

文章编号 1004-924X(2017)11-2939-08

## 基于多波段深度神经网络的舰船目标识别

刘 峰<sup>1\*</sup>, 沈同圣<sup>2</sup>, 马新星<sup>1</sup>, 张 健<sup>3</sup>

(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;

3. 91206 部队, 山东 青岛 266108)

**摘要:**考虑多波段图像的融合识别可以扩展识别系统的应用范围,本文探索并设计了一种基于卷积神经网络的融合识别方法。该方法以 AlexNet 网络模型为基础,同时对可见光、中波红外和长波红外三波段图像进行特征提取;然后,利用互信息的方法对串联的三波段特征向量进行特征选择,依据重要性排序的方式选定固定长度的特征向量;最后,依据特征提取层级的不同,分别以早期融合、中期融合和后期融合 3 种融合方式来验证算法的有效性。采用自建的三波段舰船图像数据库进行了模型的训练和测试,共包含 6 类目标,5 000 余张图像。实验结果显示,采用的 3 种融合识别方法中,中间层融合的认可准确率最高,达到 84.5%,比早期融合和后期融合分别高 5%和 7%左右。另外,在本文的应用场景下,无论何种融合方式,其融合识别的准确率均明显高于其他单波段识别的准确率。

**关键词:**目标识别;舰船识别;特征融合;卷积神经网络;多波段图像;特征选择

**中图分类号:**TP391.4 **文献标识码:**A **doi:**10.3788/OPE.20172511.2939

## Ship recognition based on multi-band deep neural network

LIU Feng<sup>1\*</sup>, SHEN Tong-sheng<sup>2</sup>, MA Xin-xing<sup>1</sup>, ZHANG Jian<sup>3</sup>

(1. Department of Control Engineering,

Naval Aeronautical and Astronautical University, Yantai 264001, China;

2. Department of National Defense Science and

Technology Information Center, Beijing 100142, China;

3. Unit No. 91206, Qingdao 266108, China)

\* Corresponding author, E-mail: liufeng\_cx@126.com

**Abstract:** The fusion recognition of multi-band images can extend the application range of recognition systems. A fusion method based on convolutional neural networks (CNN) was explored and designed in this paper. Based on the AlexNet network model, it was extracted that the ship target features of three wave band images concurrently in visible light, Middle Wave Infrared (MWIR) and Long Wave Infrared (LWIR) bands. Then, it performed the feature selection for concatenated three-band eigenvectors by using the mutual information method and determines the dimensions of fusion eigenvectors according to sorting the importance of concatenated feature eigenvectors. Finally, three fusion methods named as Early fusion, Middle fusion and Late fusion were used to verify respectively

收稿日期:2017-04-11;修订日期:2017-06-30.

基金项目:国家自然科学基金资助项目(No. 61303192)

the effectiveness of the proposed algorithm according to the features extracted from different levels. An available ship target dataset in three bands containing 6 categories of targets and more than 5 000 images was established for our experimental verification. The results show that the recognition rate from Middle fusion reaches 84.5%. Compared with Early Fusion and Late Fusion, it increases by 8% and 12%. Moreover, the recognition rates of all three fusion methods have been improve significantly as compared to that of the single band recognitions at the same application scene.

**Key words:** target recognition; feature fusion; CNN; multi-spectral images; feature selection; image quality assessmen

## 1 引 言

舰船目标识别在海上搜救、渔船监测、精确制导武器以及海上战争等方面有着广阔的应用前景和研究价值。利用可见光和红外舰船目标图像,结合计算机视觉和机器学习的相关知识,探索研究多波段图像融合技术,可以达到提高识别准确率和扩展适用范围的目的。

海面场景和目标的动态变化(如雾气影响成像效果、目标移动造成轮廓变化)以及夜晚无光条件下如何能够保证全天时、全天候地准确识别舰船目标是亟待解决的重点问题。多源图像融合可以去除各传感器获得信息的冗余性,综合多个传感器的信息以提高互补性,获取对同一目标或场景的全面详尽地表述。融合识别有针对性地提取图像中的有用信息,增加图像理解的可靠性,从而增强目标的识别能力,提高系统的实用性。

可见光和红外图像分别提供了互补的视觉信息,可见光图像包含丰富的细节纹理信息,红外图像不受光照影响,具有清晰的目标轮廓。然而,目前公开的文献中,除了少数的尝试<sup>[1]</sup>,大多数的方法均采用红外或者可见光单一波段进行检测。在诸多计算机视觉问题中,针对不同数据源的信息进行融合是一项极具挑战的任务,基于深度神经网络的多源数据融合问题已经在如图像与音频<sup>[2]</sup>,图像与文本<sup>[3]</sup>,图像与视频<sup>[4]</sup>等诸多领域得到应用。

Ngiam 等<sup>[2]</sup>利用深度置信网络(Deep Belief Network, DBNs)将学习到的隐藏层作为视频和音频的共享特征; Srivastava 等<sup>[3]</sup>将具有相似结构的网络应用于图像的分类和检索中; 王等<sup>[5]</sup>将结构保留函数应用于图像和文本的相似性目标函

数,实现了图像中的短语定位功能; Simonyan 等利用两个并行的卷积网络整合图像的时空信息,并通过光流法同时利用彩色图像和深度图像包含的信息对三维目标进行分类和检测。上述方法的研究策略是针对不同性质的图像,采用并行的神经网络结构提取特征,再对来自不同图像的特征进行融合,如特征层融合<sup>[6]</sup>,置信度融合<sup>[7]</sup>; Karpathy 等<sup>[4]</sup>研究了不同融合机制下的视频分类问题,然而,基于深度神经网络的红外和可见光图像的融合检测,采用何种融合策略可以得到最优的检测效果,这个问题仍待解决。

在舰船目标的融合识别领域, Gundogdu 等<sup>[8]</sup>首次建立了大规模的舰船目标数据集 MARVEL, 该数据库中包含超过两百万张舰船图像,数据均来源于网络的公开图像,共包含 26 大类, 109 种舰船目标。目前,数据库已经过标注并可在互联网上公开使用,可应用于目标的分类、验证、检索和识别。Bousetouane<sup>[9]</sup>针对港口中的多类舰船目标,利用精细化图像对具有高度相似特征的舰船目标进行识别分类,算法采用手工设计特征和 Faster-RCNN<sup>[10]</sup>模型相结合的思路,对目标潜在区域进行定位及特征提取,可同时多种舰船目标进行实时地分类识别。为了实现对舰船目标的全天时监测, Zhang 等<sup>[11]</sup>建立了可见光/长波红外双波段数据集,利用 MatConvNet 深度学习工具箱对不同尺度、不同角度下的舰船目标进行识别,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。

目前,针对多模数据的深度神经网络目标检测算法在行为识别、目标分类等视觉任务中已经有所应用,但是,针对深度神经网络,如何设计有效的融合策略能够得到最好的检测效果,仍然处于尝试和摸索阶段<sup>[12]</sup>。

## 2 研究框架

### 2.1 网络结构

通过以上分析,本文提出了一种基于卷积神经网络的多波段舰船目标融合识别方法,该方法探索研究了 CNN 网络模型针对不同波段的图像如何进行融合识别的方法,此外,限于多波段图像数据集的资源较少,本文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 5 000 余幅图像,6 类目标。本文设计三波段融合识别方法主要包含 3 个步骤,分别为:利用改进的 Alexnet<sup>[13]</sup> 网络实现对三波段图像并行的特征提取;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;利用不同的网络层提取的特征进行融合识别,分别为早期融合、中间层融合和后期融合 3 种方式,并验证识别准确率。算法流程图如图 1 所示。

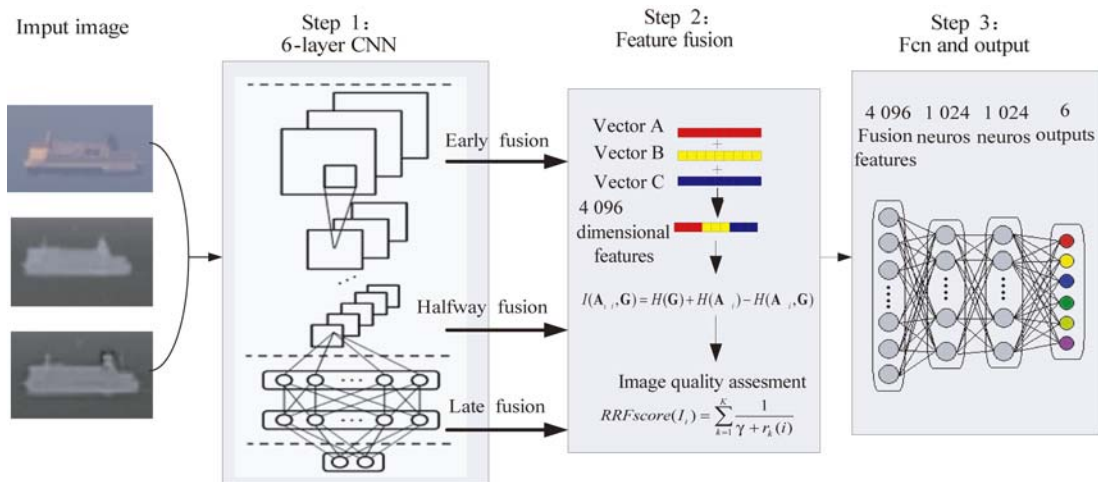


图 1 算法流程图

Fig. 1 Flow chart of our algorithm

本文以 Alexnet 网络模型作为基础,对参数进行一定的调整,网络结构如图 1 所示,在模型初始化方面,利用神经网络具有的迁移学习能力,将 ILSVRC12 中训练好的模型作为网络的初始化参数,再利用本文自建的多波段数据集进行微调。训练时,利用 3 个相同的并行网络,将同一目标的三波段图像分别作为网络的输入进行特征提取。表 1 所示为网络参数,其中,C 表示卷积层,R 为非线性激活函数(RELU)[14],L 表示局部响应标准化,P 为最大池化层,F 为全连接层,D 为 dropout 正则函数。针对不同尺寸的输入图像,需要将其映射为  $227 \times 227$  pixel 的矩形以适应网络结构的输入,文献[8]中验证了不同的映射方法对

识别准确率的影响,本文在这里采用双线性插值法。将输入图像减去像素均值利用 CNN 网络进行训练,通过前向传播逐层提取特征,在第 6 层得到 4 096 维的特征向量,记为 vector **A**。同理,利用相同的 CNN 网络对中波红外和长波红外图像进行特征提取,分别得到 vector **B** 和 vector **C**,3 组特征向量分别包含了不同波段下舰船目标的特征信息。

### 2.2 自适应权重的互信息特征选择

串联融合后的三波段图像特征维度过高,本文采用基于互信息的方法对特征向量进行自适应的特征选择,并按照重要性进行排序,该方法可根据需要任意设定阈值,确定融合向量的维度,而不再需要重新计算。

表 1 神经网络结构及参数

Tab. 1 Architecture and parameters of the proposed deep neural network

层数	网络结构					
	Conv_1	Conv_2	Conv_3	Conv_4	Fc_5	Fc_6
每层结构	C+R+L+P	C+R+P	C+R+P	C+R+P	F+R+D	F+R+D
输入尺寸	227×227×3	27×27×96	13×13×256	13×13×384	6×6×256	4 096
神经元数	96	256	384	256	4 096	4 096
卷积核尺寸	11×11	5×5	3×3	3×3	1×1	1×1
卷积步长	4	2	1	1	—	—
池化尺寸	3×3	3×3	—	3×3	—	—
池化步长	2	2	—	2	—	—

在概率论和信息论中,两个随机变量的互信息是对变量间相互依赖性的度量。两个离散随机变量  $X$  和  $Y$  的互信息可以定义为:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left\{ \frac{p(x,y)}{p(x)p(y)} \right\}, \quad (1)$$

其中:  $p(x,y)$  是  $X$  和  $Y$  的联合概率分布函数,而  $p(x)$  和  $p(y)$  分别是  $X$  和  $Y$  的边缘概率分布函数。

同时,互信息又可以等价于下式:

$$I(X;Y) = H(X) + H(Y) - H(X,Y), \quad (2)$$

其中:  $H(X)$  和  $H(Y)$  是边缘熵,而  $H(X,Y)$  是  $X$  和  $Y$  的联合熵。参考文献[15]中的特征选择方法,计算维度和标签之间的互信息是一种基于监督的方法,用图像标签去估算这维向量的值。以可见光图像特征向量  $\mathbf{A}$  为例,数据集中所有可见光图像的第  $i$  维向量为  $\mathbf{A}_{i,i}$ ,图像标签记为  $\mathbf{G}$ ,他们的互信息为  $I(\mathbf{A}_{i,i}, \mathbf{G})$ 。一般来说,互信息越大,那么这一维向量用于分类则更加有效。互信息的值是对每一维向量重要性的评估,计算公式如下:

$$I(\mathbf{A}_{i,i}, \mathbf{G}) = H(\mathbf{G}) + H(\mathbf{A}_{i,i}) - H(\mathbf{A}_{i,i}, \mathbf{G}), \quad (3)$$

其中:  $H$  是随机变量的熵,  $\mathbf{A}_i$  表示样本数量,图像标签  $\mathbf{G}$  对于不同的维数  $i$  保持不变,即  $\mathbf{G}$  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  $H(\mathbf{A}_{i,i}) - H(\mathbf{A}_{i,i}, \mathbf{G})$  即可。根据互信息的值递减对所有  $N$  维向量进行排序,如果想将  $N$  维向量降维到  $D$  维,则只需取互信息排序前  $D$  名的向量即可。

本文设定串联特征  $F_{3\text{CNN}}$  为 4 096 维,阈值的选取按照图像清晰度评价 (Image Quality Assessment, IQA) 标准,参照文献[16]中的 RRF

方法,利用公式(4),计算多种评价指标的综合得分:

$$\text{RRFscore}(I_i) = \sum_{k=1}^K \frac{1}{\gamma + r_k(i)}, \quad (4)$$

其中评价指标分别选择以下 5 种:

梯度相似性偏差 (Gradient Magnitude Similarity Deviation, GMSD)<sup>[17]</sup>: 计算像素间梯度相似性;

视觉信息保真度指数 (Visual Information Fidelity, VIF)<sup>[18]</sup>: 基于小波变换的多尺度高斯混合模型,计算图像失真程度;

特征相似性指数 (Feature SIMilarity, FSIM)<sup>[19]</sup>: 测量图像的梯度幅值和相位一致性;

颜色特征相似性指数 (FSIMC)<sup>[21]</sup>: 在对可见光图像中,相比特征相似指数,增加颜色信息;

结构相似性指数 (Structural Similarity Index, SSI)<sup>[20]</sup>: 测量图像中物体的结构失真程度。

$r_k(i)$  是图像  $I_i$  在第  $k$  个评价标准中的排名,  $\gamma$  为常数,取  $\gamma=60$ 。公式(4)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心图像的质量指标,并不是对图像清晰度的客观评价。分别对三波段图像计算  $\text{RRFscore}(I_i)$  的值,并对其进行归一化,得到不同波段特征向量的权重值,计算公式如式(5)所示

$$\widehat{F-3\text{CNN}} = \widehat{\text{RRF}(I_{\text{vis}})} \mathbf{vector A} + \widehat{\text{RRF}(I_{\text{MWIR}})} \mathbf{vector B} + \widehat{\text{RRF}(I_{\text{LWIR}})} \mathbf{vector C}, \quad (5)$$

其中:  $\widehat{\quad}$  表示归一化操作,串联后的三波段图像特征  $F-3\text{CNN}$  为 4 096 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地减少该波段图像特征的选择维数,减少

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 步中额外两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 1 024 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。

### 2.3 不同层级特征融合

卷积神经网络的不同阶段对应不同的语义层次和视觉细节特征,低层特征关注于图像的细节纹理信息,高层特征关注目标的整体轮廓。可以认为,在不同阶段的融合会导致不同的检测结果。因此,多波段检测任务可以转化为一个融合问题,即什么样的融合模型可以得到最佳检测效果。

为此,本文对 3 种融合模型进行了研究和实验,不同波段的图像分别输入并行的网络,在不同的阶段,即早期融合,中间融合和后期融合。融合模型的示意图如图 2 所示(彩图见期刊电子版)。从左至右分别为早期融合、中间层融合和后期融合。其中,红色框和黄色框分别代表卷积层和全连接层,蓝色框表示特征串联层,绿色框表示利用互信息进行特征选择,为了简单起见,图中未表示 ReLU 层、池化层和 dropout 层。

早期融合(Early Fusion):在第一个卷积层之后,对可见光和红外图像提取的特征进行串联。通过特征选择可以定量地控制串联特征的维度,按照表 1 所示模型每层的参数结构,定量地选择融合特征维度。CNN 网络中 Conv\_1 层提取的是图像低级视觉特征,如角点、边缘等,因此,早期融合模型是对图像的低级特征进行融合。

中间层融合(Middle Fusion):中间层融合同样是在网络的卷积层进行操作。与早期融合不同

的是,不同波段图像的特征向量在 Conv\_4 层后进行串联,同样使用 2.2 节中的特征选择方法,使得特征向量在固定的维度;相比于 Conv\_1 卷积层,卷积层 Conv\_4 提取的特征中包含了更多的语义信息,保留了更丰富的视觉细节。

后期融合(Late Fusion):对全连接层 Fc\_6 的特征向量进行串联,实现全连接层的融合识别。通常情况下,Fc\_6 层提取的特征可当作目标的一个新的表述,直接用于识别任务中,因此,后期融合方式属于高级特征融合。

## 3 数据集构建与训练

本文建立的数据集中,可见光图像分辨率为  $1\ 024 \times 768$ ,中波传感器工作波段为  $3.7 \sim 4.8 \mu\text{m}$ ,图像分辨率为  $320 \times 256$ ,长波传感器工作波段  $8 \sim 14 \mu\text{m}$ ,图像分辨率为  $640 \times 480$ 。数据集中共包含 6 类目标,5 187 幅图像。分别为游轮 A  $354 \times 3$  幅,游轮 B  $337 \times 3$  幅,铁路轮渡  $208 \times 3$  幅,货船  $236 \times 3$  幅,小型渔船  $291 \times 3$  幅,某型军舰  $303 \times 3$  幅。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。网络的训练采用随机梯度下降(Stochastic Gradient Descent,SGD)方法,批处理尺寸  $m=32$ ,冲量为 0.9,权重延迟为 0.000 5,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, i5-4590 处理器, GTX1080 显卡, 16 G 内存,采用 caffe 深度学习框架进行网络的构造和训练,在迭代 100 000 次的情况下,训练时间大约为 4 h。

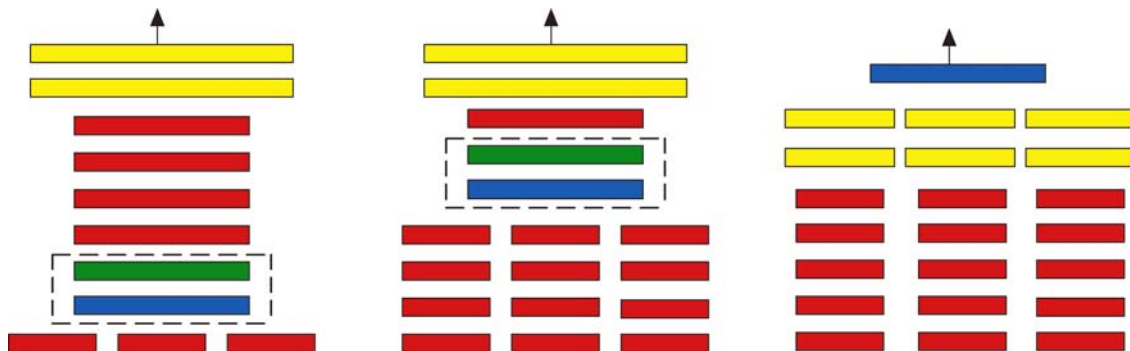


图 2 3 种特征融合模型  
Fig. 2 Three models of feature fusion

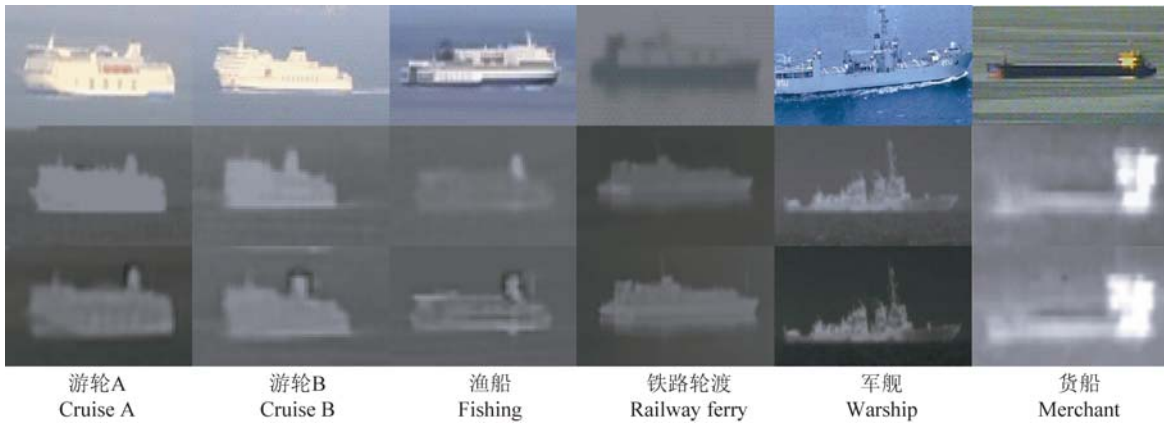


图 3 数据集中图片示例

Fig. 3 Sample pictures in database

### 4 实验验证及对比

图 3 为数据集中 3 种波段不同目标的示例图片。实验验证不同融合方法下的目标识别率,再分别利用 4 种不同的方法进行对比,并分析误识别产生的原因,分别为:

(1)HOG+SVM 识别: HOG 特征描述子选择  $64 \times 128$  的图像块,步长为 8,共 3 780 维特征向量;

(2)SIFT 特征识别:对图像划分为  $64 \times 64$  的图像块,分别提取 128 维的 SIFT 特征;

(3)Alexnet 模型: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:与 Alexnet 类似,通过微调测试网络层数增加对本文的目标识别效果。

图 4 所示为不同方法得到的目标识别率对比(彩图见期刊电子版)。其中,前 4 个柱状图为不同方法分别针对可见光、中波红外和长波红外单波段的识别率,后 3 个柱状图为本文的融合方法,分别为早期融合、中间层融合和后期融合得到的目标识别率。从图中可以发现,红外波段图像的目标识别率普遍低于可见光识别率,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明显,单独对其进行识别准确率不高,且基于“词袋模型”方法的识别准确率普遍低于神经网络的方法,主要原因在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达到较高的识别率。

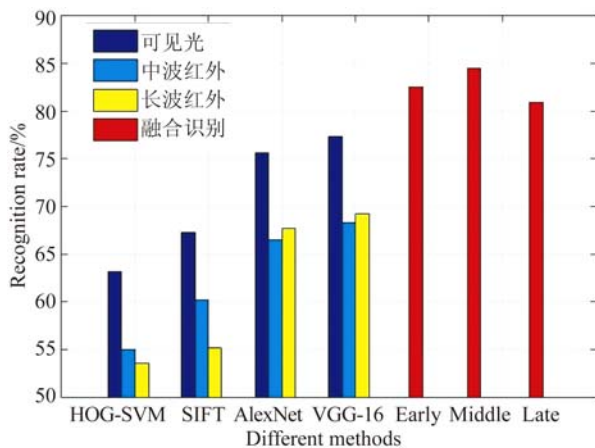


图 4 不同方法识别率对比

Fig. 4 Recogniniton rates of diffeerent methods

本文提出的融合识别模型中,充分利用了三波段图像的互补信息,又通过特征选择的方法去除对分类作用较小的无关向量,实验表明,无论何种融合方法,融合识别率均高于比单波段识别。对比 3 种融合方式,中间层融合得到了最好的识别效果,识别率达到 84.5%,早期融合识别率为 81.5%,后期融合识别率为 79.9%,分别提高了 5%和 7%左右。分析其原因,早期融合提取图像中的低级特征,接近手工设计特征中的边缘、角点等特征,在模型表述的鲁棒性上具有一定的欠缺,全连接层中提取的特征向量具有更准确的概括性,但缺失了泛化能力。图 5 所示为 3 种融合方法识别每一类目标的识别概率,对角线表示正确识别率,其余位置为误识别率。

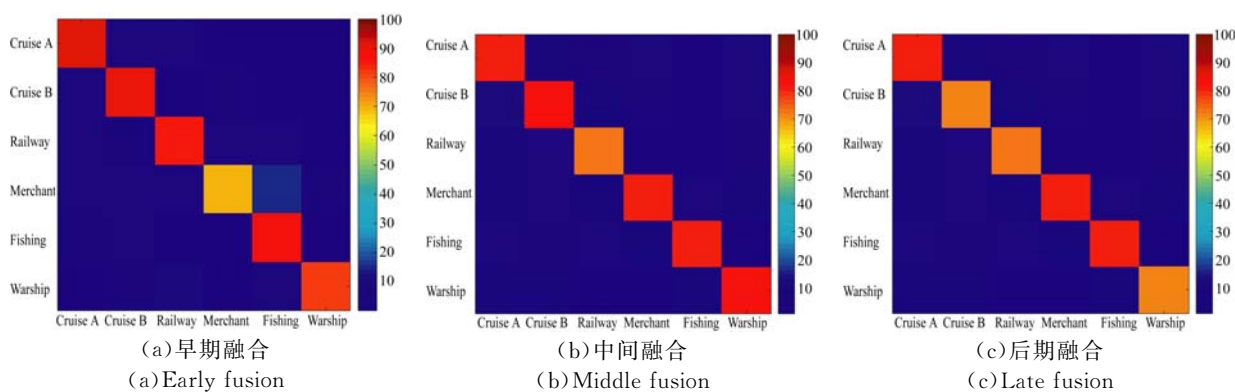


图 5 三中融合方法识别率矩阵

Fig. 5 Matrix of recognition rate for three fusion method

## 5 结 论

单波段图像包含的信息有限,合理利用多波段图像的互补信息,可以有效地提高目标融合识别的准确率。本文利用深度卷积神经网络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有效地融合。在特征的提取与选择上,利用 AlexNet 模型作为基础,利用 6 层的卷积层和 2 层的全连接层提取图像特征,再采用基于互信息的特征选择方法对特征的重要性进

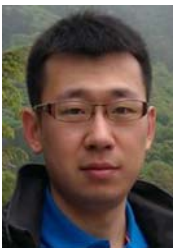
行排序,消除冗余信息,提高计算效率。在融合模型的选择上,尝试了 3 种方式,分别为早期融合、中间层融合和后期融合,不同层次提取的特征是对图像进行不同角度的表述,因此,其融合识别率也存在差别。实验部分,通过实拍图像构建多波段图像数据库并进行目标识别仿真验证,并与目前流行的多种方法进行对比。在本文的应用场景下,无论何种融合方式,融合识别的准确率明显高于其他单波段识别准确率。今后的研究中,将对多种可能的融合方法进行尝试和实验,探索和研究更多类型的融合方法。

## 参考文献:

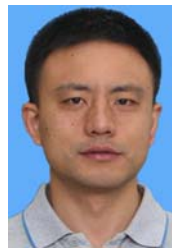
- [1] WAGNER J, FISCHER V, HERMAN M, *et al.*. Multispectral pedestrian detection using deep fusion convolutional neural networks [C]. *24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 2016:509-514.
- [2] NGIAM J, KHOSLA A, KIM M, *et al.*. Multi-modal deep learning [C]. *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011:689-696.
- [3] SRIVASTAVA N, SALAKHUTDINOV R R. Multi-modal learning with deep boltzmann machines [C]. *Proceedings of the 25th International Conference on Neural Information Processing Systems, NIPS*, 2012:2222-2230.
- [4] KARPATY A, TODERICI G, SHETTY S, *et al.*. Large-scale video classification with convolutional neural networks [C]. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014:1725-1732.
- [5] WANG LW, LI Y, LAZEBNIK S. Learning deep structure-preserving image-text embeddings [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016:5005-5013.
- [6] SOCHER R, HUVAL B, BATH B, *et al.*. Convolutional-recursive deep learning for 3d object classification [C]. *Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012:656-664.
- [7] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos [C]. *Proceedings of the 27th International Conference on Neural Information Processing Systems, MIT Press*, 2014:568-576.
- [8] GUNDOGDU E, SOLMAZ B, YÜCESOY V, *et al.*. MARVEL: A large-scale image dataset for maritime vessels [C]. *Asian Conference on Computer Vision, Springer*, 2016:165-180.
- [9] BOUSETOUANE F, MORRIS B. Fast CNN sur-

- veillance pipeline for fine-grained vessel classification and detection in maritime scenarios [C]. 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2016:242-248.
- [10] REN S Q, HE K M, GIRSHICK R, *et al.*. Faster R-CNN: Towards real-time object detection with region proposal networks [C]. *AIPS Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015:91-99.
- [11] ZHANG M M, CHOI J, DANILIAIS K, *et al.*. Vais: A dataset for recognizing maritime imagery in the visible and infrared spectrums [C]. *IEEE Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015:10-16.
- [12] 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.  
ZHOU F Y, JIN L P, DONG J. Review of convolutional neural network [J]. *Chinese Journal of Computers*, 2017, 40(6):1229-1251. (in Chinese)
- [13] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 60(6):84-90.
- [14] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines [C] *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, Omnipress, 2010:807-814.
- [15] ZHANG Y, WU J X, CAI J F. 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.
- [16] YE P, KUMAR J, DOERMANN D. Beyond human opinion scores: blind image quality assessment based on synthetic scores [C] *IEEE Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014:4241-4248.
- [17] XUE W F, ZHANG L, MOU X Q, *et al.*. Gradient magnitude similarity deviation: a highly efficient perceptual image quality index [J]. *IEEE Transactions on Image Processing*, 2014, 23(2):684-695.
- [18] 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.
- [19] 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.
- [20] 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.

#### 作者简介:



刘 峰(1988—),男,黑龙江哈尔滨人,博士研究生,2011年于华中科技大学获得学士学位,2013年于海军航空工程学院获得硕士学位,主要从事机器视觉及目标检测识别等研究。E-mail: liufeng\_cv@126.com



马新星(1982—),男,江苏如皋人,博士研究生,2005年,2013年于海军航空工程学院分别获得学士学位、硕士学位,主要从事红外图像处理、场景仿真、目标检测等研究。E-mail: xinxing\_matom.com