

基于MEAL-YOLOv8的助餐机器人食物种类识别

任诗扬, 吴伟铭, 胡冰山, 喻洪流

(上海理工大学智能康复工程研究院, 上海 200093)

摘要: 助餐机器人在辅助老年人和上肢残疾人方面展现出巨大潜力。助餐机器人通常体积小、具备良好的移动便捷性, 因此必须采用轻量化设计, 在有限的空间内高效运作并节省能耗。为此, 提出MEAL-YOLOv8轻量化视觉模型实现更高效的食物分类功能, 结合了BRA与MSDA两种注意力机制, 融合了YOLOv8与MobileNetV3的优点。在JET-SON NANO B01平台上部署后, MEAL-YOLOv8模型大小仅为2.37 MB, 内存占用为158 MB, 分类准确率达到91.4%, 具有高精度、低延迟的视觉识别能力, 使助餐机器人能在资源受限的环境中高效运作, 能实时、准确地识别食物。

关键词: 模型轻量化; YOLOv8; 助餐机器人; 食物识别

DOI: 10.11907/rjtk.251103

中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2026)004-0020-07

扫描二维码阅读全文:



Food Category Recognition for Meal-Assistance Robots Using MEAL-YOLOv8

REN Shiyang, WU Weiming, HU Bingshan, YU Hongliu

(Institute of Intelligent Rehabilitation Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Catering robots have shown great potential in assisting the elderly and people with upper limb disabilities. Food assistance robots are usually small in size and require good convenience and mobility, so they must adopt lightweight design to operate efficiently and save energy in limited space. Therefore, a lightweight visual model of MEAL-YOLOv8 is proposed to achieve more efficient food classification function, combining BRA and MSDA attention mechanisms, and integrating the advantages of YOLOv8 and MobileNetV3. After deployment on the JET-SON NANO B01 platform, the MEAL-YOLOv8 model has a size of only 2.37 MB, a memory usage of 158 MB, and a classification accuracy of 91.4%. It has high-precision and low latency visual recognition capabilities, allowing the food aid robot to operate efficiently in resource limited environments and recognize food in real-time and accurately.

Key Words: model lightweight; YOLOv8; meal-assistance robots; food recognition

0 引言

全球人口老龄化问题日益严峻, 2050年全球老年人口将超过16亿, 超过10—24岁青少年的数量^[1-2]。随着年龄的增长和身体机能下降, 一些老年人失去了独立进食的能力。此外, 脊髓损伤、运动神经元疾病及其他上肢功能障碍等情况, 也使得部分人群无法自主进食, 极大影响了生活质量。由于目前家庭治疗师和专业护理人员短缺, 助餐机器人便成为帮助这些群体独立进食的有效手段^[3]。

助餐机器人已从最初的基础功能发展到如今的智能交互。早期, Handy1、My Spoon、Winsford Feeder和Obi等机器人主要依赖于用户操作, 例如通过按钮、肢体动作或下巴控制机器人完成进食功能^[4-6]。为了提升进食体验的个性化和直观性, 研究人员开始探索视觉交互技术, 针对食物位置识别和食物质地反馈的研究已取得显著进展^[7-11]。例如, Mask R-CNN框架已被成功应用于精确定位食物区域, 但在食物分类方面仍存在明显不足, 仅将食物类别简化为基础分类体系, 即将食物简单地划分为块状(如泡菜)、碎片状(如萝卜片)等基本形态类别^[7]; HAL-UL

收稿日期: 2025-02-18

基金项目: 国家重点研发计划项目(2022YFC3601400)

作者简介: 任诗扬(2000-), 女, 上海理工大学智能康复工程研究院硕士研究生, 研究方向为基于机器视觉的智能人机交互; 吴伟铭(1997-), 男, 上海理工大学智能康复工程研究院博士研究生, 研究方向为大模型与人工智能; 胡冰山(1982-), 男, 博士, 上海理工大学智能康复工程研究院教授、博士生导师, 研究方向为康复机器人智能控制、柔性驱动机制的设计与控制、轻量级协作机械臂; 喻洪流(1966-), 男, 博士, 上海理工大学智能康复工程研究院教授、博士生导师, 研究方向为人类仿生力学与智能控制、康复机器人、人与机器智能交互。本文通讯作者: 喻洪流。

的 HSV 分割颜色空间方法依赖人工特征设计, 在强光照下分割会导致准确率下降^[8]; 现有深度学习模型如 DCF-YOLOv8s 受限于 4 类粗分类: 非牛顿流体粘性餐食、块状餐食、非牛顿流体-固体交互餐食、切块混合餐食, 缺乏供使用者选择的细分食物种类^[9]。

基于深度学习的计算机视觉算法是提升食物识别精度的有效途径。相较于传统的物体检测算法而言, 深度学习算法在特征提取能力方面优势明显^[12-14]。YOLO 算法经过多次迭代演化, 在统一框架内实现了速度和精度的平衡, 非常适合在实时场景中高效、准确地识别不同大小的食物^[15-19]。然而, 深度学习算法依赖复杂的计算, 功耗高、热量高, 因此高性能计算单元需要复杂的电源管理和散热解决方案, 无形中增加了设计的复杂性和成本。对于依赖电池供电并有尺寸限制的移动助餐机器人而言, 食物识别算法需要具备低计算复杂度、高准确度、实时性和易于部署的特点。

为此, 本文提出一种轻量化的食物分类识别算法 MEAL-YOLOv8, 以便于在嵌入式开发板上部署助餐机器人系统。同时, 为了提升模型处理效率、推理速度, 减少参数量和计算负载, 将 MobileNetV3 模型集成到 YOLOv8 架构中, 替换原有的 Darknet-53 主干网络, 并在融合算法中添加双层路由注意力 (Bi-Level Routing Attention, BRA) 和多

尺度膨胀注意力 (Multi-Scale Dilated Attention, MSDA)^[20-21]。

1 研究方法

1.1 MEAL-YOLOv8 主干网络

考虑到经济便捷的助餐机器人计算资源有限, 因此选择更轻量化的 MobileNetV3 作为 MEAL-YOLOv8 主干网络, 如图 1 所示。由此可见, Backbone 部分已替换为 MobileNetV3 的主干网络, 可减少卷积和计算量, 使 YOLOv8 更轻量化。

Backbone 部分主要由多个 Bneck 结构组成, 特征层通过一系列 Bneck 结构逐步深化, 如图 2 所示。具体地, 每个 Bneck 模块的流程始于 1×1 逐点卷积, 可有效减少通道数、显著降低后续计算成本。后续, 进行深度卷积 (Dwise) 运算, 在各个输入通道上独立执行卷积运算, 以进一步削减计算量, 提升运行效率。

同时, 为了增强模型的非线性表达能力, 还引入了激活函数。

$$\text{Relu6}(x) = \min(\max(x, 0), 6) \quad (1)$$

$$\text{h-swish}[x] = x \left(\frac{\text{Relu6}(x + 3)}{6} \right) \quad (2)$$

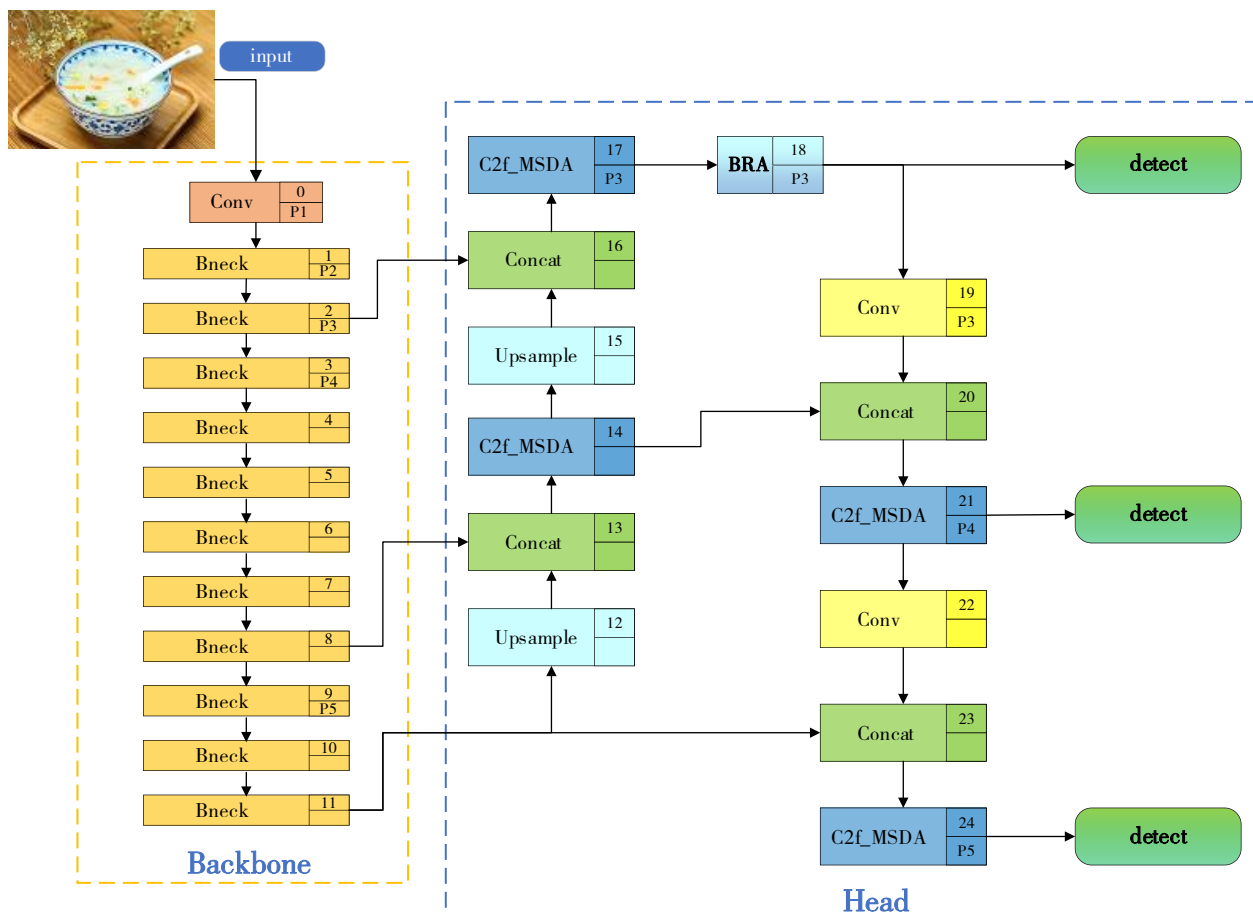


Fig. 1 Overall network structure of MEAL-YOLOv8

图 1 MEAL-YOLOv8 整体网络结构

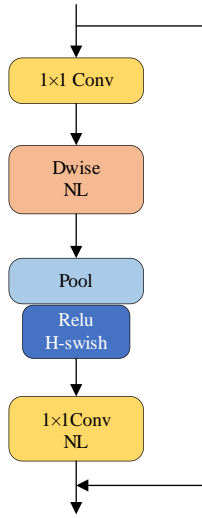


Fig. 2 Bneck structure in Backbone
图2 Backbone中的Bneck结构

式中: x 为输入特征;Relu6函数将输入值 x 范围控制在区间 $[0,6]$,抑制过大激活值;h-swish函数对 $x+3$ 进行Relu6运算,通过分段线性化降低计算成本,适合轻量化网络。

首先,对输入特征图进行全局平均池化;其次,利用 1×1 逐点卷积扩展通道数,不仅恢复了之前的特征维度,还增强了模型对不同特征的表达能。本文在Bneck结构中采用了深度可分离卷积,相较于传统卷积具有更低的计算复杂度。其中,深度可分离卷积是一种将标准卷积分解为深度卷积和 1×1 卷积的形式,深度可分离卷积的计算成本为:

$$D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F \quad (3)$$

标准卷积的计算成本为:

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F \quad (4)$$

式中:输入通道数为 M ;输出通道数为 N ;核大小为 $D_K \cdot D_K$;特征图大小为 $D_F \cdot D_F$ 。

最后,使用 3×3 大小的卷积核进行计算,通过可分离卷积降低计算复杂度,使模型既轻量化又高效,适合资源受限的嵌入式设备。理论上,深度可分离卷积约是标准卷积计算量的 $1/9$ 。

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + D_K^2 = \frac{1}{N} + \frac{1}{9} \quad (5)$$

1.2 Head部分

如图1所示,Head部分包括Upsample、Concat以及改进的C2f-MSDA和BRA模块。其中,C2f-MSDA模块是在YOLOv8原有的C2f模块中加入MSDA注意力机制,同时将BRA注意力机制加入P3流程的C2f-MSDA模块;Upsample模块将低分辨率特征图放大;Concat模块指对特征进行拼接操作。

由此可知,上采样后的特征图会与相应分辨率的特征

图进行拼接,以融合多尺度特征;C2f-MSDA模块在MEAL-YOLOv8中旨在提升模型的效率和性能。结构如图3所示。

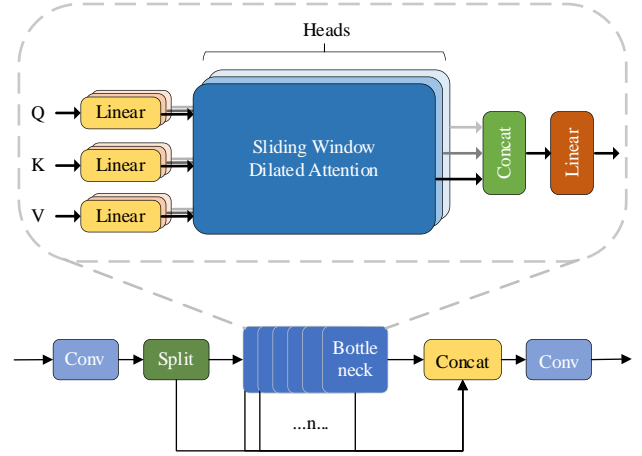


Fig. 3 Structure of C2f-MSDA
图3 C2f-MSDA结构

C2f-MSDA通过第一个卷积层后,Split操作将特征图分为两路处理:一路保持信息直接传递;另一路则历经一个或多个Bottleneck结构深化特征提取过程。在Bottleneck结构中加入MSDA模块,可帮助模型更好地理解输入数据的层次结构和细节。如图3所示,特征图 X 先经过一个卷积层,再将卷积后的图像分割为 Q, K, V 矩阵。

$$Q, K, V \in \mathbb{R}^{[B \times d \times H \times W]} \quad (6)$$

式中: B 为批量大小; H, W 分别为高度、宽度; d 为注意力头的维度。

然后,滑动窗口扩大注意力机制(Sliding Window Dilated Attention, SWDA)使用膨胀卷积从 K, V 中提取局部区域的信息,使得 Q, K, V 矩阵的范围改变为:

$$Q \in \mathbb{R}^{[B \times h \times N \times 1 \times d]} \quad (7)$$

$$K, V \in \mathbb{R}^{[B \times h \times N \times k^2 \times d]} \quad (8)$$

式中: h 为注意力头的数量; N 为空间位置数量(即 $H \times W$); k 为膨胀卷积膨胀率。

接下来,计算注意力权重。

$$A = \text{soft max} \left(\frac{Q \cdot K^T}{\sqrt{d}} \right) \quad (9)$$

式中: $Q \cdot K^T$ 表示点积操作; \sqrt{d} 为一个缩放因子,防止点积结果过大,导致softmax函数饱和。

最后,将注意力权重 A 应用于 V 上,得到重新排列和重组的输出 O ,再传递给一个线性层进行特征聚合。

$$O = A \cdot V \quad (10)$$

$$Y = \text{Linear}(O) \quad (11)$$

首先,上述过程会在不同的膨胀率下重复进行,以处理不同尺度的信息,并将所有不同膨胀率下的输出连接在一起,形成MSDA模块的输出;其次,MSDA通过不同尺度的注意力机制捕捉多尺度上下文信息,融合后的模型能在不同尺度上更好地理解、融合特征,减少不必要的冗余信

息;再次,Concat操作合并两路处理后的特征图,在通道维度整合跨层特征信息;最后,应用卷积层进行必要的特征调整与优化,以更好地匹配后续任务需求。综上,C2f_MS-DA模块构建了高效、高性能的特征处理流程。

为了解决轻量化YOLOv8算法带来的精度损失问题,在Head部分加入BRA模块。BRA是一种注意力机制,用于优化神经网络的计算效率,尤其在资源受限的设备上,通过两个级别的路由操作来实现动态的稀疏注意力分配。

首先,对输入特征图进行区域划分,并计算各区域间的相关性矩阵。

$$A' = Q'(K')^T \quad (12)$$

式中: Q' 、 K' 分别为区域级别的查询和键的平均表示。

其次,对相关矩阵 A' 进行剪枝,仅保留每个区域最相关的前 k 个区域,生成路由索引矩阵 I' 。

$$I' = \text{topIndex}(A') \quad (13)$$

再次,利用索引矩阵 I' ,通过聚合操作获取 K^g 、 V^g 。

$$K^g = \text{gather}(K, I') \quad (14)$$

$$V^g = \text{gather}(V, I') \quad (15)$$

最后,应用注意力机制得到:

$$\text{Attention}(Q, K^g, V^g) = \text{soft max} \left(\frac{Q_r K_r^T}{\sqrt{d_k}} \right) V_g \quad (16)$$

在上述计算过程中,LCE(局部上下文增强)通过深度卷积实现;Attention(注意力机制)采用缩放点积注意力;

BRA的总计算复杂度为 $O(H \cdot W \cdot S)$;传统注意力机制复杂度为 $O((H \cdot W)^2)$; H 、 W 分别为特征图的高和宽; S 为区域划分因子。

2 实验数据集与平台

2.1 实验数据

为了确保所提食物识别算法能满足进食者全面的营养摄入及饮食多样化需求,实验中精心挑选了10种典型的亚洲特色食物,涵盖碳水化合物、蛋白质及丰富的维生素来源,具体包括米饭、面条、饺子、寿司、麻婆豆腐、宫保鸡丁、粥以及香蕉、草莓、橙子等3种水果,旨在构建一个具有代表性的食物数据集,满足模型对不同类别食物的识别要求。

同时,为了模拟现实生活中的多样性与复杂性,数据集包括不同照明条件、摆放角度、食物状态及背景复杂度下的图像,共计5000幅食物图像样本,并按8:1:1的比例划分4000张训练集、500张验证集、500张测试集,旨在模拟现实生活场景中的多样性与复杂性,验证算法在复杂环境中的适应性与鲁棒性。

2.2 实验平台

本文在自主研发的KANG-Feed助餐机器人上部署和测试所提算法,机器人系统的整体工作流程如图4所示。

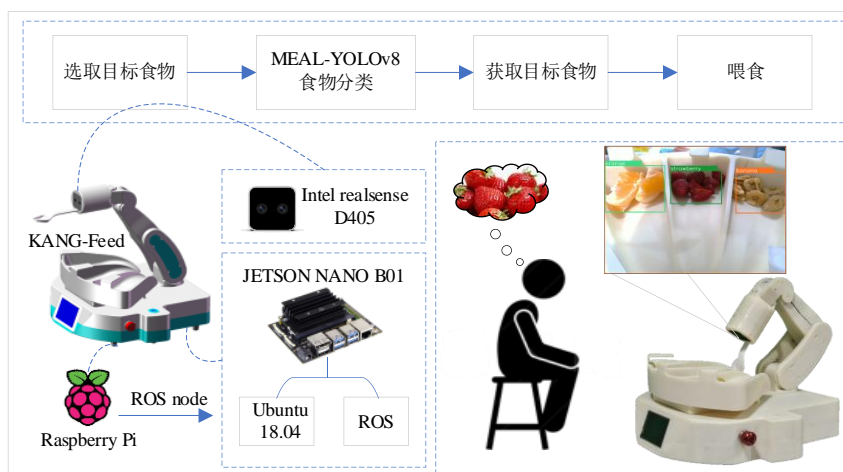


Fig. 4 Overall feeding process

图4 喂食整体流程

首先,用户选择期望获取的食物种类;其次,MEAL-YOLOv8模型对食物进行分类识别,获取目标食物并通过机械臂传递至用户。该实验平台旨在模拟实际应用场景,以满足进食者对食物识别和传递的实时需求。

实验平台的核心硬件为JETSON NANO B01开发板,由NVIDIA公司生产,采用NVIDIA Tegra X1处理器,配备四核ARM Cortex-A57 CPU和4GB LPDDR4内存,是一款高性能的嵌入式计算平台,适用于机器视觉和深度学习应用。JETSON NANO B01开发板的运行环境为Ubuntu 18.04,具备良好的软件支持,能为MEAL-YOLOv8模型的

部署提供稳定的计算平台。在该实验平台上,深度相机(Intel RealSense D405)被固定在KANG-Feed助餐机器人机械臂的末端,用于实时采集餐碗中的食物图像。RealSense D405是一款能提供准确三维数据的高精度的深度相机,可增强食物识别系统在复杂环境中的感知能力,MEAL-YOLOv8算法通过该数据能对食物进行精准分类。

此外,系统部署了机器人操作系统(Robot Operating System, ROS)以协调机器人各模块的工作,并处理任务调度,使JETSON NANO B01开发板与机器人内置的Raspberry Pi之间能进行高效的通信。为了实时传输并处理食物

识别结果,采用了以下三级优化:①通过TensorRT框架整合INT8量化实现硬件加速(3~5倍提速);②先部署GPU进行预处理,再通过CUDA重写后处理内核以消除传输瓶颈;③在系统层面动态调节分辨率、预分配内存池及锁定GPU频率。通过该配置,系统不仅具备了较强的灵活性和扩展性,还能有效提升响应速度与实时性,从而满足KANG-Feed送餐机器人在实际应用中的高效性要求。最终,识别结果通过ROS系统反馈至机器人,完成食物识别和喂食的全过程。

3 评估方法

3.1 比较模型

为了在JETSON NANO B01平台上选取最适合部署的轻量化算法,以实现高效、资源友好的性能,设计了一系列比较实验。表1为构建的比较网络配置,包括将YOLOv8的原始主干网络替换为MobileNetV3、EfficientViT、FasterNet等知名的轻量化网络。本文在资源受限环境下将MEAL-YOLOv8与上述模型进行比较,以探索模型的优化潜力。同时,为进一步验证模型的识别精度,将MobileNetV3与经典的单发多框探测器算法(Single Shot Multi-Box, SSD)结合,与YOLOv8-MobileNetV3进行比较。其中,model 7为YOLOv8-MobileNetV3模型仅加入MSD注意力机制,以期在不显著增加计算负担的前提下,最大化模型的识别能力。

Table 1 Comparison of algorithm network architectures

表1 算法网络结构比较

模型	方法	网络主干	注意力机制
model 1	YOLOv8n	DarkNet-53	\
model 2	YOLOv8x	DarkNet-53	\
model 3	YOLOv8-EfficientViT	EfficientViT	\
model 4	YOLOv8-FasterNet	FasterNet	\
model 5	YOLOv8-MobileNetV3	MobileNetV3	\
model 6	MobileNetV3-SSD	MobileNetV3	\
model 7	YOLOv8-MobileNetV3-BRA	MobileNetV3	BRA
MEAL-YOLOv8	MEAL-YOLOv8	MobileNetV3	BRA&MSDA

3.2 评估指标

为了全面评估所提算法在嵌入式平台上的性能,采用了一系列评估指标,涵盖模型复杂度、准确度及在嵌入式设备上的运行效率等方面。

在模型复杂度评估中,主要考量了参数量(Parameters)、模型大小(Model size)和计算性能(GFLOPS)。其中,较低的参数量有助于减小模型体积,使其更适合部署在资源受限的设备上;较小的模型大小能减少存储需求与加载时间,提升系统响应速度;较低的GFLOPS(每秒十亿次浮点运算)值表明模型在推理过程中所需的计算资源较少,适用于计算受限的嵌入式设备。

在准确度评估方面,采用平均精度(Average Precision, AP)和平均精度均值(mean Average Precision, mAP)

进行评价。其中,AP用于衡量模型在不同类别食物识别中的精度;mAP则通过计算在交并比(Intersection over Union, IoU)为0.5时的精度—召回曲线下的面积,以反映模型在多类别物体检测任务中的整体性能。

为了评估模型在嵌入式开发板上的部署表现,使用内存占用(Memory Usage)、推理时间(Inference Time)进行评估。其中,较低的内存占用有助于减少内存压力,提升系统稳定性和响应速度;较短的推理时间可确保模型能快速响应,满足实时性要求。

4 比较结果

4.1 模型复杂度

表2为各算法的复杂度和准确度。由此可知,MEAL-YOLOv8的复杂度最低,彰显了其轻量化设计的特点,初步验证了对YOLOv8改进的有效性,训练后MEAL-YOLOv8的参数量仅为1.09 G(G为10亿参数),参数量最低。在内存占用方面,MEAL-YOLOv8仅占用2.37 MB的内存,相较于Model 1-Model 7的模型最小,极大增强了模型的可部署性和适用范围。在计算复杂度方面,MEAL-YOLOv8的计算性能(GFLOPS)为2.8,与未加入注意力机制的model5相同,但需要的计算资源最少,进一步证明了MEAL-YOLOv8的运算效率较高。

Table 2 Complexity and accuracy of each algorithm

表2 各算法的复杂度与准确度

模型	参数量/G	模型大小/MB	计算性能	mAP
model 1	3.01	5.97	8.10	0.913
model 2	68.13	130.00	257.40	0.931
model 3	4.01	8.36	9.40	0.901
model 4	1.75	3.54	5.10	0.891
model 5	1.19	2.54	2.80	0.906
model 6	4.87	18.80	7.59	0.801
model 7	1.19	2.54	3.00	0.913
MEAL-YOLOv8	1.09	2.37	2.80	0.914

4.2 准确性分析

图5为MEAL-YOLOv8在10个食品类别中的准确性及mAP。图6为MEAL-YOLOv8与7个先进的深度学习模型进行比较的结果。由图5、图6可见,MEAL-YOLOv8在香蕉、粥、宫保鸡丁、麻婆豆腐和寿司等类别上的准确度均超过0.95,识别精度较高;MEAL-YOLOv8相较于其他模型,在所有类别中的精度处于中上水平,证明了其在食品识别任务中的稳定性与可靠性。

研究发现,MEAL-YOLOv8在10个食品类别上的mAP值为0.914(见表2),略低于model 2的0.931,但在实时性、嵌入式平台适应性方面优势明显。特别是,在计算资源受限的环境中,MEAL-YOLOv8能有效平衡精度与处理速度,满足实时处理的需求,而model 2的骨干网络较复杂,导致其内存占用、计算资源需求较高,在嵌入式开发板上运行速度较慢,难以满足实时应用的要

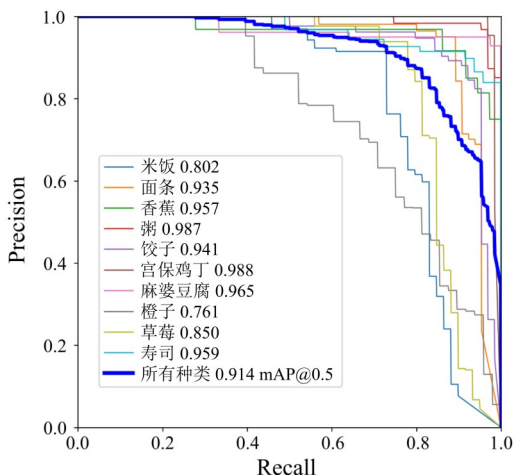


Fig. 5 P-R curve

图 5 P_R 曲线

求, 限制了其在实际部署中广泛应用。

此外, 将 MEAL-YOLOv8 部署到嵌入式开发板得到的平均实时检测精度为 0.822, 如图 7 所示。由此可见, 尽管

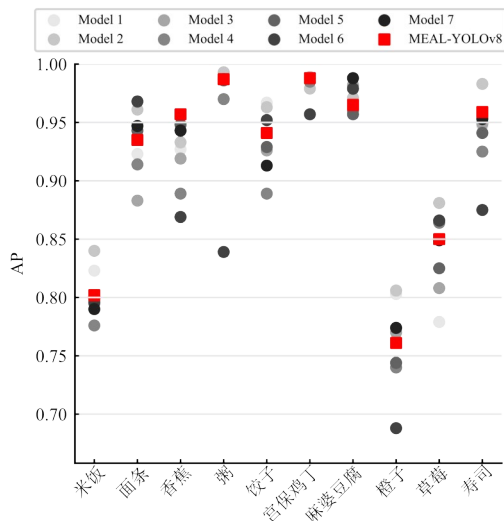


Fig. 6 Scatter of AP values for various food types in each model

图 6 各模型食物种类 AP 值散点

该模型在精度上略低于 model 2, 但在实时检测中能准确识别食品类型, 可满足实时处理的要求。

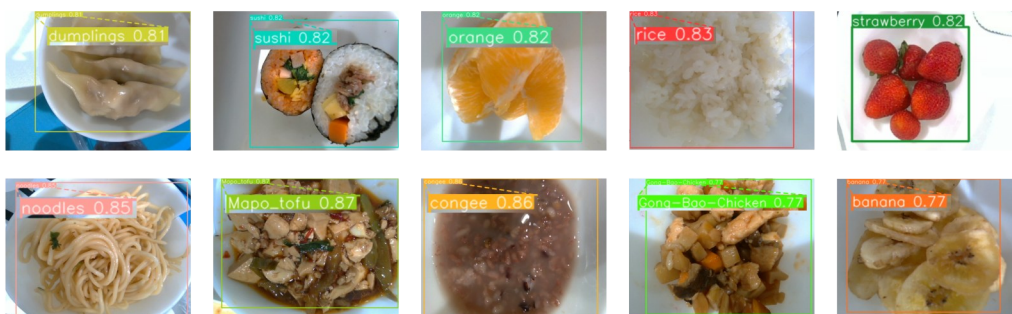


Fig. 7 Real-time food category detection

图 7 实时食物种类检测

更重要的是, MEAL-YOLOv8 在精度与性能之间实现了有效的平衡, 特别适用于计算资源受限且对实时性要求较高的应用场景, 例如嵌入式系统、边缘计算平台。

4.3 模型部署分析

表 3 为将 MEAL-YOLOv8 模型部署到 JETSON NANO B01 后的实验结果, 平台搭载了 NVIDIA Tegra X1 处理器, 具有较低的计算能力和内存容量, 进一步证明了其高效的资源利用能力。由此可知, MEAL-YOLOv8 模型大小仅为 2.37 MB, 在比较模型中最低, 尤其适配于存储资源有限的环境。在实时检测食物种类时, MEAL-YOLOv8 的内存占为 158 MB, 在比较模型中最低, 较低的内存需求确保了模型在内存配置较紧张的系统, 也能维持稳定的性能表现, 证明了其在资源管理方面的高效性。MEAL-YOLOv8 模型的末端视觉停留平均时间 ≥ 1.5 s, 563.79 ms 的推理耗时约占停留时间的 37%, 不会形成任务阻塞、影响喂食节奏。尽管, MEAL-YOLOv8 的推断速度并非最快, 但平均推断时间可兼顾较小模型大小与低内存占用, 在平衡性能与资源消耗方面的优势明显。

综上, MEAL-YOLOv8 的卓越之处在于其能在保证高效推断速度的同时, 实现最小化的模型体积与内存占用, 这种特质使得该模型适合在资源受限场景下应用。

Table 3 Comparison results of different models

表 3 不同模型比较结果

模型	模型大小/MB	内存占用/MB	推理时间/ms
model 1	5.97	329	611.29
model 2	130.00	816	10 457.31
model 3	8.36	327	1245.67
model 4	3.54	337	541.73
model 5	2.54	345	597.83
model 6	18.8	300	456.37
model 7	2.54	368	836.51
MEAL-YOLOv8	2.37	158	563.79

4.4 消融实验

表 4 为模型在不同主干网络下的实验结果。由此可知, 将 YOLOv8 的主干网络替换为 MobileNetV3 的主干网络后, 在参数量减少 60% 的情况下, 仅牺牲了 0.77% 的 mAP, 实现了参数量与精度之间的平衡。表 5 为在替换主干网

络的前提下为模型加入不同注意力机制的结果。由此可知,在仅加入BRA机制的情况下,mAP提升0.7%,但参数量未发生变化;在加入BRA+MSDA注意力机制后,mAP提升0.1%、参数量降低8.4%,表明MSDA可进一步优化模型体积,提升模型精度。

Table 4 Backbone network comparison

表4 主干网络比较

模型	主干网络	参数量/G	mAP
model 1	DarkNet-53	3.01	0.913
model 5	MobileNetV3	1.19	0.906

Table 5 Comparison of attention mechanisms

表5 注意力机制比较

模型	注意力机制	参数量/G	mAP
model 5	/	1.19	0.906
model 7	BRA	1.19	0.913
model 7	BRA+MSDA	1.09	0.914

5 结语

本文针对现有送餐机器人在食物种类识别方面的不足,以及嵌入式平台计算能力的局限性,提出一种MEAL-YOLOv8模型,旨在提升送餐场景中的食物识别精度,实现轻量化设计,以适应资源受限的嵌入式平台。实验表明,MEAL-YOLOv8在食物识别任务中取得了91.4%的准确率,在高精度食物分类方面的性能优秀。

轻量化设计后,MEAL-YOLOv8模型的大小被压缩至2.37 MB,可确保其能在计算资源有限的JETSON NANO B01嵌入式平台上高效运行。运行时,模型的内存消耗控制在158 MB以内,进一步凸显了其在嵌入式设备上的优势。综上,MEAL-YOLOv8在实际部署中食物识别精度高,有效降低了对计算资源的依赖,为送餐机器人系统的普及和应用提供了技术支持。

参考文献:

- [1] United Nations. World population prospects 2022: summary of results [EB/OL]. https://www.un.org/development/desa/pd/sites/www.un.org/development/desa/pd/files/wpp2022_summary_of_results.pdf.
- [2] RUDNICKA E, NAPIERAŁA P, PODFIGURNA A, et al. The world health organization (WHO) approach to healthy ageing [J]. *Maturitas*, 2020, 139: 6-11.
- [3] GERLACH H K, JANSEN T A, LEONHARDT S. A survey on robotic devices for upper limb rehabilitation [J]. *Journal of Neuroengineering and Rehabilitation*, 2014, 11: 1-29.
- [4] TOPPING M. The development of Handy 1, a robotic aid to independence for the severely disabled [C]// *IEEE Colloquium Mechatronic Aids for the Disabled*, 1995: 1-2.
- [5] SOYAMA R, ISHII S, FUKASE A. 8 selectable operating interfaces of the meal-assistance device "my spoon" [C]// *Advances in Rehabilitation Robotics, Lecture Notes in Control and Information Science*, 2006: 155-163.
- [6] HERMANN R P, PHALANGAS A C, MAHONEY R M, et al. Powered feeding devices: an evaluation of three models [J]. *Archives of Physical Medicine and Rehabilitation*, 1999, 80: 1237-1242.
- [7] CHOI I, KO K, SONG H, et al. A meal-assistance robot system for Asian food and its food acquisition point estimation and user interface based on face recognition [J]. *Applied Science*, 2023, 13(5): 3216.
- [8] KAWAMOTO H, SHIRAKI T, OTSUKA T, et al. Meal-assistance by robot suit HAL using detection of food position with camera [C]// *IEEE International Conference on Robotics and Biomimetics*, 2011: 889-894.
- [9] FAN Y, ZHANG L, ZHENG C, et al. Real-time and accurate meal detection for meal-assisting robots [J]. *Journal of Food Engineering*, 2024, 371: 111996.
- [10] ZHAO X, CHEN D, XIANG G, et al. A vision-based target localization system for the meal assistance robot [C]// *International Conference on CYBER Technology in Automation, Control, and Intelligent Systems*, 2022: 769-774.
- [11] PERERA C J, LALITHARATNE T D, KIGUCHI K. EEG-controlled meal assistance robot with camera-based automatic mouth position tracking and mouth open detection [C]// *IEEE International Conference on Robotics and Automation*, 2017: 1760-1765.
- [12] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [DB/OL]. <https://arxiv.org/abs/1409.1556>.
- [13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// *IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.
- [14] GIRSHICK R. Fast R-CNN [C]// *IEEE International Conference on Computer Vision*, 2015: 1440-1448.
- [15] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]// *Las Vegas: IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [16] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]// *Honolulu: IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [17] REDMON J, FARHADI A. Yolov3: an incremental improvement [DB/OL]. <https://arxiv.org/abs/1804.02767>.
- [18] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection [DB/OL]. <https://arxiv.org/abs/2004.10934>.
- [19] GE Z, LIU S, WANG F, et al. Yolox: exceeding YOLO series in 2021 [DB/OL]. <https://arxiv.org/abs/2107.08430>.
- [20] JIAO J, TANG Y M, LIN K Y, et al. DilateFormer: multi-scale dilated transformer for visual recognition [J]. *IEEE Transactions on Multimedia*, 2023, 25: 8906-8919.
- [21] ZHU L, WANG X, KE Z, et al. Biformer: vision transformer with bi-level routing attention [C]// *Vancouver: IEEE Conference on Computer Vision and Pattern Recognition*, 2023.

(责任编辑:刘嘉文)