doi:10.6041/j.issn.1000-1298.2023.10.028

基于超轻量化孪生网络的自然场景奶牛单目标跟踪方法

刘月峰　刘　博　暴　祥　刘好峰　王　越
（内蒙古科技大学信息工程学院，包头 014010）

摘要：针对跟踪模型泛化能力差、跟踪模型正样本选取质量低、深层模型参数量大不利于部署等问题，本文提出了超轻量化孪生网络模型 Siamese－remo。首先结合传统随机采样方法和 go－turn 方法，设计出新型的正负样本选取策略，增加模型泛化能力；其次采用 shiftbox－remo 的数据增强方式均匀正样本分布，并提升正样本采集质量；然后通过改进后的超轻量化 Mobileone－remo 网络提取特征，一定程度减少深层网络对跟踪平移不变性的破坏，并预设不同特征融合参数，单独训练网络分类和回归；最终加入 Center－rank loss 函数，根据样本点位置影响置信度、IOU 排名，对网络分类回归策略进行优化。实验证明，自然场景下奶牛单目标跟踪模型期望平均重合度（Expected average overlap，EAO）达到 0.475，相对于基线模型提升 0.078，与现有跟踪器对比取得了较好的成绩，且参数量仅为现有主流算法的 1/20，为后续自然场景下奶牛身份识别与目标跟踪系统提供了技术支持。

关键词：奶牛；单目标跟踪；特征融合；孪生网络；轻量化模型
中图分类号：TP391.4　　**文献标识码**：A　　**文章编号**：1000-1298(2023)10-0282-12　　**OSID**：

Single Target Tracking Method for Dairy Cows in Natural Scenes

LIU Yuefeng　LIU Bo　BAO Xiang　LIU Haofeng　WANG Yue
(*School of Information Engineering*, *Inner Mongolia University of Science and Technology*, *Baotou* 014010, *China*)

Abstract: The cow single target tracking technology is a new technology proposed for intelligent management of dairy farms and it is the basis for the research of cow multi-objective tracking. The presence of padding in the deep network will destroy the translation invariance of the tracking model, the number of redundant parameters, and other addressing issues such as low quality of positive sample selection for tracking models, poor generalization ability of tracking models will also affect the cow tracking performance. Thus a high-performance cow single-target tracking method was proposed. Firstly, Siamese－remo model was used to extract features by improving Mobileone network to reduce the damage of tracking translation invariance by deep network to some extent, and different feature fusion parameters were preseted to train network classification and regression respectively; secondly, traditional method and go－turn method were combined to design a positive and negative sample selection strategy to improve the quality of positive sample collection; then special data enhancement was used to increase the generalization ability of the model; finally, Center－rank loss function was added to optimize the network classification and regression strategy according to the sample point location affecting confidence and IOU ranking. The experiment proved that the expected average overlap (EAO) of the cow single target tracking model in natural scenes reached 0.475, which was improved by 0.078 relative to the baseline model, and achieved better results compared with existing trackers. The number of parameters was only one-twentieth of the existing mainstream algorithms, which provided strong technical support for the subsequent cow identification and target tracking system.

Key words: dairy cows; single target tracking; feature fusion; siamese networks; lightweight models

0　引言

随着经济的发展和社会的进步，消费者对于牛肉的品质、奶牛的奶质有着更高的要求，奶牛养殖场需要向大规模、科学绿色养殖的方向发展[1-2]。自然场景下奶牛身份识别和跟踪系统是奶牛养殖场智

收稿日期：2023－04－11　修回日期：2023－06－05
基金项目：内蒙古自治区自然科学基金项目(2022MS06008)
作者简介：刘月峰(1977—)，男，副教授，博士，主要从事机器学习与大数据研究，E-mail：liuyuefeng01035@163.com

能化管理的重要内容[3-6]。对于需要着重关注的奶牛个体，例如刚治愈的奶牛、行为不正常的奶牛等，需要进行单目标跟踪，并且可以为接下来奶牛多目标跟踪奠定基础。单目标跟踪技术是近年来热门的研究工作，主要研究方向为基于相关滤波的方法[7]和基于 Siamese FC[8-10]的孪生网络方法。基于 Siamese FC 的孪生网络方法由模板分支和搜索分支组成，模板由第 1 帧得到的 Anchor 获得，推理阶段将模板图像在搜索图像中进行局部搜索，类似于局部单次检测框架。基于孪生网络的方法分为 Anchor-base 方案和 Anchor-free 方案。Anchor-base 方案大多基于多尺度测试，预设一定数目的 Anchor 在网络中进行训练，而 Anchor-free 方案大多通过分类和回归直接对目标进行跟踪，获取其位置和预测框。LI 等[11]提出了区域特征提取网络(Siamese region proposal network，Siamese-RPN)，它由特征提取的子网络和包括分类回归分支的区域提议子网络构成，在当时公开数据集上取得了领先的跟踪性能指标。LI 等[12]随后又将 Resnet 深层网络作为孪生网络特征提取网络逐层聚合，证明了先前由于深层网络存在 padding 的原因破坏了跟踪平移不变性的要求导致跟踪失败，并加入深度交叉相关实现模板特征与搜索图之间的特征匹配，进一步提升了跟踪性能。ZHANG 等[13]提出了一种 Anchor-free 的方案预测目标的位置和大小，引入特征对比模块，从预测的边框中学习对象感知特征，进一步帮助跟踪器对目标和背景进行分类。GUO 等[14]使用逐像素卷积代替分离通道卷积，并加入 Center-ness 中心惩罚项进行跟踪，取得了较高的性能评估指标。CHEN 等[15]使用 Resnet50 作为骨干网络，去掉了最后两个卷积块的降采样操作，采用不同的扩张率提高模型的感受野，用分类模块和回归模块组成自适应头部，超过了当时所有跟踪器的跟踪效果。为了解决背景干扰大、分类和回归样本不匹配的问题，FENG 等[16]设计了基于排序的优化损失函数，包括分类和回归排名损失函数，进一步加强了跟踪的性能。上述方法采取的特征提取网络大多基于 Resnet50 网络进行改进，包含较大的参数量，选取一种轻量化模型提取特征是本文研究的重点。

传统正负样本选取策略[14-16]将视频数据前后相邻帧图像随机抽取 1 幅作为正样本，其他视频段中随机抽取 1 幅作为负样本输入模型训练，将图像数据经过翻转、平移、亮度变换等数据增强处理后输入训练。go-turn[17]方法根据目标运动轨迹设计出一种运动增广策略，正负样本靠近目标真实框中心分布密集，向四周发散分布。这两种选取策略对于帧速率高、视频流稳定的摄像头效果明显，然而若出现丢帧或目标相邻帧位移较大的情况，这两种策略效果较差，故设计合适的正负样本选取策略直接决定了本文跟踪器性能。

通用跟踪器[15-16]正负样本点划分区域方法各异，主要包括根据真实框(ground-truth)作为划分依据和根据真实框设计椭圆作为划分依据的方法。前者将真实框内部作为正样本点选取区域，外部作为负样本点选取区域，由于大部分物体真实框边界存在大量背景干扰，若将背景作为正样本传入网络则会增大模型学习难度。后者结合通用跟踪对象外形特征，设计两个椭圆作为样本点代选区域，增加无关样本点的选取，巧妙地将物体边缘较难学习位置忽略，提升了跟踪精度。

现有跟踪器方法使用的特征提取网络大多基于浅层网络 Alexnet 和深层网络 Resnet 系列网络，Alexnet 网络参数少但特征提取能力较差，Resnet 网络有较强的特征提取能力却包含大量的冗余参数。Mobileone 网络基于 MobileNet 网络[18]改进，是一种轻量型架构，它的特点是低参数量、高效率完成深度学习任务，合并冗余参数的设计压缩了网络结构，大大提升了推理速度，是一种十分适用于部署移植的网络架构[19-20]。

本文旨在研究一种适合在自然场景下部署的奶牛单目标跟踪器，“自然场景”即饲养奶牛的牛舍场景，其中包含奶牛间遮挡、牛舍栏杆遮挡、昼夜光线变化以及复杂的背景噪声等实际饲养场景。为提升数据样本采集多样性，还加入公开数据集的奶牛数据，并提高正样本质量来增强模型学习能力，最后将跟踪器轻量化压缩。

1　材料和方法

1.1　研究方案

本文首先将获取到的视频转换为图像数据后制作单目标跟踪数据集，并加入部分公开数据集中“牛”、“马”的跟踪数据，进行多数据集联合训练。首先进行正负样本的选取，结合传统方法和 go-turn 方法，将图像相邻 n 帧随机抽取 2 幅图像作为正样本，从其他视频序列随机抽取 6 幅图像作为负样本；接着进行样本预处理工作，将 2 幅正样本采用 shiftbox-remo 的数据增强方式，每幅图像随机增强 11 次，均匀正样本的分布，增加样本多样性，共组成 24 对正样本对，6 对负样本对，并进行一定概率的遮挡、亮度变换、翻转操作；然后传入改进后的 backbone 特征提取网络 Mobileone-remo，将 Mobileone 中步长(stride)为 2 的双、三支结构重参

数化为单支结构,处理速度更快、参数量更少;预设2组自适应权重,将1/8、1/16、1/32尺度下的特征层进行融合,一组用于回归分支,一组用于分类分支;再采用分离通道卷积的方式传给分类分支和回归分支;最后模型通过分类损失、回归损失、中心排序损失(Center-rank loss)联合优化网络参数,完成奶牛单目标跟踪器的设计工作,本文具体研究方案流程图如图1所示,跟踪器Siamese-remo网络模型如图2所示。

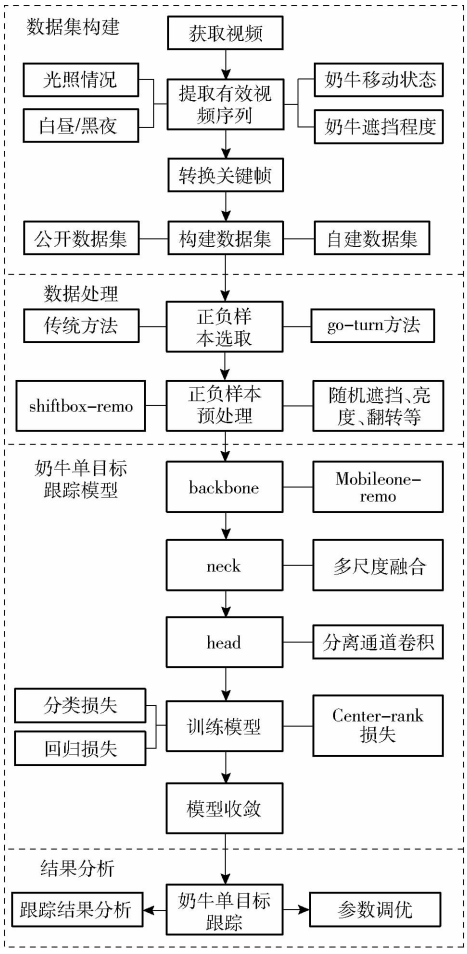


图1 研究方案流程图

Fig. 1 Flow chart of research program

1.2 数据材料获取和数据集构建

1.2.1 数据材料获取

本文使用的数据集由两部分构成,一部分为2020年内蒙古自治区包头市某奶牛养殖场采集到的52头奶牛视频。视频共2 596段,每段60 min,视频格式为MPEG4,视频帧高度为1 080像素,宽度为1 920像素,码率为1 639 kb/s,传输速率为60 f/s。另一部分为公开数据集中牛类、马类视频和图像。由于牛和马的体型相似,且为了增添训练样本的多样性,本文扩充一定规模的数据,将搜集到的公开数据集中牛类、马类的单目标跟踪视频、图像加入训练集。

1.2.2 数据集构建

本文结合自然场景下奶牛养殖场的视频图像,制作了符合单目标跟踪的数据集。由于奶牛在养殖场中行动缓慢,且处于进食状态的奶牛位置变化较小,故首先将奶牛处于进食状态的视频去除,仅保留奶牛处于移动状态的视频图像;由于奶牛在牛场移动缓慢,故将原视频每10帧抽取1帧图像;然后本文使用Labelme软件进行数据标注,将每段视频中每头奶牛的行动轨迹标注信息放在一个路径下,最终得到63段视频,1 890段奶牛跟踪序列;最后将数据文件进行裁剪和统一图像大小,整理成与GOT10K格式相同的数据形式,即以真实框中心坐标为中点,经过设计好的长宽计算方式裁剪出大小为127像素的图像作为模板图像,大小为511像素的图像作为搜索图像,若裁剪窗口超出图像范围,则用平均RGB值进行填充,如图3所示。

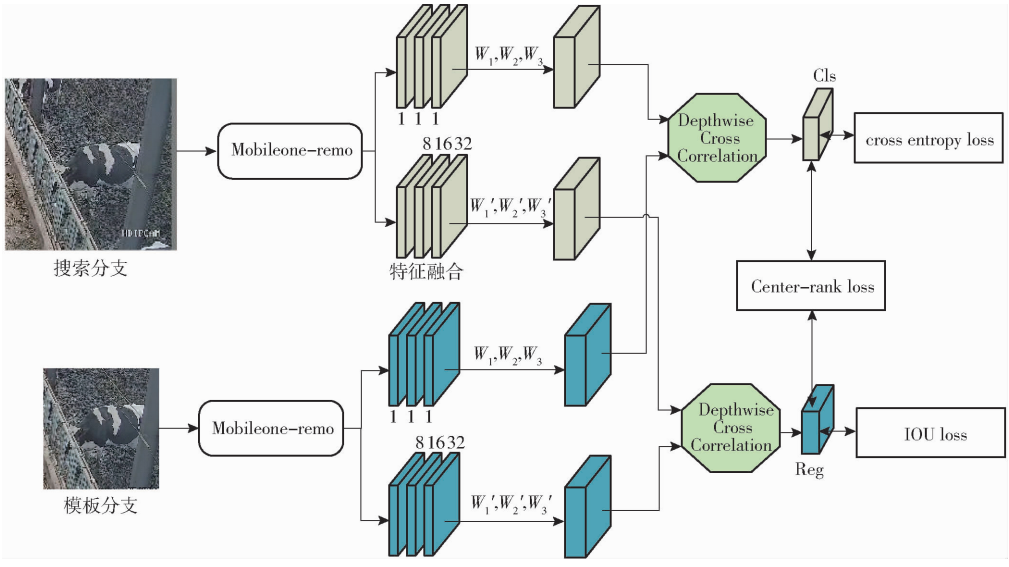


图2 Siamese-remo网络模型示意图

Fig. 2 Schematic of Siamese-remo network model

(a) 模板图像　　(b) 搜索图像

图 3　数据集预处理后的样例

Fig. 3　Sample after datasets preprocessing

由于从自然场景下获得的上述数据规模较小，难以完成单目标跟踪的要求，故本文选择将 DET[21]、COCO[22]、GOT10K[23]、VID[21]、YTB[24]、LASOT[25]公开数据集中标注为“牛”和“马”类的数据加入到训练集，模型根据不同数据集保存图像的方式分别读取到跟踪序列的真实框。

1.3　实验方法

1.3.1　正负样本选取策略

本文的正负样本选取策略通过结合 Siamban 方法[15]和 go－turn 方法[17]来增加网络泛化性能。孪生网络训练样本分为 2 个分支：模板分支和搜索分支，从数据集中随机抽取 1 幅模板图像，首先从其所在视频跟踪序列对应帧前后 frame-range 帧中随机抽取 2 幅图像，每幅图像进行 12 次 shiftbox－remo 图像增广操作后得到一组正样本队列；然后从其所在不同视频跟踪序列帧中随机取 6 幅图像，进行 shiftbox－remo 图像增广操作后作为负样本，即 1 幅模板图像对应 24 幅正样本，6 幅负样本。

模板帧图像对应的正样本搜索图像区域中，根据图像中真实框划分区域，分为正样本点、负样本点和无关样本点，分别记为 1、0、－1，如图 4 所示，中间小矩形面积包含的样本点为正样本点，大矩形外侧的样本点为负样本点，2 个矩形中间部分为无关样本点，设计无关样本点的目的是图像中真实框边缘样本包含较多复杂背景噪声干扰，且理论上边缘信息网络较难学习，故将其设置为无关样本不参与损失计算。经过实验对比论证，奶牛单目标跟踪模型设计 2 个正方形区域划分正负样本点边界效果最佳，正样本取正样本区域内所有正样本点计算损失，负样本随机取 3 倍正样本数的负样本点计算损失。

1.3.2　正负样本预处理

根据 1.3.1 节的描述，数据集包含尺寸为 127 像素×127 像素的原始模板图像和尺寸为 511 像素×511 像素的原始搜索图像，本文根据原始模板图像，在搜索图中裁剪出相应像素的搜索图。首先以搜索图中真实框为基准，假设真实框宽高分别为 w、h，裁剪出的搜索图宽 w_{crop}、高 h_{crop} 分别为

图 4　正负样本点选取策略示意图

Fig. 4　Schematic of positive and negative sample point selection strategy

$w+0.5(w+h)$、$h+0.5(w+h)$，为了增加泛化性能，对宽高进行小幅度形变处理。

自然场景下奶牛单目标跟踪受遮挡因素影响严重，为了解决这个问题，本文首先对裁剪框位置进行随机选取，模拟出奶牛部分区域未受遮挡时的真实场景，实现跟踪框“局部—整体”的跟踪能力。采用 shiftbox－remo 的裁剪方式，假设真实框左上角、右下角坐标分别为 (x_1, y_1)、(x_2, y_2)，裁剪框左侧可选择区域即 $(x_1 - w_{crop}, x_2)$，裁剪框上侧可选择区域即 $(y_1 - h_{crop}, y_2)$，若超出图像边界，则将坐标极值设为边界坐标，裁剪框位置范围如图 5 所示，正方形为原始搜索图，红色框为真实框，虚线框为裁剪框，A、B、C、D 为裁剪框移动范围极限位置，本文为了提升正样本质量，选择将裁剪框与真实框之间的交并比 $I_{ou} > 0.3$ 的图像作为搜索图像，最终将裁剪后的图像统一尺寸为 160 像素×160 像素的搜索图。

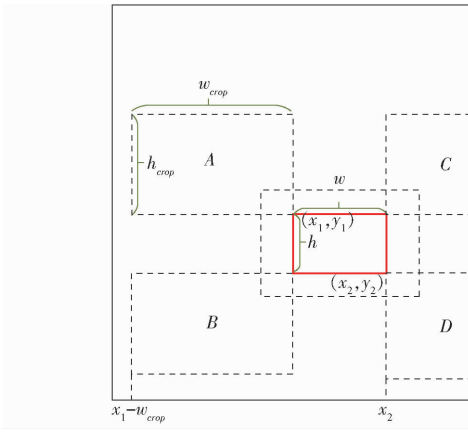
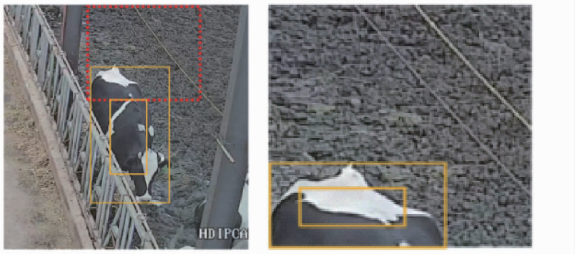


图 5　shiftbox－remo 裁剪方式示意图

Fig. 5　Schematic of shiftbox－remo cropping method

为了进一步解决遮挡问题对跟踪模型的影响，设计了适用于上述裁剪方式的正样本区域选取方式，如图 6 所示，图 6a 为裁剪框位置和原正负样本点选取区域示意图，回归分支正负样本点选取区域如黄色矩形所示，图 6b 为分类分支正负样本点选取区域随真实框更新示意图，红色区域为裁剪框位置。根据 1.2.2 节正负样本点选取策略，两个黄色矩形

中间部分样本点将作为无关样本忽略,然而经过裁剪、resize 操作后仅存在无关样本和负样本传入网络,无法学习到遮挡情况下局部正样本信息,这与本文实现“部分—整体”的跟踪目标相悖,故将分类分支中物体真实框坐标随着图像裁剪操作而更新位置,这样可以提升正样本多样性并提升其质量;回归分支仍保留裁剪操作之前的坐标,这样可以使网络具有预测“部分—整体”的能力,而并非仅可以预测局部位置。



(a) 裁剪框位置和原正负样本点选取区域　(b) 随真实框更新后的正负样本点选取区域

图 6　正负样本点划分区域示意图

Fig. 6　Schematics of dividing positive and negative sample points into regions

考虑到自然场景下昼夜变换亮度不同,且数据集光线较暗,本文对图像亮度进行数据增广,提升模型在夜间的跟踪能力,本文对模板图像和搜索图像进行了一定程度的翻转、旋转、随机擦除等数据增强方式,提升了模型泛化能力。

1.3.3　特征提取网络——Mobileone - remo

Mobileone 网络是一个轻量化的深层网络模型,相较于 Resnet 系列网络等深层网络模型,具有简单、高效、即插即用的特点。如图 7 所示,为了进一步压缩网络模型结构,并尽可能减小深层网络中 padding 对于平移不变性的影响,类比于 Siamese RPN + + 模型对 Resnet50 网络的处理[12],本文将 Mobileone 中 stride 为 2 的 3 ×3 卷积 padding 设置为 0,由于 scale 分支、3 ×3 卷积分支、skip 分支尺度不同无法相加,故将 3 个分支进行重参数化操作。对于仅有 scale 分支和 skip 分支的结构块,实验发现将其重参数化后并不会影响跟踪性能,反而可以进一步压缩模型,减少运算成本,故对 Mobileone - remo 同样进行重参数化操作。

1.3.4　多尺度预测

深层网络不同层可以提取到图像不同尺度的信息,较浅层可以获得图像高分辨率信息,例如颜色、位置,而较深层可以提取到图像丰富的语义信息,跟踪任务需要计算出跟踪位置和跟踪对象,故采取多尺度特征自适应融合方式训练网络。首先预设两组训练权重 W_i 和 $W_i'(i=0,1,2)$,分别与 1/8、1/16、1/32 尺度特征相乘,一组用于分类分支训练,一组用于回归分支训练,由深层网络不同层提取图像信息特点可知分类分支深层网络权重占比较大,回归分支浅层网络权重占比较大。

1.3.5　多功能特征头

Siamese - remo 将模板图像和搜索图像融合后的特征进行分离通道卷积,包含两个功能头进行跟踪,一个用于分类,一个用于回归。考虑到模板帧在图像预处理阶段会将一定范围背景裁剪保留,本文经过仿真统计出中心 9 ×9 区域数据仍可以捕捉到完整的跟踪模板信息,故卷积前会对模板帧特征进行中心裁剪[12,15],以真实框中点为中心裁剪大小为 9 ×9 的区域,然后输入分离通道卷积网络,如图 8 所示。在分类分支中,本文将图像信息分为前景和背景,故输出通道数为 2;在回归分支中,回归信息为训练样本点与真实框 4 条边的距离,分别记为 L、

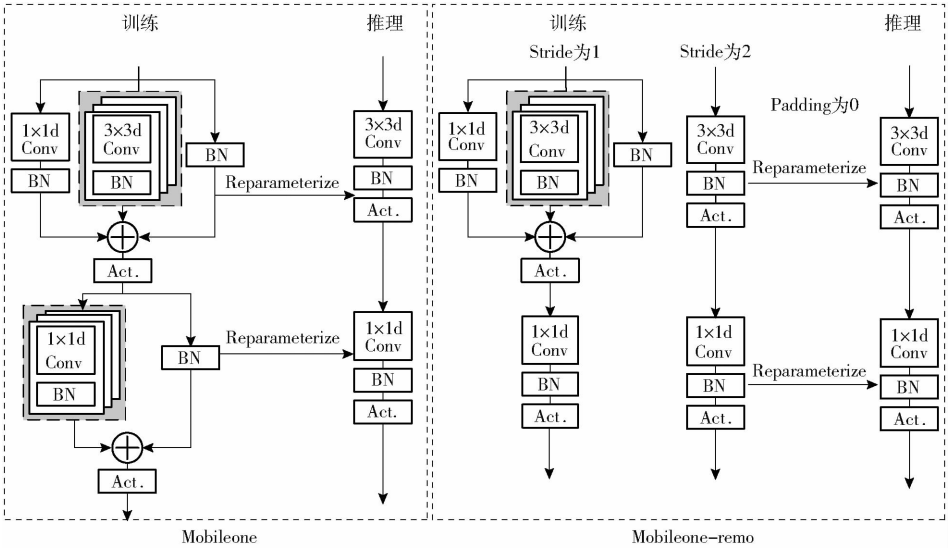


图 7　基线模型 Mobileone 与本文模型 Mobileone - remo 的结构图

Fig. 7　Baseline model Mobileone and model Mobileone - remo

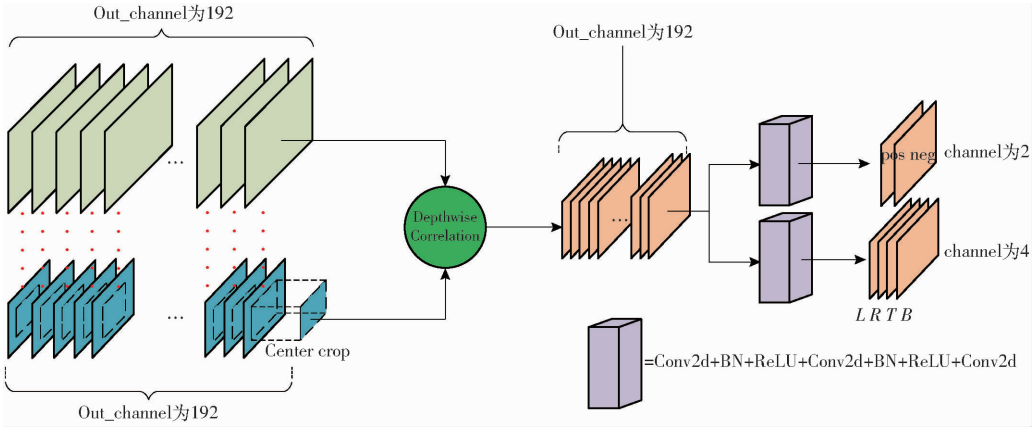


图 8　分离通道卷积和多功能头示意图

Fig. 8　Depth-with cross correlation and multifunctional head

R、T、B，故输出通道数为 4。

1.3.6　损失函数

本文在分类分支使用 cross entropy loss 计算；在回归分支使用 IOU loss 计算；两种损失函数权重组合自适应调优，联合优化训练网络。Loss 计算公式为

$$Loss = \alpha Loss_1 + \beta Loss_2 \tag{1}$$

其中

$$Loss_1 = -\sum_{i=1}^{N} y^{(i)} \lg \hat{y}^{(i)} + (1 - y^{(i)}) \lg(1 - \hat{y}^{(i)}) \tag{2}$$

$$Loss_2 = 1 - \frac{A \cap B}{A \cup B} \tag{3}$$

式中　α、β——网络自适应学习权重，初始值取 1

$Loss_1$——cross entropy loss

$Loss_2$——IOU loss

N——标签样本总数

$\lg \hat{y}^{(i)}$——第 i 个标签样本为正样本的概率

$y^{(i)}$——样本为正样本的标签

$\lg(1 - \hat{y}^{(i)})$——第 i 个标签样本为负样本的概率

$1 - y^{(i)}$——样本为负样本的标签

A——预测框　　B——真实框

本文创新性地设计了一种基于真实框中心点位置距离的排序损失——Center－rank loss。考虑到本文研究对象为奶牛，目标一定会占据真实框中心点附近大面积区域，故根据坐标位置对目标的分类、回归得分进行排序，靠近目标中心的样本点置信度高于较远位置的样本点置信度；同理，越靠近目标中心的样本点 IOU 高于较远位置的样本点 IOU。由于正样本点数量过多导致排序训练时间过长，且这种强制排名可能带来某些样本点排序的不合理性，本文选择距中心位置一定区域范围内，随机选 n 个样本点进行排序，可以在一定程度上提高模型预测能力。假设正样本 i，$j \in A_{pos}$，Center－rank loss 的计算公式为

$$Loss_{\text{center-rank}} = \frac{1}{N_{pos}} \sum_{i,\, j \in A_{pos},\, d_i > d_j} \exp(-\gamma(p_i - p_j)) + \frac{1}{N_{pos}} \sum_{i,\, j \in A_{pos},\, d_i > d_j} \exp(-\gamma(v_i^{iou} - v_j^{iou})) \tag{4}$$

式中　d_i——正样本 i 与真实框中心点的距离

d_j——正样本 j 与真实框中心点的距离

v_i^{iou}——正样本 i 的 IOU 预测值

v_j^{iou}——正样本 j 的 IOU 预测值

p_i——正样本 i 的前景置信度

p_j——正样本 j 的前景置信度

γ——超参数控制损失值

N_{pos}——选取的正样本数目，当 $d_i > d_j$ 时，p_i 排名在 p_j 之前，v_i^{iou} 的排名在 v_j^{iou} 之前

总损失函数为

$$Loss_{\text{all}} = Loss + Loss_{\text{center-rank}} \tag{5}$$

最终本文 Center－rank loss 正样本点选取范围为原正样本选取区域的 1/4，选取点数为 15。

2　实验与结果分析

2.1　实验系统环境和参数设置

本实验操作系统为 Ubuntu 18.04，CPU 为 AMD EPYC 7543 32－Core Processor，主频 3 400 MHz，GPU 为 NVIDIA GeForce GTX 3090 × 4，运行内存为 24 GB。奶牛身份跟踪模型训练共 20 个训练周期，对于维度为（N，C，H，W）的特征向量采用 dropout 方法防止过拟合，概率参数设为 0.3 对通道维度 C 进行冻结操作，并对 $H \times W$ 维度也按照类似 dropout 方式进行参数为 0.05 概率的冻结，以模拟出某些非全局特性，使模型学习到一定程度的局部特征。初始学习率为 0.001，经 5 个训练周期的学习率预热达到 0.005，backbone 权重衰减系数为 0.001，全局权

重衰减系数为 0.000 5,batch_size 设置为 4,num_workers 设置为 16,预训练模型使用 ImageNet 训练网络模型。

2.2 评估指标

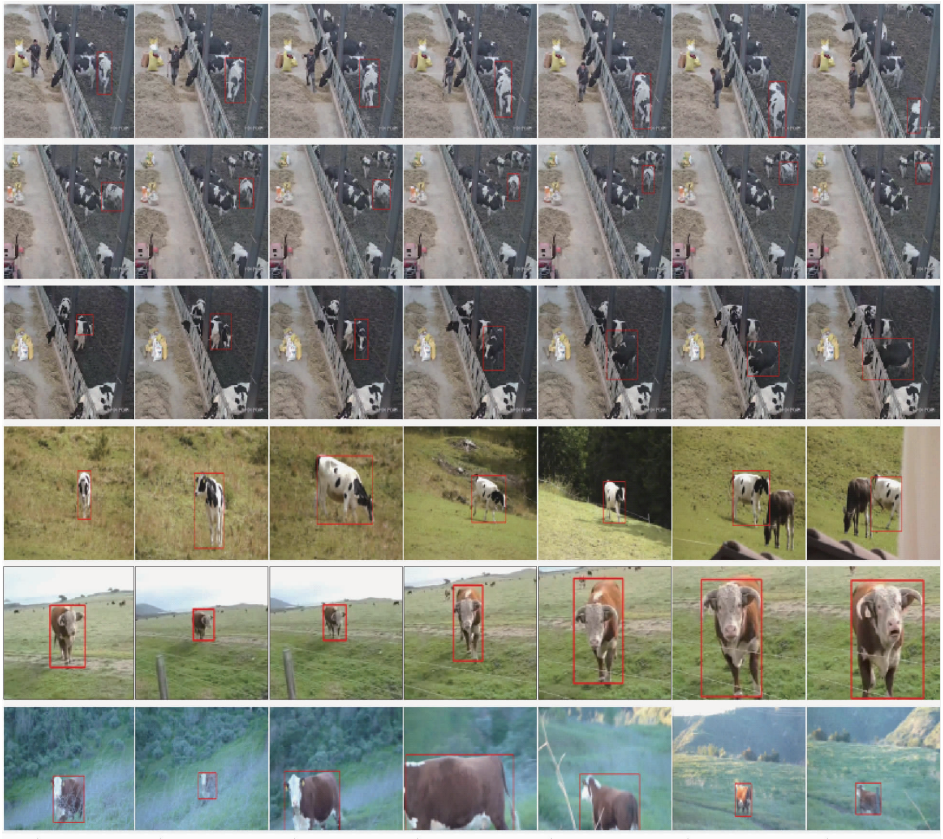
现有的用于单目标跟踪评价的指标有准确率(Accuracy)、鲁棒性(Robustness)、期望平均重合度(EAO)、查准率(Precision)、成功率(Success plot)等。鲁棒性是体现跟踪器稳定性的指标,数值越大稳定性越差,定义为每个视频序列上跟踪失败的视频帧占总帧数的比例,平均鲁棒性即所有视频序列平均跟踪失败比例。

EAO 结合跟踪器平均重合度和鲁棒性,是一个更全面的单目标跟踪性能评价指标,EAO 数值越大跟踪性能越好。查准率为预测框中心点位置与真实框中心点欧氏距离小于一定阈值的视频帧百分比,以像素为单位,根据不同的阈值得到不同的百分比,该评估指标可以反映目标位置的准确性,但是无法反映目标大小与尺度变化。成功率含义为重合率得分,即 IOU 超过设定阈值即为跟踪成功的帧。

2.3 单目标跟踪实验

本文提出的单目标跟踪模型在奶牛测试集中准确率达到 59.4%,鲁棒性达到 0.172,EAO 达到 0.475,查准率为 63.1%,成功率为 52.1%,模型参数量达到 2.7×10^6,在大幅缩小模型规模的前提下保持了较高的精度。本文还对模型在其他场景下的跟踪结果进行实验,验证模型的泛化能力,由于公开数据集中奶牛单目标跟踪数据遮挡情况较少且光线较亮,跟踪效果更优,结果准确率达到 62.1%,鲁棒性达到 0.162,EAO 达到 0.512,查准率为 67.4%,成功率为 54.4%。跟踪结果如图 9 所示,其中包含本文数据集场景和其他场景的可视化跟踪效果。为了更好地对比本文模型的优势,本文对比现在较为流行的一些单目标跟踪器训练本数据集的结果,采用相同的参数调优策略,尽可能达到该研究方法的最优结果,EAO 值如表 1 所示,各跟踪器在本文测试集的查准率和成功率指标如图 10 所示。

从图 9 可以看出,本文模型对于解决目标受复杂背景因素影响、遮挡因素影响等问题具有较好的处理能力。从表 1 可以看出,仅有 SiamFC 和 SiamRPN 模型采用浅层网络 Alexnet 作为特征提取网络,而其余模型所采用特征提取网络皆为 Resnet50 框架。本文选用改进的 Mobileone 超轻量化模型提取特征,在参数量较大缩减的情况下,通过上文的改进策略,Siamese - remo 超出了大部分



(a) Frame1 (b) Frame40 (c) Frame80 (d) Frame120 (e) Frame160 (f) Frame200 (g) Frame240

图 9 奶牛跟踪模型结果可视化效果图

Fig. 9 Visualizations of dairy cow tracking model results

表 1　不同跟踪器跟踪性能 EAO 结果比较

Tab. 1　Comparison of tracking performance EAO results of different trackers

参数	跟踪器								
	SiamFC	SiamRPN	SiamRPN + +	SiamMASK	OCean	SiamBAN	SiamACM	SiamRBO	Siamese－remo
准确率/%	49.2	53.4	58.4	56.4	59.5	61.5	60.9	59.6	59.4
鲁棒性	0.251	0.223	0.182	0.201	0.170	0.164	0.157	0.180	0.172
EAO	0.368	0.419	0.459	0.468	0.477	0.487	0.494	0.474	0.475
参数量		5.240×10^6			5.472×10^7	5.491×10^7	5.720×10^7		2.700×10^6

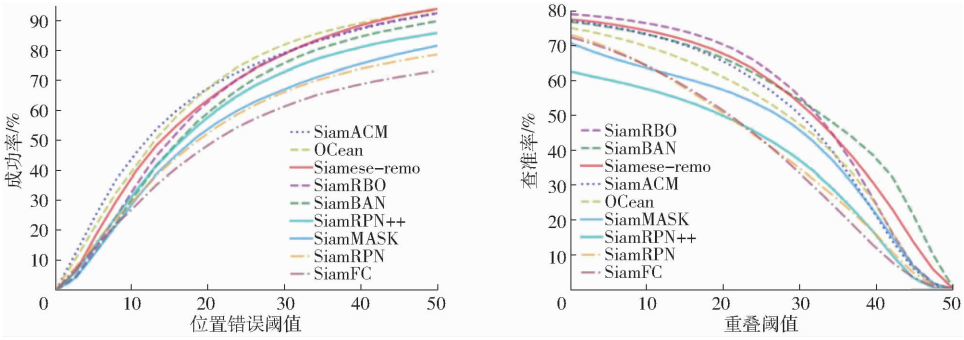


图 10　不同跟踪器的成功率和查准率结果对比

Fig. 10　Comparison of success plot and precision plot results of different trackers

Resnet50 模型的 EAO，较性能最高的模型相比 EAO 仅落后 2.1%，参数量却大大缩减。在对成功率和查准率的结果比较中（图 10），本文模型较最优模型低 1.1 个百分点和 5.2 个百分点，进一步证明了本文模型的有效性。

2.4　消融实验

2.4.1　正负样本选取及预处理策略实验

基线模型采用随机抽帧的方式从相邻帧抽取正样本，对图像不进行预处理；go－turn 方法采用运动裁剪方式，模拟物体运动轨迹，对搜索图像随机裁剪，重复 11 次，构成 12 对正样本对；本文结合基线模型和 go－turn 模型样本选取方法，抽取 2 幅正样本，随机进行 11 次 shiftbox－remo 裁剪方式，构成 24 对正样本对。本文还对模板图像裁剪大小、形状进行对比，假设真实框宽高分别为 w、h，裁剪中心为真实框中点，超出位置根据 RGB 均匀填充。裁剪方式共分为 4 种：①将 h、w 放大两倍，然后统一到 160 像素 ×160 像素。②将 h、w 分别放大至 $h+0.5(h+w)$、$w+0.5(h+w)$，并比较 $h+0.5(h+w)$ 与 $w+0.5(h+w)$ 像素，选择数值大的值裁剪正方形区域，并统一至 160 像素 ×160 像素。③在原图直接裁剪 160 像素 ×160 像素大小区域。④本文方法将 h、w 分别放大至 $h+0.5(h+w)$、$w+0.5(h+w)$，然后统一到 160 像素 ×160 像素。实验结果如表 2 所示。

由表 2 可得，对于宽高比例较大的奶牛目标，裁剪方式①会进行较大的形变处理，影响实验结果；裁剪方式②、③对目标没有形变处理，导致泛化性能较差，而本文方法对奶牛目标进行基于宽、高的形变，形变尺度比例适中，取得了最优效果。本文模拟了 10 000 幅图像通过 3 种裁剪方式的真实框中心点位置，如图 11 所示，可以看出基线模型没有对裁剪位置平移，故中心点位置全部落到中央；使用运动增广方式对裁剪框位置进行处理，模拟物体运动方向，但数据预处理后样本分布范围较小，泛化能力较差；使用本文的裁剪方式物体可以均匀分布在图像各区域位置，由于需要保证裁剪框与真实框之间 IOU 大于 0.3，故中心点落在图像角落区域的概率逐渐降低，实验结果证明使用本文裁剪方法跟踪效果最佳。

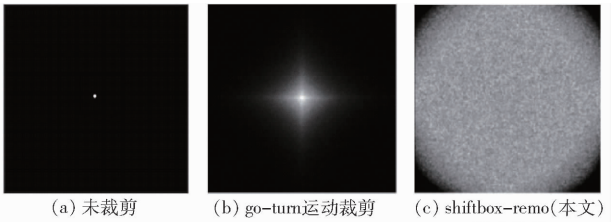
表 2　不同正负样本选取及预处理策略的 EAO 比较

Tab. 2　EAO comparison of experimental results for different positive and negative sample selection and preprocessing strategies

裁剪方式	基线模型	go－turn 模型	Siamese－remo 模型
①	0.443	0.413	0.459
②	0.437	0.394	0.431
③	0.402	0.393	0.424
④	0.451	0.435	0.475

2.4.2　正负样本点划分方式及选取策略实验

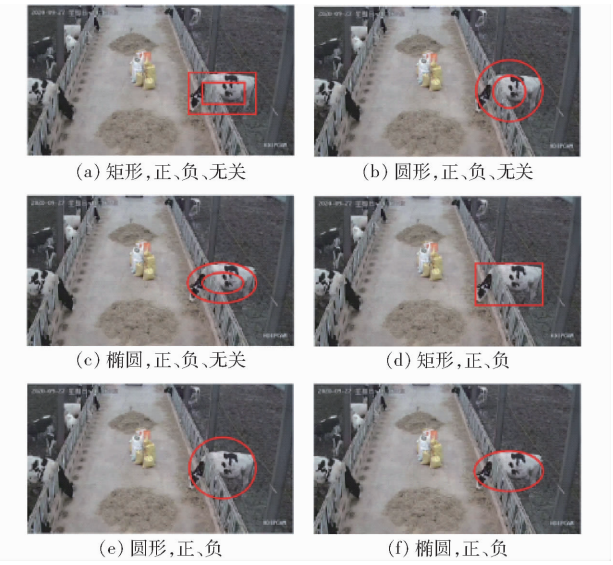
实验对比采用图 12a 的正负样本划分方式，对比不同负样本点数目对跟踪结果的影响，包括随机取 24 个正样本点，随机取 72 个负样本点；取全部正样本点，取全部负样本点；取全部正样本点，随机取正样本点 3 倍数目负样本点；随机取 24 个正样本点，取全部负样本点。还对比 3 种正负样本点划分方式的有效性，分别为椭圆、圆、矩形，并且对于是否



(a) 未裁剪 (b) go-turn运动裁剪 (c) shiftbox-remo(本文)运动裁剪

图 11 经图像裁剪后真实框中心点位置仿真结果示意图

Fig. 11 Schematics of simulation results of ground truth center point position after image cropping



(a) 矩形,正、负、无关 (b) 圆形,正、负、无关 (c) 椭圆,正、负、无关 (d) 矩形,正、负 (e) 圆形,正、负 (f) 椭圆,正、负

图 12 正负样本点采样划分方式示意图

Fig. 12 Positive and negative sample point division method

添加无关样本进行研究,对比实验如图 13 所示。图 13 中,p 为正样本,n 为负样本,i 为无关样本,r 为矩形,c 为圆形,e 为椭圆形,all p 为全部正样本,all n 为全部负样本,3num(p)n 为 3 倍正样本数目的负样本数。

实验证明,由于本文跟踪器为特定类别实例跟踪,对于奶牛个体,外观更接近于矩形,椭圆和

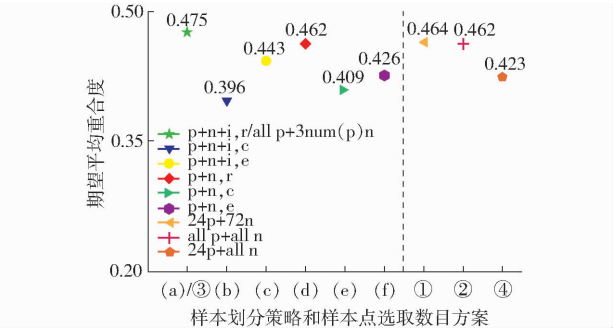


图 13 正负样本点选取区域划分方式及样本点选取数目实验结果

Fig. 13 Experimental results on division of positive and negative sample point selection regions and number of sample points selected

圆形会将奶牛边缘位置部分特征点定义为负样本点进行学习,影响跟踪器的准确性。在是否加入无关样本的实验中,加入无关样本的跟踪器 EAO 比不加无关样本的跟踪器 EAO 高 1.3%,因为无关样本的存在将物体边缘难以学习到的样本忽略,这样可以提升正样本数据的质量,并且可以减少由于边缘背景噪声带来的影响。根据图 13 可知,正样本点全部选取,随机选取正样本数量 3 倍的负样本效果最佳,并且结合测试结果可知正样本数量越多对于跟踪器的学习效果越好。由于自然场景下背景复杂,负样本如果全部选取会在一定程度对边缘位置正样本有抑制作用,小概率情况下将导致跟踪过程预测框略小。

2.4.3 特征提取网络及多尺度预测实验

本文对不同模型特征提取网络 backbone 进行比较,包括 Alexnet、Resnet18、Resnet34、Resnet50、MobilenetV1、MobilenetV2、Mobileone、Mobileone - remo,利用本文制作的数据集分别训练上述模型,实验结果如表 3 所示。

表 3 不同模型跟踪性能

Tab. 3 Results of tracking performance indicators for different models

参数	模型							
	Alexnet	Resnet18	Resnet34	Resnet50	MobilenetV1	MobilenetV2	Mobileone	本文模型
准确率/%	54.9	58.2	59.8	59.7	57.2	56.7	59.1	59.4
鲁棒性	0.251	0.221	0.224	0.234	0.203	0.195	0.167	0.172
EAO	0.402	0.479	0.492	0.489	0.453	0.447	0.473	0.475
参数量	3.750×10^6	1.255×10^6	2.266×10^7	4.553×10^7	4.220×10^6	3.470×10^6	5.530×10^6	2.080×10^6

从表 3 可以发现,由于浅层网络无法获得深层网络的语义信息,相较深层网络回归准确率较差;而深层网络中 Resnet 系列网络精度明显高于轻量化网络模型,但网络模型包括大量参数,参数量为轻量化网络模型的 10 ~ 30 倍;相较于其他深层轻量化网络模型,Mobileone - remo 具有跟踪准确率更高,参数量更少的优点,在 Mobileone 的基础上缩小一半的参数量,由于对步长为 2 的 Padding 置零,可以尽可能减小对跟踪模型平移不变性的破坏,故跟踪性能有所提升。

为了探究多尺度特征对跟踪模型的影响，以及采用两套初始化权重分别对分类回归进行训练的作用，设计相关消融实验，实验结果如表 4 所示。

表 4　消融实验结果

Tab. 4　Ablation experiment resultts

L3	L4	L5	两套初始化权重	准确率/%	鲁棒性	EAO
√				54.8	1.201	0.430
	√			55.6	1.197	0.440
		√		54.1	1.204	0.426
√	√			56.1	1.174	0.458
√		√		55.6	0.173	0.452
	√	√		57.6	0.171	0.469
√	√	√		58.4	0.182	0.462
√			√	58.0	1.186	0.450
	√		√	57.7	1.188	0.453
		√	√	57.4	1.192	0.447
√	√		√	58.7	1.174	0.458
√		√	√	58.4	0.178	0.452
	√	√	√	59.2	0.171	0.469
√	√	√	√	59.4	0.172	0.475

注：L3、L4、L5 分别表示模型输出 1/8、1/16、1/32 尺度特征；“√”表示使用该处理方法。

实验结果表明，经过对多尺度特征进行融合，效果明显优于仅使用单一尺度特征跟踪，浅层网络提取到高分辨率特征和深层网络提取到的语义信息共同对跟踪网络起作用，故采用 3 种尺度特征自适应融合效果最佳。现有孪生网络跟踪器对于分类分支和回归分支采用相同的权重参数进行训练，并不能很好地利用多尺度特征完成不同任务的优势，本文采用不同初始化参数单独训练分类和回归，网络自适应训练后打印权重信息发现，在回归任务上深层网络权重占比较高，在分类任务上浅层网络权重占比较高，实验结果证明该方法对跟踪性能有一定程度的提升。

2.4.4　不同损失函数实验

本文比较了不同损失函数训练对跟踪性能的影响，仅对算法损失函数部分进行改动，实验数据、模型以及训练方法不变。SiamBAN 模型[15]使用二元交叉熵损失函数用于分类，使用 IOU 损失用于回归，按两者所占比重 1∶1进行权重计算，记为 type 1，实验结果如图 14 所示；本文采用了基于样本点与真实框中心点距离进行分类回归排序，记为 type 2；Siamese－CAR 模型[14]在分类分支中加入 center－ness 并用于分类损失，记为 type 3；回归模型计算真实框坐标(x_1, y_1, x_2, y_2)替代 L、R、T、B，使用 L1 损失，记为 type 4；根据 Siamese－Mask 模型[26]加入二值分割分支损失函数，记为 type 5；根据 Siamese－RBO 模型[16]在分类分支加入了基于 IOU 和置信度的动态排名损失，记为 type 6。

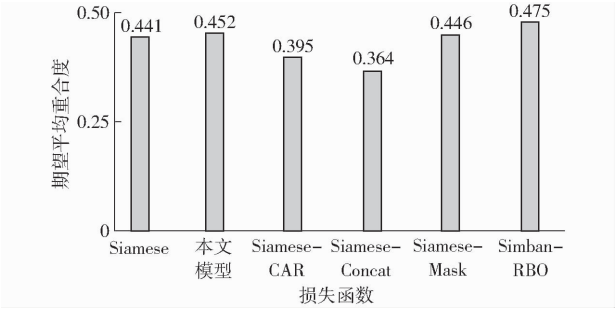
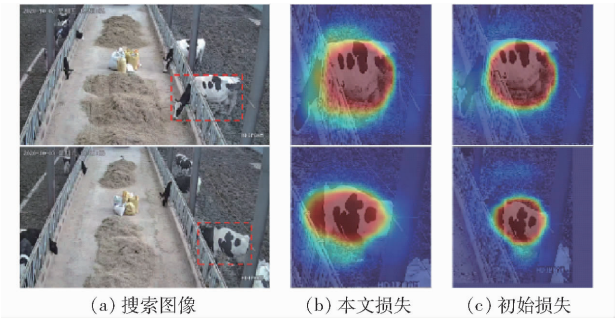


图 14　6 种不同损失函数跟踪结果

Fig. 14　Tracking results of six different loss function

通过图 14 可以看出，本文提出的基于中心位置的排序损失评估结果仅次于基于 IOU 和置信度动态排序的损失评估指标。结合图 15 可视化分类效果，分析可得加入 Rank－remo loss 后，由于分类得分排序受距离影响，分类响应在目标中心至边缘区间有一定梯度的缓慢下降；而原始损失函数由于没有距离的影响，分类响应仅在目标边缘突然下降，与距离无关。本文选取分类得分最高值点进行回归训练，选取到的最高值点距离目标中心越近，则越有利于目标回归学习，而原始方式选取到分类得分最高值点位置区域更大，当选取到样本点接近目标边缘位置时会影响回归效果，故说明在解决跟踪问题的过程中，基于样本点与真实框中心点距离对其分类和回归结果进行重新排序是有效果的。基于 IOU 和置信度动态排序的方法，需要根据样本点置信度排名调整 IOU 排名，根据样本点 IOU 排名调整置信度排名，这种动态算法无疑需要更大的计算量，严重影响了训练时间，这与本文设计网络算法的初衷相违背，故设计统一排序标准——按与真实框中心点的距离进行排序，有效地降低了算法复杂度，减少了一半的训练时间，且跟踪性能也取得了较好的结果。



(a) 搜索图像　(b) 本文损失　(c) 初始损失

图 15　Rank－remo loss 与原始损失函数分类计算结果可视化示意图

Fig. 15　Visualization diagram of Rank－remo loss and original loss function classification calculation results

2.4.5 数据增强实验

本文数据为自然场景下的奶牛图像,存在大量遮挡、光线变换等场景,为了扩充样本多样性,本文进行了尺度变换(SCALE)、灰度变换(GRAY)、模糊处理(BLUR)、翻转(FLIP)、随机擦除(ERASE)等数据增强工作,有效提升了模型性能,实验结果如表 5 所示,经实验对比各数据增强操作得出最优超参数,并设置数据增强概率 SCALE 为 0.5, GRAY 为 0.4, BLUR 为 0.2, FLIP 为 0.5, ERASE 为 0.5。

表 5 数据增强结果

Tab.5 Results of data enhancement

SCALE	GRAY	BLUR	FLIP	ERASE	EAO
					0.453
√					0.455
√	√				4.465
√	√	√			0.465
√	√	√	√		0.468
√	√	√	√	√	0.475

注:"√"表示使用该处理方法。

从表 5 可以看出,经过采用 5 种常见的数据增强工作,实验 EAO 提升 0.022,加入灰度和随机擦除方式的效果最为明显,这是由于本文实验数据集背景为牛舍,夜间光线较暗,加入灰度增广来丰富数据多样性,对于解决夜间跟踪"误跟"、"漏跟"问题效果明显。而牛舍中存在大量遮挡场景,加入随机擦除的方式也有利于模型的性能提升。

2.4.6 其它实验

本文对推理阶段模板帧是否更新进行比较,在模板帧更新模型中,将前一帧作为后一帧跟踪模型的模板图像进行处理[27]。实验发现,模板帧更新会导致跟踪失败,实验结果较差,模型跟踪失败后也缺乏纠错能力,当预测框位置准确率较低时,会影响模板帧更新后质量,导致跟踪失败。还比较了不同预训练模型对实验的影响,分别包括使用 ImageNet[28] 预训练模型、通用多类别跟踪数据集预训练模型、奶牛目标检测数据预训练模型等,比较发现使用 ImageNet 预训练模型效果较好。

3 结束语

提出了一种自然场景下奶牛单目标跟踪模型,基于传统孪生网络算法,设计了一种新型的正负样本选取策略,提升了模型样本的多样性,并进行 shiftbox - remo 数据增强处理,提升正样本采集质量。然后使用改进后的 Mobileone - remo 网络提取特征,融合 1/8、1/16、1/32 尺度特征,并分别输入分类分支和回归分支,采用超轻量化模型提取到高质量特征。最后加入了中心点排序损失函数进行训练,根据样本点与真实框中心点距离优化模型参数。实验证明,本文提出的跟踪器在奶牛测试数据集的 EAO 评估指标达到 0.475,模型参数量缩小至 1/20,节省了计算资源,提高了计算效率,验证了本文方法的有效性,为奶牛身份识别与目标跟踪系统的研究提供了技术支持。

参 考 文 献

[1] 张永红. 科学饲养提高奶牛养殖效益[J]. 中国畜禽种业,2019, 15(2): 101 - 102.
ZHANG Yonghong. Scientific feeding to improve the efficiency of dairy farming[J]. The Chinese Livestock and Poultry Breeding, 2019, 15(2): 101 - 102. (in Chinese)

[2] 刘月峰,边浩东,何滢婕,等. 基于幅值迭代剪枝的多目标奶牛进食行为识别方法[J]. 农业机械学报,2022,53(2):274 - 281.
LIU Yuefeng, BIAN Haodong, HE Yingjie, et al. Detection method of multi-objective cows feeding behavior based on iterative magnitude pruning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(2): 274 - 281. (in Chinese)

[3] 张宏鸣,孙扬,赵春平,等. 反刍家畜典型行为监测与生理状况识别方法研究综述[J]. 农业机械学报,2023,54(3):1 - 21.
ZHANG Hongming, SUN Yang, ZHAO Chunping, et al. Review on typical behavior monitoring and physiological condition identification methods for ruminant livestock[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(3): 1 - 21. (in Chinese)

[4] DAVISON C, MICHIE C, HAMILTON A, et al. Detecting heat stress in dairy cattle using neck-mounted activity collars[J]. Agriculture, 2020, 10(6): 210.

[5] 张宏鸣,汪润,董佩杰,等. 基于 DeepSORT 算法的肉牛多目标跟踪方法[J]. 农业机械学报, 2021, 52(4): 248 - 256.
ZHANG Hongming, WANG Run, DONG Peijie, et al. Beef cattle multi-target tracking based on DeepSORT algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(4): 248 - 256. (in Chinese)

[6] 涂淑琴, 刘晓龙, 梁云,等. 基于改进 DeepSORT 的群养生猪行为识别与跟踪方法[J]. 农业机械学报, 2022, 53(8):345 - 352.
TU Shuqin, LIU Xiaolong, LIANG Yun, et al. Behavior recognition and tracking method of group housed pigs based on improved DeepSORT algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(8): 345 - 352. (in Chinese)

[7] 黄月平, 李小锋, 杨小冈,等. 基于相关滤波的视觉目标跟踪算法新进展[J]. 系统工程与电子技术, 2021, 43(8):2051 - 2065.
HUANG Yueping, LI Xiaofeng, YANG Xiaogang, et al. New progress in visual object tracking algorithms based on correlation filtering[J]. Systems Engineering and Electronics, 2021, 43(8): 2051 - 2065. (in Chinese)

[8] LUCA B, JACK V, JOAO F H, et al. Fully-convolutional siamese networks for object tracking[C]// European Conference on Computer Vision. Amsterdam: Springer, 2016: 850 - 865.
[9] LUCA B, JACK V, JOAO F H, et al. Learning feed-forward one-shot learners[C]// Advances in Neural Information Processing Systems, 2016.
[10] LUCA B, JACK V, JOAO F H, et al. Staple: complementary learners for real-time tracking[C]// Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[11] LI B, YAN J J, WU W, et al. High performance visual tracking with siamese region proposal network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8971 - 8980.
[12] LI B, WEI W, WANG Q, et al. Siamrpn + + : evolution of siamese visual tracking with very deep networks[C]// Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 4282 - 4291.
[13] ZHANG Z P, PENG H W, FU J L, et al. Ocean: object-aware anchor-free tracking[C]// European Conference on Computer Vision, 2020: 771 - 787.
[14] GUO D Y,WANG J,CUI Y, et al. Siamcar: siamese fully convolutional classification and regression for visual tracking[C]// Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 6269 - 6277.
[15] CHEN Z D, ZHONG B N, LI G R, et al. Siamese box adaptive network for visual tracking[C]// Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 6668 - 6677.
[16] FENG T, QIANG L. Learning to rank proposals for siamese visual tracking[J]. IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society, 2021, 30:8785 - 8796.
[17] HELD D, THRUN S, SAVARESE S. Learning to track at 100 fps with deep regression networks[C]// Conference on Computer Vision and Pattern Recognition (ECCV), Springer, 2016: 749 - 765.
[18] SANDLER M, HOWARD A, ZHU M L, et al. Mobilenetv2: inverted residuals and linear bottlenecks[C]// Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 4510 - 4520.
[19] MA N N, ZHANG X Y, ZHENG H T, et al. Shufflenet v2: practical guidelines for efficient CNN architecture design[C]// European Conference on Computer Vision, 2018.
[20] KUMAR P, VASU A, JAMES G, et al. An improved one millisecond mobile backbone[J]. arXiv preprint:2206.04040, 2022.
[21] OLGA R, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211 - 252.
[22] LIN T Y, MAIRE M, SERGE B, et al. Microsoft COCO: common objects in context[C]// European Conference on Computer Vision, 2014: 740 - 755.
[23] HUANG L H, ZHAO X, HUANG K Q. GOT - 10k: a large high-diversity benchmark for generic object tracking in the wild [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.
[24] REAL E, SHLENS J, MAZZOCCHI S, et al. YouTube-BoundingBoxes: a largehigh-precision human-annotated data set for object detection in video[C]// Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 5296 - 5305.
[25] FAN H, LIN L T, YANG F, et al. LaSOT: a high-quality benchmark for large-scale single object tracking[C]// Conference on Computer Vision and Pattern Recognition, 2019: 5374 - 5383.
[26] WANG Q, ZHANG L, LUCA B, et al. Fast online object tracking and segmentation: a unifying approach[C]// Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 1328 - 1338.
[27] JACK V, LUCA B, JOAO H, et al. End-to-end representation learning for correlation filter based tracking[C]// Conference on Computer Vision and Pattern Recognition, 2017: 2805 - 2813.
[28] SANH V, WOLF T, RUSH A M. Movement pruning: adaptive sparsity by fine-tuning[C]// Proceedings of the Advances in Neural Information Processing Systems, 2020: 20378 - 20389.

(上接第 245 页)

[22] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]// European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213 - 229.
[23] DUAN K, BAI S, XIE L, et al. Centernet: keypoint triplets for object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6569 - 6578.
[24] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLO v4: optimal speed and accuracy of object detection[J]. arXiv Preprint, arXiv:2004.10934, 2020.
[25] MISRA D. Mish: a self regularized non-monotonic activation function[J]. arXiv Preprint, arXiv:1908.08681, 2019.
[26] GE Z, LIU S, WANG F, et al. YOLOx: exceeding yolo series in 2021[J]. arXiv Preprint, arXiv:2107.08430, 2021.
[27] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLO v7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464 - 7475.
[28] JEONG E J, KIM J, HA S. Tensorrt-based framework and optimization methodology for deep learning inference on jetson boards[J]. ACM Transactions on Embedded Computing Systems (TECS), 2022, 21(5): 1 - 26.