

# 基于 CMYolov8 的单环刺螠洞口检测

尹骏杨<sup>1</sup> 闵莹<sup>1</sup> 尹延峰<sup>2</sup> 纪元<sup>1</sup> 1 通讯作者

1 中国农业大学烟台研究院, 山东烟台, 264670;

2 马庄镇农业综合服务中心, 山东泰安, 271000;

**摘要:** 本研究针对单环刺螠自动采捕和产量预测需求, 提出了一种基于深度学习的单环刺螠洞口识别方法, 并设计了适用于自动采捕船嵌入式设备的架构。该方法提出了一种新颖的 CSPHet 架构, 取代了 YOLOv8 中的 C2f 结构。CSPHet 通过 HetConv 实现异构核卷积, 显著减少计算量和参数数量, 同时保持高效的特征表示能力。模型通过  $1 \times 1$  卷积预处理, 分支网络分别保留原始特征和进行 HetConv 特征提取, 并通过 CSPHet\_Bottleneck 模块进一步深化特征表达。此外, 模型融合了 FPN-PAN 结构及 MetaNeXt 模块, 通过将大核卷积设计为 Inception 分支, 增强位置信息的整合能力, 提升了检测精度并减轻硬件部署负担。实验结果表明, 改进后的网络在参数量和计算量上分别减少了 16.7% 和 18.8% 的同时, map50 和 map95 上分别上涨了 6.7%, 13.9%。为实际应用场景下模型部署提供了模型基础。

**关键词:** 单环刺螠洞口; 目标检测; 深度学习; YOLO v8

**DOI:** 10.69979/3041-0673.25.04.054

## 引言

单环刺螠又称海肠、海肠子, 隶属螠虫动物门螠纲无管螠目刺螠科, 是无管螠目在我国沿海分布的唯一种类。单环刺螠体型肥大、肉红色, 是一种具有重要经济价值和潜在药用价值的海洋无脊椎动物<sup>[1]</sup>。目前我国仍采用手工挖捕成体的方式, 其缺陷较多, 容易造成单环刺螠死亡或残损, 且采捕效率低、成本高, 影响池塘内刺参的生态混养等<sup>[2]</sup>。

自动化采捕技术为此提供了一个有效的解决方案, 能够显著提高采捕效率, 减少人工成本, 同时对生态环境的影响较小。然而, 单环刺螠栖息环境复杂, 其“U”型洞穴结构和强大的潜沙能力为自动采捕带来了技术挑战。为了应对这些挑战, 我们选择应用计算机视觉技术。

YOLO (You Only Look Once) 系列模型因其出色的检测速度和较高的精度, 在实时目标检测中得到了广泛应用。YOLOv8 继承了 YOLO 的快速特性, 在处理小规模数据集和实时检测任务时表现出色。然而, 在某些特殊场景下, 如单环刺螠洞口检测任务中, 仍需进一步优化其模型结构以应对复杂的环境特征。

为此, 本研究提出了一种基于轻量化深度学习架构的洞口检测方法, 基于 YOLOv8 改进得到 CMYolov8。首先提出了 CSPHet 架构<sup>[3]</sup>, 该架构结合了 YOLOv8 的高效性与 HetConv 异构核卷积的特性, 显著减少了模型的计算复杂度和参数量, 适应嵌入式设备的部署。此外, 模型还结合了 FPN-PAN 结构与 MetaNeXt 模块<sup>[4]</sup>, 通过大核

卷积设计为 Inception 分支, 增强了位置信息的整合能力, 进一步提升了检测精度, 并降低了硬件资源的占用。

## 1 材料与方法

### 1.1 数据集



图 1: 数据集

在实验室设置 12 个规格为  $0.5 \times 0.35 \times 0.25$  的泡沫养殖箱。模拟实际场景下的单环刺螠养殖, 定期投喂, 换沙, 保持适宜温度、盐度, 并记录孔数变化。通过手机拍摄记录单环刺螠养殖周期内图片共 744 张, 如图 1 所示。数据采用 labelme 标注工具对图像进行标注, 并按照约 8:2 划分训练集和验证集。具体划分表 1 所示。

表 1: 数据集划分

类别	训练集	验证集	总共
单环刺螠洞口	597	147	744

### 1.2 CMYolov8

YOLOv8 是一种高效的目标检测网络, 具备卓越的实时检测能力和准确性。我们在 YOLOv8 中整合了两个先进的模块——CSPHet (图 2) 和 MetaNeXtStage, 得到了 CMYolov8 以期望在减少模型参数量的同时进一步

提升模型的检测性能。

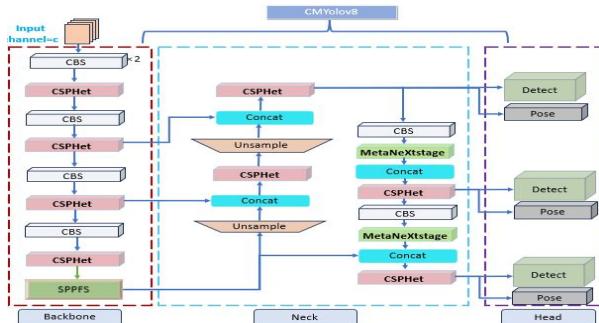


图 2: Yolov8 网络结果

### 1.2.1 CSPNet

本研究提出了一种新颖的深度学习架构 CSPNet (图 3)<sup>[5]</sup>, 该架构替代了 YOLOv8 中原有的 C2f 结构。CSPNet 借鉴了 HetConv 原理<sup>[6]</sup>, 利用异构核卷积 (HetConv) 在保持高表示效率的同时, 减少了计算量 (FLOPs) 和参数数量。

该模型首先通过一个  $1 \times 1$  卷积对输入进行预处理, 将特征分成两个分支。一个分支直接传递原始特征, 以保留基本信息, 而另一个分支则通过两个 HetConv 层进行处理, 用于有效地学习和提取特征中的重要信息。使模型能够在不同层次上进行深入的特征提取和融合, 从而提高模型的表达能力和泛化能力。

随后, 模型引入了多个 CSPNet\_Bottleneck 模块, 每个模块内部包含两个 HetConv 层, 用于进一步深化特征表示。这些模块的设计使模型能够更好地理解图像的语义信息和上下文关系。

最后, 通过  $1 \times 1$  卷积对处理后的特征进行整合和压缩, 以产生最终的输出。

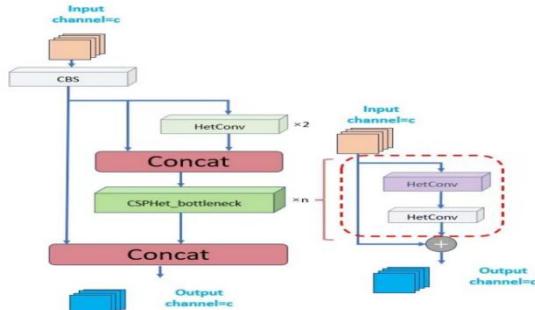


图 3: CSPNet

### 1.2.2 MetaNeXt

YOLOv8 的 Neck 部分采用了 FPN-PAN<sup>[7]</sup>双金字塔结构来融合不同尺度的特征信息。FPN 通过自顶向下的方式传播信息流, 将通过深层卷积得到的强语义特征用于增强浅层特征。然而, 随着网络层数的加深, 特征图的尺寸逐渐被压缩, 这导致位置信息的损失。因此, APN 采用了自下而上的设计, 通过利用浅层特征中的位置信息来补充深层特征中丧失的位置信息。然而, 由于 A

PN 在经过 FPN 增强其语义信息后, 位置信息已经被减弱, 从而在补充位置信息时的能力也受到削弱。

本研究提出了一种简单高效的 CNN 特征融合模块 ConvNeXt。在每个 ConvNeXt 块中, 输入 X 首先通过深度卷积进行处理, 以沿空间维度传播信息。虽然这种深度卷积操作只消耗少量的 FLOPs, 但由于高内存访问成本, 它在高性能计算设备上的效率显著降低<sup>[8,9]</sup>。因此, 为了在保持性能的同时加速大核卷积 CNN 模型的训练, 本研究借鉴了 Transformer (ViT) 中的 Inception 方法, 将  $7 \times 7$  大核卷积分解为沿通道维度的四个并行分支, 包括残差链接以保留原始的特征信息、两个正交卷积核分别处理两个维度上的特征信息以及一个  $3 \times 3$  的卷积模块来提取空间特征。基于这种新的 Inception 深度卷积, 我们构建了 MetaNeXt 模块, 如图 4 所示, 并将其放置在 APN 的卷积之后, 以提升对位置信息的整合能力。

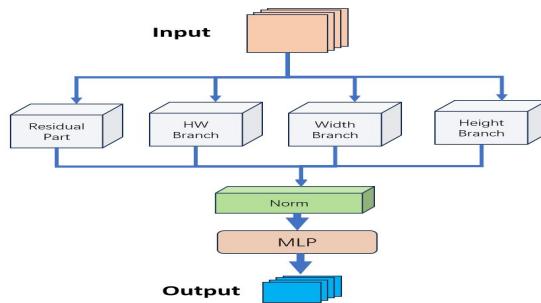


图 4: MetaNeXt

## 1.3 实验环境

实验在一台配备 AMD Core 7945HX 处理器、Nvidia RTX4060 图形处理器 (8 GB 内存) 的电脑上进行。深度学习框架为 PyTorch 12.1, 编程平台为 PyCharm, 编程语言为 Python 3.9. 所有比较算法在相同的环境下进行。

## 1.4 参数设置

实验参数设置如表 2 所示

表 2 参数设置

参数设置	
Total epoch	350
Batch size	32
Initial learning rate	0.01
Optimizer	Adam
Momentum	0.937
Weight decay	0.0005
Warmup epochs	3.0
Image size	640×640

## 1.5 评价指标

本研究提出了一种单环刺螠洞口的识别模型。分别使用平均精度均值 (mAP) 和准确率作为评估指标，同时为了评估模型的复杂性，我们使用浮点运算量 (FLOPs) 和参数量 (Params) 作为评估指标。AP 的计算公式详见公式(1-3)，准确率的公式详见公式(4)。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \#(1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \#(2)$$

$$\text{AP} = \int_0^1 (\text{Precision} \times \text{Recall}) \, d\text{r} \#(3)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \#(4)$$

这些公式中的参数定义如下：

TP：真正例数量，即模型将实际为正的样本正确分类为正的样本数量。

FP：假正例数量，即模型将实际为负的样本错误分类为正的样本数量。

TN：真负例数量，即模型将实际为负的样本正确分类为负的样本数量。

FN：假负例数量，即模型将实际为正的样本错误分类为负的样本数量。

## 2 实验

### 2.1 检测结果

从图5可以看出本研究所提出的模型能够准确的检测到孔洞

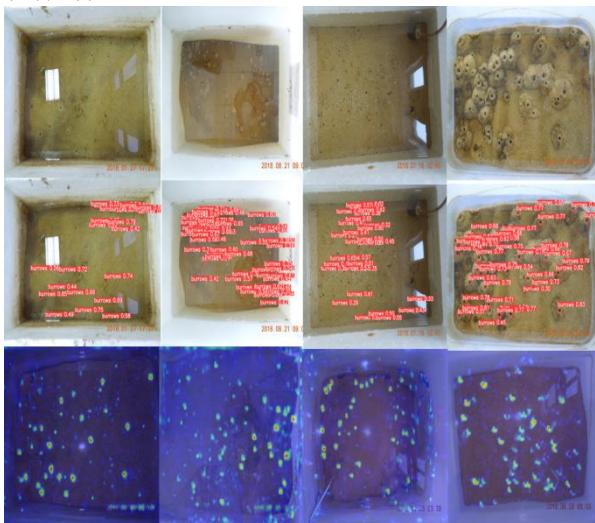


图5：第一行为原始图像，第二行为经过 CMYOLOv8 检测后的输出结果

### 2.2 消融实验

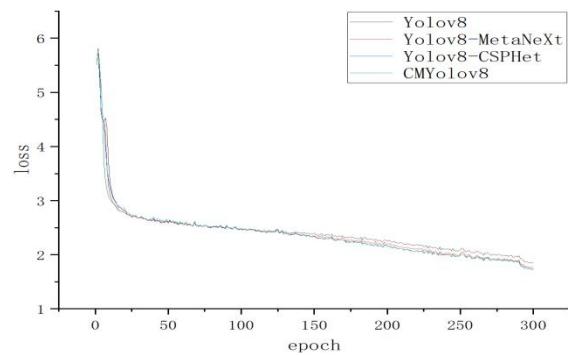


图6 Loss 训练曲线

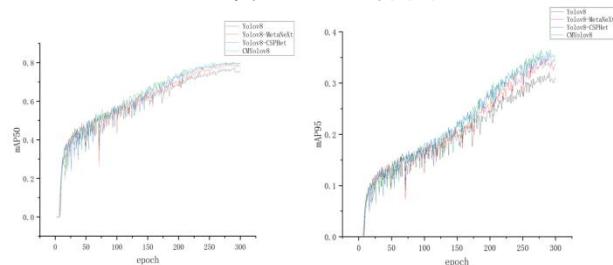


图7 mAP50 训练曲线和 mAP95 训练曲线

表3：改机模块消融实验

模型	Params	Gflops	mAP50	mAP95
Yolov8 (baseline)	2789091	6.9	0.761	0.310
Yolov8-CSP	2156835	5.3	0.798	0.347
Het	(-11.9%)	(-23.2%)	(+4.9%)	(+10.7%)
Yolov8-	2955283	7.2	0.786	0.335
MetaNeXt	(-5.9%)	(+4.3%)	(+3.3%)	(+8.1%)
CMYOLOv8 (ours)	2323027	5.6	0.812	0.353
	(-16.7%)	(-18.8%)	(+6.7%)	(+13.9%)

为了验证我们提出的 CSPHet 和 MetaNeXt 改进模型的有效性，设计了对比实验，使用参数量、计算量 (GFLOPs)、mAP50 和 mAP95 作为评估指标，其中 mAP 的 IoU 阈值为 0.5。

根据图6所示，所有模型的损失函数都可以在50轮左右快速收敛，而我们改进之后的 CMYOLOv8 模型取得了最低的损失基于 YOLOv8n 的基线模型 (YOLOv8n) 进行对比分析。从图7可以看出，CSPHet 模块的引入显著减少了参数量 (-11.9%) 和计算量 (-23.2%)，同时在 mAP50 和 mAP95 上分别提升了 4.9% 和 10.7%。MetaNeXt 模块通过大核卷积和 Inception 设计，增强了位置信息的整合能力，使得计算量略有增加 (+4.3%)，但 mAP50 和 mAP95 的提升幅度也分别达到 3.3% 和 8.1%。最终，如表3所示，我们提出的综合模型 (CMYOLOv8) 在进一步减少参数量 (-16.7%) 和计算量 (-18.8%) 的同时，mAP50 提升了 6.7%，mAP95 提升了 13.9%，证明了改进

方案在性能和轻量化方面的优越性。

### 3 讨论

本研究的实验结果展示了CMYOlov8模型在参数量、计算量和精度上的显著改进，验证了CSPNet和MetaNeXt模块的有效性。相较于原始YOLOv8模型，CMYOlov8不仅在轻量化上取得了显著进展，还在检测精度方面取得了显著提升。这种双重优化使模型在嵌入式设备上的部署变得更加现实。

然而，尽管模型在实验室环境中表现出色，未来的研究需要进一步验证其在复杂环境中的适应性。例如，在水下光线变化、遮挡以及洞口周围干扰物较多的情况下，模型的鲁棒性仍需优化。

### 4 总结

本研究提出了一种针对单环刺螠自动采捕和产量预测需求的深度学习洞口识别方法。通过将CSPNet和MetaNeXt模块引入YOLOv8框架，改进了模型的结构以提升其检测精度和轻量化性能。CSPNet模块使用异构核卷积大幅减少了参数量和计算量，同时保持高效的特征表示能力。MetaNeXt模块通过引入大核卷积与Inception设计，增强了模型对位置信息的整合能力。实验结果表明，改进后的CMYOlov8模型在参数量减少16.7%和计算量减少18.8%的同时，mAP50提升了6.7%，mAP95提升了13.9%，证明了模型的高效性与轻量化。

### 参考文献

- [1]曹梅,辛益,陈吉圣,王兴强,梁笛,韦寿永,孙美超,徐明,于润川.单环刺螠特征特性及养殖技术[J].现代农业科技,2021, No. 802(20):178-180.
- [2]大连海洋大学.一种海肠采捕船:CN111528192A[P].2020-08-14.
- [3]Wang, C.-Y.; Mark Liao, H.-Y.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W.; Yeh, I-Hau. CSPNet: A New Backbone That Can Enhance Learning Capability of CNN. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 2020. <https://doi.org/10.1109/cvprw5049.2020.00203>.
- [4]W. Yu, P. Zhou, S. Yan and X. Wang, "InceptionNeXt: When Inception Meets ConvNeXt," 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2024, pp. 5672-5683, doi: 10.1109/CVPR52733.2024.00542. keywords: {Performance evaluation; Training; Degradation; Costs; Computational modeling; Memory management; Network architecture; CNN; convolution; efficient neural networks},
- [5] Liu, S., & Wang, Z. (2020). "CSPNet: A New Backbone that Can Enhance Learning Capability of CNN." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3907-3915. DOI: 10.1109/CVPR42600.2020.00397.
- [6] Singh, P.; Vinay Kumar Verma; Rai, P.; Namboodiri, V. P. HetConv: Heterogeneous Kernel-Based Convolutions for Deep CNNs. Pure (University of Bath) 2019. <https://doi.org/10.1109/cvpr.2019.00497>.
- [7] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). "Feature Pyramid Networks for Object Detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2117-2125. DOI: 10.1109/CVPR.2017.106.
- [8] Xie, S., & Girshick, R. (2017). "Exploring Spatial Pyramid Pooling for Object Detection." IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4), 820-834. DOI: 10.1109/TPAMI.2016.2552714
- [9] Tan, M., & Le, Q. V. (2020). "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 6105-6114. DOI: 10.1109/CVPR42600.2020.00615

作者简介：尹骏杨（2002-），女，汉族，山东泰安人，本科在读，研究方向：水产养殖。

闵莹（2002-），女，汉族，山东枣庄人，本科在读，研究方向：水产养殖；尹延峰（1977-），男，汉族，山东泰安人，本科，研究方向：农业学。

通讯作者：纪元（1981-），女，汉族，湖北武汉人，博士学历，研究方向：特种水产养殖。