

## 面向侦收脉冲序列的相控阵雷达波位划分方法：专家知识与强化学习混合框架

余壮<sup>①②</sup> 凌青<sup>\*①②</sup> 闫文君<sup>\*①②</sup> 董世鹏<sup>③</sup> 张立民<sup>①②</sup> 商允力<sup>④</sup>

<sup>①</sup>(海军航空大学信息融合研究所 烟台 264001)

<sup>②</sup>(山东省海空信息感知与处理技术重点实验室 烟台 264001)

<sup>③</sup>(西安电子科技大学雷达信号处理全国重点实验室 西安 710126)

<sup>④</sup>(中国人民解放军92038部队 青岛 266071)

**摘要：**相控阵雷达因其波束灵活扫描、多模式快速切换和参数捷变特性，使得传统基于参数聚类的雷达信号分析方法面临特征参数不稳定、参数空间重叠等问题。基于此，该文从波位划分角度入手进行相控阵雷达信号分析，即从混合脉冲流中还原出属于不同波束位置的脉冲子序列，创造性地提出了一种专家知识与强化学习混合(EK-HRL)框架。该框架首先基于脉冲幅度动态门限对波位进行初步划分，然后将初步划分结果输入人在回路强化学习环境，结合专家知识引导与置信度评估，最终实现波位的精细划分。仿真数据集实验表明：所提方法的波位划分精确率达到92.7%，置信度评估模型表现出良好的校准性，该方法为人机协同解决复杂电磁信号处理问题提供了一种有效的技术路径。

**关键词：**相控阵雷达；波位划分；专家知识；强化学习；置信度

**中图分类号：**TN957.51

**文献标识码：**A

**文章编号：**2095-283X(2026)x-0001-15

**DOI:** 10.12000/JR25283

**CSTR:** 32380.14.JR25283

**引用格式：**余壮, 凌青, 闫文君, 等. 面向侦收脉冲序列的相控阵雷达波位划分方法：专家知识与强化学习混合框架[J]. 雷达学报(中英文), 待出版. doi: 10.12000/JR25283.

**Reference format:** YU Zhuang, LING Qing, YAN Wenjun, *et al.* Phased-array radar beam position partitioning for intercepted pulse sequences: An expert knowledge-based hybrid reinforcement learning framework[J]. *Journal of Radars*, in press. doi: 10.12000/JR25283.

## Phased-array Radar Beam Position Partitioning for Intercepted Pulse Sequences: An Expert Knowledge-based Hybrid Reinforcement Learning Framework

YU Zhuang<sup>①②</sup> LING Qing<sup>\*①②</sup> YAN Wenjun<sup>\*①②</sup> DONG Shipeng<sup>③</sup>  
ZHANG Limin<sup>①②</sup> SHANG Yunli<sup>④</sup>

<sup>①</sup>(*Institute of Information Fusion, Naval Aviation University, Yantai 264001, China*)

<sup>②</sup>(*Key Laboratory of Sea-Air information Perception and Processing Technology of Shandong Province, Yantai 264001, China*)

<sup>③</sup>(*National Key Laboratory of Radar Signal Processing, Xidian University, Xi'an 710126, China*)

<sup>④</sup>(*Unit 92038, People's Liberation Army, Qingdao 266071, China*)

收稿日期: 2025-12-31; 改回日期: 2026-03-23; 网络出版: 2026-05-06

\*通信作者: 凌青 [linqing19870522@163.com](mailto:linqing19870522@163.com); 闫文君 [wj\\_yan@foxmail.com](mailto:wj_yan@foxmail.com)

\*Corresponding Authors: LING Qing, [linqing19870522@163.com](mailto:linqing19870522@163.com); YAN Wenjun, [wj\\_yan@foxmail.com](mailto:wj_yan@foxmail.com)

基金项目: 国家自然科学基金(62371465), 泰山学者工程专项经费基金(ts201511020), 山东省青创团队资助(2022KJ084)

Foundation Items: The National Natural Science Foundation of China (62371465), The Taishan Scholars Project Special Fund (ts201511020), Youth Innovation Teams in Shandong Province Fund (2022KJ084)

责任编辑: 易伟 Corresponding Editor: YI Wei

©The Author(s) 2026. This is an open access article under the CC-BY 4.0 License

(<https://creativecommons.org/licenses/by/4.0/>)

**Abstract:** The characteristics of phased-array radars—including flexible beam scanning, rapid multimode switching, and parameter agility—pose challenges to traditional radar signal analysis methods based on parameter clustering, causing feature parameter instability and parameter space overlap. To address these issues, this paper analyzes phased-array radar signals from the perspective of beam position partitioning. In particular, we reconstruct pulse subsequences corresponding to distinct beam positions from mixed pulse streams and an innovative expert-knowledge and hybrid-reinforcement-learning framework is proposed. This framework first performs preliminary partitioning using dynamic pulse amplitude thresholds. It subsequently feeds the preliminary results into a human-in-the-loop reinforcement-learning environment by integrating expert knowledge guidance with confidence assessment to ultimately achieve fine-grained beam position partitioning. Experimental results obtained using simulated datasets demonstrate that the proposed framework achieves a partitioning precision of 92.7%, indicating excellent calibration of the confidence assessment model. This work provides an effective technical pathway for human-machine collaboration in solving complex electromagnetic signal processing problems.

**Key words:** Phased array radar; Beam position partitioning; Expert knowledge; Reinforcement learning; Confidence

## 1 引言

相控阵雷达作为现代电子战中最重要信息获取手段,其信号分析对于威胁识别、态势感知和电子对抗具有关键意义。准确分析相控阵雷达信号能够有效识别雷达类型、推断工作模式、预测威胁等级,为电子支援和电子对抗提供决策依据,直接关系到战场生存能力和作战效能<sup>[1,2]</sup>。然而,相控阵雷达的参数捷变、波束快速切换等特性,使传统基于参数聚类的信号分析方法面临特征参数不稳定、参数边界模糊等挑战<sup>[3,4]</sup>。在多波束交织、参数重叠的复杂电磁环境下,仅依靠脉冲描述字(Pulse Descriptor Word, PDW)的统计特征已难以理清信号组织结构<sup>[5]</sup>。

在这一背景下,波位(Beam Position)作为相控阵雷达资源调度与任务执行的基本时空单元,其重要性日益凸显。Zhang等人<sup>[6]</sup>研究指出,现代雷达资源管理已从固定扫描周期转向基于任务优先级的动态驻留调度,这意味着波位级序列才是雷达战术意图在物理层的直接体现,而单一脉冲缺乏独立的战术语义。在执行搜索、跟踪、边扫边跟等多功能任务时,相控阵雷达会在不同波位间快速切换,形成参数复杂、时序交织的波位序列<sup>[7,8]</sup>,这些序列直接映射了雷达的扫描策略及对特定空域的关注程度。Krishnamurthy等人<sup>[9]</sup>在认知电子战研究中强调,只有准确还原波位序列,才能推理识别雷达的战术意图,从而为后续的干扰决策提供依据。因此,波位划分(Beam Position Partitioning),即从混合脉冲流中还原出属于不同波位的脉冲子序列,成为连接底层信号分选与上层意图识别的关键纽带,其准确性直接影响后续工作模式识别、特定辐射源识别等任务的成败<sup>[10]</sup>。

尽管波位划分具有重要价值,但该问题在相控阵雷达信号分析领域尚未得到系统性研究。当前研究主要方向是从雷达系统设计角度研究波束管理和资源调度算法<sup>[11,12]</sup>,侧重于优化雷达性能而非信号分析。

波位划分可以被视为一个高维参数空间中的聚类问题,即试图将具有相似参数特征的脉冲归为同一类。然而,现有的基于参数聚类的方法在处理现代相控阵雷达信号时具有一定的局限性,主要体现在以下3个方面:一是参数空间重叠与快速捷变,相控阵雷达采用脉冲重复间隔参差、频率捷变等反侦察措施,导致不同波位参数边界模糊<sup>[13,14]</sup>。二是忽视时序动态性,静态方法将脉冲视为独立点,丢失了波位切换的时序关联信息<sup>[15]</sup>。三是超参数敏感,传统聚类算法难以自适应确定最佳参数,在多辐射源或高噪环境下泛化能力差<sup>[16]</sup>。

针对上述静态聚类方法的局限性,Tan等人<sup>[17]</sup>尝试深度学习方法来处理复杂调制信号,利用双谱特征从多通信场景中提取雷达信号特征。进一步,研究者开始引入具备长程依赖捕捉能力的序列模型,如Transformer架构,用以解决复杂序列分割问题<sup>[18]</sup>。然而,将这些深度模型直接应用于波位划分仍面临两大困境:一是纯数据驱动模型难以理解雷达信号背后的物理机理,容易产生不可解释的错误决策<sup>[19]</sup>;二是高质量的波位标注数据在军事领域极其稀缺,依靠海量数据训练的大模型难以适应“小样本”场景<sup>[20,21]</sup>。

为突破现有静态聚类与纯监督学习方法的局限,本文摒弃了将波位划分视为0-1标签分配等传统方法,将其建模为粗划分、精修正的序贯决策过程。在此视角下,强化学习(Reinforcement Learning, RL)提供了理想的解决方案:一方面,RL允许通过

内嵌物理约束的收益函数来引导策略寻优，从而摆脱了对海量脉冲级真值标签的依赖，有效克服了非合作侦察环境下高质数据稀缺的难题；另一方面，RL固有的序贯决策机制能够模拟人类专家观察、研判、修正的认知回路，实现复杂场景下波位的精细划分。此外，针对传统强化学习在探索初期因缺乏物理常识导致学习效率低下、复杂样本精度不足的问题，本文借鉴“人在回路”(Human-in-the-Loop, HITL)机制<sup>[22]</sup>在机器人控制<sup>[23]</sup>、自动驾驶<sup>[24]</sup>、游戏AI<sup>[25]</sup>等领域成果的经验，将HITL机制引入波位划分任务。该机制通过将专家反馈转化为奖励函数信号，能够有效弥补纯数据驱动方法的局限，成为突破上述性能瓶颈的关键。

基于上述现状，本文针对阵面物理固定、波束二维电子扫描的相控阵雷达信号，下文简称“二维固定相控阵雷达”信号，提出一种基于专家知识与强化学习混合(Expert Knowledge-based Hybrid Reinforcement Learning, EK-HRL)框架的相控阵雷达波位划分方法，该方法通过将人类专家的知识判断融入机器学习流程，有效缓解了数据标注困难和模型可信度低的问题。本文主要工作包括：根据相控阵雷达波束扫描时脉冲幅度随波位切换而变化的物理特性<sup>[26]</sup>，提出基于脉冲幅度动态门限的波位划分基线(Pulse-Amplitude Dynamic-Threshold-Based Beam Position Partitioning, PADT-BPP)模型，提供初始划分结果。在此基础上，构建了基于人在回路强化学习的波位划分(Human-in-the-Loop Reinforcement Learning-based Beam Position Partitioning, HIRL-BPP)模型，以及加速HIRL-BPP模型收敛的基于统计量的波位置信度评估(Statistics-based Beam Position Confidence Assessment, SBPCA)模型。本文的主要创新为：

针对强化学习中稀疏奖励导致的信用分配问题<sup>[27]</sup>，提出了一种“物理-知识双驱动”的四维复合奖励函数：以专家反馈形成的基础奖励为主干，引入波位内参数稳定性、相邻波位边界可分性和波

位规模合理性3类物理约束分量，并由SBPCA置信度评估模型对各分量权重进行自适应调节。该设计将置信度从结果评估指标提升为参与策略优化的核心变量，在保证物理可解释性的同时缓解了奖励稀疏问题，加速强化学习收敛，并提升模型在复杂参数交叠场景下的波位划分鲁棒性。

## 2 问题建模与理论基础

波位划分的核心目标是从截获的相控阵脉冲序列中，分离出属于不同波位的脉冲子序列，得到互不相交的独立波位集合 $\mathcal{G}_{\text{seq}} = \{G_t\}_{t=1}^N$ ，其中每个子集 $G_t$ 对应一个独立的雷达波位。波位划分过程如图1所示。

为将波位划分问题转化为序贯决策问题，本文将建模为单智能体马尔可夫决策过程(Markov Decision Process, MDP) 5元组 $(S, A, P, R, \gamma)$ ，其中， $S$ 表示状态空间，状态 $s_t \in S$ 是对第 $t$ 个待评估波位提取的13维统计特征向量，具体定义如表1所示；动作空间定义为 $A = \{\text{Keep, Merge, Split}\}$ ，分别表示保持、合并、分割当前波位 $G_t$ ，作用对象是波位片段，仅在“Split”时需要额外给出一次切分位置 $k$ ，即在第几个脉冲进行分割； $R$ 为奖励； $P$ 为状态转移概率； $\gamma$ 为折扣因子，用于平衡即时奖励与长期奖励。该MDP决策过程目标是寻找一个最优策略 $\pi: S \rightarrow A$ 使得从任意状态 $s_t \in S$ 出发的累计折扣奖励的期望得以最大化，即 $V(s) = \max_{a \in A} Q(s_t, a)$ ，该策略满足贝尔曼最优方程：

$$Q(s_t, a) = E_{s' \sim P(s'|s, a)} [r(s_t, a) + \gamma \cdot V(s')] \quad (1)$$

其中， $V(s)$ 表示最优状态价值函数， $Q(s_t, a)$ 为最优状态-动作价值函数， $E_{s' \sim P(s'|s, a)}$ 为状态转移期望， $r(s_t, a)$ 为即时奖励函数。

其中， $x \in \{\text{PW, RF, PA, PRI}\}$ ，PW为脉宽(Pulse Width)，RF为射频(Radio Frequency)，PA为振幅(Pulse Amplitude)，PRI为脉冲重复间隔(Pulse Repetition Interval)， $n$ 为当前波位脉冲数量。

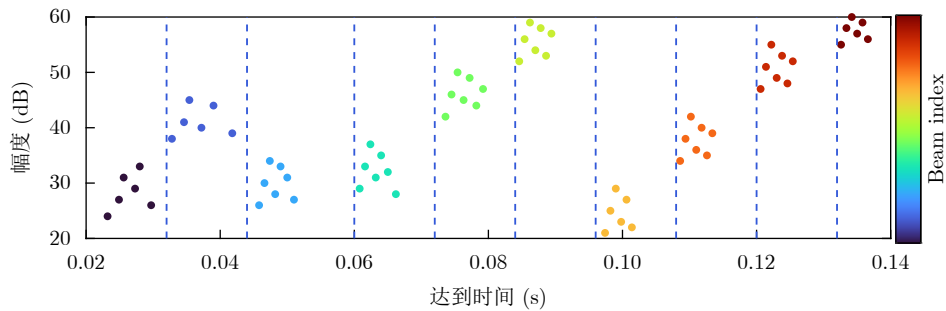


图 1 波位划分概念图

Fig. 1 Schematic diagram of beam position partitioning

表 1 13维特征  
Tab. 1 The 13-dimensional features

特征	计算公式
波位内脉冲数	$n$
均值	$\mu_x = \frac{1}{n} \sum_{i=1}^n x_i$
标准差	$\sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)^2}$
变异系数	$CV_x = \frac{\sigma_x}{\mu_x + \varepsilon}$

MDP建模依赖于状态空间的马尔可夫属性<sup>[28]</sup>, 本文参照文献<sup>[29]</sup>对脉冲序列的处理, 对波位序列进行马尔可夫属性合理性分析。

(1) 历史无关性: 通过分析信息转移链  $\mathcal{I}_t \rightarrow \mathbf{s}_t \xrightarrow{a_t} \mathcal{I}_{t+1} \rightarrow \mathbf{s}_{t+1}$  可知历史信息已通过特征提取压缩至状态  $\mathbf{s}_t$ , 使得当前状态只取决于上一状

态, 而与其历史状态无关, 满足历史无关性, 其中  $\mathcal{I}_t = \{G_{t-1}, G_t, G_{t+1}\}$  为局部信息集。

(2) 特征完备性: 本文假设上述13维特征能够充分捕捉支持波位划分决策的相关信息, 该假设的有效性将在第4节通过实验进行验证。

综上, 通过定义MDP 5元组, 本文将相控阵雷达侦收信号的波位划分问题转化为可通过强化学习求解的序贯决策问题。

### 3 专家知识与强化学习混合框架

在将波位划分问题建模为MDP决策过程的基础上, 本文提出专家知识与强化学习混合(EK-HRL)框架, 该框架采用离线训练、在线推理的工作模式, 通过将领域专家的先验知识与强化学习的自适应决策能力相结合, 创造性地将人在回路强化学习方法应用于波位划分问题。如图2所示, 专家知识

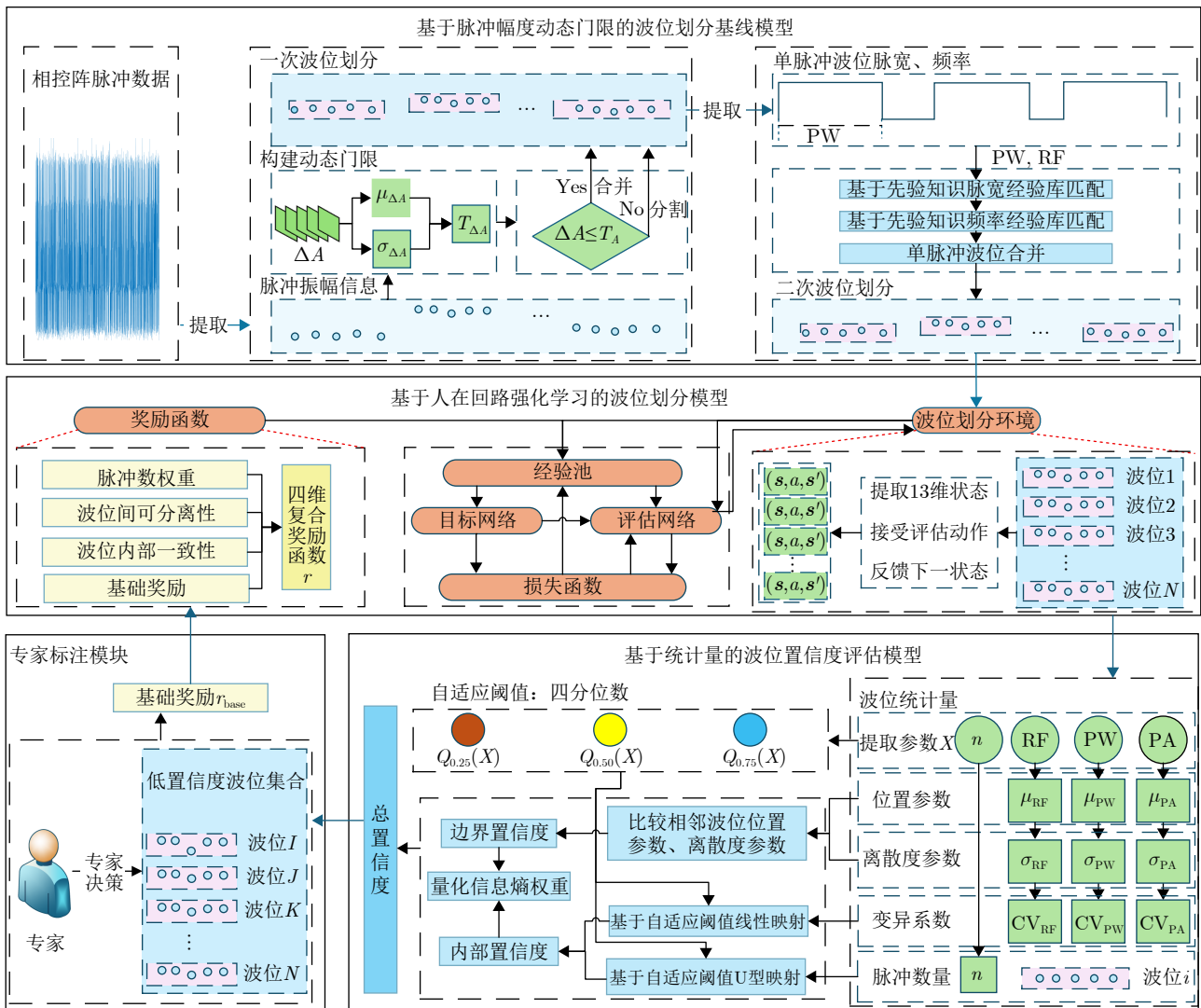


图 2 专家知识与强化学习混合框架

Fig. 2 Hybrid framework of EK-HRL

与强化学习混合框架由4部分构成：基于脉冲幅度动态门限的波位划分基线(PADT-BPP)模型、基于人在回路强化学习的波位划分(HIRL-BPP)模型、基于统计量的波位置信度评估(SBPCA)模型以及专家标注模块。

具体而言，在离线训练阶段，基线模型首先对原始脉冲数据流进行预处理，生成初步划分的波位子集  $C = \{C_1, C_2, \dots, C_N\}$ ，其中， $N$ 表示波位总数， $C_t$ 表示第 $t$ 个波位片段。该划分结果作为HIRL-BPP模型的初始输入。HIRL-BPP模型阶段，专家介入采用轮级的批处理式介入机制，置信度评估模型会评估每轮波位划分结果的置信度，专家将针对性修正低置信度波位划分决策。值得强调的是置信度评估模型在为决策主体提供反馈的同时，也通过奖励函数塑形学习策略，从而提升模型的收敛速度和划分精度。在线部署阶段，模型独立完成波位划分，无需专家实时参与。

### 3.1 波位划分基线模型

本文根据二维固定相控阵雷达波束扫描时脉冲幅度随波位切换而变化的物理特性，结合脉宽、射频等多维参数聚类，按照幅度相近、时间连续的原则，构建了PADT-BPP模型，该模型在克服传统静态阈值、单参数聚类局限的同时，又规避了深度学习的数据依赖问题。该模型计算复杂度低，可为后续智能优化提供高质量的初始解。该模型的具体计算方法如下：

(1) 构建动态门限 $T_{\Delta A}$ 作为波位划分的自适应阈值。构建一个滑窗对信号幅度差值 $\Delta A_j = |A_{j+1} - A_j|$ 进行分析，其中 $A_j$ 为第 $j$ 个脉冲的幅度值，动态计算门限以适应信号变化。滑窗设定为 $W$ 个脉冲，则动态门限计算方法如下：

$$T_{\Delta A} = \mu_{\Delta A} + k\sigma_{\Delta A} \quad (2)$$

其中， $\mu_{\Delta A}$ 与 $\sigma_{\Delta A}$ 为窗口内幅度差的均值和标准差， $k$ 为门限因子。

(2) 相邻脉冲幅度差值的比较，对于待测信号序列的任意两个相邻脉冲，计算其幅度差值 $\Delta A$ ，并进行以下操作：如果 $\Delta A \leq T_{\Delta A}$ ，则认为脉冲幅度相近，归为一个波位；如果 $\Delta A > T_{\Delta A}$ ，则认为脉冲幅度变化过大，开始新的波位。

(3) 基于脉宽频率特征对波位进行二次波位划分与合并，结合脉宽、频率典型范围，对初步划分存在明显参数分布差异的，进一步分割。若初步划分的单脉冲波位的脉宽、频率与相邻波位相近，则进行合并。

仅仅依赖波位参数进行划分，往往存在抗噪声

能力不足的短板，对于信噪比低，异常值突出的波位往往会出现过分割与欠分割的问题，基于这一问题本文进一步引入强化学习进行波位划分。

### 3.2 基于人在回路强化学习的波位划分模型

本文采用深度Q网络(Deep Q-Network, DQN)进行波位划分决策，不同于监督学习对静态特征的拟合，DQN将波位划分视为一个在非平稳环境下的连续决策问题。智能体通过与波位划分环境交互，学习的是“如何修正波位”的策略而非固定的划分结果，这使得模型在面对雷达参数捷变时具有更强的适应性。但纯数据驱动的DQN面临探索盲目性高、样本利用率低等问题。为此，本文引入人在回路强化学习(HITL-RL)，通过专家指导信号改善学习效率，构建了HIRL-BPP模型。该模型有效解决异常值突出环境下波位划分准确率低的问题。

具体而言，专家介入采用置信度触发机制：当SBPCA模型输出的波位置信度 $C_{\text{total}}(\mathbf{s}_t, a_t) \leq \tau_{\text{conf}}$ 时，其中 $\tau_{\text{conf}}$ 为决策阈值，触发专家对当前波位片段进行审核。对于二维固定相控阵雷达，波束在单一波位驻留期间指向恒定，接收幅度相对稳定，波位切换时指向改变导致幅度突变。基于该物理特性，专家判决遵循“幅度相近、时间连续”原则，对易混淆波位结合脉宽与频率进行辅助决策。专家给出修正动作 $a_{\text{expert}} \in \{\text{Keep, Merge, Split}\}$ ，若选择Split则需额外指定切分位置 $k$ 。该修正动作通过两条路径融入学习过程：一是作为监督样本 $(\mathbf{s}, a, r, \mathbf{s}')$ 写入经验回放池，供DQN采样学习；二是通过专家决策奖励 $r_{\text{base}}$ 进入奖励函数，引导策略优化。当 $C_{\text{total}}(\mathbf{s}_t, a_t) \geq \tau_{\text{conf}}$ 时，该波位具有较高置信度，由物理约束奖励主导自主探索，无需专家介入。

模型结构如图3所示，其中相控阵波位划分环境作为智能体的交互对象，输出当前波位的状态特征 $\mathbf{s}$ ，并接收评估网络的动作决策 $a$ ，反馈下一状态 $\mathbf{s}'$ 与环境奖励 $r$ ，构成 $(\mathbf{s}, a, r, \mathbf{s}')$ 的交互闭环。

模型学习框架由评估网络、目标网络、经验池3部分构成。评估网络采用64维隐藏层的全连接架构，将13维的状态向量，映射为动作价值函数 $Q(\mathbf{s}_t, a_t; \theta)$ ，并采用 $\epsilon$ -greedy探索策略平衡探索与利用。目标网络参数定期从评估网络参数复制，用于计算TD目标值 $y_t$ ，通过解耦网络参数以抑制训练震荡。经验池存储元组 $(\mathbf{s}, a, r, \mathbf{s}')$ 并支持随机采样。最终，模型通过梯度下降最小化预测 $Q(\mathbf{s}, a; \theta)$ 与目标 $Q_{\text{target}}(\mathbf{s}', a'; \theta')$ 值之间均方误差，实现参数优化。

### 3.3 置信度驱动的四维复合奖励函数

在波位划分问题中，奖励函数的设计是HITL-

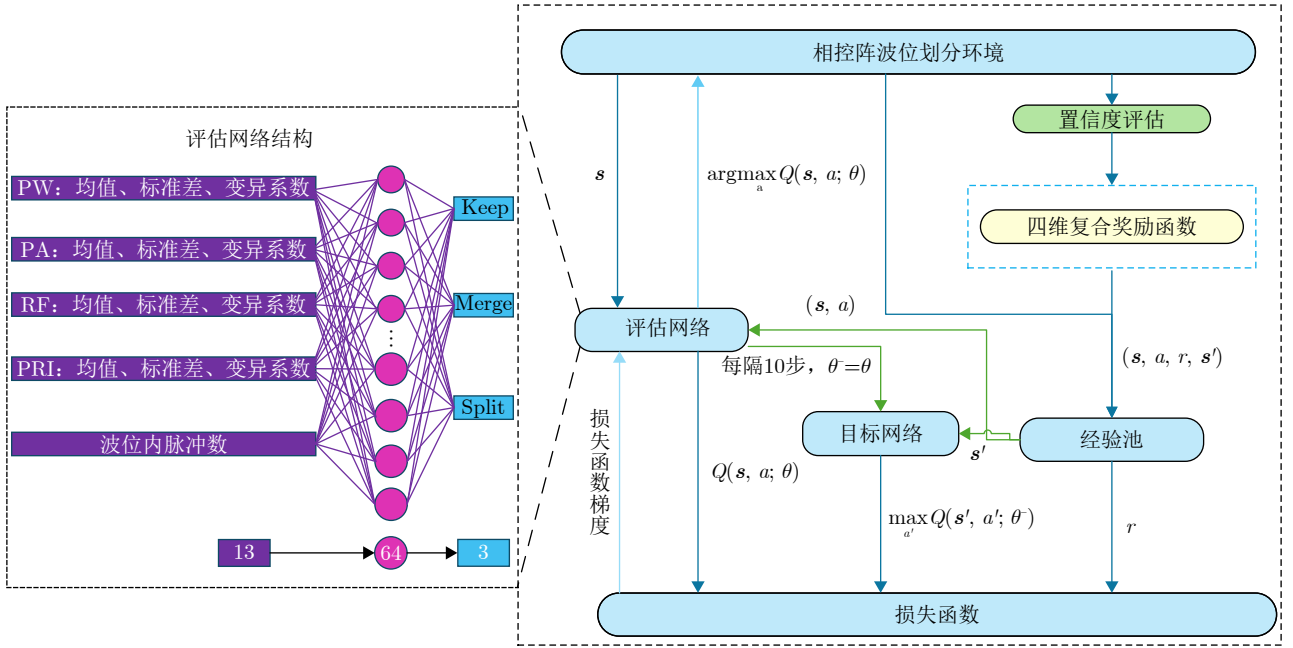


图3 基于人在回路强化学习的波位划分模型

Fig. 3 HIRL-BPP model

RL在复杂物理系统中落地应用的核心瓶颈。针对HITL-RL在波位划分任务中稀疏二进制反馈效率低及物理先验知识缺失问题，划分结果违背信号固有属性、忽略波位尺度偏差导致划分准确度低等不足，本文受物理信息机器学习<sup>[30]</sup>与奖励塑形<sup>[31]</sup>理论的启发，创造性地设计了一种置信度驱动的四维复合奖励机制，将SBPCA模型的置信度评估转化为智能体的内在动机，构建融合专家引导与物理一致性约束的奖励函数。其构建过程包含置信度评估与奖励函数合成两个阶段。

### 3.3.1 波位置置信度评估模型

本节构建SBPCA模型，旨在将领域专家对雷达信号“波位内参数稳定”与“边界清晰”的定性认知，转化为定量评价指标。

(1) 物理特征映射与置信度量：模型首先统计波位 $G_t$ 的PW, PA, RF以及PRI四维参数的变异系数 $CV_x$ ，其中对RF进行归一化处理，以消除量纲影响。此外，为消除离群点对置信度评估的偏差，引入Winsorization截尾处理技术<sup>[32]</sup>。当变异系数低于下四分位数 $Q_1$ 时，波位划分质量保持稳定；超过上四分位数 $Q_3$ 后，表示划分质量显著下降。为捕捉该特性，本文构建了基于经验分位数的分段映射函数 $M_{stab}(CV_x)$ ：

$$M_{stab}(CV_x) = \begin{cases} 1, & CV_x \leq Q_1 (\text{Tolerance Region}) \\ \frac{Q_3 - CV_x}{Q_3 - Q_1}, & Q_1 < CV_x < Q_3 (\text{Penalty Region}) \\ 0, & CV_x \geq Q_3 (\text{Cutoff Region}) \end{cases} \quad (3)$$

该函数通过容忍区间保持稳定、惩罚区间线性衰减、截断区间施加硬约束的3段式设计，能够有效平衡对系统误差的鲁棒性与对异常波动的敏感性。

此外，波位划分过程中常见过分割与欠分割问题：过分割指将单个真实波位错误地切分为多个片段，欠分割则指将多个波位的脉冲错误地合并为单个片段。过分割与欠分割问题都会使划分结果的脉冲数量产生偏差，基于此，本文采用倒U型约束函数 $M_{size}(n)$ 评估波位 $G_t$ 脉冲数量 $n$ 是否处于合理范围：

$$M_{size}(n) = \exp\left(-\frac{(n - \mu_{opt})^2}{2\sigma_{tol}^2}\right) \quad (4)$$

其中， $\mu_{opt} = (Q_{0.25}(n) + Q_{0.75}(n))/2$ 为最优规模期望； $\sigma_{tol} = (Q_{0.75}(n) - Q_{0.25}(n))/2$ 为总体规模容差标准差， $Q_{0.25}(n)$ 和 $Q_{0.75}(n)$ 分别为划分波位脉冲数分布的第一、第三四分位数。该函数对脉冲数 $n$ 进行评估：当 $n$ 接近最优值 $\mu_{opt}$ 时输出高置信度，偏离越远置信度衰减越快。

(2) 信息融合：为消除主观赋权偏差，构建完全数据驱动的评价体系：

内部置信度，引入Shannon信息熵理论<sup>[33]</sup>量化各特征的权重。首先，对各特征映射值进行归一化处理：

$$p_{xt} = \frac{M_{stab}(CV_x(G_t))}{\sum_{i=1}^N M_{stab}(CV_x(G_i))} \quad (5)$$

其中， $N$ 为波位样本总数， $G_t$ 为第 $t$ 个波位片段。其次，计算特征 $x \in \{\text{PW}, \text{RF}, \text{PA}, \text{PRI}\}$ 的信息熵： $H_x = -(\ln N)^{-1} \sum_{t=1}^N p_{xt} \ln p_{xt}$ ，然后计算信息效用值 $U_x = 1 - H_x$ ，对规模特征 $M_{\text{size}}(n)$ 同理计算 $U_n$ ，最后归一化得到自适应权重： $\omega_x = U_x / \sum U$ ， $\omega_n = U_n / \sum U$ ，其中 $\sum U = \sum U_x + U_n$ ，则内部置信度：

$$C_{\text{internal}}(G_t) = \sum \omega_x \cdot M_{\text{stab}}(\text{CV}_x(G_t)) + \omega_n \cdot M_{\text{size}}(n) \quad (6)$$

边界置信度，量化相邻波位 $G_t$ 与 $G_{t+1}$ 的区分程度，构建波位 $G_t$ 的八维特征向量 $v(G_t) = [\mu_{\text{PW}}, \mu_{\text{RF}}, \mu_{\text{PA}}, \mu_{\text{PRI}}, \sigma_{\text{PW}}, \sigma_{\text{RF}}, \sigma_{\text{PA}}, \sigma_{\text{PRI}}]_{G_t}$ ，其中 $\mu_x$ 和 $\sigma_x$ 分别表示波位 $G_t$ 内特征 $x$ 的均值和标准差。则相邻波位在特征空间中的欧氏距离：

$$D(G_t, G_{t+1}) = \sqrt{\sum_{d=1}^8 \left( \frac{v_d(G_t) - v_d(G_{t+1})}{\sigma_d} \right)^2} \quad (7)$$

其中， $\sigma_d$ 表示第 $d$ 维特征在全局数据集上的标准差。最后，利用Sigmoid函数将该距离映射为置信度得分：

$$C_{\text{boundary}} = \frac{1}{1 + \exp(-\gamma \cdot (D - D_{\min}))} \quad (8)$$

其中， $D_{\min}$ 为最小邻域距离， $\gamma$ 为陡峭度参数，本模型 $\gamma$ 取 $2/\sigma_D$ 。

最后进行全局置信度融合。 $C_{\text{internal}}(G_t)$ 与 $C_{\text{boundary}}(G_t)$ 分别刻画波位内部参数稳定性与相邻边界可分性，两者信息来源互补且非严格独立。为降低主观赋权偏差，此处采用式(6)相同计算方法计算融合权重 $\alpha$ 与 $\beta$ ，进一步则可得该波位的全局置信度：

$$C_{\text{total}}(G_t) = \alpha \times C_{\text{internal}}(G_t) + \beta \times C_{\text{boundary}}(G_t) \quad (9)$$

### 3.3.2 四维复合奖励函数

基于构建的SBPCA模型，设计了专家知识与物理驱动的复合奖励函数。对于波位 $G_t$ ，智能体获得的复合奖励为 $r_t$ ，定义为专家决策奖励 $r_{\text{base}}$ 与物理约束奖励 $r_{\text{phy}}$ 的动态组合：

$$r_t = \lambda_t \cdot r_{\text{base}}(\mathbf{s}_t, \mathbf{a}_t) + (1 - \lambda_t) \cdot r_{\text{phy}}(\mathbf{s}_t, \mathbf{a}_t) \quad (10)$$

其中， $\lambda_t \in [0, 1]$ 为介入判决系数，由波位置置信度 $C_{\text{total}}(G_t)$ 与决策阈值 $\tau_{\text{conf}}$ 定义：

$$\lambda_t = \mathbb{I}(C_{\text{total}}(\mathbf{s}_t, \mathbf{a}_t) \leq \tau_{\text{conf}}) \quad (11)$$

当 $\lambda_t = 1$ 时，专家资源将集中处理低置信度波位，通过这些困难样本上引入强监督信号，能够有效规避训练初期因盲目探索导致的训练效率低的

问题；当 $\lambda_t = 0$ 时，该波位划分具有较高置信度，采用由物理约束主导的自主探索。

(1) 专家决策奖励 $r_{\text{base}}$ ，引入汉明距离 $D_{\text{H}}$ 量化模型动作 $\mathbf{a}_t$ 与专家决策 $\mathbf{a}_{\text{expert}}$ 的差异：

$$r_{\text{base}}(\mathbf{s}_t, \mathbf{a}_t) = -\eta \cdot D_{\text{H}}(\mathbf{a}_t, \mathbf{a}_{\text{expert}}) \quad (12)$$

采用汉明距离直接度量专家修正决策与模型决策的差异，以确保离散动作的精确比较<sup>[34]</sup>。

(2) 物理约束奖励 $r_{\text{phy}}$ ，将波位内参数稳定性、波位间边界可分性、规模合理性3类物理约束映射至 $[-0.5, 0.5]$ 的对称区间，引入隐式基线 $b = 0.5$ ，则可得

$$r_{\text{consistency}} = C_{\text{internal}}(G_j) - 0.5 \quad (13)$$

$$r_{\text{separability}}(\mathbf{s}_t) = C_{\text{boundary}}(G_j) - 0.5 \quad (14)$$

$$r_{\text{size}} = M_{\text{size}}(n) - 0.5 \quad (15)$$

则进一步物理约束奖励为

$$r_{\text{phy}} = w_1 \cdot r_{\text{consistency}} + w_2 \cdot r_{\text{separability}} + w_3 \cdot r_{\text{size}} \quad (16)$$

其中， $w$ 为参数权重，其设计遵循雷达信号分选的处理优先顺序。正如Cheng等人<sup>[5]</sup>指出：在处理混合雷达信号时，应优先依据周期一致性锁定并剥离有效脉冲序列，随后再对剩余信号进行统计分离。基于此，本文在超参数空间内进行系统化网格搜索，通过验证集收敛残差评估不同权重组合，最终确定帕累托最优解 $\mathbf{w} = [0.4, 0.3, 0.3]$ 。

## 4 实验设计与结果分析

### 4.1 数据集描述

本数据集是在领域专家的指导下，通过高保真仿真构建而成的。数据建模严格遵循“阵面物理固定、波束二维电扫捷变”相控阵雷达的运行机制，即在单波位驻留期内，波束指向保持恒定；而在波位切换时，受雷达资源调度策略的影响，波束指向呈现随机跳变，使幅度变化表现出显著的非周期性随机特征。

数据以波位为基本单元构建。为模拟雷达的多模式工作特性，预设了4种波形参数配置，每种配置载频范围为5300~5700 MHz、脉宽范围1~2  $\mu\text{s}$ 、每个波位脉冲个数4~7个、PRI体制包含固定和参差两种类型，信噪比为15~35 dB，均值25 dB。在波位调度过程中，每个波位从这4种配置中随机选择一种并随机决定波束指向。

为验证算法鲁棒性，引入以下异常：波位丢失率为10%，脉冲分裂率为20%，随机插入50个干扰脉冲，PDW参数抖动服从高斯分布，标准差与信

号幅度呈负相关。本实验训练数据与测试数据集分别包含16229, 15417个按时间排序的脉冲样本, 分别调度3565, 3103次波位, 由此得到波位级样本数3565, 3103, 数据集根据波位切换过程标注真值波位标签, 为波位划分算法的性能评估提供参考, 并按照8:2比例将训练数据分为训练集与验证集。

## 4.2 评估指标体系

(1) 波位划分准确性指标: 采用标准的精确率(Precision)、召回率(Recall)以及F1-Score衡量波位划分的有效性与检测能力。

(2) 置信度评估模型性能指标: 为量化模型置信度的可靠性, 引入期望校准误差(Expected Calibration Error, ECE)、置信度-准确率相关系数(Confidence-Accuracy Correlation, CAC)以及置信度区分度(Confidence Discrimination, CD)。其中期望校准误差用于衡量预测置信度与真实准确率的偏差, 值越小表示校准性越好, 定义为

$$ECE = \sum_{m=1}^M \frac{|B_m|}{N} |\text{acc}(B_m) - \text{conf}(B_m)| \quad (17)$$

其中,  $M$ 为置信度区间数,  $B_m$ 为第 $m$ 个置信度区间的波位数,  $N$ 为总波位数,  $\text{acc}(B_m)$ ,  $\text{conf}(B_m)$ 分别为第 $m$ 个置信度区间内波位划分的实际准确率与平均置信度。置信度-准确率相关系数定义为置信度 $C$ 与准确率 $A$ 的皮尔逊相关系数 $CAC = \rho(C, A)$ , 用于评估二者的线性相关程度。置信度区分度, 表征模型区分正确与错误划分波位的能力, 定义为正确划分与错误划分波位的平均置信度之差:

$$CD = E[C|y = 1] - E[C|y = 0] \quad (18)$$

## 4.3 决策阈值敏感性实验

为确定EK-HRL框架最佳决策阈值 $\tau_{\text{conf}}$ , 分析不同 $\tau_{\text{conf}}$ 对于模型精度的影响并量化训练过程中专家介入频次, 本实验在训练过程中分别设定不同的决策阈值 $\tau_{\text{conf}}$ 进行训练, 验证在相同测试集上的划

分性能。EK-HRL框架配置如表2所示, 实验结果如图4所示。

由图4分析可知, 随着阈值增大, 训练过程平均累计标注量呈指数增长趋势,  $\tau_{\text{conf}} = 0.9$ 时, 达到472.1波位/千脉冲, 约为 $\tau_{\text{conf}} = 0.2$ 的19.7倍。收敛轮次呈U型分布, 且图5直观展示了各阈值训练过程中Precision的变化趋势。在 $\tau_{\text{conf}} \leq 0.5$ 的低阈值区收敛缓慢且最终性能受限, 在 $\tau_{\text{conf}} \geq 0.8$ 的高阈值区虽初期提升较快但后期出现明显的性能回落现象,  $\tau_{\text{conf}} = 0.7$ 在第25轮收敛时达到全局最优。

从探索-利用权衡角度分析, 低阈值策略会导致仅对极低置信度样本触发专家介入, 而大量中等置信度的边界模糊样本只能依赖物理约束奖励 $r_{\text{phy}}$ 进行判决, 缺乏专家决策奖励 $r_{\text{base}}$ 的精确指导, 导致模型在困难样本上的学习效果受限。高阈值策略则使专家过度介入高置信度样本, 模型过拟合专家决策模式而丧失自主探索能力, 造成标注资源浪费的同时影响准确率。综合分析 $\tau_{\text{conf}} = 0.7$ 实现监督信号与自主探索的最优平衡。

## 4.4 专家偏差敏感性分析

为评估EK-HRL对专家偏差的鲁棒性, 本文设计专家偏差敏感性分析实验: 在训练中, 对 $\tau_{\text{conf}} \leq 0.7$ 低置信度样本以5%, 10%, 20%进行错误标注, 训练结束后在固定测试集上评估Precision。结果如图6所示, 结果表明即使在20%偏差下, EK-HRL仍高于HRL-BPP的0%偏差下的表现, 表明框架具有良好的鲁棒性。专家偏差仅影响低置信度样本子集, 而高置信度样本由物理约束奖励 $r_{\text{phy}}$ 驱动, 有效抑制专家偏差对整体训练效果的影响。

## 4.5 对比实验

为比较本文EK-HRL框架与传统强化学习PPO模型<sup>[35]</sup>、SAC-D模型<sup>[36]</sup>在波位划分模型上的表现, 进行整体性能分析, 其中, PPO代表经典的On-policy策略梯度方法, SAC-D代表典型Off-policy方法,

表 2 对比实验参数

Tab. 2 Parameters of comparative experiments

方法	网络结构	奖励函数	置信度	其他参数
PADT-BPP模型	---	---	---	滑窗大小 $W = 5$ 、门限因子 $k = 2.0$ ;
SAC-D模型	13→64→3	$r_{\text{phy}}$	---	学习率0.0003, 批次大小64, 折扣因子 $\gamma=0.99$ , 经验池1,000,000, 更新步长4, 更新轮次1, 目标网络更新频率8000, 初始随机步数20,000, 奖励裁剪 $[-1,+1]$ ;
PPO模型	13→64→3	$r_{\text{phy}}$	---	学习率0.0003, 批次大小256, 更新步长1024, 折扣因子 $\gamma=0.99$ , 裁剪系数0.1, 更新轮数3, 熵系数0.01, 总迭代50轮;
DQN模型	13→64→3	$r_{\text{phy}}$	---	
HRL-BPP模型	13→64→3	$r_t$	---	学习率0.001, 批次大小32, 折扣因子 $\gamma=0.99$ , 经验池10,000, $\epsilon$ -greedy探索概率 $\epsilon$ 从1.0→0.1, 总迭代50轮, EK-HRL框架决策阈值 $\tau_{\text{conf}} = 0.7$
EK-HRL框架	13→64→3	$r_t$	SBPCA模型	

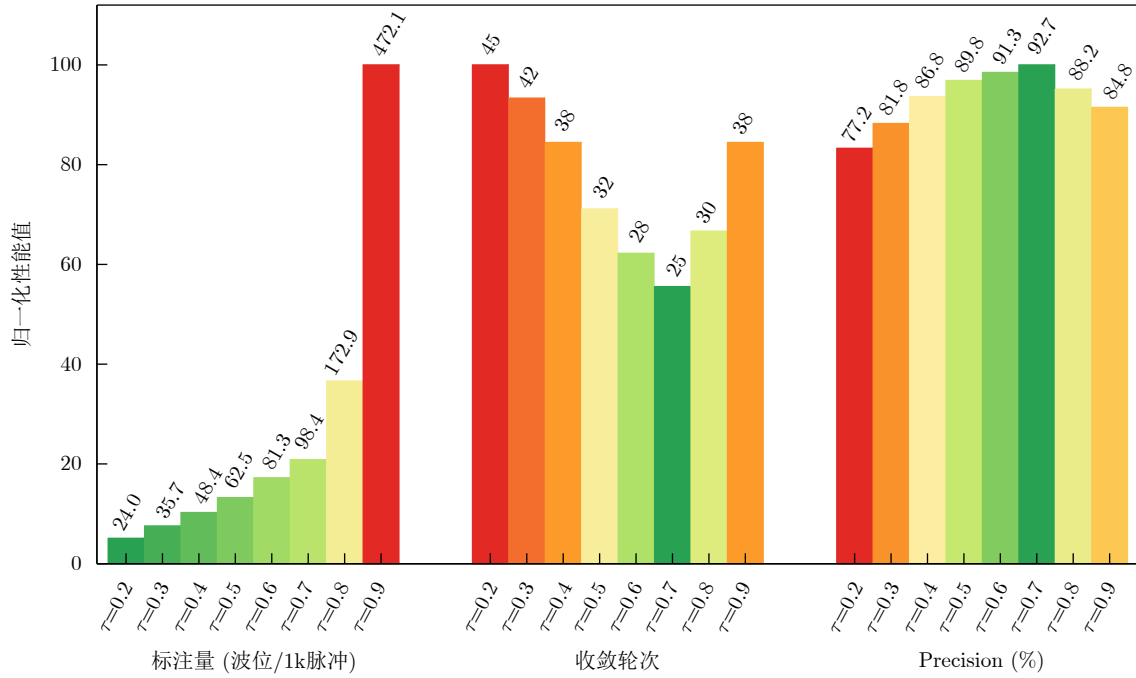


图 4 不同阈值下各统计量对比

Fig. 4 Comparison of statistical metrics under different confidence thresholds

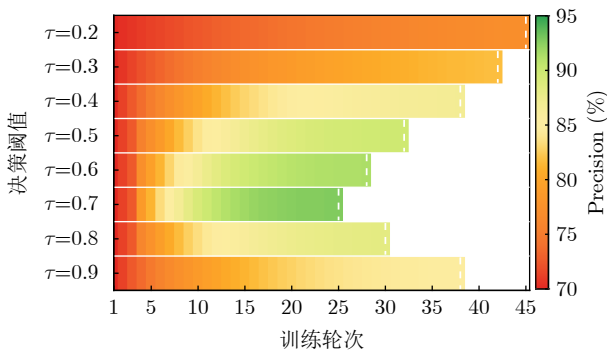


图 5 各阈值训练过程Precision热力图

Fig. 5 Heatmap of precision evolution during training across different decision thresholds

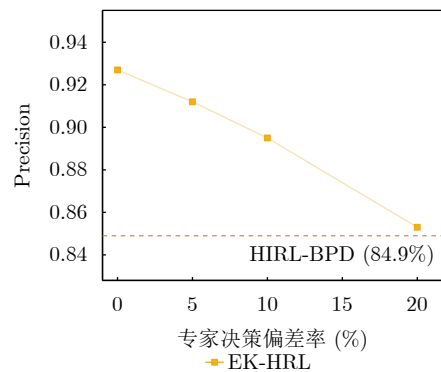


图 6 专家偏差敏感性分析

Fig. 6 Expert bias sensitivity analysis

确保实验完备性。此外，为验证EK-HRL框架核心模块的贡献，对EK-HRL框架进行消融实验，分析PADT-BPP模型、DQN模型、HIRL-BPP模型的作用与必要性。各模型参数配置如表2所示，其中PPO参数依照文献[35]中离散任务参数设定、SAC-D模型参数依照文献[36]进行设定。

#### 4.5.1 整体性能分析

在独立重复条件下各进行20次实验，得到如图7所示的Precision-训练轮次曲线。结果表明，EK-HRL框架收敛速度最快，约25轮即达到稳定状态，且误差带最窄，训练稳定性最优。将训练好的模型在统一的测试集上验证得到实验结果如表3所示，其中推理时间是模型在测试集上的计算耗时，不含专家决策时间。从图8可以清晰地观察到，PPO模型、

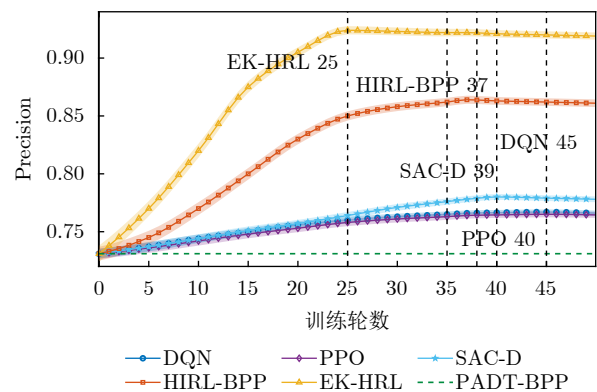


图 7 多次重复实验的Precision-训练轮次曲线

Fig. 7 Precision-epoch curves over multiple runs

SAC-D模型虽然计算开销上较小，在“推理速度”维度略占优势，但其在精确率维度显著低于EK-HRL框架。相比之下，本文提出的EK-HRL框架波位划

分性能最佳,划分精确率达到92.7%。图9进一步显示EK-HRL框架的划分结果在高置信度区间集中度最高,表明该框架在保证满足雷达实时处理精度要求的同时,实现了多性能指标的均衡最优。

实验结果表明,基于13维特征的EK-HRL框架在划分性能上均优于对比方法,说明所构建的13维特征能够有效捕捉支持波位划分决策的关键信息,验证了第2节提出的特征完备性假设。

### 4.5.2 各模块性能分析

为了验证EK-HRL框架中各核心组件的有效性,特别是专家知识与置信度评估机制对模型性能的贡献,本节设计了逐步递进的消融实验。对比分析显示,从代表传统规则的PADT-BPP模型、纯强化学习的DQN模型到人机回环的HIRL-BPP模型,最终至专家与置信度深度融合的EK-HRL框架,波位划分的各项指标均实现了稳步提升。实验结果有力地证明了,单纯依赖数据驱动或规则引导均难以应对复杂的波位边界,而EK-HRL通过置信

度评估与专家修正机制,显著提升了决策的准确率,验证了该框架设计的合理性与必要性。

如表3所示,PADT-BPP模型Precision为71.2%,F1-Score为0.706,推理时间最短。该方法能够一定程度上划分真实波位边界,但因缺乏神经网络与强化学习能力,仅依赖规则化机制,在波位边界模糊区域存在大量误判。如图10所示,竖线表示错误划分边界,相比波位标签信息Ground Truth,PADT-BPP的决策序列出现大量错误分割,图11进一步显示PADT-BPP模型波位聚类效果较差,边界混叠严重,这正是引入强化学习方法的核心动机。

标准DQN模型相比PADT-BPP模型在波位划分性能上有略微突破,Precision达到75.8%,F1-Score为0.741,推理时间为25.2ms/1k脉冲,收敛轮数为45轮。DQN通过神经网络学习波位片段的13维特征,能够捕捉传统方法难以建模的复杂模式。然而,标准DQN的性能仍显著低于后续方

表3 实验结果

Tab. 3 Results of experiments

方法	Precision (%)	Recall (%)	F1-Score	推理时间
PADT-BPP模型	71.2	68.9	0.706	18.7 (ms/1k脉冲)
DQN模型	75.8	72.4	0.741	25.2 (ms/1k脉冲)
PPO模型	76.5	73.5	0.750	24.9 (ms/1k脉冲)
SAC-D模型	78.0	75.1	0.769	25.7 (ms/1k脉冲)
HIRL-BPP模型	84.9	82.6	0.838	25.2 (ms/1k脉冲)
EK-HRL框架	92.7	91.8	0.921	28.9 (ms/1k脉冲)

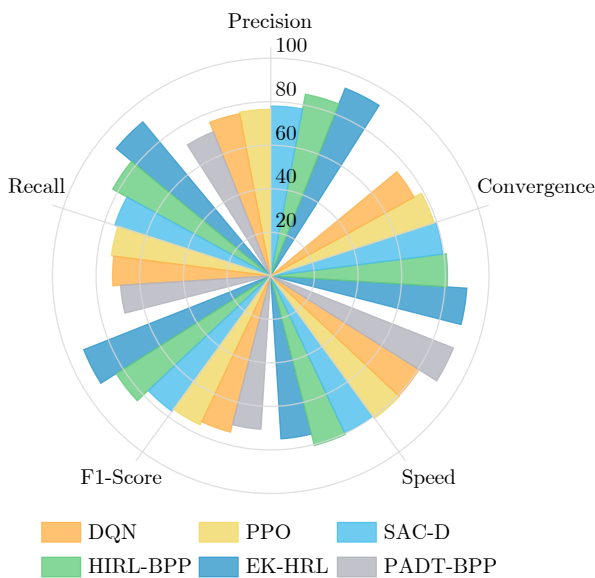


图8 各模型在5个维度上的性能对比雷达图

Fig. 8 Performance comparison radar chart of different models across five dimensions

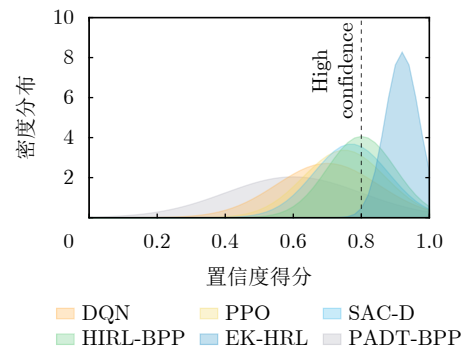


图9 决策样本的置信度得分分布密度

Fig. 9 Confidence score distribution density of decision samples

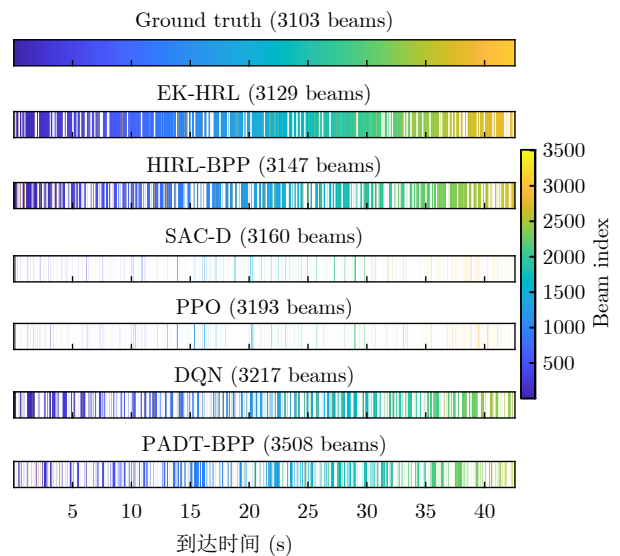


图10 不同波位划分方法的序列决策结果对比

Fig. 10 Comparison of sequential decision results for beam position division using different models

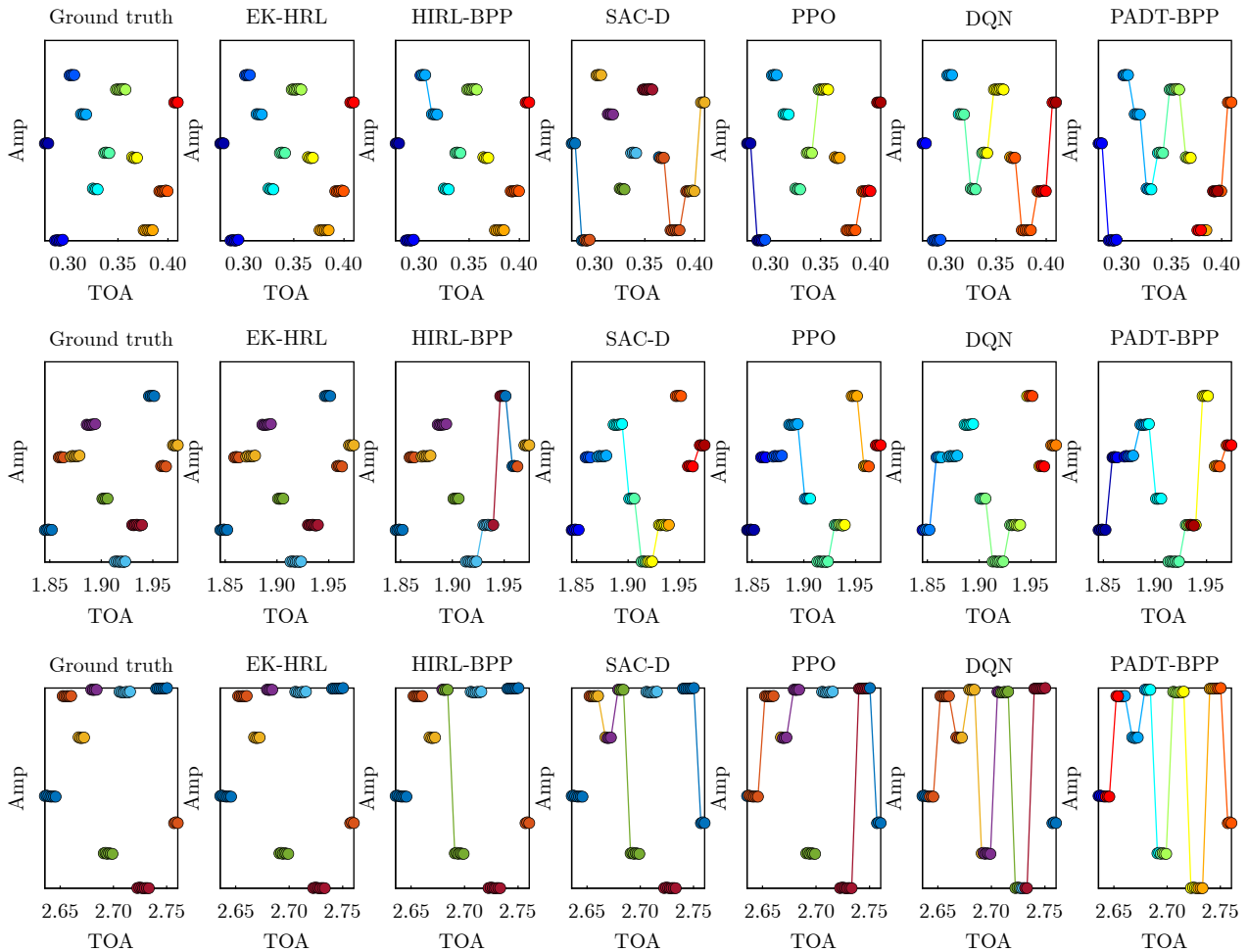


图 11 不同方法波位划分结果的时序可视化对比

Fig. 11 Time-series visualization of beam position partitioning results using different methods

法，主要原因是缺乏专家知识指导，依赖随机探索效率低，且无置信度机制约束，对低可靠性决策的过滤能力不足。如图9所示，DQN模型的划分结果置信度分布较为扁平，未能形成有效的高置信度峰值。这种不确定性直接反映在图11的可视化结果中，虽然相比PADT-BPP划分效果有所改善，但仍存在较多错误边界。

HIRL-BPP模型的Precision跃升至84.9%，F1-Score为0.838，性能相比DQN进一步提升。性能提升得益于专家对模型难以判别波位的决策修正。从图8雷达图可见其准确率指标覆盖面积明显大于DQN。然而，由于该模型中SBPCA模型仅参与奖励函数计算，未为决策主体提供反馈、未标注高置信度训练样本，性能与收敛效率仍有提升空间。

EK-HRL的Precision达到92.7%，F1-Score为0.921，收敛轮数仅25轮。推理时间为28.9ms/1k脉冲。在线推理阶段，SBPCA模块仍用于输出各波位的置信度评分，为下游任务提供决策可信度参考，由此产生3.7 ms的额外开销。如图11所示，EK-

HRL框架展现了良好的波位划分性能，同一波位的脉冲点高度聚集，不同波位间边界清晰。图12展示了EK-HRL框架的波位划分局部细节，其中每个三角形标记代表一个脉冲，不同颜色对应不同波位。右侧局部放大框展示了波位内脉冲幅度分布情况，进一步验证了模型的鲁棒性。该框架的性能优势得益于SBPCA模块的双重作用机制：一方面，通过置信度评估筛选出低置信度波位，确保专家集中标注模型不能准确处理的样本，提升专家知识的针对性；另一方面SBPCA模型为模型标注高置信度训练样本，过滤低置信度样本的错误探索。

#### 4.6 基于统计量的波位置置信度评估(SBPCA)模型有效性验证

为验证SBPCA模型的准确性与可靠性，本实验从校准性能、误差分布两个维度展开分析。将训练好的模型在测试集上的划分结果按置信度分为10个区间，统计每个区间的波位数量、平均置信度、实际准确率及绝对误差。实验结果如表4所示。

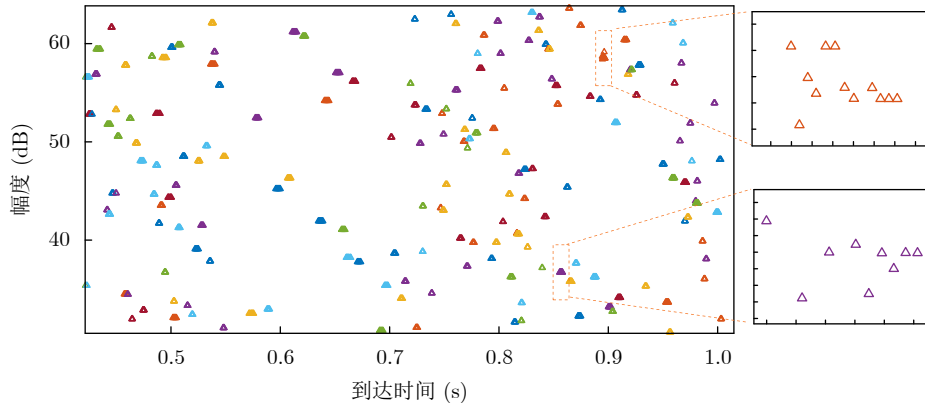


图 12 EK-HRL框架的波位划分细节展示

Fig. 12 Detailed illustration of beam position partitioning in the EK-HRL framework

表 4 SBPCA模型有效性验证实验结果

Tab. 4 Validation results of the SBPCA model

置信度区间	波位数量	平均置信度	实际准确率	绝对误差
0~0.1	14	0.068	0.210	0.142
0.1~0.2	22	0.195	0.183	0.012
0.2~0.3	30	0.293	0.236	0.057
0.3~0.4	40	0.348	0.453	0.105
0.4~0.5	53	0.451	0.531	0.080
0.5~0.6	75	0.547	0.676	0.129
0.6~0.7	108	0.653	0.821	0.168
0.7~0.8	223	0.748	0.933	0.185
0.8~0.9	1170	0.861	0.949	0.088
0.9~1.0	1368	0.947	0.980	0.033

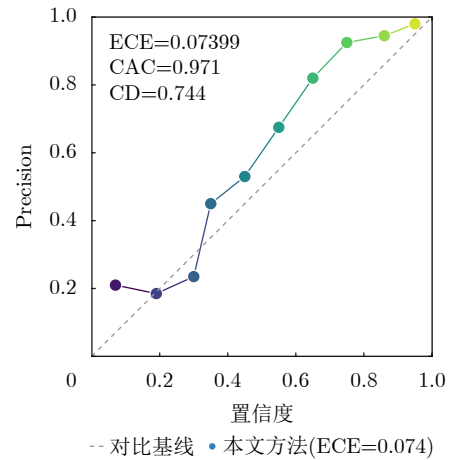


图 13 SBPCA模型置信度校准曲线

Fig. 13 Confidence calibration curve of the SBPCA model

整体校准性能分析：实验表明，SBPCA模型实现了优异的校准性能，期望校准误差(ECE)为0.074，置信度-准确率相关系数(CAC)达到0.971，置信度区分度(CD)为0.744。其中CAC为0.971，表明置信度评分能够可靠地反映波位划分的实际准确率。从图13观察到SBPCA模型在不同置信度区间呈现非线性校准模式。低置信度区间(0~0.1)表现出“欠自信”特征，反映了模型在波位边界模糊处的不确定性。中等置信度区间(0.3~0.7)展现渐进式校准发散，偏差从0.105升至0.168，但在0.7~0.8区间出现0.185的局部峰值，这种非单调模式反映了强化学习训练中探索-利用权衡的动态过程。高置信度区间(0.8~1.0)实现较为理想的校准，平均偏差仅6.0个百分点，0.9~1.0区间偏差仅0.033，满足良好校准标准。

样本分布与误差协同分析：如图14所示，气泡图直观展示了置信度与Precision之间的强正相关关系。通过二次多项式拟合，发现Precision随置信度增长而增高。样本分布呈正偏态，81.7%集中在高

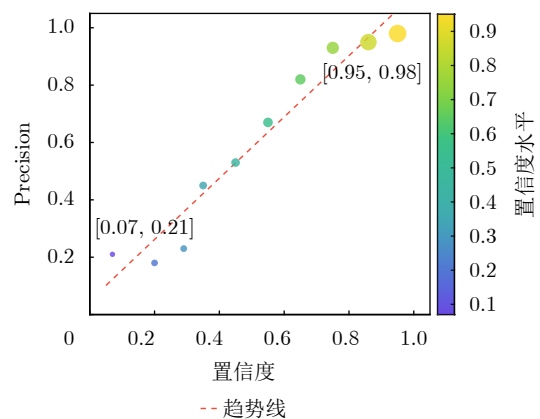


图 14 置信度与Precision关系气泡图

Fig. 14 Bubble chart of confidence versus precision

置信度区间(0.8~1.0)，低置信度区间(0~0.6)仅占7.57%。这种分布反映了SBPCA模型在不同置信度区域的样本难度差异：低置信度区域对应参数边界模糊的困难样本，而高置信度区域则对应模型的正确决策。37.71%高置信度样本误差仅0.088，形成“高置信度低误差”的理想配置。

综上所述, SBPCA模型通过实现 $ECE=0.074$ 的良好校准、 $CAC=0.971$ 的强线性关系以及 $CD=0.744$ 的有效区分能力, 为EK-HRL框架的置信度驱动奖励塑形机制提供了坚实的实证依据。模型在校准性能和误差分布健康性两个维度的优异表现, 验证了统计量驱动的置信度评估方法的有效性。

## 5 结语

本文系统性提出了相控阵雷达波位划分的概念, 并从波位角度入手进行相控阵雷达信号分析, 将波位划分定义为从混合脉冲流中还原属于不同波位的脉冲子序列, 创新性地提出了融合专家知识与强化学习的混合框架(EK-HRL)。主要贡献包括: 将波位划分问题建模为序列决策问题, 提出了基于脉冲幅度动态门限的基线划分模型, 为智能优化提供高质量初始解; 构建了人在回路的强化学习架构, 有效融合专家知识提升决策可信度; 设计了基于统计量的波位置信度评估模型与四维复合奖励函数, 通过奖励塑形机制显著加速模型收敛并提升划分精度。实验表明该方法取得了92.7%的划分精确率, 且置信度评估表现出良好的校准性。然而, 该方法仍存在对专家标注的初始依赖、极端参数交叠场景性能受限及多辐射源场景扩展性未充分验证等局限。未来研究将致力于开发半监督学习机制以降低标注依赖, 探索多模态信号融合以提升复杂环境鲁棒性, 并将框架扩展至多辐射源混合信号处理场景, 验证实测数据中的有效性, 推动其在实际作战系统中的可靠应用。

**利益冲突** 所有作者均声明不存在利益冲突

**Conflict of Interests** The authors declare that there is no conflict of interests

## 参 考 文 献

- [1] 王雪松, 王占领, 庞晨, 等. 极化相控阵雷达技术研究综述[J]. 雷达科学与技术, 2021, 19(4): 349–370. doi: [10.3969/j.issn.1672-2337.2021.04.001](https://doi.org/10.3969/j.issn.1672-2337.2021.04.001).  
WANG Xuesong, WANG Zhanling, PANG Chen, et al. Review on polarimetric phased array radar technologies[J]. *Radar Science and Technology*, 2021, 19(4): 349–370. doi: [10.3969/j.issn.1672-2337.2021.04.001](https://doi.org/10.3969/j.issn.1672-2337.2021.04.001).
- [2] GURBUZ S Z, GRIFFITHS H D, CHARLISH A, et al. An overview of cognitive radar: Past, present, and future[J]. *IEEE Aerospace and Electronic Systems Magazine*, 2019, 34(12): 6–18. doi: [10.1109/MAES.2019.2953762](https://doi.org/10.1109/MAES.2019.2953762).
- [3] GOK G, ALP Y K, and ARIKAN O. A new method for specific emitter identification with results on real radar measurements[J]. *IEEE Transactions on Information Forensics and Security*, 2020, 15: 3335–3346. doi: [10.1109/TIFS.2020.2988558](https://doi.org/10.1109/TIFS.2020.2988558).
- [4] 王颖, 郭睿, 梁毅. 实测双极化雷达压制干扰的特性分析与抑制[J]. 海军航空大学学报, 2025, 40(1): 163–170, 196. doi: [10.7682/j.issn.2097-1427.2025.01.007](https://doi.org/10.7682/j.issn.2097-1427.2025.01.007).  
WANG Ying, GUO Rui, and LIANG Yi. Characteristics analysis and suppression of measured dual-polarization radar blanketing jamming[J]. *Journal of Naval Aviation University*, 2025, 40(1): 163–170, 196. doi: [10.7682/j.issn.2097-1427.2025.01.007](https://doi.org/10.7682/j.issn.2097-1427.2025.01.007).
- [5] CHENG Wenhai, ZHANG Qunying, DONG Jiaming, et al. An enhanced algorithm for deinterleaving mixed radar signals[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2021, 57(6): 3927–3940. doi: [10.1109/TAES.2021.3087832](https://doi.org/10.1109/TAES.2021.3087832).
- [6] ZHANG Peng, YAN Junkun, PU Wenqiang, et al. Multi-dimensional resource management scheme for multiple target tracking under dynamic electromagnetic environment[J]. *IEEE Transactions on Signal Processing*, 2024, 72: 2377–2393. doi: [10.1109/TSP.2024.3390119](https://doi.org/10.1109/TSP.2024.3390119).
- [7] CHEN Baoxin, CHEN Xiaolong, HUANG Yong, et al. Transmit beam pattern synthesis for the FDA radar[J]. *IEEE Antennas and Wireless Propagation Letters*, 2018, 17(1): 98–101. doi: [10.1109/LAWP.2017.2776957](https://doi.org/10.1109/LAWP.2017.2776957).
- [8] 闫文君, 刘康晟, 凌青, 等. 跨场景辐射源个体识别技术综述[J]. 雷达学报(中英文), 待出版. doi: [10.12000/JR25166](https://doi.org/10.12000/JR25166).  
YAN Wenjun, LIU Kangsheng, LING Qing, et al. Survey of cross-scenario specific emitter identification technology[J]. *Journal of Radars*, in press. doi: [10.12000/JR25166](https://doi.org/10.12000/JR25166).
- [9] KRISHNAMURTHY V, PATTANAYAK K, GOGINENI S, et al. Adversarial radar inference: Inverse tracking, identifying cognition, and designing smart interference[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2021, 57(4): 2067–2081. doi: [10.1109/TAES.2021.3090901](https://doi.org/10.1109/TAES.2021.3090901).
- [10] 刘松涛, 赵帅, 汪慧阳. 雷达辐射源识别技术新进展[J]. 中国电子科学研究院学报, 2022, 17(6): 523–533. doi: [10.3969/j.issn.1673-5692.2022.06.002](https://doi.org/10.3969/j.issn.1673-5692.2022.06.002).  
LIU Songtao, ZHAO Shuai, and WANG Huiyang. New development on the technology of radar emitter identification[J]. *Journal of China Academy of Electronics and Information Technology*, 2022, 17(6): 523–533. doi: [10.3969/j.issn.1673-5692.2022.06.002](https://doi.org/10.3969/j.issn.1673-5692.2022.06.002).
- [11] RECHT B. A tour of reinforcement learning: The view from continuous control[J]. *Annual Review of Control, Robotics, and Autonomous Systems*, 2019, 2: 253–279. doi: [10.1146/annurev-control-053018-023825](https://doi.org/10.1146/annurev-control-053018-023825).
- [12] AL KASSIR H, ZAHARIS Z D, LAZARIDIS P I, et al. A review of the state of the art and future challenges of deep learning-based beamforming[J]. *IEEE Access*, 2022, 10: 80869–80882. doi: [10.1109/ACCESS.2022.3195299](https://doi.org/10.1109/ACCESS.2022.3195299).

- [13] 石荣, 吴聪. 基于PRI信息的雷达脉冲信号分选技术研究综述[J]. 电讯技术, 2020, 60(1): 112–120. doi: [10.3969/j.issn.1001-893x.2020.01.019](https://doi.org/10.3969/j.issn.1001-893x.2020.01.019).  
SHI Rong and WU Cong. Review on technology research about radar pulse signal deinterleaving based on PRI information[J]. *Telecommunication Engineering*, 2020, 60(1): 112–120. doi: [10.3969/j.issn.1001-893x.2020.01.019](https://doi.org/10.3969/j.issn.1001-893x.2020.01.019).
- [14] YAO Yu, LIU Haitao, MIAO Pu, *et al.* MIMO radar design for extended target detection in a spectrally crowded environment[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(9): 14389–14398. doi: [10.1109/TITS.2021.3127727](https://doi.org/10.1109/TITS.2021.3127727).
- [15] 张嘉翔, 张凯翔, 梁振楠, 等. 一种基于深度强化学习的频率捷变雷达智能频点决策方法[J]. 雷达学报(中英文), 2024, 13(1): 227–239. doi: [10.12000/JR23197](https://doi.org/10.12000/JR23197).  
ZHANG Jiexiang, ZHANG Kaixiang, LIANG Zhenan, *et al.* An intelligent frequency decision method for a frequency agile radar based on deep reinforcement learning[J]. *Journal of Radars*, 2024, 13(1): 227–239. doi: [10.12000/JR23197](https://doi.org/10.12000/JR23197).
- [16] FAWAZ H I, FORESTIER G, WEBER J, *et al.* Deep learning for time series classification: A review[J]. *Data Mining and Knowledge Discovery*, 2019, 33(4): 917–963. doi: [10.1007/s10618-019-00619-1](https://doi.org/10.1007/s10618-019-00619-1).
- [17] 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*, 2023, 59(1): 292–310. doi: [10.1109/TAES.2022.3184619](https://doi.org/10.1109/TAES.2022.3184619).
- [18] PAPA L, RUSSO P, AMERINI I, *et al.* A survey on efficient vision transformers: Algorithms, techniques, and performance benchmarking[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(12): 7682–7700. doi: [10.1109/TPAMI.2024.3392941](https://doi.org/10.1109/TPAMI.2024.3392941).
- [19] TU Ya, LIN Yun, ZHA Haoran, *et al.* Large-scale real-world radio signal recognition with deep learning[J]. *Chinese Journal of Aeronautics*, 2022, 35(9): 35–48. doi: [10.1016/j.cja.2021.08.016](https://doi.org/10.1016/j.cja.2021.08.016).
- [20] YAN Wenjun, LING Qing, YU Keyuan, *et al.* A pseudolabel method with semantic drift for specific emitter identification[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2025, 61(3): 6217–6235. doi: [10.1109/TAES.2025.3527960](https://doi.org/10.1109/TAES.2025.3527960).
- [21] WANG Yu, GUI Guan, LIN Yun, *et al.* Few-shot specific emitter identification via deep metric ensemble learning[J]. *IEEE Internet of Things Journal*, 2022, 9(24): 24980–24994. doi: [10.1109/JIOT.2022.3194967](https://doi.org/10.1109/JIOT.2022.3194967).
- [22] MOSQUEIRA-REY E, HERNÁNDEZ-PEREIRA E, ALONSO-RÍOS D, *et al.* Human-in-the-loop machine learning: A state of the art[J]. *Artificial Intelligence Review*, 2023, 56(4): 3005–3054. doi: [10.1007/s10462-022-10246-w](https://doi.org/10.1007/s10462-022-10246-w).
- [23] TANG Chen, ABBATEMATTEO B, HU Jiaheng, *et al.* Deep reinforcement learning for robotics: A survey of real-world successes[J]. *Annual Review of Control, Robotics, and Autonomous Systems*, 2025, 8: 153–188. doi: [10.1146/annurev-control-030323-022510](https://doi.org/10.1146/annurev-control-030323-022510).
- [24] WU Jingda, HUANG Zhiyu, HU Zhongxu, *et al.* Toward human-in-the-loop AI: Enhancing deep reinforcement learning via real-time human guidance for autonomous driving[J]. *Engineering*, 2023, 21: 75–91. doi: [10.1016/j.eng.2022.05.017](https://doi.org/10.1016/j.eng.2022.05.017).
- [25] WURMAN P R, BARRETT S, KAWAMOTO K, *et al.* Outracing champion Gran Turismo drivers with deep reinforcement learning[J]. *Nature*, 2022, 602(7896): 223–228. doi: [10.1038/s41586-021-04357-7](https://doi.org/10.1038/s41586-021-04357-7).
- [26] 宋新超, 吴连慧, 王星宇. 基于侦察幅度信息的雷达行为及特征分析[J]. 舰船电子对抗, 2019, 42(3): 48–51. doi: [10.16426/j.cnki.jcdzdk.2019.03.011](https://doi.org/10.16426/j.cnki.jcdzdk.2019.03.011).  
SONG Xinchao, WU Lianhui, and WANG Xingyu. Radar behavior and feature analysis based on reconnaissance amplitude information[J]. *Shipboard Electronic Countermeasure*, 2019, 42(3): 48–51. doi: [10.16426/j.cnki.jcdzdk.2019.03.011](https://doi.org/10.16426/j.cnki.jcdzdk.2019.03.011).
- [27] RAO Jinjun, XU Xiaoqiang, BIAN Haoran, *et al.* A modified random network distillation algorithm and its application in USVs naval battle simulation[J]. *Ocean Engineering*, 2022, 261: 112147. doi: [10.1016/j.oceaneng.2022.112147](https://doi.org/10.1016/j.oceaneng.2022.112147).
- [28] ARULKUMARAN K, DEISENROTH M P, BRUNDAGE M, *et al.* Deep reinforcement learning: A brief survey[J]. *IEEE Signal Processing Magazine*, 2017, 34(6): 26–38. doi: [10.1109/MSP.2017.2743240](https://doi.org/10.1109/MSP.2017.2743240).
- [29] TAO Jin and ZHANG Xindong. Radar emitter signal recognition method based on improved collaborative semi-supervised learning[J]. *Journal of Systems Engineering and Electronics*, 2023, 34(5): 1182–1190. doi: [10.23919/JSEE.2023.000126](https://doi.org/10.23919/JSEE.2023.000126).
- [30] KARNIADAKIS G E, KEVREKIDIS I G, LU Lu, *et al.* Physics-informed machine learning[J]. *Nature Reviews Physics*, 2021, 3(6): 422–440. doi: [10.1038/s42254-021-00314-5](https://doi.org/10.1038/s42254-021-00314-5).
- [31] MANNION P, DEVLIN S, MASON K, *et al.* Policy invariance under reward transformations for multi-objective reinforcement learning[J]. *Neurocomputing*, 2017, 263: 60–73. doi: [10.1016/j.neucom.2017.05.090](https://doi.org/10.1016/j.neucom.2017.05.090).
- [32] BOUDT K, TODOROV V, and WANG Wenjing. Robust distribution-based winsorization in composite indicators construction[J]. *Social Indicators Research*, 2020, 149(2):

- 375–397. doi: [10.1007/s11205-019-02259-w](https://doi.org/10.1007/s11205-019-02259-w).
- [33] LOMBARDI O, HOLIK F, and VANNI L. What is Shannon information?[J]. *Synthese*, 2016, 193(7): 1983–2012. doi: [10.1007/s11229-015-0824-z](https://doi.org/10.1007/s11229-015-0824-z).
- [34] ZHANG Donglin, WU Xiaojun, XU Tianyang, *et al.* DAH: Discrete asymmetric hashing for efficient cross-media retrieval[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(2): 1365–1378. doi: [10.1109/TKDE.2021.3099125](https://doi.org/10.1109/TKDE.2021.3099125).
- [35] SCHULMAN J, WOLSKI F, DHARIWAL P, *et al.* Proximal policy optimization algorithms[EB/OL]. <https://arxiv.org/abs/1707.06347>, 2017.
- [36] CHRISTODOULOU P. Soft actor-critic for discrete action settings[EB/OL]. <https://arxiv.org/abs/1910.07207>, 2019.

### 作者简介

余 壮，硕士生，主要研究方向为相控阵雷达信号分析。

凌 青，教授，主要研究方向为信号智能处理。

闫文君，副教授，主要研究方向为信号智能处理。

董世鹏，硕士生，主要研究方向为雷达信号处理。

张立民，教授，主要研究方向为卫星信号处理及应用、信号智能处理。

商允力，高级工程师，主要研究方向为电子对抗。

(责任编辑：于青)