

基于 Co-PSPNet 的轻量级水下鱼体图像分割算法

李晓雯, 李海涛, 高树静, 张俊虎
(青岛科技大学 信息科学技术学院, 山东 青岛 266061)

摘要: 为解决水下环境中图像模糊和资源有限等问题, 提出了一种基于 Co-PSPNet 网络的轻量级水下鱼体图像分割算法; 主干网络使用了 MobileNetV2, 以降低网络的计算复杂度和参数量; 引入了 CoordConv 模块, 以增强网络对鱼体边界等细节信息的感知能力; 将全局池化后的特征作为注意力机制网络的输入, 以增强具有较高语义信息的特征; 经过大量的实验评估, 该算法在公开的水下鱼体图像数据集上取得了优越的性能; 实际应用中, 该算法满足了水下生态研究和水下机器人领域对水下鱼体图像分割的工程需求; 通过对水下环境中图像模糊和资源限制等问题的解决, 该算法为水下生态研究和水下机器人领域的应用提供了有效的图像分割解决方案。

关键词: 图像分割; 轻量级算法; PSPNet; CoordConv 模块; 水下鱼体; 注意力机制

A Lightweight Underwater Fish Body Image Segmentation Algorithm Based on Co-PSPNet

LI Xiaowen, LI Haitao, GAO Shujing, ZHANG Junhu

(School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: To solve the problems of image blurring and limited resources in underwater environments, a lightweight underwater fish segmentation algorithm based on the Co-PSPNet network is proposed. The main network employs MobileNetV2 to reduce computational complexity and parameter count. The CoordConv module is introduced to enhance the network to the perception of detailed information like fish boundaries. The features obtained after the global pooling are used as inputs for an attention mechanism network, enhancing the features with higher semantic information. Through the extensive experimental evaluation, the algorithm has a superior performance of publicly available underwater fish image datasets. In practical applications, this algorithm meets the engineering requirements of underwater fish image segmentation in the domains of underwater ecology research and underwater robotics. By solving the image blurring and resource limitations in underwater environments, the proposed algorithm provides an effective solution for image segmentation in underwater ecology research and underwater robotics applications.

Keywords: image segmentation; lightweight algorithm; PSPNet; CoordConv module; underwater fish; attention mechanics

0 引言

真实水下场景中的鱼类语义分割是一项具有挑战性的任务, 也是各种处理步骤的重要前提。水下鱼体图像分割在水下生态研究^[1]和水下机器人领域具有重要意义。准确地分割水下鱼体图像可以为生态学研究提供宝贵的数据, 并为水下机器人的自主导航和目标识别提供支持^[2]。然而, 由于水下环境的特殊性, 水下鱼体图像分割面临着一些挑战和限制^[3]。资源有限、光线衰减、水下散射等因素会影响图像的质量和鱼体的可见性, 使得传统的分割方法在水下环境中表现不佳。此外, 水下鱼体的形变和细节变化也增加了分割的难度。

语义分割方法可分为传统图像处理方式和深度学习方

式^[4-5]两种。近年来, 随着深度学习和计算机视觉的进展, 传统的图像分割方法, 如阈值分割、边缘检测方法、区域生长、区域分裂与合并、边界分割、基于聚类的方法等^[6-8], 通常在处理复杂、噪声较多或者场景变化较大的图像时效果有限。相比之下, 基于深度学习方法能够通过自动学习特征和语义信息, 更准确地处理复杂的图像分割任务, 适应多样性的场景和物体变化。文献 [9] 探讨了两种深度学习模型在低光摄像头下的鱼类库存监测应用中进行鱼类分割的适用性。通过修剪这些网络并采用不同的编码器, 它们更适用于硬件有限的系统, 如远程操作或自主操作的水下载具。文献 [10] 提出了一种双池聚合注意网络, 通过新颖的位置和通道注意模块, 高效地捕获长距离依赖关系, 从而在提升特征表示的同时显著改善了分割性能。

收稿日期: 2023-08-08; 修回日期: 2023-08-16。

基金项目: 山东省重点研发计划(科技示范工程)课题(2021SFGC0701); 青岛市海洋科技创新专项(22-3-3-hygg-3-hy)。

作者简介: 李晓雯(1999-), 女, 硕士研究生。

通讯作者: 高树静(1976-), 女, 博士, 副教授。

引用格式: 李晓雯, 李海涛, 高树静, 等. 基于 Co-PSPNet 的轻量级水下鱼体图像分割算法[J]. 计算机测量与控制, 2024, 32(2): 268-275, 308.

文献 [11] 提出了一种预处理 CNN, 用于声纳图像中鱼类分割, 该预处理 CNN 与条件随机场集成, 旨在分离学习鱼类实例和学习鱼类养殖环境。这种方法可以改进声纳图像中鱼类的 Mask R-CNN 分割, 并简化在鱼类养殖环境中应用 Mask R-CNN。文献 [12] 探索了在嘈杂低分辨率图像数据集中, 采用深度学习和卷积神经网络 (CNN) 进行鱼类分割的方法。通过使用成像声纳生成的距离-方位位置, 实现了鱼类与非鱼类概率预测, 并在丹麦海峡和法罗群岛的自采集数据上训练和测试模型, 展示了在小规模数据集上实现满意性能和泛化能力的技术。

本文旨在研究深度学习方法对真实水下场景中检测到的鱼类进行分割和轮廓提取的适用性。除了关于图像中存在什么样的对象以及它们位于何处的信息外, 一个成功的语义分割揭示了每个像素属于哪一类。因此, 在图像中额外提取感兴趣目标的轮廓及其覆盖的简洁区域成为可能。鱼类的精确分割是自动确定形态特征的重要前提, 如总长度, 进而可用于确定鱼类体重。近年来, 对于在有限硬件条件下成功应用深度学习算法的需求不断增长。对于打算使用水下机器人、遥控车辆等应用的需求尤为重要。因此, 本文主要研究了轻量级分割模型。具体而言, 我们采用了改进版本的“金字塔场景解析网络 (PSPNet, pyramid scene parsing network)”^[13]来进行包含鱼的图像分割任务, 以实现鱼类和背景类的二值分割。PSPNet 作为一种高效的语义分割算法, 通过利用金字塔池化模块 (PPM, pyramid pooling module) 来捕获不同尺度的上下文信息, 取得了显著的成果。然而, 在水下鱼体图像分割中, 传统的 PSPNet 仍然存在一些限制。

为了进一步提高分割精度和算法的轻量级特性, 本文提出了一种基于 PSPNet 改进的轻量级水下鱼体图像分割算法。首先, 我们选择将主干网络替换为 MobileNetV2^[14], 以降低算法的参数量和计算复杂度。MobileNetV2 是一种轻量级的网络结构, 具有良好的性能和较低的计算开销, 适用于资源受限的环境。其次, 为了增强算法对水下鱼体图像的空间信息表示能力, 我们引入了 CoordConv 模块^[15]。

CoordConv 模块通过将坐标信息作为额外的输入通道, 使得网络可以更好地感知和利用像素的位置信息, 从而提升分割算法对水下鱼体图像中复杂形态和纹理的建模能力。最后, 我们对 PSPNet 的金字塔池化模块进行改进, 将全局池化后的特征作为全注意力机制网络的输入, 注意力机制能够帮助模型更加关注重要的特征, 并抑制不重要的特征, 从而进一步提升分割的性能。

通过以上改进策略, 我们旨在提高水下鱼体图像分割算法的准确性和效率, 为水下生态研究和水下机器人应用提供更可靠的工具和方法。在接下来的章节中, 将详细介绍改进的算法, 并通过实验验证其在水下鱼体图像分割任务中的优越性能。

1 水下鱼体语义分割模型

1.1 改进的 Co-PSPNET 网络

改进的 Co-PSPNET 算法主要分为 4 个部分, 分别为主干网络、CoordConv 模块、基于 CBAM 注意力机制的金字塔池化模块、解码器。

算法框架结构如图 1 所示。

第一部分为主干网络。假设输入图片的尺寸为 $H \times W$ (高度 \times 宽度), 图片经过主干网络 MobileNetV2 进行特征提取。MobileNetV2 的卷积层在处理图片时, 会进行下采样, 通常会将输入图片的尺寸缩小为 $H/32 \times W/32$ 。通过初始卷积层和多个 Bottleneck 块进行特征提取。每个 Bottleneck 块包含 1×1 的扩展卷积层、 3×3 的深度可分离卷积层和 1×1 的投影卷积层, 最后通过上采样将特征图恢复到输入图片大小。

第二部分为 CoordConv 模块。首先, CoordConv 模块生成一个与特征图大小相同的坐标网格。这个坐标网格的形状是 $H \times W$, 每个像素点对应一个二维坐标 (x, y) 。接下来, CoordConv 模块将坐标网格的 (x, y) 坐标信息分别扩展为两个 3 维张量, 分别是形状为 $1 \times H \times W$ 的 X 坐标张量和 Y 坐标张量。然后, CoordConv 模块将上述生成的 X 坐标张量和 Y 坐标张量分别与原始特征图进行通道拼接。这样, 特征图的每个像素点都会添加两个额外的通道, 分

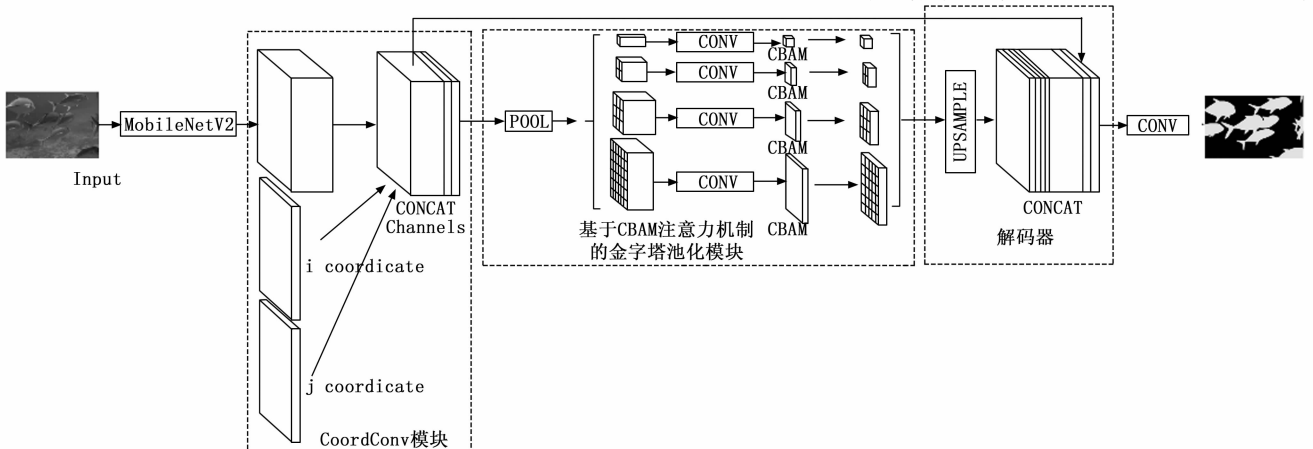


图 1 改进 Co-PSPNet 网络

别对应该像素点的 X 坐标和 Y 坐标信息。最终，特征图的形状变为 $(C+2) \times H \times W$ ，其中 C 个通道对应于原始特征图的通道，而多出来的 2 个通道是 X 坐标和 Y 坐标信息。

第三部分是基于 CBAM 注意力机制的金字塔池化模块。针对 CoordConv 模块输出特征图的全局以及 $1/4$ 、 $1/9$ 、 $1/36$ 子区域分别进行全局平均池化操作。每个池化后，特征向量经过 CBAM 模块处理，通过通道注意力和空间注意力分别学习通道权重和空间权重，分别对不同通道的特征和不同空间位置的特征进行加权，以增强重要通道和位置的特征表示能力。

第四部分为解码器，经过金字塔池化模块后得到的特征图，通过解码器进行上采样和融合操作，以恢复原始图像尺寸 $H \times W$ 的语义分割结果。解码器使用双线性插值的方式来实现上采样，并与 CoordConv 模块输出的特征图在通道维度上进行拼接，实现高层语义信息和低层空间信息的有机结合。

最后，解码器生成的语义分割结果通过 1×1 卷积核的卷积层进行通道维度上的降维操作，最终输出的通道数为类别数量，用于生成语义分割的预测结果。这样，模型就可以将每个像素点分类到对应的语义类别，并完成整个语义分割任务。

1.2 特征提取网络

传统的 PSPNet 主干网络使用的是 ResNet 等较为复杂的网络结构，具有较高的计算复杂度和参数量。为了实现轻量级水下鱼体图像分割算法，本算法通过实验对比多种不同的轻量级分类网络见 3.4 节，我们选择了 MobileNetV2 作为替代方案。MobileNetV2 具有轻量级的结构和高效的特征提取能力，能够在保持准确性的同时减少计算开销，适应资源受限的水下环境。

MobileNetV2^[14] 是一种轻量级的卷积神经网络架构，目的是在移动设备和嵌入式系统上实现高效的图像识别和计算。它是 Google 团队在 2018 年提出的 MobileNet 系列的第二个版本。MobileNetV2 的设计目标是在保持高精度的同时，显著减少模型的大小和计算复杂度。它采用了一系列的设计策略和技术，以实现这一目标。下面是 MobileNetV2 的一些关键特点和技术：MobileNetV2 引入了一种称为“*Inverted Residuals*”（反向残差模块）的模块，它包含了轻量级的深度可分离卷积（Depthwise separable convolution）^[16]。这种模块的结构与传统的残差模块相反，先使用 1×1 的卷积进行降维，然后应用深度可分离卷积进行特征提取，最后再使用 1×1 的卷积进行升维。这种结构能够有效地减少参数数量和计算复杂度。MobileNetV2 引入了一个宽度乘法器，用于动态地调整模型的宽度（即通道数）。通过调整宽度乘法器的值，可以在精度和计算复杂度之间进行权衡。较小的乘法器值将减少模型的参数和计算量，但可能会降低模型的性能。MobileNetV2 可以适应不同的输入分辨率，从而在不同的应用场景中灵活应用。通过改变输入图像的分辨率，可以在模型大小和推理速度之

间进行权衡。MobileNetV2 中的深度可分离卷积具有线性瓶颈（linear bottlenecks）特性，即在升维和降维的过程中使用了线性激活函数。这可以避免非线性激活函数引入的额外计算开销。

总体而言，MobileNetV2 通过结合多种优化策略和技术，实现了在移动设备上高效而准确的图像识别。它在参数数量和计算复杂度方面较小，适用于资源受限的环境。

1.3 CoordConv 模块

由于在水下环境中，鱼体与背景之间的边界通常模糊不清，边缘信息不明显。通过引入位置信息，CoordConv 模块^[15]可以帮助算法更好地捕捉到鱼体边界的位置和形状，提高对鱼体的精确分割能力。

传统的卷积操作在处理图像时只考虑了像素的局部邻域信息，而忽略了像素的位置信息。然而，对于水下鱼体图像分割这样的任务，像素的位置信息对于区分不同鱼体、准确分割边界等至关重要。CoordConv 模块是一种通过引入坐标信息的卷积操作，能够提供更丰富的空间上下文信息，对于处理水下鱼体图像中的形态变化和纹理细节非常有帮助。

CoordConv 模块的基本思想是在卷积操作的输入特征图中增加两个额外的通道，分别表示像素的横坐标和纵坐标。这样，每个像素的输入特征向量就不仅包含了原始图像的颜色值，还包含了其在原始图像中的位置信息，如图 2 所示。在 CoordConv 模块中，这两个额外的通道被称为 X 通道和 Y 通道。

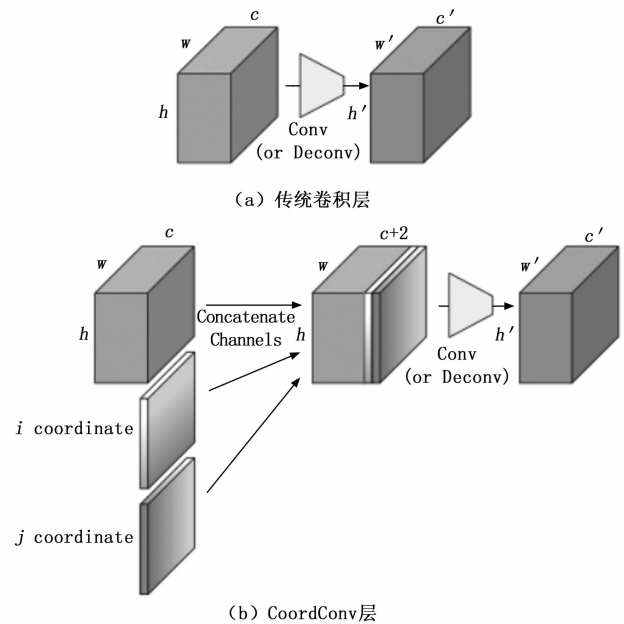


图 2 两种卷积层

具体地，CoordConv 模块可以表示为以下的数学形式，如公式 (1)、(2) 所示：

$$X_i = \frac{X}{W} \tag{1}$$

$$Y_i = \frac{Y}{H} \quad (2)$$

其中: X_i 和 Y_i 分别表示像素的归一化横坐标和纵坐标, X 和 Y 分别表示像素的原始横坐标和纵坐标, W 和 H 分别表示图像的宽度和高度。通过将归一化的坐标值作为额外的通道输入到卷积操作中, CoordConv 模块可以使网络更好地感知和利用像素的位置信息。

通过在我们的改进算法中引入 CoordConv 模块, 我们能够充分利用水下鱼体图像中的位置信息, 增强算法的鲁棒性和分割精度。该模块的引入为我们的轻量级水下鱼体图像分割算法提供了一种有效的机制, 使得算法能够更好地适应水下环境下的图像特点和挑战, 提高分割的准确性和稳定性。

1.4 注意力机制

传统的神经网络在处理输入时, 对所有的特征都以相同的权重进行处理, 无法有效地区分和利用不同特征之间的重要性。而注意力机制通过动态地调整特征的权重, 使网络能够自适应地关注输入中具有更高重要性的部分。本算法通过实验对比多种不同的注意力机制见 3.5 节, 发现引用“卷积块的注意力模块 (CBAM, convolutional block attention module)”^[17] 对该算法性能提升效果最好。如图 3 所示, CBAM 是一种综合了空间注意力和通道注意力的注意力模块, 用于增强卷积神经网络对空间和通道间相关性的关注, 它通过自适应地调整特征图的权重, 使网络能够更有针对性地利用输入特征中的重要信息。

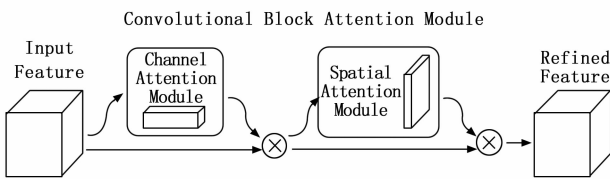


图 3 CBAM 注意力机制

空间注意力用于关注输入特征图的空间相关性。它通过对特征图在空间维度上进行池化操作, 得到每个空间位置的特征向量。然后, 通过全连接层学习每个位置的权重, 从而获取每个空间位置的重要性。最后, 通过广播乘法将空间注意力权重与特征图进行逐元素相乘, 使得网络能够在不同空间位置上有针对性地调整特征的权重。为了汇总空间特征, 主要采用了全局平均池化和最大池化这两种方法, 以利用不同的信息。

对于一个输入特征 F , 其尺寸为 $H \times W \times C$ 。首先, 我们对其进行全局平均池化和最大池化, 分别得到两个 $1 \times 1 \times C$ 的通道描述。接着, 这两个描述分别经过一个共享的两层神经网络处理: 第一层包含 C/r 个神经元, 激活函数为 Relu; 第二层包含 C 个神经元。此后, 得到的两个特征进行相加, 经过 Sigmoid 激活函数产生权重系数 M_c 。最终, 通过将权重系数 M_c 与原始特征 F 相乘, 得到按比例缩放的新特征, 详见公式 (3):

$$M_c(F) = \sigma\{MLP(Pool_{avg}(F)) + MLP[Pool_{max}(F)]\} = \sigma\{W_1[W_0(F_{max}^C)] + W_1[W_0(F_{max}^C)]\} \quad (3)$$

通道注意力用于聚焦输入特征图中的通道相关性, 首先通过在通道维度上进行全局平均池化操作, 提取每个通道的全局特征; 然后, 通过两个全连接层学习每个通道的权重, 从而获取每个通道的重要性; 最后, 通过广播乘法将通道注意力权重与特征图进行逐元素相乘, 使得网络能够自适应地调整通道的权重。

与通道注意力类似, 对于一个输入特征 F , 其尺寸为 $H \times W \times C$ 。首先分别对通道维度进行平均池化和最大池化操作, 以获得两个 $H \times W \times 1$ 的通道描述, 然后将这两个描述在通道上连接在一起。接下来, 通过一个 7×7 的卷积层, 使用 Sigmoid 作为激活函数, 计算出权重系数 M_s 。最终, 将权重系数与特征 F 相乘, 得到经过缩放的新特征, 具体参考式 (4):

$$M_s(F) = \sigma\{f^{7 \times 7}[Pool_{avg}(F), Pool_{max}(F)]\} = \sigma[f^{7 \times 7}(F_{avg}^S, F_{max}^S)] \quad (4)$$

通过综合空间注意力和通道注意力, CBAM 注意力模块能够在卷积网络的每个块 (block) 中增加对空间和通道相关性的关注。这种注意力机制使网络能够更好地捕捉输入特征中的关键信息, 减少对无关信息的依赖, 从而提升模型的性能和泛化能力, 如图 4 所示。

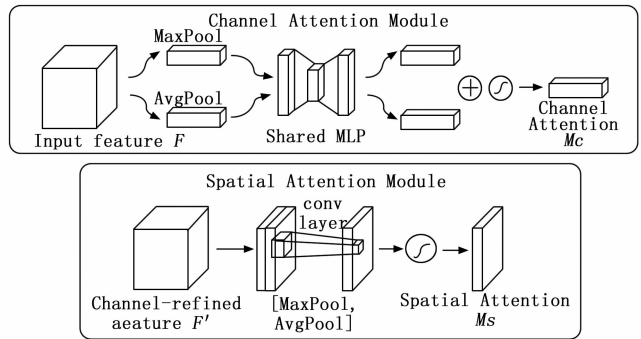


图 4 通道注意力和空间注意力模块

2 水下鱼体分割实验

2.1 实验环境

实验在 Linux 操作系统下, 基于 GPU、PyTorch 和 CUDA 框架完成的, 具体参数如表 1 所示。

表 1 实验平台软硬件配置

名称	版本
操作系统	Ubuntu 18.04
CPU	Inter(R) Xeon(R) Silver 4210R CPU @ 2.40 GHz
GPU	NVIDIA GeForce RTX 3080 Ti
Python	3.7.11
PyTorch	1.13.1+cu116
CUDA	11.4
CuDNN	8.2

2.2 数据集介绍

在这项研究中，网络使用 SUIM^[18]数据集进行了评估。SUIM 数据集共包含包括鱼类在内的多个类别的 1 525 张水下图像及其真实语义标签。SUIM 为每个类别提供带有单独注释的测试拆分，它还包括一个包含 110 张图像的测试集。因此，在实验中，我们使用鱼类和其他脊椎动物类别，并将这些数据用于鱼类分割。来自 SUIM 数据集的示例图像如图 5 所示。

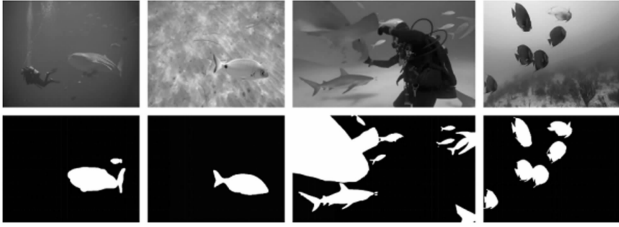


图 5 SUIM 样本图像与相应的分割标签图像。

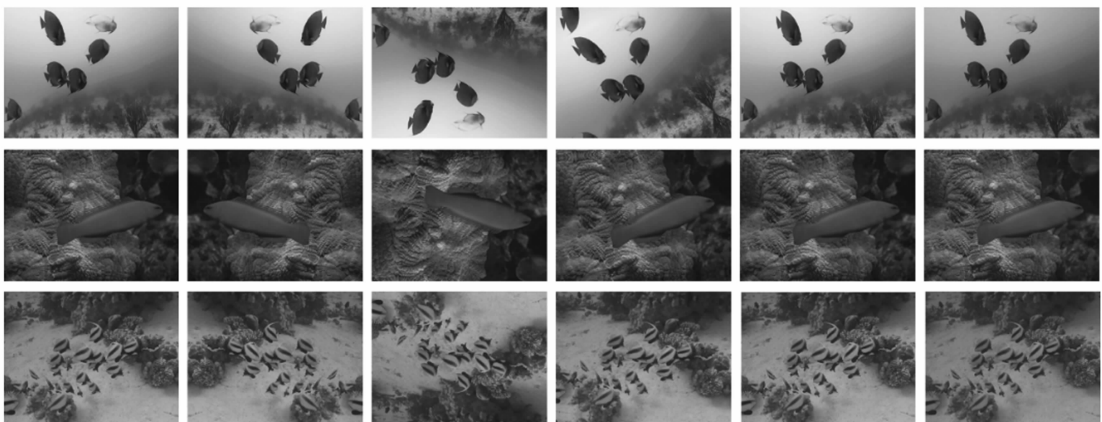
从鱼类和其他脊椎动物类别中挑选出鱼类图像仅 738 张，为提高模型精度和增加模型鲁棒性，对原始数据集进行样本增强^[19]，采用了基于几何变换、颜色空间变换 2 类数据扩充方案，具体方法包括水平翻转、垂直翻转、随机旋转、放大缩小、调整亮度和对比度这 5 种扩充方式，结果如图 6 所示。扩充后数据集总数为 4 428 张，其中训练集、验证集和测试集按照 8 : 1 : 1 进行分配。

2.3 实验评价指标

为验证本文方法的有效性，采用以下评价指标：平均像素准确率 (MPA, mean pixel accuracy)、平均交并比 (MIoU, mean intersection over union)、像素准确度等，同时使用参数量、FLOPS、模型的大小来衡量模型的复杂度。

MPA: 平均像素准确度是像素准确度的一种改进指标，考虑了每个类别在图像中的出现频率。它计算每个类别的像素准确度，然后对它们求平均，计算如式 (5) 所示：

$$MPA = \frac{1}{k} \sum_{i=1}^k \frac{p_{ii}}{\sum_{j=1}^k p_{ij}} \quad (5)$$



(a) 原图 (b) 水平翻转 (c) 垂直翻转 (d) 随机旋转 (e) 放大缩小 (f) 调整亮度和对比度

图 6 部分 SUIM 数据集数据扩充图片

MIoU: 交并比 (IoU, intersection over union) 是衡量两个集合重叠程度的指标。在语义分割中，交并比是指预测的分割结果与真实分割结果的交集与并集之比。平均交并比计算每个类别的交并比，然后对它们求平均，计算如式 (6) 所示：

$$MIoU = \frac{1}{K} \sum_{i=1}^k \frac{p_{ii}}{\sum_{j=1}^k p_{ij} + \sum_{j=1}^k p_{ji} - p_{ii}} \quad (6)$$

式中， k 表示像素的类别数； p_{ii} 表示实际类别为 i 、预测的类别也为 i 的像素的数目； p_{ij} 表示实际类别为 i 、预测的类别为 j 的像素的数目； p_{ji} 表示实际类别为 j 、预测的类别为 i 的像素的数目。

准确率 (accuracy) 表示预测结果中正确的占总预测值的比例，如式 (7) 所示：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

式中， TP 为被划分为正类且判断正确的个数； TN 为被划分为负类且判断正确个数； FP 为被划分为正类且判断错误的个数； FN 为被划分为负类且判断错误个数。

2.4 训练参数及模型参数设置

在训练过程中，我们采用了以下参数设置。我们选择了批量大小为 16，每次迭代使用 16 个样本进行模型参数的更新。我们的模型经过了 300 轮的训练，每一轮中模型遍历整个训练集一次，并进行参数更新。为了优化模型，我们采用了 Adam 优化算法，它是一种自适应学习率的优化算法，它结合了梯度的一阶矩估计 (均值) 和二阶矩估计 (方差)，可以更有效地调整学习率，加速模型的收敛过程。初学习率设置为 0.01，并通过余弦退火 (cosine annealing) 的方式，在训练过程中降低学习率。

由于水下鱼体语义分割是一个二分类问题，只有鱼体和背景两个类别，所以我们采用了二分类交叉熵损失函数，计算如式 (8) 所示。二分类交叉熵损失函数可以度量模型预测结果与真实结果之间的差异，并推动模型学习到更准确的预测。通过最小化交叉熵损失，模型能够逐渐调整权重和偏差，以使预测结果与真实结果尽可能接近。

$$L_{bce} = \frac{1}{N} \sum_i - \{y_i \cdot \log(p_i) + [1 - y_i] \cdot \log(1 - p_i)\} \quad (8)$$

式中, y_i 表示样本 i 的真实值, p_i 表示样本 i 预测为正类的概率, N 为图像像素点的总数目。

2.5 可视化训练过程

在我们的训练过程中, 我们追踪了损失函数随着训练迭代次数的变化。图 7 是在预先训练好的 MobileNetV2 网络模型的基础上添加水下鱼体样本进行再训练过程中的 loss 变化图。观察图 7 发现, 初始阶段, 损失函数迅速下降, 表明模型在学习数据的特征和模式; 在中期阶段, 我们可以观察到损失函数出现轻微的震荡或波动, 这可能是由于模型在权衡不同样本和特征的时候产生了一些波动, 但整体上保持在一个相对稳定的范围内; 最终, 损失函数在一个稳定的范围内波动, 不再有明显的下降或增加趋势, 说明在当前训练参数及模型参数设置下, 分割网络模型的性能也就接近最优。

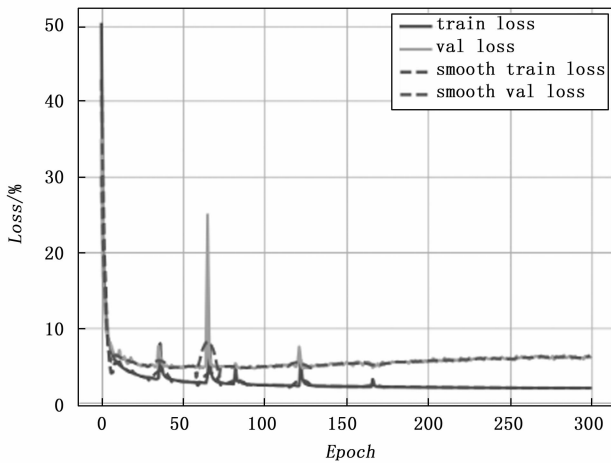


图 7 损失函数变化图

3 实验结果与分析

3.1 模型性能实验结果对比

在本研究中, 我们比较了 5 种不同模型的性能, 即 UNet^[20]、DeepLabv3+^[21]、HRNet^[22]、PSPNet 和 Co-PSPNet, 对水下鱼体分割数据集进行训练、验证和测试, 测试集的 $mIoU$ 、 mAP 和 $Accuracy$ 结果如表 2 所示。

表 2 不同模型的测试结果

模型	特征提取器	$mIoU/\%$	$mPA/\%$	$Accuracy/\%$
UNet	VGG16	87.16	91.89	96.86
DeepLabv3+	Xception	89.92	95.03	97.50
HRNet	hrnetv2_w18	91.64	95.21	97.99
PSPNet	Resnet50	90.57	92.63	97.24
Co-PSPNet	MobileNetV2	93.21	95.59	98.07

由表 2 可知, Co-PSPNet 模型的 $mIoU$ 达到 92.21%, 相比于常用的语义分割网络模型 UNet、DeepLabv3+、HRNet 和 PSPNet 分别高出 5.05、2.29、0.57 和 1.64 个

百分点。首先, UNet 模型使用 VGG16 作为特征提取器, 在 $mIoU$ 、 mPA 和 $Accuracy$ 方面表现良好, 但相较于其他模型, 其性能稍显不足。DeepLabv3+ 模型采用了 Xception 作为特征提取器, 并取得了出色的性能, 超越了 UNet 模型在各项指标上的表现。HRNet 模型以 hrnetv2_w18 作为特征提取器, 获得了相对较高的得分, 这归功于其高分辨率的特征融合和多尺度处理能力。PSPNet 模型利用 Resnet50 作为特征提取器, 虽然在 $mIoU$ 和 $Accuracy$ 方面略低于其他模型, 但在 mPA 方面表现相对较好。然而, Co-PSPNet 模型采用了轻量级的 MobileNetV2 作为特征提取器, 并在所有评价指标上取得了最佳性能。这表明我们的改进措施在水下鱼体分割任务中是有效的。综上所述, 通过对这些模型的性能进行比较分析, Co-PSPNet 模型以其出色的性能在水下鱼体分割任务中脱颖而出。

3.2 模型复杂度比较

在本研究中, 我们比较了 5 种不同模型的复杂度, 即 UNet、DeepLabv3+、HRNet、PSPNet 和 Co-PSPNet。表 3 总结了这些模型的参数数量、浮点运算数 (FLOPs) 以及模型大小。相比之下, 我们提出的 Co-PSPNet 的模型在复杂度方面表现出明显的优势。我们的模型具有更少的参数数量 (仅为 3.3 M)、更低的 FLOPs (760.14 M) 和更小的模型大小 (13.5 MB)。这意味着我们的模型在计算资源和存储空间方面要求较低, 更加轻量级, 适合在资源有限的水下环境中应用。

表 3 5 种模型的复杂度对比

模型	特征提取器	Params/M	FLOPs/M	Model size/M
UNet	VGG16	24.89	43 226.42	99.6
DeepLabv3+	Xception	54.71	15 967.63	219.9
HRNet	hrnetv2_w18	9.64	3 571.57	39.4
PSPNet	Resnet50	46.71	11 376.35	187.2
Co-PSPNet	MobileNetV2	3.3	760.14	13.5

3.3 分割结果可视化

各模型部分输出结果可视化如图 8 所示, 从图中可以清楚地看出, 在水下鱼体语义分割任务中, Co-PSPNet 模型相较于 UNet、DeepLabv3+、HRNet 和 PSPNet 模型, 呈现出更全面、准确的检测结果。

首先, 通过观察图 8 中的可视化结果, 我们可以发现改进的模型相对于 UNet 和 DeepLabv3+ 模型, 能够更好地提取和利用特征信息, 提供更精确的分割结果。

其次, 相较于 HRNet 模型, 改进的模型在水下鱼体语义分割任务中展现出更好的综合性能。尽管 HRNet 模型在捕捉图像中的全局和局部信息方面表现出色, 但在鱼体的特征表达和语义分割方面存在一定的限制。相比之下, 改进的模型通过引入新的特征提取和融合注意力机制, 能够更好地结合全局和局部信息, 提高鱼体语义分割的准确性和鲁棒性。

最后, 与 PSPNet 模型相比, 改进的模型在处理不同尺

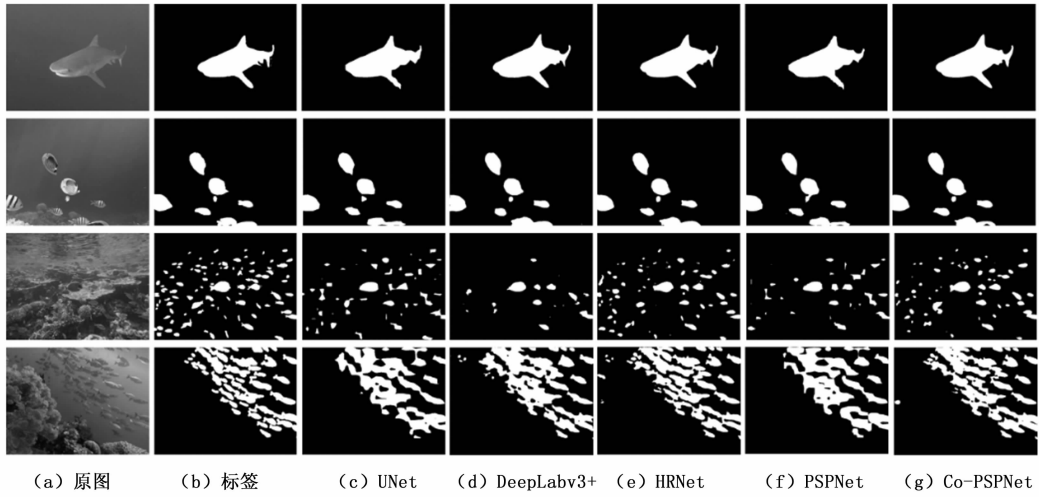


图 8 不同算法的可视化结果

度鱼体时具有更好的适应性。水下环境中，鱼体的尺寸和形状可能存在较大的变化。PSPNet 模型使用金字塔池化机制来捕捉不同尺度的上下文信息，但在处理尺度差异较大的鱼体时可能存在一定的限制。改进的模型通过引入 CoordConv 模块和注意力机制，能够更好地处理尺度变化，提高鱼体语义分割的多样性和泛化能力。

综上所述，通过图 8 的可视化结果，我们可以得出结论：在水下鱼体语义分割任务中，Co-PSPNet 具有更全面、准确和稳健的性能。这些结果为改进模型在水下生态研究、水下监测和保护等领域的应用提供了坚实的基础。未来的研究可以进一步探索改进模型的潜力，并将其应用扩展到更广泛的水下场景，如海洋生物学研究、水下资源勘探和水下遗址保护等，为未来水下目标检测和分割技术的发展提供了新的思路 and 方向。

3.4 不同主干网络的对比实验

为了能够在有限硬件条件下进行鱼体图像分割，我们通过选择轻量级主干网络对 PSPNet 网络进行改进，主要对比了 ShuffleNetV1^[23]、ShuffleNetV2^[24]、MobileNetV1^[25]、MobileNetV2、MobileNetV3^[26] 网络，并在 SUIM 数据集上进行了实验。

表 4 不同主干网络性能对比

Backbone	<i>MIoU</i> / %	<i>mPA</i> / %	<i>Accuracy</i> / %	<i>Params</i> / M	<i>FLOPs</i> / M	<i>Model size</i> / M
ShuffleNetV1	80.33	86.03	95.09	7.02	2 233.64	32.4
ShuffleNetV2	82.74	87.44	95.79	4.4	396.38	18.5
MobileNetV1	80.6	85.81	95.22	8.89	832.58	40.2
MobileNetV2	86.5	91.26	96.69	2.38	578.23	9.8
MobileNetV3	83.5	87.53	95.26	2.96	230.59	22.7

从表 4 的结果来看，将 MobileNetV2 网络作为 PSPNet 的主干网络获得了最好的性能。此外，在参数量和模型大小方面也具有优势。虽然在 FLOPs 方面相对于 ShuffleNetV2 和 MobileNetV3 稍有逊色，但综合来看，MobileNetV2 网络在 PSPNet 中表现较为出色。

3.5 不同注意力机制的对比实验

注意力机制本质是通过计算相应的权重值，让卷积神经网络识别出需要重点关注的有用特征向量，忽略不重要的特征信息。从而在避免无用特征干扰拟合结果的同时，还对运算速度有一定的改善。我们基于 MobileNetV2 主干网络，在特征金字塔后面引入 3 个不同的注意力机制进行对比，分别为“坐标注意力机制 (CA, coordinate attention)”^[27]、“全局注意力机制 (GAM, global attention mechanism)”^[28]、CBAM 注意力机制。

表 5 不同注意力机制性能对比

Backbone	Attention	<i>MIoU</i> / %	<i>mPA</i> / %	<i>Accuracy</i> / %
MobileNetV2	CA	91.33	94.85	97.93
MobileNetV2	GAM	91.31	94.8	95.91
MobileNetV2	CBAM	91.47	94.97	97.96

通过表 5 中对比实验结果，我们可以得出以下结论：引入注意力机制可以有效地提升水下鱼体图像分割算法的性能。不同的注意力机制在性能上有所差异，CBAM 注意力机制表现出最佳的性能。这表明加入通道和空间注意力可以更好地捕捉到关键鱼体特征，提高分割的准确性和鲁棒性。

3.6 不同模块的消融实验

为证明 MobileNetV2、CoordConv 模块、CBAM 注意力机制等各模块的有效性，利用控制变量法设计了 4 组消融实验，以 *mIoU*、*mAP* 和 *Accuracy* 作为实验评价指标，实验数据如表 6 所示。其中采用的是预训练的 MobileNetV2 主干网络。

表 6 不同模块的消融实验结果

MobileNetV2	CoordConv	CBAM	<i>mIoU</i> /%	<i>mPA</i> /%	<i>Accuracy</i> /%
✓			91.09	94.67	97.86
✓	✓		91.94	95.43	98.0
✓		✓	91.47	94.97	97.96
✓	✓	✓	93.21	95.59	98.07

4 结束语

在本研究中, 我们提出了一种基于 Co-PSPNet 网络的轻量级水下鱼体图像分割算法, 通过将 PSPNet 的主干网络替换为 MobileNetV2, 加入 CoordConv 模块和注意力机制模块, 我们成功地在保证精度的同时降低了网络的计算复杂度和参数量, 从而提高了算法的效率, 使其能够更好地适应水下环境的特点。

通过在公开的水下鱼体图像数据集上进行大量实验, 我们验证了我们提出的算法在鱼体图像分割任务上的优越性能。实验结果表明, 我们的方法能够有效地从水下图像中提取出鱼体的准确轮廓, 为水下生态研究和水下机器人的应用提供了重要支持。

尽管我们的算法取得了令人满意的结果, 但仍存在一些改进的空间。一个方向是实现多种类的分割, 即将算法扩展到识别和分割水下环境中的多种鱼类, 当前的研究侧重于鱼体图像分割任务, 但在实际应用中, 可能会遇到多种鱼类共存的情况, 因此, 将我们的算法扩展到多种类的分割将进一步提高其实用性和适应性。还有就是可以进一步优化注意力机制的设计, 以提高网络对关键鱼体特征的关注程度。

参考文献:

[1] 虞净萍, 张文亮. 水下摄影监测生态水路枯水期亚临界流况鱼类族群的分布 [J]. 农业工程学报, 2010, 56 (4): 83-97.

[2] BINGHAM B, FOLEY B, SINGH H, et al. Robotic tools for deep water archaeology: surveying an ancient shipwreck with an autonomous underwater vehicle [J]. Journal of Field Robotics, 2010, 27: 702-717.

[3] 林 森, 赵 颖. 水下光学图像中目标探测关键技术研究综述 [J]. 激光与光电子学进展, 2020, 57 (6): 18-29.

[4] CHEN L C, PAPANDEOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40 (4): 834-848.

[5] BADRINARAYANAN V, HANDA A, CIPOLLA R. Segnet: a deep convolutional encoder decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (12): 2481-2495.

[6] 杜海顺, 汪凤泉. 一种快速的模糊 C 均值聚类彩色图像分割方法 [J]. 计算机工程与应用, 2009, 45 (33): 138-140.

[7] 沈志忠, 王 硕, 曹志强, 等. 基于边缘检测的图像分割方法及其在机器鱼中的应用 [J]. 机器人, 2006, 28 (4): 361-366.

[8] 刘 超, 蔡文华, 陆 玲. 图像阈值法分割综述 [J]. 电脑知识与技术: 学术版, 2015 (1): 140-142.

[9] BÖER G, VEERAMALLI R, SCHRAMM, H. Segmentation of fish in realistic underwater scenes using lightweight deep learning models [J]. Proceedings of the 2nd International Conference on Robotics, Computer Vision and Intelligent Systems, 2021: 158-164.

[10] ZHANG W, WU C, BAO Z. DPANet: Dual pooling-aggregated attention network for fish segmentation [J]. IET Computer Vision, 2022 (1): 16.

[11] CHANG C C, WANG Y P, CHENG S C. Fish segmentation in sonar images by mask R-CNN on feature maps of conditional random fields [J]. Sensors, 2021, 21 (22): 7625.

[12] ENVELOPE J H C, ENVELOPE L V M, ENVELOPE O R. Deep learning based segmentation of fish in noisy forward looking MBES images [J]. IFAC-PapersOnLine, 2020, 53 (2): 14546-14551.

[13] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881-2890.

[14] HU L, GE Q. Automatic facial expression recognition based on MobileNetV2 in real-time [J]. Journal of Physics Conference Series, 2020, 1549: 022136.

[15] EL JURDI R, PETITJEAN C, HONEINE P, et al. CoordConv-Unet: investigating coordconv for organ segmentation [J]. Innovation and Research in Biomedical Engineering: IRBM, 2021 (6): 42.

[16] 高淑萍, 赵清源, 齐小刚, 等. 改进 MobileNet 的图像分类方法研究 [J]. 智能系统学报, 2021, 16 (1): 11-20.

[17] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C] // Proceedings of the European Conference on Computer Vision, 2018: 3-19.

[18] ISLAM M, J EDGE C, XIAO Y, et al. Semantic segmentation of underwater imagery: dataset and benchmark [C] // Proceeding of the IEEE International Conference on Intelligent Robots and Systems, Las Vegas, NV, USA, 2020: 1769-1776.

[19] 邢予权, 潘今一, 王 伟, 等. 基于语义分割与迁移学习的手势识别 [J]. 计算机测量与控制, 2020, 28 (4): 196-199.

[20] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C] // International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015: 234-241.

[21] CHEN L C, ZHU Y, PAPANDEOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C] // Proceedings of the European Conference on Computer Vision, 2018: 801-818.

(下转第 308 页)