Intrinsic Morphological Relationship Guided 3D Craniofacial Reconstruction Using Siamese Cycle Attention GAN

Junli Zhao
College of Computer Science & Technology, Qingdao University
China
zhaojl@yeah.net

Chengyuan Wang
College of Computer Science & Technology, Qingdao University
China
wcyorange@163.com

Yu-Hui Wen
School of Computer and Information Technology, Beijing Jiaotong University
China
yhwen1@bjtu.edu.cn

Fuqing Duan
School of Artificial Intelligence, Beijing Normal University
China
fqduan@bnu.edu.cn

Ran Yi
School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University
China
ranyi@sjtu.edu.cn

Yong-Jin Liu*
Department of Computer Science and Technology, BNRist, MOE-Key Laboratory of Pervasive Computing, Tsinghua University
China
liuyongjin@tsinghua.edu.cn

Qingdong Long
College of Computer Science & Technology, Qingdao University
China
2944073120@qq.com

Zhenkuan Pan
College of Computer Science & Technology, Qingdao University
China
zkpan@126.com

Xianfeng Gu
Department of Computer Science, Stony Brook University
United States of America
gu@cs.stonybrook.edu

Abstract

Craniofacial reconstruction is essential in forensic science and has widespread applications. It is challenging due to the detailed facial geometry, complex skull topology, and nonlinear skull-face relationship. We propose a novel approach for 3D craniofacial reconstruction using a Siamese cycle attention mechanism within Generative Adversarial Networks (GAN). Benefiting from the cycle attention mechanism, our method focuses on high-frequency features and morphological connections between the skull and face. Additionally, a Siamese network preserves its identity consistently. Extensive experiments demonstrate superior accuracy and high-quality details of our approach.

CCS Concepts

• **Computing methodologies** → *Shape analysis*; • **Applied computing** → *Evidence collection, storage and analysis.*

Keywords

Craniofacial reconstruction, Cycle Attention GAN, Intrinsic morphological relationship, Siamese network

*Corresponding author.

SA Technical Communications '24, December 03–06, 2024, Tokyo, Japan

ACM ISBN 979-8-4007-1140-4/24/12
https://doi.org/10.1145/3681758.3698016

ACM Reference Format:
Junli Zhao, Chengyuan Wang, Yu-Hui Wen, Fuqing Duan, Ran Yi, Yong-Jin Liu, Qingdong Long, Zhenkuan Pan, and Xianfeng Gu. 2024. Intrinsic Morphological Relationship Guided 3D Craniofacial Reconstruction Using Siamese Cycle Attention GAN. In *SIGGRAPH Asia 2024 Technical Communications (SA Technical Communications '24), December 03–06, 2024, Tokyo, Japan.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3681758.3698016

1 Introduction

Craniofacial reconstruction (CR) aims to estimate a corresponding face based on its skull for identification, which is extensively applied in forensic science, archaeology, and criminal investigation. Due to the hardness and uniqueness of the skull, CR often becomes the last method for identifying a decomposed or mutilated body [Rinchon et al. 2018]. The use of CR technology to make the ancients reappear dates back to the Neolithic Age, which was performed by manually applying clay to the skull. However, traditional manual reconstruction requires highly skilled sculptors with professional knowledge, and the results are prone to the sculptor's subjectivity, making it difficult to reproduce.

In recent years, computer-assisted craniofacial reconstruction has emerged as a flexible, fast, and efficient method [Wen et al. 2020]. Existing linear statistical methods [Madsen et al. 2018] learn the relationships between the skull and face by building a statistical model, but often fail to capture complex nonlinearities, leading to a lack of high-frequency details. Recently, some researchers have used the powerful feature extraction capabilities of deep learning to perform craniofacial reconstruction [Zhang et al. 2022] showing promising results in 3D shape generation. However, challenges

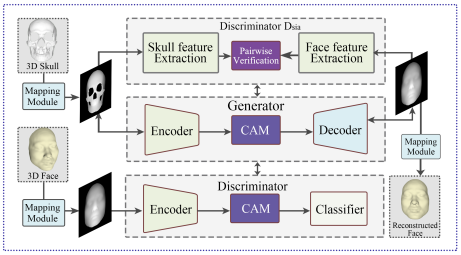

Figure 1: Overview of the proposed approach: (a) Representing 3D craniofacial data as depth maps; (b) Translating skulls to faces with cycle attention GAN and Siamese discriminator; (c) Reconstructing 3D facial shapes from depth maps.

remain in accurately representing complex topologies and preserving morphological connections. In addition, these methods often overlook the intrinsic morphological relationship between the skull and the face, making it difficult to ensure identity consistency.

To address these problems, we construct an end-to-end craniofacial reconstruction network by employing a Siamese cycle attention network to constrain the paired relationships between the skull and the reconstructed face. The main contributions include: (1) We introduce a residual Siamese network to focus on the intrinsic morphological relationship between the skull and the face to ensure the consistency of their identities. (2) We design a novel cycle attention mechanism in a single generator that focuses on craniofacial morphological features and effectively preserves high-frequency details during reconstruction. (3) We propose cosine distance loss and identity preserving loss to constrain the generator, ensuring the reconstructed face closely matches the real face, and experiments show the superiority of our method in accuracy, identity preserving, and high-frequency details.

2 Method

Our approach addresses craniofacial reconstruction by formulating it as a mapping problem and learning the mapping function that translates from skulls to faces. Firstly, we use the projection method to obtain depth maps representing the 3D craniofacial model. Then, a cycle attention module is introduced to the generator to improve the reconstructed details. An additional Siamese network is employed as a discriminator to constrain the identity consistency between the reconstructed face and the target skull. Finally, we convert the generated face depth image into a 3D face.

2.1 Architecture of Our Model

Our model consists of one generator and three discriminators (Figure 1). The generator uses a cycle attention mechanism to enhance reconstruction details. Discriminators D_A and D_B optimize craniofacial generation, while an additional discriminator D_{sia} ensures matching between the reconstructed face and the target skull.

2.1.1 Skull-face Cycle Attention Generator G. Generator G uses cycle consistency constraints and an attention module to learn morphological correlations accurately and focus on high-frequency details. G comprises Z space, an encoder, a Channel Attention Module (CAM), and a decoder.

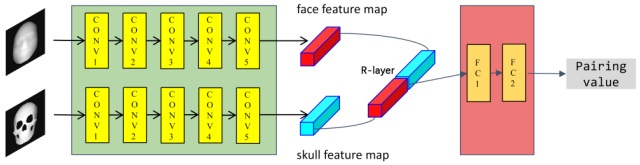

Figure 2: The residual Siamese network D_{sia} takes face and skull data, extracts features, and outputs a matching score.

Encoder and decoder. The Skull-face cycle attention generator G is an encoder-decoder structure. The encoder downsamples input images and enhances features with residual blocks. The decoder includes adaptive residual blocks and upsampling. The former utilizes Adaptive Layer-Instance Normalization (AdaLIN) for normalization, ensuring the converted face image is natural and accurate. Finally, the converted face image is obtained by up-sampling. The model connects local region feature maps to learn the non-linear relationship between the skull and the face.

CAM. We introduce an attention module to make the network focus on the high-frequency details of the face based on local area features. We implement the attention mechanism via CAM and auxiliary classifiers, which perform global discrimination to identify differences between skulls and faces and extract weights of crucial information. The CAM uses these weights as the weights of the corresponding feature maps to extract and enhance high-frequency local features of the skull and face.

2.1.2 Skull-face Siamese Discriminator D_{sia}. We employ a Siamese network as discriminator D_{sia} to determine whether the skull and the face belong to the same person. We feed the face data into one branch of the Siamese network and the skull data into the other branch to comprehensively extract the features of the same pair of skulls and faces and calculate their similarity to judge the consistency of their identity.

As shown in Figure 2, we design a five-layer convolutional network to extract the skull and face features. Each convolution layer is followed by a max-pooling layer, using a convolution kernel of size 3. The number of output channels of each convolution layer is 64, 128, 256, 512, and 512, respectively. After the convolutional layer, a flattened layer is adopted, transforming the face and skull data into two one-dimensional tensors T_f and T_s, respectively. Subsequently, both one-dimensional tensors are passed to the craniofacial residual layer (*R*-layer), where the residual L is calculated using the following formula $L = |T_f - T_s|$. Then, the tensor obtained by the craniofacial residual layer is sent to two fully connected layers. The output of the last fully connected layer is a Sigmoid activation function, resulting in a similarity value that ranges between 0 and 1. Taking 0.5 as the threshold, if the value is less than 0.5, the craniofacial data are considered paired data; otherwise, they are considered unpaired data. Since the pair problem of skull and face belongs to a binary problem, the binary cross-entropy loss function is taken as the loss function of the network.

2.1.3 Skull-face Cycle Discriminator D. The skull-face discriminator D aims to ensure that the generated face images closely resemble real facial features. Here, we set up two discriminators D_A and D_B.

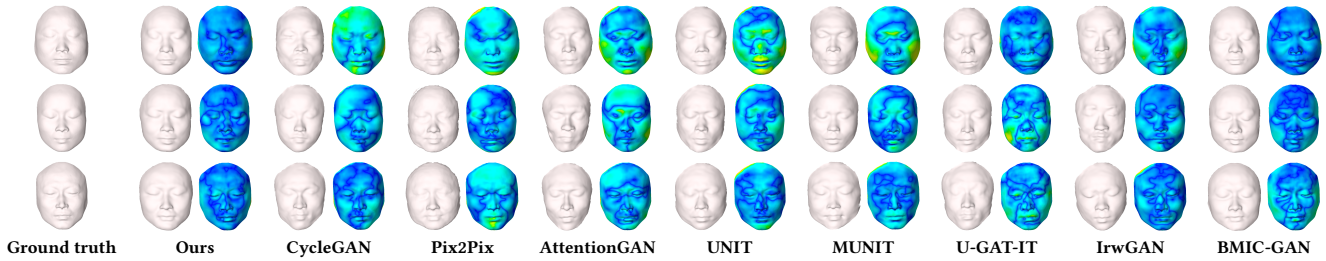

Figure 3: Reconstruction quality comparison of different methods.

Discriminator D_A identifies the mapping from face to skull, and D_B identifies the mapping from skull to face. Both D_A and D_B share an identical structure, comprising a global and a local discriminator. The global discriminator has a receptive field of up to 286, enabling deeper feature compression on the input image (256 x 256). The discriminators D_A and D_B consist of an encoder, a CAM of the Discriminator, and a classifier. Unlike the generator, the discriminator's encoder has only six downsampling layers, and it employs the Leaky-ReLU activation function. The final classifier consists of one convolutional layer with a Sigmoid activation function.

2.2 Loss function

To ensure the identity consistency between the skull and reconstructed face, we introduce the Siamese loss L_{sia}, the identity preserving loss L_{ipl} and the cosine distance loss L_{cos}. We also utilize the cycle consistency loss L_{cyc} [Zhu et al. 2017], the CAM loss L_{cam}, and the adversarial loss L_{adv} [Mirza and Osindero 2014] to ensure the quality and improve the details of the reconstructed face. Therefore, the Generator loss L_G used for optimizing the generator G is a weighted sum of these losses:

$$L_G = w_1 L_{sia} + w_2 L_{adv} + w_3 L_{cyc} + w_4 L_{cam} + w_5 L_{ipl} + w_6 L_{cos}. \quad (1)$$

Siamese loss. Siamese loss measures the matching degree of the reconstructed face and target skull. After D_{sia} extracts features and compares their differences, the final output similarity between skull and face is $v, v \in [0, 1]$. We send the similarity value v as a penalty to the generator G, and the specific formula is as follows:

$$L_{sia} = E_v[-log(v)]. \quad (2)$$

Identity preserving loss. The identity-preserving loss constrains the details of the generated and original samples to be consistent. Given a skull image $x \in X_s$, we assume that the skull data s gets the feature $En(x)$ through the encoder of the generator and reconstructs the corresponding face f. Because of the cycle consistency, the generated face will also get the feature $En(G_{s\to f}(x))$ through the encoder of the generator. We calculate the L1 loss of $En(x)$ and $En(G_{s\to f}(x))$ as the identity preservation loss:

$$L_{ipl} = \| En(x) - En(G_{s\to f}(x)) \|_1 \quad (3)$$

Cosine distance loss. We provide the following label consistency constraint formula (4) to control the generated face to be as close to the target face as possible. Specifically, it penalizes the cosine distance of the labeled points between the generated face image and the real face image:

$$L_{cos}(\theta_G) = E_{G_{s\to f}(x),x}[1 - cos(\varphi(G_{s\to f}(x)), \varphi(G(x)))] \quad (4)$$

Table 1: Reconstruction error statistics of models tested in the ablation study experiment for 60 tests

Method	*OneG*	L_{ipl}	L_{cos}	L_{sia}	**Ours**
Mean	0.007846	0.008376	0.008634	0.007776	**0.007446**
Variance	0.000043	0.000045	0.000051	0.000041	**0.000039**

where $\varphi(\cdot)$ is the feature vector obtained through 68 key feature points of the face.

3 Results

We perform craniofacial reconstruction with a single generator based on a cycle attention module. We train a Siamese network with a test accuracy of 96.24%, to enforce the compatibility between input skulls and their corresponding reconstructed faces.

3.1 Experimental Data

It isn't easy to obtain a large amount of craniofacial data due to its privacy. Our experiments are conducted on 209 pairs of craniofacial data, 60 pairs are randomly selected for testing, and the remaining 149 pairs are for training. We perform data augmentation by rotating each 3D mesh around the X, Y, and Z axes at random angles in the range $(-3°, 3°)$ nine times, followed by projection to obtain 1,341 pairs of depth images for training.

3.2 Ablation Study

We conduct ablation studies focusing on the network structure and loss function, encompassing five key aspects: (1) Cycle attention mechanism controlled by a single generator (*OneG*), which is equivalent to AttentionGAN [Tang et al. 2019] and serves as the baseline of our ablation study; (2) Use identity preserving loss (L_{ipl}); (3) Use the cosine distance loss (L_{cos}); (4) Use the trained Siamese model for matching identification (L_{sia}); (5) Our method uses all four of the above methods. Table 1 summarizes the average error and variance of these five methods on the reconstruction results of 60 test data sets. Experimental results show that both L_{ipl} and *OneG* effectively constrain generators to extract common features of the face and skull, with *OneG* performing better. The L_{cos} penalty is reduced due to the prior registration of craniofacial data, leading to similar distributions of feature vectors from facial key points. The residual Siamese model accurately extracts global features of the skull and face, returning a pairing value, making L_{sia} the most effective method. Therefore, we integrate these four modules into our method for optimal results.

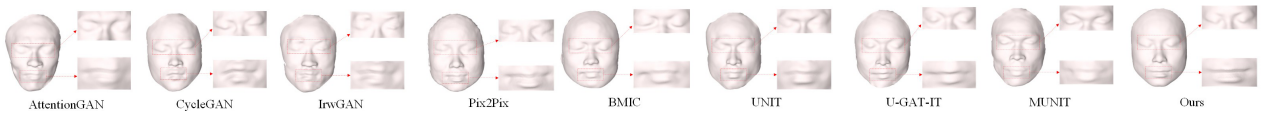

Figure 4: Qualitative comparison of reconstruction details between our method and other GANs mapping methods.

Table 2: Average reconstruction error and variance over 60 reconstructions compared to other end-to-end mappings GANs

Method	Mean	Variance
CycleGAN [Zhu et al. 2017]	0.011369	0.000134
Pix2Pix [Isola et al. 2016]	0.017590	0.000072
AttentionGAN [Tang et al. 2019]	0.013211	0.000105
UNIT [Liu et al. 2017]	0.012791	0.000096
MUNIT [Huang et al. 2018]	0.013254	0.000101
U-GAT-IT [Kim et al. 2019]	0.008816	0.000049
IrwGAN [Xie et al. 2021]	0.010285	0.000093
BMIC-GAN [Zhang et al. 2022]	0.007725	0.000036
Ours	**0.007446**	**0.000039**

Table 3: Reconstruction Error and Variance over 60 Results

Method	**HF-GGR**	**PCA**	**Ours**
Mean	0.017380	0.055044	**0.007446**
Variance	0.000177	0.001091	**0.000039**

3.3 Comparison with Other GANs

In this section, we compare our method with eight end-to-end GANs methods and the reconstruction results comparisons and errors are shown in Figure 3. The reconstruction error statistics of 60 sets of test data are summarized in Table 2. The reconstruction results of our method have the lowest average error compared with other GAN methods.

3.4 Comparison with Statistical Model Methods

In this section, we compare the effectiveness of our method with two classical statistical model methods (PCA [Desvignes et al. 2006], HF-GGR [Jia et al. 2021]) in craniofacial reconstruction. Table 3 show the reconstruction error on test data. Many high-frequency details are lost by these two statistical methods, resulting in a significant reconstruction error with the original face.

3.5 Comparisons on Reconstruction Details

Several representative examples are illustrated in Figure 4, qualitatively showing that the reconstruction results achieved by our method exhibit the highest degree of similarity to the real faces, while other methods have certain defects in the eyes, nose, and mouth reconstruction process.

4 Conclusion

In this paper, we propose a GAN model based on a cycle attention module and Siamese pairwise identification for 3D craniofacial reconstruction. Through the attention module, the network focuses more on the craniofacial structure feature, making the reconstruction of the eyes, nose, and other parts more realistic. A residual Siamese network is trained to facilitate the pairing of the skull and face and establish their nonlinear morphological connection, thereby enhancing the effectiveness of facial reconstruction. Comprehensive experimental results demonstrate that our method achieves superior accuracy and high-quality details in craniofacial reconstruction.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant Nos. 62172247, 62002258, 62302297, 61702293; Beijing Natural Science Foundation L222008, Natural Science Foundation of Shandong Province ZR2024MF087; Young Elite Scientists Sponsorship Program by CAST 2022QNRC001, NSF 2115095, NSF 1762287, NIH 92025 and NIH R01LM012434; The Talent Fund of Beijing Jiaotong University 2023XKRC045 and Shanghai Sailing Program 22YF1420300. We also thank Xianyang Hospital for providing craniofacial data.

References

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.

Xun Huang, Ming-Yu Liu, Serge J. Belongie, and Jan Kautz. 2018. Multimodal Unsupervised Image-to-Image Translation. In *European Conference on Computer Vision*.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2016. Image-to-Image Translation with Conditional Adversarial Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 5967–5976.

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.

Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwanghee Lee. 2019. U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation. (2019).

Ming-Yu Liu, Thomas M. Breuel, and Jan Kautz. 2017. Unsupervised Image-to-Image Translation Networks. In *Neural Information Processing Systems*.

Dennis Madsen, Marcel Lüthi, Andreas Schneider, and Thomas Vetter. 2018. Probabilistic Joint Face-Skull Modelling for Facial Reconstruction. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), 5295–5303.

Mehdi Mirza and Simon Osindero. 2014. Conditional Generative Adversarial Nets. *ArXiv* abs/1411.1784 (2014).

S Rinchon, S Arpita, S Mahipal, and K Rajeev. 2018. 3D forensic facial reconstruction: A review of the traditional sculpting methods and recent computerised developments. *Int J Forens Sci* 3, 1 (2018), 1–8.

Hao Tang, Hong Liu, Dan Xu, Philip H. S. Torr, and N. Sebe. 2019. AttentionGAN: Unpaired Image-to-Image Translation Using Attention-Guided Generative Adversarial Networks. *IEEE Transactions on Neural Networks and Learning Systems* 34 (2019), 1972–1987.

Yang Wen, Zhou Ming-Quan, Lin Pengyue, Geng Guo-hua, Liu Xiaoning, and Li Kang. 2020. Craniofacial Reconstruction Method Based on Region Fusion Strategy. *BioMed Research International* 2020 (2020).

Shaoan Xie, Mingming Gong, Yanwu Xu, and Kun Zhang. 2021. Unaligned Image-to-Image Translation by Learning to Reweight. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), 14154–14164.

Niankai Zhang, Junli Zhao, Fuqing Duan, Zhenkuan Pan, Zhongke Wu, Mingquan Zhou, and Xianfeng Gu. 2022. An End-to-End Conditional Generative Adversarial Network Based on Depth Map for 3D Craniofacial Reconstruction. *Proceedings of the 30th ACM International Conference on Multimedia* (2022).

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), 2242–2251.