



## A Oil Painting Style Migration Algorithm Based on Convolutional Neural Network

---

Jiang Min, Sheng Ran, Zhu De, Duan Yunsheng, Li Fangfang  
and Sun Dong

EasyChair preprints are intended for rapid  
dissemination of research results and are  
integrated with the rest of EasyChair.

April 23, 2019

# 一种基于卷积神经网络的油画风格迁移算法

江敏<sup>1</sup>, 盛然<sup>1</sup>, 竺德<sup>1</sup>, 段运生<sup>1</sup>, 李芳芳<sup>2</sup>, 孙冬<sup>2</sup>

<sup>1</sup>安徽大学网络信息中心, 安徽 合肥 230601

<sup>2</sup>安徽大学电气工程与自动化学院, 安徽 合肥 230601

(zhude@ahu.edu.cn)

## A Oil Painting Style Migration Algorithm Based on Convolutional Neural Network

Jiang Min<sup>1</sup>, Sheng Ran<sup>1</sup>, Zhu De<sup>1</sup>, Duan Yunsheng<sup>1</sup>, Li Fangfang<sup>2</sup> and Sun Dong<sup>2</sup>

<sup>1</sup>(Network Information Center, Anhui University, Hefei, 230601)

<sup>2</sup>(College of Electrical Engineering and Automation, Anhui University, Hefei, 230601)

**Abstract** The oil painting style is to add the sand painting style information of an image to any image, and maintain the semantic content of the image to produce a new ornamental image, mainly introducing a convolutional neural network based image. This paper mainly introduces an image style algorithm based on convolutional neural network, which can separate and reorganize the image content and style of natural pictures, and then realize the oil painting migration of pictures. The algorithm combines the content of any image with many well-known oil painting styles to achieve high-perceptual quality artwork. The experimental results show that the algorithm can achieve deep fusion of image content and optimization style, which proves the effectiveness of the algorithm.

**Key words:** convolutional neural network; image stylization; oil painting

**摘要** 油画风格是把一张图像的油画风格信息加到任意某张图像上, 并保持该图像的语义内容, 产生一个新的具有观赏性的图像。本文主要介绍了一种基于卷积神经网络的图像风格算法, 可以分离和重组自然图片的图像内容与风格, 进而实现图片的油画化迁移。该算法可以将任意图像的内容与众多知名油画风格相结合, 获得高感知质量的艺术作品。通过实验验证该算法能够实现图像内容与优化风格的深度融合, 证明了该算法的有效性。

**关键词** 卷积神经网络; 图像风格化; 油画

中图法分类号 TP18

油画起源于 15 世纪的欧洲, 由荷兰人发明, 凭借颜料的遮盖力和透明性能充分的表现描绘对象, 色彩丰富、立体质感强。500 多年来, 油画这一绘画形式一直被画家们所青睐, 尤其是《蒙娜丽莎》、《最后的审判》、《亚当与夏娃》等作品深受人们的喜爱。油画是用亚麻仁油或核桃油、罂粟油等一些快干油及经过处理的玛蒂树脂或达玛树脂等, 调合各种颜色粉制成油画颜料和媒介剂在亚麻布、木板或厚纸板上作画

的一种绘画形式。油画的绘制不仅需要繁复的材料准备工作, 更需要扎实的绘画基础, 对于普通人来说, 想完成一幅专业的油画作品是非常困难的。随着互联网发展迅猛, 人们利用计算机将图片转换成不同的风格样式, Photoshop 等专业处理图像软件可以对照片进行色彩渲染和曲线调整, 将照片渲染成油画风格。近期, 来自俄罗斯莫斯科的 PRISMA 实验室开发团队研发了一款照片编辑器 Prisma, 通过图像风格化将普

收稿日期: yyyy-mm-dd [Date 六号] 修回日期: yyyy-mm-dd [Date 六号]

基金项目: 赛尔网络下一代互联网技术创新项目(NGII20180312)。

通的照片与众多油画艺术家作品的风格结合起来,转变成一幅具有高感知、美化性、多风格性等的油画艺术照。

图像风格化也叫风格迁移,传统的风格迁移通常通过建立数学或统计模型、笔触、纹理等对图像进行风格化处理。Winnemoller<sup>[1]</sup>等人提出一种自动实时图像抽象框架,通过调整亮度和颜色对比度(对高对比度部分使用更高的视觉显著性,平滑低频部分),创造出具有卡通风格的图像,但是在风格多样性方面受到限制。Hertzmann 等人<sup>[2]</sup>是提出的已知一对源图像和目标图像图像类比算法,通过学习源图像示例之间的映射关系,得到目标图像的风格化效果图,但是图像类比算法只捕获到图像的低级特征,无法获得有效的样式,效果不理想。Ashikhmin<sup>[3]</sup>提出的快速纹理传输致力于传输高频纹理信息,同时保留目标图像的粗尺度。Lee<sup>[4]</sup>等人通过附加地通知具有边缘方向信息的纹理转移来改进该算法。由于仅仅使用了像素级的低层图像特征,在灵活性、有效性、多样性等方面都具有很大的局限性,因此这些算法的普适性不强。尽管这些方法,并且针对特定的任务目标,往往需要大量的手工调参。

本文介绍的风格化算法其核心是 Gatys<sup>[5]</sup>等人提出的基于卷积神经网络的艺术风格算法,卷积神经网络中内容和风格的表示是可分离的,进而在此基础上将两种表示独立处理,产生新的图像的风格迁移。本文包括以下主要研究内容:第1部分主要介绍卷积神经网络模型-VGG16模型;第2部分重点讲解图像油画风格迁移算法;第3部分通过实验验证不同的沙画作品与不用的图像采用本文算法进行风格迁移结果进行对比分析,测试该算法的正确性和合理性;最后总结本文算法的概括与总结。

## 1.卷积神经网络—VGG16模型

卷积神经网络在图像分类、语义分割、特征提取、风格迁移等方面有着广泛应用。如 chen<sup>[6]</sup>等人提出了一种基于注意力机制的卷积神经网络解决复杂场景中的图像分类问题,该模型精通细粒度分类问题,具有更好的鲁棒性。卷积的过程就是特征提取的过程,每一个卷积核代表着一种特征,如果图像中某块区域与某个卷积核的结果越大,那么该区域就与该卷积核越相似。图像风格化选用的是不带全连接层的VGG16模型,包含13个卷积层和5个最大池化层,通过不断加深网络结构来提高性能<sup>[7]</sup>。VGG16模型共分

5层,前两层中每层经过两次卷积后再池化,后三层是经过三次卷积后再池化,每次卷积后都伴有激活函数,在这里更多的卷积核可使决策函数更具有辨别能力。整体的卷积核大小均为 $3\times 3$ ,3个卷积层串联的效果相当于一个 $7\times 7$ 的卷积层,可以减少卷积层的参数量,同时也拥有更多的非线性变换,使CNN对特征的学习能力更强。这里采用池化是最大池化,池化核的大小为 $2\times 2$ ,maxpool更容易捕捉图像上的变化,梯度的变化,带来更大的局部信息差异性,更好地描述边缘、纹理等构成语义的细节信息,这点尤其体现在网络可视化上,通过在输入空间的梯度上升,可视化VGG16的滤波器。该模型是公开的,可以在caffe-framework<sup>[8]</sup>中进行下载,每个滤波器的可视化结构图1所示,可以看出低层的卷积核对颜色,边缘信敏感度更高,越高层的卷积核,内容也越抽象复杂。

通过卷积神经网络提取图像特征,用不同的卷积核提取不同的特征。给定输入图像在CNN中的每个处理阶段被处理成一组特征图像。其中底层主要反映像素级的底层特征,而高层主要反映图像的边缘结构。

## 2 图像油画风格迁移算法

风格化的关键之处是该系统使用神经表示来分离和重组任意图像的内容和风格,提供神经算法创造艺术形象,通过找到同时匹配照片的内容表示和相应艺术品的风格表示的图像来合成图像。

### 2.1 内容和风格重建

卷积神经网络以前馈方式分层处理视觉信息。每层单位为图像滤波器的集合,每个图像滤波器从输入图像中提取某个特征,输入不同的滤波器,得到不同的特征映射。卷积神经网络经过训练后物体信息沿着处理层次越来越明确,沿着网络的处理层次,输入图像越来越注重图像的实际内容与其详细的像素值相比较。通过仅从该层中的特征图重建图像<sup>[9]</sup>来可视化每层包含的关于输入图像的信息。卷积神经网络的风格化算法是一种纹理传输算法,它通过最先进的卷积神经网络的特征表示来约束纹理合成方法。通过执行预图像搜索来匹配示例图像的特征表示来生成新图像。这种通用方法之前已经用于纹理合成<sup>[10]</sup>的背景,并且提高了对深度图像表示<sup>[11]</sup>的理解。样式转换算法结合了卷积神经网络的参数纹理模型<sup>[12]</sup>,通过这种方法来转换它们的图像表示。

网络中的较高层根据对象及其在输入图像中的排列捕获高级内容,来自较低层的重建简单地重现了原始图像的精确像素值。我们将网络高层中的特征响应称为内容表示。我们可以在CNN的不同处理阶段

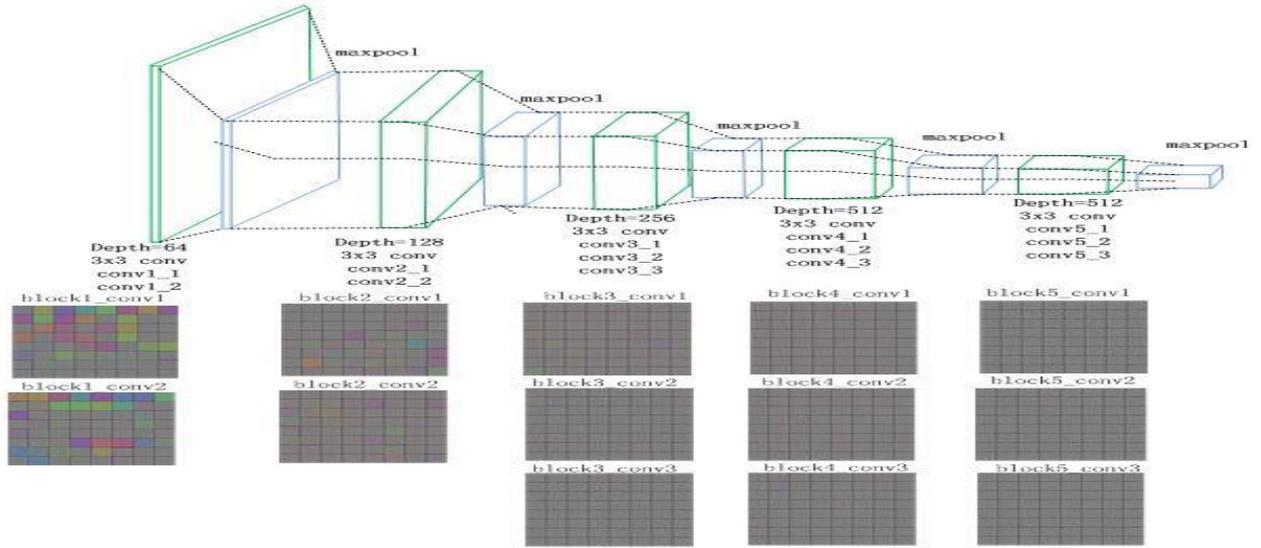


图1 VGG16 可视化模型结构图

Fig.1 VGG16 visualization model structure

将信息可视化, 只需知道网络在特定层中的响应即可重建输入图像, 如图2所示, 上半部分是风格重建, 越是高层的特征, 风格重建的就越粗粒度化, 下半部分是内容重建, 越是底层的特征, 重建的效果就越精细, 越不容易变形, 高层神经网络在提取了图片的风格, 比如色彩和排列, 而底层神经网络提取的则是图像的内容, 也就是图像的细节。从原始图层的“conv1” - “conv5”重建输入图像 VGG 的网络。我们发现从下层重建几乎是完整的 (“conv1” - “conv3”)。在网络的较高层中, 详细的像素信息丢失, 而图像的高级内容被保留 (“conv4” - “conv5”)。为了获得输入图像的风格表示, 我们使用了特征空间, 这些特征空间最初是用来捕捉纹理信息的。该特征空间由每层滤波器的响应来构建。它由特征映射的空间范围内的不同滤波器响应之间的相关性组, 通过在每一层加入这种相互关系, 我们能够关于输入图像的一个稳定的、多尺度的表示, 该表示可以捕获纹理信息而不是全局布置。

## 2.2 内容与风格表示

合成图像的风格内容与另外两幅分别提供内容和风格的图像有些差异, 图像内容和风格不能完全分开, 通常不存在同时完全匹配两个约束的图像, 我们会根据关于内容和风格的损失函数来对生成的图像进行优化。网络中的每个层都定义了非线性滤波器组, 其复杂度随着网络中层的位置而增加。给定的输入图像  $x$  通过对该图像的滤波器响应被编码在卷积神经网络的每一层中。例如在第  $l$  层有  $N_l$  个不同滤波器, 经过卷积会得到  $N_l$  个不同的特征图, 每个特征图的大小为  $M_l$ 。层中的响应可以存储在矩阵  $F^l \in R^{N_l \times M_l}$  中,

其中  $r_{ij}^l$  是第  $l$  层中位置  $j$  处的第  $i$  个滤波器的激活。

为了可视化在层次结构的不同层上卷积图像信息, 可以在白噪声图像上执行梯度下降, 以找到与原始图像的特征响应相匹配的另一图像<sup>[13]</sup>。假设  $\bar{p}$  和  $\bar{x}$  是原始图像和生成的图像,  $p_l$  和  $F_l$  是分别是它们在第  $l$  层中的特征表示, 用两个特征表示之间的平方误差损失评价内容损失, 内容损失相对于第  $l$  层激活的导数如公式 (2), 可以使用标准误差反向传播来计算相对于图像  $\bar{x}$  的梯度, 根据梯度改变最初的随机图像  $\bar{x}$ , 直到它在卷积神经网络某一层产生与原始图像  $\bar{p}$  相同的响应。

$$L_{content}(\bar{p}, \bar{x}, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^l - P_{ij}^l)^2 \quad (1)$$

$$\frac{\partial L_{content}}{\partial F_{ij}^l} = \begin{cases} (F_{ij}^l - P_{ij}^l) & \text{if } F_{ij}^l > 0 \\ 0 & \text{if } F_{ij}^l < 0 \end{cases} \quad (2)$$

在网络每一层的卷积响应建立一个样式表示法, 用于计算不同滤波器响应之间的相关性, 这些特征相关性由格雷姆矩阵  $G_l \in R^{N_l \times N_l}$  表示,  $G_{ij}^l$  为矢量化特征图之间的内积

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l \quad (3)$$

为了生成与输入图像样式相匹配的纹理, 使用白噪声图像的梯度下降来更新与输入原始图像样式表示符合的另一个图像, 通过最小化输入原始图像的格拉姆矩阵与生成图像的格拉姆矩阵之间距离的平方来表示。

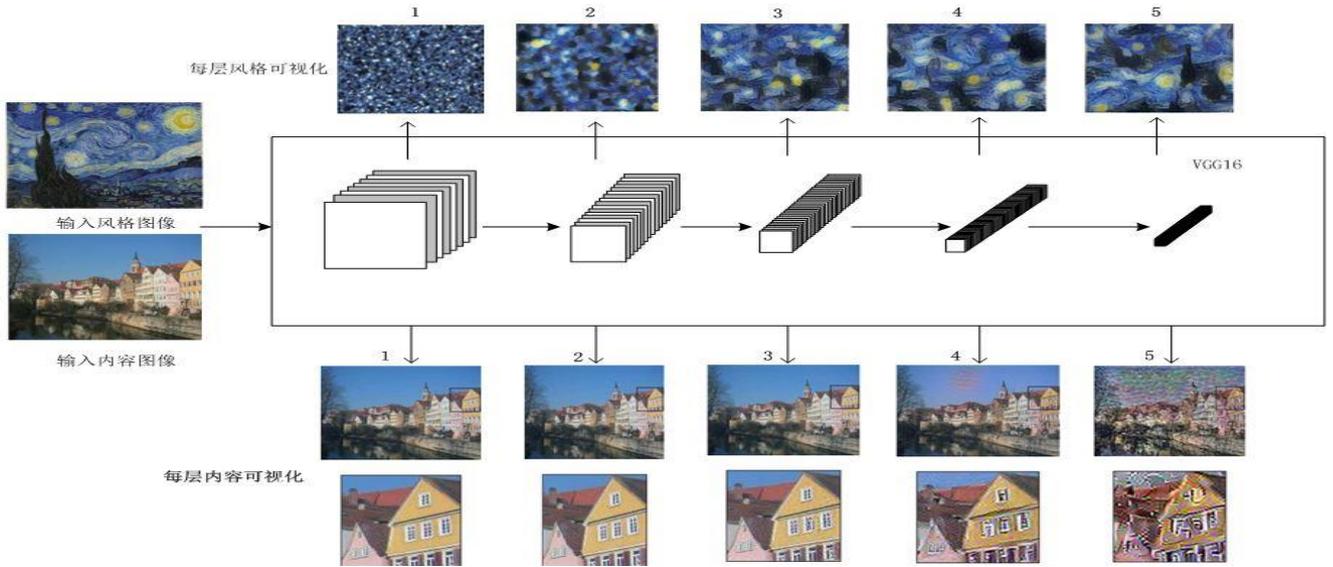


图2 卷积神经网络的图像可视化表示

Fig.2 Image Visualization of Convolutional Neural Networks

假设  $\bar{a}$  和  $\bar{x}$  是原图像和生成图像,  $A_l$  和  $G_l$  分别表示第  $l$  层样式, 则该层的风格损失和图像风格总损失分别为如下所示, 其中  $w_l$  是第  $l$  层在总损失中所占的权重

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

$$L_{style}(\bar{a}, \bar{x}) = \sum_{l=0}^L w_l E_l \quad (5)$$

损失相对于  $l$  层激活函数的导数如下, 使用标准误差反向传播计算  $E_l$  相对于像素值  $x$  的梯度。

$$\frac{\partial E_l}{\partial F_{ij}^l} = \begin{cases} \frac{1}{N_l^2 M_l^2} ((F^l)^T (G^l - A^l))_{ji} & \text{if } F_{ij}^l > 0 \\ 0 & \text{if } F_{ij}^l < 0 \end{cases} \quad (6)$$

为了生成了一个能够同时匹配  $\bar{p}$  的内容表示和  $\bar{a}$  的风格表示的图像, 通过最小化白色噪声图像来表示某层原始的内容特征与卷积神经网络多个层上

定义的沙画风格相结合。定义的最小化的损失函数为:

$$L_{total}(\bar{p}, \bar{a}, \bar{x}) = \alpha L_{content}(\bar{p}, \bar{x}) + \beta L_{style}(\bar{a}, \bar{x}) \quad (7)$$

其中  $\alpha$  和  $\beta$  分别为内容和样式损失的权重因子。

对于  $\partial L_{total} / \partial \bar{x}$  的梯度可以用作一些数值优化策略的输入。这里使用的是 L-BFGS<sup>[4]</sup>, 我们发现它最适合图像合成。如果我们使用不同的网络架构和优化算法, 预计图像合成中也会出现一些差异, 沙画风格化算法过程如图 3。

### 3 实验结果

本文算法中采用的卷积神经网络中内容和风格的表示可以很好地分离, 因此可以独立处理这两种表示, 以产生新的在感知上有意义的图像。以下是根据编码实现的油画效果图, 我们将不同的图片内容表现形式与多个油画艺术品的风格表现形式相结合, 不同的图像与各种油画作品相结合的效果图如图 4 所示。

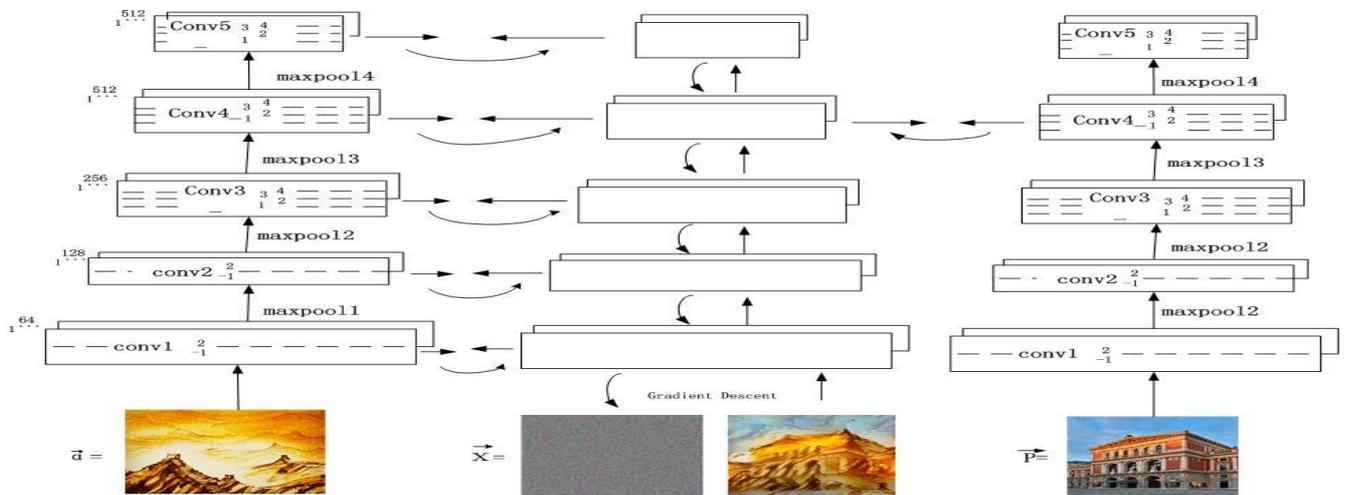


图3 沙画风格化算法

Fig.3 Sand painting stylization algorithm

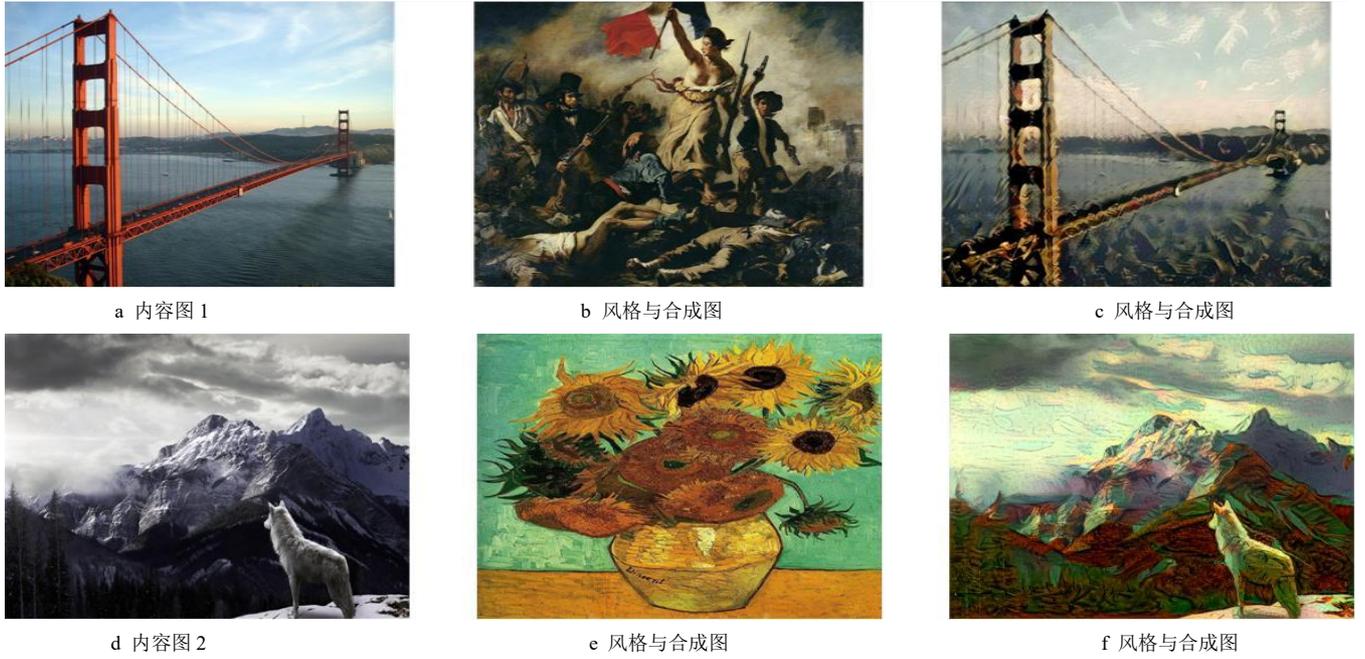


图4 不同内容的图像与沙画风格化后的合成图

Fig.4 Different contents and synthetic drawings after sand painting

图像风格化的一个影响因素是内容与风格的比率即  $\alpha/\beta$ 。图 5 显示的是图 4 中内容图 (d) 通过  $\alpha/\beta$  分别选用  $1 \times 10^{-2}$ 、 $1 \times 10^{-4}$  和  $1 \times 10^{-6}$  沙画风格化的合成图。从图 5 中三幅图像内容与风格的比率  $\alpha/\beta$  的比值依次减小, 即左边的合成图像强调内容, 可以清楚的识别内容, 但是图片的风格效果不是很好, 右边的图像风格效果很明显, 但是所显示的内容不易识别, 通常采用折中的方法调整内容和样式之间的比例来创造视觉上的更美观的图像。

图像风格化的另一个影响因素是卷积特征图层的选择。在上面说过样式表示是包含多层神经网络的多尺度表示, 这些图层的数量和位置决定了风格匹配的局部尺度, 从而导致不同的视觉体验。将样式表示与网络中的较高层相匹配可以保持局部图像结构的

规模越来越大, 从而导致更平滑和更持续的视觉体验<sup>[15]</sup>。因此, 风格图像通常将样式表示匹配到网络中的最高层。为了分析使用不同图层来匹配内容特征的效果 (使用图 4 的 d、e 两图作为源图片), 我们将其他的参数设置为相同的数量 ( $1 \times 10^4$ ) 来对图片进行风格转移, 如图 6 所示。当匹配网络较低层的内容时, 该算法会匹配照片中的大部分详细像素信息, 生成的沙画图像看起来感觉像是在图片上混合了艺术作品的纹理 (如 Conv1\_2); 中间是 Conv3\_2, 当匹配网络的较高层上的内容特征时, 照片的详细像素信息没有如此强烈的约束, 建筑物的内容和与沙画的纹理被融合在一起 (如 Conv5\_2), 图像的边缘和色彩等精细结构会发生很大变化, 更加突出了风格特征。



图5  $\alpha$  与  $\beta$  不同比值的效果图

Fig.5 The effect of different ratio of  $\alpha$  and  $\beta$

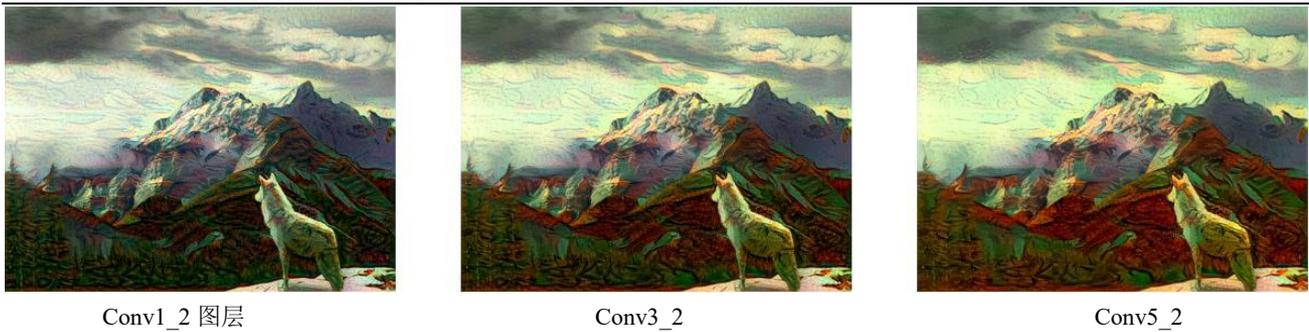


图6 匹配不同内容特征层效果图

Fig.6 Match different content feature layers renderings

## 4 总结

本文主要介绍了卷积神经网络在沙画风格化上的应用,展示了如何使用高性能卷积神经网络的特征表示在图像之间传输图像样式。虽然合成了具有高感知质量的视觉图像,但是该算法仍然存在一些技术限制,最明显的不足就是合成的图形分辨率较低,清晰度大大降低,在实验运行过程速度较慢,完成一轮图像的优化需要一个小时左右,最近研究的实时风格转换<sup>[16]</sup>提高了运行速率,虽然在沙画风格化方面取得了很好的效果,但是到目前为止并不清楚图像的风格究竟是由什么定义,因此图像内容与风格的分离不是一个明确定义的问题,它可能是绘画中的笔触、色彩、形状或者是场景的构图以及图像主题的选择等。在以后的风格化研究中,可以不断尝试使用人工神经网络对风格化处理,来取得更好的视觉效果。

## 参 考 文 献

[1]Winnemoller H, Olsen S C, Gooch B. Real-time video abstraction [C]. ACM Transactions On Graphics (TOG), ACM, 2006, 25(3): 1221-1226.

[2]Hertzmann A, Jacobs C E, Oliver N, Curless B, Salesin D H. Image analogies [J]. Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM, 2001: 327-340.

[3]Ashikhmin N. Fast texture transfer. IEEE Computer Graphics and Applications, 2003, 23(4):38-43.

[4]Lee H, Seo S, Ryoo S, et al. Directional texture transfer [C]. International Symposium on Non-photorealistic Animation & Rendering. 2010: 43-48.

[5]Chen Y, Zhao D, Lv L, et al. A visual attention based convolutional neural network for image classification [C]. Intelligent Control & Automation. IEEE, 2016: 764-769.

[6]Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4):640-651.

[7]Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. Computer Science, 2014.

[8]Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding[J]. 2014. Proceedings of the ACM International Conference on Multimedia. ACM, 2014: 675-678.

[9]Heeger D J, Bergen J R. [ACM Press the 22nd annual conference - Not Known (1995. -. )] Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, - SIGGRAPH '95 - Pyramid-based texture analysis/synthesis[J]. 1995:229-238.

[10]Efros A A. Image Quilting for Texture Synthesis and Transfer[C]. Conference on Computer Graphics & Interactive Techniques. ACM, 2001: 341-346.

[11]Simonyan K, Vedaldi A, Zisserman A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps [J]. Computer Science, 2013.

[12]Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding [J]. 2014: 675-678.

[13]Mahendran A, Vedaldi A. Understanding Deep Image Representations by Inverting Them [J]. 2014.

[14]Zhu C, Byrd R H, Lu P, et al. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization[J]. ACM Transactions on Mathematical Software, 1997, 23(4):550-560.

[15]Gatys L A, Ecker A S, Bethge M. Texture synthesis using convolutional neural networks [C]. International Conference on Neural Information Processing Systems. MIT Press, 2015.

[16]Johnson J, Alahi A, Fei-Fei L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution[J]. European Conference on Computer Vision. 2016: 694-711.