

# 一种改进的生成对抗网络的图像上色方法研究

蒋文杰, 罗晓曙\*, 戴沁璇

(广西师范大学 电子工程学院, 广西 桂林 541004)

**摘要:**利用手持云台对黑白照片/图像进行拍摄再对其进行上色是耗时耗力的工作,为了提高对黑白照片/图像上色的效率和视觉效果,利用深度学习的卷积神经网络提出了一种基于改进生成对抗网络的上色算法。采用了原模型 pix2pix (image-to-image translation with conditional generative adversarial networks) 的 U 型结构的生成器并在其中引入了自注意力机制提高输出图像的彩色多样性;其次是使用实例归一化对网络结构作进一步的优化处理,生成器和判别器在网络训练期间不断进行相互对抗学习,同时模型不断学习并优化黑白照片/图像到对应彩色图像的映射关系;最后实现了对黑白照片/图像的自动化上色,同时使用客观量化指标(MSE、PSNR 和 SSIM)对实验结果进行评价。实验结果表明:该算法能快速有效实现对黑白照片/图像的无监督上色,同时保持良好的视觉效果。

**关键词:**黑白照片/图像;上色;卷积神经网络;无监督学习;生成对抗网络

中图分类号:TP183

文献标识码:A

文章编号:1673-629X(2020)07-0056-04

doi:10.3969/j.issn.1673-629X.2020.07.013

## Research on an Improved Method of Generative Adversarial Networks Image Coloring

JIANG Wen-jie, LUO Xiao-shu\*, DAI Qin-xuan

(School of Electronic Engineering, Guangxi Normal University, Guilin 541004, China)

**Abstract:** It is time-consuming and labor-intensive to take a black and white photo/image and then paint it with a hand holder. In order to improve the efficiency and visual effect of black-and-white photo/image, we propose a coloring algorithm based on improved generative adversarial networks with deep learning convolutional neural network. The U-shaped structure of the original model pix2pix (image-to-image translation with conditional adversarial networks) is adopted and a self-attention mechanism is introduced to improve the color diversity of the output image. Followed by the use of instance normalization to further optimize the network structure, the generator and discriminator continue to compete against each other during network training, while the model continuously learns and optimizes the mapping of black and white photo/image to corresponding color images. Finally the automatic coloring of black and white photo/image is realized, and the experimental results using objective quantitative indicators (MSE, PSNR and SSIM) are evaluated, which show that the proposed algorithm can quickly and effectively achieve the unsupervised coloring of black and white photo/image while maintaining better visuals.

**Key words:** black and white photo/image; coloring; convolutional neural network; unsupervised learning; generative adversarial network (GAN)

## 0 引言

近年来,深度学习在图像处理领域表现出了巨大的优越性,目前手工对其进行上色将花费巨大的时间和人力成本,所以还无法做到对黑白照片/图像的批量上色处理。如果能够借助深度学习的方法将上色这一过程无监督化,将在效率和时间上给黑白照片/图片带来质的提升,提高其艺术观赏价值。

文献[1-2]提出了使用人工干预和灰度直方图对灰度图进行上色,效率不高,很难进行大规模推广应用。文献[3-5]提出一种艺术风格转化的方法,这种方法是通过卷积神经网络分别提取两张图片的特征,然后再将得到的特征结合起来组合成新的图片,并没有实现真正的上色功能。在文献[6]中针对高分辨率图片上色的问题,突破了难以生成高分辨率图像的限

收稿日期:2019-08-08

修回日期:2019-12-11

基金项目:广西科技重大专项(桂科 AA18118004)

作者简介:蒋文杰(1992-),男,硕士生,研究方向为深度学习;通讯作者:罗晓曙(1961-),男,博士,硕导,教授,研究方向为非线性电路理论及非线性控制系统、人工智能理论与应用。

制,虽然实现了对高分辨率图片上色,但也增加了训练难度同时对计算机硬件提出了更高的要求。

生成对抗网络作为深度学习中的一种概率生成模型,最早由 Ian Goodfellow 等人<sup>[7]</sup>于2014年提出,它能够自动学习图片的输入到输出的映射,不需要构造复杂的目标函数,构建了泛化模型。GAN由两个网络构成,分别是生成器 $G$ ,其作用是抓取数据分布特征,另外一个判别器 $D$ ,其作用是估计图片来自训练数据还是来自于生成器 $G$ 生成的图片。两者进行极大极小值博弈<sup>[7]</sup>最终达到动态平衡,具体表达式如下所示:

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

对黑白照片/图像进行上色需要考虑两个因素:输出的彩色图像与原来的黑白照片/图片的底层轮廓信息不变,另外就是上色的合理性。文中在兼顾底层信息不变和上色合理性的前提下,构建了基于改进生成对抗网络的黑白照片/图片无监督上色模型。该模型包含了带有U型<sup>[8]</sup>结构的生成器模型和以块的方式(PatchGAN)<sup>[9]</sup>的判别器模型。该模型是以黑白照片/图像作为生成器模型的输入,然后生成器输出彩色图像,接下来通过判别器判断该输出彩色图是来自真实分布还是生成器输出图像,生成器和判别器两者以对抗的方式相互进行训练,不断优化黑白图像和彩色图像的映射关系,最后得到稳定的网络模型。由于黑白照片/图像到彩色照片的数据集较少,文中所用的数据集是从各大网站抓取并进行加工处理后得到,使用tensorflow在GPU上对模型进行训练。

## 1 研究方法

利用GAN对黑白照片/图像样本集进行无监督上色训练,最终输出彩色图像。训练好的GAN模型虽然能够将输入的黑白图像转化为彩色图像,但是却无法保持上色的合理性。为了解决这个问题,这里引入了条件对抗网络CGAN(conditional generative adversarial nets)<sup>[10]</sup>。该网络是在GAN的生成器和判别器都额外添加一个条件 $y$ ,即生成器必须要生成和条件 $y$ 匹配的样本,判别器不仅要判别图像是否真实,还要判别图像和条件 $y$ 是否匹配。CGAN的基本结构如图1所示。

生成器 $G$ ,输入为一个噪声 $z$ ,输出一个图像 $G(z)$ ,判别器 $D$ ,输入为一个图像 $x$ ,输出该图像为真实的概率 $D(x)$ ,CGAN的损失函数可以定义为:

$$L_{\text{CGAN}}(G, D) = E_{x, y} [\log D(x, y)] + E_{z, y} [\log(1 - D(x, G(z, y)))] \quad (2)$$

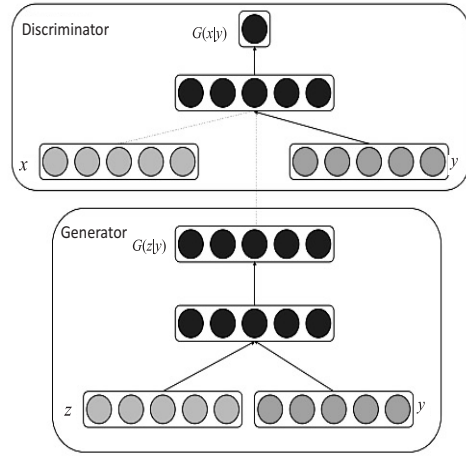


图1 CGAN基本结构

文中算法在CGAN基础上进行改进,原始的CGAN为了能够加快训练网络时的收敛速度以及归纳统一样本的统计分布性,采用批量归一化BN(batch normalization)<sup>[11]</sup>,它能够将数据强行归一化为均值为0、方差为1的正态分布上,这样有利于数据分布一致,也避免了梯度消失。BN的缺点也很明显,首先,它对样本的数量的大小比较敏感,因为每次计算的方差和均值都在一个批量上,所以如果批量太小,这会导致方差和均值不足以代表整个数据分布。其次BN只对固定深度的前向神经网络很方便。

基于此,文中采用了实例归一化IN(instance normalization)<sup>[12]</sup>替换BN,这是因为在无监督上色中,输出图像主要依赖于某个图像实例,IN是对一个批次中单张图片进行归一化,而不是像BN对整批图片进行归一化然后提取平均值,所以IN更适合于无监督上色,提高输出图片上色的合理性和更好地保留底层信息,IN的计算公式如下:

$$y_{ijk} = \frac{x_{ijk} - \mu_{ti}}{\sqrt{\sigma_{ti}^2 + \epsilon}} \quad (3)$$

$$\mu_{ti} = \frac{1}{HW} \sum_{l=1}^W \sum_{m=1}^H x_{ilm} \quad (4)$$

$$\sigma_{ti}^2 = \frac{1}{HW} \sum_{l=1}^W \sum_{m=1}^H (x_{ilm} - m\mu_{ti})^2 \quad (5)$$

其次,该算法还在生成器中引入了自注意力Self-Attention<sup>[13]</sup>机制,传统的生成对抗网络的问题表现在卷积核的大小选取上,小的卷积核很难发现图像中的依赖关系,但是大的卷积核又会降低卷积网络的计算效率。

为了提升上色效果,该算法将Self-Attention加载到生成器网络中,具体网络如图2所示,即在前一层的特征图上加入Self-Attention机制,使得生成对抗网络在生成时能区分不同的特征图。

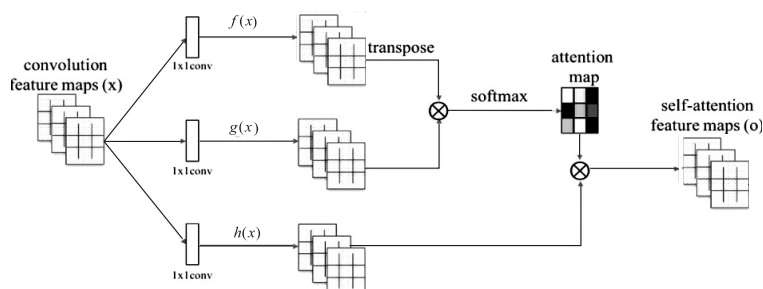


图 2 Self-Attention 机制结构

图中  $f(x) = W_f x$ ,  $g(x) = W_g x$ ,  $W_f$ 、 $W_g$  都是网络的参数, Self-Attention 机制的注意力计算如下:

$$\beta_{j,i} = \frac{\exp(s_{ij})}{\sum_{i=1}^n \exp(s_{ij})}, s_{ij} = f(x_i)^T g(x_j) \quad (6)$$

注意力层的输出如下所示:

$$o_j = \sum_{i=1}^n \beta_{j,i} h(x_i), h(x_i) = W_h x_i \quad (7)$$

最后得到加入了 Self-Attention 机制的特征图:

$$y_i = \gamma o_i + x_i \quad (8)$$

这里  $\gamma$  的值初始化为 0, 再逐渐增大权重, 由于 Self-Attention 一开始可能训练得不太好, 用它作指引效果不好, 所以等 Self-Attention 层训练得越来越好后再加大  $\gamma$  的值。

## 2 网络结构

在对图像彩色化的问题上, 为了保持原图像的底层轮廓信息不变和上色的合理性, 一般是采用 U-Net 结构, 如图 3 所示, 其好处是不需要所有的信息都通过生成器的所有层, 这样就能使得输入和输出图像的底层信息和突出边缘位置信息能够被共享。U-Net 型网络对于提升输出图像的细节起到了良好的效果。

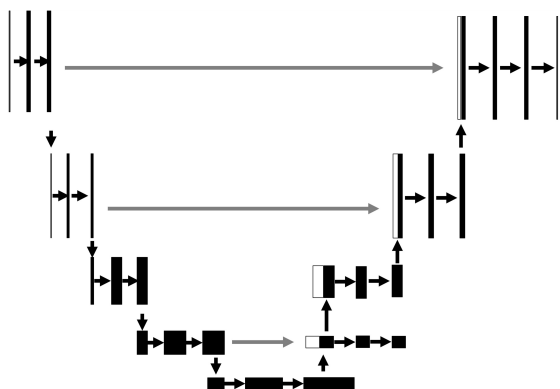


图 3 生成器模型结构

对于判别器  $D$ , 文中采用 PatchGAN 的方法, 即对每张输出图像的每个  $N \times N$  小块 (Patch) 计算概率, 然后再将这些概率求平均值作为整体的输出, 其优点在于可以加快模型收敛和计算的速度。判别器模型结构如图 4 所示。

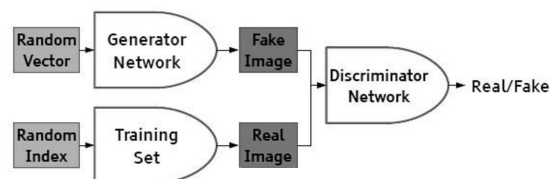


图 4 判别器网络结构

## 3 损失函数

文中算法的损失函数基于 CGAN 进行改进, 和 CGAN 相比少了噪声输入, 文中定义了和 CGAN 相同的损失, 另外为了加快模型的收敛速度并提高图像的清晰度, 在 CGAN 损失函数的基础上还添加了 L1 损失:

$$L_{LI(G)} = E_{x,y,z} [\|y - G(x,z)\|_1] \quad (9)$$

最终得到的目标函数为:

$$G = \arg \min(G) \max(D) L_{CGAN}(G, D) + \lambda L_{LI}(G) \quad (10)$$

## 4 实验

### 4.1 实验平台

由于在模型训练中需要大量的卷积计算, 相对于 CPU 而言, GPU 的运用将极大地缩短训练时间, 加快训练速度。本实验中所使用的计算机配置是双 E5-2637 v4 CPU, 操作系统为 Ubuntu 16.04, 同时还使用了 GTX1080Ti 显卡、12 GB 内存来加快训练。所使用的平台是谷歌开发的机器学习框架 tensorflow1.9。

### 4.2 实验效果

该算法所用到的图像数据集是使用爬虫从各大网站上随机抓取的, 经过筛选并剔除不适用于训练的图像, 最后剩下大约 2 万张的数据图像, 最终典型的实验效果如图 5 所示。



图 5 上色方法效果比较

### 4.3 实验对比

为了进一步验证该算法的有效性, 这里进行了简

化模型测试,使用控制变量法单独分出了 CGAN 模型,分别使用 IN 项和 GN(group normalization)<sup>[14]</sup>项代替原模型的 BN 项,最终的实验效果如图 6 所示。

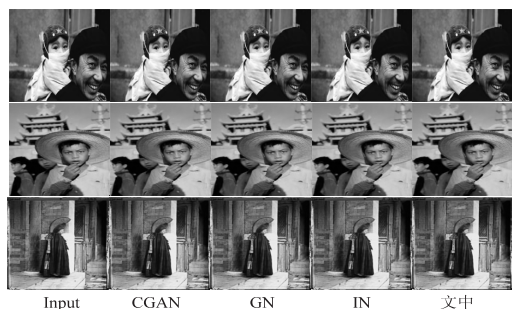


图 6 不同算法效果比较

从上面对比分析可知,CGAN 模型经过训练后虽然也能对黑白图像进行上色,但色彩较为单一;将 CGAN 模型的 BN 项替换成 IN 项,其色彩要比原模型鲜艳,但色彩分布不均匀;采用 GN 项替换后虽然色彩分布较为均匀但是色调偏暗,最后文中把 IN 项和添加的 Self-Attention 项结合起来,其效果比原先算法都要好。

## 5 量化对比分析

为了更加准确地分析生成图片的质量好坏,该算法采用了衡量图片质量的若干指标:均方差(MSE)<sup>[15]</sup>,数值越小表示越接近原图,峰值信噪比(PSNR)<sup>[16]</sup>和结构相似度(SSIM)<sup>[17]</sup>,数值越大表示画质越好。该算法选取了测试集作综合比较,最终将实验结果进行定量计算,结果如表 1 所示。

表 1 不同算法对数据平均客观指标对比

指标	CGAN	GN	IN	文中
MSE	363.86	442.58	353.55	303.51
PSNR	22.71	21.71	22.87	24.92
SSIM	0.61	0.49	0.65	0.72

从表 1 可知,使用实例归一化 IN 进行网络优化以及在生成器中引入自注意力 Self-Attention 机制的效果相较于 CGAN、GN 和 IN 在测试集的评价结果,MSE 分别减少了 60.35,139.07 和 50.04,PSNR 分别提高了 2.21 dB,3.91 dB 和 2.03 dB,SSIM 指标略有提高,这是因为实例归一化能够较好地保留底层信息,另外由于黑白图像到彩色图像的映射关系复杂,引入自注意力机制能够提高生成图像的上色效果,进一步体现了该算法的优越性。

## 6 结束语

文中研究了基于改进生成对抗网络的上色算法,实现了对黑白照片/图像的上色。该算法通过理论分析及对比试验发现,当采用实例归一化 IN 以及在生成器中引入自注意力机制 Self-Attention,最终的实验效

果在指标评价上均要比原始模型有较大提高,这充分说明了该算法的可行性和有效性。由于目前的上色方法中,研究对象均是低分辨率图像,后续将对高分辨率图像上色作进一步研究。

## 参考文献:

- [1] 姚晨. 彩色化和色彩转移图像处理关键技术研究[D]. 上海:上海交通大学,2012.
- [2] 靳坤. 黑白视频的色彩迁移技术的研究与实现[D]. 呼和浩特:内蒙古大学,2016.
- [3] 陈添丁,金炜炜,陈英旦,等. 真实图像转换的水墨图像绘制模拟[J]. 中国图象图形学报,2014,19(6):946-954.
- [4] 孙玉红. 线条画风格转换与定制方法的研究[D]. 济南:山东大学,2005.
- [5] 卢少平,张松海. 基于视觉重要性的图像油画风格化绘制算法[J]. 计算机辅助设计与图形学学报,2010,22(7):1120-1125.
- [6] 朱薇,刘利刚. 保色调的黑白卡通图像着色方法[J]. 计算机辅助设计与图形学学报,2011,23(3):392-398.
- [7] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Advances in Neural Information Processing Systems, 2014, 3:2672-2680.
- [8] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]//International conference on medical image computing and computer-assisted intervention. Munich:Springer, 2015:234-241.
- [9] 袁帅,秦贵和,晏婕. 应用残差生成对抗网络的路况视频帧预测模型[J]. 西安交通大学学报,2018,52(10):146-152.
- [10] 唐贤伦,杜一铭,刘雨微,等. 基于条件深度卷积生成对抗网络的图像识别方法[J]. 自动化学报,2018,44(5):855-864.
- [11] 杨天祺,黄双喜. 改进卷积神经网络在分类与推荐中的实例应用[J]. 计算机应用研究,2018,35(4):74-77.
- [12] 董伟. 生成对抗网络研究及其在图像风格转换的应用[D]. 宁波:宁波大学,2018.
- [13] 郑玉昆,李丹,范臻,等. 一种基于自注意力机制的多任务深度阅读理解模型[J]. 中文信息学报,2018,32(11):128-134.
- [14] 曹志义,牛少彰,张继威. 基于半监督学习生成对抗网络的人脸还原算法研究[J]. 电子与信息学报,2018,40(2):323-330.
- [15] 曹继华,王兆华. 基于均方误差最小的图像错误隐藏内插算法[J]. 通信学报,2000,21(12):1-4.
- [16] 黄小乔,石俊生,杨健,等. 基于色差的均方误差与峰值信噪比评价彩色图像质量研究[J]. 光子学报,2007,36:295-298.
- [17] 王宇庆,刘维亚,王勇. 一种基于局部方差和结构相似度的图像质量评价方法[J]. 光电子·激光,2008,19(11):1546-1553.