Generating Diverse 3D Reconstructions from a Single Occluded Face Image

CVPR 2022

Rahul Dey* Vishnu Boddeti

CVPR JUNE 19-24 2022 NEW ORLEANS LOUISIANA

MICHIGAN STATE UNIVERSITY

Occlusions are a nuisance for monocular 3D face reconstruction

Target Image

FLAME, SIGGRAPH 2017

DECA, TOG 2021

CFR-GAN, WACV 2022

Occ3DMM, IJCV 2018

Extreme3D, CVPR 2018

Problem with occlusions in 3D reconstruction

Target Image

FLAME
SIGGRAPH 2017

DECA
TOG 2021

CFR-GAN
WACV 2022

Occ3DMM
IJCV 2018

Extreme3D
CVPR 2018

Output 1

Output 2

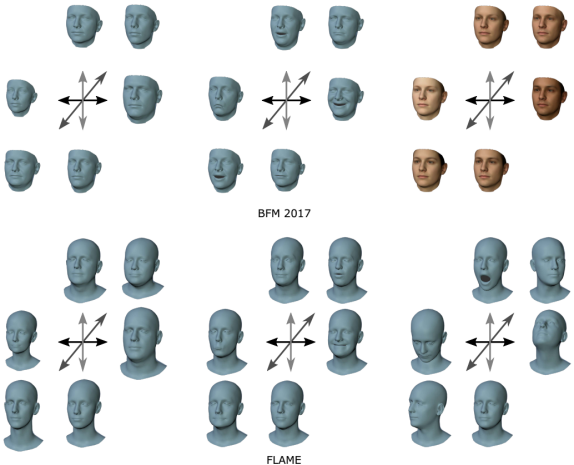
Output 3

Output 4

Output 5

Which one of these is correct?

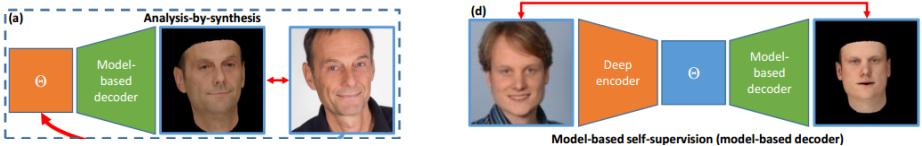
Background - How are face 3D models represented?



Source: Egger, et al. "3D Morphable Face Models — Past, Present, and Future." TOG, 2020.

Existing approaches

$$\Phi = (\beta, \theta, \psi)$$



Fitting based **Learning based**

Source: Egger, et al. "3D Morphable Face Models — Past, Present, and Future." TOG, 2020.

Problem with existing approaches

- Global model to fit to the entire head/face

$$T(\beta, \theta, \psi) = \bar{\mathbf{T}} + B_S(\beta; \mathcal{S}) + B_P(\theta; \mathcal{P}) + B_E(\psi; \mathcal{E})$$

Problem with existing approaches

- Global model to fit to the entire head/face

$$T(\beta, \theta, \psi) = \bar{\mathbf{T}} + B_S(\beta; \mathcal{S}) + B_P(\theta; \mathcal{P}) + B_E(\psi; \mathcal{E})$$

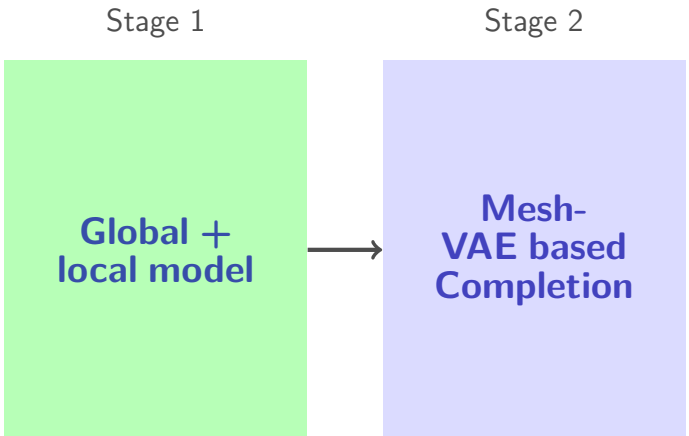
- Singular solution rather than a plurality of solutions

Our proposed solution

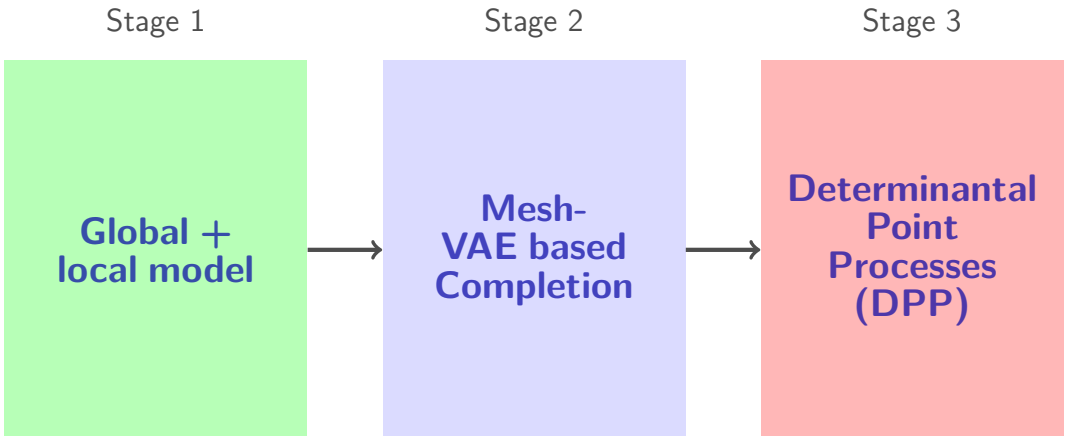
Stage 1

Global + local model

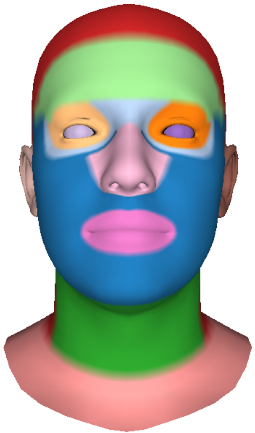
Our proposed solution



Our proposed solution



Global + local model



- Fit FLAME on the the **D3DFACS** and **CoMA** datasets

Global + local model

- Fit FLAME on the the **D3DFACS** and **CoMA** datasets
- Retain the contributions from the top 10 shape and expression bases

$$\tilde{S}^{coarse} = \bar{\mathbf{T}} + \sum_{n=1}^{N_S} \beta_n \mathcal{S}_n + \sum_{n=1}^{N_E} \psi_n \mathcal{E}_n$$

Global + local model

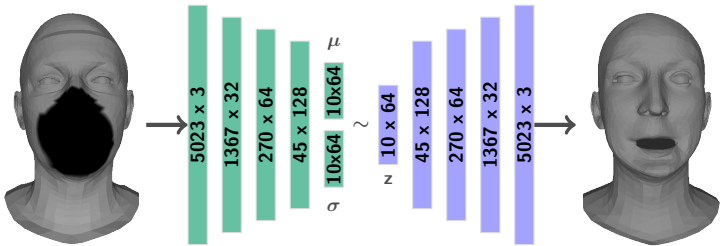
- Fit FLAME on the the **D3DFACS** and **CoMA** datasets
- Retain the contributions from the top 10 shape and expression bases

$$\tilde{S}^{coarse} = \bar{\mathbf{T}} + \sum_{n=1}^{N_S} \beta_n \mathcal{S}_n + \sum_{n=1}^{N_E} \psi_n \mathcal{E}_n$$

- Compute local PCA models using the residual errors

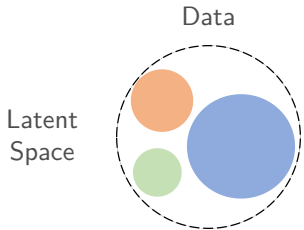
$$PCA(S^{gt}_{\mathcal{R}} - \tilde{S}^{coarse}_{\mathcal{R}}) \rightarrow (\mathcal{S}^{\mathcal{R}}, \mathcal{E}^{\mathcal{R}})$$

Mesh-VAE

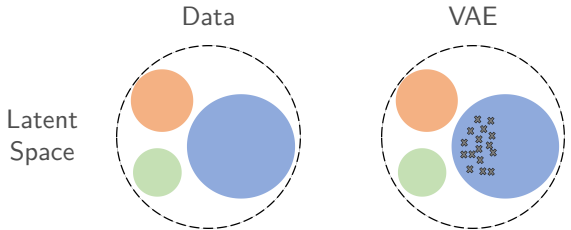


$$p(\mathbf{S}_c, \mathbf{z}|\mathbf{S}_m) = p(\mathbf{z}|\mathbf{S}_m)p(\mathbf{S}_c|\mathbf{z}, \mathbf{S}_m)$$

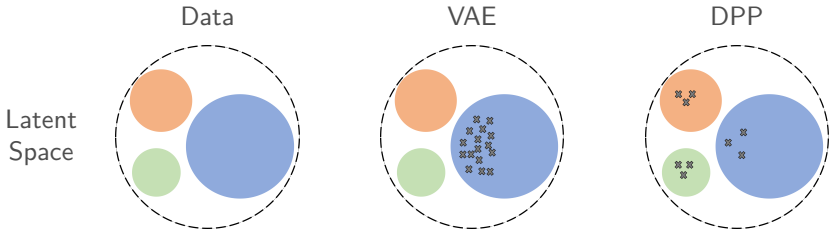
VAE is not sufficient for diversity



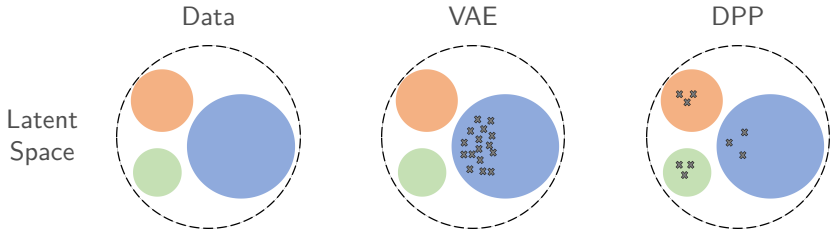
VAE is not sufficient for diversity



VAE is not sufficient for diversity

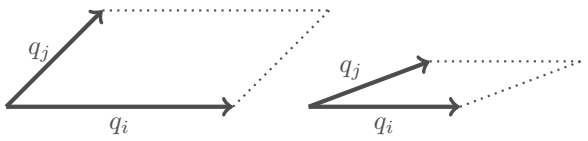


VAE is not sufficient for diversity

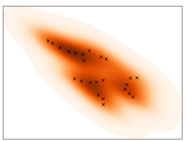


- Determinantal Point Processes:

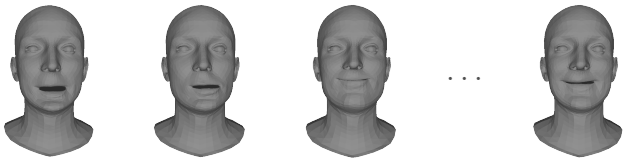
$$L_{i,j} = q_i S_{i,j} q_j,$$



Our formulation of the DPP kernel



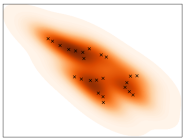
$\mathbf{z}$



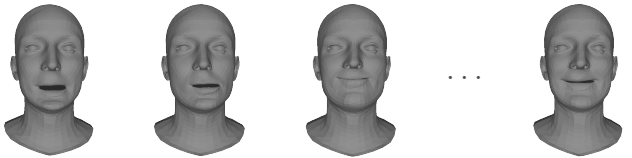
- **Similarity**

$$S_{i,j} = \exp\left(-\frac{k}{\text{median}_{i,j}(dist_{i,j})} dist_{i,j}\right)$$

Our formulation of the DPP kernel



$\mathbf{z}$



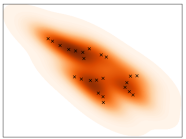
- **Similarity**

$$S_{i,j} = \exp\left(-\frac{k}{\text{median}_{i,j}(dist_{i,j})} dist_{i,j}\right)$$

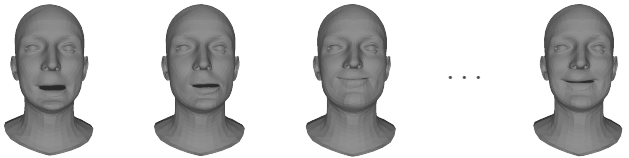
- **Quality**

$$q_i = \exp(-\max(0, \mathbf{z}_i^T \mathbf{z}_i - 3\sqrt{d}))$$

Our formulation of the DPP kernel



$\mathbf{z}$



- **Similarity**

$$S_{i,j} = \exp\left(-\frac{k}{\text{median}_{i,j}(dist_{i,j})} dist_{i,j}\right)$$

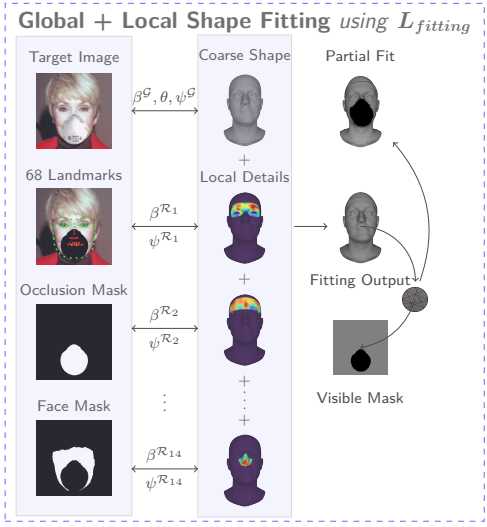
- **Quality**

$$q_i = \exp(-\max(0, \mathbf{z}_i^T \mathbf{z}_i - 3\sqrt{d}))$$

- Expected cardinality as **DPP loss**

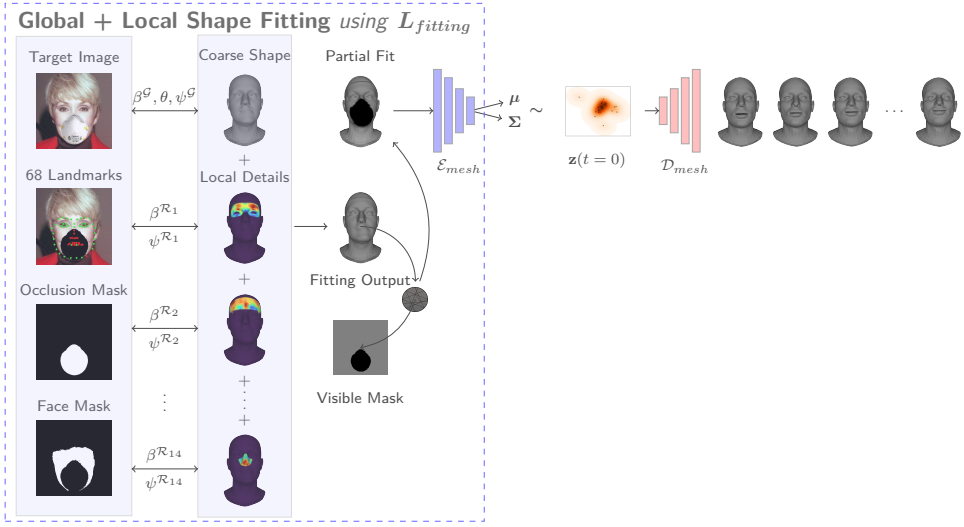
$$L_{dpp} = -tr\left(\mathbf{I} - (\mathbf{L} + \mathbf{I})^{-1}\right)$$

Diverse3DFace - the overview



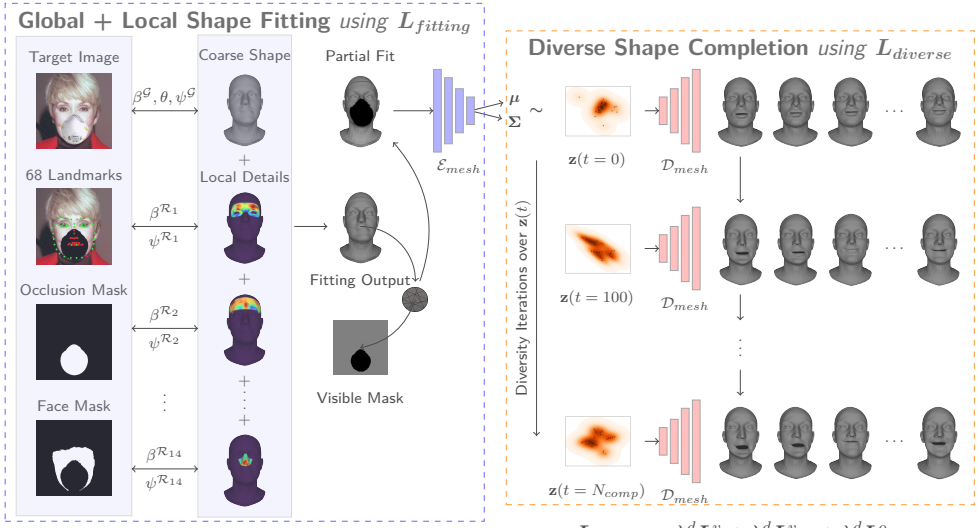
$$\boldsymbol{L}_{fitting} = \lambda_1^f \boldsymbol{L}_{lmk}^v + \lambda_2^f \boldsymbol{L}_{pho}^v + \lambda_3^f \boldsymbol{L}_{reg}$$

Diverse3DFace - the overview



$$L_{fitting} = \lambda_1^f L_{lmk}^v + \lambda_2^f L_{pho}^v + \lambda_3^f L_{reg}$$

Diverse3DFace - the overview



$$\boldsymbol{L}_{fitting} = \lambda_1^f \boldsymbol{L}_{lmk}^v + \lambda_2^f \boldsymbol{L}_{pho}^v + \lambda_3^f \boldsymbol{L}_{reg}$$

$$\boldsymbol{L}_{diverse} = \lambda_1^d \boldsymbol{L}_S^v + \lambda_2^d \boldsymbol{L}_{pho}^v + \lambda_3^d \boldsymbol{L}_{dpp}^o$$

Qualitative results - Face mask

Qualitative results - Face mask

Li *et al.*, 2017

Feng *et al.*, 2021

Ju *et al.*, 2022

Egger *et al.*, 2018

Trán *et al.*, 2018

Qualitative results - Face mask

Li *et al.*, 2017

Feng *et al.*, 2021

Ju *et al.*, 2022

Egger *et al.*, 2018

Trán *et al.*, 2018

Diverse 3D Reconstructions by Our Approach (**Diverse3DFace**)

Qualitative results - Eyeglasses

Li *et al.*, 2017

Feng *et al.*, 2021

Ju *et al.*, 2022

Egger *et al.*, 2018

Trán *et al.*, 2018

Diverse 3D Reconstructions by Our Approach (**Diverse3DFace**)

Qualitative results - Random occlusion

Li *et al.*, 2017

Feng *et al.*, 2021

Ju *et al.*, 2022

Egger *et al.*, 2018

Trán *et al.*, 2018

Diverse 3D Reconstructions by Our Approach (**Diverse3DFace**)

Quantitative evaluation metrics

- **Closest Sample Error (CSE):** Mean-vertex error between the ground-truth and the closes reconstructed shape ($\downarrow$)

Quantitative evaluation metrics

- **Closest Sample Error (CSE):** Mean-vertex error between the ground-truth and the closes reconstructed shape ($\downarrow$)
- **Average Self Distance (ASD):** Mean-vertex error between a sample and its closest neighbor, averaged across all the samples

Quantitative evaluation metrics

- **Closest Sample Error (CSE):** Mean-vertex error between the ground-truth and the closes reconstructed shape ($\downarrow$)
- **Average Self Distance (ASD):** Mean-vertex error between a sample and its closest neighbor, averaged across all the samples
  - **Average Self Distance-Visible (ASD-V)** ($\downarrow$)

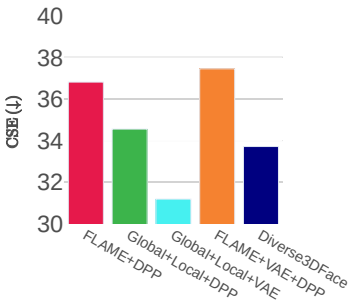
Quantitative evaluation metrics

- **Closest Sample Error (CSE):** Mean-vertex error between the ground-truth and the closes reconstructed shape ($\downarrow$)
- **Average Self Distance (ASD):** Mean-vertex error between a sample and its closest neighbor, averaged across all the samples
  - **Average Self Distance-Visible (ASD-V)** ($\downarrow$)
  - **Average Self Distance-Occluded (ASD-O)** ($\uparrow$)

Quantitative evaluation of Diverse3DFace

Baselines:

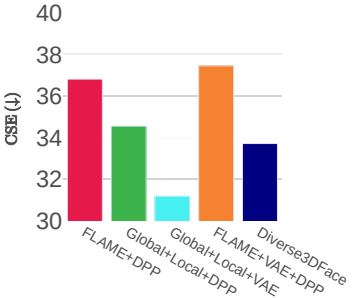
1. **FLAME+DPP** - Step 1 (with FLAME) + Step 3
2. **Global+Local+DPP** - Step 1 + Step 3
3. **Global+Local+VAE** - Step 1 + Step 2
4. **FLAME+VAE+DPP** - Step 1 (with FLAME) + Step 2 + Step 3
5. **Diverse3DFace** - Step 1 + Step 2 + Step 3

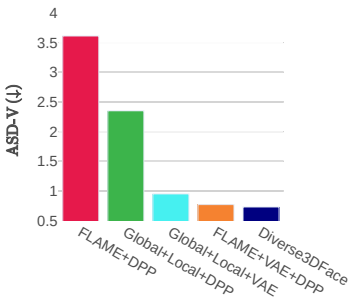


Quantitative evaluation of Diverse3DFace

Baselines:

1. **FLAME+DPP** - Step 1 (with FLAME) + Step 3
2. **Global+Local+DPP** - Step 1 + Step 3
3. **Global+Local+VAE** - Step 1 + Step 2
4. **FLAME+VAE+DPP** - Step 1 (with FLAME) + Step 2 + Step 3
5. **Diverse3DFace** - Step 1 + Step 2 + Step 3

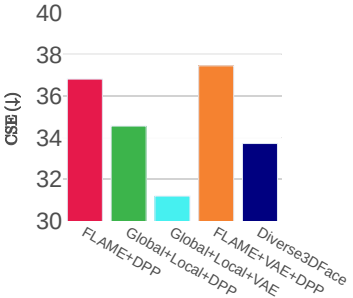


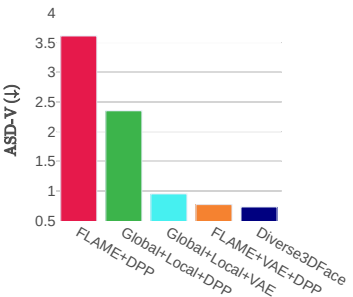


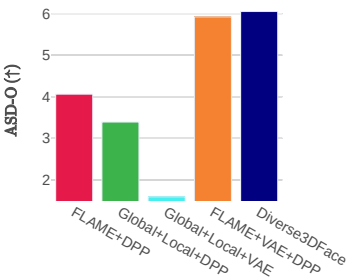
Quantitative evaluation of Diverse3DFace

Baselines:

1. **FLAME+DPP** - Step 1 (with FLAME) + Step 3
2. **Global+Local+DPP** - Step 1 + Step 3
3. **Global+Local+VAE** - Step 1 + Step 2
4. **FLAME+VAE+DPP** - Step 1 (with FLAME) + Step 2 + Step 3
5. **Diverse3DFace** - Step 1 + Step 2 + Step 3







Qualitative evaluation on real occlusions

Target Image

Fitting by Global-local model

3D Reconstructions by Diverse3DFace

Qualitative evaluation on real occlusions

Target Image

Fitting by Global-local model

3D Reconstructions by Diverse3DFace

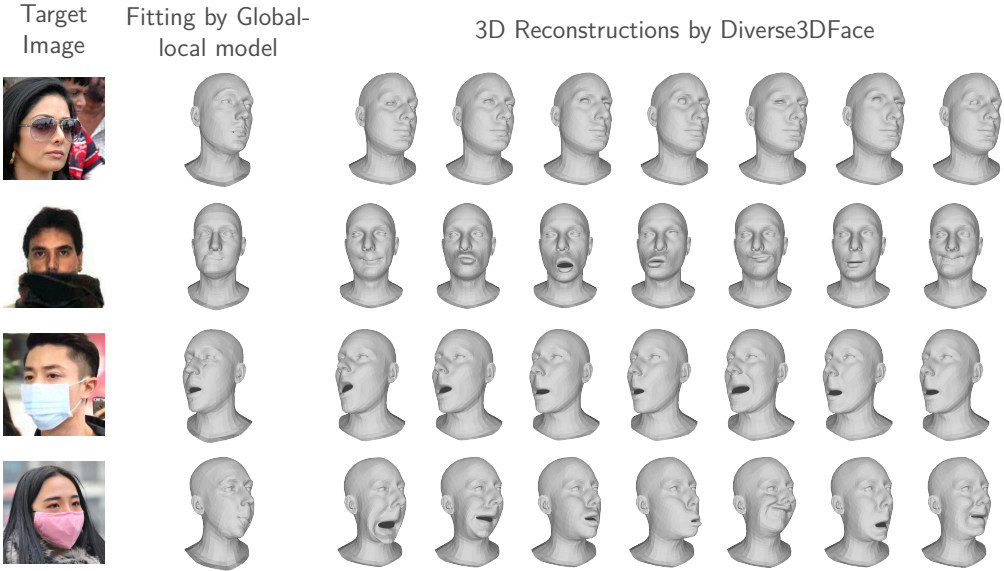
Qualitative evaluation on real occlusions

Target Image

Fitting by Global-local model

3D Reconstructions by Diverse3DFace

Qualitative evaluation on real occlusions



Conclusions

- Robustness and diversity as the desired objectives

Conclusions

- Robustness and diversity as the desired objectives
- Proposed Diverse3DFace to achieve the aforementioned objectives

Conclusions

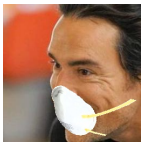
- Robustness and diversity as the desired objectives
- Proposed Diverse3DFace to achieve the aforementioned objectives
- Proposed a three step solution including a global+local shape model, Mesh-VAE based shape completion and DPP based diversification

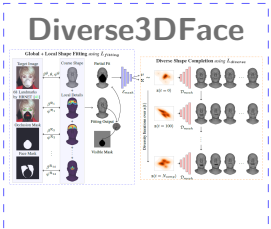
Conclusions

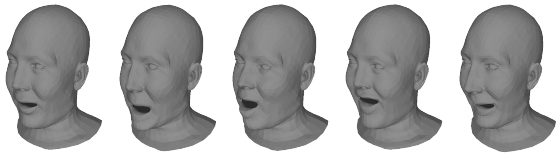
- Robustness and diversity as the desired objectives
- Proposed Diverse3DFace to achieve the aforementioned objectives
- Proposed a three step solution including a global+local shape model, Mesh-VAE based shape completion and DPP based diversification
- Quantitative and qualitative experiments comparisons against several baselines show the efficacy of the proposed approach

Conclusions

- Robustness and diversity as the desired objectives
- Proposed Diverse3DFace to achieve the aforementioned objectives
- Proposed a three step solution including a global+local shape model, Mesh-VAE based shape completion and DPP based diversification
- Quantitative and qualitative experiments comparisons against several baselines show the efficacy of the proposed approach
- Limitations:
  - Dependence on the initial landmark or face-mask estimates







Poster Session 1.1/Tuesday 21 June **Poster ID**: 152a

Title: Generating Diverse 3D Reconstructions from a Single Occluded Face Image

Code: `https://github.com/human-analysis/diverse3dface`