



上海建桥学院

本科毕业设计(论文)

题目 基于图像识别的快递信息识别算法的设计与实现

学生姓名	<u>楼上之</u>
学 号	<u>1920127</u>
指导教师	<u>强珏娴</u>
学 院	<u>信息技术学院</u>
专 业	<u>计算机科学与技术（智能应用）</u>
交稿日期	<u>2023.3.14</u>

教务处制

上海建桥学院毕业设计（论文）学术诚信声明

本人郑重声明：所呈交的毕业设计（论文），是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本毕业设计（论文）不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

作者签名：

日期： 年 月 日

上海建桥学院毕业设计（论文）版权使用授权书

本毕业设计（论文）作者同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权上海建桥学院可以将本毕业设计（论文）的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本毕业设计（论文）。

保 密 ☐，在____年解密后适用本授权书。

本论文属于

不保密 ☐。

（请在以上方框内打“√”，如作者未做出选择的情况下，按不保密处理。）

作者签名：

指导教师签名：

日期： 年 月 日

日期： 年 月 日

基于图像识别的快递信息识别算法的设计与实现

摘 要

随着互联网经济和快递业务的迅猛发展，人们享受到了电商带来的高效生活，但新冠疫情所带来的挑战也给快递系统带来了难题。经过调研发现，由于人工操作的短板，快递图像识别效率低，且误判率高，导致物流运输效率较低。为解决这些问题，本研究通过图像识别技术与深度学习算法的结合，设计优化了一种快递分拣系统。主要研究和结果如下：

首先，本研究设计了一种基于 Faster R-CNN 的快递价值单信息的检测方法，并采用改进的 FHT 算法 MFHT 对特定样本目标进行识别，并对目标信息进行深度特征提取和正样本补充训练。在自建数据集上验证，结果表明该算法不仅速度较快，且准确性高，能够快速定位快递单上的有价值区域。

其次，本研究使用 CTPN 模型，结合值信息区域识别算法，实现对快递信息码和地址等价值信息的快速识别。并通过 DenseNet 文字特征提取网络来提取特征序列，然后采用改进的加权 CTC 方法将其转译为文字序列。结果表明改进后的图像识别精度和速度得到明显提高，文本识别运也更加高效。

最后，本研究结合以上目标信息算法及文本提取模型，设计和优化了后的快递分拣系统，实现准确提取输入快递单图像的价值信息，使得快递自动分拣系统具有更高的识别精度和处理速度。

关键词：快递分拣，FHT，Faster R-CNN，CTPN，CTC

The Format Criterion of Master's Degree Paper of DUT

Abstract

With the rapid development of Internet economy and express business, people enjoy the efficient life brought by e-commerce, but the challenges brought by the new crown epidemic also bring difficulties to the express system. After research, it is found that due to the need for manual operation, the courier image recognition is inefficient and the misjudgment rate is high, leading to a sharp decrease in logistics and transportation efficiency. In order to solve these problems, this paper designs an optimized courier sorting system by using image recognition technology combined with deep learning algorithms.

Firstly, this paper designs a Faster R-CNN based courier value list information detection method, and adopts the improved FHT algorithm MFHT for specific sample target recognition with deep feature extraction and positive sample supplemental training for text targets. Validated on a self-built dataset, it is demonstrated that the algorithm is not only faster and more accurate, but also can quickly locate the valuable regions on the courier order.

Secondly, this paper uses the CTPN model, combined with the value information region recognition algorithm, to quickly achieve the recognition of value information such as information code and address. This paper uses DenseNet text feature extraction network to extract feature sequences, and then uses an improved weighted CTC method to translate them into text sequences. The research results in this paper can improve the accuracy and speed of image recognition and make the text recognition operation more efficient.

Finally, this chapter combines the research presented in the previous two chapters to design and optimize the post-courier sorting system, which can achieve accurate extraction of value information of input courier order images and make the automatic courier sorting system with higher recognition accuracy and processing speed.

Keywords: Express sorting, FHT, Faster R-CNN, CTPN, CTC

目录

1 绪论.....	1
1.1 研究的背景及意义.....	1
1.2 国内外研究现状.....	2
1.2.1 快递自动分拣技术.....	2
1.2.2 目标检测技术.....	3
1.2.3 文本检测技术.....	4
1.2.4 OCR 字符识别技术.....	4
1.3 研究内容与章节安排.....	5
2 快递单价值信息识别.....	7
2.1 传统快递价值信息识别.....	7
2.2 基于深度学习的快递价值信息识别.....	7
2.2.1 神经网络.....	8
2.2.2 卷积神经网络.....	9
2.3 快递信息编码规则介绍.....	10
3 融合 FTH 算法的价值信息检测.....	12
3.1 问题分析.....	12
3.2 基于 DenseNet 的特征提取网络.....	13
3.3 改进的正样本补充方法.....	15
3.4 FHT 算法.....	16
3.5 实验结果.....	18
3.5.1 数据集.....	18
3.5.2 实验环境和实验参数.....	18
3.5.3 实验结果展示分析.....	18
4 文字字符识别.....	20
4.1 模型结构设计及算法优化.....	20
4.1.1 CTPN 网络.....	20
4.1.2 加权 CTC.....	21
4.2 实验结果及分析.....	22
结 论.....	23
参 考 文 献.....	24
致 谢.....	26

1 绪论

1.1 研究的背景及意义

根据国家邮政局在 2023 年 1 月发布的 2022 年邮政行业运行情况报告显示,截至 2022 年底,邮政行业完成了 1391.0 亿件的寄递业务量,其中,快递业务量达到了 1105.8 亿件,邮政寄递服务业务量为 285.2 亿件。邮政行业业务收入累计完成了 13509.6 亿元,其中,快递业务收入累计完成了 10566.7 亿元。2022 年,快递业务量完成 1105.8 亿件,业务收入完成 1.06 万亿元,连续 9 年位居全球第一。此外,行业最高日处理能力超过 7 亿件,年人均快件量近 80 件。

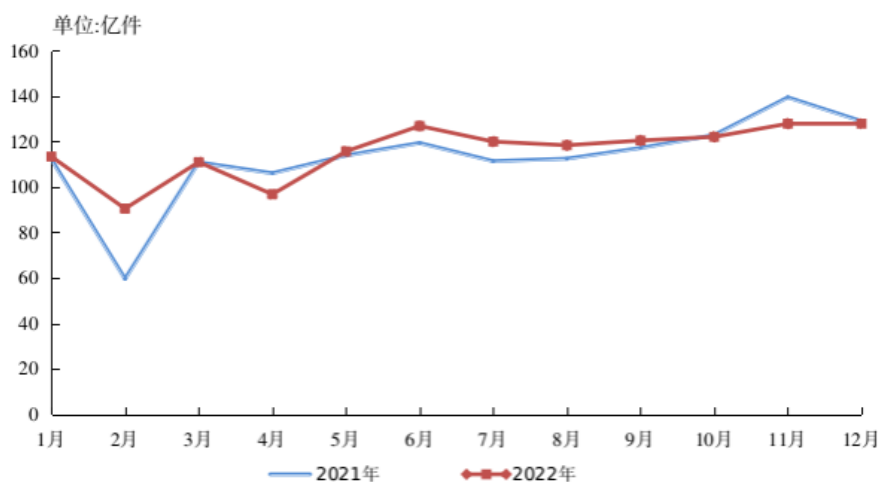


图 1.1 2022 年邮政行业寄递业务量

面对国内快递市场不断增长的包裹量,继续依靠目前的人工运输、扫码、分拣方式将会增加分拣错误率和快递件的滞留率,从而导致快递公司客户流失。近年来,由于新冠疫情的影响,快递输送链成为了新冠传播与感染的一大风险,许多需要人力参与的环节受到重大冲击。因此,亟需建立高效的无人化快递链。

本研究通过实地调查,并与企业相关负责人交流后发现,国内各类快递分拣公司基本采用相同的处理方式:每个员工搭配一台终端,扫描快递单号以确定包裹的配送状态。但是,应用于工业的图像识别环节效率低下,会存在一些问题:

(1) 由于送入扫描的快递大小和形状各异,快递单与相机的距离也不同,导致每张照片中快递单的大小也会不同。在扫描小件时,单号太小常常导致识别失败,从而增加了区域划分、识别快递价值信息的难度。

(2) 传入图片方向角度不一,因而存在识别错误或无法识别的情况。形状大小不同

的快递堆叠后，导致拍摄的识别图中快递单占比小

(3) OCR 文字识别环节速度较慢。因此，在操作过程中，员工常常需要人工多次将快递重新放置于指定区域才能识别，对于形状不规则的快递无法进行识别，导致效率低下和错误率高，严重制约了物流运输效率。在进行区域扫描或 OCR 文字扫描前，需要将快递单面对摄像头，确保传输至系统的照片正向。然而，在日常人工操作中，为了提高效率，员工很难每次都准确地将快递单正确地放入扫描区域，有时需要多次操作才能成功。为了更突出收件人的信息，在快递表单的表面，常常会有好几种字体，字号也会有不同。所以会导致不一样的文字区域很难被划分，提炼出有价值的信息，因此增加了 OCR 识别的难度。

1.2 国内外研究现状

1.2.1 快递自动分拣技术

在 20 世纪中期，西方发达国家的邮政系统开始探索快递自动分拣技术。企业家和技术人员开始研究邮政行业的自动化，并随着算法和科技的进步，不断取得了许多成果。这些自动分拣设备逐渐替代了部分机械化操作，从而解放了大量劳动力。

在经过数十年的技术迭代之后，像皇家邮政、联邦快递等发达国家的快递企业，使得快递自动化更加智能和便捷。在 1966 年，日本的大福公司开发了立体自动化仓库，并在 1995 年提出一套完善的物流分拣系统^[5]。这套快递系统为日本物流行业带来了极大的益处，使其在行业内遥遥领先。自动化物流系统在中国开始于上世纪 70 年代，当时就研发出了第一个自动化仓储系统。自 1985 年起，我国陆续颁布了相关技术标准。1995 年起至今，我国物流厂商和供应商借鉴发达国家先进技术，不断改进创新，开始了更加智能的发展。

目前，智能快递分拣领域中的工业机器人只能够完成固定模式的物体分拣任务。对于一些具有复杂场景的情况，例如存在堆放杂乱无章和互相遮挡等情形，需要通过人工方式来完成拣选任务。在国外，智能分拣系统已经被广泛应用多年，如美国康乃士的 VisionPro 3D，提供准确实时的 3D 信息。MUJIN 公司主要涉及工业领域中的 3D 物体检测识别与抓取、机械臂运动学计算以及机械臂运动规划等方面。此外，国内也出现了多家企业，如奥尔森、梅卡曼、库柏特等，为工业、金融、物流等行业提供智能分拣解决方案。

1.2.2 目标检测技术

目标检测是计算机视觉领域中的一个重要研究方向，主要应用方向是在图像或视频中准确识别和定位出特定对象的位置。目标检测是许多实际应用的基础，例如自动驾驶、智能安防、医学影像分析等。目前，目标检测算法主要分为两大类：传统目标检测和深度学习目标检测。

传统方法的目标检测主要包括基于滑动窗口的目标检测和基于区域建议的目标检测两种方式。其中，基于滑动窗口的目标检测需要在不同大小的窗口上进行分类器的训练和测试，计算量较大；而基于区域建议的目标检测则是通过生成一组候选框并对其进行分类器的训练和测试来完成目标检测，因此速度相对较快。在 2014 年以前是传统目标检测时期^[6]，传统目标识别算法，常通过二值化，倾斜校正，版面分析等预处理办法，提取区域文本图像的边缘特征信息，从而进行文本检测或目标识别。2002 年，Matas 等人提出基于文本连通域的算法，非极大值抑制算法（MSER）^[3]，能够连续使用不同的灰度值对图像进行二值化，粗略定位到图像的文本区域，检测识别各个精度程度的区域。2010 年，Epshtein 等人提出了 SWT（Stoke Width Transform）^[8]文本边缘提取算法，在 ICDAR 的数据集上试验后，实验结果表明该模型对复杂背景、不同文字尺寸识别有较好的泛化性，缺点是对噪声较为敏感，在背景涂抹比较严重和字符重叠断裂的文本区域，识别率很低，传统检测算法至此进入了饱和的发展期。

2014 年后，基于深度学习的目标检测方法则是利用卷积神经网络（CNN）中的特征提取和分类功能来完成目标检测。这类方法通常采用区域提取网络（RPN）和特征金字塔网络（FPN）等网络结构来提高检测精度和效率。其中，一些比较典型的基于深度学习的目标检测算法包括 Faster R-CNN、YOLO 等。这些方法广泛应用于目标识别和文字检测领域。

近年来，涌现出不少性能优秀的物体检测方法。例如，Girshick 团队利用卷积神经网络研究了选择性搜索区域建议方法，并将其与 AlexNet 网络集成，设计出通用的目标检测框架 R-CNN^[1]。然而，由于 R-CNN 需要多阶段复杂训练，耗费大量算力和时间，整体训练过程缓慢且难以优化。此外，SVM 分类器在回归训练时占用空间也十分庞大，导致对每张图片进行检测时速度缓慢且占用算力巨大。为解决这些问题，Girshick 团队在 2015 年提出了 Fast R-CNN 和 Faster R-CNN^[2]，通过修改 RCNN 和 SPP-Net，在解决

R-CNN 缺点的同时，将速度提高了 200 倍。同年，S.Ren 等人提出了 Faster RCNN 网络^[8]，实现了端对端的目标检测，将特征提取边界回归集成到了同一个框架中，成为后续各类文本检测模型改进的基础。

在模式识别和计算机视觉领域领域中，霍夫变换（Hough Transform, HT）给二值图像进行直线检测提供了优秀的图像处理和计算机视觉中的特征提取技术，它可以对图像中的直线、圆等形状进行检测和描述。1962 年，霍夫变换由 Paul Hough 发明的^[3]，得益于该算法在抗噪声性能等方面优秀的特性，HT 算法可以广泛应用在任意形状区域的边界提取。鉴于上述算法在快递单区域识别环节中也能实现良好的特性，同时也为了降低算法复杂性、提高应用 HT 算法计算的效率和检测精度。虽然霍夫变换具有许多优点，但也存在一些局限性，例如对噪声和曲线断裂敏感，且计算复杂度较高。2001 年，基于 HT 算法，孙丰荣和刘积仁^[4]提出了 FHT 和 MFHT（Modified Fast Hough Transform, MFHT），在探索一些研究先驱采用并行处理的策略后，如 Cypher, Synder, Sanz, Hambrush 等分别在^[10,11]中实现计算量为 $O(\log_2 N)$ ，设计了一种顺序算法，计算量仅为 $O(N^2 \log_2 N)$ ，可以并行实现，得到更高的计算精度。为了解决快递单价值区域的识别划分，本文也是在 FHT 算法设计的思路进行了融合创新。

1.2.3 文本检测技术

目标检测技术衍生出了文本检测技术，在目标检测模型中的目标对象的长宽通常为 1:1，而文本行是一较长矩形形式存在，因此受到目标检测的启发，可以将目标检测的模型的网络稍作修改就可以实现文本检测。

2016 年 ECCV（European Conference on Computer Vision）会议提出了 CTPN^[9]，该模型是基于 Faster RCNN 改进而来，能够有效检测水平的文字。但是在很多场景中送入检测的文本并非严格意义上的水平，甚至很多是完全倾斜的，本文研究的快递单文本信息识别，在工业场景中便常常会因为文本区域环境恶劣导致识别和错误了很高。2019 年，Nabati 等人提出了 RRPN 文本检测模型^[12]，它将 Faster RCNN 改进，利用可旋转的锚框，在回归中加入了角度的回归，使得该模型能够检测得到倾斜的文本行。

1.2.4 OCR 字符识别技术

在二十世纪初，OCR（Optical Character Recognition，光学字符识别）技术在国外就已经开始萌芽。1929 年，德国科学家 Tausheck 首次提出 OCR，随后 Handel 利用机器实

现了对文字识别的想法。上个世纪 80 年代,西方国家成立了许多 OCR 研究中心实验室,这些研究中心的成立推动了 OCR 技术的发展。其中,惠普在英国布里斯托实验室研发了非常著名的 OCR 识别引擎—Tesseract。时至今日,该算法的识别精度依旧可靠,且使用率很高。1998 年,LeCun 提出了卷积神经网络中的早期模型 Lenet-5,这一模型能够成功地实现对手写数字的识别。深度学习的字符识别发展到今日,算法逐渐以 CNN+RNN+CTC 或 CNN+RNN+Attention 为主流,识别准确率和效率也得到了很大的提高。OCR 技术也得到广泛应用,例如图书数字化、手写体识别、车牌识别等领域。

为了能够识别整个文本单词,2013 年,Goel 等人^[16]将输入文本和预定义语义库的单词结合,运用加权动态时间规整算法,从合成图像集中寻找与输入文本相符合的项。同年,Rodriguez-Serrano 等人使用结构化的 SVM 来学习从图像空间到标签空间的投影,可以是文本图像的特征向量比其他不正确标签的特征向量更接近。2015 年,Gordo 等人^[18]进一步拓宽了该框架,不断提升了在不同数据集下文本识别的准确率,但是由于过于以来特定语料库,因此未能广泛使用。

尽管起步较晚,我国在上世纪七十年代才开始进行相关技术的研究。然而,随着台式扫描仪和自动化办公的普及,国内 OCR 市场不断扩大,这也对 OCR 识别技术的正确率和速度提出了更高的要求。如今,我国在光学字符识别领域处于国际前列。北京大学高良才教授团队^[27]、清华大学丁晓青教授团队^[28]等长期致力于光学字符技术的开发,并取得了长足的创新和突破。同时,阿里巴巴、海康威视、腾讯、科大讯飞等公司对 OCR 识别技术的大力开发,也促进了我国科研成果的应用水平。

1.3 研究内容与章节安排

面对市场上大多中小物流分拣公司传统的快递分拣,显然还存在着很大的改进空间。本文希望基于深度学习,利用图像识别技术,来进一步减少快递分拣过程中人工的参与度,同时提高各个环节的效率。

本文的研究内容基于快递分拣系统,模拟了快递进入传送带的模型当中环节。目的是设计出良好的算法,能从快递单图像中提取出有价值的信息,便于后续环节的识别和分拣。研究内容主要分为三个环节:目标识别、文本行检测、OCR 字符识别。本文主要由六个章节构成,本文结构和内容安排如下:

第一章,绪论。在如今物流行业快递发展的背景下,提出了,目前快递分拣中存在

的几个主要的问题。概述了快递自动分拣、目标检测、以及文本检测的发展过程，并阐述了本论文的结构安排和主要研究任务。

第二章，主要阐述了基于深度学习的价值信息目标识别的探究方法和快递信息编码的规则介绍。同时在开篇对经典的快递价值信息提取和深度学习框架进行了必要的介绍，以此为基础，引出了对 CRNN 图像文字识别技术和目标检测技术的介绍，为后文引出 FHT 算法和 RCNN 模型做铺垫，本章在此提供了前置知识的必要说明。

第三章，本文分析了 CRNN 模型的优缺点，并阐述了使用 Faster RCNN 模型进行快递价值信息识别的必要性。为了减少背景和噪音的干扰，在针对快递单识别场景中，我们提出了采用 Faster RCNN 模型的方案。参考张震提出的联级 Faster RCNN 模型的设计思路，本文采用 FHT 算法来提高模型速度和精度。同时，通过结合正样本补充方法并验证数据集，我们证实了该方法可以较为准确地定位快递图像中的价值信息。

第四章，设计了基于 CTPN+DenseNet 网络来进行文字特征提取，结合加权 CTC 方法进行解码识别文本信息。在数据集上测试得到，该模型并具有良好的鲁棒性。可以在快递分拣各个过程中根据需求提供给快递员或系统准确、有价值的信息，在扫描过程中将极大地提高快递运输的效率。比如，在快递员派送过程中，无需再人工识别用户快递单的号码信息，只需对比之前快递分拣过程中系统中识别的号码数据库即可进行比对，既节省时间，也能降低错误率。

第五章，总结与展望，首先总结了本文设计算法思路中一些人需要优化或者是改进的问题，展望下一步的优化工作。

2 快递单价值信息识别

2.1 传统快递价值信息识别

快递单中的价值信息,指的是收件人或者寄件人信息的文本,在快递分拣的过程中,每个环节会根据其中的某一部分文本进行分类,如在中转站时,通过对收件人收件区域划分,选择后续进入配送的卡车。我们将这部分能对分拣或者配送有价值的信息,称为价值信息。

近年来随着人工智能技术的迅猛发展,我们可以根据是否使用深度学习技术将快递价值信息提取分为两个阶段:传统提取方法和基于深度学习的提取方法。在深度学习时代之前,在传统的快递单信息提取过程中,主要分为以下几个步骤:(1)图像获取。(2)图像预处理。(3)价值信息定位。(4)价值信息识别。图像预处理环节主要包含数据增强、归一化、标准化和降噪等,将快递单价值信息区域于背景分离,然后用过基于投影等方法,价值区域中包含文字行的字符分割为单个文字,最后基于对单个字符进行文字分类,以实现价值信息的提取。

在传统快递单识别初始,文本识别的研究主要集中在在自然环境中对单个字符的识别^[14,15],这两篇文献主要采用了一种自下而上的方法,使用字符分类器对单个字符检测识别,然后整合到整合文本图像得到文本检测。但是由于外界影响因素过多和图像中大小不一的字符导致识别率很低。综上所述,在传统的价值信息提取过程中,对于非固定结构的快递单,其定位效果往往不尽如人意。此外,传统的价值信息识别方法需要手工分割字符进行分类识别,当快递单内容较多、字符大小不一或者字符出现断裂时,识别效果也会受到很大的影响。为了解决这些问题,基于深度学习的方法被提出并已经在快递价值信息提取领域取得了显著进展。接下来,本文将着重介绍基于深度学习的快递价值信息提取技术。

2.2 基于深度学习的快递价值信息识别

深度学习是一种强大的技术,它能够从原始的输入数据中提取出越来越抽象的特征表征,同时具备着良好的泛化能力。不仅如此,深度学习还克服了过去被认为难以解决的一些人工智能问题。2013年,谷歌发表^[19]论文带火了深度学习,随着算力设备如CPU, GPU 逐渐支持 DL 的计算,深度学习开始广泛出现在各个领域。下面主要介绍深度学习

中的目标检测技术和文字识别技术：

2.2.1 神经网络

人工神经网络（ANN）是一个由大量神经元相互连接组成的自适应非线性动态网络系统^[22]。单个神经元如图 2.1 所示，神经网络由图 2.2 所示。

神经网络（NN）是指多个神经元在一层中链接到前一层的所有输出，同时这些神经元的输出也可以作为下一组神经元的输入。神经网络通过对神经元权重的学习来实现功能，其学习过程主要包括前向传播、误差计算、反向传播和权重更新。换句话说，神经网络的学习本质上就是对神经元权重进行的学习。令 w 为权重向量， x 为神经元的输入向量， g 为激活函数， b 为偏置项。神经元为 NN 输入加权总和，然后传递给非线性函数，如公式 2.1 所示：

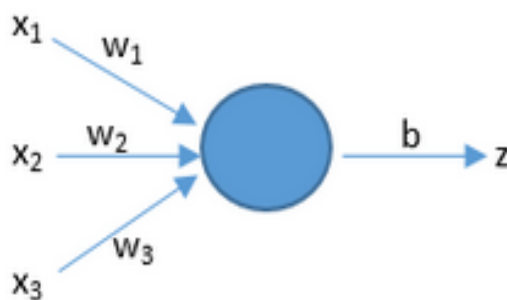


图 2.1 单个人工神经元

$$y = g\left(\sum_j w_j x_j + b\right) \quad (2.1)$$

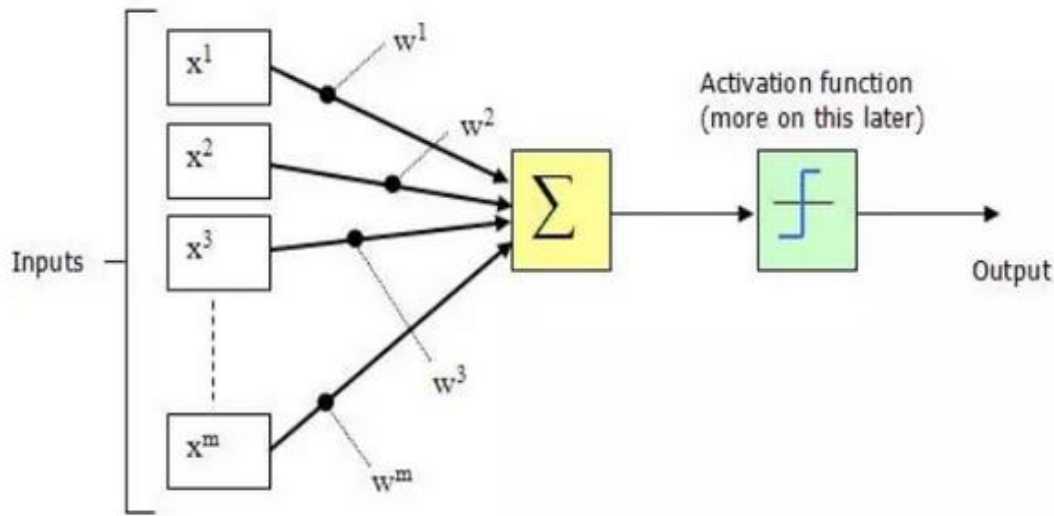


图 2.2 简单的神经网络

2.2.2 卷积神经网络

卷积神经网络是一种广泛应用于图像分类、目标识别、文字识别等多个领域的人工神经网络。上世纪 80 年代，卷积神经网络 CNN（Convolutional Neural Networks）的概念被提出。现如今随着卷积算法在工程学、数学等不同领域大放异彩，但其根源还是严格取自图像处理^[20]。CNN 的结构包括输入层，输出层以及多个隐层。卷积网络结构图如图 2.3 所示。输入和卷积核进行卷积，从而生成特征图。卷积函数如 2.2 所示。

$$h(x) = \int_{-\infty}^{+\infty} f(x)g(x-\gamma)d\gamma = (f \times g)(x) \quad (2.2)$$

普遍观点认为，在图像识别任务中，由于深层卷积网络拥有极高的速度和海量的参数，这使得它具备了很出色的性能表现^[21]。不论是残差网络还是密集网络都存在特征练级情况，导致大量参数需要迭代，导致训练过程对算力的要求很高，同时非常复杂。因此，在此基础上研究出一种轻量级的网络，进行文字图像的特征提取是很有价值的。

由于 CNN 是独立于每张图像的独立区域提取特征的，这个给大规模检测带来了挑战。特别是当网络结构较深的网络，如 AlexNet，整个识别过程是及其缓慢的。

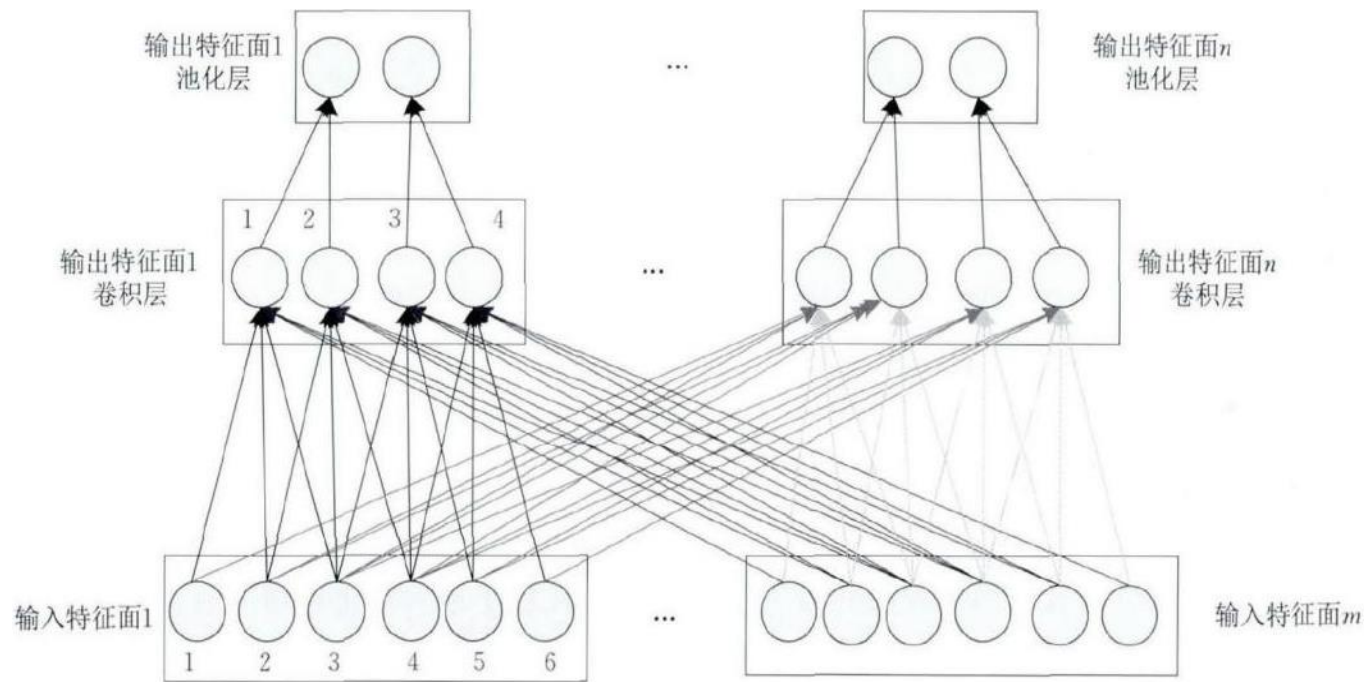


图 2.3 卷积神经网络

2.3 快递信息编码规则介绍

快递的信息码如图 2.4 所示。其前身是邮政编码，由于汉字匹配机制非常复杂，表述的地址也过于庞大，给系统造成了巨大的计算量，同时识别准确率也非常低，难以满足市场的需求，因此，需要采用数字化编码来代替汉字地址。

信息码的结构分为三段，第一段表述的是快递公司在全国的大小分拣中心代码，如表 2.1 所示。

表 2.1 A 码位的含义

地区	所含省份	编码
东北	辽宁省、吉林省、黑龙江省	2
华北	北京市、天津市、河北省、山西省、内蒙古自治区	1
华东	上海市、浙江省、江苏省、安徽省、山东省、江西省、福建省	3、4、5
华中	河南省、湖北省、湖南省	7
华南	广东省、广西省、海南省	6
西南	重庆市、四川省、贵州省、云南省、西藏省	8
西北	陕西省、甘肃省、青海省、宁夏省、新疆	9

第二段码，代表个县级行政单位内存在的快递分拣中心。第三段则表示最后一级网点下属区域的派件员。由于信息码是按照分拣流程来设计的，在通过扫描之后便可以自动形成地址，有效地提高了分拣的效率和准确性。



图 2.4 快递信息码区域

但是在实际运输过程中发生的意外，会导致该部分区域字体缺失，或者涂抹断裂导致在识别过程中无法定位到这块区域。所以需要基于深度学习识别提取快递单价值区域，是本文研究的重点。

3 融合 FTH 算法的价值信息检测

本章主要研究了基于深度学习框架的快递单价值信息检测，在自然场景中，详细分析了经典模型和算法的不足。在此基础上，提出了一种该方法基于 Faster RCNN 框架，融合 FHT 算法的价值定位方法，来准确定位价值信息。

3.1 问题分析

在第二章中提到，我国的快递分拣系统主要采用了信息码的形式，如图 3.1 所示。传统的表单价值信息检测方法主要是基于联通域技术，通过从快递单的图像中获取表单位置，然后使用二值化、灰度处理等操作来对表单内容进行分割。然而，这类方法往往对图像的质量要求较高，且难以处理如 3.2 所示的图片。因此，这种方法存在可移植性和准确性方面的缺陷。



图 3.1 快递表单图像

若要设计在自然场景中高性能的快递单价值信息定位任务，这需要要求模型有良好的鲁棒性和抗干扰性以及较高的检测准确率。可以把表单的价值看作是物体，用目标检测的逻辑去设计，将文本定位任务转变为目标检测任务。目标检测模型不仅可以得到快递价值信息的准确位置，还可以在其他分拣环节应用识别其他类别的价值信息。

虽然单阶段的检测方法速度快，但是准确率较低，不适用于物流分拣系统。所以本文采用双阶段检测方法—Faster RCNN 处理图像，在模型中速度可以通过其他算法弥补改进，相对来说准确率更为重要。但是在若直接使用目标检测模型进行定位，又会产生

以下问题:

第一,如图 3.1 中快递表单所示,在快递单中存在多个价值区域,例如寄件人和收件人的地址、电话等信息。这些信息在识别过程中会相互干扰。如果将整张图像输入到物体检测模型中,可能会导致出现多个结果,从而出现多检的情况。



图 3.2 背景有笔迹的快递表单

第二,对于物流传送带上拍摄的照片,需要通过人工协助拍摄识别。由于拍摄时会存在复杂的背景干扰,因此模型得到的结果会产生很大的偏差,导致在训练模型时出现正负样本不均匀的问题,从而导致模型的准确率很低。

为了能够解决以上问题,我们需要对 Faster RCNN 模型进行修改,使它能够适应快递价值信息定位的任务。

首先,我们需要解决 Faster RCNN 模型速度的不足,错误率高的问题。通常情况下,特征提取的第一步是最为关键的。网络所提取的特征越重要性,模型的识别效果就会越好,从而减小模型的错误率和漏检率下文就特征提取网络的选取和创新做了详细说明。

其次,对于模型是非结构化的,所以会导致检测定位出多个价值区域。在图像送入检测模型之后,我们需要排除多余的区域的干扰。

3.2 基于 DenseNet 的特征提取网络

特征提取是图像处理中最重要的步骤,它影响到一个深度模型是否能充分学习目标特征。同样,特征提取也是文本识别过程中的一个关键阶段,因为定义良好的特征图

可以促进识别,提高准确率^[23]。人们普遍认为,深度卷积网络既要依靠网络的深度,又要依靠大量的参数,才能取得优异的性能。然而,随着网络深度的增加,训练过程中需要优化的参数数量也相应增加,训练网络需要具有更多算力的硬件设备,网络训练的难度也随之增加。

DenseNet 是一个具有密集连接的卷积神经网络。该架构包含一个卷积层、一个池化层和一个密集连接模块,其中增长率为 4 是为了保持通道特征维度的适度,以防止规模过大。DenseNet 网络使用密集连接的方式将先前层的特征用于后续层的输入,该网络降低了梯度消失的影响,提高特征的传播,实现特征重用,极大地减少模型的参数量,网络分解如图 3.3 所示。

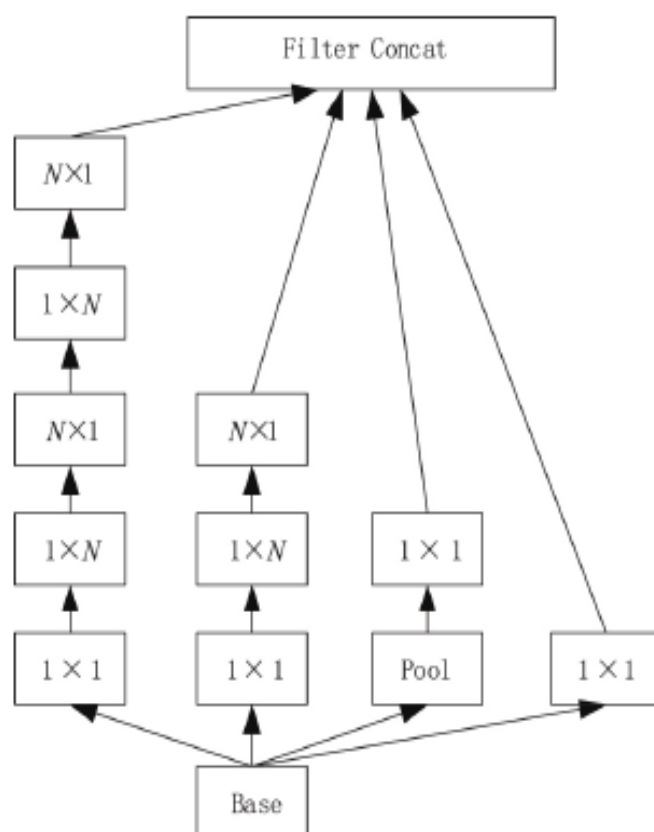


图 3.3 $N \times N$ 卷积结构分解

DenseNet 网络结构并没有像以前那样从更深的网络或更广的网络开始,而是对特征序列进行了调整。由于该网络每一层都有很浅的特征图或者很少的卷积核,所以该网络有加强特征传播,更有效利用特征,很少的参数量的特点。DenseNet 网络结构如图 3.4 所示。

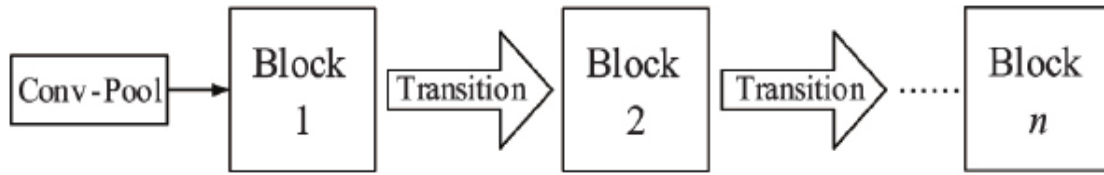


图 3.4 DenseNet 网络结构示意图

DenseNet 网络使用密集连接的方式将先前层的特征用于后续层的输入，该网络降低了梯度消失的影响，提高特征的传播，实现特征重用，极大地减少模型的参数量。由于该网络每一层都有很浅的特征图或者很少的卷积核，所以该网络有加强特征传播，更有效利用特征，很少的参数量的特点，计算高效。

3.3 改进的正样本补充方法

为了避免正样本补充方法带来性能下降，张震^[13]提出了一种改进的正样本补充方法。如果负样本的实框与其候选框的交集比 IOU 值都大于一定的阈值，则将该候选框放入正样本补充集，当正样本的数量没有达到设定的阈值 128 时，则从正样本补充集中随机抽取候选框来补充正样本。其中，锚点框 a_i 与重更新定义的真实框 g_{ii} 直接的交并如公式 3.1 所示。

$$IOU = (a_i, gt_{ij}) = \frac{area(a_i) \cap area(gt_{ij})}{area(a_i \cup gt_{ij})} \quad (3.1)$$

通过这种阳性样本补充方法，将更多含有目标信息的阴性样本候选框作为阳性样本的替换，使参与训练的整体阳性样本含有更多的目标信息，从而使模型学习到更多的目标信息，解决了模型收敛速度慢的问题，可能在一定程度上降低模型的准确度，减少误分类。如图所示图 3.5 为重新定义的真实框。

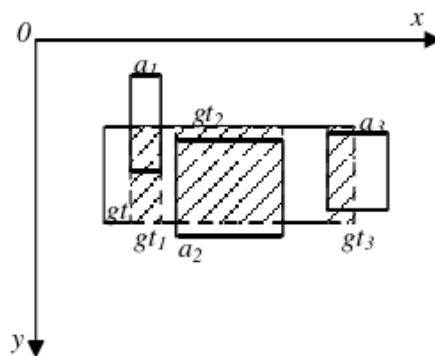


图 3.5 重新定义的真实框

3.4 FHT 算法

由于 HT 算法具有良好的抗噪声特性和对部分掩蔽现象不敏感，因此在模式识别领域被广泛使用。例如，直线检测、任意形状区域的边界提取、二维或三维运动参数的估计等等。相应地，HT 算法的许多特性也成为学术研究的热点。例如，HT 的严格数学定义问题，算法的复杂性问题，以及应用的并行实现问题。标准 HT 变换算法数学流程图如图 3.6。

为了解决 Faster RCNN 速度慢的问题，本实验融合了 FHT 算法，孙丰荣和刘积仁^[4]提出了 FHT 和 MFHT (Modified Fast Hough Transform, MFHT)，在探索一些研究先驱采用并行处理的策略后，如 Cypher, Synder, Sanz, Hambrush 等分别在^[10,11]中实现计算量为 $O(\log_2 N)$ ，设计了一种顺序算法，计算量仅为 $O(N^2 \log_2 N)$ ，可以并行实现，得到更高的计算精度。

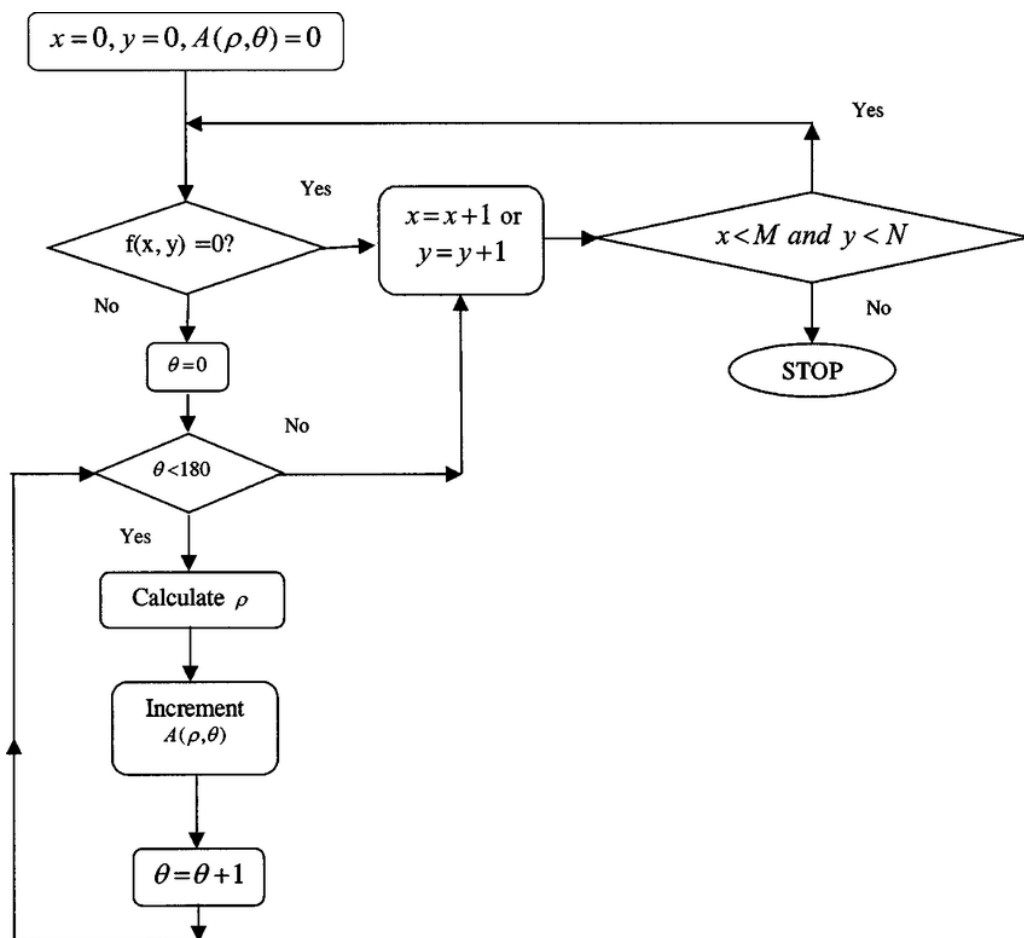


图 3.6 标准 HT 变换算法流程图

以检测直线为例，对于一条直线 l ，它可以用斜率 k 和截距 b 的形式表示为 $l: y=kx+b$ 。在参数空间中，每个点都表示一条直线，可以用极坐标来 (r,θ) 表示，其中 r 表示直线到原点的距离， θ 表示直线与 x 轴的夹角。对于每个像素点 (x,y) ，将其对应到参数空间中的直线上，可以得到一个曲线。在曲线上，经过的点越多，表示对应的直线出现的可能性越大。霍夫变换的实现过程中，需要设置阈值来过滤曲线上的噪声，同时需要进行后处理来合并相似的直线，得到最终的直线检测结果。

在 FHT 变换算法的基础上，提出了 MFHT 算法，据文献^[4]可得到结论，算法的误差分析及数值计算得出，该算法具有良好计算精度，可以给快递单价值信息检测提取提供很大的应用价值。实验结果表明，在图像规模为 1000,2000,3000,5000 时候，影响因子 k 为 2,4,8 时，结果如表 3.1 所示。在图像参数越大的情况下，MFHT 展示出的水平误差上界越优秀。

表 3.1 FHT，MFHT 算法的水平误差上界

图像规模	FHT	MFHT(k=2)	MFHT(k=4)	MFHT(k=8)
1000	4.45	3	1.988	1.125
2000	5	3.24	2	1.235
3000	5.24	3.39	2.06	1.266
4000	5.45	3.5	2.11	1.292

3.5 实验结果

3.5.1 数据集

本文实验中采用的照片由人工采集于学校快递中心（仅用于研究）。共有 5322 张照片，以上照片均先经过了压缩处理和灰度处理。其中参考文献^[13]中的评价检验方法，采用其中 2000 张照片用于训练，其中 300 张用于验证，400 张用于测试模型。

3.5.2 实验环境和实验参数

本文实验所用到的硬件参数如表 3.2 所示。

表 3.2 实验环境配置

设备	参数配置
中央处理器	AMD Ryzen 5 4600H with Radeon Graphics
显卡	NVIDIA GeForce GTX 1650
内存	64 GB DDR4
操作系统	Ubuntu 20.04

在训练快递单价值信息检测模型前，本文先使用与训练 DenseNet 的权重初始化提取特征提取网络，RPN 和 Faster RCNN 的权重使用均值为 0 且标准偏差为 0.01 的高斯分布随机权重初始化^[13]。

3.5.3 实验结果展示分析

如图 3.7 所示即为对快递单价值区的部分识别结果，利用 FHT 算法可以实现对信息码区域的快速准确识别，从而为后续 OCR 识别提供了良好的素材样本。

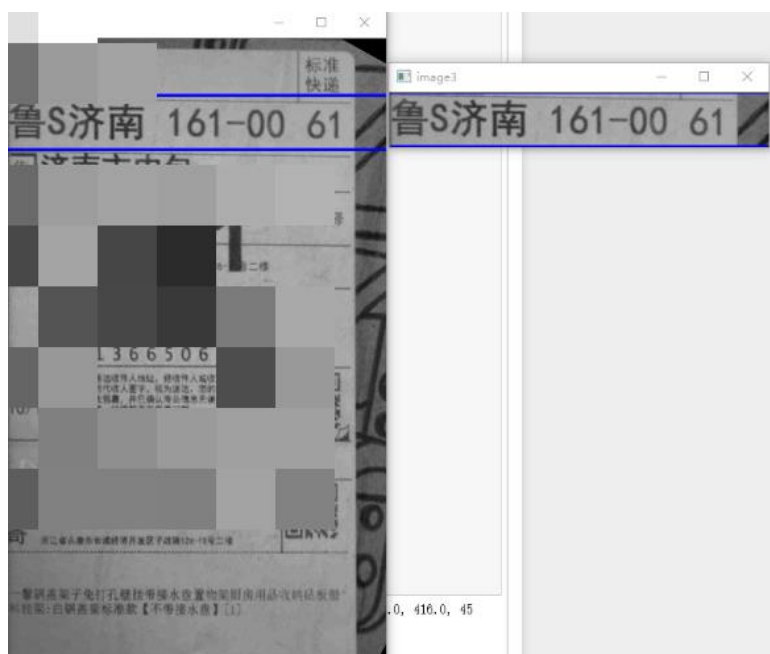


图 3.7 基于 FHT 算法的识别结果

为了解决快递单价值信息多检、错检的问题,本文基于一种改进的正样本补充方法,将负样本对应的真实框作为正样本的补充集合。本文对改进正样本补充方法做了有效性分析,参考文献^[13]中,通过 IOU (见公式 3.1) 的值,真实框和预测框的数量比,两项数据来判断预测框是否正确,截图如图 3.8 所示。



图 3.8 正样品补充的预测框和真实框

4 文字字符识别

4.1 模型结构设计及算法优化

4.1.1 CTPN 网络

由于在自然场景中识别快递单信息识别，传入模型中的图像常常是倾斜的，所以能够采用具有倾斜角度文本的检测网络十分重要的。2015 年，Lawgali^[23]在文献中展示不同特征提取网络在准确率不同，该现象说明特征提取网络的重要性。本文将了 CTPN^[24]网络应用到了文本检测任务中，该算法可以直接定位到卷积层的文本序列。CTPN 的算法结构如图 4.1 所示。

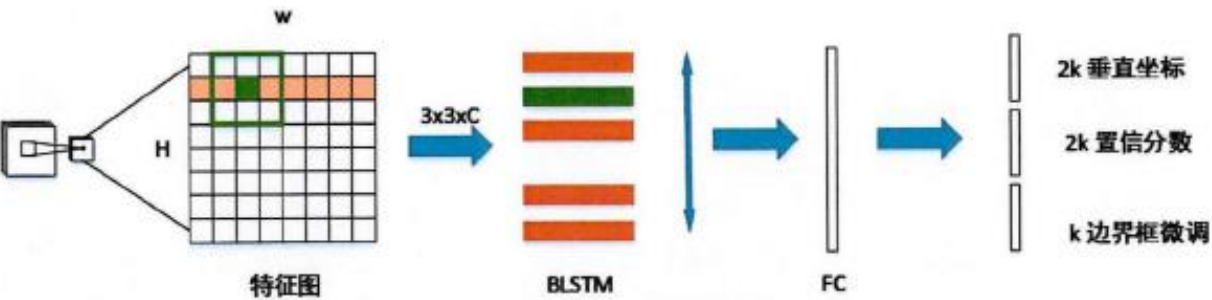


图 4.1 CTPN 网络结构

CTPN 提出了一种新的基于 RPN 网络的锚框回归机制，首先将文本区域细分为各个小区域，然后预测文本在各个区域的垂直位置和文本位置，最后以较高的精度获得文本区域位置信息。该算法采用递归神经网络连接卷积获得特征映射，这种无缝连接使得网络可以获得文本线的上下文信息，使得检测自然场景中的文本线成为可能。该算法可用于处理多尺度和多语言文本，避免了后续的过滤和细化操作。

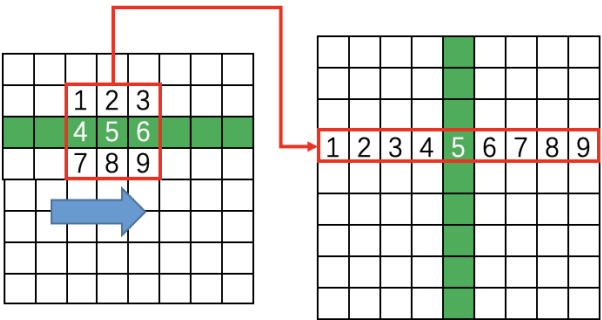


图 4.2 CTPN 网络结构原理示意

4.1.2 加权 CTC

CTPN 模型也存在缺点，由于只可以检测到水平方向的文本，竖直方向就会出现字符间断开的现象。同时，汉字的构成极其复杂，细微的笔画变化就会造成识别的错误，如“大”和“太”，“间”和“问”等。越是细小的变化，对模型的特征提取能力和细节学习能力要求就越高。传统的 CTC 和 CRNN 针对英文文本有良好的识别率，但是相比与组合数量较少的英文字母，庞大的汉字体系对模型提出了更高的要求。

在第二章中介绍了基于 CTC 的文字识别模型。CTC 机制常用于文字识别系统，解决序列标注问题中输入标签与输出标签的对齐问题，通过映射函数将其转换为预测序列，无序数据对齐处理，减少了工作量，可以广泛用于图像文本识别的损失函数计算，多用于网络参数的优化。

文献^[25]中引入了时间步长损失 L_{ts} ，目标函数 $L_1(X)$ 可以表示为公式 4.1。

$$L_{ts}(X) = \frac{1}{T} \sum_{i=1:T} -\log P(z_i/x_i) \quad (4.1)$$

$$L_1(X) = \text{CTC}(X) + \lambda L_{ts}(X) \quad (4.2)$$

CTC 无需对输入序列进行分割，这就不需要定位固有的模糊的标签边界，也不明确地标记标签间的依赖关系，从而使 CTC 解码更快。

CTC 基于条件独立的假设对标签之间的依赖关系进行建模，因此在字符层面对语言进行建模的能力有限，这通常需要在后端增加一个语言模型进行校正。

4.2 实验结果及分析

如图 4.2 所示，为在 CTPN 网络下，利用加权 CTC 算法优化得到的文本识别结果。

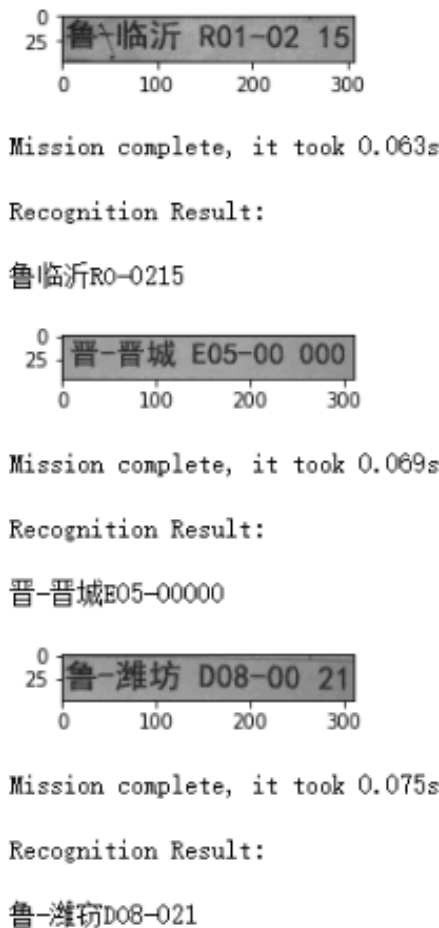


图 4.2 基于 CTC 算法优化后得到的具体参数

由于实验周期紧张，为了缩短模型训练的时间，本文参考文献^[26]，在 ImageNet 上的预训练权重来初始化参数，其余模块采用正太分布数据初始化。训练的模型具有阶梯化，每一个训练好的模型可以利用新数据面，对上一轮不完善的地方进行训练。几个参数如下：阶梯下降的动量 0.9，学习率 0.001，最后共训练 38 个 epoch。可以得到序列结果基本在 0.06~0.075s 之间。

结 论

由随着电商和快递行业的不断发展和普及,快递分拣作为其中的关键环节,也得到了越来越多的关注和研究。在这个过程中,基于深度学习的算法逐渐成为了主流。本文针对传统表单价值信息检测方法中存在的问题,提出了一种基于深度学习的快递分拣算法,并对其进行了详细介绍和实现。通过该算法,可以有效解决在物流传送带摄像头拍摄的照片中存在的复杂背景干扰问题,从而提高模型的准确率和可靠性。

本文设计了基于 Faster RCNN 的算法采用了基于卷积神经网络 DenseNet 的目标检测技术,通过对图像进行预处理、特征提取和分类等步骤,实现了对快递单中的各种信息的自动识别和分拣。针对多个价值区域、正负样本不均匀等问题,通过正样本补充实现了多目标检测和数据增强等功能,以进一步提高模型的准确性和鲁棒性。它可以有效解决传统表单价值信息检测方法中存在的问题,例如对图像质量的要求较高、无法处理复杂的背景干扰等。

OCR 模块设计了 CTPN+DenseNet+CTC 的结构。由于数据量大,在快递单信息识别中的 DenseNet 发挥了很好的作用,提升了识别的速度。在自建数据集上的验证,很好的证明了模型优秀的检测准确度和相对优秀的识别速度。虽然该算法已经在一定程度上解决了快递分拣中的关键问题,并实现了良好的效果,但是在实际应用中仍存在着一些挑战和值得进一步改进的方面。

由于随着电商和快递行业的不断发展和创新,快递单的种类和格式也会不断变化和扩展,因此如何应对更多样化的数据挑战,提高算法的鲁棒性和可适应性,将是一个重要的研究方向。其次,由于在物流传送带摄像头拍摄的照片中,存在着丰富的场景和背景干扰,因此如何更好地探索并利用这些信息,从而进一步提高模型的准确性和分类能力,也需要在未来的研究中得到更多的关注和探讨。

最后,在该算法的实现过程中,还存在着一些技术细节和工程实现上的问题,例如如何更好地采用更优秀的算法对不同区域的价值信息进行分类等。因此,加强这些方面的研究和实践,将有助于该算法的推广和应用在实际快递分拣场景中。

参 考 文 献

- [1] Girshick R, Donahue J, Darrell T, et al. Region-based convolutional networks for accurate object detection and segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 38(1): 142-158. Girshick R. Fast r-cnn[C]. Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
- [2] Ranjitha P, Shamjiith, Rajashekar K. Multi-Oriented Text Recognition and Classification in Natural Images using MSER[C]. 2020 International Conference for Emerging Technology(INCET). 2020
- [3] Hough P V C. Method and means for recognizing complex patterns: U.S. Patent 3,069,654[P]. 1962-12-18.
- [4] 孙丰荣, 刘积仁. 快速霍夫变换算法[J]. 计算机学报, 2001(10): 1102-1109.
- [5] 笑天. 浅谈日本包裹分拣[J]. 中国邮政, 1981(03): 25-27+44.
- [6] 杨涛. 传统图像分类与深度学习分类算法比较研究[J]. 荆楚理工学院学报, 2020, 35(02): 27-34. DOI: 10.14151/j.cnki.jclgxyxb.2020.02.005.
- [7] Jiang D, Li G, Tan C, et al. Semantic segmentation for multiscale target based on object recognition using the improved Faster-RCNN model[J]. Future Generation Computer Systems, 2021, 123(1)
- [8] 宋文, 肖建于. 基于笔画宽度变换的自然场景文本检测方法[J]. 计算机工程与应用, 2013, 49(009): 190-192.
- [9] 孙光民, 关世奎, 李煜, 郑鲲, 刘军华. 基于改进 CTPN 算法的试卷手写文本检测[J]. 信息技术, 2020, 44(09): 94-98. DOI: 10.13274/j.cnki.hdzj.2020.09.019.
- [10] Cypher R E, Sanz J, Snyder L. The Hough transform has $O(N)$ complexity on N - N mesh connected computers[J]. SIAM Journal on Computing, 1990.
- [11] Davies E R. Image space transforms for detecting straight edges in industrial images[J]. Pattern Recognition Letters, 1986, 4(3): 185-192.
- [12] Nabati R, Qi H. RRPN: Radar Region Proposal Network for Object Detection in Autonomous Vehicles[J]. IEEE, 2019.
- [13] 张震. 基于深度学习的快递表单信息处理及应用[D]. 南京邮电大学.
- [14] Hase H, Shinokawa T, Tokai S, et al. A robust method of recognizing multi-font rotated characters[C]// Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. IEEE Computer Society, 2004.
- [15] Campos T, Babu B R, Varma M. Character Recognition in Natural Images[C]// VISAPP 2009 - Proceedings of the Fourth International Conference on Computer Vision Theory and Applications, Lisboa, Portugal, February 5-8, 2009 - Volume 2. 2009.
- [16] Goel V, Mishra A, Alahari K, et al. Whole is Greater than Sum of Parts: Recognizing Scene Text

Words[C]// 2013 12th International Conference on Document Analysis and Recognition. IEEE Computer Society, 2013.

[17] Rodriguez J A . Label embedding for text recognition[C]// Bmvc. 2013.

[18] Gordo A . Supervised mid-level features for word image representation[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.

[19] Le Q V , Ranzato M , Monga R , et al. Building high-level features using large scale unsupervised learning[J]. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013.

[20] Khan S , Rahmani H , Shah S , et al. A Guide to Convolutional Neural Networks for Computer Vision[J]. Synthesis Lectures on Computer Vision, 2018, 8(1):1-207.

[21] Urban G , Geras K J , Kahou S E , et al. Do Deep Convolutional Nets Really Need to be Deep (Or Even Convolutional)?[J]. Nature, 2016, 521.

[22] 周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(06):1229-1251.

[23] Lawgali A . A Survey on Arabic Character Recognition[J]. International Journal of Signal Processing Image Processing & Pattern Recognition, 2015.

[24] He P , He T , Huang W , et al. Detecting Text in Natural Image with Connectionist Text Proposal Network:, 2016.

[25] Hochreiter S , Schmidhuber J . Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.

[26] 刘杰 . 快递单信息码图像检测和识别算法研究 [D]. 辽宁科技大学,2022.DOI:10.26923/d.cnki.gasgc.2022.000478.

[27] Hao L , Gao L , Yi X , et al. [IEEE 2016 12th IAPR Workshop on Document Analysis Systems (DAS) - Santorini, Greece (2016.4.11-2016.4.14)] 2016 12th IAPR Workshop on Document Analysis Systems (DAS) - A Table Detection Method for PDF Documents Based on Convolutional Neural Networks[J]. 2016.

[28] Zeng X , Xiang D , Peng L , et al. [IEEE 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) - Kyoto (2017.11.9-2017.11.15)] 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) - Local Discriminant Training and Global Optimization for Convolutional Neural Network Based Handwritten Chinese Character Recognition[J]. 2017:382-387.

致 谢

首先,我要感谢我的导师,郑老师和强老师。他们不仅为我提供了宝贵的指导和建议,也让我深入了解了研究的过程和方法。导师的支持和知识性的指导给予了我无限的动力和信心,在整个研究过程中给予了我无微不至的关心和帮助。

其次,我还要感谢上海建桥学院实验组的所有成员,他们给予了我很多探索的途径和思路。我们一起分享经验、讨论问题,在困难时鼓励我继续前进,在成功时与我分享喜悦,他们是我攻克项目期间最重要的伙伴之一。

此外,我也要感谢我的家人和朋友,在我学习和工作过程中,他们对我的支持和关爱始终坚定不移。他们的鼓励和支持,给予了我精神上的力量和安心感,让我更加专注和投入到研究之中。

总之,在完成这篇论文的过程中,我受到了众多人的帮助和支持,他们的付出是我不断进步和成长的动力源泉,他们的精神鼓舞是我追求卓越和突破自我的动力。在此,再次对所有人表示由衷的感激和敬意。