

专注于深度学习 OCR 原创技术突破

读光 OCR

原子能力产品白皮书 (V1.0.0)



目录

CATALOG

1. 前言	6
2. 产品概述	8
2.1 产品简介	8
2.2 产品技术架构	9
2.3 产品服务模式	10
3. 产品介绍	12
3.1 通用文档类 OCR 识别	12
3.1.1 全文识别高精版	18
3.1.2 文档结构化识别	19
3.1.3 英文专项识别	19
3.1.4 网络电商图文识别	20
3.1.5 多语言文字识别	21
3.2 卡证票据类自动分类及识别	22
3.2.1 个人卡证类识别	22
3.2.2 企事业资质证明类识别	29
3.2.3 发票凭证类识别	32
3.3 泛化场景 OCR 识别	35
3.3.1 卡证票据混帖识别	36
3.3.2 试题作业识别	36
3.3.3 批量文档识别及比对	37
3.3.4 Video OCR 识别	37
3.4 自定义自学习 OCR 识别	38
3.4.1 KV 结构模板	38
3.4.2 自学习 OCR 平台	41
3.4.3 图文内容逻辑顺序识别	42
4. 典型应用场景	43
4.1 办公自动化场景	43
4.2 电商治理场景	43
4.3 智慧教学场景	44
4.4 RPA 场景	45
5. 产品规格	45

5.1 公有云产品规格.....	45
5.2.1 性能指标.....	45
5.2.2 订购及授权.....	46
5.2.3 调用方法说明.....	46
5.2 私有化产品规格.....	46
5.2.1 性能指标及硬件要求.....	47
5.2.2 服务部署说明.....	47
5.2.3 订购及授权.....	48

| 01 前言

FOREWORD

1. 前言

光学字符识别 (Optical Character Recognition , OCR) 是最早的人工智能领域技术，主要是将图片及各种印刷手写文稿上的文字内容智能地提取并识别成可编辑的文本信息。

自 2012 年以来，随着深度学习 (神经网络) 算法和计算机视觉技术的高速崛起，以及基于 GPU 并行矩阵运算芯片和 NPU 专用 AI 芯片的广泛应用，结合海量数据积累和持续的算法训练优化迭代，基于深度学习的各类人工智能技术已经非常成熟并取得了巨大的商业化落地应用。

在信息化、数据化、智能化的时代大背景下，越来越多的企业正在寻求通过引入先进技术能力，不断优化业务流程来提升企业的生产工作效率，其中，OCR 识别为企业的数据化革命提供了强有力的技术支撑。近年来，随着 OCR 算法识别精度的飞速提升，以及在文字理解能力方面的不断突破，OCR 技术已相当成熟并，并被广泛地运用在金融、电商、医疗、政务、互联网平台等多个行业领域，为企业和个人带来巨大的价值，OCR 技术俨然已成为人工智能领域飞速前进的三驾马车之一。

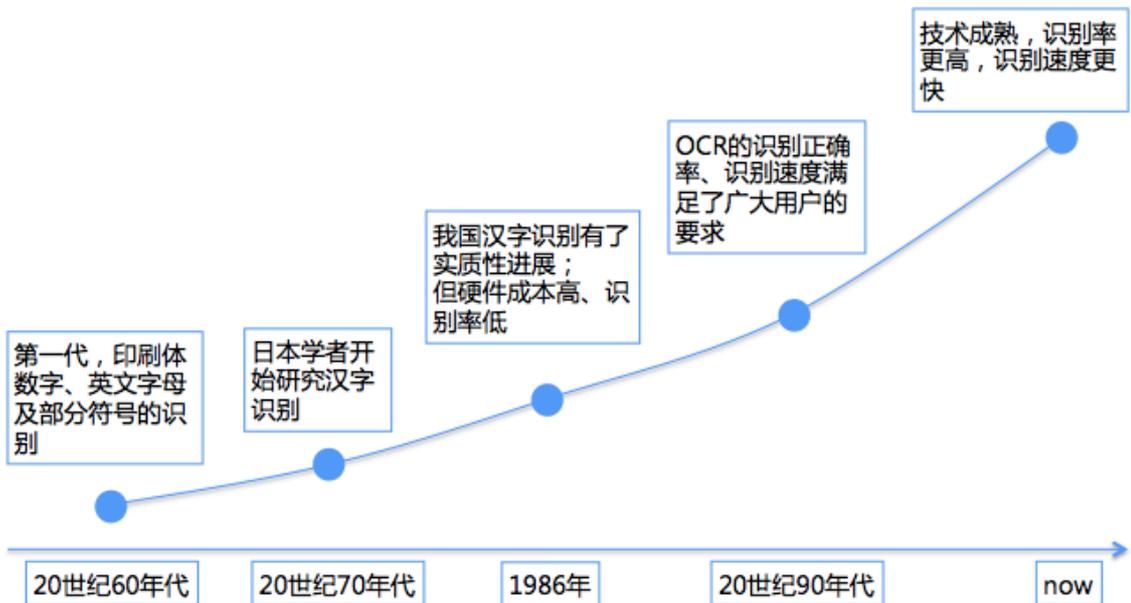


图 1-1 文字识别技术发展趋势图

| 02 产品概述

INTRODUCTION

2. 产品概述

2.1 产品简介

读光 OCR 是一款由阿里巴巴达摩院权威 AI 团队打造的全栈全场景 OCR 文本识别类产品。多年来, 通过持续整合前沿 AI 技术和行业实战经验, 读光 OCR 打磨出了能够承载跨行业敏捷应用的技术架构, 具备图像文字定位、文字识别和文字理解的全栈技术体系, 形成了包含通用文本识别、卡证票据自动化分类及结构化识别、卡证票据混帖识别、表格文档批量结构化识别以及试题作业智能识别等全量 OCR 原子能力矩阵, 支持公有云调用、私有化部署、终端 SDK 等云边端全业务场景服务能力。

基于业界领先的深度学习技术, 读光 OCR 能够提供多场景、多语种、自学习、支持智能检测分类、图文语义理解以及自定义结构化模板的高精度整文字检测和识别服务, 多项 ICDAR 指标居世界第一。

读光 OCR 于 2010 年孵化于阿里巴巴集团内部, 历经近十年阿里巴巴集团以及各行业客户的海量商业化实战应用洗礼, 在 OCR 识别准确率和处理性能、产品成熟度和能力覆盖范围等方面均稳居业界领先水平。目前, 读光 OCR 公有云持续为阿里巴巴集团和海量行业客户提供日均访问量超亿次, 累计处理量超 2000 亿的高精度、高性能、高稳定性的文字识别服务。

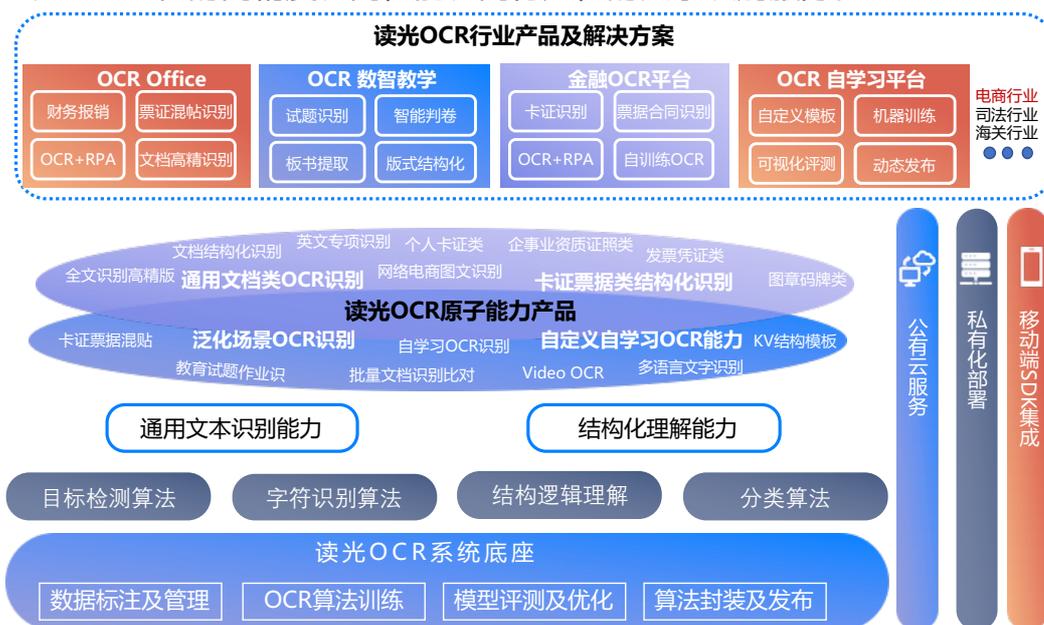


图 2-1 读光 OCR 整体业务矩阵

2.2 产品技术架构

读光 OCR 依托全文识别和结构化识别两大基石功能，能够灵活高效可扩展的提供多场景、多业务流和高度自主性的识别覆盖能力。相比于市场上其他友商的产品，读光 OCR 以其轻量接入和高灵活度的特征赢得客户的喜爱和市场的认可。无论是自动化办公场景（人工审核、录入、信息比对）、信息监管场景（鉴黄鉴暴、恶意广告的语言甄别）还是行业数字化转型需求，读光 OCR 都能轻松应对，助力政企客户提升业务效率，实现数据价值化。

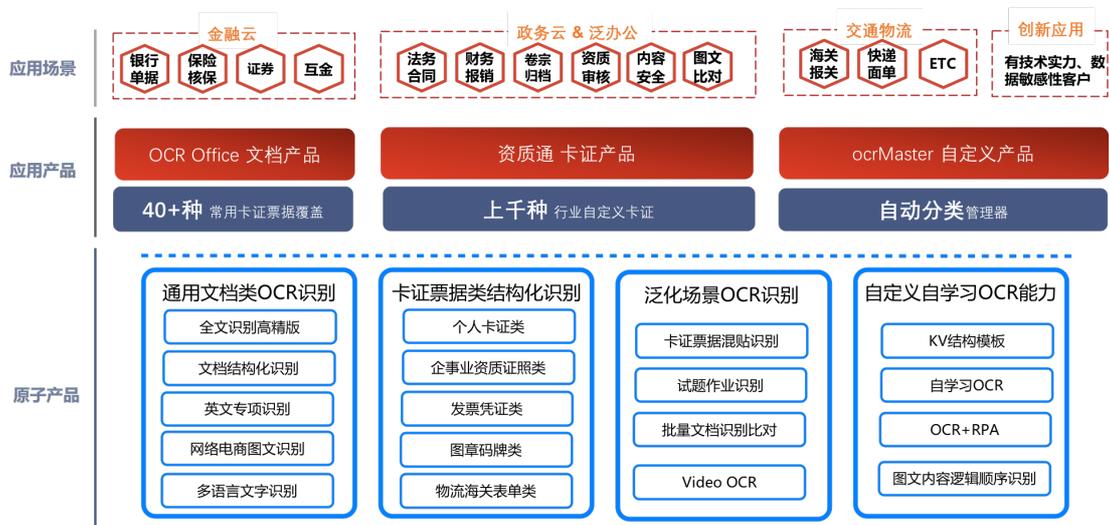


图 2-2 读光 OCR 整体业务架构体系

读光 OCR 通过强大的算法能力和系统化的工程设计，不断加强 OCR 识别的系统性能，并通过自然语言处理和计算机识别相关算法提升 OCR 识别的准确率和易用性。读光 OCR 的高安全性方案和高性能算法能力，为读光 OCR 产品提供了有力的技术支撑，使得 OCR 服务和 to C 应用更具灵活性。



图 2-3 读光 OCR 技术架构

2.3 产品服务模式

读光 OCR 目前可提供三种产品服务模式：读光公有云调用服务、私有云部署服务及客户端 SDK 服务。

- 读光公有云：依托阿里公有云，提供符合 RESTful 规范的全能力调用 API 接口，同时提供移动端 SDK 及端图片预处理和优化能力，服务可用性高达 99.99%。
- 读光专有云：基于标准 Docker 容器化封装服务，能够便捷的在客户自有 GPU 服务器或私有云上部署相关 OCR 识别服务，为客户提供安全易用、稳定实时、全生命周期管理的一站式 OCR 私有化产品解决方案。
- 客户端 SDK：满足客户在各类移动 APP 及智能终端开发过程中对客户端侧实时、精准、离线和高效的 OCR SDK 集成服务需求。

| 03 产品介绍

PRODUCT DESCRIPTION

3. 产品介绍

读光 OCR 基于底层字符定位及识别、旋转校正、版面结构化、图片分类及语义理解等核心技术能力，结合具体业务应用场景，构建了包括通用文档类 OCR 识别、卡证票据类结构化识别、泛化场景下的智能 OCR 识别及可进一步满足客户高度定制化 OCR 业务需求的自定义结构化模板识别等全量全场景 OCR 识别原子能力产品，可全面满足各行业对文本数字化的智能需求。

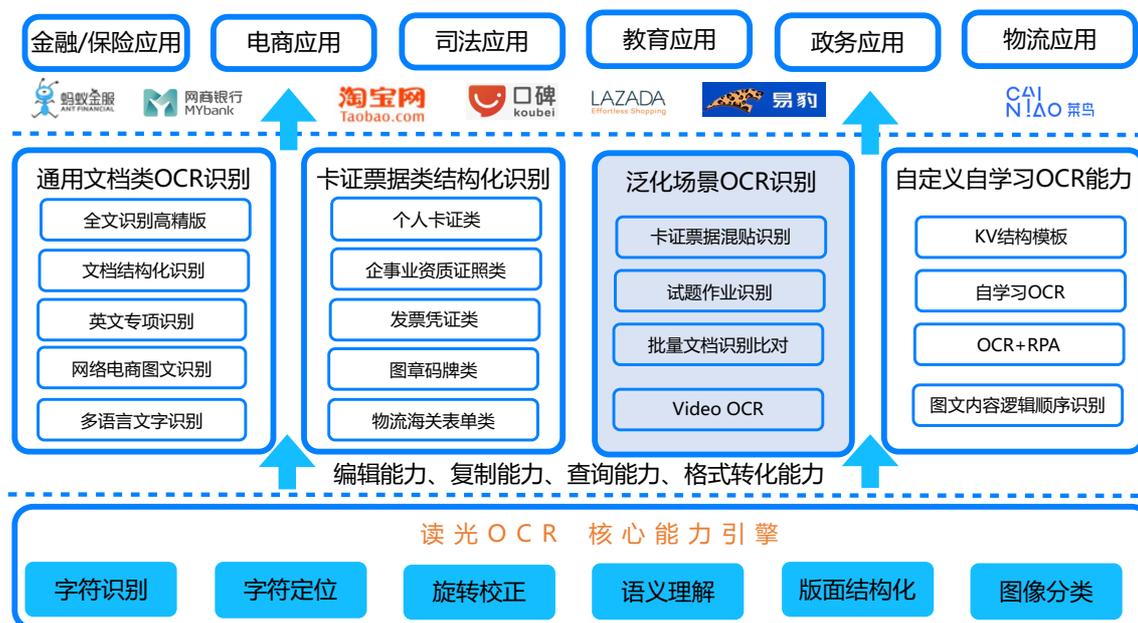


图 3-1 读光 OCR 全量原子能力产品矩阵

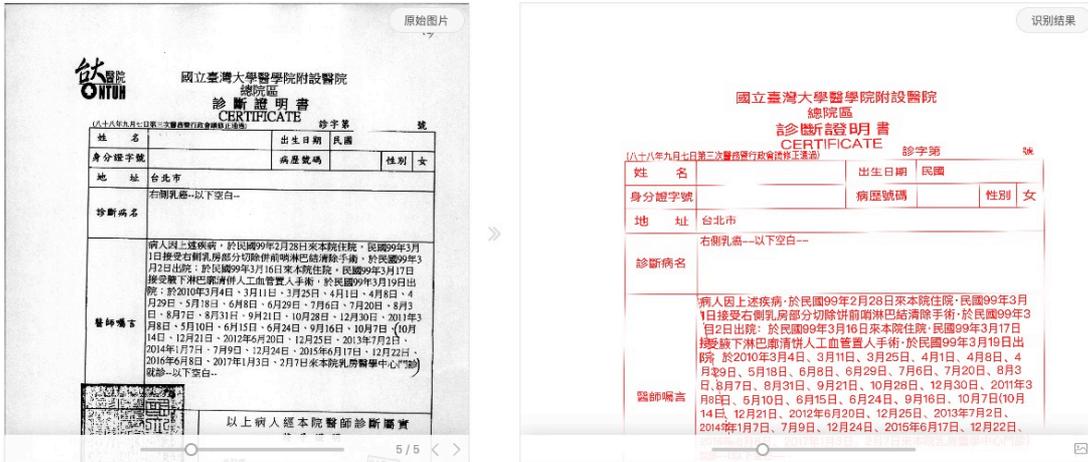
3.1 通用文档类 OCR 识别

读光通用文档类 OCR 识别产品，可将各类常见文档图片或文档扫描件中的文字信息按照文档原有的格式进行文本识别和还原。为了能够更好的还原文字信息和文档结构，读光文档识别在通用全文识别能力（文字定位、行分析、文字识别）的基础上，增加了文档结构的版面分析和文档图像处理能力，使得文档类图像也能按照结构化的方式进行文档元素提取，进一步提升文档识别的产品体验。

读光通用文档类识别具备如下功能特性：

◆ 表格自动识别

若文档中含有表格，可将表格准确识别并做提取还原，同时返回对应的表格单元格内容。



◆ 智能旋转校正

读光 OCR 可检测到文字是否偏斜，并对倾斜的文字进行矫正（0~360 度任意角度）。



◆ 单字坐标返回

读光通用文档识别不仅能够识别出文字块坐标，还能够返回出单个文字的坐标位置，使得识别更加精准，便于业务方对于 OCR 识别结果的提取和二次校验。



结果返回

```

全文还原  JSON返回
{
  "code": 200
  "data": {
    "angle": 0
    "height": 389
    "width": 593
    "prism_wordsInfo": [
      0: {
        "charInfo": [...]
        "direction": 0
        "pos": [...]
        "prob": 99
        "word": "欢迎使用读光体验馆"
      }
    ]
  }
}

```

◆ 版面智能理解（文档智能还原）

通过对文档版面的智能分析结合自然语言处理技术，能够有效地将文档识别结果还原成原始文档格式，包括对标题、段落、分栏目以及印章等内容。

原图

识别结果

印章擦除

标题提取

单行输出

表格还原

段落输出

元素还原

位置还原

内容还原

顺序还原

语义还原

◆ 印章智能检测

读光 OCR 具备全量印章智能检测和分层识别能力，包括印章定位、去除印章等，并能够实现对印章覆盖文字的修复识别。

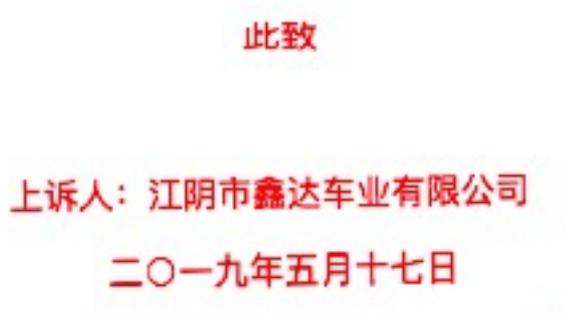
原图：



印章遮挡导致识别内容错乱：



去除印章后识别效果：



◆ 低置信度提示

为了方便用户的二次文本校对，读光 OCR 提供对置信度较低的文字进行标红处理的功能，用户可重点关注标红文字，进行人工二次校对。

◆ 双层 pdf 功能

读光 OCR 支持双层 pdf，便于用户对识别文字进行直接编辑、复制和搜索。

1.1技术方案设计

贵公司进行防腐喷涂的锅炉均为循环流化床锅炉，作为一种新型燃烧技术，由于炉型结构和设计参数等特点，在环保性能、燃烧效率、燃料适应性、高效脱硫及灰渣综合利用等方法，具有很大的优越性。然而，CFB锅炉受热面的磨损问题十分严重，贵公司决定采用超音速电弧金属喷涂技术来解决锅炉受热面磨损问题，以减少锅炉的非计划停运。我公司技术服务组根据全国各电厂的防腐防护情况，针对锅炉运行的详细状况、工艺参数以及失效因素，并依据我公司多年来从事该行业的宝贵工程经验，经我公司技术服务组专业人员的认真分析研究，并对上述部位的防护

1)问题分析:

锅炉卫燃带、炉膛出口处、焊缝处的磨损、减薄与气流中固体物料浓度、烟气速度、颗粒的特性硬度和流道几何形状等密切相关，而在CFB锅炉中，固体物料的浓度巨大，通常可达煤粉炉的几十倍到上百倍，并且烟气流速大，颗粒硬且棱角尖锐，因而在高速烟气的带动下，对CFB锅炉水冷壁等受热面部位的冲刷磨损极为严重；尤其在护墙根部水冷壁部位，由于位处密相区边缘区，不但受到严重的高速高浓度含床料、燃料气流的强烈冲刷、磨损，而且存在严重的涡流效应、切割效应和离心作用。涡流效应在炉膛四角部位，由于该处形成边壁流，物料汇集此处较多，由于固体颗粒的惯性作用，局部磨损作用尤为明显，而切割效应体现在护墙根部水冷壁处，其原因是由于防护墙的顶部提供了一个平台，当焦渣以较高的速度下降到该平台时产生反弹，其中往水冷壁管侧反弹部分，对水冷壁管就产生了严重的切割效应，离心作用是由于颗粒运行时受到烟气离心作用而引起。

其次还易受到高温氧化和硫酸盐及硫、硫化物的热腐蚀。水冷壁管具备了高温氧化和高温腐蚀条件，其烟气温度高，且是富氧燃烧，实践证明，在300℃以上，管外表温度每升高50℃，腐蚀速度增加1倍。锅炉在运行过程中受热面管表面首先发生高温氧化，表面生成Fe₂O₃，其次燃料灰中的Na₂O和K₂O与烟气中的SO₃化合生成硫酸盐，其捕捉飞灰形成结渣和流渣，此时烟气中

◆ 手写体定位及识别

基于业界领先的深度学习技术，针对图片内手写体文字“字体无规则、字迹潦草、模糊”等特点，进行识别能力增强，有效实现单字符定位、分割及精准识别。支持多场景、任意版面下整图文字中的手写中英文、字母、数字、常见字符的识别。

原图

目录	
笔记一	字音考查频率最高的227个词语·1
笔记二	字形考查频率最高的248个词语·5
笔记三	初中生应掌握的成语·8
笔记四	巧记40对中考常考易混成语·25
笔记五	病句辨析知识总结·32
笔记六	初中生应必备的标志符号用法总结·37
笔记七	中考常考古诗文默写·42
第一节	必记中国古代诗词曲重点句·42
第二节	必记文言文名篇重点句·46
第三节	必记现代文名篇重点句·49
第四节	必会30道中考名句默写题·51
笔记八	中考常考作家及作品并称·59
第一节	中考常考作家并称·59
第二节	中考常考作品并称·60
笔记九	文言文阅读知识点梳理·61
第一节	中考应掌握的通假字·61
第二节	必记70个中考常考古今异义词·66
第三节	中考常考文言文一词多义·75
第四节	文言虚词的用法及其举例·87
第五节	中考常考文言文特殊句式·93
笔记十	古诗词赏析·96
第一节	诗词中常见的意象·96

识别结果

目录	
笔记一	字音考查频率最高的227个词语·1
笔记二	字形考查频率最高的248个词语·5
笔记三	初中生应掌握的成语·8
笔记四	巧记40对中考常考易混成语·25
笔记五	病句辨析知识总结·32
笔记六	初中生应必备的标志符号用法总结·37
笔记七	初中常考古诗文默写·42
第一节	必记中国古代诗词曲重点句·42
第二节	必记文言文总篇重点句·46
第三节	必记现代文名篇重点句·49
第四节	必会30道中考名句默写题·51
笔记八	中考常考作家及作品并称·59
第一节	中考常考作家并称·59
第二节	中考常考作品并称·60
笔记九	文言文阅读知识点梳理·61
第一节	中考应掌握的通假字·61
第二节	必记70个中考常考古今异义词·66
第三节	中考常考文言文一词多义·75
第四节	文言虚词的用法及其举例·87
第五节	中考常考文言文特殊句式·93
笔记十	古诗词赏析·96
第一节	诗词中常见的意象·96

◆ 识别结果转 word/Excel 格式

读光 OCR 能够将各类图片文档输出为可编辑的 word/Excel 格式，以便于进行二次编辑和及校验。

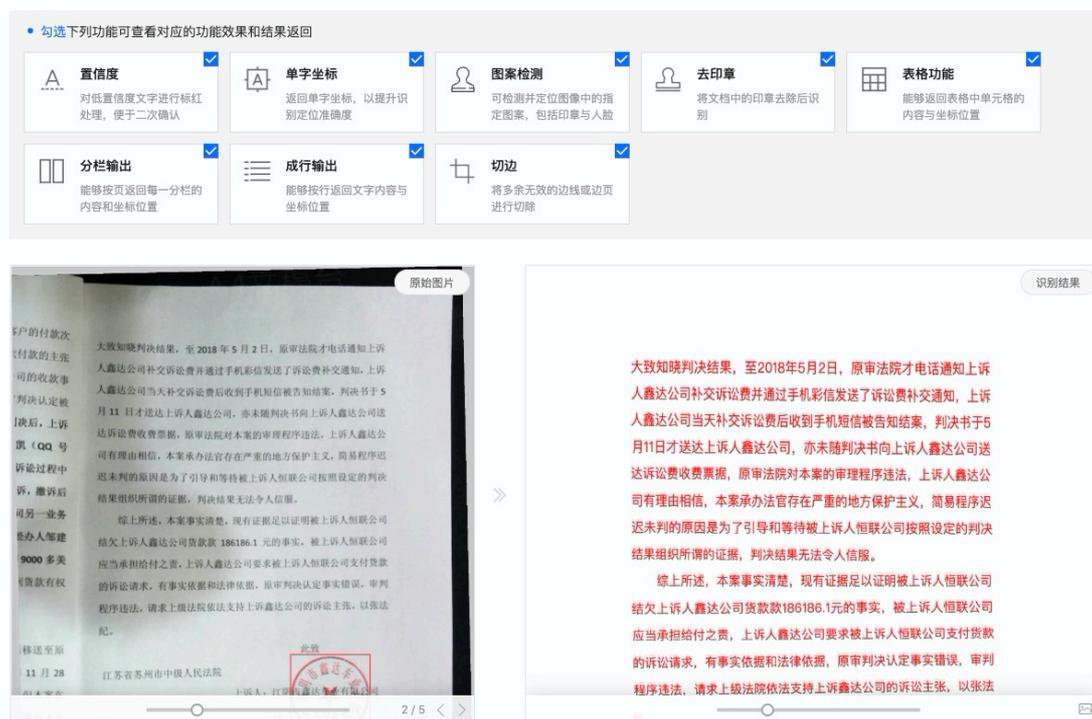


读光通用文档类 OCR 识别具体包括全文识别高精版、文档结构化识别、英文专项识别、网络电商图文识别以及多语言文字识别等原子能力产品。

3.1.1 全文识别高精版

读光 OCR 全文识别高精版提供了业内领先的基于深度学习技术的全版面文字检测和识别服务，集表格识别、旋转校正、生僻字识别、常见字符、中英文、手写体等多功能为一体，为客户提供高性价比、高精度的多场景通用文字识别能力。

效果测试

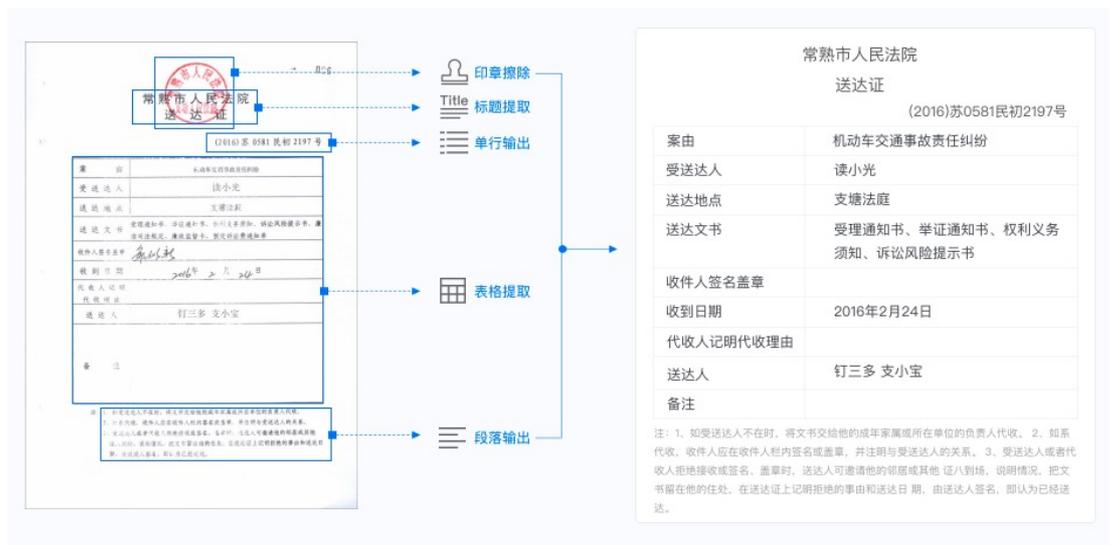


能力特性：

- ◆ **超高精度**：基于海量实战业务数据训练优化的深度学习 OCR 识别算法，文档识别准确率超过 99.7%。
- ◆ **超快识别速度**：实现 GPU 下平均 1 个字符 2ms 左右，CPU 下平均 1 个字符 6ms 左右；
- ◆ **高鲁棒性**：支持多场景、任意版面、任意背景，可容忍透视畸变、光照不均等识别环境，并可实现自动裁边、修正倾斜、表格提取、手写体识别等。
- ◆ **支持覆盖文字编辑、低置信度过滤、图案检测**等高阶功能。

3.1.2 文档结构化识别

读光 OCR 文档结构化识别重点针对各类表格、分栏目、报告表等特定格式文档，能够根据文档固有格式将文档中的文字元素（单字、文字块、行、表格等）和版面格式抽离并按顺序输出。

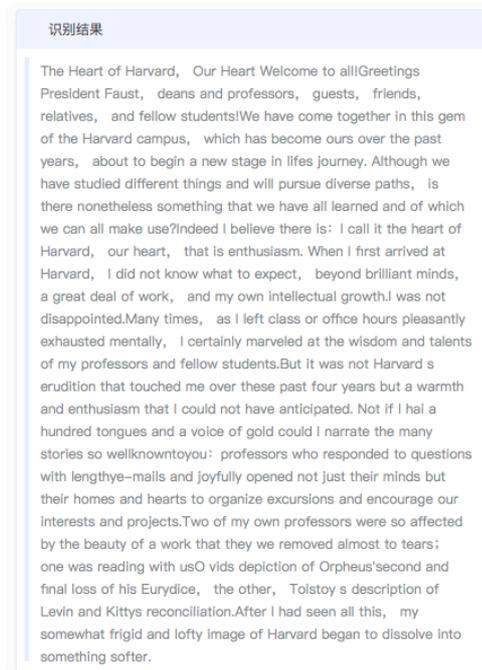
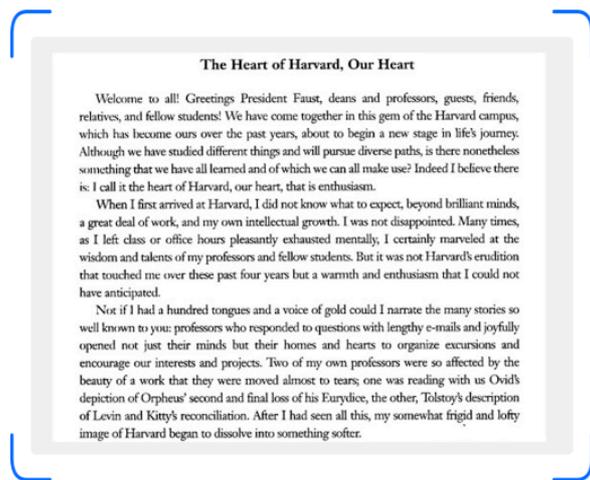


能力特性：

- ◆ 表格文档文字识别准确率超过>97.8%；
- ◆ 支持多格式版面、复杂文档背景和光照环境的精准识别；
- ◆ 可实现跨页图片分栏切边、修正倾斜、表格提取、手写体识别等优化能力；
- ◆ 针对有印章、手印的文档，可实现去印章识别。
- ◆ 支持覆盖文字编辑、低置信度过滤、图案检测等高阶功能。

3.1.3 英文专项识别

读光 OCR 英文专项识别是针对全英文图片文档场景下英文手写体及印刷体字符高效检测和识别的原子能力产品，具备英文专项识别和英文分词功能，支持旋转、表格、文字坐标等多项基础功能。

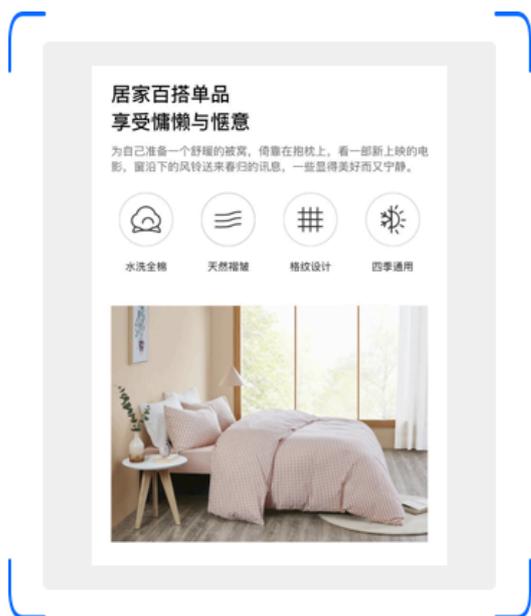


能力特性：

- ◆ 全英文文档字符识别准确率超过 >99%；
- ◆ 支持多格式版面、复杂文档背景和光照环境的精准识别；
- ◆ 可实现跨页图片分栏切边、修正倾斜、表格提取、手写体识别等优化能力。
- ◆ 支持覆盖文字编辑、低置信度过滤、图案检测等高阶功能。

3.1.4 网络电商图文识别

读光 OCR 网络电商图文识别是专门针对电商商品宣传图片、社区贴吧图片、网络 UGC 图片等网络场景下图片字符快速精准识别的原子能力产品，在违规广告识别、信息审核管理和网络安全治理等场景下具有极大应用价值。



文字块	识别内容
1	居家百搭单品
2	享受慵懒与惬意
3	为自己准备一个温暖的被窝，倚靠在抱枕上，看一部新上映的电
4	影，窗沿下的风铃送来春归的讯息，一些显得美好而又宁静。
5	水洗全棉
6	天然褶皱
7	格纹设计
8	四季通透

能力特性：

- ◆ 在常见电商宣传图片下的识别准确率高达 98.8%，超过人工录入准确率 98%；
- ◆ 针对电商海量图片内容核查场景进行特定优化，只输出文字块内容及坐标极大提升识别效率。

3.1.5 多语言文字识别

读光 OCR 多语言文字识别支持国际主流的几大语系，覆盖十余个国家的语言检测与识别，包括日语、俄语、拉丁语、泰语和韩语等，适用于国际化所需的各类图文识别与信息翻译场景。



能力特性：

- ◆ 基于达摩院强大的深度学习算法和 OCR 技术，算法整体识别准确率和召回率均超 95%；
- ◆ 支持多格式版面、复杂文档背景和光照环境的精准识别。

3.2 卡证票据类自动分类及识别

读光 OCR 卡证票据类自动分类及识别主要针对各类个人及企事业单位常用证件、执照、票据、单据及码牌等标准格式图片进行自动检测分类及关键信息结构化识别。经过多年业务技术积累，读光 OCR 已具备几十种国内外主要卡证票据分类识别能力，按业务属性可归类为个人卡证类、企事业资质证照类、发票凭证类、图章码牌类以及物流海关表单类。

3.2.1 个人卡证类识别

基于读光 OCR 的深度学习，个人证照类识别提供个人身份识别所需的身份证、名片、行驶证、驾驶证、护照、户口本、不动产权证、房产证等证件的结

结构化识别服务，且读光 OCR 可满足此十类卡证的自动分类功能，即无需提前进行卡证分类，系统可自动判断所属卡证类型并返回结构化信息。

1. 身份证识别

国内身份证 OCR 识别支持二代身份证正反面所有字段的识别，包括姓名、性别、民族、出生日期、住址、公民身份证号码、签发机关、有效期限，为用户提供最方便快捷的身份证信息录入体验。



key	value
公民身份证号码	123456202001011234
姓名	支小宝
性别	女
民族	汉
住址	山东省肥城市王庄镇花园
出生日期	2020年01月01日

能力特性：

- ◆ 基于达摩院强大的深度学习算法和 OCR 技术，各字段精度均处于业界领先水平，其中身份证号码识别准确率达到 99.9% 以上，姓名识别准确率 98% 以上。
- ◆ 成熟稳定，历经阿里巴巴集团内部海量复杂业务场景的实战应用考验；
- ◆ 适应模糊、光照不均、透视畸变、任意背景等实际应用中存在的各种情况，并可实现自动裁边、修正倾斜等。
- ◆ 实快速响应，极大提升用户体验；
- ◆ 支持身份证正反面的识别，一次扫描识别页面所有字段。

2. 银行卡识别

读光 OCR 可精准识别各类银行卡中的银行卡卡号和有效期，且支持横卡、竖卡及银行卡任意角度偏斜情况的识别与提取，为用户提供最方便快捷的身份证信息录入体验。

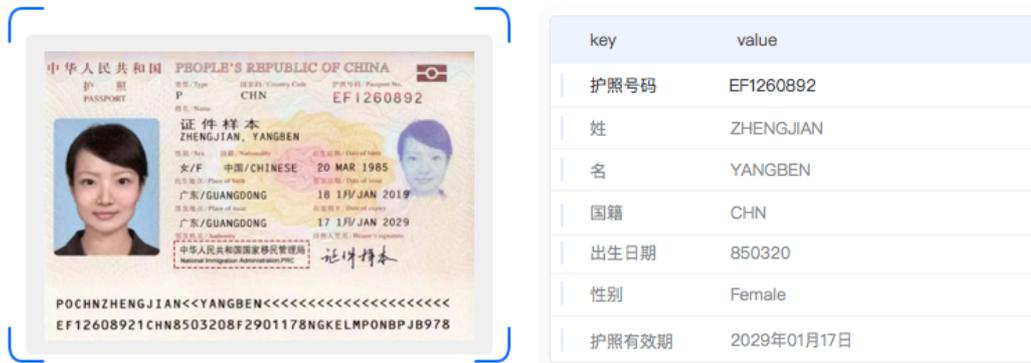


能力特性：

- ◆ 基于达摩院强大的深度学习算法和 OCR 技术，各字段精度均处于业界领先水平，其中卡号、有效期的准确率、召回率均超过 90%；
- ◆ 支持国内大多数银行，以及各种位数、凸字卡面、平面卡面等的识别；
- ◆ 成熟稳定，历经阿里巴巴集团内部海量复杂业务场景的实战应用考验；
- ◆ 适应模糊、光照不均、透视畸变、任意背景等实际应用中存在的各种情况，并可实现自动裁边、修正倾斜等。

3. 护照识别

读光 OCR 护照识别基于业界领先的深度学习技术，支持中国大陆护照、国外护照、中国香港护照及泰国护照等多个内容检测识别功能，其中中国大陆居民护照识别，已支持字段包括出生地、出生日期、国籍英文、性别英文、护照号码、有效期至、签发国、签发地英文、签发日期、类别、英文名、英文姓等；国外护照、香港护照及泰国护照识别，已支持字段包括：英文姓名、国籍、签发日期、性别、护照号码等字段，可以应用于出入境审查、国外人员身份核验等各种需要提取护照信息的场景。



能力特性：

- ◆ 基于达摩院强大的深度学习算法和 OCR 技术，各字段精度均处于业界领先水平，其中国籍、姓名、出生日期、性别、有效期至等关键信息识别准确率达到 99%；
- ◆ 已支持识别国内外多个国家的护照识别，并已支持大部分的常用字段，包括英文姓名、国籍、签发日期、性别、护照号码等字段等信息；
- ◆ 适应模糊、光照不均、透视畸变、任意背景等实际应用中存在的各种情况，并可实现自动裁边、修正倾斜等；
- ◆ 实快速响应，极大提升用户体验。

4. 驾驶证行驶证识别

读光 OCR 驾驶证/行驶证识别支持支持对证件主页和副页所有关键字段的自动定位与识别，包含证号、姓名、性别、国籍、住址、出生日期、初次领证日期、准驾车型、有效期限、发证单位、证号、姓名、档案编号等。



key	value
号牌号码	鲁R18D81
所有人	代用名
车辆类型	小型轿车
住址	山东省什么市什么区
使用性质	非营运
品牌型号	北京现代BH7200MX
车辆识别代号	LB EFA JA28X214554
发动机号码	88158976
注册日期	2015-04-27
发证日期	2015-04-27



key	value
证号	370098198902014321
姓名	李旦
性别	男
国籍	中国
住址	山东省肥城市王庄镇花园
发证单位	山东省泰安市公安局交通警察支队
出生日期	1989年2月1日
初次领证日期	2015-05-29
准驾车型	C1
有效期限	2015-05-29至2021-05-29

能力特性：

- ◆ 基于达摩院强大的深度学习算法和 OCR 技术，各字段精度均处于业界领先水平，其中行驶证的总体准确率和召回率在 93%以上，驾驶证的总体准确率和召回率在 95%以上；
- ◆ 支持行驶证主页和副页、驾驶证主页和副页所有字段的自动定位和识别；
- ◆ 适应模糊、光照不均、透视畸变、任意背景等实际应用中存在的各种情况，并可实现自动裁边、修正倾斜等；
- ◆ 实快速响应，极大提升用户体验。

5. 不动产证及房产证识别

读光 OCR 不动产证及房产证识别可准确识别房产证/不动产证中的各项关键信息，包括户主信息、房屋地址、面积大小、土地权利类型等，能够适用于全国各地的不同房产证识别，具有高鲁棒性。



能力特性：

- ◆ 基于达摩院强大的深度学习算法和 OCR 技术，各字段精度均处于业界领先水平，整体字段准确率超过 90%；
- ◆ 支持房产证及不动产证各主要字段信息的全面结构化识别，满足各种应用场景；
- ◆ 适应模糊、光照不均、透视畸变、任意背景等实际应用中存在的各种情况，并可实现自动裁边、修正倾斜等；
- ◆ 实快速响应，极大提升用户体验。

6. 名片识别

读光 OCR 支持对名片各字段的自动定位与识别，包含姓名、电话、手机号、邮箱、公司、部门、职位、网址、地址、QQ、微信及二维码等。

效果图：



姓名	张凡
职位	售前咨询顾问
公司	商汤科技
地址	上海市徐汇区龙腾大道2835号
网址	zhangfan3asensetime.com
网址	www.sensetime.com
手机	+8613911029967

能力特性：

- ◆ 基于达摩院强大的深度学习算法和 OCR 技术，各字段自动识别准确率处于业界领先水平，其中姓名、部门、电话等关键字段信息的识别准确率可达 99%以上；
- ◆ 可识别信息完备，支持姓名、电话、手机号、邮箱、公司、部门、职位、网址、地址、QQ、微信及二维码等信息字段的自动定位和识别，以及中英文、简繁体的识别；
- ◆ 适应模糊、光照不均、透视畸变、任意背景等实际应用中存在的各种情况，并可实现自动裁边、修正倾斜等；
- ◆ 实快速响应，极大提升用户体验。

7. 港澳台通行证识别

读光 OCR 支持港澳台通行证的关键字段识别，包括：类型、姓名、证件号、出生日期、英文姓名、有效期限、性别、签发地点、签发机关、机读码等。

能力特性：

- ◆ 基于达摩院强大的深度学习算法和 OCR 技术，各字段自动识别准确率处于业界领先水平，其中其中性别、英文姓名、姓名、证件号等识别准确率达到 95%以上；

- ◆ 可识别信息完备，港澳台通行证识别已支持识别大部分的常用字段，包括签签发地点、签发机关、有效期限、性别、出生日期、英文姓名、姓名、证件号等信息；
- ◆ 适应模糊、光照不均、透视畸变、任意背景等实际应用中存在的各种情况，并可实现自动裁边、修正倾斜等；
- ◆ 实快速响应，极大提升用户体验。

3.2.2 企事业资质证明类识别

读光企事业资质证明类 OCR 识别提供企事业单位、机构、公司业务开展过程中经常使用到的营业执照、商标注册证、食品经营许可证、食品生产许可证、银行开户许可证、企业证书识别、组织机构代码识别、企事业单位法人证书识别等资质证明的结构化识别服务，且读光 OCR 可满足此十多类资质证明的自动分类功能，即无需提前进行资质证明的手动分类，系统可自动判断所属类型并返回结构化信息，广泛应用于办公自动化、行业资料数字化等价值场景。

能力特性：

- ◆ 基于达摩院强大的深度学习算法和结构化模板优化技术，实现各类企事业资质证明文档的结构化输出以及关键字段信息的精准识别，其中企事业名称、法人代表等文字信息准确率超过 95%，营业执照注册号等数字信息准确率超过 98%；
- ◆ 成熟稳定，历经阿里巴巴集团内部海量复杂业务场景的实战应用考验，广受内外部客户好评；
- ◆ 适应模糊、光照不均、透视畸变、任意背景等实际应用中存在的各种情况，并可实现自动裁边、修正倾斜等。

1. 营业执照识别

读光 OCR 可快速精准的识别企业营业执照中的注册号、公司名称、地址、主体类型、法定代表人、注册资金、组成形式、成立日期、营业期限和经营范围等关键有效字段。



key	value
统一社会信用代码	913284JA8LR2048405
名称	厦门室但丁贸易有限公司
类型	法人商事主体【有限责任公司(自然人投资或控股)】
住所	厦门市集美区浔江路189
法人/负责人	江建豪
注册资本	壹佰万元整
成立日期	2017年05月04日
营业期限	自2017年05月04日至2067年05月03日
经营范围	商事主体的经营范围、经营场所、投资人信息、年报信息和监管信息等请至厦门市商事主体登记及信用信息公示平台查询。经营范围中涉及许可审批经营项目的，应在取得有关部门的许可后方可经营。

2. 银行开户许可证识别

读光 OCR 可快速精准的识别银行开户许可证中的账号、核准号、企业名称、法人姓名以及开户行等关键信息。



key	value
账号	1103857495
法定代表人	王芳
开户银行	平安银行上海长宁支行
国籍	中国
核准号	210038277101
名称	焦糖旅行社第三分社
编号	1989年2月1日
标题	开户许可证

3. 商标注册证识别

读光 OCR 可快速精准的识别商标注册证中所包含的商标名称、注册人、注册人地址以及有效期限、核定服务项目等关键有效字段信息。



key	value
商标名称	读光识字
注册人	阿里巴巴读光团队
注册人地址	天津空港物流加工区外环北路1号2-A008室
知识产权编号	3917439
商标编号	2C3917439ZC
核定服务项目	第43类
注册有效期限	自公元2006年10月07日至2016年10月06日止
局长	安青虎

4. 食品经营/生产许可证识别

读光 OCR 可快速精准的识别食品经营/生产许可证所包含经营者名称、社会信用代码、法定代表人姓名、地址、经营场所、经营项目、有效期、许可证编号等关键字段信息。



key	value
社会信用代码	91348205M2D4032852
法定代表人	陈云可
经营者名称	北京小雏菊餐饮服务有限公司
住所	北京市昌平区东小口镇龙锦三街23号
经营场所	北京市昌平区东小口镇龙锦三街23号
主体业态	食品销售经营者(贸易商)
经营项目	预包装食品销售, 含冷藏冷冻食品***
有效期至	2023年06月25日
许可证编号	JY11114071928575



key	value
社会信用代码	91348205M2D4032852
法定代表人	杨文君
生产者名称	中山市华夏中药饮片有限公司
住所	广东省中山市火炬开发区南苑
生产地址	广东省中山市火炬开发区雅柏南路3号
食品类别	蔬菜制品(食用菌制品); 水果制品(水果制品); 水产制品(非即食水产品); 其他食品(汤料)***
许可证编号	SC11644200000389
日常监督管理机构	中山市食品药品监督管理局

3.2.3 发票凭证类识别

读光发票凭证类 OCR 识别提供企事业单位、机构在业务开展过程中经常使用到的增值税发票、购车发票、火车票、出租车发票、定额发票、机票行程单、完税证明、汽车票、过路费发票、通用机打发票等多种报销凭证的结构化识别能力，且读光 OCR 支持发票凭证的自动分类功能，即无需提前进行手动分类，系统可自动判断所属类型并返回结构化信息。

能力特性：

- ◆ 基于达摩院强大的深度学习算法和结构化模板优化技术，确保各字段精度均处于业界领先水平，其中各类票据的发票代码、价税合计、合计金额、购买方识别号、开票日期等关键信息的识别准确率达到 99%以上；出租车票关键字段发票代码、发票号码识别准确率达到 95%以上；火车票关键字段姓名、出发时间、出发站、识别、座位号等识别准确率达到 96%以上。
- ◆ 成熟稳定，历经阿里巴巴集团内部海量复杂业务场景的实战应用考验，广受内外部客户好评；
- ◆ 适应模糊、光照不均、透视畸变、任意背景等实际应用中存在各种情况，并可实现自动裁边、修正倾斜等。

子图类型	增值税发票
发票代码	011001800111
发票号码	35315123
开票日期	2018年07月10日
发票金额	332
校验码	07122942791187744475
大写金额	叁佰叁拾贰圆整
发票税额	9.67
不含税金额	322.33
购买方名称	(中国)软件有限公司
购买方税号	913301007682254698
购买方地址、电话	杭州余杭区五常街道五常大道168号二...
购买方开户行及账号	招商银行杭州高新支行571906593810...
销售方名称	北京市：居酒屋
销售方税号	911101081020962778
销售方地址、电话	北京市海淀区知春路56号82112648
销售方开户行及账号	中国银行海淀支行345456019669
发票类型	电子普通发票
发票详情	货物或应税劳务、服务名称：*餐饮服务... 数量：1 单价：322.33009709 金额：322.33 税率：3% 税额：9.67

子图类型	定额发票
发票代码	111001877011
发票号码	98292817
大写金额	陆拾元整
小写金额	¥ 60.00
费用发生地	北京

1. 增值税发票识别

读光 OCR 增值税发票识别,已支持全字段识别,包括价税合计、发票代码、发票号码、合计税额、合计金额、密码区、开票日期、税率、购买方识别号、销售方识别号等。



key	value
发票代码	011001800111
发票号码	35315123
开票日期	2018年07月10日
发票金额	332
校验码	07122942791187744475
大写金额	叁佰叁拾贰圆整
发票税额	9.67
不含税金额	322.33
购买方名称	(中国)软件有限公司
购买方税号	913301007682254698
购买方地址、电话	杭州余杭区五常街道五常大道168号二层0571-85022088
购买方开户行及账号	招商银行杭州高新支行571906593810800
销售方名称	北京市: 居酒屋
销售方税号	911101081020962778
销售方地址、电话	北京市海淀区知春路56号82112648
销售方开户行及账号	中国银行海淀支行345456019669
发票类型	电子普通发票
发票详情	货物或应税劳务、服务名称: *餐饮服务*餐费 数量: 1 单价: 322.33009709 金额: 322.33 税率: 3% 税额: 9.67

2. 火车票识别

读光 OCR 火车票识别,已支持火车票全部字段的识别,包括:编号、席别、票价、姓名、座位号、出发时间、出发站、到达站、车次等。



key	value
票号	Z26E006360
出发站	杭州东站
到达站	上海虹桥站
开车时间	2016年11月16日12:02开
票价	73.0
座位类型	二等座
旅客信息	4201061982****0457许飞
座位号	05车14C号
车次	G7540

3. 机动车统一销售发票识别

读光 OCR 机动车统一销售发票识别, 支持购车发票及二手车购车发票的识别, 可识别发票中的全部字段, 包括: 机打号码、机打代码、发票号码、发票代码、身份证号、发票日期、车辆识别代号、销货单位名称、购买方名称、不含税价、税率、价税合计、价税合计(小写)、增值税税额等。



key	value
发票代码	137111725306
发票号码	00049513
开票日期	2017-12-31
发票金额	76900.00
增值税税率或征收税	17%
增值税税额	11173.50
不含税金额	65726.50
合格证号	WAA461711294258
购买方名称	谢玉
购买方身份证号/代码	2326231970042329
车辆类型	轿车
厂牌型号	森雅牌CA7167R
销售单位名称	日照汽车销售有限公司
销售方税号	913711006693468718
销售方地址	日照市南昌路26号(204国道东侧、锻管厂南侧)
销售方开户行	日照银行开发区支行
销售方单位账号	371300201020005015
主管税务机关	日照经济技术开发区国家税务局税源管理一科
主管税务代码	13711013800

4. 出租车机打发票识别

读光 OCR 出租车机打发票识别, 已支持的关键字段包括, 发票号码、发票代码、金额、日期等识别。



3.3 泛化场景 OCR 识别

读光泛化场景 OCR 识别系列为读光特有的创新技术产品。该产品以读光 OCR 全文高精度识别、格式提取和自动分类技术为基础，实现从定式结构到开放结构的图文识别和语义理解功能，能够有效的解决用户在更为复杂场景下的文字内容提取、结构化还原与信息理解等问题，帮助用户解决复杂图文识别场景下的智能分类、智能提取、批量文档识别及比对、视频文字提取、信息检索等高价值 OCR 需求。



3.3.1 卡证票据混帖识别

读光卡证票据混帖识别，基于对客户实际业务流应用场景中对 OCR 智能化、语义化、泛在化识别能力的需求，有效整合读光通用高精度文本识别、结构理解、检测分类及自然语义理解等关键技术，有效实现在各类财务票据混帖、个人证照混帖等场景下的自动分类、关键有效信息精准识别和结构化提取，极大提升 OCR 能力在财务自动化、办公自动化及各类行业应用中的技术效能。



3.3.2 试题作业识别

读光试题作业 OCR 识别产品能力，主要针对教育应用场景中对试题题目、数学公式、速算题目等信息的智能化识别需求，通过对通用 OCR 高精度识别能力的教育场景迭代优化，为用户提供数学试题图片中题目文本及数学公式的识别、速算题目文字的检测和识别等服务，并返回题目框位置与内容，为智慧教学场景下的拍照搜题、板书识别、自动阅卷等应用提供关键基石技术能力。

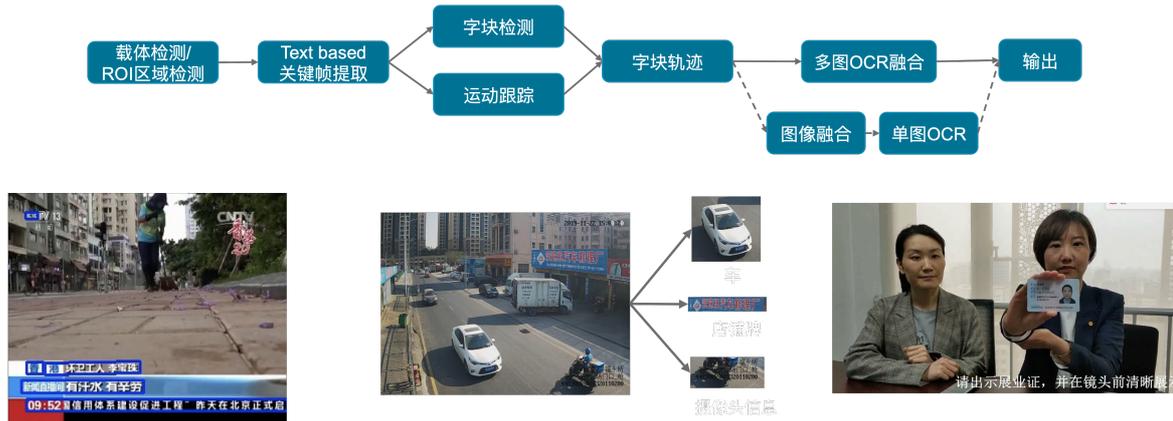
3.3.3 批量文档识别及比对

读光批量文档识别及比对产品重点针对各行业办公自动化流程中，对纸质合同、文书、协议等版本修订审查场景下，关键有效信息的批量识别和差异化比对需求，实现对上传的文档进行批量识别、自动分类、文字差异比对，降低人工审核成本，提升流程效率。用户需要上传标准文档和比对文档，系统将以标准文档为准，寻找比对文档的差异化特征。

The screenshot displays a document comparison tool. On the left, there are two panels: '标准文档' (Standard Document) and '对比文档' (Comparison Document). Each panel contains a list of numbered clauses (e.g., 3.2, 3.3, 3.4, 4., 5., 5.1, 5.2, 5.3, 5.4, 5.5, 6., 6.1, 6.2) with corresponding text. A double-headed arrow indicates the comparison between the two documents. On the right, a '差异列表' (Difference List) shows the results of the comparison, listing differences such as '删除' (Delete), '增加' (Add), and '修改' (Modify) with corresponding text snippets.

3.3.4 Video OCR 识别

Video OCR 识别能力是读光产品的特色原子能力，其以视频作为输入对象，在静态图片 OCR 的基础上加入跟踪轨迹提取、多帧结果融合等技术，实现对视频中富文本物体的跟踪检测识别，主要应用场景包括长视频字幕识别、直播短视频弹幕识别、自然场景监控字符识别以及视频证件文件录入等，能够极大提升 OCR 在动态视频中的识别准确率以及有效关键信息的价值挖掘。



3.4 自定义自学习 OCR 识别

读光 OCR 除通用文档类、卡证票据等结构化识别以及泛化场景下的 OCR 识别能力外，还重点针对不同行业客户的各类个性化定制需求，设计了包括 KV 结构模板、自学习 OCR 平台、OCR+RPA 以及图文内容逻辑顺序识别等高阶原子能力产品，让 OCR 能力真正能够深度适配各行业业务流场景，驱动技术创造生产力价值。

3.4.1 KV 结构模板

KV 结构模板是阿里读光 OCR 提供的一款用于对关键要素进行定向提取的自定义模板产品，主要针对文档、表单、票据、卡证等类型的图片。用户可以使用 KV 结构模板对图片进行模板配置和数据标注，经过训练、评测和发布后，可获得对应的模板 ID，并使用 API 接口批量调用同类图片来进行关键要素的 key-value 键值的结构化输出。

读光KV结构模板可配置卡证



配置读光 KV 结构模板具体流程：

1. **模板配置**。通过对一张具有代表性的图片进行 key/value 的规则属性配置，形成初始化的模板；作为配置的图片最好选择清晰、无遮挡、无反光的图片更有利于模板的配置。



Step.1 模板配置

通过对一张具有代表性的图片对key/value进行规则配置以形成初始化的模板

2. **创建数据集**。通过标注相同类型的图片进行模板的数据训练，因此第二步需要创建数据集，并上传至少 30 张同类图片作为训练集，上传至少 10 张图片进行测试集。上传图片后，分别对上传的图片进行标注，通过标注训练图集上的 value 值，形成一批可用的训练数据。这些已标注好的数据会按比例分成训练数据和评测数据。



Step.2 创建数据集

通过标注同类型的图片进行模板的数据训练，至少需要进行30张以上的数据标注

3. **模板训练**。模板训练是基于深度学习算法进行的机器自学习过程，系统会根据标注数据自动强化模板的识别率，提高模板的容错性。训练集的数据越多

对模板训练的效果就越佳。训练是个反复的过程，可以分多个批次进行训练，每个批次建议上传更多的图片进行标注以达到自学习的目的。



Step.3 模板训练

模板训练是基于深度学习算法进行的机器学习过程，系统会根据标注数据自动强化模板的识别率，提高模板的容错性

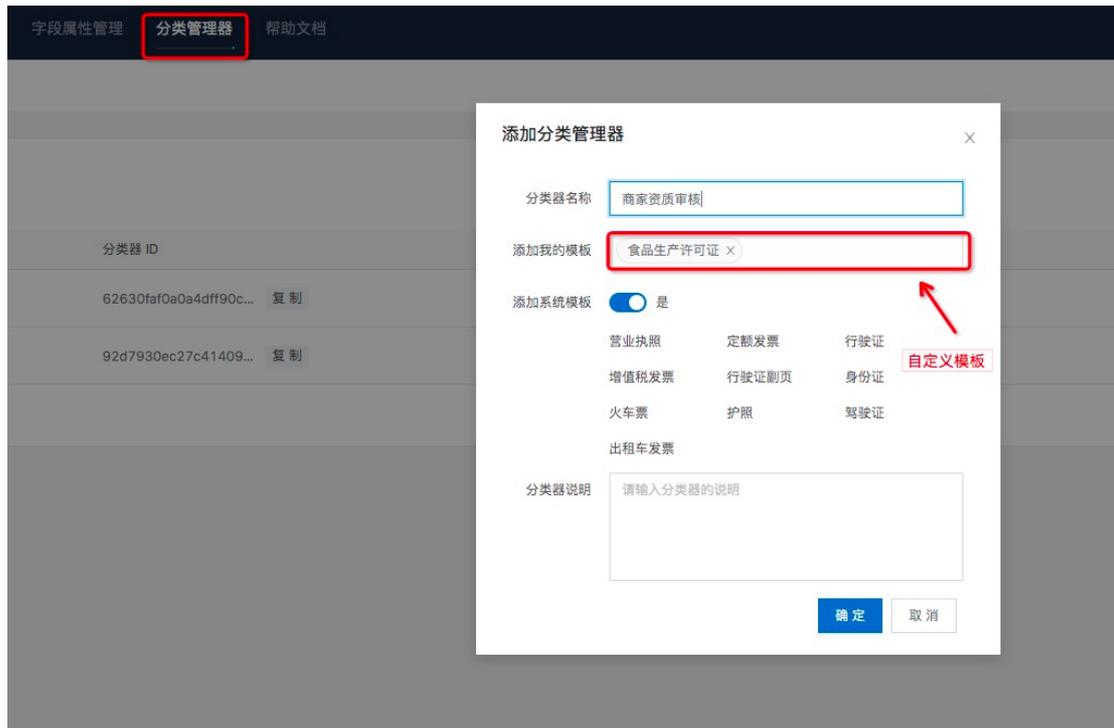
4. **模板评测与发布**。模板评测是用于对经过训练的模板进行效果评测打分的步骤。在每个训练批次中可实时查看效果评估对比。评测指标分为 OCR 识别率和 KV 字段识别率，OCR 识别率是文字本身的识别效果，KV 字段识别率是指模板在评测过程中所有字段的平均识别结构。通常情况下，KV 字段识别率/OCR 字段识别率 > 90%，是较为可用的状态，当然用户可根据自己的标准进行效果评估。根据评测结果，用户可以选择最为理想的训练版本作为发布版本，并发布上线。发布成功后即可获得模板 ID，使用对应的 API 接口，即可调用该模板进行批量识别。



Step.4 模板评测与发布

根据评测结果，用户可以选择最为理想的训练版本作为发布版本，并发布上线。发布后即可使用对应的API接口，进行批量识别

5. **模板分类管理。** 若您对多个模板有自动分类聚合的需求，可以创建分类管理器，添加已经发布的模板，系统可进行自动判断识别属于哪个模板，并调用对应的结构化接口进行识别。同时，你也可以选择系统模板，当遇到命中系统模板的卡证时，无需配置即可享受系统模板的结构化识别结果。



3.4.2 自学习 OCR 平台

读光自学习 OCR 训练平台 (OCRMaster) 以打造解决用户端到端个性化 OCR 识别需求为目的。用户可以在 OCRMaster 上基于读光 OCR 丰富的基础原子算法和深度学习训练系统进行自有场景下相关卡证、票据、表单的结构化识别，通过 OCRMaster 打造出自定义的模板产品而不需要通过定制开发等高成本手段来实现个性化需求。同时，OCRMaster 不仅可针对结构化需求进行自定义，更强大的是可甄别出特定场景下识别效果不佳的图片，并针对这些 badcase 也进行个性化训练，从而打造出从检测、识别到结构化都完全定制的自主端到端 OCR 模板产品。

OCRMaster 具有低成本、高效率、高识别率、深度定制的产品特征，用户用 50 张基础图片即可训练出一个高可用的结构化模板，用户可实时参与到整个模板研发过程中，通过人机交互对自定义模板进行实时调优，可视化见证识别效果的提升，自主可控性强。产品链路流畅透明，技术小白用户也能轻松上手，四步即可完成模板创建。

OCRMaster 不仅仅是一个研发工具，也是针对 ocr 的运营平台。在研发侧，用户可通过 OCRMaste 进行结构化模板的配置、图片分类定义、模板评测与发布；在运营侧，用户可对生产调用情况进行数据看板管理，对 badcase 设置回流，并将回流数据反哺至模型中，以此形成自循环的 ocr 生产系统。



3.4.3 图文内容逻辑顺序识别

读光 OCR 结合 NLP 自然语言理解、图文格式识别等技术，能够有效地根据图文的具体内容、形式、结构实现文字内容按一定的逻辑顺序识别输出，以保证在各类泛在场景下文本识别的逻辑可用性，如电商图文内容审核、论文数字化扫描、教育场试题识别等场景。

效果测试



4. 典型应用场景

4.1 办公自动化场景

读光 OCR 除了支持阿里集团内部各类财务、流程、资质审核等业务场景外，已在法务、金融、海关、财务、咨询、医疗、政务等多行业的办公自动化流程中进行了能力集成和业务流打通。通过在办公业务流程中集成各类 OCR 识别能力，能够整体提升线下人工录入审核效率，节约 80% 的用户时间成本，极大提升企业生产效率，降低人力成本，为企业赋能。



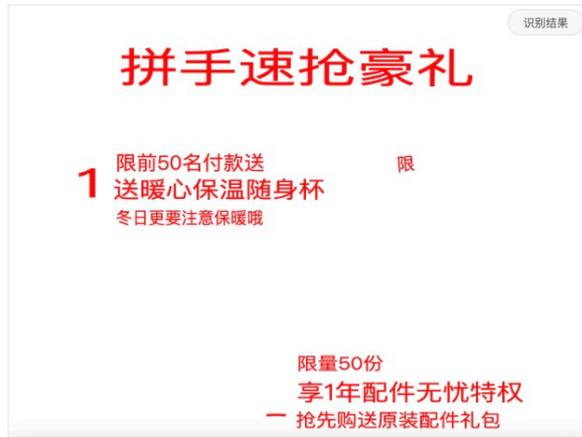
4.2 电商治理场景

读光 OCR 产品能力在电商治理场景有非常广泛的应用，覆盖了从电商图文的合规性审查、内容知识挖掘、到运营效率提升等各个业务环节。

1. 电商图文内容合规性审查：基于 OCR 图文逻辑识别能力，能够精准高效的将海量电商宣传图片中各式各样的宣传标语和商品介绍信息解析为结构化文本，进而配合电商合规管理策略进行违规商品宣传图文的识别和过滤。
2. 商品内容知识挖掘：阿里电商体系庞大，海量的商品宣传信息中，80%以上是图片格式，多年以来对于图片宣传内容平台运营方都很难有效地挖掘其

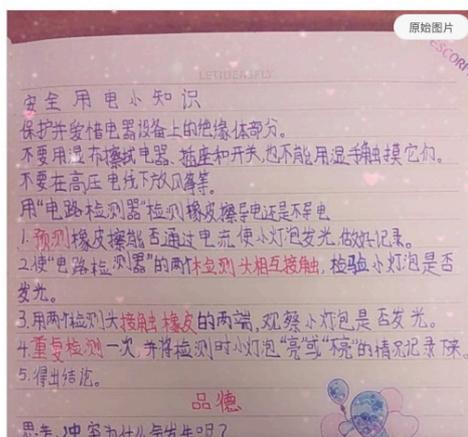
内容的价值。自读光 OCR 应用于内容知识挖掘以来，已经计算了千亿量级的详情图数据，产出了 10 倍于商品现有的文本信息，为电商平台的大数据挖掘和精准运营提供了基石能力。

- 运营效率提升：读光 OCR 各类卡证票据识别能力，能够加大提升电商平台日常运营活动中对商家资质、申请单、交易票据等的自动化内容识别和审核，能够极大地代替人工运营，有效提升运营效率，节省运营成本。



4.3 智慧教学场景

读光 OCR 在智慧教学场景下有极大地应用价值，从在线教育中的扫题搜索、公式/题目识别等应用，到线下教学场景中的板书识别、自动阅卷等应用，OCR 识别助力智慧教学触达价值落地红线。



4.4 RPA 场景

OCR+NLP 是 RPA (机器人流程自动化) 的技术基石 , RPA 在解决不同业务系统自动化处理过程中 , 需要对财务、金融、税务、人力资源、信息技术、保险、客服等业务数据中相关文档通过 OCR 识别完成数字化 , 进而实现自动化。

广告复核
效率对比

人工操作
00:00:02:24

机器人操作
00:00:02:24

▶ 打开待复核的广告Excel文件

▶ 一键运行，机器人自动操作

5. 产品规格

5.1 公有云产品规格

5.2.1 性能指标

1. 公有云单用户默认 10QPS , 可根据用户需求弹性扩容 ;
2. 服务调用成功率超 99.96% ;
3. 服务平均响应时间在 400ms 左右 ;

5.2.2 订购及授权

读光公有云通过阿里云 API 网关提供 OCR 服务，用户可以直接通过阿里云云市场购买使用。（订购地址：<https://market.aliyun.com/>）

公有云服务具备安全、购买使用简单、接入成本低等特点：

1. 安全：用户上传的所有图片读光不会进行持久化保存（阅后即焚），保证用户图片的私密性。服务请求协议使用 https，保证请求过程中的数据和图片的安全性；
2. 购买使用简单：直接在阿里云云市场上通过支付宝等方式购买流量包，购买后即可快速在线调试使用；

5.2.3 调用方法说明

请阅读阿里云官网 API 网关文档：

<https://duguang.aliyun.com/document/%E9%AB%98%E7%B2%BE%E7%89%88.html>

使用方法相关问题请联系钉钉支持群，群名称：API 网关-客户服务群，群号：21734896

5.2 私有化产品规格

读光专有云服务是以 Docker 镜像的方式为客户提供各类标准化或定制化 OCR 产品，部署运行在用户自身网络环境中的 GPU/CPU 服务器上。

读光专有云服务主要针对对网络环境有私密性要求，对图片网络传输延迟有更高要求，拥有或可以购买 GPU/CPU 机器的客户。

专有云服务具备服务定制化，完美的网络和图片私密性，更低的网络延迟等特点：

1. 服务定制化：相对于公有云提供的通用化服务，专有云服务可以完全按照用户的图片内容和结构化需求进行定制，来满足用户对各种各样图片的识别需求，并且根据具体的图片内容对识别结果进行优化；
2. 完美的网络私密性：读光运行的网络环境可以是用户的专有云或者局域网，不需要与互联网进行任何交互，所以具备完美的网络和图片的私密

性；

3. 更低的网络延迟：因为在用户的专有云或者局域网内，那么图片的网络延迟势必会更低，请求的 RT 时间也会更低。

5.2.1 性能指标及硬件要求

GPU 型号	单卡 Tesla P4	单卡 Tesla T4
GPU 显存	8G	12G
CPU 核数	16 核	16 核
内存大小	32G 最小，推荐 64G	
操作系统	CentOS 7.5	
QPS 参考 (文档识别类)	5 QPS	10 QPS

详细性能指标和精度指标，可联系阿里云进行商洽或 poc 测试

5.2.2 服务部署说明

为了屏蔽各机器环境不同导致的限制问题，读光 OCR 的私有化部署以 docker 化的部署方式为主，通过 license 硬件锁或 license 软件锁的方式提供双重安全验证。用户采购读光私有云服务后，会由读光团队寄送加密的硬件锁，用户需要自行将硬件锁插入对应的运行服务器上，并根据读光部署文档进行安装操作即可使用。

具体服务部署流程总结如下：

1. 确认 GPU 机器准备就绪；
2. 安装运行环境
 - a) 确认机器显卡驱动已正确安装；
 - b) 安装 cuda8.0 (选择性安装)；
 - c) 安装 docker
 - d) 安装 nvidia-docker
 - e) 安装 license 驱动
3. 安装 license 硬件锁
4. 导入本地镜像
5. 运行镜像，启动容器
6. 启动读光服务
7. 验证读光服务接口协议
8. 测试服务是否启动成果
9. 如何停止读光服务

部署文档参考《读光产品部署文档》

5.2.3 订购及授权

1. 产品及服务订购方式

读光 OCR 针对专有云形态提供两种采购方式，按照 QPS 购买或者按整套服务购买，详细定制方式和价格咨询，请与阿里云客服咨询，或直接邮件联系我们：duguang_support@list.alibaba-inc.com

- 1) 以路数售卖（1 路 = 1qps）

采购方式：直接选择所需产品功能服务

- 2) 以整套售卖（1 套是指按 1 张 p4 单卡或者 1 张 p100 单卡的方式进行售卖）

采购方式：直接选择所需产品功能服务

2. 产品授权方式

读光 OCR 私有化服务通过 license 软件锁的方式提供安全验证及授权。