

基于改进 YOLOv11 的高原牦牛检测模型研究

牛国辉¹ 钟小莉^{1,2} 宫宁宁¹
NIU Guohui ZHONG Xiaoli GONG Ningning

摘要

针对高原牦牛检测中存在的小目标检测难、多尺度特征融合不充分等问题,文章提出一种基于改进 YOLOv11 的高原牦牛检测模型——YOLOv11-MSAG。该模型在 YOLOv11 基础上,引入多尺度空洞注意力机制(MSDA)增强不同尺度特征提取能力;利用 ASFF 辅助特征融合模块改进检测头,提升多尺度特征融合效果;采用 YOLOv9 的 GELAN 模块替换 C3k2 结构优化网络性能。在自建高原牦牛数据集上的实验结果表明,YOLOv11-MSAG 的 mAP@0.5 和 mAP@0.5: 0.95 相较于原模型分别提高了 1.1% 和 1.9%,达到了 94.4% 和 67.7%,有效提升了高原牦牛检测的准确性和鲁棒性。

关键词

高原牦牛检测; YOLOv11; 多尺度空洞注意力; 辅助特征融合; GELAN 模块

doi: 10.3969/j.issn.1672-9528.2026.04.004

0 引言

牦牛作为青藏高原的标志性畜种,在当地生态、经济和文化中具有不可替代的重要地位。高原地区独特的地理环境和气候条件,如高海拔、低温等,使得牦牛的生存与繁衍面临诸多挑战。准确掌握牦牛的数量、分布及活动情况,对于牦牛资源的合理保护、科学管理以及当地畜牧业的可持续发展至关重要。

传统的牦牛数量统计和监测方法主要依赖人工实地观察,这种方式不仅耗费大量的人力、物力和时间,而且在复杂的高原地形和恶劣的气候条件下,效率低下且准确性难以保证。随着计算机视觉技术的飞速发展,目标检测算法为高原牦牛检测提供了新的解决方案。YOLO (you only look once) 系列模型以其高效的检测速度和良好的检测精度,在众多的目标检测任务中得到了广泛应用。然而,高原牦牛检测场景具有其特殊性,存在小目标牦牛难以检测、不同尺度牦牛特征融合不充分等问题,使得现有的 YOLOv11 模型在该场景下的性能有待进一步提升。

目前目标检测领域主要分为两阶段和单阶段检测器两大类,两阶段检测器以 Faster R-CNN (faster region-based convolutional neural network) 为代表^[1],在复杂背景下具备较高精度,但推理速度较慢、部署成本较高^[2]。单阶段检测器如 SSD (single shot multibox detector) 与 RetinaNet 强调端到端实时性,在速度-精度权衡上表现更好^[3-4]。YOLO 系列作为主流方法,其中 YOLOv8 采用解耦检测头设计,训练与部署

便捷,但对极小目标检测不足^[5];YOLOv9 通过泛化高效层聚合网络 (generalized efficient layer aggregation network, GELAN) 强化特征聚合,在相近算力下取得更优的效率-精度平衡,但结构更复杂^[6];YOLOv10 采用去非极大值抑制的近似端到端范式,显著降低后处理延迟,但训练配置更为敏感^[7];YOLOv11 在数据增广、训练配方、骨干和检测头细节上进一步优化,但对小目标的上限仍受输入分辨率与金字塔设计制约^[8-9]。基于 Transformer 的 DETR (detection transformer) 虽然实现端到端检测,但收敛较慢、对算力敏感^[10]。

为了更好地适应高原牦牛检测需求,本文对 YOLOv11 模型进行了改进。通过引入注意力机制与使用特征融合模块改进检测头,对不同尺度牦牛特征的提取与融合能力得到提升,从而提高检测的准确性与鲁棒性,为高原牦牛资源的科学管理和保护提供支撑。

1 改进 YOLOv11 模型

1.1 引入 MSDA 注意力机制

在目标检测任务中,模型常面临两个关键问题:一是难以聚焦于目标的关键特征(如高原牦牛的轮廓、毛色等),易受复杂背景的干扰;二是不同层级特征间的信息传递存在损耗,导致浅层细节特征(如小牦牛的边缘)与深层语义特征(如牦牛群的整体形态)难以有效结合。

注意力机制通过对关键信息赋予更高权重、降低无关信息的影响,能够有效解决上述问题。因此,本文引入多尺度空洞注意力机制 (multi-scale dilated attention, MSDA),MSDA 的主要思想是通过线性投影得到特征图 X 的相应查询、键和值。然后,将特征图的通道分成 n 个不同的头部,并在不同的头部中以不同的扩张率执行多尺度滑动窗口膨胀注意

1. 青海民族大学智能科学与工程学院 青海西宁 810007

2. 青海民族大学公共计算机实验教学示范中心

青海西宁 810007

力来提高模型的处理效率和检测精度。其核心是在自注意力机制的基础上,结合浅层特征的局部性和稀疏性特点,通过多尺度空洞卷积扩展感受野,实现对不同尺度语义信息的高效聚合,同时减少计算冗余^[11]。MSDA 结构如图 1 所示。

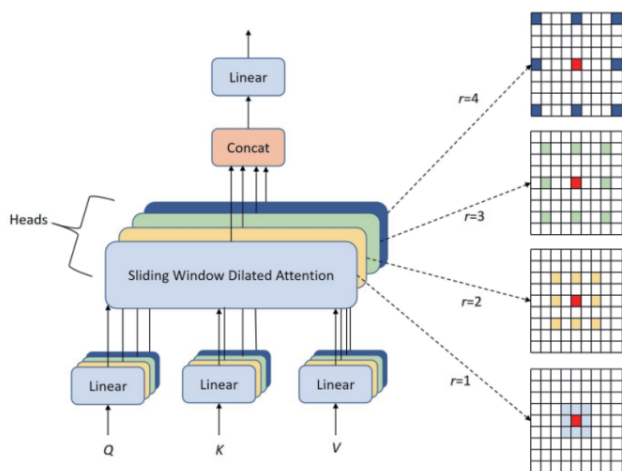


图 1 MSDA 结构图

SWDA 表示为:

$$H = \text{SWDA}(Q_i, V_i, K_i, r_i), \quad 1 \leq i \leq n \quad (1)$$

式中: Q_i 为来自解码器的向量; V_i 和 K_i 为来自编码器的向量。对于特征图上的位置 (x, y) , SWDA 以 (x, y) 为中心,在大小为 $s \times s$ 的滑动窗口内稀疏选择键 (K_i) 和值 (V_i) 进行自注意力计算,通过膨胀率 r_i 控制稀疏度 (本文设置膨胀率 $r_i = 1, 2, 3, 4$, 对应不同注意力头的有效感受野大小为 3×3 、 5×5 、 7×7 和 9×9)。最终拼接各 SWDA 输出,送入线性层完成特征聚合,让模型更聚焦关键信息,适配复杂场景下的特征传递需求。

1.2 辅助特征融合模块 ASFF 改进检测头

自适应空间特征融合 (adaptively spatial feature fusion, ASFF) 方法针对单次对象检测任务提出,它的主要创新是采用自适应空间融合方式,动态过滤特征中的冲突信息,有效增强模型的尺度不变性。根据 2019 年 Liu 等^[12] 的研究表明,将 ASFF 应用于 YOLOv3 可以显著提高在 MS COCO 数据集上的检测性能,实现了速度与准确性的平衡。ASFF 方法可以通过反向传播进行训练,与模型无关,并且引入的计算开销很小,使其成为现有对象检测框架的一种实用增强。ASFF 结构如图 2 所示。

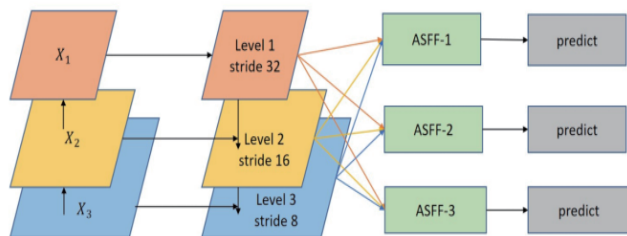


图 2 ASFF 结构图

自适应融合特征 ASFF 由不同层的特征分别乘上权重参数 α^l 、 β^l 和 γ^l 后相加得到,用公式表示为:

$$y_{ij}^l = \alpha_{ij}^l X_{ij}^{(1 \rightarrow l)} + \beta_{ij}^l X_{ij}^{(2 \rightarrow l)} + \gamma_{ij}^l X_{ij}^{(3 \rightarrow l)} \quad (2)$$

式中: y_{ij}^l 为新的融合特征; α_{ij}^l 、 β_{ij}^l 、 γ_{ij}^l 为不同层的权重参数; X_{ij} 为不同层的特征值。

权重参数 α 、 β 和 γ 是通过特征图进行 1×1 卷积操作得到的,经过 concat 操作将三者通道维度上拼接,然后通过 softmax 函数处理,使得它们的取值范围都在 0~1 之间,且总和为 1,具体用公式表示为:

$$\alpha_{ij}^l = \frac{e^{\lambda_{\alpha_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}} \quad (3)$$

$$\beta_{ij}^l = \frac{e^{\lambda_{\beta_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}} \quad (4)$$

$$\gamma_{ij}^l = \frac{e^{\lambda_{\gamma_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}} \quad (5)$$

$$\alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l = 1 \quad (6)$$

式中: $\lambda_{\alpha_{ij}}^l$ 、 $\lambda_{\beta_{ij}}^l$ 和 $\lambda_{\gamma_{ij}}^l$ 为 1×1 卷积操作的结果。通过这种方式,ASFF 能够根据高原牦牛目标在图像中的具体尺度和位置,自适应地调整不同层级特征的融合权重,从而有效提升对不同尺度高原牦牛的检测能力。

1.3 GELAN 模块替换 C3k2 结构

在现有的 C3k2 模块中,存在特征聚合方式相对固定,梯度路径较少且传播深度有限,易在深层网络中出现信息衰减;对多尺度与长程依赖的建模能力不足,导致小目标与复杂场景识别受限等问题。因此,YOLOv9 中提出了 GELAN 模块,该模块基于高效层聚合网络

(efficient layer aggregation network, ELAN) 与跨阶段局部网络 (cross stage partial network, CSPNet) 设计而成。其核心在于模块化设计,使得网络可以根据不同计算需求与资源限制,灵活地调整自身结构。这使得它既能应用于计算资源受限的轻量级模型,也能拓展到对准确性要求更高的复杂模型中^[13]。GELAN 结构图如图 3 所示。

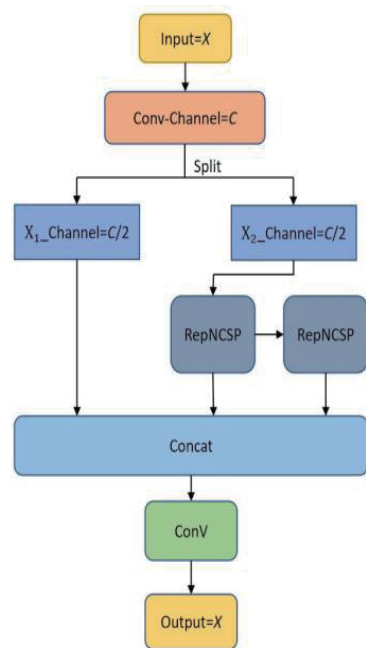


图 3 GELAN 结构图

1.4 改进后的 YOLOv11-MSAG 模型结构

通过引入 MSDA 注意力机制、用 ASFF 模块改进检测头以及使用 GELAN 模块替换 C3k2 结构的方式，对原有 YOLOv11 模型进行改进，改进后得到的 YOLOv11-MSAG 模型结构图如图 4 所示。

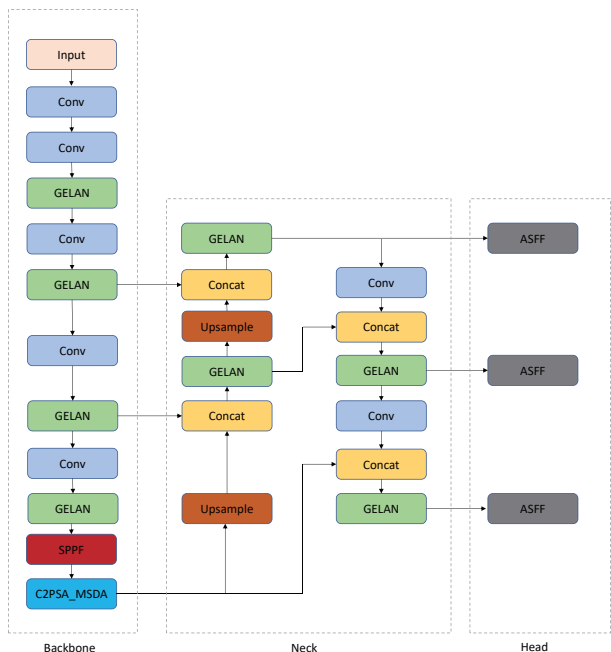


图 4 YOLOv11-MSAG 模型结构图

2 实验与结果分析

2.1 实验数据集

实验使用的数据集为自建的高原牦牛数据集，通过在高原地区实地拍摄以及收集相关影像资料构建而成。总共包含 3 230 张图片，将数据集图片按照 7:2:1 的比例随机划分为训练集、验证集和测试集。

2.2 评价指标

采用精确率（Precision）、召回率（Recall）、平均精度均值 mAP（mean average precision）和 F_1 分数作为评价指标^[14]。其中，精确度（P）表示所有预测为正类的样本中，实际为正类的比例；召回率（R）表示所有真实为正类的样本中，成功预测为正类的比例；mAP 表示不同类别精度的平均值，用于综合评估模型在多个类别上的表现； F_1 分数表示精确率和召回率的调和平均数。其计算公式分别为：

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N AP_i \quad (9)$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

式中：TP（True Positives）表示实际为正类且正确预测为正类的样本数；FP（False Positives）表示实际为负类但误判为正类的样本数；FN（False Negative）表示实际为正类但误判为负类的样本数^[15]；AP（Average Precision）表示精确度与召回率曲线下，纵轴为精度、横轴为召回率所围成的面积；N 表示数据集中类别的总数。

本文将改进后的 YOLOv11-MSAG 模型与 YOLOv8n、YOLOv9t、YOLOv10n 和 YOLOv11n 等多个模型在同一个数据集下进行了对比实验，实验结果如表 1 所示。

表 1 模型对比实验

模型	P/%	R/%	F_1	mAP@0.5 /%	mAP@0.5:0.95 /%
YOLOv8n	91.8	86.5	0.89	92.8	64.4
YOLOv9t	89.2	89.3	0.89	93.6	65
YOLOv10n	89.7	87	0.88	92.5	64.7
YOLOv11n	90	88.7	0.89	93.3	65.8
YOLOv11-MSAG	92.5	89	0.91	94.4	67.7

从表 1 中可以看出，改进后的 YOLOv11 模型在各项指标上均有显著提升。mAP 值相较于原模型分别提高了 1.1% 和 1.9%，精确率、召回率和 F_1 分别提升了 2.5%、0.3% 和 0.02。这表明模型通过改进，有效地增强了对高原牦牛特征的提取和融合能力，提高了检测的准确性和鲁棒性。

2.3 消融实验

为了对改进部分的有效性进行验证，进行了消融实验。分别对仅引入 MSDA 模块、仅使用 ASFF 改进检测头、仅替换 GELAN 模块以及同时采用这三种改进措施的模型进行训练和测试，结果如表 2 所示。

表 2 消融实验结果

模型	P/%	R/%	F_1	mAP@0.5 /%	mAP@0.5:0.95 /%
YOLOv11n	90	88.7	0.89	93.3	65.8
+ MSDA	92.3	89.4	0.91	94.2	67.1
+ ASFF	91	88	0.89	94.3	62.9
+ GELAN	91.1	88.1	0.90	93.7	65.4
YOLOv11-MSAG	92.5	89	0.91	94.4	67.7

从表 2 中可以看出，单独引入 MSDA 模块、使用 ASFF 改进检测头以及替换 GELAN 模块都能在一定程度上提升模型性能，但同时采用这三种改进措施时，模型性能提升最为

显著。说明各个改进部分之间具有协同作用，共同提高了模型对高原牦牛的检测能力。

2.4 可视化分析

通过可视化工具对比改进前后模型的检测结果，如图5~6所示，可以看出牦牛检测置信度提高，体现了 MSDA 注意力模块的有效性；改进后边界框更精准贴合牦牛轮廓，印证了 ASFF 特征融合的价值；同时，面对远距离小牦牛、半遮挡牦牛等情况，检测效果更稳定，凸显了 GELAN 模块对复杂场景的适配性。这些结果表明，改进后的模型，有效解决了高原牦牛检测中的痛点问题，检测精度与鲁棒性显著提升，展现出更优的检测性能。

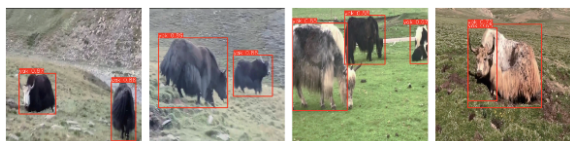


图5 YOLOv11n

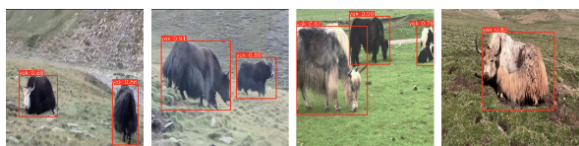


图6 YOLOv11-MSAG

3 结论

本文针对高原牦牛检测任务，对 YOLOv11 模型进行了改进。通过引入 MSDA 多尺度空洞注意力模块，提升了模型对不同尺度牦牛特征的提取能力；利用 ASFF 模块改进检测头，优化了多尺度特征融合效果；采用 GELAN 模块替换 C3k2 结构，增强了网络的特征处理能力。在自建的高原牦牛数据集上的实验结果表明，改进后的模型在平均精度均值等指标上相较于原 YOLOv11 模型有显著提升，有效提高了高原牦牛检测的准确性和鲁棒性。

参考文献

- [1] 赵凤, 李永恒, 李晶, 等. 基于改进 YOLOv4-tiny 的轻量化室内人员目标检测算法 [J]. 电子与信息学报, 2022, 44(11): 3815-3824.
- [2] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137-1149.
- [3] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Computer Vision-ECCV 2016. Berlin: Springer, 2016: 21-37.
- [4] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense

- object detection[C]// 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 2980-2988.
- [5] TALIB M, AL-NOORI A H Y, SUAD J. YOLOv8-CAB: improved YOLOv8 for real-time object detection[EB/OL]. (2024-01-24)[2026-04-07]. <https://kijoms.uokerbala.edu.iq/home/vol10/iss1/5/>.
- [6] WANG C Y, YE H I H, MARK LIAO H Y. YOLOv9: Learning what you want to learn using programmable gradient information[C]//Computer Vision-ECCV 2024. Berlin: Springer, 2024: 1-21.
- [7] WANG A, CHEN H, LIU L H, et al. YOLOv10: real-time end-to-end object detection[EB/OL]. (2024-10-30)[2026-04-07]. <https://doi.org/10.48550/arXiv.2405.14458>.
- [8] KHANAM R, HUSSAIN M. YOLOv11: an overview of the key architectural enhancements[EB/OL]. (2024-10-30)[2026-04-07]. <https://doi.org/10.48550/arXiv.2410.17725>.
- [9] HE L H, ZHOU Y Z, LIU L, et al. Research on object detection and recognition in remote sensing images based on YOLOv11[J]. Scientific reports, 2025, 15: 14032.
- [10] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Computer Vision - ECCV 2020. Berlin: Springer, 2020: 213-229.
- [11] JIAO J Y, TANG Y M, LIN K Y, et al. DilateFormer: multi-scale dilated transformer for visual recognition[J]. IEEE transactions on multimedia, 2023, 25: 8906-8919.
- [12] LIU S T, HUANG D, WANG Y H. Learning spatial fusion for single-shot object detection[EB/OL]. (2019-11-25)[2026-04-07]. <https://doi.org/10.48550/arXiv.1911.09516>.
- [13] 曹建林, 周健, 陈雪元, 等. 基于轻量化 GELAN 的多种类管道病害检测 [J]. 市政技术, 2025, 43(2): 117-125.
- [14] 戴激光, 张晓宇, 吴玉洁. 复杂场景下违规停放汽车的细粒度检测方法 [J]. 测绘科学, 2024, 49(8): 164-172.
- [15] 王国明, 贾代旺. 基于 YOLOv8 的小目标检测模型的优化 [J]. 计算机工程, 2025, 51(12): 294-303.

【作者简介】

牛国辉 (2002—), 男, 河南焦作人, 硕士, 研究方向: 图像处理。

钟小莉 (1975—), 女, 陕西旬阳人, 硕士, 教授, 研究方向: 计算机应用和图像处理。

宫宁宁 (1998—), 女, 山东聊城人, 硕士, 研究方向: 图像处理。

(收稿日期: 2025-09-10 修回日期: 2026-04-09)