

中图分类号:TP391

盲审编号: 10406_081203_2004081200109_LW



南昌航空大学

硕士 学位 论文

题 目

基于图像分割的细粒度变形表格结构识别

学 科、专 业 计算机科学技术

专 业 代 码 XXXXXX

学 位 类 型 学术学位硕士

分类号:TP391

学校代码: 10001
学号: 2004081200109

南昌航空大学
硕士学位论文
(学术学位研究生)

基于图像分割的细粒度变形表格结构识别

硕士研究生:

导 师:

申请学位级别:

学科、专业: 计算机科学与技术

所在单位:

答辩日期:

授予学位单位:

Research Based on Transformer and CNN

A Dissertation

Submitted for the Degree of Master

On the Computer Science and Technology

By Long Liu

Under the Supervision of

A.Prof. Cihui Yang

School of Information Engineering

Nanchang Hangkong University, Nanchang, China

May, 2026

摘要

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
关键词： XXX;XXX;XXX

Abstract

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Keywords:XXX;XXX;XXX

目录

摘要	I
Abstract	II
目录	III
第 1 章 绪论	1
1.1 研究背景	1
1.2 研究意义与目的	2
第 2 章 预期目标	2
第 3 章 研究的主要内容	2
3.1 单元格物理坐标定位	2
3.2 基于单元格的细粒度分割模型的构建	2
3.3 单元格逻辑坐标定位	3
3.4 表格结构映射还原	3
第 4 章 研究方案	4
4.1 数据集的构建	4
4.2 单元格逻辑坐标定位	4
4.2.1 最短基准投影重叠率	4
4.2.2 质心扩展矩形	4
4.2.3 基于分治思路的单元格行列划分	5
第 5 章 拟解决的关键问题与创新点	5
第 6 章 研究计划进度	5
参考文献	6
攻读硕士期间发表论文及参加科研情况	7
致谢	8

第1章 绪论

1.1 研究背景

在信息化的时代，随着扫描、拍照等手段的普及以及大数据和人工智能的推动，纸质文档逐渐被电子文档取代。传统的纸质文档多依赖人工输入和手动整理，需要消耗大量人力资源与时间成本，而电子文档则得益于计算机技术的发展，能自动存储、编辑及检索，极大提升了信息处理之效率与准确性。

表格中不仅包含文字信息，还包含复杂的行列关系、合并单元格以及多种对齐方式等，这些因素使得光学字符识别（OCR）^[1,2]技术在提取表格数据时面临不小的挑战。比如在一些复杂表格中，虽然OCR能够识别出单元格中的文本内容，但难以准确推断出单元格之间的层级关系和从属结构，这在一定程度上限制了其在表格数据提取中的应用。

为了应对文档中复杂表格结构的解析需求，准确获取获取表格中的行列关系、单元格合并、对齐方式等特性，研究者们开始探索表格结构识别方法，表格结构识别指的是从文档图像中自动提取表格的布局信息，并重建其行列关系。一个标准的表格由若干行列组成，当前，表格结构识别的研究方法主要可以分为两类：基于规则的传统方法和基于深度学习的现代方法。基于规则的传统方法通常依赖于图像处理技术，例如边缘检测、投影分析、连通域分析等^[3]。这些方法通过寻找表格的行列线条来判断结构，但在实际应用中，表格往往存在复杂的合并单元格、倾斜、扭曲等形变，这些都使得基于线条的传统方法效果不佳。因此，很多新方法开始尝试不依赖明确的线条，而是通过单元格的定位来重建表格结构。通过定位单元格的起始和结束位置，可以有效避免线条缺失或干扰的影响，从而实现对复杂表格的识别。

近年来，深度学习方法在表格结构识别中的应用取得了突破性进展^[3]。深度学习模型尤其是卷积神经网络（CNN）在图像分割与特征提取方面展现了巨大的优势。对于表格结构的识别，部分基于深度学习的方法依然采用了线条检测来恢复表格的行列结构，例如通过检测图像中的水平和垂直线条来推测行列划分。这类方法通常能够应对大多数标准表格，但对于复杂的表格（如包含合并单元格或倾斜表格），识别效果却并不理想。而另外一些方法则不直接依赖于线条，而是通过单元格区域的检测和定位来判断表格的结构^[3]。这类方法通常依赖于深度神经网络的端到端训练，直接从图像中学习表格的布局和结构，并能自动进行表格的恢复和数据提取。

如今，表格图像的结构识别依然面临诸多难题。首先，表格的多样性和复杂性对识别提出了严峻挑战。现代文档中的表格形式多样，不仅包括基本的网格结构，还可能包含合并单元格、斜线分隔符、嵌套表格等复杂形式，传统表格识别算法往往难以处理。此外，形变表格的出现进一步增加了识别难度。例如，由于扫描角度的变化或内容被拉伸、倾斜等原因，表格图像可能出现变形，导致传统识别方法无法准确提取结构信息。尤其是表格中的单元格划分和内容对齐，在图像形变或变形后容易发生偏移，这使得表格的检测和重建变得更为复杂。因此，如何高效、准确地识别和重建表格结构，已成为亟待解决的难题。

1.2 研究意义与目的

在文档存储数据的方式中，表格占据着举足轻重的地位，尤其是在财务报表、学术论文、统计分析等领域，它们不仅是信息传递的重要载体，也是数据分析的基础。因此，从图像文档中自动提取表格信息并重建其结构的过程，逐渐成为计算机视觉和图像处理领域备受关注的研究方向。这项工作的核心挑战在于如何高效且准确地从复杂的图像背景中分离出表格区域，识别表格内部的行列布局，并进一步精准提取每个单元格内的数据内容。这不仅有助于提升信息处理的自动化水平，减少人工干预所带来的误差，同时也为后续的数据分析提供了更加可靠的支持。此外，随着移动设备和数字化办公环境的普及，能够快速准确地处理各种形式存在的表格图像，对于促进跨平台的信息共享与交流也具有重要意义。

第 2 章 预期目标

通过对本文提出的问题的研究，结合现有的表格识别方法，预期达到以下目标：

对于存在形变的有网格线表格图像，设计并实现一种基于深度学习的图像分割方法，尽最大可能识别和分割表格中每个单元格的细粒度的坐标，此外，本课题还将设计一种算法，用于确定单元格之间的相对位置，能获取单元格在表格中行、列以及合并信息。

第 3 章 研究的主要内容

3.1 单元格物理坐标定位

在表格结构识别中，单元格物理坐标用于精确描述单元格在图像中的位置，合适的单元格物理坐标的表示方法为表格结构识别提供了精确的定位基础，尤其是在表格结构重建和数据提取过程中，它们能够确保单元格位置的准确性，也为后续的单元格逻辑坐标定位提供了基础。

3.2 基于单元格的细粒度分割模型的构建

传统的表格结构提取方法主要依赖于线段检测和几何模型推理，这些方法虽然在处理标准化表格时表现良好，但在面对形变严重的表格时，其局限性逐渐显现。尤其是直线检测算法，通常通过霍夫变换等技术检测图像中的直线，以识别表格的行和列。尽管这种方法在结构清晰的表格中运作良好，但一旦表格因拍照或扫描产生了形变，行列的线条可能出现弯曲或断裂，从而导致算法无法有效识别，最终影响表格结构的提取精度。于是，有的研究转向通过四个关键角点来表示单元格物理坐标。



图 3.1 关键点检测与图像分割

然而，基于关键点推理的模型也存在一定局限性，如图3.1，单元格的边界可能形变过于严重，导致定位出得到的单元格在后期的OCR识别丢失了大量信息，也给逻辑坐标还原带来了巨大的挑战。这种情况下，单纯依靠几何形状的描述已无法满足实际需求，为了解决这些问题，本课题关注基于图像分割的技术。与关键点推理模型的方式不同，图像分割通过对图像进行像素级的处理，能够捕捉到每个单元格的边界信息，并保留更多细节。特别是在处理复杂形状和形变严重的情况下，图像分割算法展现出强大的能力，这种细粒度的识别方法能最大程度上保留单元格的内容信息区域。

3.3 单元格逻辑坐标定位

与单元格物理坐标不同，单元格逻辑坐标主要描述单元格之间的相对位置和从属关系，通常用于表示表格的布局结构和单元格之间的逻辑关系。通过获取单元格逻辑坐标，可以更加清楚的了解表格中单元格之间的关系，如单元格本身行列信息、单元格的合并信息、单元格的跨行跨列信息等。这对于表格结构映射还原具有重要意义。

3.4 表格结构映射还原

表格重组是表格识别中的一个辅助步骤，目的是通过物理坐标和逻辑坐标的结合，将表格中的单元格从图像中的位置恢复为规整的、标准化的表格布局。这一过程帮助系统将单元格准确排列，确保不同单元格之间的空间和结构关系正确。

第4章 研究方案

4.1 数据集的构建

4.2 单元格逻辑坐标定位

单元格逻辑坐标的定位是表格结构识别中的关键步骤之一。其基本思路是通过分析相邻关系，识别所有属于同一行或同一列的单元格，并将这些单元格归入相应的集合中。对于非跨行单元格，它只能出现在一个行集合中；而对于跨行单元格，则可以出现在多个行集合中，以表示该单元格跨越的行范围。列集合同理。为了解决判断相邻关系这一关键问题，本文提出了两个概念：最短基准投影重叠率和质心扩展矩形。

4.2.1 最短基准投影重叠率

首先，为了判断单元格是否相邻，需要引入最短基准投影重叠率（Minimum Projection Overlap Ratio, MPoR）来判断某个方向上线的同属程度，设两条线段 L_1 和 L_2 ，它们在指定方向上的投影区间为 $[a_1, b_1]$ 和 $[a_2, b_2]$ ，则最短基准投影重叠率定义为：

$$\mathcal{P}(L_1, L_2) = \frac{\min(b_1, b_2) - \max(a_1, a_2)}{\min(b_1 - a_1, b_2 - a_2)} \quad (4.1)$$

其中 $\min(b_1, b_2) - \max(a_1, a_2)$ 实际上就是投影的重叠长度（如果出现负数，那么说明在该方向上是没有重叠的，两条线之间也就没有同属关系）。

通过单元格的边界线段判断单元格是否在 y 轴方向上是否隶属同一行或者 x 轴方向上是否隶属同一列（后文如果没有强调，这里使用的都是两个矩形相邻的边计算 MPoR 值）。

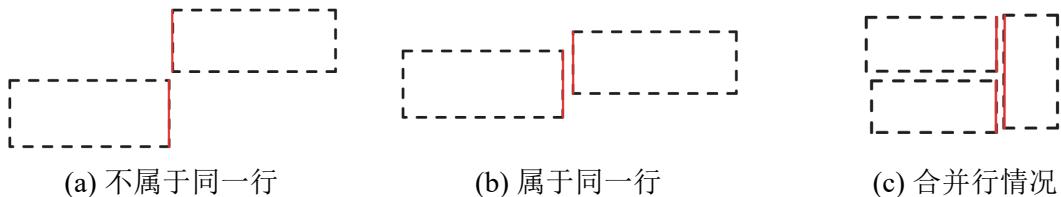


图 4.1: (a) 是左右两个矩形在水平方向的 MPoR 值小于给定的阈值 t ，所以可以判定不属于同一行，(b) 是左右两个矩形在水平方向的 MPoR 值大于给定的阈值 t ，所以可以判定属于同一行，(c) 是左边两个矩形与右边矩形在水平方向的 MPoR 值均大于给定的阈值 t ，所以可以判定右边的矩形是合并单元格，并且既属于左上矩形所在行又属于左下矩形所在行

4.2.2 质心扩展矩形

实例分割的轮廓是复杂的掩码，为了更加方便的确定单元格的逻辑坐标，需要引入质心扩展矩形（Centroid-Expansion Rectangle, CER）来简化掩码，首先计算掩码的质心，通过质心向上下左右四个方向扩展，计算质心四个方向的射线与轮廓的交点，于是就会得到掩码在垂直高度与水平宽度上的参考值，结合宽高和质心可以得到质心扩展矩形，如图4.2(a)，质心扩展矩形的宽高具有参考作用，但是也会存

在误差，所以最终参考的的垂直范围和水平范围是通过修正的，具体的修正方式如图4.2(c) 和4.2(b) 所示，将质心的四边界向内（质心方向）收缩 δ 个像素，并计算收缩后的四边界所在位置的延长线被轮廓所截取的范围，修正之后的高度范围选取左右截取范围的最大值，修正之后的宽度范围选取上下截取范围的最大值。

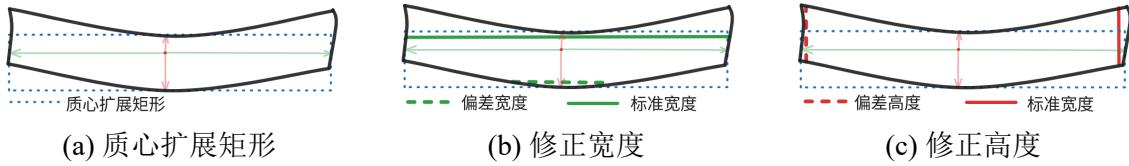


图 4.2 质心扩展矩形

4.2.3 基于分治思路的单元格行列划分

第 5 章 拟解决的关键问题与创新点

第 6 章 研究计划进度

参考文献

- [1] 陈登原. 国史旧闻: 第 1 卷[M]. 北京: 中华书局, 2000: 29.
- [2] 哈里森 • 沃尔德伦. 经济数学与金融数学[M]. 谢远涛, 译. 北京: 中国人民大学出版社, 2012: 235-236.
- [3] 北京市政协民族和宗教委员会, 北京联合大学民族与宗教研究所. 历代王朝与民族宗教[M]. 北京: 民族出版社, 2012: 112.

攻读硕士期间发表论文及参加科研情况

这里写科研成果

致谢

这里写致谢

南昌航空大学硕士学位论文原创性声明

本人郑重声明：所呈交的硕士学位论文，是我个人在导师指导下，在南昌航空大学攻读硕士学位期间独立进行研究工作所取得的成果。尽我所知，论文中除已注明部分外不包含他人已发表或撰写过的研究成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中作了明确地说明并表示了谢意。本声明的法律结果将完全由本人承担。

签名：_____ 日期：_____

南昌航空大学硕士学位论文使用授权书

本论文的研究成果归南昌航空大学所有，本论文的研究内容不得以其它单位的名义发表。本人完全了解南昌航空大学关于保存、使用学位论文的规定，同意学校保留并向有关部门递交论文的复印件和电子版本，允许论文被查阅和借阅。本人授权南昌航空大学，可以采用影印、缩印或其他复制手段保存论文，可以公布论文的全部或部分内容。同时授权中国科学技术信息研究所将本学位论文收录到《中国学位论文全文数据库》，并通过网络向社会公众提供信息服务。

(保密的学位论文在解密后适用本授权书)

签名：_____ 导师签名：_____ 日期：_____