

基于深层特征嵌入的高分辨率人脸图像重建*

缪志辉¹, 张永爱², 林志贤², 林坚普^{1**}

(1. 福州大学 先进制造学院, 福建泉州 362200; 2. 福州大学 物理与信息工程学院, 福州 350116)

摘 要: 提出了一种基于深层特征嵌入的高分辨率人脸图像重建方法, 利用预训练的 StyleGAN2 生成对抗网络模型作为人脸图像生成器, 并将深层特征嵌入到 StyleGAN2 的 W^+ 空间中, 通过梯度下降法优化 w^+ 向量, 并将优化后的 w^+ 向量输入到 StyleGAN2 中生成分辨率为 1024×1024 的重建人脸图像。实验结果表明, 重建图像与对应真实人脸图像不仅在视觉上有着较高的相似性, 且在同一特征提取网络下, LFW 和 ColorFeret 数据集的重建图像在 FAR 为 0.1% 时 II 型评估 TAR 分别为 96.04% 和 100.00%, 并且在两种不同的活体检测程序下的通过率分别达到了 88.67% 和 74.67%。所提方法在实现重建高分辨率人脸图像的同时, 与真实人脸图像的特征相似度方面也表现优秀。

关键词: 人脸图像重建; 深层特征; 生成对抗网络; 梯度下降

中图分类号: TP391 **文献标志码:** A **文章编号:** 1005-488X(2023)01-0017-09

High Resolution Face Image Reconstruction Based on Deep Feature Embedding

MIAO Zhihui, ZHANG Yongai, LIN Zhixian, LIN Jianpu

(1. School of Advanced Manufacturing, Fuzhou University, Quanzhou Fujian 362200, CHN;

2. School of Physics and Information Engineering, Fuzhou University, Fuzhou 350116, CHN)

Abstract: A method based on deep feature to reconstruct face images was proposed in this paper. The deep features were first embedded into the W^+ space of StyleGAN2, and the embedded w^+ vector was optimized by gradient descent and fed into a pre-trained StyleGAN2 model to generate face image with resolution of 1024×1024 . The experimental results showed that the reconstructed images had high visual similarity with the corresponding real face images. When using the same extraction network, the type-II evaluation TAR of the reconstructed images from LFW and ColorFeret datasets was 96.04% and 100.00% respectively when the FAR was equal to 0.1%, and the pass rates of live-

收稿日期: 2022-09-26

* 基金项目: 国家重点研发计划(2021YFB3600603); 福建省自然科学基金(2019J01221, 2020J01468); 福建省教育厅中青年教师教育科研项目(JAT210030)

作者简介: 缪志辉(1998—), 男, 硕士研究生, 研究方向为数字图像处理技术、人工智能技术; (E-mail: 1169840337@qq.com)

张永爱(1977—), 男, 研究员, 研究方向为数字图像处理技术、3D显示技术;

林志贤(1975—), 男, 教授, 研究方向为显示技术、电路与系统、数字图像处理技术;

林坚普(1989—), 男, 讲师, 研究方向为新型显示技术、人工智能技术、3D立体显示技术。(E-mail: ljplp@fzu.edu.cn)

** 通讯作者

ness detection were 88.67% and 74.67% in two products. The proposed method could reconstruct face images in high resolution and achieve high feature similarity between real and generated face images.

Key words: face image reconstruction; deep feature; generative adversarial network; gradient descent

引 言

近年来,深度学习技术的不断发展使其在许多生物特征识别和计算机视觉应用中取得了巨大的成功。生物特征识别是指根据个体的生物和行为特征(如面部,指纹、声音等)对个体进行身份识别。其中,由于基于深层卷积神经网络的人脸识别算法具有高效性、稳定性和便捷性等特点,人脸识别因此成为生物特征识别中的热门邻域,得到了广泛的应用^[1-3]。通常考虑到存储人脸图像所需的空间以及安全性,拥有人脸识别应用的互联网公司一般并不会直接存储人脸图像在他们的数据库中,而是存储通过深层卷积神经网络提取的人脸深层特征数据,通过对比深层特征之间的相似性进行身份识别。同时,深层特征也能够被重建为人脸图像直观反映用户的外貌信息,而深层特征能够何种程度上重建出人脸图像对其保护工作的研究具有重大价值。

早期基于深层特征重建人脸图像主要使用传统优化算法,如爬山算法^[4]和径向基函数回归法^[5]。随着深度学习的发展,基于深度学习的方法也被用于人脸图像重建,Zhmoginov 等人^[6]使用神经网络来反转 FaceNet 网络^[7]提取的低维人脸特征,同时生成图像,并在优化嵌入损失函数中加入了基于特征提取模型的中间层构建的额外正则化损失。Cole 等人^[8]利用面部身份特征对姿态、照明和表情的不变性,将重建问题定位为从特征向量到均匀照明、正面、中性表情的人脸映射,通过学习从面部识别网络提取的特征生成面部关键点和纹理实现图片重建。Mai 等人^[9]提出一种基于深层特征的邻域反卷积神经网络(Neighborly De-convolutional Neural Networks, NbNet),使用生成对抗网络(Generative Adversarial Networks, GAN)合成的人脸图像以及两个标准人脸数据集对 NbNet 进行训练,利用反卷积的上采样能力实现人脸图像重建。上述方法只能重建出低分辨率图像,且重建图像中均有伪影,很容易被人眼辨别为人工合成图像,虽然重建图像与原图像在特征相似度上接近,却很难通过人脸活

体检测程序。由于 GAN 在图像生成方面有着优秀的表现^[10-12],因此被作为图像重建的另一种思路^[13-14],Dong 等人^[15]使用 StyleGAN2 模型^[16]生成人脸,利用随机向量生成图片并将图片输入到特征提取网络中得到特征向量,以此生成大量的 StyleGAN2 输入到深层特征的向量对,然后使用这些数据训练从特征向量映射到 StyleGAN2 输入向量的一个全连接网络,从而生成重建人脸。该方法虽然使用了 StyleGAN2 作为图像生成器,能够生成高分辨率无伪影的图像,但是由于训练数据分布的不平衡性且映射向量维度较高,使得生成图片在视觉效果上与真实图像的相似性较差,且从二者提取的深层特征之间的距离也较大。

文章提出了一种基于深层特征重建高分辨率人脸图像的算法,使用 StyleGAN2 作为人脸图像生成器重建人脸,将人脸深层特征嵌入 StyleGAN2 的 W^+ 空间,通过梯度下降法优化 w^+ 向量后生成重建人脸图像,并搭建了一个模拟人脸识别系统的框架,利用重建的人脸图像访问该模拟框架,从而评估重建人脸的有效性。

1 基于深层特征嵌入的高清人脸图像重建

1.1 嵌入空间的选择

生成对抗网络在图像生成方面有着丰富的应用,DCGAN 在原始 GAN 图像加入 CNN 卷积层进行图像转换,但生成图像的分辨率较低,为 256×256 ,且生成人脸图像质量不稳定,很容易生成面部扭曲的人脸图像,因此不适合作为重建人脸的图像生成器。PGGAN 中加入了渐进增大特征图的方式训练生成图像,虽然生成人脸图像的质量较高,分辨率能够达到 1024×1024 ,但是由于采用逐级生成图片,没有对各层分别增添控制,导致其控制生成图像的特征的能力非常有限,输入向量在某一维度上的细微变化就可能生成完全不一样的图片,使

得图像重建算法难以收敛。而StyleGAN2不仅生成图像的质量高且分辨率为 $1\ 024 \times 1\ 024$,并对生成网络中的每一级特征图都增添控制,输入向量的微小变化也不会造成生成图像的明显变化,因此文章采用在Flickr-Faces-High-Quality(FFHQ)高清人脸数据集上预训练好的StyleGAN2模型作为重建人脸的图像生成器,根据已知的深层特征向量重建出与真实人脸图像在视觉上相似的人脸图像。图1为StyleGAN2生成网络结构图,初始向量 $z \in Z$, Z 为符合标准正态分布的 1×512 维向量空间, z 向量经过8层全连接层后得到解纠缠的 1×512 维 W 向量空间, w 向量($w \in W$)经过复制并通过lerp函数插值后组成一个 18×512 维的 W^+ 向量空间,这里lerp函数的定义如下:

$$\text{lerp}(\bar{w}, w^+, \text{weight}) = \bar{w} + (w^+ - \bar{w}) \times \text{weight} \quad (1)$$

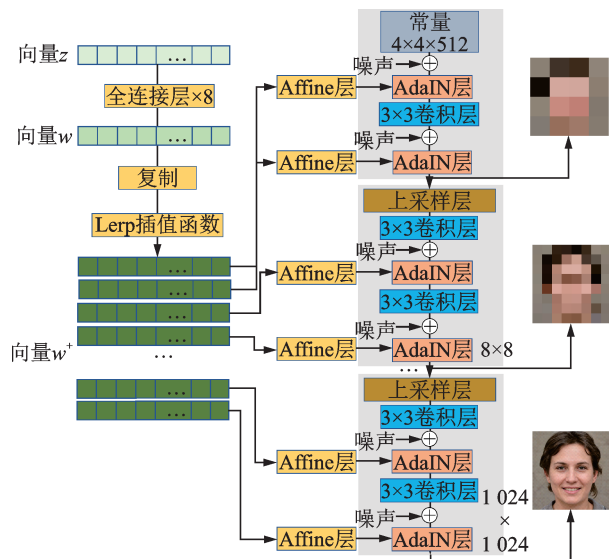


图1 StyleGAN2生成网络结构图

Fig.1 Generator network structure of StyleGAN2

其中 \bar{w} 为StyleGAN2训练过程中记录下的平均 w 向量,weight为区间 $[0, 1]$ 之间的实数,用于控制其返回值受 \bar{w} 影响的偏移程度,该函数能够约束 w^+ 向量中的每个维度向量的分布,从而防止生成扭曲失真的图像。随后 w^+ 向量的每个维度经过Affine层后得到生成图像的风格样式 $y = (y_s, y_b)$ 并输入AdaIN层生成图像的特征图,从而控制生成图像,AdaIN层操作定义为:

$$\text{AdaIN}(x, y) = y_s \left(\frac{x - \mu(x)}{\sigma(x)} \right) + y_b \quad (2)$$

其中 x 为特征图,不同的AdaIN层生成的特征图用于控制生成图像的不同特征,顶层分辨率最低,主要控制生成图像中人脸的姿态,面部朝向等,而分

辨率越高的层,对图像的控制越细微。在StyleGAN2生成图像过程中,先生成低分辨率图像,然后再逐渐生成高分辨率图像,从而实现对生成图像特征的逐层控制。

由上可知,StyleGAN2生成网络中有多个可供选择的嵌入空间,分别为 Z 、 W 、 W^+ ,文中考虑将深层特征向量嵌入到某一潜在空间中以生成重建人脸图像。实验对比了将深层特征分别嵌入这三个空间后的重建图像,实验结果如图2所示,由图可知 Z 空间对最终生成图像的风格样式最不敏感,重建出的人脸图像与原图差异较大, W 空间虽然能够重建出图像的部分特征,但在面部轮廓,肤色以及发型上的表现较差,而 W^+ 空间重建的图像在五官以及发型上的重建效果明显优于前二者。根据StyleGAN2生成网络的结构可知, w^+ 向量直接影响了生成图像的特征图以控制最终生成图像的风格样式,因此文章选择将深层特征向量嵌入到StyleGAN2的 W^+ 向量空间中直接跳过 Z 空间以及 W 空间,通过优化 w^+ 向量生成重建人脸。



图2 不同嵌入空间的重建图像

Fig.2 Reconstructed images in different embedding spaces

1.2 特征嵌入方法

将向量或者图像嵌入到StyleGAN2的潜在空间 $[17]$ 有两种方法:一种是训练一个将输入数据映射到潜在空间的神经网络模型,通过训练好的StyleGAN2网络生成一定数量的图像,并记录下生成图像的对应StyleGAN2潜在空间向量,将图像或图像经过另一网络后得到的向量作为输入,对应StyleGAN2潜在空间向量作为输出,由此训练一个从向量或图像到潜在空间的神经网络模型。另一种方法是针对某一随机或特定的输入,通过优化算法和约

束条件去优化输入。第一种方法通常存在泛化问题,训练模型只是在当前训练数据集上效果较好。因此,文章选择第二种方法将深层特征向量嵌入到 StyleGAN2 的潜在空间中,该方法虽然较第一种需要花费更多时间,但泛化能力更强,更稳定。对于优化问题来说,选择一个合适的初始点十分重要,一个恰当的初始点能够大大减少算法的收敛步数,而距离最优解较远的初始点会导致算法收敛步数增加,甚至导致算法难以收敛并更容易陷入局部最优解。为了寻找到一个合适的初始向量,文中通过预训练的 StyleGAN2 模型生成十万张图像,并记录下这十万张图像对应的 w^+ 向量,将这十万个 w^+ 向量的平均值作为文中优化问题的初始值,并与符合均匀分布的随机初始值进行比较,实验结果如图 3 所示。以平均 w^+ 向量作为初始值进行优化的重建图像在视觉上与真实图像更为相似,而以随机初始值进行优化的重建图像由于其初始 w^+ 向量分布与原始 W^+ 潜在空间的分布相差较大,不仅在视觉上相似度更低,而且还会生成过度平滑且不真实的图像。

图 4 为重建算法流程框架以及评估流程框架,文章将已知深层特征与 StyleGAN2 生成图像的深层特征之间的均方误差 (Mean Square Error, MSE) 作为最小优化的约束条件:

$$Loss = MSE(v', V) = \sum_{i=1}^n \frac{1}{n} (v'_i - V_i)^2 \quad (3)$$

其中 n 为深层特征向量的维度, V 为已知深层特征



图 3 不同初始化 w^+ 向量重建结果图

Fig.3 Reconstruction results for different initialized w^+

向量, v' 为重建图像的深层特征向量,优化后的 w^+ 向量可表示如下:

$$w^{+*} = \operatorname{argmin}_{w^+} MSE(f(g(w^+)), V) \quad (4)$$

其中 $f(\bullet)$ 为特征提取网络, $g(\bullet)$ 为 StyleGAN2 的生成网络,文章跳过 StyleGAN2 的 Z 和 W 空间,直接使用梯度下降法优化 w^+ 向量, w^+ 向量的优化过程为:

$$w^{+*} = w^+ - \eta H\left(\frac{dMSE(v', V)}{dw^+}\right) \quad (5)$$

其中 $H(\bullet)$ 为随机梯度下降优化器,文章的优化目标是找到使得已知深层特征与 StyleGAN2 生成图像的深层特征之间的均方误差最小的 w^+ 向量,并将该 w^+ 向量输入到 StyleGAN2 生成网络中生成重建图像。

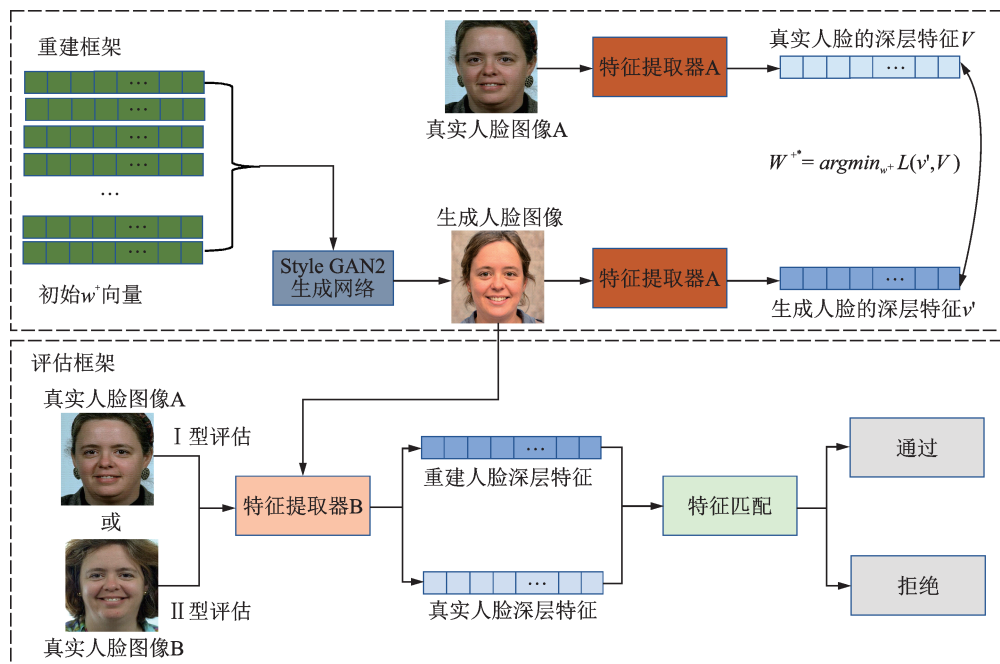


图 4 重建算法流程框架以及评估流程框架图

Fig.4 Framework for reconstruction algorithm and evaluation

2 实验与分析

2.1 实验设置

文章实验配置环境如下:Linux操作系统,CPU为Intel Core i9-10900X、GPU为NVIDIA GeForce GTX 3080(10 GB显存),内存128 GB。基于TensorFlow2.0深度学习框架,通过GPU加速平台CUDA11.1实现文中方法。

文中采用在FFHQ高清人脸数据集上进行预训练后的StyleGAN2模型作为人脸图像生成器,并使用ArcFace网络^[18]作为特征提取网络,利用两种不同主干网络的ArcFace模型来评估重建人脸图像,这两个主干网络分别为Xception^[19]和Inception^[20]。此外,采用了Labeled Faces in the Wild(LFW)、ColorFeret以及FFHQ三个人脸数据集作为特征提取的图片对象,从而方便对比重建图像与真实图像在视觉上的相似度。对于优化算法,文中采用学习率为0.1的随机梯度下降法,迭代停止条件为Loss在连续50步内没有下降或总步数达到2000步。

2.2 评价指标

将数据集中的图像通过特征提取网络提取到的特征向量作为已知深层特征,将初始化的 w^+ 向量输入StyleGAN2中生成人脸图像,再将生成的图像经过特征提取网络提取深层特征,随后使用随机梯度下降法迭代优化 w^+ 向量,最终生成重建人脸图像。这里使用到的真实人脸图像只是为了方便在视觉上对比真实人脸和重建人脸的相似度,实际上文中方法只需根据人脸的深层特征重建图像,而非真实人脸图像。为了评估文中的方法,生成的人脸图像会输入到搭建的评估框架中,计算与真实人脸图像特征的相似度,在人脸识别测试中,识别算法预测结果会分成四种情况,第一种为同人图像对被正确预测为同一个人(True Positive, TP);第二种为不同人图像对被预测为不同一人(True Negative, TN);第三种为不同人图像对被错误预测为同一人(False Positive, FP);最后一种为同人图像对被错误预测为不同人(False Negative, FN)。以此文中计算在不同错误接受率(False Accept Rate, FAR)下的正确接受率(True Accept Rate, TAR), FAR在数值上与假阳性率(False Positive Rate, FPR)相等,可表示为:

$$FAR = FPR = \frac{FP}{FP + TN} \quad (6)$$

TAR在数值上与真阳性率(True Positive Rate, TPR)相等,可表示为:

$$TAR = TPR = \frac{TP}{TP + FN} \quad (7)$$

2.3 结果评估

分别在LFW和ColorFeret数据集上选择了150个人的脸图像以及FFHQ高清人脸数据集上50个人的脸图像进行重建。图5为在LFW、ColorFeret以及FFHQ数据集上的部分重建人脸图像结果。可以看到,文中部分重建的图像与对应真实人脸图像在脸部轮廓以及五官等方面都较为相似,个别图像在发型以及发色上也具有一定的相似性。图6为文中重建图像与其他现有方法重建图像的视觉对比,文献6方法重建图像存在严重的伪影以及色彩失真;而文献8方法虽然利用特征提取器内部层信息使得重建图像在视觉上相似度较高,但其背景部分仍然存在一定的伪影;文献9方法由于在重建过程中采用了反卷积操作,造成了混叠效应,在图像全局上都有严重的棋盘格伪影;文献15方法虽然在图像生成的质量方面较高,但由于采用神经网络映射方法且映射向量维度较高,重建图像在视觉相似度上表现较差。

2.3.1 不同FAR下的TAR表现

使用了Mai等人^[9]提出的两种评估方法,I型评估为将重建图像与用于生成深度特征向量的真实人脸图像进行特征提取后比较,II型评估为将重建图像与对应同一个人的不同人脸图像进行特征提

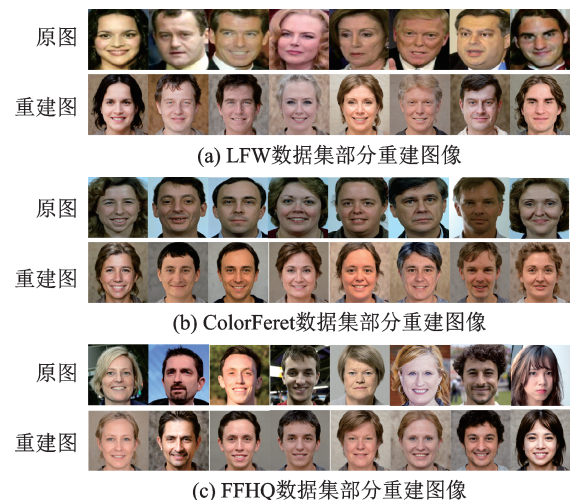


图5 部分重建人脸图像

Fig.5 Partially reconstructed face images



图 6 与现有方法重建图像的视觉对比

Fig.6 Visual comparison with existing methods

取后比较,分别评估在不同 FAR 的情况下重建人脸

的 TAR 。

在重建框架与评估框架都使用以 Inception 为主干网络的 Arcface 特征提取网络时的 TAR 结果如表 1 所示。对于 LFW 数据集,在 FAR 等于 0.1% 的情况下 I 型评估和 II 型评估的正确接受率分别达到 100% 和 96.04%;对于 ColorFeret 数据集,在 FAR 等于 0.1% 的情况下 I 型评估和 II 型评估的正确接受率都达到了 100%,这表明文章的方法在重建框架与评估框架使用同一个特征提取网络时具有有效性。

表 1 重建和评估框架特征提取器同为 Inception 时的 TAR Tab.1 TAR when feature extractor for reconstruction and evaluation frameworks both are Inception

数据集	$FAR/(%)$	阈值	正常 $TAR/(%)$	I 型评估 $TAR/(%)$	II 型评估 $TAR/(%)$
LFW	0	0.531 6	79.77	100	64.00
	0.1	0.438 2	98.12	100	96.04
	1	0.400 0	99.42	100	99.17
	10	0.352 2	99.86	100	99.72
ColorFeret	0	0.528 0	99.46	100	98.92
	0.1	0.465 5	100	100	100
	1	0.423 9	100	100	100
	10	0.371 2	100	100	100

在重建框架以及评估框架分别使用以 Inception、Xception 为主干网络的 Arcface 特征提取网络时的 TAR 结果如表 2 所示,相较于表 1 的结果有所下降,这是因为在不同特征提取网络下,相同图像提取到的特征向量并不相同。文中重建图像是在 Inception 特征提取网络下,根据 StyleGAN2 生成图像的深层特征和已知深层特征优化得到的,因此在以其他特征提取网络为评估框架下的特征相似度表现差于在以相同特征提取

网络为评估框架下的特征相似度。但是由于重建图像在视觉上仍与对应原图像具有相似性,对于 LFW 数据集,在 FAR 等于 0.1% 的情况下 I 型评估和 II 型评估的正确接受率仍然能够分别达到 74% 和 48.71%;对于 ColorFeret 数据集,在 FAR 等于 0.1% 的情况下 I 型评估和 II 型评估的正确接受率也能分别达到 70% 和 58.06%,这说明了文中方法在不同特征提取网络下的评估框架里仍具有一定有效性。

表 2 重建框架和评估框架特征提取器分别为 Inception、Xception 时的 TAR Tab.2 TAR when feature extractor for reconstruction and evaluation frameworks are Inception and Xception respectively

数据集	$FAR/(%)$	阈值	正常 $TAR/(%)$	I 型评估 $TAR/(%)$	II 型评估 $TAR/(%)$
LFW	0	0.493 3	91.90	40	23.11
	0.1	0.446 3	97.76	74	48.71
	1	0.401 8	99.41	92	76.24
	10	0.352 2	99.90	100	96.96
ColorFeret	0	0.507 8	99.64	54	43.55
	0.1	0.481 9	99.82	70	58.06
	1	0.430 0	100	90	86.56
	10	0.373 8	100	100	99.46

文中重建方法与当前比较有代表性的两种重建方法进行了比较,分别是基于神经网络映射到 StyleGAN2 输入向量^[15](Mapping)以及 NbNet^[9],LFW 数

据集的重建图像在 FAR 等于 0.1% 情况下的 TAR 比较结果如表 3 所示。Mapping 方法虽然采用 StyleGAN2 重建图像,图像分辨率能够达到 1024×1024 ,

但是由于 Mapping 方法采用神经网络映射的方法受训练数据分布的不平衡性且映射向量维度较高的影响,使得神经网络很难成功映射到 StyleGAN2 的输入向量,从而在 TAR 上表现不佳。NbNet 方法重建的图像在 TAR 上表现效果较好,但由于采用了反卷积,

造成重建图像具有伪影,且分辨率较低。文中方法重建图像由于同样采用 StyleGAN2 作为图像生成器,因此分辨率与 Mapping 方法相同,但 TAR 较 Mapping 方法有巨大提升,且与 NbNet 方法相比,文中方法重建图像在 II 型评估 TAR 上也有一定提升。

表3 LFW数据集上不同方法在0.1% FAR下的TAR
Tab.3 TAR of different methods on LFW dataset at 0.1% FAR

方法	特征提取器	分辨率	I 型评估 TAR/(%)	II 型评估 TAR/(%)
本文	ArcFace	1 024×1 024	100.00	96.04
Mapping	ArcFace	1 024×1 024	1.42	0.46
NbNet	FaceNet	160×160	95.20	53.91

2.3.2 归一化特征相似度分布情况

图7和图8对重建人脸图像在 LFW 和 ColorFeret 数据集上 I 型评估以及 II 型评估的特征归一化相似度分布进行可视化,图中横坐标表示图像间的特征相似度,横坐标越大表示特征相似度越高,纵坐标表示对应特征相似度的概率密度。黄色部分表示不同人图像对之间的特征相似度分布情况,蓝色部分表示同人图像对之间的特征相似度分布情况,红色部分表示文章重建的图像与对应人图像之间的特征相似度分布情况。从图中可以看到,真实人脸图像下同一个人(蓝色部分)和不同人(红色部分)之间的归一化特征相似度的分布相差较大,这表明基于该

特征提取网络的人脸识别系统能够有效判断两张图像之间的相似性。图7为重建框架与评估框架均为同一个特征提取网络 Inception 时的特征分布情况,其中图7(a)、图7(b)分别为 LFW 数据集在 I 型评估和 II 型评估下的特征分布情况,图7(c)、图7(d)分别为 ColorFeret 数据集在 I 型评估和 II 型评估下的特征分布情况。文章方法重建的人脸图片在 I 型评估中的特征相似度分布比同一个人的特征相似度分布更靠右,意味着文章的重建图像与用于生成深度特征向量的真实人脸图像之间的平均特征相似度大于同人图像对之间的特征相似度,这是因为文中的重建图像是基于对应图像的深层特征优化而来的。但

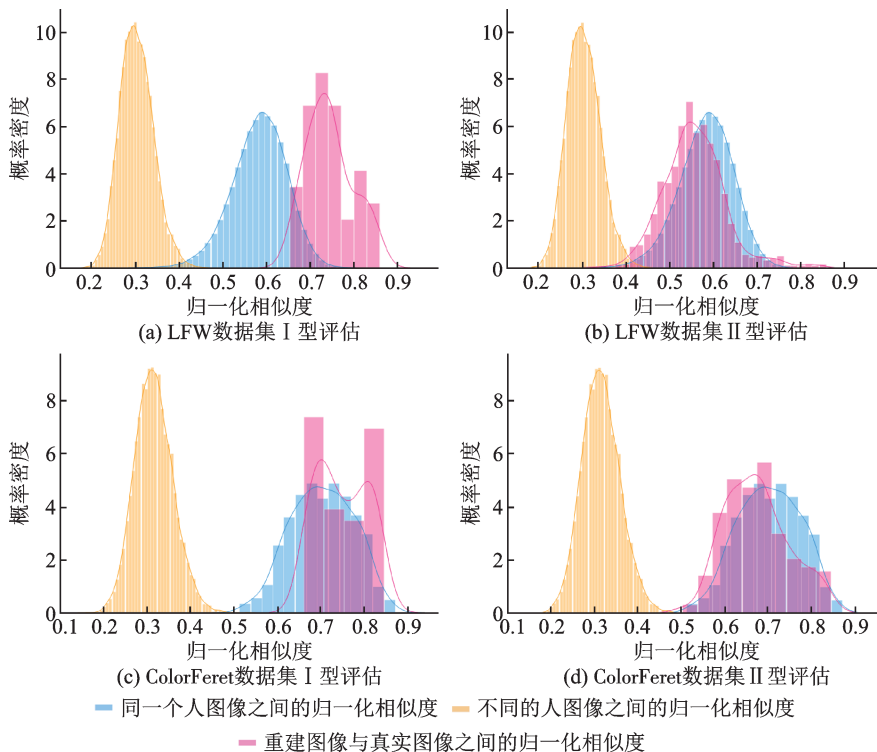


图7 重建框架和评估框架同用 Inception 特征提取器下的特征相似度分布情况

Fig.7 Distribution of feature similarity when feature extractor for reconstruction and evaluation frameworks both are Inception

在实际情况下,人脸识别系统的工作更类似于Ⅱ型评估,因为在人脸识别系统中,进行识别的人脸图像与人脸信息录入系统时的图像不为同一张图像,而文中重建图像在Ⅱ型评估中的特征相似度与同人图像对的特征相似度分布存在大量重叠部分,说明文章重建图像与原图像对应人的其他图像仍保持较大的特征相似度,这表明文中的方法在特征归一化相似度分布方面具有有效性。

图8表示重建框架与评估框架分别使用Inception和Xception特征提取网络时的特征分布情况,其中图8(a)、图8(b)分别为LFW数据集在Ⅰ型评估和Ⅱ型评估下的特征分布情况,图8(c)、图8(d)

分别为ColorFeret数据集在Ⅰ型评估和Ⅱ型评估下的特征分布情况。可以看到,文中的重建图像在两个数据集中的Ⅰ型评估和Ⅱ型评估下的特征分布介于黄色区域与蓝色区域之间,这是因为同一图像在不同特征提取器下提取到的特征向量不同导致重建图像与对应原图之间的特征相似度下降。但是红色部分都与蓝色区域存在一部分重叠,说明文中的重建图像在不同特征提取网络的评估下的特征相似度整体上高于不同人对图像之间的特征相似度,且仍有部分能够达到同人图像对之间的特征相似度水平,进一步说明了文中的方法在不同特征提取网络下的评估框架里仍具有一定有效性。

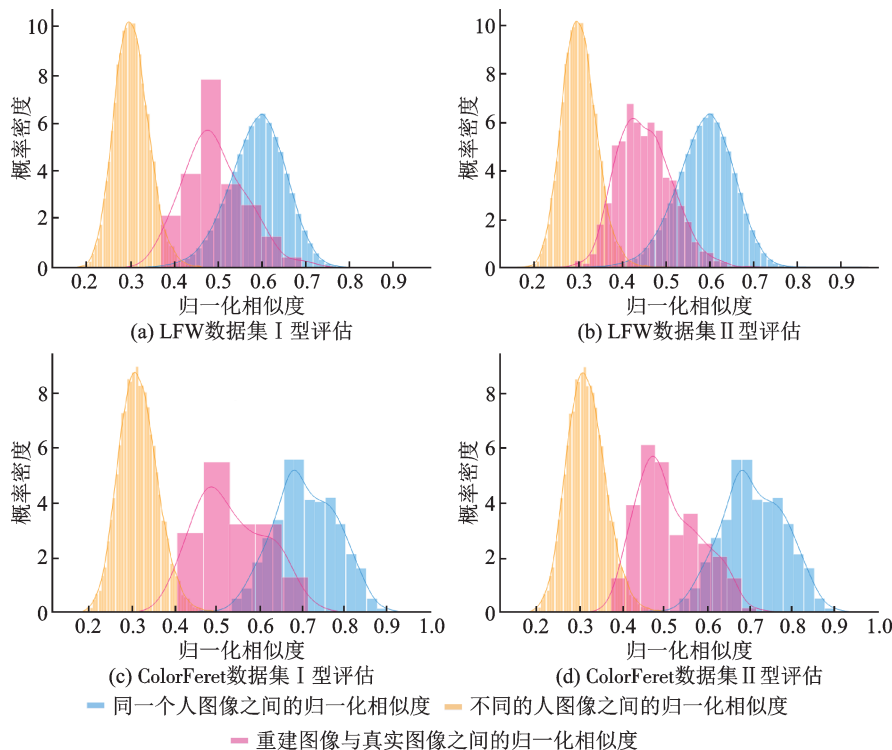


图8 重建框架和评估框架分别用Inception和Xception特征提取网络下的特征相似度分布情况

Fig.8 Distribution of feature similarity when feature extractor for reconstruction and evaluation frameworks are Inception and Xception respectively

2.3.3 重建图像在活体检测程序上的通过率

文中重建的图像在视觉上与原图有着较高的相似性,如图9。与NbNet方法重建的图像相比,文章重建的图像的分辨率更高,而NbNet方法的重建图像分辨率仅为 160×160 ,且存在大面积的伪影,Mapping方法重建图像虽然分辨率和文中重建图像相同,但是在视觉相似性上表现差。文章选取了市面上常见的两个活体检测的应用程序编程接口(Application Programming Interface, API)来对重建图像进行活体检测,如表4所示。文中重建的图像



图9 文章方法与NbNet方法以及其他方法的重建结果图
Fig.9 Reconstruction results of this paper and other methods

在活体检测API 1中的通过率为88.67%,在活体检测API 2中的通过率为74.67%,而NbNet方法重建

表4 重建图像在活体检测API上的通过率

Tab.4 The pass rate of reconstructed images on living detect APIs

活体检测 API	重建图像方法	通过图像数	未通过图像数	通过率/(%)
API 1	NbNet	0	150	0
	Mapping	129	21	86
	本文	133	17	88.67
API 2	NbNet	70	80	46.67
	Mapping	110	40	73.33
	本文	112	38	74.67

的图像由于存在伪影,很容易分辨出为人工合成图像,在活体检测 API 1 的通过率仅为 0,在活体检测 API 2 上的通过率为 46.67%, Mapping 方法和文中一样选用 StyleGAN2 进行图像重建,因此重建图像在两个活体检测 API 上的表现与文中相近,分别为 86% 和 73.33%。

3 结 论

文章提出一种基于深层特征嵌入的高分辨率人脸图像重建算法,分析了 StyleGAN2 中三个潜在嵌入空间, Z 、 W 、 W^+ , 最终选择将深层特征嵌入到 W^+ 空间中,并利用 StyleGAN2 生成十万张图像的平均 w^+ 向量作为初始向量,将已知深层特征与 StyleGAN2 生成图像特征之间的距离作为最小约束条件,利用梯度下降进行优化,最终生成分辨率为 1024×1024 的重建人脸图像。在同一特征提取网络下, LFW 和 ColorFeret 数据集的重建图像在 FAR 为 0.1% 时 II 型评估 TAR 分别为 96.04% 和 100%, 并且在两种不同的活体检测程序下的通过率分别达到 88.67% 和 74.67%, 满足高分辨率要求的同时也保证其与真实人脸图像之间的高特征相似度。

参 考 文 献

- [1] Wang M, Deng W H. Deep face recognition: A survey [J]. Neurocomputing, 2021, 429: 215-244.
- [2] 姚立平, 潘中良. 基于改进的 HOG 和 LBP 算法的人脸识别方法研究 [J]. 光电子技术, 2020, 40(2): 114.
- [3] Meng Q, Zhao S C, Huang Z D, et al. Magface: A universal representation for face recognition and quality assessment [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Kuala Lumpur, Malaysia, 2021: 14225-14234.
- [4] Adler A. Sample images can be independently restored from face recognition templates [C]. CCECE 2003-Canadian Conference on Electrical and Computer Engineering. Toward a Caring

and Humane Technology (Cat. No. 03CH37436), Montreal, Canada, 2003, 2: 1163-1166.

- [5] Mignon A, Jurie F. Reconstructing faces from their signatures using RBF regression [C]. British Machine Vision Conference 2013, Bristol, UK:BMVA, 2013: 103.1-103.12.
- [6] Zhmoginov A, Sandler M. Inverting face embeddings with convolutional neural networks [R/OL]. <https://arxiv.org/pdf/1606.04189.pdf>:(2016-07-07)[2022-09-25].
- [7] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 815-823.
- [8] Cole F, Belanger D, Krishnan D, et al. Synthesizing normalized faces from facial identity features [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, IEEE, 2017: 3703-3712.
- [9] Mai G C, Cao K, Yuen P C, et al. On the reconstruction of face images from deep face templates [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(5): 1188-1202.
- [10] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 4401-4410.
- [11] 高文超, 任圣博, 田 驰, 等. 多层次生成对抗网络的动画头像生成方法研究 [J]. 计算机工程与应用, 2022, 58(9): 230-237.
- [12] 张希仁, 张恒静, 罗亚玲, 等. 基于生成对抗网络的计算光学散射成像研究 [J]. 光电子技术, 2021, 41(3): 185.
- [13] 李 诚, 张 羽, 黄初华. 改进的生成对抗网络图像超分辨率重建 [J]. 计算机工程与应用, 2020, 56(4): 191-196.
- [14] 刘郭琦, 刘进锋, 朱东辉. 基于生成对抗网络的图像超分辨率重建算法 [J]. 液晶与显示, 2021, 36(12): 1720-1727.
- [15] Dong X B, Zhe J, Guo Z H, et al. Towards generating high definition face images from deep templates [C]. 2021 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 2021: 1-11.
- [16] Karras T, Laine S, Aittala M, et al. Analyzing and improving the image quality of stylegan [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 8110-8119.
- [17] Abdal R, Qin Y P, Wonka P. Image2stylegan: How to embed images into the style gan latent space? [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, Long Beach, USA, 2019: 4432-4441.
- [18] Deng J K, Guo J, Xue N N, et al. Arcface: Additive angular margin loss for deep face recognition [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 4690-4699.
- [19] Chollet F. Xception: Deep learning with depthwise separable convolutions [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 1251-1258.
- [20] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 1-9.