

上海大学学报(自然科学版)
Journal of Shanghai University(Natural Science Edition)
ISSN 1007-2861,CN 31-1718/N

《上海大学学报(自然科学版)》网络首发论文

题目: 一种利用年龄编辑改进年龄估计的方法
作者: 张辰昱, 徐树公, 黄剑波
收稿日期: 2020-03-23
网络首发日期: 2020-09-15
引用格式: 张辰昱, 徐树公, 黄剑波. 一种利用年龄编辑改进年龄估计的方法. 上海大学学报(自然科学版).
<https://kns.cnki.net/kcms/detail/31.1718.n.20200915.0938.002.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

一种利用年龄编辑改进年龄估计的方法

张辰昱^{1,2}, 徐树公², 黄剑波^{1,3}

- (1. 上海大学 上海电影学院, 上海 200072;
2. 上海大学 上海先进通信与数据科学研究院, 上海 200444;
3. 上海大学 上海电影特效工程技术研究中心, 上海 200072)

摘要：人脸年龄估计技术(face age estimation)作为一种新兴的生物特征识别技术，是计算机视觉中一个经典的学习问题。基于生成对抗网络(generative adversarial network, GAN)强大的数据生成能力，本工作提出了一种利用年龄编辑改进年龄估计的方法。通过分阶段联合训练年龄编辑网络 StarGAN 和年龄估计网络 SSR-Net(soft stagewise regression network)，扩增适用于年龄估计的训练数据。实验证明，在非受限条件下，该方法取得了年龄估计最好的结果。

关键词：年龄编辑；年龄估计；生成对抗网络；数据扩增；

中图分类号： 391.4 **文献标志码：** A

A method for improving age estimation using age editing

ZHANG Chenyu^{1,2}, XU Shugong², HUANG Jianbo^{1,3}

- (1. Shanghai Film Academy, Shanghai University, Shanghai 200072, China;
2. Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444
China;
3. Shanghai Engineering Research Center of Motion Picture Special Effects,
Shanghai University, Shanghai 200072, China)

Abstract: As an emerging biometric identification technology, Face age estimation is a classic learning problem in computer vision. Based on the powerful data generation ability of Generative adversarial network, this work proposes a method to improve age estimation by using age editing. The training data suitable for age estimation were amplified by combining the training age editing Network StarGAN and the age estimation Network SSR-Net (Soft Stagewise Regression Network) in stages. Experimental results show that the method achieves the best age estimation results under unrestricted conditions.

Key words: Age editor; age estimation; Generative adversarial network; data amplification;

收稿日期： 2020-3-23

基金项目： 上海大学电影学高峰学科和上海电影特效工程技术研究中心研究项目(16dz2251300)

通信作者： 黄剑波 (1980—)，男，讲师，硕士生导师，博士，研究方向为艺术学理论、图像处理等。

1 介绍

人脸年龄估计,即提供一张图片,自动识别图中人物的年龄,属于生物特征识别的范畴,对于人脸识别具有辅助作用。这项技术有很多应用,如视频监控、产品推荐、人机交互、市场分析、用户画像、年龄变化预测等^[1]。由于人脸的多样性、年龄的欺骗性,估计结果容易受光照、姿态、表情等外界因素的影响,人脸年龄估计是一个困难的任务。

传统的年龄估计基于图像处理知识,对人脸图片进行特征提取,再利用如 SVM(Support Vector Machine, 支持向量机)^[2]、Bayes(贝叶斯理论)^[3]等机器学习的理论进行年龄估计,这种方法运算耗时长,准确率低。

深度学习理论的提出、网络大数据时代的来临,基于深度卷积神经网络的人脸年龄估计技术已成为研究热点。尽管基于深度学习的人脸年龄估计研究取得了巨大的进展,但非受限条件下年龄估计的效果仍不能满足实际需求,由于实际应用场景复杂,容易导致人脸年龄估计效果不佳^[4]。

另一方面,虽然人脸识别的数据集很多,但专注于年龄估计任务的数据集却很少,并且许多数据集年代久远,存在标注错误、标注不清晰等问题^[1,4]。生成对抗网络的强大数据生成能力,让我们意识到 GAN 可以在年龄估计任务中起到数据扩增的作用,提升年龄估计准确率。

2 相关工作

2.1 年龄估计

基于深度学习的年龄估计算法大致分三类:用输入图片直接进行分类或回归;使用预测值与概率相乘求期望;引入排序和概率分布信息等^[12]。

直接对人脸图片进行特征提取并判断年龄值是最基本的年龄估计算法,问题在于:连人眼都难以分辨准确的年龄值,卷积神经网络(Convolutional Neural Network, CNN)也是如此,简单的将年龄估计作为回归任务直接预测年龄值是不合理的。[5]首次提出将年龄值划分为几个区间,减少年龄的离散度,提高准确率,把年龄估计作为一个分类任务。

但是,分类任务无法得到准确的年龄值,DEX^[6]提出把年龄估计这一回归问题用分类问题解决,通过预测概率和标签值相乘的求期望方法,可得到准确的年龄值。目前,很多表现优秀的年龄估计算法^[9, 11, 12]都是通过预测值与概率相乘求期望的方法进行预测的。本工作采用的 SSR-Net^[11]就是典型的第二种年龄估计算法。

根据人脸识别中判断一张人脸身份信息困难,但判断两张人脸是否为同一人容易^[7]的理论,OR-CNN^[8]首次将年龄回归问题转为一系列有序的多个分类的子问题,每个子问题只对是否比某个年龄大进行判别,按照有序的回归问题进行年龄估计。

目前,大多数年龄估计算法仅利用人脸检测出的区域进行预测,人脸的其他区域例如头发部分则被丢弃,C3AE^[6]认为这些信息对于年龄估计同样重要,提出利用三种不同尺寸的区域同时进行学习的方法,取得了比之前所有算法更优异的结果,本质上,C3AE 也是一种利用数据扩增改进年龄估计的方法。

由于现有人脸年龄估计数据集的限制,如何在不进行大量数据采集和标注的前提下实现数据扩增是未来人脸年龄估计研究的重要方向之一。

2.2 年龄编辑

年龄编辑 (Age editor), 即输入一张人脸图片, 通过某些方法改变该人脸的年龄并输出。大多数基于 GAN 的图像翻译 (Image translation) 算法^[13, 14, 15, 16]都可以应用于年龄编辑中。

目前, 利用 GAN 进行年龄编辑的算法都是在 CelebA (Large-scale CelebFaces Attributes Dataset)^[17]数据集上进行的, 这类年龄编辑的典型算法有 CycleGAN^[15]、StarGAN^[18]等。CelebA 数据集由香港中文大学在 2016 年提出, 包含 40 种性别、发色、是否微笑等人脸属性, 是主流的人脸图片编辑的数据集。

但由于 CelebA 数据集的标注只有年轻或年老两种, 并不包含具体的年龄值, 因此这类算法只能学习单一映射, 无法进行指定年龄段人脸图片的生成。同时, 大多数算法只专注于年龄编辑后生成图片的视觉效果, 忽略了生成的图片在人脸年龄估计可以提供的的作用。

本工作提出了一种利用年龄编辑改进年龄估计的方法, 创新点在于:

1. 联合训练图像生成网络 StarGAN 和年龄估计网络 SSR-Net, 提出一种适应于年龄编辑的分阶段训练策略。在第一阶段训练 StarGAN 进行年龄编辑时, 加入了 SSR-Net 作为另一个鉴别器进行年龄的约束, 同时对 StarGAN 中的鉴别器结构进行改进, 使其在判断图片是否为真和预测图片年龄域的同时, 也可以预测图片的身份信息; 在第二阶段利用 StarGAN 的生成数据与原数据一同训练 SSR-Net, 扩增用于年龄估计的数据集;

2. 最终实验结果表明, 我们的方法提高了非受限条件下的年龄估计准确率, 在 Adience 数据集上达到了目前最好的结果。

3 方法

3.1 网络结构

我们的目标是训练 StarGAN 在不丢失身份信息的同时学习不同年龄域之间的映射, 并利用网络生成的数据扩增数据集, 提升年龄估计准确率。为了实现这一点, 我们提出了一种分阶段联合训练的网络结构。整个网络使用了一个生成器 G (Generator)、一个鉴别器 D (Discriminator) 和一个 SSR-Net 网络。

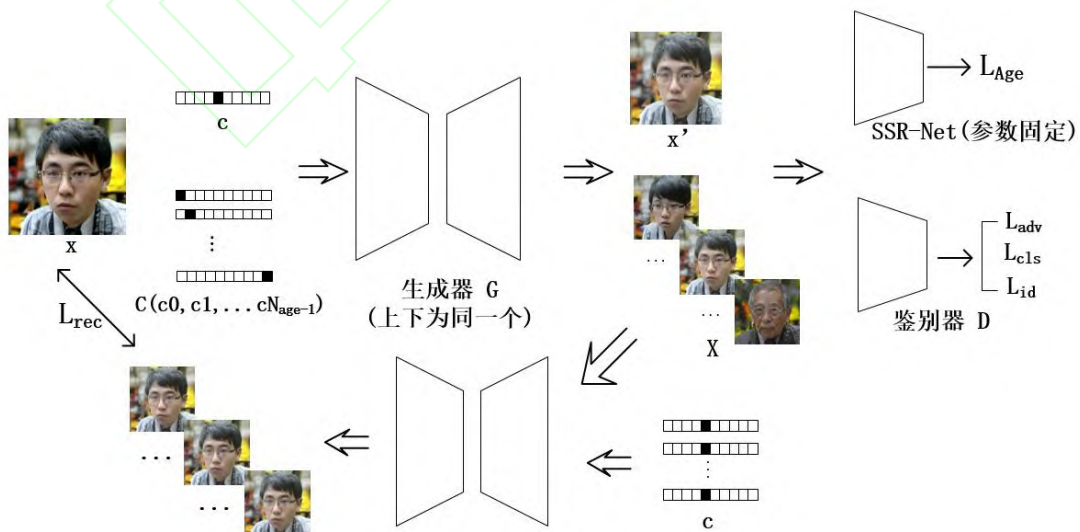


图 1 图像生成网络结构图

Fig.1 image production network structure

3.1.1 图像生成

如图 1 所示，图片生成阶段分为三个部分：图像重建、图像年龄编辑和循环一致。

步骤一：图像的重建部分，原图 x 和标注 c 作为输入送到生成器 G 进行图片的重建，生成与原图一致的生成图片 x' ；图片 x 和 x' 作为输入送至鉴别器 D 进行真假判定，并预测其身份和年龄信息；图片 x 和 x' 作为输入送至 SSR-Net 网络中进行年龄估计，两张图片的预测年龄应一致。

步骤二：图像的年龄编辑部分，原图 x 与 one-hot 形式的年龄标注集 $C(c_0, c_1, \dots, c_{N_{age}-1})$ 作为输入送至生成器 G 生成不同年龄段的人脸图片；为了平衡数据，我们将原图复制 N_{age} 份，并与这些生成的图片一同作为输入送至鉴别器 D 中进行真假判定，并预测其身份和年龄信息，该过程同时受 SSR-Net 对图片年龄预测结果的约束。

步骤三：为了保证图片在编辑前后保持内容不变，只改变年龄相关的部分，我们仿照 CycleGAN 的结构^[15]，在网络中增加了循环一致部分，将步骤二生成的图片和原标注 c 作为输入再次送入生成器 G 中，生成符合原年龄域标注 c 的人脸图片，并与原图作比较。

StarGAN 网络结构如图 2 所示，生成器由 3 个下采样模块，3 个卷积 Bottleneck 层(瓶颈)，3 个反卷积 Bottleneck 层和 3 个上采样模块组成。每个卷积结构都由卷积层、BN 层和 Relu 激活函数组成。Bottleneck 层卷积核为 3×3 ，步长为 2，周围填充为 1，其余模块除第一层和最后一层卷积核为 7×7 之外，其余采样模块的卷积核均为 4×4 ，步长均为 2，周围填充为 1。鉴别器由 6 个卷积层组成，每一层之间均有 BN 层和 Relu 激活函数。卷积核为 4×4 ，步长为 2，周围填充为 1。最终经全连接层映射至三个输出，分别为 2×1 ， $N_{age} \times 1$ 和 $N_{class} \times 1$ 。

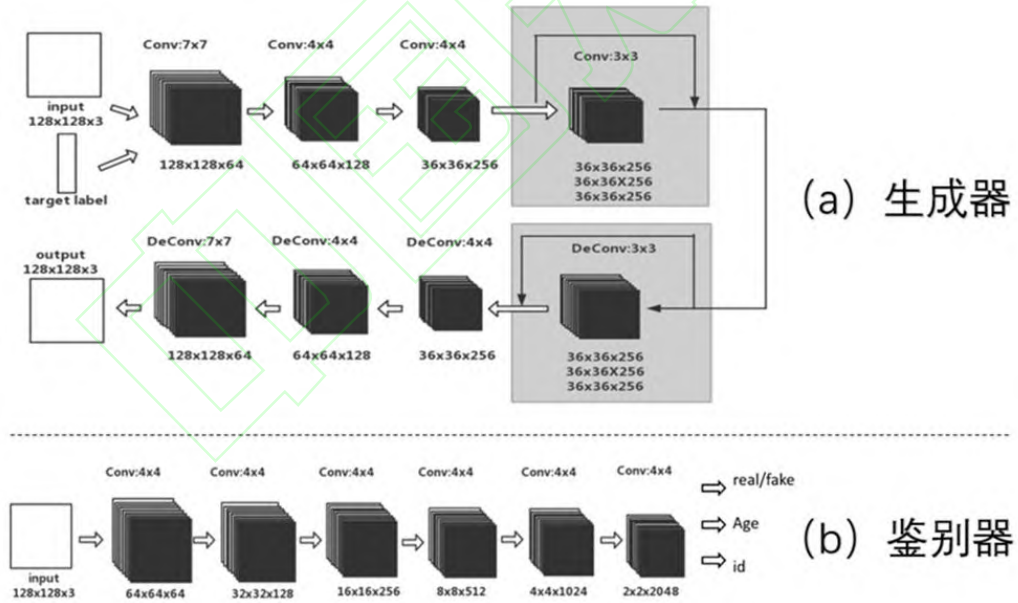


图 2 StarGAN 网络结构

Fig.2 StarGAN network structure

3.1.2 年龄估计

由于年龄信息的零散性和不易区分性，以及大多数年龄数据集的分布特点，将年龄划分 0-3、4-7、8-14、15-21、22-32、33-43、44-54、55-65 和 66+这 9 个区间，年龄值转换为对应年龄区间为 1，其余区间为 0 的 one-hot 形式的年龄标注。

SSR-Net 网络采用预测值与概率相乘求期望进行年龄估计，与上述年龄分段想法一致，更好约束，便于收敛；其次，SSR-Net 网络结构简单，即使加入复杂的 GAN 网络，也不会使得整个网络架构过大而出现无法训练的情况，且网络易于配置，是目前最流行的年龄估计框架；最后，SSR-Net 有提供预训练模型，便于网络训练，因此我们采用 SSR-Net 进行年龄估计。

在图片生成阶段中，SSR-Net 充当另一鉴别器的作用，将网络预测的年龄值与原图像年龄进行比较，作为 StarGAN 网络训练的一个约束，使在图像生成过程中，更多的年龄信息被关注。

在年龄估计阶段，如图 3 所示，原图 x 与年龄标注 c 原图 x 和标注 c 作为输入送到固定网络参数的生成器 G 进行图像的重建，生成与原图一致的图片 x' ；图片 x' 作为扩增数据与原图 x 作为输入一同送至 SSR-Net 进行年龄估计的训练。在这个过程中，用于年龄估计的数据集直接被扩充为原来的二倍，且增加的数据并不是基于原图的简单旋转变换，而是适用于年龄估计网络的含有隐藏年龄信息的人脸图片。

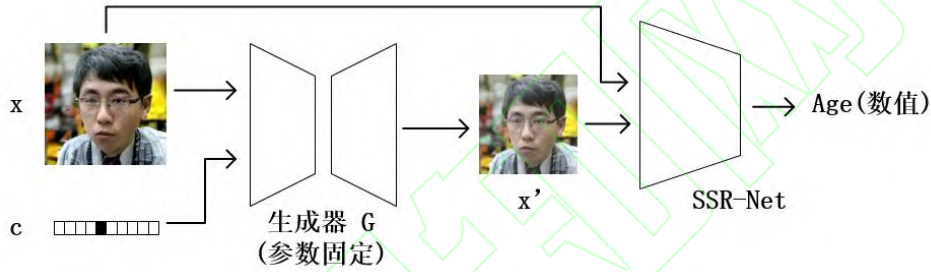


图 3 年龄估计网络结构图

Fig.3 Age estimation network structure

3.2 损失函数

为了使生成的图片与真实图片难以区分，我们采用了一种对抗损失的方法：

$$L_{adv} = E_x[\log D_{src}(x)] + E_{x,c}[\log(1 - D_{src}(G(x, c)))] \quad (1)$$

其中， $G(x, c)$ 代表输入原图 x 和目标年龄标签 c 生成器生成的图片， $D_{src}(x)$ 是鉴别器对图片是否真实的预测值，越大代表越真实。生成器试图令 L_{adv} 最小，而鉴别器试图令 L_{adv} 最大，整个网络的训练是一个博弈的过程。

对于一个给定的输入图片和一个目标年龄域标签，我们的目标是将输入图片转化为输出图片并正确分类到目标域。为了达到这个目标，我们在鉴别器上添加一个输出用于判断图片的年龄，用于约束整个训练过程。因此，训练的目标分为了两个，优化生成器和鉴别器。利用真实图片优化鉴别器，分类损失函数定义如下：

$$L_{cls}^r = E_{x,c}[-\log D_{cls}(c|x)] \quad (2)$$

其中， c 代表输入图片的原年龄域， $D_{cls}(c|x)$ 代表鉴别器 D 预测的年龄域标签的概率分布。最小化 L_{cls}^r ，鉴别器就学习到了如何正确分类一个真实的图片的年龄域。另一方面，利用生成图片优化生成器的分类损失函数定义如下：

$$L_{cls}^f = E_{x,c'}[-\log D_{cls}(c'|G(x,c'))] \quad (3)$$

其中, c' 代表要转换到的年龄目标域, $G(x,c')$ 代表输入原图 x 和目标年龄标签 c' 生成器生成的图片。最小化 L_{cls}^f , 生成器就可以生成能被正确分类到目标年龄域的图片。

通过最小化对抗性损失 L_{adv} 和分类损失 L_{cls}^r 、 L_{cls}^f , 可以使得生成器在训练时生成真实的符合目标年龄域的人脸图片, 然而, 这并不保证转换后的图片保留其输入图片的内容, 而只改变与目标年龄域相关的部分。为了缓解这个问题, 本工作对生成器应用一个循环一致性损失函数^[15], 定义为:

$$L_{rec} = E_{x,c,c'}[\|x - G(G(x,c'),c)\|_1] \quad (4)$$

其中, G 取转换后的图片 $G(x,c')$ 和原始域标签 c 作为输入, 并尝试重建原始图片 x 。我们采用 L1 范数作为重建损失。值得注意的是, 这个过程我们使用了两次生成器, 首先将原始图片转换为目标年龄域中的图片, 然后利用转换后的图片重建原始图片。

为了保证图片转换前后人物的身份不变, 另外增加了一个身份信息分类损失函数, 我们使用与尺度相关的 Softmax 函数进行身份信息的判断, 定义为:

$$L_{id} = -\log \frac{\exp(D_{id}(x)_i)}{\sum_j \exp(D_{id}(x)_j)} \quad (5)$$

另外, SSR-Net 作为另一鉴别器对整个训练过程也起到约束作用, 将图片输入 SSR-Net 中得到的年龄预测值与原年龄值比对, 其年龄损失函数定义为:

$$L_{age} = y - \sum_{k=1}^K \vec{p}^{(k)} \cdot \vec{\mu}^{(k)} = y - \sum_{k=1}^K \sum_{i=0}^{s_k-1} p_i^{(k)} \cdot i \left(\frac{V}{\prod_{j=1}^k s_j} \right) \quad (6)$$

$\vec{p}^{(k)}$ 、 $\vec{\mu}^{(k)}$ 都是向量, $\vec{p}^{(k)}$ 代表图片属于每个 bin (年龄区间) 的可能性, $\vec{\mu}^{(k)}$ 代表第 k 段 bin 所代表的年龄, 每一个 bin 的宽度为 $\frac{V}{\prod_{j=1}^k s_j}$, i 是 bin 的索引。

综上所述, 整个网络的优化函数为:

$$L_D = -L_{adv} + \lambda_{cls} L_{cls}^r + \lambda_{id} L_{id} \quad (7)$$

$$L_G = L_{adv} + \lambda_{cls} L_{cls}^f + \lambda_{rec} L_{rec} + \lambda_{id} L_{id} + \lambda_{age} L_{age} \quad (8)$$

$$L_{SSR-Net} = L_{age} \quad (9)$$

这里 λ_{cls} 、 λ_{rec} 、 λ_{id} 、 λ_{age} 均是控制各损失函数的超参数, 在实验中, 我们设置 $\lambda_{cls} = 5$ 、 $\lambda_{rec} = 10$ 、 $\lambda_{id} = 0.1$ 、 $\lambda_{age} = 1$ 。

3.3 训练策略

参考 YinXi^[21] 分段训练的思想, 整个联合训练过程分为两个过程: 图像生成和年龄估计。

在第一阶段图像生成的时候，SSR-Net 网络参数保持固定，充当另一鉴别器的作用，只改变 StarGAN 生成器和鉴别器的参数，做对抗学习；在第二阶段年龄估计时，StarGAN 生成器参数保持固定，利用其生成的图片作为扩增数据和原图训练 SSR-Net，做年龄估计，此时鉴别器处于失活状态。我们的两阶段交替训练过程的概述如图 4 所示。

Algorithm 1: Alternating training scheme for Age estimation

Stage 1: Images generated

```

Fix SSR-Net, train G and D using Eqn. 7 and Eqn. 8
C = CreatLabel()
For i=1, ..., Niter1 do
    | train 1st batch through G(x, c), without Lrec
    | train 2nd batch through G(x, C) and G(G(x, C), c)

```

Stage 2: Age estimation

```

Fix G, train SSR-Net
x' = G(x, c)
X = G(G(x, C), c)
For i=1, ..., Niter2 do
    | random samples from x, x' and X
    | train SSR-net using Eqn. 6

```

Function C = CreatLabel()

```

Init C = zero(numpy(Nage, Nage))
For i=0, ..., Nage-1 do
    c(i, i) = 1
Return C

```

alternate stage 1 and 2 until convergence

图 4 整体算法

Fig.4 Overall Algorithm

4 实验

我们在 Adience、MegaAge-Asian、Morph2 以及 FG-NET 上进行测试，分别验证本工作提出的方法在非受限条件和受限条件下的效果。所有网络模型均使用 Adam 优化^[11]，配置为： $\beta_1 = 0.5$ ， $\beta_2 = 0.999$ 。在训练过程中为了增加数据，我们以 0.5 的概率水平翻转图片。参照^[20]的做法，我们在五个鉴别器更新之后执行一个生成器更新，每次送入网络的 batch=32。

在训练阶段具体参数设置为：在步骤一图片生成阶段，固定 SSR-Net 网络参数，更新生成器和鉴别器的参数，设置 lr=0.001（Learning rate，学习速率），进行 10000 次迭代。在步骤二年龄估计阶段，固定生成器和鉴别器的参数，更新 SSR-Net 的参数，设置 lr=0.0001，进行 5000 次迭代。根据 L_{age} 的收敛情况决定重复步骤一、步骤二的次数。

4.1 年龄估计

测试在受限条件下年龄估计的准确率，评价指标为 MAE (mean absolute error, 平均绝对误差)，表示测试图片的年龄估计值和年龄真值的平均误差。MAE 越小，年龄估计算法的性能越好。结果如表 1 所示，现有的年龄估计算法在网络复杂度上分为两类，越复杂的网络，其性能表现越好。可以看到，我们的方法在不增加网络空间复杂度的同时，对 SSR-Net 在准确率方面获得了提升，同时估计时间也和原网络时间一致，唯一的不同仅体现网络的训练过程中，增加了年龄编辑部分的耗时。

对于非限制场景下的 Adience 数据集，由于判断具体年龄值较为困难，而判断一个大致的年龄区间较为简单，因此我们采取单一分类准确率和 1-off 分类准确率的评教指标^[22]。具体为：

$$A_{cc} = \frac{Num_{acc}}{Num_{test}} \quad (10)$$

$$A_{CC1-off} = \frac{Num_{1-off}}{Num_{test}} \quad (11)$$

式中， A_{cc} 和 $A_{CC1-off}$ 分别为单一分类准确率和 1-off 分类准确率， Num_{acc} ， Num_{1-off} ，和 Num_{test} 分别表示预测为正确类别图片数、预测为正确或相邻类别图片数和测试集图片总数。与单一分类准确率不同，1-off 准确率将预测类别与实际类别相邻的两类一致也判别为正确。测试结果如表 2 所示，我们的方法在非受限条件的数据集上表现优秀，原因在于在 StarGAN 进行人脸图像重建过程中增加年龄的约束，使得原图中无关年龄的信息被剔除，整个图像生成网络更多专注于年龄相关特征的恢复。

表 1 年龄估计方法在受限条件数据集上的结果

Table 1 Results of age estimation methods on constrained data sets

Type	Methods	MAE (MORPH2)	MAE (FG-Net)	Memory	Parameters
Compact	OR-CNN	3.27	---	1.7MB	479.7K
	Dense-Net	5.05	---	1.1MB	242.0K
	Mobile-Net-V1	6.5	---	1.0MB	226.3K
	SSR-Net	3.16	4.78	0.32MB	40.9K
	C3AE	2.78	4.09	0.25MB	39.7K
	C3AE (IMDB-WIKI)	2.75	2.95	0.25MB	39.7K
Bulky	Ranking-CNN	2.96	---	2.2GB	500M
	DEX	3.25	4.63	530MB	138M
	DEX (IMDB-WIKI)	2.68	3.09	530MB	138M
	AP	2.52	---	530MB	138M
	MV	2.41	4.1	530MB	138M
	MV (IMDB-WIKI)	2.16	2.68	530MB	138M
	ours	2.98	4.23	0.32MB	40.9K
	ours (IMDB-WIKI)	2.79	3.15	0.32MB	40.9K

表 2 年龄估计方法在 Adience 数据集上的结果

Table 2 Results of age estimation method on Adience data set

方法	准确率/%	1-off准确率/%
3-conv_2-fc+softmax ^[5]	50.7 ± 5.1	84.7 ± 2.2
Chained net	54.5	84.1
VGG+EMD2 loss	61.1	94
FFAM	61.8 ± 2.1	95.1 ± 0.03
VGG+EMD loss	62.2	94.3
DEX ^[6]	64.0 ± 4.2	96.6 ± 0.9
RoR ^[23]	67.34 ± 5.1	97.51 ± 0.67
AL ^[24]	67.83 ± 2.98	97.53 ± 0.59
SSR-Net	62.7	95.1

ours	68.3 ± 1.3	98.77 ± 0.8
------	----------------	-----------------

在非受限条件 MegaAge-Asian 数据集上进行的实验结果如表 3 所示，我们采用允许预测年龄与实际年龄误差在 ± 3 、 ± 5 的准确率作为评测指标，结果显示，利用 StarGAN 生成的图片扩充数据集的方法，可以获得年龄估计准确率的进一步提升，尤其是在非受限条件的情况下。同时为了验证本方法中各个部件的效果，我们进行了消融实验。在训练图像生成时去掉年龄估计网络 SSR-Net，使得网络生成的图片不能用于年龄估计，反而还会降低年龄估计准确率；同时，去掉人脸身份信息的约束，年龄估计的准确率也会下降，证明我们的方法提出的每一种约束在年龄估计任务中都是有效的。

表 3 年龄估计方法在 MegaAge-Asian 数据集上的结果

Table 3 Ablation experiments on MegaAge-Asian data sets

	MegaAge-Asian (3)	MegaAge-Asian (5)
SSR-Net	0.549	0.741
w/o SSR-Net	0.433	0.68
w/o L_{id}	0.521	0.715
ours	0.562	0.779

将年龄编辑的范围控制为在 0、 ± 1 、 ± 2 和 ± 3 个区间，以 2、4、6、8 倍扩充原数据集分别进行年龄估计准确率的实验，结果如表 4 所示，当编辑范围为 1 个区间，以 4 倍扩充原数据时达到了最好的实验结果，在 Adience 数据集上年龄估计准确率提升 1.3%，MegaAge-Asian 上有 3.8% 的提升。随着年龄编辑范围的扩大，年龄估计准确率反而下降，因此并不是用于训练的图片数据越多越好。我们猜测原因是经过年龄编辑之后的人脸图片仍是不真实的，增加大量的错误训练数据反而会降低网络的识别精度，当跨度较小时我们的方法可以有效提升年龄估计准确率，但年龄跨度过大时，我们的方法不再适用。

表 4 不同程度年龄编辑实验结果

Table 4 Experiment results were edited at different ages

	年龄估计网络训练集	Adience 准确率	MegaAge-Asian (3)	MegaAge-Asian (5)
SSR-Net	原数据	0.627	0.549	0.741
Ours	+重建数据	0.683	0.558	0.770
	+重建数据+跨 1 区间年龄编辑数据	0.696	0.562	0.779
	+重建数据+跨 2 区间年龄编辑数据	0.652	0.527	0.753
	+重建数据+跨 3 区间年龄编辑数据	0.610	0.501	0.728

4.2 年龄编辑

我们在 MegaAge-Asian 数据集上进行年龄编辑的可视化效果验证，将原图片与年龄标注矩阵 $C(c_0, c_1, \dots, c_8)$ 一同送入训练好的 StarGAN 中，结果如图 5 所示，第 1 列为原图片，后面 9 列为年龄编辑结果。如图 5 中第二行显示，我们的方法可以应对较为复杂的背景。对于年龄估计任务，一致的看法是并不需要分辨率特别高的图片，因此对于年龄编辑的效果，我们不再做优化。

从图片的年龄编辑效果可以看出,我们的方法在进行年龄编辑时只是对原输入图像增加老化或者幼化的人脸特征,本质上还是原输入图片。当我们进行区间跨越过大的年龄编辑时,得到的人脸图片与现实中位于该区间的真实人脸并不一致,导致年龄估计的准确率反而下降的实验结果;在年龄区间跨越度不大时,比如 0 或者 ± 1 个区间,得到的人脸图片与现实中位于该区间的真实人脸基本一致,因此我们的方法在判定人脸图片年龄时是有效的。



图 5 在 MegaAge-Asian 数据集上的年龄编辑结果
Fig.5 Age editing results on the MegaAge-Asian dataset

5 结束语

本工作提出了一种利用年龄编辑改进年龄估计的方法,深入研究了在图片转换领域表现优秀的 StarGAN 在年龄估计中的作用,根据年龄值的零散性和不易区分性,将年龄按区间分类,形成 one-hot 形式的年龄标注进行人脸图像的年龄编辑工作,可以生成任意阶段的年轻人脸年龄图片。同时年龄估计网络 SSR-Net 作为另一鉴别器对于 StarGAN 在生成人脸图片时具有指导意义,使得生成的人脸图片拥有了更多隐含的年龄信息。原图与经过 StarGAN 生成的扩增数据共同作为训练集去训练 SSR-Net,提高了年龄估计的准确率。

实验证明,本方法对于受限条件下的年龄估计准确率提升有限,但对于非受限条件下的年龄估计提升明显,在 Adience 上达到了目前最好的结果。本工作的年龄估计网络 SSR-Net 结构参数少,训练好之后易于部署。但总体来说,年龄估计的准确率相比人脸识别准确率还是比较低,在实际应用中仍存在许多问题。在今后的研究中,如何在不对大量数据采集和标注的前提下实现数据扩增,获得更高的年龄估计准确率,是未来人脸年龄估计研究的重要方向之一。

参考文献:

- [1] Adeel Yousaf, Muhammad Junaid Khan, Muhammad Jaleed Khan, Adil M. Siddiqui, Khurram Khurshid. A robust and efficient convolutional deep learning framework for age - invariant face recognition[J]. Expert Systems, 2020, 37(3).
- [2] Chen Y , Zhou X S , Huang T S . One-class SVM for learning in image retrieval[C]// International Conference on Image Processing. IEEE, 2001.
- [3] Zhu S C , Yuille A L . Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1996, 18(9):884-900.
- [4] Zhang K, Wang X S, Guo Y R, Su Y K, He Y X. Survey of deep learning methods for face age estimation [J]. Journal of Image and Graphics, 2019, 24(08): 1215-1230. [张珂, 王新胜, 郭玉荣, 苏昱坤, 何颖宣. 人脸年龄估计的深度学习方法综述 [J]. 中国图象图形学报, 2019, 24(08): 1215-1230.] [DOI: 10. 11834/jig. 180653]
- [5] Gil Levi and Tal Hassner. Age and Gender Classification Using Convolutional Neural Networks. IEEE Workshop on Analysis and Modeling of Faces and Gestures (AMFG), at the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, 2015.
- [6] Rothe R , Timofte R , Van Gool L . Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks[J]. International Journal of Computer Vision, 2016.
- [7] Florian Schroff, Dmitry Kalenichenko, James Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering. at the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, 2015.
- [8] Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, and Gang Hua. Ordinal regression with multiple output cnn for age estimation. In CVPR, 2016.
- [9] Zhang Y , Liu L , Li C , et al. Quantifying Facial Age by Posterior of Age Comparisons[J]. 2017.
- [10] Zhang Z , Song Y , Qi H . Age Progression/Regression by Conditional Adversarial Autoencoder[J]. 2017.
- [11] Tsun-Yi Yang, Yi-Hsuan Huang, Yen-Yu Lin, Pi-Cheng Hsiu, and Yung-Yu Chuang. SSR-Net: A compact soft stagewise regression network for age estimation. In IJCAI, 2018.
- [12] Zhang C, Liu S, Xu X, et al. C3AE: Exploring the Limits of Compact Model for Age Estimation[J]. 2019.
- [13] Mirza M, Osindero S. Conditional Generative Adversarial Nets[J]. Computer Science, 2014:2672-2680.
- [14] Isola P, Zhu J Y, Zhou T, et al. Image-to-Image Translation with Conditional Adversarial Networks[J]. 2016.
- [15] Zhu J Y , Park T , Isola P , et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks[J]. 2017.
- [16] Karras T, Aila T, Laine S, et al. Progressive Growing of GANs for Improved Quality, Stability, and Variation[J]. 2017.
- [17] Liu Z , Luo P , Wang X , et al. Deep Learning Face Attributes in the Wild[J]. 2014.
- [18] Choi, Yunje, Choi, Minje, Kim, Munyoung, et al. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation[J].
- [19] D. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
- [20] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved training of wasserstein gans. arXiv preprint arXiv:1704.00028, 2017.
- [21] Yin, X., Yu, X., Sohn, K., Liu, X. M., Chandraker, M.: Feature Transfer Learning for Deep Face Recognition with Long-Tail Data. CVPR. 2019.
- [22] Eidinger E, Enbar R , Hassner T. Age and gender estimation of unfiltered faces[J] . IEEE Transactions on Information Forensics and Security, 2014, 9(12): 2170-2179. [DOI: 10. 1109/TIFS. 2014. 2359646]
- [23] Zhang K, Gao C, Guo L R, et al. Age group and gender estimation in the wild with deep R o R architecture [J] . IEEE Access, 2017, 5: 22492-22503. [DOI: 10. 1109/ACCESS. 2017. 2761849]

- [24] Zhang K, Liu N, Yuan XF, et al. Fine-grained age group classification in the wild [C] //Proceedings of the 24th International Conference on Pattern Recognition. Beijing, China: IEEE, 2018. [DOI: 10. 1109 /ICPR. 2018. 8545333]
- [25] Karras, Tero, Laine, Samuli, Aila, Timo. A Style-Based Generator Architecture for Generative Adversarial Networks[J].

