

# 基于Mask R-CNN的轻量化羊只计数研究

杨雯茜, 李婷, 李美安\*

内蒙古农业大学计算机与工程学院, 内蒙古 呼和浩特

收稿日期: 2024年8月18日; 录用日期: 2024年11月19日; 发布日期: 2024年11月28日

## 摘要

为了提高羊只计数的准确性和实用性, 本文结合计算机视觉技术, 提出了一种基于Mask R-CNN轻量级羊只计数算法。针对数据集的制作, 前往内蒙古呼和浩特白塔村的养殖户进行数据采集, 制作了羊只图像分割数据集。在对模型的轻量化部分, 首先, 将特征提取网络的部分替换为 Inverted Residual模块并加入SE注意力机制, 在保证模型分割准确度不下降的情况下降低模型的规模。其次, 使用空间卷积池化金字塔ASPP进一步对模型的特征融合部分进行优化, 最后利用改进后Mask R-CNN生成的掩膜进行计数。结果表明: 改进后的Mask R-CNN-InvertedResidual-SE-ASPP羊只计数模型, 计数准确率达到96.27%, 较基准模型参数量减少38.46%, 计算量减小26.14%, 体积减小34.52%, 单帧推理速度提升22.12%。说明, 改进后的Mask R-CNN更适合实际应用中的高效羊只计数。

## 关键词

Mask R-CNN, SE注意力机制, ASPP, 羊只技术

# Lightweight Sheep Counting Study Based on Mask R-CNN

Wenqian Yang, Ting Li, Mei'an Li\*

College of Computer and Engineering, Inner Mongolia Agricultural University, Hohhot Inner Mongolia

Received: Aug. 18<sup>th</sup>, 2024; accepted: Nov. 19<sup>th</sup>, 2024; published: Nov. 28<sup>th</sup>, 2024

## Abstract

To enhance the accuracy and practicality of sheep counting, this paper proposes a lightweight sheep counting algorithm based on Mask R-CNN combined with computer vision technology. For data set creation, we collected sheep images from local farms in Baita Village, Hohhot, Inner Mongolia, to create a sheep image segmentation dataset. To lighten the model, we replaced parts of the feature

\*通讯作者 Email: limeian1973@126.com

文章引用: 杨雯茜, 李婷, 李美安. 基于 Mask R-CNN 的轻量化羊只计数研究[J]. 人工智能与机器人研究, 2024, 13(4): 920-929. DOI: 10.12677/airr.2024.134094

extraction network with Inverted Residual modules and incorporated Squeeze-and-Excitation (SE) attention mechanisms to maintain segmentation accuracy while reducing the model's size. Additionally, we optimized the model's feature fusion part using the Atrous Spatial Pyramid Pooling (ASPP) structure. Finally, we used the masks generated by the improved Mask R-CNN for counting. The results show that the improved Mask R-CNN-Inverted Residual-SE-ASPP sheep counting model achieved an accuracy rate of 96.27%, reduced the number of parameters by 38.46%, decreased computational complexity by 26.14%, reduced the model size by 34.52%, and increased single-frame inference speed by 22.12%. This indicates that the improved Mask R-CNN is more suitable for efficient sheep counting in practical applications.

## Keywords

Mask R-CNN, SE Attention Mechanism, ASPP, Sheep Counting

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

羊在我国经济中占据重要地位,尤其是在肉类、羊绒和羊奶的生产中。内蒙古作为主要的养羊区域,2022年羊肉总产量的21%来自该地区[1]。随着集中养殖模式的推广以及政策目标的推动,到2025年,规模养殖的比重预计将达到50%。在这一背景下,准确的羊只计数技术变得尤为重要,它不仅有助于防止羊只丢失,还能精准计算采食量和监测母羊产仔情况,从而提升经济效益和养殖管理水平。

基于计算机视觉的计数方法主要分为传统机器学习和深度学习两类。随着技术的进步,深度学习逐渐成为主流,其方法包括目标检测、多目标跟踪和图像分割[2]。基于目标检测的计数方法中,王毅恒等人[3]使用YOLOv3对奶牛进行计数,准确率达到100%,但在目标数量多或有遮挡的情况下效果较差。宁远霖等人[4]通过改进YOLOv5,引入注意力机制和优化器,提高了猪只检测性能,但在密集和严重遮挡场景下存在漏检问题。这些方法通过矩形框标记目标位置并进行分类,适用于实时性和位置信息要求较高且目标分布稀疏的场景。

在基于多目标跟踪的技术中,Zhang等人[5]提出了一种结合YOLO检测和DeepSORT跟踪的系统,通过全局运动补偿模块将计数准确率提高到93%。该方法对目标检测要求较高,不准确的检测会影响跟踪和计数精度。Chen等人[6]则在YOLOv7中引入GAM注意力机制和Alpha-IoU损失函数,结合DeepSORT,取得了95.1%的mAP。但Alpha-IoU的泛化能力较差,影响了对其他目标的准确性。Zheng等人[7]提出的YOLO-BYTE方法结合YOLOv7和ByteTrack,计数准确率为91.6%,但在处理密集目标时性能下降,并且泛化能力较差。这些方法在动态计数场景中表现良好,但需要在准确度和算力需求之间进行平衡。

在基于图像分割的计数方法中,Xu等人[8]使用Mask R-CNN和四轴飞行器捕获的图像,计数准确率达到96%,对遮挡和光照不足具有较强鲁棒性,但硬件成本较高。胡云鸽等人基于Mask R-CNN提出的生猪计数网络,通过改进预处理和特征提取,取得了86%的准确率,但对小目标和高密度目标效果较差。Liu等人[9]提出的ICNet模型具有高计算速度和低资源占用,但计数精度仅为71.4%,适用于对速度要求高的场景。HUANG等人[10]提出的两阶段中心聚类网络,平均绝对误差为0.43,但每幅图像处理时间为4.3秒,不适合需要快速计数的场景。基于图像分割的计数方法可以详细获取目标的位置、轮廓和大

小等信息，对密集目标进行较准确的计数，适用于养殖场等密集场景。然而，由于其较高的参数量，模型的训练和推理速度通常较慢，限制了其在实际应用中的普及。因此，研究更轻量化的图像分割计数方法以提高实用性显得尤为重要。

## 2. 材料与方法

### 2.1. 数据集的建构

为了应对羊只图像分割数据集较为稀缺的现状，构建了一个专用的羊只图像分割数据集。选择位于呼和浩特市白塔村的中小型养殖场作为数据采集点，拍摄了四段斜俯视角的视频，作为研究的基础数据。首先，对采集的羊只视频数据进行帧采样，本文设置的抽帧的时间间隔设置为每 5 秒一帧，总共抽取了 400 张待标注的图片。本文使用 Labelme 标注工具，首先要对每个待标注的对象用打点的方式，来选取对象的轮廓，在确定好一个对象的外轮廓之后，对每个标好的对象进行类别标记。用于区分对象是不是一类物体，并且将类别名保存在 JSON 文件中。为了能更好地模拟真实场景与增强模型的泛化能力，本文对原始数据进行了数据增强的操作，数据由原来的 400 张变为 3600 张。数据集的部分示例如图 1 所示。



Figure 1. Example of a sheep image segmentation dataset

图 1. 羊只图像分割数据集示例

### 2.2. 基准网络选择

为了选择一种算法作为羊只计数方法改进的基础模型，设计了对比实验，并使用公共数据集 VOC2012 对三个经典的图像分割模型进行了训练和评估，以分割平均精度均值(mAP)作为评价标准。实

验结果如表 1 所示。

**Table 1.** Deep learning image segmentation experiment comparison table  
**表 1.** 深度学习图像分割实验对比表

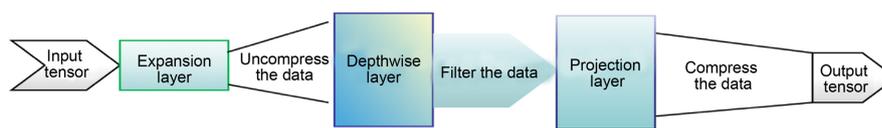
算法	mAP (%)
FCN	41.62
Unet	40.13
Mask R-CNN	47.37

从实验结果分析可得，Mask R-CNN 的 mAP 最高，为 47.37。在养殖场景中的羊只图像分割计数任务中，需要准确的分割结果，并且考虑目标的定位和检测。因此，选择使用 Mask R-CNN 作为基础模型。

## 2.3. 基准网络轻量化

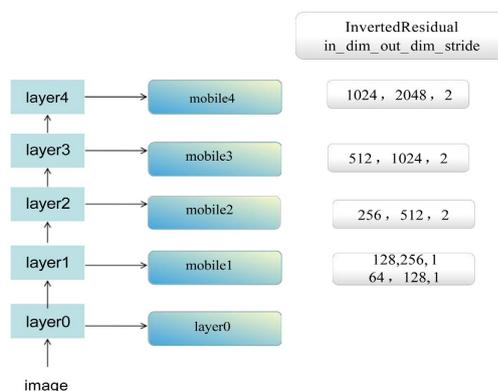
### 2.3.1. 主干网络轻量化

在选择 MaskR-CNN 作为基础模型后，需要对其主干网络进行轻量化改进。本文选取 Inverted Residual 对 Mask R-CNN 的主干网络进行轻量化改进。倒残差结构 Inverted Residual 最早由 MobileNetV2 引入，是一种在轻量级卷积神经网络中常用的模块，它的作用是提高模型的效率和性能。倒残差采用的是先升维再降维的反向操作，其结构如图 2 所示。



**Figure 2.** Inverse residual network structure diagram  
**图 2.** 倒残差网络结构图

本文在替换过程中，保留了 ResNet50 中 layer0 部分，由于这部分负责对输入图像做预处理，本文希望新的网络结构中保留相同功能，所以没有进行替换。其余四个模块均按照图 3 中的结构进行了替换。其中，in-dim 为输入维度，out-dim 为输出维度，stride 表示步幅，当 stride = 1 时卷积核每次移动一个像素，输出特征图的空间尺寸就与输入相同，当 stride = 2 时进行下采样，虽然减小了特征图的空间分辨率，但增加了感受野。通过这样的改进，可以得到一个轻量化网络结构如图 3 所示。



**Figure 3.** Schematic diagram of the introduction position of the backward residual module  
**图 3.** 倒残差模块引入位置示意图

### 2.3.2. SE 注意机制

考虑到主干网络参数量减少而造成准确度下降的情况，本文在主干网络中引入 SE 注意力机制来缓解这一情况。由图 4 可知，SE 模块主要包含以下几个步骤。

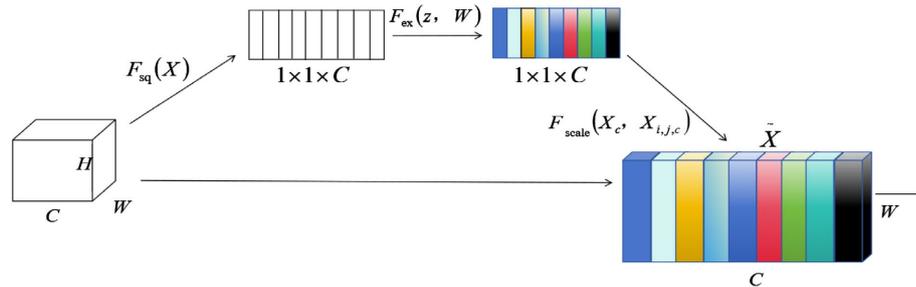


Figure 4. SE attention mechanism  
图 4. SE 注意力机制

1) Squeeze 对输入特征图进行全局平均池化，将每个通道的特征图转换为一个单一的值。假设输入特征图为  $X$ ，其形状为  $H \times W \times C$  (高度、宽度和通道数)。全局平均池化后得到的特征向量  $z$  形状为  $1 \times 1 \times C$ 。

$$z_c = F_{sq}(X) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{i,j,c}$$

2) Excitation 通过两个全连接层来学习每个通道的重要性权重。首先是一个带有 ReLU 激活函数的全连接层，然后是一个带有 sigmoid 激活函数的全连接层。

$$s = F_{ex}(z, W) = \delta[W_2 \cdot \text{ReLU}(W_1 \cdot z)]$$

其中， $W_1$  和  $W_2$  是两个全连接层的权重矩阵， $\delta$  是 sigmoid 激活函数， $s$  是学到的每个通道的权重。

3) Scale 将学到的权重  $s$  重新作用到输入特征图的每个通道上。具体地，使用广播机制将每个通道的特征图乘以对应的权重。

$$\hat{X}_{i,j,c} = F_{scale}(s_c, X_{i,j,c}) = s_c \cdot X_{i,j,c}$$

对于 Mask R-CNN-I 模型，底层通常捕捉基础且简单的特征，如边缘、颜色和纹理，这些特征对图片理解很重要，但缺乏抽象性。因此，在底层加入注意力机制效果不明显，因为特征几乎同等重要。注意力机制对深层网络的抽象特征进行加权，可以突出对任务最有帮助的特征。因此，我们在高层的 mobile2、mobile3、mobile4 模块中加入注意力机制，使网络对任务更敏感。加入后的网络结构如图 5 所示。

### 2.3.3. 特征金字塔网络的改进

FPN 是一种特征金字塔结构，它通过自下而上和自上而下的水平连接来提取多尺度的特征。在特征融合部分将 FPN 与 ASPP (空洞空间卷积池化金字塔) 结合起来。这样的组合充分利用了 FPN 的多尺度特征金字塔结构，以及 ASPP 在局部和全局感受野上的有效信息捕捉能力。这有助于提高模型对目标的理解和分割性能。

ASPP 用于增加感受野并捕获多尺度的上下文信息，同时保持特征图的空间分辨率。一般情况下，它被加在网络的深层部分，以捕捉高级语义信息而不是边缘或纹理等信息，这有助于改善网络对复杂对象和场景的理解。对于像素级的图像分割任务，在深层网络中加入 ASPP 可以扩大感受野，而在浅层加入可能会忽略有用的细节信息。因此，本文选择在 FPN 的 M5、M4 层引入 ASPP，以利用网络已经提取的

复杂特征,在不增加模型规模的情况下,增强对不同尺度结构的理解和处理能力。其引入位置如图6所示。

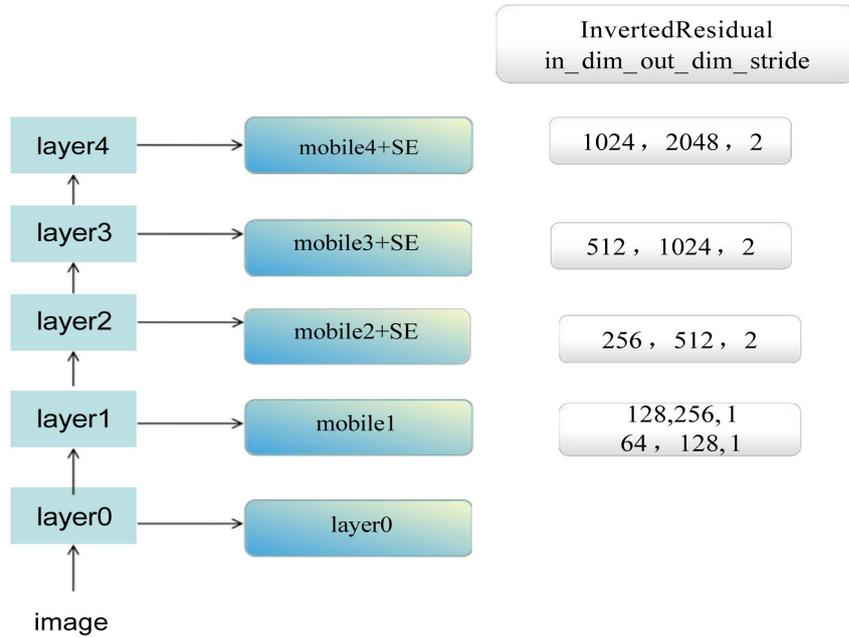


Figure 5. Schematic diagram of where SE attention mechanism is introduced  
图 5. SE 注意力机制引入位置示意图

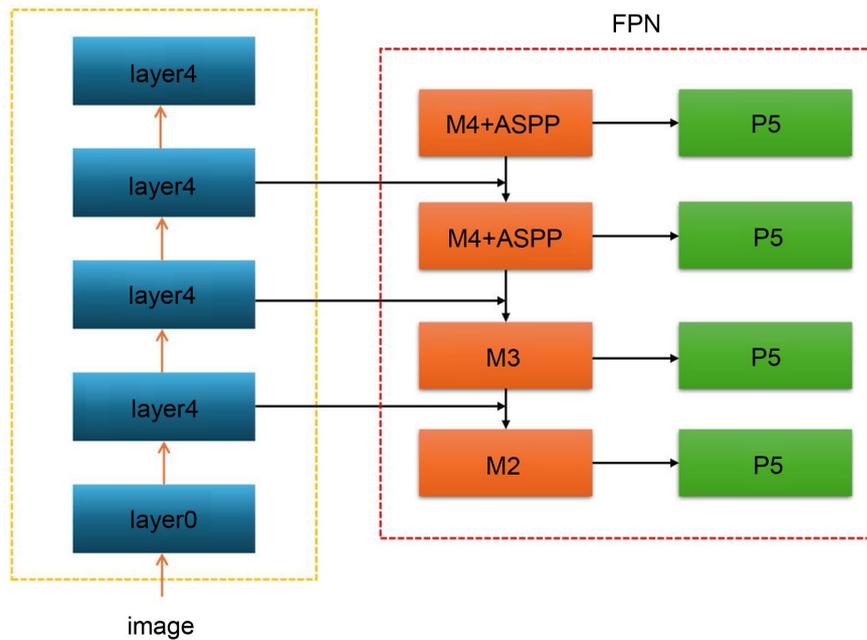


Figure 6. Schematic diagram of the ASPP introduction location  
图 6. ASPP 引入位置示意图

通过对 Mask R-CNN 的主干网络进行了轻量化改进, SE 注意力机制的引入, 以及在特征融合部分加入了 ASPP 空洞金字塔池化, 提升模型对多尺度信息的捕获能力。本文提出的模型总体架构如图7所示。

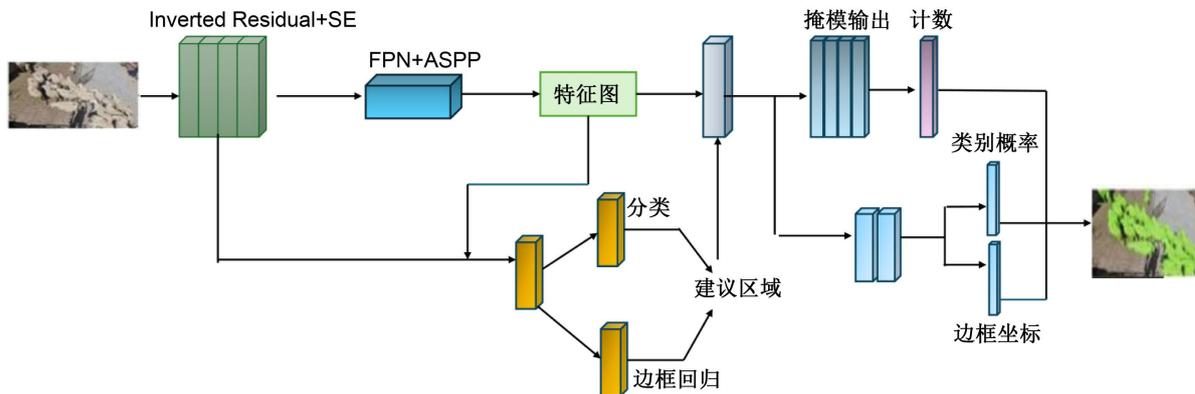


Figure 7. Model architecture diagram  
图 7. 模型架构图

## 2.4. 实验设计

数据集：本文选取在 1.1 节中构建的数据集，共计 3600 张图片。按照 8:2 的比例划分为训练集和测试集。

评价指标：本文使用的评价指标如下：*Params/M* (参数量)：表示模型中需要优化的可训练参数的数量，*Macc/G* (计算量)：表示模型的计算复杂度，单帧时间/s：表示推理单帧图像所需的时间，*MS/M* (模型大小)：表示模型所占用的存储空间大小，*Count\_Accuracy/%* (计数准确度)：评价羊只计数的准确度。其中 *Count\_Accuracy* 的计算公式如下：

$$Count\_Accuracy = \left( 1 - \frac{\sum_{i=1}^n |total\_projects_i - annotated\_count_i|}{annotated\_count_i} \right) \times \frac{1}{total\_annotated} \times 100\%$$

其中 *total\_projects* 为模型预测的羊只数量，*annotated\_count* 为真实标注的羊只数量，*total\_annotated* 为标注数据的数量。

超参数设置：算法开发环境的硬件配置包括 Xeon(R) Silver 4214R 处理器、RTX 3080 显卡、4 GB 显存，操作系统为 Windows 11。软件方面使用了 PyCharm 2021、CUDA 11.3、CUDNN 8.2、Pytorch 框架以及 Python 3.8。训练参数设置为：Epochs 为 110，Batch Size 为 8，学习率为 0.004，优化器为 SGD，权重衰减为  $1e-4$ ，动量为 0.9。

基线模型：本文对基于 YOLOv5 [12]、SSD [13]和 DeepLabv3Plus [14]的计数方法进行了重复试验，以进行性能比较。

### 2.4.1. 消融实验结果与分析

为了验证 Mask R-CNN-I-S-A 模型中各个组件的有效性，本文进行了消融实验。通过逐步移除模型中的关键组件来评估这些组件对模型性能的影响。在本文中，我们主要关注以下几个组件的消融实验：Inverted Residual 模块，SE 注意力机制模块，ASPP 模块。其中 Mask R-CNN：基础模型，不包含任何新增模块。Mask R-CNN-I：在基础模型中添加 Inverted Residual 模块。Mask R-CNN-I-S：在基础模型中添加 Inverted Residual 模块和 SE 注意力机制模块。Mask R-CNN-I-S-A：完整模型，包含所有新增模块。消融实验结果如表 2 所示。

根据表中的数据，逐步引入 Inverted Residual 模块、SE 注意力机制模块和 ASPP 模块后，模型在轻量化的同时也实现了计数准确率的提升。具体而言：参数数量从基础模型 Mask R-CNN 的 23.51 M 减少

**Table 2.** Results of ablation experiments**表 2.** 消融实验结果

模型	Params/M	Macc/G	单帧时间/s	MS/M	计数准确率
Mask R-CNN	23.51	21.58	0.2216	339	96.09%
Mask R-CNN-I	8.37	11.13	0.1368	220	92.23%
Mask R-CNN-I-S	11.12	11.14	0.1687	241	94.32%
Mask R-CNN-I-S-A	14.47	15.94	0.1726	222	96.27%

至 Mask R-CNN-I-S-A 的 14.47 M，减少了 32.95%。单帧推理时间从 Mask R-CNN 的 0.2216 秒减少至 Mask R-CNN-I-S-A 的 0.1726 秒，减少了 22.12%。计数准确率从 Mask R-CNN 的 96.09% 提升至 Mask R-CNN-I-S-A 的 96.27%，提高了 0.18 个百分点。这些结果表明，通过逐步加入 Inverted Residual 模块、SE 注意力机制模块和 ASPP 模块，模型不仅实现了轻量化，而且在保持较高计数准确率的同时，显著提升了推理速度。具体原因如下：Inverted Residual 模块采用深度可分离卷积(Depthwise Separable Convolution)，有效减少了模型的参数数量，从而降低了模型的复杂度。SE 注意力机制模块的引入有助于提高计数准确率，因为它能够更好地聚焦于重要特征，增强了模型对关键特征的关注。ASPP 模块的加入提升了多尺度物体的检测能力，通过不同尺度的空洞卷积捕捉多尺度上下文信息，增强了模型的泛化能力和鲁棒性。

综合来看，这些改进措施不仅提升了模型在羊只计数任务上的性能，而且在资源受限的环境下也保持了高效的推理能力。尤其在羊只密集的场景中，这种轻量化和高效推理的特性显得尤为重要。在羊只密集的场景中应用。分割效果如图 8 所示。

**Figure 8.** Mask R-CNN-I-S-A effect**图 8.** Mask R-CNN-I-S-A 效果

#### 2.4.2. 对比试验结果与分析

本文将改进后得到的 Mask R-CNN-InvertedResidual-SE-ASPP 简记为 Mask R-CNN-I-S-A，并与基于 YOLOv5、SSD 及 DeepLabv3Plus 的计数方法进行了重复试验比较，以便更全面地评估它们的性能。评估指标包括每种方法的参数量、计算量、单帧推理时间、模型大小及计数准确率。对比实验结果如表 3 所示。

**Table 3.** Comparative experimental results  
**表 3.** 对比实验结果

模型	Params/M	Macc/G	单帧时间/s	MS/M	计数准确率
Mask R-CNN-I-S-A	14.47	15.94	0.1726	222	96.27%
YOLOv5	21.2	24.5	0.1193	44.3	80.33%
SSD	25.08	15.46	0.2120	95.71	83.07%
DeepLabv3Plus	38.9	31.35	0.3174	304	92.35%

从表中数据可知根 Mask R-CNN-I-S-A 模型的参数数量为 14.47 M，相比于 YOLOv5 (21.2 M)、SSD (25.08 M) 以及 DeepLabv3Plus (38.9 M) 明显较少。这意味着，在保持高性能的同时，该模型实现了轻量化的设计。此外，Mask R-CNN-I-S-A 的计算量为 15.94 G，与 SSD (15.46 G) 接近，但低于 YOLOv5 (24.5 G) 和 DeepLabv3Plus (31.35 G)。这一结果表明该模型在计算效率上具有良好的表现。在处理时间方面，Mask R-CNN-I-S-A 单帧处理时间为 0.1726 秒，介于 YOLOv5 (0.1193 秒) 和 SSD (0.2120 秒) 之间，并且优于 DeepLabv3Plus (0.3174 秒)，这说明在保持较高准确度的同时，其处理速度也较快。在内存占用方面，Mask R-CNN-I-S-A 为 222MB，介于 YOLOv5 (44.3 MB) 和 SSD (95.71 MB) 之间，但低于 DeepLabv3Plus (304 MB)，显示了该模型在内存使用上的优势。最重要的是，Mask R-CNN-I-S-A 的计数准确率达到 96.27%，超过了 YOLOv5 (80.33%)、SSD (83.07%) 以及 DeepLabv3Plus (92.35%)。这一优异性能主要归功于 Inverted Residual 模块的有效使用，它减少了模型的参数数量；SE 注意力机制模块的引入帮助模型更好地关注关键特征；而 ASPP 模块则提升了模型对不同尺度物体的检测能力，增强了整体的泛化能力和鲁棒性。

综上所述，Mask R-CNN-I-S-A 在羊只计数任务上表现突出，尤其是在保证高准确率的同时实现了轻量化设计，这使其非常适合在资源受限的环境中部署和应用。

### 3. 讨论与结论

本文通过对基于 Mask R-CNN 的轻量化羊只计数模型进行改进，引入了倒残差模块以轻量化特征提取网络，并结合 SE 注意力机制提升了特征提取能力。此外，通过将 ASPP 空洞金字塔池化模块融入 FPN 网络中，增强了特征融合效果。改进后的模型在计数准确率上达到了 96.09%，相比基准模型，mAPseg 提升了 0.7%，mAPbbox 提升了 0.9%。同时，模型的参数量减少了 38.46%，计算量减少了 26.14%，体积减少了 34.52%，单帧推理速度提升了 22.12%。这一改进有效地解决了羊只计数在密集场景下所面临的时效性问题，提高了模型在实际应用中的价值。尽管取得了显著进展，但仍存在数据集覆盖范围有限的问题，可能导致模型在某些特定情况下泛化能力不足；此外，尽管模型已经实现了较大幅度的轻量化，但在模型压缩技术(如剪枝和量化)的应用方面仍有改进空间。未来的研究可以进一步探索这些技术，以实现更深层次的模型轻量化，并通过收集更多样化的数据来提升模型的泛化能力。

### 基金项目

内蒙古自然科学基金项目(2023LHMS06012)；内蒙古自治区直属高校科研基本业务费重点项目(项目编号：BR231407)。

### 参考文献

- [1] 中华人民共和国国家统计局. 中国统计年鉴[M]. 北京: 中国统计出版社, 2023.
- [2] Saleem, M.H., Potgieter, J. and Arif, K.M. (2021) Automation in Agriculture by Machine and Deep Learning Techniques: A Review of Recent Developments. *Precision Agriculture*, **22**, 2053-2091. <https://doi.org/10.1007/s11119-021-09806-x>

- 
- [3] 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.
- [4] 宁远霖, 杨颖, 李振波, 等. 基于改进 YOLOv5 的复杂跨域场景下的猪个体识别与计数[J]. 农业工程学报, 2022, 38(17): 168-175.
- [5] Zhang, G., Zhao, Y., Fu, P., Luo, W., Shao, Q., Zhang, T., *et al.* (2024) A Reliable Unmanned Aerial Vehicle Multi-Target Tracking System with Global Motion Compensation for Monitoring *Procopra przewalskii*. *Ecological Informatics*, **81**, Article 102556. <https://doi.org/10.1016/j.ecoinf.2024.102556>
- [6] Chen, X., Pu, H., He, Y., Lai, M., Zhang, D., Chen, J., *et al.* (2023) An Efficient Method for Monitoring Birds Based on Object Detection and Multi-Object Tracking Networks. *Animals*, **13**, Article 1713. <https://doi.org/10.3390/ani13101713>
- [7] Zheng, Z., Li, J. and Qin, L. (2023) YOLO-BYTE: An Efficient Multi-Object Tracking Algorithm for Automatic Monitoring of Dairy Cows. *Computers and Electronics in Agriculture*, **209**, Article 107857. <https://doi.org/10.1016/j.compag.2023.107857>
- [8] Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Sun, Z., *et al.* (2020) Livestock Classification and Counting in Quadcopter Aerial Images Using Mask R-CNN. *International Journal of Remote Sensing*, **41**, 8121-8142. <https://doi.org/10.1080/01431161.2020.1734245>
- [9] Liu, S., Zhao, C., Zhang, H., Li, Q., Li, S., Chen, Y., *et al.* (2024) ICNet: A Dual-Branch Instance Segmentation Network for High-Precision Pig Counting. *Agriculture*, **14**, Article 141. <https://doi.org/10.3390/agriculture14010141>
- [10] Huang, E., Mao, A., Gan, H., Camila Ceballos, M., Parsons, T.D., Xue, Y., *et al.* (2021) Center Clustering Network Improves Piglet Counting under Occlusion. *Computers and Electronics in Agriculture*, **189**, Article 106417. <https://doi.org/10.1016/j.compag.2021.106417>
- [11] He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017) Mask R-CNN. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 2980-2988. <https://doi.org/10.1109/iccv.2017.322>
- [12] Ultralytics (n.d.) YOLOv5. <https://github.com/ultralytics/yolov5>
- [13] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., *et al.* (2016) SSD: Single Shot MultiBox Detector. *Computer Vision—ECCV 2016*, Amsterdam, 11-14 October 2016, 21-37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)
- [14] Chen, L., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Computer Vision—ECCV 2018*, Munich, 8-14 September 2018, 833-851. [https://doi.org/10.1007/978-3-030-01234-2\\_49](https://doi.org/10.1007/978-3-030-01234-2_49)