

面向工业质量检测的云边协同增量学习方法

梁秉豪, 张传刚*, 袁明明

浪潮通信信息系统有限公司, 山东 济南

收稿日期: 2026年1月7日; 录用日期: 2026年3月3日; 发布日期: 2026年3月17日

摘要

针对计算机视觉模型在工业质量检测场景中泛化能力弱、检测精度低、推理时延高等技术瓶颈, 本文提出了一种云边协同增量学习方法。该方法融合迁移学习与增量学习技术构建分层训练框架: 首先采用提出的加权焦点损失函数训练云端基础模型, 平衡节点间与节点内数据不平衡问题; 其次在边缘节点通过冻结骨干网络的小样本迁移学习, 快速适配本地业务场景得到边缘模型; 设计双时间周期机制实现模型增量更新与数据上传; 最终云端完成增量数据融合与模型迭代优化。该方法有效平衡了工业场景中模型性能与部署需求, 为智能质量检测提供可靠技术方案。

关键词

云边协同, 增量学习, 计算机视觉, 工业质量

Cloud-Edge Collaborative Incremental Learning Method for Industrial Quality Inspection

Binghao Liang, Chuangang Zhang*, Mingming Yuan

Inspur Communication Information System Co., Ltd., Jinan Shandong

Received: January 7, 2026; accepted: March 3, 2026; published: March 17, 2026

Abstract

Aiming at the technical bottlenecks of weak generalization ability, low detection accuracy, and high inference latency of computer vision models in industrial quality inspection scenarios, this paper proposes a cloud-edge collaborative incremental learning method. This method integrates transfer learning and incremental learning technologies to construct a hierarchical training framework:

*通讯作者。

文章引用: 梁秉豪, 张传刚, 袁明明. 面向工业质量检测的云边协同增量学习方法[J]. 人工智能与机器人研究, 2026, 15(2): 464-469. DOI: 10.12677/airr.2026.152045

first, the proposed weighted focal loss function is used to train the cloud-based base model, balancing the inter-node and intra-node data imbalance issues; second, on the edge nodes, few-shot transfer learning with a frozen backbone network is adopted to quickly adapt to local business scenarios and obtain edge models; a dual-time cycle mechanism is designed to realize incremental model updates and data uploads; finally, the cloud completes incremental data fusion and model iterative optimization. This method effectively balances model performance and deployment requirements in industrial scenarios, providing a reliable technical solution for intelligent quality inspection.

Keywords

Cloud-Edge Collaboration, Incremental Learning, Computer Vision, Industrial Quality

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着人工智能与 5G 通信技术的快速演进, 基于计算机视觉的智能质量检测已成为推动制造业转型升级的核心支撑技术[1]。通过人工智能(AI)模型替代传统人工检测, 不仅能突破人工操作的主观局限性, 还可显著提升生产效率与产品质量稳定性[2]。然而, 工业生产环境的复杂性给模型部署与应用带来独特挑战: 其一, 机械设备高速运转对检测系统的推理时延提出毫秒级要求, 传统云端集中式推理难以满足实时性需求[3]; 其二, 不同工厂、不同生产线的光照条件、成像设备及工艺参数存在差异, 导致数据分布异质性显著, 单一模型的跨场景泛化能力受限[4]; 其三, 生产过程中原材料替换、工艺优化等动态变化会引发数据分布漂移, 固定模型的检测精度会随时间持续衰减。这些问题导致传统 AI 部署模式难以充分适配工业质量检测的实际需求, 亟需构建兼顾泛化性、精度与实时性的新型技术框架。

云边协同架构通过云端算力优势与边缘端实时性优势的互补, 成为解决工业场景部署难题的有效路径[5]。增量学习则能够实现模型在不重新训练全量数据的前提下, 通过持续学习新数据完成性能迭代, 有效应对数据分布漂移问题[6]。本文将云边协同架构与增量学习技术深度融合, 提出分层训练与迭代优化机制, 同时针对工业数据的不平衡特性优化损失函数设计, 最终实现模型泛化能力、检测精度与推理实时性的协同提升。

2. 相关工作

目前云边协同架构是工业场景特别是质量检测等任务的核心技术方向, 现有研究主要通过构建检测器、边缘节点和云端节点三级架构实现。何毓芬等人[7]针对工业缺陷检测在数据质量和安全性上的问题, 系统综述和分析了不同类型视觉模型的核心应用场景和优劣势。吴子强等人[8]通过异步任务调度实现模型自动化部署, 确保数据不出域并降低网络负载。马昊[9]针对光伏企业缺陷检测的需求, 基于 Faster RCNN 作为检测器进行缺陷检测, 相比传统检测方法检测效率提升了 2 倍。梁道君等人[10]提出了 LightDets 模型作为检测器, 对空调外观的不良产品进行检测, 通过云边端资源的分布式协同调度, 实现了轻量化模型在小型终端上的部署和运行。上述方法均依赖云端训练、边缘推理的整体架构, 缺乏动态优化机制。针对因数据分布变化造成的模型漂移, 目前还没有成熟的解决方案, 武星等人[11]提出了一种增量学习方法, 并对比了增量学习和全量学习的模型精度和资源消耗, 证明了增量学习是实现了模型的

持续优化和升级的高效方法。

3. 云边协同架构

为了解决上述问题，本文主要基于云端统筹，边缘适配的分层架构，构建了训练、适配、更新、筛选、迭代的闭环流程，主要包括了五大核心环节：云端基础模型训练、边缘模型适配、边缘增量更新、云端数据筛选、云端模型迭代。整体架构如图 1 所示。

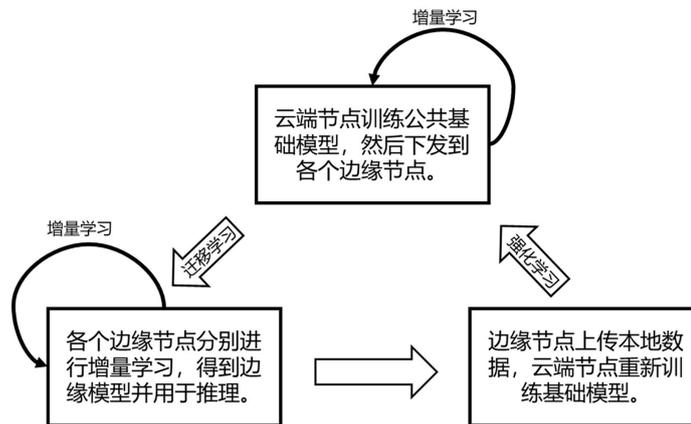


Figure 1. Overall architecture diagram
图 1. 整体架构示意图

3.1. 整体架构

本文所提出的云边协同方法主要由 1 个云端节点和多个边缘节点组成。云端节点具备较充足的智算力资源和存储空间，可以完成跨场景通用模型的训练与迭代，同时负责对边缘侧上传增量数据的筛选。边缘节点主要部署于工业生产环境中，具备轻量级智算力资源，可以实时完成工业数据采集以及特定场景模型的适配与增量更新，以及推理任务执行。边缘节点需要向云端上传筛选后的增量数据，云端节点向边缘节点下发更新后的基础模型参数。

本文系统主要由 1 个云端节点和多个边缘节点组成，各节点功能分工如下。云端节点：配备高性能 GPU 集群与 PB 级存储设备，负责跨场景通用基础模型的训练与迭代，以及边缘侧上传增量数据的融合处理；边缘节点：部署于工业生产现场，采用嵌入式计算单元，负责实时数据采集、本地模型适配、增量更新及推理任务执行，仅向云端上传经筛选的增量数据特征。

3.2. 云端模型训练

云端基础模型是边缘模型适配的基础，通过融合多个边缘节点的初始数据，可以有效提升模型的跨场景泛化能力，主要包括数据采集和模型训练两个环节。

(1) 数据采集

首先从各个边缘节点收集原始工业图像数据，需要涵盖工业质检任务所需的全类型缺陷(如裂纹、划痕、变形、缺件等)，构建模型训练所需的初始数据集 D_0 。由于各个工业视觉原件采集到的图像数据有较大的差异性，需要对数据进行标准化的预处理，包括数据归一化、数据增强和数据标注。将图像尺寸统一调整为模型训练所需的大小，然后将像素值归一化至 $[0, 1]$ 区间，消除前端图像采集装置导致的尺寸与亮度等差异对模型训练的影响。同时组合采用随机旋转、水平翻转、高斯噪声等数据增强策略，扩充数据集并降低过拟合风险。最后需要利用统一的缺陷类别标签对数据样本进行标注。

(2) 模型训练

云端基础模型主要采用单阶段目标检测架构, 兼顾工业场景中的检测速度与精度, 适配实时性需求, 模型结构分为三部分: 骨干网络: 采用 CSPDarknet 结构, 通过残差连接与跨层特征融合, 在不同粒度上提取图像特征, 为后续检测提供通用特征基础。颈部网络: 采用 PANet 结构, 通过自上而下的特征传递与自下而上的特征融合, 整合骨干网络输出的多尺度特征, 提升小尺寸缺陷的检测能力。头部网络: 包含分类分支与回归分支, 分类分支输出缺陷类别概率, 回归分支输出缺陷边界框坐标与置信度, 最终生成检测结果。

本文主要基于工业数据特征, 对模型训练过程中的损失函数进行改进, 提出加权焦点损失(Weighted Focal Loss, WFL)函数, 在传统焦点损失的基础上引入边缘节点性能权重。

$$\text{WFL}_k(p_t) = -(1-p_t)^{\frac{F_{1k}}{\bar{F}_1}} \log(p_t)$$

其中 $\text{WFL}_k(p_t)$ 表示第 k 个边缘模型上类别 t 的损失函数, p_t 表示该类别预测的置信度, 置信度越高产生的损失越少, \bar{F}_1 表示各个边缘模型宏观 F_1 得分的平均值, F_{1k} 表示第 k 个边缘模型的宏观 F_1 得分。若边缘模型 F_1 得分较高, 那么难易样本带来的损失差异较小, 反之, 难以样本带来的损失差异较大。通过引入带权聚焦损失函数作为模型训练的损失函数, 一方面通过 $(1-p_t)$ 项降低易分类样本的损失权重, 迫使模型聚焦难分类样本, 平衡难易样本带来的影响, 解决边缘节点内的样本不平衡问题。另一方面, 通过 $\frac{F_{1k}}{\bar{F}_1}$ 解决边缘节点间的不平衡问题, 若某边缘节点的 F_{1k} 较低, 则 $\frac{F_{1k}}{\bar{F}_1}$ 值较小, 难分类样本的损失权重进一步放大, 使云端训练过程更关注低精度节点的数据, 平衡各节点的模型性能。

3.3. 边缘模型适配

云端基础模型需要在边缘侧个性化工业数据集中进行本地化适配, 利用边缘侧自有小样本数据, 快速优化模型以适配本地检测需求, 同时避免边缘侧全量训练的算力消耗, 主要包括模型下发和迁移学习两个环节。云端将训练完成的基础模型参数压缩后下发至各边缘节点, 降低传输带宽需求。边缘侧接收模型后, 冻结骨干网络的所有参数, 仅保留颈部网络与头部网络的参数可更新, 减少边缘侧训练参数量, 降低算力消耗。同时边缘节点采用自有小样本数据作为训练数据集, 采用传统焦点损失作为训练损失函数, 通过梯度下降法训练颈部与头部网络, 训练完成后, 得到适配该边缘节点业务场景的边缘模型。

3.4. 边缘模型更新

在工业场景应用过程中, 边缘节点会实时生成新的增量数据, 为了避免模型因数据分布变化导致性能漂移, 本文设计了双时间周期增量更新机制, 实现边缘模型的持续优化。短周期(T_{edge})为边缘模型的增量更新周期, 通过根据业务需求进行设定, 在该时间周期, 边缘节点仅更新本地模型, 不向云端上传数据。长周期(T_{cloud})为边缘节点向云端上传增量数据的周期, 满足 $T_{\text{cloud}} = n * T_{\text{edge}}$ 。其中 n 为正整数, 当边缘节点每完成 n 次本地增量更新后, 向云端上传一次汇总的增量数据。通过双周期设计, 可以实现边缘模型的快速迭代, 及时适配本地数据变化; 另外也有助于减少数据上传频次, 降低网络传输压力。

在每个短周期结束时, 边缘节点依次完成增量数据采集、模型训练和模型切换。增量数据采集过程中收集该周期内生成的所有新的样本数据, 形成增量数据集 $d_{t,k}$, 其中 t 为周期序号, k 为边缘节点编号。在模型训练过程中首先需要冻结骨干网络与颈部网络参数, 仅更新头部网络参数, 然后以增量数据 $D_{t,k}$ 为输入, 以焦点损失为损失函数, 训练得到更新后的边缘模型 θ'_k 。最后完成模型切换, 将当前推理任务的模型从 θ_k^{-1} 切换为 θ'_k 。

在每个长周期结束时, 边缘节点依次完成数据预处理, 特征提取和数据上传。首先需要对数据进行

预处理，将 n 个 $d_{i,k}$ 进行合并，去除重复与低质量样本，得到增量数据集 $D_{i,k}$ 。特征提取过程利用边缘模型 θ_k^i 的骨干网络提取 $D_{i,k}$ 的特征向量，最终只上传特征向量而非原始图像，提升数据传输效率。最后将 $D_{i,k}$ 对应的特征向量和标签，通过本系统上传至云端节点。

3.5. 云端模型更新

云端节点接收各边缘节点上传的 $D_{i,k}$ 后，将增量数据与数据集 D_{T-1} 进行融合得到数据集 D_T ，通过采用带权聚焦损失对模型进行重新训练，得到更新后的云端基础模型 Θ_T ，整体流程如图 2 所示。

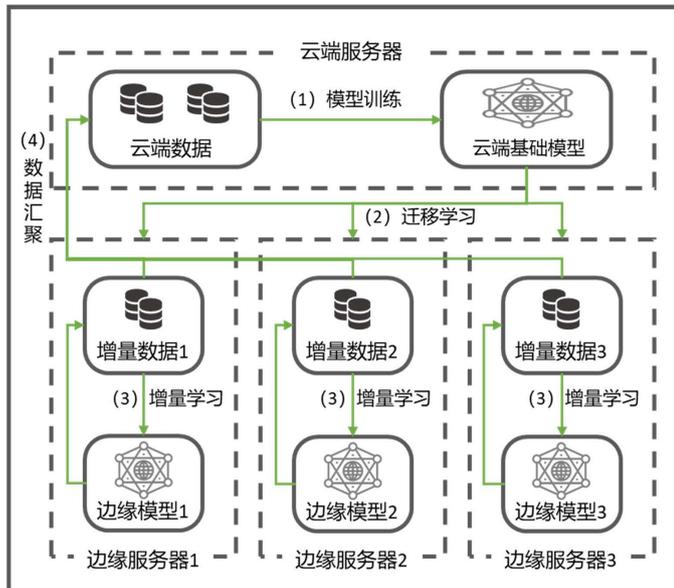


Figure 2. Schematic diagram of cloud-edge collaborative incremental learning process
图 2. 云边协同增量学习流程示意图

4. 总结与展望

本文针对工业质量检测中模型泛化性、精度与时延难以兼顾的核心问题，提出一种云边协同增量学习方法。通过构建云端/边缘分层架构，云端聚焦跨场景泛化能力提升，边缘侧专注本地场景适配与实时推理，对泛化性、精度和时延均进行了优化。此外，提出加权焦点损失函数，平衡节点间性能不平衡与节点内样本不平衡问题，增强云端基础模型的跨场景适配能力。基于双时间周期增量更新机制，通过短周期实现边缘模型的实时优化，通过长周期降低数据上传开销，适配工业数据动态生成特性，通过云边协同实现模型持续优化，有效抵抗数据分布漂移。后续研究需要考虑多模态数据融合，提升复杂缺陷检测能力，结合大模型的通用表征能力，探索云端大模型预训练，边缘小模型微调的协同模式，进一步提升模型的小样本适配能力与泛化性能。

基金项目

泰山产业领军人才项目(No. tscx202312006); 山东省博士后创新项目(No. SDCX-ZG-202400307)。

参考文献

- [1] 王耀南, 陈铁健, 贺振东, 等. 智能制造装备视觉检测控制方法综述[J]. 控制理论与应用, 2015, 32(3): 273-286.
- [2] 李苒笙, 李志强, 贾北洋. 基于大模型的工业质检系统关键技术及应用[J]. 广东通信技术, 2025, 45(5): 35-39+44.

-
- [3] 董文轩, 梁宏涛, 刘国柱, 等. 深度卷积应用于目标检测算法综述[J]. 计算机科学与探索, 2022, 16(5): 1025-1042.
 - [4] 杜松杰, 李诺亚, 潘卫国, 等. 基于反射图引导和交叉注意力机制的低光照目标检测方法[J]. 云南大学学报(自然科学版), 2025(47): 1-10.
 - [5] 钟炜强. 工业产品表面质量检测智能算法研究与系统实现[D]: [硕士学位论文]. 北京: 北京邮电大学, 2023.
 - [6] 范存君, 金松燕, 金楠, 等. 基于深度学习的裂缝识别与量化分析研究[J]. 工业建筑, 2024, 54(8): 126-132.
 - [7] 何毓芬, 何赞泽, 尹勇, 等. 大模型在工业缺陷检测领域的应用现状与展望[J]. 仪器仪表学报, 2025, 46(10): 22-41.
 - [8] 吴子强, 刘敏, 张义伟, 等. 深度学习模型的云边协同训练和部署[J]. 电子产品世界, 2022, 29(12): 42-45.
 - [9] 马昊. 基于云边协同的光伏组件缺陷检测应用[J]. 通信企业管理, 2024(10): 76-80.
 - [10] 梁道君, 袁东风, 张海霞, 张明高. LightDets: 端边云协同的产品质量在线检测方法[J]. 计算机集成制造系统, 2025, 31(9): 3187-3196.
 - [11] 武星, 殷浩宇, 姚骏峰, 等. 基于增量学习的工业软件: 转炉图像异常识别系统[J]. 武汉大学学报(理学版), 2024, 70(3): 293-301.