



聚



籍



聚 籍 ———

基于深度学习的  
古籍数字化平台

# 目录

摘要 .....	4
一、    市场背景 .....	6
二、    产品简介 .....	6
三、    核心技术 .....	9
1. 技术介绍 .....	9
2. 创新点 .....	10
第一章 系统背景和可行性分析 .....	13
1.1 项目背景分析 .....	13
1.1.1 古籍阅读者的增加 .....	13
1.1.2 两岸关系进入新纪元 .....	13
1.1.3 深度学习的发展 .....	14
1.1.4 古籍数据化发展空间巨大 .....	15
1.2 可行性分析 .....	16
1.2.1 市场可行性分析 .....	16
1.2.2 技术可行性 .....	17
第二章 产品介绍及项目技术介绍 .....	19
2.1 产品简介 .....	19
2.2 功能介绍 .....	19
2.2.1 网站 聚籍网 .....	19
2.2.2 APP 聚籍 .....	22
2.2.3 微信小程序 繁体识别 .....	25
2.3 核心技术介绍 .....	25
2.3.1 系统框架及步骤 .....	26
2.3.2 图像预处理 .....	28
2.3.3 文字定位 .....	30
2.3.4 文字分割 .....	34
2.3.5 文字识别 .....	35
第三章 公司分析与发展战略 .....	39
3.1 公司基本情况 .....	39
3.2 组织结构 .....	39
3.3 部门职能及构成 .....	40
3.4.人力资源管理 .....	41
3.4.1 创新机制 .....	41
3.4.2 激励机制 .....	43

3.5. 总体发展战略 .....	43
3.6. 分阶段发展战略 .....	44
第四章 市场分析与营销推广 .....	46
4.1 市场分析 .....	46
4.1.1 行业概况 .....	46
4.1.2 环境分析 .....	49
4.1.3 目标市场选择 .....	51
4.1.4 竞争分析 .....	52
4.2 营销推广 .....	54
4.2.1 营销理念 .....	54
4.2.2 营销策略——4ps 理论 .....	55
4.2.3 推广方式 .....	56
第五章 商业模式和盈利模式 .....	59
4.1 商业模式 .....	59
4.1.1 客户细分 .....	59
4.1.2 价值主张 .....	59
4.1.3 客户关系 .....	60
5.2 盈利模式 .....	60
5.2.1 前期盈利 .....	60
5.2.2 中期盈利 .....	61
5.2.3.后期盈利 .....	62
第六章 财务分析 .....	63
6.1 融资计划 .....	63
6.1.1 前期融资方案 .....	63
6.1.2 中期融资方案 .....	64
6.1.3 后期融资方案 .....	65
6.2 财务预测 .....	65
6.2.1 基本财务假设 .....	65
6.2.2 公司前期投资预测 .....	67
6.2.3 未来五年收支预测 .....	69
6.4 主要财务指标分析 .....	74
6.4.1 静态投资回收期 .....	74
6.4.2 投资净现值（NPV） .....	74
6.4.3 内涵报酬率（IRR） .....	74
第七章 风险管理 .....	75



7.1 财务风险 .....	75
7.1.1 概述.....	75
7.1.2 产生原因 .....	75
7.1.3 对策.....	76
7.2 技术风险 .....	76
7.2.1 概述.....	76
7.2.2 产生原因 .....	76
7.2.3 对策.....	77
7.3 市场与竞争风险.....	77
7.3.1 概述.....	77
7.3.2 产生原因 .....	77
7.3.3 对策.....	78
7.4 成员管理风险.....	78
7.4.1 概述.....	78
7.4.2 产生原因 .....	78
7.4.3 对策.....	79

## 图索引

图 1 “聚籍”网在线古籍阅读界面 .....	7
图 2 “聚籍”app 论坛界面 .....	8
图 3 上传图片识别转换（识别） .....	9
图 4 上传图片识别转换（转换） .....	9
图 5 “聚籍”网在线阅读古籍界面 .....	20
图 6 “聚籍”网论坛社区界面 .....	20
图 7 “聚籍”网实体书商城界面 .....	21
图 8 上传图片识别转换（识别） .....	22
图 9 上传图片识别转换（转换） .....	22
图 10 “聚籍”app 界面 .....	23
图 11 “聚籍”app 繁体识别界面 .....	24
图 12 繁体识别微信小程序界面 .....	25
图 13 系统框架图 .....	27
图 14 图像识别流程图 .....	28
图 15 图像预处理流程图 .....	28
图 16 二值化处理过的图片 .....	29
图 17 多次膨胀处理 .....	30
图 18 一次膨胀处理 .....	30
图 19 原图 .....	30
图 20 SSD 的网络结构 .....	31
图 21 SSD Feature Map .....	32
图 22 Conv9 输出特征图 5x5 .....	32
图 23 定位效果图 .....	33
图 24 原图 .....	34
图 25 基于神经网络图像分割大体步骤 .....	35
图 26 CRNN 的网络架构 .....	36
图 27 感受野 .....	37
图 28 系统识别出的图片 .....	38
图 29 公司组织形式 .....	39
图 30 古文化产业占 GDP 比重变化趋势 .....	50
图 31 马斯洛需求层次理论 .....	51
图 32 SWOT 分析图 .....	53
图 33 线上推广图 .....	56
图 34 公司股本结构与规模图 .....	64

图 35 公司前期投资方向 .....	67
图 36 风险种类图 .....	75

## 表索引

表 1 部门职能及其构成表 .....	40
表 3 预收费标准表 .....	61
表 4 公司股本结构与规模 .....	63
表 5 平台前期投资预算表 .....	68
表 6 未来五年现金支出预测表 .....	69
表 7 未来五年折旧与摊销预测表 .....	71
表 8 未来五年收入情况预测表 .....	71
表 9 利润表 单位（元） .....	72
表 10 资产负债表 单位（元） .....	73

# 摘要

## 一、市场背景

### 1、古籍数据化发展空间巨大

数字图书馆和大型电子文献资料库的开发和建设,近年来业已成为学术界和相关技术部门的一个热门话题。在这个话题中,古籍的数字化也得到了高度重视。并且,伴随着传统文化的传播,数字化的古籍在学术研究中的利用率也在不断提高。而古籍数字化的价值,也正是在这样的使用中一点一点体现出来。

但目前古籍数据库的发展还有很多的不足,比如文件格式过多,不方便用户阅读,技术上的缺陷等等。而中国古籍的数量巨大,范围宽广,形式丰富,都表明了古籍数据化的发展空间巨大。

### 2、古籍阅读者的增加

随着高等教育的范围逐步扩大以及弘扬中国传统文化思想的传播,热爱历史古籍的人数也不断增加,对于刚接触古籍的初学者,古籍中复杂的繁体字无疑是阅读的一大障碍,我们的平台将为其提供繁体字与简体字对比版本,既能保留古籍的完整性,又能很好的帮助他们减少阅读障碍,降低入门门槛。注释功能也有利于阅读者的学习和理解。

### 3、现存繁体字识别系统局限性大

古籍中使用的大多是繁体字,目前网络上已经有不少的繁体字转换系统可以使用了,但这些系统还存在不少问题,比如只能针对一种繁体字体系进行转换,转换时只能输入文字,无法智能识别图片上的文字进行转换等。而我们平台独创了一套繁体字识别系统,基于深度学习,能做到用户仅需上传一张图片,我们的系统就能智能识别图片中的繁体字,进行转换,同时系统还能不断学习,不断提高识别的准确率。

## 二、产品简介

“聚籍”是专注于古籍数字化和古籍文化交流的综合服务平台,包含网站、



app 及相应的小程序。平台结合我们独创的繁体识别转换系统，推出了在线阅读古籍、论坛社区、上传图片文字识别等功能，希望为用户带来更好的古籍阅读体验，推动优秀传统文化的传播。

1. 在线阅读古籍

在线阅读古籍旨在为用户提供优质的古籍阅读体验，满足其古籍学习研究的需求。阅读界面中将逐页显示古籍的图片，客户可再根据自身需求选择显示繁体字或简体字文本以便对比阅读，将显著提高用户的使用效率，大大拓展用户的信息视角，为研究工作带来便利。个人用户和机构用户注册过个人账户的用户可以对阅读的古籍进行批注，同时个人中心还有收藏、检索记录、阅读历史等辅助用户进行阅读的个性化功能。

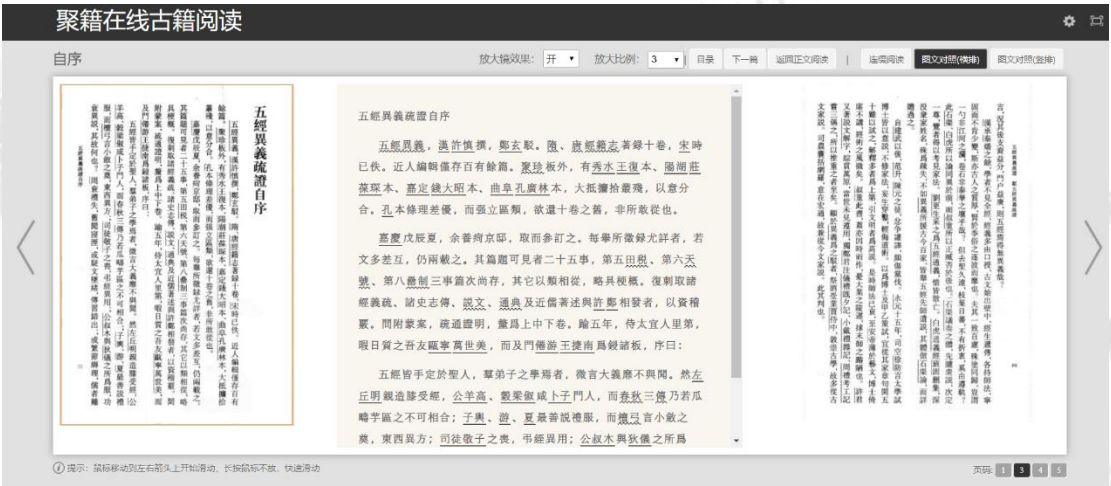


图 1“聚籍”网在线古籍阅读界面

2. 古籍论坛社区

古籍论坛社区是“聚籍”推出希望将学界与大众结合，推进估计学术研究的良性生态系统。在论坛社区中用户可以发表文章，自由发问和建组讨论，并根据用户的关注进行个性化推送。用户可以在论坛社区上分享随笔、研究心得等个性文章，随时在线提问，互相解决问题；也可以根据个性化的需求结成组群，进行内部讨论和分享。用户可以提出一些自己关于繁体字、古籍、古文化的问题和看法，互相勉励交流信息，共同学习进步。





图 2“聚籍”app 论坛界面

### 3. 上传图片识别转换

鉴于一部分古籍图片有轻微破损、模糊不清，不易阅读；且因古籍大多使用繁体字，对初学者阅读有困难。我们在工具中推出了上传图片识别转换的功能供用户使用。该功能利用了我们的核心技术，用户可自主上传古籍图书的图片，得到准确率高的繁简体识别文本，工具可以识别一些模糊不清的图片以及手写体，以帮助用户更好的阅读理解古籍。

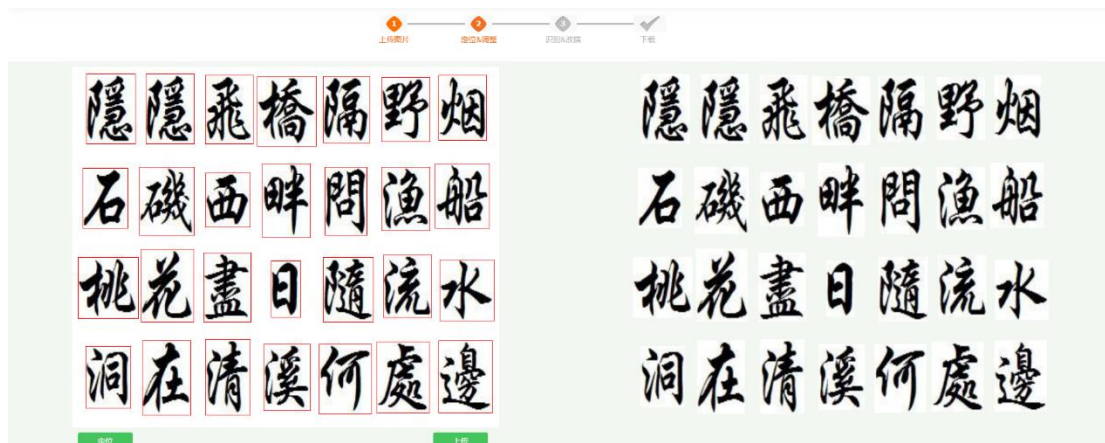


图 3 上传图片识别转换（识别）



图 4 上传图片识别转换（转换）

此外，我们还有相关新闻资讯、市集等功能模块，满足不同用户的阅读与研究需求。

### 三、核心技术

#### 1. 技术介绍

##### 1. SSD 定位模块

基于 SSD 模型，将一个图像中的文字切分出来，并且将他进行分类，打上属于他的标签，形成一个字的大量训练样本集。SSD，一种用于多个类别的快速单次对象检测器。SSD 模型的一个关键特点是使用多尺度卷积边界框输出附加到网络顶部的多个特征图。这种表示允许我们有效地模拟可能的框形状空间。提出了一种使用单个深层神经网络检测图像中对象的方法。方法名为 SSD，将边界框的

输出空间离散化为一组默认框，该默认框在每个特征图位置有不同的宽高比和尺寸。首先，我们可以利用 SSD 模型实现图像训练样本的采集，可以将一个个文字切分出来，并且打上标签，形成样本的训练库，接着我们就可以训练模型。SSD 还有一个作用就是对于用户上传的照片中的文字进行定位，将每个文字用一个个小框给框出来，每个文字扔进已经训练好的卷积网络，之后卷积网络进行卷积和池化的过程，最后得出相应的结果。

## 2. 卷积网络识别模块

采用卷积神经网络进行繁体字的识别转换。多层卷积与池化的模型，对其进行特殊的针对繁体字的模型构建，使构建出的模型可以通过识别转换繁体字的改错过程不断的迭代训练，准确率得到不断地提升，最终稳定在一个较好的准确率的程度。

## 3. 产品前后端交互模块

采用 HTML、CSS、JS 等完成的前端的页面的搭建，因为项目是深度学习实现的，采用了 python 的代码完成了繁体字的识别转换，所以在搭建前后台交互的过程中，选用了 python 的 flask 框架，完成前后台的交互过程。后端响应处理，调用已经训练好的识别模型，还有 SSD 定位模型，完成繁体字的定位和识别转换的整个过程。如果当登陆或者是查询用户的识别记录的时候，前端发起 URL 请求，后端 request.form.get 到相应的提交的表单的数据，之后执行 mysql 语句，进行数据的查询操作，或者是进行数据库的添加字段，之后将查询到的结果进行后端的处理或者是使用 return\_template 返回查询到的数据，前端使用 {{var}} (var 为后端返回的变量名) 进行后端数据返回到前端数据的获取。然后数据用于各种场景，可以用于 js，也可以用于前台数据的显示。

## 2. 创新点

系统采用的最关键的技术便是深度学习。由两部分构成，基于深度学习的 SSD 定位模型和基于深度学习的繁体字识别转换模型。其次还有 JS、CSS、HTML、python 的 flask 框架等相关技术完成。基于深度学习的 SSD 定位模型可以通过训练不断的提高模型的识别的准确率，在目前还是比较少的应用在一些系统中的。



基于 SSD 模型，将一个图像中的文字切分出来，并且将他进行分类，打上属于他的标签，形成一个字的大量训练样本集。第二个模型是基于深度学习的卷积神经网络的构建，通过构建出专门为繁体字识别的分类模型，经过训练后，会得到一种比较好的准确率，能够达到较准确的识别效果。

### 1. 基于深度学习的 SSD 智能定位模型

为什么说这是一个创新点那？经过知网论文的索引，很少有看到关于这个智能的定位模型的介绍，但是在系统中，这个智能定位模型的表现并不俗，能够达到一个比较好的定位效果，更加重要的是，这个模型和识别转换模型一样，都具有一定的智能性，它可以通过打标签，训练，变得更加的定位准确，目前的定位效果还是可观的，近段时间，我们正在人工努力的打一些标签，训练他们，想使 SSD 定位更加的出色，但是最终达到一个很好的情况是丝毫不慌的，因为根据我们的盈利模式，并不对主要识别功能收费，相比较 OCR 技术实现的软件，我们存在这样的优势。正是由于基于深度学习的定位模型，所以其的可学习性的优点也就展现出来了，伴随模型的不断进化，会使识别准确率慢慢的提高，最终收敛于一个比较好的状态。

### 2. 基于深度学习的卷积神经网络模型

我们使用 CRNN 神经网络对文字定位后得到的 proposal 进行训练，分类，最后得到我们的识别结果。其中，我们使用到的 CRNN 的网络架构又三部分组成，包括卷积层，循环层和转录层。1) 卷积层，从输入图像中提取特征序列；2) 循环层，预测每一帧的标签分布；3) 转录层，将每一帧的预测变为最终的标签序列。

在 CRNN 的底部，卷积层自动从每个输入图像中提取特征序列。在卷积网络之上，构建了一个循环网络，用于对卷积层输出的特征序列的每一帧进行预测。采用 CRNN 顶部的转录层将循环层的每帧预测转化为标签序列。再经过特征序列提取，网络训练就可以达到最终的识别。

### 3. 可以识别手写的繁体字

此款基于深度学习的繁体字识别的一大优势便是，能够对手写的繁体字进行识别转换，转换成对应的简体字，模型具有灵活的适应性，适用不同人手写的繁

体字，只要做一个可变范围的繁体字的训练，便可以使模型轻松的识别需要识别的繁体字，识别我们自己写的繁体字不是梦。

#### **4. 繁体字的识别率会不断提升**

深度学习的模型，使它可以不断的通过用户的纠正，提高识别的准确率。对于用户上传的照片，后台进行切分成不同的字，展现给用户，但是可能会存在部分的定位切分错误，引导用户进行更改，之后后台保存用户纠正后的一个个不同字的图片。这样的正确的定位信息保存到与上传图片的名字相同的 XML 文档中。供以后训练 SSD 定位模型用。用户跳转到识别的界面，运行识别程序，后台传回识别出来的结果，用户更改已经识别出的不对的文字，后台接到已经经过用户大体分辨的识别结果，给对应的字的图片，以识别出的简体字作为对应的文件名，存储在后台，经过简单的人工审核，供以后拿去训练现在的识别模型，使识别转换的准确率不断地提升。

#### **5. 识别图片中竖版字，并输出简体中文正确语序段落**

市面上大多数的识别文字是不具有识别竖版字的功能的，因为他们识别的广泛性，所以他们并不知道哪个是竖版字，所以他们统一按照横板字的方式进行排版的，所以在获取结果的时候，识别出的繁体字就会非常的混乱，不能够有一种正确的排版。而我们的系统是专门针对繁体字的，尤其是针对一个图片上的繁体字翻译成语序正确的一段话的，能够给使用者较高的体验。经过 js 的逻辑变化，判断行列坐标，转而实现正确的语序，方便用户下载阅读。

# 第一章 系统背景和可行性分析

## 1.1 项目背景分析

### 1.1.1 古籍阅读者的增加

随着高等教育的范围逐步扩大以及弘扬中国传统文化思想的传播，热爱历史古籍的人数也不断增加，对于刚接触古籍的初学者，古籍中复杂的繁体字无疑是阅读的一大障碍，我们的平台将为其提供繁体字与简体字对比版本，既能保留古籍的完整性，又能很好的帮助他们减少阅读障碍，降低入门门槛。注释功能也有利于阅读者的学习和理解。

### 1.1.2 两岸关系进入新纪元

2018年7月13日，中共中央总书记习近平在会见中国国民党前主席连战率领的台湾各界人士参访团时发表重要讲话，从四个方面揭示了近40年来两岸关系发展的主流和大势，强调我们对两岸关系未来充满信心，因为推动两岸关系和平发展、携手致力民族复兴，是符合民族整体利益、顺应时代潮流、造福两岸同胞、得到两岸同胞拥护的正确道路；因为不管经历多少风雨，两岸同胞在民族、文化认同和情感上从未分离；因为尽管数十年来两岸关系跌宕起伏，但总体趋势是向前发展的；因为两岸是密不可分、休戚与共的命运共同体。两岸同胞对更加美好生活的共同追求，对两岸关系走近走好的一致向往，是任何人都阻挡不了的。这些谈话指引两岸同胞看清大势，提振信心，团结前行。只要两岸同胞有信心、有恒心，心往一处想、劲往一处使，两岸关系就能不断克难前行，迎来光明前景。

近些年来，两岸在科技、文化、教育、新闻、出版等领域的交流与合作日益广泛和深入，造成海峡两岸信息交换的剧增。但是文字制度上几十年的差异，造成了两岸在常用语、专业术语、外国人地名翻译等各方面都存在着差异，另外繁体字和简体字数量都比较大，限于普通人的文字知识水平，难免造成交流上的困难，所以迫切需要一套汉字简繁转换系统。而市场上现存的一些小程序、APP等



等，大多只能将繁体字识别出来且正确率低，对于一些急需要将繁体字直接转换成简体字的情况却无法实现。

### 1.1.3 深度学习的发展

神经网络的发展总体可以概括为三大阶段：第一代神经网络（1958~1969）、第二代神经网络（1986~1998）以及第三代神经网络-DL（2006-至今）。深度学习-deeplearning 正是第三代神经网络。

2006 年，DL 元年。是年，Hinton 提出了深层网络训练中梯度消失问题的解决方案：无监督预训练对权值进行初始化+有监督训练微调。其主要思想是先通过自学习的方法学习到训练数据的结构（自动编码器），然后在该结构上进行有监督训练微调。但是由于没有特别有效的实验验证，该论文并没有引起重视。

2011 年，ReLU 激活函数被提出，该激活函数能够有效的抑制梯度消失问题。

2011 年，微软首次将 DL 应用在语音识别上，取得了重大突破。

2012 年，Hinton 课题组为了证明深度学习的潜力，首次参加 ImageNet 图像识别比赛，其通过构建的 CNN 网络 AlexNet 一举夺得冠军，且碾压第二名（SVM 方法）的分类性能。也正是由于该比赛，CNN 吸引到了众多研究者的注意。

AlexNet 的创新点：

（1）首次采用 ReLU 激活函数，极大增大收敛速度且从根本上解决了梯度消失问题；

（2）由于 ReLU 方法可以很好抑制梯度消失问题，AlexNet 抛弃了“预训练+微调”的方法，完全采用有监督训练。也正因为如此，DL 的主流学习方法也因此变为了纯粹的有监督学习；（3）扩展了 LeNet5 结构，添加 Dropout 层减小过拟合，LRN 层增强泛化能力/减小过拟合；（4）首次采用 GPU 对计算进行加速；

2013,2014,2015 年，通过 ImageNet 图像识别比赛，DL 的网络结构，训练方法，GPU 硬件的不断进步，促使其在其他领域也在不断的征服战场。

2015 年，Hinton，LeCun，Bengio 论证了局部极值问题对于 DL 的影响，结果

是 Loss 的局部极值问题对于深层网络来说影响可以忽略。该论断也消除了笼罩在神经网络上的局部极值问题的阴霾。具体原因是深层网络虽然局部极值非常多，但是通过 DL 的 BatchGradientDescent 优化方法很难陷进去，而且就算陷进去，其局部极小值点与全局极小值点也是非常接近，但是浅层网络却不然，其拥有较少的局部极小值点，但是却很容易陷进去，且这些局部极小值点与全局极小值点相差较大。论述原文其实没有证明，只是简单叙述，严密论证是猜的。

2015, DeepResidualNet 发明。分层预训练, ReLU 和 BatchNormalization 都是为了解决深度神经网络优化时的梯度消失或者爆炸问题。但是在对更深层的神经网络进行优化时, 又出现了新的 Degradation 问题, 即”通常来说, 如果在 VGG16 后面加上若干个单位映射, 网络的输出特性将和 VGG16 一样, 这说明更深次的网络其潜在的分类性能只可能 $\geq$ VGG16 的性能, 不可能变坏, 然而实际效果却是只是简单的加深 VGG16 的话, 分类性能会下降(不考虑模型过拟合问题)“Residual 网络认为这说明 DL 网络在学习单位映射方面有困难, 因此设计了一个对于单位映射(或接近单位映射)有较强学习能力的 DL 网络, 极大的增强了 DL 网络的表达能力。此方法能够轻松的训练高达 150 层的网络。

图像识别作为深度学习中的重要一块, 一直都是块热门的研究领域。每年都有许多研究人员投入大量时间与精力去研究, 推出新的技术。图像识别现在还处于高速发展的时候, 相信未来会有更大的市场空间。

#### 1.1.4 古籍数据化发展空间巨大

数字图书馆和大型电子文献资料库的开发和建设, 近年来业已成为学术界和相关技术部门的一个热门话题。在这个话题中, 古籍的数字化也得到了高度重视。并且, 伴随着传统文化的传播, 数字化的古籍在学术研究中的利用率也在不断提高。而古籍数字化的价值, 也正是在这样的使用中一点一点体现出来。

但目前古籍数据库的发展还有很多的不足, 比如文件格式过多, 不方便用户阅读, 技术上的缺陷等等。

而中国古籍的数量巨大, 范围宽广, 形式丰富, 都表明了古籍数据化的发展空间巨大。

## 1.2 可行性分析

### 1.2.1 市场可行性分析

#### 1、繁体字转换的重要性

中国推行简体字作为官方语言已经许多年了，我们对简体字的认识已经根深蒂固了，以至于我们对繁体字已经有些陌生了。对于一些简单的繁体字，我们还能通过看字形，看上下文意来猜测出是什么字，但有些复杂的繁体字，我们已经很难辨认出来了。在两岸关系不断发展，文化交流不断加强的现在，我们有更大的概率会与繁体字打交道。一个优秀的繁体字转换系统能帮助我们更加高效的阅读繁体字。同时，对于古籍研究者而言，他们每天需要面对无数复杂的繁体字，如果是经验丰富的研究者，那可能还能比较流畅的阅读，但对于初入这一行的研究者，阅读这些古籍似乎有不小的障碍，一个优秀的繁体字转化系统能有效降低古籍研究者的入门门槛，让更多对中国传统文化有兴趣，对研究古籍有兴趣的人，不会因为过高的入门门槛而放弃研究古籍，也能更好的传承我们的传统文化。

#### 2、现存的繁体字转换系统的缺陷

目前网络上已经有不少的繁体字转换系统可以使用了，但这些系统都有一个致命的缺陷。繁体字转换系统最重要的几点是什么？是准确率以及方便性。而目前网络上的繁体字转换系统在方便性的设计上都显得不够充分，目前的繁体字转换系统都支持输入文字，然后在进行转换。但是很多情况下，用户想要进行转换的内容都是存在图片这一载体上的繁体字，面对这种情况，目前网络上的繁体字转换系统都没有给出很好的解决手段，它们均不支持提取图片中的繁体字进行识别转换，而这一点就是我们系统的优势所在。我们通过研究 SSD 算法，做到了智能识别并提取图片中的繁体字，然后再进行转换。这样就能很好的满足用户在方便性上的需求，对于古籍，只需要将书籍通过扫图，上传到系统中，就能得到转换后的文本。更甚者，在日常生活中碰到了不认识的繁体字，只需要用手机拍摄一张照片，然后上传到系统，就能很快得到转换后的简体字。



### 3、古籍数据库的需求

目前古籍数据库也有一些，但是很多是未经标点的原始版本，且识别出来的错误较多、质量低，不方便大家的学习与研究。

因为原版古籍的价格高、获取难，一些专家学者进行学术研究时很不方便。而通过技术手段、数字化手段将古籍转移到其他载体既有利于保护古籍，也有利于学者及时的翻阅查找资料。在很多学者看来，互联网改变了做学问的方式，古籍数字化让学术研究发生了革命性变化。

#### 1.2.2 技术可行性

《SSD: Single Shot MultiBox Detector》我们从这篇论文中能够看到用 SSD 来定位的物体的准确性还是相当可观的，通过训练的模型，经过不断的测试，能够发现 SSD 在实现定位方面具有良好的定位效果。进而得到了学者们的广泛研究。所以在学术界 SSD 定位得到了一定的认可，在技术可行性上，具有很强的可行性。

再说基于深度学习的卷积神经网络识别转换繁体字，自打阿尔法围棋（AlphaGo）第一次击败人类职业围棋选手、第一次战胜围棋世界冠军。这个人工智能机器人就得到了广泛的关注，更多的科研人员关注了其实现的核心技术-卷积神经网络。我们可以搭建各种不一样的神经网络，用于识别不同的物体，并进行分类。我们的卷积神经网络模型，是根据繁体字特点，专门搭建的一个卷积神经网络模型。再经过一定数据集的训练，能达到一种比较好的识别准确率。由于现在系统借助简单的前端页面已经在运行了，也收获了 90%多的识别准确率，所以在实现方面具有可行性。

虽然不算很成熟的 flask 后台框架，但是在一些前后端的交互的过程中，还是展示出了不错的前后台交互，可以正确的响应前台的 URL，还有后台数据往前台的传递。这个框架也是被广大使用 Python' 框架的程序员接受的。存在很好的使用性的。所以在技术上是存在很大的可行性的。其实我们现在系统已经能够利用 flask 框架，完成正常的前后台的交互的。

还有 js、css、html 等就更不用说了，已经存在与业界很久了，每一位开发 web 的程序员必备的一项技术，可以说技术上实现是相当没有问题的。

注：我们已经用上述技术，完成了整个系统，只差前端的美化，以及完善。



## 第二章 产品介绍及项目技术介绍

### 2.1 产品简介

随着深度学习的推广，深度学习被应用到了很多的地方，我们的项目是基于深度学习的古籍整理与数字化综合服务平台，包含了网站、app 以及引流的小程序。我们的平台希望通过核心技术（繁体识别及转换）与相关出版社、大学、图书馆甚至是博物馆合作，推出覆盖面广、专业权威的古籍整理数据库，以满足不同用户的阅读与研究需求。

同时，我们将提供专注于古籍传统文化的论坛社区，用户可以在论坛中发表自己的随笔、研究心得，也可以自由发问，互助解决问题。根据自身需求结成小组，进行内部讨论和分享。我们希望论坛能够将学界与用户结合，为古籍整理和学术研究的推进建立良性的生态系统。

此外，我们还有学界相关的新闻资讯发布，与相关出版商合作的实体书商城，以及相关的小工具（包括古文字典、纪年转换以及上传图片识别转换）。

### 2.2 功能介绍

#### 2.2.1 网站 聚籍网

##### 1.在线古籍阅读

在线古籍阅读根据古籍的内容分为“经史子集”四大部类，每个大部类底下还有若干子部类。用户可以直接输入书名、作者、出版社、出版时间、古籍版本、数目来源、提要、子目等关键字来检索相关古籍。也可以选择相关领域的进行查看。阅读网站内的古籍需要成为网站会员或直接购买相关类别，我们对非会员用户提供 3 页的试读页数。





图 5“聚籍”网在线阅读古籍界面

阅读界面中将逐页显示古籍的图片，客户可再根据自身需求选择显示繁体字或简体字文本以便对比阅读，将显著提高用户的使用效率，大大拓展用户的信息视角，为研究工作带来便利。个人用户和机构用户注册过个人账户的用户可以对阅读的古籍进行批注，同时个人中心还有收藏、检索记录、阅读历史等辅助用户进行阅读的个性化功能。

## 2. 论坛社区

在论坛社区中用户可以发表文章，自由发问和建组讨论，并根据用户的关注进行个性化推送。用户可以在论坛社区上分享随笔、研究心得等个性文章，随时在线提问，互相解决问题；也可以根据个性化的需求结成组群，进行内部讨论和分享。



图 6“聚籍”网论坛社区界面

### 3.新闻资讯

在新闻资讯的页面中我们将提供古籍相关的新闻资讯供用户阅读了解，其中包括古籍出版整理的新闻及相关会议、学术讨论的咨询，让用户可以更加快速及时地获取古籍传统文化的相关资讯。

### 4.实体书商城

实体书商城是平台与各大古籍出版商合作的网上古籍购买平台。用户可以在上面购买正版专业古籍以及古籍相关书籍，会员用户在购买时还可以获得一定折扣。



图 7“聚籍”网实体书商城界面

### 5.工具

平台提供的工具包括在线上传图片识别转换、字典查询和纪年转换。

上转图片识别转换利用了我们的核心技术，用户可自主上传古籍图书的图片，得到准确率高的繁体识别文本，工具可以识别一些模糊不清的图片以及手写体，以帮助用户更好的阅读理解古籍。

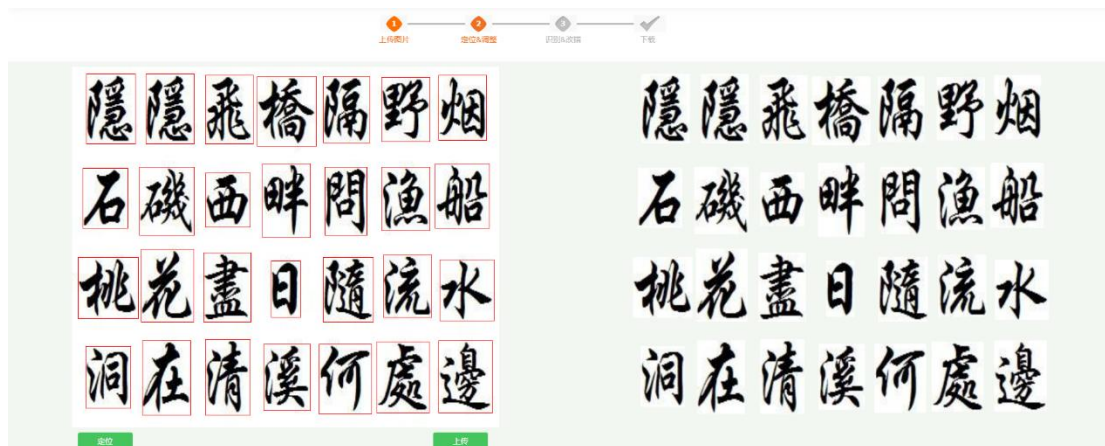


图 8 上传图片识别转换（识别）



图 9 上传图片识别转换（转换）

字典查询中用户可以输入想要查询的繁（简）字进行查询，字典将显示其读音、释义及出处。

纪念转换则可以通过输入数字年份得到对应的干支、时期、政权等相关信息，以便更好的理解对应年代的古籍信息。

### 2.2.2 APP 聚籍

页面设置以简洁丰富、分类明确为主，采用柔和的色系，降低用户阅读时被其他影响的概率。

首页的下方菜单栏分为五大板块：首页、发现、市集、我的、繁体识别。



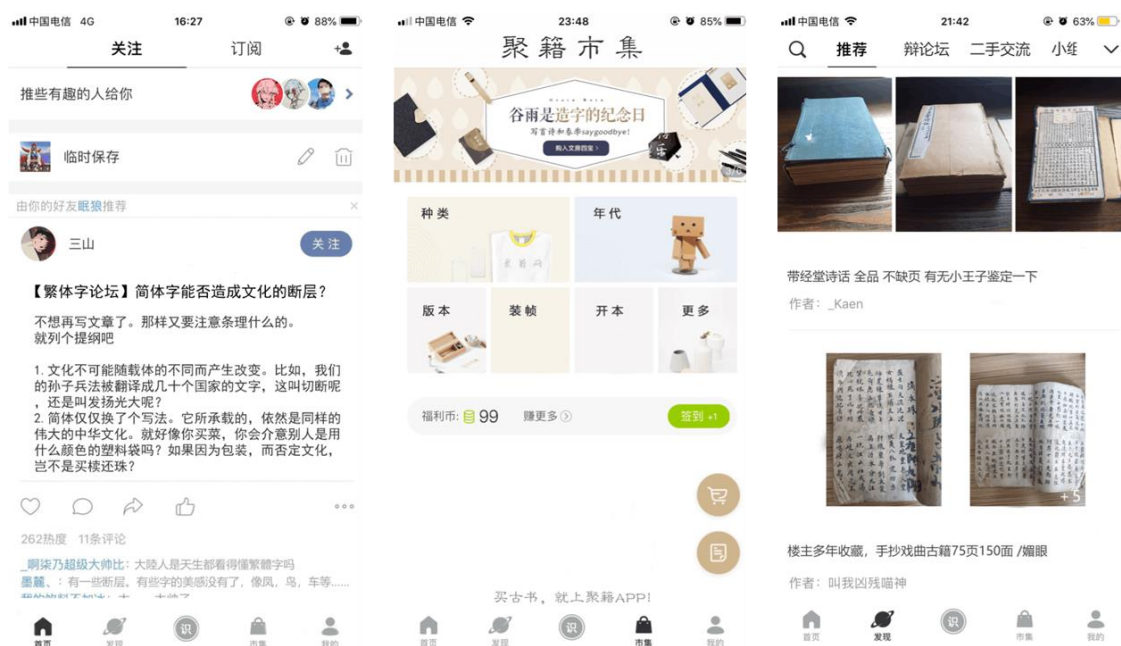


图 10“聚籍”app 界面

## 1. 前端

(1) 登陆注册功能：用于客户登陆使用。

(2) 首页

顶端设置一个搜索栏，方便用户快捷的搜索定向的信息。在搜索栏下方预留较大幅的位置设置当日的头条，头条推送最有价值的信息，让用户一目了然。头条下方设置数据库分栏板块，方便用户找寻书目。

(3) 发现

发现的界面主要设置为两个板块，一个是分享讨论，一个是众包。

其中分享讨论的主要功能是可以提出一些自己关于繁体字、古籍、古文化的问题和看法，互相勉励交流信息，共同学习进步。用户也可以关注圈子的人为好友。

众包的主要功能是有相关专业的用户可以在软件上接我们下方的任务，如给识别出来的繁体加标点等，平台会给予用户一定的积分、优惠或货币。

在分类板块下是动态浏览，包括用户关注的好友、专家动态。且用户本人也可以发布动态。

(4) 市集

商城类似其他互联网电商平台，设置搜索栏、商品分类、购物车等基本功能。旨在联结产品与消费者，并且商城的商品推送可根据用户的相关资料进行个性化

推送。帮助用户快捷方便的选购到所需的各类产品。也可以向好友分享商品。主要贩卖实体书籍、文具、古风周边等。

## （5）我的

“我的”相当于个人中心，可以设置头像、昵称等基本模式的设置。其中主打五大板块——消息、达人认证、打赏记录、在线客服。

消息一栏可以查看好友发来的消息。，以及回答收到的评论、喜欢等。

达人认证主要是回答的问题，或任务的完成达到一定程度会获得平台授予的荣誉。

打赏记录主要功能是回答或分享获得的其他用户的打赏。

在线客服主要是用户及时反馈问题的渠道。

## （6）繁体识别

这一功能与小程序的功能一致，用户点击下方拍照按钮，就可拍照或提交照片识别图中的繁体字，方便简洁。



图 11"聚籍"app 繁体识别界面

## 2. 后台

（1）登陆功能：后台管理人员登陆入口。

（2）商家执行管理功能：对品种进行分类或者添加、删除品种分类以及对

宣传方面进行管理。

(3) 品牌及服务管理功能：平台对每一个入驻专家、学者进行一定的资质审核及认证。

### 2.2.3 微信小程序 繁体识别

页面设置以简洁明了为基准，功能单一。为了方便用户使用，小程序只为大家提供了拍照识别功能。

用户可以通过搜索关键词直接进入小程序。

点击页面下方的“拍照识别”，上传繁体字照片进行识别，得到结果。

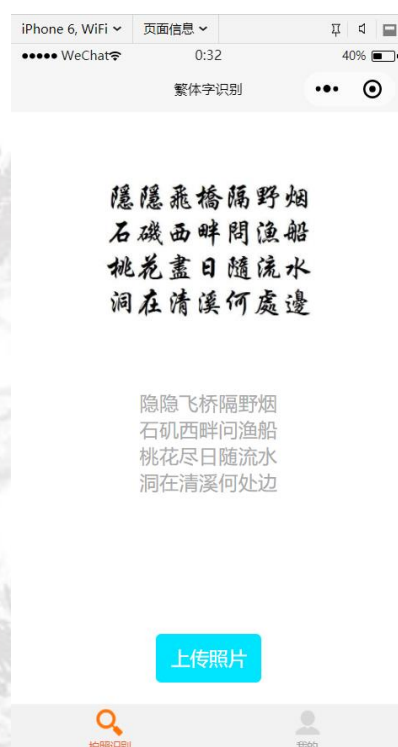


图 12 繁体识别微信小程序界面

## 2.3 核心技术介绍

本系统主要是基于深度学习的目标检测与识别方法来识别繁体字的，即基于回归的目标检测与识别算法 SSD(Single Shot MultiBox Detector)。SSD 算法识别效果相对于传统文字识别方法要好，原因在于基于深度学习的方法克服了传统文字识别的一些不可逾越的缺陷，例如传统文字的考虑文字的纹理特征或者颜色、倾



斜度等，而采用深度学习的方法则会弱化或者不考虑这些问题。目前该方法处于蓬勃的发展当中，效果也越来越突出。

其中 SSD 算法是本系统的核心。SSD 算法是一种直接预测 bounding box 的坐标和类别的 object detection 算法，没有生成 proposal 的过程。针对不同大小的物体检测，传统的做法是将图像转换成不同的大小，然后分别处理，最后将结果综合起来，而本文的 ssd 利用不同卷积层的 feature map 进行综合也能达到同样的效果。算法的主网络结构是 VGG16，将两个全连接层改成卷积层再增加 4 个卷积层构造网络结构。对其中 5 个不同的卷积层的输出分别用两个 3\*3 的卷积核进行卷积，一个输出分类用的 confidence，每个 default box 生成 21 个 confidence（这是针对 VOC 数据集包含 20 个 object 类别而言的）；一个输出回归用的 localization，每个 default box 生成 4 个坐标值 (x, y, w, h)。另外这 5 个卷积层还经过 priorBox 层生成 default box（生成的是坐标）。上面所述的 5 个卷积层中每一层的 default box 的数量是给定的。最后将前面三个计算结果分别合并然后传递给 loss 层。

### 2.3.1 系统框架及步骤

在繁体字识别系统中，文字识别涉及到图像预处理、文字定位、文字分割和文字识别几个主要步骤，其中图像预处理包含图像剪裁、图像去噪声、图像二值化等操作；文字定位主要有基于纹理文字定位方法、基于连通域文字定位方法以及二者混合方式；文字分割主要有基于阈值的算法、基于分组算法等；文字识别主要基于特征提取、基于词典、基于神经网络等方法。下图是整个系统框架。

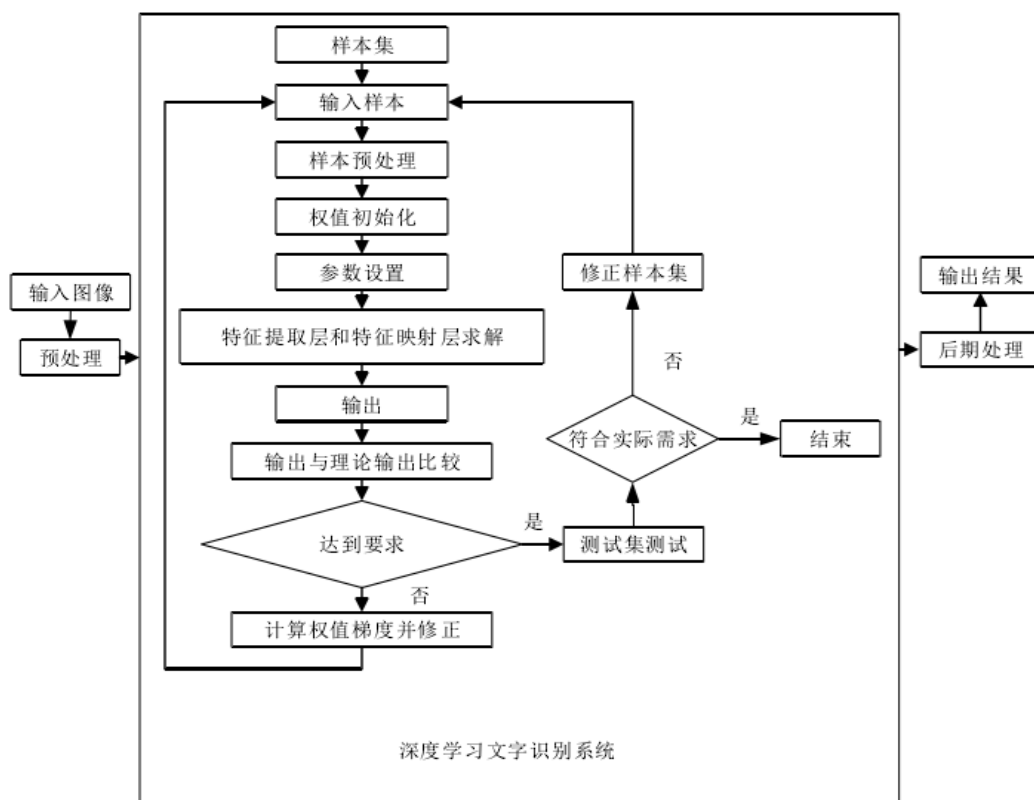


图 13 系统框架图

分为以下几个步骤：

1、搜集样本集构建一个样本库，当然也需要事先搜集一些测试集便于后期测试使用，搜集样本集尽量越多越好，这样训练的时候便于系统学习到更多的信息，获得更好的学习的效果；

2、输入样本到神经网络模型，有时候为了收到更好效果通常需要对样本做一些处理，例如图像的尺寸大小修正、图像去噪等；

3、对神经网络中的一些参数和权值做一些初始化，参数设置一般是根据实际经验确定，对于权值参数初始化系统通常会自动调整；

4、系统根据输入数据进行特征提取和特征映射，期间权重参数会自动确定，每一次输入权重参数会自动确定；

5、对于输出结果需要与预期结果作比较，如果误差没有达到指定要求还需要重新反馈到神经网络，进行重新训练，直到符合要求；

6、达到指定要求后，一般还需要利用测试集进行测试，如果输出符合期望值则这个系统基本可以使用到实际应用中，如果不符合要求，还需要对神经网络

进行重新训练，直到再次测试符合要求即可；

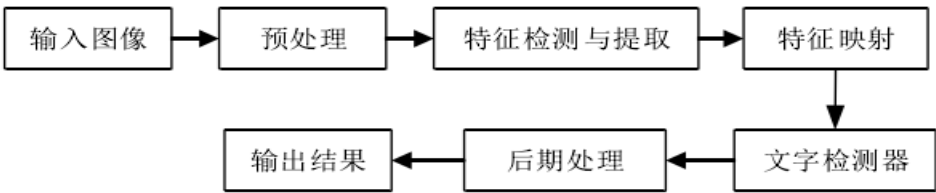


图 14 图像识别流程图

### 2.3.2 图像预处理

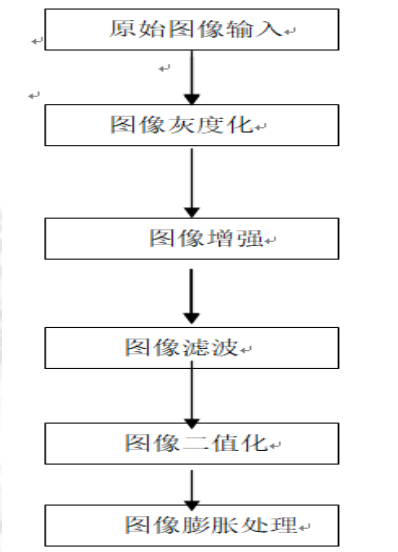


图 15 图像预处理流程图

#### 1. 图像灰度化

灰度图像上每个像素的颜色值又称为灰度，指黑白图像中点的颜色深度，范围一般从 0 到 255，白色为 255，黑色为 0。所谓灰度值是指色彩的浓淡程度，灰度直方图是指一幅数字图像中，对应每一个灰度值统计出具有该灰度值的像素数。

灰度就是没有色彩，RGB 色彩分量全部相等。如果是一个二值灰度图像，它的像素值只能为 0 或 1，我们说它的灰度级为 2。用个例子来说明吧：一个 256 级灰度的图像，如果 RGB 三个量相同时，如：RGB(100,100,100)就代表灰度为 100，RGB(50,50,50)代表灰度为 50。

现在大部分的彩色图像都是采用 RGB 颜色模式，处理图像的时候，要分别对 RGB 三种分量进行处理，实际上 RGB 并不能反映图像的形态特征，只是从光



学的原理上进行颜色的调配。

图像灰度化处理可以作为图像处理的预处理步骤，为之后的图像分割、图像识别和图像分析等操作做准备。

### 2.图像二值化

图像二值化（ Image Binarization）就是将图像上的像素点的灰度值设置为 0 或 255，也就是将整个图像呈现出明显的黑白效果的过程。

下图即是二值化处理过的图片。

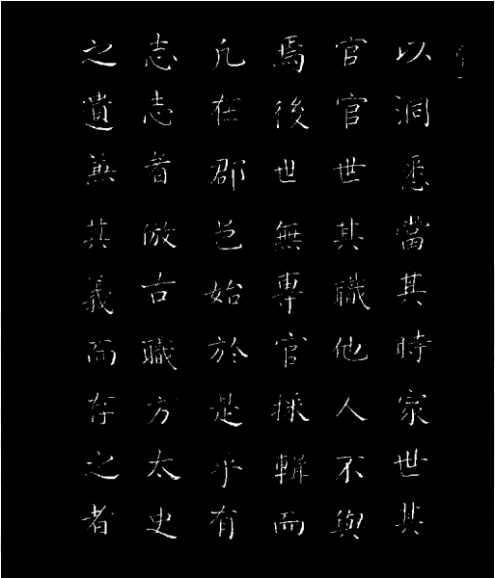


图 16 二值化处理过的图片

### 3.膨胀处理

形态学变换膨胀采用向量加法对两个集合进行合并。膨胀的算符是  $\oplus$ ，图像  $X$  用结构元素  $B$  来膨胀写作  $X \oplus B$ ，是所有可能向量加之和的集合，向量加法的两个操作数分别来自于  $X$  和  $B$ ，并且取到任意可能的组合。膨胀用来填补物体中小的空洞和狭窄的缝隙。它使物体的尺寸增大。下图是经过膨胀处理。

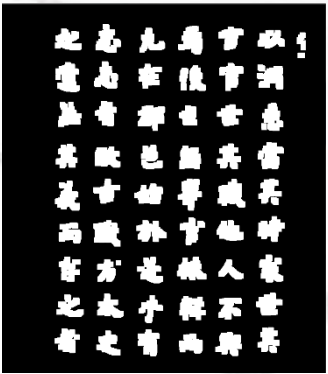


图 17 多次膨胀处理

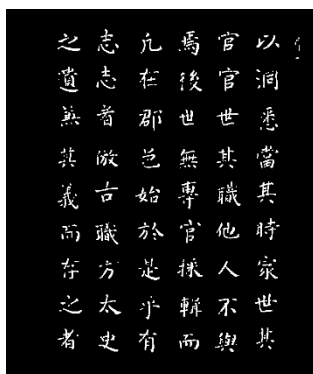


图 18 一次膨胀处理

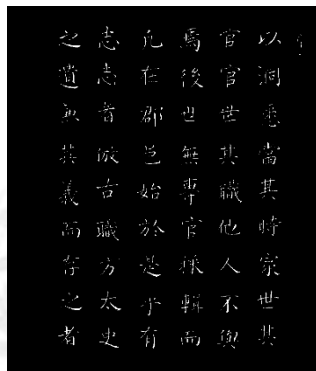


图 19 原图

### 2.3.3 文字定位

图片经过预处理后，满足 SSD 的需求，接下来便可通过 SSD 算法实现文字定位。文字定位是整个系统中很重要的一环，定位的准不准确直接影响识别的准确与否，从而也关系着整个系统的识别正确率。

#### 1. 网络架构

SSD 网络主体设计的思想是特征分层提取，并依次进行边框回归和分类。因为不同层次的特征图能代表不同层次的语义信息，低层次的特征图能代表低层语义信息(含有更多的细节)，能提高语义分割质量，适合小尺度目标的学习。高层次的特征图能代表高层语义信息，能光滑分割结果，适合对大尺度的目标进行深入学习。所以作者提出的 SSD 的网络理论上能适合不同尺度的目标检测。

所以 SSD 网络中分为了 6 个 stage，每个 stage 能学习到一个特征图，然后进行边框回归和分类。SSD 网络以 VGG16 的前 5 层卷积网络作为第 1 个 stage，然

后将 VGG16 中的 fc6 和 fc7 两个全连接层转化为两个卷积层 Conv6 和 Conv7 作为网络的第 2、第 3 个 stage。接着在此基础上，SSD 网络继续增加了 Conv8、Conv9、Conv10 和 Conv11 四层网络，用来提取更高层次的语义信息。如下图所示就是 SSD 的网络结构。在每个 stage 操作中，网络包含了多个卷积层操作，每个卷积层操作基本上都是小卷积。

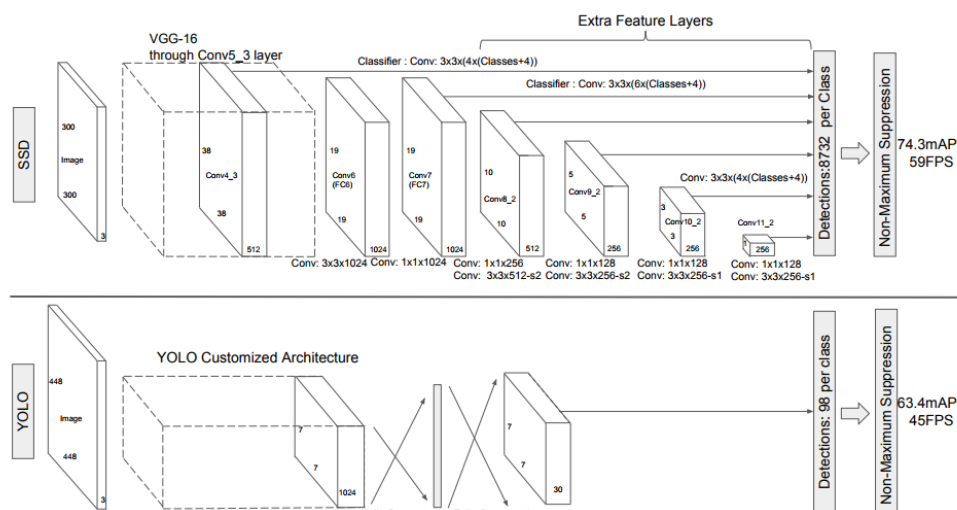


图 20 SSD 的网络结构

骨干网络：SSD 前面的骨干网络选用的 VGG16 的基础网络结构，如上图所示，虚线框内的是 VGG16 的前 5 层网络。然后后面的 Conv6 和 Conv7 是将 VGG16 的后两层全连接层网络(fc6, fc7)转换而来。

另外：在此基础上，SSD 网络继续增加了 Conv8 和 Conv9、Conv10 和 Conv11 四层网络。图中所示，立方体的长高表示特征图的大小，厚度表示是 channel。

## 2.设计要点

### Default Boxes 生成

在目标检测网络设计中，Default Boxes 生成都是重中之重，也许直接决定了网络能针对的任务以及检测的性能。在 SSD 中，作者充分的吸取了 Faster R-CNN 中的 Anchors 机制，在每个 Stage 中根据 Feature Map 的大小，按照固定的 Scale 和 Radio 生成 Default Boxes。这里为了方便阐述，选取了 Conv9 的 Feature Map，输出大小为 5x5。SSD 网络中我们设置 Conv9 的每个点默认生成 6 个 box，如下图所示。因此在这一层上，共生成了  $5 \times 5 \times 6 = 150$  个 boxes。



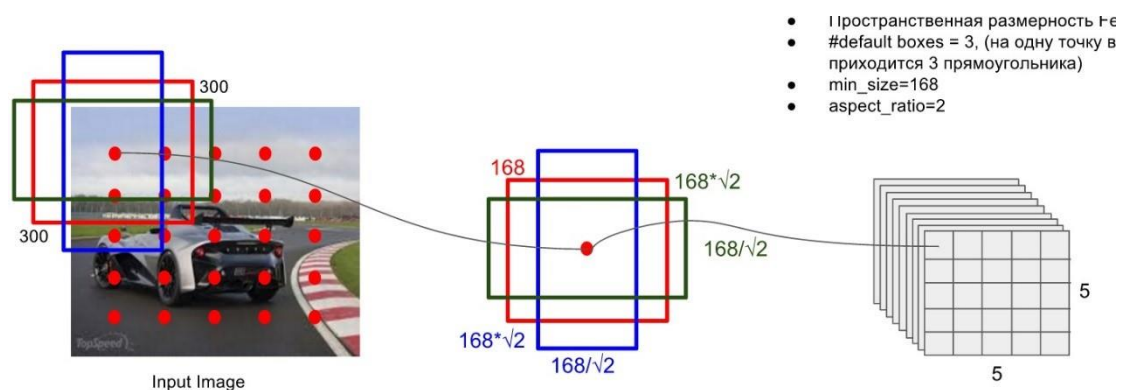


图 21 SSD Feature Map

### 1) 特征向量生成

在每张特征图上得到许多 Default Box 后还需要生成相应的特征向量，用来进行边框回归和分类。对于边框回归，只需要 4 维向量即可，分别代表边框缩放尺度(坐标轴两个方向)和平移向量(坐标轴两个方向)。对于分类，SSD 网络采取为每个类别进行打分的策略，也就是说对于每个 Default Box，SSD 网络会计算出相应的每个类别的分数。假设数据集类别数为  $c$ ，加上背景，那么总的类别数就是  $c+1$  类。SSD 网络采用了  $c+1$  维向量来分别代表该 Default Box 对于每个类别所得到的分数。这里，假设是 VOC 数据集，那么每个 Default Box 将生成一个  $20 + 1 + 4 = 25$  维的特征向量。同样，以 Conv9 输出特征图  $5 \times 5$  为例。

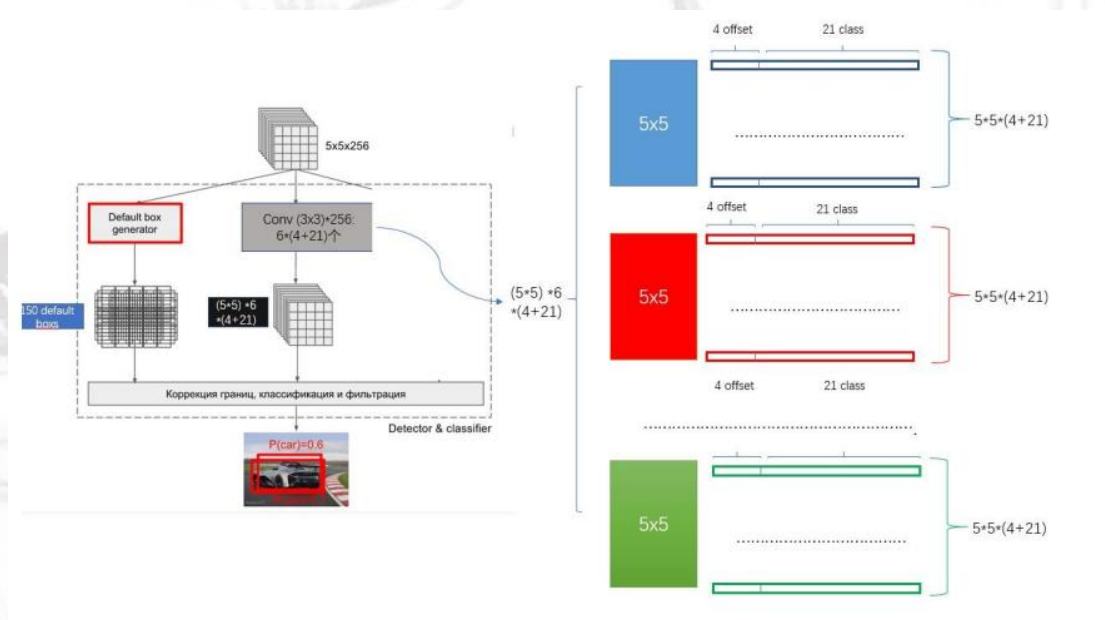


图 22 Conv9 输出特征图  $5 \times 5$

### 2) 新增卷积网络

SSD 网络在 VGG 基础上新增加了几个卷积网络(如网络结构中所述)。SSD 网

络总共增加了 Conv8、Conv9、Conv10 和 Conv11 四个卷积网络层。新增的这些网络都是通过一些小的卷积核操作。引用论文所说，这些小的卷积核操作是 SSD 网络性能优秀的核心。下面本报告将简单的阐述一下作者对于卷积网络的简单配置。

#### 卷积核配置

- 假设 Feature Map 通道数为 P，SSD 网络中每个 Stage 的卷积核大小统一为  $3 \times 3 \times P$ 。其中 padding 和 stride 都为 1。保证卷积后的 Feature Map 和卷积前是一样大小。

#### 卷积滤波器

- 每个 Feature Map 上  $m \times n$  个大小的特征点对应 K 个 Default Boxes，假设类别数+背景=c，最终通过卷积滤波器得到  $c+4$  维特征向量。那么一个 Feature Map 上的每个点就需要使用  $k \times (c+4)$  个这样的滤波器。

下图是定位效果图



图 23 定位效果图

之志凡焉官以  
遺志在後官洞  
兼者郡世世悉  
其倣邑無其當  
義古始專職其  
而職於官他時  
存方是採人冢  
之太乎輯不世  
者史有而與其

图 24 原图

#### 2.3.4 文字分割

文字分割就是将文字区域像素与背景像素分离开来，以便于后续对文字区域处理，文字分割是文字识别前重要的一个步骤，所以文字分割的结果会影响到文字识别的效果。通常文字分割的算法基于灰度图像的，由于场景文字图像背景的复杂性，文字颜色有时候也会考虑进去。一般关于文字分割的方法有如下 3 种：基于阈值方法、基于分组方法和基于统计模型方法。我们使用的是基于分组方中的基于学习方法。

对于基于学习方法，我们借鉴了最近的流行的神经网络模型，通过构建的神经网络，例如卷积神经网络，学习大量的样本，完成字符分割的任务。基于学习的方法不同于传统方法，其利用了类似于人类神经元网络结构，在学习过程中可以自动的优化图像分割遇到的问题，只要通过大量的样本训练，最后可以得到好的分割效果。

下图是基于神经网络图像分割大体步骤。



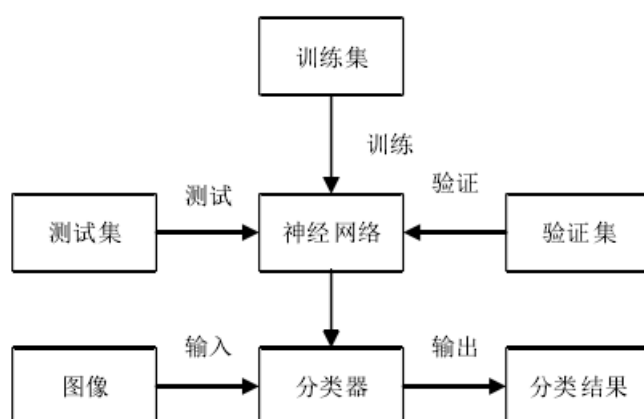


图 25 基于神经网络图像分割大体步骤

### 2.3.5 文字识别

在完成文字定位以后，往往还不能说完成了文字检测任务，虽然文字定位是关键的一个步骤。但是定位以后的文字区域仍然以像素的方式呈现。可以说在文字定位阶段是以文字区域为主要研究目的，而在文字提取技术中，是在文字定位基础之上，进一步得到真正的文字信息。

我们使用 CRNN 神经网络对文字定位后得到的 `proposal` 进行训练，分类，最后得到我们的识别结果。

#### 1. CRNN 的网络架构

其中，我们使用到的 CRNN 的网络架构又三部分组成，包括卷积层，循环层和转录层。

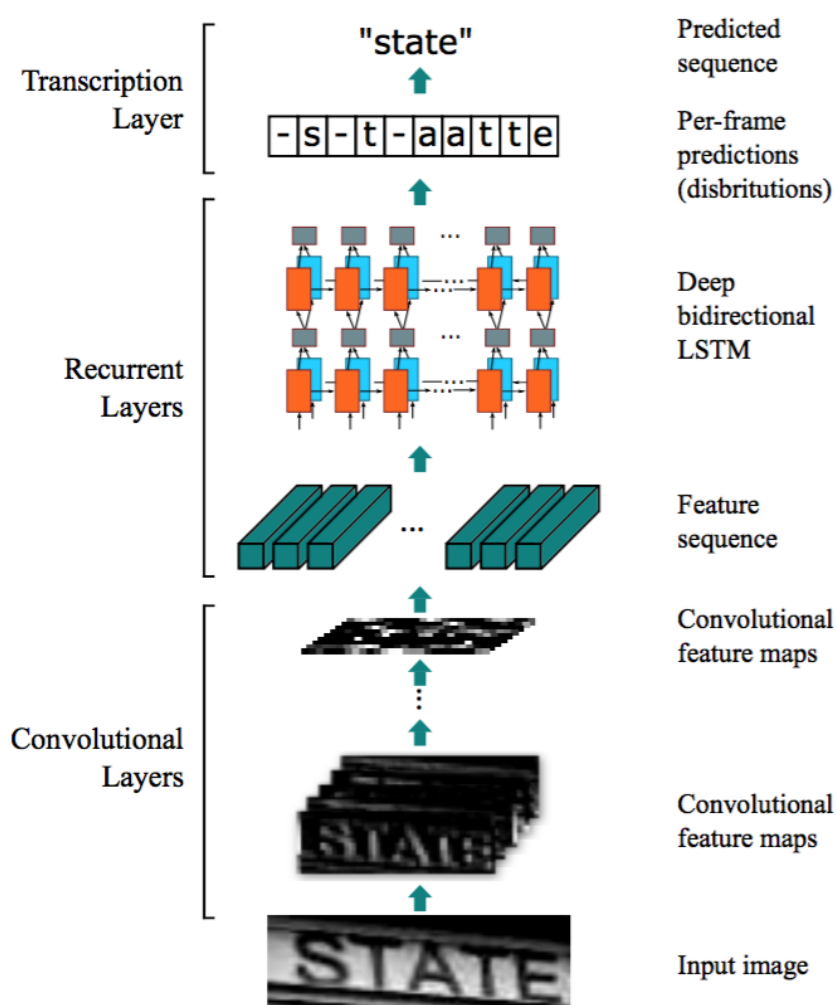


图 26 CRNN 的网络架构

1) 卷积层，从输入图像中提取特征序列；2) 循环层，预测每一帧的标签分布；3) 转录层，将每一帧的预测变为最终的标签序列。

在 CRNN 的底部，卷积层自动从每个输入图像中提取特征序列。在卷积网络之上，构建了一个循环网络，用于对卷积层输出的特征序列的每一帧进行预测。采用 CRNN 顶部的转录层将循环层的每帧预测转化为标签序列。

## 2、特征序列提取

在 CRNN 模型中，通过采用标准 CNN 模型（去除全连接层）中的卷积层和最大池化层来构造卷积层的组件。这样的组件用于从输入图像中提取序列特征表示。在进入网络之前，所有的图像需要缩放到相同的高度。然后从卷积层组件产生的特征图中提取特征向量序列，这些特征向量序列作为循环层的输入。具体地，特征序列的每一个特征向量在特征图上按列从左到右生成。这意味着第  $i$  个特征向

量是所有特征图第  $i$  列的连接。在我们的设置中每列的宽度固定为单个像素。

由于卷积层，最大池化层和元素激活函数在局部区域上执行，因此它们是平移不变的。因此，特征图的每列对应于原始图像的一个矩形区域（称为感受野），并且这些矩形区域与特征图上从左到右的相应列具有相同的顺序。如图所示，特征序列中的每个向量关联一个感受野，并且可以被认为是该区域的图像描述符。

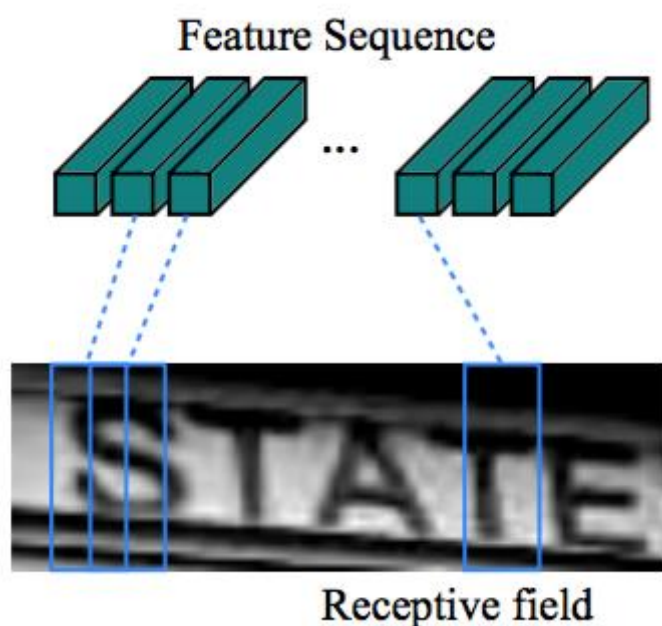


图 27 感受野

### 3.网络训练

目标函数直接从图像和它的真实标签序列计算代价值。因此，网络可以在成对的图像和序列上进行端对端训练，去除了在训练图像中手动标记所有单独组件的过程。网络使用随机梯度下降（SGD）进行训练。梯度由反向传播算法计算。特别地，在转录层中，误差使用前向算法进行反向传播。在循环层中，应用随时间反向传播（BPTT）来计算误差。我们对 1000 张图片进行了标注，训练。最终有 96,78% 的正确率。





图 28 系统识别出的图片

## 第三章 公司分析与发展战略

### 3.1 公司基本情况

名称：基于深度学习的古籍数字化平台

愿景：成为智能古文识别开创领导者

使命：为大众提供知识服务，传播优秀传统文化

宗旨：便捷、高效、实惠、多功能

概述：本公司发明提供了一种基于图像识别技术的古文字智能识别系统,其中所述的方法,包括如下步骤:建立用于图像识别技术的古文字智能识别平台的古文字图像及资料数据库、客户端和服务端;通过服务端接收客户端的图片并进行识别;将识别的信息与古文字图像及资料数据库中信息进行对比并将对比信息反馈到客户端。公司将此技术作为核心技术亮点开发了古籍数字化及古籍文化交流平台,用户可以使用浏览器、手机 APP 随时随地的访问、使用平台阅读古籍，尽情的享受科技带来的乐趣和便捷，为古文研究工作提供便利。

### 3.2 组织结构

公司初期拟采取直线制的组织形式，易于明确各自的之策；决策迅速到位；

同时要求高管具有国人的管理手段和专业知识。公司拟设董事会，总经理以及技术部，行政部，市场部和财务部。如下图：



图 29 公司组织形式

### 3.3 部门职能及构成

表 1 部门职能及其构成表

主要构成	主要职能
董事会	1.对内决定总经理的人选。 2.负责制定公司总体发展战略、制作。 3.保存董事会的议事录、备置公司章程和各种簿册并及时向股东大会报告资本的盈亏情况等。 4.对外代表公司。
总经理	1.负责公司的日常经营事务的经营管理。 2.对董事会负责，提出任免各部门经理的人选，协调各部门之间的关系。
技术部	1.负责产品部门的一系列方案运用技术来实现整体的功能。 2.控制从技术开发到最后形成的整个研发的过程，处理与产品有关的技术问题。 3.负责产品在运营中的技术问题。
行政部	1.负责起草公司内部管理的各项规章制度、文件。 2.负责公司的网站建设及维护。 3.负责公司的日常工作和协调各部门的工作。 4.负责公司人员的招聘选拔培训上岗工作。 5.负责公司考核标准制定，员工绩效考核，激励工作。 6.负责公司劳动合同管理、劳务管理、人事档案工作等。
市场部	1.制定产品管理制度与方案、规划产品发展方向以及长期竞争策略。 2.发展市场调研和分析，对产品提出整改计划。 3.管理产品的全过程，监管产品发展情况，处理产品服务问题，及时提出改进意见或降低成本，以提升产品价值准确把握市场动向。 4.负责包括市场分析，广告，公共关系、小手客户服务等一系列事务。
财务部	1.负责公司资金的筹集、使用和分配，日常会计工作与税收管理，每个会计年度末向CEO 汇报本年财务情况并规划



	<p>下面财务工作。</p> <p>2.对岗位设置、人员配备、核算组织程序等提出相关方案。</p> <p>3.结合本公司的具体情况建立规范的财务模式,指导建立健全相关财务的核算制度,同时负责对公司内部财务管理制度的执行情况进行检查和考核,提高经济效益。</p>
--	--

### 3.4.人力资源管理

在企业的核心竞争资源中,人力资源是个重要元素,是企业核心竞争力的基础动力之一。如何有效的配置人力资源最大程度的发挥人力资源优势,成为企业倾情关注的课题。21 世纪理性营销时代的到来,使个人英雄无法再在营销舞台上独唱主角,依靠个人力量叱咤风云、劲舞弄潮的日子一去不返。团队,这个营销时尚名词,开始被越来越多的企业探讨钻研。团队管理,正被纳入企业人力资源管理的治新领域。在本计划当中人力资源的管控还包括创新机制和激励机制。

#### 3.4.1 创新机制

##### 1.建立激励创新的工作机制

一是建立激励创新的工作机构。企业应设立专职部门。配备专职人员。负责创新工作的评价与考核。激励创新制度的制定、创新工作培训等。二是做好创新产品的保护工作。做好创新成果的保护工作,防止创新成果当事人的利益受损。三是建立岗位定期轮换制度,以增强员工的新奇感和创新的主动性。

##### 2.建立科学有效的创新考核评价体系

企业要建立一套衡量创新成果考核评价体系。这是对创新成果进行考核奖励的基础。考核评价体系要体现公平性和客观性。以有利于奖惩兑现为前提。企业可引进国内外先进的创新评价模型,再根据自身实际进行适当修改和完善。使评价模型更适合本单位实际。评价体系的内容涵盖要全,评价指标要尽可能量化。以减少人为评价产生的误差。

### 3.健全创新激励方式

加强对创新工作的考核奖惩兑现。健全创新激励方式，达到员工的创新成果价值与其所得合理匹配。从而激发员工的创新积极性。企业应当从物质鼓励、精神激励和文化建设三个方面加强和完善。一是注重物质激励。奖励的形式主要有：工资、奖金、红利、福利以及人事待遇。一项创新成果产生后，企业的创新管理部门应根据其成果所产生的预期价值，或者根据创新成果运用于实际工作所产生的当期价值，确定一个比例值作为员工创新成果的收入分红。二是精神激励。企业可制定一定的衡量标准，对达到某个标准的。可授予员工某种称号，并享受一定的权利，如评先评优时予以加分等。在职务级别晋升上，也可将创新能力纳入考核内容。三是文化建设。将创新纳入企业文化建设的内容，在干部员工中营造一种创新光荣、创新受尊重的良好企业文化氛围，体现对创新人才人格的尊重，使他们产生对信用社的归属感。最大限度地利用所掌握的专业知识、技术知识。激发他们创新的内在动力。

### 4.各类新的创新机制

首先是企业制度创新，采用法人治理结构，企业的治理结构科学规范。已实现股东会、董事会、经理层的各司其职、相互制约的企业领导体制。与制度创新的配套展开相关的基本制度，如人事制度、分配制度、财务制度、投资管理制度等也已经制定。

其次是企业市场创新，在原有的大学生需求市场中挖掘痛点以满足消费者需求欲望。通过创新在各方面具有不同程度的服务，可以满足或创造个性化的需求，从而创造大学生市场的“新组合”。综合采用产品创新、市场领域创新、营销手段创新以及营销观念创新，实行最优化组合，以达到本企业经营的目标。

然后是企业组织创新，实行创新型的管理、监督和决策程序，不断倡导创新精神，激发创新意识、引导创新方向、鼓励创新行为、提升创新能力，对从事创新项目的人员进行奖励。奖励既有精神又有物质，从而提高团队积极性，并且把管理思想转换到领导、激励、支持和信任上来。

最后是企业管理创新，提升企业的运营能力，从而在竞争中立于不败之地。以顾客价值为中心、丰富顾客价值、提供个性化产品和服务组合；整合企业内部

和外部与经营过程相关的资源，创造和发挥资源杠杆的竞争优势。

### 3.4.2 激励机制

激励机制是企业实现远大理想的基础，是团队继续发展的动力，也是保证在竞争中处于优势的原因，因此我们将从精神激励、薪酬激励、荣誉激励、工作激励四个方面，依据五项原则激励员工不断进步，不断创新，团结协作，共创美好未来！

1) 精神激励：对员工已经取得的阶段性成果和工作进展及时进行评判和鼓励，向员工阐明组织的目标、组织所倡导的价值观

2) 薪酬激励：向员工阐明组织的奖酬内容、绩效考核标准和行为规范，不断改善工作环境和安全条件，依靠晋升和薪酬的增加来激励员工，提高工作积极性。

3) 荣誉激励：论功行赏。

4) 工作激励：给员工分配的工作要适合他们的工作能力和工作量，要善于听取员工自己对工作的评价。通过基本和高级的培训计划，提高员工的工作能力，并且从公司内部选拔有资格担任领导工作的人才。

## 3.5. 总体发展战略

公司总体战略以通过共享服务以及依托微信小程序、朋友圈等的撬动方式，促进在各个古文识别需求群体的兴起以及使用，以较短的时间在公众面前大幅度地提高平台的知名度。依据市场预估铺点数，我们平台将从 2018 年进入浙江省的多个省市发展到 2021 年基本进入全国古文化市场。同时，我们的整体战略采取以古文专业大学生为中心逐渐向外辐射的市场占领模式。下面将会详细的讲解分阶段的战略目标。前期主要战略目标是往考古专家，古文学者发展，同时完善公司筹备、双向推广、制度完善及技术改进。中期的主要战略目标是往宗教信仰者发展，同时做好技术创新，建立自己的品牌。后期的主要战略目标是往外国有人古文学习爱好者及欧美地区的发展，将业务进行拓展，做到服务对象多元化及



成为一个标杆企业，帮助各个古文学习者快速智能地进行深度学习，逐渐将知名度扩展到全球。

### 3.6. 分阶段发展战略

#### 第一阶段发展战略（公司成立第 1-2 年）

第一阶段时，公司处于未完善体制架构的市场活动时期，此阶段公司主要处于对市场的试水阶段。这一阶段也是重要的积累期，公司积累在古文化市场中的用户以及合适的运营人员，主要以拍照智能识别作为古文深度学习市场的切入口，通过免费使用的模式吸引人的眼光，提高用户流量。

第一阶段以增长型战略为主，致力于在各个用户群体中推广免费识别服务，吸收用户量，同时借助用户的反馈内容训练样本模型。

在第一阶段的目标：

完成在古文学大学生的多个学习使用铺点，设立市场一部；

服务铺设点突破 400 个；

平台活跃用户达 20, 000 人次；

常用运营人员不少于 200 人。

#### 第二阶段发展战略（公司成立 3-4 年）

第二阶段市场主要是以经文使用者的宗教领域发展，主要目标客户有青年、中年、乃至老年等，设立市场二部、三部、四部。通过市场扩张及资源整合实现平台服务多样化，提升平台的市场竞争力，积累流量，逐渐开始建立社群服务。此时在第一阶段的基础上，模型准确率得到一定提高，取消全体用户免费识别的权限，替代为每位用户每年限制十次识别机会。

在第一阶段的发展过程基础上，公司的第二阶段期末发展目标：

完成宗教领域内的 300 个铺点；

平台活跃用户达 1, 000, 000 人次；

常用运营人员达到 20, 000 人。

### 第三阶段发展战略（公司成立 5-6 年）

第三阶段发展战略将平台推广至海外人群，规划占领的区域主要是美国、英国、瑞士、泰国、越南、日本等。服务内容主要分为两方面，一方面通过提供更多的增值服务，提升产品价值，增强用户体验；另一方面为多方利益相关者提供基于智能古文识别的特殊定制服务。

平台第三阶段目标：

完成全国范围内的项目铺点；

使用人次达到 100,000 人次；

平台活跃用户达 100, 000, 000 人次。

## 第四章 市场分析与营销推广

### 4.1 市场分析

#### 4.1.1 行业概况

数字化是古籍再生性保护的重要手段。古籍数字化属于古籍整理的范畴，代表着古籍整理的未来方向。“古籍数字化”是指利用现代信息技术对古籍文献进行加工处理，使其转化为电子数据形式，通过光盘、网络等介质保存和传播。现在网络版古籍主要是将数字化的古籍资源在网络上有偿或无偿发布，供互联网用户使用，这是目前古籍数字化的主要目标。

古籍数字化是对古籍或古籍内容的再现和加工，属于古籍整理的范畴，是古籍整理的一部分。

##### 1. 古籍数字化的现状

古籍的数字化最早是从计算机技术发达的美国开始的。20 世纪 70 年代末，OCLC 和 RLIN 先后推出了《朱熹大学章句索引》、《王阳明传习录索引》、《戴震原善索引》等数据库。到 20 世纪 80 年代以后，中国的台湾、香港和大陆才相继开始了中文古籍数字化项目。目前古籍的数字化可谓“繁花似锦”，除专门的制作公司外，一些单位和个人也投放了不少注意力，并且也确实在各自的工作领域取得了一些重要成果。

##### 一、 丰富多样的文件格式

从文件格式上来看，目前的数字化古籍除常见的 txt、doc、html 格式外，还有 exe、pdf、wdl、pdg、ebk、edb 等。此外，还有位图形式和多媒体格式的数字化的古籍。如此众多的图书格式，显示出中国古籍数字化的途径非常多样。而且，也正是通过这些途径，数字化了的古籍才得以快速地进入了人们的视野，并为广大爱好者和专业学术研究人员所注意。

##### 二、 便捷灵活的传播途径

数字化的古籍大范围进入研究者的视域，应该归功于网络。网上书库、网络



全文检索、网上图书馆和相对专业的个人网站相得益彰，为专业研究者提供了古籍电子文献的不同的使用路径。它们在推进中国古籍数字化上的作用不应被忽视。就广大的中国文化爱好者甚至是专业研究者而言，其对古籍数字化之价值的认识，并非靠专门制作的大型数字化系列图书，而恰恰来自他们简单的网页浏览。

### 三、恢宏阔大的开发规模

中国古籍的数字化主要依赖于学术研究机构和学校，专门技术公司和出版社，图书馆，并且已经有了不少优秀的成果问世。此外，像超星数字化图书馆和地方上一些公共图书馆等也在其数字化图书中不断增添新的古籍，一些古籍研究机构和个人也在工作中将相关文献变成电子文本，也是古籍数字化队伍中的不可小视的力量。这些产品的开发，不仅更加有效地保护了古籍，而且在丰富人们的学习资源、提高学术研究的质量以及扩大中国文化的影响上，无疑也会发挥积极的作用。

## 2.古籍数字化现存问题

古籍数字化的最终目的，是更大范围、更便捷、更有效地使这些文献为人们所利用。而由于用户群体及其需求的差异，这个“用”也显示出不同的层次。因此，满足这些用户群体的需求，也应该是古籍数字化所需要考虑的问题。从这方面看，目前的古籍数字化产品还有不少可以完善的地方。

### 一、缺乏统筹规划

古籍的数字化是一项系统工程，需要通过科学合理的规划和长远的发展战略进行统筹。然而目前我国古籍数字化工作缺乏一定的宏观管理和调控，在此环境下，建设单位多根据自身业务需求开展古籍的数字化工作，缺乏分工协作的意识。同时，各单位已有的古籍数字化建设，也呈现出诸多弊端：第一，选题出现偏颇，资源集中于热点项目，尤其一些商业机构在资源开发时通常集中在少数常用古籍，对文化遗产和学术研究同样重要的一些冷门项目却无人问津；第二，资源集中导致大量重复建设，极大地浪费人力、物力和财力，如《文渊阁四库全书》《二十五史》等古籍资源存在多种电子版本。

### 二、技术问题凸显

已有技术尚不能满足古籍数字化事业，文字处理、检索方式选择、图像制

作、古籍文字识别、辅助功能选择等问题都亟待解决。其中，文字处理问题是古籍数字化工作中最早出现的问题，古籍中使用的文字多是繁体字，同时含有大量生僻字、异体字、通假字、避讳等，而在字符集收录文字不够全的情况下，难免对古籍的数字化工作造成困扰，进一步给检索的准确性带来麻烦，容易造成检索结果失真。现阶段古籍资源库的检索技术是基于文本形式的关键词检索，但古籍的数字化还存在图形、图像等格式，因此现有检索技术的能力在检索深度上还很有限。

### 三、 专业人才匮乏

古籍数字化本质上是古籍整理的延伸，是数字化和古籍整理的结合。因此需要通晓古籍知识的人才参与，同时需要精通计算机信息技术的人员助力。为满足上述条件，我国古籍数字化工作的参与人员一般由从事古籍整理和从事信息技术的专业人员构成。但是“隔行如隔山”，随学科分类日趋精细，日渐增大的学科鸿沟要求两者兼通的复合型人才，就目前而言，这类人才十分稀缺，并不能满足实际工作需求。

## 3.古籍数字化发展趋势

### 一、 加强标准化建设

在大数据的框架下必须加强标准化建设。在确保数字古籍资源格式统一、数据规范、长期可读、便于共享的原则指引下制定标准、统一的规范。在古籍文献数字化管理工作中，可以根据数字化基础，参考国际标准和国外成功的经验，制定适合古籍文献数字化开发的管理标准。

### 二、 建立古籍文献数字化标准数据库

古籍数字化标准数据库是对古籍数字化基础性先天不足的补充，可以解决互联网新技术冲击下暴露出的各种问题，使古籍数字化工作能为社会认知、为用户认可。古籍文献数字化标准数据库便于进入市场。古籍文献数字化标准数据库的出现方便了商业公司，既降低了开发成本，减少开发时间，也可以集中精力寻找社会舆情与古籍文献的相关点，便于把握市场热点。以古籍文献标准数据库为基础的商业公司可以结合社会热点，适时推出与之相对应的产品。产品不仅局限于数据库，可以及时分类推送到使用各种移动终端的用户手中。通过商业运作

引发社会关注，借助市场手段使资源持续流入数字化产业，推动大数据时代古籍文献数字化的进一步发展。

### 4.1.2 环境分析

#### 1.政治环境

政策大力扶持。党和政府历来重视古籍的保护，21世纪以来，对古籍的数字化工作更在政策上给予高度重视和支持。中央号召，地方响应，鼓励学界、业界各方力量积极参与投入，为我国古籍数字化工作的繁荣发展打下坚实基础。

古籍是中华文明的重要载体。自2007年以来，文化部等相关部门实施了"中华古籍保护计划。十年来，在社会各界的共同努力下，古籍保护工作取得了显著成果。2017年初印发的《中共中央办公厅 国务院办公厅关于实施中华优秀传统文化传承发展工程的意见》以及《国家"十三五"时期文化发展改革规划纲要》《文化部"十三五"时期文化发展改革规划》等都对古籍保护工作提出了明确要求。国家发改委“十三五”100项任务中华古籍保护计划也名列其中。《"十三五"时期全国古籍保护工作规划》，经文化部党组会审议通过，于2017年8月正式印发。这也是我国第一个古籍保护规划。

《"十三五"时期全国古籍保护工作规划》的重要任务之一就是加强古籍整理出版和数字化建设。鼓励和支持各古籍收藏单位加快古籍数字化步伐，借助互联网、大数据、云服务等高新技术，率先对馆藏特色文献和珍贵古籍进行数字化，加快建立中华古籍数字资源库和中华古籍综合信息数据管理平台，扩大古籍数字资源开放，促进资源共享，提高利用效率。

#### 2.经济环境

在当代，文化已成为世界范围内经济社会发展的价值维度。经济文化一体化发展、环境与文化的协调发展，都已成为当今全球化发展的重要趋势。文化产业在社会发展中的地位和作用已得到普遍认同，其中古文化更是其中不可或缺的一部分。2017年，古文化产业增加值占GDP的比重进一步提升，达到4.27%。



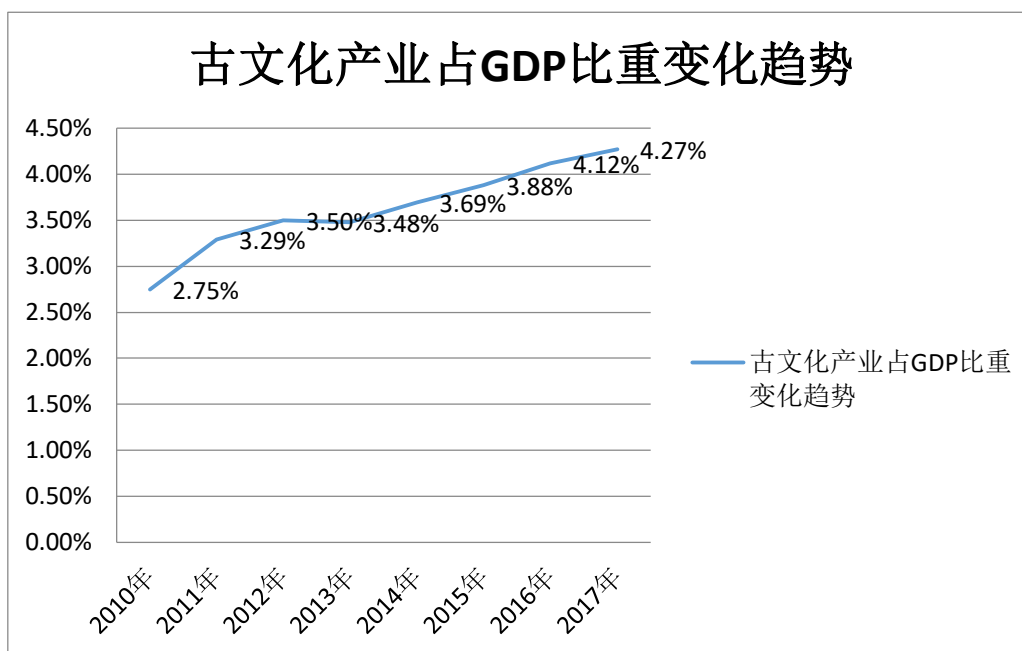


图 30 古文化产业占 GDP 比重变化趋势

### 3.社会和文化环境

中共中央总书记、国家主席、中央军委主席习近平来到北京师范大学看望教师和同学期间，特别提到了关于部分小学语文教材删除古诗词的问题，他表示：“我很不赞成把古代经典诗词和散文从课本中去掉，‘去中国化’是很悲哀的。”次日，据央视官方微博“央视新闻”消息，习近平在专机上再次提到：语文课应该学古文经典，把中华优秀传统文化不断传承下去。”主席的两次表态都在力挺古文以及我国的传统文化，相信社会各界也会很快做出响应，而在传播传统文化的渠道上，除了课程传授外，最主要也最容易为大众接触、接受的应属文化产业的传播方式。

随着教育的广泛普及和深化，学习古文字的人口数逐年增长，主要密度分布在亚洲国家，欧美等发达国家也占据一部分且呈上升趋势。物质生活得到保障后，人们越来越注重精神层面的学习，要求速度快，效率高。

### 4.技术环境

目前市场上的古文识别技术并不成熟，因此带来机遇。市面上现有的古文字处理系统可以根据用户输入的汉字简体字查询出各个历史朝代对应的古汉字，但是，这些古文字查询系统只能够根据简体字查询古文字，而不能通过古文字查询简体字或者形近字。

### 4.1.3 目标市场选择

#### 1.市场细分

当前，知识付费的快速发展正推动着传统出版业的产业化升级，利用互联网对整个出版业进行重构，形成新的知识传播体系。著名的马斯洛需求层次理论提出，人类需求是阶梯式的，由低到高按层次可分为生理需求、安全需求、社交需求、尊重需求、自我实现需求，当较为基础的需求得到满足后才会出现更高级的、社会化程度更高的需求。将马斯洛的需求层次理论套用于付费行业，知识付费是自我实现、自我成长的需求。我们将自己的市场细分为知识付费方向，针对想要研读古籍的人群，包括学生、专业学者及传统文化爱好者，通过收取一定费用向其提供更加优质的古籍阅读服务。

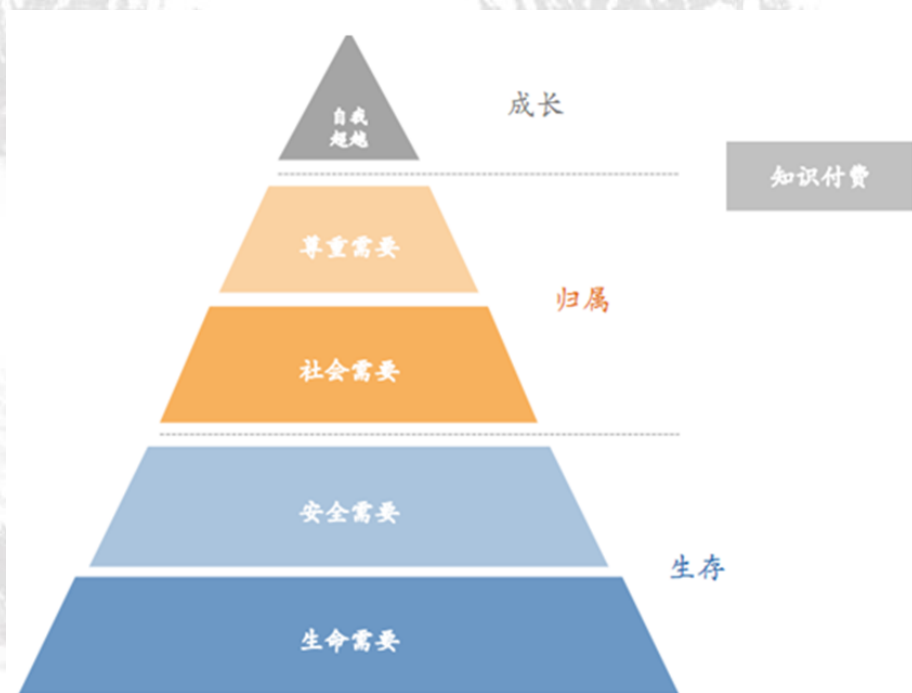


图 31 马斯洛需求层次理论

#### 2.目标市场

我们的定位的市场是古籍数字化及古籍文化交流平台，在现有的古文字查询系统中，几乎所有都只能够根据简体字查询古文字，如篆体字，先秦篆体字，金文，甲骨文等，而不能通过古文字查询简体字或者形近字，或是存在不能顺利识别，识别错误，识别速度慢等问题。古文字需求者往往需要手动查阅资料，但古

籍因其特殊性实体书数目并不多，很多甚至因年代久远易损坏无法直接接触，这也更体现出古籍数字化平台的重要性。我们的平台正是为了解决上述问题而生，正是我们的目标市场所在。

### **3.市场定位**

#### **（1）古文学相关学生**

这类群体多数为青年人群，追求时间效率和识别的准确性，需要识别的古文也较为高难度。当识别器识别缓慢或错误率较高，甚至无法识别时，容易出现不稳定，不耐烦的情绪，降低对平台印象分。我们的平台具备的速度快、准确率高、等识别功能，提高效率和用户满意度。

#### **（2）古文研究者**

这类群体精通更多的古文，在功能上更注重进一步文学内涵的展开延伸，或是对识别内容的进一步整理归纳。市场现有的技术不支持该群体的需求。我们的平台具备的对识别内容的归纳整理，对延伸内容的推荐使我们对于该细分群体的一大竞争力。

#### **（3）考古专家**

该类群体追求识别的准确性，所要识别的文字相比而言更为冷硬生僻，且出土文物模糊不清是阅读困难，这要求平台更加完善和智能化。我们的平台可以方便考古学家地分析新出土的文字的。

#### **（4）外国友人**

这类群体热爱汉语和中国传统文化，希望在识别古文的同时获取更多关于文字含义、用法方面的详细注释，因汉语能力的有限不善于自主查阅纸质文献资料翻译古文，我们的平台具备的拍照扫描自动识别转换功能可提高外国友人的学习效率，并且获取额外的学习信息，提高用户满意度。

## **4.1.4 竞争分析**

### **1.SWOT 分析**



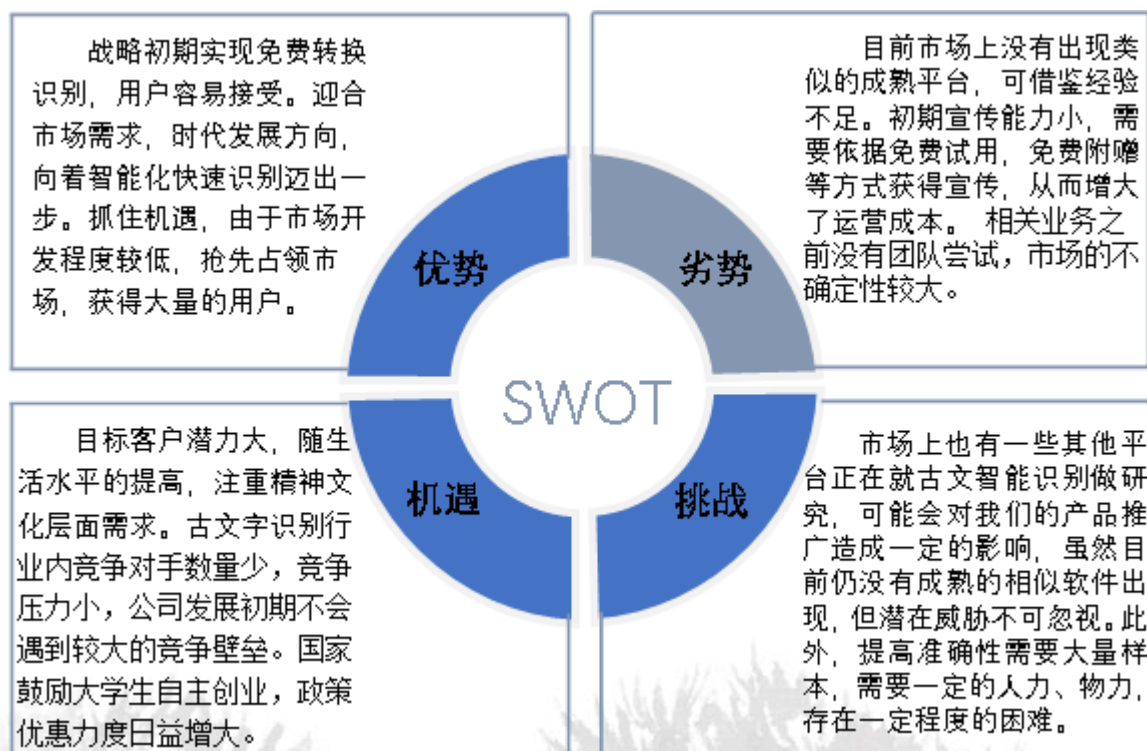


图 32 SWOT 分析图

## 2.波特五力模型

### ● 潜在进入者竞争能力分析：

- 1). 行业内平台已具有一定经济规模，而此时潜在进入的竞争者竞争能力较小。
- 2). 大学生创业政策扶持，潜在进入这竞争较大。
- 3). 提高自身产品不可复制性，打造核心功能，移动跟踪识别，智能识别分割等。

### ● 供应商议价能力分析：

- 1). 供方具有比较稳固市场地位而不受市场剧烈竞争困扰的企业所控制，其平台的用户很多，以致于每一单个用户都不可能成为供方的重要客户。
- 2). 供方的产品各具有一定特色，技术存在一定难度以致于买主难以转换或转换成本太高，或者很难找到可与供方企业产品相竞争的替代品。

### ● 购买者议价能力分析：

- 1). 购买者的总数较多，而每个购买者的购买量较小，不会占据卖方服务的很大比例。
- 2). 古文字智能识别为创新类产品，同行中的产品拥有自动拍照和 360 度扫

描识别的功能，但还不能实现移动跟拍的效果，而较高端的技术成本较大，购买者在其中的议价能力较低。

- 同行业竞争能力分析：

- 1). 需要提高技术的核心功能，具有一定的不可复制性，竞争力较弱。
- 2). 优先占据目标市场，随着社会发展，竞争力必然加大。

- 替代品竞争能力分析：

- 1). 随着社会的不断发展替代品生产者的侵入，使得企业必须不断提高识别速度或者通过降低成本来降低会员价格等、或者使其平台更具有特色，否则其用户数量与利润增长的目标就有可能受挫。

- 2). 不断完善而结合多功能，低价、高质、用户转换成本越低，我们的竞争力就强。

## 4.2 营销推广

### 4.2.1 营销理念

营销理念在营销活动中具有指导性意义，是实现营销目标的基本策略和手段。

聚籍平台（网站和 APP）的主要使用者是对古籍有需求的组织和个人，而繁体识别（小程序）的主要使用者则是对古文字实时性有需求的研究群体和求知群体。但最终的目标市场是便于他们的深度学习。服务价格（如会员会费、数据库使用费等）、用户的接受程度及同类平台服务都在我们的考虑范围之内。只要通过自主推广、合作推广、依托微信进入市场，之后和各个出版社建立长期的合作关系。当然我们深知开拓市场的不易。因此，公司计划每年投入一定量的市场营销费，通过各种方式进行推广。

为了保持平台中关键技术繁体识别的领先地位，公司不断的对现有的数据样本进行扩容和改进，同时通过自主研发和技术引进相结合的方式开发出衍生产品，始终保持对同类企业的竞争优势。总之，公司将脚踏实地，逐步建立健全古籍平台和用户反馈系统，并拓展产品线。

## 4.2.2 营销策略——4ps 理论

通过对营销理念的分析，根据目标市场定位及产品特点，我们选择了以顾客需求为导向的 4ps 营销理论作为项目制定营销策略的指导理念。我们将从产品（Product）、价格（price）、渠道（Place）、促销（Promotion）四个方面进行具体阐述。

### 1. 产品

为了满足不同消费者的需求，我们将提供多元化、多种形式的产品，并且为消费者提供增值服务与售后服务，做到给消费者一个好的购物体验。

①增值服务：我们将不定期的为客户提供系统维护方案，通过数据分析，充分运用最新的信息技术，了解系统的运行状态，优化系统模块，并且为需要的用户提供古籍的阅读与检索等服务。

②售后服务：我们将从建设固定有效的机制出发，保证客户反馈、投诉及相应的信息渠道的通畅，收集用户使用平台后进行的一系列评价，进行分析和针对性地改进，简言之努力打造信息上升渠道，使一线使用者的需要直接反馈给产品中心，从客户群体中获得宝贵的“资源”。为了更好地回馈我们的用户，我们每一年将选出符合相关标准的用户，提供他们免费会员服务或其他增值服务。

目前古籍阅读领域可以说仍是一个蓝海，所以积极发现和挖掘用户需求就尤为重要。在项目从前期构想到投入市场运营阶段，都需要与用户进行不断的互动从而迅速发现和准确捕捉目标市场中的用户需求信息，让用户参与项目产品的完善与升级，与大部分用户一起实现项目产品的协同定制，使平台真正满足用户所需，提供更高的客户价值。

### 2. 价格

根据不同的市场定位，制定不同的价格策略，产品的定价依据是企业的品牌战略，注重品牌的含金量。

①差别化定价：适度让利于合作的高校等长期合作组织，针对不同的使用时间与范围制定不同的价格，以充足的满足消费者的需求。

②分期价格管理：产品初步进入市场时应以比正常价格较低的价格定价，占



领市场份额，中后期发展到一定阶段恢复到正常价格。

③优惠策略：为长期的客户提供一定的优惠、折扣等。

### 3.渠道

由于我们主要以线上销售为主，并且产品为数字化形式，公司将更注重线上销售网络的建设，并不断完善。

### 4.促销

公司将以利用各种信息传播手段刺激消费者购买欲望，促进产品销售的方式来实现我们的营销目标，其中包括对同促销有关的广告、营业推广等可控因素的组合和运用。

①体验式营销：每本古籍的页面将为消费者提供试读页面，更好的吸引顾客消费，也有利于消费者判断自己的需求。

②事件营销：公司将定期参加与古文化、古籍、古文字相关的线下活动，甚至为其赞助来扩大产品的知名度。

## 4.2.3 推广方式

### 1.线上推广



图 33 线上推广图

#### （1）利用浏览器进行推广

浏览器是互联网的一大奇迹，它使浏览者可以方便地在互联网这个信息大海

洋中找到自己所需的信息，也给信息提供者提供了一种受众广、针对性强且效率高的发布途径，越来越多的企业和个人都通过搜索引擎来发现新客户，利用搜索引擎广告或者通过搜索引擎优化。公司每月将会花费一定的费用用 SEO 竞价排名，使我们产品在搜索引擎中的排名靠前，以增加客户发现并了解产品的可能性。

## （2）利用微博进行推广

微博是近两年新兴的媒体，随着他的火热催生了有关营销方式，即微博营销。首先我们将选择新浪微博与腾讯微博建立账号矩阵和链式传播系统，用于品牌传播、建立公共关系和用户互动等。同时，用户在完成日志填报服务功能后，软件会自动弹出微博分享请求，若用户点击“分享”，则由软件自动生成状态在微博上发布，并带有软件下载地址链接并提醒项目微博主页这些状态再通过好友转发，让“熟人的熟人”也获取软件的信息。以这种病毒式方法推广宣传手机软件，传播速度是惊人的。

## （3）利用论坛进行手工推广

网络的普及推动了论坛的迅猛发展，几乎每个门户网站都设有论坛，中国互联网论坛的总数超过 130 万个，位居全球第一。论坛强调的是互动，有共同爱好、共同需求的网友们可以在各类不同的论坛里就自己感兴趣的主题进行交流，坦诚相见、互通有无，相对于商业媒体而言，论坛可以说是网民心中的一处“净土”。

在人人网、豆瓣、论坛、贴吧等主要 SNS 社区以及 BBS 建立官方主页。网站与软件的新动态都会在官方主页上不断更新，包括优惠政策等，通过新鲜事分享发送给所有的粉丝，持续实现品牌升温，建立客户品牌忠诚度。讨论区狙击了项目的用户和潜在用户，讨论关于安全生产管理等信息，工作热暖可以随时和用户们进行交流，在线调查系统可以进行长时间持续的一线信息了解，而且信息面大，使用方便。调查费用几乎为 0。

## （4）利用网络广告进行推广

网络广告是指在互联网上发布的所有以广告宣传为目的的信息，如图像式网络广告、网络联盟广告、关键词广告、邮件广告等，随着互联网的迅猛发展，网络广告已经成为网络推广的一种主要形式。与传统广告相比，网络广告有很大的

优势，如传播范围广、不受时空限制、交互性强、效果可量化、能有效监控、投放灵活有针对性、有文字语音视频等多种载体、费用相对较低等。在选择网络广告时，应根据自己的产品情况、经济能力选择合适的网站、合适的广告位和时段进行投放。企业经理、总监比较喜欢通过网络看视频节目，例如“世界青年说”收视率攀升，而过两期节目我们可以发现，凡是在节目中或视频播放前植入广告的产品均取得了不错的增长率，我们可以使用该种方式，直接在播放视频投放网络广告或者植根于节目中间，以快速取得提升。

#### （5）利用社交软件进行推广

常见的推广软件有邮件群发软件、QQ 群发软件、论坛群发软件、搜索引擎登录软件等，通过大量发贴，让更多的浏览者知道自己网站或产品的相关信息，但注意不要滥发未经许可的垃圾邮件，一定要提供给接收人有用的信息。

#### （6）通过自有的小程序产品宣传

微信小程序除了放其他广告也可以放我们 APP 和网站的广告。因为小程序我们将提供给顾客免费拍照识别服务，有利于向 APP 和网站进行引流。

网络推广的方法很多，不同的方法各有自己的优缺点，经常需要多种方法综合去用，不能单纯地只用一种方法，而网络是个虚拟的世界，到底哪些方法的组合最适合自己、最有效需要进行长期的测试，找到以后再加大这个组合的投资，把效果放大，这样才能达到事半功倍的效果。

## 2. 线下推广

#### （1）提供赞助

积极协助古文化学习网站活动的举办并为他们提供相关赞助，使公司品牌名称能够在古籍、古文字行业取得一定的影响力与知名度。

#### （2）参加行业峰会

在公司拥有一定的行业声望时，参加古文行业峰会与不同类型的古文化方向企业形成战略合作，有利于进一步扩大公司的渠道。



## 第五章 商业模式和盈利模式

### 4.1 商业模式

商业模式，在综合了各种概念的共性的基础上，提出了一个包含十个要素的参考模型。这些要素包括价值主张，消费者目标群体，分销渠道，客户关系，价值配置，核心能力，价值链，成本结构，收入模型，裂变模型。本节，我们主要将从客户细分、价值主张和客户关系介绍我们的商业模型。

#### 4.1.1 客户细分

我们针对不同的客户享有不同的具体服务主要将客户大致分为普通客户和会员客户。

普通用户在我们的平台在线阅读古籍可以试读，了解平台优秀的阅读模式及资源，并可以根据自己的需求购买单独的数据库。同时可以自由参与平台的论坛社区，查看新闻资讯，在实体书商城购买商品。

对于平台的会员客户，主要是对公司平台有所了解并期望更多使用权限的客户，公司将按其需求收费，具体按日或按月或者按年由客户自由选择收取费用。会员客户享有平台所有权限，可以自由地使用平台所有功能。

#### 4.1.2 价值主张

“聚籍”是一家专注于古籍数字化的平台，我们希望通过我们特有的繁体识别转换技术推动古籍数据化的发展，为大众提供知识服务，传播优秀传统文化。

通过市场分析，我们发现目前市面上的繁体识别系统不具备智能识别功能，且转换速度慢，准确性低，识别范围狭窄。为了解决以上问题，我们公司秉持着创造更加优质的服务这个愿景，不断地探索古文识别行业的盲区，希望可以为顾客提供更加优质的服务。定位客户类型，读懂对方潜在的心理活动，进而带给客户更好地服务体验。

### 4.1.3 客户关系

公司客户主要为古文研究者和相关专业学生等。作为一个服务类型为主的企业，主要为其提供在线古籍阅读研究的服务。公司发展后期，我们的数据较之初期更为丰富和准确，将寻求更多方向的合作，例如博物馆、图书馆等等。随着公司发展，我们将愈来愈深刻地了解客户的需求并进行客户导向的投资，双方人员共同探讨行动计划，企业对竞争对手形成了很高的进入壁垒。客户将这一关系视为垂直整合的关系，客户企业里的成员承认两个企业间的特殊关系，他们认识到企业的产品和服务对他们的意义，有着很强的忠诚度。在此关系水平上，价值由双方共同创造，共同分享，企业对客户成功地区别于其竞争对手、赢得竞争优势发挥重要作用。双方对关系的背弃均要付出巨大代价。企业对客户信息的利用表现在战略层面，关系的核心由价值的分配转变为新价值的创造。

## 5.2 盈利模式

公司主要的盈利模式为流量变现、会员会费、网络广告、数据库购买收入以及特殊用户相关增值服务。前期的盈利项目主要包括流量变现、会员会费等。在公司运营状况、财务状况稳定之后，中期平台建立了一定的影响力，留存了一定数量的核心用户，在保持并发展前期盈利模式的情况下增加网路广告收入以及出版社与学校等组织合作收入。后期平台已具备一定的市场份额，影响力较大，增加推广收入，古籍售卖等，实现公司盈利模式的多元化发展。

### 5.2.1 前期盈利

#### 1.流量变现

微信小程序、朋友圈等是我们营销中的重要途径。我们公司用户免费体验后的分享推广在各自领域，获取更多不同类型的客户群体，最终汇聚流量。

#### 2.会员会费

由于古籍等资源特殊性，我们公司将根据不同的需要向广大的用户开放期限不同的会员服务（网站和 APP 的会员同步）。并且根据部分读者需求不同按照

数据库分别提供会员服务。会员可享受的服务及价格如下表所示：

表 2 会员可享受的服务及价格

时间	服务	价格
1 天	全部数据库阅读	18 元
1 月	全部数据库阅读	480 元
一年	全部数据库阅读	3380 元
一年	单个数据库阅读	180/280 元

### 3.与用户实现共赢

微信小程序与 APP 为用户通过免费智能识别古文识别的功能，同时用户在使用过程中，用识别结果反馈和纠正，因此我们的平台在前期原有的数据库基础上，随着用户数量和使用次数的不断扩张，对原有数据库进行改错、纠正和完善，丰富自己的数据库，为中后期进一步的盈利奠定基础。

## 5.2.2 中期盈利

### 1.广告收入

在公司自己的网站/APP 之中接入广告，赚取额外的利润。广告收入有多种方式，在不影响界面整洁的情况下我们在项目宣传网页上放置一些有相关 API 服务供应商家的广告，主要形式有横幅式广告、边区窄广告和其他植入式广告，借此收取广告费。由于该盈利方式会影响用户体验，因此该方式在产品投入市场的第四年开始实行。在 APP 端，开发时便嵌入广告平台软件，寻求第三方广告平台合作，按客户点击广告数获得广告分成。在网页端，初期同样依靠广告联盟的方式，采取按点击计费 and 展示次数计费的方式。

表 2 预收费标准表

首页广告	200 元/天；
其他区域窄广告	150 元/天/条；
植入广告费用	80-100 元/天/条；

### 2.合作收入

中期平台建立了一定的影响力，留存了一定数量的核心用户。可以寻求与各个出版社的合作，获取出版社相关古籍数据库的同时，也为出版社提供相关的古



籍识别服务，达到双赢。也可以与高校等进行长期合作，为其踢狗更加优质、优惠的服务。

### **5.2.3.后期盈利**

当公司进入稳定期，已积累一定数量的会员和资金，有了一定的知名度和美誉度，订阅流量也十分可观，公司已经有能起向多元化的盈利模式拓展，更多地盈利方式引入，公司进入后盈利模式时期。

该时期的变更或增加的盈利模式具体内容如下：

#### **1.推广收入**

在后期，我们以免费的方式与一些大型教育系统和智能翻译识别网站进行合作；而加盟商一般是急需推广尚未成名机构本网站会收取一定的推广费用。

#### **2.特殊定制需求收入**

该盈利模式针对有特殊需求的客户，如对古文识别内容有特殊需求，需要定制特殊服务的用户，或者是对数据库有特殊要求和想要更高特权的客户或需要某种解决方案，公司将以直接商谈的方式确定为其提供的特殊应用服务以及收费方式。

#### **3.实体书售卖收入**

后期公司的古籍数据库逐渐完善，公司可以引入实体书售卖。在获得版权的前提下，售卖相关的古籍的繁体或简体版本。也可以与出版社合作，将出版社的书放在我们的平台销售，增加他们的销售渠道，获取相关收入。

## 第六章 财务分析

### 6.1 融资计划

#### 6.1.1 前期融资方案

平台成立之初，结构精简、人员精炼，注册资本预计为 100 万元，自筹部分 60 万，剩余部分根据公司状况可以选择向金融机构借款或风险投资，考虑到金融借款要求的严格，我们选择风险投资的方式。公司自筹基金为 60 万，在公司的资金结构中，60%的资金来源于公司创业团队成员自筹，剩余 40%寻求风险投资。

为了给本公司今后的运营创造良好的条件，避免风险投资资金退户造成的运营风险与损失，我们拟风险投资主要从大学生创业基金、大公司、风险资本家和天使投资人等处取得。这样不仅更加容易筹集到资金，而且有利于化解风险，为公司创造更好的发展前景及理财环境。

公司的股本结构及规模如下表所示：

表 3 公司股本结构与规模

股本来源	自筹资金	风险投资
金额	60万	40万
比例	60%	40%

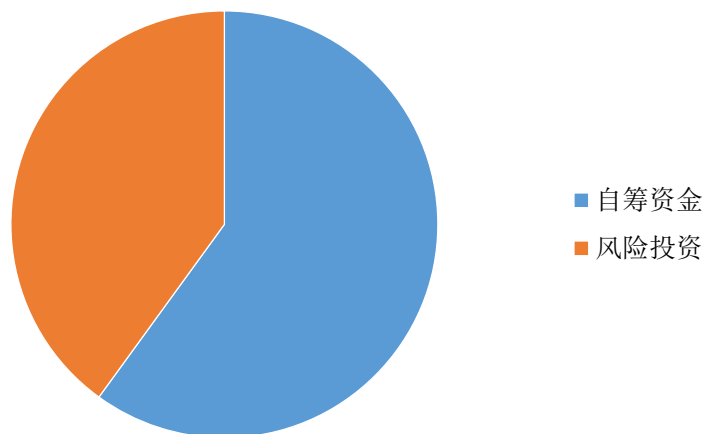


图 34 公司股本结构与规模图

### 6.1.2 中期融资方案

公司计划在总体战略发展阶段进行一定规模的扩大，需要更多的资金。中期所需的大量资金主要来源如下：

#### 1. 公司前期留存收益

为保证公司正常运营以及规模扩大所需资金，公司拟定在运营第一阶段、第二阶段不进行分红或者少分红的利润分配方式，增加未分配利润与盈余公积，为公司的业务拓展提供必要的支持。

#### 2. 进行新一轮股东筹资或风险投资

考虑到公司资产结构，公司向银行借款存在一定困难。因此在上市受阻或上市后流动资金仍不能满足公司快速发展时，公司将选择向股东进行新一轮的筹资并适当引入一定的风险投资。

#### 3. 上市融资

将经营公司的全部资本等额划分，经批准后在深交所创业板上市流通，公开发行。短时间内可筹集到巨额资金。此举不仅能够取得大量的资金，公司同时可以通过股市的情况为公司免费做广告，从而在无形中提高股价，发挥上市的明星



效应。

#### **4. 银行贷款**

在公司经营情况良好的情况下，可向银行进行小额贷款。

#### **5. 出版社合作**

公司将在发展中期寻求与出版社的合作，互相丰富数据库的内容，并吸引其投资。

### **6.1.3 后期融资方案**

公司将在战略成熟阶段进行规模的扩大和服务方向多元化，需要更多的资金。拓展业务所需资金主要来源如下：

#### **1. 向银行贷款**

经过较长时期发展，相信平台已具备相当水平的盈利能力及声誉。此时，向银行进行较大金额的贷款也具有一定的可行性。在再融资情况不理想或创业团队的股份被过多稀释的情况下，考虑向银行贷款，以短期贷款为主，能比较及时地获得所需的资金，但是此方式相对风险较大。必要时考虑少量长期贷款。

#### **2. 增发新股**

随着后期公司规模的扩大，公司将选择增发股份。对于定向增发及公开发股的选择及分配方式适现实情况而定。

## **6.2 财务预测**

### **6.2.1 基本财务假设**

1. 会计制度：公司执行《企业会计制度》
2. 会计年度：以公历每年一月一日至十二月三十一日止为一个会计年度
3. 记账本位币：以人民币为记账本位币
4. 记账基础和计价原则：以权责发生制为记账基础，以历史成本为计价原则
5. 固定资产折旧：固定资产估计使用寿命 5 年，按 5 年计提，残值 5%，按直线折旧法计算；

6.无形资产按 10 年来直线摊销。

7.应收应付款及应付职工薪酬处理：

A.预计企业第二年的应收账款为该年营业收入的 10%，之后 2-5 年的应收账款为当年营

业收入的 20%。

B.应付账款记为当年销售费用的 10%

C.每年的应收、应付款均于下一年年初收回和支付

D.应付职工的薪酬的 10%在下一年年初支付

8.企业将在每年年末对无形资产进行减值测试，预计五年内无形资产不发生减值

9.公司试用的税种和税率列示如下：

A.公司营业税及附加税（5%）

B.企业所得税（25%）

10.法定盈余公积计提比例按当年净利润的 10%

11.为保证公司正常运营以及规模的扩大，公司拟定前五年不分红。

### 6.2.2 公司前期投资预测



图 35 公司前期投资方向

公司成立初期资金主要集中投入于电商平台开发这一模块，巩固硬件基础。因为公司经营基本以电商为主，在公司建立前期我们考虑把计算机、服务器等固定资产购置作为资金的第一投入点。其次技术投入（无形资产）为资金的第二投入点。然后技术人员的开发费用为资金的第三投入点。最后广告投入的推广费用以增大知名度是资金的第四投入点。在平台发展中后期，根据市场需求进行再一次的计算机、服务器等网络设备的添置。在平台发展成熟阶段，我们考虑扩展为其他互联网商务沟通软件提供服务以及系统维护与升级。



表 4 平台前期投资预算表

前期	投资项目	单价	数量	总价
硬件设施支出				
	办公地点租赁费	60000	1	60000
	计算机	5000	10	50000
	服务器租金	8000	2	16000
	办公设备	30000	1	30000
	装修费用	20000	1	20000
技术投入				
	无形资产	200000	1	200000
推广费用				
	广告投放			200000
	推广活动			100000
	推广人员差旅费			20000
人力成本				
	开发技术人员工资		3人	150000
	财务人员		1人	40000
	市场人员		若干	150000
	其他人员		若干	50000

编制说明：

①固定资产按照设备的市场估值计算。

②办公设备包括电脑设备、办公座椅、固定电话、打印机、办公易耗品等。

### 6.2.3 未来五年收支预测

表 5 未来五年现金支出预测表

项目	第一年	第二年	第三年	第四年	第五年
主营业务成本					
服务器租金	16000	16000	24000	36000	48000
技术开发费用	100000	200000	300000	300000	350000
系统维护与升级	50000	70000	80000	90000	100000
其他	15000	20000	25000	30000	35000
小计	181000	306000	429000	456000	533000
管理费用					
财务人员	50000	100000	150000	180000	200000
差旅费	30000	30000	40000	50000	80000
人力资源	40000	45000	60000	70000	90000
其他	20000	20000	30000	40000	50000
小计	140000	195000	280000	340000	420000
销售费用					

市场人员	80000	80000	100000	120000	140000
广告投放	200000	300000	400000	450000	480000
推广活动	100000	120000	150000	160000	180000
其他	20000	25000	30000	35000	40000
小计	400000	525000	680000	765000	840000
其他支出	8000	10000	12000	15000	20000
合计	729000	1036000	1401000	1576000	1813000

编制说明：

1. 主营业务成本：公司的主营业务成本主要是对赞助宝平台技术研发人员的薪酬费用及硬件成本。

2. 其他职工薪酬：由于公司初期规模较小，且人员以开发技术人员为主，因此公司人员数量不大，初期预估人员 12 名，在公司发展中后期再依据需求适当地扩大员工人数。

3.销售费用：公司在日常运营过程中，会产生一定金额的费用支出，如销售人员与合作方、客户等签约交涉发生的业务招待费。

4.管理费用：公司正常的日常运营离不开决策与管理部門的统筹规划与安排，决策与管理部門的日常消耗其他费用将作为管理费用支出，如管理与决策部門因公出差所需的差旅费等等。



表 6 未来五年折旧与摊销预测表

项目	第一年	第二年	第三年	第四年	第五年
计算机	4000	4000	4500	5000	6000
办公设备	2000	2000	2500	3000	3000
其他	1000	1000	1500	2000	2000
折旧合计	7000	7000	8500	10000	11000
无形资产	6000	6000	6000	7000	7000
摊销合计	6000	6000	6000	7000	7000
折旧与摊销合计	13000	13000	145000	17000	18000

编制说明：

公司固定资产主要有电脑设备、办公桌、空调、服务器等，固定资产预计可使用年限为 10 年，预计净残值为 0，采用年限平均法计提折旧。公司的长期摊销主要是办公场地的装修费用，公司初次装修费用分 3 年平均摊销，第二次装修费用分 2 年平均摊销。无形资产预计可使用年限为 20 年，加速折旧。

表 7 未来五年收入情况预测表

项目	第一年	第二年	第三年	第四年	第五年
网站广告	0	500000	800000	1000000	1200000
出版社等合作	10000	150000	300000	800000	900000
活动主办商	0	400000	500000	700000	1000000
政府支持	100000	100000	100000	100000	100000
其他增值业务	0	0	10000	50000	100000
合计	110000	1150000	1710000	2150000	3300000

### 6.3 财务报表

表 8 利润表 单位（元）

项目	第一年	第二年	第三年	第四年	第五年
营业收入	30000	890000	1210000	1615000	1265000
减：营业成本	181000	306000	429000	456000	533000
管理费用	140000	195000	280000	340000	420000
销售费用	400000	525000	680000	765000	840000
其他	8000	10000	12000	15000	20000
利润总额	-619000	114000	309000	574000	1487000
减：所得税费用	0	28500	772501	143500	371750
净利润	-619000	85500	231750	430500	1115250

表 9 资产负债表 单位（元）

项目	第一年	第二年	第三年	第四年	第五年
流动资产：					
货币资金	350000	368000	497400	466000	347700
应收账款	0	23000	12000	0	330000
流动资产合计	350000	391000	509400	266000	677700
非流动资产：					
固定资产	385200	3561000	334900	453000	409600
长期待摊费用	0	82000	67000	54000	48700
非流动资产合计	385200	3643000	401900	507000	458300
资产合计	735200	829100	911300	973000	1136000
应付账款	0	35870	23360	0	11200
短期借款	10000	0	0	0	0
应付职工薪酬	125200	153230	204500	196800	245000
应交税费	0	0	0	0	0
负债合计	135200	179100	227860	196800	256200
实收资本	600000	600000	600000	600000	600000
盈余公积	0	0	10000	40000	100000
未分配利润	0	50000	73440	136200	179800
所有者权益合计	600000	650000	683440	776200	879800
负债及所有者权益合计	735200	829100	911300	973000	1136000



## 6.4 主要财务指标分析

### 6.4.1 静态投资回收期

投资回收期是指以投资项目经营净现金流量抵偿原始总投资所需要的全部时间，是评价投资可行性的重要指标，投资回收期越短，说明投资风险越小。

根据公式：得  $t=3.95$ 。通过净现值流量、投资额等数据计算出回收期为 3.95 年。说明回收期较短，该项目在财务上可行。

### 6.4.2 投资净现值（NPV）

投资净现值考虑的是资金的时间价值和整个项目的寿命周期的现金流量情况，从动态角度反应了该项目的投入和净产出的关系。

假设资本成本为 10%，由净现值  $NPV = \text{未来现金流入现值} - \text{未来流出现金现值}$  可得：

以五年的现金流计算： $NPV=268716.36$

因为  $NPV$  大于 0，即投资后可望获得的投资报酬率将超过投资者要求的报酬率，因而在财务上此方案可行。

### 6.4.3 内涵报酬率（IRR）

内部报酬率是指投资后实际可望达到的投资报酬率。从动态的角度直接反映了投资项目的直接受益水平，它是指能够使未来现金流入量现值等于未来现金流出量现值的折现率。

根据预计现金流，以 5 年为时间跨度，计算得  $IRR=26.8\%$

计算可得，公司投资的内涵报酬率为 26.8%，大于资本成本率，说明在财务上可行。

## 第七章 风险管理



图 36 风险种类图

### 7.1 财务风险

#### 7.1.1 概述

财务风险指企业由于不同的资本结构而对企业投资者的收益产生的不确定影响。本公司的财务风险主要表现为产品推广的投资成本和资本回收过程所面临的风险，产品推广的成功与否直接关系到产品在未来的发展，而资本回收的周期长短关系到公司的生产运营，从而影响公司的盈利水平。

#### 7.1.2 产生原因

(1) 产品的推广是一个产品前期成长的必经过程，其周期较长，耗用资金和时间成本较大，面临的财务风险较大。

(2) 资本回收率影响本公司的盈利状况，从而影响本公司的资产结构，若运营不善将会导致本公司处于亏损状态，会增加公司的借入资金，提高了借入资产比例，这会提高财务风险。

(3) 市场利率的变化会增加借入资本利息差额的不确定性，这可能会提高

### 7.1.3 对策

(1) 实行严格的资金借贷和运用审批制度，根据公司发展情况和资金市场变化，调节结构。

(2) 使投资项目尽快产生收益，提高资产盈利能力，降低投资风险。

(3) 加强对业务收入、业务支出、日常现金等的管理，在保持较高的流动性基础上减少资金占用，为公司扩大投资提供现金流。

(4) 加大资本运营的力度，构建和拓宽畅谈的融资渠道，为企业资金供应建立稳固的渠道，为公司的发展不断输入资金。

(5) 建立相应的风险预警机制，加强内部管理，严格规章制度，把可能发生的损失降到最低程度。

(6) 为避免企业在发生意外及其他不可抗拒因素给企业带来损失，将在财务预算中拨出专款，购买各种保险以规避可能遇到的风险。

## 7.2 技术风险

### 7.2.1 概述

本项目的技术风险主要表现在产品的技术周期短，容易被替代，技术创新过程中，质量达不到预期，或无法及时完成创新任务，造成资金短缺。创新投资难以收回而陷入技术困境风险。没有了解市场需求，而过度研发，造成资金浪费，成本价格过高，失去市场。研发过程中，资金流突然中断风险，开发中不可避免会遇到各种各样棘手的问题需要解决。再有就是知识产权和商标方面，如果项目被他人抄袭模仿，有可能会造成巨大的灾难。

### 7.2.2 产生原因

(1) 古文识别缺乏相关成功可参考模式，新兴技术使用经验不足，在开发



中不可避免会遇到各种各样的问题。

(2) 当今社会，科技变化日新月异，更新换代发展较快。在技术革新和发展方面也存在一定的风险。

(3) 由于大学生对于知识产权的保护意识不是很强。稍有不慎便会造成巨大的损失。

### **7.2.3 对策**

(1) 积极请教在技术方面有丰富经验的指导老师，做好面对问题迎难而上解决问题的心理准备，技术开发人员首先要保持对项目开发的乐观心态。

(2) 加大研发力度和资金支持，提高技术人员的薪酬待遇和激励机制，留住人才。

(3) 定期对团队成员进行知识产权保护培训，增加团队成员对知识产权保护的意识，每当产品有重大进展时便着手申请专利和软著，在产品研发初期就要完成商标的注册事宜。

## **7.3 市场与竞争风险**

### **7.3.1 概述**

目前创新产品每年都有很多，但是能被市场接受的很少，到时公司上市因股市价格、利率、汇率等的变动而产生的市场风险。此外，古文识别技术的内部网站和软件运营模式容易被复制，知识产权保护相对困难，在推广的后期可能会出现服务类似本项目的产品来参与竞争。所以如何应对潜在竞争者威胁也在我们的考虑之中。

### **7.3.2 产生原因**

(1) 潜在竞争对手的威胁。当开发技术比较成熟，市场有较可观的盈利性情况下，容易有大量同质团队与公司涌入此行业，占据部分市场份额。

(2) 市场认可度较低。作为大学生创业企业的创新产品，初期市场认可度较低，市场对新进企业需要有一个了解和认可的过程。

### 7.3.3 对策

(1) 要不断进行古文识别技术改进方面的市场调查，进行分析和收集资料，把握行业波动规律。与此同时，建立市场的信息反馈机制。建立市场信息反馈机制，及时、准确地了解合作商和用户的需求以及竞争对手的变化情况。充分利用行业开拓者的优势，尽快扩大市场份额。

(2) 进行一段时间的试运行，在总结分析各项数据后，在结合之前的市场调研

研、企业反馈情况和专家分析，最后确定我们是否进入目标市场。

(3) 在一定的技术支持下将新型模式注册知识产权。

## 7.4 成员管理风险

### 7.4.1 概述

作为一支大学生创业团队，我们有多样化人才组合，每个人都各有所长，恰恰是这样，这支队伍就缺一不可。这支队伍可能会出现队员士气低下、缺乏凝聚力的情况。也可能出现队员忠诚度降低导致的人才流失。最重要的一点是，学生缺乏创业经验，在实际运营的组织与管理过程中易存在缺陷。

### 7.4.2 产生原因

(1) 知识储备不够充足，对生产管理这方面的认知还存在一定的缺陷。

(2) 人才缺乏、市场营销策略不当或管理经验不足导致运营达不到预期效果的风险。

(3) 缺乏必要的规范、惩罚制度与合适的激励措施，增加了工作失误率，重复工作，降低工作质量和工作效率，降低项目成员积极性。

### 7.4.3 对策

（1）对团队成员进行专业的生产管理培训，提高自身综合素质，掌握先进的管理知识和科学技术知识。并运用科学的管理理念和一些管理工具平衡好技术服务与市场营销之间的工作，使项目的运作达到最佳的状态。

（2）创建更多反馈途径，让项目成员有更多的机会与管理层沟通，管理层也要充分重视项目成员意见并认真考虑他们所反映的问题，以物质或非物质的形式肯定项目成员的工作，让项目成员知道自身对于项目的重要性。

（3）加强团队建设，建立具有充分弹性、敏感性和适应性的团队组织。同时加强成员之间的合作精神，了解各自在安全生产管理和软件技术研发等方面的工作内容，使团队成员在组织分工上能够有很好的配合。