

深度学习第二次项目报告

画风迁移

2017011558 自 72 张扬

2017010549 自 73 桑金楠

目录

1 Background.....3

2 Related Work.....3

3 Model And Methods.....4

4 Experiment Results6

5 Summary7

1 Background

在近年的研究中，画风迁移一般都通过神经网络的训练与优化，保证图像的图形内容不被破坏的情况下，赋予其另一幅图像的艺术风格。这种技术可以将我们日常生活中的图像赋予含有艺术感的画风。

在画风迁移的方法出现之前，图像艺术渲染的工作主要是通过传统机器学习大样本训练的方法来实现的，构建的模型通过提取样本中的特征元素，在处理新图像时通过类比特征来创建新作品。这种方法的一个重大缺陷在于目前风格作品样本的不足。而画风迁移的工作与此不同，仅需一幅图片的实例和一张任意的风格艺术图即可完成工作。

目前做画风迁移的主流思想分为两种。第一种是基于描述性的模型，即利用预训练好的模型直接对目标图片进行迭代优化，典型代表作 Leon A. Gatys 等人提出的基于 VGG 深度卷积网络优化目标图像的工作[1]。第二种是基于生成性的模型，即利用图片信息训练优化模型，训练好的模型将目标图片作为输入，输出直接作为最终结果，比较有代表性的工作有 Cycle-GAN 等。然而 Cycle-GAN 也并不能表现出比较令人满意的结果。

虽然描述性思想具有直观性和良好的结果，但这种算法因为需要不断迭代优化，不到比较大的迭代次数就不会获得较好的迭代结果，时间上消耗代价过大，因此在本次项目中我们更希望能够找到时间消耗代价少，同时也保证效果良好的基于生成性的模型。

因此在以上的背景之下，我们的最终想法就是找到一个 Autoencoder 将输入的示例图片和指定艺术风景图分别编码，融合后通过 decoder 输出风格迁移图，最终结果的评估决定由我们自身以及周边同学的主观评判来实现。

2 Related Work

在以上背景的情况下，我们做了一些前期的调研，首先是基于描述性的模型，Leon A. Gatys 等人提出了利用 VGG-19 通过迭代优化目标图像以完成画风迁移，VGG-19 提取出待迁移图像特征和指定风格的图像特征，优化目标为最小化两者的协方差矩阵[1]。在文中给出的效果如图 1 所示。

除此之外，在 Gatys 工作的基础之上，Michael Elad 等人改进了原有的基于纹理合成的算法，使其达到接近于卷积神经网络的优化效果，同时也对内容界定、风格界定等重要问题做出了回答，给出的效果如图 2 所示。

其次我们还尝试寻找了一些基于生成性的模型，最为契合我们最终想法的就是 WCT 模型，具体介绍详见“3 Model And Methods”。

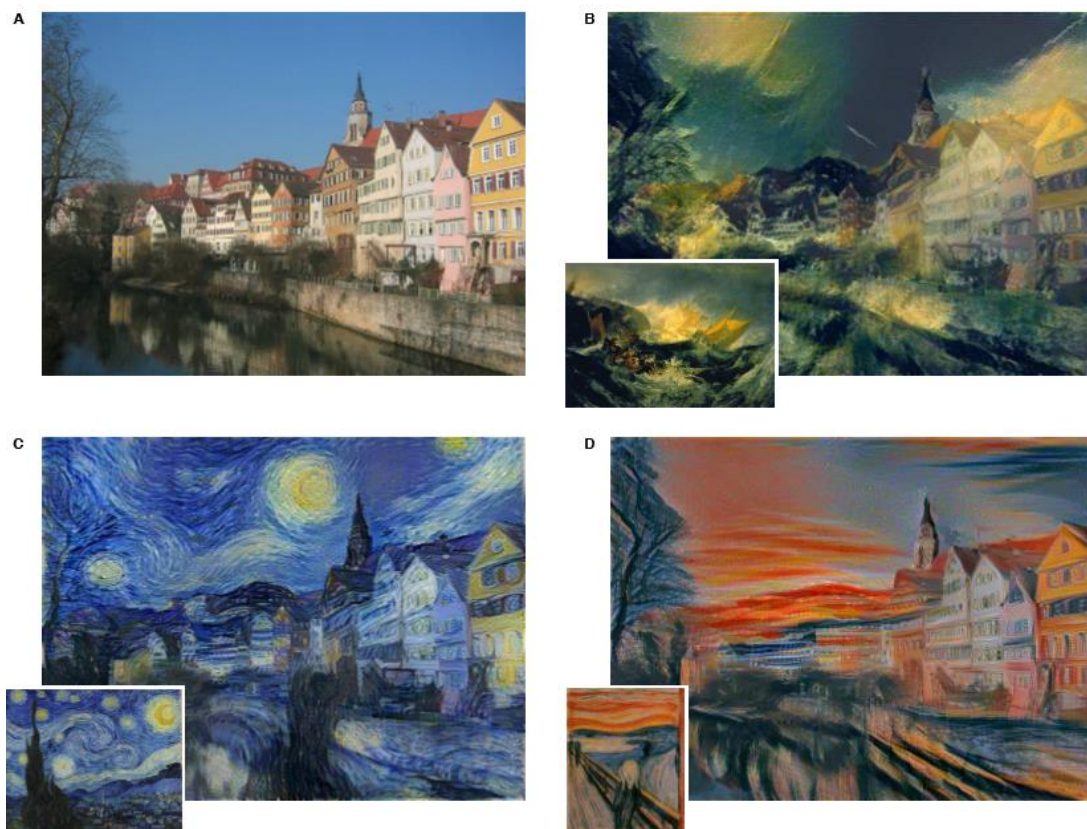


图 1



图 2

3 Model And Methods

本次项目中我们最终采用的是类似于 Autoencoder 结构的 WCT(Whitening and coloring transform)模型[3], WCT 模型是将画风迁移问题转换成了图像重建的问题, 同样是利用 VGG-19 作为 encoder 将图像提取为特定空间下的特征表示, 为了能够将输入某个特征表示准确解码成应显示的图片, 采用对称的结构定义出 decoder, 训练优化目标是尽可能将编码后的特征复原出跟原图接近甚至一样的图。为了能够充分利用上五层特征提取, 对 VGG-19 的五次特征提取分别训练出对应的 decoder, 转换成一个流水线重建图的过程。

Decoder 的 loss function 比较通俗易懂, 为

$$L = ||I_o - I_i||_2^2 + \beta ||\phi(I_o) - \phi(I_i)||_2^2$$

其中 I_i 和 I_o 分别是输入图片和重建图片， $\phi(I_i)$ 和 $\phi(I_o)$ 分别是 VGG 编码器提取出的输入和输出图特征映射，因此 loss 我们可分为两部分，一部分称为 pixel reconstruction loss，一部分成为 feature loss。

以上只是针对图像重建的问题的方法，接下来才是实现风格迁移的原理，给定一对内容图片 I_c 和风格图片 I_s ，我们首先提取出各自在某层(例如 Relu_4_1)对应的 VGG 特征映射 $f_c \in R^{C \times H_c \times W_c}$ 和 $f_s \in R^{C \times H_s \times W_s}$ ，其中 $H_c, W_c(H_s, W_s)$ 是内容(风格)特征映射的高和宽，C 是通道数，我们希望能够利用 f_s 的某些统计特性调整 f_c ，WCT 的目的就是直接对 f_c 做线性变换使其匹配上 f_s 自身的协方差矩阵，其中包含了两步线性变换：

- **Whitening Transform.** 对 f_c 做零均值化，然后对 f_c 做线性变换如下

$$\hat{f}_c = E_c D_c^{-\frac{1}{2}} E_c^T f_c$$

其中 $f_c f_c^T = E_c D_c E_c^T$ ， D_c 是 $f_c f_c^T$ 的特征值对角阵，而 E_c 是特征向量正交化后的正交阵，变换后的 \hat{f}_c 满足 $\hat{f}_c \hat{f}_c^T = I$ 。这样处理的原因是因为作者发现了变换后的 \hat{f}_c 解码对应的图片仅包含少量原有图片的颜色信息，而大部分信息是基本的图片内容框架，如图 3 所示



图 3

至此画风迁移问题近似简化为仅剩下如何上色的问题需要解决。

- **Coloring Transform.** 同样是对 f_s 做零均值化，然后对之前获得的 \hat{f}_c 做线性变换如下

$$\hat{f}_{cs} = E_s D_s^{\frac{1}{2}} E_s^T \hat{f}_c$$

其中满足 $f_s f_s^T = E_s D_s E_s^T$ ， D_s 是 $f_s f_s^T$ 的特征值对角阵，而 E_s 是特征向量正交化后的正交阵，变换后的 \hat{f}_{cs} 满足 $\hat{f}_{cs} \hat{f}_{cs}^T = f_s f_s^T$ ，这样一来相当于变换后的特征具有和风格特征相同的自协方差。最后由于 f_s 做了零均值化，需要逆变换回原来的特征描述空间，即

$$\widehat{f}_{cs} = \hat{f}_{cs} + m_s$$

同时为了能够让用户指定不同大小的风格迁移程度，利用 \widehat{f}_{cs} 和 f_c 做一次线性插值公式，中和两者的作用，即

$$\widehat{f}_{cs} = \alpha \widehat{f}_{cs} + (1 - \alpha) f_c$$

最后自然是利用 decoder 解码对应的 \widehat{f}_{cs} ，即可获得对应的风格迁移图像，再加上之前提及的流水线重建图像流程，最终模型结构如图 4 所示

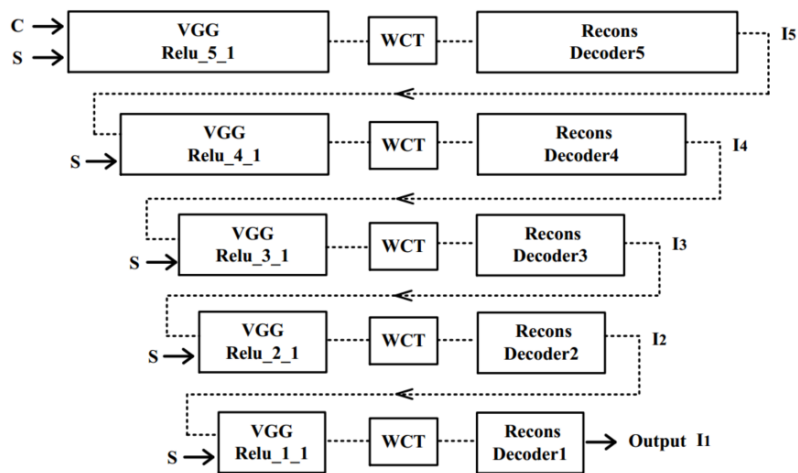


图 4

4 Experiment Results

本次笔者采用了 COCO2017 的 test 中五万张图片进行训练 decoder，每个 decoder 优化迭代 15000 次，收敛过程如图 5、6 所示(以训练 relu3_1 的 decoder 为例)

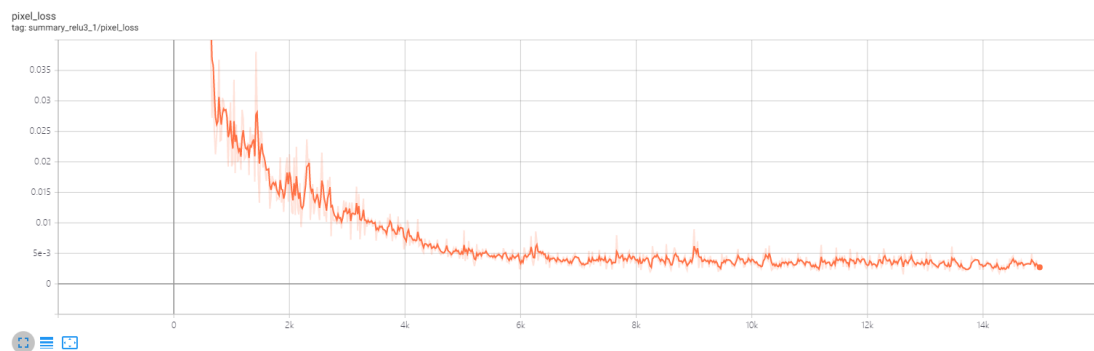


图 5: pixel loss

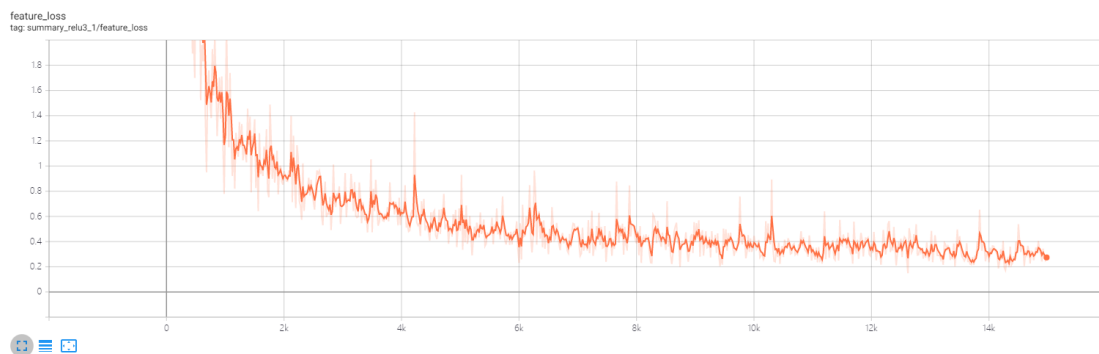


图 6: feature loss

最终结果不仅实现了画风迁移，也实现了纹理风格迁移，选取了一部分结果展示如下所示

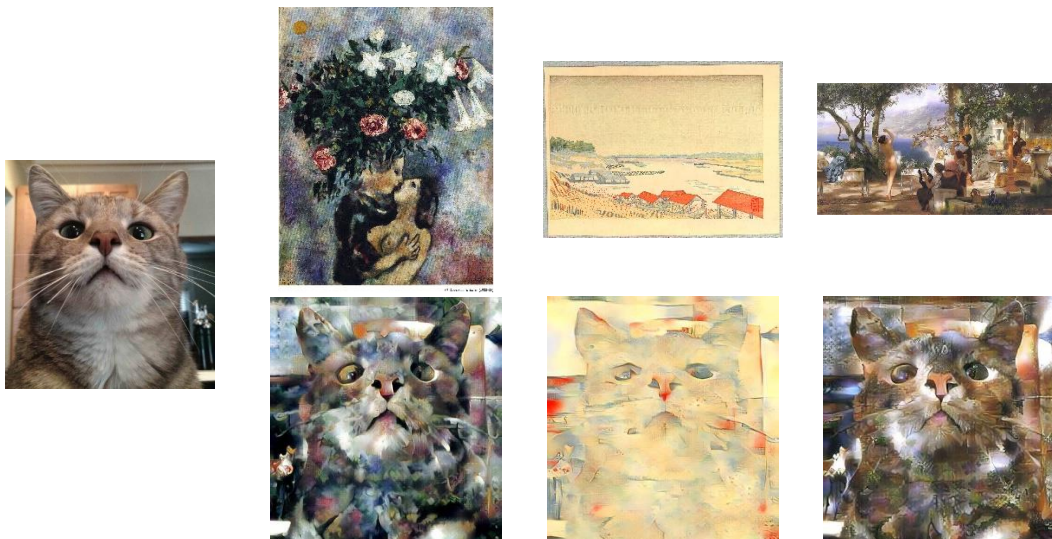


图 7：画风艺术风格迁移



图 8：纹理风格迁移

5 Summary

虽然本次项目可能仅仅是做了复现的工作，但深入地了解关于图像风格迁移这个领域的知识，还是让笔者收获颇多，特别仔细研读了三四篇其中的原理介绍，更是让笔者对这些方法的可行性和做出贡献的地方有了非常深的印象。

本次项目没能做出一些创新改进的地方，既因为考试一直到大作业 DDL 提交前两天，也主要是因为自身对于这个领域的了解不多，画风迁移之后的图片评判本就是一个掺杂主观判断的问题，很难利用一个形式化的公式描摹出来，助教曾经也跟笔者说过可利用信噪比之类的指标进行评判，但信噪比本质上只是表征信号中掺杂噪声的比例，笔者本学期在《数字图像处理》课上曾学到过，噪声的适当加入可以改善对图片的观感，简单来说就是看起来可能会更舒服一点，所以笔者思前想后还是没办法加入对项目的改进。

总而言之，感谢助教们对笔者的悉心指导。

Reference

- [1] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge(2016). Image Style Transfer Using Convolutional Neural Networks. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2414-2423
- [2] Michael Elad, Peyman Milanfa. Style Transfer Via Texture Synthesis(2017). *IEEE Transactions on Image Processing (Volume: 26 , Issue: 5 , May 2017)*, pp. 2338-2351
- [3] Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.H.: Universal style transfer via feature transforms. *In: NIPS. (2017)* 385–395