

一种基于改进SSD网络的猪个体目标检测方法研究

黄 静, 张 健

(浙江理工大学信息学院, 浙江 杭州 310018)

✉syhj_sy@163.com; gxzj126@126.com



摘 要: 猪个体的目标检测对实现猪养殖过程的精细化管理, 促进猪养殖业的智能化与信息化升级具有重要意义。针对实际猪舍环境光照较暗, 猪个体被遮挡导致检测困难的实际问题, 同时兼顾实时性监测的需求, 提出一种基于SSD(Single Shot MultiBox Detector)网络改进的猪个体目标检测算法。首先利用大量猪个体图像对改进后的SSD网络进行训练, 然后对猪个体目标检测网络进行评估, 最后使用真实的猪舍视频对该网络模型进行测试。结果显示, 改进后的SSD网络在对猪个体进行检测时有96.38%的平均精度, 在实际具有光线变化和目标遮挡情况的视频中平均能达到21.6 FPS。通过对图像帧的随机抽样发现, 漏检率降低了3.65%, 误检率降低了1.45%。

关键词: 目标检测; SSD; 特征金字塔; 注意力机制

中图分类号: TP391.4 **文献标识码:** A

Research on a Target Detection Algorithm of Individual Pig based on Improved SSD Network

HUANG Jing, ZHANG Jian

(School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

✉syhj_sy@163.com; gxzj126@126.com

Abstract: Target detection of individual pig is of great significance to realizing the fine management of pig breeding process and promote the intelligent and information upgrading of pig breeding industry. Aiming at the detection difficulty of individual pig blocking in the dark pig house, this paper proposes an improved pig target detection algorithm based on SSD (Single Shot MultiBox Detector) network, taking into account the needs of real-time monitoring. Firstly, a large number of individual pig images are used to train the improved SSD network; then, the individual pig target detection network is evaluated; finally, the network model is tested with real pig house video. Results show that the improved SSD network has an average accuracy of 96.38% in the detection of individual pigs, reaching an average of 21.6 FPS (Frames Per Second) in the actual video with illumination change and target occlusion. Through random sampling of image frames, it is found that the missed detection rate is reduced by 3.65%, and the false detection rate is reduced by 1.45%.

Keywords: target detection; SSD; feature pyramid; attention mechanism

1 引言(Introduction)

随着我国农业转向高质量发展阶段, 畜牧养殖业也向着标准化、专精化、信息与智能化的发展方向转变, 出现了智能化养殖的新理念^[1]。畜禽的健康养殖要求养殖人员对其养殖的每一只畜禽要有及时和准确的了解, 以实现智能化的管理^[2]。与此相关的自动化和智能化既是相关产业发展的迫切需求, 也是相关科研人员研究的重点之一^[3]。生猪养殖业就是这种情

况的典型代表, 通过图像进行猪个体的目标检测, 为后续实现猪养殖过程的精细化管理提供了技术支持。

对于猪个体目标检测这样的问题, 近年来, 深度学习技术应用于该领域的算法发展很快, 各种优秀的网络模型被提出来。基于神经网络的目标检测算法又可以分为两种: One-stage目标检测算法^[4]与Two-stage目标检测算法^[5]。One-stage目标检测算法事先预设候选框, 因此其计算速度比较

快, 计算量相对较低, 缺点是准确度相对较低。YOLO系列目标检测算法是其代表, 目前已经有YOLOv1^[6]、YOLOv2^[7]、YOLOv3^[8]、YOLOv4^[9]、YOLOv5^[10]。Two-stage目标检测算法的过程则是在输入后, 第一步确定候选框, 然后对这些候选框进行识别, 以确定其类别。因此, 这些算法得到的检测结果往往比较精确, 但是计算量较大, 目标检测的速度比较慢。其代表网络算法则为RCNN系列, 包括R-CNN^[11]、Fast R-CNN^[12]、Faster R-CNN^[13]及Mask R-CNN^[14]等。

目前, 卷积神经网络已经用于猪个体目标检测问题。例如, PSOTA等^[15]和宋伟先^[16]都利用卷积神经网络进行猪个体目标检测; 房俊龙等^[17]搭建了改进的CenterNet网络; 孙东来等^[18]搭建了改进的YOLOv3网络等。

上述模型采集的数据图像大多采集于大型养殖场, 图像质量较高, 未能充分考虑复杂猪舍环境中光照较暗, 猪易被遮挡导致的图像区域面积较小、特征匮乏的问题。同时, 实际猪个体目标检测出现漏检错检问题也大都与这两个原因有关。

针对该问题, 本文通过在原始SSD算法的基础上引入FPN结构^[19]和CBAM注意力机制^[20], 并引入Focal Loss损失函数, 在不显著降低原始算法计算实时性的同时, 提高算法的猪个体目标检测精度和对猪遮挡及光线不足情况的适应性, 更好地实现在实际猪舍环境下的猪个体目标检测。

2 SSD算法模型(The model of SSD algorithm)

SSD算法是一种经典的目标检测算法, 也是目前目标检测算法一种重要的检测框架。原始的SSD算法性能相对平均, 其与以精准确度见长的Faster R-CNN算法相比检测速度快, 与以速度著称的YOLO系列算法相比精准性更好。

2.1 SSD算法的网络结构

SSD网络的结构如图1所示。

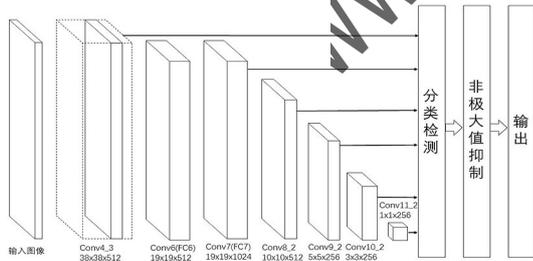


图1 SSD网络结构图

Fig.1 Diagram of SSD network structure

SSD的网络模型使用VGG网络作为主干网络, 与VGG16网络一样有16层, 包含五组卷积。其中, 有两组卷积都分别由两个卷积层组成, 剩下的三组卷积则分别由三个卷积层以及三个全连接层构成。但是和原始的VGG16网络不同的是, 起初为全连接层的FC6和FC7转为卷积层, 还一并摒弃了全部的Dropout层与FC8层。此外, 在VGG16的特征提取层后又新增了Conv6、Conv7、Conv8、Conv9四个卷积层。

2.2 SSD算法的目标预测和分类过程

SSD算法使用六个卷积层对应的大小各异的特征图作为预测部分的输入, 分别是第四卷积层中的最后一次卷积的特征图、FC7中卷积的特征图、第六和第七以及第八和第九卷积层中的中间的卷积对应的特征图。SSD默认框从这六张特征图中生成先验框。先验框由两部分组成: 先验框中心位置信息和尺寸信息。先验框中心能够把特征图划分为许许多多尺寸固定的网格, 其数量的多少与特征图的大小成正相关, 这就相当于将原始的输入图像根据特征图的尺寸分别分割成网格。

SSD算法会在每个特征图中生成四个或者六个先验框, 然后进入先验框的微调和筛选过程。在一张特征图中, SSD算法会先根据交并比选出和真实框重合度超过某个事先确定的阈值的先验框, 这些先验框被称为正样本。而其他的先验框则被称为负样本。在这种情况下, 可能会产生正样本的数量远远少于负样本的问题。因此, SSD采用了和R-CNN网络同样的难例挖掘方法, 也就是首先使用正样本和负样本的一个子集去训练SSD网络, 然后立即用这个网络模型进行预测并将那些被错误预测的负样本(这些负样本的正确的分类应该是正样本)收集起来, 形成难样本集, 最后再使用这个难样本集去训练网络模型, 如此反复。在整个过程中, 保持正负样本的数量比例接近1:3。完成后, SSD算法使用非极大值抑制的方法进行先验框的筛选, 将不合适的先验框去掉, 留下的即为检测结果。

2.3 SSD算法的损失函数

SSD算法的损失函数包含预测类别置信度损失和位置回归损失两部分。损失函数是通过这两者的结合产生的, 具体如式(1)所示:

$$L(x, c, l, g) = \frac{1}{N} (L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)) \quad (1)$$

其中, N 代表所有先验框的总数, 既囊括了正样本也囊括了负样本, 如果其值为0, 则将损失设置为0。 $L_{\text{loc}}(x, l, g)$ 表示位置回归损失, $L_{\text{conf}}(x, c)$ 表示预测类别置信度损失, α 用于调整两种损失之间的比例, 其默认值为1。边框回归损失部分采用了Smooth L1损失函数, 可以量化先验框和真实框之间位置上的误差, 如式(2)一(7)所示:

$$L_{\text{loc}}(x, l, g) = \sum_{i \in \text{Pos}} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k \text{smooth}_{\text{L1}}(l_i^m - \hat{g}_j^m) \quad (2)$$

$$\text{smooth}_{\text{L1}}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5\alpha & \text{otherwise} \end{cases} \quad (3)$$

$$\hat{g}_j^{\text{cx}} = (g_j^{\text{cx}} - d_i^{\text{cx}}) / d_i^w \quad (4)$$

$$\hat{g}_j^{\text{cy}} = (g_j^{\text{cy}} - d_i^{\text{cy}}) / d_i^h \quad (5)$$

$$\hat{g}_j^w = \log\left(\frac{g_j^w}{d_i^w}\right) \quad (6)$$

$$\hat{g}_j^h = \log \left(\frac{g_j^h}{d_j^h} \right) \quad (7)$$

其中，后面四个公式的作用是将真实框的坐标值转化成能够计算损失的编码值。 \hat{g}_j^x 和 \hat{g}_j^y 表示真实框中心x坐标和y坐标的损失值， \hat{g}_j^w 和 \hat{g}_j^h 则表示真实框宽度和高度的损失值。

置信度损失函数如式(8)和式(9)所示：

$$l_{\text{conf}}(x, c) = - \sum_{i \in \text{Pos}} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in \text{Neg}} \log(\hat{c}_i^0) \quad (8)$$

$$\hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)} \quad (9)$$

其中，Pos和Neg是正负样本的英文缩写。式(8)中第一项是第*i*个先验框与第*j*个真实框关于类别*p*进行匹配的损失值， \hat{c}_i^p 是第*i*个先验框预测该类别的概率值。式(8)中第二项表示预测框预测为背景的损失值， \hat{c}_i^0 表示第*i*个先验框预测为背景的概率值。 x_{ij}^p 取值为0或者1，其值为1时表示第*i*个先验框成功匹配到第*j*个类别为*p*的真实框，其值为0时则表示没有匹配到任何真实框。

3 SSD算法的改进(Improvement of SSD algorithm)

由于猪图像的光照条件差、猪个体被遮挡等造成检测困难的直接原因是图像中的目标区域较小，特征不够明显。为了解决这个问题，需要加强算法的特征提取能力，增强其对较小目标和困难目标的检测能力。同时，通过改进损失函数，可以使网络更加专注于这些难以检测的目标，提高该类目标的检测成功率。基于此，本文做了如下改进。

3.1 CBAM注意力机制

在算法中加入注意力机制可以使其无视无足轻重的特征信息，着重注意重要的特征信息，增强算法的表征能力，进而提高算法的目标检测能力。

CBAM作为混合域注意力机制，将通道和空间这两种因素结合在一起构建注意力模块，使其更加全面，也更加高效。同时，作为一种轻量化模块，加入CBAM模块不会显著增加卷积神经网络的计算量，却可以提高目标检测的准确性。

CBAM模块的流程是：首先在卷积神经网络中选出特定的特征图送入通道注意力模块中，将通道注意力结果和初始的特征图进行乘法操作；将所得到的中间结果送入空间注意力模块中，然后将空间模块的输出结果与原来的中间结果相乘，即可得出最后的输出结果。

3.2 特征金字塔结构

SSD目标检测算法是一款经典的目标检测算法，既拥有较快的计算速度，满足大多数场景下实时性的需求，又可以拥有不错的准确率，而且与Faster R-CNN相比，其训练过程无须中间权重的存储过程，降低了训练成本。SSD网络所采用的多特征图检测方式当时较为先进，但从当下看仍有较大的改进空间。

单从算法结构方面来说，SSD算法通过改进VGG16网络，提取其中六个特征图，分别对这些特征图进行直接识别和分类，可能会出现多个特征图识别到同一个语义信息的问题，造成重复和浪费。由于不同尺度的特征图对不同目标的表征能力不同，而SSD算法又没有当下已经应用很普遍的多特征融合，造成了其目标的检测能力有较大进步空间。

特征金字塔网络结构(FPN)是一种常常被用来解决目标多尺度检测问题的重要手段，是多特征融合的主要结构之一。其在YOLOv3以来的YOLO系列算法中与PAN结构相结合的应用已经十分成熟。

FPN结构只需在已经成熟的网络模型基础上更改网络层之间连接的方式，在尽量降低对算法计算量和实时性的影响的情况下，有效提高网络模型对小目标的检测能力。

特征金字塔结构将深层特征图(包含丰富的语义特征信息)和浅层特征图(包含丰富的纹理信息)通过三种连接方式(自上而下、自下而上和横向连接)进行相互结合，使得各不同尺度特征图都拥有丰富的特征信息，然后再在这些不同层级的特征图上进行检测。

将特征金字塔、CBAM模块和原始的SSD算法进行有机结合，得到改进的SSD算法，其网络结构如图2所示。

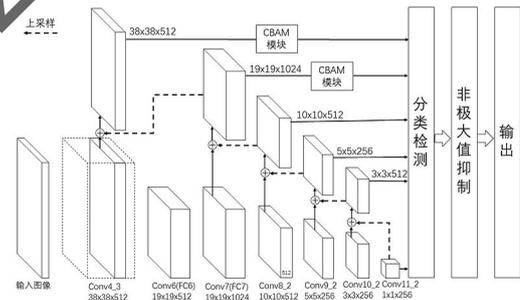


图2 改进后的SSD网络结构图

Fig.2 Diagram of improved SSD network structure

3.3 Focal Loss损失函数

对于因为光线条件较差，猪被遮挡导致的难以识别的猪个体，其检测网络模型应该进行重点关注。同时，猪个体目标检测问题只需分辨猪个体和背景这两种类别，目标数量较少，这样就会造成正负样本的数量不平衡。虽然SSD算法事先设定了正负样本的比例，避免被过多的负样本在训练时浪费掉模型分辨样本的能力，但还需要进一步改进。

为此，引入Focal Loss损失函数。该损失函数可以有针对性地解决这个问题。

Focal Loss损失函数如式(10)所示：

$$L = -y \log y' - (1-y) \log(1-y') = \begin{cases} -\log y' & y=1 \\ -\log(1-y') & y=0 \end{cases} \quad (10)$$

其中， y' 是网络输出的预测类别概率， y 是真实样本的标签。当样本数据量比较大时，损失函数的更新迭代速度会变得很

慢，且不容易达到全局最优。

初步改进后的Focal Loss损失函数如式(11)所示：

$$L_n = \begin{cases} -(1-y')^\gamma \log y', & y = 1 \\ -y'^\gamma \log(1-y'), & y = 0 \end{cases} \quad (11)$$

式(11)在原有的公式中添加了一个 $(1-y')^\gamma$ 作为权重因子，此时，预测出正确样本时网络输出的概率越大，这个权重因子就会越小，即分类正确时，样本的损失权重较小；分类错误时，样本的损失权重较大。这样网络就不能通过对简单样本识别的不停优化来降低总体损失，却放弃对难分类样本的辨别，使得模型不得不重点关注比较难以分类的样本。

同时，再加入一个平衡权重因子 α ，用来解决正负样本数量比例不均问题，如式(12)所示：

$$L_n = \begin{cases} -\alpha(1-y')^\gamma \log y', & y = 1 \\ -(1-\alpha)y'^\gamma \log(1-y'), & y = 0 \end{cases} \quad (12)$$

这样就既能够解决正负样本数量不平衡问题，也能够使得模型更加注重难分类的样本。在本次实验中，这些样本就是因为光照、遮挡等问题造成的难以检测到的样本。

将Focal Loss损失函数置换原本的预测类别置信度损失函数，得到的改进SSD网络结构的损失函数如式(13)所示：

$$L(x, c, l, g) = \frac{1}{N} (L_n(x, c) + \alpha L_{loc}(x, l, g)) \quad (13)$$

4 实验分析(Experimental analysis)

4.1 数据集制作

本次实验通过网络爬虫程序获得网络上公开的猪目标图片，以及从互联网公开视频和自己实地拍摄的视频中截取猪目标图片共5,122张，其中特别加入含有猪被遮挡、光照条件差的图片约1,000张组成实验数据集。每张图片至少包含一个目标，可能包含多个目标。整个数据集按照Pascal VOC 2007数据集格式制作，在统一将图片尺寸缩放为416×416后，利用Imglab软件对图片中的目标进行标注，并将结果保存为XML文件。标注完成后，随机按8：1：1将其划分成训练集、验证集和测试集。

4.2 实验平台条件

本次实验的软硬件条件及其相关配置如表1所示。

表1 实验环境及配置

Tab.1 Experimental environment and configuration

实验环境	配置
CPU	i9-7900X
显卡	RTX 2080Ti
内存	12 GB
系统	CentOS 7
编程语言	Python 3.7.6
深度学习框架	Pytorch 1.7.1

4.3 评价指标

结合本次实验的实际情况，决定采用平均精度指标作为评价标准，其是精确率P关于召回率R的积分。两者分别如式(14)和式(15)所示：

$$P = \frac{TP}{TP + FP} \quad (14)$$

$$R = \frac{TP}{TP + FN} \quad (15)$$

其中，TP表示正样本被正确检测的个数；FN表示正样本被错误检测成负样本的个数；FP表示负样本被错误检测成正样本的个数。精确率表示的是正确分类的样本占所有样本的比重；召回率则表示的是检测正确的正样本占所有正样本的比重。

另外三个评价指标为检测速度、误检率、漏检率。检测速度指其在数据集或者视频流中每一秒检测的图像张数或者帧数。误检率为将不是目标的图像检测为目标数目占总应检出目标数目的比值。漏检率为本应是目标的图像检测为不是目标的数目占总应检出目标数目的比值。

4.4 实验结果分析

分别对原始SSD算法和改进后的SSD算法进行训练，待两者的损失函数收敛后完成训练。分别使用测试集对这两个已经训练好的模型进行对比验证，得到的结果如表2所示。

表2 SSD网络模型改进前后性能指标对比表

Tab.2 Comparison table of performance indicators before and after the improvement of SSD network model

模型	检测速度/FPS	平均精度/%
原始SSD算法	39.07	93.41
改进SSD算法	38.41	96.38

为了验证改进后的SSD算法模型的实际效果，取存在光照较差、猪被遮挡情况的猪舍视频进行视频检测。检测过程中，检测平均速度为21.6 FPS，典型检测结果截图如图3所示。

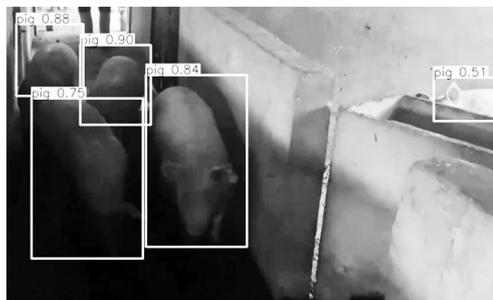


图3 视频检测结果图

Fig.3 Diagram of video detection results

为了定量对改进后的SSD算法模型的检测效果进行析，将两种算法模型对上述视频的检测结果随机进行截取，获得约800张图像。随机选取其中292张含有猪目标的图像进行计数，计算出两种算法对该视频的误检率和漏检率结果，如表3所示。

表3 SSD网络模型改进前后实际视频检测结果对比表

Tab.3 Comparison table of actual video detection results before and after the improvement of SSD network model

模型	漏检率/%	误检率/%
原始SSD算法	10.17	8.81
改进SSD算法	6.52	7.36

从表2和表3可以看出,改进的SSD算法有96.38%的平均精度,相较于原始SSD算法,其FPS指标降低了1.7%,但平均精度却提高了2.97%。平均精度的提高得益于改进的SSD算法通过FPN结构与CBAM机制的联合应用,提高了网络对猪个体目标的特征提取能力。同时,漏检率降低了3.65%,误检率降低了1.45%,两者的降低不仅与上述网络结构的改进有关,改进的损失函数使网络模型更加关注这类目标也起了一定作用。整体改进后虽然算法的实时性略有降低,但在进行视频检测时仍能有21.6 FPS,基本可以满足可能的视频实时监测需求。同时,从图3中可以发现,对于左侧光线较暗且猪个体存在一定的相互遮挡的情况,以及右侧猪大部分被遮挡的情况,改进后的SSD算法均可以对其进行有效识别,但部分猪目标的检测框与实际个体之间的空隙稍大。综上说明,改进后的SSD算法对于光线较暗和猪遮挡情况均有良好的适应性,对因此造成的猪目标难以检测问题有一定的改善。

5 结论(Conclusion)

本文基于SSD目标检测算法,对猪个体目标检测问题进行探究。首先,通过广泛收集数据,构建了较为全面的数据集,同时对训练集、验证集和测试集进行了划分。然后,为了提高原始算法对猪个体目标的检测能力,使用FPN结构和CBAM模块对网络进行改进,使用Focal Loss损失函数对原有损失函数进行改进。在对改进后的网络完成训练后,使用测试集数据和猪舍视频进行了测试,最终的实验结果表明,相较于原算法,改进后的SSD网络模型在猪个体目标检测问题上有96.38%的平均精度,检测平均速度为21.6 FPS,并对光线变化和目标遮挡情况有更好的适应性。但同时也发现,部分检测框与实际个体之间的空隙稍大,需要进一步探究和改进。

参考文献(References)

- [1] 李凯.山东省水产绿色健康养殖现状及发展建议[J].渔业研究,2022,44(02):178-186.
- [2] 李宁波,周琼.工厂化集约化生猪智能养殖技术初探[J].畜牧兽医科技信息,2022(03):161-162.
- [3] 孙一平,李丹,蔺旭鸿,等.计算机视觉技术在家禽养殖与公鸡选种中应用综述[J].农业机械学报,2021,52(S1):219-228.
- [4] 包晓敏,王思琪.基于深度学习的目标检测算法综述[J].传感器与微系统,2022,41(04):5-9.

- [5] 吴雨露,张德贤.基于深度学习的目标检测算法综述[J].信息与电脑(理论版),2019(12):46-48.
- [6] 林健巍.YOLO图像检测技术综述[J].福建电脑,2019,35(09):80-83.
- [7] 薛月菊,黄宁,涂淑琴,等.未成熟芒果的改进YOLOv2识别方法[J].农业工程学报,2018,34(07):173-179.
- [8] 刘林芽,吴送英,左志远,等.基于YOLOv3算法的山区铁路边坡落石检测方法研究[J].计算机科学,2021,48(S2):290-294.
- [9] 王滢暄,宋煊生,梁浩翔,等.基于改进的YOLOv4高速公路车辆目标检测研究[J].计算机工程与应用,2021,57(13):218-226.
- [10] 杨晓玲,江伟欣,袁浩然.基于yolov5的交通标志识别检测[J].信息技术与信息化,2021(04):28-30.
- [11] 董静,耿达,郭迎港,等.室内环境下基于R-CNN的光照自适应物体检测[J].计算机工程与应用,2019,55(02):168-173,252.
- [12] 车翔玖,刘华罗,邵庆彬.基于Fast RCNN改进的布匹瑕疵识别算法[J].吉林大学学报(工学版),2019,49(06):2038-2044.
- [13] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6):1137-1149.
- [14] 吴金亮,王港,梁硕,等.基于Mask R-CNN的舰船目标检测研究[J].无线电工程,2018,48(11):947-952.
- [15] PSOTA E T, MITTEK M, PÉREZ L C, et al. Multi-pig part detection and association with a fully-convolutional network[J]. Sensors, 2019, 19(4):852.
- [16] 宋伟先.基于深度学习的猪只目标检测及状态分析[D].长春:东北农业大学,2019.
- [17] 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.
- [18] 孙东来,王继超,陈科,等.基于Ghost-YOLOv3-2算法的2尺度猪目标检测[J].江苏农业科学,2022,50(07):189-196.
- [19] 宋智勇,潘海鹏.基于空间注意力多尺度特征融合的织物缺陷分类算法[J].激光与光电子学进展,2022,59(10):114-120.
- [20] 王美华,吴振鑫,周祖光.基于注意力改进CBAM的农作物病虫害细粒度识别研究[J].农业机械学报,2021,52(04):239-247.

作者简介:

黄 静(1965-),女,博士,教授.研究领域:通信工程,大数据,深度学习.

张 健(1996-),男,硕士生.研究领域:目标检测,智能信息处理.