

# FaceCom: Towards High-fidelity 3D Facial Shape Completion via Optimization and Inpainting Guidance

Yinglong Li<sup>1</sup>, Hongyu Wu<sup>1\*</sup>, Xiaogang Wang<sup>3</sup>, Qingzhao Qin<sup>2</sup>,  
Yijiao Zhao<sup>2</sup>, Yong Wang<sup>2</sup>, Aimin Hao<sup>1</sup>

<sup>1</sup>State Key Laboratory of Virtual Reality Technology and Systems, Beihang University

<sup>2</sup>Peking University School and Hospital of Stomatology

<sup>3</sup>College of Computer and Information Science, Southwest University

{dragonylee, whyvrlab, ham}@buaa.edu.cn, wangxiaogang@swu.edu.cn,  
2211210502@pku.edu.cn, {kqcadcs, kqcadc}@bjmu.edu.cn

## Abstract

*We propose FaceCom, a method for 3D facial shape completion, which delivers high-fidelity results for incomplete facial inputs of arbitrary forms. Unlike end-to-end shape completion methods based on point clouds or voxels, our approach relies on a mesh-based generative network that is easy to optimize, enabling it to handle shape completion for irregular facial scans. We first train a shape generator on a mixed 3D facial dataset containing 2405 identities. Based on the incomplete facial input, we fit complete faces using an optimization approach under image inpainting guidance. The completion results are refined through a post-processing step. FaceCom demonstrates the ability to effectively and naturally complete facial scan data with varying missing regions and degrees of missing areas. Our method can be used in medical prosthetic fabrication and the registration of deficient scanning data. Our experimental results demonstrate that FaceCom achieves exceptional performance in fitting and shape completion tasks. The code is available at <https://github.com/dragonylee/FaceCom.git>.*

## 1. Introduction

The shape completion of incomplete facial scans presents a significant challenge in the fields of computer vision, graphics, and medicine. On one hand, limitations in scanning devices and occlusions during the capture process often result in incomplete scans. On the other hand, clinical scenarios often involve patients with facial defects, for whom conventional maxillofacial repair methods such as mirrored reconstruction [2] or the selection of pre-designed

prosthetics from a database [31] may lack personalized shapes and require manual adjustments to fit the patient's facial contours. With the rapid development of deep learning and the increasing availability of three-dimensional facial datasets, there is a growing interest in leveraging diverse individual samples and deep learning techniques for personalized and precise facial shape completion.

While recent years have seen numerous works on image inpainting for faces [22, 27, 30, 46], there have been relatively few studies directly addressing shape completion for three-dimensional facial data. Most voxel-based shape completion approaches [12, 35] and point cloud based methods [41, 42] have demonstrated promising results on various 3D datasets, such as ShapeNet [8]. However, when applied to 3D facial datasets, their completion results for incomplete data are often rough and inaccurate. These results may be attributed to the high similarity between 3D facial shapes, which requires networks to learn subtle shape variations. While point cloud-based methods can learn different shapes, they are insensitive to small-scale deformations and struggle to generate smooth results due to the lack of explicit surface adjacency. Similarly, voxel based methods, although offering regular results and ease of construction and training, face challenges in generating high-precision results due to memory and resolution limitations. Therefore, we propose a novel approach that operates directly on facial mesh data for generation purposes.

Our proposed method involves a graph convolutional neural network structure, specifically an autoencoder, that operates directly on explicit meshes to learn various facial shape features and perform parameterized generation tasks. Our approach aims to learn the facial shapes of as many identities as possible, focusing on neutral faces without considering variations of expressions. To address the limited number of identities in existing 3D facial datasets, we mix

\*Corresponding author.

several publicly available datasets with our collected facial dataset, resulting in a comprehensive 3D facial dataset with a substantial number of identities and unified topology. Following auto-regressive training, we obtain a shape generator capable of sampling from a hypersphere distribution and generating a complete facial mesh. Additionally, leveraging the larger scale of facial image datasets, we train a 2D facial inpainting network applicable to our dataset based on the work by Lugmayr *et al.* [30]. In the case of incomplete facial inputs, we utilize an optimization method to fit a complete facial mesh, with guidance provided by the results of image inpainting. Post-processing is employed to refine and achieve the final completion results. Please refer to Fig. 5 for the visual results.

Our approach excels in generating high-fidelity completion results for various forms of incomplete input, without the need for specific vertex numbers or correspondences. In our experiments, we demonstrate that our method outperforms other state-of-the-art models in fitting neutral faces and exhibits excellent completion results for various incomplete faces. Furthermore, our method yields practical application results in clinical data completion experiments.

In summary, the main contributions of this paper include:

- Introducing a novel 3D facial completion method based on generative model, image inpainting guidance and optimization techniques, capable of generating high-fidelity completion results for free-form incomplete facial scans.
- Proposing a three-dimensional face generation network based on a multi-scale encoder-decoder architecture, outperforming existing state-of-the-art models in fitting neutral faces.
- Validating the potential applications of our completion method in clinical maxillofacial repair and non-rigid registration fields.

## 2. Related Work

**Parametric Face and Head Models.** Blanz and Vetter [1] were the pioneers in introducing 3D Morphable Models (3DMMs) for representing faces. They utilized PCA to reduce the dimensions of shape and texture spaces, enabling parametric face generation. Subsequently, 3DMMs found widespread application in face reconstruction [17, 40, 52]. Based on our research, current studies on parameterized face or head models can be broadly categorized into three types. The first type involves linear or blendshapes-based statistical models [3, 4, 6, 7, 13, 14, 25, 32, 43]. These models often use a unified topology dataset and separately model shape, expression, or texture, and are characterized by their ease of construction and processing. The second type aims to achieve non-linear deformation explicitly using deep learning methods [5, 9, 10, 33], which extract features with

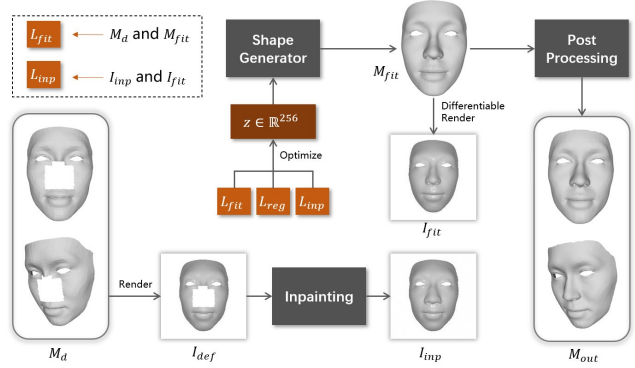


Figure 1. **FaceCom Overview.** During the shape fitting stage, the shape generator is employed to fit the incomplete facial scan  $M_d$  using an optimization method. The optimization objective comprises several loss functions, including the disparity between the input  $M_d$  and the fitting result  $M_{fit}$ , the difference between the images  $I_{fit}$  and  $I_{inp}$ , and regularization terms. Following the acquisition of the optimal complete shape relative to the incomplete input, a post-processing step is applied to achieve the final completion results  $M_{out}$ .

spectral or spiral graph convolutions. The final type represents the recent emergence of implicit neural functions for constructing continuous models [20, 38, 44, 48]. These models can finely reproduce facial expression variations but often exhibit poorer fitting performance for different identities.

**Shape Completion.** 3D shape completion, or point cloud completion, is a technique used to reconstruct missing parts of input models. In recent years, deep learning methods have shown promise in effectively addressing missing data in 3D models. Some approaches have focused on learning and generating based on voxels due to their regularity and ease of convolutional processing [12, 35, 39]. However, voxel-based methods tend to consume a significant amount of space and yield lower-quality results. Consequently, several studies have directly tackled shape completion tasks on point cloud [36, 41, 45] or have combined graph convolutions [26, 50] to incorporate more local information. Recently some novel methods have also leveraged implicit functions to avoid discretization or have integrated new generative models to achieve superior generation outcomes [11, 42]. Despite these advancements, the effectiveness of these shape completion techniques on highly detailed facial data remains limited. The work of Litany *et al.* [26] shares some similarities with our approach and also attempted facial completion experiments, yet they imposed constraints on the topological structure of incomplete input, and the completion results were found to be unsatisfactory.

### 3. Method

We propose an algorithm called FaceCom, which is designed for shape completion of incomplete 3D facial scans. The key components of our completion process, including the shape generator, image inpainting guidance, and post-processing, will be discussed in this section. The workflow of FaceCom is illustrated in Fig. 1.

#### 3.1. Shape Generator

We employ a GNN-based autoencoder framework for learning the representation of 3D facial meshes and generation, as shown in Fig. 2. The procedure involves extracting features from the input facial mesh at both the global and local levels, with the fused features producing a 256-dimensional implicit representation of the face. The shape generation phase worked inversely to the encoding phase, decoding the implicit vector to gain a complete facial shape. The encoder part is meticulously illustrated in Fig. 2, and the decoder follows a symmetric structure. In contrast to the global path, the local path removes the downsampling layers, reduces the number of channels, and preserves local geometric information by using linear layers in place of pooling layers.

Meshes retain rich local geometric information, yet their irregularity makes them challenging to process. We employed FeaStNet [18, 37] to conduct convolution operations directly on meshes, which demonstrates excellent feature extraction capabilities. Downsampling layers were implemented and improved based on the work of Ranjan *et al.* [33]. To ensure the differentiability of the downsampling process, we simplified [19] the face template  $M$ , resulting in  $M_1, \dots, M_k$ , where  $M_i \in \mathbb{R}^{N_i \times 3}$ . We assumed that a vertex’s position in  $M_{i+1}$  is a linear combination of the positions of the  $k$ -nearest vertices in  $M_i$ , allowing us to compute the transformation matrix  $M_{i+1} = D_i M_i$ , where the sparse matrix  $D_i \in \mathbb{R}^{N_{i+1} \times N_i}$ . Upsampling can also be achieved using a similar method. Within the neural network, we maintained fixed transformation matrices to facilitate differentiable upsampling and downsampling operations.

Our objective is to perform shape completion on incomplete(defect) facial scans in a clinical setting. Hence, we disregard facial expression variations, allowing the shape generator to learn as many individual facial shapes as possible. Unlike 2D images, publicly available 3D facial datasets often contain a limited number of individuals. To obtain a more diverse and aligned dataset, we applied several public 3D facial datasets [13, 43] to use their neutral faces. Additionally, we collected 400 neutral facial scans in the hospital. All the 3D facial data we utilized is presented in Tab. 1. Subsequently, we employed the non-rigid registration method proposed in [16] to achieve a dataset  $\mathcal{X}$  with a unified topological structure, which comprises numer-

Dataset	Sub. Num.	Reserved
Headspace [13, 14, 51]	1171	57
FaceScape [43]	834	41
Ours	400	20
Total	2405	118

Table 1. **3D face datasets used in our study**, all aligned to a unified topology with approximately 5% reserved for testing. Note that low-quality samples were manually excluded, resulting in a reduced number of subjects from the origins.

ous individuals, different regions, and various age groups. Around 5% of the data are reserved for subsequent testing purposes.

During training, the reconstruction loss of the network is defined as the mean squared error between the input  $\mathbf{X}$  and the output  $dec(enc(\mathbf{X}))$ , denoted as  $L_{MSE}$ . Since the autoencoder essentially serves as a data compression method, facilitating a point-to-point mapping between data samples and latent space points, its inherent nature lacks generative capabilities. As a result, we attempted to train the model using the Variational Autoencoder (VAE) approach [24]. However, we encountered challenges such as the vanishing KL divergence and non-decreasing  $L_{MSE}$ , hindering the network from reaching a satisfactory convergence. Consequently, we introduced a regularization constraint expressed by

$$L_{reg} = (\|\mathbf{z}\|_2 - 1)^2, \tag{1}$$

aiming to confine the mesh’s implicit representation, denoted as  $\mathbf{z} = enc(\mathbf{X})$ , within a hypersphere distribution. Our experimental results further supported the notion that this approach facilitates optimization-based inference procedures. Prior study [47] has illustrated that this approach can significantly enhance the autoencoder’s generative capacity and overall stability. The optimization objective was defined as  $L_{MSE} + \lambda L_{reg}$ . We initialized the learning rate at 0.001 and halved it every 50 epochs. The training process utilized the Adam optimizer [23] within the PyTorch framework, with the entire training process taking approximately 6 hours on an NVIDIA Geforce RTX4090 GPU.

#### 3.2. Shape Fitting

Many generation-based image inpainting approaches [22, 27] and point cloud shape completion techniques [12, 49] train an end-to-end network to directly complete the missing parts of the input. However, incomplete meshes lose topological information, making it challenging to seamlessly grow the completion parts using deep learning methods. Hence, we first trained a parameterized generative model (refer to Sec. 3.1), followed by using an optimization-based method to fit the input of the incomplete

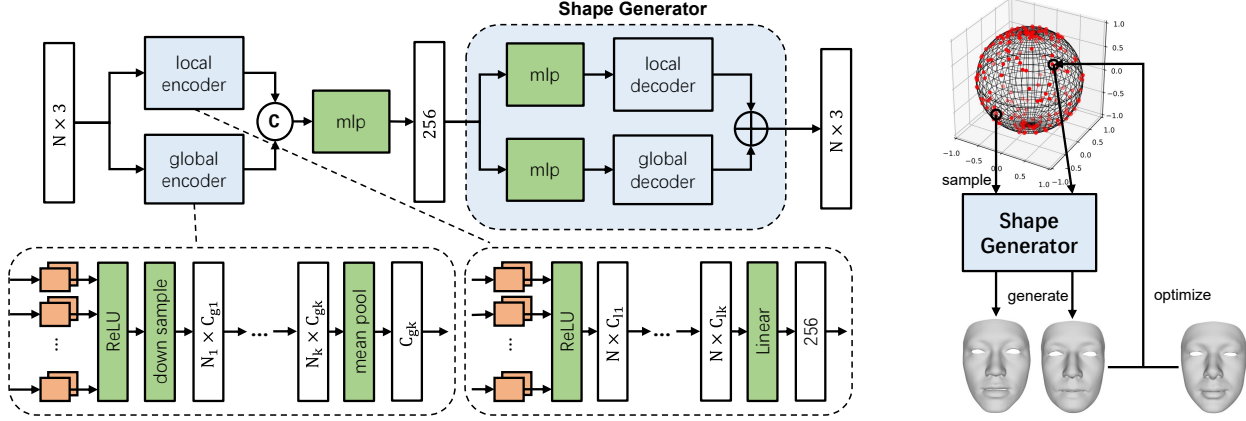


Figure 2. **Architecture of shape generator.** Left: Network used for training, which is primarily based on a graph convolutional auto-encoder structure. It conducts feature extraction and generation separately from both local and global perspectives. After training, the decoder component serves as the shape generator for FaceCom. Right: Our shape generator can produce diverse facial samples from the hypersphere space and is readily optimized within the latent space.

facial mesh.

Let  $\mathcal{D}(\mathbf{v}, \mathbf{M})$  denote the distance from vertex  $\mathbf{v}$  to the surface  $\mathbf{M}$ . For surfaces represented by triangle meshes, we calculate the closest distance from points to the triangular faces. For the incomplete facial mesh input  $\mathbf{M}_d$ , the fitting loss is defined as

$$L_{fit} = \frac{1}{|\mathbf{M}_d|} \sum_{\mathbf{v} \in \mathbf{M}_d} \mathcal{D}(\mathbf{v}, \mathcal{T}(\text{dec}(\mathbf{z}))), \quad (2)$$

which quantifies the similarity between generated mesh and the input incomplete mesh, where  $\mathcal{T}$  represents a rigid transformation. In practice, the incomplete facial mesh may exhibit inconsistencies relative to the template, so we typically exclude a certain percentage of vertices farthest from the calculation of  $L_{fit}$ . Drawing upon the above procedures, the shape fitting process can be described as

$$\min_{\mathbf{z}, \mathcal{T}} L_{fit} + \lambda L_{reg}. \quad (3)$$

During the fitting process, we implemented a differentiable  $\mathcal{D}$  based on publicly available PyTorch3D [34], obtaining the optimal fitting result  $\text{dec}(\mathbf{z})$  through iterative optimization. Notably, our optimization-based method is applicable to any form of incomplete facial input, allowing  $\mathbf{M}_d$  to be a mesh, a point cloud, or a set of key points.

### 3.3. Image Inpainting Guidance

Our facial shape generator was trained on a dataset including neutral faces from thousands of identities. However, this volume is relatively small compared to the datasets used in the field of image generation [15, 30], making it difficult to cover a wide range of practical scenarios. Therefore, a natural idea is to train an image inpainting

network customized for our dataset, which can effectively guide the three-dimensional shape completion process.

We utilize RePaint [30], essentially a denoising diffusion probabilistic model [15, 21], as a prior for our image inpainting process. The intermediate image  $x_t$  in the reverse diffusion process originates from  $x_{t+1}$  and the conditional probability  $p_\theta(x_t|x_{t+1})$ . RePaint introducing a mask  $m$  as input, known parts  $m(x_0)$  of the image are sampled during the diffusion process using the conditional probability  $q(x_t|x_0)$ , resulting  $m(x_t^{\text{known}})$ , while the unknown parts retain the results of the inverse diffusion process  $(1-m)(x_t^{\text{unknown}})$ . The combined result serves as the intermediate sampling result, effectively incorporating guided information into the image generation process. Additionally, a resampling technique is employed to improve the coherence between the repaired and known parts. Further details can be found in [30].

An image inpainting network tailored to our specific task, denoted as  $inp$ , was trained to guide the three-dimensional shape completion process. We first set up a differentiable rendering camera  $\mathcal{R}$  with fixed parameters. Subsequent, leveraging the network weights from CelebA-HQ [28] in [30] and the training approach outlined in [15], we fine-tuned the network on our dataset  $\mathcal{R}(\mathcal{X})$ , resulting in a network capable of taking incomplete rendered faces and masks as input and producing restoration outputs.

Define the loss function between the image inpainting results and the rendering results of the fitting shape as

$$L_{inp} = \text{MSE}(\mathcal{R}(\mathcal{T}(\text{dec}(\mathbf{z}))), \text{inp}(\mathcal{R}(\mathbf{M}_d))). \quad (4)$$

By integrating the image inpainting guidance, the shape completion process for incomplete facial scans can be described as

$$\min_{\mathbf{z}, \mathcal{T}} L_{fit} + \lambda_1 L_{inp} + \lambda_2 L_{reg}. \quad (5)$$

### 3.4. Post Processing

Our optimization-based approach is capable of completing various types of incomplete 3D facial input. However, from an application standpoint, inconsistencies in the non-defective regions of the completion results can lead to misalignments at the boundaries. To address this limitation, we have devised a series of engineering post-processing methods. Through post-processing, the completed results exhibit greater consistency with the defective input, leading to more precise completion outcomes.

The visualization of the post-processing is illustrated in Fig. 3. For the defective facial input Fig. 3a, we employed the techniques described in Sections Sec. 3.1 and Sec. 3.3 to obtain the fitted face Fig. 3b. Subsequently, we identified the repaired parts of Fig. 3b using a thresholding approach, as shown in Fig. 3c. We then projected the non-defective regions onto the surface represented by the defect input along the vertex normal direction, yielding the result displayed in Fig. 3d. Finally, by leveraging the K-nearest neighbors of the extended portion of the defective region (the green portion in Figure Fig. 3c), we performed weighted vertex deformation to obtain the final completion outcome Fig. 3e. The projection process may introduce some outliers, therefore there is an additional refinement step in the end.

Please note that the post-processing stage primarily serves as an engineering approach to enhance the conformity between the completion results and the defect input. It addresses inherent issues within the optimization-based method. Post-processing greatly refines the final results particularly for faces with pronounced wrinkles. While we provide a brief description of the post-processing stage here, more detailed information is available in our supplementary materials. Figs. 3f and 3g demonstrate the disparities of the fitted face and the final completion result, relative to the defective facial input, using the same scale for color mapping. It is evident that post-processing has yielded improved results.

## 4. Experiments

In this section, we conduct multiple experiments to test FaceCom thoroughly. We evaluate the fitting capability of the shape generator on neutral faces in Section 4.1, examine FaceCom’s proficiency in completing irregular incomplete facial inputs in Section 4.2, assess FaceCom’s effectiveness in designing prosthetics for clinical facial defect cases in Section 4.3, explore the potential of FaceCom for non-rigid registration in Section 4.4, and finally validate the necessity of different components of FaceCom in Section 4.5.

### 4.1. Fitting

The reserved data from Tab. 1 are resampled to create a facial scan test set, with the number of triangular

Method	CD(mm)↓	MD(mm)↓
FLAME [25]	0.622	0.309
facescape [43]	0.685	0.416
ImFace [48]	0.632	0.505
NPHM [20]	0.578	0.381
ours(w/o post-processing)	<b>0.480</b>	<b>0.275</b>
ours(w/ post-processing)		0.160

Table 2. **Fitting comparison results** on L2 chamfer distance and mean point-to-surface distance. The bottom row shows FaceCom’s outcomes for non-rigid face registration.

faces in each mesh randomly ranging between 10k and 20k. We conducted fitting experiments on the test set using the shape generator trained in Sec. 3.1 and compared the results with state-of-the-art 3DMM models FLAME [25] and FaceScape [43], and the implicit parametric model ImFace [48] and NPHM [20]. Some visual examples are shown in Fig. 4.

FLAME [25] and ImFace [48] offer dedicated methods for fitting scanned data, requiring the provision of additional landmarks. We prepared compliant landmarks beforehand on the test set to facilitate the evaluation process. FaceScape [43], on the other hand, solely provides a method for fitting 3D key points, and NPHM [20] only provides a point cloud fitting method after depth observation for specific samples. We modified their code to achieve the fitting of scanned data after spatial coordinate alignment. Our evaluation included the assessment of the L2 Chamfer Distance and the average point-to-surface distance, both are unidirectional as the fitted mesh typically has a larger scope. The Chamfer Distance(CD) is significantly influenced by vertex density, so we employed Loop Subdivision [29] to refine the fitted meshes to approximately 200k faces before calculation. The mean point-to-surface distance(MD) corresponds to the calculation of the average  $\mathcal{D}$  mentioned in Sec. 3.2. The results, outlined in Tab. 2, demonstrate the superior performance of our shape generator in fitting neutral faces compared to existing state-of-the-art 3D facial generation models.

### 4.2. Shape Completion

To the best of our knowledge, there have been few works specifically addressing facial shape completion in the field of computer vision. The experiments of point cloud completion methods for facial datasets have yielded unsatisfactory results. Please refer to our supplementary materials for details. Therefore, in this section, we did not conduct comparative experiments. Instead, we tested FaceCom’s completion capability by constructing several incomplete facial inputs on the test set with various regions and extents, and presented visual results.

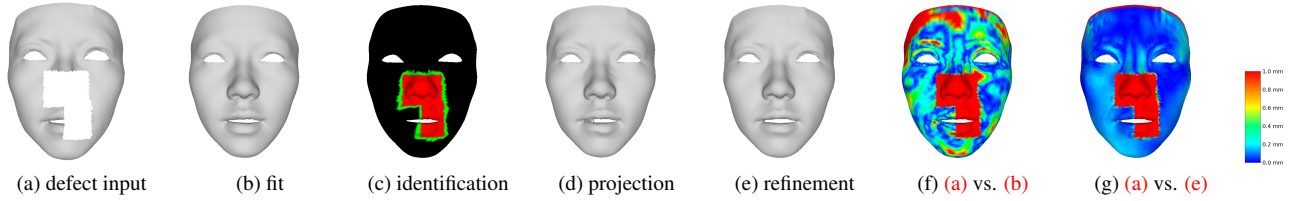


Figure 3. **Visualization of post-processing technique.** (a) depicts the incomplete facial scan input, (b) demonstrates the fitting result, (c) to (e) illustrate each step of the post-processing process. (f) and (g) showcase the disparities between the results before and after post-processing and the incomplete input.

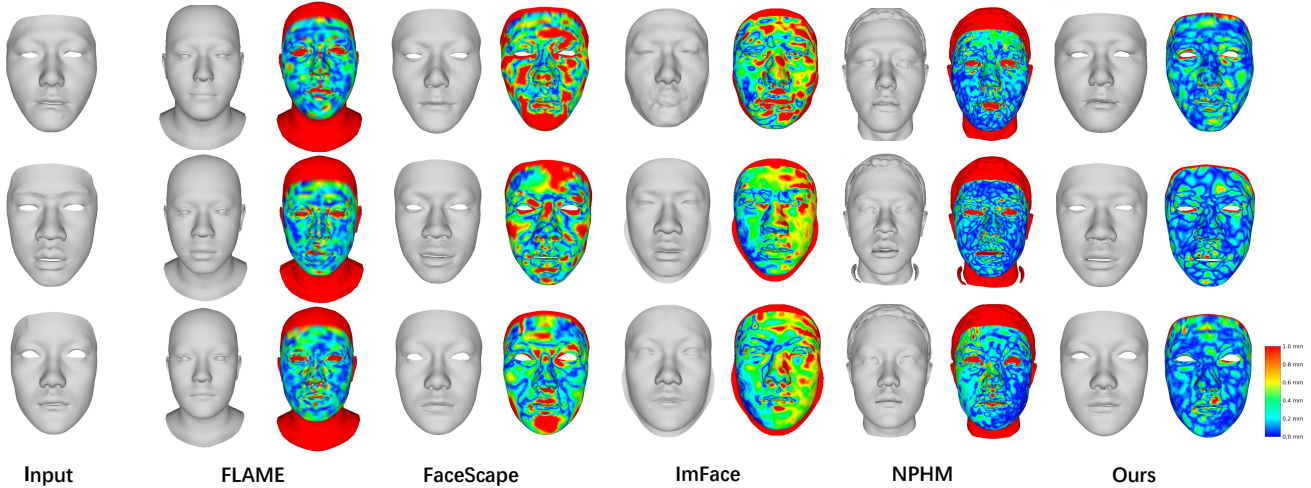


Figure 4. **Fitting experiment examples.**

The results of facial shape completion experiments are presented in Fig. 5. The visualizations demonstrate the efficacy of FaceCom in achieving natural and high-fidelity completion results for defects located in different facial areas (Figs. 5a to 5f), extensive defect areas (Figs. 5g to 5k), and cases involving elderly subjects with pronounced wrinkles (Figs. 5c and 5k). Even with extremely limited inputs, such as partial information about specific regions (Figs. 5l and 5n), scattered fragments (Fig. 5m), or just key points (Fig. 5o), FaceCom is capable of generating natural and authentic completion results. As the size of the defect increases, the information provided decreases, resulting in larger discrepancies between the completion results and the ground truth. For example, in Figs. 5l and 5n, it is challenging to determine the original facial fullness based on the available information, leading to completion results that do not entirely match the actual condition. However, the non-defective regions remain consistent, resulting in a harmonious completion. It is important to note that our optimization-based completion strategy focuses on generating a complete facial structure rather than directly generating defect regions. While this approach proves effective for a wide range of irregular defect inputs and yields ex-

cellent fitting results, there might be slight mismatches in non-defective areas. We attempted to address this concern by employing post-processing techniques.

The defect test data we manually constructed were aimed at highlighting the capability of our method to effectively complete various irregular defect inputs. It should be noted that some of the defect styles, such as the one illustrated in Fig. 5k, are not typically encountered in clinical cases. To evaluate the practical application of our solution in a clinical setting, we tasked medical professionals with collecting several samples of nasal defects that conform to clinical conditions. One of these examples is depicted in Fig. 5p. The experimental results and analysis are outlined in Sec. 4.3.

### 4.3. Clinical Experiment

In clinical facial prosthetics, two main requirements are typically considered. Firstly, the prosthetic margins should seamlessly adhere to the contours of the patient’s defect area. This criterion is evaluated by the root mean square distance of sample points, noted as margin fitness, and is deemed the most critical factor. Secondly, the prosthesis should blend naturally with the patient’s original facial fea-

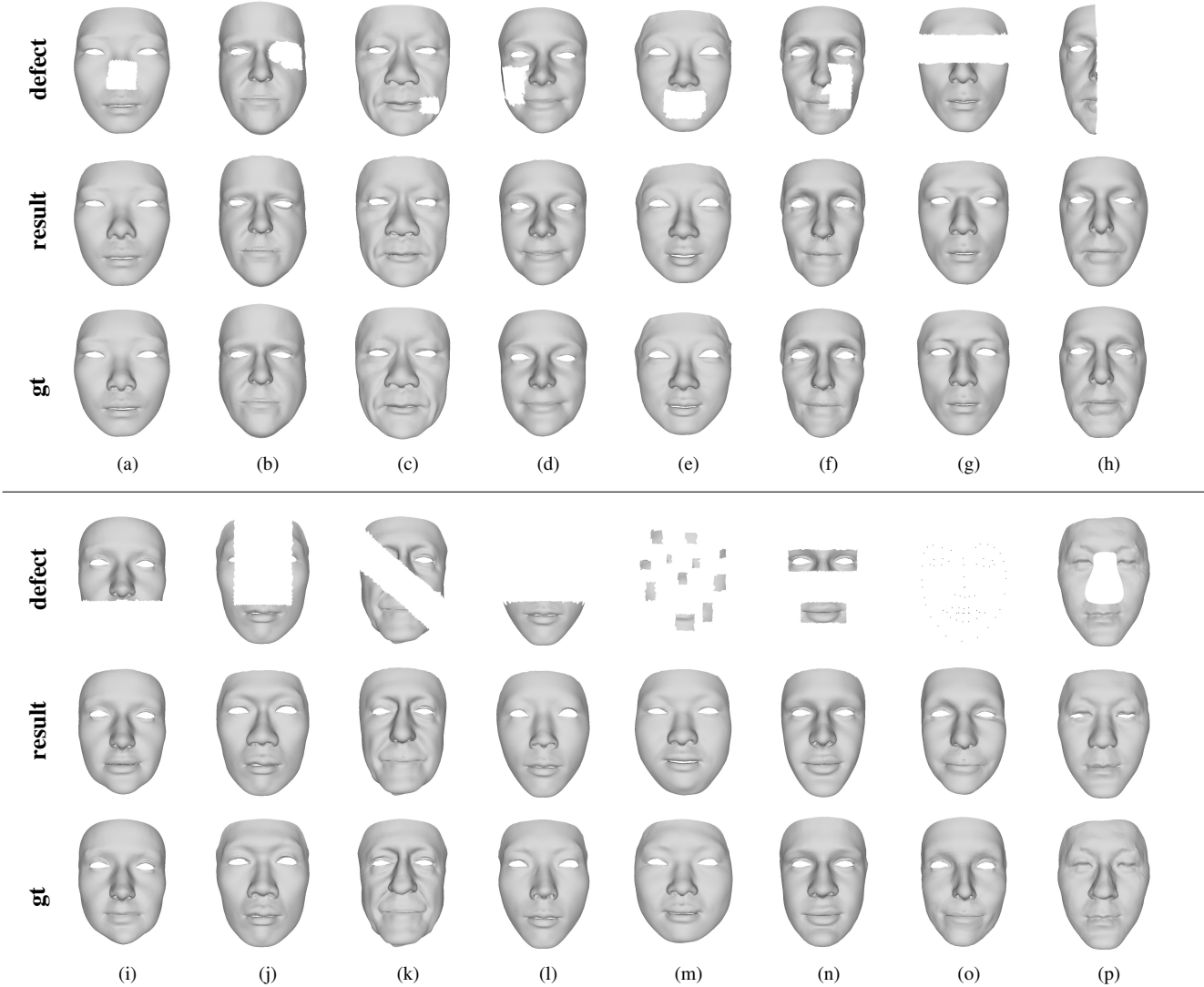


Figure 5. **Examples of facial shape completion experiment.** FaceCom demonstrates high-quality completion results for facial inputs with defects of various positions and sizes, even for samples with numerous wrinkles or lower quality. Please refer to our supplementary materials for more detailed views.

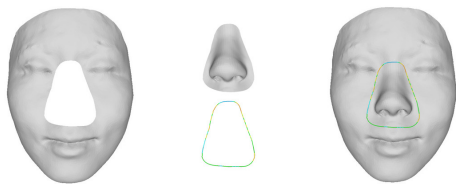


Figure 6. **Clinical Experiment.** Left: nasal defect input. Middle: extract nasal region from completion result. Right: combine and calculate margin fitness.

tures, which is challenging to quantify and is often assessed visually.

To test the practical application of FaceCom in a clinical setting, we recruited 20 clinical patients in the hospital and scanned their facial data to construct nasal defects. These 20 cases of nasal defects were all handled by professional physicians and met the standards for clinical data collection. Following the completion process by FaceCom, we extracted the nasal region for evaluation, as illustrated in Fig. 6. The average margin fitness of the completed results for the 20 cases of nasal defects was  $0.33 \pm 0.21$  mm, indicating their practicality for manufacturing and use. Additionally, the natural and seamless appearance of the completed results has been acknowledged by the physicians. Our method shows promise in providing assistance to facial defect patients in the medical field.

Method	reconstruction	regularization
Ours(global-only)	0.336	<b>0.240</b>
Ours(local-only)	0.208	0.419
Ours(full)	<b>0.153</b>	0.419

Table 3. Ablation experiment results for shape generator.

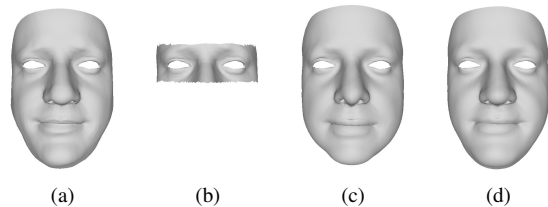


Figure 7. Ablation experiment example for image guidance. (a) Ground truth. (b) Constructed large-scale defect. (c) Fitting result without image guidance. (d) Fitting result with image guidance.

#### 4.4. Non-rigid Registration

While our method is primarily designed for completing defective 3D facial scans, it can also be utilized as a non-rigid face registration tool for facial scans with potential holes or missing areas. This is because FaceCom maintains topological consistency of outputs across various facial scans. To assess FaceCom’s performance in non-rigid facial registration, we conducted a simple test. Using the same test data as in Sec. 4.1, we applied the full completion approach to achieve non-rigid registration results. The quantitative results of the registration are shown in the last row of Tab. 2. Although we did not conduct a comparative experiment for non-rigid registration, as it falls outside the primary focus of this paper, our method achieves a precision of 0.16mm for the mean surface error, similar to the facial non-rigid registration algorithm proposed in [16], without the need for landmarks. This underlines the potential of our method in the field of non-rigid facial registration.

#### 4.5. Ablation Study

In this section, we conduct ablation experiments on the shape generator to assess the significance of its local and global modules. Additionally, we examine the role of image inpainting guidance on the 3D facial shape completion process.

For the shape generator, we isolate the global and local modules individually while adjusting the channel numbers to maintain a consistent parameter count. We trained each of them under the same training parameters. The experimental results, including the quantitative reconstruction loss and regularization loss, are summarized in Tab. 3. Notably, we observe that under similar training conditions, the network with only the global module tends to prioritize op-

timizing the regularization term, whereas the full network demonstrates superior mesh data reconstruction.

For the image inpainting guidance module, we explore its influence on FaceCom’s completion results and provide qualitative analysis. Our findings suggest that for cases with smaller defect areas, the image inpainting guidance has minimal impact. Conversely, for scan data with more extensive defect areas, as illustrated in Fig. 7, this method achieves results that are closer to the ground truth, effectively steering the shape generator’s generation process.

## 5. Discussion

Due to the intended application of our research in providing assistance for patients with facial defects, we primarily considered only basic shape variations. However, this approach may limit its applicability in other contexts. For future improvements, incorporating changes such as expressions and textures into our facial shape completion solution could be explored. Additionally, addressing the issue of less effective deformation handling at the edges during geometric processing is another aspect that requires attention.

The publicly available datasets used in this research were obtained through formal applications and adhered to their respective agreements. Additionally, the clinical facial data we collected was obtained with the consent of the patients and has been approved for scientific research by the relevant Institutional Review Board. We have provided the corresponding approval documents in the supplementary materials.

## 6. Conclusion

We propose a facial shape completion solution FaceCom in this paper, involving training a shape generator on diverse individual datasets, followed by the generation of a complete face from the incomplete input using optimization and image inpainting guidance, and finally refining the completed face through post-processing. The entire process is automated and efficient, yielding high-fidelity completion results for irregular incomplete facial scans. Our approach demonstrates promising results in fitting and completion, and it can be practically applied to the design of clinical facial prostheses.

**Acknowledgment.** The authors would like to thank the anonymous reviewers for their suggestions in improving this paper. They would also like to thank Junlin Chang for his insightful ideas. This work is supported in part by the National Natural Science Foundation of China under Grant 82271039 and 62132021, in part by the Open Project Program of Peking University School of Stomatology under Grant PKUSS20220202.



## References

- [1] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 157–164, 2023. 2
- [2] Sophia Bockey, Philipp Berssenbrügge, Dieter Dirksen, Kai Wermker, Martin Klein, and Christoph Runte. Computer-aided design of facial prostheses by means of 3d-data acquisition and following symmetry analysis. *Journal of Cranio-Maxillofacial Surgery*, 46(8):1320–1328, 2018. 1
- [3] Timo Bolkart and Stefanie Wuhler. A groupwise multilinear correspondence optimization for 3d faces. In *Proceedings of the IEEE international conference on computer vision*, pages 3604–3612, 2015. 2
- [4] James Booth, Anastasios Roussos, Allan Ponniah, David Dunaway, and Stefanos Zafeiriou. Large scale 3d morphable models. *International Journal of Computer Vision*, 126(2):233–254, 2018. 2
- [5] Giorgos Bouritsas, Sergiy Bokhnyak, Stylianos Ploumpis, Michael Bronstein, and Stefanos Zafeiriou. Neural 3d morphable models: Spiral convolutional networks for 3d shape representation learning and generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7213–7222, 2019. 2
- [6] Alan Brunton, Timo Bolkart, and Stefanie Wuhler. Multilinear wavelets: A statistical shape space for human faces. In *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part I 13*, pages 297–312. Springer, 2014. 2
- [7] Chen Cao, Yanlin Weng, Shun Zhou, Yiyong Tong, and Kun Zhou. Facewarehouse: A 3d facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):413–425, 2013. 2
- [8] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 1
- [9] Zhixiang Chen and Tae-Kyun Kim. Learning feature aggregation for deep 3d morphable models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13164–13173, 2021. 2
- [10] Shiyang Cheng, Michael Bronstein, Yuxiang Zhou, Irene Kotsia, Maja Pantic, and Stefanos Zafeiriou. Meshgan: Non-linear 3d morphable models of faces. *arXiv preprint arXiv:1903.10384*, 2019. 2
- [11] Yen-Chi Cheng, Hsin-Ying Lee, Sergey Tulyakov, Alexander G Schwing, and Liang-Yan Gui. Sdfusion: Multimodal 3d shape completion, reconstruction, and generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4456–4465, 2023. 2
- [12] Angela Dai, Charles Ruizhongtai Qi, and Matthias Nießner. Shape completion using 3d-encoder-predictor cnns and shape synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5868–5877, 2017. 1, 2, 3
- [13] Hang Dai, Nick Pears, William Smith, and Christian Duncan. Statistical modeling of craniofacial shape and texture. *International Journal of Computer Vision*, 128:547–571, 2020. 2, 3
- [14] Hang Dai, Nick Pears, William AP Smith, and Christian Duncan. A 3d morphable model of craniofacial shape and texture variation. In *Proceedings of the IEEE international conference on computer vision*, pages 3085–3093, 2017. 2, 3
- [15] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 4
- [16] Zhenfeng Fan, Silong Peng, and Shihong Xia. Towards fine-grained optimal 3d face dense registration: An iterative dividing and diffusing method. *International Journal of Computer Vision*, pages 1–21, 2023. 3, 8
- [17] Yao Feng, Haiwen Feng, Michael J Black, and Timo Bolkart. Learning an animatable detailed 3d face model from in-the-wild images. *ACM Transactions on Graphics*, 40(4):1–13, 2021. 2
- [18] Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019. 3
- [19] Michael Garland and Paul S Heckbert. Surface simplification using quadric error metrics. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 209–216, 1997. 3
- [20] Simon Giebenhain, Tobias Kirschstein, Markos Georgopoulos, Martin Rünz, Lourdes Agapito, and Matthias Nießner. Learning neural parametric head models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21003–21012, 2023. 2, 5
- [21] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 4
- [22] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics*, 36(4):1–14, 2017. 1, 3
- [23] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 3
- [24] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 3
- [25] Tianye Li, Timo Bolkart, Michael J Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4d scans. *ACM Transactions on Graphics*, 36(6):194–1, 2017. 2, 5
- [26] Or Litany, Alex Bronstein, Michael Bronstein, and Ameesh Makadia. Deformable shape completion with graph convolutional autoencoders. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1886–1895, 2018. 2
- [27] Hongyu Liu, Ziyu Wan, Wei Huang, Yibing Song, Xintong Han, and Jing Liao. Pd-gan: Probabilistic diverse gan for image inpainting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9371–9381, 2021. 1, 3
- [28] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of*

- the *IEEE international conference on computer vision*, pages 3730–3738, 2015. 4
- [29] Charles Loop. Smooth subdivision surfaces based on triangles. 1987. 5
- [30] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11461–11471, 2022. 1, 2, 4
- [31] David Palousek, Jiri Rosicky, and Daniel Koutny. Use of digital technologies for nasal prosthesis manufacturing. *Prosthetics and orthotics international*, 38(2):171–175, 2014. 1
- [32] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter. A 3d face model for pose and illumination invariant face recognition. In *2009 sixth IEEE international conference on advanced video and signal based surveillance*, pages 296–301. Ieee, 2009. 2
- [33] Anurag Ranjan, Timo Bolkart, Soubhik Sanyal, and Michael J Black. Generating 3d faces using convolutional mesh autoencoders. In *Proceedings of the European conference on computer vision*, pages 704–720, 2018. 2, 3
- [34] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3d deep learning with pytorch3d. *arXiv:2007.08501*, 2020. 4
- [35] David Stutz and Andreas Geiger. Learning 3d shape completion from laser scan data with weak supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1955–1964, 2018. 1, 2
- [36] Lyne P Tchampi, Vineet Kosaraju, Hamid Rezaatofghi, Ian Reid, and Silvio Savarese. Topnet: Structural point cloud decoder. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 383–392, 2019. 2
- [37] Nitika Verma, Edmond Boyer, and Jakob Verbeek. Feastnet: Feature-steered graph convolutions for 3d shape analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2598–2606, 2018. 3
- [38] Daoye Wang, Prashanth Chandran, Gaspard Zoss, Derek Bradley, and Paulo Gotardo. Morf: Morphable radiance fields for multiview neural head modeling. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–9, 2022. 2
- [39] Xiaogang Wang, Marcelo H Ang, and Gim Hee Lee. Voxel-based network for shape completion by leveraging edge generation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 13189–13198, 2021. 2
- [40] Erroll Wood, Tadas Baltrušaitis, Charlie Hewitt, Matthew Johnson, Jingjing Shen, Nikola Milosavljević, Daniel Wilde, Stephan Garbin, Toby Sharp, Ivan Stojiljković, et al. 3d face reconstruction with dense landmarks. In *European Conference on Computer Vision*, pages 160–177. Springer, 2022. 2
- [41] Peng Xiang, Xin Wen, Yu-Shen Liu, Yan-Pei Cao, Pengfei Wan, Wen Zheng, and Zhizhong Han. Snowflakenet: Point cloud completion by snowflake point deconvolution with skip-transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5499–5509, 2021. 1, 2
- [42] Xingguang Yan, Liqiang Lin, Niloy J Mitra, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. Shapeformer: Transformer-based shape completion via sparse representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6239–6249, 2022. 1, 2
- [43] Haotian Yang, Hao Zhu, Yanru Wang, Mingkai Huang, Qiu Shen, Ruigang Yang, and Xun Cao. Facescape: a large-scale high quality 3d face dataset and detailed riggable 3d face prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 601–610, 2020. 2, 3, 5
- [44] Tarun Yenamandra, Ayush Tewari, Florian Bernard, Hans-Peter Seidel, Mohamed Elgharib, Daniel Cremers, and Christian Theobalt. i3dmm: Deep implicit 3d morphable model of human heads. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12803–12813, 2021. 2
- [45] Xumin Yu, Yongming Rao, Ziyi Wang, Zuyan Liu, Jiwen Lu, and Jie Zhou. PointR: Diverse point cloud completion with geometry-aware transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12498–12507, 2021. 2
- [46] Yanhong Zeng, Jianlong Fu, Hongyang Chao, and Baining Guo. Learning pyramid-context encoder network for high-quality image inpainting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1486–1494, 2019. 1
- [47] Junbo Zhao, Yoon Kim, Kelly Zhang, Alexander Rush, and Yann LeCun. Adversarially regularized autoencoders. In *International conference on machine learning*, pages 5902–5911. PMLR, 2018. 3
- [48] Mingwu Zheng, Hongyu Yang, Di Huang, and Liming Chen. Imface: A nonlinear 3d morphable face model with implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20343–20352, 2022. 2, 5
- [49] Linqi Zhou, Yilun Du, and Jiajun Wu. 3d shape generation and completion through point-voxel diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5826–5835, 2021. 3
- [50] Liping Zhu, Bingyao Wang, Gangyi Tian, Wenjie Wang, and Chengyang Li. Towards point cloud completion: Point rank sampling and cross-cascade graph cnn. *Neurocomputing*, 461:1–16, 2021. 2
- [51] Wojciech Zielonka, Timo Bolkart, and Justus Thies. Towards metrical reconstruction of human faces. In *European Conference on Computer Vision*, pages 250–269. Springer, 2022. 3
- [52] Michael Zollhöfer, Justus Thies, Pablo Garrido, Derek Bradley, Thabo Beeler, Patrick Pérez, Marc Stamminger, Matthias Nießner, and Christian Theobalt. State of the art on monocular 3d face reconstruction, tracking, and applications. 37(2):523–550, 2018. 2