



《数字图像处理》课程报告

PixelForge

组 员：
李鹏达 10225101460
吴泽霖 10225101428

2024 年 6 月 - 7 月

摘要

进入二十一世纪以来，随着计算机技术的不断发展，数字图像处理技术也得到了相应的提高。同时，该技术的应用领域也不断拓展，从改善人的视角，发展到工业，农业，军事等领域的广泛应用。在这种背景之下，个人的数字图像处理需求也逐渐增加。

本项目利用 OpenCV，Pytorch 等开源的图像处理库，旨在制作一个个人可用的数字图像处理软件。同时，依据 VGG19，CycleGAN 等模型，此项目实现了更多趣味性的功能，为个人用户提供了新的创作与编辑能力。

关键词: 图像处理，风格迁移

目 录

摘要	i
第一章 项目简介	1
1.1 项目概述	1
第二章 基础功能实现	1
2.1 具体功能	1
第三章 具体场景应用	2
3.1 图像矫正	2
3.1.1 边缘提取	2
3.1.2 纠正	2
3.2 图像扫描效果	2
第四章 风格迁移	4
4.1 基于 VGG19 的风格迁移	4
4.2 基于 CycleGAN 的人脸卡通化风格迁移	4
第五章 总结	5
参考文献	6

第一章 项目简介

1.1 项目概述

本项目旨在开发一个综合性的数字图像处理平台，涵盖基本的图像处理功能，基于人脸的卡通化技术及多对多的风格迁移功能。

基本的图像处理功能包括但不限于：图像增强、滤波、形态学处理、几何变换等。同时还包括了图像的矫正和扫描功能。卡通化技术通过图像转换算法把现实的人脸部照片转化为具有卡通风格的人像。风格迁移功能则允许用户将一种艺术风格迁移到另一种艺术风格上。项目意在为用户提供一种全面而直观的数字图像处理体验，同时拥有具有实用性和创新性的图像处理技术。

第二章 基础功能实现

2.1 具体功能

该部分我们主要依据课程中的相应内容，实现了基础的图像处理功能。具体的功能如下：

1. 基本功能：包括图像的读取、显示、保存，以及灰度图、二值化和彩色图像之间的转换。
2. 图像计算：包括图像的加减乘除、缩放、平移、旋转，以及图像的仿射变换、傅里叶变换和其频谱的绘制等
3. 图像增强：图像的对数变换，线性变换和直方图均衡化等操作。
4. 形态学处理：包括腐蚀、膨胀、开运算、闭运算等操作。
5. 图像恢复：包括高斯噪声，椒盐噪声。还包括图像的排序滤波、均值滤波、高斯滤波、选择滤波等操作。
6. 图像分割：包括阈值分割、线性变化检测、边缘检测等操作。
7. 图像锐化：包括空域的锐化，频域的锐化。空域包括：Roberts 算子，Sobel 算子，Prewitt 算子，和 Laplacian 算子。频域包括：理想高通滤波器，ButterWorth 滤波器和高斯高通滤波器。
8. 图像平滑：包括空域的平滑，频域的平滑。空域包括：邻域平均法，中值滤波法（ $3*3$ ）和中值滤波法（ $5*5$ ）。频域包括：理想低通滤波器，ButterWorth 滤波器和高斯低通滤波器。

第三章 具体场景应用

3.1 图像矫正

图像的矫正是为了提取到图中所需要的子图像，并且将其矫正为矩形的形状。那么这里就分为两个部分来实现，第一个是边缘提取，第二个是矫正。

3.1.1 边缘提取

首先我对图像进行一个闭运算，消除图像的文字部分，方便后续的处理。然后，我们使用 OpenCV 提供的 GrabCut 算法，对图像的主体部分进行大致的提取，并进行掩膜操作。

然后我们进行边缘提取，这里我们先对图像进行高斯滤波，减少噪点，然后使用 Canny 算法进行边缘提取。这样我们就可以提取出大致的边缘^[1]。

3.1.2 矫正

在边缘提取之后，我们需要对图像进行矫正，也就是使用仿射变换。由于我们处理的图像一般是四边形，所以我们需要提供四个角点。在进行边缘提取以后，大部分情况下可以直接找到对应的点。对于其他的情况，我们可以使用霍夫变换，找到图中的直线，并求取这些直线的交点，作为角点。但是这样的变化有时候，并不能取得四个点，对于点缺失的情况，我们默认使用图片的边缘来代替。在大致的选点之后，我们还提供了手动选点的功能，方便用户自行调整想要选取的部分。在选点之后我们还需要对点进行排序，这样才能进行仿射变换。最后我们使用 OpenCV 提供的仿射变换函数，对图像进行矫正。



图 3.1 边缘提取与矫正效果

3.2 图像扫描效果

这一部分参考网上的博客，我首先使用了高通滤波之后，再使用了二值化，然后在去噪，来表现这种效果。

```
# Filter and denoise
gray = cv.cvtColor(dst, cv.COLOR_BGR2GRAY)
# Apply adaptive thresholding
binary = cv.adaptiveThreshold(gray, 255,
                               cv.ADAPTIVE_THRESH_MEAN_C, cv.THRESH_BINARY_INV, 21, 10)

kernel = np.ones((3, 3), np.uint8)
```

```

binary = cv.morphologyEx(binary, cv.MORPH_CLOSE, kernel)
return binary

```

实际使用过后发现，这样的方法在处理简单的图像时效果还不错，但是在处理复杂的图像时效果不佳。所以我选择使用了另一种方法，也就是卷积化的方法，这种方法可以更好的保留图像的细节，同时也可以更好的去噪。

```

warped = cv.cvtColor(dst, cv.COLOR_BGR2GRAY)
# 获取局部区域的阈值
T = threshold_local(warped, 11, offset=10, method="gaussian")
warped = (warped > T).astype("uint8") * 255
laplacian_kernel = np.array([[0, -1, 0],
                            [-1, 5, -1],
                            [0, -1, 0]], dtype=np.float32)
# 对图像进行卷积运算
sharpened = cv.filter2D(warped, -1, laplacian_kernel)
# 确保像素值在合理范围内
sharpened = np.clip(sharpened, 0, 255).astype(np.uint8)
return sharpened

```

还有一种效果最好的方法就是矩阵的除法，再对画面进行高亮处理，但是这种技术并不适合全部的图像，会导致图像缺失，所以最终没有采用。

```

src1 = image.astype(np.float32)
gauss = gaussian_filter(image, sigma=101)
# 高斯滤波
gauss1 = gauss.astype(np.float32)
dst1 = (src1 / gauss1) * 255

```



图 3.2 扫描效果对比(最左侧为原图，左侧为后去噪的滤波法，中间为先去噪的滤波法，最右侧为卷积法)

第四章 风格迁移

4.1 基于 VGG19 的风格迁移

VGG19 作为 2015 年由 **Karen Simonyan** 和 **Andrew Zisserman** 提出的深度卷积神经网络，其具体小卷积的特点，即其利用的是一个 3×3 的卷积核，这样的卷积核可以更好的提取图像的特征。主流的代码都是利用 Keras 实现的，我们这里使用 Pytorch 来重新调取模型^[2]。

VGG19 模型在风格迁移中主要针对艺术风格的迁移，我们在进行风格迁移对其进行十次左右的迭代，并且取其中生成器的 loss 最小的图像作为最终的结果。对于特征不明显的风格，VGG19 的效果并不是很理想。

4.2 基于 CycleGAN 的人脸卡通化风格迁移

CycleGAN 属于是无配对数据 (unpaired) 产生的图片，也就是说你有一些名人名家的作品，也有一些你想转换风格的真实图片，这两种图片是没有任何交集的。CycleGAN 的原理是通过两个生成器和两个判别器，通过两个生成器的互相迁移，来实现风格的迁移^[3]。

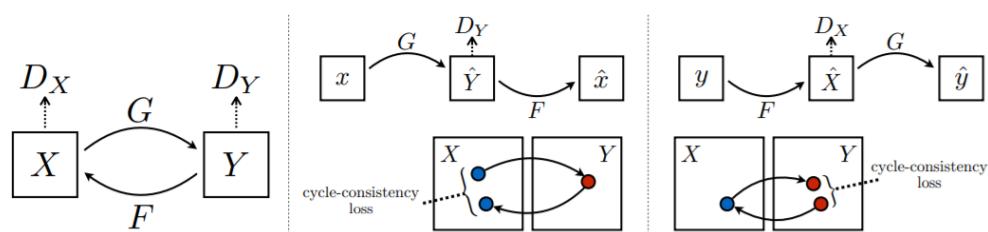


图 4.1 训练原理

我们希望有两个输入 X 与 Y ，通过两个生成器 G 与 F ，分别将 X 转换为 Y ， Y 转换为 X ，然后通过两个判别器 D 与 D' ，分别判断 X 与 Y 的真实性。通过两个生成器的对抗，来实现风格的迁移。所以我们的损失函数应为两个生成器的损失函数的和，判别器类似。具体如下：

$$\begin{aligned} Loss_{GAN} &= \mathcal{L}_{GAN}(G, D, X, Y) + \mathcal{L}_{GAN}(F, D', Y, X) \\ &= \mathbb{E}_{y \sim p_{data}(y)}[\log D(y)] + \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D(G(x)))] \quad (4.1) \\ &\quad + \mathbb{E}_{x \sim p_{data}(x)}[\log D'(x)] + \mathbb{E}_{y \sim p_{data}(y)}[\log(1 - D'(F(y)))] \end{aligned}$$

我们这里使用的是 Pre-trained 的 CycleGAN 模型，通过人脸识别的开源的大模型，在一组照片中提取出人脸，然后通过双判别器的对抗，训练出一个可以将真实照片转换为卡通风格的生成器。



图 4.2 人脸识别效果

在训练的过程中，我们发现两个损失函数存在鲁棒性的问题。分析发现，CycleGAN 对于人类的肤色和性别都极其敏感，对于广泛人脸数据集下的训练效果甚微。于是我们使用开源的卡通化图集，针对亚洲女性的人脸进行训练，效果明显提升。在测试 1k 个轮次之后，两个损失函数的值都得到了大幅度降低。由于时间的限制，我们没有进行更深入训练，对于人的眼睛以及头发与皮肤的交界处的效果仍然不太稳定，并且卡通化的人脸带有明显的女性化特征，我们将在后续进行偏向男性的卡通化训练^[4]。

第五章 总结

在该项目中，我们实现了课程中的基础的图像处理功能，并且对相关代码进行了优化。在此基础上，我们实现了文档的矫正和扫描功能。在风格迁移方面，我们主要使用开源的模型，并且进行了一定量的模型训练。此外，我们还使用 Avalonia 框架使我们的程序更加的美观。

参 考 文 献

- [1] Rother C, Kolmogorov V, Blake A. " GrabCut" interactive foreground extraction using iterated graph cuts[J]. ACM transactions on graphics (TOG), 2004, 23(3): 309-314.
- [2] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [3] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.
- [4] Kim J, Kim M, Kang H, et al. U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation[J]. arXiv preprint arXiv:1907.10830, 2019.