课程论文

对 StyleCLIP 的人脸图像编辑的优化研究

姓　　名：杨馥冰

学　　号：19307130304

课 程 名 称：数字图象处理

课 程 序 号：COMP130032.01

专 业 名 称：计算机科学与技术

二〇二二年　五月

对 StyleCLIP 的人脸图像编辑的优化研究

摘要

人脸编辑在视觉效果和电子商务领域的应用非常广泛，但是目前能直接通过自然语言进行人脸图像编辑的研究依然很少。StyleCLIP 通过对图像生成模型 StyleGAN 和多模态图文模型 CLIP 的结合，可以以任意自然语言描述对应的图像文本编辑操作，有着无穷的潜力。但在实践中,经过模型编辑后的图像质量并不十分稳定,需要花费一定的时间用于参数调整。我们通过实验探索分析了几种超参数对图像编辑效果的影响，并通过 Mapper 提升了图像编辑的速度。可视化结果证明，我们对参数分析的结论有助于提升生成更高质量的图像。

1. 引言

近年来，深度学习快速发展，尤其是生成网络的出现，为人脸编辑任务提供了全新的思路。采用深度生成模型的人脸编辑技术不仅速度快，泛化能力也更强。但是许多深度生成模型都仅在图像上进行操作，或者编辑操作受限，或者需要通过其他示例图片指定编辑操作，或者需要对隐状态进行调参，因此在使用上存在一定的困难。受到 StyleGAN[1]的启发，StyleCLIP[2]的工作集中在利用 StyleGAN 的隐空间来生成高度逼真的图像，并通过引入 CLIP[3]模型强大的语义控制能力，来开发一个基于自然语言文本的图像编辑器，由于引入了自然语言作为人机交互的接口，打通了图文之间信息传递的隔阂，可以直接进行零样本学习的图像编辑，因此拥有广泛的应用前景。

本文探究并实践了 StyleCLIP 模型在人脸图像编辑任务中的应用。StyleCLIP 的一大优势在于其无需训练即可快速部署,因此可以快速的验证在人脸图像编辑任务上的性能。此外，StyleCLIP 依然存在一些权重参数需要进行调整，引入了参数调整的工作量，因此我们也探索了一些模型超参数是如何在对图像编辑结果产生影响,且应该如何在调参的过程中合理的设置超参数。最后，StyleCLIP 的迭代优化模式处理一张图需要在 GPU 环境下耗费约 1 分钟的时间，我们也尝试通过 Mapper 来加速图像编辑速度。本文后续部分组织如下：

a) 第二章介绍了相关工作，包括近几年内流程的人脸图像生成模型与编辑模型，以及多模态图文模型 CLIP。

b) 第三章介绍了 StyleCLIP 模型，分析了模型结构和损失函数组成。

c) 第四章介绍了我们所进行的实验，其中包括对公开代码库的直接复现，改造代码库

后对几种不同图像编辑操作的尝试，对超参数的调整与可视化，以及 Mapper 优化。

d) 第五章进行总结并阐述了这次研究的贡献点。

2. 相关工作

a) 人脸图像编辑模型

人脸图像编辑，其目标在于修改感兴趣的人脸属性，例如头发颜色，表情，年龄等，属于图像迁移任务。其应用非常广泛，包括电影特效，视频特效，人物美颜，等等。人脸编辑的过程，需要合理地修改感兴趣的属性，保持与修改属性无关的其他区域不变的同时，高质量、逼真地完成对图像的修改。因此，任务的难点在于，简单粗暴的向图上堆砌新属性，容易使得图像失真，也容易改变与属性无关的图像细节，如人脸上的身份特征，图像背景，光照，等等。

对人脸图像编辑的任务的处理框架，通常分为三个阶段。首先，通过构建编码器并反演得到能表征原始人脸图像的隐变量，随后，根据需要的人脸编辑任务来编辑隐变量，使其包含编辑任务所要求的元素的同时，尽可能保留原有图片的信息，最后，通过编辑后的隐变量生成编辑后的人脸图像，通常会使用在 2019 年被提出的 StyleGAN，它有着优秀的人脸图像融合、迁移与生成能力，在许多任务中都能得到应用，并在后续不断的得到更新。

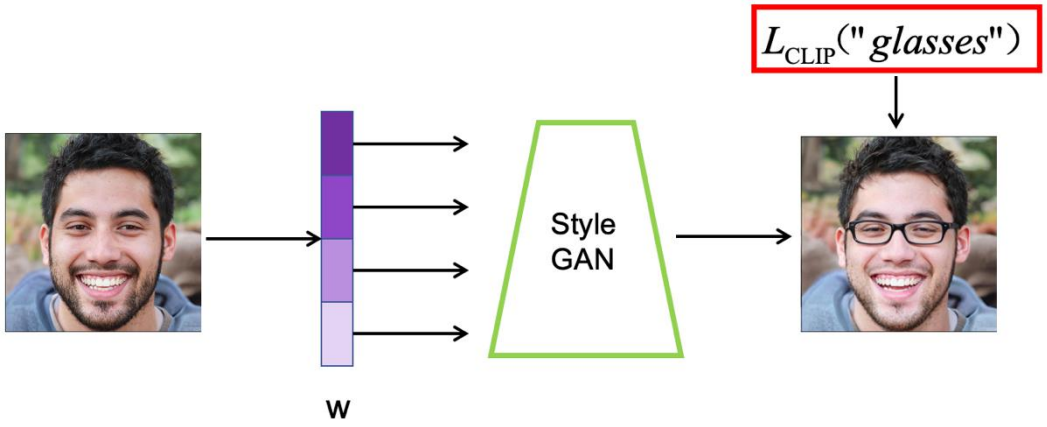
传统的人脸编辑模型要求图片的隐状态表示是属性独立的,根据编辑操作加入对应属性的向量，并将融合后的隐状态进行解码。由于人脸包含的许多属性之间都具有关联性，因此传统的编辑方式通常会导致信息丢失,产生图像畸变,并明显限制了隐状态的表示能力。2019 年的 AttGAN[4]不再要求隐状态是属性独立的，而是在后续的重建解码部分添加约束，通过分类模型监督产生想要的图片，保证了生成图片包含编辑操作的属性，从而可以按强度控制属性的变化程度，且不改变属性无关区域。2020 年的 InterFaceGAN[5]则更直接的将隐状态想象成欧式空间，将隐状态和语义特征通过线性变换联系起来，可以直接使用训练完成的 GAN 模型进行图像编辑。2018 年的 ELEGANT[6]模型，通过交换两张图片的隐空间特征，来快速完成对多属性的图像编辑任务。但这些研究都集中在对单模态的图像样本进行编辑上，具体的编辑操作对应的向量需要通过隐状态的方式给出，使用者的交互体验很差。

b) CLIP 模型

在 2021 年，OpenAI 通过进行大规模的图文多模态对比学习预训练，提出了 CLIP 模型。CLIP 由一个图像编码器和文本编码器组成，输入的多模态样本通过编码器，得到图像表征和文本表征，通过对比学习拉进同一样本的图文表征。由于训练样本中的文本模态信息均由

互联网上的自然语言构成，因此，利用预训练的 CLIP 模型，可以以零样本学习的形式计算图片与文本的相似度，从而打通图文模态之间信息传递的桥梁。

3. StyleCLIP 模型介绍



$$\arg\min_{w \in \mathcal{W}+} D_{\text{CLIP}}(G(w), t) + \lambda_{\text{L2}} \|w - w_s\|_2 + \lambda_{\text{ID}} \mathcal{L}_{\text{ID}}(w),$$

图 1 StyleCLIP 模型结构

StyleCLIP 的想法即是源于 Clip 和 StyleGAN 的混合。StyleGan 有着强大的从隐状态生成人脸图片的能力，但只能对图片进行融合，或在图像之间进行风格迁移，人类使用者与模型之间缺少直接的交互能力。CLIP 可以进行零样本学习，在图文之间的信息交流上有着得天独厚的优势，可以搭建起图文之间的桥梁。StyleClip 的基本设计，就是通过对图像隐状态加一个较小的扰动，使扰动后生成的图像，与使用者提供的文本描述，在 CLIP 模型上的相似度尽量大。

如图 1 所示，初始的图像送入 StyleGAN 的编码器得到隐状态 Ws，我们需要将其修改为隐状态 w，并通过 StyleGAN 的生成器得到修改后的图像，通过 CLIP 模型计算图像嵌入和文本嵌入的相似度，并通过梯度回传优化隐状态 w。整个过程不需要涉及对网络的训练，全部使用公开的 StyleGAN 和 CLIP 的预训练参数。因此架设非常容易，不需要训练成本。

模型的损失函数如图 1 下方所示，分为三个部分，第一歌部分是 CLIP 计算的相似度，第二个部分是 L2 正则项损失，保证新的隐状态在旧的隐状态附近，使得图片不会有较大的修改。第三个部分是 ArcFace Network[7]，可以提取图像中的人脸特征，计算与原始图像人脸特征的差异。三者通过加权求和优化待求的隐状态 w，即是 StyleCLIP 的做法。由于图像从隐状态直接生成得到，不会引入生硬的 PS 痕迹，从而可以尽量避免模型失真。

4. 实验

a) 实验配置与公开代码复现

由于 StyleCLIP 的作者已经提供了公开的代码库，我们选择以此为基础进行实验。代码库提供了 Optimization 和 Mapper 方法的入口。由于 Optimizaiton 方法模型结构更简单，编辑速度更快捷，因此是后续实验的主要方法。

代码库中自带的 Optimization 流水线并不完整。由于模型的训练流程中仅需要对图片的隐状态进行优化，因此代码库可以选择读取事先存储的图像隐状态，或者使用一个随机生成的图像隐状态，而没有提供图像编码器接口用于人脸图片的编码。我通过调用随机生成图像隐状态的接口，设置了几种不同的文本描述进行人脸编辑，下载了公开的模型预训练权重，使用了代码库的默认参数，即 L2Loss 权重取 0.008，IDLoss 权重取 0（即不使用），学习率取 0.1 优化 300 个步骤，并不加人工选择的以训练过程中最后一张图片（通常也对应了稳定后的最低训练损失）作为结果，以避免引入过多的人工干预。生成的六张图片示例如图 2 所示，其中每张子图的子标题对应了修改的文本描述，左侧为随机生成的原始图片，右侧为对应的修改后图片。

Glasses　　Female face　　Female face

Sad face　　Donald Trump　　A face with makeup

图 2 StyleCLIP 随机人脸编辑图像示例

从图 2 可以观察到，模型按照给定的文本描述对这些图像进行了正确的编辑。模型可以对于一些简单的需求，如第一张图对应的加上眼镜“Glasses”，模型可以很自然的给人物添加上眼镜。对于复杂的需求，模型也能理解其对应的含义，如第五张图需要将人物变得像特朗普，模型对应的抓住了特朗普的胖脸，金发，眯眯眼等特征，，最后一张图对应的描述是对人物上妆“makeup”，模型则相应的进行美白，打眼影，说明其能理解“上妆”所对应的具体含义。

虽然如此，也可以注意到有些图片的质量不如人意。例如第二张图片将小女孩修改为成年女性后，虽然可以预料到五官会发生较大的变动，但对皮肤的上色也变得非常不均。第三张图片将男性图片修改为女性后，还顺带修改了肤色，发色，衣服颜色，人物背景，且生成后的人物肤色不均，质感很差。第四张图成功调整了女性表情的同时，也大幅度改变了肤色和发色，这些都不是期望的目标。因此还需要对模型的参数进行更深的分析。

b） 代码改造与简单编辑操作

由于 StyleCLIP 的官方公开代码库中不能指定图片进行定量对比实验，因此需要对代码库进行改造，将 StyleGan 的编码器部分接入流水线，将指定的图像转换成隐状态并存储后，再通过 StyleCLIP 加载存储的隐状态数据并进行转换。经过这样的改造，模型已经可以对指定的图片和文本生成对应的编辑结果。我们先在五张人物图片上验证了三种修改较小的操作：开心，伤心，添加刘海。实验结果如图 3 所示。

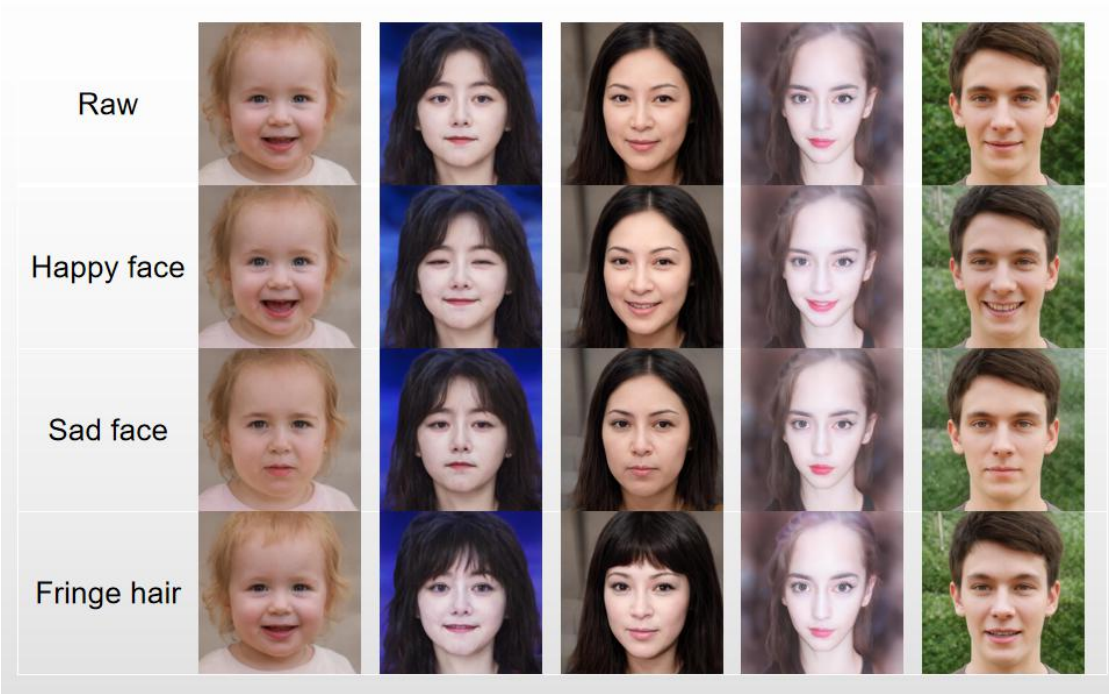


图 3 StyleCLIP 对三种简单操作的图像编辑结果

从实验结果中可以观察到，StyleCLIP 在开心和伤心两种图像编辑操作上效果都很好，可以从对应的两行编辑图像中明显看到图中人物的情绪变化，也符合对应指定的文字描述。最后一个添加刘海的操作，在不同图片上效果有所不同，前三张图都在额头添加了明显的头发，符合我们对刘海的定义，倒数第二张图则只在额头附近增加了少许头发，并没有形成我们认知中的刘海，最后一张图一开始额头就有较厚的头发，模型也仅仅是将这部分头发变得更厚。我们也尝试过对后两张图进行一些参数搜索，但并没有得到什么收益。可以猜测，CLIP

模型对刘海概念的理解，仅仅是额头附近有一片较多的头发，而对刘海的具体形状则学的并不是很好，因此没有引导 StyleCLIP 得出正确的图像编辑结果。

c） 复杂编辑操作

在前一部分实验结束后，我们也探索了一些更多的图像编辑操作，但在进行瘦脸编辑时，使用代码库默认的参数在所有图像上几乎都看不到图像变化，编辑效果不好。可以猜想是由于 L2Loss 控制了模型能改变的隐状态信息量，在限制优化范围后模型找不到一个较好的解。通过减少 L2Loss 的权重，可以完成更多的图像编辑操作。因此这一节的实验包含了一些较难的操作，包括前文提到的瘦脸，以及“艾莎公主”，老年女性共三种操作。实验结果如图 4 所示。其中第 2 行即对应了前文所提到的瘦脸操作，L2Loss 的权重从 0.008 被降低到 0.006。通过对比脸型和宽度，我们可以确认到瘦脸成功的在几张图片上都发挥了作用。

图 4 StyleCLIP 对三种复杂操作的图像编辑结果

我们也可以尝试一些更复杂的操作，比如“艾莎公主”编辑操作，其难点在于，直观的文字描述中没有包含任何和具体操作相关的信息，且“艾莎公主”从语义上又包含了多种不同的修改角度，比如西方成年女性，蓝眼金发，水蓝色服装等等具体特征，因此需要模型对这些语义有所理解。在将正则项权重放低到 0.004 后的图片编辑结果如第 3 行所示，五张图片都或多或少出现了“艾莎公主”的元素。其中第 2 张与第 3 张图片中对应的黑色头发女性，可能是由于修改发色会涉及的隐状态变化太大，因此保持了原始的发色，但能明显感受到五官向着西方女性的角度靠近，且肤色变得更白，符合“艾莎公主”编辑操作。其中效

果最好的是第 4 张图片，不仅是头发完全变成了金色，眼睛对应上了妆，衣物也被识别并修改成了蓝色。

最后的操作需要将人物变为老年女性。虽然这个操作在各种应用中都显得有些司空见惯了，但由于年轻女性和老年女性在面部特征上相差较大，因此需要对图像进行的编辑其实很多，诸如皱纹法令纹，肤色肤质发质，以及面部肌肉的松弛，因此权重实际上需要被放的更宽。在将权重降低到 0.002 之后，也能正确的得到五张编辑后的图片。如第 4 行所示，五张图片的人物在正确的变成老年女性的同时，也保持了各自的面部特征。

d) 难操作“深色皮肤”

在这一部分，我们展示了一个在实验中遇到的难操作，即变为“深色皮肤”。图 5 展示了调整 L2 正则化损失权重的过程中对应的图像变化。

图 5 StyleCLIP 对“深色皮肤”操作的图像编辑结果

首先可以明显观察到，损失权重越小，编辑后图片和原图差别越大，和文本描述越相近。这符合 StyleCLIP 模型训练损失的定义，即正则项会限制模型从文本中获取的信息量。

除此之外，图中能很明显的体现从预训练模型中产生的数据偏置。例如，当权重被降低到 0.004 时，第 1 张图的婴儿面部特征发生了明显的变化，而将权重降低到 0.002 时，甚至出现了耳钉和唇钉，这些既不包含在原始图像中，也与文本内容无关的内容。第 3 张图则直接呈现了更多非洲人的特点，整个面部特征都产生了明显变化。可以想到这是由于 CLIP 模型在训练过程中见过的大部分深色皮肤的人的照片，或多或少的都涉及了非洲血统，因此

模型认为“深色皮肤”包含了一些非洲民族的常见饰物，从而在图像编辑时引入了这些变化。很明显的，我们并不希望这些额外的情况。

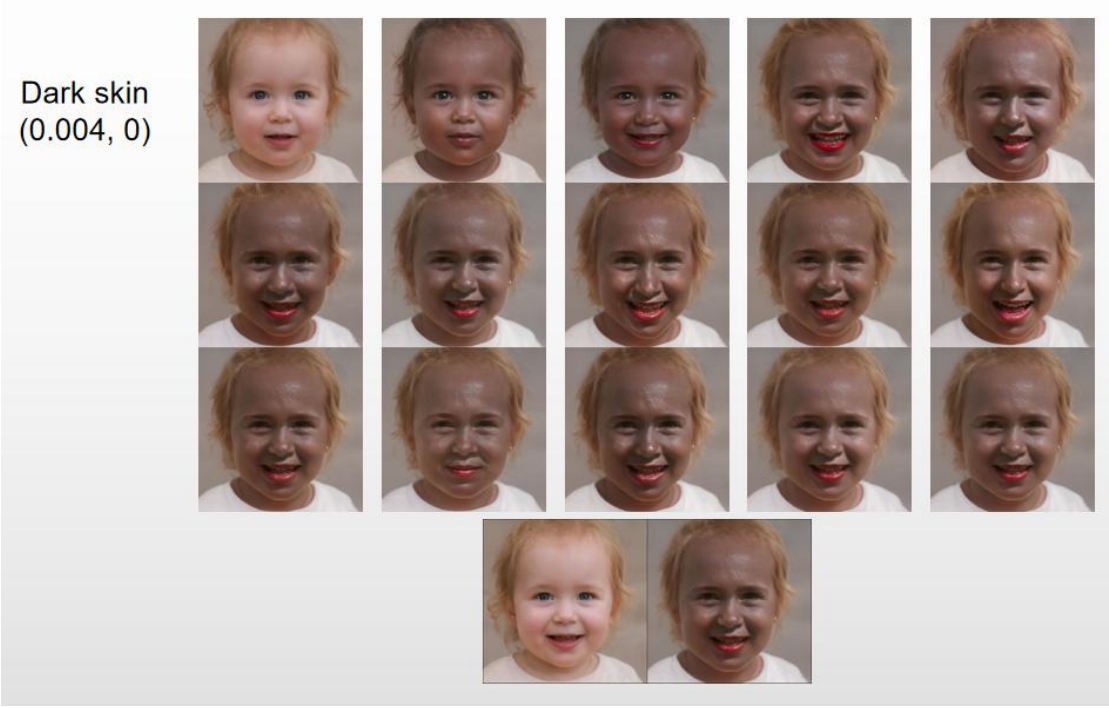


图 6 StyleCLIP 对婴儿图像进行“深色皮肤”编辑的过程

图 6 可以更全面的展示在固定 L2 正则损失为 0.004 时，StyleCLIP 对婴儿图像进行“深色皮肤”编辑的具体优化过程，以从上到下，从左到右的顺序，以 20 个训练步骤为间隔显示一张图，共 300 个训练步骤。可以观察到，模型最初生成一张图就是期望的“深色皮肤”图片，但在训练过程中，小孩的面部表情逐渐变形扭曲，差异变大。可以想到，StyleCLIP 模型寻找的是在 L2 正则项损失的限制下，模型认为最符合“深色皮肤”描述的图片，又由于训练集中对应满足条件的图片大多为成年黑人，从而将小孩的整体面部特征向相应方向进行引导。我们希望增加一个损失，能够尽量控制人物面部特征不变，也就是 ArcFace Network 所要做的事情。

e) 在 ArcFace Network 限制下的难操作“深色皮肤”

我们在这一版实验中引入 ArcFace Network 的 IDLoss 损失，从 L2 正则项和 IDLoss 两个损失的权重同时进行网格搜索。图像编辑的生成结果如图 7 所示。

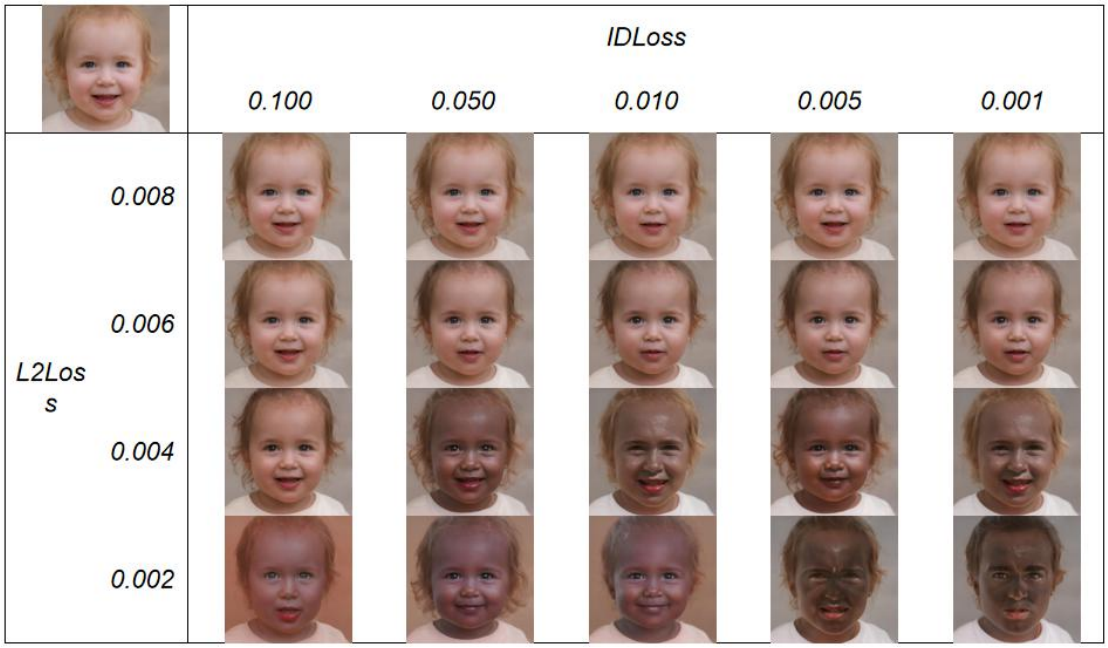


图 7 StyleCLIP 对婴儿图像进行“深色皮肤”的参数网格搜索结果

在 L2 正则项权重较小时，很明显的可以观察到，随着 IDLoss 的权重增加，生成的图像在人物特征上逐渐变得更接近于原始的图像。尤其是在 L2 正则项权重取 0.002，IDLoss 选择 0.001 或 0.005 时，生成的图片皮肤呈现煤黑色，面部表情扭曲，看不出与原始图像的联系。随着 IDLoss 的逐渐增加，生成的人物图像就在保持深色皮肤的同时，逐渐变得更接近原始的图片。

另外，IDLoss 也会一定程度上限制图片的变化，在这个角度上它与 L2 正则项起的作用类似。如 L2Loss 取 0.004，IDLoss 取 0.1 的图像，仅稍稍变深了一些肤色，而在 IDLoss 降低到 0.05 时，就会明显使肤色变深，因此细致调参时也要综合考虑两个权重的影响。

f) “深色皮肤” 婴儿照片结果展示

我们最终选择的 L2Loss 权重为 0.004，IDLoss 权重为 0.05，并通过连续图片展示了随着模型的学习不断更新图片隐状态，导致婴儿肤色变化的过程。为了更好的展现变化过程，让模型变化缓慢，我们将学习率降低到原本的 1/10，也就是 0.01。我们以 10 张图片为间隔进行采样生成连续图片，如图 8 所示。可以观察到婴儿的肤色逐渐变得越来越黑。此外，在前几帧婴儿肤色变深的同时，发型变得稀疏简单，这是由于隐状态未收敛，在向“深色皮肤”方向靠近的同时，忽略了图像细节导致的，随着后续不断的优化，婴儿肤色继续变深的同时，发型也逐渐恢复到和原始图像相似，这是由于 L2 正则项损失限制了图像在接近原图的范围内，因此为了更加靠近“深色皮肤”的描述，就会尽量只修改与皮肤有关的隐状态，而保持其他不变，这也符合图像编辑的目的。我们也以视频的形式展示了编辑过程，300 个

训练步骤生成的对应图像以每秒 10 帧的速度播放，共 30 秒，放在论文附件中。

图 8 StyleCLIP 在（0.004，0.0）权重下对婴儿图像进行“深色皮肤”的编辑结果

g） “幼稚”与“衰老”

这一部分通过连续图片的方式,具体展示了将我的真实证件照逐渐变得幼稚和衰老的过程，对应使用的文字描述分别是“baby”和“old female”，结果如图 9 和图 10 所示。由于希望在这一阶段展示人物年龄逐渐变化的过程，因此没有使用固定的 L2 正则项权重，而是使其随训练过程而逐渐变小，这样就可以让视频在初期更接近原始图像，在后期更接近文字描述的一个平滑渐变过程。

在“baby”侧的实验中，由于婴儿脸型和成人脸型相差较大，因此没有设置 IDLoss，使用的 L2 正则项权重从 0.004 线性衰减到 0.001。为了避免在后期无法收敛，学习率使用固定的 0.01，而非默认的余弦衰减。从图中可以观察到变化过程比较平滑，人物特征一直保持不变，但编辑前后能明显看到人物的五官变得年轻。

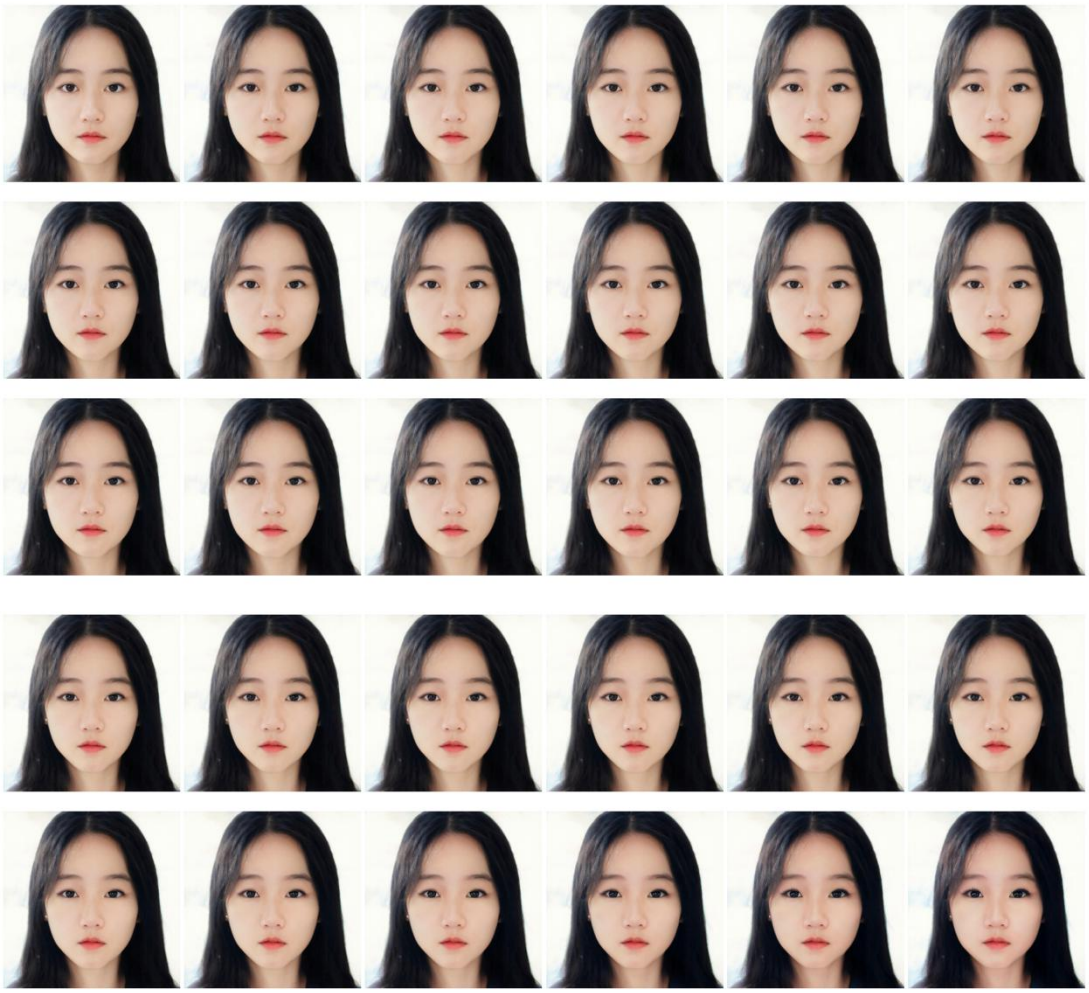


图 9 StyleCLIP 对证件照的“幼稚”编辑操作

而在“Old female”一侧的实验中， IDLoss 被设置为 0.05，用于避免人脸老化后严重变形失去辨识度，且使用的 L2 正则项权重从 0.006 线性衰减到 0.001。如图所示，人物从五官开始逐渐老化，到达一定程度后不再明显改变五官，而是开始调整发色使其逐渐变白，同时慢慢加深法令纹，再逐渐扩展到五官整体的衰老，符合我们对衰老过程的认知，同样的体现了一个渐变的过程，且没有出现 “头发乌黑发亮却五官衰老面布皱纹”的异常。由于 StyleGan 和 Clip 在训练时引入了大量西方人脸的数据,导致生成的人脸普遍具有少量西方人的特征，形成明显的偏置，这在应用上是难以避免的。

图 10 StyleCLIP 对证件照的“老年女性”编辑操作

h) Mapper 优化

在之前的实验中，StyleCLIP 使用 StyleGAN 和 CLIP 的预训练模型直接进行优化，整个过程中不需要经过模型训练阶段，而是直接迭代更新图像的隐状态。这个流程使得 StyleCLIP 架设方便，易于实现，应用简单，但也导致每次编辑图像都需要重新进行迭代，会耗费较多的 GPU 用时。在实际操作时每生成一张图片就需要花费约 1 分钟的时间，如果需要编辑大批量的图片，不仅会需要大量的 GPU 计算资源，计算延时上也会有很大的问题，因此在实际的应用中表现有限。

论文中包含了对 StyleCLIP 的两种优化，分别称为 Mapper 和 Global Direction，因为后者较为复杂，且耗时较多，我们选择了前者进行实现。如图 10 所示，Mapper 优化即是通过神经网络学习对应给定文本描述的映射,这个映射可以直接得到输入图像隐状态编辑后的编辑图像隐状态，而不用再经过反复梯度回传来迭代更新隐状态，从而大大提升对图像的编辑速度。经过 Mapper 优化后，模型编辑图像的时间可以降到几十毫秒的级别，作为代价，模型将会失去输入图像编辑操作文本描述的能力，一个 Mapper 只能对应一个固定图像编辑操作，且需要几小时的训练时间，以及同样需要一定的调参工作。

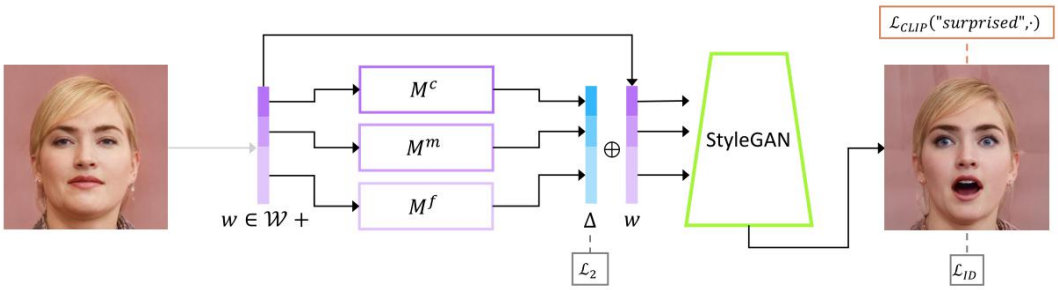


图 11 StyleCLIP Mapper 优化

我在实验中尝试了 Mapper 优化，指定“sad face”作为图像编辑操作，并使用了默认参数进行训练，即训练集包含 24176 个样本，测试集包含 2824 个样本，以批次大小为 2 训练 50000 个步长。训练完成后对 2824 个样本进行编辑，花费 7 分 49 秒，平均每个样本花费 166ms，相对于原本近一分钟的迭代时间，编辑效率提升了好几百倍。我也对测试集中的数据进行了可视化展示，抽样 20 条，如图 11 为“sad face”的对应编辑结果，

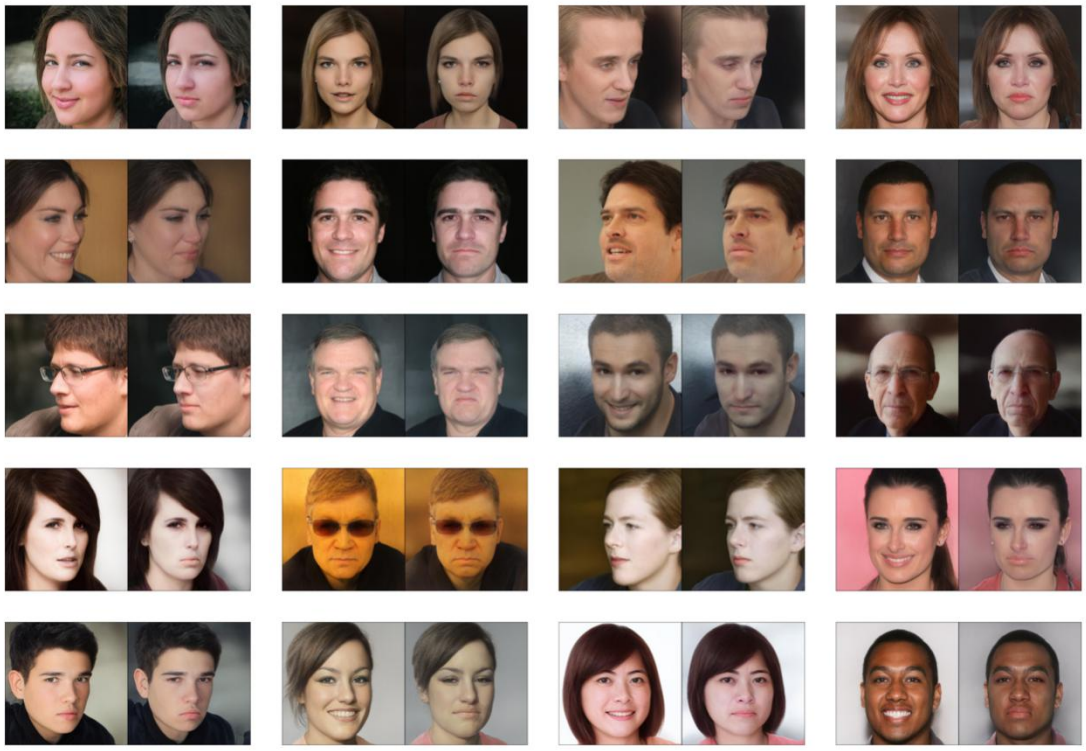


图 11 StyleCLIP Mapper “sad face”编辑结果采样

5. 结论

在本次研究中，我们复现了近年一个具有新意的研究 StyleCLIP，其可以通过文字描述进行图像编辑，跨越了图文信息传递的鸿沟。在直接应用代码库生成图像编辑示例之外，我们也改造了流水线用于定量研究。我们在多种不同的图像编辑操作上测试了 L2 正则项损失

权重和 IDLoss 损失权重的影响，其中 L2 正则项直接约束了模型对图片的修改程度，IDLoss 约束了模型对人脸特征的修改程度，在可能导致任务面部特征大幅度变化的操作时可以关闭 IDLoss，否则需要设置权重。在实验中观察到了预训练数据分布的偏置，如模型会倾向于产生偏西方的人脸，等等。我们还通过 Mapper 优化了模型进行编辑的速度，一张图的处理时间从一分钟降低到几百毫秒。最终，我们通过总结的调参经验，在多种图像编辑操作上得到了对应的优质生成图片，并以连续图片和视频的方式展示了对真实图像的图像编辑变化过程。

参考文献

[1] 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.

[2] Patashnik O, Wu Z, Shechtman E, et al. Styleclip: Text-driven manipulation of stylegan imagery[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 2085-2094.

[3] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//International Conference on Machine Learning. PMLR, 2021: 8748-8763.

[4] He Z, Zuo W, Kan M, et al. Attgan: Facial attribute editing by only changing what you want[J]. IEEE transactions on image processing, 2019, 28(11): 5464-5478.

[5] 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.

[6] 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.

[7] 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.

代码运行方式：

模型代码库基于 https://github.com/orpatashnik/StyleCLIP 。

添加的图像编码器通过 python test.py 启动，并会将读入的图像转换成对应隐状态。输入输出均硬编码，根据需求使用。

模型运行时需要加载预训练参数，均可以从网络上下载同名文件得到。将文件放在 args 默认参数的对应位置。因为参数过大，因此没有打包到代码文件夹中。

模型训练脚本如下，分别对应生成幼稚化的图像和老年化的图像。

CUDA_VISIBLE_DEVICES=2 python optimization/run_optimization.py --description "baby" --l2_lambda 0.001 --id_lambda 0.000 --results_dir results-baby-dynamic-300 --latent_path latent-0.torchsave --lr 0.01 --step 300 --save_intermediate_image_every 1 &

CUDA_VISIBLE_DEVICES=3 python optimization/run_optimization.py --description "old\ female" --l2_lambda 0.001 --id_lambda 0.050 --results_dir results-old-female-dynamic-300 --latent_path latent-0.torchsave --lr 0.01 --step 300 --save_intermediate_image_every 1 &

Mapper 训练

python -m mapper.scripts.train --exp_dir mapper/exp/ --description "sad face"

Mapper 测试

python -m mapper.scripts.inference --exp_dir mapper/inference/ --checkpoint_path mapper/exp/checkpoints/best_model.pt --latents_test_path test_faces.pt --couple_outputs

视频

三个视频：

rebuild-dark-skin.mp4 对应将婴儿照片变成深色皮肤；

rebuild-300-young.mp4 对应将证件照变幼稚；

rebuild-300-old-female.mp4 对应将证件照变成老年女性。