

基于改进 SSD 的水下目标检测算法研究

王伟¹, 贺昱曜¹, 郭玉锦², 李宝奇^{3,4}, 何灵蛟¹

(1.西北工业大学 航海学院, 陕西 西安 710072; 2.西安轻工业钟表研究所有限公司, 陕西 西安 710061; 3.中国科学院 声学研究所, 北京 100190; 4.中国科学院先进水下信息技术重点实验室, 北京 100190)

摘要:随着人类对海洋的不断深入探索,准确、快速地检测水下环境中的鱼类、仿生体及其他智能体对完善水下防御体系显得越来越重要。针对水下复杂环境下目标检测准确率低、实时性差的问题,提出一种基于改进 SSD 的目标检测算法。该算法用 ResNet 卷积神经网络代替 SSD 的 VGG 卷积神经网络作为目标检测的基础网络,并在基础网络中利用所提出的深度分离可变形卷积模块进行特征提取,提高对水下复杂环境下目标检测的精度及速度。所提出的深度分离可变形卷积主要是在可变形卷积获取卷积核偏移量的过程中融合深度可分离卷积,以减少参数量来达到提升网络运行速度的目的,同时通过稀疏表示来提升网络的鲁棒性。实验结果显示,相比 ResNet 作为基础网络的 SSD 检测模型,利用深度分离可变形卷积改进的 SSD 检测模型检测水下目标的准确率提升了 11 个百分点,检测时间减少了 3 ms,证明新算法的有效性。

关键词:水下目标检测;SSD;深度可分离卷积;可变形卷积

中图分类号:TP273 **文献标志码:**A **文章编号:**1000-2758(2020)04-0747-08

水下目标识别对于海洋资源开发、海洋环境监测、海洋军事应用等具有重大意义^[1]。在实际应用中,由于水下环境的复杂性,增加了水下目标的检测难度。

目前典型的水下目标检测算法有 Fabic 等^[2]提出的基于形状分类的算法,该算法利用斑点计数法和形状分析法对水下视频序列中的鱼类进行分类,但其针对性强,缺乏通用性,难以应用于其他目标。Oliverm 等^[3]利用哈里斯和黑森检测器,对物体进行区域检测与匹配来检测水下目标,但当水体存在非均匀光照时,该算法性能衰减很快。

图像的检测、识别主要涉及行人检测^[4]、面部识别^[5]、图像分割^[6]等。在众多图像的检测、识别算法中,卷积神经网络(convolutional neural network, CNN)表现尤为突出。卷积神经网络^[7]是通过端到端的方法进行训练的机器学习模型,该模型采用梯度下降法来对参数进行调节。模型首先需要在特定数据集上进行训练,在训练过程中,模型可以逐渐学习提取图像目标特征,最后利用提取的特征完成图

像分类、目标检测或语义分割等任务。Wang 等^[8]提出基于尺度空间法的卷积神经网络目标识别算法,该算法在目标检测准确率方面优于传统算法,但其检测实时性较差,并且对复杂环境下水下目标的检测准确率有待进一步提高。

针对文献[8]存在的水下复杂环境中目标检测准确率低、实时性差的问题,本文在可变形卷积(deformable convolutional)^[9]的基础上融合深度可分离卷积(depthwise separable convolution)^[10]思想,首先提出一种称为深度分离可变形卷积的新型卷积结构,然后对 SSD 目标检测模型进行优化改进并提出一种基于改进 SSD 的水下目标快速检测算法(DD-SSD)。该算法首先利用残差卷积神经网络(ResNet)代替原始 SSD 目标检测模型的基础网络 VGG-16,然后在网络后端用深度分离可变形卷积适当替换标准卷积层,在提高网络模型检测精度的同时保证了检测速度。该算法模型为真正端到端的检测模型,可以直接将水下目标图像或者视频输入训练好的网络模型进行目标对象的检测。实验结果证

明,本文所提算法对复杂环境下的水下目标检测准确率有较大提高,而且检测速度也明显提高。

1 深度分离可变形卷积

可变形卷积网络是卷积神经网络的一种变体,他对于解决复杂的视觉任务,学习密集空间变化非常有效。深度可分离卷积主要通过对空间信息和深度信息进行去耦合来达到模型加速的目的,在实际应用中可以进一步减少模型的检测时间。本文主要在可变形卷积过程中融合深度可分离卷积,提出了深度分离可变形卷积模块,它可以使网络模型的特征提取能力进一步增强,同时减少网络模型的参数量,降低模型的大小,提高网络模型的运行速度。

1.1 可变形卷积

传统卷积神经网络的卷积核的几何结构是固定的,例如使用等宽等高的正方形卷积核,这种卷积方式限制了网络模型几何变换建模的能力。如图 1a)所示,3 * 3 的卷积核大小、形状固定,卷积过程中几何变换建模的能力不足。可变形卷积网络模型^[9]主要针对卷积神经网络普遍对几何变换建模能力的不足,提出一种新思路,即在标准卷积核上加入一个可以学习的偏移量使采样网络可以自由变形。如图 1b)~1d)所示,可变形卷积核的每个元素都按一定规则偏移,采样位置也就随之变化,从而提升卷积神经网络的几何变换建模能力,提高对不规则物体、非刚性目标及复杂环境目标的识别、检测能力。

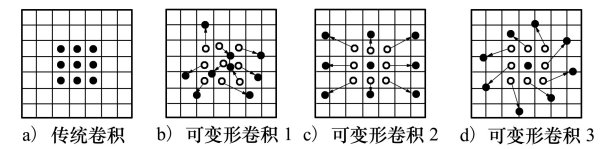


图 1 3 * 3 卷积核采样点的位置图

1.2 深度可分离卷积

Howard 等于 2017 年提出 MobileNet^[11] 移动端、轻量化的模型,该网络模型由 13 个卷积层、平均池化层、全连接层以及输入输出层组成,其核心是深度可分离卷积。深度可分离卷积主要将传统卷积分解为一个深度卷积和一个 1 * 1 的传统卷积,这样处理有助于减少模型参数、降低计算量。深度可分离卷积如图 2 所示。

图 2a)表示传统的标准卷积,图 2b)、2c)表示

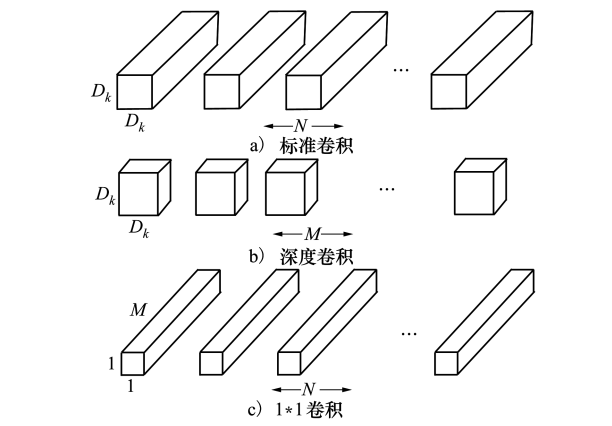


图 2 标准卷积与深度可分离卷积说明

深度可分离卷积。如果卷积模块的输入为 $D_F \cdot D_F \cdot M$,输出为 $D_F \cdot D_F \cdot N$,卷积核为 $D_K \cdot D_K$,那么标准卷积的计算量为

$$D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F \tag{1}$$

而深度可分离卷积的计算量为

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F \tag{2}$$

从(1)至(2)式可以明显看出深度可分离卷积通过解耦空间信息和深度信息,使模型参数量大大减少,提升了模型运算效率。

1.3 深度分离可变形卷积

针对可变形卷积会增加卷积网络的运行时间,以及对复杂场景特征提取能力不足的问题,本文提出深度分离可变形卷积。可变形卷积通过加入偏移卷积实现标准卷积核变形学习的能力。但在偏移卷积的过程中,可变形卷积使用的是一般的标准卷积,本文所提的深度分离可变形卷积,是在可变形卷积学习偏移量的过程中将卷积的空间和通道进行分离,一方面深度可分离卷积适当减少模型参数,使模型具有更强的鲁棒性,提高了卷积网络模型的特征提取能力;另一方面通过减少模块参数量可以加快模块运行速率,从而提升整个卷积网络模型的运行速度。

深度分离可变形卷积整个过程可分为 3 步:首先在输入特征图 x 上用深度可分离卷积进行采样,获得 x 上每个像素点的偏移量;然后利用双线性插值算法得到每个像素点偏移的像素量,相当于在卷积核上产生形变达到不固定形状采样的目的;最后在输入特征图 x 上用带有偏移量的卷积核进行可变形卷积操作,即将带有权重的采样值进行加和计算。

假设输入特征图尺寸为 $D_F \cdot D_F$, 数量为 M , 深度卷积核大小为 $D_K \cdot D_K$, 数量为 N , 经过深度可分离卷积后的特征图 y_{offset} 大小不变, 为 $D_F \cdot D_F$, 数量为 $2N$, 输出 y_{offset} 是指输入特征图 x 中每个像素点的偏移量。这里得到的偏移量不是整数, 需要利用双线性插值算法进一步得到偏移量在输入特征图上对应的偏移像素量。采用双线性插值算法既可以得到尽可能精确的像素偏移量, 又可以使网络进行反向传播。最后在带有偏移像素量的输入特征图 x_{offset} 上进行卷积操作, 相当于在原始特征图 x 上利用可变形的卷积核进行卷积操作。整个深度分离可变形卷积过程可用 (3) 式表示

$$y(p_0) = \sum_{p_n \in S} w(p_n) \cdot \sum_q G(p, q) \cdot x(q) \quad (3)$$

式中: p_0 表示输出特征图 y 中的每个像素点位置; S 表示卷积核权重矩阵; p_n 代表矩阵中的每个权重值; G 代表二维双线性插值核; $p = p_0 + p_n + \Delta p$ 代表通过深度卷积学习到的随机偏移量; q 为映射到输入特征图 x 的整体空间位置。

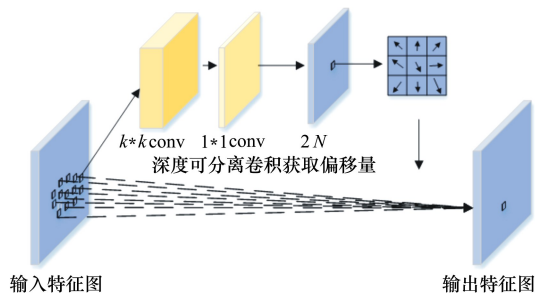


图 3 深度分离可变形卷积

图 3 为深度分离可变形卷积的卷积过程示意图, 在输入层的特征映射基础上利用深度可分离卷积获得偏移采样特征提取, 然后通过双线性插值的

算法将带有偏移量的采样点集中, 由于变形后的卷积核与原卷积核具有相同的空间分辨率及扩张维度, 所以输出偏移域和输入映射具有相同的空间分辨率, 即 N 个二维通道分别对应 N 个二维偏移量, 在网络模型的训练过程中, 梯度通过双线性运算进行反向传播, 原卷积核和偏移卷积核共同学习用于生成输出特征的权值参数量。

深度分离可变形卷积参数量相比可变形卷积有所减少, 网络运行速度加快, 同时由于参数量的减少使网络稀疏化, 网络的特征提取能力进一步加强。

本文在 ResNet-50 网络模型的基础上进行改进, 主要在网络后端利用改进的可变性卷积替换标准卷积, 目的是进一步提高特征提取能力, 同时减少模型参数, 最终使网络模型在保证检测精度的情况下达到实时性检测的要求。

2 基于改进 SSD 的水下目标检测算法

2.1 SSD 目标检测模型

与基于候选区域 (region proposal) 的目标检测模型 (如 RCNN、Fast-RCNN、Faster-RCNN) 不同, SSD (single shot multiBox detector)^[12] 是基于多边框 (bounding boxes) 回归的目标检测模型, 它是在 YOLO^[13] 的基础上发展而来的。Faster RCNN^[14] 需要 2 步完成检测, 首先通过 CNN 得到候选框, 然后再进行分类与回归; 而 YOLO 与 SSD 只需要一步就可以完成检测。另外, YOLO 需要在全连接层后检测, 而 SSD 直接采用卷积层进行检测。

SSD 以 VGG-16 为基础模型, 在其基础上新增加了 4 个卷积层来获取不同尺度的特征图, 其结构示意图如图 4 所示。

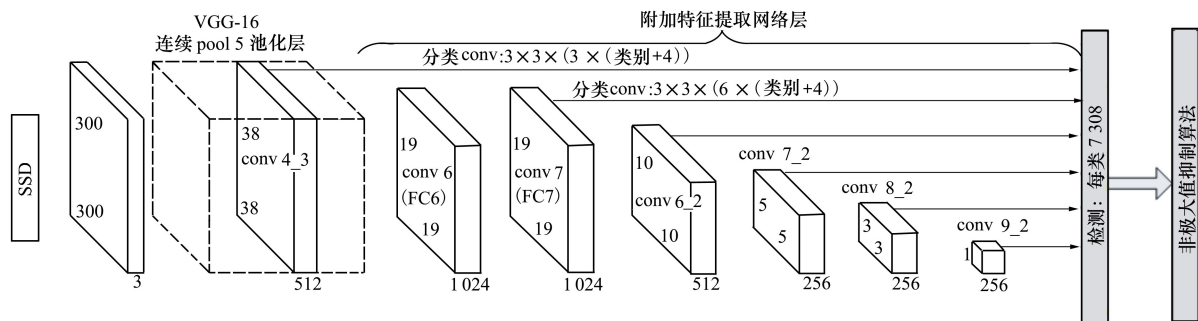


图 4 SSD 网络结构图

SSD 模型的输入图像尺寸主要为 300×300 , 图像经过输入层进入 VGG-16 卷积网络进行特征提取, 最后通过输出层计算损失量, 其损失函数包含两部分, 分别是定位损失和回归损失, 这两部分共同对网络结构进行评估, 总体损失还是如公式(4)所示

$$L(x, c, a, b) = \frac{1}{N}(L_{\text{conf}}(x, c)) + \gamma L_{\text{loc}}(x, a, b) \quad (4)$$

(4) 式的第一部分 $\frac{1}{N}(L_{\text{conf}}(x, c))$ 表示置信度损失, 其中 c 表示置信度, N 表示与完整目标框匹配的框数; 第二部分 $\gamma L_{\text{loc}}(x, a, b)$ 表示定位损失, 其中 x 表示目标框的中心坐标, a 表示框的宽度, b 表示框的高度。该网络通过 3×3 的卷积核进行卷积操作, 最后分类器判别, 输出 8 732 个结果, 包括类别概率及特征图单元框回归后的坐标值, 从而完成对图像中的目标进行检测的任务。

2.2 特征提取网络的改进

通过增加卷积神经网络的深度可以提升网络的特征提取能力, 而随着网络的加深会出现梯度消失致使网络很难训练。针对这个问题, He 等^[11]提出了残差网络 ResNet, 它在网络不断加深的时候使网络训练变得容易, 提高了卷积神经网络的特征提取能力。

由于水下环境的复杂性及目标对象的特征多变性, 本文首先利用 ResNet-50 替换 SSD 的特征提取基础网络 VGG-16 用以提升一般对象的特征提取能力, 然后将 ResNet-50 网络的最后 3 个卷积层 (Conv5) 换成可变形卷积模块, 进一步提升水下复杂环境下柔性目标的特征提取能力, 最后在可变形卷积的 offset 过程中融合深度可分离卷积的思想, 达到降低网络运行时间的目的; 同时利用深度可分离卷积的稀疏表示增加网络的鲁棒性进一步提升网络特征提取能力。表 1 为改进 SSD 基础网络模型的结构, 其中 conv 5 采用本文提出的深度分离可变形卷积模块。

将深度可分离卷积思想融入可变形卷积中的 offset 过程中, 可以提升可变形卷积模块对非刚性目标复杂特征的提取能力, 同时可以进一步减少模型参数, 保证整体模型的鲁棒性的同时提高模型运行速度。

表 1 特征提取网络结构

层数	输出尺寸	结构/类型
conv 1	114 * 114	7 * 7, 64, stride 2
pool 1	57 * 57	3 * 3, max, stride 2
conv 2	57 * 57	$\begin{pmatrix} 1 * 1, 64 \\ 3 * 3, 64 \\ 1 * 1, 128 \end{pmatrix} \times 3$
conv 3	29 * 29	$\begin{pmatrix} 1 * 1, 128 \\ 3 * 3, 128 \\ 1 * 1, 512 \end{pmatrix} \times 4$
conv 4	15 * 15	$\begin{pmatrix} 1 * 1, 256 \\ 3 * 3, 256 \\ 1 * 1, 1\ 024 \end{pmatrix} \times 6$
conv 5	8 * 8	$\begin{pmatrix} 1 * 1, 256 \\ 3 * 3, 256, \text{ddf} \\ 1 * 1, 1\ 024 \end{pmatrix} \times 3$
pool 5	2 * 2	7 * 7, 平均池化, 步长 1
fc1000	/	InnerProduct

2.3 基于改进 SSD 的目标检测模型 (DD-SSD)

针对复杂环境下的水下目标检测任务, 本文提出一种优化的基于改进 SSD 的目标检测模型 DD-SSD。该网络模型结构首先用 2.2 节改进的特征提取网络替换原 SSD 的特征提取网络, 在减少模型参数、提高模型检测速度的同时, 通过自学习产生形变的不规则卷积核对复杂环境下的水下目标进行更好的特征提取, 增加模型在复杂环境下的水下目标检测精度; 然后在特征提取网络后端增加 4 个卷积层, 充分利用不同尺度下的特征图进行检测。改进后的网络结构如图 5 所示。

在基于改进 SSD 的目标检测网络模型 DD-SSD 中, 图像首先输入到改进的特征提取网络, 利用深度可分离可变形卷积提升网络对复杂特征的提取能力, 然后在网络后端增加 4 个卷积层, 并抽取 ResNet-50 后端的 2 个卷积层 (res4f、res5c) 与模型最后的 4 个卷积共同组成 6 个卷积层作为检测目标使用。6 个卷积层对应特征图的尺度依次减小, 并使用不同的预测锚框。其中在第一个卷积层 (res4f) 对应特征图的每个像素点设置 4 个锚框; 第二个卷积层 (res4f) 对应特征图的每个像素点设置 6 个锚框。依次, 第三、第四个卷积层对应特征图的每个像素点设置 6 个锚框, 最后 2 个卷积层设置 4 个

锚框。锚框大小的生成规则与原始 SSD 一样。深度分离可变形卷积在保证目标特征信息完整的前提下,一方面增加了模型的多样性、压缩了模型计算量,同时减少了模型结构,使模型检测实时性更好;

另一方面增加了模型几何变换的建模能力,对复杂场景下的图像目标特征提取能力更强,可以进一步提升复杂环境下的水下目标检测精度。

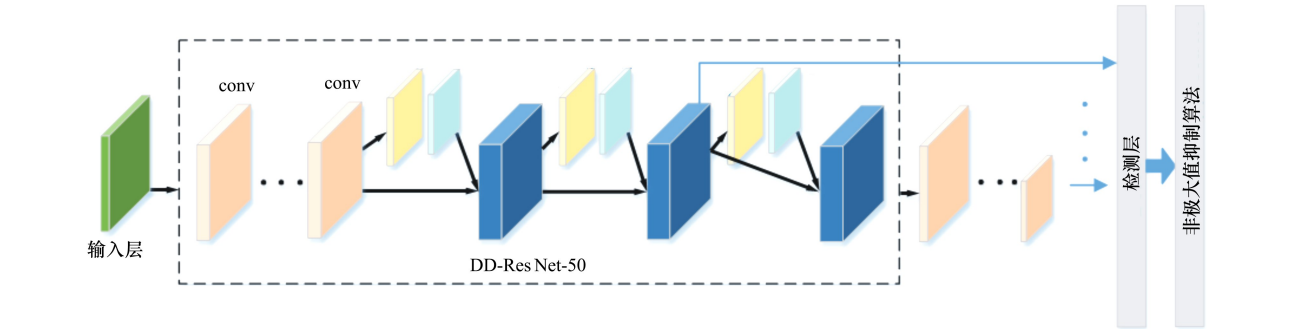


图 5 DD-SSD 网络结构

3 实验及结果分析

为了验证所提算法的优越性,本文设计了以下实验。首先采集水下柔性目标图像进行标注,制作实验数据集。本文选取了 6 种水下生物模型作为检测的目标,然后通过对比在不同网络模型下的训练、测试结果验证 DD-SSD 模型对水下目标检测的准确性和实时性。实验采用 GPU 训练方式,并利用 CUDNN 进行加速处理。实验计算机所采用 CPU 为 core i3 处理器,GPU 为 GTX1070。

实验中,首先利用公开数据集 ImageNet 对改进的特征提取网络模型预训练,然后用自制水下目标数据集对基于改进 SSD 的水下目标检测模型进行训练,并测试模型检测速度及检测精度。最后,在公开数据集 VOC 上对用于检测的单层和多层附加卷积层进行实验,进一步验证本文所提算法的有效性。

3.1 实验数据集

实验采集了 6 种水下柔性目标共 1 133 张图片进行了标注工作,然后按比例进行划分,其中测试数据 227 个,训练数据 906 个。表 2 为详细实验采集数据。

表 2 实验数据集						
类别	shrimp	crab	dolphin	Shark1	Shark2	Shark3
数量	195	183	133	205	196	221

3.2 改进前后基础网络对图像分类准确率的对比

实验中,首先对各目标检测的基础网络进行图像分类的模型预训练。训练、测试数据集选用 Imagenet 数据集,学习率设置为 0.05,训练方式选择 multistep,训练批次 batch_size 为 14,最大训练次数 max_iter 为 600 000,每训练 10 000 次进行一次测试,最终各目标检测的基础网络对 Imagenet 图像数据集分类的测试结果如表 3 所示。

实验结果表明加入可变形卷积对图像分类精度有一定提升,但是在融入深度可分离卷积思想后的改进网络结构的图像分类精度最高,比原始 ResNet-50 提升超过 2 个百分点,比加入可变形卷积的 ResNet-Deform 网络结构提升超过 1 个百分点,说明改进的基础网络特征提取能力更强。

表 3 基础网络图像分类结果比较(分类准确率) %			
分类准确度	ResNet	ResNet-Deform	ResNet-DD
TOP1	57.04	58.30	59.84
TOP5	80.75	81.63	82.79

3.3 改进前后模型检测精度及检测速度的实验对比

实验模型的训练及测试采用自建的水下目标数据集,在相同的训练环境及训练方式下,对各实验模型各训练 10 000 次,同时,实验对各模型均进行 100 次的运行时间测试,最后取平均值作为最终模型单次运行时间,表 4 为实验的最终结果。

表 4 检测精度及速度比较

模型结构	检测精度/%	检测时间/ms	模型参数/MB
ResNet-SSD	83.55	48.94	127
ResNet-Deform-SSD	90.65	49.42	131
DD-SSD	95.06	45.91	107.2

通过表 4 可以发现,加入可变形卷积后的网络模型(ResNet-Deform-SSD)对复杂水下的柔性目标检测精度比基础网络为 ResNet 的 SSD 网络模型(ResNet-SSD)提高了 7 个百分点,而最终优化后的网络模型(DD-SSD)检测精度最高,达到 95.06%;同时,可以发现加入可变形卷积之后网络模型(ResNet-Deform-SSD)的运行时间比基础网络为 ResNet 的 SSD 网络模型(ResNet-SSD)增加了 0.5 ms,模型参数量增加了 4 MB,而经过加入深度可分离卷积最终优化后的网络模型(DD-SSD)运行时间减少了 3.5 ms,参数量减少 23.8 MB,检测帧率约为 22 frame/s,基本满足实时目标检测要求,实验总体结果符合预期目标。

图 6 为水下目标检测测试数据的检测效果图,其中图 6a)为原图,图 6b)为优化后网络模型(DD-SSD)的检测效果图。从图中可以看到,图 6a)中的



图 6 实验检测效果图

水下目标被图 6b)中的矩形框准确框出来,同时在目标框的左上角显示目标种类及可信度。说明本文提出的基于改进 SSD 的水下目标检测模型对水下目标检测的有效性和准确性,可以实现对特定水下环境目标的检测。

3.4 不同卷积层的检测实验结果

为了进一步说明改进可变形模块对网络特征提取有提升能力,实验选取不同层对比 VOC 数据集的目标检测结果,表 5 为实验结果对比。

表 5 特征提取网络结构(检测精度) %		
附加特征提取网络	改进前	改进后
1 层(res5c)	34.52	41.95
6 层	48.99	53.48

表 5 中,1 层(res5c)是指仅用 res5 残差卷积层进行检测,6 层是指用 res4f、res5c 以及后加的 4 个卷积层共 6 层进行检测,改进前是指原始网络模型(ResNet-SSD),改进后是指最终优化后的网络模型(DD-SSD)。实验结果表明在基础网络模型中添加深度分离可变形卷积模块可以提升对一般目标检测的准确率,其中仅用 1 层(res5c)作为检测时,改进后算法相比改进前准确率提高了 7.4 个百分点;而用 6 层作为检测用时,改进后算法相比改进前算法准确率提高约 4.5 个百分点。说明本文提出的深度分离可变形卷积确实可以提升网络模型对图像特征的提取能力。

4 结 论

本文针对复杂环境下的水下目标检测,提出一种基于改进 SSD 的水下目标检测模型 DD-SSD。实验结果显示,利用本文算法对复杂环境下的水下目标检测效果明显优于以 ResNet 为基础网络的 SSD 目标检测算法,检测精度(mAP)达到 95.06%,检测速度达到 22 frame/s,满足实际应用的检测速度及检测精度的基本要求。证明了本文所提算法对复杂水下环境目标检测的有效性和实用性。

参考文献:

- [1] CHANG R, WANG Y, HOU J, et al. Underwater Object Detection with Efficient Shadow-Removal for Side Scan Sonar Images [C]//Oceans, 2016: 1-5
- [2] FABIC J N, TURLA I E, CAPACILLO J A, et al. Fish Population Estimation and Species Classification from Underwater Video Sequences Using Blob Counting and Shape Analysis[C]//Underwater Technology Symposium, 2013: 1-6
- [3] OLIVER K, HOU W. Image Feature Detection and Matching in Underwater Conditions[C]//Proceeding of Society of Photo-Optical Instrumentation Engineers, 2010
- [4] LI J, LIANG X, SHEN S M, et al. Scale-Aware Fast R-CNN for Pedestrian Detection[J]. IEEE Trans on Multimedia, 2018, 20(4): 985-996
- [5] PARKHI O M, VEDALDI A, ZISSERMAN A. Deep Face Recognition[C]//British Machine Vision Conference, 2015: 1-12
- [6] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs[J]. Computer Science, 2014(4): 357-361
- [7] WEIDEMANN A D, GRAY D J, ARNONE R A, et al. Comparison and Validation of Point Spread Models for Imaging in Natural Waters[J]. Optics Express, 2008, 16(13): 9958-9965
- [8] WANG S H, ZHAO J W, CHEN Y Q. Robust Tracking of Fish Schools Using CNN for Head Identification[J]. Multimedia Tools and Applications, 2017, 76(22): 23679-23697
- [9] DAI J, QI H, XIONG Y, et al. Deformable Convolutional Networks[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 764-773
- [10] CHOLLET F. Xception: Deep Learning with Depthwise Separable Convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251-1258
- [11] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications [J/OL] (2017-04-06) [2019-05-15]. <http://arxiv.org/pdf/1704.04861.pdf>
- [12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot Multibox Detector[C]//European Conference on Computer Vision, 2016: 21-37
- [13] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788
- [14] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149
- [15] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]//Conference on Computer Vision and Pattern Recognition, 2016: 770-778

Exploring Underwater Target Detection Algorithm Based on Improved SSD

QIANG Wei¹, HE Yuyao¹, GUO Yujin², LI Baoqi^{3,4}, HE Lingjiao¹

1.School of Marine engineering, Northwestern Polytechnical University, Xi'an 710072, China;

2.Xi'an Horological Research Institute or Light Industry Corporation Ltd., Xi'an 710061, China;

3.Institute of Acoustic, Chinese Academic of Sciences, Beijing 100190, China;

4.Key Laboratory of Science and Technology on Advanced Underwater Acoustic Signal Processing, Beijing 100190, China)

Abstract: As the in-depth exploration of oceans continues, the accurate and rapid detection of fish, bionics and other intelligent bodies in an underwater environment is more and more important for improving an underwater defense system. Because of the low accuracy and poor real-time performance of target detection in the complex underwater environment, we propose a target detection algorithm based on the improved SSD. We use the ResNet convolution neural network instead of the VGG convolution neural network of the SSD as the basic network for target detection. In the basic network, the depthwise-separated deformable convolution module proposed in this paper is used to extract the features of an underwater target so as to improve the target detection accuracy and speed in the complex underwater environment. It mainly fuses the depthwise separable convolution when the deformable convolution acquires the offset of a convolution core, thus reducing the number of parameters and achieving the purposes of increasing the speed of the convolution neural network and enhancing its robustness through sparse representation. The experimental results show that, compared with the SSD detection model that uses the ResNet convolution neural network as the basic network, the improved SSD detection model that uses the depthwise-separated deformable convolution module improves the accuracy of underwater target detection by 11 percentage points and reduces the detection time by 3 ms, thus validating the effectiveness of the algorithm proposed in the paper.

Keywords: underwater target detection; SSD; depthwise-separated convolution; deformable convolutional