CCF 会议人工智能会议论文阅读报告

学号：21851099

姓名：汤 凯 凯

论文选择：Learning Face Age Progression: A Pyramid Architecture of GANs

论文领域：Face Age Progression

发表会议：2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018. IEEE Computer Society 20

一． 背景与问题描述

1.1 背景介绍

本文的任务是 Age Progression，它可以改变给定人脸图像的年龄，可以应用在娱乐，跟踪失踪人员等等方面上。现有一些解决方案包括建模面部变化，使用数据建立模型等。

其中，使用深度生成网络的方法在灵活性和效果上有良好的性能，但也面临一些问题，包括生成的图片的身份持久性（identity permanence），图片的详细性能以及时间序列数据的收集。大多数现有方案优先考虑年龄信息和身份特征信息即 identity 信息，导致在生成不同年龄时不能很好地保存 identity。

本文提出了一个基于 GAN（生成式对抗网络）的模型，并致力于解决以上问题。

1.2 主要贡献点

本文的主要贡献是：

1. 解决了生成图像年龄准确性和 identity 信息保留的问题。
2. 在实验过程中，注意到保留脸部的前额和头发可以增强实验效果，因此使用完整的而不是裁剪的图片作为数据集。
3. 在各种数据集上取得了良好的效果，并且在遮挡和化妆的情况下表现出强大的稳健性。

1.3 两个亮点

文章方法的两个亮点：

1. 多 loss，GAN 网络的 loss+ identity 保留的 loss+pixel 的 loss，
2. Discriminator 的多尺度提取特征，也就是文章标题的金字塔结构，特征提取器的跨级并联结构。

1.4 实验介绍

这篇文章的实验非常充分，对于年龄的准确性，identity permanence 程度以及金字塔结构的贡献都做了大量的实验，所用的数据集为 fg-net，morph 和 CACD.

二． 算法简介

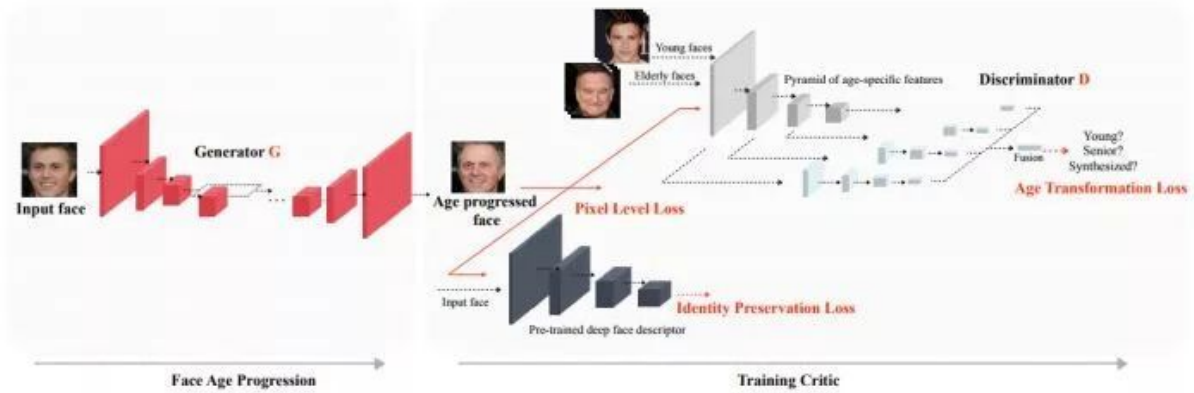

Figure 2. Framework of the proposed age progression method. A CNN based generator *G* learns the age transformation. The training critic incorporates the squared Euclidean loss in the image space, the GAN loss that encourages generated faces to be indistinguishable from the training elderly faces in terms of age, and the identity preservation loss minimizing the input-output distance in a high-level feature representation which embeds the personalized characteristics.

2.1 算法思路

本文通过 vgg16 用于提取年龄特征和 deep face descriptor 用于提取 identity，提出了一个基于 GAN（生成式对抗网络）的模型，分别模拟 identity 和年龄特征带来的面部变化相对于经过时间的约束，确保生成的面部呈现所需的老化效果，同时保持 identity 稳定。 此外，为了产生更逼真的面部细节，由合成面部传达的高级年龄特征通过多尺度的金字塔形对抗性 Discriminator 来估计，其以更精细的方式模拟老化效应。

2.1.1 AcGans 介绍

作为 IP 模式识别的 CNN 的初始模型作为简单的判别模式识别而存在，并且作为基本模型被扩展到各个方向。基本功能是图像判别模型，此后基于 Loc+CNN 的检测模型-分离式、end2end、以及 MaskCNN 模型，而后出现基于 CNN 的预测模型-AcGans [1]。

CNN 作为一个基本判别式模型简化为数学模型依然为一个函数映射 f(x)->y; 基于 CNN 的检测模型数学模型为 L(x)+f(x)->y,其中 L(x)依然为判别式，给出 loc 信息，二维的为(y1,y2)点对； 基于 CNN 的 Mask 给出每个 Pixel 的类别信息，数学模型可以简化为 k(x).f(x)—k(x).y，其中 K(x)为一个与点位置线性相关的函数；

到了 AcGans, 例如基于年龄的预测，CNN 为其组成部分之一，而生成式为主要目的服务，数学模型可以简化为 g(f0(f2)*f2(x))—y，把一个判别式 f(x)分离为维持不变性的 f0(x)和用于分离的 f2(x)，其中 f0(x)满足生成式约束不变性， f2(x)满足特征提取-数据输入不变性约束，以满足使用数据完成训练生成模型所要求，以及处理输入的特征提取模型。

通过训练的模型，数据流为 f2(x)*X—>f2(X), 通过特征提取函数，生成纹理特征；f0*f2(X)—>f0(f2(X))

2.1.2 Generator

Generator 是目前 CNN 的常规配置。是一种先 encode 再 decode 的一个 U 型结构，通过三层卷积层，四层残差，和三层转置卷积，每一层卷积层后都接一个 BN 和 ReLU，整个网络都没有 pooling，而是大小为 3，步长为 2 的卷积层下采样。

Generator 的 loss 由三部分组成：

LG = λaLGAN G + λpLpixel + λiLidentity

Loss 部分作为亮点在第二小节中重点分析。

2.1.3 Discriminator

在本文中提出了特征提取器来提出特定的特征，因为作者认为同一年龄组的不同人脸有相同的特定信息即 identity，并且提取器提取这些特征。此外，分类器由年龄分类任务预先训练。

Discriminator 中有两个负样本，分别是生成的图片和真实的老化图片，正样本是年轻的图片，也就是说 Discriminator 是一个三分类器，这比原先的一个正样本，一个负样本性能要提升不少。

Discriminator 的结构是金字塔形，首先它的主体是一个预训练的 vgg16，接着它从 2，4，7，10 四层中提取 feature map，并分别经过不同的卷积层得到分类结果，拼接起来得到 12*3 的最终结果和 label 进行比较。这样一种多尺度的提取特征方式使得每个尺度都可以选择自己着重观察的特征，fusion 的思想使结果更为准确。

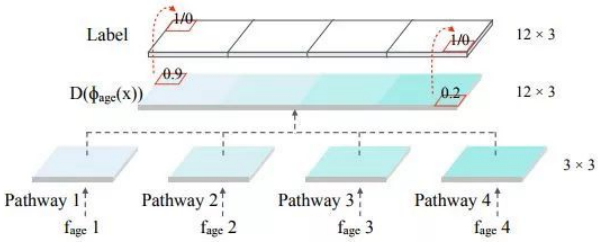

Figure 3. The scores of 4 pathways are finally concatenated and jointly estimated by the discriminator D (D is an estimator rather than a classifier; the *Label* does not need to be a single scalar).

作者考虑到了 low-level 和 high-level 将第 2 层、第 4 层、和第 7 层等信息合并作为 d 的输入。

identity 信息的保留和上一个 extractor 类似，在人脸分类数据集上预训练，然后拿来直接当 extractor。

$$\mathcal{L}_{pixel} = \frac{1}{W \times H \times C} \|G(x) - x\|_2^2 \tag{6}$$

where x denotes the input face and W, H, and C correspond to the image shape.

Finally, the system training loss can be written as:

$$\mathcal{L}_G = \lambda_a \mathcal{L}_{GAN\_G} + \lambda_p \mathcal{L}_{pixel} + \lambda_i \mathcal{L}_{identity} \tag{7}$$

$$\mathcal{L}_D = \mathcal{L}_{GAN\_D} \tag{8}$$

We train G and D alternately until optimality, and finally G learns the desired age transformation and D becomes a reliable estimator.

独立训练最优 G 和 D，然后得到 G 学习到一个年龄变换，D 作为一个可靠的分类函子

2.2 Loss 部分介绍

Generator 的 loss 由三部分组成：

LG = λaLGAN G + λpLpixel + λiLidentity

2.2.1 Identity preservation Loss

这里提到了条件 GAN，即生成的图片是依赖于输入的，实际上侧重于对条件分布进行建模，但在训练过程中是不知道这个条件分布的，因为使用的数据集不要求包含每个人的人脸随年龄变化的序列。那么，如何保证能够学到这样的条件分布呢，以前提出的方法有很多，比如 Cycle-GAN 等中使用的循环一致性，在本文中使用的是 Identity preservation Loss，是一种基于编码一致性的方法，需要找到一个特征空间并减小在这个特征空间上输入和输出的距离，而这个特征空间需要对身份变化（identity change）敏感而对其它变化（如 Age，Background）不敏感。作者通过在一个包含同一个体在不同场景下的图片上的数据集预训练一个深度网络得到 deep face descriptor——一个提取人脸特征表现很好的网络，对原图和生成后图片的人脸特征计算欧几里得距离。

2.2.2 GAN Loss

GAN 的部分并没有采取常规的 log 形式，而是用 Least Square 的形式。 在 G 和 DiscriminatorD 之间使用了一个在 age estimation 任务上预训练的 VGG-16 来连接，并使用了基于金字塔的网络架构。

后续实验也说明了金字塔结构通过融合 high-level feature 到 pixel-level feature 对生成图片的效果的影响。在 DiscriminatorD 的训练过程中，Young face, Elderly face, Synthesized face 均送入 DiscriminatorD，其中在 D 的梯度更新阶段，Elderly face 给与 real label，其它为 fake label。由于 Discriminator 相当于对不同年龄的人脸进行分类，所以在 age estimation 任务上预训练有助于进一步的正则化以提高模型的稳定性。

2.2.3 Pixel Level Loss

pixel 部分比较普通，就是每个像素的平均欧几里得距离。

为了进一步保证输入和输出图片只有年龄上的变化，本文还使用了输入和输出图片在原始空间上的 L2 距离进行进一步的约束。

三. 思考与扩展建议

3.1 思考

本文的亮点是多 loss，多尺度，灵活借鉴了前人的预训练网络：vgg16 用于提取年龄特征和 deep face descriptor 用于提取 identity。采用的 GAN 网络可以看作 Least Square GAN 和 DCGAN 的结合。

在本文中已经进行了不少实验来证明性能优越性。 通过年龄聚类获得三个不同年龄组的年龄组，并训练他们改变年龄。 输入面部图像的年龄小于这三个年龄组。 除了直接显示结果外，实验还分析了模型对非年龄特征的保留效果，如姿势，眼镜，头发的年龄变化，生成的面部图像的年龄估计与 GAN 学习分布的比较，以及其他方法比较等等。

3.2 扩展建议

Generator 中 loss 的 GAN 部分采取用 Least Square 的形式，还可以尝试 Wasserstain，进行比较实验结果后进行取舍。

参考

[1] wishchin：https://www.cnblogs.com/wishchin/

[2] 李光睿: https://zhuanlan.zhihu.com/p/35661176

[3] CVPR 2018 值得一看的 25 篇论文: http://www.sohu.com/a/229526356_500659