覃宜霜, 陶雯, 贾建双,等. 基于改进 YOLOv5 的卷烟品规展示视频检测研究[J]. 智能计算机与应用,2025,15(6):127-133.
DOI:10.20169/j.issn.2095-2163.25022201

基于改进 YOLOv5 的卷烟品规展示视频检测研究

覃宜霜, 陶 雯, 贾建双, 陈 杰, 覃琼慧
(广西中烟工业有限责任公司, 南宁 530000)

摘 要: 在卷烟营销管理中,精准统计烟草零售终端柜台中卷烟品规数量对于库存控制和产品陈列优化具有至关重要的作用。为了实现高效的卷烟产品识别与统计,本文提出了一种基于 YOLOv5 的轻量化卷烟品规展示视频检测方法。首先,为了降低模型复杂度,在 YOLOv5 中引入轻量化的 MobileNetV3 主干网络,以减少计算量和参数规模;其次,采用视频帧匹配技术融合帧级检测结果,以确保跨帧目标的准确匹配和统计。实验结果表明,相较于原始 YOLOv5-1 模型,在检测准确率几乎不变的情况下,改进后的模型参数量减少了 87.81%,计算量 (*FLOPs*) 减少了 89.82%。整体结果表明,本文方法在确保高精度检测的同时,显著降低了计算开销,为卷烟品规的自动化统计与智能管理提供了有效的技术支撑。此外,卷烟产品陈列位置的可视化记录为工业企业的管理决策提供了有力支持,有助于优化库存管理与品规陈列布局。

关键词: 卷烟检测; YOLOv5; MobileNetV3; SIFT; 视频计数
中图分类号: TP183 **文献标志码**: A **文章编号**: 2095-2163(2025)06-0127-07

Improve YOLOv5 object detection algorithm for shelf tobacco videos

QIN Yishuang, TAO Wen, JIA Jianshuang, CHEN Jie, QIN Qionghui
(China Tobacco Guangxi Industrial Co., Ltd., Nanning 530000, China)

Abstract: In the management of cigarette marketing, accurate statistics of cigarette quantity in the counter of tobacco retail terminal plays a vital role in inventory control and product display optimization. In order to achieve efficient cigarette product identification and statistics, this paper proposes a video detection method of lightweight cigarette specification display based on YOLOv5. Firstly, in order to reduce the complexity of the model, the lightweight MobileNetV3 backbone network is introduced in YOLOv5 to reduce the amount of calculation and parameter size; Secondly, the video frame matching technology is used to fuse the frame level detection results to ensure the accurate matching and statistics of cross frame targets. The experimental results show that compared with the original YOLOv5-1 model, the parameters of the improved model are reduced by 87.81% and the amount of calculation (*FLOPs*) is reduced by 89.82% with almost the same detection accuracy. The overall results show that this method not only ensures high-precision detection, but also significantly reduces the computational cost, and provides effective technical support for the automatic statistics and intelligent management of cigarette quality regulation. In addition, the visual record of cigarette product display location provides strong support for the management decision-making of industrial enterprises, and helps to optimize inventory management and product specification display layout.

Key words: cigarette detection; YOLOv5; MobileNetV3; SIFT; video counting

0 引 言

卷烟零售业是全球消费品市场的重要组成部分,其发展也与时代变化息息相关。在卷烟的生产、包装、运输和销售等各环节中,卷烟产品的管理与检测具有重要作用[1]。然而,传统的卷烟管理与检测方式主要依赖于人工检查和基于规则的图像处理技术。人工检查方法虽然灵活性较强,但效率低下且容易受到疲劳和主观因素的干扰;基于规则的图像处理技术通过固定算法检测卷烟外观,适应性明显不足。这些传统方式难以应对现代卷烟零售业对高效性、精准性和智能化的更高要求。

在这种背景下,视频检测技术应运而生,成为解决上述问题的有效途径。与静态图像检测相比,视频检测不仅可以利用时间序列信息实现对卷烟产品的持续跟踪和识别,还能通过多帧数据融合提高检

作者简介: 覃宜霜(1980—),女,硕士,中级经济师,主要研究方向:卷烟市场营销。Email:zhiweizhu@mail.nwpu.edu.cn。
收稿日期: 2025-02-22

测的稳定性和准确性[2]。深度学习的快速发展为卷烟检测提供了新的解决方案。文献[3]提出了一种基于计算机视觉和机器学习的卷烟包装真伪鉴别模型,通过图像处理和机器学习模型,实现了高准确率的真伪卷烟包装鉴别。文献[4]设计了一套由工业相机、旋转编码器、光源和图像处理系统组成的工业化系统,实现对卷烟烟盒商标纸表面缺陷检测。然而上述模型并不具有计数和盘点的功能,且模型复杂。为了解决这些问题,文献[5]利用 Mask R-CNN 模型实现卷烟烟盒可视区域检测。文献[6]提出了一种改进的 YOLOV5 模型应用于复杂场景下的卷烟烟盒识别任务。文献[7]在 YOLOv3 模型的基础上加入多空间金字塔池化结构。文献[8]结合深度残差网络(ResNet)和自适应选择卷积网络(SKNet),以提高烟盒识别的准确性和效率。尽管现有深度学习模型在静态图像的烟盒检测、分类和规格识别方面取得了可观成效,但在动态视频场景中的目标检测与追踪任务中仍存在明显不足。与静态图像检测相比,视频扫描货架具有显著优势。视频数据在多时间点上提供多样化视角和背景信息,全面覆盖货架上卷烟不同位置,节省资源的同时有效降低了漏检率,克服了依赖单一静态图像所带来的局限性。然而,现有深度学习方法在动态视频场景下仍难以兼顾效率与精度。此外,现有模型普遍依赖大量内存和计算资源,进一步限制了其在移动设备上的部署和应用。

为解决上述问题,本文对 YOLOv5 主干网络进行了优化,采用了轻量化的 MobileNetV3[9] 结构。MobileNetV3 通过深度可分离卷积(Depthwise Separable Convolution)可显著降低模型的计算复杂度,同时保留了较强的特征提取能力。这种优化降低了模型的资源需求,提高对卷烟的适应性,使其更适合在嵌入式设备和动态场景中运行。此外,结合图像处理技术,本文进一步实现了动态视频场景下卷烟产品的高效检测与计数,有效提升了生产管理效率。

综上所述,本文基于 YOLOv5 框架和 MobileNetV3 网络,对卷烟视频检测与计数进行研究与改进,主要贡献如下:

(1)构建常见卷烟数据集。采集市场上常见卷烟产品的数据,进行高质量标注,构建了一套适用于深度学习的卷烟检测数据集。

(2)模型优化。基于 YOLOv5 模型进行改进,采用了轻量化的 MobileNetV3 结构对 YOLOv5 的主干网络进行优化,提升模型在复杂背景环境下的性能,降低了模型推理计算量,同时使其更加适合卷烟检测任务。

(3)视频计数。提出基于视频帧间匹配的计数方法,解决了卷烟视频检测中目标计数的问题,实现了高效、精准的视频目标计数功能。

(4)可视化界面。提供了可视化界面以展示模型的检测结果。

本研究为卷烟行业提供了一套智能化的视频检测与计数解决方案,同时为深度学习技术在工业领域的应用探索提供了重要借鉴。

1 相关技术与理论

1.1 目标检测算法

目标检测是计算机视觉领域的核心任务之一,旨在识别图像或视频中的目标对象。当前目标检测方法主要分为两大类:基于候选区域的检测方法(Two-Stage)和单阶段检测方法(One-Stage)。典型的 Two-Stage 目标检测方法包括 R-CNN[10] 系列,如 Fast R-CNN[11]、Faster R-CNN[12] 等。Faster R-CNN 通过引入区域建议网络,有效提升了目标检测效率。然而,该类方法通常计算量较大,难以满足实时性要求。相比之下,单阶段检测方法摒弃了候选区域生成步骤,直接回归目标类别与位置。代表性方法包括 YOLO(You Only Look Once)[13] 系列、SSD(Single Shot multiBox Detector)[14] 等。SSD 采用多尺度特征融合方式,YOLO 框架则采用全局回归策略,将目标检测任务转化为回归问题,具有更高的检测速度。

YOLO 算法核心流程包含 3 个关键阶段。首先进行网格化空间划分。随后,每个网格同步预测目标的空间参数与语义信息。最后,采用非极大值抑制技术,从中筛选出置信度最高的边界框,并移除其他冗余框,以确保最终检测结果的准确性和简洁性。

近年来,YOLO 系列不断发展[15]。然而,YOLOv5 仍具有明显优势。首先,YOLOv5 相较于 YOLOv8 计算复杂度更低,虽然 YOLOv8 检测精度提升,但其网络结构更深,推理速度较慢,难以在资源受限的设备上高效运行。此外,YOLOv5 的检测框架更为稳定,经过长期优化和工程实践,YOLOv5 在 ONNX 等推理加速框架中兼容性较好[16]。因此,本文选择 YOLOv5 作为基础检测框架,并结合轻量化主干网络 MobileNetV3,以进一步提升在移动终端设备下的检测性能和实时性。

1.2　图像匹配算法

由于卷烟视频中存在烟盒目标密集、排列紧凑的情况，传统的视频跟踪方法已难以有效应对，因此采用了基于尺度不变特征变换（Scale－Invariant Feature Transform，SIFT）算法[17]的图像特征匹配方法。SIFT 是一种经典的局部特征提取算法，能够从图像中提取出显著性关键点，并在相邻帧之间进行目标匹配。在特征匹配过程中，使用快速最近邻搜索库（Fast Library for Approximate Nearest Neighbors，FLANN）[18]来提高匹配效率。FLANN 是一种高效的算法，能够在较大的数据集中快速找到最接近的特征点。为了提高匹配的准确性，采用了最近邻比率法（Nearest Neighbor Ratio Test）对匹配质量进行筛选。在筛选出高质量的匹配点对后，进一步计算单应性矩阵（Homography Matrix）[19]来实现目标框的坐标变换。单应性矩阵是描述图像几何关系的矩阵，通过匹配点对推导出两帧图像之间的透视变换。然而，由于图像中的某些匹配点可能存在误差，导致计算不准确，为了剔除这些异常匹配点并提高结果的准确性，引入了随机采样一致性（Random Sample Consensus，RANSAC）[20]算法。RANSAC 通过反复随机选择匹配点对，计算单应性矩阵，从而剔除错误匹配点，确保最终的目标框坐标变换具有较高的鲁棒性和精度。

通过上述方法，能够在视频帧之间精确地匹配目标框，从而有效避免重复计数或漏计。这种方法克服了传统视频跟踪方法在密集目标、遮挡等情况下的局限性，从而实现高效准确的目标追踪与计数。

2　所提算法

整体流程如图 1 所示。首先，系统接收视频输入并提取关键帧，关键帧的数量与视频长度成正比。将提取的关键帧输入识别网络，进行卷烟检测。随后对检测到的目标框进行坐标转换与数据融合，确保数据准确对齐。最后，品规位置和数量信息以可视化形式进行呈现，便于直观展示分析。

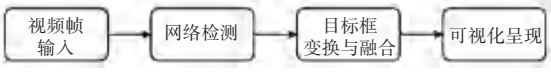

图 1　算法工作流程

Fig. 1　Algorithm workflow

2.1　MobileNetV3 主干网络检测模块

传统 YOLOv5 框架中采用的 CSPDarkNet 主干网络通过大量卷积模块提取特征，在计算资源上的消耗较大。本文在 YOLOv5 的框架下引入 MobileNetV3 作为轻量化骨干网络，以提高特征提取的效率。网络整体架构如图 2 所示。

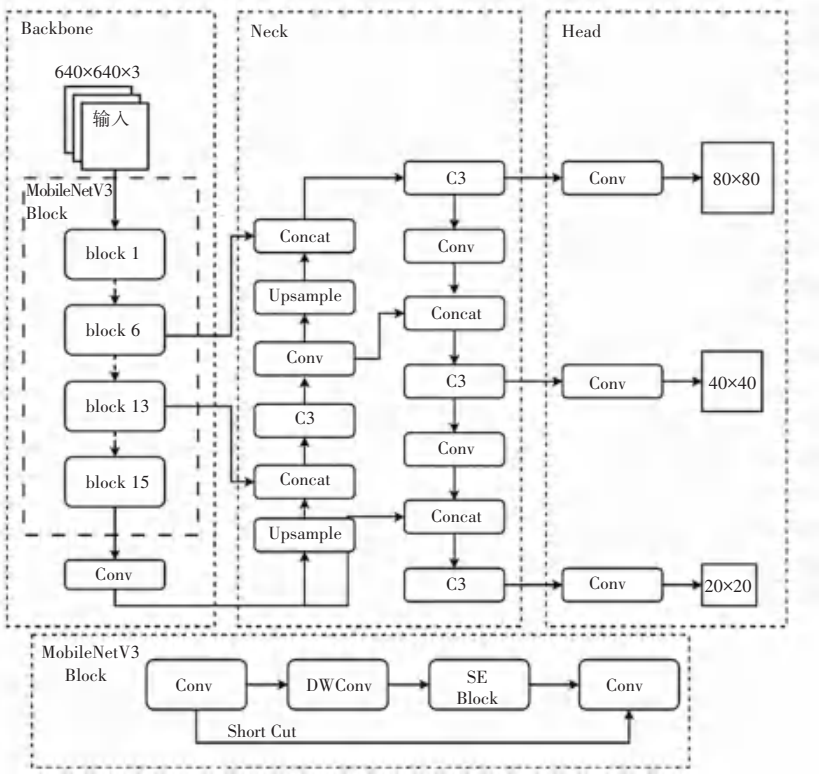

图 2　所提算法网络结构图

Fig. 2　Network structure diagram of the proposed algorithm

主干网络(Backbone)包含多个 MobileNetV3 模块和 1 个起连接作用的卷积模块。每个 MobileNetV3 模块由 2 个 1×1 的普通卷积和 1 个深度卷积以及 SE 注意力模块组成。1×1 的普通卷积用于改变特征图的深度。深度卷积用于特征提取,其通道数始终为 1,这减少了特征提取时卷积核的深度,从而极大减少模型推理计算量。SE 注意力模块在识别过程中自动为特征图赋予权重,以加强对关键特征的识别与关注,并提升模型对不同通道特征的区分能力。为了进一步减少计算量并加快模型推理速度,MobileNetV3 模块在设计时采用了 Shortcut 结构。

特征融合网络(Neck)和检测头(Head)与原 YOLOv5 框架保持一致。特征融合网络(Neck)处理后,将多尺度的特征图传递到检测头(Head),实现目标的分类和定位。特征融合网络主要采用特征金字塔网络(Feature Pyramid Network, FPN)和路径聚合网络(Path Aggregation Network, PAN)的组合结构,通过自上而下和自下而上的特征融合方式,有效捕捉不同尺度目标的特征信息。其中,自上而下路径用于增强高层语义信息对低层特征的指导能力,自下而上路径则用于补充底层细节信息,提升小目标的检测性能。在检测头(Head)部分,网络通过多尺度预测策略有效处理小、中、大目标的检测任务。检测头通过设置多个不同尺度的输出层,结合锚框机制和边界框回归,分别对多尺度特征图进行分类和定位,输出目标的类别概率和位置坐标。这种设计提高了模型对目标大小差异的适应能力,使其在多种复杂场景下都能保持良好的检测性能。每个输出层包含 1 个卷积层,用于进一步处理特征图,以生成精确的检测结果。在此过程中,引入了 CIoU(Complete Intersection over Union)损失函数,用于优化边界框。这种多层次的优化策略,使得检测头在实时性和精确性方面均表现出色。

2.2 基于特征匹配的目标框变换与融合方法

为解决多帧识别中的重复计数问题,本研究采用了图像特征点匹配技术,通过在相邻视频帧之间进行目标框的坐标变换与融合,去除重复信息,从而准确追踪烟盒位置并计数。具体而言,使用 SIFT 算法进行特征提取,并结合 FLANN 进行高效特征匹配,最后通过单应性矩阵实现目标框坐标的变换。

在完成目标框坐标变换后,进一步将变换后的目标框与前一帧的目标框进行比较,通过计算交并比来判断两者是否属于同一目标。当交并比值大于设定的阈值时,认为当前目标框与前一帧目标框为同一目标,并进行融合;若未找到匹配的目标框,则认为该目标为新目标,并将其加入追踪列表。该算法采用视频序列抽帧处理策略,通过对间隔采样的非连续帧进行跨帧目标关联和时序上下文建模,在降低计算量的同时有效地确保了目标位置的追踪精度。

通过精准的目标框匹配与融合,该方法能够有效减少因烟盒目标相似性或重复现象导致的计数误差。最终,变换与融合后的目标框及其类别信息被输出为图像和文本文件,文本文件可供进一步分析与处理;而将所有识别的目标框合成一幅完整图像,可直观展示不同品规卷烟在视频中的位置分布与变化趋势,为计数任务以及后续商业分析提供了可靠支持。

3 实验分析

3.1 数据集与实验平台

本实验数据来源于陕西省西安市卷烟销售点的零售柜台拍摄视频,总计 43 条视频,涵盖市场上最常见的 44 种卷烟品规。卷烟随机进行摆放以尽可能符合实际销售环境。从这些视频中抽取图片 214 张,并使用 LabelImg 标注工具对其进行标注,形成初始数据集。在抽取的 214 张图片中,覆盖 44 种卷烟品规,由于柜台陈列方式的随机性,每张图片中包含的卷烟种类数量并不固定。通过添加随机高斯噪声、调整亮度、随机裁剪(Cutout)、随机旋转以及平移和镜像的数据增强方法对初始数据集进行扩展,最终得到的数据集共包含 1 926 张图像,采用 4∶1 的比例将数据集划分为训练集和测试集。

本文的实验环境基于 Windows 11 操作系统,硬件配置包括 Intel i7-13700KF CPU 和 NVIDIA GeForce RTX 4090D GPU。模型使用 Python 3.8.19 和 PyTorch 2.2.2 框架构建,CUDA 版本为 11.8。

在训练过程中,模型的具体设置如下:迭代次数为 600 次,批量大小(Batch Size)为 16,初始学习率设为 0.01,并采用余弦退火机制动态调整学习率。此外,训练过程中使用随机梯度下降优化算法对模型参数进行优化。

3.2 评价指标

本文从以下方面对模型的综合性能进行评估:精度(*Precision*, *P*)、召回率(*Recall*, *R*)和均值平均精度(mean Average Precision, *mAP*)。数学定义公式如下:

$$\begin{cases} P = \dfrac{TP}{TP + FP} \\ r = \dfrac{TP}{TP + FN} \\ mAP = \dfrac{1}{N}\sum_{i=1}^{N} AP_i \\ AP = \int_0^1 P(r)\,\mathrm{d}r \end{cases} \tag{1}$$

其中，*TP* 表示正确检测到的正样本；*FP* 表示错误检测到的正样本；*FN* 表示未被检测到的正样本；*N* 表示数据集内样本类别总数。

平均精度（Average Precision，*AP*）表示模型在不同召回率值下精度的平均值。通常 *mAP@X* 则表示在交并比（Intersection over Union，*IoU*）阈值设定为 *X*% 时的 *mAP*。本文采用混淆矩阵（Confusion Matrix）对各个类别间的检测与分类效果进行可视化分析。

为了衡量模型的复杂程度，本文还从参数量（Number of Parameters）和浮点运算次数（Floating Point Operations，*FLOPs*）两个方面进行了量化分析。

3.3　实验结果分析

在本文构建的卷烟数据测试集上，所提出的检测模型在 44 个卷烟类别上的平均检测精度 *P* 达到 0.981，平均召回率 *R* 为 0.971。同时，*mAP@*0.5 达到 0.988，而 *mAP@*0.50 – 0.95 为 0.783。结果表明，该模型在大多数类别上的检测性能均达到较高水平，能够有效区分不同卷烟品规。检测结果如图 3 所示。各类别间的混淆矩阵如图 4 所示，除背景类（background）外，对角线元素均高于 0.75，表明模型在绝大多数类别上的检测精度较高。其中，大部分类别的检测精度超过 0.95。然而，编号为 FJQPL016 的卷烟品规由于样本数量较少（仅为其他品规卷烟的四分之一），检测精度相对较低，未超过 0.8。值得一提的是，即便使用 YOLOv5-s 和 YOLOv5-l 版本，该类别的检测精度仍未超过 0.8，这进一步验证了样本分布不均衡对模型检测精度的影响，表明在小样本类别上仍存在一定的优化空间。

（a）视频帧　　（b）该帧检测结果

图 3　检测结果展示

Fig. 3　Display of test results

为了验证本文所提出模型在性能和复杂度上的优势，将其与原 YOLOv5 中的 s（small）和 l（large）版本进行对比分析，结果见表 1。实验结果表明，在保证检测精度基本一致的前提下，本文模型显著降低了计算复杂度和参数规模。其中，与 YOLOv5-s 相比，计算量 *FLOPs* 减少了约 31.90%，参数量减少了约 20.78%；而与 YOLOv5-l 相比，计算量减少了约 89.82%，参数量减少了约 87.81%。这一结果表明，引入 MobileNetV3 主干网络可有效降低模型复杂度，同时保持较优的检测性能，使得模型在资源受限的环境下（如嵌入式设备或移动端应用）具备更强的部署能力。

表 1　各模型性能和复杂度对比

Table 1　Comparison of performance and complexity of different models

模型	精度（*P*）	召回率（*R*）	*mAP@*0.5	*mAP@*0.50 – 0.95	参数量	*FLOPs*
所提模型	**0.981**	0.971	0.988	0.783	**5 652 790**	**11.1**
YOLOv5-l	**0.981**	**0.977**	**0.990**	**0.784**	46 364 464	109.0
YOLOv5-s	0.980	0.967	0.982	0.781	7 135 600	16.3

为了统计整个视频中的卷烟类别数量，本文采用视频帧匹配技术对帧级检测结果进行融合，以获得视频级的统计结果。在测试集上的实验表明，最终统计得到的视频级检测准确率达到 97.75%，且漏检率为 0。该结果表明，本文方法能够从视频中高效提取并准确统计目标类别，进一步验证了其在视频数据分析任务中的有效性和优越性。

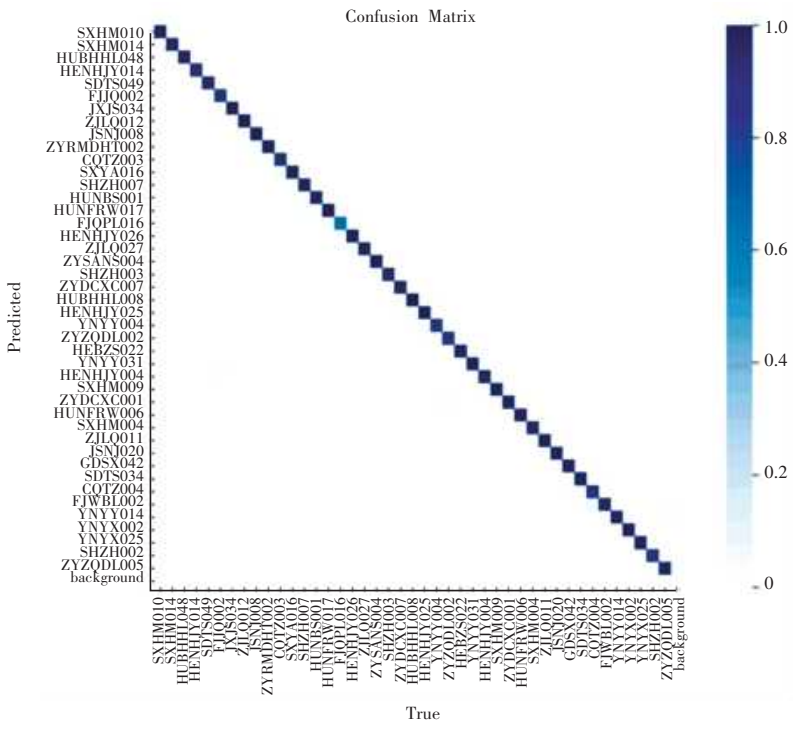

图 4 混淆矩阵

Fig. 4 Confusion matrix

3.4 实验结果界面展示与应用

本研究通过展示界面以更直观地呈现视频级别的检测结果,如图 5 所示。通过该界面,用户可以清晰地看到每个卷烟品规在视频中的实际分布情况。可视化界面不仅提供了直观的统计结果展示,还能够辅助卷烟工业企业更有效地进行库存管理和营销策略优化。具体阐释如下。

(1)实现库存监测,提供决策依据:利用本研究的视频检测技术,能精准识别零售终端柜台上不同品规的库存数量,提升库存盘点效率和准确性的同时,帮助工业企业了解产品的销售状态,合理调整投放节奏,确保产品的库存保持在合理水平。

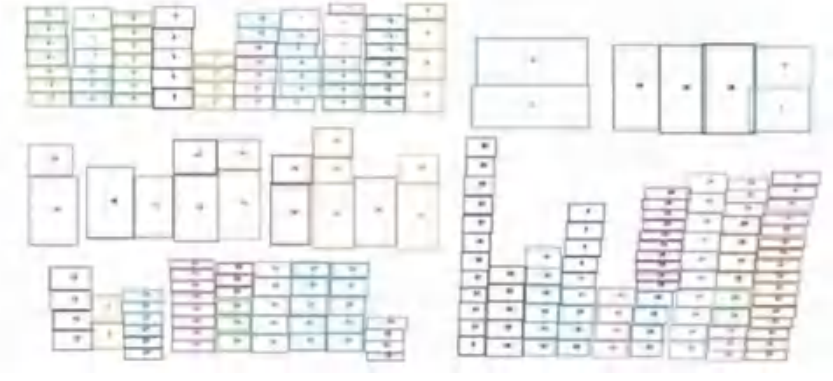

图 5 实验结果界面展示

Fig. 5 Experimental result interface display

(2)终端陈列指导与优化:本研究的视频检测可视化界面能够直观反映各种品牌卷烟在零售终端销售区域的摆放情况,使管理者能够定期检查各种及竞品卷烟陈列情况,从而准确分析不同品规的陈列方式对消费者的购买决策的影响,指导卷烟以更优的陈列方式进行展示,有助于提升品牌形象,促进销售。

(3)精准营销指导:通过分析零售终端不同产品的陈列方式、库存等信息,为工业企业制定营销策略、调整产品结构提供数据支撑,使营销策略更加精准有效,提高销售效率。

4 结束语

本文围绕卷烟品规展示检测任务,提出了一种

基于深度学习的视频级卷烟展示检测方法。通过改进 YOLOv5 模型并引入 MobileNetV3 主干网络，保证高检测精度的同时，大幅降低了计算复杂度，使模型更适用于实际业务场景。此外，结合视频帧匹配技术，对帧级别的检测结果进行融合，以统计整个视频中的卷烟品规数量。实验结果表明，该方法在卷烟数据测试集上达到了 97.75% 的视频级检测准确率，证明了其在实际应用中的有效性。

此外，可视化界面为卷烟工业企业提供了基于数据驱动的决策支持，这对于优化卷烟的库存控制、陈列布局以及营销策略等应用具有重要意义，有助于进一步提高管理效率和市场响应能力，从而在激烈的市场竞争中占据有利位置。未来，研究可进一步结合多摄像头数据融合、时序分析等技术[21]，以提升系统的智能化水平，为卷烟及相关行业的智能管理提供更加高效、精准的解决方案。

参考文献

[1] 叶永生. 基于 RFID 技术的烟草物流系统分析与设计[J]. 智能计算机与应用，2012，2(1)：66-68.

[2] 左国才，陈明丽，匡林爱，等. 基于深度学习的智能交通视频多目标检测研究[J]. 智能计算机与应用，2020，10(8)：180-182.

[3] 钟宇，徐燕，刘德祥，等. 基于计算机视觉和机器学习的真伪卷烟包装鉴别[J]. 烟草科技，2020，53(5)：83-92.

[4] 刘浩，贺福强，李荣隆，等. 基于机器视觉的卷烟小盒商标纸表面缺陷在线检测技术[J]. 中国烟草学报，2020，26(5)：54-59.

[5] 单宇翔，龙涛，楼卫东，等. 基于深度学习的复杂场景中卷烟烟盒检测与识别方法[J]. 中国烟草学报，2021，27(5)：71-80.

[6] 申玉鹏. 基于深度学习的香烟种类识别算法研究[D]. 南宁：广西民族大学，2023.

[7] 淡卫波，朱勇建，黄毅. 基于深度学习的烟包识别与分类[J]. 包装工程，2023，44(1)：133-140.

[8] 赵志成，罗泽. 基于注意力机制和深度残差网络的烟盒规格识别[J]. 计算机应用与软件，2023，40(9)：242-247.

[9] HOWARD A，SANDLER M，CHU G，et al. Searching for mobilenetv3 [C]//Proceedings of the IEEE/CVFInternational Conference on Computer Vision. Piscataway，NJ：IEEE，2019：1314-1324.

[10] GIRSHICK R，DONAHUE J，DARRELL T，et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，NJ：IEEE，2014：580-587.

[11] GIRSHICK R. Fast R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway，NJ：IEEE，2015：1440-1448.

[12] REN Shaoqing，HE Kaiming，GIRSHICK R，et al. Faster R-CNN：Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence，2017，39(6)：1137-1149.

[13] REDMON J，DIVVALA S，GIRSHICK R，et al. You only look once：Unified，real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，NJ：IEEE，2016：779-788.

[14] LIU Wei，ANGUELOV D，ERHAN D，et al. SSD：Single shot multibox detector [C]//Proceedings of the 14th European Conference on Computer Vision(ECCV). Cham：Springer，2016：21-37.

[15] JIANG Peiyuan，ERGU D，LIU F，et al. A review of Yolo algorithm developments[J]. Procedia Computer Science，2022，199：1066-1073.

[16] QI Jiangtao，LIU Xiangnan，LIU Kai，et al. An improved YOLOv5 model based on visual attention mechanism：Application to recognition of tomato virus disease [J]. Computers and Electronics in Agriculture，2022，194：106780.

[17] LOWE D G. Distinctive image features from scale-invariantkeypoints[J]. International Journal of Computer Vision，2004，60：91-110.

[18] MUJA M，LOWE D G. Fast library for approximate nearestneighbors[EB/OL]. (2015-11-17). https://www.cs.ubc.ca/research/flann/.

[19] DUBROFSKY E. Homography estimation [D]. Vancouver：Univerzita Britské Kolumbie，2009.

[20] DERPANIS K G. Overview of the RANSAC algorithm [J]. Image Rochester NY，2010，4(1)：2-3.

[21] ZHU Xingkui，LYU S，WANG Xu，et al. TPH-YOLOv5：Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway，NJ：IEEE，2021：2778-2788.