

基于改进 Mask R-CNN 的笼养死鸭识别方法

柏宗春^{1,2} 吕胤春^{1,3} 朱一星¹ 马肄恒¹ 段恩泽^{1,2}

(1. 江苏省农业科学院农业设施与装备研究所, 南京 210014;

2. 农业农村部长江中下游设施农业工程重点实验室, 南京 210014; 3. 江苏大学农业工程学院, 镇江 212013)

摘要: 针对规模化笼养肉鸭舍内死鸭识别采用人工作业方式时,存在作业效率低、劳动强度大、养殖成本高等问题,以层叠式笼养肉鸭为研究对象,提出了一种基于深度学习的笼养死鸭识别方法。为了采集数据,首先面向立体层叠式养殖环境设计了一款适用于肉鸭舍的自主巡检装备。针对笼养肉鸭舍铁丝网遮挡严重的问题,基于机器视觉对笼网进行修复,基于 OpenCV 对图像进行增强处理。构建了一种基于 Mask R-CNN 的死鸭识别模型,采用 Swin Transformer 对模型进行优化,解决了 Mask R-CNN 网络缺乏整合全局信息能力的问题。对比分析了 SOLO v2、Mask R-CNN 和 Mask R-CNN + Swin Transformer 模型识别笼内死鸭准确率。实验结果表明,在平均精度均值为 90% 的条件下,Mask R-CNN + Swin Transformer 模型对笼内死鸭总体识别准确率可达 95.8%,在自主巡检装备上的检测效果优于其他主流的目标检测算法。

关键词: 机器视觉; 笼养肉鸭; 死鸭识别; Mask R-CNN

中图分类号: S834.8

文献标识码: A

文章编号: 1000-1298(2024)07-0305-10

OSID:



Dead Duck Recognition Method Based on Improved Mask R-CNN

BAI Zongchun^{1,2} LÜ Yinchun^{1,3} ZHU Yixing¹ MA Yiheng¹ DUAN Enze^{1,2}

(1. Institute of Agricultural Facilities and Equipment, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China

2. Key Laboratory of Protected Agriculture Engineering in the Middle and Lower Reaches of Yangtze River, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China

3. College of Agricultural Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: Traditional manual methods for identifying dead ducks within large-scale stacked cage poultry houses have proven to be inefficient, labor-intensive, and costly. Focusing on stacked cage housing for meat ducks, a deep learning-based method was proposed for dead duck recognition. To collect the necessary dataset, a specialized autonomous inspection system tailored for meat duck housing within three-dimensional stacked environments was initially designed. To address the issue of severe wire mesh obstruction within the cage housing, machine vision techniques were employed to repair the cage mesh and enhance images by using OpenCV. A dead duck recognition model was constructed based on Mask R-CNN, and further optimized with the Swin Transformer to overcome the limitation of Mask R-CNN's global information integration. The accuracy of dead duck recognition among the SOLO v2, Mask R-CNN, and Mask R-CNN + Swin Transformer models was compared and analyzed. Experimental results demonstrated that under the condition of mAP value of 90%, the Mask R-CNN + Swin Transformer model achieved an overall dead duck recognition rate of 95.8% within the duck cages, outperforming other mainstream object detection algorithms on the autonomous inspection equipment.

Key words: machine vision; caged ducks; dead duck recognition; Mask R-CNN

0 引言

高效识别鸭舍中的死鸭已成为肉鸭自动化养殖

模式中亟待解决的重要问题之一^[1]。养殖场鸭苗的日均死亡率通常在 0.05% ~ 0.1%^[2],在一个存栏量为 1×10^5 只的肉鸭厂,平均每日需拣出 50 ~

收稿日期: 2023-11-06 修回日期: 2023-12-21

基金项目: 江苏省现代农业重大核心技术创新项目(CX(22)1008)

作者简介: 柏宗春(1981—),男,副研究员,博士,主要从事畜禽养殖技术与装备研究,E-mail: vipmaple@126.com

通信作者: 段恩泽(1995—),男,助理研究员,博士,主要从事畜禽养殖技术与装备研究,E-mail: 952510993@qq.com

100 只死鸭,由于肉鸭的高密集养殖,死鸭的识别仍然依赖人工操作,这导致效率低下、劳动强度大和养殖成本高昂。一方面管理人员的行为会导致肉鸭应激,死亡率提高,降低生产效率;另一方面鸭舍内死鸭对空气的污染可能会对工作人员身体造成伤害。因此,开展笼内死鸭识别模型的研究对于家禽养殖业至关重要^[3],为了采集数据需要研发一款适用于肉鸭舍的自主巡检装备。

目前机器视觉在畜禽养殖中应用广泛,显著改善了劳动力短缺的情况,不仅可以代替人类从事危害健康的工作,还能够降低生产成本和节约能源^[4]。薛鸿翔^[5]设计的死鸡自动识别系统,采用热红外采集技术,基于卷积神经网络模型识别鸡场存在的死鸡,但该方法主要用于平养肉鸡,对于笼养目标识别准确率不高;段恩泽等^[6]基于 Mask R-CNN 和 LiteFlowNet 网络构建了一种多目标背景下基于视频关键帧的死兔识别模型,但该方法检测速度有待优化;姜来等^[7]基于机器人技术、红外热成像技术及图像处理等技术设计一种基于温度判断的死鸡识别算法,但总体识别率只有 80%;赵一名等^[8]以 YOLO v5s 目标检测算法为基础,通过加入 SE 注意力模块,构成改进后的 YOLO v5s-SE 模型,检测准确率达到 97.7%,但缺乏适配的图像采集工具;此外,还有一些修复图像遮挡的创新技术,GUO 等^[9]基于深度卷积神经网络(CNN)引入 3 个跳跃连接反卷积和卷积运算,对单幅图像去除雨痕遮挡的效果达到 80% 以上;徐旻等^[10]以 YOLO v4 算法为基础,引入注意力机制,融合图像修复,对遮挡目标的检测精度达到 79.63%,但算法检测速度下降,削减了算法实时性。尽管这些技术在其他养殖领域得到了应用,但关于肉鸭识别模型的研究报道较少。

本文在立体层叠式笼养肉鸭环境下,使用自主巡航机器人自动追踪群养肉鸭,采集肉鸭图像,采用深度学习的方法实现图像的处理,并在此基础上验证死鸭识别模型准确率。

1 材料与算法

1.1 实验设计

实验于 2023 年 7—8 月在南京市六合区的一个肉鸭养殖场进行,共持续 42 d。鸭舍采用立体层叠式笼养模式,共有 3 层笼具,中间过道宽度为 0.8 m,每个单笼尺寸为 160 cm × 70 cm × 40 cm,养殖密度为 15 只/m²,并且配备自动化饮水设备以减少人对动物的干扰,如图 1 所示。

自主巡检机器人的主要任务是根据需求在鸭舍内按照规定路线行走,采集所需位置的图像。自主

巡检机器人建模图和实物如图 2 所示,总体尺寸为 64 cm × 48 cm × 210 cm。

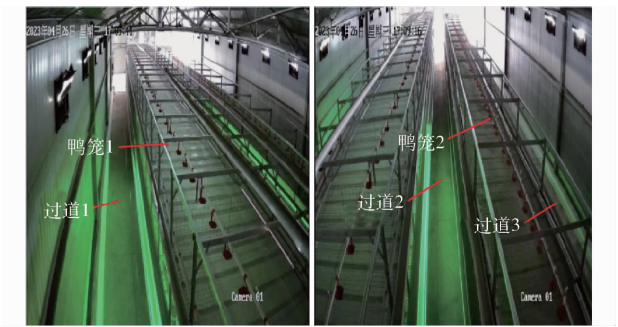


图 1 肉鸭养殖场工作环境

Fig. 1 Working environment of duck farm

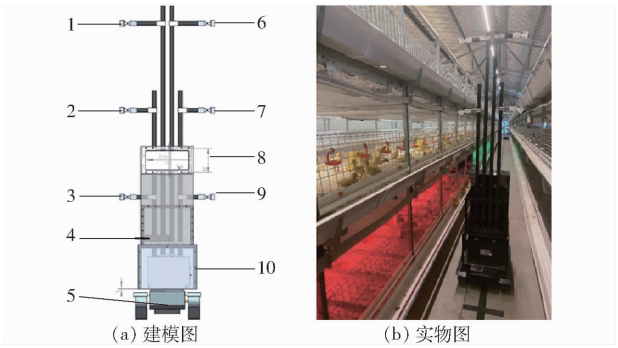


图 2 自主巡检机器人建模图及实物图

Fig. 2 Modeling and physical drawing of autonomous inspection robot

1. 相机 1 2. 相机 2 3. 相机 3 4. 主机 5. 小车底盘 6. 相机 4
7. 相机 5 8. 显示屏 9. 相机 6 10. 电源

机器人采用履带式结构以适应鸭舍环境,并降低巡检时噪声。两电机差速底盘作为驱动结构,以确保机器人可以在鸭舍内的任何地点自由转向。机器人采用磁条导航来保证巡检平台始终沿着铺设的磁条运行。

自主巡检机器人依照事先铺设的磁条轨道进行系统巡检。当机器人行驶于磁条轨道时,一旦磁条传感器探测到磁条宽度发生变化并且达到指定宽度时,上位机将发出指令,车轮停止旋转,并同时向工业相机发出图像采集指令。相机完成图像采集任务后,将反馈结果。若发现鸭笼内有死亡禽类,系统将通过通信模块将此信息传送至终端,随后继续巡检下一个鸭笼。若未检测到异常情况,机器人将持续执行巡检任务。

摄像头采用 USB 输出的 IMX588 相机模组,相机固定在机器人侧方,距离笼底面 0.8 m,倾斜 45° 俯拍,位置如图 3 所示。在图像采集过程中,确保肉鸭始终在拍摄范围内,并尽量减小镜头畸变的影响。相机通过 USB 线连接到巡检机器人内部的一台计算机上,采用 Python 算法实现图像的自动抓取。6 个摄像头依次采集,采集速率为 6 f/min,间隔 10 s。

最终,获得 1 075 幅死鸭图像以及大量正常肉鸭图像,用于构建死鸭目标识别模型数据集。

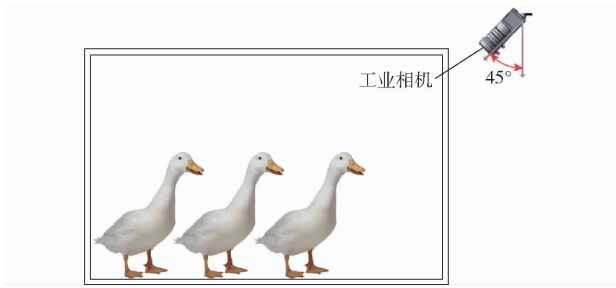


图 3 图像采集示意图
Fig.3 Image acquisition diagram

1.2 基于改进 Mask R - CNN 的掩膜分割网络

1.2.1 Mask R - CNN 网络模型

卷积神经网络 (Convolutional neural network, CNN)^[11]是一种深度学习模型,广泛应用于处理图像和视觉数据的计算机视觉任务中^[12]。Mask R - CNN 是一种广泛应用于目标检测和实例分割的深

度学习模型。它扩展了 Faster R - CNN 框架,通过添加额外的分割分支来生成目标实例的精确掩码^[13]。

Mask R - CNN 的核心思想是将目标检测和实例分割任务结合在一起,使模型能够检测目标的位置并为每个目标实例生成像素级别的掩码,从而提供更详细的目标分割^[14]。在死鸭识别和检测任务中,CNN 通过卷积和池化操作自动学习和提取图像中的特征,如边缘、纹理、形状等,从而准确地识别死鸭。Mask R - CNN 网络是一种基于深度学习的实例分割算法,不仅可以实现对图像的目标检测、目标分类和语义分类,还能对每个单独目标的像素进行分类。它是一种典型的两阶段网络算法,包括了区域提取和边界框回归及分类两个阶段^[15]。其结构包括主干网络、区域提取网络、RoIAlign (Region of interest Align) 和功能型网络,如图 4 所示。

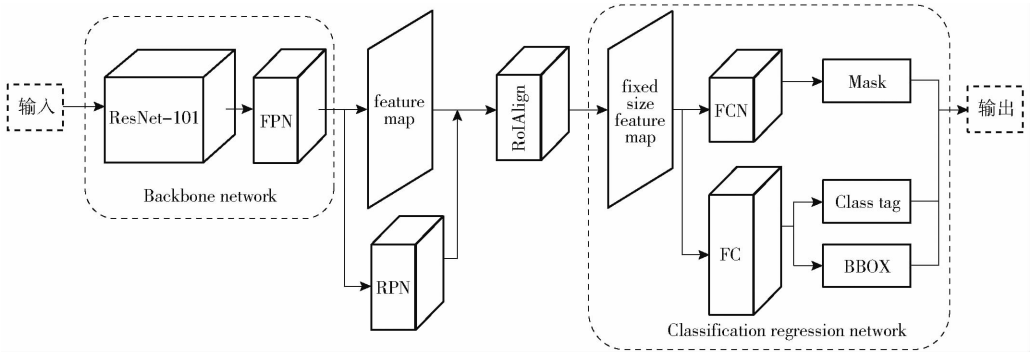


图 4 Mask R - CNN 网络结构图
Fig.4 Mask R - CNN network structure

采用 ResNet - 101 深度残差网络和特征金字塔 (Feature pyramid network, FPN) 的组合作为特征提取网络。ResNet - 101 具有较深的网络结构,通过残差连接能够有效地提取图像中的高级语义特征。而 FPN 通过构建多层特征金字塔,将不同尺度和语义层次的特征进行融合,从而更好地捕捉目标肉鸭的细节和上下文信息;区域候选网络负责生成候选目标区域 (RoIs),通过利用锚框和候选框回归来定位可能包含肉鸭的图像区域;RoI 分类网络根据 RoIs 中的特征进行分类,为每个 RoI 预测其所属的类别;掩膜生成网络在 RoIs 的基础上生成目标的像素级掩膜,实现精确分割。

1.2.2 基于 Swin Transformer 的网络改进

受到养殖策略的限制,群养肉鸭相互遮挡的现象严重。肉鸭舍光环境较暗,光照强度在 10 ~ 20 lx,光线不足使图像中群养肉鸭的个体、边界线变得模糊。上述现状使神经网络在提取图像特征时产生大量噪声干扰,降低了像素间的远程依赖关系。

Mask R - CNN 卷积神经网络能够有效提取局部信息,但缺乏整合全局信息的能力。因此,本文模型的优化策略主要考虑给模型增加自注意力机制,能够包含全局特征,一方面降低因遮挡导致的目标漏检误检概率,另一方面优化复杂环境下目标丢失的漏检问题。

Transformer 可以弥补这一缺陷,通过多头注意力机制能够包含全局特征^[16]。Swin Transformer 是一种新型的视觉 Transformer 模型,专门设计用于图像处理任务^[17]。它采用分层分组的注意力机制,将图像分割成非重叠的块,然后进行自注意力计算,以有效捕捉不同尺度的信息。Swin Transformer 在图像分类、目标检测和实例分割任务中表现出色^[18]。Swin Transformer 提出了一种层次化的 Transformer 结构,通过移位窗口方案将自注意力计算限制在不重叠的局部窗口的同时,还允许跨窗口拼接特征图^[19]。这种分层架构具有更高的计算时效性和更低的计算复杂度。

1.2.2.1 针对骨干网络的优化

将 Swin Transformer 的特征提取部分嵌入到 Mask R-CNN 模型中,以替代 Mask R-CNN 的传统卷积骨干网络,整合过程可能需要修改输入和输出以匹配 2 个模型,以及设计一个适用于多任务(目标检测和实例分割)的损失函数。

Swin Transformer 使用位置编码来引入图像中的位置信息。位置编码是一个与嵌入向量具有相同维度的矩阵,允许模型了解每个图块在原始图像中的位置。

每个 Swin Transformer 模块首先对输入特征进行标准化(Layer normalization, LN),然后计算窗口注意力(W-MSA),通过残差连接将结果加到输入上,并再次进行标准化。经过多层感知机(MLP)进行非线性变换后,得到变换后的特征表示。在后续的网络结构中,只需要将 W-MSA 替换为移位窗口注意力(SW-MSA),其他步骤保持不变。与传统 ViT 中的多头自注意(MSA)^[20]相比,MSA 缺乏跨窗口的连接,Swin Transformer 中的 W-MSA 作为一个单元控制窗口中的计算区域,减少了网络计算的数量,降低了图像的复杂度,如图 5 所示。SW-MSA

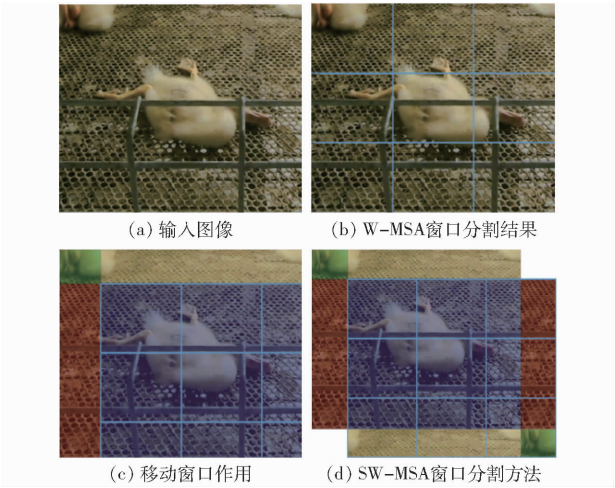


图 5 移动窗口作用流程

Fig. 5 Flow of moving window action

的位置与 W-MSA 层相连。因此,SW-MSA 需要在之后提供不同的窗口分割方法(W-MSA)实现跨窗口通信。

通过 W-MSA 对输入图像的窗口分割结果如图 5b 所示。图像的每个周期向上和向左移动一半窗口大小,然后将图 5c 中的蓝色和红色区域移至图像下方和右侧,如图 5d 所示。W-MSA 虽然降低了计算复杂度,但是不重合的窗口之间缺乏信息交流,在 2 个连续的 Swin Transformer Block 中交替使用 W-MSA 和 SW-MSA 解决了不同窗口的信息交流问题。移动窗口的划分方式使上一层相邻的不重合窗口之间引入连接,大大增加了感受野,极大提高了 Swin Transformer 的计算效率,使 2 个窗口有了交互,从而达到了全局建模的能力。

Swin Transformer 模型的特征提取过程分为 4 个阶段,采用了分层自注意力机制,信息可以在不同的层次之间流动,有助于模型从局部到全局捕获不同尺度的深层次特征。网络结构如图 6 所示。

阶段 1:包括 1 个线性编码模块和 2 个经过修改的 Swin Transformer 模块。这些 Swin Transformer 模块具备自注意力机制,用于捕捉图块之间的关系和上下文信息。

阶段 2:这个阶段引入 1 个区域合并模块,以及 2 个 Swin Transformer 模块。区域合并模块通过融合相邻区域,促进区域之间的信息交互,从而防止信息丢失。

阶段 3:继续提取特征,包括 1 个区域合并模块和 6 个 Swin Transformer 模块。这一阶段有助于捕捉更多的上下文信息和全局特征。

阶段 4:最后阶段包括 1 个区域合并模块和 2 个连续的 Swin Transformer 模块,这有助于扩大感受野,提取更深层次的特征,以获取全局信息。

通过这 4 个阶段,Swin Transformer 逐步构建了

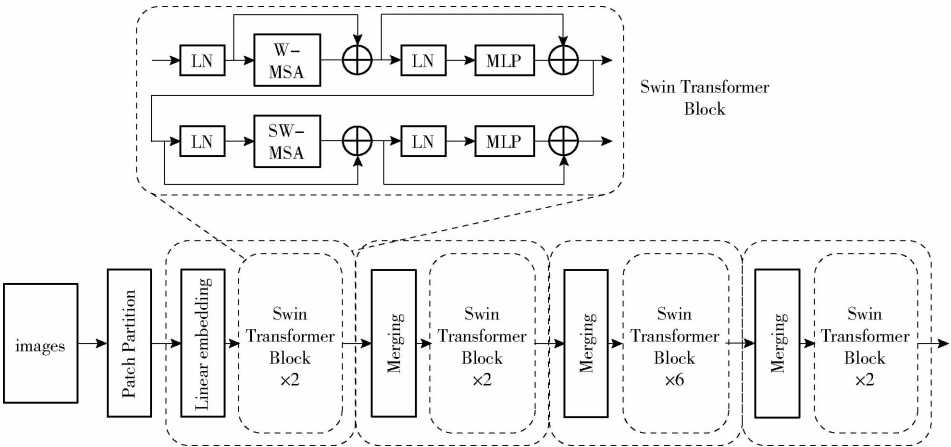


图 6 Swin Transformer 模型的特征提取流程图

Fig. 6 Feature extraction flow chart of Swin Transformer model

图像特征表示,从浅层到深层,有效地捕获了图像结构和上下文信息。

每个 Swin Transformer 模块首先对输入特征进行标准化,然后计算窗口注意力(W - MSA),通过残差连接将结果加到输入特征上,并再次进行标准化。使用移位窗口多头自注意力层(SW - MSA),通过进行窗口的偏移达成窗口之间的信息交流,增加相对位置编码,注意力的计算限制在每个窗口内,进而减少了计算量。注意力计算公式为

$$A_u(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax} \left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}} + \mathbf{B} \right) \mathbf{V} \quad (1)$$

式中 $\mathbf{Q}$ ——查询矩阵

$\mathbf{K}$ ——键矩阵 $\mathbf{V}$ ——值矩阵

$\mathbf{B}$ ——偏置矩阵,表示相对位置

d ——键矩阵维度

在每次 Swin Transformer 模块提取特征后,通过弱监督选择器对特征图进行弱监督选择,保留有效

特征参与特征融合,然后将融合后的特征输入分类器进行分类。

获取预训练的 Swin Transformer 模型权重。创建 Mask R - CNN 模型的框架,包括目标检测、分割头和损失函数,将 Swin Transformer 的特征提取部分替换为 Mask R - CNN 的传统卷积骨干网络,并且使用 Swin Transformer 的预训练权重初始化新骨干网络。

1.2.2.2 输入和输出适配优化

修改输入分辨率、通道数等确保新的骨干网络的输入和输出与 Mask R - CNN 的其他部分兼容。Swin Transformer 需要一批图像作为输入,修改 Mask R - CNN 的数据加载和处理代码以满足 Swin Transformer 的输入要求。根据 Swin Transformer 的特征输出,调整 FPN 层级,以确保与 Swin Transformer 输出兼容。基于 Swin Transformer 的改进 Mask R - CNN 网络结构如图 7 所示。

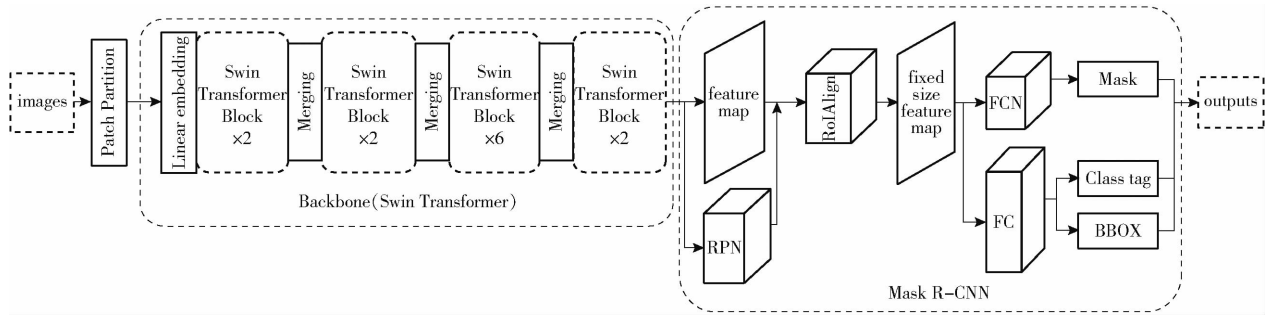


图 7 基于 Swin Transformer 的改进 Mask R - CNN 网络结构图

Fig.7 Improved Mask R - CNN network structure based on Swin Transformer

1.2.2.3 损失函数优化

本文引入交叉熵损失函数 Cross - Entropy Loss^[21]作为损失函数。该函数用于训练分类模型,通过最小化损失,模型可以更好地进行类别预测,处理肉鸭密集、肉鸭遮挡严重等复杂因素环境下的对象检测和分割任务。总损失函数为

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C (y_{ij} \lg p_{ij}) \quad (2)$$

式中 N ——样本数量

C ——类别数量

y_{ij} ——真实标签向量中的元素,表示样本 i 属于类别 j 的真实概率

p_{ij} ——模型预测概率向量中的元素,表示模型对样本 i 属于类别 j 的预测概率

1.3 图像增强与笼网修复

采集到的图像需要经过增强处理,以提升模型训练效果。首先,图像中包含周围笼具内的肉鸭图像信息,需要提取出感兴趣区域;其次,相机采样图像整体偏暗,导致背景与前景的特征差异不显著,通

过图像增强可以提高图像的可读性和训练精度。

对图像进行剪裁。原始图像尺寸为 1 920 像素 × 1 200 像素,拍摄时目标笼具位于图像正中心,因此以原图像中心点为基准,将图像剪裁至 960 像素 × 720 像素。

为了提高后续死鸭检测的可靠性,基于 OpenCV 对采集到的图像进行图像增强处理。处理过程包括增加亮度、增强对比度以及增加饱和度的批处理操作。具体结果如图 8 所示。

由于侧方俯拍的方式,部分肉鸭会被笼网遮挡,为了提高死鸭识别准确率,对遮挡图像的笼网进行修复,该算法使用计算机视觉和图像处理技术,能够自动化地进行图像修复,以获得清晰且完整的肉鸭图像。

采用 Labelme 工具标注肉鸭图像中的笼网部分并保存为 JSON 格式文件,利用 OpenCV 和 Python 构建图像处理函数,处理使用 Mask R - CNN 网络模型训练得到的遮挡笼网部分的掩膜部分,将训练得到的遮挡笼网部分 HSV 颜色空间下界和 HSV 颜色

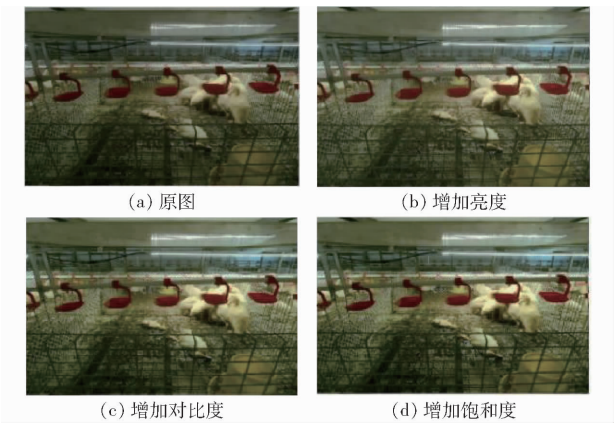


图 8 图像增强处理过程

Fig. 8 Image enhancement process

空间上界作为输入参数。

函数读取训练得到的遮挡笼网图像,并将其转换为 HSV 颜色空间^[22]辅助实现更准确地选择和处理目标颜色范围,提高遮挡笼网部分修复的准确性。

在图像处理过程中,算法首先获取图像尺寸,然后通过顺时针由外到内逐层遍历图像中的每个像素点,遍历过程中,算法通过以下步骤对每个像素点进行处理:

- (1)判断当前像素点是否位于指定的颜色区域内,即通过掩膜检查当前像素点的颜色,如果当前像素点在颜色区域内,则该像素点为待修复像素点。
- (2)以待修复像素点为圆心,5 个像素点单位为半径,创建一个小圆形模板,这个模板用于确定哪些像素点将影响待修复像素点的颜色。
- (3)将小圆形模板与颜色区域的掩膜进行交集计算,以获取小圆形模板与颜色区域的重叠区域。
- (4)从小圆形模板中减去交集区域,得到剩余区域,这个剩余区域包含在小圆形模板内,但不包括在颜色区域内的像素点。
- (5)计算剩余区域像素点颜色均值,将这一均值颜色用作修复像素点新颜色,并更新颜色掩膜。
- (6)处理完该像素点后,顺时针对下一个像素点进行上述处理。

完成对图像中所有像素点的遍历后,修复后的图像将去除遮挡笼网部分,其中,采用不同颜色的笼网掩膜修复效果不同,修复结果对比如图 9 所示。

由图 9 可以看出,绿色掩膜修复效果最佳。将该算法应用于肉鸭图像的处理,可以有效地去除遮挡笼网部分,为后续死鸭目标识别提供清晰可靠的图像基础。

1.4 死鸭识别模型

1.4.1 数据采集

使用肉鸭自主巡检机器人作为图像采集工具,采集肉鸭图像数据。图像首先经过裁剪和增强处

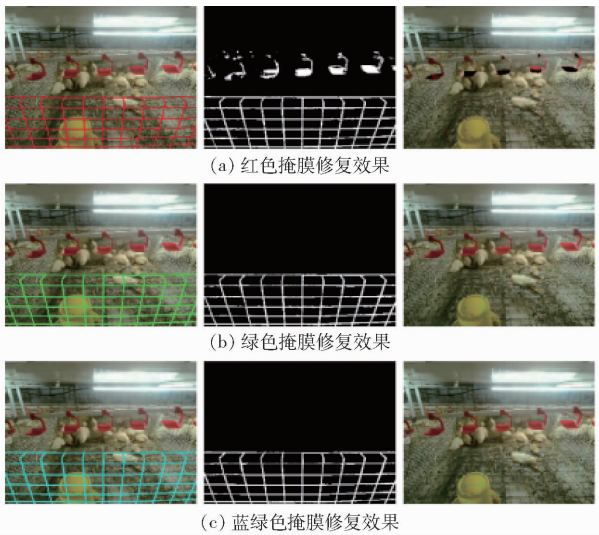


图 9 不同颜色的笼网掩膜修复过程

Fig. 9 Repair process of cage mesh mask with different colors

理,应用图像修复算法对遮挡笼网部分进行去除。随后使用 SAM - Tool 半自动标注工具快速标注死鸭目标,标签类别名称为“dead_duck”,创建包含死鸭类别信息的 JSON 文件。最终获得构建死鸭目标识别模型数据集图像共 1 075 幅,其中 915 幅为训练集,160 幅为验证集。

1.4.2 数据集制作

针对死鸭目标识别任务分别采用 Mask R - CNN、SOLO^[23]、SOLO v2^[24]、QueryInst^[25] 和 Mask R - CNN + Swin Transformer 实例分割模型进行识别训练,输入图像尺寸为 960 像素 × 720 像素,为了优化每个模型的性能以及有效完成死鸭目标识别任务,分别设置每个模型学习率、批量大小、迭代次数和优化器选择等训练参数,如表 1 所示。每个模型共训练 120 轮,每迭代 8 个轮次保存一次模型权重,最终选择识别精度最高的模型作为结果。

表 1 训练模型超参数设置

Tab. 1 Training model hyperparameter setting

模型	学习率	批量大小	迭代次数	优化器
Mask R - CNN	0. 002	16	120	SGD
SOLO	0. 002	16	120	SGD
SOLO v2	0. 001	16	120	SGD
QueryInst	0. 000 1	16	120	AdamW
Mask R - CNN + Swin Transformer	0. 000 1	16	120	AdamW

2 死鸭识别模型实验

2.1 死鸭识别评估

选择损失值和死鸭识别平均准确率 (Average precision, AP) 作为评价死鸭识别模型性能指标。在目标检测和分割任务中,评估模型性能主要考察边界框平均精度均值 (Bounding box mean average

precision, bbox_mAP) 和分割均值平均精度 (Segmentation mean average precision, segm_mAP) 两性能指标。bbox_mAP 衡量模型在检测和定位目标方面的准确性和精度, segm_mAP 衡量模型在分割目标方面的准确性和精度。

由图 10 可知, 经过 2×10^4 次迭代后, SOLO v2 和 Mask R - CNN 模型损失值分别达到 0.097 和 0.115, Mask R - CNN + Swin Transformer 和 SOLO 模型损失值则分别为 0.156 和 0.206。

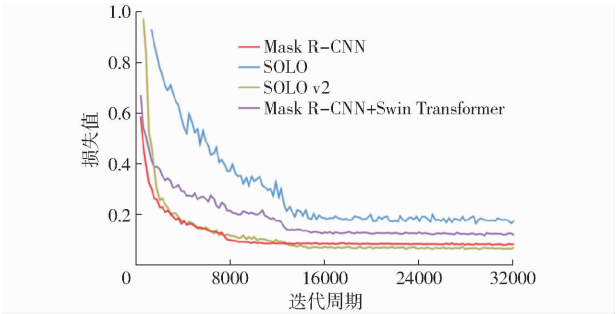


图 10 模型训练损失值曲线

Fig. 10 Loss value curves of model training loss function

由图 11a 可知, 训练经过 75 个迭代周期后, Mask R - CNN + Swin Transformer、Mask R - CNN 和 QueryInst 模型的 bbox_mAP 分别达到 90.6%、86.6% 和 89.2%。由图 11b 可知, 训练 90 个迭代周期后, Mask R - CNN + Swin Transformer 和 SOLO v2 模型的 segm_mAP 值最高, 达到 90.1% 和 88.9%, 其次是 Mask R - CNN 和 QueryInst 模型分别为 86.8% 和 86.4%, SOLO 模型 segm_mAP 值最低为 83.3%。

为了评估 Mask R - CNN、SOLO、SOLO v2、QueryInst 和 Mask R - CNN + Swin Transformer 在死鸭目标识别任务中的性能, 设计了死鸭目标识别准确率试验, 分别使用死鸭目标识别误检率、漏检率和准确率作为评价指标。

(1) 死鸭识别误检率 (False positive rate, FPR) 是指模型错误识别正常肉鸭为死鸭样本数量与所有正常肉鸭样本数量之比, 衡量模型在正常肉鸭中的误检率。

(2) 死鸭识别漏检率 (False negative rate, FNR) 是指模型未能正确识别死鸭的样本数量与所有死鸭样本数量之比, 衡量模型在死鸭样本中的漏检率。

(3) 死鸭识别平均准确率 (Average precision, AP)^[26] 是指模型正确识别为死鸭的样本数量与模型共识别为死鸭的样本数量之比, 衡量模型在所有死鸭预测中的准确性。

实验使用 120 幅测试集图像来测试模型的死鸭识别正确率, 并设置不同的置信度阈值的平均精度

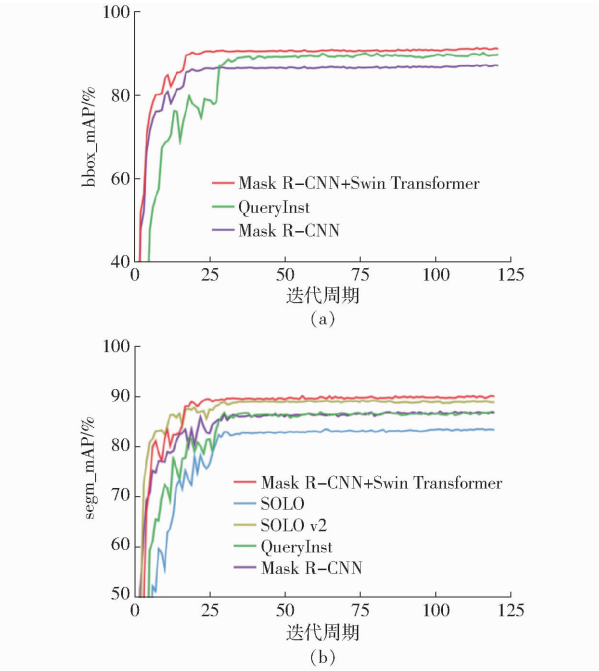


图 11 模型训练阶段 bbox_mAP 曲线和 segm_mAP 曲线

Fig. 11 Model training stage bbox_mAP curves and segm_mAP curves

均值 (Mean average precision, mAP), 分别为 30%、60% 和 90%, 通过在不同的 mAP 下进行测试, 能够评估模型在不同置信度下的表现, 模型的 mAP 越高, 表示它的高置信度条件下的性能越好。通过实验选择适合鸭舍环境的 mAP。

由表 2 可知, SOLO 和 QueryInst 网络模型随着 mAP 的提高, 死鸭漏检率提升, SOLO 模型在置信度阈值为 90% 时, 死鸭识别准确率仅为 83.3%, QueryInst 模型为 91.7%。如 mAP 设置过低, 模型

表 2 模型不同置信度死鸭识别准确率
Tab.2 Dead duck recognition accuracy with different confidence levels

模型	mAP/%	死鸭识别	死鸭识别	死鸭识别
		误检数	漏检数	准确率/%
Mask R - CNN	30	5	1	93.3
	60	2	1	97.5
	90	1	2	97.5
SOLO	30	8	6	88.3
	60	4	7	90.8
	90	0	20	83.3
SOLO v2	30	2	0	98.3
	60	2	1	97.5
	90	1	5	95.0
QueryInst	30	2	2	96.7
	60	2	4	95.0
	90	0	10	91.7
Mask R - CNN + Swin Transformer	30	1	1	98.3
	60	1	2	97.5
	90	1	4	95.8

可能会更容易检测出死鸭,但这可能会导致误检率上升,以 SOLO 模型为例,当 mAP 为 30% 时,误检率达到 6.7%。SOLO v2、Mask R - CNN + Swin Transformer 和 Mask R - CNN 模型在死鸭识别任务中表现较好,在 mAP 调整为 90% 时,死鸭识别准确率分别为 95.0%、95.8% 和 97.5%。

2.2 笼网去除前后死鸭检测效果对比

为评估在死鸭图像中去掉笼网前后,对死鸭检测效果的影响,分别对去掉笼网前后的死鸭图像进行标注,制作了数据集,用于训练 Mask R - CNN + Swin Transformer 算法模型。得到 2 个权重文件,并在验证集上进行了检测。将 mAP 调整为 90%,选择死鸭目标识别误检率、漏检率和准确率作为评价指标。

根据表 3 可以看出,使用去掉笼网后的图像作为训练集对死鸭模型检测效果有正面的提升。在使用去掉笼网后的图像训练模型验证时,相较于模型在训练集和验证集都未去除笼网的情况下,提升效果显著,达到了 95.8%,相比于 89.2%,提升了 6.4 个百分点。然而,当训练集使用未去除笼网的图像,而验证集使用去除笼网的图像时,死鸭检测效果却出现了下降,死鸭识别误检数增加。这种情况可能是由于光线不足,死鸭与笼网之间的过拟合导致的。

表 3 笼网去除前后死鸭识别准确率

Tab.3 Dead duck identification accuracy before and after cage net removal				
训练集是否 去除笼网	验证集是否 去除笼网	死鸭识别 误检数	死鸭识别 漏检数	死鸭识别 准确率/%
否	否	4	8	90.0
	是	6	7	89.2
是	否	2	6	93.3
	是	1	4	95.8

2.3 不同生长阶段死鸭检测效果对比

本实验为评估自主巡检装备在肉鸭不同生长阶段对死鸭进行识别的性能。在肉鸭养殖实验期间,每日 08:00 对肉鸭舍进行巡检,在 0~14 日龄段共发现死亡肉鸭 7 只,15~35 日龄段共发现死亡肉鸭 1 只,由于肉鸭出现死亡的情况主要发生在 0~14 日龄段,因此针对 0~7 日龄和 8~14 日龄的死鸭进行巡检测试。每当在肉鸭笼内发现死亡肉鸭时,使用巡检装备对死鸭对象进行时长 1 h 的视频录制,以捕捉到更多的死鸭特征和姿态信息。在 0~7 日龄段,共有 4 只死亡肉鸭,分布在 3 号笼、6 号笼、7 号笼和 8 号笼中,选择其中 3 只分别位于 6 号笼、7 号笼和 8 号笼的死鸭作为实验对象,采用截取视频帧的方式从每只死鸭中获取 20 幅图像用于测试。同样,在 8~14 日龄段,共有 3 只死亡肉鸭,分布在 4

号笼、5 号笼和 7 号笼中,采用相同的视频帧截取方法,从每只鸭笼中各获取 20 幅测试图像。使用 SOLO v2、Mask R - CNN 和 Mask R - CNN + Swin Transformer 模型对 0~7 日龄和 8~15 日龄出现的死鸭分别进行识别实验,采用死鸭识别平均准确率作为评价指标,实验结果如表 4 所示。

表 4 0~7 日龄和 8~15 日龄死鸭识别准确率

Tab.4 Identification accuracy of dead ducks aged 0~7 days and 8~15 days							%
模型	0~7 日龄			8~15 日龄			
	6 号笼	7 号笼	8 号笼	4 号笼	5 号笼	7 号笼	
Mask R - CNN	95	100	100	100	95	100	
SOLO v2	100	90	100	95	90	100	
Mask R - CNN + Swin Transformer	100	100	100	100	100	100	

由表 4 可知,Mask R - CNN 模型和 SOLO v2 模型在在识别 0~7 日龄和 8~15 日龄死鸭时均出现漏检或错检的情况,Mask R - CNN 模型出现将健康肉鸭的脚错检成死鸭的问题,SOLO v2 模型出现预测框检测错误的问题,Mask R - CNN + Swin Transformer 模型对 0~7 日龄和 8~15 日龄的死鸭识别准确率均为 100%,无漏检与错检现象。因此应用 Mask R - CNN + Swin Transformer 模型在自主巡检装备上检测效果最优。

3 讨论

从 Mask R - CNN、SOLO、SOLO v2、QueryInst 和 Mask R - CNN + Swin Transformer 模型训练结果的损失值曲线可以看出,Mask R - CNN 与 QueryInst 的训练结果略优于 Mask R - CNN + Swin Transformer 模型。但从模型训练的 bbox_mAP 曲线以及 segm_mAP 曲线中可以看出,Mask R - CNN + Swin Transformer 明显优于其他模型。置信度阈值设置为 90% 进行训练,其中 SOLO v2、Mask R - CNN + Swin Transformer 和 Mask R - CNN 模型在死鸭识别任务中表现较好,死鸭识别准确率分别为 95.0%、95.8% 和 97.5%。但随后将其部署在死鸭自主巡检装备上,Mask R - CNN 模型出现将健康肉鸭的脚错检成死鸭的问题,SOLO v2 模型出现预测框检测错误的问题,Mask R - CNN + Swin Transformer 模型能够实现对 0~7 日龄和 8~15 日龄的死鸭准确识别,无漏检与错检现象。出现这种现象的原因可能是机器学习中的过拟合现象导致的,过拟合发生时,模型在训练数据中学到了数据中的噪声和随机变化,而不能很好地泛化到新的、未见过的数据,这导致模型对于训练数据表现出高度适应性,但对于测

试数据的泛化性能较差。因此 Mask R - CNN + Swin Transformer 模型在实际应用中效果最优,且自注意力机制包含全局特征,增加了整合全局信息的能力,一方面降低遮挡导致的目标漏检误检概率,另一方面优化复杂环境下目标丢失的漏检问题。

4 结束语

构建了一种高度适应性的死鸭识别模型,通过

基于深度学习的遮挡笼网去除,基于 OpenCV 的图像增强处理以及 Swin Transformer 与 Mask R - CNN 算法模型的融合应用,以及基于 Swin Transformer 的增加自注意力机制和移动窗口的划分方式,改进 Mask R - CNN 模型,并开展了死鸭识别评估实验,得出 Mask R - CNN + Swin Transformer 模型在实际应用中效果最优的结论,为养殖场优化养殖工艺提供了技术和理论支撑。

参 考 文 献

[1] 王明宇,王勇生.我国肉鸭养殖业发展面临的挑战[J].当代畜牧,2021(8):71-73.

[2] 祖全亮,刘宏,杨景晔.当前肉鸭产业发展形势与建议对策[J].家禽科学,2022(4):41-43.

[3] 李明阳,应诗家,戴子淳,等.新型肉鸭养殖模式生产性能及经济效益对比分析[J].中国家禽,2020,42(4):80-85.

[4] 胡惠玥,杨小玲,刘仁鑫,等.畜禽养殖机器人研究现状与展望[J].南方农机,2023,54(19):1-6,10.

[5] 薛鸿翔.基于红外热成像技术的死鸡识别系统设计与实现[D].南京:南京农业大学,2022.

XUE Hongxiang. Design and implementation of dead broiler identification system based on infrared thermal imaging technology [D]. Nanjing: Nanjing Agricultural University, 2022. (in Chinese)

[6] 段恩泽,王粮局,雷逸群,等.基于实例分割和光流计算的死兔识别模型研究[J].农业机械学报,2022,53(2):256-264,273.

DUAN Enze, WANG Liangju, LEI Yiqun, et al. Dead rabbit recognition model based on instance segmentation and optical flow computing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(2): 256-264, 273. (in Chinese)

[7] 姜来,王文娣,霍晓静,等.死鸡识别机器人系统设计与试验[J].中国农机化学报,2023,44(8):81-87.

JIANG Lai, WANG Wendi, HUO Xiaojing, et al. Design and experiment of dead chicken recognition robot system[J]. Journal of Chinese Agricultural Mechanization, 2023, 44(8): 81-87. (in Chinese)

[8] 赵一名,沈明霞,刘龙申,等.基于改进 YOLO v5s 和图像融合的笼养鸡死鸡检测方法研究[J].南京农业大学学报,2024,47(2):369-382.

ZHAO Yiming, SHEN Mingxia, LIU Longshen, et al. Study on the method of detecting dead chickens in caged chicken based on improved YOLO v5s and image fusion[J]. Journal of Nanjing Agricultural University, 2024, 47(2): 369-382. (in Chinese)

[9] GUO Y, LI C, LIU Q. R2N: a novel deep learning architecture for rain removal from single image[J]. Comput. Mater. Contin., 2019, 58(3): 829-843.

[10] 徐旸,史金光,郑子琦,等.融合图像修复的遮挡目标检测算法[J].电光与控制,2023,30(1):21-28,86.

XU Yang, SHI Jinguang, ZHENG Ziyu, et al. Occlusion target detection algorithm with image inpainting[J]. Electronics Optics & Control, 2023, 30(1): 21-28, 86. (in Chinese)

[11] 孙志军,薛磊,许阳明,等.深度学习研究综述[J].计算机应用研究,2012,29(8):2806-2810.

SUN Zhijun, XUE Lei, XU Yangming, et al. Overview of deep learning[J]. Application Research of Computers, 2012, 29(8): 2806-2810. (in Chinese)

[12] 周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(6):1229-1251.

ZHOU Feiyan, JIN Linpeng, DONG Jun. Review of convolutional neural network[J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251. (in Chinese)

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

[14] 李旭冬,叶茂,李涛.基于卷积神经网络的目标检测研究综述[J].计算机应用研究,2017,34(10):2881-2886,2891.

LI Xudong, YE Mao, LI Tao. Review of object detection based on convolutional neural networks[J]. Application Research of Computers, 2017, 34(10): 2881-2886, 2891. (in Chinese)

[15] 李彦冬,郝宗波,雷航.卷积神经网络研究综述[J].计算机应用,2016,36(9):2508-2515,2565.

LI Yandong, HAO Zongbo, LEI Hang. Survey of convolutional neural network[J]. Journal of Computer Applications, 2016, 36(9): 2508-2515, 2565. (in Chinese)

[16] PARMAR N, VASWANI A, USZKOREIT J, et al. Image transformer [C] // International Conference on Machine Learning. PMLR, 2018: 4055-4064.

[17] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.

[18] LIANG J, CAO J, SUN G, et al. Swinir: image restoration using swin transformer [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 1833-1844.

[19] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks [C] // International Conference on Machine Learning. PMLR, 2019: 7354-7363.

[20] BODENHOFER U, BONATESTA E, HOREJS-KAINRATH C, et al. MSA: an R package for multiple sequence alignment [J]. Bioinformatics, 2015, 31(24): 3997 – 3999.

[21] ZHANG Z, SABUNCU M. Generalized cross entropy loss for training deep neural networks with noisy labels[J]. Advances in Neural Information Processing Systems, 2018,31:8792 – 8802.

[22] 梁建安,刘斌,梁美彦,等. 基于彩色偏振图像的 HSV 空间目标增强方法[J]. 应用光学,2023,44(3):548 – 555.
LIANG Jian'an, LIU Bin, LIANG Meiyang, et al. Target enhancement method in HSV color space based on color polarized image[J]. Journal of Applied Optics,2023,44(3):548 – 555. (in Chinese)

[23] WANG X, KONG T, SHEN C, et al. Solo: segmenting objects by locations [C] // Computer Vision-ECCV 2020: 16th European Conference, 2020: 649 – 665.

[24] WANG X, ZHANG R, KONG T, et al. Solov2: dynamic and fast instance segmentation[J]. Advances in Neural Information Processing Systems, 2020, 33: 17721 – 17732.

[25] FANG Y, YANG S, WANG X, et al. Instances as queries [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 6910 – 6919.

[26] YUE Y, FINLEY T, RADLINSKI F, et al. A support vector method for optimizing average precision [C] // Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2007: 271 – 278.

(上接第 304 页)

[20] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.

[21] 何东宇,朱荣光,范彬彬,等. 倒置残差网络结合注意力机制的掺假羊肉分类检测系统构建[J]. 农业工程学报,2022, 38(20):266 – 275.
HE Dongyu, ZHU Rongguang, FAN Binbin, et al. Construction of a classification detection system for adulterated mutton by inverted residual network combined with attention mechanism [J]. Transactions of the CSAE,2022,38(20):266 – 275. (in Chinese)

[22] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [J]. IEEE Computer Society, 2017, 106(1): 936 – 944.

[23] 马钦,万传峰,卫建,等. 基于 U – Net 和特征金字塔网络的秸秆覆盖率计算方法[J]. 农业机械学报,2023,54(1):224 – 234.
MA Qin, WAN Chuanfeng, WEI Jian, et al. Calculation method of straw coverage based on U – Net network and feature pyramid network [J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(1):224 – 234. (in Chinese)

[24] GUO C, FAN B, ZHANG Q, et al. AugFPN: improving multi-scale feature learning for object detection [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.

[25] KEMAL O, BARIS C, EMRE A, et al. Rank & sort loss for object detection and instance segmentation [C] // ICCV 2021 Oral,2021.

[26] 孟浩,刘强. 基于 FPGA 的卷积神经网络训练加速器设计[J]. 南京大学学报(自然科学版), 2021, 57(6):1075 – 1082.
MENG Hao, LIU Qiang. Design of FPGA-based training gas pedal for convolutional neural networks [J]. Journal of Nanjing University (Natural Science), 2021, 57(6):1075 – 1082. (in Chinese)

[27] XU Z H, LI J X, DENG S Y, et al. Optimization of real-time detection and classification model for plant leaf diseases based on TensorRT [J]. Computer Systems and Applications, 2023, 32(2): 94 – 101.