文章编号：2095-4980(2023)06-0734-11

辐射源个体识别的一种可解释性测试架构

刘文斌[a]，范平志[a]，李雨锴[b]，王钰浩[b]，孟　华[*b]

(西南交通大学 a.信息科学与技术学院；b.数学学院，四川 成都 611756)

摘　要：由于射频信号种类多，电磁环境复杂，特征提取难度大，现有的基于人工特征的射频辐射源个体识别方法的鲁棒性、适用性难以满足应用需求。数据驱动的深度学习方法虽然可以提供更灵活的辐射源个体识别模式，但深度学习方法自身可解释性差，而且缺乏通用测试模式来评价一个深度学习方法的优劣。本文在电磁大数据非凡挑战赛目标个体数据集的基础上，探索了基于该数据集的深度学习模型测试方法，提出面向辐射源个体识别神经网络模型的通用测试系统架构。该构架通过信号特征遮掩、生成对抗网络(GAN)、欺骗信号汇集、信道模拟等方法构造仿真测试样本，并把测试样本与原样本数据导入深度模型进行识别结果对比测试。基于测试结果分析了深度模型聚焦的信号关键特征位置，分析模型的鲁棒性，揭示信道环境对识别性能的影响，从而解释了深度学习网络模型的性能。

关键词：辐射源个体识别；可解释性；生成对抗网络；无线信号欺骗

中图分类号：TN92　　**文献标志码：**A　　**doi：**10.11805/TKYDA2022243

An interpretable testing architecture for specific emitter identification

LIU Wenbin[a]，FAN Pingzhi[a]，LI Yukai[b]，WANG Yuhao[b]，MENG Hua[*b]

(a.School of Information Science & Technology; b.School of Mathematics, Southwest Jiaotong University, Chengdu Sichuan 611756, China)

Abstract：Due to the diversity of RF signals, the complexity of the electromagnetic environment, and the difficulty of feature extraction, the robustness and applicability of the existing artificial features-based RF-specific emitter identification methods cannot meet the application requirements. Although the data-driven deep learning methods can provide a more flexible mode of specific emitter identification, they are less interpretable and lack a general test mode to evaluate their advantages and disadvantages. An evaluation method is explored for the deep learning model on the target individual dataset of the Electromagnetic Big Data Super Contest, and a general testing system architecture is proposed for the specific emitter identification model based on deep neural networks. The framework constructs the simulation test samples through signal feature masking, Generative Adversarial Network (GAN), deception signal collection, channel simulation and other methods, and imports the test samples and original data into the deep model to compare the recognition results. The test results are employed to judge the location of the signal key features extracted by the deep model, to analyze the robustness of the model, and to reveal the impact of the channel environment on the recognition performance, thus the performance of the deep learning model can be interpretable.

Keywords：specific emitter identification；interpretability；Generative Adversarial Network(GAN)；wireless signal spoofing

信号识别包括信号检测、信号类型识别、辐射源个体识别(Specific Emitter Identification，SEI)3 个层次[1]。信号检测一般指信号存在性检测。随着检测的频段带宽逐步加大，在存在性检测的基础上还需识别出观测范围内

收稿日期：2022-12-09；**修回日期：**2023-02-04

基金项目：国家自然科学基金资助项目(62276218)

[*]**通信作者：**孟　华　email:menghua@swjtu.edu.cn

各个信号的频率、带宽等参数以及信号的类别，该过程升级为信号类型识别。如果能进一步识别出是哪个具体辐射源发出的信号，则称为辐射源个体识别。传统的信号类型识别包括信号的体制类型的识别和基于信号波形变化的行为识别。由于信号的调制识别对信号而言尤其重要，因此自动调制识别已经成为深度学习在信号识别领域的研究热点。鉴于不同于软件定义的调制特征，辐射源个体特征一般是由射频通道等硬件决定的，因此辐射源个体识别需要对发送同类信号的不同个体对象进行区分，主要用于非合作形式下的身份认证，如入侵检测、信号侦察等。

传统的辐射源个体识别方法，一般是先提取信号的瞬态或稳态高阶特征，再基于特征进行个体区分。引入深度学习方法之后，可以通过卷积神经网络、循环神经网络等模型自动提取信号的数值特征并应用于个体分类[1−3]，当训练样本足够充分时，基于深度学习的特征提取与分类通常具有更强的识别性能。深度学习模型的性能与训练样本的数量和质量高度相关，而过去辐射源个体识别领域缺乏可靠的大规模基准信号数据集，这严重阻碍了识别模型的构建、评估及应用。同时，深度学习方法提取的特征向量是针对损失函数进行模型调参后获取的数值特征，因此对人类而言，这些特征是难以理解的，因此深度学习模型的可解释性是制约模型广泛应用的重要瓶颈[4−6]。另一方面，随着对抗智能方法的深入，面向辐射源个体识别的攻击方法也引起了学界的注意[7]，即便深度学习方法在公开数据集上有更高的识别正确率，但只需要对目标网络进行针对性的梯度攻击即可让网络函数在辐射源识别时进行误判，而传统识别方法在对抗梯度攻击上反而更加鲁棒。由于与辐射源特征识别相关的工程应用对鲁棒性和安全性有着很高的要求，因此有必要引入对个体识别深度学习网络的通用且简单可行的鲁棒性测试方法，通过模型的鲁棒性来解释个体识别样本集分布特性以及深度学习模型的性能边界。

关于辐射源数据集的收集整理和智能分类方法的研究是近年来领域的研究热点。Timothy 使用 GNU Radio 等构建了开源的多个版本的调制数据集，样本长度从 128 增长到了 1 024，并在此基础上开展了大量的机器学习与深度学习模型的设计与测试[8]。随着研究的深入，更多的如：宽带调制数据、4G、5G、WiFi、NB−IoT 等民用数据集不断地被整理发掘并为人工智能的应用提供重要的支撑[9−11]。另一方面，如何利用已有的数据产生更多的数据样本，即数据增强技术也被引入到信号处理领域[12]。

辐射源特征提取与识别是信号处理领域的研究热点，传统的辐射源个体识别主要是基于辐射源硬件的物理特性对辐射源进行分类或模式识别，虽然也有采用功放仿真等数值方法来模拟产生数据进而训练分类算法，但该模式受限于特定对象，因此应用存在局限。为了拓展辐射源识别的方法，构建基于人工智能和数据驱动的辐射源识别技术引起学者们的广泛关注。美国东北大学 Shawabka 等介绍了自己构建的 20 余台 USRP(Universal Software Radio Peripheral)设备的数据集，以及国防高级研究计划局(Defense Advanced Research Projects Agency，DARPA)构建的 5 117 个 WiFi 设备和 5 000 个 ADS−B 设备的数据集，分析了不同信道环境下的个体识别测试结果，验证了无线信道的时变性会对识别准确率产生显著影响，发现深度学习学到的更多的是信道特征而不是辐射源特征，并提出了基于信道均衡的方法来提高识别精确度[13]。生成对抗网络(GAN)是一个包含判别器与生成器的深度学习模型[14−17]。它可以无监督地在复杂数据集上探索数据的分布规律，并仿照已知样本生成新的样本，在信号处理领域，GAN 网络常用于对信号进行样本增强、调制识别、噪声干扰消除等方面[18]，并尝试应用于辐射源个体识别[19−21]。美国霍普金斯大学无线物理实验室(Applied Physics Laboratory，APL)指出，DARPA 的自适应电子战行为学习(Behavioral Learning for Adaptive Electronic Warfare, BLADE)、自适应雷达对抗(Adaptive Radar Countermeasures, ARC)等项目已将机器学习成功应用于捷变通信信号(agile communications signals)和威胁雷达信号，但是在未知新波形识别、复杂时频波形数据表示、结合信号特征与其他背景数据（如方向、波极化或地理位置）的辐射源个体识别等方面，现有的算法、模型与实际工程应用存在差距，并建议对错误识别的样本来源与出错原因进行分析[22]。

上述研究表明，数据驱动与人工智能相结合的辐射源特征提取与识别成为新兴的研究方向，但离实际应用仍有一定的差距，主要原因是缺乏开源的数据集和能适用于工程应用的可解释性强的测试方法。2022 年 6 月举行的第一届电磁大数据非凡挑战赛(Electromagnetic Big Data Super Contest，EBDSC)，给出了开源的目标个体识别数据集[23]。优质的公共数据集是讨论信号特征和网络分类能力的重要基础，因此该比赛项目一方面为基于深度学习方法的辐射源特征提取与识别提供了重要的推动；另一方面也为评价深度学习方法的鲁棒性、测试模型的可解释性提供可能。赛事中目标个体识别数据集包含 10 万条样本共计 10 个类别。赛题聚焦电磁目标个体的细微特征挖掘，通过设计智能算法，完成对电磁目标个体的识别。结合工程应用需求和无线信道的复杂性，本文基于该数据集，设计了一类针对辐射源识别模型的通用鲁棒性测试框架，利用多种精心设计的数据生成、数据欺骗等手段检查模型性能边界，评估模型的鲁棒性能与可解释性，进而为模型的工程应用提供必要的风险评估。

本测试框架也可以应用到其他各类辐射源个体识别的可解释性分析测试，探索人工智能方法是否适合处理各类辐射源数据，并通过智能分类模型、生成模型的实验解构分析该数据集的特性与分布特征，如通过高相似样本生成来解释特征提取与学习机理，分析样本分布及其关联性，揭示信道环境对识别性能的影响，从而提升对目标样本特征和被测网络能力边界的可解释性。

本文提出了一种测试方法，首先构造辐射源个体识别深度学习网络，对目标数据集的多个目标个体对象进行区分；然后对原始样本进行针对性的遮挡、重构和加噪，进而对新生成样本进行识别测试。通过对比识别结果，分析得出原始样本数据集的个体分布特征和识别网络的能力边界。

1 可解释性测试方法与系统架构

1.1 样本集

与调制信号特征由程序设计产生相比，辐射源个体特征一般与其射频、功放、天线等硬件相关，难以模拟生成，实际数据获取困难，因此也缺少开源的数据集；研究普遍采用的USRP设备以及ADS-B设备、WiFi设备样本，其数据集未开放，且不同对象信号样本差异很大，识别方法的迁移性差，且很难解决特定目标的识别问题。

电磁大数据非凡挑战赛[8]提供了目标个体识别的开源数据集，其辐射源数据共10个类别，每个类别有10 000条样本(训练测试划分为8:2)；每条样本含3 000个采样点，其中信号段约占2 000个采样点；覆盖5 dB~14 dB共10种信噪比(Signal-to-Noise Ratio，SNR)，即每个类别每种信噪比有1 000条样本；未提供采样率、调制类型等参数，但经过信号特征分析可以得出，信号包含两类脉冲，前5个对象为单一频率信号，后5个对象为线性调频(Linear Frequency Modulation，LFM)信号。

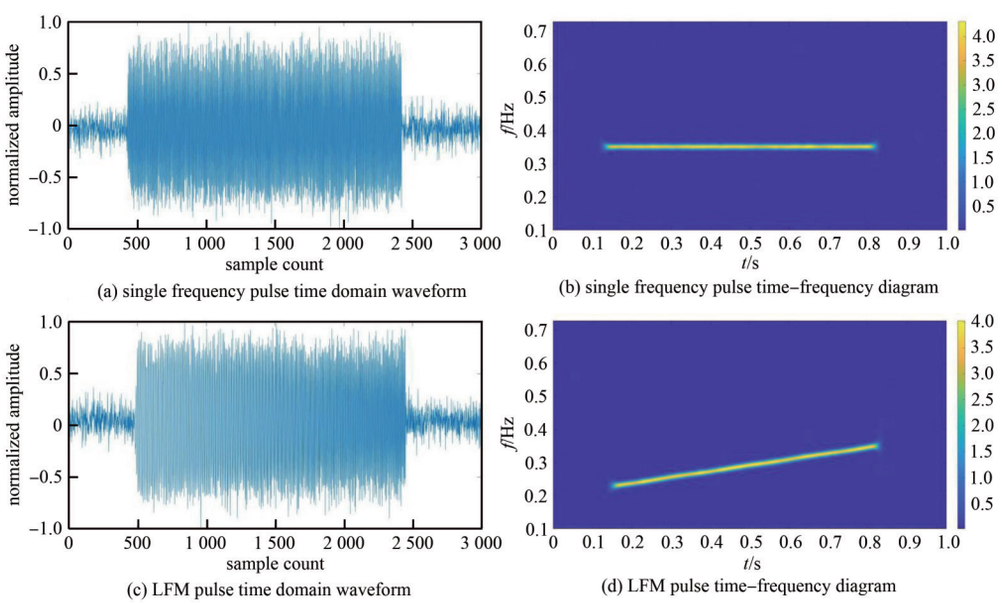

Fig.1 Original signal sample (SNR: 14 dB)
图1 原始样本信号 (信噪比: 14 dB)

1.2 实例识别网络

本文为了展示整个测试框架，首先搭建了一个针对辐射源个体识别的实例网络—尺度转移检测网络(Scale-Transferrable Detection Network, STDN)。该网络模型采用了自适应的时域特征抽取模型—转移窗注意力汇聚网络。一般的Transformer模型处理长时序信号时复杂度高且难以学习信息规模变化大的数据，这导致其强大的表示能力难以得到发挥。为充分地提取特征，网络以自注意力机制作为特征提取的基础组件，设计逐块的自注意力计算模块来减少冗余的参数和计算量；同时，受卷积操作的启发，在每次降采样之前间歇地应用转移窗口来增大感受野；最后，利用扩展通道式的跨步卷积来代替传统的池化实现降采样，避免特征丢失的问题。上述一系列方法解决了Transformer模型不适配辐射源个体识别的问题。基于时域Transformer模型提取的特征输入一个全连接层和Softmax输出层，得到最终的预测概率。辐射源个体识别网络模型结构如图2所示。

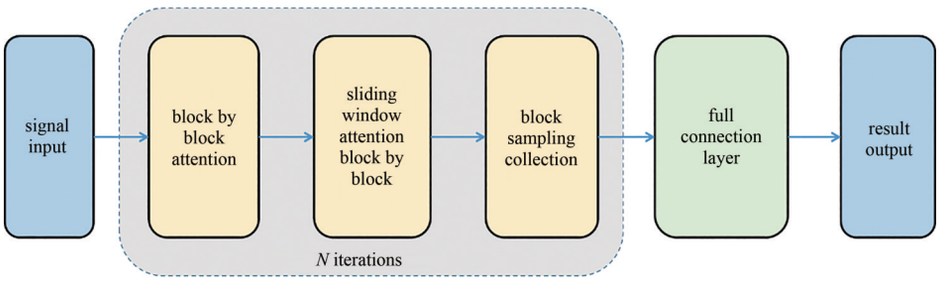

Fig.2 Specific Emitter Identification model—STDN
图2 辐射源个体识别网络模型STDN

1.3 基于生成对抗网络的数据生成方法

生成对抗网络(GAN)是一种深度学习模型，模型通用框架中一般包含生成器(Generative Model)和判别器(Discriminative Model)，本文采用的模型结构如图 3 所示。生成器通过随机信息产生仿真样本，而判别器对抗性地去区分真实样本和仿真样本。在互相博弈学习的过程中不断提升生成器的性能，进而找到样本数据可能的生成模式和分布规律。

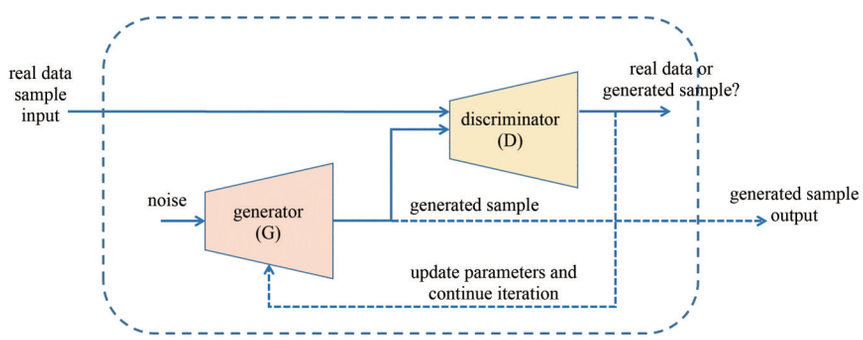

Fig.3 Sample generation network based on GAN
图3 基于GAN的样本生成网络

在信号处理领域，不论是通过生成式对抗网络增强原始样本训练集的离线训练方式，或是在识别网络训练过程中动态地生成对抗样本来进行在线的对抗训练，其目的都是通过数据增强的方法提升识别网络的能力。本文侧重于基于样本生成方法提供一个通用的测试系统，并提升对样本集和被测网络的分析能力。

1.4 针对辐射源识别深度网络模型的测试框架

深度学习模型往往是从训练样本中通过计算探索有益于分类的数值特征。这些数值特征对人类而言非常不直观。为了分析和解释网络的表征能力与分类能力，本文提出了如图4所示的测试框架。该架构包含信号特征遮掩、GAN网络的信号生成、信道耦合解耦、欺骗信号汇集等模块来测试和解析模型的能力。各模块的功能分析如下。

1) 信号特征遮掩：采用时频特征遮挡(掩码)的方法生成样本，并把生成样本送给识别模型进行判别。通过检查模型对遮挡前后样本的特征提取与分类情况，判断遮挡位置信号的特征强度：如果模型能对带遮挡的样本正确分类，说明特征提取器与分类边界对遮挡部分不敏感；如果遮挡样本无法正确分类，则说明特征提取与识别与该部分信息高度相关。

2) 生成模型：可以采用包含深度学习在内的各种生成模型来生成样本。合理的数据增强能帮助模型应对可能出现的复杂信号，以提升模型的泛化性能，如利用GAN作为可学习的数据增强。该模块的出发点是利用这种数据增强来构建攻击测试方法，丰富测试样本，支撑测试平台。本文采用简单全连接层的生成器(G)和判别器(D)来构建GAN网络，如果通过GAN网络生成的信号能很好地被分类，则说明整个模型的泛化性能强，样本距离分类边界比较远，模型有更强的抗干扰能力，也一定程度上展示了模型找到了目标对象的核心特征，被测模型具有更好的可信度和可解释性。

3) 信道耦合解耦：信道对辐射源个体特征提取有极大的影响，包括衰减、多径、噪声等。在基于深度学习的辐射源特征提取与分类时，因为训练样本采集的时空以及规模限制，识别模型极有可能学习到的是信道特征而不是辐射源射频的指纹特征，为此需要分析并减轻信道的影响，称之为去卷积或解耦合。本文以高斯白噪声(Additive White Gaussian Noise, AWGN)信道为例，测试信噪比对模型识别结果的影响，并支撑通过降噪来提升个体识别性能。

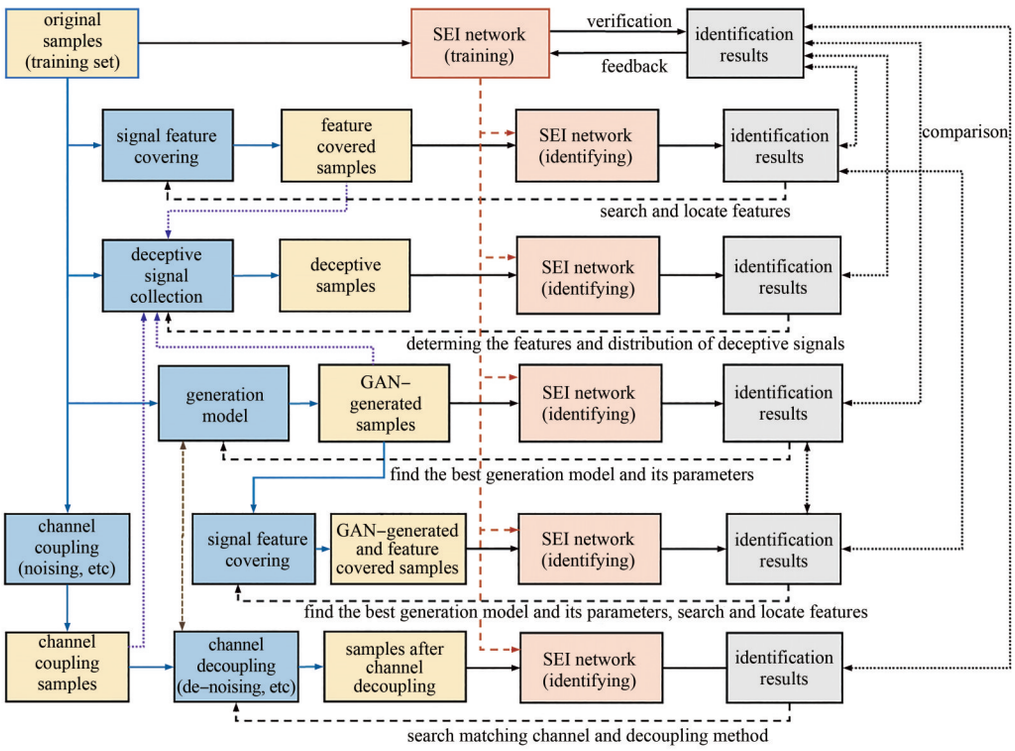

Fig.4 Deep learning model performance test architecture for Specific Emitter Identification

图 4 辐射源识别深度学习模型性能测试架构

4) 欺骗信号汇集：通过汇集原始信号、遮掩信号、GAN 生成信号、信道耦合信号等样本中造成模型误判的“欺骗性”样本，对这些样本的特征、生成参数进行分析，跟踪样本“变异”的识别结果并得到这种映射关系的知识图谱。通过对比分析正常样本与欺骗样本的差异，从而分析被测模型的能力与脆弱性。

该测试框架是一个通用的测试模式，任何关于辐射源特征提取与识别的深度网络模型都可以通过该框架分析、解释网络模型的能力与鲁棒性。

2 基于 GAN 的辐射源个体识别测试实验

2.1 测试环境与基线测试

1) 测试环境

模型训练及测试环境：操作系统 Windows sever 2019，CPU 为 Intel E5–2630，GPU 为 NVIDIA TITAN Xp×3。

2) 个体识别基线测试

针对 1.1 节介绍的原始样本训练集和测试集，针对 1.2 节提出的识别网络 STDN 进行测试，10 个对象的识别结果混淆矩阵如图 5 所示。训练集中，总样本数为 80 000 条，前 5 个对象中误识别的样本为 4 213 条，后 5 个对象中误识别的样本为 6 条，总的识别准确率为 94.7%；测试集中，总样本数为 20 000 条，前 5 个对象中误识别的样本为 1 078 条，后 5 个对象中误识别的样本为 2 条，总的识别准确率为 94.6%。训练集和测试集的准确率结果相似，误识别分布结果相似，说明模型没有出现明显的过拟合现象，并且识别网络找到了数据差异的核心特征，但是约 5% 的误分类说明，有些样本越过了分类边界。通过单类正确率来看，后 5 个对象的识别准确率接近 100%，说明其特征易辨别；前 5 个和后 5 个对象之间没有交叉识别出错的情况，说明深度学习到的两组对象之间特征差异大；前 2 个对象中，3#(即图中 inst–3)识别准确率最高。

3) 特征遮挡测试

针对每一个对象选取训练集中信噪比为 14 dB 的前 256 个样本，分别对样本 3 000 个样点中的头部(1~1 000 段)、腰部(1 001~2 000 段)、尾部(2 001~3 000 段)，从各段的起始位置开始设置窗口进行掩码遮挡（即样点值替换成 0)，遮挡窗口大小按 200、400、600、800、1 000 的长度依次增加。掩码后的样本送给识别网络进行判断，结果如图 6 所示，其中每个对象的准确率的计算方式为输出结果中识别正确的样本数量与输入的该对象样本数量的比值。在测试结果中，识别准确率下降较快的遮挡位置可认为是关键特征位置。考虑到每个样本 3 000 个样点中的起止样点分布有一定的随机性，因此测试采用统计分析。

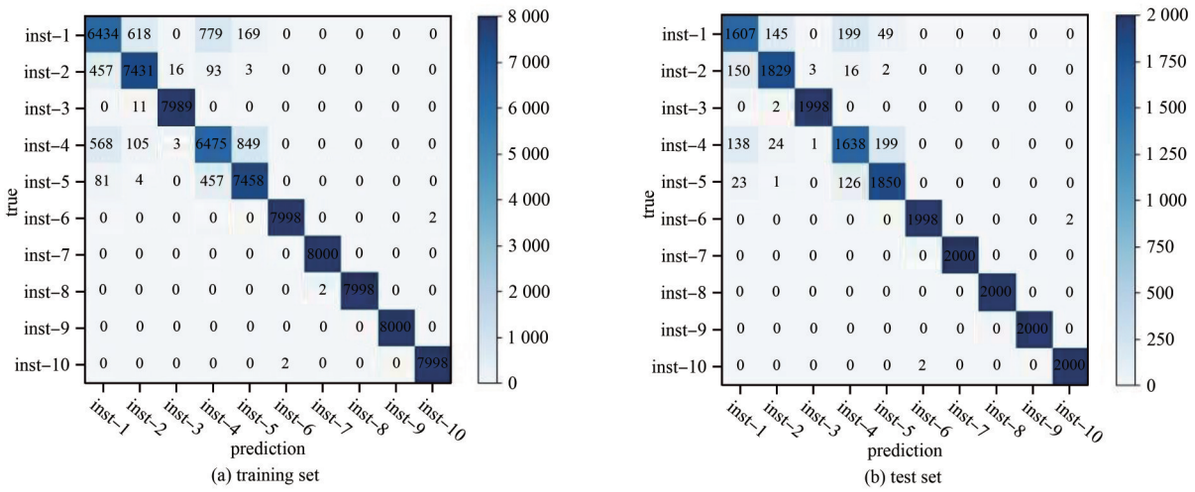

Fig.5 Confusion matrix of identification result

图5 识别结果混淆矩阵

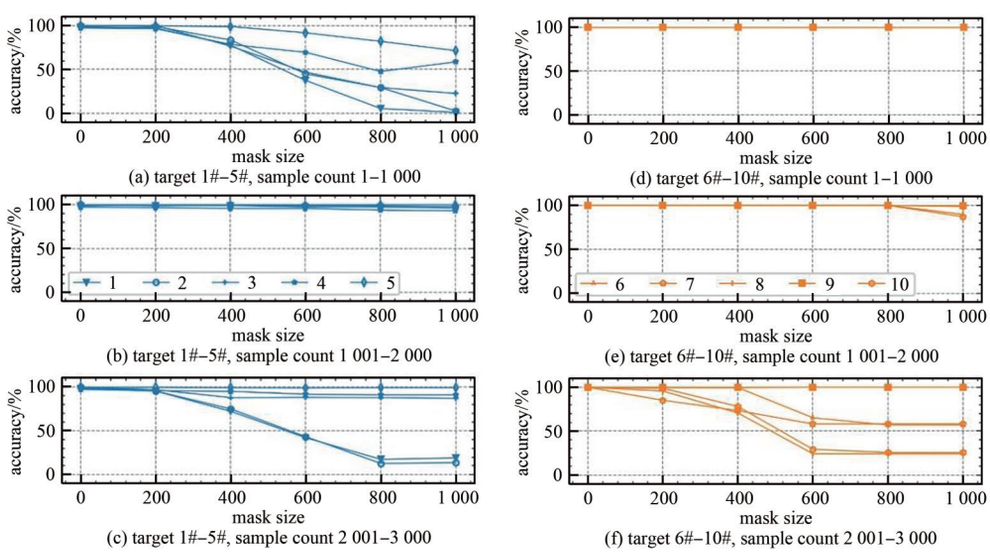

Fig.6 Locating test results when covering original sample features

图6 原始样本特征遮挡定位测试结果

从图6可以看出1#–5#对象中：1#和2#在遮挡头部和尾部时识别准确率都有显著下降，说明特征集中在头尾2个位置；3#和4#在遮挡头部或尾部时识别准确率有一定程度下降，且头部下降更为明显，说明其头部特征更明显；5#在遮挡头部时识别准确率有一定程度下降，但在尾部遮挡时未下降，说明其尾部特征不明显；各对象的腰部被遮挡时识别准确率下降不明显，说明各对象腰部特征不明显。

6#–10#对象中：头部被遮挡时识别准确率都未下降，说明所有对象的关键特征位置没有局限在头部；腰部被遮挡时7#和8#有轻微下降；尾部被遮挡时8#和10#降幅较大，6#和7#有一定降幅，9#未下降，说明对象个体间差异较大。

从测试结果可以分析得出，特征遮挡能够快速发现个体关键特征及其位置。发现6#–10#只有尾部特征会影响个体识别；而1#–5#头部和尾部都会影响个体识别；发现纯噪声段的信号遮掩对个体识别的结果几乎无影响，说明特征是从样本的信号段学习得到的；发现线性调频信号比单频信号更容易识别，且头部遮挡对其识别结果影响小，说明频率较宽的信号特征部位更多，更易于被模型学习到。

4) 信噪比影响测试

针对1#、10#训练集中的每种信噪比800条样本和验证集的每种信噪比200条样本，分别送给个体识别网络，计算其识别准确率，结果如图7所示。1#低信噪比时准确率较低，随信噪比的升高而增大；验证集与训练集的结果相似，而训练集因为样本量较大所以曲线更平滑。10#各信噪比时的识别准确率都较高，验证集与训练集的结果曲线几乎重合。

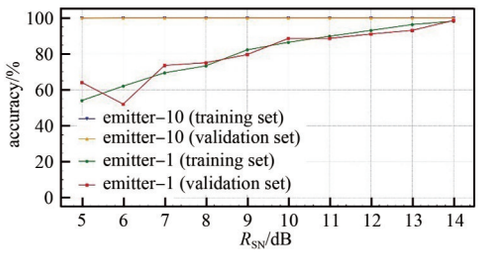

Fig.7 Identification accuracy under different SNRs for 1# and 10#
图7 1#和10#在不同信噪比下的识别准确率

2.2 基于GAN生成样本的测试

1) 时频相似度

GAN网络设置的参数为：学习率0.000 2；包含1个输入层、1个隐藏层和1个输出层，使用全连接层架构；Batch Size为256，每次迭代输出256个样本(每个样本有3 000个采样点)。使用14 dB的原始样本作为输入，生成的样本如图8所示，发现其时频特征与原始样本(见图1)均非常相似。

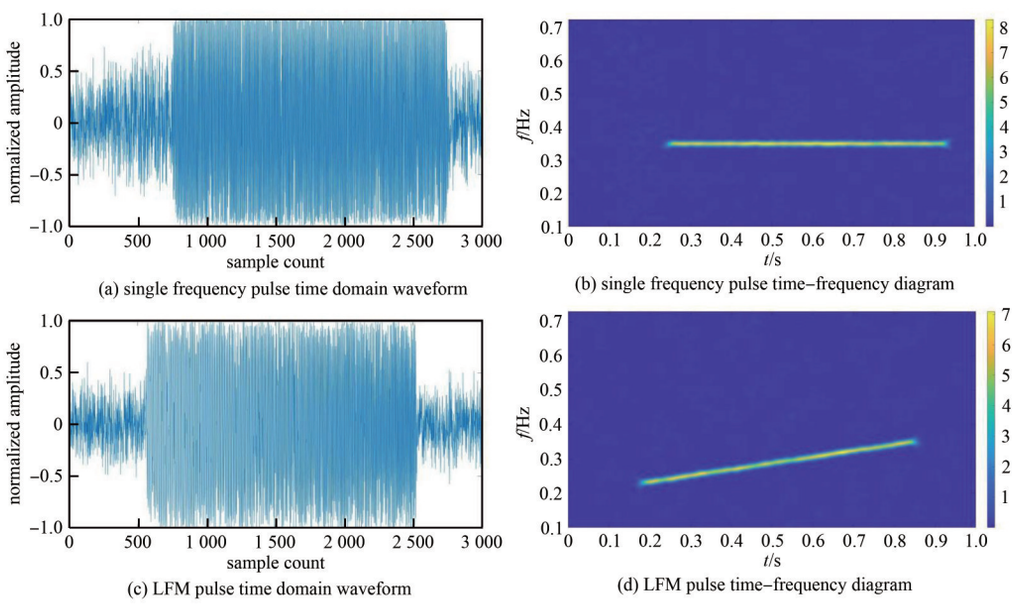

Fig.8 Signal generated by GAN(SNR: 14 dB)
图8 GAN生成信号图(信噪比: 14 dB)

2) 个体差异测试

使用相同的网络参数，基于4#、5#、9#、10#的14 dB原始样本作为输入进行GAN生成，各迭代1 000次，间隔10次记录一次迭代生成的样本，即共记录100次。生成样本送识别模型进行判断，其预测的准确率变化如图9所示，可以看出，4#、5#辐射源个体相比9#、10#个体难生成，其中，4#个体比5#个体更难生成。这与图5混淆矩阵中原始样本体现的趋势一致，说明生成的学习难度与识别的学习难度是相关的，其本质上都是对样本的特征学习。

3) 特征遮挡测试

进一步，基于各对象的14 dB原始样本作为GAN输入，把第200次后迭代生成且能被识别网络判别为真的样本构建新的样本集，每个对象样本数量为256条。对这些样本进行如3.1相同方法的遮挡测试，其结果如图10所示。

可以看出，1#–4#在遮挡头部和尾部时识别准确率都有显著下降；5#在遮挡头部时识别准确率有一定程度下降，但在尾部遮挡时没有下降；1#–5#腰部被遮挡时识别准确率部分对象有一定程度下降，但大都下降不明显。6#–10#中，头部被遮挡时识别准确率都未下降；腰部被遮挡时6#、7#和8#有一定下降；尾部被遮挡时7#、8#和10#降幅较大，6#有少量降幅，9#一直未下降。这些结果与原始样本的遮挡测试结果相似，说明生成样本能够学习到个体特征，且与原始样本特征部位相似。

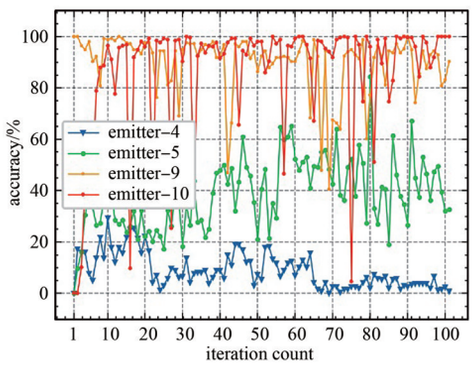

Fig.9 Sample generation tests for different emitters by GAN
图9 不同个体样本GAN生成测试

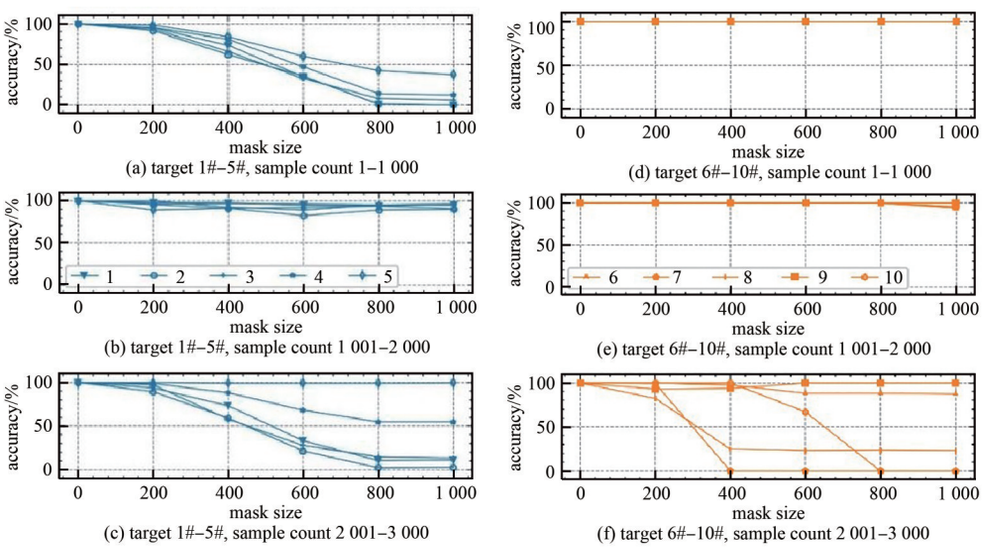

Fig.10 Locating test results when covering GAN generated sample features
图10 GAN生成样本的特征遮挡定位测试结果

4) 信噪比影响测试

计算同一个体不同信噪比时的GAN生成的难度，通过GAN生成样本的识别准确率来进行直观的评估，并通过迭代过程中的变化来更形象化地呈现。选择10#对象，输入GAN网络5 dB、9 dB、13 dB三种信噪比，分别进行500次GAN生成迭代，每10次记录1次迭代生成的样本并进行识别，结果如图11所示。可以看出，信噪比越高，生成样本的效果就越好。但因为GAN网络不稳定，迭代过程中生成样本的效果波动较大。

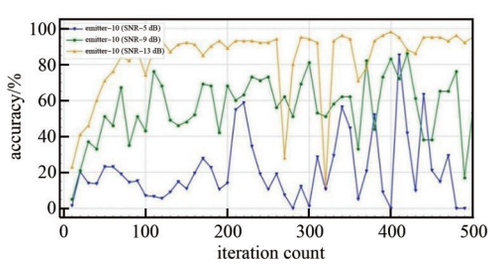

Fig.11 GAN iterative process under different SNRs
图11 不同信噪比下GAN迭代过程

在1#和10#各800条14 dB样本基础上添加高斯白噪声，构建5 dB~13 dB范围信噪比的加噪样本。把这些样本送给识别网络进行预测，同时构建矩阵来跟踪误识别结果，评估结果如图12所示。可以看出，随着信噪比的降低，识别准确率也会随之降低，但10#识别准确率一直远高于1#。预测结果与图7的原始样本测试结果相似，说明加噪操作成功模拟了原始样本的信道特征。从测试结果矩阵还可以发现信道变化对误判结果的显著影响，同时可以分析出信道特点和具体对象个体特征的鲁棒性。

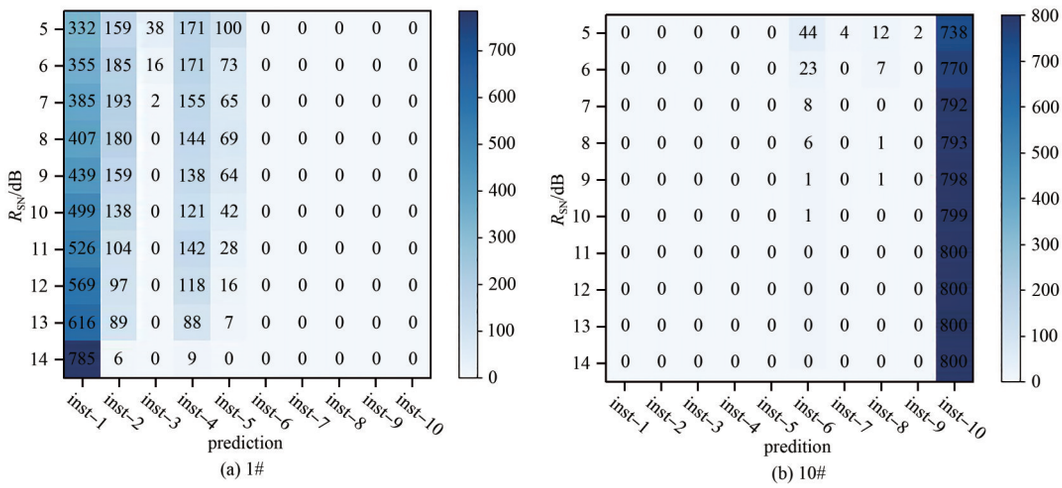

Fig.12 Identification distribution of signal samples after adding noise manually

图 12 人工加噪后信号样本识别结果分布

2.3 测试结果分析

通过面向辐射源个体识别网络构建可解释性测试架构，并把基于GAN生成样本的测试结果与基线测试对比，本文有几个主要的发现：

1) 不同个体对象的GAN生成难度不同，且生成难度与识别难度相匹配，这说明GAN本质上是对特征的学习与描述。因此可以用目标对象的GAN生成难度来评估个体识别的特征提取难度。

2) GAN生成样本在进行不同部位的掩码遮挡测试后，其结果与原始样本遮挡测试结果相似。这说明了被测辐射源具有可解释的个体特征和特征部位，简单全连接层设计的GAN网络对被测信号个体特征画像具有较强的学习与描述能力。

3) 信号样本信噪比越低，其特征提取与识别的难度越大，同时GAN学习与生成的难度也越大，均表现为模型对样本的预测准确率降低。

4) 通过加噪模拟生成的信号样本，与原始各信噪比样本的测试结果基本一致。

由此可以看出，基于本文的测试架构来进行模拟评估，既能挖掘并定位目标对象的个体特征，对特征的学习难度进行表征，还能模拟信道环境来拓展样本，从而能更为全面地测试辐射源个体识别深度学习网络的能力与缺陷。

3 结论

本文搭建了一个针对个体识别深度学习网络的测试系统架构，提出了特征遮挡、GAN生成、信道模拟等样本生成方法，并通过新的样本生成所带来的识别结果变化来搜索、展示被测辐射源的特征规律和被测个体识别网络的能力边界。经测试，本文所提出的方法能够搜索定位关键的个体特征，能模拟并分析信道影响，能展示造成误识别的“欺骗”样本分布，从多个角度提高了解释能力，从而有助于增强对目标个体样本和被测网络的理解。

从物理层信号识别的对抗学习方面考虑，系统可进一步收集并分析造成误识别的生成样本，构建更具针对性的欺骗性样本和测试方法；还能基于此系统，对GAN迭代过程误判别结果进行动态可视化呈现，跟踪学习与生成过程并设计选择生成参数。后续将考虑进一步深入搜索并区分稳态及瞬态特征，降低信道对个体识别结果的耦合影响，提升个体样本不均匀、信道环境变化大等情形下的样本生成能力；同时通过对测试样本的细分与跟踪，高分辨力、高动态展示个体特征与信道变化对测试结果的影响。后续希望在本文搭建的对抗性测试平台及其架构基础上，能逐步升级通用化的测试方法，促进辐射源个体识别网络的高可解释性工程应用。

参考文献：

[1] 李润东. 基于深度学习的通信信号智能盲检测与识别技术研究[D]. 成都:电子科技大学, 2021. (LI Rundong. Research on intelligent blind detection and recognition of communication signals based on deep learning[D]. Chengdu, China: University of Electronic Science and Technology of China, 2021.)

[2] CHEN Shichuan,ZHENG Shilian,YANG Lifeng,et al. Deep learning for large-scale real-world ACARS and ADS-B radio signal classification[J]. IEEE Access, 2019(7):89256-89264.

[3] WANG Guanhua,ZOU Cong,ZHANG Chao,et al. ACARS signal source generation and recognition based on convolutional neural network[C]// 2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting(BMSB). Chengdu,China: IEEE, 2021. doi:10.1109/BMSB53066.2021.9547081.

[4] YANG Zhou,LIU Ninghao,HU Xiaben,et al. Tutorial on deep learning interpretation: a data perspective[C]// Proceedings of the 31st ACM International Conference on Information & Knowledge Management. Georgia:[s.n.], 2022: 5156-5159.

[5] MOHAN S. Managing expectations:explainable a.i.and its military implications[Z]. ORF Issue Brief No.570, 2022.

[6] 梁先明,倪帆,陈文洁,等. 基于时频Grad-CAM的调制识别网络可解释研究[J/OL]. 西南交通大学学报, 2022. https://kns.cnki.net/kcms/detail/51.1277.u.20220608.1636.008.html. doi:10.3969/j.issn.0258-2724.20210791. (LIANG Xianming,NI Fan, CHEN Wenjie,et al. Interpretability of modulation recognition network based on time frequency Grad-CAM[J/OL]. Journal of Southwest Jiaotong University, 2022. https://kns.cnki.net/kcms/detail/51.1277.u.20220608.1636.008.html.) doi:10.3969/j.issn.0258-2724.20210791.

[7] SHI Yi, DAVASLIOGLU K, SAGDUYU Y E. Generative adversarial network in the air: deep adversarial learning for wireless signal spoofing[J]. IEEE Transactions on Cognitive Communications and Networking, 2020,7(1):294-303.

[8] SHEA T O, WEST N. Radio machine learning dataset generation with GNU radio[C]// Proceedings of the 6th GNU Radio Conference. Boulder:[s.n.], 2016:69-74.

[9] WEST N, SHEA T O, ROY T. A wideband signal recognition dataset[C]// IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications(SPAWC). Lucca,Italy:[s.n.], 2021:6-10.

[10] SHEBERT S R,MARTONE A F,BUEHRER R M. Open set wireless standard classification using convolutional neural networks[C]// MILCOM 2021 Track 5—Special Topics in Military Communications. San Diego,CA,USA:[s.n.], 2021:757-762.

[11] 李靖超,应雨龙. 基于功率谱密度的通信辐射源个体识别方法[J]. 太赫兹科学与电子信息学报, 2021,19(4):596-602. (LI Jingchao,YING Yulong. Individual identification method of communication radiation source based on power spectral density[J]. Journal of Terahertz Science and Electronic Information Technology, 2021,19(4):596-602.) doi:10.11805/TKYDA2021140.

[12] LI Mingxuan,LIU Guangyi,LI Shuntao,et al. Radio classify generative adversarial networks:a semi-supervised method for modulation recognition[C]// The 18th IEEE International Conference on Communication Technology, Chongqing,China:[s.n.], 2018:669-672.

[13] AL-SHAWABKA A,RESTUCCIA F,D'ORO S,et al. Exposing the fingerprint: dissecting the impact of the wireless channel on radio fingerprinting[C]// IEEE INFOCOM 2020—IEEE Conference on Computer Communications. Toronto, ON, Canada: IEEE, 2020:646-655.

[14] GOODFELLOW I,POUGET-ABADIE J,MIRZA M,et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11):139-144.

[15] MIRZA M,OSINDERO S. Conditional generative adversarial nets[J]. arXiv preprint arXiv:1411.1784, 2014.

[16] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.

[17] ARJOVSKY M,CHINTALA S,BOTTOU L. Wasserstein GAN[J]. arXiv:1701.07875. 2017.

[18] CHEN Shengyi, SHANGGUAN Wangyi, TAGHIA Jalal, et al. Automotive radar interference mitigation based on a generative adversarial network[C]// 2020 IEEE Asia-Pacific Microwave Conference(APMC). Hong Kong,China:IEEE, 2020:728-730.

[19] GONG Jialiang, XU Xiaodong, QIN Yufeng, et al. A generative adversarial network based framework for specific emitter characterization and identification[C]// 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP). Xi'an,China:IEEE, 2019:1-6.

[20] GONG Jialiang, XU Xiaodong, LEI Yingke. Unsupervised specific emitter identification method using radio-frequency fingerprint embedded InfoGAN[J]. IEEE Transactions on Information Forensics and Security, 2020(15):2898-2913.

[21] TAN Kaiwen, YAN Wenjun, ZHANG Limin, et al. Semi-supervised specific emitter identification based on bispectrum feature extraction CGAN in multiple communication scenarios[J]. IEEE Transactions on Aerospace and Electronic Systems, 2022. doi: 10.1109/TAES.2022.3184619.

[22] SHARMA P, SARMA K K, MASTORAKIS N E. Artificial intelligence aided electronic warfare systems—recent trends and evolving applications[J]. IEEE Access, 2020(8):224761-224780.

[23] 电磁大数据非凡挑战赛[EB/OL]. [2022-06-16]. http://ebdsc.cn/,2022-06-16. (Electromagnetic big data super contest[EB/OL]. [2022-06-16]. http://ebdsc.cn/,2022-06-16.)

作者简介：

刘文斌(1983–)，男，在读博士研究生，高级工程师，主要研究方向为电磁信号智能识别.email: bingege389@sina.com.

范平志(1955–)，男，博士，教授，博士生导师，主要研究方向为大规模物联网、高移动无线通信、信息理论与编码技术.

李雨锴(1994–)，男，学士，工程师，主要研究方向为电磁信号识别与生成.

王钰浩(1999–)，男，在读硕士研究生，主要研究方向为对抗训练与智能识别.

孟　华(1982–)，男，博士，副教授，主要研究方向为拓扑学、知识表示与推理、机器学习.

（上接第724页）

[6] 何波坪,黄文伟,胡庭棪. 信息化条件下联合作战协同对数据链系统的需求[J]. 国防科技, 2011,32(1):51–54. (HE Boping, HUANG Wenwei,HU Tingqi. Requirement of data link system for joint operation coordination under information condition[J]. Defense Science and Technology, 2011,32(1):51–54.)

[7] 张春磊,裴琴,易楷翔. 美电磁频谱作战技术体系与应对策略研究[J]. 中国电子科学研究院学报, 2022,17(5):439–444. (ZHANG Chunlei,PEI Qin,YI Kaixiang. Study on the technological architecture of US electromagnetic spectrum operations and countermeasures[J]. Journal of Chinese Academy of Electronic and Information Technology, 2022,17(5):439–444.)

[8] 高岩,于博. 复杂电磁环境特性[J]. 四川兵工学报, 2008,29(1):19–21. (GAO Yan, YU Bo. Characteristics of complex electromagnetic environment[J]. Journal of Sichuan Ordnance, 2008,29(1):19–21.)

[9] 汪连栋. 复杂电磁环境概论[M]. 北京:国防工业出版社, 2015. (WANG Liandong. Introduction to complex electromagnetic environment[M]. Beijing:National Defense Industry Press, 2015.)

[10] 李原,杨明川,王钢,等. 认知电子战理论及关键技术研究[C]// 第三届中国指挥控制大会. 北京:[s.n.], 2015:73–79. (LI Yuan, YANG Mingchuan,WANG Gang,et al. Research on the theory and crucial technology of cognitive electronic warfare[C]// The Third China Command and Control Conference. Beijing:[s.n.], 2015:73–79.)

[11] 祝学军,赵长见,梁卓,等. OODA智能赋能技术发展思考[J]. 航空学报, 2021,42(4):16–25. (ZHU Xuejun,ZHAO Changjian, LIANG Zhuo, et al. Development of OODA intelligent empowerment technology[J]. Acta Aeronautica et Astronautica Sinica, 2021,42(4):16–25.)

[12] 张澎,张成,管洋阳,等. 关于电磁频谱作战的思考[J]. 航空学报, 2021,42(8):87–98. (ZHANG Peng,ZHANG Cheng,Guan Yangyang,et al. Views on electromagnetic spectrum operation[J]. Acta Aeronautica et Astronautica Sinica, 2021,42(8):87–98.)

作者简介：

方胜良(1968–)，男，博士，教授，博士生导师，主要研究方向为人工智能、电磁态势可视化.email: eeifslyl@163.com.

胡豪杰(1992–)，男，在读硕士研究生，主要研究方向为电磁态势可视化.