特征融合的卷积神经网络多波段舰船目标识别

刘 峰[1], 沈同圣[2], 马新星[1]

[1]海军航空工程学院控制工程系, 山东 烟台 264001;

[2]中国国防科技信息中心, 北京 100142

摘要 针对海面背景舰船目标单一波段图像识别率低的问题,提出了一种基于卷积神经网络(CNN)的融合识别方法。该方法提取可见光、中波红外和长波红外 3 个波段舰船目标特征进行融合识别。模型主要分为 3 个步骤:通过设计的 6 层 CNN,同时对三波段图像进行特征提取;利用基于互信息的特征选择方法对串联的三波段特征向量按照重要性进行排序,并按照图像清晰度评价指标选取固定长度的特征向量作为目标识别依据;通过额外的 2 个全连接层和输出层进行回归训练。采用自建的三波段舰船图像数据库进行模型的训练和测试,共包含 6 类目标,5000 余张图像。实验结果表明,本文方法识别率达到 84.5%,与单波段识别方法相比有明显提升。

关键词 机器视觉; 目标识别; 特征融合; 卷积神经网络; 多波段图像; 特征选择; 图像清晰度

中图分类号 TP391 **文献标识码** A

doi: 10.3788/AOS201737.1015002

Convolutional Neural Network Based Multi-Band Ship Target Recognition with Feature Fusion

Liu Feng[1], Shen Tongsheng[2], Ma Xinxing[1]

[1] *Department of Control Engineering, Naval Aeronautical and Astronautical University, Yantai, Shandong* 264001, *China*;

[2] *China Defense Science and Technology Information Center, Beijing* 100142, *China*

Abstract In order to improve the recognition rate in single-band images of ship targets with complex background, we propose a new fusion recognition method based on convolutional neural networks (CNN). This method extracts the ship target features of images in three wave bands, which are visible light, medium-wave infrared and long-wave infrared images. The model is divided into three steps. Firstly, a 6-layer CNN model is designed to extract the image features of three bands simultaneously. Secondly, a feature selection method based on mutual information is used for sorting the concatenated features according to the importance, and then the feature vectors of fixed dimension can be chosen depending on the indicator of image clarity evaluation. The dimension-reduced feature vector is regarded as the basis of target recognition. Finally, a 2-layer fully connected networks and an output layer are designed for training and regression. We build a triple-band ship target dataset for our experimental verification, which contains 6 categories of targets and more than 5000 images. The experimental results show that the recognition rate of the proposed method can reach 84.5%, which is improved significantly compared to that of the single-band recognition method.

Key words machine vision; target recognition; feature fusion; convolutional neural networks; multi-band images; feature selection; image definition

OCIS codes 150.0155; 110.2970; 110.4234; 100.3008

收稿日期: 2017-04-10; **收到修改稿日期**: 2017-05-15

基金项目: 国家自然科学基金(61303192)

作者简介: 刘 峰(1988—),男,博士研究生,主要从事计算机视觉、图像处理、目标检测等方面的研究。E-mail: liufeng_cv@126.com

导师简介: 沈同圣(1966—),男,博士,教授,主要从事精确制导、目标系统智能化等方面的研究。E-mail: tongsheng_shen@163.com

1 引　言

随着海洋环境的开发和利用日渐增多，海上舰船目标的准确识别无论在军事还是民用领域都得到广泛的应用，如海上搜救、渔船监控、精确制导武器以及多方面的潜在海洋威胁等。可见光图像分辨率高，细节纹理清晰，并且对目标的区分度好。红外图像不受光照情况影响，可满足夜间无光情况下的工作需要，若能利用不同传感器成像的优点进行融合识别，可以有效扩展复杂条件下多波段图像目标识别的适用范围，并提高识别率。

目前常用的目标识别方法主要分为两类：基于特征的方法和基于神经网络的方法。前者主要依赖人工设计的特征向量，并结合目标自身具有的结构、纹理、颜色等特征作为判别依据，进行目标的识别和分类。常用的有方向梯度直方图(HOG)特征[1-3]、尺度不变特征变换(SIFT)算法[4]、Fisher 向量[5]等。

张迪飞等[1]利用 HOG 特征，结合支持向量机(SVM)的方法，对海面上的红外舰船目标进行识别分类，在一定程度上克服了背景的干扰，但是实验中并没有对多种目标及形变、光照等变化进行测试。Feineigle 等[4]利用 SIFT 算法描述子对港口中的舰船目标进行识别，通过对目标局部特征的描述和匹配实现了对光照和角度的不变性，但是利用滑动窗口对目标进行特征提取造成维数过高，计算效率较低。Sánchez 等[5]利用 Fisher 向量结合高斯混合模型对大规模数据集进行线性分类，目标类别包含千种以上，通过对最损失函数进行优化得到了当时最好的分类准确率。Smeelen 等[6]利用基于协方差的融合方法，将可见光和红外图像特征向量进行特征级融合，但该方法仅限于较低维度的特征融合，高维情况会使计算量显著增加。

基于卷积神经网络(CNN)的识别方法是一种端到端的模型结构，使图像可以直接作为网络的输入，避免了传统识别算法中复杂的特征提取和数据重建过程。多层的卷积结构设计使得该网络对平移、比例缩放、倾斜或者其他形式的形变具有高度不变性[7]。目前，该方法在图像分类[8-10]、目标检测[11-12]、显著性分析[13]等众多计算机视觉领域取得了突破性的进展。Krizhevsky 等[8]提出的 AlexNet 模型在图像分类挑战赛中赢得了当年的第一名，证明了 CNN 在复杂条件下的有效性，并使用图形处理器(GPU)使大数据训练在可接受的时间范围内得到了结果。在此基础上 He 等[10]将 CNN 的模型深度扩展到 152 层，使得 ImageNet 大规模视觉识别挑战赛(ILSVRC)的目标分类识别率已经达到甚至超越了人类的识别能力。Bousetouane 等[14]针对港口中的舰船目标，提取目标候选区域，利用超快区域卷积神经网络(Faster-RCNN)[11]方法进行训练提取目标特征，可同时识别多种舰船目标及背景。Zhang 等[15]建立可见光/长波红外双波段数据集，采用牛津大学视觉研究组提出的 VGG-16[9]神经网络，并且在单波段图像无法获取目标时，利用另一种波段图像对目标进行识别。通过足够多转换的组合，可以学习到更加复杂的函数表达。

以上述分析为基础分析，本文提出一种基于 CNN 的多波段舰船目标融合识别方法。该方法可以分别提取三波段图像特征，并利用基于互信息的特征选择方法和图像清晰度指标对串联特征进行自适应降维，最后通过全连接层实现目标的分类。

2 融合算法模型

2.1 算法总体框架

本文算法一方面克服了目前 CNN 模型仅针对单一波段图像作为数据输入的限制，可同时提取三波段图像的目标特征，并对其进行有效的融合识别；另一方面限于多波段图像数据集的资源较少，在对模型的研究过程中，自建三波段舰船目标数据集进行实验验证，该数据集共包含 6 类目标，5000 余张图像。设计的三波段融合识别方法主要包含 3 个步骤：1)利用改进的 AlexNet 网络实现对三波段图像的并行特征提取；2)利用基于互信息的特征选择方法对串联的融合特征进行降维，去除无关的特征向量；3)添加 2 个全连接层和 1 个输出层对网络进行回归训练，得到目标分类结果。本文算法流程如图 1 所示。

2.2 特征提取网络结构

将改进的 AlexNet 作为网络特性提取的基本结构，利用神经网络的迁移学习能力，将 ILSVRC12 中训练好的模型作为网络的初始化参数，再利用自建的多波段数据集进行微调。设计的网络结构如图 2 所示。

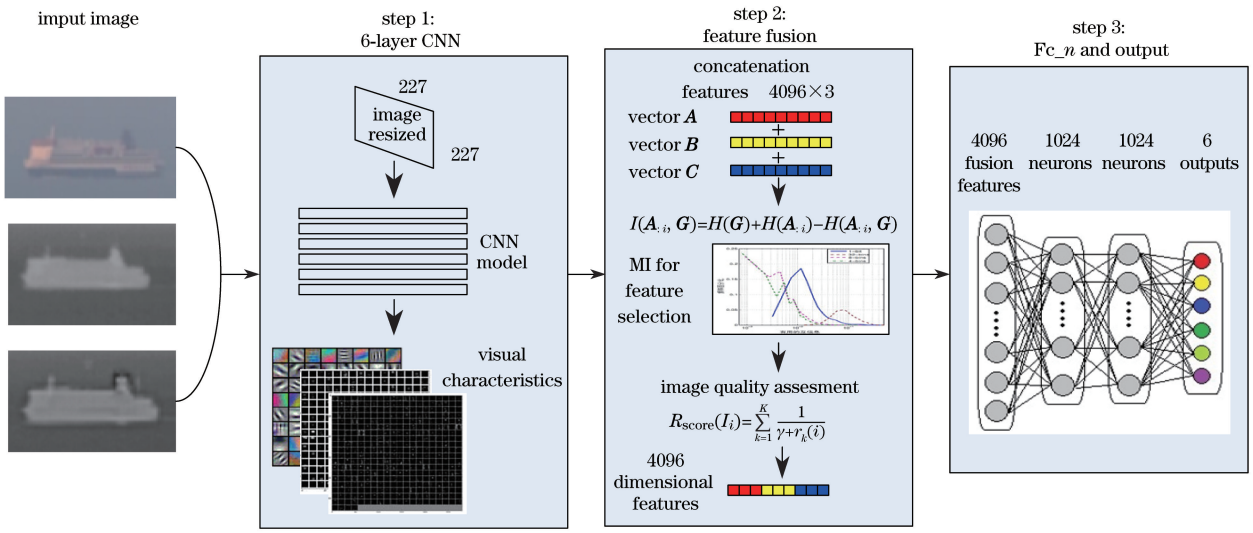

图 1　本文算法流程

Fig. 1　Flow of the proposed algorithm

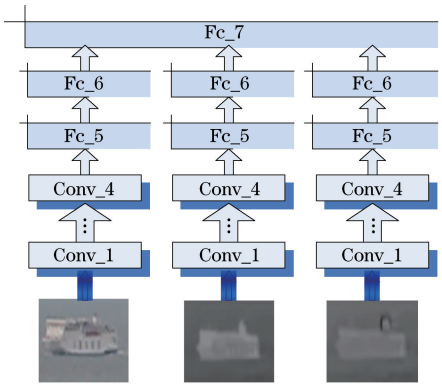

图 2　网络结构示意图

Fig. 2　Diagram of the network structure

将同一目标的三波段图像并行送入 3 个相同的神经网络进行特征提取，网络的具体参数如表 1 所示，每层结构略有不同但都包含待学习参数。

表 1　神经网络结构及参数

Table 1　Structure and parameters of the neural network

Parameter	Structure of network					
Layer	Conv_1	Conv_2	Conv_3	Conv_4	Fc_5	Fc_6
Structure	C+R+L+P	C+R+P	C+R+P	C+R+P	F+R+D	F+R+D
Input	227×227×3	27×27×96	13×13×256	13×13×384	6×6×256	4096
Neuron	96	256	384	256	4096	4096
Kernal size	11×11	5×5	3×3	3×3	1×1	1×1
Stride	4	2	1	1		
Pooling	3×3	3×3		3×3		
Pooling stride	2	2		2		
Train parameter	96×(11×11×3+1)	256×(5×5×96+1)	384×(3×3×256+1)	256×(3×3×384+1)	4096×(3×3×256+1)	4096×(4096+1)

表 1 中 C 表示卷积层，通过卷积运算，可以使原图像特征增强，并且降低噪声。R 表示非线性激活函数(RELU)[16]，与传统的 sigmoid 激活函数相比，RELU 可以加速收敛过程，使得网络自行引入稀疏性，等效于对网络进行无监督学习的预训练。L 表示局部响应标准化，在提取低层特征时增加网络的泛化能力，用 $a^{i}_{x,y}$

表示位置(x,y)处用卷积核i计算时得到的活跃度，应用RELU，则响应标准化的活跃度$b^i_{x,y}$表示为

$$b^i_{x,y}=\frac{a^i_{x,y}}{\left[k+\alpha\sum\limits_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)}(a^i_{x,y})^2\right]^{\beta}},\tag{1}$$

式中n为求和操作遍历的空间位置相邻特征映射数，N为同一层中核的个数。P表示最大池化操作，计算特征图中的局部最大值，相邻的池化单元通过移动一行或者一列从小块上读取数据，减少表达的维度并使数据具有平移不变性。F表示全连接层，D表示dropout正则技术，该方法随机地将某些单元隐藏，隐藏的单元不参与CNN的训练过程。因此，当每次有输入时，网络采样一个随机结构，该方法降低了神经元之间的共适应性，可有效防止网络发生过拟合。

在训练过程中，针对不同尺寸的输入图像，需要将其映射为227 pixel×227 pixel的矩形，以适应网络结构的输入。Krizhevsky等[8]验证了不同的映射方法对识别率的影响，本文采用双线性插值法。将输入图像减去像素均值后利用CNN进行训练，通过前向传播逐层提取特征，在第6层得到4096维的特征向量，记为向量$\boldsymbol{A}$。同理，利用相同的CNN对中波红外和长波红外图像进行特征提取，分别得到向量$\boldsymbol{B}$和$\boldsymbol{C}$。将3组向量按顺序进行串联组成融合特征向量，该向量包含了不同波段下目标的特征信息。原始图像中通常缠绕着高度密集的特征，如果能够解开特征间缠绕的复杂关系，转换为稀疏特征，则特征具备稳健性。此外，如图3所示，深度神经网络中每层待优化参数主要集中在全连接层。利用合理的降维方法去除高维特征向量中的冗余和噪声信息，在减小计算量的同时还可以提高识别的准确率。

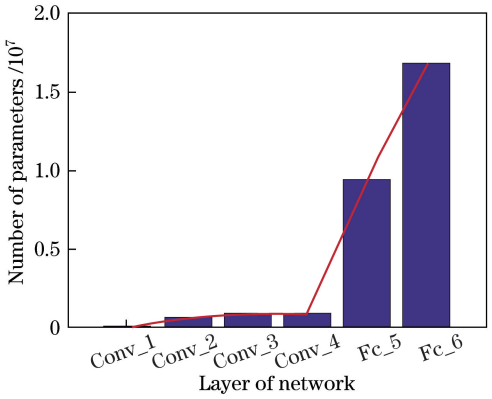

图3　不同层网络参数对比

Fig. 3　Contrast of parameters for different layers

2.3　自适应权重的互信息特征选择

特征压缩是通过投影将所有信息进行压缩，保留的信息仍含有一定的冗余和噪声，而特征选择的过程是通过舍弃冗余信息而保留对分类贡献较高的有用信息。采用基于互信息特征选择的方法对串联特征进行降维，并按照重要性进行排序，该方法可根据需要任意设定阈值，选择不同维度的特征向量，而不再需要重新计算。串联融合的特征按照可见光、中波红外、长波红外的顺序进行排列，在降维过程中3类图像的特征向量彼此互不干扰。特征的排序与选择仅在同一幅图像的特征向量中进行。为了确保融合后特征向量具有固定的维数，利用图像清晰度评价指标可以自适应确定不同波段图像的阈值。

利用文献[17]中基于互信息的特征选择方法计算维度和标签之间的互信息是一种基于监督的方法，以可见光图像特征向量$\boldsymbol{A}$为例，数据集中所有可见光图像的第i维向量为$\boldsymbol{A}_{:,i}$，标注的图像类别标签为$\boldsymbol{G}$，其互信息为$I(\boldsymbol{A}_{:,i},\boldsymbol{G})$。一般来说，互信息越大，则这一维向量用于分类越有效。互信息的值是对每一维向量重要性的评估，计算公式为

$$I(\boldsymbol{A}_{:,i},\boldsymbol{G})=H(\boldsymbol{G})+H(\boldsymbol{A}_{:,i})-H(\boldsymbol{A}_{:,i},\boldsymbol{G}),\tag{2}$$

式中H为随机变量的熵，$\boldsymbol{A}_{:}$为样本数量。图像标签$\boldsymbol{G}$对于不同的维数i保持不变，即$\boldsymbol{G}$为定值向量矩阵，因此它的熵也保持不变，则互信息的计算排序只需要求解$H(\boldsymbol{A}_{:,i})-H(\boldsymbol{A}_{:,i},\boldsymbol{G})$即可。根据互信息的值按照降序对所有$N$维向量进行排序，若要将$N$维向量降到$D$维，只需取互信息排序前$D$名的向量即可。

利用定量的离散变量方法代替核密度分布计算随机变量$\boldsymbol{A}_{:,i}$的信息熵。一种典型而有效的计算随机变量$\boldsymbol{A}_{:,i}$熵的方法是通过核密度分布估计其概率分布，但是对于大规模数据计算量过大，这里采用标准化的离散变量方法计算熵值，即通过特征向量中最大值和最小值，平均划分为n个空间。信息熵可以表示为

$$H(x)=-\sum_j p_j \mathrm{lb}\ p_j, \tag{3}$$

式中p_j表示x被划分到n维空间中第j维的概率。对特征向量按照重要性排序之后，可以通过阈值选择特定长度的三波段特征向量串联组成新的特征向量$\boldsymbol{F}_{3\mathrm{CNN}}$。设定串联特征$\boldsymbol{F}_{3\mathrm{CNN}}$为4096维，阈值的选取按照图像清晰度评价(IQA)标准，参照文献[18]中的倒数排名融合(RRF)方法，可计算得到多种评价指标的综合得分为

$$R_{\mathrm{score}}(I_i)=\sum_{k=1}^{K}\frac{1}{\gamma+r_k(i)}, \tag{4}$$

式中$r_k(i)$为图像I_i在第k个评价标准中的排名；γ为常数，取$\gamma=60$。

评价指标如下：

1) 梯度相似性偏差(GMSD)[19]，计算像素间梯度相似性；

2) 视觉信息保真度指数(VIF)[20]，基于小波变换的多尺度高斯混合模型，计算图像失真程度；

3) 特征相似性指数(FSIM)[21]，测量图像的梯度幅值和相位一致性；

4) 颜色特征相似性指数(FSIMC)[21]，在可见光图像中，与特征相似指数相比增加了颜色信息；

5) 结构相似性指数(SSI)[22]，测量图像中物体的结构失真程度。

(4)式是对多种评价方法进行综合考虑，得到单张图像相对于整个数据集中图像的质量指标，并不是对图像清晰度的客观评价。分别对三波段图像计算$R_{\mathrm{score}}(I_i)$值，并对其进行归一化，得到不同波段特征向量的权重值，计算公式为

$$F_{3\mathrm{CNN}}=\overbrace{R_{\mathrm{score}}(I_{\mathrm{vis}})}^{A}\boldsymbol{A}+\overbrace{R_{\mathrm{score}}(I_{\mathrm{MWIR}})}^{B}\boldsymbol{B}+\overbrace{R_{\mathrm{score}}(I_{\mathrm{LWIR}})}^{C}\boldsymbol{C}, \tag{5}$$

式中$\overbrace{\quad}$表示归一化操作，串联后的三波段图像特征$\boldsymbol{F}_{3\mathrm{CNN}}$为4096维。当某个波段拍摄的图像清晰度较差时，通过归一化的串联特征选择可以有效减少该波段图像特征的选择维数，减小不清晰波段图像对目标识别造成的影响。选择后的特征利用图2网络中第3步中额外2层全连接层及输出层，对融合后的特征向量进行回归训练，输出不同目标的类别概率，其中全连接层每层包含1024个神经元，输出层利用softmax函数对不同类别舰船目标进行分类。

2.4　数据集构建与训练

利用具有共视轴的三轴经纬仪对海上舰船目标进行拍摄，采样帧频均为1 s，同一时刻拍摄的三波段图像作为一个整体进行存储。可见光图像分辨率为1024 pixel×768 pixel，中波传感器工作波段为3.7～4.8 μm、图像分辨率为320 pixel×256 pixel，长波传感器工作波段8～14 μm、图像分辨率为640 pixel×480 pixel。拍摄海面上行驶的舰船在不同时刻、不同背景下的图像，构建多波段舰船图像目标数据库，共包括6类目标，5187幅图像。数据库中包括游轮A 354×3幅，游轮B 337×3幅，铁路轮渡208×3幅，货船236×3幅，小型渔船291×3幅，某型军舰303×3幅。在训练之前需要对数据集中的目标进行类别标注，并按照随机采样的方式将其按照50%，20%和30%的比例划分为训练集、验证集和测试集。网络训练采用随机梯度下降(SGD)方法，批处理尺寸$m=32$，冲量为0.9，权重延迟为0.0005，初始学习率为0.01，当代价函数趋于稳定后学习率降低为0.001，学习周期为100。仿真验证平台为Ubuntu14.0，处理器为i5-4590，显卡为gtx1080，内存为16 GB，采用caffe深度学习框架进行网络的构造和训练，在迭代10^5次的情况下，训练时间约为4 h。

3　实验验证及对比

图4为数据集中3种波段不同目标的示例图片。实验验证分为两部分：1)验证不同维度的融合特征向量对识别率的影响，确定选取的融合特征维度；2)分别利用4种不同的识别方法与本文方法进行对比，并分析误识别产生的原因。

图 4　目标识别数据库示例图片。(a)游轮 A;(b)游轮 B;(c)渔船;(d)铁路轮渡;(e)军舰;(f)货船

Fig. 4　Examples in target identification database. (a) Cruise A; (b) cruise B; (c) fisher; (d) railway ferry; (e) warship; (f) merchant ship

三波段图像的融合特征维度直接影响本文算法的识别率和计算时间，通过实验确定融合特征的特征维度 $\boldsymbol{F}_{3\mathrm{CNN}}$。串联后的三波段图像串联特征共 12288 维，从 $\boldsymbol{F}_{3\mathrm{CNN}}=2048$ 开始，以 256 维间隔选取一次，共取 41 个不同的串联维度测试模型的识别率。如图 5 所示，横坐标为串联的特征向量维度，上方曲线表示不同维度下的识别率，对应左侧纵坐标，下方直方图表示不同维度特征向量所对应的全连接层神经元数量，对应右侧纵坐标。由图 5 可以看出，随着串联特征维度的增加，识别率趋于平缓，甚至出现逐渐下降的趋势。这是由于串联的特征向量中包含的无用噪声信息对识别造成了干扰，若不进行有效的特征筛选，识别效果与单波段识别类似，达不到融合识别的目的。综合考虑特征维度对识别率和计算量的影响，选取 $\boldsymbol{F}_{3\mathrm{CNN}}=4096$ 作为三波段图像融合特征的维度。

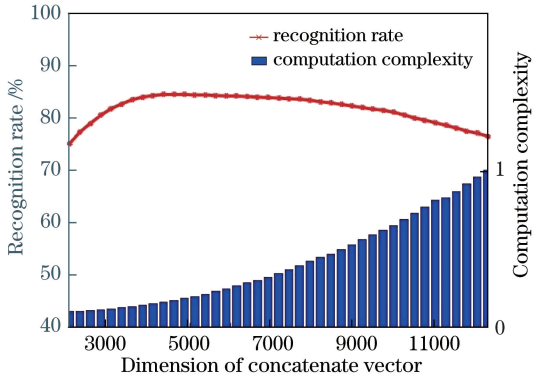

图 5　融合特征维数选择

Fig. 5　Dimension selection of fusion features

利用构建的多波段舰船目标数据集进行实验验证，所有方法采用相同的训练集和测试集，并将本文方法与其他 4 种方法进行对比，分别为：

1) HOG+SVM 识别[1]，HOG 特征描述子选择 64 pixel×128 pixel 的图像块，步长为 8，共 3780 维特征向量；

2) SIFT 特征识别[4]，将图像划分为 64 pixel×64 pixel 的图像块，分别提取 128 维 SIFT 特征；

3) AlexNet 模型[8]，利用 ILSVRC12 中训练得到的参数进行初始化，再用本文数据进行微调；

4) VGG-16 模型[9]，与 AlexNet 类似，通过微调测试网络层数增加对本文目标识别效果的影响；

5）本文方法，单波段识别为不经过特征选择、直接利用 6 层网络提取的特征进行识别的概率。

表 2 为不同方法得到的目标识别率对比，红外波段图像的目标识别率普遍低于可见光图像的识别率，这是因为拍摄的红外图像分辨率相对较低，细节纹理等特征不如可见光图像明显，单独对其进行识别，识别率不高。

表 2　不同方法识别率对比

Table 2　Recognition rate comparison of different methods

Method	Recognition rate /%			
	Visible light	Medium-wave infrared	Long-wave infrared	Fusion recognition
HOG+SVM	63.2	55.0	53.5	
SIFT	67.3	60.2	55.2	
AlexNet	75.6	66.5	67.7	
VGG-16	77.3	68.3	69.2	
Proposed	75.1	67.2	68.1	84.5

在 5 种识别方法中，基于词袋模型（BOW）的方法识别率普遍低于神经网络方法，主要原因在于人工提取的特征是独立存在的，不包含语义信息，在进行匹配时缺少目标之间的关联性。基于 AlexNet 和 VGG-16 的神经网络识别方法在大规模识别任务中取得了较为理想的结果，但是该模型只能对单波段图像分别进行目标识别，不能充分利用多波段图像间的融合特征，在图像清晰度不高的情况下难以取得较高的识别率。本文方法在利用 CNN 特征提取的基础上，利用基于互信息的方法对融合特征按照图像清晰度指标进行重要性排名，消除单一波段中存在的干扰，最终得到了 84.5%的识别率，相比于前 2 类识别方法的识别率分别提高了近 40%和 20%。图 6 所示为 5 种方法识别每一类目标的识别概率，对角线表示识别率，其余位置为误识别率。

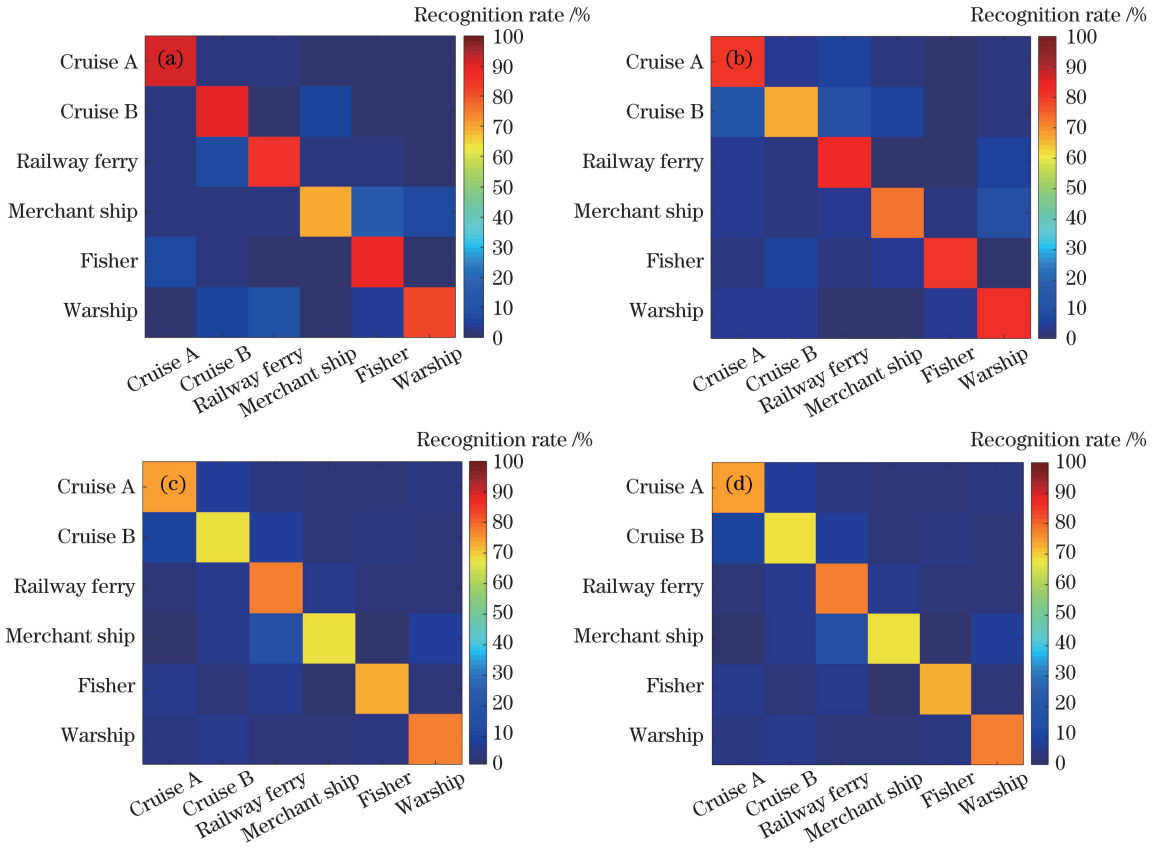

图 6　本文方法识别率矩阵。(a)融合识别；(b)可见光；(c)中波红外；(d)长波红外

Fig. 6　Recognition rate matrices of the proposed method. (a) Proposed fusion recognition; (b) visible light; (c) medium-wave infrared; (d) long-wave infrared

本文方法适用于可见光无法获得精细成像的情况，在测试的数据集中，可见光图像受海上水雾及光照影响，存在大量噪声且细节纹理特征缺乏，红外图像可用信息较少难以进行分类识别。图 7 所示为本文方法得到的部分误识别图像，右侧为判断该类目标的概率值，直方图表示将该目标识别为各类目标的概率。从图像中进行直观的分析可知，主要有两方面原因导致误识别：1）舰船目标在航行过程中因转弯、掉头而产生部分遮挡时，由于训练图片过少或设计的特征向量维数较低，不能充分描述舰船在不同角度下的特征，导致匹配失效；2）某波段图像出现模糊等情况时会影响识别的准确率。

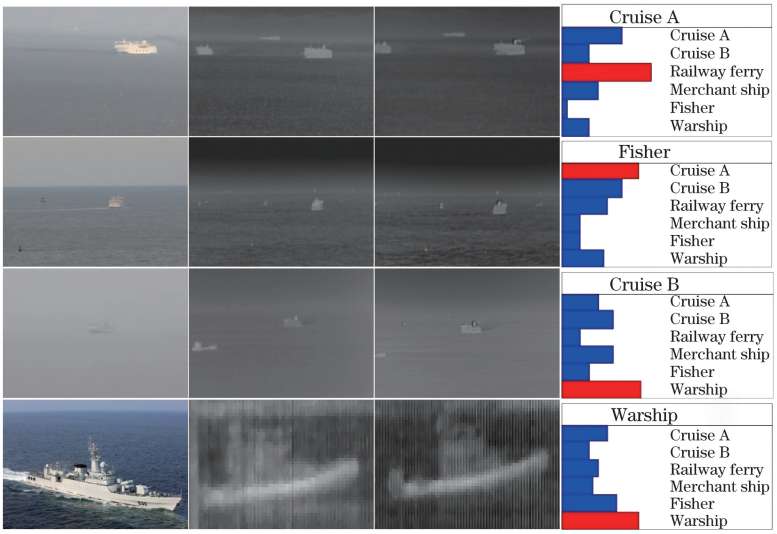

图 7　部分误识别图像

Fig. 7　Partial images of false recognition

4　结　　论

在单波段图像无法获得精细成像的情况下，对多波段舰船目标的融合识别问题进行研究。利用深度 CNN 在目标分类上的优势，设计合理的网络模型对三波段图像进行特征提取并进行有效融合。合理的特征融合方法不但可以提高识别率，同时还可以消除冗余信息，提高计算效率。在本文应用场景下，多波段图像融合后的目标识别率明显高于其他方法和单波段目标识别率。目前构建的数据集还不够完善，样本集中的目标应包含不同光照、角度、尺度、遮挡、复杂背景等多种情况下的训练数据，才能具有更好的泛化能力和实验说服力。在今后的研究中，将进一步丰富数据集并优化算法模型。

参　考　文　献

[1]　Zhang Difei, Zhang Jinsuo, Yao Keming, *et al*. Infrared ship-target recognition based on SVM classification[J]. Infrared and Laser Engineering, 2016, 45(1): 0104004.
张迪飞, 张金锁, 姚克明, 等. 基于 SVM 分类的红外舰船目标识别[J]. 红外与激光工程, 2016, 45(1): 0104004.

[2]　Ramanan D, Zhu X. Face detection, pose estimation, and landmark localization in the wild[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2879-2886.

[3]　Wojek C, Dollár P, Schiele B, *et al*. Pedestrian detection: an evaluation of the state of the art[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(4): 743-761.

[4]　Feineigle P A, Morris D D, Snyder F D. Ship recognition using optical imagery for harbor surveillance[C]. Proceedings of AUVSI Unmanned Systems North America Conference, 2007: 249-263.

[5]　Sánchez J, Perronnin F, Mensink T, *et al*. Image classification with the fisher vector: theory and practice[J]. International Journal of Computer Vision, 2013, 105(3): 222-245.

[6]　Smeelen M A, Schwering P B W, Toet A, *et al*. Semi-hidden target recognition in gated viewer images fused with thermal IR images[J]. Information Fusion, 2014, 18: 131-147.

[7]　Zhou Feiyan, Jin Linpeng, Dong Jun. Review of convolutional neural network[J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251.

周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.

[8] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]. International Conference on Neural Information Processing Systems, 2012: 1097-1105.

[9] Liu Dawei, Han Ling, Han Xiaoyong. High spatial resolution remote sensing image classification based on deep learning[J]. Acta Optica Sinica, 2016, 36(4): 0428001.
刘大伟, 韩玲, 韩晓勇. 基于深度学习的高分辨率遥感影像分类研究[J]. 光学学报, 2016, 36(4): 0428001.

[10] He K, Zhang X, Ren S, *et al*. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[11] Ren S, He K, Girshick R, *et al*. Faster R-CNN: towards real-time object detection with region proposal networks[C]. Advances in Neural Information Processing Systems, 2015: 91-99.

[12] Redmon J, Divvala S, Girshick R, *et al*. You only look once: unified, real-time object detection[J/OL]. (2016-05-09) [2017-01-05] https://arxiv.org/abs/1506.02640.

[13] Kuen J, Wang Z, Wang G. Recurrent attentional networks for saliency detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 3668-3677.

[14] Bousetouane F, Morris B. Fast CNN surveillance pipeline for fine-grained vessel classification and detection in maritime scenarios[C]. IEEE International Conference on Advanced Video and Signal Based Surveillance, 2016: 242-248.

[15] Zhang M M, Choi J, Daniilidis K, *et al*. VAIS: a dataset for recognizing maritime imagery in the visible and infrared spectrums[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015: 10-16.

[16] Nair V, Hinton G E. Rectified linear units improve restricted Boltzmann machines[C]. Proceedings of the 27th International Conference on Machine Learning, 2010: 807-814.

[17] Zhang Y, Wu J, Cai J. Compact representation for image classification: to choose or to compress?[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 907-914.

[18] Ye P, Kumar J, Doermann D. Beyond human opinion scores: blind image quality assessment based on synthetic scores[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 4241-4248.

[19] Xue W, Zhang L, Mou X, *et al*. Gradient magnitude similarity deviation: a highly efficient perceptual image quality index[J]. IEEE Transactions on Image Processing, 2014, 23(2): 684-695.

[20] Sheikh H R, Bovik A C, de Veciana G. An information fidelity criterion for image quality assessment using natural scene statistics[J]. IEEE Transactions on Image Processing, 2005, 14(12): 2117-2128.

[21] Zhang L, Zhang L, Mou X, *et al*. FSIM: a feature similarity index for image quality assessment[J]. IEEE Transactions on Image Processing, 2011, 20(8): 2378-2386.

[22] Wang Z, Bovik A C, Sheikh H R, *et al*. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.