

# 基于改进 AdaIN 的图像风格迁移

吴岳, 宋建国

(山东科技大学 计算机科学与工程学院, 山东 青岛 266590)

**摘要:** 图像风格迁移技术是指给定内容图和风格图, 利用机器学习算法将内容图渲染成具有艺术风格的画作。针对传统图像风格迁移算法无法兼顾速度与生成图像质量等问题, 基于 AdaIN 算法, 提出 AdaIN 改进算法。在原始 AdaIN 网络中加入内容图像的深度信息计算模块, 提取内容图像的深度图, 通过将迁移图像数据与深度图归一化处理后的数据按元素相乘的方法, 突出内容图的深度信息, 使得输出的风格迁移图像各深度下具有不同的风格化程度。实验表明, 相较于 Gatys 等多种传统风格迁移算法, AdaIN 改进算法在运行时间上可降低约 11%; 不需要针对每种风格单独训练网络, 避免了模型重复训练; 内容图像深度信息得以保存, 提高了图像渲染质量。

**关键词:** 图像风格迁移; 机器学习; AdaIN 算法; 深度信息

**DOI:** 10.11907/rjdk.192783

**中图分类号:** TP317.4

**文献标识码:** A

**开放科学(资源服务)标识码(OSID):**

**文章编号:** 1672-7800(2020)009-0224-04



## Image Style Transfer Based on Improved AdaIN

WU Yue, SONG Jian-guo

(School of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China)

**Abstract:** The technology of image style transfer is to render the content image into a painting with artistic style by using machine learning algorithm given the content image and style image. Aiming at the problem that the traditional image style migration algorithm can't take into account the speed and the quality of the generated image, this paper proposes an improved AdaIN algorithm based on the AdaIN algorithm designed by Xun Huang et al. By adding the depth information calculation module of the content image into the original AdaIN network, the depth map of the content image is extracted. After the normalization of the migrated image data and the depth map, the data is processed by elements. The method of multiplication highlights the depth information of the content map, which makes the output style migration image with different stylization degree in different depth. Experimental results show that the improved algorithm described in this paper can reduce the running time by about 11% compared with many traditional style migration algorithms such as Gatys. It does not need to train the network separately for each style, avoiding the repeated training of the model. The depth information of the content image can be saved and the rendering quality of the image is therefore improved.

**Key Words:** image style transformation; machine learning; AdaIN; depth information

## 0 引言

图像风格迁移是指利用机器学习算法将一张图像渲染成有艺术风格的画作<sup>[1]</sup>。如图1所示, 图像风格化算法输入有二个, 分别是(a)内容图和(b)风格图, 输出有一个, 为(c)风格迁移图像, 输出图像既保留内容图的内容, 又融合了风格图的艺术风格。

图像风格化起源于非真实感渲染和纹理生成, 早期方

法<sup>[2-3]</sup>依赖于低级图像特征而构建, 常常无法捕获图像语义结构, 而利用深度学习构建的模型可以从数据中自适应地提取图像特征, 使其在处理这类难以手工构建特征的问题上具有先天优势。因此, 利用深度学习技术研究图像的风格迁移也逐渐成为主流。

Gatys 等<sup>[4]</sup>提出了一种新的风格转换算法, 该算法相较于早期方法可以灵活地提取任意风格图片的样式, 并对任意的内容图片进行风格化。该算法是在图像经过预训练的 VGG-19 网络时的特征表达 (Feature Map) 上计算 Gram

收稿日期: 2020-01-14

作者简介: 吴岳 (1994-), 男, 山东科技大学计算机科学与工程学院硕士研究生, 研究方向为计算机图形学、计算机视觉; 宋建国

(1995-), 男, 山东科技大学计算机科学与工程学院硕士研究生, 研究方向为空间聚类、大数据与云计算。本文通讯作者: 吴岳。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

矩阵, 利用 Gram 矩阵表示一种纹理。然而, 该种算法对算力要求较高, 即使是在现代 GPU 上生成中等分辨率的单张图像也需要几分钟时间。为了加快转换速度, Ulyanov 等<sup>[5]</sup>在后续研究中提出通过转换网络实现对内容图像的风格化, 经过训练的转换网络会绑定到一种特定风格上, 这种方法加快了转换速度, 但对于每种样式都要训练单独的转换网络。风格迁移任务也可以借用 GAN 网络完成, Zhu & Zhang 等<sup>[6-7]</sup>利用 GAN 网络训练模型, 增加了图像转换的多样性, 但该算法一般需要更多的风格图片数据集去完成风格重建。Liu 等<sup>[8]</sup>通过加入正则化对传统方法<sup>[2]</sup>进行了扩展, 并且利用 Chen<sup>[9]</sup>的方法保留了内容图像的深度图信息, 由于图像深度信息在人类对图像的感知和分析中必不可少, 因此该算法显著提高了图像渲染质量。在之后的研究中, 如 AdaIN<sup>[10]</sup>利用 Encoder、Decoder 结构, 允许在不训练单独网络的情况下传输任意样式, 但由于该方法未能保留内容图像深度信息, 因而渲染质量不佳。

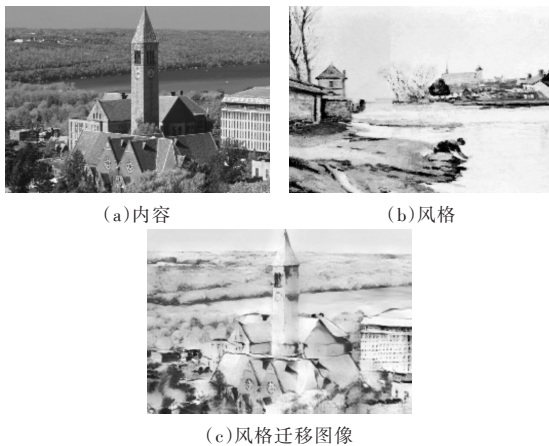


图 1 风格迁移

本文主要基于 AdaIN 方法进行扩展改进, 在保留其 Encoder、Decoder 结构的前提下融合内容图的深度计算模块, 使得最后输出的风格化图像兼顾效率和深度信息, 从而提高渲染质量。

## 1 相关算法

### 1.1 AdaIN

AdaIN 算法结构如图 2 所示。AdaIN<sup>[10]</sup>算法输入包括内容图像  $I_c$  和风格图像  $I_s$ , 输出为风格迁移后的图像  $I_{out}$ 。

其算法流程可用式(1)表示。

$$I_{out} = D(AdaIN(E(I_c), E(I_s))) \quad (1)$$

其中,  $E()$  和  $D()$  分别表示编码器 Encoder 和解码器 Decoder。为方便表示, 定义经过编码器处理后的数据如式(2)所示。

$$x = E(I_c), y = E(I_s) \quad (2)$$

$$x \in R^{c \times w \times h}, y \in R^{c \times w \times h}, AdaIN(x, y) \in R^{c \times w \times h}$$

其中,  $AdaIN(x, y)$  是对 IN (Instance-Normalization) 这一规范化方法<sup>[14]</sup>的改进。如式(3)所示, IN 对每个样本在每个通道进行规范化, 通过对特征图  $x$  作减去均值和除以标准差处理, 使其成为均值为 0、方差为 1 的数据分布后, 再进行仿射变换。

$$IN(x) = \gamma \left( \frac{x - \mu(x)}{\sigma(x)} \right) + \beta \quad (3)$$

如式(4)所示, AdaIN 与 IN 的不同之处在于其仿射变换参数来源于样本, 具体到风格迁移任务中, 其参数源自于风格图像  $I_s$ 。

$$AdaIN(x, y)_{cij} = \sigma_c(y) \left( \frac{x_{cij} - \mu_c(x)}{\sigma_c(x)} \right) + \mu_c(y), \quad c = 1, 2, 3 \dots C$$

$$\mu_c(x) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W x_{cij}, \quad \sigma_c(x) = \sqrt{\frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W (x_{cij} - \mu_c(x))^2}$$

Encoder 是 VGG 网络<sup>[18]</sup>的全部卷积层部分(直到 Re-lu4\_1 部分), VGG 网络在 ImageNet<sup>[13]</sup>数据集上进行预训练(其网络参数在接下来的解码器训练中固定)。Decoder 与 Encoder 对称, 逐步对图像数据进行上采样操作, 这就保证了最后 Decoder 输出既与内容图像维度相同, 又融合了风格图像的艺术风格。

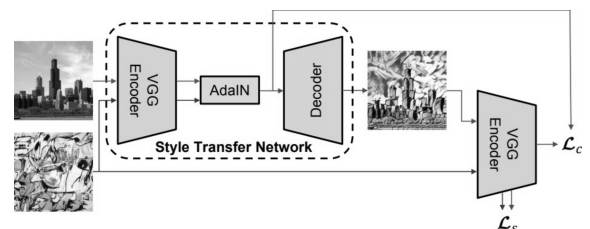
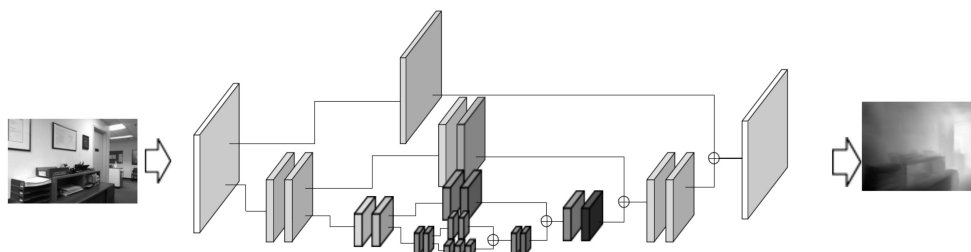


图 2 AdaIN 算法结构

### 1.2 Depth Estimation

Chen<sup>[9]</sup>等设计的端到端深度估计算法针对单张图片。该算法利用相对深度学习, 并且只使用图像本身进行预测。

如图 3 所示, 该算法主要思想是构造一个从输入图像到像素级深度图的函数映射 (Image-to-Depth)。其使用

图 3 Chen<sup>[9]</sup>提出的网络结构

Hourglass<sup>[15]</sup>网络的一个变体,包含一系列卷积下采样、反卷积和上采样,从更高的分辨率加入特征。这种对称型网络设计方式保证了输出图与输入图的分辨率相同。该模型是基于相对深度数据训练而成,因此其输出只能通过像素值的相对大小判断图像上各像素的远近。

## 2 AdaIN 算法改进

针对 AdaIN 算法生成的风格迁移图像因无法保存内容图深度信息而导致的渲染质量下降问题,选择对 AdaIN 算法加以改进,融合内容图的深度计算模块,基于人对近处物体比远处物体更敏感的视觉处理特点,对近处物体减弱其风格化程度,以更详尽的方式进行渲染,物体越远,风格化减弱程度越低。AdaIN 改进结构如图 4 所示。

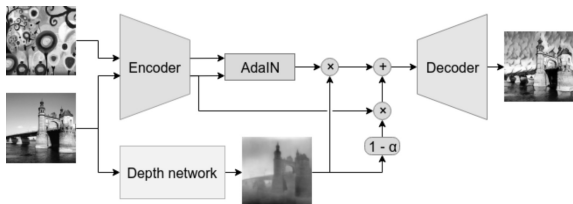


图 4 AdaIN 改进结构

具体做法是:基于深度图表示相对深度的特点,对于内容图像的深度图  $I_d$  作归一化处理(减去最小值,除以最大值与最小值的差值),再将数据映射到  $[0, 1]$  区间,因此 0 对应最接近的像素,1 对应最远像素。如式(5)所示。

$$\alpha = \frac{I_d - \min(I_d)}{\max(I_d) - \min(I_d)}, \quad \alpha \in R^{h \times w} \quad (5)$$

修改原始 AdaIN 算法流程,融合内容图深度信息。如图 5 所示,通过一个实例展示深度信息如何控制风格强度,其中深度渐变图从左到右数值逐渐加大(数值范围  $[0, 1]$ ),即深度逐渐增加,对应到风格迁移图像上,从左到右风格化程度逐渐加深。通过该实例可以发现,可将风格迁移图像与  $[0, 1]$  之间任意深度值相结合,构造出不同程度

的风格化效果。

利用式(6)可以控制图像各深度的风格化程度。其中,  $\otimes$  表示按元素相乘,由于  $\alpha$  维度为  $h \times w$ ,而图像维度为  $c \times w \times h$ ,因此在相乘之前利用 Pytorch 或 Numpy 中的广播机制对  $\alpha$  进行维度扩展,使其成为  $c \times w \times h$  的数据。

$$I_{out} = D(\alpha \otimes AdaIN(E(I_c), E(I_s))) + (1 - \alpha) \otimes E(I_c) \quad (6)$$

对于改进的算法模型, Encoder、Decoder、Depth-Net-work 都可以选择使用经过预训练的网络。该计算方法的优势在于不需要对新的风格进行额外训练,可以有效节约计算资源,并加快图片风格化速度。

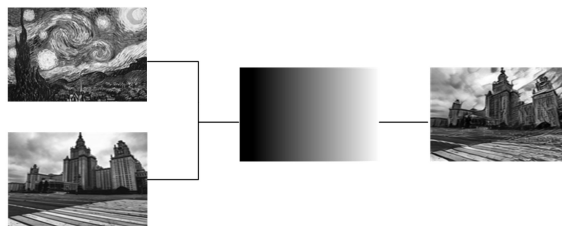


图 5 渐变深度信息控制风格化强度

## 3 实验与结果分析

### 3.1 实验环境搭建

实验在 Linux 环境下使用深度学习框架 Pytorch 搭建网络,该框架 Torchversion 库中含有经过预训练的 VGG-19 网络<sup>[17]</sup>,有助于加快整个网络训练速度。

为验证不同内容、不同艺术风格图片的迁移效果,笔者进行了大量实验,并与当前主流风格迁移算法原始 AdaIN、Gatys 等方法进行对比。风格迁移效果比较如图 6 所示。

### 3.2 结果分析

通过对比可以发现,本文方法相较于 AdaIN 方法和 Gatys 方法具有更好的风格迁移效果。AdaIN 等方法对整个图像以相同的风格化强度进行均匀渲染;而本文方法通过突出内容图像的深度信息,使得输出的风格迁移图像在

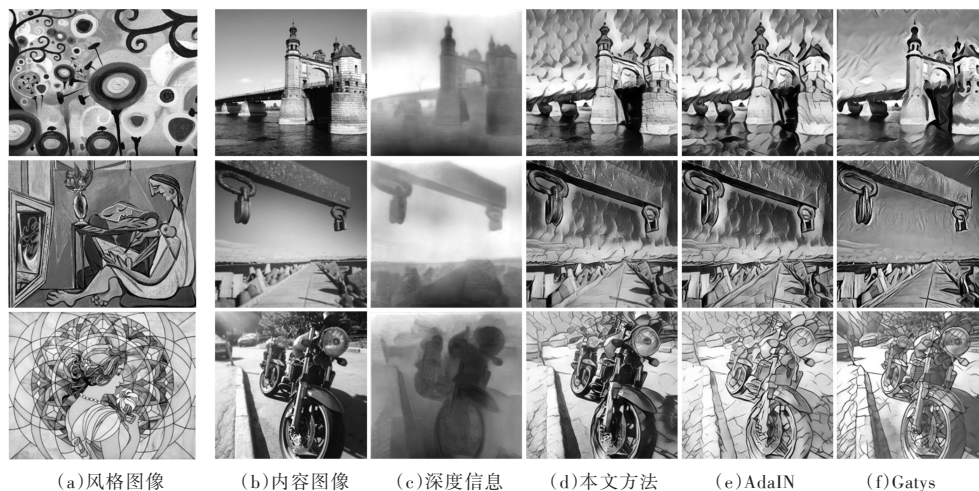


图 6 风格迁移效果对比



各深度下具有不同的风格化程度,图像中的近处物体比远处物体应用了较少的风格处理,而是以更详细的方式渲染它们,从而产生了令人满意的视觉效果。

为了进行更全面的对比,考虑以下几个方面:①实时性:模型是否可以通过神经网络一次前向传播生成风格迁移图像;②免学习性:对于新的风格类别,是否需要重新训练一种与之对应的新模型;③深度保存性:是否可以保存内容图像的深度信息,使得风格迁移图像渲染质量更高。

表1 算法优势对比

算法	实时性	免学习性	深度保存性
Gatys的方法	否	否	否
Jonson的方法	是	否	否
Dumoulin的方法	是	否	否
Ghiasi的方法	是	否	是
AdaIN方法	是	否	是
Liu的方法	是	是	否
本文方法	是	是	是

## 4 结语

本文针对AdaIN算法在风格迁移任务中不能保存内容图深度信息的弊端,提出AdaIN改进算法。这一改进算法不但保留了AdaIN算法的优点,即快速实时的风格迁移图像生成、无需为新风格训练新的模型,而且其生成的风格迁移图像允许从内容图中保留图像深度信息,从而提高了生成图像的渲染质量。下一步工作中,将尝试对该算法作进一步改进,加入局部特征处理功能,即只对图像中某一特定物体进行风格化处理,保留其余物体的原始风格不变。

### 参考文献:

- [1] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521: 436-444.
- [2] GOOCH B, GOOCH A. Non-photorealistic rendering[M]. Boca Raton: AK Peters/CRC Press, 2001.
- [3] THOMAS S, SCHLECHTWEIG S. Non-photorealistic computer graphics: modeling, rendering and animation[M]. San Francisco: Morgan Kaufmann, 2002.
- [4] GATYS L A, ECKER A S, BETHGE M. Texture and art with deep neural networks[J]. Current Opinion in Neurobiology, 2017, 46: 178-186.
- [5] ULYANOV D, VEDALDI A, LEMPITSKY V. Instance normalization: the missing ingredient for fast stylization[DB/OL]. <https://arxiv.org/abs/1607.08022>, 2016.
- [6] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]. Proceedings of the IEEE international conference on computer vision, 2017: 2223-2232.
- [7] ZHANG L M, JI Y, LIN X, et al. Style transfer for anime sketches with enhanced residualu-net and auxiliary classifier gan[C]. 2017 4th IAPR Asian Conference on Pattern Recognition, 2017: 506-511.
- [8] LIU X C, CHENG M M, LAI Y K, et al. Depth-aware neural style transfer[C]. Proceedings of the Symposium on Non-Photorealistic Animation and Rendering, 2017: 1-10.
- [9] CHEN W F, FU Z, YANG D, et al. Single-image depth perception in the wild[C]. International Conference on Advances in Neural Information Processing Systems, 2016: 730-738.
- [10] HUANG X, BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 1501-1510.
- [11] GHIASI G, LEE H, KUDLUR M, et al. Exploring the structure of a real-time, arbitrary neural artistic stylization network[DB/OL]. <https://arxiv.org/abs/1705.06830>, 2017.
- [12] LI Y J, FANG C, YANG J M, et al. Universal style transfer via feature transforms[C]. International Conference on Advances in Neural Information Processing Systems, 2017: 386-396.
- [13] DENG J, DONG W, SOCHER R, et al. Imagenet: a large-scale hierarchical image database[C]. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 248-255.
- [14] ULYANOV D, VEDALDI A, LEMPITSKY V. Improved texture networks: maximizing quality and diversity in feed-forward stylization and texture synthesis[DB/OL]. <https://arxiv.org/pdf/1701.02096.pdf>.
- [15] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation[DB/OL]. <https://arxiv.org/abs/1603.06937>.
- [16] JOHNSON J, ALAHI A, LI F F. Perceptual losses for real time style transfer and super resolution[C]. European Conference on Computer Vision, 2016: 694-711.
- [17] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[DB/OL]. <https://arxiv.org/abs/1409.1556>, 2019.
- [18] 张建华, 孔繁涛, 吴建寨, 等. 基于改进VGG卷积神经网络的棉花病害识别模型[J]. 中国农业大学学报, 2018, 23(11): 161-171.
- [19] 许哲豪, 陈玮. 基于生成对抗网络的图片风格迁移[J]. 软件导刊, 2018, 17(6): 207-209, 212.
- [20] 乔丽莎. 基于深度学习的图像风格艺术化[D]. 西安: 西安理工大学, 2018.

(责任编辑: 孙娟)