DOI: 10.19650/j.cnki.cjsi.J2514810

基于改进 RTMPose 的水下光阵关键点检测*

赵世泉,刘佳兴,王宇超
(哈尔滨工程大学智能科学与工程学院 哈尔滨 150001)

摘 要:视觉光学导引是无人潜航器(UUV)近端回收的主流方案,但其检测精度与鲁棒性易受机载设备算力限制及复杂水下环境干扰。针对刚体光阵目标结构特性与水下光学导引任务需求,提出一种基于改进 RTMPose 的水下光阵关键点实时检测方法。核心贡献在于:1)建立水下光阵骨架模型与水下光阵关键点检测数据集,将人体姿态估计自顶向下范式适配至光学导引任务;2)设计融合重参数化技术和坐标注意力机制的轻量化骨干网络,通过解耦训练-推理过程和嵌入分方向的位置信息融合机制,显著降低参数量和计算复杂度的同时,补偿轻量化带来的性能损失,提升模型对水下图像光学衰减退化、遮挡等干扰的鲁棒性。实验结果表明,本方法在自建数据集上平均精度(AP)达到 90.8%,平均端点误差(EPE)为 0.828 pixels;相比于基线模型,改进模型参数量与浮点运算数(Flops)分别降低约 22.8% 和 26.6%;在 Jetson AGX Orin(64 GB)嵌入式平台上,改进模型端到端推理延迟较基线降低了 1.14 ms。泛化性实验表明,该方法能通过训练快速适配不同光阵形态与多种水域条件,具备良好的通用性与可靠性。本研究提出的水下光阵关键点检测方法在检测精度、速度以及鲁棒性这 3 方面取得良好平衡,满足水下光阵关键点检测任务精度及实时性要求,为 UUV 等机载设备上实现高鲁棒性光学导引提供有效的解决思路。

关键词: 无人潜航器;光学导引;关键点检测;RTMPose;重参数化;坐标注意力
中图分类号: TH39 **文献标识码**: A **国家标准学科分类代码**: 520.20

The keypoint detection of underwater optical guidance arrays based on the enhanced RTMPose method

Zhao Shiquan, Liu Jiaxing, Wang Yuchao
(*College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China*)

Abstract: The visual optical guidance is a mainstream solution for the near-end recovery of unmanned underwater vehicles (UUVs), whose accuracy and robustness are susceptible to the limited computational power of onboard devices and interferences of complex underwater environments. In order to address the structural characteristics of rigid optical array targets and the requirements of underwater optical guidance tasks, this study proposes a real-time keypoint detection method for the underwater optical arrays based on the improved RTMPose method. The core contributions are as follows: 1) Establishing a skeleton model for the underwater optical arrays, constructing a dataset for the underwater optical array keypoint detection and adapting the top-down human pose estimation paradigm for the optical guidance tasks; 2) Designing a lightweight backbone network integrating re-parameterization technology and a coordinate attention mechanism. By decoupling the training-inference process and embedding the directional positional information fusion mechanism, the number of parameters and computational complexity are significantly reduced with the compensating ability of performance loss caused by lightweight design, thereby enhancing the model's robustness against the underwater image degradation, occlusion, and other interferences. Experimental results show that the proposed method achieves an average precision (AP) of 90.8% and a mean pixels error of 0.828 with our dataset. Compared to the baseline model, the proposed model reduces the number of parameters and floating-point operations by approximately 22.8% and 26.6% with the Jetson AGX Orin (64 GB) embedded platform, the corresponding end-to-end inference latency is also reduced by 1.14 ms compared to the baseline. The generalization experiments demonstrate that the method can quickly adapt to different optical array configurations and various water conditions through training, which exhibits the fair versatility and reliability. In conclusion, the

收稿日期:2025-12-29 Received Date: 2025-12-29
*基金项目:国家自然科学基金面上项目(52271313)资助

proposed underwater optical array keypoint detection method achieves a favorable balance among detection accuracy, speed, and robustness, which meets the precision and real-time requirements for the underwater optical array keypoint detection tasks and provides an effective solution for the high-robustness optical guidance on UUVs and other onboard devices.

Keywords: UUV; optical guidance; keypoint detection; RTMPose; reparameterization; coordinate attention

0　引　　言

无人潜航器(unmanned underwater vehicle,UUV)水下回收技术是保障其高隐蔽性、高机动性的关键。通过定期水下回收,UUV 可隐蔽地进行能源补充和数据传输,从而显著提高作业范围与作业效率。水下回收一般包括中远端导引与近端导引两个阶段。其中,近端导引要求 UUV 精确驶入回收站,具有较高的定位精度要求,直接决定回收任务的成败[1-2]。基于单目相机的光学导引是 UUV 水下回收近端导引中常用的方法,通过检测安装在回收基站上的信号灯或特殊图像,匹配二维图像与三维空间的关键点,从而获取回收基站与 UUV 的相对位置关系,为运动控制器提供导引信息。

光学导引中检测光阵关键点的方法主要分为两类:传统数字图像处理方法和深度学习方法。传统数字图像处理方法主要根据待检测目标在相机中的成像特征(如形状、几何分布、颜色等),结合实际物理信息筛选图像轮廓,进而确定关键点位置[3]。例如,徐硕等[4]将光源数量设定为迭代条件实现自适应二值化阈值,并通过 K 均值聚类算法剔除伪光源。在自适应阈值分割的基础上,朱志鹏等[5]通过分析水下光线传播特性设计水下彩色图像通道加权补偿法,先增强图像再阈值分割。熊攀等[6]设计暗通道先验去雾算法进行图像增强,改进 Canny 边缘检测,以最大轮廓面积为约束条件迭代更新阈值,最终利用最小包围圆法定位圆心关键点。王秋生等[7]在改进 Canny 算法的基础上,利用根据距离自适应调整阈值的方法分割光源轮廓。由于水体对光的吸收和衰减特性,导引光源常采取蓝绿色。在复杂水下环境中,相机采集的导引光源图像通常呈现出蓝绿色背景、对比度低、光照不均匀、曝光过度、亮斑干扰和生物遮挡等问题,传统数字图像处理方法常因其参数、阈值固定或自适应算法的局限,无法适应不同水域条件,无法满足 UUV 水下回收任务的高鲁棒性需求[8]。

基于深度学习方法通过对大量水下标注样本数据的拟合,通常具有较好的鲁棒性,具有广阔的应用前景[9]。目前主流方法主要有两种:1)利用目标检测器检测光阵区域,剔除光阵外围的干扰,然后在区域内部利用传统图像算法分割光源提取关键点;2)利用目标检测模型或图像分割模型直接检测出所有属于光阵的光源,利用基于光阵几何特征或尺度不变特征转换(scale-invariant feature transform,SIFT)[10]等算法匹配特征。周冰[11]设计基于异色光源的对接站,先利用 YOLOv5(you only look once version 5)网络检测对接站,然后利用光源间的颜色差异区分并匹配关键点,通过设计光源颜色的方式有效降低关键点误匹配的概率。Liu 等[12]提出对接神经网络(docking neural network,DoNN)框架并公开水下暗光图像数据集(underwater dark image dataset,UDID),采用先设计目标区域检测器,再通过自适应分割与 K 均值聚类提取关键点,其算法没有考虑关键点缺失的情况。祝志坤等[13]考虑到水下光学成像模糊等问题,设计结合暗通道先验去雾和 YOLOv9(you only look once version 9)目标检测网络,检测增强后的图像中的每个独立光源,算法可以适应不同水质。但这些方法仍未将光阵的结构信息融入神经网络中,依然无法解决遮挡或光源缺失等特殊情况。

基于自顶向下(top-down)范式[14-15]的实时多人姿态估计(real-time multi-person pose,RTMPose)[16]模型是一种用于多人复杂场景下实时检测人体关节关键点的先进模型,其能在多人相互遮挡的场景下,高精度且快速地完成关键点检测。针对 UUV 机载平台算力有限及水下复杂环境的挑战,更全面地利用光阵结构信息,同时解决水下光阵部分光源成像模糊、丢失以及被遮挡的问题,本文将人体姿态估计 top-down 范式应用到水下光阵关键点检测场景,提出基于 RTMPose-LC 模型的水下目标关键点检测算法。首先,建立包含 8 个关键点的水下光阵骨架模型,并建立光阵关键点数据集;其次,基于实时人体姿态估计模型 RTMPose-s,重构其骨干网络以降低参数量和计算复杂度;然后,嵌入坐标注意力(coordinate attention,CA)[17]模块,使模型更加关注空间位置特征,形成新模型 RTMPose-LC。最后,在自制数据集上进行实验,对比分析验证改进策略的有效性。

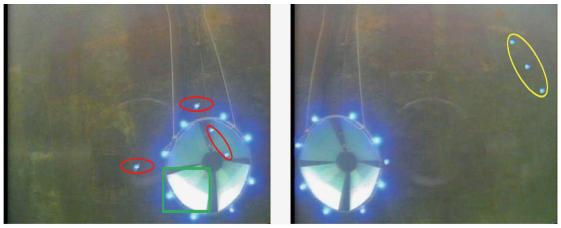
1　基于 RTMPose-LC 的水下光阵关键点检测算法

1.1　基于 top-down 范式的关键点检测框架

1)关键点检测方法框架

在 UUV 自主回收任务中,回收装置上安装的导引光源成像质量受复杂水体环境影响显著。水体对不同波长光的衰减特性差异会导致图像产生色偏;同时,水中悬浮颗粒对光的散射、水中生物及回收装置表面对光的反射

会引入随机光斑与噪声干扰。如图 1(a)所示,回收装置附近出现气泡和回收装置本体反射形成的干扰(图 1(a)中以椭圆形和矩形标注);如图 1(b)所示,图像中出现随机的光斑干扰(图 1(b)中椭圆形标注)。



(a) 气泡和本体反光 (b) 随机光斑
(a) Bubbles and body reflection (b) Random speckle pattern

图 1 回收装置水下成像

Fig. 1 Underwater imaging of the recovery device

受上述两类噪声干扰,当噪声出现在回收装置附近或其形状、亮度与实际导引光源相似时,传统数字图像处理方法难以实现可靠的关键点检测,易发生误检与漏检。对于先利用深度学习检测器确定目标区域,再利用传统数字图像处理方法在目标区域中提取关键点的策略,虽剔除回收装置范围外的噪声干扰,但对于目标区域内出现噪声干扰的情况依然无法有效分割。以上两种思路存在共同的局限性:前者依赖底层像素处理,难以充分利用导引光阵的整体物理信息;后者在前者的基础上缩小处理范围,但本质上依然未能充分利用真实物理信息,没有建立光源之间的位置拓扑关系模型。

为克服上述局限性,本文借鉴人体姿态估计任务思想,将目标骨骼信息导入模型。人体姿态估计包括 top-down 和自底向上(bottom-up)两种范式,前者适用于 5 人以下,无多人重叠且精度要求高的场景,人数增加会降低推理速度;后者适用于多人同时出现的场景,推理速度不受人数影响,但易出现关键点误关联,易受噪点干扰。光学导引任务中不会出现 3 个以上回收装置出现在同一图像中的情况且易出现噪点光源干扰。综合两种范式的优缺点,本文采用基于 top-down 范式的关键点检测框架,将导引光阵的骨架信息(如光源相对位置、拓扑连接、光源顺序等)显式融入检测模型,以提升模型在复杂干扰下的鲁棒性。算法整体框架如图 2 所示。

图 2 基于 Top-down 范式的关键点检测框架

Fig. 2 Keypoint detection framework based on Top-down paradigm

该框架分为两个阶段:(1)目标检测阶段,目标检测器在采集图像中筛选出目标区域,剔除远离光阵区域的噪声干扰,并缩小后续处理范围;(2)姿态估计阶段,在裁剪出的目标区域内,姿态估计器利用光阵的骨架信息回归各个光源关键点,增强对区域内干扰的分辨能力。

2)光阵骨骼模型

姿态估计是计算机视觉领域中的一项基本任务。在人体姿态估计任务中,通过对图像或视频中人体关节位置和角度的估计,推测人体的姿势和动作。由于人体具有柔韧性,其关节相对位置和角度多变,且关键点的可见性受衣物、视角和拍摄环境影响,该任务颇具挑战性。不同于人体骨骼,光阵骨架是不可形变的刚体,其图像形状仅受相机拍摄角度影响。此外,在 UUV 水下回收场景中,通常图像中仅出现一个回收装置,不存在多目标相互遮挡的问题。因此,将人体姿态估计的思想应用到水下光阵关键点检测任务具备合理性和可行性。

如图 3 所示,COCO 数据集中建立的 17 个关键点人体骨骼模型,包括四肢和躯干关键点。类似地,本文建立 8 关键点刚性光阵骨架模型。以右上角靠近回收装置拉线的导引灯为 0 号关键点,按顺时针方向依次为 1~7 号关键点。各关键点的重要性相同,且成对称分布(0~7、1~6、2~5、3~4 分别对称)。

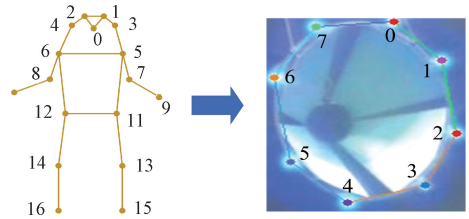


图 3 人体骨骼关键点和光阵骨架关键点

Fig. 3 Human skeletal keypoints and optical array skeleton keypoints

1.2 基于改进 RTMPose 的关键点检测

1)RTMPose-s 模型结构

RTMPose 是由上海人工智能实验室联合 OpenMMlab 团队开发的一种高性能实时多人 2D 姿态估计模型,采用 top-down 范式,强调关键点检测的实时性和准确性的平衡。其摒弃了传统的热图(heatmap)回归方法[18],引入 SimCC(simple coordinate classification)检测头[19],将坐标预测任务转为 X、Y 方向上的分类任务,将图像空间离散为多个坐标区间(bin),通过 Softmax 分类进行预测,使用一维高斯向量进行监督与优化,完全去除热图生成与上采样过程,保证高精度的同时大幅降低推理耗时。在骨干网络部分,采用高效的 CSPNeXt(cross stage partial next)架构[20],由卷积模块(ConvModule)、跨阶段部分连接层(cross stage partial layer,CSPLayer)和空间金字塔池化快速瓶颈层(spatial pyramid pooling fast bottleneck,SPPFBottleneck)等核心模块组成,强调高分辨率、低延迟

和嵌入式设备部署友好性。头部网络包括 7×7 卷积、全连接层(fully connected layer,FC)和门控注意力单元(gated attention unit,GAU)。骨干网络提取的特征先由大核卷积层压缩通道,再经 FC 层扩展维度,利用 GAU 整合全局和局部信息,最终通过坐标分类器实现关键点定位。

2)骨干网络重构

UUV 机载嵌入式平台相比 GPU 服务器,计算资源严重受限,难以直接部署复杂的图像检测或分割算法。此外,UUV 需要执行多种复杂水下任务[21-23],其自主对接回收功能模块应尽可能减少对系统整体计算资源的占用。本研究的任务背景为 UUV 水下对接,不同于多人姿态估计任务,本任务场景中不存在识别多目标的情况,并且采集图像背景相对单一,干扰源相对有限。因此,在保证关键点检测较高精度的前提下,使算法在推理时占用尽可能低的计算资源,对 UUV 系统的实际部署具有重要意义。

模型轻量化有助于降低计算成本并提高推理速度,虽然 RTMPose 采用部署友好型的网络结构,强调实时性,但其原设计面向更复杂的应用场景,直接用于水下光阵关键点识别任务在性能上存在冗余,存在进一步轻量化的空间。同时,RTMPose 采用 hard-swish、hard-sigmoid、leak-relu 等新型高效算子,这些算子在多种嵌入式平台(尤其是国产平台)上兼容性较差。为提升模型在不同嵌入式设备上的通用性,应采用传统的算子,如 Conv2d、ReLU 等。传统算子具有更好的适配性,但其效率低于新型算子,达到同等性能一般需要更为复杂的结构,参数量必然会大幅增加。为降低计算成本并保证较高精度,本文利用结构性重参数化技术[24-26],解耦训练和推理过程,引入重参数化卷积模块(re-parameterizable convolution block,RepConvBlock)[27]替换原 CSPLayer。该模块核心思想是在训练时通过多分支结构充分提取特征,而在推理阶段将多分支结构等效为单一分支,从而实现高精度和高效率。在训练时,RepConvBlock 采用由 3×3 卷积、1×1 卷积和恒等连接组成的多分支并行结构;在推理时,首先将卷积层和归一化层融合为带偏置的融合卷积,如式(1)~(3)所示。

$$\mathrm{Conv}(\boldsymbol{x}) = \boldsymbol{W}(\boldsymbol{x}) + \boldsymbol{b} \tag{1}$$

$$\mathrm{BN}(\boldsymbol{x}) = \boldsymbol{\gamma} \times \frac{\boldsymbol{x} - \mathrm{avg}(\boldsymbol{x})}{\sqrt{\boldsymbol{\sigma}^2(\boldsymbol{x})}} + \boldsymbol{\beta} \tag{2}$$

$$\mathrm{BN}(\mathrm{Conv}(\boldsymbol{x})) = \boldsymbol{\gamma} \times \frac{\mathrm{Conv}(\boldsymbol{x}) - \mathrm{avg}(\boldsymbol{x})}{\sqrt{\boldsymbol{\sigma}^2(\boldsymbol{x})}} + \boldsymbol{\beta} =$$

$$\boldsymbol{\gamma} \times \frac{\boldsymbol{W}(\boldsymbol{x}) + \boldsymbol{b} - \mathrm{avg}(\boldsymbol{x})}{\sqrt{\boldsymbol{\sigma}^2(\boldsymbol{x})}} + \boldsymbol{\beta} =$$

$$\frac{\boldsymbol{\gamma} \times \boldsymbol{W}(\boldsymbol{x})}{\sqrt{\boldsymbol{\sigma}^2(\boldsymbol{x})}} + \left(\frac{\boldsymbol{\gamma} \times (\boldsymbol{b} - \mathrm{avg}(\boldsymbol{x}))}{\sqrt{\boldsymbol{\sigma}^2(\boldsymbol{x})}} + \boldsymbol{\beta} \right) \tag{3}$$

式中:$\mathrm{Conv}(\boldsymbol{x})$ 表示卷积层;$\boldsymbol{W}(\boldsymbol{x})$ 表示卷积层权重;$\boldsymbol{b}$ 表示卷积层偏置;$\mathrm{BN}(\boldsymbol{x})$ 表示归一化层;$\boldsymbol{\gamma}$ 表示缩放参数;$\boldsymbol{\beta}$ 表示偏置参数;avg 表示滑动均值;$\boldsymbol{\sigma}$ 表示滑动方差。

融合后卷积的权重为 $\frac{\boldsymbol{\gamma} \times \boldsymbol{W}(\boldsymbol{x})}{\sqrt{\boldsymbol{\sigma}^2(\boldsymbol{x})}}$,卷积偏置为 $\frac{\boldsymbol{\gamma} \times (\boldsymbol{b} - \mathrm{avg}(\boldsymbol{x}))}{\sqrt{\boldsymbol{\sigma}^2(\boldsymbol{x})}} + \boldsymbol{\beta}$;将并行的 1×1 卷积分支和恒定连接分支都扩充为 3×3 卷积分支;最后根据卷积操作的可加性,将等效的 3 个 3×3 融合卷积分支的权重和偏置相加得到 1 个 3×3 融合卷积。该设计在训练时通过多分支结构保证模型的高表征能力,在推理时等效为单路 3×3 卷积,实现更快的推理速度,更低的内存占用,提升模型部署友好性。本文重构的轻量级特征提取骨干网络如图 4 所示,包含 Stem 层、Stage 模块和下采样模块(DownSample)。

为了降低计算成本,Stem 层使用两个 3×3 的降采样卷积将特征图压缩至输入图像大小的四分之一,特征图空间维度大幅减小,通道数增加。随后,特征图依次通过多个 Stage、DownSample 结构。前馈网络(feed-forward network,FFN)包括两个 1×1 卷积、ReLU 激活函数和残差连接,可以有效提升非线性特征表达能力并保留细节特征。在每个 Stage 后采用 DownSample 模块进行下采样,进一步压缩特征图尺寸,逐步提取从低级细节到高级语义的层次化特征。同时,删除原模型中使用多尺寸卷积核的 SPPFBottleneck 模块,以保证模型低计算资源占用的特性。

3)CA 注意力机制

模型轻量化后往往伴随性能上的损失。为补偿该损失,可在轻量级网络上嵌入适当的注意力机制。删除 SPPFBottleneck 模块后,模型在多尺度特征融合和全局感受野性能上均会下降,需要增加轻量型注意力模块来补偿。CA 由 Hou 等[17]提出,可以将位置信息嵌入通道注意力中,且参数量小,非常适合轻量级网络性能优化,其结构如图 5 所示。

图 5 中,先分别对尺寸为 $[C,H,W]$ 的输入特征图按照 X 和 Y 方向进行平均池化,生成对应方向上的特征图,如式(4)、(5)所示。

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leqslant i < W} \boldsymbol{x}_c(h,i) \tag{4}$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leqslant j < H} \boldsymbol{x}_c(j,w) \tag{5}$$

式中:$z_c^h(h)$ 表示高度为 h 的第 c 个通道的特征图输出;$z_c^w(w)$ 表示宽度为 w 的第 c 个通道的输出;W 表示宽度;i、j 表示对应的坐标位置;H 表示高度。

再将特征图进行变换和拼接,通过共享的 1×1 卷积进行特征融合,再经 BN 层与非线性激活函数层完成

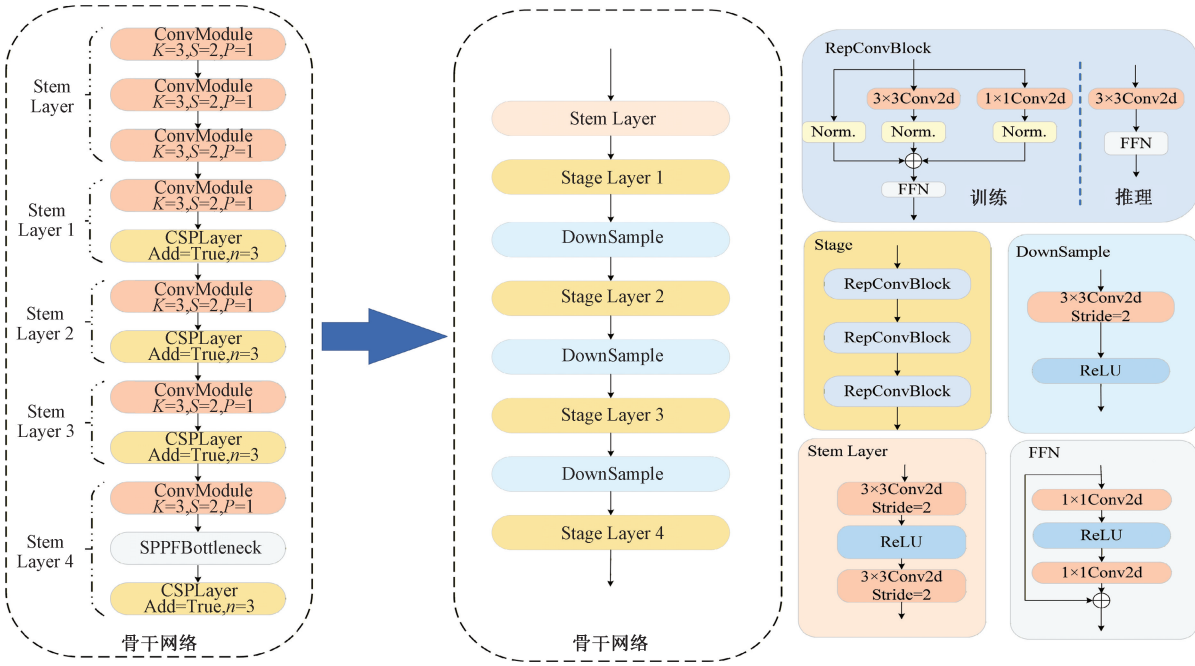


图 4 重构 Backbone 结构

Fig. 4 Restructured Backbone architecture

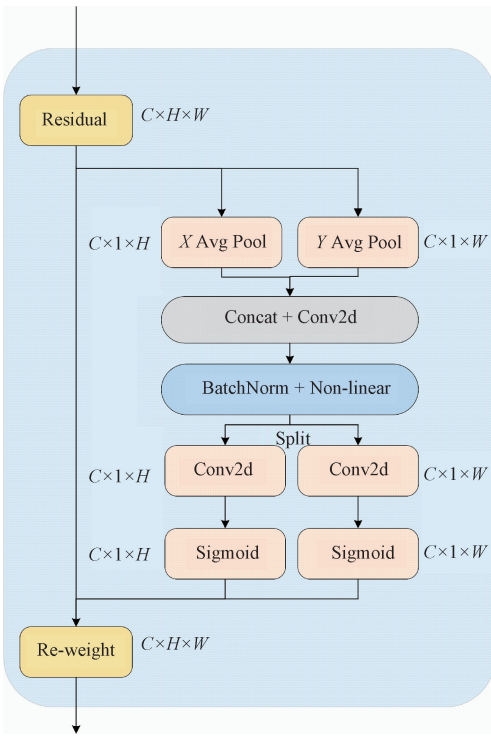


图 5 CA 注意力机制结构

Fig. 5 Structure of the coordinate attention mechanism

归一化,如式(6)所示。然后,沿着空间维度拆分特征图,分别用二维卷积变换到输入相同的通道数,通过 sigmoid 激活函数得到注意力向量,如式(7)、(8)所示。最后,原始特征图与注意力向量相乘得到最后输出,如式(9)所示。

$$\boldsymbol{f}=\delta(\boldsymbol{F}_1([z^h,z^w])) \tag{6}$$

$$\boldsymbol{g}^h=\sigma(\boldsymbol{F}_h(\boldsymbol{f}^h)) \tag{7}$$

$$\boldsymbol{g}^w=\sigma(\boldsymbol{F}_w(\boldsymbol{f}^w)) \tag{8}$$

$$\boldsymbol{y}_c(i,j)=\boldsymbol{x}_c(i,j)\times\boldsymbol{g}_c^h(i)\times\boldsymbol{g}_c^w(j) \tag{9}$$

式中:$\boldsymbol{f}$ 表示在 X 和 Y 方向编码空间信息的中间特征图;δ 表示非线性激活函数;$\boldsymbol{F}_1$,$\boldsymbol{F}_h$,$\boldsymbol{F}_w$ 表示 1×1 卷积;$\boldsymbol{g}$ 表示注意力向量;σ 表示 Sigmoid 激活函数;$\boldsymbol{y}_c(i,j)$ 表示最终输出向量;$\boldsymbol{x}_c(i,j)$ 表示原有特征向量。

在防水相机拍摄的水下光阵图像中,受到水体对光的衰减与散射影响,图像会呈现曝光过度、模糊等情况。引入 CA 注意力机制可提升模型对位置信息的敏感度,弥补轻量化带来的性能损失。CA 会将通道注意力分解为沿水平方向和垂直方向的 1D 特征编码,将位置信息分方向融入注意力图中,从而加强输入特征在特定方向上的长距离依赖性。这样的处理方式与本文中头部网络中的坐标分类器适配性较高。

本文将坐标注意力机制模块 CA 嵌入到 RepConvBlock 的多分支结构与 FFN 网络中间,得到 RepConvCABlock。在融合不同层级特征时更加关注空间位置信息,弥补由于模型轻量化导致的性能损失,如图 6 所示。

4)改进 RTMPose-LC 模型总体结构

改进后的模型命名为 RTMPose-LC,其总体结构如图 7 所示。模型主要包括基于重参数化技术重构的骨干

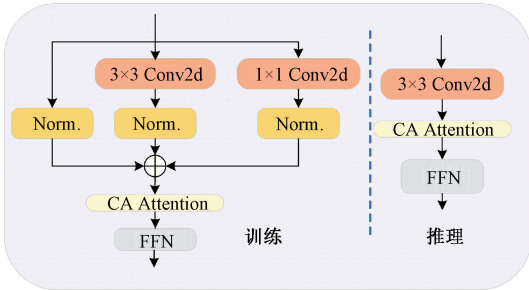


图 6　RepConvCABlock 结构

Fig. 6　Structure of the RepConvCA block

网络和 SimCC 坐标分类头。为避免结构冗余,本文只在部分 RepConvBlock 模块中嵌入 CA 注意力模块,即把每个 Stage 层的第 2 个 RepConvBlock 模块替换为 RepConvCABlock 模块。

RTMPose-LC 模型沿用原 RTMPose-s 的高效坐标分类头,以保证其高精度坐标回归能力,为进一步提升关键点坐标回归精度,在原始 RTMPose 的 Kullback-Leibler 散度(Kullback-Leibler divergence,KLD)损失函数的基础上加权融合均方误差(mean-square error,MSE)损失作为新的损失函数。总的损失函数如式(10)所示。

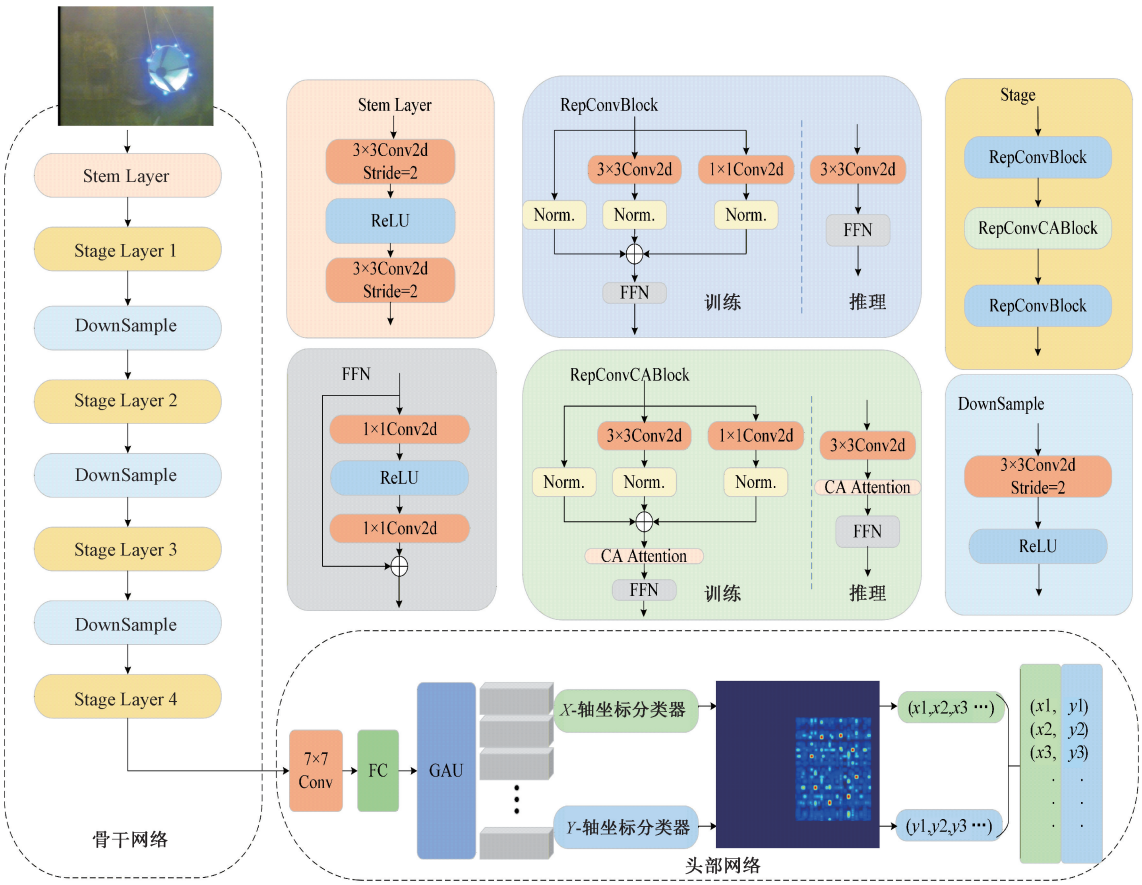


图 7　改进模型总体结构

Fig. 7　Overall structure of the improved model

$$L_{\text{total}} = \eta L_{\text{kld}} + \xi L_{\text{mse}} = \eta \sum_{i=1}^{k} P(i) \lg \frac{P(i)}{Q(i)} + \xi \frac{1}{8} \sum_{i=1}^{8} (y_i - \hat{y}_i)^2 \tag{10}$$

式中:$P(i)$ 表示第 i 个关键点真实概率分布;$Q(i)$ 表示第 i 个关键点预测概率分布;y_i 表示真实坐标值;$\hat{y}_i$ 表示预测坐标值;η,ξ 表示加权系数,本文中分别取 1.0 和 0.5。

骨干网络采用重构的轻量型特征提取网络,利用重参数卷积模块解耦模型训练和推理过程。训练阶段采用多分支并行结构充分提取、融合各层特征;推理阶段将多分支结构等效为单路 3×3 融合卷积,保证推理精度,同时提升推理速度,降低内存占用,更能满足嵌入式平台的部署要求。

2　水下光阵关键点检测实验及结果分析

2.1　水下光阵关键点数据集及标注

中国科学院沈阳自动化研究所 UDID 开源数据集提供 10 000 张水下相机实时采集图像,可以作为本文关键

点检测数据集的基础数据。为降低低质量图像对模型训练的影响,从 UDID 数据集中筛选出 2 000 张不同拍摄角度、不同拍摄距离且不存在拖影的图像,分辨率 448×448。采用 labelme 工具进行目标类别和关键点位置信息标注,标注遵循 COCO 数据集格式,标注可见性,目标检测框和关键点。标注示例如图 8 所示,矩形框标注待检测光阵区域,不同灰度的点依次标注待检测光源圆心。将数据集按 8∶1∶1的比例划分为训练集、验证集和测试集。

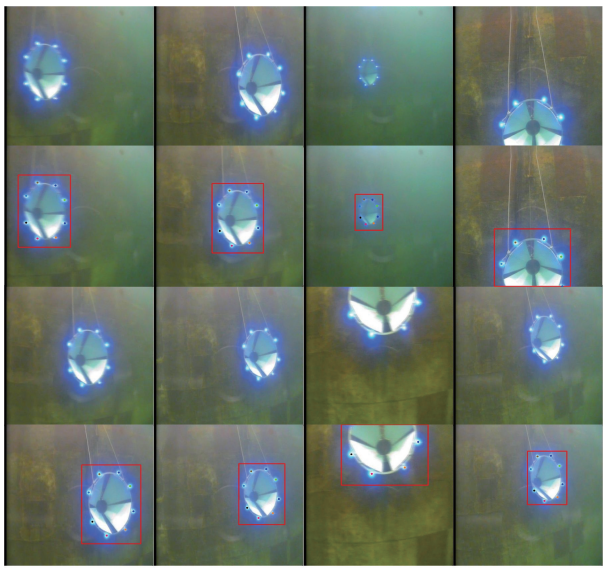


图 8 数据集标注示例

Fig. 8 The annotation examples of dataset

2.2 评价指标

类比于二维人体姿态估计,本文沿用其评价指标,选取平均精度(average precision,AP)与平均召回率(average recall,AR)两个指标。本文自建数据集格式为 COCO 数据集通用格式用于关键点检测,通过计算目标关键点相似度(object keypoint similarity,OKS)来计算 AP 和 AR,如果单个预测关键点坐标和真实关键点坐标在像素坐标系的欧式距离在一定阈值范围内,则认为此次预测是正确的。目标关键点相似度公式如式(11)所示。

$$OKS_p = \frac{\sum_i \exp\{-d_{pi}^2/2S_p^2\sigma_i^2\}\delta(v_{pi} > 0)}{\sum_i \delta(v_{pi} > 0)} \tag{11}$$

式中:p 表示目标所属类别(如 Person、Lights 等);d_{pi} 表示 p 类第 i 个关键点预测值与真实值的欧式距离;S_p 为尺度因子,一般选取目标边界框的面积;σ_i 表示第 i 个关键点的标准差;v_{pi} 代表第 i 个关键点的可见性。

根据计算出的光阵所有预测关键点的 OKS 可以得到 AP 与 AR。不同的 OKS 阈值下的 AP 与 AR 能全面反映出模型的性能,本文重点关注平均精度均值(mean average precision,mAP)、中等目标 AP(M)两个指标。

为更准确地评估关键点检测精度,本文还增加平均端点误差(end point error,EPE)评价指标,通过计算预测关键点与真实关键点的平均欧氏距离,更加直观地显示逐点的绝对误差。EPE 计算如式(12)所示。

$$EPE = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i^{\text{pred}} - x_i^{\text{gt}})^2 + (y_i^{\text{pred}} - y_i^{\text{gt}})^2} \tag{12}$$

式中:N 为关键点数量;$(x_i^{\text{pred}}, y_i^{\text{pred}})$ 表示第 i 个点的预测坐标;$(x_i^{\text{gt}}, y_i^{\text{gt}})$ 表示第 i 个点的真实坐标。

2.3 模型训练

本文所有的模型训练均基于 PyTorch 和 MMPose 工具箱,实验平台配置为 Intel CORe i7-13620H、NVIDIA RTX 4060 8 G 显卡。模型训练环境为 ubuntu18.04、Anaconda3+Python3.7+CUDA10.0+ PyTorch1.1.0。输入图像尺寸为 256×192,设置初始学习率为 0.004,最大训练轮数为 420,batch size 为 128,前 5 个训练轮次使用线性预热,第 200~420 轮使用余弦退火学习率调度,最小学习率设置为初始学习率的 5%。为提升网络的泛化性,在训练阶段增加图像增强,如随机水平翻转、颜色抖动、随机遮挡、模糊增强等。

2.4 对比实验与结果分析

1)对比实验

为验证本文提出的 RTMPose-LC 模型在水下光阵关键点检测任务中的准确性与实时性,将本文模型与部分先进的两阶段关键点检测模型进行比较,所使用的目标检测器均为 RTMDet-s,结果如表 1 所示。

表 1 不同模型在自建数据集上的关键点检测结果

Table 1 Keypoint detection results of different models on a self-built dataset

网络模型	*GFLOPs*	参数量/M	*AP*/%	*AP*(*M*)/%	端点误差/pixels	帧率/fps
Litehrnet	0.265	1.130	78.9	59.5	1.197	100.72
DeepPose	0.307	2.319	80.9	43.1	0.945	162.72
RTMPose-s	0.665	5.247	90.5	68.1	0.831	169.52
本文模型	0.488	4.051	90.8	76.4	0.828	221.81

表 1 中,在相同条件下,选取 3 个基于 top-down 范式的轻量化模型与本文模型对比。其中,Litehrnet(lightweight high-resolution network)为基于热力图的模型;DeepPose 为基于直接回归的模型,其骨干网络选用 mobilenetV2 轻量化网络;RTMPose-s 为本文的 baseline 模型。对比实验结果可知,在 AP 指标上,本文模型继承 RTMPose-s 的高精度属性,分别比 litehrnet 和 DeepPose 提高 11.9 和 9.9 个百分点;在 AP(M)指标上提升更为显著,分别比 litehrnet、DeepPose 和 RTMPose-s 提升 16.9、33.3 和 8.3 个百分点。参数量和计算复杂度相较于 baseline 分

别降低 22.8% 和 26.6%,FPS(frames per second)分别比 litehrnet、DeepPose 和 RTMPose-s 提升约 120%、36.4% 和 30.7%。RTMPose-LC 模型在拥有最高准确度的同时有更快的推理速度,实现推理阶段的精度与速度的同时提升。

针对水下光阵关键点识别任务,为全面评估不同范式算法,选取基于单阶段的模型 RTMO-s、YOLOPose-s 与本文模型对比,实验结果如表 2 所示。

表 2　不同范式模型关键点检测结果

Table 2　Keypoint detection results for the models of different paradigms

网络模型	参数量/M	*AP*/%	*AP*(*M*)/%
YOLOPose-s	15.100	67.4	34.6
RTMO-s	9.000	73.9	31.7
RTMPose-s	5.247	90.5	68.1
本文模型	4.051	90.8	76.4

表 2 中,RTMPose-s 和本文模型关键点检测准确率均明显高于 YOLOPose-s 和 RTMO-s(real-time multi-person pose estimation),在检测小目标时精度均超出约 30%。表 2 实验数据均是在理想检测器条件下得出,top-down 范式的检测模型在实际部署时会受到目标检测器精度限制,后续章节进行讨论。

2)关键点检测性能影响分析

针对 top-down 范式的检测模型在实际部署时精度受目标检测器性能限制的问题,RTMPose-LC 在训练时针对输入的目标检测框数据进行数据增强,包括缩放、旋转、平移以及部分遮挡,模拟目标检测器检测结果失准的情况。同时,还采用基于关键点生成扩展边界框的方法减少对目标检测器的依赖,保障在检测器失准情况下模型的性能。为评估上述训练策略效果,在姿态估计阶段的目标框输入数据随机加上 10%~40% 的偏移,检测结果如图 9 与表 3 所示。

图 9 为偏移 10% 的测试集部分数据,从左至右依次为原图、正常推理结果和偏移检测框推理结果,在偏移检测结果中标记出每个关键点的置信度。在目标检测框外面的区域,模型仍然具备检测关键点的能力,且偏离目标框较远的位置关键点检测置信度低(图中最低 0.46、0.39、0.58)。通过训练,姿态估计器对目标检测器的依赖性降低,在框外区域也具有检测能力,实际部署可以通过置信度筛选关键点。

表 3 中,模型在检测框偏移 20% 以下时,AP 下降 2.6%,AP(M)下降 4.8%,EPE 低于 2 pixels,精度能够保持;偏移超过 40%时,准确率下降严重(下降 26%),EPE 增加 5 倍,模型性能受检测器影响很大。所以,在检测器发生 20%以下的检测偏差时,本文模型具有良好的鲁棒性,

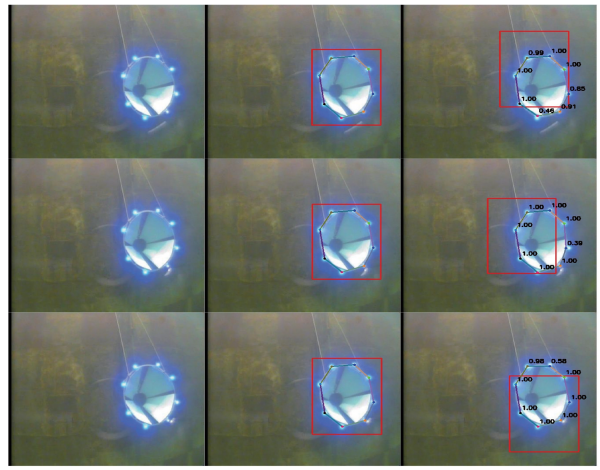


图 9　检测器依赖分析结果

Fig. 9　The dependency analysis result chart of detector

表 3　检测框偏移测试结果

Table 3　Test results of bounding box offset

偏移量	*AP*/%	*AP*(*M*)/%	端点误差/pixels
10%	89.8	75.4	0.846
20%	88.2	71.6	0.883
30%	81.7	66.6	1.643
40%	64.8	53.2	5.551

经过后端关键点置信度筛选能够适配关键点检测任务;在检测器出现大幅偏差、误检或失效时,模型检测精度大幅下降,需要增加后处理逻辑参照前后帧数据才能适配关键点检测任务。此外,为具体分析本文所提出的各个模块对模型性能的影响,设计消融实验,结果如表 4 所示。

表 4　消融实验结果对比

Table 4　Comparison of ablation experiment results

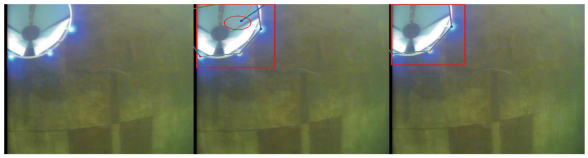
模型构成	*AP*/%	*AP*(*M*)/%	参数量/M	*GFLOPs*	端点误差/pixels
RTMPose-s	90.5	68.1	5.247	0.665	0.831
+RepConv	88.7	63.8	4.038	0.487	0.864
+RepConv+CA	90.2	73.6	4.051	0.488	0.843
+RepConv+CA+MSE	90.8	76.4	4.051	0.488	0.828

分析表 4 中各个评价指标可以得出,通过重构 Backbone 和重参数化技术,模型的计算复杂度(Flops)从 0.665 G 下降到 0.487 G,降低约 26.6%,参数量(Params)从 5.247 M 下降至 4.038 M,降低约 22.8%。Backbone 重构后 AP 下降 1.8%,AP(M)下降 4.3%,表明此种模型轻量化的方法会导致性能下降。嵌入 CA 注意力模块后,AP(M)提升至 73.6%,超过基线模型 5.5 个百分点,CA 模

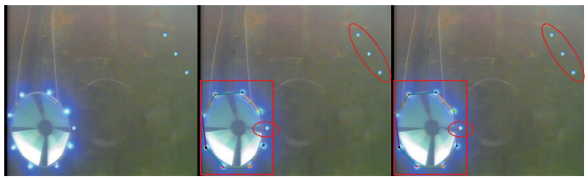
块给模型带来更强的位置感知能力,提升模型对于模糊图像的检测能力。最后,融合 MSE 损失函数进一步提升坐标回归的准确性。模型计算量和参数量基本稳定,且中等目标检测准确率提升较大,性能提升明显。

3)对比模型推理效果

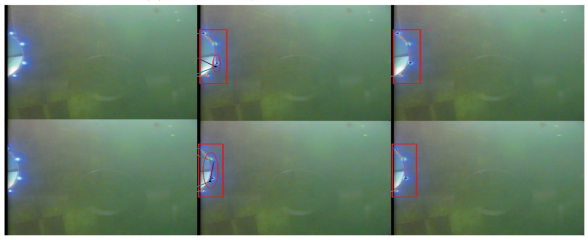
对训练好的模型推理效果进行对比,从 UDID 数据集中选取自建数据集外的数据进行模型推理测试,推理效果如图 10 所示。



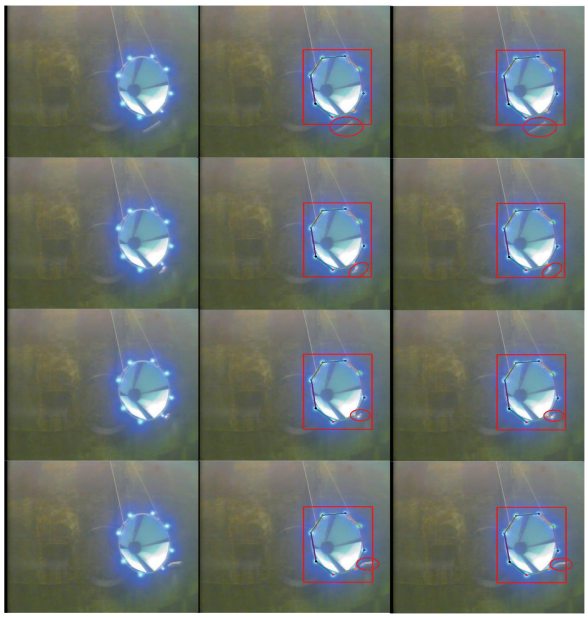
(a) 本体反光
(a) Body reflection



(b) 随机光源干扰
(b) Random light source interderence



(c) 连续帧出现大面积遮挡
(c) Large-area occlusion occurs in consecutive frames



(d) 鱼类遮挡光源
(d) Occlusion of the light source by fish

图 10 推理结果

Fig. 10 Inference results

图 10 中,选取 8 张测试图片进行推理,共 a、b、c、d 这 4 组推理结果,每组图片从左至右为原图、RTMPose-s 推理结果和本文模型推理结果。图 10(a)代表光阵本体表面反光的情况,图中椭圆形标记处,RTMPose-s 出现误检,本文模型推理结果正确;图 10(b)代表出现随机干扰光源情况(图中椭圆形标出),两个模型均具有分辨干扰光源的能力;图 10(c)代表光阵被大面积遮挡情况,图中椭圆形标记处,RTMPose-s 出现误检,本文模型推理结果正确;图 10(d)中光阵附近出现鱼类干扰(椭圆形标记),两个模型同样具备分辨动态干扰物的能力。根据推理结果可以得出:在检测目标较为完整的情况下,RTMPose-s 模型和本文模型均具备区分干扰光源、水泡干扰和动物动态干扰的情况。相较于 RTMPose-s 模型,本文模型在检测目标存在大面积遮挡和大面积反光的情况下,得益于 CA 模块更强的位置感知能力,具有更好的抗遮挡性能。

4)模型泛化性分析

为评估 RTMPose-LC 模型的泛化性,本文增设清水池实验和 Gazebo 仿真环境实验。这两种环境与 UDID 数据集采集图片成像效果存在显著差异,清水池采集图像如图 11 所示,Gazebo 采集图像如图 12(b)所示。因水质不同清水池采集的图像整体偏蓝,UDID 采集图像整体偏绿。

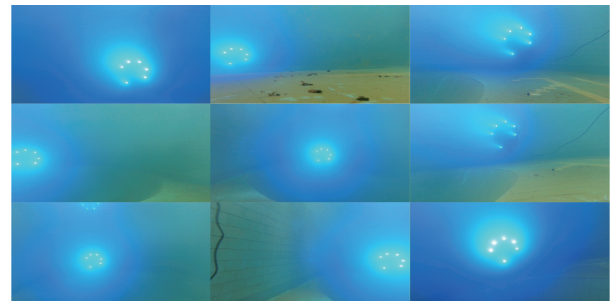
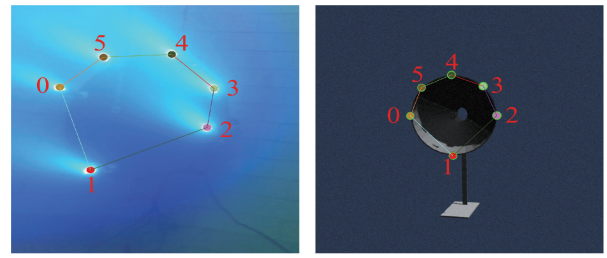


图 11 清水池采集图像

Fig. 11 Images collected from a clean water pool



(a) 清水池6关键点模型
(a) Pool 6-keypoint model

(b) Gazebo 6关键点模型
(b) Gazebo 6-keypoint model

图 12 Pool-datasets 关键点和 simulator-datasets 关键点

Fig. 12 Keypoints of pool-datasets and keypoints for the simulator-datasets

本节实验采用重新设计的 6 关键点模型,分别在水池和 VVV-Simulator 仿真环境采集图像,图像分辨率为

1 920×1 080,建立 Pool-datasets(2 000 张)和 Simulator-datasets(2 000 张),6 关键点模型如图 12(a)、(b)所示。

图 12 中,6 个关键点呈现圆形分布,0~2、3~5 呈左右对称布局,1~4 连线与 0~2 连线为垂直关系,3 与 5 分别位于其相邻两点的角平分线与圆交点处。Pool-datasets 采用与 2.1 节相同类型的标注方式,Simulator-datasets 直接将仿真环境实际像素坐标作为标注信息。实验设置如下:UDID 代表相对复杂、浑浊度较高的自然水体,Pool-datasets 代表浑浊度低的真实水体,Simulator-datasets 代表理想条件下的水下环境。采用与前文相同的训练策略,不同数据集训练结果如表 5 所示。

表 5　RTMPose-LC 在不同数据集上性能对比

Table 5　Performance comparison of RTMPose-LC across different datasets

数据源	光阵模型	图像分辨率	*AP*/%	端点误差/pixels
UDID	8 关键点	448×448	90.5	0.828
Pool	6 关键点	1 920×1 080	95.6	1.841
Gazebo	6 关键点	1 920×1 080	99.8	0.938

表 5 中,对比模型在 UDID 数据、Gazebo 仿真数据与水池数据 3 个自建数据集的关键点检测性能。受实验条件限制,无法获取 6 关键点光阵在 UDID 环境下的图像数据,本节重点分析清水池和仿真环境中的数据。模型在 Pool 环境中 AP 为 95.6%,在 Gazebo 环境中 AP 为 99.8%,这表明当水体环境趋于理想化时,模型能保持较高精度的检测能力,对水质、光照变化具有良好的适应性。新增两个环境中,Pool 环境 EPE(1.841),Gazebo 环境 EPE(0.938)数值高于 UDID 环境 EPE(0.828)。Simulator-datasets 直接使用仿真环境的真值作为标注信息,标注精度远远高于 Pool-datasets 的人工标注方式且仿真环境噪声极少,所以在同样分辨率下精度较高。此外,受 SimCC 分类头分类方式的影响,在更高分辨率的图像上,单个像素表示的物理区域变小,导致计算 EPE 时数值偏大,而实际物理距离并未增大。表 5 中水池数据分辨率由 448×448 变为 1 920×1 080,EPE 数值有限增加,由 0.828 增加至 1.841 是合理的。泛化性对比试验表明,本文模型在不同水质水池和虚拟仿真环境下性能稳定,并且通过训练能适应目标形态变化,具有良好的泛化能力。

5)模型部署

为检验 RTMPose-LC 模型在嵌入式平台的推理速度,本文将在 PC 端训练的模型迁移到 Jetson AGX Orin(64 GB)高性能嵌入式 AI 模组上,测试基本环境为:Jetpack5.1.2、CUDA11.4、Cudnn8.6.0、Tensorrt8.6.9、OpenCV4.5,如图 13 所示。

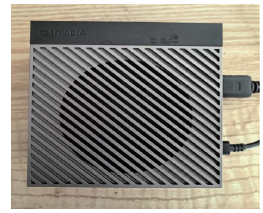


图 13　Jetson AGX Orin(64 GB)

Fig. 13　Jetson AGX Orin(64 GB)

PC 端训练完成的模型权重 pth 文件转换为 onnx 文件,再经 Tensorrt 转换为 trt 文件(FP16)部署到嵌入式设备上,推理速度测试结果如表 6 所示。

表 6　模型部署推理速度对比

Table 6　Comparsion of model deployment inference speed

网络模型	输入尺寸/pixels	推理延迟/ms	端到端延迟/ms	帧率/fps
RTMPose-s	256×192	3.28	6.43	155.44
RTMPose-LC	256×192	2.85	5.29	188.89

表 6 中,推理延迟表示使用随机数据测量模型本身的推理延迟;端到端延迟表示使用真实图像数据,包括图像加载、预处理和推理后端数据处理等步骤,模拟实际使用时总的推理延迟。分析表 6 可知,得益于重参数卷积技术,将模型骨干网络的多分支特征提取结构等效为单分支结构(3×3 混合卷积),本文模型在嵌入式设备上的推理速度优于 RTMPose-s。

3　结　　论

本文针对 UUV 水下关键点检测任务中模型鲁棒性和精度难以兼顾的问题,将人体姿态估计的思想迁移至 UUV 水下回收光学导引任务中,提出一种实时、准确的关键点检测方法。设计了一种针对机载设备算力受限难题的轻量化网络结构,利用重参数化技术重构骨干网络并通过轻量级注意力机制提升模型性能,解决了轻量化模型位置感知能力弱的问题。实验表明,本文提出的 RTMPose-LC 模型在精度、速度和鲁棒性这 3 个方面取得了良好平衡。此外,本文还进行了模型泛化性分析,设计了清水池和虚拟仿真环境对比试验,更改目标关键点分布,证明了本文模型具有良好的通用性,能够更好地满足 UUV 光学导引任务实时性和抗干扰性的需求,是解决水下光阵关键点识别的有效方案。

参考文献

[1] YANG Q SH, LIU H T, HONG L, et al. Anti-disturbance control strategy in capture stage for AUV dynamic base docking with optical guided constraints[J]. Ocean Engineering, 2024, 311: 118946.

[2] TAN X G, ZHANG G M. Research on surface defect detection technology of wind turbine blade based on UAV image[J]. Instrumentation, 2022, 9(1): 41-48.

[3] ABRO G M E, MAHMOUD E. A hybrid PSO-ACO algorithm for precise localization and geometric error reduction in industrial robots [J]. Instrumentation, 2025, 12(1): 70-76.

[4] 徐硕, 姜言清, 李晔, 等. 智能水下机器人自主回收的双目视觉定位[J]. 哈尔滨工程大学学报, 2022, 43(8): 1084-1090.
XU SH, JIANG Y Q, LI Y, et al. A stereo vision localization method for autonomous recovery of autonomous underwater vehicle [J]. Journal of Harbin Engineering University, 2022, 43(8): 1084-1090.

[5] 朱志鹏, 张晓文, 韩祥. 基于单双目切换的自主式水下机器人视觉定位方法[J]. 船舶工程, 2023, 45(2): 131-139, 144.
ZHU ZH P, ZHANG X W, HAN X. AUV visual localization method based on monocular and binocular switching[J]. Ship Engineering, 2023, 45(2): 131-139, 144.

[6] 熊攀, 齐向东, 孙岩林, 等. 多特征融合的 AUV 末端视觉自主对接技术研究[J/OL]. 中国舰船研究, 1-13 [2026-01-06].
XIONG P, QI X D, SUN Y L, et al. Research on multi-feature fusion-based terminal visual autonomous docking technology for AUV [J/OL]. Chinese Journal of Ship Research, 1-13[2026-01-06].

[7] 王秋生, 齐向东, 薛晨阳, 等. AUV 终端对接的单目视觉导引方法研究(英文)[J]. Journal of Measurement Science and Instrumentation, 2024, 15(2): 216-223.
WANG Q SH, QI X D, XUE CH Y, et al. Monocular visual guidance method for terminal docking of an autonomous underwater vehicle[J]. Journal of Measurement Science and Instrumentation, 2024, 15(2): 216-223.

[8] 李学龙, 孙哲, 吴国俊. AUV 水下回收光学导引(特邀)[J]. 光学学报, 2025, 45(12): 9-30.
LI X L, SUN ZH, WU G J. Optical guidance for AUV underwater docking (invited) [J]. Acta Optica Sinica, 2025, 45(12): 9-30.

[9] WANG Z Y, ZHOU X Q, YANG Y, et al. Robust under-water docking visual guidance and positioning method based on a cage-type dual-layer guiding light array[J]. Sensors, 2025, 25(20): 6333.

[10] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[11] 周冰. 面向 AUV 回收的异色光源阵列视觉导引方法研究[D]. 长春: 吉林大学, 2024.
ZHOU B. Research on visual guidance method of heterochromatic light source array for AUV recovery[D]. Changchun: Jilin University, 2024.

[12] LIU SH, OZAY M, OKATANI T, et al. A vision based system for underwater docking[J]. ArXiv preprint arXiv: 1712.04138,2017.

[13] 祝志坤, 卢丙举, 李一辰, 等. 基于单双目融合的 AUV 坐落式回收光视觉引导算法[J]. 控制与决策, 2025, 40(1): 28-37.
ZHU ZH K, LU B J, LI Y CH, et al. Light visual guidance algorithm for AUV situated recovery based on monocular and binocular fusion [J]. Control and Decision, 2025, 40(1): 28-37.

[14] ZHANG L X, HUANG W T, WANG CH L, et al. Improved multi-person 2D human pose estimation using attention mechanisms and hard example mining [J]. Sustainability, 2023, 15(18): 13363.

[15] 张小娜, 吴庆涛. 基于深度学习的自顶向下人体姿态估计算法[J]. 电子测量技术, 2021, 44(9): 105-109.
ZHANG X N, WU Q T. Top-down human pose estimation algorithm based on deep learning [J]. Electronic Measurement Technology, 2021, 44(9): 105-109.

[16] JIANG T, LU P, ZHANG L, et al. RTMPose: Real-time multi-person pose estimation based on MMPose[J]. ArXiv prepint arXiv: 2303.07399,2013.

[17] HOU Q B, ZHOU D Q, FENG J SH. Coordinate attention for efficient Mobile network design[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.

[18] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5693-5703.

[19] LI Y J, YANG S, LIU P D, et al. SimCC: A simple coor-dinate classification perspective for human pose estim-ation[C]. Computer Vision-ECCV 2022, 2022: 89-106.

[20] LYU CH Q, ZHANG W W, HUANG H AN, et al. RTMDet: An empirical study of designing real-time

object detectors[J]. ArXiv preprint arXiv: 2212.07784, 2022.

[21] SONG CH ZH, NIU L M. Target path tracking method of intelligent vehicle based on competitive cooperative game[J]. Instrumentation, 2022, 9(1): 31-40.

[22] 徐会希, 吕凤天, 石凯, 等. 深海长期驻留自主水下机器人系统研究现状与发展趋势[J]. 机器人, 2023, 45(6): 720-736.
XU H X. LYU F T, SHI K, et al. Research status and development trends of deep-sea long-term resident autonomous underwater vehicle systems [J]. Robot, 2023, 45(6): 720-736.

[23] 周道先, 张吟龙, 徐高飞, 等. 基于形变卷积和深层聚合网络的水下文物检测[J]. 仪器仪表学报, 2023, 44(11): 185-195.
ZHOU D X, ZHANG Y L, XU G F, et al. Underwater cultural artifact detection based on deformable convolution and deep layer aggregation network[J]. Chinese Journal of Scientific Instrument, 2023, 44(11): 185-195.

[24] DING X H, ZHANG X Y, MA N N, et al. RepVGG: Making VGG-style ConvNets great again [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13728-13737.

[25] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.

[26] 韩萍, 白继睿, 周杰龙, 等. 基于轻量化改进和模型剪枝的 SAR 图像飞机目标检测[J]. 仪器仪表学报, 2025, 46(9): 110-124.
HAN P, BAI J R, ZHOU J L, et al. Lightweight SAR image aircraft target detection based on lightweight improvement and model pruning[J]. Chinese Journal of Scientific Instrument, 2025, 46(9): 110-124.

[27] WEN SH Y, GAO Y, HU B R, et al. AARPose: Real-time and accurate drogue pose measurement based on monocular vision for autonomous aerial refueling [J]. Chinese Journal of Aeronautics, 2025, 38(6): 103307.

作者简介

赵世泉,2013 年于哈尔滨工程大学获得学士学位,2020 年于比利时 Ghent University 和哈尔滨工程大学获得博士学位,现为哈尔滨工程大学副教授,主要研究方向为复杂过程控制系统、船舶运动控制系统、预测控制、人工智能等。

E-mail:zhaoshiquan@hrbeu.edu.cn

Zhao Shiquan received his B.Sc. degree from Harbin Engineering University in 2013, and his Ph.D. degree from Ghent University and Harbin Engineering University in 2020, respectively. He is currently an associate professor at Harbin Engineering University, specializing in research areas of complex process control systems, ship motion control systems, predictive control, artificial intelligence, etc.

王宇超(通信作者),2009 年于哈尔滨工程大学获得硕士学位,2011 年于哈尔滨工程大学获得博士学位,现为哈尔滨工程大学副教授,博士生导师,主要研究方向为水下无人系统智能感知、海洋运载器自主决策与智能控制、嵌入式系统设计与集成等。

E-mail:wangyuchao@hrbeu.edu.cn

Wang Yuchao (Corresponding author) received his M.Sc. degree in 2009 from Harbin Engineering University, his Ph.D. degree in 2011 from Harbin Engineering University. He is currently an associate professor and doctoral supervisor at Harbin Engineering University. His main research interests include intelligent perception for underwater unmanned systems, autonomous decision-making and intelligent control for marine vehicles, embedded system design and integration, etc.