DOI:10.19651/j.cnki.emt.2519200

基于交叉注意力和自适应损失的奶牛识别方法*

王雨蝶[1] 陈零壹[1] 韩 雷[1] 苏 新[1] 陆晓春[2]
(1.河海大学信息科学与工程学院 常州 213200; 2.河海大学人工智能与自动化学院 常州 213200)

摘 要:唯一性的身份认证对于奶牛养殖场农业保险的实施极为重要,但目前没有准确且可靠的奶牛识别方法,存在骗保事件,保险覆盖比较困难,针对此问题本文提出交叉注意力机制与自适应损失函数,并基于 YOLOv7 模型框架对养殖场复杂环境中的奶牛进行检测。通过交叉注意力机制提取图像不同方向上的关联信息,融合图像的深层和浅层特征,用于适应养殖场不良光照条件和拍摄角度带来的尺度变化。针对数据集中不同样本图像的质量不一的问题,通过自适应损失函数调节简单样本和困难样本的权重,使模型在训练过程中更加关注困难样本,增加了检测模型的鲁棒性和泛化性能。实验结果表明,提出的交叉注意力机制和自适应损失函数模型在奶牛检测和识别任务准确率达到了 94.63%,相较于 YOLOv7 原模型提高了 11.42%。

关键词:奶牛识别;交叉注意力;目标检测;自适应损失函数
中图分类号:TP391.41;TN911.73 **文献标识码**:A **国家标准学科分类代码**:520.60

Cross-attention and adaptive loss based cow identification method

Wang Yudie[1] Chen Lingyi[1] Han Lei[1] Su Xin[1] Lu Xiaochun[2]
(1. College of Information Science and Engineering, Hohai University, Changzhou 213200, China;
2. College of Artificial Intelligence and Automation, Hohai University, Changzhou 213200, China)

Abstract: Uniqueness-based identity authentication is crucial for the implementation of agricultural insurance in dairy farms. However, there is currently no accurate and reliable method for cow identification, leading to incidents of insurance fraud and difficulties in coverage. To address this issue, this paper proposes a cross-attention mechanism and an adaptive loss function, built upon the YOLOv7 model framework, to detect cows in the complex environments of dairy farms. The cross-attention mechanism extracts correlation information from different directions in the images, integrating both deep and shallow features to adapt to scale variations caused by poor lighting conditions and shooting angles in farm settings. To tackle the inconsistency in image quality across the dataset, the adaptive loss function adjusts the weights of easy and hard samples, enabling the model to focus more on challenging samples during training, thereby enhancing the robustness and generalization performance of the detection model. Experimental results indicate that the proposed cross-attention mechanism and adaptive loss function model achieved an accuracy rate of 94.63% in the task of dairy cow detection and recognition, which is an improvement of 11.42% compared to the original YOLOv7 model.

Keywords: cow identification; cross attention; object detection; adaptive loss function

0 引 言

2022 年河南省人民政府《河南省肉牛奶牛产业发展行动计划》中提出,要大力推广农业保险的覆盖,降低奶牛养殖的风险和损失,并指出农业保险是减少奶牛养殖的经济损失的主要方法。在农业保险实施过程中,受保的奶牛需要进行身份识别来进行受保认定,且身份信息需要随奶牛的生长发育及时更新[1]。然而,奶牛识别没有较为可靠的方法,存在农户私自更换标签骗保的可能,进而使保险机构经营亏损,违背了农业保险农户与银行双赢的初衷。在目前常见的奶牛识别方法中,嵌入式标签需要的识别器识别距离短,标签侵入牛体,有疫病的风险,处理不当时标签易被带上餐桌;液氮冻结编号需要的成本较高;信息耳标易丢失,易被替换。因此,基于图像的奶牛识别以其较高的准确

收稿日期:2025-06-26
*基金项目:国家自然科学基金(62371181)项目资助

率、较低的成本以及方便快捷的认定方式超越了众多传统的奶牛识别方式，成为了众多识别方式中最有潜力的一种。

人工智能已经在图像领域取得了出色的成果，主要得益于人工神经网的应用。卷积神经网络（convolutional neural networks，CNN）通过它不断变大的规模，不断深化的神经元连接和越来越复杂的卷积形式，已经演化地越来越强大。目标检测算法研究通过不断优化的网络架构、更高效的特征提取方法和越来越精细的检测策略，已经取得了显著的进展。现有的牛只识别方案多基于 CNN 和 YOLO 等检测框架。Yang 等[2]为解决奶牛因样貌和姿势差异导致的识别困难问题，使用 YOLOv8 模型实现快速的奶牛识别，并结合可变形卷积与基于坐标的注意力机制。实验结果该方法模型的平均精度（mean average precision，mAP）可达 72.9%。Qiao 等[3]使用了 Mask R-CNN 作为深度学习框架的主干网络，用于解决真实饲养场环境中牛的实例分割和轮廓提取的问题，mAP 达到了 92%。基于自注意力机制的 Transformer 模型问世后，其在图像领域的应用模型 Vision Transformer 及其变种 Swin Transformer 等使得人工神经网络在图像领域的应用效果得到了进一步的提高。Peng 等[4]为实现多种鸡肉部件的分类检测，解决传统人工分类易出错的问题，使用优化的 Swin-Transformer 模型，借助 Transformer 的自注意力结构捕捉鸡肉部件图像更全面的高层视觉语义信息，在鸡肉部件验证集 mAP 达 97.21%。作为图像分类算法和目标检测算法最先进的模型，CNN、YOLO、transformer 等深度学习模型有很大的潜力应用在奶牛识别的任务上。

然而，现有的奶牛识别并未关注样本中环境复杂和个体姿态变化的问题，同时忽略了奶牛类别及其独特花纹特征。现实中养殖场环境复杂，存在栅栏遮挡、饲料堆叠等干扰物，奶牛常呈现聚集状态，且拍摄环境时常出现光照不亮以及奶牛图像的尺度剧烈变化的情况，导致传统检测模型易产生误检或漏检。上述模型难以在复杂环境下捕捉到奶牛鉴别的关键性生物特征，且奶牛数量庞大，奶牛之间的差距变小，面部特征、花纹形状、瞳孔虹膜等鉴别特征易受姿态变化影响，实际情况中奶牛的行为姿态又难以控制。因此，深度学习大模型不能简单地应用到奶牛识别任务上，需要调整和创新以应用到实际的奶牛识别任务中。

本文将研究视角放在奶牛识别系统的前端环节，聚焦复杂场景下的奶牛目标检测问题，提出结合交叉注意力机制和自适应损失的目标检测网络。通过提取模型对图片交叉的方向上不同位置之间的关联信息，利用交叉注意力增强模型在低光照、阴影交错以及拍摄距离差异导致的尺度剧烈变化等复杂条件下完成目标检测任务。同时，又提出了自适应损失函数应对难以识别的样本和类别，实现模型对不同难度的检测任务的灵活的参数更新机制，增强模型对于困难任务的注意力。

1　相关工作

目前目标检测算法已经在奶牛识别领域有了较多的研究成果。Yang 等[2]为解决奶牛因样貌和姿势差异导致的识别困难问题，使用 YOLOv8 模型实现快速的奶牛识别，并结合可变形卷积与基于坐标的注意力机制。实验结果该方法模型 mAP 可达 72.9%。然而，文献[2]的实验场景覆盖不足，训练数据多为光照充足条件下且无遮挡的图片对极端遮挡（如奶牛完全被饲料或栅栏遮挡）的鲁棒性仍有提升空间。为进一步提升复杂环境下的检测性能，Qiao 等[3]使用了 Mask R-CNN 作为深度学习框架的主干网络，用于解决真实饲养场环境中牛的实例分割和轮廓提取的问题。相较于文献[2]得到了更好的实验结果，mAP 达到了 92%，但这个模型的计算资源消耗大，对低分辨率图像的分割效果较差。基于这个理论，Qiao 等[5]构建了一个复杂的数据集，使用 YOLOv5 模型来对图片进行特征的提取，额外使用了自适应空间特征融合算法（adaptively spatial feature fusion，ASFF）来自适应地为每张特征图进行特征融合，通过空间维度上的自适应权重分配，模型能更全面地提取特征信息，mAP 达到了 94.7%。Wang 等[6]将三重注意模块（triplet attention module，TAM）整合到骨干网络中，利用 TAM 对图片的跨纬度信息交互能力增强网络对奶牛的注意力，提出了基于 YOLOv8n（you only look once v8 nano）的改进模型 estrus-yolo（E-YOLO）。在图片中奶牛的体型较小，文中将完全交并比（complete intersection over union，CIoU）损失替换为衡量归一化 Wasserstein 距离（normalized wasserstein distance，NWD）的损失，以模型降低对目标奶牛位置偏差的敏感性，即减少模型对于奶牛在图片中位置的关注度。该模型的 mAP 达到了 95.7%。文献[6]通过 TAM 模块有效增强了模型的特征提取能力，并利用 CIoU 损失减少了模型对于无效信息的关注度，其准确度优于文献[2-3,5]，但未考虑环境因素，实际应用中仍存在漏检和误检问题。Zheng 等[7]针对检测中可能出现的漏检和误检问题，提出了一种多目标跟踪方法。该方法使用 YOLOv7 作为骨干网络来提取图片中的特征，增加了自注意力和卷积混合模块来解决奶牛空间分布不均匀和目标奶牛的尺度变化的问题。实验表明，文献[7]在检测的准确度上达到了进一步的提升，在奶牛目标检测的数据集上 mAP 达到了 97.3%。但它依赖于检测模块的精度，一旦出现漏检和误检，跟踪阶段难以恢复。

其他基于目标检测的奶牛识别研究也基本使用了 YOLO 模型来作为奶牛识别的主干网络。Guo 等[8]通过奶牛身体、身份 ID、眼窝热像共 3 个识别网络对奶牛进行跟踪识别，但这种方式需要多重身份信息，采集数据步骤繁琐。Yu 等[9]将 DenseResNet 和 YOLO 模型相结合，丰富尺度信息的交互，增强了模型的特征提取能力，相比 YOLOv4 的 mAP 提升 1.7%，但因 YOLO 系列已迭代多

个版本,仅与旧版本对比难以充分证明其先进性。Hao等[10]将高效的多尺度注意力模块集成到YOLO模型中,显著提高检测头部和腿部等较小目标的性能,使模型的检测精度达到了95.1%,但缺乏对奶牛整体检测的研究。Bello等[11]采用最先进的目标检测模型Mask YOLOv7模型,将Mask机制嵌入到YOLOv7算法的主干,对奶牛进行整体检测,检测精度达到了95%。但它未优化动态模糊即无人机移动导致的图像模糊。Ahmad等[12]利用特征注意力机制、挤压激励模块和数据增强技术改进了YOLOv8架构,提出了改进的带有特征注意机制的YOLOv8(improved YOLOv8 with feature attention mechanism,IYOLO-FAM),对于奶牛行为识别的平均准确率mAP达到了88%。各目标检测模型的比较如表1所示。

表1 目标检测模型在奶牛识别任务的比较

Table 1 Comparison of object detection models for cattle recognition tasks

文献	模型	改进模型	数据集规模	mAP/%	检测时间
[2]	YOLOv8	无	1 196张图像	72.90	16.00 ms
[3]	Mask RCNN	无	30帧视频	92.00	0.73 s
[5]	YOLOv5	YOLOv5-ASFF	1 000张图像	94.70	16.00 ms
[6]	YOLOv8n	E-YOLO	132段25帧视频	95.70	8.10 ms
[7]	YOLOv7	YOLO-BYTE	375段25帧视频	97.30	21.00 ms
[8]	YOLOv3	无	3 152张图像	96.00	—
[9]	YOLOv4	DRN-YOLO	54段25帧视频	96.91	22.65 ms
[10]	YOLOv5	YOLOv5-EMA	8 024张图像	95.10	—
[11]	YOLOv7	无	1 000张图像	95.00	—
[12]	YOLOv8	IYOLO-FAM	10 000张图像	88.00	—

以上模型虽然在各自的实验条件下表现结构优异,检测精度高,但是在养殖场的实际环境中,由于存在栅栏遮挡、饲料堆积等干扰因素,加上奶牛通常成群聚集活动,以及拍摄时光照条件不理想、拍摄距离不一导致的图像尺度变化显著,这些复杂情况使得传统检测模型容易出现误判和漏检的问题。现有的模型在如此复杂的环境中难以有效提取奶牛的关键鉴别特征,而大规模养殖又使得个体间差异变得细微,奶牛的面部特征、体表花纹、虹膜纹理等关键识别特征极易受到动物姿态变化的影响。鉴于奶牛行为姿态难以人为控制,直接将现有的深度学习大模型应用于奶牛识别任务显然存在局限性,必须进行针对性的改进和创新。

为此,本文针对复杂场景下奶牛检测难题,提出一种融合交叉注意力机制和自适应损失函数的目标检测网络。该网络通过捕捉图像不同区域间的关联特征,利用交叉注意力机制提升模型在低光照、阴影交错以及大幅尺度变化等挑战性条件下的检测性能。同时,引入的自适应损失函数能够根据样本识别难度动态调整参数更新策略,从而增强模型对困难检测任务的关注度,实现更精准的目标检测。

实验结果表明所提出的交叉注意力机制以及自适应损失函数能够很好的运用于奶牛目标检测和识别的任务中,并且在最后进行了多项消融实验的对比,进一步证明了交叉注意力机制在不同维度上的正向作用,以及自适应损失函数对本文数据集的较好适配性,能够很好地应对奶牛目标检测和识别的实际任务中养殖场光照条件不良、遮挡物较多、样本图像质量较差的问题。

2 基于目标检测的奶牛识别模型

养殖场环境中存在栅栏遮挡、饲料堆叠等干扰物,奶牛常呈现聚集状态,且拍摄环境时常出现光照不亮以及奶牛图像的尺度剧烈变化的情况,导致传统检测模型易产生误检或漏检。针对养殖场景的特殊性挑战,现有算法在复杂背景干扰下的鲁棒性亟待提升。因此,本论文提出交叉注意力机制,通过提取模型对图片交叉的方向上不同位置之间的关联信息,利用交叉注意力增强模型在低光照、阴影交错以及拍摄距离差异导致的尺度剧烈变化等复杂条件下完成目标检测任务。

从个体精准识别的角度考虑,常规检测算法难以捕捉对奶牛鉴别的关键性生物特征。不同于普遍的物体检测问题,奶牛的面部特征、花纹形状、瞳孔虹膜等鉴别特征易受姿态变化影响,而实际情况中奶牛的行为姿态又难以控制。因此,本文又提出了自适应损失函数以应对难以识别的样本和类别,实现模型对不同难度的检测任务的灵活的参数更新机制,增强模型对于困难任务的注意力。

图1为交叉注意力YOLO网络模型的整体网络结构图。模型的基本框架来自YOLOv7,相比于其他版本的YOLO模型,YOLOv7的多尺度特征提取的结构最能适应养殖场环境中尺度变化的问题。模型基于YOLOv7框架进行改进,针对养殖场环境中目标尺度多变、光照不均及

动物姿态干扰等问题，引入以下核心创新模块：Cross Transformer 模块、ELANC（efficient layer aggregation networks cross Transformer）模块和 SPPCSPC（spatial pyramid pooling cross stage partial connections）模块。并在 Cross Transformer 模块和 ELANC 模块引入交叉注意力机制，俩个模块以并行分支形式协同工作。

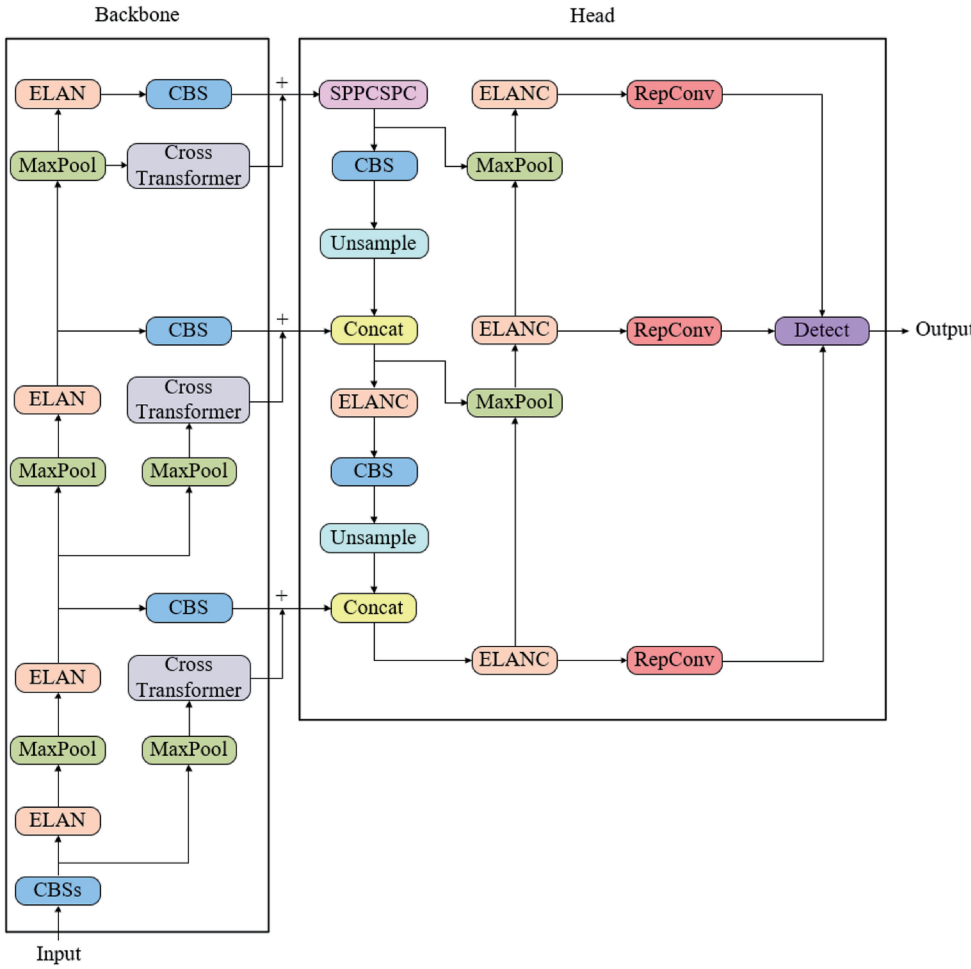

图 1　交叉注意力 YOLO 网络模型结构

Fig. 1　Cross-attention YOLO network architecture

CBS 模块为卷积层，是主干网络中用于特征提取的主要结构；ELAN（efficient layer aggregation networks）模块为高效层聚合网络[12]，通过多尺度特征融合增强主干网络性能；MaxPool 模块为最大池采样，特别的是本模型中的池化层带有一定量的参数便于给 Transformer 层做线性映射；Unsample 模块为上采样层，用于融合深层与浅层特征；Concat 模块为拼接层，将两个不同的特征沿图片的通道维度拼接起来；RepConv 模块是一个重参数化结构[13-14]，能够加快模型的推理速度；Detect 模块用于将不同尺度的特征输出为最终的检测结果。

Cross Transformer 模块为基于交叉注意力机制的 Transformer 模块。Cross Transformer 的自注意力计算在特征图的纵向维度上，负责捕捉同一列内像素的长度依赖关系，如奶牛身体的垂直结构、腿部姿态等，且垂直方向注意力可关联牛背部的光照高光区与腹部阴影区，通过整体垂直结构判定目标完整性，有效缓解光照不均导致的检测误差。在横向维度上，提取行内像素的关联信息，如奶牛面部花纹的水平连续性、背部轮廓等，且水平注意力对奶牛转头、侧身等姿态变化不敏感，例如无论牛头左转或右转，其面部花纹的水平分布模式均可被有效捕获，显著降低了奶牛运动姿态变化对检测的干扰。SPPCSP 模块整合了空间金字塔池与跨阶段部分连接[15]。ELANC 模块为 ELAN 网络的交叉注意力版本，它额外将图片的不同通道间的信息进行关联，增强跨通道信息交互，提升多尺度特征融合能力。

3　面向奶牛识别的交叉注意力

面向养殖场复杂环境的横向与纵向交叉注意力和通道交叉注意力，以及其他相关模块的作用。本文提出的交叉注意力机制应用于两个核心模块：Cross Transformer 和

ELANC。其中,Cross Transformer 以并行分支形式与 ELAN 模块协同工作,在特征提取阶段增强主干网络的特征提取性能,ELANC 则是将 ELAN 提取出的特征进行交叉注意力计算后的改进网络。

3.1 面向养殖场复杂环境的交叉注意力

横向与纵向交叉注意力是先将图像中横向与纵向方向的关联信息分别提取出来再进行融合的注意力机制。Transformer 模型关注的是一个序列,目标是提取序列中不同位置之间的关联信息。自注意力计算方程实际上是将输入向量复制为查询向量、键向量和值向量 3 个副本,然后将它们映射为一个输出。输出向量为值向量加权求和的结果,其中每个值对应的权重由查询向量与对应键向量的相关性函数计算得出,公式为:

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{1}$$

式中:$\boldsymbol{Q}$ 为查询向量,$\boldsymbol{K}$ 为键向量,$\boldsymbol{V}$ 为值向量。查询向量表示当前需要关注的目标序列,键向量则用于与查询向量匹配以获取目标序列内部的关联信息,值向量是用于生成输出的信息载体。

为增强模型的主干网络对于图片特征的提取能力,本文设计了全新的 Cross Transformer 模块用于关注图像的纵向和横向关联信息,并将它们结合起来,形成交叉的关联信息网络。Cross Transformer 模块的结构图如图 2 所示。Cross Transformer 的自注意力计算在特征图的纵向维度上,负责捕捉同一列内像素的长度依赖关系,如奶牛身体的垂直结构、腿部姿态等,且垂直方向注意力可关联牛背部的光照高光区与腹部阴影区,通过整体垂直结构判定目标完整性,从而加强检测精度,很好地解决了阴影问题;在横向维度上,提取同一行内像素的关联信息,如奶牛面部花纹的水平连续性、背部轮廓等,且水平注意力对奶牛转头、侧身等姿态变化不敏感,例如无论牛头左转或右转,其面部花纹的水平分布模式均可被有效捕获,也在一定程度上解决了奶牛易动的问题。

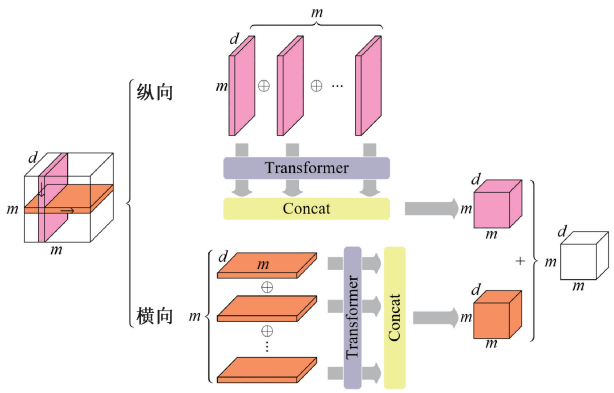

图 2 Cross Transformer 模块结构

Fig. 2 Cross Transformer module architecture

3.2 融合深浅层特征的通道交叉注意力

通道交叉注意力是将特征图的通道维度间的关联信息提取出来,并最终与特征图的信息融合起来的注意力机制。为了关注深层信息,本文使用通道自注意力机制处理 ELAN 模块的输出特征图,以提取特征图中通道之间的关联信息。

ELAN 是一种通过多路径特征融合增强卷积神经网络表征能力的模块,其核心作用在于实现多尺度特征的高效提取与融合,特别适用于处理复杂场景下的目标检测任务。在浅层路径,通过直接保留 1×1 卷积的输出,维持原始特征图的分辨率与局部细节信息。在深层路径,通过将 Path 2 的输出再输入双层卷积层,每个双层卷积包含两次连续的 3×3 卷积操作,逐步扩大感受野并提取更抽象的语义特征。最终的拼接操作则保留了从图像细节到高层语义的多层次信息。ELAN 的多路径特征融合方法,很好地使模型适应了养殖场中奶牛因距离远近呈现的显著尺度差异问题,使模型在单次前向传播中同时适配不同尺度的目标检测需求,避免传统单一路径网络因固定感受野导致的漏检问题。

ELANC 模块为 ELAN 模块级联通道自注意力(channel self-attention, CSA)模块。

CSA 模块通过分组策略和通道自注意力机制,在多尺度特征融合中实现了高效的通道间信息交互。它不仅补偿了上采样导致的信息损失,还通过跨尺度通道关联,增强了模型对复杂场景,尤其是多尺度目标的检测能力。这种结构设计一方面防止了图片输入 Transformer 模块可能带来的极大计算成本,而且显著提升了特征表达的准确性。

3.3 面向难易样本的自适应损失函数

为解决样本不平衡问题,本文提出自适应损失的解决方案,通过动态划分样本的做法和权重分配的机制缓解了这一问题。

通过统计当前训练批次所有候选正样本(IoU>0 的样本)的 IoU 均值,将其作为动态阈值 μ。正样本为预测框与真实框的 IoU> μ 的样本,负样本为预测框与真实框的 IoU≤μ 的样本。在正样本中,靠近阈值的样本被区分为困难样本,其他为简单样本。本文根据样本的类型给所有样本的损失函数添加不同的权重来让模型自适应调整对不同样本的关注度,对负样本模型给予最低的权重,让模型不关注此类样本,对简单样本模型给予中等的权重且权重值随 IoU 自适应调整,对困难样本模型给予最高的权重。权重方程为:

$$f(x) = \begin{cases} 1, & x \leqslant \mu \\ e^{1-\mu}, & \mu < x < \mu + 0.1 \\ e^{\frac{1-\mu}{0.9-\mu}(1-x)}, & x \geqslant \mu + 0.1 \end{cases} \tag{2}$$

本文将简单样本与困难样本的阈值设置为 $\mu + 0.1$,

因此，样本预测框与真实框的 IoU 值在（μ，$\mu+0.1$）范围内时，模型给予最高的权重 $e^{1-\mu}$ 给困难样本，IoU$\geqslant\mu+0.1$ 时，模型给予可变的权重给简单样本。当 IoU=1 时，由于预测框与真实框完全相同，所以将权重值设置为 1，与负样本的权重相同，不需要模型再对完全正确的样本进行学习。

4　仿真实验结果讨论与分析

4.1　实验数据集选取与训练参数设置

本文主要使用的数据集名为 Cows2021 数据集，来自于公共数据集[15]，该数据集包含 186 头牛，在英国布里斯托大学的温德赫斯特农场拍摄，摄像机从离地面 4 m 的地方向下指向一条通道进行图像的收集。该数据集分辨率为 1 280 pixel×720 pixel，包含 10 402 张奶牛牛背的图像，光线条件相对统一，遮挡较少，背景简单。数据集拥有真实标签，包括真实的物体框和类别标签。数据集图像如图 3 所示。本文还收集了来自河南多个奶牛养殖场的数据，共 2 917 张图像，图像拍摄多角度，环境更复杂，存在栅栏、饲料堆等遮挡，奶牛姿态多样，聚集情况常见，包含更多的困难样本，同时标注了奶牛身份信息。数据集图像如图 4 所示。

图 3　Cows2021 数据集

Fig. 3　Cows2021 dataset

图 4　河南多个奶牛养殖场数据集

Fig. 4　Henan multiple cow farm dataset

4.2　模型训练参数设置

1）网络参数设置

在图 1 中的 3 个 Cross Transformer 模块中，各个模块的主要参数有：嵌入向量的维度，Transformer 基本模块数量，注意力头数量。其中自注意力头用来分割嵌入向量，将嵌入向量输入 Transformer 基本模块进行自注意力计算前，需要将嵌入向量分为若干份，分别对它们进行自注意力计算，最后再将嵌入向量拼接为原来的尺寸。

本文的 Cross Transformer 模块的参数配置在整个模型中的位置由浅到深如表 2 所示。

在 ELANC 中的 Transformer 模型的参数配置则根据 ELAN 模块的输出特征图的维度来进行调整，嵌入向量的维度等于该输出特征图维度，而其他参数则与表 2 中对应嵌入向量维度相同的模块配置相同。

表 2 Cross Transformer 模块参数配置

Table 2 Cross Transformer module parameter configuration

在模型中深度	嵌入向量维度	基本模块数量	注意力头个数
浅	128	4	2
中	256	4	4
深	512	4	8

2)训练参数设置

本节实验将在公共数据集 Cows2021 和本文特有的数据集上进行实验,训练参数如下:采用 SGD 优化器进行 1 000 轮的迭代训练,权重衰减值设为 0.001。训练过程中首先在最初的 5 轮中缓慢将学习率从 0 提高到 0.05,然后使用余弦退火算法将学习率缓慢降低到 0.005。两个数据集的图片都先进行尺寸的调整,变为 640×640×3,再输入模型,训练批次大小设置为 256。

4.3 不同算法性能对比

本文使用的所有对比算法都按照原论文介绍的模型结构和训练方法进行复现,并且在相同的实验环境和训练参数下进行。由于本文的目的是检测奶牛并识别奶牛,需要作为保险业务的身份识别认定,因此准确度的优先级高于检测框位置的精确度,本文先展示准确度的实验结果,准确度的计算方法为正样本中预测类别正确的数量与所有正样本的数量的比值。表 3 展示了多种目标检测算法在 Cows2021 数据集和本文特有的数据集上的准确率。

表 3 中的数据前两名用黑体标出,第一名用下划线强调。表 3 中的 LHM 模型为 Cows2021 数据集的作者提出的针对 Cows2021 的目标检测算法,在 Cows 数据集上的准确度在所有模型中排行最高,而在本文自有数据集上的准确度稍低于其他优秀算法。本文提出的算法在 Cows2021 数据集上的准确度略低于其他优秀算法,但在本文数据集上优于其他所有算法。

表 3 多种目标检测算法在 Cows2021 数据集和本文特有数据集上的准确率

Table 3 Accuracy benchmarking of object detection algorithms on Cows2021 and proprietary datasets

模型	Cows2021	自有数据集
MaskRCNN[3]	86.25	89.27
YOLOv5-ASFF[5]	89.71	90.57
YOLO-BYTE[7]	91.91	**94.01**
DRN-YOLO[9]	91.70	93.80
YOLOv5-EMA[10]	90.11	93.12
Mask YOLOv7[11]	82.87	83.21
IYOLO-FAM[12]	85.76	90.15
LHM	**92.44**	92.87
YOLO-World[16]	83.23	85.45
RT-DERT[17]	83.67	84.98
Relation DETR[18]	88.93	90.31
DINO[19]	81.60	81.81
本文	**92.17**	**94.63**

本文提出了自适应损失算法,并特别将正样本进行了区别,其中正样本被分为了简单样本和困难样本,因此,上述计算准确度的方法也要随着正样本的分割而改变。除了常规的准确度实验之外,本文进行了另外一项区别简单样本和困难样本的准确度实验,其中简单样本的准确度为简单样本中预测类别正确的数量与所有简单样本的比值,困难样本的准确度为困难样本中预测类别正确的数量与所有困难样本的比值。实验结果如表 4 所示。

表 4 不同算法在简单与困难样本的准确度

Table 4 Comparative accuracy analysis of algorithms on easy and hard samples

模型	Cows2021		自有数据集	
	简单样本	困难样本	简单样本	困难样本
MaskRCNN[3]	89.47	78.25	91.32	84.51
YOLOv5-ASFF[5]	92.23	82.37	94.89	88.39
YOLO-BYTE[7]	**96.14**	**89.11**	**96.01**	89.83
DRN-YOLO[9]	95.33	87.68	93.39	**91.22**
YOLOv5-EMA[10]	93.87	86.14	93.91	90.19
Mask YOLOv7[11]	86.93	79.31	88.43	82.41
IYOLO-FAM[12]	89.45	82.34	90.84	87.90
LHM	**98.21**	88.77	**95.98**	87.40
YOLO-World[16]	86.77	80.16	87.47	82.90
RT-DERT[17]	87.37	79.95	85.84	80.71
Relation DETR[18]	91.03	82.56	91.12	83.29
DINO[19]	84.60	77.83	85.65	76.43
本文	94.33	**90.32**	93.48	**93.07**

表 4 中的数据前两名用黑体标出，第 1 名用下划线强调。LHM 模型在 Cows2021 数据集的简单样本中准确度远超其他算法，因此 Cows2021 在表 3 中的最高准确度的原因也可由此推测出来：Cows2021 数据集的图像大多类似，而本文数据集的难度参差不齐，多有遮挡的图像，因此 LHM 在简单样本上的高性能就适合于 Cows2021 这样难度统一的数据集。但是本文提出的算法在 Cows2021 数据集的困难样本上取得了最高准确度，这得益于自适应损失给困难样本提供了最高的关注度，特别是在本文的数据集上，本文提出的算法在困难样本上的准确度远超其他优秀算法，因此也在一定程度上提高了整体的准确度。相反，在两个数据集的简单样本上，本文提出的算法准确度略低于其他优秀算法。其他最新的优秀目标检测算法在它们各自的文献中表现出色，但在奶牛识别的任务中，效果相对来说较差一些，因此本文提出算法能够针对奶牛识别任务的特殊性，得到更加出色的效果。

4.4　消融实验

本文分别在 Cows2021 数据集和本文数据集上进行各模块的消融实验，实验结果如表 5 所示。

表 5　在 Cows2021 数据集和本文数据集上的消融实验

Table 5　Ablation experiments on Cows2021 dataset and proposed dataset

YOLOv7	Cross T	ELANC	自适应损失	Cows2021	自有数据集
√	×	×	×	82.87	83.21
√	√	×	×	84.32	89.93
√	×	√	×	88.28	87.74
√	√	√	×	91.42	89.14
√	×	×	√	83.86	88.68
√	√	√	√	92.17	94.63

由表 5 中数据可知，当 Cross Transformer 模块单独作用时，在 Cows2021 数据集上的效果较差，而在本文数据集上的效果较好，并且这种差距的提升相当明显，简单分析可以得知原因是 Cows2021 数据的图像都是自上而下进行拍摄的，大多是奶牛的背部照片，光影差距带来的影响并不明显，因此 Cross Transformer 这种专注于提取水平和垂直方向信息的模块带来的增益就较为有限；而对于本文数据集，场景中的阴影较多，垂直方向上奶牛的光影条件变化较大，因此 Cross Transformer 模块就能够自行调整对于垂直光影条件的注意力，从而带来更大的增益。当 ELANC 模块单独作用时，对于准确度的提升都相当可观，可见图片在通道维度的信息也相当重要，将深层与浅层的信息融合起来能够带来相当大的性能提升。最后是当自适应损失单独作用于模型时，由于 Cows2021 数据集的样本间相似度较高，因此对不同难度样本调整损失权重提升不大，而对本文数据集，这种方法就能够带来较大的改善。

为进一步研提出的交叉注意力机制与自适应损失之间的关系，本文对于简单样本和困难样本分别进行了消融实验。由于 Cows2021 数据集对于自适应损失方法的敏感度不高，因此此实验只在本文数据集上进行，实验结果如表 6 所示。

表 6　简单样本和困难样本的消融实验

Table 6　Ablation experiments on easy and hard samples

YOLOv7	Cross T	ELANC	简单样本	困难样本
√	×	×	88.43	82.41
√	√	×	91.34	88.48
√	×	√	90.02	87.24
√	√	√	94.68	92.07

由表 6 中数据可知，交叉注意力机制对于困难样本的提升更大，因此证明了本文提出的交叉注意力机制很适合本文数据集，也与自适应损失非常契合。

4.5　检测结果

本文的检测结果如图 5 所示，图 5 中的矩形框为模型的检测结果。可以看出，本文提出的模型在奶牛目标检测任务中表现出较好的性能。具体而言，模型能够准确地识别出图像中大部分奶牛的位置，并生成紧密包围目标的检测框。检测框的置信度普遍较高，表明模型对奶牛类别的判定具有较高的准确性。

另外，为表现本文的交叉注意力机制的作用效果，在实验过程中收集注意力网络的相关参数，绘制了如图 6 所示的热力图。由实验热力图可知，注意力机制对奶牛花纹的敏感度较大，模型对奶牛的花纹区域表现出显著的注意力集中现象，特别是在黑白斑块交界处呈现高激活值，这与生物视觉系统中“边缘检测”的机制相吻合，说明模型能够自动学习到最具判别性的局部特征。另外，在相同光照条件下，模型对阴影区域和明亮区域的注意力权重分布均匀，没有表现出明显的偏好差异，这表明模型具有光照不变性的特征学习能力，因此模型不会因为阴影而失去重要的身份信息的提取，符合交叉注意力机制的设计初衷，交叉注意力通过建立特征通道间的依赖关系，自适应地强调重要区域而抑制了无关背景信息。

图 5 奶牛检测结果

Fig. 5 Dairy cow test results

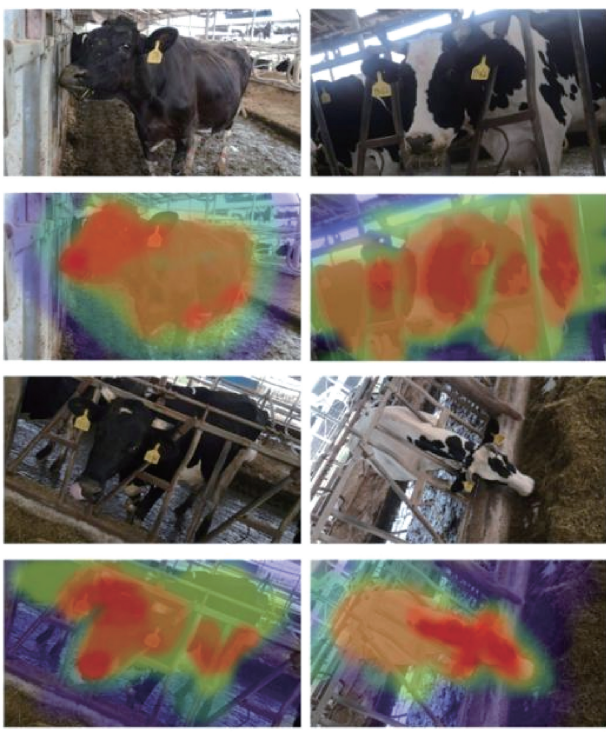

图 6 交叉注意力机制热力图

Fig. 6 Cross-attention mechanism heatmap

5 结 论

本文针对奶牛识别任务中的目标检测问题,根据奶牛以及养殖场环境的特性提出了交叉注意力机制,并将其融合到 YOLOv7 基础模型中,提出了自适应损失函数的方法来解决数据集中的样本难度不一致的问题,通过自动调整不同难度的样本的损失函数值来使模型更关注困难样本,重点克服了养殖场光照条件不良、遮挡物较多、样本图像质量较差的问题。在最后进行了多项消融实验的对比,进一步证明了交叉注意力机制在不同维度上的正向作用,以及自适应损失函数对本文数据集的较好适配性。

参考文献

[1] 张南,张旭光. 我国奶牛保险承保特点与优化[J]. 黑龙江畜牧兽医,2022(12):7-11.
ZHANG N, ZHANG X G. Characteristics and optimization of dairy cow insurance underwriting in China[J]. Heilongjiang Animal Science and Veterinary Medicine,2022(12):7-11.

[2] YANG W J, WU J CH, ZHANG J L, et al. Deformable convolution and coordinate attention for fast cattle detection[J]. Computers and Electronics in Agriculture, 2023, 211: 108006.

[3] QIAO Y L, TRUMAN M, SUKKARIEH S. Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming[J]. Computers and Electronics in Agriculture,2019,165:104958.

[4] PENG X H, XU CH CH, ZHANG P, et al. Computer vision classification detection of chicken parts based on optimized Swin-Transformer [J]. CyTA-Journal of Food, 2024, 22(1): 2347480.

[5] QIAO Y L, GUO Y Y, HE D J. Cattle body detection based on YOLOv5-ASFF for precision livestock farming[J]. Computers and Electronics in Agriculture, 2023, 204: 107579.

[6] WANG ZH, HUA ZH X, WEN Y CH, et al. E-YOLO: Recognition of estrus cow based on improved YOLOv8n model [J]. Expert Systems with Applications, 2024, 238: 122212.

[7] ZHENG ZH Y, LI J W, QIN L F. YOLO-BYTE: An efficient multi-object tracking algorithm for automatic monitoring of dairy cows[J]. Computers and Electronics in Agriculture, 2023, 209: 107857.

[8] GUO SH S, LEE K H, CHANG L Y, et al. Development of an automated body temperature detection platform for face recognition in cattle with YOLO V3-tiny deep learning and infrared thermal imaging[J]. Applied Sciences, 2022, 12(8): 4036.

[9] YU ZH W, LIU Y H, YU S F, et al. Automatic

detection method of dairy cow feeding behaviour based on YOLO improved model and edge computing[J]. Sensors, 2022, 22(9): 3271.

[10] HAO W L, REN CH, HAN M, et al. Cattle body detection based on YOLOv5-EMA for precision livestock farming[J]. Animals, 2023, 13(22): 3535.

[11] BELLO R W, OLADIPO M A. Mask YOLOv7-based drone vision system for automated cattle detection and counting[J]. Artificial Intelligence and Applications, 2024, 2(2): 115-125.

[12] AHMAD M, ZHANG W H, SMITH M, et al. IYOLO-FAM: Improved YOLOv8 with feature attention mechanism for cow behaviour detection[C]. 2024 IEEE 15th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference(UEMCON), 2024: 0210-0219.

[13] WANG CH Y, BOCHKOVSKIY A, LIAO H Y. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2023: 7464-7475.

[14] DING X H, ZHANG X Y, MA N N, et al. RepVGG: Making VGG-Style ConvNets great again[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2021: 13733-13742.

[15] ZHANG X D, ZENG H, GUO SH, et al. Efficient long-range attention network for image super-resolution[C]. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 649-667.

[16] CHENG T H, SONG L, GE Y X, et al. YOLO-world: Real-time open-vocabulary object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2024: 16901-16911.

[17] ZHAO Y, LYU W, XU SH L, el al. DETRs beat YOLOs on real-time object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.

[18] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning[J]. Advances in Neural Information Processing Systems, 2016, 29: 1-9.

[19] ZHANG H, LI F, LIU SH L, et al. DINO: DETR with improved denoising anchor boxes for end-to-end object detection[J]. ArXiv preprint arXiv: 2203.03605, 2022.

作者简介

王雨蝶，硕士研究生，主要研究方向为目标检测与识别、光通信等。

E-mail：1229974638@qq.com

陈零壹，硕士研究生，主要研究方向为特征提取、图像修复等。

E-mail：varasekie@163.com

韩雷，硕士研究生，主要研究方向为特征提取、图像修复等。

E-mail：1355425356@qq.com

苏新，教授，硕士生导师，主要研究方向为移动通信、边缘/雾计算、智慧海洋等。

E-mail：leosu8622@163.com

陆晓春(通信作者)，讲师，博士，主要研究方向为嵌入式人工智能、测量技术及仪器。

E-mail：Lu213022@163.com