

任务感知的多尺度小样本SAR图像分类方法

张睿, 王梓祺, 李阳, 王家宝, 陈瑶

引用本文

张睿, 王梓祺, 李阳, 王家宝, 陈瑶. 任务感知的多尺度小样本SAR图像分类方法[J]. 计算机科学, 2024, 51(8): 160-167.

ZHANG Rui, WANG Ziqi, LI Yang, WANG Jiabao, CHEN Yao. Task-aware Few-shot SAR Image Classification Method Based on Multi-scale Attention Mechanism [J]. Computer Science, 2024, 51(8): 160-167.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于原型修正的小样本半监督语义图像翻译算法](#)

Few-shot Semi-supervised Semantic Image Translation Algorithm Based on Prototype Correction
计算机科学, 2024, 51(8): 224-231. <https://doi.org/10.11896/jsjcx.230500038>

[基于伪标签依赖增强与噪声干扰消减的小样本图像分类](#)

Few-shot Image Classification Based on Pseudo-label Dependence Enhancement and Noise Interference Reduction
计算机科学, 2024, 51(8): 152-159. <https://doi.org/10.11896/jsjcx.230500066>

[基于拉格朗日对偶的小样本学习隐私保护和公平性约束方法](#)

Lagrangian Dual-based Privacy Protection and Fairness Constrained Method for Few-shot Learning
计算机科学, 2024, 51(7): 405-412. <https://doi.org/10.11896/jsjcx.230500012>

[基于特征手性的数据无关模型评估方法](#)

Data-free Model Evaluation Method Based on Feature Chirality
计算机科学, 2024, 51(7): 337-344. <https://doi.org/10.11896/jsjcx.230500179>

[基于跨域小样本学习的SAR图像目标识别方法](#)

SAR Image Target Recognition Based on Cross Domain Few Shot Learning
计算机科学, 2024, 51(6A): 230800136-7. <https://doi.org/10.11896/jsjcx.230800136>

任务感知的多尺度小样本 SAR 图像分类方法

张睿 王梓祺 李阳 王家宝 陈瑶

陆军工程大学指挥控制工程学院 南京 210007

(3959966@qq.com)

摘要 针对 SAR 图像分类时存在的带标注样本较少的问题,提出了一种任务感知的多尺度小样本 SAR 图像分类方法。为了能够充分挖掘局部特征并关注具体任务下的关键局部语义 patches,引入了两种有效的注意力机制,获得了更加高效且丰富的特征表示。首先,在特征提取阶段使用互补注意力模块(CSE Block),关注原始特征中不同语义部分的显著特征,从被抑制的特征中提取次级显著特征并与主要显著特征融合,得到更加高效且丰富的特征表示。随后,利用自适应情景注意力模块(AEA Block)获得整个任务中的关键语义 patches,增强任务间的区分信息,提升小样本 SAR 图像分类任务的精度。结果表明,在 SAR 图像分类标准数据集 MSTAR 上,5-way 1-shot 任务分类精度相较于次优方法精度提升了 2.9%,并且该方法在两项任务中的运行时间与其他度量学习方法相比水平相当,未额外增加过多的计算资源,验证了其有效性。

关键词: 多尺度注意力机制;小样本学习;SAR 图像分类;度量学习

中图分类号 TP391

Task-aware Few-shot SAR Image Classification Method Based on Multi-scale Attention Mechanism

ZHANG Rui, WANG Ziqi, LI Yang, WANG Jiabao and CHEN Yao

Command and Control Engineering College, Army Engineering University of PLA, Nanjing 210007, China

Abstract Aiming at the problem of the lack of labeled samples in SAR image classification, this paper proposes a task-aware few-shot SAR image classification method based on multi-scale attention mechanism. In order to fully mine local features and focus on the key local semantic patches under specific tasks, this paper introduces two effective attention mechanisms to obtain more efficient and rich feature representation. First, in the feature extraction stage, the complemented squeeze-and-excitation attention block(CSE Block) is used to focus on the salient features of different semantic parts of the original features. It can extract secondary salient features from the suppressed features and merge them with the main salient features, which can obtain more efficient and rich feature representation. Subsequently, an adaptive episodic attention block(AEA Block) is used to obtain key semantic patches in the entire task, which can enhance the differentiated information between tasks and improve the accuracy of SAR image classification tasks. The results show that the classification accuracy of the 5-way 1-shot task is 2.9% higher than that of the sub-optimal task on the SAR image classification standard MSTAR dataset. In the two tasks, the runtime of the proposed method is the same as other metric-learning methods, without additional excessive computing resources, which verifies its effectiveness.

Keywords Multi-scale attention mechanism, Few-shot learning, SAR image classification, Metric learning

1 引言

合成孔径雷达(Synthetic Aperture Radar, SAR)是一种主动式对地观测系统,可安装在飞机、卫星等飞行平台上全天时、全天候对地实施观测,并具有一定的地表穿透能力^[1]。因此,经过几十年的发展,SAR 已经成为遥感技术中的一种重要手段,在军事和民用领域均得到了广泛应用^[2]。图像分类是计算机视觉领域一项重要的研究课题,随着深度学习的快速发展,尤其是卷积神经网络(CNN)的提出,使得图像分类

任务的精度取得了显著的提升。然而,CNN 需要借助大量的带标注的数据来获得丰富的图像表征,从而提升图像识别、图像分割等下游任务的性能。但现有带标注的 SAR 图像样本很少,通过收集并标注的方式扩充数据集又需要大量的专家知识,成本高昂^[3]。因此,在 SAR 图像分类任务中难以获得大量带标注的样本来训练网络模型,这就就会产生模型泛化能力不足和过拟合等问题,导致在识别只有少量样本的新类时出现精度坍塌的现象^[4]。如何在小样本困境下处理好 SAR 图像分类问题,从而提升 SAR 图像

到稿日期:2023-05-25 返修日期:2023-09-13

基金项目:江苏省自然科学基金(BK20200581)

This work was supported by the Natural Science Foundation of Jiangsu Province, China(BK20200581).

通信作者:李阳(solarleon@outlook.com)

的实用价值,成为了亟待研究的课题。

与传统的分类方法中每个待识别目标需要数百个标记样本不同,小样本学习(Few-shot Learning, FSL)旨在解决仅使用少数标记样本完成图像分类的问题^[5]。目前小样本学习在图像分类领域已有很多成果,但大多数方法均针对可见光图像分类问题进行研究。当迁移到 SAR 图像领域时,面临着难以适配 SAR 图像特性、特征表示不够高效以及难以提取到关键特征等问题,分类精度明显下降。如图 1 所示,左侧柱状图表示小样本学习方法在可见光数据集 miniImageNet 上的分类精度,右侧柱状图表示相应方法在 SAR 图像分类数据集 MSTAR 上的分类精度。从图中可以看出,这些方法迁移到 SAR 图像分类领域时,分类精度出现了较大程度的下降。现阶段已有学者开始关注小样本 SAR 图像分类问题^[6],但是这些方法存在使用的数据集不统一以及无法验证其泛化性的问题^[7]。此外,现有小样本 SAR 图像分类方法主要利用先验知识去适配数据,旨在设计领域知识引导的小样本学习方法^[8],而不是提升分类方法对 SAR 图像特征的提取能力。

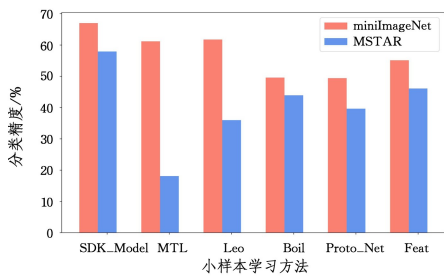


图 1 小样本学习方法在不同数据集上的分类精度

Fig. 1 Classification accuracy of few-shot learning methods on different datasets

为解决以上问题,本文提出了一种任务感知的多尺度小样本 SAR 图像分类方法。该方法通过引入多尺度的注意力机制获得更加高效且丰富的特征表示,引入自适应的情景训练,不需要人为地加入领域知识,有效改善了现有小样本方法不适配 SAR 图像的问题。并且分析后续实验结果发现,仅 ATL_Net^[9]与本文方法在小样本 SAR 图像的 5-way 1shot 任务中分类精度超过 70%,即自适应的情景训练可适配 SAR 图像特性,提升小样本 SAR 图像的分类精度。

针对 SAR 图像在分类时存在的带标注样本较少的问题,本文提出了一种任务感知的多尺度小样本 SAR 图像分类方法。本文的主要贡献在于:

1) 提出了一种任务感知的多尺度小样本 SAR 图像分类方法,该方法能够针对 SAR 图像特点自适应地发现关键特征,提升小样本 SAR 图像的分类性能。

2) 设计了一种多尺度注意力机制,该机制包含互补注意力模块(Complemented Squeeze and Excitation Block, CSE Block)和自适应情景注意力模块(Adaptive Episodic Attention Block, AEA Block)。两者构成的多尺度注意力机制不仅能够发现图像样本的关键特征,还可以自适应感知 SAR 图像分类任务的特点。

3) 在 SAR 图像分类标准数据集 MSTAR 上与 16 种小样本学习方法对比。在 5-way 1-shot 任务上,本文方法的分类精度相较于没有自适应情景训练的最优方法 DN4 提高了 7.56%,相较于有自适应情景训练的最优方法 ATL_Net 提高了 2.9%。

2 相关工作

2.1 基于度量学习的小样本学习方法

小样本学习指在训练样本较少的情况下完成对目标的学习任务。2003 年斯坦福的 Li 等首次提出了 One-shot Learning 的概念,并希望采用贝叶斯框架对单样本学习问题予以解决^[10]。而随着深度学习的兴起,多种行之有效的基于深度学习的小样本图像分类方法被提出^[11]。目前解决小样本学习图像分类问题的方法主要包括:基于微调的方法、基于元学习的方法和基于度量学习的方法^[12]。其中,基于度量学习的小样本学习方法是当前的研究热点。

基于度量学习的小样本学习方法是一种端到端的训练方法。相较于元学习方法,度量学习方法具有不需要在预训练模型上再次微调以达到拟合条件的优点,因此可以快速适应新的任务^[13]。基于度量的小样本学习方法的基本思路是:利用每个类别的数据,学习一个特征嵌入空间,使得在此特征嵌入空间中模型能更有效地度量样本之间的相似度^[14]。Sung 等^[15]于 2018 年提出的关系网络(Relation Network, RN)是一种经典的度量学习方法,该方法在匹配网络的基础上对度量网络的结构进行了简化。RN 分为两部分,分别为特征嵌入模块和关系模块。特征嵌入模块用于提取样本特征;关系模块采用直接拼接的方式融合测试集样本与支持集样本特征,关系分数最大的融合特征为预测结果。Li 等^[16]在 2019 年提出深度最近邻神经网络(Deep Nearest Neighbor Neural Network, DN4),该方法通过对卷积特征图的深层局部描述符进行最近邻搜索,并利用基于局部描述符的图像到类的度量方法,替换原有的图像级特征的度量方法。Dong 等^[9]于 2020 年基于 DN4 提出了一种自适应任务感知局部表示网络 ATL_Net,该方法通过引入 AEA Block,自适应地感知整个任务中重要的局部 patches 来模拟人类识别过程,进一步挖掘了局部表示(Local Representations, LRs)的潜力,在多个小样本学习任务中取得了良好的性能。

然而,大部分基于度量学习的小样本学习方法未能很好地适配 SAR 图像特性,在特征提取时忽视了被抑制特征中有用的次级显著特征,导致在小样本 SAR 图像分类任务中未能取得较好的分类精度。本文利用 CSE Block 有效地提取到了被抑制特征中的次级显著特征,获得了更加高效的特征表示^[17]。

2.2 小样本 SAR 图像分类方法

传统 SAR 图像分类方法通常依靠专家知识设计不同的特征提取器,再进行调参和优化。该类方法的局限性在于成本昂贵、鲁棒性差且难以泛化^[18]。为解决训练时带标注样本不足的问题,基于深度学习的小样本学习方法成为 SAR 图像分类的研究热点^[19]。

小样本 SAR 图像分类研究兴起于 2018 年。Liu 等^[6]设计了一种基于辅助样本训练的光偏振法原型网络,该方法提升了标记图像严重受限时雷达目标识别的精度。Sun 等^[7]在 2022 年针对雷达目标图像的散射特性提出了散射特性分析网络 (Scattering Characteristics Analysis Network, SCAN)。该网络设计了一个散射提取模块,将目标成像机制与网络相结合,通过显式监控来学习每种目标类型散射点的数量和分布。同时,SCAN 考虑到 SAR 目标成像的变异性,设计了由角度自适应分类器和频率嵌入式模块组成,由目标姿态角引导的元学习网络。角度自适应分类器引导网络关注具有不同目标姿态角的正样本对,频率嵌入式模块将脉冲余弦变换与网络训练过程有效地结合在一起,以丰富频域信息。该网络在自建 SAR 飞机数据集上取得了良好的分类效果。Zhang 等^[8]在 2022 年针对 SAR 目标车辆识别问题,提出了以领域知识为驱动的双流深度网络 (Domain Knowledge Powered Two-Stream Deep Network, DKTS-N)。该网络结合了车辆方位角、振幅和相位数据有关的 SAR 领域知识,成为了小样本 SAR 车辆目标识别的开创性工作。DKTS-N 通过提取整个图像和图像 patches 的特征,有效地利用了 SAR 领域知识。同时,为了测量车辆的全局特征和局部特征之间的结构信息距离,DKTS-N 改进了推土距离 (Earth Mover's Distances, EMD),实现了双流深度网络的特征度量。然后,考虑到 SAR 车辆识别中方位角的敏感性,DKTS-N 用最近邻分类器代替结构化的全连接层进行分类,解决了不同方位角下同一目标的 SAR 图像存在差异的难题。该方法充分利用 SAR 车辆域知识的概念(两栖步战车具有防浪板、不同型号的主战坦克炮管大小具有差异等),用于辅助识别任务,取得了较好的识别效果。

然而,现有的小样本 SAR 图像分类方法偏重于利用领域先验知识,仅在特定的数据集上可以取得较好的分类性能。此外,现有方法未采用统一的数据集和实验设置,使得不同方法的泛化性能难以判别。本文利用 AEA Block 实现了自适应的任务感知,未使用领域知识引导,而是根据具体任务自适应地感知关键的语义 patches,有效提升了小样本 SAR 图像分类的精度。

2.3 多尺度注意力机制

注意力机制的设计源自对人类视觉的研究。通过注意力机制,人类可以有选择地关注到目标物体的部分信息,同时忽略其他信息。多尺度注意力机制 (Multi-scale Attention) 的目的是在面对给定任务时,在模型的不同阶段应用不同的注意力模块完成更高效且更丰富的图像特征提取^[20]。Fan 等^[21]于 2020 年通过引入自注意力机制,自适应地融合局部特征及其全局依赖关系,提出了一种多尺度注意力网络 MA-Net。该方法在肝脏和肿瘤图像分割任务中取得了良好的效果。Zhang 等^[22]于 2022 年针对视网膜水肿中最特殊的生物标志物硬性渗出物 (Hard Exudates) 图像分割任务,提出了一个多尺度注意力融合模块。该模块可以有效地为微小病变和边界检测提供有用的细节特征,并在两个公共病变数据集 IDRiD 和 EOphtha 上取得了良好的分割效果。在 EOphtha 数据集

上,与已有方法相比,该方法的分割率和召回率提高均超过了 3%。Chen 等^[23]于 2022 年针对手势识别任务,提出了一种融合多模态数据的多尺度注意力三维卷积网络。该网络实现了局部和全局的注意力机制,其中局部注意力机制利用手部检测器提取的手部信息来关注手部区域,减少了手势无关因素的干扰,有效提升了手势识别任务的精度。

本文受多尺度注意力机制的启发,设计了一种任务感知的多尺度小样本 SAR 图像分类方法。该方法能够针对 SAR 图像特点自适应地发现关键特征,提升小样本 SAR 图像的分类性能。

3 本文方法

3.1 问题定义

在小样本 SAR 图像分类任务中,设 S 表示一个支持集 (Support Set), Q 表示一个查询集 (Query Set), 均为 MSTAR 数据集中的 SAR 图像样本。在支持集 S 中应该包含 N 个不同的 SAR 图像类,每个类中含有 K 个带标记的 SAR 图像样本。小样本 SAR 图像分类的目标是根据支持集 S 对查询集 Q 中的每个未标记样本进行分类,此时称之为一次 N -way K -shot 任务^[24]。但是当支持集 S 中的每个类只有少量的样本时,模型将难以有效地学习用于 Q 中样本分类的知识。为了解决这个问题,小样本学习常常借助包含大量的类和标记样本的辅助集 A 来学习可迁移的知识,以改进在查询集 Q 上的分类,但是辅助集 A 相对于支持集 S 应具有不相交的类标记空间^[25]。具体地,在训练阶段的每次迭代中,通过模拟小样本学习任务来构造 episode 以训练分类模型,在训练过程中会构造上万个 episode 来训练分类模型,即 episode 训练。在测试阶段,利用支持集 S ,学习的模型可以直接用于对查询集 Q 中的每个 SAR 图像进行分类。

本文使用多尺度注意力机制,在浅层网络中使用 CSE Block 提取更加高效和丰富的特征表示。随后利用 AEA Block 对输出的特征图提取更加关键的语义 patches,从而改善小样本学习方法,适应小样本 SAR 图像分类任务。

3.2 任务感知的多尺度小样本 SAR 图像分类方法

本文针对小样本 SAR 图像分类问题提出了任务感知的多尺度小样本 SAR 图像分类方法,方法框架如图 2 所示。本文方法选用常见的 Conv64F 作为骨干网,其由 4 个卷积模块组成(每个卷积块包含一个卷积层、批归一化和 LeakyReLU 激活)。在此基础上加入了 CSE Block 与 AEA Block,形成多尺度的特征表示增强。选用 CNN 而不是近来在多项视觉任务中取得优异表现的 Vision Transformer (ViT) 作为骨干网,是因为 CNN 通过卷积核移动固定步长来提取特征,相较于 Transformer 可以更好地提取到图像的局部特征,从而挖掘更加丰富的局部表示。 F_{φ} 表示加入了 CSE Block 后的嵌入模块,提取出带有次级显著特征的局部特征图,且 $F_{\varphi}(x) \in \mathbb{R}^{C \times H \times W}$,即可以对每张输入图像求得 HWC 维的局部表示,进而求得总共 $NKHWC$ 维的支持集局部表示,并计算局部关系矩阵 \mathbf{M}^R ,计算公式为:

$$\mathbf{M}_{i,j}^R = g(L_i^q, L_j^s) \quad (1)$$

其中, $i \in \{1, \dots, HW\}$, $j \in \{1, \dots, NKHW\}$, L^q 表示查询集样本特征, L^s 表示支持集图像样本特征, $g(\cdot, \cdot)$ 表示相似性度量。本文方法采用余弦相似度进行相似性度量。

AEA Block 生成自适应的情景注意力, 用于为特定的任务选择具有区分性的关键 patches, 本文利用 AEA Block 生成情景注意力矩阵 M^A 。随后通过逐元素乘法将 M^A 应用到关系矩阵 M^R 上以消除噪声, 然后从处理后的关系矩阵中得到

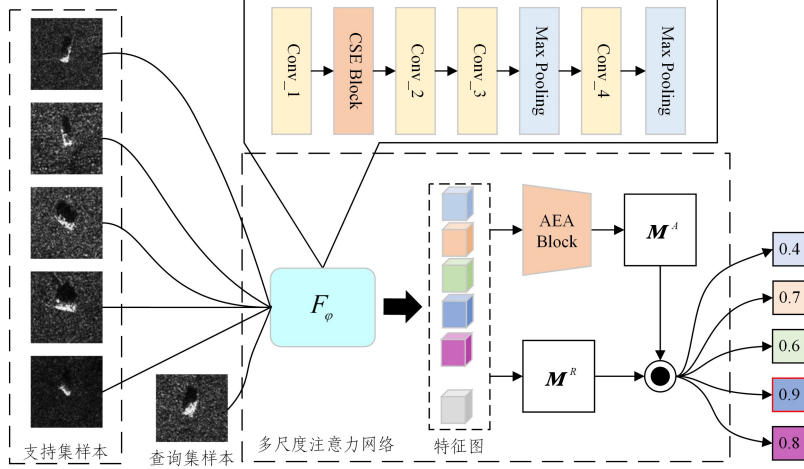


图2 本文方法框架

Fig. 2 Framework of our method

3.3 多尺度注意力机制

本文设计了一种多尺度注意力机制, 该模块包含 CSE Block 和 AEA Block。CSE Block 既可以关注目标样本主要信息, 从中提取显著的局部特征, 又可以关注被抑制的剩余通道信息, 从中提取次级显著特征, 从而充分提取目标中有效的显著特征; AEA Block 可以为特定的任务自适应地选择其中关键的语义 patches 作为分类的依据, 利用任务之间的共享参数建立起任务间的关联关系, 增强任务间的区分信息, 从而提升具体任务的分类精度。两者构成的多尺度注意力机制不仅能够发现关键特征, 还可自适应适配 SAR 图像特点。

3.3.1 互补注意力模块(CSE Block)

CSE Block 基于 SE 注意力机制^[26] 设计, 包括两个子模块, 即主要显著特征提取模块和次级显著特征提取模块。如图 3 所示, 本文用 x , S_m , S_c 和 S_s 表示 CSE Block 的输入特征张量、主要显著特征张量、次级显著特征张量和被抑制的特征张量, 用 m 和 n 分别表示通道权重向量和全为 1 的向量。利用激励-压缩模块, 主要特征提取模块首先通过压缩特征 U 获得通道描述符 r , 即:

$$r = [r_1, \dots, r_c] \quad (3)$$

$$r_c = \frac{1}{WH} \sum_{w=1}^W \sum_{h=1}^H u_c(w, h)$$

然后对通道描述符 r 进行激励, 得到对应的权重向量 m 。

然后利用求得的权重向量 m 计算得到主要显著特征 S_m , 即:

$$S_m = [m_1 u_1, \dots, m_c u_c] \quad (4)$$

其中, S_m 表示模块输出的主要显著特征张量, u_c 表示特征 U 的第 c 个通道。

最终的预测得分, 并将查询图像归类到预测得分最高的类别当中, 预测得分计算式为:

$$Score_a = \frac{V_s}{HW} \sum_{i=1}^{HW} \sum_{j=Z_i^s}^{Z_i^{kHW}} (M^A \odot M^R)_{i,j} \quad (2)$$

其中, V_s 表示交叉熵损失的温度, Z_i^s 表示支持集 S 的第 n 类中的第 k 个 HW 维特征关系。最终, 我们将查询样本 q 划分到预测得分最高的类别中, 完成一次分类任务。

同时, S_m 和 S_s 是关于特征 U 的互补特征, 即:

$$S_m + S_s = U \quad (5)$$

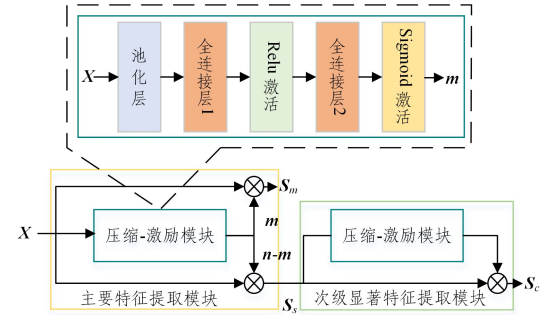


图3 互补注意力模块

Fig. 3 Schematic of CSE block

与提取主要显著特征 S_m 类似, 通过激励-压缩模块可以从被抑制的特征 S_s 中寻找可以正确区分目标样本的有效特征, 提取方式与主要显著特征提取基本相同, 即通过提取被抑制特征 S_s 的“主要显著特征”得到次级显著特征 S_c 。经过上述两次特征提取工作, CSE Block 同时提取了原样本中的主要显著特征 $S_m \in \mathbb{R}^{HWC}$ 和次级显著特征 $S_c \in \mathbb{R}^{HWC}$, 并通过对两类特征进行融合得到了目标样本最终的特征表示 $F = F(S_m + S_c)$ 。具体过程如算法 1 所示。

算法 1 互补注意力机制

输入: 输入特征张量 x

输出: 互补注意力融合特征 F

1. while x do
2. 通过压缩 x 和特征 U 获得通道描述符 r
3. 利用式(3)计算

$$\mathbf{r} = [\mathbf{r}_1, \dots, \mathbf{r}_c], \mathbf{r}_c = \frac{1}{WH} \sum_{w=1}^W \sum_{h=1}^H \mathbf{u}_c(\mathbf{w}, \mathbf{h})$$

4. 激励通道描述符 \mathbf{r} 获得对应权重向量 \mathbf{m}
5. 计算主要显著特征向量 \mathbf{S}_m
6. 利用式(4)计算 $\mathbf{S}_m = [\mathbf{m}_1 \mathbf{u}_1, \dots, \mathbf{m}_c \mathbf{u}_c] \in \mathbb{R}^{HWC}$
7. 同理计算次级显著特征 $\mathbf{S}_c \in \mathbb{R}^{HWC}$
8. 计算互补注意力融合特征 $\mathbf{F} = \mathbf{F}(\mathbf{S}_m + \mathbf{S}_c)$
9. end while

本文中所使用的 CSE Block 是针对任务层级的注意力机制。CSE Block 会根据不同的分类任务提取更加丰富的特征表示,如在 MSTAR 数据集中,CSE Block 会引导模型在关注坦克的炮管特征(主要显著特征)的同时,在其他被抑制的特征中发现履带作为次级显著特征;在 SAR 飞机数据集中,CSE Block 会引导模型在关注飞机的机翼特征(主要显著特征)的同时,在其他被抑制的特征中发现起落架等作为次级显著特征。该方法通过提取主要显著特征和次级显著特征作为互补特征用于指导最后的分类任务,从而有效提升小样本 SAR 图像分类模型的性能。

3.3.2 自适应情景注意力模块(AEA Block)

AEA Block 是以任务为导向设计的,该模块通过情景训练和学习一个可变的阈值来选择信息最丰富的关键语义 patches。如图 4 所示,该模块输入为情景任务 $T = \{A_s, A_Q\}$,输出为情景注意力关系矩阵 \mathbf{M}^A 。输入的支持集和查询集样本特征图分别通过共享参数的两个 1×1 卷积块得到两者间的关系矩阵。同时,AEA Block 使用多层感知机(MLP)来自适应地预测每个查询集图像的阈值 V_c ,该部分可以针对不同的 patches 学习不同的阈值 V_c ,用于选择信息最丰富且关键的语义信息。 V_c 的计算式为:

$$V_c = \sigma(F_r(L_q^q)) \quad (6)$$

其中, $\sigma(\cdot)$ 表示 sigmoid 激活函数; F_r 以每张查询图像 q 的局部表示作为输入,利用多层感知机(Multi-Layer Perceptron, MLP)自适应预测 q 的阈值 V_c 。

最后将前面得到的关系矩阵乘以对应的阈值 V_c ,得到情景注意力关系矩阵 \mathbf{M}^A ,其计算式为:

$$\mathbf{M}_{i,j}^A = \frac{I(\mathbf{M}_{i,j}^q)}{\sum_j I(\mathbf{M}_{i,j}^q)} \quad (7)$$

$$I(x) = x / (1 + \exp^{-k(x-v_c)})$$

其中, $\mathbf{M}_{i,j}^q$ 表示在原始局部表示上利用两个 1×1 的卷积层进行特征变换后,学习到的一个特征矩阵; $I(x)$ 是一个超参数 k 的 sigmoid 激活函数变体。具体过程如算法 2 所示。

算法 2 自适应情景注意力机制

输入:情景任务 $T = \{A_s, A_Q\}$

输出:情景注意力关系矩阵 \mathbf{M}^A

1. while 不收敛 do
2. 求局部特征 $L^S \leftarrow F_0(A_s)$
3. 求局部特征 $L^Q \leftarrow F_0(A_Q)$
4. for L^q in L^Q do
5. 通过式(1)获得关系矩阵 \mathbf{M}^R
6. 通过式(6)计算 L^q 的自适应阈值 V_c
7. 通过式(7)构建情景自适应注意力关系矩阵 \mathbf{M}^A

8. 计算相似性得分 Score,从而得到 L^q 的概率 P
9. end for
10. 更新损失函数 $L \leftarrow \sum y \log(P)$
11. 更新模型相关参数
12. end while

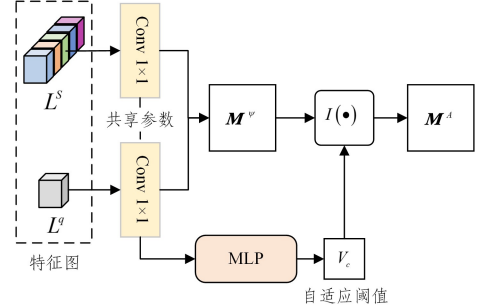


图 4 自适应情景注意力模块

Fig. 4 Schematic of AEA block

也即,该模块通过两个矩阵乘法、一个元素乘法,以及一些卷积运算很容易地实现了对关键语义信息的输出。将得到的情景注意力关系矩阵 \mathbf{M}^A 与前文提到的局部关系矩阵 \mathbf{M}^R 进行元素乘法,得到加权关系矩阵后送入 softmax 函数,通过式(2)得到最终的相似性分数。

4 实验结果与分析

4.1 实验设置

本文使用 SAR 图像分类标准数据集 MSTAR^[27] 作为实验数据集。MSTAR 是一个合成孔径雷达图像数据集,常用于 SAR 图像分类任务。该数据集由 10 个类组成,每个类有 200 多幅图像,每幅图像的大小为 84×84 。在实验时,我们将 MSTAR 划分为训练集和测试集,两者各包含 5 类样本,且互不相同。

本文实验基于 Pytorch 框架,在一台配有一块 NVIDIA GTX 1080 的内存为 8GB 的显卡上完成。在情景训练时,从每个 episode 中选择 15 张查询图像,例如在 5-way 1-shot 设置条件下,每个任务包含 5 张支持集图像和 75 张查询集图像。整个模型训练过程中,未使用数据增强与权重衰减,也没有在测试阶段进行微调。本文方法选用批处理大小为 128 的 Conv64F 作为骨干网络,选用 Adam 优化器^[28] 和交叉熵损失来训练网络,初始学习率设置为 0.001,整个训练过程迭代 50 个 epoch。在目前主流的关于小样本学习的文献中,均选用 5-way 1-shot 和 5-way 5-shot 任务作为通用的测评标准。为了公平比较本文方法与其他模型的性能,我们选取 N 为 5、 K 为 1 和 5 进行实验。

4.2 对比分析

我们从分类精度与运行时间两方面综合比较了本文模型与对比模型的性能。表 1 列出了本文模型与 17 种小样本 SAR 图像分类模型的精度和运行时间的对比结果。可以看到,本文方法在 5-way 1-shot 任务中取得了最佳分类精度,相较于次优方法 ATL_Net 提升了将近 3%,在 5-way 5-shot 任务中也取得了次优分类精度,与最优精度非常接近。我们认为,现有的基于微调的方法分类精度低于本文方法的原因

在于,该类方法在可见光数据集上预训练得到的初始参数,跨越在 SAR 图像上微调时,没有很好地利用领域知识。而基于元学习的方法分类精度较差的原因在于,现有带标注的 SAR 图像数据集太小,元学习器未能学到足够的先验知识便过拟合,因此基于元学习的方法迁移到 SAR 图像领域未取得较好的分类精度。

在对运行时间进行比较时,由于所有方法在训练时均迭代 50 个 epoch,因此本文取每一轮训练的平均时间进行比较

较^[42]。在基于度量学习的方法中,本文方法与其他 17 种方法的运行时间相当,没有太大差异。基于微调的方法中,分类精度最佳的 Baseline++ 与本文方法接近。但是由于该类方法需要通过学习新的知识来微调模型参数,而此过程需要耗费很长的训练时间,因此该类方法运行时间普遍较长^[43]。而基于元学习的方法由于带标注的 SAR 图像训练样本过少,因此元学习器在训练阶段未能学习到足够的知识,从而导致训练时间相应减少。

表 1 实验结果比较

Table 1 Comparison of experimental results

方法	方法来源	方法类型	骨干网	分类精度				运行时间/min	
				5-way		5-way		5-way	
				1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Baseline* ^[29]	ICLR 2019	模型微调	Conv64F	54.44	83.13	0.90	2.18		
Baseline++* ^[29]	ICLR 2019	模型微调	Conv64F	59.98	86.37	5.85	15.27		
SKD_Model* ^[30]	BMVC 2021	模型微调	Conv64F	57.92	78.39	0.48	0.52		
MAML* ^[31]	ICML 2017	元学习	Conv64F	19.87	60.67	0.10	0.18		
Versa* ^[32]	ICLR 2019	元学习	Conv64F	66.96	68.01	0.72	0.76		
R2D2* ^[33]	ICLR 2019	元学习	Conv64F	63.99	68.88	0.58	0.68		
MTL* ^[34]	CVPR 2019	元学习	Conv64F	18.13	47.07	0.08	0.10		
Leo* ^[35]	ICLR 2019	元学习	Conv64F	36.00	44.00	0.08	0.10		
ANIL* ^[36]	ICLR 2020	元学习	Conv64F	20.99	61.91	0.62	0.84		
Boil* ^[37]	ICLR2021	元学习	Conv64F	43.94	54.50	0.38	0.52		
Proto_Net* ^[38]	NIPS 2017	度量学习	Conv64F	39.68	42.72	0.46	0.52		
Feat* ^[39]	CVPR 2020	度量学习	Conv64F	46.11	56.36	0.60	0.74		
Relation Net* ^[15]	CVPR 2018	度量学习	Conv64F	64.84	77.51	0.64	0.78		
DN4* ^[16]	CVPR 2019	度量学习	Conv64F	67.37	85.15	0.64	0.80		
CovaMNet* ^[40]	AAAI 2019	度量学习	Conv64F	58.75	45.75	0.62	0.76		
ATL_Net* ^[9]	IJCAI 2020	度量学习	Conv64F	<u>72.03</u>	88.81	0.63	0.76		
门控多尺度匹配网络 ^[41]	系统工程与电子技术 2022	度量学习	Conv64F	70.50	84.20	—	—		
本文方法		度量学习	Conv64F	74.93	<u>88.57</u>	0.64	0.80		

注:表中加粗字体表示各列最优结果,下划线表示次优结果,带*的小样本学习方法精度为在数据集 MSTAR 上重测的结果。

4.3 消融实验

表 2 列出了本文方法消融实验的结果。在加入由 CSE Block 与 AEA block 两种注意力机制构成的多尺度注意力机制后,模型在 5-way 1-shot 任务中分类精度提升了 2.9%,在 5-way 5-shot 任务中精度几乎与最佳精度持平。在删除 CSE Block 后,在 5-way 1-shot 任务中分类精度出现了明显的下降;在删除 AEA Block 后,5-way 1-shot 和 5-way 5-shot 任务的分类精度均出现明显下降,在 5-way 1-shot 任务中下降超过 6%;而同时删除 CSE Block 和 AEA Block 后,5-way 1-shot 和 5-way 5-shot 任务分类精度在 4 种情况中处于最低。由此可见,多尺度注意力机制对于小样本 SAR 图像分类起到了良好的提升效果。

表 2 消融实验结果

Table 2 Results of ablation experiment

方法	分类精度		运行时间	
	5-way		5-way	
	1-shot	5-shot	1-shot	5-shot
本文方法	74.93	88.57	0.64	0.80
w/o CSE Block	72.03	88.81	0.63	0.76
w/o AEA Block	68.47	86.17	0.62	0.82
w/o CSE&AEA Block	67.37	85.15	0.62	0.76

而在删除 CSE Block 后,模型在 5-way 5-shot 任务上的

精度反而提升了 0.24%,可以推测随着单次任务中 shot 数量的增加,CSE Block 在对特征进行拼接时引入了噪声,但对最终的分精度未形成大的干扰。

4.4 讨论

如图 5 所示,在固定 5-way 设置且匹配训练的条件下,分别测试了本文方法与 ATL_Net 从 1-shot 到 5-shot 设置下的分类精度。可以发现,当 shot 数小于 4 时,本文方法的精度均优于 ATL_Net,但随着 shot 数的递增,尤其从 1-shot 到 3-shot,本文方法的优势逐渐变小。

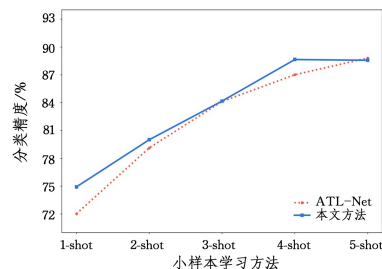


图 5 5-way 匹配训练条件下不同 shot 设置下的分类精度

Fig. 5 Classification accuracy of different shot settings under 5-way matching training condition

由此可推测,CSE Block 在 shot 数较少时很好地利用了从被抑制的特征中提取到的次级显著特征,从而提升了在

训练样本极少时的分类精度。随着训练阶段 shot 数量的增加,较多的样本很好地支持了情景训练^[44],从而减少了对次级显著特征的依赖,并且过多样本的次级显著特征反而可能引入了噪声,导致本文方法的分类精度下降。

结束语 针对小样本 SAR 图像分类问题,本文提出了一种任务感知的多尺度小样本 SAR 图像分类方法。通过设计的多尺度注意力机制,本文方法可以获得更高效且关键的特征表示。实验证明,本文方法在 SAR 图像分类标准数据集 MSTAR 上取得了良好的效果,尤其是在 5-way 1-shot 任务中有效地提升了小样本 SAR 图像的分类精度。下一步工作可以进一步探究标签噪声对小样本学习的影响,设计具有噪声感知能力的鲁棒小样本学习方法。

参 考 文 献

- [1] LI Y, WANG J B, XU Y L, et al. DeepSAR-Net: Deep convolutional neural networks for SAR target recognition[C]// 2017 IEEE 2nd International Conference on Big Data Analysis(ICBDA). 2017:740-743.
- [2] GUO W W, ZHANG Z H, YU W X, et al. Perspective on explainable SAR target recognition[J]. Journal of Radars, 2020, 9(3):462-476.
- [3] SU S H, CUI Z T, GUO W W, et al. Explainable analysis of deep learning methods for SAR image classification[C]// IEEE International Geoscience and Remote Sensing Symposium (IGARSS). 2022.
- [4] GE Y Z, LIU H, WANG Y, et al. Survey on deep learning image recognition in dilemma of small samples[J]. Journal of Software, 2022, 33(1):193-210.
- [5] LIU Y, LEI Y B, FAN J L, et al. Survey on image classification technology based on small sample learning[J]. Acta Automatica Sinica, 2021, 47(2):297-315.
- [6] LIU D Y, GAO X Z, SHEN Q M. Prototypical network for radar image recognition with few samples[C]// Journal of Physics: Conference Series. IOP Publishing, 2020.
- [7] SUN X, LV Y X, WANG Z R, et al. SCAN: Scattering characteristics analysis network for few-shot aircraft classification in high-resolution SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60:1-17.
- [8] ZHANG L B, LENG X G, FENG S J, et al. Domain knowledge powered two-stream deep network for few-shot SAR vehicle recognition[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60:1-15.
- [9] DONG C Q, LI W B, HUO J, et al. Learning task-aware local representations for few-shot learning[C]// Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence. 2021:716-722.
- [10] LI F F, FERGUS R, PERONA P. A Bayesian approach to unsupervised one-shot learning of object categories[C]// Proceedings of the IEEE International Conference on Computer Vision. 2003:1134-1141.
- [11] PENG Y C, QIN X L, ZHANG L G, et al. Survey on Few-shot Learning Algorithms for Image Classification[J]. Computer Science, 2022, 49(5):1-9.
- [12] ZHAO K L, JIN X L, WANG Y Z. Survey on few-shot learning[J]. Journal of Software, 2021, 32(2):349-369.
- [13] ZHANG R, YANG Y X, LI Y, et al. Multi-task few-shot learning with composed data augmentation for image classification[J]. IET Computer Vision, 2022, 17(2) 211-221.
- [14] LIU X, ZHOU K R, HE Y L, et al. Survey of metric-based few-shot classification[J]. Pattern Recognition and Artificial Intelligence, 2021, 34(10):909-923.
- [15] SUNG F, YANG Y X, ZHANG L, et al. Learning to compare: Relation network for few-shot learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018:1199-1208.
- [16] LI W B, WANG L, XU J L, et al. Revisiting local descriptor based image-to-class measure for few-shot learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:7260-7268.
- [17] ZHAO X, WANG J B, LI Y, et al. Complemented attention method for fine-grained image classification[J]. Journal of Image and Graphics, 2021, 26(12):2860-2869.
- [18] ZHAO P F, HUANG L J, XIN Y, et al. Multi-aspect SAR target recognition based on prototypical network with a small number of training samples[J]. Sensors, 2021, 21(13):4333.
- [19] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553):436-444.
- [20] AL-ANTARY M T, ARAFA Y. Multi-scale attention network for diabetic retinopathy classification[J]. IEEE Access, 2021, 9:54190-54200.
- [21] FAN T L, WANG G L, LI Y, et al. Ma-net: A multi-scale attention network for liver and tumor segmentation[J]. IEEE Access, 2020, 8:179656-179665.
- [22] ZHANG J Y, CHEN X S, QIU Z X, et al. Hard exudate segmentation supplemented by super-resolution with multi-scale attention fusion module[C]// 2022 IEEE International Conference on Bioinformatics and Biomedicine(BIBM). 2022:1375-1380.
- [23] CHEN H Z, LI Y N, FANG H J, et al. Multi-scale attention 3D convolutional network for multimodal gesture recognition[J]. Sensors, 2022, 22(6):2405.
- [24] WANG Y Q, YAO Q M, KWOK J T, et al. Generalizing from a few examples: A survey on few-shot learning[J]. ACM Computing Surveys, 2020, 53(3):1-34.
- [25] JIAN Y R, TORRESANI L. Label hallucination for few-shot classification[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2022:7005-7014.
- [26] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018:7132-7141.
- [27] WANG H P, CHEN S Z, XU F, et al. Application of deep-learning algorithms to MSTAR data[C]// 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). 2015:3743-3745.
- [28] KINGMA D P, BA J. Adam: A method for stochastic optimization[C]// International Conference on Learning Representations (ICLR). 2015.

- [29] CHEN W Y, LIU Y C, KIRA Z, et al. A closer look at few-shot classification[C]//International Conference on Learning Representations(ICLR). 2019.
- [30] RAJASEGARAN J, KHAN S, HAYAT M, et al. Self-supervised knowledge distillation for few-shot learning[C]//British Machine Vision Conference(BMVC). 2021.
- [31] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//International Conference on Machine Learning. 2017:1126-1135.
- [32] GORDON J, BRONSKILL J, BAUER M, et al. Meta-learning probabilistic inference for prediction[C]//International Conference on Learning Representations(ICLR). 2019.
- [33] BERTINETTO L, HENRIQUES J F, TORR P H S, et al. Meta-learning with differentiable closed-form solvers[C]//International Conference on Learning Representations(ICLR). 2019.
- [34] SUN Q, LIU Y Y, CHUA T S, et al. Meta-transfer learning for few-shot learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:403-412.
- [35] RUSU A A, RAO D, SYGNOWSKI J, et al. Meta-learning with latent embedding optimization[C]//International Conference on Learning Representations(ICLR). 2019.
- [36] RAGHU A, RAGHU M, BENGIO S, et al. Rapid learning or feature reuse? Towards understanding the effectiveness of maml[C]//International Conference on Learning Representations (ICLR). 2020.
- [37] OH J, YOO H, KIM C H, et al. BOIL: Towards representation change for few-shot learning[C]//International Conference on Learning Representations(ICLR). 2021.
- [38] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning[J]. arXiv:1703.05175, 2017.
- [39] YE H J, HU H, ZHAN D C, et al. Few-shot learning via embedding adaptation with set-to-set functions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:8808-8817.
- [40] LI W B, XU J L, HUO J, et al. Distribution consistency based covariance metric networks for few-shot learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019:8642-8649.
- [41] LIU Q, ZHANG X Y, LIU Y X. Few-shot SAR target recognition method based on gated multi-scale matching network[J]. Systems Engineering and Electronics, 2022, 44(11):3346-3356.
- [42] LI W B, WANG Z Y, YANG X S, et al. LibFewShot: A comprehensive library for few-shot learning[J]. arXiv:2109.04898, 2021.
- [43] ZHANG H Y, ZHANG J, HUANG J. Multi-label Image Classification Model Based on Graph Attention Network[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2022, 39(1):34-41.
- [44] LI X X, AN W J, WU J J, et al. Channel attention bilinear metric network[J]. Journal of Jilin University(Engineering and Technology Edition), 2024, 54(2):524-532.



ZHANG Rui, born in 1977, Ph.D, professor, Ph.D supervisor. His main research interests include data engineering and information fusion.



LI Yang, born in 1984, Ph.D, associate professor, is a senior member of CCF (No. D24215). His main research interests include computer vision, deep learning and image processing.

(责任编辑:何杨)