

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2025)04-1155-15

论文引用格式: Hu J P and Zhou Y. 2025. 3D stylized portrait synthesis and structured face modeling. Journal of Image and Graphics, 30(4): 1155-1169(胡佳平, 周漾. 2025. 三维风格化人脸生成与结构化建模. 中国图象图形学报, 30(4): 1155-1169)[DOI:10.11834/jig.240380]

三维风格化人脸生成与结构化建模

胡佳平, 周漾*

深圳大学计算机与软件学院, 深圳 518060

摘要: 目的 现有的三维人脸风格化方法难以生成较大相机姿态的人脸视图, 或止步于生成多角度的人脸视图, 而非结构化的三维网格模型。鉴于此, 提出一种基于样例的三维人脸风格化与结构化建模方法。**方法** 该方法不仅能够合成新的全相机视角下的风格化人脸视图, 还具有生成结构化的人脸三维网格模型的能力, 即包括人脸三维网格以及对应的纹理贴图。具体来说, 提出了一个两阶段的结构化三维风格人脸生成框架, 主要包括三维感知人脸生成器域迁移、基于多视图约束的人脸纹理优化两个步骤。首先, 利用二维人脸风格化数据增强策略微调三维感知生成器, 然后通过一个视图对齐策略对齐基于隐式神经场的渲染视图以及基于三维网格的渲染视图, 再利用多视图约束的梯度回传优化人脸模型的纹理贴图, 最后通过融合多幅纹理贴图得到最终的纹理贴图。**结果** 本文三维感知人脸生成结果FID(Fr chet inception distance)指标优于对比实验中基于扩散模型的方法, 并能在多种风格域中生成整洁高质量的结构化人脸模型。**结论** 该方法能够有效构建高质量的结构化三维风格人脸模型, 生成高质量的全角度风格人脸视图与纹理贴图。此外, 显式构建的结构化人脸模型能够更为便捷地用于三维人脸相关下游任务。**关键词:** 视觉内容生成; 三维风格人脸生成; 三维感知生成; 域迁移; 纹理优化

3D stylized portrait synthesis and structured face modeling

Hu Jiaping, Zhou Yang*

Collage of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China

Abstract: Objective Facial image stylization and 3D face modeling are important tasks in the fields of computer graphics and vision, with significant applications in virtual reality and social media, including popular technologies such as virtual live streaming, virtual imaging, and digital avatars. This work focuses on 3D facial stylization and generation to produce novel and stylized facial views from a given real face image and a style reference. The novel views can be rendered at corresponding angles by inputting camera poses in 3D space. Meanwhile, these views need to maintain good 3D multi-view consistency while expressing the exaggerated geometry and colors characteristic of the given artistic style reference. Facial stylization and facial modeling are prominent tasks in the fields of computer graphics and vision, with significant applications in virtual reality and social media. These applications include popular technologies such as virtual live streaming, virtual imaging, and digital humans. This paper addresses the task of 3D facial stylization generation, aiming to produce facial views from corresponding angles by inputting camera poses in 3D space. These views need to maintain good 3D multi-view consistency while expressing the exaggerated geometry and colors characteristic of artistic styles. Existing methods for 3D facial generation can be broadly categorized into two types: those based on 3D deformable models and those based on

收稿日期: 2024-07-04; 修回日期: 2024-09-05; 预印本日期: 2024-09-12

* 通信作者: 周漾 zhouyangvcc@szu.edu.cn

基金项目: 广东省高等学校创新团队项目(2022KCXTD025); 深圳大学教学改革重点项目(JG2024018)

Supported by: DEGP Innovation Team (2022KCXTD025); Shenzhen University Teaching Reform Key Program (JG2024018)

implicit neural representations. Methods based on 3D deformable models often struggle to express non-facial components such as hairstyles and glasses, which severely limits the quality of the generated results. On the other hand, methods based on implicit neural representations, while capable of achieving good generation results, tend to produce severely distorted facial views under large camera poses, such as side profiles. Additionally, the results of implicit methods typically include only facial geometry and multi-view facial views, making it difficult to integrate them with mature rendering pipelines. This limitation hinders their application in practical scenarios. Consequently, both existing 3D facial generation methods face challenges in producing high-quality 3D stylized facial models with good structured modeling, i. e., 3D facial meshes and topologically complete texture maps. **Method** To address the shortcomings of existing methods, this paper proposes a novel approach for 3D stylized facial generation and structured modeling. The goal of this paper is to train a 3D aware stylized facial generator within the style domain of a specified artistic facial sample. This generator should be capable of producing high-quality 3D facial views from any angle in the specified style, including large-pose side profiles and back views. Furthermore, based on multi-view facial data, the generator should produce structured 3D facial models, including facial geometric mesh models and corresponding texture maps. To achieve this, the paper proposes a two-stage method for 3D stylized facial generation and structured modeling. The method comprises two main steps: 3D aware facial generator domain transfer and multi-view constrained facial texture optimization. In the first stage, the paper utilizes 2D facial stylization prior methods to perform data augmentation on artistic style samples, generating a small-scale artistic style facial dataset. Subsequently, the camera poses and facial masks of the facial images in this dataset are extracted sequentially. The annotated stylized facial dataset is then used to fine-tune a 3D aware generator in the natural style domain. The fine-tuned 3D aware generator can generate high-quality multi-view facial views and 3D mesh models. The focus of the second stage is to optimize facial textures using multi-view images from a set of directions. The paper first performs smoothing and UV unwrapping on the facial mesh. To align the volumetric rendered facial views with the differentiable rendered facial views for pixel-level loss optimization of facial textures, the paper proposes a simple and effective facial view alignment strategy based on mask affine transformation. Finally, multi-view facial supervision is used to optimize facial textures, and the final facial texture map is obtained through texture fusion. **Result** To demonstrate the superiority of the proposed method, the paper compares the two-stage 3D facial generation and structured modeling method with existing advanced baseline methods. This comparison illustrates the quality of 3D aware stylized facial generation and the effectiveness of structured facial mesh generation. Additionally, to demonstrate the effectiveness of each stage component of the proposed method, the paper includes ablation studies for the key components of the method. The use studies illustrate the correctness of the proposed method. Qualitative and quantitative experiments show that the proposed method can effectively construct high-quality structured 3D stylized facial models and generate high-quality stylized facial views. Moreover, the explicitly modeled structured facial models can be more conveniently applied to downstream tasks related to 3D faces. The results indicate that the proposed method not only achieves superior performance in generating stylized facial views but also ensures the structural integrity and applicability of the generated 3D facial models in practical scenarios. **Conclusion** In general, this work presents a comprehensive approach to 3D stylized facial generation and structured modeling, addressing the limitations of existing methods. By leveraging a two-stage process that includes domain transfer and multi-view constrained texture optimization, the proposed method achieves high-quality results in both facial view generation and structured modeling. The effectiveness of the method is demonstrated through extensive experiments, highlighting its potential for practical applications in virtual reality, social media, and other related fields.

Key words: visual content synthesis; 3D portrait stylization; 3D aware generation; domain adaptation; texture optimization

0 引言

三维风格化人脸生成是指根据二维人脸图像，

合成具有指定风格的三维人脸模型。在工业应用方面,三维人脸风格化或者三维人脸内容生成可以为多种场景提供高质量、高效率以及高自由度的数字化服务,例如视频聊天、视频会议、虚拟现实、增强现

实、娱乐游戏等,这些场景对于提升人们的数字交流体验、满足个性化需求、增强社会互动和娱乐性等方面都具有重要作用。

现有基于 StyleGAN (style-based generative adversarial network) (Karras 等, 2019) 或扩散模型 (diffusion model) (Ho 等, 2020) 的二维人脸风格化方法 (Yang 等, 2022; Su 等, 2021; Chong 和 Forsyth, 2022; Ruiz 等, 2023; Song 等, 2021; Xue 等, 2022) 已经相当成熟,但是这些 2D 方法受限于缺乏三维空间理解,因此具有严重的多视角不一致。隐式神经表达 (Park 等, 2019; Mildenhall 等, 2021) 的发展极大地改善了三维人脸相关问题中的多视角不一致问题,但这些方法受限于算力,只能产生分辨率较低的人脸结果,抑或只能在一个小视角范围内生成高质量人脸,难以满足风格化任务的几何多样性。

一部分方法 (Han 等, 2023; Ye 等, 2023; Zheng 等, 2022; Wu 等, 2018) 与结构化的人脸三维可变形模型 (Blanz 和 Vetter, 1999) 结合,依靠构建好的人脸三维网格数据集,可以得到统一、高度结构化的三维人脸模型,这些模型不仅包括高质量的人脸几何,还具备对应的纹理贴图。统一的拓扑结构为三维人脸模型的应用提供显著的便利。然而受限于统计模型的表达能力,这些人脸生成方法往往只能在一个小视角范围内生成高质量人脸,难以刻画包括后脑勺、头部发型等非面部的人脸组件信息,因此不能满足高质量人脸风格化的需求。

随着可微分渲染 (Kato 等, 2018) 与隐式神经表达 (Park 等, 2019; Mildenhall 等, 2021) 的发展,基于三维感知的生成模型取得了快速的发展,这些方法 (OrEl 等, 2022; Chan 等, 2022; Sun 等, 2022a, b; An 等, 2022; Abdal 等, 2023) 使用隐式神经表达记录场景的空间信息,再与成熟的二维图像生成方法 StyleGAN (Karras 等, 2019) 结合,因此能够产生高质量的多视角一致人脸图像,在新视角合成问题上取得了令人惊叹的进展。同时得益于其在二维图像生成问题中强大的表达能力,这些方法往往只需要极少的数据集。然而,这些三维感知的生成模型存在一些局限性,以人脸生成为例,这些方法在正对摄像机视角下的生成表现较好,但在观察角度增大时可能出现失真的问题。此外,相较于结构化的三维模型,这些生成模型的结果更难以直接集成到传统渲染和建模管线中。

针对这些问题,本文提出了一种全新的基于样例的风格化三维人脸生成方法,只需给定单幅自然人脸图像以及特定的风格样例,如迪士尼卡通、讽刺画、日本漫画风格等,无需任何额外的结构化数据,如人脸关键点或者深度图,也无需任何三维数据监督,即可合成高质量的风格化人脸模型。相比于近年的三维感知生成方法仅能生成人脸侧前方的人脸视图,本文方法能够合成全角度的风格化三维人脸视图,这些视图包括头部整体信息,以及头发、眼镜等非人脸组件信息。此外,本文生成结果具有高质量的三维网格以及相应的纹理贴图。相较于现有的三维人脸生成方法,该方法能够更为便捷地与工业界较为成熟的三维渲染技术对接,用于虚拟数字人等 3D 内容生成领域。

1 相关工作

三维人脸内容生成问题是计算机视觉与计算机图形学领域较为热门的话题。为了生成多视角一致的人脸图像,许多方法将传统 2D 图像神经网络与三维空间感知的技术相结合,其中主要分为基于传统三维可变形网格模型 (3D morphable model, 3DMM) 的方法 (Blanz 和 Vetter, 1999) 与基于隐式神经表达 (implicit neural representation, INR) 的方法 (Park 等, 2019; Mildenhall 等, 2021)。Han 等人 (2023) 提出一个显式的风格化三维人脸生成框架,该框架基于统一拓扑的三维人脸数据集,将生成过程拆分为结构化的三维网格几何变形以及纹理贴图优化两部分,分别对应三维人脸的几何风格与纹理风格。具体而言,其将传统的三维可变形模型与微分渲染结合,利用 FaceWareHouse (Cao 等, 2014) 从单幅人脸图像得到结构化的自然人脸模型,再依次利用几何变形与微分渲染逐步优化生成三维人脸的几何与纹理贴图。Ye 等人 (2023) 则通过构建一个 3D 讽刺画风格人脸的大规模数据集,在 3D 风格化空间中构建一个三维风格人脸的主成分分析 (principal component analysis, PCA) 模型。随后,通过建立 PCA 人脸表达与三维讽刺画空间之间的对应关系,由输入人脸照片得到一组人脸表达系数,即可重建一个完整的三维讽刺画风格人脸。受 Park 等人 (2019) 的启发,ImFace (Zheng 等, 2022) 将三维可变形模型与隐式神经表达有向距离场 (signed distance field, SDF)

(Oleynikova等, 2016)结合, 提出一种在隐式连续空间控制人脸属性的方法。然而, 这些人脸生成方法的性能受限于3DMM统计模型属性的表达能力, 因此在生成质量上难以令人满意。总体而言, 这些方法都无法处理非人脸组件, 如发型、耳朵等区域的信息。

随着隐式神经表达的发展, 大量工作将其与二维的图像生成方法StyleGAN(Karras等, 2019)结合, 这种策略已被证明能够得到高质量的人脸, 还能保证2D方法难以实现的三维多视角一致性问题。这些方法从非结构化的二维人脸图像中构建一个隐式神经场, 利用体渲染方法得到一个较低分辨率的人脸视图以及特征图, 再通过基于StyleGAN2(Karras等, 2020)的上采样网络得到高质量的人脸视图。其中, Chan等人(2022)提出的三平面特征能够更高效地表达3D场景, 因此具有生成高质量的视图一致图像的能力。此外, 得益于Marching Cube(Lorensen和Cline, 1987)方法, 隐式神经表达的人脸三维场景能够便捷地转换为三维网格模型。在三维风格化人脸生成方法中, Abdal等人(2023)和Song等人(2023)利用迁移学习的策略微调三维感知生成器, 前者使用更多生成对抗网络(generative adversarial network, GAN)正则项来稳定风格域风格化人脸的生成, 后者则采用一个风格先验来指导三维感知生成器的迁移学习, 但两者都局限于视角受限的人脸数据集, 因此只能产生角度有限的人脸视图, 且生成人脸图像内容与背景紧紧地耦合在一起。An等人(2023)提出一个自然人脸风格域的全角度三维人脸生成方法, 通过在三维感知生成器的训练数据集中添加大量带有标注的大姿态人脸图像数据, 其具有生成全角度人脸视图的能力。虽然能够生成高质量的全角度人脸视图, 但是在人脸生成问题中, Song等人(2023)依然止步于生成多视角的人脸图像, 而非显式的结构化人脸模型。考虑到其在生成大姿态人脸视图任务中的杰出表现, 本文在其模型基础上构建三维感知风格化人脸生成器。

随着扩散模型迅速发展, 大量方法利用扩散模型来生成人脸相关内容, 得益于扩散模型卓越的生成能力, 这些方法能够产生高质量的人脸视图。Wang等人(2023)将隐式神经表达与扩散模型结合, 将神经辐射场表示为多个2D特征映射, 并将这些映射展开到单个2D特征平面中, 再利用三维感知的扩

散模型生成全角度的人脸头部。得益于扩散模型的强大能力, 这种方法产生令人惊叹的结果。然而, 这种方法不仅需要大规模的训练, 且生成效率严重受限于扩散模型的计算消耗。同时, 使用文本提示深度神经网络不如图像风格样例直观便捷。此外, 缺少对于隐空间的控制使得模型对风格化程度的控制不如基于StyleGAN(Karras等, 2019)的方法便捷有效。现有方法止步于合成新视角图像, 而本文方法进一步显式地构建结构化的风格人脸模型, 包括三维网格模型以及对应的纹理贴图, 这比体渲染多视角人脸结果更方便于相关下游任务, 嵌入到成熟的网格渲染流程中。

2 本文方法

本文目标是使用单幅自然人脸输入生成具有特定风格的三维人脸模型, 同时得到高质量的纹理贴图, 即

$$(\mathbf{M}, \mathbf{T}) = F(\mathbf{I}_s, \mathbf{z}, \mathbf{c}) \quad (1)$$

式中, $\mathbf{M}$ 代表生成的风格化人脸的三维网格, $\mathbf{T}$ 代表网格对应的纹理贴图, $\mathbf{I}_s$ 代表给定风格样例图像, $\mathbf{z}$ 、 $\mathbf{c}$ 分别代表采样噪声以及相机姿态。

本文方法总体流程图如图1所示, 主要包括三维感知人脸生成器域迁移、基于多视图约束的人脸纹理优化两个阶段。第1阶段采取一种风格化人脸数据增强策略, 用于在特定风格域微调三维感知生成器。在微调生成器之后, 高维三维感知空间的深度特征即对应一个风格化的三维人脸。由此即可得到高质量的人脸视图以及人脸网格。在本文方法第2阶段中, 基于第1阶段高质量全角度人脸视图, 进一步优化人脸网格纹理贴图。通过在微分渲染过程中约束对齐后的人脸视图, 最终得到人脸三维网格的纹理贴图, 也即一个完整的结构化三维人脸模型, 包括三维网格 $\mathbf{M}$ 以及纹理贴图 $\mathbf{T}$ 。

考虑到自然人脸域现有三维感知生成方法较为成熟, 基于体渲染能够生成高质量、全姿态自然人脸视图。因此, 本文方法第1阶段目标是将其迁移到给定风格域。其难点在于: 1)人脸风格化通常只有少量的风格参考样例, 模型极容易过拟合; 2)在微调三维感知生成器过程中, 风格样例的背景容易对生成结果的几何产生干扰。对此, 本文采取了一个数据增强的策略, 先利用二维人脸风格化方法构建

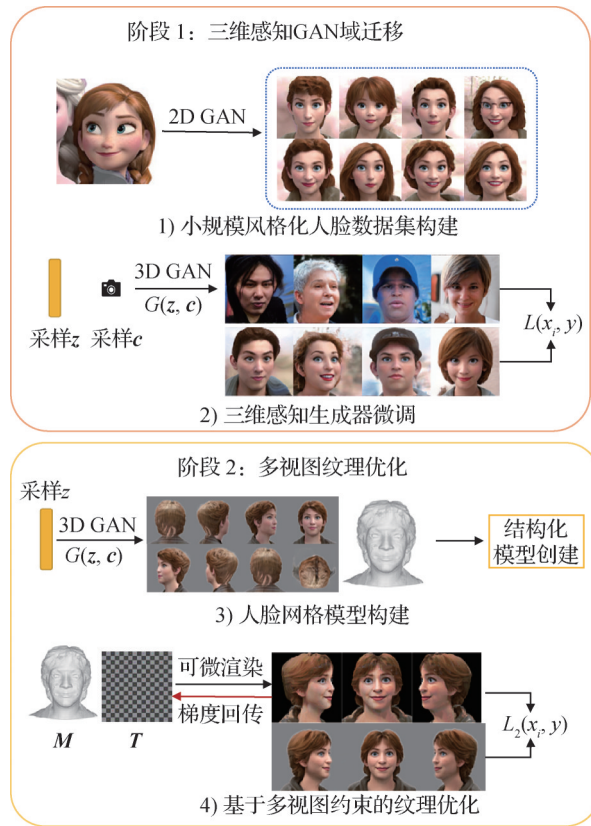


图1 本文方法整体流程

Fig. 1 Overview of our work

一个较小的风格数据集,再对这些图像去除背景色,提取人脸掩膜,并利用一个标准的人脸姿态估计方法对这个小型人脸数据集进行相机姿态标注,以便进行后续的三维感知生成器的微调。然而,直接使用风格化数据集微调三维感知生成器会严重打乱原模型的空间理解,这是由于风格化人脸的人脸分布要远比自然人脸杂乱、多样,训练过程中难以自主学习三维空间感知。为此,本文还设计了一种半监督的微调策略,这种策略能够将三维感知生成器的生成域逐步迁移到样例指定的风格域,同时最大程度地保留人脸几何结构变形的稳定。

第2阶段目标是利用第1阶段体渲染多视角结果获得高质量的三维网格,并优化其对应的纹理贴图。由于体渲染生成器的上采样结构,低维空间的深度特征对应一幅空间中的高质量视图,这种生成方法带来了无可避免的多视角歧义。此外,现有方法并未对齐自然网格以及体渲染表达,仅依靠多视图体渲染结果进行纹理贴图会导致严重的瑕疵和歧义。为此,本文提出了一种视图对齐策略,能够最大程度地利用体渲染生成的高分辨率视图,再通过可

微分渲染优化网格的纹理细节,利用对齐后的目标视图与微分渲染的人脸视图计算像素级损失,得到每个视角下的人脸纹理优化结果,再通过一个简洁有效的融合策略,即可得到最终的纹理贴图。

本文2.1至2.3节对应第1阶段,2.4至2.5节对应第2阶段。

2.1 小样本风格人脸数据增强

为了解决风格化人脸数据集难以收集,以及对风格化人脸数据的标注费时费力的问题,本文提出一种简单可行的针对三维风格化人脸的数据增强策略。受到Yang等人(2022)提出的二维人脸图像风格化方法的人脸图像内容生成质量启发,利用其进行给定风格样例的数据增强。具体而言,如图2中阶段1的三维感知生成器域迁移流程图所示,首先利用PSP(pixel2style2pixel)(Richardson等,2021)人脸图像编码器将风格样例映射到StyleGAN的特征空间,然后固定其对应的编码作为二维人脸风格化网络的外部风格编码,用于控制生成人脸图像的艺术风格。随后,随机采样一组内部风格编码,这些风格编码主要影响人脸生成器的人脸身份特征,由此能够得到一组小规模的风格图像数据集,该数据集的人脸图像与输入样例保持着相同的艺术风格,同时具有各异的身份特征。

这种策略的另一个好处是,在这个过程中能够轻易得到对应自然人脸域中的人脸图像,只需要在风格编码融合过程中,将主要作用于融入样例风格的外部风格编码的权重置为0。随后,通过自然人脸图像提取对应的人脸相机姿态参数,即可解决风格化数据集难以标注的问题。利用二维先验方法得到一组风格人脸图像后,需要对该小规模人脸数据集进行数据预处理,包括人脸对齐、相机姿态预测和人脸掩膜提取。

如图3所示,人脸对齐主要包括两个步骤:首先需要对人脸图像进行关键点检测,预测关键点的坐标,关键点的信息包括人脸的眼睛、眉毛、鼻子、嘴巴以及人脸轮廓。之后需要根据预测的关键点坐标对人脸进行对齐,即通过图像变换方法将关键点对准到一个统一的固定位置上。本文采用预训练的MTCNN(multi-task convolutional neural network)(Xiang和Zhu,2017)网络预测人脸图像的68个关键点,再进行人脸对齐处理。由于MTCNN由自然人脸图像数据集训练得到,因此仅适用于自然人脸图像,

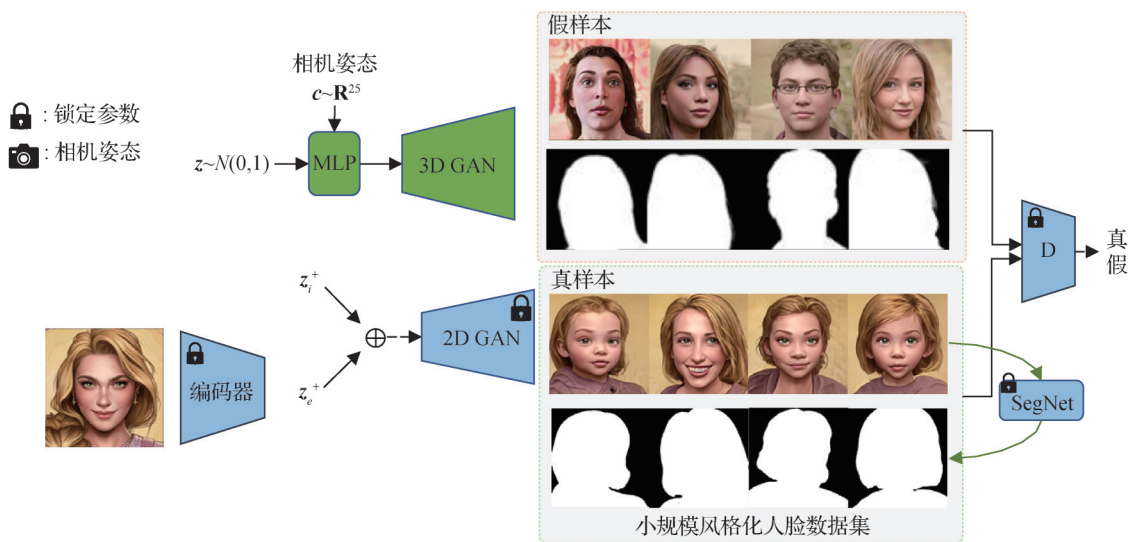


图2 阶段1:三维感知GAN域迁移流程示意图

Fig. 2 Pipeline of stage one: 3D-aware GAN domain adaptation

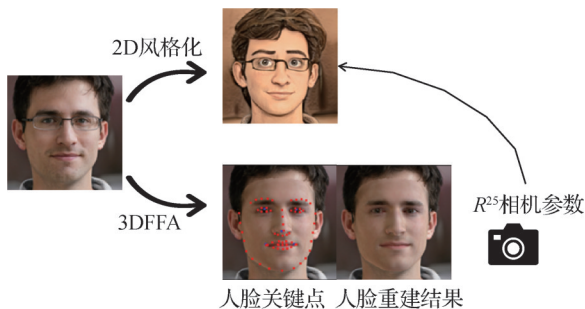


图3 人脸数据集预处理示意图

Fig. 3 Preprocessing dataset of frontal face

在此本文使用数据集中对应自然风格的人脸数据进行人脸对齐,以便后续的相机姿态预测。

为了从二维人脸图像中训练三维感知生成器,需要从人脸图像中预测相机姿态。在具体实现中,人脸的相机姿态为一个25维的向量 c ,包括一个展平的 4×4 相机外参矩阵以及一个展平的 3×3 的相机内参矩阵。为了标注人脸三维信息,本文利用现有人脸姿态预测模型3DFFA(3D dense face alignment)(Guo等,2020)预测人脸图像的三维相机姿态。

最后,为了从人脸图像中分离出人脸区域,消除背景对于三维感知生成器的影响,需要提取人脸图像的掩膜,本文使用预训练的SegNet(segment network)(Badrinarayanan等,2017)模型提取人脸图像的掩膜。对于一幅给定的人脸图像,SegNet能够从中预测出人脸各个区域的类别,输出结果为一个19通道的 512×512 像素的二分类图像,在本文中只使用返回中的第2个通道,即人脸整体区域。如此,即

可获得带有相机视角标注以及人脸掩膜的风格化人脸数据集,构建的风格化人脸数据集将用于自然人脸域的三维感知生成器域迁移到对应风格域。

2.2 风格域三维感知人脸生成器微调

对三维感知生成器进行微调的目的在于将生成器的自然人脸生成域迁移到由一个2D数据集定义的特定风格。在本文任务中,该数据集即为由风格样例图像数据增强获得的小规模风格数据集。本文将该微调任务定义为:给定一个先验的三维感知生成器,即Panohead中的自然人脸域 T_{target} 的生成器 G ,任务目标为生成一个风格域 T_{target} 的风格化人脸生成器 G' ,在训练前后,需要保持人脸在不同风格域的身份特征统一,同时 G' 需要尽可能地表现样例人脸图像的风格,包括人脸色彩、三维几何和语义等信息。三维感知生成器微调的流程图如图2所示。其中 z_e^+ 、 z_i^+ 分别意为控制注入艺术风格的外部风格编码、控制人脸身份特征内部风格编码。通过固定 z_e^+ 并随机采样 z_i^+ 即可生成大量相同艺术风格的人脸图像。

本文使用标准的对抗损失对生成器进行微调,同时为了保证训练的稳定性,本文延续StyleGAN2(Karras等,2022)的训练策略,利用标准的正则项约束梯度,以防止训练过程中的梯度爆炸与过拟合问题。微调过程中学习目标为

$$\argmin_{G,D} D(G'(z \sim N(0,1), c), c) + [1 - D(I', c')] \quad (2)$$

式中, $G'(z, c)$ 代表三维感知生成器由随机采样的噪声 z 在相机视角 c 下的生成结果, I' 代表由数据增强

策略构建的小规模数据集中的样本, c' 则代表与人脸视图相对应的相机姿态参数。

2.3 输入人脸图像三维感知风格化人脸生成

在将三维感知生成器迁移到目标风格域后, 下一步需要在隐空间中采样以得到风格化的三维人脸。然而由用户输入的自然人脸图像得到对应的三维风格人脸依然是一项具有挑战性的任务。需要将真实的输入图像投影到微调后的三维感知生成器所在的生成空间, 以得到对应人脸身份的风格化三维人脸模型。

为此, 本文采用关键枢纽调整 (pivotal tuning inversion, PTI) (Roich 等, 2023) 中的 GAN 反演策略。首先, 使用基于优化的方法找到一个隐编码 z , 使得风格化图像与真实的输入图像最为接近, 使用逐像素的 L2 损失以及图像级的 LPIPS (learned perceptual image patch similarity) (Zhang 等, 2018) 损失来量化这种相似性。随后, 为了进一步提高重建人脸的质量, 固定优化得到的隐编码 z , 然后通过反向传播优化生成器的参数, 使得生成器结果尽可能接近原图像。由此, 即可得到对应给定输入人脸的三维感知风格化人脸结果。在实际程序实现中, 优化 z 与优化生成器 G' 的迭代次数分别设置为 200 和 200。

2.4 结构化三维人脸模型构建

为了显式地优化人脸的纹理, 需要首先创建结构化的三维人脸模型, 以便将其与传统的图形学渲染管线结合应用。利用多视角人脸图像进行纹理优化的前提是整洁平滑的人脸网格模型以及对应的纹理贴图。得益于 Marching Cube 方法, 通过提取神经辐射场 (Heusel 等, 2017) 中的零势面能够获得三维

人脸的几何。然而, 由于这种方式得到的网格往往带有一定的瑕疵, 尤其对于风格化人脸这种几何更为多样化的场景, 因此, 需要首先对由神经辐射场得到的人脸网格模型进行平滑处理。

如图4所示, 首先由 Marching Cube 方法 (也称为等值面提取方法) 从神经辐射场中提取三维网格, 随后将其顶点数统一规范化到 30 000, 高分辨率会导致渲染需要的时长大大增加。随后, 根据三维人脸网格的连接关系, 三维网格被划分为多个区块, 而顶点数目最多的区块被保留。这种方法能够有效地消除三维网格模型的孤岛以及瑕疵, 由于人脸网格的统一性, 这种策略对于三维网格中人脸相关内容影响甚微。

由于曲面的平滑程度提高能够有效提升网格纹理映射的质量, 因此, 在对三维人脸网格进行结构化创建前, 对由 Marching Cube 方法得到的粗人脸网格进行平滑处理。平滑处理三维网格的目的在于约束表面的法线尽可能连续, 这能够大大提高纹理映射以及后续纹理优化的质量。为使三维网格的表面尽可能平滑, 三维网格首先被采样为一定顶点数量的点云, 再由泊松点云重建得到三维网格, 以缓解人脸模型表面复杂区域法线突变的程度。经过处理后的网格更为整洁, 这对于利用微分渲染的视图反向优化纹理贴图发挥重要作用。此外, 统一处理后三维人脸网格被规范化为统一顶点数量, 这种策略能够降低微分渲染反向传播过程的计算消耗, 提高后续利用多视图约束进行纹理优化的效率。

由于现有的自动化纹理展开方法, 如 xatlas (<https://github.com/jpcy/xatlas>), 难以获得整洁平滑

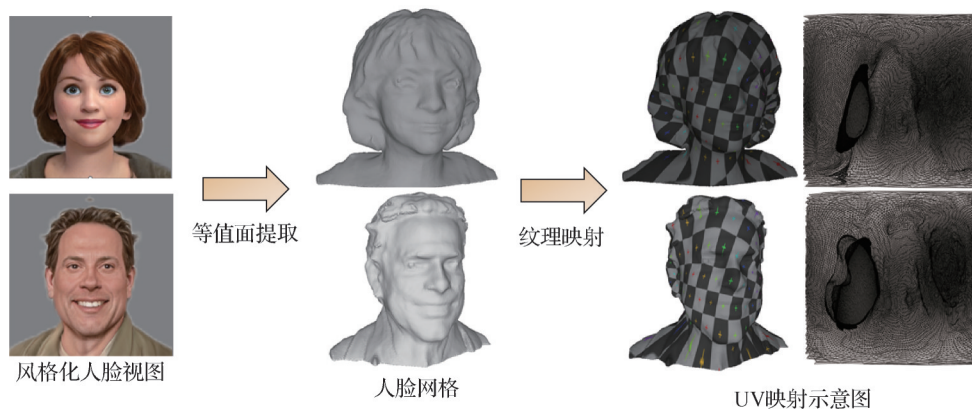


图4 人脸模型结构化处理示意图

Fig. 4 Structural processing of face mesh

的UV展开结果,综合考虑纹理展开效率与人类头部模型的特殊性,本文采用blender(<https://www.blender.org/>)的球形纹理展开方式进行UV映射。UV展开后的拓扑结构如图4所示,其中UV展开图中的重影区域为人脸模型的底部,由于多视图纹理优化过程中,用于监督的人脸视图来自于水平视角以及头顶视角,因此重影部分不会在纹理优化过程中产生影响。

2.5 基于多视图约束的纹理优化

由于体渲染与基于网格的可微渲染的成像方式不同,造成人脸视图与视图不对应,因此,如果直接使用体渲染的人脸视图通过神经渲染反向优化纹理贴图,语义上的不对应会导致网格模型的纹理难以由人脸视图提取正确的像素值。

为了解决这一问题,本文提出了一种基于掩膜对齐的多视角纹理优化方法。如图5所示。具体而言,首先采样一定数量的相机视角,在水平方向的环形轨道内,采样间距为 60° 的6个相机视角,随后在体渲染中得到对应的高质量人脸视图以及相对应的语义掩膜。为了避免未对齐的人脸视图带来的纹理瑕疵,通过一种基于仿射变换的策略将语义掩膜对齐。这种策略的数学形式表达式为

$$\argmin \sum_{i=1}^n \left\| mask_{nr}^i, A \left(mask_{vr}^i, \begin{pmatrix} R_1 & R_2 & t_1 \\ R_3 & R_4 & t_2 \end{pmatrix} \right) \right\|_2 \quad (3)$$

式中,仿射变换矩阵 (R, t) 为二维图像仿射变换的系数, n 代表相机视角的个数, A 代表仿射变换操作,而

$mask_{vr}^i, mask_{nr}^i$ 分别代表来自体渲染、可微渲染的视图掩膜。对于 $mask_{nr}^i$,可通过将网格对应的纹理贴图像素值置为 $(1, 1, 1)$ 得到。得益于体渲染与可微渲染的视角方向相同,因此视角的错位主要包含平移方向的错位,以及少量的边缘畸变,因此,利用这种优化策略能够高效得到两个语义掩膜之间的仿射变换系数。随后,将优化得到的仿射变换矩阵引用到对应视图中,即可将其与可微渲染的人脸视图语义掩膜对齐。

为了由多视角的人脸视图得到人脸模型的纹理贴图,本文基于可微分渲染实现纹理优化。如图5流程所示,体渲染得到的人脸视图经过仿射变换后,即可为可微渲染模块提供约束,反向优化纹理贴图。在具体程序实现中,本文选择了环绕人脸环形轨道的6个水平视角,以及为了提取头顶的信息的俯视角视图。

这些视图可表示为 $I_{target}^i = A^i \times I_{vr}^i, i = (0, 1, \dots, 6)$,其中 i 代表相机姿态 v^i 的编号, A^i 代表该视图对应的仿射变换参数, I_{vr}^i 则代表由体渲染得到的对应人脸视图。随后,由可微渲染模型渲染出对应相机视角的视图,即 $I_{nr}^i = R(m, t, c^i), i = (0, 1, 2, \dots, 6)$,表示在相机姿态 c^i 下,可微渲染器 R 由人脸网格 m 以及纹理贴图 t 得到的渲染视图。纹理优化的目标为,使得渲染模块在各个视角下的生成结果 I_{nr}^i 尽可能接近 I_{target}^i 。为了最大程度保留 I_{target}^i 的信息,纹理优化环节只使用逐像素的 L_2 损失约束纹理图像,即

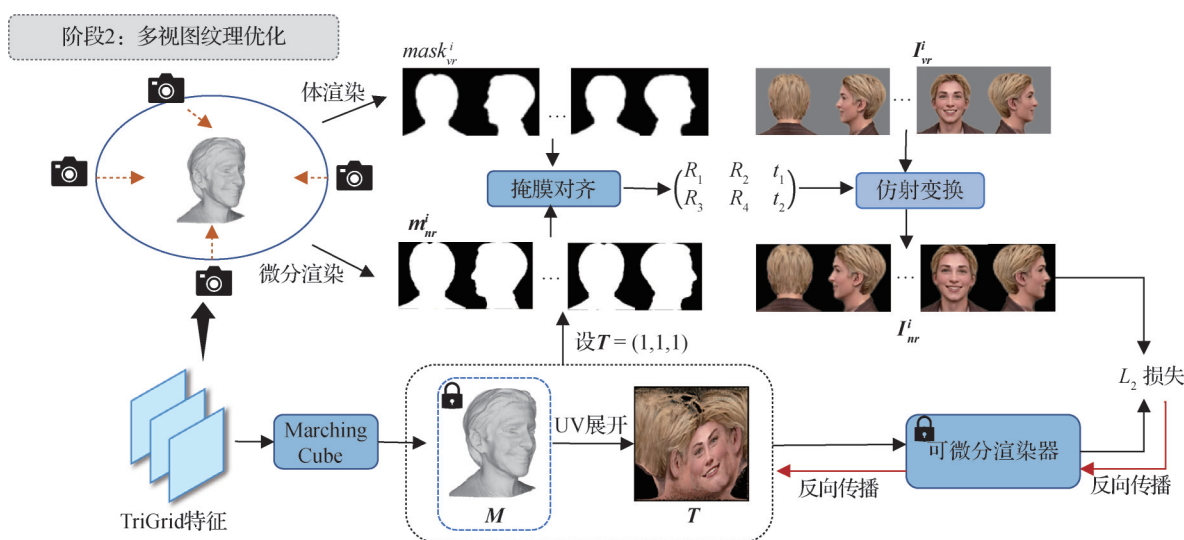


图5 阶段2:多视图纹理优化流程示意图

Fig. 5 Pipeline of stage two: texture optimization with multi-view constraints

$$t_i = \arg \min_i \|I_{nr}^i, I_{target}^i\|_2 \quad i = 0, 1, \dots, 6 \quad (4)$$

式中, t_i 为在相机视角 c^i 下的纹理贴图优化结果。在每个人脸视图优化过程前, 纹理贴图被初始化为 $(0, 0, 0)$, 使得优化过程模型独立地学习每一个相机视角下的视图。相比于同时优化这些人脸视图, 独立优化能够避免纹理贴图由于多视角歧义造成模糊或鬼影。在获取所有视角对应的纹理优化结果后, 最后融合各个视角对应的纹理贴图。

由于多视角视图彼此之间存在歧义, 即视图中相同的部位, 例如眼睛, 虽然对应着相同的 UV 坐标, 却有着不同的颜色值。为了消除这种多视角歧义, 同时尽量保留优化结果中质量较好的区域, 如图 6 所示, 本文采用一种简洁有效的融合策略: 首先, 对每幅纹理结果进行腐蚀操作, 这种策略能够有效消除非纹理区域的异常值。随后由正对相机视角开始, 逐步由其他相机视角下的纹理贴图补全其中的零值。这种融合策略能够有效保留优化结果中的细节, 同时消除多视角歧义。

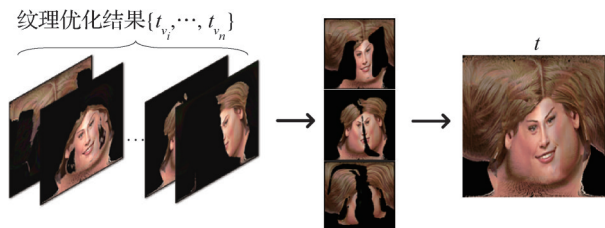


图6 纹理融合示意图

Fig. 6 Pipeline of blending textures

3 实验与分析

3.1 实验配置与评价指标

为了实现三维感知生成器由真实人脸域到艺术风格人脸域的迁移, 本文使用 Yang 等人(2022)的二维先验方法进行数据增强, 通过在二维风格化人脸的特征空间中采样, 构建一个小规模的人脸数据集, 小规模数据集人脸图像的数量设置为 100。二维人脸风格化先验提供了多种风格样例, 如皮克斯动画、迪士尼卡通、讽刺画等。选择其中一些风格鲜明的风格人脸图像作为三维人脸风格化的风格样例图像。本文方法采用的自然风格人脸图像来源于 FFHQ (flickr-faces-high-quality) 数据集, 其中包括 70 000 幅高质量的 $1\,024 \times 1\,024$ 像素的人脸正脸图

像。受限于 GPU (graphic processing unit) 显存的容量, 本文的数据集以及生成人脸图像的分辨率统一为 512×512 像素。在实验中, 对于每个特定风格, 利用数据增强得到的风格化人脸图像, 对自然人脸风格的三维感知生成器进行微调。使用 Nvidia 官方的 StyleGAN2 网络 (Karras 等, 2020) 以及 panohead (Song 等, 2023) 的三维感知生成器。在训练过程中, 沿用 StyleGAN2 的损失函数和权重设置, 网络训练过程使用的批次大小为 4, 学习率设置为 0.002。微调过程在 500 代时基本收敛, 训练图像数量为 16 000, 花费时间大概为 10 min。微调完成后, 对于每个给定的 z 编码以及相机视角 c , 网络仅需 0.106 s 即可生成一幅人脸视图。

为了定量比较, 本文使用 IS (inception score) (Salimans 等, 2016) 与 FID (Fréchet inception distance) (Heusel 等, 2017) 作为定量评估模型生成质量的评价指标。其中 IS 指标使用图像分类模型来评估生成图像的分布质量。IS 值越大, 代表生成图像的清晰度与多样性越好。由于 IS 指标不能反映模型过拟合的问题, 因此本文另外使用 FID 指标来评估模型生成人脸的多样性。FID 直接使用预训练的深度神经网络计算图像的深度特征, 以此衡量生成结果与真实图像之间的距离。FID 指标越小, 代表模型生成图像的分布与真实图像的分布越接近, 即模型生成质量越好。

3.2 本文方法结果

从 Yang 等人(2022)提出的风格人脸数据集中选择了部分具有代表性的风格人脸图像作为样例, 展示本文方法的三维感知人脸生成结果, 结果如图 7 所示, 其中每一行的左侧代表指定的风格样例人脸图像, 右侧为由域迁移后的三维感知生成器生成的该风格下三维感知多视图人脸图像, 即环绕人脸的环形轨道中各个角度的人脸视图, 以及该人脸对应的三维网格。通过观察可以发现, 本文方法的三维感知风格化人脸生成结果保留了给定风格样例的艺术特征, 能够生成高质量的全相机角度人脸视图, 这些视图具有高度的多视角一致性的同时, 能够表达包括后脑、发型、眼镜等非人脸组件的信息。此外, 人脸的几何结构整洁自然, 这是后续纹理优化的基础。为了说明本文方法在单幅输入人脸图像场景下的生成能力, 图 8 展示了域迁移后的三维感知生成器由单幅自然环境下的人脸图像输入得到的多视



图 7 基于风格样例的三维感知风格化人脸生成结果

Fig. 7 Results of 3D aware stylized face generation based on style exemplars

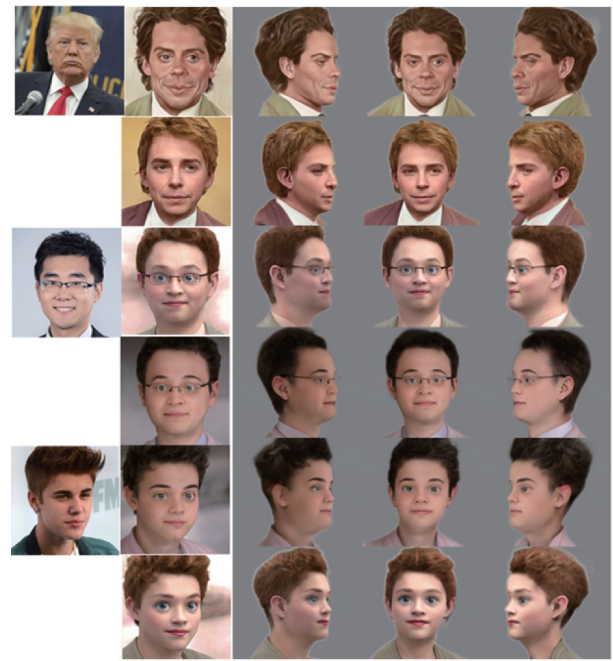


图 8 输入人脸图像的风格化三维人脸生成结果

Fig. 8 Results of 3D aware facial mesh generation with input portraits

角人脸视图。可以看出,域迁移后的生成器依然能够保持较高的身份一致性,同时生成高质量的人脸视图。图9所示为结构化的三维人脸生成结果,也即

显式的三维人脸网格以及UV贴图。可以看出,本文纹理优化方法能够充分利用三维感知生成器生成的多视角视图中的信息,融合得到整洁合理的UV贴图。

3.3 方法对比

为与现有的三维人脸风格化方法进行对比分析,分别选择基于三维可变形模型的传统方法3DPS(3D portraits synthesis)(Han等,2021),以及三维感知人脸生成方法EG3D(efficient geometry-aware 3D generative adversarial network)(Chan等,2022)进行对比。3DPS提出的人脸风格化方法不便于大量生成三维人脸,因此本文在统一输入的情况下,比较两种方法生成的三维风格化人脸几何以及视图质量。EG3D是基于隐式表达的人脸生成方法中的代表,与本文方法类似,可通过在隐空间采样隐编码生成三维风格化人脸。

为了保证对比的公平性,采用相同的风格数据集以及迁移学习策略,在统一的实验环境下对本文方法与EG3D(Chan等,2022)方法进行了训练和测试,并计算生成结果的FID、IS评价指标,如表1所示,为了计算微调后的三维感知人脸生成器的FID与IS指标,选定一种代表性卡通风格,随机生成50 000幅图像用于计算生成图像分布,需要特别提



图9 结构化三维风格人脸生成结果

Fig. 9 Results of structural 3D portrait stylization

到的是,生成图像中视角分布情况与训练数据集保持一致。这些生成图像用于与目标分布即原样本分布进行计算,再与基准方法进行对比。定性的实验结果可视化如图 10 所示。可以看出,与 3DPS 这种基于隐式神经表达的三维人脸生成方法相比,本文方法能够极大提高生成人脸的视觉效果,高质量地刻画发型等头部区域的信息。与 EG3D 相比,本文

方法显著提高了大姿态下人脸视图的质量,且能够产生真实、整洁的人脸几何。

为了说明本文方法生成结构化三维人脸模型的优越性,进一步对比了现有成熟的三维形状生成方法。Meshy (<https://www.meshy.ai/>) 是一个成熟的三维网格模型生成商业软件,能够通过提示词或参考图像生成具有纹理贴图的三维模型。在对比实验

表 1 三维感知人脸生成域迁移结果定量对比
Table 1 Quantitative comparison of the domain adaptation results of 3D aware face generation

方法	IS $\uparrow$	FID $\downarrow$
EG3D based	3.53	6.42
本文	5.04	5.37

注:加粗字体表示各列最优结果。“ $\uparrow$ ”表示值越高越好,“ $\downarrow$ ”表示值越低越好。

中,相同的多视角人脸图像被输入到生成模型中,用于优化具有相同纹理空间的模型贴图。如图 11 所示,Meshy 生成的人脸形状在纹理复杂处具有难以忽视的瑕疵,而本文方法能够产生整洁正确的纹理贴图,即使在眼部等纹理复杂处依然能够保持较好的细节。

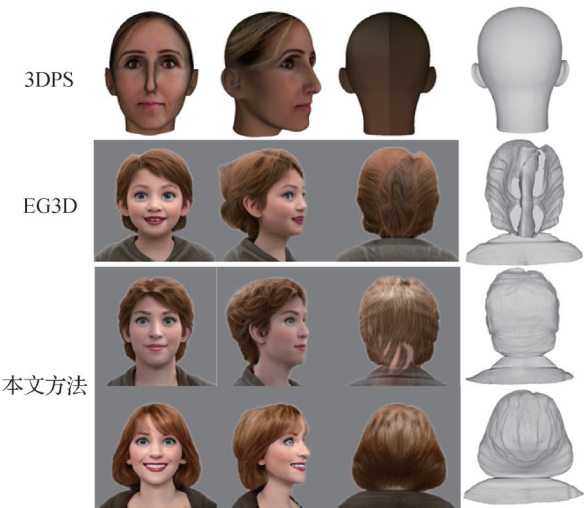


图 10 不同三维人脸生成方法的效果对比
Fig. 10 Qualitative comparison of different 3D face generation methods



图 11 不同结构化人脸生成方法的效果对比
Fig. 11 Qualitative comparison of different structural face model generation methods

综合以上实验表明,本文方法不仅能够生成高质量的多视角人脸视图,视图对齐与纹理融合策略能够显著地提高人脸纹理优化的结果,最终产生对应给定样例风格的高质量结构化三维人脸模型。定性和定量对比表明,本文的三维人脸风格化和结构化建模质量均优于现有方法。另外,本文方法能够显式地建模三维风格人脸的结构化网格模型,这使得本文方法能够更为便捷地融入到现有成熟的网格渲染管线中。

时间效率上,对于单幅给定的风格样例,本文方法第 1 阶段首先构建一个小规模的数据集,然后将其用于微调三维感知生成器,总体花费约 12 min,即可在 0.4 s 内生成任意三维感知风格化人脸视图。受限于基于三维网格可微渲染的效率,本文方法第 2 阶段,即基于多视图的纹理优化,总体需要约 15 min,最终优化得到结构化的三维人脸网格。

3.4 消融实验

为了验证小样本风格人脸数据增强策略与纹理优化中视图对齐策略的有效性,本文设计了消融实验。如图 12 与表 2 所示,当小样本数据集规模 (N_{imgs}) 为 10 幅、50 幅时,域迁移后的三维感知生成器不能生成足够高质量的人脸图像。如规模为 10 幅时,场景中依然有不可忽视的鬼影,当规模为 50 幅时,人脸模型的几何依然存在较大的瑕疵。当数据集规模提高后,生成器的 FID 指标不断降低,这意味着生成人脸图像的多样性和图像质量随之提

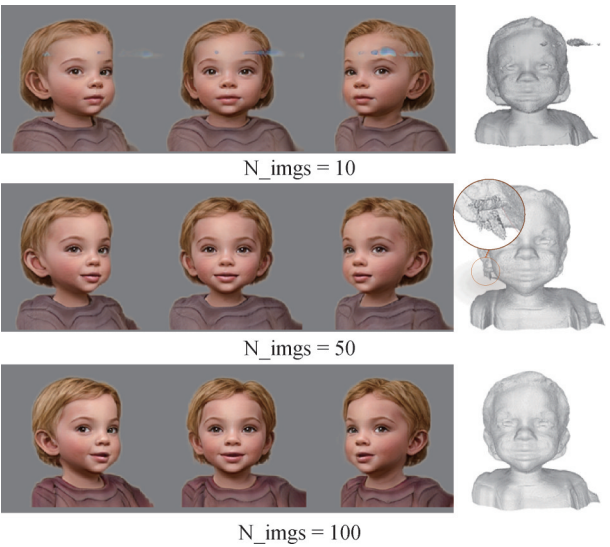


图 12 不同数据增强规模对生成器域迁移的影响
Fig. 12 Results of generator domain adaptation with different data augmentation scales

高。由此可以看出,通过二维风格化人脸数据增强的策略,三维感知生成器能够更为稳定地迁移到风格域中,在保持人脸视图生成质量的同时,构建整洁真实的三维几何。

表 2 不同数据增强规模的模型表现对比
Table 2 Performance of model with different data augmentation scales

小样本数据集规模/幅	FID ↓
10	10.39
50	6.79
100	5.37

注:“↓”表示值越低越好。

此外,针对纹理优化环节中的视图对齐策略,本文也设计了去除视图对齐过程对模型生成纹理优化结果的影响。使用多视角视图进行纹理融合的目标在于将三维感知风格化人脸的高质量纹理迁移到三维网格中。由于体渲染与网格表达的差异,在使用像素级损失优化纹理贴图的过程中,人脸视图的语义对齐至关重要。如图 13 所示,当未经对齐的视图直接用于纹理优化过程时,生成人脸模型的渲染视图具有严重的瑕疵。这是由于视图监督过程中不正确的语义对应导致 UV 纹理无法提取准确的信息,如将背景色误认为发色,而这严重影响了纹理优化过程中多视图监督的效果,使得最终得到的纹理信息错位。由于利用可微渲染反向优化纹理过程中采用的是逐像素的 L2 损失函数,因此,由神经辐射场渲染得到的人脸视图与由三维网格渲染得到的目标视图之间的对应关系,对纹理优化的最终结果至关重要。

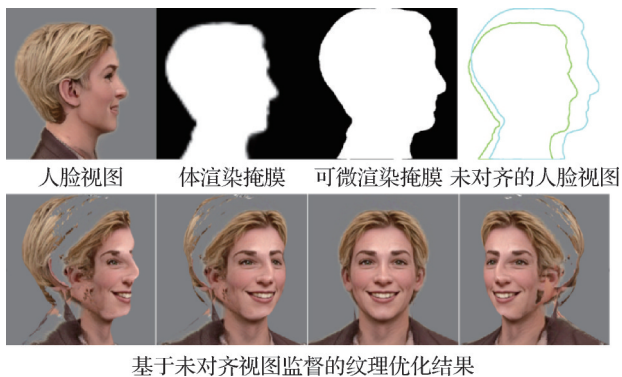


图 13 未采用视图对齐策略的纹理优化结果

Fig. 13 Results of texture optimization without aligning views

4 结 论

本文提出了一种基于样例的三维人脸风格化与结构化建模方法,只需指定一幅风格样例,该方法可高效率地根据给定艺术人脸图像样例生成对应风格的三维人脸模型。与此同时,为了便于三维人脸相关应用,该方法能够生成三维网格模型以及与之对应的高质量纹理贴图。得益于风格化人脸数据增强策略,自然人脸域的三维感知生成器能够稳定地迁移到样例的艺术风格域中,同时保持良好的几何感知以及人脸身份保持。域迁移后的三维感知生成器能够生成全视角的高分辨率风格化人脸视图。更进一步地,为了得到对应三维网格以及纹理贴图,本文提出一种基于掩膜对齐的多视角纹理优化方法。首先将平滑、简化处理后的三维网格进行结构化处理,再利用多视图的微分渲染模块反向优化纹理贴图,使之充分利用体渲染的高清视图。为了避免纹理优化过程中的多视角歧义,本文提出一种有效的基于仿射变换的人脸视图对齐策略,用于消除视图的错位。

实验中展示了多种风格下的结构化三维风格化人脸生成结果,包括具有良好多视角一致和生成质量的人脸视图以及对应的纹理贴图。然而本文框架仍然有以下问题。本文方法能够有效迁移纹理与色彩风格,但人脸几何变形程度依然有限,这是由于几何变形不仅体现在纹理色彩变化上,更需要人脸三维网格进行对应的几何变形。由于本文方法缺少对网格变形的相关约束,使得三维感知生成器难以学习到准确的三维场景变形,在较为夸张的艺术风格时容易导致模型产生脏乱的人脸几何。由于人脸几何风格表达的不足,如何高质量地保持人脸身份信息仍是当前一大挑战。另外,本文纹理优化的结果依然受限于纹理展开的质量,在处理更为夸张的人脸风格时,仍然含有难以忽视的瑕疵,未来通过引入更为整洁的纹理展开方法将能够有效改善本文多视图纹理优化的结果。

参考文献(References)

Abdal R, Lee H Y, Zhu P H, Chai M L, Siarohin A, Wonka P and Tulyakov S. 2023. 3DAvatarGAN: bridging domains for personal-

- ized editable avatars//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 4552-4562 [DOI: 10.1109/CVPR52729.2023.00442]
- An S Z, Xu H Y, Shi Y C, Song G X, Ogras U Y and Luo L J. 2023. PanoHead: geometry-aware 3D full-head synthesis in 360°//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 20950-20959 [DOI: 10.1109/CVPR52729.2023.02007]
- Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]
- Blanz V and Vetter T. 1999. A morphable model for the synthesis of 3D faces//Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. Los Angeles, USA: ACM Press/Addison-Wesley Publishing Co.: 187-194 [DOI: 10.1145/311535.311556]
- Cao C, Weng Y L, Zhou S, Tong Y Y and Zhou K. 2014. FaceWarehouse: a 3D facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics*, 20(3): 413-425 [DOI: 10.1109/TVCG.2013.249]
- Chan E R, Lin C Z, Chan M A, Nagano K, Pan B X, de Mello S, Gallo O, Guibas L, Tremblay J, Khamis S, Karras T and Wetzstein G. 2022. Efficient geometry-aware 3D generative adversarial networks//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 16102-16112 [DOI: 10.1109/CVPR52688.2022.01565]
- Chong M J and Forsyth D. 2022. JoJoGAN: one shot face stylization//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 128-152 [DOI: 10.1007/978-3-031-19787-1_8]
- Guo J Z, Zhu X Y, Yang Y, Yang F, Lei Z and Li S Z. 2020. Towards fast, accurate and stable 3D dense face alignment//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 152-168 [DOI: 10.1007/978-3-030-58529-7_10]
- Han F Z, Ye S Q, He M M, Chai M L and Liao J. 2023. Exemplar-based 3D portrait stylization. *IEEE Transactions on Visualization and Computer Graphics*, 29(2): 1371-1383 [DOI: 10.1109/TVCG.2021.3114308]
- Heusel M, Ramsauer H, Unterthiner T, Nessler B and Hochreiter S. 2017. GANs trained by a two time-scale update rule converge to a local Nash equilibrium//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 6629-6640
- Ho J, Jain A and Abbeel P. 2020. Denoising diffusion probabilistic models//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 6840-6851
- Karras T, Laine S and Aila T. 2019. A style-based generator architecture for generative adversarial networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 4396-4405 [DOI: 10.1109/CVPR.2019.00453]
- Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J and Aila T. 2020. Analyzing and improving the image quality of StyleGAN//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 8107-8116 [DOI: 10.1109/CVPR42600.2020.00813]
- Kato H, Ushiku Y and Harada T. 2018. Neural 3D mesh renderer//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3907-3916 [DOI: 10.1109/CVPR.2018.00411]
- Lorensen W E and Cline H E. 1987. Marching cubes: a high resolution 3D surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 21(4): 163-169 [DOI: 10.1145/37402.37422]
- Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2021. NeRF: representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99-106 [DOI: 10.1145/3503250]
- Oleynikova H, Millane A, Taylor Z, Galceran E, Nieto J and Siegwart R. 2016. Signed distance fields: a natural representation for both mapping and planning//RSS 2016 Workshop: Geometry and Beyond-Representations, Physics, and Scene Understanding for Robotics. Ann Arbor, USA: University of Michigan [DOI: 10.3929/ethz-a-010820134]
- OrEl R, Luo X, Shan M Y, Shechtman E, Park J J and Kemelmacher-Shlizerman I. 2022. StyleSDF: high-resolution 3D-consistent image and geometry generation//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 13493-13503 [DOI: 10.1109/CVPR52688.2022.01314]
- Park J J, Florence P, Straub J, Newcombe R and Lovegrove S. 2019. DeepSDF: learning continuous signed distance functions for shape representation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 165-174 [DOI: 10.1109/CVPR.2019.00025]
- Richardson E, Alaluf Y, Patashnik O, Nitzan Y, Azar Y, Shapiro S and Cohen-Or D. 2021. Encoding in style: a StyleGAN encoder for image-to-image translation//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 2287-2296 [DOI: 10.1109/CVPR46437.2021.00232]
- Roich D, Mokady R, Bermano A H and Cohen-Or D. 2023. Pivotal tuning for latent-based editing of real images. *ACM Transactions on Graphics*, 42(1): #6 [DOI: 10.1145/3544777]
- Ruiz N, Li Y Z, Jampani V, Pritch Y, Rubinstein M and Aberman K. 2023. DreamBooth: fine tuning text-to-image diffusion models for

- subject-driven generation//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 22500-22510 [DOI: 10.1109/CVPR52729.2023.02155]
- Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A and Chen X. 2016. Improved techniques for training GANs//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc.: 2234-2242
- Song G X, Luo L J, Liu J, Ma W C, Lai C P, Zheng C X and Cham T J. 2021. AgileGAN: stylizing portraits by inversion-consistent transfer learning. *ACM Transactions on Graphics*, 40(4): #117 [DOI: 10.1145/3450626.3459771]
- Song G X, Xu H Y, Liu J, Zhi T C, Shi Y C, Zhang J F, Jiang Z H, Feng J S, Sang S and Luo L J. 2023. AgileGAN3D: few-shot 3D portrait stylization by augmented transfer learning [EB/OL]. [2024-07-04]. <https://arxiv.org/pdf/2303.14297.pdf>
- Su H, Niu J W, Liu X F, Li Q F, Cui J H and Wan J. 2021. MangaGAN: unpaired photo-to-manga translation based on the methodology of manga drawing//Proceedings of the 35th AAAI Conference on Artificial Intelligence. Virtually: AAAI: 2611-2619 [DOI: 10.1609/aaai.v35i3.16364]
- Sun J X, Wang X, Shi Y C, Wang L Z, Wang J and Liu Y B. 2022a. IDE-3D: interactive disentangled editing for high-resolution 3D-aware portrait synthesis. *ACM Transactions on Graphics*, 41(6): #270 [DOI: 10.1145/3550454.3555506]
- Sun J X, Wang X, Zhang Y, Li X Y, Zhang Q, Liu Y B and Wang J. 2022b. FENeRF: face editing in neural radiance fields//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 7662-7672 [DOI: 10.1109/CVPR52688.2022.00752]
- Wang T F, Zhang B, Zhang T, Gu S Y, Bao J M, Baltrusaitis T, Shen J J, Chen D, Wen F, Chen Q F and Guo B N. 2023. RODIN: a generative model for sculpting 3D digital avatars using diffusion//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 4563-4573 [DOI: 10.1109/CVPR52729.2023.00443]
- Wu Q Y, Zhang J Y, Lai Y K, Zheng J M and Cai J F. 2018. Alive caricature from 2D to 3D//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7336-7345 [DOI: 10.1109/CVPR.2018.00766]
- Xiang J and Zhu G M. 2017. Joint face detection and facial expression recognition with MTCNN//Proceedings of the 4th International Conference on Information Science and Control Engineering (ICISCE). Changsha, China: IEEE: 424-427 [DOI: 10.1109/ICISCE.2017.95]
- Xue Y, Li Y H, Singh K K and Lee Y J. 2022. GIRAFFE HD: a high-resolution 3D-aware generative model//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 18419-18428 [DOI: 10.1109/CVPR52688.2022.01789]
- Yang S, Jiang L M, Liu Z W and Loy C C. 2022. Pastiche master: Exemplar-based high-resolution portrait style transfer//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 7683-7692 [DOI: 10.1109/CVPR52688.2022.00754]
- Ye Z P, Xia M F, Sun Y, Yi R, Yu M J, Zhang J Y, Lai Y K and Liu Y J. 2023. 3D-CariGAN: an end-to-end solution to 3D caricature generation from normal face photos. *IEEE Transactions on Visualization and Computer Graphics*, 29(4): 2203-2210 [DOI: 10.1109/TVCG.2021.3126659]
- Zhang R, Isola P, Efros A A, Shechtman E and Wang O. 2018. The unreasonable effectiveness of deep features as a perceptual metric//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 586-595 [DOI: 10.1109/CVPR.2018.00068]
- Zheng M W, Yang H Y, Huang D and Chen L M. 2022. ImFace: a non-linear 3D morphable face model with implicit neural representations//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 20311-20320 [DOI: 10.1109/CVPR52688.2022.01970]

作者简介

胡佳平,男,硕士研究生,主要研究方向为人脸内容生成。

E-mail: kexin039@163.com

周漾,通信作者,男,副教授,主要研究方向为纹理合成、图像处理、几何分析和三维重建。E-mail:zhouyangvcc@szu.edu.cn