



+



## 挑战

随着电商行业的快速发展，快递业务总量增长迅猛。据统计，2022年上半年，全国快递服务企业的业务量累计完成512.2亿票，同比增长3.7%<sup>1</sup>。为应对激增的业务量并提升快件揽收量与派送效率，上海韵达货运有限公司（以下简称为“韵达”）实施了网格仓<sup>2</sup>策略，借助由视觉人工智能（AI）驱动的智能共配<sup>3</sup>分拣系统对快件进行自动分拣。同时，由于快递行业重时效的特性，推进全链路时效优化也十分必要，为此，韵达正大力推广分拨视频分析平台以分析包括格口在内的分拨视频，进而优化激励机制和业务管理能力。然而，在项目推进过程中，韵达遇到了以下挑战：

- **网格仓承载快件数量庞大：**据公开数据显示，韵达2022年5月完成的业务量为16.12亿票，日均业务量超5000万票<sup>4</sup>。由此可见，网格仓每日处理的快件数量庞大。例如，上海某网点的网格仓试点每日需要处理的快件数量高达3万票<sup>5</sup>。
- **智能分拣线须兼顾高准确度和低时延：**分拣线会对快件面单上的信息进行三段码OCR (Optical Character Recognition) 检测，由于识别的结果决定着快件在分拣线上的去向，因此该步骤对分拣效率至关重要，对系统的识别准确度和时延有较高要求。一方面，当识别出现错误，就会造成分拣线回流，即快件需重新再过分拣线或需要人工分拣；而当错误率较高时，更会加重网点的操作压力。故而，韵达提出系统识别的准确度须达到95%以上。另一方面，系统时延与分拣线传输速度息息相关，检测算法的时延会直接影响分拣传输的效率，过慢就会造成快件积压，进而影响分拣速度和派送时效，韵达经过评估后发现智能分拣线系统的时延必须要小于130ms才能满足他们对高效率的需求。
- **视频流分析体量庞大，对算力要求高：**分拨视频分析平台的全网部署会涉及到来自超过10万<sup>6</sup>个摄像头的视频流，数据体量庞大，对韵达IT基础设施的算力和带宽都提出了更高要求。

案例简介  
 物流  
 视觉 AI  
 英特尔® 数据中心 GPU Flex 系列  
 英特尔® 分发版 OpenVINO™ 工具套件

## 韵达基于英特尔® 数据中心 GPU Flex 系列，优化视觉 AI 方案，践行智慧物流战略



### 解决方案与成果

针对以上挑战，韵达基于英特尔® 数据中心 GPU Flex 系列 170 和英特尔® 分发版 OpenVINO™ 工具套件来优化其三段码 OCR 检测系统和分拨视频分析平台性能。

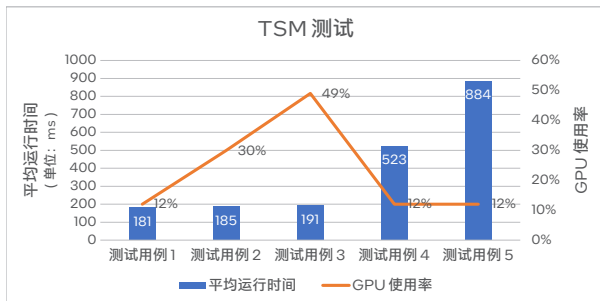
英特尔® 数据中心 GPU Flex 系列 170 是英特尔面向视觉 AI 和智能云计算等场景而打造的、基于英特尔® X<sup>e</sup> 架构的 GPU，拥有高达 512 个执行单元，能够保证多线程处理的吞吐量，同时支持 H.264、H.265 (HEVC) 硬件编码/解码和 AV1 编码/解码<sup>6</sup>，高度契合韵达视觉 AI 方案对图片与视频处理的需求。

英特尔® OpenVINO™ 工具套件包含模型优化器和推理引擎两大组件。韵达利用 OpenVINO™ 模型优化器 MO (Model Optimizer) 可将 Caffe、TensorFlow、Pytorch 和 PaddlePaddle 等多种常见框架的模型转换为 OpenVINO™ 中间数据格式 (IR, Intermediate Representation) 的离线模型，并且对这些模型的性能进行优化；推理引擎则可以为跨英特尔多种芯片 (包括 CPU、GPU 和 FPGA 等) 的计算机视觉异构计算提供加速支持。

为验证方案性能，韵达进行了三段码 OCR 测试和 TSM (Temporal Shift Module) 测试。

在三段码 OCR 测试中，共对 2450 张图片进行了识别，测试结果显示平均运行时间为 114 ms<sup>7</sup>，优于韵达 130 ms 的期望标准，可很好地满足韵达对低时延的需求。此外，根据实验室测试结果，经优化后，三段码 OCR 识别的准确度能够达到 97 %-98%<sup>7</sup>，也优于韵达 95% 的预期基准。

在 TSM 测试中，选用了 1200 个视频并分不同实例和批量大小做了测试，测试结果如图 1 所示。



测试用例 1: 1 实例, 批量大小=1  
 测试用例 2: 2 实例, 批量大小=1  
 测试用例 3: 3 实例, 批量大小=1  
 测试用例 4: 1 实例, 批量大小=3  
 测试用例 5: 1 实例, 批量大小=5

图 1. 基于英特尔® 数据中心 GPU Flex 170 的 TSM 测试结果<sup>7</sup>

综合测试和网点试点结果, 基于英特尔® 数据中心 GPU Flex 系列 170 和英特尔® 分发版 OpenVINO™ 工具套件的视觉 AI 方案, 从算力、时延、准确度、并发能力、稳定性和散热能力等多个方面都能很好地满足韵达的需求, 并为韵达带来了以下业务优势:

- 提升业务效率并降低成本: 智能分拣系统的部署显著提升了分拣线效率, 进而提升了派送效率和派送时效; 此外, 高效的智能分拣系统还帮助韵达实现了人力和成本节约。
- 优化业务管理和决策: 英特尔® 数据中心 GPU Flex 系列 170 带来的高算力让韵达实现了更高效的分拨视频流分析, 可帮助韵达在跨年度/跨季度预测、合理定价、网点时效提升和奖惩制定等方面做出优化。

## 为什么选择英特尔

### 软硬件全面创新, 满足各类需求

从算力提升到框架指令的适配, 英特尔着眼 AI 应用开发与部署的各个环节, 全面创新, 旨在满足用户的不同需求。硬件层面, 英特尔不仅在持续增强其 CPU 英特尔® 至强® 可扩展处理器的内置 AI 加速能力, 还推出包括 GPU、FPGA 和 VPU 在内的各类专用加速器以满足不同场景对算力的特定需求。软件层面, 从模型、框架到底层库, 英特尔也在不断创新以适配各类硬件, 赋能用户更好地基于英特尔® 架构实现 AI 加速。例如, 在英特尔® oneAPI 和 OpenVINO™ 工具套件的支持下, 模型可实现跨英特尔® CPU 和 GPU 的无缝切换, 且几乎不会对应用层造成任何影响。

### 专业的技术支持, 加速 AI 应用开发与部署

AI 应用的开发并非易事, 尤其是在异构计算兴起的今天, 更是变得越来越复杂。英特尔拥有强大且专业的技术团队, 可在整个项目周期和项目结束后为用户提供专业支持。因此, 企业即使开发能力有限, 亦可实现快速开发与部署。

### 更多信息

- 有关韵达的更多信息, 请访问: <http://www.yundaex.com/cn>。
- 有关英特尔® 数据中心 GPU Flex 系列的更多信息, 请访问: <https://www.intel.cn/content/www/cn/zh/products/details/discrete-gpus/data-center-gpu/flex-series.html>。
- 有关英特尔® 分发版 OpenVINO™ 工具套件的更多信息, 请访问: <https://docs.openvino.ai/cn/latest/index.html>。
- 有关英特尔® 数据中心 GPU Flex 系列的其他案例研究, 请访问: <https://www.intel.cn/content/www/cn/zh/customer-spotlight/cases/arcvideo-tech-efficient-ai-system-old-film-repair.html>。



<sup>1</sup> 《国家邮政局公布 2022 年上半年邮政行业运行情况》, <https://www.spb.gov.cn/gjyzj/c100015/c100016/202207/1c0f0bd13af647c49a0819f577cb4a22.shtml>。  
<sup>2</sup> 网格仓: 一种介于分拨中心与网点的站点。与分拨中心相比, 其不具备中转功能; 与网点相比, 是由网点老板和韵达总部共同投资, 同时受两者监管。  
<sup>3</sup> 指快递共配, 即通过把各个快递品牌整合起来, 共用场地、运力等资源, 缩小快递员的派送区域, 提高派送密度, 把单一品牌包裹派送转变为多品牌包裹配送, 从而提升配送能力。  
<sup>4</sup> 《A 股快递公司 5 月成绩出炉: 韵达日单量超 5000 万创记录、行业单票价格降幅收窄》, <http://www.nbd.com.cn/articles/2021-06-18/1798575.html>。  
<sup>5</sup> 数据由韵达提供, 如欲了解详情, 请与韵达联系。  
<sup>6</sup> 英特尔® 数据中心 GPU Flex 系列 170 规格: <https://www.intel.cn/content/www/cn/zh/products/sku/230019/intel-data-center-gpu-flex-170/specifications.html>。  
<sup>7</sup> 性能测试结果基于韵达于 2022 年 10 月进行的测试。配置详情: 单节点, 双路英特尔® 至强® 金牌 6348 处理器 (28 核/路, 56 线程/路), 启用超线程, 启用睿频; GPU: 英特尔® 数据中心 GPU Flex 系列 170; 内存总容量: 256 GB (16 x 16 GB, DDR 2933); 操作系统: Ubuntu 20.04; 内核版本: 5.10.54; 工作负载: dlstreamer; 编译器: gcc; 库: 英特尔® oneAPI 工具套件; 其他软件: 英特尔® OpenVino™ 工具套件 2022.2 版。

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。  
 描述的成本降低情景均旨在在特定情况和配置中举例说明特定英特尔产品如何影响未来成本并提供成本节约。情况均不同。英特尔不保证任何成本或成本降低。  
 英特尔技术特性和优势取决于系统配置, 并可能需要支持的硬件、软件或服务得以激活。产品性能会基于系统配置有所变。没有任何产品或组件是绝对安全的。更多信息请从原始设备制造商或零售商处获得, 或请见 [intel.cn](http://intel.cn)。  
 性能测试中使用的软件和工作负荷可能仅在英特尔微处理器上进行了性能优化。诸如 SYSmark 和 MobileMark 等测试均系基于特定计算机系统、硬件、软件、操作系统及功能。上述任何要素的变动都有可能对测试结果的变化。请参考其他信息及性能测试 (包括结合其他产品使用时的运行性能) 以对目标产品进行全面评估。更多信息, 详见 [www.intel.cn/benchmarks](http://www.intel.cn/benchmarks)。  
 描述的成本降低情景均旨在在特定情况和配置中举例说明特定英特尔产品如何影响未来成本并提供成本节约。情况均不同。英特尔不保证任何成本或成本降低。  
 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。  
 © 英特尔公司版权所有。