ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 39,No. 1,Jan. 2024,pp. 167－180
DOI:10. 16337/j. 1004-9037. 2024. 01. 015

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

面向畸变扭曲文档的两种图像矫正网络

冯　瑾，池　越，周亚同，何静飞

（河北工业大学电子信息工程学院，天津 300401）

摘　要：由于文档纸张的几何形变、拍摄场景的干扰及拍摄角度不理想导致的透视失真，移动设备获取的文档图像的光学字符识别（Optical character recognition，OCR）性能受到很大挑战。针对折叠和扭曲的畸变文档图像预处理问题，设计了两种基于自编码器的网络结构，以实现自适应性图像矫正并提高文字识别正确率。首先提出空洞残差块和非对称卷积残差块两种残差块，然后将残差块与自编码器相结合，设计了一种非对称空洞自编码器网络；同时利用空间金字塔池化代替全连接层，并用非对称卷积残差块实现特征提取，设计了另一种空间金字塔自编码器网络。实验结果表明，与畸变图像相比，经非对称空洞自编码器网络矫正后的图像在OCR正确率、OCR召回率和文本相似度上分别提高了26.3%、20.4%和12.3%，而经空间金字塔自编码器网络矫正后的图像在正确率、召回率和文本相似度上分别提高了27.7%、22.0%和15.5%。与RectiNet等其他图像矫正网络相比，这两种网络可以自适应矫正多种类型的畸变文档图像，且矫正后的图像在文字识别上表现更为优异。本文提出的两种矫正网络能有效提高图像文字识别正确率、召回率和文本相似度，同时在鲁棒性、泛化性等方面与现有矫正网络相比具有明显的优势。

关键词：图像矫正；畸变文档图像；机器学习；自编码器；卷积残差块；空间金字塔池化

中图分类号：TP391　　**文献标志码**：A

Two Image Rectification Networks for Distorted and Warped Documents

FENG Jin，CHI Yue，ZHOU Yatong，HE Jingfei

（School of Electronic Information Engineering，Hebei University of Technology，Tianjin 300401，China）

Abstract：Due to the geometric distortion of the document paper，the interference from the shooting scene，and perspective distortion brought on by the unfavorable shooting angle，the optical character recognition（OCR）quality of document photos taken by mobile devices has been severely hampered. Two networks based on auto-encoder are created to perform adaptive image correction and increase the accurate rate of text recognition in order to handle pre-processing distorted document images with folding and distortion. First，we propose two different types of residual blocks：dilated residual blocks and asymmetric convolutional residual blocks，and then combine the residual blocks with the auto-encoder to create an asymmetric dilated auto-encoder. In the meantime，we create a spatial pyramid auto-encoder by using spatial pyramid pooling instead of fully connected layers and implementing feature extraction with asymmetric convolutional residual blocks. Experimental results show that，compared with distorted images，the corrected images by the asymmetric dilated auto-encoder respectively improve by 26.3%，

基金项目：国家自然科学基金（61801164）。
收稿日期：2022-12-19；**修订日期**：2023-02-27

20.4% and 12.3% in OCR precision, OCR recall, and text similarity. Besides the corrected images by the spatial pyramid auto-encoder respectively improve by 27.7%, 22.0% and 15.5% in OCR precision, OCR recall, and text similarity. Compared with other image rectification networks such as RectiNet, the corrected images by these two auto-encoders perform much better on optical character recognition. The corrected document images of both asymmetric dilated auto-encoder and spatial pyramid auto-encoder are effectively improved in terms of OCR precision, OCR recall, and text similarity. Not only that, they have relatively obvious advantages over existing networks in terms of robustness and generalizability.

Key words: image rectification; distorted document image; machine learning; auto-encoder; convolutional residual block; space pyramid pooling

引　言

随着数码设备和移动设备的发展,阅读方式的载体已经由书籍、报纸等纸质材料变成手机、电脑和iPad等电子设备。对于电子书籍,文档图像是其中一个重要形式。但在现实生活中,文档自身存在折叠扭曲情况,如纸张会因受潮变成波浪状,摊开较厚的书本会使书页产生弯曲[1]。在拍摄过程中,无疑是希望获取文档的全部文字信息,此时图像的边界会包括周围的环境,或者由于拍摄姿势或角度的限制,图像中文档发生倾斜和扭曲形变。畸变扭曲图像会严重影响版面分析、表格提取及文字识别等后续应用效果[2]。因此矫正扭曲畸变文档图像是文字识别前一个重要的预处理过程。

为提高畸变扭曲文档的可读性和易分析性,国内外针对矫正进行了研究并取得了丰硕成果。现有的矫正方法可以分为基于图像处理和基于机器学习的方法。基于图像处理的方法大多利用边界检测[2]、基准线检测[3-4]、连通域分割等图像处理方式来实现图像矫正。这些方法处理精细,但是只能处理轻微透视和扭曲的图像或某种特定的畸变,不具备自适应性[5]。以下重点介绍基于机器学习的矫正方法。

由于图像的纹理特征不会随图像倾斜、旋转而改变,对噪声有很好的抵抗性,机器会自动学习图像的纹理特征规律,实现自动矫正畸变图像。原有的文本行、基准线特征在机器学习的方法中仍然适用[6-7]。Mohammad等[8]通过估计基线形状和字符的倾斜角来获得文本每行的翘曲形状。同样连通域检测的方法在机器学习中也有应用[9]。针对表面不平整引起的像素变化,Garai等[10]提出用翘曲控制参数(Warping control parameter, WCP)和翘曲位置参数(Warping position parameter, WPP)来衡量畸变程度,用卷积神经网络(Convolutional neural network, CNN)[11]估计WCP,用文本行弯曲度等特征计算WCP。Xie等[12]利用网络提取语义信息预测畸变图像的控制点和矫正图像参考点,然后在控制点和控制点之间使用插值、映射等操作实现图像矫正。这些方法的效果很大程度上取决于数据集的全面性,无法做到自适应性的矫正图像。

在基于机器学习的诸多矫正方法中,还有一种端到端的方法[13]。这种方法把图像矫正理解为图像到图像的翻译问题,并且利用神经网络实现矫正[14]。DocUNet[15]是第一个实现端对端矫正畸变图像的方法,它使用两个堆叠的U-Net[16]结构把图像矫正问题转化为像素坐标位置的回归预测问题。与传统方法相比,其优势是可矫正任意弯曲畸变的图像。但是该方法对整个图像集进行训练,当尺寸变大像素点变多时训练成本骤增,且需要大量的样本集支持。DewarpNet[17]增加了大量3D图像特征,其矫正效果相对DocUNet具有显著提高。但是DewarpNet对设备要求很高不具有推广性,而且矫正后的图像容易产生撕裂从而引入新的形变。RectiNet[18]提出一种具有分支的堆叠U-Net网络,分支网络用于辅助识别文档的边界和线。该网络可以有效地去除背景,但只是粗暴的矫正边界,内部文本依旧畸变,甚

至在矫正过程中会使得文本更加扭曲。Li等[19]提出了一种图像流切割缝合方法,并基于补丁设计了文档几何矫正和照明矫正网络。补丁方法可有效增强数据集、缩短训练成本,但是该方法采用学习率轮数减缓方式容易导致训练不到位,矫正效果仍有待提升。

目前已有诸多研究虽然在一定程度上实现了文档图像矫正,但是存在着过度关注图像矫正前后图像相似度,忽略文字内部结构反而引入新的形变的问题。此外还有些方法加入过多先验知识,无法做到自适应矫正多种类型的畸变。为解决上述问题,本文设计了基于补丁的自编码器网络学习畸变图像到正常图像的图像流。图像流表征图像间的像素级位移流向。基于图像流的网络对图像的中间或边界等各个区域的畸变情况一视同仁,矫正过程中不易产生新的形变。该网络不受先验条件变化的影响,可以实现自适应性的图像矫正。同时为改善畸变文档图像对后续处理操作的影响,提高文字识别的正确率,充分扩大了改进网络的感受野,并进行特征融合。在合成验证图像集以及现实畸变图像集分别评估网络,并与其他4个网络进行对比实验。实验结果显示,与畸变图像相比,经本文网络矫正后的图像在文字识别正确率、召回率、文本相似度上得到显著提高,与其他矫正网络相比,本文网络具有更好的鲁棒性和泛化性。

本文主要贡献如下:

(1)设计了非对称卷积残差块和空洞残差块分别用于特征提取和增大感受野。

(2)将非对称卷积残差块用于下采样,并利用空洞残差块结合自编码器设计了非对称空洞自编码器(Asymmetric dilated auto encoder,ADAE)网络,重点聚焦并矫正文本水平方向。

(3)利用空间金字塔池化代替全连接层,结合非对称卷积残差块设计了另一种空间金字塔自编码器(Spatial pyramid auto encoder,SPAE)网络,增加网络鲁棒性减少计算量。

1　畸变扭曲图像矫正理论基础

1.1　机器学习图像矫正

文档图像会有其特定的结构信息,同样的畸形图像中也会有一些不规则的结构信息,如曲折的文本行走向、非矩形的纸张轮廓以及其他非文本信息。这些结构信息可以表征畸变特征,是矫正图像的有利线索。同光流的含义类似,畸形图像可以看作是正常图像中每一个像素点通过“运动”(即映射)生成了畸形图像,这个像素点的“运动”被称为图像流。因此,假定网络可以从输入畸变文档中自动提取有利特征,并提出一种网络用来学习畸形文档图像到正常图像中的图像流。

对于畸变图像中的任意一个像素点$\boldsymbol{D}(x_1,y_1)$,利用映射变换到对应的正常文档图像的像素点$\boldsymbol{R}(x_2,y_2)$,即

$$\boldsymbol{R}(x_2,y_2)=\boldsymbol{G}_{\text{flow}}\boldsymbol{D}(x_1,y_1) \tag{1}$$

式中:$\boldsymbol{G}_{\text{flow}}$为图像流,是关于$(x_1,y_1)$的二维向量,因此对于高为$H$、宽为$W$的图像,其张量形状为$(H,W,3)$,相应图像流形状则为$(H,W,2)$。

与其他直接对图像进行操作的网络不同,本文所提出的网络学习是从畸变图像到矫正图像流的过程。因此它具有很强的自适应性,可以表示各种类型的畸变,包括弯曲、折叠和透视形变等。

1.2　自编码器

自编码器是一种处理无标签数据的自监督学习方法。自监督学习通过构建辅助任务来获取数据中的监督信息,如单词预测随机打乱单词顺序、图像重组将图像切割为补丁。图像矫正则是随机对图像进行扭曲变形处理,以此来构建辅助任务,用相应图像流监督信息。

自编码器由编码器、隐藏层和解码器组成。在训练的时候编码器会将高维的输入数据压缩为低维向量，这相当于施加一个“瓶颈”，迫使原始输入压缩知识，只保留原始数据中关键特征。与主成分分析不同的是，编码器可以利用神经网络的非线性特征提取能力，使得输出的低维隐藏层变量更具有代表性。解码器则对隐藏层变量升维重构，一般自编码器优化目标是最小化输入和输出之间的误差，即最小重构误差。

1.3 卷积残差块

残差网络是一种可缓解网络退化和梯度消失的网络，可提高网络的有效深度[20]。残差网络由多个残差块组成。残差块采用跳跃连接的形式，分成主路径(Main path)和捷径路径(Shortcut path)两条路。标准残差块(Identity block)的输入与网络层输出相连，两者结合完后再通过激活函数。卷积残差块(Convolution residual block, Conv block)是另一种类型的残差块，与标准残差块不同的是，卷积残差块捷径路径中添加了卷积层，其结构如图1所示。

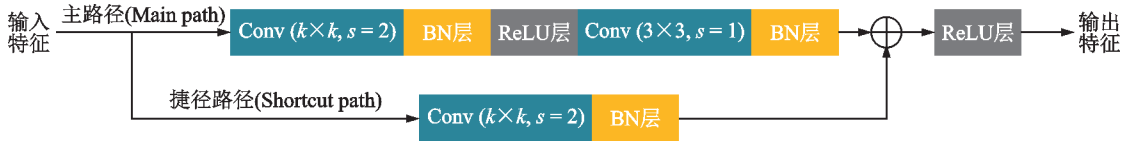


图1 卷积残差块结构图

Fig.1 Structure diagram of convolution residual block

标准残差块不能调节特征维度，卷积残差块弥补了这个不足。两者结合使用，既增加了网络深度，提高模型学习效率和准确率，又可以改变网络维度实现特征降维，使得低维向量更具备代表性。

2 图像矫正网络及改进

2.1 残差块设计

卷积过程是对图像局部像素点的非线性特征提取，而卷积核的尺寸影响图像局部加权卷积区域，也就是感受野。虽然畸变图像变形复杂多样，其弯曲或折叠情况也是完全随机的，但是文字走向可以表征形变走向，这个特征不变。增大感受野可以增大卷积层输出的特征图在输入图像上的映射区域，从而覆盖更多的文字区域。

下面将以增大模型感受野、提取畸变图像特征为出发点，在原有标准残差块的基础上改进并设计两种残差块用于自编码器的编码部分。

2.1.1 非对称卷积残差块

理论上说，提高卷积核尺寸是增大感受野的方法之一，常用的卷积残差块中卷积核尺寸为3×3[21]。对于步长为1的卷积层，如果把3×3的卷积核换成7×7的卷积核，感受野从9变到了49，但是参数量也随之增多。为平衡感受野和参数训练成本之间的平衡，提出一种非对称卷积残差块(Asymmetric convolution residual block, Asymmetric conv block)。先前的工作表明，标准的$k\times k$卷积分成$1\times k$和$k\times 1$的两个卷积，虽然最后的实验精度略有下降，但是大幅降低了参数量[22]。同时，人们的阅读习惯为逐行阅读，对垂直方向上的不工整具有一定包容性，而水平方向上的杂乱无章会严重影响阅读体验。故提出非对称卷积残差块，扩大了卷积核水平方向上的尺寸，着重于提高文档水平方向上的感受野，结构如图2所示。扭曲图像分成两支，一支通过7×5的非对称下采样卷积层和3×3卷积层，另一支只通过非对称下采样卷积层，最后结果相加进入网络的下一层。理论感受野[23]的计算公式为

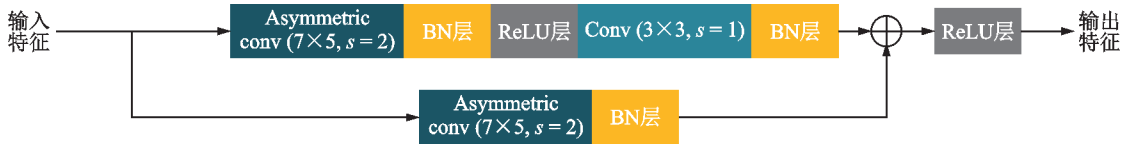


图2　非对称卷积残差块结构图

Fig.2　Structure diagram of asymmetric convolution residual block

$$TRF_l = TRF_{l-1} + stride_{l-1} \times (kernel_l - 1) \quad (2)$$

$$TRF_{2D} = TRF_x \times TRF_y \quad (3)$$

式中:TRF表示理论感受野值;l表示层数且$l \geqslant 2$;stride表示步长;kernel表示卷积层的卷积核尺寸;TRF_{2D}为2D表示的理论感受野大小;TRF_x和TRF_y分别表示横向和纵向感受野大小。不同卷积残差块的每层理论感受野值如表1所示。

表1　不同残差块的横向、纵向以及理论感受野

Table 1　Transverse, vertical and theoretical receptive fields of different residual blocks

卷积残差块	TRF_x	TRF_y	TRF_{2D}
3×3卷积残差块	7	7	49
非对称卷积残差块	11	9	99

根据表1可知,图像经过一个残差块,感受野提高了将近两倍,且更加聚焦畸变文档水平方向上的特征。将原有卷积残差块的卷积层替换成非对称卷积获得了非对称卷积残差块,在限制网络参数增长的基础上扩大了理论感受野,聚焦图像水平方向畸变特征。

2.1.2　空洞残差块

空洞卷积可以在不损失信息且不引入更多参数量的情况下增大有效感受野,使得输出的特征图包括更大范围的特征信息,因此提出一种空洞残差块(Dilated convolution residual block, Dilated conv block)用于编码,结构如图3所示。特征向量依次通过空洞卷积层、BN层、ReLU层、空洞卷积层、BN层与它本身相加,最后通过ReLU层作为下一层的输入向量,其中空洞卷积用尺寸为3×3、步长为2、空洞率为2的卷积核。使用空洞卷积的残差块与不使用的相比,横向和纵向的有效感受野从5扩大到9,2D感受野从25扩大到81。与标准残差块相比,空洞残差块不再使用普通卷积改为空洞卷积,充分增大了模型的有效感受野。

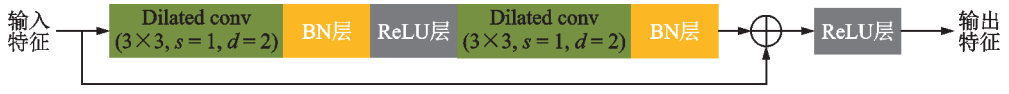


图3　空洞残差块结构图

Fig.3　Structure diagram of dilated convolution residual block

2.2　非对称空洞自编码器

传统的自编码器结构大多使用卷积层或残差块对图像编解码。为扩大模型的感受野,在传统自编码器结构中融合非对称卷积残差块和空洞卷积残差块,设计出非对称空洞自编码器(ADAE),如图4所示。ADAE网络利用编码器1(ADAE-encoder 1)对局部补丁图像进行特征提取,学习图像的畸变特征;利用编

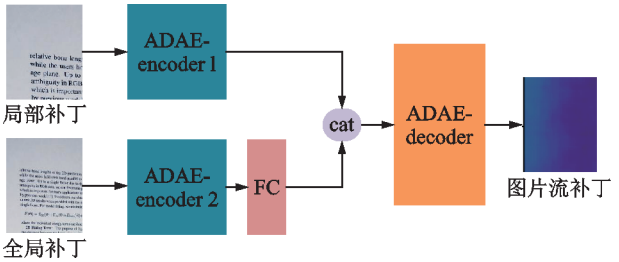


图4　非对称空洞自编码器结构图

Fig.4　Structure diagram of ADAE

码器2(ADAE-encoder 2)对全局补丁图像进行特征提取,为网络引入局部形变补丁周围图像特征,添加上下文信息促进畸变特征表征学习、改善失真估计;然后将对全局特征向量进行全连接操作将其压缩,拼接至局部特征向量上实现特征融合;最后利用解码器(ADAE-decoder)对融合后的特征向量上采样预测相应的局部图像流补丁。

ADAE-encoder 1结构如图5(a)图所示,包括2个卷积层、2个标准残差块、3个卷积残差块(1个非对称和1个对称)和2个空洞卷积残差块。其中第一层卷积层采用7×7卷积核,后面的卷积层采用3×3,主要目的是增大通道数;使用标准残差块旨在缓解梯度消失降低损耗;使用卷积残差块的目的是实现空间下采样和特征提取,非对称卷积残差块也会对特征向量进行降维提取,但是会更加聚焦图像横向畸变特征;空洞卷积残差块用于加深网络,增大感受野。在残差块中每个卷积层后面都会添加归一化层和激活函数(为减少篇幅图中并未画出),用于改善训练。ADAE-encoder 2结构如图5(b)图所示,包括1个卷积层、1个标准残差块、2个非对称卷积残差块和4个卷积残差块,各部分功能与ADAE-encoder 1相似。解码器使用3个双线性插值模块实现空间上采样,同时伴随卷积层进行通道数减小。

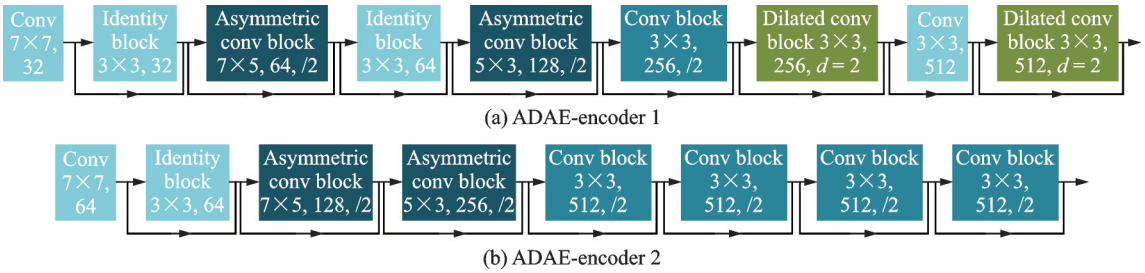


图5 非对称空洞自编码器网络详细编码结构

Fig.5 Detailed coding structure of ADAE network

2.3 空间金字塔编码器

提高文档图像分辨率,可以增加可利用数据,但也会伴随着训练成本上升的代价。受超分辨率图像重建启发,对图像插值生成到高分辨率图像补充缺失值,为后续卷积操作提供了更多的特征信息。实验结果发现,ADAE网络矫正后的图像出现了细微的缝隙情况。因此在编码过程中对局部补丁图像重建,融合局部补丁图像的多尺度特征,最终设计出空间金字塔编码器(SPAE)。SPAE的数据处理结构(图6)与ADAE相似,不同的是SPAE的全局补丁分支依次经过编码器(SPAE-encoder 2)、标准残差块(Identity block)、多尺度融合模块(Multi-scale fusion module, MSFM),生成全局补丁特征量再与局部补丁特征量相融合。

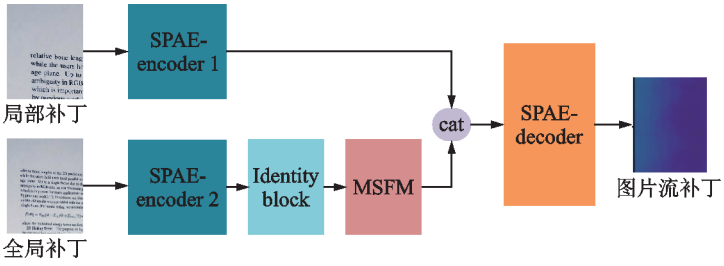


图6 空间金字塔自编码器结构图

Fig.6 Structure diagram of SPAE

SPAE-encoder 1由图像重建模块(Image reconstruction block)、两个非对称卷积残差块、两个对称卷积残差块的组成。如图7所示,与ADAE-encoder 1相比,SPAE-encoder 1引入图像重建模块。在该模块中,对特征向量应用双三次插值(bicubic),再依次通过卷积残差块和非对称残差块实现特征下采样。双三次插值与双线性插值相比,增多了用于插值计算的像素点,使得插值效果更加平滑且准确。由于双三次插值要通过周围16个像素点插值,这种先升维再用卷积残差块降维,明显增大了输出特征

向量上的某一个像素点在输入特征向量上的映射区域[24]。在SPAE的全局补丁分支中,SPAE-encoder 2与ADAE-encoder 2结构相同。图8中标准残差块(Identity block)具体结构同1.3小节,MSFM采用空间金字塔池化。最大池化的滑动窗口和步长依次为4、2、1,对应MSFM从上到下的特征图,最后由于特征向量的尺寸从525×1变到512×1,前后数量较小且数值相近,故用Linger层微调尺寸。虽然全连接层的作用在于特征降维,但是它也会带来参数爆炸的弊端,用卷积叠加池化层代替大型全连接层可以有效地减少计算量提升网络性能,MSFM模块提供多尺度化的全局补丁特征,其中标准残差块调节特征向量通道数,为送入池化层做准备。空间金字塔池化可以从不同的角度提取特征,再将其拼接聚合在一起,充分实现特征融合以增加网络鲁棒性。

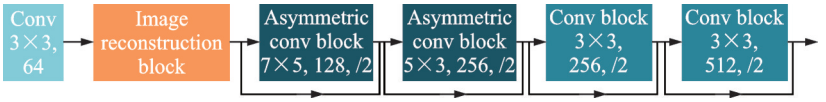


图7　空间金字塔自编码器网络详细编码结构

Fig.7　Detailed coding structure of SPAE network

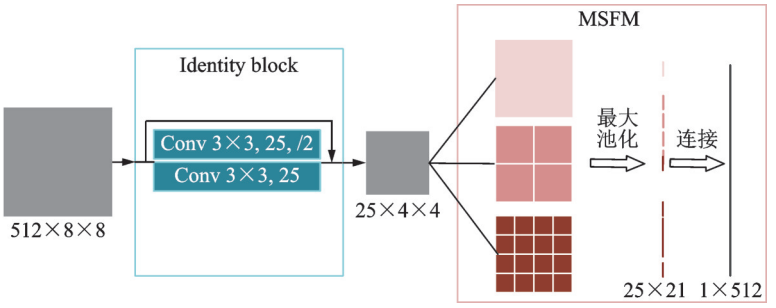


图8　Identity block和MSFM详细结构及特征向量尺寸变化

Fig.8　Detailed structure and feature vector size changes of identity block and MSFM

2.4　损失函数

网络采用最小化损失函数的方式进行训练优化。参考光流估计,将端点误差(End-point error,EPE)作为损失函数。损失函数定义为预测估计图像流与真实图像流中各像素位移向量的欧式距离的均值,即

$$L=\frac{1}{HW}\sqrt{\sum_{i=0}^{H}\sum_{j=0}^{W}(\boldsymbol{P}(i,j)-\boldsymbol{G}(i,j))^2} \tag{4}$$

式中:H和W分别为图像的高和宽,(i,j)表示图像中某一个像素点的位置,$\boldsymbol{P}(i,j)$表示预测估计图像流中相应像素点的位移,$\boldsymbol{G}(i,j)$表示真实图像流中像素点位移。损失函数数值越小,表示预测估计图像流与真实图像流两者越接近,效果越好。

3　实验结果与分析

3.1　网络训练

本实验在Ubuntu20.04操作系统上,基于Python3.8和PyTorch1.10框架实现。硬件环境为15核AMD EPYC 7543 32-Core Processor(内存80 GB),RTX3090(24显存)。

将文献[19]中的畸变图像、图像流数据集,切割成畸变局部图像补丁及其对应的图像流补丁、全局图像补丁。它们的尺寸均为256×256,局部图像补丁间有64个像素点重叠,全局补丁图像的覆盖区域包括局部补丁及其周围区域的2×2倍图像,是512×512大小的图像缩小的结果。随机分配约70 000个补丁数据用于网络训练,20 000个补丁数据用于测试,根据损失函数和测试结果选出最优训练模型。

网络采用Adam[25]优化器控制梯度下降,Adam可以自适应地为网络中不同参数设置学习率,在局

部机器学习中具有高效的性能。学习率初始化为0.000 1,并采用指数衰减的方式,学习率公式为

$$\text{learn\_rate} = \text{base\_learn\_rate} \times \alpha^{\lfloor \text{epoch/step} \rfloor} \tag{5}$$

式中:learn_rate表示当前迭代轮次的学习率;epoch∈(1, 200)表示训练时迭代轮次;base_learn_rate表示学习率初始值;$0<\alpha<1$表示学习率指数衰减的倍数,即每次衰减都乘以α;step表示步长,$\lfloor \text{epoch/step} \rfloor$表示向下取整,即每过step个轮次学习率衰减一次。base_learn_rate=0.000 1,α=0.5, step=2,具体应用为每过2个迭代轮次,学习率衰减至原来的1/2。

3.2 评估标准

不同的网络用于评估文档图像矫正效果的指标不尽相同,Ma等[15]采用多尺度结构相似性(Multi-scale structural similarity, MS-SSIM)和局部失真(Local distortion, LD),Li等[19]使用光学字符识别(Optical character recognition, OCR)识别正确率,Das等[17]采用OCR识别文本与真实文本间最小编辑距离(Edit distance, ED)、字符错误率(Character error rate, CER)、结构相似性(Structural similarity, SSIM)。由于本文主要针对文档图像进行矫正,故从图像相似度和光学字符识别(OCR)效果两种角度进行评估,且重点关注后者。在OCR效果指标中,OCR分数、OCR正确率、OCR召回率体现文字级别的识别效果,文本相似度则展现了文本整体的相似度,评价指标详细说明见表2。

表2 图像矫正评价指标描述

Table 2 Description of evaluation index for image correction

评估角度	评价指标	英文表示	描述
OCR效果	OCR分数	ocrScore	初始值为0,对于真实文本的某一个字符,如果预测文本中存在则加1,否则加0,最后除以真实文本的长度。ocrScore∈[0,1],越大表示两文本越相似。
	OCR正确率	ocrPrecision	初始值为0,与ocrScore判断过程相似,不同的是若预测文本中存在则加1,并去掉预测文本中该字符,同样最后除以真实文本的长度。ocrPrecision∈[0,1],越大表示两文本越相似。
	OCR召回率	ocrRecall	初始值为0,与ocrPrecision判断过程相似,不同的是最后除以预测文本的长度。ocrRecall∈[0,1],越大表示两文本越相似。
	文本相似度	TextSimilarity	首先计算两个文本去除关键词后的Levenshtein(LS)距离,然后用1减去归一化后的LS。 TextSimilarity∈[0,1],越大表示两文本越相似。
图像相似度	直方图相似性	HistSimilarity	计算两张图像对应直方图的相关性,结果位于(−1,1),越大表示图像越相似。
	峰值信噪比	PSNR	PSNR>0,越大表示两张图像越相似。
	结构相似性	SSIM	SSIM≤1,越大表示两张图像越相似。

3.3 验证集评估

为验证ADAE、SPAE网络的有效性,利用不同的网络对验证集的畸变图像进行矫正,并对比矫正后图像和扫描得到真实图像之间的图像相似度。采用Tesseract[26]引擎对矫正后图像进行文字识别,将识别文本与从PDF文件中复制获得的真实文本对比评估OCR识别效果,对比结果如表3所示,其中加粗字体为每个指标的前两名。

验证集包含205张分辨率为1 680像素×2 400像素合成畸变图像。在电子资源库中获取了大量期刊的PDF格式文件,并且为保证数据的多样性,这些电子资源涵盖了医学、政治、通信等多个领域的期

表3　不同网络的评价结果对比

Table 3　Comparison of evaluation results of different networks

网络	ocr-Score/%	ocr-Precision/%	ocr-Recall/%	Text-Similarit/%	HistSimilarity	PSNR	SSIM
畸变图像	84.874 1	49.885 9	54.412 3	43.144 1	—	—	—
Li等[19]	88.725 1	59.361 6	58.606 8	46.701 3	0.976 9	9.999 6	0.159 40
Kim等[27]	85.847	60.463 8	65.717 4	48.088 7	0.969 67	6.599 7	0.112 83
RectiNet[18]	89.994 3	63.664 7	60.431 1	49.121 4	**0.988 14**	10.160 9	**0.187 27**
Xie等[12]	90.378 4	73.208 1	72.436 1	53.358 6	**0.982 53**	**10.901 3**	0.155 81
ADAE(本文)	**90.681 5**	**76.169 2**	**74.792 4**	**55.460 6**	0.975 56	**10.188 5**	**0.190 17**
SPAE(本文)	**91.241 2**	**77.629 2**	**76.434 7**	**58.660 8**	0.975 94	10.141 2	0.187 00

刊。由于文档图像矫正是提高后续文字识别的正确率,故舍弃一些插图表格较多的文档。将PDF文件转换为同一尺寸的图像,利用Ma等[15]的方法随机合成畸变图像,作为验证集进行评估。

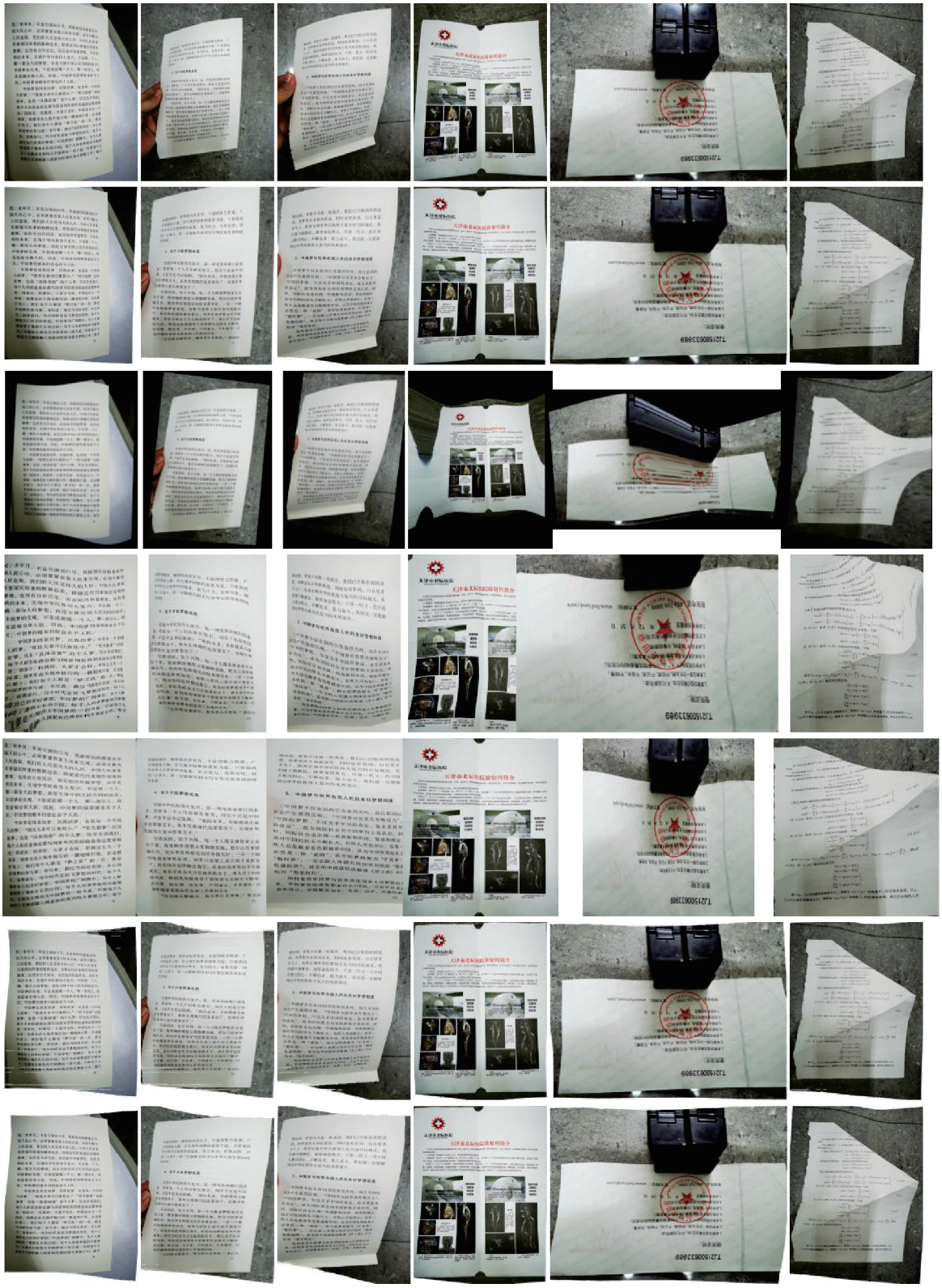
对比分析未矫正的畸变图像,使用ADAE网络矫正后图像在OCR正确率、OCR召回率和文本相似度上依次提高了26.3%、20.4%和12.3%,使用SPAE网络矫正后图像在同样指标上依次提高了27.7%、22.0%和15.5%。实验结果验证了ADAE、SPAE网络对畸变图像矫正的有效性。

对比分析其他图像矫正网络。从图像相似度角度,相较与彩色图像,验证集都是黑白的文档图像,且大多为黑色文字分布具有规律性且色调单一。因此图像相似度的各种指标中,各网络表现效果虽有不同但是差异不大。RectiNet[18]针对文档边界进行辅助矫正,Xie等[12]提出的控制点参考点映射也包括边界轮廓。这两种方法矫正的图像可以去除背景,直方图相似性指标略为优秀。综合分析图像相似度,SSIM指标中ADAE网络表现最为良好,而其他指标ADAE、SPAE网络与对比网络相差较少,表现也很优异。从OCR效果角度,Kim等[27]基于文本行对页面卷曲和拍摄时相机角度造成的透视变换进行矫正,因此在本验证集中同时具有折叠和弯曲的畸变图像上表现效果不好。RectiNet[18]、Xie等[12]对图像边界轮廓给予关注有效去除背景,但是对于图像内部文字矫正注意较少,因此效果较差。最终实验结果表明,ADAE、SPAE网络矫正后的图像在OCR效果上明显优于对比网络,而在两种网络中SPAE更胜一筹。

3.4　现实图像矫正评估

由于训练集、测试集和验证集均是合成畸变图像,为落实实际应用,进一步验证网络泛化性。在现实场景中模拟拍摄了扭曲文档的多种畸变类型图像(包括书籍太厚、手持文档姿势不当、拍摄角度不当、文档折叠、文档弯曲且有干扰及文档折叠且形状不规则)。图9为ADAE、SPAE和对比网络矫正结果。针对Kim等[27]矫正图像裁剪掉部分无效区域,其余图像只调整大小没有改变纵横比。

Li等[19]对每种类型畸变文档都有所矫正,但是每种效果都不如ADAE和SPAE效果好。虽然Kim等[27]方法在曲面弯曲(如图9(a,b)第3行)矫正效果突出,但是对于其他类型的畸变基本无效。RectiNet[18]有效剔除了文档图像背景,但是从图9(a)第4行左下角、图9(c,f)第4行图像可以明显发现该网络过度关注边界信息而忽视了内部结构,从而引入了新的畸变。图9第5行,Xie等[12]的控制点方法设定文档是长方形,故对于有干扰、不规则文档(图9(e,f)第5行)效果很差,同样也引入了新的畸变。ADAE、SPAE网络把握全局信息,重点聚焦文本横向走向,可以很好地实现文字矫正,并且这两种网络可以做到自适应矫正严重透视、折叠、文本行弯曲、形状不规则等畸变文档。



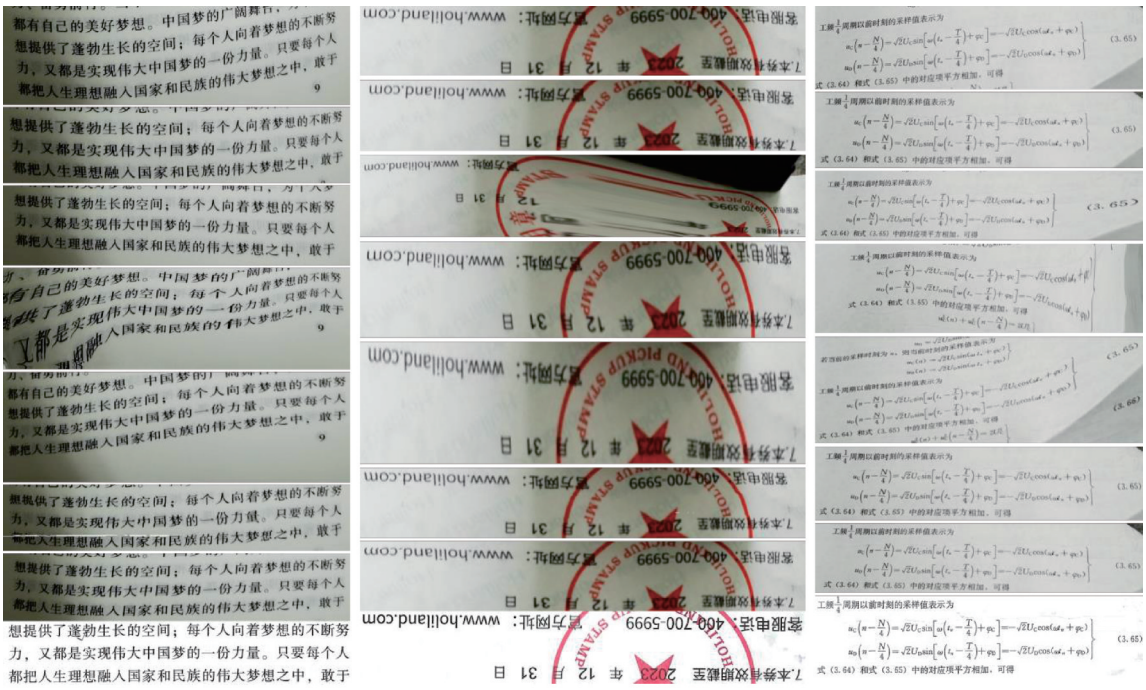
(a) Books that are too thick　(b) Documents held in improper postures　(c) Improper shooting angles　(d) Documents that are folded　(e) Documents that are bent causing distorted text lines　(f) Documents that are folded and irregularly shaped

图9　各网络对真实畸变文档的矫正情况

Fig.9　Rectification results of real distortion documents by various networks

注:从第1到第8行分别为不同网络矫正后的图像:原始畸变图像;Li等[19];Kim等[27];RectiNet[18];Xie等[12];ADAE;SPAE;文档扫描图像。

图10为部分类型畸变图像的矫正结果局部放大图。对于图10(a),Li等[19]、Xie等[12](图10(a)第2、5行)矫正后的图像文本行仍然存在一定程度的弯曲;RectiNet[18](图10(a)第4行)中“又都是”等汉字内部结构发生了严重改变,以至于无法识别。对于图10(e),Kim等[27](图10(e)第3行)完全做不到自适应,Xie等[12](图10(e)第5行)中“网址”“WWW”等文字出现了轻微的模糊。对于图10(f),Kim等[27]、RectiNet[18]、Xie等[12](图10(f)第3、4、5行)矫正后文字出现了不同情况的扭曲。通过观察各网络文本行走向,不难发现ADAE和SPAE矫正后的图像在不改变或少改变文字内部结构的基础上,调整文字



(a) Books that are too thick　(b) Documents that are bent causing distorted text lines　(c) Documents that are folded and irregularly shaped

图10　各网络对部分畸变文档的局部矫正结果

Fig.10　Local rectification results of some distorted documents by various networks

注:从第1到第8行分别为不同网络矫正后的图像:原始畸变图像;Li等[19];Kim等[27];RectiNet[18];Xie等[12];ADAE;SPAE;文档扫描图像。

像素点位置使得文本行更加笔直,充分印证了上述结果。除此之外,Xie等[12]直接对图像操作,输出尺寸固定且清晰度明显下降,由于ADAE、SPAE预测图像流再映射矫正,故矫正前后图像分辨率不变,而图像清晰度越高OCR效果越好。ADAE矫正部分图像在补丁缝合处出现裂痕,初步判断是空洞卷积导致,相比之下,SPAE网络并未出现上述弊端。现实图像矫正结果表明,ADAE、SPAE可以自适应矫正现实场景的多种畸变文档,具有良好的泛化性。

实验结果表明,ADAE、SPAE网络矫正后图像文字识别效果有大幅提升,且效果优于其他对比网络。其中ADAE网络同时使用非对称卷积残差块和空洞残差块,从理论感受野和有效感受野两个维度进行提高,因此在图像相似度指标上表现得比SPAE网络优异,而SPAE网络在OCR角度的各项指标上都比ADAE网络更胜一筹。现实场景实验表明,ADAE、SPAE网络具有稳健的自适应能力,虽然ADAE在补丁缝合处出现了轻微缝隙,但SPAE的改进中增加了图像重建和多尺度融合模块,有效避免了上述情况,不过两者对繁杂畸变文档都展现出良好的泛化性。

4 结束语

本文提出了两种残差块,并基于此设计了两种不同的图像矫正网络ADAE和SPAE。在ADAE网络中,非对称卷积残差块用于特征提取、空洞残差块用于加深网络,聚焦图像横向畸变特征。SPAE网络增加了图像重构模块,通过非对称卷积残差块实现下采样,并用空间金字塔池化代替全连接层,既融合了多尺度特征,又控制了网络参数量。ADAE侧重于网络感受野的扩大,矫正后的图像与扫描图像相似度更高;SPAE则侧重于融合多尺度特征,矫正后的图像OCR正确率更高。相较于畸变图像,使用ADAE、SPAE矫正后的图像,在图像相似度与OCR性能上均取得了显著的效果。面对复杂多变的畸变图像时,ADAE、SPAE具有更为稳定的鲁棒性、更好的文字矫正效果。面对现实场景的畸变图像,ADAE、SPAE自动学习文档畸变特征,表现出良好的自适应性和泛化性。

在后续研究中,将尝试空洞卷积与特征融合模块搭配使用,进一步研究补丁缝隙的成因,并结合两模块的优点。用机器学习端对端矫正文档图像时,文字会出现轻微变形,同时在文字识别过程中,文档图像上的不规则照明也会影响OCR性能。因此在接下来的工作中将尝试对图像流进行平滑处理,并实现文档的光照矫正,力图复原畸变文档图像,为下一步的图像处理工作夯实基础。

参考文献:

[1] ZHU Y, SUN C, WANG Y. Research on the influence of perspective angle on document image correction results[C]// Proceedings of 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC). Chongqing, China: IEEE, 2020: 771-775.

[2] TYMCHENKO O, KHAMULA O, HAVRYSH B. Information technology development in correction method of geometric distortions of text images[C]//Proceedings of 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT). Zbarazh, Ukraine: IEEE, 2020: 199-202.

[3] LI Y, ZOU F, YANG S. Research on improving OCR recognition based on bending correction[C]//Proceedings of 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). Chongqing, China: IEEE, 2020: 833-837.

[4] AL-SHATNAWI A M. A skew detection and correction technique for Arabic script text-line based on subwords bounding[C]// Proceedings of 2014 IEEE International Conference on Computational Intelligence and Computing Research. Coimbatore, India: IEEE, 2014.

[5] ZHANG K, CUI T. Research on text correction technology based on secondary transformation[C]//Proceedings of 2021 International Conference on Computer, Internet of Things and Control Engineering (CITCE). Guangzhou, China: IEEE, 2021: 78-81.

[6] XUE C, TIAN Z, ZHAN F. Fourier document restoration for robust document dewarping and recognition[C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 4563-4572.

[7] JIANG X, LONG R, XUE N. Revisiting document image dewarping by grid regularization[C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 4533-4542.

[8] MOHAMMAD K, QAROUSH A, WASHHA M. An adaptive text-line extraction algorithm for printed Arabic documents with diacritics[J]. Multimedia Tools and Applications, 2021, 80(2): 2177-2204.

[9] 鲁湛, 丁晓青. 基于笔段方向信息的联机手写汉字倾斜矫正算法[J]. 模式识别与人工智能, 2000, 13(4): 378-382.
LU Zhan, DING Xiaoqing. Rectifying algorithm for slanting online Chinese character based on the strokes' direction information[J]. Pattern Recognition and Artificial Intelligence, 2000, 13(4): 378-382.

[10] GARAI A, BISWAS S, MANDAL S. Dewarping of document images: A semi-CNN based approach[J]. Multimedia Tools and Applications, 2021, 80(28/29): 36009-36032.

[11] VARGAS-HÁKIM G A, MEZURA-MONTES E, ACOSTA-MESA H G. A review on convolutional neural network encodings for neuroevolution[J]. IEEE Transactions on Evolutionary Computation, 2022, 26(1): 12-27.

[12] XIE G W, YIN F, ZHANG X Y. Document dewarping with control points[C]//Proceedings of Document Analysis and Recognition—ICDAR 2021. Lausanne, Switzerland: Springer International Publishing, 2021: 466-480.

[13] DAS S, SINGH K Y, WU J. End-to-end piece-wise unwarping of document images[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE, 2021: 4248-4257.

[14] BANDYOPADHYAY H, DASGUPTA T, DAS N. RectiNet-v2: A stacked network architecture for document image dewarping[J]. Pattern Recognition Letters, 2022, 155: 41-47.

[15] MA K, SHU Z, BAI X. DocUNet: Document image unwarping via a stacked U-Net[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 4700-4709.

[16] SHAZIYA H, SHYAMALA K, ZAHEER R. Automatic lung segmentation on thoracic CT scans using U-Net convolutional network[C]//Proceedings of the 2018 IEEE International Conference on Communication and Signal Processing (ICCSP). New York: IEEE, 2018: 643-647.

[17] DAS S, MA K, SHU Z. DewarpNet: Single-image document unwarping with stacked 3D and 2D regression networks[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019: 131-140.

[18] BANDYOPADHYAY H, DASGUPTA T, DAS N. A gated and bifurcated stacked U-Net module for document image dewarping[C]//Proceedings of 2020 25th International Conference on Pattern Recognition (ICPR). Milan, Italy: IEEE, 2021: 10548-10554.

[19] LI X, ZHANG B, LIAO J. Document rectification and illumination correction using a patch-based CNN[J]. ACM Transactions on Graphics, 2019, 38(6): 168.

[20] 周涛, 刘赟璨, 陆惠玲, 等. ResNet及其在医学图像处理领域的应用:研究进展与挑战[J]. 电子与信息学报, 2022, 44(1): 149-167.
ZHOU Tao, LIU Yuncan, LU Huiling, et al. ResNet and its application to medical image processing: Research progress and challenges[J]. Journal of Electronics & Information Technology, 2022, 44(1): 149-167.

[21] RASHID M, KHAN M A, ALHAISONI M. A sustainable deep learning framework for object recognition using multi-layers deep features fusion and selection[J]. Sustainability, 2020, 12(12): 5037.

[22] DING Xiaohan, GUO Yuchen, DING Guiguang, et al. ACNet: Strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV 2019). Seoul, South Korea: IEEE, 2019: 1911-1920.

[23] LUO W J, LI Y J, URTASUN R, et al. Understanding the effective receptive field in deep convolutional neural networks[C]//Proceedings of Advances in Neural Information Processing Systems 29 (NIPS 2016). La Jolla: Neural Information Processing Systems (NIPS), 2016.

[24] ARAUJO A, NORRIS W, SIM J. Computing receptive fields of convolutional neural networks[EB/OL]. [2022-11-14]. https://distill.pub/2019/computing-receptive-fields/.

[25] ZHANG Z. Improved Adam optimizer for deep neural networks[C]//Proceedings of 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). Banff, AB, Canada: IEEE, 2018.

[26] SMITH R. An overview of the Tesseract OCR engine[C]//Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR 2007). Curitiba, Brazil: IEEE, 2007: 629-633.

[27] KIM B S, IL KOO H, CHO N I. Document dewarping via text-line based optimization[J]. Pattern Recognition, 2015, 48(11): 3600-3614.

作者简介:

冯瑾(1998-),女,硕士研究生,研究方向:数字图像处理、智能信息处理与机器学习,E-mail: fj031717xiamo@163.com。

池越(1977-),**通信作者**,男,副教授,硕士生导师,研究方向:数字图像处理、模式识别与人工智能,E-mail:chiyueliuxin@126.com。

周亚同(1973-),男,教授,博士生导师,研究方向:机器学习与模式识别、数字图像与视频处理、地震信号处理。

何静飞(1988-),男,副教授,博士生导师,研究方向:高维信号处理、机器学习。

(编辑:王静)