

自监督学习的单幅透射图像恢复

徐金东¹ 马咏莉¹ 梁宗宝¹ 倪梦莹¹

摘要 现有基于学习的单幅透射图像恢复方法常需要大量成对的标签数据来训练模型, 因缺乏成对图像集的监督约束, 致使透射图像恢复效果欠佳, 限制了其实用性. 提出了一种基于自监督学习的单幅透射图像恢复方法, 利用循环一致性生成对抗网络的循环结构和约束转移学习能力实现非成对图像的模型训练, 通过设计自学习模块, 从大规模的无监督数据中挖掘自身的监督信息对网络进行训练, 以此形成有效的从浅层到深层的特征提取, 提高透射图像正面内容的纹理、边缘等细节信息恢复质量, 实现单幅图像的透射去除. 实验结果表明, 该方法在合成图像数据集、公共图像数据集以及真实图像数据集上都取得了较好的透射图像恢复结果.

关键词 特征提取, 透射去除, 图像恢复, 自监督, 生成对抗网络

引用格式 徐金东, 马咏莉, 梁宗宝, 倪梦莹. 自监督学习的单幅透射图像恢复. 自动化学报, 2023, 49(1): 219–228

DOI 10.16383/j.aas.c220165

Single Bleed-through Image Restoration With Self-supervised Learning

XU Jin-Dong¹ MA Yong-Li¹ LIANG Zong-Bao¹ NI Meng-Ying¹

Abstract The existing single bleed-through image restorations based on learning methods often need a large number of paired images to train the model, due to the lack of paired images supervision constraints, the restoration effect of bleed-through image is not well, which limits their practicability. In this paper, a single bleed-through image restoration method based on self-supervised learning is proposed, the method uses the circulatory structure and constraint transfer learning ability of cycle-consistent generative adversarial networks to realize self-supervised single bleed-through image removal of unpaired images. Through the design of self-learning module, mining self-supervised information from large-scale unsupervised data are used to train the network, so as to form an effective feature extraction module from shallow layer to deep layer, which improve the restoration quality of texture, edge and other detail information of the front content of the bleed-through image. Finally, the bleed-through content has been removed for a single image. The experimental results show that this method achieves good results for bleed-through image restoration on synthetic image datasets, public image datasets and real image datasets.

Key words Feature extraction, bleed-through removal, image restoration, self-supervised, generative adversarial network

Citation Xu Jin-Dong, Ma Yong-Li, Liang Zong-Bao, Ni Meng-Ying. Single bleed-through image restoration with self-supervised learning. *Acta Automatica Sinica*, 2023, 49(1): 219–228

扫描仪、相机和摄像机等设备对纸质内容进行成像时, 背面信息常会透射叠加到正面文字内容中, 这不仅降低了图像内容的可读性, 还会影响图像的后续处理, 如电子阅卷、历史文档数字恢复^[1]和场景文本识别^[2]等. 透射图像恢复, 即透射去除, 是从含有背面和正面内容的混合图像中恢复出正面内容的过程, 可建立如式 (1) 所示模型.

$$I = (1 - \alpha)F + \alpha g(T) \quad (1)$$

其中, I 表示有透射的混合图像, F 表示正面图像, α 是透射混合参数, T 表示背面图像, $g(\cdot)$ 为透射衰减函数. 因 T 、 α 和 $g(\cdot)$ 均未知, 从 I 中去除 T 、恢复 F 是一个不适定的 NP (Non-deterministic polynomial) 难问题. 而且, 背面图像的内容结构和属性常与正面图像相似, 导致难以在去除背面透射信息的同时恢复正面文档图像中的内容和细节.

现有文档图像的恢复可以分为两大类: 基于阈值处理的方法和基于学习的方法. 阈值处理法是传统的文档图像恢复方法^[3–5], 通过像素阈值判分正面或背面内容, 文献 [6] 提出了一种结合局部图像对比度和局部图像梯度的自适应对比度图, 采用局部区域中检测到的边缘来估计局部阈值, 但由于透射文档图像质量参差不齐, 在计算局部和全局阈值时需要大量的经验参数, 系统结构复杂、运行效率较

收稿日期 2022-03-10 录用日期 2022-07-21
Manuscript received March 10, 2022; accepted July 21, 2022
国家自然科学基金 (62072391, 62066013) 资助
Supported by National Natural Science Foundation of China (62072391, 62066013)
本文责任编辑 陈胜勇
Recommended by Associate Editor CHEN Sheng-Yong
1. 烟台大学计算机与控制工程学院 烟台 264005
1. School of Computer and Control Engineering, Yantai University, Yantai 264005

低。为减轻参数设计和调整的负担,基于学习的方法通过训练的思路来获取图像恢复模型,文献 [7] 和文献 [8] 通过引入不同的分类器直接对图像特征进行分类,减少参数的数量,提高了文档二值化方法的效率。近几年神经网络在不少应用场景中取得良好性能,基于深度学习透射图像恢复的方法^[9-12] 备受研究者的青睐,这类方法能够实现端到端的从透射图像中恢复出正面内容,但需要大量的成对数据集,且常存在过拟合现象。最近,生成对抗网络 (Generative adversarial networks, GAN)^[12] 在合成图像方面大放异彩,并且一部分科研人员将其用于图像恢复相关处理任务,文献 [13] 利用 GAN 扩充训练数据集,处理图像二值化任务,文献 [14] 使用 Pix2Pix GAN 去除了光学音乐识别的五线谱,文献 [15] 引入了一种包含两个判别器网络的双判别器 GAN 结构,以结合全局和局部信息,而 Castellanos^[16] 提出了一种使用无监督区域自适应的神经网络方法将文档图像进行二值化操作,与其他学习的方法相比,这些基于 GAN 的模型会产生更好的结果,但是仍需要大量的成对数据进行训练,难以泛化。

为了解决以上问题,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)^[17], 提出了一种基于自监督学习的单幅透射图像恢复方法 (Self-supervised learning based on cycle-consistent generative adversarial networks, S-CycleGAN), 主要由特征提取模块和自学习模块组成,特征提取模块通过跳跃链接融合全局和局部特征,以增强去透射模型的表达能力,自学习模块能对不同

的特征映射赋予不同的权重并进行自适应学习,进而实现透射图像的高质量恢复。

本文的主要创新概括为:

- 1) 将 S-CycleGAN 应用于文档图像恢复,提出了一种基于自监督学习的单幅透射图像恢复方法,并用于单幅图像中透射内容去除,该方法不依赖于先验知识,在训练过程中不需要成对数据集;
- 2) 设计了一种新颖的去透射生成器,它结合了自学习模块和特征提取模块来自监督训练网络提取图像特征,尽可能保留了图像的细节内容;
- 3) 通过设计特征提取模块、自学习模块和跳跃链接,融合全局和局部特征,增加了 S-CycleGAN 的深度,提高了文本内容的表达能力,产生视觉上满意的恢复效果。

1 自监督学习的单幅透射图像恢复

S-CycleGAN 是将未配对图像的去透射问题转化为图像到图像的生成问题,利用自监督学习的循环一致性网络生成无透射的图像。S-CycleGAN 整体网络架构如图 1 所示,涉及的核心内容包括:生成器、判别器和损失函数。

S-CycleGAN 由两个分支组成: 1) 透射-透射分支: $x \rightarrow G_y(x) \rightarrow G_x(G_y(x))$, 利用有透射图像 x 生成无透射图像 $G_y(x)$, 再重构有透射图像 $G_x(G_y(x))$; 2) 无透射-无透射分支: $y \rightarrow G_x(y) \rightarrow G_y(G_x(y))$, 利用无透射图像 y 生成有透射图像 $G_x(y)$, 再重构无透射图像 $G_y(G_x(y))$ 。

S-CycleGAN 的两个分支中的每一个分支都有一个判别器, 分别是 D_y 和 D_x , 这两个分支共享两

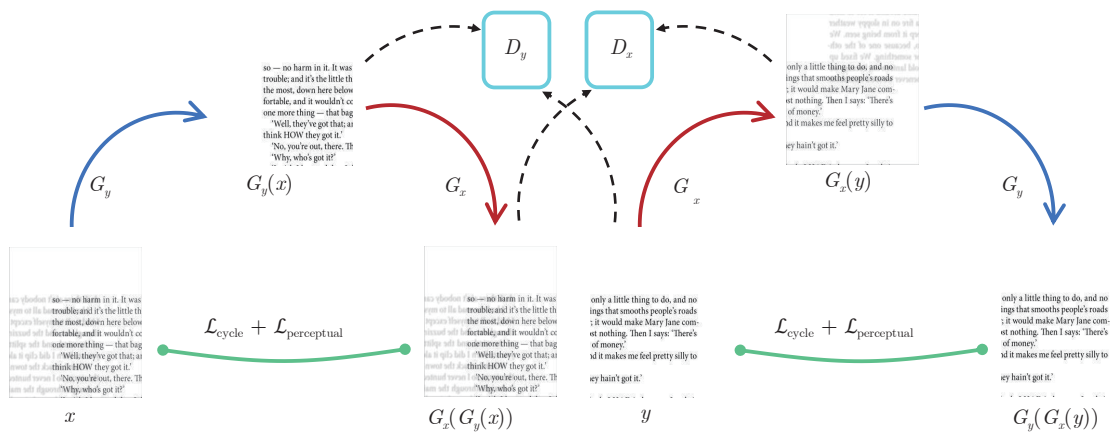


图 1 S-CycleGAN 的网络结构 (G_y 和 G_x 表示生成器, D_y 和 D_x 表示判别器. x 和 y 分别表示输入的透射图像和无透射图像, $\mathcal{L}_{\text{cycle}}$ 和 $\mathcal{L}_{\text{perceptual}}$ 分别表示循环一致性损失和感知损失)

Fig.1 Structure of S-CycleGAN (G_y and G_x are generators while D_y and D_x are discriminators, x and y represent the input bleed-through image and non-bleed-through image respectively, $\mathcal{L}_{\text{cycle}}$ and $\mathcal{L}_{\text{perceptual}}$ represent cycle consistency loss and perceptual loss respectively)

个生成器 G_y 和 G_x . G_y 是将透射图像生成 (恢复) 为无透射图像, 而 G_x 是将透射部分添加到无透射图像中生成有透射图像.

1.1 生成器网络结构

生成器 G_x 采用与 CycleGAN 相似的编码器-解码器网络结构, 本文 G_x 使用了 6 个残差块. 为了实现自监督学习以及增强特征表示, 本文方法在 CycleGAN 生成器 G_x 的基础上构建生成器 G_y , 在 G_y 网络结构中加入特征提取模块和自学习模块, 从大规模的无标签数据中挖掘自身的监督信息, 从而得到更好的去透射结果. 生成器 G_y 的作用是恢复正面的图像内容, 在去除透射过程中保留细节, 网络结构如图 2 所示.

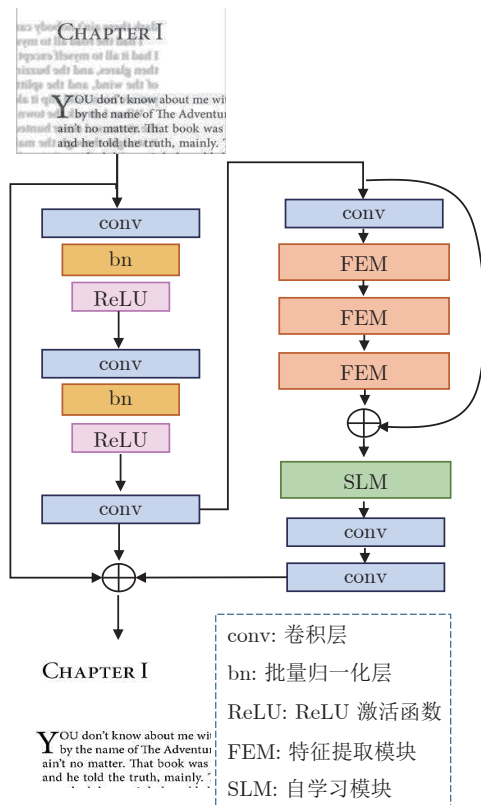


图 2 生成器 G_y 的网络结构

Fig.2 The network structure of G_y

G_y 网络的输入是一幅带有透射的图像, 首先经过三层卷积来增强输入图像纹理细节, 提高正面图像与背面图像的对比度, 然后馈送到三个具有跳跃链接的特征提取模块 (Feature extractor module, FEM) 中, 以保留低层信息并将其传递到深层网络中, 三个特征提取模块的输出特征通过自学习模块 (Self-learning module, SLM) 以及两层卷积融合获得一个无透射的输出. 其中, 自学习模块的作

用是获得的自适应权值, 使得 S-CycleGAN 更加关注混合内容中较严重区域和正面纹理等细节信息.

1.1.1 特征提取模块-FEM

FEM 的网络结构如图 3 所示, 包括卷积层、ReLU 层和 SLM. 在 S-CycleGAN 的设计中, 生成器 G_y 的目的是将带有透射的输入图像生成无透射的图像, 并且不需要估计中间参数, 为了实现这个目标, 生成器 G_y 应该尽可能地保留正面图像内容和细节, 同时消除透射信息. 因此, 生成器 G_y 中的特征提取模块应充分利用透射图像的特征信息.

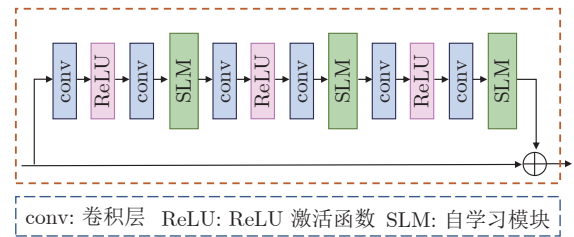


图 3 FEM 的网络结构

Fig.3 The network structure of FEM

FEM 由多个卷积、ReLU 和 SLM 的功能层组成, 以层次递进方式反复提取特征信息. 经过图 3 所示的由浅入深的功能层后, 会交换大量不必要的冗余信息, 从而导致一些有用特征信息丢失. 因此, 在特征提取块中增加了跳跃链接, 将浅层特征与深层特征融合, 使网络保留更多的有效特征信息, 从而更有利于恢复清晰的图像. 另外, 为了平衡网络性能和时间复杂度, 将 FEM 的所有卷积层的通道数设为 64, 卷积核大小设为 3×3 .

1.1.2 自学习模块-SLM

SLM 的网络结构如图 4 所示, 主要用于关注图像的不同特征. 输入的特征图 F 大小由 $C \times H \times W$ 变为 $C \times 1 \times 1$, 其中 C 表示通道数, $H \times W$ 表示图像大小. 通过两个卷积层和 ReLU、sigmoid 激活函数来获取特征, 与输入的特征图 F 做元素相乘操作得到 F_c , 在这一阶段的卷积核大小为 1×1 , 通道数分别为 8 和 64, 步长为 1. 过程可如式 (2) 所示.

$$F_c = F \otimes [\sigma (\text{Conv} (\delta (\text{Conv} (\text{pool} (F)))))] \quad (2)$$

其中, $\text{pool}(\cdot)$ 表示平均池化操作, $\text{Conv}(\cdot)$ 表示卷积操作, $\sigma(\cdot)$ 是 sigmoid 函数, $\delta(\cdot)$ 是 ReLU 函数, \otimes 是元素相乘, F_c 表示此阶段的输出.

考虑到透射到正面的背面内容“深浅不一”, 透射图像的背面内容常不均匀地叠加在正面图像上, 因此需要 SLM 能够捕捉特征图 F_c 中的高响应区域, 使模型更加关注相应区域并对此进行处理. 因此, 特征图 F_c 通过池化、卷积和 Sigmoid 激活操作

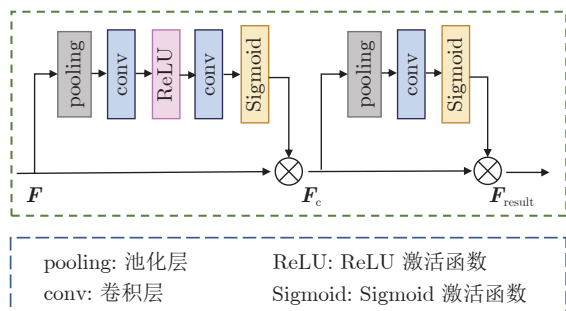


图 4 SLM 的网络结构 (F 表示自学习模块的输入, F_c 表示自学习模块的中间输出, F_{result} 表示自学习模块的输出)

Fig. 4 The network structure of SLM (F is the input to the self-learning module. F_c is the intermediate output of the self-learning module. F_{result} is the output of the self-learning module)

后, 大小由 $C \times H \times W$ 变为 $1 \times H \times W$, 通道数降为 1, 将会得到区域特征信息, 与输入特征图 F_c 做乘法操作, 获得最终的关注不同相应区域的权重特征图, 在这一阶段的卷积核大小为 3×3 , 通道数为 1, 步长为 1. 处理过程如式 (3) 所示.

$$F_{result} = F_c \otimes [\sigma(\text{Conv}(\text{pool}(F_c)))] \quad (3)$$

其中, F_{result} 表示自学习模块的最终输出.

为了解释自学习模块在网络中的有效性, 对特征提取模块结构输出的特征权重图进行可视化. 图 5 显示了尺寸为 4×64 的特征图, 可以清楚地看到, 网络以不同的权重自适应地学习不同的特征映射.

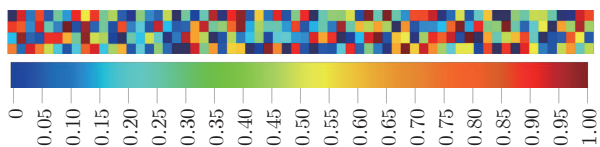


图 5 自学习模块权重图

Fig. 5 Self-learning module weight map

1.2 判别器网络结构

判别器用于区分生成图像的“真假”, 其输入是生成器网络生成的图像. 在 S-CycleGAN 中, 有两个判别器, 即 D_y 和 D_x . D_y 用于区分生成的无透射图像和真实的无透射图像, D_x 用于区分生成的透射图像和真实的透射图像. 判别器 D_y 和 D_x 采用相同的网络结构, 如图 6 所示. 判别器的前 4 组模块用于提取特征, 最后一组模块用于确定生成图像的“真假”. 网络中 5 个卷积层通道数依次为 64、128、256、512 和 1, 每一层的卷积核大小为 4×4 、步长为 2.

1.3 损失函数

由于采用非成对的数据来监督网络, 生成的图

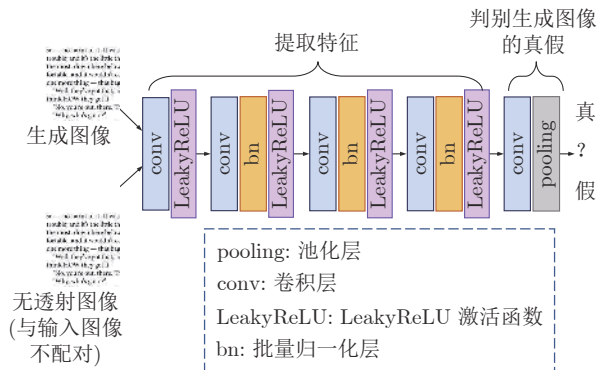


图 6 判别器的网络结构

Fig. 6 The network structure of discriminator

像无法保持图像中的颜色和结构信息, 故引入循环一致性损失^[17]来最小化透射图像 x 和其重构的透射图像 $G_x(G_y(x))$ 、无透射图像 y 和其重构的无透射图像 $G_y(G_x(y))$. 循环一致性损失 (\mathcal{L}_{cycle}) 定义为式 (4).

$$\mathcal{L}_{cycle} = E_{x \sim P_{data(x)}} [\|G_y(G_x(y)) - x\|_1] + E_{y \sim P_{data(y)}} [\|G_x(G_y(x)) - y\|_1] \quad (4)$$

其中, x 和 y 表示不需要配对的透射图像和无透射图像, E 表示数学期望, $x \sim P_{data(x)}$ 表示透射数据集中样本的概率分布, $y \sim P_{data(y)}$ 表示无透射数据集中样本的概率分布, $\|\cdot\|_1$ 表示 L_1 范数.

判别器 D_y 的作用是最大化损失, 并以此区分生成的无透射图像和真实的无透射图像, 而生成器 G_y 要使损失最小化, 使生成的无透射图像接近真实的无透射图像. 因此, D_y 的对抗性损失 (\mathcal{L}_{D_y}) 定义为式 (5).

$$\mathcal{L}_{D_y} = E_{y \sim P_{data(y)}} [\ln(D_y(y))] + E_{x \sim P_{data(x)}} [\ln(1 - D_y(G_y(x)))] \quad (5)$$

同样, 判别器 D_x 的对抗损失 (\mathcal{L}_{D_x}) 定义为式 (6).

$$\mathcal{L}_{D_x} = E_{x \sim P_{data(x)}} [\ln(D_x(x))] + E_{y \sim P_{data(y)}} [\ln(1 - D_x(G_x(y)))] \quad (6)$$

循环一致性损失和对抗性损失能够约束图像中的颜色和结构信息, 但对文档图像的细节信息没有考虑. 因此, 本文使用感知损失^[18]来使生成的图像在语义细节上更接近目标图像, 在生成器 G_y 和生成器 G_x 的约束下的感知损失 ($\mathcal{L}_{perceptual}$) 如式 (7) 所示.

$$\mathcal{L}_{perceptual} = \|\varphi(x) - \varphi(G_x(G_y(x)))\|_2^2 + \|\varphi(y) - \varphi(G_y(G_x(y)))\|_2^2 \quad (7)$$

其中, $\|\cdot\|_2$ 表示 L_2 范数, φ 表示 VGG-16 网络^[19]的特征提取器. 本文使用 VGG-16 网络的第 2 和第

5 个池化层中提取的特征按照式 (7) 进行组合, VGG-16 网络是由 ImageNet^[20] 预先训练的模型。

通过综合循环一致性损失 $\mathcal{L}_{\text{cycle}}$ 、对抗损失 \mathcal{L}_{D_y} 和 \mathcal{L}_{D_x} 以及感知损失 $\mathcal{L}_{\text{perceptual}}$, 形成 S-CycleGAN 的最终目标损失函数 ($\mathcal{L}_{\text{total}}$) 如式 (8) 所示。

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cycle}} + \mathcal{L}_{D_y} + \mathcal{L}_{D_x} + \omega \times \mathcal{L}_{\text{perceptual}} \quad (8)$$

其中, ω 是感知损失函数的权重系数。

2 实验结果与分析

2.1 实验设置

2.1.1 对比算法和数据集

为了充分测试 S-CycleGAN 的性能, 实验采用的数据集为: DIBCO 2013^[21]、DIBCO 2012^[22]、DIBCO 2014^[23]、DIBCO 2017^[24]、DIBCO 2011^[25]、DIBCO 2009^[26]、H-DIBCO 2016^[27] 和真实的透射图像。仿真合成图像按照式 (1) 进行混合, 其中 $g(\cdot)$ 为高斯模糊操作, 滤波器大小为 5×5 , 标准差为 2, α 在 $[0.15, 0.25]$ 间随机选择, 以尽可能模拟实际透射情况。

实验选择了 6 种具有代表性的文本图像恢复方法: 经典的 Otsu^[4] 算法和 Ntirogiannis^[5] 方法, 基于深度学习的 SAGAN^[28]、DD-GAN^[15]、Castellanos^[16] 和 Sungho^[29] 方法。

2.1.2 评价指标

为定量评价透射图像恢复效果, 使用文档图像二值化国际竞赛中的评估参数^[21-27]: 峰值信噪比 (Peak signal to noise ratio, PSNR)、F-measure (FM), pseudo-F-measure (pFM) 和距离倒数失真 (Distance reciprocal distortion, DRD)。其中, 前三个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。

1) PSNR

$$PSNR = 10 \lg \left(\frac{MAX_I^2}{MSE} \right) \quad (9)$$

其中, $MSE = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - B(i, j)]^2$, I 表示原始纯净图像, B 表示去透射图像, MAX_I 表示图片可能的最大像素值, 此处取 255。 (i, j) 表示对应像素的位置, $m \times n$ 表示图像大小。 $PSNR$ 越大, 表示去透射效果越好。

2) FM

$$FM = \frac{2 \times Precision \times Recall}{(Precision + Recall)} \quad (10)$$

其中, $Recall = \frac{TP}{(TP+FN)}$, $Precision = \frac{TP}{(TP+FP)}$, TP 、 FP 和 FN 分别表示真阳性、假阳性、假阴性值。

3) pFM

$$pFM = \frac{2 \times pRecall \times Precision}{(pRecall + Precision)} \quad (11)$$

其中, $pRecall$ 表示生成的结果图像相比于标准二值化图像检测到文字信息完整性的百分比。

4) DRD

$$DRD = \frac{\sum_q DRD_q}{NUBN} \quad (12)$$

其中, DRD 用于测量二值图像中的视觉失真。 $NUBN$ 是标准二值化结果图像中非均匀 (并非所有黑色或白色像素) 8×8 像素块的数量, DRD_q 表示在 5×5 像素块邻域内, 标准二值化图像与生成的结果图像第 q 个翻转像素之间的权重加和的差值。

2.1.3 参数敏感性分析

为了对比感知损失函数的权重系数 ω 取值不同对文档图像透射去除效果的影响, 本节以合成透射数据集为例, 结合 FM 和 pFM 评价指标, 通过设置不同的权重系数 ω 进行参数敏感性分析实验。

不同的权重系数 ω 对 FM 和 pFM 评价指标的影响如图 7 所示, 可见当感知损失函数的权重系数 ω 越小时, FM 和 pFM 评价指标的数值越小, 当 ω 在 0.8 附近时, FM 和 pFM 的值达到最大, 此时文档图像透射去除效果最好, 所以本文设定 $\omega = 0.8$ 。

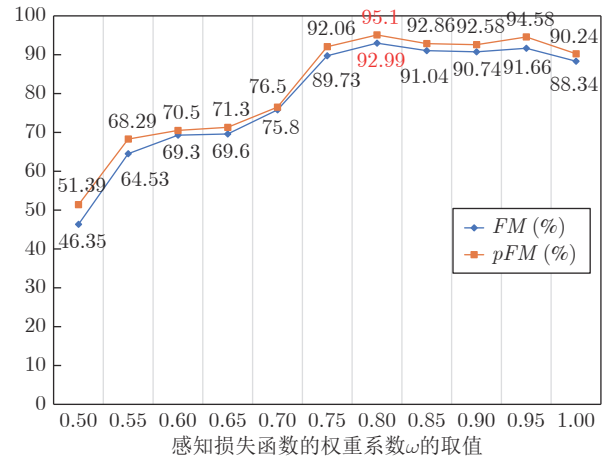


图 7 不同权重系数 ω 对 FM 和 pFM 评价指标的影响

Fig. 7 Influence of different weight coefficient ω for FM and pFM

2.2 DIBCO 数据集实验结果与分析

表 1 为 3 个 DIBCO 系列数据集的定量评价结果 (获得的最佳结果用粗体标出)。与已有的方法相比, 本文的 S-CycleGAN 在 4 个评价指标中均有 3 项取得了最佳结果。

为了直观展示视觉对比效果, 图 8 给出了不同

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

数据集	方法	PSNR (dB)	FM (%)	pFM (%)	DRD
DIBCO 2011	Otsu	18.52	67.81	74.08	17.45
	Ntirogiannis	16.34	85.34	86.06	8.18
	SAGAN	21.55	87.47	92.28	5.97
	DD-GAN	22.67	88.90	91.19	5.53
	Castellanos	22.95	89.40	91.78	5.62
	Sungho	23.54	89.67	91.03	5.59
DIBCO 2009	S-CycleGAN	24.36	89.71	91.62	5.49
	Otsu	14.90	72.57	73.51	23.67
	Ntirogiannis	14.30	84.60	88.40	6.34
	SAGAN	19.64	89.75	90.85	6.35
	DD-GAN	21.14	92.53	92.59	4.86
	Castellanos	21.95	90.00	91.68	6.03
DIBCO 2016	Sungho	22.56	87.73	92.09	5.35
	S-CycleGAN	22.83	90.98	92.65	4.54
	Otsu	15.52	70.44	73.03	20.45
	Ntirogiannis	18.14	83.72	87.49	10.98
	SAGAN	20.35	91.64	92.71	5.64
	DD-GAN	21.54	90.48	93.63	3.17
DIBCO 2011	Castellanos	22.30	91.13	92.28	3.05
	Sungho	21.96	90.27	92.69	2.63
	S-CycleGAN	22.35	91.90	93.79	3.53

方法在 DIBCO 2011^[25] 数据集中的一个样本图像的处理结果. 如图所示, 图 8(a) 中大面积背面内容叠加到正面图像中且右上部分透射内容较重 (强透射区域). 因直方图没有明显的双峰特征, 且图像整体亮度偏低, Otsu^[4] 算法计算出的阈值较小, 从而导致计算输出的二值图像引入了大量噪声. Ntirogiannis^[5] 方法虽然能抑制一定的噪声, 但也丢失了部分文本内容. 基于深度学习的方法 SAGAN^[28]、DD-GAN^[15]、Castellanos^[16] 和 Sungho^[29] 方法的恢复结果要优于经典方法, 对文字的恢复较为准确, 但仍有较多背面内容在恢复过程中被当作正面的内容, 导致恢复结果出现了噪声误差. 本文方法 S-CycleGAN 给出了较好的视觉效果, 更加接近真值图像 (Ground truth), 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.

图 9 给出了不同方法对 H-DIBCO 2016^[27] 数据集中的一个样本图像的恢复结果. Otsu^[4]、DD-GAN^[15] 和 Sungho^[29] 的实验结果中存在大量噪声信息. Ntirogiannis^[5] 方法可以更有效地去除阴影和类似正面文本的背面噪声, 但存在文本信息错分的情况. SAGAN^[28] 和 Castellanos^[16] 方法恢复的实验结

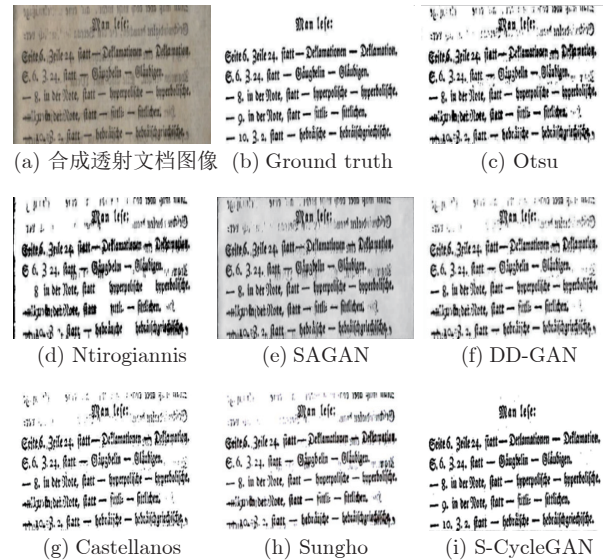


图 8 各方法在 DIBCO 2011 数据集内一个样本的恢复结果
Fig. 8 Experiment results of one sample in DIBCO 2011 datasets by different methods



图 9 各方法在 H-DIBCO 2016 数据集的一个样本恢复结果
Fig. 9 Experiment results of one sample in H-DIBCO 2016 datasets by different methods

果中存在文字笔画不连续的现象. S-CycleGAN 不仅能更好地去除阴影和噪声, 而且更好地保留了正面文本细节.

2.3 合成数据集和真实图像实验结果与分析

2.3.1 合成数据集实验结果

本文方法 S-CycleGAN 与其他几种去除透射

算法在合成数据集上的定量评价如表 2 所示. 粗体为恢复方法中最佳的指标结果, 显然可见, S-CycleGAN 在四个客观度量指标上均优于其他去除透射方法.

表 2 合成数据集的文档透射图像恢复定量评价

Table 2 Quantitative evaluation of document bleed-through image restoration of synthetic datasets

数据集	方法	PSNR (dB)	FM (%)	pFM (%)	DRD
合成数据集	Otsu	16.35	88.37	89.59	4.94
	Ntirogiannis	19.30	89.21	90.68	8.87
	SAGAN	16.05	87.61	91.28	5.21
	DD-GAN	20.45	90.51	90.01	4.73
	Castellanos	19.95	90.65	93.78	4.06
	Sunggho	21.03	90.53	92.67	3.86
	S-CycleGAN	22.66	92.99	95.10	2.93

合成数据集中一个样本图像的恢复结果对比如图 10 所示, 可见经典的 Otsu^[4] 和 Ntirogiannis^[5] 方法无法消除文档的背面透射内容, 这是因为此类方法难以在正面和背面内容区分中找到一个有效阈值, 且 Ntirogiannis^[5] 方法恢复的内容中标点符号严重缺失. Castellanos^[16] 正面的文档内容没有完全恢复且造成了一定的内容细节损失, 恢复效果不理想. DD-GAN^[15] 和 Sunggho^[29] 方法产生了相对较好的视觉效果, 但存在纹理细节不清晰等问题. 综合来看, S-CycleGAN 生成的结果更为自然, 纹理细节也较为清晰.

2.3.2 真实透射图像实验结果

真实图像为“京瓷 FS-1020MFP”扫描获取的全国大学英语六级试卷和某初中数学试卷的数字图像, 截取其中有明显透射的样本进行恢复, 各种对比方法的实验结果分别如图 11 和图 12 所示.

由视觉效果综合比较可见, Ntirogiannis^[5] 和 Sunggho^[29] 方法的恢复结果中存在部分内容缺失的问题, 例如图 12(g) 中“-6”的“-”丢失, 恢复的图像与原始文档内容不一致. Otsu^[4] 和 DD-GAN^[15] 方法可以将透射内容进行一定程度的去除, 但存在恢复正面字迹不清晰、边缘不够平滑的问题. Castellanos^[16] 和 SAGAN^[28] 方法保留了正面文本内容但仍存在一定的背面噪声. 与其他方法相比, 本文方法 S-CycleGAN 不仅能更好地去除透射内容和噪声, 而且清晰、平滑、正确的保留了正面文本内容.

2.4 消融实验

第 3.2 节和第 3.3 节的对比实验结果表明, 基于 S-CycleGAN 的图像恢复方法取得了良好的性



图 10 各方法在合成数据集的一个样本恢复结果

Fig. 10 Experiment results of one sample on synthetic document bleed-through datasets by different methods

能. 为了更好地理解该方法中不同组成部分的作用, 测试每个组成部分在网络中的重要性, 进行了有无特定成分的消融实验, 以验证每个模块在 S-CycleGAN 的功能. 消融实验在合成数据集和 H-DIBCO 2016^[27] 数据集上进行, 主要对比: CycleGAN、无 SLM 的 S-CycleGAN 和本文方法 S-CycleGAN.

表 3 给出了 S-CycleGAN 及其两种消融变体在合成数据集和 H-DIBCO 2016 数据集上的客观度量指标计算结果. 由表 3 可见, 加入了本文设计的功能模块后的 S-CycleGAN 的实验指标得到较大提升, 结果最优, 且部分指标相差较大. 图 13 给出了 S-CycleGAN 及其两种消融变体在合成数据集中一个样本图像的恢复结果. 可以看出网络中加入自学习模块能够关注到图像的不同区域特征, 以“非均匀”的方式处理透射到正面的“深浅不一”内容, 更好的恢复出了正面图像内容的纹理、边缘等信息, 不仅有效地去除了透射内容, 还能完整保留正面图像的文字细节. 从表 3 的客观指标评价和图 13 的实验结果来看, 在 CycleGAN 中加入了特征提取模块和自学习模块对透射文档图像的恢复十分有效.

3 结论

本文提出了一种自监督学习的单幅透射图像恢复网络 (S-CycleGAN), 该网络可以直接对非配对

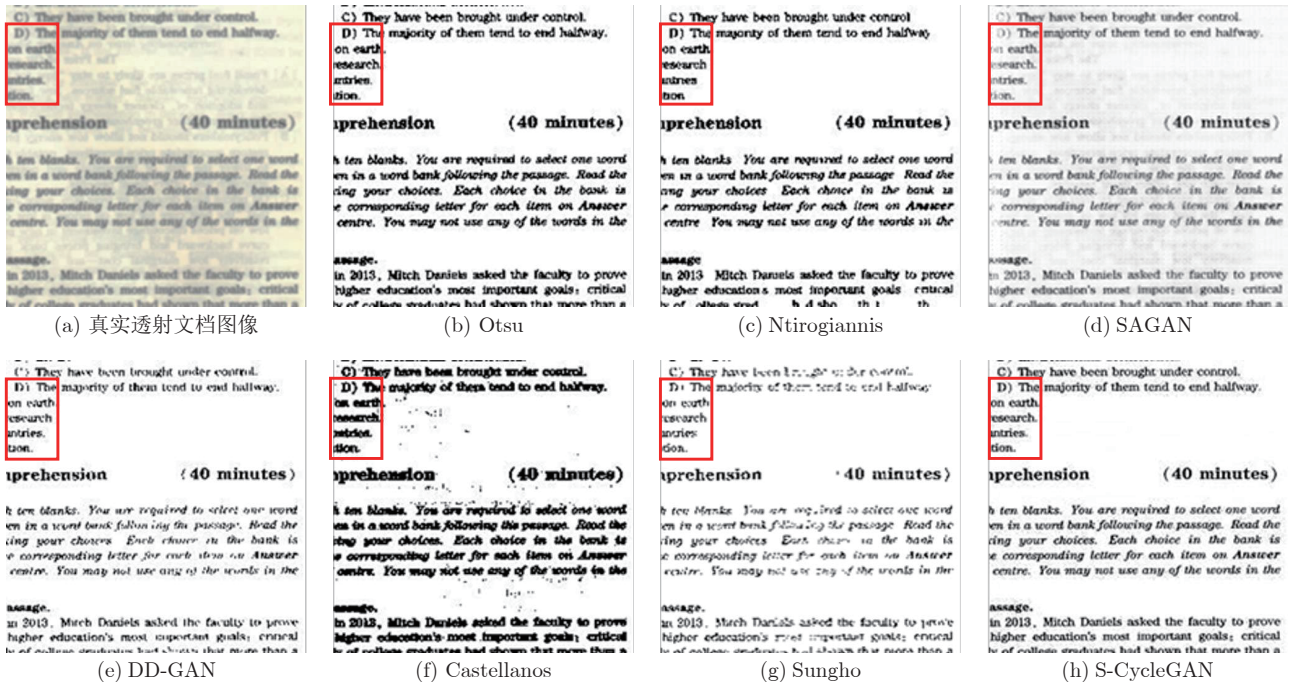


图 11 不同方法在全国大学英语六级试卷透射图像的恢复结果
 Fig. 11 Experiment results of CET-6 bleed-through datasets by different methods

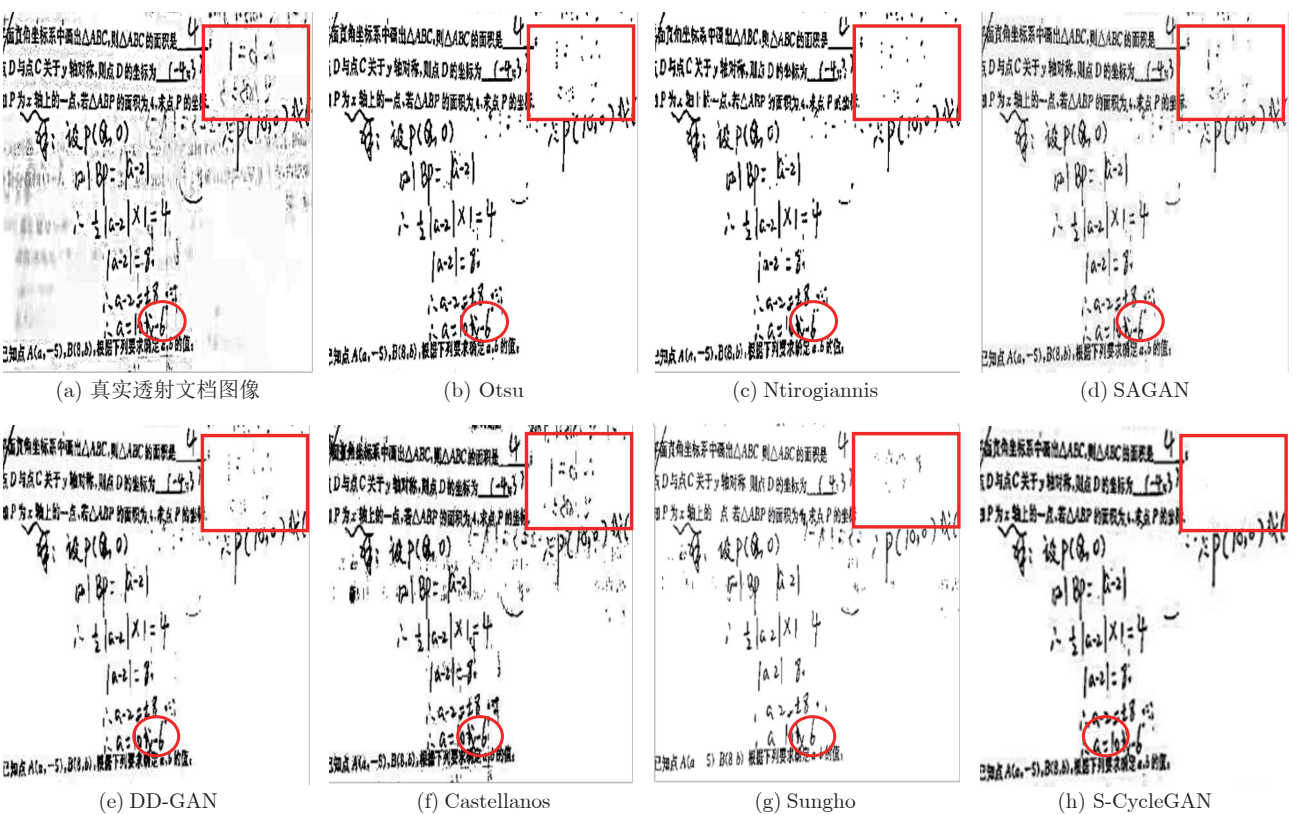


图 12 不同方法在某初中数学试卷透射图像上的恢复结果
 Fig. 12 Experiment results of test papers bleed-through datasets by different methods

表 3 S-CycleGAN 模块有效性客观评价指标对比
Table 3 Objective evaluation indexes comparison for the modules in S-CycleGAN

数据集	方法	PSNR (dB)	FM (%)	pFM (%)	DRD
合成数据集	CycleGAN	12.48	62.42	65.51	20.95
	无 SLM	19.75	88.80	92.50	3.95
	S-CycleGAN	22.66	92.99	95.10	2.93
H-DIBCO 2016	CycleGAN	11.41	69.71	71.33	16.31
	无 SLM	18.21	86.60	88.80	4.36
	S-CycleGAN	22.35	91.90	93.79	3.53

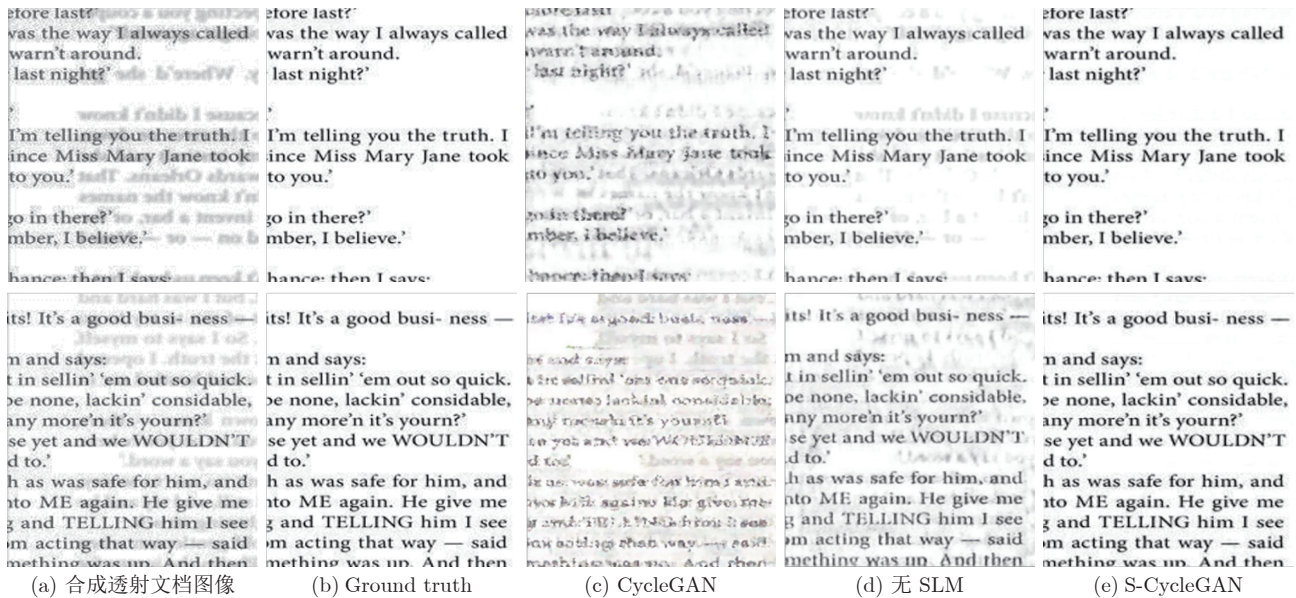


图 13 不同网络结构在合成数据集上的消融实验

Fig. 13 Ablation experiments of different network structures on synthetic datasets

的透射文档图像执行图像恢复任务, 且无需混合比例、阈值等任何先验参数. 通过设计有效的生成器网络, 针对文档图像中透射分布不均匀、文字印记深浅不一等问题, 在网络中加入自学习模块, 提高关注点的表示和感兴趣内容的表现力, 以获得更好的透射图像恢复效果和重建细节内容. 在合成数据集、DIBCO 数据集和真实图像上与现有方法进行实验对比, 结果表明 S-CycleGAN 在客观度量指标和视觉效果上均取得了较好的结果, 有望集成于扫描仪、相机等实际成像设备.

References

- 1 Rasyidi H, Khan S. Historical document image binarization via style augmentation and atrous convolutions. *Neural Computing and Applications*, 2020, **33**: 7339–7352
- 2 Gupta N, Goyal N. Machine learning tensor flow based platform for recognition of hand written text. In: Proceedings of the International Conference on Computer Communication and Informatics. Coimbatore, India: IEEE, 2021. 1–6
- 3 Han Y H, Wang W L, Liu H M, Wang Y Q. A combined approach for the binarization of historical tibetan document images. *International Journal of Pattern Recognition and Artificial Intelligence*, 2019, **33**(14): 1954038
- 4 Otsu N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems Man Cybernetics-Systems*, 2007, **9**(1): 62–66
- 5 Ntirogiannis K, Gatos B, Pratikakis I. Performance evaluation methodology for historical document image binarization. *IEEE Transactions on Image Processing*, 2013, **22**(2): 595–609
- 6 Su B, Lu S, Tan C L. Binarization of historical document images using the local maximum and minimum. In: Proceedings of the International Work-shop on Document Analysis Systems. Boston, USA: Work-shop on Document Analysis Systems, 2010. 154–160
- 7 Tensmeyer C, Martinez T. Document image binarization with fully convolutional neural networks. In: Proceedings of the International Conference on Document Analysis and Recognition. Kyoto, Japan: IEEE, 2017. 99–104
- 8 Wu Y, Rawls S, Abdalimageed W, Natarajan P. Learning document image binarization from data. In: Proceedings of the IEEE International Conference on Image Processing. Phoenix, USA: IEEE, 2016. 3763–3767
- 9 He S, Schomaker L. Deepotsu: document enhancement and binarization using iterative deep learning. *Pattern Recognition*, 2019, **91**: 379–390
- 10 Kang S, Iwana B K, Uchida S. Complex image processing with less data-document image binarization by integrating multiple

- pre-trained u-net modules. *Pattern Recognition*, 2020, **109**: 107577
- 11 Mondal R, Chakraborty D, Chanda B. Learning 2d morphological network for old document image binarization. In: Proceedings of the International Conference on Document Analysis and Recognition. Sydney, Australia: IEEE, 2019. 65–70
 - 12 Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *Advances in Neural Information Processing Systems*, 2014, **3**: 2672–2680
 - 13 Reed S, Akata Z, Mohan S, Tenka S, Schiele B, Lee H. Learning what and where to draw. In: Proceedings of the Neural Information Processing Systems. Barcelona, Spain: Curran Associates, 2016. 217–225
 - 14 Konwer A, Bhunia A K, Bhowmick A, et al. Staff line removal using generative adversarial networks. In: Proceedings of the International Conference on Pattern Recognition. Beijing, China: IEEE, 2018. 1103–1108
 - 15 De R, Chakraborty A, Sarkar R. Document image binarization using dual discriminator generative adversarial networks. *IEEE Signal Processing Letters*, 2020, **27**: 1090–1094
 - 16 Castellanos F J, Gallego A J, Jorge C Z. Unsupervised neural domain adaptation for document image binarization. *Pattern Recognition*, 2020, **119**: 108099
 - 17 Zhu J Y, Park T, Isola P. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 2223–2232
 - 18 Sajjadi M, Scholkopf B, Hirsch M. EnhanceNet: single image super-resolution through automated texture synthesis. In: Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 4501–4510
 - 19 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the International Conference on Learning Representations. California, USA, 2015. 1–14
 - 20 Jia D, Dong W, Socher R, Li L J, Kai L, Li F F. Imagenet: A large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 248–255
 - 21 Pratikakis I, Gatos B, Ntirogiannis K. Icdar 2013 document image binarization contest. In: Proceedings of the International Conference on Document Analysis and Recognition. Washington, USA: IEEE, 2013. 1471–1476
 - 22 Pratikakis I, Gatos B, Ntirogiannis K. Icfhr 2012 competition on handwritten document image binarization. In: Proceedings of the International Conference on Frontiers in Handwriting Recognition. Bari, Italy: IEEE, 2012. 817–822
 - 23 Ntirogiannis K, Gatos B, Pratikakis I. Icfhr 2014 competition on handwritten document image binarization. In: Proceedings of the International Conference on Frontiers in Handwriting Recognition. Hersonissos, Greece: IEEE, 2014. 809–813
 - 24 Pratikakis I, Zagoris K, Barlas G, Gatos B. Icdar2017 competition on document image binarization. In: Proceedings of the International Conference on Document Analysis and Recognition. Kyoto, Japan: IEEE, 2017. 2379–2140
 - 25 Pratikakis I, Gatos B, Ntirogiannis K. Icdar 2011 document image binarization contest. In: Proceedings of the International Conference on Document Analysis and Recognition. Beijing, China: IEEE, 2011. 1506–1510
 - 26 Gatos B, Ntirogiannis K, Pratikakis I. Icdar 2009 document image binarization contest. In: Proceedings of the International Conference on Document Analysis and Recognition. Barcelona, Spain: IEEE, 2009. 1375–1382
 - 27 Pratikakis I, Zagoris K, Barlas G, Gatos B. Icfhr 2016 Hand-

written document image binarization contest. In: Proceedings of the International Conference on Frontiers in Handwriting Recognition. Shenzhen, China: IEEE, 2016. 2167–6445

- 28 Zhang X, Goodfellow I, Metaxas D, Odena A. Self-attention generative adversarial networks. In: Proceedings of the International Conference on Machine Learning. California, USA, 2019. 7354–7363

- 29 Suh S, Kim J, Lukowicz P, Lee Y O. Two-stage generative adversarial networks for document image binarization with color noise and background removal. 2020, arXiv: 2010.10103



徐金东 烟台大学计算机与控制工程学院教授。主要研究方向为盲源分离, 图像处理和模式识别。本文通信作者。
E-mail: xujindong@ytu.edu.cn
(**XU Jin-Dong** Professor at the School of Computer and Control Engineering, Yantai University. His

research interest covers blind source separation, image processing and pattern recognition. Corresponding author of this paper.)

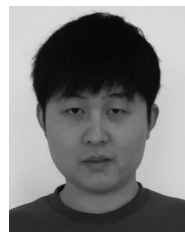


马咏莉 烟台大学计算机与控制工程学院硕士研究生。主要研究方向为盲源分离和图像处理。

E-mail: mayonglim@163.com

(**MA Yong-Li** Master student at the School of Computer and Control Engineering, Yantai University.

Her research interest covers blind source separation and image processing.)

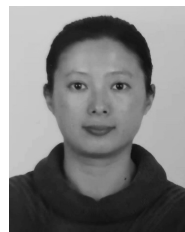


梁宗宝 烟台大学计算机与控制工程学院硕士研究生。主要研究方向为盲源分离和图像处理。

E-mail: liangzongbao_ytu@163.com

(**LIANG Zong-Bao** Master student at the School of Computer and Control Engineering, Yantai Uni-

versity. His research interest covers blind source separation and image processing.)



倪梦莹 烟台大学物理与电子信息学院讲师。主要研究方向为信号处理和模式识别。

E-mail: nimengying@ytu.edu.cn

(**NI Meng-Ying** Lecturer at the School of Physics and Electronic Information, Yantai University. Her

research interest covers signal processing and pattern recognition.)