


基于 Cow - DETR 与深度图像的非接触式奶牛体质量评估

沈维政<sup>1</sup> 张哲<sup>1</sup> 戴百生<sup>1</sup> 王鑫杰<sup>1</sup> 赵凯旋<sup>2</sup> 李洋<sup>3</sup>

(1. 东北农业大学电气与信息学院, 哈尔滨 150030; 2. 河南科技大学农业装备工程学院, 洛阳 471023;  
3. 东北农业大学动物科学技术学院, 哈尔滨 150030)

**摘要:** 针对当前牧场奶牛体质量(体重)称量效率低,人工参与容易引发奶牛应激等问题,提出了一种基于改进 DETR(Detection transformer)网络的端到端式奶牛体质量评估方法(Cow - DETR),实现利用奶牛背部深度图像进行非接触式奶牛体质量评估。首先设计并搭建实验数据采集装置,利用 Intel RealSense D435 深度相机和体重秤采集奶牛背部深度图像和体质量数据;然后,通过边缘平滑滤波器和孔洞填充滤波器对深度图像进行补全处理,减少深度数据缺失对体质量评估的影响;最后,以 DETR 网络为基础建立奶牛体质量评估模型,通过在预测模块中添加含有交替全连接层的体质量预测单元,提升奶牛体质量相关的特征信息提取能力,实现端到端式奶牛背部定位的同时进行奶牛体质量非接触式评估。结果表明,本文方法可以实现较高精度的奶牛体质量评估,通过 5 倍交叉验证,在含有 139 头奶牛数据的数据集中,平均绝对误差不超过 17.21 kg,平均相对误差不超过 3.71%,单幅图像平均识别时间为 0.026 s。通过与现有体质量评估方法相对比,本文方法比其他 6 种方法在更多的奶牛头数的数据集中取得了更低的平均绝对误差和平均相对误差,同时本文方法对奶牛站立姿势要求较低,更符合牧场实际生产需要,为奶牛体质量评估提供了新的解决思路。

**关键词:** 奶牛; 体质量评估; 目标检测; 深度图像; 深度学习

中图分类号: S823; TP391.41      文献标识码: A      文章编号: 1000-1298(2023)08-0277-09      OSID: 

Non-contact Predicting Method of Dairy Cow Weight Based on Cow - DETR and Deep Image

SHEN Weizheng<sup>1</sup> ZHANG Zhe<sup>1</sup> DAI Baisheng<sup>1</sup> WANG Xinjie<sup>1</sup> ZHAO Kaixuan<sup>2</sup> LI Yang<sup>3</sup>

(1. College of Electrical Engineering and Information, Northeast Agricultural University, Harbin 150030, China  
2. College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471023, China  
3. College of Animal Science and Technology, Northeast Agricultural University, Harbin 150030, China)

**Abstract:** In order to solve the existence of some problems such as the low weighing efficiency of dairy cows in the current pasture, and being easy to cause the stress of dairy cows by manual participation, an end-to-end method of dairy cow weight estimation (Cow - DETR) based on improved detection transformer (DETR) network was proposed. The non-contact estimation on the dairy cow weight was carried out by using the depth image of dairy cow back. Firstly, a data acquisition device was designed and built, with which the cow's back depth image and weight data were collected by using the Intel RealSense D435 depth camera and the weight scale. Then, deep image data was filled by using the edge flat filter and hole filling filter to reduce the impact of deep data loss on weight estimation. Finally, by adding the weight prediction unit with an alternate fully connection layer (AFC) to the prediction module of DETR to establish a cow weight estimation model. AFC was added to improve the ability of dairy cow weight-related feature extracting. It implemented the end-to-end dairy cow back positioning while performing a non-contact estimation of dairy cow weight by Cow - DETR model. The data of 139 cows were used to evaluate the model, and the results through 5-fold cross validation showed that the weight

收稿日期: 2022 - 12 - 03    修回日期: 2023 - 02 - 11

**基金项目:** 国家自然科学基金项目(32072788、31902210、32002227)、国家重点研发计划项目(2019YFE0125600)、黑龙江省重点研发计划项目(2022ZX01A24)和财政部和农业农村部:国家现代农业产业技术体系项目(CARS36)

**作者简介:** 沈维政(1977—),男,教授,博士生导师,主要从事智慧畜牧研究,E-mail: wzhen@neau.edu.cn

**通信作者:** 戴百生(1986—),男,副教授,博士,主要从事计算机视觉和智慧畜牧研究,E-mail: bsdai@neau.edu.cn



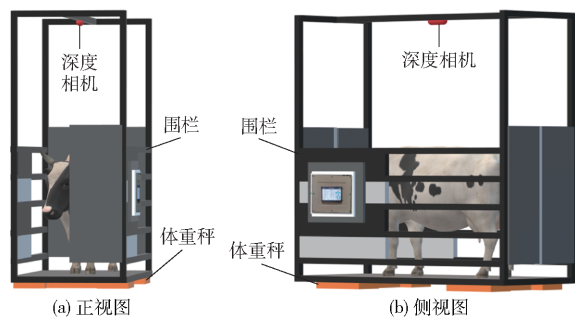


图 1 奶牛数据采集装置图

Fig. 1 Cow data acquisition device diagrams

## 1.2 数据补全处理

由于牧场环境强光和奶牛背部黑色花纹对光线的吸收等影响,奶牛背部数据产生部分不规律的残缺。如图 2a 所示,奶牛臀部和肩部出现不同大小和形状的残缺,其包含错误的深度信息,会严重影响模型的体质量评估精度<sup>[26-28]</sup>。针对这一问题,本文利用 RealSense SDK 2.0 提供的边缘平滑滤波器和孔洞填充滤波器对深度图像数据进行补全处理<sup>[29]</sup>。首先采用边缘平滑滤波器对采集图像深度噪声进行平滑处理,确保图像保持较为合理且稳定的边缘。该滤波器通过对  $x$  轴和  $y$  轴进行扫描,获取目标像素深度值,计算一维指数移动平均线 (Exponential moving average, EMA) 以确定合适的平滑量,计算公式为

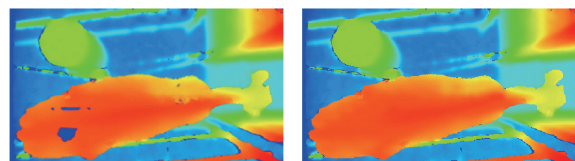
$$S_t = \begin{cases} Y_t & (t = 1) \\ \alpha Y_t + (1 - \alpha) S_{t-1} & (t > 1 \text{ 且 } \Delta = |S_t - S_{t-1}| < \delta_{\text{thresh}}) \\ Y_t & (t > 1 \text{ 且 } \Delta = |S_t - S_{t-1}| > \delta_{\text{thresh}}) \end{cases} \quad (1)$$

式中  $\alpha$ ——加权减少程度系数

$Y_t$ ——新记录的瞬时值 (视差或深度)

$S_{t-1}$ ——时间段  $t - 1$  的 EMA 值

$\delta_{\text{thresh}}$ ——深度阈值



(a) 原始深度图像 (b) 滤波后深度图像

图 2 深度图像滤波效果

Fig. 2 Depth image filtering effect diagrams

接下来进行空间域的孔洞填充,由于 D435 深度相机采用左成像器作为立体匹配算法的参考,因此本研究采用左侧有效像素 (Left valid pixel) 对深度数据缺失部分进行空间域的孔洞填充;最后进行时间域的孔洞填充,采用与边缘平滑滤波器相同类型的指数移动平均线滤波器,通过对深度图像的像

素在时间域上应用指数滑动平均来进行孔洞填充,经过多次试验,系数  $\alpha$  取 0.5、阈值参数  $\delta_{\text{thresh}}$  取 20 时可取得较好效果。最终补全效果如图 2b 所示,图 2a 中背部缺失部分得到有效填充,取得较完整的奶牛背部深度数据。运用上述方法对所有采集到的原始图像进行处理,得到用于模型训练的深度图像数据集。

## 1.3 数据标注与数据增强

为了提升本实验算法的可重复性和实验结果的可信度,本研究模型均采用 5 倍交叉验证的方式进行训练和测试。将采集到的 139 头奶牛数据,共计 14 421 幅图像,根据奶牛数量平均分成 5 份,每份包含约 20% 样本。选择其中 4 份约 111 头奶牛作训练集,其余作测试集,为方便网络模型训练,本实验将体重秤采集到的体质量数据写入文件名中,文件命名格式为:index\_weights\_.png,如:000001\_596\_.png。本实验需要进行两个类别的目标分类检测,分别为奶牛“cow”和默认背景类“background”,采用 LabelImg 标记软件进行标注,为了减少奶牛头部对实验结果的影响,标注时标注框不包含牛头部,并将图像标注内容记录到 XML 格式文件中,包括图像索引、图像大小、对象类别、位置坐标和体质量真值。

为了丰富训练集的图像数据,避免过拟合现象的发生,本文通过镜像和旋转来扩大训练数据集,其中镜像的对称轴为图像中心线。经过镜像和旋转处理后,训练集图像总数扩大近一倍。将训练集和测试集名称写入 TXT 格式文件中,用于后续奶牛体质量评估模型的训练。

## 1.4 Cow - DETR 奶牛体质量评估模型

为了实现奶牛目标检测的同时进行奶牛个体体质量评估,并且减少检测框内除奶牛背部外的背景等杂物对体质量评估的影响,本研究提出一种改进的 DETR 网络的端到端式奶牛体质量评估模型——Cow - DETR。该网络包含 3 个模块,分别为骨干网络模块、Transformer 编码器与解码器模块、预测模块。其中骨干网络模块为基础特征提取模块,负责对图像进行特征提取;Transformer 编码器与解码器模块采用注意力机制,对骨干网络模块的输出进行特征提取,通过注意力机制,能够关注奶牛背部区域,减少背景中栏杆、粪便等杂物信息的干扰,提高后续预测模块的预测精度;预测模块在 DETR 网络原有基础之上,增加了含有交替全连接层的体质量预测单元,接受 Transformer 编码器与解码器模块传来的 Token,对其进行检测框预测、类别预测和体质量预测,实现了高精度牛背目标检测的同时预测奶牛体质量。整体网络结构如图 3 所示。

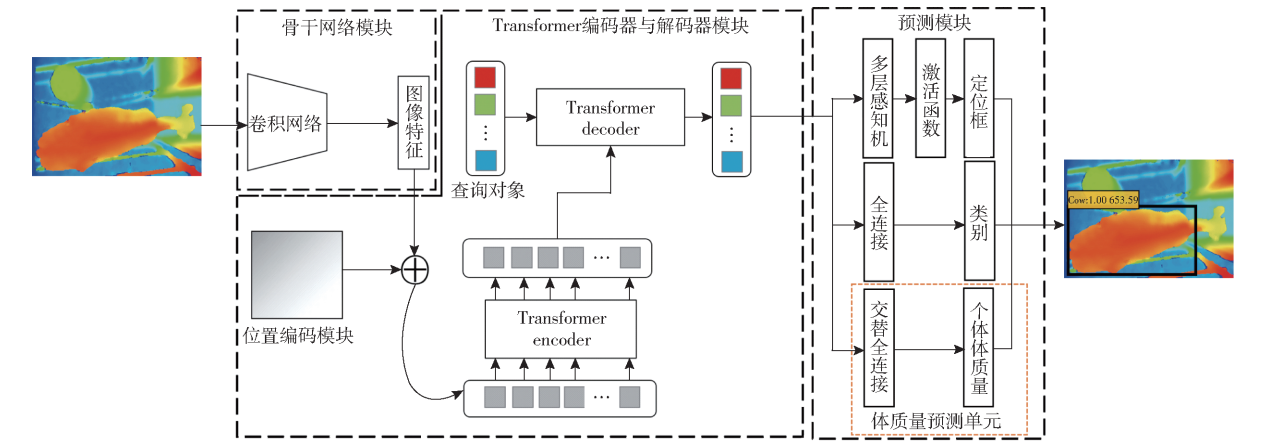


图 3 整体网络结构

Fig. 3 Structure of neural network

1.4.1 骨干网络模块

骨干网络为基础特征提取模块,通过对深度图像进行卷积,获得图像特征向量。为了更好地提取特征同时避免随机梯度爆炸、消失等问题,本研究选用 ResNet-50 作为基础特征提取网络,具体结构如表 1 所示。

表 1 ResNet-50 网络结构

Tab. 1 ResNet-50 network structure

层数	输出尺寸/(像素×像素)	网络结构
C1	424×240	7×7,64,步长为2 3×3 max pool,步长为2
C2	212×120	$\begin{bmatrix} 1\times1,64 \\ 3\times3,64 \\ 1\times1,256 \end{bmatrix} \times 3$
C3	106×60	$\begin{bmatrix} 1\times1,128 \\ 3\times3,128 \\ 1\times1,512 \end{bmatrix} \times 4$
C4	53×30	$\begin{bmatrix} 1\times1,256 \\ 3\times3,256 \\ 1\times1,1024 \end{bmatrix} \times 6$
C5	26×15	$\begin{bmatrix} 1\times1,512 \\ 3\times3,512 \\ 1\times1,2048 \end{bmatrix} \times 3$

通过卷积神经网络从输入图像  $X_{\text{png}} \in \mathbf{R}^{1 \times H_0 \times W_0}$  (单通道数据) 中获取低分辨率的特征映射  $f \in \mathbf{R}^{C \times H \times W}$  ( $C=2\,048, H=\frac{H_0}{32}, W=\frac{W_0}{32}$ )。之后将输出特征进行  $1 \times 1$  卷积降维,展平处理以得到特征序列,并输入到 Transformer 编码器与解码器模块。

1.4.2 Transformer 编码器与解码器

编码器模块包含 6 层,每层由一个多头自注意力层和一个前馈网络(Feed forward network,FFN)组成,由于编码器对排列数字不敏感,因此需要对特征序列中的向量进行位置编码,将位置信息和特征向量一并输入编码器,编码器通过对骨干网络的输出

进行特征提取,利用注意力机制,更好地关注奶牛背部相关的特征信息,进而减少背景中栏杆等杂物对后续预测的干扰,结构如图 4 左侧所示。

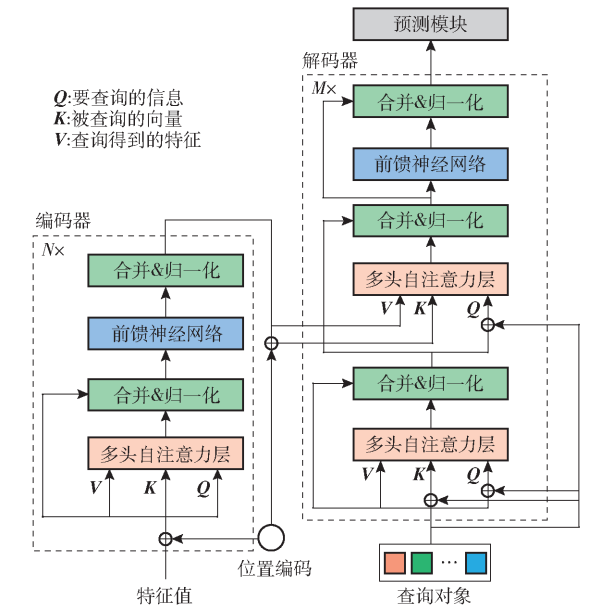


图 4 编码器与解码器结构图

Fig. 4 Image of transformer encoder and decoder

解码器结构和编码器类似,通过对编码器输出的特征图进行特征提取,将编码器的输出和  $N$  个查询对象作为输入,同时将固定数量的位置编码嵌入到  $N$  个查询对象中,对  $N$  个查询对象根据编码器提取到的特征进行重构。输出  $N$  个包含不同信息的 Token,通过将  $N$  个 Token 输入后续的预测模块,进行检测框预测、目标分类预测和体质量预测,结构如图 4 右侧所示。

1.4.3 预测模块设计

预测模块部分由 3 个并行的全连接模块构成,其共享输入特征,分别输出检测框坐标,类别置信度和奶牛体质量归一化值。该模块接受编码器解码器输出的  $N$  个 Token,对不同 Token 分别进行预测,具体网络结构如图 5 所示。其中检测框预测

单元通过多层感知机 (Multilayer perceptron, MLP) 与 Sigmoid 函数结合的方式实现,输出神经元个数设置为 4,最终输出检测框的中心坐标 ( $x$  和  $y$ ) 以及检测框的长和宽 ( $h$  和  $w$ )。目标分类预测单元

通过单个全连接层 (Fully connected, FC) 实现,由于本研究是背景和奶牛背部的二分类预测,预测单元输出维度设置为 2,最终输出的类别为奶牛或背景。

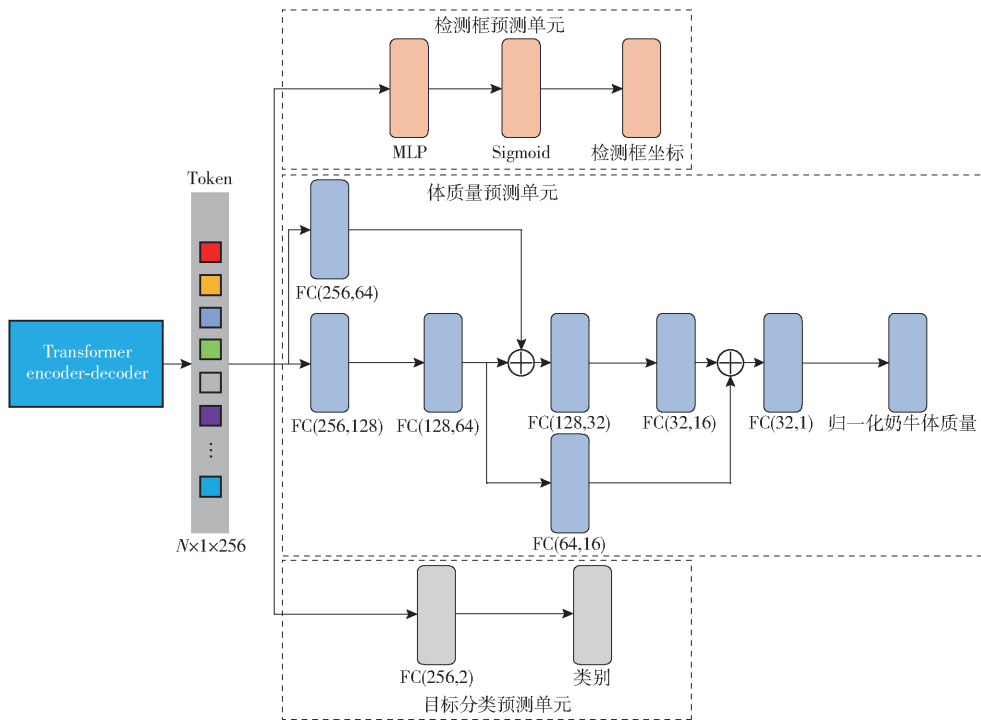


图 5 预测模块网络结构图

Fig. 5 Image of prediction module network structure

对于体质量预测单元,本研究引入全连接层构建回归分支进行体质量预测,为了更好地提取奶牛体质量相关特征信息,以提高体质量预测的精度,本文构建了一个全新的交替全连接层,用于体质量预测,如图 5 中体质量预测单元所示,模块输入是 256 维数据,首先分为 2 个分支,一个分支是 256 维作为输入、64 维作为输出的单全连接层,另一个是 256 维作为输入、64 维作为输出的双全连接层。然后将双全连接层的输出作为输入,输入到 64 维作为输入、16 维作为输出的单全连接层。同时将 2 个 64 维的输出进行结合,作为后续 128 维作为输入、16 维作为输出的双全连接层的输入。最后将 2 个输出为 16 维的全连接层的输出结合,输入到最后一个全连接层,进行归一化奶牛体质量的预测。与单个全连接层直接预测归一化奶牛体质量相比,交替全连接层可以分层次进行特征提取,同时结合上层特征,有利于得到准确的回归预测。

由于本研究采集的奶牛体质量数据范围跨度较大,这样未经处理的数据输入到神经网络会对体质量评估造成干扰。为了减少干扰,加快网络的收敛速度,提高模型训练效率,本研究将奶牛图像对应的真实体质量标签输入到网络训练之前,对其进行归一化处理,使其映射到  $[0,1]$ ,计算公式为

$$Y^{(i)} = \frac{y^{(i)} - y_{\min}}{y_{\max} - y_{\min}} \tag{2}$$

式中  $Y^{(i)}$ ——奶牛  $i$  真实体质量归一化值  
 $y^{(i)}$ ——奶牛  $i$  的体质量真值  
 $y_{\min}$ ——测试集奶牛体质量最小值  
 $y_{\max}$ ——测试集奶牛体质量最大值

体质量估计值为网络输出层输出值经过反归一化计算所得的数值,计算公式为

$$\hat{y} = \hat{Y}(y_{\max} - y_{\min}) + y_{\min} \tag{3}$$

式中  $\hat{Y}$ ——网络预测的奶牛体质量归一化值  
 $\hat{y}$ ——网络预测的奶牛体质量反归一化值

### 1.5 损失函数设计

损失函数  $L$  主要由检测框预测损失  $L_{\text{box}}$ 、目标分类预测损失  $L_{\text{label}}$  和体质量预测损失  $L_{\text{weight}}$  组成。本研究的体质量评估模型采用联合训练方式,当真实标签和候选区域之间的 IOU (交并比) 大于 0.7 时,该 Token 被认为是正例。当且仅当 Token 为正例,该特征值才会进行损失函数计算,计算公式为

$$L = \lambda_1 L_{\text{box}} + \lambda_2 L_{\text{label}} + \lambda_3 L_{\text{weight}} \tag{4}$$

式中  $\lambda_1, \lambda_2, \lambda_3$ ——损失函数权重

#### 1.5.1 检测框预测损失

本研究的检测框损失函数由损失函数  $L_1$  和  $L_{\text{iou}}$



组成,  $L_1$  不仅提高检测准确率、避免梯度爆炸还可以快速收敛到更高的精度, 提高模型的鲁棒性。但是  $L_1$  在求解预测框时仅关注 4 个点的损失值, 并没有关注 4 个值之间的关联性, 不能反映真实框和预测框之间的包含关系, 为此本研究引入损失函数  $L_{iou}$ , 将预测框作为一个整体进行回归。具体损失函数计算公式为

$$L_{\text{box}}(b_{\sigma(i)}, \hat{b}_i) = \lambda_{\text{iou}} L_{\text{iou}}(b_{\sigma(i)}, \hat{b}_i) + \lambda_{L_1} \|b_{\sigma(i)} - \hat{b}_i\|_1$$

(5)

$$L_{\text{iou}}(b_{\sigma(i)}, \hat{b}_i) = 1 - \left( \frac{|b_{\sigma(i)} \cap \hat{b}_i|}{|b_{\sigma(i)} \cup \hat{b}_i|} - \frac{|B(b_{\sigma(i)}, \hat{b}_i) \setminus b_{\sigma(i)} \cup \hat{b}_i|}{|B(b_{\sigma(i)}, \hat{b}_i)|} \right)$$

(6)

式中  $\lambda_{\text{iou}}, \lambda_{L_1}$ ——权重参数  
 $b_{\sigma(i)}$ ——第  $i$  个索引的真实框  
 $\hat{b}_i$ ——第  $i$  个索引的预测框  
 $B(b_{\sigma(i)}, \hat{b}_i)$ —— $b_{\sigma(i)}$  和  $\hat{b}_i$  的最小包围矩形  
 $B(b_{\sigma(i)}, \hat{b}_i) \setminus b_{\sigma(i)} \cup \hat{b}_i$ —— $B(\cdot)$  中没有覆盖  $b_{\sigma(i)}$  和  $\hat{b}_i$  的面积

1.5.2 目标分类预测损失

本研究的目标分类损失  $L_{\text{label}}$  采用交叉熵损失函数, 交叉熵损失函数可以很好地描述输出值与真实值之间的距离, 通过不断学习、优化模型预测的每个类别的概率与真实标签类别之间的距离, 以达到正确分类的目的, 具体计算公式为

$$L_{\text{label}} = - \sum_{i=0}^n \lg(\hat{p}_{\sigma(i)} c_i)$$

(7)

式中  $n$ ——测量样本总数  
 $\hat{p}_{\sigma(i)}$ ——预测类别输出  
 $c_i$ ——真实类别输出

1.5.3 体质量预测损失

体质量预测损失采用传统的损失函数  $L_1$ , 该损失函数有稳定的梯度, 不会导致梯度爆炸问题, 具有较为稳健性的解, 常用于回归预测。具体损失函数计算公式为

$$L_{\text{weight}} = \frac{1}{n} \sum_{i=1}^n |\hat{Y}^{(i)} - Y^{(i)}|$$

(8)

式中  $\hat{Y}^{(i)}$ ——第  $i$  个奶牛体质量预测归一化值

2 实验设置

2.1 实验环境

实验在 Linux 系统环境下运行, 采用能够支持 GPU 加速和动态神经网络的深度学习开源框架 Pytorch 1.10, 配合 CUDA 11.0 进行训练。服务器搭载的处理器为 Intel Xeon E5 - 2678V3 CPU,

2.50 GHz, 48 核, 内存为 270 GB, 显卡为 RTX3090 × 8。

2.2 参数设置

本研究通过解析视频流文件, 获取到 848 像素 × 480 像素、16 位、单通道、PNG 格式的图像数据作为数据集, 将采集到的 139 头奶牛数据, 按奶牛头数平均分成 5 份, 选择其中 4 份作为训练集, 其余奶牛数据组成测试集, 采用 5 倍交叉验证的方式进行训练和测试。批量大小 (Batch size) 在显卡内存允许范围内设置得尽可能大, 模型迭代次数共计 60 周期。优化器为 Adam, 其中骨干网络模块的学习速率为  $1 \times 10^{-5}$ , 其他模块学习速率为  $1 \times 10^{-4}$ , 权值衰减为  $1 \times 10^{-5}$ , 解码器的查询对象设为 100。根据经验损失函数的权重  $\lambda_1, \lambda_2, \lambda_3$  分别设置为 5、1、1。

训练网络使用在 COCO 数据集上预训练的网络模型进行迁移学习, 参数在预训练模型的基础上根据本实验图像数据进行微调, 通过此方式进行训练可以有效减少总体需要训练的迭代次数, 大约在 40 次迭代后, 模型收敛。

2.3 评价指标

为了评估本文方法并与其他非接触式体质量评估模型进行对比, 本文采用平均绝对误差 (Mean absolute error, MAE) 和平均相对误差 (Mean relative error, MRE) 作为判断标准。此外, 本文还利用单幅图像平均预测时间, 来评估算法的时效性。

2.4 交叉验证

为了评估本文模型的性能, 提升本文实验算法的可重复性和实验结果的可信度, 本文模型均采用 5 倍交叉验证的方式进行训练和测试, 首先将采集到的 139 头奶牛数据, 共计约 14 421 幅图像, 按奶牛头数平均分成 5 份, 每份包含大约 20% 的样本。每次实验顺序选择其中 4 份约 111 头奶牛作为训练集, 其余作为测试集, 数据集划分结果如图 6 所示。利用训练集对模型进行训练, 并在对应的测试集中进行测试, 共进行 5 次训练和测试, 将 5 次测试集的测试结果的平均值作为该模型的最终预测结果, 就可以得到整个数据集的预测结果。

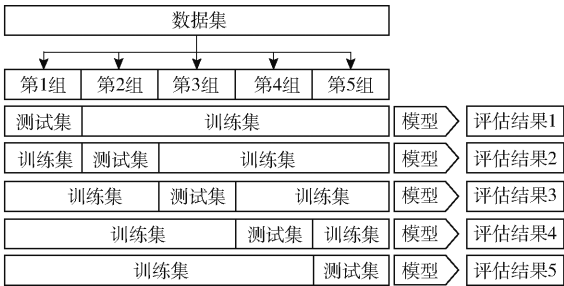


图 6 交叉验证数据集划分

Fig. 6 Cross-verification data set division

### 3 结果与分析

#### 3.1 交替全连接层有效性验证

为了验证本文提出的交替全连接层对奶牛体质量评估精度提升的有效性,使用本文实验构建的数据集对以下模型进行训练和测试:体质量预测模块为交替全连接层的 Cow - DETR 模型、体质量预测模块为单全连接层的 Cow - DETR (w/o AFC)模型、文献[23]所提出的基于 Faster R - CNN 的体质量评估模型与其体质量预测模块替换为交替全连接层的改进 Faster R - CNN (w/AFC)体质量评估模型,文献[24]所提出的基于 BotNet 的体质量评估模型与其体质量预测模块替换为交替全连接层的改进 BotNet (w/AFC)体质量评估模型。测试比对结果如表 2 所示。

表 2 不同体质量评估模型性能对比

Tab.2 Results of test dataset

体质量评估网络	MAE/ kg	MRE/%	单幅图像平均 预测时间/s
Faster R - CNN	26.83	5.99	0.025
Faster R - CNN (w/AFC)	24.32	5.04	0.029
BotNet	22.87	4.75	0.026
BotNet (w/AFC)	20.51	4.21	0.028
Cow - DETR (w/o AFC)	22.11	4.53	0.022
Cow - DETR	17.21	3.71	0.026

从表 2 可以看出,交替全连接层可以大大减少 MAE,有效提高体质量评估精度,其中 Faster R - CNN (w/AFC)的 MAE 为 24.32 kg,相比 Faster R - CNN 提升 9.36%,MRE 为 5.04%,相比 Faster R - CNN 减少 0.95 个百分点;BotNet (w/AFC)的 MAE 为 20.51 kg,相比 BotNet 提升 10.32%,MRE 为 4.21%,相比 BotNet 减少 0.54 个百分点;体质量预测模块为交替全连接层的 Cow - DETR 体质量评估效果最好,MAE 为 17.21 kg,与 Cow - DETR (w/o AFC)相比,精度提升 22.16%,MRE 为 3.71%,相对于 Cow - DETR (w/o AFC)减少 0.82 个百分点。Cow - DETR 与 Faster R - CNN (w/AFC)、BotNet (w/AFC)相比,MAE 分别提升约 29.24% 和 16.09%;MRE 分别减少 1.33、0.5 个百分点,主要原因是本文提出的 Cow - DETR 模型采用注意力机制,可以更加关注奶牛背部深度信息,有效地避免除目标奶牛背部外的栏杆、粪便等杂物信息对体质量评估精度的影响。

综上,交替全连接层可以提升奶牛特征提取的能力,有效地提高体质量评估精度,本文提出的 Cow - DETR 在 4 种网络模型中具有最低的 MAE 和 MRE,单幅图像平均预测时间为 0.026 s,可以实现

奶牛体质量实时且较高精度评估。

#### 3.2 与分段式体质量评估模型对比

为了对比端到端的体质量评估方法与分段式的体质量评估方法的效果,本研究设计了分段式体质量评估模型对比实验,首先通过单独训练 DETR 目标检测网络进行个体目标检测,进而基于目标检测结果手工截取检测框内的奶牛深度图像;然后基于截取得到的深度图像数据集,设计单独的回归神经网络,预测奶牛体质量。与本文提出的 Cow - DETR 端到端式奶牛体质量评估模型进行对比,经过 5 倍交叉验证,在测试集中对比结果如表 3 所示。

表 3 与分段式体质量评估模型性能对比

Tab.3 Comparison with segmented weights evaluation

体质量评估网络	model performance		
	MAE/ kg	MRE/ %	单幅图像平均 预测时间/s
DETR + Regression Network	28.50	5.96	0.200
Cow - DETR	17.21	3.71	0.026

从表 3 可以看出,Cow - DETR 与 DETR + Regression Network 相比,其 MAE 提升 39.61%,MRE 减少 2.25 个百分点。主要原因是,本文提出的 Cow - DETR 奶牛体质量评估模型,采用含有注意力机制的端到端的预测方式,相比于分阶段式体质量评估模型单独训练的 DETR 目标检测网络,本文方法可以利用体质量信号对 DETR 网络进行监督训练,使网络的注意力模块更加关注奶牛体质量之间的差异,更好地对奶牛特征进行提取和对体质量进行预测,且单幅图像平均预测时间更短,符合牧场实际需求<sup>[30]</sup>。

#### 3.3 与现有其他体质量评估模型对比

为了体现本文提出的 Cow - DETR 相比现有奶牛体质量评估模型的优势,将本文模型与现有奶牛体质量评估模型进行对比,对比结果如表 4 所示。

表 4 与其他奶牛体质量评估模型对比

Tab.4 Comparison with other cow weights evaluation

体质量评估模型	models		
	MAE/ kg	测试集 数量/头	数据集 数量/头
Mask R - CNN + CNN <sup>[9]</sup>	20	12	63
基于 RGB 图像体尺预测 + 多元 线性回归 <sup>[10]</sup>	17.5	24	24
三维重建 + 二元线性回归 <sup>[18]</sup>	23.34	29	29
三维重建 + 多元线性回归 <sup>[19]</sup>	18.2	6	6
基于三维点云体尺预测 + 偏最 小二乘法回归 <sup>[20]</sup>	24	10	10
Cow - DETR	17.21	139	139

由表 4 可以看出,文献[10]和文献[19]相对于

其他现有体质量评估模型,其预测精度较高,但是其采集奶牛数据集数量相对较少,本实验使用的数据集共计 139 头奶牛,并采用 5 倍交叉验证的方式进行训练和测试,将 5 次测试集的测试结果的平均值作为该模型的最终预测结果,可以得到整个数据集的测试结果,且本研究在测试集中的 MAE 为 17.21 kg,与文献[10]和文献[19]提出的方法相比分别提升 1.66% 和 5.44%。

值得注意的是,表 4 中的现有体质量评估模型<sup>[9-10,18-20]</sup>都使用了分阶段的方式预测奶牛体质量,该方式首先通过第 1 阶段的网络预测对奶牛体尺或面积等特征值进行提取,之后利用上述特征值单独训练体质量回归模块进行奶牛体质量评估。但是,前段网络提取的特征容易产生不可避免的误差,影响后续体质量评估精度。而本文提出的 Cow-DETR 奶牛体质量评估模型,采用端到端的预测方式,避免了分段式的误差问题,再次验证了本文所提出的奶牛体质量评估模型可以进行较高精度的奶牛体质量评估。

3.4 Cow-DETR 体质量评估模型评估结果与分析

为了进一步分析本文所提出的体质量评估模型的预测精度,对该模型在 5 个不同测试集上的 5 次评估结果进行分析。本研究采用 5 倍交叉验证的方式进行训练和测试,可以得到整个数据集的测试结果,因此测试集共 139 头牛,共计 14 421 幅图像。本研究给出数据集中每头奶牛的体质量估计值和真实值,如图 7 所示。为了更清楚地描述本实验模型体质量评估值和真实值之间的接近程度,本研究在测试集中随机选取 800 幅图像,并给出了每幅图像体质量评估值和真实值之间的绝对误差和相对误差,误差曲线如图 8 所示。图 8 表明,体质量评估的绝对误差一般小于 23 kg(用红线标注),仅有少数测量误差较大,但也小于 29 kg;体质量评估的相对误差较低,一般小于 4.8%(用红线标注),主要在 3.6% 上下浮动,绝对误差较大的图像对应的相对误差也不超过 5.2%。通过计算 14 421 幅测试图像的体质量评估结果,平均绝对误差为 17.21 kg,平均相对误差为 3.71%,结果表明,本文所提出的奶牛体质量评估模型的体质量评估值与真实值相近,模型预测精度较高。图 9 显示了本文模型在真实奶牛活动场景中目标检测和体质量评估效果,其中“Cow”表示待测对象的类别,类别对应的数值表示类别预测的概率,最后数值表示体质量评估结果。本文提出体质量评估模型对待测对象站立姿势和奶牛站立方向要求较低,仅需奶牛完整的出现在摄像区域,就可以检测出目标类别,并对其进行定位,以及预测出

目标的体质量参数。待测对象的类别及其概率以及相应体质量评估值在图中目标检测框的左上角区域进行显示。通过与真实值进行对比证明,本文所提出的模型可以同时实现奶牛背部定位和体质量评估且评估结果较为准确。

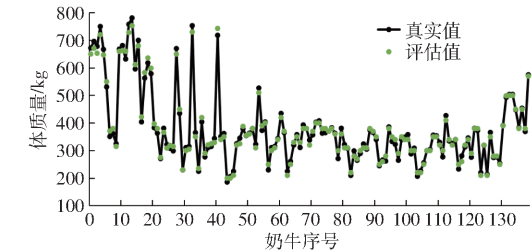


图 7 测试集检测结果

Fig.7 Detection results of test set

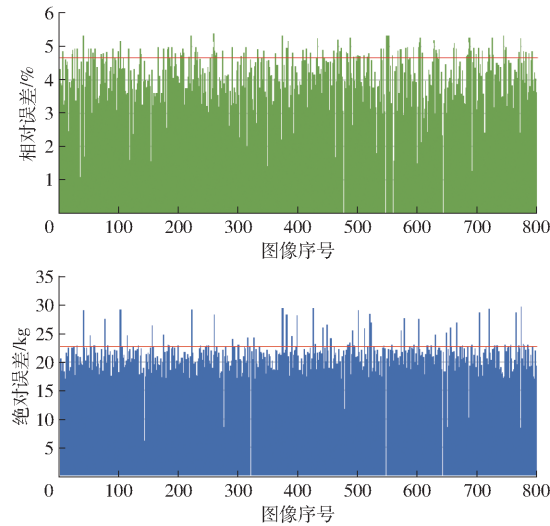


图 8 相对误差和绝对误差

Fig.8 Relative error and absolute error

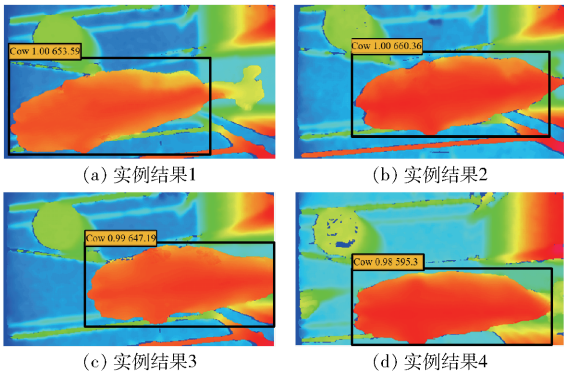


图 9 实例结果

Fig.9 Results of example

4 结论

(1)提出了一种基于奶牛背部深度图像和改进 DETR 算法的端到端式奶牛体质量评估方法(Cow-DETR),利用奶牛背部曲面变化的信息,在实现了非接触式端到端的高精度奶牛体质量评估的同时减少对奶牛位姿的要求,以及避免分阶段



的体质量评估方式对奶牛体质量评估精度的影响。

(2)设计了一种深度图像补全处理方法,通过边缘平滑滤波器和孔洞填充滤波器,对奶牛背部深度图像进行补全处理,有效降低深度图像缺失对体质量评估的影响。

(3)引入了 DETR 网络,借助其注意力机制,更加关注奶牛背部,减少背景杂物信息对体质量评估的影响;同时在 DETR 网络预测模块中添加了含有交替全连接层的体质量预测单元,能够对奶牛体质量

量相关特征信息进行更好的提取,实现了在检测奶牛个体的同时对奶牛体质量进行预测。

(4)与现有其他体质量评估模型对比,本文体质量评估方法在奶牛数量更多的预测集中取得了更好的效果,其中 MAE 为 17.21 kg,MRE 为 3.71%,单幅图像平均预测时间为 0.026 s。结果表明,本文提出的 Cow-DETR 体质量评估方法可以利用奶牛背部深度图像对奶牛的体质量进行较高精度的非接触式评估,符合规模化牧场对奶牛体质量评估的实际需求。

# 参 考 文 献

- [1] 马聪,张建华,朱丹. 牛场称重系统的发展现状、存在问题及对策分析[J]. 宁夏农林科技, 2018, 59(7): 33-35, 63. MA Cong, ZHANG Jianhua, ZHU Dan. Present situation, problems and countermeasures of weighting system in cattle farm[J]. Journal of Ningxia Agriculture and Forestry Science and Technology, 2018, 59(7): 33-35, 63. (in Chinese)
- [2] ZANTON G I, HEINRICHS A J. Meta-analysis to assess effect of prepubertal average daily gain of Holstein heifers on first-lactation production[J]. Journal of Dairy Science, 2005, 88(11): 3860-3867.
- [3] KEOWN J F, EVERETT R W, EMPET N B, et al. Lactation curves[J]. Journal of Dairy Science, 1986, 69(3): 769-781.
- [4] 徐启明. 不同因素对奶牛产奶量的影响研究[J]. 安徽农学通报, 2014, 20(16): 113-114. XU Qiming. Study on the effect of different factors on milk yield of dairy cows[J]. Anhui Agricultural Science Bulletin, 2014, 20(16): 113-114. (in Chinese)
- [5] BRICKELL J S, POLLOTT G E, CLEMPSON A M, et al. Polymorphisms in the bovine leptin gene associated with perinatal mortality in Holstein-Friesian heifers[J]. Journal of Dairy Science, 2010, 93(1): 340-347.
- [6] GJERGJI M, DE MORAES WEBER V, SILVA L O C, et al. Deep learning techniques for beef cattle body weight prediction [C]// International Joint Conference on Neural Networks (IJCNN). IEEE, 2020: 1-8.
- [7] DOHMEN R, CATAL C, LIU Q. Computer vision-based weight estimation of livestock: a systematic literature review[J]. New Zealand Journal of Agricultural Research, 2022, 65(2-3): 227-247.
- [8] JUN K, KIM S J, JI H W. Estimating pig weights from images without constraint on posture and illumination[J]. Computers and Electronics in Agriculture, 2018, 153: 169-176.
- [9] DOHMEN R, CATAL C, LIU Q. Image-based body mass prediction of heifers using deep neural networks[J]. Biosystems Engineering, 2021, 204: 283-293.
- [10] ASTIKA I W, HIDAYAT R. Prediction of cow's body weight by measurement of the body dimensions with image analysis[C]// IOP Conference Series: Earth and Environmental Science. IOP Publishing, 2020, 542(1): 012070.
- [11] ALONSO J, CASTAÑÓN Á R, BAHAMONDE A. Support vector regression to predict carcass weight in beef cattle in advance of the slaughter[J]. Computers and Electronics in Agriculture, 2013, 91: 116-120.
- [12] SCHOFIELD C P. Evaluation of image analysis as a means of estimating the weight of pigs[J]. Journal of Agricultural Engineering Research, 1990, 47: 287-296.
- [13] 武尧,刘振宇,谷亚宁,等. 基于机器视觉的母猪体重估测[J]. 电子技术与软件工程, 2020(1): 100-101. WU Xiao, LIU Zhenyu, GU Yaning, et al. Sow weight estimation based on machine vision[J]. Electronic Technology & Software Engineering, 2020(1): 100-101. (in Chinese)
- [14] MARCHANT J A, SCHOFIELD C P. Extending the snake image processing algorithm for outlining pigs in scenes[J]. Computers and Electronics in Agriculture, 1993, 8(4): 261-275.
- [15] SCHOFIELD C P, MARCHANT J A, WHITE R P, et al. Monitoring pig growth using a prototype imaging system[J]. Journal of Agricultural Engineering Research, 1999, 72(3): 205-210.
- [16] KUZUHARA Y, KAWAMURA K, YOSHITOSHI R, et al. A preliminarily study for predicting body weight and milk properties in lactating Holstein cows using a three-dimensional camera system[J]. Computers and Electronics in Agriculture, 2015, 111: 186-193.
- [17] SONG X, BOKKERS E A M, VAN DER TOL P P J, et al. Automated body weight prediction of dairy cows using 3-dimensional vision[J]. Journal of Dairy Science, 2018, 101(5): 4448-4459.
- [18] 初梦苑,刘刚,司永胜,等. 基于三维重建的奶牛体重预估方法[J]. 农业机械学报, 2020, 51(增刊 1): 378-384. CHU Mengyuan, LIU Gang, SI Yongsheng, et al. Predicting method of dairy cow weight based on three-dimensional reconstruction[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(Supp. 1): 378-384. (in Chinese)