Computer Science and Application 计算机科学与应用, 2025, 15(12), 313-325
Published Online December 2025 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2025.1512346

Hans汉斯

面向集运MaaS多式联运的多模态大模型研究与应用

刘　隐

上海海勃数科技术有限公司，上海

收稿日期：2025年11月9日；录用日期：2025年12月12日；发布日期：2025年12月19日

摘　要

多模态大模型作为人工智能的前沿方向，正逐步成为推动智慧物流升级的核心动力。在集运MaaS多式联运这一复杂场景中，传统单模态方法难以融合船舶AIS轨迹、港区监控视频、电子单证与气象文本等多源异构数据，导致跨模态语义理解与决策支持不足。针对上述问题，本文提出一种基于多模态大模型的智能系统架构，并构建四大核心应用场景：智能单证识别、货物异常识别、智能规划决策与业务数据问答。系统利用Qwen2.5-VL-32B实现单证自动解析与结构化输出，通过图文一致性比对提升货物核验效率，整合实时多源数据为客户提供多式联运路径优化方案，并以Qwen3-235B-A22B支撑自然语言查询与多轮交互。研究表明上述应用在提升物流效率、优化运营决策与降低作业成本方面具有显著潜力。

关键词

多模态大模型，智慧物流，多式联运，跨模态融合，人工智能

Research and Application of MLLMs for Container Shipping MaaS Multimodal Transportation

Yin Liu

Shanghai Harbor Digital Technology Co., Ltd., Shanghai

Received: November 9, 2025; accepted: December 12, 2025; published: December 19, 2025

Abstract

Multimodal large models, as a frontier of artificial intelligence, are increasingly becoming a principal driving force for advancing intelligent logistics. In the complex context of container consolidation

文章引用: 刘隐. 面向集运 MaaS 多式联运的多模态大模型研究与应用[J]. 计算机科学与应用, 2025, 15(12): 313-325.
DOI: 10.12677/csa.2025.1512346

under Mobility-as-a-Service (MaaS) multimodal transport, traditional unimodal approaches struggle to integrate heterogeneous data sources—such as vessel AIS trajectories, port surveillance video, electronic documents, and meteorological texts—resulting in insufficient cross-modal semantic understanding and decision support. To address these challenges, this paper proposes an intelligent system architecture based on multimodal large models and develops four core application scenarios: intelligent document recognition, cargo anomaly detection, intelligent planning and decision-making, and business-data question answering. The system employs Qwen2.5-VL-32B to automate document parsing and structured output, leverages image-text consistency checks to enhance cargo verification efficiency, integrates real-time multisource data to provide multimodal route-optimization solutions, and utilizes Qwen3-235B-A22B to enable natural-language queries and multi-turn interaction. Empirical analysis demonstrates that the proposed applications have significant potential to improve logistics efficiency, optimize operational decision-making, and reduce operational costs.

Keywords

Multimodal Large Language Models, Smart Logistics, Multimodal Transport, Cross-Modal Fusion, Artificial Intelligence

1. 引言

1.1. 研究背景及意义

随着全球贸易的持续增长与供应链网络的日益复杂，航运业正处于数字化与智能化转型的关键阶段。当前，国际集装箱运输及多式联运体系在运力调度、路径规划、港口协同与资源配置等方面仍面临数据割裂及智能决策不足等问题，严重制约了航运系统的整体运行效率与服务质量。为实现集运 MaaS 模式下的高效衔接与智能调度，亟需具备跨模态感知、综合分析与决策能力的新型智能系统。

人工智能技术的快速发展为航运领域的智能化变革提供了重要支撑，其中大型语言模型作为近年来的突破性技术，展现出了前所未有的语言理解与生成能力[1]。传统的人工智能系统大多专注于单一模态的信息处理，使其难以应对航运场景中涉及的多源异构数据。在此背景下，多模态大模型应运而生。该类模型通过融合视觉、语言及其他感知模态，突破了传统人工智能系统的单模态限制，具备更强的语义理解与跨模态推理能力[2]。深度学习尤其是 Transformer 架构的提出为多模态技术的发展奠定了坚实基础，基于上述技术积累，多模态大模型正逐步成为推动智慧航运与多式联运体系高质量发展的关键技术支撑。

1.2. 国内外研究现状

多模态大模型正逐步成为推动智慧航运与多式联运体系高质量发展的关键技术支撑。多模态大模型发展建立在各个单模态大模型技术积累的基础之上，其演进历程体现了从单一模态向多模态融合的技术路径。国外在该领域处于领先地位，主要研究机构包括 Google、OpenAI、Meta、Anthropic 等科技公司，以及斯坦福大学、MIT 等学术机构。

在语言大模型方面，Transformer 架构的提出彻底改变了自然语言处理领域的局面。Vaswani 等人[3]

在 2017 年提出的“Attention Is All You Need”奠定了现代大型语言模型的基础架构。GPT 系列模型代表了自回归语言模型发展的重要里程碑，从 GPT-1 到 GPT-3，模型规模呈现指数级增长[4]。Brown 等人提出的 GPT-3 模型在多种自然语言处理任务上展现了强大的零样本和少样本学习能力。BERT 系列模型则代表了双向编码器架构的发展路径，Devlin 等人提出的 BERT 模型通过掩码语言建模和下一句预测任务，实现了对上下文信息的双向理解。

在视觉大模型领域，Vision Transformer 的提出标志着重要转折点。Dosovitskiy 等人将 Transformer 架构成功应用于图像分类任务，ViT 在 ImageNet 等大规模数据集上的优异表现证明了 Transformer 架构在视觉任务中的有效性。基于 ViT 的改进方案进一步提升了模型性能和效率，Liu 等人[5]提出的 Swin Transformer 通过层次化的窗口注意力机制实现了多尺度特征提取。

在视觉–语言大模型方面，CLIP 模型的提出标志着视觉–语言预训练进入了新的发展阶段。Radford 等人[6]通过对比学习的方式在大规模图文对数据上进行预训练，学习图像和文本的联合表示空间。生成式视觉–语言模型代表了另一个重要发展方向，DALL-E 系列模型通过文本到图像的生成任务展现了多模态生成的巨大潜力。Flamingo 模型在少样本 VQA 任务上取得了显著性能，其核心架构通过引入可训练的感知重采样模块与门控交叉注意力层，连接了冻结的预训练视觉编码器与语言模型。

MLMs 研究呈现出蓬勃发展的态势。早期的工作如 Alayrac 等人提出的 Flamingo 模型，通过在大型语言模型中注入视觉信息，实现了强大的少样本视觉语言理解能力。随后，Liu 等人提出的 LLaVA 模型通过视觉指令微调，进一步提升了模型的多模态理解和指令遵循能力。2023 年底，各大科技公司相继发布了更强大的 MLMs，如 OpenAI 的 GPT-4V [7]、Google 的 Gemini 等，这些模型在多模态理解和生成方面展现出了接近人类水平的能力，这些技术突破正在航运领域得到实质性应用，例如中远海运集团基于大模型开发的智能客服系统能够处理复杂的航运业务咨询，部分港口企业利用多模态模型实现集装箱状态自动识别和异常检测，显著提升了港口作业效率。

与此同时，多模态模型 Qwen-VL [8]在架构设计上融合了当代先进模型的优势，在视觉-语言理解方面取得突破，支持图像描述、视觉问答、文档解析等多模态任务，其强大的多模态理解能力特别适合于处理航运业务中复杂的图像、文本和数据融合任务，为实现从单证识别到货物追踪的全流程智能化提供了核心技术支持[9]。

2. 集运 MaaS 多式联运平台概述

2.1. 平台建设背景与目标

集运 MaaS 平台的建设背景源于集装箱运输行业长期存在的“跨主体、跨部门、跨地区”数据共享难题。传统物流模式下，货主、船公司、港口、车队、海关等多方主体信息系统互不联通，导致信息孤岛现象严重，业务流程繁琐、效率低下。在国际集装箱运输过程中，涉及出口流程节点和进口流程节点，包括订舱、报关、装箱、运输、到港、清关等多个环节，传统方式需人工频繁切换不同系统，处理效率低且易出错。在此背景下，以上海市交通委员会、上海海事局及上海海关为代表的政府监管机构共同参与顶层设计与政策支持，推动建设集运 MaaS 平台，以“安全共享、数据引领、聚合共赢、分步推进”为指导理念，打造以港口服务为核心的统一公共服务平台。

平台旨在通过数字化手段，整合集装箱运输“全链条”物流信息，建立“全融合”性的业务办理门户，为货主、船公司、港口运营商、物流企业及政府监管部门提供数字供应链服务，最终提升港口作业效率、优化客户服务体验，推动港口行业数字化转型，为国际航运中心建设赋能。此平台不仅破解了长期困扰行业的跨主体、跨部门、跨地区数据共享难题，更为上海国际航运中心建设提供了重要支撑，推动了港航生态圈向更加现代化、透明化的方向发展。

2.2. 核心功能体系

如表 1 所示，集运 MaaS 平台以“一门式查询、一站式服务、一体化赋能、一单制运行”为核心，整合船舶动态、货物状态与多式联运路线等全程物流信息，通过统一界面实现箱货车船全程跟踪，初步构建跨区域海铁联运信息网络。平台提供无纸化换单、港区计划申报、电子设备交接单和电子放货等在线业务功能，并引入分布式数字身份认证与区块链技术，确保操作安全可靠、流程不可篡改，显著提升物流效率与数据可信度[10]。

平台还借助数字孪生技术对码头周边道路进行建模，深度融合港区内外部数据，构建仿真环境以辅助交通监管与资源配置，并通过分布式存储计算保障多源数据融合与系统稳定运行。同时，积极推动跨运输方式单证互认，开展海铁联运“一单制”业务，契合国际多式联运信息服务标准。此外，平台聚合多方监管与运营数据，形成港航数据底座，并接入基于 AI 大模型的智能客服系统，具备自然语言理解和多轮对话能力，可全天候处理复杂业务咨询，进一步提升用户体验与服务效率。

Table 1. “Four-in-One” functional system of the consolidated shipping MaaS platform
表 1. 集运 MaaS 平台“四个一”功能体系

功能体系	核心功能	关键技术
一门式查询	一站式查询船舶、货物、多式联运路线等全程物流信息	数据整合与可视化技术
一站式服务	无纸化换单、EDO、EIR 等在线业务办理	区块链、数字身份认证
一体化赋能	数字孪生建模、数据驱动决策支持、智能预测	数字孪生、大数据分析、AI
一单制运行	跨运输方式单证互认，铁海联运全程物流产品	区块链、API 集成、数据标准

3. 集运 MaaS 多模态大模型技术原理

3.1. 集运 MaaS 多模态大模型概括

在全球航运体系数字化与智能化持续推进的背景下，集装箱运输正朝着 MaaS 模式加速演进。该模式旨在通过多式联运的全链条协同，实现运输资源的动态调度与服务一体化。然而，集运 MaaS 场景中存在数据来源多样、结构复杂、模态差异显著等特点，如船舶 AIS 轨迹、港区监控视频、电子单证、气象与航线文本信息等，这些多源异构数据对智能化分析与决策提出了更高要求。为实现对复杂航运场景的全面感知、语义理解与智能推理，构建具备跨模态信息融合能力的多模态大模型成为关键技术路径。

多模态大模型(Multimodal Large Language Models, MLLMs)为集运 MaaS 平台的智能化升级提供了技术基础，MLLMs 是基于大型语言模型扩展形成的人工智能技术体系，核心在于整合文本、图像、音频、视频等异质模态信息，实现跨模态感知、语义理解与逻辑推理。与传统单模态模型相比，MLLMs 具备三大核心特质：一是参数规模庞大，通常达数十亿至千亿级，可捕捉复杂多模态语义关联；二是训练范式创新，通过多模态预训练与指令微调，显著提升模型泛化能力与任务适配性；三是能力涌现，如无需光学字符识别的数学推理、跨模态因果分析等，可支撑复杂场景应用。

3.2. 多模态大模型关键技术

3.2.1. 多模态大模型基础架构

多模态大模型的基础架构通常由三个核心组件构成：模态编码器、连接器和语言模型。

模态编码器负责将原始多模态数据转换为高维特征表示。视觉编码器主要采用卷积神经网络(CNN)

架构或 Transformer 架构。研究表明，ViT 在图像分类任务上的准确率较高，这主要得益于其自注意力机制对全局依赖关系的有效捕捉。文本编码器则通常采用 BERT、RoBERTa 等预训练语言模型，这些模型在大规模文本语料上训练，能够生成高质量的文本表征。编码器的核心功能是将图像、音频等原始多模态信息压缩为更紧凑的特征表示。目前主流的实现方式并非从零开始训练编码器，而是优先采用已与其他模态完成初步对齐的预训练编码器——这一选择可大幅降低后续与 LLM 对齐的难度。基于这类已完成初步对齐的编码器，后续仅需通过对齐预训练，即可高效实现其与 LLM 的模态匹配。模态特征编码是 MLLMs 的基础，需将异质模态数据转换为统一维度的特征向量，核心包括视觉编码与文本编码两类技术。视觉编码主流采用 ViT 架构，将图像分割为 P × P 像素 patch 后线性投影并添加位置编码，公式如下：

$$x_{\text{vis}} = \text{PE}\left(W_p \cdot \text{patch}\left(I\right) + b_p\right) \tag{1}$$

其中，I 为输入图像，$\text{patch}(\cdot)$ 为图像分块操作，W_p、b_p 为线性投影参数，$\text{PE}(\cdot)$ 为正弦位置编码。

文本编码以预训练语言模型为核心，如 BERT、GPT 系列，其核心依托 Transformer 自注意力机制[3]捕捉文本上下文依赖，Transformer 架构以自注意力机制为核心，计算公式为：

$$Attention\left(Q,K,V\right) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

其中 Q、K、V 分别为查询、键、值矩阵，d_k 为维度缩放因子。

连接器的作用是将视觉特征处理成语言模型可理解的形式。目前主流连接器设计包括 Q-Former 和线性投影层。Q-Former 通过一组可学习的查询向量与视觉特征进行交互，能够有效提取与文本相关的视觉信息；线性投影层则通过全连接层将视觉特征映射到语言模型的嵌入空间，实现模态间的特征对齐。

语言模型作为多模态信息的综合处理器，负责生成最终的响应输出。在多模态大模型的构建中，直接从零训练 LLM 的成本极高，因此更高效且实用的方案是基于已完成预训练的 LLM 进行扩展。

3.2.2. 跨模态对齐与融合技术

跨模态对齐与融合是多模态大模型实现多源信息统一理解与互补增强的核心环节，旨在弥合不同模态之间的语义差异，提升模型在复杂场景下的感知与推理能力。主要技术包括对比学习、跨模态注意力机制以及动态门控融合方法。

对比学习：通过 InfoNCE 损失函数拉近正样本对距离，推远负样本对。CLIP 模型采用大规模图像－文本对进行对比学习，实现零样本迁移能力其损失函数如下：

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp\left(\frac{s_{ii}}{\tau}\right)}{\sum_{j=1}^{N}\exp\left(\frac{s_{ij}}{\tau}\right)} \tag{3}$$

其中 s_{ij} 为相似度矩阵，τ 为温度参数。

跨模态注意力是自注意力机制在多模态场景的扩展，通过动态计算文本与视觉模态的注意力权重，实现局部信息精准交互。以 Flamingo 模型为例，其跨模态注意力计算公式如下：

$$\text{CrossAttention}\left(Q_t, K_v, V_v\right) = \text{softmax}\left(\frac{Q_t K_v^T}{\sqrt{d_k}}\right)V_v \tag{4}$$

其中，Q_t 为文本模态生成的查询矩阵，K_v、V_v 为视觉模态生成的键、值矩阵，d_k 为维度缩放因子。

3.2.3. 模型训练与优化策略

MLLMs 训练采用“预训练–指令微调–对齐微调”三阶段流程，逐步优化模型能力。

预训练作为基础阶段，实现不同模态的语义对齐与多模态知识注入，通常依赖大规模“模态–文本对”数据，指令微调旨在教会模型更好地理解用户指令并完成指定任务，通过该方式优化的大语言模型(LLMs)可依据新指令泛化至未见过的任务，进而提升零样本性能；多模态指令样本通常以指令 I、多模态输入 M、真实响应 R 的三元组形式存在，模型训练时，基于参数 θ 根据指令 I 和多模态输入 M 预测答案，即

$$A = f(I, M; \theta) \tag{5}$$

其中 A 为预测答案；训练目标采用 LLMs 常用的自回归目标，鼓励模型预测响应的下一个 token，目标函数可表示为

$$L(\theta) = -\sum_{i=1}^{N} \log p\left(R_i \middle| I, R_{<i}; \theta\right) \tag{6}$$

式中 N 为真实响应 R 的长度，(R_i) 为真实响应的第 i 个 token，$(R_{<i})$ 为第 i 个 token 之前的序列。

4. 多模态大模型在集运 MaaS 平台应用设计

4.1. 系统架构选型

集运 MaaS 平台采用面向服务的架构(SOA)设计理念，将多式联运业务逻辑拆分为多个独立的微服务模块，通过定义良好的接口实现服务间松耦合协作。该架构以多模态大模型为核心智能引擎，通过 API 网关对外提供统一服务接入，支持与港口 TOS 系统、海关报关平台、船公司调度系统等多源异构系统的无缝集成。平台采用分层架构设计，包括基础设施层、数据中台层、AI 能力层和应用服务层，其中数据中台层通过分布式数据湖技术整合船舶 AIS 数据、视频监控流、电子单证等多模态数据，为上层智能应用提供统一数据支撑。这种架构设计既保证了系统的高可用性和可扩展性，又通过服务治理机制实现了业务功能的快速迭代和灵活复用，为多式联运业务的数字化转型提供了坚实的技术基础。

4.2. 系统技术路线

平台采用分层协同的架构体系，如图 1 所示，在技术实现上体现为前端展示层、逻辑编排层、权限中转层、向量数据库层和硬件模型层的五级技术栈。

前端展示层基于 Vue.js 框架开发，支持桌面端 Web 和移动端 H5 访问，提供登录页、聊天页和知识库页等交互界面，实现用户与系统的无缝对接。逻辑编排层依托 Dify 平台进行核心逻辑编排，集成日志记录、数据标注和 API 访问等功能，确保业务流程的高效调度与可追溯性。权限中转层采用 Fastapi 框架处理文件上传等安全管控任务，通过轻量级服务网关实现权限校验与数据中转。向量数据库层基于 Ragflow 构建，支持向量数据库的编辑与 chunk 查询，并集成文档解析器、OCR 技术、文档布局分析和表格结构识别等功能，实现对多源异构数据的智能化处理与关联建模。

硬件模型层部署 Nvidia A800 GPU、48 核 CPU、100G 内存和 1T 存储等硬件资源，运行 Qwen3-235B-A22B、Qwen2.5-VL-32B 等大语言模型以及 Bge-large-zh、Bge-reranker-v2-m3 等嵌入与重排序模型，为系统提供强大的智能计算核心。整个架构通过 API 网关统一服务接口，确保各层间松耦合协作，支持与外部系统的快速集成和业务协同。

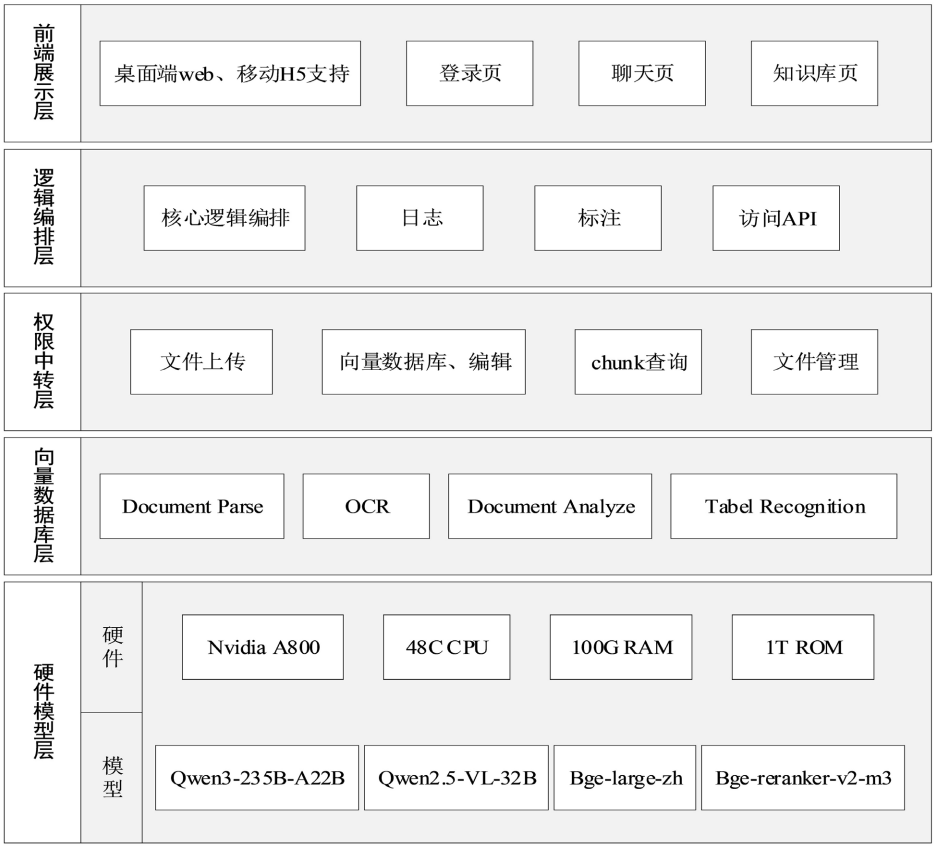

Figure 1. MaaS consolidated shipping platform system architecture diagram
图 1. MaaS 集运平台系统架构图

4.3. 系统功能设计

4.3.1. 智能单证识别

在 MaaS 集运平台的业务体系中，单证识别是实现运输链路高效运作和信息流通的关键环节。提单、货物清单、报关单以及运输委托书等文件不仅种类繁多，而且在不同国家和地区存在显著的格式差异与语言多样性，这对自动化识别提出了极高要求。Qwen2.5-VL-32B 凭借其跨模态理解与指令理解能力，为单证图像识别提供了新的技术路径，使平台能够直接基于原始文档影像与文本模态进行建模与解析，避免了繁琐的人工操作。

在识别流程中，流程图如图 2 所示。平台首先将业务环节产生的各类单证影像统一接入，由 Qwen2.5-VL-32B 作为核心多模态大模型执行跨模态解析。模型通过视觉特征与语言特征的深度融合，不仅能够识别文档中的文本信息，还能结合文本模态生成符合业务需求的结构化输出。在典型的提单场景中，模型能够抽取诸如提单号、发货人、收货人、箱号、货物描述、日期与金额等关键要素，并以统一的 JSON 格式返回，同时附带字段在图像中的位置信息。这一能力使得平台不仅能将结果直接写入核心业务系统，还能为后续的人工验证提供直观参考。

最终，经过模型解析的结构化结果将直接流入多式联运 MaaS 平台的核心业务系统，包括运输管理、清关处理与结算模块，实现从单证影像到业务数据的端到端转换。与传统方案相比，该流程显著提高了识别的自动化程度和准确性，降低了人工录入与复核的工作负担，并有效提升了跨境运输场景下的数据一致性与业务响应速度。

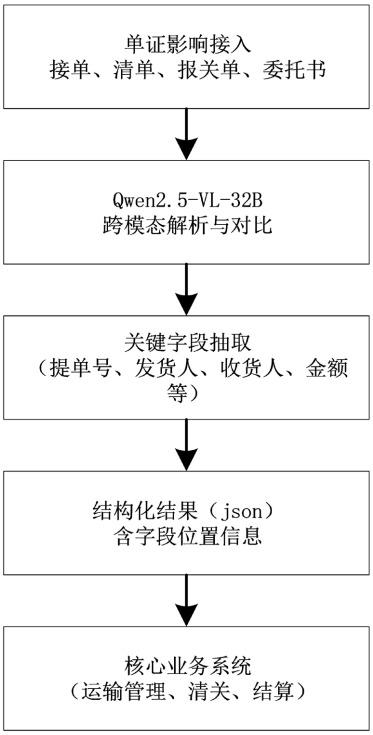

Figure 2. Document recognition process of the MaaS platform
图 2. MaaS 平台单证识别流程

基于 Qwen2.5-VL-32B 的单证识别流程通过跨模态建模与人工纠正相结合的方式，摒弃了传统图像识别算法，建立了一个智能化的单证处理流程。这一设计不仅为多式联运业务的数字化与智能化发展提供了坚实的技术支撑，也展现出多模态大模型在实际产业场景中广阔的应用前景与发展潜力。

4.3.2. 货物异常识别处理

在多式联运 MaaS 集运平台中，货物信息的一致性是保障运输链条高效运作与合规流转的核心要求。货物在装运、仓储及转运环节往往需要与提单、清单及报关单等多类单证信息进行对应，但在实际操作中，常常出现登记信息与货物实物不符的情况，例如重量、体积、箱号、标签标识与系统记录存在偏差。这类不一致若未能及时发现，不仅会导致运输调度和清关环节的延误，还可能引发合规风险。传统的核验方式主要依赖人工抽检，效率低且易受人为疏漏影响。

在技术框架上，流程图如图 3 所示。Qwen2.5-VL-32B 通过结合图像模态与文本模态，能够实现货物实物状态与登记信息的跨模态一致性校验。平台在货物进出仓、装载或换装环节采集货物图像及关联文本记录，如箱号、重量标签和清单条目。模型对图像进行解析，识别出关键标识与数值信息，同时对文本登记内容进行语义解析与结构化抽取。随后，系统将两类信息进行比对，一旦发现图像识别结果与文本登记存在明显差异，如重量数值不符、箱号与标签不一致或清单条目缺失，模型便会生成异常标记并反馈至业务系统。

多模态大模型识别出的异常信息直接流入 MaaS 平台的管理模块。例如，若检测到货物标签上的重量与系统登记不一致，系统会自动触发复核流程；若发现箱号识别与清单记录不符，则生成提示任务要求人工进行校正。平台通过这种方式实现了从货物实际状态到登记数据的自动化对比，大幅减少了人工逐项核对的工作量。

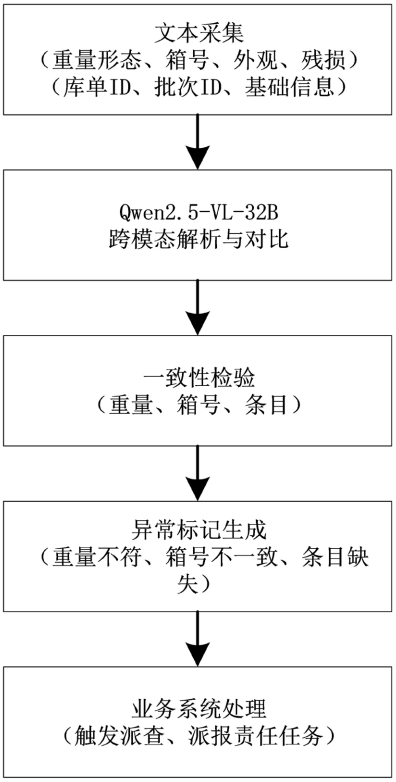

Figure 3. Cargo anomaly recognition process of the MaaS platform
图 3. MaaS 平台货物异常识别流程

基于 Qwen2.5-VL-32B 的货物异常识别流程，将货物实际信息与登记信息的一致性核验作为核心任务，实现了自动化、智能化的比对与纠正机制。这一设计不仅提升了多式联运场景下的信息透明度与运作效率，还显著降低了因信息不符导致的合规风险与运营延误，为 MaaS 集运平台的数字化与智能化发展提供了重要支撑。

4.3.3. 智能规划和决策

在集运 MaaS 平台中，面向货主与货代客户提供高效、透明且可靠的智能运输方案是实现用户价值的关键环节。系统基于 Qwen2.5-VL-32B 多模态大模型，实现对船舶 AIS 轨迹、港区监控视频、气象文本与铁路班列时刻等多源异构数据的深度融合理解，有效应对极端天气、政策变动等动态干扰，为客户提供包括路径规划、运输方案推荐、运费估算和船期选择在内的综合决策支持，全面提升多式联运服务的响应能力与客户满意度。

在智能规划流程中，流程图如图 4 所示。平台依托 Qwen2.5-VL-32B 实时接入并解析各类非结构化数据：将港区监控视频转化为拥堵指标，气象图像转换为可读的风力与影响范围文本，提取班列到港时间窗口等关键参数，并统一生成结构化特征。在方案生成阶段，模型融合视觉感知与业务规则，识别堆场拥堵、泊位异常等视觉事件，结合跨境运输策略为客户生成多条可行路径及其详细方案，包括运输时长、运费预估和船期选项。一旦数据发生更新，系统实时触发方案重算，经校验后通过 API 调整结果同步至客户平台及业务系统，确保信息透明与决策及时。

该方案有效解决了传统模式下数据整合困难、响应迟缓的问题。Qwen2.5-VL-32B 实现从多模态数据到标准化决策参数的全自动转换，显著减少人工干预，提升数据处理与方案生成效率；同时，系统具备

强大的动态响应能力，可依据实时数据持续优化运输方案与船期选择，为客户提供更高准确率的运费估算和更可靠的物流决策支持，助力企业实现降本增效与服务升级。

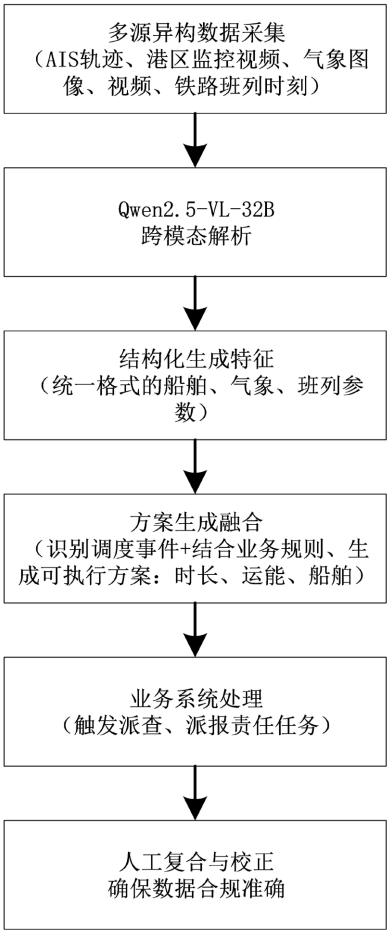

Figure 4. Intelligent planning and decision-making for multimodal transportation
图 4. 多式联运智能规划和决策

4.3.4. 业务数据问答助手

在集运 MaaS 平台中，面向货主、货代及运营人员提供高效且准确的数据查询与业务问答能力，已成为提升用户体验与操作效率的核心功能。该 AI 助手基于 Qwen3-235B-A22B 大语言模型，深度融合船期数据、靠泊动态、集装箱状态与货物跟踪信息等多维度业务数据，并接入持续更新的知识库，涵盖航线表、港口资料与运输政策等相对稳定的内容，实现对用户自然语言提问的智能解析与精准回答。该系统可广泛应用于箱货船期查询、航线详情展示、费用构成说明及节点跟踪等多种典型场景，为用户提供一站式数据与决策支持。

在智能问答系统构建过程中，流程图如图 5。在系统构建过程中，平台采用多层次技术架构处理数据集成与知识管理。数据预处理阶段整合多源异构数据，包括结构化数据与非结构化资料。预处理流程采用专用模型进行数据清洗与标准化，采用 BGE-large-zh 嵌入模型对文本数据进行向量化表示，构建语义索引使用 BGE-reranker-v2-m3 重排序模型对检索结果进行精度优化结合 Qwen2.5-VL 多模态模型处理文档中的表格与结构化布局信息这些预处理后的数据通过向量化技术构建为可高效检索的多模态业务知识

库，形成系统的数据底座。

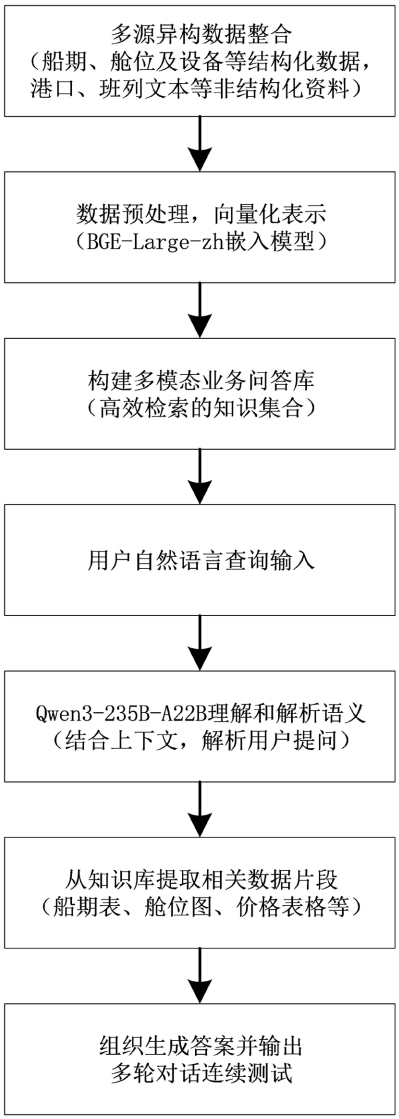

Figure 5. Q&A and dialogue for the multimodal transportation knowledge base
图 5. 多式联运知识库问答与对话

该 AI 助手有效解决了传统业务查询模式中数据分散、响应缓慢和理解偏差等痛点。基于 Qwen3-235B-A22B 的智能系统实现了从非结构化文本与表格数据到用户查询的端到端语义理解与答案生成。用户无需在不同系统间多次手动检索，极大提升了查询效率与答案一致性。同时，该系统还具备处理复杂多轮对话的能力，能够保持上下文语义连贯并在必要时进行意图澄清，尤其适用于跨境多式联运等高专业性场景，为用户提供可靠、即时且易于理解的全球物流信息支持，有效推动企业数字化服务转型与客户满意度提升。

4.4. 系统应用示例

基于多模态大模型的核心能力，集运 MaaS 平台实现了智能单证识别、货物异常识别、智能规划决

策与业务数据问答四大应用功能的深度融合。其将原始单证影像进行跨模态解析，自动抽取提单号、箱号等关键字段并以高亮框形式在原图上定位标注，同时在右侧面板结构化展示识别结果与置信度，提供数据核验与导出功能。

货物异常识别功能对货物图像与文本登记信息进行一致性比对，在报表界面中以色块与图标形式对异常数据进行显著标识，同步展示图像证据与系统记录以供对照，支持用户快速筛选与复核。

智能规划决策功能在地图界面上动态展示多式联运路径方案，结合港区视频、气象文本等多源数据，以可视化方式呈现运输时长、运费预估等关键指标，支持多方案对比与实时优化。

其通过对话式界面响应用户自然语言提问，在对话区域清晰记录用户问题与系统生成的答案，关键数据予以突出显示，并支持多轮交互。

5. 结论与展望

本文系统研究了多模态大模型在集运 MaaS 多式联运平台中的应用，围绕其技术原理、关键模型特性及落地实践展开了深入探讨。多模态大模型通过整合文本、图像等多源信息，实现了跨模态语义理解与交互，为传统物流行业数字化转型提供了强大的技术驱动力。研究表明，MLLMs 不仅能够有效应对集装箱运输中的单证处理、货物异常识别、路径动态调度与全流程追踪等核心业务需求，还在提升处理效率、降低人工成本、增强系统智能性等方面展现出显著优势。

具体而言，本文首先剖析了多模态大模型的基础架构与关键技术，包括模态编码、跨模态对齐与融合机制、长序列处理及多阶段训练策略等，明确了其技术演进路径与性能瓶颈。进而，结合集运 MaaS 平台的业务特点，提出了智能单证处理、货物异常识别、智能规划及智能助手四大典型应用场景，平台实现了从“被动响应”到“主动感知”、从“单点操作”到“全局优化”的转变，显著提升了物流链路的透明度、响应速度与可靠性。尽管本研究验证了多模态大模型在集运 MaaS 平台中的应用潜力，但仍存在若干局限性，需在后续研究与实践中予以关注。

在物流运输这类高可靠性要求的场景中，模型可能生成看似合理但不符合实际业务逻辑的内容，例如错误识别货物状态、生成不准确的运输路径建议等。为缓解此类风险，系统应引入人工审核机制；同时，可构建置信度评估模块，对模型输出的可靠性进行量化评分，低置信度结果自动触发人工介入，从而在自动化与风险控制之间取得平衡。

其次，多模态大模型通常依赖大规模计算资源，对硬件基础设施提出较高要求，可能限制其在中小型物流企业的推广。此外，模型推理延迟可能影响实时调度与异常响应效率。未来应着力开发轻量化、领域专用的多模态模型，通过知识蒸馏、动态稀疏化等技术降低计算开销，提升边缘部署能力；同时，探索模型并行与增量推理机制，优化系统在高并发场景下的吞吐能力。

最后，当前系统在人机协同决策流程方面仍有优化空间。尽管模型在多个场景中实现了自动化处理，但复杂异常或跨企业协调等情境仍需人类经验介入。未来研究应聚焦于构建更可靠的人机协同框架，明确机器与人工的职责边界，设计可解释的决策辅助界面，并探索多智能体协商机制在跨组织物流协同中的应用潜力，从而在提升效率的同时，保障复杂业务链路的鲁棒性与透明度。

参考文献

[1] 刘畅行, 陈思衡, 杨峰. 基于 MLMs 的智能无人机系统: 总结与展望[J]. 无线电工程, 2024, 54(12): 2923-2932.

[2] 车万翔, 窦志成, 冯岩松, 等. 大模型时代的自然语言处理: 挑战、机遇与发展[J]. 中国科学: 信息科学, 2023, 53(9): 1645-1687.

[3] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.

[4] 刘静, 郭龙腾. GPT-4 对多模态大模型在多模态理解、生成、交互上的启发[J]. 中国科学基金, 2023, 37(5): 793-802.

[5] Zhang, H., Li, F., Liu, S., *et al.* (2022) DINO: DETR with Improved De-Noising Anchor Boxes for End-To-End Object Detection. arXiv: 2203.03605.

[6] Radford, A., Kim, J.W., Hallacy, C., *et al.* (2021) Learning Transferable Visual Models from Natural Language Supervision. *Proceedings of the* 38*th International Conference on Machine Learning*, 18-24 July 2021, 8748-8763.

[7] Peng, B., Li, C., He, P., *et al.* (023) Instruction Tuning with GPT-4. arXiv: 2304.03277.

[8] Bai, J., Bai, S., Yang, S., *et al.* (2023) Qwen-VL: A Versatile Vision-Language Model for Understanding, Localization, Text Reading, and Beyond. arXiv: 2308.12966.

[9] 梁生龙, 范秋霞. 基于大模型的生成式数字孪生体建模[J/OL]. 图学学报: 1-7. https://link.cnki.net/urlid/10.1034.T.20250812.1556.004, 2025-10-13.

[10] 张雨薇, 王民. 基于提示词与多模态大模型选择的环境地图评价优化路径研究[J]. 地理教学, 2025(13): 20-24.