

doi:10.6041/j.issn.1000-1298.2022.03.030

# 基于改进 YOLO v3 的肉牛多目标骨架提取方法

张宏鸣<sup>1,2</sup> 李永恒<sup>1</sup> 周利香<sup>1</sup> 汪润<sup>1</sup> 李书琴<sup>1,2</sup> 王红艳<sup>2,3</sup>(1. 西北农林科技大学信息工程学院, 陕西杨凌 712100; 2. 宁夏智慧农业产业技术协同创新中心, 银川 750004;  
3. 西部电子商务股份有限公司, 银川 750004)

**摘要:** 针对肉牛行为识别过程中,多目标骨架提取精度随目标数量增多而大幅降低的问题,提出了一种改进 YOLO v3 算法(Not classify RFB-YOLO v3, NC-YOLO v3),在主干网络后引入 RFB(Receptive field block)扩大模型感受野,剔除分类模块提高检测效率,结合 8SH(8-Stacked Hourglass)算法实现实际养殖环境下的肉牛多目标检测与骨架提取。实验为肉牛骨架设置 16 个关键节点形成肉牛骨架点位信息,通过对图像多尺度和多方向训练提高检测精度。针对多目标骨架提取模型检测的关键点信息进行统计分析,提出一种对肉牛站立和卧倒行为识别的方法。实验结果表明:在目标检测方面,NC-YOLO v3 模型的召回率可达 99.00%,精度可达 97.80%,平均精度可达 97.18%。与原模型相比,平均精度提高 4.13 个百分点,去除的网络参数量为 13.81 MB;在单牛骨架提取方面,采用 8 层堆叠的 Hourglass 网络检测关键点位置,平均精度均值可达 90.75%;在多牛骨架提取方面,NC-YOLO v3 构建的模型相对于 YOLO v3 构建的模型,平均精度均值提高 4.11 个百分点,达到 66.05%。

**关键词:** 肉牛;多目标骨架提取;目标检测;关键点检测;RFB;YOLO v3

中图分类号: TP391.4 文献标识码: A 文章编号: 1000-1298(2022)03-0285-09

OSID:



## Multi-target Skeleton Extraction Method of Beef Cattle Based on Improved YOLO v3

ZHANG Hongming<sup>1,2</sup> LI Yongheng<sup>1</sup> ZHOU Lixiang<sup>1</sup> WANG Run<sup>1</sup> LI Shuqin<sup>1,2</sup> WANG Hongyan<sup>2,3</sup>

(1. College of Information Engineering, Northwest A&amp;F University, Yangling, Shaanxi 712100, China

2. Ningxia Intelligent Agricultural Industry Technology Collaborative Innovation Center, Yinchuan 750004, China

3. West China Electronic Bussiness Co., Ltd., Yinchuan 750004, China)

**Abstract:** In view of the problem that the extraction accuracy of beef cattle skeleton was decreased greatly with the increase of targets in the process of beef cattle behavior recognition, an improved YOLO v3 algorithm (Not classify RFB-YOLO v3, NC-YOLO v3) was proposed. After the backbone network, receptive field block (RFB) was introduced to expand the receptive field of the model, and the classification module was eliminated to improve the detection efficiency. Combining 8SH (8-Stacked Hourglass) algorithm to realize multi-target detection and skeleton extraction of beef cattle in actual breeding environment. In the experiment, totally 16 key nodes were set for the beef cattle skeleton to form the beef cattle pose point information, and the detection accuracy was improved through multi-scale and multi-direction training of the image. Based on the statistical analysis of key points of multi-target skeleton extraction model, a method for beef cattle standing and lying down behavior recognition was proposed. Experimental results showed that in terms of target detection, the recall of the NC-YOLO v3 model can reach 99.00%, the precision can reach 97.80%, and the average precision can reach 97.18%. Compared with the original model, average precision was increased by 4.13 percentage points, and the amount of network parameters removed was 13.81 MB; in terms of single-ox skeleton extraction, the 8-Stacked Hourglass network was used to detect key point positions, and the mean average precision can reach 90.75%. In terms of multi cattle skeleton extraction, compared with the model constructed by YOLO v3, the mean average precision of the model constructed by NC-YOLO v3 was increased by 4.11 percentage points to 66.05%.

**Key words:** beef cattle; multi-target skeleton extraction; target detection; key point detection; receptive field block; YOLO v3

收稿日期: 2021-10-26 修回日期: 2022-01-11

基金项目: 宁夏智慧农业产业技术协同创新中心项目(2017DC53)、国家重点研发计划项目(2020YFD1100601)和国家自然科学基金项目(41771315)

作者简介: 张宏鸣(1979—),男,教授,博士,主要从事空间大数据管理与智慧农业研究,E-mail: zhm@nwsuaf.edu.cn

## 0 引言

牛的行为影响着牛的健康和产出<sup>[1-4]</sup>。通过骨架提取可以获知牛的姿态,为牛的反刍、运动、进食、站立、卧倒等行为识别<sup>[5-10]</sup>提供信息支持。基于养殖场的监控视频进行的多牛骨架提取,对大型肉牛养殖场的精细化管理与精准化饲喂具有重要意义。

多目标骨架提取<sup>[11-12]</sup>是多牛姿态信息获取的重要方法。基于深度学习的多目标骨架提取方法<sup>[13-14]</sup>分为自上而下和自下而上两类<sup>[15-16]</sup>。自上而下包括:CHEN等<sup>[17]</sup>在目标检测器后连接基于GlobalNet和RefineNet的级联金字塔网络(Cascaded pyramid network, CPN)模型进行多目标骨架提取;FANG等<sup>[18]</sup>提出了一种由SSTN(Symmetric spatial transformer network)、NMS(Non-maximum-suppression)和PGPG(Pose-guided proposals generator)组成的RMPE(Regional multi-person pose estimation)框架用于多目标骨架提取;PAPANDREOU等<sup>[19]</sup>以Faster RCNN作为目标检测器,使用ResNet检测目标的热图和偏移量,最后将二者融合得到关键点位置。

自下而上包括:CAO等<sup>[20]</sup>提出了PAFs(Part affinity fields)方法学习身体部位和对应个体之间的关联用于多人骨架提取;CHENG等<sup>[21]</sup>提出了HigherHRNet方法进行人体多目标骨架提取,缓解了传统的自下而上方法对于小尺度人体骨架提取错误率较高问题;PISHCHULIN等<sup>[22]</sup>提出的DeepCut方法首先提取关键点候选区域,对关键点进行分类,在人体多目标骨架提取任务中表现良好。

人体的多目标骨架提取技术已经相当成熟<sup>[23-24]</sup>,但对于大型动物如肉牛的多目标骨架提取技术报道相对少见。在大型动物的多目标骨架提取方面,宋怀波等<sup>[25]</sup>基于自下而上的思路,提出了基于PAFs的奶牛骨架提取模型。该模型的置信度在单目标方面为78.90%;双目标方面为67.94%;3个及以上目标方面为48.59%。该模型在多目标方面的骨架提取精度较低,有待进一步改进。

本文提出一种基于YOLO v3<sup>[26-28]</sup>和SH(Stacked Hourglass)<sup>[29]</sup>的自上而下肉牛多目标骨架提取算法,解决在一定目标数量范围内随着目标数量的增加模型精度急剧降低的问题。利用真实大型牛场监控视频,提取关键帧构建目标检测数据集和骨架提取数据集;设计16个关键点全面标识肉牛姿态情况;使用改进YOLO v3算法构建目标检测模型并结合骨架提取模型实现肉牛的多目标骨架提取,将为肉牛的精准化养殖提供技术支持。

## 1 材料和方法

### 1.1 材料

#### (1) 实验数据

选用某肉牛养殖场棚内和场外两种不同环境下的西门塔尔牛和安格斯牛实际监控视频为实验数据,摄像头位于养殖大棚上,以斜向下角度拍摄,拍摄范围为整个肉牛场地。每段肉牛视频长35 min,格式为mp4,分辨率为1920像素×1080像素,帧率为24 f/s。

#### (2) 实验环境

本文使用CentOS 7系统服务器,其硬件配置为Intel Xeon E5-2650 v4 CPU 2.20 GHz处理器和NVIDIA Corporation GP100GL [Tesla P100 PCIe 16GB] (rev a1)显卡,软件配置为Python 3.7.5编程语言、Pytorch 1.2.0深度学习框架和Cuda 10.0加速程序。

### 1.2 方法

本研究旨在使用深度学习的方法,通过处理肉牛监控视频获得肉牛骨架,加以分析进而对肉牛的站立和卧倒两种行为进行识别。技术路线如图1所示,包括4个环节:①数据集构建。从采集的视频中提取关键帧,对关键帧进行标注和剪裁,构建目标检测、骨架提取和多目标骨架提取数据集。②目标检测模型训练。利用目标检测数据集训练YOLO v3和NC-YOLO v3模型,进行评估对比。③骨架提取模型训练。利用骨架提取数据集正常训练和多尺度训练骨架提取模型,进行评估对比。

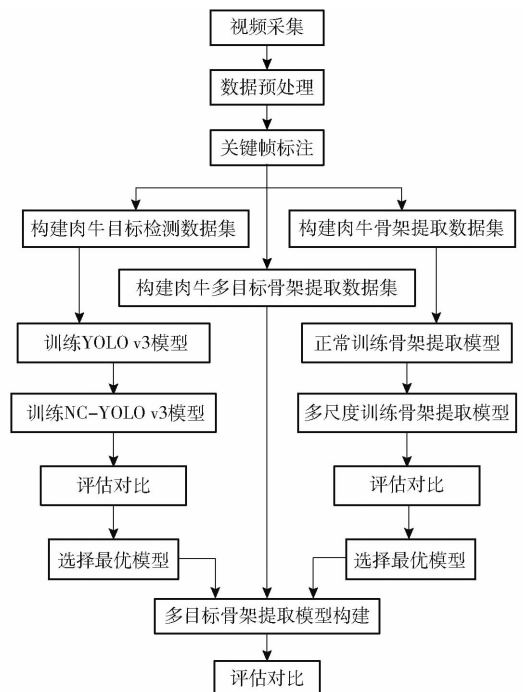


图1 技术路线图

Fig. 1 Technology road map

训练骨架提取模型,进行评估对比。④多目标骨架提取模型构建。结合目标检测和骨架提取最优模型,实现肉牛的多目标骨架提取。

### 1.2.1 数据集构建

#### (1)数据预处理

对视频进行筛选,去掉夜间、无牛等情况的视频。由于视频中相邻帧具有很高的重合度,为避免数据的重复性,提取出视频中较为清晰、目标分明、内容差异较大的帧。最后共形成 10 804 幅肉牛图像。

#### (2)肉牛体框标注

对 9 786 幅图像中所有需要检测的肉牛标注覆盖全身的体框,如图 2a 所示。



(a) 体框标注 (b) 体框和关键点标注

图 2 数据标注

Fig. 2 Data annotations

#### (3)肉牛体框和关键点标注

对 1 018 幅图像标注肉牛体框,对标注出体框的每头肉牛进行关键点标注(图 2b)。考虑到肉牛的头部、四肢和脊椎位置是肉牛姿态情况的重要组成部分,本研究着重对这 3 部分进行标注。在肉牛头部于脖子、眉间设置 2 个关键点,在肢体部分 3 个关节设置 3 个关键点,在脊椎的椎首和椎尾设置 2 个关键点,共计 16 个关键点表示肉牛姿态情况。具体关键点标注如图 3 所示。

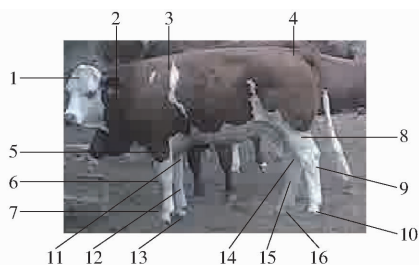


图 3 关键点设置

Fig. 3 Key point setting

1. 眉间 2. 脖子 3. 椎首 4. 椎尾 5. 前左腿根 6. 前左膝
7. 前左蹄 8. 后左腿根 9. 后左膝 10. 后左蹄 11. 前右腿根
12. 前右膝 13. 前右蹄 14. 后右腿根 15. 后右膝 16. 后右蹄

#### (4)标注数据的处理

将仅标注体框的肉牛图像制作为 COCO 格式的目标检测数据集,训练目标检测模型。将 918 幅标注体框和关键点的肉牛图像进行裁剪,制作为共有 3 414 幅单牛图像的骨架提取数据集,训练骨架提取模型。将 100 幅标注体框和关键点的肉牛图像制作为多目标骨架提取数据集,进行多目标骨架提取模

型的评估。为了验证目标数量对检测精度的影响,制作 20 幅单牛、双牛、4 牛、6 牛、10 牛和 15 牛及以上数量的多目标骨架提取数据集,多目标骨架提取数据集共 220 幅。3 种数据集名称、图像数量等如表 1 所示。

表 1 数据集图像数量分布

Tab. 1 Data set image distribution

数据集	训练集/幅	测试集/幅	比例	总数/幅
目标检测	7 829	1 957	8:2	9 786
骨架提取	2 732	682	8:2	3 414
多目标骨架提取	0	220	0:220	220

### 1.2.2 目标检测模型训练

本研究选用 YOLO v3 算法作为目标检测器。YOLO v3 是一步产生目标位置坐标和类别概率的单阶段目标检测算法,使用 Darknet53 深度神经网络作为主干部分,提取尺寸为  $Y \times Y$  的特征图。特征图中每个单元格负责检测落于此位置目标的  $A$  个体框坐标和类别概率。体框坐标是 5 维向量  $(x, y, w, h, c_{onf})$ ,其中,  $x, y, w, h, c_{onf}$  分别对应体框中心点  $x$  坐标、 $y$  坐标、体框宽度、体框高度和物体置信度。类别概率是介于 0 与 1 之间的  $n$ (类别总数)维向量,因此特征图维度为  $A \times Y \times Y \times (5 + n)$ 。Darknet53 输出 3 种不同尺寸的特征图,利用 3 个 YOLO 层分别处理,实现图像的多尺度检测,具有优秀的检测能力和准确度。由于养殖场肉牛数量密集互相遮挡,某些肉牛身体部分被遮挡导致露出部分较为分散,且监控画面中某些肉牛因距离摄像头太近而尺寸太大,模型感受野不足将会导致检测未被完全遮挡和太大尺寸肉牛的效果较差。传统的目标检测具有提取目标位置和对目标分类两种功能,而养殖场通常只存在一种养殖动物,因此,本文所用目标检测器可以不必对目标进行分类。为了加强模型的感受野和特征提取能力,同时提高目标检测效率和减小网络参数量,对 YOLO v3 模型做了如下改进:

(1)在 Darknet53 后引入 RFB 模块<sup>[30]</sup>。RFB 通过模拟人类视觉的感受野以扩大模型的感受野,提高对部分遮挡和较大尺寸肉牛的检测精度。RFB 由多分支卷积和空洞卷积构成,基于 SSD 提出的 RFBNet 的平均精度均值(mAP)达到 82.20%。RFB 模块如图 4 所示。

(2)去除网络分类功能提高目标检测效率,减小网络参数量。

改进后的网络结构如图 5 所示。训练改进前的 YOLO v3 和改进后的 NC-YOLO v3 目标检测模型,学习率均为  $1 \times 10^{-5}$ ,数据处理批次均为 16,均采用

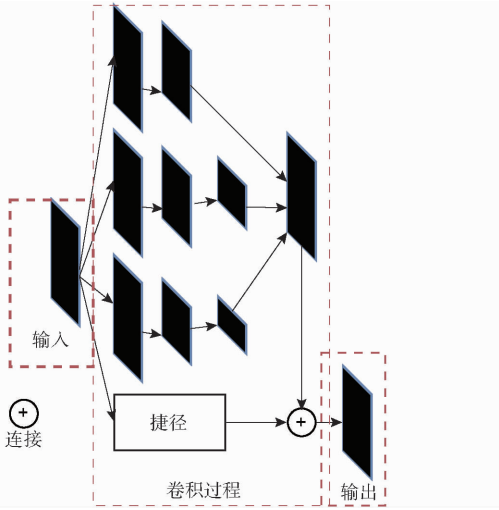


图4 RFB 结构图

Fig.4 RFB structure diagram

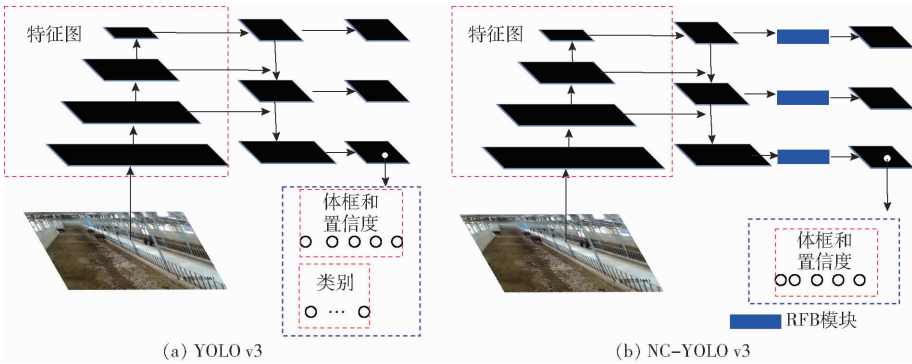


图5 YOLO v3 改进前、后结构图

Fig.5 Structure diagrams before and after improvement of YOLO v3

head length, PCKh) 达到 90.90%, 在单目标关键点检测方面的准确度较高。其由全卷积网络构成, 初级模块为残差网络, 具有高效的特征提取能力。SH 可由多个 Hourglass 网络堆叠而成, 前一个 Hourglass 的输出作为后一个的输入, 可以对关键点相对位置特征进行提取。由于肉牛的毛色复杂多样、所处环境复杂多变、四肢高度相似、遮挡严重, 数量较少的堆叠 Hourglass 网络对于细节特征感知能力较差, 通过单牛图像预测关键点位置较为困难。本文采用 8 层堆叠的 Hourglass (8SH) 网络进行关键点检测, 构建骨架提取模型。其模型结构如图 6 所示。

设置滤除置信度滤除最大置信度较小的关键点。模型输出热图中的最大值大于此置信度, 则表示该热图所预测的关键点可见, 反之, 则其不可见, 并将其滤除。

正常训练和多尺度训练骨架提取模型。训练所用数据处理批次为 4, 学习率为  $1 \times 10^{-5}$ 。正常训练所用图像尺寸为 256 像素  $\times$  256 像素, 多尺度训练每个数据处理批次图像尺寸不同, 从 192 像素  $\times$  192 像素、256 像素  $\times$  256 像素、320 像素  $\times$  320 像素 3 种尺寸中随机选择。

多尺度训练。

目标检测模型参数设置如下: 非最大性抑制阈值设为 0.5, 物体置信度设为 0.5, 数据处理批次设为 1, 图像尺寸设为 416 像素  $\times$  416 像素。通过训练过程中的损失、平均精度变化情况和目标检测多种评价指标, 对 YOLO v3 和 NC-YOLO v3 模型进行评估对比。

### 1.2.3 骨架提取模型训练

骨架提取是通过检测目标身体关键点, 然后将关键点依序连接进而形成骨架信息。本研究选用 SH 算法作为关键点检测器。SH 算法在 MPII 数据集上的正确关键点比例 (Percentage of correct key points, PCK) 达到 99.00%, 基于头部长度的正确关键点比例 (Percentage of correct key points based on

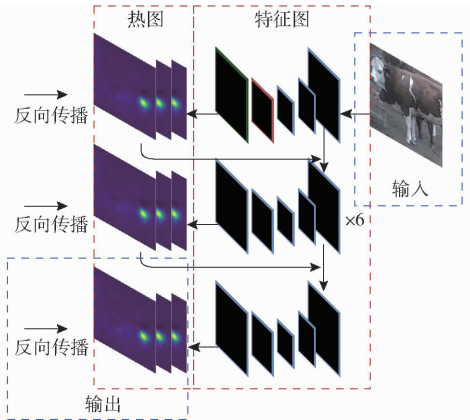


图6 8SH 结构图

Fig.6 8SH structure diagram

骨架提取模型参数设置如下: 数据处理批次设为 1, 关键点滤除置信度设为 0.5, 图像尺寸设为 256 像素  $\times$  256 像素。通过训练过程中的损失变化情况和骨架提取评价指标对 2 种不同训练方式下训练的模型进行评估对比。

### 1.2.4 多目标骨架提取模型构建

本文基于自上而下的思路实现肉牛的多目标骨架提取。训练目标检测和骨架提取模型后, 经过选择得到两个最优模型, 进而基于最优模型构建多目

标骨架提取模型。多目标骨架提取模型主要包含4个步骤,模型流程如图7所示。①通过目标检测模型检测输入图像肉牛体框。②从原图中根据肉牛体框裁剪出单牛图像。③通过骨架提取模型检测单牛图像的关键点。④将单牛图像关键点映射至原图像所在位置,依序连接关键点进而形成多牛骨架。

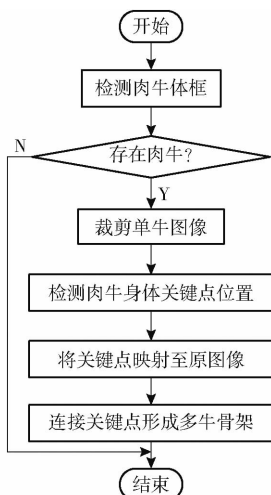


图7 多目标骨架提取模型流程图

Fig.7 Flow chart of multi-target skeleton extraction model

使用多目标骨架提取数据集,对以最优 YOLO v3 模型构建的多目标骨架提取模型和以最优 NC-YOLO v3 模型构建的多目标骨架提取模型根据评价指标进行对比。根据单牛、双牛、4牛、6牛、10牛和15牛及以上数量的数据对最优多目标骨架提取模型按照评价指标进行对比。采用非最大性抑制阈值0.5、物体置信度0.5、数据处理批次1和图像尺寸416像素×416像素进行目标检测。采用数据处理批次1、关键点滤除置信度0.5和图像尺寸256像素×256像素进行骨架提取。

## 2 评价指标

### 2.1 目标检测评价指标

衡量一个目标检测算法,通常根据其目标检测的准确度、分类的正确率和检测的目标数量进行评估。本文所用目标检测模型不需要进行分类,因此只需依据其目标检测的准确度和检测目标数量进行评估。

模型的输出为一系列的目标体框坐标,需要将其与真实坐标进行对比,得到其准确率。使用计算模型输出的体框坐标与真实体框坐标的交并比(Intersection over union, IoU)评价两者之间的相似程度。利用交并比判断模型检测结果的正确性。本文在评估目标检测模型时,设置交并比阈值为0.5,

即交并比大于0.5的预测结果认为是正确预测,反之,则认为是错误的预测结果。精度(Precision,  $P$ )即为正确的预测占有所有预测的百分比,值越大说明输出越准确。召回率(Recall,  $R$ )即为正确的预测占有所有真实框的百分比,值越大说明检测覆盖程度越好。

使用 Precision-recall (PR) 曲线计算平均精度(Average precision, AP)。AP即PR曲线与坐标轴所围的面积。AP调和了准确率和召回率,对目标检测模型进行综合衡量。在此,不需要计算目标检测的mAP。mAP为多个类别AP的平均值,本文所使用目标检测模块已经剔除了分类功能。

将目标检测模型的运行时间和模型参数文件所占存储空间作为评价指标。

### 2.2 单牛骨架提取评价指标

骨架提取模型的评估主要是评估关键点的相似度。本文中的骨架提取模型输出16幅热图,经过处理后得到16个关键点坐标,将其与真实坐标进行对比,评估骨架提取模型的效果。物体关键点相似度(Object key point similarity, OKS)是常用的骨骼关键点检测算法的评估指标。其计算公式为

$$O_{ks} = \frac{\sum_i e^{-\frac{d_i}{2\sigma_i^2} \delta}}{\sum_i \delta} \quad (1)$$

式中  $O_{ks}$ ——骨架提取中关键点相似度

$\delta$ ——关键点的可见程度,0表示不可见,1表示可见

$S$ ——单目标图像像素面积,像素

$d_i$ ——关键点  $i$  真实坐标与预测坐标的欧氏距离

$\sigma_i$ ——关键点  $i$  的归一化因子

将测试集输入模型后计算OKS,得到所有图像的OKS,人工给定一个阈值,计算出AP。

给AP设置不同阈值,得到多个AP,对多个AP求平均,最终获得mAP。

### 2.3 多牛骨架提取评价指标

使用OKS计算的mAP衡量肉牛多目标骨架提取模型,mAP从关键点的相似度和召回率综合评价多目标骨架提取模型的精度水平。

假设一幅图像中共有  $M$  头牛,多牛骨架提取模型预测出  $N$  头牛。将  $M$  头牛与预测出的  $N$  头牛计算关键点相似度,得到了一个  $M \times N$  的OKS矩阵。找出矩阵中每一行的最大值作为当前真实肉牛关键点的OKS,最终每一头真实肉牛都会得到一个OKS,设置不同阈值计算AP,最后得出mAP。

### 3 实验与结果分析

#### 3.1 目标检测模型结果分析与评估

在测试数据集上,取 IoU 为 0.5, YOLO v3 与 NC-YOLO v3 在评价指标上的对比如表 2 所示。

表 2 不同目标检测算法效果对比

Tab.2 Comparison of different target detection algorithms

模型	平均精度/%	精度/%	召回率/%	时间/ms	模型参数
					文件所占存储空间/MB
YOLO v3	93.05	92.12	95.56	67.70	234.96
NC-YOLO v3	97.18	97.80	99.00	51.40	257.31

NC-YOLO v3 的平均精度、精度、召回率分别比 YOLO v3 高 4.13、5.68、3.44 个百分点,时间比 YOLO v3 短 16.30 ms,召回率可达 99.00%,精度可达 97.80%,平均精度可达 97.18%。本文提出的 NC-YOLO v3 算法比 YOLO v3 算法效果更好。

由于添加了 RFB 模块,导致改进后模型参数量增加。改进前后模型参数量如表 3 所示。去除模型分类功能后可使模型参数量减小 13.81 MB,便于模型的存储。

表 3 改进前后模型参数量

Tab.3 Model parameters before and after improvement

模型	模型参数文件所占存储空间/MB
YOLO v3	234.96
添加 RFB 的 YOLO v3	271.12
NC-YOLO v3	257.31

损失值和平均精度变化曲线如图 8 所示。由图 8a 可知,NC-YOLO v3 的损失值收敛速度快于 YOLO v3,并且收敛后的损失值小于 YOLO v3。由图 8b 可知,NC-YOLO v3 的平均精度收敛值高于 YOLO v3。

实验结果表明,NC-YOLO v3 在 YOLO v3 的基础上提高了检测精度和检测效率,并减小了网络参数量,适合作为本文的多目标骨架提取算法的目标检测模型。

#### 3.2 骨架提取模型结果分析与评估

两种训练方式的损失值曲线如图 9 所示。

由图 9a 可知,模型训练时在第 25 次迭代开始收敛,其收敛损失值维持在 25 左右。由图 9b 可知,模型训练在第 600 次才开始收敛,其收敛损失值维持在 5 左右,表明多尺度训练模型相对于正常训练模型具有较高的精度。

在骨架提取测试数据集上,取 0.42 ~ 0.60 之间

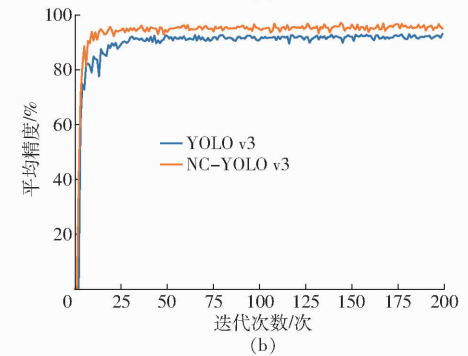
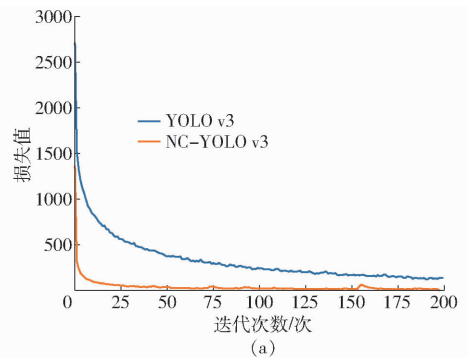


图 8 目标检测损失值和平均精度变化曲线

Fig.8 Target detection loss curves and average precision curves

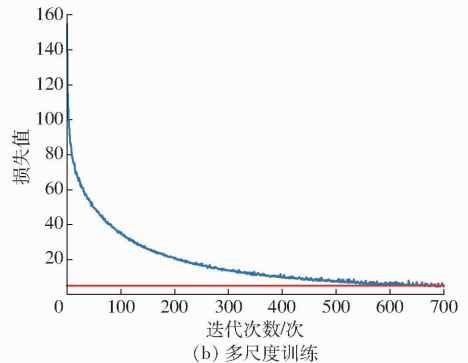
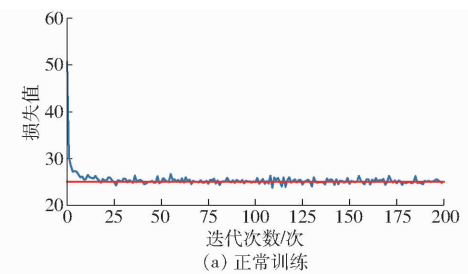


图 9 骨架提取损失值变化曲线

Fig.9 Skeleton extraction loss value change curves

共 10 个阈值,计算 mAP,结果如表 4 所示。

实验结果表明,多尺度训练下的模型,其 mAP 可达 90.75%,高于正常训练下的模型。在测试集数据上取得了较好的结果,适合作为本文多目标骨架提取算法的单牛骨架提取模块。将一幅单牛图像输入骨架提取模型,结果如图 10 所示。

#### 3.3 多牛骨架提取模型分析与评估

取 0.42 ~ 0.60 之间共 10 个阈值计算 mAP,原

表 4 不同训练方式下模型的平均精度对比  
Tab. 4 Comparison of model average precision under different training methods

训练方式	阈值										mAP
	0.42	0.44	0.46	0.48	0.50	0.52	0.54	0.56	0.58	0.60	
正常训练	91.03	90.12	87.21	85.63	89.75	72.57	69.94	64.18	59.25	51.62	76.13
多尺度训练	98.54	96.37	96.28	94.33	92.39	91.30	87.25	85.99	83.69	81.34	90.75

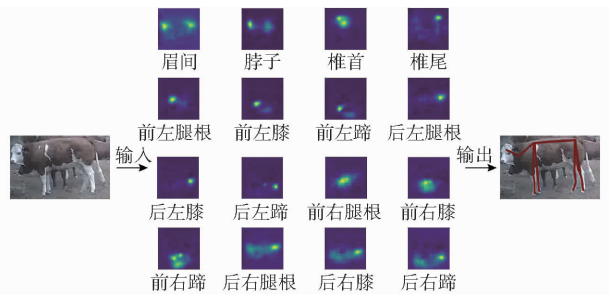


图 10 单牛骨架提取结果

Fig. 10 Results of single cattle skeleton extraction

始模型和改进模型对比结果如表 5 所示。

NC - YOLO v3 模型的 mAP 比 YOLO v3 提高了 4.11 个百分点,达到了 66.05%。

在目标数量为 1、2、4、6、10、大于等于 15 的情况下,NC - YOLO v3 构建的多牛骨架提取模型平均精度均值分别为 69.25%、70.45%、69.21%、68.49%、56.31%、49.02%。目标数量从 1 提升至 6 时,模型精度波动幅度较小,说明在这个区间内目标数量对模型精度影响较小。目标数量增至 10 头直

表 5 不同模型的平均精度对比

Tab. 5 Comparison of model average precision of different algorithms

模型	阈值										mAP
	0.42	0.44	0.46	0.48	0.50	0.52	0.54	0.56	0.58	0.60	
YOLO v3	71.12	69.58	69.14	65.11	62.00	59.31	56.21	54.49	53.65	53.33	61.94
NC - YOLO v3	78.21	75.38	71.24	69.47	67.45	65.39	62.41	58.99	56.84	55.45	66.05

至大于 15 头时,模型精度大幅降低,这是由目标太过密集导致遮挡严重、目标检测算法精度急剧下降所导致。

将相同的 3 幅图像用两种不同算法进行多目标骨架提取,结果如图 11 所示。白色方框代表误

检,黄色方框代表漏检。由此可见,本文算法对于被遮挡目标的检测体框更容易覆盖全身(如图 11 a 白色方框所示)且更逼近目标轮廓,漏检目标更少。因此本文算法更适用于肉牛多目标骨架提取。

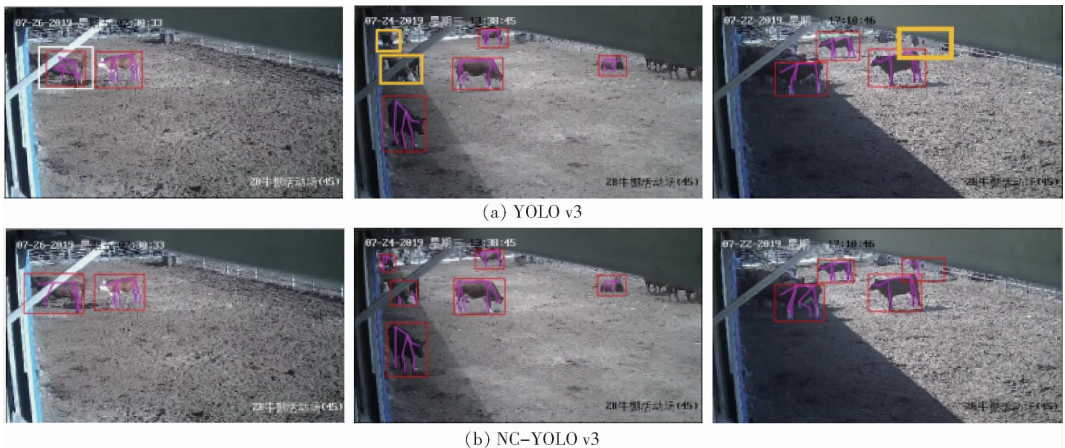


图 11 多目标骨架提取结果

Fig. 11 Multi-target skeleton extraction results

#### 4 模型识别效果验证

对肉牛多目标骨架提取模型检测出的肉牛骨架数据进行分析,实现肉牛的卧倒和站立行为的识别。

利用统计学方法,对卧倒与站立行为的肉牛关键点分布进行统计,共统计处于卧倒和站立行为肉

牛各 50 头,统计数据如表 6 所示。

识别角度是肉牛身体关键点所对应的骨架夹角,从表 6 可以发现 8 个识别角度对肉牛的卧倒和站立行为的影响较大。由于模型检测存在一定误差,根据一个或几个识别角度判断肉牛行为,会导致误差扩大,因此设计如下步骤对肉牛行为进行识别:

表6 关键点分布统计

Tab.6 Key point distribution statistics (°)

识别角度	站立	卧倒
角4 平均值	161	89
角7 平均值	158	85
角10 平均值	160	89
角13 平均值	164	90
角5 平均值	168	41
角8 平均值	177	45
角11 平均值	170	53
角14 平均值	172	47

①滤除由于相关关键点不存在而导致无法计算的识别角度。②计算识别角度,若不存在可计算的识别角度,则无法识别。③若识别角度大于或等于 $135^\circ$ 的关键点数目大于识别角度小于 $135^\circ$ 的数目,则认定肉牛处于站立行为,反之,则认定肉牛处于卧倒行为。

识别结果如图12所示,绿色方框表示肉牛处于站立状态,红色方框表示肉牛处于卧倒状态,黑色方框是由于多目标骨架提取模型检测出的关键点不足而导致无法通过此算法检测肉牛行为。由图12可知,通过分析多目标骨架提取模型检测出的肉牛骨架数据,可实现对肉牛站立与卧倒行为的识别。

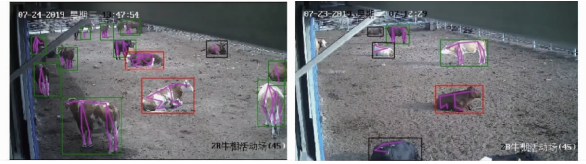


图12 肉牛行为识别结果

Fig.12 Results of beef cattle behavior recognition

## 5 结论

(1)利用养殖场监控视频,建立了肉牛目标检测数据集、肉牛单目标骨架提取数据集、肉牛多目标骨架提取测试数据集。通过对YOLO v3添加RFB、剔除分类功能,建立了NC-YOLO v3模型。实验结果表明,在肉牛目标检测数据集上,NC-YOLO v3比YOLO v3具有更高的检测精度。结合NC-YOLO v3模型与多尺度训练的8SH模型建立了肉牛多目标骨架提取模型。验证了本文模型在一定的目标数量内,随着目标数量的增多,模型精度不会大幅降低,而是小幅度波动。

(2)通过对肉牛关键点分布信息进行统计,得出了肉牛处于站立状态与卧倒状态下其识别角度的特征。通过肉牛多目标骨架提取模型检测出的肉牛骨架,计算识别角度,分析其特征进而可以识别其站立与卧倒行为。实验结果表明,本文模型可以为肉牛行为识别提供技术支持。

## 参 考 文 献

- [1] KING M T M, LEBLANC S J, PAJOR E A, et al. Cow-level associations of lameness, behavior, and milk yield of cows milked in automated systems[J]. *Journal of Dairy Science*, 2017, 100(6): 4818-4828.
- [2] 何东健, 孟凡昌, 赵凯旋, 等. 基于视频分析的犊牛基本行为识别[J]. *农业机械学报*, 2016, 47(9): 294-300.  
HE Dongjian, MENG Fanchang, ZHAO Kaixuan, et al. Recognition of calf basic behaviors based on video analysis[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2016, 47(9): 294-300. (in Chinese)
- [3] POURSAEBERI A, BAHR C, PLUK A, et al. Real-time automatic lameness detection based on back posture extraction in dairy cattle: shape analysis of cow with image processing techniques[J]. *Computers and Electronics in Agriculture*, 2010, 74(1): 110-119.
- [4] 赵圣波, 王淞, 林佳琳. 肉牛食道阻塞的临床症状和综合疗法[J]. *现代畜牧科技*, 2017(8): 126.  
ZHAO Shengbo, WANG Song, LIN Jialin. Clinical symptoms and comprehensive treatment of esophageal obstruction in beef cattle[J]. *Modern Animal Husbandry Technology*, 2017(8): 126. (in Chinese)
- [5] 宋怀波, 李通, 姜波, 等. 基于Horn-Schunck光流法的多目标反刍奶牛嘴部自动监测[J]. *农业工程学报*, 2018, 34(10): 163-171.  
SONG Huaibo, LI Tong, JIANG Bo, et al. Automatic detection of multi-target ruminant cow mouths based on Horn-Schunck optical flow algorithm[J]. *Transactions of the CSAE*, 2018, 34(10): 163-171. (in Chinese)
- [6] 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]. *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]. *Transactions of the Chinese Society for Agricultural Machinery*, 2017, 48(6): 145-151. (in Chinese)
- [7] 张宏鸣, 武杰, 李永恒, 等. 多目标肉牛进食行为识别方法研究[J]. *农业机械学报*, 2020, 51(10): 259-267.  
ZHANG Hongming, WU Jie, LI Yongheng, et al. Recognition method of feeding behavior of multi-target beef cattle[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(10): 259-267. (in Chinese)
- [8] MATTACHINI G, RIVA E, PROVOLO G. The lying and standing activity indices of dairy cows in free-stall housing[J]. *Applied Animal Behaviour Science*, 2011, 129(1): 18-27.
- [9] 宣传忠, 马彦华, 武佩, 等. 基于声信号特征加权的设施养殖羊行为分类识别[J]. *农业工程学报*, 2016, 32(19): 195-202.  
XUAN Chuazhong, MA Yanhua, WU Pei, et al. Behavior classification and recognition for facility breeding sheep based on acoustic signal weighted feature[J]. *Transactions of the CSAE*, 2016, 32(19): 195-202. (in Chinese)
- [10] 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]. *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]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(5): 231 - 244. (in Chinese)
- [11] LAXMAN K, PRERANA M. AnimePose: multi-person 3D pose estimation and animation [J]. Pattern Recognition Letters, 2021, 147: 16 - 24.
- [12] JIAN P, YA S. An improved algorithm for detection and pose estimation of texture-less objects [J]. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2021, 25(2): 204 - 212.
- [13] 周雅婷, 许童羽, 陈春玲, 等. 基于神经网络算法的肉牛采食行为检测方法 [J]. 沈阳农业大学学报, 2016, 47(6): 752 - 757. ZHOU Yating, XU Tongyu, CHEN Chunling, et al. Detection method of beef cattle feeding behavior based on neural network algorithm [J]. Journal of Shenyang Agricultural University, 2016, 47(6): 752 - 757. (in Chinese)
- [14] 田富洋, 王冉冉, 刘莫尘, 等. 基于神经网络的奶牛发情行为辨识与预测研究 [J]. 农业机械学报, 2013, 44(增刊1): 277 - 281. TIAN Fuyang, WANG Ranran, LIU Mochen, et al. Oestrus detection and prediction in dairy cows based on neural networks [J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(Supp. 1): 277 - 281. (in Chinese)
- [15] 卢健, 杨腾飞, 赵博, 等. 基于深度学习的人体姿态估计方法综述 [J/OL]. 激光与光电子学进展: 1 - 27 [2021 - 04 - 15]. <http://www.opticsjournal.net/Articles/Abstract/lop/58/24/2400005.cshtml>. LU Jian, YANG Tengfei, ZHAO Bo, et al. A review of deep learning-based human pose estimation [J/OL]. Laser & Optoelectronics Progress: 1 - 27 [2021 - 04 - 15]. <http://www.opticsjournal.net/Articles/Abstract/lop/58/24/2400005.cshtml>. (in Chinese)
- [16] 邓益依, 罗键欣, 金凤林. 基于深度学习的人体姿态估计方法综述 [J]. 计算机工程与应用, 2019, 55(19): 22 - 42. DENG Yinong, LUO Jianxin, JIN Fenglin. Overview of human pose estimation methods based on deep learning [J]. Computer Engineering and Applications, 2019, 55(19): 22 - 42. (in Chinese)
- [17] CHEN Y, WANG Z, PENG Y, et al. Cascaded pyramid network for multi-person pose estimation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7103 - 7112.
- [18] FANG H S, XIE S, TAI Y W, et al. RMPE: regional multi-person pose estimation [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 2334 - 2343.
- [19] PAPANDREOU G, ZHU T, KANAZAWA N, et al. Towards accurate multi-person pose estimation in the wild [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4903 - 4911.
- [20] CAO Z, SIMON T, WEI S E, et al. Realtime multi-person 2D pose estimation using part affinity fields [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7291 - 7299.
- [21] CHENG B, XIAO B, WANG J, et al. HigherHRNet: scale-aware representation learning for bottom-up human pose estimation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 5386 - 5395.
- [22] PISHCHULIN L, INSAFUTDINOV E, TANG S, et al. DeepCut: joint subset partition and labeling for multi person pose estimation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4929 - 4937.
- [23] QIN Z, SHELTON C R. Social grouping for multi-target tracking and head pose estimation in video [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(10): 2082 - 2095.
- [24] 罗梦诗, 徐杨, 叶星鑫. 融入双注意力的高分辨率网络人体姿态估计 [J/OL]. 计算机工程: 1 - 10 [2021 - 05 - 06]. <https://doi.org/10.19678/j.issn.1000-3428.0060493>. LUO Mengshi, XU Yang, YE Xingxin. Human pose estimation based on high resolution network with binocular attention [J/OL]. Computer Engineering: 1 - 10 [2021 - 05 - 06]. <https://doi.org/10.19678/j.issn.1000-3428.0060493>. (in Chinese)
- [25] 宋怀波, 李振宇, 吕帅朝, 等. 基于部分亲和场的行走奶牛骨架提取模型 [J]. 农业机械学报, 2020, 51(8): 203 - 213. SONG Huaibo, LI Zhenyu, LÜ Shuaichao, et al. Skeleton extraction model of walking dairy cows based on partial affinity field [J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(8): 203 - 213. (in Chinese)
- [26] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779 - 788.
- [27] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263 - 7271.
- [28] REDMON J, FARHADI A. YOLO v3: an incremental improvement [J]. arXiv preprint arXiv:1804.02767.
- [29] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation [C] // European Conference on Computer Vision, 2016: 483 - 499.
- [30] LIU S, HUANG D, WANG Y. Receptive field block net for accurate and fast object detection [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 385 - 400.