



西南交通大学
Southwest Jiaotong University

基于YOLO的智能自动结算购物车

指导老师：龙治国

小组成员：陶炳赫，徐耕洋，
王烨澄，张凌屹，刘浩轩



精勤求學
敦篤勵志
果毅力行
忠恕任事

CONTENT 目录

01 | 项目简介

02 | 项目创新点

03 | 项目执行过程

04 | 成果演示

05 | 总结与展望

目录

01

项目简介

- 项目背景
- 研究目标
- 项目意义



- 目前零售商品多以人工结账和自助扫码结账为主，顾客浪费大量时间在排队上，商家浪费金钱在排队上。（尤其是新兴的零食店）



图1. 零食有鸣的排队现象

为了解决这一现象，我们期望设计出一种可以一边购物一边结账的购物车。

- 要有零售商品识别的功能。
- 也要有散装称重的功能，尤其是散装零食的称重。
- 可以在购物的过程中结账
- 要拿出商品时需要检查是否真的拿出商品

- 这可以帮助顾客节省大量时间，增加购物体验，还可以帮助商家节省更多运营成本，提高收益。
- 为未来的智能化购物提供参考。

目录

02

项目创新点

- 软件的创新
- 硬件的创新



利用多尺度空洞注意力(MSDA)，改进YOLOv8的网络结构。

- MSDA 能够模拟小范围内的局部和稀疏的图像块交互，这些发现源自于对 ViTs 在浅层次上全局注意力中图像块交互的分析。作者发现在浅层次上，注意力矩阵具有局部性和稀疏性两个关键属性，这表明在浅层次的语义建模中，远离查询块的块大部分无关，因此全局注意力模块中存在大量的冗余。

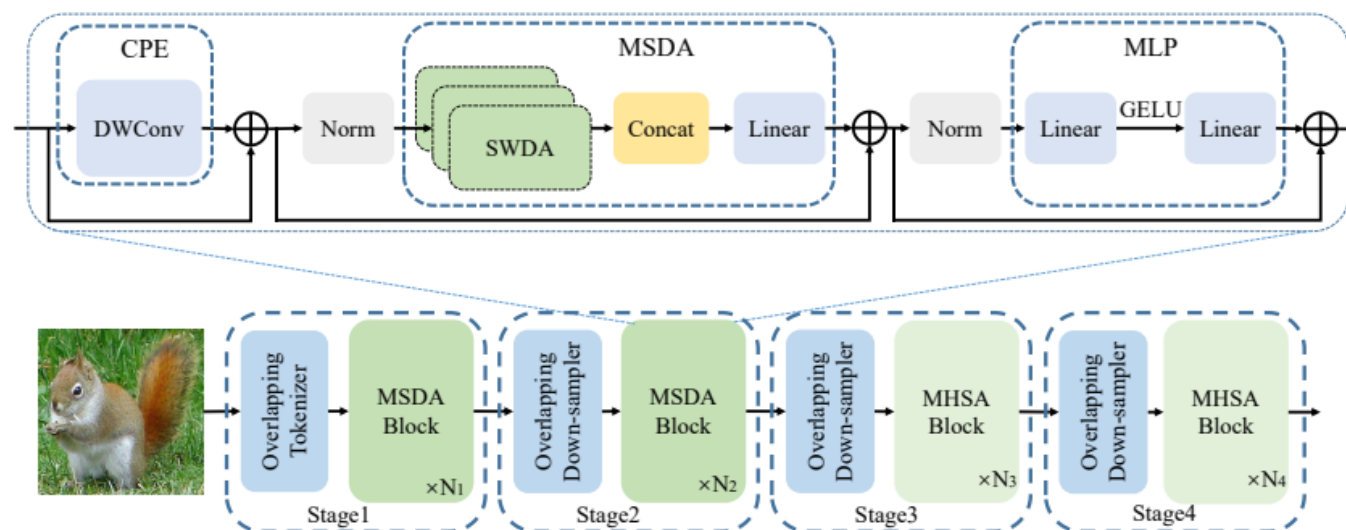


Fig. 3: The overall architecture of our DilateFormer. The top part shows the proposed Multi-Scale Dilated Attention (MSDA) block, consisting of DwConv, Multi-Scale Sliding Window Dilated Attention operation (SWDA) and MLP. The bottom part shows DilateFormer, consisting of Overlapping Tokenizer, Overlapping Downsampler, Multi-Scale Dilated Attention (MSDA) block and Multi-Head Self-Attention (MHSA) block.

CSDN @阿_旭

图2. MSDA网络结构

利用多尺度空洞注意力(MSDA)，改进YOLOv8的网络结构。

- 通过在不同头部 (heads) 中使用不同空洞率 (dilation rates) 的滑动窗口注意力，捕捉多尺度特征信息，同时减少全局注意力计算的冗余。将特征图的通道分成多个头部，每个头部独立计算注意力，并通过特征拼接和线性变换整合输出。

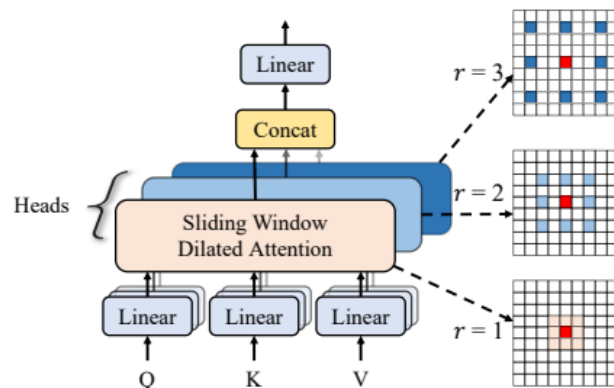


Fig. 4: **Illustration of Multi-Scale Dilated Attention (MSDA)**. First, the channels of the feature map are split into different heads. Then, the self-attention operation is performed among the colored patches in the window surrounding the red query patch, using different dilation rates in different heads. Besides, features in different heads are concatenated together and then fed into a linear layer. By default, we use a 3×3 kernel size with dilation rates $r = 1, 2$ and 3 , and the sizes of attended receptive fields in different heads are 3×3 , 5×5 and 7×7 .
CSDN @阿_旭

图3. MSDA工作原理

利用多尺度空洞注意力(MSDA), 改进YOLOv8的网络结构。

➤ MSDA在目标检测和图像分类等任务中表现出色, 能够提升模型对多尺度目标的检测能力, 同时优化计算效率。

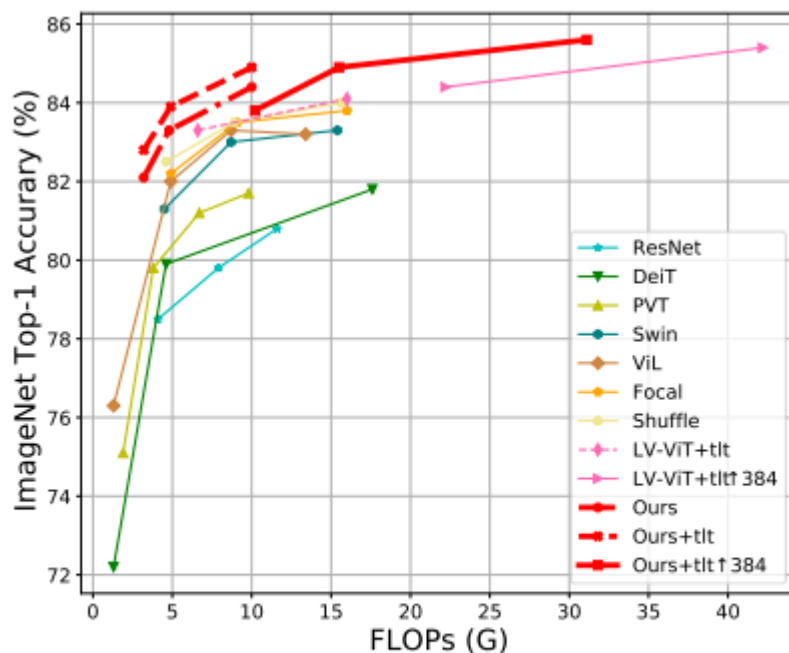


Fig. 1: Performance comparisons with respect to FLOPs on ImageNet-1K classification. Without extra training data, our DilateFormer variants achieve comparable or even better performance with fewer FLOPs.

CSDN @阿_旭

图4. 与其他改进的性能对比

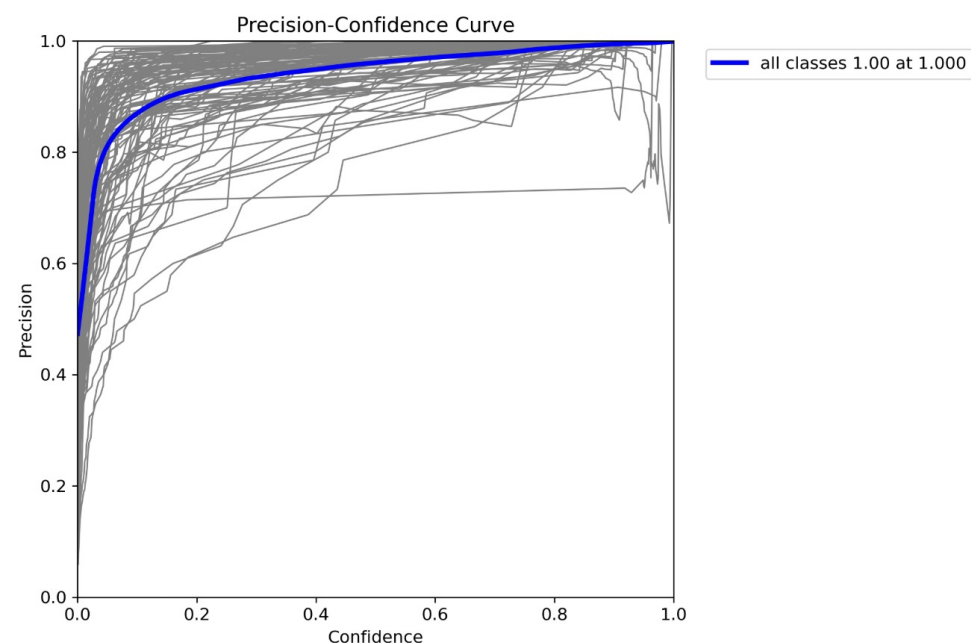


图4.1. 本项目的训练结果

视觉识别结合称重模块

- 采用hx711称重模块，精度可以精准到1g，实现对无标签产品的定量结算，解决了传统条形码结算平台需手动输入编码。



图4.3 称重模块

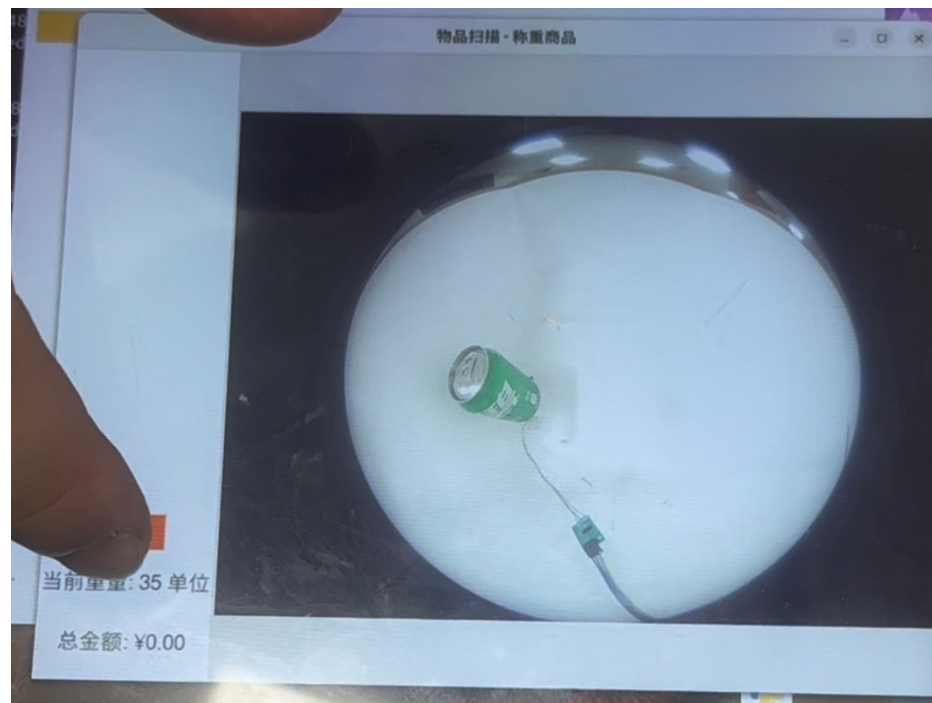


图4.4 称重结果

激光指示模块

- 通过二自由度云台和舵机驱动板实现对摄像头前区域的全覆盖指示。
- 如果称重结算时出现多种商品，程序会将激光打在需要移除的物品上，避免称重结算结果混乱，同时也可以给消费者明确的指示。

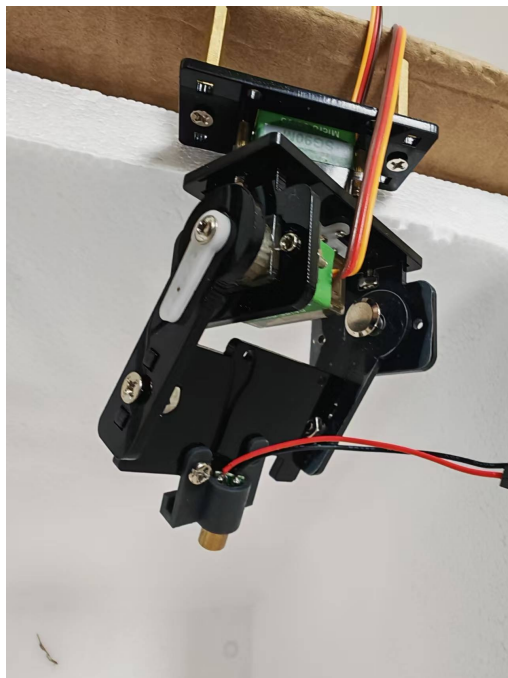


图4.5 激光指示模块



图4.6 激光指示到多余的商品

目录

03

项目执行过程

- 小组分工
- 小组合作



小组成员根据专业进行分工，负责擅长的部分。

陶炳赫（计算机）：负责yolo模型训练，添加注意力机制，优化代码等。

徐耕洋（电子信息）：负责指示灯，称重，摄像头等硬件模块的搭建与调试。

王烨澄（机械）：负责硬件的组装和设备容器的搭建。

张凌屹（计算机）：负责GUI的开发和指示灯算法。

刘浩轩（电子信息）：负责硬件组装的设计和设备容器的设计。

小组成员除了定期与老师汇报进度，同时不定期私下会面推动项目进度。



图5. 推进硬件组装的会面

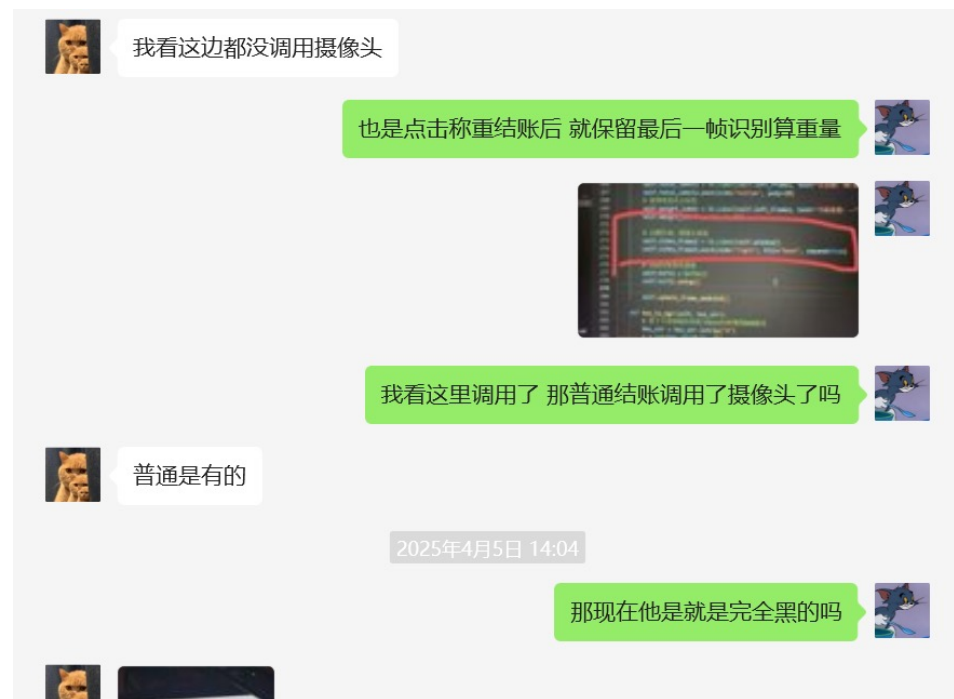


图6. 线上讨论软件问题

目录 04

成果演示



目录 05

总结与展望



困难与挑战

- 由于小组第一次独立来实现一个硬件系统，面临着报错无法处理，软件适配难以解决等问题；在软件上由于第一次完整的做一个项目，在训练数据集等过程遇到了不同的困难。

未来展望

- 当前我们的购物车系统只是个原型，未来可以投入更多时间完善该系统，包括收集更多数据来训练等，使其真正做到可以商用，达到节省顾客时间的需求。同时经历诸多困难，我们在开展其他项目时也会更加得心应手。



西南交通大学
Southwest Jiaotong University

谢谢聆听!

汇报人：陶炳赫

2025年4月10日

