基于生成式对抗神经网络的素描路径生成系统

张么元 龚敬洋

关键字: 人工神经网络;深度学习;生成式对抗网络;绘画风格转移

1.选题背景

现在网路上风靡的相机滤镜都有素描画的图片滤镜。同时在执法和刑事案件中嫌疑人的素描画任然作为目击者提供线索的重要依据。在娱乐和社会安全两方面都有对素描画像的需要。但人工素描价格昂贵，生产效率低，而普通基于图片简单滤镜变换的素描画真实度不高，通常伴有不合适的线条或低清晰度，同时笔画线条不清楚。

最近GAN(Generative Adversarial Networks)[1] 和CNN(convolutional neural network)等神经网络技术的发展，直接产生矢量素描笔画路径的机器算法成为可能。本文将介绍我们利用GAN技术提出的素描路径生成系统的算法和框架，并进行详细的实验证明和与普通滤镜素描生成器和人工绘画的结果比较和改进方案。

2.相关工作

基于神经网络算法的发展和CUDA(Compute Unified Device Architecture)显卡计算技术的发展，原先许多必须由人工生成图像合成算法现在可以通过神经网络得出我们想要的模型。而2014年Ian Goodfellow等人提出的GAN(Generative Adversarial Networks)进一步实现了神经网络在图像处理上的进一步发展。

3.我们的方法

想象阿尔布雷 希特·丢勒(Albrecht Dürer ,1471－1528)大师在文艺复兴时期面对铜镜中的自己画出了世界上第一幅自画像，他的铅笔线条柔和流畅明暗中的人栩栩如生，如果丢勒来到现在的城市公园，为富有年轻活力的少女画画像会怎么样?

1.回归素描绘画过程

普通的机器滤镜实现的主要原理都是由基于图片灰度特征或者卷积计算后得到特征，而人在绘画中恰恰不会一开始就从整体的色调或者从整体描画出图像。通过模仿人类艺术创作的过程实现机器进行艺术创作的效果会明显比简单机器滤镜的效果要好很多，Combining Sketch and Tone for Pencil Drawing Production[2]通过绘画主体的描绘和背景的描绘拆分实现了机器素描绘画的最高效果，所以我们回归到人类人像的素描绘画过程中，实现更好的结果。

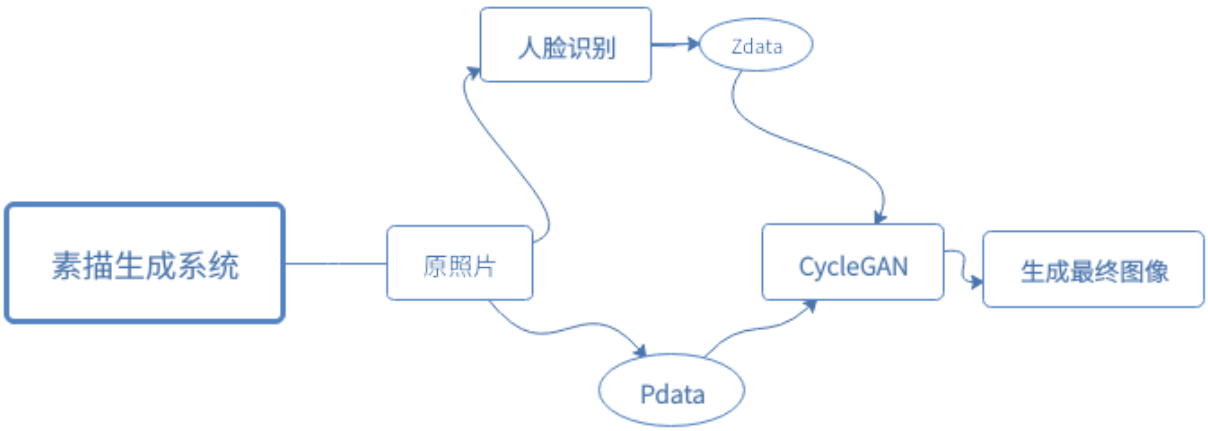
我们考虑一位素描艺术家看到了一张人脸(感谢陈同学的出镜)

陈同学

他会考虑人物的脸型轮廓，五官位置和形状，头发式样，光线强弱.....所以我们就需要根据图片提取这些特征。

3.模型建立

1.模型结构



本节将详细阐述解决方法，并提供详细的模型方案。主要模型分为人脸信息提取和神经网络模型的设计与搭建。首先由原图像提取人脸信息，随后将人脸信息和图像一起作为数据集输入模型，最后输出素描路径。

2.公式

给定由$(A_i, B_i)_{i=1}^{N}$表示原图像和草图的数据集。我们的目的是让模型学习两个函数$B' = f_{p\to s}(A)$和$A' = f_{s\to p}(B)$代表照片到素描的生成器和素描到照片的生成器。我们考虑这是一个图像到图像的翻译工作，我们使用CycleGAN[此处有论文]，假设两个网络F和G两个神经网络生成器分别代表代表照片到素描的生成器和素描到照片的生成器。F以真人图像R_A输入，生成J_B的素描路径；G以素描路径J_B输入，生成R'_A的真人图像。

所以图像转化为素描的过程可以表示为：$J_B = F(R_A), R'_A = G(J_B)$

3.人脸特征信息提取

1、人脸识别API的选择

将肖像照片转化为素描照片，首先需要提取照片中人脸中相关特征点的位置，以方便后续的GAN网络生成素描画。目前互联网上提供了大量已训练成熟的基于CNN(convolutional neural network)的面部识别API可供调用，但它们在响应时间，面部关键点数量和调用流量计费方式上均有差异。通过对国内三款主流面部识别服务供应商提供的API进行大量测试，基本可以得出三款API的相关差异。

API名称	平均响应时间	关键点数量	免费调用流量限制
百度云	315ms	72	2QPS
旷视FACE++	206ms	106	不限量，与其他用户共享QPS池
腾讯	294ms	88	1万张/月

由上表可知，对于小规模面部识别调用而言，FACE++[3]在关键点数量及响应时间上相比其他两款API均有明显优势。因此本文中我们将选择该API进行人脸特征信息的提取。

2、图片预处理

旷视FACE++对于上传图片有最大 4096×4096 像素，2MB文件大小的限制要求，而目前绝大多数拍摄设备拍出的图片文件参数均高于该值，同时为了减少因进行人脸识别而产生的流量，需要首先对图片进行适当压缩和s缩放。我们首先将图片进行适当锐化以保证其不因缩放导致锐度下降，进而影响识别成功率。随后通过OpenCV[4]图像压缩算法对文件体积进行适当压缩，并对图片进行等比例缩放以保证其大小和体积被控制在合理范围内。为了防止图片缩小时出现波纹，我们使用了像素关系重采样的方式(CV_INTER_AREA)对图片进行缩放。具体函数用法如下：

```
#通过CV_INTER_AREA方法缩放图片至目标大小
cv2.resize(SourceImage, (DXsize,DYsize), interpolation = cv2.INTER_AREA)
#适当降低照片质量以减小图片质量
cv2.imwrite(TargetFileName, SourceImage, [int(cv2.IMWRITE_JPEG_QUALITY),
QualityKeepValue])
```

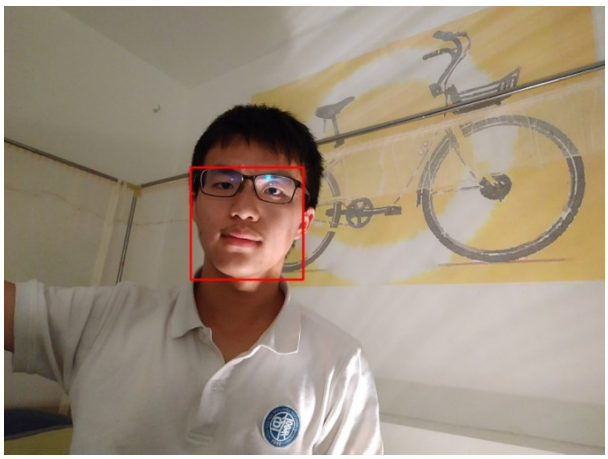
经过测试，处理后图像的文件体积已基本被控制在可接受范围内。测试数据如下：

原图片大小	原图片体积	压缩后图片大小	压缩后图片体积
4288 * 2848	3.01MB	1500 * 995	554kb
6000 * 4000	6.15MB	1500 * 1000	682kb
2048 * 2048	2.18MB	2048 * 2048	325kb

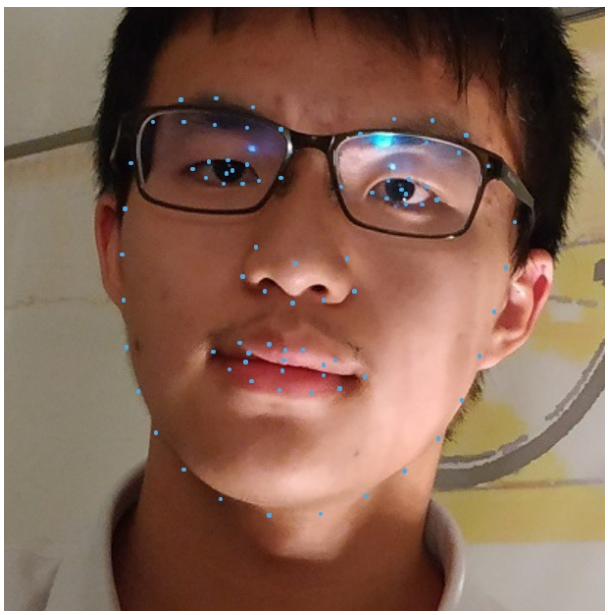
3、获得人脸特征信息

将照片进行预处理后，我们通过POST方式调用旷世FACE++的面部识别接口，并获得包含人脸特征信息的JSON数据。获得的人脸特征信息包含面部的矩形位置(face_rectangle)，面部器官的轮廓位置(landmarks)以及人脸的特征信息(attributes)。通过对JSON数据进行解析和分离，便可得到精确的人脸特征点位置信息。测试结果如下：

图3-1 原图片



(图3-2 识别结果(face_rectangle))



(图3-2 识别结果(landmarks))

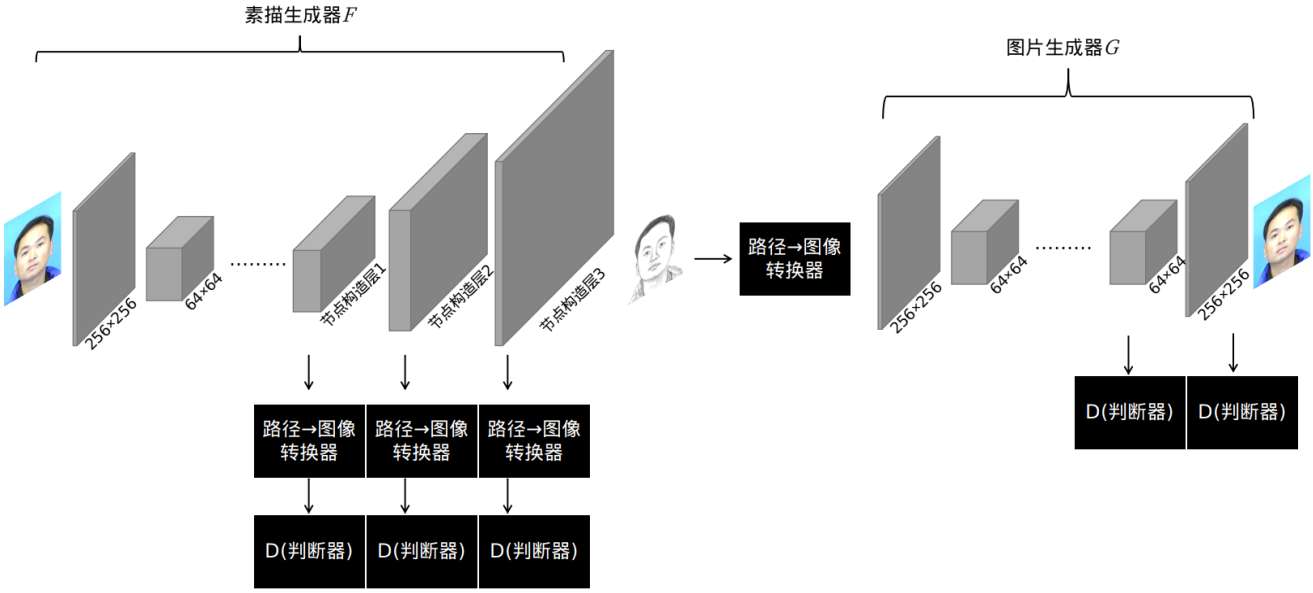
输出的数据是人物轮廓的内容和点阵位置，这就是先把握人物的整体形象。这部分作为人脸特征信息，设其为Z_i。

3.模型设计

cycleGAN[5]是一种不成对的图像到图像转换的神经网络算法，由Berkeley AI Research (BAIR) laboratory, UC Berkeley在2018年提出。 算法主要基于GAN生成式对抗网络算法。

该算法的原理可以概述为：将A风格图片转换成B风格图片[6]。也就是说，现在有两个样本空间(sample space)，X 和Y我们希望把X 空间中的样本通过cycleGAN转换成Y空间中的样本。所以我们的目的就是拟合学习一个生成器或映射，设这个映射为F，则它就对应着GAN中的生成器(Generator)，F可以将X中的样本空间x映射到Y的样本空间$F(x)$中。对于生成的图片，我们还需要GAN中的鉴别器(Discriminator)来区分它是否为理想图片，即为我们所希望要的图片，由此建立GAN(Generative Adversarial Networks)。

正如在讨论中[7], 这些伪影是由于已知的训练不稳 定性而产生的，同时产生高分辨率图像。 这些不稳定 性可能是由于自然图像分布和隐含模型分布的支持可 能在高维空间中不重叠的事实。 这个问题的严重性随 着图像分辨率的增加而增加。 因此，为了在生成逼真 图像时避免这些伪像，我们提出了一个逐级多尺度优 化框架，通过利用生成器子网络中不同分辨率的特征 映射的隐式存在。 考虑到大多数GAN框架具有与编码 器 - 解码器类型相似的生成器，其中具有一堆卷积和 最大池化层，随后是一系列解卷积层。 反卷积层将特 征映射从较低分辨率顺序上采样到较高分辨率。 来自 每个解卷积层的特征图都是通过3×3卷积层转发以生成不同分辨率的输出图像。



整个网络结构如上图所示。通过构建多重生成式对抗神经网络，在隐藏层中也添加判断器减少伪像的产生，不过最终还是需要通过像素之间的比较判断图片生成效果，所以我们使用openCV中的函数将路径(矢量图)转换成为像素图像进行比较。

所以在三个节点构造层我们有$\{J_{B1}, J_{B2}, J_{B3}\}$的输出，分别通过3个判别器$\{D_{B1}, D_{B2}, D_{B3}\}$，而$J_{B3}$作为最后一个图像将继续传递到下一个网络中，最终的产生的两张原图像$\{R'_{A1}, R'_{A2}\}$，R'_{A2}就是最终的还原人脸图像，这两张图像分别加入判别器$\{D_{A1}, D_{A2}\}$。

综合模型我们可以得出模型的损失函数:

$$L_{GAN_{A_i}} = E_{B_i \sim Pdata}[logD_{A_i}(B_i)] + E_{A_i \sim Pdata}[log(1 - D_{A_i}(F(R_A))i)]$$

$$L_{GAN_{B_i}} = E_{A_i \sim Pdata}[logD_{B_i}(A_i)] + E_{B_i \sim Pdata}[log(1 - D_{A_i}(G(R_B))i)]$$

训练的目标就是$(F(R_A))_i = F_{B_i}, (G(R_B))_j = F_{A_j}$且$i = 1, 2, 3, j = 1, 2$。为了使生成的图像更加接近我们的目标，我们需要使生成差L_{syn}最小。L_{syn}可以被定义为:

$$L_{syn_{A_i}} = ||J_{A_i} - R_{A_i}||_1 = ||G(R_B)_i - R_{A_i}||_1$$

$$L_{syn_{B_i}} = ||J_{B_i} - R_{B_i}||_1 = ||G(R_A)_i - R_{B_i}||_1$$

除了用L_{syn}减小差距，还通过在不同分辨率阶段引入L_{cyc}来实现减少了可能的映射函数的空间过多出现的伪像，其定义如下：

$$L_{cyc_{A_i}} = ||R'_{A_i} - R_{A_i}||_1 = ||G(F(R_A))_i - R_{A_i}||_1$$

$$L_{cyc_{B_i}} = ||R'_{B_i} - R_{B_i}||_1 = ||G(F(R_B))_i - R_{B_i}||_1$$

综上目标损失函数即为：

$$L(G, F, D_A, D_B) = \sum_{n=1}^{3}(L_{GAN_{A_i}} + L_{GAN_{B_i}} + \lambda L_{syn_{A_i}} + \lambda L_{syn_{B_i}} + \mu L_{cyc_{A_i}} + \mu L_{cyc_{B_i}})$$

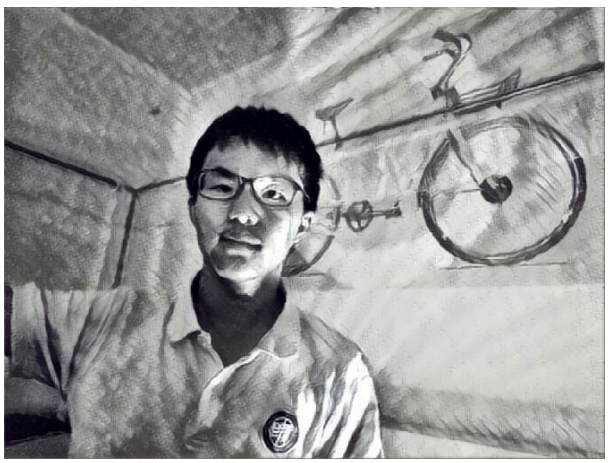
4.实验结果

所提出的方法在现有的查看草图数据集上进行评估。中大脸部素描数据库（ CUFS）[8] 是一个观看素描数据库，其中包括来自香港中文大学（ CUHK）学生数据库的188张面孔，来自AR数据库的123张面孔 [9], 以及来自XM2VTS数据库的295个人脸[10]. 对于每张脸，都有一张艺术家根据在正常照明条件下以正面姿势拍摄的照片以及中性表情绘制的草图。

经过训练最终的结果：

PnVhvT.png

最后放上陈同学经过我们的模型后产生的肖像画。



5.参考文献

[1] Ian, J, Goodfellow, Jean, Pouget-Abadie, Mehdi, Mirza, Bing, Xu, David, Warde-Farley, Sherjil, Ozair, Aaron, Courville, Yoshua, Bengio. Generative Adversarial Networks[C]. arXiv:1406.2661v1:Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, 2014.

[2] ewu, Lu, Li, Xu, Jiaya, Jia. Combining Sketch and Tone for Pencil Drawing Production[C]. The Chinese University of Hong Kong:Cewu Lu Li Xu Jiaya Jia, 2012.

[3] face++[EB/OL]. https://www.faceplusplus.com.cn/.

[4] OpenCV[EB/OL]. https://opencv.org/.

[5] Jun-Yan, Zhu, Taesung, Park, Phillip, Isola, Alexei, A, Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks[C]. arXiv:1703.10593:Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, 2018.

[6] 玄学酱. 可能是近期最好玩的深度学习模型：CycleGAN的原理与实验详解[EB/OL]. https://yq.aliyun.com/articles/229300.

[7] H. Zhang, T. Xu, H. Li, S. Zhang, X. Huang, X. Wang, and D. Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks[C]. In IEEE ICCV, 2017.

[8] X. Wang and X. Tang. Face photo-sketch synthesis and recognition. TPAMI, 31(11):1955–1967, 2009.

[9] A. Martinez and R. Benavente. The ar face database, cvc. 1998.

[10] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. Xm2vtsdb: The extended m2vts database. In Second international conference on audio and video-based biometric person authentication, volume 964, pages 965–966, 1999.