

基于YOLOv8的江豚性行为检测与识别模型

陈慧妍¹, 戈潘缘元², 郝玉江², 张荣福^{1*}

¹上海理工大学光电信息与计算机工程学院, 上海

²中国科学院水生生物研究所, 湖北 武汉

收稿日期: 2025年3月7日; 录用日期: 2025年3月31日; 发布日期: 2025年4月9日

摘要

对江豚性行为的研究有助于此濒危物种的保育工作。传统方式依靠人眼观察, 存在效率低、成本高、易疲劳等问题。随着深度学习技术的发展, 中国科学院水生生物研究所配备的视觉监控设备可借助目标检测与识别技术提高观测效率。本文首先创建YFPSB数据集, 包含2901张图像, 涵盖五种行为类别。针对水下画质模糊、江豚与背景混淆等问题, 本文建立了YOLOv8n-DBTA模型, 首先通过自动伽马校正和暗通道去雾对图像预处理, 接着设计双主干特征融合框架, 增加更多特征融合路线, 保留更多底层细节和特征以提升检测能力, 然后设计任务对齐检测头, 将位置与类别信息交互, 学习江豚行为与位置的关系, 最后进行模型实验分析, 结果表明: 模型在mAP_{0.50}和mAP_{0.50:0.95}指标分别达到97.9%和79.1%, 优于其他主流模型。

关键词

江豚, 目标检测, 深度学习

Yangtze Finless Porpoise Sexual Behavior Object Detection Model Based on YOLOv8

Huiyan Chen¹, Panyuanyuan Ge², Yujiang Hao², Rongfu Zhang^{1*}

¹School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

²Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan Hubei

Received: Mar. 7th, 2025; accepted: Mar. 31st, 2025; published: Apr. 9th, 2025

Abstract

Research on the sexual behavior of Yangtze finless porpoises contributes to the conservation of this

*通讯作者。

文章引用: 陈慧妍, 戈潘缘元, 郝玉江, 张荣福. 基于 YOLOv8 的江豚性行为检测与识别模型[J]. 建模与仿真, 2025, 14(4): 118-128. DOI: 10.12677/mos.2025.144270

endangered species. Traditional methods rely on human observation, which is inefficient, costly, and prone to observer fatigue. With advancements in deep learning technology, the visual monitoring system at The Finless Porpoise Pavilion at the Institute of Hydrobiology, Chinese Academy of Sciences, can leverage object detection and recognition techniques to improve observation efficiency. In this study, we first constructed the YFPSB dataset, consisting of 2901 images covering five behavioral categories. To address challenges such as underwater image blurriness and the confusion between finless porpoises and the background, we establish the YOLOv8n-DBTA model. First, we apply automatic gamma correction and dark channel dehazing for image preprocessing. Then, we design a dual-branch feature fusion framework that introduces additional feature fusion paths, preserving more low-level details and features to enhance detection capability. Additionally, we develop a task-aligned detection head to enable interaction between positional and categorical information, allowing the model to learn the relationship between porpoise behavior and location. Finally, we conduct model experiments and analysis. The results demonstrate that our model achieves 97.9% and 79.1% on mAP_{0.50} and mAP_{0.50:0.95}, respectively, outperforming other mainstream models.

Keywords

Finless Porpoise, Object Detection, Deep Learning

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

长江江豚 Yangtze Finless Porpoise (*Neophocaena asiaorientalis asiaorientalis*)是仅分布于中国长江中下游干流及主要的通江湖泊的现存唯一淡水鲸类动物, 种群数量约 1249 头[1]。2013 年世界自然保护联盟将其列为“极度濒危”物种, 2017 年《华盛顿公约》(CITES)将长江江豚列为附录 I 物种。深入了解动物的繁殖生物学特征以及实时准确地判断其繁殖生理状况是濒危动物保护的重要基础。建立长江豚类自然保护区, 同时开展长江江豚饲养下的行为学与繁殖学研究, 是积极有效的保种措施[2]。

传统的江豚行为学研究依靠专家人眼识别, 存在效率低、人工成本高的问题。如今, 利用摄像头来对人工饲养水生生物进行监控观测已经成为水生生物研究的重要手段。通过江豚池的监控摄像可以观察、记录江豚的性行为。中国科学院水生生物研究所白鱀豚馆配备了支持倍速回放的监控平台来缓解传统方法的不足, 但是通过直接观察监控来研究江豚性行为仍有局限性。首先, 需要观察水上和水下两个机位的数个摄像头来对江豚的性行为进行研究, 对同一时段多个机位的江豚进行观察的工作量是庞大的。其次, 观察人员需要持续观看监控, 但是大部分时间江豚并没有性行为, 肉眼监控研究江豚性行为的效率较低, 因此迫切需要引入自动化和智能化的技术手段来实时检测与识别江豚性行为。

随着神经网络及模式识别等自动化技术的发展, 图像检测与识别技术开始被运用于水生生物: 孙月平[3]等提出了一种基于改进 YOLOv8n-seg 的蟹塘水草区域分割与定位方法, 该方法引入基于 RT-DETR 的 HGNetv2 主干网络, 利用其轻量化卷积和层次化的特征提取能力, 实现了模型的轻量化与分割精度的提升, 对于水下检测等需要在资源受限的环境中运行的任务具有重要的实际意义。JI [4]等针对水下河蟹图像模糊、亮度低的特点, 提出基于多尺度金字塔融合图像增强和 MobileCenterNet 模型的河蟹目标检测方法。张铮[5]等提出一种基于图像增强与 GC-YOLOv5 的模型来针对浑浊的水下环境下的河蟹识别。

上述研究结果表明, 图像检测与识别技术对水生生物的自动化识别具有一定作用, 但水生生物的识

别往往各物种间差异较大，模型的普适性不强，且易受到观测环境影响，需要针对各类观测情况进行优化设计。基于上述背景，本文创建了首个人工饲养环境下的江豚性行为数据集(Yangtze Finless Porpoise Sexual Behaviour, YFPSB)，针对水质模糊、江豚与背景混淆等影响江豚性行为检测与识别的问题，构建了江豚性行为检测与识别模型 YOLOv8n-DBTA。

2. 数据集构建

2.1. 数据集的获取与建立

要使得识别任务精确进行，就要选用高质量的数据集对模型进行训练和测试。本文首先建立了江豚性行为数据集。图像数据采集工作于中国科学院水生生物研究所白鱔豚馆进行。中国科学院水生生物研究所白鱔豚馆(湖北武汉)的饲养设施包含 3 个饲养池，分别为肾形主养池、圆形副养池和圆形医疗池。我们选择了来自肾形主养池和圆形副养池两个池子共三个机位的监控录像，共获得 967 张江豚图像，如图 1 所示，(A)图为肾形主养池水上俯视机位，(B)图为肾形主养池水下机位，(C)图为圆形副养池的水下机位。

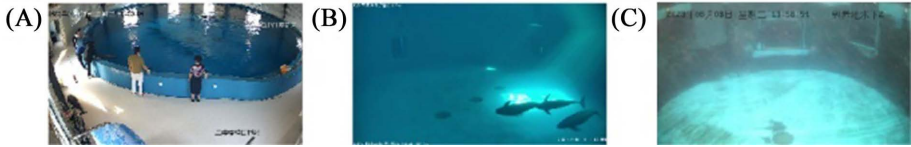


Figure 1. Surveillance footage from three cameras
图 1. 三个机位的监控录像

本文聚焦于江豚性行为的研究，对于水上机位，将江豚行为分别标注为由训练员引导的出水行为(Normal Training)、人工采精行为(Training for Semen Collection)和正常行为。对于水下机位，将江豚行为分别标注为性行为(Sexual Behavior)、自慰行为(OBJ)和正常行为。

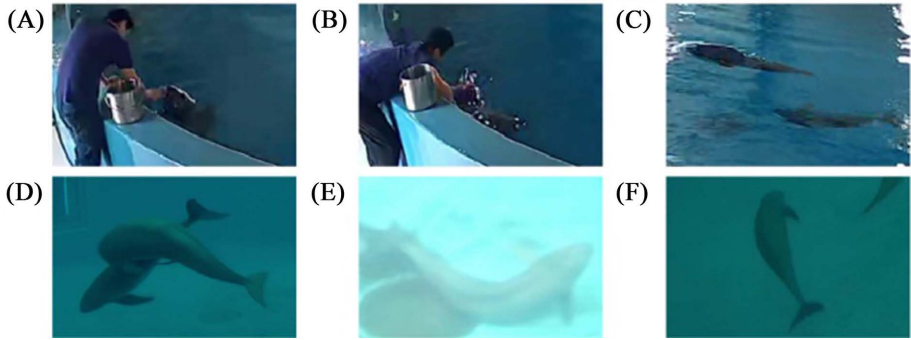


Figure 2. Yangtze finless porpoise behavior
图 2. 江豚行为

本数据集参考张长群等人[6]建立的性行为谱共标注了五种行为。

- 1) 训练员引导的行为(Normal Training): 训练员引导江豚靠近、出水等行为，如图 2(A)所示。
- 2) 人工采精行为(Training for Semen Collection): 训练员对江豚进行人工采精，如图 2(B)所示。
- 3) 性行为(Sexual Behavior): 阴茎勃起后接近或接触其他个体的任何行为，包括生殖器勃起、尝试插入、插入、尾鳍插入、接近、追逐、摩擦等十三种行为。如图 2(D)所示为尝试插入行为。

4) 自慰行为(Object Rubbing): 雄性或雌性长江江豚在池底, 池边凸出部分以及窗户边缘摩擦生殖裂, 如图 2(E)所示。对于雄性江豚, 自慰时常可见生殖器勃起。

5) 正常行为(Normal): 非训练员引导下江豚非性行为的的活动状态。如图 2(C)所示为水上正常活动的江豚, 如图 2(F)所示为水下正常活动的江豚。

2.2. 数据集的增强与扩充

在深度学习训练的过程中, 通过旋转、平移、翻转等数据增强[7]方法, 将图片由 967 张扩充至 2901 张。本文使用 labeling 软件对图片进行标注, 创建了 yolo 标注格式的江豚性行为数据集。江豚性行为 YFPSB 数据集包含 2901 张图片, 数据集按照 7:3 的比例划分为训练集和验证集, 将 2030 张图像用于训练, 871 张图像用于验证。

3. 江豚性行为检测与识别模型

为了配合白鱔豚馆的监控平台实时检测与识别江豚性行为并自动统计次数和时间段, 建立了江豚性行为目标检测与识别模型。由于单阶段的算法比双阶段的算法更适应需要兼顾实时性和准确性的江豚性行为检测与识别需求。本文基于单阶段的 YOLOv8 [8]进行改进。YOLOv8 是一种常用的目标检测与识别算法, 具有检测精度高、推理速度快等优点。它包含 5 种基础结构, 分别是 YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l 和 YOLOv8x, 这些不同的变体使得 YOLOv8 能够在精度和速度中权衡, 拥有更广阔的应用场景。其中, YOLOv8n 是 YOLOv8 网络中最轻量级的网络, 在兼顾精度的同时减少了参数量。

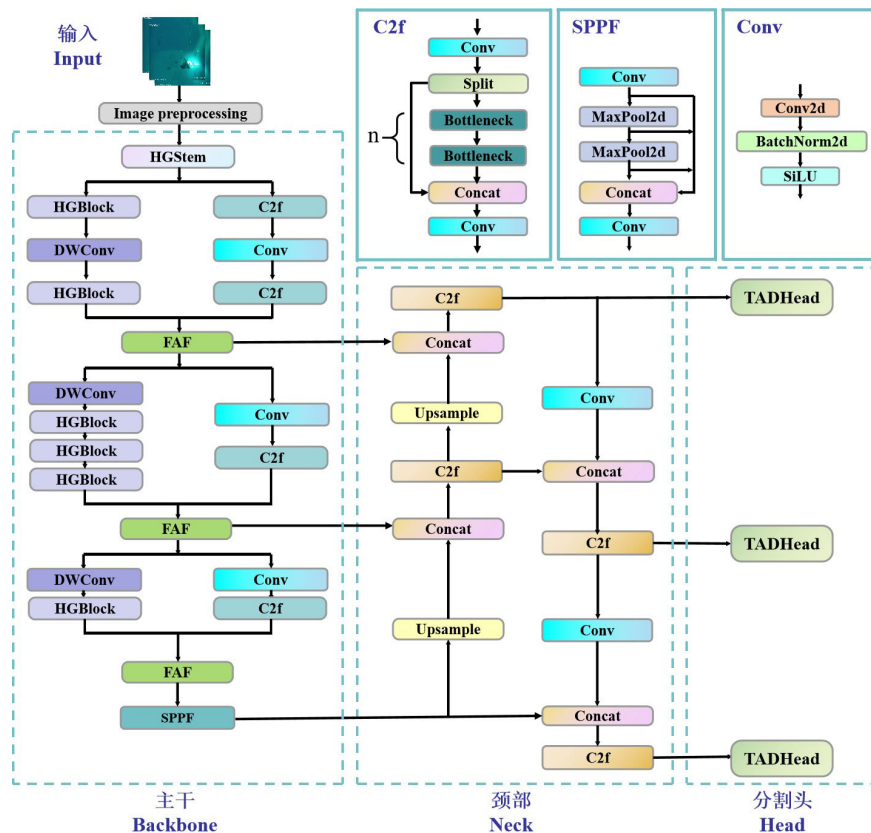


Figure 3. Overall architecture diagram of the model

图 3. 模型结构

基于 YOLOv8n, 本文提出了一种双主干的江豚性行为检测与识别模型 YOLOv8n-DBTA, 如图 3 所示为 YOLOv8n-DBTA 的结构图。模型的优化主要体现在三个方面: 1) 针对水下能见度低、水质模糊导致江豚容易与背景混淆和目前模型定位与分类解耦的问题, 在主干网络前增加了预处理模块, 即通过自适应伽马校正、暗通道先验去雾进行图像增强, 以增强江豚与豚池背景的区分度; 2) 采用双主干特征融合框架(DualBackbone), 使模型在复杂的水下背景中更好地提取出江豚的细节特征, 避免背景干扰; 3) 设计任务对齐检测头(Task Align Detection Head, TADH)生成分类与定位的联合交互特征来提高模型学习江豚性行为类别与位置间关系的能力。本研究对 YOLOv8n-DBTA 进行了消融研究, 验证改进的可行性, 与其他流行的网络结构相比, YOLOv8n-DBTA 对江豚性行为的检测与识别效果优异。

3.1. 图像预处理模块

通过观察数据集, 可以看到采集到的图像受到水下环境及研究所设备的限制, 存在画质模糊、难以明显区分江豚与背景的问题, 给模型的训练造成困难。为此, 我们首先对图像进行预处理, 以增强江豚与背景的区分度。预处理模块由自动伽马校正[9]、暗通道先验去雾[10]两个部分组成。经过图像预处理的前后对比如图 4 所示, 左侧为原始图像, 右侧为处理后的图像, 可以看到处理后缓解了水质模糊的问题, 图片中的江豚与背景更容易区分。



Figure 4. Comparison of images before (left) and after (right) processing
图 4. 图像处理前(左)后(右)对比

3.2. 双主干特征融合框架(DualBackbone)

在水下环境中, 光线、波动以及背景的复杂性都会干扰目标的检测与识别。经典的 YOLOv8n 网络可能无法有效地分离江豚与背景, 存在漏检和误检的问题, 且在处理细粒度行为分类时可能存在误差。水下环境中, 江豚有时成群出现, 原网络可能难以同时检测与识别多个目标, 尤其是当它们距离较远或部分遮挡时。为了解决这一问题, 在 YOLOv8n 的主干网络中引入了 HGNetv2 [11], 提出双主干特征融合框架(DualBackbone)。双分支结构允许网络在不同层次上提取不同特征, 在复杂的水下背景中更好地提取出江豚的细节特征, 避免背景干扰。通过融合不同尺度的信息, 双分支结构可以更好地处理多场景的变化, 尤其是在水下江豚行为检测中, 江豚可能处于不同的尺度和位置。该改进增强了对不同尺寸江豚目标的识别能力。双分支结构通过特征融合, 有助于提高对多个江豚的同时检测与识别的能力, 尤其是在

群体行为或重叠目标的情况下，能够减少漏检和误检。

为了避免在浅层特征图上消耗过多计算资源，本研究设计的 DualBackbone 共享一个 HGStem 的信息，避免计算量过大、推理时间过大的问题。如图 3 所示，图片输入后，先经过网络初始预处理层 HGStem 以在网络的早期阶段快速提取特征，并通过降低特征图的空间分辨率来减少计算量。接着进入两个不同的主干进行特征学习，主干部分的左侧为参考 HGNetv2 结构的主干，右侧为 YOLOv8 的经典主干部分。HGNetv2 网络使用深度可分离卷积 DWConv 替换传统卷积，能够减少计算负载，提高模型的推理速度，并在每个通道上能独立学习特征。这种设计不仅提高了模型的效率，还增强了对复杂环境的适应能力。

由于直接用双主干分别学习特征会导致最后得到的特征差异过大，本研究提出了 FAF 模块来解决来自不同主干的特征差异。从两个分支得到的特征在 FAF (Feature Align Fusion) 特征对齐融合模块进行动态对齐，然后将融合后的特征再给到各自的主干进行学习。如图 5 所示，来自两个不同主干学习到的特征 fea1 和 fea2 先经过一个卷积将通道数变成一样的，然后拼接到一起，通过卷积再加一个 Sigmoid 归一化函数生成动态对齐权重去调整各自的特征，增强模型对不同特征的适应能力，然后经过 split 将特征切分回去，得到的动态对齐特征与之前的特征 fea1、fea2 分别相乘，然后用经过梯度更新、可以根据输入特征动态调整的通道权重 param1 和 param2 (两个参数都初始化为 0.5，参数限制在 -1、1 的范围避免过大过小) 对特征相乘，调整各自的特征后相加。最后再经过一个卷积层输出，如式(1)、式(2)、式(3)和式(4)所示。

$$fea1' = \text{Conv}(fea1) \quad (1)$$

$$fea2' = \text{Conv}(fea2) \quad (2)$$

$$\text{weights1, weights2} = \text{Split}\left(\sigma\left(\text{Conv}\left(\text{Concat}\left(fea1', fea2'\right)\right)\right)\right) \quad (3)$$

$$\text{output} = \text{Conv}\left(\text{param1} \times \text{weights1} \times fea1' + \text{param2} \times \text{weights2} \times fea2'\right) \quad (4)$$

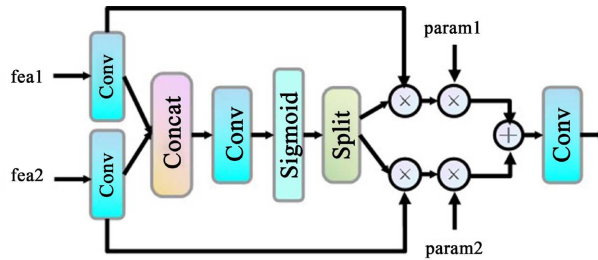


Figure 5. FAF module structure

图 5. FAF 模块结构

3.3. 任务对齐检测头(Task Align Detection Head)

江豚的行为类别与其位置相关，如江豚在自慰时通常出现在井盖等地、靠近水面的时候通常是正常行为。因此，经典 YOLOv8 的解耦结构虽然在分类和定位上表现优秀，但它将分类头和检测头完全分离，在模型的训练和推理过程中，分类头和检测头各自独立地处理输入数据，没有直接的信息交互。这在复杂场景下可能导致性能下降。为了解决这个问题，本研究提出了共享参数的任务对齐检测头(Task Align Detection Head, TADH)。参照 TOOD [12] 的思想引入任务对齐机制，使得定位分支和分类分支之间可以相互学习，使模型更适合江豚性行为检测与识别。经典 YOLOv8 的检测头部分用到了 12 个 3×3 卷积和 6 个 1×1 卷积，参数量为 897,664。约占计算量的 1/5，这是经典 YOLOv8 检测头庞大的原因，为了解决

这个问题。本研究采用共享的方法，使检测头轻量，同时用 Conv_GN [13]来弥补特征提取能力的损失。

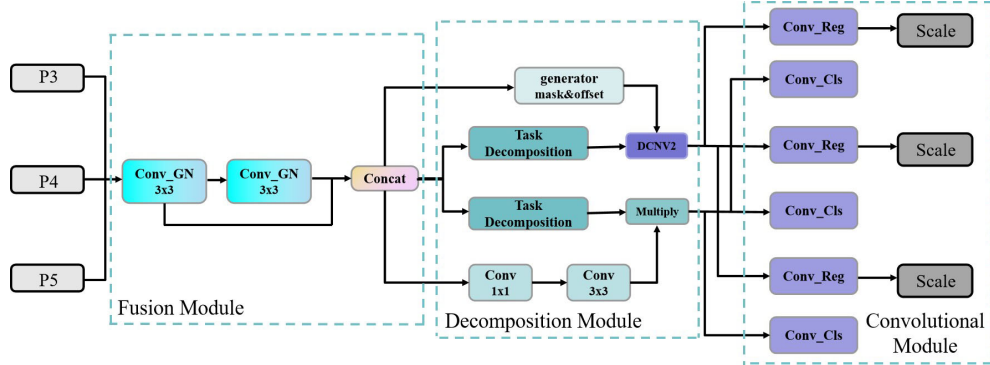


Figure 6. TADH module structure

图 6. TADH 模块结构

如图 6 所示，P3、P4、P5 输入到两个共享的 3×3 卷积层进行处理后进行拼接，这样做的目的是增强定位和分类任务之间的交互性。拼接后的特征图在进入真正的检测头之前，已包含了跨任务的交互信息。为了控制计算量和保证特征提取能力，拼接后的特征图会通过 Conv_GN 进行通道数压缩，最终的输出通道数与输入的相同。通过生成的 mask 和 offset 以及可变形卷积[14] (DCNV2)，定位分支能够处理更加复杂的目标形状和位置，增强了定位能力。分类分支则通过 1×1 卷积和 3×3 卷积进行通道压缩，并利用交互特征生成动态的注意力权重，进一步提升了分类的精确度。交互特征送入 Task Decomposition 模块进行任务拆解，两个任务拆解模块 Task Decomposition 模块间不共享参数。Task Decomposition 模块参考了 TOOD 中 TAP 的 layer attention 结构。由于定位和分类任务的关注点不同，交互特征可能会引入任务间的冲突。

$$X_k^{\text{task}} = w_k \cdot X_k^{\text{inter}}, \forall k \in \{1, 2, \dots, N\} \quad (5)$$

$$w = \sigma \left(fc_2 \left(\delta \left(fc_1 \left(x^{\text{inter}} \right) \right) \right) \right) \quad (6)$$

式 5 中，为特征对每个任务的计算值，是第 k 层的注意力权重，是任务交互的特征。式 6 中， σ 是 sigmoid 函数， fc 为全连接层，是指 relu 函数，对由拼接后得到的使用平均池化后得到。

Task Decomposition 模块动态计算每个任务特定的特征，从而有效地将两个任务的特征进行分解以避免冲突。最后在使用共享卷积的同时，为了应对每个检测头所检测的目标尺度不一致的问题，使用 Scale 层对特征进行缩放。通过上述改进，TADH 不仅减少了参数量，还通过共享参数和任务对齐机制，增强了模型在水下江豚性行为识别任务中的性能，定位和分类的交互性得到了极大的提升。

4. 结果与分析

4.1. 评价指标

本研究选择平均精度[15] (mAP)、模型参数量、浮点计算量(GFLOPs)三个指标来评价模型精度与计算效率。计算公式如式(7)、式(8)、式(9)和式(10)所示。

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

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

$$AP = \int_0^1 P(r) dr \quad (9)$$

$$mAP = \frac{1}{N} \sum_{n=1}^N AP_n$$

4.2. 实验环境与训练参数

本文实验均在 Linux 系统上进行,操作系统版本为 Ubuntu 18.04, CPU 为 Intel(R) Core(TM) i9-10900K, RAM 为 62GB, 显卡为 NVIDIA GeForce RTX 3090。本文的实验配置的环境的 Python 版本为 3.8.8, PyTorch 版本为 2.1.0 + cu121, CUDA 版本为 11.1。在实验中, 训练 epoch 数设置为 400, batch size 设置为 16, 初始学习率设置为 0.01。不使用预训练权重, 网络输入图像的大小默认设置为 640×640 。

4.3. 消融实验

4.3.1. 图像预处理模块的消融实验

在进行图像预处理模块的消融实验中, 我们对比了几种不同配置的 YOLOv8n 模型的性能。实验结果如表 1 所示。其中, “w/o IP” 表示没有使用图像预处理模块, “w/IP” 表示使用图像预处理模块, “DBTA” 表示我们提出的改进模型。从表格中可以看出, 使用图像预处理模块的模型(YOLOv8n (w/IP) 和 YOLOv8n-DBTA (w/IP))在 $mAP_{0.50}$ 和 $mAP_{0.50:0.95}$ 两个指标上都优于没有使用图像预处理模块的模型(YOLOv8n (w/o IP)和 YOLOv8n-DBTA (w/o IP))。这表明图像预处理模块对于提高模型性能是有帮助的。我们的改进模型 YOLOv8n-DBTA 在使用图像预处理模块后, $mAP_{0.50}$ 从 0.973 提高到了 0.979, $mAP_{0.50:0.95}$ 从 0.778 提高到了 0.791。通过这个消融实验, 我们验证了图像预处理模块对于提高 YOLOv8n 模型性能的有效性, 这是因为预处理模块使用了自动伽马校正与暗通道先验去雾技术, 使得图像的亮度得到增强, 背景与目标更加分明, 改善了模型在水下复杂环境下的检测精度与鲁棒性。同时, 我们的改进模型 YOLOv8n-DBTA 在图像预处理模块的帮助下, 表现出了更好的性能。

Table 1. Results of ablation experiments with the IP module
表 1. 预处理模块消融实验结果

模型	$mAP_{0.50}$ (%)	$mAP_{0.50:0.95}$ (%)
YOLOv8n (w/o IP)	96.8	77.4
YOLOv8n (w/ IP)	97.2	77.5
YOLOv8n-DBTA (w/o IP)	97.3	77.8
YOLOv8n-DBTA (w/ IP)	97.9	79.1

4.3.2. 模型改进的消融实验

改进前后的实验结果如表 2 所示。从表中可以看出, 改进前模型的正常训练和人工采精的识别精度相对较低, 分别为 95.1%和 98.2%。这主要是因为这些行为在监控视频中较为常见, 且动作较为复杂, 容易受到背景干扰。正常行为的识别精度最低, 为 94.3%。这主要是因为正常行为的动作较为多样, 且在水下环境中容易与背景混淆。改进后这些类别的精度都有显著提升。通过引入双主干网络缓解背景干扰, 通过将位置信息与类别信息进行交互, 提升正常训练和人工采精等与位置有关联的行为类别的精度

消融实验结果如表 3 所示。从表中可以看出, 采用原始 YOLOv8n (w/o IP)模型分类的平均精度值 $mAP_{0.50}$ 和 $mAP_{0.50:0.95}$ 分别为 96.8 和 77.4, 参数量和浮点计算量分别为 3.01 M 和 1.82×10^9 次。与原始 YOLOv8n 模型相比, 方案 1 用共享参数的耦合头 TADH 替换原来的检测头后, 参数量下降了 25.2%,

Table 2. Experimental results of the model
表 2. 模型的实验结果

试验方案	Normal training	Sexual behavior	Obiect Rubbing	Normal	Training for semen collection
YOLOv8n	95.1	98.4	99.2	94.3	98.2
Ours	96.7	98.7	99.5	95.9	98.5

浮点计算量上升了 5.6%， $mAP_{0.50}$ 和 $mAP_{0.50:0.95}$ 分别提升到 97.6%和 80.3%。方案 2 的双主干模型框架，使参数量和计算量上升，平均进度值相比原始模型略有提升，表明 DualBackbone 改进对模型性能也有一定的正面影响。方案 3 同时采用 DualBackbone 和 TADH，参数量与原模型相比上升了 50.5%，浮点计算量上升了 47.0%，模型的平均精度值 $mAP_{0.50}$ 和 $mAP_{0.50:0.95}$ 分别达到 97.9 和 79.1，与改进前相比分别提升了 1.1%和 1.7%，为所有方案中最优，说明两种改进方法的结合能够最大化提升模型性能。这表明，水下图像处理技术的采用通过自动伽马校正与暗通道先验去雾的方式，改善图像的亮度并缓解了水体浑浊的问题，改进了图像的背景区分度。DualBackbone 和 TADH 两种改进方法均对模型性能有积极影响，且二者的结合可以取得更好的效果。

综上所述，通过消融实验，我们验证了图像预处理模块、DualBackbone 和 TADH 三种改进方法对 YOLOv8n 模型性能的提升作用。其中，YOLOv8n-DBTA (w/IP)同时引入了 DualBackbone 和 TADH，在保证参数量和计算量可控的前提下，取得了最佳的性能。

Table 3. Results of ablation experiments
表 3. 消融实验结果

试验方案	DualBackbone	TADH	参数量(M)	浮点计算量 ($\times 10^9$ 次, 1 s)	$mAP_{0.50}$ (%)	$mAP_{0.50:0.95}$ (%)
YOLOv8n (w/o IP)	×	×	3.01	8.1	96.8	77.4
方案 1 (w/IP)	×	√	2.24	8.6	97.6	80.3
方案 2 (w/IP)	√	×	5.28	11.4	97.5	78.5
方案 3 (w/IP)	√	√	4.52	11.9	97.9	79.1

4.4. 不同模型对江豚性行为的识别性能比较

为了进一步验证所提模型在江豚性行为检测与识别任务中相对于其他主流模型的优越性，对改进模型与当前主流目标检测与识别模型进行了对比实验，实验结果如表 4 所示。YOLOv8n-DBTA (w/IP)模型对江豚性行为的检测与识别的效果高于其他 3 种模型。这是因为本文提出的模型从图像预处理、双主干特征融合框架(DualBackbone)和任务对齐检测头(TADH)这三个方面对江豚的性行为识别做了针对性增强，其中：

- 1) 图像预处理模块通过自动伽马校正和暗通道先验去雾技术，增强了图像的亮度并缓解了水体浑浊的问题，改善了图像的背景区分度，从而提高了模型在复杂水下环境下的检测精度和鲁棒性；
- 2) 双主干特征融合框架(DualBackbone)增强了对不同尺寸和位置的江豚目标的识别能力，尤其在群体行为或重叠目标的情况下，减少了漏检和误检；
- 3) 任务对齐检测头(TADH)通过共享参数和任务对齐机制，将分类任务与定位任务进行交互，增强了模型对江豚性行为类别与位置关系的学习能力，提升了定位和分类的交互性。

本实验证明了这些改进的有效性。

Table 4. Comparison results of recognition performance of different models

表 4. 不同模型识别性能对比结果

模型	参数量(M)	浮点计算量($\times 10^9$ 次, 1 s)	mAP _{0.50} (%)	mAP _{0.50:0.95} (%)
YOLOv5n	2.50	7.1	97.1	76.0
YOLOv10n	2.27	6.5	97.0	73.6
YOLOv8n	3.01	8.1	96.8	77.4
YOLOv8n-DBTA (w/IP)	4.52	11.9	97.9	79.1

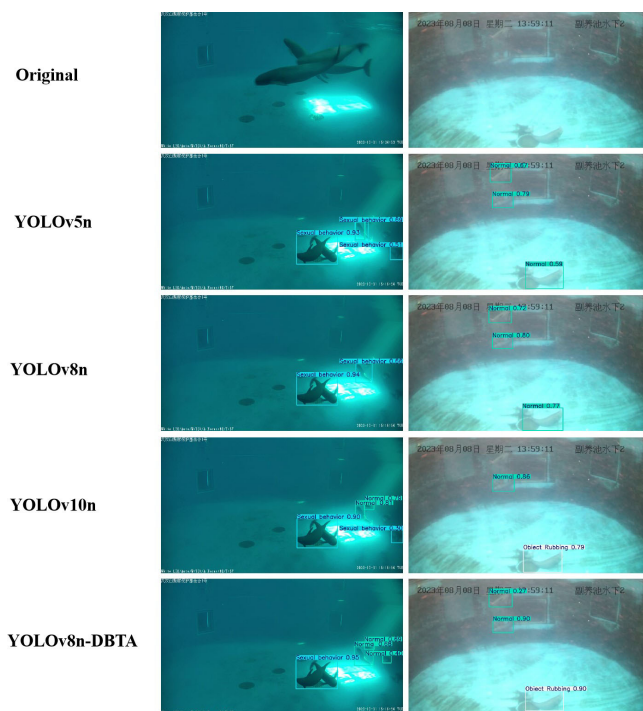


Figure 7. The recognition effect of different models

图 7. 不同模型的识别效果

为了更直观地反映模型在江豚性行为检测与识别任务中相对于其他主流模型的优越性，在江豚数据集中选取两张水质模糊、江豚目标较小或部分被遮挡的图片进行推理。如图 7 所示，YOLOv5n、YOLOv8n 和 YOLOv10n 都出现了不同程度的漏检和误检。其中，YOLOv5n、YOLOv8n 将自慰误检为普通行为，而我们的模型将定位与类别耦合，成功检测出自慰行为。对于与背景容易混淆的目标较小的江豚，我们的模型的效果也均优于其他流行模型。

综上所述，本文提出的改进模型与不同的目标检测与识别模型相比，在检测精度上具有一定的优势，同时并未增加过多的参数量，使得本模型的部署相对便利。

5. 总结

长江江豚是濒危动物，对长江江豚性行为的研究可以促进其人工繁殖的成功以及种群快速增加和恢复。传统的江豚性行为识别靠专家人眼识别，存在效率低、人工成本高的问题，提出一种双主干的江豚

性行为检测与识别模型 YOLOv8n-DBTA。首先增加了图像预处理模块来增加图像的对比度、缓解水体浑浊的问题,接着设计双主干特征融合框架(DualBackbone)以保留更多底层特征图细节和特征来提高模型对江豚的检测能力。最后设计任务对齐检测头(TADH),将位置与类别信息进行交互来更好地学习江豚的行为与位置的关系。我们还与其他主流模型进行了对比试验,证明我们的模型比其他模型更适合江豚性行为检测。本模型为监控江豚性行为提供了一种先进的技术手段,为今后深入开展江豚的繁育保护及繁殖生物学研究提供支撑,为相关领域的科研和水生物保护工作提供有力支持。

参考文献

- [1] 陈炳耀, 信誉, 路方婷, 等. 长江江豚监测现状及展望[J]. 中国环境监测, 2023, 39(2): 1-10.
- [2] 于道平. 长江江豚保护生物学研究进展[J]. 安徽大学学报: 自然科学版, 2003(4): 98-103.
- [3] 孙月平, 刘勇, 郭佩璇, 等. 基于改进 YOLOv8n-seg 的蟹塘水草区域分割与定位方法[J]. 农业工程学报, 2024, 40(17): 224-233.
- [4] Ji, W., Peng, J., Xu, B. and Zhang, T. (2023) Real-Time Detection of Underwater River Crab Based on Multi-Scale Pyramid Fusion Image Enhancement and MobileCenterNet Model. *Computers and Electronics in Agriculture*, **204**, Article 107522. <https://doi.org/10.1016/j.compag.2022.107522>
- [5] 张铮, 鲁祥, 胡庆松. 基于图像增强与 GC-YOLO v5s 的水下环境河蟹识别轻量化模型研究[J]. 农业机械学报, 2024, 55(11): 124-131.
- [6] 张长群, 郑杨, Platto Sara, 等. 人工饲养长江江豚的同性性行为 and 自慰行为[J]. 兽类学报, 2015, 35(3): 241-252.
- [7] 尚志军, 张华, 曾成, 等. 基于机器视觉的枳壳自动定向方法与试验[J]. 中国农机化学报, 2019, 40(7): 119-124.
- [8] Varghese, R. and M., S. (2024) YoLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. 2024 *International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, Chennai, 18-19 April 2024, 1-6. <https://doi.org/10.1109/adics58448.2024.10533619>
- [9] Babakhani, P. and Zarei, P. (2015) Automatic Gamma Correction Based on Average of Brightness. *Advances in Computer Science*, **4**, 156-159.
- [10] He, K., Sun, J. and Tang, X. (2009) Single Image Haze Removal Using Dark Channel Prior. 2009 *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, 20-25 June 2009, 1956-1963. <https://doi.org/10.1109/cvpr.2009.5206515>
- [11] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., et al. (2024) DETRs Beat YOLOs on Real-Time Object Detection. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 16965-16974. <https://doi.org/10.1109/cvpr52733.2024.01605>
- [12] Feng, C., Zhong, Y., Gao, Y., Scott, M.R. and Huang, W. (2021) TOOD: Task-Aligned One-Stage Object Detection. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 3490-3499. <https://doi.org/10.1109/iccv48922.2021.00349>
- [13] Tian, Z., Shen, C., Chen, H. and He, T. (2019) FCOS: Fully Convolutional One-Stage Object Detection. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 9626-9635. <https://doi.org/10.1109/iccv.2019.00972>
- [14] Zhu, X., Hu, H., Lin, S. and Dai, J. (2019) Deformable ConvNets V2: More Deformable, Better Results. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 9300-9308. <https://doi.org/10.1109/cvpr.2019.00953>
- [15] Mohapatra, D., Das, N. and Mohanty, K.K. (2022) Deep Neural Network Based Fruit Identification and Grading System for Precision Agriculture. *Proceedings of the Indian National Science Academy*, **88**, 228-239. <https://doi.org/10.1007/s43538-022-00079-0>