文章编号：2095-2163(2020)11-0154-05 中图分类号：TP 391 文献标志码：A

基于深度学习方法的手写文本行提取综述

杨益暄，田益民，崔圆斌，齐千慧，韩利利
（北京印刷学院 信息工程学院，北京 102600）

摘 要：文本行提取一直是手写文档图像分析与识别领域的热点研究课题。随着深度学习的发展，越来越多的方法涌现出来，通过对近几年的相关文献分析整理，本文按照全卷积神经网络、编解码器、循环神经网络、生成对抗网络的基于深度学习的手写文本行提取方法进行总结和分析，列举了每种方法的代表性实例，并对常用训练数据集进行了介绍。分析了各类方法的特点与不足，并对未来可研究方向进行展望。

关键词：手写文档图像；深度学习；手写文本行提取

A Review of Handwritten Text Line Extraction Based on Deep Learning Methods

YANG Yixuan, TIAN Yimin, CUI Yuanbin, QI Qianhui, HAN Lili
(Beijing Institute of Graphic Communication, Beijing 102600, China)

【Abstract】 Text line extraction, as a crucial step in image analysis and recognition of handwritten documents, has always been a hot research topic in this field. With the development of deep learning, more and more methods have emerged. Through the analysis and sorting of relevant documents in recent years, this article is based on deep learning based on full convolutional neural networks, codecs, recurrent neural networks, and generative adversarial networks. Summarize and analyze the handwritten text line extraction methods, list representative examples of each method, and introduce commonly used training data sets. Finally, the four types of methods listed are summarized, the characteristics and shortcomings of each method are analyzed, and the future research directions are prospected, and three suggestions are put forward.

【Key words】 handwritten document image; deep learning; handwritten text line extraction

0 引 言

信息技术的飞速发展无时无刻不在影响着人们的生活。多媒体数字网络的迅猛发展，使传统的纸质媒体受到了剧烈的冲击。然而仍然有很多价值珍贵的或对个人而言意义重大的资料保存记录在纸质媒介上。如杂志、手写文稿、历史文稿等。当在成千上万的文档中搜索信息时，纸质文档的不适用性就会体现出来。文档分析与识别(Document Analysis and Recognition, DAR)[1]系统是一项复杂的工程任务，其一般包含文本与非文本部分分离、文本行或单词提取、文本倾斜和偏移的校正、字符或单词识别等步骤。其中，文本行是文档图像中的重要实体，因此正确提取文档中的文本行将直接影响单词或字符识别的准确程度。

当今流行的文本行提取方法分为两大类：传统方法和学习方法。传统手写文本行提取方法主要基于机器学习和启发式算法，又可细分为3类[5]：自顶向下的方法、自底向上的方法和混合型方法。自顶向下的方法对文本行的字符序列部分进行分组，并将属于不同文本行的各个组成部分进行拆分，从而实现文本行的定位。如文献[6]，用接缝裁剪法获取图像特征后，通过计算能量图分割文本行；Alaei[7]等人则先确定行间间隙，然后在每个间隙上应用分段过滤，最后使用细化算法分割手写文本行。自底向上的方法则是根据局部特征对像素或相连的部分进行分组，以形成文本行。如，文献[8]使用超像素法获得图像连通区域(Connected components, CCs)，并通过最小化能量函数对 CCs 进行分割提取文本行。Ryu 等人[9]通过改进文献[8]的 CCs 分割方法和最小化能量函数，克服了少量 CCs 无法正确提取文本行的问题，提出了一种不受语言影响的文本行提取方法。混合型方法则将自顶向下和自底向上文本行提取方法相结合。如，Louloudis 等人[10]采用了先得到 CCs 后，再用 Hough 变换检测文本行，最后通过后期处理以纠正错误。Deshmukh 等人[11]则利用全局阈值和各行的局部阈值分割文本行，再

基金项目：国家自然科学基金(6378001)。

作者简介：杨益暄(1996-)，男，硕士研究生，主要研究方向：图像处理、深度学习；田益民(1966-)，男，博士，教授，硕士生导师，主要研究方向：数字图像处理与数值分析。

通讯作者：田益民 Email：tym66105@163.com

收稿日期：2020-09-17

通过后期基于分治和水平投影的方法，从剩余未分割的区域中提取文本行。

时至今日，已有许多基于传统机器学习的文本行提取方法。如基于 Hough 变换的方法[2]、基于模拟水流的方法[3]、基于图像接缝裁剪[4]方法等等。而随着深度学习的兴起，越来越多的工作者结合深度学习开展相应的研究。

本文将针对目前主要的基于深度学习的文本行提取方法分析研究，并对常用的数据集做简单介绍。

1　基于深度学习提取方法

随着深度学习的兴起，人们注意到基于深度学习的文本行提取方法可以解决许多传统方法面临的问题。许多研究人员指出，从文档中提取文本行最有效的方法是搭建深度神经网络，大量的参数和网络隐层数使其拥有很强的非线性拟合能力以及自学习能力。此外，基于深度学习的方法在应对文档图像中的不同噪声和古籍纸张自然老化或污损有着较好的鲁棒性。

1.1　基于全卷积神经网络方法

图像分割作为计算机视觉领域的三大任务之一，一直存在着极大的挑战性。由 Long 等人[12]提出的全卷积神经网络(Fully Convolutional Networks，FCN)在这方面取得了显著的进步。FCN 将传统卷积神经网络的全连接层替换为卷积层，实现了单张图像像素级的分类，从而解决语义级别的图像分割(semantic segmentation)问题。FCN 分为二部分：卷积层和反卷积层。卷积层可以接受任意尺寸的输入图像，之后采用反卷积层对最后一个卷积层产生的特征图进行上采样，使它恢复到输入图像相同的尺寸，从而预测每一个像素的类别，同时保留了原始输入图像中的空间信息，最后在上采样的特征图上进行逐像素分类。

由于 FCN 在图像分割领域的优秀表现，研究人员将此网络应用于文档或历史手稿图像的文本行提取中。FCN 作为一种端到端的图像分割方法，可以通过反卷积层得到的热图并使用不同的分割方法来提取文本行。Vo 等人[13]通过 FCN 对手写文档图像进行了文本行提取；Baraket 等人[14]同样使用 FCN 对具有挑战性历史手稿图像进行了文本行提取，得到了比传统方法更好的效果。但是，原始的 FCN 结构在反卷积过程对图像细节的处理不到位，丢失了许多细节信息。Renton 等人[15-16]分别对比了反卷积、上池化和空洞卷积在手写文本行提取的应用效果，发现空洞卷积增大感受野，提高了对文本信息的识别精度。因此提出了一种新的架构，将卷积层和最大池化层替换为空洞卷积。此外引入 X 高度作为文本行的标签进行训练，减少文本行之间字符粘连的影响，在所用数据集上达到了不错的效果。

1.2　基于编解码器方法

为了实现医学图像的分割，Ronneberger 等人[17]于 2015 年提出一种编解码结构的网络模型 U-Net。U-Net体系结构由两个对称部分组成，即收缩路径和扩展路径。收缩路径进行特征提取，扩展路径通过组合从收缩路径捕获的图像上下文信息来保证准确定位。U-Net 体系结构既充当编码器又充当解码器。U-Net 作为 FCN 的变体，可以将可变大小的图像作为 U-Net 结构的输入，而且，训练阶段不需要大量的图像。另外，U-Net 在对文档图像语义分割的多项工作中显示出有效的效果。

基于原始的 U-Net，Mechi 等人[18]提出了一种自适应 U-Net 结构的历史手稿图像文本行分割方法。该方法在解码器阶段使用反卷积操作，以在网络架构的输入和输出上保持相同的分辨率。同时将原网络结构收缩路径所设置的卷积核减少到一半，消除训练阶段的过度拟合问题。Gruning 等人[19]提出了一种基于 ARU-Net 的历史手稿文本行检测方法，该架构是 U-Net 的扩展。通过注意模型和残差结构构建 U 型结构，旨在及时处理任意大小的图像，以考虑所有空间上下文信息。其使用的空间注意机制允许 ARU-Net 专注于不同位置和比例的图像内容。此外，还可以从头开始训练。利用数据增强方法，不需要过多地手动标注示例图片。Neche 等人[20]提出了结合循环神经网络的 RU-Net，这种方法比 ARU-Net 训练简单，仅需要较少的处理步骤，即可达到更好的效果。

1.3　基于循环神经网络方法

循环神经网络(Recurrent Neural Network，RNN)于 20 世纪 80 年代提出，随着不断地改进和 GPU 性能的提升，逐渐在自然语言处理、目标检测等方面取得了诸多成果。

基于 RNN 的手写文本行提取方法受到目标检测方法的启发，结合 CNN 和根据 RNN 改进的长短期记忆神经网络(Long Short-Term Memory，LSTM)对文本行进行定位检测。Moysset 等人[21-22]对文本行周围的 bounding box 进行打分，再利用分类器定位每个文本行的起点并标记，最后得到文本行的边界框。在文献[23]中，Moysset 通过 MLSTM 改良了之前的方法，提高了这种定位方法的精度，对具有高

度差异性的数据集进行测试,显示出了良好的效果。

1.4 基于生成式对抗网络方法

生成式对抗网络(Generative Adversarial Networks, GAN)于 2014 年由 Goodfellow 等人[24]提出。GAN 基于零和博弈的思想,构造出一个生成器和一个判别器。生成器从随机信号分布中合成一些有意义的数据矩阵,判别器则区分真实分布和虚假分布,通过不断的对抗来优化网络的结构。目前,GAN 已经在图像编辑、图像生成、视频预测、图像超分辨率等诸多领域大放异彩。

由于 GAN 架构优秀的生成能力,Kundu 等人[25]首次将 GAN 引入文本行提取领域。受 Isola 等人[26]提出的 pix2pix 启发,以 Encoder-Decoder 和 U-Net 分别作为两个 GAN 的生成器,以 Patch-GAN 作为判别器。在实验过程中,以 U-Net 为生成器的 GAN 在迭代对抗训练中能够更精确的分割文本图像。这种 pix2pix 结构有效地学习了文本行的特征,为手写文本行提取领域注入了新的思路。但 GAN 对输入的超参数极其敏感。此外,则需要更多的数据集利用其他方法和人工来标注真实标签依然是应用方面的关键问题。

2 数据集

本节简要整理了常用的基于深度学习手写文本行提取应用的数据集,见表 1。

表 1 数据集

Tab. 1 Dataset introduction

名称	图像数量	备注
ICDAR 2013 手写分割竞赛数据集[27]	200 张训练集图像和 150 张测试集图像	含有英语、希腊语和孟加拉语的二值图像,英语和希腊语图像共 250 张,孟加拉语图像 100 张。
HIT-MW[28]	853 幅中文手写图像和 8 664 文本行	图像使用 Otsu 算法进行二值化,并保存为 bmp 图像。
DIVA-HisDB[29]	150 幅带注释的手稿图像	历史手稿数据库,包括 3 种具有挑战性中世纪手搞
cBAD[30]	2 036 幅手写档案文档图像	具有高分辨率(3000×4000),但质量差异很大。
KHATT[31]	两组 2 000 个简短的手写段落	该数据库由阿拉伯手写文档图像组成,由 1 000个来自不同年龄、性别和国籍的人书写
IAM-HisDB[32]	包含 47 页的 Parzival 中世纪德国手稿,60 页瑞士圣高尔大教堂手稿和 20 页乔治华盛顿的手稿。	数据量小,但具有高分辨率,质量差异明显。

当需要对算法的可行性进行验证时,可选用国际文档分析与识别会议(International Conference on Document Analysis and Recognition, ICDAR)的分割竞赛和 HIT-MW 等数据集。此类数据手写文本排列整齐,图像噪声和伪影较少,预处理方法简单。当实验目的为具有挑战性的历史手稿时,可选用 DIVA-HisDB 类数据集。

3 结束语

文本行提取领域经过了几十年的发展,虽然已经拥有长足的发展和实用的算法,但在大数据时代面对海量的文档图片数据仍然捉襟见肘,尤其对于历史手稿的图片处理更是一大难题,时下大热的深度学习为该领域探索了新的出路。基于深度学习的手写文本行提取,涵盖了各种不同的方法,每种算法都有各自的特点。RNN 根据目标检测的原理对文本行进行定位,这种方法新颖而且不需要标记文本行的边界,但其缺点也很明显,在处理繁重的任务时无法起到更好的效果,并且对于历史手稿类的图像,难以提取其倾斜的甚至曲线状的文本行。FCN 作为计算机视觉领域著名的图像分割网络能够端到端对图像分割,易于对布局较为简单的普通手写文本图像进行提取。但其反卷积过程中对图像粗糙处理的缺点会在文本行分割之后丢失文字的细节信息。对于不同语言文字的保存会减少准确性,而之后的文字或单词提取也会面临诸多困难。与 FCN 相比,U-Net 在上采样阶段进行了比较大的改动,结合了下采样时的低分辨率信息和上采样时的高分辨率信息提高分割精度。GAN 的方法则是结合了纳什均衡和图像分割的思想对手写文本行进行提取。从以上方法可以看出,FCN、U-Net 等基于分割的文本行

提取方法是深度学习方法的主流。

目前的方法在一定程度上达到了需求，但仍有很大的提升空间。以下提出 3 点对未来研究的展望：

(1)本文提及的 4 种神经网络都存在各自的局限性，探索不同网络结合的效果会是一条可行的途径。

(2)GAN 方法应用不够广泛，还有很大的空间可以提升，可以使用其他的 GAN 网络和更多数据集进行验证。

(3)由于不同历史手稿的特殊性和差异性，对于监督学习的深度学习方法来说为大量图像添加标签是一个亟待解决的问题。因此无监督学习的方法会是未来研究的一大热点。

参考文献

[1] WATANABE T. Document Analysis and Recognition[J]. IEICE Transactions on Information and Systems, 1999, 82(3): 601-610.

[2] LIKFORMANSULEM L, HANIMYAN A, FAURE C, et al. A Hough based algorithm for extracting text lines in handwritten documents[C]//International conference on document analysis and recognition, 1995: 774-777.

[3] 蒋勇,陈晓静. 一种多方向手写文本行提取方法[A]. 中国自动化学会控制理论专业委员会(Technical Committee on Control Theory,Chinese Association of Automation).第二十七届中国控制会议论文集[C].中国自动化学会控制理论专业委员会(Technical Committee on Control Theory,Chinese Association of Automation):中国自动化学会控制理论专业委员会,2008:4.

[4] KESIMAN M W, VALY D, BURIE J, et al. Southeast Asian palm leaf manuscript images: a review of handwritten text line segmentation methods and new challenges [J]. Journal of Electronic Imaging, 2016, 26(1): 11011.

[5] SAABNI R, ASI A, EL-SANA J. Text line extraction for historical document images[J]. Pattern Recognition Letters, 2014, 35(1):23-33.

[6] ZHANG X, TAN C L. Text Line Segmentation for Handwritten Documents Using Constrained Seam Carving[C]//International conference on frontiers in handwriting recognition, 2014: 98-103.

[7] ALAEI A, PAL U, NAGABHUSHAN P, et al. A new scheme for unconstrained handwritten text-line segmentation[J]. Pattern Recognition, 2011, 44(4): 917-928.

[8] KOO H I, CHO N I. Text-Line Extraction in Handwritten Chinese Documents Based on an Energy Minimization Framework[J]. IEEE Transactions on Image Processing, 2012, 21(3): 1169-1175.

[9] RYU J, KOO H I, CHO N I, et al. Language-Independent Text-Line Extraction Algorithm for Handwritten Documents[J]. IEEE Signal Processing Letters, 2014, 21(9): 1115-1119.

[10] LOULOUDIS G, GATOS B, PRATIKAKIS I, et al. Text line and word segmentation of handwritten documents[J]. Pattern Recognition, 2009, 42(12): 3169-3183.

[11] DESHMUKH M S, PATIL M P, KOLHE S R, et al. A hybrid text line segmentation approach for the ancient handwritten unconstrained freestyle Modi script documents[J]. The Imaging Science Journal, 2018, 66(7): 433-442.

[12] LONG J, SHELHAMER E, DARRELL T, et al. Fully convolutional networks for semantic segmentation[C]//Computer vision and pattern recognition, 2015: 3431-3440.

[13] VO Q N, LEE G. Dense prediction for text line segmentation in handwritten document images[C]//International conference on image processing, 2016: 3264-3268.

[14] BARAKAT B K, DROBY A, KASSIS M, et al. Text Line Segmentation for Challenging Handwritten Document Images using Fully Convolutional Network[C]//International conference on frontiers in handwriting recognition, 2018: 374-379.

[15] RENTON G,CHATELAIN C, ADAM S, et al. Handwritten Text Line Segmentation Using Fully Convolutional Network[C]//International conference on document analysis and recognition, 2017: 5-9.

[16] RENTON G, SOULLARD Y, CHATELAIN C, et al. Fully convolutional network with dilated convolutions for handwritten text line segmentation[J]. International Journal on Document Analysis and Recognition, 2018, 21(3): 177-186.

[17] RONNEBERGER O, FISCHER P, BROX T, et al. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]//Medical image computing and computer assisted intervention, 2015: 234-241.

[18] MECHI O, MEHRI M, INGOLD R, et al. Text Line Segmentation in Historical Document Images Using an Adaptive U-Net Architecture[C].//International conference on document analysis and recognition, 2019.

[19] GRUNING T, LEIFERT G, STRAUS T, et al. A two-stage method for text line detection in historical documents[J]. International Journal on Document Analysis and Recognition, 2019, 22(3): 285-302.

[20] NECHE C, BELAID A, KACEMECHI A, et al. Arabic Handwritten Documents Segmentation into Text-Lines and Words using Deep Learning[C]//International conference on document analysis and recognition, 2019: 19-24.

[21] MOYSSET B, KERMORVANT C, WOLF C, et al. Paragraph text segmentation into lines with Recurrent Neural Networks[C]//International conference on document analysis and recognition, 2015: 456-460.

[22] MOYSSET B, ADAM P, WOLF C, et al. Space displacement localization neural networks to locate origin points of handwritten text lines in historical documents[C]//Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing. 2015:1-8.

[23] MOYSSET B, KERMORVANT C, WOLF C, et al. Full-Page Text Recognition: Learning Where to Start and When to Stop[C]//International conference on document analysis and recognition, 2017: 871-876.

[24] GOODFELLOW I, POUGETABADIE J, MIRZA M, et al. Generative Adversarial Nets[C]//Neural information processing systems, 2014: 2672-2680.

[25] KUNDU S, PAUL S, BERA S K, et al. Text-line extraction from handwritten document images using GAN[J]. Expert Systems with Application, 2020, 140(2):112916.1-112916.12.

[26] ISOLA P, ZHU J, ZHOU T, et al. Image-to-Image Translation with Conditional Adversarial Networks[C]//Computer vision and pattern recognition, 2017: 5967-5976.

（下转第 160 页）