doi:10.6041/j.issn.1000-1298.2022.03.043

基于单阶段目标检测算法的羊肉多分体实时分类检测

赵世达[1]　王树才[1,2]　郝广钊[1]　张一驰[1]　杨华建[3]

(1.华中农业大学工学院,武汉 430070;2.农业农村部长江中下游农业装备重点实验室,武汉 430070;
3.青岛建华食品机械制造有限公司,胶州 266300)

摘要:针对输送带场景中羊肉分体需要进一步分类检测问题,提出一种基于单阶段目标检测算法的羊肉多分体实时分类检测方法。在羊屠宰车间环境下采集包含多类、多个羊肉分体图像,经图像增广及归一化后建立羊肉多分体图像数据集,其中训练集 7 200 幅,测试集 1 400 幅,验证集 400 幅。利用单阶段目标检测算法 YOLO v3 引入迁移学习对羊肉多分体图像数据集展开训练并获得最优模型,基于最优模型返回图像中各羊肉分体的类别及其位置,从而实现羊肉分体的分类检测。选用平均精度及单幅图像平均处理时间作为评判模型检测精度与速度指标。然后通过更换羊肉多分体识别模型的特征提取网络优化检测速度。另外设置包含亮、暗两种亮度水平的附加光照数据集以及代表羊肉分体遮挡情形的附加遮挡数据集,分别验证优化后模型的泛化能力与抗干扰能力,并通过多尺度特征明显的颈部与腹肋肉测试优化后模型的鲁棒性。最后引入 Mask R-CNN、Faster R-CNN、Cascade R-CNN 和 SSD 4 种常用目标检测算法针对不同数据集分别进行对比试验,在此基础上,进一步更换特征提取网络为 MobileNet V1、ResNet34 和 ResNet50 验证优化后模型的综合检测能力。试验结果表明,优化后模型的检测速度较原始模型提升 48.53%,同时对光照、遮挡复杂环境下羊肉多分体识别具备较强的泛化能力与抗干扰能力,以及对多尺度特征显著的羊肉分体检测具有良好的鲁棒性,针对羊肉多分体图像验证集,优化后羊肉多分体识别模型的平均精度达到 88.05%,单幅图像处理时间为 64.7 ms,综合检测能力优于其他算法,说明该方法具备较高的检测精度和良好的实时性,能够满足实际生产需求。

关键词:羊肉多分体;目标检测;YOLO v3;优化;试验

中图分类号:S24;TP391　　**文献标识码**:A　　**文章编号**:1000-1298(2022)03-0400-12　　**OSID**:

Mutton Multipartite Real-time Classification and Detection Based on Single-stage Object Detection Algorithm

ZHAO Shida[1]　WANG Shucai[1,2]　HAO Guangzhao[1]　ZHANG Yichi[1]　YANG Huajian[3]

(1. *College of Engineering*, *Huazhong Agricultural University*, *Wuhan* 430070, *China*
2. *Key Laboratory of Agricultural Equipment in Mid-lower Yangtze River*,
Ministry of Agriculture and Rural Affairs, *Wuhan* 430070, *China*
3. *Qingdao Jianhua Food Machinery Manufacturing Co.*, *Ltd.*, *Jiaozhou* 266300, *China*)

Abstract: Aiming at the problem that mutton multipartite needs to be further classified and detected in the conveyor belt scene, a real-time classification and detection method for mutton splits based on a single-stage object detection algorithm was proposed. In the sheep slaughter workshop environment, multiple types and multiple mutton split images were collected. After image augmentation and normalization, a mutton multipartite image data set was established, including 7 200 training sets, 1 400 test sets, and 400 verification sets. Using the single-stage object detection algorithm YOLO v3 to introduce transfer learning to train the mutton multipartite image data set and obtain the optimal model. Based on the optimal model, the category and position of each mutton split in the image were returned, so as to realize the classification and detection of mutton multipartite. The average accuracy mAP and the

收稿日期:2021-03-25　修回日期:2021-04-13

基金项目:国家重点研发计划项目(2018YFD0700804)

作者简介:赵世达(1992—),男,博士生,主要从事家畜产品智能加工与检测研究,E-mail:zsd@webmail.hzau.edu.cn

通信作者:王树才(1966—),男,教授,博士生导师,主要从事农产品智能化检测与装备研究,E-mail:wsc01@mail.hzau.edu.cn

average detection time of a single image were selected as the accuracy and speed indicators for judging the detection effect of the model. Then, the detection speed was optimized by replacing the feature extraction network of the mutton multipartite recognition model. In addition, an additional illumination data set containing two brightness levels of "bright" and "dark" and an additional occlusion data set representing the occlusion situation of mutton were set to verify the generalization ability and anti-interference ability of the optimized model, and the robustness of the optimized model was tested through the neck and abdominal rib with obvious multi-scale features. Finally, four commonly used object detection algorithms: Mask R－CNN, Faster R－CNN, Cascade R－CNN, and SSD were introduced to conduct comparative experiments on different data sets. On this basis, the feature extraction network was further replaced with MobileNet V1, ResNet34 and ResNet50 to verify the optimized model's comprehensive testing capabilities. The test results showed that the detection speed of the optimized model was 48.53% higher than that of the original model. At the same time, it had strong generalization ability and anti-interference ability for multi-part recognition of mutton under complex environment of light and shading, and it had good robustness to the mutton multi-part detection with multi-scale features. It was optimized for the verification set of mutton multipartite image. The mAP value of the optimization model reached 88.05%, and the processing time of a single image was 64.7 ms, the comprehensive detection ability was better than that of other algorithms, which indicating that this method had high detection accuracy and good real-time performance, and can meet actual production needs.

Key words: mutton multipartite; object detection; YOLO v3; optimization; experiment

0　引言

羊肉分体是依据家畜屠宰行业分割标准将羊胴体分割后得到的产物，一般为 6 分体、8 分体两大类，与大众饮食习惯息息相关，同时与家畜产品增值存在直接关系。目前，我国羊肉产品生产加工企业大多采用半自动化生产工艺，特别是羊肉分体分拣环节，全部依赖人工完成，工作量大，环境恶劣，且存在一定的食品安全隐患[1-2]。因而对羊肉分体进行自动化分拣很有必要，但完成以上目标的先决条件在于羊肉分体种类的准确、快速获取，因此，实现羊肉多分体的实时分类检测具有重要意义。

近年来，在羊体目标检测研究中，广泛采用的方法大多基于图像处理、光谱技术结合机器学习展开。即根据羊体或各羊肉分体在图像中所呈现的不同特征表达进行特征提取，例如颜色、纹理、轮廓、区域中心坐标等，再将各特征以机器学习的方式构建分类器[3-6]。这些方法虽然可以实现特定羊体参数准确识别，但存在人工提取目标特征困难，工作繁琐，选用机器学习设计的模型适应性差、实时性不足的缺点，特别是羊肉分体种类多样、自然特征表达复杂、部分特征具有相似情形，同时，输送带场景中包含多个、多类且可能遮挡的羊肉分体，以上特点对上述文献方法的展开带来困难。

基于深度卷积神经网络的目标检测方法具备对简单、抽象特征的强大自主学习能力，其准确率、泛化能力均优于基于传统 HOG[7]、SIFT[8]等人为特征结合机器学习的检测方法，能够较好地解决上述问题。现阶段，众多学者将基于深度学习的目标检测应用至各家畜目标识别任务中[9-14]。基于深度学习的目标检测算法在家畜养殖领域中的良好表现，使得探讨将该方法用于羊肉多分体实时分类检测成为可能。

目前，基于深度卷积神经网络的目标检测主要包括两类，一类是先生成候选区域，再对候选区域进行分类和回归的二阶段检测器，检测精度高，但速度较慢[15]。代表有 Faster R－CNN[16]、Mask R－CNN[17]、TridentNet[18]；另一类是不需要产生候选区域，基于回归思想直接进行目标分类和回归的单阶段检测器，较二阶段检测器具备较优的实时性，但精度较差[19-20]，例如 YOLO v1－v3[21]、SSD[22]、RetinaNet[23]等。本文首先建立羊肉分体图像数据集，然后基于检测速度较快的单阶段目标检测算法 YOLO v3 针对羊肉多分体图像数据集展开研究，获取输送带场景下图像中的多个、多类羊肉分体的类别与位置，然后更换主干网络为轻量级神经网络 MobileNet V3[24]优化羊肉多分体识别模型的检测速度。在此基础上，建立包含亮、暗两种亮度水平的附加光照数据集以及代表羊肉分体遮挡情形的附加遮挡数据集，用以验证优化后模型的泛化能力以及抗遮挡能力，并且通过选用形体差异较为显著的颈部、腹肋肉分体测试优化后模型的鲁棒性。最后引入 Mask R－CNN、Faster R－CNN、Cascade R－CNN、SSD 4 种常用的目标检测算法进行对比试验，并验证 MobileNet V1[25]、ResNet34、ResNet50[26]共 3 种特征提取网络下模型的识别性能。

1 材料与方法

1.1 试验材料与图像采集

试验以成年波尔山羊经分割后得到的羊肉 6 分体作为研究对象,图像采集于内蒙古自治区美洋洋食品股份有限公司的羊胴体分割生产车间。在输送带场景下,固定华谷动力科技公司生产的 WP-UC600 型 CCD 相机于羊肉分体样本正上方 0.6 m 处,搭配 Z4S-LE-SV-1214H 型欧姆龙镜头获取图像,期间无特定背景及光源,不限定拍摄时间。由于同分割批次的羊胴体存在质量、体态、年龄相似情况,为增大羊肉各分体之间的差异性以提高模型的泛化能力,随机采集 6 批次羊胴体,共计 2 100 副。最终经人工筛选后得到 3 600 幅包含多个、多类的羊肉分体图像,分辨率为 2 448 像素 ×3 264 像素。图像采集装置示意图与样本示例如图 1、2 所示。

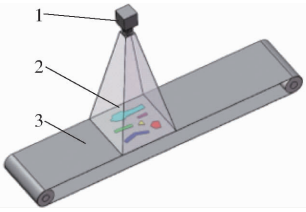


图 1 羊肉多分体图像采集装置示意图

Fig. 1 Illustration of image acquisition device

1. 相机 2. 羊肉多分体 3. 输送带



图 2 羊肉多分体图像样本示例

Fig. 2 Example of multipartite images of mutton

本文羊肉 6 分体分别为腰椎骨、胸椎骨、颈部、腹肋肉、肩胛骨和腿骨,其中,腰椎骨与胸椎骨是制作羊蝎子和法排的原材料,腹肋肉使用场景更为广泛,其它 3 类也有其独特价值,如图 3 所示。可以看出各类羊肉分体表现出的颜色、纹理、轮廓等特征差异明显,并且尺寸区别较大。另外,通常情况下由于羊体尺寸不同导致相同种类的羊肉分体尺寸也存在较大的差异,多尺度特征显著。以上特点要求羊肉多分体识别方法需要具备多尺度特征检测能力。

1.2 样本预处理及生成数据集

基于深度卷积神经网络的目标检测精度与样本数据集规模显著相关,在保证图像自然特征表达情况下,扩充训练图像数量不仅能够提高模型准确率,还可以满足深层网络对数据的要求,减少过拟合现象[27],因此,对本试验小规模羊肉分体图像进行图像增广以扩充数据量很有必要。本文采用翻转、旋转、平移 3 种方式进行图像数据扩充,再以比例不变性原则将图像缩放至 412 像素 ×412 像素,建立羊肉多分体图像数据集,其中训练集 7 200 幅,测试集 1 400 幅,验证集 400 幅。另外,参考 VOC 数据集格式使用 LabelImg 图像标注工具对 6 类羊肉分体进行标记,设定腰椎骨、胸椎骨、颈部、腹肋肉、肩胛骨和腿骨的标签分别为 Sheep1、Sheep2、Sheep3、Sheep4、Sheep5 和 Sheep6。最终,据统计,羊肉多分体图像数据集中上述各类羊肉分体的数量比例分布趋近于 5∶4∶3∶6∶2∶7,各分体的数量如表 1 所示。

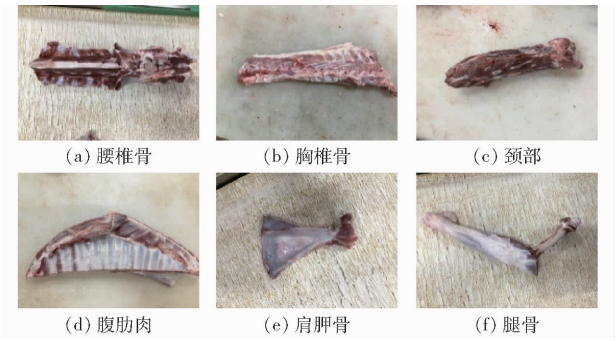


图 3 羊肉各分体图像样本示例

Fig. 3 Sample image samples of each split of mutton

表 1 羊肉分体数据集种类及其数量

Tab. 1 Types and quantities of mutton multi-part datasets

数据集	腰椎骨	胸椎骨	颈部	腹肋肉	肩胛骨	腿骨
训练集	3 796	3 003	2 509	4 732	1 781	4 875
测试集	801	590	513	986	366	1 139
验证集	245	163	134	317	98	382
总计	4 842	3 756	3 156	6 035	2 245	6 396

1.3 YOLO v3 算法原理

1.3.1 YOLO v3 网络结构

YOLO v3 属于单阶段目标检测算法的一种,该网络结构由 DarkNet53 特征提取网络与 YOLO v3 多尺度分支两部分组成。其中 DarkNet53 特征提取网络通过设置阶跃连接加深网络深度,以提高特征提取能力,同时加快模型收敛。YOLO v3 多尺度分支部分引入特征金字塔(Feature pyramid networks, FPN)结构将网络划分为 3 个分支,分别对应检测小目标、中目标和大目标,利用获取目标多尺寸特征图感受野的形式,增强多尺度目标检测能力,提升模型的鲁棒性。YOLO v3 网络结构如图 4 所示。

首先,将尺寸为 412 像素 ×412 像素的羊肉多分体图像输入 DarkNet53 特征提取网络,经特征通道数为 32 与 64、卷积核尺寸为 3 ×3 的卷积核以步长 1 和步长 2 遍历图像。然后卷积层(Conv)输出

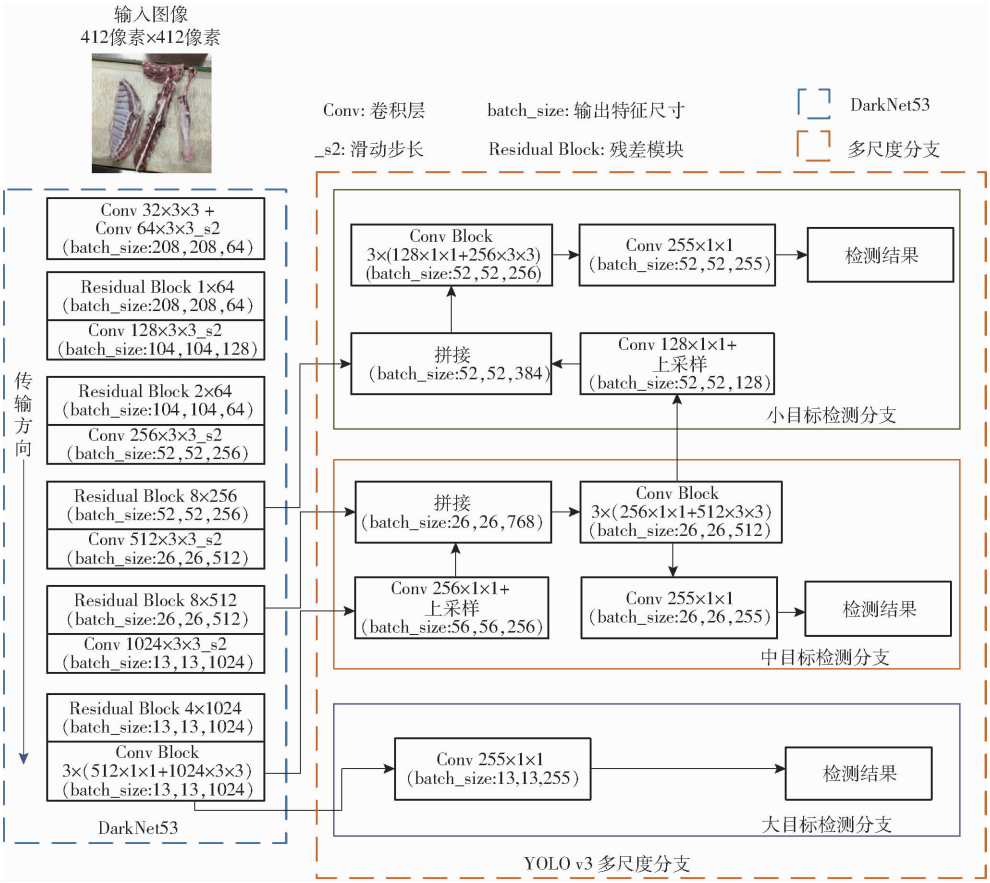


图 4　YOLO v3 网络结构图

Fig. 4　Architecture for YOLO v3 network

特征由批归一化(BN)和 Leaky_ReLU 激活函数处理后得到 208 × 208 × 64 的特征图。最后输入卷积核尺寸为1 × 1 的残差模块,通过下采样输出尺寸为 104 × 104 × 128 的特征图,之后交替卷积,直至输出尺寸为 13 × 13 × 1 024 的特征向量至 YOLO v3 多尺度分支的大目标和中目标分支。其中,中目标分支将特征向量上采样后与 DarkNet53 网络中输出尺寸为26 × 26 × 512 的特征图拼接,经多次卷积后一部分输入至小目标分支,另一部分输出尺寸为26 × 26 × 255 的特征结果。输入至小目标分支的特征向量上采样后与 52 × 52 × 256 的残差模块输出特征拼接,最终得到 52 × 52 × 255 的特征图,由于该特征图感受野较小,使得 YOLO v3 对小尺寸目标能够保持较高的检测精度。

1.3.2　边界框预测及损失函数

YOLO v3 依据多尺度分支的 3 种特征图尺寸(S = 13,26,52),将输入图像划分为 $S \times S$ 个网格,并对落入各网格中的目标进行滑动窗口计算,输出被检目标的边界框及其置信度。该置信度由各网格包含目标的概率与边界框准确度共同决定[28]。

YOLO v3 采用 9 种锚点框(Anchor box),提高对多尺度目标的检测精度,各锚点框尺寸分别为 10 × 13、16 × 30、33 × 23、 30 × 61、62 × 45、59 × 119、116 × 90、156 × 198、373 × 326。其中,前 3 个锚点框适用于检测小尺寸的肩胛骨,中间 3 个锚点框对检测尺寸相对中等的腰椎骨、颈部和腿骨较为敏感,剩余锚点框更适用于检测尺寸较大的胸椎骨和腹肋肉。然后基于锚点框得到各检测目标预测边界框,再将置信度小于阈值的边界框剔除,采用非极大值抑制算法(Non-maximum suppression,NMS)[29],获取最终的被检目标边界框(Bounding boxes),本文设定阈值为 0.5。该边界框尺寸与位置由 4 种参数 b_x、b_y、b_w、b_h 决定,分别对应边界框相对于特征图的中心点坐标以及宽、高。图 5 为YOLO v3 边界框原理图,其中 C_x、C_y 代表预测边界框中心所在网格左上角网格的坐标,P_w、P_h 为该单元格对应先验框的宽和高,t_x、t_y 为预测边界框中心距所处网格左上角的相对宽、高偏移量。b_x、b_y、b_w、b_h 计算式为

$$\begin{cases} b_x = \sigma(t_x) + C_x \\ b_y = \sigma(t_y) + C_y \end{cases} \tag{1}$$

$$\begin{cases} b_w = P_w e^{t_w} \\ b_h = P_h e^{t_h} \end{cases} \tag{2}$$

式中　σ——sigmoid 函数

t_w——预测边界框的宽方向偏移量

t_h——预测边界框的高方向偏移量

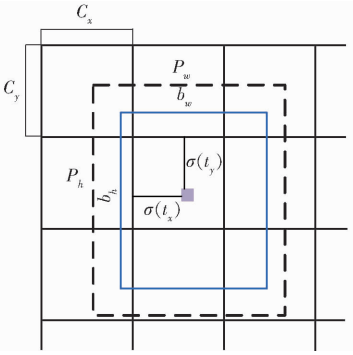


图 5 YOLO v3 边界框原理图

Fig. 5 Schematic of YOLO v3 boundary frame

损失函数用来判断模型的预测值与真实值之间的差异程度,与网络的性能直接相关。YOLO v3 网络训练过程中的损失值主要由边界框定位损失、边界框尺寸损失、置信度损失和类别损失 4 部分组成,该值越小则反映模型的训练效果越好,鲁棒性越强。其计算式为

$$\begin{aligned}L=&\lambda_{\text{coord}}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{\text{obj}}[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2]+\\&\lambda_{\text{coord}}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{\text{obj}}[(w_i-\hat{w}_i)^2+(h_i-\hat{h}_i)^2]+\\&\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{\text{obj}}[C_i\lg\hat{C}_i+(1-C_i)\lg(1-\hat{C}_i)]+\\&\lambda_{\text{noobj}}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{\text{noobj}}(C_i-\hat{C}_i)^2+\\&\sum_{i=0}^{S^2}I_{ij}^{\text{obj}}\sum_{C\in\text{classes}}(p_i(C)-\hat{p}_i(C))^2\end{aligned}\tag{3}$$

式中 I_{ij}^{obj}、I_{ij}^{noobj}——第 i 个网格范围内第 j 个锚点框负责检测目标、不负责检测目标

B——每个网格中预测边界框的个数

x_i、y_i——第 i 个网格边界框中心横、纵坐标

$\hat{x}_i$、$\hat{y}_i$——实际边界框中心横、纵坐标

w_i、h_i——第 i 个网格边界框宽度、高度

$\hat{w}_i$、$\hat{h}_i$——实际边界框中心宽度、高度

C_i——预测目标类别 $\hat{C}_i$——实际目标类别

p_i——预测类别概率

$\hat{p}_i$——实际类别概率 λ_{coord}——权重系数

λ_{noobj}——惩罚权重系数

1.4 判断标准

本文采用目标检测任务中常用的平均检测精度(mAP)作为量化分析羊肉多分体识别模型对羊肉分体分类检测效果的评判指标,该指标反映的是模型对腰椎骨、胸椎骨、颈部、腹肋肉、肩胛骨和腿骨 6 类羊肉分体类别检测精度(AP)的平均值,数值越大则表征模型的准确度越高。

为满足生产线实际需求,羊肉多分体识别模型不仅需要具备较高的检测精度,还要求良好的实时性。因此,除获取检测精度之外,测试模型的检测速度同样重要。本文选用模型单幅图像平均处理时间作为第二指标,判断模型的实时性。

2 试验与结果分析

基于单阶段目标检测算法的羊肉多分体实时分类检测主要包含 4 个步骤:采集羊肉多分体图像,经过预处理后建立羊肉多分体图像数据集;引入迁移学习针对羊肉多分体图像数据集训练 YOLO v3,返回检测精度最优模型,并更换最优模型特征提取网络为轻量级神经网络 MobileNet V3,优化模型的检测速度;分别通过附加亮度、遮挡数据集验证优化后模型的泛化能力和抗遮挡能力,以及测试优化后模型的鲁棒性;与常用目标检测算法进行对比试验,并进一步更换模型的特征提取网络验证优化后模型的综合检测能力。流程如图 6 所示。

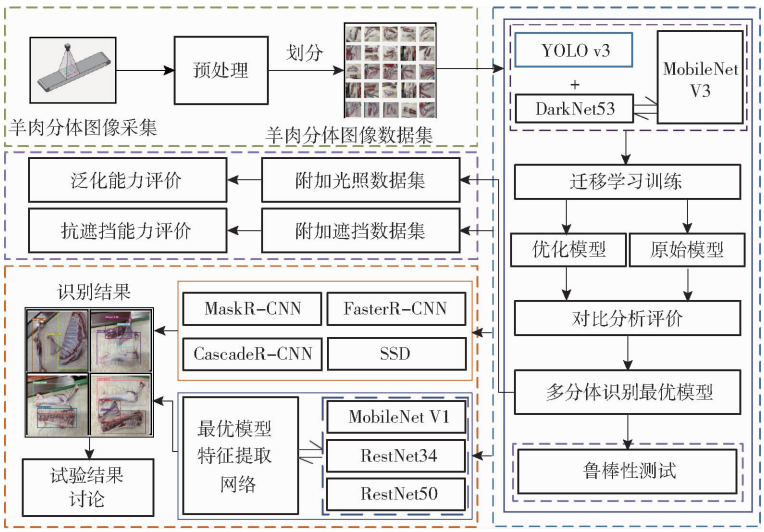


图 6 羊肉多分体实时分类检测试验流程图

Fig. 6 Technical flowchart of experiment for real-time classification and detection test of mutton multipartite

2.1　试验环境

本文试验基于深度学习框架 Tensorflow 在 Dell T5810 塔式图形处理工作站上完成，计算机系统为 Windows 10 专业版操作系统，搭载 Intel core64 至强 3.70 GHz W－2145 处理器，内存和显卡分别为 64 GB、NVIDIA P4000－8G。

2.2　基于迁移学习的 YOLO v3 网络训练

面对小规模数据样本的深度学习模型训练时，引入迁移学习可以降低过拟合及加速模型收敛，提升训练效果。因此，本文试验建立的所有模型在训练过程中均加载了基于 ImageNet 数据集的预训练权值。采用随机梯度下降（Stochastic gradient descent，SGD）优化器进行梯度下降，设置初始学习率及动量因子分别为 0.001 和 0.9，batchsize 和 gamma 系数为 8 和 0.1。另外，设定训练过程中每迭代 2 000 次保存一次结果，共计 40 000 次，将保存的 20 个训练结果比较后，输出最优模型用于后期羊肉多分体图像数据集的输入。针对羊肉多分体图像数据集，引入迁移学习和未引入迁移学习的 YOLO v3 网络损失值及平均检测精度随迭代次数与循环次数的变化趋势如图 7、8 所示。

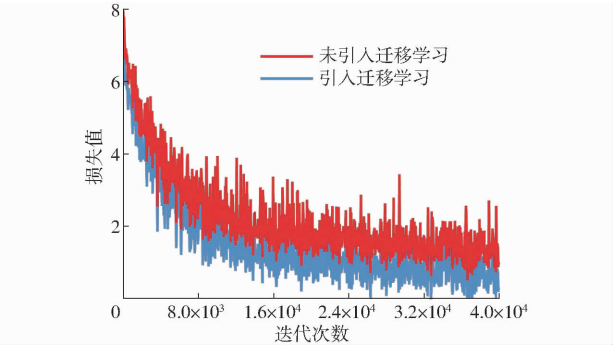


图 7　引入迁移学习与未引入迁移学习损失值随迭代次数变化曲线

Fig. 7　Loss value changes with number of iterations during transfer learning and non-transfer learning

由图 7、8 可知，针对羊肉多分体图像数据集，加载预训练权值的 YOLO v3 网络损失值在训练初期迅速降低，当迭代次数达到 8 000 次时缓慢下降随即平滑，直至收敛于 1.07。未加载预训练权值网络的训练损失值随迭代次数变化趋势与加载训练相似，均在训练开始时迅速减小，之后缓慢收敛，但其损失值明显大于采用迁移学习的训练方式。另外，针对羊肉多分体图像测试集，迁移学习训练得到的平均检测精度达到 88.72%，较未引入迁移学习提升 6.78 个百分点，说明基于迁移学习方法训练网络可以降低羊肉多分体识别模型的损失值，提升识别精度。因此，为保证模型的识别准确率，试验后期所采用模型的训练均基于迁移学习方式完成。

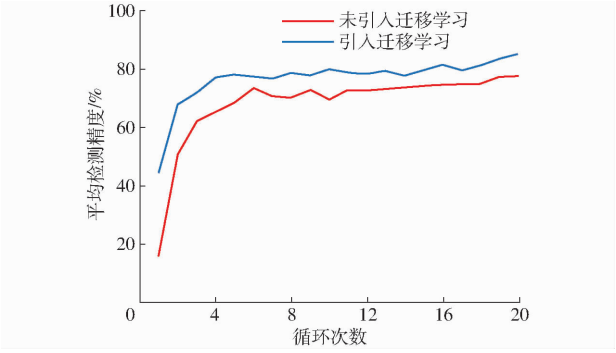


图 8　引入迁移与未引入迁移学习平均检测精度随循环次数变化曲线

Fig. 8　The mAP changes with number of epoch during transfer learning and non-transfer learning

2.3　YOLO v3 对羊肉多分体数据集的识别结果

基于训练得到的最优模型针对羊肉多分体验证集展开测试，获取图像中 6 类羊肉分体的类别和位置，得到识别平均检测精度及单幅图像处理时间，以判断 YOLO v3 模型对羊肉多分体的识别精度和实时性。其中，获取单幅图像处理时间的方式为分别记录验证集各图像识别时间再求其均值。部分识别结果如图 9 所示。

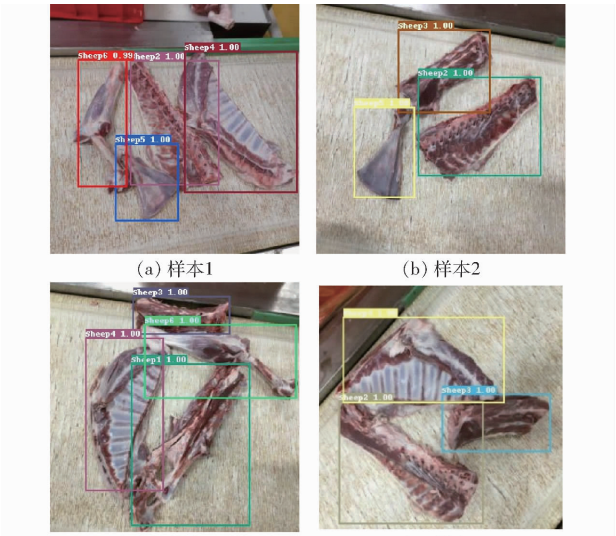


图 9　针对羊肉多分体验证集 YOLO v3 模型识别结果

Fig. 9　Recognition results of YOLO v3 model for mutton multipartite validation data sets

实际生产过程中，羊体分割以间歇性、周期性的作业方式完成，导致分割得到的羊肉分体在输送带上呈现出随机片区的形式，表现为单幅图像包含多个、多类且重复的羊肉分体现象，如图 9 所示。图中，YOLO v3 模型检测结果中没有出现错检与漏检的现象，可以以较高的置信度检测出各羊肉分体的类别，同时，各羊肉分体的边界框标记完整、准确，特别是图 9a、9b 中小尺寸的肩胛骨以及图 9c、9d 中尺

寸明显不同的颈部均可以准确识别,以上结果说明 YOLO v3 羊肉多分体模型具备较高的检测精度以及良好的多尺度检测能力。最终,YOLO v3 对羊肉多分体验证集的平均检测精度和单幅图像平均处理时间分别达到 87.79% 和 96.1 ms,精度满足生产线的需求,但处理速度略慢,实时性有待提高。

2.4 YOLO v3 网络检测速度优化

YOLO v3 的特征提取部分为包含 53 个卷积层的卷积神经网络 DarkNet53,该网络较深的多层结构有利于提取羊肉分体的多种抽象特征用于目标定位从而提升模型的识别精度,但随之增加的浮点计算量也加大了模型推断耗时,为实时性要求较高的识别任务带来困难。如果采用轻量级神经网络替代 DarkNet53,则可以优化 YOLO v3 的检测速度。MobileNet V3 网络属于轻量级神经网络中的一种,该网络在 MobileNet V1 深度可分离卷积模块(Depthwise separable convolution, DSC)和 MobileNet V2 逆残差结构(Inverse residual structure,IRS)的基础上添加了引入注意力机制的逆残差模块(Squeeze-and-excitation,SE),极大地减少网络参数量,有益于缩短模型的计算时间,同时采用 H-Swish 激活函数增加网络提取低维特征的非线性,例如羊肉各分体的颜色、纹理等特征,其结构如图 10 所示。因此,采用轻量级神经网络 MobileNet V3 替换 DarkNet53 作为 YOLO v3 的特征提取网络,试图在保证模型识别精度的前提下提升实时性。

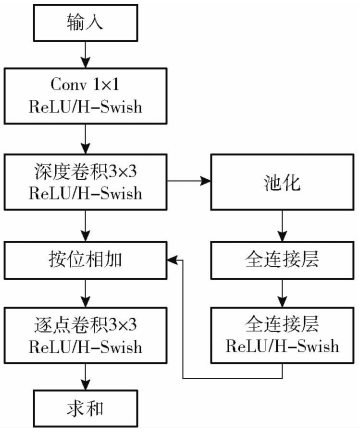


图 10 引入注意力机制的逆残差模块结构

Fig. 10 Structure of inverse residual module with SE

本文基于 MobileNet V3 替换特征提取网络 DarkNet53 得到优化后的 YOLO v3-MobileNet V3 模型,引入迁移学习进行训练,另外为保证模型收敛,将迭代次数扩充至 60 000 次,其余超参数设置与 YOLO v3 原始模型保持一致,之后针对羊肉多分体图像验证集展开识别精度和速度测试。训练过程中 YOLO v3-MobileNet V3 模型的平均检测精度随迭代次数变化趋势如图 11 所示,随着循环次数的增加,平均检测精度相应迅速提升然后平滑,直到稳定于循环 25 次左右。最终,针对验证集,训练得到最优 YOLO v3-MobileNet V3 模型的平均检测精度达到 88.05%,与原始模型差距不显著,仍可以准确识别图像中羊肉多分体的类别与位置,但单幅图像平均处理时间为 64.7 ms,较原始模型缩短 31.4 ms,检测速度较原模型提升 48.53%,说明更换特征提取网络为 MobileNet V3,能够有效缩短对羊肉多分体的检测时耗,提升实时性,但对检测精度无明显影响。

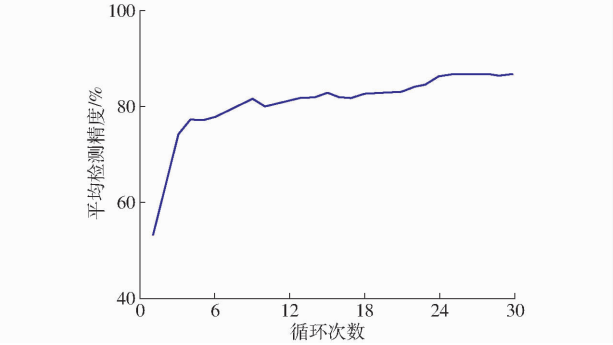


图 11 优化模型平均检测精度随循环次数变化曲线

Fig. 11 Change trend of mAP value of optimization model

2.5 不同光照下羊肉多分体识别

为测试优化后羊肉多分体识别模型的泛化能力,随机选取 600 幅归一化后的羊肉分体图像,建立附加光照数据集,并将其 RGB 色彩空间转换为 HSV 色彩空间,基于 Matlab 图像处理工具箱设置 1.3 倍和 0.7 倍的图像亮度,以模拟亮、暗两种光照水平场景,进行羊肉多分体识别模型的泛化能力试验,其中调亮、调暗图像各 300 幅。针对附加光照数据集,优化后模型的部分识别结果如图 12 所示。

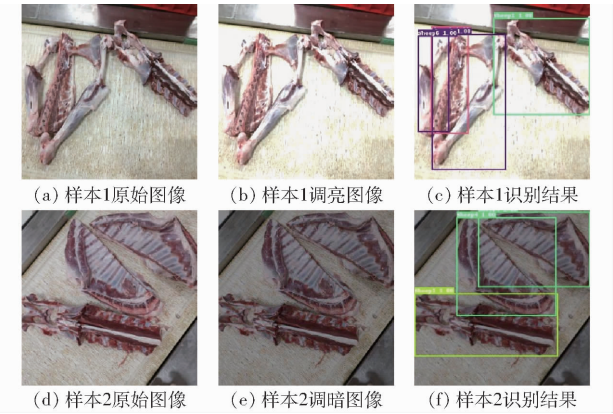


图 12 针对附加光照数据集优化模型识别结果

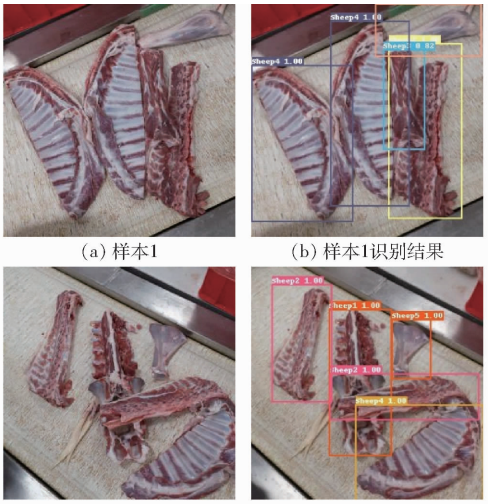
Fig. 12 Recognition results of optimize model for additional lighting data sets

根据图 12 样本 1、样本 2 识别结果可以看出,羊肉多分体在亮、暗两种光照强度下,优化后模型仍

然可以准确识别图像中 6 类羊肉分体的类别，同时检测得到的各分体边界框标记正确、完整，羊肉分体真实轮廓与标记边界框无明显误差现象。最终，针对附加验证集，优化后羊肉多分体识别模型的平均检测精度达到 85.16%，说明光照强度变化对该模型的识别效果影响并不明显，能够适应于不同亮度场景的羊肉多分体识别，具备较强的泛化能力。

2.6　遮挡状态下羊肉多分体识别

分割完成后的羊肉分体随机散落在呈匀速运动的输送带上，这种不确定的散落方式使得输送带上的片区羊肉分体存在少量叠放状况，造成部分羊肉分体被遮挡。如图 13a 所示，可以看出颈部分体叠落在胸椎骨与腹肋肉之间，造成腹肋肉部分特征丢失，图 13c 中胸椎骨分体置于腰椎骨上方，遮挡了部分腰椎骨，特别是颈部分体和腹肋肉分体的颜色、纹理特征十分相似，这种区域相交且特征相似的情况不利于被遮挡状态下的羊肉多分体识别。但是，图中羊肉分体未被遮挡部分仍保留了清晰、独立的自然特征，有利于排除羊肉分体遮挡造成的识别干扰。为测试羊肉多分体识别模型对遮挡状态下羊肉分体的识别效果，人工选取带遮挡特征的羊肉多分体图像 200 幅，作为附加遮挡数据集，对模型的综合识别检测能力展开测试，部分识别结果如图 13 所示。



(a) 样本1　(b) 样本1识别结果

(c) 样本2　(d) 样本2识别结果

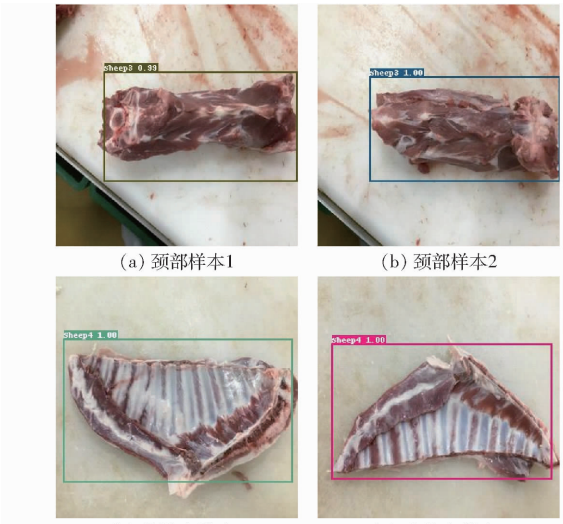
图 13　针对附加遮挡数据集优化模型识别结果

Fig. 13　Recognition results of optimize model for additional occlusion data sets

根据图 13a 样本 1、图 13c 样本 2 遮挡状态下的羊肉多分体识别结果，优化后的羊肉多分体识别模型可以准确区分叠落、遮挡状态下的颈部-胸椎骨、胸椎骨-腰椎骨。这可能是胸椎骨被颈部遮挡的像素区域面积相对未被遮挡区域较小，且存在粘连，对胸椎骨真实边界框定位影响较弱，从而识别模型仍能排除干扰，确定胸椎骨在图像中的位置，如图 13a、13b 所示。而图 13c 中腰椎骨被胸椎骨截断为中、小尺寸 2 部分，不利于单一目标的识别，但 YOLO v3 结构中的中目标识别分支、小目标识别分支，使得模型对截断后的多种尺寸部分敏感，仍然可以提取特征用于分类识别排除遮挡干扰。另外，YOLO v3 的 9 种尺寸不同的锚点框能够充分囊括同一类目标被遮挡后的剩余部分，便于在求取真实边界框过程中将像素交并比及边界框置信度进行比较和统一，降低遮挡部分的误差。最终，针对附加遮挡数据集，优化后的羊肉多分体识别模型的平均检测精度达到 83.47%，能够准确识别被遮挡羊肉多分体的类别及其位置，具备较强的抗遮挡能力与良好的识别性能。

2.7　优化后模型鲁棒性测试

羊肉多分体中同类分体之间存在个体尺寸差异较大的情况，为测试优化后模型对尺寸差异羊肉分体的识别能力，选择形体区别相对明显的颈部和腹肋肉进行鲁棒性试验。采集 2 组颈部与腹肋肉样本各 200 个，共计 400 个样本，建立附加多尺寸数据集。基于训练过程保存的最优优化后模型针对附加多尺寸数据集展开测试，部分识别结果如图 14 所示。



(a) 颈部样本1　(b) 颈部样本2

(c) 腹肋肉样本1　(d) 腹肋肉样本2

图 14　针对附加多尺寸数据集优化模型识别结果

Fig. 14　Recognition results of optimize model for additional multi-dimensional data sets

由图 14 可以看出，相同类别的颈部和腹肋肉在外形轮廓、体尺寸上存在明显不同，特别是图 14a 颈部样本 1 中，颈部右侧端部的锁骨特征较颈部样本 2 不显著，图 14c 腹肋肉样本 1 与腹肋肉样本 2 相比，尺寸偏大且相对饱满，以上特点不利于模型的精准识别。而根据图 14 的识别结果，优化后模型可以有效排除上述因多尺寸差异带来的识别干扰，并且能够以较高的置信度得到分体类别，同时预测边界

框位置误差相对较小。最终,针对附加多尺寸数据集,优化后模型的平均检测精度为 80.79%,说明优化后模型针对多尺度特征明显的羊肉分体具备良好的鲁棒性。

2.8 与其他识别算法对比试验

目前,随着卷积神经网络的发展,基于该技术的目标识别方法愈发多样,但面对不同对象与任务,不同方法的识别性能往往不同。为探究优化后羊肉多分体识别模型较其他常用识别算法的优劣性,引入 Mask R-CNN、Faster R-CNN、Cascade R-CNN、SSD 进行对比试验,另外,更换 YOLO v3 网络的特征提取网络为 MobileNet V1、ResNet34 和 ResNet50 与优化后羊肉多分体识别模型作进一步对比分析。以上算法的训练超参数与 YOLO v3 原始模型保持一致,然后利用分别得到的最优模型针对羊肉多分体验证集、附加亮度数据集、附加遮挡数据集展开测试,得到相应的 mAP 和单幅图像处理时间,用以评判各方法的识别效果。部分识别对比结果如图 15 所示。

由图 15a 可以看出,在原始、调亮、调暗和遮挡 4 种情况下,Mask R-CNN、Faster R-CNN、Cascade R-CNN 均能够实现各羊肉分体的准确识别,无漏检现象,同时预测边界框定位正确,较目标真实位置无明显误差。特别是在调暗图像中,相邻腹肋肉分体紧密贴合处于黏连状态形成整体,以上方法同样能够以较高的置信度预测得到被检分体相应的类别和边界框位置。另外,在遮挡情况下,叠落状态的肩胛骨对胸椎骨分体部分形成遮挡,上述方法依然可以准确识别被遮挡的胸椎骨分体。而 SSD 面对羊肉多分体原始、调暗和遮挡状态时,产生漏检现象,主要表现为未能准确识别出相邻的肩胛骨和腿骨,另外,在调亮状态中,SSD 方法预测得到的边界框与真实位置误差明显,腿骨端部部分没有完整标记,同时调暗情况下尺寸相对较小的肩胛骨预测类别概率只有 75%。以上现象说明,针对羊肉多分体数据集,Mask R-CNN、Faster R-CNN、Cascade R-CNN 具备较高的识别精度以及良好的泛化性和抗遮挡能力,在识别精度上能够满足实际的生产需求,而 SSD 识别能力较弱。其原因可能是基于候选区域思想设计 Mask R-CNN、Faster R-CNN、Cascade R-CNN 在图像目标检测过程中,先进行候选区域提取,区分可能存在目标的区域与背景,再对候选区域进行卷积、池化等运算进行边框回归与分类,有利于排除非目标区域的干扰,在目标密度较大时仍能保持较高的准确率。而 SSD 直接对输入图像进行处理,得到目标位置坐标值以及类别概率,再对位置进行修正,易受特征相似情况的影响,同时 SSD 的特征金字塔层特征全部来自于特征提取网络的最后一层,造成浅层包含小尺寸物体的特征信息丢失,因而对小尺寸的肩胛骨检测准确率较低。

图 15b 为更换特征提取网络为 MobileNet V1、ResNet34 和 ResNet50 后的 YOLO v3 识别模型与优化后模型对羊肉多分体图像在原始、调亮、调暗和遮挡 4 种状态下的识别结果。可以看出,4 种模型均未出现漏检和错检现象,但在识别精度上存在差异,例如,YOLO v3-MobileNet V1 和 YOLO v3-ResNet34 在调亮和调暗状态下对颈部分体、腿骨分体与肩胛骨的边界框与真实位置误差明显,而 YOLO v3-ResNet50 和优化后模型上述现象并不显著,说明 YOLO v3-ResNet50 与优化后模型在识别精度和泛化性、抗遮挡能力上均优于 YOLO v3-MobileNet V1 和 YOLO v3-ResNet34。其原因可能是 MobileNet V1 采用的 ReLU 激活函数会破坏网络提取低维空间特征的非线性,导致线性瓶颈,而 MobileNet V3 则通过 H-Swish 激活函数和注意力机制逆残差模块改善了以上缺点,从而使得采用 MobileNet V3 的 YOLO v3 模型较 MobileNet V1 对羊肉分体特征定位更准确。另外 ResNet50 较 ResNet34 更深的神经网络结构,有利于深层抽象特征的提取,从而提高模型的识别精度。

实际生产线对羊肉多分体识别方法除要求具备较高的识别精度之外,还需要较短的处理时耗,因此,判断各方法针对羊肉多分体的识别实时性同样重要。获取了 Mask R-CNN、Faster R-CNN、Cascade R-CNN、SSD、YOLO v3-MobileNet V1 等方法针对羊肉多分体图像验证集的单幅图像处理时间以及平均检测精度,如表 2 所示。

由表 2 可知,针对羊肉多分体图像验证集,Mask R-CNN 单幅图像平均处理时间最久,为 694.4 ms,其次为 Faster R-CNN(575.8 ms),Cascade R-CNN 在二阶段目标检测算法中耗时最少,为 237.5 ms,但仍比单目标检测算法中耗时最久的 SSD 多 146.7 ms,其原因可能是二阶段目标检测算法先确定目标候选框再进行回归和预测的推理过程相对直接对图像进行回归计算的单目标检测算法更为复杂,造成耗时较久的情况。单目标检测算法中,优化后模型对单幅图像平均处理时间最短,为 64.7 ms,然后是 YOLO v3-MobileNet V1(67.2 ms),另外,YOLO v3 在特征提取网络为 ResNet34 和 ResNet50 情况下单幅图像平均处理时间相差不大,可能是 ResNet50 网络深度虽然比 ResNet34 较深,但采用了 1×1 和 3×3 两种尺寸的



图 15　优化模型与其他模型和不同特征提取网络下的识别对比结果

Fig. 15　Recognition results of optimized model compared with other models and different feature extraction networks

表 2 各模型分类检测结果

Tab. 2 Classification and detection results of each model

模型	特征提取网络	平均检测精度/%	单幅图像平均处理时间/ms
Mask R - CNN	ResNet50	92.65	694.4
Faster R - CNN	ResNet50	93.78	575.8
Cascade R - CNN	ResNet50	92.96	237.5
SSD	VGG16	67.28	90.8
YOLO v3 - MobileNet V1	MobileNet V1	84.35	67.2
YOLO v3 - ResNet34	ResNet34	80.23	74.1
YOLO v3 - ResNet50	ResNet50	87.28	76.6
优化后模型	MobileNet V3	88.05	64.7

卷积核,使得其网络参数量与只采用 3×3 尺寸卷积核的 ResNet34 相差不大,从而两种模型在处理时耗上较为接近。以上说明在实时性方面,优化后模型性能最优,能够满足实际生产需求,而两阶段目标检测算法耗时较久,与羊肉多分体的实时分类检测存在差距。在检测精度方面,优化后模型的平均检测精度最高,为 88.05%,较 SSD、YOLO v3 - MobileNet V1、YOLO v3 - ResNet34、YOLO v3 -ResNet50 分别高 20.77、3.70、7.82、0.77 个百分点。对比试验表明,优化后羊肉多分体识别模型可以兼顾较高的检测精度和良好的实时性,综合检测能力最优。

3 结论

(1)基于单阶段目标检测算法的羊肉多分体识别模型以 YOLO v3 为主要网络,能够实现输送带场景图像中多个、多类羊肉分体类别的准确检测,同时边界框预测正确、完整,具备一定的多尺度检测能力,对羊肉多分体图像验证集的平均检测精度达到 87.79%,在精度上能够满足实际生产线的需求。

(2)通过更换羊肉多分体识别模型的主干网络为 MobileNet V3,可以有效提升实时性,但对检测精度无显著影响。针对羊肉多分体图像验证集,优化后模型单幅图像处理耗时为 64.7 ms,较原始模型提升 48.53%,说明本文方法在羊肉多分体实时分类检测任务中可以兼顾较高的识别精度和良好的实时性。

(3)设计的基于单阶段目标检测算法的羊肉多分体识别模型针对光照数据集、遮挡数据集、多尺度数据集平均检测精度分别达到 85.16%、83.47% 和 80.79%,对光照变化、遮挡状态实际生产过程中常见的干扰因素及叠落情形具有较强的泛化能力和抗干扰能力,对多尺度特征显著的羊肉分体具备良好的鲁棒性。

(4)与 Mask R - CNN、Faster R - CNN、Cascade R - CNN、SSD 相比,对羊肉多分体图像验证集,优化后羊肉多分体识别模型较前 3 种模型识别精度相差不大,但单幅图像平均处理时间缩短,与引入 MobileNet V1、ResNet34 和 ResNet50 3 种不同特征提取网络的识别模型比较,优化后模型的平均检测精度为 88.05%,较 SSD、YOLO v3 - MobileNet V1、YOLO v3 - ResNet34、YOLO v3 - ResNet50 分别高 20.77、3.70、7.82、0.77 个百分点,综合检测能力最优。

参 考 文 献

[1] 丁存振, 赵瑞莹. 我国肉羊屠宰加工业现状、问题及对策[J]. 肉类研究,2004,28(3):31 - 35.
DING Cunzhen, ZHAO Ruiying. Current status, issues and countermeasures of sheep slaughtering and processing industry in China[J]. Meat Research, 2014, 28(3): 31 - 35. (in Chinese)

[2] 赵世达, 王树才, 白宇,等. 基于生成对抗网络与 ICNet 的羊骨架图像实时语义分割[J]. 农业机械学报,2021,52(2):329 - 339,380.
ZHAO Shida, WANG Shucai, BAI Yu, et al. Real-time semantic segmentation of sheep skeleton image based on generative adversarial network and ICNet[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021,52(2):329 - 339, 380. (in Chinese)

[3] 孟令峰, 朱荣光, 白宗秀,等. 基于手机图像的不同贮藏时间下冷却羊肉的部位判别[J]. 食品科学,2020,41(23):21 - 26.
MENG Lingfeng, ZHU Rongguang, BAI Zongxiu, et al. Discrimination of chilled lamb from different carcass parts at different storage times based on mobile phone images[J]. Food Science, 2020,41(23):21 - 26. (in Chinese)

[4] 李振强, 王树才, 赵世达,等. 基于机器视觉和机器学习的羊骨架自动分割方法[J]. 食品与机械,2020,36(6):125 - 132.
LI Zhenqiang, WANG Shucai, ZHAO Shida, et al. The calculation methods of goat trunk's segmentation trajectory based on machine vision and machine learning[J]. Food & Machinery,2020,36(6):125 - 132. (in Chinese)

[5] 张丽娜, 杨建宁, 武佩,等. 羊只形态参数无应激测量系统与试验[J]. 农业机械学报, 2016, 47(11): 307 - 315.
ZHANG Li'na, YANG Jianning, WU Pei, et al. Design and experiment of non-stress measuring system for sheep's conformation parameters[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016,47(11):307 - 315. (in Chinese)

[6] KAMRUZZAMAN M, ELMASRY G, SUN D W, et al. Application of NIR hyperspectral imaging for discrimination of lamb muscles[J]. Journal of Food Engineering, 2011, 104(3):332 - 340.

[7] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D C, 2005: 886 - 893.

[8] BAY H, TUYTELAARS T, VAN GOOL L. SURF: speed up robust features[C]// Proc. of the 9th European Conference on

Computer Vision. Berlin,2006:404 – 417.
[9] 赵凯旋, 何东健. 基于卷积神经网络的奶牛个体身份识别方法[J]. 农业工程学报, 2015,31(5):181 – 187.
ZHAO Kaixuan, HE Dongjian. Recognition of individual dairy cattle based on convolutional neural networks[J]. Transactions of the CSAE, 2015,31(5):181 – 187. (in Chinese)
[10] 刘岩, 孙龙清, 罗冰,等. 基于改进 CNN 的多目标生猪检测算法[J]. 农业机械学报, 2019,50(增刊):283 – 289.
LIU Yan, SUN Longqing, LUO Bing, et al. Multi-target pigs detection algorithm based on improved CNN[J]. Transactions of the Chinese Society for Agricultural Machinery,2019,50(Supp.):283 – 289. (in Chinese)
[11] 叶长文, 康睿, 戚超,等. 基于 Faster – RCNN 的肉鸡击晕状态检测方法[J]. 农业机械学报, 2019,50(12):255 – 259.
YE Changwen, KANG Rui, QI Chao, et al. Stunning state recognition method of broiler chickens based on faster region convolutional neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019,50(12):255 – 259. (in Chinese)
[12] 赵世达, 王树才, 李振强,等. 基于 U 型卷积神经网络的羊肋排图像分割[J]. 食品与机械, 2020,36(9):116 – 121,154.
ZHAO Shida, WANG Shucai, LI Zhenqiang, et al. Image segmentation of sheep ribs based on U-shaped convolutional neural network[J]. Food & Machinery,2020,36(9):116 – 121,154. (in Chinese)
[13] 刘楷东, 谢斌, 翟志强,等. 基于 R2U – NET 和空洞卷积的羊后腿分割目标肌肉区识别[J]. 农业机械学报,2020,51(增刊 2):507 – 514.
LIU Kaidong, XIE Bin, ZHAI Zhiqiang, et al. Target muscle region recognition in ovine hind leg segmentation based on R2U – NET and atrous convolution algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(Supp. 2):507 – 514. (in Chinese)
[14] 李振强, 王树才, 赵世达,等. 改进 DeepLabV3 + 和 XGBoost 的羊骨架切割方法[J]. 计算机工程与应用,2021,57(18):263 – 269.
LI Zhenqiang, WANG Shucai, ZHAO Shida, et al. The cutting methods of sheep's trunk based on improved DeepLabV3 + and XGBoost[J]. Computer Engineering and Applications,2021,57(18):263 – 269. (in Chinese)
[15] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C] //Computer Vision and Pattern Recognition. IEEE, 2014:580 – 587.
[16] REN S Q, HE K M, GIRSHICK R, et al. Faster R – CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,39(6):1137 – 1149.
[17] HE K, GKIOXARI G, DOLLAR P, et al. Mask R – CNN[C] //2017 IEEE International Conference on Computer Vision. Washington D.C. IEEE, 2017:2961 – 2969.
[18] LI Y, CHEN Y, WANG N, et al. Scale-aware trident networks for object detection[C] //2019 IEEE International Conference on Computer Vision. Seoul, arxiv: 1901.01892.
[19] 罗会兰, 陈鸿坤. 基于深度学习的目标检测研究综述[J]. 电子学报, 2020,48(6):1231 – 1239.
LUO Huilan, CHEN Hongkun. Survey of object detection based on deep learning[J]. Acta Electronica Sinica, 2020,48(6): 1231 – 1239. (in Chinese)
[20] 刘俊明, 孟卫华. 基于深度学习的单阶段目标检测算法研究综述[J]. 航空兵器,2020,27(3):45 – 53.
LIU Junming, MENG Weihua. Research on single-stage object detection algorithm based on deep learning[J]. Aero Weaponry,2020,27(3):45 – 53. (in Chinese)
[21] 刘小刚, 范诚, 李加念,等. 基于卷积神经网络的草莓识别方法[J]. 农业机械学报,2020,51(2):237 – 244.
LIU Xiaogang, FAN Cheng, LI Jianian, et al. Identification method of strawberry based on convolutional neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(2):237 – 244. (in Chinese)
[22] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C] // European Conference on Computer Vision. Amsterdam,2016:21 – 37.
[23] LIN T Y,GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2017,99:2999 – 3007.
[24] ANDREW H, MARK S, GRACE C, et al. Searching for MobileNetV3[J]. arXiv:2019,1905.02244.
[25] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [J]. arXiv:2017,1704.04861.
[26] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C] // IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, arXiv: 2016, 1512.03385.
[27] 李庆旭, 王巧华, 肖仕杰,等. 基于深度学习的群体种鸭蛋受精信息检测方法[J]. 农业机械学报,2021,52(1):193 – 200.
LI Qingxu, WANG Qiaohua, XIAO Shijie, et al. Detection method for fertilizing information of group duck eggs based on deep learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021,52(1):193 – 200. (in Chinese)
[28] 崔文靓, 王玉静, 康守强,等. 基于改进 YOLOv3 算法的公路车道线检测方法[J/OL]. 自动化学报,2019. http://kns.cnki.net/kcms/detail/11.2109.TP.20190605.1051.001.html.
[29] NEUBECK A, VAN GOOL L. Efficient non-maximum suppression[C] // International Conference on Pattern Recognition, IEEE Computer Society, Hong Kong, 2006:850 – 855.