

## 摘 要

随着对海洋水下资源开发及环境探索的需求水平不断提高，水下目标识别技术引起越来越多研究者的重视。然而，水下环境的特殊性带来了一系列挑战，包括数据采集的困难、图像质量的低下等，这些都限制了水下目标识别技术的进步。为了克服这些障碍，研究者们提出了一系列针对水下场景优化的深度学习模型，以期提高识别的准确性和鲁棒性。本文通过研究 YOLOv7 网络，实现对水下目标的识别，并通过多种方法改善模型的识别性能。

首先，本文介绍水下目标识别的背景和当前研究情况。接着，着重介绍基于深度学习的水下目标识别方法，重点解释其中关键的技术——卷积神经网络。同时也讨论了该类方法的分类和评价指标。本文采用两个模块完成识别任务：YOLOv7 目标识别模块与图像增强模块。在目标识别模块，先介绍了基于 YOLOv7 的目标识别网络，为了优化网络性能，本文将卷积块注意力机制（Convolutional Block Attention Module, CBAM）与 Wise-IoU 损失函数融合进原网络中，并通过实验说明了这些改进的有效性。在图像增强模块，本文采用了一种基于 Inception 网络与注意力机制的方法，使用了有利于目标识别的增强图像作为数据集中的高质量图像，主要通过校正图像颜色来进行图像增强。应用图像增强模块后，提高了原 YOLOv7 网络的识别效果。最后，本文得到了一个引入图像增强的改进 YOLOv7 目标识别系统，通过两阶段的训练，测试与实验结果对比，在本实验所用数据集上，该策略下的模型较原 YOLOv7 模型的 mAP 值提高了 2.2%。

**关键词：**水下目标识别；YOLOv7 网络；注意力机制；图像增强

## ABSTRACT

As the demand for the development of underwater resources and environmental exploration in the ocean increases, underwater target recognition technology has attracted more and more attention from researchers. However, the particularity of the underwater environment brings a series of challenges, including difficulties in data collection and poor image quality, all of which limit the progress of underwater target recognition technology. To overcome these obstacles, researchers have proposed a series of deep learning models optimized for underwater scenarios, in the hope of improving the accuracy and robustness of recognition. This paper studies the YOLOv7 network to achieve recognition of underwater targets, and improves the recognition performance of the model through various methods.

First, this paper introduces the background and current research status of underwater target recognition. Then, it focuses on introducing the underwater target recognition methods based on deep learning, with a particular emphasis on explaining the key technology—convolutional neural networks. Additionally, the classification and evaluation metrics of this method are also discussed. This paper uses two modules to complete the recognition task, including the YOLOv7 target recognition module and the image enhancement module. In the target recognition module, the target recognition network based on YOLOv7 is introduced. To optimize network performance, the paper integrates the convolutional block attention module and Wise-IoU loss function into the original network and demonstrates the effectiveness of these improvements through experiments. In the image enhancement module, a method based on the Inception network and attention mechanism is used, utilizing enhanced images favorable for target recognition as high-quality images in the dataset, mainly through correcting image colors to enhance images. After applying the image enhancement module, the recognition effect of the original YOLOv7 network is improved. Finally, the paper developed an enhanced YOLOv7 target recognition system with image enhancement. Through two-stage training, testing and comparison, the model under this strategy improved the mAP value by 2.2% compared to the original YOLOv7 model on the dataset used in this experiment.

**Keywords:** underwater target recognition; YOLOv7 network; attention mechanism;  
image enhancement

# 目 录

第一章 绪论 .....	1
1.1 研究背景及意义 .....	1
1.2 国内外研究现状 .....	1
1.2.1 传统水下目标识别方法 .....	2
1.2.2 基于深度学习的水下目标识别方法 .....	2
1.3 本文主要研究内容和结构安排 .....	3
1.3.1 主要研究内容 .....	3
1.3.2 本文结构安排 .....	3
第二章 神经网络与目标识别相关理论 .....	5
2.1 卷积神经网络 .....	5
2.2.1 卷积层 .....	5
2.2.2 池化层 .....	6
2.2.3 激活函数 .....	6
2.2.4 损失函数 .....	7
2.2 基于深度学习的目标识别算法 .....	8
2.2.1 两阶段检测算法 .....	8
2.2.2 单阶段检测算法 .....	9
2.3 目标识别的相关评价指标 .....	9
2.3.1 IoU .....	9
2.3.2 精确率与召回率 .....	10
2.3.3 平均精度 .....	11
2.3.4 平均精度均值 .....	11
2.4 本章小结 .....	12
第三章 基于 YOLOv7 算法的水下图像目标识别方法 .....	13
3.1 YOLOv7 网络结构 .....	13
3.1.1 Backbone 层 .....	14
3.1.2 Neck 层 .....	15
3.1.3 Head 层 .....	15

---

3.2 算法改进 .....	15
3.2.1 基于注意力机制的网络结构改进 .....	16
3.2.2 坐标损失函数改进 .....	19
3.3 实验结果分析 .....	21
3.3.1 数据集准备 .....	21
3.3.2 实验环境及参数设置 .....	22
3.3.3 融合注意力机制的对比实验结果 .....	22
3.3.4 消融实验 .....	23
3.4 本章小结 .....	25
<b>第四章 基于神经网络的水下图像增强方法 .....</b>	<b>26</b>
4.1 图像增强流程 .....	26
4.1.1 通道级别的颜色校正 .....	27
4.1.2 全局颜色校正 .....	27
4.1.3 数据降维输出 .....	27
4.2 网络结构 .....	27
4.2.1 Inception 模块 .....	27
4.2.2 CAM 模块 .....	28
4.3 实验及结果 .....	29
4.3.1 数据集准备 .....	29
4.3.2 实验结果分析 .....	30
4.4 本章小结 .....	31
<b>第五章 引入图像增强的改进 YOLOv7 水下目标识别系统 .....</b>	<b>32</b>
5.1 整体网络设计与训练流程 .....	32
5.1.1 网络结构 .....	32
5.1.2 训练流程 .....	32
5.2 实验结果 .....	33
5.2.1 实验结果分析及对比 .....	33
5.2.2 消融实验 .....	37
5.3 可视化界面设计 .....	39
5.4 本章小结 .....	42

总结与展望 .....	43
-------------	----

参考文献 .....	45
------------	----

致谢 .....	48
----------	----

# 第一章 绪论

## 1.1 研究背景及意义

海洋作为地球上最大的生态系统，对人类社会和自然环境的发展都起着至关重要的作用<sup>[1]</sup>。它不仅在调节全球气候、维持生物多样性方面发挥着关键作用，而且蕴含着丰富的矿产、能源和生物资源，是人类可持续发展的重要物质基础。随着对海洋开发需求的不断扩大，利用程度不断加深，如何准确、快速、稳定地识别水下目标显得十分迫切与重要。

过去的海洋水下目标观测需要派遣潜水员进行水下作业，对水下目标图像的采集需要相关研究者手动完成，这些方式均比较低效且危险<sup>[2]</sup>。而利用载人潜水艇收集数据则有成本太高的问题。随着科技的发展，人类逐渐开始使用自动化无人装备进行水下作业，无人水下设备离不开水下目标识别系统，只有通过定位水下目标，设备才能完成任务。

传统的水下目标识别主要使用水下声呐技术，虽然声呐技术具有较好的穿透能力，但其分辨率有限，且容易受到环境噪声的影响。相较于声呐技术，光学图像具有更不易受到干扰，分辨率高，采集更加迅速，所占空间内存较少与包含信息更加丰富等优势，具有更广泛的应用前景。

在这一背景下，研究和开发先进的基于光学图像的水下目标识别算法，利用机器学习、深度学习等人工智能技术，提高对水下图像中目标特征的提取和识别能力，已成为海洋科学和工程领域的研究热点。这些研究不仅能够推动海洋资源的高效开发利用，还能为海洋环境保护、海洋灾害预警、水下考古、水下工程监测等领域提供强有力的技术支撑。

## 1.2 国内外研究现状

随着设备计算能力的不断提升，深度学习方法的提出极大地改变了各个领域任务的性能和效率。在水下目标识别领域，这一变革同样显著。本节以此为界，分别介绍了传统的水下目标识别方法和基于深度学习的目标检测算法。

### 1.2.1 传统水下目标识别方法

传统的目标识别方法主要集中在利用声纳或光学图像中提取的手工设计特征并结合传统的机器学习算法进行目标分类。这些特征可以是形状、纹理、颜色等方面的特征，常用的算法包括支持向量机、K 近邻算法等。这些方法的优点在于对特定领域有较强的解释性，但需要人工设计特征，难以适应复杂的环境变化和目标形态多样性。

Olmos<sup>[3]</sup>提出了一个用于检测无约束的海底视频中的人为制造物体的系统，由于水下图像中纹理和颜色等特征不易提取，该方法选定轮廓为识别特征，首先自动确定提取轮廓的最优尺度，接着提取轮廓特征，并使用贝叶斯分类器作为分类器判断当前物体是否属于目标；Mehdi Fatan<sup>[4]</sup>等提取图像的边缘特征，使用多层感知器和支持向量机利用纹理信息对特征进行分类，并运用霍夫变换进行水下电缆探测；Tri Susanto 等人<sup>[5]</sup>使用了 RGB 和 HSV 颜色检测作为参考，提出了一种基于颜色特征的水下目标检测方法。

### 1.2.2 基于深度学习的水下目标识别方法

随着深度学习技术的发展，特别是卷积神经网络（CNN）的成功应用，研究者开始尝试利用深度学习方法进行水下目标识别。凭借越来越先进的计算资源，这些方法通过直接自动从原始数据中学习特征表示和目标分类，大大提升了识别的准确率和速度。

Li X 等先基于 Fast R-CNN<sup>[6]</sup>，在 ImageCLEF 鱼类图像数据集上取得良好效果，mAP 值可达 81.4%，单张图像的检测效果比 R-CNN 快近 80 倍；后又基于 Faster R-CNN<sup>[7]</sup>，采用一个区域建议网络（RPN）生成高质量的目标候选对象，并与原始检测网络共享这些网络来实现检测加速，在同一数据集上的 mAP 值再次提高了 3.9%，检测速度也更快了；Zhang<sup>[8]</sup>基于 SSD 算法做出两点改进：引入注意力模块和进行多尺度特征融合，提出 MFFSSD 水下目标识别算法，通过注意力模块从通道和空间两方面突出有用信息，多种卷积方法提取融合了不同层次的特征，克服了水下图像中常见的物体较小且模糊的问题；Hou C<sup>[9]</sup>等在骨干网络中加入基于 gnConv 的自注意子层 HorBlock 模块，并提出了基于遗传算法的超参数演化方法获得模型最优参数，大大提高了检测精度。



## 1.3 本文主要研究内容和结构安排

### 1.3.1 主要研究内容

本文的主要研究内容如下：

（1）通过 YOLOv7 网络完成水下目标识别任务。考虑到水下环境复杂，存在许多小尺寸，遮挡和重叠的问题，在网络中加入 CBAM 注意力机制，并考虑到水下数据集质量较差的问题，使用了新的损失函数 Wise-IoU。通过这两种方法对 YOLOv7 网络进行了改进，训练得到的新网络实现了水下目标识别性能的提升。

（2）针对水下图像的比例退化，不均匀的光衰减导致的低可见度和颜色信息丢失的问题，本文采用了基于 Inception 神经网络融合 CAM 注意力机制的方法进行水下图像增强。

（3）当前的图像增强算法多是以提升视觉效果为目标，在这个过程中可能会反而会导致有利于目标检测的图像信息丢失<sup>[10]</sup>，为了避免这个问题，在训练过程中，本文的图像增强模块使用了不同类型的数据集，由使用基于视觉标准的增强图像更改为有利于目标识别的增强图像，更改了原定的 Ground Truth，使原图向着能够提升最终识别效果的方向进行增强。

（4）两阶段地完成了图像增强与目标识别模型网络的调整与训练。首先训练完成了图像增强网络，并将已训练好的模型作为图片预处理的步骤，将图像经过增强后再输入到目标识别网络中，完成训练。最终得到的引入图像增强的 YOLOv7 网络实现了性能的提升。

### 1.3.2 本文结构安排

本文的具体章节安排如下：

第一章为绪论，主要阐述了水下目标识别的研究背景及意义，并从传统方法和基于深度学习的方法两个方面简要介绍了水下目标识别任务的发展与现状，最后概括说明了本文的主要研究内容。

第二章介绍了本文主要涉及的技术——卷积神经网络的相关知识，包括它的网络基本结构与思想，并对目标识别的两类方法和评价指标进行了介绍。

第三章阐述了基于 YOLOv7 网络的水下目标识别方法。首先分析了 YOLOv7

网络的架构与各模块的作用。接着，为了改善水下目标识别任务的性能，对原 YOLOv7 模型在特征提取与回归损失函数方面做了改进，并对该改进后的模型进行了实验结果的讨论和分析。

第四章介绍了本文用到的另一个模型：基于神经网络的图像增强模块。该部分从图像质量入手提高识别任务的效果。首先介绍了图像增强模型的设计思想与网络架构，然后在训练中，使用了一个全新的数据集，以有利于目标识别任务的高质量图像作为原图像的 Ground Truth，增强了对目标检测任务的针对性。最后实验验证了图像增强模块对于目标识别任务的改善效果。

第五章将上述两个模型整合，得到了引入图像增强的水下目标识别系统，深入讨论本文方法的实验结果，并通过对比与消融实验的方法对结果做了进一步的分析。最后展示了本系统的可视化界面。

最后一章是总结与展望。在总结部分，对本文的主要工作进行了概括；在展望部分，探讨了本文可继续探索的改进方向。

## 第二章 神经网络与目标识别相关理论

卷积神经网络（Conventional Neural Network, CNN）是基于深度学习的目标识别的重要工具之一。本节分别介绍了卷积神经网络及目标识别方法的相关理论知识，作为本文算法的基础。

### 2.1 卷积神经网络

尽管卷积神经网络的变体层出不穷，但卷积层、池化层和全连接层仍然是构建这些网络的基础元素。同时，激活函数和损失函数也作为网络的核心组件，赋予了网络模型学习和预测的能力。

#### 2.2.1 卷积层

卷积层在神经网络中的作用是提取特征，生成输入数据的特征图。每个卷积层通常包含多个卷积核，这些不同的卷积核可以学习到不同的特征，例如边缘、纹理、颜色等，使得网络能够理解和表示输入数据的结构和内容。

在图像处理中，由于图像是二维数据，所以卷积核也是二维的，通常是一个小的正方形矩阵，在卷积操作中，卷积核通过在输入数据上滑动，并计算输入数据与卷积核之间的点积，生成输出特征图。卷积核的尺寸、步长及是否填充等各种参数用于控制在所输出的特征图的大小，而卷积核的数量可以控制输出的深度。卷积核的不同决定了提取到的特征不同，因此如何设计各层中的卷积核是至关重要的问题。

在一般处理图像的卷积神经网络中，卷积操作涉及图像的宽度，高度和通道数，典型的卷积层往往呈现三维结构，如下图 2.1 所示，其中注意  $W^p$  中的  $p$  表示卷积核的个数，输出特征  $Y^p$  的  $p$  表示通道数，两者是相等的。

相比于全连接层，卷积层的优越性主要体现在参数共享的思想。在全连接层中，输入节点上的各维信息都会传递到后面的每一个节点中，虽然最大程度上保证每个节点都不会遗失该维度上的信息，但参数太多，导致网络的训练过程非常缓慢。基于图像的局部不变性，卷积网络中各通道中图像的不同位置使用的卷积核对应参数共享，从而降低了网络总参数量。

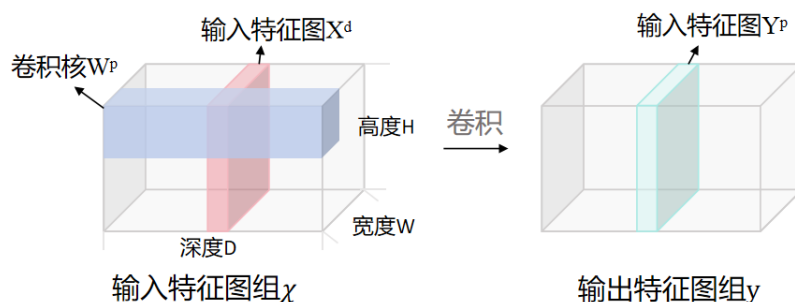


图 2.1 三维卷积层示意图

### 2.2.2 池化层

池化层通常跟在卷积层之后，用于减少特征图的空间维度，同时保留最重要的信息。池化的过程是一个移动窗口在输入矩阵上滑动，滑动过程中每个窗口按特定规则得到输出。池化层的作用在于减少特征图的尺寸，从而降低模型的参数数量和计算复杂度，并且有助于提取更加显著的特征。池化操作通常分为最大池化（Max Pooling）和平均池化（Average Pooling）两种常见类型，最大池化就是取窗口中数据矩阵的最大值，平均池化则是取平均值，如图 2.2 所示，对一张  $4 \times 4$  的特征图分别做步长为 2 的最大池化和平均池化的操作，卷积核的大小为  $2 \times 2$ ，可以得到不同的结果，从而提取到图像不同的特征，一般最大池化常用于提取图像的边缘信息，平均池化则可以将图像特征与背景区分开来。

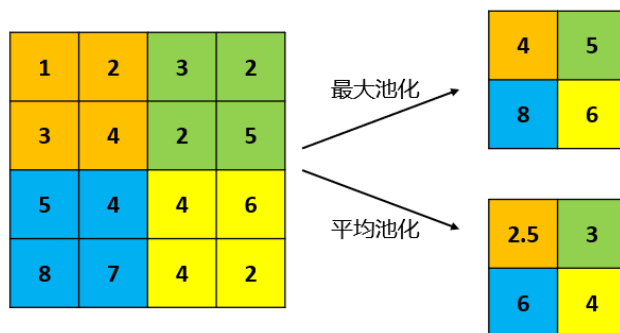


图 2.2 最大池化与平均池化过程示例

### 2.2.3 激活函数

一般情况下，激活函数应用在卷积层、全连接层等网络层的输出上，用于对前一层网络的线性输出进行非线性变换，为模型引入非线性的特性，从而使得网络具备对任意函数的表达能力，能适应更丰富，更复杂的任务。一般的激活函数

包括 ReLu 函数，Sigmoid 函数和 Softmax 函数等。

#### 2.2.4 损失函数

损失函数用于定量地计算期望的输出和模型的实际预测输出之间的差异，在模型训练过程中是模型进行学习的重要参考。神经网络通常都以最小化损失函数为目标得到最优的模型权重，损失函数的值越小，说明模型和数据的拟合效果越好。根据算法特点、应用场景等的不同，研究人员会设计不同的损失函数，常见的损失函数可以分为针对连续型变量的回归损失和针对离散型变量的分类损失，前者包括均方误差（MSE），Smooth L1 loss，各种 IoU 损失函数等，后者包括交叉熵损失函数，hinge 损失函数等。

##### （1）均方误差

均方误差（MSE）也称为 L2 Loss，对于 N 个样本，其定义如下：

$$MSE = \frac{1}{N} \sum_{i=1}^N (f(x_i) - y_i)^2 \quad (2.1)$$

其中  $f(x_i)$  表示样本的预测值， $y_i$  表示样本的真实值。取平方的操作避免了差出现正负相抵的情况。

##### （2）交叉熵损失函数

交叉熵损失函数可用于二分类和多分类的任务，其计算方法为：

$$CE = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log p_{i,c} \quad (2.2)$$

其中  $N$  表示样本总数， $C$  表示类别总数， $y_{i,c}$  表示第  $i$  个样本是否属于第  $c$  类，是则 1，否则取 0， $p_{i,c}$  表示第  $i$  个样本属于  $c$  类的预测概率。交叉熵函数的值也是越小，说明模型的拟合效果越好。

##### （3）IoU 损失函数

IoU 损失检测中重要的损失函数之一<sup>[9]</sup>可以用来计算预测边框与真实边框之间的差异，其中 IoU 的概念和计算方法会在后面小节进行详细介绍。在 IoU 损失提出之前，算法大都是基于目标边界框中的 4 个坐标点信息之间分别进行回归损失计算的，但由于同一边框信息之间肯定是具有相关性的，因此，IoU 将边界框的 4 个边界信息作为一个整体进行回归计算，显然更具有合理性，其计算方法为：

$$\text{IoU Loss} = -\ln \text{IoU} \quad (2.3)$$

很多情况下也定义为：

$$\text{IoU Loss} = 1 - \ln \text{IoU} \quad (2.4)$$

常见的，在计算出损失函数的值之后，网络会基于此计算不同参数的梯度，再根据该梯度寻找下一个能够使损失函数下降得最快的训练参数的值，用于下一轮训练。这个过程反复迭代进行，直到网络收敛，即模型达到较好的分类预测性能。

## 2.2 基于深度学习的目标识别算法

在目标识别领域，基于深度学习的方法已经得到了广泛的应用，同时也取得了良好的效果。常见的，基于算法思想不同，可以将基于深度学习的目标识别算法分为两阶段目标检测算法和单阶段目标检测算法。

### 2.2.1 两阶段检测算法

相比于单阶段检测算法，两阶段目标检测算法起步较早。它的算法流程可以分为两个阶段，第一阶段称为“区域建议”（Region Proposal），根据输入图像生成目标候选区域，下一阶段则是对这些候选区域进行分类和边界框回归。这类算法的代表算法有 RCNN,以及在它的基础上进行改进得到的 SPP-Net, Faster-RCNN 等。

以 R-CNN<sup>[1]</sup>为例，算法的大致流程如下，

（1）候选框提取：R-CNN 采用选择性搜索算法，提取 1000-2000 个可能包含物体的候选区域。而单阶段检测算法由于直接处理整幅图像，提取的候选框多达上万个。

（2）特征提取：将输入图片调整为相同大小后，通过卷积神经网络 AlexNet 进行特征提取，得到每个候选框的特征信息。

（3）SVM 分类：基于特征信息，使用 SVM 分类器完成目标分类任务。

（4）回归器精修：最后需要对候选框的位置进行精修，通过 NMS 去除多余的框，得到最终的检测结果。

两阶段目标检测算法由于先对图像的不同区域进行了筛选，可以起到将前景分开的作用，一定程度上解决了图片中的目标太少，背景信息太多，即正负样本不平衡的问题，通常具有更高的准确性，但速度较慢。

### 2.2.2 单阶段检测算法

单阶段目标检测算法将目标检测任务作为直接从图像中预测目标类别和边界框的回归问题。典型的单阶段目标检测算法包括 YOLO 系列（如 YOLOv3、YOLOv4）、SSD 等。

YOLO 系列是经典的单阶段目标检测算法，目前研究人员也在不断改进和扩大 YOLO 系列，目前已有 YOLOv9 问世。最初的 YOLO<sup>[12]</sup>网络的设计思想大致为：

- （1）输入图像，将其划分为  $S \times S$  的网格。
- （2）对于每个网格都预测  $B$  个边框，包括边界框的位置、尺寸、置信度以及边界框中的物体属于各个类别的概率。
- （3）根据上一步可以预测出  $S \times S \times B$  个目标窗口，然后需要根据阈值去除可能性比较低的目标窗口。

单阶段的方法实现了端到端的训练，速度快，可以满足实时性的应用需求，但参数量庞大，训练过程效率不高，最终的精确度较两阶段的方法可能普遍较低。

## 2.3 目标识别的相关评价指标

### 2.3.1 IoU

$\text{IoU}^{[13]}$  是一种衡量目标检测准确度的标准，用于过程中涉及预测范围的任务，可以得出真实框（Ground Truth）和预测框（Bounding Box）之间的重合程度从而判断预测结果是否可以接受。

$\text{IoU}$  的计算公式为：

$$\text{IoU} = \frac{S_{\text{交}}}{S_{\text{并}}} \quad (2.5)$$

可以理解为是真实框与预测框的重叠部分除以其并集部分得出的结果。

在算法实现过程中，会为  $\text{IoU}$  设定阈值，将每组预测框与真实框的  $\text{IoU}$  值与该值进行比较，通常在当前  $\text{IoU}$  大于设置阈值时，可以认为检测出现的预测框是“Positive”，即成功检测到目标。因此， $\text{IoU}$  的阈值对以下介绍的 TP、FN、FP、TN 等会产生较大影响。

### 2.3.2 精确率与召回率

#### （1）混淆矩阵

混淆矩阵是分类问题评价指标的基础。对于二分类模型，混淆矩阵如表 2.1 所示：

表 2.1 二分类混淆矩阵

混淆矩阵		预测类别	
		正类	负类
真实类别	正类	TP	FN
	负类	FP	TN

表 2.2 多分类混淆矩阵

混淆矩阵		预测类别		
		类别 1	类别 2	类别 3
真实类别	类别 1	C11	C12	C13
	类别 2	C21	C22	C23
	类别 3	C31	C32	C33

其中 TP（True Positive）为“真正例”，表示预测结果为正类的正类目标数量，FN（False Negative）为“假负例”，表示预测结果为负类的正类目标，FP（False Positive）为“假正例”，表示预测结果为正类的负类目标，TN（True Negative）为“真负例”，表示预测结果为负类的负类。

而本文所进行的水下目标检测任务是多目标的，因此，上述二分类下的混淆矩阵可拓展为表 2.2（以 3 个类别为例）。

#### （2）精确率（Precision, P）

精确率又称为查准率，反映出随机误差，即方差的大小，描述了实际值的扰动情况。具体来讲，它衡量了模型在所有被判定为正例的样本中，实际为正例的比例，二分类任务下精确率的计算公式为

$$P = \frac{TP}{TP+FP} \quad (2.6)$$

多分类下，每个类别的精确率计算公式为：



$$P_i = \frac{TP_i}{TP_i + FP_i} \quad (2.7)$$

其中  $i$  表示类别， $TP_i$  表示预测为第  $i$  类中真实类别也为第  $i$  类的“真正例”数量， $FP_i$  表示预测为  $i$  类中实际为其它类别的“假正例”数量。

精确率越高，模型的性能越好。

### （3）召回率（Recall, R）

召回率又称为查全率，反映出系统误差，即偏差的大小，描述了实际值与真实结果的偏离程度。它则衡量了模型在所有真实正例中，成功预测为正例的比例，二分类下的计算公式为

$$Recall = \frac{TP}{TP + FN} \quad (2.8)$$

多分类下，每个类别的召回率计算公式为：

$$R_i = \frac{TP_i}{TP_i + FN_i} \quad (2.9)$$

其中  $i$  表示类别， $FN_i$  表示实际为第  $i$  类，而被预测为其它类别的“假反例”数量。

召回率越高，模型的性能越好。

### 2.3.3 平均精度

精确率和召回率是一对矛盾的度量，一般来讲，其中一个指标的上升意味着另一个的下降。所以为了综合两者来衡量模型的性能，提出了 P-R 曲线，它以召回率为横轴，精确率为纵轴。P-R 曲线与坐标轴围成区域的面积越大，说明模型的性能越好。平均精度（Average Precision, AP）的值即为 P-R 曲线与坐标轴围成的面积。

### 2.3.4 平均精度均值

平均精度均值（mean Average Precision, mAP）是评估目标检测模型性能，尤其是多类别检测的一个重要指标，它是每个类别的 AP 的平均值。mAP 通常与不同的 IoU 阈值结合使用，以衡量模型在不同精度水平下的性能。mAP@.5 和 mAP@.5:.95 就是两种不同 IoU 阈值范围内的 mAP 计算方式，前者只考虑 IoU 大于或等于 0.5 的预测框，后者则更广泛地考虑了 IoU 的阈值从 0.5 到 0.95（步长为 0.05）上的平均 mAP 值。

mAP 的值越高，模型性能越好。

## 2.4 本章小结

本章首先介绍了卷积神经网络的基础知识，具体包括卷积层，池化层，激活函数及损失函数的原理。接下来，介绍了目标识别领域的理论知识，按照两阶段检测算法，单阶段检测算法的分类方法将目标识别算法分为两类，并分别简要阐述了每一类算法的思想与代表算法。最后介绍了目标识别领域常用的评价指标。

## 第三章 基于 YOLOv7 算法的水下图像目标识别方法

本节主要介绍了 YOLOv7 网络架构与本文在原网络基础上做出的改进，改进包括两方面，一是将卷积块注意力机制（Convolutional Block Attention Module, CBAM）融合进网络中，二是采用了 Wise-IoU 损失函数，最终分析了改进后的网络用于目标检测的实验结果。

### 3.1 YOLOv7 网络结构

YOLO 算法是单阶段目标检测方法，将目标检测任务视为一个回归问题，直接从图像像素到边界框坐标、类别概率的映射，可以实现端到端的训练。

YOLOv7<sup>[14]</sup>是 YOLO 系列网络的一个改进版本，它的主要网络结构可以分为 Backbone 层，Neck 层及 Head 层三个部分，在目标检测中分别发挥着不同的作用，整个网络的工作流程大致是：

（1）输入图像。

（2）特征提取：图像首先进入 Backbone 层，Backbone 层中的多种类型的卷积层以多尺度的方式提取图像特征。该层的输出是一系列特征图，这些特征图随后将被用于更精细的分析。

（3）特征优化：Neck 层的输入是 Backbone 层输出的特征图，在 Neck 层中，不同尺度的特征图可以进一步融合，以提炼到图像上下文信息等更丰富的特征。

（4）目标预测：Head 层根据之前网络提取到的特征，进行目标分类和定位。

YOLOv7 整体网络结构如图 3.1 所示，

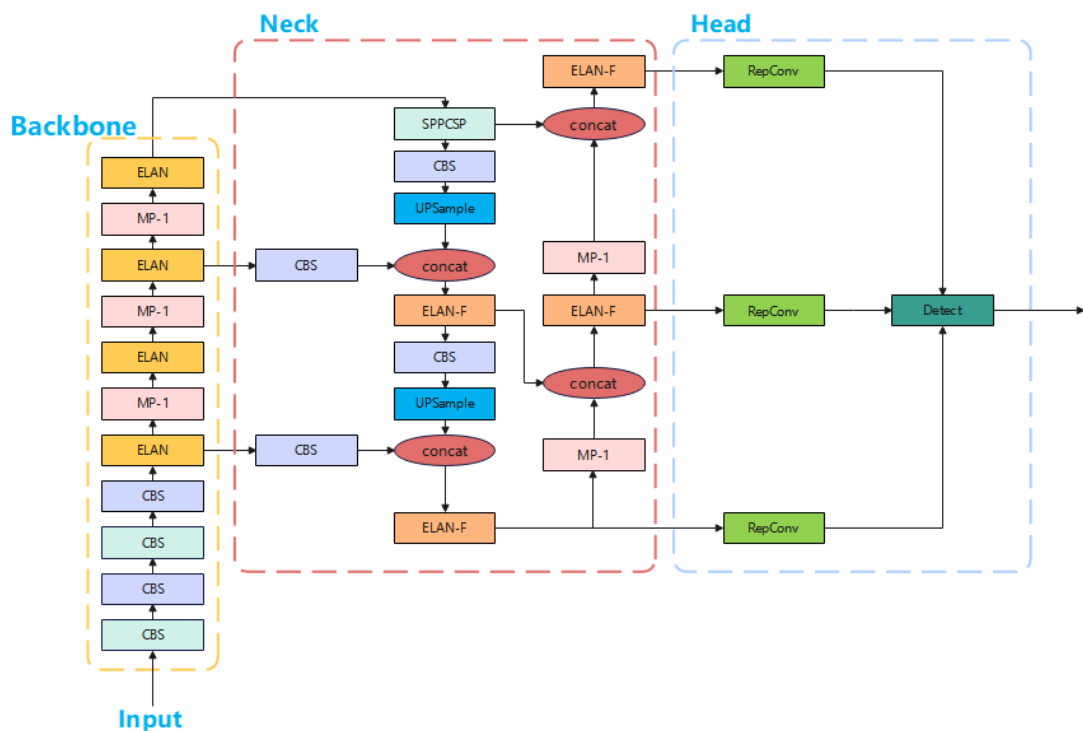


图 3.1 YOLOv7 网络结构

### 3.1.1 Backbone 层

Backbone 层是 YOLOv7 网络中用于特征提取的部分。它由 CBS 模块，扩展高效层聚合网络（ELAN）模块，多路径卷积（MP-1）模块组成。

#### （1）CBS 模块

该模块包括卷积，批量归一化和 ReLU 激活函数，可用于特征提取，并采用下采样操作来选择要提取的特征。

#### （2）ELAN 模块

它通过控制最长和最短两个梯度路径，使网络能够学习更多的特征，从而具有更好的泛化能力。它的一个分支通过  $1 \times 1$  卷积模块改变通道数，另一个分支还要再经过四个  $3 \times 3$  卷积模块进行特征提取后，引入残差结构来叠加特征。

#### （3）MP-1 模块

该模块将输入分为两部分，第一部分通过最大池化的操作进行下采样，然后通过  $1 \times 1$  卷积模块压缩通道；另一分支首先通过  $1 \times 1$  卷积压缩通道，然后使用步长为 2 的  $3 \times 3$  的卷积进行下采样操作。最终再将两个分支的结果叠加。

### 3.1.2 Neck 层

Neck 层是 YOLOv7 网络中用于特征融合的部分。它采用传统的 PAFPN 结构，将来自 Backbone 层的三个特征层的特征进行融合，通过结合不同尺度的特征，让模型具有更强的鲁棒性。

#### （1）SPPCSPC 模块

这个模块首先将特征分为两部分，一部分进行卷积的常规处理，另一部分进行 SPP 操作，即通过 1x1、5x5、9x9 和 13x13 四个不同尺寸的最大池化层，目的是获得不同的感知区域，帮助网络区分大小不同的目标。

#### （2）ELAN-F 模块

ELAN-F 与 Backbone 中的 ELAN 结构类似，但在第一分支中通过将每个输出部分相加来增加输出数量，这允许在更深层的网络结构中更有效地学习和收敛。

### 3.1.3 Head 层

Head 是 YOLOv7 网络中用于目标分类和边界框回归的部分。它使用增强的特征层来进行目标的分类和定位。

#### （1）IDetect 检测头

YOLOv7 选择了一个具有大、中、小三种目标尺度的 'IDetect' 检测头，用于将网络中间层提取到的特征转换成最终的边界框和类别预测结果。

#### （2）RepConv 模块

RepConv 引入了一个可重参数化的卷积结构的概念。在训练时，RepConv 使用三个分支：顶部分支是一个用于特征提取的 3x3 卷积，第二分支是一个用于特征平滑的 1x1 卷积，在输入和输出大小相等时添加一个 Identity 的残差结构，最后融合并求和这三个部分。在推理时，只有一个 3x3 卷积，它是从训练模块重参数化的。

## 3.2 算法改进

本实验采取了一系列创新的策略来提升 YOLOv7 网络在水下目标检测任务中的表现。首先，在 YOLOv7 的 Neck 层中巧妙地融合了 CBAM 注意力机制。CBAM 作为一种先进的注意力模型，能够有效地识别和强调图像中的关键特征，

同时抑制不重要的信息<sup>[15]</sup>。通过这种融合，显著增强了模型在特征融合阶段的能力，使得网络能够更加精确地捕捉到水下环境中目标的细节和轮廓。

此外，实验还对网络的损失函数进行了改进。传统的损失函数在处理水下图像时存在一定的局限性，尤其是在图像质量不佳的情况下，往往难以达到理想的检测效果。为了解决这一问题，本实验模型采用了 Wise-IoU 损失函数，替代了原有的损失函数。Wise-IoU 通过动态地调整不同质量图像之间的权重，优化了边界框回归的准确性，从而达到改善模型在水下环境中的检测性能的目的。

改进的 YOLOv7 网络结构如图 3.2 所示：

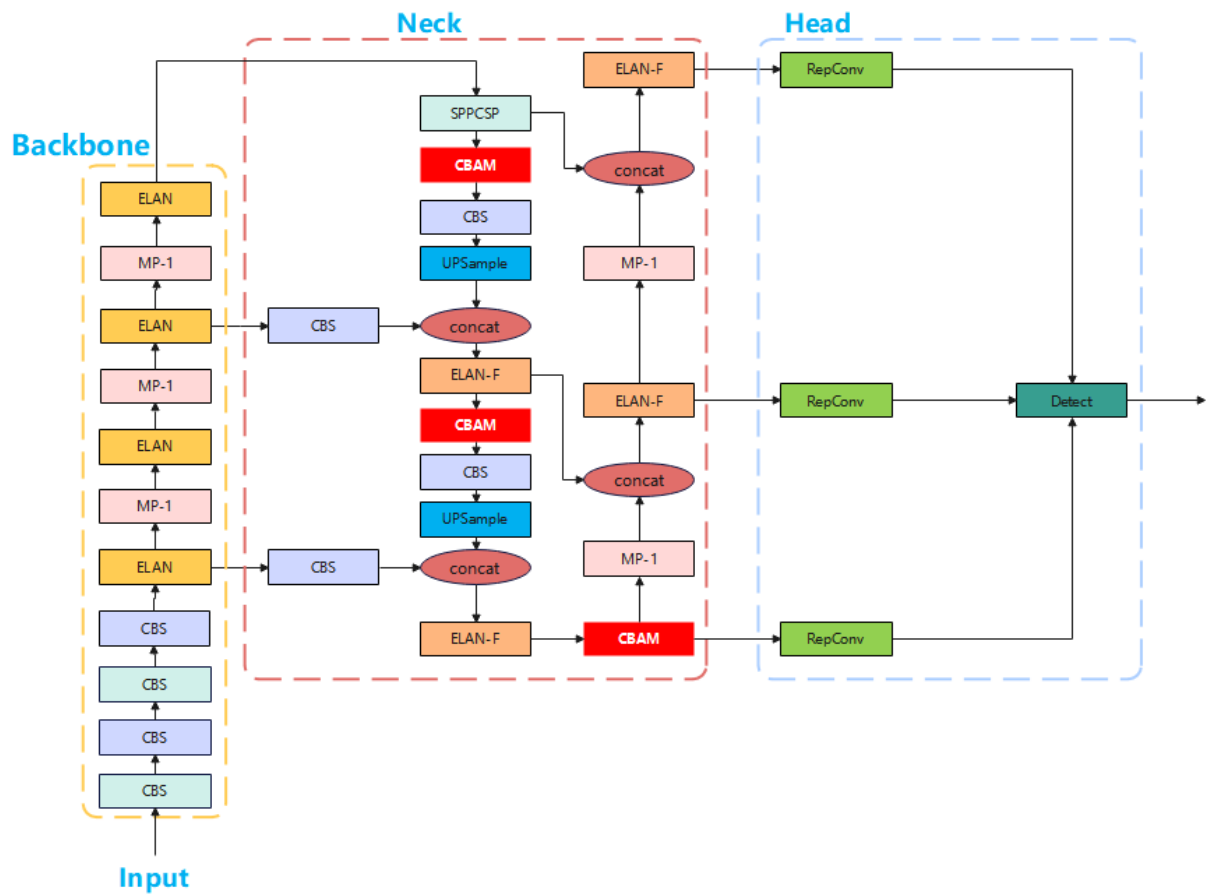


图 3.2 改进的 YOLOv7 网络结构

### 3.2.1 基于注意力机制的网络结构改进

#### (1) CBAM 注意力机制

注意力机制是一种在深度学习模型中模拟人类视觉注意力的技术，它的核心思想就是让模型从关注全部到关注重点，允许模型动态地聚焦于输入数据的特定部分。具体通过为输入图像的不同区域分配一个权重，即“注意力分数”，来实

现这一点，进而让机器根据评分将注意力放在待检测的目标及其周围区域，优化有限的计算能力下的资源分配。

常用的注意力机制包括通道注意力，空间注意力和混合域注意力机制，它们从不同方面入手提高模型检测性能。CBAM 是一种专为卷积神经网络设计的注意力模块，它结合了通道注意力和空间注意力<sup>[15]</sup>，是一种混合域的注意力机制给出参考文献。其整体流程如图 3.3 所示，

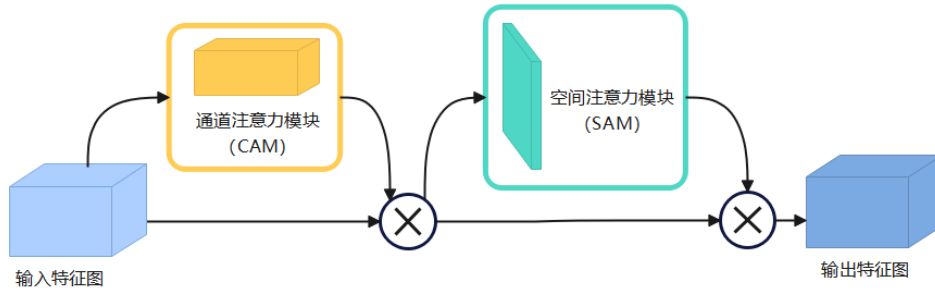


图 3.3 CBAM 注意力机制整体流程

大致流程为：

- （a）输入特征图：输入一张特征图  $F$ ；
- （b）通道注意力：先经过通道注意力模块（Channel Attention Module, CAM），得到基于通道注意力机制的通道注意力特征图  $M_c(F)$ ，再将其与输入特征图相乘，得到  $F'$ （这里的“相乘”指逐元素相乘）。
- （c）空间注意力：通道注意力通过空间注意力模块（Spatial Attention Module, SAM），得到相应的注意力特征图  $M_s(F')$ ，同样与该模块的输入特征图相乘，得到  $F''$ 。
- （d）输出特征图：输出经过两次注意力增强的特征图  $F''$ ，可能被送入下一个卷积层继续进行其它处理。

通道注意力用于选择图像中“什么”是有意义的，空间注意力用于选择图像中“哪里”是有意义的，两者能够从三个维度充分处理输入的特征图。结合上述流程，整个过程按照算式可以表示为：

$$F' = M_c(F) \otimes F \quad (3.1)$$

$$F'' = M_s(F') \otimes F' \quad (3.2)$$

其中 $\otimes$ 是逐元素相乘的计算方式。

CBAM 的实现主要在于其中的 CAM 模块和 SAM 模块两个子模块。

CAM 子模块的具体结构如图 3.4 所示，

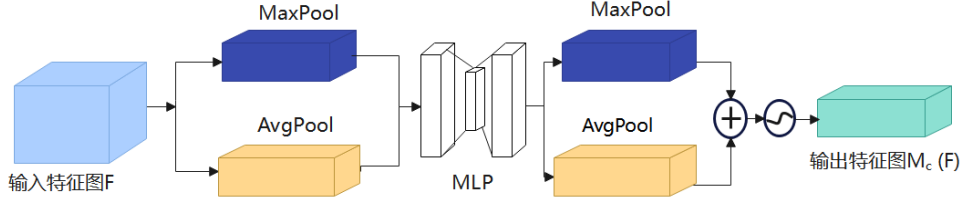


图 3.4 CAM 子模块结构

在该模块中，首先对输入大小为  $C \times H \times W$  的特征图进行平均池化和最大池化，对特征图进行空间维度的压缩，结果输入多层感知机(multi-layer perceptron, MLP)，融合这两种空间描述从而得到通道注意力权重，然后经过激活函数输出大小为  $C \times 1 \times 1$  通道注意力特征图，是一维的。 $M_c(F)$ 与  $F$  相乘后可以得到大小仍为  $C \times H \times W$  的特征图。该模块的计算公式如下，

$$M_c(F) = \sigma (MLP(AvgPool(F) + MaxPool(F))) \quad (3.3)$$

SAM 子模块的具体结构如图 3.5 所示，

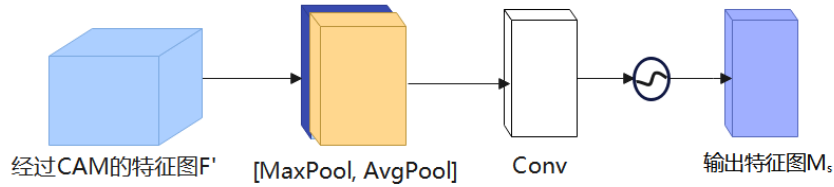


图 3.5 SAM 子模块结构

通道注意力处理完后将特征图输入到空间注意力机制模块，在该模块中，特征图进行基于通道的全局最大池化和全局平均池化，得到两个大小为  $1 \times H \times W$  的特征图，将两个结果进行拼接后通过卷积操作将维度降低到一个通道，即为  $1 \times H \times W$ ，通过激活生成空间注意力特征图。然后还需要与输入相乘得到输出结果，输出经过注意力增强的特征图大小仍为  $C \times H \times W$ 。该模块的计算公式为：

$$M_s(F) = \sigma (f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \quad (3.4)$$

## (2) 网络结构改进

CBAM 加在什么位置会带来最好的效果，这个问题根据水下目标检测任务



的多样性以及所使用的水下目标数据集的特性等，会得到不同的答案。因此，根据 YOLOv7 的网络结构，本文将 CBAM 分别集成到 Backbone 层，Neck 层和 Head 层，目的是为了探究在哪个部分引入注意力机制能够最大程度地提升本实验模型的检测性能，并将带来最优检测效果的模型作为本章最终采用的网络结构。

如图 3.6 左，将 CBAM 加入 BackBone 层，该层负责从输入图像提取初级特征，在这一层次集成 CBAM，可以帮助网络更有效地提取和筛选有用信息，为后续的检测任务打下坚实的基础。

如图 3.6 中，将 CBAM 加入 Neck 层，Neck 层在 YOLOv7 中扮演着特征融合的角色，它将不同尺度的特征图进行有效整合，以便于检测不同尺寸的目标。在这一层次引入 CBAM，可以进一步提升特征的表达能力，使网络更加敏感于水下环境中细微的目标变化。

如图 3.6 右，将 CBAM 加入 Head 层，Head 层作为网络的输出部分，负责对融合后的特征进行最终的分类和边界框回归。在这一层次应用 CBAM，可以优化决策过程，提高目标检测的准确性。

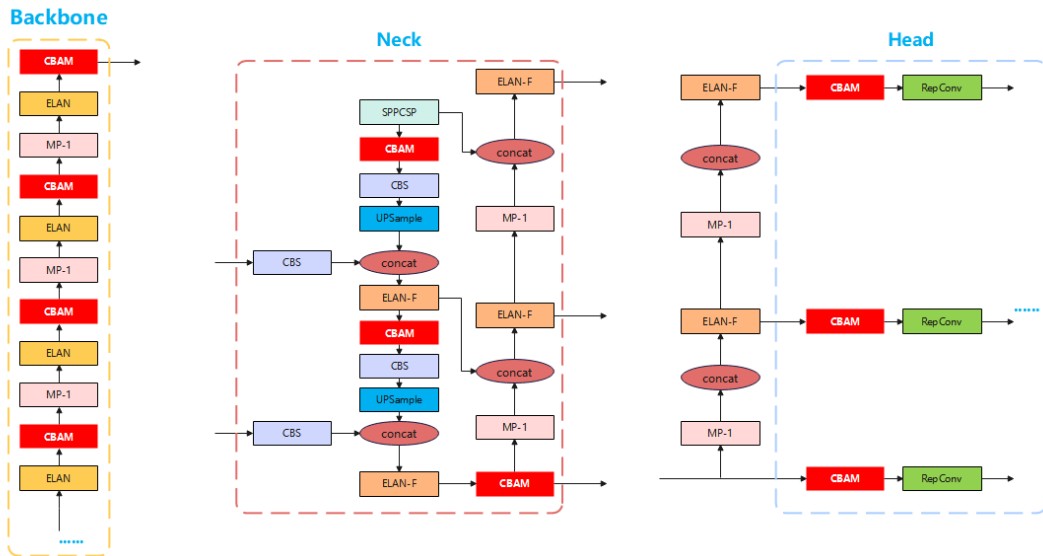


图 3.6 不同位置分别融合 CBAM 后的局部网络结构

### 3.2.2 坐标损失函数改进

在目标检测领域，损失函数的设置直接影响目标检测结果的准确性。YOLOv7 中使用了多种的损失函数，包括类别损失，目标置信度损失和坐标损失三部分，最后的损失函数结果是三者之和。其中类别损失和目标置信度损失均通

过交叉熵损失函数实现，坐标损失采用 CIOU 损失。CIOU 损失函数的计算方法大致为：

$$L_{CIOU} = 1 - IOU(A, B) + \rho^2(A_{ctr}, B_{ctr})/c^2 + \alpha v \quad (3.5)$$

$$v = \frac{4}{\pi^2} (\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h})^2, \alpha = \frac{v}{(1-IOU)+v} \quad (3.6)$$

公式中，A，B 代表两个框，A<sub>ctr</sub>，B<sub>ctr</sub> 代表 A 和 B 的中心点。相较于最基本的 IoU 损失函数，CIOU 引入了对检测框长宽比的考量，即公式中的  $\alpha v$ 。

考虑水下真实情况下采集到的数据集质量通常参差不齐，所以本文采用 Wise-IoU 作为新的坐标损失函数。Wise-IoU 的特性在于其能够依据检测框的质量差异，灵活地调整权重<sup>[16]</sup>。这种方法减少了高质检测框的竞争性，同时降低了劣质样本可能导致的负面梯度，使得 Wise-IoU 能够专注于中等质量的锚点框，从而增强了整个检测系统的性能，比较适合本实验的任务场景。

Wise-IoU 有三个版本，分别为 WIoU v1、WIoU v2 和 WIoU v3。在 WIoU v1 中，除了考虑不同锚框之间的差异，还认为当预测框与目标框重和较好时，应减少对锚框几何度量的关注，因此基于距离构造了两级注意力机制，其具体计算方法为：

$$L_{WIoU_{v1}} = R_{WIoU} L_{IoU} \quad (3.7)$$

$$R_{WIoU} = \exp\left(\frac{(x-x_{gt})^2 + (y-y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (3.8)$$

其中，x，y 是锚框的中心坐标，x<sub>gt</sub>，y<sub>gt</sub> 表示目标框的中心坐标，W<sub>g</sub>，H<sub>g</sub> 是目标框的长宽。R<sub>WIoU</sub> 可以显著放大普通质量锚框的 L<sub>IoU</sub>，L<sub>IoU</sub> 则可以显著降低高质量锚框的 R<sub>WIoU</sub>，并在上述情况下削弱对锚框与目标框中心距离的关注。其中，上标\*表示将 W<sub>g</sub> 与 H<sub>g</sub> 从计算图中进行分离，可避免影响收敛速度。

WIoU v2 基于 WIoU v1，参考 Focal Loss，引入单调聚焦机制，有效降低了简单示例对损失值的贡献。这使得模型能够聚焦于困难示例，获得分类性能的提升。

WIoU v3 在 WIoU v1 的基础上，引入非单调聚焦系数 r，以实现根据当前训练情况动态地调整梯度增益分配策略。首先，它通过定义“离群度” $\beta$  来描述检测框质量：

$$\beta = \frac{L_{IoU}^*}{L_{IoU}} \quad (3.9)$$

其中 $\overline{L_{IoU}}$ 是损失函数的滑动平均值，用于平稳地跟踪 IoU 损失函数的变化趋势，从而在优化过程中更加稳定地更新模型参数，它保证了训练过程中增益策略更新的动态性。 $\beta$  的值越小，说明锚框的质量越高。即在离群度过大或者过小时，为其分配一个较小的梯度增益，避免高低质量锚框的影响。

基于  $\beta$  构造非单调聚焦系数  $r$ ：

$$r = \frac{\beta}{\delta \cdot \alpha^{\beta-\delta}} \quad (3.10)$$

则 WIoU v3 的损失计算可定义为：

$$L_{WIoU_{v3}} = r L_{WIoU_{v1}} \quad (3.11)$$

其中  $\alpha$  与  $\delta$  为超参数，可以根据模型情况设置。当回归框的离群度满足  $\beta=C$ （ $C$  为预设值）时，回归框可以获得最高的梯度增益。由于 $\overline{L_{IoU}}$ 是动态的，锚框的质量划分标准也是动态的，这使 WIoU v3 能够及时调整当前梯度增益的分配策略，做出最优选择。

本实验使用了 WIoU v3 版本。

### 3.3 实验结果分析

#### 3.3.1 数据集准备

本文用于水下目标检测的数据集一共有 4377 张 bmp 格式的水下图片及其标注文件，按照 8: 2 的比例将其划分为训练集与验证集，即训练集包含 3501 张水下图片，验证集包含 876 张水下图片。数据集中包含 6 类目标，分别为 HumanDrivers, Wrecks/ruins, Fish/vertebrates, Plant/sea-grass, Robots, Fish\_group。

该数据集的标注使用 xml 格式，而 YOLO 模型需要 txt 格式的标注文件，因此需要先进行格式的转换。转换后，YOLO 标注文件的格式为：

`<object-class> <x> <y> <width> <height>`

其中，object-class 是类别的标签索引， $x$ ， $y$  表示目标框的中心坐标，width，height 表示目标检测框的宽和高， $x$ ， $y$ ，width，height 均已做归一化处理。以下是一个具体的标注实例：

0 0.8160 0.6129 0.2992 0.5468

### 3.3.2 实验环境及参数设置

本实验所使用语言为 python，使用 pytorch 深度学习框架进行模型的训练与测试，且本次实验均基于 AutoDL 算力云平台。本实验所使用 GPU 配置为 NVIDIA GeForce RTX 4090，具体环境配置为 python3.8+pytorch1.13.0+cuda117，并利用 numpy，opencv-python，pillow 等库。

AutoDL 算力云是一个专注于提供 GPU 云服务的平台，它允许用户按需租用 GPU 来进行深度学习和其他计算密集型任务，为本次实验提供了可用的 GPU 算力资源。

在模型的训练过程中，有一些超参数的设置也会影响结果，本实验模型在训练过程中的参数设置如表 3.1 所示：

表 3.1 模型参数设置

超参数	值
批处理量	32
训练轮数	100
初始学习率	0.001
图片大小	640×640

在批处理量的设置上，为了最大程度地利用 GPU 资源并且保证合适的收敛速度，最终选择为 32；在训练轮数上，按{50, 100, 150, 200}分别对原始 YOLOv7 网络进行训练，参考模型训练曲线的收敛情况，最后选择 100 为最大迭代次数。

本文所有实验始终使用相同实验环境及模型基本设置进行。

### 3.3.3 融合注意力机制的对比实验结果

表 3.2 展示了在 YOLOv7 中的不同位置加入注意力机制的实验结果，YOLO\_Bb、YOLO\_Nk、YOLO\_Hd 及 YOLO\_All 分别表示在 Backbone 层、Neck 层、Head 层及三层全部加入的四个模型，在不同层中添加 CBAM 机制的方式如图 3.6 所示。分析结果可得，在 Neck 层加入 CBAM 后，模型取得了最高的 mAP 值 61.4%，比原模型高 0.8%，精确率高于原模型 1.5%，召回率高于原模型 1.2%，且它的精确率是五个模型中最高的，召回率是五个模型中第二高的，综合效果最好。

Neck 层之所以能够获得最佳的嵌入结果，主要原因可能在于在 Neck 层中，CBAM 通过对不同维度的特征图进行加权，使得网络能够更加关注那些包含目标信息的特征，而忽略那些不太相关的特征。这样，网络不仅能够保留从底层提取到的丰富特征信息，还能够在高层进行有效的特征融合。这种有侧重的融合通过对特征进行重新组合和优化，让模型获得了更加细粒度的语义信息，从而获得最显著的效果。

四个改进模型都不同程度地提高了模型的召回率，这可能是因为 CBAM 通过注意力机制增强了模型对图像中关键特征的响应，这可能使得模型能够捕捉到之前未能注意到的细微目标，从而更多地检测到正确的目标物体，但相比于召回率的提升水平，精确率的提升水平显得稍低，CBAM 与 Backbone 层的融合甚至还使模型精确率降低了，可能是由于水下图片质量较低，背景和目标区分程度小，通过引入注意力机制只能在一定程度上影响误检率。

而在 YOLO\_All，即三个层都加入了 CBAM 的模型中，精确率稍有提升，但其对于召回率的贡献相较于其它改进模型反而下降了，甚至 mAP 值相比原模型出现了下降，说明 CBAM 机制并不是加得越多对于网络识别结果的改善越大。

结果表明，CBAM 在网络的不同部分所起的作用确实有较大差异。如何在网络中加入注意力机制的改进实现网络性能的提升，需要具体问题具体分析，通过实验验证不同改进策略的效果。

表 3.2 网络不同位置加入 CBAM 后模型的性能对比

模型	P	R	mAP@.5	mAP@.5:.95
YOLOv7	69.3%	55.7%	60.6%	40.3%
YOLO_Bb	68.2%	57.8%	60.6%	40.9%
YOLO_Nk	70.8%	56.9%	61.4%	41.3%
YOLO_Hd	70.5%	56.4%	60.7%	40.3%
YOLO_All	70.5%	55.8%	60.4%	40.3%

### 3.3.4 消融实验

由于在 Neck 层取得了最好的效果，本节最终使用将 CBAM 融入网络 Neck 层中的策略。另外，本实验还将 Wise-IoU 加入网络进行训练与测试。

为了验证不同改进方法对于水下目标识别结果的影响，本节通过消融实验进行了分析，其中 YOLOv7\_A 表示只在 Neck 层加入 CBAM，YOLOv7\_B 表示只将损失函数调整为 Wise-IoU，YOLOv7\_C 表示两者都加入，最终得到包括 YOLOv7 原模型在内的四个识别网络的性能指标数据，如下表 3.4 所示。

表 3.4 消融实验结果

序号	模型	CBAM	WIoU	P	R	mAP@.5	mAP@.5:.95
1	YOLOv7	×	×	69.3%	55.7%	60.6%	40.3%
2	YOLOv7_A	√	×	70.8%	56.9%	61.4%	41.3%
3	YOLOv7_B	×	√	70.5%	59.0%	61.5%	44.3%
4	YOLOv7_C	√	√	70.8%	59.2%	61.7%	43.7%

根据表 3.4，对比实验 1、4，融合了 CBAM 机制和 Wise-IoU 的模型取得了最高的 mAP 结果，比原模型提高了 1.1%，精确率提高了 1.5%，召回率提高了 3.5%，效果显著。同时，每个加入改进措施的模型都在原模型的性能上有所提升，说明了本实验两个改进策略的有效性，CBAM 的引入赋予了网络获取更多有价值特征进行融合的能力，Wise - IoU 可以使网络更专注于有效特征，并对不同质量的图像有更好的权重选择。

可以分别对比实验 1、3 与实验 2、4，对 Wise-IoU 对模型的效果进行分析。从精确率，使用 Wise - IoU 后，相对于原模型精确率提升较大，达到 1.2%，但相比于已加入 CBAM 的模型，精确率没有提升。而且结合实验 1、2，融合 CBAM 对模型的精确率提升效果可以达到 1.8%，优于 Wise-IoU。从召回率，可以发现 Wise-IoU 对召回率的提升显著，相比原模型可以提升 3.3%，相比于已加入 CBAM 的模型可以提升 2.3%。从 mAP，加入 Wise-IoU 后，模型的综合性能都得到了提升，说明通过改变损失函数，对不同质量的数据集有所侧重，能够优化模型的决策过程。值得一提的是，加入 Wise-IoU 的模型的 mAP@.5:.95 值也大大提升了，说明在其它较高的 IoU 阈值上，Wise-IoU 的提升效果很好，可能更有益于小目标检测，因此在后续模型的研究中，可以继续调整 Wise-IoU 的超参数  $\alpha$  与  $\delta$ ，并关注模型在不同 IoU 阈值上的表现。

结果表明，在 Neck 层加入 CBAM 注意力机制并改进损失函数为 Wise-IoU 确实能够改善水下目标识别模型的性能。

### 3.4 本章小结

本章介绍了 YOLOv7 网络的基础架构，对原网络实现了两方面的改进，一方面，引入 CBAM 注意力机制增强模型对特征的处理能力，分别在网络的 Backbone 层、Neck 层和 Head 层加入该机制，探究 CBAM 在网络中的不同位置对最终检测性能的影响；一方面，修改原损失函数，采用 Wise-IoU 这一更为先进的损失函数弥补水下数据集质量参差不齐的缺点。最终完成了多个模型的训练与实验，根据目标检测的指标，得出 YOLOv7+Neck 层 CBAM+Wise-IoU 为最优模型，实现了对原网络的水下目标识别效果的改善。

## 第四章 基于神经网络的水下图像增强方法

由于水下环境特殊，采集到的水下图像图片普遍存在模糊，噪声，色彩失真等问题，低质量的水下图像通常会造成特征信息不清晰，误导模型对噪声敏感等问题，影响目标检测算法的精度和泛化能力。一般的图像增强方法的思想包括：基于先验经验，增强图像的对比度或边缘信息等特征；基于物理模型，建模图片的退化过程从而实现对现有图像的增强；以及基于深度学习，使模型学习到从低质量图像到高质量图像的映射。

本文实验主要针对水下图像颜色失真的问题，具体包括颜色的比例退化和不均匀的光衰减问题<sup>[17]</sup>。颜色的比例退化主要影响图像中不同区域颜色的可见度比例，即距离相机较近的内容比距离相机较远的内容更清晰，这意味着水下图像中不同区域的颜色的退化速率不同。不均匀的光衰减导致的颜色失真指不同颜色的光波长不同，其在水中的衰减速度也不同，红色的溶解速度明显快于绿色和蓝色<sup>[18]</sup>。

考虑到这两个问题，需要从每个颜色通道提取不同比例的特征，采用自适应加权策略才能更合理地恢复图像特征<sup>[17]</sup>。

### 4.1 图像增强流程

图 4.1 展示了图像增强模块的整体增强流程与网络结构，图像增强流程可以分为三部分，分别为通道级别的颜色校正，全局颜色校正和数据降维输出，其中颜色校正均通过 Inception 网络模块+CAM 注意力模块实现。

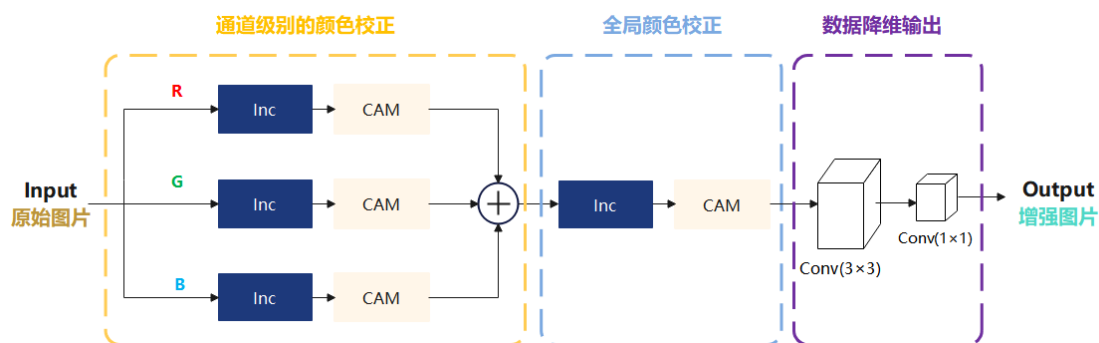


图 4.1 图像增强模型整体流程



#### 4.1.1 通道级别的颜色校正

由于颜色的退化程度是依赖通道的，所以该阶段分别从 R、G、B 三个通道中提取特征，分别自适应地学习图像增强的权重策略。在各通道中，网络继续处理了不同尺度下的比例退化问题，在这个过程中，先利用 Inception 模块提取图像不同尺度的特征，再将这些信息输入通道注意力模块<sup>[17]</sup>(Channel-wise Attention Module, CAM)，让模型识别哪些特征图对于颜色校正更为重要，从而为不同比例的退化分配不同的权重系数。

#### 4.1.2 全局颜色校正

在第二阶段，将提取的特征图串联后，从组合的颜色特征图中提取特征。这一阶段能够分析各通道间的关系以捕获更高层次的退化，CAM 模块让模型在信息丢失率更高的颜色通道对应的特征图上赋予了更多的权重。

#### 4.1.3 数据降维输出

经过前两步的处理后，图像的保持了较高的特征空间维度，包括高度、宽度和通道，为了将其转换到最终的仅含高度和宽度的二维 RGB 图像，在该部分首先使用了  $3 \times 3$  的卷积核来减少通道数。这种卷积操作有助于在减少输出特征图数量的同时，保护信息的完整性，接着通过激活函数调整特征图的输出值，使其拥有正确的强度值范围。

### 4.2 网络结构

本节分别介绍了颜色校正所使用的主要网络模块：Inception 模块和 CAM 模块的设计与工作机制。

#### 4.2.1 Inception 模块

Inception 模块作为一个多尺度特征提取器，利用不同大小的卷积核提取图像不同尺度的特征。如图 4.2 所示，其中包含  $1 \times 1$ ， $3 \times 3$  和  $5 \times 5$  且步长为 1 的卷积核，使用更大的卷积核，能提取到更为细粒度的信息。另外，模块中的最大池化层用于提取图像的边缘信息。最后将三种大小的卷积核与池化层提取到的四个特征集拼接起来得到 Inception 模块的输出结果。

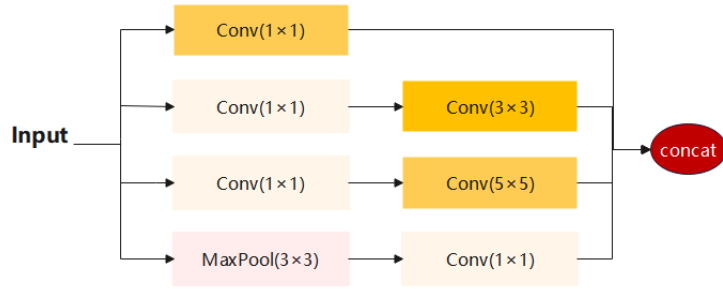


图 4.2 Inception 模块结构

#### 4.2.2 CAM 模块

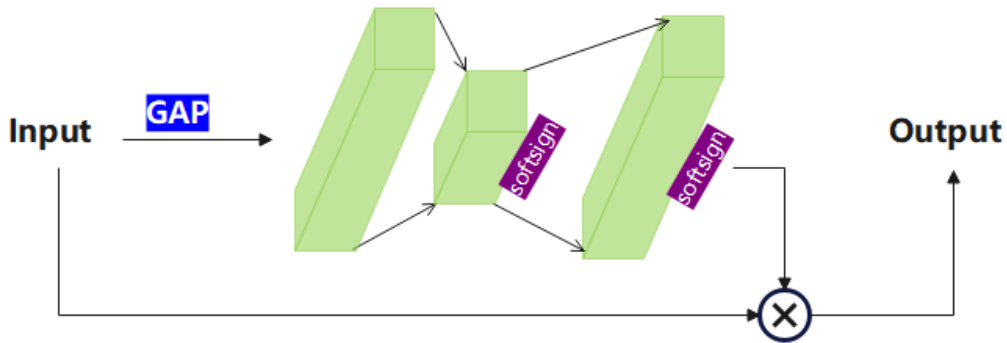


图 4.3 CAM 模块结构

增强网络中的 CAM 模块与 CBAM 中的 CAM 模块不同，虽然它们均基于通道学习增强特征表示，但本节中的 CAM 模块在生成权重的策略上在原 CAM 模块的基础上做了改动，引入了 Excitation 策略，以适应水下图像增强的任务。本节 CAM 模块的结构如图 4.3 所示。

CAM 模块首先使用全局平均池化( GAP ) 将每个特征图在其空间维度上缩减到一个单独的系数，来获取输入特征图的全局信息，即进行获取初始化系数的过程。然后，CAM 采用 Excitation 策略，它是一种用于生成权重系数的技术，该策略的核心在于引入非线性<sup>[19]</sup>。它具体实现方式是设计了两个连续的全连接层，第一个全连接层用于降维，第二个全连接层用于升维，并在这两个层之间使用一个非线性的激活函数。此外，原 Excitation 模块使用了 ReLU 函数和 Sigmoid 函数，而本实验中将其更换为 Softsign 激活函数。因为 Sigmoid 函数的输出范围控制在 0 到 1 之间，而 Softsign 函数使获得的权重系数在 -1 到 1 之间，这使模型也能学习到负权重，从而抑制某些特征，以便更好地调整特征图。具体来说，在水下情境中，负权重可以帮助抑制某些颜色通道，比如明显的蓝色和绿色。

上述过程 CAM 模块的工作流程可以表示为：

$$\text{Coeff}_m = \psi(C_2\psi(C_1\text{GAP}_m)) \quad (4.1)$$

其中， $C_1$ ， $C_2$  分别表示降维和升维的操作， $\text{GAP}_m$  是经过全局平均池化后的特征图， $\psi$  表示 Softsign 激活函数。

## 4.3 实验及结果

### 4.3.1 数据集准备

图像增强模块的所需的训练数据集包括同一水下场景中的高、低质量图片这两类，其中，使用低质量图片作为输入图片，高质量图片作为 Ground Truth。

但根据文献<sup>[20]</sup>的研究结果，并不是所有水下目标检测方法都能从图像增强技术中获益。实际上，大多数水下目标检测模型在应用现有图像增强算法对图片进行预处理后，其性能并没有提升，反而出现了下降的情况。这可能是因为现有图像增强算法是针对图像的视觉质量，而视觉质量的提高并不意味着模型性能的增强。在图像增强过程中，如平滑，去噪等，可能反而会使图片丢失了可用于目标检测的特征信息，干扰了检测算法的正常工作。

因此，为了保证图像增强模块能有利于后续的目标检测任务，本实验修改了 Ground Truth 数据，从原模型所用的基于图像增强指标的高质量图片，改为基于目标识别指标的高质量图片，确保本实验模型能够将原始图片向着提高目标检测性能的方向进行增强。

一张基于目标识别的高质量图片的制作流程大致如下：

（1）训练好 Faster RCNN，Cascade RCNN，ATSS，Yolov3 和 Retinanet 五种目标识别模型。

（2）分别使用 Wk\_Net、LCNet、TACL、TUDA、FUIEGAN、LANet、STSC、HE\_Prior 及 TOPAL 九种经典的图像增强算法对原始图片进行图像增强。

（3）使用九张经过不同方法得到的增强图片与原始图片，共十张图片，分别输入上述的五个目标检测模型中进行测试，取每张图片在这些模型中得到的 mAP 的平均值作为评估该图片所对应的增强方法下的目标检测效果的定量指标。

（4）根据（3）中的指标，取最好的平均 mAP 所对应的增强图像作为该图片所对应的高质量图片。

本模块所用数据集包含 2000 对水下高低质量图片，其中低质量图片均选取自上一目标检测模块所用数据集中的图片，再经过上述流程得到相应的 2000 张高质量图片用于模型的训练。

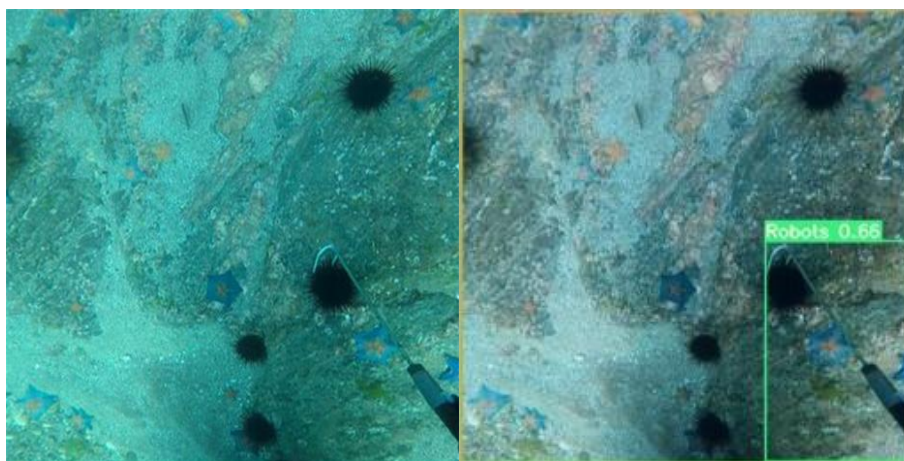
#### 4.3.2 实验结果分析

表 4.1 图像增强前后模型结果对比

模型	P	R	mAP@.5	mAP@.5:.95
YOLOv7	69.3%	55.7%	60.6%	40.3%
YOLOv7_Enh	71.4%	56.1%	61.5%	41.2%

根据评估结果，图像增强策略全面提升了模型的检测效果，精确率，召回率和 mAP 值分别提升了 2.1%，0.4%，0.9%，说明了该图像增强方法对于提高目标识别效果的可行性。其中精确率提升水平瞩目，可能是进行颜色的校正后，不同目标间、目标与背景间的特征差异明显，降低了误检率。

图 4.4 中的 robots 目标经过原 YOLOv7 模型及第三章中得到的模型都无法识别出来，如图 4.4（a），但经过图像增强后使用原 YOLOv7 网络能够识别出该目标，如图 4.4（b）。这可能是因为原图像整体色彩偏绿，增加了目标与背景的区别难度。经过图像增强，使环境颜色恢复到与它的真实颜色相近，而背景的真实颜色特征其实与目标相差较大，一定水平上恢复了环境与目标物体间存在的差异，有利于识别。



(a) 未进行增强的模型识别后的图像 (b) 进行增强的模型识别后的图像

图 4.4 经过图像增强的目标识别效果图

## 4.4 本章小结

本章阐述了所使用的图像增强模块的工作原理,为了保证该模块的图像增强能够有利于后续目标检测任务,使用了有利于目标检测的高质量图像作为 Ground Truth。通过实验验证了该图像增强模块对的 YOLOv7 网络的水下目标识别的性能确实能够带来提升。

## 第五章 引入图像增强的改进 YOLOv7 水下目标识别系统

### 5.1 整体网络设计与训练流程

#### 5.1.1 网络结构

引入了图像增强的改进 YOLOv7 水下目标识别网络的整体结构如图 5.1 所示，该网络包括图像增强模块与目标识别模块。

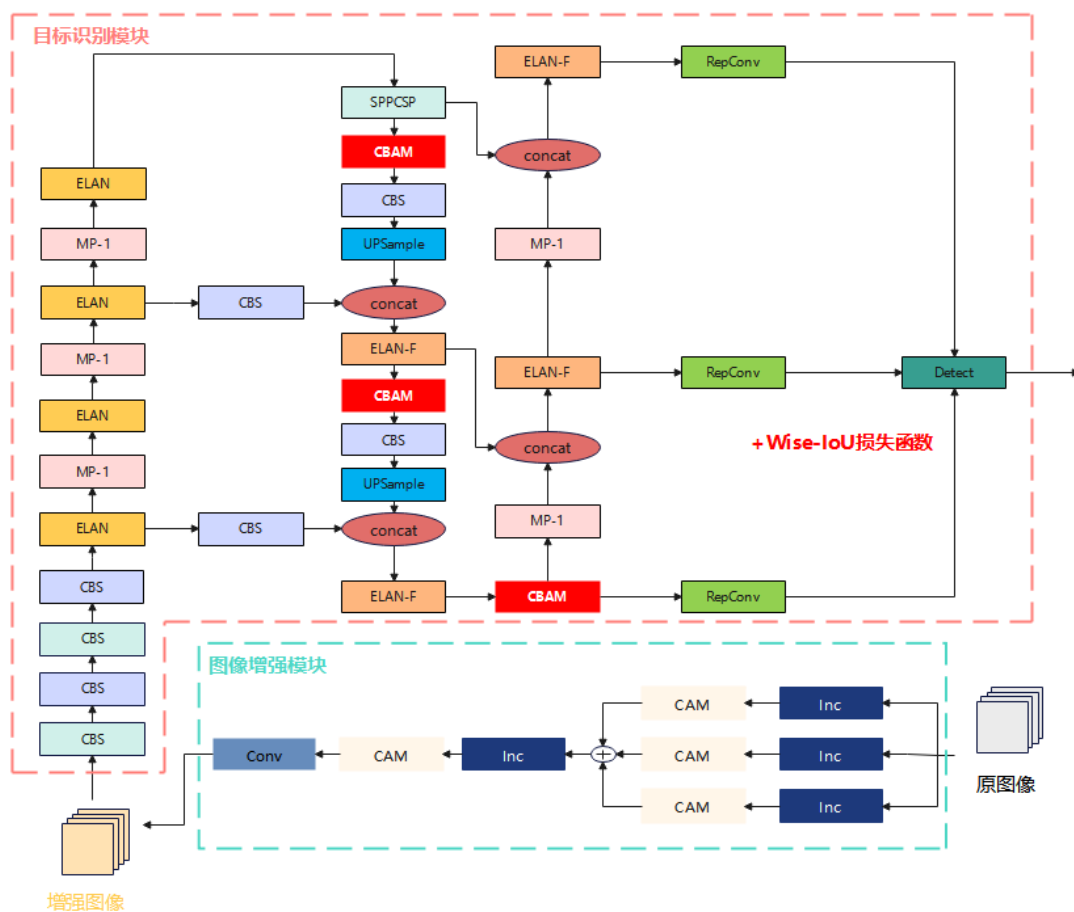


图 5.1 引入图像增强的改进 YOLOv7 水下目标检测网络结构

#### 5.1.2 训练流程

由于本实验分为两个模块，整体网络的训练也需要分两阶段进行。

第一阶段，图像增强。使用含有水下高低质量图像对的数据集训练图像增强模型。

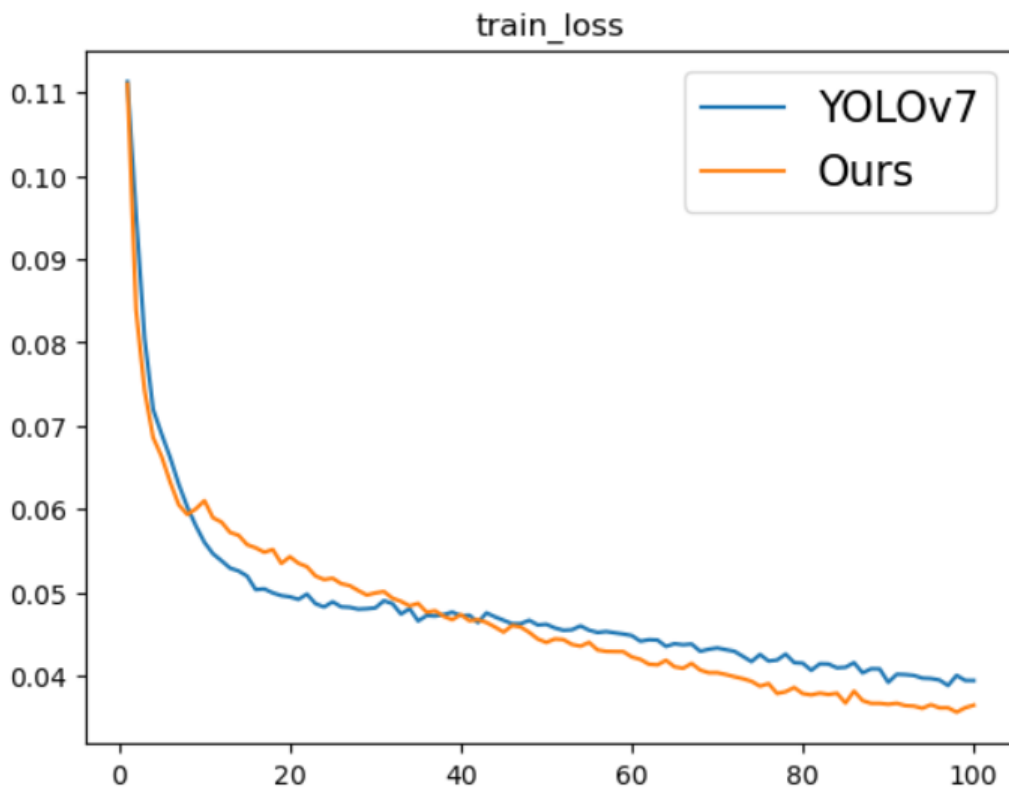
第二阶段，该阶段包括数据预处理与模型的训练。数据预处理是指将目标检

测所用数据集先全部输入训练好的图像增强模型中进行图像增强，得到经过增强的图像后，将增强图像用于 YOLOv7 网络的训练，最终可以得到新的水下目标识别模型。

## 5.2 实验结果

### 5.2.1 实验结果分析及对比

YOLOv7 原网络与当前模型在相同数据集和环境设置下分别进行训练后，得到训练过程中的训练集和验证集的总损失值变化曲线，如图 5.2 所示。可以发现，改进后的模型有较快的收敛速度，最终也达到一个更低的损失值。



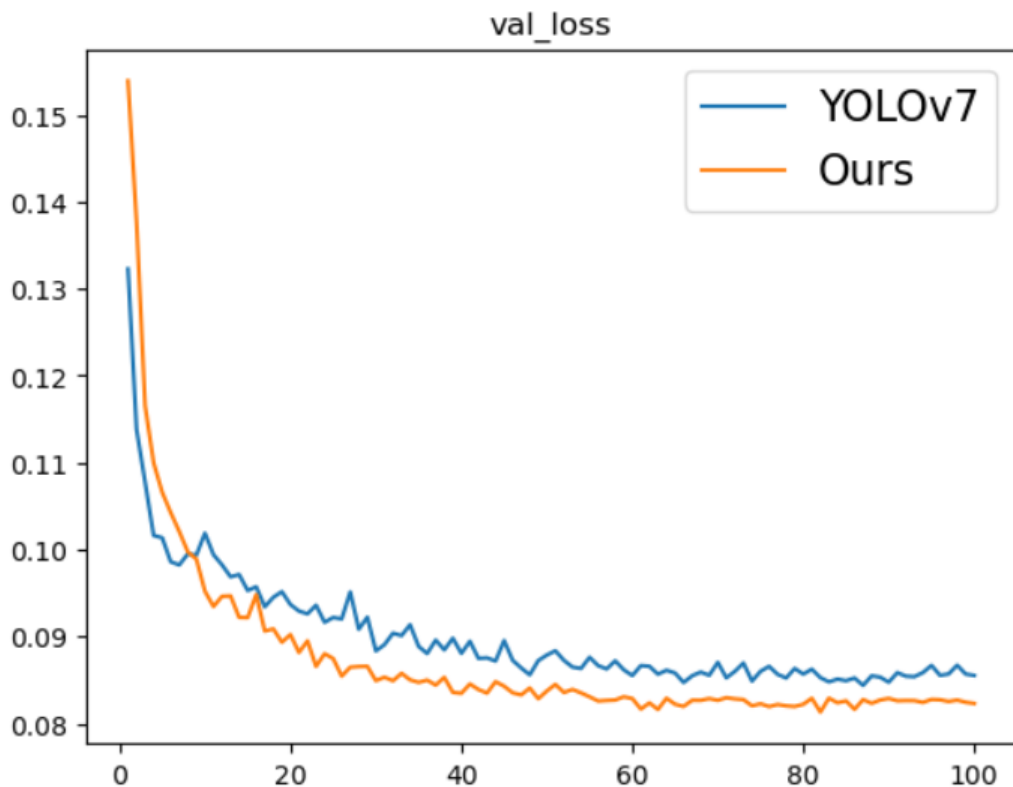


图 5.2 训练过程中的损失函数曲线

接下来，本节分别对各类目标的识别精确率，召回率与 mAP 结果进行了对比与分析。

表 5.1 改进前后精确率对比

模型	Human Divers	Wrecks/ ruins	Fish/ver tebrates	Plant/se a-grass	Robots	Fish_gr oup
YOLOv7	90.1%	82.9%	68.5%	43.6%	73.3%	57.5%
YOLOv7_Ours	91.2%	76.9%	82.5%	47.7%	71.1%	61.2%
提升效果	+1.1%	-6.0%	+1.4%	+4.1%	-2.2%	+3.7%



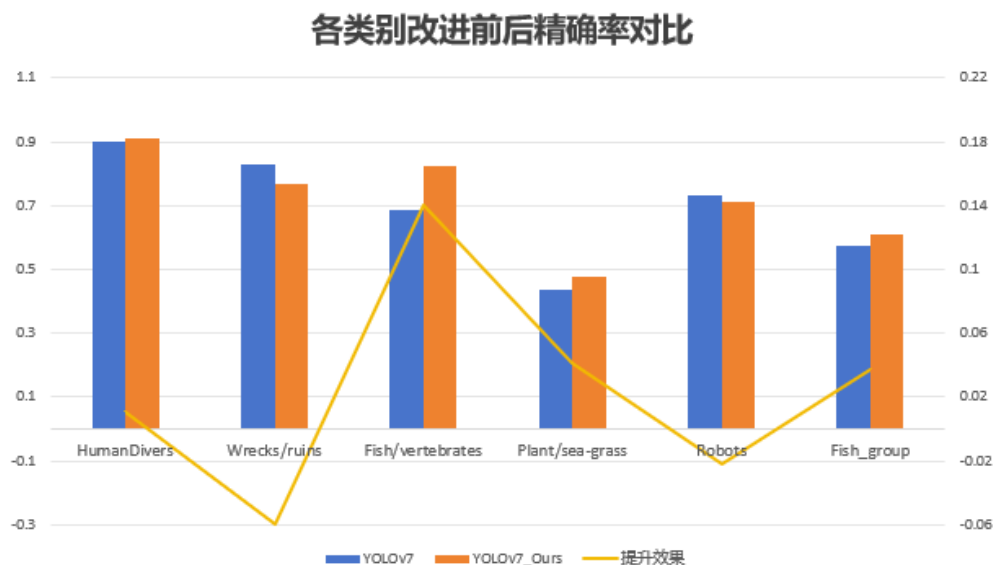


图 5.3 各类别改进前后精确率对比

表 5.2 改进前后召回率对比

模型	Human Divers	Wrecks/ruins	Fish/vertebrates	Plant/sea-grass	Robots	Fish_group
YOLOv7	86.3%	56.8%	63.9%	22.9%	58.3%	44.4%
YOLOv7_Ours	89.8%	6.21%	62.2%	25.0%	61.7%	50.0%
提升效果	+3.5%	+5.3%	-1.7%	+2.1%	+3.4%	+5.6%

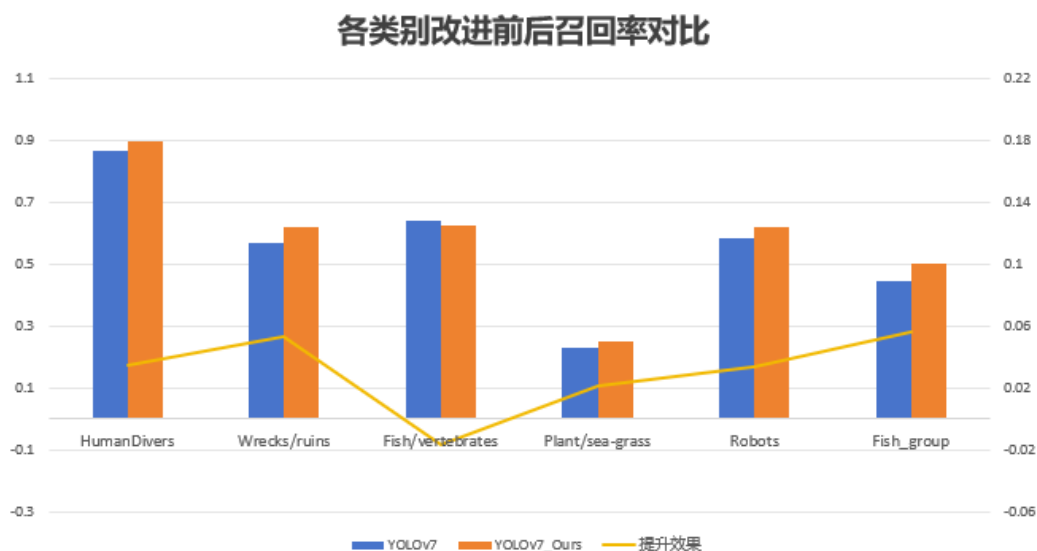


图 5.4 各类别改进前后召回率对比

表 5.3 改进前后 mAP 值对比

模型	Human Divers	Wrecks/ruins	Fish/vertebrates	Plant/sea-grass	Robots	Fish_group
YOLOv7	89.6%	65.6%	69.0%	22.7%	68.0%	44.8%
YOLOv7_Ours	91.5%	67.8%	70.8%	23.8%	69.2%	53.8%
提升效果	+1.9%	+2.2%	+1.8%	+1.1%	+1.2%	+9.0%

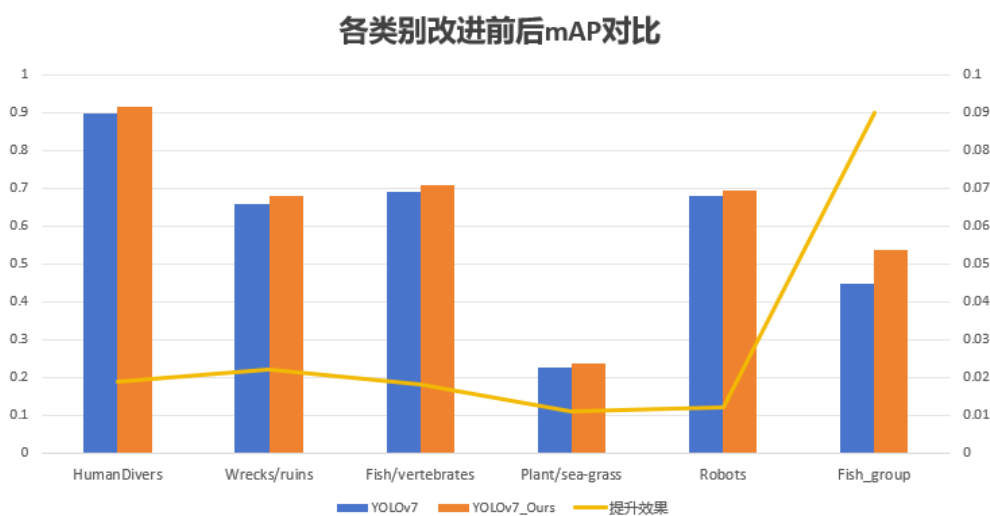


图 5.5 各类别改进前后 mAP 对比

根据表 5.3-5.5，经过改进的模型取得了较好的目标识别性能，对六类目标的识别精确度与召回率大部分都提升了，各类别的 mAP 值全部取得了提升。

根据类别横向比较模型结果可以发现，其中对于 HumanDivers 的识别准确度普遍较高，观察数据集可以发现，可能是因为 HumanDivers 具有更易识别的生物学特征，与环境差异比较大，同时图像可用性也较高。而对于 Plant/sea-grass 和 Fish\_group 目标，模型识别效果较差，Plant/sea-grass 识别的 mAP 值甚至都在 30% 以下。Plant/sea-grass 识别率较低可能是因为水下图片颜色较暗，且水草本身颜色就与水下环境相近，另外在许多图片中（如图 5.6（b）中第三张例图），它受附近水下物体的遮挡严重，所以不易识别；Fish\_group 识别率低可归因于在任务的其它目标中有 Fish 的选项，两者容易混淆，模型常常将 Fish\_group 识别为多个 Fish 目标，大大影响了对 Fish\_group 的识别效果。图 5.6 展示了包含 HumanDivers，Plant/sea-grass 与 Fish\_group 类别的训练集中的示例图片。



(a) HumanDivers



(b) Plant/sea-grass



(c) Fish\_group

图 5.6 三类目标的图像

而根据不同模型纵向对比,可以发现经过改进的模型在上述两类目标的识别的三项指标中均有稳定提升,说明该改进策略确实在当前数据中对水下难以识别的目标有一定程度上的效果。

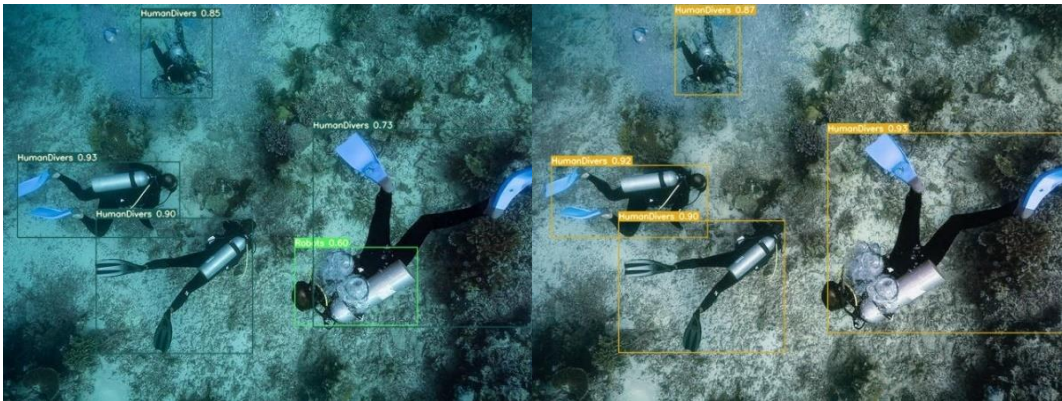
### 5.2.2 消融实验

实验结果如表 5.4 所示,对比实验 1、4,最终引入图像增强的改进 YOLOv7 与原模型相比, mAP 值提升了 2.2%,精确率提升了 2.5%,召回率提升了 2.8%,模型性能得到较大改善。

对比实验 2、3,可以发现在本实验的数据及环境下,数据增强策略和对 YOLOv7 网络结构的改进策略对于网络识别效果的提升效果其实有所不同,两者各有所长,数据增强策略能更大地提升模型的精确率,而第三章中对网络结构的改进对召回率有更大改善。两者的综合性能,即 mAP 值相近。

表 5.4 消融实验结果

序号	模型	图像 增强	网络 改进	P	R	mAP@. 5	mAP@. 5:95
1	YOLOv7	×	×	69.3%	55.7%	60.6%	40.3%
2	YOLOv7_Enh	√	×	71.4%	56.1%	61.5%	41.2%
3	YOLOv7+CBAM +WIoU	×	√	70.8%	59.2%	61.7%	43.7%
4	YOLOv7_Ours	√	√	71.8%	58.5%	62.8%	44.3%



(a)



(b)

图 5.7 识别效果对比

图 5.7 展示了改进后模型的水下目标识别效果图，左侧为原始 YOLOv7 模型检测结果，右侧为引入图像增强并进行识别网络结构改进的 YOLOv7 模型检测结果。在（a）中，改进模型避免了对 Robots 的误检，对 HumanDivers 的预测



置信度平均较高；在（b）中，对属于 Fish/vertebrates 和 Robots 两个类别的目标识别置信度均提高了。

### 5.3 可视化界面设计

本实验设计的海洋水下目标识别系统界面是基于 PyQt5 框架实现的，该框架是一个创建跨平台 GUI 应用程序的流行工具。系统界面的设计旨在提供一个直观、用户友好的操作环境，使用户能够轻松地进行图像选择、检测和结果保存。

系统的可视化界面下图 5.8-5.12 所示，图 5.8 展示了初始界面，右侧矩形框是原始图片的显示区，左侧则是经过增强及检测的结果图片显示区。



图 5.8 初始界面



图 5.9 开始检测步骤报错界面



图 5.10 检测成功界面



图 5.11 保存结果步骤界面



图 5.12 保存步骤报错界面

可视化界面的工作流程如下：

（1）运行 **demo** 程序，加载原始界面。在原始界面点击“选择图片”按钮，系统会弹出文件选择界面如图 5.8 所示，默认路径为该文件下的“inference/images”文件夹。

（2）选择所要检测的水下图片，并点击“开始检测”。如果在没有选择图片时点击“开始检测”，页面会弹出提示“请先导入图片”，如图 5.9 所示。

（3）开始检测后，系统首先运行增强模型的测试程序，加载图像增强模型，将图片输入增强模型中进行处理，处理完后已增强的图片会保存在当前文件夹下。系统接着会自动将已增强的图片输入目标检测模型中，并运行目标检测模型的检测程序，加载目标检测模型，在已增强的图片上进行目标检测，最终输出带有检测标签及检测框的图片，如图 5.10 所示。

（4）点击“保存图片”按钮，可以设置图片的保存名称和路径，完成设置后，可以将目标检测的图片结果保存下来。确定后，会显示“保存成功”，如图 5.11，证明检测图片已保存。如果未导入图片或进行检测时就点击“保存结果”，页面会提示“请先进行图片检测”，如图 5.12 所示。

## 5.4 本章小结

本章根据第三、四章得到的目标检测模块与图像增强模块，得到了引入图像增强的 YOLOv7 水下图像目标识别模型，将目标检测数据集经过增强后再输入改进的 YOLOv7 网络中进行训练和测试，通过对比与消融实验，对比各目标类别与整体的 mAP 等检测指标，说明了该模型能够在一定程度上提高对水下目标的识别性能。



## 总结与展望

基于光学图像的水下目标识别技术是海洋科学、资源开发、环境保护、军事应用和搜救行动等领域的关键技术。随着对海洋探索的不断深入，对水下目标的精确识别和定位需求也日益增长。海洋资源的开发需要对海底地形、地质结构和生物分布进行详细了解，正需要水下目标识别技术来提供这些信息。在环境保护方面，该技术有助于监测海洋污染、珊瑚礁破坏和生物多样性变化，为保护海洋生态提供科学依据。在军事领域，水下目标识别对于潜艇探测、水雷清除和反潜作战至关重要，增强了国防安全。技术的进步，包括成像技术、图像处理算法和计算能力的提升，使得水下目标识别技术更加精准和高效，能够适应多变的水下环境。

本文首先对基于光学图像的目标检测算法相关理论知识做了阐述，着重介绍了基于深度学习的方法。本实验主要实现了一个引入图像增强的改进 YOLOv7 水下目标识别系统，该系统包含基于 YOLOv7 的目标识别与基于深度学习的图像增强两个模块。在基于 YOLOv7 的目标识别模块，通过在不同位置加入 CBAM 注意力机制的实验从而得出最优策略，也尝试了 Wise-IoU 损失函数，最后得到了在 Neck 层加入 CBAM 注意力机制并采用 Wise-IoU 损失函数的最优组合改进方式。在图像增强模块，使用了包含有利于目标检测的高质量图片数据集，从而确保图像向着有利于后续检测任务的方向进行增强。经过实验，使用该模块处理过的增强图片进行目标识别，确实提升了原 YOLOv7 模型的检测效果。最后，将两个模块结合，将经过增强处理的图片输入到改进的 YOLOv7 模型中，获得的模型性能良好，mAP 值较原模型提升了 2.2%。此外，本实验也制作了一个可视化界面，可用于水下图片的目标识别效果展示，在后端会自动实现对输入图像的增强与识别并输出结果，实现了端到端的应用。

尽管引入图像增强的水下目标识别模型取得了不错的性能提升，但是显然，仍存在很多可以改进之处：

(1) 最终将训练与预测过程分为两个阶段，没有实现端到端的训练。之后可以更加深入地研究增强与识别的网络结构，将其调整为一个完整的增强与识别网络。

（2）图像增强模块的训练过分依赖数据集，同时用于训练图像增强和目标识别的数据集规模较小，未来可以在完善增强与识别策略的基础上，尝试使用更大的水下图像数据集。

（2）由于时间及设备的限制，本文进行的实验不够充足，可以进行更多的对比实验来增强说服力。

## 参考文献

- [1]陈谦. 基于深度学习的水下目标检测方法研究[D].北京: 北京石油化工学院,2024.
- [2]苗馨方. 改进 YOLO 算法研究及其在水下生物识别中的应用[D].长春: 长春工业大学,2023.
- [3]Olmos E A. Detecting man-made objects in unconstrained subsea videos [J]. British Machine Vision Conference, 2003: 21-30.
- [4]Fatan M, Daliri M R, Shahri A M. Underwater cable detection in the images using edge classification based on texture in-formation[J].Measurement,2016, 91:309-317.
- [5]Susanto T,Mardiyanto R,Purwanto D.Development of Underwater Object Detection Method Base on Color Feature [C]. 2018 International Conference on Computer Engineering,Network and Intelligent Multimedia (CENIM).Indonesia:IEEE, 2018:254-259.
- [6]Li X, Shang M, Qin H, et al. Fast accurate fish detection and recognition of underwater images with Fast R-CNN[C]. Oceans, 2015: 1-5.
- [7]Li X , Shang M , Hao J ,et al. Accelerating fish detection and recognition by sharing CNNs with objectness learning[J]. OCEANS 2016 - Shanghai, 2016: 1-5.
- [8]Zhang J, Zhu L, et al. MFFSSD: An Enhanced SSD for Underwater Object Detection[J]. Chinese Automation Congress (CAC), 2020: 5938-5943.
- [9]Hou C, Guan Z, Guo Z, et al. An improved YOLOv5s based scheme for target detection in a complex underwater environment[J].Journal of Marine Science and Engineering,2023,11(5):1041.
- [10]Yu M, Shen L, Wang Z, Hua X. Task-Friendly Underwater Image Enhancement for Machine Vision Applications[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62(1): 1-14.
- [11]Virasova A, Klimov D I, Khromov O, Gubaidullin I R, Oreshko V V.

Rich feature hierarchies for accurate object detection and semantic segmentation[J]. Radioengineering, 2021.

[12]Redmon J, Divvala S K, Girshick R B, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 779-788.

[13]Yu J, Jiang Y, Wang Z, Cao Z, Huang TS. UnitBox: An Advanced Object Detection Network[C]. Proceedings of the 24th ACM international conference on Multimedia, 2016.

[14]Wang C, Bochkovskiy A, Liao H 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 (CVPR), 2022: 7464-7475.

[15]Woo S, Park J, Lee J, Kweon I. CBAM: Convolutional Block Attention Module[EB/OL]. (2018-07-17)[2024-05-11]. <https://arxiv.org/abs/1807.06521>.

[16]Tong Z, Chen Y, Xu Z, Yu R. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism[EB/OL]. (2023-01-24)[2024-05-11]. <https://arxiv.org/abs/2301.10051>.

[17]Tolie H F, Ren J, Elyan E. DICAM: Deep Inception and Channel-wise Attention Modules for underwater image enhancement[J]. Neurocomputing, 2024.

[18]Schettini R, Corchs S. Underwater Image Processing: State of the Art of Restoration and Image Enhancement Methods[J]. EURASIP Journal on Advances in Signal Processing, 2010, 2010: 1-14.

[19]Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation Networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017: 7132-7141.

[20]Wang Y, Guo J, He W, Gao H, et al. Is Underwater Image Enhancement All Object Detectors Need? [J]. IEEE Journal of Oceanic Engineering

49, 2023: 606-621.

[21]方笑海. 基于神经网络的水下目标图像识别算法研究[D].杭州电子科技大学,2023.DOI:10.27075/d.cnki.ghzdc.2022.000075.

[22]郑棣文. 基于深度学习的水下光学图像目标检测算法研究[D].哈尔滨工程大学,2023.

[23]梁秀满,李然,于海峰,刘振东. 改进 YOLOv7 的水下目标检测算法[J]. 计算机工程与应用,2024,60(06):89-99.

[24]王钧,麻方达,符朝兴. 基于改进 YOLOv8 的小目标零件识别方法[J]. 青岛大学学报(工程技术版):1-9.

[25]朱攀. 基于深度学习的水下光学图像目标检测系统的设计与实现[D].东南大学,2024.

[26]李云龙. 基于 Faster-RCNN 网络的水下光学图像目标检测算法研究[D].哈尔滨工程大学,2024.

[27]Shorten C, Khoshgoftaar T M. A survey on Image Data Augmentation for Deep Learning[J]. Journal of Big Data, 2019, 6: 1-48.

[28]Chen X, Yuan M, Yang Q, Yao H, Wang H. Underwater-YCC: Underwater Target Detection Optimization Algorithm Based on YOLOv7[J]. Journal of Marine Science and Engineering, 2023.

[29]Fu C, Liu R, Fan X, et al. Rethinking general underwater object detection: Datasets, challenges, and solutions[J]. Neurocomputing, 2022, 517: 243-256.

[30]Han F, Yao J, Zhu H, Wang C. Underwater Image Processing and Object Detection Based on Deep CNN Method[J]. J. Sensors, 2020, 6707328: 1-20.

[31]Alzubaidi L, Zhang J, Humaidi A J, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions[J]. J Big Data, 2021, 8: 53.

## 致谢

随着毕业设计即将完成，我的大学时光也进入了尾声。在本科四年的学习与生活中，我经历颇丰，收获颇多。

在完成毕业设计的过程中，我要感谢公舒艺，杨蕾等我的朋友，室友及其他我身边的人，她们在我完成毕设与实习的艰难时光里，陪伴我的生活，倾听我的烦恼，在我不管大小的麻烦时，都真诚地为我提供力所能及的帮助；感谢毕设期间为我提供指导的学长曹磊，在很多学术问题上，他积极地为我答疑解惑，提出了有益的建议，为我的毕设提供了很多关键性帮助；感谢指导老师沈礼权老师，给予我很多宝贵的建议，为我的毕设指明方向；最后我也要感谢我的父母，他们的理解与支持始终是我坚强后盾。

在本科四年的生活中，从同学到老师，从朋友到家人，很多人都为我的生活提供了助力，增添了色彩，都是对我来说很宝贵的人，值得我奉上最诚挚的谢意。

在未来的征途上，我将携带着这份感恩之心，继续前行。

雷萌

上海大学

2024年5月13日