基于双目视觉和改进YOLOv11-pose的河蟹水下原位质量估测方法

李澳强[1,2,3], 戴航宇[1,2,3], 郭 亚[1,2,3*]

(1. 江南大学智能光学感知与应用国际联合研究中心,江苏无锡 214122,中国; 2. 轻工过程先进控制教育部重点实验室,江南大学,江苏无锡 214122,中国; 3. 江南大学 物联网工程学院,江苏无锡 214122,中国)

摘 要: [**目的/意义**] 河蟹质量的精准估测对于河蟹精准化养殖具有重要的意义,但是传统的河蟹质量人工测量效率低且易损伤蟹体,面向河蟹水下原位质量估测方法研究非常缺乏。为此研究开发高效无损的河蟹水下原位质量估测方法。[**方法**] 基于双目视觉建立融合关键点检测与三维测量的质量估测方法,克服现有三维重建技术在水下动态场景中关键点定位精度不足的瓶颈,进而实现河蟹水下原位质量精准估测。首先,改进YOLOv11框架,通过将MBConv模块与EffectiveSE注意力机制融合重构C3K2特征提取模块,并引入空间动态特征融合机制表面细节融合模块,构建适应养殖水体浑浊环境的河蟹目标检测模型;其次,建立关键点检测匹配算法替代传统全局立体匹配,结合三角测量原理实现甲壳三维参数精准测量;最后,构建双层反向传播神经网络,融合甲壳形态参数与性别特征进行质量回归预测。[**结果和讨论**] 改进后的目标检测模型在交并比为0.5的平均精度均值(Mean Average Precision, mAP)达97.2%,关键点检测精度mAP50达96.7%;河蟹甲壳三维测量平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)仅为2.68%;系统整体质量预测MAPE为7.1%。[**结论**] 建立的非接触式测量方法能够实现水下河蟹形态参数的精准获取与质量估测,为河蟹养殖智能化监测提供了关键技术参数,具有重要的理论和应用价值。

关键词: 河蟹;关键点检测;双目视觉;YOLOv11;质量估测;原位检测

中图分类号: TP391.4 **文献标识码:** A **文章编号:** SA202505019

引用格式:李澳强, 戴航宇, 郭亚. 基于双目视觉和改进YOLOv11-pose的河蟹水下原位质量估测方法[J]. 智慧农业(中英文), 2026, 8(1): 192-202. DOI: 10.12133/j.smartag.SA202505019

LI Aoqiang, DAI Hangyu, GUO Ya. Underwater *Insitu* Weight Estimation Method for Chinese Mitten Crab Based on Binocular Vision and Improved YOLOv11-pose[J]. Smart Agriculture, 2026, 8(1): 192-202. DOI: 10.12133/j. smartag. SA202505019 (in Chinese with English abstract)

0 引 言

在河蟹集约化养殖体系中,个体质量的精准估测是精准投喂、疾病防控和收获决策优化等的关键[1]。现行河蟹质量监测技术主要依赖人工定期捕捞结合物理称量法,该方法不仅存在劳动强度大、操作效率低等固有缺陷,更可能诱发河蟹的应激反应,导致附肢自切等生理损伤,进而影响个体生长与养殖经济效益[2]。随着机器视觉与数字图像处理技术的突破性发展,基于机器视觉的非接触式检测技术已逐步取代传统人工检测方式[3],在水下河蟹的目标识别等领域具有重要的价值[4,5],显著推动了水产养殖的智能化进程[6]。

基于机器视觉的河蟹质量估测方法通常遵循以下技术路径:首先通过图像分析获取甲壳形态学参数,继而构建甲壳参数与体质量的非线性回归模型实现质量估测。在甲壳形态参数-质量关系建模领域,唐永成等[7]通过生物学统计分析揭示了河蟹甲壳形态参数与体质量间具有显著相关性。和飞等[8]在此理论基础上建立了雌雄异型的甲壳长宽

收稿日期:2025-05-19

基金项目:国家自然科学基金(51961125102);江苏省现代农业项目(BE2022366)

作者简介:李澳强,硕士研究生,研究方向为机器视觉。E-mail:liaoqiang919@gmail.com

*通信作者:郭 亚,博士,教授,研究方向为系统建模与控制、传感器与仪器。E-mail:guoya68@163.com

copyright©2026 by the authors (登录www.smartag.net.cn免费获取电子版全文)

二项式回归模型，验证了形态参数法估测河蟹质量的理论可行性。在视觉测量技术层面，SUN等[9]采用YOLOv5-seg网络实现河蟹甲壳实例分割，通过模板匹配算法计算甲壳长宽尺寸，获得1.5%的平均绝对百分比误差。CHEN等[10]则设计了基于YOLOv8的关键点定位网络，在蟹壳图像中精准检测解剖参考点，结合背景标定实现尺寸自动测量，进一步验证了关键点检测法的工程适用性。然而需注意的是，上述研究均基于离水实验室静态环境构建，尚未解决水下原位检测的问题。

针对水下生物尺寸的视觉测量，当前主流方案多采用双目立体视觉系统，结合三维重建与图像分割技术实现尺寸检测[11]。目前已在鱼类[12-14]、对虾[15]和海参[16]等生物上取得了良好的测量效果。证明了使用双目视觉方案测量水下生物方案的可行性。但是在实际河蟹高密度养殖环境中存在高浓度藻类、悬浮泥沙及河蟹的特殊水下活动等干扰因素，使水体浑浊度显著升高，导致水下拍摄的图片出现严重的模糊、偏色和对比度降低的问题。

传统双目匹配算法（如Semi-Global Block Matching（SGBM）、Block Matching（BM））的核心原理是通过计算左右视图对应像素的相似性来估计视差，进而重建三维信息。其技术流程主要包括代价计算、代价聚合和视差优化3个阶段。传统双目匹配算法依赖图像中丰富的纹理特征（如边缘、角点）构建唯一的匹配对应关系。河蟹生活的底层沙质池底本就缺乏明显纹理特征，同时浑浊水体中的悬浮物（如泥沙、藻类）散射入射光，导致目标表面纹理模糊，边缘梯度下降[17]。研究表明，浊度每增加1 NTU，图像边缘清晰度降低约12%[18]。在低纹理区域，匹配代价分布平坦，无法通过聚合（如SGBM的路径代价累积）确定最优视差，导致视差估计在目标边缘处出现断裂，影响三维重建完整性[19]。再者，在静态场景中背景稳定，匹配误差可通过全局优化抑制，而在浑浊的水下环境中，大量悬浮颗粒在水中随机运动，在左右视图中引入时变噪声，破坏空间连续性约束[20]。动态噪声在代价体积中导致局部极值，全局优化算法因噪声干扰难以收敛，导致视差估计偏离真实值。

值得注意的是，崔海鹏等[21]在鱼类研究中创新采用关键点匹配替代传统双目匹配策略，通过建立鱼类特征点对应关系将三维测量平均相对误差控制在7%以内，这为本研究提供了重要启示。但此方案高度依赖关键点定位的精度，而已有的关键点检测方法缺乏对于浑浊水域的环境鲁棒性，使得其在目标识别精度和关键点定位准确性方面均面临严峻挑战[22, 23]。

针对上述技术瓶颈，本研究提出系统性解决方案。首先，基于YOLOv11架构整合MBConv模块与EffectiveSE注意力机制，重构C3K2特征提取模块以提升模型表征能力，同时引入空间动态特征融合机制表面细节融合模块（Surface Detail Fusion Module, SDFM），增强模型的关键点检测效果。其次，建立特征点双目匹配算法替代传统三维重建流程。最后，构建双层反向传播（Back Propagation, BP）神经网络模型，整合甲壳长宽、性别等多源特征实现质量预测。研究中建立的整套河蟹原位质量预测方法，弥补了河蟹水下原位无接触测量的空白，减少了人力的消耗和对河蟹正常生长环境的干扰，能够持续地对水下河蟹质量进行监控，以期为河蟹的精细化养殖提供强有力的支撑，助力河蟹养殖向更加智能化阶段发展。

1　研究材料

1.1 研究对象与采集系统

本研究基于河蟹的生态习性与养殖实践，设计了1套可控环境下的数据采集方案。实验在1.5 m×1.5 m×1.0 m的水箱中进行，为模拟实际养殖场景，在箱底放入食台，并在食台上投放饲料，然后抽取河蟹养殖池中的池水并混入伊乐藻等，形成包含泥沙悬浮物、植物残片及饲料颗粒的拟真水下环境。图像采集系统采用MindVision MV-MGE134GC工业级双目相机模组，该设备分辨率像素为1 280×1 024，搭配3.37 mm焦距镜头，相机基线长度为12 cm。相机模组封装于定制化防水外壳，通过支架浸没于水下。相机采集到的图像通过以太网接口实时传输至工作站。

研究选取宜兴河蟹养殖基地的河蟹530只（雄性240只，雌性290只）作为样本，质量分布为2.45～229.30 g，头胸甲长度（L）15.95～74.70 mm、宽度（W）16.5～79.7 mm。

1.2 数据集构建

1.2.1　水下河蟹关键点检测数据集构建

实验中使用感知哈希算法剔除相似度高于90%的重复图像，继而采用Laplacian梯度分析法排除运动模糊样本，最终保留3 200张有效图像构建YOLO-pose数据集。采用Labelme 5.5.0软件进行标注，

建立双层次标注体系。在形态学层面，定义4个关键特征点：头胸甲前缘中央缺刻顶点（如图1中关键点1）、左右第4侧齿顶点（如图1中关键点2、关键点3），以及后缘中点（如图1中关键点4）；在性别分类层面，构建最小外接矩形框标定甲壳和蟹钳区域，并根据原始记录信息进行性别二元分类（Male/Female）。标注数据经Python脚本转换为YOLO格式，并按8∶1∶1比例随机划分训练集（2 560张）、验证集（320张）与测试集（320张）。

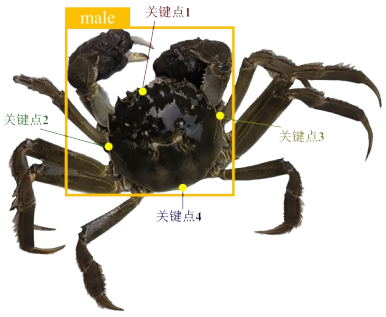


图1 河蟹关键点标注示例

Fig. 1 Illustration of keypoint annotation for Chinese mitten crab

1.2.2 河蟹生物计量复合数据集构建

为建立形态参数与质量之间的量化关系，本研究对530只河蟹进行了系统性生物测量。采用精度为0.01 mm的游标卡尺测量河蟹的背甲长和背甲宽，使用精度为0.01 g的电子天平称量河蟹的质量，并记录其性别。基于上述测量数据，构建了包含性别、形态尺寸和质量的复合数据集，并按照8∶1∶1的比例划分为训练集（424只）、验证集（53只）和测试集（53只）。

2 研究方法

本研究建立了一个基于改进YOLOv11算法和三维重建技术的螃蟹质量估测方法，其技术路线主要包含以下3个环节。1）模型架构优化。通过在YOLOv11网络架构中引入Efficientnet中的MBConv与EffectiveSE模块重构C3K2基础单元，并设计SDFM特征融合机制，显著提升了水下复杂环境中目标检测的鲁棒性及关键点定位精度。2）基于河蟹关键点匹配双目立体视觉系统。将双目相机采集的左右视图分别输入改进后的YOLOv11网络进行推理，获取甲壳关键点坐标与性别分类信息，基于关键点分类和匹配机制建立特征点对应关系，结合相机标定参数通过三角测量原理实现关键点的三维坐标解算，进而构建甲壳形态的关键点三维点云模型并精确测量其空间维度参数。3）构建双层BP神经网络，输入上述测量的甲壳长宽和检测到的性别信息预测质量。水下河蟹质量估测的完整流程如图2所示。

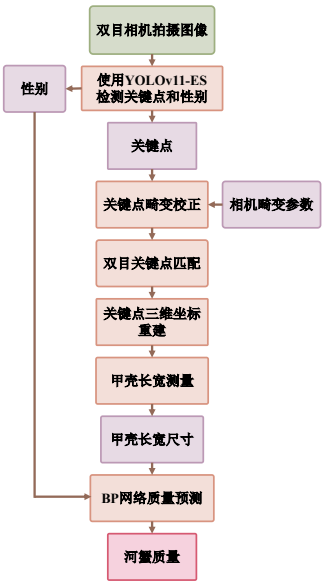


图2 水下河蟹质量估测流程

Fig. 2 Flowchart of underwater Chinese mitten crab weight estimation

2.1 河蟹目标识别及关键点检测算法改进

本研究选用YOLOv11[24]作为基础检测框架，YOLOv11采用C3k2模块优化特征提取过程，通过2×2小卷积核的密集部署，在降低22%参数量的同时增强局部特征捕获能力[25]，在浑浊水下环境中，河蟹轮廓的边缘细节往往因悬浮颗粒和光散射而变得模糊，YOLOv11原生的C3K2结构可以在基础特征提取阶段对低对比度目标有较好的响应能力。然而，该模型在水下河蟹检测场景中仍存在显著局限：水下色偏削弱了颜色特征的有效性；悬浮颗粒物引发的泊松噪声使检测置信度大幅波动[26]，误检率升高；同时，现有注意力机制对水草摆动引起的时序性遮挡缺乏动态建模能力[27]。针对这些挑战，本研究提出两个改进方向：一是使用MBConv和EffectiveSE注意力机制改进原有的C3K2模块，构建C3K2-EMBC模块；二是添加SDFM特征融合机制，增强网络在强干扰下的特征提取能力。

MBConv深度可分离卷积与EffectiveSE通道注意力机制对C3K2进行了重构，使得模型在减少噪声干扰、抑制水体折射光斑的同时，还能强化与关键点位置相关的特征通道响应。同时YOLOv11的路径聚合网络（Path Aggregation Network, PAN）

+C3K2多尺度特征融合框架与SDFM特征融合机制契合：PAN结构负责跨层信息传递，而SDFM又在浅层细节与深层语义之间建立了自适应加权，二者相辅相成，在水下弱纹理、强噪声的情况下能更精准地聚焦甲壳边缘与关键点区域。

2.1.1　C3K2_EMBC模块改进

在真实河蟹养殖场景中，水下图像目标识别和关键点检测面临多重挑战。首先，河蟹养殖水体浊度高，叠加光衰减效应引发图像模糊和色彩失真；其次，传统卷积算子对高频噪声敏感，常规特征提取网络难以有效捕获甲壳缺刻等关键特征。针对上述问题，本研究引入EfficientNet的MBConv模块[28]，其深度可分离卷积结构通过解耦空间与通道维度特征学习，其相较于标准卷积减少了网络参数量，在具有高噪声的真实图像中能有效减轻过拟合风险，增强网络对于输入噪声的鲁棒性。结合扩展率为4的逐点卷积层构建“扩展-压缩”特征流，通过通道维度的提升再压缩显著提升网络对低对比度目标的表征能力。针对水面波纹折射和水草遮挡引发的不均匀光斑问题，传统注意力机制存在定位偏差缺陷[29]。本研究融合EffectiveSE注意力机制[30]，该机制通过自适应平均池化建立通道注意力权重，使网络能动态调节特征图通道响应强度。

具体改进中，本研究对C3k2模块进行系统性重构，结构如图3所示。首先，基于SANDLER等[31]在MobileNetV2中提出的倒残差思想，改进后的C3k2_EMBC模块采用级联式EMBC单元替代原标准卷积层。首先，通过扩展连接构建多尺度特征金字塔，利用4倍通道扩展率捕获甲壳形态的细微差异；其次，集成EffectiveSE通道注意力机制，在特征图生成过程中动态抑制水体折射形成的光斑等干扰噪声；最后，引入概率为0.1的Dropout层，通过随机失活增强模型对遮挡的鲁棒性。

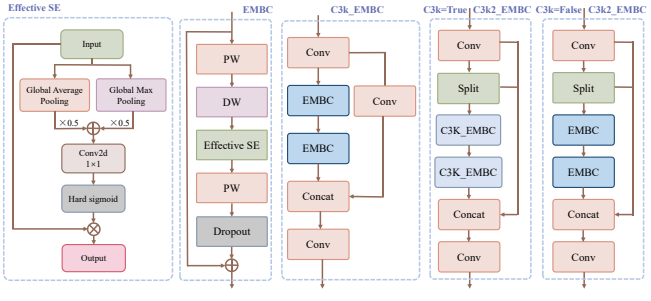


图3　C3K2_EMBC网络结构图

Fig. 3　C3K2_EMBC network architecture diagram

2.1.2　SDFM融合机制

水下图像受光线散射、吸收及悬浮颗粒影响，常呈现低对比度、颜色偏差和细节模糊的问题。SDFM[32]通过双分支特征的自适应校准与注意力加权机制，可有效增强甲壳纹理、边缘等微观特征的显著表达[33]。具体而言，其Recalibrate子模块利用通道注意力对输入特征进行全局重要性重标定，抑制噪声通道并强化与关键点相关的特征响应。YOLOv11的Neck采用PAN结构聚合多尺度特征，但传统C3K2模块侧重于深层语义特征的提取，忽略了浅层细节的空间关联。SDFM引入局部空间注意力与全局通道注意力的协同机制，通过动态权重分配实现跨层特征的精细化融合。例如，在甲壳边缘检测中，局部注意力聚焦于形态学梯度变化区域，而全局注意力则维持整体结构的连续性，二者的乘积权重可平衡细节保留与语义一致性。水下环境中螃蟹常与水草、砂石等背景混杂，传统检测器易受相似纹理干扰。SDFM的特征校准阶段通过残差连接保留原始特征分布，避免梯度消失问题；后续的加权融合则通过软选择机制整合多源特征的优势。

具体的网络结构中，当两个特征图输入SDFM模块时，首先沿通道维度拼接，并通过包含自适应平均池化和卷积的Recalibrate模块生成通道权重，对拼接特征进行自校正并添加残差连接；随后将校正后的特征拆分为两个新的特征图F_a和F_b，再通过通道聚合压缩为单通道特征，分别利用局部空间注意力和全局通道注意力计算权重，经Sigmoid激活后形成融合权重，最终以加权求和方式融合F_a和F_b，其中权重直接作用于F_a，F_b则通过互补权重（1−w）实现自适应特征融合。将SDFM融合机制添加到YOLOv11原模型中完整的YOLOv11-ES网络结构如图4。

2.2　背甲尺寸测量和质量预测

2.2.1　相机的标定和图像校正

针对实际双目视觉系统存在的非线性畸变效应及空间位姿偏差问题，在研究中采用基于“张正友标定法”的系统参数优化方案。使用规格为12×9方格（间距1 cm）的棋盘格作为标定靶，通过水下相机在距相机30～60 cm的水下采集50组立体图像对。利用MATLAB R2023b的Stereo Camera Calibrator工具解算系统参数，包括内参矩阵、外参矩阵和畸变系数。通过OpenCV的initUndistortRecti-

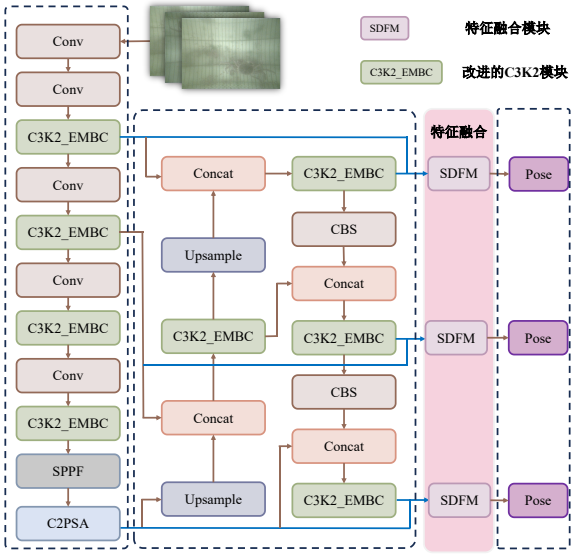


图 4 YOLOv11-ES 网络结构图

Fig. 4 Network architecture diagram of YOLOv11-ES

fyMap 函数生成像素级校正映射表，建立双目图像几何变换模型，有效消除光学畸变并实现立体像对行对准，为后续立体匹配建立精确的立体视觉模型，进一步使用校正映射表将检测到的关键点坐标进行变换。

2.2.2 双目图像特征点匹配

特征点的匹配是根据双目图像中识别到的同一只河蟹的解剖关键点位置，但由于实际拍摄的图像中同一张图中可能有多只螃蟹，同时由于两个相机的视场不同，因而在特征点匹配时需要采取一定的策略，避免图像的两个特征点误匹配。具体的匹配策略如下：

1）检测双目视觉的两侧图像是否都至少含有 1 个存在 4 个特征点的矩形框，且性别相同。如果成立就进入下一步，否则丢弃图片。

2）读取每个矩形框的纵坐标，并向外扩充 5 个像素，计算每个纵坐标的交集，若交集范围大于 90%，那么将两个矩形框归为 1 个集合。

3）计算同一个集合中检测到两只螃蟹关键点 1 和关键点 4 所构成的向量夹角是否小于 5°，且性别类型相同就判定为同一只螃蟹，进一步检测两个矩形框中是否同时含有 4 个关键点，若都含有 4 个关键点，即可完成关键点匹配。

2.2.3 河蟹甲壳测量

在研究中将关键点 1 和关键点 4 之间的实际距离定义为河蟹的甲壳长，关键点 2 和关键点 3 之间的实际距离定义为河蟹的甲壳宽。求解两关键点在空间中距离即可求出河蟹的甲壳长宽。双目测距的示意图如图 5 所示，关键点的空间距离确定主要包含以下步骤。

根据两幅图像中的视差求解深度信息，如公式（1）所示。

$$Z = \frac{f \times B}{d} \tag{1}$$

式中：f为焦距，mm；B为基线长度，mm；d为视差，mm。

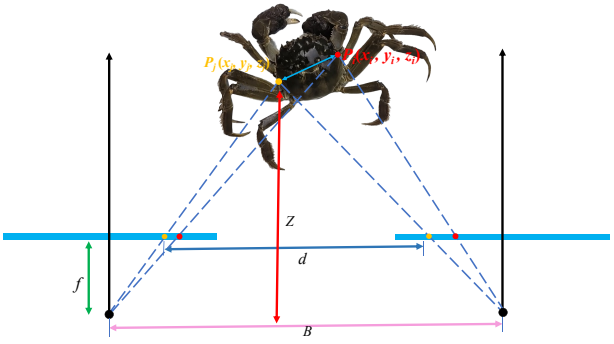


图 5 双目相机测量河蟹甲壳尺寸示意图

Fig. 5 Stereo camera-based measurement of Chinese mitten crab carapace dimensions

三维坐标计算：对于图像上一点(u, v)可得空间坐标$P_r(X, Y, Z)$，表示为公式（2）。

$$\begin{cases} X = \dfrac{(u - c_x) \times Z}{f_x} \\ Y = \dfrac{(v - c_y) \times Z}{f_y} \\ Z = Z \end{cases} \tag{2}$$

式中：(c_x, c_y)为相机内参中的主点坐标；f_x和f_y分别为相机在像素坐标系沿水平方向和垂直方向上的等效焦距。

两个关键点$P_j(x_j, y_j, z_j)$和$P_i(x_i, y_i, z_i)$在空间中的距离如公式（3）所示。

$$D = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \tag{3}$$

式中：D为空间中两点的实际距离，mm。

2.2.4 质量预测

研究中搭建了包含 192 个神经元的两层 BP 神经网络来预测河蟹的质量，将河蟹的甲壳长宽和性别作为网络的输入，各层间使用 ReLU 激活函数，采用 Huber 损失函数平衡平均绝对误差（Mean Absolute Error, MAE）与均方误差（Mean Squared Error, MSE）的优势。为增强模型泛化能力，在网络中嵌入 Dropout 正则化，有效防止神经元协同适应。模型训练采用河蟹生物计量复合数据集中的数据。

2.3 参数设置和评价指标

本研究的实验在硬件配置上采用 Intel（R）Xeon（R）CPU E5-1620处理器、NVIDIA GeForce RTX 2080 Ti 显卡及64 GB内存，以确保高效的计算与数据处理能力；软件环境基于Windows 10操作系统，编程语言为Python 3.9.20，深度学习框架采用 PyTorch 1.12.1，并通过CUDA 12.6实现硬件加速。YOLO模型训练过程中，最大迭代次数（Epoch_max）设置为200，batch size设置为32，初始学习率设为 0.001。而在BP网络的训练过程中，Epoch_max和Batch size分别设为250和16，初始学习率则为0.000 1。

为全面评估系统性能，实验采用多维评价指标：目标检测任务通过精确率（P）、召回率（R）、交并比为0.5时的平均精度均值（Mean Average Precision 50, mAP50）及F_1分数（精确率与召回率的调和均值）衡量检测精度；关键点检测任务沿用相同指标，但聚焦于关键点定位精度。三维测距任务通过平均绝对误差和平均绝对百分比误差（Mean Absolute Percentage Error, MAPE）量化尺寸测量偏差，其中，MAE计算测量值与真实值的绝对误差均值；MAPE表征相对误差的百分比均值。

3 结果与分析

3.1 河蟹目标识别及关键点检测效果分析

3.1.1 消融实验

为了验证两种改进策略的有效性，在保证参数和运行环境一致的情况下，对比分析了不同改进方案的性能表现，实验结果如表1所示。

表1 改进YOLOv11模型河蟹关键点检测消融实验结果对比

Table 1 Comparison of ablation experiment results for keypoint detection of Chinese mitten crab using the improved YOLOv11 model

序号	EMBC	SDFM	GFLOPS	BOX				Pose			
				P/%	R/%	mAP50/%	F_1/%	P/%	R/%	mAP50/%	F_1/%
1	×	×	6.6	85.3	84.7	92.8	85.0	85.8	85.1	93.1	85.4
2	×	√	9.7	90.2	91.2	95.7	90.7	89.9	90.8	95.8	90.3
3	√	×	6.4	89.8	93.2	95.1	91.5	89.1	92.3	94.4	90.7
4	√	√	9.5	91.7	94.7	97.2	93.2	91.3	94.4	96.7	92.8

注：√表示使用该模块；×表示未使用该模块。

消融实验结果验证了本研究改进方案的有效性。通过引入SDFM跨尺度特征融合机制，目标检测任务的性能指标得到显著提升：P、R、mAP50和F_1分数分别提高4.9、6.5、2.9和5.7个百分点。同步地，关键点检测任务对应指标也分别实现4.1、5.7、2.7和4.9个百分点的提升，证实了跨尺度特征融合在捕获复杂视觉模式方面的有效性，但伴随3.1 GFLOPs的计算量增加。

在模型轻量化改进方面，将C3K2模块重构为嵌入EffectiveSE的MBConv结构后，关键点检测的R和F_1分数分别提升7.2和5.3个百分点，同时模型计算量降低0.2 GFLOPs，这表明深度可分离卷积与通道注意力机制的协同设计能够有效平衡特征表达能力和计算效率。

值得注意的是，当两种改进策略联合应用时产生显著协同效应：目标检测任务中P、R、mAP50和F_1分数分别提升6.4、10.0、4.4和8.2个百分点；关键点检测指标同步提升5.5、9.3、3.6和7.4个百分点。实验数据表明，模块组合效果优于任意单一改进方案，最终为河蟹性别分类与甲壳关键点定位任务提供了精度与效率兼顾的解决方案。

3.1.2 对比实验

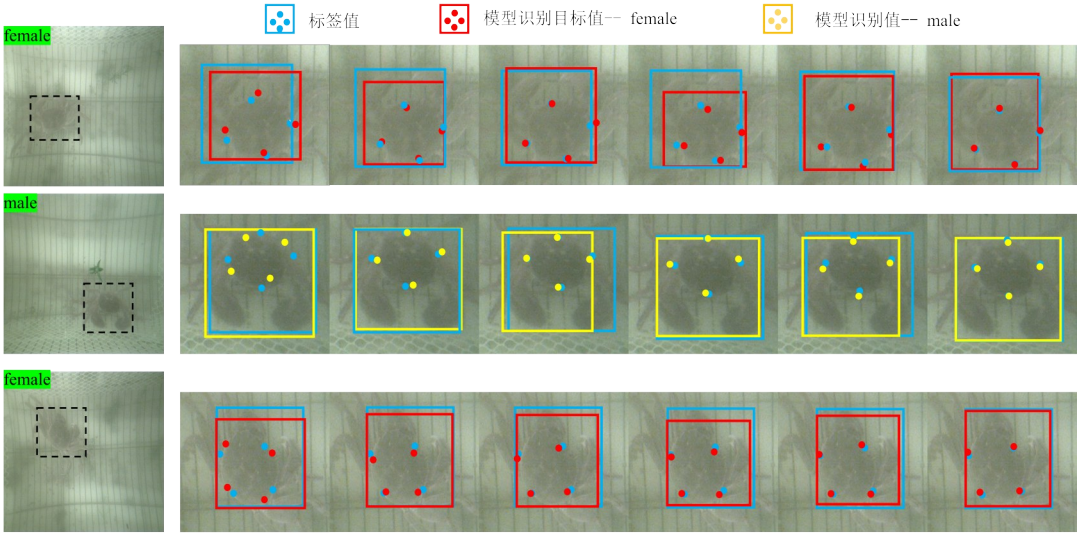
为了验证本研究中改进模型在螃蟹性别分类和关键点检测任务中的优越性，研究中还与YOLOv5、YOLOv8n、YOLOv10n、YOLOv11n和YOLOv12n模型在相同数据集和参数的条件下做了对比实验，结果如表2所示，检测结果的可视化对比如图6所示。

在河蟹目标识别和关键点检测任务中，YOLOv11-ES在多项关键指标上显著优于其他主流轻量级模型。具体而言，其GFLOPS为9.5，虽略高于YOLOv8n（8.3）和YOLOv10n（8.0），但在检测性能上实现了全面突破：目标检测任务的P达到91.7%，R提升至94.7%，mAP50和F_1分数分别达到97.2%与93.2%，较原版YOLOv11n（P=85.3%，R=84.7%，mAP50=92.8%）提升显著。此外，YOLOv11-ES在关键点检测中同样表现优异（mAP50=96.7%），较YOLOv5（92.3%）和YOLOv8n（93.7%）分别提升4.4和3.0个百分点，证明了改进策略在多任务场景下的泛化能力。实验表明，通过

表2 不同模型在河蟹关键点检测实验中的结果对比

Table 2 Comparison results from different models in Chinese mitten crab keypoint detection experiments

模型	GFLOPS	BOX				Pose			
		P/%	R/%	mAP50/%	F_1/%	P/%	R/%	mAP50/%	F_1/%
YOLOv5	7.3	83.2	87.5	92.4	85.3	88.3	82.2	92.3	85.1
YOLOv8n	8.3	89.3	85.4	93.7	87.3	89.3	85.4	93.7	87.3
YOLOv10n	8.0	81.4	83.4	86.0	82.4	81.4	83.4	86.2	82.4
YOLOv11n	6.6	85.3	84.7	92.8	85.0	85.8	85.1	93.1	85.4
YOLOv12n	6.6	79.4	86.0	88.7	82.6	80.7	86.5	89.1	83.5
YOLOv8-ES	9.7	90.6	87.7	95.4	89.1	90.3	87.3	94.9	88.8
YOLOv11-ES	9.5	91.7	94.7	97.2	93.2	91.3	94.4	96.7	92.8



a. 原图 b. YOLOv5 c. YOLOv8n d. YOLOv10n e. YOLOv11n f. YOLOv12n g. YOLOv11ES

图6 不同模型在河蟹生物计量复合数据集上的检测结果可视化对比

Fig. 6 Visualization comparison of detection results from different models on the Chinese mitten crab biometric composite dataset

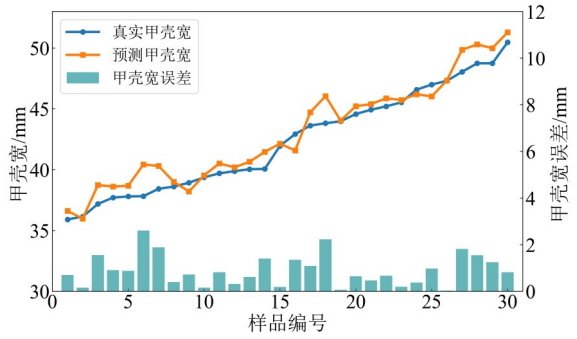
结构优化与计算资源合理分配，YOLOv11-ES在可控复杂度下显著提升了综合性能。为了验证YOLOv11作为基底对改进策略的兼容优势，实验中将相同的MBConv+EffectiveSE与SDFM改进策略迁移到效果最好的YOLOv8骨干上，观察其在河蟹水下关键点检测与目标识别任务中的性能提升情况，通过对比YOLOv8-ES与YOLOv11-ES在相同数据集和评价指标下的表现，实验结果显示YOLOv11-ES在各项指标中均优于YOLOv8-ES，同时改进后的YOLOv8-ES在各项指标中又显著优于YOLOv8n，这充分证明了论文中提出的模块化改进方案的结构通用性和场景适应性。

3.2 河蟹甲壳测量精度分析

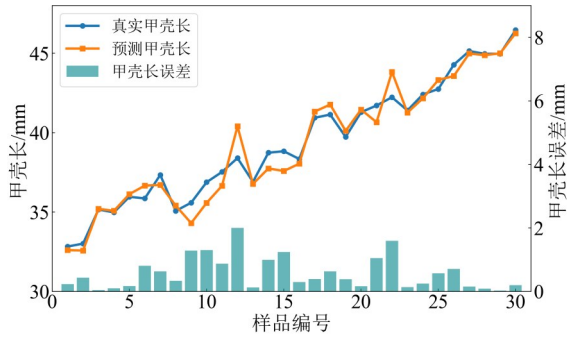
为验证本研究算法对河蟹甲壳尺寸的实际测量精度和鲁棒性，实验中额外测量30只河蟹，在1.1节中的环境中对每只河蟹采集不少于20组图像（涵盖河蟹在视野中的不同位置、多种姿态，以及强光和暗光的条件），并对应手动测量河蟹的甲壳长宽和质量，同时记录河蟹的性别，以此作为真实值。将每只河蟹的20组图像通过YOLOv11-ES网络同步预测关键点坐标及性别特征。然后依据2.2节建立的几何测量模型，对各图像进行甲壳长宽参数解析，并通过计算20组测量值的算术平均值消除由体位姿态、空间位置等偶然因素引起的测量偏差。采用配对样本对实测值与人工标注真值进行对比，预测的甲壳长度和宽度与真实值之间的差异如图7所示。

实验数据显示，甲壳宽度测量的MAE为0.893 mm，MAPE为2.68%；长度测量MAE为0.574 mm，MAPE为1.48%。回归分析表明，测量值与真实值呈显著线性相关（宽相关系数R = 0.977，长相关系数R = 0.981）。该结果验证了本研究算法在水体环境下的测量稳定性和工程适用性。

在尺寸测量与三维重建过程中，水体浑浊度和



a. 甲壳宽预测值和真实值对比



b. 甲壳长预测值和真实值对比

图7　预测河蟹甲壳长宽和真实甲壳长宽的对比

Fig. 7　Comparison between predicted and actual carapace length and width of Chinese mitten crab

光照强度等环境变量会直接影响相机成像质量，继而对测量精度产生系统性干扰。图像质量能够综合反映上述环境因素的变化，因此在实验中构建了图像质量指标与尺寸测量误差之间的定量关联分析。具体地，对30只河蟹分别拍摄了共计750张水下图像（每张图像中河蟹的数量范围为1～5只），并采用通用水下图像质量评价指标（Underwater Image Quality Measure, UIQM）对每幅图像的颜色失真、清晰度和对比度进行打分。所选图像的UIQM值分布在［1.0，2.5］，将该区间等距划分为15个子区间（步长0.1），每个区间对应50张图像。

对于每个区间，计算每个河蟹样本甲壳长（L_i）与宽（W_i）之和的测量值与人工标注真实值之间的相对误差，如公式（4）所示。

$$E_i = \frac{|(L_i + W_i)_{测量} - (L_i + W_i)_{真值}|}{(L_i + W_i)_{真值}} \times 100\% \quad (4)$$

式中：$(L_i + W_i)_{测量}$为每个河蟹样本使用双目测距测得的甲壳长宽之和；$(L_i + W_i)_{真值}$为人工手动测量的每个河蟹样本的甲壳长宽之和。

进一步计算50张图像中测得的所有河蟹样本误差的平均值，如公式（5）所示。

$$E = \frac{1}{M}\sum_{i=1}^{M} E_i \quad (5)$$

式中：M为每个区间中河蟹样本的总数。

实验结果如图8所示，曲线清晰地揭示了平均相对误差随UIQM指标变化的趋势，为评估算法在不同成像条件下的环境鲁棒性提供了依据。

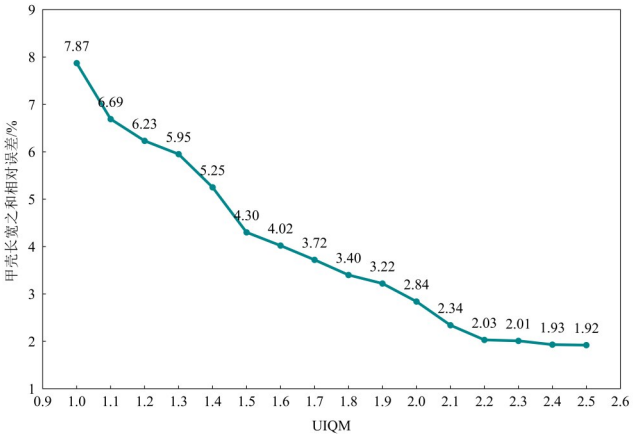


图8　河蟹甲壳长宽之和相对误差随图像质量变化折线图

Fig. 8　Line chart of the relative error in the sum of carapace length and width of Chinese mitten crabs versus image quality variation

如图8所示，甲壳长宽之和的平均相对误差随着UIQM值的提升呈显著下降趋势。在低质量区间（UIQM≈1.0～1.4），图像因色彩失真和对比度低而导致关键点检测困难，平均测量误差最高可达7.87%。当UIQM进入中等范围（≈1.5～2.2）时，误差迅速下降至5%以内，表明即使是适度的清晰度和色彩保真度提高也能一定程度地增强测量精度。而当UIQM超过2.2后，曲线趋于平缓，误差进一步下降幅度收窄，最终在1.9%左右趋于稳定。这一现象说明，在确保关键特征可见性的前提下，图像质量的额外提升对几何测量精度的边际贡献有限，但在低质量条件下，环境因素仍是影响测量鲁棒性的主要瓶颈。值得注意的是，在图像中存在2只以上河蟹的情况中，关键点匹配准确率达99.3%，只有个别出现严重重叠的样本出现了误匹配问题，且由于误匹配后计算的三维距离远超过正常范围，因而可以通过设置正常的河蟹甲壳尺寸范围来剔除问题样本。

3.3　系统整体效果验证分析

为评估水下河蟹质量估测系统的工程实用性，本研究将3.2节中使用双目系统测量的甲壳尺寸数据与YOLOv11-ES网络预测出的性别信息组合后输入到训练好的BP神经网络中，输出系统的最终质量预测数据并与人工测量数据进行对比，实验结果

如图9所示。

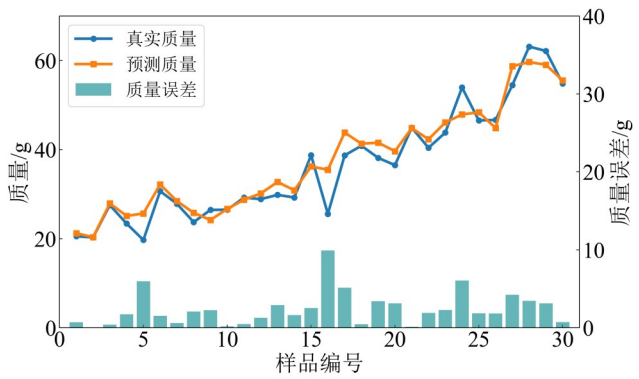


图9 系统整体预测河蟹质量和真实质量的对比

Fig. 9 Overall comparison between the system-predicted and actual weight of Chinese mitten crabs

图9的质量估测结果可视化表明，在独立样本的测试集上，系统呈现稳定的质量预测能力。质量估测的绝对误差MAE为2.39 g，MAPE为7.1%。上述数据充分说明在水下环境中本研究建立的基于关键点检测的河蟹质量估测方法具有良好的效果。

4 讨论与结论

传统河蟹质量测量依赖人工捕捞与物理称重，存在效率低、易损伤蟹体的问题，而现有基于机器视觉的方法多局限于实验室静态环境或清澈水体，难以应对实际养殖场景中的复杂干扰。本研究建立的方法通过融合关键点检测、双目视觉与神经网络建模技术，实现了非接触式、高精度的水下原位质量估测，在浑浊水体环境下展现出显著的优势，为河蟹养殖的智能化监测提供了新的技术路径。

针对水下动态场景中目标检测与关键点定位的挑战，本研究对YOLOv11框架进行了系统性优化，通过轻量化重构C3K2_EMBC模块与SDFM实现显著性能提升。其中，MBConv模块采用深度可分离卷积解耦空间与通道维度的特征学习，在降低参数量的同时有效降低噪声干扰；EffectiveSE注意力机制则通过自适应通道权重分配动态增强关键点相关特征响应，显著提升几何精度。SDFM机制通过双分支特征自适应校准与跨尺度权重融合，弥补传统PAN结构在浅层细节与深层语义整合的不足：局部空间注意力聚焦甲壳边缘梯度变化区域，使水下场景中误检率大幅降低；全局通道注意力则通过动态调整通道权重可以很好地维持甲壳结构连续性。MBConv与SDFM的协同应用进一步构建“局部精细特征提取-全局语义一致性保持”的优化闭环，联合作用下目标检测与关键点定位的mAP50分别提升4.4与3.6个百分点，验证了噪声抑制能力与多尺度特征互补性对复杂水下场景的关键支撑作用。

关键点匹配策略与双目视觉的结合是本方法的核心要素。传统三维重建技术依赖全局图像特征匹配，在浑浊水体中易因特征缺失或误匹配导致测量误差增大。本研究通过关键点的稀疏匹配与性别分类约束，大幅降低计算复杂度的同时，将甲壳三维测量的平均相对误差控制在2.68%。尽管该方法在动态场景下对快速移动或姿态多变的河蟹仍存在偶然误差，但其在低可见度环境中的稳定性显著优于依赖高分辨率点云或完整轮廓提取的现有方案。质量预测模型通过双层BP神经网络整合甲壳形态参数与性别特征，实现了7.1%的平均相对误差。

综上所述，本研究提出了一种基于关键点检测与双目视觉的水下河蟹质量估测方法，通过改进YOLOv11目标检测框架、设计关键点双目匹配算法及构建BP神经网络预测模型，实现了高效、非侵入式的河蟹质量水下估测。实验表明，系统在关键点检测（mAP50 = 96.7%）、甲壳三维测量及质量预测中均达到养殖场景实用化要求。相较于传统方法，本研究的方案显著降低人工干预成本，减少蟹体应激损伤，为水产养殖智能化提供了可靠技术支撑。由于河蟹养殖实际环境及生长过程的复杂性，本研究中的养殖气象条件、水体环境条件、河蟹的品系、测试样本数量有限，难以覆盖所有实际情况。在未来的研究中，可以不断扩充各种场景下的数据，以及通过迁移学习等技术手段，不断提高模型的精度和泛化性能。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献：

[1] 段延娥, 李道亮, 李振波, 等. 基于计算机视觉的水产动物视觉特征测量研究综述[J]. 农业工程学报, 2015, 31(15): 1-11.
DUAN Y E, LI D L, LI Z B, et al. Review on visual characteristic measurement research of aquatic animals based on computer vision[J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(15): 1-11.

[2] LI D L, LIU C, SONG Z Y, et al. Automatic monitoring of relevant behaviors for crustacean production in aquaculture: A review [J]. Animals, 2021, 11(9): 2709.

[3] ZHAO Y X, QIN H X, XU L, et al. A review of deep learning-based stereo vision techniques for phenotype feature and behavioral analysis of fish in aquaculture[J]. Artificial Intelligence Review, 2024, 58(1): 7.

[4] 张铮, 鲁祥, 胡庆松. 基于图像增强与GC-YOLOv5s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, 374.
ZHANG Z, LU X, HU Q S. Lightweight model for river crab detection based on image enhancement and improved YOLOv5s[J].

Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(11): 124-131, 374.
[5] JI W, PENG J Q, XU B, et al. Real-time detection of underwater river crab based on multi-scale pyramid fusion image enhancement and MobileCenterNet model[J]. Computers and Electronics in Agriculture, 2023, 204: 107522.
[6] LIU C H, WANG Z Y, LI Y C, et al. Research progress of computer vision technology in abnormal fish detection[J]. Aquacultural Engineering, 2023, 103: 102350.
[7] 唐永成, 彭姣, 赵运林, 等. 池养中华绒螯蟹不同性别形态及质量差异分析[J]. 渔业科学进展, 2019, 40(6): 114-120.
TANG Y C, PENG J, ZHAO Y L, et al. Morphological attributes and quality parameters of different sexes of *Eriocheir sinensis* cultured in a pond[J]. Progress in Fishery Sciences, 2019, 40(6): 114-120.
[8] 和飞, 王志忠, 卢红, 等. 黄河口中华绒螯蟹成蟹形态性状与体质量的相关性及通径分析[J]. 水产学杂志, 2024, 37(3): 31-36.
HE F, WANG Z Z, LU H, et al. Correlation and path analysis on morphometric traits and body weight for adult Chinese mitten handed crab (*Eriocheir sinensis*) from Yellow River Delta[J]. Chinese Journal of Fisheries, 2024, 37(3): 31-36.
[9] SUN D W, LI J T, LI Z, et al. Grading related feature extraction of Chinese mitten crab based on machine vision[J]. BIO Web of Conferences, 2024, 142: 02016.
[10] CHEN K, CHEN Z Q, WANG C B, et al. Improved YOLOv8-based method for the carapace keypoint detection and size measurement of Chinese mitten crabs[J]. Animals, 2025, 15(7): 941.
[11] HUO G, WU Z, LI J, et al. Underwater target detection and 3D reconstruction system based on binocular vision[J]. Sensors, 2018, 18(10): 3570.
[12] KONG M R, LI B B, ZHANG Y H, et al. Non-intrusive mass estimation method for crucian carp using instance segmentation and point cloud processing[J]. Computers and Electronics in Agriculture, 2024, 226: 109445.
[13] ZHOU M G, SHEN P F, ZHU H, et al. In-water fish body-length measurement system based on stereo vision[J]. Sensors, 2023, 23 (14): 6325.
[14] SHI C, WANG Q B, HE X L, et al. An automatic method of fish length estimation using underwater stereo system based on LabVIEW[J]. Computers and Electronics in Agriculture, 2020, 173: 105419.
[15] SETIAWAN A, HADIYANTO H, WIDODO C E. Shrimp body weight estimation in aquaculture ponds using morphometric features based on underwater image analysis and machine learning approach[J]. Revue d'Intelligence Artificielle, 2022, 36(6): 905-912.
[16] 董鹏, 周烽, 赵悰悰, 等. 基于双目视觉的水下海参尺寸自动测量方法[J]. 计算机工程与应用, 2021, 57(8): 271-278.
DONG P, ZHOU F, ZHAO C C, et al. Automatic measurement of underwater sea cucumber size based on binocular vision[J]. Computer Engineering and Applications, 2021, 57(8): 271-278.
[17] LI Q, WANG H J, XIAO Y, et al. Underwater unsupervised stereo matching method based on semantic attention[J]. Journal of Marine Science and Engineering, 2024, 12(7): 1123.
[18] 汤忠强, 周波, 戴先中, 等. 基于改进DCP算法的水下机器人视觉增强[J]. 机器人, 2018, 40(2): 222-230.
TANG Z Q, ZHOU B, DAI X Z, et al. Underwater robot visual enhancements based on the improved DCP algorithm[J]. Robot, 2018, 40(2): 222-230.
[19] 王新伟, 孙亮, 雷平顺, 等. 用于海洋宏生物原位观测的水下激光雷达相机[J]. 红外与激光工程, 2021, 50(6): 37-45.
WANG X W, SUN L, LEI P S, et al. Underwater light ranging and imaging for macro marine life *in situ* observation and measurement[J]. Infrared and Laser Engineering, 2021, 50(6): 37-45.
[20] HU K, WANG T Y, SHEN C W, et al. Overview of underwater 3D reconstruction technology based on optical images[J]. Journal of Marine Science and Engineering, 2023, 11(5): 949.
[21] 崔海朋, 秦朝旭, 马志宇. 基于深度学习的鱼类特征点检测与体征识别方法[J]. 中国农机化学报, 2024, 45(6): 201-207.
CUI H P, QIN C X, MA Z Y. Fish key feature point detection and sign identification based on deep learning[J]. Journal of Chinese Agricultural Mechanization, 2024, 45(6): 201-207.
[22] JIAN M W, YANG N, TAO C, et al. Underwater object detection and datasets: A survey[J]. Intelligent Marine Technology and Systems, 2024, 2(1): 9.
[23] DIAZ-GARCIA P, ESCALONA F, CAZORLA M. UKDM: Underwater keypoint detection and matching using underwater image enhancement techniques[EB/OL]. arXiv: 2504.11063, 2025.
[24] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements[EB/OL]. arXiv: 2410.17725, 2024.
[25] 牛子昂, 裘正军. 基于改进YOLOv11-Pose的玉米植株骨架及表型参数提取方法[J]. 智慧农业(中英文), 2025, 7(2): 95-105.
NIU Z A, QIU Z J. Extraction method of maize plant skeleton and phenotypic parameters based on improved YOLOv11-pose [J]. Smart Agriculture, 2025, 7(2): 95-105.
[26] FU C P, FAN X, XIAO J W, et al. Learning heavily-degraded prior for underwater object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(11): 6887-6896.
[27] 陶洋, 钟邦乾, 赵文博, 等. 融合显示视觉中心与注意力机制的水下目标检测算法[J]. 激光与光电子学进展, 2024, 61(12): 441-450.
TAO Y, ZHONG B Q, ZHAO W B, et al. Underwater object detection algorithm integrating explicit visual center and attention mechanism[J]. Laser & Optoelectronics Progress, 2024, 61(12): 441-450.
[28] TAN M X, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[EB/OL]. arXiv: 1905.11946, 2019.
[29] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2018: 7132-7141.
[30] LEE Y, PARK J. CenterMask: Real-time anchor-free instance segmentation[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2020: 13903-13912.
[31] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2018: 4510-4520.
[32] TANG L F, ZHANG H, XU H, et al. Rethinking the necessity of image fusion in high-level vision tasks: A practical infrared and visible image fusion network based on progressive semantic injection and scene fidelity[J]. Information Fusion, 2023, 99: 101870.
[33] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block attention module[M]// Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.

Underwater *Insitu* Weight Estimation Method for Chinese Mitten Crab Based on Binocular Vision and Improved YOLOv11-pose

LI Aoqiang[1,2,3], DAI Hangyu[1,2,3], GUO Ya[1,2,3*]

(1. *International Joint Research Center for Intelligent Optical Sensing and Applications at Jiangnan University*, *Wuxi* 214122, *China*; 2. *Key Laboratory of Advanced Process Control for Light Industry*, *Ministry of Education*, *Jiangnan University*, *Wuxi* 214122, *China*; 3. *School of Internet of Things Engineering*, *Jiangnan University*, *Wuxi* 214122, *China*)

Abstract: [Objective] With the accelerated development of large-scale and intelligent aquaculture, accurate estimation of the body weight of individual Chinese mitten crabs is critical for tasks such as precise feeding, disease prevention, and optimization of harvest decisions. Traditional methods of manually catching and weighing crabs are time-consuming, labor-intensive, and can cause stress or injury to the crabs, while also failing to provide real-time monitoring. To address the challenges posed by turbid water conditions in aquaculture, which lead to poor image quality and difficulty in feature extraction, a method is proposed for estimating Chinese mitten crab weight that combines binocular vision with deep learning‐based keypoint detection. This approach achieves high-precision detection of anatomical keypoints on the crab, providing new technical support for precision aquaculture and intelligent management. **[Methods]** Based on a lightweight YOLOv11 framework, in its C3K2 module, MBConv depthwise-separable convolutions were incorporated to significantly reduce computational complexity and improve feature extraction efficiency. An EffectiveSE channel attention mechanism was introduced to adaptively emphasize important channel-wise features. To further enhance cross-scale information fusion, a spatial dynamic feature fusion module (SDFM) was added. The SDFM adaptively and weightedly fused local spatial attention with global channel attention, enabling detailed extraction of crab shell edges and anatomical keypoints. The improved YOLOv11-ES model could simultaneously output the crab's bounding box, the positions of four anatomical keypoints, and the crab's sex classification in a single forward pass. In the 3D reconstruction stage, calibrated stereo camera parameters were used, and a sparse keypoint matching strategy guided by the crab's sex and spatial geometric constraints was employed. High-confidence keypoint pairs were selected from the left and right views, and the true 3D coordinates of the crab's carapace length and width were computed by triangulation. Finally, the obtained carapace length, width, and sex label data were fed into a two-layer back-propagation (BP) neural network to perform a regression prediction of the individual crab's weight. **[Results and Discussion]** To validate the effectiveness and robustness of the proposed method, a dataset of Chinese mitten crab images with annotated keypoints was constructed under varying water turbidity and lighting conditions, and both ablation and comparative experiments were conducted. The YOLOv11-ES achieved a mean average precision at intersection over union (IOU) threshold of 0.5 (mAP@50) of 97.2% on the test set, which was 4.4 percentage points higher than the original YOLOv11 model. The keypoint detection component reached an mAP@50 of 96.7%, which was 3.6 percentage points higher than that of the original YOLOv11 model. In comparative experiments, YOLOv11-ES also demonstrated significant advantages over other models in the same series. Moreover, in a full-system evaluation using images of 30 individual crabs, the mean absolute percentage error (MAPE) for carapace width measurements was only 2.68%, and for carapace length it was 1.48%. The Pearson correlation coefficients between the measured and manually obtained true values for both carapace length and width exceeded 0.977, indicating high accuracy in the 3D reconstruction and minimal measurement error. Experiments analyzing the influence of image quality on measurement accuracy showed that when the underwater image quality measure (UIQM) reached at least 1.5, the combined MAPE of carapace length and width errors could be kept below 5%. When UIQM reached at least 2.2, the MAPE dropped to about 1.9%. These results confirmed the robustness of the method against variations in water turbidity and lighting conditions. For weight regression prediction, the BP network trained on carapace length, width, and sex features achieved a mean absolute error (MAE) of 2.39 g and a MAPE of 7.1% on an independent test set, demonstrating high-precision estimation of individual crab weight. **[Conclusions]** The proposed method, which combines an improved YOLOv11 object detection network, binocular sparse keypoint matching, and a two-layer BP regression network, enabled high-precision, low-error, real-time, non-contact estimation of Chinese mitten crab weight in complex turbid aquatic environments. This approach featured a lightweight model, high computational efficiency, excellent measurement accuracy, and strong adaptability to varying environmental conditions. It provided key technical parameters for intelligent Chinese mitten crab farming. In the future, this approach could be extended to other aquaculture species and complex farming scenarios. Combined with transfer learning and online adaptive calibration techniques, its generalization capability could be further improved and integrated with intelligent monitoring platforms to achieve large-scale, all-weather underwater crab weight estimation, contributing to the sustainable development of smart aquaculture.

Key words: Chinese mitten crab; keypoint detection; binocular vision; YOLOv11; weight estimation; *insitu* detection

Foundation items: National Natural Science Foundation of China (51961125102); Modern Agriculture Resersch Funds of Jiangsu Province (BE2022366)

First author: LI Aoqiang, E-mail: liaoqiang919@gmail.com
***Corresponding author:** GUO Ya, E-mail: guoya68@163.com