doi:10.6041/j.issn.1000-1298.2026.18.030

基于改进 YOLO v8n－pose 的多视角羊脸关键点检测方法

李书琴　马红福

（西北农林科技大学信息工程学院，陕西杨凌 712100）

摘要：羊脸关键点检测是羊脸识别、面部表情及头部姿态分析技术的重要基础。针对自然养殖环境下羊脸面部朝向多变、尺度不一等检测难点，本研究提出基于 YOLO v8n－pose 模型改进的 YOLO v8n－DCL－pose 模型，该模型在保持轻量化结构的基础上显著提升了对多视角、多尺度羊脸关键点检测精度。采用改进的可变形卷积（Deformable convolution network v4）改进 c2f 模块，利用动态采样机制增强模型在羊脸形变以及羊脸尺度不一情况下的特征自适应能力；在特征融合层新增 1 个 160 像素×160 像素的小目标检测层，提升模型对远距离小尺度羊脸关键点的检测能力。采用轻量化非对称解耦检测头（Lightweight asymmetric decoupled head，LADH－Head）将原有的任务耦合检测头拆分为独立的分类、回归和 IoU（Intersection over union）分支，提高关键点检测精度。在自建数据集上，改进的 YOLO v8n－DCL－pose 模型在保持计算效率的同时，较 YOLO v8n－pose 参数量仅增加 6×10^5，AP 提升 5.2 个百分点，mAP50 提升 5.1 个百分点，mAP95 提升 9.9 个百分点，表现出更优的关键点检测性能。YOLO v8n－DCL－pose 模型能够有效提取不同视角羊脸关键点信息，可为后续羊脸识别等任务提供关键信息支撑。

关键词：羊脸；关键点检测；YOLO v8n－pose；DCNv4

中图分类号：TP391.4　　**文献标识码**：A　　**文章编号**：1000-1298(2026)18-0355-10　　**OSID**：

Multi-view Key Point Detection Method For Sheep Faces Based on Improved YOLO v8n－pose

Li Shuqin　Ma Hongfu

(*College of Information Engineering*, *Northwest A&F University*, *Yangling*, *Shaanxi* 712100, *China*)

Abstract: Key point detection of sheep faces was regarded as a fundamental step for sheep face recognition, facial expression analysis, and head posture estimation. To address the challenges posed by variable orientations and inconsistent scales of sheep faces in natural breeding environments, an improved YOLO v8n－pose network model, named YOLO v8n－DCL－pose, was proposed. This model significantly improved the detection accuracy of key points on multi-view and multi-scale sheep faces while maintaining a lightweight structure. Specifically, several enhancements were made: firstly, an improved deformable convolution module, DCA, was employed to replace the original c2f module. A dynamic sampling mechanism was utilized to enhance the model's adaptability to sheep face deformations and varying scales. Secondly, a 160 pixel×160 pixel small-object detection layer was added at the feature fusion stage to strengthen the model's ability to detect key points on small-scale, long-distance sheep faces. Finally, a lightweight asymmetric decoupled head (LADH－Head) was adopted, in which the original coupled detection head was decomposed into separate branches for classification, regression, and IoU prediction, thereby improving the precision of key point detection. Experiments on a self-built dataset showed that the improved YOLO v8n－DCL－pose model increased the parameter size by only 6×10^5, while maintaining computational efficiency. The overall AP was improved by 5.2 percentage points, mAP50 by 5.1 percentage points, and mAP95 by 9.9 percentage points compared with that of the original model, demonstrating superior performance in key point detection. These results indicated that the proposed YOLO v8n－DCL－pose model served as an effective key point extractor, which was capable of providing critical feature support for downstream tasks such as sheep face recognition.

Key words: sheep face; key point detection; YOLO v8n－pose; DCNv4

收稿日期：2025－06－10　修回日期：2025－08－15

基金项目：国家重点研发计划项目（2022YFD1300200）

作者简介：李书琴（1965—），女，教授，博士生导师，主要从事农业信息化与智能信息系统相关研究，E-mail：lsq_cie@nwsuaf.edu.cn

0 引言

羊只脸部关键点检测是头部姿态估计、羊脸识别的关键,也是羊采食行为分析、健康状况预警、疾病检测等重要支撑技术,因此羊脸部关键点检测对羊智慧养殖具有重要意义[1]。

随着深度学习发展和广泛应用,关键点检测技术取得了较大发展。目前主流的检测方法主要有:以 MTCNN[2-3] 为代表的级联式算法、以 Openpose[4-5] 为代表的两阶段式算法、以 HRNet[6-7] 为代表的并行式算法、以 YOLO-pose[8-10] 系列为代表的单阶段回归式算法等。这些方法在关键点检测和姿态估计任务中表现出色,尤其为人脸关键点检测技术的发展奠定了坚实基础。

然而,羊脸的生物特征与人脸不同,羊只面部耳朵和毛发常常遮挡面部特征,其姿态变化和遮挡问题更加复杂,特别在复杂光照和多视角环境下,导致传统人脸检测方法难以直接应用。近年来已有诸多针对家畜关键点检测的模型算法,并取得了显著成果。Chen 等[11] 利用图卷积网络结合坐标分类头,将羊只体态关键点定位转换为坐标分类问题,保留了关键点的空间信息,通过捕获结构位姿信息提高精度,模型平均精度(AP)比 HRNet 提高 1.2 个百分点。Hua 等[12] 基于 YOLOX[13] 设计了一种名为 YOLOX-pose 的端到端奶牛关键点检测模型,使用 OKS(Object keypoint similarity)损失函数优化 OKS 评估指标,在自建奶牛姿势数据集上取得了速度和精度的较好平衡。刘刚[14] 等通过在 YOLO v5-pose 中融合 CBAM 注意力、替换 C3Ghost 模块,并引入 DyHead 检测头的方法来提升猪只体尺关键点的特征提取能力,模型 AP 达 92.6%。Yang 等[15] 提出了 CowK-Net 奶牛骨骼关键点提取网络,引入 Keypoint Refine Machine (KPRM) 和混合编码器,优化了跨尺度信息融合与关键点处理,实现了较高精度的关键点检测。宋怀波等[16] 基于 ShuffleNetV2 与 SimCC 的组合提出了一种轻量化奶牛关键点检测器,分类准确率可达 84.56%。Wu 等[17] 将空洞空间金字塔池化(ASPP)模块引入 ResNet101 的浅层网络,并且在卷积层中使用不同的膨胀率扩大模型的感受野,提高模型对奶牛不同维度大小和遮挡关键点特征的提取能力。黄志杰等[18] 改进 YOLO v5-Face 用于猪脸关键点检测,采用 MobileOne 骨干网络降低参数量,其平均检测精度相比较原模型提升 2.48 个百分点。黄小平等[19] 在 YOLO v7-pose 模型中引入 SPPFCSPCL 结构以改进模型对牛脸关键点的特征提取能力,改进后的模型较原模型 AP 提升 5 个百分点。张惠莉等[20] 使用嵌有 CA 注意力机制的 FasterNet 特征提取网络改进 SimCC 模型,以提升对长距离高猪只面部关键点的特征捕获能力,模型 AP 达到 76.12%。

上述对家畜关键点检测研究,大多集中在正脸或姿态受限下的检测,针对多视角研究较少,且存在模型复杂度高、检测准确率低的问题。为此,本研究提出一种基于改进 YOLO v8n-pose 的多视角羊脸关键点检测方法,在体征提取部分引入可变形卷积 DCNv4 (Deformable convolution v4)[21] 和 CAA (Context anchor attention)[22] 注意力机制,以增强模型对不同视角的羊脸特征提取能力;新增 1 个小目标检测层以提高对小目标特征关键点的提取能力;使用轻量化的解耦检测头替换原检测头,以进一步提高关键点检测精度。

1 羊脸关键点检测模型

YOLO v8-pose 是一种端到端的关键点检测算法,具有较低的计算复杂度和轻量化设计。按照模型尺寸缩放的不同,可分为 YOLO v8n-pose、YOLO v8s-pose、YOLO v8m-pose、YOLO v8l-pose 以及 YOLO v8x-pose。为满足模型轻量化需求,选择 YOLO v8n-pose 模型进行改进。

YOLO v8n-pose 模型是针对 COCO[23] 数据集的 17 个人体关键点检测的模型,而本研究聚焦于羊脸面部的左眼、右眼、左鼻角、右鼻角、鼻尖、左嘴角以及右嘴角 7 个生物特征关键点;此外,由于羊脸面部特征与人脸差异较大,且在不同视角和尺度下表现出较强的形变,导致 YOLO v8n-pose 模型在处理羊脸特征时效果不佳。为此,对 YOLO v8n-pose 模型进行改进:在骨干网络和特征提取网络部分,用改进的可变形卷积(DCA)模块改进 c2f 模块,提升模型对不同大小和形状羊脸的特征适应能力;在特征融合层增加 1 个 160 像素×160 像素的小目标特征检测层,提升模型对细小关键点的检测精度;在检测层使用轻量化解耦检测头(Lightweight asymmetric detection head, LADH-Head)[24] 替换原结构中检测头,通过分离分类与回归的特征提取路径,消除任务间干扰,从而提升关键点的检测效率。改进的模型称为 YOLO v8n-DCL-pose,模型结构如图 1 所示。

1.1 c2f_DCA 模块

c2f 模块在 YOLO v8n-pose 模型中主要作用是通过多个卷积操作增加模型的特征提取能力,以提高对关键点的检测和定位能力。然而,普通的 Conv 卷积操作中,固定的卷积核尺寸和形状在面对羊脸

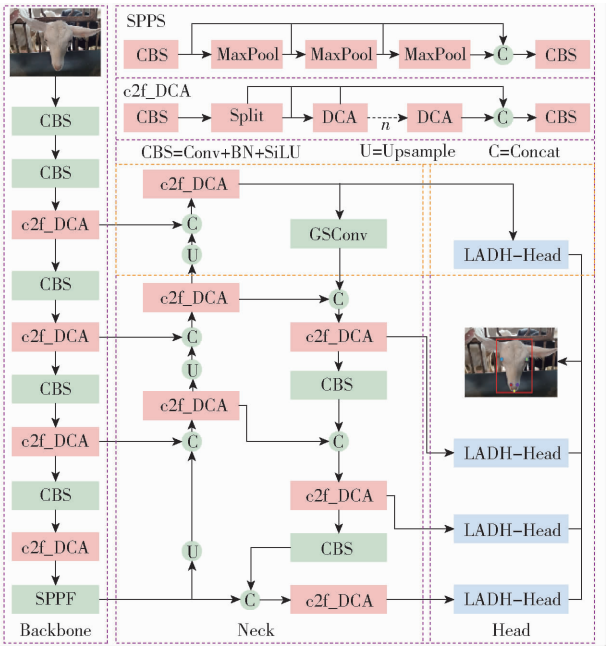

图 1　YOLO v8n－DCL－pose 模型结构

Fig. 1　YOLO v8n－DCL－pose network architecture

这种具有多变视角和不规则结构的目标时，无法充分提取出细节特征。

DCNv4 通过标准卷积提取图像局部特征，用以生成偏移量，并据此调整采样位置，从而提升模型对形变目标的定位能力；但其缺乏对图像全局信息的有效建模与融合。为此，对 DCNv4 进行改进，设计了 DCA 模块，替代 c2f 结构中的 BottleNeck 模块，得到 c2f_DCA 模块，以增强模型在不同面部视角下对卷积核形状和位置的自适应调整能力，提升对羊脸面部细节和关键点的特征提取效果。

图 2 为可变形卷积示意图，其通过引入 1 个专用的偏移量预测分支，使得卷积操作更具灵活性。该分支通过额外的卷积层对输入特征图进行分析，动态生成采样点在 X 和 Y 二维空间方向上的偏移场(offset field)。这种设计使得卷积核能够根据特征的实际分布，自适应地调整感受野的位置，从而更加精准地聚焦于羊脸特征区域。其数学计算式为

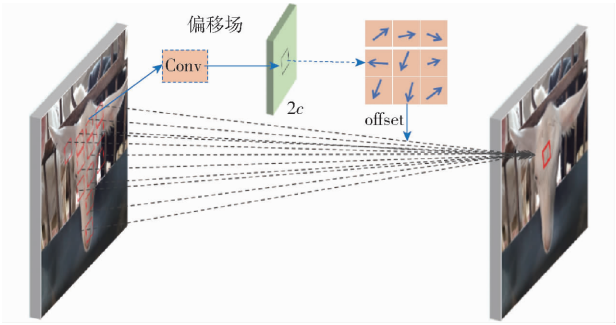

图 2　可变形卷积示意图

Fig. 2　Diagram of deformable convolution network

$$y(p_0)=\sum_{p_n\in R} w(p_n)x(p_0+p_n+\Delta p_n)$$

式中　$y(p_0)$——当前采样点

x——输入的特征图

Δp_n——采样偏移量

$w(p_n)$——p_n位置卷积核权重函数

$x(p_0+p_n+\Delta p_n)$——特征图采样位置

对输入的图像首先使用一个标准卷积网络生成共享偏移 Δp_n；随后，对每个采样点 n，使用双线性插值计算整幅输入图在偏移位置 $p_0+p_n+\Delta p_n$的值，得到 n 个特征图；最后，沿新维度堆叠这 n 个图并通过 1×1 卷积进行加权求和得到输出的特征图。

生成共享偏移 Δp_n的标准卷积操作过程主要聚焦于羊脸图像局部特征(如鼻部、纹理等)，缺乏对图像全局信息的综合依赖和融合，导致其在不同视角和尺度的羊脸图像中特征适应能力较弱。因此，在 DCNv4 模块基础上引入 CAA 注意力机制增强图像上下文的全局信息对偏移量的影响，从而提高采样点准确性。改进的 DCA 结构如图 3 所示。CAA 模块首先通过全局平均池化操作提取羊脸图像的全局语义信息，捕捉面部整体结构特征；采用 1×1 卷积提取局部特征，保留眼部、鼻角等关键区域的细节特征。对特征图分别沿宽度和高度方向进行深度可分离卷积操作：水平方向采用$(1\times k_b)$卷积核建模水平方向的远程依赖，垂直方向采用$(k_b\times1)$卷积核建模对垂直方向的远程依赖。依赖范围通过动态扩展公式$k_b=11+2n$ 确定(n 为当前网络层数)：浅层($n=1$、$k_b=13$)捕获局部细节特征之间的依赖关系(如眼睛与鼻子之间的相对位置关系)，深层($n=4$、$k_b=19$)捕获边缘轮廓对细节特征的位置关系，通过这种动态调整机制使得网络自适应地关注不同尺度和形变的特征关联。最后通过 1×1 卷积融合通道与空间信息，生成融合上下文关系的注意力权重特征图。

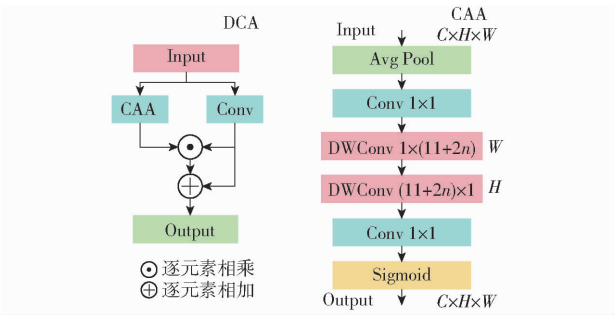

图 3　DCA 模块

Fig. 3　Module of DCA

DCA 通过将 CAA 提取的全局上下文注意力权重特征与 Conv 卷积提取的特征做逐元素相乘加权，

再通过残差逐元素相加完成特征融合,生成最终的动态卷积的采样偏移量,从而增强模型对高形变图像的适应性,提升模型对不同视角和尺度下的羊脸面部关键点检测能力。

1.2 小目标检测层

在 YOLO v8n - pose 模型中,随着 Backbone 网络深度增加,输入图像(640 像素 ×640 像素)被转换为 P1(320 像素 ×320 像素)、P2(160 像素 ×160 像素)、P3(80 像素 ×80 像素)、P4(40 像素 ×40 像素)、P5(20 像素 ×20 像素)一系列尺度递减的特征图。其中,由浅层(P1 和 P2)生成的大尺寸特征图空间分辨率高,包含丰富的图像细节信息;由深层(P3、P4 和 P5)生成的小尺寸特征图则包含丰富的语义信息,但其空间分辨率较低。在目标检测阶段,Neck 网络仅融合 Backbone 网络深层的多尺度特征,为后续检测头(Head)提供特征表示。

然而,在远距离拍摄场景中,羊脸面部关键点变小,这时由于深层特征图的分辨率较低,对于图像中的细节却缺乏足够的表达能力,且易受背景噪声干扰。这会导致关键点检测精度下降。因此,针对这一问题,通过增加 1 个 160 像素 ×160 像素的小目标尺度检测层增强较小羊脸面部关键点检测精度,结构如图 4 所示。

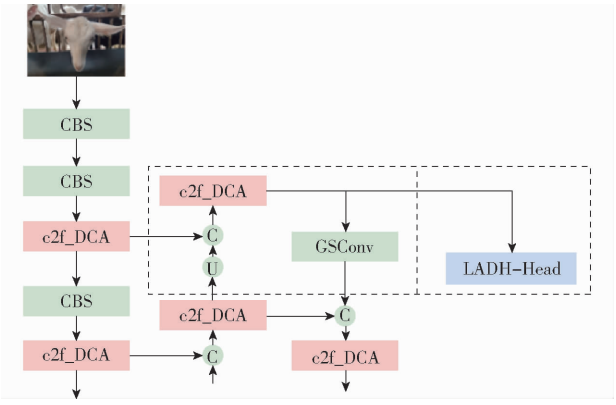

图 4 小目标尺度检测层结构

Fig. 4 Small-target scale detection layer structure

为实现浅层细节与深层语义的高保真融合,选择将深层特征图通过上采样(upsample)操作提升至 160 像素 ×160 像素(分辨率)。这一特定尺度与主干网络第 1 层 c2f_DCA 模块输出的浅层特征图尺寸一致,可直接进行拼接(Concat)操作实现两者的特征融合,最大程度保留浅层特征中重要的细节信息,避免因分辨率不匹配导致的插值失真。通过将浅层细节和深层语义特征进行融合,增强模型对小目标的关注,使得模型能够更好地提取细节特征,确保小目标关键信息不被丢失。

在多尺度特征融合过程中,随着网络深度增加以及特征图多次缩放操作,部分语义信息可能会丢失,同时也会产生冗余特征。为解决此问题,选择 GSConv[25] 卷积对特征图进行降维操作,在减少冗余信息的同时,尽可能保持网络语义信息。其操作流程如图 5 所示。

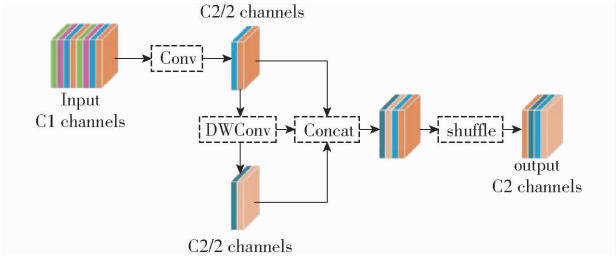

图 5 GSConv 模块

Fig. 5 GSConv module

GSConv 结合了 Conv 和 DWConv 以及混洗操作的优势。通过将 Conv 生成的通道信息均匀渗透到 DWConv 输出的特征图中,保持了通道之间的交互,保留了 DWConv 的计算效率。提升了模型准确性,同时保持了轻量级设计,优化了性能,确保模型能够继续高效地进行后续的推理任务。

1.3 LADH - Head 检测头

检测头(Head)作为 YOLO v8n - pose 模型核心模块之一,负责输出物体边界框和关键点位置,是帮助模型在进行目标定位、精准预测每个关键点坐标的关键模块,直接影响模型检测准确性。LADH - Head 检测头是基于非对称去耦头[26](ADH)的一种轻量级非对称检测头,其结构如图 6 所示。

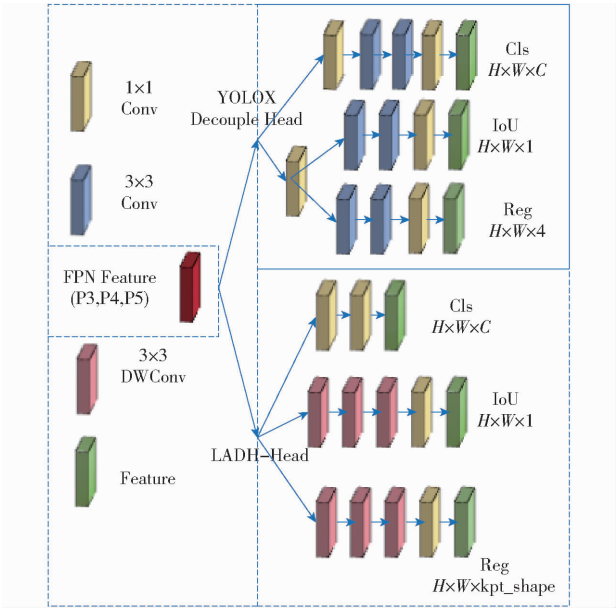

图 6 LADH - Head 与 YOLOX - DH 结构

Fig. 6 Structure of LADH - Head and YOLOX - DH

与传统的 YOLOX 解耦头不同,LADH - Head 通过将分类、回归和 IoU 任务解耦至不同通道,确保每个任务独立优化,避免了任务间的干扰。使得分类

任务专注于识别羊脸类别，回归任务负责精准定位关键点，IoU 分支则独立优化目标框与实际羊脸的重叠度。这种任务解耦提高了羊脸关键点检测精度。

LADH－Head 采用 3×3 深度可分离卷积（DWConv）替换 YOLOX 解耦头中的 3×3 普通卷积，通过 DWConv 将卷积操作分解为深度卷积和逐点卷积，模型参数量减少 9×10^5（见 3.3 节消融试验部分），同时有效地增大了卷积的感受野，有助于更好地捕捉羊脸全局和局部关键点特征。此外，逐点卷积通过 1×1 卷积核操作混合不同通道间的信息，在不改变特征图空间维度情况下，使得不同通道在同一位置的特征可以有效融合，增强通道间的特征融合能力。从而在处理羊场复杂场景和多尺度目标时，能够提升模型捕捉羊脸面部关键位置信息的敏感度。

2　试验数据

2.1　数据采集与处理

数据采自西北农林科技大学萨能奶山羊养殖基地和宁夏地区农户养殖场，选取奶山羊、湖羊和滩羊 3 个品种作为研究对象。使用 Canon EOS 200D Ⅱ型数码相机拍摄，分辨率为 1 920 像素×1 080 像素，帧率为 25 f/s，采用手持相机在近距离（0.3～1 m）、远距离（1～3 m）条件下采集 7 只滩羊、8 只湖羊及 15 只奶山羊个体的视频数据。拍摄时覆盖正脸与侧脸视角，并动态调整角度以保证数据集具有足够的多样性。通过 FFmpeg 对视频进行关键帧提取，针对连续帧图像相似度过高的问题，采用结构相似度指数[27]（Structural similarity index measure，SSIM）进行图像过滤。SSIM 通过比较图像方差、协方差和平均强度衡量图像之间的相似度，其取值为 0～1，值越接近 1 表示图像越相似。当 SSIM 相似度指数大于阈值 0.8 时，自动筛选并移除重复图像，之后通过人工检查去除羊脸模糊或面部信息缺失的样本。最终，筛选出 8 154 幅羊脸图像，包括正脸 2 658 幅、三分之一侧脸 2 788 幅和三分之二侧脸 2 708 幅。按比例 8∶1∶1 划分为训练集（6 523 幅）、验证集（815 幅）和测试集（816 幅）。图 7 展示了不同场景下部分数据。

在羊脸面部关键点选择上，参照 COCO 数据集对动物脸部关键点的选取格式，以及考虑到关键点选取要确保在不同姿态、表情、光照条件下都能稳定识别，同时要对姿态、表情等变形具有较强的适应性[28－29]。因此，对羊只正脸选取左眼、右眼、左鼻角、右鼻角、鼻尖 5 个中心部位为关键点；对羊只三分之一侧脸选取可见眼、可见鼻角、可见嘴角 3 个中心部位为关键点；对羊只三分之二侧脸选取可见眼、可见鼻角、可见嘴角、鼻尖 4 个中心部位为关键点。

采用 Labelme 标注软件对羊脸面部框和面部关

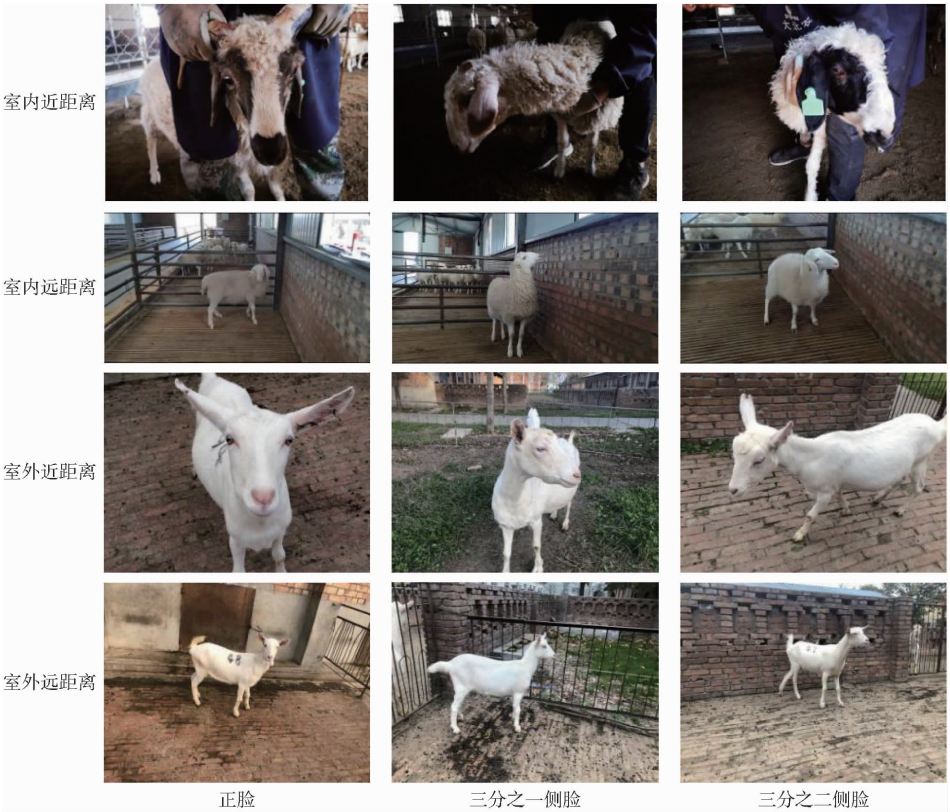

图 7　不同场景下羊脸数据

Fig. 7　Data of sheep faces in different scenarios

键点进行标记,将标注结果作为关键点真实值。

2.2 试验环境

模型训练所用计算机硬件配置:处理器为 AMD 锐龙 R5 5600,显卡为 NVIDIA GeForce RTX 4070,采用 Windows 10(64 位)操作系统。训练软件环境:Python 版本为 3.8,PyTorch 版本为 1.12.1,采用 CUDA 11.6 深度学习计算框架。

2.3 关键点检测评价指标

使用 OKS[23](Object keypoint similarity)通过欧氏距离量化预测关键点与真实关键点向量之间的相似性。采用精确率(Precision, P)衡量模型查准率;采用召回率(Recall, R)衡量模型查全率;采用平均精度(Average Precision, AP)衡量模型在全召回范围内的整体检测精度;采用平均精度均值(mean Average Precision, mAP)作为多类别检测任务的综合评价指标,即所有关键点类别平均精度(AP)平均值,用于全面评估模型全局检测性能。mAP50 指 IoU 为 0.5 时平均精度均值,mAP95 指 IoU 为 0.95 时平均精度均值,参数量指网络需要训练调整的全部权重数量。采用帧率(Frames Per Second, FPS)衡量模型实时推理性能。

3 试验结果与分析

3.1 不同检测模型性能测试

为验证本文模型同其他模型性能优势,选用 YOLO v5s-pose、HarNet[30]、YOLO v8n-pose、YOLO v10n-pose、Openpose[31] 共 5 个模型在测试集上分别进行试验,结果如表 1 所示。

表 1 不同模型性能测试结果

Tab.1 Comparative experiments of different models

模型	AP/%	mAP50/%	mAP95/%	参数量	帧率/(f·s^{-1})
YOLO v5s-pose	71.9	90.4	60.8	1.37×10^7	150.3
HarNet	75.3	91.6	67.1	1.31×10^7	132.1
YOLO v8n-pose	79.9	92.4	68.8	1.28×10^7	157.6
YOLO v10n-pose	80.0	92.7	67.4	1.44×10^7	148.8
Openpose	78.5	89.4	57.3	2.21×10^7	55.3
YOLO v8n-DCL-pose	85.1	97.5	78.7	1.34×10^7	146.5

由表 1 可知,YOLO v5s-pose 平均精度为 71.9%,帧率为 150.3 f/s,相较其他模型,其检测精度较低,但推理速度较高。主要因其特征提取模块较为简单,不能有效处理复杂的目标细节和小的特征关键点。HarNet 模型 AP 较 YOLO v5s-pose 有所提升,达到 75.3%,但推理速度有所下降。这主要是基于 HarNet 模型引入调和注意力机制,通过增加模型参数,模型能够较好地捕捉图像中的长程依赖和局部细节特征,从而提升检测精度。YOLO v8n-pose 平均精度达到 79.9%,mAP95 达到 68.8%。原因是 YOLO v8n-pose 在 YOLO v5s 的基础上加强了特征提取和模型设计的优化,同时引入多尺度特征融合技术,能够保留更多面部细节信息,使得模型在处理细小关键点特征和复杂背景时,保持较高的准确性。虽 YOLO v10n-pose 平均精度超过 YOLO v8n-pose,但其 mAP95 不及 YOLO v8n-pose,原因可能是 YOLO v10n-pose 模型架构更侧重于轻量化设计,导致对小目标的优化不足,在细小目标的关键点检测中有局限性。Openpose 平均精度为 78.5%。由于其采用两阶段设计,导致模型参数量较大,达到 2.21×10^7,计算开销较高,推理速度较慢,影响了整体性能。YOLO v8n-DCL-pose 模型在保障实时性(146.5 f/s)的同时,平均精度为 85.1%,mAP50 和 mAP95 分别为 97.5% 和 78.7%,在试验的所有模型中表现最优。验证了本研究模型有效性。图 8 展示了不同模型的关键点检测对比效果。由图 8 可看出,YOLO v5s-pose、HarNet、YOLO v8n-pose、Openpose 模型对于正脸的面部关键点,基本可准确检测到。然而对于侧脸的关键点,几个模型都出现不同程度偏移。YOLO v8n-DCL-pose 模型对于不同视角的羊脸关键点基本能检测到,且位置与手工标记的差异较小。表明了 YOLO v8n-DCL-pose 模型可以更好地检测出不同视角下羊脸关键点,在处理多视角羊脸关键点检测任务时具有显著优势。

3.2 不同注意力机制性能测试

为验证在改进可变形卷积时选取 CAA 模块作为特征提取的有效性,在模型同一位置,添加 ECA[32](Efficient channel attention)、CA[33](Coordinate attention)、CBAM[34](Convolutional block attention module)、SE[35](Squeeze and excitation)、CAA 进行多组试验,结果如表 2 所示。

由表 2 可知,引入不同类型的注意力机制后,检测模型精度都略有提升。引入 ECA 后,模型精确率和召回率略有提升,mAP50 也有所提高,达到 93%,mAP95 为 69.8%,表明 ECA 通过增强通道特征的自适应权重分配,提高了重要特征的关注度。采用 CA 模块后,精确率和召回率进一步提升,mAP50 为 93.6%,mAP95 为 71.5%,表明其自适应的空间注意力机制有助于细节特征的捕捉,因此关键点检测精度也有所提高。CBAM 在通道和空间维度的注意力融合上取得了平衡,但由于空间注意力的冗余处理,使得在引入 CBAM 模块后,模型虽在召回率上有所提高,但在精确率上的表现不及 CA。

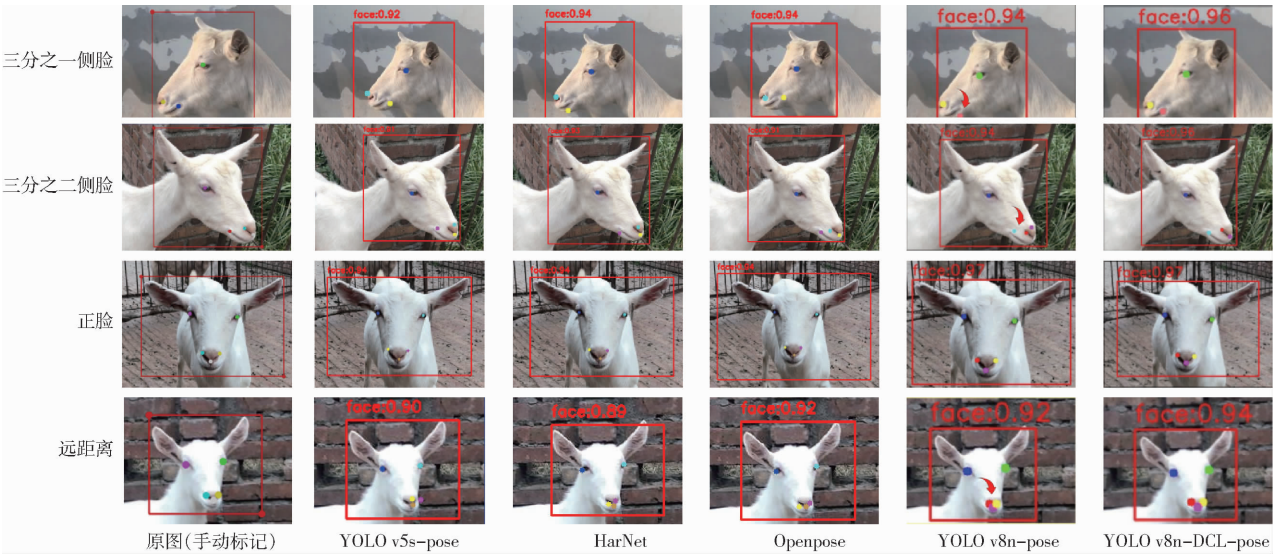

图 8　不同模型关键点检测结果

Fig. 8　Results of key point detection in different models

表 2　注意力机制性能测试结果

Tab. 2　Performance test results of attention mechanisms

%

注意力机制	精确率	召回率	mAP50	mAP95
无	81.3	92.1	92.7	69.0
ECA	82.1	92.9	93.0	69.8
CA	82.7	93.0	93.6	71.5
CBAM	82.3	92.8	93.0	70.9
SE	82.1	92.7	92.9	69.4
CAA	83.5	93.4	94.0	72.4

SE 模块通过自适应通道调节增强了通道表达能力，但因空间特征优化较弱，关键点 AP 仅小幅提升。CAA 结合空间和通道的注意力优化，提升模型捕捉细节特征的能力，在引入该模块后，模型精确率提升至 83.5%，召回率达 94.0%，mAP95 达 72.4%，取得了最优性能。因此选择融入 CAA 注意力模块作为改进可变形卷积模块的优化方案更为合理。

3.3　消融试验

在 YOLO v8n－pose 模型基础上逐个加入本文的改进方法进行测试，以分析本文提出的不同优化模块对 YOLO v8n－pose 模型性能的影响。试验结果如表 3 所示。

表 3　消融试验结果

Tab. 3　Ablation experiment results

LADH－Head	DCA	小目标检测层	精确率/%	mAP50/%	mAP95/%	参数量	帧率/(f·s^{-1})
×	×	×	79.9	92.4	68.8	1.28×10^7	157.6
√	×	×	81.7	93.6	69.9	1.19×10^7	183.2
×	√	×	83.5	94.0	72.4	1.30×10^7	130.4
×	×	√	82.9	93.2	74.0	1.32×10^7	154.7
√	√	√	85.1	97.5	78.7	1.34×10^7	146.5

注：√表示采用该方法，×表示未采用该方法。

由表 3 可知，LADH－Head、DCA 和小目标检测层的引入不同程度地提升了模型性能。模型在改进前的精确率为 79.9%，mAP50 为 92.4%，mAP95 为 68.8%，帧率为 157.6 f/s。在加入 LADH－Head 后，精确率提升至 81.7%，mAP50 提高 1.2 个百分点，mAP95 提升 1.1 个百分点，参数量减少 9×10^6，推理速度提升 16.2%。验证了 LADH－Head 作为轻量化解耦检测头，可以通过解耦每个任务，减少任务相互干扰，从而提升分类和回归任务精度，增强羊脸关键点检测精度。加入 DCA 后，推理速度有所下降，但精确率提高至 83.5%，mAP50 提高 1.6 个百分点，mAP95 提升 3.6 个百分点。表明 DCA 可通过优化卷积操作，提高模型对羊脸形变和复杂特征的适应能力。加入小目标检测层后，模型推理速度稍有下降，精确率提升至 82.9%，mAP50 提高 0.8 个百分点，mAP95 提升 5.2 个百分点，是 3 个方案中 mAP95 提升最明显的。这是因为模型引入浅层特征，提升了特征图空间细节保留能力。使模型在处理远距离、多形态的羊脸关键点时，能更敏感地捕捉微小面部特征（如眼角、鼻尖），从而提升关键点定

位精度,尤其在高 IoU 阈值要求下改善显著。最终,将 3 个改进策略全部引入后,精确率达到 85.1%,mAP50 提升 5.1 个百分点,mAP95 提高 9.9 个百分点。显著提高了模型整体性能,尤其在 mAP95 的表现,这是由于 LADH - Head 通过任务解耦减少分类与回归的相互干扰,DCA 利用动态卷积适应羊脸形变特征,小目标检测层则通过高分辨率特征增强羊脸部微小关键点的定位能力。三者协同形成“结构优化-特征适应-细节增强”的互补机制,显著提升了关键点检测精度,验证了本文对 YOLO v8n - pose 模型改进的有效。

3.4 可视化分析

为验证改进模型在羊脸关键点检测任务中的有效性,从测试集中随机选取 3 幅不同面部朝向的羊脸图像,采用公理的梯度加权类激活映射(Axiom-based grad-cam,XGrad - CAM)[36]方法对羊脸关键点特征区域绘制热力图进行可视化,结果如图 9 所示。

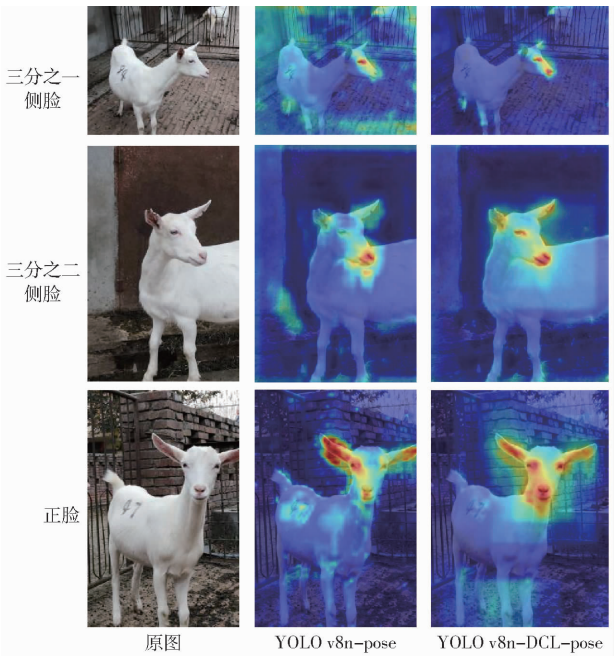

图 9 热力图效果

Fig. 9 Heatmap visualization

从图 9 可看出,YOLO v8n - pose 模型在检测羊脸面部关键点时,对于正脸图像能够较好地聚焦于面部关键点附近的特征区域,检测效果较为理想。然而,在处理 2 种不同侧脸图像时,模型易受到周围环境因素干扰,无法充分聚焦于面部关键点区域,甚至会遗漏部分关键点特征信息。YOLO v8n - DCL - pose 模型在检测羊脸正脸和侧脸图像中的关键点时,能够更加精准地聚焦于眼部、鼻部、嘴角等关键点附近的特征区域,同时对周围环境的无关特征关注显著减少。该结果进一步验证了 YOLO v8n - DCL - pose 模型在多视角羊脸关键点检测任务中所具备的优异定位能力与特征感知能力。

4 结论

(1)基于 YOLO v8n - pose 模型进行改进,构建了新的关键点检测模型 YOLO v8n - DCL - pose。采用改进的动态可变形卷积 DCNv4 模块优化模型特征提取部分,提升模型对不同大小和形状目标的特征适应能力;通过在特征提取层新增 1 个 160 像素 ×160 像素的小目标检测层,提升特征图的空间细节保留能力,保留羊脸面部更多局部细节信息,显著提升在高阈值(mAP95)要求下,羊脸面部特征关键点的检测效率;采用轻量化解耦检测头(LADH - Head)替换原结构中的检测头,分离分类与回归的特征提取路径从而消除任务间干扰,进一步提升关键点检测效率。

(2)改进的 YOLO v8n - DCL - pose 模型平均精度、mAP50、mAP95 分别达到 85.1%、97.5% 以及 78.7%,模型参数量仅有 1.34×10^7。相较于 YOLO v8n - pose 模型的 AP 提升 5.2 个百分点,mAP50 提升 5.1 个百分点,mAP95 提升 9.9 个百分点。改进的 YOLO v8n - DCL - pose 模型有效地提升了在羊脸面部朝向多变,尺度不一情况下羊脸关键点的检测精度,可为基于多视角的羊脸个体身份识别、表情识别、头部姿态估计等研究提供技术支持。

参 考 文 献

[1] 魏斌,Billah Masum,王美丽,等. 基于深度学习的羊脸检测与识别方法[J]. 家畜生态学报, 2022, 43(3): 47 - 50.
Wei Bin, Billah Masum, Wang Meili, et al. Sheep face detection and recognition method based on deep learning[J]. Journal of Livestock Ecology, 2022, 43(3): 47 - 50. (in Chinese)

[2] 杨玉洁,陈天星,石林坤,等. 偏转角度情况下 MTCNN 人脸检测算法改进[J]. 计算机与数字工程, 2023, 51(9): 2074 - 2078.
Yang Yujie, Chen Tianxing, Shi Linkun, et al. Improvement of MTCNN face detection algorithm under deflection angle[J]. Computer and Digital Engineering, 2023, 51(9): 2074 - 2078. (in Chinese)

[3] Guo Q, Wang Z, Fan D, et al. Multi-face detection and alignment using multiple kernels[J]. Applied Soft Computing, 2022, 122: 108808.

[4] 赵小虎,王轲鑫,孟献峰,等. 基于人体姿态估计的羽毛球动作评估方法研究[J]. 华中科技大学学报(自然科学版), 2024,52(11):110 - 116.

Zhao Xiaohu, Wang Kexin, Meng Xianfeng, et al. Research on badminton action evaluation method based on human pose estimation[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2024,52(11):110－116. (in Chinese)

[5] 赵俊杰.人体骨骼点提取方法及其在车内人员分类中的应用[D].长沙:湖南大学,2023.
Zhao Junjie. Human skeletal point extraction method and its application to vehicle occupant classification[D]. Changsha:Hunan University, 2023. (in Chinese)

[6] Fan Q, Liu S, Li S, et al. Bottom-up cattle pose estimation via concise multi-branch network[J]. Computers and Electronics in Agriculture, 2023, 211: 107945.

[7] Lu B, Sun Y, Yang Z, et al. HRNet: 3D object detection network for point cloud with hierarchical refinement[J]. Pattern Recognition, 2024, 149: 110254.

[8] Dong C, Tang Y, Zhang L. MDA－YOLO Person: a 2D human pose estimation model based on YOLO detection framework[J]. Cluster Computing, 2024, 27(9): 12323－12340.

[9] Xia Z, Zhou H, Yu H, et al. YOLO－MTG: a lightweight YOLO model for multi-target garbage detection[J]. Signal, Image and Video Processing, 2024, 18(6): 5121－5136.

[10] 徐红梅,杨浩,李亚林,等. 基于改进 YOLO－pose 的复杂环境下拖拉机驾驶员关键点检测[J]. 农业工程学报, 2023, 39(16): 139－149.
Xu Hongmei, Yang Hao, Li Yalin, et al. Key point detection of tractor drivers in complex environments based on improved YOLO－pose[J]. Transactions of the CSAE, 2023, 39(16): 139－149. (in Chinese)

[11] Chen L, Zhang L, Tang J, et al. GRMpose: GCN-based real-time dairy goat pose estimation[J]. Computers and Electronics in Agriculture, 2024, 218: 108662.

[12] Hua Z, Wang Z, Xu X, et al. An effective poseC3D model for typical action recognition of dairy cows based on skeleton features[J]. Computers and Electronics in Agriculture, 2023, 212: 108152.

[13] Zhang Y, Zhang W, Yu J, et al. Complete and accurate holly fruits counting using YOLOX object detection[J]. Computers and Electronics in Agriculture, 2022, 198: 107062.

[14] 刘刚,曾雪婷,刘晓文,等. 基于改进 YOLO v5－pose 的群养生猪体尺自动测量方法[J]. 农业机械学报, 2025, 56(5): 455－465.
Liu Gang, Zeng Xueting, Liu Xiaowen, et al. Automatic measurement method of body size for group health-preserving pigs based on improved YOLO v5－pose[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(5): 455－465. (in Chinese)

[15] Yang G, Qiao Y, Deng H, et al. One-stage keypoint detection network for end-to-end cow body measurement[J]. Engineering Applications of Artificial Intelligence, 2025, 146:110333.

[16] 宋怀波,华志新,马宝玲,等. 基于 SimCC－ShuffleNetV2 的轻量化奶牛关键点检测方法[J]. 农业机械学报, 2023, 54(10): 275－281.
Song Huaibo, Hua Zhixin, Ma Baoling, et al. Key point detection method for lightweight dairy cows based on SimCC－ShuffleNetV2[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(10): 275－281. (in Chinese)

[17] Wu S, Han S, Zhang J, et al. Multi-scale keypoints detection and motion features extraction in dairy cows using ResNet101－ASPP network[J]. Journal of Integrative Agriculture, 2024, 23(12):3121－3136.

[18] 黄志杰,徐爱俊,周素茵,等. 融合重参数化和注意力机制的猪脸关键点检测方法[J]. 农业工程学报, 2023, 39(12): 141－149.
Huang Zhijie, Xu Aijun, Zhou Suyin, et al. Key point detection method for pig face integrating heavy parameterization and attention mechanism[J]. Transactions of the CSAE, 2023, 39(12): 141－149. (in Chinese)

[19] 黄小平,侯现坤,郭阳阳,等. 基于改进 YOLO v7－pose 的牛脸关键点检测与姿态识别方法[J]. 农业机械学报, 2024, 55(11): 84－92.
Huang Xiaoping, Hou Xiankun, Guo Yangyang, et al. Key point detection and pose recognition method of cow face based on improved YOLO v7－pose[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(11): 84－92. (in Chinese)

[20] 张惠莉,王光远,员玉良,等. 基于 FCM－SimCC 的猪只面部关键点定位方法[J]. 农业机械学报, 2025, 56(4): 397－407.
Zhang Huili, Wang Guangyuan, Yuan Yuliang, et al. Key point localization method for pig face based on FCM－SimCC[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(4): 397－407. (in Chinese)

[21] Xiong Y, Li Z, Chen Y, et al. Efficient deformable convnets: rethinking dynamic and sparse operator for vision applications[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2024: 5652－5661.

[22] Cai X, Lai Q, Wang Y, et al. Poly kernel inception network for remote sensing detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 27706－27716.

[23] Lin T, Maire M, Belongie S, et al. Microsoft coco: common objects in context[C]// European Conference on Computer Vision, 2014: 740－755.

[24] Zhang J, Chen Z, Yan G, et al. Faster and lightweight: an improved YOLO v5 object detector for remote sensing images[J].

Remote Sensing, 2023, 15(20): 4974.
[25] Li H, Li J, Wei H, et al. Slim-neck by GSConv: a lightweight-design for real-time detector architectures[J]. Journal of Real-Time Image Processing, 2024, 21(3):111.
[26] Huang L, Li W, Tan Y, et al. YOLOCS: object detection based on dense channel compression for feature spatial solidification [J]. Knowledge-Based Systems, 2025, 310(1):436 - 441.
[27] Wang Z, Li Q. Information content weighting for perceptual image quality assessment[J]. IEEE Transactions on Image Processing, 2011, 20(5): 1185 - 1198.
[28] Yoshiki I. HOKEM: human and object keypoint-based extension module for human-object interaction detection[C]// 2023 IEEE International Conference on Image Processing (ICIP). IEEE, 2023: 271 - 275.
[29] Yu H, Xu Y, Zhang J, et al. AP - 10K: a benchmark for animal pose estimation in the wild[PP/OL]. V1. arXiv(2021 - 08 - 28)[2026 - 05 - 29]. https://arXiv.org/abs/2108.12617.
[30] Yu C, Xiao B, Gao C, et al. Lite-hrnet: a lightweight high-resolution network[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2021: 10440 - 10450.
[31] Cao Z, Hidalgo G, Simon T, et al. Openpose: realtime multi-person 2D pose estimation using part affinity fields[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(1): 172 - 186.
[32] Wang Q, Wu B, Zhu P, et al. ECA - Net: efficient channel attention for deep convolutional neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020: 11534 - 11542.
[33] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2021: 13713 - 13722.
[34] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3 - 19.
[35] Hu J, Shen L, Sun G, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,42(8):2011 - 2023.
[36] Fu R, Hu Q, Dong X, et al. Axiom-based Grad - CAM: towards accurate visualization and explanation of CNNs (BMVC2020 Oral)[PP/OL]. V1. arXiv(2020 - 08 - 05)[2025 - 05 - 29]. https://doi.org/abs/2021.02312.