

# 论文推荐: CartoonGAN

魏张鉴

21821100

计算机学院

975777376@qq.com

## 摘要

本文对发表在CVPR2018的作者为Yang Chen、Yu-Kun Lai、Yong-Jin Liu的论文《CartoonGAN: Generative Adversarial Networks for Photo Cartoonization》进行了论文推荐。

全文分为被推荐论文简介、推荐理由、自行测试以及总结和展望四大部分。

被推荐论文简介中作者简单介绍了该论文的目的、创新的研究方法、相比现有方法的优越性;推荐理由部分,分为论文特色、创新性求解思路、重要实验结论三个部分,讲述了为何选取并推荐了这篇论文;自行测试部分使用作者提供的已训练神经网络,验证了风格化效果;最后在总结与展望部分总结全文,并分析了未来发展趋势。

## 1. 被推荐论文简介

这篇论文提出了一种将真实世界场景照片转化为卡通风格图像的解决方案,这在计算机视觉和计算机图形中具有一定的价值和挑战性[1]。

众所周知,以艺术方式对图像进行风格化在非真实感渲染领域已经被广泛研究了[2]。传统方法只能为特定样式开发专用算法。然而,需要付出巨大努力来制作模仿特定个体艺术家的细分风格。近年来,基于学习的风格转移方法已经引起了相当大的注意,

本篇的解决方案恰恰是基于机器学习的方法,这种方法最近在绘画等艺术形式的图像风格化很受欢迎。然而,现有的方法并不能产生令人满意的卡通化效果,因为下列原因:

- (1) 卡通风格具有独特的特点,伴随着高度的简化和抽象;
- (2) 卡通图像往往有清晰的边缘,平滑的着色和相对简单的纹理,这对现有方法中使用的基于纹理描述的损失函数提出了重大挑战。

因此作者提出了CartoonGAN,一个用来卡通风格化的生成对抗网络(GAN)框架。

这篇论文中采用了未配对的照片和卡通图像进行训练。有两种适合卡通化的新颖的损失评估方式被作者

提出了:

(1) 语义内容损失,在VGG网络的高级特征图中表现为稀疏正则化,以应对照片和卡通之间的实质性风格变化;

(2) 边缘促进对抗性损失,可以用来保持清晰的边缘。

此外,作者还引入了一个不错的初始化方法,以改善网络到目标形态的收敛。

他们的方法也比现有方法能更高效地训练。实验结果表明,该方法能够从真实世界的照片中生成高质量的卡通图像(即遵循特定艺术家的风格,边缘清晰,着色平滑),并优于现存的最先进的方法。

## 2. 推荐理由

之所以我在浩如烟海的CVPR 2018、ECCV 2018、CVPR 2019论文集中选中了该篇论文,自然是有相应的理由的。

### 2.1. 论文特色

文章整体通俗易懂,遣词造句平易近人,对数学符号的解释说明也是不厌其烦,娓娓道来。

除了文本方面易于理解,该篇论文还做到了图文并茂,直观地显示了神经网络结构,以及他们的结果与现有方法对比,体现出了CartoonGAN的优越性。

最重要的是,该论文有多处创新点,具体的创新点将会在2.2和2.3部分进一步介绍。

此外,此文选题、选材新颖,图像风格化是近年来流行的机器视觉方向,训练、测试数据来源于知名动画,趣味性高。

### 2.2. 创新性求解思路

#### 2.2.1 总损失函数

该论文在评估损失时,结合了两两种新颖的损失函数,尤其是边缘促进对抗性损失的求解方式非常巧妙。

CartoonGAN的总损失函数 $L(G, D)$ 由两部分组成: (1) 对抗性损失 $L_{adv}(G, D)$ ,它驱动生成器网

络以实现所期望的图集变换，(2) 内容丢失 $\mathcal{L}_{con}(G, D)$  (第 3.2.2 节)，它在卡通风格化期间保留图像内容。该论文使用一种简单的加法形式来表示这个损失函数：

$$\mathcal{L}(G, D) = \mathcal{L}_{adv}(G, D) + \omega \mathcal{L}_{con}(G, D),$$

其中 $\omega$ 是平衡两个给定损失的权重。较大的 $\omega$ 导致来自输入照片的更多内容信息被保留，因此导致产生具有更细节纹理的风格化图像。在我们所有的实验中，我们设置 $\omega = 10$ ，这样可以实现风格和内容保持的良好平衡。

看似只使用了简单的加权相加，但巧妙地将两个损失相结合，用以训练GAN。看似轻描淡写地取 $\omega = 10$ ，也许背后包含了作者大量试错心血地结晶。

## 2.2.2 对抗性损失函数

而至于对抗性损失 $\mathcal{L}_{adv}(G, D)$ 的求解，就离不开对卡通图像的边缘做一些特殊处理了。

因为边缘像素占整张图片像素比例非常小，但卡通画的一大特征就是拥有清晰可见的边缘，所以必须另辟蹊径了。

该论文应用了下列三个步骤：(1)使用标准的Canny边缘检测器[2]检测边缘像素，(2)扩展边缘区域，以及(3)在扩展的边缘区域中应用一个高斯平滑。



图 1. 对卡通图片进行特殊的边缘平滑处理

图 1 显示了一个一张卡通图片（左侧）和一张边缘平滑后的修改版（右侧）的例子。对于每张在照片图集 $P$ 中的照片 $p_k$ ，生成器 $G$ 输出一张生成得到的图片 $G(p_k)$ 。在CartoonGAN中，训练鉴别器 $D$ 的目标是最大化给予 $G(p_k)$ ， $G(p_k)$ 就是这些没有清晰边缘的卡通图片（即 $c_i \in S_{data(c)}$ ），和真实的卡通图片（即 $c_i \in S_{data(c)}$ ）正确的标签的可能性，这样一来生成器 $G$ 可以被正确引导地转换输入内容到正确的目标图集。因此，作者定义边缘强化对抗损失为：

$$\begin{aligned} \mathcal{L}_{adv}(G, D) = & \mathbb{E}_{c_i \sim S_{data(c)}} [\log D(c_i)] \\ & + \mathbb{E}_{e_j \sim S_{data(e)}} [\log(1 - D(e_j))] \\ & + \mathbb{E}_{p_k \sim S_{data(p)}} [\log(1 - D(G(p_k)))]. \end{aligned} \quad (3)$$

## 2.2.3 内容损失函数

至于内容损失评估方面，该论文采用了VGG网络来提取高级特征，将内容损失定义为：

$$\mathcal{L}_{con}(G, D) = \mathbb{E}_{p_i \sim S_{data(p)}} [\|VGG_l(G(p_i)) - VGG_l(p_i)\|_1] \quad (4)$$

(2) 即使使用输入照片和生成的卡通图像之间的VGG特征图的 $\ell_1$ 稀疏正则化来定义语义内容损失。据作者所说， $\ell_1$ 稀疏正则化能够比标准的 $\ell_2$ 标准化更好地应对代表性和区域特征发生巨大变化的风格差异。

说到VGG网络，VGG模型是2014年ILSVRC竞赛的第二名，第一名是GoogLeNet。但是VGG模型在多个迁移学习任务中的表现要优于googLeNet。而且，从图像中提取CNN特征，VGG模型是首选算法。因此，在CartoonGAN中作者选用VGG提取特征之间的距离来评估内容相似度，也是中规中矩的选择。

## 2.2.4 初始化阶段

众所周知，对神经网络进行合理的初始化，避免陷入局部最优非常关键，CartoonGAN也不例外。

为了帮助改善其收敛性，作者们提出了一个新的初始化方法。生成器网络 $G$ 的目标是在保持语义内容的同时以卡通风格重建输入照片。他们使用生成器开始对抗性学习框架，该生成器仅重建输入图像的内容。在初始化阶段，他们仅使用语义内容损失 $\mathcal{L}_{con}(G, D)$ 来预训练发生器网络 $G$ 。

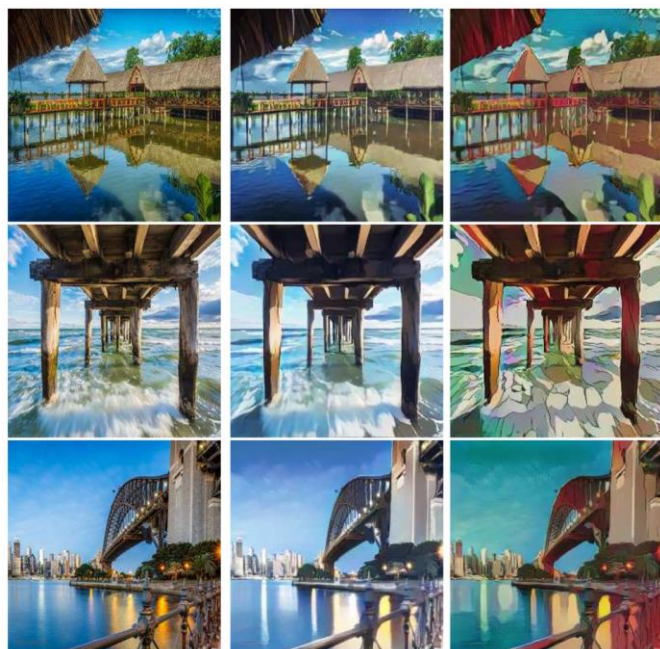
论文中的实验结果表明，这个简单的初始化阶段有助于CartoonGAN快速收敛到良好的配置，而不会过早收敛。

## 2.3. 重要实验结论

作者通过大量的测试数据，得出了CartoonGAN的方法，训练效率更高，产生的风格化图像效果更好，无论是语义内容保留还是卡通风格层面。

如图 2 所示，CartoonGAN可以顺利生成不同艺术家风格的卡通画作品，并且风格吻合度极高。

如图 3 所示，CartoonGAN相比现有的NST、CycleGAN、Gatys等方法，在不同方面均有优越的表现。



(a)输入 (b) Shinkai风格 (c)Hayao风格  
图 2。CartoonGAN生成的不同艺术风格的结果

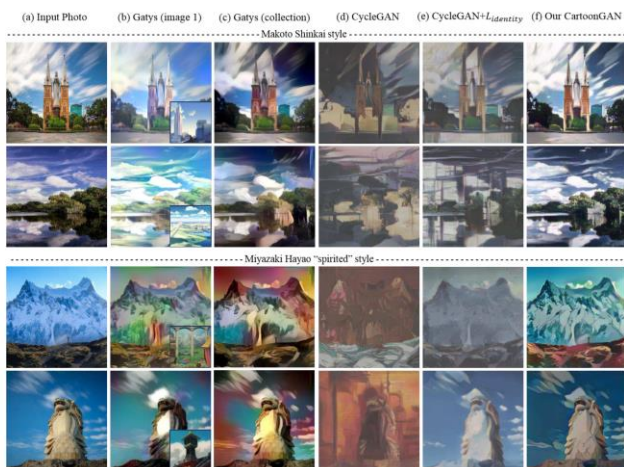


图 3。比较CartoonGAN和NST和CycleGAN对于新海诚（顶端）和宫崎骏（底端）风格的效果。Gatys image 1) and Gatys (collection) 是NST两次调整后的结果，分别是卡通图片拥有与输入照片接近的内容和所有卡通图片都被用于训练的两种情况。

此外，本文还通过切除实验研究了CartoonGAN各部分的作用。下图显示了损失函数的切除示例，其中所有结果都由Makoto Shinkai风格的数据训练。以下结果表明每个部分都在CartoonGAN中起着重要作用。首先，初始化阶段帮助生成器G快速收敛到合理的形态。如图b所示，在没有初始化的情况下，尽管显示出了一些关键特征，但是风格远非期望的。其次，即使使用合适的VGG层，由于大量的风格差异，输入的特征图中

卡通风格图像巨大且通常是局部的差异仍然是需要的。使用高级VGG特征图的 $\ell_1$ 稀疏正则化（而不是 $\ell_2$ ）有助于应对卡通图像和照片之间的实质风格差异。最后，精心设计的边缘损失引导生成器G在结果中产生清晰的边缘，从而产生更好的卡通风格图像。

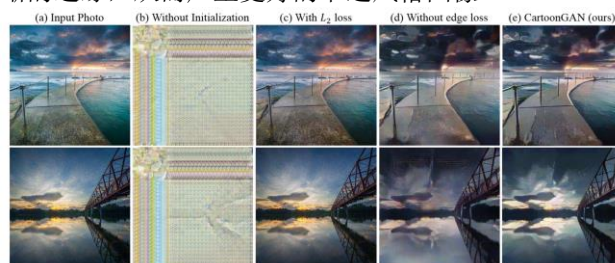


图 4。移除/改变CartoonGAN的损失函数的各部分导致的结果

### 3. 自行测试

使用作者训练好的神经网络，我利用Pytorch+Cuda进行了风格化效果测试，根据输入的真实世界照片，我得到了符合论文描述的卡通风格化图像输出结果，如下图所示。



图 5。自行测试CartoonGAN，由输入照片得到了卡通风格化照片

可以看出风格迁移效果还是可圈可点的，的确如作者描述所述，产生了清晰的边缘，平滑的着色效果。

### 4. 总结和展望

卡通风格迁移领域非常有前景，相信不久的将来，会助力于艺术创作，有较大的商业价值。

未来研究神经风格迁移，有前景的方向主要集中在两个方面。一是解决前面提到的当前算法所面临的难题[3]，即参数微调问题、笔触方向控制问题和神经风格迁移中“快”与“更快”的问题。第二个有前景的方向是

关注神经风格迁移的新扩展（比如时尚风格迁移，字符风格迁移，还有本推荐论文的卡通风格迁移）。

除了卡通风格化以外，类似的GAN技术还能应用到其他领域，并且对相应的参数，损失函数，初始化方法做出更精妙的调整，值得后继研究者深入挖掘。

## 参考文献

- [1]Yang Chen,Yu-Kun Lai,Yong-Jin Liu, 《CartoonGAN: Generative Adversarial Networks for Photo Cartoonization》,CVPR2018
- [2]P. L. Rosin and J. Collomosse. *Image and Video-Based Artistic Stylisation*. Springer, 2013.
- [3]Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, Mingli Song, *Neural Style Transfer: A Review*, arXiv:1705.04058