

面向嵌入式平台的轻量级光学遥感图像舰船检测

王慧赢¹, 王春平¹, 付强^{1*}, 韩子硕², 张冬冬¹¹陆军工程大学石家庄校区电子与光学工程系, 河北 石家庄 050003;²中国人民解放军 32356 部队, 青海 西宁 710003

摘要 针对现有基于深度学习的轻量级目标检测算法对复杂遥感场景图像中舰船目标检测精度低、检测速度慢的问题, 提出了一种面向嵌入式平台的轻量级光学遥感图像舰船实时检测算法(STYOLO)。首先, 针对主干网络内存访问成本较高的问题, 利用高效网络架构 ShuffleNet v2 作为主干网络对图像进行特征提取, 降低内存访问成本, 提高网络并行度; 其次, 利用 Slim-neck 特征融合结构作为特征增强网络, 以融合较低层级特征图中的细节信息, 增强对小目标的特征响应, 在多尺度信息融合区域施加坐标注意力机制, 强化目标关注以提高较难样本检测以及抗背景干扰能力; 最后, 提出一种跨域迁移和域内迁移相结合的学习策略, 减少源域与目标域的差异性, 提升迁移学习效果。实验结果表明: 基于光学遥感图像舰船检测公开数据集 HRSC2016, 与同类型快速检测算法 YOLOv5s 相比, 所提算法的检测精度提高了 2.7 个百分点, 参数量减少了 61.77%, 在嵌入式平台 Jetson Nano 上检测速度达到 102.8 frame/s, 能够有效实现对光学遥感图像中舰船目标的实时、准确检测。

关键词 光学遥感图像; 舰船检测; 实时检测; 嵌入式平台; 注意力机制; 迁移学习

中图分类号 TP751 **文献标志码** A

DOI: 10.3788/AOS221689

1 引言

舰船目标检测在国防安全、港口动态监视、海运交通管理等军用和民用领域均发挥着重要作用^[1]。当前, 随着航天遥感技术的快速发展, 高分辨率光学遥感图像数量呈指数级增长, 为研究舰船目标检测技术奠定数据基础的同时, 也要求检测系统必须具备实时准确的特点以匹配遥感图像数量的增长速度^[2]。

传统目标检测方法主要通过建立数学模型^[3]或利用目标显著性来完成^[4]。但这类算法大多依赖专家先验知识, 具备一定的局限性, 无法应对背景的复杂多变以及目标的多模态、多差异性。近年来, 伴随着深度学习技术的快速发展, 基于卷积神经网络(CNN)的目标检测方法因其学习能力强、检测精度高而被广泛应用^[5]。当前, 主流的深度学习目标检测模型主要分为双阶段网络^[6-9]和单阶段网络^[10-14]两类。通常情况下, 双阶段网络检测精度较高, 但计算量大、消耗时间长, 在嵌入式设备上部署较为困难。

单阶段网络 YOLO 系列检测算法因其简洁的网络结构, 且同时兼顾检测精度和检测速度而受到广泛关注与应用, 但由于嵌入式设备较差的计算能力和有限的内存资源, 导致单阶段检测模型很难直接应用于

嵌入式设备上对目标进行实时检测, 因此, 在面向嵌入式的目标检测任务中, 文献[15]通过将检测图片平均分割成多个单元格并采用深度可分离卷积实现小型目标检测; 文献[16]通过精简网络与改进多尺度预测来实现遥感目标检测; 文献[17]提出改进主干网络及锚框机制实现实时苹果检测; 文献[18]将主干网络与 GhostNet 进行融合以捕获和细化特征实现快速舰船检测; 文献[19]提出了 MobileNet v2 结合深度可分离卷积实现轻量级水下目标检测; 文献[20]通过精简网络并加入空间和通道维度注意力机制实现小型舰船检测; 文献[21]结合 MobileNet 和聚类思想实现实时目标检测等。上述研究在检测速度上有所提升, 但检测精度较低, 无法有效应对复杂遥感场景图像中舰船目标的实时、准确检测问题: 1) 主干网络内存访问成本较高、并行度低、实时性差; 2) 使用深度可分离卷积的特征增强网络对有云层遮挡、尺度差异大且密集分布等较难检测的舰船增强能力弱; 3) 网络学习过程中由于源域与目标域相似性差, 检测效果不佳。

基于以上分析, 在算力资源有限的嵌入式平台上, 为了保证检测精度的同时提高光学遥感图像中舰船的检测效率, 提出了一种面向嵌入式平台的轻量级目标检测算法 ShuffleNet Tiny-YOLO(简称 STYOLO)。

收稿日期: 2022-09-07; 修回日期: 2022-10-01; 录用日期: 2022-10-27; 网络首发日期: 2022-11-04

基金项目: 军内科研项目(KYSZJWJK2236)

通信作者: *1245316750@qq.com

该算法以 YOLOv5s 为基础框架,首先,利用高效网络架构 ShuffleNet v2 搭建主干网络降低内存访问成本,提高网络并行度;其次,利用 Slim-neck 结构作为特征增强网络克服深度可分离卷积带来的输入通道信息在计算过程中被分离的缺点,并使网络中每一个特征层都能兼顾深层特征的语义信息和浅层特征的细节信息,增强对小目标的特征响应,在多尺度融合区域施加坐标注意力机制,强化目标关注以提高较难样本检测以及抗背景干扰能力;最后,在训练阶段提出了一种跨域迁移和域内迁移相结合的学习训练策略,减少源域

与目标域的差异性,提升迁移学习效果,进一步提高网络检测精度。

2 STYOLO 目标检测算法

STYOLO 光学遥感图像舰船检测模型,如图 1 所示。主要包括 4 部分:输入端、主干网络、特征增强网络、输出端。输入端保留 YOLOv5s 的一系列图像预处理操作来加快推理速度提升网络检测能力;主干网络进行特征提取;特征增强网络对主干网络所提取的特征进行增强;输出端利用增强后的特征层获取预测结果。

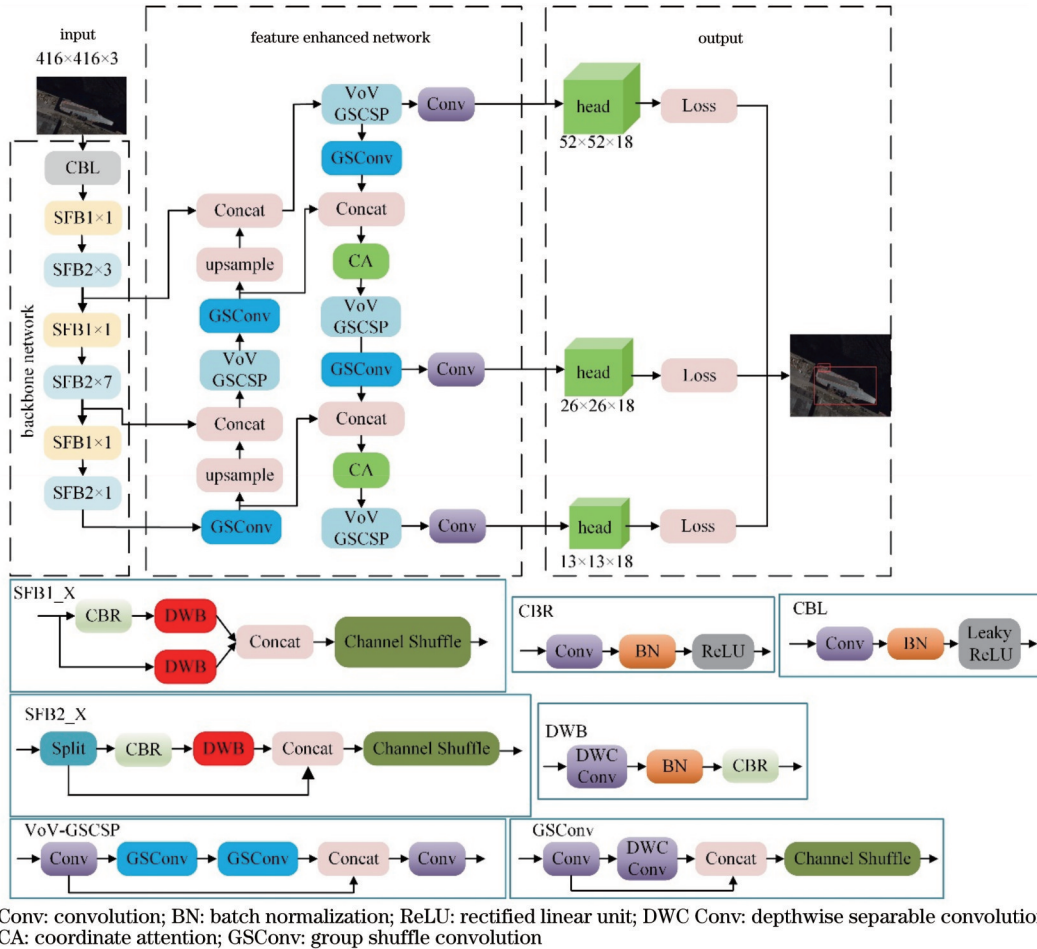


图 1 STYOLO 结构图

Fig. 1 Structure diagram of STYOLO

2.1 主干网络

深度可分离卷积^[22](DSC)可有效降低网络参数数量和运算浮点数(FLOPs),进而实现轻量化。但在进行DSC操作时,输入图像的通道信息在计算过程中是分离的,相对于标准卷积(SC)而言,DSC虽然速度上有很大提升,但精度上却有所下降。在实际应用中速度和精度同样重要,针对DSC的这一缺陷,MobileNet^[23-24]采用1x1密集卷积融合分离的通道信息,但1x1的密集卷积会占用大量计算资源,对计算能力较差的嵌入式设备不够友好;ShuffleNet v1^[25]使用大量组卷积没有考虑内存访问成本。为了解决以上

问题,文献[26]提出了高效网络设计指导原则,并设计了 ShuffleNet v2,其有效结合了 MobileNet 与 ShuffleNet v1 的优点,添加了 Channel Split 操作巧妙地实现了速度与精度的良好平衡。ShuffleNet 结构图如图 2 所示。

高效网络设计指导原则:1) 相等的通道宽度可有效降低内存访问成本;2) 合理应用组卷积避免增加内存访问成本;3) 降低网络碎片化提高并行度;4) 减少元素级操作。

如图 2(c) 所示,ShuffleNet v2 引入了 Channel Split,在每个单元的开始处,输入特征通道被分为 2 个

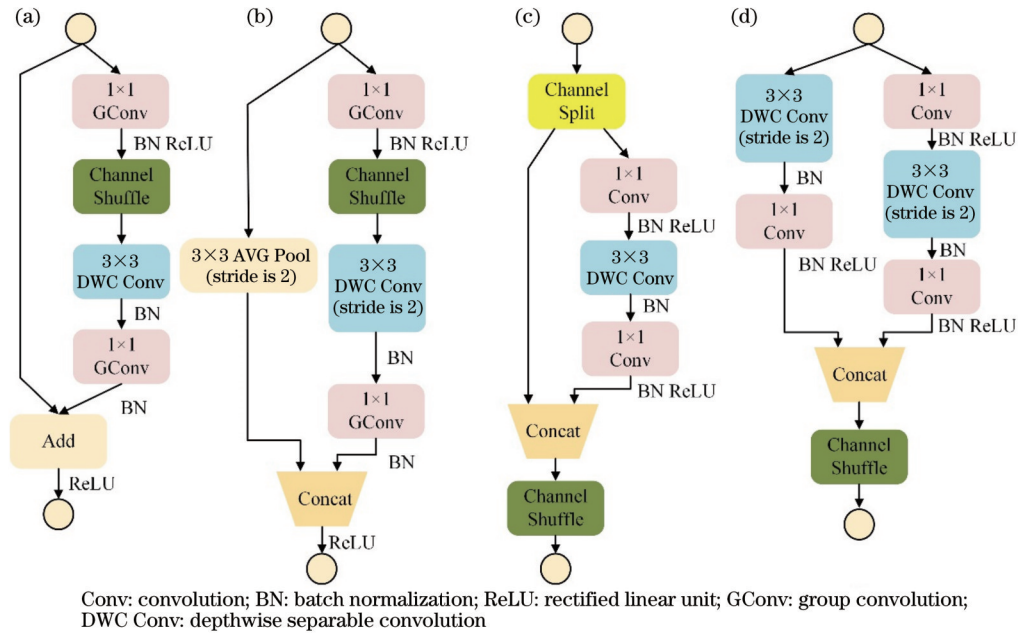


图 2 ShuffleNet 结构图。(a)(b) ShuffleNet v1 构建块; (c)(d) ShuffleNet v2 构建块

Fig. 2 Structure diagrams of ShuffleNet. (a)(b) ShuffleNet v1 building block; (c)(d) ShuffleNet v2 building block

分支,满足第3条指导原则,一个分支为恒等变换,另一个分支由3个卷积组成,且具有相同的输入和输出,满足了第1条指导原则;2个卷积不再是分组的满足第2条指导原则;卷积后2个分支拼接在一起,通道数保持不变满足了第1条指导原则;使用“通道洗牌”操作来实现2个分支的信息通信,洗牌后,下一个单元开始,在 ShuffleNet v2中“Add”操作不再存在,ReLU激活函数和沿深度卷积的元素级操作仅存在于一个分支,此外,3

个连续的逐元素操作“Concat”“Channel Shuffle”“Channel Split”被合并为单个逐元素操作,满足第4条指导原则。如图2(d)所示,空间下采样单元稍作修改,删除了通道拆分操作,使输出通道的数量加倍。图2(c)、图2(d)构建块的堆叠构成了 ShuffleNet v2。根据以上分析并结合实际应用,选用 ShuffleNet v2 搭建主干网络,如图3所示,图中的①、②、③为主干网络提取到的特征层。

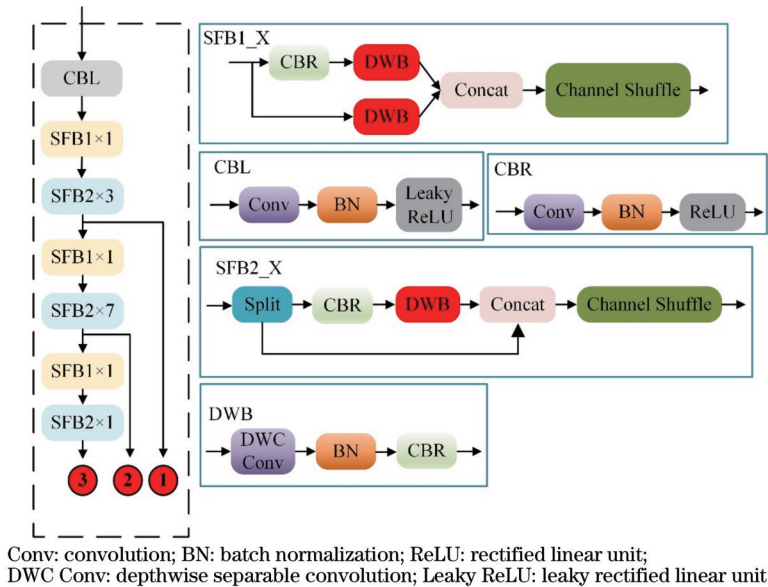


图 3 主干网络

Fig. 3 Backbone network

2.2 特征增强网络

针对光学遥感图像中密集分布的中、小型舰船、有云层遮挡的舰船、尺度差异大且密集分布的舰船难以

准确检测问题,在特征增强网络的设计上既要考虑到计算速度也要兼顾精度。DSC操作在网络轻量化设计上作用显著,但在一定程度上会对精度造成负面影

响。为了使DSC在节省计算资源的情况下,精度逼近SC,文献[27]提出了GSCConv模块克服了DSC的固有

缺点,并设计了Slim-neck特征融合网络,来确保检测器更高的计算成本效益。

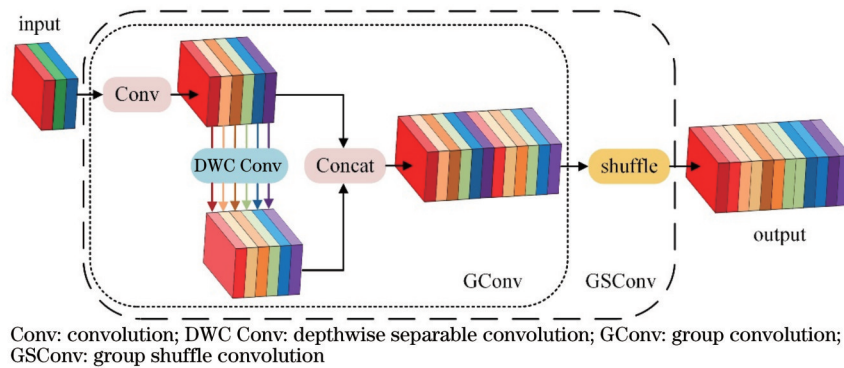


图4 GSCConv模块结构图

Fig. 4 GSCConv module structure diagram

由图4可知,使用shuffle将SC生成的特征信息逐一渗透到DSC生成的特征信息中,这种方法使得经过DSC得到的通道信息拥有来自SC的全部信息,使输出信息更加完备,同时也降低了计算成本。图5为YOLOv5s的第10层也就是YOLOv5s的Neck起

始端分别应用SC、DSC、GSCConv运算得到的光学遥感图像舰船特征图。由图5可知,由GSCConv运算生成的特征图更加接近SC运算生成的特征图,将GSCConv模块应用到本文任务中可以有效提高舰船的检测精度。

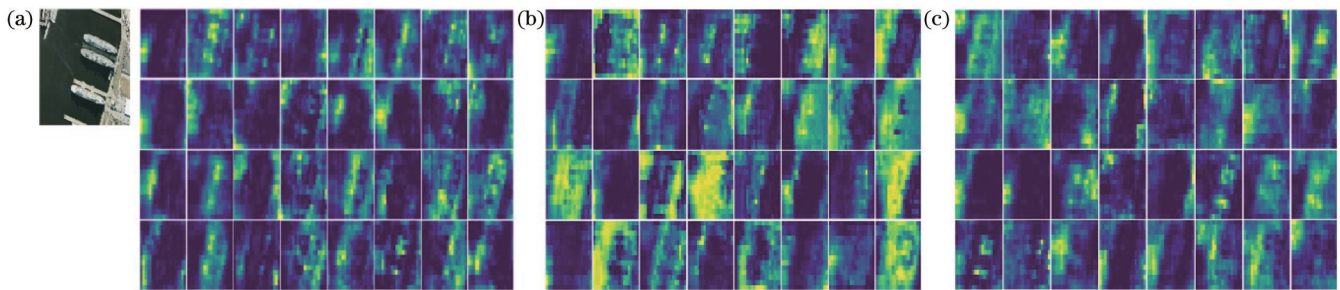


图5 YOLOv5s第10层特征图。(a) SC运算生成的特征图;(b) DSC运算生成的特征图;(c) GSCConv运算生成的特征图

Fig. 5 Characteristic diagrams of 10th layer of YOLOv5s. (a) Characteristic diagram generated by SC operation; (b) characteristic diagram generated by DSC operation; (c) characteristic diagram generated by GSCConv operation

输入图像经过网络逐层卷积下采样处理,造成小型舰船特征大量丢失甚至被忽略,考虑到嵌入式平台有限的计算资源,使用一次性聚合方法设计跨级部分网络模块(VoV-GSCSP),保持了特征增强能力,又降低了计算消耗,将SC操作替换为GSCConv操作进一步减少网络参数量。GSCConv操作结合VoV-GSCSP对主干网络提取到的特征层进行增强,使每一个特征层都能同时兼顾深层特征的语义信息和浅层特征的细节信息,提高对小目标的特征响应,特征增强网络如图6所示。

在深度神经网络中应用注意力机制可以使网络性能得到很大的提升,但有些注意力机制应用在模型较小的移动网络中时性能会明显落后于大型网络,这是由于注意力机制嵌入到轻量型的网络会带来巨大的计算开销,如self-attention。轻量化网络模型中常用的注意力机制主要有Squeeze-and-excitation(SE)^[28]、Bottleneck attention module(BAM)^[29]、Convolutional

block attention module(CBAM)^[30]。SE忽略了目标的位置信息,只考虑到了网络内部的通道信息;BAM、CBAM通过在通道上进行全局池化操作来引入位置信息,但这种方法无法获取目标的全局位置信息。因此,文献[31]提出了一种适用于轻量型网络的坐标注意力机制(CA),CA通过嵌入位置信息到通道注意力使网络获取大范围的区域信息,同时避免引入过大的计算开销。

光学遥感图像中的舰船有些呈密集分布或较大舰船旁边分布较小舰船,造成模型检测精度下降,施加坐标注意力机制对任务中较难识别的舰船目标分配较高权重,增加对其的关注度,分配低权重抑制自然背景,提高对密集分布的较小舰船、较大舰船旁边分布的较小舰船等较难检测目标的识别精度。坐标注意力机制如图7所示,为了避免全局池化操作导致的位置信息丢失,CA将注意力分别放在宽度和高度两个维度上,便于有效利用输入特征图的空间坐标信息。CA包含

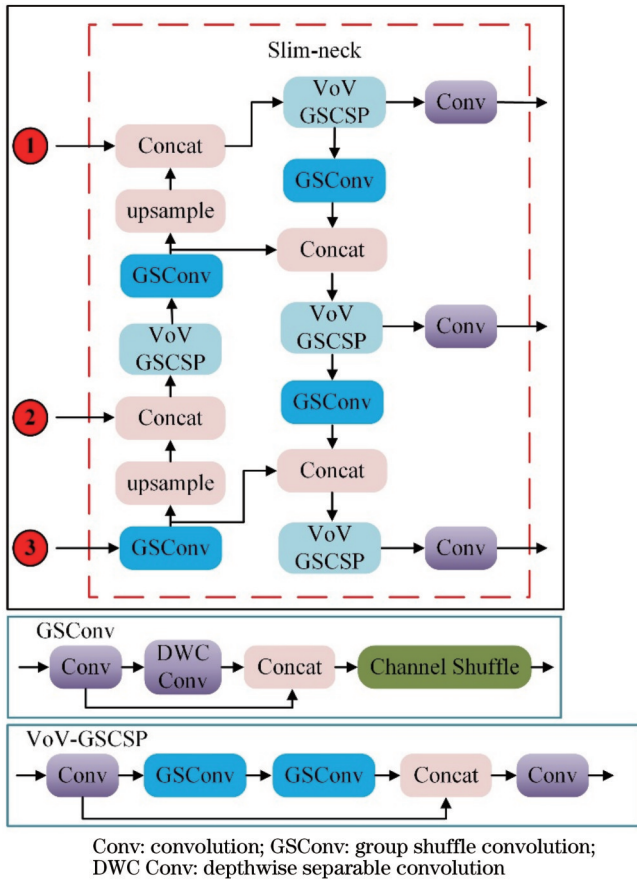
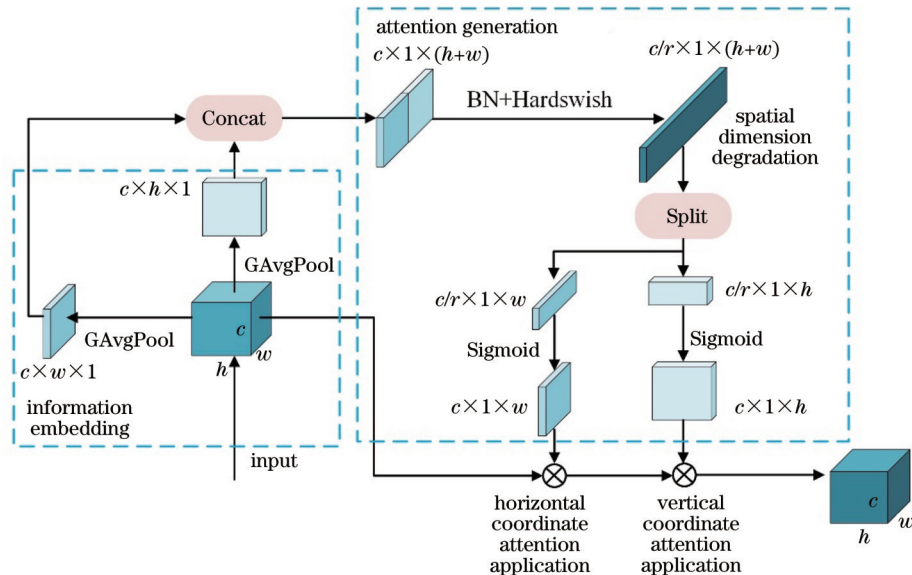


图 6 特征增强网络

Fig. 6 Feature enhancement network

信息嵌入和注意力生成两部分。在信息嵌入阶段对输入的特征图的所有通道分别沿垂直坐标方向、水平坐标方向进行全局平均池化操作获得 $c \times h \times 1$ 、 $c \times w \times 1$ 的特征图, 然后进行特征汇聚输入到注意力生成阶段;



BN: batch normalization; Hardswish: hardswish activation function; GAVgPool: global average pooling

图 7 坐标注意力机制

Fig. 7 Coordinate attention mechanism

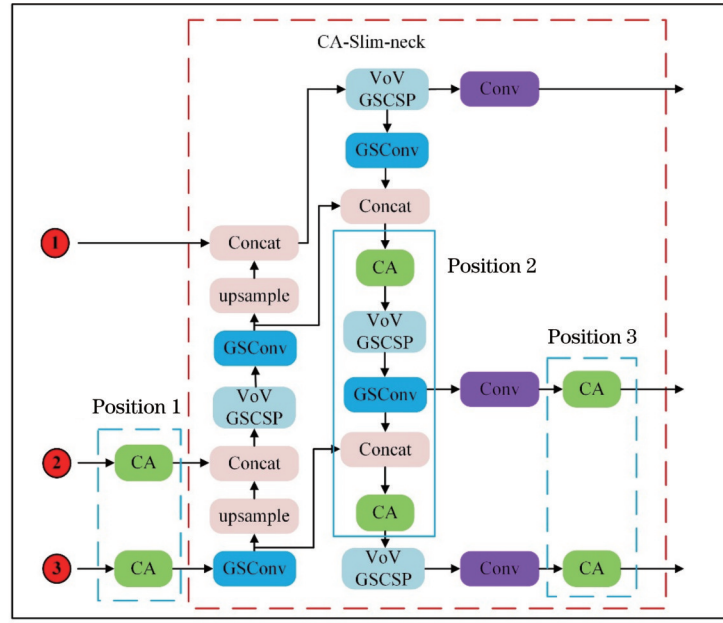
在注意力生成阶段首先将输入的特征图拼接为 $c \times 1 \times (h+w)$ 的特征图, 通过收缩率 r 将特征图从 c 维压缩至 c/r 维, 其次将特征图沿空间维分解为 $c/r \times 1 \times w$ 的水平注意力张量和 $c/r \times 1 \times h$ 的垂直注意力张量, 然后将其分别扩张为 $c \times 1 \times w$ 、 $c \times 1 \times h$, 最后与输入特征图相乘。

如图 8 所示, 将坐标注意力机制引入到 Slim-neck 结构中位置 2 处, 使注意力机制充分获取不同尺度的特征信息, 增强网络对较难检测目标的识别能力, 网络中在位置 1 及位置 3 处施加的 CA 模块仅用于对比实验, 不作为最终网络结构的一部分。

2.3 跨域迁移学习与域内迁移学习相结合的训练策略

利用大量的数据训练可以提升网络模型的检测精度, 针对光学遥感图像舰船数据集缺乏的问题, 通常采用数据增强方式扩充数据集, 并结合跨域迁移的方式进行知识迁移。由于源域与目标域相似性低, 若仅采用跨域迁移学习方法进行知识的学习, 网络很难快速收敛且检测精度也不佳。因此, 针对本文检测任务提出了一种跨域迁移学习结合域内迁移学习的训练策略^[32], 在通用数据集与光学遥感图像舰船数据集之间添加仅含有舰船特征的数据集作为过渡域, 并采用缩放、翻转等方式进行数据扩充, 丰富数据集中不同方向、不同尺度下的舰船特征, 减少与目标域的差异, 促进迁移学习的学习效果, 提升模型的检测精度。

跨域迁移与域内迁移相结合的训练策略分为 2 步: 1) 进行跨域迁移学习, 使用通用大型数据集训练得到的参数对主干网络进行初始化, 并利用仅含有舰船特征的数据集对模型进行微调, 充分学习光学遥感



Conv: convolution; CA: coordinate attention; GSCConv: group shuffle convolution

图 8 施加坐标注意力机制的 Slim-neck 结构图

Fig. 8 Slim-neck structure diagram with coordinate attention mechanism

图像中的舰船特征;2) 进行域内迁移学习,利用光学遥感图像舰船数据集在步骤 1) 训练好的模型上进行微调,学习到复杂环境影响的舰船特征。

3 网络训练与检测试验

3.1 数据集

试验所采用的数据集分为 2 种。数据集 1 使用公开数据集 HRSC2016^[33],该数据集是光学遥感图像中舰船目标检测唯一公开数据集,其中所有图像均来自 Google Earth 的 6 个著名的港口,分辨率在 2 m 到 0.4 m 之间,像素尺寸在 300 pixel×300 pixel 到 1500 pixel×900 pixel 之间,多为 1000 pixel×600 pixel,总共包含 2976 张舰船目标;数据集 2 使用自建数据集,由于 HRSC2016 数据集中仅包含 6 个港口,场景缺乏多样性,因此,从互联网、Google Earth 上共搜集 1100 张图像作为样本,其中轻度云层遮挡样本、目标与背景较难区分样本、目标尺度差异较大样本、较小目标样本、密集分布样本分别各占数据集的 20%,使用 LabelImg 图像标注工具对数据集进行标注,并以 8:1:1 的比例划分为训练集、测试集、验证集,如表 1 所示。

表 1 自建数据集数量

Table 1 Number of self-built dataset

Dataset	Image	Ship
Training set	880	2392
Test set	110	345
Validation set	110	250

数据集的格式均为 VOC 格式,应用 k -means 算法对 HRSC2016 数据集中真实目标框的尺寸进行聚类,如图 9 所示,最终得到 9 个聚类中心,分别为(9, 14)、

(13, 35)、(27, 19)、(61, 30)、(33, 63)、(26, 153)、(112, 67)、(64, 146)、(128, 169),并将聚类的结果设置为网络的先验框。

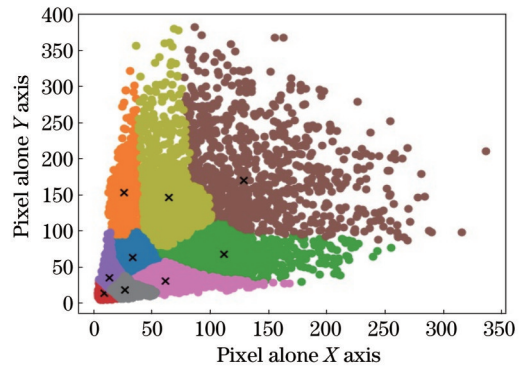


图 9 HRSC2016 数据集聚类结果

Fig. 9 Clustering result of HRSC2016 dataset

为了提高第二阶段域内迁移学习的效果,解决源域与目标域的特征分布差异导致的负迁移,对 HRSC2016 数据集进行增强以提高源域与目标域的相似性,数据增强过程如图 10 所示:1) 将数据集上所有真实目标从原始图像中裁剪下来;2) 将裁剪下来的图像进行随机缩放、翻转;3) 随机选 4 张图像拼接成原始图像大小。

拼接过程中不局限于一张图像上的目标,且图像与图像间不相接,好处为不需要对图像边缘作特殊处理。

3.2 损失函数

模型的损失函数主要由 3 类损失函数组成:边框损失、类间损失、目标损失。由于本研究任务只有一类目标,因此,去除类间损失只保留边框损失和目标损失。YOLOv5s 采用更全面的检测框与真实框之间的

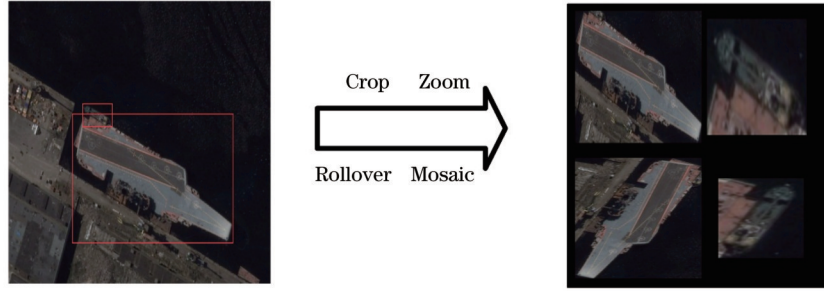


图 10 数据增强过程

Fig. 10 Data enhancement process

交并比 (Generalized intersection over union, GIoU) 作为边框损失函数, 可以解决检测框与真实框之间的交并比 (Intersection over union loss, IOU) 无法解决的梯度失效问题, 但当 2 个目标边界框高度和宽度一致时 GIoU 损失函数就会退化为 IOU 损失函数, 并且 GIoU 损失函数还存在着收敛速度慢、回归不精确的问题。光学遥感图像中较小舰船居多且多呈密集状态分布, GIoU 损失函数虽然计算较为简单, 但缺乏回归定位损失函数应有的长宽比这一重要参数, 应用在本任务时, 很容易出现密集分布目标漏检的问题, 故针对本文舰船目标分布特点, 借鉴 YOLOv4 的损失函数方法, 采用更完善的检测框与真实框之间的交并比 (Complete intersection over union, CIOU) 损失函数为本文网络的损失函数。GIoU 损失函数、CIOU 损失函数的计算过程如下:

$$f_{IoU} = \frac{|A \cap B|}{|A \cup B|}, \quad (1)$$

$$L_{IoU} = 1 - f_{IoU}, \quad (2)$$

$$f_{GIoU} = f_{IoU} - \frac{|A_c - (A \cup B)|}{|A_c|}, \quad (3)$$

$$L_{GIoU} = 1 - f_{GIoU}, \quad (4)$$

$$f_{CIOU} = f_{IoU} - \left(\frac{\rho^2(b, b^{gt})}{c^2} + av \right), \quad (5)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2, \quad (6)$$

$$a = \frac{v}{(1 - f_{IoU}) + v}, \quad (7)$$

$$L_{CIOU} = 1 - f_{CIOU}, \quad (8)$$

式中: A 为真实框; B 为预测框; A_c 为最小闭包区域的面积; ρ 为预测目标框中心点 B 与真实目标框中心点 A 之间的欧氏距离; b, b^{gt} 分别为预测目标框与真实目标框的中心点; c 为能够同时包含预测目标框与真实目标框的最小闭包区域对角线距离; w^{gt} 为真实目标框的长; h^{gt} 为真实目标框的宽; w 为预测目标框的长; h 为预测目标框的宽; a 为权重函数; v 用来度量长宽比的相似性; L_{IoU} 为 IoU 损失函数; L_{GIoU} 为 GIoU 损失函数; L_{CIOU} 为 CIOU 损失函数。

非极大值抑制 (NMS) 常被用来处理一个目标产生多个候选框的问题, YOLOv5s 使用的是加权 NMS 的方法, 可以解决传统 NMS 每次迭代后选出的最大得分框未必是准确的问题, 但加权 NMS 属于顺序处理模式, 运算效率较低, 并且对一些密集分布的目标容易漏检, 而光学遥感图像中的舰船中密集分布的小舰船居多。综合考虑了真实目标框、预测目标框、先验框之间的距离、重叠情况、尺度的交并比 (Distance intersection over union, DIoU) 损失函数使目标框回归更稳定, 因此, 为了进一步提升网络的检测精度采用 DIoU-NMS。

3.3 实验平台

实验平台分为 2 种: 模型训练平台为台式计算机, Ubuntu 18.04 系统, 配置为 Intel(R) Xeon(R) CPU E5-2630 v3 @2.40 GHz, 内存为 64 GB, GPU 为 NVIDIA GeForce RTX 2080Ti/PCIe/SSE2, 显存为 1T, Python 版本为 3.6, CUDA 版本为 11.1, CUDNN 版本为 8.0.5, 在 PyTorch 深度学习框架上进行模型训练。模型优化后的推理过程在 Jetson Nano 嵌入式平台上进行, 中央处理器 (CPU) 为 Qual-Core ARM Cortex-A57 64-bit@1.42 GHz, 图形处理器 (GPU) 为 128-core Maxwell, 内存为 4 GB 64-bit LPDDR4, 配置系统为 ARM 版 Ubuntu 18.04 系统, 模型运行环境配置为 JetPack4.4、DeepStream5.1、TensorRT7.1。

3.4 模型训练

网络模型的训练分为 2 阶段: 1) 跨域迁移学习, 利用在大规模数据集 COCO 训练好的 ShuffleNet v2 网络预训练权重对网络模型参数进行初始化, 并利用 HRSC2016 数据集对网络进行微调; 2) 域内迁移学习, 利用自建数据集对网络进一步微调。

每阶段训练分为两步: 首先, 冻结主干网络, 训练轮数为 50、训练批次为 16、学习率为 1×10^{-2} ; 接着, 解冻训练, 训练轮数为 50、训练批次为 8、学习率为 1×10^{-4} 。选用随机梯度下降 (Stochastic gradient descent, SGD) 优化器, 学习率的下降方式为余弦退火。训练日志使用 Tensorboard 记录, 每迭代一次写入训练集损失, 每训练一轮写入验证集损失, 共训练 100 轮, 将后 50 轮中验证集损失最低的模型作为训练结果进行分析。

训练过程使用 Tensorboard 可视化进行记录, 每进行一次迭代的损失加入到训练集损失中; 每训练一轮加入验证集损失中, 并保存模型效果最好的一次权重。

为验证模型的有效性, 在相同测试集及验证集下分别与 ShuffleNetv2-YOLOv5s、YOLOv5s、MobileNetv3-YOLOv5s、YOLOv5n 进行训练迭代 100 轮得到的训

练集边框损失 (train/box_loss)、训练目标损失 (train/obj_loss)、验证集边框损失 (val/box_loss)、验证目标损失 (val/obj_loss)、精确率 (P)、召回率 (R)、 $mAP@0.5$ [交并比阈值为 0.5 时的平均精度均值 (mAP)]、 $mAP@0.25$ (交并比阈值在 0.5~0.95 之间每隔 0.05 取一次 mAP , 并将这些值再取平均) 如图 11 所示。

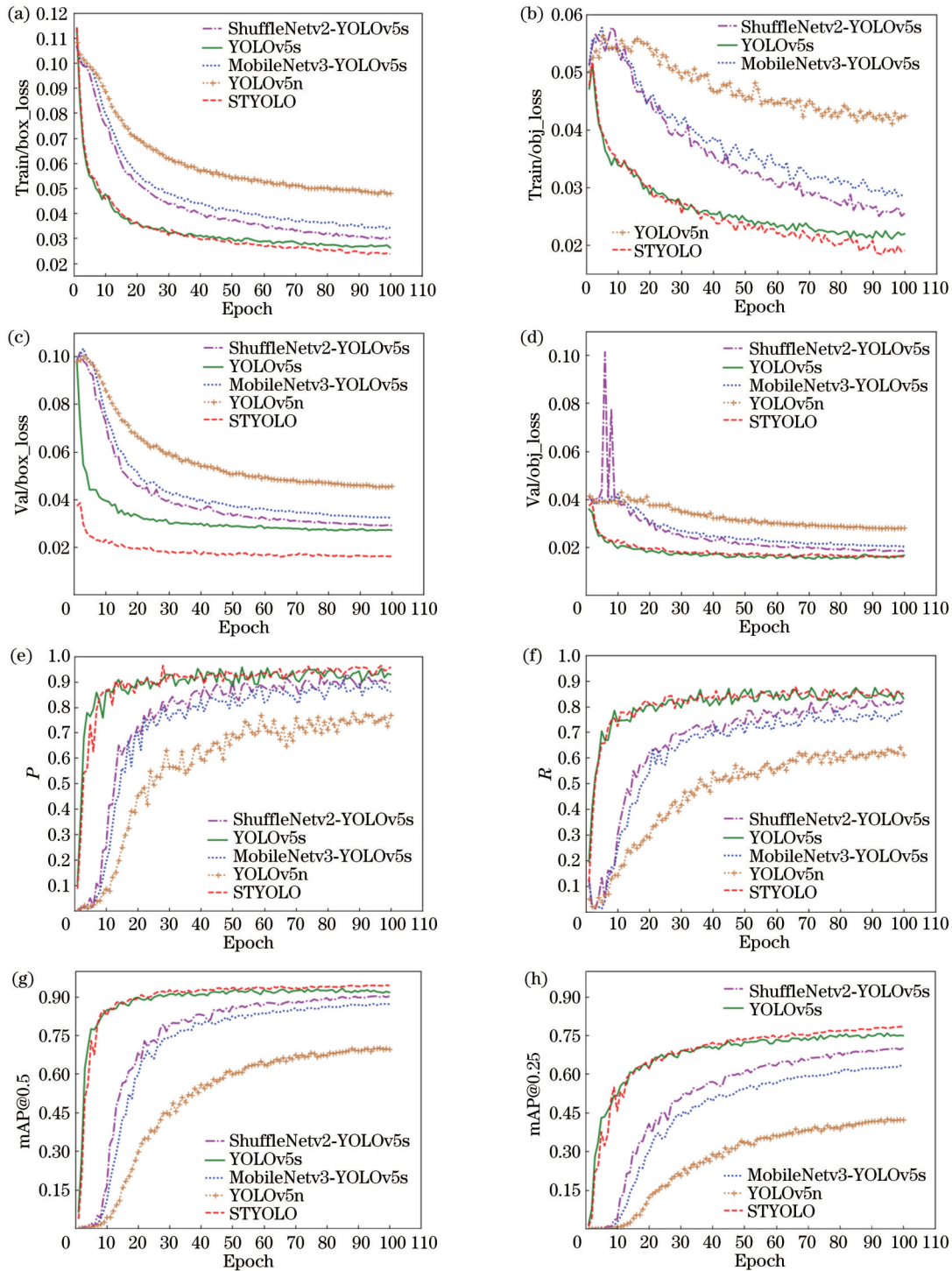


图 11 5 种检测模型训练曲线。(a) 训练集边框损失; (b) 训练目标损失; (c) 验证集边框损失; (d) 验证目标损失; (e) 精确率; (f) 召回率; (g) $mAP@0.5$; (h) $mAP@0.25$

Fig. 11 Training curves of five detection models. (a) Training set box loss; (b) validation set box loss; (c) training object loss; (d) validation object loss; (e) precision; (f) recall; (g) $mAP@0.5$; (h) $mAP@0.25$

由图 11 的训练集边框损失及目标损失可知, YOLOv5s 及本文模型在 0 到 30 轮损失值下降迅速, 随着迭代的进行损失值下降缓慢, 说明模型已趋于收敛, 且曲线迅速下降波动平稳, 说明模型较稳定; ShuffleNetv2-YOLOv5s、MobileNetv3-YOLOv5s、YOLOv5n 均在 60 轮后模型趋于收敛, 并由训练集目标损失可知, 3 种模型相较于 YOLOv5s 及本文模型曲线波动较大; 由验证集边框损失可知, 本文模型下降最快、最为平稳且损失值达到了最低; 由目标损失曲线可知, YOLOv5s 与本文模型下降最快, ShuffleNetv2-YOLOv5s 在训练 10 轮左右目标损失出现较大波动; 由 P 、 R 可知, YOLOv5s 与本文模型均收敛较快且曲线波动较为平稳; 由 mAP 曲线可知, YOLOv5s 与本文模型波动较为平稳, 本文的 mAP@0.5 及 mAP@0.25 均表现最佳, mAP@0.5 达到了 94.33%, mAP@0.25 达到了 78.32%, YOLOv5s 的 mAP@0.5 达到了 91.63%,

表 2 网络轻量化对模型的影响

Table 2 Effect of network lightweighting on the model

Model	P / %	R / %	mAP@0.5 / %	FLOPs/B	Parameter/MB
YOLOv5s	90.25	76.80	87.64	15.8	7.01
STYOLO	84.12	78.36	85.52	6.0	2.66

3.5.2 坐标注意力机制施加位置对网络检测效果的影响

在网络轻量化的基础上, 施加 CA, 在 Slim-neck 中的不同位置施加 CA 对网络检测能力的影响, 如表 3 所示。

表 3 在不同位置施加 CA 的检测结果对比

Table 3 Comparison of detection capabilities of CA applied at different positions

Application position	mAP@0.5 / %	Parameter/MB
Baseline	85.52	2.66
Position 1	80.05	2.67
Position 2	90.46	2.68
Position 3	85.66	2.67

由表 3 可知, CA 施加在位置 1 处使检测模型的 mAP 降低了 5.47 个百分点; CA 施加在位置 2 处使检测模型的 mAP 提高了 4.94 个百分点; CA 施加在位置 3 处使检测模型的 mAP 提高了 0.14 个百分点, 说明 CA 施加在 Slim-neck 网络中不同位置并不一定会带来网络检测性能的提升。位置 1 处为主干网络的输出端, 其负责对图像特征的提取过程, 在位置 1 施加 CA 会给 Slim-neck 带来大量错误信息, 因此, 网络检测精度会有所下降; 位置 2 处是特征增强网络中不同尺度信息的交汇处, 能够使 CA 获得更加丰富的位置信息, 提升网络特征增强性能; 位置 3 处是 Slim-neck 网络的末端, 特征增强任务基本完成, 在此位置施加注意力机制对检测模型精度的提升作用不大。

mAP@0.25 达到了 74.86%, 2 种模型曲线收敛迅速波动较少, 说明模型较稳定。

3.5 实验结果

为了验证 STYOLO 的有效性, 采用 mAP、 P 、 R 作为衡量模型精度的指标, 同时采用整个检测模型产生的参数量、浮点数 (FLOPs)、每秒传输帧数作为衡量模型的大小、运算量、运行速度的指标。

3.5.1 网络轻量化对比实验

由表 2 可知, 基于 YOLOv5s 框架, 应用 ShuffleNet v2 作为主干网络, Slim-neck 作为特征增强网络后, 应用跨域迁移学习方式对两种检测模型进行训练, 相较于 YOLOv5s 模型, 轻量化后 mAP@0.5 下降了 2.12 个百分点, 浮点数降低了 62.02%, 参数量降低了 62.05%。说明网络轻量化可以有效降低参数量、计算量, 但会损失一定的检测精度。为了提高网络的检测精度, 还需进一步优化, 提升综合能力。

另外, CA 施加在位置 2 处的内存占用较大, 参数量相较于没有添加 CA 的检测模型增加了 0.75%, 但精度提升较高, 而在内存空间及算力可以承受的范围内提高检测模型的精度较为重要, 故最终选择在位置 2 处施加 CA。

3.5.3 施加不同注意力机制对模型检测效果的影响

在网络轻量化的基础上, 位置 2 处施加不同注意力机制, 对比不同注意力机制对网络检测能力的影响, 如表 4 所示。

由表 4 可知, 施加 SE 在收缩率 r 为 32 时检测模型的 mAP 最高 (87.46%), 提升了 1.94 个百分点; 施加 CBAM 在收缩率 r 为 32 时检测模型的 mAP 最高

表 4 施加不同注意力机制的检测能力对比

Table 4 Comparison of detection capabilities with different attention mechanisms

Attention mechanism	r	mAP@0.5 / %	Parameter / MB
Baseline	—	85.52	2.66
	8	86.01	2.72
	32	87.46	2.67
SE	8	86.32	2.72
	16	86.88	2.68
	32	87.26	2.67
CBAM	8	89.32	2.75
	16	88.68	2.71
	32	90.46	2.68

(87.26%),提升了1.74个百分点;施加CA在收缩率 r 为32时检测模型的mAP最高(90.46%),提升了4.94个百分点。说明施加CA应用于检测任务可有效提高舰船的检测精度。

3.5.4 不同训练方式对网络检测效果的影响

对比不同训练方式对检测模型性能的影响,如表5所示。

表5 不同迁移学习方式检测能力对比

Table 5 Comparison of detection capabilities with different transfer learning methods

Training method	Dataset	mAP@0.5 / %
Cross-domain	COCO dataset+self-built dataset	90.46
In-domain	HRSC2016 dataset+self-built dataset	80.16
Cross-domain+in-domain	COCO dataset+HRSC2016 dataset+self-built dataset	94.33

应用COCO预训练模型在自建数据集上进行训练,作为跨域迁移的学习方式;对模型进行随机初始化并先后在HRSC2016数据集和自建数据集上进行训练,作为域内迁移的学习方式;应用COCO预训练模型先后在HRSC2016和自建数据集上进行训练,作为跨域迁移结合域内迁移的学习方式。由表5可知,应用跨域迁移学习结合域内迁移学习的训练方式检测模型的mAP最高(94.33%),相较于单独应用跨域迁移学习与域内迁移学习的训练方式mAP分别提高了3.87、14.17个百分点,这是因为跨域迁移学习结合域内迁移学习的训练方式相较于跨域迁移学习的训练方式,添加过渡域可以缓解因源域与目标域相似性低所带来的负面影响;相较于域内迁移学习的训练方式,应用通用特征对检测模型进行初始化可以使检测模型避免因权值太过随机导致的特征提取效果差的问题。

3.5.5 不同检测模型对比实验

表6为本模型分别与ShuffleNetv2-YOLOv5s、YOLOv5s、MobileNetv3-YOLOv5s、YOLOv5n在台式计算机与嵌入式平台Jetson Nano端进行对比实验。由表6可知,本模型检测精度为94.33%,相较于

YOLOv5s提高了2.7个百分点,相较于ShuffleNetv2-YOLOv5s提高了4.19个百分点,相较于MobileNetv3-YOLOv5s提高了7.27个百分点,相较于YOLOv5n提高了24.61个百分点;本模型的参数量为2.68 MB,约为YOLOv5s的38.23%,ShuffleNetv2-YOLOv5s的48.38%,MobileNetv3-YOLOv5s的75.71%,YOLOv5n的1.51倍;本文模型的浮点数为6.2 B,相较于YOLOv5s减少了60.76%,相较于ShuffleNetv2-YOLOv5s减少了45.61%,相较于MobileNetv3-YOLOv5s减少了1.59%;相较于YOLOv5n增加了51.22%;在台式计算机的图像每秒传输帧数为200.6 frame/s,约为YOLOv5s的2.1倍,ShuffleNetv2-YOLOv5s的1.33倍,MobileNetv3-YOLOv5s的1.66倍,YOLOv5n的1.47倍;在Jetson Nano端的图像每秒传输帧数为102.8 frame/s,约为YOLOv5s的2.21倍,ShuffleNetv2-YOLOv5s的1.36倍,MobileNetv3-YOLOv5s的1.70倍,YOLOv5n的1.50倍。综上结果分析,本模型在光学遥感舰船检测任务上实现了检测速度和检测精度的平衡,整体性能均表现较好。

表6 不同模型检测性能比较

Table 6 Comparison of detection performance with different models

Model	mAP@0.5 / %	Parameter / MB	FLOPs / B	Desktop computer (FPS) / (frame·s ⁻¹)	Jetson Nano (FPS) / (frame·s ⁻¹)
YOLOv5n	69.72	1.77	4.1	136.7	68.4
YOLOv5s	91.63	7.01	15.8	95.4	46.5
ShuffleNetv2-YOLOv5s	90.14	5.54	11.4	150.6	75.5
MobileNetv3-YOLOv5s	87.06	3.54	6.3	120.8	60.4
STYOLO	94.33	2.68	6.2	200.6	102.8

对有云层遮挡、尺度差异大且密集分布,与背景区分度较小等复杂场景下舰船目标的检测效果,如图12所示。其中实线矩形框为预测结果,虚线矩形框为漏检结果,点划线矩形框为误检结果。由图12可知,本算法可有效识别密集分布的中、小型舰船、有轻度云层

遮挡的舰船、尺度差异大且密集分布的舰船、与背景区分度较小的舰船。但对重度云层遮挡的舰船、密集分布的小型舰船仍存在漏检,后续需对此问题进一步研究。

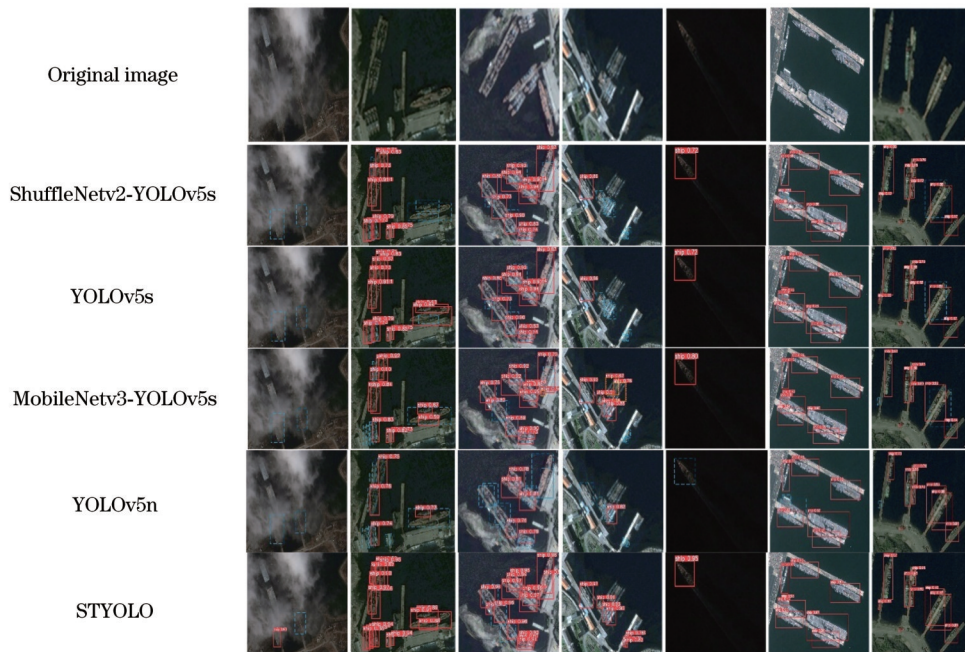


图 12 不同检测模型检测效果对比

Fig. 12 Comparison of detection effects with different detection models

4 结 论

针对现有轻量级目标检测算法无法实现对复杂遥感场景中舰船实时、准确检测问题,提出了一种面向嵌入式平台的轻量级光学遥感图像舰船实时检测算法 STYOLO。算法通过使用高效网络架构 ShuffleNet v2 构建主干网络,在保证模型特征提取能力的同时,降低内存访问成本,提高网络并行度;使用 Slim-neck 作为特征增强网络克服深度可分离卷积给网络检测精度带来的负面影响,并使网络中每一个特征层都能同时兼顾深层特征的语义信息和浅层特征的细节信息,增强对小目标的特征响应,在多尺度融合区域施加坐标注意力机制,强化目标关注以提高较难检测样本识别以及抗背景干扰能力;提出了一种跨域结合域内迁移学习的训练方式,减少源域与目标域的差异性,提升迁移学习效果,进一步提高模型的检测精度。与当前应用于嵌入式的主流检测算法进行对比实验,STYOLO 在保证高精度的同时有效提高了检测速度,在 Jetson Nano 端检测速度达到 102.8 frame/s,约为 YOLOv5s 的 2.21 倍, ShuffleNetv2-YOLOv5s 的 1.36 倍, MobileNetv3-YOLOv5s 的 1.70 倍, YOLOv5n 的 1.50 倍;精度达到了 94.33%,相较于 YOLOv5s 提高了 2.7 个百分点,相较于 ShuffleNetv2-YOLOv5s 提高了 4.19 个百分点,相较于 MobileNetv3-YOLOv5s 提高了 7.27 个百分点,相较于 YOLOv5n 提高了 24.61 个百分点,可满足在嵌入式平台上对光学遥感图像中舰船准确、实时的检测。

在遥感图像舰船检测任务中,可见光图像容易受

自然环境的影响导致目标特征被削弱,难以提升算法精度,利用红外图像结合可见光图像实现融合检测以提升弱目标检测精度是研究的重点方向。

参 考 文 献

- [1] 薛俊达, 朱家佳, 张静, 等. 基于 FFC-SSD 模型的光学遥感图像目标检测[J]. 光学学报, 2022, 42(12): 1210002.
Xue J D, Zhu J J, Zhang J, et al. Object detection in optical remote sensing images based on FFC-SSD model[J]. Acta Optica Sinica, 2022, 42(12): 1210002.
- [2] 农元君, 王俊杰. 基于注意力和强化学习的遥感图像描述方法[J]. 光学学报, 2021, 41(22): 2228001.
Nong Y J, Wang J J. Remote sensing image caption method based on attention and reinforcement learning[J]. Acta Optica Sinica, 2021, 41(22): 2228001.
- [3] Dai H, Du L, Wang Y, et al. A modified CFAR algorithm based on object proposals for ship target detection in SAR images[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(12): 1925-1929.
- [4] Tao D, Anfinson S N, Brekke C. Robust CFAR detector based on truncated statistics in multiple-target situations[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(1): 117-134.
- [5] Zhang D D, Wang C P, Fu Q. CAFC-net: a critical and align feature constructing network for oriented ship detection in aerial images[J]. Computational Intelligence and Neuroscience, 2022, 2022: 3391391.
- [6] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [7] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.
- [8] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine

- Intelligence, 2017, 39(6): 1137-1149.
- [9] Dai J F, Li Y, He K, et al. R-FCN: object detection via region-based fully convolutional networks[C]//Advances in Neural Information Processing Systems 29, December 5-10, 2016, Barcelona, Spain. Copenhagen: MLR Press, 2016: 379-387.
- [10] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [11] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [12] Redmon J, Farhadi A. YOLOv3: an incremental improvement [EB/OL]. (2018-04-08) [2022-09-07]. <https://arxiv.org/abs/1804.02767>
- [13] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [14] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2022-05-08]. <https://arxiv.org/abs/2004.10934>.
- [15] 崔家华, 张云洲, 王争, 等. 面向嵌入式平台的轻量级目标检测网络[J]. 光学学报, 2019, 39(4): 0415006.
Cui J H, Zhang Y Z, Wang Z, et al. Light-weight object detection networks for embedded platform[J]. Acta Optica Sinica, 2019, 39(4): 0415006.
- [16] 农元君, 王俊杰. 基于嵌入式的遥感目标实时检测方法[J]. 光学学报, 2021, 41(10): 1028001.
Nong Y J, Wang J J. Real-time object detection in remote sensing images based on embedded system[J]. Acta Optica Sinica, 2021, 41(10): 1028001.
- [17] Yan B, Fan P, Lei X Y, et al. A real-time apple targets detection method for picking robot based on improved YOLOv5 [J]. Remote Sensing, 2021, 13(9): 1619.
- [18] Liu T, Zhou B J, Zhao Y S, et al. Ship detection algorithm based on improved YOLO V5[C]//2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE), July 15-17, 2021, Dalian, China. New York: IEEE Press, 2021: 483-487.
- [19] Zhang M H, Xu S B, Song W, et al. Lightweight underwater object detection based on YOLO v4 and multi-scale attentional feature fusion[J]. Remote Sensing, 2021, 13(22): 4706.
- [20] Hu J M, Zhi X Y, Shi T J, et al. PAG-YOLO: a portable attention-guided YOLO network for small ship detection[J]. Remote Sensing, 2021, 13(16): 3059.
- [21] Liu Y, Gao M F, Zong H M, et al. Real-time object detection for the running train based on the improved YOLO V4 neural network[J]. Journal of Advanced Transportation, 2022, 2022: 4377953.
- [22] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1800-1807.
- [23] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2018-10-15) [2017-01-17]. <https://arxiv.org/abs/1704.04861>.
- [24] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.
- [25] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6848-6856.
- [26] Ma N N, Zhang X Y, Zheng H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer Vision-ECCV 2018. Lecture Notes in Computer Science. Cham: Springer, 2018, 11218: 122-138.
- [27] Li H L, Li J, Wei H B, et al. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles[EB/OL]. (2022-06-06)[2022-08-06]. <https://arxiv.org/abs/2206.02424>.
- [28] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [29] Park J, Woo S, Lee J Y, et al. BAM: bottleneck attention module[EB/OL]. (2018-07-17)[2022-05-08]. <https://arxiv.org/abs/1807.06514>.
- [30] Woo S, Park J, Lee J Y, et al. Cbam: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer Vision-ECCV 2018. Lecture Notes in Computer Science. Cham: Springer, 2018, 11211: 3-19.
- [31] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 13708-13717.
- [32] Bukhsh Z A, Jansen N, Saeed A. Damage detection using in-domain and cross-domain transfer learning[J]. Neural Computing and Applications, 2021, 33(24): 16921-16936.
- [33] Liu Z K, Yuan L, Weng L B, et al. A high resolution optical satellite image dataset for ship recognition and some new baselines[C]//Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods, February 24-26, 2017, Porto, Portugal. Setúbal: SCITEPRESS - Science and Technology Publications, 2017: 324-331.

Lightweight Ship Detection Based on Optical Remote Sensing Images for Embedded Platform

Wang Huiying¹, Wang Chunping¹, Fu Qiang^{1*}, Han Zishuo², Zhang Dongdong¹

¹*Department of Electronic and Optical Engineering, Shijiazhuang Campus of Army Engineering University, Shijiazhuang 050003, Hebei, China;*

²*32356 Troops of the Chinese People's Liberation Army, Xining 710003, Qinghai, China*

Abstract

Objective Ship detection plays an important role in military and civilian fields such as defense security, dynamic port monitoring, and maritime traffic management. With the rapid development of space remote sensing technologies, the number of high-resolution optical remote sensing images is increasing exponentially, which lays the data foundation for research on ship detection techniques. Meanwhile, it is required that detection systems should have real-time accuracy to match the growth rate of the number of remote sensing images. Traditional methods for object detection are mainly accomplished by the construction of mathematical models or the use of object saliency. However, most of these algorithms rely on the prior knowledge of experts and have certain limitations, which cannot cope with the complex and variable background and the multimodal and heterodyne objects. Recent years have seen the rapid development of deep learning technology. The object detection method based on convolutional neural networks (CNNs) is widely used because of its strong learning ability and high detection accuracy. Currently, mainstream object detection models based on deep learning are mainly divided into two categories, i. e., two-stage networks and single-stage networks. In general, two-stage network detection has high accuracy but is difficult to deploy on embedded devices due to a large amount of computation and huge time consumption. The YOLO series, single-stage network detection algorithms, have received extensive attention and applications due to their simple network structure and consideration of both detection accuracy and detection speed. However, due to the poor computing power and limited memory resources of embedded devices, it is difficult to directly apply single-stage detection models to embedded devices to detect objects in real time. Hence, we expect to deploy a high-performance model to detect ships in optical remote sensing images on equipment terminals with limited resources and space and achieve a lightweight ship detection network for complex remote sensing scene images to promote the landing of the model.

Methods As the existing lightweight object detection algorithms based on deep learning have low detection accuracy and slow detection speed for ships in complex remote sensing scene images, a lightweight real-time ship detection algorithm STYOLO is proposed for embedded platforms. The algorithm uses YOLOv5s as the basic framework. First of all, considering the high memory access costs in the backbone network, the efficient network architecture ShuffleNet v2 is used as the backbone network to extract image features, which reduces memory access costs and improves network parallelism. Secondly, the Slim-neck feature fusion structure is used as the feature enhancement network to fuse the detailed information in the lower-level feature maps to enhance the feature response to small objects. In addition, the coordinate attention mechanism is applied in the multi-scale information fusion region to strengthen object attention and thus improve the ability to detect difficult samples and resist background interference. Finally, a learning strategy combining cross-domain and in-domain transfers is proposed to reduce the difference between source and target domains and improve the transfer learning effect.

Results and Discussions After 100 training iterations with ShuffleNetv2-YOLOv5s, YOLOv5s, MobileNetv3-YOLOv5s, and YOLOv5n on the same test and validation sets, all the evaluated metrics have good performance (Fig. 11), which verifies the effectiveness of the proposed algorithm. On the basis of the YOLOv5s framework, ShuffleNet v2 is used as the backbone network, and Slim-neck is used as the feature enhancement network; the two detection models are trained by cross-domain transfer learning. Compared to the YOLOv5s model, the lightweight model has reduced the detection accuracy, the number of floating points, and the number of parameters by 2.12 percentage points, 62.02%, and 62.05% (Table 2), respectively. To improve the detection accuracy of difficult samples and the ability to counter background interference, we employ the coordinate attention mechanism at the intersection of different information scales in the feature enhancement network. Compared with the results of the detection model without the coordinate attention mechanism, the mAP of the proposed algorithm is improved by 4.94 percentage points, and the number of parameters is raised by 0.75% (Table 3). When different attention mechanisms are applied at the intersection of different information scales in the feature enhancement network, it is found that the model applied with the coordinate attention mechanism has

the highest mAP of 90.46% at a shrinkage rate of 32, an increase of 4.94 percentage points (Table 4). A learning strategy that combines the cross-domain transfer with the in-domain transfer is proposed to reduce the discrepancy between source and target domains and improve transfer learning. The mAP of the proposed algorithm with the above learning strategy is 94.33%, which is 3.87 and 14.17 percentage points higher than that with the training methods of cross-domain transfer learning and in-domain transfer learning, respectively (Table 5). The proposed algorithm is compared with ShuffleNetv2-YOLOv5s, YOLOv5s, MobileNetv3-YOLOv5s, and YOLOv5n on desktop computers and the Jetson Nano terminal. The proposed algorithm achieves a good trade-off between detection speed and detection accuracy in the optical remote sensing ship detection task, and the overall performance is good (Table 6 and Fig. 12).

Conclusions To address the problem that existing lightweight object detection algorithms cannot achieve real-time accurate detection of ships in complex remote sensing scenes, we propose a lightweight real-time algorithm to detect ships in optical remote sensing images for embedded platforms, called STYOLO. Compared to current mainstream detection algorithms used in embedded systems, STYOLO can effectively improve detection speed while ensuring high accuracy. On the Jetson Nano terminal, it has a detection speed of 102.8 frame/s, which is approximately 2.21 times faster than YOLOv5s, 1.36 times faster than ShuffleNetv2-YOLOv5s, 1.70 times faster than MobileNetv3-YOLOv5s, and 1.50 times faster than YOLOv5n. The precision reaches 94.33%, 2.7 percentage points higher than YOLOv5s, 4.19 percentage points higher than ShuffleNetv2-YOLOv5s, 7.27 percentage points higher than MobileNetv3-YOLOv5s, and 24.61 percentage points higher than YOLOv5n, which can meet the requirements of accurate and real-time detection of ships in optical remote sensing images on embedded platforms. In the detection tasks of ships in remote sensing images, visible images are vulnerable to the natural environment, which leads to the weakening of the target features and difficulty in improving the accuracy of the algorithm. Hence, improving the accuracy of weak object detection by combining infrared and visible images for fusion detection is a key research direction.

Key words optical remote sensing image; ship detection; real-time detection; embedded platform; attention mechanism; transfer learning