基于区分显著性的无源域适应辐射源个体识别

韩昌芝[1] 陈健威[1] 王 闯[1] 俞 璐[1]
HAN Changzhi CHEN Jianwei WANG Chuang YU Lu

摘 要 针对无法通过源域数据进行分布对齐和对抗训练，以及利用目标域数据标签训练分类器的问题，深入研究利用目标域数据的结构特性进行域适应，关注无源无监督场景下的通信辐射源个体识别任务。将类内散度小、类间散度大的思想引入到无源无监督域适应问题，提出一种基于区分显著性的无源无监督域适应通信辐射源个体识别方法。利用源域模型生成目标域数据的伪标签，并基于伪标签计算各类质心，进而构造区分显著性约束条件，引导特征提取层向目标域特征类内散度小、类间散度大的方向优化，通过提高目标域特征的可区分性提升迁移效果。在公开数据集及自建数据集上的实验结果均优于对比方法，验证了方法的可行性。

关键词 通信辐射源个体识别；无监督域适应；无源域适应；区分显著性；散度

doi：10.3969/j.issn.1672-9528.2024.11.004

0 引言

随着人工智能技术的快速发展，深度学习技术在通信辐射源个体识别上的应用日益广泛。传统机器学习方法通常假设训练集与测试集数据分布相同，然而，由于不同信号之间的相互影响，空间噪声的干扰，以及信号采集设备和环境的差异，实际得到的测试数据往往与训练数据分布不同，这种现象称为分布偏移。为解决此问题，通常采用迁移学习的方法，将从训练数据中学习到的知识迁移到测试数据中，以缓解分布偏移对模型性能的影响。作为迁移学习方法中的一个重要分支，域适应方法更加关注在相同任务和特征空间条件下的迁移学习问题，在应对分布偏移问题中表现出强大性能。

在一些真实场景中，受数据的隐私性、安全性、存储成本和计算成本等因素的制约，在模型训练过程中无法访问源域数据，这无疑给模型训练带来巨大挑战。对此，很多研究者开始关注无源无监督域适应方法，该方法不访问源域数据，仅使用源域数据预先训练好的模型和目标域无标签数据完成模型预测任务。本研究正是考虑这一场景，提出一种在无源无监督下的通信辐射源个体识别方法。具体而言，本文主要贡献如下。

（1）提出了一种基于区分显著性的无源无监督域适应方法，引入目标域特征的区分显著性约束，通过最小化类内散度和最大化类间散度的方式，提高目标域特征的可区分性，从而提升模型迁移效果，以解决源域和目标域数据分布不同导致的性能降低的问题。

（2）本文在 Oracle 射频指纹数据集和实采射频指纹数据集上均验证了本方法的有效性，并进一步在 Oracle 射频指纹数据集上进行了消融实验，验证算法中各部分的作用。

1 无源域适应研究现状

随着深度学习的不断发展，该技术在计算机视觉、医学数据分析、自然语言处理和辐射源个体识别等领域 [1-4] 中都取得了显著成效。深度神经网络的典型学习过程是将在训练集上学习的模型直接应用于测试集，这种学习过程依赖于训练和测试数据分布相同的基本假设。当训练和测试数据之间存在分布差异时，就会出现域偏移 [5]，导致模型预测性能的大幅下降。为解决此类问题，研究人员提出域适应算法 [6]。其中，无监督域适应旨在将从源域中学习的知识迁移到目标域上，而无需访问目标域标签信息。早期的域适应方法 [7-9] 主要采用矩阵匹配的方式对齐源域和目标域的特征分布。在引入对抗性学习方法后，文献 [10] 以多个源域数据信息进行对抗性训练；文献 [11] 通过对违反聚类假设的部分添加惩罚的正则化项进行对抗性训练。此外，文献 [12] 采用了多个可学习的分类器，通过它们之间的预测多样性，实现源域和目标域之间的局部或类别级特征对齐；文献 [13] 提出了一种简单的自训练策略改进域偏移下的粗糙伪标签问题。

上述域适应方法需要在自适应期间始终访问源域数据。然而，在一些实际场景中，由于数据涉及隐私具有较高安全性要求，或者数据量过大导致存储、传输和计算成本过大等

1. 陆军工程大学 江苏南京 210000

问题，往往难以访问源域数据。很多研究开始关注无源无监督域适应方法，并取得了很大进展。文献 [14] 通过合成额外的训练样本获得紧凑的决策边界，有利开放类目标的检测以及目标适应；文献 [15] 设置了诱饵和锚点，促使源域分类器更好地在目标域数据中发挥作用；文献 [16] 提出了邻域聚类，该聚类加强了局部邻域之间的预测一致性；文献 [17] 提出为硬样本学习额外的目标特定分类器，并采用对比类别匹配模块对目标特征进行聚类。

2 基于区分显著性的辐射源个体识别方法

本文所用模型包括 1 个特征提取器 G 和 1 个分类器 H。在源域模型训练过程中，仅访问源域数据，并使用标准交叉熵损失训练特征提取器和分类器；在目标域模型训练过程中，不再访问源域数据，仅使用源域训练好的模型参数进行初始化，并通过目标域无标签数据进行模型训练，如图 1。具体来说，目标域数据经特征提取器和分类器得到数据样本的软标签分类结果，之后通过最小化类内散度和最大化类间散度对整个模型进行优化。

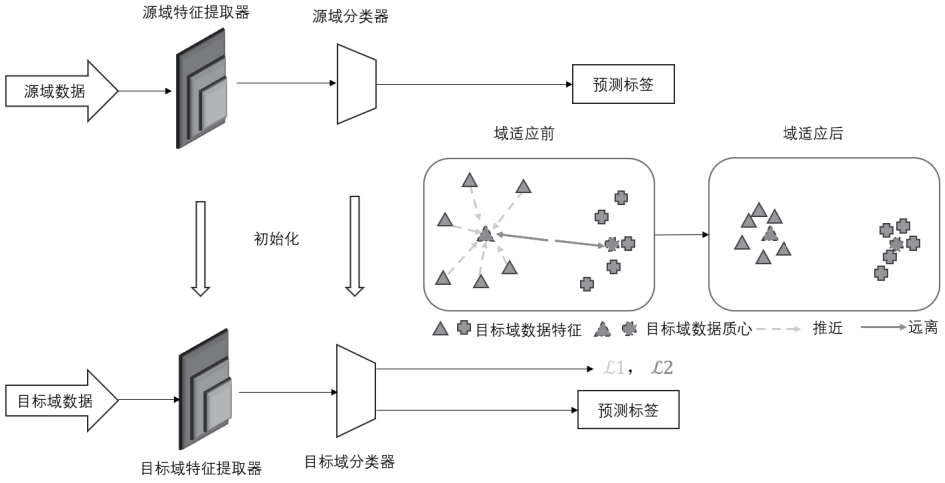

图 1 基于区分显著性的无源无监督域适应模型架构

2.1 源域模型训练

首先给出所提方法的形式化定义。给定一组带标签源域数据样本 $D_s=\{x_s^i,y_s^i\}_{i=1}^{n_s}$ 和一组不带标签目标域数据样本 $D_t=\{x_t^i\}_{i=1}^{n_t}$，其中 x_s^i 为源域第 i 个样本，$y_s^i\in\mathcal{Y}\subseteq\mathbb{R}^K$ 为样本真实标签，K 是标签集 $C=\{1,2,...,K\}$ 的类别数，x_t^i 表示目标域第 i 个样本（不含标签），其具有与 D_s 相同的标签集 C，n_s、n_t 分别为源域和目标域样本数量。在无源无监督域适应中，可以访问源域训练好的模型 $F(G(\cdot))$，其中 G 是 CNN 特征提取器，F 是线性分类器。在训练过程中只有 D_t 中的数据可以使用，而 D_s 中的数据不能使用。本文模型中，源域和目标域任务相同，即 $\mathcal{X}_s\to\mathcal{Y}_s$ 与 $\mathcal{X}_t\to\mathcal{Y}_t$ 相同（$\mathcal{X}_s$、$\mathcal{Y}_s$、$\mathcal{X}_t$、$\mathcal{Y}_t$ 分别表示源域样本空间、源域标签空间、目标域样本空间和目标域标签空间）。无源无监督域适应的目标是利用源域训练好的模型和目标域无标签样本，训练一个目标域模型，以此实现对目标域样本的分类预测。

为实现无源无监督域适应算法，首先使用源域数据训练源域模型（源域数据在源域模型训练完成后将不再被访问）。所提方法通过最小化交叉熵损失训练一个深度神经网络 f_s：$\mathcal{X}_s\to\mathcal{Y}_s$ 作为源域模型。

$$\mathcal{L}_{src}(f_s;\mathcal{X}_s,\mathcal{Y}_s)=-\mathbb{E}_{(x_s,y_s)\in\mathcal{X}_s\times\mathcal{Y}_s}\sum_{k=1}^{K}q_k log\,\delta_k(f_s(x_s)) \tag{1}$$

式中：$\delta_k(a)=\frac{\exp(a_k)}{\Sigma_i\exp(a_i)}$ 表示 K 维向量 $\boldsymbol{a}$ 的 softmax 输出的第 k 个元素，q_k 是源域数据标签的独热编码向量，对于分类正确的维度，q_k 为 1，对于其余的维度，q_k 为 0。

源域模型的深度神经网络由特征提取器 $g_s:\mathcal{X}_s\to\mathbb{R}^d$ 和分类器 h_s: $\mathbb{R}^d\to\mathbb{R}^K$ 两部分组成。即 $f_s(x)=h_s(g_s(x))$，其中 d 表示输入特征的尺寸。已有的域适应方法通过使用最大均值差异 (maximum mean discrepancy，MMD) 或领域对抗性对齐，匹配特征空间 $\mathbb{R}^d$ 中的数据分布来对齐不同的领域。然而，这两类方法都假设源域和目标域共享相同的特征编码器，并且需要在域适应期间访问源域数据，这与无法访问源域数据的前提相悖。因此在源域模型后，将模型的特征提取器 g_s 和分类器 h_s 保存，并用于目标域模型的训练。

2.2 目标域模型训练

在目标域模型训练前，先使用源域上训练好的模型参数对目标域模型进行初始化，将源域模型的特征提取器和分类器的网络参数赋予目标域对应模块，即 $\theta(g_s)=\theta(g_t)$、$\theta(h_s)=\theta(h_t)$。受 SHOT 的启发，首先将目标样本 x_t 送入模型，得到特征提取器的输出 $g_t(x_t)$ 和分类器 h_t 的输出 $h_t(g_t(x_t))$。考虑到 one-hot 形式的硬标签会造成数据原始信息的丢失，例如，来自第一类目标样本的网络输出为 [0.3, 0.4, 0.1, 0.1, 0.1]，其 one-hot 形式的硬标签可能被强制输出为 [0.0, 1.0, 0.0, 0.0, 0.0]，从而损失大量信息。为尽可能多保留样本信息，本文使用 softmax 层的输出 $\delta(h_t(g_t(x_t)))$ 作为目标域数据属于各个类别的概率，以便更好地保留目标域数据原有信息。接下来，计算目标域每一类样本的质心：

$$c_k = \frac{\sum_{x_t \in \mathcal{Z}_t} \delta_k(f_t(x_t)) g_t(x_t)}{\sum_{x_t \in \mathcal{Z}_t} \delta_k(f_t(x_t))} \tag{2}$$

式中：$\delta(h_t(g_t(x_t)))$ 表示目标域样本通过模型输出的 K 维 softmax 向量的第 k 个分量，即 x_t 属于第 k 类的概率。所得质心 c_k 是由特征提取器从目标域样本中提取到的特征表示，可以稳定且可靠地表征目标域不同类别数据的分布。

随后，计算目标域每个样本的特征到质心 c_k 的距离 d:

$$d_k(x_t) \triangleq d(g_t(x_t), c_k) \tag{3}$$

式中：$x_t \in \mathcal{X}_s$，$k = 1, 2, \cdots, K$。常见的距离采用 1- 范数距离、2- 范数距离等，本文采用的是 2- 范数距离。

结合目标域样本在 softmax 层的输出 $\delta_k(h_t(g_t(x_t)))$，得到目标域各样本到各类样本质心的距离总和作为模型的损失函数 $\mathcal{L}_1$。

$$\mathcal{L}_1 = \sum_{k=1}^{K} \sum_{x_t \in \mathcal{X}_t} \delta_k(f_t(x_t)) d_k(x_t) \tag{4}$$

在使目标域样本尽可能靠近质心的同时，还要确保这些质心彼此之间尽可能地远，以此使分类器得到的分类结果更好地接近真实结果。为了实现上述目的，计算目标域各类样本质心之间的距离之和作为损失 $\mathcal{L}_2$。

$$\mathcal{L}_2 = \sum_{k=1}^{K} \sum_{j=1}^{K} d\left(c_i, c_j\right) \tag{5}$$

$\mathcal{L}_1$ 损失旨在将目标域样本靠近其 softmax 层输出较大的那一类的质心，以便目标域样本实现更好的分类。同时，$\mathcal{L}_2$ 损失旨在将目标域各个类的质心尽可能地远离，以尽可能将不同类样本分离开来。由此得到目标模型训练总的损失为：

$$\mathcal{L}(f_t) = \mathcal{L}_1 - \lambda \mathcal{L}_2 \tag{6}$$

2.3 算法流程

本节主要介绍基于区分显著性的无源无监督域适应通信辐射源个体识别方法的基本算法，如表 1 所示。

表 1 所提方法算法实现

1. 预搭建深度模型，输入源域样本 x_s 以及源域标签 y_s。
2. 依据式（1）计算损失 $\mathcal{L}_{src}$，训练源域模型 f_s，保存 g_s、h_s。
3. 输入目标域样本 x_t，迭代次数 T，初始化 g_t=g_s、h_t=h_s。
4. for t in 1:T do：
5. x_t 输入 g_t，得到 $g_t(x_t)$
6. 将 $g_t(x_t)$ 输入 h_t，得到 $h_t(g_t(x_t))$
7. 将 $h_t(g_t(x_t))$ 输入 softmax 层，得到 $\delta_k(h_t(g_t(x_t)))$
8. 依据 $\delta_k(h_t(g_t(x_t)))$，得到目标样本的标签
9. 根据式（2）计算质心 c_k
10. 根据式（3）（4）计算各目标样本到其质心的距离 $\mathcal{L}_1$
11. 根据式（5）计算各质心之间的距离，作为损失 $\mathcal{L}_2$
12. 根据式（6）最小化损失 $\mathcal{L}$，更新 g_t 和 h_t
13. end

它由两个过程组成：源域模型训练过程为步骤 1~2，目标域模型训练过程为步骤 3~13。

3 实验

3.1 实验条件

实验的硬件环境配置：使用的 CPU 为 Inter (R) Xeon (R) Silver 4210R CPU @2.40 GHz，使用的 GPU 为 Nvidia Geforce RTX 3090，内存为 DDR60G×2000。

实验的软件环境配置：Python 版本为 3.10，torch 版本为 1.11.0，Cuda 版本为 11.3。

3.2 数据集

本实验使用了 Oracle[18] 射频指纹数据集（Oracle RF fingerprinting dataset，ORFD），并进行了改造，以形成适合域适应任务的数据集。此外，为了更贴近实际应用场景，本实验还采集了专用于域适应任务的射频指纹数据集。

在 Oracle 射频指纹数据集中，为了确保数据集的丰富性和完整性，针对每类发射机信号，保留了共计 2000 个 2×256 大小的样本，作为原始数据集 D_{RAW}，而后通过向原始数据集 D_{RAW} 中分别添加 0 dB、5 dB、10 dB 和 15 dB 的噪声，生成 $D_{0\,dB}$、$D_{5\,dB}$、$D_{10\,dB}$ 和 $D_{15\,dB}$ 四个数据集，每个数据集均有 16 个类别的样本，每类样本 2000 个，总共有 32 000 个样本。

实验中，设置在不同调制方式、中心频率和发送频率条件下，采集 5 个射频指纹数据集，以模拟和探究各种环境变化对辐射源个体识别的影响。每组数据均包含 10 类样本，每类样本中均含有 2000 个尺寸为 2×256 的样本，具体设置如表 2 所示。

表 2 数据集 D_1、D_2、D_3、D_4 和 D_5 信号参数设置

名称	中心频率 /GHz	发送速率 /(kbit·s^{-1})	调制方式	距离间隔 /m
D_1	1	20	8PSK	1
D_2	0.8	20	8PSK	1
D_3	1	40	8PSK	1
D_4	1	20	4QAM	1

3.3 对比实验

当前，在辐射源个体识别任务中，很少研究是基于无源域适应场景，因此本文选取在其他领域均具有较好表现的无源域适应代表性方法作为对比，并将其应用在前一节中提到的数据集中。

本文主要选择两种方法作为对比，分别是基于源域假设的迁移方法（source hypothesis transfer，SHOT）和基于特征聚类的方法（attracting and dispersing，AaD）。CNN 代表不使用迁移学习方法，Source-free 代表使用无源域适应方法分别在 Oracle 标准数据集和实验室实采数据集上进行实验。表 3 对比了在信道变化条件下各方法的准确率，表 4 对比了在接收条件变化下各方法的准确率。

表 3 在 Oracle 数据集上不同方法分类准确率

单位：%

标准	方法	$D_{RAW}\rightarrow D_{0\,dB}$	$D_{RAW}\rightarrow D_{5\,dB}$	$D_{RAW}\rightarrow D_{10\,dB}$	$D_{RAW}\rightarrow D_{15\,dB}$	平均识别率
CNN	CNN	10.94	38.59	86.66	94.88	57.77
Source-free	SHOT	59.06	89.31	94.02	94.58	84.24
	AaD	48.31	89.20	94.24	94.65	81.6
	ours	60.18	89.73	94.70	94.97	84.86

表 4 在自采数据集上不同方法分类准确率

单位：%

标准	方法	$D_1\rightarrow D_2$	$D_1\rightarrow D_3$	$D_1\rightarrow D_4$
CNN	CNN	21.49	31.74	41.98
Source-free	SHOT	24.98	43.41	44.32
	AaD	21.82	38.80	45.01
	Ours	25.76	45.81	46.23

为了更好展示对比实验结果，实验绘制了 $D_{RAW}\rightarrow D_{5\,dB}$ 任务下各方法的准确率和损失函数变化曲线，如图 2、图 3 所示。

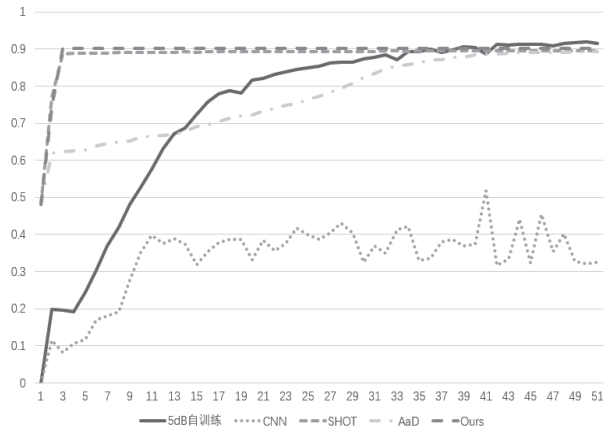

图 2 $D_{RAW}\rightarrow D_{5\,dB}$ 任务各方法准确率变化曲线

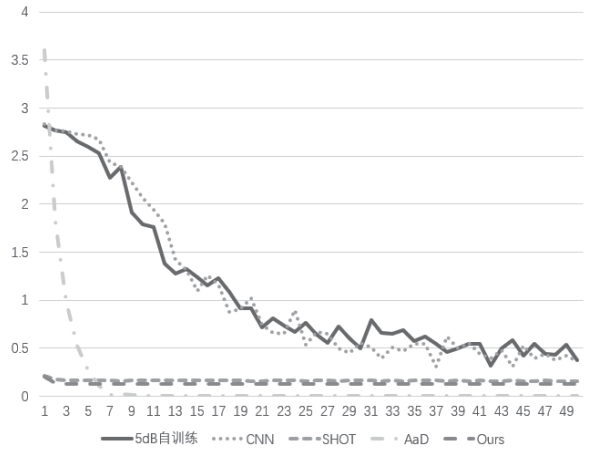

图 3 $D_{RAW}\rightarrow D_{5\,dB}$ 任务各方法损失函数变化曲线

3.4 消融实验

为进一步探究并验证本文提出的无源域适应中各个模块存在的必要性，本节进行消融实验以评估缺失不同模块条件下的实验效果。为研究信道环境发生变化时各个模块的必要性，使用 Oracle 射频指纹数据集，并在不同信噪比高斯噪声传输条件下进行实验，具体实验结果如表 5 所示。除此之外，对超参数 λ 进行了实验。实验结果如表 6 所示。

表 5 所提方法不同组件上的消融实验

单位：%

方法	$D_{RAW}\rightarrow D_{0\,dB}$	$D_{RAW}\rightarrow D_{5\,dB}$	$D_{RAW}\rightarrow D_{10\,dB}$
$\mathcal{L}1$	59.58	89.58	94.69
$\mathcal{L}2$	59.75	89.68	94.68
$\mathcal{L}1$、$\mathcal{L}2$	60.18	89.73	94.70

表 6 所提方法在不同 λ 上的消融实验

单位：%

方法	$D_{RAW}\rightarrow D_{0\,dB}$	$D_{RAW}\rightarrow D_{5\,dB}$	$D_{RAW}\rightarrow D_{10\,dB}$
λ=0.3	59.75	89.70	94.68
λ=1	59.77	89.72	94.70
λ=1	60.18	89.73	94.70
λ=5	60.04	89.73	94.70
λ=10	59.72	89.63	94.68

4 总结

本文研究了基于区分显著性的无源无监督域适应通信辐射源个体识别方法，通过迁移源域模型并引入目标域特征的区分显著性约束，促使特征提取模型向更有利于目标域特征类内聚类、类间分离的方向优化，从而提升识别的迁移效果。

该方法在 Oracle 射频指纹标准数据集和实采数据集上都取得了良好的识别效果。

参考文献：

[1] VOULODIMOS A, DOULAMIS N, DOULAMIS A, et al. Deep learning for computer vision: a brief review[J]. Computational intelligence and neuroscience, 2018, 2018:1-13.

[2] SHEN D, WU G, SUK H I. Deep learning in medical image analysis[J]. Annual review of biomedical engineering, 2017, 19: 221-248.

[3] OTTER D W, MEDINA J R, KALITA J K. A survey of the usages of deep learning for natural language processing[J]. IEEE transactions on neural networks and learning systems, 2020, 32(2): 604-624.

[4] 秦嘉 . 基于深度学习的通信辐射源个体识别 [D]. 北京：北京邮电大学 ,2019.

[5] LI D, YANG Y X, SONG Y Z, et al. Deeper, broader and artier domain generalization[C]//2017 IEEE International Conference on Computer Vision. Piscataway:IEEE, 2017: 5542-5550.

[6] WANG M, DENG W H. Deep visual domain adaptation: A survey[J].Neurocomputing, 2018,312(10):135-153.

[7] SUN B C, FENG J S, SAENKO K. Return of frustratingly easy domain adaptation[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence and the Twenty-Eighth Innovative Applications of Artificial Intelligence Conference,v.3. Palo Alto:AAAI Press, 2016:2058-2065.

[8] TZENG E, HOFFMAN J, ZHANG N, et al. Deep domain confusion: maximizing for domain invariance[DB/OL]. (2014-12-10)[2024-03-21]. https://doi.org/10.48550/arXiv.1412.3474.

[9] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks[J]. JMLR, 2016,17(1):2096-2030.

[10] SHU R, BUI H H, NARUI H, et al. A dirt-t approach to unsupervised domain adaptation[C/OL]//ICLR 2018.(2018-05-23)[2024-04-13]. https://doi.org/10.48550/arXiv.1802.08735.

[11] LEE C Y, BATRA T, BAIG M H, et al. Sliced wasserstein discrepancy for unsupervised domain adaptation[C]// 2019 IEEE/CVP Conference on Computer Vision and Pattern Recognition,[v.15]. Piscataway: IEEE,2019:10285-10295.

[12] LU Z H, YANG Y X, ZHU X T, et al. Stochastic classifiers for unsupervised domain adaptation[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 9111-9120.

[13] KUNDU J N, VENKAT N, BABU R V. Universal source-free domain adaptation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition,[v.7]. Piscataway: IEEE, 2020:4543-4552.

[14] KUNDU J N, VENKAT N, RAHUL M V, et al. Towards inheritable models for open-set domain adaptation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition,[v.7]. Piscataway: IEEE, 2020: 12376–12385.

[15] HUANG J X, GUAN D Y, XIAO A R, et al. Model adaptation: Historical contrastive learning for unsupervised domain adaptation without source data[C/OL]// Neural Information Processing Systems 34 (NeurIPS 2021). Red hook NY:Curran Associates Inc., 2021. [2024-04-21]. https://doi.org/10.48550/arXiv.2110.03374.

[16] YANG S Q, WEIJER J V D, HERRANZ L, et al. Exploiting the intrinsic neighborhood structure for source-free domain adaptation[C/OL]// NeurIPS 2021. (2021-11-29)[2024-04-17]. https://doi.org/10.48550/arXiv.2110.04202.

[17] XIA H F, ZHAO H D, DING Z M. Adaptive adversarial network for source-free domain adaptation[C]// 2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 9010-9019.

[18] SANKHE K, BELGIOVINE M, ZHOU F, et al. ORACLE: optimized radio classification through convolutional neural networks[C]//IEEE INFOCOM 2019-IEEE Conference on Computer Communications. Piscataway: IEEE, 2019: 370-378.

【作者简介】

韩昌芝（1999—），男，山东青岛人，硕士研究生，研究方向：迁移学习、模式识别。

陈健威（2000—），男，广东江门人，硕士研究生，研究方向：迁移学习、模式识别。

王闯（1995—），男，安徽滁州人，硕士研究生，研究方向：迁移学习、模式识别。

俞璐（1973—），女，吉林长春人，博士，副教授，研究方向：多媒体信息处理、模式识别、图像处理等。

（收稿日期：2024-08-06）