

# 基于局部区域相关信息的图像编辑算法<sup>①</sup>



王志茹<sup>1</sup>, 蔺素珍<sup>1</sup>, 王彦博<sup>1</sup>, 李大威<sup>2</sup>, 侯晓伦<sup>1</sup>

<sup>1</sup>(中北大学 计算机科学与技术学院, 太原 030051)

<sup>2</sup>(中北大学 电气与控制工程学院, 太原 030051)

通信作者: 蔺素珍, E-mail: [lsz@nuc.edu.cn](mailto:lsz@nuc.edu.cn)

**摘要:** 针对现有图像编辑方法存在编辑结果不自然、不能很好地模拟图像闭塞内容的问题, 提出基于局部区域相关信息的图像编辑算法. 首先, 获取图像编辑区域和剩余区域的掩码, 使用预训练的 StyleGAN2 模型和掩码分别得到编辑区域和剩余区域的雅可比矩阵; 然后, 基于图像编辑区域变化最大和剩余区域变化最小的编辑思想, 将局部区域的语义发现制定为双层优化问题; 最后, 借助最优运输理论, 构建能够利用图像编辑区域和剩余区域相关信息来获取语义方向的算法, 该算法不仅具有闭合解而且效率较高. 实验结果表明, 与主流的局部图像编辑算法 LowRankGAN、ReSeFa 和 SDFlow 算法相比, 在编辑人脸图像中的嘴巴、眼睛等局部区域时, 本文算法能够很自然地模拟图像闭塞内容以及实现自然的图像编辑, 在保证编辑后的图像质量的同时, 也实现了图像局部区域编辑, 证实了算法的可控性和有效性.

**关键词:** 图像编辑; 生成对抗网络; 隐空间; 语义方向; 最优运输

引用格式: 王志茹, 蔺素珍, 王彦博, 李大威, 侯晓伦. 基于局部区域相关信息的图像编辑算法. 计算机系统应用, 2025, 34(6): 107-117. <http://www.c-s-a.org.cn/1003-3254/9847.html>

## Image Editing Algorithm Based on Local Region Correlation Information

WANG Zhi-Ru<sup>1</sup>, LIN Su-Zhen<sup>1</sup>, WANG Yan-Bo<sup>1</sup>, LI Da-Wei<sup>2</sup>, HOU Xiao-Lun<sup>1</sup>

<sup>1</sup>(School of Computer Science and Technology, North University of China, Taiyuan 030051, China)

<sup>2</sup>(School of Electrical and Control Engineering, North University of China, Taiyuan 030051, China)

**Abstract:** Addressing issues such as unnatural editing results and the inability to effectively simulate occluded content in images, an algorithm is proposed for image editing by discovering the semantic direction of local image regions. First, the masks of the editing region and the remaining region are obtained, and their Jacobian matrices are computed using a pre-trained StyleGAN2 model and the corresponding masks. Then, based on the principle that the editing region undergoes the most significant changes while the remaining region experiences minimal alterations, the semantic discovery of the local region is formulated as a two-layer optimization problem. Finally, leveraging optimal transport theory, an algorithm is proposed to determine the semantic direction by utilizing the correlation between the editing and remaining regions. This approach not only provides a closed-form solution but also ensures high efficiency. Experimental results demonstrate that, compared to mainstream local image editing algorithms such as LowRankGAN, ReSeFa and SDFlow, the proposed algorithm effectively simulates occluded content and achieves natural image editing, particularly when modifying facial features such as the mouth and eyes. While maintaining the quality of the edited image, it also enables precise local region editing, confirming the algorithm's controllability and effectiveness.

**Key words:** image editing; generative adversarial network (GAN); latent space; semantic direction; optimal transportation

① 基金项目: 国家自然科学基金 (62271453); 国家自然科学基金联合基金 (U21A20524); 山西省自然科学基金 (202303021211147); 山西省基础研究计划 (20210302123025)

收稿时间: 2024-09-11; 修改时间: 2024-10-10, 2024-11-19; 采用时间: 2024-12-06; csa 在线出版时间: 2025-04-30

CNKI 网络首发时间: 2025-05-06

## 1 引言

作为计算机视觉领域的新热点, 图像编辑已成为图像智能再创造和生成具有特殊用途新颖图像的主要手段. 随着高质量、多样化的深度生成模型<sup>[1]</sup>不断涌现, 如变分自编码器 (variational auto-encoder, VAE)<sup>[2]</sup>、生成对抗网络 (generative adversarial network, GAN)<sup>[3]</sup>以及扩散模型 (diffusion model)<sup>[4]</sup>, 很多基于不同生成模型的高效图像编辑方法应运而生.

现有代表性的生成模型中, VAE 将图像从高维空间编码至低维空间, 而低维数据难以模拟图像细节信息, 导致图像生成模糊<sup>[1]</sup>; 扩散模型在计算密度函数时采样困难导致效率低下, 且扩散模型的本质是通过像素之间的扩散来实现图像变化, 因此其生成结果可能会缺乏逼真度和创造力<sup>[5]</sup>; GAN 能够从随机噪声中生成不同风格、不同特征的图像, 更有利于生成多样性且创造力丰富的图像<sup>[6]</sup>, 更受到研究者重视. 现阶段利用 GAN 实现图像编辑大致有基于图像转换生成的编辑、基于拖拽式的编辑和基于隐空间语义信息的编辑 3 种方法. 其中, 图像转换生成的编辑方法通常基于图像域转换的思想来实现图像到图像的转换, 如输入手绘草图<sup>[7]</sup>、文字<sup>[8]</sup>等系列参考图像的属性 and 样式, 输出生成和编辑的图像, 其缺点是网络训练中往往需要配对数据, 但并非所有实际问题都能够轻松获得配对数据; 基于拖拽式的编辑方法通过定义多个控制点和对应目标点, 并将控制点的内容拖拽到目标点<sup>[9]</sup>, 但控制拖拽多个点以及精准拖拽到目标位置具有很大难度; 基于隐空间语义信息的编辑方法认为 GAN 的隐空间中包含丰富的语义信息<sup>[10-12]</sup>, 可以通过将隐向量朝向语义方向的移动来改变相应的图像内容. 由于隐空间本身包含的语义信息比训练模型学习到的图像特征要更加贴合图像本身的特征, 通过探索固定 GAN 模型隐空间内的可解释语义来改变隐向量能够更好地实现可控编辑, 所以这种方法一经提出, 迅速得到了认可.

基于隐空间语义信息编辑方法的关键是发现隐空间语义信息. 目前, 对于隐空间语义信息的发现已有很多尝试, 最直接的方法是根据图像的目标属性标记, 通过监督训练找到隐空间的分离边界<sup>[13,14]</sup>, 文献<sup>[13]</sup>通过支持向量机 (support vector machine, SVM) 算法找到能够分离一个特定二值语义属性的超平面, 该超平面的法向量将引导隐向量朝向特定语义方向移动, 控制图像生成. 然而, 这种有监督的学习方法需要目标属性的

清晰定义以及相应的手动注释, 这严重限制了它们在实践中中的应用. 鉴于此, 新近很多工作以无监督的方式来寻找隐空间的可引导方向, 如文献<sup>[15]</sup>通过研究将隐向量带入 GAN 生成器的全连接层的基本作用, 提出了一种用于发现潜在语义信息的通用封闭形式分解方法; 文献<sup>[16]</sup>基于主成分分析在隐空间和特征空间的应用中识别出重要的语义方向, 证明了大量可解释的控制可以由沿主方向的分层摄动定义. 但它们将整幅图像作为整体去发现全局语义, 然后通过用该全局语义改变隐向量来实现对整个输出图像的操作, 但某些全局语义会影响图像中的多个方面, 导致编辑不够精准.

图像局部编辑因其实用性而受到更多研究人员的关注, 并且在发现图像的局部语义方面取得进展. 例如, Zhu 等人<sup>[11]</sup>提出了 LowRankGAN 方法, 通过将图像区域和 GAN 的隐子空间相联系获取局部语义信息, 但這些子空间的发现依赖于松弛因子的低秩分解, 对一些敏感超参数的依赖导致对所选区域的鲁棒性不足. 文献<sup>[12]</sup>重新审视了使用预训练 GAN 的局部操作任务并提出 ReSeFa 方法, 并将基于区域的语义发现制定为对偶优化问题. 但 ReSeFa 不能充分利用不同图像区域之间的相关性信息导致编辑效果很不自然, 例如在编辑嘴巴等弯曲程度较大的区域内容时, 不能很好地控制弯曲弧度, 也不能很好地模拟闭塞的内容, 例如嘴巴里面的舌头、牙齿.

为此, 考虑到 StyleGAN 家族<sup>[17-19]</sup>具有丰富的隐空间, 便于通过操纵隐空间中的语义方向很好地控制生成图像的风格和特征, 本文借助预训练 StyleGAN2 模型的图像生成能力及最优运输理论能够学习和获取不同域之间的相关性信息的特点<sup>[20]</sup>, 基于图像编辑区域内容变化最大和剩余区域内容变化最小的思想, 提出了一种基于 StyleGAN2 的可控编辑算法 OT-ReSeFa (regional semantic discovery based on optimal transport). OT-ReSeFa 通过一阶 Taylor 展开将经过最优运输映射的图像编辑前后, 前景与背景的差异变化建模为 Rayleigh 熵优化问题, 通过最优运输实现编辑区域变化最大化和剩余区域变化最小化的全局视角, 以及其区域相关性的局部视角的交互, 以更好地利用编辑区域和剩余区域中某些像素点的相关性, 实现更加自然的编辑效果.

总的来说, 本文的主要工作如下.

(1) 提出了 OT-ReSeFa 算法. 此算法在利用区域之

间相关性的基础上得到隐空间的语义方向,并借助预训练的 StyleGAN2 模型实现可控的局部图像编辑任务.通过考虑图像编辑区域和剩余区域间的相关性,使得图像内容变化更自然,更好地模拟图像闭塞内容.

(2)在数据集 FFHQ<sup>[17]</sup>和 MetFaces<sup>[21]</sup>上验证所提算法有效性.与代表性算法相比,实验结果表明,本算法在编辑图像指标 KID (kernel inception distance)<sup>[22]</sup>和 FID (Fréchet inception distance)<sup>[23]</sup>上均有一定降低.

## 2 相关工作

### 2.1 StyleGAN2 隐空间的操纵模型

使用 StyleGAN2 学习的先验知识进行图像编辑已被广泛探索<sup>[24,25]</sup>.很多研究表明 StyleGAN2 隐空间中的隐向量具有一定算术运算性质,可以根据此性质及获取到的语义方向向量,通过在 StyleGAN2 学习的低维隐空间中执行矢量算法来修改图像的语义特征<sup>[11-13]</sup>.

具体来说,给定一个训练有素的 StyleGAN2 模型,生成器  $G$  可以表述为确定性函数  $G: Z \rightarrow X$ ,其中  $Z$  代表  $Z$  空间,隐向量  $z \in Z$  服从高斯分布,  $X$  代表图像空间,每个样本  $x \in X$  都具有一定语义信息,通过语义方向  $n$  改变隐向量  $z$  会导致生成图像  $G(z)$  中的相应语义变化.在生成图像  $G(z)$  上实现语义操作的过程可以表述为:

$$x^{\text{edit}} = G(z + \alpha n) \quad (1)$$

其中,  $\alpha$  表示编辑程度,这里,  $n$  通常被假定为单位向量<sup>[15]</sup>,即  $n^T n = 1$ .

由于 StyleGAN2 隐空间中具有丰富的语义,如何合理利用预训练生成器发现语义方向引起了相关研究者的强烈关注.文献<sup>[13]</sup>训练了一个属性分类器,可以使用图像注释来寻找更有意义的语义方向;在无监督的情况下发现和获取隐空间中的语义方向研究也取得进展,文献<sup>[16]</sup>利用 StyleGAN2 以分层的方式自发地编码图像丰富的语义,并且使用分层输入来控制编辑结果,文献<sup>[26]</sup>证明了 StyleGAN2 可以学习到每层表示的可转移性,并且其高度结构化的语义层次结构特性可以促进更加精准地进行可控编辑.因此,本文基于预训练的 StyleGAN2 模型发现和获取局部语义方向,进而实现局部图像编辑.

### 2.2 最优运输理论

最优运输理论是用凸几何、非线性偏微分方程解决概率统计问题的一个重要工具<sup>[27]</sup>.一言以蔽之,就是

以最小化运输成本的思想来计算不同概率分布之间的差异,具体如下.

给定两个概率空间  $x \in X$  和  $y \in Y$  以及在空间上的分布  $\mu$  和  $\nu$ ,期望找到映射  $T: X \rightarrow Y$ ,将服从分布  $\mu$  的随机变量变换为服从分布  $\nu$  的随机变量,同时极小化运输代价  $c(x, y)$  的期望.符合要求的映射  $T$  被称为最优运输映射.

最优运输映射的运输代价可以称为是两个概率分布之间的 Wasserstein 距离、Monge-Kantorovich 距离或推土机距离.计算机视觉领域,在处理一些机器学习任务时, Wasserstein 距离通常优于其他距离的表现. Wasserstein 距离最优运输映射的运输代价定义如下:

$$c(x, y) = \|x - y\|_2^2 \quad (2)$$

其中,  $\|\cdot\|_2$  表示向量的 2 范数.

借助最优运输理论能够分析经验分布间的全局和局部相互影响.具体来说,通过采用最优运输理论,能够全面捕捉数据之间的交互关系,既包括全局的分布尺度,也涵盖局部的样本尺度.这种混合视角能够同时考虑数据的局部和全局特征,从而提供更为精准和全面的结果.为此,本文借助最优运输理论,利用图像的编辑区域和剩余区域的相关性,获得更自然的编辑效果.

## 3 本文方法

现有的局部区域的图像编辑算法在编辑图像时,不能自然地编辑动作幅度较大的图像区域.例如,在编辑人脸图像时,对于嘴巴微笑弧度较大的图像区域,不能很好地控制嘴巴弧度缩回,在控制嘴巴张大的过程中,也不能很好地模拟牙齿随之抬高,舌头随之显露的自然变化过程.对于现存的这些问题,我们考虑编辑区域(前景)和剩余区域(背景)的像素点具有一定的相似性,即它们之间存在一定的相关性,在获取隐空间语义方向的过程中利用它们的相关性能够使得编辑效果更加自然.

本文通过引入最优运输思想实现了前景变化最大化和背景变化最小化的全局视角,以及考虑前景背景相关性的局部视角的交互,提出了局部图像区域编辑算法 OT-ReSeFa,本文算法具有更自然的编辑效果.第 3.1 节介绍用于构建优化问题的前景和背景雅可比矩阵为何以及如何计算;第 3.2 节介绍 OT-ReSeFa 算法

优化问题及如何构建;第3.3节介绍优化问题求解方法.

### 3.1 计算前景与背景的雅可比矩阵

#### 3.1.1 利用雅可比矩阵的原理

对于图像的局部编辑,理想的操作是只改变图像  $x$  中某个区域的内容,剩余的区域尽可能地保持不变.这里选择想要编辑的图像区域称作前景,表示为  $x_f$ ,剩余的图像区域称作背景,表示为  $x_b$ .  $x_f$  和  $x_b$  形成图像  $x$  内所有像素的分区,即  $x_f \cup x_b = x$  且  $x_f \cap x_b = 0$ . 对式(1)加以背景  $x_b$  保持不变的约束,这样,式(1)表示的图像的全局变化就重新规划为图像前景  $x_f$  的内容发生变化和背景  $x_b$  的内容保持不变的局部编辑任务,重新制定式(1)以适应局部编辑任务:

$$\begin{cases} x^{\text{edit}} = G(z + \alpha n) \\ \text{s.t. } \|x_b^{\text{edit}} - x_b\| = 0 \end{cases} \quad (3)$$

对于任意隐向量  $z$ ,生成器  $G(\cdot)$  相对于  $z$  的雅可比矩阵  $J_z$  可以定义为  $(J_z)_{j,k} = \frac{\partial \{G(z)\}_j}{\partial z_k}$ . 这里,  $j$  表示图像中的像素位置,  $k$  表示隐空间  $Z$  的维度. 通过泰勒级数,编辑结果的一阶近似可以写成:

$$G(z + \alpha n) = G(z) + \alpha J_z n + o(\alpha) \quad (4)$$

根据式(4),式(3)的约束可近似为:

$$\|x_b^{\text{edit}} - x_b\|_2^2 = \|G(z + \alpha n)_b - G(z)_b\|_2^2 \approx \alpha^2 n^T J_b^T J_b n \quad (5)$$

其中,  $\|\cdot\|_2$  表示向量的 2 范数,  $J_b$  是背景像素值对隐向量的导数,即  $(J_b)_{j,k} = \frac{\partial \{G(z)\}_j}{\partial z_k}$ .

同理,前景的变化也可类似表示为:

$$\|x_f^{\text{edit}} - x_f\|_2^2 \approx \alpha^2 n^T J_f^T J_f n \quad (6)$$

如上推理,图像前景和背景的变化可由雅可比矩阵表示,所以本文利用雅可比矩阵来构建 OT-ReSeFa 算法的优化问题.

#### 3.1.2 计算雅可比矩阵的实现

在实践中,只要生成器  $G(\cdot)$  是可微的,雅可比矩阵就可以很容易地计算出来,而不管其架构如何.

首先,初始化预训练 StyleGAN2 模型的生成器,定义下采样生成函数来对生成器生成的图像进行下采样,控制返回的生成图像大小为  $256 \times 256$  以节省计算时间和存储空间. 随机生成隐向量  $z$ , 经过生成器的映射网络后成为具有更加解纠缠能力的隐向量  $w$ . 以隐向量

$w$  以及下采样生成函数作为雅可比矩阵计算函数的输入,计算并返回整幅图像的雅可比矩阵  $J$ .

如图1所示,获取图像前景和背景的掩码,分别用于计算前景和背景的雅可比矩阵. 具体实现为,从整幅图像的雅可比矩阵  $J$  中,根据像素索引提取前景和背景像素的雅可比值,其像素索引分别为前景掩码中数值为 1、0 的部分. 前景和背景的雅可比矩阵如式(7)和式(8)所示:

$$J_f = \{J(i, j, 0) \mid (i, j) \in M_f\}, M_f = \{(i, j) \mid M_{ij} = 1\} \quad (7)$$

$$J_b = \{J(i, j, 0) \mid (i, j) \in M_b\}, M_b = \{(i, j) \mid M_{ij} = 0\} \quad (8)$$

其中,  $J$  形状为  $[H, W, 1, 512]$ ,  $M_f$  为前景掩码数值为 1 的像素索引,  $M_b$  为前景掩码数值为 0 的像素索引.

### 3.2 构建获取局部语义的优化问题

如图1中蓝色区域所示,由第3.1节分析和计算得到前景、背景的雅可比矩阵,并基于最优运输将前景变化最大和背景变化最小的思想构建为统一求最大化的优化问题.

令  $J_F = J_f^T J_f$  及  $J_B = J_b^T J_b$ , 对于  $J_F = [f_1, \dots, f_p]^T$  和  $J_B = [b_1, \dots, b_q]^T$  样本间的两两欧氏距离矩阵定义为:

$$M_{J_F, J_B} := \left[ \|f_i - b_j\|_2^2 \right]_{i,j} \in R^{p \times q} \quad (9)$$

其中,  $p, q$  的值分别为 512, 512.

设  $\langle A, B \rangle := \text{tr}(A^T B)$  为矩阵的 Frobenius 点积, 本文使用正则化参数  $\lambda$  在距离计算中控制局部信息的强度,且  $\lambda \geq 0$ .

最优运输提供了一个强有力的工具箱来计算两个经验概率分布之间的距离,通过考虑这两个度量之间的所有概率耦合来做到全局和局部视角的混合,以选择一个表示为  $T$ , 即对于给定的标准是最优的. 这种耦合描述了全局和局部尺度上的相互作用,反映在运输权重  $T_{i,j}$  上. 基于最优运输理论,本文中采用正则化 Wasserstein 距离来测量前景和背景之间的相异性,如式(10)所示:

$$W_\lambda(J_F, J_B) := \langle T_\lambda, M_{J_F, J_B} \rangle \quad (10)$$

其中,  $T_\lambda$  是熵平滑最优运输问题的解决方案.

设  $U_{pq}$  是  $p \times q$  非负矩阵分解的多面体,使得它们的行和列边缘分别等于  $1_p/p$  和  $1_q/q$ , 则:

$$U_{pq} := \{T \in R_+^{p \times q} : T1_q = 1_p/p, T^T 1_p = 1_q/q\} \quad (11)$$

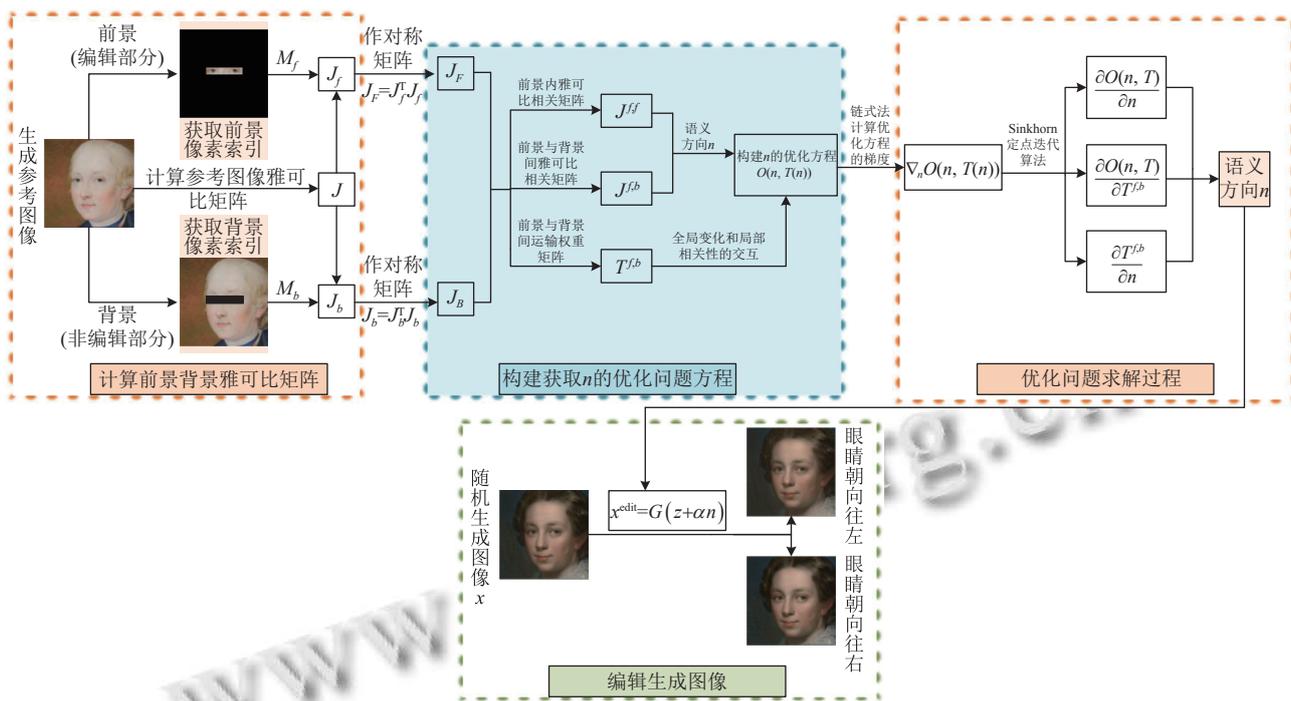


图1 OT-ReSeFa 算法实现图 (以获取眼睛姿态朝向语义为例)

根据式 (11),  $T_\lambda$  可表示为式 (12):

$$T_\lambda := \arg \min_{T \in U_{pq}} \lambda \langle T, M_{J^f, J^b} \rangle - \Omega(T) \quad (12)$$

其中,  $\Omega(T)$  表示  $T$  的熵, 可被视为离散联合概率分布, 即  $\Omega(T) := - \sum_{i,j} t_{ij} \log(t_{ij})$ ,  $t_{ij}$  是运输权重矩阵  $T$  的元素.

由上可知, 获取隐空间语义方向  $n$  的优化问题可以写成:

$$\begin{cases} \max_n \left\{ O(n, T(n)) = \frac{\sum_{f,b>f} \langle n^T n, J^{f,b} \rangle}{\sum_f \langle n^T n, J^{f,f} \rangle} \right\} \\ \text{s.t.} \begin{cases} J^{f,b} = \sum_{i,j} T_{i,j}^{f,b} (f_i - b_j)(f_i - b_j)^T, \forall f, b \\ T^{f,b} = \arg \min_{T \in U_{pq}} \lambda \langle T, M_{nJ^f, nJ^b} \rangle - \Omega(T) \end{cases} \end{cases} \quad (13)$$

### 3.3 优化问题求解过程

(1) 链式法计算优化方程的梯度

式 (13) 可以被重新制定为以下双层优化问题:

$$\begin{cases} \max_n O(n, T(n)) \\ \text{s.t.} T(n) = \arg \min_{T \in U_{pq}} E(T, n) \end{cases} \quad (14)$$

其中,  $T$  是前景和背景之间的运输矩阵, 内部问题函数  $E$  定义为:

$$E(T, n) = \sum_{f,b>f} \lambda \langle T^{f,b}, M_{nJ^f, nJ^b} \rangle - \Omega(T^{f,b}) \quad (15)$$

目标函数  $J$  可以表示为:

$$O(n, T(n)) = \frac{\langle n^T n, J^{f,b} \rangle}{\langle n^T n, J^{f,f} \rangle} \quad (16)$$

其中,  $J^{f,b}$ ,  $J^{f,f}$  是前景内的雅可比相关矩阵, 背景与前景间雅可比相关矩阵, 即  $J^{f,b} = \sum_{i,j} T_{i,j}^{f,b} (f_i - b_j)(f_i - b_j)^T$ ,  $J^{f,f} = \sum_{i,j} T_{i,j}^{f,f} (f_i - f_j)(f_i - f_j)^T$ .

事实上,  $O$  相对于  $n$  是可微的. 这是由于式 (13) 中的优化问题是严格凸的, 使得问题的解是唯一的, 因此  $T(n)$  是光滑可微的. 因此, 可以使用链式法则直接计算  $O$  的梯度, 如式 (17) 所示:

$$\nabla_n O(n, T(n)) = \frac{\partial O(n, T)}{\partial n} + \sum_{f,b>f} \frac{\partial O(n, T)}{\partial T^{f,b}} \frac{\partial T^{f,b}}{\partial n} \quad (17)$$

(2) 计算优化方程梯度中的每一项元素

如图 1 右侧所示, 求解优化问题转化为求解优化方程的梯度, 则需计算梯度公式 (式 (17)) 中的 3 项元素, 即  $\frac{\partial O(n, T)}{\partial n}$ ,  $\frac{\partial O(n, T)}{\partial T^{f,b}}$  和  $\frac{\partial T^{f,b}}{\partial n}$ .

计算式 (17) 中的第 1 项, 假设  $T$  是恒定的, 可以计算如式 (18) 所示:

$$\frac{\partial O(n, T)}{\partial n} = n \left( \frac{2}{\sigma_{f,f}^2} J^{f,b} - \frac{2\sigma_{f,b}^2}{\sigma_{f,f}^4} J^{f,f} \right) \quad (18)$$

其中,  $\sigma_{f,f}^2 = \langle n^T n, J^{f,f} \rangle$  且  $\sigma_{f,b}^2 = \langle n^T n, J^{f,b} \rangle$ , 为计算式 (17) 中的第 2 项, 我们分成式 (19) 和式 (20) 两种情况:

$$\frac{\partial O(n, T)}{\partial T^{f,b}} = \text{vec} \left( \frac{1}{\sigma_{f,f}^2} M_{nX^f, nX^b} \right) \quad (19)$$

$$\frac{\partial O(n, T)}{\partial T^{f,f}} = -\text{vec} \left( \frac{\sigma_{f,b}^2}{\sigma_{f,f}^4} M_{nX^f, nX^f} \right) \quad (20)$$

式 (19) 和式 (20) 中,  $\text{vec}(\cdot)$  表示将矩阵按列向量化, 式 (17) 的主要困难在于计算  $\frac{\partial T^{f,b}}{\partial n}$ , 因为最优传输矩阵无法以封闭形式表达, 使用围绕 Sinkhorn 定点迭代算法包裹的自动微分方法可解决这个问题。

将  $k$  次迭代后获得的解作为  $f, b$  对关于  $n$  的函数, 记为  $T^k(n)$ , 如式 (21) 所示:

$$T^k(n) = \text{diag}(\mu^k) e^{-\lambda M} \text{diag}(v^k) \quad (21)$$

式 (21) 中,  $\text{diag}(\cdot)$  表示对角矩阵, 其中  $v^k$  和  $\mu^k$  迭代  $k$  次后分别更新为:

$$v^k = \frac{1_q/q}{e^{-\lambda M^T} \mu^{k-1}} \quad (22)$$

$$\mu^k = \frac{1_p/p}{e^{-\lambda M} v^k} \quad (23)$$

式 (22) 和式 (23) 中,  $M$  是关于  $n$  的距离矩阵,  $k$  次迭代后的  $T(n)$  的直接微分为:

$$\begin{aligned} \frac{\partial T^k}{\partial n} &= \frac{\partial [\mu^k 1_q^T]}{\partial n} \circ e^{-\lambda M} \circ 1_p(v^k)^T + \mu^k 1_q^T \circ \frac{\partial e^{-\lambda M}}{\partial n} \circ 1_p(v^k)^T \\ &+ \mu^k 1_q^T \circ e^{-\lambda M} \circ \frac{\partial [1_p(v^k)^T]}{\partial n} \end{aligned} \quad (24)$$

其中,  $\circ$  是矩阵元素乘积。

根据式 (22) 和式 (23),  $\mu^k$  和  $v^k$  在第  $j$  行的一个分量的梯度分别为:

$$\frac{\partial \mu_j^k}{\partial n} = -\frac{1/p}{[e^{-\lambda M} v^k]_j^2} \left( \sum_i \frac{\partial e^{-\lambda M_{ji}}}{\partial n} v_i^k + \sum_i e^{-\lambda M_{ji}} \frac{\partial v_i^k}{\partial n} \right) \quad (25)$$

$$\begin{aligned} \frac{\partial v_j^k}{\partial n} &= \\ &- \frac{1/q}{[e^{-\lambda M^T} \mu^{k-1}]_j^2} \left( \sum_i \frac{\partial e^{-\lambda M_{ij}}}{\partial n} \mu_i^{k-1} + \sum_i e^{-\lambda M_{ij}} \frac{\partial \mu_i^{k-1}}{\partial n} \right) \end{aligned} \quad (26)$$

对于矩阵  $e^{-\lambda M}$  中第  $i$  行第  $j$  列元素的梯度为:

$$\frac{\partial e^{-\lambda M_{ij}}}{\partial n} = -2e^{-\lambda M_{ij}} n (f_i - b_j) (f_i - b_j)^T \quad (27)$$

通过在每次迭代中跟踪所有矩阵, 然后通过连续应用这些方程得到  $\frac{\partial T^{f,b}}{\partial n}$ 。

### (3) 算法总体实现步骤

本文算法的求解步骤如算法 1 所示。

#### 算法 1. 求解语义方向 $n$

1. 初始化  $L=0, n^0 \sim N(0,1)$
2. 重复
3. 计算式 (15), 得到  $T^{f,b}, T^{f,f}$
4. 根据  $T^{f,b}, T^{f,f}$ , 计算得到  $J^{f,b}, J^{f,f}$
5. 根据  $J^{f,b}, J^{f,f}$ , 计算式 (18), 得到  $\frac{\partial O(n,T)}{\partial n}$
6. 计算式 (19) 和式 (20), 得到  $\frac{\partial O(n,T)}{\partial T^{f,b}}$  和  $\frac{\partial O(n,T)}{\partial T^{f,f}}$
7. 设置 Sinkhorn 定点迭代次数  $k=10$ , 基于自动微分法计算式 (25) 和式 (26), 得到迭代  $k$  次后的  $\frac{\partial \mu_j^k}{\partial n}$  和  $\frac{\partial v_j^k}{\partial n}$
8. 计算式 (27)
9. 计算式 (24), 得到  $k$  次迭代后的  $\frac{\partial T^L}{\partial n}$
10. 使用上述所有元素计算式 (17), 得到梯度  $G^L = \nabla_n O(n^L, T(n^L))$
11. 计算梯度下降方向  $D^L = \Pi_{\Delta}(n^L - G^L) - n^L$
12. 根据线搜索选取合适的步长  $\alpha_L$
13.  $n^L \leftarrow \Pi_{\Delta}(n^L + \alpha_L D^L)$
14.  $L \leftarrow L+1$
15. 直至收敛

## 4 实验与结果分析

### 4.1 数据集

本文中我们使用了两个人脸图像数据集, 分别是 FFHQ<sup>[17]</sup> 和 MetFaces<sup>[21]</sup> 数据集。这两个数据集主要用于提取特征来计算评价指标。

FFHQ<sup>[17]</sup> 数据集最初作为生成对抗网络的训练数据集创建, 后期被用作 StyleGAN2 的训练数据集。FFHQ 是一个高质量人脸数据集, 包含 70 000 张 1024×1024 分辨率的高清人脸图像, 其在年龄、种族和图像背景上丰富多样且差异明显, 在人脸属性上也拥有非常多的变化, 拥有不同的年龄、性别、种族、肤色、表情、脸型、发型等。因此该数据集适用于开发一些人脸属性分类或者人脸属性编辑算法。

MetFaces<sup>[21]</sup> 数据集是从艺术品中提取的人脸图像数据集, 最初是用于训练生成对抗网络。该数据集包含 1 336 张分辨率为 1024×1024 的高质量 PNG 图像。

本文实验是对生成图像进行编辑, 通过预训练的 StyleGAN2 模型的生成器生成数据, 所以数据集主要用于计算评价指标。本文通过 1 000 张 FFHQ 数据集中

的图像以及 1 000 张 MetFaces 数据集中的图像作为提取特征计算指标的真实图像数据,通过生成 500 张编辑后的 FFHQ 图像及 MetFaces 图像作为提取特征计算指标的生成图像数据。

## 4.2 实验设置

实验环境: 操作系统为 Ubuntu 20.04.1, 开发环境为 PyTorch 1.8.0、CUDA 11.1、Python 3.7, 硬件平台为 NVIDIA GeForce RTX 3090 服务器。

超参数设置: 算法 1 中的迭代参数  $L$  设置为 100,  $k$  设置为 10, 正则化参数  $\lambda$  设置为 2。

## 4.3 对比方法及评价指标

为了验证本文方法在图像编辑任务中的有效性,在图像编辑结果的定性、定量对比实验中,采用基于语义编辑的两个方法作为基准模型,分别是 LowRankGAN<sup>[11]</sup>、ReSeFa<sup>[12]</sup>和 SDFlow<sup>[25]</sup>。

本文采用 KID (kernel inception distance)<sup>[22]</sup>和 FID (Fréchet inception distance)<sup>[23]</sup>这两个评价指标来估计编辑图像的质量。其计算方法为利用 VGG16 提取两幅图像的特征并计算距离。FID、KID 计算真实图像分布和

编辑图像分布之间的距离,其值越小表明编辑图像保留的原始图像内容越多。本文中使用了 256×256 的 1 000 张 FFHQ 数据集的人脸图像作为计算评价指标所需的真实数据集。

## 4.4 实验结果分析

### 4.4.1 对比实验结果与分析

#### (1) 定性结果分析

本文算法与 LowRankGAN、ReSeFa 和 SDFlow 算法的比较如图 2 所示,图 2(a)和图 2(b)分别展示了各个方法在 FFHQ 数据集上对嘴巴张开程度大小、眼睛大小以及鼻子大小的编辑结果。

观察图 2(a)和图 2(b)的第 1 列,这两组中嘴巴闭合的编辑结果,可以发现,LowRankGAN 在控制嘴巴关闭时,下巴会变宽,人脸会显得臃肿,与真实图像的下巴及面部相差很多。ReSeFa、SDFlow 在编辑嘴巴弧度较大的图像时,关闭嘴巴会使得嘴巴的弧度更大,牙齿也不能很好隐藏,导致嘴巴无法关闭,编辑结果非常不自然。而本文方法能够很好地控制嘴巴的弧度,不会造成脸部臃肿、下巴变宽。

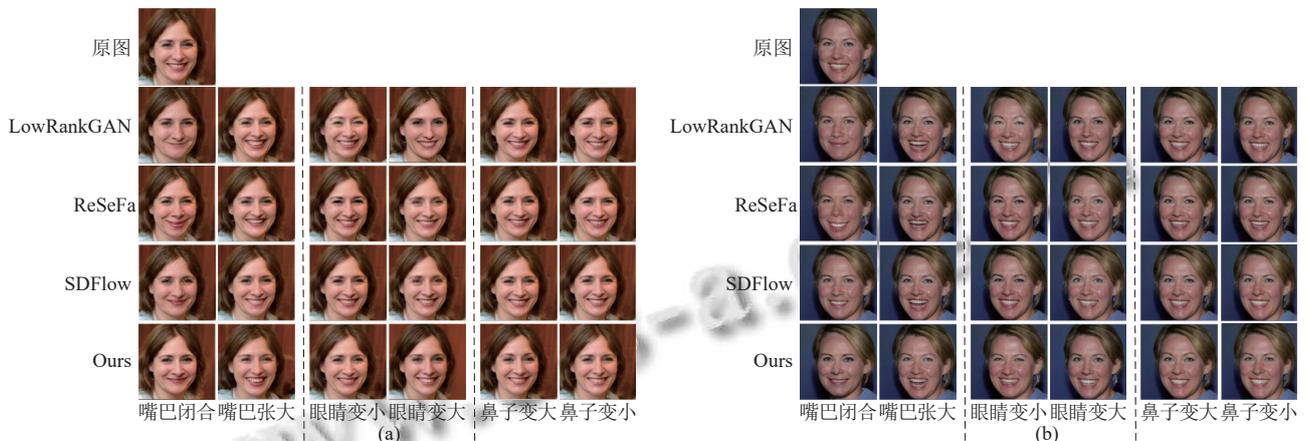


图 2 本文算法与 LowRankGAN、ReSeFa 及 SDFlow 算法的定性比较

观察图 2(a)和图 2(b)的第 2 列,这两组中嘴巴张大的编辑结果,我们可以发现 LowRankGAN、SDFlow 在控制嘴巴张大时,编辑结果比较自然。而 ReSeFa 的编辑结果中,下巴、下嘴唇以及牙齿显得非常的不自然。而本文方法很好地模拟了嘴巴张大时的牙齿、舌头以及下巴的状态,使得编辑结果非常自然。

观察图 2(a)和图 2(b)的第 3 列,这两组中眼睛变小的编辑结果,可以发现,LowRankGAN 在控制眼睛变小时,眼睛周围肤色大面积变白,面部臃肿,与原图

相差较大。ReSeFa、SDFlow 在控制眼睛变小时,眼珠颜色变暗,而本文方法不会出现眼睛周围肤色变化,眼珠变暗,编辑更加自然。

观察图 2(a)和图 2(b)的第 4 列,这两组中眼睛变大的编辑结果,LowRankGAN 在控制眼睛变大时,眼睛变化较自然但面部变窄。ReSeFa、SDFlow 在编辑眼睛时,眼睛周围睫毛会变得稀疏且皱纹加深。本文方法编辑结果相比于其他算法要更加自然且不突兀。

观察图 2(a)和图 2(b)的第 5 列,这两组中鼻子变

大的编辑结果, LowRankGAN、SDFlow 和 ReSeFa 控制鼻子变大时, 更偏向控制鼻子变宽, 与其他算法相比, 本文方法更倾向于鼻子变宽变长。

观察图 2(a) 和图 2(b) 的第 6 列, 这两组中鼻子变小的编辑结果, LowRankGAN、SDFlow 和 ReSeFa 控制鼻子变小时, 更偏向控制鼻子变窄, 与其他算法相比, 本文方法更倾向于鼻子变窄变短。

综上所述, 本文的编辑算法能够自然地编辑嘴巴、眼睛弧度较大的区域, 能够很好地模拟编辑过程中牙齿、舌头、眼部等局部的变化, 证明了本文算法

能够利用编辑区域和剩余区域的相关性实现自然的局部可控编辑。

## (2) 定量结果分析

为从客观指标上进一步对编辑结果进行量化对比, 选择 FFHQ 数据集上的嘴巴、眼睛和鼻子区域的正负语义方向进行编辑。基于 FID 和 KID (按 1E4 缩放) 两个指标, 将本文方法与基于语义的图像编辑方法 LowRankGAN、ReSeFa 及 SDFlow 进行对比表 1 为不同方法编辑嘴巴闭合和张大、眼睛变小和变大以及鼻子变小和变大的对比结果。

表 1 本文算法与 LowRankGAN、ReSeFa 及 SDFlow 算法的定量比较

数据集	方法	嘴巴闭合		嘴巴张大		鼻子变小		鼻子变大		眼睛变小		眼睛变大	
		FID	KID										
FFHQ	LowRankGAN	52.715	<u>1.831</u>	52.271	2.230	52.589	2.132	52.289	2.116	53.096	2.193	52.182	2.194
	ReSeFa	50.422	1.842	51.333	2.286	50.636	2.081	50.305	1.886	51.267	2.172	51.258	2.185
	SDFlow	<u>49.953</u>	1.836	<u>50.479</u>	<u>2.217</u>	<u>49.911</u>	<u>2.025</u>	<b>49.137</b>	<b>1.836</b>	<u>50.557</u>	<u>2.154</u>	<u>50.139</u>	<u>2.136</u>
	Ours	<b>48.825</b>	<b>1.823</b>	<b>49.728</b>	<b>2.196</b>	<b>49.782</b>	<b>2.021</b>	<u>49.164</u>	<u>1.847</u>	<b>50.349</b>	<b>2.146</b>	<b>50.041</b>	<b>2.127</b>
MetFaces	LowRankGAN	52.879	2.021	52.432	2.310	52.738	2.123	52.671	2.156	52.936	2.188	53.670	2.204
	ReSeFa	51.303	<u>1.834</u>	51.164	2.228	51.254	2.081	50.414	1.901	51.610	2.163	51.477	2.168
	SDFlow	<u>49.947</u>	1.842	<u>50.732</u>	<u>2.213</u>	<b>50.346</b>	<b>2.043</b>	<u>49.411</u>	<u>1.874</u>	<u>50.824</u>	<u>2.132</u>	<u>50.305</u>	<u>2.159</u>
	Ours	<b>48.833</b>	<b>1.824</b>	<b>50.319</b>	<b>2.206</b>	<u>50.400</u>	<u>2.045</u>	<b>49.268</b>	<b>1.871</b>	<b>50.717</b>	<b>2.126</b>	<b>50.180</b>	<b>2.157</b>

从表 1 可以看出, 本文方法的编辑效果基本优于其他方法。以 FFHQ 数据集上编辑嘴巴闭合和眼睛变小为例, 编辑嘴巴闭合的 FID 和 KID 相较于 LowRankGAN 方法分别降低了 7.37% 和 0.44%, 相较于 ReSeFa 方法分别降低了 3.16% 和 1.03%, 并相较于 SDFlow 方法分别降低了 1.86% 和 0.7%。编辑眼睛变小的 FID 和 KID 相较于 LowRankGAN 方法分别降低了 4.79% 和 2.1%, 相较于 ReSeFa 方法分别降低了 1.4% 和 1.19%, 并相较于 SDFlow 方法分别降低了 0.41% 和 0.37%。

但从表 1 可以看出, 本文方法在编辑鼻子时, FID 和 KID 高于 SDFlow 方法, 编辑效果劣于 SDFlow 方法。SDFlow 更加关注不同语义的纠缠问题, 比如在编辑鼻子的属性时, 不希望扰动其他的属性发生变化。而本文方法通过考虑前景 (鼻子区域) 背景 (其他区域) 中相似像素点的相关性使得图像内容变化更加自然流畅, 却忽视了在选取编辑区域时产生的属性纠缠问题, 使得某些属性效果不够好。

但总体上, 本文所提算法与目前经典算法相比, 编辑的图像效果有所改善。

### 4.4.2 单语义编辑实验

为了验证本文方法在单语义编辑上的有效性, 我

们使用 StyleGAN2 模型分别在 FFHQ 和 MetFaces 数据集上进行了实验, 选择眼睛、鼻子、嘴巴、头发等作为局部区域来获取它们的语义。

图 3 展示了本文方法在 FFHQ 数据集上的实验结果。图 3(a)–(d) 分别展示了嘴巴、鼻子、眼镜、年龄的两组编辑结果。每组编辑结果由 3 幅图像组成, 中间图像是原始图像, 左边和右边分别是根据语义方向得到的正负两端的编辑图像。如图 3(a) 所示, 这两组关于嘴巴的编辑结果, 从左到右展示了嘴巴张开程度由小到大的变化。如图 3(b) 所示, 这两组关于鼻子的编辑结果, 从左到右展示了鼻子由小到大的变化。图 3(c) 展示了眼镜大小、镜框粗细的变化。图 3(d) 展示了人脸年轻到年老的变化过程。

图 4 展示了本文方法在 MetFaces 数据集上的实验结果。图 4(a)–(d) 分别展示了眼睛、嘴巴、鼻子和头发的两组编辑结果。如图 4(a) 所示, 这两组关于眼睛的编辑结果, 从左到右展示了眼睛朝向逐渐向右的变化过程。如图 4(b) 所示, 这两组关于嘴巴的编辑结果, 从左到右展示了嘴巴由窄到宽的变化。图 4(c) 从左到右展示了鼻子由小到大的变化。图 4(d) 展示了头发颜色由深到浅的变化过程。

图3和图4说明本文方法可以揭示对合成图像进行合理局部控制的语义方向,能够对特定语义进行编辑且编辑结果符合人类日常生活经验,比较自然.



图3 FFHQ数据集上的单语义编辑结果

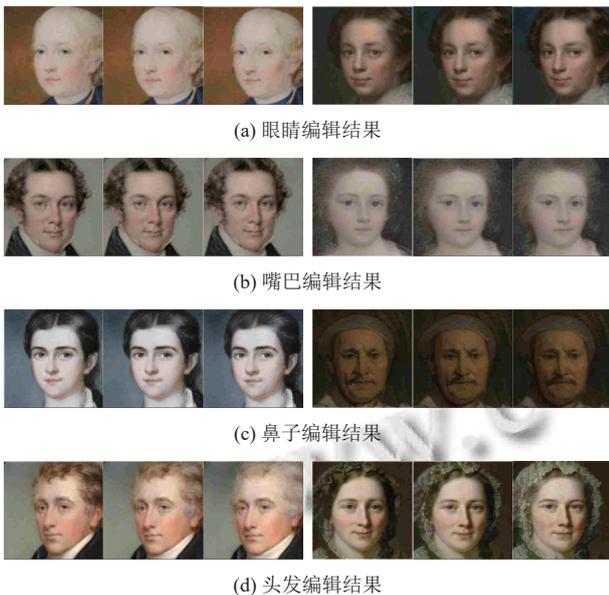


图4 数据集 MetFaces 上的单语义编辑结果

#### 4.4.3 多语义编辑实验

基于单语义编辑的思想,通过对隐向量执行多个语义方向向量的叠加操作,可以实现多语义编辑,多语义编辑的公式为:

$$z^{\text{edit}} = z + \alpha_1 n_1 + \alpha_2 n_2 + \dots + \alpha_i n_i \quad (28)$$

为验证本文多语义编辑的有效性,使用式(28)对

隐向量进行了年龄、嘴巴、鼻子相关的语义方向向量的叠加,其编辑结果如图5所示.图中的每一行是不同语义方向向量叠加的过程,第1-3行分别对应年龄、嘴巴、鼻子上的变化.隐向量 $z$ 沿着年龄语义方向向量的正方向移动,人脸中的胡子增加、皱纹加深,明显比原图中的人脸年长很多;接着沿着嘴巴语义方向向量的负方向移动,嘴巴张大程度明显变大;再沿着鼻子语义方向向量的正方向移动,鼻子明显变小;最后得到了一张年龄变大、嘴巴张大和鼻子变小的人脸图像.由于每个语义方向向量并不正交,会引起其他语义的变化,从第3行可以看出眼睛的些许变化.但从整体上看,这种叠加多个语义方向向量的多语义编辑方法,能够控制多个语义的变化,证明了式(28)的有效性.

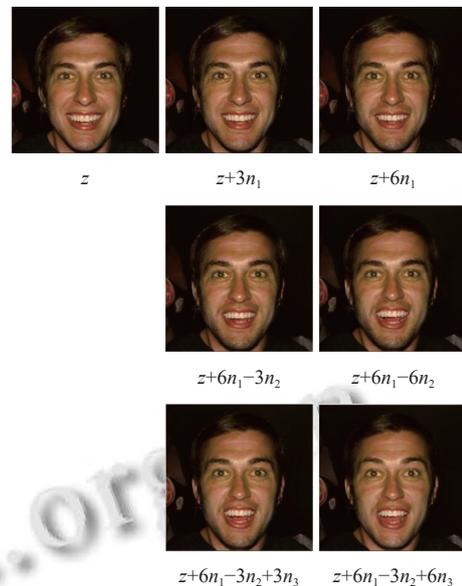


图5 多语义编辑结果

#### 4.4.4 控制因子对语义编辑程度的影响实验

为了观察控制因子 $\alpha$ 对语义修改程度的影响,本文在使用单语义方向向量对人脸的局部区域进行编辑时,通过设定不同的 $\alpha$ 对同一张图像进行编辑. $\alpha$ 取负值时,隐向量 $z$ 沿着语义方向的负方向移动, $\alpha$ 越小,编辑的程度越大; $\alpha$ 取正值时,隐向量 $z$ 沿着语义方向的正方向移动, $\alpha$ 越大,编辑的程度越大.如图6所示,图6(a)中展示了 $\alpha$ 从-10-0和从0-10控制年龄的变化过程, $\alpha$ 取的负值越小,人脸越老; $\alpha$ 取的正值越大,人脸越年轻.图6(b)中展示了 $\alpha$ 从-4-0和从0-4控制嘴巴张开程度的变化过程, $\alpha$ 取的负值越小,嘴巴张开越大; $\alpha$ 取的正值越大,嘴巴张开越小.通过控制因子可以观察像

序列一样的语义变化过程,这对于在实际生活中应用是十分有益的。

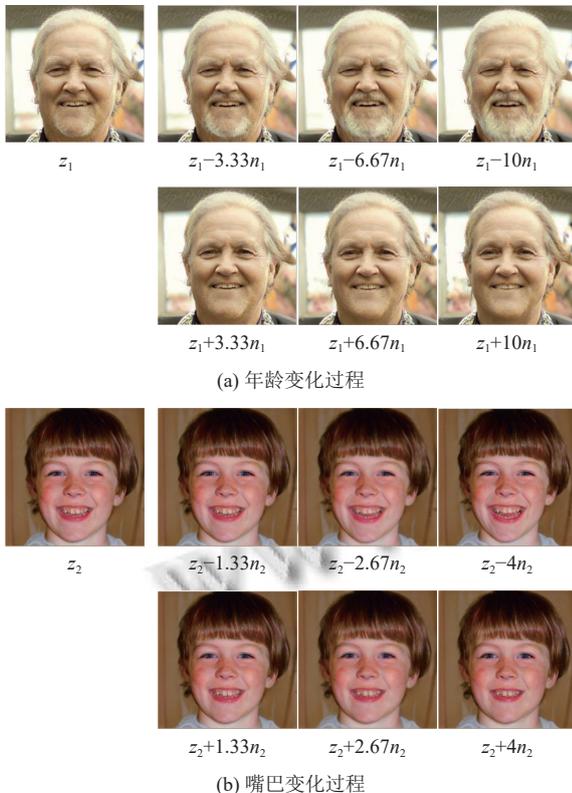


图6 不同控制因子下的单语义编辑结果

## 5 结语

本文着眼于当前基于生成对抗网络的语义图像编辑的研究趋势,深入剖析了主流方法在图像编辑方面的局限.为了突破这些限制,我们提出利用最优运输能够学习不同域之间的相关信息的特点,充分挖掘编辑区域(前景)与剩余区域(背景)之间的内在联系,使得编辑结果更为自然,且能够合理地模拟图像中的闭塞内容.经过实验验证,本文算法在图像编辑任务中展现出出色的局部区域控制能力,相较于其他基线模型,所编辑的图像更为逼真自然.目前,本文方法在细粒度编辑上效果欠佳.在未来的工作中,我们将进一步探索细粒度的编辑研究,以期实现更为精准和多样的图像编辑效果.

## 参考文献

1 胡铭菲,左信,刘建伟.深度生成模型综述.自动化学报,2022,48(1):40-74.

- 2 Asesh A. Variational autoencoder frameworks in generative AI model. Proceedings of the 24th International Arab Conference on Information Technology (ACIT). Ajman: IEEE, 2023. 1-6.
- 3 Brophy E, Wang ZW, She Q, *et al.* Generative adversarial networks in time series: A systematic literature review. ACM Computing Surveys, 2023, 55(10): 199.
- 4 Croitoru FA, Hondru V, Ionescu RT, *et al.* Diffusion models in vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(9): 10850-10869. [doi: 10.1109/TPAMI.2023.3261988]
- 5 Yang L, Zhang ZL, Song Y. Diffusion models: A comprehensive survey of methods and applications. ACM Computing Surveys, 2024, 56(4): 105.
- 6 Bond-Taylor S, Leach A, Long Y, *et al.* Deep generative modelling: A comparative review of VAEs, GANs, normalizing flows, energy-based and autoregressive models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(11): 7327-7347. [doi: 10.1109/TPAMI.2021.3116668]
- 7 Bashkirova D, Lezama J, Sohn K, *et al.* MaskSketch: Unpaired structure-guided masked image generation. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 1879-1889.
- 8 段亚茹,赵嘉雨,何立明.基于生成对抗网络的文本生成图像算法.计算机系统应用,2023,32(1):348-357. [doi: 10.15888/j.cnki.csa.008910]
- 9 Pan XG, Tewari A, Leimkühler T, *et al.* Drag your GAN: Interactive point-based manipulation on the generative image manifold. Proceedings of the 2023 ACM SIGGRAPH Conference. Los Angeles: ACM, 2023. 78.
- 10 Liu KL, Cao GF, Zhou F, *et al.* Towards disentangling latent space for unsupervised semantic face editing. IEEE Transactions on Image Processing, 2022, 31: 1475-1489. [doi: 10.1109/TIP.2022.3142527]
- 11 Zhu JP, Feng RL, Shen YJ, *et al.* Low-rank subspaces in GANs. Proceedings of the 35th International Conference on Neural Information Processing Systems. Curran Associates Inc., 2021. 1273.
- 12 Zhu JP, Shen YJ, Xu YH, *et al.* Region-based semantic factorization in GANs. Proceedings of the 39th International Conference on Machine Learning. Baltimore: PMLR, 2022. 27612-27632.
- 13 Shen YJ, Yang CY, Tang XO, *et al.* InterFaceGAN: Interpreting the disentangled face representation learned by

- GANs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(4): 2004–2018. [doi: [10.1109/TPAMI.2020.3034267](https://doi.org/10.1109/TPAMI.2020.3034267)]
- 14 陶玲玲, 刘波, 李文博, 何希平. 有闭解的可控人脸编辑算法. *计算机应用*, 2023, 43(2): 601–607.
- 15 Shen YJ, Zhou BL. Closed-form factorization of latent semantics in GANs. *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021. 1532–1540.
- 16 Härkönen E, Hertzmann A, Lehtinen J, *et al.* GANSpace: Discovering interpretable GAN controls. *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2020. 825.
- 17 Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 4401–4410.
- 18 Karras T, Laine S, Aittala M, *et al.* Analyzing and improving the image quality of StyleGAN. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020. 8107–8116.
- 19 Karras T, Aittala M, Laine S, *et al.* Alias-free generative adversarial networks. *Proceedings of the 35th International Conference on Neural Information Processing Systems*. Curran Associates Inc., 2021. 66.
- 20 曹杰彰, 莫朗元, 杜卿, 等. 基于最优传输理论的联合分布匹配方法及应用. *计算机学报*, 2021, 44(6): 1233–1245. [doi: [10.11897/SP.J.1016.2021.01233](https://doi.org/10.11897/SP.J.1016.2021.01233)]
- 21 Karras T, Aittala M, Hellsten J, *et al.* Training generative adversarial networks with limited data. *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2020. 1015.
- 22 Bińkowski M, Sutherland DJ, Arbel M, *et al.* Demystifying MMD GANs. *Proceedings of the 6th International Conference on Learning Representations*. Vancouver: OpenReview.net, 2018. 1–36.
- 23 Heusel M, Ramsauer H, Unterthiner T, *et al.* 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: Curran Associates Inc., 2017. 6629–6640.
- 24 Melnik A, Miasayedzenkau M, Makaravets D, *et al.* Face generation and editing with StyleGAN: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(5): 3557–3576.
- 25 Li BL, Huang ZH, Shan HM, *et al.* Semantic latent decomposition with normalizing flows for face editing. *Proceedings of the 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Seoul: IEEE, 2024. 4165–4169.
- 26 Xu YH, Shen YJ, Zhu JP, *et al.* GH-Feat: Learning versatile generative hierarchical features from GANs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(6): 7395–7411.
- 27 马丽涛, 边伟. 最优传输理论及其在图像处理中的应用. *运筹学学报*, 2019, 23(3): 109–125.

(校对责编: 王欣欣)