

## 基于生成对抗网络的图像艺术风格迁移

董 伟, 赵杰煜\*

(宁波大学 信息科学与工程学院, 浙江 宁波 315211)

**摘要:** 提出了一种新颖的图像艺术风格化算法, 利用结构相似性指数和最小二乘生成对抗网络, 搭建图像艺术风格化模型. 通过对模型生成器和判别器的对抗训练以及重建约束, 该模型可以生成一幅逼真的风格化作品. 根据在人脸肖像素描 sketch-photo 数据集和中国水墨画风格 beihong-photo 数据集实验表明, 与目前流行的 DualGAN 算法、CycleGAN 算法、Pix2Pix 算法和 GAN 算法相比, 本文提出的方法具有更好的风格化效果.

**关键词:** 结构相似性指数; 最小二乘生成对抗网络; 图像艺术风格化; 非真实感绘制

中图分类号: TP301.6

文献标志码: A

文章编号: 1001-5132 (2019) 05-0030-06

随着大数据时代的来临与深度学习技术的兴起, 人工智能领域取得了快速发展, 使得研究者对人工智能的关注度以及社会大众对人工智能的憧憬得到空前提升<sup>[1-2]</sup>. 图像艺术风格化是近年来非真实感绘制领域的研究热点之一, 其主要考虑如何利用计算机模拟某种艺术风格绘制素描、水彩以及油画等<sup>[3]</sup>. 因此, 在海量图像数据下, 运用深度学习技术来实现图像艺术风格化具有重要意义.

已有学者提出了一些图像艺术风格化算法和基于生成对抗网络的图像风格转换算法. 如 Gatys 等<sup>[4]</sup>首次运用深度学习来实现图像的艺术风格转换, 通过卷积神经网络中间层输出特征图的统计特性来对图像内容以及风格特征进行提取和量化, 虽然效果惊艳, 但每幅图像的渲染都要经过漫长的学习过程, 在多轮迭代计算中逐像素地施加纹理. 文献[5-6]在 Gatys 的基础上提出了一种实时图像艺术风格化算法, 通过预训练 VGGNet 提取图像的高维抽象特征用以构建损失函数, 同时使用图像转换网络来存储风格的纹理特征. 该算法实验效果良好, 但每次仅仅学习到单张图片的艺术风格. Isola 等<sup>[7]</sup>提出了一种监督性风格转换框架 Pix2Pix. 该算法要求训练样本配对, 即已知输入图像的风格化结果, 因此为了使生成器产生更逼

真的结果, 该算法引入  $L_1$  范数来惩罚生成结果与真实结果之间的差异. 为了解决 Pix2Pix 训练数据需要配对的问题, Zhu 等<sup>[8]</sup>提出了无监督风格转换框架 CycleGAN, 引入循环一致性理论, 使用两组生成器和判别器, 将源域中一幅图像依次经过源域到目标域的映射, 以及目标域到源域的映射后得到源图像的重建图像, 并使用  $L_1$  范数约束重建图像与源图像越相近越好. Yi 等<sup>[9]</sup>提出了基于对偶学习和  $L_1$  范数的无监督图像风格转换 DualGAN. 近期, 谢志峰等<sup>[10]</sup>提出了一种基于 CycleGAN 生成高低曝光图片, 并结合输入图像, 实现了 HDR 风格迁移. 林嘉骏等<sup>[11]</sup>提出了一种基于像素级 GAN 的灰度图像彩色化模型, 目标函数使用  $L_1$  范数作为彩色化优化项. 上述基于 GAN 的图像风格转换算法都使用  $L_1$  范数作为生成器损失函数的一部分, 但  $L_1$  范数只是从像素角度去衡量图像之间的差异性, 并没有充分利用图像的结构信息.

本文结合结构相似性指数 (Structural Similarity Index Measurement, SSIM)<sup>[12]</sup>和最小二乘生成对抗网络 (Least Squares Generate Adversarial Nets, LSGAN)<sup>[13]</sup>, 提出一种在图像内容特征和风格特征保持平衡的图像艺术风格化算法, 使得在图像细节生成上有更出色的表现.

收稿日期: 2018-08-03.

宁波大学学报(理工版) 网址: <http://journalg.nbu.edu.cn/>

基金项目: 国家自然科学基金 (61571247); 浙江省自然科学基金 (LZ16F030001); 浙江省国际合作项目 (2013C24027).

第一作者: 董伟 (1993-), 男, 安徽宣城人, 在读硕士研究生, 主要研究方向: 机器学习与深度学习. E-mail: 1511082629@nbu.edu.cn

\*通信作者: 赵杰煜 (1965-), 男, 浙江宁波人, 教授, 主要研究方向: 计算机图像处理、机器学习与神经网络. E-mail: zhao\_jieyu@nbu.edu.cn

## 1 生成对抗网络

2014年Goodfellow等<sup>[14]</sup>首次提出了生成对抗网络(Generative Adversarial Nets, GAN), 成为深度学习结合生成模型和判别模型的成功典范. GAN将一个对抗性判别器模型(判别器)巧妙地结合到一个生成模型(生成器)中, 生成器生成数据, 判别器判断该数据的真伪, 二者相互竞争, 相互促进, 生成器产生的结果愈来愈真, 达到以假乱真的程度; 判别器的能力也愈来愈强, 对于真假数据的判别越来越强. GAN训练过程是一个全自动的非指导性学习过程, 几乎无需人工干预. 已有的研究表明<sup>[15-18]</sup>, GAN在诸如图像生成、图像超分辨率和半监督学习等各种任务中发挥着重要作用.

GAN的基本思想是同时训练一个判别器 $D$ 和一个生成器 $G$ , 判别器 $D$ 旨在区分真实样本和生成样本; 而生成器 $G$ 试图尽可能“欺骗”判别器, 使得判别器误认为生成样本为真实样本. 生成器 $G$ 与判别器 $D$ 均可以采用深度神经网络<sup>[19]</sup>. GAN的优化过程可以描述为一个“二元极小极大”问题, 目标函数为<sup>[12]</sup>:

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} (\log D(x)) + E_{z \sim p_z(z)} (\log(1 - D(G(z)))) \quad (1)$$

式中:  $p_z$  是随机噪声的分布;  $p_{\text{data}}$  是真实样本的分布;  $E(\cdot)$  表示计算期望值.

对于生成器 $G$ , 目标是 $\min_G V(D, G)$ , 由于 $D(x)$ 与目标无关, 因此仅需要最大化 $D(G(z))$ , 即生成器生成的结果需要成功“误导”判别器的输出结果. 对于判别器 $D$ , 目标是 $\max_D V(D, G)$ , 因此需要最大化 $D(x)$ 和最小化 $D(G(z))$ , 即判别器需要区分样本是真实样本还是生成样本.

## 2 基于GAN的图像艺术风格迁移

### 2.1 网络搭建

构建一个基于SSIM和LSGAN的残差网络进行图像艺术风格化学习. 网络的搭建借鉴了深度残差网络(ResNet)<sup>[20]</sup>和LSGAN. 其中生成器网络结构包含3个部分: 卷积层、残差块以及转置卷积层, 分别具有不同的卷积步长和模板数, 在对抗训练过程中完成对输入图像内容和目标图像风格特征的提取以及二者之间的融合过程. 判别器网络

结构由卷积层构成, 在对抗训练过程中完成对生成样本和真实样本的区分.

本研究的图像艺术风格化模型如图1所示, 由2组生成器和判别器构成. 在模型训练学习过程中, 人脸图像经过生成器 $G_{X2Y}$ 生成具有素描绘制风格的人脸肖像, 判别器 $D_Y$ 学习正确区分当前作品是否由生成器生成. 同时, 人脸素描图片经过生成器 $G_{Y2X}$ 转换为对应现实状态下的人脸图像, 判别器 $D_X$ 学习正确区分当前图像是否由生成器 $G_{Y2X}$ 转换过来. 本模型生成器的网络结构搭建借鉴了He等<sup>[20]</sup>提出的深度残差网络, 包括3个部分: 前部由3个卷积层构成; 中部由9个残差块构成; 尾部由2个转置卷积层和1个卷积层构成. 除第一层和最后一层卷积层使用了 $7 \times 7$ 的卷积核, 滑动步长取1, 其余卷积层均采用 $3 \times 3$ 的卷积核, 滑动步长取2. 生成器 $G_{Y2X}$ 与生成器 $G_{X2Y}$ 均采用相同的网络结构.

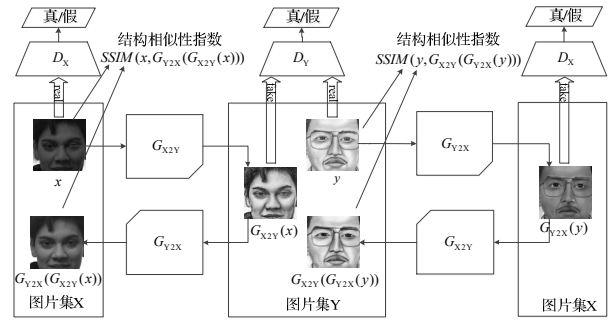


图1 基于GAN图像艺术风格化模型示意图

图像艺术风格化模型的判别器网络结构是由5个卷积操作组成的深度卷积神经网络. 输入为生成器产生的生成样本和真实样本, 输出为判断结果. 判别器 $D_X$ 与判别器 $D_Y$ 均采用相同的网络结构. 生成器和判别器的网络结构信息见表1. 由表1可知, 1~25层为生成器, 26~30层为判别器. 生成器除转置卷积操作外, 其余卷积操作之前均采用边缘镜像填充进行补零. 除生成器输出卷积层外, 其余所有卷积操作后都添加一个实例归一化层和ReLU激活函数层, 生成器输出层采用Tanh作为激活函数. 判别器除输出卷积层外, 其余卷积操作后均添加一个实例归一化层和LeakyReLU层.

在判别器中使用LeakyReLU作为激活函数, LeakyReLU是ReLU激活函数的改进版本.

$$f(x) = \begin{cases} 0.02x, & x < 0, \\ x, & x \geq 0. \end{cases} \quad (2)$$

表1 基于生成对抗网络的图像艺术风格化模型

| 序号 | 层类 | 卷积核 | 步长 | 模板数 | 序号 | 层类   | 卷积核 | 步长 | 模板数 |
|----|----|-----|----|-----|----|------|-----|----|-----|
| 0  | 输入 | -   | -  | 1   | 16 | 卷积   | 3×3 | 1  | 256 |
| 1  | 卷积 | 7×7 | 1  | 64  | 17 | 卷积   | 3×3 | 1  | 256 |
| 2  | 卷积 | 3×3 | 2  | 128 | 18 | 卷积   | 3×3 | 1  | 256 |
| 3  | 卷积 | 3×3 | 2  | 256 | 19 | 卷积   | 3×3 | 1  | 256 |
| 4  | 卷积 | 3×3 | 1  | 256 | 20 | 卷积   | 3×3 | 1  | 256 |
| 5  | 卷积 | 3×3 | 1  | 256 | 21 | 卷积   | 3×3 | 1  | 256 |
| 6  | 卷积 | 3×3 | 1  | 256 | 22 | 卷积   | 3×3 | 1  | 256 |
| 7  | 卷积 | 3×3 | 1  | 256 | 23 | 转置卷积 | 3×3 | 2  | 128 |
| 8  | 卷积 | 3×3 | 1  | 256 | 24 | 转置卷积 | 3×3 | 2  | 64  |
| 9  | 卷积 | 3×3 | 1  | 256 | 25 | 卷积   | 7×7 | 1  | 1   |
| 10 | 卷积 | 3×3 | 1  | 256 | 26 | 卷积   | 4×4 | 2  | 32  |
| 11 | 卷积 | 3×3 | 1  | 256 | 27 | 卷积   | 4×4 | 2  | 64  |
| 12 | 卷积 | 3×3 | 1  | 256 | 28 | 卷积   | 4×4 | 2  | 128 |
| 13 | 卷积 | 3×3 | 1  | 256 | 29 | 卷积   | 4×4 | 1  | 256 |
| 15 | 卷积 | 3×3 | 1  | 256 | 30 | 卷积   | 1×1 | 2  | 1   |

## 2.2 损失函数

在确定模型的目标函数时, 借鉴文献[13]提出的最小二乘生成对抗网络的相关理论, 在生成器  $G_{X2Y}$  和判别器  $D_Y$  训练过程中生成器  $G_{X2Y}$  试图尽可能“欺骗”判别器  $D_Y$ , 使判别器  $D_Y$  误认生成样本为真实样本, 即学习最小化  $(D_Y(G_{X2Y}(x)) - 1)^2$ . 同时, 判别器  $D_Y$  通过不断地学习, 提升识别生成样本与真实样本的能力, 使其对生成样本的评分更低, 真实图评分更高, 即学习最小化  $(D_Y(y) - 1)^2$  和  $(D_Y(G_{X2Y}(x)))^2$ .

生成器  $G_{X2Y}$  的损失函数为:

$$\min_{G_{X2Y}} V_{GAN}(G_{X2Y}) = \frac{1}{2} E_{x \sim p_{data}(x)} [(D_Y(G_{X2Y}(x)) - 1)^2]. \quad (3)$$

判别器  $D_Y$  的损失函数为:

$$\min_{D_Y} V_{GAN}(D_Y) = \frac{1}{2} E_{y \sim p_{data}(y)} [(D_Y(y) - 1)^2] + \frac{1}{2} E_{y \sim p_{data}(y)} [(D_Y(G_{X2Y}(x)))^2]. \quad (4)$$

同理, 生成器  $G_{Y2X}$  的损失函数为:

$$\min_{G_{Y2X}} V_{GAN}(G_{Y2X}) = \frac{1}{2} E_{y \sim p_{data}(y)} [(D_X(G_{Y2X}(y)) - 1)^2]. \quad (5)$$

同理, 判别器  $D_X$  的损失函数为:

$$\min_{D_X} V_{GAN}(D_X) = \frac{1}{2} E_{x \sim p_{data}(x)} [(D_X(x) - 1)^2] + \frac{1}{2} E_{y \sim p_{data}(y)} [(D_X(G_{Y2X}(y)))^2]. \quad (6)$$

为了进一步优化 GAN 的精度, 本文在 LSGAN 的损失函数的基础上, 在生成器的损失函数中添加重建损失函数作为损失函数的一部分. 为衡量生成器输入图像与其重建图像之间的差异程度, 引入 SSIM 来量化二者的相似性. 对输入图像  $x$  与其重建图像  $G_{Y2X}(G_{X2Y}(x))$  的结构相似度定义为:

$$SSIM(x, \hat{x}) = \frac{(2\mu_x \mu_{\hat{x}} + c_1)(2\sigma_x \sigma_{\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)}, \quad (7)$$

$$\hat{x} = G_{Y2X}(G_{X2Y}(x)), \quad (8)$$

则生成器  $G_{X2Y}$  重建损失函数项为:

$$L_{rec} = E_{x \sim p_{data}(x)} [1 - SSIM(\hat{x}, x)] + E_{y \sim p_{data}(y)} [1 - SSIM(\hat{y}, y)]. \quad (9)$$

最终, 生成器的损失函数为:

$$\min_{G_{X2Y}} V_{GAN}(G_{X2Y}) = \lambda L_{rec} + \frac{1}{2} E_{x \sim p_{data}(x)} [(D_X(G_{Y2X}(y)) - 1)^2], \quad (10)$$

式中  $\lambda$  用于控制对抗损失项与重建损失项间的相对重要性.

## 2.3 算法流程

**Step 1** 输入图片集  $X$ 、图片集  $Y$ 、生成器  $G_{X2Y}$  及其对应参数  $\theta_{X2Y}$ 、判别器  $D_Y$  及其对应参数  $\omega_Y$ 、生成器  $G_{Y2X}$  及其对应参数  $\theta_{Y2X}$ 、判别器  $D_X$  及其对应参数  $\omega_X$ 、权重系数  $\lambda$ 、最大循环次数  $N_{epoch}$ 、当前迭代次数  $t$ 、开始减小学习率的回合数  $N_{offset}$ 、生成器初始学习率  $\eta^g$  和判别器初始学习率  $\eta^d$ .

**Step 2** 随机初始化  $\theta_{X2Y}$ 、 $\omega_Y$ 、 $\theta_{Y2X}$  和  $\omega_X$ .

**Step 3** 更新迭代次数  $t=t+1$ .

**Step 4** 从集合  $X$  和  $Y$  中采样一张图像, 将其输入到对应的生成器得到相应的生成样本, 对真实样本以及生成样本分别打上 1 和 0 标签.

**Step 5** 将生成样本输入到对应的生成器, 得到图像的重建图像, 并计算各生成器重建损失.

**Step 6** 生成样本与真实样本分别输入判别器, 计算判别器目标函数, 最小化该目标函数, 同时根

据误差, 采用误差反传和改进后 Adam 优化更新判别器的权值; 同理, 对判别器进行权值更新。

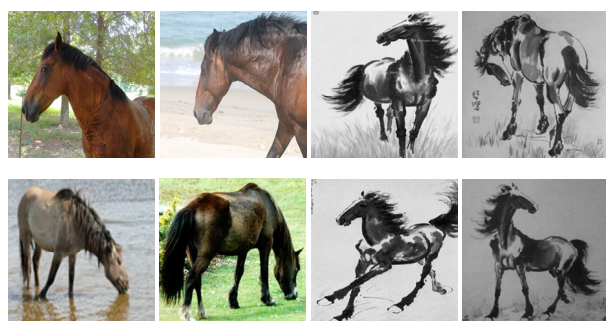
**Step 7** 对生成器进行优化, 同时最小化生成器目标函数, 同样采取误差反传与改进后的 Adam 优化更新生成器的权值; 同理, 对生成器进行优化并更新权值。

**Step 8** 判断当前迭代次数  $t$  与  $N_{\text{offset}}$  的大小关系, 然后更新学习率的值。

**Step 9** 循环计算 Step 3 至 Step 9, 直至达到最大循环次数。

### 3 实验

将本文提出的方法在 2 个不同艺术风格的数据集上进行实验, 通过与 DualGAN 算法<sup>[9]</sup>、CycleGAN 算法<sup>[8]</sup>、Pix2Pix 算法<sup>[7]</sup>和 GAN 算法<sup>[8]</sup>的对比, 验证本文算法的有效性, 并将该方法应用到具有中国特色的水墨画进行风格转换。在水墨画风格转换实验中, 图片集 X 采用 horse2zebra 数据集中马的图片, 图片集 Y 采用从百度图片爬取的徐悲鸿所绘水墨画马的图片(图 2)。利用 Python 的 OpenCV 库对爬取到的图片进行预处理, 将所有图片大小裁剪到像素  $128 \times 128$ 。同时, 采用数据增强技术对图片集 Y 中的数据进行扩充。实验环境为 Intel Xeon Silver 4116 2.1 GHz CPU, 128 G 内存, GPU 为 NVIDIA TITAN XP, 深度学习框架 TensorFlow 1.8.0。



图片集 X

图片集 Y

图2 中国水墨画风格数据集 beihong-photo

#### 3.1 实验数据

采用人脸素描风格数据集 sketch-photo 和中国水墨画风格数据集 beihong-photo 进行实验, 其中 beihong-photo 由宁波大学计算科学与技术研究所收集并创建。sketch-photo 和 beihong-photo 的训练

集和测试集包含的样本数量见表 2。

表2 实验使用数据统计

| 数据集           | 训练集 X | 训练集 Y | 测试集 X | 测试集 Y |
|---------------|-------|-------|-------|-------|
| sketch-photo  | 995   | 995   | 199   | 199   |
| beihong-photo | 800   | 400   | 200   | 150   |

#### 3.2 超参数

本文提出方法基于 TensorFlow 深度学习框架实现, 使用 GPU 进行运算加速。优化算法采用改进后 Adam, 生成器初始学习率  $\eta^g$  为 0.000 2, 判别器初始学习率  $\eta^d$  为 0.000 1。前 100 回合学习率不发生变化, 后 100 回合学习率开始线性衰减。最大循环次数  $N_{\text{epoch}}$  为 200, 重建损失函数项的比重  $\lambda$  为 4。

#### 3.3 结果分析

图 3 给出了 5 组实验在 sketch-photo 数据集的风格转换结果。从图 3 可以发现, GAN 算法和 Pix2Pix 算法生成的素描肖像比较模糊, 尤其在眼睛部位; DualGAN 算法在输入图像色系偏暗情况下, 生成的素描肖像不太理想, 出现大片黑块; CycleGAN 算法在输入图像佩戴眼镜情况下, 生成的素描肖像在眼镜部分发生了扭曲, 没有将眼镜与眉毛二者之间的细节信息很好地保存。本文提出的算法能够很好地保留人脸五官细节信息, 同时可以保证生成的素描肖像轮廓不会扭曲。



输入 GroundTruth 本文算法 DualGAN CycleGAN Pix2Pix GAN

图3 本文算法与其他方法在人脸素描数据集 sketch-photo 上的实验结果对比

选取 Kernel MMD (Maximum Mean Discrepancy, MMD)<sup>[21]</sup>、Wasserstein 距离、FID (Fréchet Inception Distance, FID)<sup>[22]</sup>以及 IS (Inception Score, IS)<sup>[23]</sup> 4 种评价指标, 对本文算法与 DualGAN 算法、CycleGAN 算法在 sketch-photo 数据集上进行定量

评估, 评估结果见表 3. Kernel MMD 评估指标是在给定核函数下度量了真实分布与生成分布之间的差异; Wasserstein 距离评估指标是通过计算 Wasserstein 距离来区分真实分布与生成分布的差异性; FID 是将真实分布与生成分布建模为高斯随机变量, 并计算 Fréchet 距离来衡量真实分布与生成分布之间的差异性; MMD、Wasserstein 距离以及 FID 指标分别从不同角度来衡量真实分布与生成样本之间的差异性, 值越小模型越好. 从表 3 可发现, 本文提出的方法能更好地拟合目标数据集的真实样本分布. IS 评估指标是计算生成图像在 Inception 网络中产生的 logit 响应和边缘分布的 KL 散度来衡量 GAN 生成图片的质量以及多样性, 值越大模型越好. 从表 3 可发现, 本文算法与其他 2 种方法在该指标上差异不大, IS 是评价现实图片与生成图片的区别, 而用在绘画风格上效果不明显.

表 3 模型评估结果

| 模型       | MMD   | Wasserstein | FID   | IS    |
|----------|-------|-------------|-------|-------|
| 本文算法     | 0.056 | 0.059       | 0.034 | 1.323 |
| DualGAN  | 0.347 | 0.176       | 0.187 | 1.339 |
| CycleGAN | 0.130 | 0.078       | 0.060 | 1.326 |

将本文提出方法应用到具有中国特色的水墨画风格转换中, 结果如图 4 所示. 从图 4 可发现, 本文算法可以学习到徐悲鸿先生的绘画风格, 尤其在马的鬃毛和马尾等部位的处理.



图 4 本文算法在水墨画风格数据集

beihong-photo 上的实验结果

#### 4 结语

本文结合结构相似性指数和最小二乘生成对抗网络, 提出一种在图像内容特征和风格特征保持平衡的图像艺术风格化算法. 实验结果表明, 在

成功训练完一个 GAN 后, 利用生成器的强大生成能力, 可以获得高质量风格化作品. 下一步将对本模型进行改进, 使得模型能够学习到更加抽象的艺术绘制风格.

#### 参考文献:

- [1] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553):436-444.
- [2] Bengio Y. Learning deep architectures for AI[J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-55.
- [3] 曹毅. 基于图像的水墨画绘制方法的研究[D]. 长春: 吉林大学, 2012.
- [4] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Communication Society, 2016:2414-2423.
- [5] Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution[C]//European Conference on Computer Vision. Berlin: Springer, 2016:694-711.
- [6] Ulyanov D, Lebedev V, Vedaldi A, et al. Texture networks: Feed-forward synthesis of textures and stylized images [C]//Proceedings of 33rd International Conference on Machine Learning. New York: ACM, 2016:1349-1357.
- [7] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Computer Society, 2017: 5967-5976.
- [8] 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. New York: IEEE Computer Society, 2017:2242-2251.
- [9] Yi Z, Zhang H, Tan P, et al. DualGAN: Unsupervised dual learning for image-to-image translation[C]//IEEE International Conference on Computer Vision. New York: IEEE Computer Society, 2017:2868-2876.
- [10] 谢志峰, 叶冠桦, 闫淑真, 等. 基于生成对抗网络的 HDR 图像风格迁移技术[J]. 上海大学学报(自然科学版), 2018, 24(4):524-534.
- [11] 林嘉骏, 诸葛晶晶, 张晴. 基于像素级生成对抗网络的复杂场景灰度图彩色化[J]. 计算机辅助设计与图形学学报, 2019, 31(3):439-446.
- [12] Wang Z, Bovik A C, Sheikh H R, et al. Image quality

- assessment: From error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [13] Mao X, Li Q, Xie H, et al. Least squares generative adversarial networks[C]//IEEE International Conference on Computer Vision. New York: IEEE Computer Society, 2017:2813-2821.
- [14] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Massachusetts: MIT Press, 2014: 2672-2680.
- [15] Nguyen A, Clune J, Bengio Y, et al. Plug & play generative networks: Conditional iterative generation of images in latent space[C]//IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Computer Society, 2017:3510-3520.
- [16] Li C, Wand M. Precomputed real-time texture synthesis with markovian generative adversarial networks[C]//European Conference on Computer Vision. Berlin: Springer, 2016:702-716.
- [17] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Computer Society, 2017:105-114.
- [18] Zhang H, Xu T, Li H. StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks[C]//IEEE International Conference on Computer Vision. New York: IEEE Computer Society, 2017:5908-5916.
- [19] 王坤峰, 苟超, 段艳杰, 等. 生成式对抗网络 GAN 的研究进展与展望[J]. 自动化学报, 2017, 43(3):321-332.
- [20] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Computer Society, 2016:770-778.
- [21] Gretton A, Fukumizu K H, Teo C, et al. A kernel method for the two-sample-problem[C]//Advances in Neural Information Processing Systems 19 Proceedings of the 2006 Conference. Massachusetts: MIT Press, 2007:1672-1678.
- [22] Heusel M, Ramsauer H, Unterthiner T, et al. GANs trained by a two time-scale update rule converge to a local nash equilibrium[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Massachusetts: MIT Press, 2017: 6626-6637.
- [23] Salimans T, Goodfellow I J, Zaremba W, et al. Improved techniques for training GANs[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Massachusetts: MIT Press, 2016: 2232-2242.
- [18] Zhang H, Xu T, Li H. StackGAN: Text to photo-realistic

## Image stylization transfer using generative adversarial neural networks

DONG Wei, ZHAO Jieyu\*

( Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China )

**Abstract:** This paper proposes a novel neural style transfer algorithm. The Structural Similarity Index Measurement (SSIM) with the Least Squares Generate Adversarial Nets (LSGAN) is adopted to build a model for image repainting with neural styles. Through the adversarial training and reconstruction constraints, the model can generate a more realistic and reasonable stylized picture. Two groups of experiments are performed with CHUK sketch-photo dataset and a Chinese ink-painted style dataset (named beihong-photo which is constructed by the authors' lab). Experimental results show that the presented method achieves a more acceptable state-of-art effect compared with those from the most popular algorithms, including DualGAN, CycleGAN, Pix2Pix and the classic GAN.

**Key words:** structural similarity index measurement; least squares generative adversarial nets; image art stylization; non-photorealistic rendering

(责任编辑 史小丽)