

文章编号 1004-924X(2023)24-3618-12

细粒度遥感舰船开集识别

柳长源*, 李 婷, 兰朝凤

(哈尔滨理工大学 测控技术与通信工程学院, 黑龙江 哈尔滨 150080)

摘要:为了解决传统深度卷积神经网络在舰船图像细粒度分类中的局限性,本文设计了细粒度遥感舰船开集识别模型。首先,引入了基于注意力机制的STN模块,加在特征提取网络前用来过滤背景信息;然后在STN模块后接一个多尺度的并行的卷积结构,强化网络对不同尺度的局部区域的特征提取能力;接着将提取到的特征分别输入基分支和元嵌入分支,用来增大类间方差和减小类内方差,同时强化模型对尾类小样本的学习;最后对两个分支的分类结果进行决策融合,根据设定的阈值判别已知类和未知类进一步对已知类进行细分。在平衡与不平衡分布的FGSCR-42数据集上进行了4种开放度实验,结果表明:在平衡分布的数据集上4种开放度的平均准确率为90.5%,86.3%,85.7%,85.1%,不平衡分布数据集的平均准确率为90.0%,85.1%,84.3%,84.1%。与当前主流的舰船识别方法相比,本文方法分类具有更高的识别准确率和更好的泛化能力。

关键词:注意力机制;细粒度分类;开集识别;决策融合

中图分类号:TP391.41 **文献标识码:**A **doi:**10.37188/OPE.20233124.3618

Fine-grained remote sensing ship open set recognition

LIU Changyuan*, LI Ting, LAN Chaofeng

(College of Measurement and Control Technology and Communication Engineering,
Harbin University of Science and Technology, Harbin 150080, China)

* Corresponding author, E-mail: liuchangyuan@hrbust.edu.cn

Abstract: In this study, a fine-grained remote sensing ship open-set recognition model is designed to address the limitations of traditional deep convolutional neural networks in fine-grained classification of ship images. First, a STN module based on attention mechanism is introduced before the feature extraction network to filter background information. In addition, a multi-scale parallel convolution structure is added after the STN module to enhance the feature extraction ability of the network for local regions of different scales. The extracted features are input into the base and meta-embedded branches, to increase inter-class variance and reduce intra-class variance, strengthening the model's learning of the tail class small samples concomitantly. Finally, the classification results of the two branches are fused; known and unknown classes are distinguished according to the set threshold; and known classes are subdivided. Four types of openness experiments were conducted on the FGSCR-42 datasets with balanced and unbalanced distributions. The results show that the average accuracies of the four types of openness in the balanced distribution dataset are 90.5%, 86.3%, 85.7%, and 85.1%; the corresponding average accuracies of the unbalanced distribution dataset are 90.0%, 85.1%, 84.3%, and 84.1%. Compared with the current

收稿日期:2023-05-16;修订日期:2023-07-05.

基金项目:黑龙江省自然科学基金资助项目(No. F2016022);国家自然科学基金资助项目(No. 11804068)

mainstream ship recognition methods, the proposed method has higher recognition accuracy and better generalization ability.

Key words: attention mechanism; fine-grained classification; open set recognition; decision fusion

1 引言

舰船在现代化海域中扮演着重要的角色。军事方面,精准的舰船识别可以增强近海防御和远洋打击能力,民用方面,舰船识别技术可以帮助相关部门进行海上搜救和交通管制等活动^[1]。所以对舰船目标精准、高效地识别是智能化海洋监控应用中非常重要的研究方向^[2]。

由于不同类别的舰船之间形态较为相似而同类别的舰船之间形态差异较大,并且遥感舰船图像中的目标主体较小很难从周围环境中辨识出来,因此对舰船的精细识别具有较高的难度。目前,遥感舰船细粒度分类的关键在于局部微小的差异,而常规方法受到先验知识和弱特征表达的限制,难以处理复杂背景下的图像,且提取到的特征大多是全局特征,缺少对局部微小差异部分的重点关注^[3]。因此,需要将细粒度目标的识别方法引入到舰船的细粒度分类中。

Lin 等人^[4]提出了双线性卷积神经网络(Bilinear Convolutional Neural Network, BCNN),模型由两个子网络组成,一个子网络用于提取目标区域,另一个用于获取区域内的特征,以此来捕获不同子类间的微小差异,但是该模型忽略了不同卷积层之间的交互。Jianlong FU 等人^[5]结合了区域定位和区域内特征学习提出了循环注意力卷积神经网络(Recurrent Attention Convolutional Neural Network, RA-CNN)。该网络由三个尺度构成,每个尺度均由注意力网络(Attention Proposal Sub-Network, APN)和分类子网络构成,其中注意力网络用于定位注意力区域,此模型的缺点是三个尺度参数不共享,模型计算较复杂。Chen 等人^[6]提出了破坏与重构模型(Destruction and Construction Learning, DCL),使用区域混乱机制(Region Confusion Mechanism, RCM)扰乱了由输入图像分割成的多个区域,迫使网络更多的关注判别性区域以此提高分类准确性。以上方法虽然对于细粒度遥感舰船的

类取得了较好的效果,但是受客观因素的限制,训练深度神经网络时几乎不可能收集到所有的舰船类别,而传统的细粒度遥感舰船分类方法都具有闭集性,即对于未知的类别,分类器会作出错误的分类,将其强行划分成某种已知的类别,最终降低识别的准确率。

为了使细粒度舰船分类更好适用于现实场景,本文设计了细粒度遥感舰船开集识别模型。通过带有几何空间注意模块的主干网络提取深度特征,利用基分支学习更紧凑的类中心分布,元嵌入分支重新定位类中心,将细粒度遥感舰船的分类问题推向开放环境,打破了传统的基于闭集假设下分类问题的局限。

2 相关方法

2.1 开集识别

传统的卷积神经网络(Convolutional Neural Networks, CNN)由于使用 Softmax 函数将输出类标签对应于训练类的最大规范化概率作为预测结果。因此,当出现训练数据集中没有的未知类别时,分类器会错误地将其归为与已知类别最相似一类。由于现实场景中的物体多种多样很难获取所有的类别用于训练,为此许多学者提出开放集识别(Open-set Recognition, OSR)的概念^[7],并提出了许多基于深度学习的 OSR 方法,以缓解传统 CNN 的闭集特性。开放集识别描述了一种场景,即在拒绝未知的类别的同时还能精准识别出已知类别,如图 1 所示。OSR 方法原理

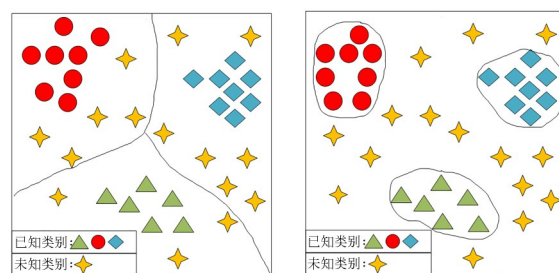


图 1 开放集识别场景

Fig. 1 Open set identification scenario

是优化最小化训练数据集 U 和开放空间风险以及经验风险,其定义如公式(1)所示:

$$f = \arg \min_{f \in H} \{R_o(f) + \lambda_r R_e(f(U))\}, \quad (1)$$

其中, λ_r 是一个正则化常数。

现有的 OSR 方法有判别模型^[8-9]和生成模型两种^[10-11]。判别模型侧重于校准模型的潜在特征表示或模型的后验概率分布。而生成模型一般是通过变分自动编码器(Variational Autoencoder, VAE)或者生成式对抗网络(Generative Adversarial Networks, GAN)对数据施加先验概率分布。虽然以上两种模型将 CNN 应用到一个未知的开放世界中,但是细粒度遥感舰船具有类内的多样性和类间的相似性,并且舰船通常与周围的环境相融合。将其应用于细粒度遥感舰船的识别更为复杂,识别的准确率也不高。

2.2 注意力机制

注意力机制是使计算机能够像人类一样自主获取重点关注区域,从而过滤掉无关的信息。现有的注意力有两种分别为硬注意^[12]和软注意^[13-16]。由于软注意力具有可微性使其在网络的正向传播和反向传播过程中能够实时进行更新,因此在 CNN 中被广泛使用。

Hu 等人^[13]提出了可自由嵌入到其他网络中的压缩激活模块(Squeeze and Excitation Network, SE-Net),该模块通过挤压、激励和尺度操作在通道维度上对特征进行处理,使得信息量较大的通道特征被更多的关注。借鉴 Hu 等人的思想 Wang 等人^[14]提出了一种高效通道注意力模块(Efficient Channel Attention Module, ECA),该模块是不降维式的跨通道交互模块,仅仅利用了通道维度上的特征。Woo S^[15]等人整合了通道注意力和空间注意力提出了轻量级的注意力模块(Convolutional Block Attention Module, CBAM),虽然以上注意力模块在分类任务中表现良好,但是不能将所关注的重点区裁剪出来使输入图片变成理想的图片。为了在过滤背景信息的同时更突出目标主体,Jader-berg 等人^[16]提出了一种空间变换网络(Spatial Transformer Networks, STN),在空间域内进行空间变换来提取关键信息。经过训练的 STN 可以找到感

兴趣区域(Region Of Interest, ROI)的区域,并具有旋转、缩放和变换的功能。因此,本文将 STN 引入到特征提取网络前,用于去除舰船背景信息的干扰,将舰船目标主体裁剪出来并放大到标准尺寸。

3 算法设计

本文的分支融合网络模型由基分支和元嵌入分支构成。首先采用 STN 提取细粒度遥感舰船图像中的舰船目标,之后使用特征提取网络提取舰船目标图像的深度特征分别作为基分支和元嵌入分支的输入,基分支用于减小舰船图像的类内差异,元嵌入分支用于增大舰船图像的类间差异并加强对小样本的学习能力,同时也对基分支的分类结果进行校准。最后对两个分支的分类结果决策融合,根据设定的阈值判别已知类和未知类,若为已知类则进一步细分出具体类别。模型的总体框图如图 2 所示。

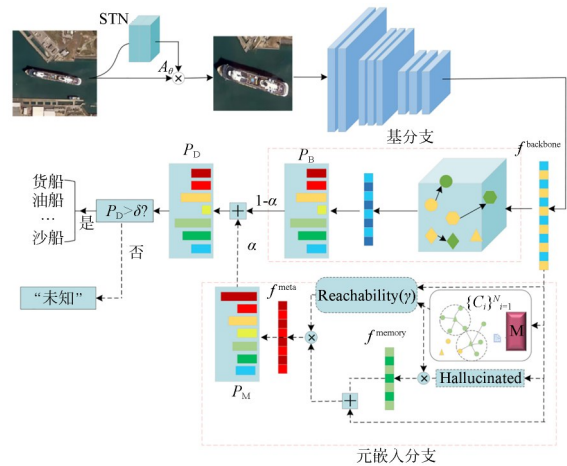


图 2 模型整体结构框图

Fig. 2 Model overall structure block diagram

3.1 多尺度 STN-Resnet 网络

STN 主要由三部分构成分别是:参数预测(localisation net)、坐标映射(grid generator)和像素采集(sampler),基本架构如图 3 所示。

参数预测模块的输入 U , 输出变换矩阵, 决定 STN 模块对输入图像或特征图做哪种变换; 坐标映射用于做矩阵运算, 该运算以目标图像 V

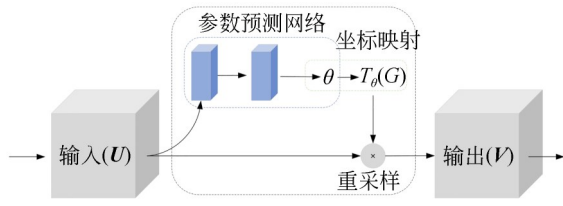


图3 STN基本架构

Fig. 3 Basic architecture of STN

中的所有坐标点为自变量, θ 为参数计算输入图像 U 的坐标点, 具体计算过程如式(2)所示:

$$\begin{pmatrix} x_i^s \\ y_i^s \\ 1 \end{pmatrix} = \theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix}, \quad (2)$$

其中: (x_i^s, y_i^s) 是原图像坐标, (x_i^t, y_i^t) 是输出特征映射中匹配的目标图像的坐标。输入和输出图像的尺寸可以相同也可以不同, 不同时即为缩放操作, 缩放程度主要依赖于 θ_{11} 和 θ_{22} 。经过坐标变换后有部分坐标可能是小数, 而像素采集通过插值的方式来处理经过坐标变换后得到的小数坐标。

STN可以任意位置、任意数量的插入到现有的卷积神经网络中, 使得网络能够自主的对整张图像或特征图进行平移、缩放、剪裁和旋转等空间变换。由于该模块在选择图像中重点关注区域的同时, 还可以将该区域放大到标准尺寸。故本文将其插入到特征提取网络 Resnet34^[17]前, 用于提取舰船目标主体。

由于卷积核大小不同获得的输入图像中的感知域也不同, 而细粒度舰船的分类关键在于细小差异。仅使用一种尺度的卷积核, 提取到的信息比较单一, 在一定程度上限制了细粒度遥感舰船的分类准确性。因此, 本文使用 1×1 , 3×3 和 5×5 不同大小的卷积核构建了两个并行的卷积通道来提取不同尺度的信息, 其中第一条通道由 1×1 , 3×3 卷积和BN层构成; 第二条通道由 1×1 , 5×5 卷积和BN层构成, 最后将两个通道的输出进行拼接经过一个 1×1 卷积降维输入到 Resnet34 网络。并行卷积模块结构示意图如图4所示。

3.2 基分支结构

对于细粒度分类而言仅利用特征提取网络得到不同类别之间的潜在特征, 区分性不是很好, 而利用STN先去除背景信息的干扰, 由特征

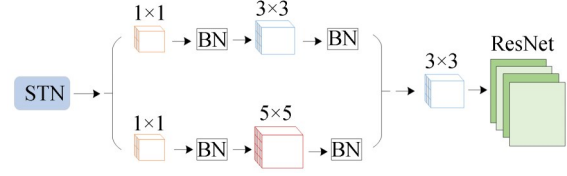


图4 并行卷积模块结构示意图

Fig. 4 Structure diagram of parallel convolution module

提取网络得到的深度特征辨别能力依旧不强, 以此进行分类也很难达到令人满意的效果。Wen 等人在 2016 年提出中心损失函数 (Center Loss)^[18]。Center Loss 最初应用在人脸这种复杂的数据分布, 使得同种类别间的特征更加紧凑, 公式如式(3)所示:

$$L_c = \frac{1}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2, \quad (3)$$

其中: x_i 是由特征提取网络提取到的第 i 个特征, $c_{y_i} \in \mathbf{R}^d$ 是类别 y_i 的特征中心, d 是特征维数, m 为批量处理的大小。

特征中心 c_{y_i} 随深度特征的变化而变化, 越靠近最近类中心的特征, 越属于相应的类。为了提高网络的训练效率, 训练时基于小批量 m 更新中心, 并在每次迭代的过程中, 通过平均相应类的特征来计算中心, 目的是避免更新所有类中心导致网络训练时间的增加。

L_c 梯度和类中心的更新方程^[18]如式(4)和式(5)所示:

$$\frac{\partial L_c}{\partial x_i} = x_i - c_{y_i}, \quad (4)$$

$$\Delta c_j = \frac{\sum_{i=1}^m \delta(y_i=j) \cdot (c_j - x_i)}{1 + \sum_{i=1}^m \delta(y_i=j)}, \quad (5)$$

其中: $y_i=j$ 为判断 δ 状态的条件, 当条件成立, $\delta=1$; 如果不成立, $\delta=0$ 。当 y_i 和 c_j 的类别 j 不同时, 不需要更新, 只有类别相同时才更新。如果单独使用 Center Loss 来监督网络的学习, 会使得深度学习的特征和类中心退化为零, 导致损失非常小。又由于该损失函数和 Softmax Loss 一样不需要对样本进行复杂的重组, 两者的要求几乎相同。因此, 本文将 Center Loss 加到基分支中, 和 Softmax Loss 联合监督, 加强对辨别特征的学习, 提高识别的准确率。故基分支中的总损失函数 L_{Base} 为 L_{SoftMax} 和 L_c 的加和, 如式(6)和式

(7)所示:

$$L_{\text{Base}} = L_{\text{SoftMax}} + \lambda_c L_c, \quad (6)$$

$$L_{\text{SoftMax}} = - \sum_{i=1}^m \log \frac{e^{\mathbf{W}_j^T \mathbf{x}_i + b_j}}{\sum_{j=1}^n e^{\mathbf{W}_j^T \mathbf{x}_i + b_j}}, \quad (7)$$

其中: $\mathbf{W}_j \in \mathbf{R}^d$ 表示最后一个全连接层中权重 $\mathbf{W} \in \mathbf{R}^{d \times n}$ 的第 j 列, d 是特征维度, n 是类别数, $b \in \mathbf{R}^n$ 是偏置项, 标量 λ_c 在 $[0, 1]$ 中受限, 主要用于平衡两个损失函数。 λ_c 的值影响了不同特征之间的可分性, λ_c 值越大, 会导致同类之间的特征更紧凑, 反之, 同类间的特征更分散。

该分支的输入为 Resnet 网络最后一个卷积层的输出特征 $\mathbf{f}^{\text{backbone}}$, 采用双损失共同监督, 其中, Center Loss 用于减小同类间的差异, Softmax Loss 用于学习类间的差异, 两者的联合加强了对深度特征的辨别能力的学习, 使网络更适用于类间差异小的目标。

3.3 元嵌入分支

细粒度舰船具有较大的类内差异的同时具有较小的类间差异。虽然基分支有效解决了同类别舰船图像之间差异较大的问题, 但是由于不同类别舰船之间的差异很小, 单独使用基分支对有细微差异的舰船分类准确率较低。为了减小舰船图像类内变化的同时增加其类间的差异, 使特征在不受相应类中心影响的同时还增大了与其他类中心之间的距离。本文的元嵌入分支借鉴类结构分析方法^[19], 采用判别中心作为基本构造块。将训练的主干特征 $\mathbf{f}^{\text{backbone}}$ 作为元嵌入分支的输入, 让 $M = \{c_i\}_{i=1}^N$ 表示所有训练数据的视觉记忆, N 为训练类别数目。在训练过程中, 随机抽取一定的样本数据用于更新中心 c_i , 最大限度地提高训练样本和其他类中心之间的距离。

由于细粒度舰船图像采集困难, 所以数据分布大多数是不均衡的。对于开放环境下的数据量较少的类别缺少监督更新的能力。借鉴文献^[20]的思想, 在元嵌入分支中使用记忆特征 $\mathbf{f}^{\text{memory}}$ 丰富主干特征。使用幻觉(hallucinated)系数 $O \in \mathbf{R}^N$ 评估每个类别的中心和 $\mathbf{f}^{\text{backbone}}$ 之间的权重关系如式(8)所示:

$$\mathbf{f}^{\text{memory}} = O^T C = \sum_{i=1}^N o_i c_i. \quad (8)$$

幻觉系数通过一个轻量级的全连接网络从

主干特征中获得的, 并通过 γ 对特征 $\mathbf{f}^{\text{memory}}$ 进行缩放, 得到元特征 \mathbf{f}^{meta} 。如式(9)所示:

$$\mathbf{f}^{\text{meta}} = (1/\gamma) \cdot \mathbf{f}^{\text{memory}}, \quad (9)$$

其中, γ 为主干特征 $\mathbf{f}^{\text{backbone}}$ 与其最近的类中心之间的距离如式(10)所示:

$$\gamma = \text{reachability}(\mathbf{f}^{\text{backbone}}, C) = \min_i \|\mathbf{f}^{\text{backbone}} - c_i\|_2. \quad (10)$$

当 γ 较大时认为输入的样本是未知的即远离每个类的中心, 并且它的元特征在所有维度上均被缩到近似于零, 这进一步体现了开放集类样本的特征表示与已知训练类样本的特征有显著差异。相反, γ 较小认为输入的是一个已知类的样本, 此时每个维度的差异都保持不变。元嵌入分支作为训练的第二阶段, 总损失函数 L_{meta} 为交叉熵损失函数 L_{CE} 与最大边际损失函数 L_M 的加和, 如式(11)所示:

$$L_{\text{meta}} = L_{\text{CE}} + \lambda_m L_M(\mathbf{f}^{\text{backbone}}, \{c_i\}_{i=1}^N), \quad (11)$$

其中, λ_m 为最大边际损失的权重。

此分支使用由余弦相似度设计的分类器^[21]替换传统的 Softmax 分类器, 这使得该模型能够统一对已知类和未知类进行识别, 其中分类结果 P_M 由余弦相似度算子计算, 如公式(12)所示:

$$P_M = \tau \cdot \cos(\mathbf{f}^{\text{meta}}, \mathbf{w}_k) = \tau \cdot \overline{\mathbf{f}^{\text{meta}}} \overline{\mathbf{w}_k}^T, \quad (12)$$

其中: $\overline{\mathbf{f}^{\text{meta}}} = \frac{\mathbf{f}^{\text{meta}}}{\|\mathbf{f}^{\text{meta}}\|}$, $\overline{\mathbf{w}_k} = \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}$ 为特征和第 k 类

权重向量的 L2 正则化表示, τ 是控制 Softmax 概率分布峰值的参数。

3.4 分支融合

在大多数情况下, 先融合多个特征, 再送入分类器进行分类。这种特征融合的方法在闭集假设下的分类任务中取得了较好的效果, 但是对于开集的分类任务效果并不理想。开集的分类任务常常会出现未知的类别。按照人类对未知物体的感知角度, 融合它最终的分类结果, 而不是它应该属于的类别即特性。因此, 本文通过端对端融合两个分支的最终识别结果来确定舰船的所属类别。最终两个分支融合的分类结果 P_D 如式子(13)所示, 即将基分支的分类结果 P_B 和元嵌入分支的分类结果 P_M 进行决策融合。

$$P_D = \alpha P_M + (1 - \alpha) P_B, \quad (13)$$

其中, α 表示校准度。

为了更好地拒绝未知类别,在融合分类结果 P_D 后设定一个阈值,用于判别是否为已知类别,最终分类结果 y 如式(14)所示:

$$y = \begin{cases} \arg \max_{i \in \{1, \dots, N\}} P_D(C_i|x) & P \geq \delta \\ \text{未知} & \text{otherwise} \end{cases}, \quad (14)$$

其中, δ 是区分已知类和未知样本的鉴别概率阈值。

4 测量实验与结果

4.1 数据集预处理

FGSCR-42 数据集^[22]是一个广泛的光学遥感图像中细粒度舰船分类的公共数据集,包含类别较多的军用舰船和民用舰船的图像。为了验证本文方法的有效性,选取了 FGSCR-42 数据集的一部分进行评估。考虑到民用船易于采集、军用舰船采集困难,选取 7 种民用船为已知类别作为训练样本,将多种军用舰船的混合作为未知类别和 7 种已知类的民用船一起作为测试类别。每类样本示例如图 5 所示。

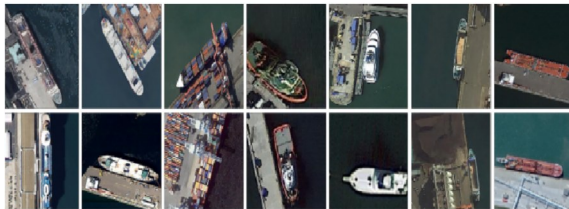


图 5 样本示例

Fig. 5 Sample example

OSR 方法中的测试集类要比训练集类多,故在本研究中将多种军用舰船的混合统一当作未知类别,和 7 种已知类别一起作为测试类。为了更适用于现实场景,在实验中将数据集设置为平衡和不平衡分布两种,平衡分布的训练数据集中每个类别有 400 个,共 2 800 个样本如表 1 所示,不平衡的训练数据集如表 2 所示。

由于遥感图像的目标识别需要不同视角的识别能力,因此每类选取 60 多个方向的测试样本,按照开放度公式^[23]如式(15),设置了 4 种不同的开放度分别为 9.25%, 14.16%, 18.35% 和 21.98%,每种开放度下的对应的未知类别中分

别包含 3, 5, 7 和 9 种军用舰船。

$$O = 1 - \sqrt{\frac{2 \times |C_{TR}|}{|C_{TR}| + |C_{TE}|}}, \quad (15)$$

其中: C_{TR} 为训练类别的标签数量, C_{TE} 为测试类别的标签数量。

表 1 平衡分布数据集

Tab. 1 Dataset of balanced distribution

类别序号	类别	训练数量
0	油船(Tank ship)	400
1	巨型船(Megayacht)	400
2	砂船(Sand ship)	400
3	货船(Cargo)	400
4	集装箱(Container)	400
5	拖船(Towing ship)	400
6	民船(Civil ship)	400
0~6	总数	2 800

表 2 不平衡分布数据集

Tab. 2 Datasets of the unbalanced distribution

类别序号	类别	训练数量
0	油船(Tank Ship)	220
1	巨型船(Megayacht)	180
2	砂船(Sand Ship)	250
3	货船(Cargo)	300
4	集装箱(Container)	400
5	拖船(Towing Ship)	750
6	民船(Civil Ship)	700
0~6	总数	2 800

测试集的设置见表 3 所示,以此评估本文方法的有效性。

表 3 测试数据集

Tab. 3 Test dataset

开放度/%	未知类包含种类	总测试数量
9.25	3	600
14.16	5	720
18.35	7	840
21.98	9	960

对处理后的数据集采取常用的图像处理方法,首先对细粒度遥感舰船图像进行水平翻转和垂直翻转并旋转 90°, 180° 和 270° 来增加数据的多

样性。之后对光学 RGB 图像中的每个像素用平均值 μ_x 和标准差 σ_x 进行标准化,如式(16)所示,得到泛化效果:

$$\hat{x} = \frac{x - \mu_x}{\sigma_x}, \quad (16)$$

其中, x 和 \hat{x} 分别表示原始图像和标准化图像的像素值。

4.2 实验环境及参数设置

实验硬件环境采用 IntelCore i9-13900K 处理器, RTX4080 魔鹰显卡, 显存为 16 G。软件环境为 Ubuntu22.04.2 和 CUDA11.7, 使用 pytorch1.8 深度学习框架。初始学习率设置为 0.001, 每 50 轮收缩 1/10, 一阶段训练 100 轮, 二阶段训练 200 轮, 在这个阶段中冻结主干网络的参数, 即优化器只更新与元记忆分支相关的参数。以上两个阶段中 batch-size 均设置为 24。 λ_c 和 λ_m 控制了来自相应类中心的吸引力, 分别设置为 0.01 和 0.1。 α 决定元嵌入分支对基分支分类结果的校准程度, 将 α 从 0 变化到 0.5, 同样在两种数据集上进行实验如图 6 所示。

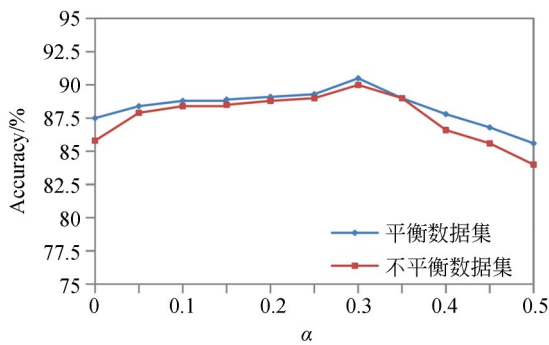


图 6 不同 α 下的识别准确率曲线图

Fig. 6 Graph of recognition accuracy under different α

通过观察发现随着 α 值的变化而最终的识别精度并不那么稳定, α 值越高, 识别的精度越低, 而 α 值越低, 不会显著提高精度, 综合考虑将其设置为 0.3。

鉴别概率阈值 δ 表示已知类和未知类之间的界限, 在一定程度上影响网络的识别准确率。而本文中已知类占比较多, 因此同样在低开放度下将从 0.5 到 0.9 在平衡分布和不平衡分布的两种数据集上进行实验, 将识别的准确率用折线图表示如图 7 所示。

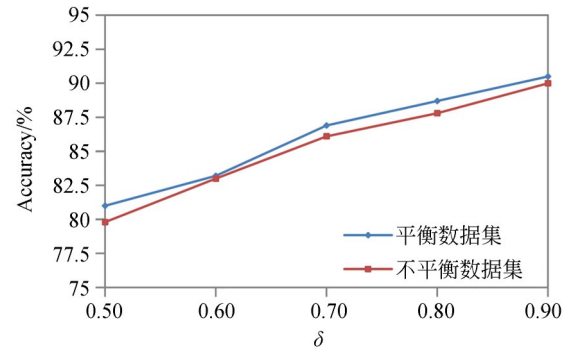


图 7 不同 δ 下的识别准确率曲线图

Fig. 7 Different δ recognition accuracy curves

通过观察发现 δ 为 0.9 时在两种数据集下的识别准确率均较好, 故本研究 δ 选取为 0.9。

4.3 消融实验

为了验证 STN 模块和中心损失函数 L_c 对该模型整体性能的贡献, 本文在不平衡数据集上用准确率作为评价指标来评估该模型性能。因为现实场景中所能采集的不同类别图像数量很难达到均衡, 所以采用不平衡数据集进行验证, 更加符合实际的情况, 实验的结果如图 8 所示。

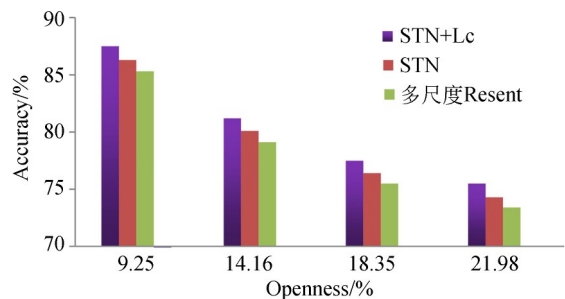


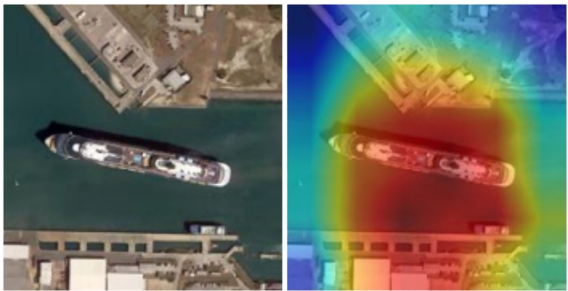
图 8 不同开放度下 STN 实验对比图

Fig. 8 Comparison of STN experiments under different openness

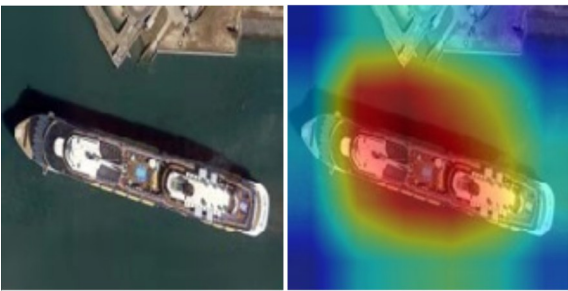
从图 8 中可以看出, 加入 STN 模块后 SResNet34 网络的识别准确率整体高于多尺度 Resnet 网络的识别准确率, 可见背景信息的干扰较大的影响了舰船识别。在加入 STN 模块的基础上使用 L_c 中心损失函数聚合类内特征, 使得网络的识别准确率得到了进一步提升, 由此可知, L_c 也能够提升该模型整体性。

为了更直观地看出 STN 模块对网络性能的影响, 将输入 ResNet34 网络和 SResNet34 网络的原始图像和注意力图进行对比, 如图 9 所示(彩图

见期刊电子版)。注意力图中红色部分表示网络更多关注的区域,由图可知加入 STN 模块后的 SResNet34 网络受背景信息干扰明显减小,更多关注目标主区域。



(a) ResNet34网络的原始图像和注意力图
(a) Raw images and attention maps of the ResNet34 network



(b) SResNet34网络的原始图像和注意力图
(b) Raw images and attention maps of the SResNet34 network

图 9 ResNet34 网络和 SResNet34 网络的原始图像注意力图
Fig. 9 ResNet34 network and SResNet34 network original image attention diagram

元嵌入分支中的元特征主要由 reachability 和幻觉(hallucinated)系数两个模块构成。为了验证这两个模块对该模型整体性能的贡献。同样在不平衡分布的数据集上进行验证,实验结果如图 10 所示。

通过观察,可以发现当未知类数量较少时,仅包含幻觉(hallucinated)系数的模型比仅包含 reachability 模块的模型具有较好的识别结果。然而,随着未知类的增加,它的性能不如包含 reachability 模块的网络模型好,因为当未知类别较少时,该模型的精度主要取决于已知类别的分类准确率。由图 10 可知,单独使用这两者并没有显著提高网络模型的性能,但将这两者的结合对模型的性能有很大提升。

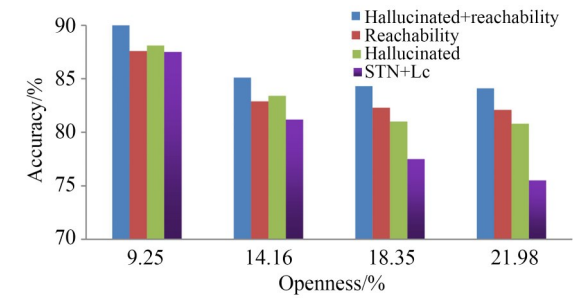


图 10 不同开放度下 hallucinated 和 reachability 实验对比图
Fig. 10 Experimental comparison of hallucinated inated and reachability under different degrees of openness

4.4 对比实验及结果分析

为了更好地验证本文方法的泛化性能,首先在 21.98% 的开放度下将本文模型分别应用在处理后的细粒度遥感舰船 FGSC-23 数据集和 FG-SCR-42 数据集上,并对其结果进行比较,如表 4 所示。可以看到本文方法在 FGSC-23 和 FG-SCR-42 数据集上的识别准确率均高于其他几个对比模型。

表 4 不同数据集上准确率的比较

Tab. 4 Comparison on accracy of different data sets (%)

方法/数据集	FGSCR-42	FGSC-23
Resnet	44.9	59.8
BCNN	45.7	61.3
OpenMax	73.8	78.8
OLTR	76.2	80.4
CenterLoss	78.3	81.6
本文方法	85.1	87.7

为了全面评价本文提出的细粒度图像分类与开集融合方法的有效性,单独将本文模型应用到处理后的细粒度遥感舰船图像 FGSCR-42 数据集上,评价指标为准确率、F1-score 和混淆矩阵。表 5 表示在平衡数据集下不同模型的性能比较。观察结果发现,Resnet 和 BCNN 的总体精度接近于测试样本中已知类样本的比例,这充分证明了传统的基于卷积神经网络的细粒度遥感舰船分类方法具有闭集性,即全部开放的未知类样本均被归为已知类样本;Openmax, Center Loss 和 OLTR 方法改善了传统的基于卷积神经网络的图像分类的闭集性,使其更适应于开放环境。

然而,随着开放类的增加,准确率和 F1-score 的性能都有显著下降。由表 5 可知,本文的方法整体性能最好,在 9.25% 开放度下相比于 Center Loss 模型准确率提升 1.5%,F1-score 提升 0.8%;21.98% 开放度下准确率提升 6.8%,F1-score 提升 3.6%。本文模型随着未知类的增加平均准确率和 F1-score 值降低的较少。

在不平衡数据集上训练的不同模型的性能比较见表 6 所示。由于训练数据分布不平衡,各模型的准确性在一定程度上有所下降,因为此时

除了区分已知类样本和未知类样本外,尾部类和未知类的有效区分也增加了识别的难度。从表中可以观察到,基于闭集假设下的 Resnet 和 BCNN 模型对细粒度遥感舰船分类精度较低,有些尾部类被识别为未知样本。OpenMax,Center Loss 和 OLTR 等方法随着开放程度的增加识别的准确率和 F1-score 值均比对应开放程度下基于闭集假设的识别准确率高,但与平衡数据集下的识别准确率相比降低的较多,由此可知以上 3 种方法易受数据分布的影响。

表 5 平衡数据集上各方法的比较
Tab. 5 Comparison of methods on the balanced datasets

方法/开放度	9.25%		14.16%		18.35%		21.98%	
	Accracy	F1-score	Accracy	F1-score	Accracy	F1-score	Accracy	F1-score
Resnet	72.4%	75.8%	60.3%	66.4%	50.2%	62.5%	44.9%	60.1%
BCNN	73.6%	76.5%	61.2%	68.2%	51.0%	64.1%	45.7%	63.1%
OpenMax	83.7%	87.2%	78.9%	84.0%	76.3%	81.0%	73.8%	77.9%
OLTR	88.7%	90.8%	82.4%	86.5%	78.5%	82.5%	76.2%	80.1%
CenterLoss	89.0%	91.4%	83.2%	88.2%	80.1%	84.6%	78.3%	82.8%
本文方法	90.5%	92.2%	86.3%	89.4%	85.7%	87.9%	85.1%	86.4%

表 6 不平衡数据集上各方法的比较
Tab. 6 Comparison of methods on the unbalanced datasets

方法/开放度	9.25%		14.16%		18.35%		21.98%	
	Accracy	F1-score	Accracy	F1-score	Accracy	F1-score	Accracy	F1-score
Resnet	72.0%	75.5%	60.3%	66.2%	50.1%	62.4%	44.3%	59.4%
BCNN	73.3%	85.2%	61.1%	68.0%	51.0%	63.8%	45.5%	62.9%
OpenMax	81.6%	86.3%	75.5%	81.2%	74.3%	78.8%	72.7%	76.5%
OLTR	85.2%	88.1%	77.8%	83.5%	75.5%	82.0%	74.5%	79.8%
CenterLoss	85.5%	87.6%	78.5%	83.1%	77.8%	82.6%	76.5%	81.1%
本文方法	90.0%	91.9%	85.1%	88.4%	84.3%	85.8%	84.1%	85.3%

图 11 表示开放度为 9.25% 时的混淆矩阵,其中(a)表示在平衡分布的数据集下标准化的混淆矩阵,(b)表示在不平衡数据集下标准化的混淆矩阵。每个类别的召回率由混淆矩阵的对角线表示,且每一行的概率加和为 1。

由混淆矩阵可直观地看出开放度相同的情况下模型在平衡与不平衡分布的数据集上识别

准确率相差较小。进一步证明了本文的方法受数据分布影响较小。

综上所述,本文的方法无论在平衡分布的数据集下还是不平衡分布的数据集下均优于先进的判别 OSR 方法和细粒度图像分类方法,并且在未知类的数量增加和数据不均衡时具有较强的泛化能力。

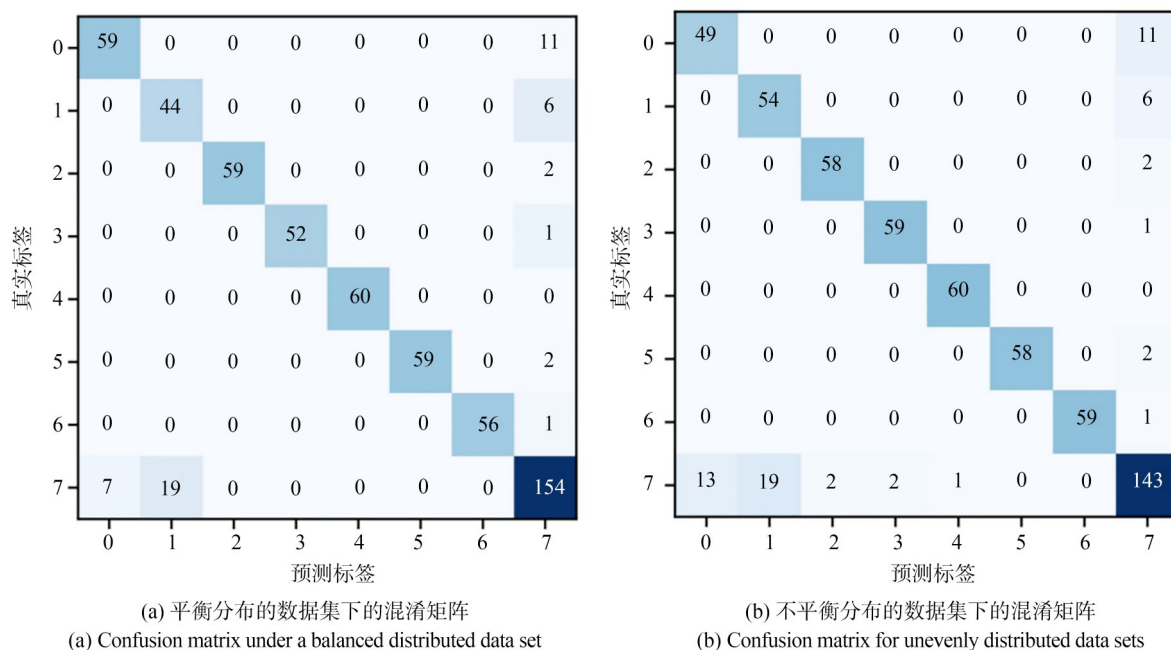


图 11 开放度 9.25% 时的混淆矩阵

Fig. 11 Confusion matrix at 9.25% openness

5 结 论

本文将开集识别方法融入到细粒度图像分类中,有效解决了传统细粒度遥感舰船分类的局限性。通过STN模块提取舰船目标主体,利用损失函数减小类内差异的同时增加类间差异,最后将基分支和元嵌入分支的分类结果进行决策融合。在开放程度不同时对平衡分布和不平衡分布的数据集进行了充分实验,结果表明,本文方法在开放度为 9.25% 时平衡分布和不平衡分布的数据集上识别准确率偏差最小是 0.5%,开放度为

18.35% 时识别准确率偏差最大是 1.4%,与基于判别模型 OSR 中表现较好的 CenterLoss 模型相比,在平衡数据集的 4 种开度下准确率分别提升 1.5%,3.1%,5.6%,6.8%,F1-score 提升 0.8%,1.2%,3.3%,3.6%,不平衡的数据集下准确率分别提升 4.5%,6.6%,6.5%,7.6%,F1-score 提升 4.3%,5.3%,3.2%,4.2%。综上,本文方法具有较高的准确率和较好的泛化能力。在未来的工作中,可考虑雨雪和浓雾天气,引入更加复杂背景,再结合本文模型设计一个整体的识别系统,使得其更符合实际情况。

参考文献:

- [1] 杜梅,李景景. 复杂背景遥感影像中小目标船舶细粒度智能识别[J]. 舰船科学技术, 2021, 43(20): 205-207.
DU M, LI J J. Fine-grained intelligent recognition of small target ships in remote sensing images with complex background[J]. *Ship Science and Technology*, 2021, 43(20): 205-207. (in Chinese)
- [2] 刘忻伟,朴永杰,郑亮亮,等. 面向航天光学遥感复杂场景图像的舰船检测[J]. 光学精密工程, 2023, 31(6): 892-904.
LIU X W, PIAO Y J, ZHENG L L, *et al.* Ship de-

- tection for complex scene images of space optical remote sensing[J]. *Opt. Precision Eng.*, 2023, 31(6): 892-904. (in Chinese)
- [3] 肖术明,张叶,常旭岭,等. 面向航天光学遥感场景压缩感知测量值的舰船检测[J]. 光学精密工程, 2023, 31(4): 517-532.
XIAO S M, ZHANG Y, CHANG X L, *et al.* Ship detection oriented to compressive sensing measurements of space optical remote sensing scenes[J]. *Opt. Precision Eng.*, 2023, 31(4): 517-532. (in Chinese)
- [4] LIN T Y, ROYCHOWDHURY A, MAJI S. Bi-

- linear CNN Models for Fine-Grained Visual Recognition[C]. 2015 *IEEE International Conference on Computer Vision (ICCV)*. 7-13, 2015, Santiago, Chile. IEEE, 2016: 1449-1457.
- [5] FU J L, ZHENG H L, MEI T. Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 4476-4484.
- [6] CHEN Y, BAI Y L, ZHANG W, *et al.* Destruction and construction learning for fine-grained image recognition [C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 15-20, 2019, Long Beach, CA, USA. IEEE, 2020: 5152-5161.
- [7] KODAMA Y, WANG Y N, KAWAKAMI R, *et al.* Open-set recognition with supervised contrastive learning[C]. 2021 *17th International Conference on Machine Vision and Applications (MVA)*. 25-27, 2021, Aichi, Japan. IEEE, 2021: 1-5.
- [8] DAI W, DIAO W H, SUN X, *et al.* CAMV: class activation mapping value towards open set fine-grained recognition [J]. *IEEE Access*, 2021, 9: 8167-8177.
- [9] CAI Z W, VASCONCELOS N. Cascade R-CNN: high quality object detection and instance segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43 (5) : 1483-1498.
- [10] VENDRAMINI M, OLIVEIRA H, MACHADO A, *et al.* Opening deep neural networks with generative models[C]. 2021 *IEEE International Conference on Image Processing (ICIP)*. 19-22, 2021, Anchorage, AK, USA. IEEE, 2021: 1314-1318.
- [11] YANG Y, HOU C, LANG Y, *et al.* Open-set human activity recognition based on micro-Doppler signatures [J]. *Pattern Recognition*, 2019, 85: 60-69.
- [12] 张婷. 基于注意力机制的图像内容理解与视觉推理算法研究[D]. 成都: 电子科技大学, 2021.
- ZHANG T. *Research on Image Content Understanding and Visual Reasoning Algorithm Based on Attention Mechanism*[D]. Chengdu: University of Electronic Science and Technology of China, 2021. (in Chinese)
- [13] 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. IEEE, 2018: 7132-7141.
- [14] WANG Q L, WU B G, ZHU P F, *et al.* ECA-Net: efficient channel attention for deep convolutional neural networks[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 13-19, 2020, Seattle, WA, USA. IEEE, 2020: 11531-11539.
- [15] WOO S, PARK J, LEE J Y, *et al.* CBAM: Convolutional Block Attention Module[M]. Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [16] JADERBERG M, SIMONYAN K, ZISSERMAN A, *et al.* Spatial Transformer Networks [EB/OL]. 2015; *arXiv*: 1506.02025. <https://arxiv.org/abs/1506.02025.pdf>
- [17] ZHANG H L, XU H, LIN T E. Deep open intent classification with adaptive decision boundary [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(16): 14374-14382.
- [18] WEN Y D, ZHANG K P, LI Z F, *et al.* A Discriminative Feature Learning Approach for Deep Face Recognition[M]. Computer Vision - ECCV 2016. Cham: Springer International Publishing, 2016: 499-515.
- [19] XU X W, ZHANG X L, ZHANG T W. Multi-Scale SAR Ship Classification with Convolutional Neural Network [C]. 2021 *IEEE International Geoscience and Remote Sensing Symposium IGARSS*. 11-16, 2021, Brussels, Belgium. IEEE, 2021: 4284-4287.
- [20] LIU Z W, MIAO Z Q, ZHAN X H, *et al.* Large-scale long-tailed recognition in an open world[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 15-20, 2019, Long Beach, CA, USA. IEEE, 2020: 2532-2541.
- [21] LUO X, YU Z J, ZHAO Z G, *et al.* Effective short text classification via the fusion of hybrid features for IoT social data[J]. *Digital Communications and Networks*, 2022(6): 942-954.

- [22] DI Y H, JIANG Z G, ZHANG H P. A public dataset for fine-grained ship classification in optical remote sensing images [J]. *Remote Sensing*, 2021, 13(4): 747.
- [23] GENG C X, HUANG S J, CHEN S C. Recent advances in open set recognition: a survey [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(10): 3614-3631.

作者简介:



柳长源(1970—),男,黑龙江哈尔滨人,博士,副教授,硕士生导师,1993年于吉林大学获得理学学士学位,2004年于哈尔滨理工大学获得硕士学位,2013年于哈尔滨工程大学获得博士学位,2016-2017年在美国普渡大学做访问学者,自动化学会会员。研究方向为图像处理、人工智能、模式识别等。
E-mail: liuchangyuan@hrbust.edu.cn