

# 基于条件对抗自动编码器的跨年龄人脸合成

程志康<sup>1,2</sup>, 孙锐<sup>1,2</sup>, 孙琦景<sup>1,2</sup>, 张旭东<sup>1,2</sup>

(1. 合肥工业大学 计算机与信息学院, 合肥 230009; 2. 工业安全与应急技术安徽省重点实验室, 合肥 230009)

**摘要:** 跨年龄人脸合成是指通过已知特定年龄的人脸图像合成其他年龄段的人脸图像, 在动漫娱乐、公共安全、刑事侦查等领域有广泛的应用。针对跨年龄人脸合成图像容易产生器官变形扭曲、人脸局部特征保持效果不佳等问题, 提出一种基于条件对抗自动编码器的合成方法。通过在解码器结构中引入通道关注和空间关注模块, 分别从通道域和空间域提取重要信息, 使模型在训练过程中忽略背景等无关信息, 聚焦人脸图像变化的区域, 有效解决合成图像器官扭曲变形等问题。此外, 设计一种多尺度特征损失网络, 从多个尺度更深层次地约束人脸图像的局部结构特征, 从而保持人脸合成过程中局部特征结构的稳定性。在UTKFace跨年龄人脸数据集上的实验结果表明, 与CAAE方法相比, 该方法有效避免了人脸器官变形扭曲问题, 能够更好地保持人脸局部结构特征, 具有较佳的人脸合成效果和细节保持能力。

**关键词:** 跨年龄人脸合成; 条件对抗自动编码器; 通道关注模块; 空间关注模块; 多尺度特征损失网络

开放科学(资源服务)标志码(OSID):



**中文引用格式:** 程志康, 孙锐, 孙琦景, 等. 基于条件对抗自动编码器的跨年龄人脸合成[J]. 计算机工程, 2022, 48(6): 304-313.

**英文引用格式:** CHENG Z K, SUN R, SUN Q J, et al. Cross-age face synthesis based on conditional adversarial autoencoder[J]. Computer Engineering, 2022, 48(6): 304-313.

## Cross-age Face Synthesis Based on Conditional Adversarial Autoencoder

CHENG Zhikang<sup>1,2</sup>, SUN Rui<sup>1,2</sup>, SUN Qijing<sup>1,2</sup>, ZHANG Xudong<sup>1,2</sup>

(1. School of Computer and Information, Hefei University of Technology, Hefei 230009, China;

2. Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei 230009, China)

**[Abstract]** Cross-age face synthesis involves synthesizing facial images of other age groups from facial images of known specific age groups. It has a wide range of applications in the fields of animation entertainment, public safety, criminal investigation, and so on. To solve the problems of organ distortion and poor local feature preservation in cross-age face image synthesis, a cross-age face image synthesis method based on a Conditional Adversarial AutoEncoder (CAAE) is proposed. By introducing both channel and spatial attention into the decoder structure, more important parts are taken from the channel and spatial domains, respectively, so that the model ignores irrelevant information such as the background in the training process, focuses on the changing area of the face image, and effectively avoids the distortion and deformation of organs in synthetic images. In addition, a multi-scale feature loss network is designed to constrain the local structural features of face images from multiple scales to maintain the stability of the local feature structure in the face synthesis process. The experimental results from the UTKFace cross-age face dataset show that compared with the CAAE method, this approach effectively prevents the deformation and distortion of facial organs, can better maintain the local structural features of the face, and has a better face synthesis effect and detail retention ability.

**[Key words]** cross-age face synthesis; Conditional Adversarial AutoEncoder (CAAE); channel attention module; spatial attention module; multi-scale feature loss network

DOI: 10.19678/j.issn.1000-3428.0062018

## 0 概述

跨年龄人脸合成技术旨在预测特定人脸图像过

去或者未来的变化, 能大幅增强人脸识别系统的性能, 并在动漫娱乐、公共安全、寻找失踪儿童等领域具有广泛的应用。但近年来上述领域的研究一直面

**基金项目:** 国家自然科学基金(61471154, 61876057); 安徽省重点研发计划科技强警专项(202004d07020012)。

**作者简介:** 程志康(1997—), 男, 硕士研究生, 主研方向为计算机视觉、机器学习; 孙锐(通信作者), 教授、博士; 孙琦景, 硕士研究生; 张旭东, 教授、博士。

收稿日期: 2021-07-08

修回日期: 2021-09-03

E-mail: 2789094552@qq.com

面临着数据稀缺的巨大挑战,很多研究工作都需要同一个人不同年龄段的多幅人脸图像,但这在实际生活中很难实现,导致训练难以达到最优效果。此外,由于生成网络训练的不稳定性,因此合成的跨年龄人脸图像容易出现器官扭曲变形、人脸的特征结构保持效果不佳等问题。

传统的跨年龄人脸合成方法大致可分为基于物理模型的方法和基于原型的方法这2种方法。基于物理模型的方法通过对面部肌肉、皱纹、皮肤、面部轮廓等生物学面部变化进行复杂的建模来模拟衰老效果。LANITIS等<sup>[1]</sup>和RAMANATHAN等<sup>[2]</sup>将面部结构建模为物理衰老模型进行跨年龄人脸合成, RAMANATHAN等<sup>[3]</sup>和BERG等<sup>[4]</sup>探索了跨年龄人脸合成中衰老面孔的纹理变化, SUO等<sup>[5]</sup>通过捕获脸部肌肉的相关信息进行跨年龄人脸合成。上述方法会产生粗略的老化效果,需要同一个人大量并且年龄跨度很大的照片,且对相关参数的复杂度较高。基于原型的方法<sup>[6-7]</sup>将面部图像分为不同的年龄组,并学习各组之间的衰老模式,因此在一定程度上可以放宽对同一个人年龄跨度较大的配对样本需求。该方法以每个年龄段的平均人脸为原型,将原型之间的差异视为衰老模式。WANG等<sup>[8]</sup>设计了一个循环面孔老化(Recurrent Face Aging, RFA)模型,该模型捕获了相邻年龄组之间的中间演化状态,并采用2层门控循环单元(Gate Recurrent Unit, GRU)来建模复杂的动态外观变化,能够合成一些具有老化迹象的图像,但较依赖于成对样品的可用性,然而这些样本难以收集且成本很高。另外,基于原型的方法以平均人脸作为原型,导致难以捕捉到每个人个性化的人脸特征,并且由于使用了平滑纹理,因此无法很好地捕捉高频细节(皱纹、斑点等)。为更好地保持人脸的个性化特征, SHU等<sup>[9]</sup>提出一种基于字典学习的人脸老化方法,将每个年龄组的年龄模式学习到相应的子字典中。给定的人脸将会被分解为年龄模式和个人模式两个部分,通过子字典将年龄模式转换为目标年龄模式,然后使用综合目标年龄模式和个人模式生成老化的面孔,但该方法会出现严重的重影伪影。

近年来,深度生成模型在图像合成中展现了较好的性能,部分学者也开展了基于深度生成网络的跨年龄人脸合成研究。DUONG等<sup>[10]</sup>构建一种基于时间深度限制的玻尔兹曼机器的年龄老化模型,能够捕捉非线性老化的变化。DUONG等<sup>[11]</sup>提出一种时间无量保存(Temporal Non-Volume Preserving, TNVP)老化方

法,该方法具有易于处理的密度函数,可以生成高质量的按年龄划分的人脸图像,但由于在建模时没有关于人脸个性的任何信息输入,因此按照完整的衰老顺序合成的人脸图像在颜色、表情、身份上均有很大差异。值得注意的是,尽管上述方法均有一定效果,但需要成对的训练数据以确保合成高质量人脸图像。为了解决收集成对实验训练样本的难题, ZHANG等<sup>[12]</sup>提出一种新的跨年龄人脸合成算法。该算法结合了对抗性自动编码器<sup>[13]</sup>和生成对抗网络(Generative Adversarial Network, GAN)<sup>[14]</sup>的优点,在没有配对的输入输出图像整体框架中实现了跨年龄人脸合成。相对于只进行人脸老化单方向的合成工作,该方法能同时实现跨年龄人脸图像的老化和去龄化。在训练网络时,该方法只需要将带有年龄标签的人脸图像输入到网络模型中;在测试网络时,结合需要转化的人脸图像和目标年龄标签就能得到指定年龄的人脸合成图像。该方法不仅能产生具有年龄变化效果的合成人脸图像,而且还保留了人脸的个性化信息。但基于传统的条件对抗自动编码器跨年龄人脸合成模型所生成的图像常出现器官扭曲变形以及人脸局部特征结构保持效果不佳的问题。

本文通过设计一种多尺度特征损失网络,对输出人脸图像的局部特征结构进行约束,优化生成的人脸图像局部特征结构。针对生成网络模型会出现生成图像器官扭曲变形的问题,对编码器解码器进行改进,并将通道关注和空间关注模块引入到解码器结构中,改善合成图像的效果。

## 1 传统的条件对抗自动编码器模型

对抗自动编码器可以看成是生成对抗网络和变分自动编码器的一种结合体。GAN包含生成网络 $G$ 和判别网络 $D$ 这2个相互对抗的网络,采用博弈论的原理,通过2个网络之间不断进行对抗博弈,使生成网络 $G$ 能够学习到真实数据的分布。生成网络 $G$ 主要从一个随机的噪声 $z$ (随机数)生成可以欺骗判别网络 $D$ 的假图片,判别网络 $D$ 主要对生成网络 $G$ 生成的假图片和真实图片进行判别,判别两者是一张图片的概率。生成网络 $G$ 和判别网络 $D$ 在不断对抗博弈中能达到一种平衡状态,称之为纳什均衡<sup>[15]</sup>。理想情况是生成网络能够生成足够真的图片,判别网络难以判定生成图片的真实性。变分自动编码器保留了基本的编码器解码器结构,与传统的自编码器通过数值的方式描述潜在空间不同,它以概率的方式描述对潜在空间的观察,在图片重构上具有较

为广泛的应用。对抗自动编码器不仅可以像变分自动编码器一样保留自动编码器网络,代替 GAN 从随机噪声生成图像,而且可以像 GAN 中的对抗性网络

那样代替 KL 散度损失。  
基于传统的条件对抗自动编码器的跨年龄人脸合成模型的原理如图 1 所示。

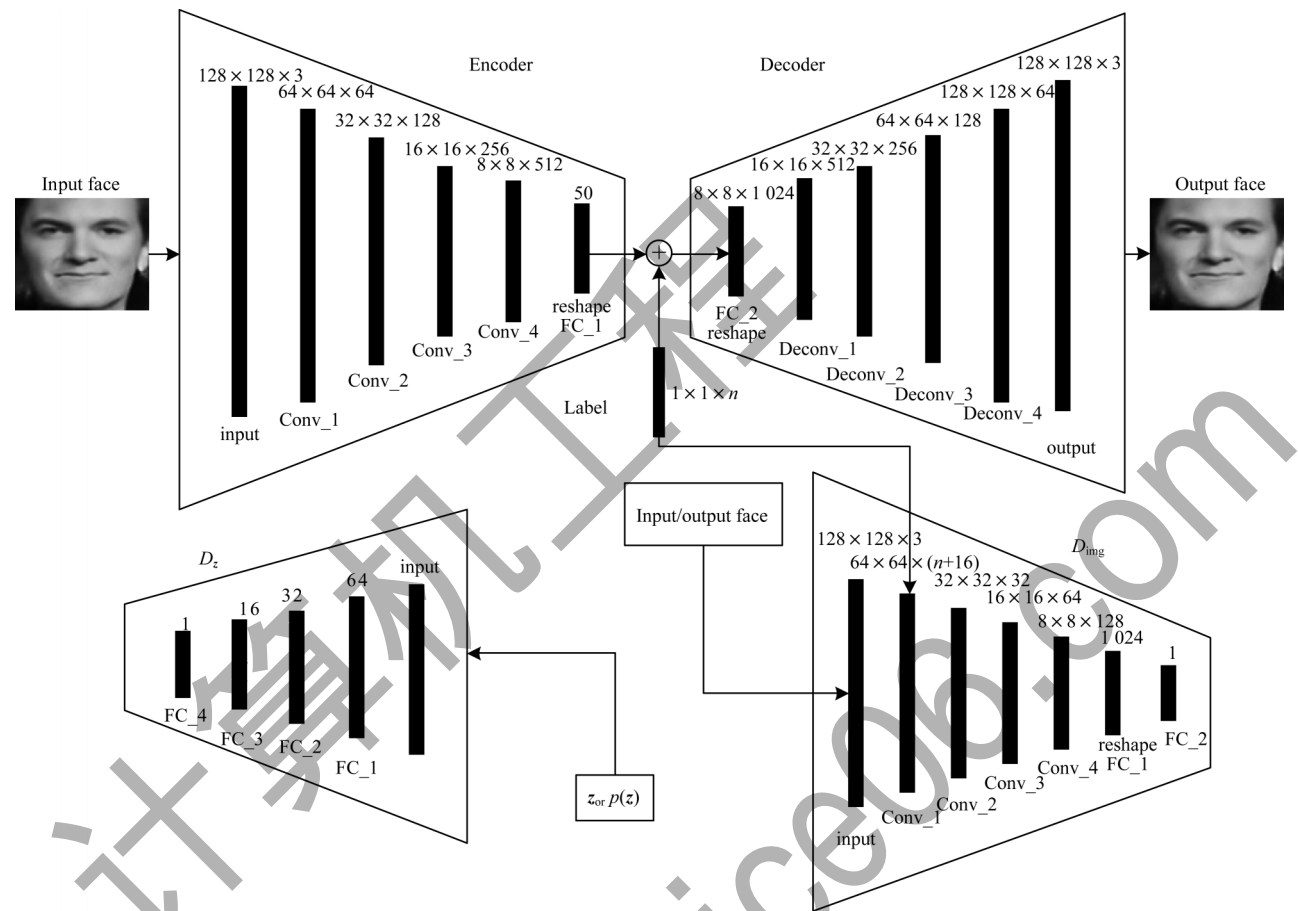


图 1 传统的条件对抗自动编码器模型网络框架

Fig.1 Network framework of traditional conditional adversarial autoencoder model

从如图 1 可以看出,编码器解码器作为主体结构主要完成输入人脸图像的重构,2 个判别器能使模型在跨年龄人脸合成过程中生成更加逼真的图像。在训练阶段,一开始输入的  $128 \times 128 \times 3$  像素的人脸图像会被编码器映射到低维空间,在低维空间中可以得到一个具有人脸个性化特征的 50 维度特征向量  $z$ 。对于输入的具有年龄标签的人脸图像,在低维空间中将大小为  $1 \times 1 \times n$  的年龄标签向量与具有个性化人脸特征的特征向量  $z$  进行拼接。经过拼接的具有年龄和人脸特征信息的特征向量通过解码器重新恢复到高维空间中,同样得到一个  $128 \times 128 \times 3$  像素的人脸图像。为了使跨年龄人脸合成的图像更加逼真,在低维空间中具有人脸个性化特征的特征向量  $z$  上以及输入输出人脸图像上分别施加判别器。施加在特征向量  $z$  上的判别器  $D_z$  可以施加先验分布,例如均匀分布,使在潜在低维空间中的  $z$  具有均匀的分布,能够产生更加均匀

的人脸图像。施加在面部图像上的鉴别器  $D_{img}$  能够保证生成器生成更加逼真的人脸图像。在测试过程中只有编码器和解码器工作,将人脸图像输入网络结构中,只需要添加特定的年龄标签,就可以产生特定年龄的人脸合成图像。

2 本文方法

2.1 网络框架

本文方法的网络框架如图 2 所示,主要包括 1 个编码器、1 个解码器、2 个判别器  $D_z$  和  $D_{img}$  以及 1 个多尺度特征损失网络。其中:编码器解码器构成跨年龄人脸合成的基础结构,能够实现人脸图像的转换工作;2 个判别器能够约束生成的人脸图像更加逼真,保证合成的效果;中间连接器将映射到低维空间中的人脸特征向量  $z$  与人脸年龄标签进行连接;多尺度特征损失网络从多个尺度约束输入人脸图像和输出人脸图像之间的局部特征结构,使生成的人脸图像能够保留局部特征结构。



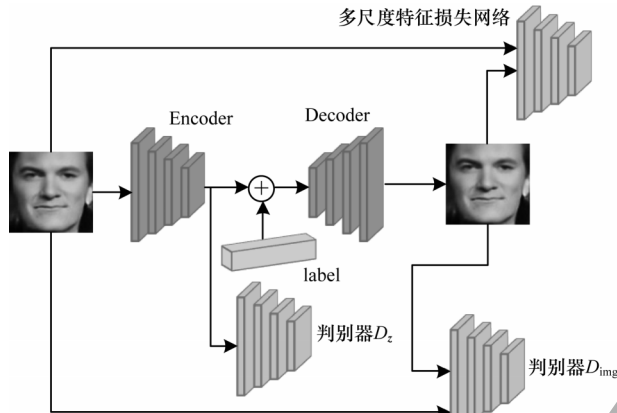


图2 本文方法的网络框架

Fig.2 Network framework of method in this paper

## 2.2 区域关注编码器-解码器

在跨年龄人脸合成过程中,需要关注脸部区域,忽略背景信息。本文对编码器解码器结构进行改进,在解码器中引入区域关注模块,从而在通道和空间中取得更加重要的部分,使人脸生成的效果更好。区域关注模块将输入的信息进行权重的标定,赋予重要信息更大的权重,减少不重要信息的权重,从而对关键位置的信息进行提取,比较符合人类视觉观察事物的特性。如图3所示是本文所改进的编码器解码器结构。

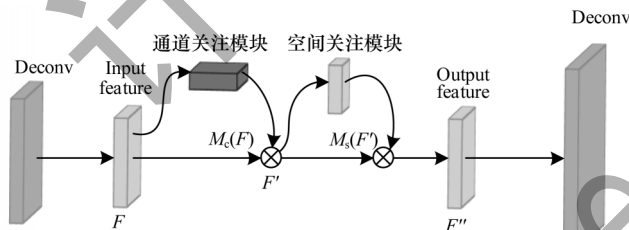


图3 区域关注的编码器解码器结构

Fig.3 Structure of encoder and decoder of region attention

在解码器中的2个反卷积层中引入区域关注模块,分别进行通道关注和空间区域关注,并令经过反卷积层所得到的特定特征图  $F \in \mathbb{R}^{C \times H \times W}$  分别经过一个一维的通道关注图  $M_c \in \mathbb{R}^{C \times 1 \times 1}$  和一个二维空间关注图  $M_s \in \mathbb{R}^{1 \times H \times W}$ 。通道关注通过利用通道间的关系生成通道注意力图,由于特征图的每个通道均被视为特征检测器<sup>[16]</sup>,因此通道关注于一张输入图像是否有意义。空间区域关注则是利用特征之间的空间关系生成空间区域注意力图,它更加关注于输入图片的重要信息部分。两者相辅相成,能够抑制不重要信息的干扰,对图像合成效果具有提升作用。总的进程可以表示为:

$$F' = M_c(F) \otimes F \quad (1)$$

$$F'' = M_s(F') \otimes F' \quad (2)$$

其中:  $\otimes$  表示逐像素相乘;  $F$  是通过解码器得到的人脸特征图;  $M_c(F)$  表示特征图  $F$  经过通道关注得到的特征图;  $F'$  是经过通道关注模块之后的中间特征图;  $M_s(F')$  表示特征图  $F'$  经过空间关注得到的特征图;  $F''$  是中间特征图经过空间关注模块之后得到的最终的人脸特征图。在训练过程中,区域关注模块会分别推断出通道和空间2个方向的特定权重,使人脸合成过程减少对不必要的背景等信息的关注,聚焦于人脸重要区域的合成工作。

## 2.3 多尺度特征损失网络

为了在跨年龄人脸合成过程中保持人脸局部特征结构,只进行年龄相关的变化,本文设计了一个多尺度特征损失函数网络,如图4所示。

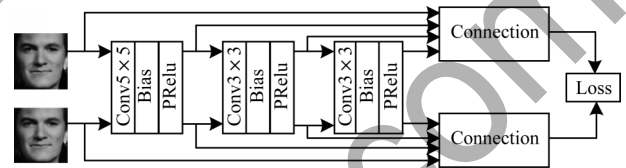


图4 多尺度特征损失网络

Fig.4 Multi-scale feature loss network

在跨年龄人脸合成训练过程中,每幅人脸图像都会加入年龄标签进行重构,在此过程中使合成的图像和输入的人脸图像具有更相似的局部特征结构尤其重要,为此需要对两幅图像进行局部特征结构的约束。但是通过简单的  $L_1$  损失函数或  $L_2$  损失函数往往不能很好地约束人脸的个性化特征保持不变,因此本文设计了一个多尺度特征损失网络对两幅人脸图像进行多重的局部特征结构约束,保证人脸个性化特征的保持效果更好,避免人脸图像局部特征结构变形等情况的发生。如图4所示,将输入的图像和经过跨年龄合成主体网络后的合成图像一同送入多尺度特征损失网络,分别经过3层设计好的卷积神经网络。对于原始输入的2张图像以及经过每层卷积层后的特征图分别进行 Charbonnier 损失函数约束,最终得到多尺度损失函数的表达式如式(3)所示:

$$L_{cb} = \frac{1}{L} \sum_{i=1}^{i=L} \sqrt{[F_i(x) - F_i(x')]^2 + \varepsilon^2} \quad (3)$$

其中:  $L$  取4;  $x$  和  $x'$  分别表示输入的人脸图像和合成的人脸图像;  $F_i(x)$  表示获取人脸图像的特征图;  $i$  表示第  $i$  个人脸特征图;  $\varepsilon$  是一个很小的常量,目的是增强损失函数计算的稳定性,一般取值为  $1 \times 10^{-3}$ 。

本文网络采用 PReLU<sup>[17]</sup> 作为激活函数,该激活函数在 ReLu 的基础上进行了改进,能够避免人脸相关特征的丢失。ReLu 的表达式如式(4)所示:

$$f_{\text{ReLU}}(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (4)$$

ReLU 激活函数以 0 作为阈值,与 Sigmoid 和 tanh 2 个激活函数相比,在梯度下降上有更快的收敛速度,并且在一定程度上能有效抑制梯度消失现象。PReLU 激活函数的表达式如式(5)所示:

$$f_{\text{PReLU}}(x) = \begin{cases} x, & x > 0 \\ ax, & x \leq 0 \end{cases} \quad (5)$$

ReLU 激活函数将负数强行置 0 可能会导致相关特征的丢失,PReLU 激活函数相对于 ReLU 激活函数多加了一个参数  $a$ ,避免了这种情况的发生,其中, $a$  是一个可以学习的值。

## 2.4 损失函数

传统的条件对抗自动编码器的基本结构包括编码器、解码器、鉴别器  $D_z$ 、鉴别器  $D_{\text{img}}$ ,本文所提出的结构在此基础上增加了一个多尺度特征损失网络。本文所提网络结构的损失函数在原来网络的基础上用多尺度特征损失网络的损失函数代替了原来的  $L_2$  损失函数。多尺度特征损失网络能够保证输入人脸和合成人脸的人脸局部特征结构的一致性,并且比基本的  $L_2$  损失函数具有更好的效果,多尺度特征损失网络得到的损失函数如式(3)所示。

鉴别器  $D_z$  对低维空间中的人脸特征  $z$  施加约束,均匀分布被施加在  $z$  上作为先验分布。用  $p_{\text{data}}(x)$  表示训练数据的分布, $z$  的分布表示为  $q(z|x)$ ,假设  $p(z)$  是一个先验分布, $z^* \sim p(z)$  表示从  $p(z)$  随机采样。所以使鉴别器  $D_z$  认为  $z$  来自先验分布增加的对抗性损失可以定义为:

$$L_{\text{GD } z} = E_{z^* \sim p(z)}[\log_a D_z(z^*)] + E_{x \sim p_{\text{data}}(x)}[\log_a (1 - D_z(E(x)))] \quad (6)$$

鉴别器  $D_{\text{img}}$  帮助合成的人脸更加逼真,同样,带有年龄标签  $I$  的解码器施加一个对抗损失可以定义为:

$$L_{\text{GD img}} = E_{x, I \sim p_{\text{data}}(x, I)}[\log_a D_{\text{img}}(x, I)] + E_{x, I \sim p_{\text{data}}(x, I)}[\log_a (1 - D_{\text{img}}(G(E(x), I)))] \quad (7)$$

其中: $(x, I)$  表示年龄为  $I$  的人脸图像  $x$ ;  $G(E(x), I)$  表示经过编码器得到的向量和年龄标签向量  $I$  在拼接之后送到解码器所得到的人脸图像。

最终得到总的损失函数为:

$$L_{\text{total}} = \lambda_1 L_{\text{cb}} + L_{\text{GD } z} + L_{\text{GD img}} \quad (8)$$

其中: $L_{\text{cb}}$  表示多尺度特征损失。

## 3 实验结果与分析

本节主要介绍本文方法的实现细节以及验证本文所提模型相较于传统的条件对抗自动编码器(Conditional Adversarial AutoEncoder, CAAE)模型以及人脸转换

(Face Transformer, FT)模型<sup>[18]</sup>的效果。实验分别从定性和定量的角度验证本文所提模型的优越性。

### 3.1 实验数据集

实验使用 UTKFace 人脸数据集,图 5 是部分数据集中的人脸图像。UTKFace 数据集的人脸图像超 20 000 张,均具有年龄和性别标签,年龄跨度为 1 ~ 78 岁。本文剔除了 UTKFace 数据集中的极度不清晰图像,并通过上网搜集了部分清晰图像作为补充。通过人脸的 68 个特征点对收集到的人脸图像进行人脸检测<sup>[19-20]</sup>,使图像满足实验的要求。对于某些未知年龄的图像,采用年龄分类器<sup>[21]</sup>进行测量,给定每幅图像相应的年龄标签。此外,对实验数据进行年龄分类,使其拥有不同的年龄标签。由于人脸在低年龄段的变化较大,因此最终的分组规则是将人脸年龄分为 10 类,分别是 0 ~ 5 岁、6 ~ 10 岁、11 ~ 15 岁、16 ~ 20 岁、21 ~ 30 岁、31 ~ 40 岁、41 ~ 50 岁、51 ~ 60 岁、61 ~ 70 岁和 70 岁以上。



图5 UTKFace数据集人脸图像示例

Fig.5 Face image example of UTKFace data set

### 3.2 实验环境和参数设置

本文的实验环境:硬件平台为 PC: Intel Core i7-8700 CPU,型号为 Nvidia GeForce GTX1070Ti 的显卡,内存为 16 GB,使用的语言是 python 语言。针对本文的数据集,损失函数的参数  $\lambda_1$  取值为 100。本文采用 ADAM<sup>[22]</sup> 学习策略来动态调节学习率,其中  $\alpha = 0.0002$ ,  $\beta_1 = 0.5$ 。

### 3.3 定性分析

为更直观地观察本文方法生成的人脸图像质量,本文挑选出来自不同年龄段、不同性别的人脸测试图像分别进行跨年龄人脸合成。对每输入一张测试人脸的图像,都分别输出 0 ~ 5 岁、6 ~ 10 岁等 10 个年龄段的合成人脸图像。调用预训练模型进行训练,实验结果如图 6 所示。如图 6 所示,第 1 行分别是输入的特定年龄和性别的人脸图像,第 1、3、5、7、9 列的合成图像是 CAAE 方法生成的,而第 2、4、6、8、10 列的合成图像是基于本文方法生成的。对比可知,本文方法更好地保持了跨年龄人脸合成过程中人脸的局部特征结构,解决了合成过程中出现的人脸扭曲、眼睛等器官变形的的问题。图 6 中方框标注的几幅对比图在人脸局部特征结构的保持效果上更加明显,由这几幅图可知,尤其是眼部器官及周围的区域,本文方法相较于 CAAE 方法能更好地避免人脸器官变形情况的发生。



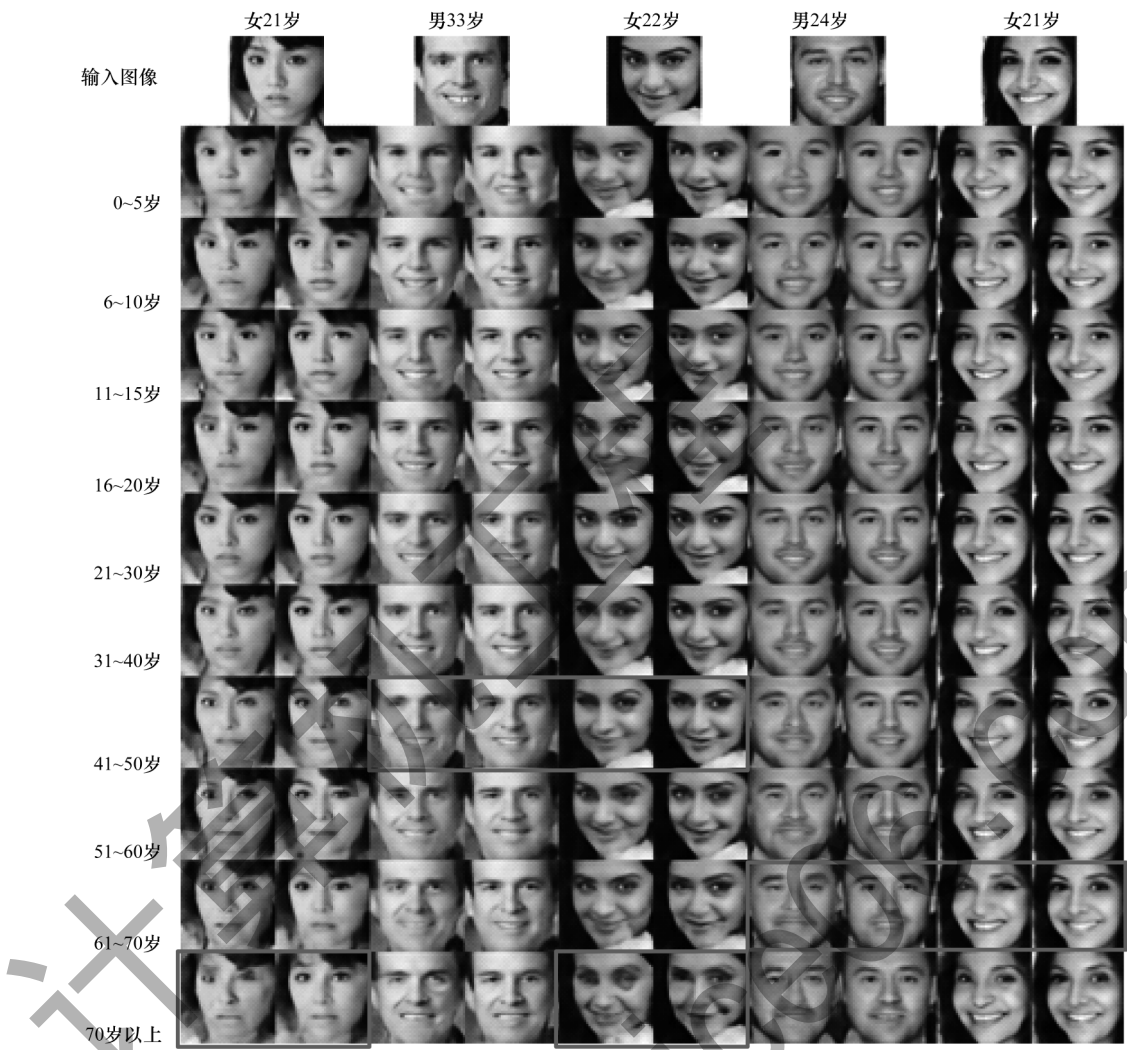


图 6 本文方法与 CAAE 方法的结果对比

Fig.6 Result comparison between the method in this paper and CAAE method

将本文方法与 FT 方法进行对比, 结果如图 7 所示。由图 7 可知, 本文方法在跨年龄人脸合成过程中能够较好地保持人脸的特征结构, 避免了人脸器官的扭曲现象, 尤其是图中标记方框的人脸图

像。如图 7(c) 所示, FT 方法产生的人脸图像眼睛部位产生了扭曲变形的情况, 图 7(a) 所示 FT 方法产生的人脸图像脸颊部位甚至出现了特征缺失的现象。

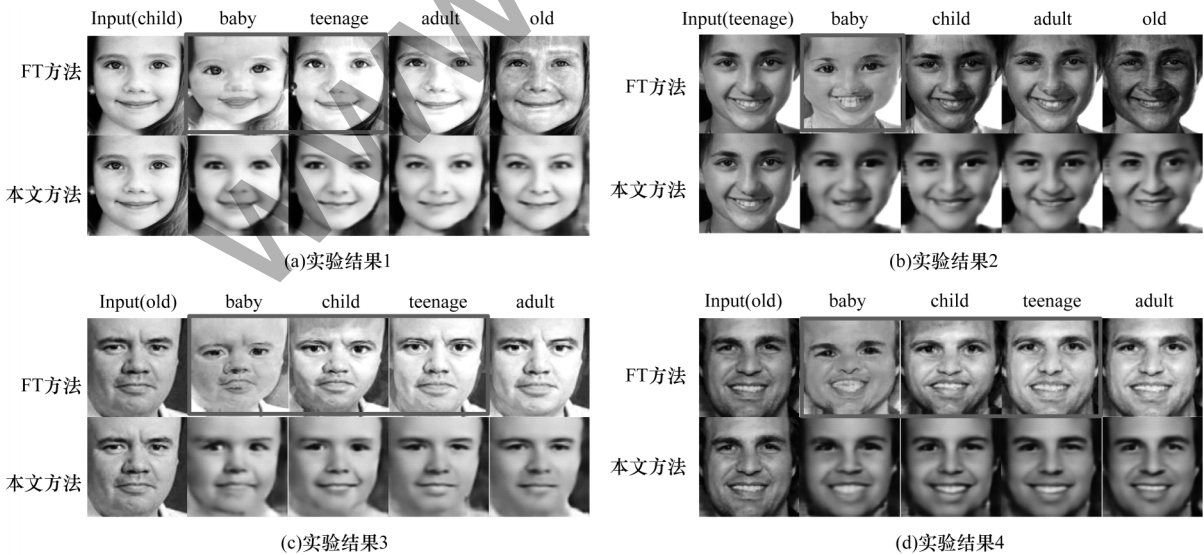


图 7 本文方法与 FT 方法的结果对比

Fig.7 Result comparison between the method in this paper and FT method

为证明本文所提区域关注模块在跨年龄人脸合成中的有效性,选取5组特定年龄和性别的人脸图像作为输入图像,分别观察人脸图像的合成效果。调用预训练模型分别进行训练,实验结果如图8所示。

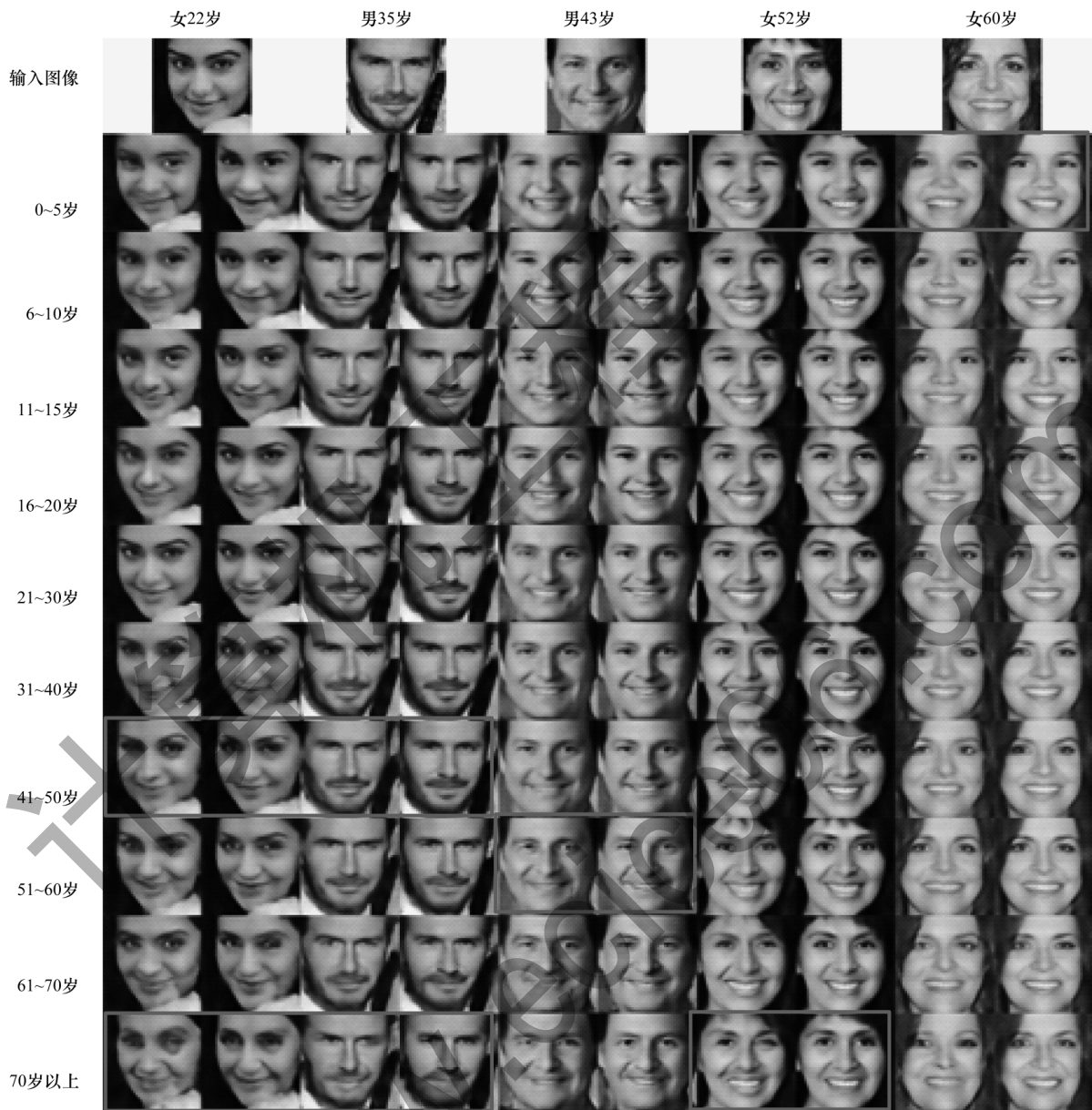


图8 区域关注模块的消融实验结果

Fig.8 Ablation experiment result of region attention module

在图8中,第1、3、5、7、9列的合成图像是网络中没有添加区域关注模块所生成的,而第2、4、6、8、10列的合成图像是网络中添加了区域关注模块所生成的。对比可知,基于区域关注的编码器解码器结构所合成的人脸图像更好地避免了人脸的扭曲现象和眼睛等器官的变形现象,图8中方框标注的图像更能体现这一点,证明了区域关注模块在跨年龄人脸合成任务中的有效性。

基于传统的条件对抗自动编码器模型仅仅采用简单的 $L_2$ 损失来约束合成的跨年龄人脸图像,这往往并不能很好地保持合成人脸图像的局部特征结构。本文所设计的多尺度特征损失网络从多个尺度约束输入人脸图像和输出人脸图像之间的局部特征结构,具有更好的人脸局部特征结构保

持效果。为证明本文所设计的多尺度特征损失网络有效性,对添加多尺度特征损失网络以及不添加多尺度特征损失网络(仅仅使用简单的 $L_2$ 损失函数约束)分别进行实验。分别调用预训练模型进行训练,实验结果如图9所示。图9第1行是输入的特定性别和年龄的人脸图像,第1、3、5、7、9列分别是未添加多尺度特征损失网络所合成的人脸图像,第2、4、6、8、10列分别是添加了多尺度特征损失网络所合成的人脸图像。对比可知,添加了多尺度特征损失网络所合成的人脸图像更能够保持人脸的局部特征结构。由图9中标方框的几幅对比图可知,本文所提出的多尺度损失网络更好地保持了人脸图像眼部的局部特征结构,验证了多尺度特征损失网络的有效性。



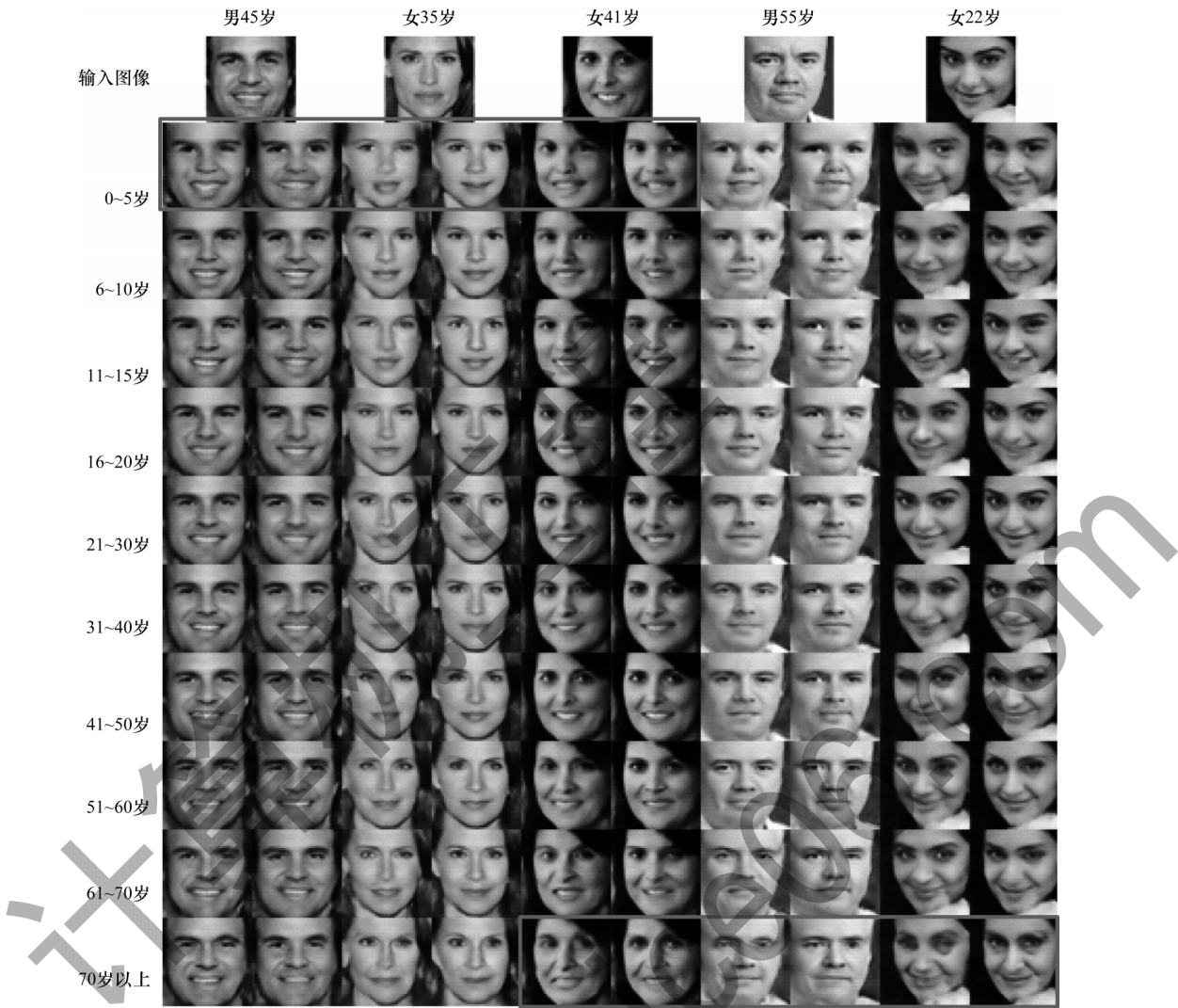


图9 多尺度特征损失网络的消融实验结果

Fig.9 Ablation experiment result of multi-scale feature loss network

3.4 定量分析

为了从定量指标上进一步验证本文方法相较于CAAE方法、FT方法的优越性,分别采用志愿者评价、余弦相似度(Cosine Similarity, CS)、结构相似度(Structural Similarity, SSIM)、年龄估计精度这4个指标进行定量评估。

3.4.1 志愿者评价

为更好地评估本文方法所生成的跨年龄人脸图像合成效果,邀请了100名志愿者参与生成图像质

量评价,在评价指标打分上要高于CAAE方法以及FT方法,验证了本文提出算法的有效性。

表1 不同方法的志愿者评价结果对比

Tabe1 Comparison of volunteer evaluation results of different methods

方法	得分
CAAE方法	8.9
FT方法	8.8
本文方法	9.1

3.4.2 余弦相似度评价

余弦相似度是通过计算2个向量A和B的夹角余弦值来评估它们的相似度,计算公式如式(9)所示:

$$\cos \theta = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \tag{9}$$

其中:A<sub>i</sub>和B<sub>i</sub>表示向量A和B的各分量。零度角的

量的相关评估。给定每个志愿者输入的人脸图像、用CAAE方法生成的跨年龄人脸图像、用FT方法生成的跨年龄人脸图像以及本文方法生成的跨年龄人脸图像,让志愿者根据给定的输入图像分别对2种方法生成的每个年龄段图像进行人脸局部特征保持度评价指标的打分,判别合成的人脸图像产生的器官扭曲变形程度。将分值定在0~10分之间,分数越高代表人脸局部特征结构保持得更好,能避免人脸器官扭曲变形等情况。对志愿者的打分求平均值,结果如表1所示。由表1可知,本文方法在志愿



余弦值是1,而其他任何角度的余弦值均不大于1,最小值是-1,两幅图像之间的余弦相似度指标越接近于1,表示两者之间的相似度越高。本文通过计算输入的人脸图像和生成图像的余弦相似度指标,评估本文方法的生成图像和输入图像人脸相似度。选取多张待测试的人脸图像,分别计算输入图像和不

同方法所合成的人脸图像之间的余弦相似度,最后分别对余弦相似度求均值,结果如表2所示。由表2可知,本文方法的余弦相似度指标要高于CAAE方法,验证了本文方法在跨年龄人脸合成过程中人脸局部特征结构的保持效果更好,避免了人脸器官扭曲变形的问题。

表2 本文方法与CAAE方法的余弦相似度对比

Table 2 Comparison of cosine similarity between method in this paper and CAAE method										
方法	0~5岁	6~10岁	11~15岁	16~20岁	21~30岁	31~40岁	41~50岁	51~60岁	61~70岁	70岁以上
CAAE方法	0.949 72	0.941 05	0.960 51	0.952 55	0.971 56	0.963 83	0.959 52	0.962 83	0.944 51	0.959 48
本文方法	0.969 96	0.968 39	0.980 12	0.979 36	0.989 36	0.980 62	0.985 28	0.981 28	0.967 82	0.973 52

由于FT方法只有baby、child、teenage、adult、old 5个年龄段的人脸合成图像,所以本文方法与FT方法的余弦相似度对比只取这5个年龄段的人脸图像。对于选取的人脸图像,分别计算输入的人脸图像和2种方法生成的人脸图像间的余弦相似度值,最后对50个实验结果求平均值,实验结果如表3所示。

表3 本文方法与FT方法的余弦相似度对比

Table 3 Comparison of cosine similarity between method in this paper and FT method					
方法	baby	child	teenage	adult	old
FT方法	0.802 46	0.931 62	0.948 91	0.942 31	0.802 87
本文方法	0.970 21	0.974 55	0.971 32	0.980 73	0.975 23

由表3可知,本文方法相较于FT方法的余弦相似度更高,能产生更好的人脸特征保持效果。

3.4.3 结构相似度评价

为验证本文方法在人脸的局部特征结构保持上的优越性以及避免人脸扭曲变形问题上的有效性,采取结构相似度分别从结构、亮度、对比度这

3个方面来度量图像间的相似性。对于给定的2个图像 $x$ 和 $y$ ,2个图像间的SSIM表达式如式(10)所示:

$$S_{SSIM}(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_{2x}+\mu_{2y}+C_1)(\sigma_{2x}+\sigma_{2y}+C_2)} \quad (10)$$

其中: $\mu_x$ 是 $x$ 的平均值; $\mu_y$ 是 $y$ 的平均值; $\sigma_{2x}$ 表示 $x$ 的方差; $\sigma_{2y}$ 表示 $y$ 的方差; $\sigma_{xy}$ 是 $x$ 和 $y$ 的协方差。 $C_1=(k_1L)^2$ 和 $C_2=(k_2L)^2$ 是用来维持稳定的常数; $L$ 是像素值的动态范围; $k_1=0.01$ ; $k_2=0.03$ 。SSIM值越接近1,说明重建后的图像与原图结构越相似,重建效果越好。

挑选多张人脸图像,对合成的人脸图像和输入的人脸图像进行SSIM指标的测量,最后对所得到的50次结果求均值,SSIM的值越大,表示2幅图像在局部特征结构上的相似性越高。表4所示为CAAE方法和本文方法合成人脸图像的SSIM值对比结果。由表4可知,本文方法相比于CAAE方法具有更高的结构相似度指标,验证了本文方法在人脸局部特征保持上的优越性以及避免人脸扭曲变形等问题的有效性。

表4 本文方法与CAAE方法的结构相似度对比

Table 4 Comparison of structural similarity between method in this paper and CAAE method										
方法	0~5岁	6~10岁	11~15岁	16~20岁	21~30岁	31~40岁	41~50岁	51~60岁	61~70岁	70岁以上
CAAE方法	0.732 6	0.740 9	0.763 4	0.765 0	0.773 2	0.775 9	0.767 7	0.744 2	0.738 2	0.730 3
本文方法	0.741 4	0.750 1	0.779 2	0.773 4	0.788 9	0.781 5	0.770 4	0.752 6	0.745 3	0.739 4

取多张人脸图像,并分别测量输入人脸图像和不同方法合成的人脸图像间的结构相似度,最后对50次测量结果取均值,实验结果如表5所示。由表5可知,与FT方法相比,本文方法的结构相似度更高,验证了其具有更好的人脸特征保持效果。

表5 本文方法与FT方法的结构相似度对比

Table 5 Comparison of structural similarity between method in this paper and FT method					
方法	baby	child	teenage	adult	old
FT方法	0.461 2	0.558 6	0.750 2	0.758 5	0.613 1
本文方法	0.750 1	0.751 0	0.774 2	0.786 5	0.743 5

3.4.4 年龄估计精度评价

为验证本文方法在跨年龄生成效果上的优势,本文采用预训练的排序卷积神经网络对生成的图像进行年龄估计。以估计结果与目标年龄标签之间的平均绝对值误差(Mean Absolute Error,MAE)作为评价指标,MAE值越低,表示年龄跨越的准确性越高。选取多张人脸图像,分别对不同方法合成的人脸图像进行测试,并对50次测量结果取平均值。为了兼顾FT方法,分别测试了0~5岁、6~10岁、16~20岁、21~30岁、51~60岁共5个年龄分段,实验结果如表6所示。由表6可知,本文方法在年龄估计精度上优于MAE和FT方法,验证了本文方法能够生成与目标年龄段年龄更接近的人脸图像。

表6 不同方法的年龄估计精度对比

Table 6 Comparison of age estimation accuracy of different methods

年龄段	0~5岁	6~10岁	16~20岁	21~30岁	51~60岁
FT方法	5.43	8.01	5.26	4.01	9.61
CAAE方法	11.34	7.62	4.34	3.62	8.13
本文方法	11.12	6.54	4.06	3.44	7.99

#### 4 结束语

针对跨年龄人脸合成过程中出现的合成图像局部特征结构保持效果不佳、容易产生器官扭曲变形等问题,本文提出一种基于条件对抗自动编码器的跨年龄人脸合成方法。在传统的条件对抗自动编码器模型的基础上设计一种基于区域关注的编码器解码器结构,在解码器结构中引入通道关注和空间关注模块,使模型在人脸合成过程中忽略背景信息,更加关注人脸变化区域,减少生成的人脸图像器官扭曲变形等情况发生。此外,设计一种多尺度特征损失网络对跨年龄人脸合成过程进行约束,从多个尺度约束输入人脸图像和输出人脸图像之间的局部特征结构,使合成人脸局部特征结构得到保持。实验结果表明,与CAAE方法相比,本文方法不仅具有较好的人脸局部特征保持效果,而且更好地解决了合成过程中出现的器官扭曲变形等问题。

#### 参考文献

- [1] LANITIS A, TAYLOR C J, COOTES T F. Toward automatic simulation of aging effects on face images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4): 442-455.
- [2] RAMANATHAN N, CHELLAPPA R. Modeling age progression in young faces [C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2006: 387-394.
- [3] RAMANATHAN N, CHELLAPPA R. Modeling shape and textural variations in aging faces [C]//Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition. Washington D. C., USA: IEEE Press, 2008: 1-8.
- [4] BERG A C, PERALES LOPEZ F J, GONZALEZ M. A facial aging simulation method using flaccidity deformation criteria [C]//Proceedings of the 10th International Conference on Information Visualisation. Washington D. C., USA: IEEE Press, 2006: 791-796.
- [5] SUO J L, ZHU S C, SHAN S G, et al. A compositional and dynamic model for face aging [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(3): 385-401.
- [6] TIDDENMAN B, BURT M, PERRETT D. Prototyping and transforming facial textures for perception research [J] IEEE Computer Graphics and Applications, 2001, 21(5): 42-50.
- [7] KEMELMACHER-SHLIZERMAN I, SUWAJANAKORN S, SEITZ S M. Illumination-aware age progression [C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2014: 3334-3341.
- [8] WANG W, CUI Z, YEN Y, et al. Recurrent face aging [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 2378-2386.
- [9] SHU X, TANG J, LAI H, et al. Personalized age progression with aging dictionary [C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2015: 3970-3978.
- [10] DUONG C N, LUU K, QUACH K G, et al. Longitudinal face modeling via temporal deep restricted Boltzmann machines [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 5772-5780.
- [11] DUONG C N, QUACH K G, LUU K, et al. Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition [C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 3755-3763.
- [12] ZHANG Z, SONG Y, QI H. Age progression/regression by conditional adversarial autoencoder [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 4352-4360.
- [13] MAKHZANI A, SHLENS J, JAITLY N, et al. Adversarial autoencoders [EB/OL]. [2021-06-06]. <https://arxiv.org/abs/1511.05644>.
- [14] GOODFELLOW I J, POUGET ABADIE J, MIRZA M, et al. Generative adversarial nets [EB/OL]. [2021-06-06]. <https://arxiv.org/abs/1406.2661>.
- [15] 杨林,王永杰. 基于单点多步博弈的网络防御策略选取方法 [J]. 计算机工程, 2021, 47(1): 154-164.
- [16] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2014: 818-833.
- [17] HE K M, ZHANG X Y, REN S Q. Delving Deep into rectifiers: surpassing human-level performance on ImageNet classification [C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2015: 1026-1034.
- [18] Face Transformer (FT) demo [EB/OL]. [2021-06-06]. <http://cherry.dcs.aber.ac.uk/transformer/>.
- [19] Dlib C++ Library [EB/OL]. [2021-06-06]. <http://dlib.net/>.
- [20] KAZEMI V, SULLIVAN J. One millisecond face alignment with an ensemble of regression trees [C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2014: 1867-1874.
- [21] ROTHE R, TIMOFTE R, GOOL L V. DEX: deep expectation of apparent age from a single image [C]//Proceedings of 2015 IEEE International Conference on Computer Vision Workshop. Washington D. C., USA: IEEE Press, 2015: 252-257.
- [22] KINGMA D P, BA J. ADAM: a method for stochastic optimization [EB/OL]. [2021-06-06]. <https://arxiv.org/pdf/1412.6980.pdf>.