

CycleGAN-SN: 结合谱归一化和 CycleGAN 的图像风格化算法

余艳杰, 孙嘉琪, 葛思擘, 杨清宇
(西安交通大学电子与信息学部, 710049, 西安)

摘要: 为解决 CycleGAN 算法图像风格化质量不高、网络稳定性不强的问题, 提出了 CycleGAN-SN 算法。在 CycleGAN 算法判别网络的每一个卷积层后添加谱归一化层, 通过幂迭代法估算卷积层参数矩阵的谱范数, 采用随机梯度下降法更新卷积层参数。由于参数在每一次更新中的变化量很小, 只需迭代一次即可快速估算出矩阵的最大奇异值。根据得到的最大奇异值, 对卷积层参数进行归一化处理, 使得整个判别网络满足 1-Lipschitz 连续。在 4 个常用风格图像数据集上进行实验, 并与 CycleGAN 算法进行对比, 结果表明: 所提算法能够在保留原有图像细节的基础上, 生成色彩鲜艳、纹理清晰、风格渲染充分的风格化图像; 在训练过程中的损失函数振荡幅度小, 能够使用更大的学习率进行训练, 稳定性较强; 能够有效减少网络收敛所需的步数, 收敛速度较快; 在测试阶段一次性风格化 751 幅图像时, 时间最多仅增加 0.63 s, 几乎没有额外的时间消耗。

关键词: 图像风格化; 谱归一化; CycleGAN

中图分类号: TP18 **文献标志码:** A

DOI: 10.7652/xjtuxb202005018 **文章编号:** 0253-987X(2020)05-0133-09



OSID 码

CycleGAN-SN: Image Stylization Algorithm Combining Spectral Normalization and CycleGAN

YU Yanjie, SUN Jiaqi, GE Sibao, YANG Qingyu

(Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: To solve the problems of lower image stylization quality of CycleGAN algorithm and weaker network stability, a CycleGAN-SN algorithm is proposed. This algorithm adds a spectral normalization layer after each convolutional layer of the discriminator network of CycleGAN algorithm, then estimates the spectral norm of the parameter matrix of the convolutional layer by a power iteration method, and updates the parameters of the convolutional layer with a random gradient descent method. The parameter change in each update is small, so it only needs to iterate once to estimate the maximum singular value of the matrix. The maximum singular value obtained is used to normalize the parameters of the convolution layer, which makes the entire discriminative network meet the 1-Lipschitz continuity. Experiments are performed on 4 commonly used style image datasets. Compared with CycleGAN algorithm, the results show that the proposed algorithm can generate higher-quality stylized pictures with bright colors, clear textures, and sufficient style renderings while retaining the details of the original pictures. The loss function of the proposed algorithm in the training process has a small oscillation amplitude. The network can be trained at higher learning rate and with greater stability. The proposed

algorithm effectively reduces the number of steps required for network convergence and has a faster convergence rate. The proposed algorithm stylizes 751 images at a time during the test phase with protracting the period only by a maximum of 0.63 seconds, almost no extra time consumption.

Keywords: image stylization; spectral normalization; CycleGAN

图像风格化是指将一张普通内容图像转变为一张具有艺术风格的风格化图像。Gatys 等在 2016 年创造性地用深层卷积网络提取图像特征,利用格拉姆矩阵刻画图像风格,迭代优化噪声图像,生成了视觉效果良好的风格化图像^[1]。图像风格化因此受到工业界和学术界的广泛关注。

但是,Gatys 等所提的风格化方法比较耗时,因此 Justin 等和 Ulyanov 等利用转换网络将迭代优化过程转移到训练阶段,实现了快速单风格转换^[2-3]。为实现多风格转换,Dumoulin 等在转换网络中加入条件实例归一化(CIN)结构,利用实例归一化(IN)^[4-5]中的仿射变换参数实现了多风格转换^[6]。在任意风格转换方面,Ghiasi 等在 CIN 结构的基础上,提出风格预测网络求解对应的仿射变换参数,实现了任意风格转换^[7]。

此外,Huang 等提出了 AdaIN 结构,将风格图像的均值和方差作为仿射变换参数,也可快速完成任意风格转换^[8]。Li 等证明了采用格拉姆矩阵刻画图像风格的理论依据,该依据表明,图像风格化可以看成是一个领域自适应^[9]问题,并利用其他几种形式的风格损失函数来实现图像风格化^[10]。Li 的结论促使越来越多的科研人员利用生成对抗网络(GAN)来丰富图像风格化技术^[11]。

Zhu 等提出了 CycleGAN 算法,作为一种无监督学习算法,该算法可一次性对整个数据集内的所有图像进行风格化^[12]。但是,该算法为稳定网络的训练过程而采用最小二乘函数^[13]作为判别网络的损失函数,导致生成样本的多样性不足,图像风格化质量不高。当学习率增大时,算法的稳定性也会受到影响。

谱归一化(SN)^[14]通过求解矩阵的谱范数来归一化矩阵参数。归一化过程能够使判别网络的每一个卷积层从理论上完全满足 1-Lipschitz 连续,从而增强网络的稳定性。

本文在 CycleGAN 算法的判别网络中引入谱归一化,提出 CycleGAN-SN 算法。与 CycleGAN 算法相比,CycleGAN-SN 算法的泛化能力强,有效地解决了生成样本多样性不足的问题,提高了风格

化图像的质量,增强了网络的稳定性,提升了网络的收敛速度。

1 CycleGAN 算法

CycleGAN 算法不需要将数据集提前配对,即可实现两个数据集间的相互转化。在转换时,内容图像数据集 X 通过生成网络 G 和判别网络 D_Y ,生成风格化图像数据集 Y ,实现风格化过程。

图像风格化对应两部分内容:提取并融合风格图像的风格,与对抗损失函数相对应;保留内容图像的信息,与循环一致性函数相对应。

对抗损失函数为

$$L_{\text{GAN}}(\mathbf{G}, D_Y, X, Y) = E_{y \sim P_{\text{data}}(y)}(\log D_Y(y)) + E_{x \sim P_{\text{data}}(x)}(\log(1 - D_Y(\mathbf{G}(x)))) \quad (1)$$

式中: x, y 分别表示内容图像和风格图像, $x \in X, y \in Y$; $E_{i \sim P_{\text{data}}(i)}$ 表示在 i 服从 $P_{\text{data}}(i)$ 的情况下求均值, $P_{\text{data}}(i)$ 表示数据集 I 的分布, $i = x, y, I = X, Y$ 。式(1)与 GAN 中的损失函数定义方式相似,但采用最小二乘损失函数替代最大对数似然函数。

循环一致性损失函数为

$$L_{\text{cyc}}(\mathbf{G}, \mathbf{F}) = E_{x \sim P_{\text{data}}(x)}(\|\mathbf{F}(\mathbf{G}(x)) - x\|_1) + E_{y \sim P_{\text{data}}(y)}(\|\mathbf{G}(\mathbf{F}(y)) - y\|_1) \quad (2)$$

式(2)用于减少生成网络 G 可能的映射函数空间,保留转换后图像的内容信息。

根据对抗损失函数和循环一致性函数,可得网络的总损失函数为

$$L(\mathbf{G}, \mathbf{F}, D_X, D_Y) = L_{\text{GAN}}(\mathbf{G}, D_Y, X, Y) + L_{\text{GAN}}(\mathbf{F}, D_X, Y, X) + \lambda L_{\text{cyc}}(\mathbf{G}, \mathbf{F}) \quad (3)$$

式中 λ 为控制参数,控制着两部分损失函数的相对比重。

2 CycleGAN-SN 算法

CycleGAN 算法采用最小二乘损失函数替代最大对数似然函数以缓解梯度消失问题,该函数对离群样本惩罚过大,导致生成样本只能对真实样本进行简单模仿和细微改动,生成样本的多样性下降。图像风格化是指在原来内容图像的基础上融入风格图像的风格,包括色彩、纹理等多种信息,不仅仅是

对内容图像的简单模仿。因此,CycleGAN 算法产生的图像风格化质量不高。

CycleGAN-SN 算法由 CycleGAN 算法改进而来,通过在判别网络的每一个卷积层后引入谱归一化层,提高风格化图像的质量,增强网络的稳定性,并进一步提升网络的收敛速度。

2.1 优势分析

在提高图像风格化的质量方面,CycleGAN-SN 算法使用谱归一化给判别网络施加一个全局化的正则,防止网络参数向某一个方向强化。线性算子的 Lipschitz 连续性常数仅由最大奇异值决定,进而允许参数矩阵能够使用尽可能多的特征去生成图像,增加生成样本的多样性,提高图像风格化的质量。

在提升网络稳定性方面,由于 CycleGAN 算法以 GAN 的结构为基础,所以稳定性不强的原因与 GAN 的相同,都是由于判别网络的控制性能差而导致的。GAN 稳定性定理指出,当判别网络的输入和输出满足 1-Lipschitz 连续时,判别网络的控制性能增强,GAN 的训练过程就会稳定^[15]。

神经网络一般由多层组成,经过多层神经网络复合而成的判别网络本质上相当于一个复合函数。复合函数的 Lipschitz 理论指出,只要单个函数满足 1-Lipschitz 连续,则由它们所组成的复合函数也一定满足 1-Lipschitz 连续。

CycleGAN 算法的判别网络激活函数均为 Leaky Relu 函数^[16],该函数满足 1-Lipschitz 连续。因此,只要判别网络的每一个卷积层满足 1-Lipschitz 连续,整个判别网络就满足 1-Lipschitz 连续。CycleGAN-SN 算法将谱归一化应用到判别网络的每一个卷积层中,使得卷积层满足 1-Lipschitz 连续,进而增强网络的稳定性。详细分析如下。

对于卷积层 t :输入为 $\mathbf{h}$,卷积层参数矩阵为 $\mathbf{A}$,矩阵 $\mathbf{A}$ 的谱范数 $\sigma(\mathbf{A})$ 的计算方式为

$$\sigma(\mathbf{A}) = \sup_{\mathbf{h} \neq 0} \frac{\|\mathbf{A}\mathbf{h}\|_2}{\|\mathbf{h}\|_2} = \sup_{\|\mathbf{h}\|_2 \leq 1} \|\mathbf{A}\mathbf{h}\|_2 \quad (4)$$

式中: $\sigma(\mathbf{A})$ 等于矩阵 $\mathbf{A}$ 的最大奇异值;卷积层 t 的 Lipschitz 连续性常数等于该卷积层参数矩阵的谱范数。通过求解卷积层参数矩阵 $\mathbf{W}$ 的谱范数 $\sigma(\mathbf{W})$,使得归一化后的卷积层参数矩阵 $\mathbf{W}_{\text{SN}}$ 的最大奇异值为 1。 $\mathbf{W}_{\text{SN}}$ 的计算方式为

$$\mathbf{W}_{\text{SN}} = \mathbf{W} / \sigma(\mathbf{A}) \quad (5)$$

卷积层经过谱归一化后,其最大奇异值均为 1,卷积层的输入输出均满足 1-Lipschitz 连续,由此得出判别网络的输入输出满足 1-Lipschitz 连续。

2.2 最大奇异值快速估计

直接计算判别网络中的每一个卷积层参数矩阵的最大奇异值,所需计算量较大。当矩阵的主特征向量不存在多重性,并且第一个右奇异值特征向量不正交于第一个左奇异值特征向量时,可以使用幂迭代法估算矩阵的最大奇异值。为此,从球状的均匀分布上随机采样第一个右奇异值特征向量,此时它与第一个左奇异值特征向量正交的概率为 0。

随机初始化一个向量 $\mathbf{u}$ 作为参数矩阵 $\mathbf{W}$ 的右奇异值向量,则左奇异值向量 $\mathbf{v}$ 为

$$\mathbf{v} = \frac{\mathbf{W}^T \mathbf{u}}{\|\mathbf{W}^T \mathbf{u}\|_2}; \quad \mathbf{u} = \frac{\mathbf{W} \mathbf{v}}{\|\mathbf{W} \mathbf{v}\|_2} \quad (6)$$

式(6)经过多次迭代以后,就可以估算出矩阵 $\mathbf{W}^T \mathbf{W}$ 最大特征值 λ_1 的开方,也就是 $\mathbf{W}$ 的最大奇异值,即

$$\sqrt{\lambda_1} = \mathbf{u}^T \mathbf{W} \mathbf{v} \quad (7)$$

2.3 算法描述

当用随机梯度下降法更新卷积参数矩阵 $\mathbf{W}$ 时, $\mathbf{W}$ 在每一轮更新中的变化量很小。在这种情况下,向量 $\mathbf{u}$ 和 $\mathbf{v}$ 的多次迭代计算只需一次迭代即可,既提高了运算速度,网络也同样可以获得良好的性能。

本文提出的 CycleGAN-SN 算法中的谱归一化步骤如下。

(1)随机初始化一个向量 $\mathbf{u}_l$ (各个层的 $\mathbf{u}_l$ 采样于各向同性分布)。

(2)采用幂迭代的方法对判别网络的每一个卷积层参数矩阵 $\mathbf{W}^l$ 进行归一化处理,公式为

$$\left. \begin{aligned} \mathbf{v}_l &\leftarrow (\mathbf{W}^l)^T \mathbf{u}_l / \|\mathbf{W}^l\|^T \mathbf{u}_l\|_2 \\ \mathbf{u}_l &\leftarrow \mathbf{W}^l \mathbf{v}_l / \|\mathbf{W}^l \mathbf{v}_l\|_2 \end{aligned} \right\} \quad (8)$$

(3)计算归一化以后的卷积层参数矩阵 $\mathbf{W}_{\text{SN}}$,公式为

$$\left. \begin{aligned} \sigma(\mathbf{W}^l) &= \mathbf{u}_l^T \mathbf{W}^l \mathbf{v}_l \\ \mathbf{W}_{\text{SN}}^l(\mathbf{W}^l) &= \mathbf{W}^l / \sigma(\mathbf{W}^l) \end{aligned} \right\} \quad (9)$$

(4)在数据集 D_M 上,采用随机梯度下降的方法更新卷积层参数矩阵 $\mathbf{W}^l$,公式为

$$\mathbf{W}^l \leftarrow \mathbf{W}^l - \alpha \nabla_{\mathbf{W}^l} l(\mathbf{W}_{\text{SN}}^l(\mathbf{W}^l), D_M) \quad (10)$$

式中 α 为学习率。

在判别网络中,由谱归一化操作所组成的层称为谱归一化层,简称 SN 层。SN 层的位置在卷积层之后,激活函数之前。

3 实验结果对比分析

实验平台 CPU 为 Intel(R) Core(TM) i7-7820X,显卡为丽台 P5000 16 GB 独显,操作系统为

64 位 ubuntu16.04。在 tensorflow-gpu 1.1 下进行仿真实验。

仿真实验所采用的 4 个风格图像数据集分别为 monet2photo、vangogh2photo、ukiyo2photo、cezanne2photo,可以在 WikiArt 中下载,4 个风格图像训练集的数据量依次为 1 072、400、563、525 幅。1 个内容图像数据集从 Flickr 数据集中获取,对应的训练集由 6 287 幅风景图像组成。总损失函数中的 λ 为 10,采用 Adam 函数,batch 的大小为 1。

生成网络采用 Justin 等的网络结构^[2],该网络由两个步长为 2 的卷积、若干残差块和两个步长为 1/2 的分数卷积构成,利用 IN^[5] 替代批量归一化 (BN)^[17],使用 9 个残差块^[18] 进行图像风格化。判别网络采用 70×70 的 PatchGANs^[19-21],可以同时判别图像中 70×70 个重叠补丁区域中的真假情况。

本文采用 Shrivastava 等采取的措施,将过去一阶段生成的图像输入判别网络,而不是将生成网络最新生成的图像作为判别网络的输入^[22]。

为了刻画 CycleGAN-SN 算法相对于 CycleGAN 算法的改进效果,从视觉效果、网络稳定性和网络收敛速度共 3 个方面进行评价。

3.1 视觉效果

在 GAN 生成图像样本的视觉效果评价方面,有 inception score^[23]、frechet inception distance^[24] 等量化指标,通过评估生成图像的清晰程度、多样性、与训练集的近似程度等指标,刻画 GAN 模型性能。在图像风格化方面,风格化图像的清晰程度与风格图像的纹理背景有关。在风格图像背景朦胧、纹理不清晰时,风格化图像也会不清晰。因此,清晰程度很难准确反映图像风格化的效果。在多样性和近似程度方面,风格化图像不同于内容图像和风格图像,而是融合了两者的特征,无法采用多样性和近似程度指标去衡量。

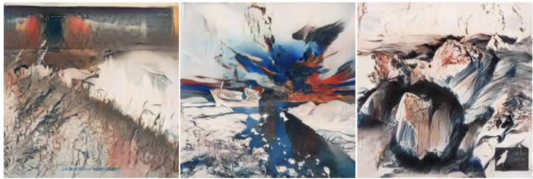
所以,在评价图像风格化视觉效果时,主要是根据人的主观感受。本文利用风格渲染程度、纹理细节清晰程度评价风格提取和融合的效果,利用内容细节评价内容图像的信息保留程度,从而得到图像风格化视觉效果的主观评价。风格渲染程度指图像色彩、图像纹理整体效果和亮度等信息,纹理细节指图像的清晰程度,内容细节指内容图像的内容信息保留程度。

在训练轮数为 200、学习率为 0.000 2 的条件下,CycleGAN 与 CycleGAN-SN 算法在 ukiyo2photo 和 monet2photo 数据集上产生的风格化图像如图 1

和图 2 所示。



(a)原图



(b)CycleGAN 算法产生的风格化图像



(c)CycleGAN-SN 算法产生的风格化图像

图 1 在 ukiyo2photo 数据集上产生的风格化图像



(a)原图



(b)CycleGAN 算法产生的风格化图像



(c)CycleGAN-SN 算法产生的风格化图像

图 2 在 monet2photo 数据集上产生的风格化图像

从图 1 可以看出,CycleGAN-SN 算法在风格渲染程度和内容细节方面均明显优于 CycleGAN 算法。在风格渲染程度方面,CycleGAN-SN 算法产生的图像颜色分布均匀,整体纹理效果好,CycleGAN

算法产生的图像颜色分布不均匀,存在黑色聚集现象,整体纹理效果一般。在内容细节方面,CycleGAN 算法产生的图像内容细节丢失,内容图像中的水流、森林等信息在风格化图像中均没有体现,而 CycleGAN-SN 算法则完整地保留了内容图像的信息。此外,CycleGAN-SN 算法在纹理细节上也略优于 CycleGAN 算法。

从图 2 可以看出,CycleGAN-SN 算法在风格渲染程度和纹理细节方面均明显优于 CycleGAN 算法。在风格渲染程度方面,CycleGAN-SN 算法产生的图像色彩鲜艳,亮度适中,而 CycleGAN 算法产生的图像色彩单调,亮度偏低。在纹理细节方面,CycleGAN 算法产生的图像纹理细节丢失,看起来很朦胧,CycleGAN-SN 算法产生的图像纹理细节清晰,效果较好。在内容细节上,两者都取得了不错的效果。

经过图 1 和图 2 的分析可知,使用 CycleGAN-SN 算法风格化的图像渲染充分、色彩鲜艳、纹理清晰,能够保留原来内容图像中的细节,具有良好的视觉效果。

3.2 稳定性

本节首先根据网络在训练过程中损失函数的变化情况进行分析,然后增大学习率,根据产生的风格化图像的视觉效果来综合判断网络稳定性。

在训练轮数为 200、学习率为 0.000 2 的条件下,CycleGAN 与 CycleGAN-SN 算法在 monet2photo 和 ukiyoe2photo 数据集上训练过程中的损失函数变化情况如图 3 和图 4 所示。

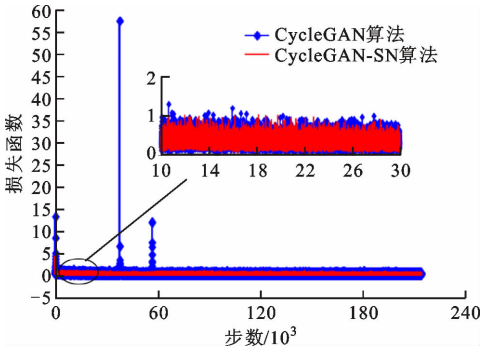


图 3 不同算法在 monet2photo 数据集上的损失函数变化情况

从图 3 可以看出:CycleGAN 算法在训练开始时,具有很大的损失函数,收敛过程中产生了很高的脉冲,因此网络的稳定性下降,收敛速度降低;CycleGAN-SN 算法仅在训练初始时有较大的损失函数,而后快速收敛,具有很强的稳定性和很快的收敛

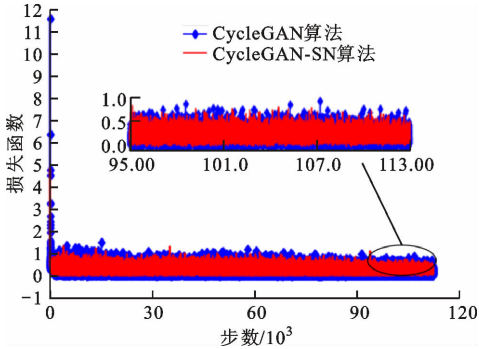


图 4 不同算法在 ukiyoe2photo 数据集上的损失函数变化情况

速度。

从图 4 可以看出:CycleGAN 算法在训练开始时产生了很大的损失函数,收敛过程整体趋于稳定,但在部分地方产生了小幅脉冲;CycleGAN-SN 算法在训练开始时的损失函数和训练过程中的损失函数振荡幅度均小于 CycleGAN 算法的,收敛速度更快,稳定性更强。

经过图 3 和图 4 的分析可知,CycleGAN-SN 算法不仅具有更强的稳定性,而且网络收敛所需的训练步数降低。

在训练轮数为 200 时,不同学习率下 CycleGAN 与 CycleGAN-SN 算法在 vangogh2photo 数据集上产生的风格化图像如图 5 所示。

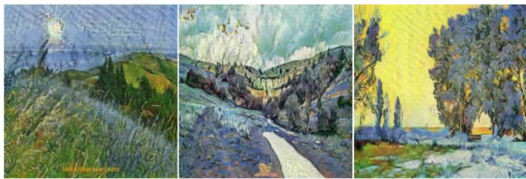
图 5b 和 5c 是增大学习率至 0.001 0 时的风格化图像,可以看出,CycleGAN-SN 算法在风格渲染程度上明显优于 CycleGAN 算法。在学习率增大的情况下,CycleGAN 算法不仅出现了纹理整体效果差、风格渲染不充分等问题,还出现了颜色分散、图像泛蓝光等新问题,说明算法的稳定性已经受到了破坏。但是,CycleGAN-SN 算法产生的风格化图像依然色彩鲜艳、亮度适中,且纹理整体效果好,并没有受到学习率增大的影响,说明 CycleGAN-SN 算法的稳定性更强。

图 5d 和 5e 是进一步增大学习率至 0.005 5 时的风格化图像,可以看出,CycleGAN 算法已经无法生成风格化图像,网络的稳定性已经被严重破坏,而 CycleGAN-SN 算法依然能够产生质量很高的风格化图像,CycleGAN-SN 算法具有很强的稳定性,能够有效抵御干扰。

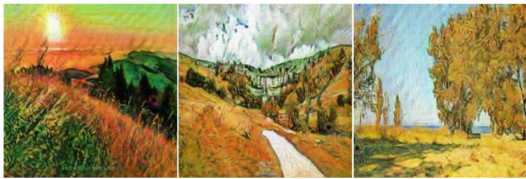
经过图 3~5 的分析可知,与 CycleGAN 算法相比,CycleGAN-SN 算法具有更强的稳定性,在训练时允许网络使用更大的学习率,网络的收敛速度也进一步加快。



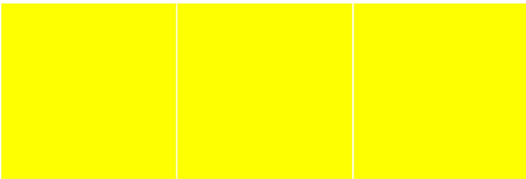
(a)原图



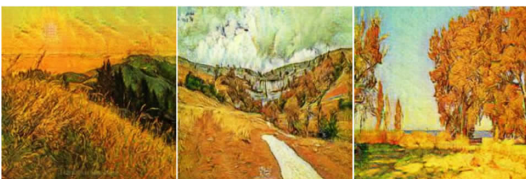
(b)学习率为 0.001 0 时 CycleGAN 算法产生的风格化图像



(c)学习率为 0.001 0 时 CycleGAN-SN 算法产生的风格化图像



(d)学习率为 0.005 5 时 CycleGAN 算法产生的风格化图像



(e)学习率为 0.005 5 时 CycleGAN-SN 算法产生的风格化图像

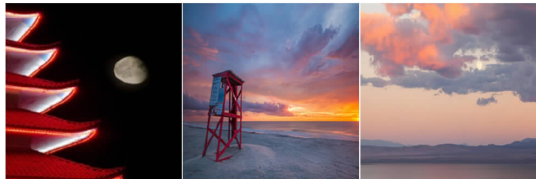
图 5 不同学习率时 2 种算法在 vangogh2photo 数据集上产生的风格化图像

3.3 收敛速度

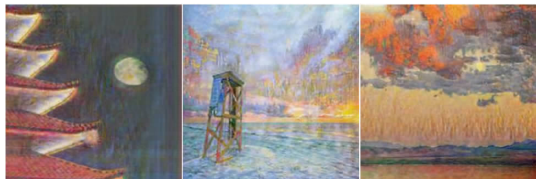
3.3.1 不同训练轮数下的视觉效果 在学习率为 0.000 2 的情况下,CycleGAN 算法分别训练 130、160、180 轮,CycleGAN-SN 算法训练 100 轮,得到风格化图像如图 6~8 所示。

从图 6 可以看出,CycleGAN-SN 算法在风格渲染程度、纹理细节和内容细节上均略优于原算法。

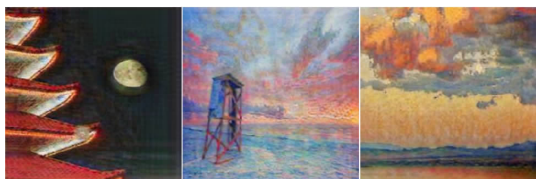
CycleGAN-SN 算法产生的图像纹理清晰、内容完整、风格渲染充分。



(a)原图



(b)训练轮数为 130 时 CycleGAN 算法产生的风格化图像

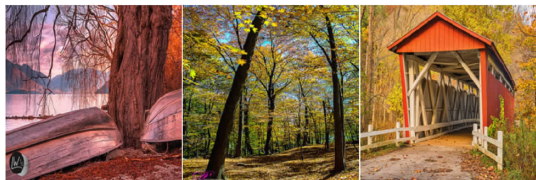


(c)训练轮数为 100 时 CycleGAN-SN 算法产生的风格化图像

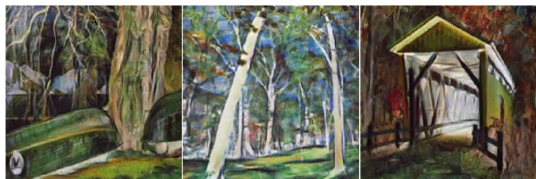
图 6 不同训练轮数时 2 种算法在 monet2photo 数据集上产生的风格化图像

从图 7 可以看出,CycleGAN 算法产生的图像颜色灰暗,内容细节部分丢失,纹理不清晰,CycleGAN-SN 算法则有效地避免了这些问题,在纹理细节和内容细节上均优于原算法。

从图 8 可以看出,两种算法在纹理细节、内容细节上的效果几乎相同,但 CycleGAN-SN 算法产生



(a)原图



(b)训练轮数为 160 时 CycleGAN 算法产生的风格化图像



(c)训练轮数为 100 时 CycleGAN-SN 算法产生的风格化图像

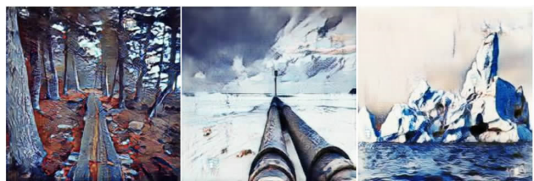
图 7 不同训练轮数时 2 种算法在 cezanne2photo 数据集上产生的风格化图像



(a)原图



(b)训练轮数为 180 时 CycleGAN 算法产生的风格化图像



(c)训练轮数为 100 时 CycleGAN-SN 算法产生的风格化图像

图 8 不同训练轮数时 2 种算法在 ukiyoe2photo 数据集上产生的风格化图像

的风格化图像色彩更加鲜艳,在风格渲染程度上略优于原算法。

经过图 6~8 的分析可知,CycleGAN-SN 算法在训练轮数为 100 的情况下,风格化图像的视觉效果优于训练轮数为 130、160、180 的 CycleGAN 算法的。当 CycleGAN-SN 算法的训练用时少于 CycleGAN 算法的时候,CycleGAN-SN 算法仍然具有更快的收敛速度。

3.3.2 收敛速度分析 CycleGAN-SN 算法在判别网络的每一个卷积层后加入了谱归一化层。CycleGAN-SN 算法只需一次迭代即可近似得到矩阵的最大奇异值,网络前向传播过程的计算量几乎不

变,但反向传播进行梯度更新时,计算量有所增加,导致 CycleGAN-SN 算法完成一轮训练所需要的时间比 CycleGAN 算法有所增加。CycleGAN 和 CycleGAN-SN 算法在 4 个数据集上训练一轮所需的时间如表 1 所示。

从表 1 可以看出,数据集内的图像数量从 400 到 1 072 不等,每轮训练所增加的时间依次从 5.09 s 到 16.52 s 不等。数据集内的图像数量越多,每轮训练增加的时间也相对增多。表 2 是 CycleGAN 和 CycleGAN-SN 算法在 4 个数据集上训练所需的总时间,可以看出,CycleGAN 算法在训练轮数为 130、160、180 时所需的总时间分别是 CycleGAN-SN 算法在训练轮数为 100 时的 1.27 倍、1.56 倍、1.75 倍。

表 1 CycleGAN 和 CycleGAN-SN 算法在 4 个数据集上训练一轮所需的时间

数据集	时间/s		增加的时间/s
	CycleGAN	CycleGAN-SN	
vangogh2photo	196.26	201.35	5.09
ukiyoe2photo	276.77	284.77	8.00
cezanne2photo	257.71	266.14	8.43
monet2photo	524.68	541.20	16.52

表 2 CycleGAN 和 CycleGAN-SN 算法在 4 个数据集上训练所需的总时间

数据集	时间/h			
	CycleGAN-SN(100 轮)	CycleGAN(130 轮)	CycleGAN(160 轮)	CycleGAN(180 轮)
vangogh2-photo	5.59	7.09	8.72	9.81
ukiyoe2-photo	7.90	9.99	12.30	13.84
cezanne2-photo	7.39	9.31	11.45	12.89
monet2-photo	15.03	18.95	23.32	26.23

从表 1 和表 2 可知,在 4 个数据集上,CycleGAN-SN 算法相对于 CycleGAN 算法的收敛速度均有明显提升。

3.4 测试用时

在测试阶段,只需一次前向传播即可生成风格化图像,CycleGAN 和 CycleGAN-SN 算法在 4 个数据集上的测试时间如表 3 所示。4 个风格图像数据

集的测试集数据量分别为 58、121、263、400,内容图像测试集的数据量为 751 幅。从表 3 可以看出,在测试阶段一次性风格化 751 幅图像时,CycleGAN-SN 算法相对于 CycleGAN 算法增加的时间最多为 0.63 s,最少为 0.18 s,几乎没有额外的时间消耗。

经过 3.1~3.3 节的分析可知,与 CycleGAN 算法相比,CycleGAN-SN 算法能够产生视觉效果更好的风格化图像,且该算法具有更强的稳定性和更快的收敛速度。这是因为谱归一化给判别网络施加了一个全局化的正则,防止网络参数向某一个方向强化,增强了网络的泛化能力,进而提高了图像风格化的质量。而且,谱归一化使判别网络的输入输出之间严格满足 1-Lipschitz 连续性要求,增强了网络的稳定性,提高了网络的收敛速度。

表 3 CycleGAN 和 CycleGAN-SN 算法在 4 个数据集上的测试时间

数据集	时间/s		增加的 时间/s
	CycleGAN	CycleGAN-SN	
vangogh2photo	35.45	35.63	0.18
ukiyo2photo	35.50	35.76	0.26
monet2photo	35.75	36.18	0.43
cezanne2photo	36.12	36.75	0.63

4 结 论

本文针对 CycleGAN 算法因采用最小二乘损失函数影响图像风格化质量以及在学习率增大时稳定性受到影响的问题,设计了 CycleGAN-SN 算法。该算法在判别网络的每一个卷积层后引入谱归一化层,仅需一次幂迭代即可估算出矩阵的最大奇异值,并对卷积层参数进行归一化处理,确保判别网络的输入输出满足 1-Lipschitz 连续。从视觉效果、网络稳定性、收敛速度和测试用时共 4 个角度出发,在 4 个风格图像数据集上进行大量实验,得出以下结论:

- (1)在视觉效果方面,CycleGAN-SN 算法能够生成色彩鲜艳、纹理清晰、风格渲染充分的风格化图像;
- (2)在网络稳定性方面,CycleGAN-SN 算法在训练过程中损失函数振荡幅度小,无明显尖峰脉冲,能够允许网络采用更大的学习率进行训练;
- (3)在收敛速度方面,CycleGAN-SN 算法训练 100 轮的视觉效果优于 CycleGAN 算法训练 130、160、180 轮的,有效地减少了算法收敛所需的步数;
- (4)在测试用时方面,CycleGAN-SN 算法一次

性风格化 751 幅图像时,与 CycleGAN 算法相比最多增加了 0.63 s,最少增加了 0.18 s,几乎没有额外的时间消耗。

在图像转换领域,CycleGAN 算法是近两年的主流做法,图像风格转换仅是图像转换领域的其中一个方面。本文将谱归一化应用到 CycleGAN 算法中,仅验证了其在图像风格转换上的作用。未来应进一步通过大量实验来验证谱归一化在图像转换领域各个方面的作用,以此证明其对图像转换的相关应用具有广泛的适用性。

参考文献:

- [1] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks [C] // Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2016: 2414-2423.
- [2] JUSTIN J, ALEXANDRE A, LI F F. Perceptual losses for real-time style transfer and super-resolution [C]//Proceedings of the 14th European Conference on Computer Vision. Cham, Germany: Springer, 2016: 694-711.
- [3] ULYANOV D, LEBEDEV V, VEDALDI A, et al. Texture networks: feed-forward synthesis of textures and stylized images [C]//Proceedings of the 33rd International Conference on Machine Learning. Princeton, NJ, USA: IMLS, 2016: 2027-2041.
- [4] ULYANOV D, VEDALDI A, LEMPITSKY V. Improved texture networks: maximizing quality and diversity in feed-forward stylization and texture synthesis [C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2017: 6924-6932.
- [5] LI Yanghao, WANG Naiyan, SHI Jianping, et al. Revisiting batch normalization for practical domain adaptation [C/OL] // Proceedings of the 5th International on Learning Representations. London, UK: ICLR, 2017. [2019-08-01]. <https://arxiv.org/pdf/1603.04779v4.pdf>.
- [6] DUMOULIN V, SHLENS J, KUDLUR M. A learned representation for artistic style [C/OL] // Proceedings of the 5th International on Learning Representations. London, UK: ICLR, 2017. [2019-08-01]. <https://arxiv.org/pdf/1610.07629.pdf>.
- [7] GHIASI G, LEE H, KUDLUR M, et al. Exploring the structure of a real-time, arbitrary neural artistic stylization network [C/OL] // Proceedings of the 2017

- British Machine Vision Conference. Guildford, UK: BMVA Press, 2017. [2019-08-01]. <https://arxiv.org/pdf/1705.06830.pdf>.
- [8] HUANG X, BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2017: 1501-1510.
- [9] PAN S J, YANG Q. A survey on transfer learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.
- [10] LI Y, WANG N, LIU J, et al. Demystifying neural style transfer [C/OL]// Proceedings of the 26th International Joint Conference on Artificial Intelligence. California, USA: IJCAI, 2017. [2019-08-01]. <https://arxiv.org/abs/1701.01036>.
- [11] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]// Proceedings of the 27th International Conference on Neural Information Processing Systems. Cham, Germany: Springer, 2014: 2672-2680.
- [12] 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. Piscataway, NJ, USA: IEEE, 2017: 2223-2232.
- [13] MAO X, LI Q, XIE H, et al. Least squares generative adversarial networks [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2017: 2794-2802.
- [14] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks [C/OL]// Proceedings of the 6th International on Learning Representations. London, UK: ICLR, 2018. [2019-08-01]. <https://arxiv.org/pdf/1802.05957.pdf>
- [15] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks [C]// Proceedings of the 34th International Conference on Machine Learning Princeton, NJ, USA: IMLS, 2017: 298-321.
- [16] MAAS A L, HANNUN A Y, NG A Y. Rectifier nonlinearities improve neural network acoustic models [C/OL]// Proceedings of the 30th International Conference on Machine Learning. Princeton, NJ, USA: IMLS, 2013. [2019-08-01]. http://robotics.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf.
- [17] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]// Proceedings of the 32nd International Conference on Machine Learning. Princeton, NJ, USA: IMLS, 2015: 448-456.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [19] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2017: 1125-1134.
- [20] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2017: 4681-4690.
- [21] LI C, WAND M. Precomputed real-time texture synthesis with Markovian generative adversarial networks [C]// Proceedings of the 14th European Conference on Computer Vision. Cham, Germany: Springer, 2016: 702-716.
- [22] SHRIVASTAVA A, PFISTER T, TUZEL O, et al. Learning from simulated and unsupervised images through adversarial training [C]// Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2017: 2107-2116.
- [23] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training GANs [C]// Proceedings of the 2016 Conference of Advances in Neural Information Processing Systems. Vancouver, Canada: NIPS, 2016: 2234-2242.
- [24] 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 2017 Advances in Neural Information Processing Systems. Vancouver, Canada: NIPS, 2017: 6629-6640.

(编辑 陶晴)