

关于图像风格迁移的简单综述

唐一峰

21824088

航空航天大学

21824088@zju.edu.cn

摘要

风格迁移是一种非常有趣的图像处理方法，主要目的为将一种图像的抽象风格特征抽取出来，并迁移至另一张图像上。一种普遍的应用是，迁移名画的风格到一张普通的图像上，使得这张被处理的图像类似于名画。有多个小有名气的 App，如 DeepArt 和 Prisma，都使用了这种图像处理方法而一夜爆红。本文将以风格迁移处理内容数与风格数为主线，主要介绍三种风格迁移的算法框架。

1 风格迁移的传统方法

在深度学习流行之前，图像风格迁移的主要方法为非参数法：分析某一种风格的图像，并针对其建立一个数学或统计模型，再对目标图像做出改变使其更符合建立的数学模型，主要基于物理模型的绘制和纹理的合成。

Hertzmann 等人^[1]提出了一种基于类推思想的方法，通过图像特征映射关系合成具有新纹理的图像。Efros 等人^[2]提出了一种简单的纹理算法，通过对样本纹理进行拼接和重组以合成新的纹理。首先对风格图像的纹理进行手工提取，并建立数学模型，之后算法的思路非常简单，即使用动态规划法，改变目标图像的像素使得目标图片与风格图像的纹理契合的最好，即重叠误差最小。

$$E_{ij} = e_{ij} + \min(E_{i-1,j-1}, E_{i-1,j}, E_{i-1,j+1})$$

最终实现的图例如图 1 所示，可以看出对于类似于左图纹理风格简单的图像处理效果很好，而右图具有复杂纹理风格的图像则效果不佳。传统方法的一大缺点即在于此，难以适应复杂的图像变化。其主要原因在于非参数的图像风格迁移方法只能提取图像的底层特征，而非高层抽象特征，因此在处理颜色和纹理较

复杂的图像时，最终的图像合成效果较为粗糙。

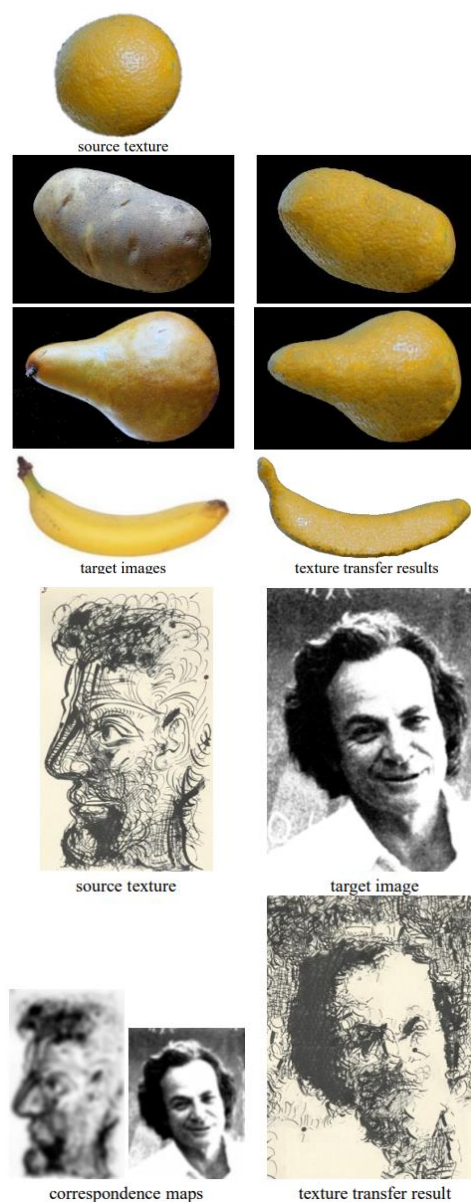


图 1 基于纹理算法的图例

2 深度学习用于风格迁移的先例

基于传统风格迁移研究的实际应用非常有限，深度学习的兴起改变了这一现状。2015 年第一个基于神经网络的图像风格迁移算法由 Gatys^{[3][4]}实现了：使用卷积神经网络来将

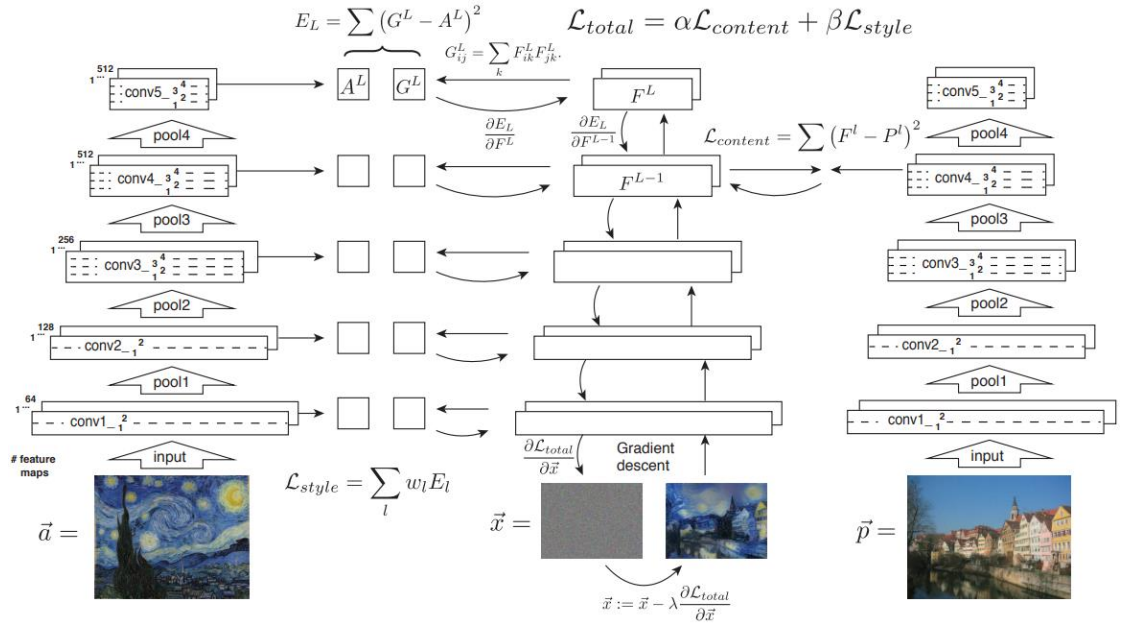


图 2 卷积神经网络应用于风格迁移的实现过程

图像的内容抽象特征表示和风格抽象特征表示进行分离，然后从随机噪声图像开始，通过迭代优化的方式生成具有目标图像内容和新风格的合成图像，来有效地实现图像风格迁移，因此这种方法也被称为基于优化的方法。

论文中使用的框架是 VGG19，如图 2 所示有三个框架和三个输入图片。a 是风格图，p 是内容图，x 是随机生成的噪声图。该过程的总体思想就是通过对噪声图像的参数调整，使其不断的同时趋近于内容图也趋近于风格图。论文中使用的是两种损失函数：风格和内容损失。内容损失 $\mathcal{L}_{content}$ ，是由内容图和随机噪声图经过卷积滤波后，在第四层分别的特征图距离的平方和；风格损失 \mathcal{L}_{style} 是对风格图和噪声图的每一层滤波的特征图分别求 gram 矩阵，再求其距离的平方和，再将 5 层的结果加权求和。最后将两种损失加权求和求得总的损失 \mathcal{L}_{total} 。

$$\mathcal{L}_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^l - P_{ij}^l)^2$$

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2$$

在内容损失中选择第四层的原因，作者做了大量试验，发现前三层的局部特征保存很好，从第四层开始局部特征开始有一些损失，

而全局特征并没有很多损失，这正是我们所需要的结果，因此选择了第四层。

作者选择将 CNN 应用于风格迁移的思路是很有眼光的，而且也是顺理成章的。传统的风格迁移需要研究者手工提取纹理/特征，而 CNN 的一个很大的优点在于可以自动提取图像的特征或者纹理，因此可以应用 CNN 可以大大简化提取纹理的工作，这也是最大的创新点。但是基于优化的方法最大的缺点是速度慢，大概 200-300 次迭代才能得到比较理想的结果。

3 深度学习用于风格迁移的加速

针对速度慢的缺点，Justin Johnson 等^[5]提出了一种改进方法来加速风格迁移。与 Gatys 的方法不同的是，不再使用机生成的噪声图，而是在输入图像前搭建一个转换网络，即生成的图像是从内容图片中得到的。

从图 3 可以看出网络分为 Image Transform Net 和 Loss Network 两种，在训练中，Transform 网络用来对图像进行转换，它的参数是变化的，而 Loss 网络，则保持参数不变，Transform 的结果图，风格图和内容图都通过 Loss Net 得到每一层的 feature 激活值，并以此进行 Loss 计算，输入 $x=y_c$ 是内容图片。

使用了逐像素求差损失函数的那些方法无法抓住输入及输出图像在感知上的差距。举

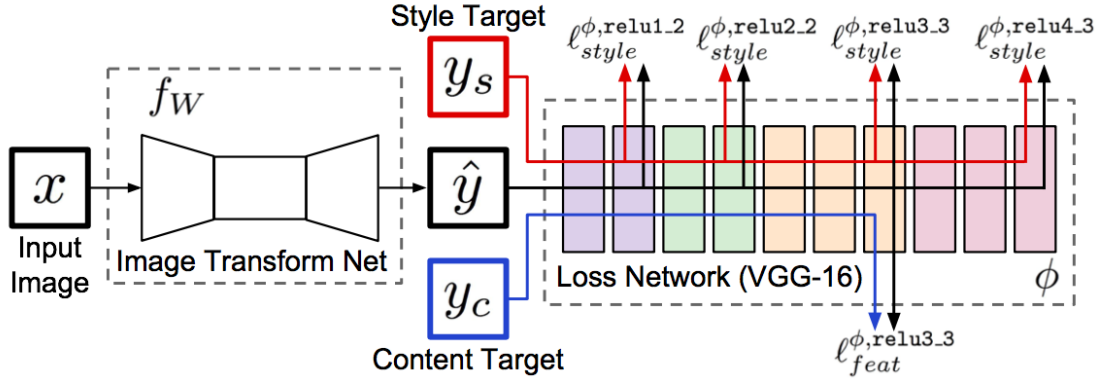


图3 具有前馈转换网络的网络架构图

个例子，考虑两张一模一样的图像，只有1像素偏移上的差距，尽管从感知上这俩图片一模一样，但用逐像素求差的方法来衡量的话，这两张图片的结果差别很大。这篇论文结合已有的感知方法，即通过建立感知损失函数（不基于逐像素间的差距，取而代之的是从预训练好的CNN中提取高层次的图像特征来求差），图像通过使损失函数最小化来生成。

因此这篇文章的主要思路为：先训练一个用于图像转换任务的前馈网络，且不用逐像素求差构造损失函数，转而使用感知损失函数，从预训练好的网络中提取高级特征，在测试的过程中，图像转换网络能做到实时转换。定义了两个 perceptual loss 函数分别用来衡量两张图像的高级感知差距和语义差距，损失计算如下。

内容特征损失依然使用了网络的深层，只保留图像的内容和空间结构：

$$\ell_{feat}^{\phi,j}(\hat{y}, y) = \frac{1}{C_j H_j W_j} \left\| \phi_j(\hat{y}) - \phi_j(y) \right\|_2^2$$

风格损失依然为多层网络生成损失总和：

$$G_j^{\phi}(x)_{c,c'} = \frac{1}{C_j H_j W_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} \phi_j(x)_{h,w,c} \phi_j(x)_{h,w,c'}$$

计算的损失函数训练只是为了训练前面的图像转换网络，这个模型是采用SGD的方法更新权重，当模型训练好之后，用于生成风格迁移的只是这部分图像转换网络。

这种基于感知损失的方法出现，大大加速了风格迁移的速度，在速度上比基于优化的

Gatys的方法快三个数量级，而质量上则达到了相同水平，个人认为大部分的生成图像略差于Gatys的生成图像。

4 用一个计算网络实现多种风格的迁移

前一篇Johnson的论文解决了运行速度的问题，但仍有比较大的局限性：每个模型只能进行一种风格图像的迁移，每增加一张新的风格图需要训练一个新的模型。

基于多种风格迁移的目的Dongdong Chen等^[6]提出了一个由多个卷积滤波器组组成的架构StyleBank，每个滤波器组代表一种样式。为了将图像转换为特定的样式，将相应的滤波器与单个自动编码器产生的中间特征嵌入进行卷积，将原始图像分解为多个特征响应映射。

论文提出的前馈网络中，风格库和自动编码器是共同学习的。它不仅可以实现同时学习各种风格库，而且为新的图像样式提供了非常有效的增量学习方法。后者可以只学习一个新的过滤器风格，同时保持自动编码器的参数固定。

如图4所示，论文构建了一个基于简单图像自动编码器的前馈网络，它首先通过编码器将输入图像(即内容图像)变换为特征图。接着受texton概念的启发，将通过学习风格图像得到的类比风格表示法应用于StyleBank。

网络架构由图像编码器E、StyleBank层K和图像解码器D三个模块组成，它们构成了两个学习分支，自动编码器分支(即E→D)和风格化分支(即E→K→D)。两个分支共享同

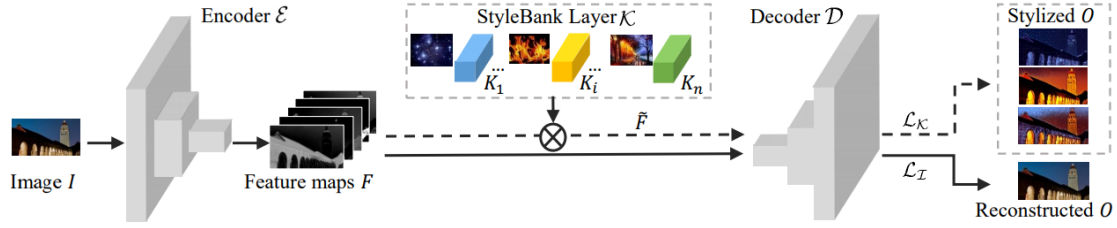


图 4 StyleBank 的架构图

一个编码器和解码器。网络需要内容图像 I 作为输入，然后通过编码器 $F = E(I)$ 将图像转换为多层特征图 F 。对于自动编码器分支，训练自动编码器生成尽可能接近输入图像的图像，即图像 I 。同时，对于风格化分支，在 E 和 D 之间添加了一个中间的风格化库 K 层。在这一层中，StyleBank 的 n 个样式分别与特征图 F 进行卷积，得到转换后的 \tilde{F} 。最后，对风格 \tilde{F} 进行解码得到最终的结果图。

这样的设计有两个明显的优势：1. 增加新的风格时可以采用增量训练法。保持编码与解码器 E, D 不变，初始化一个新的 K 进行训练，速度很快。2. 可以比较简单的进行 Region-specific style fusion，即将照片不同的区域转化成不同的风格。



图 5 与基于优化方法的对比图



图 6 与基于前馈架构的方法对比图

这篇论文最后得到的结果与基于优化的 Gatys 方法比较（如图 5 所示）可以看出表现得结果更优，与基于前馈架构的 Johnson 方法比较（如图 6 所示）可以看出结果并不逊与单

风格生成的模型方法。

5 小结

本文以实现风格迁移的内容数和风格数作为线路，主要分析了三篇经典的论文。第一篇实现了单一风格单一内容的风格迁移，第二篇实现了单一风格多内容的风格迁移，第三篇则实现了多风格多内容的风格迁移。对风格迁移的前期研究做了一个简单的综述，总体而言，图像风格迁移是一个具有挑战性的有趣课题，它不仅受到了学术界的广泛关注，而且在工业界中也有很大的应用和需求，具有广阔的研究和应用前景。

参考文献

- [1] Hertzmann, A., Jacobs, C. E., Oliver, N., Curless, B., & Salesin, D. H. (2001, August). Image analogies. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques (pp. 327-340). ACM.
- [2] Efros, A. A., & Freeman, W. T. (2001, August). Image quilting for texture synthesis and transfer. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques (pp. 341-346). ACM.
- [3] Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576.
- [4] Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2414-2423).
- [5] Johnson, J., Alahi, A., & Fei-Fei, L. (2016, October). Perceptual losses for real-time style transfer and super-resolution. In European

conference on computer vision (pp. 694-711).
Springer, Cham.

- [6] Chen, D., Yuan, L., Liao, J., Yu, N., & Hua, G.
(2017). Stylebank: An explicit representation for
neural image style transfer. In Proceedings of the
IEEE Conference on Computer Vision and
Pattern Recognition (pp. 1897-1906).