

年龄变换项目报告

李亦杨 10195101467 刁泽皓 10195101470

摘要：在此处输入中文摘要（字数一般不少于 500 字）。摘要必须反映全文中心内容，内容应包括目的、过程及方法、结论。要求论述简明、逻辑性强、尽量用短句。采用第三人称的写法，并请用过去时态叙述作者工作，用现在时态叙述作者结论。

1 生成对抗网络

生成对抗神经网络 (Generative Adversarial Networks) 于 2014 年提出，其可以学习数据的分布规律，并创造出可以以假乱真的图像和文本。^[?]

GAN 的网络结构分为生成器和辨别器两部分。生成器读取一个随机的噪音输入，并输出一个图片，辨别器判别图片是否为训练集中的图片。当辨别器无法分辨真假（即判断正确的概率为 50%）时，训练结束。生成器和辨别器要分别搭建神经网络模型，可以根据实际需求选择卷积神经网络、递归神经网络或全连接神经网络。

GAN 的模型训练过程可以大致概括为以下几个步骤：

1. 初始化生成器 G 和辨别器 D 两个网络的参数。
2. 从训练集中抽取 n 个样本，同时让生成器利用定义的噪声 z 分布生成 n 个样本。固定生成器 G ，并训练辨别器 D ，使其尽可能区分真假。
3. 循环更新 k 次辨别器 D 之后，更新 1 次生成器 G ，使辨别器尽可能地区分不了真假。

在进行多次更新迭代后，理想状态下，只要最终辨别器 D 无法区分出图片到底是来自真实的训练样本集合，还是来自生成器 G 生成的样本即可。此时辨别的概率为 0.5，完成训练。

迭代过程中的关键点在于损失价值函数的计算，Generative Adversarial Networks 一文中提出了其计算公式^[?]：

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

其中真实数据 x 服从 p_{data} 分布， $p_z(z)$ 、 $G(z)$ 是将噪音输入 z 映射到数据空间拟合真实图像后的分布， $D(x)$ 为得到真实数据（换言之，即辨别器无法分辨真假的数据）的概率。训练中需要取得使得 D 最大的 x 分布，同时最小化伪造的概率 $\log(1 - D(G(z)))$ ，以这一条件来训练生成器。

文章中严格证明了：

- 最优的生成器 G 存在满足 $P_G = P_{data}$ 的唯一解，换言之当生成的图像分布与原数据分布近似的时候存在最优生成器。
- 最优辨别器需要满足 $D(x) = \frac{P_{data}}{P_{data} + P_G}$ ，此时可以保证 $V(D, G)$ 能够取得唯一的极大值 0.5。

- 只要有足够的训练数据和正确的环境，训练过程一定会收敛到最优。

2 现行的主流年龄变化方案

在 Lifespan Age Transformation Synthesis 之前，主流的年龄变化训练方案主要是为衰老的特定子因素建立单独的模型，其通常包含皱纹、纹理等代表性因素，同时着重于在离散的特定年龄下的转换。这些模型通常仅限于特定的域子集，即以风格或纹理差异为特征的同质域，换言之，这些转换常常基于风格迁移的思想。诸如 DiscoGAN 或 CycleGAN 一类的网络可以很好地完成这类图像翻译任务。在生成器部分，需要利用若干个卷积层构建编码器，用若干个卷积层构建解码器，同时在两者之间利用一个能够产生变化的深度网络来担当特征转换器。DiscoGAN 使用 CNN 编码器和解码器，使用全连接网络当转换器，而 CycleGAN 则使用了 ResNet 充当转换器。^[?]

基于此思路之上，基于 CycleGAN 实现的 S2GAN 在拟真性方面作出了改进，其通过抽象出身份信息，将个人的衰老基础以及传统年龄变换方案运用的群体化的特定年龄转变模式相结合，以生成具有个性化特征的人脸老化方案。S2GAN 在编码器部分提取出个人的衰老模式作为个人特征，在解码器将人脸衰老特征解码，通过对几个年龄组进行加权得到目标图像，并在中间部分将线性组合系数与个人衰老特征相乘（而不是使用传统的连接）来完成年龄变化操作。^[?]

然而以图像翻译作为年龄变换的主要思路，忽视了在年龄变化过程中头部发育带来的结构、形状变化，以及跨年龄段的外观变化问题。这也使得这些模型给出的年龄变换后的输出图像，相较于输入图像，往往只能呈现出在脸部各个子区域上的皮肤材质纹理（如皱纹）方面的变化。

3 另一种年龄变化方案

3.1 网络结构

由于针对年龄转换这一应用场景，缺少捕捉人不同年龄段图像的大数据集，所以监督学习变得很困难，故转向对抗性学习，使用六个锚定的年龄类别来近似连续的老化过程，并提出了新的生成式神经网络结构。该网络由一个条件生成器和一个判别器组成，条件生成器负责跨年龄组的转换。

整个网络结构和训练过程如下图：

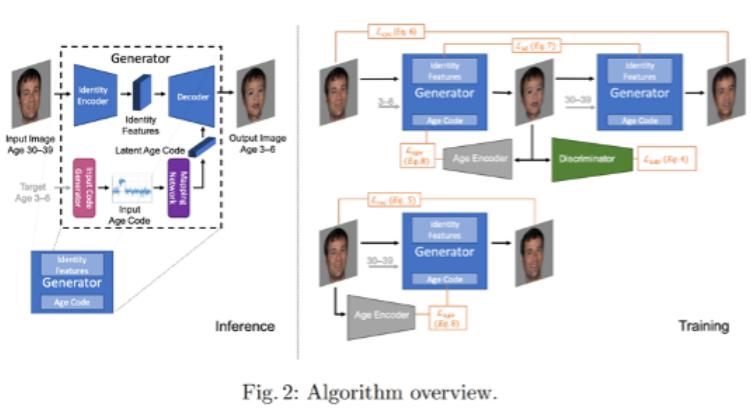


Fig. 2: Algorithm overview.

图 1: 网络架构

在预处理阶段，图像的背景和服饰元素通过使用相应的语义掩码被去除。年龄输入空间 Z 由一个 $50 * n$ 的向量组表示，其中 n 是年龄类别的数量，当输入年龄种类为 i 时，我们生成一个向量 z_i :

$$z_i = \mathbf{1}_i + v_i \quad v \sim N(0, 0.2^2 \cdot I)$$

其中 $\mathbf{1}_i$ 是一个 $50 \cdot n$ 元素向量，它包含元素 $50 \cdot i$ 到 $50 \cdot (i+1) - 1$ 上的 1 和其他地方的 0， I 是单位矩阵。一个生成器被用来生成所有的目标年龄，它由三部分组成：身份编码器、潜在映射网络和解码器。训练过程中还使用了一个年龄编码器用来将真实和生成的图像嵌入年龄潜伏空间。

身份编码器 E_{id} 读取输入图像 x 并提取一个身份特征张量 w_{id} ，有 $w_{id} = E_{id}(x)$ 。

映射网络 $M : Z \rightarrow W_{age}$ ，将年龄输入向量 Z 嵌入到统一的年龄潜在空间 W_{age} 。 M 是 8 层 MLP 网络， W_{age} 是 256 元素的潜在向量。映射网络学习了最佳的年龄潜在空间，使之能够在年龄集群之间实现平滑的过渡和插值。

解码器读取年龄潜在编码和身份特征，并生成输出图像 $y = F(w_{id}, wage)$ 。其中身份特征 w_{id} 由样式化卷积块处理，为减少水滴伪影，模型将 AdaIN 归一化成替换为 StyleGAN2 中提出的调制卷积层，并且在每个调制卷积层后有一个像素范数层用来进一步减少伪影。

综上，从输入图像 x 和输入目标年龄向量 z_t 到输出图像 y 的整体生成器映射为：

$$y = G(x, z_t) = F(E_{id}(x), M(z_t))$$

年龄编码器将输入图像 x 映射到其在年龄向量空间 Z 中的正确位置。产生一个年龄向量 $z_s = E_{age}(x)$ ，对应与图像 x 的源年龄簇 s 。鉴别器，使用具有具有小批量标准偏差的 StyleGAN 鉴别器，并将最后一个全连接层修改为具有 n 个输出，以区分多个类别，对第 i 类的真实图像只会惩罚第 i 个输出。

3.2 训练过程

为了为了补偿年龄集群之间的不平衡，在每次训练迭代中，我们首先对源集群 s 和目标集群 t 进行采样，然后从每个类别中采样一张图片，执行三个前向传递：

$$\begin{aligned} y_{gen} &= G(x, z_t) \\ y_{rec} &= G(x, z_s) \\ y_{cyc} &= G(y_{gen}, z_s) \end{aligned}$$

这里， y_{gen} 是目标年龄 t 的生成图像， y_{rec} 是源年龄 s 的重建图像，并应用一个循环从年龄 t 生成的图像 y_{gen} 重建源年龄 s 的 y_{cyc} 。这些通道提供了所有必要的信号，以最小化以下损失函数，即：

1. 对抗性损失
2. 自我重建损失
3. 循环损失
4. 身份特征损失

5. 年龄向量损失

4 代码实际运行结果

由于设备原因，我们使用已经训练好的模型进行图像年龄转换。我们分别使用 StarGAN（基于 CycleGAN）与 Lifespan Age Transformation Synthesis 一文中提出网络，对 S2GAN 一文中使用的图片进行年龄转化。

原始图片如下：



图 2: 原始图片

所得出的结果如下图：

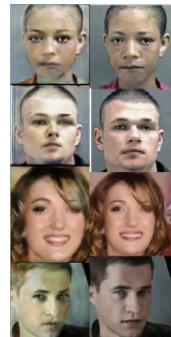


图 3: StarGAN, 原始图片在最右侧



图 4: S2GAN, 原始图片在最左侧

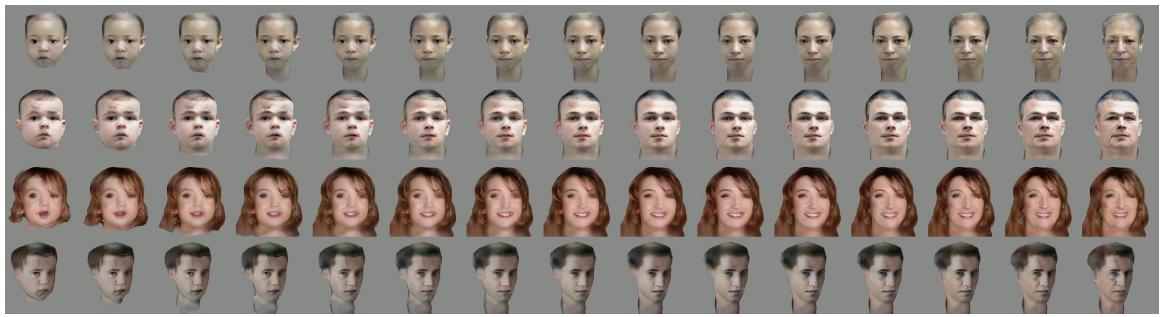


图 5: Lifespan Age Transformation Synthesis

可以发现，论文中提出的这一新模型在模拟年龄变化上具有更好的真实性，即人类成长过程中身体结构的变化也被考虑到了。然而，一些特定角度下的照片可能无法较好地模拟，同时一些纹理上的表现可能有所不足。