

---

---

# 基于 CR-GAN、RaR 与 LAT 的侧脸年龄编辑

余今 (211240074、211240074@smail.nju.edu.cn)

(南京大学 匡亚明学院, 南京 210023)

**摘要:** 本论文旨在尝试基于 CR-GAN、RaR 与 LAT 的侧脸年龄编辑方法。针对人脸属性编辑技术在模拟失踪案件侦破等领域的实际应用需求, 本研究以年龄属性编辑为主题, 旨在保持人物身份信息的同时改变年龄。通过对正脸年龄编辑模型的基础上, 提出了两种实现思路: LAT\_CR-GAN 与 RaR\_LAT\_RaR。LAT\_CR-GAN 模型主要通过 CR-GAN 网络实现正脸到侧脸的视角转换, 同时进行年龄编辑, 而 RaR\_LAT\_RaR 则通过 Rotate-and-Render 模块将侧脸转换为正脸, 再进行年龄编辑和视角转换。实验结果表明, 虽然存在数据集与硬件限制, 但两种思路均取得一定成效, 但仍存在较大改进空间。

**关键词:** 人脸年龄编辑; 人脸视角编辑; CR-GAN; Rotate-and-Render; Lifespan age transformation

## 1 引言

### 1.1 研究背景和意义

相信大家在网上冲浪时, 或多或少会见过这种小程序, 只要输入一张照片就可以看到自己老去时的模样。这种任务被称为人脸属性编辑 (Facial Editing), 在这里主要是对年龄的编辑, 通过将年龄和身份特征解耦, 达到保持人物的身份信息而改变年龄的目的。最初, 人脸属性编辑技术多用于相机滤镜等娱乐效果, 但随着编辑技术的准确化与精细化, 使得它可能被用于模拟失踪和被拐卖人口的外观, 具有协助失踪案件侦破的现实意义。因此, 该研究兼具学术与现实意义。

### 1.2 目标和研究方法

本项目的目标是在借助来自斯坦福大学、华盛顿大学与 Adobe 研究所的正脸年龄编辑模型 Lifespan age transformation<sup>[1]</sup>的基础上, 建立能生成不同年龄的 90 度侧脸图片的模型。由于大作业要求并未明确规定, 考虑到输入图片可能为正脸与侧脸两种情况以及训练数据与硬件设备的限制, 笔者分别参考了 CR-GAN 网络<sup>[2]</sup>与 Rotate-and-Render 模块<sup>[3]</sup>, 设计了两种实现思路并作出对比。

考虑到公开的带有年龄标签的脸部数据特别是侧脸数据均不足, 人工标注成本较大, 直接训练侧脸年龄编辑模型较为困难, 笔者决定利用现有的年龄编辑模型与多视角的人脸数据集, 将设计重点放在脸部视角转换上, 两种思路应运而生。第一种思路为 LAT\_CR-GAN: 在输入图片为正脸时, 将 Lifespan age transformation (LAT) 模型与 CR-GAN 网络拼接, 前者承担年龄编辑任务, 后者承担脸部视角转换任务 (即将不同年龄的正脸转换为侧脸)。第二种思路为 RaR\_LAT\_RaR: 在输入图片为侧脸时, 由 Rotate-and-Render (RaR) 模块将侧脸转换为正脸, 用 LAT 模型进行年龄编辑, 再将 LAT 输出的正脸输入 RaR, 转换为 90 度侧脸作为输出。

## 2 相关研究

### 2.1 传统GAN与StyleGAN<sup>[10]</sup>

GAN (Generative adversarial network, GNN) 是深度生成型网络的一种, 旨在训练生成器网络  $G$  对数据映射  $G: Z \rightarrow X$  进行估计, 其中  $Z$  代表先验分布为  $p_z$  的输入隐空间,  $X$  代表分布为  $p_x$  的目标数据空间 ( $p_x$  未知)。为了提升生成器输出图像的真实性, GAN 引入了判别器网络  $D$  对生成器输出的分布  $p_g$  和  $p_x$  进行区分。  $G$  和  $D$  的参数可以在对抗训练过程中通过交替优化得到, 其目标函数为:

$$\min_G \max_D L_{adv} = \mathbb{E}_{\mathbf{x} \sim p_x(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

当训练过程理想地收敛时,  $p_g$  与  $p_x$  将无法被区分, 也即模型可以生成以假乱真的图片。

2019 年, Karras 等人<sup>[11]</sup>提出了基于风格的 GAN 模型, 即 StyleGAN。与传统 GAN 模型最大的区别在于, StyleGAN 的生成器  $G$  中除了由 CNN 构成的合成网络  $h$  以外, 还包含一个用全连接模块搭建的映射网络  $f$ 。为了得到一张合成图像, StyleGAN 先用  $f$  将隐变量  $\mathbf{z} \in Z$  映射为中间变量  $\mathbf{w} \in W$ , 然后用  $N$  个仿射变换模块  $A_i (i = 1, 2, \dots, N)$  分别计算不同尺度层级的风格系数  $\mathbf{s}_i = A_i(\mathbf{w})$ 。最后, 这些风格系数被送入自适应实例正则化模块(Adaptive Instance Normalization, AdaIN)中, 对生成网络  $h$  中的卷积特征图进行调制。除了风格系数外,  $h$  还以一个可学习的张量  $\mathbf{c}$  作为底层输入, 如图 1 所示。除了隐空间  $Z$ 、 $W$  和  $S$  以外, 一些 GAN Inversion 的相关研究也将  $\mathbf{w}$  的复制构成的空间记做  $W^+$ , 常用于基于 StyleGAN 的真实图像编辑任务。<sup>[10]</sup>

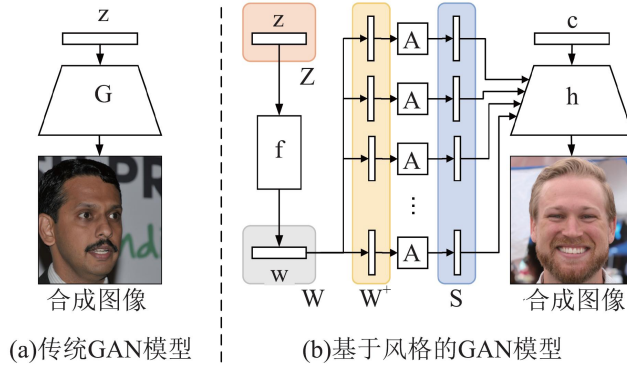


图 1 StyleGAN 模型结构示意图

### 2.2 人脸年龄编辑

人脸年龄编辑大多基于 GAN 实现, 如基于图像域间映射 (如双年龄组、多年龄组等) 的方法、基于年龄特征分解的方法和基于隐空间插值的方法<sup>[10]</sup>。近年来, 基于 GAN 的人脸编辑技术已经取得了显著的进展, 在分辨率、图像纹理和骨骼变化、对多视角训练数据的依赖性等方面都有很大的进步, 但仍存在缺乏大规模、高质量训练数据, 年龄变化时序连续性较差和极端年龄 (如婴儿时期与老年时期) 逼真度较差等问题。

除了 GAN 以外, Stable Diffusion 作为近几年发展迅猛的新模型, 在图像生成方面也展示了巨大的优势, 具有根据描述生成精细多年龄段人脸的能力, 与之对应地, 也需要大量的数据与算力进行训练。由于技术与硬件的限制, 本项目重点关注 Lifespan age transformation 模型的人脸编辑方式。

## 2.3 人脸视角转换

人脸视角转换任务，又被称为人脸矫正任务，常有两种实现方法。

第一种较为传统，充分利用具有鲁棒性的局部描述符（如 Gabor, Haar, LBP）解析局部失真并采用度量学习的方法来实现姿态不变性。相比之下，深度学习的方法通常采用池化来处理位置变化并采用三元损失或对比损失来确保大类内间距的不变性。如李海彦等结合仿射变换和成像原理，对多姿态人脸图像进行姿态调整为近似于正面人脸的方法<sup>[5]</sup>；王科俊等结合图像旋转公式实现了人脸图像的平面旋转校正<sup>[6]</sup>等。然而，由于需要平衡不变性和可分辨性，这些方法不能很好的处理大姿态的问题，且生成效果会存在局部扭曲问题，并不理想。

第二种是利用 3D 渲染工具，先将 2D 图像与一组通用的或者特定的 3D 模型做对齐，通过 3D 几何变换来渲染正脸。这些方法在小 pose 人脸能取得较好的效果，但是对于大 pose 的人脸效果不好，原因是较大的纹理损失。如牛津大学视觉几何组的对 3D 可变形物体的无监督模型<sup>[4]</sup>和基于 Rotate-and-Render 训练策略的通用人脸视角重建模型<sup>[3]</sup>等。近几年，有人提出了基于深度学习的方法改变数据驱动的方式来还原正脸。但是矫正的图像缺少细节并导致在大姿态的情况下产生模糊。

还有一种经典侧脸校正模型 TP-GAN<sup>[7]</sup>，提出了两种途径合成正面照的深度网络结构。两种途径分别关注推理全局结构和转换局部纹理，相应的特征层融入最后的合成中，验证了一种“生成式识别”网络的可能性，并且在大姿态的情况下取得较好的识别。本项目则关注 CR-GAN，其提出的双路径框架可以使得模型处理来自未知空间的数据能力更强。

## 3 本文所用辅助模型简介

### 3.1 Lifespan age transformation (LAT) 的年龄编辑

该论文<sup>[1]</sup>是一个 unpair multi-domain transfer 的图像翻译方法。作者将年龄分成 6 个段，作为单独的 6 个域，能实现一个样本在这 6 个域的连续渐变迁移。巧妙的地方在于，不同于其他人脸年龄编辑的 CGAN 方式，作者使用了一个 mlp 编码年龄向量  $z$ ，进入到一个隐空间。不同年龄段的 latent code 进行插值，用插值之后的 latent code 生成图像，则图像的效果就介于两个域之间。比如说，该论文选择人脸年龄编辑这个方向实验。数据集中各个年龄被化分为几个段： $\{0:6\}, \{16:30\}, \dots$ 。如果想得到  $\{7:15\}$  岁的人脸，则将  $\{0:6\}$  和  $\{16:30\}$  这两个区间对应的 latent code 进行按比例插值即可。如果没有 mlp，现存的方式的做法是， $z$  将会被视作 condition，直接送入 G 中。但是，如果直接在  $z$  空间插值，得到  $z'$ ， $z'$  是无法直接起到作用的。因为在训练阶段，G 就没有见过  $z'$  这种数据分布。在隐空间插值，绕过了这个问题。

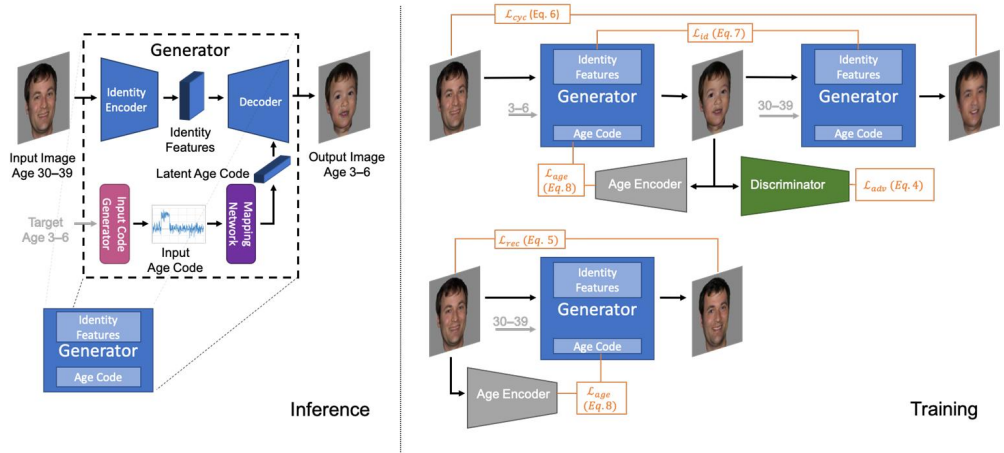


图 2 Lifespan age transformation 的算法示意图<sup>[1]</sup>

此外，由于该工作使用的 FFHQ 数据集<sup>[12]</sup>包含年龄段的性别不平衡。为了防止在输出中引入这些偏见，例如为女性生产男性面部特征或反之亦然，训练了两个独立的模型，一个是男性模型，一个是女性模型。应用哪个模型的决定留给用户自己去决定。

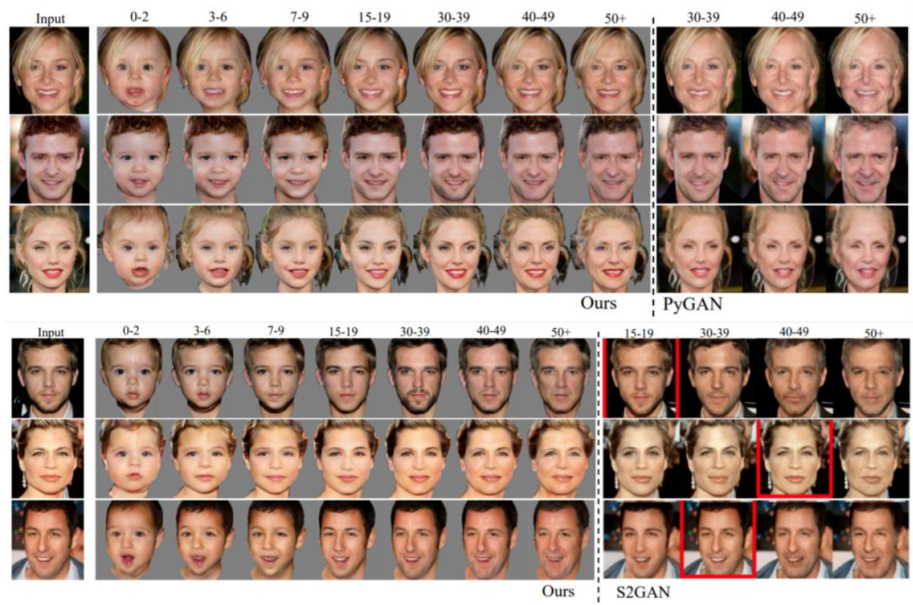


图 3 LAT 的结果与部分 SOTA 模型的比较

比较结果表明，LAT 与 PyGAN 相比，新模型得到了更为清晰和真实的照片预测结果，不同年龄之间的转换也更为顺滑；与 S2GAN 相比，能够产生更清晰的皱纹，相比 S2GAN 来说更适用于年龄较大的人群。

### 3.2 CR-GAN的人脸视角转换

在面对单视角向多视角的转换时，Yu Tian 等人提出了在单一的重建路径之外，引入一条生成路径来保

持学习的嵌入空间的完整性，通过两种参数共享路径的协作和竞争，提高模型对“看不见的”数据集的泛化能力。<sup>[2]</sup>

具体地，在生成路径中，训练生成器  $G$  与判别器  $D$ ，固定编码器  $E$ ，目的是使  $G$  能将随机噪声  $\mathbf{z}$  拟合到真实图片  $\mathbf{x}$  的空间分布，即  $D$  目标为最小化：

$$\mathbb{E}_{\mathbf{z} \sim \mathbb{P}_{\mathbf{z}}} [D_s(G(v, \mathbf{z}))] - \mathbb{E}_{\mathbf{x} \sim \mathbb{P}_{\mathbf{x}}} [D_s(\mathbf{x})] + \lambda_1 \mathbb{E}_{\hat{\mathbf{x}} \sim \mathbb{P}_{\hat{\mathbf{x}}}} \left[ (\|\nabla \hat{\mathbf{x}} D(\hat{\mathbf{x}})\|_2 - 1)^2 \right] - \lambda_2 \mathbb{E}_{\mathbf{x} \sim \mathbb{P}_{\mathbf{x}}} [P(D_v(\mathbf{x}) = v)] ,$$

$G$  目标为最大化：

$$\mathbb{E}_{\mathbf{z} \sim \mathbb{P}_{\mathbf{z}}} [D_s(G(v, \mathbf{z}))] + \lambda_3 \mathbb{E}_{\mathbf{z} \sim \mathbb{P}_{\mathbf{z}}} [P(D_v(G(v, \mathbf{z})) = v)]$$

在重建路径中，固定  $G$ ，训练  $E$  与  $D$ ，目的是使  $E$  学习  $G$  的逆，从而学习潜在的、“看不见”的真实图片的空间分布，同时，要求  $E$  应学习到输入图片的身份特征。所以， $D$  应当最小化：

$$\mathbb{E}_{\mathbf{x}_i, \mathbf{x}_j \sim \mathbb{P}_{\mathbf{x}}} [D_s(\hat{\mathbf{x}}_j) - D_s(\mathbf{x}_i)] + \lambda_1 \mathbb{E}_{\hat{\mathbf{x}} \sim \mathbb{P}_{\hat{\mathbf{x}}}} \left[ (\|\nabla \hat{\mathbf{x}} D(\hat{\mathbf{x}})\|_2 - 1)^2 \right] - \lambda_2 \mathbb{E}_{\mathbf{x}_i \sim \mathbb{P}_{\mathbf{x}}} [P(D_v(\mathbf{x}_i) = v_i)]$$

$E$  应当最大化：

$$\mathbb{E}_{\mathbf{x}_i, \mathbf{x}_j \sim \mathbb{P}_{\mathbf{x}}} [D_s(\hat{\mathbf{x}}_j) + \lambda_3 P(D_v(\hat{\mathbf{x}}_j) = v_j) - \lambda_4 L_1(\hat{\mathbf{x}}_j, \mathbf{x}_j) - \lambda_5 L_v(E_v(\mathbf{x}_i), v_i)]$$

经过自监督训练，使得最终 CR-GAN 能从单一视角生成多视角的人脸图片。

### 3.3 Rotate-and-Render (RaR) 的人脸视角转换

人脸旋转（转正）任务的一大障碍是对高质量的多视角配对训练数据的依赖。因此，它们的生成结果受到数据量以及数据分布的限制，从而会造成对特定数据的过拟合。RaR 方法的出发点是摆脱多视角的监督，从而解决由于多视角监督带来的 domain 泛化能力问题。RaR 提出一种新颖的无监督框架，利用三维人脸建模，通过反复的旋转和渲染操作构建自监督，并用一个普通的 Pix2Pix 生成最终图像。该框架不依赖于同一人的多视角图像，就可以生成其他视角的高质量图像，并且框架适用于各个非限定的场景。此外，将生成的数据用于人脸识别模型的训练，对于人脸识别的性能也有显著提升。

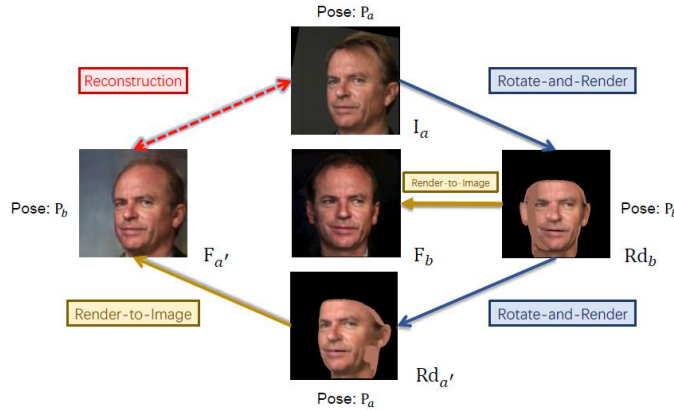


图 4 RaR 提出的无监督训练框架<sup>[3]</sup>

整个框架主要包含三个模块，三维人脸重建模块，通过旋转和渲染构建训练数据的 Rotate-and-Render 模块，以及将渲染图像重建到真实图像分布的 Render-to-Image 模块。其中，三维人脸重建技术在这里只是作为工具使用，核心为后两个模块。模型通过 Rotate-and-Render 来构建训练数据对，通过 Render-to-Image 来消除前面生成过程中的伪影，生成符合真实图像分布的图像。



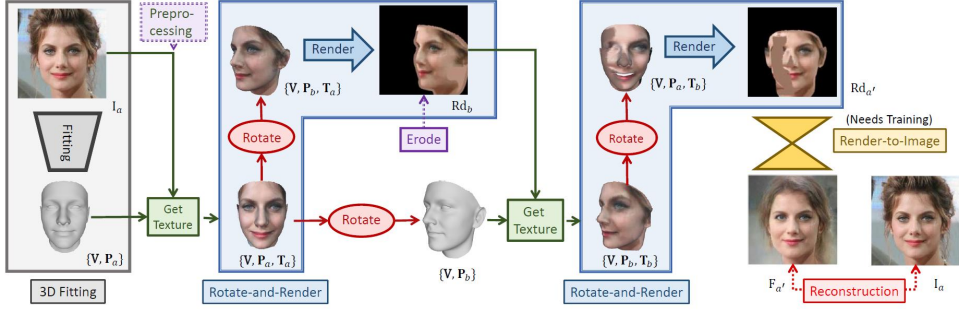


图 5 Rotate-and-Render 的训练数据构建策略<sup>[3]</sup>

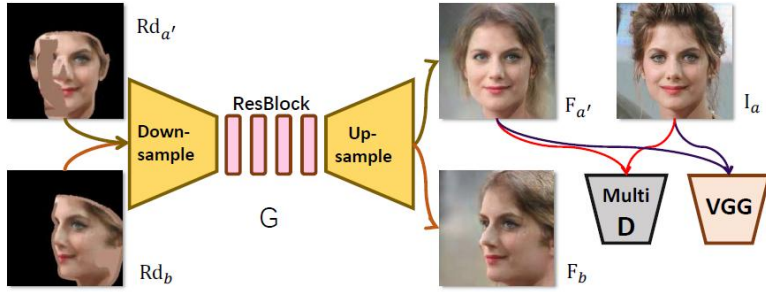


图 6 Render-to-Image 模块<sup>[3]</sup>

此外，在这些模块之外，RaR 对生成器和损失函数 loss function 等网络结构的构建则不是重点，借助了其他现有的工作。在与部分 SOTA 模型对比后可以看出，RaR 的生成效果质量更好。论文还将生成的图片用于人脸识别，也得到了不错的效果。但是由于网络构建不精细，较难生成高分辨率图像。

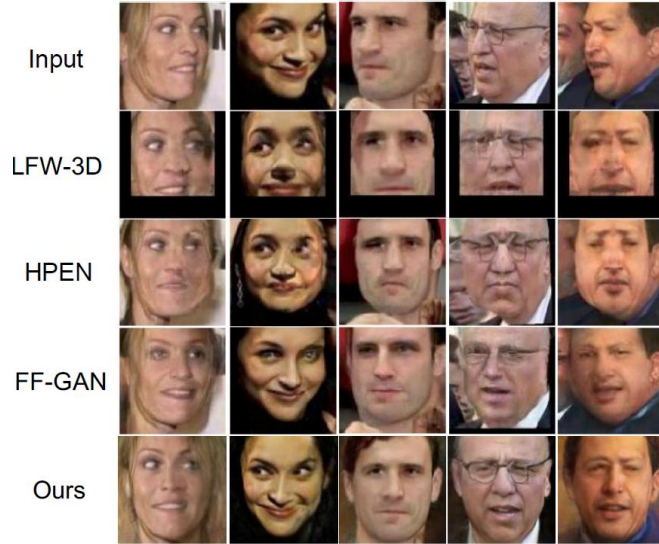


图 7 RaR 的生成效果对比<sup>[3]</sup>

## 4 实验设计

### 4.1 数据集 (list.txt, list\_name.txt)

本项目使用的是来自卡内基梅隆大学的人脸数据 Multi-PIE<sup>[8]</sup>，并经过了 CR-GAN 项目<sup>[2]</sup>的裁剪，如图片大小裁剪、使人脸五官对齐摆正等。由于版权问题，只有 250 位受试者的脸部数据。根据编号的不同，包含多种表情、灯光与角度等类型，本项目提取了其中的正面与左侧 90 度的视角（051: frontal, 正面标签 0; 010: profile, 左侧 90 度标签 1），每人两种视角各约 40 张，共 19920 张身份、表情、灯光等条件均一一对应的正面侧脸图像作为项目的训练集。

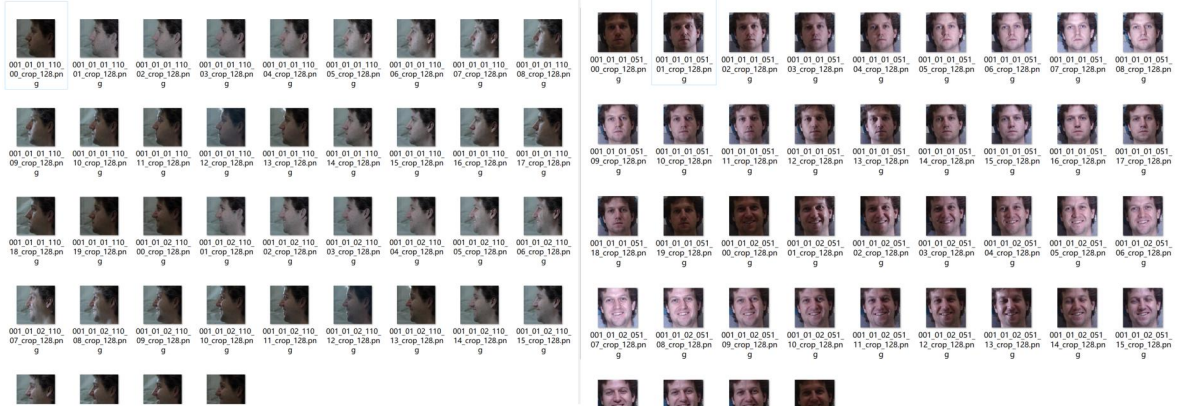


图 8 裁剪后 Multi-PIE 被试 1 的正侧面图

为了观察训练效果，本项目还使用了 CFPW 数据集<sup>[9]</sup>，该数据集包括 500 位名人 (list\_name.txt) 的各 10 张正脸与 4 张侧脸照片（共 7000 张），但正侧面的拍摄条件并非完全对应，且部分侧脸不符合 90 度的要求，故仅作为效果对比用的数据。此外，以上数据集缺少婴幼儿与老年的脸，为了效果对比，会从网络上选取部分照片进行展示。

### 4.2 思路1结构与过程：LAT\_CR-GAN (model1.py)

该思路主要构建了一个 CR-GAN 模型来完成从正脸到侧脸的视角转换，仅对 LAT 进行了 fine-tune。

#### 4.2.1 CR-GAN 模型架构

首先是生成器，在建构代码时为 Gen\_view\_feature\_x 类，在训练时被初始化为 G\_vzx，v 为视角特征共 2 维，z 为其余特征共 128-v=126 维。生成器接受 v 与 z 后生成图片 x，netron 可视化后具体模型结构如下（中间共 5 个残差块）

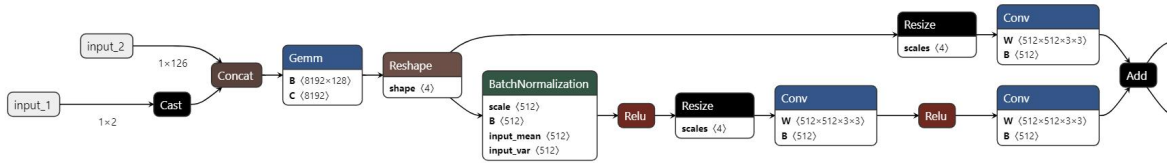


图 9 CR-GAN 生成器 G 残差块 1

图 10 CR-GAN 生成器 G 残差块 2、3

图 11 CR-GAN 生成器 G 残差块 4、5

其次为生成器的逆，即编码器 E，接受图片并生成视角  $v$  与特征  $z$ ，在代码中为 `Gen_x_view_feature`，训练时实例化为 `G_xvz`，其结构经过 `netron` 可视化后为

图 12 CR-GAN 编码器 E 残差块 1、2

图 13 CR-GAN 编码器 E 残差块 3、4

图 14 CR-GAN 编码器 E 残差块 5

判别器在模型中为 `Dis_x_view_score` 类，训练时实例化为 `D_xvs`，与 `G/E` 类似，也含有 5 个残差块，由于不是重点，此处便略去展示。

### 4.2.2 CR-GAN 训练过程

同 3.2 中的介绍，在生成路径中，训练  $G_{vzx}$  与  $D_{xvs}$ ，完成这样的路径：

```
(v, z)-->G  vxz-->x  bar-->D  xvs( (v,x  bar), (v,x  real) )
```

其中，分类损失采用 WGAN 的计算方式，并加入梯度 Lipschitz 惩罚项。损失函数计算如下所示（具体数值见 train.py 代码）。

```
# 判别器的目标是能识别img1的视角，区分真实图片与G_vzx生成的图片
d_xvs_loss = x_bar_loss_s - x_loss_s + 10. * gp_loss + v_loss_x
```

图 15 D xvs 损失函数



```
# 生成器G_vzx总损失
g_vzx_loss = -x_bar_loss_s + v_loss_x_bar
```

图 16 G\_vzx 损失函数

在重建路径中，训练 D\_x（实际为 D\_xvs）与 G\_xvz

```
# 判别器的目标是能识别另一视角，区分G_xvz+G_vzx生成的图片（可能是G_vzx从随机图片生成的v1视角A/从img1生成的v2视角B）
# A与B都是对G_xvz特征提取能力的考验，B还需考验G_vzx根据v1的feature生成v2的能力
d_x_loss = x_bar_bar_loss_s - x_loss_s + 10. * gp_loss + v_loss_x
```

图 17 D\_x 损失函数

```
# 对G_xvz，目标是最大化判别器得分，最小化x_bar_bar与原图的图像差距和视角差距，最小化v_bar与vv1的差距
g_xvz_loss = -x_bar_bar_loss_s + 4*x_l1_loss + v_loss_x_bar_bar + 0.01*v_loss_x
```

图 18 G\_xvz 损失函数

采用 Adam 优化器，学习率为 1e-4，动量参数为[0.5, 0.9]，迭代训练 150 轮，得到 CR-GAN。

#### 4.3 思路2结构与过程：RaR\_LAT\_RaR

由于数据集、技术与硬件条件的限制，难以复现出比原模型效果更好或类似的模型，所以这个思路仅仅进行了模型拼接，即 90 度左侧脸→RaR→正脸→LAT→不同年龄正脸→RaR→不同年龄左侧脸。经过了漫长的配环境与调试过程，也得到了结果。

## 5 结果对比

### 5.1 思路1LAT\_CR-GAN

#### 5.1.1 CR-GAN 训练结果

在训练了 150 轮后，各部分的损失曲线变化如图 19 所示。

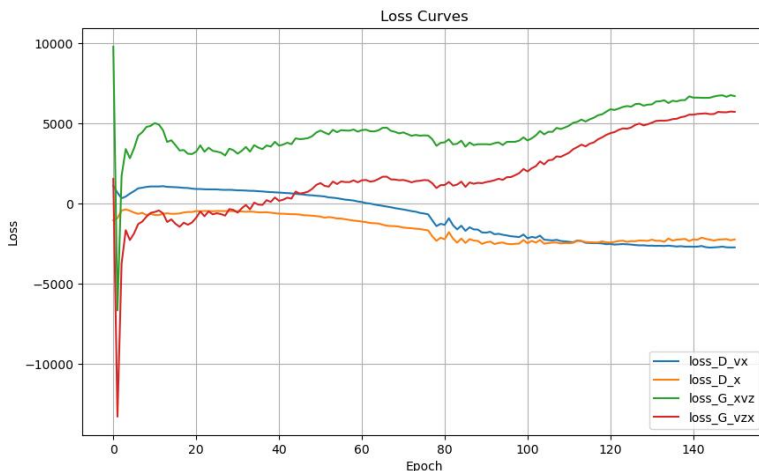


图 19 CR-GAN 各部分损失变化

可以看出，模型并未收敛，但考虑到训练时长开销过大，所以在得到了较好的效果以后，提前终止训练。根据各损失的意义，似乎在第 80 轮时模型已经收敛，但是从输出效果来看，先不论拟合度，仅从生成图片的分辨率来看，第 150 轮时更高，所以仍选择第 150 轮的训练结果作为展示。从图 20 可以看出，生成器将随机向量拟合到真实图片空间的能力已经较强，虽然图片仍会存在一些伪影和扭曲，但总的来说效果很好。从图 21 可看出，虽然分辨率不够高，但编码器 E 也较好地学习到了 G 的逆。

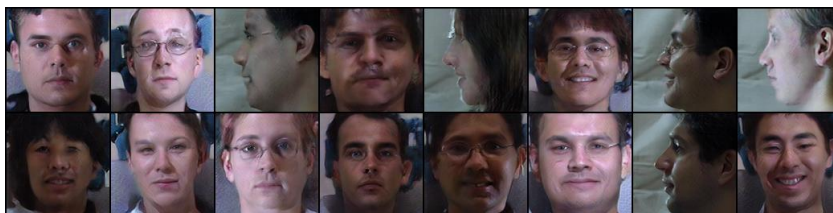


图 20 生成器从随机向量生成真实图片的效果



图 21 生成器 G 将编码器 E 的输出生成真实图片的效果

### 5.1.2 LAT\_CR-GAN 生成结果

从生成结果来看，效果不理想。一方面可能是 CR-GAN 未收敛，另一方面可能是由于 CR-GAN 的训练数据缺少婴幼儿时期与老年时期，无法将输入拟合到婴幼儿和老年的分布空间，导致 LAT 的生成结果再经过 CR-GAN 后，年龄特征被抹去，而都集中分布于青中年时期。如果时间充足，应该在训练数据集中加入更多 0-20 岁与 60 岁+的训练图像进行训练。



图 22 LAT\_CR-GAN 生成结果

## 5.2 思路2PaR\_LAT\_RaR

在调试完成后，原始图片与生成图片如下所示。可以看出，思路 2 生成的图片年龄特征比思路 1 更明显，身份特征保留得也更好，但是伪影更多，精细程度一般。

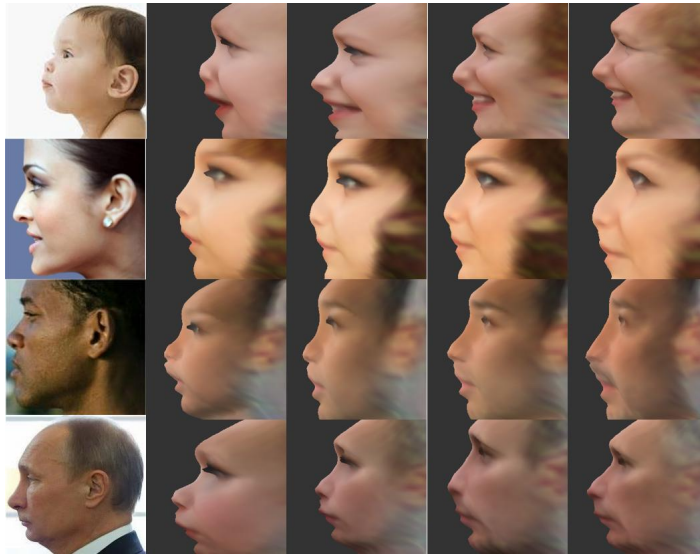


图 23 PaR\_LAT\_RaR 生成结果

## 6 讨论与改进

### 6.1 数据集与硬件限制

1. 难以找到完全符合标准的数据集：基于 GAN 的人脸年龄编辑模型效果需要大规模、高质量的训练数据，现在常用的公开人脸年龄编辑数据集较多，如 FG-NET、MORPH、CACD、IMDB-WIKI、UTKFace、FFHQ-Aging 等<sup>[10]</sup>。这些数据集有许多包含跨度足够大的年龄范围（UTKFace 为 0-116 岁）、种族、性别、姿态和受控拍摄条件等，但大多为人物正脸或非标准 90 度侧脸，难以同时兼具人脸视角编辑的任务，若要从头开始训练模型，需要大规模的数据提取、裁剪、处理和人工标注等操作。所以，本次作业只能退而求其次，将年龄编辑与脸部视角编辑任务解耦，分开实现。若有能力构建足够多受控条件下带有年龄标注的正侧脸数据集，模型可以进一步改进。

2. 数据集年龄性别比例不平衡：目前本文主要使用的数据集 Multi-PIE 存在性别和年龄比例不平衡的问题，如缺少婴幼儿和老年照片。而在测试中发现，模型对男女性脸部视角的转换效果有一定差距，参考资料 LAT 模型在青少年到中年的转换效果比幼年与老年更好。这种情况有客观原因，比如男女性、幼年与老年的脸部纹理和头骨发育变化较大等。如果能针对输入数据的性别、年龄段特化训练对应的模型，或许可以获得更好的效果。

3. 硬件限制。考虑到成本与时间问题，本文使用的 GPU 来自 Colab 与个人游戏本，性能并非目前的 top 水平，所以训练时间十分长，自行设计的 CR-GAN 在未完全收敛时只能提前停止，更不用说针对性别与年龄段特化训练模型。如果能接触到算力更高的硬件设备或是时间更宽裕，或许可以得到更好的实现效果。

## 6.2 模型结构优化

1. 结构优化：目前人脸视角编辑最经典的模型是 TP-GAN<sup>[7]</sup>，提出了一个有两条路径的深度架构，用于正面人脸图像合成。这两条路径，一条专注于推理全局结构，另一条则推理局部的纹理，分别得到两个特征地图。这两个特征图会融合在一起，用于接下来的最终合成，得到了十分优秀的生成效果。这种仿生的模型设计方式可以用于优化本文所使用的 CR-GAN，但由于这两个模型承担的任务互逆（正侧脸转换），移植与优化难度较大，由于时间与技术限制，暂未实现。若实现了这种仿生优化，可能得到更好的实现效果。

2. 设计思路优化：本文设计的两条思路并未完全控制变量，拿来直接对比并不算很合适，但由于时间、硬件与技术的限制，没有将更严谨的模型组合训练出来并对比效果。

## 6.3 效果评价指标

本文所使用的效果评价指标仍局限于主观判断，而根据参考文献<sup>[10]</sup>，目前可使用的定量指标囊括了输出图像真实性、年龄变化准确性、无关语义一致性等多种方面。若系统地使用这些指标，可以有更科学的结果呈现。除此以外，目前对侧脸进行年龄编辑的工作较少，所以暂时未找到合适的模型与本文的模型进行直接的效果对比。

## 7 感想与总结

由于目前并没有完全一样的项目可供参考，本次大作业是我对一个陌生任务的尝试与实现。在完成作业和形成论文的过程中，我经历了漫长的搜索资料、阅读文献与对比模型的过程，在最终实现和效果检验的时候，也经历了不小的波折与考验，我也在其中得到了锻炼和成长。回顾这门课程，感谢老师与助教们的课程讲解和作业设计，让我接触并尝试构建神经网络去完成一个个任务，感受深度学习的魅力并从中学习到知识。我相信，这对我未来的学习道路也有很大的帮助。

### 参考文献:

- [1] Or-El R, Sengupta S, Fried O, et al. Lifespan age transformation synthesis[C]//Computer Vision - ECCV 2020: 16th European Conference, Glasgow, UK, August 23 - 28, 2020, Proceedings, Part VI 16. Springer International Publishing, 2020: 739-755.
- [2] Tian Y, Peng X, Zhao L, et al. CR-GAN: learning complete representations for multi-view generation[J]. arXiv preprint arXiv:1806.11191, 2018.
- [3] Zhou H, Liu J, Liu Z, et al. Rotate-and-render: Unsupervised photorealistic face rotation from single-view images[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 5911-5920.
- [4] Wu S, Rupperecht C, Vedaldi A. Unsupervised learning of probably symmetric deformable 3d objects from images in the wild[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1-10.
- [5] 李海彦, 徐汀荣, 张立晓, et al. 基于仿射变换的多姿态人脸矫正和识别[J]. 计算机应用研究, 2014, 31(4). DOI:10.3969/j.issn.1001-3695.2014.04.062.
- [6] Kejun W, GuoFeng Z, Guoxia F. An approach to fast eye location and face plane rotation correction. J Comput Aided Des Comput Graph. 2013;25(6):865-872,879.

- 
- [7] Huang R, Zhang S, Li T, et al. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2439-2448.
  - [8] Ralph Gross, Iain Matthews, JeffreyCohn, Takeo Kanade, and Simon Baker. Multi-pie.ImageVision Computer, 28(5):807–813, 2010.
  - [9] S. Sengupta, J.C. Cheng, C.D. Castillo, V.M. Patel, R. Chellappa, D.W. Jacobs, Frontal to Profile Face Verification in the Wild, IEEE Conference on Applications of Computer Vision, 2016.
  - [10] LIU Yunfan,LI Qi,SUN Zhenan,TAN Tieniu. Face Age Editing Methods Based on Generative Adversarial Network: A Survey[J]. Frontiers of Data and Computing, 2023, 5(2): 2-23, <https://cstr.cn/32002.14.jfdc.CN10-1649/TP.2023.02.001>.
  - [11] KARRAS TLAIN S.AILA TA Style-Based Generator Architecture for Generative Adversarial Networks[J]IEEE Transactions on Pattern Analysis and Machine Intelligence,2021,43(12): 4217-4228.
  - [12] Karras,T.,Laine,S.,Aila,T.:A style-based generator architecture for generative adversarial networks. In:Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.pp.4401 – 4410(2019).