An End-to-End Conditional Generative Adversarial Network Based on Depth Map for 3D Craniofacial Reconstruction

Niankai Zhang[1] , Junli Zhao[1*] , Fuqing Duan[2*] , Zhenkuan Pan[1] , Zhongke Wu[2] , Mingquan Zhou[3] , Xianfeng Gu[4]

[1]College of Computer Science and Technology, Qingdao University, Qingdao 266071, China
[2]Virtual Reality Research Center of Ministry of Education, Beijing Normal University, Beijing 100875, China
[3]School of Information Science and Technology, Northwest University, Xi'an 710127, China
[4]Department of computer Science, Stony Brook University, Stony Brook, 11790, USA
* Correspondence authors: zhaojl@yeah.net, fqduan@bnu.edu.cn

ABSTRACT

Craniofacial reconstruction is fundamental in resolving forensic cases. It is rather challenging due to the complex topology of the craniofacial model and the ambiguous relationship between a skull and the corresponding face. In this paper, we propose a novel approach for 3D craniofacial reconstruction by utilizing Conditional Generative Adversarial Networks (CGAN) based on craniofacial depth map. More specifically, we treat craniofacial reconstruction as a mapping problem from skull to face. We represent 3D craniofacial shapes with depth maps, which include most craniofacial features for identification purposes and are easy to generate and apply to neural networks. We designed an end-to-end neural networks model based on CGAN then trained the model with paired craniofacial data to automatically learn the complex nonlinear relationship between skull and face. By introducing body mass index classes(BMIC) into CGAN, we can realize objective reconstruction of 3D facial geometry according to its skull, which is a complicated 3D shape generation task with different topologies. Through comparative experiments, our method shows accuracy and verisimilitude in craniofacial reconstruction results.

CCS CONCEPTS

• **Computing methodologies** → **Shape modeling**; **Reconstruction**; • **Applied computing** → **Computer forensics**.

KEYWORDS

Craniofacial Reconstruction, GANs, Depth Map, Body Mass Index Classes(BMIC), Neural Networks

ACM Reference Format:
Niankai Zhang[1] , Junli Zhao[1*] , Fuqing Duan[2*] , Zhenkuan Pan[1] , Zhongke Wu[2] , Mingquan Zhou[3] , Xianfeng Gu[4]. 2022. An End-to-End Conditional Generative Adversarial Network Based on Depth Map for 3D Craniofacial Reconstruction. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10–14, 2022, Lisbon, Portugal.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3503161.3548254

MM '22, October 10–14, 2022, Lisbon, Portugal

ACM ISBN 978-1-4503-9203-7/22/10…$15.00
https://doi.org/10.1145/3503161.3548254

1 INTRODUCTION

Craniofacial reconstruction(CFR) achieves the purpose of personal identification by reestablishing a resemblance of the facial appearance of an unknown body. When confronted with a decomposed, mutilated, or skeletonized body with all the other methods failed, craniofacial reconstruction can be the last technique to identify the unknown body[38]. The first evident endeavor of the CFR technique can date back to Neolithic times[43]. It is performed by manually putting plasters over a skull. In the 19th century, several attempts[16, 22, 45]had been made to obtain the soft tissue depth measurements of the face, which leads traditional manual reconstruction towards systematically scientific[43]. However, manual reconstruction requires the performer to have a high degree of anatomical and artistic modeling expertise, and it is also time-consuming and subjective[4].

Craniofacial reconstruction is a more difficult task than general task since it is to generated a correspondence face of a specified target skull, not a random face. Most of the craniofacial reconstruction methods are based on the relationship between the soft tissues and the underlying skull [38]. However, this relationship is nonlinear and complex, and existing methods are controversial because of the lack of complete understanding of this relationship [42]. Moreover, craniofacial data is high-dimensionally complex, and skull and skin are of different topologies. Therefore, it isn't easy to perform craniofacial reconstruction on 3D meshes directly, which are difficulty to be applied to neural networks. A common approach is to represent craniofacial data in feature space by dimensionality reduction and perform craniofacial reconstruction in feature space. These methods usually obtain accurate reconstructions. However, the reconstructions often lack high-frequency details because of the representation of craniofacial data in a low-dimensional feature space.

In order to solve the above problems, we propose to represent 3D craniofacial shapes with depth maps and use neural networks as a regression model to learn the mapping from skull to face. Neural networks have a significant advantage in solving nonlinear problems and retaining high-frequency details.Craniofacial reconstruction can be realized according the relationship between the skull and face, which is to be learned in our network. The use of depth map representation avoids the problem of the different topologies of the

skull and face. And depth map preserves the high-frequency details of craniofacial data to a large extent.

The main contributions of our work are as follow:(1) We design a novel end-to-end, CGAN-based neural networks model for craniofacial reconstruction. Meanwhile, body mass index classes(BMIC) are introduced to improve the accuracy of reconstruction. we apply an end-to-end CGAN to 3D craniofacial reconstruction successfully, which is a difficult 3D shape generation task between different typology models. (2) We selected a suitable representation for 3D craniofacial data, depth maps. Depth map is easy to generate and includes most of the craniofacial features for identification purposes. It is also convenient to apply neural networks, especially lightweighted and efficient Convolutional Neural Networks(CNNs). (3) Through comparative experimentation with existing end-to-end GANs, we demonstrate the superiority of our model in craniofacial reconstruction. We also conducted an ablation study on our approach to better understand the impact of different parts of our model on craniofacial reconstruction.

2 RELATED WORK

2.1 Computer-assisted Craniofacial Reconstruction

Computer-assisted methods can mainly be divided into knowledge-based methods and learning-based methods. Knowledge-based methods reconstruct the facial characteristics based on pre-measured craniofacial soft tissue thickness at different locations [36, 40]. This category of methods is a machinery reproduction of traditional manual reconstruction, e.g., Gietzen [11].

In learning-based methods[20, 26, 34], high-dimensional and complex craniofacial data are usually represented in a low- dimensional feature space. Later a mapping function from skull to face is obtained in the low-dimensional feature space by machine learning. Li [26] established a statistical model of craniofacial data and trained a least square support vector regression model in the parameter space. Paysan [34]used ridge regression to learn the mapping in the parameter space and specified attributes, e.g., age, weight for the reconstruction target. Duan [6] used multilinear subspace analysis to extract the features of craniofacial subspace with the attributes, e.g., age and BMI for establishing a mapping based on partial least square regression. Xiao [46] used Gaussian Process Latent Variable Models to represent craniofacial data and employed least square support vector machine regression to establish the mapping from skull to face in the latent space. Learning-based methods usually obtain accurate reconstructions, but the reconstructions often miss the details of the face due to the use of statistical models[14].

2.2 Generating 3D Shape with Neural Networks

Recently, with Convolutional Neural Networks(CNNs) [23] showing promising in image generation, manipulation, and completion, etc., some works have been trying to apply CNNs to 3D shapes. However, traditional CNNs can not be directly applied to non-Euclidean 3D shapes. Researchers put forward two kinds of methods to solve this problem: defining convolution-like structures in non-Euclidean space and representing 3D shapes in Euclidean space. The first type of method [2, 15, 28, 31, 44] is usually applied to 3D shapes with large shape differences, e.g., tables and chairs; therefore, it is not suitable for our question. CFR-GAN[35] proposed a deep generative model for craniofacial reconstruction which has the ability to generate high-fidelity face images. 3D face provides a more sufficient way for personal identification, but above method are not for 3D craniofacial reconstruction.

In the second type of method, the most common representation is images. Gilani [12] calculate the depth, azimuth, and elevation of each vertex of a mesh and use it to generate a three-channel image. Galteri [9] used the same method to get the image representation of a 3D mesh, except they replaced the azimuth value with curvature, as their work values the curvature more. Feng [8] proposed to store coordinates of points of a 3D face model in a UV position map. Instead of directly dealing with 3D shapes, some methods try to regress the parameters of a statistic model of 3D shapes by using differentiable renderers[10, 28].

In our case, a depth map is a good option for Euclidean representation, for it is easy to generate and convenient for applying neural networks. And a depth map projected from the frontal face preserves most of the facial features, which meets the need for craniofacial reconstruction to a large extent.

2.3 GAN-based Image-to-image Mapping

Recently, Generative Adversarial Network(GAN) has attracted a lot of attention because of its ability to produce high-dimensionally complex and verisimilar data. GAN proposed by Goodfellow [13], which models data probability distribution through adversarial learning. GAN also learns the mapping between the input domain and the target domain. For example, the original GAN maps the noise to the target. CGAN[32] maps the noise and label conditions to the target of different categories. GAN-based mapping between image domains has been widely studied and achieved great success. Some works conditioned the generator on the input image through L_n regression loss, e.g., image repair [33], image super-resolution [24]. Pix2Pix [19]conditioned the discriminator on input to reduce the specificity of the task. CycleGAN[47] proposed the use of cycle consistency loss for self-supervision, which is quite effective. Li et al.[27] present a craniofacial reconstruction method that synthesizes craniofacial images from 2D computed tomography scan of skull data based on deep generative adversarial nets. However, we have never seen the method to solve the problem of craniofacial reconstruction on 3D craniofacial data with CGAN.

The problem we aim to solve can be regarded as a mapping from the 3D skull domain to the 3D face domain. Therefore, a GAN-based structure would be suitable for reconstructing realistic faces. Furthermore, through mapping 3D craniofacial data to 2D depth images, the problem of complex 3D craniofacial data applied to the neural network is solved successfully.

3 METHOD

3.1 Overview

Our proposed method treats craniofacial reconstruction as a mapping problem, where a CGAN performs the mapping process. Our model learns a mapping function from the skull domain to the face domain with paired skull-face data. To better utilize neural networks, we first represent 3D craniofacial shapes with depth maps,

then used a CGAN to perform an image-to-image mapping from skull image to face image based on BMIC. After that, the generated skin images are converted back into 3D shapes, as shown in Fig. 1.

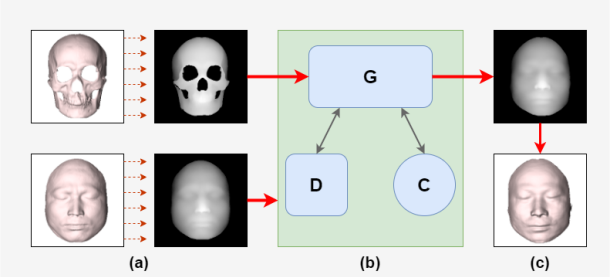


Figure 1: Overview of the proposed approach. The proposed approach contains three steps: (a)representing 3D craniofacial data with depth maps (b)training a CGAN model with paired skull-face depth maps to translate a skull into a face; Our model consists of a generator G, a discriminator D and a Classifier C (c)converting the generated face depth map back into a 3D shape.

3.2 Representation of Craniofacial Data

We represent 3D craniofacial geometry data with a 2D image, where the only channel represents the depth of the 3D shape. This representation allows us to construct our neural networks with 2D convolutional networks. It contains more comprehensive information and details than prior craniofacial representation method. Galteri [9] inspires this approach, where they used a 3-channel image to represent the geometry of a 3D shape, and the three channels contain respectively the depth value, the elevation value of the normal, and the mean curvature value. In our case, we experimentally found that the elevation channel and the curvature channel do not benefit the reconstruction. On the opposite, it introduces noises to our model and makes the reconstruction visually obscure. Hence we only adopted the depth channel for our model, and we elaborated on the tests of different representations in Sec.4.3.1.

The data used in this work are 3D skull-face mesh pairs facing directly in a positive y-axis direction. To obtain the required depth map, we employed a orthographic projection along the negative Y-axis to get the image plane and used the Y coordinate of the point as depth value, as shown in Fig.2. For the given point $P(x, y, z)\{x, y, z \in (0, 1)\}$, the index and pixel value in the image plane after projection become

$$\begin{cases} P_x = -\frac{1}{2}x \times w + \frac{1}{2}w \\ P_y = -\frac{1}{2}z \times h + \frac{1}{2}h \\ P_{depth} = y \end{cases} \tag{1}$$

, where w and h are the width and height of the image content respectively. The depth values are rescaled in the range of $[0, 255]$ to fit with the pixel value. Then a pixel value interpolation is conducted based on the triangle information to fill the void in the projected image. For a given blank pixel P, we calculate its barycentric coordinates (w_1, w_2, w_3) in the corresponding triangle (P_1, P_2, P_3) in Eq.(2).

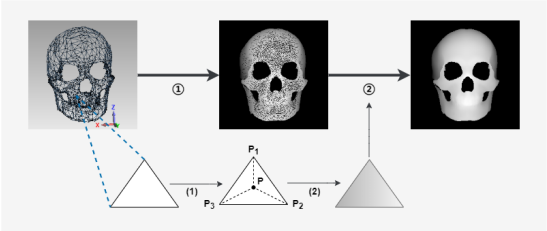


Figure 2: The depth map representation of craniofacial data was obtained by projection.①The vertices of the 3D mesh are projected to the 2D image plane.② The void in the projected image plane is filled by triangular interpolation.

$$P_{index} = w_1 P1_{index} + w_2 P2_{index} + w_3 P3_{index} \tag{2}$$

The pixel value at pixel P is calculated as the weighted average of pixel values at the triangle vertices by barycentric coordinates in Eq.(3).

$$P_{pixel} = w_1 P1_{pixel} + w_2 P2_{pixel} + w_3 P3_{pixel} \tag{3}$$

At last, a depth buffer is employed to prevent pixel value override when different points project to the same pixel.

3.3 Architecture of Our Model

Our model consists of a generator G for skull image to face image translation, a discriminator D for face image discrimination, and a classifier C for face image classification by its body mass index class(BIMC). As shown in Fig.3, generator G is trained by a combined constraint of pixel loss, adversarial loss, and BMIC loss.

3.3.1 Skull-face Generator G. Generator G is an end-to-end network for skull-to-face translation. It takes a symmetrical encoder-decoder as its structure. This symmetrical structure is constructed with three parts, starting with a convolutional layer and two downsample convolutional layers. Several residual blocks are in the middle, ending with two upsample transpose convolutional layers and another convolutional layer. We conditioned generator G on BMIC labels to generate face images with certain BMICs. To satisfy this purpose, we represent the BMIC label with a 3-channel feature map and insert the 3-channel feature map before the residual blocks. Moreover, we insert a convolutional layer to change the concatenated feature map back to the same channels. Details of generator G are shown in Tab. 1. Convolutional layers treat a local area of feature maps as a whole with connections, which allows our model to learn the relationship between skull and face as local-area to local-area mapping instead of point-to-point mapping. In which case, each point in the reconstructed face is based on the characteristics of a specific area of the target skull, rather than just a single point.

3.3.2 Face Discriminator D. Discriminator D aims to boost the generated face image towards real face image data distribution, making the generated face images more similar to real face images with more clear facial features. We use 70×70 PatchGAN[19, 24, 25, 47]for our discriminator. It consists of six convolutional layers without fully connected layers, starting with three downsample

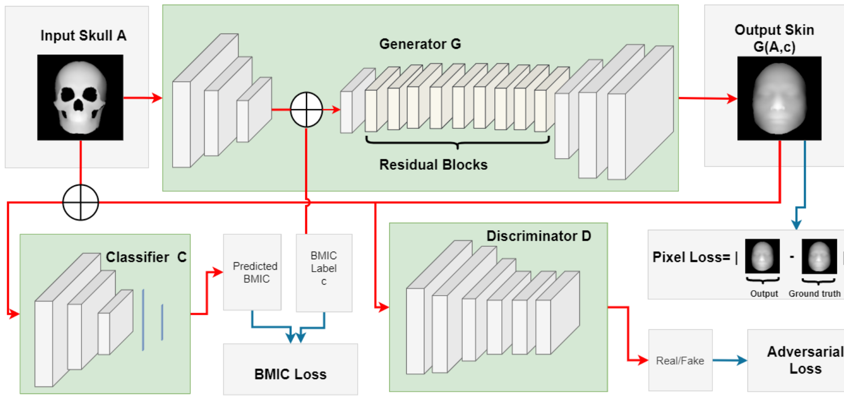


Figure 3: Detailed architecture of our networks. ⊕ denotes concatenation. Generator G translates input skull A into output face $G(A, c)$ conditioned on BMIC label c by adding c before the residual block. Classifier C predicts the BMIC of $G(A, c)$. Discriminator D discriminates whether $G(A, c)$ is real data. Generator G is trained by optimizing the weighted sum of pixel loss, BMIC loss, and adversarial loss.

convolutional layers, following one convolutional layer, in the end, a convolutional layer with the output of $70 \times 70 \times 1$. Discriminator D classifies whether every patch of all 70×70 patches is real or fake and outputs the average of all patches. In such a manner, our model would have fewer parameters and a shorter training duration.

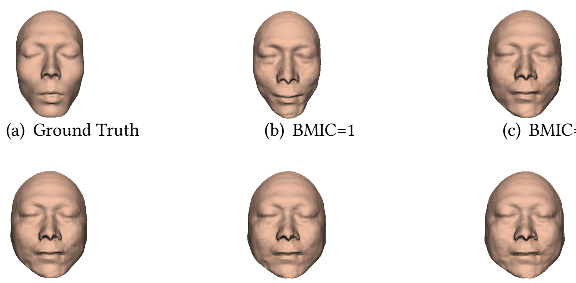
3.3.3 BMIC Classifier C. With generator G and discriminator D as our neural networks model, some of the reconstructions are accurate, but the others are less than satisfactory. We experimentally found that those test samples with good results are those with an average body mass. This situation is because the relationship between skull and face is not one-to-one, and the appearance of an individual varies with weight changing while the skull remains the same [34]. Aiming at this problem, we conditioned the input of generator G on BMIC labels. We manually divide the craniofacial data into five classes from level one to level five based on the visual estimation of an individual's body mass index(BMI) through its face. Since this classification is related to BMI, we referred to it as BMIC for short. To discriminate whether the generated faces match the right BMIC, classifier C is employed. We use LeNet[23] as our classifier, which is a classical classifier, and experimentally proved efficient to our task. We experimentally found that it is challenging for classifier C to classify with only face images as input. However, with both skull image and face image pair as input, the valuation of correct classification is high. Fig.4 showed the ground truth and the reconstructions with five different BMICs.

3.4 Loss Functions

To restrain the property of the output and guarantee wanted results, three types of losses are defined as explained below.

3.4.1 Pixel Loss. To control the generated face image comparable to the target face image, we define pixel loss in Eq.(4). Where y is the target face, and $G(x)$ is the generated face, and pixel loss L_{pix} is L_p norm between y and $G(x)$ in pixel level. In our case, we choose $p = 1$, as it has shown the best performance in our experiments.

$$L_{pix} = ||y - G(x)||_p \tag{4}$$



(a) Ground Truth (b) BMIC=1 (c) BMIC=2 (d) BMIC=3 (e) BMIC=4 (f) BMIC=5

Figure 4: Same input skull, different reconstructions with different BMICs.

Layer	Filter	Output Shape
Conv	7×7	$256 \times 256 \times 64$
Conv(downsample)	3×3	$128 \times 128 \times 128$
Conv(downsample)	3×3	$64 \times 64 \times 256$
Concat	-	$64 \times 64 \times 258$
Conv	3×3	$64 \times 64 \times 256$
ResidualBlock	3×3	$64 \times 64 \times 256$
ResidualBlock	3×3	$64 \times 64 \times 256$
ResidualBlock	3×3	$64 \times 64 \times 256$
ResidualBlock	3×3	$64 \times 64 \times 256$
ResidualBlock	3×3	$64 \times 64 \times 256$
ResidualBlock	3×3	$64 \times 64 \times 256$
ResidualBlock	3×3	$64 \times 64 \times 256$
ResidualBlock	3×3	$64 \times 64 \times 256$
ResidualBlock	3×3	$64 \times 64 \times 256$
TransConv(upsample)	3×3	$128 \times 128 \times 128$
TransConv(upsample)	3×3	$256 \times 256 \times 64$
Conv	3×3	$256 \times 256 \times 1$

Table 1: The network structure of generator G

3.4.2 Adversarial Loss. With only the pixel loss, generator G can generate face images similar to the target face images. However, the generated images are lack details, which means the generated images are unclear to human visual perception, especially in the eyes,

nose, and mouth regions of the face. To overcome this problem, we employed adversarial loss, as shown in Eq.(5). In Eq.(5), x and y are the depth map representation of the skull data and the face data. $G(x)$ denotes the face depth map generated by generator G by feeding input skull x. $D(y)$ and $D(G(x))$ denote the discriminant results of discriminator D for real face y and generated face $G(x)$, respectively. The generator G tries to trick the discriminator D by minimizing L_{adv}. The discriminator D attempts to recognize the face generated by generator G by maximizing L_{adv}. The adversarial loss pushes the generated face image towards real face image data distribution through adversarial learning of discriminator and generator.

$$\begin{aligned} L_{adv}(G,D) =& E_{y\sim p_{face}(y)}[logD(y)] \\ &+ E_{x\sim p_{skull}(x)}[log(1-D(G(x)))] \end{aligned} \tag{5}$$

3.4.3 BMIC Loss. BMIC loss is calculated between the BMIC classifying result of classifier C on the generated face and the BMIC label. The classical cross-entropy loss in the classification networks is adopted for BMIC loss, as shown in Eq.(6). $L_{BMIC}(G)$ prompts generator G to generate face images with correct BMICs. In Eq. (7), $L_{BMIC}(C)$ is the cross-entropy loss between the BMIC classifying result of classifier C on the real face image and the BMIC label, and it is for classifier C training.

$$L_{BMIC}(G) = -\sum_{i=1}^{n} c_i log(C(G(x))) \tag{6}$$

$$L_{BMIC}(C) = -\sum_{i=1}^{n} c_i log(C(y)) \tag{7}$$

When optimizing generator G, we used L_G, a sum of the weighted L_{pix}, L_{adv}, and L_{BMIC}, as the optimization target. As shown in Eq.(8), w_p, w_a and w_c are the weights of L_{pix}, L_{adv}, and L_{BMIC}, respectively.

$$L_G = w_p L_{pix} + w_a L_{adv} + w_c L_{BMIC} \tag{8}$$

4 EXPERIMENTS

In this section, we elaborated on the implementation details of the proposed method. We conducted an ablation study on our proposed approach to fully understand and evaluate the role and effect of each part for craniofacial reconstruction. Finally, we compared the proposed craniofacial reconstruction method with other end-to-end GANs and traditional machine learning methods.

4.1 Implementation Details

The coding part of our work was inspired by CycleGAN [47] and implemented in the PyTorch framework. In our networks, discriminator D uses 70×70 Patch-GANs [19, 24, 25, 47], and generator G is adapted from Johnson [21], in which BMIC conditions are added. For classifier C, we used the classical classification network LeNet [23], and test accuracy in the classification of BMIC was as high as 99%. For adversarial loss, we tested the original GAN loss [13], the LSGAN loss [30], and the WGAN loss [1], among which the LSGAN loss showed the best performance. We set the weight of the different parts of generator loss as $w_p = 1$, $w_a = 0.01$, $w_c = 0.005$, which showed the best performance in training.

For the training of generator G and discriminator D, we followed the training procedure of Shrivastav[39]. We maintained a buffer pool of generated face images of the last epoch of training to update the discriminator. As to the training of classifier C, we only used real craniofacial data. Classifier C is trained simultaneously with generator G and discriminator D but separately updating parameters.

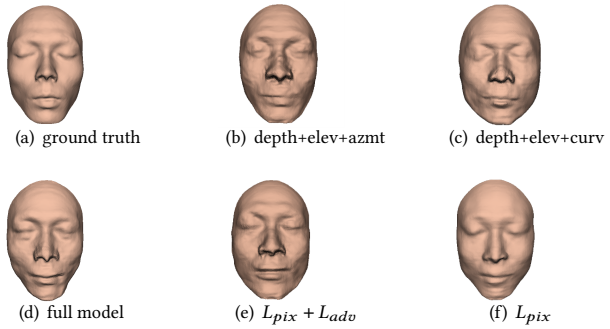
4.2 Data Augmentation

The experimental data used in this paper are head CT images. In order to obtain meshes, we first extracted the boundary of skulls and faces from the CT images[7] and then reconstructed meshes using the Marching Cube algorithm[29]. To obtain a uniform projected depth map, we adjusted the skull and face mesh to Frankfurt coordinate system [17], then performed TPS registration[18]. After the above process, the meshes of the front half skull and face, with the same scale and unified posture, are obtained.

With a total of 209 skull and skin 3D mesh pairs for experiments, among which 60 pairs are randomly selected for testing, and the remaining 149 pairs are for training. If one mesh is converted into one depth map, 149 pairs of images will be generated for training, which is far from sufficient to support the training of our networks. Therefore, we performed data augmentation by rotating the 3D mesh around X, Y, and Z-axis at random angles(-3,3) and then performed projections to obtain different depth maps. A 3D shape gets nine images, and a total of 1341 images could be obtained for the training of our model. The data augmentation we took increases not only the size of the training data but also enhances the robustness of our model for the problem of incomplete unity of face and skull posture.

4.3 Results and Analysis

4.3.1 Ablation Study. In this chapter, in order to more clearly understand the influence of different parts of our model on craniofacial reconstruction, we carried out an ablation study for this model. Fig.5 shows different reconstruction results when adding or deleting different components in our model.



(a) ground truth (b) depth+elev+azmt (c) depth+elev+curv

(d) full model (e) $L_{pix} + L_{adv}$ (f) L_{pix}

Figure 5: An ablation study on the impact of different components of our method on craniofacial reconstruction.

Reconstruction Error	depth+elev+azmt	depth+elev+curv	full model	L_{pix}	$L_{pix} + L_{adv}$
Min	0.00000092	0.00000038	**0.00000015**	0.00000040	0.00000077
Mean	0.02060362	0.01221315	0.00787630	**0.00775745**	0.01865223
Max	0.08330410	0.06238850	0.05656357	**0.05278093**	0.08557593
Variance	0.00029790	0.00011250	**0.00003635**	0.00003678	0.00021415

Table 2: Reconstruction error statics of models tested in the ablation study experiment for 60 tests

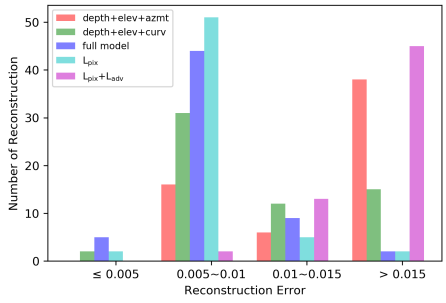


Figure 6: The histogram of the mean error for 60 reconstructions of models tested in the ablation study

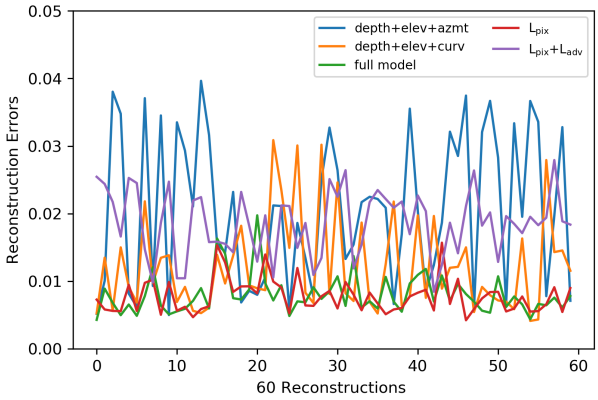


Figure 7: The mean error for 60 reconstructions of models tested in the ablation study

We conducted the ablation study from two aspects, different representations of craniofacial data and loss function. From different representations, we added geometric information, e.g., curvature, elevation, and azimuth, to the depth map to form a 3-channel image. According to the training strategy by Galteri [9], we make the discriminator accept only the depth value as input to reduce the noises that other channels introduced. However, the reconstruction results showed that the geometrical information did not make the reconstruction more accurate but introduced noise. It may suggest that curvature, elevation, and azimuth has little effect on craniofacial reconstruction.

From different loss functions, we can see from the obtained results that L_{pix} controls the generated face to have an overall similarity with the ground truth but can't guarantee clear local details. L_{adv} guarantees the generation of realistic details. For individuals with very high or very low BMI, L_{pix} and L_{adv} could not control the accuracy of the results generated, and the results were more accurate by adding L_{BMIC}.

At the same time, we conducted error statistics for 60 reconstruction results of the above five models, as shown in Fig.6,Fig.7,and Tab.2. Among these, the model only used L_{pix} has the lowest error, including mean error and maximum error. And most of the mean reconstruction errors fall in a low range, which conforms to our expectation that the overall similarity brings the lowest error. The reconstruction error increases with the addition of L_{adv}, but the high-frequency details of the reconstruction results are improved. After the addition of L_{BMIC}, the error dropped to a level only higher than that of L_{pix} alone. It is in line with our explanation in Sec.3.4.3, the addition of L_{BMIC} improved the accuracy of the reconstruction. For the input of depth, elevation, and azimuth and the input of depth, elevation, and curvature, even with the full model (with the addition of L_{BMIC}) , there is still a high error, indicating that elevation, azimuth, and curvature introduce a lot of noise.

4.3.2 Comparisons with deep learning methods. In this chapter, we compared our proposed approach with other domain-to-domain mapping GANs. We calculated the reconstruction errors to evaluate the efficiency of the proposed method. Fig.8 shows some of the reconstruction results and reconstruction error graphs represented in rainbow color scale. Here, we calculated the geometric deviation between the reconstructed face and the ground truth for the reconstruction error by employing the algorithm from [37, 41]. Fig.8 tells that the reconstruction results obtained by our method are similar to the ground truth in human visual perception, while Pix2Pix reconstructions appear blurred and CycleGAN reconstructions are less identical to the ground truth. It also shows a lower reconstruction error our method has.

Reconstruction Error	Pix2Pix	CycleGAN	Ours
Min	0.00000450	0.00000020	**0.00000015**
Mean	0.01759093	0.01136900	**0.00787630**
Max	0.07239635	0.05715283	**0.05656357**
Variance	0.00007280	0.00013470	**0.00003635**

Table 3: Reconstruction error comparison with CycleGAN and Pix2Pix, lower is better.

We conducted reconstruction error statistics on the reconstruction results of 60 sets of test data. Compared with Pix2Pix and CycleGAN, the average error of the reconstructions by our model was significantly lower, as shown in Fig.10(a). In addition, among the 60 test data, more than 80% of the average error of reconstruction results of our model fell in a low range (< 0.01), which was much higher than both Pix2Pix and CycleGAN, as shown in Fig.10(b). It

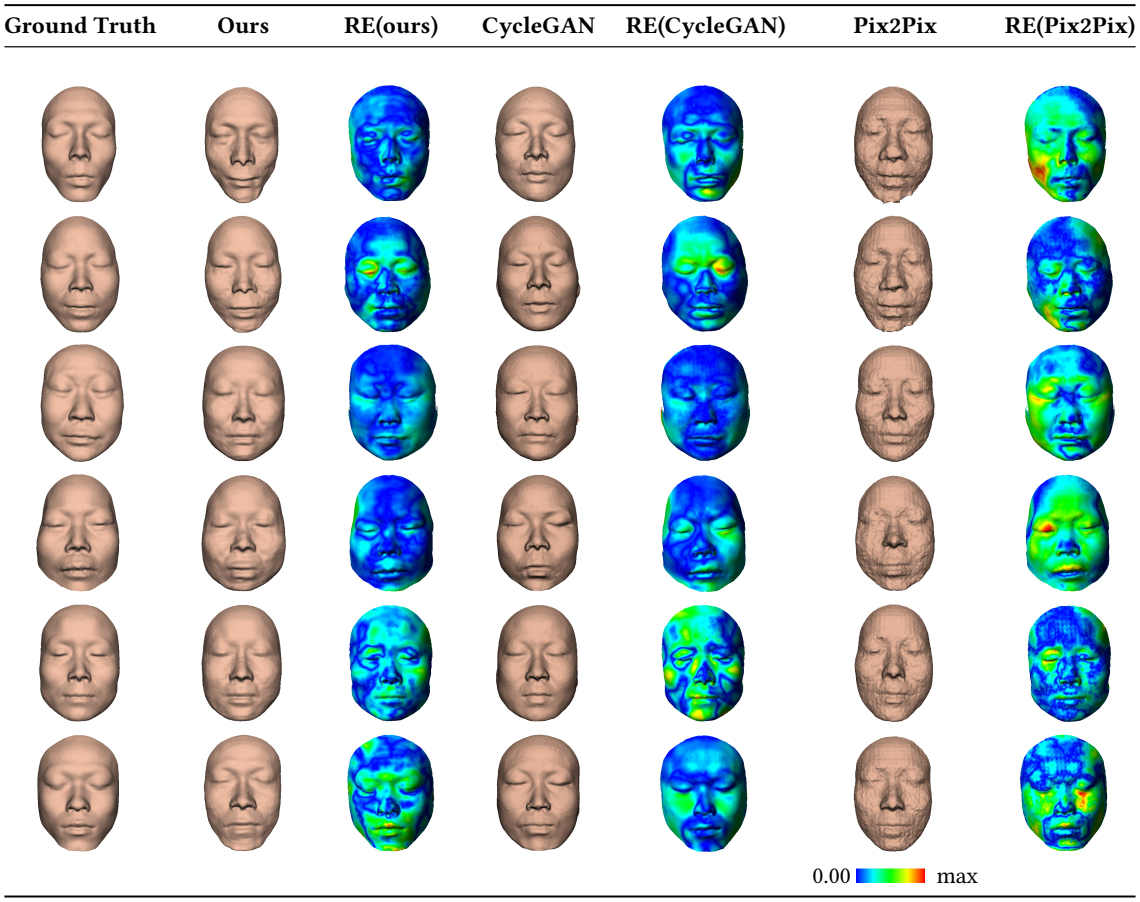


Figure 8: Reconstruction result quality comparison. RE denotes Reconstruction Error. From left to right are: the ground truth, the reconstruction results of our method, the reconstruction error graph of our method, the CycleGAN reconstruction results, the reconstruction error graphs of CycleGAN, the Pix2Pix reconstruction results, the reconstruction error graphs of Pix2Pix.

suggests the high robustness of our model. At the same time, the minimum, maximum, and variance of the reconstruction errors by our model are all lower than Pix2Pix and CycleGAN, as shown in Tab.3. It is a good indication that our approach works well for craniofacial reconstruction.

4.3.3 Comparison with traditional machine learning methods . We also compared our proposed approach with traditional machine learning methods. Fig.9 shows some of the reconstruction results and reconstruction error graphs represented in rainbow color scale. Fig.9 tells that the reconstruction results obtained by our method are much more similar to the ground truth than HF-GGR[20], FMM-GR[3], and PCA [5]. It also shows a lower reconstruction error our method has.

We conducted reconstruction error statistics on the reconstruction results of 60 sets of test data. As shown in Fig.10(c), our method has the lowest mean reconstruction error. At the same time, the minimum, maximum, and variance of the reconstruction errors by our model are all lower than HF-GGR[20], FMM-GR[3], and PCA [5], as shown in Tab.4. It also shows a lower reconstruction error our method has and illustrates the effectiveness of our approach.

Reconstruction Error	FMM-GR	HF-GGR	PCA	Ours
Min	0.00013065	0.00000050	0.00245142	**0.00000015**
Mean	0.02080330	0.01738085	0.05504495	**0.00787630**
Max	0.07067193	0.06589645	0.15359872	**0.05656357**
Variance	0.00007280	0.00017700	0.00109160	**0.00003635**

Table 4: Reconstruction error statics comparison with traditional machine learning methods, lower is better.

5 CONCLUSION

Craniofacial reconstruction plays a very important role in forensic investigations. In this paper, we propose a method for 3D craniofacial reconstruction by utilizing neural networks. With the help of the domain-to-domain mapping capability of GANs, we construct a novel end-to-end network that directly maps the input skull to the output face. In future work, we plan to apply a more accurate representation for 3D geometric craniofacial data to improve the accuracy of the reconstruction results. On this basis, we plan to generate the texture information of the corresponding face according

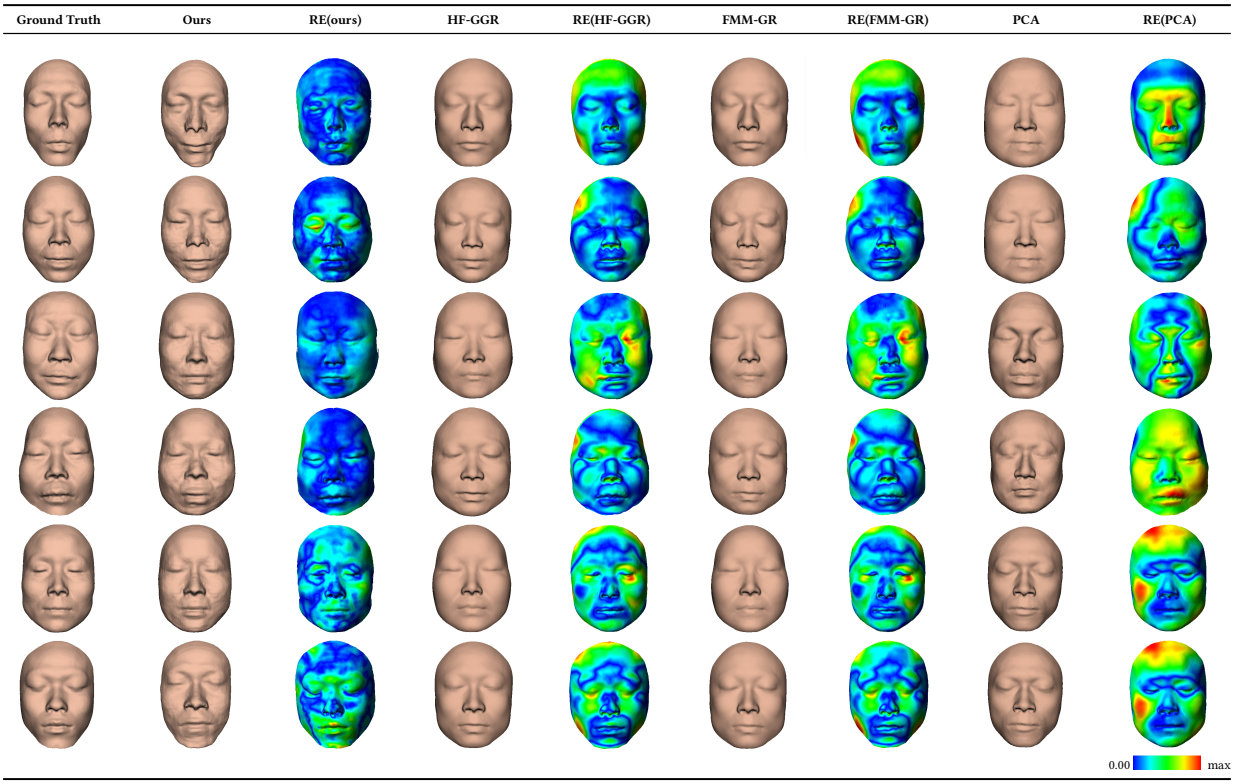
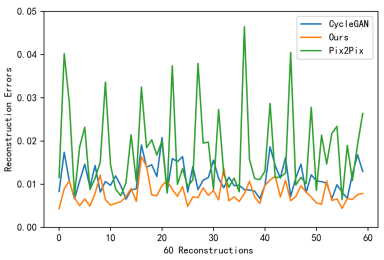
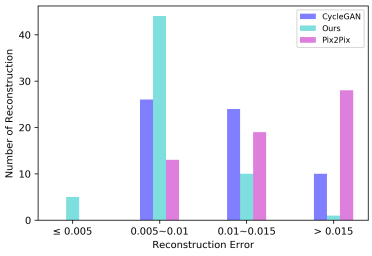


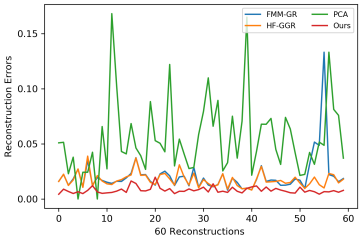
Figure 9: Reconstruction result quality comparison with machine learning methods. RE denotes Reconstruction Error. From left to right are: the ground truth, the reconstruction results of our method, the reconstruction error graph of our method, the HF-GGR reconstruction results, the reconstruction error graphs of HF-GGR, the FMM-GR reconstruction results, the reconstruction error graphs of FMM-GR, the PCA reconstruction results, the reconstruction error graphs of PCA.



(a) Pix2Pix, CycleGAN and our model



(b) Histogram of Pix2Pix, CycleGAN and our model



(c) FMM-GR, HF-GGR, PCA and our model

Figure 10: The mean error for 60 reconstructions

to some physical attributes of the input skull data, e.g., age, gender, or race, to make the result more realistic.

ACKNOWLEDGMENTS

The authors gratefully appreciated the anonymous reviewers for all of their helpful comments. This work was supported by the National Natural Science Foundation of China under Grant Nos.62172247, 61702293, National Statistical Science Research Project (No.2020 LY100) , National Natural Science Foundation of Shandong Province (No.ZR2019LZH002), NSF 2115095, NSF 1762287, NIH 92025 and NIH R01LM012434. We also thank the support of Xianyang Hospital for providing craniofacial data.

REFERENCES

[1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein generative adversarial networks. In *International conference on machine learning*. PMLR, 214–223.

[2] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. 2013. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203* (2013).

[3] Zhonghan Chen, Junli Zhao, Han Yu, Zhenkuan Pan, Bin Jia, and Jinhua Li. 2021. Craniofacial Reconstruction Based on Geodesic Regression Model. *Journal of Computer-Aided Design Computer Graphics* 33 (03 2021), 395–404. https://doi.org/10.3724/SP.J.1089.2021.18320

[4] Peter Claes, Dirk Vandermeulen, Sven De Greef, Guy Willems, John Gerald Clement, and Paul Suetens. 2010. Computerized craniofacial reconstruction: conceptual framework and review. *Forensic science international* 201, 1-3 (2010), 138–145.

[5] Michel Desvignes, Gerard Bailly, Yohan Payan, and Maxime Berar. 2006. 3D semi-landmarks based statistical face reconstruction. *Journal of computing and Information technology* 14, 1 (2006), 31–43.

[6] Fuqing Duan, Sen Yang, Donghua Huang, Yongli Hu, Zhongke Wu, and Mingquan Zhou. 2014. Craniofacial reconstruction based on multi-linear subspace analysis. *Multimedia Tools and Applications* 73, 2 (2014), 809–823.

[7] Fuqing Duan, Yanchao Yang, Yan Li, Yun Tian, Ke Lu, Zhongke Wu, and Mingquan Zhou. 2014. Skull identification via correlation measure between skull and face shape. *IEEE transactions on information forensics and security* 9, 8 (2014), 1322–1332.

[8] Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, and Xi Zhou. 2018. Joint 3d face reconstruction and dense alignment with position map regression network. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 534–551.

[9] Leonardo Galteri, Claudio Ferrari, Giuseppe Lisanti, Stefano Berretti, and Alberto Del Bimbo. 2019. Deep 3D morphable model refinement via progressive growing of conditional Generative Adversarial Networks. *Computer Vision and Image Understanding* 185 (2019), 31–42.

[10] Baris Gecer, Stylianos Ploumpis, Irene Kotsia, and Stefanos Zafeiriou. 2019. Ganfit: Generative adversarial network fitting for high fidelity 3d face reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1155–1164.

[11] Thomas Gietzen, Robert Brylka, Jascha Achenbach, Katja Zum Hebel, Elmar Schömer, Mario Botsch, Ulrich Schwanecke, and Ralf Schulze. 2019. A method for automatic forensic facial reconstruction based on dense statistics of soft tissue thickness. *PloS one* 14, 1 (2019), e0210257.

[12] Syed Zulqarnain Gilani, Ajmal Mian, and Peter Eastwood. 2017. Deep, dense and accurate 3D face correspondence for generating population specific deformable models. *Pattern Recognition* 69 (2017), 238–250.

[13] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial networks. *arXiv preprint arXiv:1406.2661* (2014).

[14] Grd, Petra, and Ena Barčić. 2021. A Survey on 3D Digital Facial Reconstruction Algorithms. In *CECIIS*.

[15] Rana Hanocka, Amir Hertz, Noa Fish, Raja Giryes, Shachar Fleishman, and Daniel Cohen-Or. 2019. Meshcnn: a network with an edge. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–12.

[16] Wilhelm His. 1895. *Anatomische Forschungen über Johann Sebastian Bach's Gebeine und Antlitz, nebst Bemerkungen über dessen Bilder, von Wilhelm His.* S. Hirzel.

[17] Yongli Hu, Fuqing Duan, Baocai Yin, Mingquan Zhou, Yanfeng Sun, Zhongke Wu, and Guohua Geng. 2013. A hierarchical dense deformable model for 3D face reconstruction from skull. *Multimedia tools and applications* 64, 2 (2013), 345–364.

[18] Ruikun Huang, Junli Zhao, Fuqing Duan, Xin Li, Celong Liu, Xiaodan Deng, Zhenkuan Pan, Zhongke Wu, and Mingquan Zhou. 2019. Automatic craniofacial registration based on radial curves. *Computers & Graphics* 82 (2019), 264–274.

[19] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1125–1134.

[20] Bin Jia, Junli Zhao, Shiqing Xin, Fuqing Duan, Zhenkuan Pan, Zhongke Wu, Jinhua Li, and Mingquan Zhou. 2021. Craniofacial reconstruction based on heat flow geodesic grid regression (HF-GGR) model. *Computers & Graphics* 97 (2021), 258–267.

[21] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*. Springer, 694–711.

[22] Julius Kollmann and Werner Büchly. 1898. *Die persistenz der rassen und die reconstruction der physiognomie prähistorischer Schädel.* Archiv f. Anthrop.

[23] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.

[24] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, and Zehan Wang. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4681–4690.

[25] Chuan Li and Michael Wand. 2016. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *European conference on computer vision*. Springer, 702–716.

[26] Yan Li, Liang Chang, Xuejun Qiao, Rong Liu, and Fuqing Duan. 2014. Craniofacial reconstruction based on least square support vector regression. In *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 1147–1151.

[27] Yuan Li, Jian Wang, Weibo Liang, Hui Xue, Zhenan He, Jiancheng Lv, and Lin Zhang. 2022. CR-GAN: Automatic Craniofacial Reconstruction for Personal Identification. *Pattern Recognition* 124, 2022 (2022), 1–12.

[28] Jiangke Lin, Yi Yuan, Tianjia Shao, and Kun Zhou. 2020. Towards high-fidelity 3D face reconstruction from in-the-wild images using graph convolutional networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5891–5900.

[29] William E Lorensen and Harvey E Cline. 1987. Marching cubes: A high resolution 3D surface construction algorithm. *ACM siggraph computer graphics* 21, 4 (1987), 163–169.

[30] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. 2017. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*. 2794–2802.

[31] Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. 2015. Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of the IEEE international conference on computer vision workshops*. 37–45.

[32] Mehdi Mirza and Simon Osindero. 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* (2014).

[33] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. 2016. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2536–2544.

[34] Pascal Paysan, Marcel Lüthi, Thomas Albrecht, Anita Lerch, Brian Amberg, Francesco Santini, and Thomas Vetter. 2009. Face reconstruction from skull shapes and physical attributes. In *Joint Pattern Recognition Symposium*. Springer, 232–241.

[35] Pengyue, Lin, Yang Wen, Xia Siyuan, Jiang Yu, Liu Xiaoning, and Geng Guohua. 2021. CFR-GAN: A Generative Model for Craniofacial Reconstruction. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 462–469.

[36] VM Phillips and NA Smuts. 1996. Facial reconstruction: utilization of computerized tomography to measure facial tissue thickness in a mixed racial population. *Forensic science international* 83, 1 (1996), 51–59.

[37] Michaël Roy, Sebti Foufou, and Frédéric Truchetet. 2004. Mesh comparison using attribute deviation metric. *International Journal of Image and Graphics* 4, 01 (2004), 127–140.

[38] Rinchon S, Arpita S, Mahipal S, and Rajeev K. 2018. 3D Forensic Facial Reconstruction: A Review of the Traditional Sculpting Methods and Recent Computerised Developments. *International Journal of Foren-sic Sciences* 3, 1, 000134.

[39] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang, and Russell Webb. 2017. Learning from simulated and unsupervised images through adversarial training. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2107–2116.

[40] Wuyang Shui, Mingquan Zhou, Qingqiong Deng, Zhongke Wu, Yuan Ji, Kang Li, Taiping He, and Haiyan Jiang. 2016. Densely calculated facial soft tissue thickness for craniofacial reconstruction in Chinese adults. *Forensic science international* 266 (2016), 573. e1–573. e12.

[41] Samuel Silva, Joaquim Madeira, and Beatriz Sousa Santos. 2005. Polymeco-a polygonal mesh comparison tool. In *Ninth International Conference on Information Visualisation (IV'05)*. IEEE, 842–847.

[42] Andrew J Tyrrell, Martin P Evison, Andrew T Chamberlain, and Michael A Green. 1997. Forensic three-dimensional facial reconstruction: historical review and contemporary developments. *Journal of Forensic Science* 42, 4 (1997), 653–661.

[43] Laura Verzé. 2009. History of facial reconstruction. *Acta Biomed* 80, 1 (2009), 5–12.

[44] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Hang Yu, Wei Liu, Xiangyang Xue, and Yu-Gang Jiang. 2020. Pixel2Mesh: 3D mesh model generation via image guided deformation. *IEEE transactions on pattern analysis and machine intelligence* (2020).

[45] Hermann Welcker. 1883. *Schiller's Schädel und Todtenmaske: nebst Mittheilungen über Schädel und Todtenmaske Kant's. Mit einem Titelbilde, 6 Lithographirten Tafeln und 29 in den Text Eingedruckten Holzstichen.* Friedrich Vieweg und Sohn.

[46] Zedong Xiao, Junli Zhao, Xuejun Qiao, and Fuqing Duan. 2015. Craniofacial reconstruction using gaussian process latent variable models. In *International Conference on Computer Analysis of Images and Patterns*. Springer, 456–464.

[47] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*. 2223–2232.