

李昊玥, 陈桂芬, 裴傲 基于改进 Mask R-CNN 的奶牛个体识别方法研究 [J]. 华南农业大学学报, 2020, 41(6): 161-168.
LI Haoyue, CHEN Guifen, PEI Ao. Research on individual recognition of dairy cows based on improved Mask R-CNN[J]. Journal of South China Agricultural University, 2020, 41(6): 161-168.

基于改进 Mask R-CNN 的奶牛个体识别方法研究

李昊玥, 陈桂芬, 裴 傲

(吉林农业大学 信息技术学院, 吉林 长春 130118)

摘要:【目的】针对传统奶牛养殖中采用人工识别奶牛个体的方法效率低且主观性强的问题, 提出一种基于改进 Mask R-CNN 的奶牛个体识别方法。【方法】该方法对 Mask R-CNN 中的特征提取网络结构进行优化, 采用嵌入 SE block 的 ResNet-50 网络作为 Backbone, 通过加权策略对图像通道进行筛选以提高特征利用率; 针对实例分割时目标边缘定位不准确的问题, 引入 IoU boundary loss 构建新的 Mask 损失函数, 以提高边界检测的精度; 对 3 000 张奶牛图像进行训练、验证和测试。【结果】改进 Mask R-CNN 模型的精度均值 (AP) 达 100%, IoU_{Mask} 达 91.34%; 与原始 Mask R-CNN 模型相比, AP 提高了 3.28%, IoU_{Mask} 提高了 5.92%。【结论】本文所提方法具备良好的目标检测能力, 可为复杂农场环境下的奶牛个体精准识别提供参考。

关键词: Mask R-CNN; SENet; IoU boundary loss; 图像识别; 奶牛识别

中图分类号: TP391.41

文献标志码: A

文章编号: 1001-411X(2020)06-0161-08

Research on individual recognition of dairy cows based on improved Mask R-CNN

LI Haoyue, CHEN Guifen, PEI Ao

(College of Information Technology, Jilin Agricultural University, Changchun 130118, China)

Abstract: 【Objective】To propose an individual cow recognition method based on the improved Mask R-CNN algorithm, and solve the problem of low efficiency and strong subjectivity of artificially identifying individual cows in traditional dairy farming. 【Method】This method optimizes the feature extraction network structure in Mask R-CNN, adopts ResNet-50 network embedded in SE block as backbone, and selects image channels by weighting strategy to improve feature utilization. For the problem of inaccurate target edge positioning during instance segmentation, a boundary weighted loss function is added to construct a new Mask loss function to improve the accuracy of boundary detection. A total of 3 000 cow images are trained, validated and tested. 【Result】The improved Mask R-CNN model had an average precision (AP) of 100% and IoU_{Mask} of 91.34%. Compared with the original Mask R-CNN model, AP increased by 3.28% and IoU_{Mask} increased by 5.92%. 【Conclusion】The proposed method has strong segmentation accuracy and robustness, and can provide a reference for accurate recognition of cow images under complex farming environment.

Key words: Mask R-CNN; SENet; IoU boundary loss; image recognition; cow recognition

收稿日期: 2020-03-29 网络首发时间: 2020-10-19 10:13:48

网络首发地址: <https://kns.cnki.net/kcms/detail/44.1110.S.20201017.1423.004.html>

作者简介: 李昊玥 (1993—), 女, 硕士研究生, E-mail: 824630206@qq.com; 通信作者: 陈桂芬 (1956—), 女, 教授, 博士, E-mail: guifchen@163.com

基金项目: 国家自然科学基金 (U19A2061)

奶牛养殖业是我国畜牧业的一个重要组成部分,近年来得到了快速发展,正逐步向规模化、智能化与精准化的方向转变^[1-2]。奶牛个体快速而准确的识别作为奶牛智能化和数字化养殖的关键,其识别方法的研究已成为当下奶牛养殖体系中的热点^[3-4]。传统的奶牛个体识别主要通过人工观测的方式进行,该方法人力成本高、主观性强且获取的信息不能及时、准确地给予反馈,不符合现代化奶牛养殖的发展需求^[5]。除人工识别方法以外,目前应用较多的个体识别方法是无线射频检测技术(Radio frequency identification, RFID),该技术将RFID标签固定在奶牛耳部,在阅读器可识别范围内即可读取标签内的奶牛个体信息,避免了人工观察标签编号的过程,但耳标的佩戴依然会对奶牛的安全造成影响而且存在标签易破损、技术成本高以及兼容性差的问题^[6]。

近年来,随着计算机技术的不断发展,基于计算机视觉的个体识别技术得到了广泛的应用。图像识别的样本通过牛场内的摄像头进行采集,所需成本低,也避免了电子耳标数量多、耳标易破损的问题^[7]。传统的图像识别技术常采用方向梯度直方图、支持向量机、和局部二值模式(Local binary pattern, LBP)等方法。如陈娟娟等^[8]使用优化方向梯度直方图(Histogram of gradient, HOG)特征对图像进行特征提取,并利用空间金字塔匹配原理(Spatial pyramid matching, SPM)和自定义直方图交叉核对15类奶牛的7500张奶牛头部图像进行分类,准确率要比单独使用HOG特征高21.3%。张满国等^[9]提出了一种基于改进方向梯度直方图与局部二值模式(HOG-LBP)特征融合的特征提取算法,并将其应用到奶牛个体识别,为实现奶牛个体识别自动化打下了基础。以上这些方法虽然在奶牛个体识别领域取得了较好的试验结果,但都对图像的拍摄角度和拍摄位置敏感,适用性较差,且图像背景复

杂或奶牛姿态不一时识别精度不高,难以实现自动识别。卷积神经网络(Convolutional neural network, CNN)^[10]的出现为图像识别提供了新思路,它具有强大的特征学习和特征表达能力,可以用于大批量图像数据的识别和检测,有良好的抗干扰能力,适用于农场复杂环境下的奶牛个体识别,更符合现代化农场信息管理的需求^[11]。2015年赵凯旋等^[12]利用卷积神经网络对奶牛视频进行分析、提取并跟踪奶牛运动过程中的躯干图像,实现了奶牛个体的精准识别。2019年王毅恒等^[13]使用YOLO v3算法快速、准确地实现了农场环境下奶牛图像复杂背景中的个体识别。

基于卷积神经网络的识别方法无需与奶牛直接接触,该方法可降低劳动成本、减小对奶牛的伤害,具有良好的应用价值。由此,本文采用卷积神经网络方法以实现奶牛个体的自动识别,并针对上述目标检测算法仅能生成奶牛图像中带有目标个体的选框而无法精确标出奶牛具体轮廓特征的问题选用Mask R-CNN算法,同时进行相应的改进,提出了一种基于改进Mask R-CNN的奶牛图像识别方法。此外,本文结合迁移学习的方式首先在COCO数据集上对提出的新算法网络的权重进行预训练,然后在实地拍摄的奶牛图像数据集上进行了仿真试验和结果评估。

1 材料与方法

1.1 数据采集与预处理

试验数据来源于吉林省某农场的实地拍摄,通过数码相机等设备采集了546张农场复杂背景下的奶牛图片,分辨率为800×600。对其中不清晰的图像加以剔除,最终选取500张作为数据集,所选样本数据包含不同光照强度、背景和拍摄角度下不同奶牛数量和奶牛姿态的图像,部分奶牛样本数据如图1所示。



图1 奶牛样本图像

Fig. 1 Cow sample image

神经网络模型的精度对数据量有着极高的要求,500张奶牛图像不足以支撑网络的训练。因此,

本文选用水平翻转、垂直翻转、顺时针旋转45°、逆时针旋转45°和灰度化5种方法对样本数据进行扩

展, 如图 2 所示。增强后的奶牛图像共 3 000 张, 按照 3:1:1 的比例划分训练集、验证集和测试集, 得到

训练图像 1 800 张, 验证图像 600 张, 测试图像 600 张。



a: 原始图像; b: 水平翻转; c: 垂直翻转; d: 顺时针旋转 45°; e: 逆时针旋转 45°; f: 灰度化

a: Original picture; b: Flip horizontal; c: Flip vertical; d: Clockwise rotation of 45°; e: Counterclockwise rotation of 45°; f: Gray picture

图 2 增强后的奶牛图像
Fig. 2 Enhanced cow image

网络的训练以监督学习的方式进行, 本文采用 Labelme 软件对增强后的数据进行标注, 标注时按照图像中奶牛的轮廓创建目标区域, 并打上标签。由于本文识别的只有奶牛 1 个类别, 因此统一用奶牛的英文 cow 作为标签名。

1.2 改进 Mask R-CNN

1.2.1 Mask R-CNN 原理 Mask R-CNN 是由 He 等^[14]于 2017 年提出的一种实例分割框架, 该方法在 Faster R-CNN^[15]的基础上增加了 Mask 分支, 在目标检测的同时对目标进行像素级的分割与分类; 并用 ROI align 替代 Faster R-CNN 中的 ROI pooling, 解决了区域的不匹配问题 (Misalignment), 具有较高的精度和速率。Mask R-CNN 一经提出就

在各个领域得到广泛的应用。

Mask R-CNN 的整体框架如图 3 所示。主要由 Backbone、RPN(Region proposal network, RPN)、ROI align 和 Classifier 共 4 个部分组成。Backbone 用于提取输入图片的特征图 (Feature map), 该特征图作为后续 RPN 和全连接层的输入; RPN 用于生成候选区域 (Region proposal), 并对每个特征区域进行类别可能性判断和框回归 (Bounding box regression) 操作; ROI align 通过收集输入的特征图和候选区域提取 Proposal feature maps, 并作为后续全连接层的输入, 进行目标类别判定; Classifier 利用 Proposal feature maps 计算候选区域的类别, 同时再次进行框回归以精确定位检测框, 并为目标生

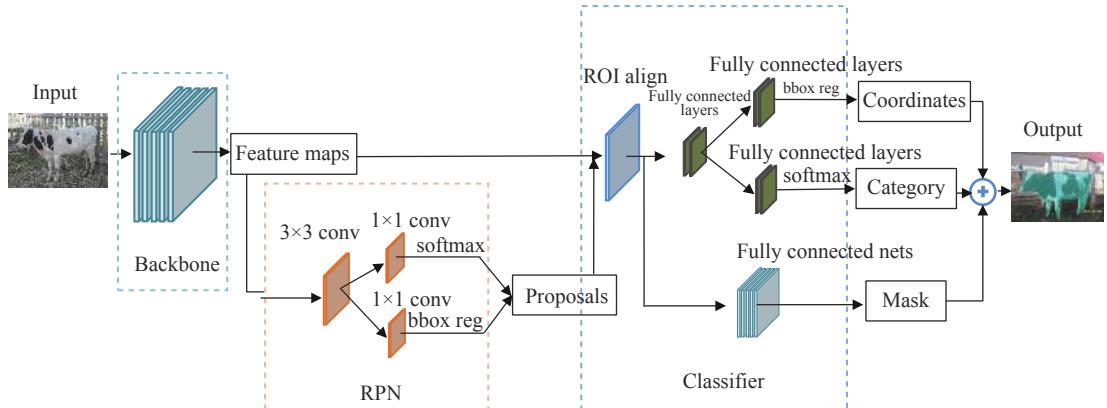


图 3 Mask R-CNN 网络结构图
Fig. 3 Network structure diagram of Mask R-CNN

成掩码。

1.2.2 SENet SENet(Squeeze-and-excitation network) 是 Hu 等^[16]于 2017 年提出的新型网络结构, 采用了一种全新的“特征重标定”策略对每个通道的重要程度进行评估。本文借鉴其做法, 将嵌有 SE block 的 ResNet^[17]作为 Mask R-CNN 模型的特征提取网络基本结构, 对奶牛图像的通道信息加以利用, 根据 Loss 值对特征的权重进行学习和重调。模型采用重要特征信息的权重大、其他特征信息的权重小的方式进行训练, 具体步骤如下:

1) 挤压 (Squeeze) 操作: 对所有通道的特征值进行全局平均池化 (Global average pooling) 操作, 将每个二维特征通道变成 1 个实数, 即 z_c , 表征在特征通道上响应的全局分布, 以获得全局感受野。 z_c 的计算公式如下:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j), \quad (1)$$

式中, H 、 W 分别表示为图片的高和宽, u 表示输入图像经过标准卷积操作后得到的特征图, u_c 表示 u 中的第 c 个二维矩阵, c 表示特征通道数, $u_c(i, j)$ 代表通道特征图矩阵 u_c 中第 i 行第 j 列的元素。

2) Excitation 操作: 通过捕捉通道之间的非线性相互作用关系对各个通道进行权重评估, 从而在通道维度上对原始特征重定标。Excitation 操作所得结果的计算公式如下:

$$s = \sigma(W_2 \delta(W_1 z)), \quad (2)$$

式中, s 表示各通道权重 s_c 的集合, δ 表示 ReLU 函数, σ 表示 Sigmoid 激活函数。 z 表示各通道经过 Squeeze 操作得到的实数 z_c 的集合, W_1 和 W_2 这 2 个权重通过学习得到, W_1 和 W_2 的维度分别是 $C \times C$ 和 $C \times \frac{C}{r}$, C 为通道数目, r 为缩放参数, 用以降低计算量, 为达到传播速度和检测准确率的平衡, 参考文献 [16], 本文中的 r 取值为 16。

3) Reweight 操作: 将 Excitation 输出的权重通过乘法逐通道加权到先前的特征上, 完成通道维度上的原始特征重标定, 从而增强对关键通道域的注意力。输入特征图 u 结合权重 s 进行重定标后的输出 \tilde{x}_c 的计算公式如下:

$$\tilde{x}_c = s_c \cdot u_c, \quad (3)$$

式中, s_c 表示各通道对应权重, u_c 为每个特征图对应的通道特征矩阵。

1.2.3 模型优化与构建 本文训练的目标仅奶牛一种, 类别相对单一, 因此对网络层数的要求较低,

为提高算法的运行效率, 选用网络较为简单的 ResNet-50 作为 Mask R-CNN 中 Backbone 网络的基础结构, 整体结构如图 4 所示。ResNet-50 虽然引入了残差学习, 具备更好的特征提取能力, 但其仍未对图像信息加以充分的利用, 仅关注了图像的空间信息而忽略了通道之间的信息, 并且其卷积层的滤波器是作用在局部视野上的, 特征图之间相互独立且重要性相同, 然而实际训练过程中不同的特征对于模型的影响并不相同。基于此, 本文在 ResNet-50 的 Shortcut 中嵌入 SE block 进行优化, 优化后的 Identity block 结构如图 5 所示。

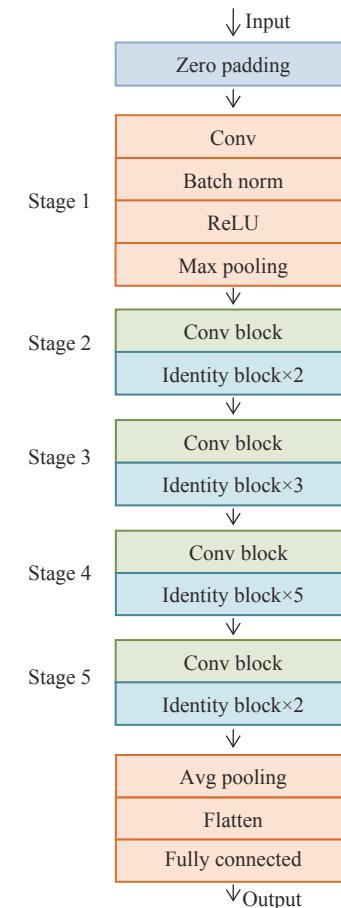


图 4 ResNet 结构

Fig. 4 Structure of ResNet

由图 5 可看出, 优化后的 Identity block 主体部分未变, 依旧先使用 Conv 层对 Input 进行 1×1 卷积操作, 将通道数减少 $1/4$, 然后对其进行 3×3 卷积操作进行特征提取, 再通过 1×1 的卷积层, 恢复通道数, 使得 Input 与 Outout 的维度相同。其中, 1×1 的卷积操作减少了卷积的参数个数与计算量, 同时改善了网络的非线性程度。在 Identity block 的 Shortcut 部分 Input 不再与主体部分得到的特征信息直接相加, 而是先经由 Global average pooling 层

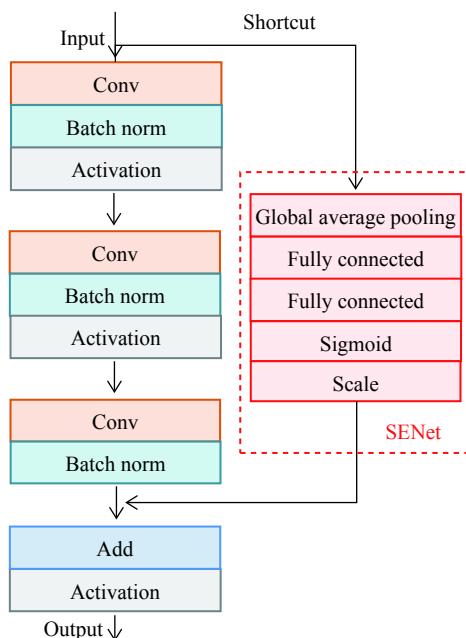


图 5 Identity block 结构

Fig. 5 Structure of identity block

进行 Squeeze 操作, 然后通过 2 个 Fully connected 层构建通道间的相关性, 并输出各个通道的权重, 接着在 Sigmoid 层对权重进行归一化, 获得 0~1 之间的权重值, 最后采用 Scale 操作将归一化权重加权到相对应的通道特征上。ResNet-50 中除了 Identity block 还有 Conv block, 优化后的 Conv block 结构同 Identity block 相似, 但由于其 Input 与 Output 的维度不一致, 因此在 Shortcut 中先加入 Conv 和 Batch norm 操作对 Input 进行降维, 使其与 Output 维度相同后再进行后续操作。

1.2.4 损失函数优化

Mask R-CNN 的损失函数 (L) 由 3 个部分组成, 分别为分类误差 (L_{cls})、检测误差 (L_{box}) 和分割误差 (L_{mask}), 其公式如下:

$$L = L_{cls} + L_{box} + L_{mask}. \quad (4)$$

L_{cls} 使用 Softmax 损失函数计算目标的分类概率, 计算公式如下:

$$L_{cls} = \text{Softmax} = -\log \left(\frac{e^{f_{vi}}}{\sum_j e^{f_j}} \right), \quad (5)$$

式中, f 表示评分向量, v_i 表示样本 i 的标签, f_j 表示分类评分向量 f 中第 j 个元素。

L_{box} 使用 Smooth L_1 函数计算边框损失, 计算公式如下:

$$L_{box} = \text{Smooth}_{L_1} = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & x < -1 \text{ or } x > 1 \end{cases}, \quad (6)$$

式中, x 为输入值。

L_{mask} 使用二值交叉熵 (Binary crossentropy) 损失函数:

$$L_{mask} = -\sum_y y \log(1 - \hat{y}) + (1 - y) \log(1 - \hat{y}), \quad (7)$$

式中, y 为期望输出, \hat{y} 为实际输出。

在人工对奶牛图像进行像素级标注时, 往往只需关注奶牛的边缘部分, 轮廓内部分复制边缘的标注信息即可, 因此边缘信息对于实例分割十分重要, 它们很好地表征了实例。原始 Mask R-CNN 在预测时忽略了边界信息, 试验中会出现边缘检测不准确的现象, 从而影响 Mask 掩码的准确度。针对这一问题, 本文借鉴文献 [18] 的做法, 在 L_{mask} 中增加 IoU boundary loss, 即 $L_{boundary}$ 对掩码部分的损失函数进行优化。首先提取 Mask 掩码的边界像素, 降低非边界像素对损失函数的影响, 然后计算真实 Mask 边界与预测 Mask 边界的重合情况。 $L_{boundary}$ 的公式如下:

$$L_{boundary} = 1 - \frac{2 |C_j \cap \hat{C}_j|}{|C_j| + |\hat{C}_j|}, \quad (8)$$

式中, $|C_j|$ 表示真实 Mask 边界像素强度之和, $|\hat{C}_j|$ 表示预测 Mask 边界像素强度之和。优化后 Mask 部分的损失函数 ($L_{mask-boundary}$) 如下:

$$L_{mask-boundary} = L_{mask} + L_{boundary} = 1 - \frac{2 |C_j \cap \hat{C}_j|}{|C_j| + |\hat{C}_j|} - \sum_y y \log(1 - \hat{y}) + (1 - y) \log(1 - \hat{y}). \quad (9)$$

1.3 模型训练

Mask R-CNN 模型的训练在 Tensorflow^[19] 深度学习框架下进行, 采用 GPU 进行加速。首先利用 Mask R-CNN 模型在 COCO 数据集上的预训练权重初始化网络参数, 然后利用自制奶牛数据集对奶牛识别模型进行训练, 模型的整个训练分为 2 个阶段:

1) 冻结 Backbone 网络, 对未使用 COCO 预训练权重的随机初始化层 (除 Backbone 外的其他所有层) 进行训练;

2) 对整个 Mask R-CNN 模型进行训练。

模型训练时依据本文的识别需求, 类别设置为 2(包含 Cow 和 Background 2 个类别), Anchor 大小设置为 32、64、128、256、512, α 权重系数设置为 1, 权重衰减系数设置为 0.000 5, 每个尺寸的 Anchor 对应生成 3 种比例 (0.5、1.0、2.0) 的锚框, 每幅图片

的 Anchor 个数为 256。此外, 模型采用非线性修正线性单元 (Rectified linear unit, ReLU) 作为激活函数, 并基于随机梯度下降法 (Stochastic gradient descend) 进行优化, 动量因子为 0.9, 网络权重参数的初始学习率为 0.001, 训练共设置 200 个 Epoch, 全部训练数据完成 1 次运算为 1 个 Epoch, 当 Epoch 达到 100 时学习率减少为 0.000 1。

2 结果与分析

2.1 评价指标

为了客观、全面地评价网络模型, 本文采用分类置信度中的精度均值 (Average precision, AP) 对模型的整体性能进行评估。AP 是预测单个目标类别的平均查准率, 由精确率 (Precision) 和召回率 (Recall) 计算得出, 以精确率为纵坐标, 召回率为横坐标, 绘制 P-R(Precision-recall) 曲线, AP 即为 P-R 曲线的积分。Precision 是指在所有被标记为奶牛的物体中, 奶牛正样本所占的比率; Recall 是测试集中所有奶牛正样本中, 被正确识别为奶牛的比例。Precision 和 Recall 的计算方式如下:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{\text{TP}}{N}, \quad (10)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (11)$$

式中, TP 表示奶牛被正确分类和定位的正样本个数, FP 表示被标记为奶牛的负样本个数, N 为样本总数, FN 为测试样本中真实的正样本数减去 TP。此外, AP 在进行计算时需要设定奶牛的预测边界框与真实边界框的交并比 (Intersection over union, IoU)^[20], 本文将其设定为 0.7, 当 IoU>0.7 时将测试样本认定为正样本, 否则为负样本。

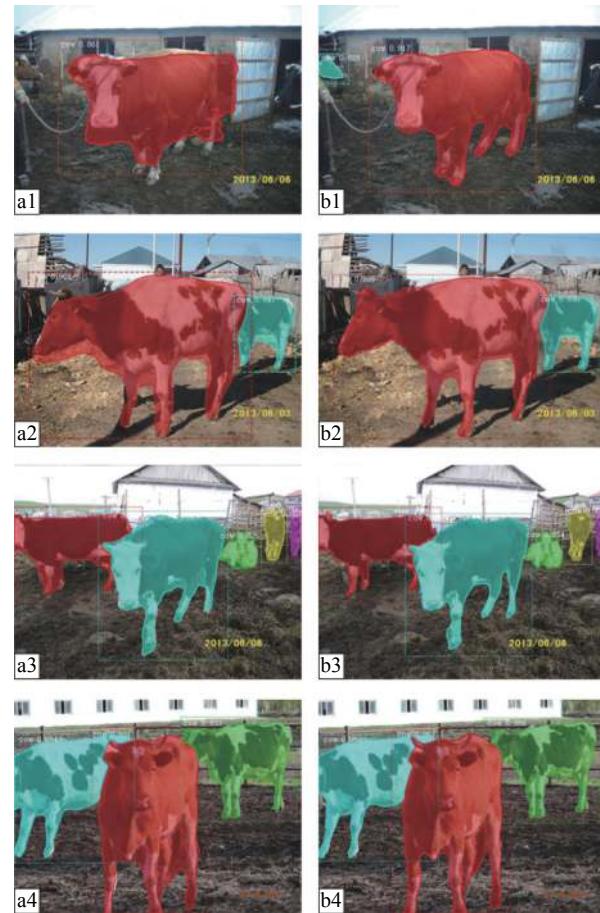
在大多数的实例分割任务中 AP 可以很好地表征模型的性能, 然而 AP 仅适用于分类置信度的评价, 无法对 Mask 的实际质量进行评估。针对这一问题, 本文借鉴文献 [21] 中的做法, 选用 IoU 值对 Mask 掩码的质量进行评估。本文中的 IoU_{Mask} 通过计算奶牛预测 Mask 区域 (C) 与其对应的人工标记 Mask 区域 (G) 之间的交集与并集之比以量化 Mask 的质量, 从而进一步衡量奶牛个体 Mask 的检测定位的精确程度, 计算方式如公式 (12) 所示:

$$\text{IoU}_{\text{Mask}} = \frac{\text{area}(C) \cap \text{area}(G)}{\text{area}(C) \cup \text{area}(G)}. \quad (12)$$

2.2 对比试验结果分析

为验证本文模型在奶牛个体识别方面的有效性, 将改进模型与原始 Mask R-CNN 模型在前文所

述的奶牛测试集上进行对比试验, 改进模型在原始 Mask R-CNN 网络的基础上同时优化了特征提取网络结构和损失函数, 训练过程中参数设置和原始 Mask R-CNN 网络保持一致, 得到的实际检测结果如图 6 所示。



a1~a4 为原始模型, b1~b4 为改进模型

a1~a4 are the original models, b1~b4 are the improved models

图 6 改进前后 Mask-R-CNN 模型试验结果对比

Fig. 6 Comparison of experimental results before and after improvement of Mask R-CNN model

由图 6 可看出, 改进前后的模型均能较为准确地识别目标的类别, 但改进后网络的 AP 值更高, 对于奶牛边缘的检测更为接近实际轮廓, 图像边界与细节更为平滑, 能有效识别图片中的小目标个体, 不同奶牛个体之间的分割也更为准确。表 1 展示了 Mask R-CNN 模型改进前后在测试集上的客观评价结果, 从中可看出改进模型在奶牛图像识别的

表 1 改进前后 Mask R-CNN 模型各项评价指标对比

Table 1 Comparison of evaluation indexes for Mask R-CNN model before and after improvement

模型 Model	AP/%	IoU _{Mask} /%
原始模型 Original model	96.72	85.42
改进模型 Improved model	100	91.34

任务上的 AP 为 100%, IoU_{Mask} 达 91.34%, 相较于原始 Mask R-CNN 模型有了较为明显的提升, 涨幅分别为 3.28% 和 5.92%。

综合主观的视觉效果与客观的定量评价, 本文方法的识别效果要优于原始的 Mask R-CNN 模型, 证明了本文方法的可行性和有效性。

2.3 模型性能影响因素分析

为呈现各个改进部分对模型性能所产生的影响, 本节在训练参数和数据集相同的情况下, 分别将仅优化网络结构的模型以及仅增加 IoU boundary loss 的模型与原始的 Mask R-CNN 网络模型进行了对比试验。

2.3.1 网络结构优化对模型性能的影响 通过在 ResNet 的 Shortcut 中增加 SE block 对模型的特征提取网络结构进行优化, 以此增强模型的特征学习能力。在相同条件设置下, 将仅优化特征网络结构的模型与原始 Mask R-CNN 模型进行对比试验, 结果如表 2 所示。

表 2 结构优化前后 Mask R-CNN 模型评价指标对比

Table 2 Comparison of evaluation indexes for Mask R-CNN model before and after structural optimization

模型	AP/%	$\text{IoU}_{\text{Mask}}/\%$
原始模型 Original model	96.72	85.42
结构优化后的模型 Model after structural optimization	100	87.51

表 2 显示结构优化后的模型在 AP 和 IoU_{Mask} 上均有提高, 分别提升了 3.28% 和 2.09%, 证明加入通道注意力机制可以有效提高奶牛个体识别的准确率。

2.3.2 IoU boundary loss 对模型性能的影响 为验证 IoU boundary loss 对模型的影响, 从其对 Mask loss 的影响和对 Mask 掩码质量的影响两方面对 IoU boundary loss 进行评估。

1) 对 Mask loss 的影响:

模型的训练过程实质上就是寻找最小 Loss 值的过程, 本节不考虑网络结构优化, 仅采用 IoU boundary loss 对 Mask R-CNN 的损失函数进行优化, 通过分析训练过程中 Mask 部分的损失函数值的变化趋势, 可以得到 IoU boundary loss 对 Mask 损失函数值的影响。改进前后的 Mask loss 的变化曲线如图 7 所示。

从图 7 可以看出, 加入 IoU boundary loss 的模型迭代至第 5 个 Epoch 后 Mask loss 的曲线趋于平

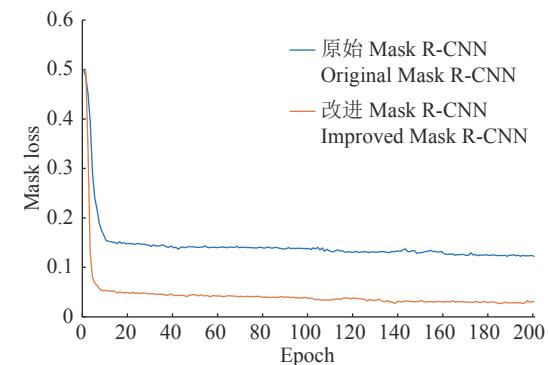


图 7 IoU boundary loss 对 Mask loss 的影响

Fig. 7 The influence of IoU boundary loss on Mask loss

稳, 迭代至第 140 个 Epoch 时曲线趋于收敛, 而原始模型在第 10 个 Epoch 后曲线才趋于平缓, 迭代至第 160 个 Epoch 时趋于收敛。相比之下, IoU boundary loss 的模型的 Mask loss 收敛速度更快, 得到的 Loss 值更低, 模型的识别效果更好。

2) 对 Mask 掩码质量的影响:

本文采用 IoU_{Mask} 指标对模型改进前后生成的 Mask 掩码质量进行定量分析, 得到的结果如表 3 所示。

表 3 损失函数优化前后 Mask R-CNN 模型 IoU_{Mask} 对比

Table 3 Comparison of IoU_{Mask} for Mask R-CNN model before and after loss function optimization

模型 Model	$\text{IoU}_{\text{Mask}}/\%$
原始模型 Original model	85.42
加入 IoU boundary loss 后的模型 Model added with IoU boundary loss	87.84

从表 3 中可以看出加入 IoU boundary loss 后的模型 IoU_{Mask} 为 87.84%, 相较于原始 Mask R-CNN 模型提高了 2.42%, 结果表明: 在 Mask loss 中加入 IoU boundary loss 可以提高边界的识别精度, 进而提升 Mask 掩码的质量, 有效地优化了训练模型。

综合上述试验的结果可以看出, 单独对网络结构进行优化和单独增加 IoU boundary loss 都可以提升模型的性能, 但提升效果有限, 而本文最终采用的改进模型同时使用了优化的网络结构和 IoU boundary loss, 融合后的改进模型在 AP 值和 IoU_{Mask} 值上均有显著的提高(表 1), 模型性能良好, 具有较强的奶牛个体识别能力。

3 结论

为解决因传统奶牛个体识别方法的局限性以及复杂的奶牛体姿态和环境导致对奶牛个体识别的有效性、准确率较低的问题, 本文提出了一种改

进 Mask R-CNN 的方法并构建了基于改进 Mask R-CNN 的奶牛个体识别模型。在奶牛测试集上进行对比试验，并对模型性能的影响因素加以分析，得到如下结论：

1) 使用嵌入 SE block 的 ResNet 作为 Mask R-CNN 的特征提取网络的基本结构对图像通道进行加权处理，可以增强网络的特征提取能力；加入 IoU boundary loss，可以提高模型对识别目标边缘信息的学习能力。

2) 改进的 Mask R-CNN 模型在奶牛图像集上的 AP 达 100%，IoU_{Mask} 达 91.34%。奶牛个体的类别能被准确识别，Mask 掩码位置也较为准确，与实际边缘更为贴合。与原始 Mask R-CNN 相比，在训练和测试条件相同的情况下，改进 Mask R-CNN 在测试集上的 AP 与 IoU_{Mask} 均超过原始 Mask R-CNN，具有良好的有效性与准确率。

参考文献：

- [1] 郑国生, 施正香, 滕光辉. 基于不同行为时间的奶牛健康状况评价[J]. *农业工程学报*, 2019, 35(19): 238-244.
- [2] 何东健, 刘冬, 赵凯旋. 精准畜牧业中动物信息感知与行为检测研究进展[J]. *农业机械学报*, 2016, 47(5): 231-244.
- [3] 刘杰鑫, 姜波, 何东健, 等. 基于高斯混合模型与 CNN 的奶牛个体识别方法研究[J]. *计算机应用与软件*, 2018, 35(10): 159-164.
- [4] 张满圃, 单新媛, 于洋, 等. 基于小波变换和改进 KPCA 的奶牛个体识别研究[J]. *浙江农业学报*, 2017, 29(12): 2000-2008.
- [5] 刘忠超, 翟天嵩, 何东健. 精准养殖中奶牛个体信息监测研究现状及进展[J]. *黑龙江畜牧兽医*, 2019(13): 30-33.
- [6] 汪开英, 赵晓洋, 何勇. 畜禽行为及生理信息的无损监测技术研究进展[J]. *农业工程学报*, 2017, 33(20): 197-209.
- [7] 孙雨坤, 王玉洁, 霍鹏举, 等. 奶牛个体识别方法及其应用研究进展[J]. *中国农业大学学报*, 2019, 24(12): 62-70.
- [8] 陈娟娟, 刘财兴, 高月芳, 等. 基于改进特征袋模型的奶牛识别算法[J]. *计算机应用*, 2016, 36(8): 2346-2351.
- [9] 张满圃, 米娜, 于洋, 等. 基于特征融合的奶牛个体识别[J]. *江苏农业科学*, 2018, 46(24): 278-281.
- [10] 杨阿庆, 薛月菊, 黄华盛, 等. 基于全卷积网络的哺乳母猪图像分割[J]. *农业工程学报*, 2017, 33(23): 219-225.
- [11] 蔡骋, 宋肖肖, 何进荣. 基于计算机视觉的牛脸轮廓提取算法及实现[J]. *农业工程学报*, 2017, 33(11): 171-177.
- [12] 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.
- [13] 王毅恒, 许德章. 基于 YOLOv3 算法的农场环境下奶牛目标识别[J]. *广东石油化工学院学报*, 2019, 29(4): 31-35.
- [14] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. *IEEE T Pattern Anal*, 2017(99): 1.
- [15] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE T Pattern Anal*, 2017, 39(6): 1137-1149.
- [16] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. *IEEE T Pattern Anal*, 2020, 42(8): 2011-2023.
- [17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. IEEE. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas: IEEE Computer Society, 2016.
- [18] LUO Z, MISHRA A, ACHKAR A, et al. Non-local deep features for salient object detection[C]. IEEE. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Hawaii: IEEE Computer Society, 2017.
- [19] LIU F, LIU P Y, LI B, et al. Deep learning model design of video target tracking based on TensorFlow platform[J]. *Laser & Optoelectronics Progress*, 2017, 54(9): 091501.
- [20] SUN X D, WU P C, HOI S C H. Face detection using deep learning: An improved faster RCNN approach[J]. *Neurocomputing*, 2018, 299: 42-50.
- [21] HUANG Z J, HUANG L C, GONG Y C, et al. Mask scoring R-CNN[C]. IEEE. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Los Angeles: IEEE Computer Society, 2019.

【责任编辑 庄 延】