doi:10.6041/j.issn.1000-1298.2023.10.025

基于弱监督数据集的猪只图像实例分割

王海燕[1,2]　江烨皓[3,4]　黎　煊[1,5]　马云龙[6]　刘小磊[1,2]

(1.华中农业大学深圳营养与健康研究院,深圳 518000;2.中国农业科学院深圳农业基因组研究所,深圳 518000;
3.华中农业大学信息学院,武汉 430070;4.岭南现代农业科学与技术广东省实验室深圳分中心,深圳 518000;
5.农业农村部智慧养殖技术重点实验室,武汉 430070;6.农业动物遗传育种与繁殖教育部重点实验室,武汉 430070)

摘要:在智慧养殖研究中,基于深度学习的猪只图像实例分割方法,是猪只个体识别、体重估测、行为识别等下游任务的关键。为解决模型训练需要大量的逐像素标注图像,以及大量的人力和时间成本的问题,采用弱监督猪只分割策略,制作弱监督数据集,提出一种新的特征提取骨干网络 RdsiNet:首先在 ResNet-50 残差模块基础上引入第 2 代可变形卷积,扩大网络感受野;其次,使用空间注意力机制,强化网络对重要特征的权重值;最后引入 involution 算子,借助其空间特异性和通道共享性,实现加强深层空间信息、将特征映射同语义信息连接的功能。通过消融实验和对比实验证明了 RdsiNet 对于弱监督数据集的有效性,实验结果表明其在 Mask R-CNN 模型下分割的 mAP_{Semg} 达到 88.6%,高于 ResNet-50、GCNet 等一系列骨干网络;在 BoxInst 模型下 mAP_{Semg} 达到 95.2%,同样高于 ResNet-50 骨干网络的 76.7%。而在分割图像对比中,使用 RdsiNet 骨干网络的分割模型同样具有更好的分割效果:在图像中猪只堆叠情况下,能更好地分辨猪只个体;使用 BoxInst 训练的模型,测试图像中掩码具有更高的精细度,这更有利于开展下游分析。

关键词:猪只;弱监督实例分割;空间注意力机制;involution 算子

中图分类号:S828;TP391.4　　**文献标识码**:A　　**文章编号**:1000-1298(2023)10-0255-11　　**OSID**:

Pig Image Instance Segmentation Based on Weakly Supervised Dataset

WANG Haiyan[1,2]　JIANG Yehao[3,4]　LI Xuan[1,5]　MA Yunlong[6]　LIU Xiaolei[1,2]

(1. *Shenzhen Institute of Nutrition and Health*, *Huazhong Agricultural University*, *Shenzhen* 518000, *China*
2. *Agricultural Genomics Institute at Shenzhen* (*AGIS*), *Chinese Academy of Agricultural Sciences*, *Shenzhen* 518000, *China*
3. *College of Informatics*, *Huazhong Agricultural University*, *Wuhan* 430070, *China*
4. *Shenzhen Branch of Guangdong Laboratory of Lingnan Modern Agricultural Science and Technology*, *Shenzhen* 518000, *China*
5. *Key Laboratory of Smart Farming for Agricultural Animals*, *Ministry of Agriculture and Rural Affairs*, *Wuhan* 430070, *China*
6. *Key Laboratory of Agricultural Animal Genetics*, *Breeding and Reproduction*, *Ministry of Education*, *Wuhan* 430070, *China*)

Abstract: In smart livestock farming research, deep learning-based method for pig image instance segmentation is crucial for downstream tasks such as individual pig recognition, weight estimation, and behavior recognition. However, the model often requires a large number of pixel-wise annotated images for training, which imposes significant manpower and time costs. To address this issue, a weakly supervised pig segmentation strategy was proposed, creating a weakly supervised dataset, and introducing afeature extraction backbone network called RdsiNet. Firstly, the second-generation deformable convolution was incorporated into the ResNet-50 residual module to expand the network's receptive field. Secondly, spatial attention mechanisms were used to strengthen the network's weight values for important features. Finally, the involution operator was introduced to enhance deep spatial information and connect feature maps with semantic information by using its spatial specificity and channel sharing mechanism. The efficacy of RdsiNet for weakly supervised datasets was demonstrated through ablation experiments and comparative experiments. The experiments showed that the mean value of mask AP under the Mask R-CNN reached 88.6%, which was higher than a series of backbone networks such as

收稿日期:2023-03-12　修回日期:2023-06-16

基金项目:国家重点研发计划项目(2022YFD1601903)、湖北省科技重大专项(2022ABA002)、华中农业大学-中国农业科学院深圳农业基因组研究所合作基金项目(SZYJY2022034)和中央高校基本科研业务费专项资金项目(2662022XXYJ009)

作者简介:王海燕(1979—),女,副教授,主要从事动物表型组和智慧养殖研究,E-mail:wanghaiyan@webmail.hzau.edu.cn

ResNet－50 and GCNet. Meanwhile, the mean value of mask AP under the BoxInst reached 95.2%, which was also higher than that of ResNet－50 which reached only 76.7%. Furthermore, the display of image segmentation results of the test set showd RdsiNet also had better segmentation effect than ResNet－50. In the case of pig stacking, RdsiNet can better distinguish each pig. When using the BoxInst for training, RdsiNet can perfectly segment the outline of pigs, which was more conducive to downstream analysis.

Key words: pig; weakly supervised instance segmentation; spatial attention mechanisms; involution operator

0 引言

随着生猪养殖规模增大,现代化养殖技术对其帮助越发重要。利用人工智能技术丰富我国智慧农场解决方案,研发生猪养殖过程中的猪只信息智能感知、个体精准饲喂、养殖环境智能调控等核心技术与装备,正成为推动我国生猪养殖业健康发展的关键因素[1－3]。近年来,深度学习的兴起不断推动计算机视觉技术发展,研究者将深度学习引入到猪场猪只个体识别跟踪、姿态行为分类及体尺体重测量等任务中,取得了令人满意的效果[4－13]。

在猪只计数、行为识别、体重体尺测量等任务中,首要任务都是将猪只从图像中分割出来。目前,以深度学习为基础的图像实例分割正逐渐取代传统的机器学习前景背景分离算法,被应用到多数研究中。李丹等[14]通过训练神经网络模型,分割得到图像中猪只的面积以识别猪只爬跨行为;胡云鸽等[15]通过人工标注 1 900 幅图像制作数据集,在 Mask R－CNN[16]中的特征金字塔[17](Feature pyramid network, FPN)模块,使用轮廓边缘特征连接高层特征,极大提升了猪只边缘模糊目标识别的效果,并且能够满足单栏饲养密度为 1.03～1.32 头/m^2的养殖场的猪只盘点需求。上述研究证明,图像实例分割在智能化养殖产业所起的作用越发重要。

由于需要对图像的深层语义信息进行提取并预测,因此实例分割不仅需要大量的图像用于神经网络训练,还需要训练样本拥有像素级别的掩码信息(需要进行精细的标注)。而在猪只图像实例分割任务中,制作一个强监督(像素级标注)的数据集相当耗费人力,特别是图像中猪只个数多、产生堆叠、光照、噪声等因素影响,都会对精细标注效率产生影响[18]。因此,摆脱对高质量数据集的需求,正在成为分割领域研究的重点工作之一。当前,已有研究人员提出弱监督学习的概念[19],通过使用弱监督数据集,即采取粗糙标注的方式制作的数据集,通过改变神经网络对特征信息的处理模式,减少图像实例分割对像素级信息的过分依赖。国内现在已有针对农业领域使用弱监督学习方法的研究,赵亚楠等[20]提出基于边界框掩码的深度卷积神经网络,通过引入伪标签生产模块,用低成本的弱标签实现玉米植株图像实例分割;黄亮等[21]结合 RGB 波段最大差异法,实现对无人遥感灯盏花的弱监督实例分割。上述研究方案在节约数据集标注成本的同时,还取得了较高的精度,这也证明了弱监督图像实例分割在猪只养殖等智能化农业领域具有很大的研究和应用价值。

为了解决猪只图像实例分割中制作强监督数据集耗时耗力的问题,本文使用粗糙标注的方法构建弱监督数据集;从优化图像特征提取和处理过程,以此提升弱监督实例分割效果的角度出发,结合第 2 代可变形卷积、空间注意力机制和 involution 算子,提出新的特征提取骨干网络 RdsiNet;通过使用 Mask R－CNN 分割模型进行训练,以验证 RdsiNet 网络改进的有效性;最后使用仅需边界框作为监督信息的 BoxInst[22]弱监督实例分割模型训练数据集,以本文的 RdsiNet 作为特征提取骨干网络,在进一步验证 RdsiNet 有效性的同时,提升猪只的分割效果。

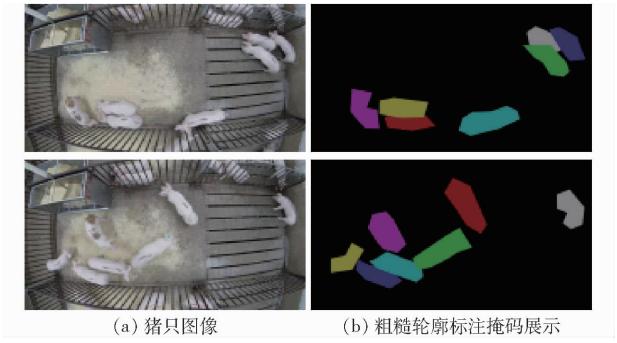
1 弱监督数据集构建及分析

1.1 数据集构建

弱监督实例分割(Weakly supervised instance segmentation)是一种使用较少的监督信息进行训练的实例分割方法。通常只需要图像级别的标签,而不需要每个像素的精确标注,根据标注方式的不同可以细分为无监督、粗监督、不完全监督等类型[23],分别对应无标注、粗糙标注和部分标注的数据集制作方法。考虑到猪舍猪只不断运动的特性,其分帧后得到的图像会带有猪只的行为信息,不同图像中同一猪只的空间信息对于实例分割神经网络模型有着重要的意义。因此,为了能为神经网络模型提供更有效的特征区域和空间信息,同时减少每幅图像的标注时间,本文采取粗糙的轮廓标注框作为数据集的标注方式。

LU 等[24]针对猪只图像分割研究,制作了一个规模较大的数据集(包括训练集 15 184 幅图像,验

证集 1 898 幅图像，测试集 1 900 幅图像）；该数据集图像由公开的猪场监控视频分帧而成[25]，本文对其进行筛选，选取其中 10 ~ 18 周龄且处于同一场景下的 7 头猪只的监控视频图像，共选出 17 980 幅猪只图像作为本文研究的原始数据。其后本文使用 Labelme 软件，对此原始数据所有图像进行基于弱监督的粗糙的轮廓标注（共标注 17 980 幅）。像素级标注要求标注框紧密贴合猪只身体轮廓，并且给不同的猪只打上专属的编号，每幅图像耗时约 10 min。图 1 为本文采用的粗糙标注方式的标注效果，和逐像素方式相比，标注框不再呈现猪只背部的几何结构，而是以图 1b 所示的多边形直接覆盖猪只，每幅图像只需 2 min 就可以完成标注，比起逐像素方式工作效率提高了 5 倍，大大节约了标注时间成本。最后在进行神经网络模型训练之前将所有标注图像进行训练集、验证集、测试集的划分，划分比例为 8∶1∶1，共得到训练集 14 384 幅图像、验证集 1 798 幅图像和测试集 1 798 幅图像。



(a) 猪只图像　　(b) 粗糙轮廓标注掩码展示

图 1　粗糙轮廓标注样式展示

Fig. 1　Display of rough contours annotation style

1.2　数据集分析

尽管粗糙的轮廓标注可以极大地节约数据集制作时间成本，但是其提供的低质量的真值标签，在实例分割神经网络训练过程中，会造成网络学习性能的下降。尤其在神经网络反向传播的过程中，一方面标注框同时包含分割实例和背景信息，会导致某些权重的梯度异常或是下降方向错乱，造成梯度稀疏、混淆等问题[25]；另一方面，逐像素的标注框能为神经网络提供更多特征信息，而粗糙的标注框却无法做到，甚至会提供错误的特征信息，最终影响训练结果[26-28]。

2　猪只实例分割模型改进

2.1　特征提取主干网络改进

2.1.1　引入第 2 代可变形卷积

因为粗监督数据集的标注框不贴合猪只，这会导致标注框内同时包含背景像素值和猪只的像素值，且这两种像素值差距较大，在神经网络反向传播过程中，会影响网络对猪只边缘信息的优化过程。为解决此问题，本文从特征提取角度出发引入可变形卷积，在特征提取过程中将更多的背景像素加入特征图中，扩大网络感受野。第 1 代可变形卷积由 DAI 等[29]提出，通过在传统卷积操作中引入偏移量概念，将传统卷积核由固定结构变为发散性结构，从而扩大特征提取的感受野，其特征值计算公式为

$$y(\boldsymbol{p}_0)=\sum_{\boldsymbol{p}_n\in\boldsymbol{R}}w(\boldsymbol{p}_n)x(\boldsymbol{p}_0+\boldsymbol{p}_n+\Delta\boldsymbol{p}_n)\tag{1}$$

式中　$\boldsymbol{p}_0$——特征图中进行卷积的采样点

$y(\boldsymbol{p}_0)$——卷积输出的特征值

$\boldsymbol{p}_n$——采样点在卷积核范围内的偏移量

$w(\boldsymbol{p}_n)$——卷积核权重

$x(\boldsymbol{p}_0+\boldsymbol{p}_n+\Delta\boldsymbol{p}_n)$——加上偏移量后采样位置的特征值

$\boldsymbol{R}$——卷积核感受野区域

尽管通过网络学习偏移量可以增大骨干网络的感受野，但网络同时也会通过可变形卷积学习许多无关信息，造成混乱。ZHU 等[30]在第 1 代的基础上，提出了第 2 代可变形卷积操作，通过增加一个权重系数 Δm_{p_n}，增大网络对于卷积操作的自由度，可以在学习中弱化或舍弃某些无关采样点权重，计算公式为

$$y(\boldsymbol{p}_0)=\sum_{\boldsymbol{p}_n\in\boldsymbol{R}}w(\boldsymbol{p}_n)x(\boldsymbol{p}_0+\boldsymbol{p}_n+\Delta\boldsymbol{p}_n)\Delta m_{p_n}\tag{2}$$

在神经网络学习的过程中，通过对 Δm_{p_n} 进行赋值，可以对学习到的特征值进行区分，将不需要的特征值舍去。

文献[31-32]为了解决传统卷积感受野不够导致对图像复杂信息提取能力差的问题，通过引入第 2 代可变形卷积操作，使得网络感受野和图像特征建立变化性关系，使其可以自适应地融合每个像素点相邻的相似结构信息，进而提高检测的准确率。因此，本文在骨干网络中使用第 2 代可变形卷积，可以使特征图包含更多背景信息，将网络感受野扩大以匹配粗监督标注框，减少错误信息带来的影响，网络通过不断地迭代和反向传播，可以提升最终分割效果。

2.1.2　空间注意力机制模块

空间注意力机制由 WOO 等[33]提出，是一种模仿人眼视觉的一种处理机制。在图像处理中，空间注意力机制通过生成权值矩阵的方式，对主干网络所提取的不同特征赋予不同的权重，以此在众多信息中选取关键的部分。如图 2 所示，输入尺寸为 $H\times W\times C$ 的特征图，通过最大池化和平均池化得到尺寸为 $H\times W\times 1$ 的两幅特征图，将这

两幅特征图按照通道维度拼接，然后再使用 7 × 7 的卷积核和 Sigmoid 函数，得到权重矩阵 $\boldsymbol{M}_s$，计算公式为

$$\boldsymbol{M}_s(F)=\sigma(f_{7\times7}([\text{AvgPool}(F),\text{MaxPool}(F)])) \tag{3}$$

式中 F——输入的初始特征图

$\boldsymbol{M}_s(F)$——空间注意力机制得到的权重矩阵

$\sigma(\,)$——Sigmoid 函数

AvgPool(　)——平均池化操作

MaxPool(　)——最大池化操作

将 $\boldsymbol{M}_s$ 与输入的特征图相乘，就为神经网络模型加入了空间注意力。

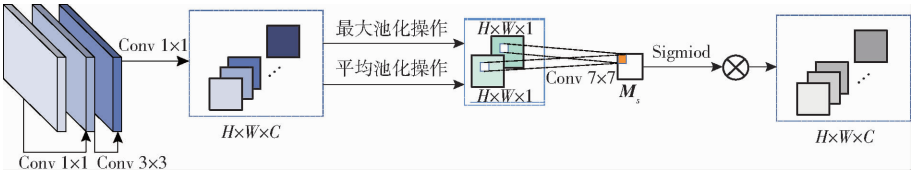


图 2　空间注意力机制

Fig. 2　Spatial attention mechanisms

俞利新等[34]针对特征图提取过程中冗余信息过多的问题，通过引入空间注意力机制以减弱源图像中的冗余信息从而突出目标，并通过消融实验验证了该方法的有效性。基于此，本文在骨干网络中加入空间注意力机制，用于对特征通道中不同特征映射赋予权重，将强有用的特征映射值如猪只轮廓、纹理、颜色等，平均到每个通道特征图中，扩大其在网络中的影响因子。

2.1.3　involution 算子

对于图像实例分割任务而言，核心思想在于对深层的抽象特征进行语义预测。但是随着神经网络层数的加深，骨干网络会失去大量的空间信息，导致网络区分不同实例能力不足。尤其对本文所使用的弱监督数据集而言，其中猪只聚集、移动等场景较多，对分割产生的挑战很大。基于此问题，本文在骨干网络中引入 LI 等[35]提出的 involution 算子，区别于传统的特征提取方式，它将空间各异性和通道共享性作为设计出发点，杨洪刚等[36]为提升神经网络模型对细粒度图像的能力，使用 involution 算子提取了图像的底层语义信息和空间结构信息进行了特征融合，并验证了其有效性，其结构如图 3 所示。

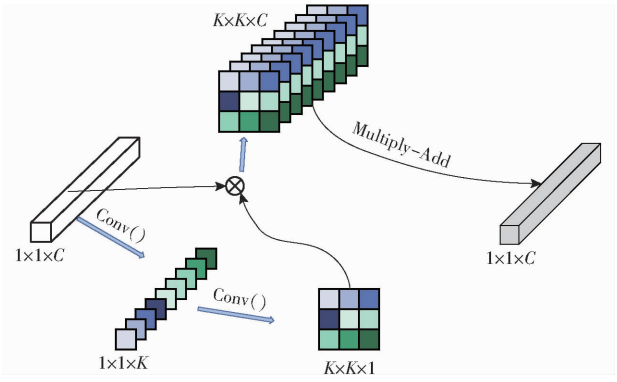


图 3　involution 算子提取特征模式图

Fig. 3　Feature pattern diagram extracted by involution operator

假设输入尺寸为 $H\times W\times C$ 的特征图，对 $1\times1\times C$ 的像素点的特征向量作下一步特征提取时，使用卷积操作先将其通道数 C 压缩至 K^2，再将获得的 K^2 个通道数作为新的大小为 K 的卷积核；其后将初始的 $1\times1\times C$ 特征向量在特征图中扩展至 $K\times K$ 大小的区域，与上一步中得到的卷积核相乘并相加，得到最终的结果。与卷积相比，involution 算子对于具体空间位置的卷积核由该位置的特征向量决定，并且对不同的输出通道使用相同的卷积核，具有了空间特异性和通道共享性。

本文通过使用 involution 算子，不仅可以解决深层网络空间信息丢失的问题，还可以将深层的语义信息和特征通道中被赋予空间注意力的信息连接，加强网络对于猪只图像分割的学习，提升分割的精度。

2.1.4　RdsiNet 特征提取骨干网络结构

本文提出的 RdsiNet 骨干网络结构如图 4 所示，其中蓝色虚线框中展示了本文通过对传统残差块加入空间注意力机制和 involution 算子后得到的残差-空间注意力机制模块和残差- involution 模块。参考 ResNet－50[37]中 3、4、6、3 层的残差模块分布概念，在 ResNet－50 的残差结构后加入空间注意力机制，提出残差-空间注意力模块，作为新的特征提取模块，并且在 Layer1 中串联使用 3 块。在 Layer 2 和 Layer 3 中，将第 2 代可变形卷积加入残差-空间注意力模块，替代原本 3 × 3 卷积操作，分别使用 4 块和 6 块；最后，将 ResNet－50 残差模块中的 3 × 3 卷积操作替换为 involution 算子，构建残差－involution 模块，在 Layer 4 中同样串联 3 个此模块。

2.2　实例分割网络模型

2.2.1　实验分割模型选择

为验证 RdsiNet 骨干网络的有效性，本文选取两种实例分割模型进行训练：需要像素级掩码标注

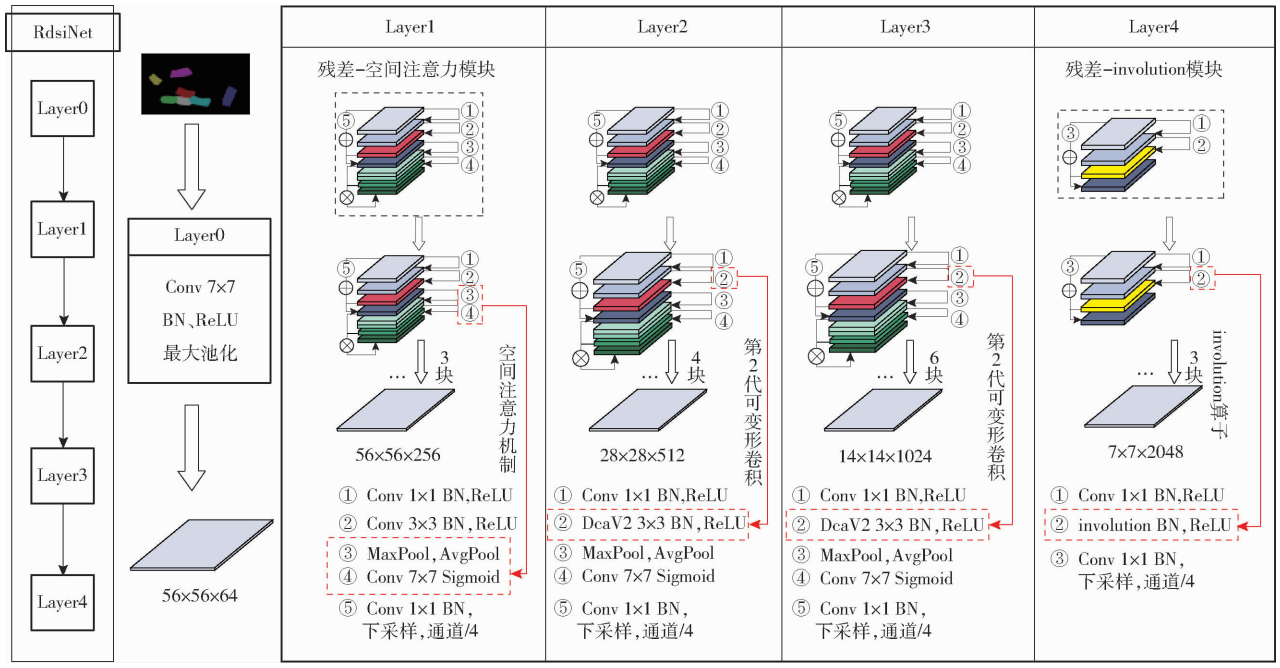


图 4　RdsiNet 骨干网络结构

Fig. 4　Structure of RdsiNet backbone network

进行训练的 Mask R－CNN；仅需要边界框标注进行训练的弱监督实例分割模型 BoxInst。

由于本文标注的掩码标签（Mask label）无法为 Mask R－CNN 提供准确、高质量的掩码监督信息（Mask ground truth），因此最终的实例分割效果不如逐像素标注的效果，但另一方面，这也更能反映不同骨干网络对于猪只图像的特征提取能力，因此本文使用 Mask R－CNN 以验证本文所提 RdsiNet 骨干网络的有效性。

2.2.2　Mask R－CNN 分割模型

图 5 展示了本文使用的 RdsiNet 的设计结构和 Mask R－CNN 实例分割模型的训练过程。图 5a 是第 2 代可变形卷积的实现过程，其发散性的特征提取方式扩大了网络的感受野，用于 Layer2 和 Layer3 层。图 5b 是猪只的轮廓纹理特征图像，图 5c 是 RdsiNet 网络特征提取过程中不同特征通道的图像展示结果，通过将轮廓纹理特征矩阵平均加至不同的特征通道内，实现增加实例分割模型对猪只轮廓的注意力，用于 Layer1、Layer2 和 Layer3 层。图 5d 展示了由特征图上某一点像素的不同特征通道所生成的卷积核，将其与该像素相乘并相加，实现特征通道和图像像素的交互，用于最后一层（Layer4 层）。

Mask R－CNN 是一种基于像素级掩码标注的全监督实例分割模型，其分割模型步骤如图 5 所示，通过对 RdsiNet 提取的特征图进行感兴趣区域（ROI）和 RoIAlign 操作，在特征图上生成感兴趣空间并将其与输入图像像素区域对齐，之后对空间内物体进行类别、边界框和掩码的预测及损失函数（包括 $Loss_{cls}$、$Loss_{box}$、$Loss_{mask}$）的反向传播，最终经不断迭代完成训练。

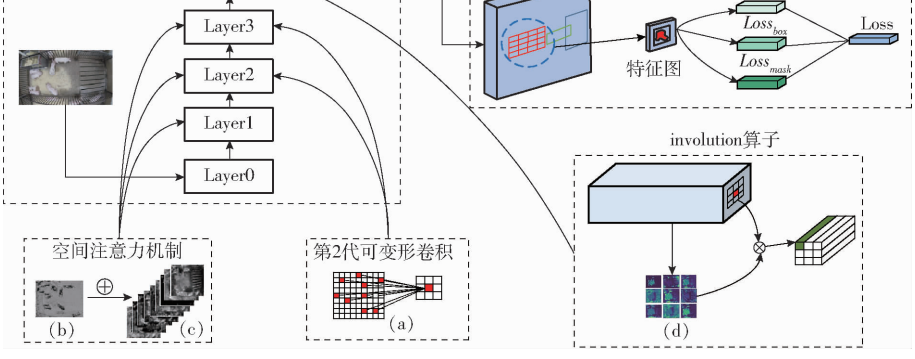


图 5　Mask R－CNN 分割模型训练框架

Fig. 5　Framework of Mask R－CNN

2.2.3 BoxInst 分割模型

BoxInst 是一种基于边界框标注(Bounding box)的弱监督实例分割模型,主要由骨干提取网络、FPN 层、共享 Head 层(Controller Head)和 Mask 预测分支(Mask Branch)组成,其仅需使用边界框的标注(Box label)作为监督信息去训练实例分割网络。本文使用 RdsiNet 作为 BoxInst 特征提取骨干网络,并通过 FPN 加强对不同尺度实例的学习能力。而对于掩码预测部分,这个过程由 2 个分支组成,分别为共享的 Head 层和 Mask 预测分支。共享 Head 层用来预测实例及其最小外接框,Mask 预测分支则用来对预测的外界框内所有像素进行前景背景预测,最终实现物体的分割。如图 6 所示,BoxInst 采用动态卷积的思想对每一个实例编码,通过共享 Head 层,对不同尺度特征图进行实例预测,获取每个实例的类别及动态生成其 Controller 参数;而在 Mask 预测分支中,将 FPN 层得到的特征图和每个实例的相对位置相加输出为总特征图,将共享 Head 层得到的每个实例参数分别作用在总特征图上以生成不同的掩码预测区域,并预测其边界框和掩码。在获取实例边界框后,一方面通过对边界框的左上角和右下角顶点坐标值进行反向传播,如图 6 所示,提升边界框的精准度;另一方面计算其内部所有像素之间的相似性,引入如图 6 所示的相邻像素颜色相似度(pairwise)属性关系进一步约束前景、背景像素,并使用 Lab 色彩空间下颜色的相似度作为真实标签,对不同的像素进行聚类,最终实现不依靠标注的掩码监督信息实现实例分割;其中 L_{proj} 表示边界框两个顶点坐标的损失值,而 $L_{pairwise}$ 表示掩码的损失值,其中边界框的损失值由数据集提供的边界框标注计算,而掩码损失值由模型迭代过程中通过学习到的像素间颜色关系计算得到。

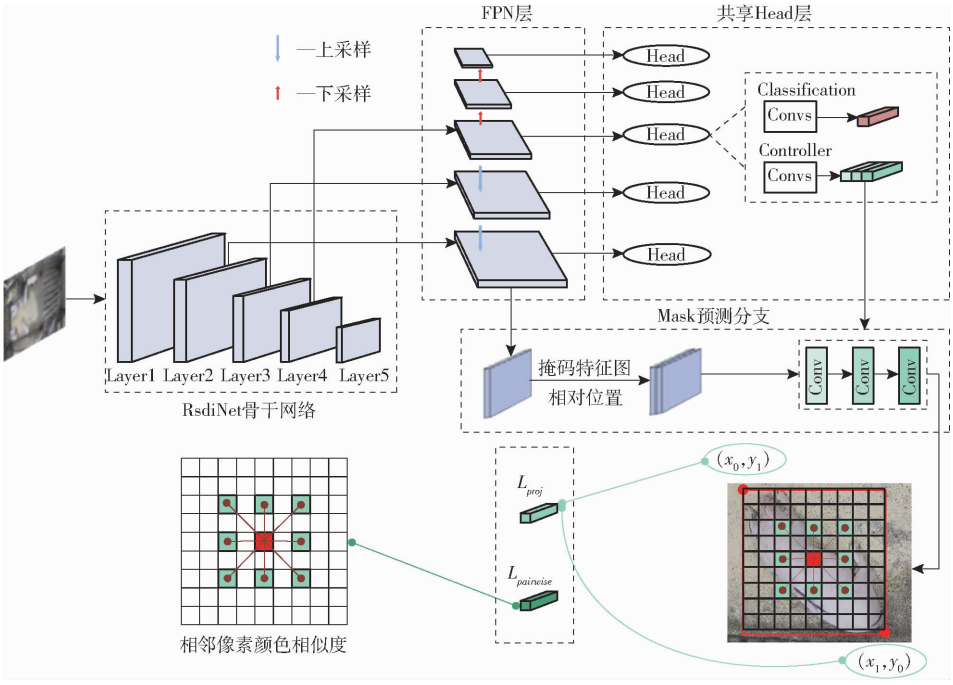


图 6 本文 BoxInst 分割模型训练框架

Fig. 6 Framework of BoxInst segmentation model

3 实验与结果分析

3.1 实验环境

本文基于 Mmdetection 框架进行实验,使用的核心计算显卡为 2 块 GeForce GTX 2080Ti,显存为 22 GB,显卡驱动 CUDA 版本为 10.1,Python 版本为 3.7.13,Pytorch 版本为 1.7.1,mmcv 版本为 1.5.0,mmdet 版本为 2.25.2。

3.2 实验参数设置

实验过程中模型训练轮数为 12 轮,学习率设为 0.001,采用 AdamW 优化器,权重衰减(weight_decay)设为 0.05。

3.3 模型训练损失值曲线

Loss 函数是评价模型性能的主要指标之一,其可以反映模型训练过程中的稳定性和衡量模型。图 7 为使用 RdsiNet 作为特征提取网络的 Mask R-CNN 和 BoxInst 的 loss 函数曲线,可以看出,随着训练轮数的增加,两种模型损失值都呈现平稳下降趋势,且曲线平滑,在迭代了 10 000 次后逐渐趋于收敛,这表明 RdsiNet 骨干提取网络设计合理,训练时间和成本可控,具有较强的鲁棒性。

3.4 模型训练结果

3.4.1 Mask R-CNN 训练结果

对于 Mask R-CNN 实例分割模型,本文分别使

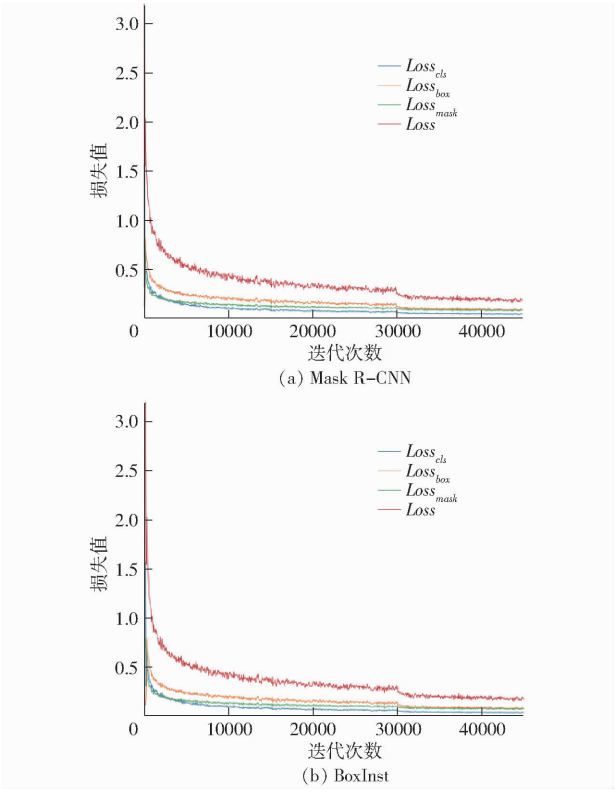


图 7　Mask R－CNN 和 BoxInst 训练 Loss 曲线

Fig. 7　Loss curve graphs of Mask R－CNN and BoxInst

用 ResNet－50、GCNet[38]、RegNet[39]、ResNeSt[40]、CotNet[41]和提出的 RdsiNet 骨干网络进行实验，实验结果如表 1 所示。在实验效果评估中，使用 mAP_{Bbox}和 mAP_{Semg}评价回归的边界框和猪只分割精度。

表 1　不同骨干网络训练结果对比

Tab. 1　Comparison of different backbone network training results

主干网络	参数量	mAP_{Bbox}/%	mAP_{Semg}/%
ResNet－50	3.490×10^7	88.8	83.1
GCNet	4.090×10^7	86.7	82.3
RegNet	3.417×10^7	88.4	82.6
ResNeSt	4.702×10^7	91.5	85.8
CotNet	4.356×10^7	92.7	87.6
RdsiNet	3.805×10^7	93.4	88.6

平均精度均值（mAP）指所有类的平均精度（AP）的平均值，用来衡量多类别目标检测效果。表 1 显示，本文改进后的骨干网络具有最高的 mAP_{Bbox}和 mAP_{Semg}值，分别为 93.4% 和 88.6%。同 GCNet、ResNeSt 和 CotNet 相比，以更少的参数获得了更好的实例分割效果，而对比 ResNet－50，在小幅提升参数量的情况下，mAP_{Bbox}和 mAP_{Semg}获得了较大的增益，体现了 RdsiNet 骨干网络的优越性。

3.4.2　Mask R－CNN 分割模型测试图像

为进一步验证 RdsiNet 的效果，本文分别使用参数量低于 4×10^7的 4 种骨干网络进行模型分割效果测试，图 8 为在猪只扎堆、粘连等条件下，ResNet－50、GCNet、RegNet 和 RdsiNet 骨干网络在 Mask R－CNN 分割模型下的图像测试效果。对比 4 种骨干网络下图像分割效果可以看出，ResNet－50、GCNet、RegNet 对于猪只聚集情况，均无法准确提取有效空间信息，以辅助分割模型判别猪只实例个数及空间位置，造成大量错检等问题；而本文所提出的 RdsiNet 网络，明显具有更强的特征提取能力，且可以准确判断聚集条件下猪只实例个数，主要体现在特征提取和处理的过程中：扩大感受野、为特征信息添加注意力、将深层语义信息和通道特征交互连接，可以更好地定位图像实例，增强分割模型对图像的学习能力。

3.4.3　BoxInst 训练结果

由于 Mask R－CNN 必须依靠像素级的掩码信息进行反向传播，才能得到优秀的实例分割效果，3.4.2 节同样说明了尽管 RdsiNet 骨干网络改善了特征提取的过程，但最终测试图像中掩码仍较为粗糙。基于此，考虑到本文制作的数据集可以提供准确的边界框信息，因此再次使用仅需边界框作为监督信息的 BoxInst 实例分割模型训练此数据集。

表 2 展示了基于 BoxInst 分割模型，ResNet－50 和 RdsiNet 骨干网络的参数，由于 BoxInst 只使用边界框作为监督信息，因此测试数据集中只计算 mAP_{Bbox}来衡量模型的性能。如表 2 所示，RdsiNet 的 mAP_{Bbox}较 ResNet－50 提升 2.2 个百分点，达到 89.6%，这说明使用 RdsiNet 骨干网络的 BoxInst 对于边界框的预测更加精准。

为进一步测试 BoxInst 分割模型的分割效果，本文在测试集中随机抽取了 50 幅图像，进行了像素级掩码标注，将标注掩码作为真值，同模型预测的掩码求不同阈值下的交并比，以此计算 mAP_{Semg}。计算结果如表 2 所示，RdsiNet 的 mAP_{Semg}为 95.2%，远高于 ResNet 的 76.7%，这体现了 BoxInst 分割模型下，RdsiNet 不仅分割效果更好，且具有更好的鲁棒性。

3.4.4　BoxInst 分割模型测试图像

图 9 展示了 BoxInst 弱监督实例分割模型在 ResNet－50 和 RdsiNet 骨干网络下最终的测试图像，可以明显看出，BoxInst 分割模型在 RdsiNet 骨干网络下具有更好的分割效果，其掩码不仅紧密地贴近猪只轮廓，呈现明显的猪只几何形状，而且在猪只移动的不同场景下依旧可以完美分割。而 ResNet－50 的图像分割效果出现较多问题，包括掩码过度覆盖、猪只漏检等，这说明本文所提出的 RdsiNet 骨干网络对于提升弱监督实例分割效果具有很大的作用。

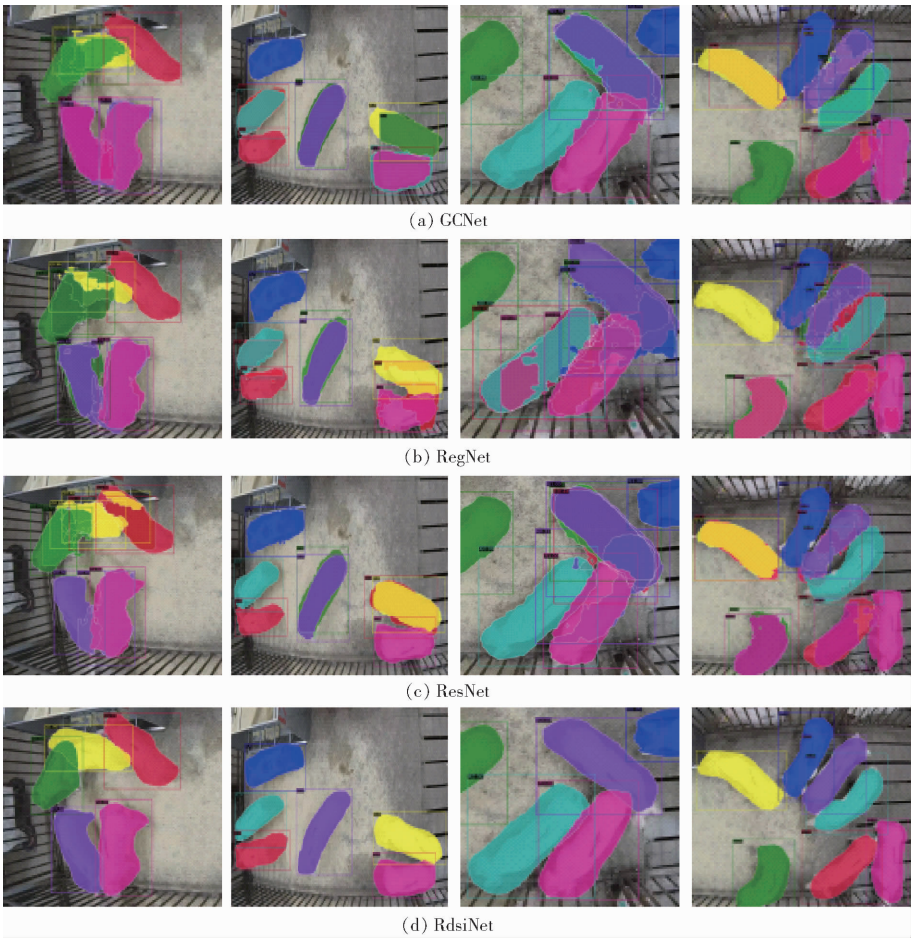


图 8 Mask R－CNN 模型中 4 种骨干网络分割效果对比

Fig. 8 Comparison of segmentation effects of four backbone networks of Mask R－CNN

表 2 2 种骨干网络训练结果对比

Tab. 2 Comparison of results by using two backbone networks

骨干网络	参数量	mAP_{Bbox}/%	mAP_{Semg}/%
ResNet－50	3.490×10^7	87.4	76.7
RdsiNet	3.805×10^7	89.6	95.2

3.5 消融实验

3.5.1 实验结果对比

本文使用 ResNet－50 骨干网络、添加空间注意力机制和第 2 代可变形卷积操作的 ResNet－50 网络以及本文提出的 RdsiNet 骨干网络，在 Mask R－CNN 分割模型上进行消融实验，表 3 是消融实验的结果。上述 3 种网络分别表示为 ResNet－50、ResNet－50＋SPA＋DCN 和 RdsiNet。如表 3 所示，空间注意力机制和第 2 代可变形卷积对图像实例分割效果的提升具有重要作用，额外增加 involution 算子之后的 RdsiNet 骨干网络相比较原始的 ResNet－50，mAP_{Bbox} 和 mAP_{Semg} 提升 4.2、4.8 个百分点，总计达到 93.4% 和 88.6%。实验结果表明 involution 算子不仅可以提升模型的性能，还可以大幅降低网络参数。表 3 中的数据表明，本文提出的骨干网络在提升分割精度的同时，还将参数量控制在合理范围内，以较低的代价换取了更好的性能。

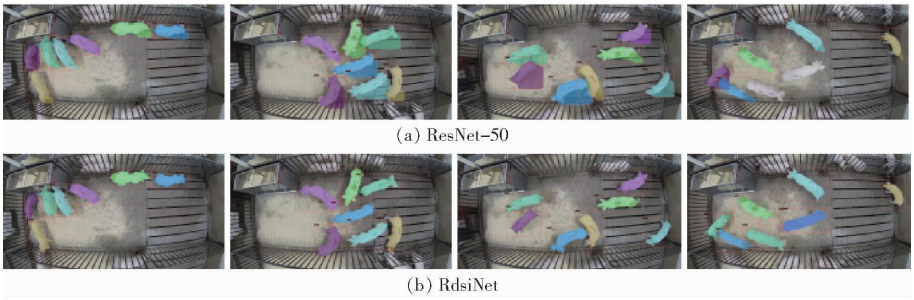


图 9 BoxInst 下两种骨干网络分割效果对比

Fig. 9 Comparison of segmentation effects of two backbone networks of BoxInst

表 3　消融实验骨干网络结果对比

Tab. 3　Comparison of backbone performance in ablation experiments

主干网络	参数量	mAP_{Bbox}/%	mAP_{Semg}/%
ResNet－50	3.490×10^7	88.8	83.1
ResNet－50＋SPA＋DCN	4.469×10^7	92.8	87.3
RdsiNet	3.805×10^7	93.4	88.6

3.5.2　类激活图

由于神经网络具有不可解释性，因此很难从正向推导的方式去判定不同特征提取方式的作用。但特征图的权重可以认为是被卷积核过滤后而保留的有效信息，其值越大，表明特征越有效，对网络预测结果越重要。基于此，本文使用 Grad－CAM[42] 对输入图像生成类激活的热力图，如图 10 所示，颜色越深红的地方表示值越大，其值越大，表明特征越有效，表示原始图像对应区域对网络的响应越高、贡献越大，对网络预测结果越重要。对比消融实验中 3 个骨干网络的类激活图，可以看出增加了第 2 代可变形卷积核空间注意力机制后，网络感受野明显增大，但无法做到对猪只有效范围的提取精度；而增加了 involution 算子的 RdsiNet 网络不仅具有更大的感受野，而且其红色范围更加准确，进一步证明了其对有效特征提取的准确度较高。

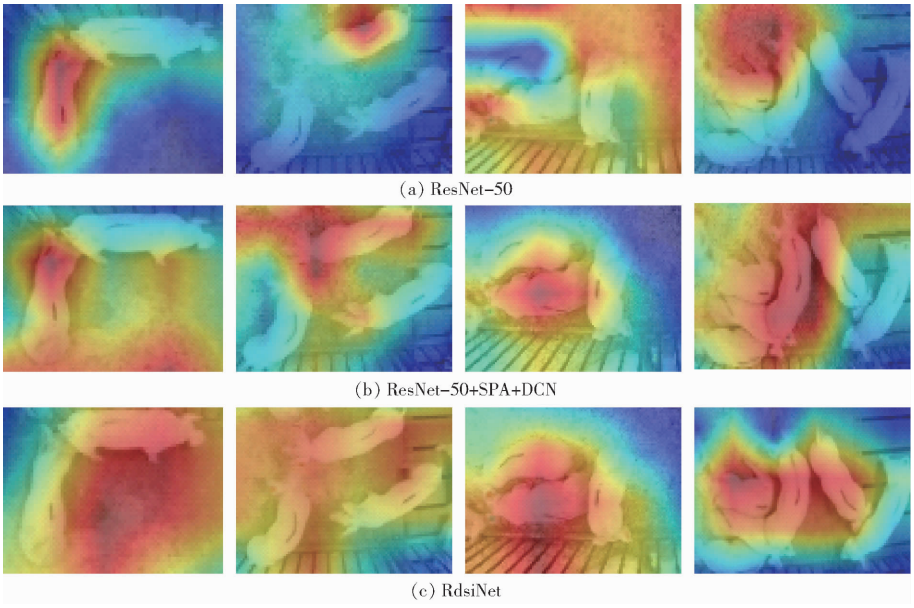


图 10　3 种骨干网络类激活图

Fig. 10　Heatmaps of three backbone networks

4　结论

（1）提出使用弱监督学习的方法进行猪只图像实例分割，制作粗糙轮廓标注的弱监督数据集，解决了逐像素标注数据集过程中具有的时间成本高、效率低、标注难等问题。同时，为解决弱监督会造成网络学习性能下降的问题，使用第 2 代可变卷积、空间注意力机制和 involution 算子搭建 RdsiNet 特征提取骨干网络，在对图像进行特征提取和处理的过程中，扩大网络感受野、加强重要特征信息和解决深层网络空间信息丢失问题，并且将骨干网络深层中提取出来的抽象语义信息和特征映射相连接，优化了猪只图像实例分割的效果。通过消融实验验证了 RdsiNet 骨干网络在弱监督数据集上的有效性。

（2）基于 Mask R－CNN 分割模型，将 ResNet－50、GCNet、RegNet、ResNeSt、CotNet 和本文提出的 RdsiNet 骨干网络做对比实验，RdsiNet 取得了最高的 mAP_{Bbox} 和 mAP_{Semg}，分别为 93.4% 和 88.6%，比 ResNet－50 分别提高 5.6、5.5 个百分点。在分割测试图像方面中，RdsiNet 同样具有最好的表现，尤其在猪只堆叠、模糊的情况下，RdsiNet 比 ResNet－50 具有更好的空间位置特征提取能力；最后通过使用消融实验和类激活图进一步验证了 RdsiNet 构建的合理性和有效性。

（3）为进一步改善分割效果，使用基于边界框作为监督信息的 BoxInst 实例分割模型，分别使用 ResNet－50 和 RdsiNet 骨干网络进行训练。对比之下，RdsiNet 不仅有更高的 mAP_{Bbox} 和 mAP_{Semg}，且具有更好的分割效果，同样表明了 RdsiNet 在图像特征提取过程中的优势，可以为猪只体重预测、个体识别跟踪等任务提供参考。

参 考 文 献

[1] 杨亮,熊本海,王辉,等. 人工智能养猪在我国的发展现状与研究展望[J]. 猪业科学,2022,39(11):41 - 44.

[2] 沈明霞,陈金鑫,丁奇安,等. 生猪自动化养殖装备与技术研究进展与展望[J]. 农业机械学报,2022,53(12):1 - 19.
SHEN Mingxia, CHEN Jinxin, DING Qi'an, et al. Current situation and development trend of pig automated farming equipment application[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(12): 1 - 19. (in Chinese)

[3] 姚继红. 智慧养殖管理模式在养猪生产中的应用[J]. 畜禽业,2022,33(9):33 - 35.

[4] CANG Y, HE H, QIAO Y. An intelligent pig weights estimate method based on deep learning in sow stall environments[J]. IEEE Access, 2019, 7: 164867 - 164875.

[5] SUWANNAKHUN S, DAUNGMALA P. Estimating pig weight with digital image processing using deep learning[C]// 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). IEEE, 2018: 320 - 326.

[6] CZYCHOLL I, HAUSCHILD E, BÜTTNER B, et al. Tail and ear postures of growing pigs in two different housing conditions [J]. Behavioural Processes, 2020, 176: 104138.

[7] LUO Y, ZENG Z, LU H, et al. Posture detection of individual pigs based on lightweight convolution neural networks and efficient channel-wise attention[J]. Sensors, 2021, 21(24): 8369.

[8] 高云,李静,余梅,等. 基于多尺度感知的高密度猪只计数网络研究[J]. 农业机械学报,2021,52(9):172 - 178.
GAO Yun, LI Jing, YU Mei, et al. High-density pig counting net based on multi-scale aware[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(9): 172 - 178. (in Chinese)

[9] RIEKERT M, KLEIN A, ADRION F, et al. Automatically detecting pig position and posture by 2D camera imaging and deep learning[J]. Computers and Electronics in Agriculture, 2020, 174: 105391.

[10] ALAMEER A, KYRIAZAKIS I, BACARDIT J. Automated recognition of postures and drinking behaviour for the detection of compromised health in pigs[J]. Scientific Reports, 2020, 10(1): 1 - 15.

[11] CHEN C, ZHU W, LIU D, et al. Detection of aggressive behaviours in pigs using a RealSence depth sensor[J]. Computers and Electronics in Agriculture, 2019, 166: 105003.

[12] 王荣,高荣华,李奇峰,等. 融合特征金字塔与可变形卷积的高密度群养猪计数方法[J]. 农业机械学报,2022,53(10):252 - 260.
WANG Rong, GAO Ronghua, LI Qifeng, et al. High-density pig herd counting method combined with feature pyramid and deformable convolution[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(10): 252 - 260. (in Chinese)

[13] HE H, QIAO Y, LI X, et al. Automatic weight measurement of pigs based on 3D images and regression network[J]. Computers and Electronics in Agriculture, 2021, 187: 106299.

[14] 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):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(Supp.): 261 - 266, 275. (in Chinese)

[15] 胡云鸽,苍岩,乔玉龙. 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. Design of intelligent pig counting system based on improved instance segmentation algorithm[J]. Transactions of the CASE, 2020, 36(19): 177 - 183. (in Chinese)

[16] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R - CNN[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961 - 2969.

[17] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117 - 2125.

[18] 栾晓梅,刘恩海,武鹏飞,等. 基于边缘增强的遥感图像弱监督语义分割方法[J]. 计算机工程与应用,2022,58(20):188 - 196.
LUAN Xiaomei, LIU Enhai, WU Pengfei, et al. Weakly-supervised semantic segmentation method of remote sensing images based on edge enhancement[J]. Computer Engineering and Applications, 2022, 58(20): 188 - 196. (in Chinese)

[19] BILEN H, VEDALDI A. Weakly supervised deep detection networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2846 - 2854.

[20] 赵亚楠,邓寒冰,刘婷,等. 基于弱监督学习的玉米苗期植株图像实例分割方法[J]. 农业工程学报,2022,38(19):143 - 152.
ZHAO Ya'nan, DENG Hanbing, LIU Ting, et al. Instance segmentation method of seedling maize plant images based on weak supervised learning[J]. Transactions of the CSAE, 2022, 38(19): 143 - 152. (in Chinese)

[21] 黄亮,吴春燕,李小祥,等. 基于弱监督语义分割的灯盏花无人机遥感种植信息提取[J]. 农业机械学报,2022,53(4):157 - 163,217.
HUANG Liang, WU Chunyan, LI Xiaoxiang, et al. Extraction of *Erigeron breviscapus* planting information by unmanned aerial vehicle remote sensing based on weakly supervised semantic segmentation[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(4): 157 - 163, 217. (in Chinese)

[22] TIAN Z, SHEN C, WANG X, et al. Boxinst: high-performance instance segmentation with box annotations[C]// Proceedings

of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2021：5443－5452.
[23] ZHOU Z H. A brief introduction to weakly supervised learning[J]. National Science Review, 2018, 5(1)：44－53.
[24] LU J, WANG W, ZHAO K, et al. Recognition and segmentation of individual pigs based on Swin Transformer[J]. Animal Genetics, 2022, 53(6)：794－802.
[25] PSOTA E T, SCHMIDT T, MOTE B, et al. Long-term tracking of group-housed livestock using keypoint detection and map estimation for individual animal identification[J]. Sensors, 2020, 20(13)：3670.
[26] OQUAB M, BOTTOU L, LAPTEV I, et al. Is object localization for free? —weakly-supervised learning with convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2015：685－694.
[27] LI Y F, GUO L Z, ZHOU Z H. Towards safe weakly supervised learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(1)：334－346.
[28] DURAND T, THOME N, CORD M. Weldon：weakly supervised learning of deep convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016：4743－4752.
[29] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]//Proceedings of the IEEE International Conference on Computer Vision,2017：764－773.
[30] ZHU X, HU H, LIN S, et al. Deformable convnets v2：more deformable, better results[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019：9308－9316.
[31] 郭子豪,董乐乐,曲志坚.基于改进 Faster RCNN 的节肢动物目标检测方法[J].计算机应用,2023,43(1):88－97.
GUO Zihao, DONG Lele, QU Zhijian. Arthropod object detection method based on improved Faster RCNN[J]. Journal of Computer Applications,2023,43(1):88－97.(in Chinese)
[32] 庄前伟,王志明,吴龙贻,等.基于改进 SOLOv2 的穴盘幼苗图像分割方法[J].南京农业大学学报,2023,46(1):200－209.
ZHUANG Qianwei, WANG Zhiming, WU Longyi, et al. Image segmentation method of plug seedlings based on improved SOLOv2[J]. Journal of Nanjing Agricultural University,2023,46(1):200－209.(in Chinese)
[33] WOO S, PARK J, LEE J Y, et al. Cbam：convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV),2018：3－19.
[34] 俞利新,崔祺,车军,等.结合结构重参数化方法与空间注意力机制的图像融合模型[J].计算机应用研究,2022,39(5):1573－1578.
YU Lixin, CUI Qi, CHE Jun, et al. Image fusion model based on structure reparameterization method and spatial attention mechanism[J]. Application Research of Computers,2022,39(5):1573－1578.(in Chinese)
[35] LI D, HU J, WANG C, et al. Involution：inverting the inherence of convolution for visual recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2021：12321－12330.
[36] 杨洪刚,陈洁洁,徐梦飞.双线性内卷神经网络用于眼底疾病图像分类[J].计算机应用,2023,43(1):259－264.
YANG Honggang, CHEN Jiejie, XU Mengfei. Bilinear involution neural network for image classification of fundus diseases[J]. Journal of Computer Applications,2023,43(1):259－264.(in Chinese)
[37] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016：770－778.
[38] CAO Y, XU J, LIN S, et al. Gcnet：non-local networks meet squeeze-excitation networks and beyond[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops,2019.
[39] RADOSAVOVIC I, KOSARAJU R P, GIRSHICK R, et al. Designing network design spaces[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020：10428－10436.
[40] ZHANG H, WU C, ZHANG Z, et al. Resnest：split-attention networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022：2736－2746.
[41] LI Y, YAO T, PAN Y, et al. Contextual transformer networks for visual recognition[J]. arXiv Preprint, arXiv：2017.12292,2021.
[42] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam：visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE International Conference on Computer Vision,2017：618－626.