doi:10.6041/j.issn.1000-1298.2021.07.014

基于改进 YOLO v3 模型的奶牛发情行为识别研究

王少华[1,2]　何东健[1,2]

(1.西北农林科技大学机械与电子工程学院,陕西杨凌 712100;
2.农业农村部农业物联网重点实验室,陕西杨凌 712100)

摘要:为提高复杂环境下奶牛发情行为识别精度和速度,提出了一种基于改进 YOLO v3 模型的奶牛发情行为识别方法。针对 YOLO v3 模型原锚点框尺寸不适用于奶牛数据集的问题,对奶牛数据集进行聚类,并对获得的新锚点框尺寸进行优化;针对因数据集中奶牛个体偏大等原因而导致模型识别准确率低的问题,引入 DenseBlock 结构对 YOLO v3 模型原特征提取网络进行改进,提高了模型识别性能;将 YOLO v3 模型原边界框损失函数使用均方差(MSE)作为损失函数度量改为使用 F_{IoU} 和两框中心距离 D_c 度量,提出了新的边界框损失函数,使其具有尺度不变性。从 96 段具有发情爬跨行为的视频片段中各选取 50 帧图像,根据发情爬跨行为在活动区出现位置的不确定性和活动区光照变化的特点,对图像进行水平翻转、±15°旋转、随机亮度增强(降低)等数据增强操作,用增强后的数据构建训练集和验证集,对改进后的模型进行训练,并依据 F1、mAP、准确率 P 和召回率 R 指标进行模型优选。在测试集上的试验表明,本文方法模型的识别准确率为 99.15%,召回率为 97.62%,且处理速度达到 31 f/s,能够满足复杂养殖环境、全天候条件下奶牛发情行为的准确、实时识别。

关键词:奶牛发情;爬跨行为;YOLO v3;锚点框优化;DenseBlock;损失函数优化

中图分类号:TP391.41;S24　　**文献标识码**:A　　**文章编号**:1000-1298(2021)07-0141-10　　**OSID**:

Estrus Behavior Recognition of Dairy Cows Based on Improved YOLO v3 Model

WANG Shaohua[1,2]　HE Dongjian[1,2]

(1. *College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi* 712100, *China*
2. *Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi* 712100, *China*)

Abstract: Aiming to improve the detection accuracy and speed of estrus behavior of dairy cows in a complex scene, a method of recognizing estrus behavior of dairy cows based on improved YOLO v3 model was proposed. To solve the problem of the cows' size was inconsistent with the size of the object in the COCO dataset, which caused the original anchors were not applicable, new anchors were obtained by clustering new data sets and optimized by using linear expansion. As cows with a big size, the small difference between individuals and associations between behaviors, which was difficult to distinguish, a DenseBlock structure was introduced to the feature extraction network of YOLO v3 model to improve its detection performance on the large objects. Considered that the original bounding box loss function of YOLO v3 model was not invariant to the object scale, the F_{IoU} and the center distance D_c of two boxes were used as the measuring method, and a new bounding box loss function was proposed to make it scale-invariant. Totally 50 images were extracted each from 96 video mounting behavior clips of dairy cows, according to the uncertainty position of cows' mounting behavior in the active area and the character of the light changing of the active area, horizontally flipped, rotated ± 15° and random brightness enhancement (decrease) were applied on them for data augmentation. The augmented data was divided into three parts as training sets, validation sets, and test sets, training sets and validation sets were used to train the improved model and the best training model was chosen as dairy cow estrus behavior

收稿日期:2020-07-10　修回日期:2020-08-25

基金项目:陕西省重点产业创新链(群)——农业领域项目(2019ZDLNY02-05)和国家自然科学基金面上项目(61473235)

作者简介:王少华(1990—),男,博士生,主要从事农业图像视频处理研究,E-mail:smarwong@163.com

通信作者:何东健(1957—),男,教授,博士生导师,主要从事智能检测与农业信息技术研究,E-mail:hdj168@nwsuaf.edu.cn

recognition model with the indicators F1, mAP, accuracy rate P, and recall rate R. The experiment on test sets showed that the accuracy rate of the model was 99.15%, the recall rate was 97.62%, and the processing speed reached 31 f/s, which could accurately and real-time identify cows' estrus behavior in a complex breeding environment under all weather. The research could also provide a reference for other large livestock behavior recognition.

Key words: dairy cow estrus; mounting behavior; YOLO v3; anchor optimization; DenseBlock; loss function optimization

0 引言

目前,我国奶牛养殖场多以传统养殖方式为主,近年来随着牛奶消费需求的增加,奶牛养殖规模和数量快速增长,奶牛场劳动力成本与经济效益之间的矛盾日益突出,因此迫切需要自动化的奶牛养殖管理方法。评估奶牛发情状况是奶牛养殖的重要环节,奶牛发情期短,错过或漏识别会给奶牛场造成经济损失。传统的奶牛发情评估主要依靠人工进行观察,需要投入大量的时间和精力,因此,研究奶牛发情行为的自动识别方法具有重要意义。

随着计算机视觉技术的发展,不少研究者提出了基于视频图像分析识别奶牛发情的方法[1-8]。近年来,深度学习理论和方法得到了极大发展,研究者基于深度学习进行了许多精准养殖方面的研究[9-20]。JIANG 等[9]在 YOLO v3 网络中加入由均值滤波算法和带泄露修正线性单元(Leaky ReLU)构成的过滤层(Filter layer),优化训练过程,提出了一种基于 FYOLO 的奶牛关键部位识别方法,识别准确率为 99.18%。ZHANG 等[10]以视频帧和光流场变化作为网络输入,提出了一种 Two-Steam 的卷积神经网络模型,用于自动识别视频监控中的猪只行为,识别准确率为 98.99%。何东健等[13]通过优化锚点框和改进网络结构提出了一种基于改进 YOLO v3 模型的挤奶奶牛个体识别方法,识别准确率为 95.91%,该方法可实现复杂环境下奶牛个体的精准识别。在奶牛发情行为识别方面,PARK 等[16]将采集的牛只在不同状态和行为时的叫声输入卷积神经网络,训练网络通过声音识别牛只包括发情在内的不同行为和状态,识别准确率为 96.20%,但该声音采集方法对环境要求较高,不适用于奶牛场复杂的养殖环境。刘忠超等[17]受 LeNet-5 启发,构建了一种多层卷积神经网络,并使用奶牛行为图像样本训练模型,提出了一种基于卷积神经网络的奶牛发情行为识别模型,模型识别准确率为 98.25%,但该模型只能用于识别已获取的奶牛行为图像。王少华等[18]采用改进的 GMM 检测运动奶牛目标,基于训练好的 AlexNet 识别被检测奶牛目标的行为,提出了一种基于机器视觉的奶牛发情行为自动识别方法,识别准确率为 100%,召回率为 88.24%,但该方法识别速度较慢。作为一种自主特征学习方法,深度学习具有自动获取图像多层次、多维度特征信息的能力,有效克服了人工提取特征向量的局限性,在农业领域具有广阔的应用前景。以 YOLO v3 模型为代表的端到端的目标检测方法,具有识别速度快、识别准确率高、模型抗干扰能力强等优点,具备在复杂环境下识别奶牛发情行为的潜力。

针对复杂环境下奶牛发情行为检测精度和速度亟待提高的问题,本文提出一种基于改进 YOLO v3 模型的奶牛发情行为识别方法。在获取奶牛行为图像的基础上,利用 LabelImg 开源工具对图像中的爬跨行为进行标注,构建奶牛发情行为样本数据集;根据样本集数据特点,从锚点框尺寸优化、特征提取网络改进、边界框损失函数优化 3 方面对 YOLO v3 模型进行改进,并基于 TensorFlow 和高性能 GPU 计算平台实现对改进后模型的训练;最后,利用优选的模型,在测试集上对模型性能进行评估与分析。

1 材料与方法

1.1 数据来源

本研究供试视频采集自陕西省宝鸡市扶风县西北农林科技大学畜牧实验基地的奶牛养殖场,通过调研发现,奶牛发情爬跨行为主要发生在养殖场的奶牛活动区,故本研究选择对奶牛活动区进行视频采集。奶牛活动区长 30 m,宽 18 m,在奶牛活动区的对角线位置各安装了 2 个分辨率为 200 万像素的监控摄像机(YW7100HR09-SC62-TA12 型,深圳亿维锐创科技有限公司),摄像机安装高度 3.3 m,向下倾斜 15.5°,安装好的摄像机及视野图像如图 1 所示。视频采用场边俯视的角度记录了 56 头具备发情能力的成年泌乳奶牛的活动情况,视频记录时间为 2016 年 12 月至 2017 年 4 月,共采集到奶牛活动视频 3 600 段,每段视频长 10 min,视频分辨率为 1 920 像素(水平)×1 080 像素(垂直),帧率 25 f/s。

1.2 图像预处理

奶牛场复杂的养殖环境以及视频图像生成、传输过程受到干扰,使采集到的视频图像中存在噪声

(a) 摄像机安装位置　　(b) 摄像机视野图像

图 1　摄像机安装位置及视野图像

Fig. 1　Camera installation location and view image

而降低图像质量。为提高后期奶牛发情行为的识别效率，对采集到的视频图像进行降噪和图像增强操作。

分析本文视频图像噪声产生原因可知，噪声类型主要为系统加性高斯噪声，故可选用高斯滤波或双边滤波来去除。考虑到高斯滤波用加权平均的方法去除噪声，会使图像边缘处出现模糊，而双边滤波使用与空间距离或灰度距离相关的高斯函数相乘作为滤波因子，在滤除噪声的同时尽可能地考虑了图像的边缘信息，因此，本研究选用双边滤波滤除图像噪声。

由于本文视频采集跨度时间长，当夜晚环境亮度降低时，拍摄到的视频图像对比度会下降，导致奶牛轮廓与背景区分不明显。本文使用分段线性函数对图像进行对比度拉伸，通过对比度拉伸可以有目的地增强奶牛像素，使视频图像中的奶牛轮廓更加清晰，对比度拉伸计算公式为

$$f(x)=\begin{cases}\dfrac{a_1'}{a_1}x & (x<a_1)\\ \dfrac{a_2'-a_1'}{a_2-a_1}(x-a_1)+a_1' & (a_1\leqslant x\leqslant a_2)\\ \dfrac{255-a_2'}{255-a_1'}(x-a_2)+a_2' & (x>a_2)\end{cases} \tag{1}$$

式中　x——输入像素　　$f(x)$——输出图像

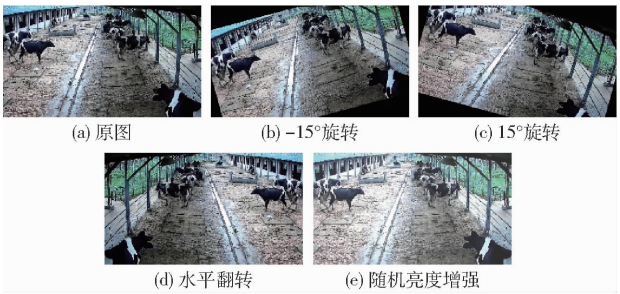
a_1、a_2——感兴趣目标灰度区间上、下限阈值，本文 $a_1=50$，$a_2=150$

a_1'、a_2'——感兴趣目标灰度区间新映射区间上、下限阈值，本文 $a_1'=40$，$a_2'=200$

其中，a_1、a_2、a_1' 和 a_2' 由预备试验确定。

1.3　供试样本集构建

观看奶牛活动区监控视频，人工筛选出具有爬跨行为的视频片段 96 段，每段长度 15 ~ 80 s 不等，利用视频帧分解技术，每 5 帧取 1 帧，得到视频图像 13 655 幅，其中，每幅图像中既包含有爬跨行为奶牛，也包含有其他行为奶牛，如站立、游走、躺卧等。由于各视频段图像数量并不均等，为避免因样本量差异对模型训练结果产生影响，将各视频段图像统一调整为 50 幅，最终构成样本集图像共有4 800 幅。为扩大样本量，提高数据泛性，根据奶牛爬跨行为在活动区出现位置的不确定性以及活动区光照变化的特点，采用对样本图像水平翻转、±15°旋转、随机亮度增强（降低）的方法进行数据增强，最终得到增强后的图像 14 400 幅，作为模型的样本增强数据集，经增强后的图像样例如图 2 所示。



(a) 原图　(b) -15°旋转　(c) 15°旋转

(d) 水平翻转　(e) 随机亮度增强

图 2　数据增强后的图像样例

Fig. 2　Sample images after data augmentation

按照训练、验证、测试之比为 7∶2∶1，对样本增强数据集进行随机划分后，训练集样本 10 800 幅、验证集样本 2 880 幅、测试集样本 1 440 幅，各类样本集间无数据重叠。采用开源工具 LabelImg 对训练集、验证集样本进行标注，标记对象为奶牛爬跨行为，标记标签为“mounting”，标记文件保存为 PascalVOC 格式，名称与图像名一一对应。

2　YOLO v3 模型构建与改进

2.1　YOLO v3 模型

2.1.1　特征提取网络

YOLO v3 模型[21]使用 Darknet－53 作为其特征提取网络。Darknet－53 共有 53 层卷积层，每个卷积层由 1×1 和 3×3 的卷积核构成，其余为残差层，均位于卷积层之后，其网络结构如图 3 所示。Darknet－53 引入残差（Residual）结构，可以很好地控制梯度的传播，使构建的网络更深，特征学习能力更强大。YOLO v3 模型的第 75 层到 105 层为特征融合层，使用上采样和多尺度特征融合方法，加强网络对小目标的检测精度。ImageNet 数据集[22]下，YOLO v3 模型特征提取网络与其他网络的性能对比试验结果表明[21]，Darknet－53 比 Darknet－19[23]速度有所下降，但精度提升，Darknet－53 性能优于 ResNet－101[24]，且速度提高了 1.3 倍，和 ResNet－152[24]性能相当，但速度提高了 1.1 倍。

2.1.2　边界框分类与预测

YOLO v3 模型基于锚点框尺寸微调实现对边界框的预测，其中 9 种锚点框尺寸由 COCO 数据集上聚类得到，并分别分配给 3 种不同的特征图，用于构建模型对不同大小目标的检测能力。锚点框尺寸与

特征图对应关系如表 1 所示。

	类型	滤波器	尺寸	输出
	Convolutional	32	3×3	512×512
	Convolutional	64	3×3/2	256×256
1×	Convolutional	32	1×1	
	Convolutional	64	3×3	
	Residual			256×256
	Convolutional	128	3×3/2	128×128
2×	Convolutional	64	1×1	
	Convolutional	128	3×3	
	Residual			128×128
	Convolutional	256	3×3/2	64×64
8×	Convolutional	128	1×1	
	Convolutional	256	3×3	
	Residual			64×64
	Convolutional	512	3×3/2	32×32
8×	Convolutional	256	1×1	
	Convolutional	512	3×3	
	Residual			32×32
	Convolutional	1024	3×3/2	16×16
4×	Convolutional	512	1×1	
	Convolutional	1024	3×3	
	Residual			16×16

图 3 Darknet－53 网络结构

Fig. 3 Structure of Darknet－53 network

表 1 锚点框尺寸与特征图对应关系

Tab. 1 Correspondence between anchor boxes size and feature maps

特征图尺寸/（像素×像素）	13×13	26×26	52×52
感受野分类	大	中	小
锚点框尺寸	116×90、156×198、373×326	30×61、62×45、59×119	10×13、16×30、33×23

YOLO v3 模型采用直接预测相对位置的方法对边界框进行预测，它将输入图像划分为 $S\times S$ 个网格，每个网格单元负责检查中心位于其内的边界框和置信度，网格单元的具体信息可以表示为 (t_x,t_y,t_w,t_h,C)，其中 t_x、t_y 表示边界框的坐标偏移量，t_w、t_h 表示尺度缩放，C 为预测结果置信度。YOLO v3 使用 t_x、t_y、t_w、t_h 作为学习信息，使用预设的锚点框尺寸经线性回归微调（平移加尺度缩放）实现边界框的预测。模型边界框预测的输出值 (t_x,t_y,t_w,t_h) 为 4 组偏移量，b_x、b_y、b_w、b_h 的计算公式为

$$\begin{cases} b_x=\sigma(t_x)+c_x \\ b_y=\sigma(t_y)+c_y \\ b_w=p_w e^{t_w} \\ b_h=p_h e^{t_h} \end{cases} \tag{2}$$

式中 σ——Sigmoid 激活函数

(c_x,c_y)——网格单元左上角坐标

t_x、t_y——边界框中心坐标偏移量

t_w、t_h——边界框宽和高的缩放尺度

p_w、p_h——锚点框宽和高在特征图上的特征映射

(b_x,b_y)——边界框中心坐标

b_w、b_h——边界框宽和高

2.2 YOLO v3 模型改进

2.2.1 锚点框聚类与优化

YOLO v3 模型的锚点框尺寸由其在 COCO 数据集上聚类得到，与本文需要检测的奶牛爬跨行为目标尺寸并不一致，为提高模型训练效果，需设置新的锚点框尺寸。使用 K-means 算法在已标注的样本集上聚类，得到新的锚点框尺寸为 46×47、61×109、76×73、94×150、114×90、136×194、140×118、191×155、232×262。

使用新的锚点框尺寸测试发现，模型训练效果一般，分析原因为：由于本文使用的样本数据均为奶牛视频图像，样本类型单一，且样本中标注框尺寸分布集中，导致聚类产生的锚点框尺寸也过于集中，无法体现出 YOLO v3 模型的多尺度检测的输出优势，且本文标注框尺寸多大于锚点框，增加了训练开销。为改善训练效果，使用线性扩展法对锚点框尺寸进行优化，本文使用线性尺度缩放的办法将锚点框尺寸向两边拉伸，线性拉伸表达式为

$$\begin{cases} w'_{\min}=\alpha w_{\min} \\ w'_{\max}=\beta w_{\max} \\ w'=\dfrac{w-w_{\min}}{w_{\max}-w_{\min}}(w'_{\max}-w'_{\min})+w'_{\min} \quad (w\neq w_{\min}、w_{\max}) \\ h'=w'\dfrac{h}{w} \end{cases} \tag{3}$$

式中 α——最小框缩小倍数，经预试验确定 $\alpha=0.5$

β——最大框扩大倍数，经预试验确定 $\beta=1.5$

w——扩展前锚点框宽度

w'——扩展后锚点框宽度

$w_{\min}$、$w_{\max}$——扩展前锚点框宽度最小值、最大值

$w'_{\min}$、$w'_{\max}$——扩展后锚点框宽度最小值、最大值

h——扩展前锚点框高度

h'——扩展后锚点框高度

优化后的锚点框尺寸为 23×23、49×87、75×72、106×170、141×111、180×257、187×157、276×224、348×393。可以看出，优化后的锚点框尺寸分布更为分散，经测试，使用优化后的锚点框训练模型拟合速度更快。最后，将新的锚点框尺寸按感受野大、中、小分为 3 组，分别替代表 1 中 3 种特征图对应的锚点框。

2.2.2 特征提取网络的改进

本文模型以奶牛发情爬跨行为作为检测目标，由于奶牛个体偏大、个体间差距小，且各行为之间关

联，较一般的目标更难区分，为使模型取得更好的识别效果，需对特征提取网络进行改进。DenseNet[25]网络结构更深，对大目标检测能力更强，受其启发本文引入 DenseBlock 结构与 YOLO v3 模型特征提取网络融合，以提高网络大目标检测性能。

对于卷积神经网络 H，假设输入图像为 m_0，经过 L 层神经网络，每一层都实现了一个非线性变换 H_i，H_i可以是多种函数操作的组合，如：批量归一化、修正线性单元 ReLU、上下采样或卷积等。传统前馈卷积神经网络将第 i 层的输出作为 $i+1$ 层的输入，可以表示为

$$m_i = H_i(m_{i-1}) \tag{4}$$

式中　m_i——第 i 层输出的特征映射

ResNet 中增加了从输入到输出的残差(Residual)结构，可以表示为

$$m_i = H_i(m_{i-1}) + x_{i-1} \tag{5}$$

使用 ResNet 的主要优势是梯度可以经恒等映射到达前面的层，但 ResNet 处理恒等映射和非线性输出的方式是叠加，这在一定程度上破坏了网络中的信息流。

为进一步优化信息流的传播，HUANG 等[25]在设计 DenseNet 网络时提出 DenseBlock 结构，如图 4 所示，引入从任何层到后续层的直接连接，这样第 i 层得到了之前所有层的特征映射 m_0、m_1、…、m_{i-1}作为输入

$$m_i = H_i([m_0, m_1, \cdots, m_{i-1}]) \tag{6}$$

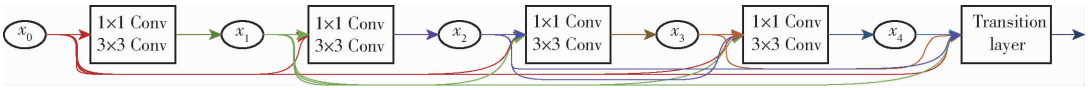


图 4　DenseBlock 结构

Fig. 4　Structure of DenseBlock

在 YOLO v3 特征提取网络后增加 2 个 DenseBlock 结构，达到加深网络的目的，增加的 DenseBlock 结构由 4 组 1×1 和 3×3 卷积层和 1 层过渡层(Transition layer)组成。其中，1×1 卷积用于降低输入特征通道数，3×3 卷积用于新的特征提取，每层的输入均是前面所有层的输出经特征通道拼接而来，过渡层的主要目的是特征融合和改变输出特征通道数。使用 DenseBlock 结构来增加网络深度而不使用残差结构的原因是 DenseBlock 结构具有更少的参数量。例如，对于通道数为 2 048 的输入特征图像，使用残差结构需要 1×1 和 3×3 卷积数量分别为 1 024 个和 2 048 个，而使用 DenseBlock 结构需要 1×1 和 3×3 卷积数量分别为 192 个和 48 个，此外，随着网络结构的加深，模型的训练更加困难，研究表明[27]，在没有预训练的前提下，DenseBlock 更易训练，这主要源于该结构密集连接的方式。改进后，YOLO v3 的特征提取网络如图 5 所示。

	类型	滤波器	尺寸	输出
	Convolutional	32	3×3	512×512
	Convolutional	64	3×3/2	256×256
1×	Convolutional	32	1×1	
	Convolutional	64	3×3	
	Residual			256×256
	Convolutional	128	3×3/2	128×128
2×	Convolutional	64	1×1	
	Convolutional	128	3×3	
	Residual			128×128
	Convolutional	256	3×3/2	64×64
8×	Convolutional	128	1×1	
	Convolutional	256	3×3	
	Residual			64×64
	Convolutional	512	3×3/2	32×32
8×	Convolutional	256	1×1	
	Convolutional	512	3×3	
	Residual			32×32
	Convolutional	1024	3×3/2	16×16
4×	Convolutional	512	1×1	
	Convolutional	1024	3×3	
	Residual			16×16
	Convolutional	2048	3×3/2	8×8
	Concatenation			8×8
4×	Convolutional	160	1×1	
	Convolutional	40	3×3	
	Avgpool	2208	2×2/2	4×4
	Concatenation			4×4
4×	Convolutional	160	1×1	
	Convolutional	40	3×3	

图 5　改进后的 YOLO v3 特征提取网络

Fig. 5　Improved YOLO v3 feature extraction network

对于改进后的特征提取网络，若输入图像尺寸为 512 像素×512 像素，则由 DenseBlock 结构输出的最小特征图尺寸为 4 像素×4 像素，是输入图像的 128 倍下采样，由于下采样倍数高，特征图感受野更大，模型对大目标检测能力更强。改进后的 YOLO v3 模型使用上采样构造特征金字塔进行多尺度特征融合，实现高低层特征信息共享，以提高模型不同尺度特征图的上下文语义。

2.2.3　边界框损失函数优化

YOLO v3 模型的损失函数由坐标误差、分类误差和置信度误差 3 部分构成，其中坐标误差使用均方差(Mean squared error，MSE)来度量，但 MSE 对目标尺度不具有不变性。预测框与真实框的交并比(Intersection over union，IoU)与 MSE 相比，具有尺度上的鲁棒性，但 IoU 作为损失函数，存在预测框与真实框不重合时梯度为 0，损失函数无法优化的问题。为使 IoU 适合作为损失函数，本文对 IoU 的计算方法做出调整，提出 F_{IoU}作为边界框回归损失函

数,其计算公式为

$$F_{IoU} = I_{IoU} - \frac{|A_c - U|}{A_c} \tag{7}$$

式中 I_{IoU}——预测框与真实框的交并比

A_c——预测框与真实框的最小闭合区域面积

U——A_c中不属于预测框和真实框的面积

F_{IoU}也具有尺度上的鲁棒性,由式(7)可知,两框不重叠时依然可以计算出 F_{IoU},解决了两框不重合梯度为 0 的问题。若将其作为边界框损失函数,则表示形式为

$$L_{IoU} = 1 - F_{IoU} \tag{8}$$

式中 L_{IoU}——使用 F_{IoU}作为度量的边界框损失函数

使用 L_{IoU}作为边界框损失函数,需同时在损失函数中加入两框距离的度量,为边界框的回归提供移动方向。受 YOLO v1[26]模型启发,使用两框的中心距离 D_c作为两框距离的度量,其计算公式为

$$D_c = (x - x')^2 + (y - y')^2 \tag{9}$$

式中 (x,y)——真实框的中心坐标

(x',y')——预测框的中心坐标

D_c——预测框与真实框的中心距离

最终的边界框损失函数表达式为

$$L = \eta_1 D_c + \eta_2 L_{IoU} \tag{10}$$

式中 L——使用 F_{IoU}和 D_c作为度量的边界框损失函数

η_1、η_2——平衡两种度量之间数值差距的权重系数,经预试验确定 $\eta_1 = 10$,$\eta_2 = 0.5$

3 试验结果与分析

3.1 试验平台

本文模型训练及验证在 CPU 为 Intel Core i9-9900KF,内存为 32 GB,操作系统为 Windows 10 的服务器上进行,使用 TensorFlow 在 GPU 上并行计算完成,试验使用的 GPU 配置为 Nvidia GeForceRTX 2080Ti,显存为 2 GB,并行计算环境为 CUDA 10.0 和 cudnn 7.6.5。试验编程语言为 Python 3.6.2。

3.2 奶牛发情行为识别训练

训练集样本 10 800 幅,验证集样本 2 880 幅。训练使用的样本批尺寸(batchsize)为 9,每 2 次迭代更新 1 次权重,训练过程采用小批量梯度下降(Mini-batch gradient descent, MBGD)方法进行优化,等效批尺寸为 18。训练共 100 个迭代周期(epoch),每个周期迭代 10 800/9 次,共计迭代 1.2×10^5次。训练的初始学习率设为 0.001,学习率调整采用 epoch-decay 策略,每训练完 1 个周期,学习率减小为原来的 0.9,最终将学习率调整到 0.000 01。

图 6 为训练过程损失值变化曲线,由图可以看出,模型在前 5 个周期迭代中损失值迅速下降,表明模型快速拟合,从第 6 个周期到第 50 个周期,损失值开始缓慢减小,在训练迭代 90 个周期后,损失值稳定在 5~6 之间,只有轻微振荡,表明模型拟合结束。

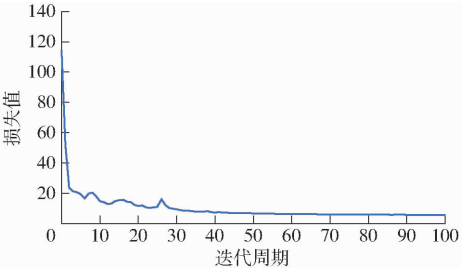


图 6 损失值随迭代周期的变化曲线

Fig. 6 Loss curve with epochs

为了防止因迭代次数过多而产生过拟合,在训练迭代 75 个周期后,每迭代 1 个周期输出 1 次权值模型,共产生 25 个模型。采用识别准确率 P、识别召回率 R、准确率和召回率综合评价指标 F1、平均精度均值(Mean average precision, mAP)作为模型性能评价指标。

在目标检测中每个类别都可以根据准确率 P 和召回率 R 绘制 $P-R$ 曲线,平均精度(Average precision, AP)是 $P-R$ 曲线与坐标轴所围面积,而 mAP 是所有类别平均精度的平均值。本文的目标检测类别为 1,因此 AP 也等于 mAP。

3.3 最优模型的确定

为更好地对模型性能进行评估,需明确各评价指标优先级。本文奶牛发情行为识别试验中,采用的评价指标优先级由大到小依次为 F1、mAP、P、R。模型训练过程中,各指标随迭代周期的变化曲线如图 7 所示。

由图 7 可知,在前 80 个周期的训练过程中,各项指标变化幅度较大,但总体趋势是增长的;在后 20 个周期的训练中,各项指标逐渐趋于稳定,在小范围内振荡。周期为 97 时,F1 有最大值,为 98.75%;周期为 83 时,mAP 最大值为 98.10%;周期为 97 时,准确率 P 最大值为 99.57%;周期为 82 时,召回率 R 最大值为 97.96%。

综合考虑 F1 和其他 3 项指标,本文最终选用第 97 个周期迭代完成后保存的模型作为奶牛发情行为识别模型,此时模型具有最高的 F1 和准确率,也具有较高的 mAP 和召回率。

3.4 识别结果分析

用筛选出的模型在测试集上进行试验,测试集

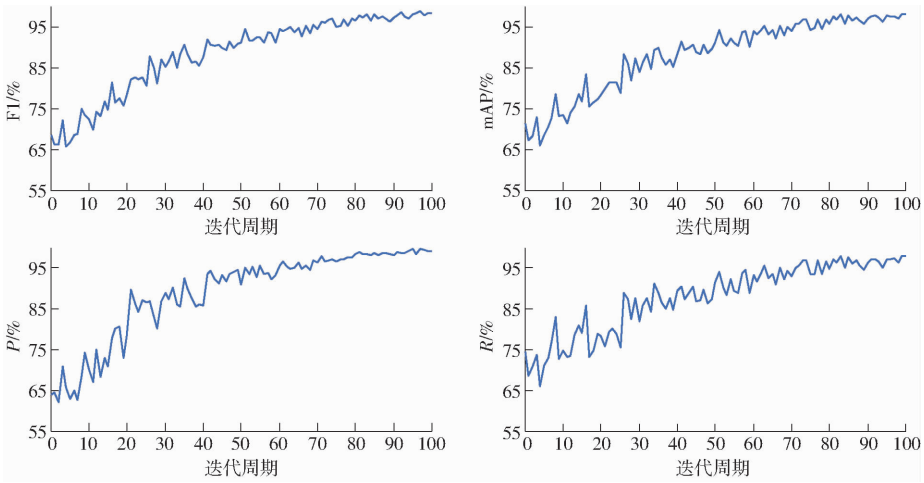


图 7　各指标随迭代周期的变化曲线

Fig. 7　Evaluation indexes variation curves with each epoch

包含 1 440 幅图像，模型共识别出 1 406 幅图像中的奶牛爬跨行为，经人工确认，其中正确识别 1 394 幅，错误识别 12 幅，识别准确率为 99.15%，有 34 幅图像中的爬跨行为未被模型识别，识别召回率为 97.62%，未被识别的主要原因为存在遮挡的爬跨行为以及少量夜间发生的爬跨行为，其中存在遮挡的爬跨行为 28 幅，夜间爬跨行为 6 幅。

对奶牛躯干的定位及检测精度，最终会影响模型的识别准确率，因此，首先对模型的奶牛躯干定位及检测精准程度进行分析。评价目标检测的效果时需先区分前景和背景，本文模型中发情奶牛躯干为目标检测的前景，图像帧的其他区域为背景。使用前景误检率 V_{ff}、背景误检率 V_{bf} 作为模型检测准确度的评价指标。

$$\begin{cases} V_{ff} = \dfrac{|A_p - A_b|}{A_p} \times 100\% \\ V_{bf} = \dfrac{|A_b - S|}{A_b} \times 100\% \end{cases} \tag{11}$$

式中　A_p——预测框区域面积

A_b——奶牛躯干最小外接矩形面积

S——预测框与奶牛躯干最小外接矩形相交区域面积

从试验结果中选取 200 幅具有有效牛身信息的识别结果进行统计，并与 YOLO v3、Faster R－CNN[27] 进行对比，如表 2 所示。由表 2 可知，本文模型前景检测的误检率为 13.28%，低于 YOLO v3 的 15.44% 和 Faster R－CNN 的 14.67%；背景误检率为 21.55%，较 YOLO v3 降低了 2.12 个百分点，较 Faster R－CNN 降低了 1.24 个百分点。前景误检率实际上表达了模型对奶牛躯干特征提取的准确性，试验结果中本文模型的前景误检率和背景误检率均低于 YOLO v3 和 Faster R－CNN，表明本文模型对奶牛躯干和非奶牛躯干特征具有更强的区分能力，模型提取到的特征更为准确。

表 2　本文模型与其他模型的奶牛躯干检测准确率

Tab. 2　Detection accuracy of cow's trunk with proposed model and other models　%

模型	前景误检率	背景误检率
本文模型	13.28	21.55
YOLO v3	15.44	23.67
Faster R－CNN	14.67	22.79

为进一步评价本文模型的识别效果，使用识别准确率 P、识别召回率 R 以及单帧识别时间作为评价指标，并与其他模型进行对比，结果如表 3 所示。

表 3　本文模型与其他模型对奶牛爬跨行为的识别效果

Tab. 3　Recognition performance of proposed model and other models on dairy cows mounting behavior

模型	识别准确率/%	识别召回率/%	单帧识别时间/s
本文模型	99.15	97.62	0.032
YOLO v3	96.52	90.34	0.025
Faster R－CNN	99.36	89.25	0.125
文献[3]模型	99.67		0.029
文献[17]模型	98.25	94.20	0.257
文献[19]模型	94.50		

由表 3 可知，本文模型比 YOLO v3 模型识别准确率提高了 2.63 个百分点，比 YOLO v3 模型召回率提高了 7.28 个百分点；与 Faster R－CNN 相比，本文模型识别准确率虽然降低了 0.21 个百分点，但召回率提高了 8.37 个百分点，单帧识别时间减少了 0.093 s。在识别对象相同或相似的情况下，本文模型比文献[3]模型识别准确率降低了 0.52 个百分点，但本文模型具有较强的迁移性，比文献[17]的模型准确率提高了 0.90 个百分点，单帧识别时间减少了 0.225 s，本文模型比文献[19]模型准确率提高了 4.65 个百分点。从识别速度上看，本文模型平均

帧率为 31 f/s,满足对奶牛发情行为识别的实时性需求。

3.5 模型性能分析

本文模型未识别的 34 幅图像中有 28 幅为奶牛发情行为存在遮挡的图像。遮挡会降低模型识别召回率的原因为:遮挡减少了模型可获取的发情爬跨行为特征,遮挡严重时模型无法提取到足够的特征信息来识别奶牛发情行为,从而发生了漏识别。图 8 为存在不同程度遮挡的奶牛发情行为识别结果,图 8a、8b 中存在部分遮挡的发情行为也能被很好识别出来,图 8c 为重度遮挡的发情行为,模型无法识别出发情行为。

(a) 轻度遮挡 (b) 中度遮挡 (c) 重度遮挡

图 8 存在不同程度遮挡的奶牛发情行为识别结果

Fig. 8 Recognition results of cows' estrous behavior with different degrees of occlusion

为了对本文模型的抗遮挡能力进行定量评估,构建存在遮挡的奶牛发情行为样本集,样本集共有 120 幅存在不同程度遮挡的奶牛发情爬跨行为图像,使用样本集对本文模型进行测试,并与 YOLO v3、Faster R-CNN 进行对比,使用 F1 作为评价指标,测试结果为:本文模型为 90.23%,YOLO v3 为 86.56%,Faster R-CNN 为 88.67%。本文模型抗遮挡能力比 YOLO v3 模型提高了 3.67 个百分点,比 Faster R-CNN 提高了 1.56 个百分点。进一步分析测试结果可知,由于奶牛发生爬跨行为时头部和背脊出现上扬,当图像中包含爬跨牛只头部和背脊信息,且爬跨行为被遮挡面积不超过 40% 时,则奶牛发情行为能较好地被本文模型所识别。

统计资料显示,奶牛夜间发情概率为 60% ~ 70%,本文模型对夜间奶牛发情也具有较好的识别效果,但仍有少量夜间发情行为未能被模型识别,测试集共有奶牛夜间发情行为图像 530 幅,其中有 6 幅图像未被识别,本文模型识别夜间发情行为的准确率为 98.87%。图 9 为奶牛夜间发情行为识别结果,由图可知,图 9a ~ 9d 中模型均能准确识别发生了奶牛爬跨行为,图 9e、9f 中模型未识别出爬跨行为。对比分析可知,出现这一情况主要是由于夜晚奶牛场光线较暗以及低亮度环境下相机记录的图像不清晰导致。后期可考虑在奶牛活动区加装照明设备或更换具有红外摄像功能的相机。

(a) 视野近处 (b) 视野远处 (c) 部分遮挡

(d) 视野远端光线较亮 (e) 视野远端光线较暗1 (f) 视野远端光线较暗2

图 9 奶牛夜间发情行为识别结果

Fig. 9 Recognition results of cows' estrous behavior at night

试验结果表明,本文模型具有较强的多尺度目标检测能力。这主要得益于特征提取网络的加深和多尺度特征融合的网络结构,本文模型通过加深特征提取网络,模型对大目标的检测性能得以提升,通过多尺度特征融合,模型增强了对不同尺度目标的检测能力。图 10a 为大目标尺寸的奶牛爬跨行为,图 10b 为目标尺寸偏小的奶牛爬跨行为,由图可知,模型均能准确识别出奶牛爬跨行为。

为进一步评价本文模型对不同尺寸目标的检测性能,构建多尺寸目标测试集,测试集共有 450 幅图像,其中小尺寸(宽:0 ~ 290 像素,高:0 ~ 150 像素)目标 102 幅,中等尺寸(宽:290 像素 ~ 660 像素,高:150 像素 ~ 380 像素)目标 189 幅,大尺寸(宽:660 像素 ~ 1 000 像素,高:380 像素 ~ 540 像素)目标 159 幅。使用测试集对本文模型进行测试,与 YOLO v3、Faster R-CNN 进行对比,以 F1 作为评价

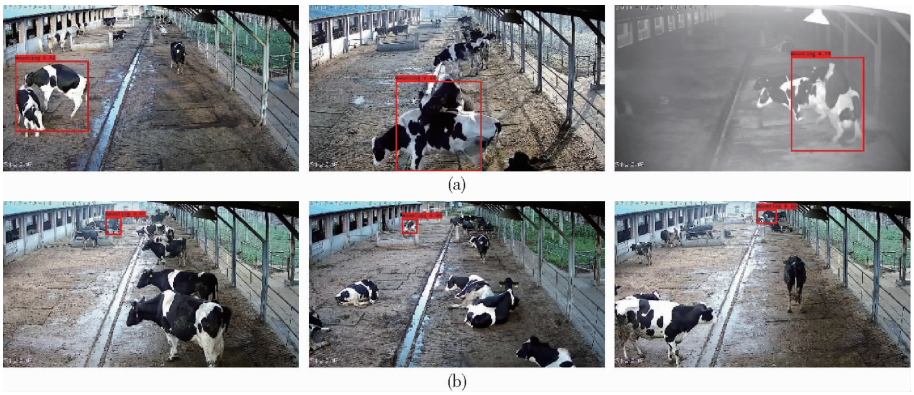


图 10　不同尺寸目标的奶牛发情行为识别结果

Fig. 10　Recognition results of different scales target of cows' estrous behavior

指标，测试结果如表 4 所示。

表 4　不同尺寸目标的奶牛发情行为识别结果对比

Tab. 4　Comparison of recognition results of different scales target of cow's estrous behavior

目标尺寸	样本数量/幅	模型	F1/%
大	159	本文模型	95.60
		YOLO v3	90.57
		Faster R－CNN	94.34
中	189	本文模型	99.47
		YOLO v3	96.29
		Faster R－CNN	97.88
小	102	本文模型	86.27
		YOLO v3	84.31
		Faster R－CNN	85.29

由表 4 可知，本文模型多尺度目标检测性能优于 YOLO v3、Faster R－CNN，尤其是在大尺寸目标检测方面，本文模型比 YOLO v3 提高了 5.03 个百分点，比 Faster R－CNN 提高了 1.26 个百分点。本文模型识别中等尺寸目标的 F1 为 99.47%，高于 Faster R－CNN 的 96.29% 和 YOLO v3 的 97.88%。本文模型识别小尺寸目标的 F1 为 86.27%，分别比 YOLO v3 和 Faster R－CNN 提高了 1.96 个百分点和 0.98 个百分点。3 种模型对小尺寸目标识别的 F1 都偏低，这是由于小尺寸目标距离摄像机较远，目标可用于提取的特征减少，模型无法获得足够的特征来识别发情行为，从而导致结果的 F1 偏低。为克服这一问题，进一步提高奶牛发情行为识别准确率，在后期研究中可考虑使用顶视角度采集奶牛活动视频。

4　结论

（1）根据奶牛发情行为数据集特点，对 YOLO v3 模型锚点框尺寸进行重新聚类和优化，同时引入 DenseBlock 改进特征提取网络，并使用由 F_{IoU} 和两框中心距离 D_c 作为度量方法的边界框损失函数，提出了一种基于改进 YOLO v3 模型的奶牛发情行为识别方法。

（2）在测试集上的试验表明，改进后的模型识别准确率为 99.15%、召回率为 97.62%，比 YOLO v3 模型识别准确率提高了 2.63 个百分点、识别召回率提高了 7.28 个百分点，模型识别的平均帧率为 31 f/s，能够满足实际养殖环境下奶牛发情行为的实时识别，与现有识别对象相同或相似的模型相比，本文模型具有较高的识别精度和较快的识别速度。

参　考　文　献

[1] 何东健，刘冬，赵凯旋. 精准畜牧业中动物信息智能感知与行为检测研究进展[J/OL]. 农业机械学报，2016，47(5)：231－244.
HE Dongjian，LIU Dong，ZHAO Kaixuan. Review of perceiving animal information and behavior in precision livestock farming [J/OL]. Transactions of the Chinese Society for Agricultural Machinery，2016，47(5)：231－244. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20160532&journal_id=jcsam. DOI：10.6041/j.issn.10001298.2016.05.032.(in Chinese)

[2] Del FRESNO M，MACCHI A，MARTI Z，et al. Application of color image segmentation to estrusc detection[J]. Journal of Visualization，2006，9(2)：171－178.

[3] TSAI D，HUANG C. A motion and image analysis method for automatic detection of estrus and mating behavior in cattle[J]. Computers and Electronics in Agriculture，2014，104：25－31.

[4] BRUYERE P，HETREAU T，PONSART C，et al. Can video cameras replace visual estrus detection in dairy cows? [J]. Theriogenology，2012，77(3)：525－530.

[5] CHUNG Y，CHOI D，CHOI H，et al. Automated detection of cattle mounting using side-view camera[J]. KSII Trans. Internet Inf. Syst.，2015，9(8)：3151－3168.

[6] BRUNASSI L D，MOURA D J，NAAS I D，et al. Improving detection of dairy cow estrus using fuzzy logic[J]. Sci. Agric.，2010，67(5)：503－509.

[7] 张子儒. 基于视频分析的奶牛发情信息检测方法研究[D]. 杨凌:西北农林科技大学,2018.
ZHANG Ziru. Research on detection method of cow estrus information based on video analysis[D]. Yangling: Northwest A&F University, 2018. (in Chinese)
[8] 顾静秋,王志海,高荣华,等. 基于融合图像与运动量的奶牛行为识别方法[J/OL]. 农业机械学报,2017,48(6):145 - 151.
GU Jingqiu, WANG Zhihai, GAO Ronghua, et al. Recognition method of cow behavior based on combination of image and activities[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(6): 145 - 151. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20170619&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2017.06.019. (in Chinese)
[9] JIANG B, WU Q, YIN X, et al. FLYOLOv3 deep learning for key parts of dairy cow body detection[J]. Computers and Electronics in Agriculture, 2019, 166:104982.
[10] ZHANG K, LI D, HUANG J, et al. Automated video behavior recognition of pigs using two-stream convolutional networks[J]. Sensors, 2020, 20(4):1085.
[11] XU B, WANG W, FALZON G, et al. Automated cattle counting using Mask R - CNN in quadcopter vision system[J]. Computers and Electronics in Agriculture, 2020, 171:105300.
[12] ZHAO K, BEWLEY J M, HE D, et al. Automatic lameness detection in dairy cattle based on leg swing analysis with an image processing technique[J]. Computers and Electronics in Agriculture, 2018, 148:226 - 236.
[13] 何东健,刘建敏,熊虹婷,等. 基于改进 YOLO v3 模型的挤奶奶牛个体识别方法[J/OL]. 农业机械学报,2020,51(4):250 - 260.
HE Dongjian, LIU Jianmin, XIONG Hongting, et al. Individual identification of dairy cows based on improved YOLO v3[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(4): 250 - 260. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20200429&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2020.04.029. (in Chinese)
[14] 杨秋妹,肖德琴,张根兴. 猪只饮水行为机器视觉自动识别[J/OL]. 农业机械学报,2018,49(6):232 - 238.
YANG Qiumei, XIAO Deqin, ZHANG Genxing. Automatic pig drinking behavior recognition with machine vision[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(6): 232 - 238. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no=20180627&flag=1. DOI:10.6041/j.issn.1000-1298.2018.06.027. (in Chinese)
[15] 康熙,张旭东,刘刚,等. 基于机器视觉的跛行奶牛牛蹄定位方法[J/OL]. 农业机械学报,2019,50(增刊):276 - 282.
KANG Xi, ZHANG Xudong, LIU Gang, et al. Hoof location method of lame dairy cows based on machine vision[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(Supp.): 276 - 282. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no=2019s043&flag=1. DOI:10.6041/j.issn.1000-1298.2019.S0.043. (in Chinese)
[16] PARK S, JUNG D, SANG H M, et al. Development of vocal recording and analysis system for laying hens and cow based-on cloud-computing[C]// 2019 ASABE Annual International Meeting. Boston, 2019.
[17] 刘忠超,何东健. 基于卷积神经网络的奶牛发情行为识别方法[J/OL]. 农业机械学报,2019,50(7):186 - 193.
LIU Zhongchao, HE Dongjian. Recognition method of cow estrus behavior based on convolutional neural network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(7): 186 - 193. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no=20190719&flag=1. DOI:10.6041/j.issn.1000-1298.2019.07.019. (in Chinese)
[18] 王少华,何东健,刘冬. 基于机器视觉的奶牛发情行为自动识别方法[J/OL]. 农业机械学报,2020,51(4):241 - 249.
WANG Shaohua, HE Dongjian, LIU Dong. Automatic recognition method of dairy cow estrus behavior based on machine vision[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(4): 241 - 249. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20200428&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2020.04.028. (in Chinese)
[19] 李丹,张凯锋,李行健,等. 基于 Mask R - CNN 的猪只爬跨行为识别[J/OL]. 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/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(Supp.): 261 - 266, 275. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=2019s041&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.S0.041. (in Chinese)
[20] 赵凯旋,何东健. 基于卷积神经网络的奶牛个体身份识别方法[J]. 农业工程学报,2015,31(5):181 - 187.
ZHAO Kaixuan, HE Dongjian. Recognition of individual dairy cattle based on convolutional neural networks[J]. Transactions of the CSAE, 2015, 31(5): 181 - 187. (in Chinese)
[21] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv:1804.02767v1, 2018.
[22] LI F. ImageNet: crowdsourcing, benchmarking & other cool things[R]. CMU VASC Seminar, March, 2010.
[23] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu, HI, USA, 2017:6517 - 6525.
[24] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas, NV, USA, 2016:770 - 778.
[25] HUANG G, LIU Z, LAURENS V D, et al. Densely connected convolutional networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu, HI, USA, 2017:2261 - 2269.
[26] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas, NV, USA, 2016:779 - 788.
[27] REN S, HE K, GIRSHICK R, et al. Faster R - CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137 - 1149.