



· 文化遗产数字化保护 ·

基于多层次瓶颈注意力模块的颅骨到面皮的生成方法

王 洁, 姜文凯, 蒋佳琪, 梁增磊, 刘晓宁, 耿国华

(西北大学 信息科学与技术学院, 陕西 西安 710127)

摘要 从未知颅骨恢复其生前面貌是考古学、法医学和刑侦学重要的研究方向。现有的计算机三维辅助复原过程繁琐,耗时长,该研究针对现有模型在颅骨到面皮(不含纹理、头发等的面貌)图像生成上存在失真、扭曲、不平滑等现象,提出一种结合生成对抗网络和多层次瓶颈注意力模块的颅骨到面皮图像生成方法。该方法的生成器由6层 AdaRes-Block 和瓶颈注意力模块组成,从通道和空间两个维度引导生成器关注更重要的区域,并根据特征自适应地调整归一化方式。同时,针对生成器模型较大的问题,引入蓝图可分离卷积减小其体积。此外,将判别器分为两部分,前几层被用来进行编码,取消传统网络中的单独编码器模块,使模型更紧凑;后几层则采用多尺度判别策略,从不同层级对图像进行分类判别,增强其准确性。实验结果表明,在颅骨到面皮图像生成任务上,该方法生成的面皮图像质量高于现有的其他方法,在视觉质量和图像质量上都取得了最高的分数,复原效果更加真实,图像定量评价指标 PSNR、SSIM 平均提升 1.115, 0.017, LPIPS 平均降低 0.026, 面皮平均相似度为 0.855。

关键词 颅面生成;生成对抗网络;图像转换;瓶颈注意力模块;蓝图可分离卷积

中图分类号: TP391 **DOI:** 10.16152/j.cnki.xdxzbz.2025-01-017

Skull-to-skin generation method based on multi-level bottleneck attention module

WANG Jie, JIANG Wenkai, JIANG Jiaqi, LIANG Zenglei,
LIU Xiaoning, GENG Guohua

(College of Information Science and Technology, Northwest University, Xi'an 710127, China)

Abstract Restoring the face from an unknown skull is an important research direction in archaeology, forensics and criminal investigation. The existing computer-aided 3D restoration process is cumbersome and time-consuming. In view of the distortion, twisting and non-smoothness of the existing model in the generation of skull-to-skin (face without texture and hair, etc.) images, this paper proposes a skull-to-skin image generation method combining a generative adversarial network and a multi-level bottleneck attention module. Specifically, the generator consists of six layers of AdaResBlock and a bottleneck attention module, which guides the generator to focus on more important areas from the two dimensions of channel and space, and adjusts the nor-

收稿日期:2024-06-08

基金项目:国家自然科学基金(62271393);陕西省重点研发计划(2021GY-028)

第一作者:王洁,女,从事机器学习、颅骨身份认证以及可视化研究。

通信作者:刘晓宁,女,教授,从事机器学习、图像处理研究, xnliu@nwu.edu.cn。

malization method according to the feature adaptiveness. At the same time, in order to solve the problem of the large size of the generator model, the blueprint separable convolution is introduced to reduce its volume. In addition, the discriminator is divided into two parts. The first few layers are used for encoding, eliminating the separate encoder module in the traditional network, making the model more compact; the latter layers adopt a multi-scale discrimination strategy to classify and discriminate images from different levels to enhance their accuracy. Experimental results show that in the task of skull-to-skin image generation, the skin images generated by this method have higher quality than other existing methods, and have achieved the highest scores in both visual quality and image quality. The restoration effect is more realistic, and the image quantitative evaluation indicators PSNR and SSIM are improved by an average of 1.115 and 0.017, and LPIPS is reduced by an average of 0.026. The average facial similarity is 0.855.

Keywords craniofacial generation; generative adversarial network; image translation; bottleneck attention module; blueprint separable convolution

从颅骨恢复死者生前面貌,即颅面复原,是根据人体颅骨特征与面部特征之间的关系来重建生前面貌形象^[1]。在考古学中,其对发掘古代遗物、证实和增补历史信息等具有重要意义,且对于人类骨骼考古学的发展具有显著推动作用。在刑事侦查中,其可以为确定无名尸体的身份提供关键线索。图 1 所示为颅骨和面皮图像,之所以称为面皮而不是面貌,是因为训练数据均采自 CT,重构后,不带皮肤纹理、头发等信息,且为闭眼状态。

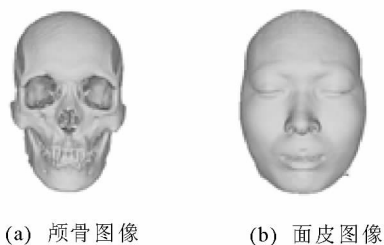


图 1 颅骨面皮图像

Fig. 1 Skull and skin images

传统的手工颅面复原是人类学家基于颅面形态关系并利用颅骨或颅骨的石膏模型,运用可塑材料(如橡皮泥、黏土、塑像蜡等)来还原生前面貌。这类方法往往主观性大,耗时长。随着计算机技术的不断发展,采用计算机三维辅助颅面复原能更高效、客观地实现颅面复原。但是,这种方法仍然面临点云数据量大,特征点标定不准确,不足以描述面皮特征等问题。

近年来,深度学习在图像生成和转换领域取得了显著进展。图像作为信息的一种表现形式,在和生活和工作中的应用愈加广泛。颅骨和面皮图像作为图像中的一种,同样蕴含着丰富的信息,并

且,颅骨和面皮大量有价值的信息集中在面部正视图部分,不需要后脑勺等其他部位的信息,因此可以将颅骨到面皮的复原过程转换为图像到图像的二维转换问题。

由于图像转换旨在学习不同图像域之间的映射,因此如何表示这些映射以生成期望的结果与生成模型显式相关。经典的图像生成模型包括变分自编码器(variational auto-encoder, VAE)^[2]、扩散模型(diffusion model)^[3]和生成对抗网络(generative adversarial networks, GAN)^[4]。其中,基于 VAE 生成的图像较为模糊,缺乏清晰度和细节;扩散模型生成的图像质量更高,但由于其依赖长马尔可夫链的扩散步骤来生成样本,导致计算资源和时间的消耗过高;而 GAN 模型通过对抗训练的方式,生成器网络和判别器网络互相协作,通常能够更快地训练且生成质量较高的图片。因此,本研究选择生成对抗网络作为颅骨到面皮图像的生成框架。

图像转换可以是将图像从 X 域(例如猫)转换到 Y 域(例如狗), X 和 Y 域语义相似,但数据分布不同^[5];也可以看作是将一个图像的风格特征转移到另一个图像上,使目标图像具有与源图像相似的内容和不同的艺术风格,例如将一张照片的风格转换为梵高的绘画风格,即风格迁移^[6];姿态转换任务^[7]也是一种图像转换,将一个人身体姿态从一个角度转换为另外一个角度,具体可分为有监督、无监督、半监督等。只要提供合适的类型和足够的数据量作为源目标映像,就可以用于许多不同类型的任务。现有的有监督框架对于颅骨到面皮的转换任务来说,生成的面皮图像缺乏全局一致性和真实性;而无监督框架则

多通过一对一映射^[8]、权重共享^[9]等方法解决无监督的非可识别性问题。由于颅骨到面皮图像的转换任务具有较高的复杂性,二者之间的关系是非线性的,现有模型生成的面皮图像不够准确和真实,还会出现不同程度的扭曲。

针对以上问题,本研究基于 GAN 设计一种带有多层次瓶颈注意力模块(multi-level bottleneck attention module, MLBAM)的颅面生成模型。该模型从通道和空间两方面学习对关键特征的感知能力,以提高生成器的表达能力和细节保留能力。同时,将判别器分为两部分,前几层被用来进行编码,取消传统网络中的单独编码器模块,缩短隐空间中域之间的转换距离,后几层则采用多尺度判别策略来增强其准确性。此外,为减小生成器模型的体积,采用蓝图可分离卷积(blueprint separable convolutions, BSConv)替代普通卷积。

1 相关工作

1.1 颅面复原

计算机三维辅助颅面复原大致分为以下两类^[10]。

1)利用计算机模拟手工复原法。该方法利用待复原颅骨与面部软组织厚度之间的关系进行颅面复原。1989 年,由 Vanezis 等人提出使用激光扫描仪和摄像机对头骨进行数字化,然后在待复原颅骨的特征点上添加对应的软组织厚度,最后添加面皮特征来实现颅面复原^[11];而人体属性可能会对软组织厚度产生影响,Degreef 等人首次采用多元线性回归法建立了软组织厚度与 BMI、年龄和性别的关系,从实际的角度出发进行颅面复原^[12];Gietzen 等人提出了一种显示的基于面部稠密软组织厚度的颅面复原技术^[13],通过寻找颅骨模型和面皮模型之间的最近点,计算面部稠密软组织厚度并添加到待复原的颅骨模型中,实现颅面复原;Vandermeulen 等人提出了通过颅骨配准将参考模型的面部软组织厚度分布复制到待复原颅骨,还采用了基于 B 样条的自由形变算法提高配准质量^[14]。

目前,颅骨特征点的标定多采用手工方式进行,耗时长,工作量大,不同专家选定特征点的位置和数量也因人而异,而显示的基于面部稠密软组织厚度方法在复原过程中要计算面部软组织厚度,可能会有处理复原结果不平滑和存在孔洞等

问题,使得该技术计算量很大。尽管颅骨配准法速度快,操作简单,但选择的参考模型会直接影响颅面复原的结果。

2)利用机器学习分析颅骨和对应面皮数据之间的形态关系,实现颅面复原。2006 年,Desvignes 等人首次使用机器学习方法实现颅面复原^[15],将其每个颅骨和面皮点云表示为一个向量,采用主成分分析将每个向量表示为平均数据,以及主成分和对应的主成分系数的线性组合。最终,通过优化求解待复原颅骨对应的面貌点云的主成分系数,实现颅面复原。此外,还有学者通过偏最小二乘回归^[16]、支持向量机^[17]、最小二乘正则相关性^[18]、金字塔变换网络(Fast-Net)^[19]等方法构建颅骨和面皮的相关信息,完成颅面复原。2021 年,Lin^[20]等人提出了 CFR-GAN 模型,该模型利用生成对抗网络解决颅面复原问题,克服了传统颅面非线性变形表示方法和缺乏面部纹理表示的缺点,将颅面复原分为粗重建和精重建两步,粗糙重建通过颅骨重建相应人头的整体结构内容,精细重建则恢复面部特征轮廓。Zhang^[21]等人设计了一个基于 CGAN 的端到端神经网络模型,然后用配对的颅面数据训练模型,以自动学习颅骨和面部之间复杂的非线性关系。

以上均是利用机器学习三维辅助颅面复原技术,大多是从颅骨和面貌的三维点云模型中发现两者的形态关系,构建颅面形态关系的数学模型,并利用该模型进行颅面复原。但是,该方法存在点云数据量大,在捕捉和建模复杂关系方面有局限性等缺点,可能无法完全准确地还原面皮的细节特征。

也有部分研究是基于二维的深度学习,例如 Li^[22]等人通过在残差块中引入 Mod-Demo 技术,缩放卷积权值来间接调整激活值,避免重构人脸图像的伪影,并在鉴别器中引入了空间金字塔池(ASPP),以提高重构人脸图像的全局相干性。但是,ASPP 模块是通过多个不同尺度的并行卷积操作来捕捉多尺度特征,这增加了模型的计算复杂度。在处理高分辨率图像或大规模数据集时,ASPP 模块的计算开销可能会变得非常显著。

1.2 图像转换

GAN 以生成器和判别器对抗的方式进行训练,使得生成器能够逐渐学习到生成逼真图像的能力,而判别器则逐渐提高对真实和生成图像的辨别能力。这种对抗学习能力使得 GAN 能够生

成高质量、逼真的图像,满足图像转换任务对真实性和质量的要求。

典型的有监督方法包括 Pix2Pix^[23],其基于条件生成对抗网络^[24],通过输入图像和目标图像的配对进行训练,但由于模型学习两者之间的像素级映射关系,生成的面貌图像不真实、缺乏合理性且分辨率低;尽管 2018 年提出的 Pix2PixHD^[25]可以生成 2 048 × 1 024 分辨率的图像,但还存在训练时间长,不能生成多模态图像的缺点。CycleGAN 引入了循环一致性损失,使得模型能够在无监督的情况下进行训练,并且在没有配对数据的情况下实现两个领域之间的图像转换^[8];UNIT 提出了一个共享潜在空间假设来处理无监督的图像转换^[9];MUNIT 的提出为无监督图像转换生成的图像提供了多样性^[26];DRIT 也是一个多模态的模型,主要贡献是提出了一个内容鉴别器来约束两个不同领域的内容特征^[27];U-GAT-IT 通过利用一个类激活图 CAM 注意力模块和新的归一化函数(AdaLIN)让模型知道在何处进行密集转换,增强模型的鲁棒性^[28];AttentionGAN 通过加

入一种注意力掩模引导模型来区分前景物体并最小化其变化^[29]。

除此之外,计算机视觉的很多分支都可以归类为图像转换。例如,图像修复^[30]、人脸属性编辑 STGAN^[31]、图像去雨 DCD-GAN^[32]等。

2 本研究方法

2.1 总体网络结构

设 X 、 Y 为两个域,分别表示颅骨图像和面皮图像。网络由两个生成器和两个判别器构成,即将图像从 X 域转为 Y 域的生成器 $G_{X \rightarrow Y}$,将图像从 Y 域转回 X 域的 $G_{Y \rightarrow X}$,二者结构相同,都由 n 层 MLBAM 和两个上采样模块构成。判别器 D_X 、 D_Y 分为编码部分 E_X^D 、 E_Y^D 和分类部分 C_X 、 C_Y 。利用 E_X^D 、 E_Y^D 进行编码,分类部分 C_X 、 C_Y 采用多尺度结构来增强模型的表达能力。总体的网络结构如图 2 所示,生成器和判别器具体结构在本文 2.2 和 2.3 中介绍。

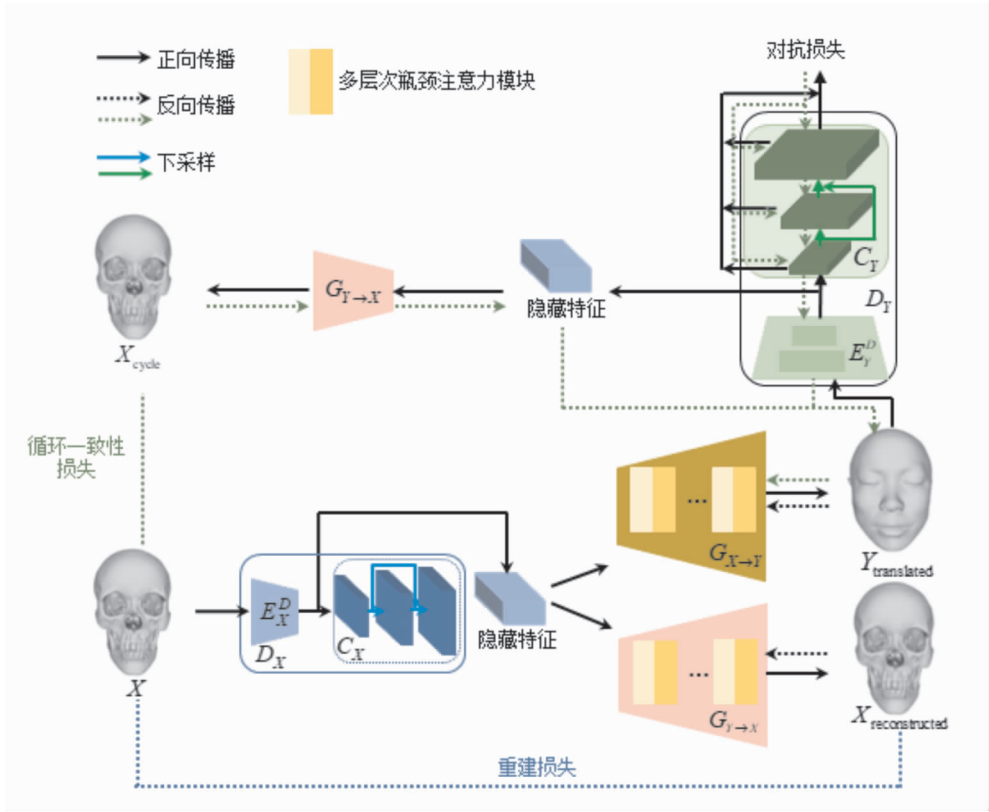


图 2 本研究方法网络结构图

Fig. 2 Themain network structure diagram

2.2 生成器网络架构

本研究设计出一种带有多层次瓶颈注意力模

块(MLBAM)的生成器结构。具体如图 3 所示, AdaResBlock (ResNet + AdaLIN) 作为生成器的

Bottleneck 结合残差连接和 AdaLIN(Adaptive Layer-Instance Normalization)操作。通过堆叠多个块的方式,分层提取和压缩输入特征并将其转换为更低维的表示,同时根据输入数据的统计特征自动调整归一化方式,以便适应不同的数据分布。在 AdaResBlock 之后,引入瓶颈注意力模块,自适应地调整特征在通道和空间两部分的的关系,进一

步提升对关键特征的表示能力。通过多层次的设计,能够更全面地捕捉不同层次的特征信息,提高生成器对于重要特征的感知和生成能力。除此之外,将该生成器的普通卷积替换为蓝图可分离卷积(BSConv),目的是减少模型的参数数量和计算量,使得模型更加轻量级且高效。

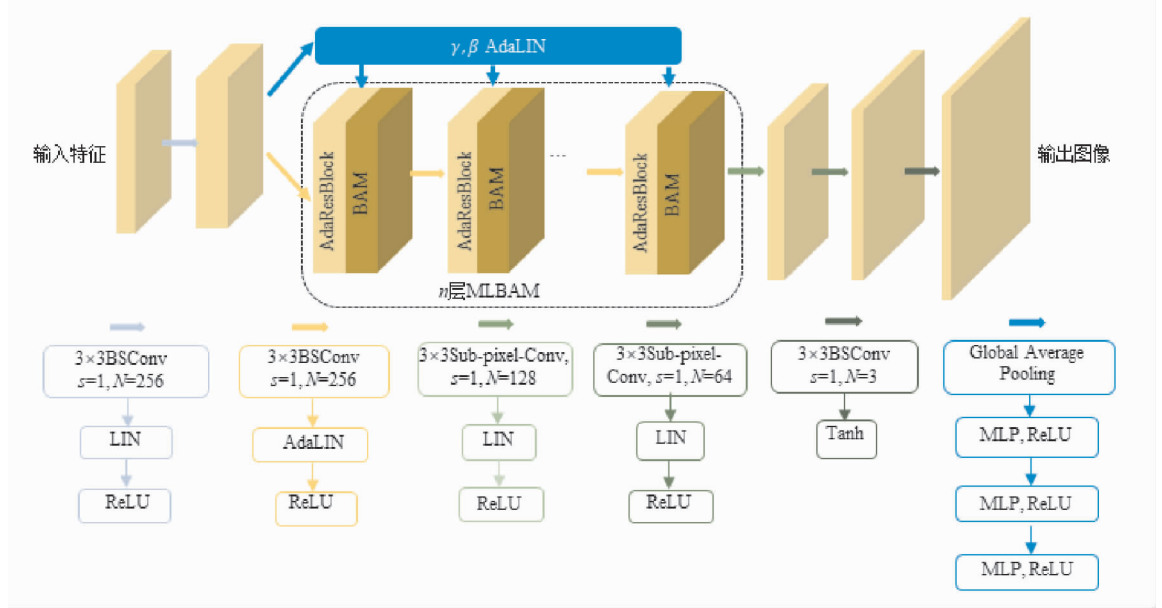


图 3 生成器结构图

Fig. 3 Generator structure diagram

2.2.1 瓶颈注意力模块

仅靠 AdaResBlock 可以提供一定程度的特征提取能力,但无法捕捉图像的细微特征、纹理和色彩细节,从而导致生成的面皮缺乏细节,出现脸颊扭曲等不真实的情况。瓶颈注意力模块(Bottleneck Attention Module, BAM)作为一种有效提高网络表征能力的方式,可放在模型的每个瓶颈处,构建一个具有多个参数的分层注意。因此,本研究在每个 AdaResBlock 后加入一个 BAM,以此来保留图像的细节,提高面皮的生成质量。通过结合 AdaResBlock 的特征提取能力和 BAM 的特征捕捉能力,生成器结构能够更好地处理复杂的图像转换任务,生成更加真实、细致的面皮图像。

瓶颈注意力模块的结构如图 4 所示,特征被输入到通道注意力分支和空间注意力分支^[33],分别得到通道特征注意力权重矩阵 $M_c(F) \in R^{C \times 1 \times 1}$ 和空间特征注意力权重矩阵 $M_s(F) \in R^{1 \times H \times W}$,再扩充到原始尺寸 $M_c(F) \in R^{C \times H \times W}$ 、 $M_s(F) \in R^{C \times H \times W}$,并相加,得到混合的特征注意力矩阵 $M(F)$:

$$M(F) = \sigma(M_c(F) + M_s(F)) \quad (1)$$

式中, σ 为 sigmoid 函数。细化后的特征 F' 由原特征 F 和 $M(F)$ 得到,具体计算为:

$$F' = F + F \otimes M(F) \quad (2)$$

其中通道注意力分支作用是强调网络应该关注什么特征,首先对于输入特征图 $F \in R^{C \times H \times W}$ 进行全局平均池化操作得到一个通道向量 $F_c \in R^{C \times 1 \times 1}$,进而经过带有一个隐藏层的多层感知器(MLP)和一个 BN(Batch Normalization)层。MLP 中的隐藏层激活大小为 $R^{C/r \times 1 \times 1}$, r 为衰减率,具体计算过程为

$$M_c(F) = \text{BN}(\text{MLP}(\text{AvgPool}(F))) = \text{BN}(W_1(W_0 \text{AvgPool}(F) + b_0) + b_1) \quad (3)$$

其中, $W_0 \in R^{C/r \times C}$, $b_0 \in R^{C/r}$, $W_1 \in R^{C \times C/r}$, $b_1 \in R^C$ 。

空间注意力分支目的是强调特征向量在不同空间位置的价值信息,对颅骨到面皮的二维图像转换来说就是对眼眶、嘴巴、鼻子等有价值信息的强调。将特征 $F \in R^{C \times H \times W}$ 用 1×1 卷积降维到

$R^{C/r \times H \times W}$, 然后通过两个 3×3 卷积(Padding = 1)有效利用上下文信息, 进而经过 1×1 卷积得到 $R^{1 \times H \times W}$ 的空间注意力图, 在最后应用 BN 调整图像尺度, 具体公式为

$$M_s(F) = \text{BN}(f_3^{1 \times 1}(f_2^{3 \times 3}(f_1^{3 \times 3}(f_0^{1 \times 1}(F))))))$$

(4)

其中, f 为卷积操作, 上角标为卷积大小。

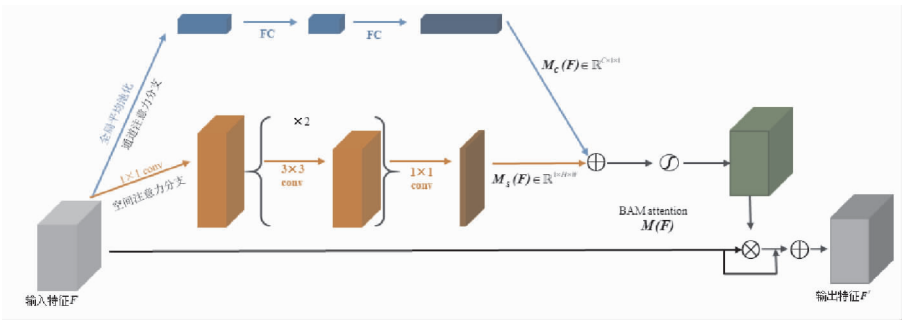


图 4 瓶颈注意力模块结构图
Fig. 4 Bottleneck attention module structure

2.2.2 蓝图可分离卷积

在训练时如果模型参数过大可能导致资源浪费或网络难以收敛的问题。蓝图可分离卷积(Blueprint Separable Convolutions, BSConv)是一种利用核内相关性的轻量级卷积, 用其替换普通卷积可以更好地捕捉卷积核内部的相关性, 灵活地控制生成图像的细节, 降低生成器的复杂度。

如图 5 所示, 标准卷积的卷积核尺寸为 $M \times S \times S$, 可训练参数为 $M \times K \times S^2$ 。蓝图可分离卷积将标准卷积核尺寸由 $M \times S \times S$ 分解为 M 个 $S \times S$ 尺寸的卷积核, 可训练参数仅有 $S^2 \times K + M \times K$, 大大减少计算量, 加快网络训练。与深度可分离卷积(DepthwiseSeparable Convolution, DSConv)相比, 蓝图可分离卷积可以看成是对其的逆运算, 即

先进行逐点卷积(Pointwise Convolution), 再进行逐通道卷积(Depthwise Convolution)。这样做是根据每个卷积核的权重沿深度轴高度相关, 呈现出相同的视觉结构, 比如有一个卷积核模板, 其深度轴上的每个卷积核都是在模板上按不同的因素缩放得到, 而这个二维模板就称为“蓝图”^[34]。DSConv 更多的是依赖于跨内核相关性, 隐式地将一个三维蓝图(在深度轴将二维蓝图组合在一起)应用于所有卷积核, 在卷积过程中内核之间的相关性占主导地位, 而 BSConv 应用二维蓝图在滤波器内部分布权重, 可以更好地捕捉滤波器内部相关性, 提升卷积分离的效果, 进而实现对图像细节的精细控制。

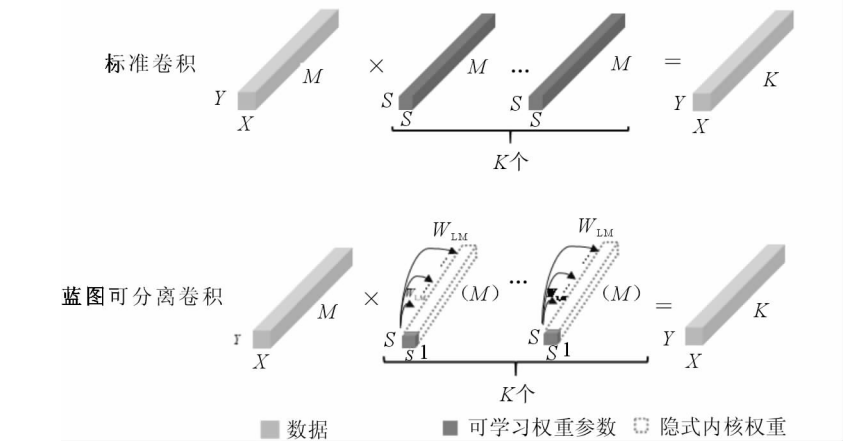


图 5 BSConv 卷积原理
Fig. 5 BSConv convolution principle

2.3 判别器网络结构

目前, 基于 GAN 的图像转换方法多采用由编码器、生成器和判别器等多个组件构成的框架

[见图 6(a)]。这种框架在一些任务上取得了良好的效果, 但其结构可能相对复杂, 在隐空间中特征的转换距离较大, 生成的面皮轮廓不平滑, 五官

相似度不高,在颅面转换这种复杂的任务上效果较差。在生成对抗网络中,判别器用于区分真实图像和生成图像,因此在图像传入判别器时,判别器的第一作用就是对图像进行编码,在 DC-GAN^[35]的研究中,也证实了判别器的编码功能。因此,本研究利用判别器的前几层对图像进行编码,后几层进行分类判别[见图 6(b)]。 D_X 、 D_Y 结构相同,这里用 D_Y 介绍,如之前所述判别器分为 E_Y^D 和 C_Y ,其中 C_Y 分为 3 个子分类器 C_Y^0 、 C_Y^1 和 C_Y^2 ,分别用于小尺度、中尺度和大尺度的判别。 E_Y^D 的输出连接子分类器 C_Y^0 ,在 E_Y^D 上进行下采样操作得到较小尺度特征图;进而分为两路,一路直

接连接 C_Y^1 ,一路再进行下采样连接到 C_Y^2 ,即将单个输入图像通过不同层级的处理,得到不同尺度的特征表示。这些特征表示看作是具有不同细节和语义级别的图像。然后,将这些特征表示按照它们对应的输入尺寸送入分类器,以进行更准确的判别。这样做不仅使整个网络更简单,而且与传统的 GAN 在训练后放弃判别器的训练方式不同,判别器的前几层仍然保留用于编码,直接通过判别器的损失进行训练而不是利用生成器的反向传播间接训练,生成的面皮图像能更好地捕获外观、纹理、形状等特征,使得生成的面皮更加接近真实面皮,具有更高的逼真度和视觉质量。

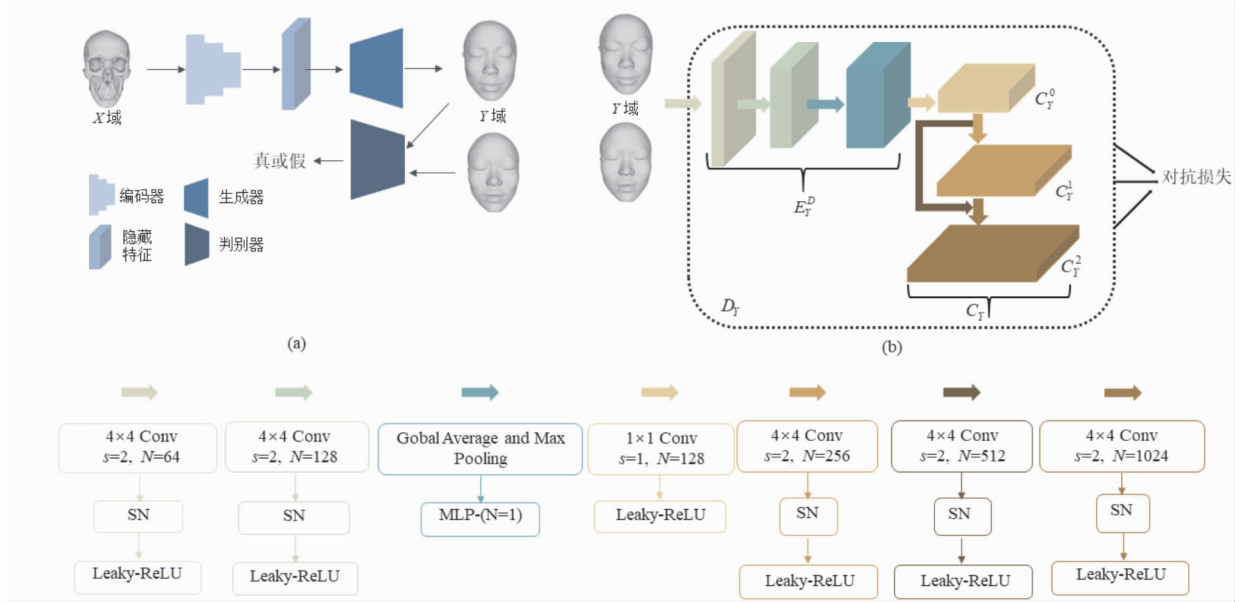


图 6 判别器结构

Fig. 6 Discriminator structure

2.4 损失函数

本研究损失函数由对抗损失、循环一致性损失和重建损失 3 部分组成,对抗损失是为了将图像在 X 域和 Y 域更好地转换,循环一致性损失和重建损失是为了解决上述提到过的不可识别性问题。

2.4.1 对抗损失

生成器和判别器的损失都采用最小二乘对抗损失^[36],可以改善梯度消失问题,避免模式崩溃,使训练更稳定,公式为:

$$\min_{G_{X \rightarrow Y}} \max_{D_Y} L_{\text{gan}}^{X \rightarrow Y} = E_{y \sim Y} [(D_Y(y))^2] + E_{x \sim X} [\log(1 - D_Y(G_{X \rightarrow Y}(E_X(x))))^2]. \quad (5)$$

通过 GAN 训练 min-max 时,编码器与生成器

一起进行最小化训练,而与判别器进行最大化训练。此时,由于现在编码器 E_X 不仅作为判别器的一部分,它的输出还被作为生成器的输入,所以最大化 $L_{\text{gan}}^{X \rightarrow Y}$ 时, E_X 是固定的, E_Y 是训练的,当最小化 $L_{\text{gan}}^{X \rightarrow Y}$ 时, E_X 、 E_Y 都固定。

2.4.2 循环一致性损失

循环一致性损失^[8]是为了让图像经过两个生成器后尽可能和原图保持一致,对于本研究来讲,就是颅骨经过 $G_{X \rightarrow Y}$ 生成面皮图像后,再经过 $G_{Y \rightarrow X}$ 逆向转换到颅骨图像,使得原颅骨图像和逆向转换后的颅骨图像尽量保持一致。其公式为:

$$\min_{G_{X \rightarrow Y}, G_{Y \rightarrow X}} L_{\text{cycle}}^{X \rightarrow Y} = E_{x \sim X} [|x - G_{Y \rightarrow X}(E_Y(G_{X \rightarrow Y}(E_X(x))))|_1] \quad (6)$$

其中, $\|\cdot\|_1$ 是 L1-norm 的值,此时 E_X 、 E_Y 都固定。

2. 4. 3 重建损失

该损失基于共享潜在空间假设,旨在通过将源域真实样本的隐藏向量输入源域生成器 $G_{Y \rightarrow X}$,并将其重构后的输出与原始输入样本接近恒等映射^[37]。这样做是为了确保生成器的输出图像能够保持与输入图像的相似性,提高图像生成的质量和准确性。具体公式为:

$$\min_{G_{Y \rightarrow X}} L_{\text{recon}}^{Y \rightarrow X} = E_{x \sim X} [\|x - G_{Y \rightarrow X}(E_X(x))\|_1]$$

(7)

其中 E_X 固定, Y 域到 X 域的损失函数与 X 域到 Y 域相同。

2. 4. 4 总损失

判别器总损失为:

$$\max_{E_X, C_X, E_Y, C_Y} \lambda_1 L_{\text{gan}}$$

(8)

生成器总损失为:

$$\min_{G_{X \rightarrow Y}, G_{Y \rightarrow X}} \lambda_1 L_{\text{gan}} + \lambda_2 L_{\text{cycle}} + \lambda_3 L_{\text{recon}}$$

(9)

其中, $L_{\text{gan}} = L_{\text{gan}}^{X \rightarrow Y} + L_{\text{gan}}^{Y \rightarrow X}$, $L_{\text{cycle}} = L_{\text{cycle}}^{X \rightarrow Y} + L_{\text{cycle}}^{Y \rightarrow X}$, $L_{\text{recon}} = L_{\text{recon}}^{X \rightarrow Y} + L_{\text{recon}}^{Y \rightarrow X}$, $\lambda_1 = 1, \lambda_2 = \lambda_3 = 10$ 。

3 实验结果与分析

3. 1 数据预处理

本实验使用的数据来自陕西中医药大学附属医院的志愿者头部 CT 扫描。利用 Marching Cube 算法对 CT 数据的颅骨和面皮图像进行三维重建,把得到的颅骨和面皮的重建模型在法兰克福坐标系中垂直映射到 xoz 平面上,得到颅骨和面皮的二维正视图,并对其进行增强(旋转 90° , 180° , 270° 和镜像翻转)。

3. 2 实验参数

实验所用的颅骨和面皮数据训练集 600 对,测试集 144 对,验证集 4 对,在进行训练时输入图片统一大小为 256×256 ,基于 python 3. 6. 3 和 Pytorch 1. 10. 2,使用 NVIDIA GeForce RTX 4080 显卡训练,利用 Adam 优化器,学习率为 0. 000 1, ($\beta_1 = 0. 5, \beta_2 = 0. 999$),所有实验的 batch size 设置为 2。

3. 3 实验结果定性评估

由于本研究是基于 GAN 的计算机二维图像转换进行从颅骨生成面皮图像,所以与几种先进的图像转换网络进行比较,包含 Pix2Pix^[23]、CycleGAN^[8]、U-GAT-IT^[28] 和 AttentionGAN^[29]。4 种

方法的生成器、判别器训练时间段数和优化策略都采用官方默认设置,实验结果如图 7 所示。具体来说,Pix2Pix 生成的图像与目标面皮基本轮廓不一致,第 4 行第 3 列最为明显,嘴巴和眼睛部分也发生变形;CycleGAN 可以生成与目标面皮相近的鼻子,但其他五官发生扭曲;U-GAT-IT 生成的图像轮廓虽与目标面皮大致相同,但是捕捉五官细节的能力还是不够,与真实面皮差距较大;AttentionGAN 生成的 4 副图像五官基本一致,区别甚小,而且脸部均出现“肿胀”现象。

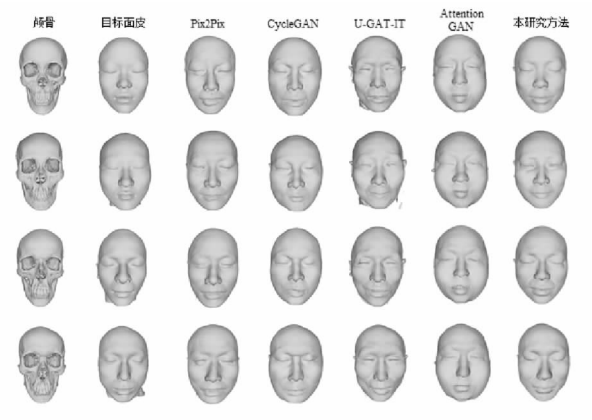


图 7 与 Pix2Pix、CycleGAN、U-GAT-IT、AttentionGAN 实验对比结果

Fig. 7 Experimental comparison results with Pix2Pix, CycleGAN, U-GAT-IT, and AttentionGAN

综上所述,对比方法生成的图像与目标面皮相差较大,不能作为识别身份的特征。本方法生成的面皮图像,轮廓和细节方面都与真实图像最接近,具有较高的准确性。

3. 4 实验结果定量评估

评价生成图像质量的高低不仅取决于视觉感受,而且需要一些客观的评价指标对结果做出定量评价。因此,本研究用 PSNR、SSIM 和 LPIPS 作为评估指标,对比网络为 Pix2Pix、CycleGAN、U-GAT-IT 和 AttentionGAN。实验结果如表 1 所示。

1) 峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR): PSNR 是基于图像的均方误差而计算得到,数值越高表示图像质量损失越小,是一种最普遍的评价方法。但是,其无法完全反应人眼对图像细节和结构的感知。

2) 平均结构相似性 (Mean Structural Similarity Index Measure, SSIM): SSIM 将图像比较为人眼感知的 3 个重要因素,计算两幅图像之间的结构相似性。其值越高,图像质量损失越小。

3)学习感知图像块相似度(Learned Perceptual Image Patch Similarity, LPIPS):LPIPS 是基于图像的 patch 级别,旨在量化图像之间的感知差异,其值越低,图像质量差异越小。

表 1 Pix2Pix、CycleGAN、U-GAT-IT、AttentionGAN 和本方法定量对比结果

Tab. 1 Quantitative comparison results of Pix2Pix, CycleGAN, U-GAT-IT, AttentionGAN and this method

方法	PSNR/dB ↑	SSIM ↑	LPIPS ↓
Pix2Pix ^[23]	18.595	0.794	0.185
CycleGAN ^[8]	19.276	0.816	0.179
U-GAT-IT ^[28]	19.004	0.801	0.197
AttentionGAN ^[29]	18.296	0.800	0.225
本方法	19.908	0.820	0.170

实验表明,本方法在颅面转换任务上不仅视觉效果优秀,而且在 PSNR、SSIM 和 LPIPS 三个指标上得分也最好。这表明,本方法不仅生成的图像质量高,且与真实图像之间的感知差异也最小,最接近真实面皮图像。

3.5 面皮相似度对比

表 2 为所用几种对比方法生成的面皮图像和本方法生成的面皮图像与真实面皮图像相似度对比,表 3 为本研究生成的面皮图像和真实面皮图像交叉对比的结果。其借助腾讯 AI Lab^[38]提供的人脸相似度计算功能,对比表 2 和表 3 相似度,结果可以看出,本研究生成的面皮图像与其对应的真实面皮图像相似度最高,说明本方法生成的面皮更接近真实面皮。

表 2 不同方法生成面皮图像与真实面皮图像相似度对比

Tab. 2 Comparison of the similarity between different methods of generating skin images and real skin images

面皮	Pix2-Pix	Cycle-GAN	U-GAT-IT	Attention GAN	本文方法
面皮 1	0.51	0.60	0.52	0.49	0.86
面皮 2	0.64	0.73	0.67	0.63	0.74
面皮 3	0.55	0.71	0.82	0.42	0.90
面皮 4	0.80	0.66	0.84	0.61	0.92

表 3 本文方法生成面皮图像与真实面皮图像相似度交叉对比

Tab. 3 Cross comparison of the similarity between the skin image generated by the method in this paper and the real skin image

真实	生成			
	面皮 1	面皮 2	面皮 3	面皮 4
面皮 1	0.86	0.41	0.56	0.47
面皮 2	0.51	0.74	0.56	0.80
面皮 3	0.43	0.54	0.90	0.48
面皮 4	0.53	0.55	0.64	0.92

3.6 消融实验

为了证明判别器编码(E_D)、瓶颈注意力模块(BAM)以及蓝图可分离卷积(BSConv)的有效性,对其进行消融实验(见图 8)。

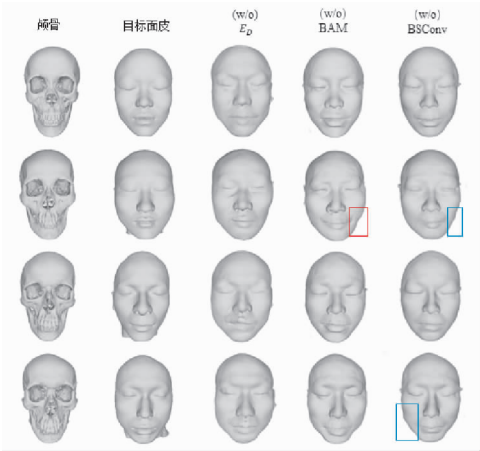


图 8 消融实验
Fig. 8 Ablation study

其中, (w/o) E_D 表示移除判别器内部的编码器组件,采用传统的独立编码器模块。通过对比可以发现,如此生成的面部轮廓显得不够流畅,甚至某些面部特征呈现出扭曲的状态。由此可知,将编码器嵌入判别器结构中的独特设计,可以缩短特征在隐空间中域之间的转换路径,从而提高生成面皮的质量。

(w/o)BAM 去除了瓶颈注意力模块,在此情况下生成的面皮图像中,嘴巴部位呈现出较高的相似性,甚至在上唇区域隐约可见牙齿的轮廓。但是,其生成的图像细节不够,且缺少一块面部结构,如红色框所圈部分,由此可证明本方法添加 BAM 模块的有效性。

(w/o)BSConv 表示去掉蓝图可分离卷积,采用普通的卷积方法。观察结果显示,仅有一张位

于最下方的面皮图像与真实图像较为接近,但其轮廓线条不够平滑,出现了“抖动”现象,如蓝色框标注所示。这一对比充分说明了 BSConv 在减轻模型参数负担的同时,能够更有效地利用卷积核内部的相关性,优化卷积分离效果,进而实现对图像细节的精准把控。

图 9 展示了 MLBAM 层数作为超参数的实验结果,从左至右依次为模型拥有 4 层、5 层、6 层、7 层的 MLBAM。由图 9 可以看出,层数过少,模型没有足够的能力捕捉到颅骨到面皮之间复杂的非线性关系,导致模糊、失真、不平滑等问题;但层数过多时,模型的复杂度会因太复杂而导致过拟合。同时,过多的层数意味着瓶颈注意力模块的增多,避免不了“过度关注”图像中的区域,导致生成的面皮质量下降。实验表明,具有 6 层的 MLBAM 效果最好。



图 9 MLBAM 层数超参数实验

Fig. 9 MLBAM layer number hyperparameter experiment

表 4 为消融实验的定量评价结果。由表 4 第 3 行可知,在仅移除 BSConv (蓝图可分离卷积)后,所有 3 个指标均表现最优,而本研究所采用的包含 6 层瓶颈注意力模块(6-MLBAM)的完整模型紧随其后,位列第二。这一结果并不直接表明 BSConv 组件是无效的。

实际上,颅骨到面皮图像生成任务的核心目标是实现视觉感官上的优化。本研究提出的方法,结合了 E_D 、BAM、BSConv,在视觉效果方面展现出了最佳表现。这一点可以从图 9 中 6-MLBAM 的生成图像得到直观验证,同时图 8 也详细对比了仅移除 BSConv 组件后的视觉表现结果。因此,尽管移除 BSConv 后的模型在某些定量指

标上略有优势(与 6-MLBAM 的平均指标差异不超过 0.06),但整体而言,带有 6 层 MLBAM 的完整方法仍然表现最为出色。这充分说明了每个模块在模型中的重要作用,它们相互协作,共同推动了模型性能的提升。

表 4 消融实验

Tab. 4 Ablation study

方法	PSNR/dB $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
(w/o) E_D	19.232	0.804	0.252
(w/o) MLBAM	19.866	0.813	0.176
(w/o) BSConv	20.078	0.822	0.170
4-MLBAM	19.518	0.812	0.196
5-MLBAM	18.497	0.781	0.217
6-MLBAM	19.908	0.820	0.170
7-MLBAM	18.256	0.774	0.222

表 5 分别为加入 E_D 、BAM、BSConv 后生成器和判别器的模型参数总量。由表 5 可见,加入 BSConv 后有效地降低了生成器的大小,使得生成器与判别器大小基本相匹配,可以更好地平衡它们之间的学习能力。因为生成器的参数过多,而判别器的参数较少,生成器可能会过于强大,导致模型难以收敛。相反,如果判别器的参数过多,而生成器的参数较少,判别器可能会过于敏感,难以提供准确的反馈信号。因此,参数量的平衡可以帮助生成器和判别器之间保持相对稳定的竞争,促进模型的学习和生成结果的质量。

表 5 网络模块的参数总数

Tab. 5 Total number of parameters for network modules

方法	生成器	判别器
baseline + E_D	68.074G	12.000G
+ E_D + BAM	68.512G	12.000G
+ E_D + MLBAM +	11.012G	12.000G
BSConv		

4 结语

本研究针对现有基于 GAN 的图像转换模型在颅面转换任务上存在捕捉五官细节不细致,轮廓不流畅导致扭曲、变形、失真等情况,提出基于多层次瓶颈注意力模块的颅面转换网络,生成器包含 6 层 AdaResBlock 和 BAM,从不同的角度引导生成器捕捉重要特征,且自适应地调整特征的归一化方式,并引入蓝图可分离卷积减小模型体积。判别器前几层用于编码代替传统的单独编码器结构,促进高维图像空间域的转换。实验表明,

本方法生成的面皮图像均优于其他网络方法,生成结果与真实面皮最接近,可为颅骨身份认证提供一定线索。

本研究的实验数据均采自 CT 重构,当颅骨采集渠道不同时,生成的结果仍有待提升。因此,下一步的研究应致力于来自不同域的颅骨图像的面皮图像生成。

参考文献

- [1] 王琳,赵俊莉,段福庆,等. 颅面复原方法综述[J]. 计算机工程, 2019, 45(12):8-18.
WANG L, ZHAO J L, DUAN F Q, et al. Survey on craniofacial reconstruction method[J]. Computer Engineering, 2019, 45(12):8-18.
- [2] REZENDE D J, MOHAMED S, WIERSTRA D. Stochastic backpropagation and approximate inference in deep generative models[C]//International conference on machine learning(ICML). Beijing: PMLR, 2014: 1278-1286.
- [3] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[J]. Advances in Neural Information Processing Systems, 2020, 33: 6840-6851.
- [4] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [5] XIE S, XU Y, GONG M, et al. Unpaired image-to-image translation with shortest path regularization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Los Angeles: IEEE, 2023: 10177-10187.
- [6] BUI N T, NGUYEN H D, BUI-HUYNH T N, et al. Efficient loss functions for GAN-based style transfer[C]//Fifteenth International Conference on Machine Vision(ICMV). Yerevan: SPIE, 2023: 373-380.
- [7] REN Y, YU X, CHEN J, et al. Deep image spatial transformation for person image generation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 7690-7699.
- [8] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision (ICCV). Venice: IEEE, 2017: 2223-2232.
- [9] LIU M Y, BREUEL T, KAUTZ J. Unsupervised image-to-image translation networks[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS). Long Beach: ACM, 2017: 700-708.
- [10] 税午阳,邓擎琼,吴秀杰,等. 颅面复原技术在体质人类学中的应用[J]. 人类学学报, 2021, 40(4): 706-720.
SHUI W Y, DENG Q Q, WU X J, et al. An overview of craniofacial reconstruction technology application in physical anthropology[J]. Acta Anthropologica Sinic, 2021, 40(4):706-720.
- [11] VANEZIS P, BLOWES R W, LINNEY A D, et al. Application of 3-D computer graphics for facial reconstruction and comparison with sculpting techniques[J]. Forensic Science International, 1989, 42(1/2): 69-84.
- [12] DEGREEF S, VANDERMEULEN D, CLAES P, et al. The influence of sex, age and body mass index on facial soft tissue depths[J]. Forensic Science, Medicine, and Pathology, 2009, 5: 60-65.
- [13] GIETZEN T, BRYLKA R, ACHENBACH J, et al. A method for automatic forensic facial reconstruction based on dense statistics of soft tissue thickness[J]. PLoS One, 2019, 14(1): 1-19.
- [14] VANDERMEULEN D, CLAES P, LOECKX D, et al. Computerized craniofacial reconstruction using CT-derived implicit surface representations[J]. Forensic Science International, 2006, 159: 164-174.
- [15] DESVIGNES M, BAILLY G, PAYAN Y, et al. 3D semi-landmarks based statistical face reconstruction[J]. Journal of Computing and Information Technology, 2006, 14(1): 31-43.
- [16] DUAN F, HUANG D, TIAN Y, et al. 3D face reconstruction from skull by regression modeling in shape parameter spaces[J]. Neurocomputing, 2015, 151: 674-682.
- [17] SHUI W, ZHOU M, MADDOCK S, et al. A PCA-Based method for determining craniofacial relationship and sexual dimorphism of facial shapes[J]. Computers in Biology and Medicine, 2017, 90: 33-49.
- [18] 周明全,杨稳,林芃樾,等. 基于最小二乘正则相关性分析的颅骨身份识别[J]. 光学精密工程, 2021, 29(1):201-210.
ZHOU M Q, YANG W, LIN P Y, et al. Skull identification based on least square canonical correlation analysis[J]. Optics and Precision Engineering, 2021, 29(1):201-210.
- [19] ZHAO L, MA L, CUI Z, et al. FAST-Net: A coarse-to-fine pyramid network for face-skull transformation[C]//International Workshop on Machine Learning in

- Medical Imaging (MLMI). Cham: Springer Nature Switzerland, 2023: 104-113.
- [20] LIN P Y, YANG W, XIA S Y, et al. CFR-GAN: A generative model for craniofacial reconstruction [C] // 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Houston: IEEE, 2021: 462-469.
- [21] ZHANG N, ZHAO J, DUAN F, et al. An end-to-end conditional generative adversarial network based on depth map for 3D craniofacial reconstruction [C] // Proceedings of the 30th ACM International Conference on Multimedia (ACM, MM). Lisboa Portugal: ACM, 2022: 759-768.
- [22] LI Y, WANG J, LIANG W, et al. CR-GAN: Automatic craniofacial reconstruction for personal identification [J]. Pattern Recognition, 2022, 124: 108400.
- [23] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Hawaii: IEEE, 2017: 1125-1134.
- [24] MIRZA M, OSINDERO S. Conditional generative adversarial nets [EB/OL]. (2014-11-06) [2022-10-23]. <https://arxiv.org/abs/1411.1784>.
- [25] WANG T C, LIU M Y, ZHU J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake: IEEE, 2018: 8798-8807.
- [26] HUANG X, LIU M Y, BELONGIE S, et al. Multimodal unsupervised image-to-image translation [C] // Proceedings of the European conference on computer vision (ECCV). Munich: Springer, 2018: 172-189.
- [27] LEE H Y, TSENG H Y, HUANG J B, et al. Diverse image-to-image translation via disentangled representations [C] // Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 35-51.
- [28] KIM J, KIM M, KANG H, et al. U-GAT-IT: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation [C] // International Conference on Learning Representations. New Orleans: Ithaca, 2019: 1-19.
- [29] TANG H, LIU H, XU D, et al. Attentiongan: Unpaired image-to-image translation using attention-guided generative adversarial networks [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 34(4): 1972-1987.
- [30] 陈晓雷, 杨佳, 梁其铎. 结合语义先验和深度注意力残差的图像修复 [J]. 计算机科学与探索, 2023, 17(10): 2450-2461.
- CHEN X L, YANG J, LIANG Q D. Image inpainting combining semantic priors and deep attention residuals [J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(10): 2450-2461.
- [31] LIU M, DING Y, XIA M, et al. Stgan: A unified selective transfer network for arbitrary image attribute editing [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 3673-3682.
- [32] CHEN X, PAN J, JIANG K, et al. Unpaired deep image deraining using dual contrastive learning [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022: 2017-2026.
- [33] PARK J, WOO S, LEE J Y, et al. Bam: Bottleneck attention module [EB/OL]. (2018-07-17) [2022-11-10]. <https://arxiv.org/abs/1807.06514>.
- [34] HAASE D, AMTHOR M. Rethinking depthwise separable convolutions: How intra-kernel correlations lead to improved mobilenets [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Virtual: IEEE, 2020: 14600-14609.
- [35] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [EB/OL]. (2015-11-19) [2023-03-15]. <https://arxiv.org/abs/1511.06434>.
- [36] MAO X, LI Q, XIE H, et al. Least squares generative adversarial networks [C] // Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 2794-2802.
- [37] CHEN R, HUANG W, HUANG B, et al. Reusing discriminators for encoding: Towards unsupervised image-to-image translation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Virtual: IEEE, 2020: 8168-8177.
- [38] YANG X, JIA X, GONG D, et al. LARNeXt: End-to-end lie algebra residual network for face recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(10): 11961-11976.