

基于卷积神经网络的图像局部风格迁移

缪永伟^{1,2} 李高怡¹ 鲍陈¹ 张旭东¹ 彭思龙³

(浙江工业大学计算机科学与技术学院 杭州 310023)¹ (浙江理工大学信息学院 杭州 310018)²
(中国科学院自动化研究所 北京 100190)³

摘要 图像风格迁移是计算机图形学和计算机视觉的一个研究热点。针对现有的图像风格迁移方法中难以对内容图局部区域进行风格迁移的难点,提出了一种基于卷积神经网络的图像局部风格迁移框架。首先,根据输入的内容图和风格图,利用图像风格迁移网络生成全局风格迁移图;然后,利用图像语义分割网络,通过自动语义分割生成的掩码确定图像前景区域与背景区域;最后,利用掩码图确定风格迁移区域并融合未迁移区域得到图像局部风格迁移结果,同时提出一种基于曼哈顿距离的图像融合算法以优化局部风格迁移对象与未迁移区域之间边界的衔接和平滑过渡。该框架综合考虑了目标区域和边界带的像素值、位置等细节信息,在 3 个公开的图像数据集上进行实验,结果表明该方法能够高效、快速并自然地实现输入内容图的局部风格迁移,生成艺术性与真实性和谐并存的视觉效果。

关键词 图像局部风格迁移,深度学习,卷积神经网络,曼哈顿距离,自动语义分割

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2019.09.039

Image Localized Style Transfer Based on Convolutional Neural Network

MIAO Yong-wei^{1,2} LI Gao-yi¹ BAO Chen¹ ZHANG Xu-dong¹ PENG Si-long³

(College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China)¹

(College of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)²

(Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)³

Abstract Image style transfer is a research hot topic in computer graphics and computer vision. Aiming at the difficulty in the style transfer of the local area of the content image in the existing image style transfer method, this paper proposed a localized image transfer framework based on convolutional neural network. First, according to the input content image and style image, the image style transfer network is used to generate the whole style transferred image. Then, the image foreground and the background area are determined by the mask generated by automatic semantic segmentation. Finally, according to style transfer result of the foreground or the background region, an image fusion algorithm based on Manhattan distance is proposed to optimize the convergence and smooth transition between the stylized object and the original area. The framework comprehensively considers the pixel values and positions of the target area and the boundary band, and experiments on three public image datasets demonstrate that the method can efficiently, quickly and naturally implement local style transfer of input content maps, and produce visual effects that are both artistic and authentic.

Keywords Localized image style transfer, Deep learning, Convolutional neural network (CNN), Manhattan distance, Automatic semantic segmentation

1 引言

图像风格迁移(即图像风格化)是指结合给定的内容图 C (Content image)与风格图 S (Style image),生成同时具有内容

特征和艺术风格特征的风格迁移图 R (Result image)^[1],它是计算机图形学和计算机视觉的研究热点。由于在油画绘制、卡通动漫制作、素描画生成、民族文化研究等方面的应用^[2-5],图像风格迁移得到了学术界和工业界的普遍关注。

到稿日期:2018-07-20 返修日期:2018-10-18 本文受国家自然科学基金项目(61272309)资助。

缪永伟(1971—),男,博士,教授,博士生导师,CCF 高级会员,主要研究方向为计算机图形学、数字几何处理、计算机视觉、机器学习,E-mail: ywmiao@zstu.edu.cn(通信作者);李高怡(1994—),女,硕士生,主要研究方向为计算机视觉、机器学习;鲍陈(1983—),男,博士生,主要研究方向为计算机图形学、计算机视觉、机器学习;张旭东(1982—),男,博士,讲师,主要研究方向为计算机图形学、计算机视觉、机器学习;彭思龙(1971—),男,博士,研究员,博士生导师,主要研究方向为计算机视觉、机器学习。

近年来,基于神经网络的深度学习技术推动了图像风格迁移方面的研究^[6-9]。利用可有效提取图像特征的卷积神经网络,Gatys等^[6]提出了图像风格迁移的自动算法,其认为最终生成的风格迁移结果图应该同时具备内容图 C 的内容特征与风格图 S 的风格特征。此后,Gatys等^[7]进一步提出保留原图色彩的图像风格迁移算法,使得最终的生成图虽得到风格图的纹理特征但仍保留了内容图的颜色分布。Johnson等^[8]提出了基于前馈网络的图像风格迁移系统,将迁移系统分为两个不同的网络:图像转换网络和损失函数网络。Vincent等^[9]构建了一种包含多个风格图的转换网络,并指出网络中多种风格的部分参数可以共享。

需要指出的是,上述图像风格迁移方法主要针对输入的整幅图像进行全局风格迁移,尽管用户操作简单方便,但是全局风格迁移需要对整幅图像进行统一操作,无法在视觉效果上突出图像中的感兴趣主体,导致图像风格迁移效果显得单调重复,难以突出重点等;此外,对于一类带有文字信息的图像,经全局风格迁移后的图像中的文字部分可能变得模糊甚至被抹去,从而变得难以辨认。因此,针对特定区域、特定目标进行图像局部风格迁移是一个值得研究的课题。针对已有的图像风格迁移方法^[6-9]中难以对内容图局部区域进行风格迁移的难点,结合自动语义分割和图像风格迁移神经网络,本文提出了一个新颖的图像局部风格迁移框架。该框架利用分开的2个深度卷积神经网络对输入图像进行语义分割与风格迁移,并利用语义分割产生的掩码图来确定图像的背景区域和前景物体;针对图像前景或背景区域进行风格迁移并融合得到图像局部风格迁移效果。该方法能够高效、快速地实现输入内容图的局部风格迁移,且经风格迁移区域边缘优化处理后的图像前景区域与背景区域可以自然地融合,生成艺术性与真实性和谐并存的视觉效果。

本文的主要贡献在于:1)结合语义分割与图像风格迁移算法,提出了一个图像局部风格迁移框架;2)设计并构建了一个端到端的用于图像局部风格迁移的卷积神经网络系统;3)在实现图像的朴素局部风格迁移的基础上,提出了基于曼哈顿距离的优化算法,实现了更好的局部风格迁移融合效果,使得风格迁移区域与原图部分能够平滑地融合;4)与以往手工进行的前景目标提取不同,本文方法无需用户的任何交互操作,亦无需预处理,利用语义分割网络可以自动实现前景与背景的分隔,减少了用户的交互工作量。

2 图像局部风格迁移

图像风格迁移旨在将一幅图像的风格迁移到另一图像上,其难点在于如何获取图像的内容与风格特征。与Gatys等^[6]类似,本文借助卷积神经网络CNN来获取输入图像的内容与风格特征,并利用语义分割产生的掩码图来确定图像背景区域与前景物体,针对图像前景或背景区域进行风格迁移并融合得到图像局部风格迁移效果。

2.1 基于神经网络的风格迁移

将一张尺寸为 $H \times W$ 的图像放入CNN作为输入的风格

图,在网络的第 l 层可获取该层的特征图并表示为 $F_l \in R^{N_l \times H \times W}$,其中 N_l 为第 l 层的卷积核个数。在卷积神经网络的较高层(一般设为Relu3-3),可直接利用该层网络输出的特征图作为图像的内容特征。图像风格特征可由网络输出特征图的统计特性来表示,而统计特性通过度量各维度特性及各维度之间关系的格拉姆矩阵(Gram Matrix)计算,如式(1)所示。

$$G_{ij}^l = \sum_{k=1}^{M_l} F_{ik}^l F_{jk}^l \quad (1)$$

其中, M_l 为第 l 层特征图的长宽乘积。

将内容图 C 、风格图 S 、初始化为白噪点的结果图 R 分别放入卷积神经网络中进行计算,得到对应特征后,通过定义内容损失与风格损失分别度量生成结果图与内容图、风格图之间的差异。定义总损失函数为二者的加权和,如式(2)所示:

$$L = \alpha L_{\text{content}} + \beta L_{\text{style}} \quad (2)$$

其中,内容损失 L_{content} 为内容图 C 与风格图 S 在网络某特定层(一般设置为Relu3-3)特征值的平方误差,风格损失 L_{style} 则考虑了Relu1-1到Relu4-3之间的每一层,其定义为风格图与结果图的各层特征图的格拉姆矩阵之间平方误差的和。

在神经网络的训练过程中,使用随机梯度下降算法^[15]来优化损失函数,当其达到最优时结果图即被迭代优化为最终的风格迁移效果图。

2.2 结合语义分割的图像局部风格迁移

针对输入原图,采用图像语义分割网络,将全连接神经网络FCN^[10]与条件随机场CRF^[11]相结合,分割得到含有语义标签的语义分割图。为了获取输入图像的前景与背景,依据分割图标签进行二值化,生成掩码图,使掩码图中每个像素值均被设置为0或1,其中值为0的像素点集对应于在风格迁移过程中希望保持不变的区域,而值为1的像素点集对应于原图中待风格化的区域。考虑到不同的风格化需求,该方法将生成2张取值相反的掩码图像,并分别用于生成前景风格迁移结果图与背景风格迁移结果图。

在优化损失函数以生成风格图时,本文借鉴Johnson等^[8]提出的快速风格迁移网络结构,将生成问题转化为转换问题来实现风格迁移。整个风格迁移过程可分为图像转换网络与损失网络两部分,其中损失网络采用了在ImageNet^[12]上训练好的VGG-19网络模型^[13],最终的损失函数采用式(2)中定义损失函数。与Gatys等^[6]不同的是,为了减少计算量、增大接收场大小,风格迁移结果图不再是从一张初始化为白噪点的图像一步步经网络迭代生成,而是经由前馈图像转换网络快速生成,该网络分为下采样块 D (Down sampling)、残差网络块 R (Residual Net)^[14]以及上采样块 U (Up sampling)3部分。下采样部分包含2个步长为2的卷积层,若输入一张大小为 $H \times W$ 的内容图,经下采样后其特征图的长宽尺寸均缩小为原来的1/4,从而减少了特征映射的空间范围;若干残差网络块增加了网络深度,提升了网络性能,但并不会改变特征图的尺寸。上采样过程含有2个步长为1/2的卷积层,经此,结构特征图又恢复到 $H \times W$ 的大小,实现了在不改变特征图尺寸的前提下获取内容特征与风格特征。

本文的图像局部风格迁移方法的流程如图 1 所示,整个框架被分为语义分割网络和图像风格迁移网络两部分。内容图像 C 与风格图像 S 经图像风格化网络得到全局风格迁移图 R_{whole} ,内容图像 C 经语义分割网络后得到分割图 C_{seg} ,利用分割图确定风格迁移区域,并融合未迁移区域得到图像局部风格迁移结果。

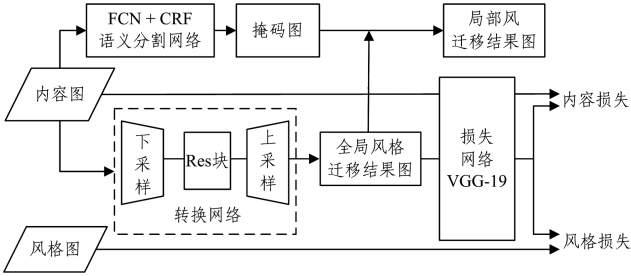


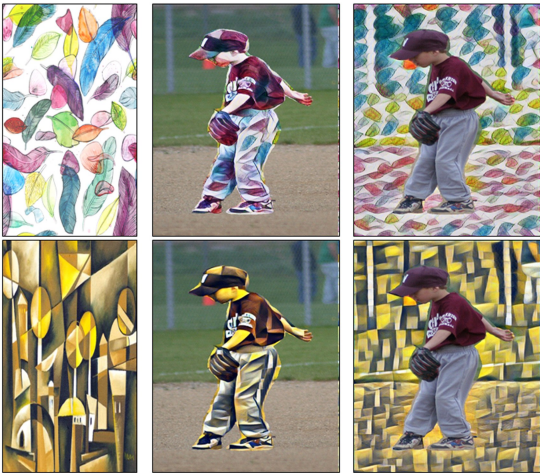
图 1 图像局部风格迁移框架

Fig. 1 Framework of localized image style transfer

图 2 给出了图像局部风格迁移的实例,其中每输入一张风格图就输出两张对应的局部风格迁移结果图。



(a) 内容图



(b) 风格图 (c) 前景风格迁移结果图 (d) 背景风格迁移结果图

图 2 图像朴素局部风格迁移的效果

Fig. 2 Effect of simple localized image style transfer

从实验结果可知,对背景进行风格迁移时,人像边界附近本该属于背景区域的像素却保留了原像素值;同样地,当对前景进行风格迁移时,部分属于背景区域的像素也出现了被风格迁移的情况。由此可见,利用图像的朴素局部风格迁移方法虽然简单快速,但最终的风格迁移效果几乎完全取决于图像语义分割的质量,其风格迁移区域与背景之间的边缘杂乱且不清晰,有时还会出现边界不规则地扩大或缩小以及许多

噪点,且边界附近有大量本该属于背景部分的像素却呈现出与前景一致的风格,该结果不符合用户对图像局部风格化的预期,前景区域边缘需平滑之后才能达到更自然的视觉效果。

2.3 基于曼哈顿距离的边界平滑优化

为了使图像局部风格迁移后的风格迁移区域像素能够平滑地融合到原始图像未被风格迁移的区域中,需要对语义分割得到的图像前景部分进行边缘优化处理,并通过优化方法实现风格迁移区域与原始图像的光滑、无缝融合。具体地说,本文在可能的像素标签中(前景或背景)为图像的每个像素选择合适的标签,融合时要求:1)风格迁移区域和原始图区域的边界应靠近原始分割图边界;2)融合时要减少突兀转变,以使得两种风格尽可能平滑地互相嵌入。本文改进方法的思路是在图像语义分割掩码的边缘处定义一个边界窄带。将窄带外的像素点的标签固定为图像语义分割结果分配到的标签,因此只需对边界窄带内的像素点进行标签赋值就可以达到平滑的效果,此处标签可分为前景和背景。对于边界窄带内坐标为 (x, y) 的像素点 P ,须得到其标签值 K 。常规做法是将标签值设置为 0 或 1 的离散值,但考虑到须得到的并不是一个保留完整边界线轮廓的前景物体,而是 2 种不同风格局部区域的平滑融合,语义分割算法无法做到精准分割,通过非黑即白的判别标准判别标签为前景或背景是不合理的,因此本文将标签 K 设定为一个定义在 $[0, 1]$ 上的连续值,使用 l^0 与 l^1 分别代表该像素的标签为背景和前景的概率。以背景风格迁移为例,边界窄带内越靠近背景区域的像素点应该有更大的风格图的比重,而越靠近前景物体的点应保留更多的原图内容,融合时像素的原始值 P_0 与风格迁移后的值 P_s 用式(3)融合为新点 P :

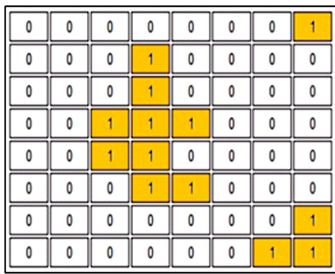
$$P = l^0 P_0 + l^1 P_s = l^0 P_0 + (1 - l^0) P_s \quad (3)$$

因此,待解决问题即为如何确定这样的边界带以及如何获取 P 点的标签概率,使得越靠近前景物体的点被判定为前景物体的概率越大,即 l^1 越大。基于此,引入曼哈顿距离:曼哈顿距离(Manhattan Distance)用于几何学的度量空间^[15],其表示为像素点 i 与像素点 j 在标准坐标系上绝对轴距的总和,具体计算公式如下:

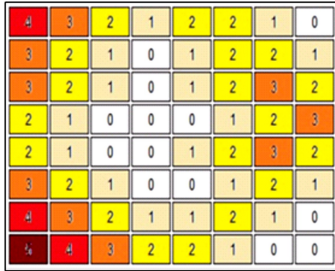
$$d_{ij} = |x_i - x_j| + |y_i - y_j| \quad (4)$$

其中, x_i 与 y_i 分别表示像素点 i 的横纵坐标值,同样地, x_j 与 y_j 分别表示像素点 j 的横纵坐标值。本文将曼哈顿距离应用到图像分割后的二值掩码图像中。如图 3 所示,图 3(a)为语义分割后处理得到的掩码图,其中前景部分的像素值为 1,其余部分为 0,则其余部分到前景物体的曼哈顿距离图 D 即如图 3(b)所示。曼哈顿距离图的求解思想可以归结为:1)原二值图中所有像素值为 1 的像素点在曼哈顿距离图 D 中的值为 0;2)对于任意原二值图中像素值非 1 的像素点 P ,设其 4-邻域(分别为上、下、左、右)像素点在距离图 D 中的值为 d_0, d_1, d_2, d_3 ,则像素点 P 在距离图 D 中的值确定为 $D(P) = \min\{d_0, d_1, d_2, d_3\} + 1$,表示 P 点必定经过 4-邻域中的一点才能达到值为 1 的像素点。这里利用动态规划方法求解曼哈顿距离图,如算法 1 所示。曼哈顿距离图求解算法的

时间复杂度为 $O(n^2)$, $n = \max(H, W)$ 。



(a) 原二值图像



(b) 曼哈顿距离图

图3 原二值图像与曼哈顿距离图的对比

Fig. 3 Comparison between original binary image and Manhattan distance map

算法1 曼哈顿距离图求解算法

输入: 背景掩码图 $MASK_{back}$

输出: 曼哈顿距离图 D

1. 新建与 $MASK_{back}$ 相同尺寸的矩阵 D 用来存放曼哈顿距离。
2. 对 $D[i, j]$ 赋初值: 遍历 $MASK_{back}$, 对于 $[i, j]=1$ 的点, 令 $D[i, j]=0$, 其余点均令 $D[i, j]=height * weight$, 即距离最大值。
3. 正向遍历 D , 若 $D[i-1, j] > D[i, j-1]$, 则 $temp = D[i, j-1] + 1$, 否则 $temp = D[i-1, j] + 1$; 接着若 $D[i, j] > temp$, 则 $D[i, j] = temp$, 否则 $D[i, j]$ 保持不变, 即从左到右、从上到下, 比较点 P 与其左、上邻域点的距离大小, 并选择 $D[i-1, j] + 1, D[i, j], D[i, j-1] + 1$ 这3个值中的最小值赋给 $D[i, j]$ 。
4. 逆向遍历 D , 若 $D[i+1, j] > D[i, j+1]$, 则 $temp = D[i, j+1] + 1$, 否则 $temp = D[i+1, j] + 1$; 接着若 $D[i, j] > temp$, 则 $D[i, j] = temp$, 否则 $D[i, j]$ 保持不变, 即从右到左、从下到上, 比较点 P 与其右邻域点和下邻域点的距离大小, 选择 $D[i+1, j] + 1, D[i, j], D[i, j+1] + 1$ 这3个值中的最小值赋给 $D[i, j]$ 。
5. 由步骤3和步骤4更新得曼哈顿距离图 D 。

根据计算得到的曼哈顿距离图, 可以按照如下方法确定图像中边界带像素的背景标签概率。设定一个阈值 M 用以获取边界带, 所有曼哈顿距离 $d < M$ 且不为0的像素点即为边界内的点。以背景风格化为例, 在 d 从 M 变为0的过程中, 像素点对应的背景标签概率 l^1 应同样从1下降为0, 为了使过渡更平滑, 将 d 与 l^1 之间的关系对应为:

$$l^1 = \sqrt{\frac{d}{M}} \quad (5)$$

因此, 边界带内像素点 P 的最终值为:

$$P = l^0 P_0 + l^1 P_s \\ = (1 - \sqrt{\frac{d}{M}}) P_0 + \sqrt{\frac{d}{M}} P_s \quad (6)$$

实验表明, 阈值 M 设为5时能产生更好的光滑、无缝的融合效果, 优化后的结果如图4所示。



(a) 优化后的前景风格迁移



(b) 优化后的背景风格迁移

图4 优化后的图像局部风格迁移效果

Fig. 4 Effect of localized image style transfer after optimization

3 实验结果与讨论

为了验证方法的有效性, 在 Ubuntu 16.04 系统下, 使用基于 Python 的 Tensorflow 框架^[16], 实现本文提出的图像局部风格迁移方法。采用预训练网络方式, 使用 ImageNet 数据集上预训练好的 VGG-19 网络^[13]。机器配置为 i7-7700 CPU, 8GB 内存和 GeForce GTX 1070 6G 显卡图形工作站。

由于目前国内外尚未有专门用于图像风格迁移的公开数据集, 为了保证方法的通用性, 本文选用计算机视觉领域的3个标准数据集: PASCAL-VOC 数据集¹⁾、COCO-2014 数据集²⁾和 CIFAR-100 数据集³⁾, 每个数据集的详细介绍见3.1节。在每个数据集中随机选取图像进行朴素局部风格迁移与优化局部风格迁移, 分析并比较实验结果。在卷积神经网络的图像风格迁移领域, 由于艺术风格的效果感受具有一定主观性, 难以用量化方法来评估其性能, 为了衡量本文方法的有效性, 检验利用本文方法得到的图像局部风格迁移的迁移区域的边缘是否清晰且平滑地嵌入, 将采用融合优化算法得到的融合效果与朴素融合方法得到的融合效果进行对比, 并将本文方法与其他方法进行比较。

3.1 数据集

1) PASCAL-VOC 数据集: 该数据集包含20个类别, 包括人类、鸟、猫、牛、狗等。该数据集共有11530张图片, 并且有27450个带有注释物体的ROI以及6929张图片的分割实例。

2) COCO-2014 数据集: 该数据集共有82783张图片, 包

¹⁾ <http://host.robots.ox.ac.uk/pascal/VOC/voc2012/>

²⁾ <http://mscoco.org/>

³⁾ <http://www.cs.toronto.edu/~kriz/cifar.html>

括 91 种类别,其特点是每一类图像较其他数据集更多,这有利于获得每类目标的更多特征。

3) CIFAR-100 数据集:该数据集含有 100 个类别的 6 万张图像,这 100 个类别又分为 20 个大类,每个图像带有一个小类的“fine”标签与一个大类的“coarse”标签。

3.2 朴素局部风格迁移和优化局部风格迁移

图像的朴素局部风格迁移效果与基于曼哈顿距离的优化融合效果如图 5、图 6 所示。实验采用了在 ImageNet^[12] 数据集上训练好的 VGG-19^[13] 网络。实验结果表明,曼哈顿距离阈值 M 取 5 时能产生较好效果。

观察图 5 与图 6 中的图(c)与图(d)可知,本文首先提出的朴素局部风格迁移虽能完成前景/背景的风格迁移工作,但在边界处常出现由于分割不精准导致的风格迁移“溢出”或“不足”的现象。图 5 中将风格图油画“星空”迁移至内容图“鸟”上,图像的朴素局部风格迁移结果显示,在鸟的身体周边出现了杂乱不清晰的风格效果(见图 5(c)),这是由于分割网络无法做到完全精准地分割。同样地,图 6(c)中小孩耳朵在背景风格迁移时呈现出与背景一致的风格,这些问题在实际运用中均应该避免。观察图 5(d)与图 6(d)可以发现:经本文提出的基于曼哈顿距离的优化融合后,风格迁移结果的视觉效果更加自然,风格迁移区域与原图区域之间的边界更平滑;且观察融合边界处的图像细节可以发现,风格迁移的区域与原图区域以细腻的相互嵌入方式进行融合,并没有出现明显的分割痕迹,局部风格迁移效果更好,可以实现很好的艺术感与真实性并存的滤镜效果。

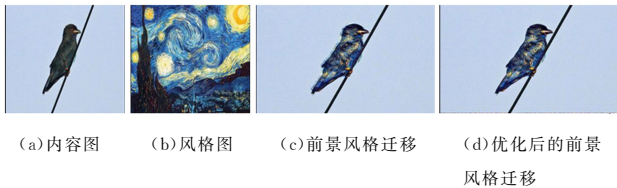


图 5 图像“鸟”分别利用朴素局部风格迁移与优化局部风格迁移的效果比较

Fig. 5 Comparison of style transfer for image “bird” using simplicity and optimizing schemes



图 6 图像“小孩”分别利用朴素局部风格迁移与优化局部风格迁移的效果比较

Fig. 6 Comparison of style transfer for image “child” using simplicity and optimizing schemes

3.3 方法比较与讨论

为了比较本文方法与其他方法的效率,实验从上文提及的 3 个数据集中随机选取 200 张图像,每张图像分别进行 3 次风格转移操作,另外 2 种方法分别为 Gatys 迭代生成法^[6]

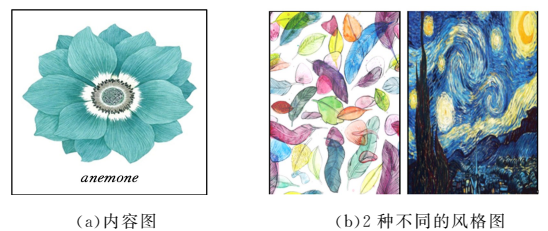
(迭代次数为 1000)和 Johnson 方法^[8],分别统计 200 张图像利用 3 种方法处理的平均耗费时间,结果如表 1 所列。从表 1 可以看出,Gatys 方法所需时间明显比其他方法要多几个数量级,本文方法在 Johnson 方法的基础上增加了语义分割与局部风格迁移的边界平滑优化,处理速度稍慢,但取得了很好的图像局部风格迁移效果。

表 1 3 种不同方法的时间统计

Table 1 Time statistics for three different methods (单位:s)

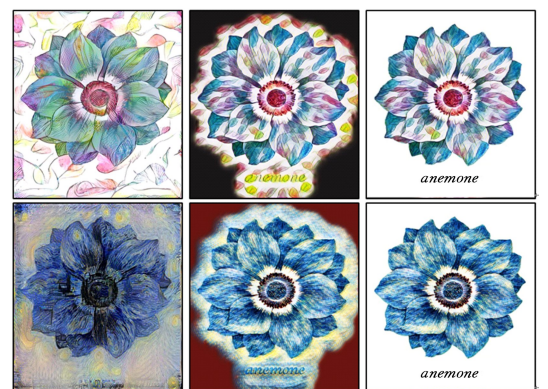
方法	Gatys ^[6]	Johnson ^[8]	Ours
平均时间	938	1.8	2.1

由于目前在基于卷积神经网络的图像风格迁移方面尚未有局部风格迁移的工作,因此这里将本文方法风格迁移效果与 2 种不同的整体风格迁移方法的效果进行比较,如图 7 所示。图 7(a)中的输入内容图是一幅背景为白色且下方带有文字的花朵图案,图 7(b)是实验分别采用的 3 组不同风格的图,利用 Gatys 方法^[6]、Johnson 方法^[8]和本文方法得到的图像风格迁移效果分别如图 7(c)、图 7(d)和图 7(e)所示。图 7(c)中,利用 Gatys 方法得到的风格迁移效果较为均匀细腻,能够精细地模拟风格图的细节,但花瓣线条发生了轻微的扭曲和形变,原图中的文字也已经消失。图 7(d)中,利用 Johnson 方法对花朵进行了有效的风格迁移,但花朵之外的背景部分出现了突兀的颜色,且背景与花朵之间的衔接部分十分粗糙,原图中的文字也变得模糊不可辨认。本文方法可以有效避免这些问题,如图 7(e)所示,利用本文方法进行图像风格迁移得到的结果图像中花朵的形状得以完整保存,且花朵部分实现了有效的风格迁移,部分与部分之间的衔接也实现了平滑精细的效果,原图中的文字在风格迁移效果图中依然完整存在。



(a) 内容图

(b) 2 种不同的风格图



(c) Gatys 方法^[6]

(d) Johnson 方法^[8]

(e) 本文方法

图 7 文字保留的风格迁移效果比较

Fig. 7 Comparisons of text preserved style transfer

结束语 针对已有的图像风格迁移中难以对内容图局部区域进行风格迁移的难点,本文结合自动语义分割和图像风格迁移2个深度卷积神经网络,提出了一种新颖的图像局部风格迁移方法。该方法在利用图像风格迁移网络生成整体风格迁移图和基于图像语义分割网络得到的语义分割图的基础上,通过图像转换网络与损失网络,利用优化内容损失和风格损失实现图像的局部风格迁移;同时提出了一种基于曼哈顿距离的图像融合算法,以优化局部风格迁移对象与相邻未迁移区域之间边界的衔接和平滑过渡。实验表明,本文方法能够高效、快速地实现输入内容图的局部风格迁移。与已有方法相比,图像的局部风格迁移减少了突兀感与不自然感,且经基于曼哈顿距离的优化融合,风格图像的前景区域与背景区域可以实现自然的融合,生成艺术性与真实性和谐并存的视觉效果。

未来将考虑针对视频的局部风格迁移问题,特别将探索如何在视频中实现对前景物体的精准定位与实时分割,解决视频中帧与帧之间的风格连贯性,克服视频中由于位置、光线、视角等变化引起的闪烁、抖动等问题。

参 考 文 献

- [1] GOOCH B, GOOCH A. Non-photorealistic rendering [M]. New York: AK Peters/CRC Press, 2001.
- [2] DECARLO D, SANTELL A. Stylization and abstraction of photographs [J]. ACM Transactions on Graphics, 2002, 21(3): 769-776.
- [3] WANG J, XU Y, SHUM H Y, et al. Video toning [J]. ACM Transactions on Graphics, 2004, 23(3): 574-583.
- [4] LI P, SUN H Q, SHENG B, et al. Image stylization with enhanced structure on GPU [J]. Science China Information Sciences, 2012, 55(5): 1093-1105.
- [5] ZANG Y, HUANG H, LI C F. Stroke style analysis for painterly rendering [J]. Journal of Computer Science and Technology, 2013, 28(5): 762-775.
- [6] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2016: 2414-2423.
- [7] GATYS L A, BETHGE M, HERTZMANN A, et al. Preserving color in neural artistic style transfer [J]. arXiv: 1606. 05897, 2016.
- [8] JOHNSON J, ALAHI A, LI F F. Perceptual losses for real-time style transfer and super-resolution [C]// Proceedings of European Conference on Computer Vision. Cham: Springer, 2016: 694-711.
- [9] DUMOULIN V, SHLENS J, KUDLUR M, et al. A learned representation for artistic style [J]. arXiv: 1610. 07629, 2016.
- [10] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(4): 640-651.
- [11] LAFFERTY J D, MCCALLUM A, PEREIRA F C N. Conditional random fields: probabilistic models for segmenting and labeling sequence data [C]// Proceedings of Eighteenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc., 2001: 282-289.
- [12] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2009: 710-719.
- [13] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv: 1409. 1556, 2014.
- [14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2016: 770-778.
- [15] MALKAUTHEKAR M D. Analysis of euclidean distance and Manhattan Distance measure in face recognition [C]// International Conference on Computational Intelligence and Information Technology. IET, 2013: 503-507.
- [16] ABADI M, AGARWAL A, BARHAM P, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems [J]. arXiv: 1603. 04467, 2016.