# 一种具有空间约束的快速神经风格迁移方法

刘洪麟 帅仁俊

(南京工业大学计算机科学与技术学院 南京 211816)

**摘 要** 在神经风格迁移(Neural Style Transfer)技术中常用Gram矩阵进行图像风格提取,简单来说就是将各个特征进行内积。这样Gram矩阵只能提取其静态特征,对图片中物体的空间序列完全没有约束。文中提出了一种具有空间约束的快速神经风格迁移方法。首先,使用残差重新设计了快速神经风格迁移的转换网络。然后,运用空间偏移的方法对Feature map(特征图)进行变换,让变换后的Feature map $T(a^l)$进行Gram矩阵计算可得到相邻位置的互相关性。此互相关性包含空间信息,即约束了物体的空间序列。最终,实验表明了该方法对空间信息的约束明显优于传统方法,可以得到效果更好的风格化图像。

**关键词** 快速神经风格迁移,残差,Gram矩阵,空间偏移,互相关性

**中图法分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.03.042

## Method of Fast Neural Style Transfer with Spatial Constraint

LIU Hong-lin SHUAI Ren-jun

(College of Computer Science and Technology, Nanjing Tech University, Nanjing 211816, China)

**Abstract** Gram matrix, a method to get the inner product in simple terms, was commonly used for image style extraction in the style-transfer techniques. The Gram matrix can only extract the static features, but it is completely unconstrained to the spatial sequence of objects in the picture. This paper proposed a fast neural style transfer method with space constraints. First, the residuals are used to redesign the transform network of fast neural style transfer. Then, the method of spatial offset is used to transform the Feature map. Feature map $T(a^l)$ are used for Gram matrix computation to get the cross-correlation, which contains the spatial information. That is to say, it can constrain the object's spatial sequence in the picture. Finally, experiments show that the method's ability of space constraint is better than traditional method, and the stylized image with better effect can be quickly obtained.

**Keywords** Fast neural style transfer, Residual, Gram matrix, Spatial offset, Cross-correlation

## 1 引言

神经风格迁移是一种使用卷积神经网络将一张图片的语义内容与不同风格的图片融合起来的技术。著名学者Gatys通过实验诠释了CNN的力量:通过将图片的内容与风格分离重组,CNN可以学习到图片的风格,创作出具有艺术魅力的作品。从此,神经风格迁移便成了学术研究和产业应用中的热门话题。最初的图像风格化是基于迭代优化的,因为这种方法需要不断地迭代来寻找最优解,所以速度很慢。后来一些研究团队从另一方面解决了风格化迭代速度较慢的问题,即使用转换网络和损失网络,这种方法与迭代优化的方案相比速度更快,被称为快速神经风格迁移[1]。

目前,神经风格迁移已经在影视和游戏领域有了较为广泛的应用,比如Prisma相机、电影特效和游戏场景设计。电影*Come swim*中的情节是一半现实和一半梦境,梦境的画面就是经过神经风格迁移技术处理得到的,这种方法得到的渲染方式显然比直接用电脑绘制具有更高的性价比。因此,制片方还写了一篇题为《运用神经风格转移实现印象派风格之在电影"Come Swim"中的应用》的论文。对于CNN来说,这只是一种风格上的模仿,在创造时并没有人类的情感,所以说这种技术创造出来的作品并不是真正的艺术。人工智能不会从人类手中夺走创作权,更不会消灭艺术,只会使这项美妙的事业进入更高的境界。但在这个艺术化、商业化的时代中,神经风格迁移技术可以节约大量的人力成本和时间成本。

## 2 相关工作

目前,有很多著名的学者在神经风格迁移方面进行研究。基于深度学习的图像风格化源于Gatys等[2]的研究工作。他们运用卷积神经网络取得了很好的纹理合成效果;同时,他们又考虑如何运用卷积神经网络将一张图像的纹理迁移到另一张图片上。于是,他们又提出了一种艺术风格的神经网络算法[3]。Risser等[4]在Gatys的基础上添加了梯度直方图进行约束,实验结果表明,纹理合成和风格迁移的效果优于Gatys的算法。Johnson等[5]使用一个转换网络解决了神经风格迁

到稿日期:2018-06-01 返修日期:2018-08-05

**刘洪麟**(1992-),男,硕士生,主要研究方向为机器学习,E-mail:534110389@qq.com;**帅仁俊**(1962-)男,副教授,硕士生导师,主要研究方向医学图像处理、大数据处理等,E-mail:srjwhy@sina.com(通信作者)。

移迭代寻优速度慢的问题，图片风格化的速度提升了 3 个数量级。但网络每次只能转换一种风格，如果要切换风格，必须重新训练转换网络。Ulyanov 等[6]在之前的基础上进行改进，提出了一种多层次的网络架构，在网络局部的细节上处理得更好。

图片风格化时间以及风格化效果是神经风格迁移技术最重要的两点。转换网络改进与图片物体空间信息约束相结合，可以得到效果更好的风格化图片。

## 3 相关理论

### 3.1 神经风格迁移

神经风格迁移，即使一张普通的图片具有另一张图片的风格，其基本原理为：假设给定一张内容图片 $c$ 和一张风格图片 $s$，生成一张图片 $g$。图片 $g$ 在内容上与图片 $c$ 相似，在风格上与图片 $s$ 相似。其中使用了 VGG-19 网络提取图片的内容与风格，该技术的原理如图 1 所示。

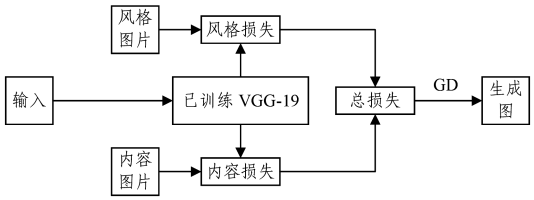



图 1 神经风格迁移原理图

Fig. 1 Schematic diagram of neural style transfer

基于迭代优化的神经风格迁移算法的步骤如下：

1)随机初始化 $g$，即随机选取像素的白噪声图作为 VGG-19 的输入；

2)求出整体损失函数 $L_{\text{total}}$；

3)使用梯度下降法优化方法求解 $L_{\text{total}}$ 的最小值；

4)如果 $L_{\text{total}}$ 收敛或小于指定阈值，则停止迭代；否则转步骤 2)。

卷积神经网络每层的输出为若干个特征图(Feature Map)，其形式为$[n_H, n_W, n_C]$。其中 $n_H$ 为特征图的高，$n_W$ 为特征图的宽，$n_C$ 为通道数，即特征图的个数。每层的输出是一个 3 维数组，本文使用 $a^l$ 表示某层卷积神经网络的输出。根据 Gatys 等的研究，使用 Gram 矩阵可以提取图片的纹理特征，即风格[7]。图片 $c$ 与生成图 $g$ 的内容损失为：

$$L_{\text{content}}=\frac{1}{n_H n_W n_C}\parallel a^l(c)-a^l(g)\parallel^2 \tag{1}$$

图片 $s$ 与生成图 $g$ 的风格损失为：

$$L_{\text{style}}=\frac{1}{(2n_H n_W n_C)^2}\parallel G(a^l(s))-G(a^l(g))\parallel_F^2 \tag{2}$$

总体损失为：

$$L_{\text{total}}=\alpha L_{\text{content}}+\beta L_{\text{style}} \tag{3}$$

其中，$\alpha$ 与 $\beta$ 是内容和风格在整体误差中的比重，对这两个值进行调整可以得到侧重点不同的生成图。

上述方法通过梯度下降法迭代优化得到生成图，这样的优化过程最少也需要 1000 次左右的迭代，就算有 GPU 加速也需要 20 min 左右，这完全达不到实时合成的需求。

### 3.2 快速神经风格迁移

快速神经风格迁移是采用转换网络和损失网络的方法将 Gatys 提出的方法进行拆分，即转换网络负责图像风格化，损失网络负责计算损失函数值并优化。这种分而治之的思想极大地提高了图形风格化的速度，满足了实时风格化的需求。快速神经风格迁移的原理图如图 2 所示。

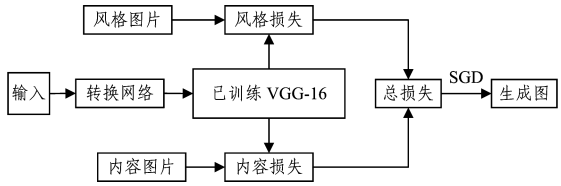



图 2 快速神经风格迁移的原理图

Fig. 2 Schematic diagram of fast neural style transfer

从图 2 可以看出，相比于 Gatys 等提出的方法，该方法多了一个转换网络，输入图像经由转换网络后，会得到一个输出。此输出输入到 VGG-16 损失网络中，类似于 3.1 节的思路，使用 VGG-16 不同层的响应结果计算出内容损失和风格损失，最终求得整体损失。转换网络需要大量的数据集进行训练，该网络只需训练一次即可，Gatys 等的方法则需要每次重新训练网络。此方法节省了大量迭代优化的时间，做到了 3 个数量级的提速。

## 4 方法实现

### 4.1 空间信息约束

为了能够约束风格化图像中物体的空间信息，文中提出了一种特征图空间偏移的方法。使用变换后的 Feature map $T(a^l)$来计算局部相关性，其中 $T$ 表示空间变换，特征图空间偏移及计算方法如图 3 所示。

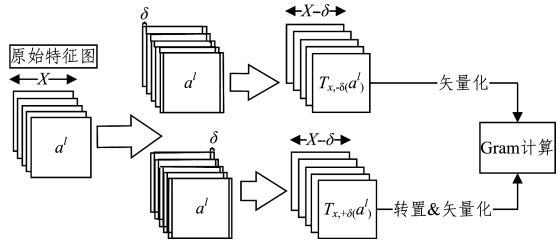



图 3 特征图空间偏移及计算方法原理图

Fig. 3 Schematic diagram of feature map space offset and calculation method

特征图 $a^l$ 的形式为$[n_H, n_W, n_C]$，$n_H$，$n_W$ 和 $n_C$ 分别为特征图的高、宽和通道数。令 $i=n_H$，$j=n_W$，$k=n_C$，$j'$为空间偏移后特征图的宽。根据 Gatys 等的研究，使用 Gram 矩阵可以提取图片的纹理特征，其表达式为：

$$G=\sum_i\sum_j a^l_{ijk}\cdot a^l_{ijk'} \tag{4}$$

如图 3 所示，文中使用 $T_{X,+\delta}$表示原始特征图前 $\delta$ 列被删除后的特征图，$T_{X,-\delta}$表示原始图后 $\delta$ 列被删除后的特征图。简单来说，空间偏移方法就是删除原始特征图的前 $\delta$ 列和后 $\delta$ 列($\delta$ 取值一般为 4)，分别得到图中两个新的特征图，将这两个新的特征图转换后再进行 Gram 计算。

$$G_\delta=Gram(T_{X,+\delta},T_{X,-\delta}) \tag{5}$$

$$G_\delta=\sum_i\sum_{j'}a^l_{ij'k}\cdot a^l_{i(j'+\delta)k'} \tag{6}$$

由式(4)可以看出，Gatys 等的方法只能够提取图片的全局静态特征，即各个特征之间进行内积。比如通道 $k=1$ 和 $k=2$ 的两个特征进行 Gram 计算$k=1$ 特征的$(1,1),(1,2),\cdots,(i,j)$位置和 $k=2$ 特征的$(1,1),(1,2),\cdots,(i,j)$位置对应点

乘求和。这种方法只在不同通道特征的同一位置进行特征提取，因此无法获取足够的空间信息。本文使用的空间偏移方法使得特征在空间上有 $\delta$ 的错位，即位置$(i,j)$和位置$(i,j+\delta)$进行 Gram 计算，这样就可以得到$(i,j)$与$(i,j+\delta)$的互相关性，相当于测量局部特征与其他相邻特征的相关性，这赋予了风格信息额外的空间信息。当然，上图所示的仅为 $X$ 轴方向上的空间偏移。此外，还要添加对 $Y$ 轴方向上的空间限制，计算特征$(i,j)$与特征$(i+\delta,j)$的互相关性，方法与 $X$ 轴方法相同。$s$ 为风格图片，$g$ 为生成图片，$X$ 轴方向偏移损失函数的计算式如下：

$$L_{X,\delta}=\sum_l \| G_{X,\delta}(a^l(s))-G_{X,\delta}(a^l(g)) \|_F^2 \tag{7}$$

$Y$ 轴方向偏移损失函数的计算式如下：

$$L_{Y,\delta}=\sum_l \| G_{Y,\delta}(a^l(s))-G_{Y,\delta}(a^l(g)) \|_F^2 \tag{8}$$

空间偏移损失函数的计算式如下：

$$L_{\delta}=\frac{1}{2}(L_{X,\delta}+L_{Y,\delta}) \tag{9}$$

本文使用此方法对图像风格化过程中物体的空间信息加以限制，为了使风格化图像显示得更为平滑，还加入了全变差损失 $L_{tv}$，参考了 Mahendran 等[8]在特征反演以及超分辨率重建上的研究。本文将快速神经风格迁移的整体损失函数改进为：

$$L_{\text{total}}=L_{\text{content}}+L_{\text{style}}+L_{\delta}+L_{tv} \tag{10}$$

### 4.2　整体架构的设计

为了减少转换网络的计算成本，本文使用残差块设计了一个结构更为精简的图像风格化转换网络，损失网络使用 VGG-19。本文设计的具有空间约束的快速神经风格迁移的整体架构如图 4 所示。

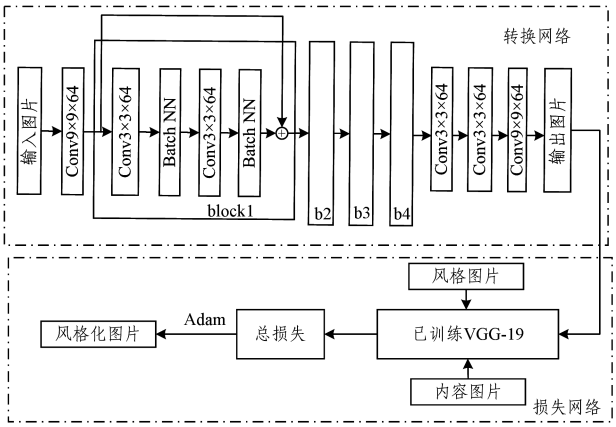



图 4　整体架构图

Fig. 4　Overall architecture

图 4 中的转换网络的第一层和最后一层使用了 9×9 的卷积核，其他所有层都使用了 3×3 的卷积核。架构在第一层之后连续加入了 4 个残差块，每个残差块中的卷积层后面都连着一个 Batch-normalization 层。转换网中所有输出的特征图都是 64 个通道的，并且使用 ReLU 作为每层的激活函数，最后一层使用 tanh 激活函数[9]。损失网络使用 VGG-19 不同层的响应结果计算出内容损失、风格损失等，最终求得整体损失[10]。

本文设计的转换网络更加精简，在保证图片质量的前提下，所有特征图的通道个数设置为 64，并且比 Johnson 等设计的网络架构少了 4 个卷积层和 2 个 Batch-normalization 层，降低了转换网络的计算成本。

## 5　实验结果及分析

### 5.1　实验环境

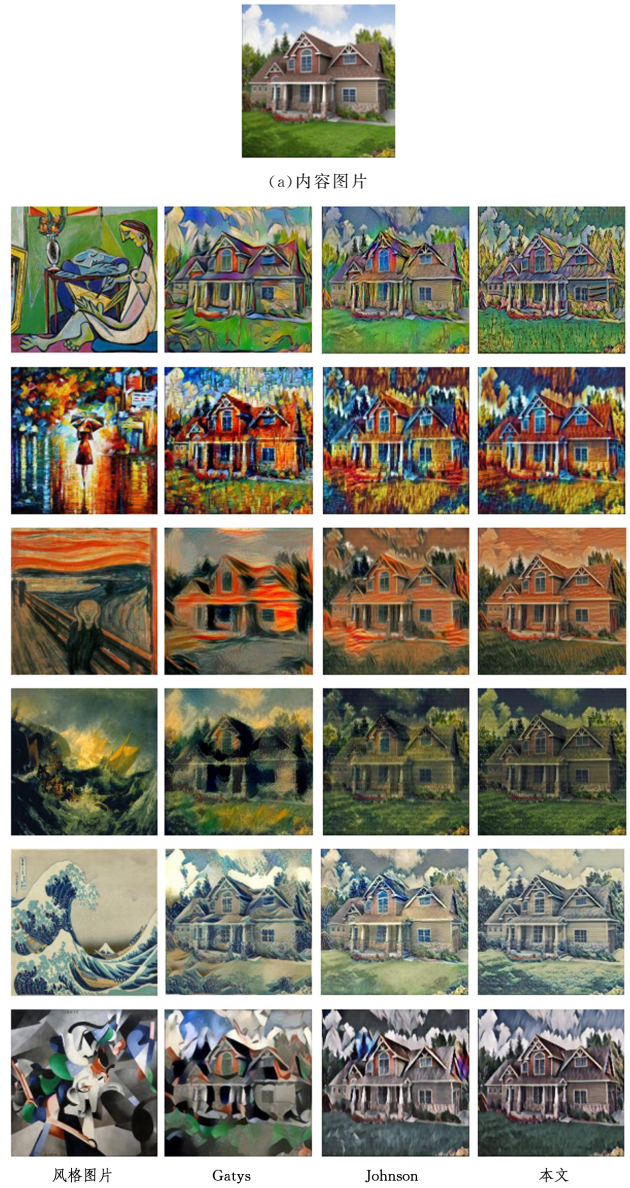
本文的实验环境是 Python3. 6. 4，PC 处理器为 i5 8600K，内存 32GB，Win10 操作系统，同时配备 GTX1060 显卡(6G 显存 GPU)。研究方案主要基于开源的机器学习框架 TensorFlow-gpu、CUDA 和 CuDNN 来实现 GPU 加速。

### 5.2　训练细节

转换网络的训练集来自 MSCOCO 数据集，共 80 000 张图片，batch_size 为 10，迭代次数为 20 000。使用 Adam 优化，学习率为 0. 0005。

### 5.3　结果及分析

将本文提出的空间约束方法与整体架构应用于图像风格化中，并与 Gatys 和 Johnson 等的图像风格化效果进行了对比，实验结果如图 5 所示。




(a)内容图片

(b)

图 5　3 种图片风格化方法的对比

Fig. 5　Comparison of three picture stylized methods

从图 5 可以看出，Gatys 等的风格化图片内的物体都存在扭曲的现象。这是由于 Gram 矩阵只能够提取图片的全局平均特征，对物体空间信息基本无限制作用。Johnson 等使用两个步长为 2 的卷积层来进行下采样，后面跟着 5 个残差块，最后使用两个步长为 1/2 的卷积层进行上采样。虽然网络架构的输入输出相同，但网络获得了较大的感受野，使得风格化过程中物体变形不会太严重。本文使用空间偏移方法，并给出了空间偏移损失函数，以此来进行物体信息空间约束。例如，在第三行实验结果中，Gatys 等的结果中草和云朵的物体信息被抹除，只保留了房子扭曲的轮廓。Johnson 等的方法表现良好，大部分物体的空间信息得到了约束，但图片中房子前的一排花朵的空间信息没有得到很明显的表达。本文风格化图片中的物体菱角分明，很少出现扭曲，物体空间信息也基本全部保留。本文使用局部特征和其他相邻特征的相关性作为物体空间的限制条件，从而赋予了风格信息额外的空间信息，因此即便是一些小的物体，它们在风格化过程中也没有被抹除。

对实验结果从定性分析和定量分析两方面进行。在定性分析中，用随机顺序让参观者观看原始内容图像、风格图像和风格化生成图，然后根据参与者给出的分数得到平均分，该方法是国际电信联盟(ITU)在标准 BT-510 中提出的。评估等级设为 5 个，分别为很差、差、一般、好和很好，用 1－5 分表示。为了减少主观随意性，邀请了 5 名图像处理方向的专业人士和 5 名非专业人士参与了实验。图 6 为观测者给出的平均分数[11]。

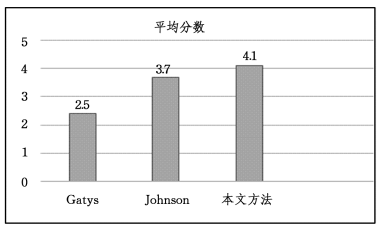



图 6 观测者对实验结果给出的评估结果

Fig. 6 Observer's assessment of experimental results

定量分析主要从图像的风格化速度方面进行比较，因为几种风格化方法的原理并不相同，Gatys 的方法是采用迭代优化，一般迭代次数为 500～1 000 才会得出结果，本文使用 Gatys 迭代 500 次的时间与 Johnson 和本文的方法进行对比，结果如表 1 所列。

表 1 风格化时间表

Table 1 Stylization time table

(单位：s)

| 图片大小 | Gatys | Johnson | 本文 |
|---|---|---|---|
| 256×256 | 21.78 | 0.027 | 0.021 |
| 512×512 | 58.34 | 0.073 | 0.052 |
| 1024×1024 | 225.43 | 0.246 | 0.193 |

本文设计的转换网络更加精简，比 Johnson 等设计的网络架构少了 4 个卷积层和 2 个 Batch-normalization 层，减少了网络的参数，并降低了转换网络的计算成本。本文架构中所有通道数都被设置为 64，参考 Ignatov 等[12]的研究，将网络中的通道数全部设置为 64，这有助于转换网络性能的提升。由图 6 和表 1 可以看出，本文为图片的快速风格化提供了一个较好的解决方案。

**结束语** 本文提出了一种具有空间约束的快速神经风格迁移方法。使用空间偏移方法对风格化图片物体空间信息加以约束，再设计结构更为精简的残差转换网络。这两者的有机结合可以更快速地得到效果更好的风格化图形。文章从定性和定量两方面进行分析，结果表明，所提方法优于一些传统的方法。接下来的工作就是继续优化风格化转换网络，并加强对风格化图片颜色的约束。

## 参 考 文 献

[1] DENG Y Y, TANG F, DONG W M. The status research of stylization of art image[J]. Journal of Nanjing University of Information Technology(Natural Science Edition), 2017, 9(6): 593-598. (in Chinese)
邓盈盈，唐帆，董未名. 图像艺术风格化的研究现状[J]. 南京信息工程大学学报(自然科学版)，2017，9(6)：593-598.

[2] GATYS L A, ECKER A S, BETHGE M. A neural algorithm of artistic style[J]. Computer Science, 2015, 111(1): 98-136.

[3] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016: 2414-2423.

[4] WILMOT P, RISSER E, BARNES C. Stable and controllable neural texture synthesis and style transfer using histogram losses[J]. Computer Science, 2017, 15(1): 96-140.

[5] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual losses for real-time style transfer and super-resolution[C]// European Conference on Computer Vision. Springer, Cham, 2016: 694-711.

[6] ULYANOV D, VEDALDI A, LEMPITSKY V. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis[C]// The IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[7] CAO J F. Research and Implementation of Stylization of Image and Video Based on Deep Learning[D]. Beijing: Chinese Academy of Sciences University, 2017. (in Chinese)
操江峰. 基于深度学习的图像与视频风格化研究与实现[D]. 北京：中国科学院大学，2017.

[8] MAHENDRAN A, VEDALDI A. Understanding deep image representations by inverting them[C]// Proceedings of the IEEE Conference on Computer Vision and PatternRecognition (CVPR). 2015.

[9] CHIKONTWE P, LEE H J. Towards Robust Face Sketch Synthesis with Style Transfer Algorithms[M]// IT Convergence and Security 2017. Singapore: Springer, 2018: 172-179.

[10] ZHU J Y, KRÄHENBÜHL P, SHECHTMAN E, et al. Generative visual manipulation on the natural image manifold[C]// European Conference on Computer Vision. Springer, Cham, 2016: 597-613.

[11] LV L Z, QIANG Y. Medical CT Image Enhancement Algorithm Based on Laplacian Pyramid and Wavelet Transform[J]. Computer Science, 2016, 43(11): 300-303. (in Chinese)
吕鲤志，强彦. 基于 Laplacian 金字塔和小波变换的医学 CT 图像增强算法[J]. 计算机科学，2016，43(11)：300-303.

[12] IGNATOV A, KOBYSHEV N, VANHOEY K, et al. DSLR-quality photos on mobile devices with deep convolutional networks[C]// The IEEE International Conference on Computer Vision (ICCV). 2017.