

DOI: 10.19650/j.cnki.cjsi.J1905798

多视角声图中水下小目标分类的深度学习神经网络方法研究

朱可卿^{1,2,3}, 田杰^{1,2}, 黄海宁^{1,2}

(1. 中国科学院声学研究所 北京 100190; 2. 中国科学院先进水下信息技术重点实验室北京 100190;
3. 中国科学院大学 北京 100049)

摘要:针对基于多视角声图的水下小目标分类问题,提出了一种深度学习神经网络多视分类方法。首先,提取声图的阴影区域,计算阴影部分的主轴斜率并匹配出与其相对应的仿真数据集。采用由这些对应仿真数据集训练的卷积神经网络分别对不同视角的待分类声图提取深度学习神经网络特征。将不同视角输出的特征向量组合起来,作为目标的特征向量,利用各个视角匹配的组合所对应的支持向量机对目标的特征向量进行预测。将分类器用于对湖、海试采集的多视角声图分类,平均正确率为93.33%,相比采用卷积神经网络、支持向量机的单视角分类方法,分别有不同程度的提升。

关键词:高分辨率声呐成像;多视角声图;深度学习神经网络;水下小目标分类

中图分类号: TP391 TH766 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Small underwater objects classification in multi-view sonar images using the deep neural network

Zhu Keqing^{1,2,3}, Tian Jie^{1,2}, Huang Haining^{1,2}

(1. Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China; 2. Key Laboratory of Science and Technology on Advanced Underwater Acoustic Signal Processing, Chinese Academy of Sciences, Beijing 100190, China;
3. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: To solve the problem of small underwater objects classification based on multi-view sonar images, a deep neural network classification method with multi-view is proposed. Firstly, the shadow area of underwater objects in sonar images is extracted. The main axis slope of shadow area is calculated, which is used to match sonar images to the corresponding simulated dataset. The convolutional neural network trained by this simulated dataset is applied to extract deep neural network features from multi-view sonar images. The achieved feature vectors from sonar images of different views are combined as a feature vector of underwater object and predicted from support vector machine. The classifier is utilized to classify multi-view sonar images collected from lake and sea trials. The average classification accuracy can reach 93.33%. The performance is improved compared with the single-view classification method using convolutional neural network and support vector machine.

Keywords: high-resolution sonar imaging; multi-view sonar images; deep neural network; underwater object classification

0 引言

随着水下成像技术的不断发展,高分辨率声图的获取为基于声图对水下小目标分类提供了可能性,逐渐成为国内外学者及工程技术人员的研究热点之一^[1-4]。通常成像声呐对目标扫测时,围绕目标完成多个航次,从不

同的视角获取目标的声图。由于这些声图来自于同一目标,它们之间存在着某种联系。又因为由不同视角照射目标而获得,各自又包含目标不同视角所特有的信息。

虽然许多基于单视角声图对水下小目标分类的研究取得了一些成果^[5-7],但是依然存在较多的不确定性。如果对小目标多个视角的声图中的信息进行有效结合,则可提高目标分类的正确率。Zerr等^[8]提出了一种基于多

视角声图的目标 3D 重建方法。他们将声图分割为背景区、高亮区和阴影区,通过阴影区重建高程图,通过高亮区重建反射贴图,最终合成目标的 3D 图像。但是,这种重建算法需要对目标沿圆周航迹进行扫测,在拖曳方式下,由于转向时平台及声呐的惯性作用,很难以这种航迹对目标进行扫测。Groen 等^[9]将多个视角的目标声图经过配准后融合成一幅图像,对融合后的图像去除阴影部分,使用所保留的高亮部分对目标进行分类,有效降低了虚警概率。这种基于像素级的融合体现了目标在不同视角下的共有信息,但融合过程损失了目标不同视角特有的信息,也可能会影响分类效果。Fawcett 等^[10]对多视角声图采用特征级融合和决策级融合两种方法进行处理。在特征级融合中,来自一个视角声图的特征与其他视角提取特征组成一个向量作为目标的多视角特征,对这个特征采用核回归分类器进行分类。在决策级融合中,将不同视角声图特征分别输入代表各自视角的分类器中,对输出结果采用 D-S 理论(dempster-shafer theory, DST)进行决策融合,取得了较好的分类结果,证明了通过特征级或决策级对多视角声图特征的融合可以有效提高分类的正确率。

近年来,深度神经网络被广泛应用于自然语言处理、图像识别^[11-13]等领域,取得了一系列突破性的成果,主要的模型有深度置信网络(deep belief network, DBN)、卷积神经网络(convolutional neural network, CNN)等。目前,已有研究将其应用于对声图的分类,如 William^[14]采用卷积神经网络基于声图对水下小目标进行了检测,取得了较好的效果。Wang 等^[15]提出一种自适应权值卷积神经网络(adaptive weights convolutional neural network, AW-CNN),解决了卷积神经网络在训练中,由于初始化随机性导致分类结果不理想的问题,并将改进后的网络应用于对声纳图像的分类,具有一定的有效性^[15]。

为了进一步提高水下小目标声图分类的正确率,结合深度网络在图像识别方面的优势,本文提出一种基于深度卷积特征的多视角声图分类方法。首先,采用 Grab-cut 算法^[16]提取声图的阴影特征,接下来计算阴影部分的主轴斜率并匹配出与其相对应的仿真数据集。采用由这些对应仿真数据集训练的卷积神经网络^[17-19]分别对不同视角的待分类声图将进行特征提取。将不同视角输出的特征向量组合起来,作为目标的特征向量。最后,利用各个视角匹配的组合所对应的支持向量机(support vector machine, SVM)^[20-21]对目标的特征向量进行预测。

本文方法利用深度神经网络对图像细节特征捕捉能力强的特点,引入卷积神经网络,可有效提高目标声图的分类效果。另外,文献[10]只考虑了斜视角不同的多视情况,在不同掠射角下,所获取的目标声图也会不同,也可能对分类效果产生影响。声图的获取需要耗费大量的

人力、物力,通过试验所获得的数据难以组成完备的训练集,训练出的分类器也很难具备对每个不同视角的声图的分类能力。针对这个问题,本文建立了仿真数据集,包括空间中各个不同斜视角、掠射角的水下小目标图像,具有完备性,深度神经网络由空间中对应视角区间的仿真数据集训练后,具有对视角区间内各个角度小目标声图的特征辨识能力。同时,由于深度神经网络输出不同视角组合向量分别训练了对应的支持向量机,本文方法适用于任意角度组合的多视角声图分类。

1 系统结构

多视角声图分类系统分为视角匹配、特征提取和分类 3 个步骤,如图 1 所示。

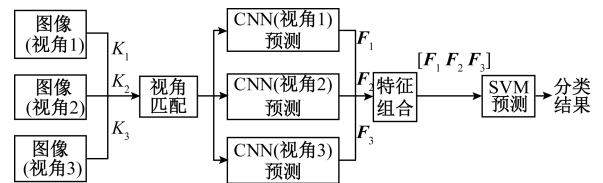


图1 分类系统结构

Fig.1 Structure diagram of the classification system

1) 视角匹配

视角匹配主要通过估计阴影部分主轴的斜率,完成待分类声图与仿真集的匹配。包括构建仿真数据集,阴影提取,计算声图中目标阴影部分主轴的斜率,对仿真集中图像基于斜视角进行分组并计算斜率平均值,将待分类声图与仿真集斜率进行视角匹配确定其对应的卷积神经网络。

2) 特征提取

特征提取主要完成基于目标阴影多视角图像得到目标的特征向量,又包括特征提取和特征组合。特征提取采用卷积神经网络来完成,视角匹配后,将目标阴影输入相对应视角区间的卷积神经网络,经预测后得到深度卷积特征。将 3 个视角的深度卷积特征进行组合得到目标的特征向量。

3) 分类

分类主要完成基于目标阴影多视角图像得到分类结果,采用支持向量机来进行预测。视角匹配后,可以得到 3 个视角区间的组合,每个组合对应唯一的支持向量机,将目标的特征向量输入到相对应支持向量机中,得到目标的分类结果。

2 视角匹配

2.1 仿真集构建与匹配

首先对水下小目标建立仿真数据集,仿真场景如

图 2 所示,设掠射角为 θ 。由几何关系可知:

$$\theta = \arctan \frac{h}{l}$$

(1)

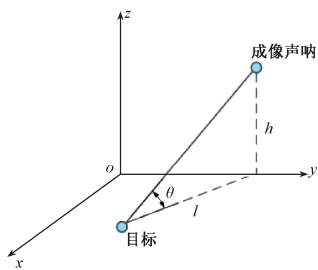


图 2 仿真场景

Fig.2 Schematic of the simulation scenario

为构造不同的掠射角,仿真设定声呐高度的取值为 30、35、40、45 m,水平斜距的取值为 120、130、140、150、160、170、180 m,两者不同组合对应的掠射角如表 1 所示。

接着确定一组声呐高度、水平斜距和掠射角的映射关系后,再针对不同斜视角 λ 生成相对应的仿真图像, $\lambda \in [0,180^\circ)$ 。其中,斜视角为入射声波与目标轴线的夹角,如图 3 所示, λ 在 $[0,180^\circ)$ 中的取值间隔为 5° ,得到 36 幅斜视角不同的仿真图像。

2.2 声图阴影主轴斜率估计

首先,采用 Grab-cut 算法对原始声图进行阴影提取,图 4 所示为圆柱体目标的合成孔径声呐 (synthetic aperture sonar, SAS)^[22-24] 图像及所提取的阴影,图像由 Kongsberg 公司的 HISAS 采集得到。

表 1 仿真图像掠射角映射表

Table 1 Simulated image grazing angle mapping table

序号	声呐高度 h/m	水平斜距 l/m	掠射角 $\theta/(\text{^\circ})$	序号	声呐高度 h/m	水平斜距 l/m	掠射角 $\theta/(\text{^\circ})$	序号	声呐高度 h/m	水平斜距 l/m	掠射角 $\theta/(\text{^\circ})$	序号	声呐高度 h/m	水平斜距 l/m	掠射角 $\theta/(\text{^\circ})$
1		120	14.0	1		120	16.3	1		120	18.4	1		120	20.6
2		130	13.0	2		130	15.1	2		130	17.1	2		130	19.1
3		140	12.1	3		140	14.0	3		140	15.9	3		140	17.8
4	30	150	11.3	4	35	150	13.1	4	40	150	14.9	4	45	150	16.7
5		160	10.6	5		160	12.3	5		160	14.0	5		160	15.7
6		170	10.0	6		170	11.6	6		170	13.2	6		170	14.8
7		180	9.5	7		180	11.0	7		180	12.5	7		180	14.0

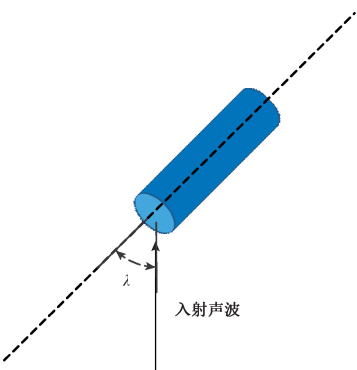
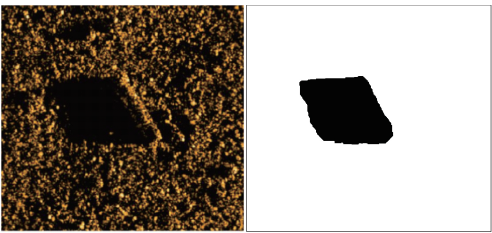


图 3 斜视角示意图

Fig.3 Schematic of squint angle

由于在仿真图像集中,声呐采集图像的斜视角为已知,可通过对声图和仿真图像阴影主轴的斜率来匹配斜视角,因此首先需要找出声图阴影的主轴。设目标的阴影部分由 k 个点组成,第 i 个点的坐标为 (x_i, y_i) ,将这些点看作是一个点集 $P = \{p_i, i = 1, 2, \dots, k\}$,其中 $p_i =$



(a) SAS 图像
(a) SAS image (b) 目标阴影
(b) target shadow

图 4 圆柱体目标

Fig.4 Cylindrical object

$(x_i, y_i), i = 1, 2, \dots, k$ 。点集的中心可以表示为 $(\bar{x}, \bar{y})$,其中:

$$\bar{x} = \sum_{i=1}^k x_i, \bar{y} = \sum_{i=1}^k y_i$$

(2)

首先,为便于计算,将点集的中心移到原点,组成新的点集 $P' = \{p'_i, i = 1, 2, \dots, k\}$ 。其中, $p'_i = (x_i - \bar{x}, y_i - \bar{y}), i = 1, 2, \dots, k$ 。

设主轴垂线方向的单位向量为 $\boldsymbol{n}$,对于每个点集中

的像素点 p'_i , 从 p'_i 到主轴的距离可以表示为 p'_i 在主轴垂线方向上的投影, 即:

$$d_i(n) = |n^T p'_i| \quad (3)$$

主轴即为对点集 P' 的最优拟合直线, 该直线可满足点集中每个点到主轴距离之和最小, 即:

$$\varepsilon^2 = \sum_{i=1}^k d_i^2(n) = (n^T p'_i)^2 =$$

$$\sum_{i=1}^k n^T p'_i p'^T_i n = n^T \left(\sum_{i=1}^k p'_i p'^T_i \right) n \quad (4)$$

同时, 上式满足:

$$n^T n = 1 \quad (5)$$

采用拉格朗日乘子来求解约束条件下的最小化问题, 定义:

$$Lagrange = n^T S n - \lambda (n^T n - 1) \quad (6)$$

其中, 为点集坐标的协方差矩阵, 即:

$$S = \sum_{i=1}^k p'_i p'^T_i \quad (7)$$

对 p'_i 求导, 令导数为 0:

$$\frac{\partial (n^T S n - \lambda (n^T n - 1))}{\partial p'_i} = 0 \quad (8)$$

可简化为:

$$S n = \lambda n \quad (9)$$

由式(9)可知, 主轴经过点集的中心点, 并会落在点集的协方差矩阵的主特征向量的方向上。 S 经特征分解, 得到 n 个特征值为 $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, 其中模最大的为主特征值, 记为 λ_p , λ_p 对应的特征向量为 n_p , 即主特征向量。

$$n_p = (x_p, y_p) \quad (10)$$

因此, 主轴的斜率为:

$$k = \frac{y_p}{x_p} \quad (11)$$

由于主轴经过中心, 可表示为:

$$y - \bar{y} = \frac{y_p}{x_p} (x - \bar{x}) \quad (12)$$

2.3 视角区间划分与匹配

以 30° 为间隔, 对 2.1 节中的 36 幅不同斜视角仿真图像再划分为 6 个视角区间, 具体如表 2 所示。

表 2 视角区间划分表
Table 2 View interval partition

视角区间序号	斜视角范围/(°)	视角区间序号	斜视角范围/(°)
1	[0, 25]	4	[90, 115]
2	[30, 55]	5	[120, 145]
3	[60, 85]	6	[150, 175]

按照 2.2 节中的方法计算出仿真数据集中所有图像中阴影的斜率, 记为: $K = \{k_1, k_2, \dots, k_n\}$, 其中 n 为仿真数据集图像数量。按照表 2 对所有仿真数据集划定视角

区间, 计算每个视角区间斜率的平均值, 分别记为 $\bar{k}_1, \bar{k}_2, \dots, \bar{k}_6$ 。

计算待分类图像中阴影主轴斜率, 记为 k_i , 则其所对应的视角区间为:

$$g_i = \operatorname{argmin}_i |k_i - k_i|, i = 1, 2, \dots, 6 \quad (13)$$

3 深度卷积特征提取

卷积神经网络相比其他神经网络, 有以下优点: 具有平移、旋转和缩放不变性, 适应性强; 网络在识别的过程中可自动提取图像的抽象特征; 通过权值共享简化网络结构, 增加了参数传递的效率等。本文采用卷积神经网络来提取小目标声图的特征。

卷积神经网络特征提取的过程主要分为卷积和池化两个步骤。

3.1 卷积运算

卷积主要采用卷积层中所包含的多个卷积核, 对输入图像进行卷积运算。一个卷积核包含一个权重系数和一个偏移量, 与前一层中位置接近的区域的多个节点相连接。该区域称为“感受野”, 其大小取决于卷积核大小。运算中, 卷积核按照一定规律扫过输入图像, 在感受野内与输入图像相同尺寸的部分进行矩阵乘法, 并叠加偏移量:

$$\begin{aligned} F_{\text{output}}(i, j) &= [F_{\text{input}} \otimes w_{\text{output}}](i, j) + b = \\ &= \sum_{k=1}^K \sum_{m=1}^L \sum_{n=1}^L [F_{\text{input}}^k(s_i + m, s_j + n) w_{\text{output}}^k(m, n)] + b, \\ (i, j) &\in \{0, 1, \dots, H_{\text{output}}\} \\ H_{\text{output}} &= \frac{H_{\text{input}} + 2p - L}{s} + 1 \end{aligned} \quad (14)$$

式中: b 为偏移量; F_{in} 和 F_{output} 分别为卷积层的输入和输出, 称为“特征图”; H_{input} 和 H_{output} 为特征图 F_{input} 和 F_{output} 的尺寸; k 为特征图的通道数; L, s, p 分别为卷积核尺寸、步长和填充层数。卷积运算主要通过稀疏连接、参数共享来提升特征提取功能。

稀疏连接受神经科学中每个细胞只对一个视觉区域内极小的一部分敏感, 而对其他部分则可以视而不见的现象启发, 网络中输出节点只与一部分输入节点相连接, 如图 5 所示。其优势在于, 相比于传统的神经网络使用矩阵乘法来建立输入与输出的连接关系, 如图 6 所示, 每一个输出单元与每一个输入单元都产生交互, 卷积网络的稀疏交互可以在核的大小远小于输入的大小的情况下提高网络的运算效率。

参数共享是指在一个模型的多个函数中使用相同的参数。在传统神经网络中, 当计算一层的输出时, 权重矩阵的每一个元素只使用一次, 当它乘以输入的一个元素后就再也不会用到了。而在卷积神经网络中, 卷积核的

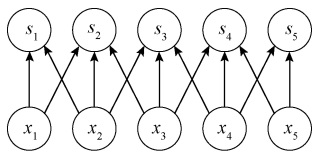


图 5 卷积神经网络中的稀疏连接

Fig.5 Sparse connectivity in convolutional neural networks

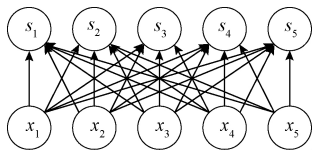


图 6 传统神经网络的连接方式

Fig.6 Connectivity in traditional neural networks

每一个元素都作用在输入的每一个位置上,即网络含有一个绑定的权重,用于一个输入的权重也会被绑定在其他权重上。

如图 7 所示,黑色实线箭头表示卷积模型中对 3 元素核中间元素的使用,因为参数共享,这个参数将被用于所有的输入位置。图 8 所示为传统神经网络参数的使用。黑色实线箭头表示在权重矩阵的中间元素参数,由于参数共享,它只使用的一次。参数共享虽然没有缩短前向传播的计算时间,但它显著地降低了模型存储的参数数量,节省了算法的内存空间。

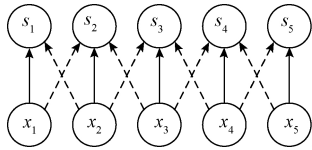


图 7 卷积神经网络中的参数共享

Fig.7 Parameter sharing in convolutional neural networks

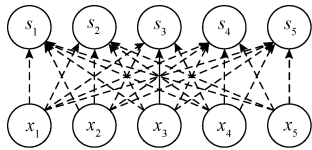


图 8 传统神经网络的参数使用

Fig.8 Parameter in traditional neural networks

3.2 池化运算

池化主要完成对卷积层所提取信息的选择和过滤。池化运算主要依靠池化层中的函数将特征图转化为各像素点相邻区域的统计量。同卷积层,池化层也是通过池化窗口对特征图进行扫描,得到新的特征图。由池化函数设计方法的不同,池化运算又包括 L_p 池化、随机池化和谱池化。其中, L_p 是卷积网络设计中长期被使用的方

法,它是一种受视觉皮层内阶层结构启发而构建的池化模型,形式为:

$$G_{\text{output}}(i,j) = \left[\sum_{m=1}^L \sum_{n=1}^L G_{\text{in}}(si+m,sj+n)^p \right]^{\frac{1}{p}} \quad (15)$$

式中: s 为步长; p 为预设定参数,当 $p = 1$ 时,在池化区域内取均值,称为“均值池化”;当 $p \rightarrow \infty$ 时,在池化区域内取极大值,称为“极大池化”。在本文中,采用极大池化,以损失特征图部分信息为代价来提取图像的背景及纹理信息。

4 分类器设计

采用卷积神经网络对图像进行特征提取,考虑到数据集较小,便于更好地拟合数据集和提高训练效率,搭建网络结构如图 9 所示。原始声图经阴影提取后,变为二值图像(1 通道),图像的尺寸设置为 64×64 ,考虑到输入图像的特点,网络卷积层和池化层参数设置如表 3 和 4 所示,全连接层共有 16 个节点,输出特征为 16 维的向量。

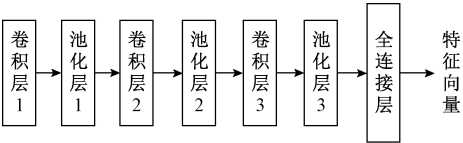


图 9 卷积神经网络

Fig.9 Schematic of convolutional neural network

表 3 卷积层结构参数

Table 3 Structure parameter of convolution layer

卷积层	卷积核数量	卷积核尺寸	步长	输出向量形式
1	24	5×5	1×1	60×60×24
2	32	3×3	1×1	28×28×32
3	48	3×3	1×1	12×12×48

表 4 池化层结构参数

Table 4 Structure parameter of pooling layer

池化层	池化窗口尺寸	步长	输出向量形式
1	2×2	1×1	30×30×24
2	2×2	1×1	14×14×32
3	2×2	1×1	6×6×48

分类器训练过程如图 10 所示,具体过程可描述如下:

首先,按照表 2 对所有仿真数据集划定方位区间,每个区间的仿真图像作为训练集,用于训练一个代表这个区间视角特性的卷积神经网络。在网络一次的训练中,

先对所有网络参数进行高斯分布 $G(\mu, \sigma)$ 的随机初始化,其中均值 $\mu = 0$, 方差 $\sigma = 1$ 。对于所有卷积层,采用修正线性单元(rectified linear unit, ReLU)^[25] 作为激活函数。从数据集中随机抽取 64 幅图像作为一个批次,输入图 9 所示的网络进行计算,利用 softmax 层将网络最后一层输出的 3 维向量转化为该样本图像在 3 种类型小目标上的概率分布,使用交叉熵^[26] 作为损失函数,计算预测值与真实标签之间的距离。最后,将学习率设置为 0.001,采用 Adam 优化器^[27] 对整个网络的参数进行优化。

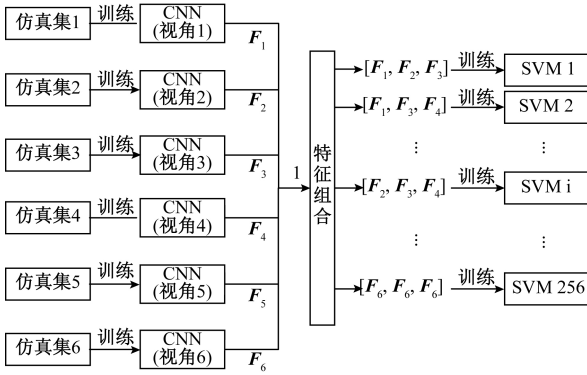


图 10 分类器训练过程

Fig.10 The training process of classifier

训练完成后,将各视角区间训练集输入训练好的卷积神经网络中,记各视角区间的深度卷积特征向量分别为 $F_1, F_2, \dots, F_6$, 将 $F_1, F_2, \dots, F_6$ 每 3 个视角区间的深度卷积特征向量任意组合, 得到代表不同情况下所获取多视角图像中目标的特征向量。

$$[F_a, F_b, F_c], a, b, c \in \{1, 2, \dots, 6\} \quad (16)$$

这样,可得到 256 组不同视角组合的目标特征向量训练集,分别用于训练各自对应的支持向量机,以完成整个分类器的训练。

5 实验

5.1 数据集

为验证本文方法的有效性,采用湖、海试所采集的真实数据集进行实验。包括声图共 153 幅,包括圆柱体、球体和圆台体 3 种类型的水下小目标各 79、30、44 幅,图像尺寸均设为 64×64 ,其中,圆柱体目标实物与 SAS 图像如图 11 所示。同时,为辅助分类器的训练和视角匹配,由 2.1 节所描述的方法构建仿真数据集,包括圆柱体、球体、圆台体 3 种类型目标图像各 1 008 幅,如图 12 所示。

5.2 斜率与斜视角、掠射角的关系

在 2.3 节中,按照斜视角范围划分了视角区间,以

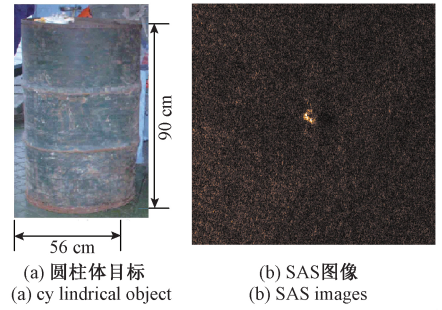


图 11 圆柱体目标及其 SAS 图像

Fig.11 Cylindrical object and its SAS image

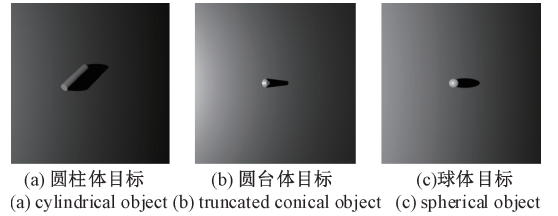


图 12 水下小目标仿真图像

Fig.12 Simulated image of underwater object

此为依据,对仿真图像进行分组,计算各组中阴影主轴的斜率平均值作为参考,与待分类图像阴影部分的主轴斜率进行匹配,以确定待分类图像所属的视角区间及相对应的卷积神经网络。在视角区间划分时,主要采用斜视角作为依据。事实上,目标的视角不仅受斜视角影响,也可能会受到掠射角的影响。分类器设计中仅考虑了斜视角对目标视角的影响,而对于掠射角,是否会对目标视角产生影响,还有待考察。因此,设置声呐高度 h 为 30 m,水平斜距 l 分别取值为 120、130、140、150、160、170、180 m,对应的掠射角分别为 14° 、 13° 、 12.1° 、 11.3° 、 10.6° 、 10° 、 9.5° ,斜视角 λ , $\lambda \in [0, 180^\circ)$, 取值间隔为 5° ,对圆柱体目标声图进行仿真。对圆柱体仿真图像提取阴影,计算阴影部分主轴斜率,绘制当斜视角、掠射角取不同值时,仿真图像阴影部分主轴斜率曲线,如图 13 所示。

图 13 中,尽管掠射角不同,各个目标阴影主轴斜率曲线几乎重合在一起,而随着斜视角增大,目标阴影斜率也单调递增。因此,声波的掠射角对图像中目标阴影主轴斜率以及图像中目标视角的影响较小。在视角区间划分时,可直接以斜视角作为依据。

5.3 实验结果

将多视角分类器应用于对水下小目标的分类,支持向量机采用 L2 正则化的线性内核,误差项的惩罚参数设置为 1.0。卷积神经网络采用如图 9 所示的结构,学习率设为 0.001,丢弃率设置为 0.8,批尺寸设为 64,随机抽取

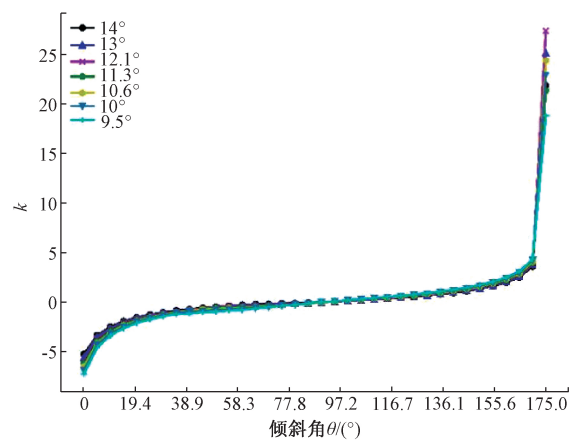


图 13 不同掠射角、斜视角的斜率
Fig.13 Slopes of different grazing angles and squint angles

50 个批次真实数据对网络进行训练。对本文提出的多视角分类方法进行 10 次实验,同时,采用如图 9 所示结构卷积神经网络进行单视角分类、采用如上所述的支持向量机进行单视角分类,进行对比,计算 3 种分类方法的平均正确率,如表 5 所示。

表 5 多视角分类算法与单视角分类算法 (CNN、SVM) 的平均正确率

Table 5 Average accuracy of multi-view classification algorithm and single-view classification algorithm (CNN, SVM)	
分类方法	正确率/%
单视角 (SVM)	79.23
单视角 (CNN)	80
多视角 (本文方法)	93.33

其中本文提出的多视角分类方法平均正确率为 93.33%, 分别比采用卷积神经网络、支持向量机的分类正确率提高了 14.1% 和 13.33%。

6 结 论

本文提出一种基于深度卷积特征的多视角声图分类方法,主要从两个方面对现有水下小目标声图分类方法进行了提升。首先,采用仿真数据集训练深度神经网络,由于仿真数据集包含空间中各个不同斜视角、掠射角的水下小目标图像,具有完备性。因此,训练得到的分类器具有对空间各个角度小目标声图的特征辨识能力。此外,考虑了所有各个视角区间特征组合的可能性,分别使用每个组合的仿真图像深度卷积特征训练了对应的支持

向量机,因此,本文方法可适用于任意角度组合的多视角声图分类。然而,本文方法在分类器的训练上需要花费较长时间,分类器的结构也较为复杂。针对多视角分类器的简化设计,还需进一步的研究。

参考文献

[1] MARCHAND B, SAITO N, XIAO H. Classification of objects in synthetic aperture sonar images [C]. 2007 IEEE/SP 14th Workshop on Statistical Signal Processing, Madison, WI, USA, 2007: 433-437.

[2] BELLO M G, DOBECK G J. The use of texture measures in improving mine classification performance [C]. Proceedings of IEEE OCEANS, 2003: 1103-1109.

[3] MCKAY J, GERG I, MONGA V, et al. What's mine is yours: Pretrained CNNs for limited training sonar ATR[C]. Proceedings of IEEE OCEANS, Anchorage, USA, 2017: 1-7.

[4] PEZESHKI A, AZIMI-SADJADI M R, SCHARF L L et al. Underwater target classification using canonical correlations[C]. Proceedings of IEEE OCEANS, 2003: 1906-1911.

[5] REED S, PETILLOT Y, BELL J. Automated approach to classification of mine-like objects in sidescan sonar using highlight and shadow information[J]. IEE Proceedings-Radar, Sonar and Navigation, 2004, 151(1): 48-56.

[6] DOBECK G J, HYLAND J C. Automated detection and classification of sea mines in sonar imagery [C]. Detection and Remediation Technologies for Mines and Minelike Targets II, International Society for Optics and Photonics, 1997: 90-110.

[7] AZIMI-SADJADI M R, KLAUSNER N, KOPACZ J. Detection of underwater targets using a subspace-based method with learning [J]. IEEE Journal of Oceanic Engineering, 2017, 42(4): 869-879.

[8] ZERR B, STAGE B. Three-dimensional reconstruction of underwater objects from a sequence of sonar images[C]. Proceedings of 3rd IEEE International Conference on Image Processing, 1996: 927-930.

[9] GROEN J, COIRAS E, WILLIAMS D P. False-alarm reduction in mine classification using multiple looks from a synthetic aperture sonar [C]. Proceedings of IEEE OCEANS, Sydney, Australia, 2010: 1-8.

[10] FAWCETT J, MYERS V, HOPKIN D, et al. Multiaspect

- classification of sidescan sonar images: Four different approaches to fusing single-aspect information[J]. IEEE Journal of Oceanic Engineering, 2010, 35(4): 863-876.
- [11] 王震,黄如意,李霖蒲,等.一种用于故障分类与预测的多任务特征共享神经网络[J].仪器仪表学报,2019, 40(7):169-177.
- WANG ZH, HUANG R Y, LI J P, et al. Multi-task feature sharing neural network used for fault diagnosis and prognosis[J]. Chinese Journal of Scientific Instrument, 2019, 40(7): 169-177.
- [12] 王永利,曹江涛,姬晓飞.基于卷积神经网络的PCB缺陷检测与识别算法[J].电子测量与仪器学报,2019, 33(8):78-84.
- WANG Y L, CAO J T, JI X F, PCB defect detection and recognition algorithm based on convolutional neural network[J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(8): 78-84.
- [13] 刘笑楠,白雨辰,尹思璐,等.基于类卷积神经网络的可见光虹膜识别方法[J].仪器仪表学报, 2017, 38(11): 2651-2658.
- LIU X N, BAI Y CH, YIN S L, et al. Iris recognition of visible light based on analogous convolutional neural network[J]. Chinese Journal of Scientific Instrument, 2017, 38(11): 2651-2658.
- [14] WILLIAM D P. Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks[C]. Proceedings of 23rd International Conference on Pattern Recognition (ICPR), 2016: 2497-2502.
- [15] WANG X M, JIAO J, YIN J, et al. Underwater sonar image classification using adaptive weights convolutional neural network[J]. Applied Acoustics, 2019, 146: 145-154.
- [16] ROTHER C, KOLMOGOROV V, BLAKE A. Grabcut: Interactive foreground extraction using iterated graph cuts[C]. ACM transactions on graphics (TOG), 2004, 23(3): 309-314.
- [17] CUN Y L, BOSER B, DENKER J S, et al. Handwritten digit recognition with a back-propagation network[C]. Neural Information Processing Systems, 1990: 396-404.
- [18] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [19] BENGIO Y. Learning deep architectures for AI[J]. Foundations and trends in Machine Learning, 2009, 2(1): 1-127.
- [20] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [21] BURGESS C J C. A tutorial on support vector machines for pattern recognition[J]. Data mining and knowledge discovery, 1998, 2(2): 121-167.
- [22] GOUGH P T, HAWKINS D W. A short history of synthetic aperture sonar[C]. Symposium Proceedings of IEEE International Geoscience and Remote Sensing (IGARSS), 1998: 618-620.
- [23] GOUGH P T, HAYES M P. Synthetic aperture sonar: The past the present and the future[C]. Proc. Inst. Acoust, 2004: 81-87.
- [24] HAYES M P, GOUGH P T. Synthetic aperture sonar: A review of current status[J]. IEEE Journal of Oceanic Engineering, 2009, 34(3): 207-224.
- [25] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[C]. Proceedings of the fourteenth international conference on artificial intelligence and statistics, 2011: 315-323.
- [26] DEBOER P T, KROESE D P, MANNOR S, et al. A tutorial on the cross-entropy method[J]. Annals of operations research, 2005, 134(1): 19-67.
- [27] KINGMA D P, BA J. Adam: A method for stochastic optimization[C]. 3rd International Conference for Learning Representations, San Diego, 2014, arXiv: 1412.6980; 2014arXiv1412.6980K.

作者简介



朱可卿,2015年于湖南师范大学获得学士学位,现为中国科学院声学研究所博士研究生,主要研究方向为水声信号处理。

E-mail: zhukeying@mail.ioa.ac.cn

Zhu Keqing received her B.Sc. degree from Hunan Normal University in 2015. She is currently a Ph. D. candidate in the Institute of Acoustics, Chinese Academy of Sciences. Her main research interests include underwater acoustic signal processing.



田杰, 1995 年于西北工业大学获得学士学位, 2002 年于北京理工大学获得博士学位, 现为中国科学院声学研究所研究员、硕士生导师, 主要研究方向为水声成像技术、水声图像处理及目标识别、水声信号处理等。

E-mail: tianjie@mail.ioa.ac.cn

Tian Jie received his B. Sc. degree from Northwest Polytechnical University in 1995, and received his Ph. D. degree from Beijing Institute of Technology in 2002. He is currently a professor and a M. Sc. advisor at Institute of Acoustic, Chinese Academy of Science. His main research interests include underwater imaging technology, underwater image processing and underwater signal processing.



黄海宁 (通信作者), 分别在 1991 年、1996 年和 1999 年于西北工业大学获得学士学位、硕士学位和博士学位, 现为中国科学院声学研究所研究员、博士生导师, 主要研究方向为阵列信号处理、水下无线传感器网络、合成孔径声纳等。

E-mail: hhn@mail.ioa.ac.cn

Huang Haining (Corresponding author) received his B. Sc., M. Sc. and Ph. D. degree all from Northwest Polytechnical University in 1991, 1996, and 1999, respectively. He is currently a professor and a Ph. D. advisor at the Institute of Acoustic, Chinese Academy of Science. His main research interests include array signal processing, underwater sensor network and synthetic aperture sonar.