

哈爾濱工業大學

人工智能数学基础大项目报告

题 目	<u>基于生成式对抗网络的人脸编辑</u>
学 院	<u>计算机科学与技术</u>
专 业	<u>人工智能</u>
学 生	<u>张智雄 李佳阳 王佳依 王诗雅</u>
任 课 教 师	<u>刘绍辉</u>

哈尔滨工业大学计算机科学与技术学院

2023. 5

目录

1、实验内容或者文献情况介绍.....	2
1.1 研究现状.....	2
1.2 项目内容.....	2
2、算法简介及其实现细节.....	2
2.1 生成式对抗网络(GAN)的基本原理	3
2.2 GAN Inversion 模型的基本原理.....	4
2.2.1 StyleGAN 的基本原理.....	4
2.2.2 Encoder 的设计.....	4
2.3 InterfaceGAN 的基本原理.....	5
2.3.1 支持向量机 SVM 的原理.....	5
2.3.2 InterfaceGAN 的原理.....	7
2.4 风格自适应模型的基本原理.....	7
3、实验设置及结果分析（包括实验数据集）	8
3.1 数据集说明.....	8
3.2 随机人脸编辑.....	9
3.3 对指定人脸的编辑.....	9
3.4 艺术人脸的编辑.....	10
4、总结.....	12
5、组内分工.....	12
参考文献.....	13

大项目：基于生成式对抗网络的人脸编辑

1、实验内容或者文献情况介绍

1.1 研究现状

人脸编辑一般可以通过特征点标定算法检测人脸图像中的关键特征点，如眼睛、嘴巴、鼻子等。而后根据编辑需求，使用各种变形算法（如三角形插值和网格变形）来对人脸进行修改。

随着深度学习的发展,越来越多的人脸编辑任务可以通过生成式对抗网络(Generative Adversarial Networks, GAN)来解决。

基于 GAN 的图像编辑，现有较为成熟的方法大多是通过操纵 GAN 隐空间来实现的。隐空间是数据的一种稀疏表示的空间，通常来说，我们直接在数据空间对图像进行修改是不现实的，因为图像属性位于高维空间中的流形中且高度耦合。但是隐空间能较好地将图像属性进行解耦，由于每一个隐变量代表了某个具体的属性，所以基于 GAN 对图像进行编辑的方法是可行的。

目前现有的人脸编辑的任务种类繁多，主要包括人脸修复、多角度人脸生成、人脸的属性编辑、表情编辑、姿态编辑等，现有的较成熟的模型如 StyleGAN、pixel2style2pixel、InterfaceGAN 等在上述任务中都有较好的表现。

1.2 项目内容

基于生成式对抗网络模型(GANs)，实现对随机生成现实人脸图像、指定现实人脸图像以及艺术人脸图像的属性编辑（年龄、配饰、面部特征）、姿态编辑（转头，抬头，低头等姿态调整）、表情编辑（睁眼、闭嘴、微笑、哭泣、愤怒）。

2、算法简介及其实现细节

对于现实人脸，InterfaceGAN 模型能够通过 SVM 分类器针对解耦较充分的 StyleGAN 的生成域 $\mathcal{W}/\mathcal{W}^+$ 隐空间找到其中某项特定属性的超平面，从而找到该属性的线性子空间，通过改变编码向量从而编辑样式图像，而后将这个已经编辑好的向量放入 GAN 生成器得到具有目标面部（属性、表情、姿态等）的新图像，而通过控制特征的迁移率的连续变化即可按不同程度对图像进行编辑。

而对于艺术人脸，现有的大多数人脸图像编辑方法都是基于真实人脸图像域的，直接应用到艺术人脸图像上会出现跨域问题，如人脸扭曲和模糊等情况。因此需要首先将难以直接编辑的艺术作品的风格、特征迁移至较为直观适应艺术图像风格的生成器上，从而将艺术图像映射到一个可编辑的域中，以确保后续的编辑不会失真。

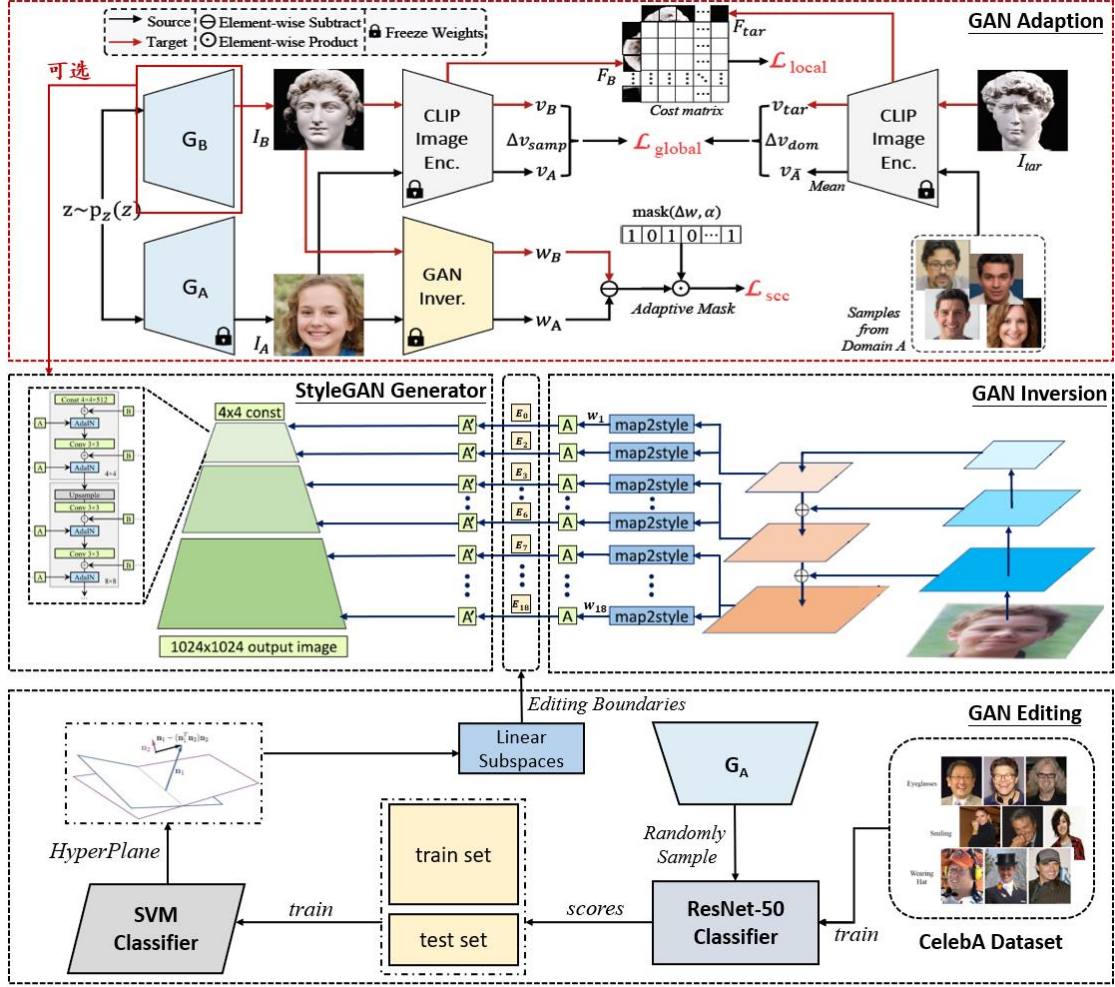


Figure 1 模型架构图

2.1 生成式对抗网络 (GAN) 的基本原理

生成式对抗网络是由生成器和判别器组成的一种深度学习框架,通过对一个简单的分布采样,学习一种映射将其变换到高维复杂的训练样本分布。

生成器通过对简单分布(如正态分布)进行采样并产生分布 P_G ;然后,判别器接收生成的数据和真实数据 P_{data} ,并尝试正确区分它们。二者通过循环迭代的对抗训练最终得到能较好拟合目标分布的生成器。

其中,对抗学习可以通过判别函数 $D(x): \mathbb{R}^n \rightarrow \{0,1\}$ 和生成函数 $G(x): \mathbb{R}^d \rightarrow \mathbb{R}^n$ 之间目标函数的极大极小值来进行表示,一种表示方式是

$$\min_G \max_D V(G, D) = E_{x \sim P_{data}} [\log D(x)] + E_{z \sim P_G} [\log (1 - D(G(z)))]$$

其中, z 是从已知简单分布(如高斯分布 $z \sim (0,1)$)中的随机采样,在训练过程中,生成器和判别器交替进行更新,使得对方的误差最大化。首先固定生成器 G ,通过最大化 $V(G, D)$ 求解此时最优的判别器 D ,此时 P_{data} 和 P_G 已知,则目标函数可以写成

$$V(G, D) = \int_x [a \log D(x) + b \log(1 - D(x))] dx$$

对其求导可以解得 $D^*(x) = \frac{P_{data}(x)}{P_{data}(x) + P_G(x)}$ ，而后固定判别器 D ，更新生成器 G 以减小目标函数。代入上述最优的 $D^*(x)$ 即可得到

$$\begin{aligned} V(G, D) &= \int_x \left[P_{data}(x) \log \frac{\frac{1}{2} P_{data}(x)}{\frac{1}{2}(P_{data}(x) + P_G(x))} + P_G(x) \log \frac{\frac{1}{2} P_G(x)}{\frac{1}{2}(P_{data}(x) + P_G(x))} \right] dx \\ &= -2 \log 2 + KL \left(P_{data} \parallel \frac{P_{data}(x) + P_G(x)}{2} \right) + KL \left(P_G \parallel \frac{P_{data}(x) + P_G(x)}{2} \right) \\ &= -2 \log 2 + 2JS(P_{data} \parallel P_G) \end{aligned}$$

可以观察得到最优情况时，当且仅当 $P_{data}(x) = P_G$ 时，此时 $D(x) = \frac{1}{2}$ ，即达到了纳什均衡。通过上述步骤交替训练生成器和判别器，GAN 网络可以逐步提高合成数据的质量，从而生成更加逼真和高质量的数据样本。

但是，基于 JS 散度的目标函数并不能很好的表示分布之间的距离，改用不同的目标函数也能够提高逼近的效果，比如 W-GAN、f-GAN 等。

2.2 GAN Inversion 模型的基本原理

2.2.1 StyleGAN 的基本原理

上述 2.1 节中提到的 GAN 由于数据量过大，并不能够生成高分辨率的图像；同时其生成具有随机性，并不具备控制生成图像的特定特征的能力。

基于此，StyleGAN 提出了一种渐进式训练的网络模型，通过从 4×4 低分辨率到 1024×1024 高分辨率的渐进训练以获得具有丰富细节的高分辨率图像。根据网络深度，可以大致分为‘粗糙’、‘中等’和‘精细’三级细节层次，不同层次对应不同特征的变化。

同时，StyleGAN 将随机简单采样 $z \in \mathcal{Z}$ 通过一系列全连接层映射到 $w \in \mathcal{W}$ 空间中，而后由 \mathcal{W} 空间非线性映射为图像特征 y ，以此避免由于训练集缺少某种特征组合导致从 \mathcal{Z} 空间到图像特征的映射扭曲。由此可以一定程度上保留不同特征的线性，因而可以通过对潜在空间向量的进行插值在图像中产生线性变化。

2.2.2 Encoder 的设计

在图像编辑任务中，直接在数据空间对图像进行修改是不现实的，因此我们需要将图像映射到一个高度解耦的 $\mathcal{W}/\mathcal{W}^+$ 空间中。

为更好的保留图像的细节，同时对特征进行解耦，pSp(pixel2style2pixel)模型和 e4e(encoder4editing)模型根据 StyleGAN 模型中‘粗糙’、‘中等’和‘精细’三级细节层次，利用金字塔结构构造特征映射，通过卷积层 map2style 提取图像编码，将已知图像映射到 StyleGAN 图像生成域上，完成对图像的解码以及反演。

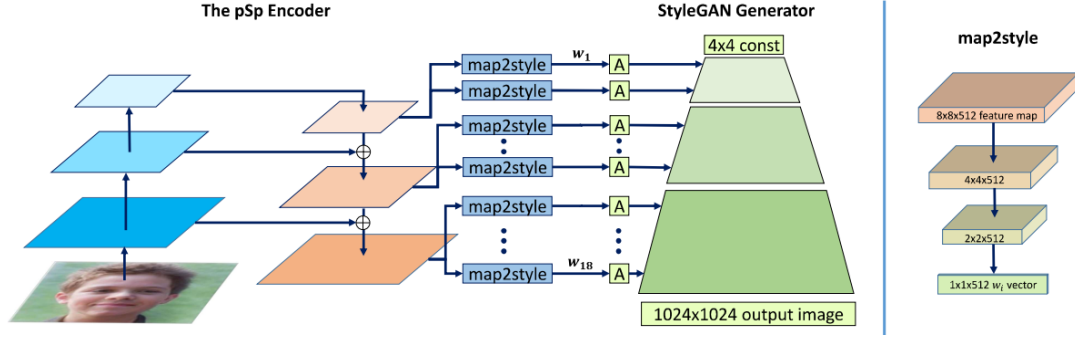


Figure 2 图像反演 pSp 模型图

训练过程中，为确保反演结果与输入尽可能相似，定义损失函数 \mathcal{L}_{loss} 如下，

$$\mathcal{L}_{loss} = \lambda_1 \underbrace{\|x - pSp(x)\|_2}_{\text{反演与原图的差异}} + \lambda_2 \underbrace{\|F(x) - F(pSp(x))\|_2}_{\text{感知相似性}} + \lambda_3 \underbrace{\|E(x) - \bar{w}\|_2}_{\text{与生成域的距离}} + \lambda_4 \underbrace{\mathcal{L}_{ID}}_{\text{身份损失}}$$

其中， $pSp(\cdot)$ 为模型反演得到的图像， $F(\cdot)$ 为图像经感知特征提取器得到结果， $E(\cdot)$ 为对图像解码结果， \mathcal{L}_{ID} 为身份损失，由人脸识别预训练模型 ArcFACE 计算。

2.3 InterfaceGAN 的基本原理

2.3.1 支持向量机 SVM 的原理

支持向量机(Support Vector Machines, SVM)是一类按有监督学习方式对数据进行二元分类的广义线性分类器，其决策边界是对学习样本求解的最大边距超平面，可以将问题化为一个求解凸二次规划的问题，求解能够正确划分训练数据集并且几何间隔最大的分离超平面。

考虑在给定特征空间上的训练数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ 的二分类问题，其中， $x_i \in \mathbb{R}^n, y_i \in \{+1, -1\}, i = 1, 2, \dots, N$ ， x_i 为训练集第 i 个样本向量， y_i 为其对应的类别。则目标是找到一个函数 $g: \mathbb{R}^n \rightarrow \mathbb{R}$ ，而后通过 $\text{sgn}(g(x))$ 映射到类别 $\{+1, -1\}$ 上。

对训练集 T ，假设其是线性可分的，即 $\exists w \in \mathbb{R}^n, b \in \mathbb{R}, \varepsilon > 0$ ，使得对于所有 $y_i = 1$ 的样本，有 $w^T \cdot x + b \geq \varepsilon$ ；而对所有 $y_i = -1$ 的样本，有 $w^T \cdot x + b \leq -\varepsilon$ 。

同时，对于给定的数据集 T 和超平面 $w^T \cdot x + b = 0$ ，定义超平面关于样本点 (x_i, y_i) 的几何间隔为

$$r_i = \frac{|w^T x_i + b|}{\|w\|} = y_i \left(\frac{w^T}{\|w\|} \cdot x_i + \frac{b}{\|w\|} \right), r = \min_{i=1,2,\dots,N} r_i$$

其中 r 为所有样本点的几何间隔的最小值。则 SVM 模型的求解最大分割超平面问题可以表示为以下约束最优化问题，也叫硬间隔(hard margin)最大化问题。

$$\max_{w,b} r, s.t. y_i \left(\frac{w^T}{\|w\|} \cdot x_i + \frac{b}{\|w\|} \right) \geq r, i = 1, 2, \dots, N$$

将约束条件两边同时除以 r ，记 $\hat{w} = \frac{w}{\|w\|r}$, $\hat{b} = \frac{b}{\|w\|r}$ ，因此最大化 r 等价于最小化 $\frac{1}{2}\|\hat{w}\|^2$ ，则问题等价于

$$\min_{w,b} \frac{1}{2} \|\hat{w}\|^2, s.t. y_i(\hat{w}^T \cdot x_i + \hat{b}) \geq 1, i = 1, 2, \dots, N$$

解这个凸二次规划问题，首先构造出拉格朗日函数，

$$L(w, b, \mu) = \frac{1}{2} \|\hat{w}\|^2 - \sum_{i=1}^N \mu_i [y_i(\hat{w}^T \cdot x_i + \hat{b}) - 1], \mu_i \geq 0$$

则问题的对偶函数及对偶问题为

$$\theta(\mu) = \min_{w,b} L(w, b, \mu); \max_{\mu} \theta(\mu), s.t. \mu \geq 0$$

首先求解 $\theta(\mu) = \min_{w,b} L(w, b, \mu)$ ，显然最小值在 L 梯度为0时得到，即

$$\begin{cases} \frac{\partial L}{\partial w} = w - \sum_{i=1}^N \mu_i y_i x_i = 0 \\ \frac{\partial L}{\partial b} = - \sum_{i=1}^N \mu_i y_i = -\mu^T y = 0 \end{cases}$$

代回 $L(w, b, \mu)$ 中求得最优解 $\mu^* = [\mu_1^*, \mu_2^*, \dots, \mu_n^*]^T$ ，计算 $w^* = \sum_{i=1}^N \mu_i^* y_i x_i$ ，由KKT互补条件知 $\mu_i^* [y_i(\hat{w}^{*T} \cdot x_i + \hat{b}^*) - 1] = 0$ ，当 $\mu_i^* > 0$ 时，有 $y_i(\hat{w}^{*T} \cdot x_i + \hat{b}^*) = 1$ 。即 w^* 和 b^* 仅由对应 $\mu_i^* > 0$ 的样本点 (x_i, y_i) 决定，这样的 x_i 即为支持向量。

而如果数据集 T 线性不可分，则存在不满足约束条件 $y_i(\hat{w}^T \cdot x_i + \hat{b}) \geq 1$ 的样本，会导致优化问题没有可行解。因此需要对目标函数和约束条件引入非负的松弛变量 ξ_i 进行修正，即变为软间隔(soft margin)最大化问题。

$$\min_{w,b} \frac{1}{2} \|\hat{w}\|^2 + C \sum_{i=1}^N \xi_i, s.t. \begin{cases} y_i(\hat{w}^T \cdot x_i + \hat{b}) \geq 1 - \xi_i, i = 1, 2, \dots, N, C > 0 \\ \xi_i \geq 0 \end{cases}$$

其中， ξ_i 为超平面加入了非线性， C 为惩罚参数，约束最优分离超平面在间隔尽量大的同时尽量保持线性。同样地，定义其凸二次规划的拉格朗日函数。

$$L(w, b, \xi, \mu, \lambda) = \frac{1}{2} \|\hat{w}\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \mu_i [y_i(\hat{w}^T \cdot x_i + \hat{b}) - 1 + \xi_i] - \sum_{i=1}^N \lambda_i \xi_i$$

则问题的对偶函数 $\theta(\mu)$ 及对偶问题为

$$\theta(\mu, \lambda) = \min_{w,b} L(w, b, \xi, \mu, \lambda); \max_{\mu, \lambda} \theta(\mu, \lambda), s.t. \mu, \lambda \geq 0$$

取 L 梯度为0代回可得 $\theta(\mu, \lambda) = -\frac{1}{2} \mu^T Q \mu + \mathbf{1}^T \mu$ ，其中 Q 为对称矩阵， $Q_{ij} = y_i y_j x_i^T x_j$ ，因而对偶问题化简为

$$\max_{\mu} -\frac{1}{2} \mu^T Q \mu + \mathbf{1}^T \mu, s.t. \begin{cases} 0 \leq \mu \leq C \cdot \mathbf{1} \\ \mu^T y = 0 \end{cases}$$

解得最优解 $\mu^* = [\mu_1^*, \mu_2^*, \dots, \mu_n^*]^T$ ，则原问题的最优解 $w^* = \sum_{i=1}^N \mu_i^* y_i x_i$ 。而由KKT互补条件 b^* 的值由 μ^* 和支持向量决定。同样地，支持向量对应 $0 < \mu_i \leq C$ 的样本。

$$\begin{cases} \mu_i = 0 \Leftrightarrow y_i(\hat{w}^T \cdot x_i + \hat{b}) \geq 1 \\ 0 < \mu_i < C \Leftrightarrow y_i(\hat{w}^T \cdot x_i + \hat{b}) = 1 \\ \mu_i = C \Leftrightarrow y_i(\hat{w}^T \cdot x_i + \hat{b}) = 1 - \xi_i \leq 1 \end{cases}$$

因此当 $0 < \mu_i < C$ 时, 有 $y_i(\hat{w}^T \cdot x_i + \hat{b}) = 1$ 成立, 又因为 $y_i^2 = 1$, 于是可得 $b^* = y_j - \sum_{support} \mu^* y_i x_i^T x_j$, 同理可以解得 $\mu_i = C$ 时对应的 b^* 。

在实际应用中, 还可以通过引入核函数, 可以使 SVM 适用于非线性分类问题。通过一个非线性映射 $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^m (m \gg n)$ 将低维的输入解耦至高维空间, 从而实现数据的线性可分。

2.3.2 InterfaceGAN 的原理

上述 2.2.1 节中 StyleGAN 生成的图像域的特征在一定程度上具有线性可分的性质, 因此我们可以通过 SVM 来寻找某特征的线性子空间, 从而实现对图像某一特定属性的编辑修改。

具体而言, 通过标注好的 Celeb-A 数据集训练一个分类器, 而后对生成器生成的数据集进行分类打分, 并随机分割为训练集和测试集。而后通过 SVM 寻找此特征的分类超平面, 从而找到一个与属性相关的线性子空间。将得到的编辑向量加在待编辑图像的编码上即可对此属性进行修改, 还可以通过修改迁移率, 以控制变化的距离。

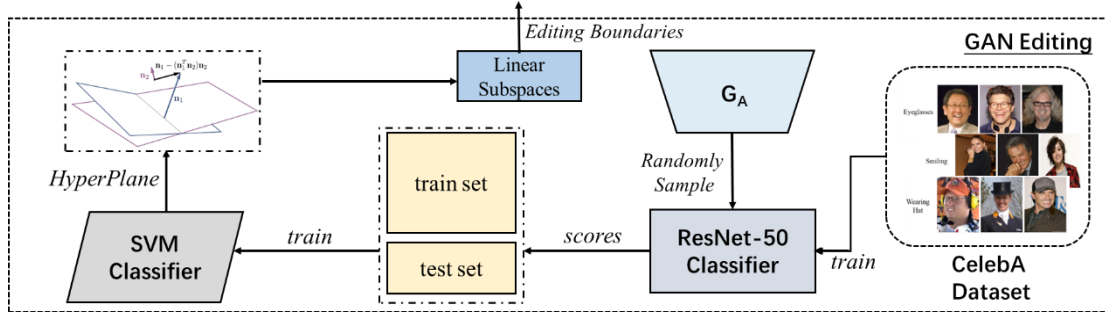


Figure 3 图像编辑 InterfaceGAN 模型图

2.4 风格自适应模型的基本原理

当需要针对目标图像的风格改换生成器时, 上述 StyleGAN 方法的训练需要大量样本。而当样本数量非常有限时, 可以采用基于单样本的学习方法(one-shot), 从给定的样本中推断出新的、未见过的样本的属性, 对模型进行泛化。

而 DIFA 模型, 基于预训练的可生成高分辨率真实人脸的 StyleGAN 模型 G_A , 通过计算目标域 B 和原始生成域 A 之间的距离实现生成器对目标图像 I_{tar} 风格的自适应。进而能够获得能够较好继承目标艺术图像的风格和特征, 同时保留人脸特征的图像生成器 G_B , 从而获得一个适应目标编辑图像的图像域, 实现生成网络对目标风格域的自适应, 以便后续编辑操作。

Figure 4 风格适应 DIFA 模型图

训练过程中，为较好的实现风格的适应并保留生成多样性，定义损失函数 \mathcal{L}_{loss} 如下

$$\mathcal{L}_{loss} = \mathcal{L}_{global} + \lambda_1 \mathcal{L}_{local} + \lambda_2 \mathcal{L}_{scc}$$

其中, \mathcal{L}_{global} 基于 CLIP(Contrastive Language Image Pretraining)模型, 描述生成域之间的距离, 以此来实现对目标域的逼近。

$$\mathcal{L}_{global} = 1 - \frac{\Delta v_{samp} \cdot \Delta v_{dom}}{|\Delta v_{samp}| |\Delta v_{dom}|}; \text{其中 } \Delta v_{dom} = v_{tar} - v_{\bar{A}}, \Delta v_{samp} = v_B - v_A$$

而 \mathcal{L}_{local} 对齐最接近的部分 token，以帮助 G_B 更好地获得 I_{tar} 的代表属性和风格，避免因内容和风格的不匹配导致可见的伪影。

$$\mathcal{L}_{local} = \max(\frac{1}{n} \sum_i \min_j C_{i,j}, \frac{1}{m} \sum_j \min_i C_{i,j}); \text{其中 } C_{i,j} = 1 - \frac{F_B^i \cdot F_{tar}^j}{|F_B^i| |F_{tar}^j|}$$

其中, C 是成本矩阵, 用于衡量从 F_B 到 F_{tar} 的标记性距离。最后, 选择性跨域一致性损失 \mathcal{L}_{scc} 则是为了促进多样性的生成, 基于上述提到的 GAN Inversion 模型计算 \mathcal{W}^+ 编码 X_A 队列和 \mathcal{W}^+ 潜伏码 X_B 队列的中心之间的差异 Δw , 并使之尽可能地相同。

$$\mathcal{L}_{SCC} = ||mask(\Delta w, \alpha) \cdot (w_B - w_A)||_1, mask(\Delta w, \alpha) = \begin{cases} 1 & |\Delta w_i| < |\Delta w_{SA_N}| \\ 0 & |\Delta w_i| \geq |\Delta w_{SA_N}| \end{cases}$$

3、实验设置及结果分析（包括实验数据集）

3.1 数据集说明

Celeb-A(Celebrities Attributes)是一个广泛用于人脸相关任务的数据集，包含了 202599 张名人照片，涵盖了大约 10177 个不同的名人身份。这些照片涵盖了各种不同的背景、姿势和表情，以及多个年龄段、性别和种族。每张照片都被标注了 40 个不同的属性，包括年龄、性别、发型等。

FFHQ(Flickr-Faces-HQ)是一个用于人脸图像生成和人脸识别研究的高质量数据集，包含了 70000 张分辨率为 1024x1024 的真实人脸图像。这些图像是从 Flickr 网站上公开可用的高质量肖像照片中采集而来。数据集中的人脸图像包含了各种年龄、性别、种族和肤色的人物。

3.2 随机人脸编辑

经由上述 InterfaceGAN 训练得到的编辑向量可以得到有关微笑、姿态、性别、头发长度等属性的编辑向量。将这些编辑向量直接加至基于 FFHQ 数据集预训练的 StyleGAN 生成器产生的 $w \in \mathcal{W}$ 向量上，生成图像如下。

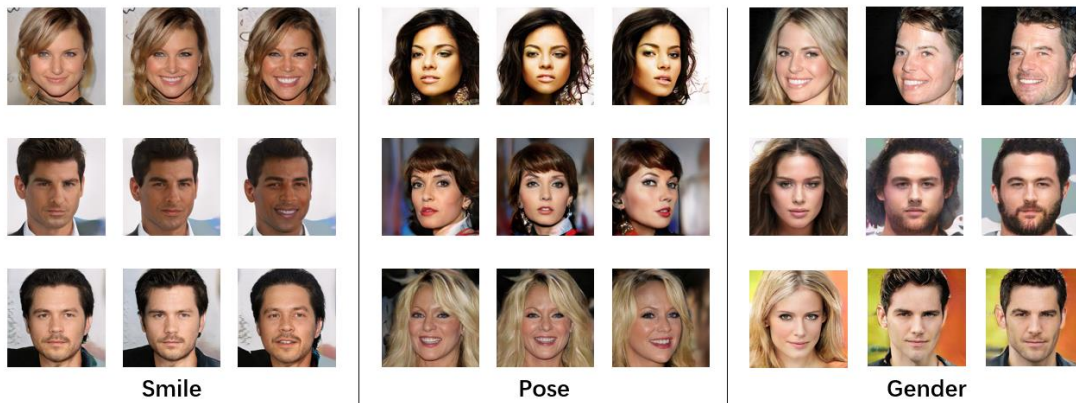


Figure 5 随机人脸编辑效果

观察效果图可以看出同一编辑向量加至不同图像上的编辑效果是不同的，并且某些属性仍与其他属性耦合度较高，如改变性别会自然而然的为男性图像加上胡须并将头发改为短发，这说明训练集内长发男性以及有胡须女性的样本缺乏，分类器不能将这些特征属性解耦到不同线性空间。

3.3 对指定人脸的编辑

而对于指定人脸，需要首先将目标图像通过上述的反演模型将其解耦至高维的隐空间 \mathcal{W}^+ ，得到其隐编码。由于 \mathcal{W}^+ 空间是拓展的 \mathcal{W} 空间，因而需要将上述获得的 \mathcal{W} 空间中的编码重复作用到图像在 \mathcal{W}^+ 空间隐编码的 18 维中（类似 StyleGAN 中将 A 向量重复投入网络层），编辑结果如下。

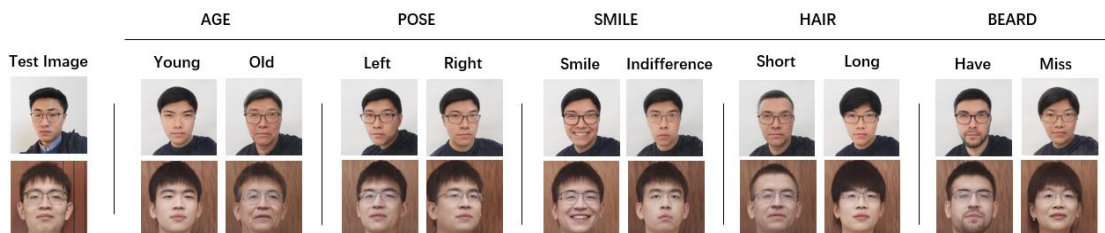


Figure 6 指定人脸编辑效果

实验发现，最终的编辑效果部分取决于反演模型的效率，同时对拍照角度有一定要求，角度的偏差可能会导致反演过程中脸型的拉长等。并且由于 pSp 模型是基于 FFHQ 数据集预训练的 StyleGAN 模型来进行图像的反演的，如果待编辑图像与 FFHQ 数据集内图像特征相差过大，会导致编辑失真等结果。

3.4 艺术人脸的编辑

同样地，我们基于 pSp 模型获得上述图片的隐空间编码，并将生成的隐空间编码再次投入预训练的 StyleGAN 生成器中，观察生成图像验证反演的准确性。为验证第一步风格迁移的必要性，我们也将此隐空间编码投入到第一步得到带有目标图像风格的 StyleGAN 生成器，将生成图像进行对比，得到结果如下。



Figure 7 图像反演对比

观察上述实验结果可以发现，同样的隐空间编码，在原本基于 FFHQ 现实人脸数据集的预训练 GAN 生成器 G_A 中生成的图像会更加具备现实人脸的特征，重建过程中其原本艺术图像的风格不能较好保留，在一些细节上会与原图存在一定差异；而利用 DIFA 模型获得的生成器 G_B 重建得到的图像会更加贴近艺术图像本身，对目标风格的适应效果更好。

但是在艺术风格的适应迁移中，对一些细节特征（如眼睛颜色等）在较为卡通动漫的艺术人物的风格迁移中不能够较好的完成，使得眼色更多的与目标图像中的背景色相适应。

而后，将上述编辑向量作用到这些艺术图像上，实验发现这些编辑向量在艺术人脸的编辑中同样适用，编辑结果如下图 8。

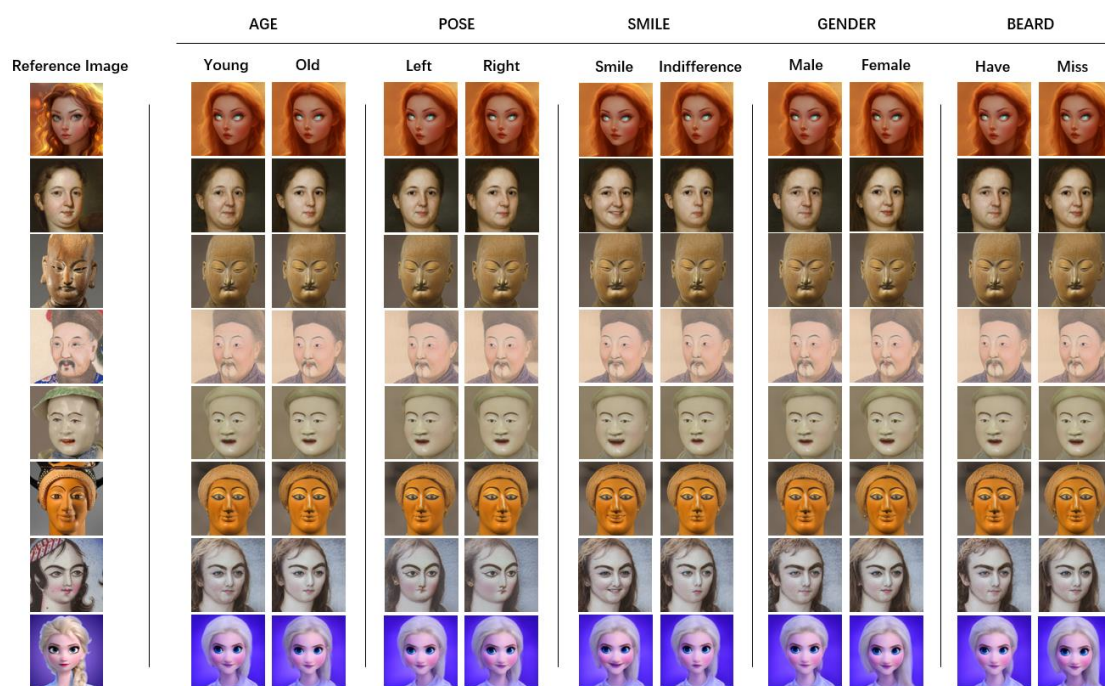


Figure 8 艺术图像编辑效果

并且，编辑的过程中图像的变化具有一定的连续性，因此我们可以通过控制特征的迁移率的连续变化即可得到一个从原始图像到目标图像连续图像，如下图所示，组合即可得到艺术图像“微笑”、“转头”或者“年龄变化”等的动画。

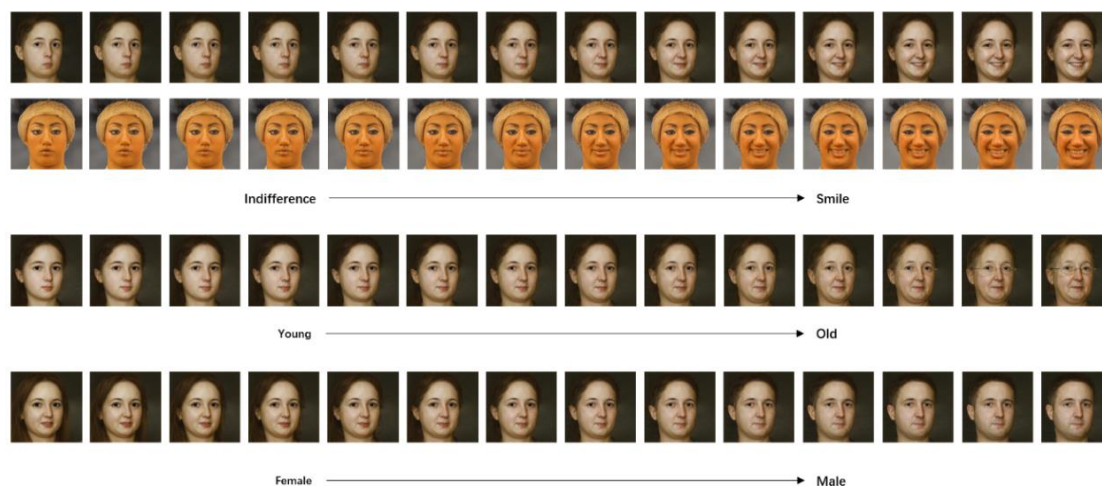


Figure 9 艺术图像连续编辑变化

同时，我们可以在同一个方向提高特征的迁移率，还可以将多个编辑向量同时作用到目标图像上，从而获得变化更多样更明显的艺术人脸（如图 10）。



Figure 10 更加多样的变化

4、总结

总体上，项目能够基本实现对面脸图像的简单属性（表情、姿态等）的编辑，并能够结合反演模型以及自适应模型拓展至指定人脸图像以及艺术人脸图像的编辑任务上，但仍存在一些不足。

图像反演中，在个别艺术人脸图像的反演重建上前后差距过大，存在失真的情况，并且由于现有预训练模型大多是基于欧美面孔训练集训练得来的，导致对一些具有中国人特征的人脸图像重建后具备一些欧美人脸特征。

图像编辑中，现有的编辑向量较少，并且由于训练集的不均匀以及语义解耦不充分，会出现在某一方向过度编辑某项属性导致另一项属性发生改变的情况（如修改年龄属性可能会出现眼镜、编辑胡须会改变性别等）。

现有编辑向量主要集中在年龄、性别、笑容、姿态等属性，对艺术人脸的编辑还有一定局限。课程结束之后可以尝试使用 StyleCLIP 等语义人脸编辑模型研究更好的 GAN 网络反演方法和挖掘属性更多样效果更好的编辑向量，提高图像的还原度以及编辑处理的自然适度。

5、组内分工

姓名	分工
张智雄	完成对 pSp 模型以及基于 SVM 的 InterfaceGAN 的文献阅读以及模型原理构建，使用 pSp 模型对图像进行反演，完成对指定人脸图像的编辑以及艺术人脸的部分编辑。
李佳阳	学习 SVM 分类器以及 GAN 网络的数学原理，并使用 InterfaceGAN 寻找微笑、年龄、姿态等属性编辑向量，完成对随机人脸图像的编辑。
王佳依	学习 SVM 分类器以及 GAN 网络的数学原理，并使用 InterfaceGAN 寻找胡须、头发、性别等属性编辑向量，完成对图像编辑的连续处理以及多种属性的编辑合成。
王诗雅	调查项目背景以及研究现状，学习 GAN 网络以及 pSp 模型的数学原理，并使用 pSp 模型验证风格迁移的必要性，完成艺术人脸的编辑。

注：风格自适应过程中，DIFA 模型的构建以及在目标图像上的测试运行在交流汇报时由张智雄、李佳阳、王佳依同学共同完成。

参考文献

- [1] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks (2014)[J]. arXiv preprint arXiv: 1406.2661, 2014, 1406.
- [2] 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. 2020: 8110-8119.
- [3] Tov O, Alaluf Y, Nitzan Y, et al. Designing an encoder for stylegan image manipulation[J]. ACM Transactions on Graphics (TOG), 2021, 40(4): 1-14.
- [4] Shen Y, Yang C, Tang X, et al. Interfacegan: Interpreting the disentangled face representation learned by gans[J]. IEEE transactions on pattern analysis and machine intelligence, 2020.
- [5] Zhang Y, Wei Y, Ji Z, et al. Towards Diverse and Faithful One-shot Adaption of Generative Adversarial Networks[C]//Advances in Neural Information Processing Systems. 2022.