

马咏莉, 李嘉培. 基于 CycleGAN 的透射文本图像复原[J]. 智能计算机与应用, 2024, 14(7): 211-215. DOI: 10.20169/j.issn.2095-2163.240733

基于 CycleGAN 的透射文本图像复原

马咏莉, 李嘉培

(郑州科技学院 信息工程学院, 郑州 450064)

摘要: 为了解决文本透射图像复原方法中实际成对图像数据集缺乏而导致的去透射效果欠佳的问题, 本文提出了一种基于 CycleGAN 的透射文本图像复原方法。利用 CycleGAN 的约束转移学习能力完成非成对图像的复原任务; 考虑到实际透射文本图像的复杂性, 结合人类视觉特征, 在 CycleGAN 中融入注意力机制, 非均匀地处理不同文本特征。实验结果表明, 所提方法在公共图像数据集以及真实图像数据上都取得了较好的透射图像恢复结果。

关键词: 图像复原; CycleGAN; 注意力机制; 透射图像恢复

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2024)07-0211-05

Bleed-through text image restoration based on CycleGAN

MA Yongli, LI Jiapei

(College of Information Engineering, Zhengzhou University of Science and Technology, Zhengzhou 450064, China)

Abstract: To solve the problem of poor de-bleed-through effect caused by the lack of paired image datasets in text bleed-through image restoration methods, a bleed-through text image restoration method based on CycleGAN is proposed. Utilize the constraint transfer learning ability of CycleGAN to complete the restoration task of non-paired images. Considering the complexity of actual bleed-through text images, combined with human visual features, attention mechanisms are incorporated into CycleGAN to non-uniformly process different text features. The experimental results show that the proposed method achieves good bleed-through image restoration results on both public image datasets and real image data.

Key words: image restoration; CycleGAN; attention mechanism; bleed-through image restoration

0 引言

将纸质内容进行成像时, 常会将页面背面信息透射叠加到正面文字内容中, 历史古籍中也存在杂质, 会给文本的使用和修复增加困难^[1-3]。恢复正面文本图像中的文字内容和细节信息是实际图像处理应用中的重要一环。

现有经典去除透射方法大多是基于阈值处理, 对阈值进行估计^[4-5]。由于正面与反面文字信息相近, 所以基于阈值估计的透射去除是典型的欠定问题, 通过阈值处理生成的图像质量取决于多种因素, 从数字化设备及其设置到文档的内在特征, 从纸张颜色和纹理到文档的手写或打印方式都会影响最终图像的质量。二值化所花费的时间也取决于文档特征, 并且在不同的算法之间变化很大。大多数用户

倾向于使用经典算法对文档图像进行透射去除, 如 Otsu 算法根据图像的灰度特性, 将图像分割成前景区域和背景区域, 根据分割的阈值计算一个类间方差, 将类间方差最大时对应的阈值作为最佳阈值^[6]; Sauvola 算法通过计算局部阈值来进行图像复原^[7]。Lins^[8]表明, 如果向算法提供彩色图像、其灰度转换图像或其 R、G、B 通道之一, 则二值化算法的性能可能不同; Mehri 和 Tuceryan^[9-10]表明, 纹理分析在文档图像处理中发挥着重要作用。

近几年生成对抗网络 (Generative Adversarial Networks, GAN) 在很多应用场景中取得了良好的性能, GAN 的输入是随机噪声, 经过一系列反卷积和上采样操作后, 随机噪声映射到图像域中的相应图像^[11]。随着 GAN 的发展, 图像生成技术迎来了新的发展机遇, 更容易实现和生成质量更好图

作者简介: 马咏莉 (1996-), 女, 硕士研究生, 主要研究方向: 盲源分离。Email: mayonglim@163.com; 李嘉培 (1997-), 女, 硕士研究生, 主要研究方向: 数据挖掘。

收稿日期: 2024-01-26

哈尔滨工业大学主办 ◆ 科技创新与应用

像^[12-14]。De^[15]提出一种拥有两个判别器的网络结构,通过结合全局和局部信息来达到图像二值化的目的;Castellanos^[16]通过无监督区域自适应的方法进行文档图像二值化处理。与传统去除透射的方法相比,这些基于GAN的方法会产生更好的结果,但其目的不是训练预测阈值的模型而是直接将正面的文本内容从透射图像中分离出来,存在训练过程中成对训练数据集采集困难且去透射效果不理想的问题;Zhu等^[17]提出通过循环一致性生成对抗网络(Cycle-consistent Generative Adversarial Networks, CycleGAN)解决图像复原中缺乏成对图像数据的问题。

本文利用无需成对数据的循环一致性生成对抗网络(Cycle-consistent Generative Adversarial Networks, CycleGAN),在CycleGAN网络结构的基础上加入注意力模块,提出一种基于CycleGAN的去透射文本图像复原的图像去透射方法(Bleed-

through text image restoration method based on CycleGAN, B-CycleGAN),利用注意力模块来提高图像中关注点的权重,进而提升图像去透射效果。

1 网络结构

1.1 整体网络结构

B-CycleGAN网络的结构如图1所示,基于CycleGAN并将注意力模块添加到生成器 G_n 中。B-CycleGAN有两个分支:

- (1)字-字分支: $w \rightarrow n_w \rightarrow w'$, 用有字图 w 生成无字图 n_w , 再重构有字图 w' ;
- (2)无字-无字分支: $n \rightarrow w_n \rightarrow n'$, 用无字图 n 生成有字图 w_n , 再重构无字图 n' 。

B-CycleGAN的两个分支各自带一个判别器 D_n 和 D_w ,并共享两个生成器:一个是生成器 G_w ,另一个是生成器 G_n 。生成器 G_n 将有字图像恢复成无字图像;生成器 G_w 将无字图像进行加字处理。

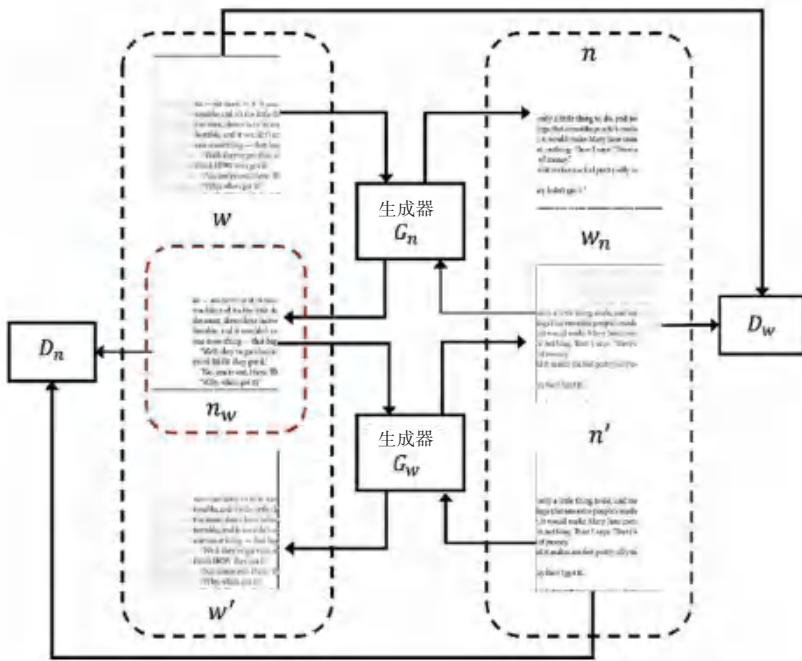


图1 B-CycleGAN的网络结构

Fig. 1 Structure of B-CycleGAN

1.2 生成网络

生成器 G_n 用于保留正面文字信息的内容并在去除透射过程中恢复正面文字细节,网络结构如图2所示。生成器 G_w 采用CycleGAN的原始体系结构。

网络的输入是一个带有透射的图像,先将图像做二值化的预处理,根据图像的灰度特性,将图像分为前景和背景两个部分,经过多层卷积处理,再将图

像传递到两个密集块中,每个密集块都结合了跳跃链接和注意力模块。两个密集块的输出特征最终通过注意力模块融合在一起,以获得无透射的输出。密集块中包含多个卷积层,增加了网络的深度,提高了特征提取能力,注意力模块使网络专注于有透射区域或其他有效信息,分主次来处理不同的特征,生成器 G_n 在处理不同内容时提供了额外的灵活性并扩展了网络的表达能力。

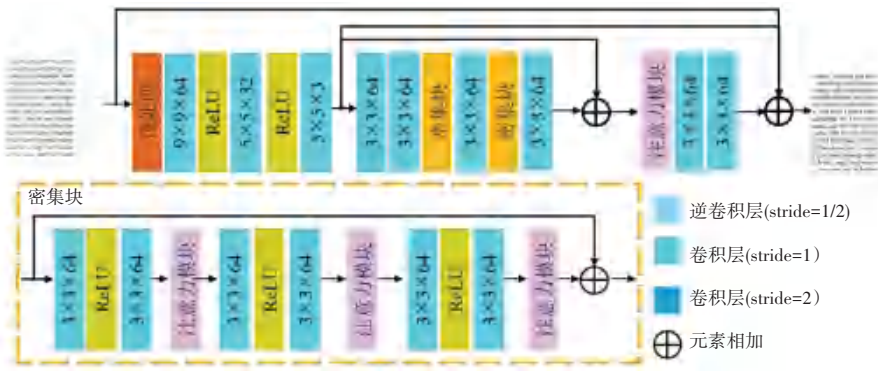


图 2 生成器 G_n 的网络结构
Fig. 2 Structure of generator G_n

1.3 注意力模块

人类视觉注意力机制表明,视觉系统总能在复杂环境中捕获到感兴趣的内容,即使在前景内容和背景内容混合的情况下,也能较好地分辨视觉目标^[18]。由于实际的文档图像中文字分布的复杂性,在网络中融入注意力模块,非均匀地处理不同特征和不同区域^[19-20]。B-CycleGAN 的注意力模块结构如图 3 所示。

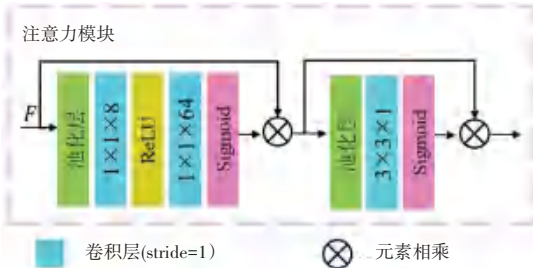


图 3 注意力模块结构
Fig. 3 Structure of attention mechanism modules

首先,将特征图形状由 $C \times H \times W$ 变为 $C \times 1 \times 1$, C 表示通道数, $H \times W$ 表示图像大小;其次,通过卷积层、ReLU、Sigmoid 激活函数和池化层来获取特征;最后,将通道特征中的通道数降到 1,从而实现不同区域文字信息的响应。

1.4 损失函数

由于没有成对训练数据集来网络训练,生成的图像可能无法保持图像中的文字结构细节信息,所以在网络中引入了循环一致性损失,用于最小化透射图像 w 与重建的透射图像 w' , 无透射图像 n 与重建的无透射图像 n' ^[17]。循环一致性损失 (\mathcal{L}_{cycle}) 定义如式(1)所示:

$$\mathcal{L}_{cycle} = E_{w \sim p_{data}(w)} [\|w' - w\|_1] + E_{n \sim p_{data}(n)} [\|n' - n\|_1] \quad (1)$$

其中, w 和 n 分别表示透射图像和无透射图像,并且不需要配对。

B-CycleGAN 中有两个判别器 D_n 和 D_w , D_n 用来区分无透射图像和生成的无透射图像, D_w 用来区分有透射图像和生成的透射图像。将 D_n 的对抗损失 (\mathcal{L}_{D_n}) 定义为式(2):

$$\mathcal{L}_{D_n} = E_{d \sim p_{data}(n)} [\log D_n(n)] + E_{h \sim p_{data}(h)} [\log(1 - D_n(D_n(w)))] \quad (2)$$

判别器 D_n 的作用是最大化目标函数并以此区分生成的无透射图像和真实的无透射图像。生成器 G_w 要最小化损失,使生成的无透射图像接近真实的无透射图像。同样,判别器 D_w 的对抗损失 (\mathcal{L}_{D_w}) 定义为式(3):

$$\mathcal{L}_{D_w} = E_{w \sim p_{data}(w)} [\log D_w(w)] + E_{h \sim p_{data}(h)} [\log(1 - D_w(G_w(h)))] \quad (3)$$

最后,结合循环一致性损失 \mathcal{L}_{cycle} , 对抗损失 \mathcal{L}_{D_d} 和 \mathcal{L}_{D_h} , 和感知损失 $\mathcal{L}_{perceptual}$, B-CycleGAN 的整体损失函数 \mathcal{L}_{total} 定义为式(4):

$$\mathcal{L}_{total} = \mathcal{L}_{cycle} + \mathcal{L}_{D_w} + \mathcal{L}_{D_n} + k * \mathcal{L}_{perceptual} \quad (4)$$

其中, k 是感知损失函数在整体损失函数 \mathcal{L}_{total} 中的权重系数。

2 实验结果与分析

2.1 实验数据及评价指标

实验采用的数据集为公共数据集以及真实的透射图像。公共数据集选择 DIBCO 2013、DIBCO 2017、H-DIBCO 2016;真实的透射图像选择学生的考试试卷部分截图。为了更客观评价透射图像的恢复效果,本文还使用 4 个评估指标:峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)、F 值(F-Measure, FM)、伪 F 值(pseudoF-Measure, pFM) 和距离倒数失真度量(Distance Reciprocal Distortion metric, DRD)。PSNR、FM 和 pFM 指标值越大说明算法准确性越高,DRD 越小说明像素恢复差错越少。

2.2 实验结果

2.2.1 公共数据集实验结果与分析

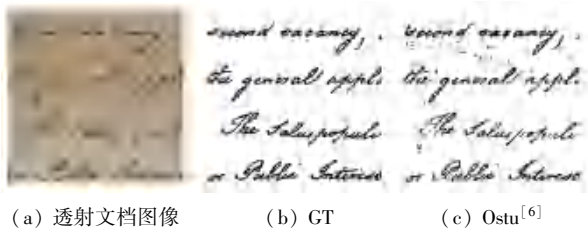
传统方法中效果较好的是 Ostu^[6] 和 Savola^[7] 基于阈值来去除透射, 基于深度学习方法中 Castellanos^[16] 方法使用生成对抗网络对图像进行复原, 将本文方法与这3种方法进行实验对比, 在公共数据集上文档透射图像恢复的定量评价结果见表1。显然, 本文方法 B-CycleGAN 在4个评价指标中取得了较好的结果。

表1 DIBCO 数据集的文档透射图像恢复定量评价

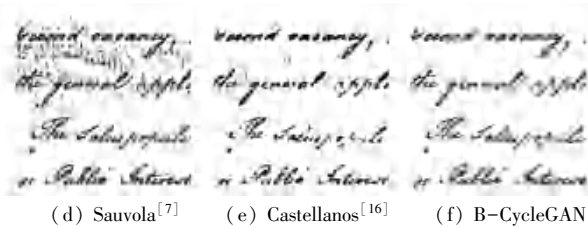
Table 1 Quantitative evaluation of document bleed-through image restoration of DIBCO datasets

方法	PSNR/ dB	FM/ %	pFM/ %	DRD
Ostu ^[6]	18.52	67.81	74.08	17.45
Sauvola ^[7]	17.58	87.56	88.25	8.35
Castellanos ^[16]	23.54	89.67	91.03	5.59
B-CycleGAN	25.28	90.15	90.53	5.43

不同去透射方法在数据集 DIBCO 2013 中的一个图像样本的恢复结果如图4所示, 其中, 图4(b) GT 为实验基准图像。从实验结果中可以看出, Ostu 方法的恢复结果正面文本内容留有大量的噪声; Sauvola 方法虽然将正面文字内容提取出来, 但是引入了大量的噪声; Castellanos 方法对文字的提取较为准确, 但仍有不必要的文字痕迹; 本文 B-CycleGAN 方法在接近真实图像方面有更好的视觉效果, 与其他方法相比, B-CycleGAN 可以在不需要成对数据集的情况下, 无论是强透射区域还是弱透射区域都可以生成更为清晰的细节, 更好地恢复透射文档图像。



(a) 透射文档图像 (b) GT (c) Ostu^[6]



(d) Sauvola^[7] (e) Castellanos^[16] (f) B-CycleGAN

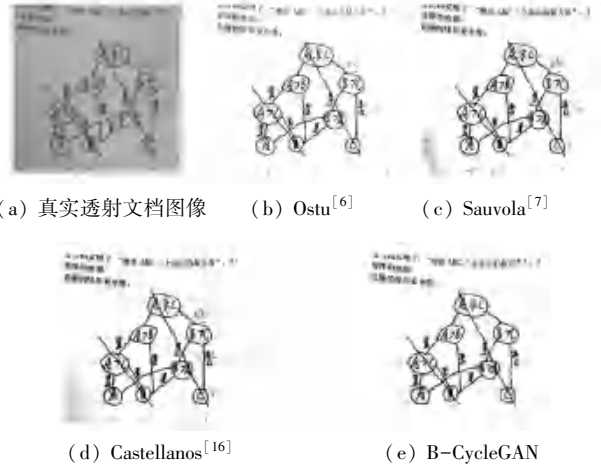
图4 不同方法在 DIBCO 2013 数据集内一个样本的恢复结果

Fig. 4 Experiment results of one sample in DIBCO 2013 datasets by different methods

2.2.2 真实透射图像实验结果

使用扫描仪获取的期末考试试卷的真实图像,

截取其中透射明显的图像样本进行恢复, 恢复结果如图5所示。Sauvola 和 Castellanos 方法的恢复结果中, 正面文字内容结果清晰, 但是有许多阴影在恢复结果中出现, 影响进行后续图像处理任务; Ostu 方法的恢复结果更好, 但仍有背面文字内容透射到正面; 本文 B-CycleGAN 方法不仅能更好地去除透射内容和噪声, 而且清晰保留了文本内容。



(a) 真实透射文档图像 (b) Ostu^[6] (c) Sauvola^[7]

(d) Castellanos^[16] (e) B-CycleGAN

图5 不同方法在期末考试试卷的扫描文档图像透射图像的恢复结果

Fig. 5 Experiment results of final exam papers bleed-through datasets by different methods

3 结束语

本文提出了一种基于 CycleGAN 的透射文本图像复原的图像去透射方法, 用来解决单幅图像的去透射问题。该方法直接在不配对的图像上执行图像去透射任务, 在训练和测试阶段, 不需要透射图像和相应的真实透射图像数据, 可以有效去除背面文档信息, 重建内容细节, 并通过注意力机制来提高恢复效果。实验结果表明, 对于合成数据集和真实的有透射图像, 该方法的性能优于其他的去透射方法, 具有更好的正面文本恢复效果。

参考文献

- [1] 吴佳芬, 马费成. 产品虚假评论文本识别方法研究述评[J]. 现代图书情报技术, 2019 (9): 1-15.
- [2] 马然. 基于深度学习的自然场景文本识别系统的设计与实现[D]. 长春: 吉林大学, 2023.
- [3] 李玮. 适配性原则在古籍修复中的实践研究-以清《重刊灾赈果报录》修复为例[J]. 兰台世界, 2018, 11: 110-113.
- [4] 陶文兵, 金海. 一种新的基于图谱理论的图像阈值分割方法[J]. 计算机学报, 2007, 30(1): 110-119.
- [5] 龙建武, 申铨京, 陈海鹏. 基于图像区域的交互式文本图像阈值分割算法[J]. 计算机研究与发展, 2012, 49(7): 1420-1431.
- [6] OTSU N. A threshold selection method from gray-level histograms [J]. Automatica, 1975, 11(285-296): 23-27.

- [7] SAUVOLA J, SEPPANEN T, HAAPAKOSKI S, et al. Adaptive document binarization [C]//Proceedings of the Fourth International Conference on Document Analysis and Recognition. IEEE, 1997: 147-152.
- [8] LINS R D, BERNARDINO R B, DA SILVA BARBOZA R, et al. Direct binarization a quality-and-time efficient binarization strategy [C]//Proceedings of the 21st ACM Symposium on Document Engineering. IEEE, 2021: 1-4.
- [9] MEHRI M, HÉROUX P, GOMEZ-KRÄMER P, et al. Texture feature benchmarking and evaluation for historical document image analysis[J]. International Journal on Document Analysis and Recognition (IJ DAR), 2017, 20(1): 1-35.
- [10] TUCERYAN M, JAIN A K. Texture analysis[J]. Handbook of Pattern Recognition and Computer Vision, 1993: 235-276.
- [11] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]//Proceedings of Neural Information Processing Systems, IEEE, 2014: 2672-2680.
- [12] 黄远,白琮,李宏凯,等. 基于条件生成对抗网络的图像描述生成方法[J]. 计算机辅助设计与图形学学报, 2020, 32(6): 911-918.
- [13] 赵树阳,李建武. 基于生成对抗网络的低秩图像生成方法[J]. 自动化学报, 2018, 44(5): 829-839.
- [14] 李庆旭,王巧华,马美湖. 基于生成对抗网络的禽蛋图像数据生成研究[J]. 农业机械学报, 2021, 52(2): 236-245.
- [15] DE R, CHAKRABORTY A, SARKAR R. Document image binarization using dual discriminator generative adversarial networks[J]. IEEE Signal Processing Letters, 2020, 27: 1090-094.
- [16] CASTELLANOS F J, GALLEGO A J, CALVO-ZARAGOZA J. Unsupervised neural domain adaptation for document image binarization[J]. Pattern Recognition, 2021, 119: 108099.
- [17] ZHU J Y, PARK T, ISOLA P. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2223-2232.
- [18] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.
- [19] QIN X, WANG Z, BAI Y, et al. FFA-Net: Feature fusion attention network for single image dehazing [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 11908-11915.
- [20] YUE Z, ZHAO Q, ZHANG L, et al. Dual adversarial network: Toward real-world noise removal and noise generation [C]//Proceedings of Computer Vision - ECCV 2020: 16th European Conference. IEEE, 2020: 41-58.