

基于生成对抗门控卷积网络的文档图像印章消除

伍贵宾, 杨宗元, 熊永平, 张兴, 王伟

引用本文

伍贵宾, 杨宗元, 熊永平, 张兴, 王伟. 基于生成对抗门控卷积网络的文档图像印章消除[J]. 计算机科学, 2024, 51(1): 198-206.

WU Guibin, YANG Zongyuan, XIONG Yongping, ZHANG Xing, WANG Wei. Seal Removal Based on Generative Adversarial Gated Convolutional Network [J]. Computer Science, 2024, 51(1): 198-206.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

基于生成式对抗网络和正类无标签学习的知识图谱补全算法

Knowledge Graph Completion Algorithm Based on Generative Adversarial Network and Positive and Unlabeled Learning

计算机科学, 2024, 51(1): 310-315. <https://doi.org/10.11896/jsjcx.230300006>

基于多尺度注意力机制的两阶段文物图像修复方法

Two-stage Method for Restoration of Heritage Images Based on Multi-scale Attention Mechanism

计算机科学, 2023, 50(6A): 220600129-8. <https://doi.org/10.11896/jsjcx.220600129>

基于激光雷达点云的3D目标检测方法综述

Review of 3D Target Detection Methods Based on LiDAR Point Clouds

计算机科学, 2023, 50(6A): 220400214-7. <https://doi.org/10.11896/jsjcx.220400214>

基于空频联合卷积神经网络的GAN生成人脸检测

GAN-generated Face Detection Based on Space-Frequency Convolutional Neural Network

计算机科学, 2023, 50(6): 216-224. <https://doi.org/10.11896/jsjcx.220400268>

一种基于强化学习的口令猜解模型

Password Guessing Model Based on Reinforcement Learning

计算机科学, 2023, 50(1): 334-341. <https://doi.org/10.11896/jsjcx.211100001>

基于生成对抗门控卷积网络的文档图像印章消除

伍贵宾¹ 杨宗元¹ 熊永平¹ 张兴² 王伟²

1 北京邮电大学计算机学院(国家示范性软件学院) 北京 100876

2 华润数科人工智能实验室 广东 深圳 518049

(wuguibin@bupt.edu.cn)

摘要 发票和文档上的印章严重影响文字识别的准确率,因此印章消除技术在文档识别和文档增强的预处理过程中发挥着重要作用。然而,现有的阈值分割方法和基于深度学习的方法存在印章消除不全以及会修改背景像素等问题。文中提出了一个两阶段式印章消除网络 SealErase。第一阶段是一个用于生成包含印章位置信息的二值化掩膜的 U 型分割网络,第二阶段是一个用于进行精细化印章消除的修复网络。由于目前缺乏公开的用于印章消除的成对数据集,现有的方法无法设计像素级的评价指标来衡量生成图像的质量。并且,利用配对的训练集训练神经网络可以有效提高网络的性能。为此,文中兼顾真实场景的泛化性以及噪声的鲁棒性构建了一个包含 8000 个样本的高仿真的印章消除数据集。其中的印章分为两种:真实文档图像中的印章和合成的印章。为了客观地评价 SealErase 的性能,文中设计了基于图像生成质量和被印章遮盖的字符识别准确率的综合评价指标用于评估 SealErase 网络的消除性能。在构建的印章消除数据集上对比了现有的印章消除模型,实验结果表明,SealErase 网络在图像生成质量的评价指标中的峰值信噪比相比最先进的方法提升了 26.79%,平均结构相似性指标提升了 4.48%。经过 SealErase 网络进行印章消除后,被印章遮盖的字符识别准确率提高了 38.86%。SealErase 在真实场景下同样可以有效消除印章并保留被遮盖的文字。

关键词: 印章消除;图像修复;印章生成;生成式对抗网络;门控卷积;SealErase

中图分类号 TP391

Seal Removal Based on Generative Adversarial Gated Convolutional Network

WU Guibin¹, YANG Zongyuan¹, XIONG Yongping¹, ZHANG Xing² and WANG Wei²

1 School of Computer Science(National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing 100876, China

2 Artificial Intelligence Lab of China Resources Digital Technology, Shenzhen, Guangdong 518049, China

Abstract Seals on invoices and documents seriously affect the accuracy of text recognition, so seal elimination techniques play an important role in the pre-processing of document analysis, and document enhancement. However, threshold-based methods and deep learning-based methods suffer from incomplete seal elimination and modification of background pixels. Thus, this paper proposes a two-stage seal elimination network, SealErase. The first stage is a U-shaped segmentation network for generating binarized masks with seal position, and the second stage is an inpainting network for refined seal elimination. Due to the lack of available public paired datasets for seal elimination, existing methods cannot design pixel-level evaluation metrics to measure the quality of the generated images. Moreover, training the neural network using paired training sets can effectively improve the performance of the network. To this end, this paper constructs a high-simulated seal elimination dataset containing 8000 samples, taking into account the generalisation to real scenes and the robustness to noise. The seals are divided into two types: seals in real document images and synthetic seals. In order to objectively evaluate the performance of SealErase, it devises a comprehensive evaluation metric based on the image generation quality and the recognition accuracy of characters obscured by seals to evaluate the elimination performance of the SealErase network. The existing seal elimination methods are compared on the seal elimination dataset, and the experimental results show that the SealErase network improve the peak signal to noise ratio by 26.79% and the mean structural similarity by 4.48% in the evaluation metric of image generation quality compared to the state-of-the-art methods. After seal elimination by SealErase network, the accuracy of recognition of characters obscured by seals is improved by 38.86%. Ex-

到稿日期:2023-05-31 返修日期:2023-09-15

基金项目:国家电网公司总部科技项目(5500-202216134A-1-1-ZN)

This work was supported by the Research on Key Technology of Multi-spectral Optical Imaging inside GIS based on Fiber Bundle Image Transmission(5500-202216134A-1-1-ZN).

通信作者:熊永平(ypxiong@bupt.edu.cn)

perimental results show that SealErase is equally effective in eliminating seals and preserving the obscured characters in real scenes.

Keywords Seal removal, Image inpainting, Seal synthesis, Generative adversarial networks, Gated Convolutions, SealErase

1 引言

光学字符识别(Optical Character Recognition, OCR)被广泛应用于文档识别、票据识别、手写文字识别等领域^[1]。然而,文档或票据上的印章(见图1)严重影响印章下文字的OCR准确率。通过技术手段去除文档或票据上的印章,可以有效提高文字的识别准确率,具有重要的应用价值。印章消除需要解决两个核心问题:印章定位与印章擦除。印章定位要求模型能够以像素级别定位到印章位置;印章擦除要求模型能够根据印章位置对印章进行擦除并最大程度恢复背景颜色与纹理,且需保留被印章遮盖的字符。



图1 真实的印章示例

Fig.1 Examples of real seals

目前,印章消除方法主要分为阈值分割方法和深度学习方法。文献[2]在HSI颜色空间中找到合适阈值将特定颜色区域提取出来,接着通过二值化、填充、去噪,将印章分割出来。文献[3]提出了一种二次分割的算法来去除印章,首先根据印章颜色进行一次粗分割,再对印章和印章下的字符进行二次分割。阈值分割方法在实际应用存在诸多问题,譬如当印章与背景颜色接近时,无法找到合理的阈值。在深度学习方法中,文献[4]提出了一种基于CycleGAN^[5]的无监督模型SealGAN。由于缺少成对的原图标注数据,SealGAN在消除过程中存在修改印章区域外像素以及印章消除不全等问题。文献[6]提出了一种基于Pix2Pix^[7]的印章消除模型,通过从已有样本图像中利用阈值分割方法提取印章来合成印章数据集,这种合成方法受到印章颜色、形状等诸多限制,导致该方法缺乏泛化性。因此我们提出了一种两阶段式的印章消除网络SealErase。第一阶段是一个U型分割网络,用于生成包含印章位置信息的二值化掩膜(Mask),第二阶段是一个基于生成式对抗网络和门控卷积的消除网络,用于进行精细化印章消除。SealErase第一阶段网络主要解决印章定位问题,而第二阶段网络主要解决印章擦除问题。

由于目前缺乏公开的用于印章消除的成对数据集,现有的方法^[3-4]无法设计像素级的评价指标来衡量生成图像的质量。并且,在神经网络训练时,利用配对的训练集对生成的图片进行监督可以有效提高网络的性能^[7]。因此,构建一个用于印章消除的配对的数据集是必要的。为此,我们构建了一个包含8000个样本的高仿真的印章消除数据集。印章消除数据集中的印章分为两种:真实文档图像中的印章和合成的印章。为了提升模型在真实应用场景的泛化性,本文利用图像编辑软件手工提取真实图像中的印章用于样本生成。为了

提升模型对噪声的鲁棒性,本文设计了一个印章合成方法来得到更多样本用于印章消除。最后,本文设计了基于图像生成质量的评价指标在像素级别上客观评价SealErase生成图像的质量及印章消除的性能。为了验证印章消除对文本识别的影响,我们也利用OCR字符识别准确率来评价SealErase的印章消除性能。

本文在构建的印章消除数据集上进行了充分的实验,结果表明,与Pix2Pix模型^[7]、CycleGAN模型^[5]以及SealGAN模型^[4]相比,SealErase网络在图像生成质量的评价指标中的峰值信噪比(Peak Signal to Noise Ratio, PSNR)提升了26.79%,平均结构相似性^[8](Mean Structural Similarity, MSSIM)指标提升了4.48%。经过SealErase网络进行印章消除后,被印章遮盖的字符识别准确率提高了38.86%。实验结果表明,SealErase在真实场景下同样可以有效消除印章并保留被遮盖的文字。

2 相关工作

2.1 文档增强

由于照明条件差、阴影、噪声、模糊、失真、老化、墨渍、渗透、水印、印章等,文档图像可能会遭受各种程度的退化(Degradation)或损坏。文档增强是一种将退化的文档图像进行恢复的技术手段,在许多文档分析和识别任务的预处理步骤中,文档增强起着重要的作用。文档增强主要包括以下几个任务^[9]:二值化(Binarization)、去模糊(Deblurring)、去噪(Denoising)、去扭曲和倾斜(Defading)、水印去除(Watermark Removal)和阴影去除(Shadow Removal)。

文献[10]针对相机对焦失败和相机抖动造成文档图像模糊的问题,提出了一种端到端的小体量的卷积神经网络。文献[11]提出了一种端到端的文档增强系统,通过输入退化的文档图像来产生干净的文档图像,其在REDNET^[12]的基础上,重新设计了一个包含15层卷积层和15层反卷积层以及8个在对应卷积层间的跳跃连接(Skip-connection)的自编码器网络结构。文献[13]提出了一种基于条件生成对抗网络^[14](Conditional Generative Adversarial Nets, cGANs)的端到端的文档增强网络DE-GAN,首次将生成对抗网络用于二值化任务。Jemni等^[15]设计了一个基于生成对抗网络的端到端的网络结构来恢复受损的图像,他们设计了一个手写文字识别器,首次将文档中的文本信息作为额外的监督信息用于文档增强任务中。文中对受损的阿拉伯文和拉丁文文档图像进行修复,结果表明该模型提高了受损文档的图像质量和可读性。

印章消除任务一般被认为是一个能够优化文档图像质量和可读性的关键技术。从这个角度出发,印章可以被视为噪声,其会对文本识别和自动化处理产生干扰。因此,在文档增强的框架下,印章消除被视为一项至关重要的子任务。然而,尽管印章消除在文档增强领域中具有很高的实用价值,但是由于印章更多地出现在中文文档图像中,同时又存在着缺乏

相关数据集和标注的不足,因此,印章消除的研究鲜有突破。目前已有一些研究在对英文文档图像的印章消除中取得了进展,但是与中文文档图像相比,这些研究还有一定的局限性。从印章本身的特点出发,印章消除可以被理解为一类特殊的水印消除任务。然而,与一般的水印特征相比,印章的特征更加复杂。印章消除时需要考虑到印章的大小、形状、颜色、模糊度、褪色、重叠、变形等多种特征,印章的复杂特征也决定了印章去除任务比水印去除任务更具有挑战性。

2.2 图像修复

图像修复需要根据图像已有部分去修复图像缺失部分^[16]。其中基于深度学习的图像修复方法可分为如下两类^[17]:基于卷积神经网络的图像修复方法和基于生成对抗网络的图像修复方法。

基于卷积神经网络的图像修复方法主要针对卷积的方式进行改进。卷积神经网络中的普通卷积(Vanilla Convolutions)会将图像中的所有输入像素都当作有效值进行计算,由于不区分有效像素和无效像素,最终生成的图片往往会产生大量的视觉伪影。为此,文献[18]提出利用部分卷积(Partial Convolutions)代替普通卷积。部分卷积在每次卷积后根据制定的规则进行一次掩膜更新来区分破损区域内的像素。该方法实现了将破损区域内的像素区别对待,可以有效解决视觉伪影等问题。但是,通过设定规则来进行掩膜更新仍然存在不合理性。例如,在掩膜更新过程中,部分卷积将含有不同数量的有效像素的区域同等对待,且无效像素将在深层逐渐消失^[19]。为了解决掩膜更新不合理的问题,文献[19]对卷积再次进行修改,提出了门控卷积(Gated Convolutions)。该方法根据输入的掩膜,在不同通道的特征图的不同空间位置乘以一个权重,用来区分前景像素和背景像素,因此可以更有效地解决修复过程中的掩膜更新问题,从而提升修复效果,并且可以在特定通道输入人为控制的掩膜,来控制修复的效果。

为了生成更精细的图像,基于生成对抗网络的图像修复模型通常采用粗制-精炼的两级模型。文献[20]提出了深度卷积生成对抗网络(Deep Convolutional Generative Adversarial Networks, DCGAN),将其将卷积神经网络和原始的生成对抗网络结合起来用于图像的生成。文献[21]将上下文编码器(Encoder-Decoder)与生成对抗网络结合起来用于图像修复。文献[22]在网络中引入空洞卷积(Dilated Convolution)并提出了一个上下文注意力机制(Contextual Attention)。空洞卷积可以扩大感受野的大小且不产生额外的开销;上下文注意力机制可以从离修复区域较远的区域中提取所需的特征,进一步弥补卷积存在局限性的不足。文献[23]提出利用小波分解将图像分解成多个频段。在低频段应用感知损失,在高频段应用对抗损失,以保证空间域与频域修复的一致性。

在已知印章位置信息的前提下,我们希望设计一个能够充分提取图像特征并专精于指定位置图像转换的模型来完成

印章擦除任务。由于擦除印章需要删除纹理并恢复相应区域的内容,因此需要一个能够同时考虑图像特征和上下文信息的模型。目前,针对图像修复任务的许多模型已经取得了较好的结果,这些模型都可以通过图像特征提取和转换来实现图像修复。在擦除印章任务中,我们也可以借鉴这些模型的优点,将其适配到擦除印章任务中。然而,印章擦除涉及的输入数据与一般图像修复略有不同。对于印章擦除任务,输入数据需要包含印章的准确位置信息,但是输入的掩膜可能不完全准确。此外,输入的图像也不能是已经被掩膜处理过的图像,因为这样无法还原被遮盖的字符。

2.3 自然场景文字消除

自然场景文字消除需要自动从自然场景图片中定位到文本位置,并进行擦除,其被广泛应用于个人隐私保护和广告植入等场景。文献[24]利用滑动窗口的方法从原图中进行裁剪,接着通过U型卷积网络对文本进行擦除。文献[25]提出了一个端到端的自然场景文本擦除模型EnsNet。EnsNet训练时使用了4个不同尺度的损失函数,用于更精细化的擦除文本并恢复背景。但由于模型没有关于文本位置的额外信息,EnsNet存在文本擦除不全等问题。为此,文献[26]提出了一个端到端的网络EraseNet,它在上采样的过程中额外引入一个分支,用于生成包含文本位置信息的掩膜。由于缺少真实的原图标注数据集,文献[26]还提出了一个自然场景文字消除的基准数据集SCUT-EnsText,用于训练和测试EraseNet。由于EraseNet是端到端模型,因此存在无法定制化消除文本的问题。

印章消除与自然场景文字消除在任务上有很多相似之处,譬如模型都需要定位到被擦出的位置并最大程度恢复背景下的复杂纹理。在提前知道自然场景文字位置的基础上,图像修复算法^[19]也能较好地恢复文本区域的背景^[26],因为在自然场景文本下的背景往往都是纹理一致的。但是对于印章消除,印章区域下的背景往往包含文字、表格线,如果简单应用图像修复算法,将丢失大量无法恢复的信息。

3 印章消除数据集

构建用于印章消除的成对数据集的核心是获得带有印章的文档图片和与之对应的无印章文档图片。为此,我们将构建印章消除数据集分为3个步骤:数据收集、印章获取、样本生成。第一步,收集足够多的真实场景下的多种文档图像作为背景图片,其中包含带有多种形状的印章的文档图像。第二步,利用图像编辑软件在像素级别上对真实印章文档图像进行印章提取,并设计一个高仿真的印章合成方法来丰富数据集中印章的多样性。第三步,将第一步收集的背景图片和第二步获取的印章掩膜进行融合,生成一个包含8000个样本的印章消除数据集。表1列出了本文提出的印章数据集的构成。

表 1 印章数据集构成
Table 1 Composition of seal datasets

数据集	背景类型 / 张					印章类型 / 张				
	发票	公文	普通文档	单据	合计	圆形	椭圆形	矩形	菱形	合计
印章数据集	450	160	160	330	1 100	3 250	2 650	1 650	450	8 000



图2 标注示例
Fig. 2 Examples of manual annotation

3.1 数据收集

构建印章消除数据集的一个关键步骤是收集背景图片和带有印章的文档图片。由于印章的非公开性,因此我们构建的印章数据集的背景图片和印章图片均来自我们内部的图片。对于背景图片和真实印章图片的选取,我们充分考虑了文档背景的多样性以及印章特征的多样性。

1)背景图片的多样性:我们主要从场景的多样性(自然拍摄场景的文档图片和扫描文档图片)和文档种类的多样性(发票、票据、办公文档等)这两个方面考虑背景图片的多样性。首先,在自然拍摄场景中,文档图片通常由手机等设备拍摄,这些文档图片往往包含文档扭曲、文档污损、文档外的背景噪声等不同的噪声。而在扫描场景下,文档通常具有较好的可读性。在实际应用中,可能遇到的带有印章的文档种类是多变的。为了提高泛化性,我们收集了足够多的不同种类的文档图片用于背景融合,如图3所示。



图3 印章数据集中不同种类的文档图片
Fig. 3 Different kinds of document images of seal datasets

2)真实印章的多样性:因为印章的特征提取的好坏直接影响着印章分割的性能,所以本文的印章消除数据集考虑了现有印章的各种形状以及各种可能的退化。形状包括圆形、椭圆形、矩形、正方形、菱形,以及一些不规则形状等。退化包括印章的形状不规则、颜色模糊、深浅不一、对比度过低等。

3.2 印章获取

神经网络在进行监督训练时需要大量成对的训练数据,但手动进行标注数据的代价是昂贵的。因此,我们通过将提取的印章掩膜和背景图片相融合的办法来得到带有印章的文档图片和与之对应的无印章文档图片。用于融合的印章图像分为两种:合成的印章和真实文档图像中的印章。合成的印章是基于我们设计的高仿真的印章合成方法合成,其中真实的印章通过图像编辑软件人工提取得到。下面具体介绍这两种印章获取方法。

1)合成印章

为了生成足够真实的印章图片,我们设计的印章合成方法主要包含3个部分:(1)印章的形状和组件,包括圆形、椭圆形、正方形、三角形、矩形、菱形。印章中的组件有五角星、国徽、文字、数字;(2)印章组件的环绕方式,印章中组件的排列方式有:沿印章边缘环绕排列,中心对齐排列,底部横向排列,

同时包含单圈印章和多圈印章两种圈数类型;(3)印章的上色。

生成印章图像分为以下两个步骤:

(1)生成二值掩模图,分为5个步骤:第一步,从6种形状中随机选择一种,并合理控制图形的大小使其能完整、居中、对齐地放在512*512像素的背景下,并随机增加圈数。第二步,随机选择印章中心的组件组合,分别有空白、国徽、数字、文字、五角星的形式。第三步,随机选择沿印章边缘环绕的文字字数与内容,其文字形式包括空白、纯数字、数字与字母组合、汉字等。我们设计了两种文字选择方式:一种是从设定好的字典中选择,其中包括“全国印章统一印制”以及各省份名等真实印章中的高频词语;另一种是从中文常用字库中随机选择。第四步,将文字合理地环绕在印章边缘,通过控制角度旋转印章并保持文字打印位置不变,使文字合理环绕在印章边缘位置。第五步,随机平移旋转印章,保证印章形状全部在图片内。图4第一行是生成的掩膜的实例。



图4 基于合成的印章及背景融合的样本示例

Fig. 4 Examples of samples based on the fusion of synthetic seals and backgrounds

(2)印章上色:本文只考虑生成红色印章,在HSV颜色空间中对(H,S,V)三通道像素值在红色的范围内进行高斯采样得到(H_0, S_0, V_0)。接着将掩模图中印章位置的像素值替换为(H_0, S_0, V_0),得到最终彩色印章图像。

2)提取真实印章

仅仅利用合成印章进行训练的效果并不理想。为了提高在真实场景中的泛化性,我们利用Photoshop对真实场景中的文档图片中的印章进行提取。采取的方法是先消除后提取,即先对印章进行消除,得到图像修复后的图片,再根据原图与修复后图片中修改像素值较大(人为设定阈值)的位置进行筛选来提取出印章。对于每一张带有印章的原始图片,我们希望尽可能去除印章并且最大程度还原印章位置下的像素以保持修复区域与周围区域纹理的一致性,最后得到一张消除印章后的文档图片以及一张印章的掩膜图。

我们使用Photoshop的图形修复功能,利用仿制印章工具和画笔工具来消除印章。原始印章图片可以分为纯色背景和复杂纹理背景。对于纯色背景,我们使用PS的图像修复功能,复制区域外的像素对选定区域进行填充来消除印章,利用画笔工具和仿制印章工具进行微调来保证区域的连续性。对于复杂纹理背景,我们不仅要消除印章,还要填充印章下的背景。而印章下往往包含文字、表格线、数字、发票水印等纹理变化剧烈的背景。图像修复工具无法对文字、表格线等

边缘锐利的位置进行修复,因此我们首先用仿制印章工具对大块背景区域进行修复,再利用画笔工具逐个像素对文字等复杂纹理位置进行修复,像素的颜色选择标准是选择与该位置最接近的像素颜色。最后,根据原图和修复后的图片进行相减,设定阈值提取出变化位置,即为印章位置,将原图中的印章区域提取出来,得到印章掩膜图。图 2 是 5 张标注数据的实例,第一行是原图,第二行是修复后的图片,右下角是提取的印章掩膜图。

3.3 样本生成

得到足够多的背景图片和印章掩膜图后,需要将两者随机进行融合。我们采用的融合的方法是对应像素加权相加的方法。首先从 0.3~0.7 进行均匀采样,将采样值作为印章图片的像素权值,将 1-采样值作为背景图片的像素权值,最终图像由两者加权求和得到。生成的样本分为两种,一种是通过合成印章融合的背景,如图 4 所示;另一种是通过真实印章融合的背景,如图 5 所示。

在提取真实印章掩膜时,我们也保留了原图与修复后的图片作为印章消除数据集的样本,用于之后的训练和测试(真实场景下的样本,即真实数据集)。

综上所述,本文成功构建了一个印章消除数据集。该数据集包括一个合成印章数据集和一个真实印章数据集。真实印章数据集仅包含真实文档图像及其相应的无印章版本,这些无印章图像均通过人工标注的方式获得。合成数据集可以分为两个部分,第一部分的印章完全由事先设定的规则生成(详见第 3.2 节“印章获取”),而第二部分则来自真实数据集

中的印章。这两部分印章均与采集的文档图像进行了背景融合,得到了最终的合成印章数据集。



图 5 基于真实印章及背景融合的样本示例
Fig. 5 Examples of samples based on the fusion of real seals and backgrounds

4 SealErase 网络

4.1 SealErase 网络结构

本文设计的印章消除网络 SealErase 的整体结构如图 6 所示,包含一个印章语义分割网络和一个印章消除网络。在第一阶段,分割网络接收带印章的原始文档图片,输出一张对印章位置进行分类的二值图。第二阶段,生成器接收原始文档图片和一阶段输出的二值图,经过粗制-精炼两级生成器输出消印章后的文档图片,最后将去印章图像送入判别器网络,与真实图像进行鉴别。在训练过程中,我们先训练分割网络,然后固定分割网络,再训练消除网络,直到收敛。

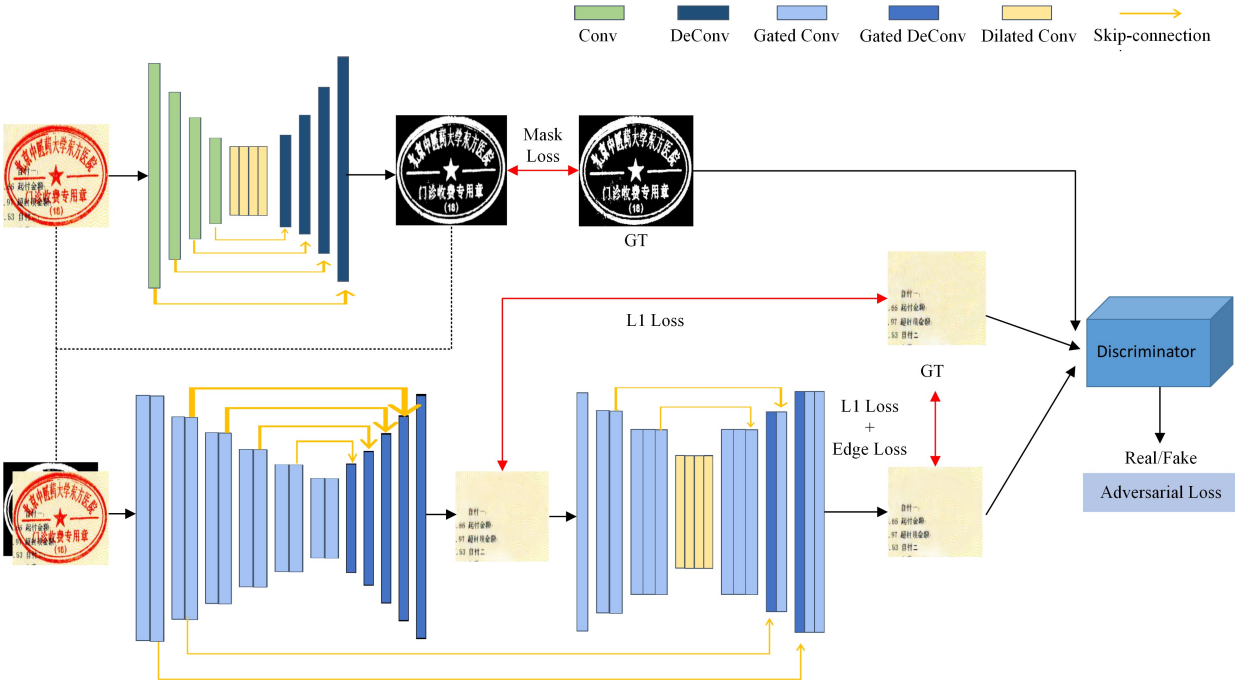


图 6 SealErase 网络结构
Fig. 6 Architecture of SealErase networks

4.2 分割网络

SealErase 网络的分割网络以编解码器 U 型网络为基本框架。编码器和解码器为对称式的 4 层卷积层和 4 层反卷积层,对应的层之间通过跳跃连接获得更好的分割效果。为了增大感受野,中间层由 4 层空洞卷积组成。在分割网络的

训练过程中,有些印章可能只占图片的一小部分区域。为了处理样本的不均匀,我们引入 Dice Loss^[27],但仅使用 Dice Loss 容易造成训练不稳定且会出现梯度饱和的情况。所以,我们采用 Dice Loss 和交叉熵损失(Cross Entropy Loss)相结合的方式训练分割网络,其中 L_{Dice} 代表 Dice Loss,定义如

式(1)所示; L_{BCE} 代表交叉熵损失,定义如式(2)所示; L_{Mask} 代表分割网络的总体损失函数,定义如式(3)所示。

$$L_{Dice} = 1 - \frac{2 \times \sum_{x,y} P_{x,y} G_{x,y}}{\sum_{x,y} P_{x,y}^2 + \sum_{x,y} G_{x,y}^2} \quad (1)$$

$$L_{BCE} = - \sum_{x,y} (G_{x,y} \log(P_{x,y}) + (1 - G_{x,y}) \log(1 - P_{x,y})) \quad (2)$$

$$L_{Mask} = \lambda_{BCE} L_{BCE} + \lambda_{Dice} L_{Dice} \quad (3)$$

其中, $P_{x,y}$ 和 $G_{x,y}$ 分别代表分割网络输出图像和印章掩膜标注在 (x,y) 处的像素值; λ_{BCE} 和 λ_{Dice} 分别是对应损失的权重。

4.3 消除网络

4.3.1 生成器

SealErase 消除网络的生成器设计为一个两阶段网络结构,包含一个粗制子网络和一个精炼子网络。在第一阶段,粗制子网络以原始印章图片和分割网络的输出作为输入,输出粗略的去印章图像。为了能够进一步提升去除性能,我们将精炼子网络串联起来作为生成器的第二阶段。精炼子网络以第一阶段的输出为输入,得到更精细的去印章图像。两个子网络内和两个子网络之间均采用跳跃连接将高级的语义信息与底层网络的输出集成,以便更好地训练生成器。在训练过程中,我们要求两个子网络的输出都尽可能和标注接近。

为了解决印章消除不全、错误修改背景等问题,我们使用了门控卷积^[19],实现对印章位置的精细化消除和修复。门控卷积定义如下:

$$\begin{cases} Gates = \text{Conv}(W_g, I) \\ Features = \text{Conv}(W_f, I) \\ o = \phi(Features) \odot \sigma(Gates) \end{cases} \quad (4)$$

其中, W_g 和 W_f 为卷积核的参数, I 为输入, $\odot$ 符号表示矩阵对应元素相乘运算, σ 为 Sigmoid 激活函数, ϕ 为一非线性激活函数,通常为 ReLU,ELU 或者 LeakyReLU。

4.3.2 判别器

消除网络的判别器采用类似 SN-PatchGAN 的结构,如图 7 所示。

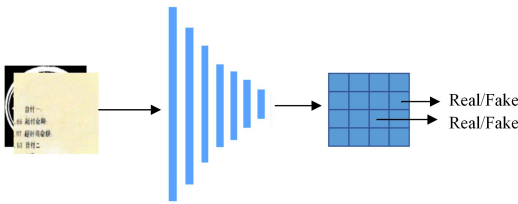


图 7 判别器网络结构

Fig. 7 Architecture of discriminator network

消除网络将谱归一化的思想应用到生成对抗网络训练中,使得判别器满足 Lipschitz 约束,限制了其参数变化的剧烈程度,从而使 GAN 模型训练更加稳定。为了更好地修复图像细节,SealErase 将原始 GAN 判别器网络的全连接层替换成卷积层,输出 $N \times N$ 矩阵,矩阵中的每一个值表示对生成图像中某一小块区域的评估结果,从而更好地关注图像细节,实现精细化的印章图像修复。

4.4 损失函数

令带有印章的原图为 I ,掩膜为 M (印章区域像素为 1,非印章区域像素为 0),不带印章的标注为 I_g ,分割网络输出为 O_M ,生成器粗略输出为 O_C ,精细输出为 O_R 。为了更好地

训练 SealErase 网络,考虑内容损失、对抗损失以及边缘损失。使用 L1 损失作为内容损失,定义如式(5)一式(7)所示:

$$L_{seal} = \| M \odot (I_g - O_C) \|_1 + 2 \| M \odot (I_g - O_R) \|_1 \quad (5)$$

$$L_{background} = \| (1 - M) \odot (I_g - O_C) \|_1 + 2 \| (1 - M) \odot (I_g - O_R) \|_1 \quad (6)$$

$$L_{content} = \lambda_{seal} L_{seal} + \lambda_{bg} L_{background} \quad (7)$$

其中, λ_{seal} 和 λ_{bg} 为对应损失的权重。将合页损失函数^[28](Hinge Loss)作为对抗损失,判别器 D 对抗损失定义如式(8)所示,生成器 D 对抗损失定义如式(9)所示。

$$L_D = E_{I_g \sim P_{data}(I_g)} [\phi(1 - D(I_g, M))] + E_{I \sim P_I(I)} [\phi(1 + D(G(I), M))] \quad (8)$$

$$L_G = -E_{I \sim P_I(I)} [D(G(I), M)] \quad (9)$$

其中, ϕ 为 ReLU 激活函数。为了保持边缘一致,使用边缘一致损失如式(10)所示:

$$L_{edge} = \| S(O_R) - S(I_g) \|_1 \quad (10)$$

其中, $S(\cdot)$ 为 Sobel 算子。最终,生成器的损失函数 L_{total} 定义如式(11)所示:

$$L_{total} = \lambda_c L_{content} + \lambda_G L_G + \lambda_{edge} L_{edge} \quad (11)$$

其中, $\lambda_c, \lambda_G, \lambda_{edge}$ 为对应损失的权重。判别器的损失函数为 L_D 。

5 实验

5.1 参数设置

我们在印章数据集以及真实数据集上评估了 SealErase 网络的性能,构建了包括 5000 张由规则合成的印章图像和 3000 张由真实数据集中的印章进行背景融合的合成图像的数据集。其中 5500 张图像被划分为训练集,2500 张图像被划分为测试集。在进行数据集划分时,应特别注意避免使用训练和测试集中的重复印章和背景图像。简而言之,对于由规则合成的印章图像,应保证每一张图片都是独立的。而对于真实数据集中的每一个印章和其对应的背景,将它们只放置于训练集或测试集之中,而不会在两个数据集中同时出现。在训练 SealErase 时,输入图片四周统一填充或缩放放到 512×512 大小。采用最大角度 15° 的随机旋转、随机图像比例和随机水平翻转来进行数据增强,提高泛化性。利用 Adam 优化器对 SealErase 网络进行参数更新。在分割网络和生成器中,将 Adam 的学习率设置为 0.00003, β_1 为 0.9, β_2 为 0.999。在判别器中,将 Adam 的学习率设置为 0.00001, β_1 与 β_2 不变。经过实验验证,我们将超参数 $\lambda_{BCE}, \lambda_{Dice}, \lambda_b, \lambda_s, \lambda_c, \lambda_G, \lambda_{edge}$ 分别设置为 1, 1, 1, 10, 5, 0.1, 1。SealErase 网络通过 PyTorch 构建。实验的批大小(Batchsize)统一设置为 1。

5.2 评价指标

为了更好地评价 SealErase 的消除性能,我们设置了内容评价指标和 OCR 字符识别准确率评价指标,使用图像修复和自然场景文字消除中常用的内容评价指标,包括:

1)均方误差(Mean Square Error, MSE):衡量两幅图片的均方损失。

2)峰值信噪比(Peak Signal to Noise Ratio, PSNR):计算一幅图片的峰值信噪比。

3)平均结构相似性^[8](Mean Structural Similarity,

MSSIM):衡量两幅图片的平均结构相似性。

4)平均灰度误差 (Average Gray Error, AGE):计算两幅图对应灰度图的绝对值误差。

5)误像素比例 (Percentage of Error Pixels, pEPs):在人为设定的阈值范围内计算两幅图不同像素的比例。

越高的 PSNR 和 MSSIM,越低的 MSE 和 pEPs,意味着更好的图像质量。消除印章的最终目的是提高发票等文档的

文字识别准确率,因此我们不仅考虑了图像的内容差异,还利用被印章遮盖字符(如图 8 中绿框所示)的 OCR 识别准确率来评价 SealErase 网络的消除性能,如式(12)所示:

$$CA=\frac{CC}{TA}$$

(12)

其中,CA 表示字符识别准确率,CC 表示正确识别的字符数,TA 表示识别的总字符数。



图 8 印章数据集结果对比(电子版为彩图)
Fig. 8 Result comparison of seal datasets

5.3 对比实验

为了验证 SealErase 网络消除印章的有效性,本文分别使用 Pix2Pix 网络^[7]、CycleGAN 网络^[5]、SealGAN 网络^[4]和 SealErase 网络在印章数据集上进行印章消除实验。其中 Pix2Pix 需要配对的训练数据,而 CycleGAN 和 SealGAN 不需要配对的训练数据。

Pix2Pix: Pix2Pix 网络^[7]利用条件对抗网络结构,在生成器的损失函数里同时使用对抗性损失和 L1 损失。对抗性损失用于最小化生成图片和真实图片的区别,而 L1 损失则用于保持生成图片和目标图片在像素级别的相似度。这种方法大大提高了生成图像的逼真度和清晰度。Pix2Pix 模型的判别器还引入了一种名为 PatchGAN 的结构,它将整张图像分割为多个子块,并对每个子块进行判别。这种结构可以使判别器对图像中细节变化的识别更加敏感,从而提高了模型的生成能力和逼真度。然而,将 Pix2Pix 直接用于印章消除会存在印章消除不全以及过度消除的问题,因为 Pix2Pix 将印章定位与印章擦除放在一个网络中进行学习,使得生成器更难定位到印章位置并实现精细化擦除。

CycleGAN: CycleGAN 网络^[5]仍使用条件对抗网络结构,其相较于 Pix2Pix 模型的改进在于引入了循环一致性

损失,使得 CycleGAN 无需成对的域之间的数据,只需要大量的单向数据即可训练出良好的转化模型。然而, CycleGAN 存在与上述 Pix2Pix 模型同样的问题,并且由于 CycleGAN 只是通过对抗训练达到最终目标,而没有将生成图像与真实图像之间的 L1 距离作为指导,因此生成的图像存在颜色偏差等问题。

SealGAN: SealGAN 网络^[4]在 CycleGAN 的基础上,在中间层以及跳跃连接处引入更多卷积层来提升模型的特征提取能力。但是,这种提升往往来自于更大的参数量,并没有直接解决印章消除的核心问题(印章定位与印章擦除)。

实验结果如表 2 所列,印章消除效果图如图 8 所示,其中 SN 表示分割网络。图 8 前两行是在合成数据集上的效果图,后 4 行是在真实数据集上的效果图。根据表格数据可以看出, SealErase 在各项内容指标上都表现优秀,超越了其他先进方法。此外,从图 8 中可以看出, SealGAN 在某些情况下存在一些问题。例如,当背景中有与印章颜色接近的文字时, SealGAN 难以准确定位印章位置,导致印章消除不彻底,同时背景内容严重损失。而当印章颜色复杂时, SealGAN 可能会留下清晰的印章轮廓或产生视觉伪影等问题,这些问题都会影响其图像翻译的质量。相比之下, Pix2Pix 模型更容易产生

大量的视觉伪影,并且在印章消除过程中,印章并没有被彻底消除,且遮盖在印章下方的字符也被消除了。除了上述问题,CycleGAN 还存在一个问题,即在修改印章区域时会修改

印章区域外的背景颜色。由于分割网络能精确输出带有印章位置信息的掩膜,SealErase 能精确定位印章位置,有效地恢复印章区域的背景像素且最大程度保留印章外区域的像素。

表 2 印章数据集内容评价指标对比

Table 2 Comparison of content evaluation on seal datasets

模型	印章数据集					真实数据集				
	MSE ↓	PSNR ↑	SSIM ↑	AGE ↓	pEPs ↓	MSE ↓	PSNR ↑	SSIM ↑	AGE ↓	pEPs ↓
Pix2Pix	0.0025	27.19	0.9227	3.90	0.0332	0.0039	25.09	0.9064	7.89	0.0987
CycleGAN	0.0028	27.78	0.9312	6.10	0.0459	0.0053	24.75	0.8994	9.43	0.1107
SealGAN	0.0018	29.38	0.9394	4.21	0.0307	0.0037	26.62	0.9068	6.42	0.0954
SealErase w/o SN	0.0008	33.86	0.9566	3.06	0.0191	0.0021	28.69	0.9408	5.10	0.0465
SealErase	0.0003	37.25	0.9815	1.33	0.0067	0.0013	30.21	0.9506	3.58	0.0236

为了更好地证明分割网络的重要性,我们将分割网络去掉,单独将消除网络在印章数据集上进行实验。将消除网络的输入去掉,分割网络的输出仅保留原始图像,其余参数及训练策略不变。实验结果如表 2 第四行和图 8 第六列所示。去掉分割网络后,由于 SealErase 没有包含印章位置信息的掩膜作为输入,消除网络同样难以定位到印章的位置,导致印章的轮廓较为明显地保留了下来。对于这种情况,SealErase 的评价指标可能会略高于 SealGAN 网络,与原 SealErase 网络相差较大。

为了验证印章消除对文本识别的影响,我们将整张测试图片(包含 1500 个被印章覆盖的字符)送入百度 OCR 进行文本识别,表 3 列出了图 8 中真实场景图片中绿框的文本识别结果,红色文本代表错误识别结果,“—”代表没有识别的文本。可以看出,带有印章的图片,文本识别出现大量错误,1500 个字符的识别准确率仅为 51.87%。通过 SealErase 网络进行印章消除后,文本识别准确率提升 38.86%。由于 Pix2Pix,CycleGAN,SealGAN 均存在过度擦除的问题,因此识别准确率仍不尽人意。

表 3 印章数据集文本识别评价指标对比

Table 3 Comparison of text recognition evaluation on seal datasets

原文本	文本识别结果					
	原图	标注	Pix2Pix	CycleGAN	SealGAN	SealErase
	省医弄付诊收费	省医疗门诊收费	省医异 1 罗收费	一方门诊事—	—	省医疗门诊收费
	浮	票	—	—	—	—
	票	票	—	—	—	票
	医疗收	医疗收	—	—	—	医行收
	Iny6 iC 9yangdyn	voice of Guangdyn	—	—	—	—)le of Guancion
	画 囊 一	方 字盖章	方 字盖章	方 字盖章	— 一盖	方 字盖章
	尚凌一	尚建新	尚—	尚影—	泡—	尚建新
						
	动军渝 x 省— 况说丽	动车检测有限 公司况说明	动车检测有限 公司况说明	动车检测有限 公司况说明	动车检测有限 公司况说明	动车检测有限 公司况说明
1500 个字符 识别准确率/%	51.87	97.27	77.00	67.20	71.00	90.73

结束语 本文针对现有印章消除方法存在消除不全、修改背景等问题,提出了两阶段的 SealErase 印章消除网络,包括一个生成印章位置掩膜的 U 型分割网络和一个基于 GAN 实现的精细擦除印章消除网络。针对目前缺乏公开的用于印章消除的成对数据集的问题,我们兼顾真实场景的泛化性以及

对噪声的鲁棒性,构建了一个包含 8 000 个样本的高仿真的印章消除数据集。其中的印章分为两种:真实文档图像中的印章和合成的印章。本文在此印章消除数据集上开展了充分的实验,并设计了基于图像生成质量和被印章遮盖字符识别准确率的评价指标来对比评估各印章消除方法的性能。实验结果表明,相比 SealGAN 等模型,SealErase 网络能够更好地

擦除印章并恢复背景的文字和纹理,对不同种类的印章消除具有更强的鲁棒性,内容评价各项指标均优于最先进的方法,其中 PSNR 指标相比最先进的方法提升了 26.79%,MSSIM 指标提升了 4.48%。在经过 SealErase 进行印章消除后,被印章遮盖的字符识别准确率提高了 38.86%。实验结果表明,SealErase 在真实场景下同样可以有效消除印章并保留被遮盖文字。

然而,印章消除仍然有很多问题没有解决,譬如真实的训练数据集缺乏,复杂印章的定位,背景中复杂的文本和纹理的修复。这仍然是一个开放的研究问题,值得更多学者的关注和探讨。

参 考 文 献

[1] SINGH A, BACCHUWAR K, BHASIN A. A survey of OCR applications[J]. International Journal of Machine Learning and Computing, 2012, 2(3): 314-319.

[2] ZHAO Y T, LI Z M, WANG H J, et al. Image preprocessed study of the seal imprint verification[J]. Chinese Journal of Scientific Instrument, 2004, 25(z3): 401-403, 410.

[3] JI J J, LOU Z. Filtering of color seal on bank notes based on re-segmentation[J]. Modern Electronics Technique, 2014, 37(22): 5-9.

[4] LI X L, ZOU C M, YANG G T, et al. SealGAN: Research on the seal elimination based on generative adversarial network[J]. Acta Automatica Sinica, 2021, 47(11): 2614-2622.

[5] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C] // Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2017: 2223-2232.

[6] WANG J, MIAO J, QING L Y, et al. Seal removal based on Pix2Pix network[J]. Journal of Beijing Information Science & Technology University, 2021, 36(4): 39-43.

[7] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 1125-1134.

[8] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[9] ANVARI Z, ATHITSOS V. A Survey on Deep learning based Document Image Enhancement[EB/OL]. <https://arxiv.org/pdf/2112.02719.pdf>.

[10] HRADIS M, KOTERA J, ZEMCIK P, et al. Convolutional neural networks for direct text deblurring[C] // Proceedings of BMVC. 2015.

[11] GANGEH M J, TIYYAGURA S R, DASARATHA S V, et al. Document enhancement system using auto-encoders[C] // Workshop on Document Intelligence at NeurIPS 2019. 2019.

[12] MAO X, SHEN C, YANG Y B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections[J]. arXiv:1603.09056, 2016.

[13] SOUBGUI M A, KESSENTINI Y. DE-GAN: a conditional generative adversarial network for document enhancement[J]. arXiv:2010.08764, 2020.

[14] MIRZA M, OSINDERO S. Conditional generative adversarial nets[EB/OL]. <https://arxiv.org/pdf/1411.1784.pdf>.

[15] JEMNI S K, SOUBGUI M A, KESSENTINI Y, et al. Enhance to read better: A Multi-Task Adversarial Network for Handwritten Document Image Enhancement[J]. Pattern Recognition, 2022, 123: 108370-108383.

[16] ZHAO L L, SHEN L, HONG R C. Survey on image inpainting research progress[J]. Computer Science, 2021, 48(3): 14-26.

[17] ELHARROUSS O, ALMAADEED N, AL-MAADEED S, et al. Image inpainting: A review[J]. Neural Processing Letters, 2020, 51(2): 2007-2028.

[18] LIU G, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions[C] // Proceedings of the European Conference on Computer Vision (ECCV). 2018: 85-100.

[19] YU J, LIN Z, YANG J, et al. Free-form image inpainting with gated convolution[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2019: 4471-4480.

[20] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. <https://arxiv.org/pdf/1511.06434.pdf>.

[21] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders: Feature learning by inpainting[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 2536-2544.

[22] YU J, LIN Z, YANG J, et al. Generative image inpainting with contextual attention[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 5505-5514.

[23] YU Y, ZHAN F, LU S, et al. WaveFill: A Wavelet-based Generation Network for Image Inpainting[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2021: 14114-14123.

[24] NAKAMURA T, ZHU A, YANAI K, et al. Scene text eraser[C] // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2017: 832-837.

[25] ZHANG S, LIU Y, JIN L, et al. 2019. Ensnet: Ensconce text in the wild[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2019: 801-808.

[26] LIU C, LIU Y, JIN L, et al. EraseNet: End-to-end text removal in the wild[J]. IEEE Transactions on Image Processing, 2020, 29: 8760-8775.

[27] MILLETARI F, NAVAB N, AHMADI S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C] // 2016 fourth International Conference on 3D vision (3DV). IEEE, 2016: 565-571.

[28] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks[EB/OL]. <https://arxiv.org/pdf/1802.05957.pdf>.



WU Guibin, born in 1982, Ph. D. His main research interests include natural language processing, image processing and deep learning.



XIONG Yongping, born in 1982, Ph.D, associate professor. His main research interests include document intelligence and OCR, visual IoT and machine vision.