Generating Diverse 3D Reconstructions from a Single Occluded Face Image

CVPR 2022

Diverse3DFace

Rahul Dey* Vishnu Boddeti

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?

Existing approaches

\[ \Phi = (\beta, \theta, \psi) \]

**Fitting based**

**Learning based**

Problem with existing approaches

- **Global model** to fit to the entire head/face

\[
T(\beta, \theta, \psi) = \bar{T} + B_S(\beta; S) + B_P(\theta; P) + B_E(\psi; E)
\]
Problem with existing approaches

- **Global model** to fit to the entire head/face

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

- **Singular solution** rather than a plurality of solutions
Our proposed solution

Stage 1

Global + local model
Our proposed solution

Stage 1: Global + local model

Stage 2: Mesh-VAE based Completion
Our proposed solution

Stage 1: Global + local model

Stage 2: Mesh-VAE based Completion

Stage 3: Determinantal Point Processes (DPP)
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

\[
\hat{S}_{\text{coarse}} = \mathbf{T} + \sum_{n=1}^{N_S} \beta_n S_n + \sum_{n=1}^{N_E} \psi_n E_n
\]
Global + local model

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

- Retain the contributions from the top 10 shape and expression bases

\[
\tilde{S}_{\text{coarse}} = \bar{T} + \sum_{n=1}^{N_S} \beta_n S_n + \sum_{n=1}^{N_E} \psi_n E_n
\]

- Compute local PCA models using the residual errors

\[
P_CA(S_{gt}^R - \tilde{S}_{\text{coarse}}^R) \rightarrow (S^R, E^R)
\]
$$p(S_c, z|S_m) = p(z|S_m)p(S_c|z, S_m)$$
VAE is not sufficient for diversity

\[ L_{i,j} = q_i S_{i,j} q_j \]
VAE is not sufficient for diversity

Determinantal Point Processes:

\[ L_{i,j} = q_i S_{i,j} q_j \]
VAE is not sufficient for diversity

Determinantal Point Processes:

\[ L_{i,j} = q_i S_{i,j} q_j \]

\[ q_i q_j q_i q_j \]
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

- **Similarity**

\[ S_{i,j} = \exp \left( - \frac{k}{\text{median}_{i,j}(\text{dist}_{i,j})} \text{dist}_{i,j} \right) \]
Our formulation of the DPP kernel

- **Similarity**
  \[ S_{i,j} = \exp \left( - \frac{k}{\text{median}_{i,j}(\text{dist}_{i,j})} \text{dist}_{i,j} \right) \]

- **Quality**
  \[ q_i = \exp(- \max(0, z_i^T z_i - 3\sqrt{d})) \]
Our formulation of the DPP kernel

• **Similarity**

\[ S_{i,j} = \exp \left( - \frac{k}{\text{median}_{i,j}(\text{dist}_{i,j})} \text{dist}_{i,j} \right) \]

• **Quality**

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

• **Expected cardinality as DPP loss**

\[ L_{dpp} = -\text{tr} \left( \mathbf{I} - (\mathbf{L} + \mathbf{I})^{-1} \right) \]
Diverse3DFace - the overview

Global + Local Shape Fitting using $L_{fitting}$

$$L_{fitting} = \lambda_1 L_{mk}^v + \lambda_2 L_{pho}^v + \lambda_3 L_{reg}$$
**Diverse3DFace - the overview**

**Global + Local Shape Fitting using** $L_{\text{fitting}}$

- **Target Image**
  - 68 Landmarks
  - Occlusion Mask
  - Face Mask

- **Coarse Shape**
  - $\beta^G$, $\theta$, $\psi^G$
  - $\beta^{R_1}$, $\psi^{R_1}$
  - $\beta^{R_2}$, $\psi^{R_2}$
  - $\beta^{R_{14}}$, $\psi^{R_{14}}$

- **Local Details**

- **Fitting Output**
  - $\mu$, $\Sigma$
  - $z(t=0)$
  - $D_{\text{mesh}}$

- **Visible Mask**

**$L_{\text{fitting}} = \lambda_1^f L_{\text{lmk}}^v + \lambda_2^f L_{\text{pho}}^v + \lambda_3^f L_{\text{reg}}$**
Diverse3DFace - the overview

Global + Local Shape Fitting using $L_{fitting}$

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

Diverse Shape Completion using $L_{diverse}$

$$L_{diverse} = \lambda_1^d L_{vS}^v + \lambda_2^d L_{pho}^v + \lambda_3^d L_{dpp}^\sigma$$
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 closest reconstructed shape \(\downarrow\).
Quantitative evaluation metrics

- **Closest Sample Error (CSE)**: Mean-vertex error between the ground-truth and the closes reconstructed shape (↓)
- **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 closest 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 metrics

- **Closest Sample Error (CSE):** Mean-vertex error between the ground-truth and the closest reconstructed shape (↓)
- **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) (↓)
  - Average Self Distance-Occluded (ASD-O) (↑)
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

Target Image

Fitting by Global-local model

3D Reconstructions by Diverse3DFace
Conclusions

- Robustness and diversity as the desired objectives
Conclusions

• Robustness and diversity as the desired objectives

• Proposed Diverse3DFace to achieve the aforementioned objectives

Limitations:
• Dependence on the initial landmark or face-mask estimates
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

• Limitations:
  • Dependence on the initial landmark or face-mask estimates
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
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
Title: Generating Diverse 3D Reconstructions from a Single Occluded Face Image

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