

文章编号: 2095-2163(2020)02-0057-05

中图分类号: TP391.41

文献标志码: A

# 基于深度学习的风格迁移算法的研究与实现

王 鹿, 曾国辉, 黄 勃

(上海工程技术大学 电子电气工程学院, 上海 201620)

**摘要:** 近年来,随着深度学习的强势崛起,计算机视觉技术的飞速发展,图像风格迁移技术也受到各界的广泛关注。图像风格迁移,是指利用算法完成对图像艺术风格的学习,最终将这种风格嫁接到其他图像上完成图像再创作的过程。本文就 VGG16, VGG19, Cycle GAN 三种风格迁移算法进行分析,利用图像深度特征与语义内容之间的联系,研究卷积神经网络中图像特征迁移。最终,通过对比 3 幅迁移图像的实际效果,证明 Cycle GAN 效果更好。

**关键词:** 风格迁移; 深度特征迁移; VGG; GAN

## Research and implementation of style transfer algorithm based on deep learning

WANG Lu, ZENG Guohui, HUANG Bo

(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

**[Abstract]** In recent years, with the strong rise of deep learning and the rapid development of computer vision technology, image style transfer technology has also received widespread attention from all walks of life. Image style transfer refers to the process of learning the artistic style of an image using an algorithm, and grafting this style onto other images to complete the image re-creation. This paper analyzes three style transfer algorithms, VGG16, VGG19, and Cycle GAN, and uses the relationship between image depth features and semantic content to study image feature transfer in convolutional neural networks. Based on the above, by comparing the actual effects of the three migration images, Cycle GAN proves to be better.

**[Key words]** style transfer; deep feature transfer; VGG; GAN

## 0 引言

近年来,随着人工智能、计算机视觉技术的迅猛发展,越来越多的机器学习应用场景在不断涌现。与此同时,目前有着较好表现的监督学习需要大量的标注数据,但标注数据是一项枯燥无味且花费巨大的任务,所以无监督的 image-to-image 式风格迁移学习即吸引了学界的更多关注,并已成为深度学习技术的重要应用之一。对于传统监督式机器学习而言,迁移学习都是基于同分布假设<sup>[1]</sup>,同时需要大量标注数据,然而实际使用不同数据集的过程中可能出现一些问题,比如数据分布差异,训练数据过期,也就是已然做过标定的数据要被丢弃,有些应用中数据分布也会随着时间的推移产生变化。因此,如何充分利用已经标注好的数据,同时又保证在新任务上的模型精度已经成为一个亟需解决的难题。基于此,本文拟对基于深度学习的迁移学习进行研究<sup>[2-3]</sup>。

目前,研究发现对一幅图像而言,风格和内容在卷积神经网络中的表达是可以分开的。也就是说,

研究中可以独立地操纵 2 种表达来产生新的、可以感受的有意义的图片。在本文中,将使用 VGG-16、VGG-19 和 Cycle GAN<sup>[4]</sup>,来实现风格迁移,对比这 3 种网络的结果,分析后发现,Cycle GAN 的效果最好,VGG-16 和 VGG-19 虽然也有一定的效果,但是合成图像的整体风格并不突出,不能满足本文研究的需求。

## 1 图像风格迁移算法

图像风格迁移技术主要有 2 种。一种是由 Gatys 等人<sup>[5]</sup>率先提出的 NAAS,该方法以 VGG 网络为基础进行损失网络的设计;另一种是 Johnson 等人<sup>[6]</sup>提出的快速图像风格迁移技术(FNST),该方法基于 Gatys 等人的研究成果,在损失网络的前端添加了一个 Image Transform Net,该网络基于残差网络进行设计。两者之间的关系如图 1 所示。

Gatys 所提出的 NAAS 是在 VGG 网络的基础上,利用梯度下降,经过若干次的迭代从而获得转换风格后的图像。该方法的整体流程如图 2 所示。

**基金项目:** 国家自然科学基金(61603242);江西省经济犯罪侦查与防控技术协同创新中心开放课题资助项目(JXJZTCX-030)。

**作者简介:** 王 鹿(1995-),男,硕士研究生,主要研究方向:图像识别、机器学习;曾国辉(1975-),男,博士,副教授,主要研究方向:机器人控制、电力电子及其控制技术;黄 勃(1985-),男,博士,讲师,主要研究方向:人工智能。

**通讯作者:** 曾国辉 Email:cherry6180@163.com

收稿日期: 2019-12-25



图1 快速图像风格迁移方法

Fig. 1 Fast image style migration method

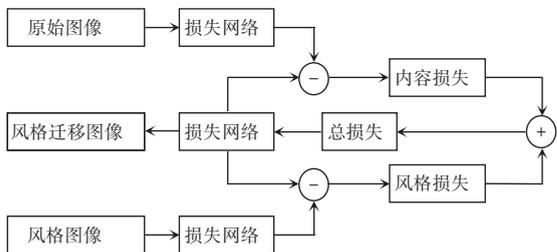


图2 NASS 工作流程图

Fig. 2 NASS workflow

### 1.1 VGG 网络

VGG (Visual Geometry Group) 隶属于牛津大学, VGG 自 2014 年先后发布了网络模型 VGG11 ~ VGG19, 该小组研究证明了增加神经网络的深度能够在一定程度上影响神经网络的性能。VGG 网络结构简单,其卷积层和池化层可以划分为不同的块 (Block),从前到后依次编号为 Block1 ~ Block5。每一个块内包含若干卷积层和一个池化层。使用同样大小的卷积核尺寸 (3×3) 和最大池化尺寸 (2×2)。此外,VGG 网络还有 5 个最大池化层,分别分布在不同的卷积层之下<sup>[7-8]</sup>。

(1) VGG16。VGG16 共有 13 个卷积层 (Convolutional Layer), 分别用 conv3-XXX 表示,3 个全连接层 (Fully connected Layer) 分别用 FC-XXXX 表示,5 个池化层 (Pool layer) 分别用 MaxPool 表示。卷积层和全连接层具有权重系数,因此也被称为权重层,总数目为 13+3=16,这即是 VGG16 中 16 的来源。考虑到池化层不涉及权重,因此不属于权重层,不被计数。VGG16 神经网络模型结构如图 3 所示。

(2) VGG19。VGG19 共有 16 个卷积层和 3 个全连接层。如图 4 所示。

以图 4 为例,在 VGG19 的 Block2 中包含 2 个卷积层,卷积核为 3×3×3,通道数都是 128,而 Block4 包含 4 个卷积层,1 个最大池化层,通道数是 512。

### 1.2 Cycle GAN

传统的 GAN 的 G 是将随机噪声转换为图片,但风格迁移中需要将图片转为图片,所以这个时候就要将图片作为 G 的输入,而 G 则是学习一种映射了。但是用单独一个 GAN 的训练并不稳定,可能导致所有照片全部映射到同一张图片的 mode collapse<sup>[9]</sup>。为此,就提出了 Cycle GAN 用来解决这个问题,

Cycle GAN 是将 2 个 GAN 组合起来,其目的是实现非成对 image 的转换,尤其适用于图像风格迁移中。为了使得 GAN 更加稳定,引入了此双向映射的机制,即 A→B 的 GAN 和 B→A 的 GAN,同时加入了一个 cycle\_loss, cycle\_loss 是采用 L1 损失设计而成<sup>[10-11]</sup>。所以研究得到 Cycle GAN 的损失函数,可将其写为:

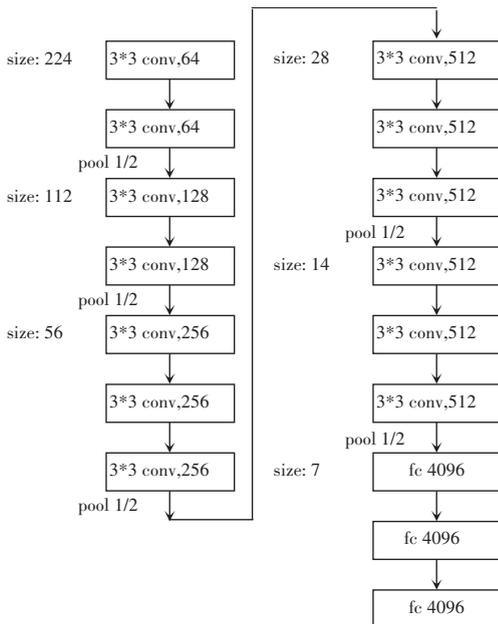


图3 VGG16 网络结构图

Fig. 3 VGG16 network structure diagram

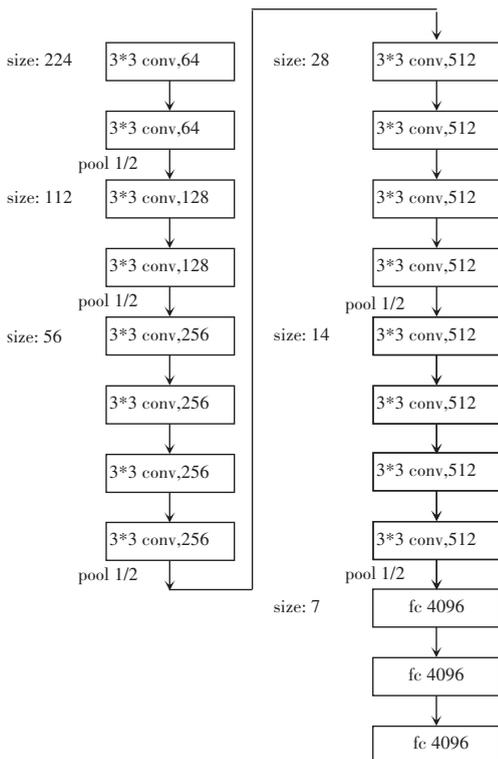


图4 VGG19 网络结构图

Fig. 4 VGG19 network structure diagram

$$L_{GAN}(F, D_X, Y, X) = E_{x \sim P_{data}(x)} [\log D_X(x)], \quad (1)$$

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim P_{data}(y)} [\log D_Y(y)], \quad (2)$$

接下来,研究得到了循环一致性损失可表示为:

$$L_{cyc}(G, F) = E_{x \sim P_{data}(x)} [\|F(G(x)) - x\|_1], \quad (3)$$

进而,推得总损失的数学公式如下:

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F). \quad (4)$$

### 1.3 损失函数

(1) 内容损失函数。假设某一层得到的响应是  $F^l \in R^{N_l * M_l}$ , 其中  $N_l$  为  $l$  层 filter 的个数,  $M_l$  为  $l$  层 filter 的大小,  $F_{i,j}^l$  表示第  $l$  层第  $i$  个 filter 在位置  $j$  的输出,  $\vec{p}$  表示提供 Content 的图像,  $\vec{x}$  表示生成的图像,  $P^l$  和  $F^l$  分别表示两者对于  $l$  层的响应, 因此  $l$  层的 Content Loss 的计算公式具体如下:

$$L_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F_{i,j}^l - P_{i,j}^l)^2, \quad (5)$$

(2) 风格损失函数。  $G^l \in R^{N_l * M_l}$ ,  $\vec{a}$  表示提供 style 的图像,  $\vec{x}$  表示生成的图像,  $A^l$  和  $G^l$  分别表示两者对于  $l$  层的响应, 因此  $l$  层的 Style Loss 的计算公式具体如下:

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{i,j}^l - A_{i,j}^l)^2, \quad (6)$$

$$L_{style}(\vec{a}, \vec{x}) = \sum_{l=0}^L w_l E_l, \quad (7)$$

综上所述,总的损失函数可由下式来计算:

$$L_{total}(\vec{p}, \vec{a}, \vec{x}) = \alpha L_{content}(\vec{p}, \vec{x}) + \beta L_{style}(\vec{a}, \vec{x}), \quad (8)$$

其中,  $\alpha$  和  $\beta$  分别是内容和风格在图像重构中的权重因子,且满足:

$$\alpha + \beta = 1. \quad (9)$$

## 2 实验

### 2.1 迁移效果的衡量指标

(1) 峰值信噪比 (Peak Signal to Noise Ratio, PSNR)<sup>[12]</sup>: 这是一种全参考的图像质量评价指标。运算中需用到如下数学公式:

$$PSNR = 10 \log_{10} \left( \frac{(2^n - 1)^2}{MSE} \right), \quad (10)$$

$$MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (X(i,j) - Y(i,j))^2, \quad (11)$$

其中,  $MSE$  (Mean Square Error) 表示当前图像  $X$  和参考图像  $Y$  的均方误差;  $H$ 、 $W$  分别表示图像的

高度和宽度;  $n$  为每像素的比特数,一般取 8,即像素灰阶数为 256。PSNR, 单位是 dB, 取值范围为  $(0, \infty)$ , 数值越大表示失真越小。

(2) 结构相似性 (structural similarity, SSIM)<sup>[13]</sup>: 这也是一种全参考的图像质量评价指标,可分别从亮度、对比度、结构三方面度量图像相似性。运算中涉及的数学公式可表示为:

$$l(X, Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1}, \quad (12)$$

$$c(X, Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2}, \quad (13)$$

$$s(X, Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3}, \quad (14)$$

其中,  $\mu_X, \mu_Y$  分别表示图像  $X$  和  $Y$  的均值;  $\sigma_X, \sigma_Y$  分别表示图像  $X$  和  $Y$  的方差;  $\sigma_{XY}$  表示图像  $X$  和  $Y$  的协方差,即:

$$\mu_X = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X(i, j), \quad (15)$$

$$\sigma_X^2 = \frac{1}{H \times W - 1} \sum_{i=1}^H \sum_{j=1}^W (X(i, j) - \mu_X)^2, \quad (16)$$

$$\sigma_{XY} = \frac{1}{H \times W - 1} \sum_{i=1}^H \sum_{j=1}^W ((X(i, j) - \mu_X)(Y(i, j) - \mu_Y)). \quad (17)$$

其中,  $C_1, C_2, C_3$  为常数,为了避免分母为 0 的情况, SSIM 取值范围  $[0, 1]$ , 值越大,表示图像失真越小。

### 2.2 实验环境

本实验选择 Python 语言和 Tensorflow, Pytorch 框架进行图像风格迁移的仿真测试。实验环境为 Intel(R) Core(TM) i5-8300H 2.1GHz CPU, GPU 为 NVIDIA。

### 2.3 训练部分

(1) VGG16 和 VGG19 训练。由于 VGG 网络深,网络架构权重数量非常大,导致训练非常缓慢,故研究用到的模型就是在 ImageNet 数据集上预训练的模型。

研究将从预训练的模型中,获取卷积层部分的参数,用于构建仿真中使用的模型,这些参数均是作为常量使用,即不再被训练,在反向传播的过程中也不会改变,此外 VGG 中的全连接层舍弃掉。参数设置见表 1。

(2) Cycle GAN。使用 vangogh2photo 数据集, Generator 采用的是文献 [6] 中的网络结构; 一个 residual block 组成的网络,降采样部分采用 stride 卷

积,增采样部分采用反卷积;Discriminator 采用的仍是 pix2pix 中的 Patch GANs 结构,大小为  $70 \times 70$ ,  $L_r = 0.0002$ 。对于前 100 个周期,保持相同的学习速率 0.0002,然后在接下来的 100 个周期内线性衰减到 0。

表 1 VGG16 与 19 相关训练参数设定

Tab. 1 VGG16 and 19 related training parameter settings

参数项	VGG16	VGG19
计算风格损失的层及其权重	“conv1_1”, 0.2	“conv1_1”, 0.2
	“conv2_1”, 0.2	“conv2_1”, 0.2
	“conv3_1”, 0.2	“conv3_1”, 0.2
	“conv4_1”, 0.2	“conv4_1”, 0.2
计算内容损失的层及其权重	“conv4_2”, 0.5	“conv4_2”, 0.5
		“conv5_2”, 0.5
训练次数	1 000	1 000
风格损失权重	1	1
内容损失权重	500	500
L2 正则化系数	0.005	0.005
dropout 的概率	0.5	0.5
噪声比率	0.5	0.5
学习率	10	10

## 2.4 实验分析

实例中的风格图片选为梵高先生的名画《星夜》,如图 5 所示。



图 5 模板艺术图《星夜》

Fig. 5 Template art illustration "Starry Night"

分别使用 Cycle GAN、VGG16、VGG19 进行实验,通过生成的图片观察不同目标函数的效果,实例中的风格图片均为梵高先生的名画《星夜》,实验运行效果见图 6,图 6 的对应数据指标见表 2。

通过对图 6 以及表 2 的对比发现,VGG16、VGG19 得到的结果相近,相较之下艺术风格并不浓烈,而 Cycle GAN 给人的直观感受就好得多,色度适

宜,艺术化表现也恰到好处。

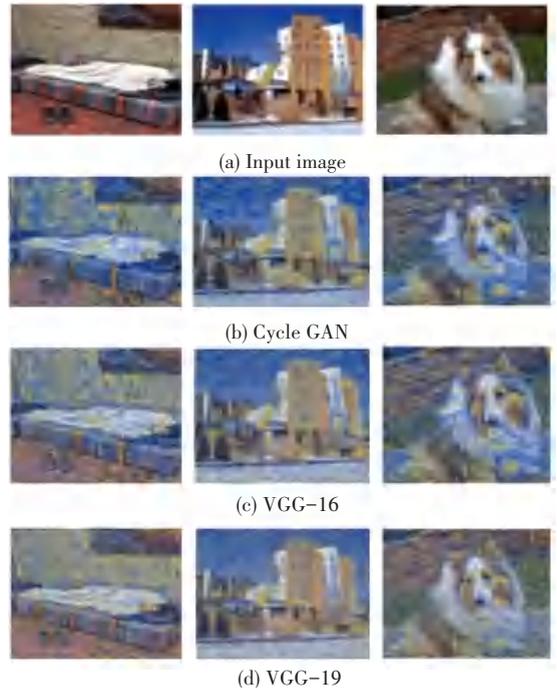


图 6 实验效果

Fig. 6 Experimental effect

表 2 实验指标

Tab. 2 Experimental index

	SSIM/PSNR		
	VGG-19	VGG-16	Cycle GAN
Image_1	0.481 7/17.473 0	0.449 4/16.868 8	0.424 6/16.190 5
Image_2	0.431 0/17.329 7	0.396 1/17.069 8	0.373 0/16.722 1
Image_3	0.432 2/16.645 3	0.459 2/17.248 0	0.428 2/16.390 3

## 2.5 实验结果分析

VGG-16 与 VGG-19 本质上并无太大区别,两者均继承了 Alex Net 的深度思想,以其较深的网络结构、较小的卷积核和池化采样域,使其能够在获得更多图像特征的同时控制参数的个数,避免过多的计算量以及过于复杂的结构,如此一来 VGG 就能更好地找出内在特性,从而提取图像的 style features。

对比之下,由 2 个镜像对称的 GAN 组成的 Cycle GAN 不仅不需要成对的数据样本,在损失函数方面也进一步考虑了防止生成器 G 与 F 生成的样本相互矛盾,保证生成的样本与真实的样本同分布,这样也就保证了生成图像的连续性。

总体来说,GAN 网络在这种 image-to-image 领域内还是存在很大的优势。

## 3 结束语

本文选用卷积神经网络模型 VGG16、VGG19 和 (下转第 65 页)