·图形图像处理·　　文章编号：1000-3428(2022)04-0231-09　　文献标志码：A　　中图分类号：TP751

基于改进U-Net的低质量文本图像二值化

王红霞[1]，何国昌[1]，李玉强[1]，陈德山[2]

（1.武汉理工大学 计算机科学与技术学院，武汉 430063；2.武汉理工大学 智能交通系统研究中心，武汉 430063）

摘　要：文本图像二值化是光学字符识别的关键步骤，但低质量文本图像背景噪声复杂，且图像全局上下文信息以及深层抽象信息难以获取，使得最终的二值化结果中文字区域分割不精确、文字的形状和轮廓等特征表达不足，从而导致二值化效果不佳。为此，提出一种基于改进U-Net网络的低质量文本图像二值化方法。采用适合小数据集的分割网络U-Net作为骨干模型，选择预训练的VGG16作为U-Net的编码器以提升模型的特征提取能力。通过融合轻量级全局上下文块的U-Net瓶颈层实现特征图的全局上下文建模。在U-Net解码器的各上采样块中融合残差跳跃连接，以提升模型的特征还原能力。从上述编码器、瓶颈层和解码器3个方面分别对U-Net进行改进，从而实现更精确的文本图像二值化。在DIBCO 2016—2018数据集上的实验结果表明，相较Otsu、Sauvola等方法，该方法能够实现更好的去噪效果，其二值化结果中保留了更多的细节特征，文字的形状和轮廓更精确、清晰。

关键词：文本图像二值化；U-Net网络；全局上下文；残差跳跃连接；DIBCO数据集

开放科学（资源服务）标志码（OSID）：

中文引用格式：王红霞，何国昌，李玉强，等.基于改进U-Net的低质量文本图像二值化[J].计算机工程，2022，48(4)：231-239.

英文引用格式：WANG H X，HE G C，LI Y Q，et al.Degraded document image binarization based on improved U-Net[J].Computer Engineering，2022，48(4)：231-239.

Degraded Document Image Binarization Based on Improved U-Net

WANG Hongxia[1]，HE Guochang[1]，LI Yuqiang[1]，CHEN Deshan[2]

（1.School of Computer Science and Technology，Wuhan University of Technology，Wuhan 430063，China；
2.Intelligent Transportation System Research Center，Wuhan University of Technology，Wuhan 430063，China）

【Abstract】Text image binarization is a key step in Optical Character Recognition（OCR）. However，the complex background noise of a degraded text image makes the extraction of its global context information and deep abstract information difficult.This results in an inaccurate segmentation of the text region and insufficient expression of features，such as text shape and contour，because of a poor binarization effect. Therefore，this paper proposes a binarization method for degraded text images based on an improved U-Net network.The segmentation network U-Net is suitable as a backbone model for small datasets，and a pretrained VGG16 is selected as the U-Net encoder to improve the model's feature extraction ability.The U-Net bottleneck layer of a lightweight global context block is fused to realize the global context modeling of feature graphs.The residual skip connection is fused in each upper sampling block of the U-Net decoder to improve the model's feature restoration ability.U-Net's improvement is based on the aforementioned three aspects：encoder，bottleneck layer，and decoder，to realize a more accurate text image binarization.Experimental results on the DIBCO 2016—2018 datasets show that the proposed method can achieve a better denoising effect，retain more detailed features in the binarization results，and provide a better delineation of characters in terms of accuracy and clarity than Otsu，Sauvola，and other methods.

【Key words】document image binarization；U-Net network；global context；residual skip connection；DIBCO dataset

DOI：10. 19678/j. issn. 1000-3428. 0060988

0　概述

低质量文本图像二值化是文本分析与识别领域的研究热点。纸质量文本在保存过程中容易受到物理条件的影响而出现墨水污渍、纸张破损、背景渗透等质量退化，或因人为破坏而产生污迹，以及因拍摄不当而导致光照不均等，严重影响了文本分析、字符识别等算法的处理效果。

基金项目：国家青年科学基金项目“基于多约束三维重构的低分辨率前视声呐目标探测研究”（51609193）。

作者简介：王红霞（1977—），女，副教授、博士，主研方向为模式识别、深度学习、图像处理；何国昌，硕士研究生；李玉强（通信作者），副教授、博士；陈德山，副研究员、博士。

收稿日期：2021-03-03　　**修回日期：**2021-04-30　　**E-mail：**15927579859@163.com

文本图像二值化是光学字符识别(OCR)技术的关键步骤,其目的是将文本图像中的文字信息从复杂图像背景中分离出来,以便通过OCR的后续步骤将目标更加集中于图像中的文字区域。对于简单规范的文本图像,其二值化较为容易,但是对于包含大量退化因素的低质量文本图像,实现其二值化则相对困难:一方面是低质量文本图像中背景噪声与文字糅杂在一起,使得背景噪声在二值化过程中易被误判为文字;另一方面则因为低质量文本图像中文字笔画粗细不一、文字的形状轮廓复杂,使得二值化过程中文字域与背景域难以精确区分。

低质量文本图像中复杂的文字区域难以确定,且图像中全局上下文信息和文字的形状、轮廓等复杂深层信息难以获取,从而导致文本图像二值化效果不佳。针对该问题,本文提出一种基于改进U-Net的低质量文本图像二值化方法。在U-Net的编码器部分使用预训练的VGG16完成图像下采样,保证下采样过程提取到足够的图像深层特征并提升模型的收敛速度;在U-Net的瓶颈层部分融合轻量级全局上下文模块实现对模型的全局上下文建模,使二值化结果保留更多的全局上下文信息;在U-Net的解码器部分向每个解码块中融合残差跳跃连接,使U-Net具有更强的特征还原能力,从而更好地恢复下采样过程所提取的深层特征。在此基础上,通过Sigmoid函数将各像素分为前景和背景两类以得到二值化图像。

1 相关工作

文本图像二值化方法总体可以分为传统法和深度学习法两类。传统法普遍基于阈值法实现,其中又分为全局阈值和局部阈值两种情况。在传统法类别中,全局阈值法是对输入图像中的所有像素点使用统一的阈值,典型代表有OTSU[1]算法、Kittler算法、简单统计法等,局部阈值法结合目标像素周边的灰度分布情况及其自身的灰度值共同计算局部阈值从而实现二值化,常见的局部阈值法包括Niblack算法、Sauvola[2]算法、Wolf算法等。

单一的阈值法精度较低,现已经很少被使用。目前,很多研究人员将阈值法与其他方法相结合以实现二值化。2015年,MANDAL等[3]提出基于形态学对比度增强的混合二值化技术,该技术使用灰度形态学工具来估计图像的背景,从而增加文本区域的对比度进而提升二值化性能。2016年,NAJAFI等[4]针对Sauvola算法计算量大以及对噪声敏感的问题,提出一种基于Sauvola算法随机实现的图像二值化方法,其提升了传统Sauvola算法的运行效率和噪声容错率。2017年,VATS等[5]提出自动化的文本图像二值化算法,该算法使用2个带通滤波方法来去除背景噪声,并使用贝叶斯优化来自动选择超参数以获得最佳的二值化结果。2018年,JIA等[6]利用结构对称像素(SSP)来计算局部阈值,并通过多个阈值的共同投票结果确定图像中某一像素是否属于前景。2019年,BHOWMIK等[7]在文本图像二值化中引入博弈论的思想,使用非零和的博弈提取图像局部信息并反馈给K-means分类器以实现像素分类。2020年,KAUR等[8]对Sauvola算法进行改进,使用笔划宽度变换自动地在图像像素之间动态计算窗口大小,减少了手动调整参数的数量,该改进方法适用于具有可变笔画宽度和文本大小的文本图像二值化。

近年来,随着深度学习技术的快速发展,大量基于深度学习的图像二值化方法被提出,使得二值化效果取得极大突破。2015年,PASTOR-PELLICER等[9]使用卷积神经网络训练低质量文本图像,实现了对文本图像中各像素的二分类。2016年,VO等[10]提出基于高斯混合马尔可夫随机场(GMMRF)的二值化方法,该方法可有效地对具有复杂背景的乐谱图像进行二值化。2017年,BRUN等[11]将卷积神经网络与图像分割算法相结合,以实现低质量文本图像二值化,其将语义分割的标签用作源和汇估计,并将概率图用于修剪图形切割中的边缘,大幅提升了二值化效果。2018年,VO等[12]提出基于深度监督网络(DSN)的文本图像二值化方法,该方法通过高层特征区分文本像素和背景噪声,通过浅层特征处理文字边缘细节等信息。2019年,ZHAO等[13]将生成式对抗网络应用于文本图像二值化,将图像二值化看作图像到图像的生成任务,并引入条件生成对抗网络(CGAN)来解决二值化任务中的多尺度信息组合问题,其进一步提升了图像二值化效果。

上述部分基于深度学习的图像二值化方法使用卷积神经网络完成文本图像的逐像素分类,从而实现图像二值化,其相较传统阈值法具有更优的分类精度。但传统卷积神经网络在图像分割时,缺乏对特征图全局信息的考虑,使得图像分割的结果图缺乏全局特性。对于文本图像二值化任务,这类方法容易使文本图像二值化结果存在文字笔画过度不自然、文字边缘像素分类不准确等问题。2015年,LONG等[14]提出只包含卷积层的神经网络模型FCN,其将图像分割问题看作像素级的分类问题。鉴于FCN在图像分割领域的优秀表现,研究人员开始将其用于文本图像二值化任务。在DIBCO 2017竞赛中,TENSMEYER等[15]提出基于FCN的文本图像二值化方法,该方法在比赛中取得了第四名的成绩,展现了全卷积网络模型优越的二值化性能。2015年,RONNEBERGER等[16]在FCN基础上提出一种改进的全卷积神经网络模型U-Net并用于医疗图像分割。U-Net采用对称的多尺度结构设计,使其在较小数据集上也能够取得高精度的分割效果。DIBCO 2017竞赛中的冠军方法将U-Net网络用于低质量文本图像二值化,其取得的成绩表明该网络非常适用于文本图像二值化分割任务。

2019年,熊炜等[17]将背景估计与U-Net网络进行融合并用于文本图像二值化,通过形态学闭操作来计算背景,并利用U-Net网络对图像进行背景与前景的分割,最后使用全局最优阈值法获得二值图。2020年,KANG等[18]将U-Net网络进行级联模块化,并将级联模块化后的U-Net用于文本图像二值化,该方法有效解决了低质量文本图像数据集规模小导致的训练不充分问题,获得了优异的二值化结果。同年,HUANG等[19]提出基于全局-局部U-Net的文本图像二值化方法,该方法将来自下采样图像的全局补丁和裁剪自源图像的局部补丁作为2个不同分支的输入,最后合并这2个分支的结果实现二进制预测。2020年,陈健[20]在U-Net上引入迁移学习方法实现低质量文本图像二值化,其在U-Net的编码器部分通过迁移学习方法迁移不同的模型作为编码器以完成图像下采样,并加载训练模型的预训练权重进行训练,解决了低质量文本图像数据集不足的问题并提升了模型的泛化能力。

综上，目前性能较好的二值化方法大多都基于U-Net实现，表明U-Net网络具有强大的二值化性能。本文对U-Net网络进行改进，针对低质量文本图像中背景干扰、噪声复杂，且受限于数据集和网络规模大小使图像全局上下文信息以及深层抽象信息难以获取，从而导致文本图像二值化效果不佳的问题，提出一种改进的低质量文本图像二值化方法。

2　本文方法

选择小数据集上表现优越的U-Net网络作为骨干模型，从编码器、瓶颈层和解码器3个部分对U-Net进行改进，以更好地满足低质量文本图像二值化的需求。编码器部分采用具有强大特征提取能力的小型网络VGG16作为特征提取器，加载VGG16预先训练好的权重进行训练；瓶颈层部分通过融合轻量级的全局上下文模块实现对文本图像的全局上下文建模；解码器部分通过融合残差跳跃连接提升特征恢复能力。将改进的U-Net模型用于低质量文本图像二值化，以期取得更精确的二值化效果。

2.1　编码器设计

本文去掉VGG16网络中最后2个全连接层，保留所有的卷积层和池化层，作为U-Net网络的编码器，其结构如图1所示。该编码器是一个含有5个尺度的下采样结构，具有很好的特征提取能力。图像的输入大小为256×256像素，通道数为3。图像通过编码器输入，每经过一个下采样编码块（Block），图像分辨率相应减小，同时图像维度增加。

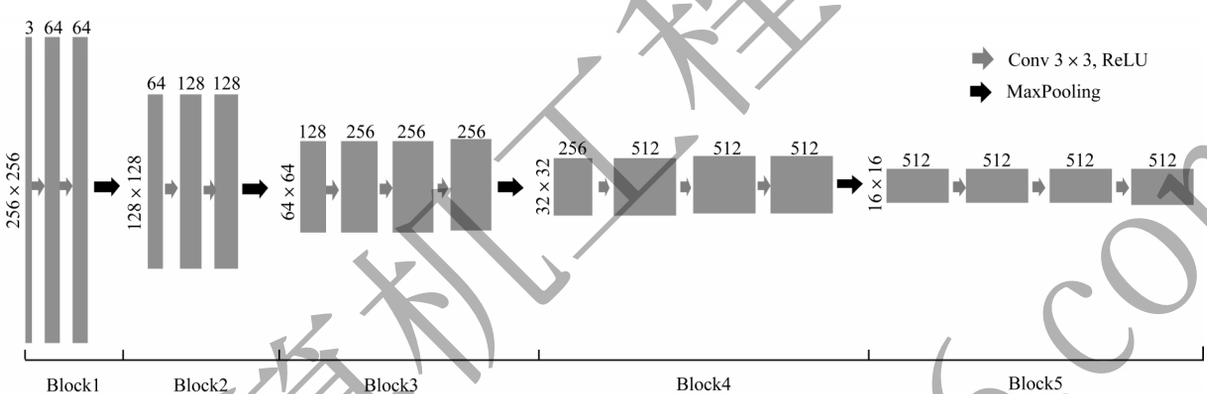


图1　编码器结构

Fig.1　Encoder structure

每个尺度的编码块提取对应尺度的二值化特征，得到不同表征的特征。浅层编码块提取高分辨率的浅层特征，如文字的边缘、纹理等细节特征，特征数量较多；深层编码块提取低分辨率的深层特征，如文字的形状、轮廓等抽象特征，这种特征抽象但较难被获取。低质量文本图像经过每个下采样编码块后输出的特征图如图2所示，图像从左往右依次为原图、Block1~Block5输出的特征图。

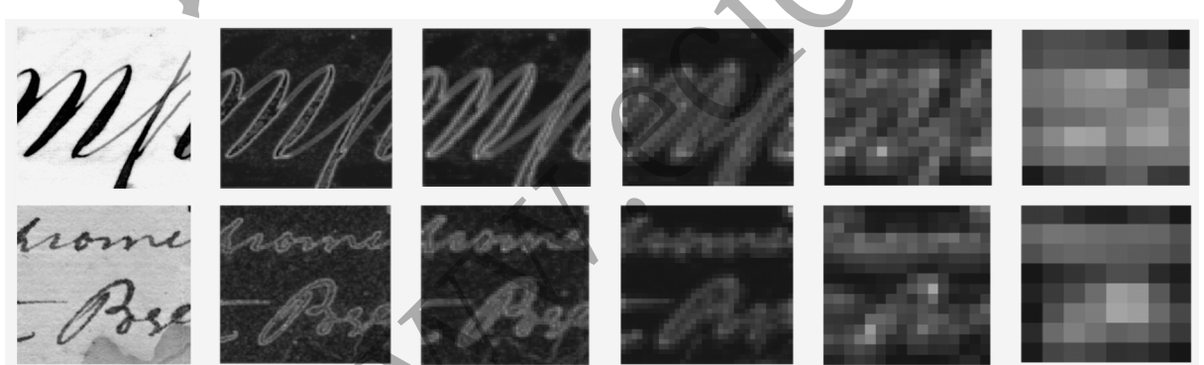


图2　编码器不同尺度的输出特征图

Fig.2　Output feature maps of encoder with different scales

2.2　瓶颈层设计

U-Net的编码器和解码器之间通过瓶颈层（bootleneck）连接，瓶颈层中包含编码过程中收集的高级语义信息，这种具有代表性的语义信息经由瓶颈层传播至解码器，并最终影响解码器还原特征图中的深层抽象特征信息。因此，瓶颈层结构对于图像分割具有重要意义。

由于低质量文本图像中存在大量的背景干扰噪声，因此判断图像中某一像素点属于文字还是背景不仅需要考虑该像素自身及其邻域的灰度值信息，更需结合图像的全局上下文信息作出综合判断。本节在U-Net的瓶颈层中引入一个轻量级的全局上下文块，构建一个可有效对图像全局上下文进行建模的U-Net瓶颈层，增强U-Net网络对长距离依赖信息的建模，从而提升U-Net的图像二值化性能。

融合轻量级全局上下文块的改进瓶颈层结构如图3所示，该瓶颈层结构可以有效地对低质量文本图像的全局上下文进行建模，从而充分学习瓶颈层高级语义

特征中的全局文字笔画区域信息,更好地确定文本图像中的文字区域,进而提升低质量文本图像二值化的精度。改进的瓶颈层结构保留了原方案中的2个卷积层用于提取更高层次的抽象语义特征,并在卷积层后紧跟一个轻量级全局上下文块用于实现对特征图的全局上下文建模,以捕获图像的长距离依赖关系。

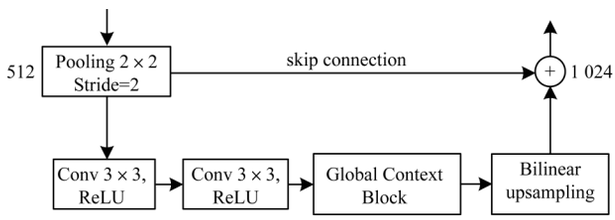


图3 融合全局上下文块的U-Net瓶颈层结构

Fig.3 U-Net bottleneck layer structure integrating global context block

轻量级全局上下文块的具体结构如图4所示,其基于文献[21]中提出的GC block构建,共由3个部分组成:1)用于上下文建模的全局注意力池,即图4中的Context Modeling部分;2)用于捕获通道依赖性的瓶颈层转换,对应图4中的Transform1部分;3)逐个广播元素相加的特征融合,对应图4中的Transform2部分。该轻量级全局上下文块是对文献[22]提出的NLNet的一种改进,其中,全局注意力池结构用于将图像中所有位置的特征聚合起来,构建特征图像的全局上下文特征;特征转换模块的作用是捕捉各通道间相互依存的关系;特征融合模块则用于将构建的全局上下文特征与不同位置的特征进行合并。全局上下文块中的训练参数全部来自1×1的卷积,并通过添加一个特征优化系数r用于进一步减少该模块中的参数量,因此,它是轻量级的,并且可以在多个卷积层之间应用,在不增加计算成本的情况下能更好地捕获图像的长距离依赖关系。

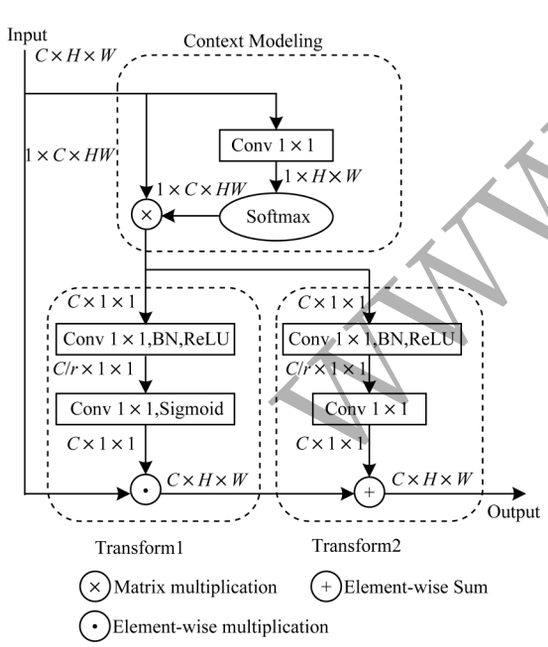


图4 全局上下文块结构

Fig.4 Structure of global context block

通过这种融合轻量级全局上下文块的瓶颈层设计替换原生U-Net中的瓶颈层结构,使得U-Net模型在训练低质量文本图像的过程中,更多地保留图像的全局上下文关系,从而更精确地区分文本图像中的文字区域与背景区域。

2.3 解码器设计

与编码器对应,解码器也由5个尺度的解码块构成,每个解码块均由若干卷积层和1个上采样层构成。解码器的作用是对编码器所提取的特征进行恢复,解码器的特征恢复能力直接决定二值化结果能否保留更多的文字细节和抽象特征,从而影响最终的二值化效果。

本文通过在解码块中融合残差跳跃连接并适当增加解码块卷积层数来提升解码器的特征恢复能力。残差跳跃连接是ResNet中引入的一种跳跃连接方式,该网络由何凯明等[23]于2015年提出,其模型引入了残差块的概念。残差网络在2015年的ILSVRC比赛中获得三项冠军,分别是图像的定位、检测与分类,性能远超其他模型。本文通过在解码块中引入残差跳跃连接,以减少卷积过程中的特征丢失,提高解码器的特征恢复能力。

残差块结构如图5所示,通过跳跃连接方式将浅层网络的输出直接传输到更深层以作为后续卷积输入的一部分,从而有效减少卷积过程中的特征丢失,提高特征利用率。残差跳跃连接的叠加方式为图像像素值叠加,必须保证叠加的2张图像具有相同尺度。当图像维度相同时为图5(a)所示的基本残差结构,维度不相同时则需先通过一个一维卷积进行维度转换后再叠加,如图5(b)所示。这2种残差结构根据具体情况分别被添加在U-Net的各解码块中,然后构建融合残差跳跃连接的解码器,其结构如图6所示。图6中解码块上采样后的图像通过U-Net的横向跳跃连接与对应尺度编码块的下采样输出堆叠,导致解码块中第一次卷积前后的图像维度不一致,无法直接添加基本残差跳跃连接,因此,本文设计成带一维卷积的残差跳跃连接来完成图像叠加,如图6中橙色的残差跳跃连接所示(彩色效果见《计算机工程》官网HTML版)。在完成第一次卷积后,图像维度与解码块卷积维度保持一致,此时,在后续卷积中再添加一个基本残差连接,如图6中蓝色的残差跳跃连接所示。

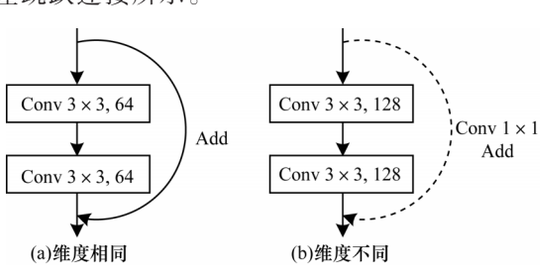


图5 残差块结构

Fig.5 Structure of residual block

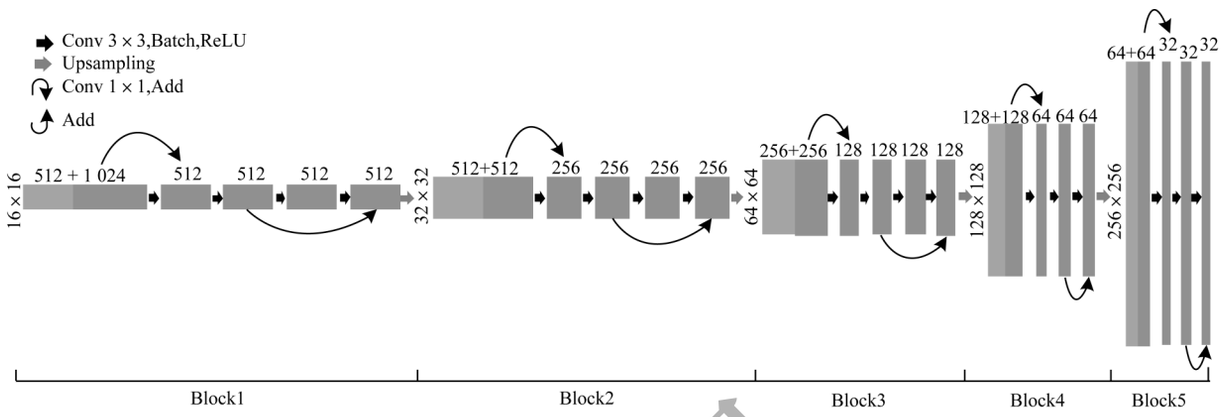


图6　融合残差跳跃连接的解码器结构

Fig.6　Decoder structure integrating residual ship connection

与编码器对应,解码器也被设计为包含5个解码块的结构,每个解码块均按图6方式设置2个残差跳跃连接。相比于对应尺度的编码块,每个解码块增设一个卷积层来提升特征恢复能力。同时,为了防止过拟合并提升模型泛化能力,在解码器的所有卷积之后都采用Batch Normalization进行规范化处理,各解码块最后通过Dropout进行强正则化输出。

融合残差跳跃连接的解码器相比常规解码器具有更好的特征恢复能力,可在二值化过程中还原更多的文字特征。常规解码器和融合残差跳跃连接的解码器所输出的特征图对比如图7所示,其中从左往右依次为原图、常规解码器输出、融合残差跳跃连接的解码器输出以及标准二值化图像。从图7可以看出,本文解码器还原的特征图保留了更多的文字区域信息,第一张图片右下角对比尤为明显,且文字轮廓更清晰,背景更纯净,说明融合残差跳跃连接的解码器比常规解码器具有更强的特征恢复能力,更好地凸显了文本图像二值化结果中文字的区域信息,从而改善了二值化效果。

图7　不同解码器的输出特征图对比

Fig.7　Comparison of output feature maps of different decoders

利用改进的编码器、瓶颈层和解码器构建U-Net网络,设计基于改进U-Net的低质量文本图像二值化方法。低质量文本图像首先在迁移VGG16的编码器中完成图像下采样;然后经过融合全局上下文块的瓶颈层结构,完成全局上下文建模;接着从融合残差跳跃连接的解码器中实现图像上采样;最后通过Sigmoid激活层输出得到最终的二值化图像。图8所示为改进的U-Net网络结构。

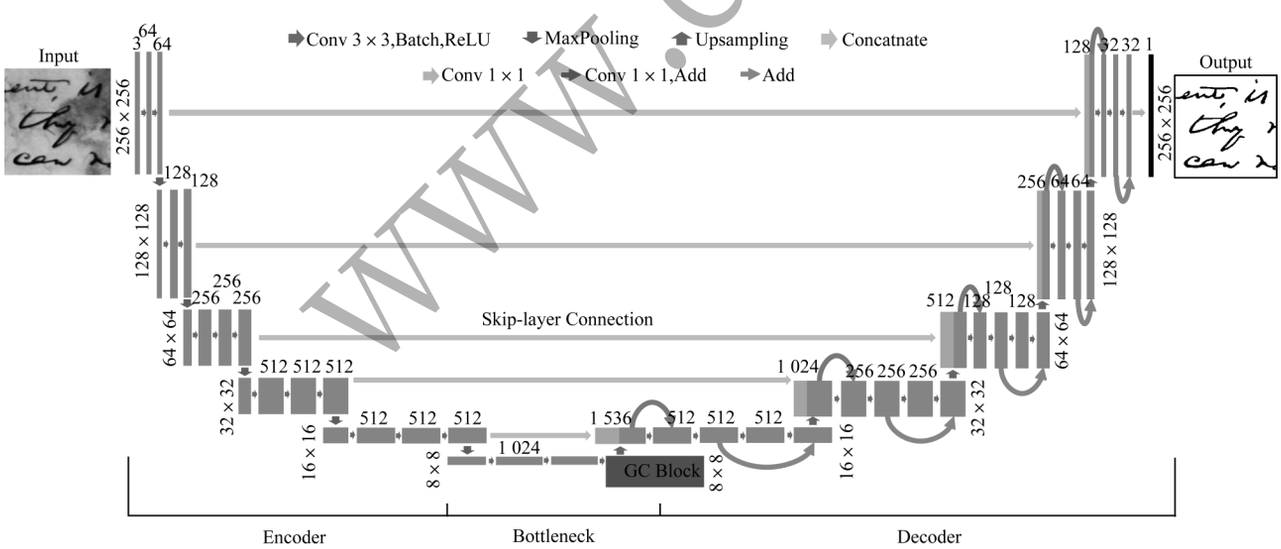


图8　改进的U-Net网络结构

Fig.8　Improved U-Net network structure

3 实验结果与分析

3.1 二值化评估指标

图像二值化结果的优劣由4项重要指标评估，分别为图像的F值（FM）、伪F值（p-FM）、峰值信噪比（PSNR）和距离倒数失真度量（DRD）。

3.1.1 图像F值

图像F值记为FM，其值大小与文本图像二值化优劣正相关，计算公式如下：

$$F_{\mathrm{FM}}=\frac{2\times P_{\mathrm{Precision}}\times R_{\mathrm{Recall}}}{R_{\mathrm{Recall}}+P_{\mathrm{Precision}}} \tag{1}$$

$$R_{\mathrm{Recall}}=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{N}}} \tag{2}$$

$$P_{\mathrm{Precision}}=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{P}}} \tag{3}$$

其中：T_{P}、F_{P}和F_{N}分别表示真实的正值、错误的正值和错误的负值。

3.1.2 图像伪F值

图像伪F值记为p-FM，其值大小与二值化结果优劣正相关，计算公式如下：

$$P_{\text{p-FM}}=\frac{2\times P_{\text{p-Recall}}\times P_{\mathrm{Precision}}}{P_{\text{p-Recall}}+P_{\mathrm{Precision}}} \tag{4}$$

其中：$P_{\text{p-Recall}}$定义为二值化图像与标准图像中字符结构的百分比。

3.1.3 图像峰值信噪比

图像峰值信噪比记为PSNR，其反映2张图像的相似度，PSNR值大小与图像二值化结果优劣正相关，计算公式如下：

$$P_{\mathrm{PSNR}}=10\log_a\left(\frac{C^2}{M_{\mathrm{MSE}}}\right) \tag{5}$$

$$M_{\mathrm{MSE}}=\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y)-I'(x,y))^2}{MN} \tag{6}$$

3.1.4 图像距离倒数失真度量

图像距离倒数失真度量记为DRD，其值大小与二值化效果优劣反相关，计算公式如下：

$$D_{\mathrm{DRD}}=\frac{\sum_{k=1}^{S}D_{\mathrm{DRD}_k}}{N_{\mathrm{NUBN}}} \tag{7}$$

$$D_{\mathrm{DRD}_k}=\sum_{i,j}[D_k(i,j)\times W_{Nm}(i,j)] \tag{8}$$

其中：D_{DRD_k}是第k个翻转像素的失真；N_{NUBN}是标准图像中不均匀色块的数量。图像的距离倒数失真度量表示图像的视觉失真程度。

本文用训练得到的模型对测试集图像进行二值化，然后用评价工具测出每张图像的各项二值化指标，最后对所有图像的指标求平均得到最终结果。输出保存的模型为训练过程中FM值最佳的模型。

3.2 结果分析

为了验证本文改进U-Net模型对低质量文本图像二值化的有效性，分别在DIBCO 2016—2018这3个数据集上和其他优秀方法进行性能对比。本文模型在训练过程中，设置epoch为256，batch size为16，学习率起始为1e-4并动态调整。将图像FM值设为模型的损失函数，每个epoch训练完毕计算FM值，若20个epoch内FM值没有增加，则将学习率乘以0.1。训练集为2009—2018的DIBCO数据集，在训练过程中，为了保证对比的公平性，当测试某个数据集时，将该年份之前的所有年份数据集作为训练样本进行训练，输出权重模型，然后加载权重模型对目标数据集进行二值化得到二值化图像，最后采用二值化评估工具对二值化图像进行指标评估。评估结果分别与DIBCO 2016—2018获胜者以及其他研究人员针对这3个数据集的经典、最新和最优成果进行对比与分析，包括二值化指标数据对比和二值化效果对比。

3.2.1 二值化评估指标对比分析

将本文方法的二值化评估指标结果与经典方法、目前最优方法的评估指标结果进行对比，对比方法包括对应年份的DIBCO竞赛冠军方法、Otsu[1]、Sauvola[2]、文献[12]方法、文献[13]方法、文献[17]方法、文献[18]方法、文献[19]方法和文献[20]方法，其中，文献[20]方法包括未添加预处理和添加预处理2种情况。实验对比数据均来自各方法的原始文献。各方法在DIBCO 2016—2018这3个数据集上的二值化评估指标对比结果分别如表1~表3所示，加粗表示最优结果。

表1 DIBCO 2016数据集上的实验结果对比

Table 1 Comparison of experimental results on DIBCO 2016 dataset

方法	FM	p-FM	PSNR/dB	DRD
Otsu[1]	86.61	88.67	17.80	5.56
Sauvola[2]	82.52	86.85	16.42	7.49
DIBCO 2016 winner	87.61	91.28	18.11	5.21
文献[12]方法	90.10	93.57	19.01	3.58
文献[13]方法	91.66	94.58	19.64	**2.82**
文献[18]方法	**93.09**	**94.85**	19.18	3.03
文献[19]方法	90.77	94.21	19.33	3.11
文献[20]方法(未预处理)	90.28	93.48	19.17	3.27
文献[20]方法(预处理)	90.83	93.58	19.28	3.12
本文方法	91.71	93.96	**19.73**	2.92

表2 DIBCO 2017数据集上的实验结果对比

Table 2 Comparison of experimental results on DIBCO 2017 dataset

方法	FM	p-FM	PSNR/dB	DRD
Otsu[1]	77.73	77.89	13.85	15.54
Sauvola[2]	77.11	84.10	14.25	8.85
DIBCO 2017 winner	91.04	92.86	18.28	3.40
文献[13]方法	90.73	92.58	17.83	3.58
文献[18]方法	91.57	93.55	15.85	2.92
文献[19]方法	92.14	93.02	18.71	2.81
文献[20]方法(未预处理)	91.74	93.14	18.43	2.68
文献[20]方法(预处理)	92.61	**94.91**	18.96	**2.35**
本文方法	**92.81**	94.12	**18.97**	2.49

表3　DIBCO 2018数据集上的实验结果对比

Table 3　Comparison of experimental results on DIBCO 2018 dataset

方法	FM	p-FM	PSNR/dB	DRD
Otsu[1]	51.54	53.05	9.74	59.07
Sauvola[2]	67.81	74.08	13.78	17.69
DIBCO 2018 winner	88.34	90.24	19.11	4.92
文献[13]方法	87.73	90.60	18.37	4.58
文献[18]方法	89.71	91.62	19.39	2.51
文献[19]方法	92.10	93.02	18.71	2.81
文献[20]方法(未预处理)	91.47	95.21	19.76	2.68
文献[20]方法(预处理)	**92.52**	**95.87**	**20.32**	**2.26**
本文方法	92.34	94.80	20.03	2.44

由表1可知，在DIBCO 2016数据集上，文献[18]方法综合二值化性能最优，该方法基于级联模块化后的U-Net实现，在FM、p-FM这2项指标上取得最优。文献[13]方法拥有最佳的DRD指标，该方法基于生成式对抗网络实现，视觉失真程度最低。本文方法拥有最佳PSNR指标及排名第二的FM指标和DRD指标，说明该方法的二值化结果与标准二值化结果的相似度最高，具有很低的视觉失真程度，其综合二值化性能仅次于文献[18]方法，比添加了图像预处理的文献[20]方法二值化性能更好。

由表2可知，在DIBCO 2017数据集上，在所有对比方法中，本文方法的FM和PSNR这2个指标最佳，p-FM和DRD这2项指标只略低于添加了预处理的文献[20]方法，这是由于文献[20]方法的预处理方式对输入图像做了锐化和光照补偿，使文字与背景区分度更高，而本文方法在没有做预处理的情况下，FM和PSNR这2个指标仍高于添加预处理后的文献[20]方法，综合二值化性能与其相当，优于其他对比方法。

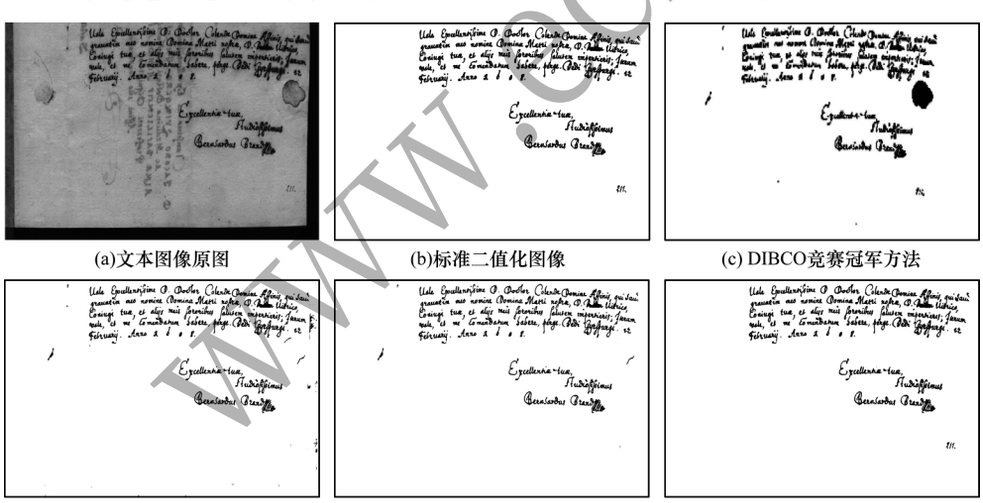
由表3可知，在DIBCO 2018数据集上，本文方法的二值化性能仅低于添加了预处理的文献[20]方法，这仍然是由于后者的图像预处理所带来的性能提升。相较其他对比方法，本文方法的各项二值化指标均具有较大幅度的提升。

通过DIBCO 2016—2018这3个数据集的测试对比可以看出，本文方法对低质量文本图像具有良好的二值化性能，验证了改进U-Net模型的有效性。

3.2.2　二值化效果图对比分析

为直观地展示本文方法的二值化效果，将本文方法与其他代表性方法的二值化效果图进行对比。

图9所示为各对比方法在DIBCO 2016—2018数据集上的部分二值化效果图。从图9可以看出：DIBCO竞赛冠军方法的二值化结果中保留了较多的背景噪声，且将图中某一渗透的墨水污渍误判为文字，未实现良好的去噪效果；未结合图像预处理的文献[20]方法的二值化效果良好，但仍保留了较多的背景噪声，在结合图像预处理后其整体二值化效果得到了较明显的提升，大部分噪声被有效去除；本文方法很好地将低质量文本图像的文字和背景分离，去除了更多的背景噪声，识别出较对比方法更精确、清晰的文字区域。



(a)文本图像原图　(b)标准二值化图像　(c) DIBCO竞赛冠军方法

(d)未结合预处理的文献[20]方法　(e)结合预处理的文献[20]方法　(f)本文方法

图9　不同方法的二值化效果对比

Fig.9　Comparison of binarization effect of different methods

本文方法所得的二值化效果图如图10所示。从图10可以看出，本文方法所得的二值化图像与标准二值化图像相似度很高，文本图像的大部分背景噪声都被有效去除，文字的形状和轮廓清晰。

(a)文本图像原图

(b)标准二值化图像

(c)本文方法二值化结果

图10 本文方法的二值化效果图

Fig.10 The binarization effect maps of this method

4 结束语

本文提出一种基于改进U-Net的低质量文本图像二值化方法。从编码器、瓶颈层和解码器3个方面对传统U-Net网络进行改进并用于低质量文本图像二值化，改进的U-Net网络不仅具有更优的特征提取能力和特征还原能力，使得图像二值化结果中保留更丰富的文字细节，同时还具有更佳的全局上下文建模能力，可在图像二值化过程中实现更好的去噪效果。实验结果验证了本文方法良好的图像二值化性能。

尽管本文方法取得了较好的文本图像二值化效果，但仍存在一定的性能提升空间。该方法通过在U-Net解码器中融合残差跳跃连接，为文本图像二值化结果保留更多的深层特征，但图像在下采样过程中分辨率逐渐减小，伴随着大量浅层特征的丢失，使得图像二值化结果中文字的边缘纹理等浅层特征细节表达不足。虽然U-Net网络通过横向跳跃将相同尺度的下采样输出与上采样输入进行堆叠，较大程度地减少了浅层特征丢失，但这部分浅层特征未被训练，不具有良好的泛化性。因此，如何更好地提取和利用文本图像浅层特征以丰富二值化结果中文字的笔画、纹理等细节，将是下一步的研究方向。

参考文献

[1] OTSU N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.

[2] SAUVOLA J, PIETIKÄINEN M. Adaptive document image binarization[J]. Pattern Recognition, 2000, 33(2): 225-236.

[3] MANDAL S, DAS S, AGARWAL A, et al. Binarization of degraded handwritten documents based on morphological contrast intensification[C]//Proceedings of ICIIP'15. Washington D. C., USA: IEEE Press, 2015: 73-78.

[4] NAJAFI M H, SALEHI M E. A fast fault-tolerant architecture for sauvola local image thresholding algorithm using stochastic computing[J]. IEEE Transactions on Very Large Scale Integration(VLSI) Systems, 2016, 24(2): 808-812.

[5] VATS E, HAST A, SINGH P. Automatic document image binarization using Bayesian optimization[C]//Proceedings of the 4th International Workshop on Historical Document Imaging and Processing. New York, USA: ACM Press, 2017: 12-23.

[6] JIA F X, SHI C Z, HE K, et al. Degraded document image binarization using structural symmetry of strokes[J]. Pattern Recognition, 2018, 74: 225-240.

[7] BHOWMIK S, SARKAR R, DAS B, et al. GiB: a game theory inspired binarization technique for degraded document images[J]. IEEE Transactions on Image Processing, 2019, 28(3): 1443-1455.

[8] KAUR A, RANI U, JOSAN G S. Modified Sauvola binarization for degraded document images[J]. Engineering Applications of Artificial Intelligence, 2020, 92: 103672.

[9] PASTOR-PELLICER J, ESPAÑA-BOQUERA S, ZAMORA-MARTÍNEZ F, et al. Insights on the use of convolutional neural networks for document image binarization[C]// Proceedings of International Work-Conference on Artificial Neural Networks. Berlin, Germany: Springer, 2015: 115-126.

[10] VO Q N, KIM S H, YANG H J, et al. An MRF model for binarization of music scores with complex background[J]. Pattern Recognition Letters, 2016, 69: 88-95.

[11] AYYALASOMAYAJULA K R, BRUN A. Historical document binarization combining semantic labeling and graph cuts[M]//SHARMA P, BIANCHI F. Image analysis. Berlin, Germany: Springer, 2017.

[12] VO Q N, KIM S H, YANG H J, et al. Binarization of

degraded document images based on hierarchical deep supervised network[J]. Pattern Recognition, 2018, 74: 568-586.

[13] ZHAO J Y, SHI C Z, JIA F X, et al. Document image binarization with cascaded generators of conditional generative adversarial networks[J]. Pattern Recognition, 2019, 96: 106968.

[14] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 3431-3440.

[15] TENSMEYER C, MARTINEZ T. Document image binarization with fully convolutional neural networks[C]//Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition. Washington D. C., USA: IEEE Press, 2017: 99-104.

[16] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[M]// NAVAB N, HORNEGGER J, WELLS W, et al. Medical image computing and computer-assisted intervention. Berlin, Germany: Springer, 2015.

[17] 熊炜，王鑫睿，王娟，等. 融合背景估计与U-Net的文档图像二值化算法[J]. 计算机应用研究，2020，37(3)：896-900.
XIONG W, WANG X R, WANG J, et al. Document image binarization algorithm based on background estimation and U-Net[J]. Application Research of Computers, 2020, 37(3): 896-900. (in Chinese)

[18] KANG S, IWANA B K, UCHIDA S. Complex image processing with less data—document image binarization by integrating multiple pre-trained U-Net modules[J]. Pattern Recognition, 2021, 109: 107577.

[19] HUANG X, LI L, LIU R, et al. Binarization of degraded document images with global-local U-Nets[J]. Optik, 2020, 203: 164025.

[20] 陈健. 基于全卷积网络的低质量文档图像二值化方法研究[D]. 武汉：武汉理工大学，2019.
CHEN J. Research on degraded document image binarization methods based on fully convolutional networks[D]. Wuhan: Wuhan University of Technology, 2019. (in Chinese)

[21] CAO Y, XU J R, LIN S, et al. GCNet: non-local networks meet squeeze-excitation networks and beyond [C]// Proceedings of IEEE/CVF International Conference on Computer Vision Workshop. Washington D. C., USA: IEEE Press, 2019: 1971-1980.

[22] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 7794-7803.

[23] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 770-778.

编辑　吴云芳