of geometric contours, we design a geometric contour constraint loss function, which uses the constraints of sparse face landmarks to improve the reconstructed geometric contour accuracy.

(3) Extensive evaluation on the AFLW2000-3D and AFLW datasets shows that our method achieves excellent performance on both tasks of 3D face reconstruction and dense face alignment.

Proposed Framework

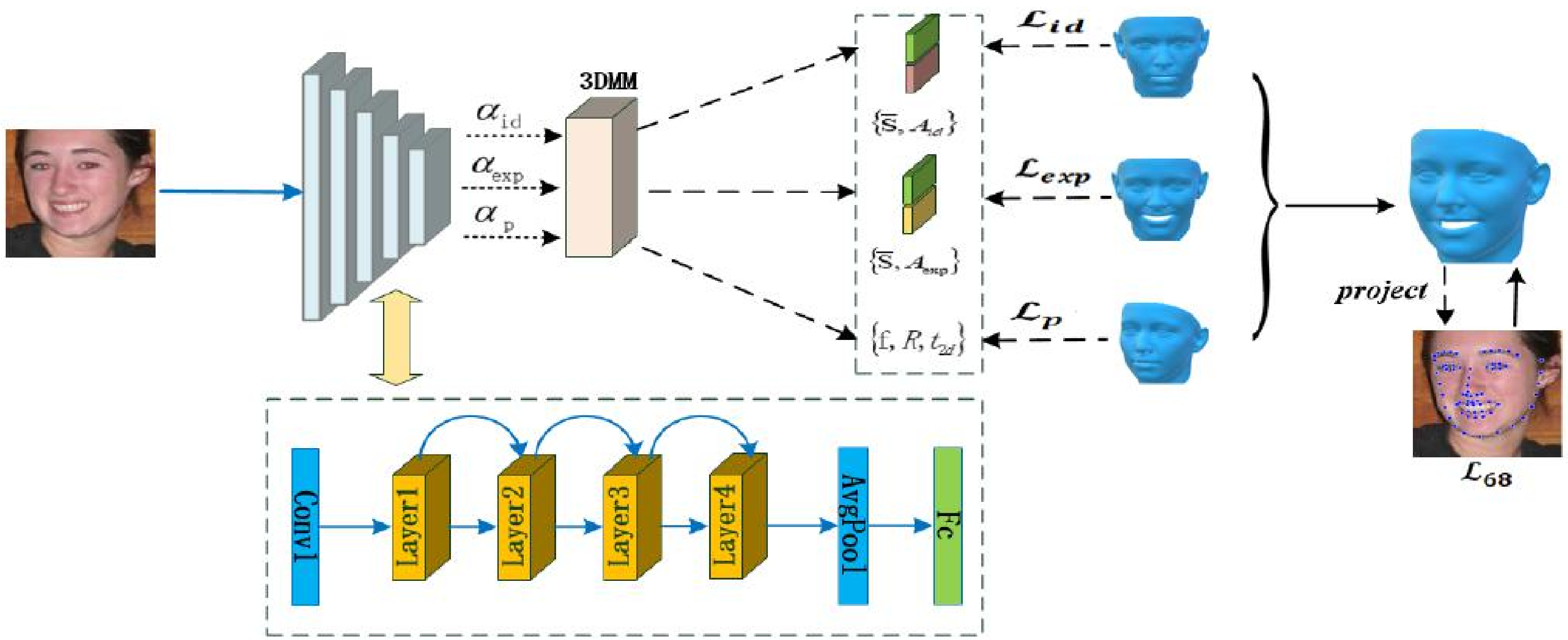


Fig. 2. The workflow of the proposed MARN framework. After extracting features of a face image by a convolutional neural network, it then predicts 62 combined parameters (identity, expression, pose) through a fully connected layer, and uses 3DMM to obtain a geometric model of the predicted face. $\mathcal{L}_{id}$, $\mathcal{L}_{exp}$, $\mathcal{L}_{p}$ are loss functions representing identity, expression, and pose attributes, respectively. These three loss functions can constrain the network to mine more face attribute information and improve the network's ability to learn each attribute. We use the predicted 3D face model to obtain the predicted landmark of the face by means of weak perspective projection. Between the predicted face landmark and the ground truth, we use the Euclidean distance loss $\mathcal{L}_{68}$ to constrain the geometric contour of the face.

Multi-Attribute Loss Function

Identity loss $\mathcal{L}_{id}$:

$$\begin{aligned}\mathcal{L}_{id} &= \left\|S\left(\alpha_{id}, \tilde{\alpha}_{exp}\right) - S\left(\tilde{\alpha}_{id}, \tilde{\alpha}_{exp}\right)\right\|^2 \\ &= \left\|A_{id}\left(\alpha_{id} - \tilde{\alpha}_{id}\right)\right\|^2\end{aligned}$$

Expression loss $\mathcal{L}_{exp}$:

$$\mathcal{L}_{exp} = \left\|A_{\exp}\left(\alpha_{\exp} - \tilde{\alpha}_{\exp}\right)\right\|^2$$

Pose loss $\mathcal{L}_{p}$:

$$\begin{aligned}\mathcal{L}_{p} = \| & \left(f * P_r * R * S\left(\tilde{\alpha}_{id}, \tilde{\alpha}_{exp}\right) + t_{2d}\right) - \\ & \left(\tilde{f} * P_r * \tilde{R} * S\left(\tilde{\alpha}_{id}, \tilde{\alpha}_{exp}\right) + \tilde{t}_{2d}\right)\|^2\end{aligned}$$

Geometric contour:

$$\begin{aligned}\mathcal{L}_{68} = \| & \left(f * P_r * R * S_{68}\left(\alpha_{id68}, \alpha_{exp68}\right) + t_{2d}\right) - \\ & \left(\tilde{f} * P_r * \tilde{R} * S_{68}\left(\tilde{\alpha}_{id68}, \tilde{\alpha}_{exp68}\right) + \tilde{t}_{2d}\right)\|^2\end{aligned}$$

Final loss function:

$$\mathcal{L} = \lambda_{id}\mathcal{L}_{id} + \lambda_{exp}\mathcal{L}_{exp} + \lambda_{p}\mathcal{L}_{p} + \lambda_{68}\mathcal{L}_{68}$$

Face Alignment Quantitative Results

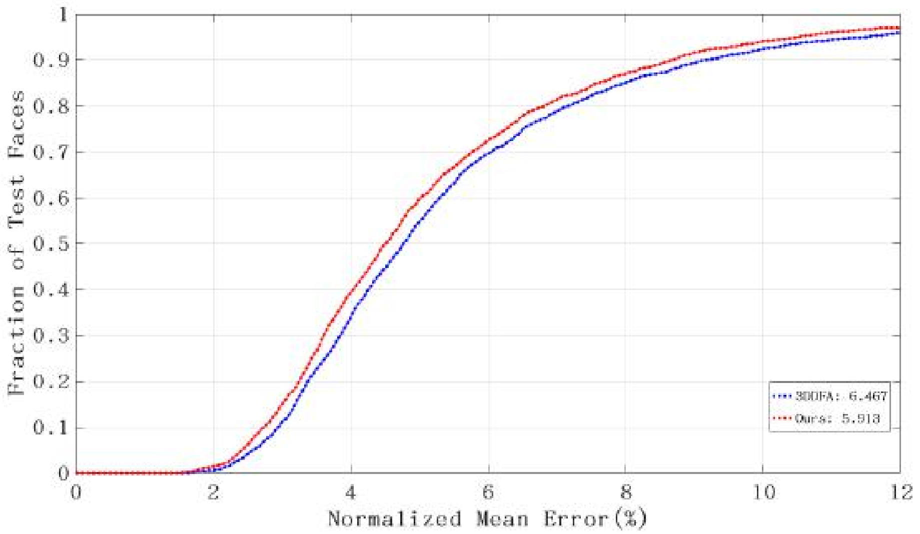
THE NME(%) OF FACE ALIGNMENT RESULTS ON AFLW AND AFLW2000-3D

	AFLW Dataset (21 pts)					AFLW2000-3D Dataset (68 pts)				
Method	[0,30]	[30,60]	[60,90]	Mean	Std	[0,30]	[30,60]	[60,90]	Mean	Std
CDM [20]	8.150	13.020	16.170	12.440	4.040	-	-	-	-	-
RCPR [21]	5.430	6.580	11.530	7.850	3.240	4.260	5.960	13.180	7.800	4.740
ESR [22]	5.660	7.120	11.940	8.240	3.290	4.600	6.700	12.670	7.990	4.190
SDM [23]	4.750	5.550	9.340	6.550	2.450	3.670	4.940	9.760	6.120	3.210
DEFA [11]	-	-	-	-	-	4.500	5.560	7.330	5.803	1.169
3DDFA(CVPR2016) [5]	5.000	5.060	6.740	5.600	0.990	3.780	4.540	7.930	5.420	2.210
Yu et al.(ICCV2017) [25]	5.940	6.480	7.960	-	-	3.620	6.060	9.560	-	-
Nonlinear(CVPR2018) [6]	-	-	-	-	-	-	-	-	4.700	-
DAMDNet(ICCVW19) [24]	4.359	5.209	6.028	5.199	0.682	**2.907**	3.830	4.953	3.897	0.837
MARN(Ours)	**4.306**	**4.965**	**5.775**	**5.015**	**0.601**	2.989	**3.670**	**4.613**	**3.757**	**0.666**

Face Reconstruction Quantitative Results

THE NME(%) OF FACE RECONSTRUCTION RESULTS ON AFLW2000-3D

	3DDFA [5]	MARN(Ours)
$[0°, 30°]$	4.877	*4.721*
$[30°, 60°]$	6.086	*5.535*
$[60°, 90°]$	8.437	*7.483*
Mean	6.467	*5.913*
std	1.478	*1.159*

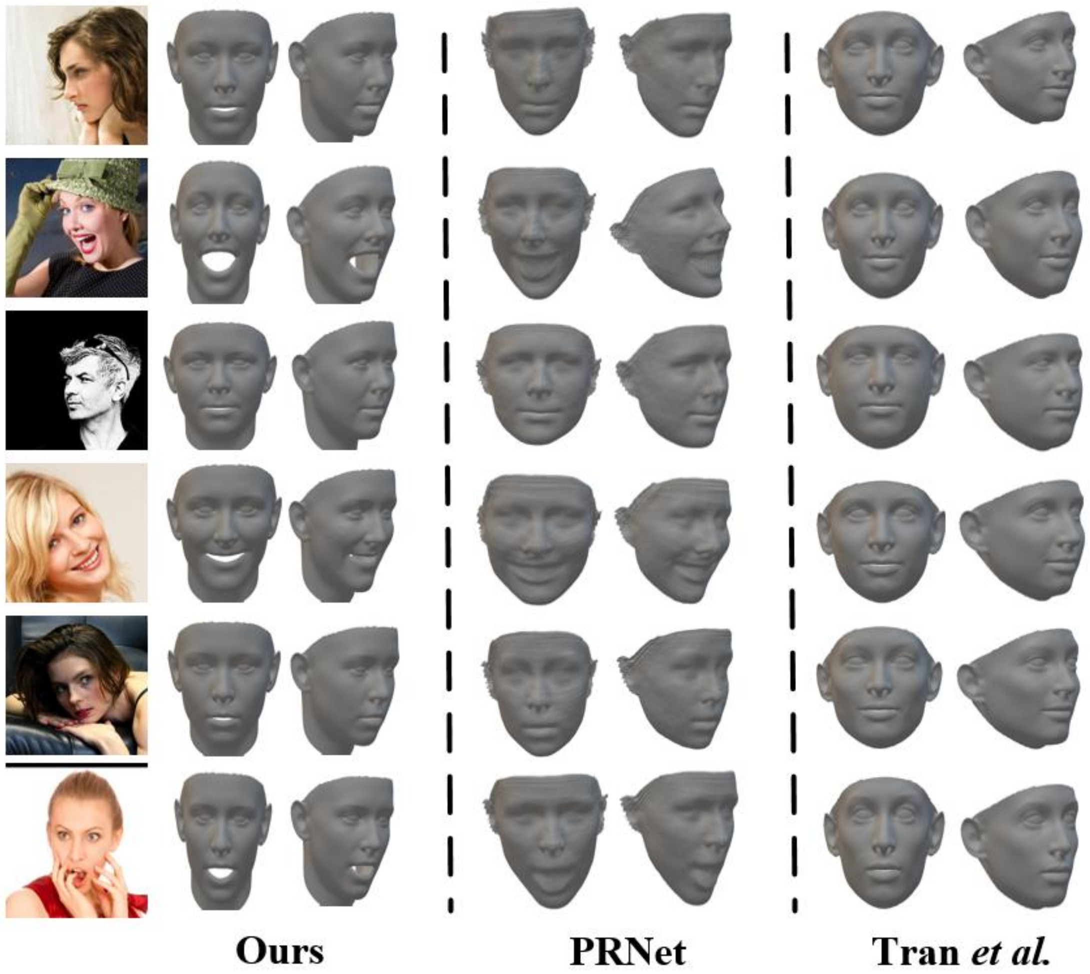


The cumulative errors distribution (CED) curves on AFLW2000-3D.

Face Alignment Qualitative Results

Comparison on 3D facial landmark detection with 3DDFA ,DAMDNet and MARN(Ours) on AFLW2000-3D.

Face Reconstruction Qualitative Results



It can be seen that the reconstructed shape of our MARN is smoother, the expression is more natural and the face has finer facial details, especially in the eyes, nose, and mouth areas.

Thanks