

基于CartoonGan的改进卡通化图片生成方法

张文天, 于 璪*

(安徽理工大学 计算机科学与工程学院, 安徽 淮南 232001)

摘要: 文章提出了一种改进的卡通图像生成网络模型,旨在增强卡通效果的同时保留语义信息。首先,设计一个显著性网络,为了防止过拟合问题以及进一步提取特征,在显著性网络上添加残差结构构成残差显著性网络,并将其拼接到CartoonGan上用来保留语义信息;其次,在前者基础上添加cbam注意力机制进一步提高卡通化效果;最后,在训练过程中为了防止训练不稳定和梯度消失使用最小二乘损失来替换交叉熵损失,并且引入显著性损失来约束显著性网络的训练。实验表明,通过在宫崎骏和新海诚两个画风的卡通数据集上进行测试,使用FID测试指标显示在两个数据集上都有一定的优化。

关键词: Cbam注意力机制;显著性网络;残差结构;CartoonGan

中图分类号: TP391.4

文献标识码: A

文章编号: 1008-9659(2024)02-0032-11

卡通艺术的起源可追溯至17世纪的荷兰,其后随着时间推移逐渐受到人们的青睐,如今已广泛应用于宣传、教育、电影、广告等众多领域。然而,传统的卡通创作方式耗时费力,需要综合考虑色彩、纹理、结构以及阴影等复杂要素。为了应对这一挑战,自动卡通图生成方法应运而生,通过输入待转换的图像,能够自动生成相应的卡通图,从而实现高效的图像生成。

在20世纪90年代中期,非逼真渲染(Non-Photorealistic Rendering, NPR)技术首次尝试了艺术风格的转换,并取得了初步的成功。在过去的十年里,深度学习方法首次在艺术风格转换领域取得突破。Gatys等人提出基于预先训练的VGG网络的特征提取方法^[1],虽然效果显著,但是其速度相对较慢。随后,Huang等学者通过引入前馈网络来加速训练过程^[2],Li等人则将卷积神经网络与马尔科夫随机场相结合,实现了更加高效的风格迁移^[3]。

2014年,生成对抗网络Gan^[4]的出现引发了剧烈轰动,尤其在图像生成方面取得了卓越的成果。随着Gan的快速发展,涌现出许多基于Gan的衍生模型。例如,CycleGan^[5],Pix2PixHD^[6],StarGan^[7],DRB-Gan^[8]等网络模型,解决了成对数据匹配,图像分辨率以及多模态等问题,进一步丰富了卡通图生成领域的研究。

尽管上述方法可以生成很好的艺术风格图片,但是在卡通图的生成上仍然不足,其原因是卡通图由清晰且锐利的边缘以及明显的颜色变化还有大量的颜色组成。因此为了更好地生成卡通图,一些专门用来生成卡通图的模型被提出。CartoonGan^[9]首先提出一种促进边缘的对抗性损失来突出边缘清晰度的卡通特征,实现了优秀的卡通迁移效果。在此之后涌现出很多优秀的卡通风格迁移网络,如AnimeGan^[10],White-Box^[11],SdpGan^[12],CartoonlossGan^[13]等,在模型轻量化,生成卡通图色块以及判别器的改动等方面都得到了很好的提升。

为了进一步优化卡通效果,文章提出一种通过引用Cbam^[14]和残差显著性网络,采用最小二乘损失替换原有的损失函数并且引入显著性损失的方法来进一步优化卡通效果的一种方法。

[收稿日期]2023-10-10

[修回日期]2023-11-01

[基金项目]2021年安徽省重点研究与开发计划项目(202104d07020010)。

[作者简介]张文天(1999-),男,硕士研究生,主要从事图像处理方面研究,E-mail:2819879788@qq.com.

*[通讯作者]于 璪(1973-),女,教授,主要从事区块链及图像处理方面研究,E-mail:3550707289@qq.com.

1 相关工作

1.1 Gan网络

生成对抗网络(Gan)是一种深度学习架构,它由两个神经网络组成:生成器(Generator, G)和判别器(Discriminator, D)。生成器接收随机噪声向量,并将其转换为逼真的数据样本。判别器的任务是区分真实数据样本和生成器产生的假样本,二者相互竞争,直到生成器生成的样本与真实数据的分布相似。Gan的损失函数如式(1)所示。

$$\min\max(D, G) = E_{x \sim P_{data}} [\log(D(x))] + E_{z \sim P_z} [\log(1 - D(G(z)))] \quad (1)$$

式(1)中的函数最大化D对真实样本判断的能力,同时最小化D对生成器G生成图片的判断能力。其中Pdata是真实样本的概率分布,Pz是随机编码z生成的生成样本的概率分布。

图1为Gan的结构图,在该图中随机噪声z经过判别器G产生生成样本G(z),随后将G(z)输入到D中,D对其进行判断。如果是真实数据那么就输出1,如果是生成的数据就输出0,最后经过不断地训练使得G生成的样本尽可能接近真实样本。

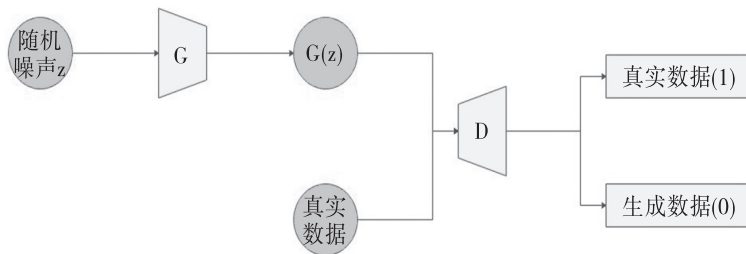


图1 Gan网络的结构图

1.2 CartoonGan模型

Chen等人提出了CartoonGan^[10],该网络包含生成器和判别器,生成器由平面卷积开始后经两个下采样模块进行空间压缩和编码,在8个布局相同的残差块构建内容和流形特征。最后经过两个上采样模块来输出和重构图像(图2)。

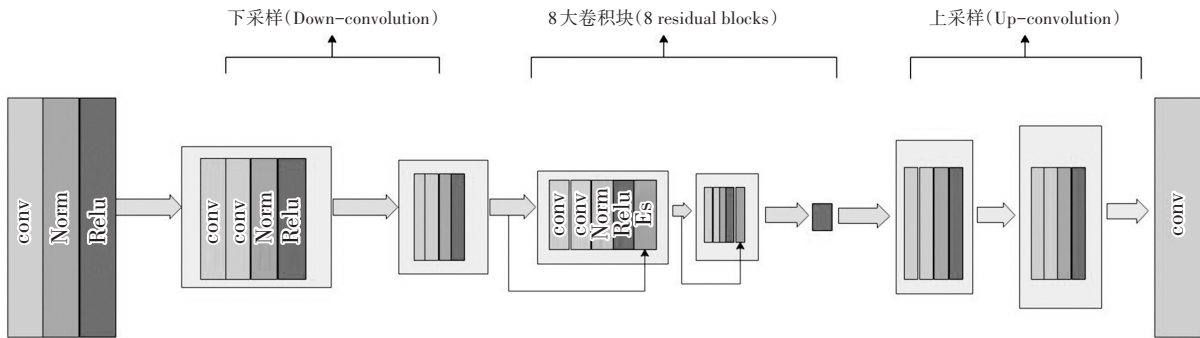


图2 CartoonGan生成器结构图

判别器相对简单,由于对卡通风格的识别依赖于图像的局部特征,因此使用补丁级别判别器来进行判别。判别器D被设计为浅层,在经过平面层阶段后,使用两个分级卷积块来降低分辨率,并提取基本的局部特征用于分类。最后,通过特征构造块和3×3卷积层来生成分类结果(图3)。

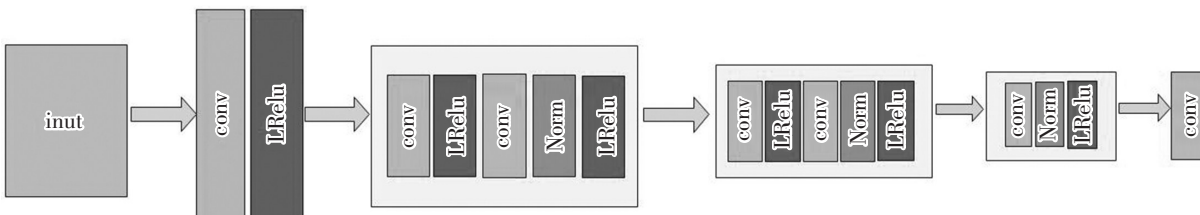


图3 CartoonGan判别器结构图

在训练过程中,首先进行预训练对图片进行预处理,正式训练采用对抗性损失函数来促进边缘信息,该损失函数强制鉴别器不仅要区分真实卡通图像和合成图像,还要区分边缘平滑的卡通图像,从而引导生成器产生更清晰的边缘并欺骗鉴别器。

2 研究方法

2.1 显著性网络

多年来,显著性检测一直受到人们的青睐,Itti 等人^[15]首次提出用来模拟人类视觉搜索的显著性模型,随着该研究的不断深入,显著性检测也被广泛应用在各个领域比如标题生成,图像分割和风格迁移等。

文章的显著性网络结构包括平面卷积层、两个上采样层和下采样层,但未包括8个残差块,其结构如图4所示。虽然使用显著性网络可以有效地保留语义内容,但过多注重内容上的保留会导致风格图片的风格迁移困难。为解决这一问题并保持图像生成质量,在拼接显著性网络时仅拼接前三层,这样可促进图像特征与显著性特征的融合,同时确保生成的图片不会过于受显著性特征的影响。

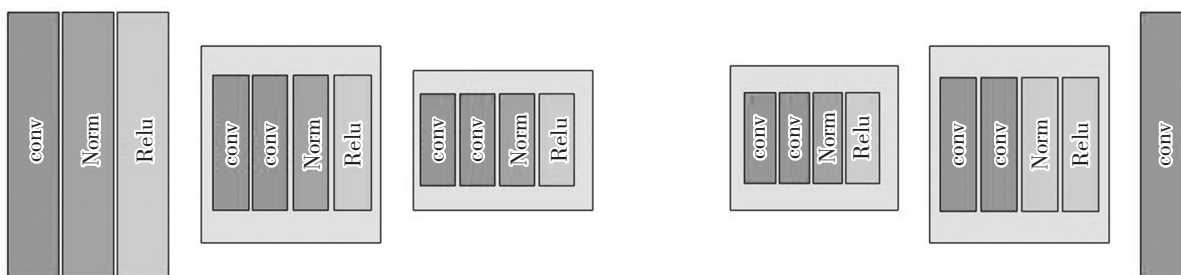


图4 显著性网络结构图

2.2 残差显著性网络

在深度学习中随着网络深度逐渐加深会遇到梯度消失和梯度爆炸。为了解决此类问题,He 等人提出了残差网络^[16],用来设计解决退化问题,同时梯度问题也有所解决,网络的性能也有所提升。

在残差结构中添加跳跃连接到第二层激活函数之前,激活函数的输入就由原来的 $H(x)=F(x)$ 变成 $H(x)=F(x)+x$ 以提高训练效率,这样的处理能够更有效地训练深度神经网络,其结构如图5所示。

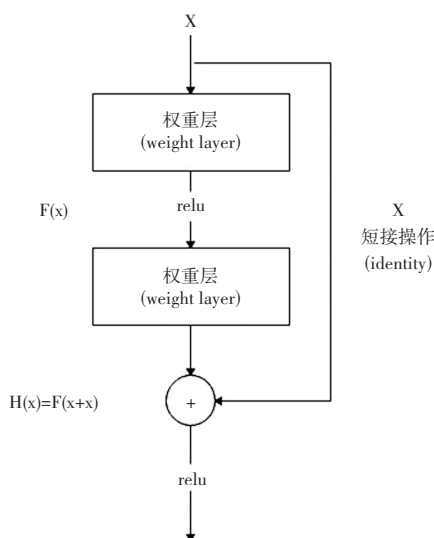


图5 残差结构图

受到该结构启发,文章在显著性网络上采用残差结构来进一步优化其能力,和显著性网络在生成器拼接的部分一样,只在该模块的前三层进行跳跃连接,上采样层保持原状。残差显著性网络结构如图6所示。

在该结构中,经过每个卷积层的特征层要使用实例归一化(InstanceNorm)进行归一化操作,其中第一个卷积层中的卷积核大小为 $7*7$,第二个卷积层中的两个卷积核大小均为 $3*3$,第三个卷积层中的两个卷积层

大小均为3*3.

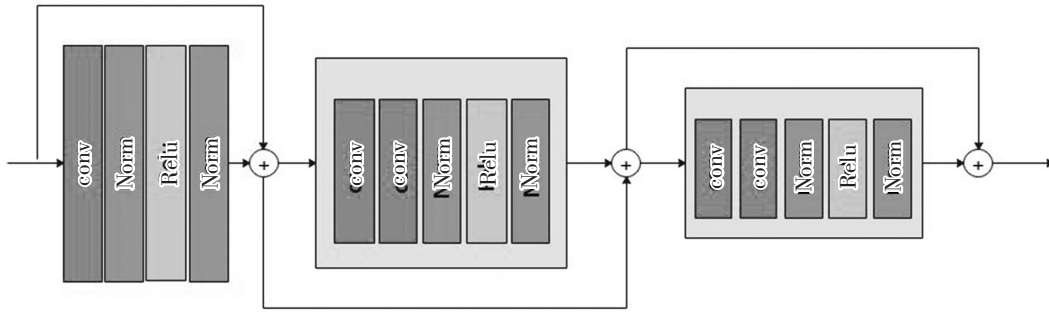


图6 残差显著性网络结构图

2.3 网络结构图

生成器部分将残差显著性网络拼接到 CartoonGan 的前三层后,受到 CBA-GAN^[17]的启发,分别在这三层的后面添加 cbam 注意力机制模块和卷积块,进而提高生成效果。文章提出的生成器模型如图 7 所示,其中 Ressel 为残差显著性网络,conv 为卷积块,Cbam 为注意力机制模块。

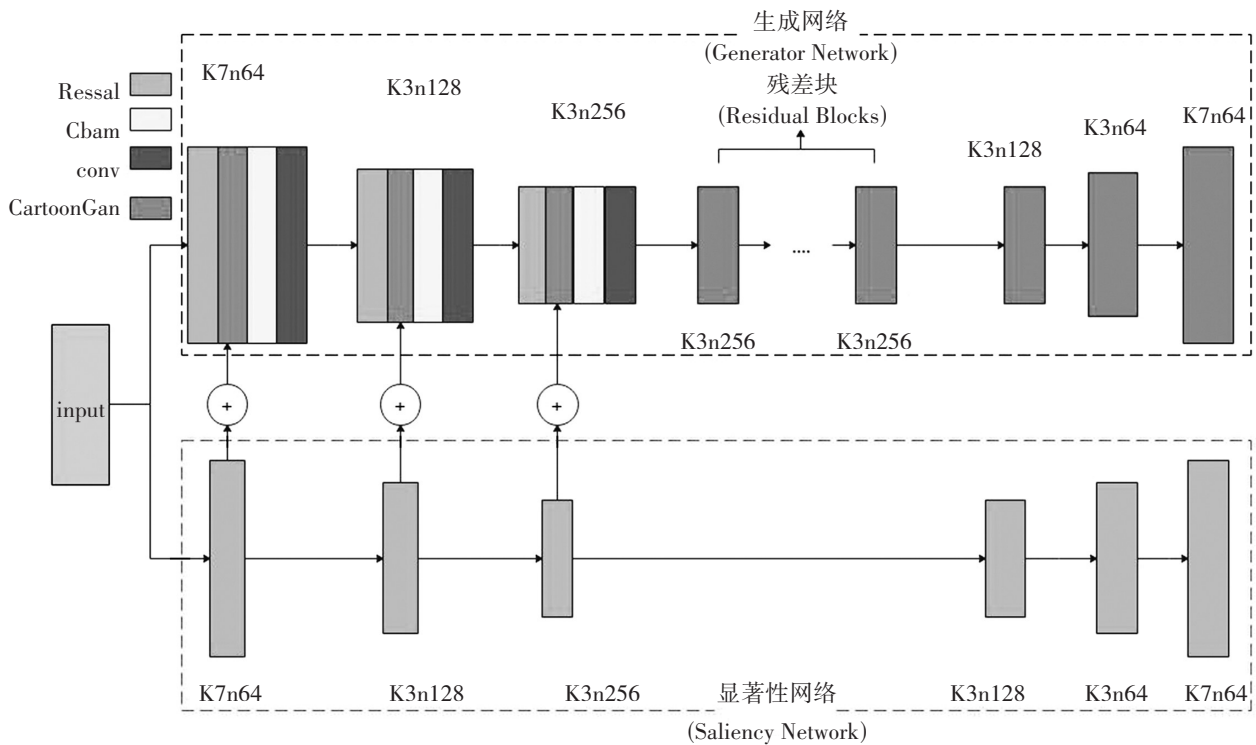


图7 生成器模型

判别器部分和 CartoonGan 判别器一样由一个卷积核大小为 3*3,步长为 1,激活函数采用泄漏线性整流函数(Leaky Relu)的卷积层开始,之后是两个步长为 2 的卷积层用来降低分辨率,最后是特征重构模块,该模块用步长为 1,卷积核大小为 3 的卷积层得到分类结果。每一层的归一化层后都使用 $\alpha=0.2$ 的泄漏线性整流函数(Leaky Relu)用作激活函数。

Cbam 是一种用于图像处理的注意力机制模块。它通过结合通道注意力和空间注意力,自适应地调整输入特征图的通道权重和空间权重,以提取更具信息量的特征表示。Cbam 结构如图 8 所示,其表示如式(2),式(3)所示。

$$F' = M_c(F) \otimes F \tag{2}$$

$$F'' = M_s(F') \otimes F' \tag{3}$$

式中,F 为输入特征,F' 为 F 经过通道注意力机制后并且和 F 相乘的结果,F'' 为 F' 经过空间注意力机制后和

F'相乘的结果。

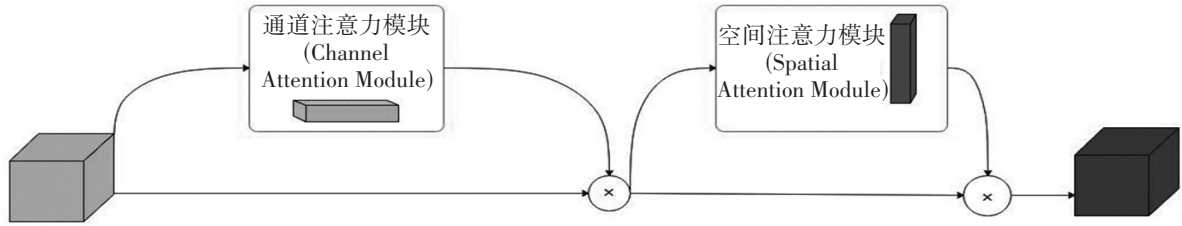


图8 cbam 结构图

Cbam 中的通道注意力机制能够自适应地学习每个通道的重要性,增强更有用的特征通道,抑制不重要的特征通道。将其拼接到生成器中,可以更好地捕捉图像的颜色,纹理等特征。通道注意力机制结构如图9所示,其表示如式(4)所示。

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (4)$$

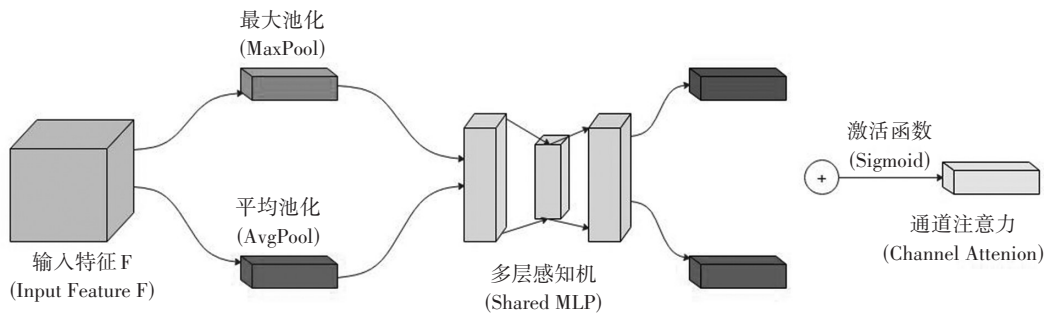


图9 通道注意力机制结构图

对卡通图来说,清晰的边缘是其重要特征,所以生成的卡通图具有良好的边缘就是一项主要的任务。cbam中的空间注意力机制通过计算并学习特征图中不同位置的权值,在训练过程中更加注重边缘特征,进而更好地生成边缘清晰的卡通图。空间注意力机制结构如图10所示,其表示如式(5)所示。

$$M_s(F) = \sigma(f^{7*7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (5)$$

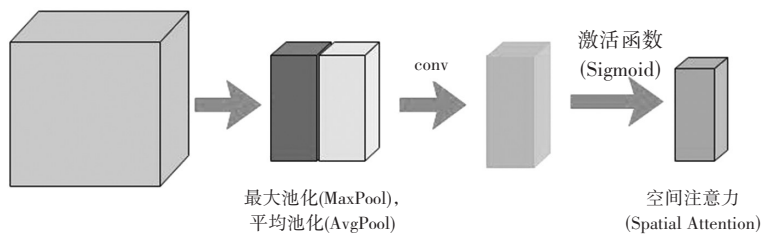


图10 空间注意力机制结构图

2.4 损失函数

将真实世界的场景映射到卡通图像的过程视为一个映射模型,其中 $S_{data}(p) = \{p_i | i = 1, 2, 3, \dots, N\}$ 为真实世界图片, $S_{data}(a) = \{a_i | i = 1, 2, 3, \dots, M\}$ 为卡通图片, $S_{data}(e) = \{e_i | i = 1, 2, 3, \dots, M\}$ 为和卡通图一一对应的卡通边缘模糊图, $S_{data}(z) = \{z_i | i = 1, 2, 3, \dots, N\}$ 为和真实世界图一一对应的显著性参考图片。为了进一步对训练过程进行限制,文章包含了3个损失函数,分别是对抗损失 L_{adv} , 用来优化生成器和判别器,内容损失 $L_{con}(G, D)$ 用来保存真实场景图片的内容,显著性损失 $L_{sali}(S)$, 用来优化显著性网络。

2.4.1 对抗损失

该损失是用来驱使生成器实现所需要的流形变换,受到相关方法^[3,18]的启发,由于CartoonGan网络模型参数较多,属于重量模型,在训练中会出现训练不稳定以及梯度消失等问题。为了缓解这一问题的出现,采用LsGan^[19]中的最小二乘损失来替换原有的交叉熵损失。原有损失函数如式(6)所示,文章采用的最小二乘损失如式(7)、式(8)所示。

$$L_{adv}(G,D)=E_{a \sim Sdata(a)}[\log D(a)]+E_{e \sim Sdata(e)}[\log(1-D(e))]+E_{p \sim Sdata(p)}[\log(1-D(G(p)))] \quad (6)$$

$$L_{adv}(G)=E_{p \sim Sdata(p)}[D(G(p))-1]^2 \quad (7)$$

$$L_{adv}(D)=E_{a \sim Sdata(a)}[(D(a)-1)^2]+E_{p \sim Sdata(p)}[D(G(p))]^2+E_{e \sim Sdata(e)}[D(e)]^2 \quad (8)$$

2.4.2 内容损失

该损失用来在图片卡通化过程中保留其图片内容。卡通风格迁移中,为了保留真实世界图片的内容信息,采用预训练好的VGG19网络来提取真实世界图片和生成卡通图片的高级语义内容, $L_{con}(G,D)$ 的表示如式(9)所示。

$$L_{con}(G,D)=E_{p \sim Sdata(p)}[\|VGG_1(G(p))-VGG_1(p)\|_1] \quad (9)$$

其中, l 表示VGG在第 l 层的特征图,在该方法中VGG网络的第1层是“conv4_4”。为了避免VGG网络中高层的真实内容图和生成卡通图的特征图之间巨大风格差异带来的影响,尤其是在特征发生显著变化的区域,采用的L1稀疏正则化来计算 $L_{con}(G,D)$ 。

2.4.3 显著性损失

显著性损失的目标是让显著性网络能够更好地检测出图片的显著性区域,该损失函数是最小化显著性网络的显著性输出和显著性参考之间的差异,其表示如式(10)所示。

$$L_{sali}(S)=E_{p \sim Sdata(p),z \sim Sdata(z)}[\|S(p)-z\|_1] \quad (10)$$

2.4.4 总体损失

总体的损失函数 $L(G,D,S)$ 由上述的三个损失函数相加得到的,其表示如式(11)所示。

$$L(G,D,S)=L_{adv}(G)+L_{adv}(D)+\omega_{con}L_{con}(G,D)+\omega_{sali}L_{sali}(S) \quad (11)$$

其中, ω_{con} , ω_{sali} 分别是内容损失权重参数和显著性损失权重参数,二者是权衡风格转换和内容保留的重要因素。根据Gatys等人的^[1]论文中提到的内容和风格权重之间比值关系可知,过大的 ω_{con} 会使图片在生成过程中保留更多原始图片的内容信息导致风格转换不明显,而过小的 ω_{con} 又会导致学习风格化过度从而使得内容损失过多。为了权衡这一过程文章设置初始 $\omega_{con-ori}$,并且设置权重变化参数 $\omega_{con-decay}$, ω_{con} 的值会根据训练次数的增加而变化,变换方式如公式(12)所示,其中epoch为训练次数。

$$\omega_{con}=\omega_{con-ori} * (\omega_{con-decay})^{(epoch/10)} \quad (12)$$

3 实验与实验结果分析

3.1 数据集介绍

训练数据集包括真实世界图片以及卡通图,真实世界图片作为内容图,卡通图片作为风格图,大小均为256*256。真实世界图片包含6659张图片,卡通图片数据集共有两个,分别来自包含2000张宫崎骏(Hayao)画风的卡通数据集,以及从电影《追逐繁星的孩子》中截取和下载的包含1860张新海诚(Shinkai)画风的卡通数据集。

3.2 实验配置

该实验的实验平台为windows10操作系统,硬件设备包括12Gen Intel(R) Core(TM)i5-12490F;显卡为NVIDIA GeForce RTX 3070,显存8G;实验环境为python3.9;pytorch1.12.1;cuda11.3,软件平台为pycharm。

3.3 实验预处理

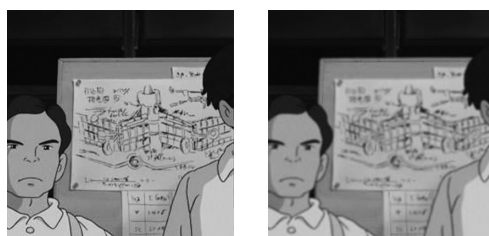
3.3.1 图片模糊处理

在训练之前需要对卡通图片进行模糊处理,首先采用Canny边缘检测,其次扩张边缘区域,最后在扩张的边缘区域使用高斯平滑得到卡通图对应的模糊卡通图。新海城风格的卡通图和对应的模糊图片如图11所示,宫崎骏风格的卡通图和对应的模糊图片如图12所示。



(a)新海诚风格卡通图 (b)对应模糊图

图11 新海诚风格的卡通图及其对应的模糊图

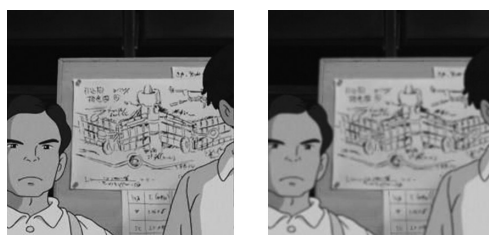


(a)宫崎骏风格卡通图 (b)对应模糊图

图12 宫崎骏风格的卡通图及其对应的模糊图

3.3.2 图片显著性处理

因为在显著性损失中需要真实世界图片的显著性参考图片,所以在训练之前需要得到其显著性参考图片。虽然鲁棒背景显著性检测(Robust Background saliency Detection, RBD)^[20]和最小障碍显著目标检测(Minimum Barrier salient object Detection, MBD)^[21]是得到显著性图片的两种常用的方法,但是如果单独的使用某一种方法可能会在一些图片的区域产生扭曲或者模糊,为了避免这一现象的发生,最终的显著性参考图片来自于这两种方法的平均值。真实世界图片(a)及其对应的显著性参考图(b)如图13所示。



(a)真实世界图片 (b)对应显著性参考图

图13 真实世界图片及其对应的显著性参考图

3.3.3 预训练

为了保证在训练过程中提高收敛性,在正式训练前进行预训练。预训练的对象是真实世界图片以及其对应的显著性参考图。在预训练中使用网络中的生成器和显著性网络来进行预训练,使用 $L_{con}(G, D)$ 来优化生成器,使用 $L_{sali}(S)$ 来优化显著性网络模块,相关参数设置为 $\omega_{con-ori} = 0.25$, $\omega_{con-decay} = 0.96$, $\omega_{sali} = 1$ 。真实世界图片(a)和其对应的经过预训练的图片(b)如图14所示,真实世界图片(a),其对应的显著性参考图(b)以及经过显著性网络预训练的图片(c)如图15所示。



(a)真实世界图片 (b)对应经过预训练图片

图14 真实世界图片及其经过预训练之后的图片



(a)真实世界图片 (b)对应显著性参考图 (c)对应显著性预训练图片

图15 真实世界图片(a)及其对应的显著性参考图(b)及经过显著性网络预训练的图片(c)

3.4 消融实验

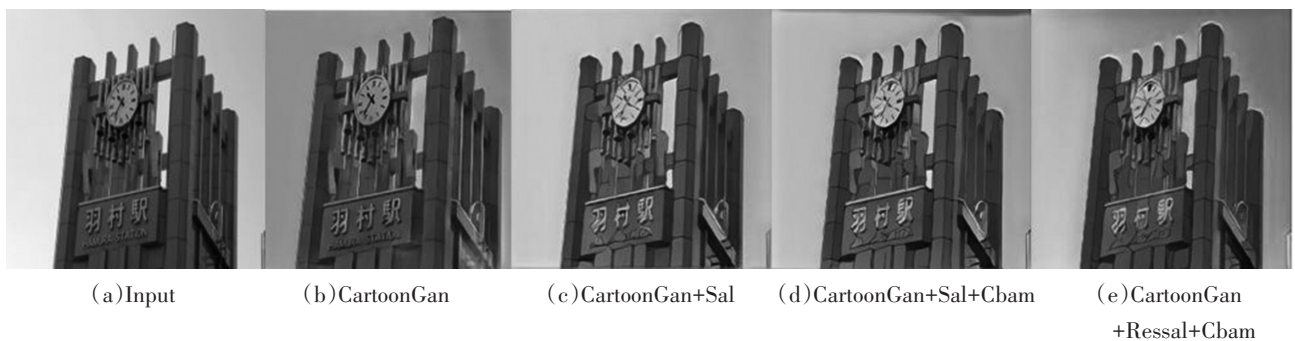
分别对两个卡通数据集上进行消融实验,卡通图如图16所示。消融实验分别是在CartoonGan基础上添加显著性网络,然后再添加Cbam,最后在显著性网络上添加残差结构。宫崎画风实验效果图如图17所示,新海城画风实验效果图如图18所示。其中Sal为显著性网络,Ressal为残差显著性网络,Cbam为注意力机制。



(a)宫崎骏画风

(b)新海诚画风

图16 卡通图



(a)Input

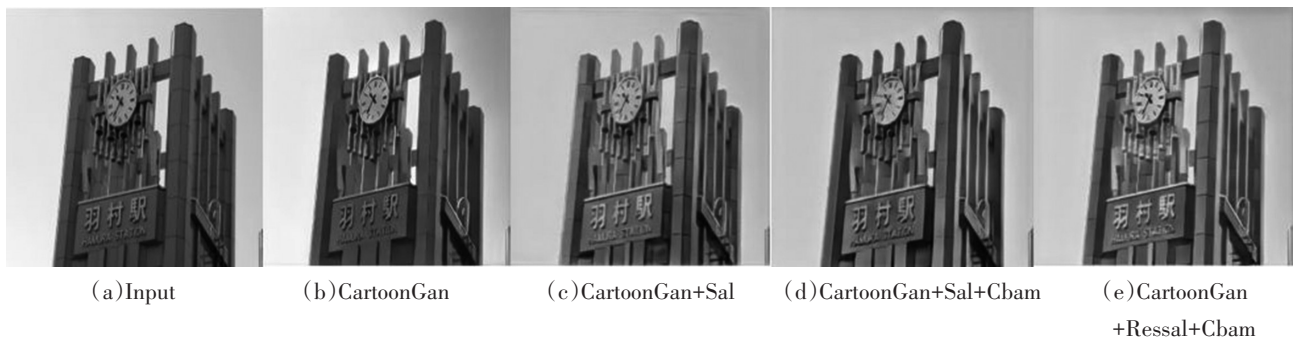
(b)CartoonGan

(c)CartoonGan+Sal

(d)CartoonGan+Sal+Cbam

(e)CartoonGan+Sal+Cbam+Ressal

图17 宫崎骏风格数据集上的消融实验效果图



(a)Input

(b)CartoonGan

(c)CartoonGan+Sal

(d)CartoonGan+Sal+Cbam

(e)CartoonGan+Sal+Cbam+Ressal

图18 新海诚风格数据集上的消融实验效果图

3.5 实验结果分析

主观分析上,通过观察消融实验的结果图可以发现,首先在语义信息略有丢失的情况下卡通效果相比原模型有所增加,这表明引入显著性损失函数和残差显著性网络可以很好地平衡内容和风格之间的关系;其次在图像边缘,轮廓上更为明显,较原始模型有了更加清晰的表示;最后图像的颜色也没出现颜色溢出或者失真现象。总体来说文章的方法可以生成语义信息较好、边缘清晰、卡通效果更好的卡通图。

客观分析上,文章采用弗雷歇感知距离(Fréchet Inception Distance, FID)测试指标,该指标是在卡通风格迁移中常用的一种测试指标,通过测试生成图像和真实图像特征分布之间的距离来衡量生成质量,值越小越好。

消融实验中生成的图片和对应卡通风格图片之间的FID结果如表1所示,其中Sal为显著性网络,Cbam为注意力机制,Ressal为残差显著性网络。和一些其他生成网络的比较结果如表2所示。图19为文章方法和其他模型在宫崎骏风格数据集上的生成效果图,图20为文章方法和其他模型在新海诚风格数据集上的生成效果图,如图19、图20所示。

表1 消融实验FID评价结果

模型	FID	
	宫崎骏风格	新海诚风格
CartoonGan	175.65	161.33
CartoonGan+Sal	169.24	155.65
CartoonGan+Sal+Cbam	165.77	152.87
CartoonGan+Ressal+Cbam	162.54	149.71

表2 相关方法的FID评价结果

模型	FID	
	宫崎骏风格	新海诚风格
White-box	171.78	156.47
AnimeGan	167.42	151.84
Ours	162.54	149.71

宫崎骏风格

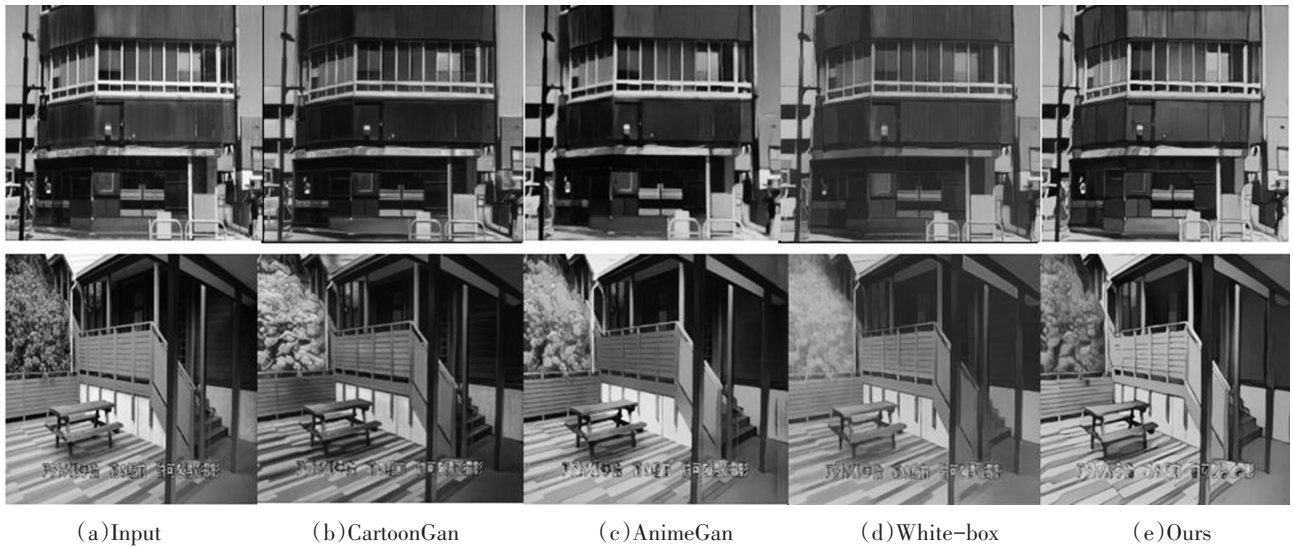


图19 文章方法和其他模型在宫崎骏风格数据集上的生成效果图

新海诚风格

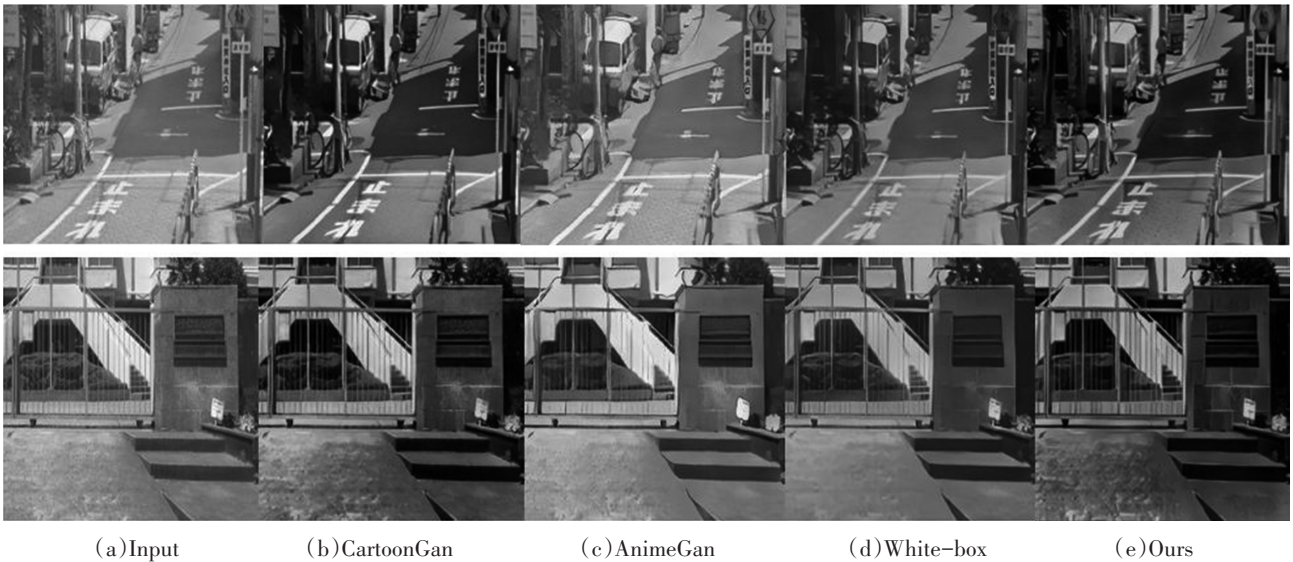


图20 文章方法和其他模型在新海诚风格数据集上的生成效果图

由表1可知,在卡通效果改进的评定中,文章采用的FID测试指标的值在消融实验中均有所降低。在最终的数据中显示,在宫崎骏和新海诚两种卡通风格上文章的模型比原始模型CartoonGan分别优化了7.4%和7.1%,而且在宫崎骏风格上要比新海诚风格上要优化得更好。由此可见较原始模型,文章模型在这两个数据集上均有优化。

由表2可知,通过FID测量指标的值可知,相比于White-box来说,文章在宫崎骏和新海诚这两个数据集上的优化分别为5.3%和4.5%,在AnimeGan上分别优化了2.9%和1.3%。因此文章模型相比一些经典的卡通图生成来说也有一定的提升。

4 结语

文章在CartoonGan基础上针对网络模型训练不稳定和梯度爆炸问题将原有的交叉熵损失修改为最小二乘损失。为了保证在卡通化的同时使得真实世界图片在内容上不会缺损太多,引入了残差显著性网络和显著性损失来平衡内容和风格之间的损失。同时在已有研究基础上引入了Cbam注意力机制进一步优化生成图像细节,以此来进一步优化卡通效果。实验结果表明,使用FID指标在宫崎骏和新海诚这两个数据集上进行测试发现均有优化,取得较好的结果,有效提升生成图像的质量,模型的性能得到了改善。

在以后的工作中,考虑到该网络依然存在参数量过多导致训练时间较长,在移动端上运行时间长等问题,未来可以考虑使用轻量化的卷积来替换原有的卷积或者采用轻量化的残差块来替换原有的残差块,并且引入相关损失函数来解决。其次,在图片卡通化的研究中,能够在减少参数量的前提下优化卡通效果且保证真实世界图片的内容信息不会损失过多是未来需要进一步的研究方向。

参考文献:

- [1] GATYS L A, ECKER A S, BETHGE M M. Image Style Transfer Using Convolutional Neural Networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2414–2423.
- [2] HUANG X, SERGE B. Arbitrary Style Transfer In Real-time with Adaptive Instance Normalization [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1501–1510.
- [3] LI C, WAND M. Combining Markov Random Fields and Convolutional Neural Networks for Images Synthesis [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2479–2486.
- [4] GOODFELLOW I, POUGETABADIE J, MIRZA M. Generative Adversarial Nets [J]. Advances in Neural Information Processing Systems, 2014, (27): 2672–2680.
- [5] LIU M Y, BREUEL T, KAUTZ J. Unsupervised Image-to-image Translation Networks [J]. Advances in Neural Information Processing Systems, 2017, (30): 700–708.
- [6] WANG T C, LIU M Y. High Resolution Image Synthesis and Semantic Manipulation with Conditional Gans [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8798–8807.
- [7] CHOI Y, CHOI M, KIM M, et al. Stargan: Unified Generative Adversarial Networks for Multi-Domain Image-to-image Translation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8789–9797.
- [8] XU W J, LONG C J, WANG R S, et al. DRB-Gan: A Dynamic Resblock Generative Adversarial Network for Artistic Style Transfer [C]// Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2021: 6383–6392.
- [9] YANG C, LAI Y K. CartoonGAN: Generative Adversarial Networks for Photo Cartoonization [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 9465–9474.
- [10] CHEN J, LIU G, CHEN X. AnimeGAN: A Novel Lightweight Gan for Photo Animation [J]. International Symposium on Intelligence Computation and Applications, 2019, (11521): 242–256.
- [11] WANG X R, YU J Z. Learning to Cartoonize Using White-box Cartoon Representations [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 8090–8099.
- [12] LI R, WANG J, LIU S C, et al. SDP-GAN: Saliency Detail Preservation Generative Adversarial Networks for High Perceptual Quality Style Transfer [J]. IEEE Transactions on Image Processing, 2021, (30): 374–385.
- [13] DONG Y S, TAN W. CartoonLossGAN: Learning Surface and Coloring of Images for Cartoonization [J]. IEEE Transactions on Image Processing, 2022, (31): 485–498.

- [14] WOO S, PARK J, KWEON S, et al. Cbam: Convolutional Block Attention Module [C]// Proceedings of the European Conference on Computer Vision, 2018: 3–19.
- [15] ITTI L, KOCH C, NIEBUR E. A Model of Saliency-based Visual Attention for Rapid Scene Analysis [J]. IEEE Transactions on Image Processing, 1998, (20): 1254–1259.
- [16] HE K M, ZHANG X Y. Deep Residual Learning for Image Recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770–778.
- [17] ZHANG F, ZHAO H H, LI Y H, et al. CBA-GAN: Cartoonization Style Transformation based on the Convolutional Attention Module [J]. Computer and Electrical Engineering, 2023, (106): 108–120.
- [18] GAO X, ZHANG Y Q, TIAN Y J. Learning to Incorporate Texture Saliency Adaptive Attention to Image Cartoonization [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2022: 2208–2232.
- [19] MAO X D, LI Q. Least Squares Generative Adversarial Networks [C]// Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2017: 2813–2821.
- [20] ZHU W J, LIANG S. Saliency Optimization from Robust Background Detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 2814–2821.
- [21] ZHANG J M, SCLAROFF S, et al. Minimum Barrier Salient Object Detection at 80 FPS [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1404–1412.
- [22] 刘冰. 基于卷积神经网络的图像风格迁移算法综述 [J]. 新疆师范大学学报(自然科学版), 2022, 41(04): 1–7.

Improved Cartoon Image Generation Method based on GartoonGan

ZHANG Wen-tian, YU Li*

(School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, Anhui, 232001, China)

Abstract: This article proposes an improved cartoon image generation network model aimed at enhancing cartoon effects while preserving semantic information as much as possible. Firstly, a saliency network was designed and in order to prevent overfitting problems and extract further features, residual structures were added to the saliency network to form a residual saliency network and concatenated onto CartoonGan to preserve semantic information; Secondly, adding a cbam attention mechanism to the former further improves the cartoonization effect; Finally, in order to prevent training instability and gradient vanishing during the training process, the least squares loss is used to replace the cross entropy loss, and the significance loss is introduced to constrain the training of the significance network. The experiment showed that through testing on cartoon datasets of Miyazaki Hayao and Makoto Shinkai, the use of FID testing indicators showed some optimization on both datasets.

Keywords: Cbam attention mechanism; Saliency network; Residual structure; CartoonGan