金融行业结构化信息抽取
电力行业结构化信息抽取
常见证件票据结构化信息抽取
其他场景结构化信息抽取
文本检测
文本识别
NER
泛安全行业结构化信息抽取
标题检测
图片检测
布局分析
表格检测
图像去噪
图像增强
图像纠偏
印章检测
半自动学习平台
自动识别
人工更正
人工审核
模型训练
模型管理
模型发布
权限管理
数据管理
资源管理
基础模型

通用文档识别

识别图片中的文字，返回包括字位置信息。同时能够识别表格、特殊字符等。

卡证识别

支持对二代居民身份证正反面所有8个字段进行结构化识别，包括姓名、性别、民族、出生日期、住址、身份证号、签发机关、有效期限。支持识别社会保障卡号、姓名、性别、出生日期、银行卡号、有效期限等。

表格识别

支持识别图片/PDF格式文档中的表格内容，覆盖各类表格样式，包括常规有线表格、无线表格、含合并单元格表格。同时，支持多表格内容识别。

图片文档还原

支持各种图片格式，及PDF、OFD文件识别。针对办公、政法、票据等常用场景的文档进行表格、文本等元素的识别和提取，并进行全文排版，输出完整的文件。

文档矫正

对图片中的文件、卡证、票据等内容进行四角点检测定位，提取主体内容并对其进行矫正。

复杂环境下的高识别度

拓尔思OCR云平台支持横向、竖向拍摄，适应透视畸变、光照不均、部分遮挡、画质模糊、加盖公章、水印以及文字宽度、粗细、颜色、下划线高度不一致等情况，大幅提升文字检测的准确度，复杂环境下具有高可用性。

命名实体识别的高匹配度

对于字段信息较为固定的文件，如身份证、社保卡、医保卡、增值发票等，可以使用模版匹配等技术。对于版式不固定的文件，例如简历、合同、办公文件等文档，可采用实体识别解决关键信息提取困难的问题，能够实现在不同场景中精准识别不同实体类型。

具备自训练的多样化配置

用户结合实际场景需求，可采用全新自主训练模式，快速定制识别表单/卡证/票据。用户还可在平台自主搭建独有的训练模型，进行独有模版文档的识别和结构化数据提取，满足多样化的场景应用需要，享受企业费用成本大幅降低、工作效率切实提高等便利。

用户身份认证

身份识别主要包括身份证识别、社保卡、医保卡、银行卡识别、行驶证/驾驶证识别。可广泛应用在政府、银行、保险、医院、证券行业等需要验证用户身份的场景中，帮助减少用户的身份信息输入，提升认证效率，提高用户体验。

票据单据识别

用于企业开具的各类票据识别，通过OCR技术提取各种纸质票据的号码、抬头、金额、票据代码等关键信息，支持数据的录入和自动核对，减少人工校验的成本。

印刷图像转文本

大量纸质印刷作品扫描电子文档后，通过OCR 技术转换成可编辑的文本，实现文档高度还原，为文字排版、信息检索提供极大的便利。

文本审核与管理

在政府、银行、电力、公安等行业中，通用OCR可以智能识别商业合同、电子卷宗、招投标文档等的文字内容，并根据不同业务场景的校验点进行文档合规性审核。也可以校对广告图中的文字内容，过滤广告中的不良信息，为企业降低业务风险。