北京理工大學

本科生毕业设计（论文）

基于真实人物多种风格漫画生成研究方法

Research on generating multiple style comic pictures based on real character image

学　　院：计算机学院

专　　业：计算机科学与技术

班　　级：07112005

学生姓名：潘宣文

学　　号：1120202720

指导教师：王崇文

2022 年 5 月 26 日

原创性声明

本人郑重声明：所呈交的毕业设计（论文），是本人在指导老师的指导下独立进行研究所取得的成果。除文中已经注明引用的内容外，本文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。

特此申明。

本人签名：　　日 期：　　年　月　日

关于使用授权的声明

本人完全了解北京理工大学有关保管、使用毕业设计（论文）的规定，其中包括：①学校有权保管、并向有关部门送交本毕业设计（论文）的原件与复印件；②学校可以采用影印、缩印或其它复制手段复制并保存本毕业设计（论文）；③学校可允许本毕业设计（论文）被查阅或借阅；④学校可以学术交流为目的，复制赠送和交换本毕业设计（论文）；⑤学校可以公布本毕业设计（论文）的全部或部分内容。

本人签名：　　日 期：　　年　月　日

指导老师签名：　　日 期：　　年　月　日

基于真实人物多种风格漫画生成研究方法

摘　要

漫画图像作为一种艺术表现形式，漫画的创作过程与人工智能生成图片的过程十分类似，基于真实人物图像的漫画风格图片生成问题的研究目的是在保持输入人物图片的关键面部身份信息的基础上，提取参考风格图像的关键信息，将真实人物图片转化为与参考风格图像色彩，结构类似即具有相同风格的图像。目前，基于真实人物图像的漫画化图像生成基本都是基于对抗生成网络（GAN）进行的。当前的对真实人物图片的漫画化生成的问题主要集中于大量的参考风格图片数据集难以获得以及漫画风格化的结果难以控制，此外，目前的对漫画风格化研究方法中，基本是以人眼主观评价为主，缺乏对生成的漫画图像的量化评估方法。

针对以上问题，本文从解决同一风格的参考风格图片数据集难以获取以及风格化强度难以控制两方面提出了一种新的 GAN 方法-MultipleComicGAN。MultipleComicGAN 的工作原理为将一或多张漫画风格图片通过 GAN 反转的方法获得隐代码，输入 StyleGAN 的使用 FFHQ 真实人脸数据集训练出的预训练模型，参考风格图片配合其输出的真实人物图片形成配对训练集进行多次迭代微调产生新的漫画风格化生成模型，达到将真实人物图片输入，输出获得多种漫画风格图像的目的。

MultipleComicGAN 还提供了一种风格控制机制，可以通过其来控制漫画化图片的色彩以及风格化的强度。在本次研究中还为 MultipleComicGAN 编写了一个用户友好型的界面，支持用户根据自己的喜好对训练过程进行参数微调。轻量化的设计使得 MultipleComicGAN 可以实现在 1 分钟以内进行 4 张照片至多 5 种风格的漫画风格化图像生成。

此外，在结果评价中，本文使用了用户评价、定性评判、定量评估三方面进行评价，在用户评价中，使用了调查问卷的形式来评判人眼对不同风格化方法的倾向性。在定量化评估中，使用 SIFID 判断对真实人物图片的身份特征信息的保持程度；使用 SSIM 来判断生成的漫画化图像与参考风格图像是否属于同一种风格，之后通过将人眼的评估以及定量化的评估结合的形式，对生成的漫画化结果进行全方位的评价。

实验结果表明，无论是从定性还是从定量的评估指标判断，本文提出的方法具有能够快速生成高质量高分辨率的漫画风格化图像；支持以多张真实人物图片为基础同时生成多张多风格漫画化图片；风格化以及训练微调耗时较短，实现了较好的轻量化处理；不依靠艺术风格训练集，只需要一或数张目标风格图片便可以生成该风格的漫画风格化图片等优点，多方面优于传统的 GAN 漫画化方法。

关键词：漫画风格化；对抗生成网络；GAN 反转

Research on generating multiple style comic pictures based on real character image

Abstract

Cartoon images, as a form of artistic expression, have a creation process very similar to the process of generating images using artificial intelligence. The research objective of generating cartoon-style images based on real character images is to maintain key facial identity information from the input human images while extracting crucial information from reference style images, transforming real human images into images that match the color and structure—or style—of the reference style images. Currently, the generation of cartoonized images from real human images is primarily based on Generative Adversarial Networks (GAN). The main challenges with the current approach to cartoonizing real human images focus on the difficulty in obtaining extensive datasets of reference style images and controlling the results of the cartoonization. Additionally, current methods for researching cartoon style primarily rely on subjective human evaluation, lacking quantitative assessment methods for the generated cartoon images.

To address these issues, this paper proposes a new GAN method—MultipleComicGAN—that tackles the challenges of obtaining datasets of reference style images in the same style and controlling the intensity of the stylization. The working principle of MultipleComicGAN involves using one or several cartoon-style images to obtain latent codes through GAN inversion, inputting these into a StyleGAN model trained on the FFHQ real human face dataset, and pairing it with real human images to form a training set for multiple iterations of fine-tuning to produce a new cartoon-style generation model. This achieves the goal of inputting real human images and outputting images in various cartoon styles.

MultipleComicGAN also provides a style control mechanism, allowing control over the color and intensity of the cartoonization. Additionally, a user-friendly interface was developed for MultipleComicGAN, allowing users to fine-tune the training process parameters according to their preferences. The lightweight design of MultipleComicGAN enables it to

generate cartoon-style images in up to five styles from four photos in under one minute.

Furthermore, in evaluating the results, this paper uses user evaluation, qualitative judgment, and quantitative assessment. For user evaluation, a survey questionnaire is used to judge human preferences for different stylization methods. For quantitative assessment, SIFID is used to determine the degree of retention of identity features of the real human images, and SSIM is used to judge whether the cartoonized image and the reference style image belong to the same style. The results from the cartoonization are then comprehensively evaluated by combining human assessments and quantitative assessments.

Experimental results show that the proposed method can quickly generate high-quality, high-resolution cartoon-style images; supports the generation of multiple cartoon-style images based on several real human images simultaneously; has short style adjustment and training fine-tuning times, achieving effective lightweight processing; and does not rely on an art style training dataset, as it can generate cartoon-style images of a particular style with just one or a few target style images. This method shows multiple advantages over traditional GAN cartoonization methods in various aspects.

Key Words: comic stylizing; Generative adversarial network(GAN); GAN inversion

目 录

摘 要 I
Abstract III
第 1 章 绪论 1
1.1 研究背景和意义 1
1.2 国内外研究现状和发展趋势 2
1.2.1 GAN 网络算法的研究现状及发展趋势 2
1.2.2 真实人物图片漫画化的研究现状以及发展趋势 4
1.2.3 对漫画化图像的评估指标的研究现状 6
1.3 本文主要研究工作 7
1.4 本文结构安排 8
第 2 章 基于真实人物图像的多风格漫画化图片生成 GAN 方法 10
2.1 GAN 网络的基本原理 10
2.2 用于人像图片生成的 GAN 11
2.2.1 StyleGAN: 基于样式的生成对抗网络 11
2.2.2 用于人脸生成的少训练样本风格化方法 12
2.3 MultipleComicGAN: 多风格漫画化图像生成 GAN 网络结构设计 14
2.3.1 生成器与判别器的设计 14
2.3.2 GAN 反转 16
2.3.3 根据反转产生的隐代码对 StyleGAN 进行微调 17
2.3.4 损失函数的设计 17
2.3.5 生成漫画风格化后的图片 18
2.4 本章小结 18
第 3 章 基于 MultipleComicGAN 的漫画风格化生成 20
3.1 漫画风格化生成 20
3.1.1 基于真实人物图像的预训练模型 20
3.1.2 控制风格化的强度 20
3.1.3 保留色彩进行风格化 23
3.1.4 迭代轮次的选择 24
3.2 训练过程 27
3.2.1 自定义训练风格图片 27
3.2.2 自定义训练参数 27

3.2.3 生成结果 27
3.3 本章小结 28
第 4 章 实验结果与分析 30
4.1 参数配置 30
4.2 图像评估指标 31
4.3 MultipleComicGAN 生成的实验结果比较 32
4.3.1 定性评估比较 34
4.3.2 定量化评估比较 35
4.4 训练时间与硬件消耗比较 38
4.5 本章小结 38
结 论 40
参考文献 42
致 谢 45

第 1 章 绪论

1.1 研究背景和意义

在近几年，随着深度学习和人工智能的发展，计算机视觉领域取得了巨大的进步。其中，图像生成技术更是其中的热点领域，通过图像生成技术，可以通过已有图像生成新的图像，也可以根据特定的输入和描述生成图像或者是修改现有的图像。而漫画作为一种艺术形式，它的风格多样，且具有强烈的个性化特征，作为一种创意表现形式，通常幽默或讽刺地夸张或简化人或物的特征。漫画作家在创作的过程中通常会去捕捉人脸的最显著特征，例如脸型，鼻子形状，嘴型，眼睛形状，发型等，然后通过放大这些独特的面部特征，加以作家自己的风格，色彩等来创作属于自己的漫画图片[1]，漫画作家在创作漫画的过程与人工智能生成图片的过程十分类似。

在此背景下，基于真实人物的漫画风格图片生成可以应用于多种场景，例如，在艺术与电子游戏创作中，如果能够通过真实的人物图像与参考风格图像直接获得该特定艺术风格下的漫画化后的人物图像，可以帮助艺术家或者是人物设计师将复杂繁琐的工作简单化；在社交网站中，很多用户既想要将自己的头像设置为现实照片，又担心会发生个人隐私信息泄露引发的一系列问题，这时便可以使用漫画风格化后的真实自拍照作为头像，既保留了用户的基本外貌特征，又有效保护了用户的个人隐私，同时给用户提供了一种更轻松愉快的社交氛围；在增强现实（AR）的场景构建中，很多时候会出现用户看到现实中的人脸而产生沉浸感不够的问题，而如果此时将这些现实中的人物使用场景的参考风格漫画风格化，可以极大的增加用户在使用时的沉浸感，增强现实的技术正在应用于教育，娱乐，宣传等多种多样的场景，将漫画化图像生成的技术应用于其中可以给用户带来更加丰富、真实的体验。如何将真实人物图像转换为具有特定风格的漫画，既保留人物的特征，又展现出特定的漫画风格是一项有挑战性且有意义的研究。例如将自己的自拍照变换为动漫游戏的漫画风格，甚至于变换为莫奈的水彩画，日本的浮世绘风格等艺术的广义上的漫画风格。在实际研究中，生成模型需要理解面部特征并且根据这些特征生成漫画风格图像。

近年来随着深度学习技术以及硬件规格的快速发展，经典的图像风格迁移方法[2-3]可以初步满足真实人物照片的漫画化图像生成。但是由于风格迁移会获取所

有风格信息，包括整个风格图像的颜色和笔触，并将其转移到整个内容图像。很多不应该被迁移的部分也被迁移到了人脸照片上，一般的图像风格迁移方法并不能完全适用于真实人物图片的漫画化风格图像生成过程。与使用 CNN 相比，GAN 是一种更成功的解决方案。为了给用户同时提供多种漫画风格化图片生成的选择，使用现有的 GAN 方法，每次对风格的训练都是从零开始训练训练一个生成模型，训练出的模型都过于特定化于某一种漫画风格[4]，以至没有办法在较短时间内获得除训练的风格之外的任何风格的图像。每增加一个新的风格选择都会大量增加对于时间和算力的消耗，大部分设备无法承担如此量级的一个模型，因此无法在各种硬件平台上都能高效运行，不具备普适性。另一方面，在现有方法中，训练形成对应的风格化模型的过程依赖于大量的风格参考图片，而通常情况下，参考风格图片难以大量获得，一位漫画家的一种风格漫画也许只有几十张甚至几张，大量的参考风格图片较为难以获取。对大量参考图片的训练过程也需要大量的内存，算力，以及存储空间。

本文提出了一种新的用于漫画生成的对抗生成网络（GAN）方法-MultipleComic-GAN。本文提出的方法旨在通过对 StyleGAN 预训练模型进行微调的方法，解决现有的基于真实人物多种风格漫画生成存在的问题，开发新的多种风格漫画生成模型，同时力求做到更好地捕捉人物的面部细节，在最大程度保留人物面部特点的前提下进行真实人物照片漫画风格图片生成，并且将人类的喜好与传统评估方法结合，评估生成的漫画化图像的质量。

1.2 国内外研究现状和发展趋势

1.2.1 GAN 网络算法的研究现状及发展趋势

在 Goodfellow[5] 等人于 2014 年提出 GAN 网络之后，国内外展开了大量的研究在 GAN 网络的改进以及应用方面，GAN 的结构如1-1所示，其灵感来源于博弈论中的二人零和博弈问题，通过训练一个生成器和一个判别器，生成器尝试生成越来越逼真图像，判别器则尝试区分真实的图像和生成器生成的图像。二者不断在对抗中进步完善，最终达到与真实的图像一致的目的。GAN 网络在计算机视觉领域应用十分广泛，例如可以应用于人脸图像生成（比如生成一张从未存在的人的照片）；图像转换（比如把马变化成斑马）；文字-图片转化；人脸属性编辑等领域。

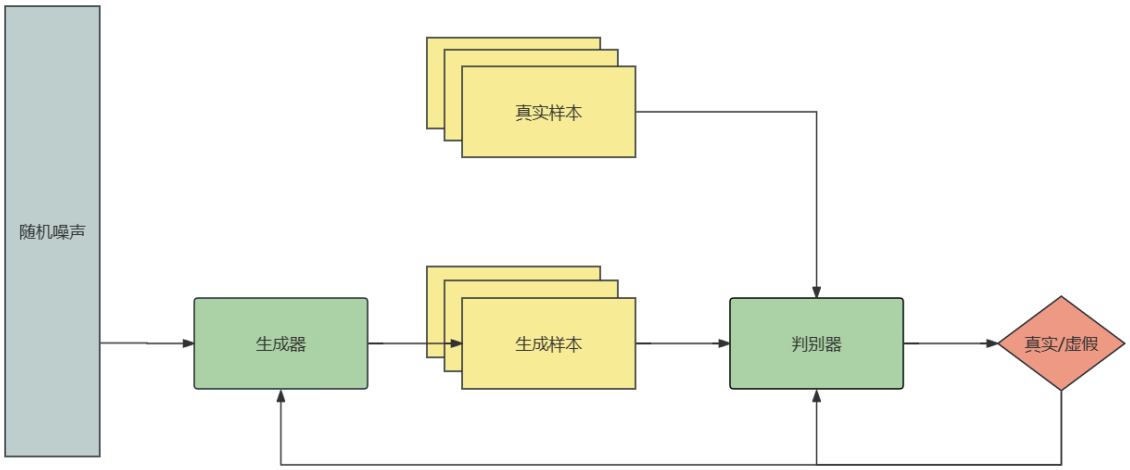


图 1-1 GAN 原型结构图

Radford[6] 等人于 2015 年提出的 DCGAN 是第一次在 GAN 中使用卷积神经网络，并取得了非常好的结果。之前，CNN 在计算机视觉方面取得了前所未有的成果。但在 GAN 中还没有开始应用 CNNs。CNN 中的反卷积层具有空间上采样功能，使 GAN 网络能够生成更高分辨率的图像。cGAN[7-8](conditional GAN) 作为有监督模型，在生成器与判别器的输入层上添加了额外的 one-hot 向量 c 作为条件信息约束网络的迭代优化方向，实现了指向性生成数据。对于图像到图像的翻译任务，Isola[9] 等人提出的 pix2pix 也显示出了令人印象深刻的结果，pix2pix 学习从输入图像到输出图像的映射, 从而实现将草图转化为照片的效果，pix2pix 的关键贡献点在于提出了用 GAN 来解决图像转换问题的通用方法，并且证明了其方法的有效性。2017 年，比起 pix2pix 更具优势训练不需要成对的数据集的 CycleGAN 被 Zhu 等人提出，CycleGAN 引入环形生成对抗结构，用于不同的图像到图像翻译。在图像质量与图像分辨率的层面上，NVIDIA 在 2017 年提出的 ProGAN[10]解决了生成高分辨率图像 (如 1024 × 1024) 的问题。ProGAN 提出了渐进式训练的方法，即从训练分辨率非常低的图像 (如 4 × 4) 的生成器和判别器开始，每次都增加一个更高的分辨率层，如图1-2所示。然而，ProGAN 仍然存在问题，与多数 GAN 一样，ProGAN 控制生成图像的特定特征的能力非常有限[11]。这些属性相互纠缠，即使略微调整输入，会同时影响生成图像的多个属性。StyleGAN (A Style-based Generator Architecture for GAN)[12-14]从 ProGAN 中演变而来，具有可基于样式的生成器，可生成更高质量的高分辨率图像。StyleGAN 在 GAN 模型基础上删除了传统输入，添加了噪声 noise，使用了自适应实例归一化

(AdaIN)[15]。StyleGAN 极大幅度地提高了研究者们对 GAN 合成的理解和可控性。

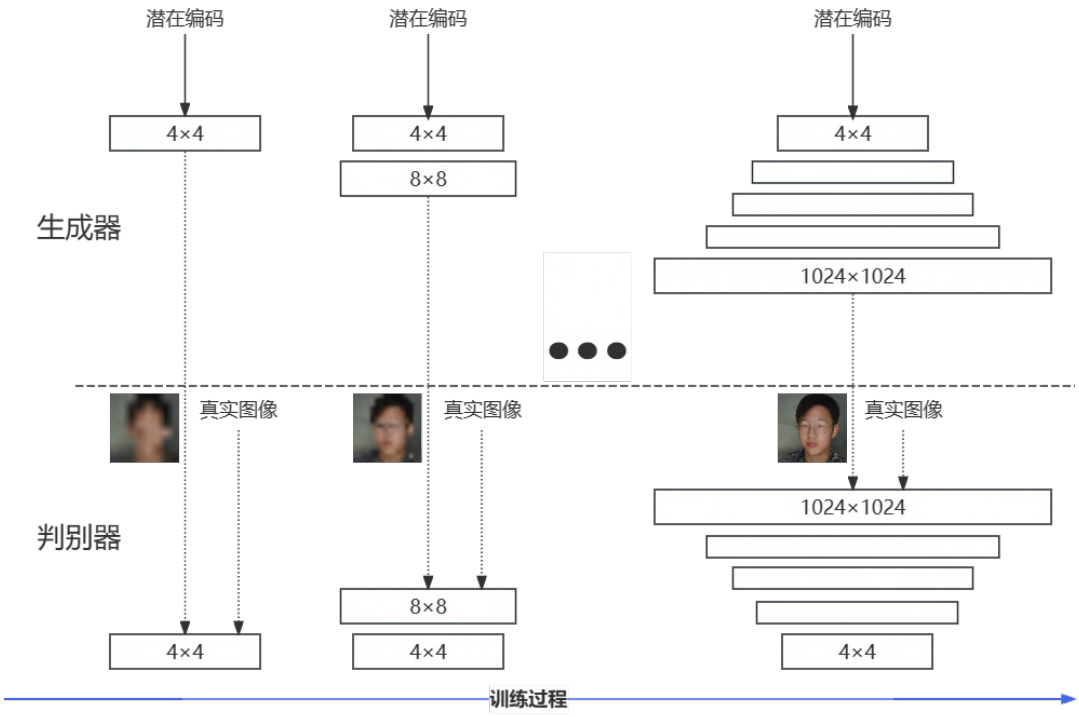


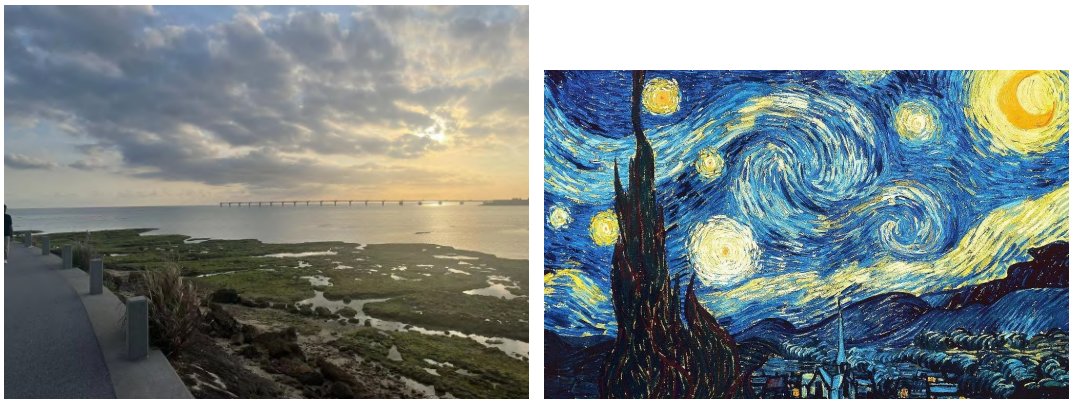
图 1-2 proGAN 结构图

总而言之，对 GAN 网络的改进主要集中在如何获得更高的图像质量，使图像生成与训练更可控以及设置损失函数以避免模式崩溃等问题[16-17]；无监督学习这几个方面。随着 GAN 网络研究的深入，GAN 的应用场景将被广泛运用于 XR，视频生成，医学图像处理等领域。

1.2.2 真实人物图片漫画化的研究现状以及发展趋势

最早的真实人物图片的漫画化可以追溯到深度学习方法出现之前的风格迁移-非真实感渲染方法，Non-Photorealistic Rendering (NPR)[18]以及图像滤波进行艺术风格模拟等。在深度学习技术得到广泛发展应用后，对风格迁移研究表明采用 CNN 的方法[19]可以做到抽取原始图像的内容信息和目标图像的风格信息以完成图像合成任务的作用，如图1-3所示。以 Gatys 等人[2]提出的神经风格迁移（Neural Style Transfer, NST）的方法为开端，对使用 CNN 进行风格迁移主要集中于提升生成风格的质量以

及缩短风格迁移所需要的时间两个方面为主。然而，使用 CNN 进行风格迁移的方法需要大量包含参考风格图像和真实图像的数据来训练模型[3]。在大多数情况下，获取特定艺术风格的多个样本极其困难，一位艺术家通常很难画出同样风格的大量作品，符合研究者期望的是能够使用尽可能少的参考样式示例的方法[20]。



(a) 需要风格迁移的风景图　　(b) 梵高星空风格图片

(c) 风格迁移后的结果

图 1-3 风格迁移的示意图

当 GAN 被引入进行类似风格迁移的工作后，对此方面研究者们进行了大量的研究。2018 年 Chen 等人[21]提出了真实世界的图转化为动漫风格的图的 CartoonGAN，该方法提出了一个专用的基于 GAN 的方法，可以有效地学习使用不成对的图片集进行训练，对现实世界照片和漫画图像建立映射，以此来生成高质量动漫风格图片。在此基础上 Wang 等人又提出了 AnimeGAN，让生成图像拥有动漫风格的纹理和线条的同时，让生成图像维持原图的颜色内容并设计了一个轻量化的 generator。2022 年 Shu 等人提出了 MSCartoonGAN[22]，基本实现了多种漫画风格进行生成的目的，然

而使用其应用到人脸上时生成的漫画图片会出现边缘模糊的问题，使得人脸与背景模糊，不符合漫画的基本特征，而且多种风格存在高度相似性。

另一方面，基于 StyleGAN 的基础上，Jang 等人在 2021 年通过对风格进行操控提出了 StyleCariGAN[23]的用于漫画图像生成 GAN 模型-StyleCariGAN，该方法能够自动地从输入照片生成真实且详细的漫画，而且可以部分选择性地控制对人脸的夸张程度和漫画风格类型，能够生成真实且详细的漫画，此外还支持其他基于 StyleGAN 的对图像的操作，例如面部表情控制。使用预训练的 GAN 作为先验可以消除图像风格化任务对大型训练数据集的需求。在 2021 年的 Mind The Gap (MTG)[24]，和 OneshotCLIP (OSC)[25] 的方法均表明，通过微调预训练的 StyleGAN2，是可以达到较为理想的真实人物漫画风格化图片生成的效果的。

总而言之，漫画化生成的方法正在不断的进展之中，目前来讲使用 GAN 的方法是更为完美的解决方案，未来的发展方向包括多模态学习、可控生成、视觉效果增强和应用领域拓展等方向，漫画化技术的发展给漫画化图片生成在各式领域的应用提供了可能性。

1.2.3 对漫画化图像的评估指标的研究现状

在对于图像的质量量化评估上的研究上，图像生成领域过去的研究者通常使用人工评估而非量化评估的方式进行[26]，在最近几年评估生成图像的质量是图像生成领域的热点问题之一，尤其是在无真实图像作为参考的无监督学习的设置中。目前，最常用的两种评估指标是 Frechet Inception Distance (FID)[27]和 InceptionScore (IS)[28]。Inception Score 是 Salimans 等人在 2016 年提出的，用于评估无监督图像生成模型的性能。IS 的基本思想是，如果生成的图像有高质量，那么它们应该与真实图像在特征空间中具有相似的分布。IS 使用预训练的 Inception 模型将生成的图像映射到特征空间，然后计算这些特征的分布与真实图像特征分布之间的 Kullback-Leibler 散度。IS 越高，生成的图像质量就越好。然而，IS 有一些已知的缺点，例如它对模式崩溃的问题判断效果不好，当生成器重复生成相同的图像时，IS 可能依然会给出很高的评价，并且可能会过度奖励多样性而不是考虑生成的图像的真实性。

基于 IS 存在的这些问题，Heusel 等人在 2017 年提出了 Frechet Inception Distance (FID)[27]。FID 通过计算生成的图像和真实图像在特征空间中的分布之间的弗雷歇特距离代替 Kullback-Leibler 散度，使得 FID 可以更好地判断生成和真实图像分布之

间的差异。FID 的具体判断方法是：FID 越低，生成的图像质量就越好。通常使用 FID 进行评估由于对真实性的评价效果更好且更加可靠性更好所以要优于使用 IS 进行量化评估。在此之后，不断有新的图像质量评判量化评估指标被提出，2018 年，Zhang 等人提出了 Learned Perceptual Image Patch Similarity (LPIPS)[29]，LPIPS 利用深度学习模型来评估图像之间的感知相似度，可以更好地反映人类对图像质量的主观评价。2019 年，在对 StyleGAN 的研究之中，Tero Karras 等人同 styleGAN[12]一起提出了 PPL，PPL 用于衡量生成图像的质量，特别是在 StyleGAN 这类生成模型中。它通过测量潜在空间中的路径来评估图像变化的平滑程度。2020 年，Gu 等人提出了 Generated ImageQuality Assessment (GIQA)[30]，GIQA 专注于对每张生成图像的质量进行单独量化评估。这使得 GIQA 能够提供更详细和个性化的图像质量分析，从而更精确地衡量和改进生成模型的性能。

1.3 本文主要研究工作

本文研究的内容是完成基于真实人物图像的多风格漫画化图像生成，因为 GAN 网络结构在解决此类问题的优越性，本文选择使用 GAN 的方法进行解决现有多种漫画风格图像生成中存在的问题针对当前漫画风格图像生成模型存在的以下问题：对多种风格生成时的适配性不够好，经常产生不同风格的照片相似度过高的问题；对人脸的细节层面以及整体结构保持的不够，会出现与风格参考图片过于相似的情况；对风格进行训练通常要消耗大量的算力，硬盘容量以及时间，获得大量同种风格的参考图片难以实现；用户可操作互动性较差，通常只适用于科研人员科研使用，对于不具备电脑基本知识的用户来说几乎无法实践生成，进行研究并提出解决方案。当前对漫画化生成结果缺乏一种行之有效的评价方法。

本文主要研究工作如下：

1. 本文采用了 Mind The Gap (MTG)[24]，和 OneshotCLIP (OSC)[25] 的思想，提出了一种新的通过对 StyleGAN2 进行微调来实现多风格漫画化生成的目的的 GAN 网络结构-MultipleComicGAN。为了解决参考风格图像难以获取的问题，通过提前对 StyleGAN2 使用 FFHQ 数据集进行预训练，之后使用对单张风格参考照片进行 GAN 反转的形式获得成对的训练数据集，后续再通过成对的训练数据集结合 StyleGAN 的样式混合特性对产生的预训练模型进行微调的方法，避免了凭空训练风格模型对算力以及时间大量消耗的问题，同时避免了从零训

练的单个生成器过于局限一种风格的问题。同时使用了 jupyter notebook 配合 markdown 文本的方式，为本文提到的 MultipleComicGAN 网络提供了一个易于交互的界面，并且将环境配置写入。即使是毫无编程经验的用户，也可以轻易输入自己的照片进行漫画风格化生成，与此同时为用户提供了可以自行调整训练参数的交互界面。

2. 针对当前对基于真实人物图片的漫画化生成结果缺乏有效的评估指标的问题，本文提出了使用定性评判，定量评估，用户评价三者结合的形式。首先使用了定性评判对整体的色彩风格等进行了一个大体的评价，之后使用了 SSIM 以及 SIFID 分别对风格以及人物身份特征信息保持程度进行进一步的量化评估，最后制作了一个调查问卷，将定性以及定量化的评估结果与用户人眼的判断相结合进行分析。以三者结合的形式，作为一种既符合主观化人类视觉偏好的，又使用了客观化的数值评估的主客观相结合的评价方法。在对方法的评价中，也综合考虑了性能指标，如训练时间，模型大小等。

1.4 本文结构安排

本文总共分为四章，以及最后的结论部分，整体内容的文章结构组织如下：

第一章为绪论部分，本章首先介绍了基于真实人物图像漫画化生成目前的研究背景及其研究意义，之后回顾了当前的 GAN 网络，漫画风格化以及漫画化图像的评估的发展现状，取得成果以及优化方向，同时也总结了当前的 GAN 网络在解决真实人物图像多风格漫画化图像生成时存在的问题。最后介绍了本文的研究内容要解决的问题以及本文的结构安排。

第二章详细阐述了“MultipleComicGAN”这一基于真实人物图像的多风格漫画化图片生成的 GAN 方法。此章节首先从 GAN 网络的基本原理开始介绍，然后深入探讨了用于人像图片生成的 GAN 技术，如 StyleGAN 等，随后详细说明了 MultipleComicGAN 的网络结构设计，包括生成器与判别器的设计、GAN 反转技术及其在风格化图像生成中的应用，以及损失函数的设计。本章还介绍了如何通过这些技术生成漫画风格化的图片，并对本章内容进行了总结。

第三章探讨了基于 MultipleComicGAN 的漫画风格化生成以及风格控制机制。首先详述了漫画风格化生成的过程，包括预训练模型的使用、GAN 反转方法的选择、控制风格化的强度，以及如何在风格化过程中保留原始色彩。接下来，章节转向介

绍生成训练的过程，包括自定义训练风格图片以及进行训练参数的调整。章节最后进行了小结。

第四章则呈现了实验结果与分析。该章节首先介绍了实验的软硬件配置，然后详细解释了用于评估图像的各种指标。随后，描述了实验的具体流程，展示了 MultipleComicGAN 生成的图像并将其与两种机理类似方法进行对比及其定性和定量以及用户的评估结果。还对训练时间和模型的轻量化进行了分析。章节末尾进行了总结。

在结论部分中，对本文的工作内容进行了总结，也指出了当前依然存在的问题，并且展望了未来的工作方向。

第 2 章 基于真实人物图像的多风格漫画化图片生成 GAN 方法

2.1 GAN 网络的基本原理

生成对抗网络（GAN, Generative Adversarial Networks）是一种深度学习模型，主要用于无监督学习，由 Ian Goodfellow 及其同事在 2014 年提出[5]。GAN 由两个核心组件组成：生成器（Generator）和判别器（Discriminator），它们在训练过程中互相对抗，从而达到提高生成图像质量的目的。其结构如1-1所示

生成器的目标是创造出尽可能真实的数据（例如图像、音频等），使其无法被判别器区分出来是假的。生成器接收一个随机噪声信号作为输入，通过这个输入信号构建出数据。这个过程类似于从一个简单的随机分布中学习如何映射到数据的复杂分布。判别器的作用是区分输入的数据是来自真实数据集还是生成器产生的。换句话说，判别器的任务是识别数据的真伪。在训练初期，判别器通常能够较容易地识别出生成器的输出，因为生成的数据质量不高。

训练 GAN 的过程可以看作是一个博弈过程，其中生成器尝试“欺骗”判别器，而判别器则尝试抵抗被欺骗。通过这种方式，生成器和判别器在训练过程中不断改进各自的性能：如果判别器识别出了生成的数据，生成器会调整其参数，尝试生成更加逼真的数据；判别器会通过区分真实数据和生成数据来优化其分类准确性。这一过程可以通过公式2-1表达

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \tag{2-1}$$

其中 $\mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)]$ 表示判别器对来自真实数据集的样本 x 正确识别为真实的平均对数概率。判别器 D 的目标是最大化这个概率，使 $D(x)$ 尽可能接近 1。$\mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))]$ 表示判别器对生成器生成的假样本 $G(z)$ 正确识别为生成的平均对数概率。判别器 D 的目标是最大化这个概率，使 $D(G(z))$ 尽可能接近 0。生成器 G 的目标是最小化 $\log(1 - D(G(z)))$，这实际上等价于最大化 $\log D(G(z))$，意味着生成器努力使其生成的假样本 $G(z)$ 被判别器误认为是真实样本，即尽量让 $D(G(z))$ 接近 1。经过足够的训练后，在理想状态下，GAN 训练会达到一个均衡点，此时判别器无法区分真假样本，即 $D(x) \approx 0.5$ 对所有的 x。这意味着生成器 G 已经非常成功地模仿了真实数据的分布。生成器能生成非常接近真实数据的样本，而判别器对真假数据的识别概率接近 50%，即它无法区分真假数据，这表示生成器与判别器达到了某种“平

衡”。通过这种方式，GAN 能学习到复杂的数据分布，生成高质量的、多样化的数据样本，广泛应用于图像生成、语音合成等领域。

2.2 用于人像图片生成的 GAN

用于人像图片生成的 GAN 已经在1.2.2中介绍过，本文选取的是使用一张参考风格图片对 StyleGAN 进行迭代微调的方法，所以详细介绍 StyleGAN 以及用于人脸生成的一次性域自适应的关键技术原理。

2.2.1 StyleGAN: 基于样式的生成对抗网络

StyleGAN，全称为样式生成对抗网络（Style-Based Generator Architecture for Generative Adversarial Networks），是由 NVIDIA 的研究团队在 2018 年提出的一种改进的 GAN 架构[12]。StyleGAN 特别适用于生成高质量、高分辨率的图像，尤其在人脸图像生成方面表现出色。其核心创新在于引入了一个新的生成器架构，该架构在生成过程中更加关注图像的风格（样式）信息，其架构如图2-1所示。

StyleGAN 的核心理念是将“风格”（或样式）的概念引入到生成过程中。在 StyleGAN 中，“风格”可以控制图像的高层属性（如脸型、发型）以及更细节的特征（如纹理）。这是通过一个映射网络（mapping network）实现的，该网络将标准的高斯噪声向量转换为中间潜在空间的表示，映射网络负责将输入的高斯噪声向量 z 转换成潜在空间向量 w，这些表示随后用于直接控制生成器的不同层。这一过程可以用公式2-2表示。

$$w = f(z) \tag{2-2}$$

其中，f 代表由多个全连接层构成的非线性映射函数。

此外 StyleaGAN 还使用了自适应归一化（AdaIN）的技术，AdaIN 调制层使得其具有很强的解耦性以及可编辑性。自适应实例归一化是 StyleGAN 中用于将潜在向量 w 应用到生成器各层的技术。每一层的特征图 x_i 通过 AdaIN 调整其样式，如公式2-3所示。

$$\text{AdaIN}(x_i, w) = w_{\sigma,i}\frac{x_i - \mu(x_i)}{\sigma(x_i)} + w_{\mu,i} \tag{2-3}$$

其中，$\mu(x_i)$ 和 $\sigma(x_i)$ 分别是特征图 x_i 的均值和标准差。$w_{\sigma,i}$ 和 $w_{\mu,i}$ 是从潜在向量 w 派生的尺度和偏移参数。

混合正则化鼓励网络在生成过程中使用来自不同 w 的样式信息，这可以视为在

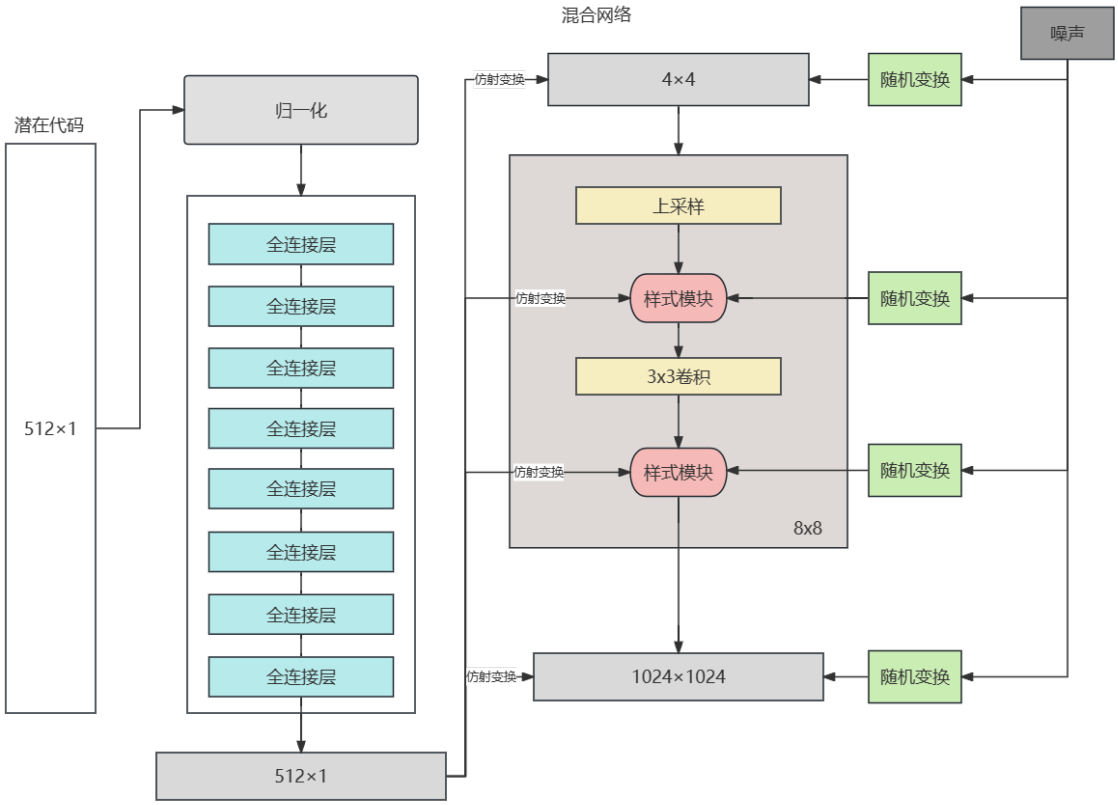


图 2-1 StyleGAN 结构图

训练过程中随机选择不同的 w 来生成每一层的样式如公式2-4所示。

$$\text{Layer Style} = \text{Rand}\left(w^{(1)}, w^{(2)}, \ldots, w^{(n)}\right) \tag{2-4}$$

Pinkney 等人首次展示了在新数据集上微调 StyleGAN 并执行层交换，使 StyleGAN 能够通过相对较小的数据集学习图像到图像的转换[31]。但即使是获取一个小的配对数据集也是困难的：收集工作既困难又昂贵；每种新风格都需要一个新的数据集；在很多情况下，同种风格图像只有数张风格参考照片，例如梵高的自画像风格根本不存在其他的。本文提到的方法同样使用了对 StyleGAN 微调的方法，从单一风格参考图片中以 GAN 反转的方法创建配对数据集，然后使用创建的数据集进行对 StyleGAN2 的微调。

2.2.2 用于人脸生成的少训练样本风格化方法

少训练样本学习现在已经在许多领域有所应用，例如图像检测分类，图像合成等。为了实现模型的轻量化处理，以及解决参考风格照片难以获得的问题，本文专注于使用少训练样本学习，主要针对于人物真实图片的少训练样本学习。通常来讲，

从极少的训练样本学习样式映射会导致过拟合问题。为了控制过拟合，文章[32-33]引入了正则化项，与此同时文章[34-35]则强制对网络权重施加约束。这些方法需要数十到数百个样式示例图像；相比之下，本文提到的方法只需一到数个。此外，这些方法由于对对抗损失的依赖，往往难以捕捉到小的风格细节。

BlendGAN[36]引入了基于 VGG 的风格编码器和权重融合模块，以在大规模的风格化面部数据集上学习任意的面部风格化，但是该方法无法捕捉到面部图像中的小但相关的风格细节。StyleGAN-NADA[37]使用 CLIP[38]根据图像提示进行图像风格化，具有非常强的泛化能力。朱等人的方法 MTG[24]使用了 GAN 反转来从参考图像中找到对应的真实面部，从而创建了一个配对数据点。MTG 使用了这个简单的配对化数据点和一些基于 CLIP 的损失函数来达到风格化的目的。朱等人使用的梯度下降反转 II2S[39]准确度极高但是较慢。一般而言，这些少训练样本风格化方法方法都从预训练的 StyleGAN 生成器开始，并使用各种损失和正则化进行微调。例如，GenDA[40]通过添加两个轻量级分类器将生成器调整到参考风格的领域，但缺乏源域和目标域之间的一对一对应关系，因此无法进行图像风格化。Ojha 等人的方法[33]引入了跨域一致性损失作为正则化项，以保持源域和目的域之间的相似性。虽然他们的方法保持了一对一的对应关系，因此可以用于图像风格化，但需要至少 10 张参考图像，并且每个模型只能容纳一个风格。

CtlGAN[41] 提出了一种少次技术，采用对比转移学习策略。OneShotCLIP[25] 使用 CLIP[38] 空间一致性损失来实现一次自适应。CLIP 模型由一个图像编码器和一个文本编码器组成，它们的目标是将语义上相似的图像和文本映射到相同的向量空间中。这样，我们就可以通过计算图像和文本向量的余弦相似度来衡量它们的语义相似性。CLIP 模型的目标函数如公式2-5所示，它试图最大化语义上相似的图像和文本之间的余弦相似度。

$$J = \max_{\theta} \frac{1}{N} \sum_{i=1}^{N} \frac{e^{s_{ii}}}{\sum_{j=1}^{N} e^{s_{ij}}} \tag{2-5}$$

其中，(s_{ij}) 是图像 (i) 和文本 (j) 的余弦相似度，它可以通过公式2-6计算：

$$s_{ij} = \frac{\mathbf{v}_i \cdot \mathbf{t}_j}{\|\mathbf{v}_i\| \|\mathbf{t}_j\|} \tag{2-6}$$

OneShotCLIP 首先使用 CLIP 模型对输入的文本和图像进行编码。在编码过程中，OneShotCLIP 还会接收一个样例图像，图像包含希望生成的图像的某些特性。OneShotCLIP 会将这个样例图像和输入的文本一起编码，以引导图像生成过程。编

码完成后，OneShotCLIP 使用 DALL-E[42]模型生成新的图像，这个输入可以表示为公式2-7。

$$\mathbf{z} = \text{concat}(\mathbf{v}, \mathbf{t}) \tag{2-7}$$

其中，$\mathbf{v}$ 是样例图像的向量表示，$\mathbf{t}$ 是文本的向量表示，$\mathbf{z}$ 是它们的组合。DALL-E 是一个生成模型，它能够根据输入的向量生成图像。在 OneShotCLIP 中，DALL-E 接收的输入向量是由 CLIP 编码的文本和样例图像的向量组合而成的。Mind the gap[24]也利用 CLIP 来确定源域和目标域之间的领域差距，并根据情况提供正则化，以防止过拟合。

Mind the Gap 和 OneShotCLIP 需要大规模的 CLIP 模块进行训练，从而增加了计算负担和非常高的训练时间。本文同样选择了从一个单一样本中创建了一个大型的配对数据集，与前文提到的选取的 GAN 反转方法不同的是，本文提出的 MultipleComicGAN 选择了使用基于 e4e[43]简单编码器的 GAN 反转，在计算效率上更高。MultipleComicGAN 可以同时一次处理多个样式，进行多风格的漫画化图像生成，每次根据不同的一或数张参考风格图片对 StyleGAN 基于真实人物图片的预训练模型进行不同的微调，并且支持用户自定义修改，以此减少过拟合以产生更好的结果。

2.3 MultipleComicGAN: 多风格漫画化图像生成 GAN 网络结构设计

本文采用了 Mind The Gap (MTG) 和 OneshotCLIP (OSC) 的思想，提出了一种新的通过对 StyleGAN2 进行微调来实现多风格漫画化生成的目的的 GAN 网络结构-MultipleComicGAN。MultipleComicGAN 的工作原理主要如以下步骤：首先对一个风格参考图像进行 GAN 反演，以获取捕获该风格的潜在代码。然后利用 StyleGAN2 的风格混合属性，生成一系列接近原始风格的风格代码，形成配对的训练集。这个过程的结构如图2-2所示。再其后使用这个新数据集对原始的 StyleGAN 模型进行微调，以更好地适应特定风格，这一过程由直接像素级损失指导。最后对于新的输入，微调后的 StyleGAN 应用学习到的风格，确保输出与参考风格保持一致。下面将详细介绍其工作步骤以及关键设计。

2.3.1 生成器与判别器的设计

MultipleComicGAN 的设计专注于风格化转换，主要通过改造和微调预训练的 StyleGAN2[13]模型来实现。MultipleComicGAN 的生成器与判别器的设计基本参考了

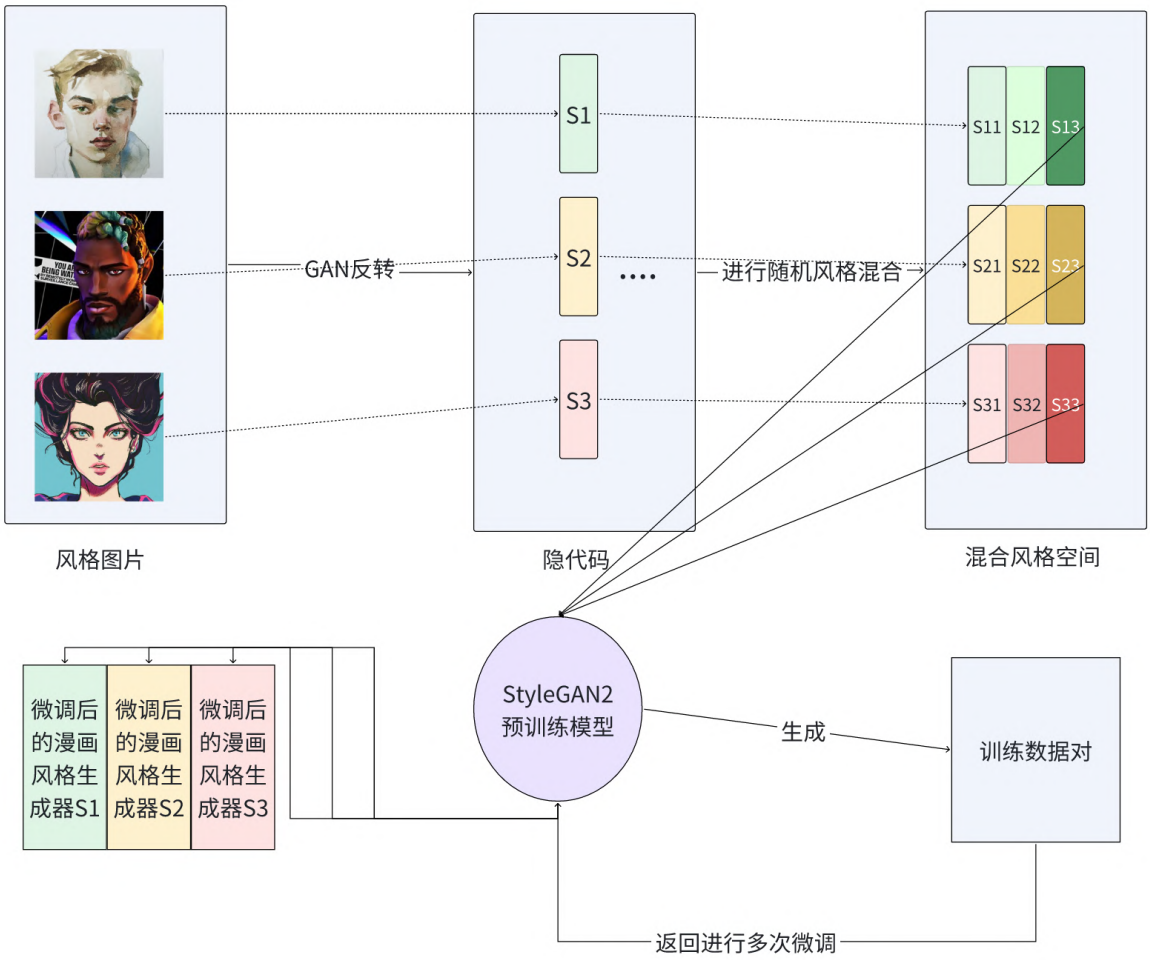


图 2-2 MultipleComicGAN 形成训练数据集训练的过程

StyleGAN2 的方法以及其对传统 GAN 的改进。MultipleComic 的核心不在于传统意义上的生成器和判别器的重新设计，而是在于如何巧妙地利用和调整预训练的 StyleGAN2 架构来实现快速和高质量的风格转移。StyleGAN2 的生成器和判别器设计强调了样式的精细控制和图像质量的提升，通过对生成过程中的每个步骤进行精确调整和优化，有效提升了图像的真实感和多样性。下面简要介绍 StyleGAN 相较于传统 GAN 的优势以及其对漫画风格化生成的适配性。

2.3.1.1 生成器

生成器的风格化关键在于 StyleGAN 增强了样式控制和图像细节的生成能力。StyleGAN2 的生成器采用样式代码来控制生成图像的每一层，通过所谓的逐层样式注入（通过 AdaIN 层实现）来影响图像的各个方面，如质感、颜色和形状。映射网络由多层感知机（MLP）组成，它将输入的潜在向量 z 映射到中间潜在空间 w。这个映

射过程通过学习到的非线性变换，提高了控制生成图像样式的能力。在 StyleGAN2 中，来自映射网络的 w 空间向量用于调制生成器中每一层的特征图的规范化参数（AdaIN）。这种方式允许不同的样式控制在不同层上实现，例如控制图像的粗略特征到细节特征。StyleGAN2 移除了 StyleGAN 中的样式混合和渐进式增长特性，改为持续整合所有分辨率，这有助于提升生成图像的整体一致性和质量。在 AdaIN（自适应实例规范化）[15]的基础上，StyleGAN2 引入了新的正则化层，使得生成的图像更加自然，减少了特定的伪影和纹理异常。

2.3.1.2 判别器

StyleGAN2 的判别器设计主要是为了更有效地识别生成图像与真实图像之间的差异，判别器采用了更为标准化的卷积神经网络架构，用于从高分辨率图像中逐步下采样到较低分辨率，直至最终输出一个单一的判别分数。与 StyleGAN 不同，在 StyleGAN2 中，判别器对所有分辨率的处理更为均匀，避免了对任何特定分辨率的偏见，增强了模型的泛化能力。生成器判别器工作目标函数如公式2-8所示。

$$\min_G \max_D V(D,G) = \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))] \tag{2-8}$$

2.3.2 GAN 反转

GAN 反转是 MultipleComicGAN 的关键技术，它允许我们找到对应于特定真实图像的潜在向量（latent vector），这个向量在输入给预训练的生成对抗网络（Generative Adversarial Network, GAN）的生成器（Generator）时，能够重建或近似这个真实图像。GAN 反转的目的是找到一个潜在空间中的点，该点通过 GAN 的生成器能够生成一个与目标图像非常相似的图像。设 G 是一个训练好的 GAN 生成器，它从潜在空间 Z 映射到图像空间 X 即 $G : \mathcal{Z} \to \mathcal{X}$，给定一个真实图像 $x \in \mathcal{X}$，使得 $G(z)$ 尽可能接近 x。这可以通过优化目标函数2-9来实现。

$$z^* = \arg\min_z \|G(z) - x\|^2 \tag{2-9}$$

在公式中，$\|G(z) - x\|^2$ 是一个损失函数，通常选择为欧氏距离的平方，用来衡量生成图像 $G(z)$ 和目标图像 x 之间的差异。MultipleComicGAN 使用 e4e 的 GAN 反转的方法，首先将一或数张参考风格图像 y 进行 GAN 反转，以获取风格代码 $w = T(y)$，并从中得到一组参数 $s(w)$，这一步使用了一个预训练好的 GAN 反转编码器。其后，需要根据原始的 $s(w)$ 生成一系列变体 $\{w_i\}$，这些变体反映了原始样式的微小变化，

以此来找到一组与 $s(w)$ 风格接近的风格代码 S，这一步通过利用 StyleGAN 的风格混合机制，既可以增加输出多样性，同时保持风格的一致性。具体的实现过程中使用了具有 26 个风格调制层的 1024×1024 分辨率的 StyleGAN2，因此 $s \in \mathbb{R}^{26\times512}$，定义固定掩码 $M \in \{0,1\}^{26}$，FC 为 StyleGAN 的风格映射层，其中 $z_i \sim \mathcal{N}(0,I)$。使用公式2-10生成新的风格代码，并对每一个批次都如此做，以此来得到一组风格代码 S 传入 StyleGAN 预训练模型对其进行微调。

$$s_i = M \cdot s + (1-M) \cdot s(FC(z_i)) \tag{2-10}$$

2.3.3 根据反转产生的隐代码对 StyleGAN 进行微调

本步骤的关键是通过对 StyleGAN 进行微调来获得一个 $\hat{\theta}$，使得 $G(s_i;\hat{\theta}) \approx y$ 成立。这个过程通过最小化生成图像与参考图像之间的差异来实现。微调的目标是调整 StyleGAN2 以便它能够更好地复现指定风格的图像。首先使用 FFHQ 真实人脸数据训练集[12]训练一个 StyleGAN2 的训练模型，该数据集包括了 70,000 张高质量的人脸图像，以此来训练一个能生成真实人物图片的 GAN 生成模型。首先假设一个经过适当训练的样式映射器可以将 S 中的 s_i 映射到 y。这个假设肯定是有效的，并且在样式映射器“减少信息”的情况下是合理的——例如，将眼睛大小或头发纹理略有不同的面部映射到同一参考图像。微调的过程是对 StyleGAN 进行微调以得到最小化的衡量生成的图像 $G(s_i;)$ 与目标样式 y 之间的差异损失函数 L，如公式2-11所示。

$$\hat{\theta} = \underset{\theta}{\text{argmin}}\,\text{loss}(\theta) = \underset{\theta}{\text{argmin}}\,\frac{1}{N}\sum_{i}^{N}\mathcal{L}(G(s_i;\theta),y) \tag{2-11}$$

对差异损失函数 L 的选取至关重要，其具体过程会在2.3.4中介绍。

2.3.4 损失函数的设计

在对于对 StyleGAN 进行微调的设计中，例如 pixel2style2pixel[44]，通常来讲可以选择 LPIPS[29]作为损失函数设定，然而，LPIPS 的设计意味着如果选择其作为损失函数的话，在风格化的过程中会丢失许多细节。LPIPS 建立在经过 224×224 分辨率训练的 VGG[45] 骨架上。为了达到我们生成高分辨率保留原人物图片细节的漫画风格化图像的目的，MultipleComicGAN 应该能够生成 1024×1024 分辨率的图像。处理这种不匹配的标准方法是在计算 LPIPS 之前将图像下采样到 256×256[13,43,46]。但这种下采样意味着我们无法控制细粒度的细节，这些细节大多数会丢失。同样，以原生

适配的 1024×1024 分辨率计算 LPIPS 会导致完全丢失细粒度的细节，因为 VGG 的过滤器（使用 224×224 分辨率训练）无法适应这种分辨率。

因此，为了满足对生成的漫画图像的要求，需要重新选取一个损失函数的值。参考 GPEN[47]对损失函数的选择方法，在此处选择使用特定层上的判别器激活之间的差异作为损失函数，该损失函数对每个图像计算一次，损失函数的公式如公式2-12所示，其中 $D(y)$ 表示 y 的激活函数。

$$\mathcal{L}(G(s_i;\theta),y) = \| D(G(s_i;\theta)) - D(y) \|_1 \tag{2-12}$$

以此损失函数，基于2.3.3中产生的成对的训练数据对 StyleGAN 进行多次迭代微调，最终得到一个微调后的 StyleGAN2 的能够生成参考风格图像同风格的图片的生成模型将其保存。

2.3.5 生成漫画风格化后的图片

用本章设计的 GAN 架构方法，对于一个输入的真实人物图片 u，我们能够得到一个漫画风格化后的漫画图像 $G(s(T(u));\hat{\theta})$，如图2-3所示。所以，漫画化图像生成的映射过程为 $G \circ s \circ T$，根据这个风格映射对输入的不同风格的参考风格图片进行对 StyleGAN2 预训练模型的微调，可以产生一个微调后的预训练模型组，保存后续交由用户进行选择生成何种风格化的图片，达到将输入的真实人物图片进行轻量级且准确保留人物细节多风格漫画化的效果。兼顾轻量化和尽可能多的生成数量的考虑，MultipleComicGAN 至多支持对 4 张输入真实人物图片进行 5 种风格的漫画风格化图像生成。

2.4 本章小结

本章详细介绍了基于真实人物图像的多风格漫画化图片生成 GAN 方法—MultipleComicGAN。此方法借鉴了 StyleGAN2 的架构，并对其进行了针对特定风格化需求的微调和优化，使其能够处理多种风格的漫画化图像生成。

首先，本章回顾了生成对抗网络（GAN）的基本原理，解释了生成器和判别器的工作方式以及它们如何通过对抗过程来提升图像生成的质量。接着，详细探讨了 StyleGAN 架构及其在人脸图像生成上的应用，尤其是介绍了 StyleGAN 的核心特点，包括样式调制、自适应实例归一化（AdaIN）技术，以及如何通过映射网络将输入的高斯噪声转换为有控制的样式代码。

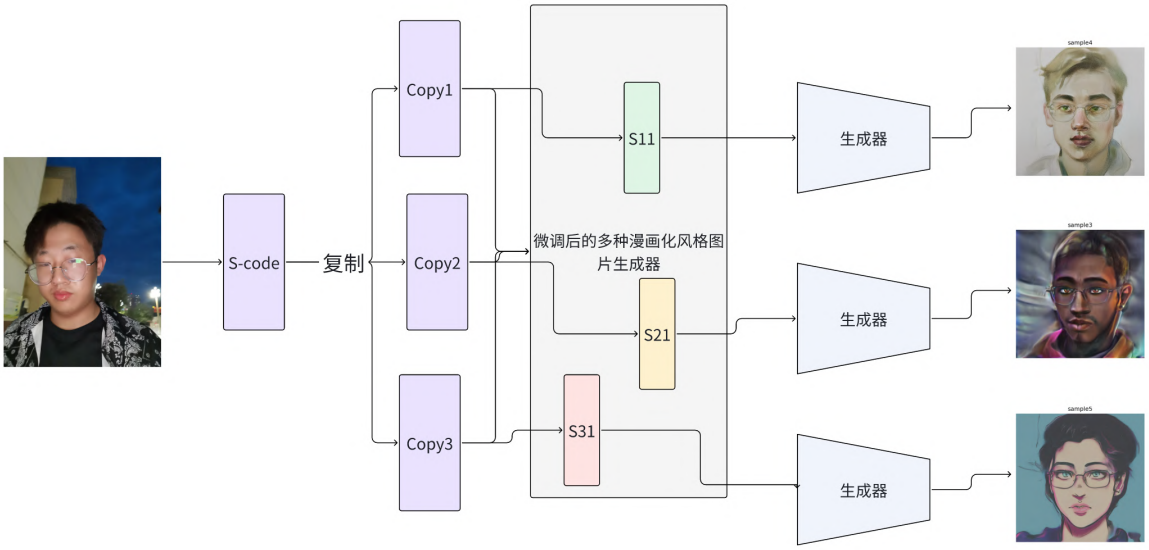


图 2-3 MultipleComicGAN 生成漫画风格化图片的过程

核心部分详述了 MultipleComicGAN 的设计思想和实现步骤，包括 GAN 反转技术的应用，它允许模型从单一或少量的风格参考图像中学习并生成新的风格代码，以及如何利用这些代码对 StyleGAN 进行微调以生成具有特定艺术风格的漫画图像。此外，本文还讨论了针对不同风格化需求如何选择合适的损失函数，以及如何利用判别器的激活差异来优化生成过程，从而更好地保留人物的细节特征。

最后，展示了通过这种新颖的 GAN 架构如何将真实人物图像转化为多种漫画风格的图像，实现了从技术理论到实际应用的跨越。MultipleComicGAN 的设计不仅提供了一种有效的技术路径来探索图像风格化的多样性，也为未来的图像生成研究提供了新的可能性，特别是在风格多样性和图像质量之间寻找最优平衡点的探索。

第 3 章 基于 MultipleComicGAN 的漫画风格化生成

3.1 漫画风格化生成

3.1.1 基于真实人物图像的预训练模型

在第 2 章的整体架构设计中已经讲到，为了实现漫画风格化的目标，我们要对基于真实人物图像训练的 StyleGAN2 预训练生成模型进行微调。在此处数据集选择了 Flickr-Faces-HQ（FFHQ）[12]高清人脸数据集。该数据集包括了 70,000 张高质量的人脸图像。这些图像已经根据不同的需要被处理成不同的的分辨率，在此处，我们的目的是得到高清晰度的漫画风格化的图像，所以选择了 1024×1024 像素的 FFHQ 数据进行训练。在训练前，所有的图像首先被归一化，使得像素值位于 [-1, 1] 区间内。此外，对图像进行中心裁剪和缩放，确保面部特征在每张图片中大致位于相似的位置，这有助于模型更好地学习和重建面部结构。通过 StyleGAN 的映射网络以及渐进式训练，达到最小化感知损失的目的，使用真实人物图像数据集训练后的 StyleGAN2 预训练模型的随机生成真实人物图片如3-1所示。

图 3-1 基于真实人物图像训练的 StyleGAN2 预训练模型的随机 6 张真实人脸照片生成结果

3.1.2 控制风格化的强度

在一些情况下，由于原风格照片过强的结构与风格，会导致我们得到的漫画风格化后的图片出现与参考风格图片过拟合的现象 (如图3-2)，可以看到由于过分保持原风格图片的色彩以及结构，导致了人物图片的一些身份特征丢失，例如头发以及脸型，同时整张图片也因为过量的色彩导致人物脸型变形观感奇怪。在这种情况下，原人物图像的很多细节直接丢失了，甚至于基本外貌特征存在了被改变的情况，为了解决此种问题，我们需要控制风格化的强度使其不能超过某个特定的阈值。与此

同时，在漫画风格化的实际应用中，不同的用户会偏好不同的风格化强度下漫画风格化的真实人物图片。因此，MultipleComicGAN 引入了能够定量化的进行风格化强度的控制的机制。为了实现对漫画风格化强度的控制，MultipleComicGAN 使用了双重机制：保持原真实人物图像人物身份的余弦相似度控制以及能够量化控制漫画风格化强度的特征插值控制机制。

图 3-2 过拟合的漫画化生成图像

为了防止一些参考风格漫画图片会使输入的真实人物图片的特征信息被扭曲，在这种参考风格严重影响生成的漫画化图像结果的情况下，将 sim 记为余弦相似度，将 F 记为预训练的面部嵌入网络 ArcFace[48]，使用如公式3-1来定义输入的真实人物的身份损失，以此来强制执行使得原输入真实人物图片的结构以及身份信息特征得以保存。

$$\mathcal{L}_{id} = 1 - sim(F(G(s_i;\theta)), F(G(s_i;\hat{\theta}))) \tag{3-1}$$

使用此方法可以强制保持原真实人物图片的结构以及人物特征。

在量化控制漫画风格化强度的过程中选择了通过特征插值控制风格强度的方式进行风格化强度的控制。特征插值[49]使我们能够定量化的改变控制漫画风格化的强

度。定量控制漫画风格化强度的特征插值法的公式如3-2所示。

$$f = (1-\alpha)f_i^A + \alpha f_i^B \tag{3-2}$$

其中 f_i^A 代表使用原始生成真实人物图像的预训练模型的原始 StyleGAN2 的网络中的第 i 层中间特征图，f_i^B 为使用了微调后的漫画风格生成模型的 MultipleComicGAN 网络的第 i 层中间特征图，α 是定义的插值因子参数。增加 α 的值会导致更强的风格化生成。在此机制下，可以通过控制定量且连续的改变 α 的值来达到定量化控制风格强度的目的如图3-3，风格强度控制机制允许用户自定义控制风格的强度，随着 α 的提升，风格化程度会变得越来越强。

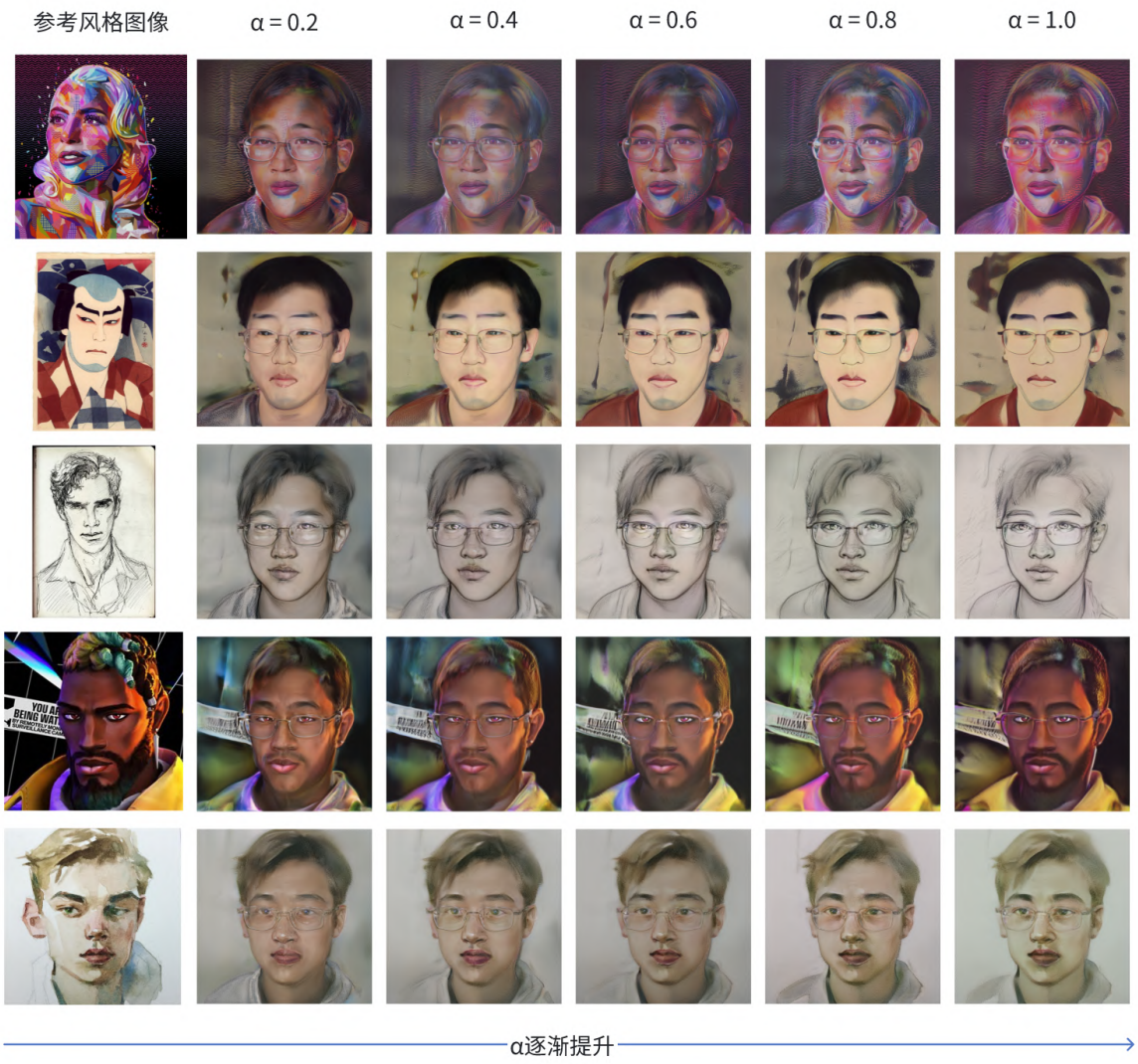


图 3-3 风格强度控制机制的结果

3.1.3 保留色彩进行风格化

色彩在漫画风格化的过程中扮演着关键但复杂的角色。尽管色彩并非漫画风格定义的全部，但它极大地影响着视觉感受和风格的传达。色彩的运用不仅可以增强视觉效果，还可以用来表达情感、设置氛围或强调特定的主题。在将真实人物图像转换为漫画风格的图像的过程中，对色彩的处理需要特别的注意，以确保最终结果既忠实于原图，又能反映出漫画的艺术风格。

在很多特定情况的任务场景中，漫画风格化会要求风格化生成的图片不仅保留了原真实人物图像的结构与面部身份特征，参考风格图像的风格，还应保留参考风格图像的色彩。色彩是否属于风格是较为模糊难以说明的，例如瞳孔的颜色大部分情况下不应该被归类到风格之中，但是发色则是属于有时可以归类于风格，有时则不属于风格的一部分，甚至于还有一些漫画创作者的风格特点之一就是夸张七彩的头发颜色。由此可见，色彩属于但不完全属于是风格的一部分，所以 MultipleComicGAN 可以自行交由用户选择是否保留原参考风格图像的色彩。控制是否保留色彩的机制原理在于简单的选择性遮罩和损失以及对 GAN 反转方法的控制。

如果 GAN 反转器从参考图像生成贴近真实人物风格，与参考风格图片相似的面部，GAN 以贴近真实人物图片的 s_i 映射到风格参考上，由此做将倾向于生成极具风格化的面部。与此同样的，如果 GAN 反转器从参考风格图像生成漫画风格化的漫画图像，微调后的 GAN 倾向于生成轻度风格化的面部，并保留输入面部的特征。这种效果可以用来控制转移的风格及其程度，通过混合反演的代码实现。

当然，实际上使用两个 GAN 反转器与构建轻量化的模型的目的不符。与此同时，平均风格代码是对不存在的的图像 y_i 的 $s(T(y_i))$ 的最大似然估计，同时也可以作为一个 GAN 反转的输出。我们通过将标准 GAN 反转器产生的代码与平均值混合，以此利用公式2-9来产生一个虚拟反演器 V(y)。调整混合使得 $G(s(V(y));)$ 具有最理想的属性。然后我们使用虚拟反演结果代替真实反演出的隐代码生成训练数据。最后如公式2-11进行微调训练，计算 $G(s(T(u);\hat{\theta})$。控制是否保留色彩进行风格化的结果如图3-4所示，如果选择保留色彩那么会生成风格化更强烈的漫画图像，对参考风格图像的色彩保留的更好；如果选择不保留色彩那么会生成轻度风格化的漫画图像，更贴近原图片的结构与色彩。

(a) 参考风格图片

(b) 真实人物图片

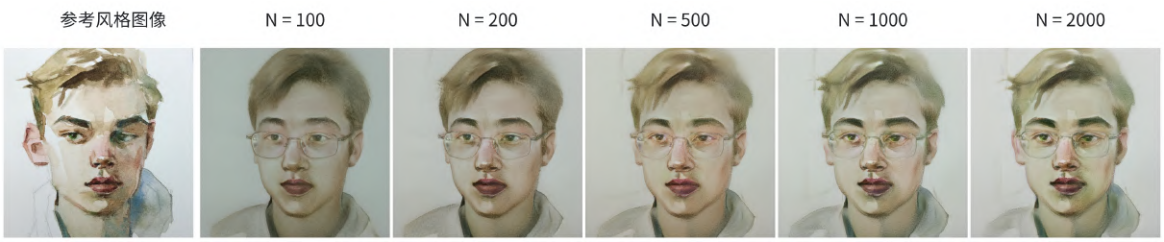
(c) 保留色彩的漫画化效果

(d) 不保留色彩的漫画化效果

图 3-4 控制是否保留色彩的生成结果

3.1.4 迭代轮次的选择

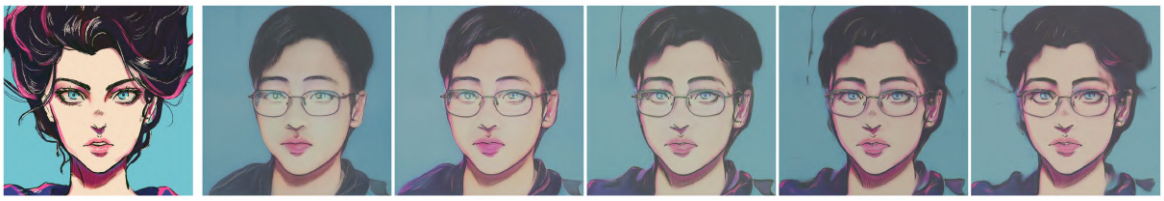
由2.3中介绍的 MultipleComicGAN 的原理可知，对 StyleGAN2 进行的迭代微调次数对于最后的漫画风格化生成模型的生成结果至关重要。选择合适的迭代轮次会影响迭代微调的效率、生成图像的质量和风格的准确性。图3-5中展示了对四种参考风格图像在不同的迭代微调次数下（N=100，200，500，1000，2000），以四种参考风格图像对 StyleGAN2 的预训练生成器进行微调，微调后的漫画生成器对真实人物图片的漫画化生成结果。



(a)

(b)

(c)

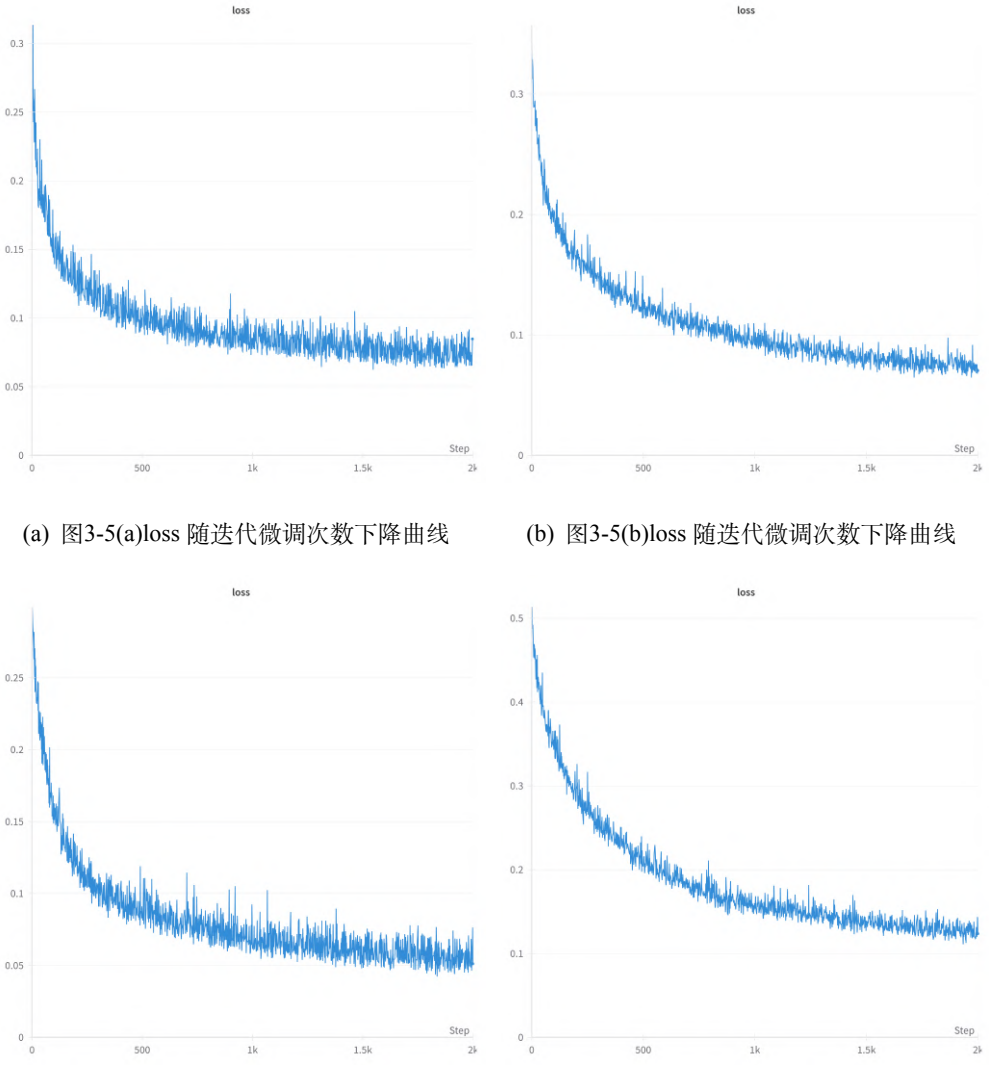


(d)

图 3-5 在不同迭代次数下的漫画化生成结果

如图3-6所示，损失函数的值随着迭代次数的增加而降低，但是在过多的迭代轮次下并不能显著降低 loss 即提高对预训练模型进行漫画化微调的效果。根据2.3.4中设计的感知损失损失函数进行图像风格化效果的判断，进行迭代微调的次数越多显然风格化的效果越好，过小的迭代轮次会导致风格化的效果较弱。然而，过多的迭代轮次下（N>1000 时），继续增加迭代微调的次数并不能够继续有效的降低损失函数的值，达到继续微调调优预训练模型的效果。与此同时，过多的迭代轮次会导致算力与时间的大量消耗浪费，与设计一个有效跨平台的轻量化的网络的目的不符。因

此，选择一个适当的迭代微调次数对于我们生成符合要求的漫画风格化的图像是十分必要的。根据对多种不同的漫画风格的损失函数的评估，通常来讲，选择迭代次数 N 处于 500-800 之间是一个合理的范围，对不同的参考风格图片会存在不同的最佳迭代次数。不同偏好的用户也可能会偏向于不同的风格化强度。因此，交互页面支持用户自由调整迭代轮次，详细内容将会在3.2.3中介绍。



(a) 图3-5(a)loss 随迭代微调次数下降曲线　(b) 图3-5(b)loss 随迭代微调次数下降曲线

(c) 图3-5(c)loss 随迭代微调次数下降曲线　(d) 图3-5(d)loss 随迭代微调次数下降曲线

图 3-6 损失函数的下降曲线

3.2 训练过程

3.2.1 自定义训练风格图片

MultipleComicGAN 允许用户自行上传至多四张自己想要进行风格化的真实人物图片，以及若干张参考风格图片进行微调训练。使用多张参考风格图片进行风格生成器的微调可以对风格映射 $G \circ s \circ T$ 有效的提升，避免3.1.2中提到的风格化强度过高问题，实现更好的风格化效果，具体在多张参考风格图片上的处理方法如公式3-3所示。在此步骤中可以选择是否使用

$$\frac{1}{M * N} \sum_{j}^{M} \sum_{i}^{N} \mathcal{L}(G(s_{ij}; \theta), y_j). \tag{3-3}$$

其中，y_k 表示第 k 个风格参考漫画图片，s_k 表示由该第 k 张风格参考漫画图片生成的风格空间代码，其中对每一个 y_k 都会构建一个 s_k。在使用完毕用户自己上传的参考风格图片对 StyleGAN 预训练模型进行微调结束后，用户可以自行选择是否使用3.1.3中提到的保留色彩风格化方法，之后自定义微调后的模型名称进行保存。

3.2.2 自定义训练参数

在 MultipleComicGAN 支持用户自定义对3-3中的 α 进行调整，以此调整风格化的强弱；决定是否选择使用3.1.3中提到的保留色彩风格化方法，决定是否保留原参考风格图像的色彩；以及对3.1.4微调迭代轮次的自行调整。

用户在输入若干张参考风格照片后，可以通过对以上三种参数的调整，自行选择训练生成符合自己喜好的漫画风格化生成模型。在一轮训练结束后，会展示以该参数设置下进行对预训练模型微调后的漫画化生成器生成的漫画化图像结果。当用户对漫画化效果不满意时，可以自行调整参数多次进行微调训练进行模型的保存。

3.2.3 生成结果

MultipleComicGAN 支持同时进行至多 4 张真实人物照片的 5 种不同风格图像生成，这些风格可以在预训练的微调好的风格模型中选择或者是使用上述的方法训练出的风格，图3-7展示了四张同一人物的不同图片使用五种不同的风格进行漫画风格化的结果，最左侧为参考风格图像。

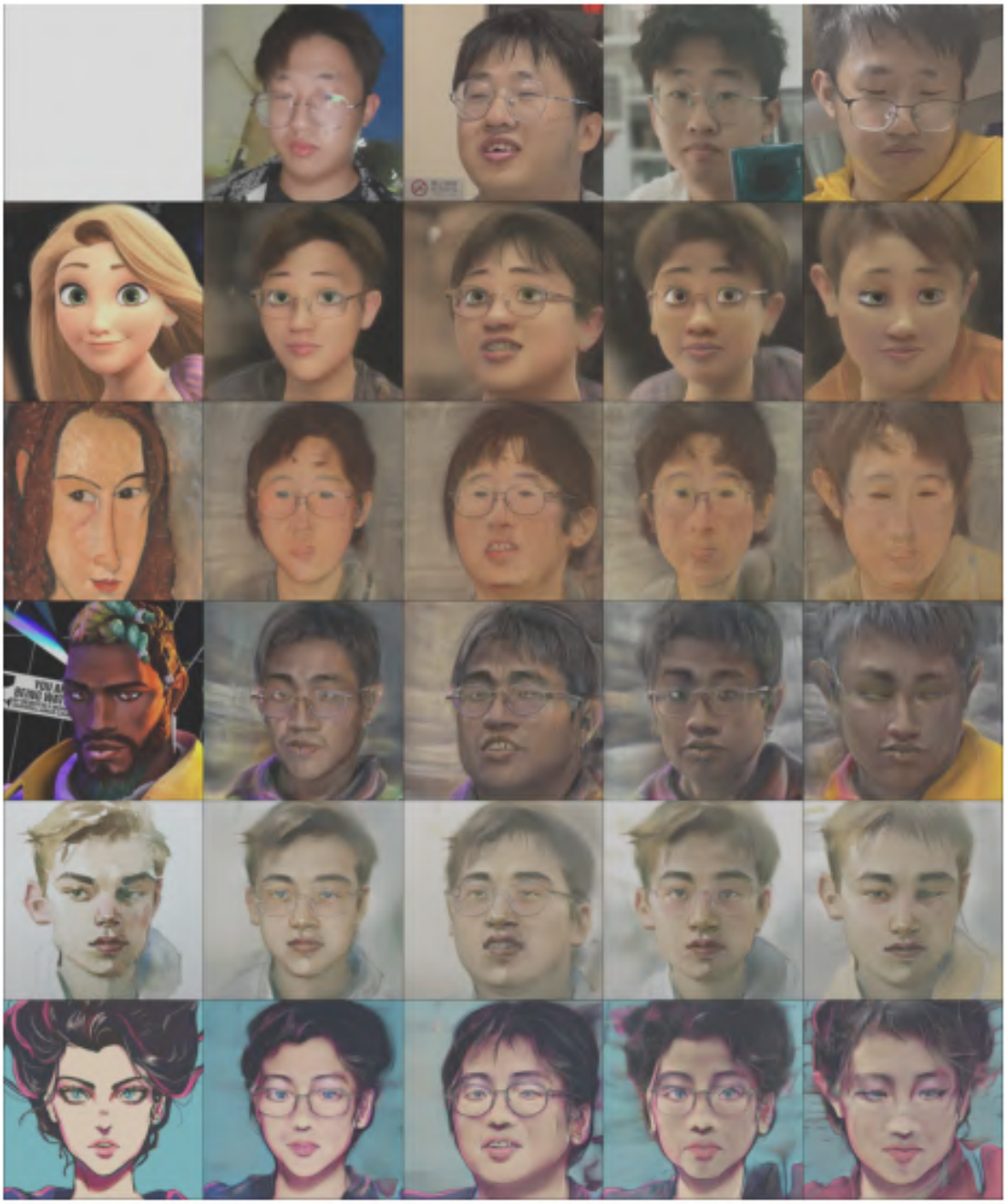


图 3-7 MultipleComicGAN 对四张真实人物图片的多种漫画风格生成结果

3.3 本章小结

本章详细介绍了基于 MultipleComicGAN 的漫画风格化生成方法及其界面设计，提供了一个全面的框架来实现真实人物图像的漫画风格化，同时确保了用户操作的简便性和生成结果的高质量。

首先，概述了基于真实人物图像的预训练模型，使用了 FFHQ 高清人脸数据集对 StyleGAN2 进行预训练，生成了高质量的真实人物图像。接着，介绍了控制风格化强度的方法，包括保持身份的余弦相似度控制和特征插值控制，解决了风格化过拟合的问题。此外，还详细讨论了保留色彩进行风格化的机制，通过选择性遮罩和虚拟反演的方法，实现了对漫画风格化色彩的控制。

然后，分析了迭代轮次的选择对风格化生成效果的影响，提供了合理的迭代次数范围，确保生成高质量的漫画风格化图像。接着，介绍了 MultipleComicGAN 支持用户自定义训练风格图片和参数，为用户提供了灵活便捷的漫画风格化生成体验，并在最后进行了 MultipleComicGAN 生成结果的展示。

本章内容展示了 MultipleComicGAN 的灵活性、可定制性和实用性，为用户提供了多样化的漫画风格化生成选择。

第 4 章 实验结果与分析

在本章节中，首先将详细介绍 MultipleComicGAN 的参数配置，具体实现细节，然后介绍本次实验使用的评价指标，并用该几项指标与最为相关的两种风格化方法-Mind The Gap(MTG) 以及 OneshotCLIP(OSC) 进行定量化的评估比较。与此同时，还将 MultipleComicGAN 与 MTG，OSC 的生成结果做成调查问卷的形式交由人眼评价进行定性化的评估比较。最后将 MultipleComicGAN 与 MTG 以及 OSC 的训练相同风格数量需要的训练时间，训练需要占用的存储空间进行了比较以及轻量化分析。

4.1 参数配置

在所有实验参数配置中，预训练模型均是从预训练的 StyleGAN2 开始，预训练数据集选择在 FFHQ 数据集上以 1024×1024 分辨率进行训练，全部使用 NVDIA 官方给出的实现的参数配置。对于 MultipleComicGAN 独有的参数设置，在本次实验参数配置中如下：使用 e4e[43]进行 GAN 反转，特征插值控制参数 α 选择固定为 1；每轮训练对 StyleGAN 的预训练模型微调次数设定为 800；损失函数的设置上选择了使用2.3.4中提到的感知损失函数；并使用 Adam 优化器在 0.002 的学习率下进行迭代训练生成器。对于 Mind the Gap 和 OneshotCLIP，选择使用它们给出的官方实现中的默认参数配置。

具体的软硬件的参数配置如4-1

表 4-1 硬件、软件环境

	指标	版本参数
硬件环境	CPU	Intel Xeon E5
	RAM	64 GB
	GPU	NVIDIA Tesla V100 16GB
软件环境	操作系统	Windows 10 Pro x86_64 Ubuntu 20.04.3 LTS
	Python	Python 3.8.6

4.2 图像评估指标

选择何种的量化图像评估指标，关键在于解决风格是什么的问题。漫画风格是一种艺术表达方式，MultipleComicGAN 要解决的是基于真实人物图片的漫画风格化问题。生成后的漫画化风格图像应该既保留了原人物真实图片的人脸长相面部特征细节，又在其中加入了参考风格漫画图片的线条色调结构。所以将风格的评价拆解为两个方面：与真实人物的图片特征细节相似度，即风格化后的图像的真实性以及与参考风格漫画图像的结构相似性，即风格化后的图像的风格保持。为了达到对这两方面进行评估的目的，选择使用 SIFID[50]以及 SSIM[51]进行定量化的评估。

SIFID 是 Shaham 等人于 2019 年提出的基于 GAN 评价常用的评价指标指标 FID 的专门针对于单张图片的评价指标。与传统的 FID 不同的是，SIFID 选择使用在 Inception 网络 [49] 的最后一个池化层 (每个图像一个向量) 之后的第二个池化层 (图中每个位置一个向量) 之前使用卷积层输出的深层特征的内部分布。传统的 FID 测量的是生成图像的深度特征分布与真实图像的分布之间的偏差，然而在本次实验中，对每一张生成的漫画化的图像只存在一张对应的真实人物图像，而不是大批量的真实人物数据集进行训练生成虚拟真实人物图片。所以，使用单图像 FID(SIFID) 更符合本次实验的漫画化图像与真实人物的图片特征细节相似度评估指标要求。

SSIM 是通过分析结构相似度来衡量图片的失真程度的感知模型。SSIM 是一个用于测量两幅图像之间视觉相似性的指标，SSIM 的思路是-人眼探测图像的结构信息是图像质量感知的重要组成部分。因此，SSIM 试图先分别通过测量两幅图像在亮度、对比度和结构这三个维度的相似性来计算它们的视觉相似度如公式4-1所示，之后将这三个值相乘得到最终的结果如公式4-2所示。

$$\begin{cases} l(x,y) = \frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1} \\ c(x,y) = \frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2} \\ s(x,y) = \frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \end{cases} \tag{4-1}$$

$$SSIM(x,y) = l(x,y) \cdot c(x,y) \cdot s(x,y) \tag{4-2}$$

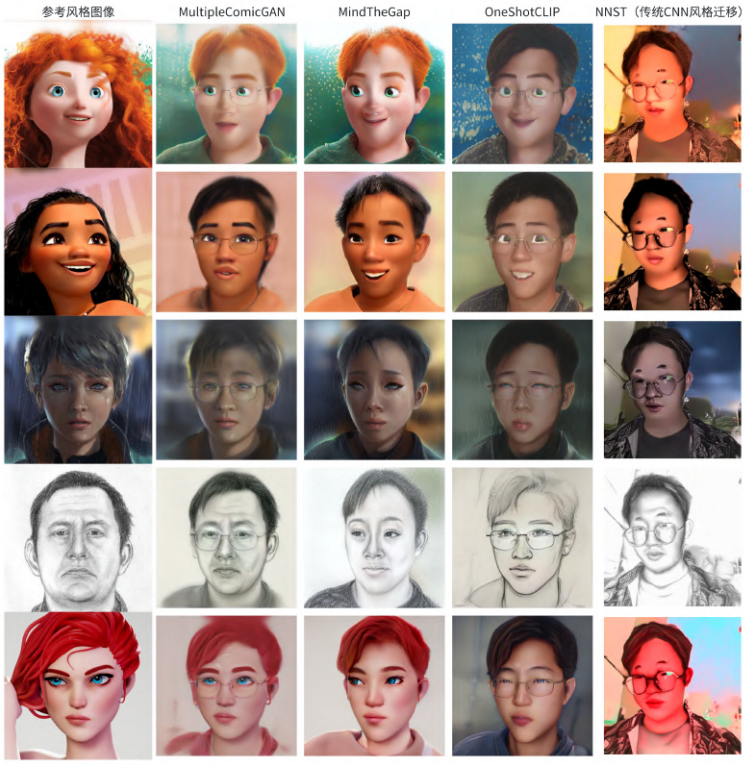
SSIM 所使用的三种维度的评估指标：亮度，对比度与结构对应于漫画风格的画风风格特点：风格可以看作这三个维度因素的结合。因此，计算漫画化的图像以及参考风格图像的结构相似性 SSIM 作为判断是否同属于同一风格的量化评估指标是符合要求的。

4.3 MultipleComicGAN 生成的实验结果比较

(a) 真实人物图片 A　　(b) 真实人物图片 B

图 4-1 两张真实人物图片

图4-2展示了对图4-1的这两张人脸照片分别使用 MultipleComicGAN，Mind The Gap(MTG)，OneshotCLIP(OSC) 以及 NNST 进行漫画风格化后的多种风格漫画化生成结果，分别进行五种使用不同风格参考风格图像的漫画化生成结果进行比较。从左到右依次是参考风格图像，MultipleComicGAN，MTG，OSC 以及 NNST 的生成结果。后续章节将对这几种方法进行详细的对比，定性分析，用户评价以及与定量化评估。



(a) 对4-1(a)中人物图片的漫画化生成结果

(b) 对4-1(b)中人物图片的漫画化生成结果

图 4-2 不同方法的漫画化生成结果

4.3.1 定性评估比较

良好的漫画风格化方法映射应该既能保留原真实人物的特点与身份特征，又能合理的把参考风格图像的风格迁移到生成的漫画化图像上。从图4-2可以看到，MultipleComicGAN 基本可以做到这两种特点，而且在一些风格上的表现比 MTG 以及 OSC 风格化的效果更好。MTG 在很多情况下会过度扭曲面部形状（如下巴、鼻子等），OSC 在很多时候会导致对人物的面部特征保持的不够良好。而 MultipleComicGAN 由于在微调中使用了像素级损失，能够生成一致的形状。还能发现与 MTG 和 OSC 相比，MultiStyleGAN 更一致地保留了输入图像的身份和表情。由于风格化本质上是主观的，通常公认的评估风格化效果的方法是进行人类评估研究。因此，为了进一步研究人眼视觉对于不同漫画风格化方法的倾向，进行了用户研究，将 MultipleComicGAN 与 MTG 和 OSC 进行比较，以制作调查问卷评估的形式来让用户评价他们更喜欢哪种风格化方法。在调查问卷中，每位参与者都会看到一个输入的真实人物图片和一个参考风格，以及三种不同方法的输出，之后选择他们最喜欢的以及最能保持原风格的结果，每位参与者会给出 10 组这样的题目。调查问卷的设计如图4-3所示，每位参与者会被给出 10 组这样的问题。

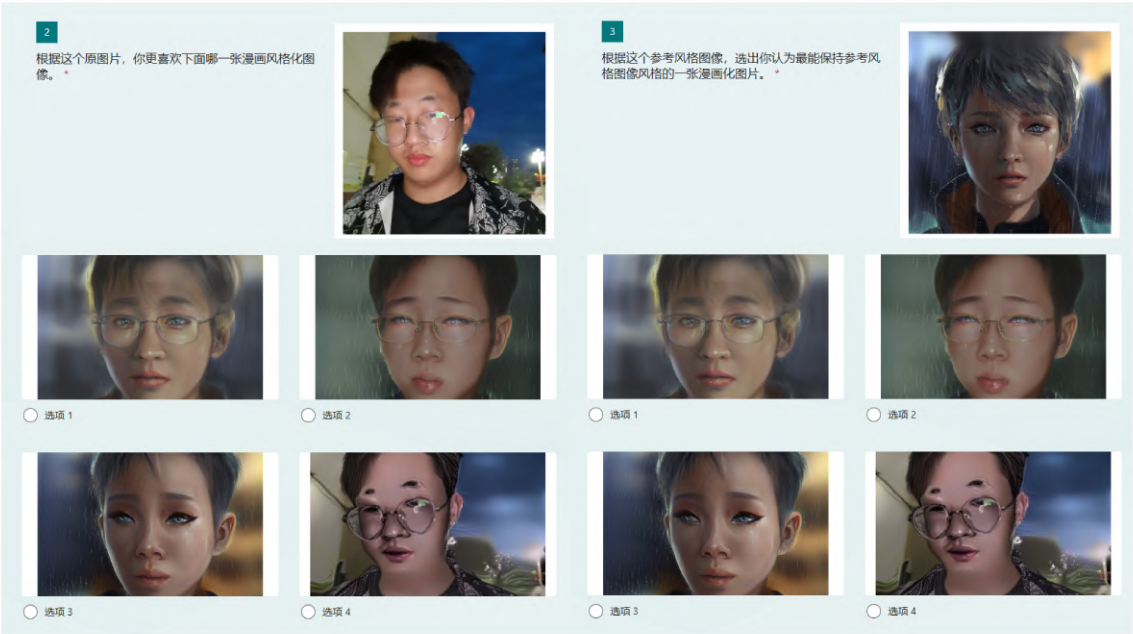


图 4-3 调查问卷的样式

最终共计收到了 124 份有效问卷，对调查问卷结果进行统计分析，最终的用户评价结果如图表4-2所示，从用户评价分析比较来看，有 37.8% 的用户更喜欢 MultipleComicGAN 生成的漫画化图像，与 OneShotCLIP 的 35.16% 大致持平，是 MindTheGap 的 3.5 倍。可见 MultipleComicGAN 在创造出人眼更偏好的漫画风格化图片上具有一定的优势，并且也有 42.58% 的用户也更倾向于认为 MultipleComicGAN 生成的漫画风格化图片与原参考风格图片属于同一种风格，与 MindTheGap 的 41.45% 持平，远远领先于 NNST 和 OneShotCLIP。

表 4-2 对调查问卷的用户评价进行的统计比较结果

方法	风格偏好得票	风格偏好占比	风格保持能力得票	风格保持能力占比
MultipleComicGAN	469	0.3782	528	0.4258
MindTheGap	136	0.1096	514	0.4145
OneShotCLIP	436	0.3516	128	0.1032
NNST	199	0.1604	70	0.0564

在问卷的结果统计中，发现用户更偏好哪种漫画化方法会根据评价的风格的不同而发生显著的不同。例如，虽然 NNST（传统的 CNN 风格迁移方法）在两项投票中的得票均为最低，然而在针对素描风格的评价中，有 57% 的用户都将“最喜欢的图像”投票给了 NNST，而在水彩风格的投票中，NNST 的偏好得票占比只有 1.5%（即 124 票里只得了 2 票）。由此可见，参考风格图像的色彩，笔触等会对用户的倾向具有较大影响。在后续对 MultipleComicGAN 进行训练调参优化的过程中，可以参考这一点，针对不同的风格类型预设不同的训练参数。

4.3.2 定量化评估比较

本文使用的定量化评估如4.2所示，对三种 GAN 方法生成的漫画化图像分别与真实人物图像求 SIFID 以及与参考风格图像求 SSIM，最后对 50 种风格的漫画化图像计算出来的值取平均值进行比较。使用的评估方法是先分别使用三种方法：MultipleComicGAN，MTG 以及 OSC 进行 50 种风格的真实人物图像的漫画化图像生成，之后对生成的漫画化图像与真实人物图像进行 SIFID 的计算，以此来判断其对人物身份特征以及脸型结构等的保持效果，之后将生成的漫画化图像与参考风格图像进

行 SSIM 值的计算，以此来判断生成的漫画风格化图像是否能够较好的保持原参考风格图像的结构，色彩等要素，最后对算出的所有数值取平均值，如图4-4所示。

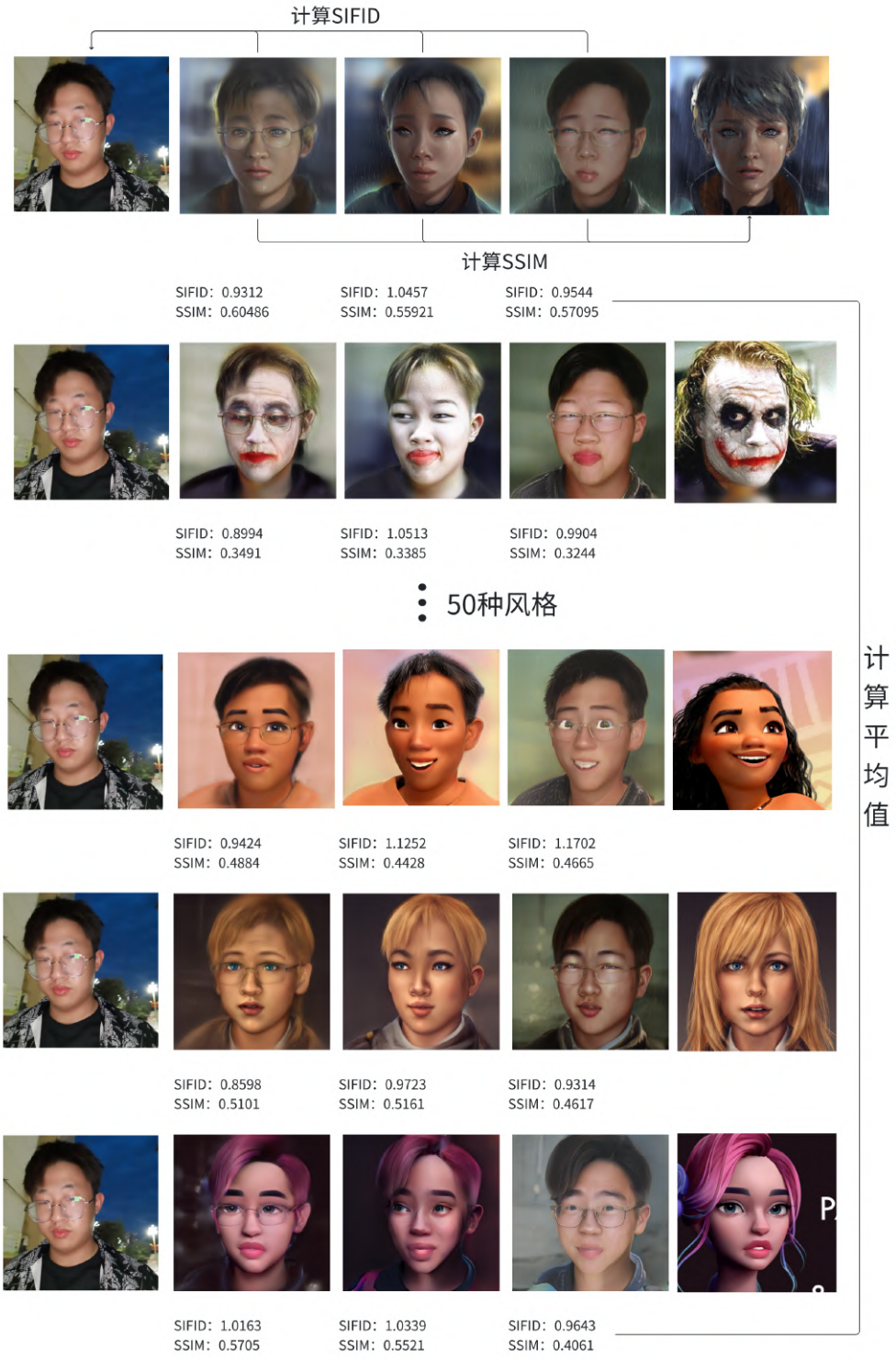


图 4-4 定量化评估的方法

在图4-3中展示了对三种漫画化方法的算出的 50 种风格的 SIFID 以及 SSIM 的平均值量化比较结果。我们可以看出，MultipleComicGAN 在 SIFID 上得到了 0.893617，大于 MTG 的 1.00532 与 OSC 的 0.986671，越低的 SIFID 证明生成的图片与真实的人物图像越相近，在这一点上定量化的评价与直观的人眼定性观测评价相同，在保持原人物面部身份特征信息这一点上 MultipleComicGAN 的表现要好于 MTG 和 OSC，MultipleComicGAN 具有更强的保留真实人物图片的身份特征的能力。

表 4-3 SIFID 以及 SSIM 的平均值

方法	SIFID	SSIM
MultipleComicGAN	0.893617	0.538774
MindTheGap	1.00532	0.456997
OneShotCLIP	0.986671	0.511803

在使用 SSIM 的对保持风格的能力进行评价上，MultipleComicGAN 取得了 0.538774 的分数，略微大于 OneShotCLIP 的 0.511803 以及 MindTheGap 的 0.456997，这表明了 MultipleComicGAN 在保持原参考风格图像的像素结构，亮度以及对比度等几个因素具有一定的优势，这表明了在使用 SSIM 衡量的风格化程度时，MultipleComicGAN 要略微好于 MTG 以及 OSC 的保持参考风格图像的风格的能力。在 SSIM 的量化表现上 MultipleComicGAN 并没有显著优于 MTG 与 OSC，三种方法的保持参考图像风格的能力基本一致，在这一点与用户评价略有不同，对于三种方法哪种更能保持风格的问题，MultipleComicGAN 和 MindTheGap 的得票基本上是持平的，均大幅度领先于 OneShotCLIP，这一点可能与 OneShotCLIP 的官方实现选择的是 e4e 的 GAN 反转方式，导致更倾向于生成更轻度风格化的漫画化图像有关。更轻度风格化的图像对人眼评估来说也许是不能保持风格化，但是涉及到量化评估指标则是良好的保持了风格化，可以在 MultipleComicGAN 的后续训练调参中应用这一点。

4.4 训练时间与硬件消耗比较

表 4-4 对三种 GAN 漫画化方法的训练微调时间

方法	总计消耗时间（训练微调 50 种风格）
MultipleComicGAN	2:17:34
MindTheGap	8:42:45
OneShotCLIP	12:9:22

在本节中，比较了 MultipleComicGAN，MTG 以及 OSC 进行 50 种风格训练需要的总时间和如表4-4。可以看出，在同样的硬件配置下，训练时间上 MultipleComicGAN 具有 3.8 倍的优于 MTG 以及 5.3 倍的优于 OSC，在轻量化节约时间和算力这一点上 MultipleComicGAN 要显著好于 MTG 和 OSC。另一方面值得注意的是，虽然三种方法的原理相似，但是 MultipleComicGAN 和 OSC 的生成模型会更加轻量级，训练出的模型需要占用的储存空间均为 130MB，在这一方面的表现基本相同，要优于于模型大小为 300MB 的 MTG。

4.5 本章小结

本章对 MultipleComicGAN 的实验结果及其与现有方法的比较进行了详细介绍和分析。首先，对 MultipleComicGAN 的参数配置和具体实现细节进行了阐述，包括使用的预训练模型、数据集、以及特有的参数设定。接着，详细介绍了本次实验使用的评价指标，包括 SIFID 和 SSIM，这两种指标旨在定量评估漫画风格化图像的质量和风格保真度。

通过与 Mind The Gap（MTG）和 OneshotCLIP（OSC）以及传统 CNN 风格方法 NNST 的定量化和定性化比较，结果显示 MultipleComicGAN 在多个方面表现优越。在定性评估中，MultipleComicGAN 更一致地保留了输入图像的身份和表情，相比之下，MTG 和 OSC 在面部形状和表情的保留上存在扭曲。用户研究进一步验证了这一点，偏好使用 MultipleComicGAN 生成的图像的参与者数量占比要更高，认为它们

更好地反映了参考风格，同时保留了输入图像的特征。

在定量化评估中，使用 SIFID 和 SSIM 作为主要工具，MultipleComicGAN 展示了在风格保持和人物身份特征保留方面的优势。这些结果表明 MultipleComicGAN 在处理漫画风格化任务时，能够有效地平衡风格转换与原图保真度。

最后，本章还对训练时间和硬件消耗进行了比较，显示 MultipleComicGAN 在资源利用效率方面具有显著优势，特别是在训练时间和计算资源消耗上，比 MTG 和 OSC 更加经济高效。这一优势使 MultipleComicGAN 不仅在生成质量上卓越，也在实际应用中更具可行性和广泛的适用性。

本章全面地展示了 MultipleComicGAN 的高效性和较为优越的风格化效果，证明了它在当前漫画风格化技术中的领先地位，并且展望了其在艺术创作、社交媒体等领域的广泛应用潜力。

结　论

在本文中，进行了对基于真实人物多种风格漫画生成的方法的研究，研究并提出了 MultipleComicGAN，一种通过对 StyleGAN2 进行微调来实现多风格漫画化生成的 GAN 网络结构。MultipleComicGAN 不仅解决了当前基于真实人物多种风格漫画生成中存在的问题，也引入了一些新颖的机制，包括对风格化强度的控制，是否保留色彩等机制，还提供了一个用户友好的交互界面，便于用户自行对训练参数进行调整，本文主要的工作及成果如下：

1. 本文提出了一个新的 GAN 架构-MultipleComicGAN，其工作原理为通过对 StyleGAN2 进行微调，使其能够生成多种漫画风格化的图像。MultipleComicGAN 将一或多张漫画风格图片通过 GAN 反转的方法获得隐代码，输入 StyleGAN 的使用 FFHQ 真实人脸数据集训练出的预训练模型，风格化隐代码配合真实人物图片形成配对训练集进行多次迭代微调来形成映射，最终产生新的漫画风格化生成模型，达到将真实人物图片输入，输出获得多种漫画风格图像的目的。

2. 为了使 MultipleComicGAN 具有广泛适应性和实用性，对其引入了风格控制机制，这一机制包括控制风格化强度以及控制是否保留参考风格图像的色彩。为了解决漫画风格化过程中与参考风格图片过拟合的问题，MultipleComicGAN 引入了保持身份的余弦相似度控制和特征插值控制，实现了对风格化强度的定量控制；为了满足特定情况下保留原参考风格图像色彩的需求，MultipleComicGAN 设计了选择性遮罩和虚拟反演的方法，允许用户选择是否保留色彩进行风格化。此外还使用 Jupyter Notebook 开发了一个轻量化的简单用户友好型交互页面，支持用户自定义训练风格图片和参数，使用户可以自己微调出最符合用户审美需求的图片。

实验结果方面采用了用户评价、定性评判、定量评估三方面进行评估比较，实验结果表明，无论是从定性还是定量评估指标判断，MultipleComicGAN 都能够快速生成高质量高分辨率的漫画风格化图像，具有多风格，轻量化，不依赖艺术风格训练集的特点。MultipleComicGAN 支持以多张真实人物图片为基础，同时生成多张多风格漫画化图片。同时也实现了较好的轻量化处理，风格化和训练微调耗时较同类

型对 StyleGAN2 进行微调的方法更短。同时，MultipleComicGAN 在训练微调的过程中不依靠艺术风格训练集，只需要一或数张目标风格图片便可以生成该风格的漫画风格化图片。

本文提出的 MultipleComicGAN，在解决基于真实人物的多风格漫画生成方法方面取得了较为满意的结果，同时在轻量化方面也取得了较为优秀的性能，但本文的方法仍存在一些局限性，可以在未来的工作中进一步对其进行完善：

1. 在对调查问卷的评价分析过程中，发现不同的风格会对用户的偏好以及风格的保持评判产生较大影响，比如水彩和素描风格用户在偏好选择上取得了截然不同的结果。在未来的工作之中，可以进一步优化生成器的设计，尝试引入一个风格判别机制，根据不同的参考风格图像自动选取不同的参数调优，GAN 反转器，迭代方式等，进一步提高生成图像的质量和多样性。

2. 目前虽然已经使用 Jupyter Notebook 为 MultipleComicGAN 开发了一个轻量级的交互界面，实现了轻量化的调参训练等，但是仍然可以结合前端技术开发一个应用程序或是增强用户交互界面，为用户提供更多的自定义选项和更加直观的操作体验，进而探索 MultipleComicGAN 在更多应用场景中的潜力，如 AR 场景构建、游戏角色设计等。

总而言之，本文提出的方法在多方面优于当前的传统 GAN 算法，提供了一种有效的技术路径来探索图像风格化的多样性，也为未来的图像生成研究提供了新的可能性。

参考文献

[1] Benson P J, Perrett D I. Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images[J]. European Journal of Cognitive Psychology, 1991, 3(1): 105-135.

[2] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2414-2423.

[3] Elad M, Milanfar P. Style transfer via texture synthesis[J]. IEEE Transactions on Image Processing, 2017, 26(5): 2338-2351.

[4] Shi Y, Deb D, Jain A K. Warpgan: Automatic caricature generation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 10762-10771.

[5] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[6] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.

[7] Mirza M, Osindero S. Conditional generative adversarial nets[J]. arXiv preprint arXiv:1411.1784, 2014.

[8] Smith K E, Smith A O. Conditional GAN for timeseries generation[J]. arXiv preprint arXiv:2006.16477, 2020.

[9] 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.

[10] 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.

[11] Gao H, Pei J, Huang H. Progan: Network embedding via proximity generative adversarial network [C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 1308-1316.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] Borji A. Pros and cons of GAN evaluation measures: New developments[J]. Computer Vision and Image Understanding, 2022, 215: 103329.

[17] Srivastava A, Valkov L, Russell C, et al. Veegan: Reducing mode collapse in gans using implicit variational learning[J]. Advances in neural information processing systems, 2017, 30.

[18] Gooch B, Gooch A. Non-photorealistic rendering[M]. AK Peters/CRC Press, 2001.

[19] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
[20] Jing Y, Yang Y, Feng Z, et al. Neural style transfer: A review[J]. IEEE transactions on visualization and computer graphics, 2019, 26(11): 3365-3385.
[21] Chen Y, Lai Y K, Liu Y J. Cartoongan: Generative adversarial networks for photo cartoonization [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 9465-9474.
[22] Shu Y, Yi R, Xia M, et al. Gan-based multi-style photo cartoonization[J]. IEEE Transactions on Visualization and computer graphics, 2021, 28(10): 3376-3390.
[23] Jang W, Ju G, Jung Y, et al. StyleCariGAN: caricature generation via StyleGAN feature map modulation[J]. ACM Transactions on Graphics (TOG), 2021, 40(4): 1-16.
[24] Zhu P, Abdal R, Femiani J, et al. Mind the gap: Domain gap control for single shot domain adaptation for generative adversarial networks[J]. arXiv preprint arXiv:2110.08398, 2021.
[25] Kwon G, Ye J C. One-shot adaptation of gan in just one clip[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.
[26] Wright M, Ommer B. Artfid: Quantitative evaluation of neural style transfer[C]//DAGM German Conference on Pattern Recognition. 2022: 560-576.
[27] Heusel M, Ramsauer H, Unterthiner T, et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium[J]. Advances in neural information processing systems, 2017, 30.
[28] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training gans[J]. Advances in neural information processing systems, 2016, 29.
[29] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 586-595.
[30] Gu S, Bao J, Chen D, et al. Giqa: Generated image quality assessment[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16. 2020: 369-385.
[31] Pinkney J N, Adler D. Resolution dependent gan interpolation for controllable image synthesis between domains[J]. arXiv preprint arXiv:2010.05334, 2020.
[32] Li Y, Zhang R, Lu J, et al. Few-shot image generation with elastic weight consolidation[J]. arXiv preprint arXiv:2012.02780, 2020.
[33] Ojha U, Li Y, Lu J, et al. Few-shot image generation via cross-domain correspondence[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 10743-10752.
[34] Robb E, Chu W S, Kumar A, et al. Few-shot adaptation of generative adversarial networks[J]. arXiv preprint arXiv:2010.11943, 2020.
[35] Mo S, Cho M, Shin J. Freeze the discriminator: a simple baseline for fine-tuning gans[J]. arXiv preprint arXiv:2002.10964, 2020.
[36] Liu M, Li Q, Qin Z, et al. Blendgan: Implicitly gan blending for arbitrary stylized face generation [J]. Advances in Neural Information Processing Systems, 2021, 34: 29710-29722.
[37] Gal R, Patashnik O, Maron H, et al. Stylegan-nada: Clip-guided domain adaptation of image generators[J]. ACM Transactions on Graphics (TOG), 2022, 41(4): 1-13.
[38] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language

supervision[C]//International conference on machine learning. 2021: 8748-8763.
[39] Zhu P, Abdal R, Qin Y, et al. Improved stylegan embedding: Where are the good latents?[J]. arXiv preprint arXiv:2012.09036, 2020.
[40] Yang C, Shen Y, Zhang Z, et al. One-shot generative domain adaptation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 7733-7742.
[41] Wang Y, Yi R, Tai Y, et al. Ctlgan: Few-shot artistic portraits generation with contrastive transfer learning[J]. arXiv preprint arXiv:2203.08612, 2022.
[42] Ramesh A, Pavlov M, Goh G, et al. Zero-shot text-to-image generation[C]//International conference on machine learning. 2021: 8821-8831.
[43] 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.
[44] 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.
[45] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
[46] Alaluf Y, Patashnik O, Cohen-Or D. Restyle: A residual-based stylegan encoder via iterative refinement[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 6711-6720.
[47] Yang T, Ren P, Xie X, et al. Gan prior embedded network for blind face restoration in the wild [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 672-681.
[48] Deng J, Guo J, Xue N, et al. Arcface: Additive angular margin loss for deep face recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4690-4699.
[49] Chong M J, Lee H Y, Forsyth D. Stylegan of all trades: Image manipulation with only pretrained stylegan[J]. arXiv preprint arXiv:2111.01619, 2021.
[50] Shaham T R, Dekel T, Michaeli T. Singan: Learning a generative model from a single natural image [C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 4570-4580.
[51] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.

致　谢

时光飞逝之间，本科四年的生活就这样结束了。这四年的学习生活中虽然遇到了各种波折，但在这四年的学习生活中，我收获颇丰，在这里感恩母校，感恩本科生活中遇到的每一个人。

首先向我的导师王崇文老师表达感谢与敬意，王老师在科研领域当中有着充足的理论知识和丰富的科研经历，在他的指导下，我在论文完成的过程中产生的诸多问题有了解答，以及对科研过程有了更深层次的理解。王老师让我完成论文的过程遇到的问题得到了解答，同时为我提供了无数的宝贵意见，在他的指导下我受益匪浅。每次汇报的时候老师都会认真倾听我汇报的内容，指出我当前存在的不足疏忽之处，并为将来的工作指明方向。也从老师对其他同学的汇报的评价分析中，学到了老师对于科研严谨的态度。在此由衷的对王崇文老师表示感谢。

其次，感谢同门师姐王艺瑾，虽然相处时间只有短短的几个月，艺瑾师姐从开题阶段便为还对工作毫无头绪的我一步步缕清思路，虽然同样是在为毕设工作而忙碌，艺瑾师姐还是抽出时间帮助我，教会了我寻找论文的方法，论文的规范化写法。在完成本次工作的过程中与艺瑾师姐的讨论以及她的建议和帮助是我完成本次工作不可或缺的。

然后，感谢我的室友崧兑旭，本文中所提到的真实人物照片基本都来自于我的室友崧兑旭的自拍照片，他对本次工作的贡献是巨大的。

还要感谢我在本科学习生活中遇到的每一位老师，让我在计算机专业领域学会了无数的基础专业知识。授课风格各异的老师还让我在学到专业知识之外，学到了无数做人的道理，也养成了终身学习的习惯。

最后，感谢我的家人们在我本科的学习中给予我的理解和物质条件支持，宽裕的物质生活让我得以安心完成学业。

四年的学习生活就这样的结束了，这篇文章既是我作为学生学习生活的终点，也是我作为研究者科研生活的起点，虽然学生时代就此结束，但是我在完成本次工作中学到的东西将会伴随我接下来的整个科研生活乃至终身。