

第四代英特尔® 至强® 可扩展处理器助力 亚信科技提升 OCR 应用性能

挑战

通过与深度学习等 AI 技术进行深度融合，RPA 系统将能够应对更为广泛的自动化处理需求，获得业务逻辑的理解能力，从基于固定规则的自动化流程平台向智能流程自动化平台演进，提升业务运营与组织管理协同效能。但与此同时，基于深度学习的 OCR 等应用也在算法、算力等层面带来了巨大的挑战：

深度学习模型推理带来巨大性能开销

在财务报销、智能质检、数据智能化采集等 RPA 的典型应用场景中，基于深度学习的 OCR 应用常常需要以模型推理的方式，对于海量的单据、文档等图片进行处理，产生较大的 AI 算法开销。

深度学习应用的 TCO 压力大

投资回报率是用户在部署 OCR 应用时重要的考虑因素之一。用户通常希望在保证目标功能、性能的前提下，尽量降低平台的 TCO。因此，尽管独立加速器能够满足基于深度学习的 OCR 应用的算力需求，但相对较高的成本也使得大量用户选择基于 CPU 的推理方案。

难以实现应用的异构化扩展

出于性能、经济性、灵活性等方面的考虑，客户最终部署的硬件常常千差万别。因此，AI 算法需要在异构化平台上进行移植，同时保证在性能方面足以满足要求，这将会导致企业将大量的成本与资金耗费在应用开发、性能优化等工作之中。

解决方案概述

为了将人力资源从单调繁琐的工作中解放出来，提升业务效率并挖掘数据价值，越来越多的企业强化了机器人流程自动化 (RPA) 方案的部署。RPA 方案能够通过光学字符识别 (OCR)，以及自然语言处理 (NLP)、语音识别、智能决策等人工智能 (AI) 技术的应用，增强在各种复杂环境与任务下的自动化处理能力。通过部署 RPA 方案，企业将能够提升业务全流程、全要素资源数字化水平，推动工作流程的自动化、智能化，但与此同时，RPA 方案中的智能 OCR 识别等负载会带来较大的 AI 算力开销，对于基础设施平台带来较大的性能与总体拥有成本 (TCO) 压力。

作为大型企业数字化转型的赋能者，亚信科技致力于推动 RPA 方案的创新与优化。利用英特尔第四代英特尔® 至强® 可扩展处理器的强大算力，以及处理器内置的英特尔® 高级矩阵扩展 (英特尔® AMX) 等高级硬件能力，亚信科技显著提升了 RPA 方案中 OCR 算法推理等关键负载的性能表现，助力用户更高效地利用 RPA 技术的价值，实现各个业务系统的数据共享与流程自动办理，提升工作效率。

亚信科技 AIRPA 平台 集成智能 OCR 应用

亚信科技机器人流程自动化平台 (AISWare AIRPA) 是通过模拟并增强人类与计算机的交互过程，实现工作流程自动化的平台。它可以模拟人类在电脑上进行操作，尤其擅长点击、复制、粘贴、输入等有固定规则且重复性较高的办公内容，而且比人类更高效、更敏捷、更准确。AISWare AIRPA 在 RPA 功能的基础上，提供了强大的智能 OCR 能力支持。

“流程自动化已经成为企业数字化转型过战略的关键部分，亚信科技智能 RPA 平台致力于数字员工的快速构建，提供简单、高效、灵活、智能的机器人流程自动化解决方案。通过采用最新的第四代英特尔® 至强® 可扩展处理器，我们进一步提升了 RPA 平台中智能 OCR 等应用的推理性能表现，帮助用户实现从流程自动化向流程智能化的转变。”

赵一鸣

亚信科技平台产品研发中心总经理



intel

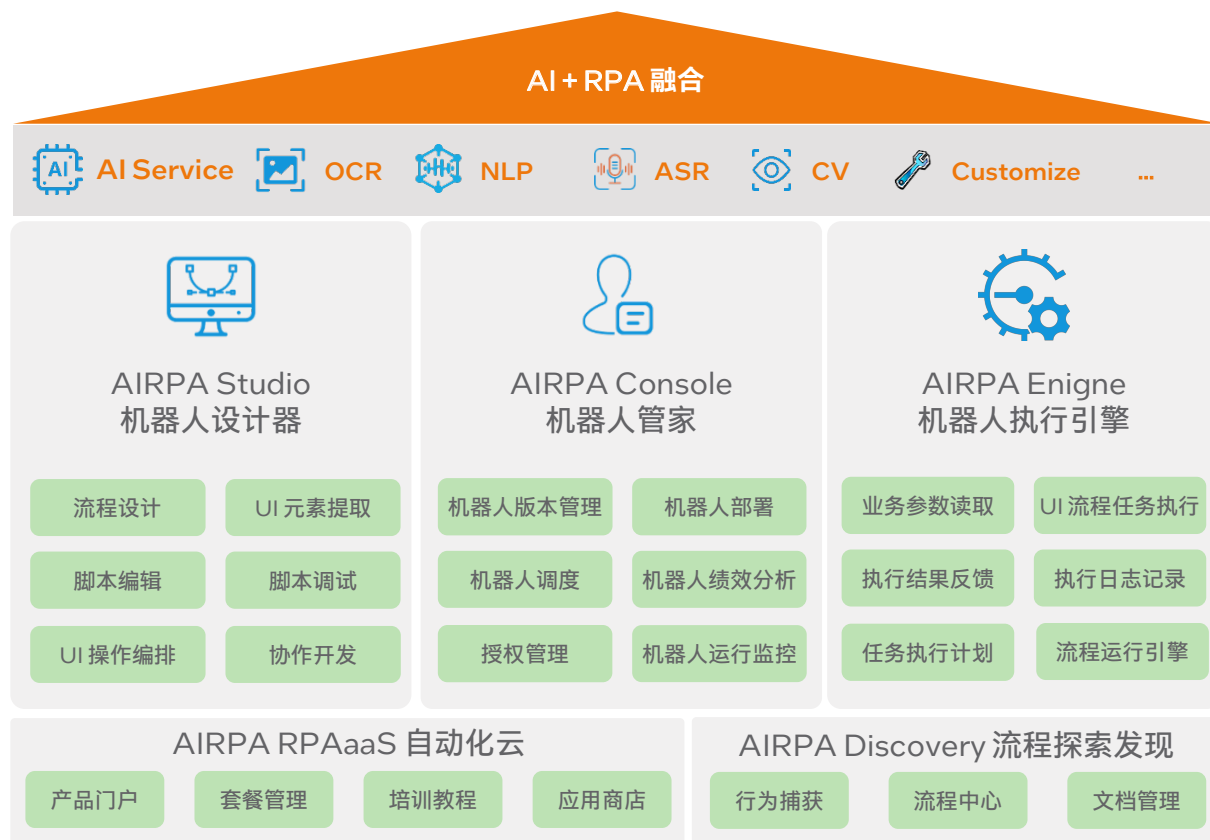


图 1: 亚信科技机器人流程自动化平台组件

该智能 OCR 应用支持用户自动化的实现票据、发票、文档等信息的录入，减少高重复性与低效率的事务性工作，其具备识别速度快、识别率高特点，极大减少人工录入工作量，且数据更加容易保存，既安全又方便，大大减少了票据存档管理费用和纸张空间。

结合 OCR 等软件流程自动化能力与亚信智慧大脑，AISWare AIRPA 面向全域提供跨平台、跨系统、跨应用的流程集成能力，通过数据智能方式实现流程数据异常自动发现、流程变化自动识别、流程知识自动积累，持续探究企业内部效率瓶颈和价值空地，推动流程智慧化重塑，提质增效、敏捷创新。

AISWare AIRPA 架构如图 1 所示，其包括 AIRPA Discovery 流程探索发现、AIRPA Studio 机器人设计器、AIRPAEngine 机器人执行引擎、AIRPARPaaS 自动化云、AI Service 机器人智能服务等组件。

采用第四代英特尔® 至强® 可扩展处理器提升 OCR 推理性能

亚信科技 AISWare AIRPA 中集成的智能 OCR 应用能够将各类常见文档图片按照文档原有的格式进行文本识别和还原，支持票据识别、卡证识别、印章识别、通用识别、手写识别、表格识别、验证码识别、智能文档处理等场景应用。为了加速智能 OCR 应用推理性能，亚信科技采用了第四代英特尔® 至强® 可扩展处理器。

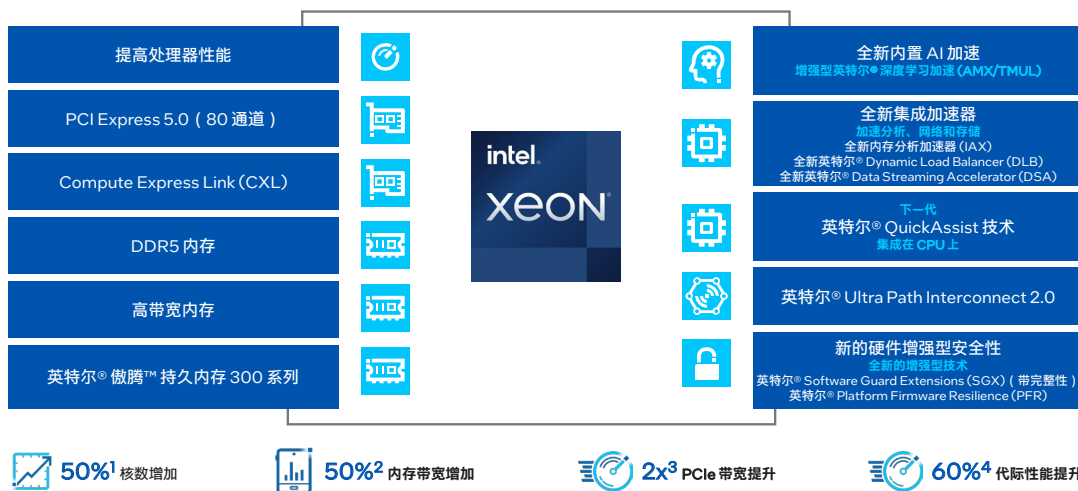
第四代英特尔® 至强® 可扩展处理器通过创新架构增加了每个时钟周期的指令，每个插槽多达 60 个核心，支持 8 通道 DDR5 内存，实现了 50%² 的内存带宽提升，并通过每 PCIe 5.0 (80 个通道) 实现了 2 倍³ 的 PCIe 带宽提升，整体可实现 60%⁴ 的代际性能提升。第四代英特尔® 至强® 可扩展处理器提供了现代性能和安全性，可根据用户的业务需求进行扩展。借助内置的加速器，用户可以在 AI、分析、云和微服务、网络、数据库、存储等类型的工作负载中获得优化的性能。通过与强大的生态系统

相结合，第四代英特尔® 至强® 可扩展处理器能够帮助用户构建更加高效、安全的基础设施。

第四代英特尔® 至强® 可扩展处理器在 AI 性能上更进一步。该处理器内置了创新的英特尔® AMX 加速引擎。英特尔® AMX 针对广泛的硬件和软件优化，通过提供矩阵类型的运算，显著增加了人工智能应用程序的每时钟指令数 (IPC)，可为 AI 工作负载中的训练和推理上提供显著的性能提升。

在智能 OCR 应用中，亚信科技通过英特尔® AMX 支持实现从 FP32 到 INT8/BF16 的量化，从而在可接受的精度损失下，增加吞吐量并加速推理。

为了验证优化后的性能提升，亚信科技对比了英特尔® 至强® 铂金 8380 处理器 (第三代英特尔® 至强® 可扩展处理器，未采用英特尔® AMX 加速) 与英特尔® 至强® 铂金 8480+ 处理器 (第四代英特尔® 至强® 可扩展处理器，采用英特尔® AMX 将模型量化至 INT8/BF16) 的性能。测试数据如图 2 所示，第四代英特尔® 至强® 可扩展处理器能够将 OCR 算法推理性能提升 3.38 倍⁵。



^{1,2,3,4} 实际性能受使用情况、配置和其他因素的差异影响。更多信息请见 www.Intel.com/PerformanceIndex

⁵ 截止 2022 年 8 月由英特尔开展的测试。测试配置 1: 单节点，双路英特尔® 至强® 铂金 8380 处理器，40 核，启用超线程，启用睿频加速技术，256 GB 总内存 (16 插槽/32 GB/3200 MHz)，SE5C620.86B.01.01.0005.2202160810，0xd000375，Ubuntu 22.04.1 LTS，5.19.0-051900-generic，gcc 11.2，AsiaInfo OCR v1，OpenVINO 2022.2.0-custom_onednn2.6_9a3a3181e7056dcf7ccd3a16e599e6882a4edc23，Yonyou OCR v1，OneDNN 2.6。测试配置 2: 单节点，双路英特尔® 至强® 铂金 8480+ 处理器，56 核，启用超线程，启用睿频加速技术，256 GB 总内存 (16 插槽/32 GB/4800 MHz)，EGSDCRB1.SYS.0085.D15.2207241333，0x2b000070，Ubuntu 22.04.1 LTS，5.19.0-051900-generic，gcc 11.2，AsiaInfo OCR v1，OpenVINO 2022.2.0-custom_onednn2.6_9a3a3181e7056dcf7ccd3a16e599e6882a4edc23，Yonyou OCR v1，OneDNN 2.6。实际性能受使用情况、配置和其他因素的差异影响。更多信息请见 www.Intel.com/PerformanceIndex

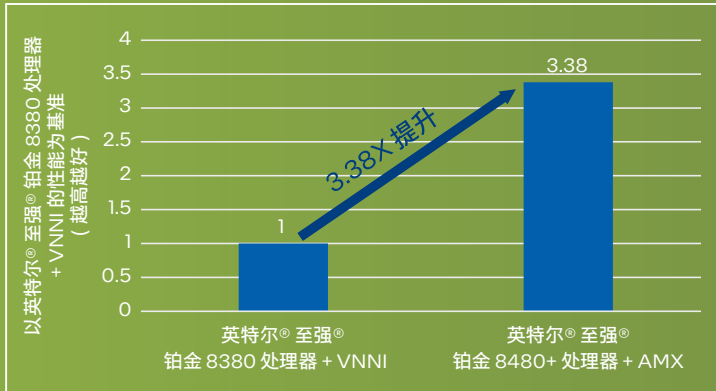


图 2. 第三代/第四代英特尔® 至强® 可扩展处理器 OCR 算法推理性能对比 (时延 <50ms)

收益：加快业务流程，赋能数字化转型

AISWare AIRPA 方案作为代替人工、降本增效的工具，广泛适用于财务报销、5G 基站建设职能质检、智慧水务防汛监控巡检、通信运营商业务批量填单、封控数据智能化采集等多种场景。在结合第四代英特尔® 至强® 可扩展处理器集成的高级硬件能力之后，能够进一步发挥方案价值，助力企业数字化转型。

- **推动降本增效、助力解放人力：**面向重复、高耗时的业务，成本降至原人工执行的 1/5 到 1/9 之间，效率提升约 5-10 倍⁶；实现 7*24 小时在线，解放人力，推动员工技能升级和价值升华，执行更有创造性的操作。
- **操作安全易用，结果准确可靠：**AISWare AIRPA 方案实现轻量化部署，支持快速交付，且方案自带开箱即用机器人，帮助用户迅速理解 RPA 能力，挖掘自身需求。

- **打破业务烟囱，联通数据孤岛：**AISWare AIRPA 方案可实现跨系统的业务流程处理，提升业务处理的连续性，加快业务信息化进程。还能够通过跨系统的数据搬运和处理加快数据运转速度，提高数据利用率。
- **流程智能化，助力企业数智化转型：**AISWare AIRPA 方案结合 AI 服务，为流程重塑提供智能化引导，推动业务流程的优化重构，充分发挥流程在企业中的价值，助力企业数智化转型。

展望

在融合第四代英特尔® 至强® 可扩展处理器集成的高级硬件能力之后，亚信科技显著提升了 AISWare AIRPA 方案中智能 OCR 应用的运行效能，能够为用户提供更具性价比的高品质 OCR 服务。除了智能 OCR 应用之外，亚信科技与英特尔还致力于将第四代英特尔® 至强® 可扩展处理器、英特尔® AMX 等特性融合到更多应用中，加速 RPA 系统运行，帮助用户充分释放 RPA 在自动化流程处理方面的价值。

面对人工智能 (AI)、大数据、高性能数据分析等应用所带来的指数级增长的算力挑战，英特尔将与亚信科技等伙伴携手，释放第四代英特尔® 至强® 可扩展处理器在不同场景的应用潜力，并通过基础设施架构优化、技术与服务创新等方式，为用户提供敏捷、灵活、高性能、高可用的解决方案，为用户数字化转型之旅提供基础能力支撑。

⁶ 数据援引自亚信科技内部测试结果。英特尔并不控制或审计第三方数据。请您审查该内容，咨询其他来源，并确认提及数据是否准确。

实际性能受使用情况、配置和其他因素的差异影响。更多信息请见 www.Intel.com/PerformanceIndex

性能测试结果基于配置信息中显示的日期进行测试，且可能并未反映所有公开可用的安全更新。详情请参阅配置信息披露。没有任何产品或组件是绝对安全的。

具体成本和结果可能不同。

英特尔技术可能需要启用硬件、软件或激活服务。

英特尔未做任何明示和默示的保证，包括但不限于，关于适销性、适合特定目的及不侵权的默示保证，以及在履约过程、交易过程或贸易惯例中引起的任何保证。

英特尔并不控制或审计第三方数据。请您审查该内容，咨询其他来源，并确认提及数据是否准确。

© 英特尔公司版权所有。英特尔、英特尔标识以及其他英特尔商标是英特尔公司或其子公司在美国和/或其他国家的商标。其他的名称和品牌可能是其他所有者的资产。