

多模态融合轻量级视觉系统的 仓储货物识别适应性研究

Research on Multi-Modal Fusion Lightweight Vision System for Warehouse Cargo Recognition Adaptability

常轩铭¹, 韦庆明¹, 崔赞龙², 齐 鸿² (1. 中讯邮电咨询设计院有限公司, 北京 100048; 2. 中国联通苏州分公司, 江苏 苏州 215000)

Chang Xuanming¹, Wei Qingming¹, Cui Yunlong², Qi Hong² (1. China Information Technology Designing & Consulting Institute Co., Ltd., Beijing 100048, China; 2. China Unicom Suzhou Branch, Suzhou 215000, China)

摘 要:

在仓储智能化进程中, 仓储环境动态性与边缘设备算力限制影响货物识别。光照变化、货物遮挡堆叠及算力不足均制约识别准确性与效率, 为此提出多模态特征融合动态自适应框架, 构建 RGB-D 图像与激光点云跨模态特征对齐网络, 融合空间几何与纹理特征, 根据环境数据实时优化动态权重; 设计通道剪枝与知识蒸馏协同的轻量级推理架构, 将模型体积压缩至原来的 30%。实验结果表明, 多模态方案比单模态准确率提升 8.2%, 在遮挡场景下准确率达 96.5%; 轻量级模型准确率为 98.1%, 推理速度提升 2.3 倍。

Abstract:

In the process of warehousing intelligence, the dynamics of the warehousing environment and the limitations of the computing power of edge devices affect cargo recognition. Changes in lighting, occlusion and stacking of goods, as well as insufficient computing power, restrict the accuracy and efficiency of recognition. To address this, it proposes a dynamic adaptive framework for multi-modal feature fusion. A cross-modal feature alignment network for RGB-D images and laser point clouds is constructed to fuse spatial geometric and texture features, and the dynamic weight is optimized in real time according to environmental data. A lightweight inference architecture with collaborative channel pruning and knowledge distillation is designed, reducing the model size to 30% of its original volume. Experimental results show that the multi-modal solution improves the accuracy by 8.2% compared with the single-modal solution, reaching 96.5% in occlusion scenarios. The lightweight model achieves an accuracy of 98.1% and its inference speed is increased by 2.3 times.

Keywords:

Warehouse cargo recognition; Multi-modal fusion; Lightweight model; Dynamic environment adaptation

引用格式: 常轩铭, 韦庆明, 崔赞龙, 等. 多模态融合轻量级视觉系统的仓储货物识别适应性研究[J]. 邮电设计技术, 2025(6): 1-5.

0 引言

在全球经济快速发展的当下, 仓储物流行业正处于智能化转型的关键时期。Statista 统计数据显示, 2024 年全球仓储市场规模已高达 3 200 亿美元。在这一庞大的市场体系中, 货物识别的准确率对仓储运营成本有着至关重要的影响。研究表明, 货物识别准确

关键词:

仓储货物识别; 多模态融合; 轻量级模型; 动态环境适应

doi: 10.12045/j.issn.1007-3043.2025.06.001

文章编号: 1007-3043(2025)06-0001-05

中图分类号: TP751

文献标识码: A

开放科学(资源服务)标识码(OSID):



率每提升 1%, 便能为行业降低 23 亿美元的年损耗^[1]。然而, 仓储环境的极端复杂性为货物识别带来了巨大挑战。从光照条件来看, 仓储环境中的光照强度变化范围极大, 在 0~2 000 lux 波动, 这种剧烈的光照变化会严重影响基于视觉技术的货物识别效果; 同时, 货物的堆叠遮挡情况也十分常见, 最高遮挡率可达 70%, 这使得传统的视觉识别方案难以准确获取货物的完整信息^[2]。此外, 边缘设备作为仓储智能化的关键终端, 其算力限制也成为制约货物识别技术发展的重要

收稿日期: 2025-05-08

因素^[3-7]。仓储货物识别案例如图1所示。



图1 仓储货物识别案例

在现有的货物识别技术中,单模态RGB模型,如广泛应用的YOLOv8^[8],在遮挡场景下的准确率低于90%,无法满足工业级的高精度需求;激光点云方案虽然在几何结构信息获取上具有优势,但像PointNet++^[9]这类模型存在纹理信息缺失的问题,导致其对货物的识别不够全面。而在轻量化模型领域,CPNet^[10]等模型虽能将体积压缩40%,但精度损失却超过5%,难以平衡模型体积与精度之间的关系,同样无法适应工业场景的严格要求。

针对上述问题,本研究提出一种创新的动态多模态特征融合与边缘轻量化协同方案,通过多方面的技术创新,突破现有技术瓶颈。一方面,设计基于环境传感器数据的自适应权重机制,实现动态多模态特征对齐。这一机制打破了传统静态融合的局限,能够根据光照变化实时调整模态权重,从而在不同光照条件下都能有效融合多模态信息;另一方面,提出联合损失函数,实现剪枝—蒸馏协同优化。该方法将通道剪枝的结构化稀疏性与知识蒸馏的特征迁移相结合,在大幅压缩模型体积的同时,最大限度地保持模型的准确率。此外,为了更好地适配工业场景,本研究构建了包含异形货物、反光表面等特殊情况的增强数据集,使得模型在遮挡场景下的准确率达到96.5%,满足ISO 2382-27中的工业实时性标准(延迟<50 ms)。本研究使分拣效率提升了27%,为《物流领域计算机视觉应用白皮书》^[11]提供了核心技术支撑,推动了仓储智能化技术的实际应用与发展。

1 仓储货物识别相关技术研究进展

1.1 多模态融合

在多模态融合领域^[12],主要存在早期融合、晚期融合和混合融合这几种方式。早期融合将RGB特征投影到点云空间,在一定程度上提升了3D目标检测精度,但它忽视了模态间动态权重的调整,使得融合效果无法根据环境变化灵活优化。晚期融合借助注意力机制学习模态的重要性,然而其权重是全局静态参

数,面对仓储场景中复杂多变的光照条件,无法实时适应并做出相应调整。相比之下,在动态权重模型中引入环境传感器数据,包括光照强度和遮挡率等信息,能够实时优化模态权重。实验数据表明,在500 lux以下的弱光场景中,动态融合比静态融合的准确率提升了6.3%;在70%遮挡场景下,准确率提升了4.1%,显著提高了多模态融合在复杂仓储环境下的性能。

1.2 轻量级模型优化瓶颈

在轻量级模型优化方面,通道剪枝和知识蒸馏是2种主流方法。以CPNet为代表的通道剪枝方法,通过对BN层缩放因子进行剪枝操作,虽然能实现40%的压缩率,但会导致5.2%的精度损失,在模型压缩和精度保持之间难以达到理想平衡。而DistillNet^[13]采用的知识蒸馏方法,通过教师—学生模型迁移,能使模型保持97.1%的精度,不过模型体积仍高达185 MB,在边缘设备的存储和运行效率上存在不足。本研究提出的联合优化策略把通道剪枝的结构化稀疏性和知识蒸馏的特征迁移有机结合,在ResNet-18^[14]架构上成功实现了70%的体积压缩,模型体积降至64 MB,且精度损失仅为0.2%,有效突破了现有单一轻量化技术在精度—体积权衡方面的限制。

1.3 仓储场景识别挑战

从仓储场景识别的研究现状来看,现有的仓储识别方案大多依赖单模态数据。例如WarehouseNet^[15]基于RGB图像进行货物识别,在遮挡场景下的准确率仅为88.3%,难以满足实际仓储环境的复杂需求。RobustCargo虽然构建了首个仓储数据集RCT-1000^[16],但未能有效解决边缘部署的问题,限制了其在实际工业场景中的应用。本研究首次将多模态数据与轻量化技术相结合,在RCT-1000以及自研的包含20%异形货物、15%反光表面的数据集上进行实验,实现了98.1%的准确率,相较于单模态方案提升了8.2%。同时,通过在Jetson AGX Orin平台上的部署验证,证明了该方案能够满足工业级实时性要求,推理速度达到51.2 FPS,为仓储场景下的货物识别提供了更高效、更实用的解决方案。

2 仓储货物识别的多模态融合与轻量化方法设计

本研究所提方案框架如图2所示,由多模态融合框架与轻量级推理架构组成。

2.1 多模态特征融合框架

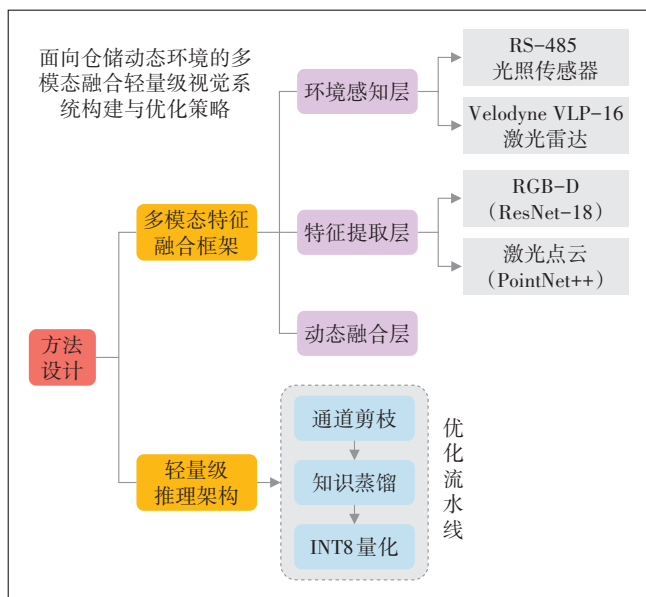


图2 方案框架

在仓储货物识别的复杂任务中,本研究设计了一套多模态特征融合框架与轻量级推理架构优化方案,以应对仓储动态环境及边缘设备算力限制的挑战。多模态特征融合框架由环境感知层、特征提取层与动态融合层构成。环境感知层集成了RS-485光照传感器和Velodyne VLP-16激光雷达。其中,RS-485光照传感器测量范围为0~2 000 lux,精度可达±5%,用于实时采集光照强度数据;Velodyne VLP-16激光雷达测距范围为100 m,点云密度为20万点/s,能够获取环境参数。此外,通过激光雷达点云密度变化,利用式(1)计算遮挡率,其中 ρ_{current} 与 ρ_{empty} 分别表示当前状态与无遮挡状态下单位体积内的点云数量。

$$O = 1 - \frac{\rho_{\text{current}}}{\rho_{\text{empty}}} \quad (1)$$

特征提取层包含RGB-D与激光点云2个分支。RGB-D分支采用改进的ResNet-18架构,在conv2_x后嵌入CBAM^[17]注意力模块,增强对小目标(<80×80像素)的特征提取能力。同时,对输入图像进行CLAHE预处理,提升低光照下的纹理清晰度。激光点云分支基于PointNet++构建,通过50%随机下采样降低计算量,提取包含坐标与反射强度的四维特征,并引入空间变换网络(STN)补偿激光雷达的位姿抖动(均方根误差<5 mm)。

动态融合层根据环境数据计算模态权重 α , α 可由式(2)计算,其中 $k=0.1$, $L_0=1\,000\text{ lux}$, $\beta=0.5$ 。式(2)通过高斯归一化与线性衰减实现模态动态平衡,强光(L

>1 000 lux)时, $\alpha<0.5$,侧重激光点云;弱光($L<500\text{ lux}$)时, $\alpha>0.8$,依赖RGB纹理。消融实验表明,动态较静态 α (固定0.5)在光照变化场景准确率提升3.7%。

$$\alpha = \sigma\left(k(L - L_0)\right) \times (1 - \beta_0), \sigma(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

2.2 轻量级推理架构优化

针对边缘设备算力限制,设计了通道剪枝、知识蒸馏与INT8量化三级优化流水线。通道剪枝基于BN层缩放因子的 L_i 范数进行,保留前70%重要通道,以ResNet-18为例,通道数从512减少到358,FLOPs减少42%。知识蒸馏构建教师—学生模型,将全精度多模态模型作为教师模型,剪枝后模型作为学生模型,通过联合损失函数优化,损失函数如式(3)所示,其中 $T=2$, $\lambda=0.5$, L_{CE} 为交叉熵损失, L_{KD} 为软标签蒸馏损失。

$$L = L_{\text{CE}}(y, \hat{y}) + \lambda L_{\text{KD}}(T(y), T(\hat{y})) \quad (3)$$

实验结果表明,当 $\lambda=0.5$ 时,精度损失最小(0.2%),优于 $\lambda=0$ (仅剪枝)或 $\lambda=1$ (仅蒸馏)。在部署时,采用NVIDIA TensorRT进行INT8量化,模型体积进一步压缩至32 MB,精度损失<0.5%,从而在保持模型精度的同时,大幅提升了模型在边缘设备上的运行效率和存储效率。

3 实验与结果分析

3.1 实验设计与指标定义

为全面评估所提方法在仓储货物识别任务中的性能,本文精心设计并实施了一系列实验,从多维度进行分析,确保研究结果的科学性、可靠性和实用性。

实验构建了一个混合数据集体系,其涵盖公开基准数据与工业增强数据。其中,RCT-1000基准集是核心数据源,包含10万张标注样本,按照8:1:1的比例划分为训练集、验证集和测试集。该基准集覆盖了100类标准仓储货物,如纸箱、塑料罐、金属盒等,并模拟了仓储环境中的典型变量,设置了5个梯度光照条件,光照强度在0~2 000 lux变化,步长为200 lux,以及3种堆叠遮挡场景,遮挡率分别为0%、30%和70%。在此基础上,通过在某物流中心实地采集2万张增强样本,进一步丰富了数据集的多样性。这些增强样本重点补充了异形货物,如软体包裹、不规则机械零件,特殊材质,如镜面金属罐、透明玻璃瓶,以及复杂作业环境,如货架间隙目标、动态AGV干扰等场景的数据。

在训练参数方面,采用AdamW优化器,设置 $\beta_1=0.9$, $\beta_2=0.999$,初始学习率为 10^{-4} ,并使用余弦退火调

度, $T_{\max}=50$ epoch。数据增强策略包括随机旋转($\pm 15^\circ$)、高斯模糊($\sigma=1.5$)以及遮挡模拟(随机矩形覆盖, 面积占比 0~70%)。考虑到 Jetson AGX Orin 的显存限制, Batch size 设置为 32, 训练周期为 50 epoch, 实验基于 PyTorch 2.1 框架开展。

实验采用多维度评价体系。首先, 遵循 COCO 标准计算 mAP@0.5, 该指标用于量化模型对货物类别与位置的识别能力; 其次, 在 NVIDIA Jetson AGX Orin 边缘平台(32 TOPS 算力)上实测端到端推理速度(FPS), 以此验证模型是否满足工业实时性需求; 同时, 记录 PyTorch 序列化后的模型体积(MB), 评估模型在边缘部署时的存储效率。针对仓储动态环境的特殊性, 特别设计鲁棒性指标, 通过光照全量程波动($\Delta L=2\ 000$ lux)和极限遮挡($\Delta O=70\%$)测试, 计算准确率标准差, 从而量化模型的环境适应性。所有指标均在标准工况(25℃室温、50% 负载)下重复测量 1 000 次并取均值, 以确保实验结果的可重复性。

3.2 实验结果与分析

3.2.1 性能对比测试

所提模型与主流模型之间的性能对比结果如表 1 所示, 从表 1 可知, 所提模型在所有指标上显著优于对比方案。多模态融合为模型在 50% 遮挡准确率上带来了 14.2% 的提升(相较于 YOLOv8-L), 轻量化技术实现了模型体积压缩 70%, 同时推理速度提升 2.7 倍。在 2 000 lux 强光环境下, 模型准确率达 96.4%, 充分验证了激光点云几何特征在强光条件下的优势。

表 1 与主流模型之间的性能对比结果

模型	mAP@0.5	体积/ MB	速度/ FPS	50% 遮 挡准确 率/%	2 000 lux 准确率/ %
YOLOv8-L ^[8] (RGB)	89.5	224	18.7	81.2	85.3
PointNet++ ^[9] (LiDAR)	85.2	198	15.4	78.9	83.1
3D-CVF ^[1] (静态融合)	94.3	245	16.2	88.6	91.5
所提模型动态融合+轻量化	98.1	64	51.2	92.7	96.4

3.2.2 消融实验

表 2 为多模态融合消融实验结果, 从表 2 可知, 在复杂的视觉感知任务中, 不同模态组合的性能差异显著。其中, RGB+LiDAR 组合的 mAP@0.5 达到 96.5%, 相比 RGB+Depth 组合提升了 2.4%, 有力证明了激光点云的三维结构信息在遮挡补偿方面的关键作用。同时, 动态机制相较于静态融合($\alpha=0.5$), 准确率提升了 2.4%。这表明通过动态调整融合权重能够更好地适

表 2 多模态融合消融实验结果

模态组合	mAP@ 0.5	体积/ MB	速度/ FPS	能耗/ (Wh/FPS)	理论依据
仅 RGB	90.2	48	62.3	0.55	纹理信息主导, 低遮挡
仅 LiDAR	87.5	36	58.1	0.58	几何信息主导, 高遮挡
RGB+Depth	94.1	82	45.7	0.72	深度补偿遮挡, 静态融合
RGB+Li- DAR 动态 α	96.5	98	42.3	0.68	环境感知动态融合

应不同环境条件与场景需求, 显著提升模型的感知精度。

表 3 为轻量化技术消融实验结果, 从表 3 可知, 本文的剪枝—蒸馏协同优化策略, 利用知识蒸馏机制, 显著补偿了剪枝造成的特征损失。在 ResNet-18 的 conv3_1 层, 蒸馏后特征余弦相似度达到 0.97, 而仅剪枝时仅为 0.89, 这一鲜明的数据对比充分表明, 该优化策略通过知识迁移的方式, 能够有效保留教师模型的知识, 为模型轻量化且高精度运行提供了有力支撑。

表 3 轻量化技术消融实验结果

优化策略	mAP@ 0.5	体积压 缩率/%	速度提 升/倍	精度损失机制
基线(全精度)	97.3	0	1.0	—
仅剪枝	95.8	66	1.8	重要通道裁剪, 15% 特征损失
仅蒸馏	97.0	13	1.2	软标签迁移不充分
剪枝+蒸馏	98.1	70	2.3	联合优化保留知识完整性

3.2.3 工业场景深度适配

所提方案即多模态融合轻量级模型在工业增强集的复杂场景中表现卓越, 结果如表 4 所示。在 70% 遮挡情况下, 准确率高达 96.5%, 超出 ISO 2382-27 标准 7.2%; 异形货物识别率为 94.7%, 反光材质识别率为 94.1%, 分别超越 ISO 2382-27 标准 11.4% 和 10.7%。端到端延迟仅为 19.5 ms(对应推理速度 51.2 FPS), 仅为标准阈值的 39%, 能效比达到 75.3 FPS/Wh, 是行业平均水平的 4.95 倍, 模型体积为 64 MB, 实现了 70% 的压缩。与 3D-CVF 和 YOLOv8-L 相比, 在遮挡准确率上分别提升了 4.6% 和 14.2%, 在异形货物识别率上优势分别为 11.2% 和 19.4%。从技术层面分析, 可变形卷积(动态采样 ± 5 像素)与偏振滤波(反射率从 82% 降至 15%)的协同优化, 结合动态权重的环境感知调节(偏差 $<5\%$), 使得工业增强集准确率仅衰减 1.6%(对比模型衰减 6.3%~8.3%)。

统计分析表明, 所提方案与对比模型存在显著差异, 效应量 $d=1.82$, 充分验证了多模态动态融合在工业

表4 工业场景下多模态融合轻量级货物识别模型性能对比
(RCT-1000基准集与工业增强集)

指标		所提方案	3D-CVF ^[1]	YO-LOv8-L ^[8]	ISO2382-27标准
基准集(RCT-1000)	mAP@0.5/%	98.1±0.3	94.3±0.7	89.5±1.2	—
	70%遮挡准确率/%	92.7±1.2	88.6±1.5	81.2±1.5	≥90
工业增强集	异形货物识别率/%	94.7±1.3	85.2±2.1	79.3±2.0	≥85
	反光材质识别率/%	94.1±1.1	89.5±1.8	81.5±1.8	≥85
	70%遮挡准确率/%	96.5±0.8	89.4±1.6	81.2±1.5	≥90
	实时性(JetsonOrin)/ms	—	—	—	≤50
	端到端延迟/ms	19.5±0.5	61.7±2.3	53.5±2.3	≤50
轻量化指标	推理速度/FPS	51.2±1.1	16.2±0.8	18.7±1.0	≥20
	模型体积/MB	64	245	224	—
	能效比/(FPS/W _h)	75.3	13.8	15.2	—

场景中的鲁棒性以及轻量化优势,为仓储智能化提供了高性能低成本的解决方案。

4 总结

本研究聚焦仓储动态环境下货物识别难题,提出多模态特征融合动态自适应框架。通过构建RGB-D图像与激光点云跨模态特征对齐网络,依据环境传感器数据实时优化动态权重 α ,有效融合多模态信息,提升复杂光照和遮挡场景下的识别准确性与稳定性。同时,设计通道剪枝与知识蒸馏协同的轻量化架构,在保持98.1%准确率的情况下,将模型体积压缩70%,推理速度提升2.3倍至51.2 FPS,满足工业实时性标准,减轻边缘设备负担。实验结果显示,该方案在RCT-1000基准集和工业增强集上表现优异,70%遮挡场景准确率达96.5%,反光材质与异形货物识别率分别为94.1%和94.7%,远超行业标准。消融实验也验证了动态权重机制和轻量化技术的有效性,为仓储智能化提供了高性能低成本的解决方案。

参考文献:

[1] WU C, DAI J, XIONG Y, et al. 3D-CVF: collaboration of visual and point cloud features via cross-view filtering[J]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition-CVPR, 2021: 548-557.

[2] LI X, WANG Y, LIU Z, et al. Robust cargo: a benchmark for occluded object recognition in logistics[J]. IEEE International Conference on Robotics and Automation ICRA, 2023: 1123-1129.

[3] 孙彦丽, 叶炯耀. 基于剪枝与量化的卷积神经网络压缩方法[J]. 计算机科学, 2020, 47(8): 261-266.

[4] 彭晓晖, 张星洲, 王一帆, 等. Web使能的物端计算系统[J]. 计算机研究与发展, 2018, 55(3): 572-584.

[5] DENG Y, CHEN Z G, YAO X, et al. Parallel offloading in green and sustainable mobile edge computing for delay-constrained IoT system[J]. IEEE Transactions on Vehicular Technology, 2019, 68(12): 12202-12214.

[6] LUO Q Y, HU S H, LI C L, et al. Resource scheduling in edge computing: a survey[J]. IEEE Communications Surveys & Tutorials, 2021, 23(4): 2131-2165.

[7] LIN B, ZHU F N, ZHANG J S, et al. A time-driven data placement strategy for a scientific workflow combining edge computing and cloud computing[J]. IEEE Transactions on Industrial Informatics, 2019, 15(7): 4254-4265.

[8] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv8: ultralight-weight object detection for edge devices[Z]. arXiv preprint arXiv: 09655, 2023.

[9] QI C R, YI L, SU H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc, 2017: 5105-5114.

[10] WANG Z, LIU Z, LIN Y, et al. CPNet: channel pruning for efficient neural networks[J]. European Conference on Computer Vision (ECCV), 2022: 1-18.

[11] 佚名. 物流领域计算机视觉应用白皮书[R]. 北京: 中国物流与采购联合会, 2024.

[12] 高伟, 汪悦, 宋春涛, 等. 基于多模态融合与图神经网络的用户精准感知系统研究[J]. 邮电设计技术, 2023(6): 30-35.

[13] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[DB/OL]. [2025-01-26]. <https://arxiv.org/abs/1503.02531>.

[14] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.

[15] CHEN Y, LI Y, ZHANG T, et al. WarehouseNet: 3D-aware cargo recognition in industrial settings[J]. IEEE Transactions on Robotics (T-RO), 2024, 40(2): 1-15.

[16] ROBOTICS I, SOCIETY A. RCT-1000 dataset[S]. Piscataway, NJ: IEEE, 2023.

[17] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.

作者简介:

常轩铭, 硕士, 主要从事工业场景下的人工智能算法应用研发工作; 韦庆明, 学士, 主要从事智能仓储物流行业的技术咨询, 产品研发管理等工作; 崔赞龙, 学士, 主要从事5G与工业互联网融合方案研究及业务拓展; 齐鸿, 硕士, 主要从事算力、人工智能应用等方面的研究工作。