基于改进 Cascade Mask R-CNN 与协同注意力机制的群猪姿态识别

王 鲁，刘 晴，曹 月，郝 霞※

（山东农业大学信息科学与工程学院，泰安 271018）

摘 要：猪体姿态识别有助于实现猪只健康状况预警、预防猪病爆发，是当前研究热点。针对复杂场景下群猪容易相互遮挡、粘连，姿态识别困难的问题，该研究提出一种实例分割与协同注意力机制相结合的两阶段群猪姿态识别方法。首先，以 Cascade Mask R-CNN 作为基准网络，结合 HrNetV2 和 FPN 模块构建猪体检测与分割模型，解决猪体相互遮挡、粘连等问题，实现复杂环境下群猪图像的高精度检测与分割；在上述提取单只猪基础上，构建了基于协同注意力机制（coordinate attention，CA）的轻量级猪体姿态识别模型（CA−MobileNetV3），实现猪体姿态的精准快速识别。最后，在自标注数据集上的试验结果表明，在猪体分割与检测环节，该研究所提模型与 Mask R-CNN、MS R-CNN 模型相比，在 $AP_{0.50}$、$AP_{0.75}$、$AP_{0.50:0.95}$ 和 $AP_{0.5:0.95\text{-large}}$ 指标上最多提升了 1.3、1.5、6.9 和 8.8 个百分点，表现出最优的分割与检测性能。而在猪体姿态识别环节，所提 CA−MobileNetV3 模型在跪立、站立、躺卧、坐立 4 种姿态类上的准确率分别为 96.9%、99.1%、99.5%和 98.6%，其性能优于主流的 MobileNetV3、ResNet50、DenseNet121 和 VGG16 模型，由此可知，该研究模型在复杂环境下群猪姿态识别具有良好的准确性和有效性，为实现猪体姿态的精准快速识别提供方法支撑。

关键词：深度学习；图像识别；实例分割；群猪姿态识别；生猪个体提取

doi：10.11975/j.issn.1002-6819.202211212

中图分类号：TP391 **文献标志码**：A **文章编号**：1002-6819(2023)-04-0144-10

王鲁，刘晴，曹月，等. 基于改进 Cascade Mask R-CNN 与协同注意力机制的群猪姿态识别[J]. 农业工程学报，2023，39(4)：144-153. doi：10.11975/j.issn.1002-6819.202211212 http://www.tcsae.org

WANG Lu, LIU Qing, CAO Yue, et al. Posture recognition of group-housed pigs using improved Cascade Mask R-CNN and cooperative attention mechanism[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(4): 144-153. (in Chinese with English abstract) doi：10.11975/j.issn.1002−6819.202211212 http://www.tcsae.org

0 引 言

近年来，生猪集约化饲养给猪场防疫带来巨大挑战，个别猪只一旦染病，会对整个猪场造成极大威胁。研究表明，姿态作为生猪的一项重要生理指标，可用于生猪疾病的预防和早期诊断[1]。而计算机视觉技术为猪体姿态识别研究提供了一种低成本、非接触式的监测新范式。通过姿态识别能及时监测猪的健康情况、及时发现猪的异常行为，从而提高猪场的经济效益。但猪场复杂环境以及生猪个体之间相互接触、遮挡等客观因素，极大增加了猪体姿态识别的难度。因此，在客观因素限制下，构建有效的猪体姿态识别模型，可为实现猪只健康状况预警、预防猪病爆发提供有效思路[2-6]。

在猪体姿态研究领域，目前常用的深度学习方法有目标检测、图像分割等。目标检测可实现复杂背景下生猪个体的框选，但此类方法在杂物遮挡、群猪粘连等场景下容易造成目标框边界模糊或漏检。因此在处理群猪姿态识别上仍存在一定限制。图像分割旨在将图像分成若干个特定的、具有独特性质的区域并提出感兴趣目标。目前，已有部分研究将基于深度学习的图像分割算法引入到饲养生猪[7]、哺乳母猪[8]、泌乳母猪[9]等多种猪只分割场景，同时基于图像分割算法构建了猪只计数[10]、姿态识别[11]等模型，有效地提高了猪只计数与识别的效率和准确率。但上述方法仅将生猪个体从养殖环境中分离出来，并不能检测出一张图片中同类物体的不同个体。

实例级检测可对同类物体中的不同个体进行有效区分，更适用于群猪姿态分割与识别。通过实例分割算法构建群猪实例分割模型解决了猪只粘连导致识别效果差的问题[12]，同时避免了复杂背景的干扰，实现了群猪实例分割[13-15]与计数[16]，验证了实例分割模型在生猪精准分割中的潜力。目前，在群猪分割领域常用的实例分割算法有 Mask R-CNN[17]、Mask Scoring R-CNN[18]、与 SOLOv2[19]等。例如，刘坤等[20]基于 Mask R-CNN、Cascade mask R-CNN 任务分割模型，通过将通道注意力[21]与空间注意力[22]相融合并嵌入特征金字塔网络中，实现了群养环境下的生猪实例检测。目前，基于实例分割的方法已广泛应用于群猪分割领域。但复

收稿日期：2022-11-25 修订日期：2023-01-30

基金项目：国家自然科学基金重大研究计划项目（91746104）；山东省重点研发计划（重大科技创新工程）项目（2022CXGC010609）

作者简介：王鲁，博士，教授，研究方向为机器学习、计算机视觉、智慧农业等。Email：wangl@sdau.edu.cn

中国农业工程学会高级会员：B041507123S

※通信作者：郝霞，博士，副教授，研究方向为计算机视觉。

Email：haoxia@sdau.edu.cn

中国农业工程学会高级会员：B041506918S

杂场景下的群猪分割仍存在一定挑战性，容易出现猪体轮廓分割缺失或模糊现象[23]。另外，上述多数研究仅完成了站立姿态下的生猪个体或者群猪分割，在其他更多生猪姿态方面的适用性仍需进一步探究。

针对上述问题，本文提出一种实例分割与协同注意力机制相结合的“两阶段”混合模型：首先，利用 HrNetV2 中的对多分辨率图像处理方法，改进 Cascade Mask R-CNN 实例分割模型，构建猪体实例分割模型，为深度分离、高度粘连、杂物遮挡等不同场景下生猪个体的高精度分割提供技术支持，为后续猪体姿态识别奠定基础；其次，基于实例结果从图像中提取出生猪个体，为猪体姿态识别提供数据集；最后，引入协同注意力机制，构建轻量级猪体姿态识别模型，为实现猪体姿态的精准快速识别提供方法支撑。

1　数据采集与预处理

1.1　数据采集

本文试验数据采自山东省泰安市岱岳区的养殖场（2021 年 6 月初至 7 月初 09:00–14:00，天气晴，光照偏弱，室内环境），拍摄品种为长白猪。猪栏大小为 4 m×2.5 m×1.2 m，每栏数量 3～8 头不等。由于多数研究普遍从俯视角度[24]采集猪体姿态，视角相对单一，并不能完全反映生猪的姿态特征。因此，本文从多视角、多角度获取生猪姿态，构建猪体姿态数据集。选取 8 栏群养生猪作为研究对象，使用相机（SONY SELP1650）采集生猪站立、坐立、躺卧、跪立四种姿态的图片。此外还从多个网站获取生猪图像数据，以提升数据集的多样性与鲁棒性。最终共采集图片 2 980 张，网络图片 621 张。

1.2　数据预处理

本研究按照以下 4 个步骤进行处理，以得到最终的试验数据集。

1）数据集筛选。采取人工方式，依据模糊、缺失等指标对图像数据集进行筛选，最终得到 2 812 张拍摄图片，498 张网络图片。

2）数据集增强。对 1)中的图像进行数据增强，每张图片以 50%的概率随机执行高斯噪声、镜面翻转、亮度增强中的 1～3 种变换，每张图片至少生成 1 张增强图片。经上述处理后，共获得图片 6 620 张。部分增强效果图见图 1。

3）数据集划分。将 2）中的数据集按照 7∶2∶1 划分为 3 部分，包括训练集 4 634 张，测试集 1 324 张，验证集 662 张。

4）数据集标注。使用 Labelme 数据标注工具对数据增强后的训练集和验证集进行标注，同时生成 json 文件，最终完成生猪实例分割数据集的制作。

a. 原始图像　a. Original image　b. 镜像翻转　b. Mirror flip　c. 亮度增强　c. Brightness enhancement　d. 添加高斯噪声　d. Adding Gaussian noise

图 1　数据增强示例

Fig.1　Data enhancement example

2　猪体姿态识别方法

2.1　模型整体框架

本研究基于实例分割与协同注意力机制相结合的两阶段方法对群猪姿态进行识别。如图 2 所示，第一阶段以实例分割数据集作为输入，以 Cascade Mask R-CNN 为基准网络对不同场景下的群猪图像进行高精度分割，并依据分割结果提取生猪个体，以构建生猪个体姿态数据集。第二阶段引入协同注意力机制，构建 CA−MobileNetV3 猪体姿态识别模型，以降低背景区域对姿态识别的干扰，增强猪体关键部分的特征学习，从而实现单只猪个体姿态的精准快速识别。

2.2　基于实例分割的生猪个体提取

2.2.1　猪体实例分割模型构建

针对复杂场景导致猪体分割效果差的问题，本文以 Cascade Mask R-CNN[25]为基准网络，构建猪体实例分

割模型。改进后的 Cascade Mask R-CNN 网络模型如图 3 所示。

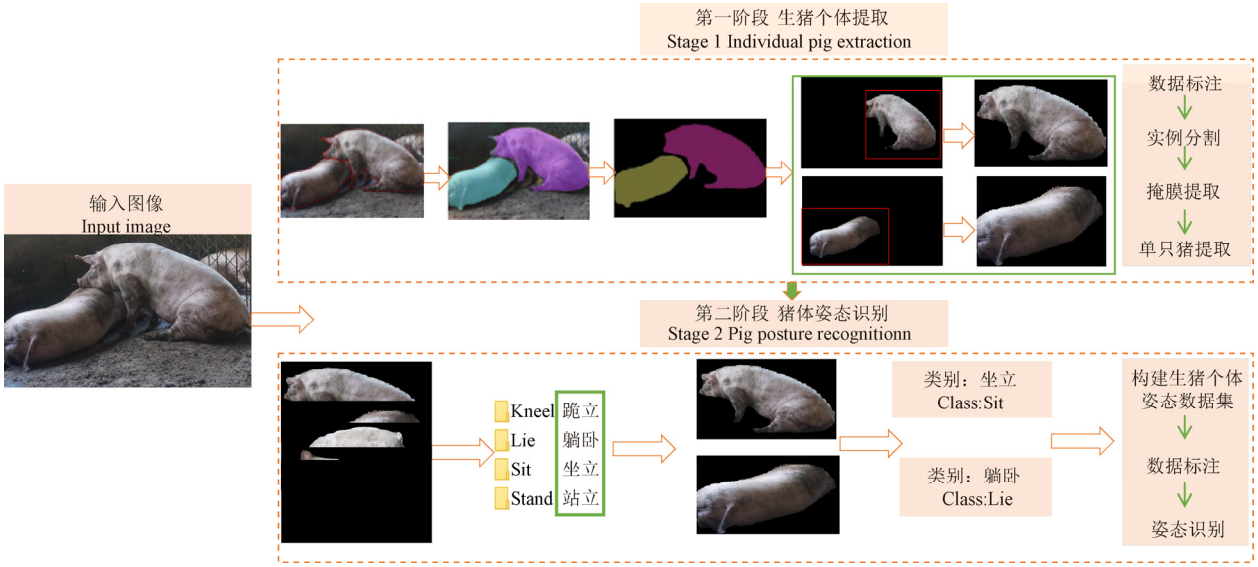
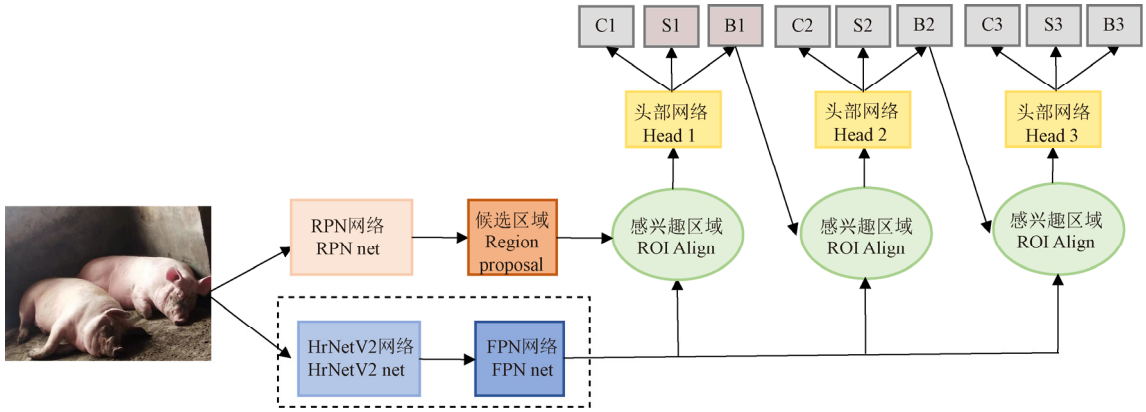


图 2 模型整体框架

Fig.2 Overall framework of the model



注：C1、S1、B1 分别代表 IOU 阈值为 0.5 的分类、检测、分割结果；C2、S2、B2 分别代表 IOU 阈值为 0.6 的分类、检测、分割结果；C3、S3、B3 分别代表 IOU 阈值为 0.7 的分类、检测、分割结果。

Note: C1, S1, and B1 represent the classification, detection, and segmentation results with IOU threshold of 0.5, respectively; C2, S2, and B2 represent the classification, detection, and segmentation results with IOU threshold of 0.6, respectively; C3, S3, and B3 represent the classification, detection, and segmentation results with IOU threshold of 0.7, respectively.

图 3 改进的 Cascade Mask R-CNN 网络模型

Fig.3 Improved Cascade Mask R-CNN Network Model

Cascade Mask R-CNN 通过串行级联多个不同 IOU 阈值头部网络（界定正负样本），使不同 IOU 值检测与其相对应的 IOU 值目标，以达到不断优化预测的目的，且前一个检测模型的输出作为后一个检测模型的输入，位置越靠后，其界定正负样本的 IOU 阈值越大。因此，该模型利用不断提高的阈值，在保证样本数不减少的情况下训练出高质量的检测器，通过级联检测网络预测结果，Cascade Mask R-CNN 计算式如下：

$$x_t^{\text{box}} = P\left(x, r_{t-1}\right), r_t = B_t\left(x_t^{\text{box}}\right) \quad (1)$$

$$x_t^{\text{mask}} = P\left(x, r_{t-1}\right), m_t = M_t\left(x_t^{\text{mask}}\right) \quad (2)$$

式中 x 表示骨干网络的 CNN 特征；x_t^{box} 和 x_t^{mask} 表示由候选 RoI 区域导出的 box 和 mask 特征；$P(\ )$为池化算子，如 RoI Align 或 RoI 池化；B_t 和 M_t 表示第 t 阶段的边界框和掩码头；r_t 和 m_t 表示对应的边界框预测和掩码预测。

HrNetV2[26]通过并行连接多个分辨率子网，在各个分支上分辨率保持不变情况下，反复融合多分辨率特征，使模型在提取大目标特征时，也能获得较细粒度的目标和边缘细节特征信息。因此，本研究引入 HrNetV2 作为 Cascade Mask R-CNN 的特征提取网络，以提高模型在不同场景下的猪体检测与分割能力。

此外，本研究还在 HrNetV2 基础上引入特征金字塔网络（feature pyramid networks，FPN），在有效融合不同尺度特征基础上，将不同尺寸的猪体特征映射为不同层次的特征图，为解决猪体间重叠、粘连等问题提供数据支撑。

2.2.2 生猪个体提取

本研究基于实例分割结果提取生猪个体，并在此基础上构建生猪个体姿态识别数据集，为后续姿态识别提供数据支持。具体方法如下：

1）生猪个体提取。如图 4 所示，首先，从掩膜图中提取出生猪个体轮廓；其次，将生猪个体轮廓图像分别

与原图进行对应映射，以提取生猪个体图像；最后，将获取到的图像进行截取，并依据轮廓等指标对截图图像进行筛选，最终共获取生猪个体数据集 8 560 张。

2）构建生猪个体姿态识别数据集。将 1）中的数据集按照跪立、站立、躺卧、坐立四种姿态进行标注。此外，为防止类间不平衡问题，采用几何变换对坐立、跪卧两类图像进行扩充，最终得到包含 9 220 张图像的生猪个体姿态数据集，并将其按照 7∶2∶1 分别划分为训练集、验证集和测试集。

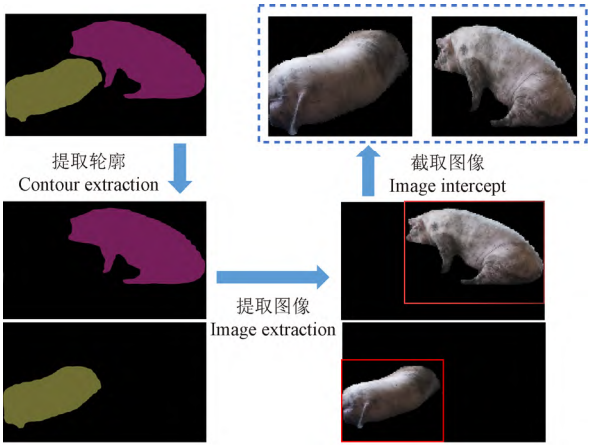


图 4　生猪个体提取过程
Fig.4　Extraction process of pig individual

2.3　猪体姿态识别模型构建

为实现猪体姿态的准确识别，本研究提出了 CA-MobileNetV3 模型，该模型选用轻量化模型 MobileNetV3[27]作为基准网络，并引入协同注意力机制（coordinate attention，CA）对其进行改进，以增强对起重要作用的位置信息的学习能力，从而提高猪体姿态识别的准确率。此外，如式（3）所示，该模型选用 Hard-swish 激活函数代替 swish 函数，旨在减少模型的运算量，提高模型的性能。

$$h-\mathrm{swish}[r]=r\frac{\mathrm{ReLU}(r+3)}{6} \tag{3}$$

式中 r 为模型输入，即由上节实例分割模型得到的生猪个体图像信息；ReLU 为激活函数。

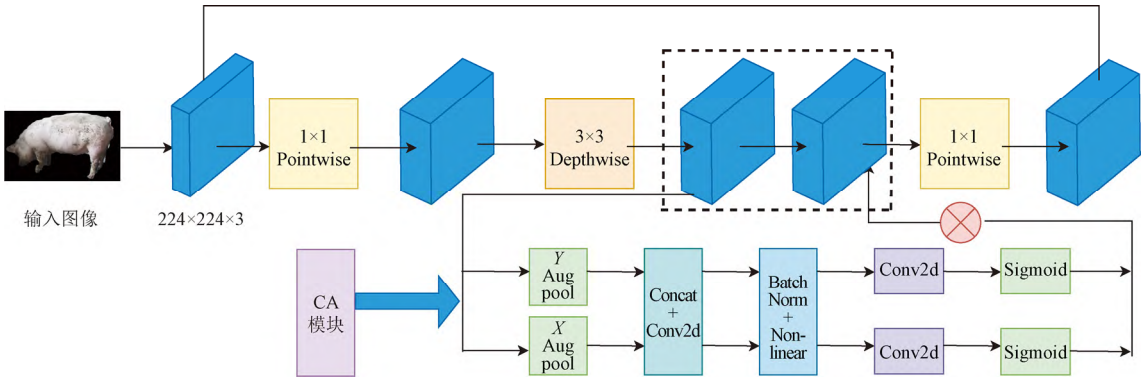
本文引入协同注意力机制[28]代替 MobileNetV3 中的压缩奖惩（squeeze−and−excitation，SE）机制[29]，以弥补 SE 模块只关注通道信息的不足问题。具体地，将通道注意力分解为 2 个一维特征编码过程，分别沿 2 个空间方向聚合特征，高度和宽度的通道输出分别为

$$Z_c^h(h)=\frac{1}{W}\sum_{0\leqslant i\leqslant W}x_{c(h,\ i)} \tag{4}$$

$$Z_c^w(w)=\frac{1}{H}\sum_{0\leqslant j\leqslant H}x_{c(j,\ w)} \tag{5}$$

式中 H 和 W 分别为高度和宽度；x_c 为通道特征映射；Z_c^h 为高度为 h 的第 c 通道；Z_c^w 为宽度为 w 的第 c 通道。

本研究使用 224×224×3 猪体图像作为 CA-MobileNetV3 模型输入。CA-MobileNetV3 的模型结构（如图 5 所示），首先，通过 1×1 的标准卷积对输入的猪体图像进行升维；其次，进行卷积核为 3×3 的深度卷积；再次，将获取的猪体特征输入到 CA 模块，得到增强后的猪体姿态注意力特征图；最后，采用 1×1 卷积核对输出特征降维，并与原来猪体特征进行融合。



注：CA 代表协同注意力；Conv2d 代表 2d 卷积操作；Pool 代表池化；*X*/*Y* Avg Pool 代表 *X*/*Y* 方向平均池化；Concat 代表特征图拼接；⊗ 代表通道级相乘；Pointwise 代表标准卷积；Depthwise 代表深度卷积。
Note: CA represents for coordinate attention; Conv2d represents 2d convolution operation; Pool represents pooling; *X*/*Y* Avg Pool represents average pooling in *X*/*Y* direction; Concat represents feature map splicing; ⊗represents channel level multiplication; Pointwise represents standard convolution; Depthwise represents deep convolution.

图 5　CA-MobileNetV3 模型结构
Fig.5　CA-MobileNetV3 model structure

3　结果与分析

3.1　猪体实例分割

3.1.1　试验参数

本研究所提模型的运行环境为 Pytorch 3.8，选用 SDG 作为优化器，猪体实例分割模型训练参数设置如表 1 所示。

3.1.2　评价指标

为衡量模型的分割性能，本研究采用平均精度（average precision, AP）作为模型的评价指标。真正例（true positive，TP）与假正例（false positive, FP）受不同 IOU 阈值的直接影响，且 TP 与 FP 值会影响精确率（precision）与召回率（recall），从而导致 AP 指标的变化。本研究

选用了 3 种不同的 IOU 阈值，即 0.50、0.75、0.50:0.95，分别表示为 $AP_{0.50}$、$AP_{0.75}$、$AP_{0.50:0.95}$，其中 $AP_{0.50:0.95}$ 表示 IOU 阈值以 0.05 为步长，取值范围在 0.50 到 0.95 之间的 AP 平均指标值。同时依据 COCO 计算指标作为目标物体面积的划分标准 ，本研究将生猪分为小目标（生猪个体区域面积<32^2 像素点）、中等目标（32^2 像素点<生猪个体区域面积<96^2 像素点）与大目标（生猪个体区域面积>96^2 像素点），并将 $AP_{0.50:0.95}$ 在计算大目标条件下的 AP 指标值记为 $AP_{0.50:0.95\text{-large}}$。

表 1 猪体实例分割模型训练参数

Table 1 Training parameters of pig instance segmentation model

参数名称 Parameter name	参数值 Parameter value
迭代次数 Epoch	50
动量因子 Momentum factor	0.9
初始优化器学习率 Initial optimizer learning rate	0.02
初始学习率 Initial learning rate	0.01
IOU 阈值 Intersection over union threshold	级联网络 IOU 阈值分别设置为 0.5、0.6 与 0.7；其他试验对比网络设置为 0.5
Core 阈值 Core threshold	0.5
权重衰减率 Weight decay rate	0.0001

3.1.3 对比试验

为验证改进 Cascade Mask R-CNN 分割模型的性能，本研究从以下两个步骤设置对比试验。

1）分别选用 ResNet50、ResNet101 和 Res2Net 等主流网络作为 Cascade Mask R-CNN 骨干网络；然后与优化后的 Cascade Mask R-CNN 模型进行比，计算其在测试集上 AP 值（表 2）。

表 2 同一任务网络（Cascade Mask R-CNN）下不同骨干网络分割的平均精度

Table 2 Average precision(AP) of different IOU thresholds for same task networks (Cascade Mask R-CNN) under different backbone networks %

骨干网络 Backbone	$AP_{0.50}$	$AP_{0.75}$	$AP_{0.50:0.95}$	$AP_{0.50:0.95\text{-large}}$
ResNet50	92.7	87.0	77.1	80.5
ResNet101	93.0	90.8	79.0	82.4
Res2Net	92.9	91.0	81.7	85.2
HrNetV2	95.8	95.8	85.2	89.2

注：AP0.50 和 AP0.75 分别表示 IOU 阈值为 0.50 和 0.75 的 AP 指标值；AP0.50:0.95 表示 IOU 阈值以 0.05 为步长，取值范围在 0.50 到 0.95 之间的 AP 平均指标值；AP0.50:0.95-large 表示 AP0.50:0.95 在计算大目标条件下的数值。

Note: AP0.50, and AP0.75 represent the AP value with IOU threshold of 0.50 and 0.75, AP0.50:0.95 means the AP value that IOU threshold interval is set to 0.05 with the range between 0.50 and 0.95, AP0.50:0.95-large indicates the value of AP0.50:0.95 under the condition of large targets.

2）将 HrNetV2 网络进行特征提取后的结果分别作为 Mask R-CNN、MS R-CNN 与 Cascade Mask R-CNN 的输入，并计算其在测试集上的 AP 值（表 3）。

由表 2 可知，当采用 HrNetV2 作为 Cascade Mask R-CNN 的骨干网络时，其分割精度最优。而当以 ResNet50 作为骨干网络时，AP 指标明显低于其他 3 种骨干网络，主要原因为：与其他 3 种骨干网络相比，ResNet50 网络层次较浅，难以对深层次特征进行更抽象化的表达与学习。相应地，ResNet101、Res2Net 网络在 AP 指标上有明显的提升。当采用 HrNetV2 网络作为骨干网络时，较 ResNet50 在 AP 相应指标上分别提升了 3.1、8.8、8.1、8.7 个百分点。由此可知，在选用不同骨干网络情况下 Cascade Mask R-CNN 任务网络性能状况不尽相同。一方面，随着骨干网络加深，级联网络 Cascade Mask R-CNN 能够抽取更丰富的特征，其效果越佳。另一方面，与其他 3 种骨干网络相比，HRNetV2 采用并联结构，在每一分支上均保持分辨率不变，细节信息损失较少。因此 HRNetV2 更适用于 Cascade Mask R-CNN。

表 3 不同任务网络下相同骨干网络（HrNetV2）分割的平均精度

Table 3 AP of different IOU thresholds for different task networks under same backbone networks(HrNetV2) %

模型 Model	$AP_{0.50}$	$AP_{0.75}$	$AP_{0.50:0.95}$	$AP_{0.50:0.95\text{-large}}$
Mask R-CNN	95.0	94.8	78.4	81.2
MS R-CNN	94.5	94.3	78.3	80.4
Cascade Mask R-CNN	95.8	95.8	85.2	89.2

由表 3 可知，Cascade Mask R-CNN-HRNetV2 网络分割精度最优。以 HRNetV2 作为骨干网络，选用 Cascade Mask R-CNN 作为任务网络比 MS R-CNN、Mask R-CNN 分别在 $AP_{0.5:0.95\text{-large}}$ 指标上提升了 8.8 和 8.0 个百分点，在 $AP_{0.50:0.95}$ 指标上分别提升了 6.8 和 6.9 个百分点，$AP_{0.75}$ 指标上分别提高了 1.0 和 1.5 个百分点。其原因可能是，在大目标问题上，级联网络 Cascade Mask R-CNN 融合多种 IOU 阈值，分割性能更优。即使在提升幅度不明显的 $AP_{0.50}$ 指标上，其性能仍能提升 0.8、1.3 个百分点。在群养生猪环境下，由于图像中存在较多干扰信息，因此选用 Cascade Mask R-CNN 作为分割任务网络一定程度上能够提取到更合理的图像特征。

3.1.4 试验结果可视化

为进一步探索不同模型在不同场景下的图像分割的性能，本文将测试集分为深度分离、相互粘连、杂物遮挡 3 种，以 3.1.3 节模型组合为例，对分割与目标检测结果进行可视化，部分结果如图 6 所示。

由于采集图像中生猪对象的信息不同，因此目标检测、分割效果都会受到不同程度的影响。从目标检测结果可看出，在深度分离的图像中，猪体轮廓比较完整清晰，所有网络均可准确地对猪体进行识别检测，且具有较好的分割效果；而在相互粘连图像中，存在猪体轮廓模糊或被遮挡现象。因此 Mask R-CNN 网络出现了错检，无法对粘连的猪体进行检测区分，且分割效果欠佳的问题，MS R-CNN 与 Cascade Mask R-CNN-HrNetV2 模型能够准确的对实例进行区分。在杂物遮挡的图像中，因杂物颜色与猪体部分颜色重叠，因此对目标猪体的检测也有一定的难度[30]，易出现错检、漏检问题。从实例分割上来看，Cascade Mask R-CNN-HrNetV2 模型在 3 种场景下的分割与检测效果最佳。对于相互粘连的情况，Cascade Mask R-CNN-HrNetV2 能够将其有效地分割出来，且猪体分

割边缘更佳精准，MS R-CNN 与 Mask R-CNN 分割结果较为粗糙，Mask R-CNN 将本不属于生猪个体的部分错误地归为一类。对于深度分离的情况，由于猪体轮廓比较清晰，模型都具有很好的分割效果；对于杂物遮挡的情况，可以看出，在细致轮廓边缘（图中的尾巴、耳朵等）上，Cascade Mask R-CNN- HrNetV2 模型分割更为细致。

上述试验结果说明在不同场景中 Cascade Mask R-CNN- HrNetV2 均能够取得最优的检测、分割性能，因此可将其作为猪体实例分割模型。

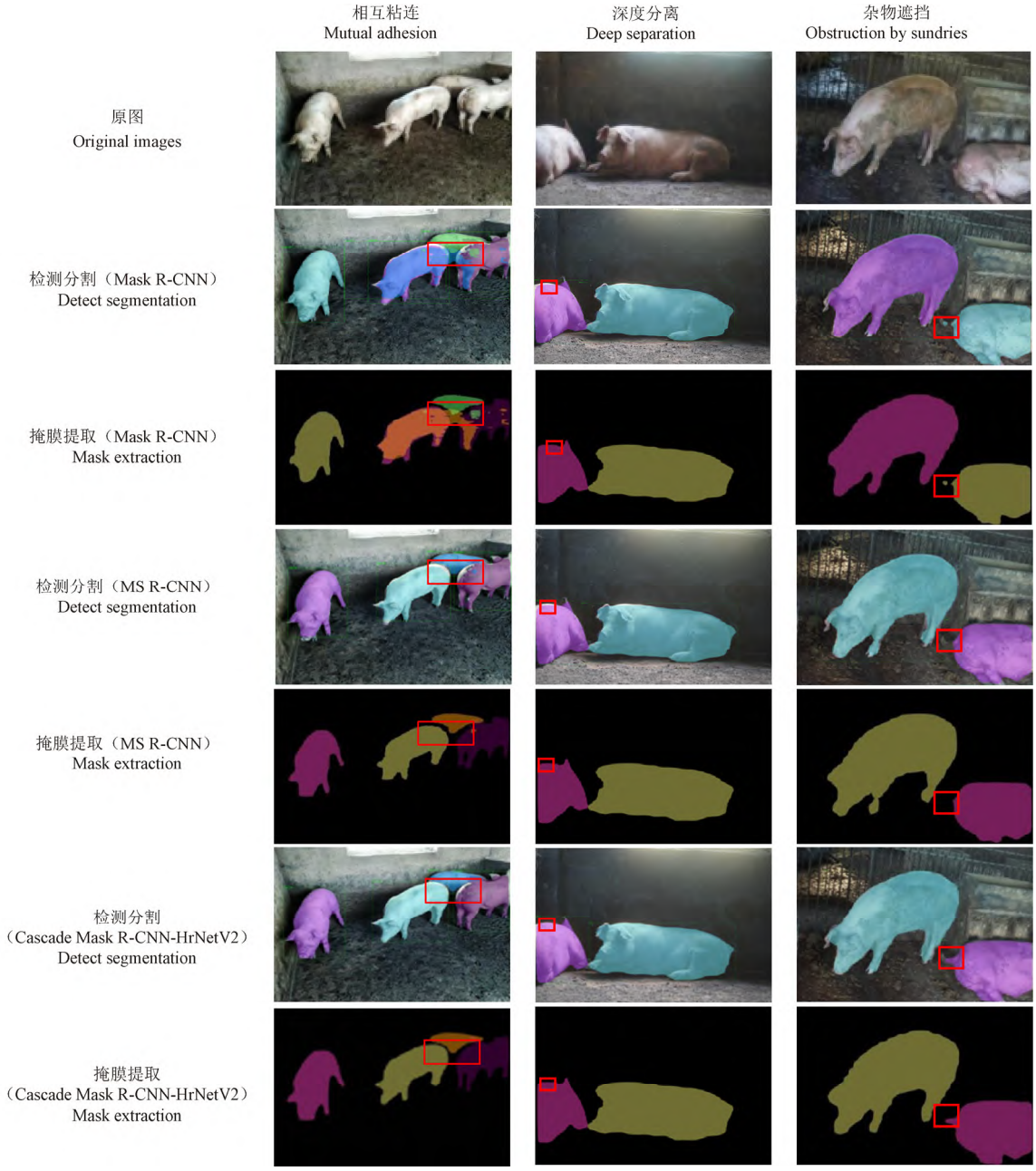


图 6　猪体实例分割部分可视化结果图

Fig.6 Visualization result diagram of pig instance segmentation

3.2　猪体姿态识别

3.2.1　试验参数

本研究模型同样采用 Pytorch 框架实现，选用 Adam 优化算法替代 SGD 优化算法，初始学习率设置为 0.001，Batch size 设置为 32，Epoch 设置为 30。

3.2.2　评价指标

本研究采用准确率、精确率、召回率和 F1 值作为猪体姿态识别的性能评估指标。

3.2.3　姿态识别结果与分析

本试验首先以生猪个体姿态图像作为模型输入，然后采用融合协同注意力机制的 MobileNetV3 网络进行猪体姿态特征提取，最后使用 softmax 函数进行猪体姿态分类。

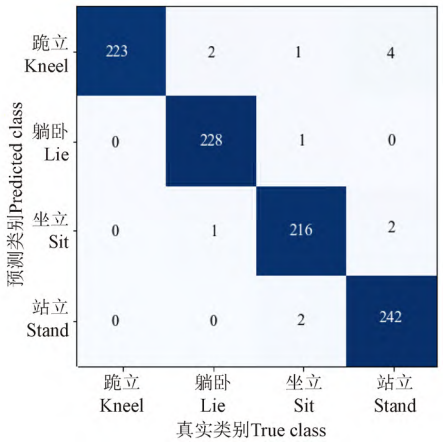
将 CA−MobileNetV3 与原 MobileNetV3 网络进行对比试验，从试验对比结果（表 4）及混淆矩阵（图 7）的对比结果可看出，CA-MobileNetV3 在准确率上比 MobileNetV3 整体提升了 3.5 个百分点；在召回率、F1

值以及精确率上，CA−MobileNetV3 在跪立、躺卧、坐立、站立 4 个姿态的试验结果均优于 MobileNetV3，特别是在较难识别的跪立、坐立姿态识别中，分别最多提升了 7.6、6.8、6.1 个百分点，这很大程度上得益于 CA-MobileNetV3 模型在通道和位置 2 个维度上关注生猪图像的重要特征，而 MobileNetV3 模型中的 SE 模块只关注了通道信息。

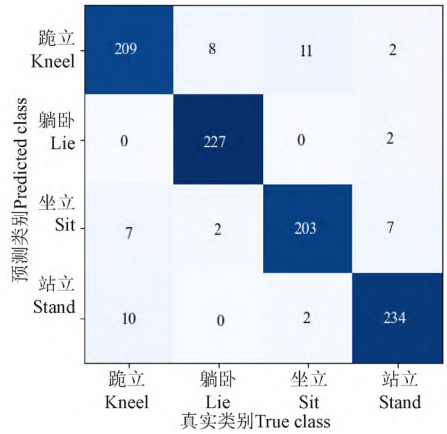
表 4 模型整体试验对比结果

Table 4 Model overall experimental comparison results %

模型 Model	召回率 Recall				F1 值 F1 value				精确率 Precision				准确率 Accuracy
	跪立 Kneel	躺卧 Lie	坐立 Sit	站立 Stand	跪立 Kneel	躺卧 Lie	坐立 Sit	站立 Stand	跪立 Kneel	躺卧 Lie	坐立 Sit	站立 Stand	
MobileNetV3	92.4	95.7	93.9	95.5	91.6	97.4	93.3	95.3	90.8	99.1	92.7	95.1	94.7
CA-MobileNetV3	100	98.7	98.6	97.6	98.4	99.1	98.6	98.4	96.9	99.6	98.6	99.1	98.2



a. CA-MobileNetV3



b. MobileNetV3

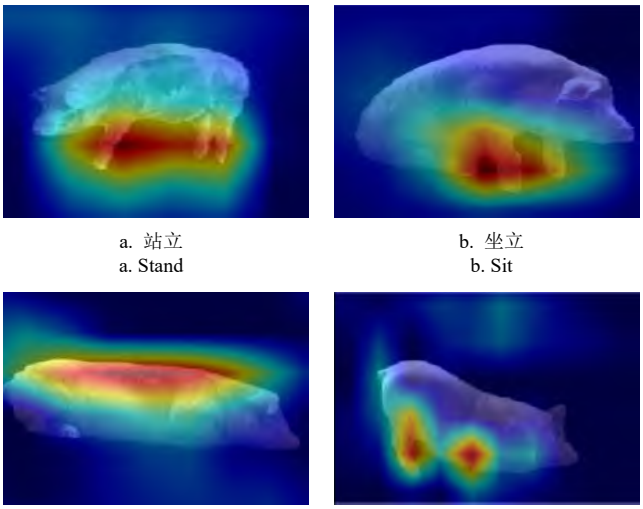
图 7 猪体姿态识别混淆矩阵结果图

Fig.7 Results of confusion matrix for pig posture recognition

为了验证 CA-MobileNetV3 模型的有效性，本研究通过类别热力图对猪体姿态特征提取部分进行了可视化。如图 8 所示，站立、坐立以及跪立姿态类的四肢、躺卧姿态类的身体部位呈高亮偏暖色调，表明协同注意力机制的引入使模型更加关注于识别目标的重要特征区域。对站立、坐立以及跪立姿态类，则主要关注四肢特征信息，通过对四肢特征信息的提取（前肢、后肢体是否弯曲），判别其对应的姿态类，对躺卧姿态类，因存在四肢遮挡的问题，所以模型更为关注身体部位，对身体特征的激活程度更高，判别其为躺卧姿态的可能性越大。

为进一步验证 CA-MobileNetV3 模型的有效性，本研究还引入了其他 3 种较经典的卷积网络模型，即 ResNet50、DenseNet121 和 VGG16，对比模型均参考模型原始框架以及参数进行设置[30]。

图 9 给出了 ResNet50、DenseNet121、MobileNetV3、VGG16 和 CA-MobileNetV3 模型的损失值变化过程。由图 9 可知，VGG16、DenseNet12 模型损失值波动性较大、不稳定。ResNet50 模型损失值波动较小，但随着训练次数增加，损失值逐渐处于稳定状态。MobileNetV3、CA-MobileNetV3 模型损失值变化相对平滑。其中，CA-MobileNetV3 模型无论从收敛速度还是模型稳定性方面都表现最优。



a. 站立 a. Stand　b. 坐立 b. Sit

c. 躺卧 c. Lie　d. 跪立 d. Kneel

图 8 猪体姿态特征提取类别热力图

Fig.8 Class activation mapping of pig posture feature extraction

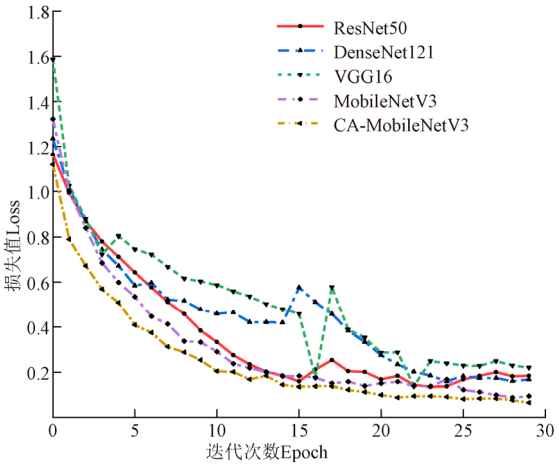


图 9 不同模型损失值变化

Fig.9 Loss value change of different models

由表 5 可知，与经典深度学习模型对比，CA-MobileNetV3 模型在识别准确率上表现最佳，达到了 98.2%，比 VGG16、DenseNet121、ResNet50 相应值分别

高出了 8.6、3.0 与 2.8 个百分点；在召回率与 F1 值上，较其他模型分别最大增加了 10.8、13.1 个百分点。在跪立、站立、躺卧 3 种姿态上，VGG16 模型在识别精确率上表现最差，原因是其网络层次较浅，受图像背景干扰信息较多，难以学习其深层特征的语义信息。DenseNet121 虽然在跪立姿态类上识别精确率略高于 CA-MobileNetV3 模型，但在其他姿态类上识别较差。ResNet50 模型在 4 类姿态识别中，其精确率、召回率、F1 值均低于 CA-MobileNetV3 模型。与其他 3 种经典模型相比，本研究所构建的 CA-MobileNetV3 模型，由于引入了协同注意力模块，增强了模型对猪体姿态识别任务中重要位置信息的学习能力，以及提取猪体显著性区域的图像特征的能力，在姿态识别性能上有明显提升，且在所有评价指标上均取得了较优的结果。

表 5　不同算法试验结果

Table 5　Experimental results of different algorithms

模型 Model	召回率 Recall rate /%				F1 值 F1 value /%				精确率 Presion /%				准确率 Accuracy/ %	模型大小 Model size/ MB	训练时间 Train time/ ms
	跪立 Kneel	躺卧 Lie	坐立 Sit	站立 Stand	跪立 Kneel	躺卧 Lie	坐立 Sit	站立 Stand	跪立 Kneel	躺卧 Lie	坐立 Sit	站立 Stand			
VGG16	89.2	96.3	85.2	89.3	85.3	95.6	87.1	91.4	80.1	95.2	89.5	94.4	89.6	528	14.50
DenseNet121	95.1	98.2	93.4	92.5	96.2	98.4	92.8	92.8	97.2	95.8	90.2	94.2	95.2	411	12.40
ResNet50	95.2	98.4	91.2	97.4	96.2	98.5	93.4	96.2	96.4	97.1	94.3	94.6	95.4	384	10.35
CA-MobileNetV3	100	98.7	98.6	97.6	98.4	99.1	98.6	98.4	96.9	99.5	98.6	99.1	98.2	151	6.28

4　结　论

本文通过生猪图像采集与标注，构建了生猪实例分割数据集和生猪个体姿态识别数据集；将 Cascade Mask R-CNN 网络与 HrNetv2、特征金字塔网络 FPN 有机结合，构建了群猪实例分割模型 Cascade Mask R-CNN-HrNetv2；采用轻量化模型 MobileNetV3 作为基准网络，引入协同注意力机制，构建了猪体姿态识别模型 CA−MobileNetV3，通过对比试验分析，得出如下结论：

1）与 Mask R-CNN、MS R-CNN 进行生猪实例分割方面对比试验，结果表明，本研究所提出的模型在 IOU 阈值分别为 0.50、0.75 的平均精度均达到 0.958，在 IOU 阈值以 0.05 为步长，取值范围在 0.50 到 0.95 之间的平均精度较对比模型提高 6.8～8.8 个百分点，表现出较好的实例分割性能。

3）与 VGG16、DenseNet121、ResNet50 等网络模型在个体姿态识别方面进行对比试验，结果表明，本研究所提出的模型在跪立、躺卧、坐立、站立 4 类姿态上的识别精确率分别达到了 96.9%、99.5%、98.6%、99.1%，在整体正确率上高于对比模型 2.8～8.6 个百分点，在姿态识别性能上有明显提升。

［参　考　文　献］

[1] 程玉兰. 非洲猪瘟背景下推进生猪产业发展的新思考[J]. 中国畜禽种业，2021，17(3)：29-30.

[2] 燕红文，刘振宇，崔清亮, 等. 基于特征金字塔注意力与深度卷积网络的多目标生猪检测[J]. 农业工程学报，2020，36(11)：193-202.
YAN Hongwen, LIU Zhenyu, CUI Qingliang, et al. Multi-target detection based on feature pyramid attention and deep convolution network for pigs[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(11): 193-202. (in Chinese with English abstract)

[3] 滕光辉. 畜禽设施精细养殖中信息感知与环境调控综述[J]. 智慧农业，2019，1(3)：1-12.
TENG Guanghui. Information sensing and environment control of precision facility livestock and poultry farming[J]. Smart Agriculture, 2019, 1(3): 1-12. (in Chinese with English abstract)

[4] ZHANG L, GRAY H, Ye X, et al. Automatic individual pig detection and tracking in pig farms[J]. Sensors, 2019, 19(5): 1188.

[5] WANG S, JIANG H, QIAO Y, et al. The research progress of vision-based artificial intelligence in smart pig farming[J]. Sensors, 2022,22(17):6541.

[6] 李丹，陈一飞，李行健，等. 计算机视觉技术在猪行为识别中应用的研究进展[J]. 中国农业科技导报，2019，21(7)：59-69.
LI Dan, CHEN Yifei, LI Xingjian, et al. Research advance on computer vision in behavioral analysis of pigs[J]. Joumal of Agricultural Science and Technology, 2019, 21(7): 59-69. (in Chinese with English abstract)

[7] 胡志伟，杨华，娄甜田，等. 采用双重注意力特征金字塔网络检测群养生猪[J]. 农业工程学报，2021，37(5)：166-174.
HU Zhiwei, YANG Hua, LOU Tiantian, et al. Instance detection of group breeding pigs using a pyramid network with dual attention feature[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(5): 166-174. (in Chinese with English abstract)

[8] YANG A, HUANG H, ZHENG C, et al. High-accuracy image segmentation for lactating sows using a fully convolutional network[J]. Biosystems Engineering, 2018, 176: 36-47.

[9] YANG A, HUANG H, ZHU X, et al. Automatic recognition of sow nursing behaviour using deep learning-based segmentation and spatial and temporal features[J]. Biosystems Engineering, 2018, 175: 133-145.

[10] LIU C, SU J, WANG L, et al. LA-DeepLab V3+: A novel counting network for pigs[J]. Agriculture, 2022,12(2):284.

[11] PSOTA E T, MITTEK M, PEREZ L C, et al. Multi-pig part detection and association with a fully-convolutional network[J]. Sensors (Basel, Switzerland), 2019, 19(4): 852.

[12] 高云，郭继亮，黎煊，等. 基于深度学习的群猪图像实例分割方法[J]. 农业机械学报， 2019， 50(4)：179-187.
GAO Yun, GUO Jiliang, LI Xuan, et al. Instance-level segmentation method for group pig images based on deep learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(4): 179-187. (in Chinese with English abstract)

[13] 孙龙清, 李玥, 邹远炳, 等. 基于改进 Graph Cut 算法的生猪图像分割方法[J]. 农业工程学报, 2017,33(16):196-202.
SUN Longqing, Li Yue, ZOU Yuanbing, et al. Pig image segmentation method based on improved Graph Cut algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(16): 196-202. (in Chinese with English abstract)

[14] LEI K, ZONG C, YANG T, et al. Detection and analysis of sow targets based on image vision[J]. Agriculture, 2022, 12(1): 73.

[15] LU J, WANG W, ZHAO K, et al. Recognition and segmentation of individual pigs based on swin transformer[J]. Anim Genet, 2022, 53(6): 794-802.

[16] 胡云鸽，苍岩，乔玉龙，等. 基于改进实例分割算法的智能猪只盘点系统设计[J]. 农业工程学报，2020，36(19)：177-183.
HU Yunge, CANG Yan, QIAO Yulong, et al. Design of intelligent pig counting system based on improved instance segmentation algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(19): 177-183. (in Chinese with English abstract)

[17] 李丹，张凯锋，李行健，等. 基于 Mask R-CNN 的猪只爬跨行为识别[J]. 农业机械学报，2019，50(S1)：261-266，275.
LI Dan, ZHANG Kaifeng, LI Xingjian, et al. Mounting behavior recognition for pigs based on mask R-CNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(S1): 261-266, 275. (in Chinese with English abstract)

[18] TU S, LIU H, LI J, et al. Instance segmentation based on Mask Scoring R-CNN for group-housed pigs[C]// International Conference on Computer Engineering and Application (ICCEA),Guangzhou, China, 2020: 458-462.

[19] WITTE J H, GERBERDIN J, MELCHING C, et al. Evaluation of deep learning instance segmentation models for pig precision livestock farming[J]. Business Information Systems, 2021, 1: 209–220.

[20] 刘坤，杨怀卿，杨华，等. 基于循环残差注意力的群养猪实例分割[J]. 华南农业大学学报，2020，41(6)：169-178.
LIU Kun, YANG Huaiqing, YANG Hua, et al. Instance segmentation of group-housed pigs based on recurrent residual attention[J]. Journal of South China Agricultural University, 2020, 41(6): 169-178. (in Chinese with English abstract)

[21] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),Salt Lake City, UT, USA, 2018: 3141-3149.

[22] HUANG Z, WANG X, HUANG L, et al. CCNet: Criss-Cross attention for semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on International Conference on Computer Vision (ICCV),South Korea, 2019: 603-612.

[23] ZHANG L, GRAY H, YE X, et al. Automatic individual pig detection and tracking in pig farms[J]. Sensors, 2019, 19(5): 1188-1208.

[24] 房俊龙，胡宇航，戴百生，等. 采用改进 CenterNet 模型检测群养生猪目标[J]. 农业工程学报，2021，37(16)：136-144.
FANG Junlong, HU Yuhang, DAI Baisheng, et al. Detection of group-housed pigs based on improved CenterNet model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(16): 136-144. (in Chinese with English abstract)

[25] WANG J,SUN K, CHENG T, et al. Deep high-resolution representation learning for Visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43: 3349-3364.

[26] CAI Z W, NUNO V. Cascade R-CNN: Delving into high quality object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018: 6154-6162.

[27] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), 2019: 1314-1324.

[28] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021: 13713-13722.

[29] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018: 7132-7141.

[30] 温长吉，王启锐，陈洪锐，等. 面向大规模多类别的病虫害识别模型[J]. 农业工程学报，2022，38(8)：169-177.
WEN Changji, WANG Qirui, CHEN Hongrui, et al. Model for the recognition of large-scale multi-class diseases and pests[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(8): 169-177. (in Chinese with English abstract)

Posture recognition of group-housed pigs using improved Cascade Mask R-CNN and cooperative attention mechanism

WANG Lu, LIU Qing, CAO Yue, HAO Xia※

(*School of Information Science and Engineering*, *Shandong Agricultural University*, *Taian* 271018, *China*)

Abstract: Pig posture recognition can greatly contribute to the early warning of pig health under the complex scenes in swine farms, even to reduce economic loss. However, it is still challenging in posture recognition, due to the mutual occlusion and adhesion of group-housed pigs in the group house. In this study, a two-stage group-housed pigs pose recognition was proposed using the combination of Instance Segmentation and Classification identification. Firstly, the instance segmentation data set was used as the input, Cascade Mask R-CNN was used as the reference network, and the HrNetV2 network was as the feature extraction network of Cascade Mask R-CNN. High-precision segmentation was carried out on the pig images in different scenes. The multi-resolution image processing in the HrNetV2 was realized to reduce the loss of spatial details of the images, in order to improve the representation of the segmentation target by the model. The FPN module was also introduced to construct the HrNetV2+FPN composite structure. The pig bodies of different sizes were mapped to the feature maps of different levels, and then to achieve the effective fusion of features of different scales. The overlapping and adhesion between pig bodies were reduced for the subsequent recognition of pig body posture. Secondly, the CA coordinate attention mechanism was introduced to design the CA-MobilenetV3 lightweight pig body pose recognition network. The pig individual pose was extracted to build the data set after the segmentation. The interference of the background region was reduced to enhance the feature learning in the key parts of the pig body. CAM visualization was used to display the extracted part of the pig body pose feature, in order to realize the accurate and rapid recognition of individual pig posture. Finally, the improved model was effectively segmented and detected the pig body, indicating the more accurate segmentation edge of the pig body. There was a rough segmentation of the conventional MS R-CNN and Mask R-CNN. By contrast, the improved Cascade Mask R-CNN model was achieved 95.8%, 95.8%, 85.2%, and 89.2% for the $AP_{0.50,}$ $AP_{0.75}$, $AP_{0.50:0.95}$, and $AP_{0.50:0.95\text{-large}}$, respectively. Therefore, the improved model performed better than the Mask R-CNN and MS R-CNN networks, in terms of detection and segmentation. In terms of pig posture recognition, the recall and F1 value of the CA-MobileNetV3 model increased by 10.8, 13.1 percentage points, respectively, compared with the rest. Furthermore, the performance of the CA-MobileNetV3 network was significantly improved to identify the sitting and lying posture class. All evaluation indexes achieved an accuracy rate of 96.9%, 99.1%, 99.5%, and 98.6%, respectively, in the kneeling, standing, lying, and sitting posture class. The performance was much better than the MobileNetV3, ResNet50, DenseNet121, and VGG16 networks of the same type. In conclusion, the improved model was superior to the current popular networks, in the aspects of pig individual segmentation and pig pose recognition. The non-contact and low-cost recognition can be expected to realize the pig attitude under different scenarios, such as deep separation, mutual adhesion, and debris shielding. The finding can also provide the model support for the practical management of pig-intensive farming.

Keywords: deep learning; image recognition; instance segmentation; group-housed pigs posture recognition; extraction of individual