



## 1 绪论

绪论：绪论相当于论文的开头，它是三段式论文的第一段（后二段是本论和结论）。绪论与摘要写法不完全相同，摘要要写得高度概括、简略，绪论可以稍加具体一点文字以 1000 字左右为宜。绪论一般应包括以下几个内容：

- ① 为什么要写这篇论文，要解决什么问题，主要观点是什么。
- ② 对本论文研究主题范围内已有文献的评述（包括与课题相关的历史的回顾，资料来源、性质及运用情况等）。
- ③ 说明本论文所要解决的问题，所采用的研究手段、方式、方法。明确研究工作的界限和规模。
- ④ 概括论文的主要工作内容。

### 1.1 研究背景和意义

随着信息技术的高速发展，人们获取知识的方式逐渐发生了改变，线上电子文件由于其传输方便、保存成本低、可即时阅读等优点，越来越获得现代人的青睐。大多数电子文档首先是以图片的形式上传到互联网中，虽然保留了原文件的信息，但是计算机无法将其直接转化为更加标准的文档格式，因为其文件中可能包含汉字、英文、特殊符号、数学公式等不同类型的字符，虽然目前对于自然语言文字的字符识别技术已经十分成熟，但是对于数学公式的识别应用还并不十分成熟。

数学公式中包含着大量的有效信息，人无法看懂用自己不懂的语言著作的书籍，但是却能通过其中的公式判断这本著作是否有翻译阅读的价值，一道数学问题仅仅通过看关键公式就可以知道解题思路甚至答案是否正确，同时对于一些理工科的学术论文，数学公式更是其中的核心内容，由此可见数学公式在教育 and 学术中举足轻重的地位。

且自 2020 年新冠疫情爆发以来，线上教育的需求呈现爆发式增长，传统教育中教师和学生的课堂互动环节和课后的作业练习批改环节都需要教师的直接或者间接把控。批改作业和考试试卷是学校主要教学工作中必不可少的一环，传统的纸质阅卷过程极其繁琐且经常出现误判，利用如今兴起的基于深度学习技术的光学字符识别技术则可以很大程度上缓解这一痛点。如果计算机可以准确识别出数学表达式，则可以将老师从某些枯燥重复的客观题批阅中解放出来，在节省时间和体力的同时也可以降低

---

<sup>①</sup> 脚注是对文中有关内容的解释、说明或补充，使用上角标（序号①、②…）标注，脚注可用小号字（一般小五号宋体）列在相应正文同一页最下部并与正文部分用细线（版面宽度的 1/4 长）隔开。（删除脚注的方法：直接删除正文中的脚注编号即可）

批改的错误率，进一步提升教学质量。

因此数学公式识别技术的相关研究能够有利线上教育的智能化管理，对线上教育、科研学习的发展都有着十分重要的现实意义。

## 1.2 数学公式识别研究历史

数学公式识别的研究受到数字图像处理技术发展的制约，总体上可以分为三个阶段：萌芽期，累积期和初步应用期。下文将简要介绍三个阶段的数学公式识别处理的发展情况（渐简明）。

### 1.2.1 萌芽期（1968 年至 1990 年）

数学公式的识别是 OCR 技术的一个重要方向，OCR 技术，即光学字符识别技术最早由德国科学家 Tausheck 博士于 1929 年提出，但首次提出数学公式的问题的是 Anderson，在他的 1968 年的博士学位论文中（注释）。在这篇论文中，他首次提出基于句法导向的数学公式识别概念，但随后的十几年时间中，仅有少量文章发表，这些文章大多是从理论角度上对数学公式进行的文法研究和结构描述，发展较为缓慢，如此长时间的发展缓慢是由于 20 世纪 60 年代是模式识别的早期阶段，理论尚不完善，OCR 技术还不够成熟，无法将工作与实际很好的结合。

### 1.2.2 累积期（1991 年至 1999 年）

直到 20 世纪 90 年代，由于深度学习的兴起，OCR 技术的进一步发展，关于数学公式识别相关的应用研究才逐渐得到重视。

这一期间有超过 40 篇的相关论文发表，研究工作包含了数学公式处理的主要研究范围，比如：定位数学公式、切割目标数学公式、字符识别以及数学公式结构的分析与重组，在这期间也有个别实现较为完整实验系统见到报道，但是这些工作的可以处理的实际情况有限，也并没有在现实生活中的进行实际检验，没有实际的测试数据说明效果。

Berkely 大学的 FateMan 从 1994 年开始致力于数学公式识别问题的研究，Blostein 于 1997 年提出数学公式的识别主要氛围两个部分：字符识别和字符重组分析。其中字符识别主要分为字符预处理、字符分割和识别三个过程，符号重组分析主要分为确立公式中所包含字符的空间关系、逻辑关系和构造数学公式意义这三个关系。在一般情况下，对数学公式的识别研究主要集中在字符识别以及符号之间的结构分析上。

此阶段的数学公式识别研究较上一阶段发展速度较快的主要原因是深度学习和 OCR 技术发展已经较为成熟，数学公式识别的研究逐渐得到重视。

### 1.2.3 应用期（2000 年至今）

虽然目前 OCR 技术已经比较成熟，在图片识别、文档图像处理识别上的识别率已经很好，但是数学公式识别的发展仍然较为落后。在 CROHME 2016 比赛中，各个竞

赛队伍中公式识别率最高的是 MyScript，为 67.65%，但是由于此队伍使用了私有的数学公式数据集，因此与其他队友并不可比性。其他队伍的识别率最高没有超过 50%。

目前还是有了一些主要用于数学公式识别的商业应用出现，其中最为著名的就是 Mathpix Snipping Tool，这款软件对于一些比较常规的公式识别效果较好，但是对于一些比较复杂的数学公式识别还是较为容易出现错误。

## 1.3 本文的主要内容和结构安排

（说明本论文所要解决的问题，所采用的研究手段、方式、方法。明确研究工作的界限和规模。）

### 1.3.1 主要内容

本论文立足于目前数学公式识别领域的研究现状，结合深度学习和自然语言处理相关知识，利用端到端的编码器——解码器模型，使用基于 Python 的 PyTorch 机器学习库，开发了一个将带有数学公式图片识别为与其对应的 LaTeX 公式的实用工具。

从数学公式识别的研究上来讲，主要有两种方法，分别是基于 OCR 的传统识别方法基于端到端的编码器——解码器识别方法；本文在深刻论述了这两种识别方法的优缺点之后，选择了使用后者进行工程实现，理由见第二章。为了明确方向，本文主要针对于将输入的纯数学公式图片识别为对应公式表达式，并不包含在文档中识别并切割公式等过程。

### 1.3.2 结构安排

论文总共分为五章，各章节内容的安排如下所示：

第一章介绍了数学公式识别研究的意识和历史研究现状，并对本论文主要工作进行概述，最后简要介绍文章结构。

第二章主要研究目前主流的两种数学公式识别方法的优缺点，深刻地比较了两种方法的不同点。并阐述了选择使用端到端的基于编码器——解码器方法的理由。

第三章主要介绍了深度学习和编码器——解码器模型中的一些相关技术，存在的缺点和优化方法。

第四章介绍了有关工程开发实现细节描述以及编码工作，测评了训练之后模型的各方面性能，并请人工对此工具的使用效果进行各方面的评价。实验表明该工具在一些非极端复杂的数学公式识别上较为良好。

第五章对本课题的各方面研究工作进行了分析和总结，并对未来工作的研究方向进行了展望和设想。

- 1) 标题 4
- (1) 标题 5
- a) 标题 6
- b) 标题 6
- (a) 标题 7

图、表、公式等一律用阿拉伯数字分章连续编号，如 图 1-3、表 2-1、(3-2) 等。图、表、公式等与正文之间间隔 0.5 行。

图应有图题，表应有表题，并分别置于图号和表号之后，图号和图题应置于图下方的居中位置，表号和表题应置于表上方的居中位置。引用图或表应在图题或表题右上角标出文献来源。

若图或表中有附注，采用英文小写字母顺序编号，附注写在图或表的下方。

#### 图：

(1) 插图须紧跟文述。在正文中，一般应先见图号及图的内容后再见图，一般情况下不能提前见图，特殊情况须延后的插图不应跨节；

(2) 提供照片应大小适宜，主题明确，层次清楚，金相照片一定要有比例尺；

(3) 图应具有“自明性”，即只看图、图题和图例，不阅读正文，就可理解图意。

通常使用的函数图采用简化形式，称为简写函数图，例如图 1-1。

图中的标目是说明坐标轴物理意义的项目，它是由物理量的符号或名称和相应的单位组成。物理量的符号由斜体字母标注，单位的符号使用正体字母标注，量与单位间用斜线隔开。例如： $I/A$ ， $\rho/\text{kg}\cdot\text{m}^{-3}$ ， $F/\text{N}$ ， $v/\text{m}\cdot\text{s}^{-1}$  等等。

(4) 图中用字为五号，如排列过密，用五号字有困难时，可小于五号字，但不得小于七号字。

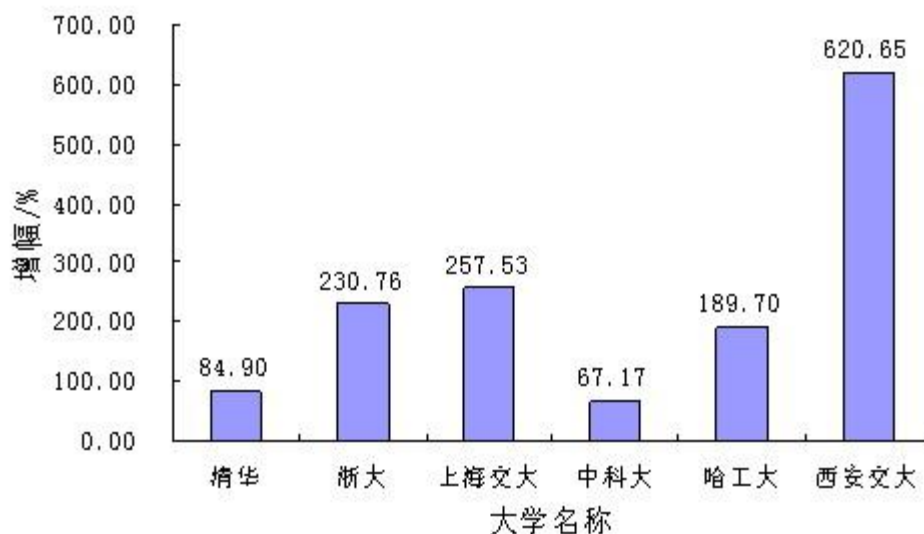


图 1-1 2005 年相对 2001 年，5 所大学 SCI-e 文献总数增幅图

(5) 图的大小一般为宽 6.67 cm×高 5.00cm。特殊情况下，也可宽 9.00 cm×高 6.75cm，或宽 13.5 cm×高 9.00cm。总之，一篇论文中，同类图片的大小应该一致，编排美观、整齐。

(6) 一幅图如有若干幅分图，均应编分图号，用(a), (b), (c), ..... 按顺序编排；且各分图的分题注直接列在各自分图的正下方，总题注列在所有分图的正下方正中，如下图所示：

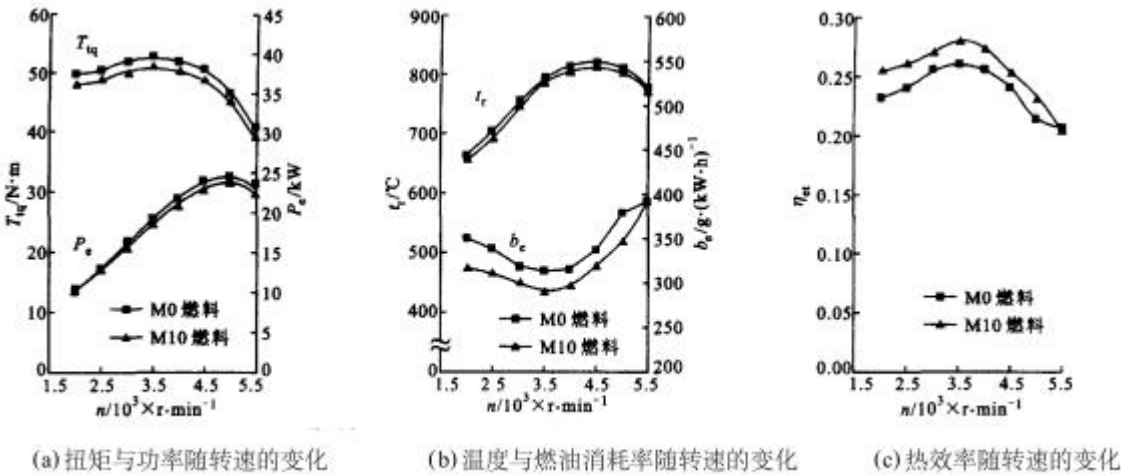


图 1 M10 燃料对汽油机全负荷速度特性的影响

表：

- (1) 如某个表需要转页接排，在随后的各页上应重复表的编号。编号后跟表题（可省略）和“（续）”，如表 1（续），续表均应重复表头和关于单位的陈述。
- 表格的设计应紧跟文述。表的编排一般是内容和测试项目由左至右横读，数据依序竖读，应有自明性。若为大表或作为工具使用的表格，可作为附表在附录中给出，论文中的表格参数应标明量和单位的符号；
- (2) 表中各物理量及量纲均按国际标准(SI) 及国家规定的法定符号和法定计量单位标注；
- (3) 一律使用三线表，与文字齐宽，线粗 1.5 磅。表内线，线粗 1 磅。例如表 1-1；
- (4) 使用他人表格须注明出处。
- (5) 表中用字为五号字体。如排列过密，用五号字有困难时，可小于五号字，但不小于七号。
- (6) 表格必须通栏，即表格宽度与正文版面平齐，如下表所示。

表 1-1 文献类型和标志代码

文献类型	标志代码	文献类型	标志代码
普通图书	M	会议录	C



汇编	G	报纸	N
期刊	J	学位论文	D
报告	R	标准	S
专利	P	数据库	DB
计算机程序	CP	电子公告	EB

在三线表中可以加辅助线，以适应较复杂表格的需要，如表 1-2 所示。

表 1-2 方弯管内流动最大速度比较

项目	层流		紊流	
	0°截面	90°截面	0°截面	90°截面
理论值 $V_{max}/m \cdot s^{-1}$	0.04	0.03	1.30	1.25
计算值 $V_{max}/m \cdot s^{-1}$	0.04	0.03	1.26	1.21
误差/%	0.00	3.12	3.07	3.20

### 公式:

(1) 公式应另起一行，居中编排，较长的公式尽可能在等号后换行，或者在“+”、“-”等符号后换行。公式中分数线的横线，长短要分清，主要的横线应与等号取平。

(2) 公式后应注明编号，直接置于小括号中，如(3-1)，右对齐顶格编排，中间不加虚线。编号前不写“公式”或“式”等字样。

(3) 公式下面的“式中”两字左起顶格编排，后接符号及其解释；解释顺序为先左后右，先上后下；解释与解释之间用“；”隔开。

范例:

$$q = k_d H^x \quad (1-1)$$

式中:  $q$  —— 灌水器流量/L·h-1;  $k_d$  —— 流量系数;  $H$  —— 工作压力/m;  $x$  —— 流态指数。

(此处, “式中: ” 为顶格输出)

(1-1) 中, .....

$$\sqrt{b^2 - 4ac} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \frac{n!}{r!(n-r)!} \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (1-2)$$

2 绪论：绪论相当于论文的开头，它是三段式论文的第一段（后二段是本论和结论）。绪论与摘要写法不完全相

---



## 2 数学公式识别方法介绍

数学公式识别相较于自然语言识别的差异十分明显，首先数学公式不同于自然语言，自然语言是一维的字符串的拼接，但数学公式是一种二维结构的语言；其次数学公式中所包含的符号集可能更加复杂，很多时候可能在数学公式中发现有各种自然语言文字出现；且数学公式中的符号是上下文敏感的，种种困难使得一般的 OCR 技术无法处理数学公式识别。

本章将主要介绍目前主流的识别数学公式识别方法，并对每种方法的差异与优缺点进行研究。

### 2.1 基于 OCR 的传统方法

传统的基于 OCR 的数学公式识别研究主要将公式识别任务分为两部分：符号识别和结构分析重组两个阶段，见图 2.1（上标：Mathematical expression recognition）

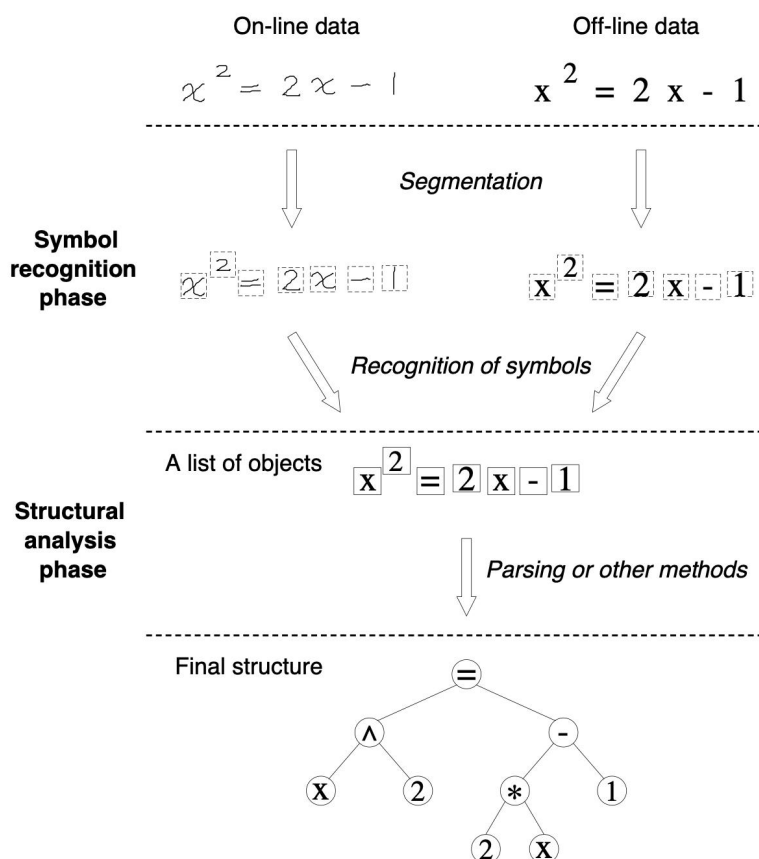


图 2-1 传统方法识别数学公式流程

### 2.1.1 符号识别

目前已经有了许多成熟的网络可以应用到数学公式符号的识别中，并且识别率可以超过 95%，但这些识别网络的输入输入往往都是已经处理好的单个的数学公式符号，即所有的符号都已经分割预处理完成。但我们想要进行完整数学公式识别的时候，输入往往是完整的数学公式图片，所以符号识别过程中也包含了图片预处理、字符分割处理等操作。用于字符分割的常用方法有投影方法、边界盒子法、连通域算法等，不管采用何种方法，字符分割等预处理过程的好坏对后续的识别效果影响十分巨大。

### 2.1.2 结构分析与重组

此阶段利用对输入图片进行预处理之后的分割字符和相关空间信息，并进行相应的语法分析，把一个个单独的识别出来的符号重构为完整的数学公式。目前已经提出了多种用于解决字符结构重组问题的方法，其中较为常见的有：基于二维 HMM 的、基于表达式树的、基于公式语法书的。其中目前表现的较为出色的方法是基于语法公式树的方法，虽然此方法表现较为出色，但是此类方法往往需要大量的先验知识来表示数学公式中的语法。

### 2.1.3 传统数学公式识别方法总结

传统数学公式识别方法中将数学公式的识别过程分为符号识别和结构分析两个阶段分开解决，虽然在最终的识别效果上可以达到要求，但是却存在着以下几点问题：

- 1) 符号识别任务中需要对输入数学公式图像进行字符分割，虽然目前对于单个字符的识别正确率已经很高，但是对于图像的字符分割效果还不够优秀，如图 2.1 可见，由于数学公式的组合的灵活性和多样性，在字符分割过程中常常会出现无法识别出需要分割或者过渡分割的情况，不管出现哪一种情况，都会对之后的字符识别过程和结构分析重组过程造成很大的影响。
- 2) 在结构分析过程中，需要大量的先验知识作为支撑，也即规定好的数学公式语法。靳简明（上标）在数学公式识别结构分析过程中提出了混合分析的方法，将需要识别的数学表达式分解为一系列子表达式，为此他总结出了 11 种通用的基本公式类型，并以此来规定每类公式的在分解和重组过程，在具体分解的过程中使用了基于结构的分析方法。
- 3) 识别过程的复杂度较高。数学公式识别任务中，特别是手写数学公式的识别中会遇到各种歧义问题，极易导致最后的识别出现误差，未来保证识别结果的正确，往往需要在识别过程中引入概率的因子进行识别评估，往往可能有多种可能性，要求出概率最高的一种，所以计算的复杂度比较高。






原图	$F(x$	$dx$	$\omega$
分割效果	  	$dx$	 
统计	正确分割	不能分割	过度分割

图 2-2 字符分割过程中可能出现的问题