论文草稿

概述：

随着网络技术的发展以及社交软件普及，人们越来越厌烦单一和同质化，希望能够展示出不同的自我。通常通过在社交软件上发布自己的日常照片来展示自己的与众不同，而如何才能在照片中展示出不一样的自我和风格，图片风格迁移技术给出了一个答案。图片风格迁移指的是将原本的图片以其他的风格展示出来例如历史上不同的油画大师们的风格，使得图片风格有了许多种可能。最初的Gatys等人的实验结果表明CNN（卷积神经网络）可以从任何照片中提取内容信息，使得基于CNN的NST（神经样式转换）得到迅速发展。本文主要探讨利用直接为预训练的CNN中的单个层构造目标活动的方法将内容结构和样式纹理结合在一起，该方法不仅限于经过专门培训的样式，还可以轻松适应任意内容和样式图像。同时将探讨该算法发改进方案，致力于得到更完美在视觉上更易接受的合理图案

关键词：图像风格迁移 补丁操作 逆网络

简介

艺术风格的概念很难进行量化，早期的研究人员使用相识性进行度量或者基于像素值的局部统计来定义样式。在神经网络之前，图像风格迁移的程序有一个共同的思路：分析某一风格的图像，给该风格建立一个数学或统计模型，再改变要做迁移的图像让它能更好的符合建立的模型。该方法做出的效果具有一定视觉上的美观和的结果，但一个很大的缺点：一个程序基本只能做某一种风格或者某一个场景。因此基于传统风格迁移研究的实际应用非常有限。

直到Gatys的论文[1]改变这种现状，它给了一种用深度学习来给纹理建模的方法，把两个不同领域的研究成果有机的结合起来，做出了令人惊艳的结果，也使得图片风格迁移这一领域得到了革命性的发展。当前的NST方法大致可分为两类[2]。一类是Gatys等利用CNN特征激活来重组给定图片的内容以及样式的方法，该方法说明图像样式和内容在某种程度上是可分离的，可任意组合图像的内容及样式，但它依赖于过慢的迭代图像优化过程，使其计算量十分庞大。另一类是Johnson和Ulyanov[3，4]等人提出的一种使用多尺度体系结构作为生成器网络，目标函数类似于Gatys等人的算法，该算法通过MOB-IR来重建风格化图像，以解决速度和计算成本问题。

针对以往方法的耗时过长问题，一种典型的加速解决方案是再训练一个神经网络，该神经网络在单次前馈传递中传递近似优化的最优值[5，6]网络无法推广其训练好的图像集，因此使用该方法的现有作品无法对任何给定样式图像执行样式转换的多功能性，这种限制使得应用程序的发展受到了极大的阻碍。本文通过重新定义优化目标，使用基于局部匹配的简单优化目标进行改进，该目标只使用CNN中的一层进行计算，该限制使得我们可以使用逆网络进行确定性的反转来自样式化层的激活，以生成样式化图像。该方法使用基于补丁的操作使用2D卷积提取补丁，通过argmax函数通道进行计算最佳匹配样式补丁，再利用2D转置卷积将最佳匹配样式补丁放置在相应的空间位置来重建完整的激活图片。

相关工作

针对样式转移的问题，早期的方法包括对线性滤波器响应进行直方图匹配[7]和非参数采样[8]，这些方法通常依赖于低级统计信息，且通常无法获得语义结构。Gatys等首次通过在CNN卷积层中匹配特征统计信息，得到相对更完美的样式转换结果。他们[1]将样式转换描述为纹理合成和内容重建两者结合在一起的优化问题，公式涉及到将附加损失函数置于预先训练的CNN的多层上，其中一些损失函数用于合成样式图像的纹理，另外一些损失函数用于重新构建内容图像。通过反向传播计算梯度，并基于梯度的优化来构建风格化图像。

另一种方法是在内容和样式图像之间使用基于补丁的相似度匹配[9]。其中Li和Wand 构建了基于补丁的损耗函数，其中每个合成补丁都具有必须匹配的最近邻居目标补丁。然后，将这种补丁匹配损失函数与Gatys等人提出的公式相加。虽然这种方法允许使用任意样式的图像，但是这种方法使用的优化框架使其生成样式化图像的速度变的相对缓慢。正如上文所说，前馈样式网络可以训练一个神经网络，该网络近似于Gatys等人针对一种或多种固定样式的损失函数的最优值。 相比于之前的方法，该方法能够获得更快的速度，但是对于每种新样式，都需要对这些方法进行重新训练 。

预训练卷积神经网络中的逆网络常被用于作为自动编码器的一部分。几乎所有现有的逆网络都要经过图像数据集和放置在RGB空间中的损失函数的训练。由于逆网络【】

不同于现有的样式转换方法，本文所研究的方法是一种直接为预训练的CNN中的单个层进行构造目标活动的方法。其中该方法有三处较大改进：

1. 该方法利用准则在激活空间中找到最佳匹配的补丁，与Li和Wand [10]提出的方法不一样的是该方法能够直接构建整个激活目标；
2. 该方法提出了一种通过反转预训练的CNN来进行前馈样式转换的程序，相比于前馈样式对于每一种新样式都要重新训练的不足，该方法不仅限于经过专门培训的样式，还可以轻松适应任意内容和样式图像；
3. 该训练方法不使用像素级损失而是使用激活损失，与现有的CNN反转方法不同，该方法通过使用特定的训练设置，该反向网络甚至能够反转超出CNN激活范围的激活。

下文中实验原理一章将会具体讲解该方法的原理和实现，实验过程主要说明代码和运行过程，实验结论一章将会和其他以往方法进行横向对比试图寻找该方法的优缺点。

实验原理及改进

目前的样式转换方法大致分为两类要么是通过迭代优化图像来传递样式，或是优化生成模型并通过单次向前生成样式化图像。本文主要讨论的方法是第二类，即利用优化生成模型的方法通过基于补丁的操作，在给定样式图像和内容图像的情况下在单层中构造激活目标，用样式图像逐个补丁替换内容图像。

实验原理：

令C和S分别代表内容图像和样式图像的RGB表示，令Φ（）表示预训练CNN的完全卷积部分所表示的函数，该函数用于将图像从RGB映射到中间激活空间。样式转换过程中首先计算出激活量Φ（C）和Φ（S），然后通过公式得到最佳匹配的样式补丁φss（C，S），之后利用最佳匹配样式补丁对图像进行激活。具体样式交换过程如下：

1.为内容和样式激活提取一组补丁，分别用φi(C) i∈nc和φj(S) j∈ns表示，其中nc和ns是从内容和样式图像中所提取补丁的数量。提取的补丁应具有足够的数量使得其可以有足够的重叠部分，并且包含激活的所有通道。  
  
2.对于每个内容激活补丁，根据归一化互相关系数确定其最佳匹配的样式补丁，  
3.用上一步中最匹配的样式补丁φss（C，S）交换每个内容激活补丁φi（C）。  
4.由于步骤3可能会产生具有不同值的重叠区域，所以应当对不同值的重叠区域进行平均计算以重建完整的内容激活补丁，用Φss（C，S）表示。  
该交换操作所形成的图像具有内容图像相同的结构，同时该图像具有从样式图像所得到的纹理特征

具体实现方案：

整个图像样式交换操作可以分为具有三个不同操作功能的网络来一一实现：（i）2D卷积层，（ii）通道argmax，（iii）2D转置的卷积层。这样的分配使得样式交换的实现可以利用现有的2D卷积和转置卷积的现有有效实现。  
实验中为了表达简便，对内容激活补丁进行重新定义表达以明确表示空间结构。其中，

1. 将d作为Φ(C)的功能通道数
2. 令φa，b(C)表示面片Φ(C)
3. a表示为a + s，b表示为b + s，1 ：d其中s是补丁大小。

其中由于内容激活补丁的标准化项对于argmax操作是恒定的，因此（1）可以重新表达为

缺少内容激活的规范化  
补丁的方法简化了计算并且可以使用2D卷积层。因此可以通过以下三个步骤进行规范化的实现，原理如图2所示：  
  
①：通过使用归一化样式激活补丁φj（S）/φj（S）作为卷积滤波器以及激活量Φ（C）作为输入，并通过单个2D卷积来计算张量K。计算出的张量K具有nc个空间位置和ns个特征通道。在每个空间位置Kab是内容激活补丁和所有样式激活补丁之间的互相关向量。  
  
②：在使用2D转置卷积计算之前，需要重新将每个互相关向量Kab放置在和最佳匹配样式激活补丁φss（C，S）相对应的一个热点向量上。  
  
③：使用2D转置卷积，其中张量K作为输入，非标准化样式激活补丁φj（S）作为滤波器。在每个空间位置，只有最佳匹配样式的激活补丁会输出，其他补丁会通过乘以零进行清理。  
  
其中，转置的卷积将会汇总重叠的补丁块的值。在输出的每个空间位置上按重叠块的数量进行逐个元素的数学期望计算以此来对这些重叠补丁求平均值。因此，对于（3）中的argmax不一定只用唯一的某个解决方案，因为可以将多个argmax解决方案简单地解释为添加更多重叠的补丁。

优化：

可以通过在目标激活Φss（C，S）的激活空间上放置损失函数来计算风格化图像的像素表示。 与以前的样式转换工作[11，21]相似，使用平方误差损失并将优化目标定义为  
  
这里我们说合成图像的维数是h乘w乘d，F是Frobenius范数，而ATV（）是广泛用于图像生成方法的总变化正则化项[1，17，26]。 由于Φ（）包含对图像进行下采样的多个最大池化操作，因此我们先将此正则化作为自然图像使用，从而获得对重新上采样的图像进行空间平滑处理的结果。 总变化正则化如下：  
  
由于函数Φ（）是预训练的CNN的一部分，并且至少是一次可微分解的，因此可以使用基于标准基于梯度的优化方法来计算（4）。

不幸的是，在诸如视频样式化之类的应用中，解决用于计算风格化图像的优化问题的成本可能太高。我们可以通过使用另一个神经网络逼近最佳值来提高优化速度。一旦经过培训，该网络便可以用于更快地生成风格化图像，并且我们将特别培训该网络以使其具有使用新内容和新样式图像的多功能性。  
我们的逆网络的主要目的是针对任何目标激活近似逼近（4）中的损失函数。因此，我们将最佳逆函数定义为：  
  
其中f代表确定性函数，H是代表目标激活的随机变量。与（4）相似，总变化正则项被添加为自然图像。  
  
由于预训练的卷积神经网络的特性，出现了几个问题。  
非内射。定义Φ（）的CNN包含卷积层，最大池层和ReLU层。这些函数是多对一的，因此没有明确定义的逆函数。类似于使用逆网络[4、25、36]的现有作品，我们改为通过参数神经网络训练逆关系的近似值。  
  
其中θ表示神经网络f的参数，而  
Hi是大小为n的数据集中的激活特征。由于不需要知道（4）的最佳值，因此该目标功能导致对神经网络的无监督训练。我们将对逆网络架构的描述放在附录中。  
  
非排斥的。由于插值，样式交换操作产生的目标激活可能超出Φ（·）的范围。这意味着，如果仅使用实像训练逆网络，则逆网络可能仅能够反转Φ（）范围内的激活。由于我们希望逆网络反转样式交换的激活，因此我们增加了训练集以包括这些激活。更准确地说，给定一组训练图像（及其相应的激活），我们使用基于图像对的样式交换激活来扩展此训练集。  
  
一旦经过训练，逆网络就可以用来代替优化程序。因此，我们建议的前馈过程包括以下步骤：  
1.计算Φ（C）和Φ（S）。  
2.通过样式交换获得Φss（C，S）。  
3.将Φss（C，S）馈入经过训练的逆网络。  
该过程如图4所示。如3.2节所述，样式交换可以实现为（不可微分）卷积神经网络。这样，整个前馈过程可以看作是具有单独训练部分的神经网络。与现有的前馈方法相比[6、17、33、34]，我们的前馈程序的最大优势在于能够仅使用一个经过训练的逆网络使用新型图像。

注释：

[1] L. A. Gatys, A. S. Ecker, and M. Bethge, “A neural algorithm of

artistic style,” ArXiv e-prints, Aug. 2015.

[2]J. E. Kyprianidis, J. Collomosse, T. Wang, and T. Isenberg, “State

of the ‘art’: A taxonomy of artistic stylization techniques for

images and video,” IEEE transactions on visualization and computer

graphics, vol. 19, no. 5, pp. 866–885, 2013.

[3] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for realtime

style transfer and super-resolution,” in European Conference

on Computer Vision, 2016, pp. 694–711.

[4] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. Lempitsky, “Texture

networks: Feed-forward synthesis of textures and stylized images,”

in International Conference on Machine Learning, 2016, pp.

1349–1357.

[5]V. Dumoulin, J. Shlens, and M. Kudlur. A learned representation

for artistic style. CoRR, abs/1610.07629, 2016. 1, 2

[6] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual Losses

for Real-Time Style Transfer and Super-Resolution. Arxiv,

2016. 1, 2, 4, 5, 7

[7] D. J. Heeger and J. R. Bergen. Pyramid-based texture analysis/

synthesis. In *SIGGRAPH*, 1995.

[8]A. A. Efros and W. T. Freeman. Image quilting for texture

synthesis and transfer. In *SIGGRAPH*, 2001

[9]M. Elad and P. Milanfar. Style-transfer via texture-synthesis.

*arXiv preprint arXiv:1609.03057*, 2016.

[10]C. Li and M. Wand. Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis. *Cvpr 2016*, page 9, 2016