

基于双流对称特征融合网络模型的海洋船舶目标识别

孙祎芸¹, 樊臻^{1†}, 董山玲^{1,2}, 郑荣濠^{1,2}, 兰剑³

(1. 浙江大学 电气工程学院, 浙江 杭州 310027;

2. 浙江大学 工业控制技术国家重点实验室, 浙江 杭州 310027;

3. 西安交通大学 电子与信息学部, 陕西 西安 710049)

摘要: 海洋船舶目标识别在民用和军事领域有着重要的战略意义, 本文针对可见光图像和红外图像提出了一种基于注意力机制的双流对称特征融合网络模型, 以提升复杂感知环境下船舶目标综合识别性能. 该模型利用双流对称网络并行提取可见光和红外图像特征, 通过构建基于级联平均融合的多级融合层, 有效地利用可见光和红外两种模态的互补信息获取更加全面的船舶特征描述. 同时将空间注意力机制引入特征融合模块, 增强融合特征图中关键区域的响应, 进一步提升模型整体识别性能. 在VAIS实际数据集上进行系列实验证明了该模型的有效性, 其识别精确度能达到87.24%, 综合性能显著优于现有方法.

关键词: 识别算法; 特征级融合; 卷积神经网络; 空间注意力机制

引用格式: 孙祎芸, 樊臻, 董山玲, 等. 基于双流对称特征融合网络模型的海洋船舶目标识别. 控制理论与应用, 2022, 39(11): 2009 – 2018

DOI: 10.7641/CTA.2022.20007

Marine ship target recognition using two-stream symmetric feature fusion convolutional neural network

SUN Yi-yun¹, FAN Zhen^{1,†}, DONG Shan-ling^{1,2}, ZHENG Rong-hao^{1,2}, LAN Jian³

(1. College of Electrical Engineering, Zhejiang University, Hangzhou Zhejiang 310027, China;

2. State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou Zhejiang 310027, China;

3. School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an Shaanxi 710049, China

Abstract: The marine ship target recognition has important strategic significance in civil and military fields, this paper proposes a two-stream symmetric feature fusion convolutional neural network model based on attention mechanism for visible and infrared images, in order to improve the comprehensive recognition performance of ship targets in complex perception environment. The model uses the two-stream symmetric network to extract visible and infrared image features in parallel. By constructing a multi-level fusion layer based on cascade average fusion, the complementary information of visible and infrared modes is effectively used to obtain more comprehensive ship feature description. At the same time, the spatial attention mechanism is introduced into the feature fusion module to enhance the response of key regions in the fusion feature map and further improve the overall recognition performance of the model. A series of experiments on the VAIS real data set have proved the effectiveness of the model, its recognition accuracy can reach 87.24%, and its comprehensive performance is significantly superior to the existing methods.

Key words: recognition algorithm; feature level fusion; convolutional neural networks; spatial attention mechanism

Citation: SUN Yiyun, FAN Zhen, DONG Shanling, et al. Marine ship target recognition using two-stream symmetric feature fusion convolutional neural network. *Control Theory & Applications*, 2022, 39(11): 2009 – 2018

1 引言

当前, 世界各国都将海洋列为战略发展领域, 海洋目标识别技术在军民两端均有良好的应用前景. 船舶

作为海上交通、渔业捕捞、海洋资源开发、军事侦察等诸多领域的重要载体, 其识别任务有着重要的研究价值^[1-3].

收稿日期: 2022-01-05; 录用日期: 2022-04-26.

[†]通信作者. E-mail: fanzhen@zju.edu.cn; Tel.: +86 571-87952706.

本文责任编辑: 闫敬.

NSFC-浙江两化融合联合基金(U1809202)资助.

Supported by the NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (U1809202).

近些年来,海洋船舶目标识别已经成为模式识别领域的研究热点.综合国内外船舶目标识别研究现状,所涉及的船舶图像按照信息来源主要分为:合成孔径雷达图像、红外图像、可见光图像等.合成孔径雷达不受天气、光照等外界环境的影响,可以获取全天候的远距离海域船舶图像,但由于斑点噪声,它的识别精确度较差,容易产生较多错误的识别;红外船舶图像的获取同样不依赖于外界光线,具有全天候的特点,但红外船舶图像对于船舶细节信息的表现一般,多数红外船舶图像分辨率较低,不能反映目标色彩信息^[4];在近距离海域,可见光船舶图像的分辨率高且获取代价低,相比较红外图像与合成孔径雷达图像,可以直观地展示船舶目标的细节信息,例如目标的纹理图案、边缘轮廓以及颜色模块等^[4],从而便于人眼直接观察并作出分析判断,但在夜间以及极端天气条件下成像效果较差.因此,多种传感器获取的船舶图像信息具有一定互补性.

在技术层面,随着深度学习技术的迅速发展,许多研究者陆续将深度学习技术引入船舶目标识别领域,基于深度卷积神经网络(convolutional neural networks, CNN)的船舶识别技术已经成为该领域的研究热点. Khellal 等人在文献[5]中提出了一种基于红外图像的船舶识别系统,通过建立基于极限学习机(extreme learning machines, ELM)的新方法来学习红外图像中有用的CNN特征,并进行快速准确的分类. Li等人在文献[6]中提出了基于CNN的船舶分类方法,并将该方法应用于AlexNet和GoogleNet网络构建两个模型,在ImageNet上利用预训练模型进行迁移学习,该方法在船舶图像上识别性能良好. Solmaz等人在文献[7]中提出基于CNN的多任务学习框架,同时考虑深度特征学习与分类改进损失函数以优化模型分类性能. Huang等人在文献[8]中融合可见光图像深层与浅层特征对船舶图像进行分类,该方法利用CNN学习图像深层特征,通过Gabor滤波器对提取的深层特征和浅层特征进行融合完备,并利用SVM进行分类识别. Shi等人在文献[9]中提出了一种多特征集成的CNN船舶识别模型,该模型在光学图像上较单特征识别模型有较大性能提升. Liu等人在文献[10]中提出了一种基于跨层多任务的船舶目标检测CNN模型,该模型通过特征共享、联合学习和跨层连接等策略,同时解决可见光船舶图像的目标分类、检测、分割问题并取得良好性能. Zhang等人在文献[11]中采用Mask R-CNN深度学习算法,构建了基于深度卷积神经网络的船舶目标特征提取和识别模型,通过对比分析验证了该模型在目标检测和分类方面的性能更好,体现了像素级识别的巨大优势. Chen等人在文献[12]中提出了一种新颖的混合深度学习方法,该方法结合了改进的生成对抗网络和基于卷积神经网络的检测方法,显著

提升了小型船舶检测精确度.

综上所述,现有的海洋船舶识别算法大多使用单源传感器获取的图像进行识别,没有考虑到综合多传感器获取图像来补充单源图像的缺失信息.多传感器信息融合可以为识别精确度的提高提供一种稳健的方式.作为最常用的两类信源,可见光传感器和红外传感器获取图像的融合可以充分利用双源图像互补信息,有望提升目标综合识别性能.不同于一般的目标识别问题,已经证实多数船舶类内差异较小,目标识别算法对其进行类别区分难度较大,尤其是所获取船舶图像复杂的背景将会为船舶识别带来更大的挑战.目前较为先进的多源融合识别模型算法数量较少且识别精度不足,有待进一步改进^[13-15].

近年来发展的视觉注意力机制通过模仿人类视觉系统,在进行图像处理时聚焦于图像空间中重要信息部分而非全部图像特征信息, CBAM注意力机制嵌入通道和空间注意力子模块来增强空间及通道维度上的特征信息^[16], NAM注意力机制利用训练模型权重的方差度量来突出显著特征^[17], GAM注意力机制提出了一种“全局”注意机制,保留信息以放大“全局”跨维度的相互作用^[18],在所有3个维度上捕捉重要特征,该机制可以引导模型学习具有重要信息的图像区域,帮助模型提取目标图像中具有区分度的关键特征信息. SimAM注意力机制无需引入额外参数为特征图推导出注意力权值^[19],而是基于著名的神经科学理论提出优化能量函数以计算注意力权值,该模块能灵活高效地学习跨通道和空间变化的注意力权重. Mobile network坐标注意力机制继承了通道注意方法的优点,该机制模拟通道间的关系,同时利用精确的位置信息捕获多尺度特征^[20].综上所述,将注意力机制引入CNN中可以抑制图像中无用的特征信息,将学习的重点聚焦于船舶目标区域.因此,本文提出了一种基于注意力机制的双流对称特征融合网络模型,通过融合可见光与红外图像特征信息提高识别精确度.

本文的主要贡献如下:

- 1) 提出了一种基于级联平均融合的多级融合方法,有效利用可见光和红外两种模态的互补信息,从而获取更加全面的船舶特征描述;
- 2) 将空间注意力机制引入特征融合模块,增强融合特征图中关键区域的响应,从而提升模型整体识别性能.

2 基于注意力机制的双流对称特征融合网络模型

基于单源光学传感器获取的图像极易受到光照条件与极端天气影响,仅依靠单源传感器图像进行船舶识别遇到极大的挑战,多源信息融合为船舶目标的精确鲁棒识别提供了一种很有前景的手段;另一方面,

深度学习算法可以自适应获取图像中更细致抽象的特征表示,而注意力机制可以获得更具分辨性能的特征表示.因此,针对海洋船舶识别问题,本文提出了一个基于注意力机制的双流对称特征融合网络模型,该网络模型主要由双流对称特征提取模块、引入注意力机制的多级特征融合模块和识别模块组成.其中,特征提取模块利用双流对称卷积网络分别提取预处理

后的可见光图像特征与红外图像特征;基于注意力机制的融合模块增强可见光和红外图像特征描述中的关键信息,并将增强后特征表达进行多级级联平均融合,改善模型特征表达;识别模块由3个全连接层和1个Softmax输出层组成,船舶识别结果通过Softmax输出层获得.具体网络结构如图1所示,本节将对该模型的设计进行详细阐述.

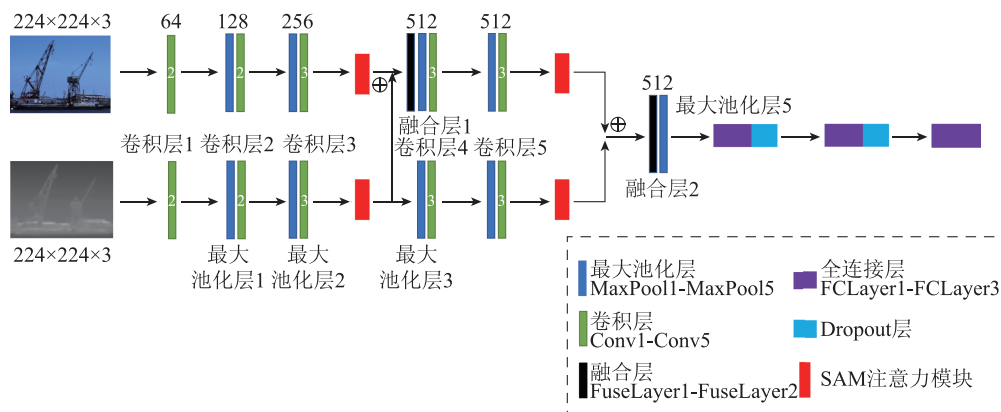


图 1 基于注意力机制的双流对称特征融合网络总体架构

Fig. 1 Architecture of two-stream symmetric feature fusion network based on attention mechanism

2.1 双流特征融合网络整体架构

当前,卷积神经网络已经在图像特征提取上表现出一定的优越性^[21],选择合适的CNN框架来有效提取船舶图像特征是提高识别性能的重要前提.本文提出的双流对称特征融合模型选取VGG16网络^[22]的13个卷积层作为特征提取模块主体结构,并行提取可见光图像特征与红外图像特征.

如图1所示,本文提出的融合模型特征提取模块结构由分别对可见光和红外热图像进行操作的两个卷积网络通道构成.其中,每个特征提取通道由5个卷积层(Conv1-Conv5)和5个池化层(Max-Pool1-MaxPool-5)搭建而成.为充分利用双源图像特征提取通道不同卷积层输出的特征描述,本文设计了基于注意力机制的多级融合模块;在卷积层Conv3之后,引入空间注意力模块(spatial attention module, SAM),随后构建第一级融合层FuseLayer1;在卷积层Conv5之后,引入SAM模块并构建第二级融合层FuseLayer2.融合层FuseLayer1输出的特征图输入可见光特征提取通道继续特征提取步骤,融合层FuseLayer2输出的特征图输入卷积核 2×2 的最大池化层MaxPool5进行特征降维,得到最终融合特征图FeatureFuse.最后,利用三级全连接层构建目标识别模块,为了防止网络过拟合,在三级全连接层中引入Dropout技术^[23],融合特征图FeatureFuse经过全连接层通过Softmax激活函数输出识别的结果概率.特征融合网络的具体结构参数如表1所示.

表 1 双流对称特征融合网络总体参数表

Table 1 Parameter table of two-stream symmetric feature fusion network

类型/层数	步长	卷积核大小及数量	输出
卷积层1 x/2	1	$3 \times 3 \times 3, 64$ $3 \times 3 \times 64, 64$	$224 \times 224 \times 64$
最大池化层1	2	$2 \times 2, 1$	$112 \times 112 \times 64$
卷积层2 x/2	1	$3 \times 3 \times 64, 128$ $3 \times 3 \times 128, 128$	$112 \times 112 \times 128$
最大池化层2	2	$2 \times 2, 1$	$56 \times 56 \times 128$
卷积层3 x/3	1	$3 \times 3 \times 128, 256$ $3 \times 3 \times 256, 256$ $3 \times 3 \times 256, 256$	$56 \times 56 \times 256$
融合层1	1	$1 \times 1 \times 512, 256$	$56 \times 56 \times 256$
最大池化层3	2	$2 \times 2, 1$	$28 \times 28 \times 256$
卷积层4 x/3	1	$3 \times 3 \times 256, 512$ $3 \times 3 \times 512, 512$ $3 \times 3 \times 512, 512$	$28 \times 28 \times 512$
最大池化层4	2	$2 \times 2, 1$	$14 \times 14 \times 512$
卷积层5 x/3	1	$3 \times 3 \times 256, 512$ $3 \times 3 \times 512, 512$ $3 \times 3 \times 512, 512$	$14 \times 14 \times 512$
融合层2	1	$1 \times 1 \times 1024, 512$	$14 \times 14 \times 512$
最大池化层5	2	$2 \times 2, 1$	$7 \times 7 \times 512$
全连接层1	/	/	4096
全连接层2	/	/	4096
全连接层3	/	/	6

本文提出的特征融合网络模型包括一个训练过程和一个测试过程.在训练过程中,该模型采用端到端

的训练,选取配准好的可见光图像与热红外图像对作为训练样本,分别将 224×224 像素大小的可见光和红外图像并行输入网络,特征提取网络模块使用VGG16预先训练的权重来训练,特征融合层以及全连接层使用随机梯度下降(stochastic gradient descent, SGD)最优化求解器训练,然后计算真实标签和Softmax函数得到的预测类标签的误差。之后,通过反向传播过程调整权重和偏差,以最小化误差。最后,保存最优模型。在测试阶段,对同一个标签下的可见光图像和红外图像也进行相同预处理,输入到双流对称特征融合网络中提取特征,并调用最优模型进行特征测试,输出船舶图像的预测标签。

2.2 多级特征融合模块

2.2.1 多级融合层构建

两个特征提取通道输入的原始图像大小均为 224×224 ,经过五级卷积层与池化层时逐渐降低分辨率并增加特征图的通道数量。

以VAIS数据集^[24]中一组可见光与红外图像数据对为例,将融合网络模型双流特征提取网络不同卷积

层输出的特征图可视化,可以得出浅层特征提取主要集中在特征纹理上,得到的特征图更接近原始图像,更深层次的提取更多地关注轮廓和形状等特征,这些特征通常更抽象、更具备代表性。

VGG16网络不同卷积层提取的特征图对应不同层次的语义信息和视觉细节,基于不同层次下的特征图搭建特征融合架构将直接影响模型识别性能,因此,多光谱图像下的船舶识别任务可以建模为卷积特征融合方法设计。根据图2-3的可视化表达,可以看出卷积层Conv1和Conv2输出的特征图 $C1, C2$ 主要捕捉低级视觉特征,如颜色、线段等;卷积层Conv3, Conv4和Conv5输出的特征图 $C3-C5$ 相比浅层特征图 $C1, C2$ 包含更多语义信息,同时也保留部分视觉细节。特征图 $C5$ 感受野较大,提取的特征包含最多抽象语义信息,有利于船舶识别,同时,根据文献[25],中间层特征融合方法可以获得更为良好的识别性能,因此,本文首先利用双流通道卷积层Conv5输出的两个特征图 $C5$ 实现特征融合。同时,考虑到仅融合高阶语义信息无法充分利用多光谱图像信息,本文聚焦关注双源图像特征表达 $C3$ 和 $C4$ 的融合效果。

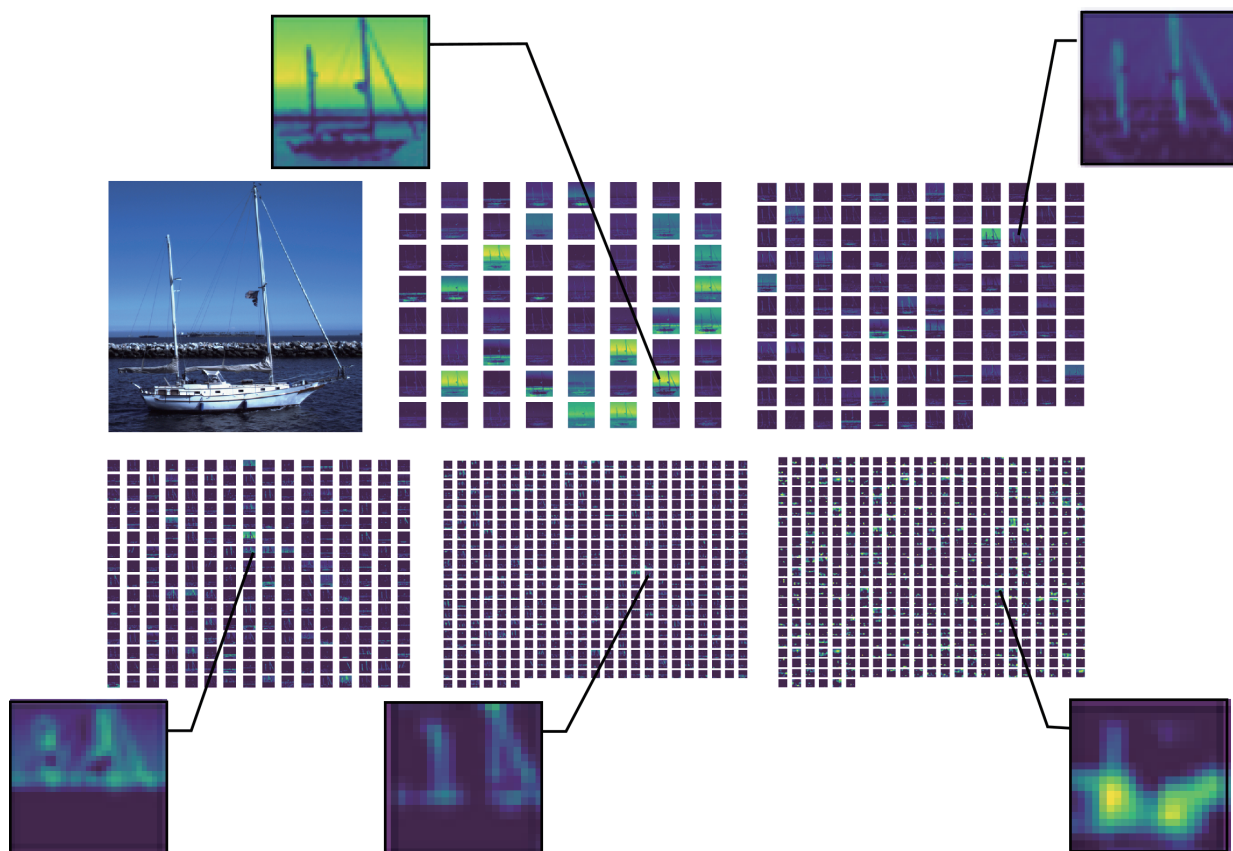


图2 融合网络模型双流特征提取网络不同卷积层输出特征图可视化(可见光图像)

Fig. 2 Normalized confusion matrix of feature fusion model on the actual VAIS data set

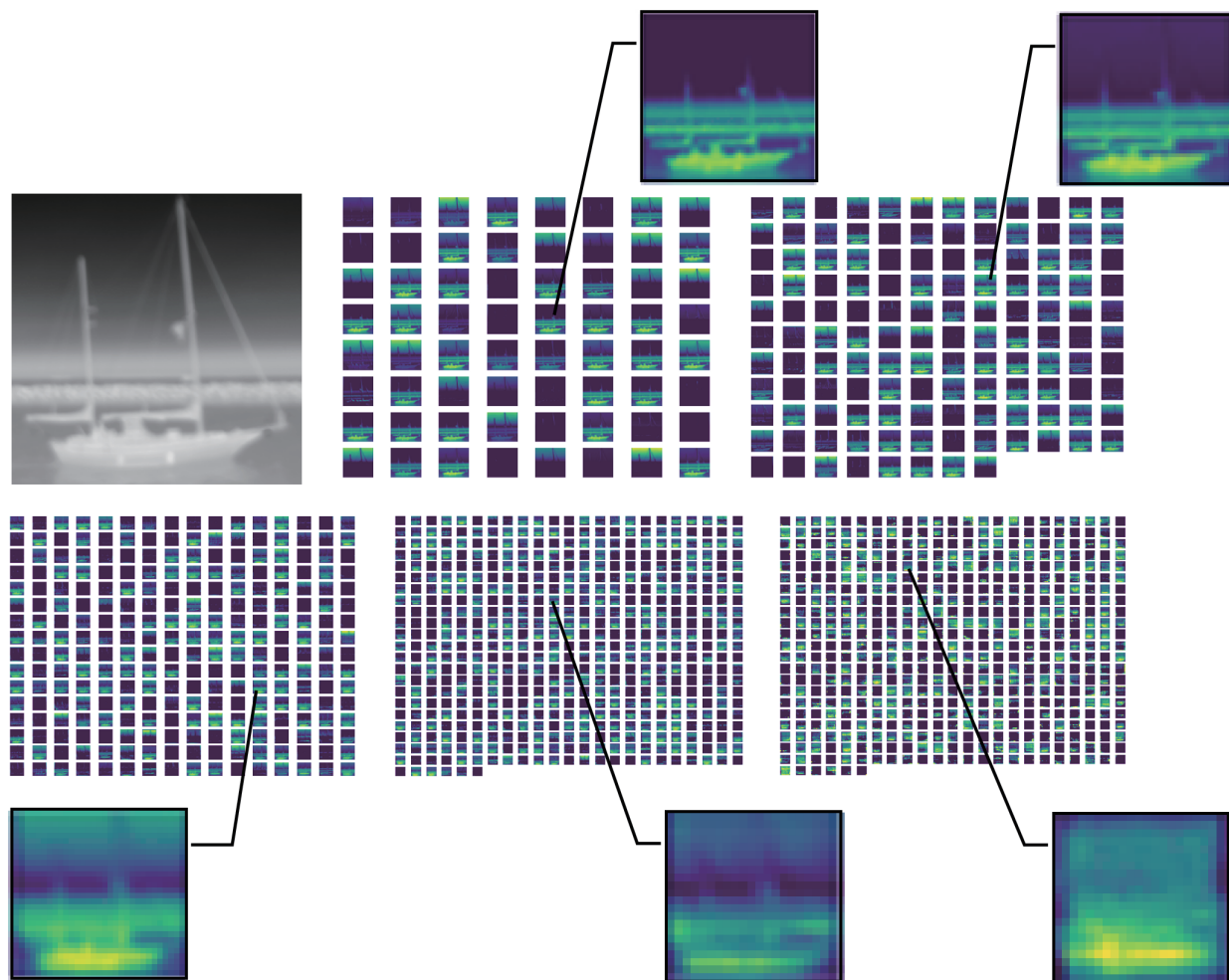


图 3 融合网络模型双流特征提取网络不同卷积层输出特征图可视化(红外图像)

Fig. 3 Normalized confusion matrix of feature fusion model on the actual VAIS data set

由于同一卷积层不同通道获取的特征表达在描述船舶图像时是彼此互补的, 所以可以通过1:1的比例将该层输出的所有通道的特征图叠加获得整体特征图. 图4和图5显示了卷积层Conv3和Conv4输出的多光谱整体特征图和相应融合后特征图的可视化比较. 图4(c)和图5(c)展示的融合特征图特征描述均得到了丰富, 但因为C3特征图特征压缩量较C4特征图更低, 包含更多有利于船舶类别区分的特征细节, 同时也能兼顾提供语义信息, 所以图4(c)中融合特征图增强包含的细节信息较多, 对船舶区域的响应更好, 对后续的特征提取工作能起到更积极的作用. 因此, 在可见光特征提取网络卷积层Conv3后设置融合层FuseLayer1, 对双流通道卷积层Conv3输出的两个特征图实现特征融合, 融合后特征图继续下一级特征提取, 该融合层与卷积层Conv5后设置的融合层FuseLayer2组成特征融合模块.

2.2.2 级联平均融合算法

常见的特征融合方法有加性融合、最大融合、级

联融合等. 将特征融合的表达式定义为

$$F = f(X, Y). \quad (1)$$

令 X, Y 分别表示待融合的特征表达; F 表示融合特征, $X, Y, F \in \mathbb{R}^{HWC}$, 其中, H, W, C 分别表示特征图的高度、宽度和通道数.

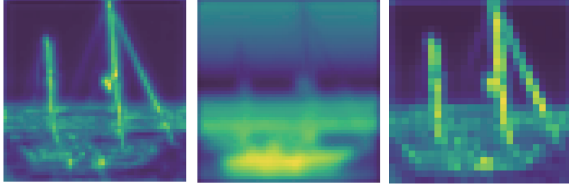
级联融合直接级联拼接所有特征图作为融合结果, 该融合方式可以应用于任意维的特征图. 相较于加性融合和最大融合, 级联融合能够更合理的保留输入特征图的所有元素, 本文基于级联融合引入平均模块, 设计级联平均融合的方法. 级联平均融合首先级联拼接所有特征图, 然后进行 1×1 的卷积运算以获取最终的融合特征图, 可见光与红外图像下级联平均融合用公式可以表示为

$$F = f^{1 \times 1}([x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n]), \quad (2)$$

其中 $f^{1 \times 1}$ 为卷积核大小 1×1 、数量为 n 的卷积操作.

本文设计的多级特征融合模块中输入第一级融合层的两个特征图大小均为 $56 \times 56 \times 256$, 级联拼接后特征图大小为 $56 \times 56 \times 512$, 平均融合后输出的特征

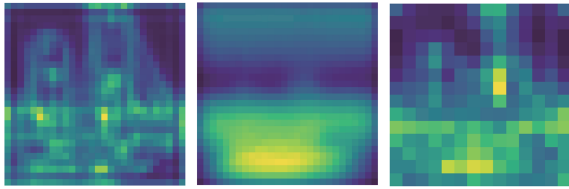
图大小仍为 $56 \times 56 \times 256$, 详细的融合过程如图6所示. 第二级融合层的两个特征图大小均为 $14 \times 14 \times 512$, 级联拼接后特征图大小为 $14 \times 14 \times 1024$, 平均融合后输出的特征图大小仍为 $14 \times 14 \times 512$.



(a) 可见光图像C3 (b) 红外图像C3整 (c) 融合后整体特征图
体特征图 体特征图 特征图

图4 卷积层Conv3输出的整体特征图和相应融合后特征图的可视化比较

Fig. 4 Conv3 convolution layer output of the overall feature map and the corresponding fusion feature map visualization comparison



(a) 可见光图像C4 (b) 红外图像C4整 (c) 融合后整体特征图
体特征图 体特征图 特征图

图5 卷积层Conv4输出的整体特征图和相应融合后特征图的可视化比较

Fig. 5 Conv4 convolution layer output of the overall feature map and the corresponding fusion feature map visualization comparison

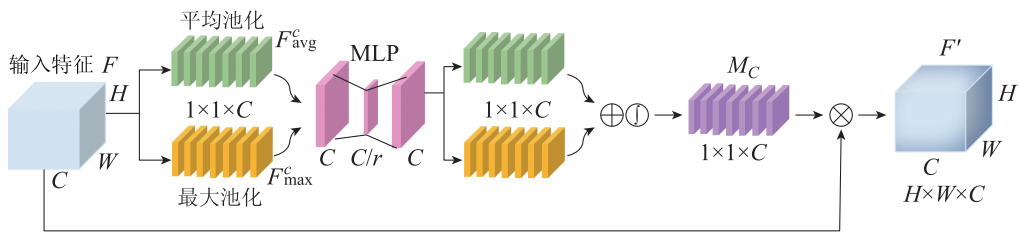


图7 CAM模块构建图

Fig. 7 Construction diagram of CAM module

如图7所示, CAM模块在空间维度上进行了压缩, 在进行Squeeze操作时引入最大池化的思路, 用最大响应的信息对均值池化的信息进行补充.

SAM模块相比起CAM模块首先使用最大池化和均值池化操作对输入的特征图进行压缩操作处理, 然后对输出的特征图在通道层面进行压缩操作处理, 即通过两次池化操作生成两个二维特征向量, 并将其按Concat操作进行拼接, 随后将拼接得到的通道数为2的特征图送入融合卷积层 (network in network, NN) 网络进行融合, 通过卷积的方式保证输出的维度与输

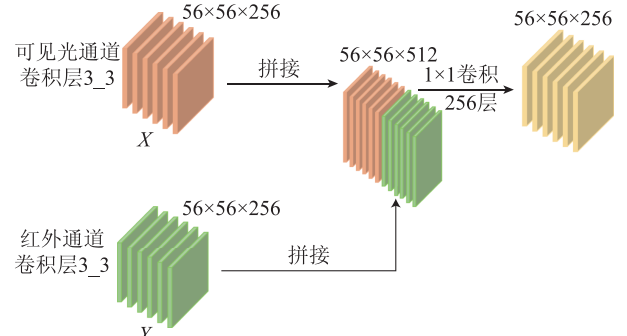


图6 融合层FuseLayer1工作流程图

Fig. 6 Flow chart of fusion layer FuseLayer1

2.3 注意力机制模块

注意力机制是一种模仿人脑的认知机制. 在深度卷积神经网络中引入注意力机制可以使模型聚焦于感兴趣的特征, 从而增强贡献率高的有用信息, 抑制无用信息. 注意力机制可以集成融入CNN网络模型, 与CNN模型一起训练. Woo等提出了一个简单但有效的注意力子网络模块CBAM^[16], 可以满足空间复杂特征的重新整合标定要求. CBAM注意力机制模块由两个独立的部分串联构成, 即通道注意力模块(channel attention module, CAM)和空间注意力模块(spatial attention module, SAM).

CAM模块结构如图7所示, 其工作原理用如下公式说明:

$$F' = M_{CAM}(F) \otimes F, \quad (3)$$

其中: $\otimes$ 符号表示元素点乘, M_{CAM} 表示通道维度上的整合处理.

入的维度一致. SAM模块结构原理如图8所示, 具体工作原理如下公式说明:

$$F'' = M_{SAM}(F') \otimes F', \quad (4)$$

其中: $\otimes$ 符号表示元素点乘, M_{SAM} 表示空间维度上的整合处理, 具体数学表达如下:

$$M_{SAM} = \sigma(f^{3 \times 3}(\text{Avgpool}(F); \text{MaxPool}(F))), \quad (5)$$

其中: σ 为激活函数, $f^{3 \times 3}$ 为 3×3 的卷积操作.

本文尝试将CBAM注意力机制模块集成到双流对称特征融合模型基础网络中, 识别性能有所提升, 对

CBAM机制进行分析, 其中SAM模块能将特征增强聚焦于空间维度上, 从而将提取特征覆盖到待识别船舶的更多部位, 关注利于船舶识别的重点信息, 例如船体的结构与形状等在舰船识别时起关键作用的特征区域, 引入SAM机制模块可以获取待识别船舶图像关键区域更强的特征响应; CAM模块聚焦关注通道层面

上哪些层会具有更强的反馈能力, 在本文模型中通道计算复杂度较低, 因此该模块的引入在增加大量通道权重参数的情况下对关键特征增强增益较小. 通过实验也可以证明CAM模块对特征融合模型性能提升贡献较小, 但其集成至模型后会带来一定程度的结构冗余与时间消耗增加.

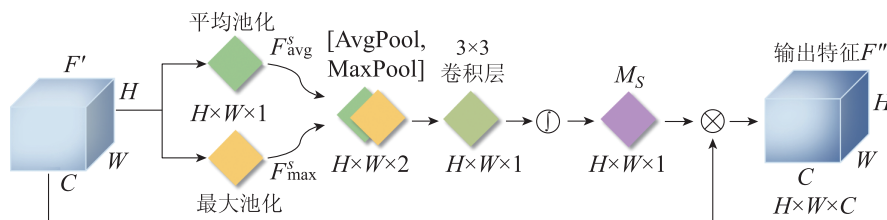


图 8 SAM模块构建图

Fig. 8 Construction diagram of SAM module

因此, 本文仅保留SAM模块并且将其集成至双流特征提取网络中. 如图9所示, SAM模块具体部署位置为特征提取网络卷积Conv3和Conv5输出后, 将获得的可见光与红外图像整合特征分别与第2.2节提出的两级融合层相结合, 两次融合输出最终的融合特征图输入识别网络, 以得到更精准的船舶识别.

Atom 1024, 光谱范围为8–12 μm . 数据集一共包括2865幅图像(1623幅可见光图像和1242幅红外图像), 其中包含1088幅可见光和红外成对图像. 该船舶数据集可以划为6个粗粒度类别, 分别为“cargo”, “medium-other”, “passenger”, “sailing”, “tug”和“small”. 本文实验仅选取配准后的1088对可见光和红外图像进行实验. 用数据集作者文献指定的方式划分训练及测试样本, 划分后的训练及测试样本信息如表2所示.

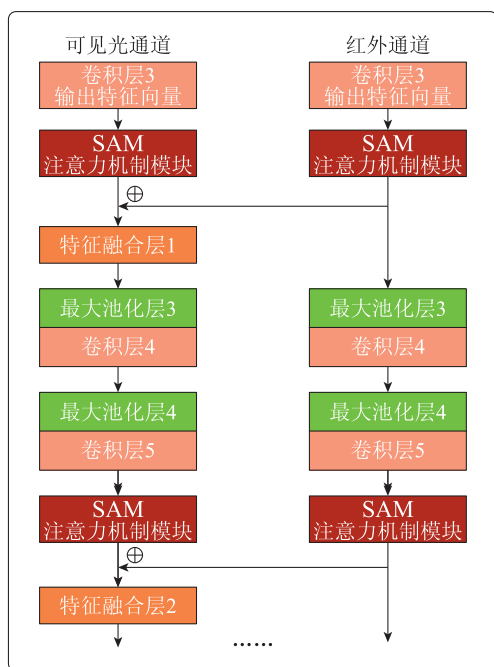


图 9 SAM模块部署图

Fig. 9 Deployment of SAM module

3 实验分析

3.1 数据集与评估方法

本次实验使用的多光谱船舶数据来自VAIS数据集^[16], 该数据集是唯一可用的成对可见光和长波红外船舶图像的公开数据集. 该数据集使用的可见光相机为ISVI IC-C25; 使用的长波红外相机为Sofradi-EC

表 2 VAIS数据集信息

Table 2 Dataset information of VAIS

序号	类别	训练集样本数	测试集样本数
1	Cargo	83	63
2	Medium-other	62	76
3	Passenger	58	59
4	Sailing	148	136
5	Tug	30	20
6	Small	158	195
	总计	539	549

本文采用的船舶识别评价指标包括识别精度 (accuracy, Acc) 和每幅图片平均模型构建时间消耗 (average model building time consumption per image, ATPi).

3.2 实验平台与参数设置

本文实验平台处理器为Intel(R) Core(TM) i9-10900K CPU, 用于深度CNN计算的图形处理器为NVIDIA GeForce RTX 2080, 计算环境是带有TensorFlow后端的Keras环境. 本文参数设置如下: 采用小批量样本方式训练模型, 批量参数设置为16, 训练迭代步数设置为6000; 模型学习速率设置为0.0025, 采用SGD算法优化损失函数, 冲量参数设置为0.9; 全连接层中Dropout参数均设置为0.5. 除此之外, 考虑到现有的VAIS数据量有限、原始数据集各类别数据量

不均衡,在模型训练前,采用随机翻转、随机旋转、变换饱和度与对比度等方式对训练数据集进行数据增强,以提高模型泛化能力,防止模型过拟合,数据增强后的训练样本信息如表3所示。

表3 VAIS数据增强后训练集信息

Table 3 Training set information after VAIS data enhancement

序号	类别	训练集样本数
1	Cargo	415
2	Medium-other	310
3	Passenger	290
4	Sailing	592
5	Tug	150
6	Small	604
	总计	2361

3.3 实验结果

为了验证本文提出的基于注意力机制的双流对称特征融合网络模型算法性能,在相同的实验条件下,将该模型算法与基线算法以及其他先进算法进行比较。

为验证模型融合模块性能,表4基于两个评价指标比较基线方法、其他特征融合方法以及本文提出的特征融合方法在VAIS数据集上的识别性能。

表4 各算法在VAIS数据集上识别性能

Table 4 Recognition performance of each algorithm on VAIS data set

算法	信源	Acc/%	ATPI/ms
VGG16(基线)	可见光	83.28±1.1	8.52
	红外	70.79±2.1	8.44
TS-CFF	可见光+红外	85.00±1.2	16.06
TS-MCFF	可见光+红外	85.95±1.3	16.15
TS-CBAM-MCFF	可见光+红外	86.31±1.7	16.75
TS-SAM-MCFF	可见光+红外	87.24±1.4	16.37

其中,本文选取的基线方法为VGG16算法,将红外和可见光图像并行输入双流网络进行特征提取,并且定义双流对称单级融合网络(two stream-cascade

feature fusion, TS-CFF)为仅在卷积层Conv5后构建单级融合层FuseLayer2得到的识别网络,定义双流对称多级融合网络(two stream-multiple cascade feature fusion, TS-MCFF)为分别在卷积层Conv3和Conv5后构建多级融合层得到的识别网络;定义引入CBAM注意力机制的多级融合网络(two stream-CBAM-multiple cascade feature fusion, TS-CBAM-MCFF)为将CBAM注意力机制引入融合模块得到的识别网络;定义本文最终构建的融合识别网络模型为引入SAM注意力机制的多级融合网络(two stream-SAM-multiple cascade feature fusion, TS-SAM-MCFF),即为将SAM空间注意力机制引入特征融合模块得到的识别网络。

由表4可以看出:相比起使用单模态图像进行目标识别,表中所有多模态图像特征融合算法的识别精度都有了很大的提升,同时每幅图像的平均模型构建时间消耗ATPI的增加都在可接受范围内,因此可以证实采用可见光和红外图像特征级融合能够有效提升船舶识别模型识别精度。

由表4继续分别对TS-MCFF算法、TS-CBAM-MCFF算法与TS-SAM-MCFF算法在实验数据上的识别精确度与ATPI消耗进行比较,可以分析得出,对于网络架构较为复杂的双流对称特征融合算法,在融合模块中引入CBAM机制能改善识别精确度,但是CBAM机制中的CAM模块对模型识别性能的提升几乎没有帮助,只能带来结构的冗余、增加ATPI时间消耗。相比之下,只引入SAM空间注意力机制可以在降低ATPI的同时更有效提高算法模型目标识别精度,这主要是因为SAM模块能强化特征图空间关键位置的学习,这将更有利于模型获取船舶图像区分性特征。

为验证本文提出的TS-SAM-MCFF模型算法在VAIS数据集各类别上的识别性能,表5比较了基线方法、其他特征融合方法与本文提出的特征融合方法在VAIS数据集各类别上的识别精确度。表格5中各算法识别精确度均为十次实验平均结果。如表5所示,与其他方法相比,本文所提出的TS-SAM-MCFF网络模型在实际VAIS数据集上实现了最好的总体识别精度。

表5 各算法在VAIS数据集各类别上的识别精确度

Table 5 Recognition accuracy of each algorithm on VAIS data set

算法	信源	各类别精确度Acc/%					
		medium-other	cargo	passenger	sailing	small	tug
VGG16(基线)	可见光	58.89	90.32	78.38	96.71	80.51	81.11
	红外	39.61	71.62	63.91	85.76	75.32	48.67
TS-CFF	可见光+红外	59.32	97.62	78.15	96.12	84.02	83.75
TS-MCFF	可见光+红外	56.87	98.03	76.16	97.29	87.25	82.91
TS-CBAM-MCFF	可见光+红外	60.39	97.82	75.56	96.84	86.18	83.75
TS-SAM-MCFF	可见光+红外	63.75	97.10	80.23	97.33	88.05	85.02

最后, 将TS-SAM-MCFF模型与近些年提出的其他先进多源特征融合模型进行性能比较, 进一步验证TS-SAM-MCFF的识别性能. 对比方法包括4种基于VAIS数据集的多源特征融合识别模型: 1) Gnostic field^[24], 2) CNN^[24], 3) Gnostic field+CNN^[24], 4) Multimodal CNN, 以及3种基于VAIS数据集的可见光图像识别模型: 1) VGG16 (基线), 2) Gabor + CNN, 3) Gabor + MS-CLBP.

从表6可以看出, TS-SAM-MCFF模型算法在可见光与红外船舶图像上的融合识别精确度较可见光船舶图像识别精确度提高3.61%, 较红外船舶图像识别精确度提高13.62%, 总体来说识别精确度提升显著. 此外, TS-SAM-MCFF与Gnostic field、CNN、Multimodal CNN等3种融合识别模型相比识别精确度有明显提升, TS-SAM-MCFF与Gnostic field + CNN模型相比单模态识别精确度提升, 双模态图像识别精确度近似持平, 但是Gnostic field + CNN模型串联组合4个分类模型实现决策级融合识别(即融合红外图像上Gnostic field、红外图像上CNN、可见光图像上Gnostic field、可见光图像上CNN的概率输出), 该模型训练与识别需要耗费大量时间, 而端到端的TS-SAM-MCFF模型在大幅度降低模型构建时间的同时近似达到该模型的识别精度, 具有更良好的综合性能. 同时, TS-SAM-MCFF模型在单模态图像上的识别精确度优于其他所有先进的多光谱识别算法与单模态识别算法.

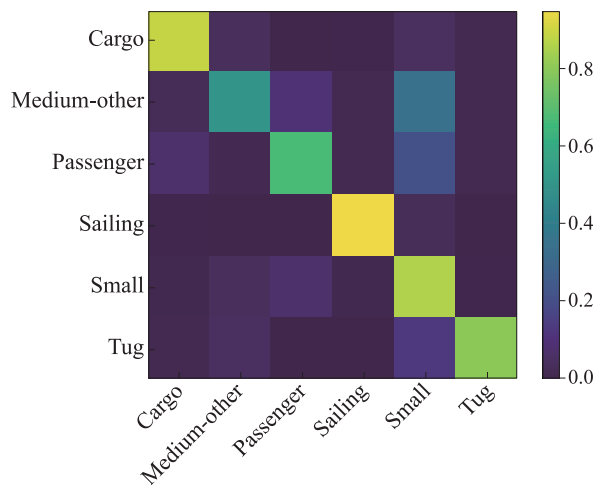
表 6 其他先进识别算法与提出的特征融合识别算法在VAIS数据集上识别性能比较

Table 6 Comparison of other advanced recognition algorithms and feature fusion recognition algorithms on VAIS data set

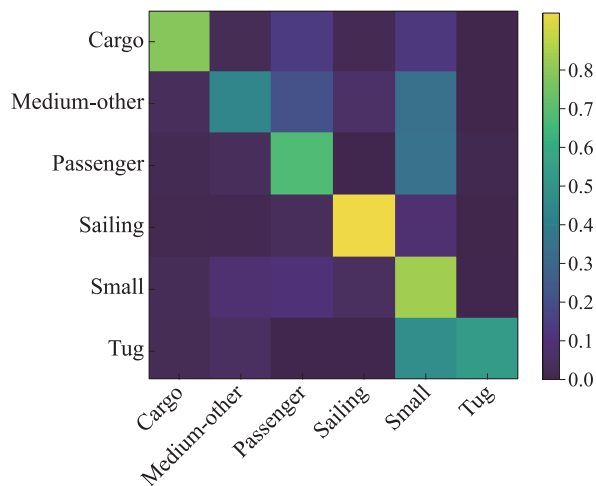
算法	精确度Acc/%		
	红外	可见光	可见光+红外
VGG16(基线)	70.79	83.28	/
Gabor+CNN[19]	/	82.4	/
Gabor+MS-CLBP[20]	/	77.73	/
Gnostic field[16]	58.7	82.4	82.4
CNN [16]	54	81.9	82.1
Gnostic field + CNN [16]	56.8	81	87.4
Multimodal CNN [21]	63.47	81.17	86.67
TS-SAM-MCFF	73.62	83.63	87.24

TS-SAM-MCFF模型在VAIS数据集上识别的归一化混淆矩阵如图10所示, 观察图10(c)可得, 相比图10(a)和(b), 融合后各类别混淆明显减少, 主要的混淆发生在“medium-other”与“small”以及“mall”与“passenger”. “medium-other”识别精确度最低, 与其他类别船舶混淆度较高的原因是该类别中包含多个细粒度, 类间特征差异较其他类别更高, 因此识别

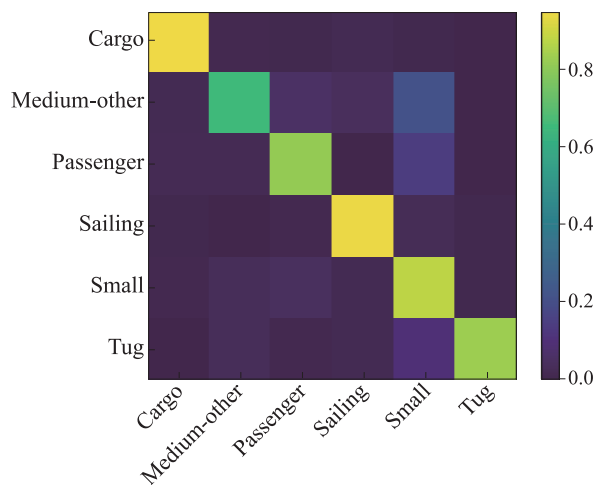
难度更高. “small”与“passenger”发生混淆的原因是这两个类别较为相似, 从视觉上看, 两个船舶类别区别度较低, 需要结合其他语义特征进行进一步识别, 这也是笔者未来的研究方向.



(a) 可见光图像混淆矩阵



(b) 红外图像混淆矩阵



(c) 融合混淆矩阵

图 10 基于注意力机制的双流对称特征融合模型在实际VAIS数据集上归一化混淆矩阵

Fig. 10 Normalized confusion matrix of feature fusion model on the actual VAIS data set

4 总结与展望

当前海洋船舶识别算法大多基于单模态图像进行研究, 识别精确度不甚理想, 为了提高海洋船舶目标识别性能, 本文提出了一种基于注意力机制的双流对称特征融合网络模型(TS-SAM-MCFF). 该模型首先利用基于VGG16的双流对称特征提取网络并行提取可见光和红外图像特征, 然后利用多级融合模块将两个特征提取通道提取的特征图进行级联平均融合, 同时, 在融合模块中集成SAM空间注意力机制增强融合特征关键信息, 输出更良好的融合特征表达, 最后利用全连接层和Softmax函数实现目标识别. 在样本数量有限的情况下, 通过数据增强与Dropout等技巧, 缓解了过拟合现象, 提升了模型的泛化能力. 通过在VAIS多光谱实际数据集上的系列实验验证了本文提出模型的性能优越性. 因为多源图像识别数据集较为缺乏, 未来希望能自行构建其他多源传感器船舶数据集进一步验证本文算法的鲁棒性. 同时, 本文构建的融合识别模型能获取更加详细的待识别物体特征描述并有效增强待其重点区域特征响应, 因此, 该算法不仅能有效提升船舶识别精度, 未来希望将其拓展到车辆、行人等其他具有挑战性的目标识别领域.

参考文献:

- [1] DING Peng, ZHANG Ye, JIA Ping, et al. Ship detection on sea surface based on visual saliency. *Acta Electronica Sinica*, 2018, 46(1): 127 – 134.
(丁鹏, 张叶, 贾平, 等. 基于视觉显著性的海面舰船检测技术. 电子学报, 2018, 46(1): 127 – 134.)
- [2] WANG Yanqing, MA Lei, TIAN Yuan. State-of-the-art of ship detection and recognition in optical remotely sensed imagery. *Acta Automatica Sinica*, 2011, 37(9): 1029 – 1039.
(王彦清, 马雷, 田原. 光学遥感图像舰船目标检测与识别综述. 自动化学报, 2011, 37(9): 1029 – 1039.)
- [3] KANJIR U, GREIDANUS H, OSTIR K. Vessel detection and classification from spaceborne optical images: A literature survey. *Remote Sensing of Environment*, 2018, 207: 1 – 26.
- [4] AISHWARYA N, THANGAMMAL C B. Visible and infrared image fusion using DTCWT and adaptive combined clustered dictionary. *Infrared Physics & Technology*, 2018, 93: 300 – 309.
- [5] KHELLAL A, MA H, FEI Q. Convolutional neural network based on extreme learning machine for maritime ships recognition in infrared images. *Sensors*, 2018, 18(5): 1490 – 1509.
- [6] LI Z Z, ZHAO B J, TANG L B, et al. Ship classification based on convolutional neural networks. *The Journal of Engineering*, 2019, 2019(21): 7343 – 7346.
- [7] SOLMAZ B, GUNDOGU E, YUCESAY V, et al. Fine-grained recognition of maritime vessels and land vehicles by deep feature embedding. *IET Computer Vision*, 2018, 12(8): 1121 – 1132.
- [8] HUANG L, LI W, CHEN C, et al. Multiple features learning for ship classification in optical imagery. *Multimedia Tools and Applications*, 2018, 77(11): 13363 – 13389.
- [9] SHI Q, LI W, TAO R, et al. Ship classification based on multifeature ensemble with convolutional neural network. *Remote Sensing*, 2019, 11(4): 419 – 440.
- [10] LIU Z, WAQAS M, YANG J, et al. A multi-task CNN for maritime target detection. *IEEE Signal Processing Letters*, 2021, 28: 434 – 438.
- [11] ZHANG D, ZHAN J, TAN L, et al. Comparison of two deep learning methods for ship target recognition with optical remotely sensed data. *Neural Comput and Applic*, 2021, 33(10): 4639 – 4649.
- [12] CHEN Z, CHEN D, ZHANG Y, et al. Deep learning for autonomous ship-oriented small ship detection. *Safety Science*, 2020, 130: 104812.
- [13] KANG X, LI C, LI S, et al. Classification of hyperspectral images by gabor filtering based deep network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017, 11(4): 1166 – 1178.
- [14] HUANG L, LI W, CHEN C, et al. Multiple features learning for ship classification in optical imagery. *Multimedia Tools and Applications*, 2018, 77(11): 13363 – 13389.
- [15] AZIZ K, BOUCHARA F. Multimodal deep learning for robust recognizing maritime imagery in the visible and infrared spectrums. *Proceedings of the International Conference Image Analysis and Recognition*. Cham: Springer, 2018: 235 – 244.
- [16] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module. *Proceedings of the European Conference on Computer Vision*, Munich: arXiv, 2018: 3 – 19.
- [17] LIU Y, SHAO Z, TENG Y, et al. NAM: Normalization-based attention module. *arXiv preprint: arXiv*, 2021: 2111.12419.
- [18] LIU Y, SHAO Z, HOFFMANN N. Global attention mechanism: Retain information to enhance channel-spatial interactions. *arXiv preprint: arXiv*, 2021: 2112.05561.
- [19] YANG L, ZHANG R Y, LI L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks. *International Conference on Machine Learning*, PMLR, 2021: 11863 – 11874.
- [20] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design. *Conference on Computer Vision and Pattern Recognition*, 2021: 13713 – 13722.
- [21] JIANG Zetao, QIN Jiaqi, ZHANG Shaoqing. Parameterized pooling convolution neural network for image classification. *Acta Electronica Sinica*, 2020, 48(9): 1729 – 1734.
(江泽涛, 秦嘉奇, 张少钦. 参数池化卷积神经网络图像分类方法. 电子学报, 2020, 48(9): 1729 – 1734.)
- [22] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint: arXiv*, 2014: 1409.1556.
- [23] BALDI P, SADOWSKI P J. Understanding dropout. *Advances in Neural Information Processing Systems*, 2013, 26: 2814 – 2822.
- [24] ZHANG M M, CHIO J, DANIILIDIS K, et al. VAIS: A dataset for recognizing maritime imagery in the visible and infrared spectrums. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Boston, MA, USA: IEEE, 2015: 10 – 16.
- [25] LIU F, SHEN T, MA X, et al. Ship recognition based on multi-band deep neural network. *Journal of the International Societies for Precision Engineering and Nanotechnology*, 2017, 25: 2939 – 2946.

作者简介:

孙祎芸 博士研究生, 目前研究方向为信息融合、多智能体系统强化学习, E-mail: 12110066@zju.edu.cn;

樊臻 博士, 讲师, 目前研究方向为水下信息感知与处理、多传感器信息融合, E-mail: fanzhen@zju.edu.cn;

董山玲 博士, 特聘研究员, 目前研究方向为信息融合、自主式水下航行器(AUV)控制、网络化混杂系统、海洋信息技术, E-mail: shanlingdong28@zju.edu.cn;

郑荣濠 博士生导师, 副教授, 目前研究方向为多智能体系统协同控制与优化, E-mail: rzheng@zju.edu.cn;

兰剑 博士生导师, 教授, 目前研究方向为多源异构数据关联与融合、目标检测识别与跟踪, E-mail: lanjian@mail.xjtu.edu.cn.