

# 计算机视觉课程报告

李晨帅

21821067

计算机学院

21821067@zju.edu.cn

## 摘要

本文介绍了CVPR 2018的文章CartoonGAN: Generative Adversarial Networks for Photo Cartoonization, 对其算法进行了简要说明, 并给出了自己推荐这篇文章的理由及其特色。

### 1. 简介

这篇文章提出了一种解决方案, 将现实世界场景的照片转换为卡通风格的图像, 这在计算机视觉和计算机图形学中具有重要价值和挑战性。这篇文章的解决方案属于基于学习的方法, 最近流行的方法是以艺术形式(如绘画)对图像进行风格化。然而, 由于(1)卡通风格具有高度简化和抽象的独特特征, 并且(2)卡通图像倾向于具有清晰的边缘、平滑的色彩阴影和相对简单的纹理, 因此现有方法不能产生令人满意的卡通化结果, 这显示出现有方法中使用的基于纹理描述符的损失函数有着显著挑战。这篇文章提出了CartoonGAN: 一种用于卡通风格化的生成对抗网络(GAN)框架。这篇文章的方法采用不成对的照片和卡通图像进行训练, 易于使用。提出了两种适合卡通化的新颖损失: (1)语义内容损失, 在VGG网络的高级特征映射中表现为稀疏正则化, 以应对照片和卡通之间的实质风格差异, 以及(2)边缘促进的对抗性损失, 用于保留清晰的边缘。这篇文章进一步引入初始化阶段, 以改善网络到目标流形的收敛。这篇文章的方法比现有方法更有效。实验结果表明, 这篇文章的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, 具有清晰的边缘和平滑的阴影), 并且优于现有方法。

#### 1.1. 应用意义

卡通是我们日常生活中广泛使用的艺术形式。除了艺术爱好, 它们在印刷媒体的出版到儿童教育的绘本中都有应用。像其他形式的艺术品一样, 许多著名的卡通图像都是基于现实世界的场景而创作的。图1展示了一个真实世界的场景, 其相应的卡通图像出现在动画电影《你的名字。》中。然而, 手动重现卡通风格的现实场

景非常费力, 并且涉及大量的艺术技巧。为了获得高质量的卡通, 艺术家必须绘制每一条线并遮蔽目标场景的每个颜色区域。同时, 具有标准特征的现有图像编辑软件/算法不能产生令人满意的卡通化结果。因此, 可以自动将真实世界的照片转换为高质量的卡通风格图像的专门的技术将非常有用, 对于艺术家来说, 可以节省大量的时间, 以便他们可以专注于更有创意的工作。这些工具还为Instagram和Photoshop等照片编辑软件提供了有用的补充。



图1: 卡通风格化的一个例子。左图: 一个真实世界的场景, 其相应的卡通图像出现在动画电影《你的名字。》中。右图: 这篇文章将左图转换为卡通风格的结果。注意, 训练集中不包含《你的名字。》中的任何图片。

### 2. CartoonGAN介绍

GAN框架由两个CNN组成。一个是生成器G, 它被训练产生愚弄判别器的输出。另一个是判别器D, 其用来判别图像是来自真实目标流形还是合成。这篇文章设计的生成器和判别器网络能适应卡通图像的特殊性, 网络详见图2。

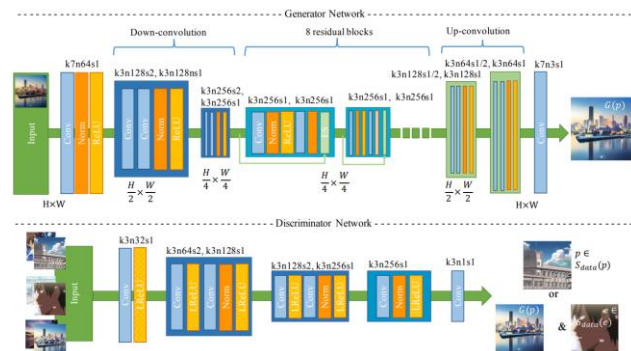


图2: 提出的CartoonGAN中的生成器和判别器网络的体系结构, 其中k是核大小, n是特征映射的数量, s是每个卷积层中的步幅, norm表示归一化层, ES表示元素化之和。

这篇文章制定了将真实世界照片转换为卡通图像的过程，将其作为将照片流形 $\mathcal{P}$ 映射到卡通流形 $\mathcal{C}$ 的映射函数。使用训练数据  $S_{data}(p) = \{p_i | i = 1 \dots N\} \subset \mathcal{P}$  和  $S_{data}(c) = \{c_i | i = 1 \dots M\} \subset \mathcal{C}$  来学习映射函数，其中 $N$ 和 $M$ 分别是训练集中的照片和卡通图像的数量。与其他GAN框架一样，通过区分卡通流形中的图像与其他图像并为 $G$ 提供对抗性损失来训练判别器函数 $D$ 以推动 $G$ 达到其目标。设 $\mathcal{L}$ 为损失函数， $G^*$ 和 $D^*$ 为网络的权重。目标则是解决最小-最大问题：

$$(G^*, D^*) = \arg \min_G \max_D \mathcal{L}(G, D)$$

损失函数定义请见第 2.1 节。

### 2.1. 损失函数

损失函数 $\mathcal{L}(G, D)$ 由两部分组成：（1）对抗性损失 $\mathcal{L}_{adv}(G, D)$ （第 2.2 节），它驱动生成网络以实现所需的流形变换，和（2）内容损失 $\mathcal{L}_{con}(G, D)$ （第 2.3 节），它在卡通风格化期间保留图像内容。这篇文章使用了一种简单的加法形式来表示损失函数：

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

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，因此导致具有更详细纹理的风格化图像。这篇文章设置 $\omega = 10$ ，这样可以实现风格和内容的良好平衡。

### 2.2. 对抗性损失

这篇文章观察到，简单地训练判别器 $D$ 以分离生成的和真实的卡通图像不足以将照片转换成卡通。这是因为清晰边缘的呈现是卡通图像的重要特征，但是这些边缘的比例在整个图像中通常非常小。因此，没有清晰再现边缘但具有正确渲染的输出图像可能使训练的判别器与标准损失混淆。

为了避免这个问题，这篇文章从训练卡通图像 $S_{data}(c)$ 中去除清晰边缘自动生成一组图像 $S_{data}(e)$ ，其中 $S_{data}(c)$ 和 $S_{data}(e)$ 分别是卡通流形和没有清晰边缘的卡通式图像的流形。图 3 展示了卡通图像和具有平滑边缘的修改版本的例子。



图 3：通过去除卡通图像中的清晰边缘，这篇文章生成相应的图像。

对抗性损失定义如下，详情可查阅文献[1]。

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

其中， $G(p)$ 表示对于照片 $p$ ，其输出的卡通化结果。而 $D(c)$ 则返回对于图片 $c$ ，其是卡通图像的概率。

### 2.3. 内容损失

内容损失定义如下，详情可查阅文献[1]。

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

与其他图像生成方法不同，这篇文章使用了输入照片和生成的卡通图像之间的VGG特征图的 $L_1$ 范数来定义语义内容丢失。这是因为卡通图像具有与照片非常不同的特征（即：清晰的边缘和平滑的阴影）。 $L_1$ 范数能够比标准的 $L_2$ 范数更好地应对这种特征。稍后将要展示，这对于重现卡通风格至关重要。

### 2.4. 初始化阶段

由于GAN模型是高度非线性的，随机初始化的话，优化可以很容易地陷入次优的局部最小值。为了帮助改善其收敛性，这篇文章提出了一个新的初始化阶段。注意，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。这篇文章使用生成器启动对抗性学习框架，该生成器仅重建输入图像的内容。为此，在初始化阶段，仅使用语义内容损失 $\mathcal{L}_{con}(G, D)$ 预训练生成器网络 $G$ 。图 4 展示了在该初始化训练阶段的 10 个时期之后的重建图像的示例，其已经产生了合理的重建。这篇文章的实验结果表明，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。

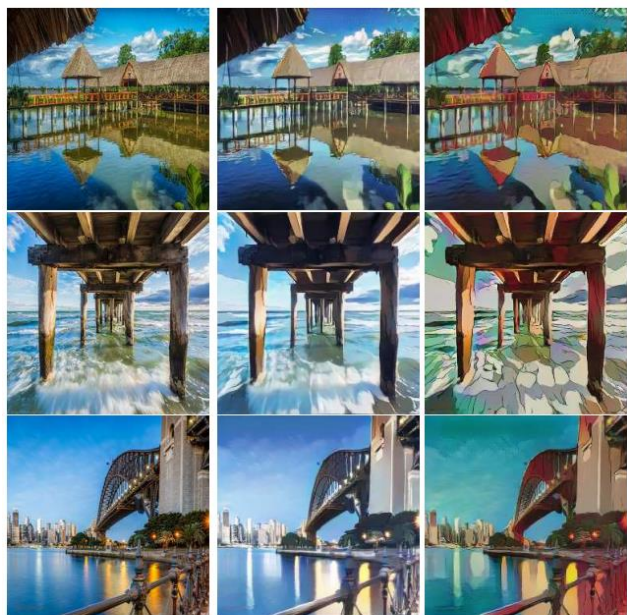


图 4：对于原始照片（左），右图是初始化阶段之后的结果。

## 3. 实验效果

CartoonGAN能够使用个别艺术家的数据进行高质量的卡通风格化，这些数据很容易从卡通视频中获得，因为这一方法不需要配对图像。不同的艺术家有他们独特的卡通风格，这些风格可以通过CartoonGAN有效地学习。由CartoonGAN生成的一些不同艺术风格的结果如图 5 所示。





(a) input photo (b) Shinkai style (c) Hayao style  
图 5: 一些由CartoonGAN产生的不同艺术风格的结果。(a) 输入的真实世界照片。(b) 新海诚风格。(c) 宫崎骏风格。

### 3.1. 和现有方法的比较

和现有方法的比较请见图 6, 详情可查阅文献[1]。

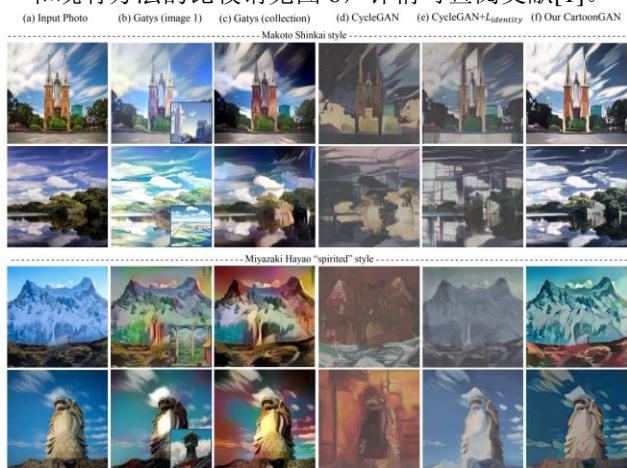


图 6 和现有方法的比较

### 3.2. 损失函数各部分的作用

损失函数各部分的作用请见图 7, 详情可查阅文献[1]。

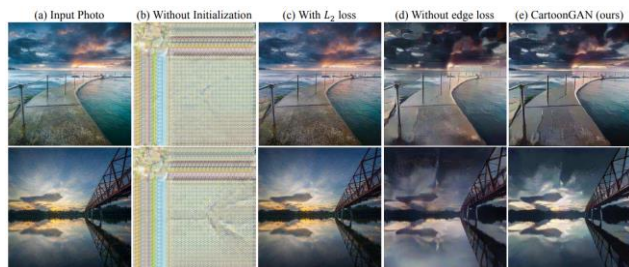


图 7 各个部分的作用

## 4. 个人评价及推荐原因

这篇文章所做的内容属于风格迁移, 然而, 目前的风格迁移对“卡通化”的效果不是很好, 这篇文章分析了其原因, 并提出了CartoonGAN, 以实现卡通化效果的风格迁移。

这篇文章创新点在于: (1) 使用了边缘损失, 将卡通图像进行模糊处理后加入了训练集, 使得边缘模糊的图片尽可能不出现; (2) 使用VGG配合了 $L_1$ 范数, 这样使得生成的图片更接近卡通化而非真实照片(图 7 可以看出使用 $L_2$ 范数会使得生成图片更接近真实照片); (3) 使用了初始化阶段, 使得生成的图片初始化后就能保证原真实场景的内容。

这篇文章提供了卡通化的工具, 可以用于各种图像处理软件, 而且由于其可以学习不同艺术风格, 因此其实用性较强, 这也是我推荐这篇文章的原因之一。

这篇文章的网络比起其他通用型的风格迁移网络而言, 更为专一, 因此更为有效。为了效果而放弃通用性, 我认为是值得的。

目前, 该算法对于风景的效果较好, 文中所给的例子也均为自然风景图。然而, 对于人脸, 该算法未必就能得到理想的结果, 这也是未来值得改进的地方。此外, 目前还没有对视频的卡通化处理, 这也是未来值得改进的地方。

## 5. 结论

CartoonGAN是一种有效的卡通化风格迁移网络, 其实现新颖、效果较好且能够有着一定的应用范围。

## 参考文献

- [1] Yang Chen, Yu-Kun Lai, Yong-Jin Liu; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 9465-9474