

刘坤, 杨怀卿, 杨华, 等. 基于循环残差注意力的群养生猪实例分割 [J]. 华南农业大学学报, 2020, 41(6): 169-178.
LIU Kun, YANG Huaqing, 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.

基于循环残差注意力的群养生猪实例分割

刘 坤¹, 杨怀卿¹, 杨 华¹, 胡志伟¹, 孟 珂²

(1 山西农业大学 信息科学与工程学院, 山西 太谷 030801; 2 吕梁学院 城乡规划系, 山西 吕梁 033001)

摘要:【目的】在群养环境下, 实现生猪粘连、杂物遮挡等不同条件下生猪个体的高精度分割。【方法】对真实养殖场景下的 8 栏日龄 20~105 d 共 45 头群养生猪进行研究, 以移动相机拍摄图像为数据源, 并执行改变亮度、加入高斯噪声等数据增强操作获取标注图片 3 834 张。探究基于 2 个骨干网络 ResNet50、ResNet101 与 2 个任务网络 Mask R-CNN、Cascade mask R-CNN 交叉结合的多种模型, 并将循环残差注意力 (RRA) 思想引入 2 个任务网络模型中, 在不显著增加计算量的前提下提升模型特征提取能力、提高分割精度。【结果】选用 Mask R-CNN-ResNet50 比 Cascade mask R-CNN-ResNet50 在 AP_{0.5}、AP_{0.75}、AP_{0.5-0.95} 和 AP_{0.5-0.95-large} 指标上分别提升 4.3%、3.5%、2.2% 和 2.2%; 加入不同数量的 RRA 模块以探究其对各个任务模型预测性能影响, 试验表明加入 2 个 RRA 模块后对各个任务模型的提升效果最为明显。【结论】加入 2 个 RRA 模块的 Mask R-CNN-ResNet50 模型可以更精确、有效地对不同场景群养生猪进行分割, 为后续生猪身份识别与行为分析提供模型支撑。

关键词: 循环残差注意力; 实例分割; 图像处理; Mask R-CNN; Cascade mask R-CNN
中图分类号: S828 **文献标志码:** A **文章编号:** 1001-411X(2020)06-0169-10

Instance segmentation of group-housed pigs based on recurrent residual attention

LIU Kun¹, YANG Huaqing¹, YANG Hua¹, HU Zhiwei¹, MENG Ke²

(1 College of Information Science and Engineering, Shanxi Agricultural University, Taigu 030801, China; 2 Urban and Rural Planning of Architecture Department, Lüliang University, Lüliang 033001, China)

Abstract: 【Objective】To realize high-precision segmentation of individual pigs under different conditions such as pig adhesion and debris shielding in a group breeding environment. 【Method】A total of 45 group-housed pigs of 20 to 105 days from eight sheds in real farming scenes were studied. Mobile camera images were used as data sources, and data enhancement operations such as changing brightness and adding Gaussian noise were performed to obtain 3 834 annotated pictures. We explored multiple models with the cross-combinations of two backbone networks ResNet50, ResNet101 and two mission networks Mask R-CNN, Cascade mask R-CNN. We also introduced the idea of recurrent residual attention (RRA) into the two major task network models to improve the feature extraction ability and segmentation accuracy of the model without significantly increasing the amount of calculation. 【Result】Compared with Cascade mask R-CNN-ResNet50, Mask R-CNN-ResNet50

improved $AP_{0.5}$, $AP_{0.75}$, $AP_{0.5-0.95}$ and $AP_{0.5-0.95-large}$ by 4.3%, 3.5%, 2.2% and 2.2% respectively. Different numbers of RRA modules were added to explore the impact on the prediction performance of each task model. The experiment showed that adding two RRA modules had the most obvious improvement effect on each task model. 【Conclusion】 The Mask R-CNN-ResNet50 model with two RRA modules can more accurately and effectively segment group-housed pigs under different scenes. The results can provide a model support for subsequent identification and behavior analysis of live pigs.

Key words: recurrent residual attention; instance segmentation; image processing; Mask R-CNN; Cascade mask R-CNN

随着生猪养殖规模的扩大, 养殖密度的不断增加, 生猪感染病患风险也大大增加, 对群养生猪健康状况监测, 直接关系到猪场的经济效益。猪只行为富含多种生物学信息, 对其进行精确快速检测可有助于发现生猪异常行为, 及时采取相应措施, 降低疾病发生率。基于机器视觉的图像分析技术可为群养生猪提供低成本、非接触式的监控方法。但生猪粘连、杂物遮挡等客观因素给猪只个体研究带来较大挑战, 对其进行有效分割是实现群猪视觉跟踪及监测的关键步骤。

基于机器视觉的多种方法在生猪个体分割领域已取得较多研究成果。传统分割方法主要分为静态分割^[1-2]与动态分割^[3-4]两大类。但其依赖于人工筛选的颜色、形状、纹理等形态学特征, 在真实养殖场景下, 关照强度、杂物遮挡、生猪粘连等因素给高质量图像获取带来较大挑战。而基于深度学习的卷积神经网络 (Convolutional neural network, CNN) 对图像特征具有强大的表征能力^[5-6], 其已在生猪姿态检测^[7-9]、目标检测^[10-12]、目标跟踪^[13-14]、行为识别^[15-16]、个体计数^[17]等领域取得较大进展。在生猪目标分割领域, 基于 CNN 的全卷积网络 (Fully convolutional network, FCN)^[18]通过反卷积层替代全连接层的方式, 能有效保留图像空间维度信息, 实现对其像素级分割。杨阿庆等^[19]以 Visual geometry group 16(VGG16)^[20]为基础网络, 构建深层抽象特征与浅层细节特征融合并执行上采样 8 倍跳跃式 FCN 结构对哺乳母猪图像进行分割; Yang 等^[21]利用哺育行为的时空信息, 通过 FCN 准确分割母猪, 并通过母猪的几何特性和仔猪长度动态计算乳房区域, 以自动识别哺育行为; Yang 等^[22]提出了一种将 FCN 和 Ostu 阈值相结合的分阶段方法, 从俯视图上分割母猪图像; Psota 等^[23]引入了一种新的数据集和方法, 用于对群养环境的生猪个体进行分割, 并使用 FCN 检测猪只的位置和方

向; 胡志伟等^[24]基于 VGG16 与 UNET^[25]相结合的全卷积神经网络模型对单只生猪个体进行分割。但上述基于 FCN 的分割方法只能将生猪个体与环境信息有效分离, 却无法分割出同一类物体中的不同个体, 不适用于分割群养猪中的粘连猪个体。实例分割主要用于分割同类物体中的不同个体, 即可将同一张图像中的生猪个体分离开来。由 He 等^[26]提出的 Mask R-CNN 可用于对图像进行高精度实例级分割, 并已在柑橘树枝^[27]、稻飞虱^[28]等个体领域得到应用, 在生猪个体研究领域, 李丹等^[29]基于 Mask R-CNN 构建猪只分割网络模型, 分别实现 94% 的分割准确率与 94.5% 的爬跨行为识别准确率; 高云等^[30]基于 Mask R-CNN 提出 PigNet 网络对粘连生猪个体进行实例级分割。但上述基于 Mask R-CNN 的方法只是将 Mask R-CNN 模型应用于生猪研究领域, 并未进行实质性模型结构调整, 而在生猪群养环境中, 粘连生猪个体在猪体、猪鼻等细节部位存在差异性, 若能使模型强化不同生猪区别性部位则可更好地提升分割边缘精度。注意力机制的出现可有效解决该问题, 其在处理信息时只关注部分利于实例分割的区域信息, 抑制次要信息以提升模型效果, 并已在开放领域的图像分类^[31]、图像分割^[32]、目标检测^[33]等任务中得到较大应用, 因此探讨将其应用于猪舍场景下群养生猪实例分割成为可能。

基于此, 本文提出一种基于深度卷积网络的非接触、低成本的群养生猪实例分割新方法, 该方法结合不同骨干网络特征提取结果与 Mask R-CNN、Cascade mask R-CNN 任务分割网络进行端到端训练, 将不同循环残差注意力 (Recurrent residual attention, RRA) 模块融入任务网络中, 在真实养殖场景下, 能实现对深度分离、高度粘连、杂物遮挡等不同条件生猪个体高精度分割, 以期对生猪身份识别与行为分析等提供模型支撑。

1 试验数据

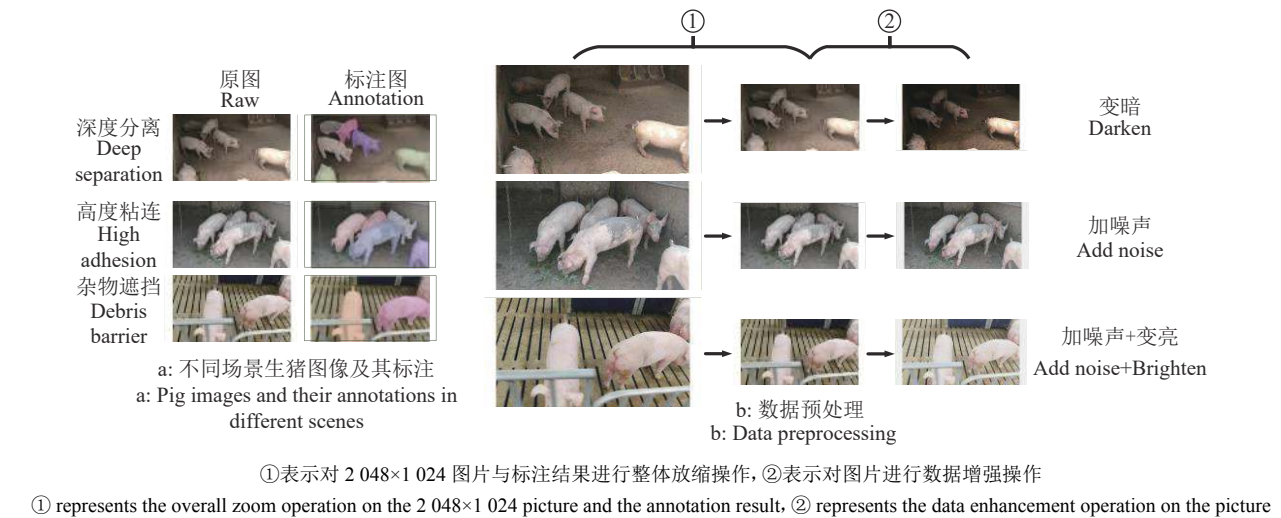
1.1 样本采集

数据采集自山西省汾阳市冀村镇东宋家庄村与山西农业大学实验动物管理中心,分别于2019年6月1日09:00—14:00(晴,光照强烈)与2019年10月13日10:30—12:00(多云,光照偏弱)进行2次采集。为检验模型对不同品种生猪个体的分割性能,选取大白、长白以及杜克混养猪作为拍摄对象,2个猪场猪栏大小分别为3.5 m×2.5 m×1.0 m和4.0 m×2.7 m×1.0 m,每栏生猪数量3~8只不等,选取其中8栏日龄20~105 d的群养生猪共计45头作为试验对象。与传统采用固定镜头俯视视角拍摄方式不同,采用平视方式进行数据采集更能有效获取到生猪的脸部、腿部等关键部位信息,这些部位更有利于生猪的精细分割,且此方式采集的数据更适用于移动端场景,为此本文采用镜

头不固定的平视视角数据采集方式,采用佳能700D防抖动摄像头进行拍摄,镜头距离生猪个体0.3~3.0 m不等,采集到不同大小及个体粘连、猪栏遮挡等不同场景数据,共拍摄得到多段时长为35 s~64 min的视频,每栏选取2段时长超过30 min的视频作为真实试验数据源以保证数据的连续性,对视频数据每隔25帧进行切割处理,得到1 920×1 080分辨率大小图片,为适应模型输入,对其按照宽高比2:1进行分辨率大小调整,对边缘采用白色像素值进行填充,最终分辨率大小调整为2 048×1 024,并采用LabelMe (<http://labelme.csail.mit.edu/Release3.0/>)进行试验数据标注。

1.2 试验数据预处理

对调整分辨率后的数据做如下预处理操作以获得群养生猪实例分割数据集,数据展示及其处理过程如图1所示:



① represents the overall zoom operation on the 2 048×1 024 picture and the annotation result, ② represents the data enhancement operation on the picture

图1 数据集展示及预处理

Fig. 1 Dataset display and preprocessing

1) 为减少模型运算量,降低训练显存占用率,对2 048×1 024图片与标注结果进行整体放缩操作,共得到1 917张512×256分辨率大小图片,如图1b中①所示。

2) 为丰富数据集,提升模型鲁棒性,对步骤1)得到的图片进行数据增强操作。仅执行改变亮度、加入高斯噪声2种增强操作,以50%的概率随机执行2种增强中的1~2种变换,每张图片生成1张增强图片,随机改变亮度值在0.8~1.2之间,大于1表示调暗,小于1表示调亮,其过程如图1b中②所示。

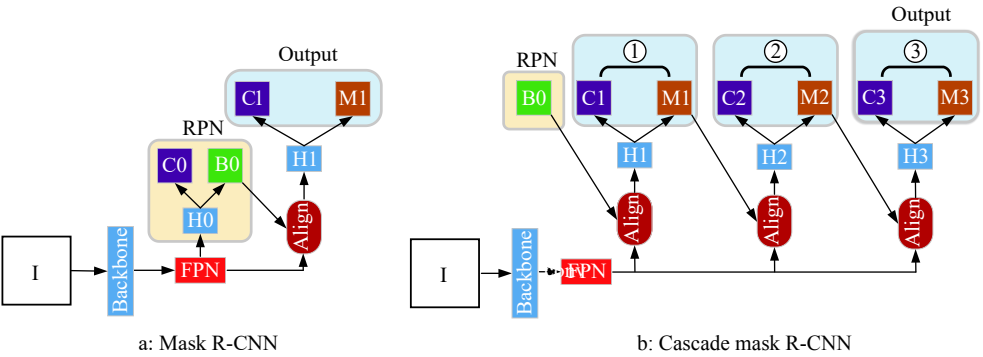
经上述处理后共获得标注图片3 834张,将其划分为训练集、验证集与测试集,其中训练集2 490

张,测试集864张,验证集480张。

2 生猪实例分割模型

2.1 Mask R-CNN 与 Cascade mask R-CNN 任务模型

Mask R-CNN是在Faster R-CNN^[34]基础上改进而来的,其由5部分构成,分别是特征提取部件(又称为骨干网络,如ResNet^[6]系列网络)、特征金字塔网络 (Feature pyramid network, FPN)、区域提交网络 (Region proposal network, RPN)、感兴趣区域对齐层 (Region of interest align, ROIAlign) 以及功能型输出,其模型结构如图2a所示,其中骨干网络主要用于对输入图像进行特征提取;FPN用于对骨干



I 表示输入图像, Backbone 表示骨干网络, Align 表示 ROIAlign, H 表示网络头部, C 表示分类结果, M 表示分割结果, ①②③分别表示 IOU 阈值为 0.5、0.6 和 0.7 的头部网络

I represents the input image, Backbone represents the backbone network, Align represents the ROIAlign, H represents the network header, C represents the classification result, M represents the segmentation result, ①②③ represent the head networks with IOU thresholds of 0.5, 0.6 and 0.7 respectively

图 2 Mask R-CNN 与 Cascade mask R-CNN 模型对比图

Fig. 2 Mask R-CNN and Cascade mask R-CNN model comparison chart

网络不同阶段输出的不同大小特征图进行金字塔信息融合, 以实现对大小差异较大的目标均能做到较好地分割; RPN^[34] 主要用于提取候选区域, 其能提取出可能感兴趣的信息但无法确定信息所属的真正类别, RPN 所提取的区域信息可用于对 FPN 特征图信息进行引导; 采用 ROIAlign 对 RPN 所截取的不同大小特征图进行尺度归一化, 以作为后续功能型部件的输入; 功能型部件由对目标个体进行分类的全连接网络以及对目标个体进行分割的全卷积网络组成。

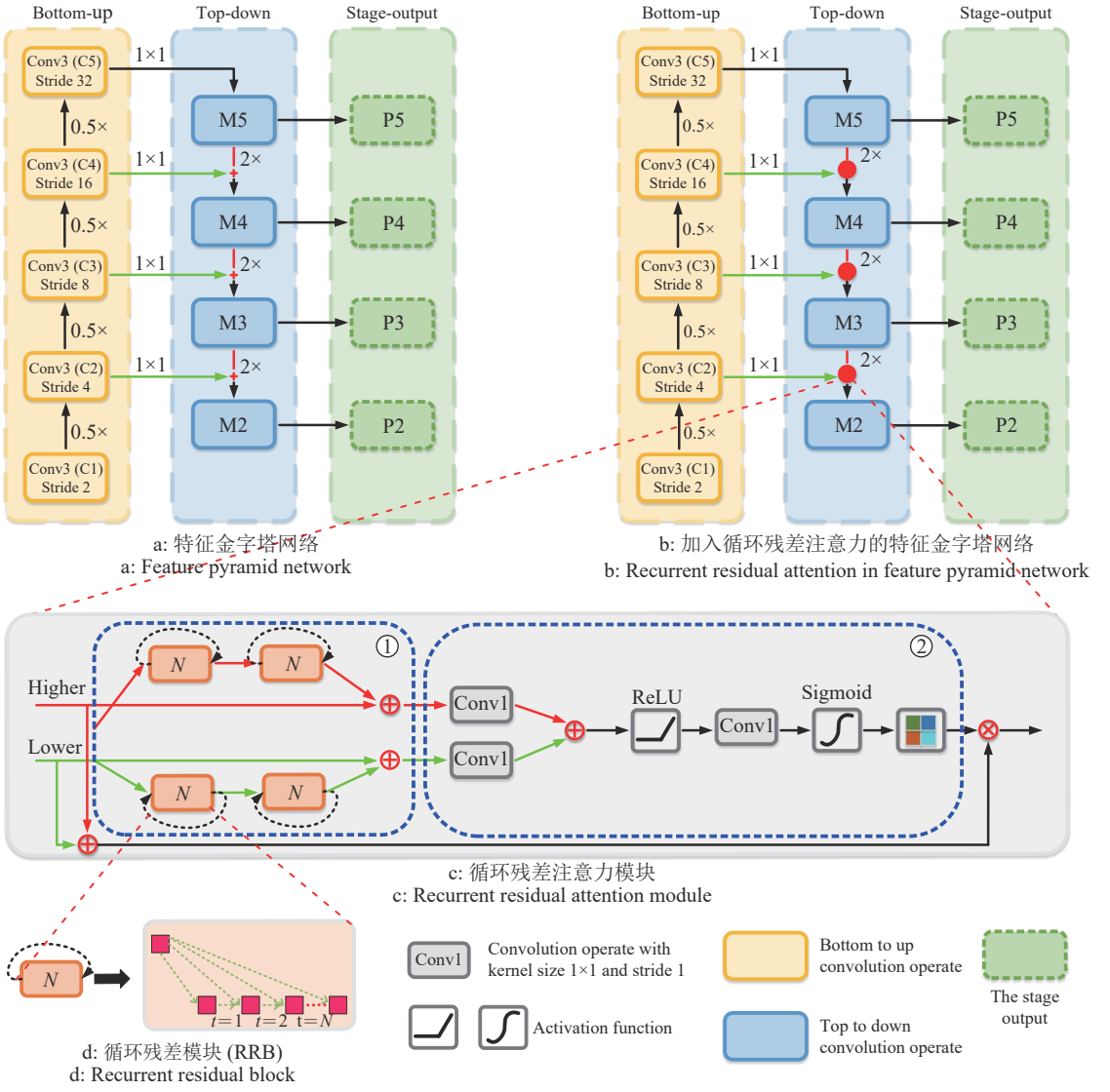
Cascade mask R-CNN 网络是由 Cai 等^[35] 提出, 其是在 Mask R-CNN 基础上改进得到的, 核心模型结构如图 2b 所示。与 Mask R-CNN 使用单一头部网络 (即图 2a 中的 H1) 进行预测不同的是, Cascade mask R-CNN 通过串行级联多个头部网络以达到不断优化预测的目的 (即图 2b 中的 H1、H2、H3), 前一个头部网络的输出作为后一个头部网络的输入, 且不同头部网络采用不同大小交并比 (Intersection over union, IOU)^[36] 阈值以确定预测结果中的正负样本, 位置越靠后的头部网络其采用的 IOU 阈值越大, 如图 2b 中①~③IOU 阈值分别对应为 0.5、0.6 与 0.7。Cascade mask R-CNN 由于采用多阈值头部网络的级联结构, 级联顺序按照阈值由小到大排列, 不断修正目标位置及其像素点区域, 避免了传统单一网络因为阈值设置误差所带来的性能损失问题。

2.2 循环残差注意力模块

常规卷积网络在特征提取过程中, 其低阶提取特征语义信息欠缺, 但目标位置准确, 高阶特征语义信息丰富, 但目标位置粗糙。FPN 引入图像金字塔思想, 能够考虑多种感受野信息, 进而提取不同

尺度特征, 对高低阶特征进行有效融合, 针对尺度较大或者较小的目标均有较强的识别效果, 其网络结构如图 3a 所示, 包括自底向上 (Bottom-up)、自顶向下 (Top-down) 以及阶段输出 (Stage-output) 三大组件, 对于 Bottom-up 组件, 其实际为骨干网络不同阶段的输出 (如 ResNet 网络), 经过每 1 次 0.5 倍 (0.5×) 下采样操作后, 特征图分辨率大小降低一半, 而对于 Top-down 组件, 每经过 2 倍 (2×) 上采样操作后, 特征图大小增为原来的 2 倍, 并将上述上采样结果与对应层级的 Bottom-up 组件进行线性相加, 得到相应层级的 Stage-output 输出。但传统 FPN 结构中不同尺度特征图间仅通过简单线性叠加的方式完成特征融合, 忽视了不同层级分支间的非线性关联关系, 注意力机制可提升利于任务相关的区域部位权重, 抑制次要信息, 进而提升模型效果, 将其引入 FPN 结构中的高低阶特征融合中, 势必将改善模型分割结果, 本文结合 FPN 与注意力思想, 提出加入 RRA 模块的 FPN, 其网络结构如图 3b 所示, RRA 模块如图 3c 所示。

RRA 模块由循环残差簇 (Recurrent residual cluster, RRC) 和注意力组件块 (Attention component block, ACB) 2 个部分构成, 分别如图 3c 中①②所示。RRC 簇中的详细操作如图 3d 所示, 对于输入特征, 其通过执行 3×3 大小卷积操作并将操作结果与输入特征进行残差融合形成循环残差模块 (Recurrent residual block, RRB), 进而捕获输入特征的长距离依赖信息, 为探究 RRC 内部 RRB 块数量对试验效果影响以寻找最适用于群养生猪实例分割的 RRC 簇, 构建包含有 N 个 RRB 块的 RRC 簇记为 RRC_N, 其内部为 N 个 RRB 块经过循环级联方式堆叠以实现多重长距离信息获取; ACB 块用于对 RRC 簇获取的高低阶特征融合进行注意力权重



M2~M5 分别表示第 2~5 阶段自顶向下组件卷积操作后的输出, P2~P5 分别表示第 2~5 阶段自底向上与自顶向下组件融合后的输出, N 表示循环的次数, Higher 与 Lower 表示高阶与低阶特征, ReLU 与 Sigmoid 表示激活函数, ①和②分别表示循环残差簇与注意力组件块

M2~M5 respectively represent the output after the top-down component convolution operation in the second to fifth stages, P2~P5 respectively represent the output after the bottom-up and top-down component fusion of the second to fifth stages, N represents the number of cycles, Higher and Lower represent high-order and low-order features respectively, ReLU and Sigmoid respectively represent activation functions, ① and ② represent recurrent residual cluster and attention component blocks, respectively

图3 传统 FPN 与加入 RRA 模块的 FPN 对比结构图

Fig. 3 Comparative structure diagram of traditional FPN and FPN added with RRA module

图计算, 其分别对高低阶特征执行 1×1 大小卷积操作并将各自操作结果进行线性叠加, 利用 ReLU 激活函数对叠加结果实现非线性化, 此外, 为进一步改善融合效果以扩大其注意力信息捕捉区域, 对 ReLU 非线性操作后的结果拼接 1×1 大小卷积运算, 并使用 Sigmoid 激活函数对其运算结果进行归一化以获取注意力结果图 Att, 并将 Att 与 RRA 模块输入的高低阶特征进行按位相乘操作, 以对模块输入特征进行深度筛选, 两大组件通过级联方式组合可用于捕捉语义与目标位置间的显著性依赖关系。上述 RRA 操作过程如式 (1)~式 (7) 所示。

$$F_{h'} = \text{RRC}_N(\text{RRC}_N(F_h)), \quad (1)$$

$$F_{l'} = \text{RRC}_N(\text{RRC}_N(F_l)), \quad (2)$$

$$F_{h''} = F_h + F_{h'}, \quad (3)$$

$$F_{l''} = F_l + F_{l'}, \quad (4)$$

$$\text{Att} = \sigma_2(\text{Conv}_{1 \times 1}(\sigma_1(\text{Conv}_{1 \times 1}(F_{h''}) + \text{Conv}_{1 \times 1}(F_{l''})))), \quad (5)$$

$$\text{RRC}_N = \begin{cases} \text{Conv}_{3 \times 3}(\text{RRC}_{N-1}) + F_{\text{rrc}}, & (N > 1) \\ \text{Conv}_{3 \times 3}(F_{\text{rrc}}) + F_{\text{rrc}}, & (N = 1) \end{cases}, \quad (6)$$

$$\text{Output} = (F_l + F_h) \text{Att}, \quad (7)$$

式中, F_h 与 F_l 分别表示 RRA 模块高阶与低阶特征输入, 分别对应于 FPN 结构中的同层 Top-down 与 Bottom-up 模块, $F_{h'}$ 与 $F_{l'}$ 分别表示经过 2 次 RRC 操

作后高低阶特征对应的输出, $F_{h''}$ 与 $F_{l''}$ 分别表示经过 RRA 模块输入与经过 RRC 簇处理后的输出结果间的融合特征, $\text{Conv}_{\text{size} \times \text{size}}(I)$ 表示对 I (例如 $F_{h''}$) 执行卷积核大小为 $\text{size} \times \text{size}$ (如 1×1 或者 3×3) 的卷积操作, 其中 I 可以是任意特征图, σ_1 与 σ_2 分别表示 ReLU 与 Sigmoid 激活函数, Att 表示经过 ACB 块处理后获取的注意力结果图, 将其与 RRA 模块输入进行按位相乘得到最终 RRA 模块输出结果 Output, F_{rrc} 表示 RRC 簇的特征输入。

3 试验平台参数及模型结果评价指标

3.1 试验平台参数

试验运行平台配置为 16 GB Tesla V100 GPU, 系统为 Ubuntu16.04, 采用 PyTorch 框架进行模型代码编写。1 个批次训练 8 张图片, 遍历完全部训练集数据称为 1 轮迭代, 试验中设置迭代轮数为 50。图像分辨率大小为 512×256 , 数据集增强前的大小为 1 917, 数据增强后总数据集大小为 3 834, 将其划分为训练集、验证集和测试集, 其中训练集大小为 2 490, 测试集大小为 864, 验证集大小为 480。采用随机梯度下降 (Stochastic gradient descent, SGD) 作为优化器, 优化器基础学习率超参数设置为 0.005~0.030 (“4.3”节会针对不同学习率大小进行对比试验), 为加速模型收敛避免其陷入局部最优解或者跳过最优解, 设置正则化权重衰减系数 (Weight decay) 为 0.000 1, 设置动量因子 (Momentum) 为 0.9。在图像输入网络前, 对 3 个通道分别进行归一化处理, 其均值分别为 123.675、116.280 和 103.530, 方差分别为 58.395、57.120 和 57.375。模型训练中仅对符合阈值约束条件的目标框内部生猪个体进行分割损失值计算, 对于 Mask R-CNN 任务网络, IOU 和 Score 阈值均设置为 0.5, 对于 Cascade mask R-CNN 任务网络, 参考自 Cai 等^[35] 关于 Cascade R-CNN 的论文, 其 3 个级联网络 IOU 阈值分别设置为 0.5、0.6 和 0.7、Score 阈值设置为 0.5。模型测试中采用 COCO(<http://mscoco.org/>) 指标进行模型优劣评价。

3.2 模型结果评价指标

模型采用实例分割领域公认的平均精度 (Average precision, AP) 作为评价标准, 以衡量模型对群养环境下生猪分割能力。AP 为精确率-召回率曲线下方面积, 其计算方法如式 (8)~(10) 所示。

$$P = \frac{TP}{TP + FP}, \tag{8}$$

$$R = \frac{TP}{TP + FN}, \tag{9}$$

$$AP = \int_0^1 PRdR, \tag{10}$$

式中, P 表示精确率, R 表示召回率, TP 表示模型预测为生猪个体区域且实际为生猪个体区域的样本数量, FP 表示模型预测为生猪个体区域但实际不为生猪个体区域的样本数, FN 表示预测为杂质背景区域但实际为生猪个体区域的样本数量。

考虑到选用不同的 IOU 阈值直接影响 TP 与假阳值 FP 给 AP 指标值带来波动, 与 COCO 指标定义相同, 选用 0.50、0.75、0.50~0.95 (其中 0.50~0.95 以 0.05 大小步长选取 10 个数) 3 种 IOU 阈值分别记为 $AP_{0.50}$ 、 $AP_{0.75}$ 、 $AP_{0.50-0.95}$ 以衡量不同条件下模型分割预测情况。同时考虑到本文生猪数据通过移动摄像头拍摄得到, 摄像头与生猪个体距离不一带来生猪个体占图像比重大小差异问题, 将生猪大小分为小目标 (生猪个体所占区域面积 $\leq 32^2$)、中等目标 ($32^2 <$ 生猪个体所占区域面积 $\leq 96^2$) 与大目标 (生猪个体所占区域面积 $> 96^2$), 并在 IOU 值为 0.50~0.95 间单独计算大目标下的 AP 指标值, 记为 $AP_{0.50-0.95\text{-large}}$ 。

4 结果与分析

4.1 不同骨干网络对任务网络的影响

骨干网络作为多种任务模型的共享结构, 常被用于特征提取操作, 为探究不同骨干网络对不同任务网络的群养生猪实例分割性能影响情况, 选取常见的 ResNet^[6] 衍生版本 (ResNet50 和 ResNet101) 作为骨干网络进行特征提取, 将提取结果分别作为 Mask R-CNN 与 Cascade mask R-CNN 任务网络输入, 并计算其在测试集上的各个分割 AP 指标值 (表 1)。

试验结果表明, 相同骨干网络对 Mask R-CNN 与 Cascade Mask R-CNN 任务网络群养生猪分割效果不一。以 ResNet50 作为骨干网络, 选用 Mask R-CNN 比 Cascade mask R-CNN 作为任务网络在 $AP_{0.50}$ 、 $AP_{0.75}$ 、 $AP_{0.50-0.95}$ 和 $AP_{0.50-0.95\text{-large}}$ 指标上分别提升 4.3%、3.5%、2.2% 和 2.2%。而将 ResNet101 作为骨干网络条件下, 选用 Cascade mask R-CNN 比 Mask R-CNN 在 $AP_{0.50}$ 、 $AP_{0.75}$ 、 $AP_{0.50-0.95}$ 和 $AP_{0.50-0.95\text{-large}}$ 指标上分别提升 17.4%、29.4%、17.3% 和 18.0%。可知, 在选用不同骨干网络情况下, Mask R-CNN 与 Cascade mask R-CNN 任务网络的性能状况呈现不同趋势, 虽从网络自身结构上 Cascade mask R-CNN 通过级联多种 IOU 阈值输出

表 1 不同骨干网络下不同任务网络分割的平均精度 (AP)¹⁾

Table 1 Average precision(AP) for different task networks under different backbone networks

模型 Model	骨干网络 Backbone network	AP _{0.50}	AP _{0.75}	AP _{0.50-0.95}	AP _{0.50-0.95-large}
Mask R-CNN	ResNet50	0.885	0.714	0.582	0.605
	ResNet101	0.660	0.348	0.368	0.384
Cascade mask R-CNN	ResNet50	0.842	0.679	0.560	0.583
	ResNet101	0.834	0.642	0.541	0.564

1)AP_x表示以x作为IOU阈值所对应的AP指标值
1)AP_x represents the AP index value using x as the IOU threshold

结果理论上可获得更佳效果,但在骨干网络较浅的情况下,其效果却弱于 Mask R-CNN。

在相同任务网络条件下,选用 ResNet50 作为骨干网络,其预测 AP 指标值均优于 ResNet101。以 Cascade mask R-CNN 为例,选用 ResNet50 作为骨干网络比 ResNet101 在各个 AP 指标上提升 0.8%~3.7%, ResNet50 骨干网络相较于 ResNet101 其深度更浅,理论上其特征提取能力弱于后者,但深层网络可能会带来梯度消失问题,导致反向传播过程中无法将梯度有效传回,且在本文群养生猪环境下,图像中仅存在生猪个体单类别信息以及猪栏、杂物等可控信息,选用较浅的 ResNet50 网络在一定程度上能够提取到合理的图像特征,此外,选用较浅的 ResNet50,其模型体积较小,训练也更为省时。

4.2 加入不同 RRA 模块对任务网络分割的影响

4.2.1 不同 RRA 模块对任务网络分割 AP 值的影响

为了能对群养环境中粘连生猪个体的猪体、猪鼻等细节部位差异性特征进行强化,以更好地提升分割边缘精度,有效将粘连生猪个体分离,避免光照、杂质等信息对分割效果的影响,本文构建了包含不同 RRB 块数的 RRA 循环残差注意力模块,将其嵌入 Mask R-CNN 与 Cascade mask R-CNN 两大任务网络中,并计算其在测试集上的分割 AP 指标值。表 2 表示不同骨干网络在引入不同数量循环残差注意力模块的任务网络上的分割 AP 指标值,其试验环境中选取的学习率大小为 0.02。

试验结果表明,在相同骨干网络条件下,加入 RRA 与未加入任何 RRA 模块的任务网络相比,其一般能提升分割预测 AP 值,即使在加入 1 个 RRA

表 2 引入不同循环残差注意力 (RRA) 模块任务网络分割的平均精度 (AP)¹⁾

Table 2 Average precision (AP) of task network segmentation with different recurrent residual attention (RRA) modules

模型 Model	骨干网络 Backbone network	RRA模块数量 No. of RRA modules	AP _{0.50}	AP _{0.75}	AP _{0.50-0.95}	AP _{0.50-0.95-large}
Mask R-CNN	ResNet50	0	0.885	0.714	0.582	0.605
		1	0.860	0.706	0.585	0.611
		2	0.892	0.769	0.636	0.660
		3	0.890	0.764	0.628	0.651
		4	0.887	0.752	0.618	0.642
	ResNet101	0	0.660	0.348	0.368	0.384
		1	0.646	0.364	0.350	0.365
		2	0.686	0.474	0.400	0.416
		3	0.684	0.461	0.396	0.409
		4	0.684	0.469	0.396	0.410
Cascade mask R-CNN	ResNet50	0	0.842	0.679	0.560	0.583
		1	0.835	0.700	0.579	0.603
		2	0.874	0.767	0.631	0.654
		3	0.873	0.749	0.623	0.647
		4	0.864	0.745	0.614	0.639
	ResNet101	0	0.834	0.642	0.541	0.564
		1	0.836	0.687	0.568	0.591
		2	0.885	0.772	0.634	0.660
		3	0.883	0.770	0.630	0.654
		4	0.873	0.756	0.617	0.641

1) AP_x表示以x作为IOU阈值所对应的AP指标值
1) AP_x represents the AP index value using x as the IOU threshold

模块后的部分试验模型下出现轻微的 AP 值降低现象 (如 Mask R-CNN-ResNet50 在加入 1 个 RRA 模块后, 其在 $AP_{0.50}$ 、 $AP_{0.75}$ 指标上与未加入 RRA 模块相比降低 2.5% 和 0.8%), 但其仍能产生有竞争力的分割结果。加入 2~4 个 RRA 模块的任务网络中, 在相同骨干网络条件下与未加入 RRA 模块的任务网络相比, 均能带来大幅度的 AP 值提升, 以 ResNet101 为骨干网络、Cascade mask R-CNN 为任务网络为例, 在加入 2~4 个 RRA 模块后, 与未加入 RRA 模块相比, 在 $AP_{0.50}$ 和 $AP_{0.75}$ 指标上分别提升 3.9%~5.1% 和 11.4%~13.0%, 证明 RRA 模块有利于群养环境下的生猪目标分割, 这是因为 RRA 模块在 FPN 中的高低阶特性融合中引入注意力信息, 对特征图中的每一点施以权重以对特征图信息进行重校准, 有效激活诸如猪鼻、猪耳等利于分割的部位信息, 抑制诸如猪粪、猪栏等杂质信息, 改善了生猪分割部位准确性及分割边缘完整性。

在相同骨干网络条件下, 加入不同 RRA 模块的任务网络其性能表现有较大差异。随着添加

RRA 模块中 RRC 簇中 RRB 块数的增加, 各个骨干网络条件下的任务网络 AP 指标值均出现先增加后减小的现象, 且在骨干网络固定时, 加入 2 个 RRA 模块的任务网络均取得最佳 AP 值。以 ResNet50 作为骨干网络为例, 将 2 个 RRA 模块加入 Cascade mask R-CNN 任务网络中, 其 $AP_{0.50}$ 、 $AP_{0.75}$ 、 $AP_{0.50-0.95}$ 和 $AP_{0.75}$ 指标值分别达到 0.874、0.767、0.631 和 0.654, 比加入 4 个 RRA 模块在各个指标上分别提升 1.0%、2.2%、1.7% 和 1.5%, 说明并非加入越深的 RRA 模块对生猪的分割越有效, 加入较浅的 RRA 模块 (如 1 个模块) 其循环残差注意力信息并未完全捕获所有有效猪体部位信息, 而加入过深的 RRA 模块极易使模型出现过拟合, 因此在实际融合 RRA 模块时以 2~3 层为宜。

4.2.2 不同 RRA 模块对任务网络分割结果可视化 为探究加入不同层级 RRA 模块的 Cascade mask R-CNN 与 Mask R-CNN 模型在测试集上的预测情况, 选用 ResNet50 作为骨干网络分别在 2 个任务网络上对部分分割结果进行可视化 (图 4)。

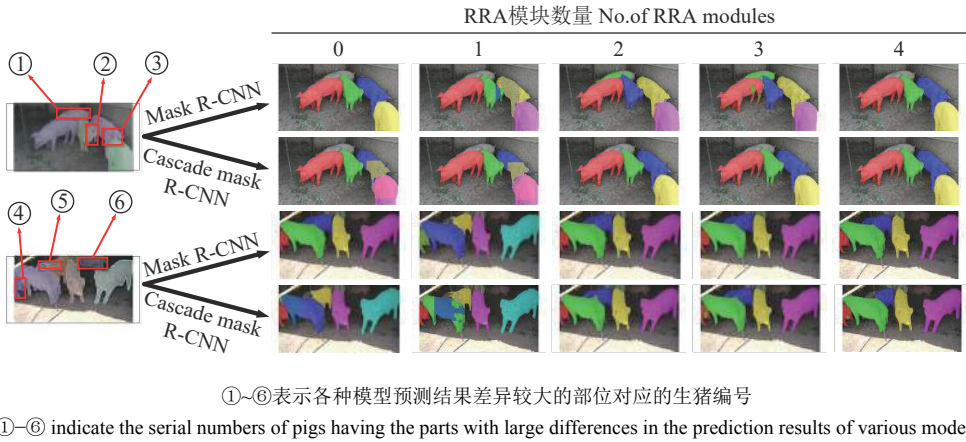


图 4 加入不同循环残差注意力 (RRA) 模块的 Cascade mask R-CNN 与 Mask R-CNN 模型测试集预测结果可视化展示

Fig. 4 Visual display of prediction results of Cascade mask R-CNN and Mask R-CNN model test sets with different recurrent residual attention(RRA) modules

如图 4 所示, 与未加入 RRA 模块相比, Cascade mask R-CNN 与 Mask R-CNN 模型在加入 RRA 模块后一般 (加入 1 个 RRA 模块的例外) 均能带来预测效果的提升, 尤其是加入 2 个 RRA 模块, 与相同条件下加入其他数量模块的模型相比均能取得最佳的分割效果。此外, 在加入相同模块条件下, 与 Cascade mask R-CNN 相比, Mask R-CNN 分割边缘更为精细、分割结果更为精确。对于偏离镜头的生猪个体, 如编号①的个体, 除 Mask R-CNN-ResNet50+2 个 RRA 模块与 Mask R-CNN-ResNet50+3 个 RRA 模块这 2 种模型外, 其余模型均未能对其进行

分割, 这可能是因为编号①的生猪个体距离镜头较远, 其脸部区域未在镜头捕捉范围内, 使得模型无法准确判断其是否属于生猪个体部位。而对于编号⑥的个体, 所有模型均未能对其进行有效分割, 这是因为编号⑥所处位置光线过于薄弱, 同时其头部位置处遮挡严重, 即使肉眼观看图片对其是否是生猪个体也会存在误判, 在本文后续研究工作中将着重解决这种多复杂条件下不存在脸部区域、光线昏暗无法辨识等场景生猪实例分割准确度提升问题。

4.3 不同场景下不同模型预测结果可视化 为进一步探究在任务网络中加入不同层级

RRA 模块的模型在不同场景下的鲁棒性能, 将生猪测试集分为深度分离、高度粘连、杂物遮挡 3 种场景, 采用表现性能最佳的模型 Mask R-CNN-ResNet50, 可视化展示在加入不同 RRA 模块后的分割效果如图 5 所示。

从图 5 中可知, 3 种场景下, 加入 2 个 RRA 模块的 Mask R-CNN-ResNet50 均能取得最佳的分割效果。对于深度分离场景中编号①的生猪个体, 其因为距离镜头较远、猪舍光照条件较差, 且被其他生猪个体部分遮挡, 导致除加入 2 个 RRA 模块的模型外其余模型均未能将其从群养环境中分离出来, 但加入 2 个 RRA 模块的模型仍能将其部分部

位分割出来, 这是因为合适层级的 RRA 模块能对有效生猪部位给予一定关注度, 提升猪体部位的分割权重, 将其与非猪体部位信息进行区别化对待, 提升了分割的精准度; 对于加入 2 个与 3 个 RRA 模块的模型在高度分离与杂物遮挡场景下的分割结果较为接近, 但加入 2 个 RRA 模块的模型分割边缘更为圆润, 如杂物遮挡场景下编号为⑤的生猪个体, 加入 2 个 RRA 模块后的模型对猪耳分割部位更为精细完整。考虑到上述不同场景中加入 2 个 RRA 模块的 Mask R-CNN-ResNet50 模型能够取得最好的分割性能, 实际中可将其作为群养生猪分割的参考模型。

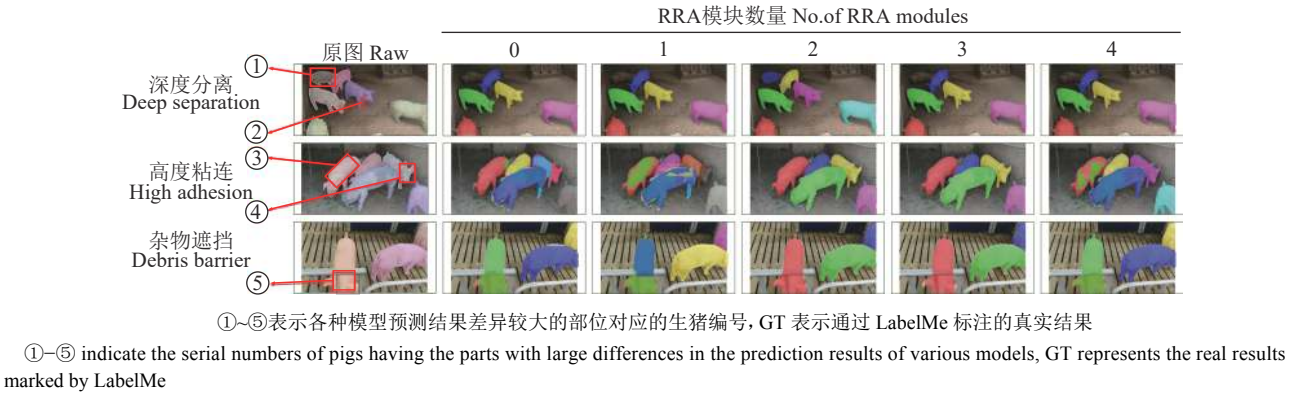


图 5 加入不同循环残差注意力 (RRA) 模块的 Mask R-CNN-ResNet50 模型对不同场景数据分割可视化展示

Fig. 5 Visualization of the segmentation of different scene data using the Mask R-CNN-ResNet50 model added with different recurrent residual attention(RRA) modules

5 结论

本文在 Mask R-CNN 与 Cascade mask R-CNN 两大任务网络中引入循环残差注意力信息, 研究不同骨干网络对任务网络、包含不同数量的循环残差块的注意力模块对分割性能的影响, 并对不同群养场景生猪分割结果进行可视化展示, 主要结论如下:

- 1) 不同骨干网络对 Mask R-CNN 与 Cascade mask R-CNN 等任务网络生猪分割效果具有一定影响。选用 ResNet50 为骨干网络, Mask R-CNN 为任务网络能够取得最佳分割结果。
- 2) 加入带有不同数量循环残差块的 RRA 模块对模型的提升幅度有所差异。加入 2 个 RRA 模块后对各个任务模型的提升效果最为明显, 且随着 RRA 内部循环残差块数量的增加, 任务模型各个预测分割 AP 指标值均出现先增加后减小的现象, 说明并非循环残差块数量越多, 其对应模型特征提取能力越强。
- 3) 对比加入不同数量循环残差块的 RRA 模块

后的 Mask R-CNN-ResNet50 模型, 其在深度分离、高度粘连、杂物遮挡等群养场景下预测可视化结果表明, 加入 2 个 RRA 模块后的分割效果最佳, 加入 1 个 RRA 模块的效果最差。

参考文献:

[1] 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.

[2] 朱伟兴, 纪滨, 秦锋. 基于伪球算子边缘模型的猪前景帧检测[J]. 农业工程学报, 2012, 28(12): 189-194.

[3] GUO Y, ZHU W, JIAO P, et al. Foreground detection of group-housed pigs based on the combination of mixture of Gaussians using prediction mechanism and threshold segmentation[J]. Biosyst Eng, 2014, 125(3): 98-104.

[4] 刘波, 朱伟兴, 霍冠英. 生猪轮廓红外与光学图像的融合算法[J]. 农业工程学报, 2013, 29(17): 113-120.

[5] KRIZHEVSKY A, SUTSKEVER I, HINTONG E. Imagenet classification with deep convolutional neural networks[J]. Commun ACM, 2017, 60(6): 84-90.

[6] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE. Proceedings of the IEEE conference on computer vision and pattern recognition.

- Las Vegas: IEEE, 2016: 770-778.
- [7] 燕红文, 刘振宇, 崔清亮, 等. 基于改进 Tiny-YOLO 模型的群养生猪脸部姿态检测[J]. 农业工程学报, 2019, 35(18): 169-179.
- [8] 薛月菊, 朱勋沐, 郑婵, 等. 基于改进 Faster R-CNN 识别深度视频图像哺乳母猪姿态[J]. 农业工程学报, 2018(9): 189-196.
- [9] NASIRAHMADI A, STURM B, EDWARDS S A, et al. Deep learning and machine vision approaches for posture detection of individual pigs[J]. Sensors, 2019, 19(17): 3738. doi: 10.3390/s19173738.
- [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(S1): 283-289.
- [11] 沈明霞, 太猛, CEDRIC Okinda, 等. 基于深层卷积神经网络的初生仔猪目标实时检测方法[J]. 农业机械学报, 2019, 50(8): 270-279.
- [12] HANSEN M, SMITH M L, SMITH L, et al. Towards on-farm pig face recognition using convolutional neural networks[J]. Comput Ind, 2018, 98: 145-152.
- [13] ZHANG L, GRAY H, YE X, et al. Automatic individual pig detection and tracking in pig farms[J]. Sensors, 2019, 19(5): 1188.
- [14] SUN L, ZOU Y, LI Y, et al. Multi target pigs tracking loss correction algorithm based on Faster R-CNN[J]. Int J Agr Biol Eng, 2018, 11(5): 192-197.
- [15] 薛月菊, 杨晓帆, 郑婵, 等. 基于隐马尔科夫模型的深度视频哺乳母猪高危动作识别[J]. 农业工程学报, 2019, 35(13): 184-190.
- [16] 高云, 陈斌, 廖慧敏, 等. 群养猪侵略性行为的深度学习识别方法[J]. 农业工程学报, 2019, 35(23): 192-200.
- [17] TIAN M, GUO H, CHEN H, et al. Automated pig counting using deep learning[J]. Electron. Agric. 2019, 163, 104840. doi: 10.1016/j.compag.2019.05.049.
- [18] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//IEEE. IEEE Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE, 2015: 3431-3440.
- [19] 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.
- [20] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. Int J Comput Vision, 2015, 115(3): 211-252.
- [21] 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]. Biosyst Eng, 2018, 175: 133-145.
- [22] YANG A, HUANG H, ZHENG C, et al. High-accuracy image segmentation for lactating sows using a fully convolutional network[J]. Biosyst Eng, 2018, 176: 36-47.
- [23] PSOTAE T, MITTEK M, PEREZ L C, et al. Multi-pig part detection and association with a fully-convolutional network[J]. Sensors, 2019, 19(4): 852. doi: 10.3390/s19040852.
- [24] 胡志伟, 杨华, 娄甜田, 等. 基于全卷积网络的生猪轮廓提取[J]. 华南农业大学学报, 2018, 39(6): 111-119.
- [25] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//MICCAI. International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: MICCAI, 2015: 234-241.
- [26] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]//IEEE. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE 2017: 2980-2988.
- [27] 林相泽, 朱赛华, 张俊媛, 等. 基于迁移学习和 Mask R-CNN 的稻飞虱图像分类方法[J]. 农业机械学报, 2019, 50(7): 201-207.
- [28] 杨长辉, 王卓, 熊龙焯, 等. 基于 Mask R-CNN 的复杂背景下柑橘树枝干识别与重建[J]. 农业机械学报, 2019, 50(8): 22-30.
- [29] 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(增刊1): 261-266.
- [30] 高云, 郭继亮, 黎焯, 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.
- [31] WANG F, JIANG M, QIAN C, et al. Residual attention network for image classification[C]//IEEE. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 3156-3164.
- [32] YU C, WANG J, PENG C, et al. BiSeNet: Bilateral segmentation network for real-time semantic segmentation[C]//ECCV. Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer International Publishing, 2018: 325-341.
- [33] 徐诚极, 王晓峰, 杨亚东. Attention-YOLO: 引入注意力机制的 YOLO 检测算法[J]. 计算机工程与应用, 2019, 55(6): 13-23.
- [34] REN S Q, HE K M, GIRSHICK R. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(6): 1137-1149.
- [35] CAI Z W, NUNO V. Cascade R-CNN: Delving into high quality object detection.[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 6154-6162.
- [36] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 658-666.