

杭州电子科技大学

本科毕业设计（论文）

（2023届）

题 目 人像个性化油画生成技术应用研究

学 院 自动化(人工智能)学院

专 业 自动化(卓越工程师计划)

班 级 19062911

学 号 19061130

学生姓名 王祥旭

指导教师 周后盘

完成日期 2023年6月

诚信承诺

我谨在此承诺：本人所写的毕业设计（论文）《人像个性化油画生成技术应用研究》均系本人独立完成，没有抄袭行为，凡涉及其他作者的观点和材料，均作了注释，若有不实，后果由本人承担。

承诺人（签名）

年 月 日

摘要

人像油画是一种具有悠久历史和广泛影响的艺术形式，而将其风格融入图像的生成中则是一个具有难度和价值的研究课题，也是人工智能技术在艺术领域的一个重要应用方向。其中，人像个性化油画生成技术是利用 AI 将普通的人像照片转换为具有个性化风格的油画作品的一种技术。

本文研究的主要内容是基于生成对抗网络实现将原始人像按指定的油画人脸肖像画进行风格融合与迁移。文中通过对 DualStyleGAN 模型进行改进，提高了图像转换在油画风格效果上的结构相似性和色彩相似性。采用能凸显原图显著特征的 $W+$ 潜在空间编码更换了 DualStyleGAN 模型中的 $Z+$ 空间编码，从而增强了内容图的人脸特征在生成图中的地位。在模型训练过程中，通过预训练的 FFHQ-StyleGAN 模型进行分阶段微调，有效减少了一次性训练模型的时间成本，并且针对油画风格任务对损失函数进行调整。

最后，设计并实现了针对用户使用的网页端人像个性化油画风格生成的应用程序，为用户提供了编码器选择、图片选择与上传、风格图选择和结构与色彩风格权重等高自由度的自定义设置，可以让用户在不具备修图能力或绘画技巧的情况下，轻松地创作出自己喜欢的个性化风格图像，丰富了用户体验。

关键词：生成对抗网络，图像风格迁移，DualStyleGAN，应用软件开发

ABSTRACT

Portrait oil painting is an art genre with a long history and extensive influence. Incorporating its style into image generation is both a challenging and valuable research topic, as well as an important application direction of artificial intelligence technology in the field of art. Personalized portrait painting generation technology is a technique that utilizes AI to transform original portrait photos into personalized oil painting photos.

The main content of this paper is based on the Generative Adversarial Network (GAN) to achieve oil painting style fusion and transfer of original portraits. By improving the DualStyleGAN model, this paper has increased the structural similarity and color similarity of the image conversion effect in the oil painting style. The \mathcal{W}^+ latent space that can highlight significant features of the original image replaces the \mathcal{Z}^+ latent space in the DualStyleGAN model to enhance the status of facial features in the generated image. In the model training process, pre-trained FFHQ-StyleGAN models were used for stage-by-stage fine-tuning, effectively reducing the time cost of one-time model training, and adjusting the loss function for oil painting style tasks.

Finally, an application program for personalized portrait painting generation on the web page was designed and implemented, providing users with high customized settings such as encoder selection, raw image selection, oil painting image selection and style weights, which allows users to easily create their favorite personalized style images with no prerequisite of photo editing skills.

Keywords: Generative Adversarial Network, Image Style Transfer, DualStyleGAN, Application Software Development

目 录

1 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状	2
1.2.1 生成对抗网络研究现状	2
1.2.2 图像风格迁移研究现状	3
1.3 本文研究的主要内容及论文结构	4
2 生成对抗网络及其在图像生成的改进	6
2.1 生成对抗网络	6
2.1.1 基本原理及其模型结构	6
2.1.2 GAN 的数学表示及证明	6
2.2 StyleGAN 模型	8
2.2.1 潜在编码与映射网络	9
2.2.2 自适应实例归一化	10
2.3 StyleGAN2 模型	10
2.4 pSp 编码器	11
2.4.1 \mathcal{W}^+ 空间	12
2.4.2 \mathcal{Z}^+ 空间	12
2.4.3 优缺点	12
2.5 DualStyleGAN 模型	13
2.6 本章小结	14
3 基于 DualStyleGAN 的人脸油画风格转换方案	15
3.1 引言	15
3.2 DualStyleGAN 的基本结构	15
3.3 面部去风格化	16
3.3.1 初始化	16
3.3.2 潜在编码的优化	16
3.4 损失函数	17
3.4.1 对风格图的损失	17
3.4.2 对内容图的损失	18
3.5 改进策略	18

3.6 本章小结	19
4 模型训练与实验分析	20
4.1 数据集的准备	20
4.1.1 FFHQ 人脸数据集	20
4.1.2 人像油画数据集	20
4.2 构建 LMDB 数据	21
4.3 环境搭建与配置	21
4.4 训练流程与结果展示	22
5 应用程序的设计与实现	25
5.1 引言	25
5.2 功能需求	25
5.3 系统设计	25
5.3.1 总体架构	25
5.3.2 执行概念	27
5.4 应用实现	28
5.4.1 细化内容图预处理	28
5.4.2 细化风格转换调整	28
5.4.3 细化前后端设计逻辑	28
5.4.4 细化实现框架和布局	29
6 总结与展望	32
6.1 全文总结	32
6.2 展望	32
致谢	33
参考文献	34
附录	37

1 绪论

1.1 研究背景及意义

随着人工智能技术的不断发展及其在各个领域的应用，艺术领域也在 AI 的加持下呈现出越来越多新颖且富有创意的可能性和挑战。人像油画作为一种具有悠久历史和广泛影响的艺术形式，自然也是 AI 技术在艺术领域的一个重要应用方向。从而人像油画风格的生成也就成为一个具有难度和价值的研究课题。其中，人像个性化油画生成技术是利用 AI 将普通的人像照片转换为具有个性化风格的油画作品的一种技术。它可以让用户在不具备修图能力或绘画技巧的情况下，轻松地创作出自己喜欢的个性化风格图像。这种技术通常伴随着图像风格迁移和图像融合的发展而不断迭代。



图 1-1 油画风格图像示例

2014 年，由 Wang 等人^[1]提出的一种基于艺术真实感驱动的图像和照片的水彩画绘制系统通过设计得出的几种图像滤波器产生艺术逼真的水彩画。虽然该算法执行效率高，易于并行化，但诸如此类的传统方法都无法提取高层抽象特征，并且都仅仅是针对一种特定风格进行建模。换言之，若想将这类算法移植为油画风、漫画风等效果是不可能的，仍然需要针对某种风格重新设计相应的滤波器。然而，随着深度学习的发展越来越多研究者尝试利用现有的神经网络，在不同层次响应图片的风格和内容，通过设计合理的损失函数进行不断迭代以达到生成目标图像

的目的。这个思路在同样发布于 2014 年的另一项研究：生成对抗网络^[2](Generative Adversarial Networks, GAN) 被 Goodfellow 等人提出后达到了高潮，GAN 在图像生成方面的研究也因此吸引着大批研究学者加入。

可以发现尽管技术和方法在不断更新，但是其研究方向大都优先着重于对特殊、夸张的风格化人脸图像进行风格迁移学习，特别是对卡通、漫画等效果的迁移效果逐渐得到可观的成果。而目前大多数深度学习算法也仍只是针对某一种风格进行学习，学习整体图像的分布而进行图像处理，有的时候这在人脸图像的风格转换上效果并不够理想。此外，大多数算法尚未引入相关的指标对效果进行有效评估。

1.2 国内外研究现状

1.2.1 生成对抗网络研究现状

对于生成模型的研究，最早可以追溯到自编码器（Auto Encoder）^[3] 的提出。自然这些生成模型在数据生成方面取得了不错效果，但在图像生成方面尚有不足，比如生成图像的质量较低、分辨率不高以及具有不可控的特性等。此外，模型的泛化能力也不够强。而 2014 年 Goodfellow 等人^[1] 提出了 GAN 模型之后，其基于零和博弈的对抗训练思想以及其可观的效果吸引着大批研究学者加入。各类学者为了解决各种领域的各种问题，也在 GAN 的基础上提出了一系列的变体，GAN 也因此在不断地优化和扩展中投入到了各种使用之中。

在生成器和判别器的改进方面，Mirza 等人^[4] 提出的 CGAN (Conditional GAN)，通过在生成器和判别器中增加约束的方法控制模型往希望的方向进行生成，解决了原始 GAN 生成结果难以控制的问题。不过，CGAN 模型也存在一些诸如训练稳定性不高、图像不清晰等缺点。为了提高生成图像的质量，Denton 等人则在 CGAN 的基础上改进生成网络的结构提出了 LAPGAN^[5]，分别使用不同的结构对图像进行上采样和下采样。Radford 等人^[6] 首次在 GAN 中引入了卷积神经网络 (Convolutional Neural Networks, CNN) 的概念，二者结合提出了 DCGAN (Deep Convolutional GAN)，得益于卷积神经网络丰富的特征表征能力，GAN 的发展（特别是图像生成上的发展），得到了极大的提高，其图像丰富性和清晰度也有所提升。Zhang 等人将自注意力机制 (Self Attention) 模块引入 DCGAN 模型中，构建了 SAGAN 模型^[7]，使得生成器和判别器可以利用所有特征线索结合上下文信息来生成和判断细节，从而提高了图像的质量和多样性。

在损失函数的设计方面，Mao 等人使用最小二乘损失函数代替原始 GAN 的交叉熵损失函数，从而提出了 LSGAN (Least Squares GAN)^[8]，有效地抑制了 GAN

训练过程中可能出现的不稳定的情况。由于原始 GAN 中通过 JS 散度来衡量两个概率分布的距离，所以在训练过程中容易出现梯度消失的问题，对此，Arjovsky 等人提出了 WGAN 模型^[9]，利用 Wasserstein 距离公式来度量真实的样本数据概率与生成数据概率分布之间的距离。Gulrajani 等人针对 WGAN 模型的简单裁剪权值策略所引起的梯度消失问题，提出了一种更加平缓的梯度惩罚函数来达到 Lipschitz 连续性的条件，从而得出了 WGAN-GP 模型^[10]。

1.2.2 图像风格迁移研究现状

经历了早期传统的基于图像滤波器的图像风格迁移技术之后，发表于 2015 年的《A Neural Algorithm of Artistic Style》^[11]一文，开启了使用神经网络进行图像风格迁移学习的新浪潮。在 GAN 的加入下，特别是加入了卷积神经网络的 CGAN，Isola 等人^[12]提出了 Pix2pix 模型实现配对图像之间的风格迁移，Pix2pix 模型结合了 GAN 其他变体的优点，优化了损失函数，不容易出现训练不收敛的情况，但训练时间较长，并且对训练集的要求较高。而 Zhu 等人提出的 CycleGAN^[13]则无需配对数据集，实现了无监督学习下就可以进行风格迁移的工作。此外，Junho Kim 等人于 2020 年发表在 ICLR 上的论文^[14]中，和 CycleGAN 一样都才采用了循环一致性损失（Cycle Consistency Loss）和身份映射损失（Identity Mapping Loss），但是却提出了一种新的无监督图像到图像转换方法，通过注意力模块来关注域与域之间的关键部分，这就是 U-GAT-IT 模型。

在传统的 GAN 模型（以及基于传统 GAN 模型进行拓展的模型）中，往往难以控制生成图像的不同特征，比如姿态、身份、发色等。这些缺陷被 Karras 等人在 2019 年提出的 StyleGAN^[15]通过引入映射网络和自适应实例归一化（AdaIN）操作得到了改善。

2021 年提出的 GANs N’ Roses (GNR)^[16]为了保证内容图片和风格的分离与多样性，使用了三个判别器：一个内容判别器，用于检测生成图像的内容是否与原始图匹配；一个风格判别器，用于检测生成图像的风格是否与目标域风格匹配；一个重建判别器，用于检测重建图像是否与原始图像相似。这三个判别器构成了对抗损失，用于指导编码器和解码器的训练。但算法需要大量的训练数据和计算资源，并且可能存在过拟合或泛化能力不足的问题。

2022 年由南洋理工大学联合实验室 S-Lab 的研究员提出的 DualStyleGAN^[17]则考虑较为周全，它是一种基于样本的高分辨率人像风格迁移的方法，在 StyleGAN 的基础上添加了一个外部风格路径，用于模拟和控制目标域的风格。提出了一个三阶段渐进微调策略，用于在不损失原始域风格样式多样性的情况下，实现目标域风格的迁移。

本文将着重于对油画肖像的风格迁移的效果优化，与其他相关领域（如图像分割、图像融合等）进行交叉融合。此外致力于结合已有的风格转移方法探索在油画方面的生成性能，实现用户的个性化选择风格与风格融合的需求，让用户能够根据自己的喜好将自己或他人的照片转换为不同风格、不同品质的油画作品。让社会能够从中感受到 AI 技术在艺术领域所带来的新鲜感和美感，并增强对传统油画艺术及其历史文化背景的了解和尊重。

1.3 本文研究的主要内容及论文结构

本文研究的主要内容是基于生成对抗网络实现将原始人像按指定的油画人脸肖像画的风格进行风格迁移，做到原始图像与目标图像的融合，使得原始图像的基本风格按照目标图像的风格进行迁移。尽可能选取和改造生成模型、调整相关参数以实现生成图像与原始图像具有较高的相似度和匹配度。

本文的各章节所述内容如下：

第一章，绪论。对本文的研究背景和意义进行分析，结合国内外人脸风格迁移技术和国内外图像生成技术的现状描述了本文研究方向的意义，并且讲述了本文研究的主要内容及论文结构。

第二章，生成对抗网络及其在图像生成的改进。详细分析了典型的生成对抗网络模型的算法原理和数学证明，阐述了针对图像生成领域而基于 GAN 模型衍生的 StyleGAN 模型，包括其主要的算法思想、网络结构和实现细则。最后介绍了由 StyleGAN 改进而来的 StyleGAN2，及基于 StyleGAN2 衍生而来的 Pixel2Style2Pixel 编码技术，再引入 AgileGAN 给出的编码器改进策略，形成对图像到图像的灵活映射转换技术。为后文的 DualStyleGAN 的提出提供了坚实的基础。

第三章，基于 DualStyleGAN 的人脸油画风格改进方案。针对人像个性化油画生成问题，提出了基于 DualStyleGAN 模型的改进措施。本章主要详细分析了在 StyleGAN2 模型和 pSp 改进编码器的技术加持下，由南洋理工大学联合实验室提出的通过划分内部风格路径和外部风格路径两条路径而进行图像的风格融合迁移研究的技术：DualStyleGAN。此外，本章还在原有的 DualStyleGAN 模型基础上，特别针对油画风格迁移任务进行了改进和调整，使得人脸重建的效果更加逼近内容图中的真实人脸，提高损失函数中对内容图的损失权重以优化生成图中的人像辨识度。

第四章，模型训练与实验分析。收集模型训练所需的人脸数据集和油画风格图像数据集，并针对具体的深度学习任务将其以 LMDB 数据进行存储。本章还对模型训练的步骤进行了划分和说明，最终展现了以人脸数据集的一张内容图和五

张风格图的风格融合迁移结果。

第五章，应用程序的设计与实现。本章从一个应用程序的需求、设计和实现三个流程进行阐述，构建了整个人像个性化油画风格生成的应用程序的系统框架和具体功能的设计和实现机制，最终展现了该应用程序完整的页面布局和功能使用画面。

第六章，总结与展望。对全文工作进行总结，针对研究过程中遇到的一些问题提出了自己的看法和见解，最后对未来相关技术研究进行展望。

2 生成对抗网络及其在图像生成的改进

2.1 生成对抗网络

为了进行图像风格转换的工作，特别是能从目前最前沿的技术中进行学习和分析，我们需要对最基本的生成对抗网络模型进行基本的原理分析和数学推导。

2.1.1 基本原理及其模型结构

生成对抗网络（Generative Adversarial Networks, GAN）主要由两个神经网络构成，分别是生成器 G （Generator）和判别器 D （Discriminator）。生成器的目标是生成服从于真实数据分布的样本，而判别器的目标则是区分输入的数据是真实数据还是生成数据。可见，二者的训练目标的存在对抗性的，而二者也正是通过这样不断的对抗训练——极小极大博弈（minimax game），使得模型能在所有可能的 G 和 D 中，求出唯一均衡解：最终 G 足以生成与训练样本相同分布的数据，相对地， D 的判别结果则稳定在了 $1/2$ 。

根据 GAN 的基本原理，我们得以绘制出其基本的模型结构。

对于服从于任意的某种分布 $P_z(z)$ 的噪声 z ，经过生成器 G 后得到数据 $G(z)$ 。我们假设 $G(z)$ 服从的分布为 $P_G(G(z))$ ，而真实数据（即数据集给出的数据）服从于 $P_{data}(x)$ ，所以我们需要训练 G 和 D 使得两种分布尽可能地一致。

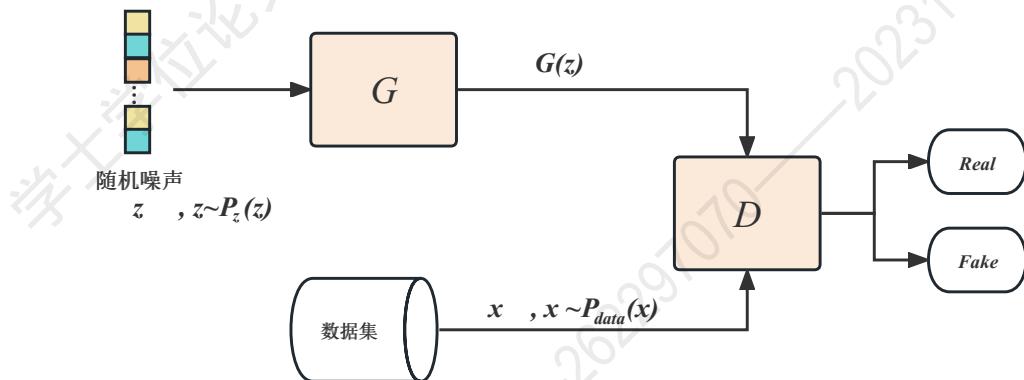


图 2-1 GAN 网络的模型结构

2.1.2 GAN 的数学表示及证明

为了指导实现 G 与 D 的参数调整，我们需要定义适当的价值函数 $V(G, D)$ 用来训练，Wang 等人则是通过引入信息熵的概念实现。为此，需要简单介绍以下信

息论中熵的相关定义和定理。

定义 2.1. 对于离散随机变量 X 的一个事件 $X = x_i$, 定义其 **信息量** 为:

$$I(x_i) = -\log(p(x_i)), \quad \forall i, x_i \in \chi, i = 1, 2, \dots, n \quad (2-1)$$

其中 $\chi = \{x_1, x_2, \dots, x_n\}$ 为随机变量 X 可能的取值范围。 $p(x_i) = P\{X = x_i\}$ 表示事件 $X = x_i$ 发生的概率。

定义 2.2. 离散随机变量 X 的 **信息熵** 被定义为其事件信息量的期望:

$$H(X) = -\mathbb{E}(\log p(x)) = -\sum_{i=1}^n p(x_i) \log(p(x_i)) \quad (2-2)$$

规定: 当 $p(x_i) = 0$ 时, $p(x_i) \log(p(x_i)) = 0$ 。

若信息熵 $H(X)$ 越小, 则说明 X 的纯度越高, 事件发生的情况更稳定。

定义 2.3. *Kullback-Leibler 散度* 或 **信息散度**, 又被称作 **相对熵**, 是两个概率分布间差异的非对称性度量。对于概率分布 P, Q , 它们的 *KL 散度* 为

$$KL(P||Q) = \sum_{i=1}^n p(x_i) \log \left(\frac{p(x_i)}{q(x_i)} \right) \quad (2-3)$$

定义 2.4. 为了解决 *KL 散度* 非对称性的问题, 在 *KL 散度* 的基础上又提出了 **JS 散度** (*Jensen-Shannon Divergence*), 作为另一种度量两个概率分布相似度的方法:

$$JS(P||Q) = \frac{1}{2}KL(P||M) + \frac{1}{2}KL(Q||M) \quad (2-4)$$

其中, $M = \frac{P+Q}{2}$ 。

不难得出, 判别器对真实数据的信息熵 $\mathbb{E}_{x \sim P_{data}(x)}[\log D(x)]$ 越大则表示判别器 D 识别真实数据的能力越强, 而判别器对生成数据的信息熵 $\mathbb{E}_{z \sim P_z(z)}[\log D(G(z))]$ 越小则表示判别器 D 将假数据误判成真数据的可能性越低; 相反, 若它的值越大则说明 D 容易误判, 也即 $G(z)$ 更接近真实数据。于是, 可得到优化目标函数如下:

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim P_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim P_z(z)}[\log(1 - D(G(z)))] \quad (2-5)$$

其中, 将 $\mathbb{E}_{z \sim P_z(z)}[\log(1 - D(G(z)))]$ 代替 $\mathbb{E}_{z \sim P_z(z)}[\log D(G(z))]$ 可以更好统一函数。

上述目标函数的求解过程主要分为两步：

- 固定 G 优化判别器 D ，使得最终得到的判别器 $D^* = \arg \max_D V(G, D)$ 。
- 固定 D 优化生成器 G ，使得最终得到的生成器 $G^* = \arg \min_G V(G, D^*)$ 。

为此，首先我们需要对价值函数 $V(G, D)$ 进行如下的化简：

$$\begin{aligned} V(G, D) &= \mathbb{E}_{x \sim P_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim P_z(z)} [\log(1 - D(G(z)))] \\ &= \int_x P_{data}(x) \log D(x) dx + \int_z P_z(z) \log(1 - D(G(z))) dz \\ &= \int_x P_{data}(x) \log D(x) dx + \int_x P_G(x) \log(1 - D(x)) dx \\ &= \int_x [P_{data}(x) \log D(x) + P_G(x) \log(1 - D(x))] dx \end{aligned}$$

从而求解 $\max_D V(G, D)$ 等价于求解该积分中被积函数的最大值。记被积函数为 $\Phi(D)$ ，在 $\Phi(D)$ 为连续可微函数时，其极值点可令其导函数为零求得。有

$$\begin{aligned} \frac{d\Phi(D)}{dD} &= \frac{P_{data}(x)}{D(x)} - \frac{P_G(x)}{1 - D(x)} = 0 \\ \Rightarrow D^*(x) &= \frac{P_{data}(x)}{P_{data}(x) + P_G(x)} \end{aligned}$$

将 D^* 代入 $V(G, D)$ 以便求解 $\min_G V(G, D^*)$ 。此时可进一步化简：

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

故当且仅当 $P_{data}(x) = P_G(x)$ 时， $V(G, D^*)$ 取得最小。这足以证明在理想情况下，最优化价值函数可以训练出与真实分布一致的生成器 G^* 。

2.2 StyleGAN 模型

为了应对 GAN 模型对人脸生成的不稳定性和图像质量的低分辨率问题，由 Karras 等人在 2017 年提出的渐进式增长的生成对抗网络 (PG-GAN)^[18] 应运而生。

它使用了渐进式增长（Progressive Growing）的方法，首先从低分辨率开始训练，然后逐步增加分辨率形成新的层：从 4×4 到 8×8 、 16×16 一直到 1024×1024 ，这些改进提高了 GAN 生成图像的质量和稳定性。PG-GAN 还通过将每个层生成的小图像标准差作为额外的特征输入到判别器 D 中，这也增强了判别器的多样性。

StyleGAN^[15] 在 PG-GAN 的基础上，通过引入映射网络 f （Mapping Network）将原来的潜在向量 $\mathbf{z} \in \mathcal{Z}$ 映射到更具有解耦性的 \mathcal{W} 空间，再依次将新的潜在向量 $\mathbf{w} \in \mathcal{W}$ 加入到生成网络中，实现了对图像不同层次特征的有效控制。StyleGAN 的整个生成器架构如下图所示。

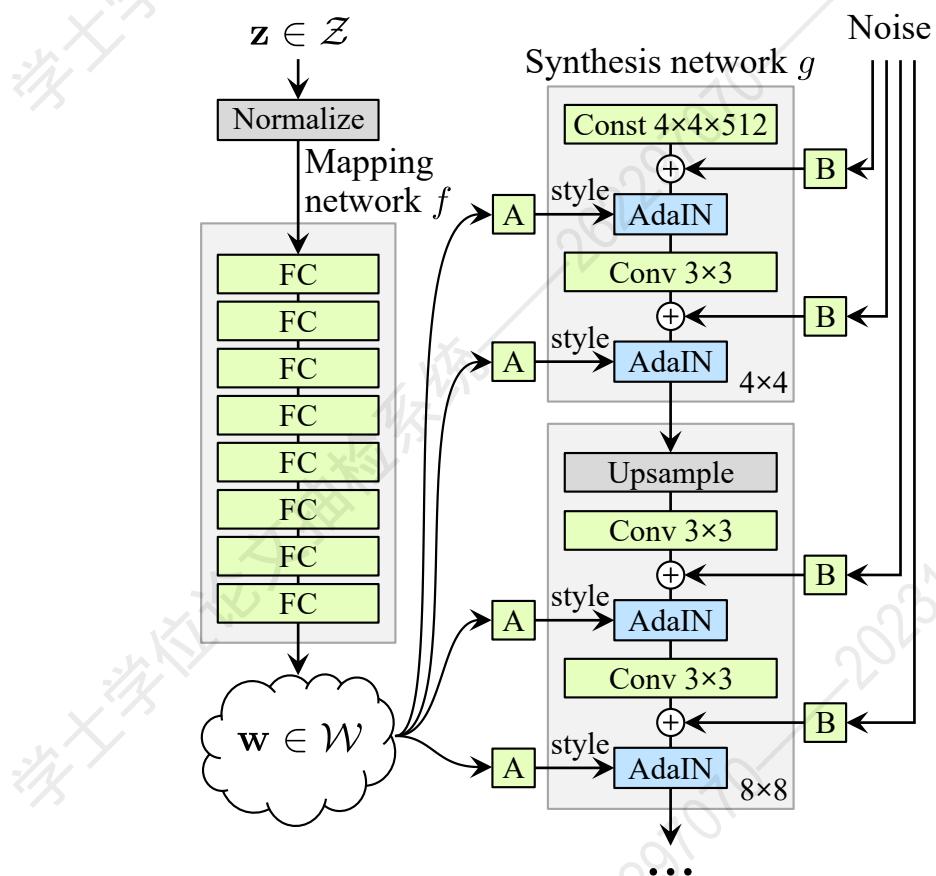


图 2-2 StyleGAN 的生成器结构

2.2.1 潜在编码与映射网络

在深度学习中，为了便于对大数据进行处理，通常我们期望能够掌握数据与数据之间的关联性；能够抓住目标对象的主要特征；能够在对数据进行压缩/编码（Encoding）之后也能恢复/解码（Decoding）出原始数据。这种需求在图像分类和

图像生成模型中同样重要，我们称通过神经网络编码或提取出来的图像特征数据为潜在编码（Latent Code）。

在 GAN 生成模型中，输入的随机噪声 \mathbf{z} 如果直接用于生成器的输入，那么生成的图像往往是不可控的，为了解决这个问题 StyleGAN 模型提出在生成之前先将属于原始空间 \mathcal{Z} 的 \mathbf{z} 映射成 \mathbf{w} ，并且保证 \mathbf{w} 所在的 \mathcal{W} 空间能对之后的生成起到控制作用，即 $\mathbf{z} \rightarrow \mathbf{w}$ 的映射网络需要实现数据的解耦。StyleGAN 采用 8 个多层感知网络（Multi-layer Perceptron Neural Networks）将 512 维的 \mathbf{z} 向同为 512 维的 \mathbf{w} 进行映射，并且提出了感知路径长度（Perceptual Path Length, PPL）和线性可分性（Linear Separability）用于衡量 \mathcal{W} 空间的解耦性。

2.2.2 自适应实例归一化

在经典的图像风格迁移算法中，自适应实例归一化^[19]（Adaptive Instance Normalization, AdaIN）的提出实现将输入的风格、纹理迁移到内容图（Content Image）中并保留内容图的主体结构。公式如下：

$$\text{AdaIN}(\mathbf{x}, \mathbf{y}) = \sigma(\mathbf{y}) \frac{\mathbf{x} - \mu(\mathbf{x})}{\sigma(\mathbf{x})} + \mu(\mathbf{y})$$

其中， \mathbf{x}, \mathbf{y} 分别表示内容图和风格图经过 VGG 卷积神经网络得到的特征。

在 StyleGAN 中，参考了 AdaIN 的思想，映射得到的潜在编码 \mathbf{w} 通过可训练的模块“A”进行仿射变换，得到与每一层的每一个特征图（Feature Map） \mathbf{x}_i 一一对应的 $\mathbf{y}_{s,i}$ 和 $\mathbf{y}_{b,i}$ ，使得潜在编码得以注入每一层中，实现风格控制。其公式如下：

$$\text{AdaIN}(\mathbf{x}_i, \mathbf{y}) = \mathbf{y}_{s,i} \frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} + \mathbf{y}_{b,i} \quad (2-6)$$

其中， $\mathbf{y} = (\mathbf{y}_{s,i}, \mathbf{y}_{b,i})$ 。

2.3 StyleGAN2 模型

AdaIN 的引入虽然使得 StyleGAN 模型在图像生成方面得到了有效的控制，但是研究者也发现 StyleGAN 的生成图像中会存在水滴状的奇怪点，并且这在特征图中更加明显。实验证明，当对合成网络中不断对特征图归一化的 AdaIN 被取消使用时，这种怪异现象得到了解决。在优化模型架构和提出一些新的指标和思考后，StyleGAN2 模型^[20] 也就应运而生了。

StyleGAN2 的生成器框架如图2-3 (d) 所示，其中 (a,b) 是第一代 StyleGAN 的实现细节。作者认为，由于 AdaIN 模块是对每一层的特征图都进行处理，所以层

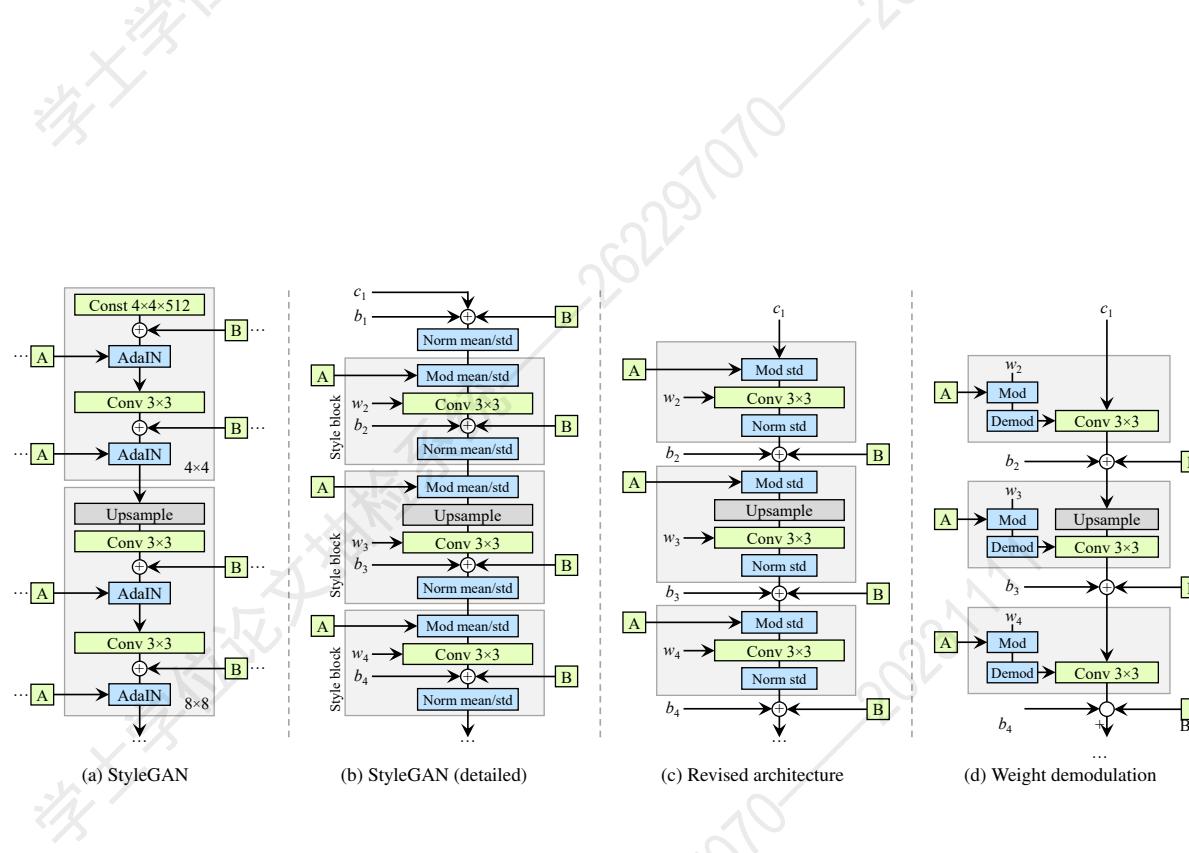


图 2-3 StyleGAN2 的生成器结构

与层之间可能存在的关系可能被忽略，从而将 AdaIN 移除转而只采用 y 的标准差时，水滴问题得到了消除。同时，作者还发现通过“B”模块注入到合成网络的噪声即使是放在层与层之间再注入，也对结果没有影响，而且还能简化一部分操作使得模型框架也更简洁、生成速度更加迅速，如(c)所示。另外，通过分析可知，对 w 进行仿射得到的向量作用给特征图之后再进行卷积的行为其实就是要对其进行了一个类似于放缩的操作，因此整个调制过程可以放到卷积层处理部分结合放缩，如(d)所示。

2.4 pSp 编码器

在 StyleGAN2 的论文^[20] 中，作者提出了能否再将一幅图映射回到潜在编码的问题，如果可行则可以通过直接对潜在编码进行修改从而实现对生成图像的局部微调。这项工作可以说将 StyleGAN 的纯图像生成任务拓展到了真正的图像风格迁移和图像转图像领域。

如果先将真实图像进行某种编码以形成潜在编码，然后再利用该编码输入给 StyleGAN 让其生成图像，那么以真实图像与生成图像的差距作为指标进行编码方式的训练学习的话，显然能够实现特定的编码器训练。但实验证明，这样的处理方法需要花费大量的训练资源并且最终重建的效果并不理想。于是，一种借助特征金字塔网络结构（Feature Pyramid Networks, FPN）来生成三个不同层级的特征最后再组合成潜在编码的框架设计方式在 2021 年被 Richardson 等人提出^[21]，命名为 Pixel2StylePixel，简称 pSp。

2.4.1 \mathcal{W}^+ 空间

类比 StyleGAN 模型的逐层生成思想，pSp 编码器将输入图像通过 FPN 特征金字塔的 ResNet 的 Skip 结构映射到 18×512 的空间之中，与 StyleGAN 提到的 \mathcal{W} 空间对应，pSp 中所映射到的空间称为 \mathcal{W}^+ 空间。并且，期间分支出了三个层级：Small Feature（对应 1 ~ 2 层）、Medium（对应 3 ~ 6 层）、Largest（对应 7 ~ 18 层）。每一个层级利用 map2style 结构将卷积层降维到 1×512 形成潜在编码，使其得以通过 StyleGAN 的模块 “A” 进行仿射变换并注入到 StyleGAN 的合成网络中。pSp 的网络结构如下图所示。

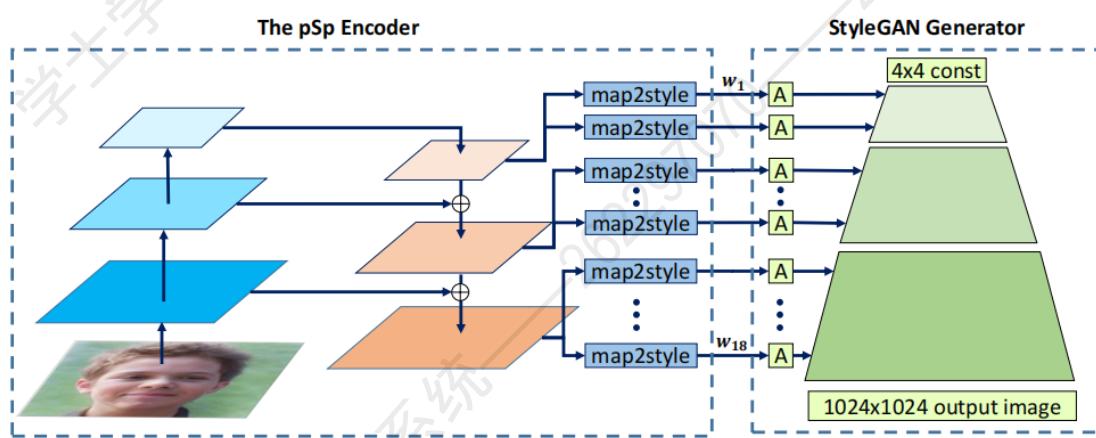


图 2-4 pSp 网络结构

2.4.2 \mathcal{Z}^+ 空间

与 pSp Encoder 同年的另一项研究：AgileGAN^[22] 的论文中，作者针对图像风格迁移这样具体的任务，提出了比 \mathcal{W}^+ 空间更适合图像到图像转换的 \mathcal{Z}^+ 空间。作者认为在做图像到图像转换，特别是不同风格的图像之间进行变换时，如果直接将原始图像映射到 \mathcal{W}^+ 空间将会丢失部分原始图像风格的隐性特征，即不一定适用于风格迁移工作。此外作者指出 \mathcal{W}^+ 空间并非高斯分布，而在训练优化的过程中若是让基于高斯分布的 \mathcal{Z} 进行编码得到的潜在空间 \mathcal{Z}^+ 的分布也服从于高斯分布，那么最终的生成效果比纯 pSp 编码要好得多。实验证明 \mathcal{Z}^+ 空间比原始的 \mathcal{W}^+ 空间更能抵抗与面部无关的背景细节和扭曲的形状。

2.4.3 优缺点

得益于 StyleGAN 的许多优良性质，pSp 也继承和发展了它的部分优点。pSp 的主要优点有：

1. 可以直接生成一系列的潜在编码（用于 StyleGAN 的 Style 输入），而不需要额外的优化；
2. 可以直接利用预训练的 StyleGAN 模型就能进行生成，节省了训练的时间和资源；
3. 可以保留输入图像的结构和语义信息的同时添加对应的风格图特征（即图像风格迁移）。

此外，pSp 编码器的缺点也很明显，主要有：

1. 依赖于预训练的 StyleGAN 模型，输出质量和多样性受到生成模型的限制；
2. 无法处理一些复杂或不常见的图像风格，比如动画或绘画风格；

2.5 DualStyleGAN 模型

如今，基于图像风格转换和图像到图像转换的人脸风格转换技术已得到了广泛研究，而目前最先进的人脸生成模型 StyleGAN 在生成高分辨率艺术肖像方面也非常具有前途。具体来说，StyleGAN 可以进行有效微调（fine-tune），通常只需要数百张人像内容图和数小时的训练时间，就能将其生成空间从人脸域转化为艺术画域。与图像风格转换和图像到图像转换模型相比，它在质量、图像分辨率、数据要求和效率方面都表现出了极大的优势。

虽然将 StyleGAN 直接用于图像风格迁移的策略行之有效，但其只能学习整体的分布转换，无法处理基于样本的风格迁移。例如，对于为了生成固定漫画风格而学习训练的 StyleGAN 模型而言，一张带有笑脸的真实人像将可能映射到嘴巴很夸张的图像风格上，这是因为 StyleGAN 主要会将其映射到具体风格域中最邻近的一张脸上。所以用户难以简单地通过改变自己的面部大小和表情来生成他们喜欢的艺术风格图像。尽管 StyleGAN 通过潜在编码的交换提供了固定的基于样本的单域风格转换但源域和目标域之间的不对齐问题可能导致在风格混合过程中出现不需要的伪影。

为了应对这些挑战，DualStyleGAN^[17] 提出了基于样本的双路风格建模和风格迁移控制策略。DualStyleGAN 保留了 StyleGAN 的控制源域风格样式的能力，文章中称其为的内在风格路径（Intrinsic Style Path），同时额外增加了一条外在风格路径（Extrinsic Style Path）用于建模和控制目标域的风格样式，从而更加自然丝滑地贴合标准风格迁移模型中的内容路径（Content Path）和样式路径（Style Path）。

2.6 本章小结

本章主要详细分析了生成对抗网络的基本原理和数学证明，阐述了基于原始的生成对抗网络模型在图像生成领域的风格控制技术和高分辨率生成技术，从而衍生的 StyleGAN 模型。对 StyleGAN 的说明包括其主要的算法思想、网络结构和实现细则，涉及潜在编码的定义、映射网络的设计以及自适应实例归一化的引入。此外针对 StyleGAN 的不足之处，介绍了由其改进而来的 StyleGAN2 模型。最后，由 StyleGAN2 衍生而来的 Pixel2Style2Pixel 编码技术（pSp 编码）给出了由纯图像生成到图像编码的全新映射技术，这在引入 AgileGAN 给出的编码器改进策略之后，形成对图像到图像的灵活映射转换技术。为后文的 DualStyleGAN 的提出提供了坚实的基础。

3 基于 DualStyleGAN 的人脸油画风格转换方案

3.1 引言

在第二章中介绍到的 DualStyleGAN 模型结合了 StyleGAN2 与 pSp 模型，构建了双路风格的图像风格迁移网络结构。该模型在夸张讽刺画和迪士尼风格动画上的风格迁移效果优异，但仍然缺乏对油画风格的探讨，而且并没有引入相关的指标对效果进行有效评估。因此，本研究将着重于对油画肖像的风格迁移的效果优化，致力于结合已有的风格转移方法探索在油画方面的生成性能，实现用户的个性化选择风格与风格融合的需求。

3.2 DualStyleGAN 的基本结构

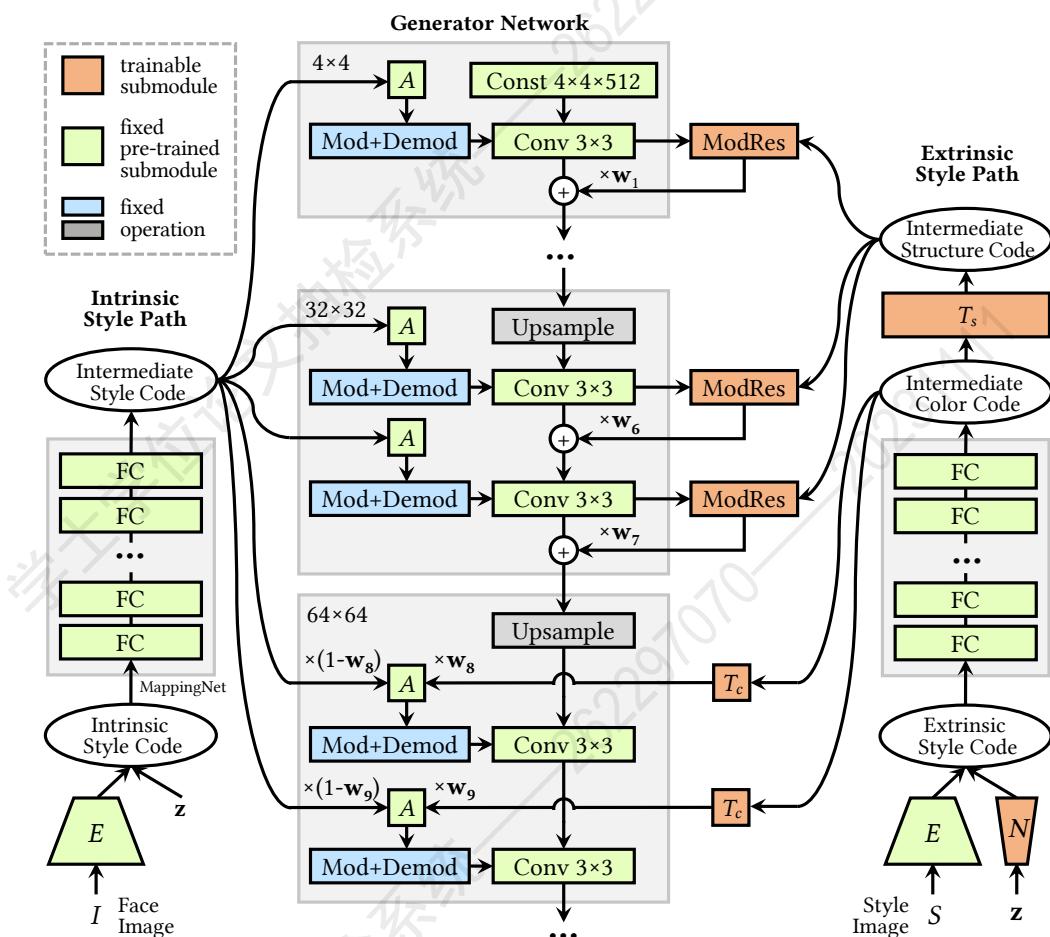


图 3-1 DualStyleGAN 网络结构

DualStyleGAN 的生成器 G 部分主要如图3-1所示，内部风格路径和生成网络构成了标准 StyleGAN 结构。内部风格编码（其实就是 StyleGAN 中的潜在编码）主要来源于真实人脸图像 I 经过编码器 E 得到的潜在编码 $\mathbf{z}^+ = E(I)$ 、随机的高斯噪声 $\mathbf{z} \in \mathbb{R}^{1 \times 512}$ 或者是某种风格肖像图通过去风格化之后而保留的人脸潜在编码 \mathbf{z}_i^+ 。

在外部风格路径中，则是将风格图 S 经过编码器 E 得到的潜在编码 $\mathbf{z}_e^+ = E(S)$ 作为外部风格编码，或者是随机的高斯噪声经过采样网络 N 映射到外部风格分布得到的潜在编码。外部风格路径的设计同样遵循了 StyleGAN 的 18 层风格层次结构：前 7 层通过线性层 T_s 并经过添加 AdaIN 的残差模块：ModRes 模块以得到残差特征之后再加到原特征中从而进行结构上的风格，而后 11 层则是通过线性层 T_c 来控制颜色上的风格。

最终，整个风格转换算法通过生成器得到所需的生成图像为 $G(E(I), E(S), \mathbf{w})$ ，其中 $\mathbf{w} \in \mathbb{R}^{18}$ ，其作用是灵活控制内部风格和外部风格的占比的权重。规定， $g(\mathbf{z}) = G(\mathbf{z}, \cdot, \mathbf{0})$ ，即单纯由随机噪声 \mathbf{z} 生成的人脸图， $\mathbf{w} = \mathbf{0}$ 表明外部风格不介入合成网络。

3.3 面部去风格化

面部去风格化旨在从目标风格图 S 中恢复其蕴含的逼真面部，即我们希望在真实人脸域中找到它们的合理对应物，以学习风格图中人脸的结构风格，实现更好的结构风格控制。由于这风格域和人脸域可能有很大的外观差异，所以这带来了一个不小的挑战，为此，DualStyleGAN 提出了一种多阶段去风格化方法来逐渐增强风格图的真实感。

3.3.1 初始话

将人脸风格图 S 通过编码器 E 映射到 StyleGAN 的潜在空间中，这里使用的是 pSp 编码器并对其进行修改，将 FFHQ 人脸到 \mathcal{Z}^+ 空间中，因为该空间比原始的 \mathcal{W}^+ 空间更能抵抗与面部无关的背景细节和扭曲的形状。当我们对使用 FFHQ 训练的 StyleGAN 输入风格图 S 之后，再用生成器 g 对其进行重建得到 $g(z_e^+)$ ，这使得其色彩风格和结构风格得到了很好的捕捉。其中， $z_e^+ = E(S) \in \mathbb{R}^{18 \times 512}$ 。

3.3.2 潜在编码的优化

在以往进行风格转换任务的工作中，通常是通过优化生成器 g 的潜在编码来重建图像，并将该编码应用于微调后的具有风格化生成能力的生成器 g' 中，从而实现人脸风格化。这里为了进行去风格化处理，可以采取相反的步骤：优化 g' 的

潜在编码 z^+ ，用新的正则化项重建 S ，并将得到的 \hat{z}_e^+ 应用到 g 中从而获得风格化版本。

$$\hat{z}_e^+ = \arg \min_{z^+} \mathcal{L}_{perc}(G'(z^+), S) + \lambda_{ID} \mathcal{L}_{ID}(G'(z^+), S) + \|\sigma(z^+)\|_1 \quad (3-1)$$

其中 \mathcal{L}_{perc} 是感知损失， \mathcal{L}_{ID} 是保持面部恒等性的身份损失（将在下一节详细介绍）， (z^+) 是 z^+ 中 18 个不同 512 维向量的标准差。此处取 $\lambda_{ID} = 0.1$ 。

3.4 损失函数

DualStyleGAN 的损失函数主要包含 GAN 模型的原始损失、感知损失（Perceptual Loss）、上下文损失（Contextual Loss）、特征匹配损失（Feature Matching Loss）以及身份损失（Identify Loss）。见式3-2。

$$\min_G \max_D \lambda_{adv} \mathcal{L}_{adv} + \lambda_{perc} \mathcal{L}_{perc} + \mathcal{L}_{sty} + \mathcal{L}_{con} \quad (3-2)$$

其中， \mathcal{L}_{adv} 即为 GAN 原始的对抗损失（见式2-5）， \mathcal{L}_{perc} 为 $G(\mathbf{z}_i^+, \mathbf{z}_e^+, \mathbf{1})$ 与输入的风格人脸图 S 之间的感知损失，即对外部风格路径中的面部去风格化的特征相似性能进行评估的损失函数。其中 $\mathbf{w} = \mathbf{1}$ 表明 \mathbf{z}_i^+ 的色彩、细节纹理等不参与损失计算，在合成网络中的比例为零。

3.4.1 对风格图的损失

DualStyleGAN 中的风格损失函数主要由上下文损失 \mathcal{L}_{CX} 和特征匹配损失 \mathcal{L}_{FM} 构成。

特征匹配损失简单来说就是将两张图 X, Y 经过 VGG19 提取特征之后直接进行 MSE Loss。而上下文损失由于在不对齐（Non-Align）数据中进行特征相似评估时展现出较好的效果^[23]，此处也能用于对 GAN 模型的训练之中。其核心是利用余弦距离进行度量，经过归一化和放大增强后得到两张图 X, Y 之间的上下文相似性： $CX(X, Y)$ 。从而我们可以定义 X, Y 在第 l 层的特征图 $\phi^l(X), \phi^l(Y)$ 的上下文损失：

$$\mathcal{L}_{CX}(X, Y, l) = -\log(CX(\phi^l(X), \phi^l(Y))) \quad (3-3)$$

为使风格图在无内容图/采用随机噪声输入的情况下获得良好的生成效果，最

终我们得到了风格图相关的损失函数：

$$\mathcal{L}_{\text{sty}} = \lambda_{\text{CX}} \mathcal{L}_{\text{CX}}(G(\mathbf{z}, \mathbf{z}_e^+, \mathbf{1}), S) + \lambda_{\text{FM}} \mathcal{L}_{\text{FM}}(G(\mathbf{z}, \mathbf{z}_e^+, \mathbf{1}), S) \quad (3-4)$$

3.4.2 对内容图的损失

为了约束生成图像得到的人脸尽可能与输入的内容图相似，DualStyleGAN 引入了人脸识别中的经典 ArcFace Loss 作为内容图的身份损失，衡量由随机噪声 \mathbf{z} 和外部风格路径的 \mathbf{z}_e^+ 生成得到的图像与纯面部重建得到的图像中，人脸的辨识度。最后，为了防止过拟合、增加模型的泛化性，还可以在损失函数之后加上对权重矩阵进行 L_2 正则化。最终内容图的损失函数如下：

$$\mathcal{L}_{\text{con}} = \lambda_{\text{ID}} \mathcal{L}_{\text{ID}}(G(\mathbf{z}, \mathbf{z}_e^+, \mathbf{1}), g(\mathbf{z})) + \lambda_{\text{reg}} \|W\|_2 \quad (3-5)$$

3.5 改进策略

对于油画风格的图像风格迁移任务来说，我们需要最终的结果图中的色彩风格更接近油画风格图，而人脸的结构风格则希望更接近内容图。由于油画风格图总体相对于抽象派画作来说其图像中的人物更接近真实人物，因此对其进行的风格图去风格化和对风格图人脸的结构融合调整的需求并不是很大，所以为了得到更好的油画风格生成效果，我们可分别对损失函数、pSp 编码器、风格权重来对 DualStyleGAN 进行改进，具体的改进措施如下：

1. 训练模型时选取偏高的 λ_{con} 和 λ_{ID} ；
2. 生成网络中，风格权重 \mathbf{w} 的前 7 层设置较低的权重以降低风格图人脸结构对生成图的影响；
3. 生成网络中，风格权重 \mathbf{w} 的后 11 层设置较高的权重以提升风格图的色彩结构对生成图的影响；
4. 训练能更好还原人脸的 $\mathcal{W}+$ 空间 pSp Encoder 替换 AgileGAN 的 $\mathcal{Z}+$ 空间来对内容图进行编码和解码。

改进后的人脸重建效果如下图所示，可见比原始 DualStyleGAN 模型的重建上能更接近内容图的人脸。



图 3-2 DualStyleGAN 人脸重建改进对比

3.6 本章小结

本章主要详细介绍了在 StyleGAN2 模型和 pSp 改进编码器的技术上，通过划分两条路径：内部风格路径和外部风格路径而进行图像的风格融合迁移研究的技术，称为 DualStyleGAN。解决了直接将 StyleGAN 用于图像风格迁移但只能学习整体的分布转换，无法处理基于样本的风格迁移任务的问题。

本章对 DualStyleGAN 模型的生成网络框架进行了详细的解释，阐明了 DualStyleGAN 在面部去风格化和损失函数的具体设计。此外，还在原有的 DualStyleGAN 模型基础上，特别针对油画风格迁移任务进行了改进和调整，包括利用原始 pSp 编码的潜在空间 $\mathcal{W}+$ 替换 DualStyleGAN 集成的 $\mathcal{Z}+$ 使得人脸重建的效果更加逼近内容图中的真实人脸，提高损失函数中对内容图的损失权重以优化生成图中人像的识别率等。

4 模型训练与实验分析

4.1 数据集的准备

为了对 DualStyleGAN 模型中的 StyleGAN 部分进行预训练，我们首先需要实现对现实人脸图像，也就是内容图的潜在编码映射和解码训练，可以采用 CV 中很经典的 FFHQ 人脸数据集进行训练。此外，对于外部风格路径的训练，我们需要找到目标风格的样本图像，这里我们采用的自然是人脸油画风格的图像。

4.1.1 FFHQ 人脸数据集

FFHQ (Flickr-Faces-HQ) 是一个从 Flickr 上爬取并通过 dlib 裁剪得到的高质量人脸数据集，包含有七万张 1024×1024 分辨率的 PNG 格式的高清人脸图像，由 StyleGAN 的研究团队收集并公布在 Github 中，在该团队后续的 GAN 图像生成质量的研究^[20] 中被使用。



图 4-1 FFHQ 人脸数据集预览

FFHQ 在人脸属性上拥有非常多的变化，拥有不同的性别、肤色、表情、发型、年龄等，还包括普通眼镜、帽子等多种人脸周边的配件，图像背景上也丰富多样且差异明显。

4.1.2 人像油画数据集

MetFaces (Metropolitan Faces) 是从肖像艺术作品中提取的人脸图像数据集，该数据集由 1336 张分辨率为 1024×1024 的高质量 PNG 图像组成，公布在 Github

中，被收集用于对有限数据训练生成对抗网络^[24]的研究，图片数据集主要取自 Metropolitan Museum of Art Collection API 并通过智能裁剪形成。

4.2 构建 LMDB 数据

LMDB 的全称为 Lightning Memory-Mapped Database，是一种利用内存映射技术，实现高效数据读取和写入的 key-value 数据库，可以将各类原始数据转换为统一的 key-value 存储。由于 LMDB 可以将图片音频等数据直接存储在内存中，而非从硬盘读取，所以在深度学习相关领域，尤其是计算机视觉任务中，使用 LMDB 可以加速图片的读取速度。

下面是在 Python 中构建 LMDB 数据的简单框架：

```
1 import lmdb
2 import torch
3
4 # 创建LMDB环境，指定路径和大小
5 env = lmdb.open('mylmdb', map_size=1099511627776)
6
7 # 创建一个写事务，将数据（键值对）写入LMDB数据库
8 with env.begin(write=True) as txn:
9     # 将键值对写入数据库，key和value可自定义
10    txn.put(key.encode('ascii'), value)
11
12 # 使用torchvision.datasets.LMDBDataset类来加载LMDB数据库，指定路径
13 # 和转换函数
14 dataset = torchvision.datasets.LMDBDataset('mylmdb', transform)
```

4.3 环境搭建与配置

本文使用了由 Google 开发的一种托管式 Jupyter 笔记本服务：Google Colaboratory（简称 Colab），可帮助传播机器学习训练和研究成果，用户可直接通过浏览器编写和执行 Python 代码，并且可以免费使用 GPU/TPU 计算资源。在经过 Colab 对模型进行训练后，移植到本地完成部署工作，本地的硬件配置如下表所示。

表 4-1 本地实验硬件环境配置

计算机型号	OMEN by HP Laptop 15-dc1xxx
CPU	Intel(R) Core(TM) i5-9300H
内存	8 GB
显卡	GeForce GTX 1660 Ti
外存	512GB + 1T

4.4 训练流程与结果展示

根据改进的 DualStyleGAN 模型训练需求，我们将其训练流程大致划分为四个阶段。绘制得到的完整的流程图如下：

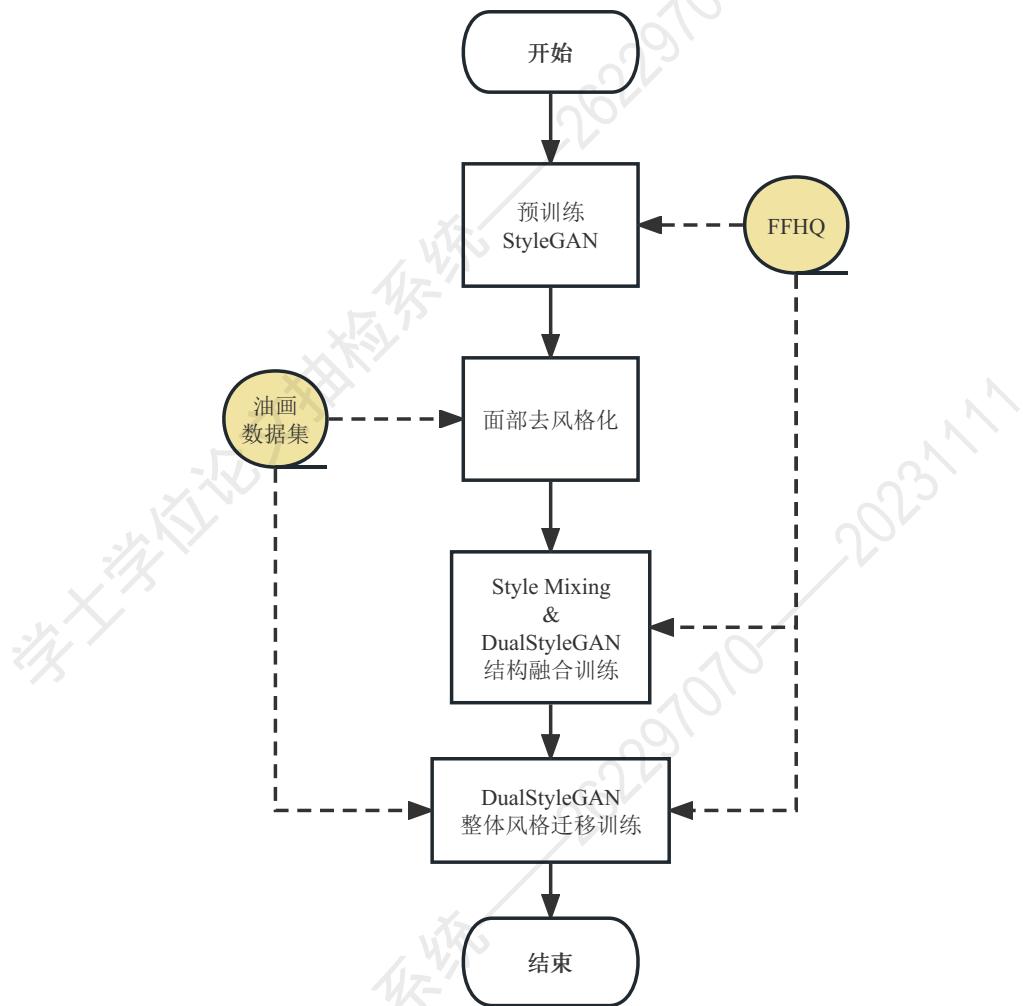


图 4-2 训练流程图

每个训练阶段的具体内容如下：

第一阶段：FFHQ-StyleGAN 的预训练。该阶段主要是利用 FFHQ 人脸数据集实现图像到 \mathcal{Z}^+ 空间的编码映射，主要参考 pSp Encoder 的特征金字塔网络结构。此外预训练模型需保证可通过 StyleGAN2 再还原出人像，该训练过程的损失函数由 GAN 原始损失和感知损失构成：

$$\min_G \max_D \lambda_{\text{adv}} \mathcal{L}_{\text{adv}} + \lambda_{\text{perc}} \mathcal{L}_{\text{perc}} (g(E(I)), I) \quad (4-1)$$

第二阶段：风格图的去风格化训练。该阶段主要是训练出合理的潜在编码，使得其能得以表征去风格化之后原风格图 S 中的人脸特征，为后续如何变形和抽象内容图的面部结构来模仿 S 提供了有效的监督。训练的优化目标函数见式3-1。

第三阶段：人脸域内间的结构迁移训练：该阶段主要是训练外部风格路径在面部结构方向上的迁移融合效果。主要训练方法是利用 Style Mixing^[15] 方法生成融合图，再利用 DualStyleGAN 的内外部风格路径进行同样的融合操作，得到生成图，通过二者的感知损失训练 DualStyleGAN 的融合能力。

Style Mixing 的操作如下：先进行随机采样得到 \mathbf{z}_1 和 \mathbf{z}_2 ，然后通过做如下融合处理：在 StyleGAN 的合成网络中，前 l 层注入的潜在编码设为 \mathbf{z}_1 ，后 $18 - l$ 层的注入 \mathbf{z}_2 ，记最终得到的潜在编码为 \mathbf{z}_l^+ ，则最后在 FFHQ-StyleGAN 的生成器 g 中生成得到融合图 $g(\mathbf{z}_l^+)$ 。

该阶段总的损失函数如下：

$$\min_G \max_D \lambda_{\text{adv}} \mathcal{L}_{\text{adv}} + \lambda_{\text{perc}} \mathcal{L}_{\text{perc}} (G(\mathbf{z}_1, \tilde{\mathbf{z}}_2, \mathbf{1}), g(\mathbf{z}_l^+)) \quad (4-2)$$

其中， $\tilde{\mathbf{z}}_2$ 在 $\{\mathbf{z}_2, E(g(\mathbf{z}_2))\}$ 中采样得到。

第四阶段：源域到风格域的迁移训练：该阶段除了训练生成图像具有相同发布的原始 GAN 模型的要求外，还有生成图像和风格图的感知损失，最主要的是包括了分别对源域的人脸相似性和风格域的风格相似性进行的训练，损失函数见式3-2。

为了展示最终成效，我们以 FFHQ 中的其中一张人脸图像为例，分别输入 4 种不同风格但又同属于油画风格的风格图进行二者的融合与迁移。如图4-3所示，其中第一列为源域的内容图，第二列为对内容图的人脸编码后通过预训练的 FFHQ StyleGAN 得到的人脸重建图，第三列为风格域的风格图，第四列为将内容图以风

格图的风格进行融合迁移后得到的生成图，权重 w 前 7 层为 0.75，后 11 层为 1。

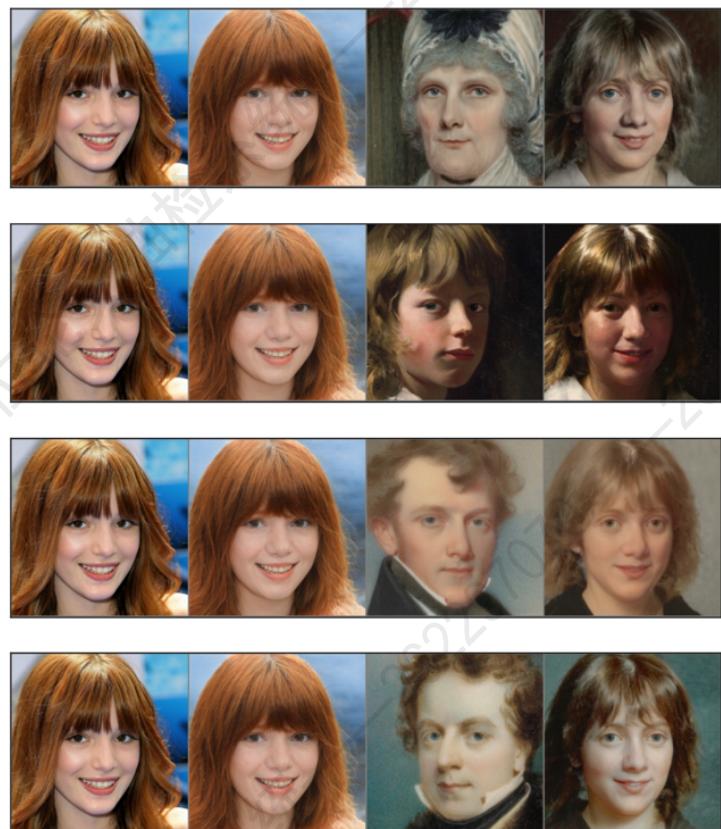


图 4-3 人脸油画风格生成效果示例

足以可见，在人脸重建上，FFHQ-StyleGAN 预训练模型及其 pSp Encoder 的 \mathcal{Z}^+ 空间优化得到了很好的效果，完整抓住了源域内容图的人脸结构特征和色彩特征。从最终的生成图来看，改进后的 DualStyleGAN 很好地把握了风格图的彩色风格和人脸特征，有效地将内容图人脸的姿态和特征保留并与风格图人脸特征进行了融合。

5 应用程序的设计与实现

5.1 引言

为了更好地向用户展示模型效果，让用户得以自主进行对目标风格图的选择和内容图的上传并最终浏览到生成图，我们需要结合软件工程相关知识，考虑绘画艺术的实际应用场景，设计并实现完成人像个性化油画生成的应用程序开发。

具体设计与实现主要分为功能需求分析、系统的结构设计以及最终的实现方法与策略。

5.2 功能需求

根据人像个性化油画生成应用的实际场景，此应用系统的实际功能可划分为以下几个部分。

1. 自主上传照片。系统应该提供用户能够在系统内自主选择和上传自己所需要进行风格变换的图片的功能。并且系统应该支持常见的图片格式，如 JPG、PNG 等。
2. 选择油画风格。系统应该提供用户能够从预设的数十种油画风格图中选择自己喜欢的风格，从而为后续的生成工作提供风格图。
3. 风格转换调整。系统应该提供用户能够从编码器选择上、人脸结构上、色彩风格上对生成图进行有效的控制和自定义的调整。
4. 个性化图像生成。系统应该结合内容图以及用户调整好的权重，将用户上传的图片和用户所选择的油画风格的图像进行风格融合与迁移。
5. 显示与下载结果。系统应该能够将处理得到的生成图片呈现给用户，并且提供用户下载该图片的功能。

5.3 系统设计

5.3.1 总体架构

人脸个性化油画风格生成应用系统的总体架构可用如图5-1所示的系统环境图（System Context Diagram, SCD）展现。整个系统环境可由用户界面处理、输入处理、输出处理、维护和自检以及主系统处理模块组成。

为了展现系统的层次结构关系，我们通过绘制该应用的具体系统结构图呈现，见图5-2。其中，主模块提供用户内容图片输入、编码器和油画风格选择等功能，最终通过图像生成模块获取处理结果。内容图输入模块包括本地文件上传和备选图选择，油画风格选择模块包括结构风格、色彩风格和目标风格图像的选择。

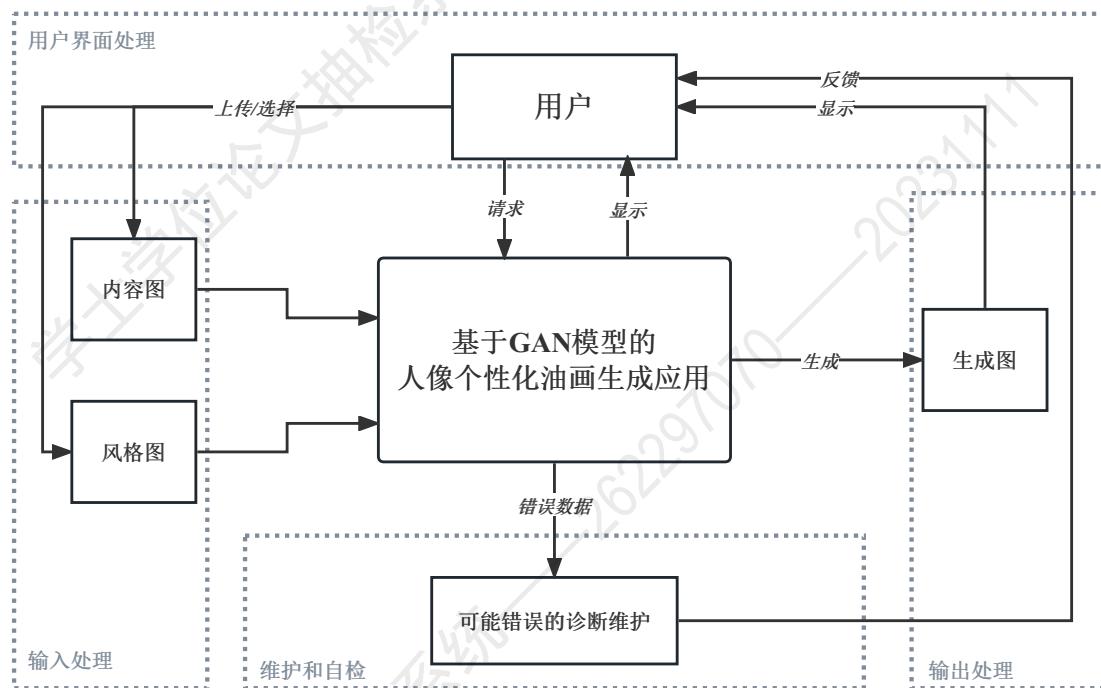


图 5-1 系统环境图

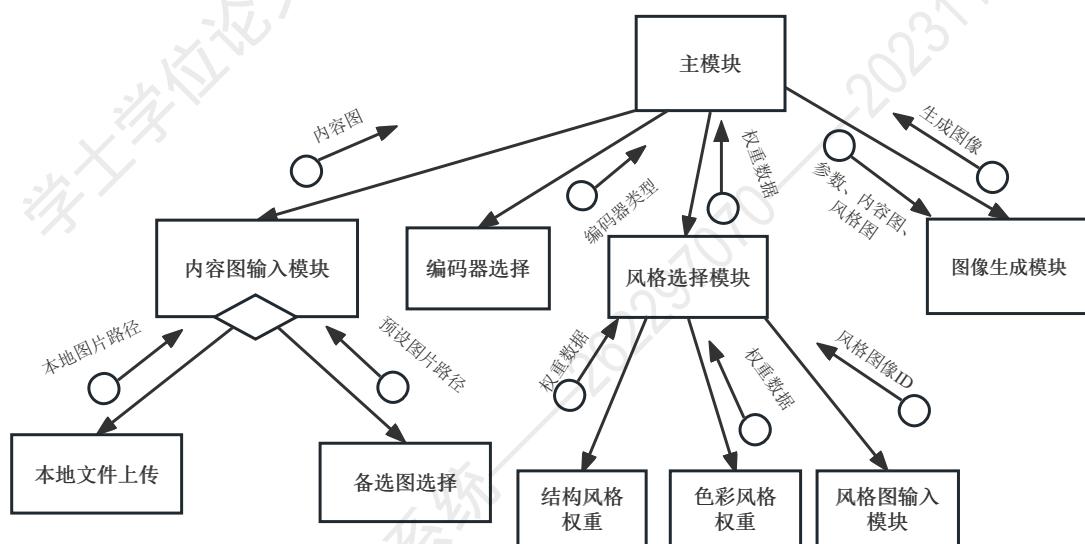


图 5-2 系统结构图

5.3.2 执行概念

为表示整个应用软件配置项之间的内在关系，我们考虑用活动图（Activity Diagram）来表示一项图像生成任务的动态处理执行流程，见图5-3。

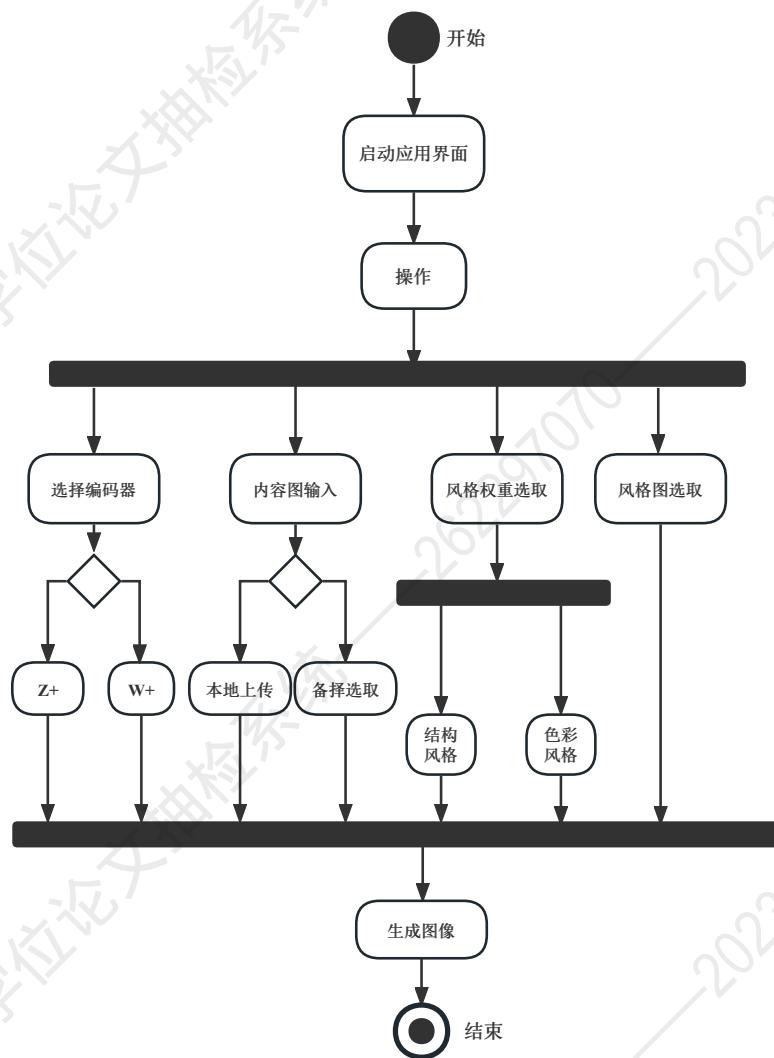


图 5-3 系统状态活动图

对于编码器的选取，我们提供两种备选训练得到的 pSp 编码器，包括 pSp Encoder 原始论文提到的 $W+$ 空间的训练以及 AgileGAN 技术的原始论文提到的 $Z+$ 空间的训练。在内容图的输入上，我们提供了数张来自 FFHQ 数据集的人脸图像作为备选参考，此外还允许用户上传自带的人脸图像进行输入。在风格权重的选取上，我们提供了结构风格权重和色彩风格权重两种可自定义的权重供用户调整出自己喜欢的生成图。在风格图的选择上，我们提供数十种已经通过 DualStyleGAN 训练成功的油画风格图像，让用户的选

5.4 应用实现

为了便于用户的快捷使用和模型的维护，我们考虑将采用 BS 架构进行具体的实现。为此我们需要细化系统框架，详细阐明各个功能的具体内容以及前端和后端设计。

5.4.1 细化内容图预处理

在具体实现中，我们提供 5 张来自 FFHQ 数据集的人脸图像给用户进行参考。其中有完全 1024×1024 大小的，也有其他尺寸但也为 $1:1$ 比例的人脸图，此外还有含有人脸但是比例并没有规格化的图片。用户也能自行上传想要进行风格变换的人脸图，规格不限。

我们的应用程序将通过以下三步对所有图片进行预处理，以便于后续的图像风格转换：

1. 通过图片所在路径读取图片以将图片转换为能被处理的数据结构用于后续处理；
2. 利用人脸 68 个特征点检测数据集训练得到的人脸识别模型对内容图进行人脸识别，并截取以人脸为中心的 $1:1$ 尺寸人脸图；
3. 将所得规格化后的人脸图像通过 FFHQ-StyleGAN 模型进行 pSp 编码后再重建，以获得可供后续风格转换所需的 \mathcal{W}^+ 或 \mathcal{Z}^+ 空间的潜在编码，此时图像已重建为了 1024×1024 大小。

5.4.2 细化风格转换调整

为了提供用户较自由的进行个性化风格生成操作，我们将作为样本的属于 Metfaces 数据集的 123 张油画风格图像在 DualStyleGAN 进行训练，这使得用户有高达 123 张油画风格图像的选择。此外，我们提供给用户结构风格权重和色彩风格权重两种可调参数，用户可分别对这两种权重在 $0 \sim 1$ 之间进行随心所欲的比例调整，为最后的生成图像提供更多的可自定义性。

5.4.3 细化前后端设计逻辑

人像个性化油画风格生成应用程序系统主要由前端、后端和 DualStyleGAN 模型三个部分组成。

前端：提供用户选择图片、上传照片、选择油画风格图、选择调控权重等功能，并向后端发送请求以获取处理结果。

后端：接收前端发来的请求，将用户上传或选择的照片和确定的油画风格和参数传递给 DualStyleGAN 模型进行处理，并将处理结果返回给前端。

模型：根据用户选择的油画风格和结构参数，将用户上传或选择的照片转换为油画风格的图像，并将结果传递给后端。

5.4.4 细化实现框架和布局

便于应用程序的开发和用户的快捷使用以及后续模型的维护，我们采用 BS 架构进行具体的实现。其中，我们选择采用发布于 2020 年 5 月的 Python 第三方库：Gradio 作为开发平台。

Gradio 是一款致力于帮助开发人员快速构建美观而且功能完善的 UI 界面、实现更好的用户体验的工具包。使用 Gradio 可以快速构建 Web 应用程序而几乎无需编写太多的 HTML、CSS 和 JavaScript 等前端的相关代码。它支持多种 UI 组件和输入类型，比如：文本框、滑块条、下拉菜单、文件上传等，能够轻松地与用户进行交互。

得益于 Gradio 较高的适配性和易用性，它可以接轨和集成诸如 Tensorflow、PyTorch、Scikit-Learn 等机器学习和深度学习相关的数据，包括文本、图像和音频等，因此它也是一个非常适合机器学习、深度学习爱好者和研究人员使用的工具。

为了呈现内容图的预处理过程，在页面布局上我们将展现备选 5 张人脸图、人脸裁剪和人脸重建的中间过程。如图5-4 所示。



图 5-4 内容图预处理布局

而在风格图的选取上，我们遵循让用户有“所见即所得”的便捷感，实现使得用户选择对应 ID 的风格图，下方就实时显示对应的风格图，如图5-5所示。



图 5-5 风格图的实时显示

其中，按钮的响应对应的函数相关的核心代码如下：

```
1 # 预处理按钮点击执行
2 preprocess_button.click(fn = run_alignment,
3                         inputs = [input_image],
4                         outputs = aligned_face)
5 # 选中样例图时
6 example_images.click(fn = set_example_image,
7                       inputs = example_images,
8                       outputs = example_images.components)
9
10 aligned_face.change(fn = reconstruct_face,
11                      inputs = [aligned_face, encoder_type],
12                      outputs = [reconstructed_face,instyle])
13
14 # 选择风格图像
15 style_index.release(fn = show_style_image,
16                      inputs = [style_index],
17                      outputs = [style_image])
18
19 # 生成风格图像按钮点击执行
```

```
20 generate_button.click(fn = image_generate,
21                         inputs = [
22                             encoder_type,
23                             style_index,
24                             structure_weight,
25                             color_weight,
26                             structure_only,
27                             instyle,
28                         ],
29                         outputs = result)
```

最终，得到整个应用程序的总体页面布局和实际画面（见图5-6）。

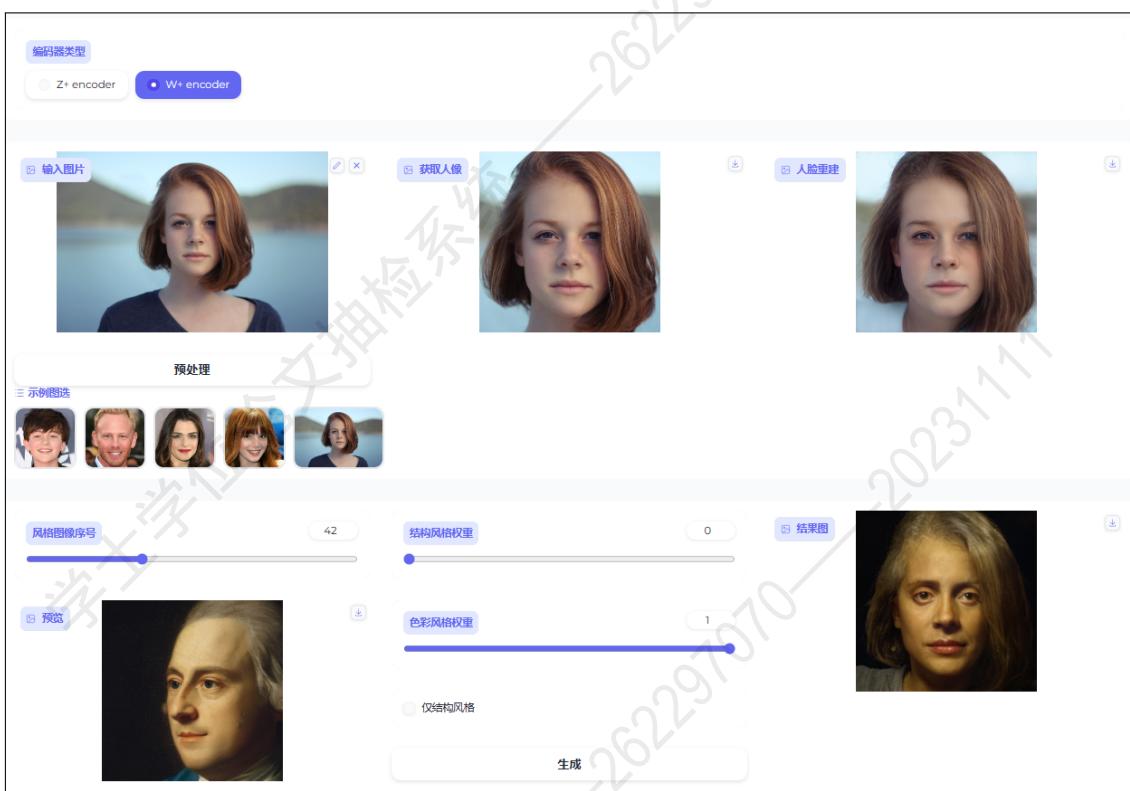


图 5-6 应用程序总体布局

6 总结与展望

6.1 全文总结

本文在阅读了大量国内外图像生成与图像风格迁移相关领域文献的情况下，分析了生成对抗网络和图像到图像转换技术的发展历程，详细分析了从经典的原始生成对抗网络的基本原理到图像的生成再到图像到图像转换的技术实现，以及如何更有效的控制生成风格的结构和样式进行的工作。针对油画风格对人脸辨识度和油画色彩风格的主要问题进行研究和改进，本文的工作总结如下：

1. 为解决图到图转换问题，分析了原始 GAN 模型和 StyleGAN 模型的网络结构及其实现图像风格控制和高分辨率生成图像的原理；分析了基于 StyleGAN 的图像编码技术 Pixel2Style2Pixel 以及 2022 年发表于 CVPR 的 DualStyleGAN。
2. 采用能凸显原图显著特征的 \mathcal{W}^+ 潜在空间编码更换了 DualStyleGAN 模型中的 \mathcal{Z}^+ 空间编码，从而增强了内容图的人脸特征在生成图中的地位。
3. 在模型训练过程中，通过预训练的 FFHQ-StyleGAN 模型进行分阶段微调，有效减少了一次性训练模型的时间成本，并且针对油画风格任务对损失函数进行调整。
4. 设计并实现了针对用户使用的网页端人像个性化油画风格生成的应用程序，为用户提供了编码器选择、图片选择与上传、风格图选择和结构与色彩风格权重等高自由度的自定义设置，丰富了用户体验。

6.2 展望

本文针对油画风格的改进措施可能存在自身知识面相对较窄以及主观或客观的其他因素而导致研究效果或许没达到最佳。

另外，现有的技术很难实现利用一个绝对通用的人工智能模型处理所有风格，泛化能力较低。不仅如此，人脸图像风格的融合与迁移这个课题的主观性较强，很难设计一个足够客观的量化指标来衡量最终生成图像的好坏，因此设计并提出更加规范的客观指标进行量化将是一项富含意义的工作。最后，利用生成对抗网络技术来实现人脸图像风格迁移技术也未必是最优解。2022 年底关注度很高的扩散模型（Diffusion Model）作为一种新的生成式模型，或许能给这个领域带来更多发展空间。

致谢

光阴似箭，一转眼大学四年的时光便悄悄来到了尾声。回顾我这平凡的四年本科生涯，学到了想学的知识、认识了一群志同道合的好友、参加了各式各样的科技竞赛、体验了前所未有的研究生备考时期和复试流程。虽然也曾有不少的遗憾和不舍，但是我仍然需要站出来，面对未知的未来和挑战。在本篇论文完成之际，我也要向所有帮助过我的人表示由衷的感谢。

首先，我要感谢我的导师周后盘老师。在整个研究期间，他给予了我无微不至的指导和关怀，提供了许多宝贵的建议和意见，帮助我克服了各种困难和挑战。他的悉心指导和鼓励对我而言是莫大的帮助。

此外，我还要感谢各位学长和身边的同学对我进行的实验操作和数据处理的技术支持，他们的帮助为我提供了必要的实验数据和方法。

最后，我也要感谢我的家人、朋友以及同学们的支持和鼓励。他们的关心和支持使我在完成论文的过程中更加坚定和自信。

在论文的完成过程中，有许多人对我给予了帮助和支持，这份感激之情将会铭刻在我的心中。因为你们的帮助，我才能完成这篇论文。再次向所有帮助过我的人表示诚挚的谢意。致以最高的敬意！

参考文献

- [1] Wang M, Wang B, Fei Y, et al. Towards photo watercolorization with artistic verisimilitude[J]. IEEE transactions on visualization and computer graphics, 2014, 20(10): 1451–1460.
- [2] Goodfellow I J, Pouget-abadie J, Mirza M, et al. Generative Adversarial Nets[J]. stat, 2014, 1050 : 10.
- [3] Hinton G E, Zemel R. Autoencoders, minimum description length and Helmholtz free energy[J]. Advances in neural information processing systems, 1993, 6.
- [4] Mirza M, Osindero S. Conditional Generative Adversarial Nets[J]. Computer Science, 2014 : 2672–2680.
- [5] Denton E L, Chintala S, Fergus R, et al. Deep generative image models using a laplacian pyramid of adversarial networks[J]. Advances in neural information processing systems, 2015, 28.
- [6] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.
- [7] Zhang H, Goodfellow I, Metaxas D, et al. Self-attention generative adversarial networks[C] // International conference on machine learning. 2019 : 7354–7363.
- [8] Mao X, Li Q, Xie H, et al. Least squares generative adversarial networks[C] // Proceedings of the IEEE international conference on computer vision. 2017 : 2794–2802.
- [9] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks[C] // International conference on machine learning. 2017 : 214–223.
- [10] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein gans[J]. Advances in neural information processing systems, 2017, 30.
- [11] Gatys L A, Ecker A S, Bethge M. A neural algorithm of artistic style[J]. arXiv preprint arXiv:1508.06576, 2015.

- [12] Isola P, Zhu J-Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017 : 1125 – 1134.
- [13] Zhu J-Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[J], 2017 : 2223 – 2232.
- [14] Kim J, Kim M, Kang H, et al. U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation[J]. arXiv preprint arXiv:1907.10830, 2019.
- [15] 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. 2019 : 4401 – 4410.
- [16] Chong M J, Forsyth D. GANs N’Roses: Stable, Controllable, Diverse Image to Image Translation (works for videos too!)[J]. arXiv preprint arXiv:2106.06561, 2021.
- [17] Yang S, Jiang L, Liu Z, et al. Pastiche master: exemplar-based high-resolution portrait style transfer[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022 : 7693 – 7702.
- [18] Karras T, Aila T, Laine S, et al. Progressive Growing of GANs for Improved Quality, Stability, and Variation[J]. arXiv preprint arXiv:1710.10196, 2017.
- [19] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization[C] // Proceedings of the IEEE international conference on computer vision. 2017 : 1501 – 1510.
- [20] 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.
- [21] Richardson E, Alaluf Y, Patashnik O, et al. Encoding in style: a stylegan encoder for image-to-image translation[C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021 : 2287 – 2296.
- [22] Song G, Luo L, Liu J, et al. Agilegan: stylizing portraits by inversion-consistent transfer learning[J]. ACM Transactions on Graphics (TOG), 2021, 40(4) : 1 – 13.

- [23] Mechrez R, Talmi I, Zelnik-manor L. The contextual loss for image transformation with non-aligned data[C] // Proceedings of the European conference on computer vision (ECCV). 2018 : 768 – 783.
- [24] Karras T, Aittala M, Hellsten J, et al. Training generative adversarial networks with limited data[J]. Advances in neural information processing systems, 2020, 33 : 12104 – 12114.

附录

王祥旭，男，2000年12月生。目前就读于杭州电子科技大学自动化学院，当前研究方向包括机器学习、计算机视觉等。

本科期间的主要成果：

1. 全国大学生数学竞赛初赛（浙江赛区）省三等奖；
2. 高教社杯全国大学生数学建模竞赛 省一等奖；
3. 美国大学生数学建模竞赛与交叉学科建模竞赛 H 奖 (Honorable Mention)；