

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)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。

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？

cycleGAN(Cycle Generative Adversarial Networks)

cycleGAN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。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 映射成为 Y 中的样本空间 $F(x)$ 。对于生成的图片，我们还需要GAN中的判别器(Discriminator)来判别它是否为理想图片，即为我

们所希望要的图片，由此建立GAN(Generative Adversarial Networks)。设这个判别器为 D_Y 这样的话，根据这里的生成器和判别器，我们就可以构造一个GAN损失(loss function)，表达式为：

$$L_{GAN}(F, D_Y, X, Y) = E_{y \sim P_{data}(y)}[T(D_Y(y))] + E_{x \sim P_{data}(x)}[T(1 - D_Y(F(x)))]$$

但单纯的使用这一个损失训练是无法达到目的的。为了追求最低的生成差异，映射 F 完全可以将所有 X 都映射为 Y 空间中的同一张图片，使损失无效化，即“循环一致性损失”(cycle consistency loss)。

所以为了保证映射结果的差异性，需要再假设映射 G 将 Y 空间中的样本空间 y 转换为 X 中的样本空间 $G(y)$ 。CycleGAN同时学习 F 和 G 两个映射，并要求 $F(G(y)) \approx y$ 并且 $G(F(x)) \approx x$ 。

根据 $F(G(y)) \approx y$ $G(F(x)) \approx x$, cycleGAN的损失函数就可以定义为：

$$L_{cyc}(F, G, X, Y) = E_{x \sim P_{data}(x)}[||G(F(x)) - x||_1] + E_{y \sim P_{data}(y)}[||F(G(y)) - y||_1]$$
 同样 $G(X)$ 也需要一个D判断器来学习，可以引入其损失函数 $L_{GAN}(F, D_X, Y, X)$

综上，我们就可以完整的创造一个cycleGAN网络，它的损失函数表达式为：

$$L = L_{GAN}(F, D_Y, X, Y) + L_{GAN}(G, D_X, Y, X) + \lambda L_{cyc}(F, G, X, Y)$$

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[此处为论文]

http://www.cse.cuhk.edu.hk/~leojia/projects/pencilsketch/pencil_drawing.htm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。

考虑一位素描艺术家看到了一张人脸，(感谢陈同学的出境)



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。

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 |
| 旷世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，矿世FACE++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

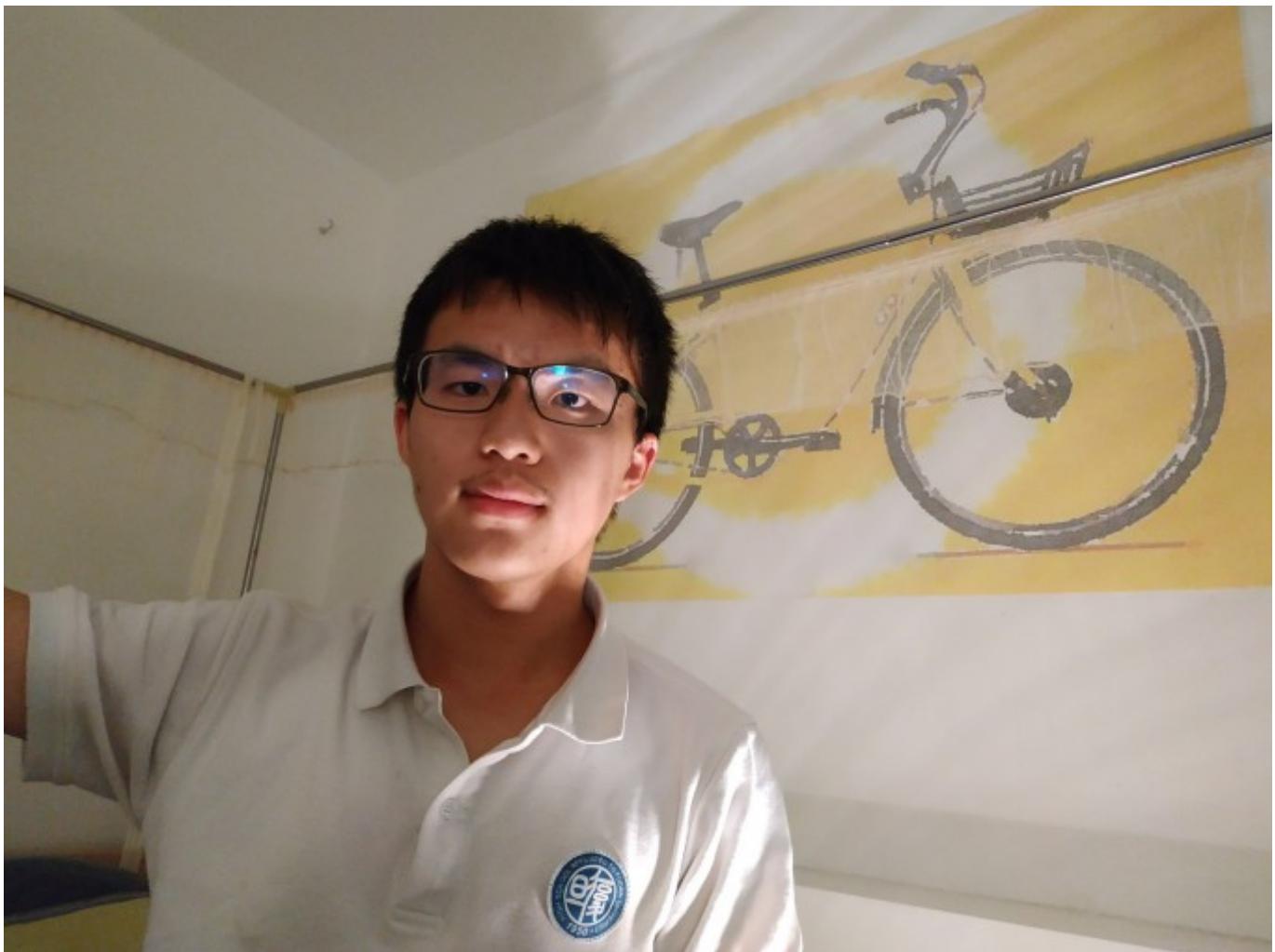
矿石FACE++对于上传图片有最大4096*4096像素，2MB文件大小的限制要求，而目前绝大多数拍摄设备拍出的图片文件参数均高于该值，同时为了减少因进行人脸识别而产生的流量，需要首先对图片进行适当压缩和缩放。我们首先将图片进行适当锐化以保证其不因缩放导致锐度下降，进而影响识别成功率。随后通过OpenCV图像压缩算法对文件体积进行适当压缩，并对图片进行等比例缩放以保证其大小和体积被控制在合理范围内。为了防止图片缩小时出现波纹，我们使用了像素关系重采样的方式(CV_INTER_AREA)对图片进行缩放。具体函数用法如下：

```
#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 |

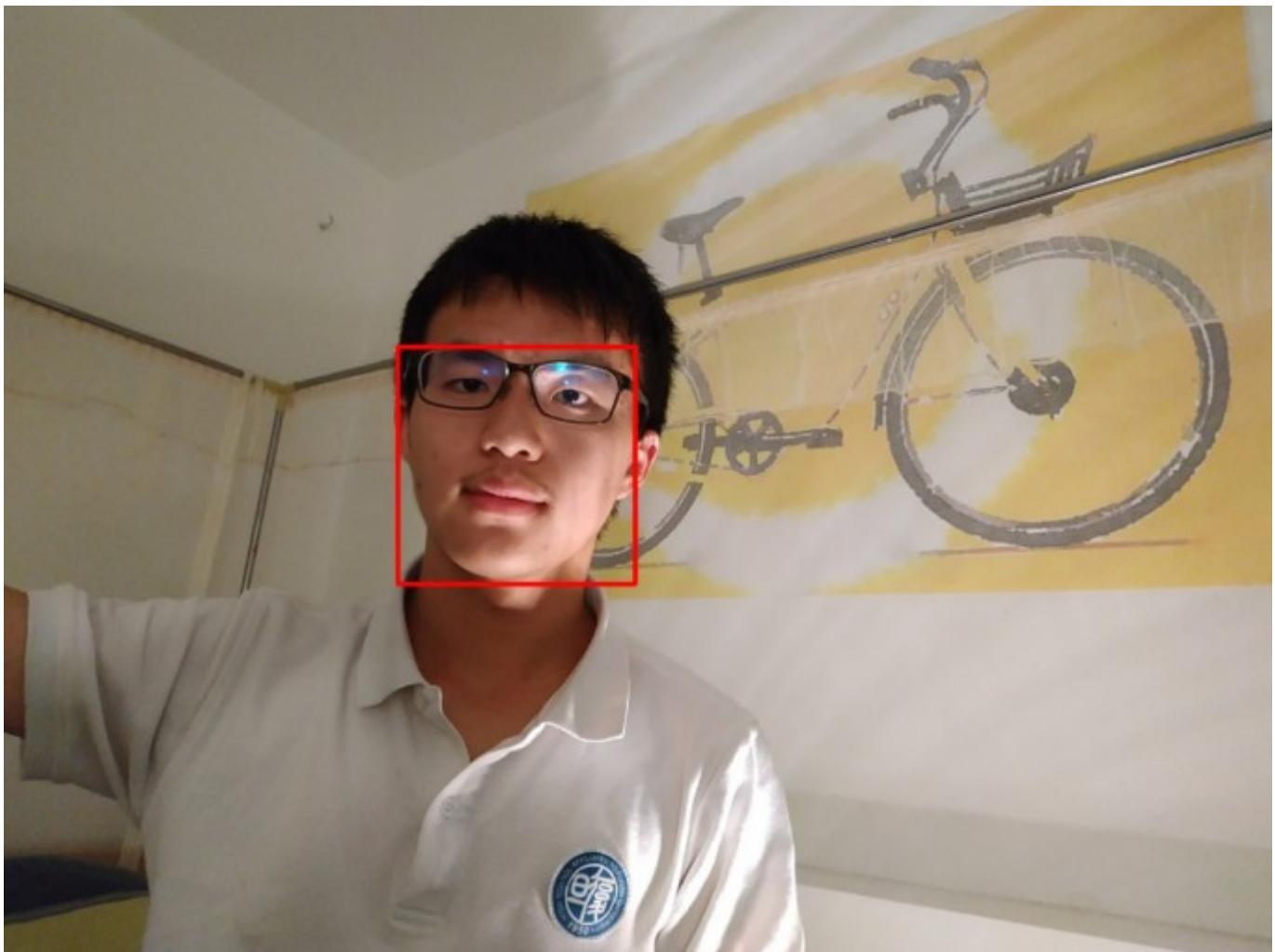
3、获得人脸特征信息

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：



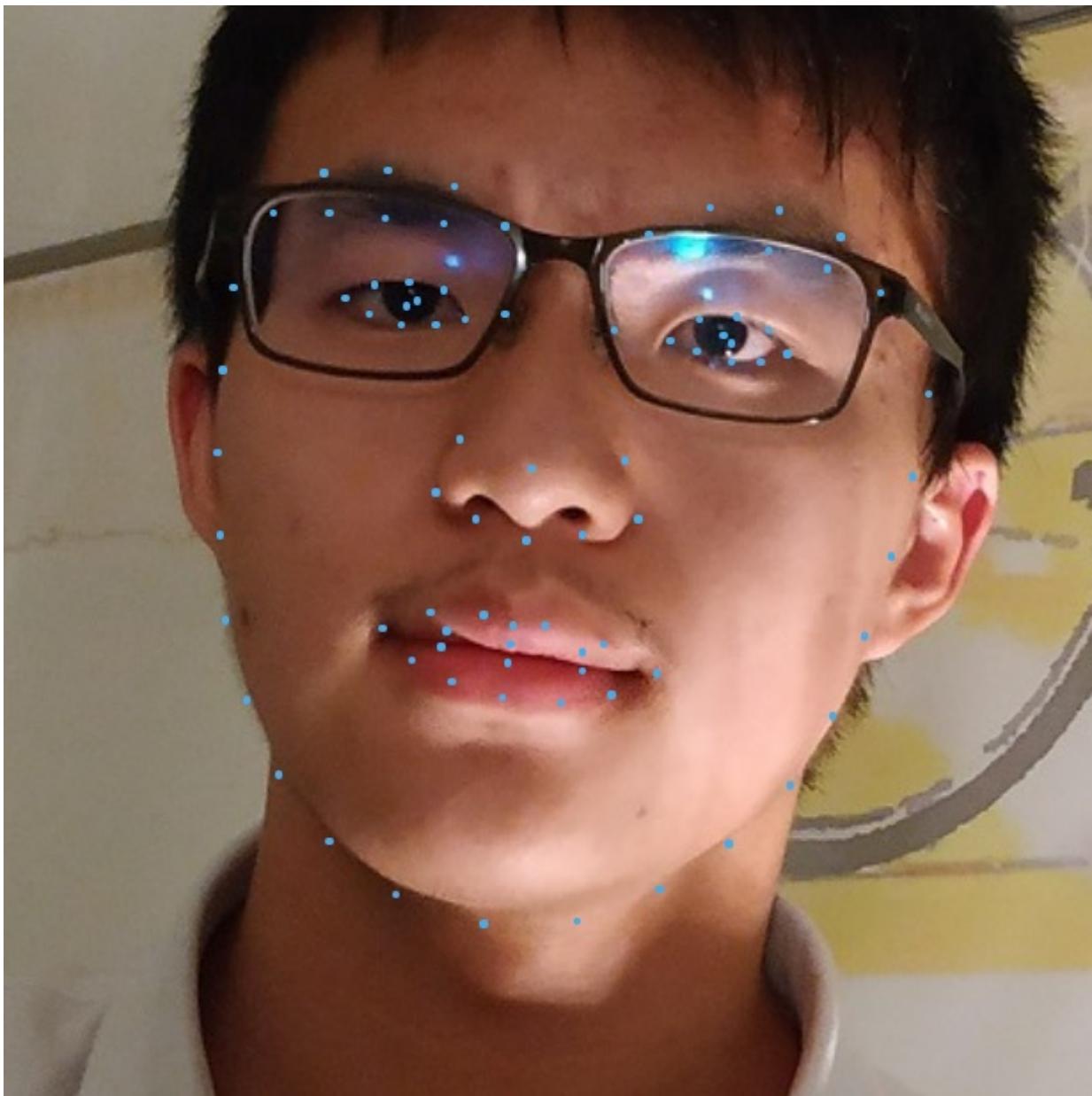
(图

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。