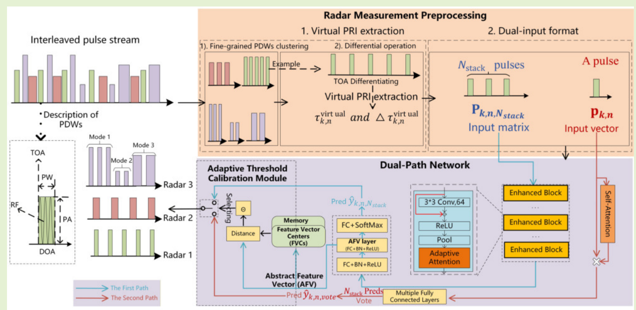
Virtual PRI-Assisted Radar Signal Sorting With a Dual-Path Network for Fine-Grained Cluster Fusion

Huajie Wu, Xingyun Chen, Ying Ma, *Member, IEEE*, Haichuan Ding, *Member, IEEE*, and Sheng Chen, *Life Fellow, IEEE*

***Abstract*—Pulse repetition interval (PRI) estimation remains critical yet challenging in radar signal sorting (RSS), particularly under complex electromagnetic environments and diverse radar operating modes. Traditional methods relying on direct PRI estimation often fail to handle complex PRI modulation patterns and unknown radar modes while exhibiting sensitivity to missing pulses. In light of this challenge, we introduce the concept of virtual PRIs, a representation derived from fine-grained clustering of pulse descriptor words (PDWs). Virtual PRIs serve as statistically robust surrogates for true PRIs that preserve temporal patterns even under complex PRI modulation and missing pulses while eliminating the need for explicit PRI estimation. This enables better handling of complex PRI modulation without prior knowledge of radar operating modes. To address the overclustering issue introduced by fine-grained clustering, where the number of obtained pulse clusters far exceeds the actual number of radar emitters, we introduce a dual-path network to merge fine-grained pulse clusters by jointly processing stacked virtual PRI sequences and individual virtual PRI pulses for global and local pulse feature extraction. The accuracy and robustness of our virtual PRI-assisted RSS schemes are validated on datasets incorporating complex PRI modulation, missing pulses, and unknown radar modes.**



***Index Terms*—Convolutional neural networks (CNNs), feature distribution shifts, pulse repetition interval (PRI), radar signal sorting (RSS).**

I. Introduction

GIVEN the increasingly sophisticated radar signals and the complex electromagnetic environments, it is more and more difficult for the electronic reconnaissance system to perfectly sort the pulses from different radars, and such imperfection in radar signal sorting (RSS) will adversely affect the effectiveness of subsequent electronic support and countermeasure operations [1], [2], [3], [4], [5], [6], [7]. Pulse descriptor words and pulse repetition interval (PRI) are widely utilized for RSS. The conventional PDW-based RSS schemes, which operate on a template-matching principle, often fail to effectively sort radar pulses when their PDWs partially overlap. Recently, driven by neural networks' capability of automatic feature extraction and nonlinear relationship modeling, we have witnessed growing interest in deep learning (DL)-based RSS, which utilizes a well-trained deep neural network for signal sorting. Unlike traditional RSS schemes, DL-based RSS attempts to separate the densely interleaved radar signals by capturing their distributions in a projected feature space such that the pulses from different radar emitters can be correctly sorted even when their parameters are partially overlapped [13], [14], [15], [16], [17].

Existing PDW/time of arrival (TOA)-based DL schemes directly work with the interleaved pulse streams, and each pulse is mainly characterized and sorted by a group of parameters called pulse descriptor words (PDWs) [16], [18], [19], [20]. Although these PDW/TOA-based DL schemes have verified the benefit of using time-domain pulse modulation behaviors for RSS, they directly rely on raw TOAs without exploiting the PRIs, which fundamentally reflect the operating modes of different radars. Consequently, this leads to suboptimal sorting performance and limited robustness in highly complex interleaved environments.

To address this limitation and fully exploit these temporal patterns, researchers have proposed PRI-based DL schemes to

Received 20 May 2026; accepted 12 June 2026. Date of publication 23 June 2026; date of current version 15 August 2026. This work was supported by the National Natural Science Foundation of China under Grant 62301052. The associate editor coordinating the review of this article and approving it for publication was Prof. Seongwook Lee. *(Corresponding author: Ying Ma.)*

Huajie Wu and Ying Ma are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: wuhuajie1120@163.com; maying@bit.edu.cn).

Xingyun Chen and Haichuan Ding are with the School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing 100081, China (e-mail: xingyunc@bit.edu.cn; hcding@bit.edu.cn).

Sheng Chen is with the School of Computer Science and Technology, Ocean University of China, Qingdao 266100, China (e-mail: sqc@ecs.soton.ac.uk).

Digital Object Identifier 10.1109/JSEN.2026.3704572

explicitly extract PRI-related information for RSS [26], [27], [28], [29], [30], [31]. Although these PRI-based DL schemes can address the lack of true PRIs to some extent, the iterative approaches for PRI extraction can lead to cumulative PRI estimation errors, where the erroneous PRI estimation can result in incorrect signal sorting and in turn affect the accuracy of subsequent PRI estimation and signal sorting. Furthermore, another significant limitation of these schemes is their poor robustness in complex reconnaissance scenarios. Specifically, the extraction of accurate PRI information becomes extremely vulnerable to multifunctional radars and complex modulation types, such as PRI jitter, stagger, and agility. This problem is further exacerbated by missing pulses in the captured pulse sequences, which severely hinders the reliable application of PRI-based schemes in practical environments.

Consequently, existing schemes either underutilize temporal PRI patterns or struggle with cumulative errors and high sensitivity to complex modulations and missing pulses during explicit PRI estimation. To circumvent these fundamental limitations, this article introduces a kind of parameter called virtual PRI, which can be efficiently extracted as a robust substitute for true PRI to assist RSS. By utilizing these substitutes, our scheme successfully captures essential temporal patterns while completely avoiding the trap of iterative PRI estimation and its associated cumulative errors. Specifically, these virtual PRIs are derived from a fine-grained PDW-based clustering result.[1] Unlike the ground-truth PRIs, virtual PRIs are the differences between the TOAs of neighboring pulses within the same fine-grained cluster. Different from the classical clustering scheme normally adopted in RSS, our fine-grained clustering scheme is designed to group the pulses into a number of clusters which is far larger than the actual number of categories [26], [32], [33], [34], [35], [36]. In doing so, we can ensure that only pulses with very similar PDWs are clustered together, and thus, the pulses within the same cluster are very likely from the same operating mode of a specific radar. Under these circumstances, the virtual PRIs derived from adjacent pulses within the same fine-grained cluster represent the accumulation of several true PRIs with high probability, even in the case of missing pulses and complex PRI modulation patterns (e.g., PRI jitter, stagger, and agility). By using the extracted virtual PRIs as surrogates for the ground-truth PRIs in pulse deinterleaving, the virtual PRI-assisted RSS scheme enables robust PRI characterization without explicit PRI estimation. Unfortunately, the fine-grained clustering also introduces an overclustering issue, where the number of obtained pulse clusters can far exceed the actual number of radar emitters. To address this issue, we propose a dual-path network to merge fine-grained pulse clusters by jointly processing stacked virtual PRI sequences and individual virtual PRI pulses for global and local pulse feature extraction. First, the dual-path network employs adaptive spatial attention, adaptive channel attention, and self-attention modules to help extract subtle differences in PDWs and virtual PRIs between different radar modes. Second, the dual-path network utilizes separate input paths to process both the PDWs and virtual PRIs associated with individual pulses and those associated with a sequence of consecutive radar pulses from the same fine-grained pulse cluster. This allows the network to explicitly extract both global and local features of radar pulse signals. The global features refer to the time-varying patterns of the virtual PRIs and PDWs extracted from the consecutive pulse sequences assigned to the same fine-grained pulse cluster, which enables the dual-path network to capture the overall distribution of the pulse sequence, for robustness against the distribution shifts in individual radar pulse features caused by noise. The local features, on the other hand, refer to fine-grained characteristics of individual pulses and enable the network to uncover consistent, subtle patterns when radar pulse features undergo overall distribution shifts due to the variations in operating modes, thereby improving generalizability. Third, to facilitate accurate RSS in the case of highly overlapped PDWs, the dual-path network is trained by incorporating interclass and intraclass losses on top of the classification loss, where the interclass loss is used to enlarge the feature differences between the pulses from different radars while the intraclass loss helps to reduce the feature differences between the signals from the same radar. The major contributions of this article are summarized as follows.

1) We introduce the concept of virtual PRIs and a fine-grained PDW-based clustering scheme for effective pulse sorting. Unlike the traditional clustering algorithm, our fine-grained clustering algorithm groups radar pulses into a much larger number of clusters than actually needed so that virtual PRI can be extracted as substitutes for the true PRI to assist RSS.
2) We propose the dual-path network to mitigate the overclustering issue introduced by the fine-grained clustering. Its dual processing paths separately model stacked virtual PRI sequences and individual virtual PRI pulses, enabling the network to jointly capture global and local pulse features for more reliable cluster merging.
3) Following the established practices in RSS research [21], [24], [37], where real-world datasets are lacking, we evaluate the effectiveness of the proposed virtual PRI-assisted RSS scheme on simulated datasets where the captured pulse sequences suffer complex PRI modulation patterns, missing pulses, and unknown radar modes. The experimental results on these datasets demonstrate that incorporating virtual PRIs yields an improvement of 14% in accuracy, and the proposed scheme consistently outperforms benchmark methods utilizing PRI estimations derived from the traditional PRI estimation algorithms.

The remainder of this article is organized as follows. Section II reviews the related work. The RSS problem and our proposed virtual PRI-assisted scheme are presented in Sections III and IV, respectively. Experimental setups and simulation results are discussed in Sections V and VI, respectively, followed by conclusion in Section VII.

[1] Although PDWs and PRI are both subject to estimation errors, unlike PDWs, the accuracy of PRI inherently depends on correct pulse-to-emitter association. Erroneous pulse-to-emitter association often leads to significant discrepancies in PRI values, and thus, the utilization of PRI in RSS requires specialized treatment.

TABLE I
COMPARISON OF EXISTING RSS APPROACHES AND THE PROPOSED METHOD

Ref.	Method Category	Core Mechanism / Feature Type	Exploit Temporal PRI Patterns	Avoid Cumulative Estimation Errors	Robust to Missing Pulses & Modulations
[8]–[10]	Conventional Schemes	Direct PRI estimation via histograms or iterative algorithms.	✓	×	×
[21]–[24]	PDW/TOA-based DL	Deep learning utilizing raw PDWs or basic TOA differences.	×	✓	×
[29]–[31]	PRI-based DL	Deep learning utilizing true PRIs or iteratively estimated PRIs.	✓	×	×
Ours	**Proposed Scheme**	**Virtual PRI via fine-grained clustering & Dual-Path fusion.**	✓	✓	✓

II. RELATED WORK

Various schemes have been explored to tackle the RSS problem. As summarized in Table I, existing studies can be broadly categorized into three groups: conventional schemes, PDW/TOA-based DL schemes, and PRI-based DL schemes. As for the conventional schemes, such as that proposed by Mardia [8], they typically rely on template-matching or iterative algorithms for PRI estimation [8], [9], [10], [11], [12]. Consequently, these conventional schemes either often fail to effectively sort radar pulses when their PDWs partially overlap, or are prone to cumulative estimation errors in PRI estimation, especially in the presence of complex PRI modulations (e.g., jitter, stagger, and agility) and missing pulses, which severely degrade their performance in complex and congested electromagnetic environments.

For the PDW/TOA-based DL schemes, Lang et al. [21] proposed a residual graph convolutional network-based RSS scheme where the adjacency matrix is constructed based on the PDWs of different pulses to capture their underlying relationships and proximities for efficient learning. Zhou et al. [22] presented a sequence-to-sequence RSS scheme where multiple self-attention mechanisms are utilized to process the embedded PDWs to extract the correlations between the pulses from the same source at the sequence level for effective signal sorting. To address the issue of seriously overlapped radar parameters, Chen et al. [23] proposed to map the PDWs of the received pulses into a 2-D lattice and utilized an instance segmentation network to cluster the PDWs for signal sorting. Given the tight connection between pulse train modulation patterns and the operating modes of radar emitters, Xiang et al. [24] utilized a recursive deinterleaving network to process the TOA, a part of the PDWs, and trained the network for RSS with the help of dual-path attention. Along the same line of thought, Chao et al. [25] proposed a semantic segmentation-based pulse deinterleaving scheme with the differences of TOAs as input. However, despite these advancements, these schemes generally underutilize essential temporal PRI patterns, rendering them less robust in the presence of missing pulses and complex PRI modulations.

For the PRI-based DL schemes, Chi et al. [26] used a limited penetrable visibility graph to capture the underlying community structure within the collected PRI measurements so that the pulses from different radars can be sorted by detecting the communities inside the graph with the label propagation algorithm and density peak clustering. On top of this work, Zhang et al. [27] attempted to further improve the sorting performance by proposing a similarity-based conditional label propagation algorithm for reliable community detection. Given the detrimental impacts of noise and outliers on label propagation, Li et al. [28] exploited the density-based spatial clustering of applications with noise (DBSCAN) to preprocess the input to the conditional label propagation algorithm so that erroneous community detection resulting from incorrect label propagation can be addressed to facilitate effective RSS. Motivated by the pixel clustering approach used in image segmentation, Nuhoglu et al. [29] mapped radar pulses to pixel points in a 2-D time-PRI image using a PRI transform-based complex autocorrelation function and developed a U-shaped convolutional neural network (U-Net)-based RSS scheme by clustering pixel points in the 2-D time-PRI image. However, the above works use the true PRI in both the training and test sets, which is not feasible in real-world environments. Therefore, some works have used iterative algorithms to extract the PRI estimates from the test pulse sequences to assist RSS [30], [31]. Benefiting from the advancements in neural networks, Liu and Philip [30] trained recurrent neural network (RNN) models using the ground-truth PRI and pulsewidth (PW) from the training set so that the PRIs of the test pulse sequences can be iteratively estimated to improve RSS performance. Along the same line of thought, Luo et al. [31] utilized the long short-term memory (LSTM) models to handle the long-term temporal dependencies in radar pulses for PRI estimation and pulse deinterleaving. However, due to their reliance on iterative processes, these schemes are fundamentally prone to cumulative estimation errors in PRI estimation, rendering them less robust in the presence of missing pulses and complex PRI modulations.

III. PROBLEM DESCRIPTION

In this article, we consider a scenario where the electronic reconnaissance system receives an interleaved pulse stream from M simultaneously active radars. These radars are indexed as $\{r_1, r_2, \ldots, r_M\}$. Each radar operates with a single or multiple modes, each employing a complex PRI modulation pattern such as stagger, jitter, or agility. The reconnaissance system measures parameters of each received pulse to obtain its

PDWs, where $\mathbf{s}_n = [t_n, \phi_n, \omega_n, \alpha_n, \theta_n]$ represents the PDWs of the nth received pulse, t_n is the TOA of the nth received pulse, ϕ_n is the radio frequency (RF) of the nth received pulse, ω_n is the PW of the nth received pulse, α_n is the pulse amplitude (PA) of the nth received pulse, and θ_n is the direction of arrival (DOA) of the nth received pulse. In congested electromagnetic environments with a multitude of radar operating modes, the values of the PDW parameters of different radars would be partially overlapped, and pulse sequences are frequently fragmented due to missing pulses. Because of missing pulses and partially overlapped PDWs, the existing RSS scheme solely based on measured PDWs could experience ambiguous and fragmented deinterleaved sequences, which calls for a new RSS scheme [8], [9], [17], [21], [22], [30]. In light of this, we introduce the virtual PRI-assisted RSS scheme as presented in Section IV.[2]

IV. Virtual PRI-Assisted RSS

In this section, we present the detailed design of the proposed virtual PRI-assisted RSS scheme, which facilitates effective RSS in scenarios with overlapped PDW parameters, missing pulses, and complex PRI modulation. The proposed scheme starts with a radar measurement preprocessing stage, which extracts the virtual PRIs and assembles the obtained virtual PRIs and the PDW parameters into a dual-input format. On top of the preprocessed data, the dual-path network utilizes two specifically designed paths to extract complementary features from the dual-input format and produce two label predictions for each fine-grained cluster. After that, an adaptive threshold calibration module fuses the two path-wise predictions of each fine-grained cluster to determine its final emitter label assignment. Finally, the fine-grained clusters assigned with the same emitter label are merged to generate the final RSS results. In Section IV-A, we will elaborate on the concept of virtual PRI and detail its extraction process. In Section IV-B, we will present the proposed dual-path network. In Section IV-C, we will discuss the training and the overall inference workflow from raw pulses to final sorting results, achieved by integrating the virtual PRI with the dual-path network.

A. Radar Measurement Preprocessing

This section presents the radar signal preprocessing stage, which extracts virtual PRIs from PDW measurements and prepares the input data of the dual-path network. First, we introduce the definition of the virtual PRI and how it could be obtained from PDW measurements using fine-grained clustering. Then, we analyze the benefits of the proposed virtual PRI. Finally, we discuss how to assemble the obtained virtual PRIs and the PDW parameters into a dual-input format, which serves as the input to the subsequent dual-path network.

[2]Our scheme is specifically designed for the challenging but frequently encountered cases where pulses from different radars have partially overlapped PDW parameters such that the received pulse stream cannot be correctly deinterleaved by directly applying traditional clustering algorithms. For the case when pulses from different radars are completely coincident in RF, PW, and DOA parameters, we might need to utilize more discriminative intrapulse features for effective RSS.

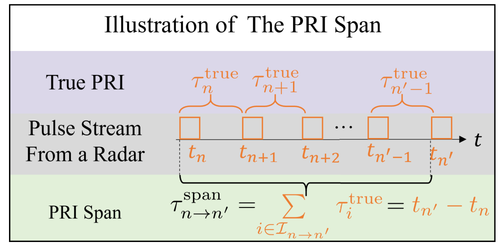


Fig. 1. Differences between the PRI span and the true PRI of a radar.

1) Virtual PRI Extraction: From Fig. 1, the true PRI is the physical PRI of a radar emitter. In contrast, to characterize the time separation between two pulses emitted by the same radar, we introduce an idealized quantity termed the PRI span. Specifically, for two pulses from the same radar with TOAs t_n and $t_{n'}$ ($t_{n'} > t_n$), the PRI span is defined as

$$\tau_{n\to n'}^{\text{span}} = \sum_{i\in\mathcal{I}_{n\to n'}} \tau_i^{\text{true}} = t_{n'} - t_n \tag{1}$$

where $\tau_{n\to n'}^{\text{span}}$ denotes the PRI span between the nth pulse and the n'th pulse, $\mathcal{I}_{n\to n'}$ denotes the index set of the consecutive true PRIs between these two pulses, and τ_i^{true} denotes the corresponding true PRI. From (1), it can be observed that the PRI span $\tau_{n\to n'}^{\text{span}}$ offers two unique advantages. First, the value of $\tau_{n\to n'}^{\text{span}}$ will not be affected by missing pulses between the nth pulse and the n'th pulse and thus serves as a valid and faithful representation of the underlying PRI timing pattern. Second, the PRI span $\tau_{n\to n'}^{\text{span}}$ can be derived as long as we can obtain the TOAs t_n and $t_{n'}$ of the nth and the n'th pulses from the same radar, regardless of the underlying PRI modulation pattern and without knowledge of the true PRIs between the intermediate pulses.

Note that, in practical RSS scenarios, $\tau_{n\to n'}^{\text{span}}$ cannot be directly obtained via the summation $\sum_{i\in\mathcal{I}_{n\to n'}} \tau_i^{\text{true}}$ as shown in (1), since the underlying true PRIs τ_i^{true} and their index set $\mathcal{I}_{n\to n'}$ are unknown from raw PDW measurements. Motivated by the second equality in (1), we propose to estimate the value of $\tau_{n\to n'}^{\text{span}}$ through fine-grained clustering. Specifically, we employ a fine-grained PDW clustering scheme to group pulses into a number of clusters far larger than the actual number of radars such that all pulses within each cluster originate from the same emitter with high probability [38], [39], [40]. Then, within each cluster, we estimate the PRI span by taking the TOA difference of two adjacent pulses (after ordering by TOA), and define this estimate as the *virtual PRI*. For the kth fine-grained cluster, the virtual PRI associated with the nth pulse is computed as

$$\tau_{k,n}^{\text{virtual}} = t_{k,n+1} - t_{k,n} \tag{2}$$

where $t_{k,n}$ and $t_{k,n+1}$ denote the TOAs of the nth and $(n+1)$th pulses in the kth fine-grained cluster, respectively. From the above discussions, the virtual PRI $\tau_{k,n}^{\text{virtual}}$ could effectively reflect the value of the PRI span as well as the underlying PRI timing pattern. Hence, the statistical distribution of virtual PRIs within a cluster is tightly linked to the underlying PRI modulation pattern. This relationship enables the merging of the fine-grained clusters to derive the final sorting result, which will be illustrated in Section IV-B.

To facilitate virtual PRI extraction, our fine-grained PDWs clustering scheme sets the number of clusters to be far larger than the actual number of radars. For instance, we use $K = 100$ clusters in our experimental scenario with 12 radars. It should be emphasized that the optimal number of clusters to be used is scenario-dependent and should be adjusted according to the complexity of the environment.[3] When compared to traditional clustering schemes, our fine-grained clustering scheme enhances intracluster purity, ensuring most pulses in each cluster come from the same radar under the same operating mode. Inspired by [24] and [26], when performing fine-grained clustering, we only utilize a subset of PDW parameters. Specifically, we exclude the TOA measurement as it increases monotonically with pulse indices, which could cause large separations between the pulses from the same radar. Meanwhile, the PA measurement is also excluded due to its susceptibility to environmental noise. Consequently, clustering is performed using the vector $\hat{\mathbf{s}}_n = [\phi_n, \omega_n, \theta_n]$, where $\hat{\mathbf{s}}_n$ denotes the vector of the PDW parameters of the nth pulse, ϕ_n is the RF of the nth pulse, ω_n is the PW of the nth pulse utilized for fine-grained clustering, and θ_n is the DOA of the nth pulse. This selective use of features enhances both the stability and effectiveness of the clustering process for the virtual PRI extraction. Our fine-grained clustering scheme can be implemented using common clustering algorithms, such as K-means [38], artificial bee colony K-means (ABC-K) [41], and particle swarm optimization K-means (PSO-K) [42], [43]. Once the clustering algorithm terminates, the TOAs of the pulses within the same cluster are used to obtain virtual PRIs through the differential operation.

In order to characterize the time-domain variations in the virtual PRIs, we further calculate the first-order differential virtual PRI associated with the nth pulse as

$$\Delta\tau_{k,n}^{\text{virtual}} = \tau_{k,n+1}^{\text{virtual}} - \tau_{k,n}^{\text{virtual}}. \tag{3}$$

2) Effectiveness of Virtual PRI: As outlined in Section IV-A, our virtual PRIs are derived through fine-grained clustering. To demonstrate the effectiveness of virtual PRIs under complex PRI modulation and missing pulses conditions, we take the pulses from the kth cluster as an example for illustration. Specifically, the virtual PRI computation process can be represented by

$$\tau_{k,n}^{\text{virtual}} = t_{k,n+1} - t_{k,n} = \sum_{i\in\hat{\mathcal{I}}_{k,n}} \tau_i^{\text{true}} + \Delta\tau \tag{4}$$

where $\tau_{k,n}^{\text{virtual}}$ denotes the virtual PRI for the nth pulse in the kth cluster. $t_{k,n}$ denotes the TOA of the nth pulse in the kth cluster. $\hat{\mathcal{I}}_{k,n}$ denotes the set of indices for the consecutive true PRIs, which are from the same radar as the nth pulse of the kth cluster within the interval $[t_{k,n}, t_{k,n+1}]$. τ_i^{true} denotes the ith involved true PRI. It should be noted that, as shown in Fig. 2, $\hat{\mathcal{I}}_{k,n}$ should include indices of missing pulses and the pulses from the same radar as the nth pulse but not assigned to the kth cluster. $\Delta\tau$ in (4) represents the interference term caused by nearby pulses originating from different radars.

[3]In scenarios where the number of radars is unknown, a density-based clustering algorithm, such as DBSCAN, can be utilized initially to roughly estimate the number of radars, thereby guiding the configuration of K.

Since virtual PRI calculation relies directly on the clustering results, we face two distinct cases: accurate cases, where consecutive pulses originate from the same radar as shown in Fig. 2(a) and (c), and inaccurate cases, where consecutive pulses stem from different radars as shown in Fig. 2(b) and (d). From Fig. 2, in the case of accurate clustering, the interference term $\Delta\tau = 0$, and the derivation of the virtual PRI under the fixed PRI involves the repetitive summation of identical true PRIs, while the derivation of the virtual PRI under PRI jitter modulation accumulates several true PRIs with varying values. Since the virtual PRIs in both scenarios are derived from the accumulation of true PRIs, the calculation of virtual PRIs remains robust even under complex PRI modulation patterns. As shown in Fig. 2, missing pulses only change the number of true PRIs accumulated for virtual PRI extraction without introducing spurious PRIs, and thus, the derived virtual PRI still faithfully reflects the actual PRI modulation pattern.

From Fig. 2, in the cases of inaccurate clustering, the obtained virtual PRI would deviate from the cumulative value of true PRIs due to the nonzero interference term $\Delta\tau$. Although this deviation might lead to degradation in RSS accuracy, the fine-grained clustering method employed in Section IV-A significantly reduces the probability that pulses from different radars are assigned to the same cluster, which effectively limits the occurrence of the inaccurate clustering cases. Moreover, we leverage both stacked sequences and individual pulses to extract the global temporal patterns and local pulse characteristics embedded in the virtual PRIs, thereby fully exploiting their complementary nature for more robust RSS. In summary, when facing missing pulses and complex PRI modulations, our virtual PRI-based scheme offers a more reliable surrogate for the actual PRI patterns, enabling effective RSS.

3) Dual-Input Format: After obtaining the virtual PRIs, the nth radar pulse in the kth cluster can be characterized by a vector, i.e.,

$$\mathbf{p}_{k,n} = \left[t_{k,n}, \phi_{k,n}, \omega_{k,n}, \alpha_{k,n}, \theta_{k,n}, \tau_{k,n}^{\text{virtual}}, \Delta\tau_{k,n}^{\text{virtual}}\right] \tag{5}$$

where $\mathbf{p}_{k,n}$ is an augmented version of the vector s_n of PDW parameters, $\tau_{k,n}^{\text{virtual}}$ is the virtual PRI extracted for the nth pulse in the kth cluster, and $\Delta\tau_{k,n}^{\text{virtual}}$ is the first-order differential virtual PRI extracted for the nth pulse in the kth cluster. To enable the subsequent dual-path network to extract features from both pulse-sequence-level and single-pulse-level features, we organize the $\mathbf{p}_{k,n}$'s associated with each fine-grained cluster of pulses into two distinct input formats, which would be fed into the two input paths of the dual-path network. The first format, designed to capture the sequence-level features, stacks the vectors $\mathbf{p}_{k,n}$ associated with N_{stack} consecutive pulses into the $N_{stack} * 7$ input matrix, i.e.,

$$\mathbf{P}_{k,n,N_{stack}} = \left[\mathbf{p}_{k,n}, \mathbf{p}_{k,n+1}, \ldots, \mathbf{p}_{k,n+N_{stack}-1}\right] \tag{6}$$

where k, n, and N_{stack} denote the index of the fine-grained cluster, the index of the corresponding pulse, and the number of stacked-pulse vectors, respectively. This stacked representation preserves the temporal relationships between adjacent

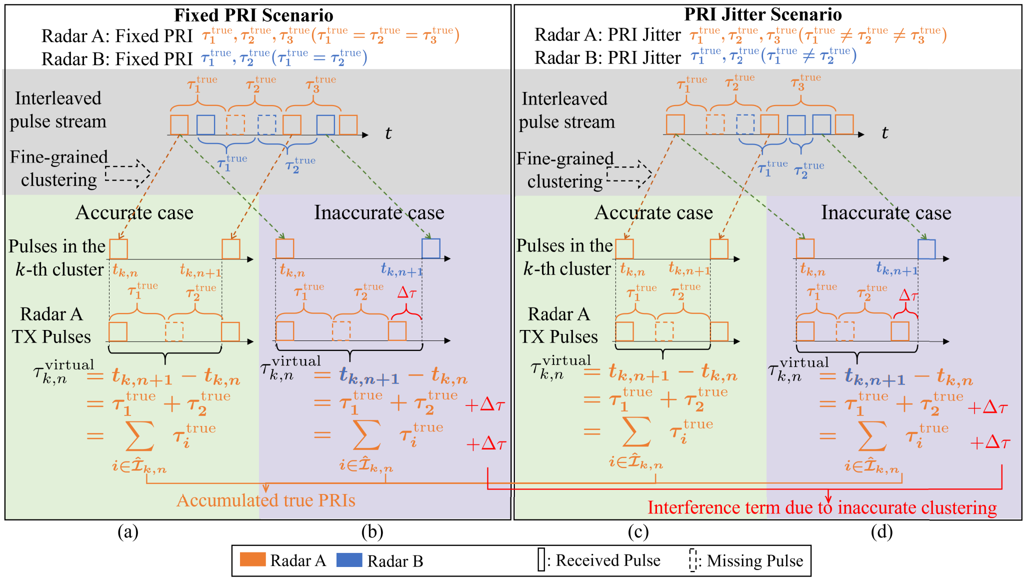


Fig. 2. Computational workflows of virtual PRIs under fixed PRI/PRI jitter scenarios. (a) and (b) Results for consecutive pulses from the same radar versus different radars within the same cluster under fixed PRI scenarios. (c) and (d) Equivalent results under PRI jitter scenarios.

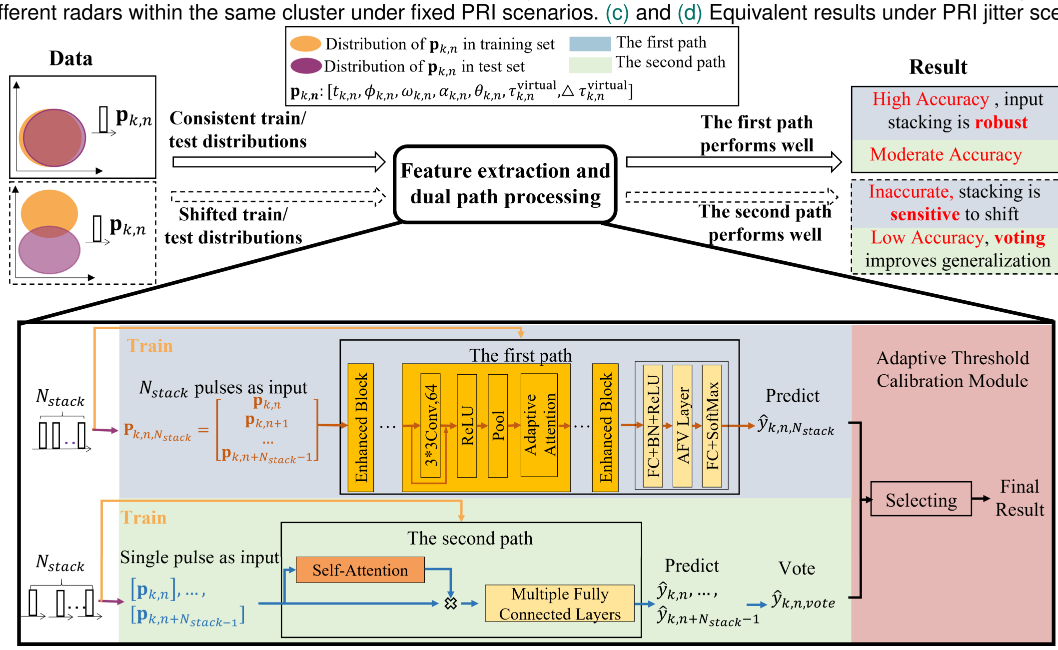


Fig. 3. Design of the dual-path network for RSS. When training and testing sets share consistent distributions, the first path ensures robustness via stacked-pulse design; under distribution shifts, the second-path single-pulse processing with voting mechanisms improves generalization.

pulses, allowing the network to learn from pulse-sequence-level patterns. The second format processes each individual pulse vector $\mathbf{p}_{k,n}$ independently, focusing on the distinctive characteristics of single pulses. By simultaneously employing these complementary representations, the dual-path network can effectively leverage both the global sequence-level patterns and the local pulse features for robust RSS.

B. Dual-Path Network Design

To extract distinct RSS-related information from different formats of the input, the dual-path network uses different structural designs, as shown in Fig. 3. Each sample input to the first path consists of the PDW- and PRI-related measurements extracted from a sequence of consecutive pulses, which facilitates the extraction of the time-varying patterns among these measurements, i.e., the global features. Each sample input to the second path is the measurement associated with a single radar pulse. This path extracts the fine-grained characteristics of individual pulses, i.e., the local features. The adaptive threshold calibration module dynamically integrates the outputs from both paths to produce the final reliable prediction. The optimization loss function jointly guides the learning of both paths. Sections IV-B1–IV-B4 elaborate on each of these components and their synergistic interactions. For more detailed input and output processes of these two paths, see Fig. 3.

1) First Path: The first path extracts global features, i.e., the time-varying patterns of the virtual PRIs and PDWs extracted from the consecutive pulse sequences assigned to the same fine-grained pulse cluster. Since these global features focus

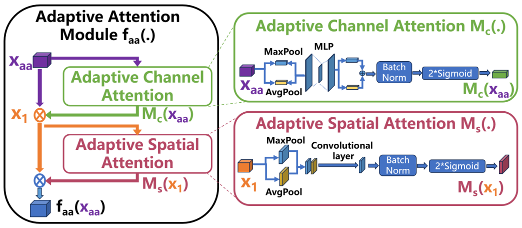


Fig. 4. Detailed structure of the adaptive attention module.

more on the overall distribution of the pulse sequence, they provide robustness against noise. The first path comprises five enhanced visual geometry group (VGG) blocks, adding an adaptive attention module [44] and a residual structure [45] to the original structure.

The structure of the enhanced block is shown in Fig. 3. The first enhanced block accepts the $\mathbf{P}_{n,k,N_{stack}}$ as its input, and subsequent enhanced blocks process the outputs of predecessor blocks. Given that the enhanced blocks share the same network structure, we denote the input of the enhanced block as $\boldsymbol{x}$ to illustrate their detailed design. First, $\boldsymbol{x}$ is processed with a convolution operation $f_{conv}(\cdot)$ to extract features, followed by a nonlinear transformation using the activation function $ReLU(\cdot)$. Next, the max-pooling operation $f_{mp}(\cdot)$ is applied to downsample the extracted features. Finally, the adaptive attention module $f_{aa}(\cdot)$ is used to refine the downsampled feature representation by emphasizing important parts, resulting in the final output $\hat{\boldsymbol{x}}$. This process can be expressed as

$$\hat{\boldsymbol{x}} = f_{aa}\left(f_{mp}\left(ReLU\left(\boldsymbol{x} + f_{conv}\left(\boldsymbol{x}\right)\right)\right)\right). \tag{7}$$

Since the first path takes multiple consecutive pulses as a single sample, its feature distribution is more complex than that of a single pulse. Therefore, an adaptive attention module is introduced to highlight RSS-related features in both channel and spatial dimensions. The detailed structure of the adaptive attention module is shown in Fig. 4. To extract features from both the channel and spatial dimensions, the adaptive attention module $f_{aa}(\cdot)$ includes an adaptive channel attention module $M_c(\cdot)$ and an adaptive spatial attention module $M_s(\cdot)$. Let $\boldsymbol{x}_{aa}$ be the input to $f_{aa}(\cdot)$. It first passes through $M_c(\cdot)$ to obtain $M_c(\boldsymbol{x}_{aa})$. Afterward, $\boldsymbol{x}_{aa}$ is multiplied element-wise with $M_c(\boldsymbol{x}_{aa})$ to produce $\boldsymbol{x}_1$, which is then input to $M_s(\cdot)$ to obtain $M_s(\boldsymbol{x}_1)$. Finally, $\boldsymbol{x}_1$ is multiplied element-wise with $M_s(\boldsymbol{x}_1)$ to generate the output $\boldsymbol{x}_2$. These operations inside the adaptive attention module can be represented by

$$\boldsymbol{x}_2 = \underbrace{M_s\left(\underbrace{M_c\left(\boldsymbol{x}_{aa}\right)\otimes \boldsymbol{x}_{aa}}_{\boldsymbol{x}_1}\right)\otimes\left(\underbrace{M_c\left(\boldsymbol{x}_{aa}\right)\otimes \boldsymbol{x}_{aa}}_{\boldsymbol{x}_1}\right)}_{f_{aa}(\boldsymbol{x}_{aa})}. \tag{8}$$

The adaptive channel attention module $M_c(\cdot)$ is used to efficiently extract a channel attention feature map by dynamically weighting each channel to enhance important features and suppress less relevant ones. The channel attention feature map is obtained through pooling, down-sampling, and nonlinear transformation on $M_c(\cdot)$ at the channel level. The operations inside $M_c(\cdot)$ can be expressed as

$$M_c\left(\boldsymbol{x}_{aa}\right) = 2\, f_s\left(f_{bn}\left(f_{mlp}\left(f_{ap}\left(\boldsymbol{x}_{aa}\right)\right) + f_{mlp}\left(f_{mp}\left(\boldsymbol{x}_{aa}\right)\right)\right)\right) \tag{9}$$

where $f_s(\cdot)$ is the sigmoid activation function, $f_{bn}(\cdot)$ represents the batch normalization (BN) layer, $f_{mlp}(\cdot)$ represents the multilayer perceptron (MLP), $f_{ap}(\cdot)$ is the average pooling operation, and $f_{mp}(\cdot)$ represents the max-pooling operation. The scaling factor of 2 in (9) is directly related to the properties of the sigmoid activation function. Since the sigmoid function constrains its outputs to the range $(0, 1)$, the subsequent scaling of its output by 2 expands the effective operating range of the attention weights to $(0, 2)$. This enables the adaptive attention module to perform both suppression (when the weights fall between 0 and 1) and amplification (when the weights fall between 1 and 2) of input features. This expanded range enhances the model's ability to adaptively emphasize informative features while suppressing less relevant ones.

Besides the adaptive channel attention module, the adaptive spatial attention module generates dynamic spatial attention maps to extract channel-wise information for effective RSS. As shown in Fig. 4, input $\boldsymbol{x}_1$ [from (8)] undergoes $f_{mp}(\cdot)$ and $f_{ap}(\cdot)$ for channel-wise dimensionality reduction. The reduced features are processed by $f_{conv}(\cdot)$ and $f_{bn}(\cdot)$ to obtain spatial weights and then nonlinearly transformed via $f_s(\cdot)$ to produce the spatial attention map. The module $M_s(\cdot)$ is expressed as

$$M_s\left(\boldsymbol{x}_1\right) = 2\, f_s\left(f_{bn}\left(f_{conv}\left(\left[f_{ap}\left(\boldsymbol{x}_1\right), f_{mp}\left(\boldsymbol{x}_1\right)\right]\right)\right)\right). \tag{10}$$

As shown in Fig. 3, the stacked measurements of N_{stack} pulses, represented by $\mathbf{P}_{k,n,N_{stack}}$, are fed into the first path to yield a prediction $\hat{y}_{k,n,N_{stack}}$ will be fused with the second-path prediction in the adaptive threshold calibration module.

2) Second Path: Unlike the first path extracting global features, the second path extracts local features, i.e., the fine-grained characteristics of individual pulses. Consequently, the second path utilizes individual pulse measurement $\mathbf{p}_{n,k}$ as input. Compared with the global features, local features enable the network to capture subtle consistent characteristics with different pulses under the same operating mode, further improving the sorting performance.

Traditional convolutional neural networks (CNNs) struggle with individual pulse features in the second path due to fixed convolution parameters, which cannot adapt to feature distribution shifts from varying radar modes. Conversely, self-attention dynamically weights features by importance, correcting shifts and capturing critical local information. Thus, the second path employs a self-attention module [46] followed by three fully connected layers. The module transforms input $\boldsymbol{x}$ into $\boldsymbol{Q}_x = f_{\boldsymbol{Q}_x}(x)$, $\boldsymbol{K}_x = f_{\boldsymbol{K}_x}(x)$, and $\boldsymbol{V}_x = f_{\boldsymbol{V}_x}(x)$ via linear transformations. It computes similarity between $\boldsymbol{Q}_x$ and $\boldsymbol{K}_x$, applies softmax normalization, and then weights $\boldsymbol{V}_x$ for final representation

$$self\text{-}attention\left(\boldsymbol{Q}_x, \boldsymbol{K}_x, \boldsymbol{V}_x\right) = Softmax\left(\frac{\boldsymbol{Q}_x \boldsymbol{K}_x{}^T}{\sqrt{d_{\boldsymbol{K}_x}}}\right)\boldsymbol{V}_x \tag{11}$$

where $d_{\boldsymbol{K}_x}$ is the dimension of $\boldsymbol{K}_x$.

The second path iteratively processes $\mathbf{p}_{k,n}$ across N_{stack} pulses, generating N_{stack} predictions

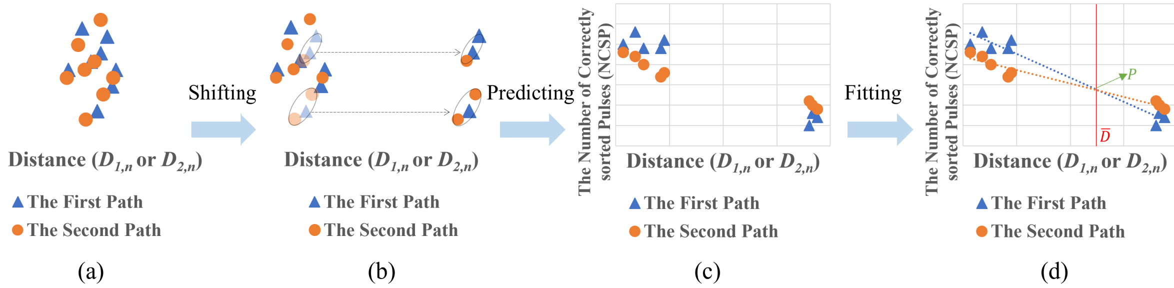


Fig. 5. Schematic of the threshold selection in the adaptive threshold calibration module. (a) Original feature distribution. (b) Distribution after shifting a randomly selected portion of the data. (c) Results of the prediction of the shifted data using the dual paths. (d) Linear fitting of the relationship between the NCSP and $D_{1,n}$ (or $D_{2,n}$) for the dual paths. P is the intersection point of the fit lines for the dual paths, and the corresponding distance is the calibration threshold $\bar{D}$.

$[\hat{y}_{k,n}, \hat{y}_{k,n+1}, \ldots, \hat{y}_{k,n+N_{stack}-1}]$. These outputs undergo voting to produce a pulse-sorting result $\hat{y}_{k,n,vote}$.

Remark 1: In the proposed method, the dual-path network is introduced to handle the complex relationship between the training set and the test set, as illustrated in Fig. 3. The global features extracted by the first path represent time-varying patterns of virtual PRIs and PDWs from consecutive pulse sequences. These patterns remain consistent when training and test data distributions align, benefiting the model's sorting task. However, as the discrepancy between PRI/PDW patterns in training and test sets increases, the global features mislead sorting decisions, reducing overall accuracy. Conversely, the second path extracts local features: fine-grained characteristics of individual pulses. By avoiding time-varying patterns, this path escapes the negative impact of global feature distribution shifts. Meanwhile, the majority voting mechanism further enhances the second-path generalization capability. Hence, by leveraging both paths, the proposed approach enhances robustness against distribution shifts. The first path effectively captures PRI sequences when consistency is maintained, while the second path provides an adaptive representation when variations occur. To mitigate sorting errors and ensure reliable performance across various training/test distributions, we will present the calibration module design Section IV-B3, which adaptively chooses between the two paths' predictions based on distribution shifts.

3) Adaptive Threshold Calibration Module: Two prediction results $\hat{y}_{k,n,N_{stack}}$ and $\hat{y}_{k,n,vote}$ are generated from the dual-path network. Given the complex feature distributions of different radar pulses, it must be carefully determined which prediction is more reliable to improve the number of correctly sorted pulses (NCSP). The adaptive threshold calibration module is designed to enable such a decision-making process. Specifically, this module takes the prediction $\hat{y}_{k,n,N_{stack}}$ of the first path as the final output if it is considered to be reliable. Otherwise, the prediction $\hat{y}_{k,n,vote}$ of the second path is used as the final sorting result.

The reliability of the outputs of the two paths is determined by quantifying the distribution shift between the training and test sets, using the distances from the features of test samples to the feature centers of the training set. Our approach has two main considerations. On the one hand, the first path, processing stacked-pulse sequences to extract global features, is expected to maintain higher reliability under minor distribution shifts since global features are inherently insensitive to minor distribution shifts. On the other hand, the second path leverages single-pulse features with voting and is expected to be more reliable under significant distribution shifts. This is because its voting mechanism can discard outliers and identify the consensus among pulses, thereby providing robustness even when the global features may be substantially altered. To determine the reliability of the two prediction results $\hat{y}_{k,n,N_{stack}}$ and $\hat{y}_{k,n,vote}$, we define $\mathbf{z}_{1,n}$ as the feature vector extracted by the first path from $\mathbf{P}_{k,n,N_{stack}}$, which is obtained as the output of the abstract feature vector (AFV) layer of the first path, i.e.,

$$\mathbf{z}_{1,n} = f_{path_1}\left(\mathbf{P}_{k,n,N_{stack}}\right) \tag{12}$$

where $f_{path_1}(\cdot)$ is the mapping of the first path. We employ the distance from the extracted first-path AFV $\mathbf{z}_{1,n}$ to the M radar-wise feature vector centers (FVCs) $\mathcal{C}_1 = \{\mathbf{C}_{1,1}, \mathbf{C}_{1,2}, \ldots, \mathbf{C}_{1,M}\}$ to quantify the degree of the distribution shift in 3PRI/PDW time-varying patterns between the training and test sets. Each of the M radar-wise FVCs in $\mathcal{C}_1$ is derived from the training set by averaging over all the extracted first-path AFVs $\mathbf{z}_{1,n}$ associated with the training pulse sequences of a specific radar, i.e.,

$$\mathbf{C}_{1,m} = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{|\mathcal{N}_{m,k}|}\sum_{n\in\mathcal{N}_{m,k}}\mathbf{z}_{1,n}, \quad m = 1, 2, \ldots, M \tag{13}$$

where $\mathcal{N}_{m,k}$ represents the set of pulse sequences associated with the mth radar within the kth cluster. Based on these FVCs, the distance $D_{1,n}$ associated with the nth test pulse sequence and the first path can be expressed as

$$D_{1,n} = \min_{m=1,2,\ldots,M} \|\mathbf{z}_{1,n} - \mathbf{C}_{1,m}\|_2, \quad n = 1, 2, \ldots, N. \tag{14}$$

The rationale behind our distribution shift quantification is that, in the same feature space, the closer the AFVs of the test samples are to the FVCs derived from the training set, the smaller the feature distribution shift is.

With the obtained distribution shift quantification $D_{1,n}$, the calibration module employs a threshold $\bar{D}$ to determine whether the distribution shift in the first path is sufficiently severe to resort to the second path for final prediction. Note that the calibration module is introduced to improve the NCSP, we propose to estimate the threshold $\bar{D}$ by identifying the feature space distance where the NCSP of the second path exceeds that of the first path. To facilitate the calculation of this threshold, we calculate the feature space distance $D_{2,n}$ between the second-path $\mathbf{z}_{2,n}$ and the $\mathcal{C}_2$ by following procedures similar to (13) and (14). Then, we attempt to determine the relationships between $D_{1,n}$ (or $D_{2,n}$) and the NCSP for both paths by simulating a shifted test set distribution with randomly selected samples from the training set. When simulating the distribution shift in the test set, we add an offset to the selected training samples in the feature spaces to ensure the inverse correlation between the amount of distribution shift and the NCSP is well captured. Once the NCSP of the shifted training samples achieved by both paths is collected, we approximate the relationship between $D_{1,n}$ (or $D_{2,n}$) and the NCSP for both paths via linear mapping and take the distance at the intersection P as the threshold $\bar{D}$ for prediction result selection. The entire threshold selection process is shown in Fig. 5. During the derivation of the calibration threshold $\bar{D}$, the BN layers in both paths normalize the extracted AFV to a distribution with zero mean and unit variance during training. This operation guarantees that the distance metrics $D_{1,n}$ and $D_{2,n}$ maintain small values when the training and test distributions remain consistent, while both exhibit significantly larger values when a distribution shift occurs. This synchronized behavior enables reliable calibration threshold selection, as now both distance measures share common distributional characteristics that make their comparison meaningful.

It should be noted that the true relationship between the feature space distance and the NCSP is likely nonlinear, making the linear mapping a practical approximation. Furthermore, when deployed in a new environment, $\bar{D}$ can be efficiently recalibrated offline using a small set of manually annotated data. Currently, $\bar{D}$ is determined empirically due to the lack of a stable mathematical criterion for dynamic optimization. While this fixed approach is highly effective, developing a fully learnable thresholding mechanism remains a valuable objective for future research.

4) Optimization Loss Function: The adaptive threshold calibration module computes each radar's feature center as the mean of AFV outputs from the first path. However, AFVs from the same radar often disperse in feature space, rendering FVCs ineffective for representing true distributions and degrading module performance. To address this, we optimize the dual-path network with a loss function concentrating same-class AFVs while dispersing different classes. The optimization loss combines classification, intraclass, and interclass losses, i.e.,

$$\mathcal{L}_{total} = \mathcal{L}_c + \gamma \mathcal{L}_{intra} - \beta \mathcal{L}_{inter} \tag{15}$$

where $\mathcal{L}_c$ is the cross-entropy loss, γ and β are the weight parameters. The parameters γ and β are empirically set to ensure that the three loss components remain within the same order of magnitude, preventing any single objective from dominating the gradient updates. The intraclass loss function $\mathcal{L}_{intra}$ is introduced to reduce the mean distance between the AFVs and the corresponding FVCs in (13), i.e.,

$$\mathcal{L}_{intra} = \frac{1}{N}\sum_{n=1}^{N} \|\mathbf{z}_{1,n} - \mathbf{C}_{1,y_n}\|_2. \tag{16}$$

The interclass loss function $\mathcal{L}_{inter}$ is introduced to enlarge the distances between the M FVCs of different radars, i.e.,

$$\mathcal{L}_{inter} = \frac{1}{M(M-1)}\sum_{m=1}^{M}\sum_{i=1,i\neq m}^{M} \|\mathbf{C}_{1,m} - \mathbf{C}_{1,i}\|_2. \tag{17}$$

The second path is optimized using the same method applied in (15), (16), and (17). Therefore, the feature and decision boundaries are simultaneously optimized in the feature space for better RSS.

C. Overall Process

This section summarizes the workflow of the proposed virtual PRI-assisted RSS scheme, including both the training and inference stages. As a prerequisite for both stages, the raw interleaved pulse stream undergoes the radar signal preprocessing procedure as described in Section IV-A. This includes fine-grained PDWs clustering to obtain high-purity pulse clusters, the extraction of the virtual PRI for each pulse, and the dual-input format reformulation of the pulse vector $\mathbf{p}_{k,n}$.

During the training stage, these dual-format inputs are utilized to train the two input paths of the dual-path network, aiming to maximize the prediction accuracy of each path independently (as described in Section IV-B). During the inference stage, an additional step is introduced where the adaptive threshold calibration module selects the final predictions output by the two paths. Specifically, the first path outputs a single prediction result $\hat{y}_{k,n,N_{stack}}$ based on the matrix $\mathbf{P}_{k,n,N_{stack}}$ of the N_{stack} stacked-pulse vectors. In parallel, the second path processes each individual pulse vector $\mathbf{p}_{k,n}$ independently and generates N_{stack} individual predictions $[\hat{y}_{k,n}, \hat{y}_{k,n+1}, \ldots, \hat{y}_{k,n+N_{stack}-1}]$ for these N_{stack} pulses. To align the dimension of the outputs from these two paths, the N_{stack} independent predictions from the second path are aggregated through a soft voting mechanism—specifically, by averaging their softmax probability vectors and selecting the class with the maximum average probability—to obtain the prediction $\hat{y}_{k,n,vote}$. The adaptive threshold calibration module then selects the more reliable prediction between $\hat{y}_{k,n,N_{stack}}$ and $\hat{y}_{k,n,vote}$ as the final result. Finally, clusters sharing the same label are merged, effectively reconstructing the complete pulse streams of each radar emitter.

V. Experimental Settings

This section describes the setup for evaluating the proposed method. We first introduce the radar signal simulation datasets. Then, the benchmark schemes used in this article for RSS are presented. Finally, the experimental setup is provided.

A. Datasets Description

To verify our method's effectiveness, two datasets are simulated with [21], [24], and [37]. Parameters for both are listed in Tables II and III. Dataset 1 evaluates the basic performance of the proposed method. Dataset 2 simulates real-world conditions with unknown radar modes to evaluate the effectiveness of the proposed RSS scheme under complex and challenging scenarios. Each radar mode contains 200 000 pulses; 10% are randomly dropped to simulate missing pulses.

TABLE II
PARAMETER SETTINGS FOR EACH RADAR EMITTER IN DATASET 1

Radar	Mode	DOA/°	PW/μs	RF/MHz	PRI/μs	Pulse number
1	i	56∼60	8.5∼9.5	3520 fixed	1000 fixed	180k
2	i	41∼45	8.4∼9.8	3030 fixed	[390,427,550,373,415,505,402, 446,506] Pulse Group Agility	180k
3	i	40∼44	8.8∼9.8	3000∼3150 Inter-Pulse Agility	1160 fixed	180k
4	i	46∼49	8.5∼9.6	3025 fixed	1505∼1565 sliding	180k
5	i	47∼50	7.6∼8.5	2890∼3050 Inter-Pulse Agility	[1339,1695,1866,2040,2053, 2140,2250,1996] stagger	180k
	ii	47∼50	8.7∼9.7	2890∼3050 Inter-Pulse Agility	[1339,1695,1866,2040,2053, 2140,2250,1996] stagger	180k
6	i	45∼48	8.75∼10.25	3040 fixed	1260(5%) jitter	180k
7	i	58∼62	9.0∼10.5	3100∼3310 Inter-Pulse Agility	900(5%) jitter	180k
8	i	59∼64	9.2∼10.2	[3410,3470,3440,3290, 3380,3350] stagger	500 fixed	180k
9	i	59∼62	7.3∼8.3	3155 fixed	[1890,1968,1905,1983,1921, 2000,1937,2008,1945,1991] stagger	180k
	ii	59∼62	9.0∼10.0	3155 fixed	[1890,1968,1905,1983,1921, 2000,1937,2008,1945,1991] stagger	180k
10	i	52∼57	8.6∼9.9	[2900,2960,2940,3000,2980, 3020,3060,3040,3080,3150] Pulse Group Agility	[1295,1315,1300,1335,1355, 1375,1350,1330,1360,1375] Pulse Group Agility	180k

Specifically, the simulated pulse streams for each radar emitter are generated independently based on their operating modes and PRI modulation patterns. These individual streams are then interleaved chronologically according to their TOAs to simulate the receiving process. In our simulation setup, we enforce an equal pulse count for each radar rather than restricting them to a unified observation time window. This design is deliberately chosen to prevent the neural network from overfitting to radar categories that would otherwise produce a significantly larger number of pulses, which would inevitably degrade the overall sorting performance. Training and test set compositions for both datasets are described as follows.

Training Sets: In Dataset 1, the training set includes all radar operating modes, with the detailed parameters provided in Table II. In Dataset 2, the training set includes only the first operating mode of each radar, with the detailed parameters provided in Table III.

Test Sets: Simulating three test scenarios (TSs) per dataset (Table IV) with operating modes labeled i–iv and a summary row for total radar counts, we verify the proposed method's pulse-sorting performance under varying radar counts. Dataset 1 maintains consistent radar modes between training and test sets, whereas Dataset 2 exhibits different modes during testing. This difference causes unseen modes and feature distribution mismatch in Dataset 2, posing greater generalization challenges and making RSS more demanding than in Dataset 1.

B. Benchmark Schemes

To demonstrate the advantages of the proposed dual-path network, we benchmark typical RSS task algorithms for performance evaluation. Existing methods are detailed as follows.

FCNN: The fully connected neural network [47] used for the RSS tasks is constructed with three fully connected layers, where each fully connected layer is followed by a normalization layer and an activation layer.

ResGCN-RSS: ResGCN-RSS [21] employs in a semi-supervised manner for RSS tasks. It constructs radar signal graphs via K-nearest neighbor (KNN) so that a three-layer ResGCN can be used for adaptive feature learning and end-to-end RSS implementation.

ABMR: The attention-based multiRNN [16] combines an attention mechanism with multiple RNNs to capture the information between pulses in complex scenarios for RSS.

SVM: Since the support vector machine (SVM) [48] can handle high-dimensional data and extract nonlinear features, we use the SVM with a polynomial kernel function as one of the benchmarks for RSS.

RNN-FCNN: The RNN-FCNN [30] iteratively sorts radar pulses by training on PRI/PW parameters, bidirectionally predicting sequences from a starting pulse, and assigning pulses to radars when predictions match measurements within a threshold until completion.

LSTM-FCNN: Building on the iterative RSS approach of RNN-FCNN [30], LSTM-FCNN [31] extends it by utilizing more features (PRI, PW, RF, and DOA) and replacing the RNN with an LSTM.

LPACD: The label propagation algorithm community detection [26] utilizes a limited penetration visibility graph to characterize the relationships between different pulse measurements. Based on the obtained graph, the label propagation algorithm and the density peak clustering algorithm are used to complete the signal sorting.

BLSTM: BLSTM [25] proposes a semantic segmentation approach for RSS. Contrasting iterative methods, it processes TOA differences and assigns semantic labels through sequence modeling. A bidirectional LSTM captures forward/backward correlations, enabling feature-agnostic radar separation.

C. Experimental Setup

We use PyTorch as the implementation platform for all the experiments. The dual-path network is trained using the

TABLE III
PARAMETER SETTINGS FOR EACH RADAR EMITTER IN DATASET 2

Radar	Mode	DOA/°	PW/μs	RF/MHz	PRI/μs	Pulse number
1	i	56∼60	8.5∼9.5	[9591∼9596,9614∼9619] stagger	[40,45,50,55,60,65,70,75,80,90,100,110, 130,140] Pulse Group Agility	180k
	ii	43-50	5.875∼7.625	[9730∼9737] jitter	[40,45,50,55,60,65,75,80,87.5,97.5,105, 117.5]Pulse Group Agility	180k
	iii	249∼254	8.3∼9.0	[9101∼9108] jitter	[40,45,50,55,60,65,75,80,87.5,97.5,105, 117.5]Pulse Group Agility	180k
	iv	60∼65	8.5∼9.5	[9603∼9608,9613∼9617] stagger	[40,45,50,55,60,65,70,75,80,90,100,110, 130,140]Pulse Group Agility	180k
2	i	41∼45	8.4∼9.8	[9610∼9620] jitter	[40,45,55,65,75,85,95,115,125,145]stagger	180k
	ii	52∼56	[7.0∼7.6,7.95∼8.6]stagger	[9102∼9108] jittered	[40,45,55,65,75,85,95,115,125,145]stagger	180k
	iii	56∼60	8.5∼9.8	[9585∼9593,9603∼9608]stagger	[40,45,55,65,75,85,95,115,125,145]stagger	180k
3	i	39∼45	8.85∼9.85	[9605.5∼9612] jitter	[35∼65] jitter	180k
	ii	58-64	6.125∼7.875	[9730∼9738] jitter	[58-74] jitter	180k
	iii	74∼80	8.0∼8.65	[9086∼9092.5] jitter	[57.5∼75.5] jitter	180k
	iv	35∼41	8.85∼9.85	[9600∼9606] jitter	[57.5∼90.1] jitter	180k
4	i	45∼50	8.45∼9.65	[9636.25∼9642.75] jitter	[40,45,50,55,60,65,75,80,90,100,110,120] Pulse Group Agility	180k
	ii	38∼44	8.5∼9.6	[9609.75∼9616] jitter	[40,45,50,55,60,65,75,80,90,100,110,120] Pulse Group Agility	180k
5	i	46∼51	8.625∼10.50	[9591∼9596,9597∼9602, 9604∼9607,9621∼9626] Pulse Group Agility	[40,45,55,65,75,85,95,110] Pulse Group Agility	180k
	ii	170∼175	7.5∼8.9	[9091∼9100] jitter	[40,45,55,65,75,85,95,110] Pulse Group Agility	180k
	iii	47∼52	8.75∼10.50	[9585∼9596.5] jitter	[40,45,55,65,75,85,95,110] Pulse Group Agility	180k
6	i	44∼49	[7.6∼8.5,8.7∼9.7]stagger	[9578.75∼9583, 9591.25∼9596.25] stagger	[40,45,55,65,75,85,95,110] stagger	180k
	ii	61-65	5.875∼8.0	[9725∼9732] jitter	[42.5,47.5,55,63,75,85,97.5,112.5]stagger	180k
	iii	270∼276	[6.7∼7.4,8.25∼8.9]stagger	[9102∼9107,9110∼9114]stagger	[40,45,55,65,70,80,95,108.5]stagger	180k
	iv	38∼43	[7.3∼8.3,9.0∼10.0]stagger	[9595∼9600,9600∼9605]stagger	[40,45,55,65,70,80,95,108.5]stagger	180k
7	i	58∼62	8.75∼10.75	[9557∼9563.5] jitter	[40,45,50,55,60,65,70,80,90,95]stagger	180k
	ii	76∼81	7.4∼8.8	[9111∼9117.5] jitter	[40,45,50,55,62.5,67.5,75,82.5,90,100] stagger	180k
	iii	36∼41	9.00∼10.75	[9579.5∼9586] jitter	[40,45,50,55,62.5,67.5,75, 80,87,97] stagger	180k
8	i	59∼64	9.2∼10.2	[9560∼9570,9574∼9579, 9612∼9620] Pulse Group Agility	[10,15,20,25,30,40] Pulse Group Agility	180k
	ii	52-57	0.875∼2.75	[9730∼9736,9736∼9742] Pulse Group Agility	[10,13,17.5,22.5,29,39] Pulse Group Agility	180k
	iii	280∼285	1.55∼2.15	[9089∼9093,9095∼9099, 9108∼9112] Pulse Group Agility	[10,13,17.5,22.5,29,39] Pulse Group Agility	180k
	iv	51∼56	9.25∼10.25	[9588∼9592.5,9590∼9594.5, 9596∼9600.5] Pulse Group Agility	[10,15,17.5,22.5,30,39] Pulse Group Agility	180k
9	i	58∼63	9.3∼10.3	[9558∼9563,9568∼9573, 9590∼9594,9596∼9601, 9608∼9613] stagger	[20,25,30,35,40,50] stagger	180k
	ii	227∼232	1.5∼2.1	[9082∼9090,9095∼9100, 9103∼9110] stagger	[20,25,30,35,40,50] stagger	180k
	iii	42∼48	9.15∼10.15	[9586∼9597,9601∼9609] stagger	[20,25,30,35,40,50] stagger	180k
10	i	52∼57	8.65∼9.95	[9574.8∼9579.8] jitter	[25∼43] jitter	180k
	ii	56∼61	8.65∼10.05	[9588.2∼9593.2] jitter	[18∼32] jitter	180k
11	i	39∼44	1.45∼2.45	[9579∼9584,9616∼9620, 9624∼9629,9648∼9652] Pulse Group Agility	[10,17.5,22.5,35] Pulse Group Agility	180k
	ii	38∼43	1.45∼2.5	[9578∼9583,9585∼9589, 9588∼9592.5,9592∼9597] Pulse Group Agility	[10,17.5,22.5,35] Pulse Group Agility	180k
12	i	50∼54	33∼38	[2824.5∼2829.5,2836∼2840.5] Pulse Group Agility	[400,435,465,510] stagger	180k
	ii	51∼56	34.5∼39	[2806.5∼2811.5,2811.5∼2816] Pulse Group Agility	[400,435,465,510] stagger	180k

TABLE IV

COMPOSITION OF TRAINING AND TEST SETS IN THE TWO DATASETS

	Dataset 1				Dataset 2			
Radar	Train	TS1	TS2	TS3	Train	TS4	TS5	TS6
1	i	i		i	i	ii	iii	iv
2	i			i	i		ii	iii
3	i		i	i	i	ii	iii	iv
4	i		i	i	i			ii
5	i\ii	i\ii	i\ii	i\ii	i		ii	iii
6	i	i	i	i	i	ii	iii	iv
7	i	i		i	i		ii	iii
8	i		i	i	i	ii	iii	iv
9	i\ii			i\ii	i		ii	iii
10	i		i	i	i			ii
11					i			ii
12					i			ii
No.	10	4	6	10	12	4	8	12

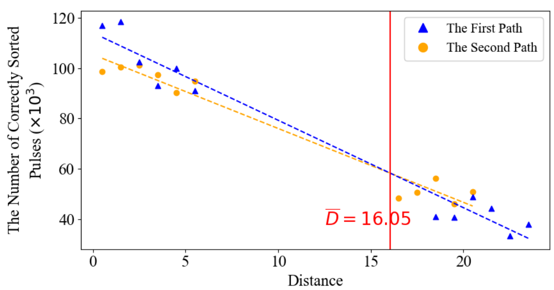


Fig. 6. Experimental results for obtaining the calibration threshold $\bar{D}$.

Adam optimizer with a batch size of 128, an initial learning rate of 0.0001, and a maximum number of iterations of 100. To determine the calibration threshold, we randomly select 30% of the data from the training set to simulate test sets with shifted feature distributions, and the calibration threshold obtained is $\bar{D} = 16.05$, as shown in Fig. 6. During the training process, γ and β in (15) are set to 0.01 and 0.001, respectively.

First, extensive experiments are conducted to evaluate the performance of the proposed virtual PRI-assisted RSS scheme in comparison to that of the benchmarks. Then, the benefits of incorporating virtual PRIs in signal sorting are studied, and the effect of fine-grained clustering with a different number of clusters K is analyzed. Next, we conduct detailed ablation experiments on each module of the dual-path network to evaluate its effectiveness. Finally, the benefit of introducing the adaptive threshold calibration module is investigated.

To ensure a fair and rigorous evaluation, all baseline methods and the proposed dual-path scheme were evaluated under a strictly unified experimental protocol. First, all models were trained and tested on the same raw datasets (Datasets 1 and 2). Regarding the input features, we strictly adhered to the original architectural designs of each baseline, providing them with their optimal required inputs (e.g., specific subsets of PDWs or full PDWs). It is worth noting that the virtual PRI feature utilized by our method is dynamically extracted from the exact same raw PDWs, ensuring that the performance comparison reflects superior feature representation capabilities rather than an unequal data advantage. Second, rather than relying on default settings, the primary hyperparameters of all baseline models were carefully fine-tuned on our specific datasets to maximize their respective convergence and potential performance. Finally, to provide a comprehensive and equitable evaluation beyond mere sorting accuracy, we further assess the models across multiple dimensions, including computational complexity and real-time inference latency, which will be detailed in Section VI-E.

VI. RESULTS AND ANALYSIS

A. Comparison With Benchmarks

We apply our virtual PRI-assisted RSS scheme and the eight benchmarks to the two datasets introduced in Section V-A. To demonstrate that our RSS scheme is robust and does not critically depend on the choice of a specific clustering algorithm for the initial fine-grained clustering step, we use three different clustering algorithms, K-means, ABC-K, and PSO-K for virtual PRI extraction, and the corresponding schemes are denoted as Dual-Path-K, Dual-Path-ABC and Dual-Path-PSO, respectively. For each clustering algorithm, the number of clusters is set to $K = 100$. The impact of K on RSS performance will be studied in Section VI-C. The experimental results are presented in Table V, for the two datasets, where the number in the bracket indicates the number of radars. The experimental results confirm the superior average sorting performance of the proposed RSS scheme. All dual-path variants (K, ABC, and PSO) surpass benchmarks in accuracy, demonstrating the scheme's flexibility across clustering algorithms for virtual PRI extraction. As shown in Table V, RSS accuracy on Dataset 2 is lower than on Dataset 1 due to unseen radar modes in the test set, causing distribution shift in PDWs and PRI measurements. This higher variation in Dataset 2 amplifies the generalization challenge.

Table V indicates that with the increasing number of radars, RSS accuracy decreases on Dataset 1 but improves on Dataset 2. This divergence occurs because, for Dataset 1, more radars increase feature distribution overlap, complicating radar distinction. Conversely, for Dataset 2, each radar exhibits varying distribution shifts. Radars with smaller shifts are more sortable, and since test scenarios randomly select radars, scenarios with more radars have a higher probability of containing easily sortable radars—explaining the accuracy increase.

As shown in Table V, most of the eight benchmarks underperform on these two datasets. The ResGCN-RSS scheme performs poorly because the KNN algorithm adopted is unable to construct a graph representation to characterize the complex relationships among the PDWs of numerous radar pulses. The LSTM-FCNN and RNN-FCNN schemes are prone to cumulative errors due to their iterative parameter estimation and pulse deinterleaving. The LPACD method adopts an unsupervised learning strategy that relies on feature distances to construct the limited penetration visibility graph and merge subcommunities, which limits its effectiveness in RSS tasks. BLSTM's sorting performance is limited by using only the TOA differences as input. In contrast, the FCNN and ABMR methods achieve high accuracy on Dataset 1 by effectively learning discriminative features from the training set's PDW distribution, while their performance significantly drops on Dataset 2 where the variations in radar operating modes cause substantial shifts in PDW parameter distributions between the

TABLE V
SIGNAL SORTING ACCURACY OF DIFFERENT METHODS ON DATASETS 1 AND 2. THE TEST SCENES ARE LABELED AS TSi(M), WHERE i IS THE SCENE INDEX AND M IS THE NUMBER OF RADARS IN THE SCENE

	Dataset 1				Dataset 2			
	TS1(4)	TS2(6)	TS3(10)	Avg. Acc	TS4(4)	TS5(8)	TS6(12)	Avg. Acc
FCNN	91.75%	84.30%	85.08%	87.04%	0.00%	25.00%	37.20%	20.73%
ABMR	89.56%	93.20%	**91.99%**	91.58%	13.20%	13.02%	33.01%	19.74%
SVM	69.99%	53.93%	52.67%	58.86%	0.00%	12.50%	30.82%	14.44%
ResGCN-RSS	24.99%	24.98%	16.67%	22.21%	25.03%	12.50%	8.34%	15.29%
LSTM-FCNN	71.32%	51.62%	43.21%	55.38%	30.70%	10.45%	7.83%	16.33%
RNN-FCNN	47.12%	29.11%	16.28%	30.84%	24.65%	9.31%	5.25%	13.07%
LPACD	50.50%	32.30%	22.50%	35.10%	**38.90%**	36.20%	25.90%	33.67%
BLSTM	81.52%	64.11%	50.88%	65.50%	33.68%	23.74%	20.95%	26.12%
Dual-Path-K	**98.90%**	**96.99%**	91.82%	**95.90%**	23.05%	45.14%	47.52%	38.57%
Dual-Path-ABC	98.90%	88.65%	90.24%	92.60%	23.05%	**47.18%**	**52.94%**	**41.06%**
Dual-Path-PSO	98.90%	96.99%	89.63%	95.17%	23.05%	44.91%	49.20%	39.05%

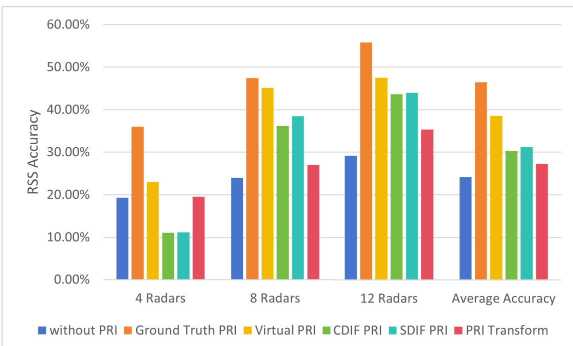


Fig. 7. RSS accuracy comparison on Dataset 2 under different PRI inputs, where the virtual PRI is replaced by: without PRI, ground-truth PRI, and the estimated PRIs produced by the PRI estimation algorithms embedded in CDIF, SDIF, and PRI transform (labeled as CDIF PRI, SDIF PRI, and PRI transform, respectively).

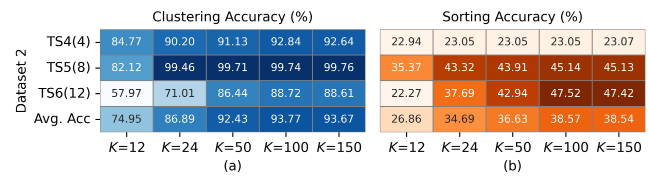


Fig. 8. Accuracy of our fine-grained clustering scheme and the sorting accuracy of our virtual PRI-assisted RSS scheme on Dataset 2 with different number of clusters K. In each subfigure, darker colors represent higher values of the corresponding metrics. (a) Clustering accuracy (%). (b) Sorting accuracy (%).

training and test data. Since these two methods are primarily optimized for the training dataset and lack specific mechanisms to maintain robustness under distribution shifts, they are expected to experience performance degradation on Dataset 2.

While the proposed framework demonstrates robust performance under complex PRI modulations, missing pulses, and unseen radar modes, realistic electromagnetic environments may also feature extreme spurious pulses and severe measurement errors. Although the fine-grained clustering and adaptive calibration module inherently provide a degree of noise robustness, a systematic evaluation under these extreme nonideal conditions remains a limitation of the current study and will be a primary focus of our future research.

B. Benefits of Using Virtual PRIs

To quantify the benefit of introducing the virtual PRI for RSS, we conduct an ablation-style comparison on Dataset 2 by substituting the virtual PRI with different PRI alternatives and evaluating the corresponding RSS accuracies, as shown in Fig. 7. The alternatives include: without PRI, ground-truth PRI, and the estimated PRIs produced by the PRI estimation algorithms embedded in cumulative difference histogram (CDIF) [8], sequential difference histogram (SDIF) [9], and PRI transform [10]. The PRI estimation algorithm embedded in CDIF estimates the PRI by constructing a CDIF from multilevel TOA differences and identifying its highest peak as the estimated PRI. The PRI estimation algorithm embedded in SDIF estimates the PRI by constructing an SDIF from single-level TOA differences and identifying its highest peak as the estimated PRI. The PRI estimation algorithm embedded in the PRI transform estimates the PRI by constructing a complex-weighted histogram from TOA differences and identifying its highest peak as the estimated PRI. Fig. 7 (using $K = 100$ clustering) shows that all PRI methods surpass non-PRI baselines by providing supplementary discrimination. Virtual PRI achieves accuracy closest to ground truth due to robustness against modulation-induced PRI variations. Traditional methods (SDIF/CDIF/PRI transform) fail with complex modulations due to fixed PRI assumptions, while virtual PRI—extracted via fine-grained clustering without direct estimation—mitigates modulation-induced distortions, demonstrating superior RSS adaptability.

C. Effect of Fine-Grained Clustering With Different Numbers of Clusters

To evaluate the impact of different numbers of clusters K on the RSS accuracy, we design an experiment on Dataset 2 using the dual-path network combined with the K-means algorithm, where K is selected from $[12, 24, 50, 100, 150]$. The clustering accuracy and sorting accuracy with different values of K for the three test scenarios of Table IV are presented in Fig. 8. The clustering accuracy evaluates the effectiveness of the K-means clustering in the virtual PRI extraction step, and the sorting accuracy represents the final sorting performance of our virtual PRI-assisted RSS scheme.

Fig. 8 shows clustering and sorting accuracy increases with K from 12 to 100 then saturates, justifying $K = 100$ for virtual PRI extraction. Larger K better isolates radar modes

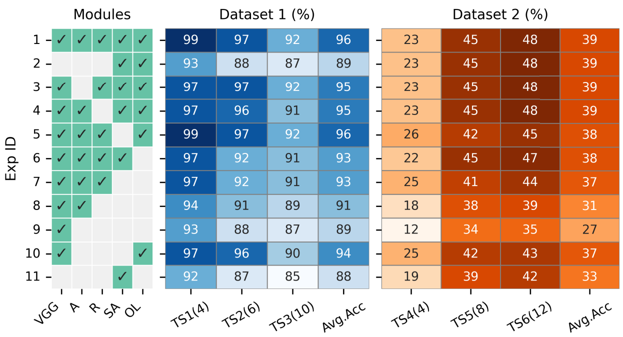


Fig. 9. RSS accuracy for ablation experiments on two datasets. In each subfigure, darker colors represent higher values of the corresponding metrics.

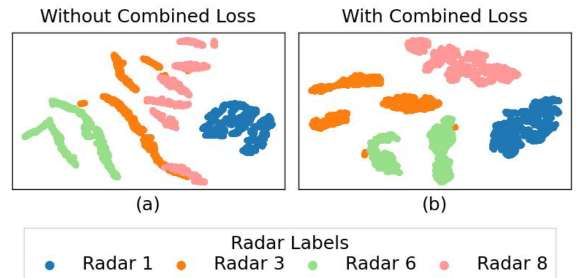


Fig. 10. t-distributed stochastic neighbor embedding (t-SNE) visualization of the AFVs for 4 radars in Dataset 2. By introducing the combined optimization loss, the learned features in (b), i.e., expID 10, form more compact and better-separated clusters when compared to the learned features in (a), i.e., expID 9.

into distinct clusters, enabling virtual PRIs to more accurately represent actual PRI modulations. Beyond a sufficient K, performance saturates as cluster separation cannot further improve pattern representation, halting accuracy gains.

It is worth discussing that while adaptive cluster number determination strategies (e.g., silhouette coefficient or gap statistic) are popular in traditional machine learning, they are not employed in our current preprocessing stage. This is because our scheme explicitly requires intentional fine-grained oversegmentation ($K \gg M$) rather than finding the true number of emitters. Furthermore, for dense radar pulse streams, computing these adaptive metrics across all data points iteratively introduces nonnegligible computational latency, making it difficult to meet the strict real-time processing constraints required in practical electronic reconnaissance. However, exploring lightweight adaptive mechanisms remains a valuable direction for future applicability enhancements.

D. Ablation Analysis of Dual-Path Network

We conduct ablation experiments to evaluate module contributions in the proposed dual-path network, which comprises five key components: "VGG" (base framework for the first path), "A" (adaptive attention module added to VGG), "R" (residual structure added to VGG), "SA" (second path with self-attention and fully connected layers), and "OL" [optimization loss function (15)]. Eleven experimental configurations combining these modules are evaluated on both datasets. Fig. 9 presents the results, where expID denotes the experiment identifier, TS indicates the test scenario, and Avg represents the average accuracy across all test scenarios per dataset.

1) Effectiveness of Adaptive Attention Module: To study the adaptive attention module's effectiveness, we compare experiments expID 1 (full modules) and expID 3 (excluding "A"). Fig. 9 shows higher average accuracy with the "A" component, demonstrating its benefit in the dual-path network's first path for improved feature selection. By emphasizing critical channel/spatial information while suppressing irrelevant features, the module yields more representative features that enhance sorting accuracy.

2) Effectiveness of Residual Structure: Comparing experiments expID 1 (full modules) and expID 4 (excluding "R") demonstrates the residual structure's effectiveness in VGG networks. Fig. 9 shows higher average accuracy with "R" component, verifying its superiority. By preserving critical features through identity mapping across layers, the structure minimizes information loss and transformation degradation, enabling distortion-free learning of intricate features for robust RSS-related feature extraction and accuracy improvement.

3) Effectiveness of Introducing Second Path: Comparing expID 5 (first-path only), expID 2 (second-path only), and expID 1 (dual-path) demonstrates the second-path effectiveness. Fig. 9 shows the first path outperforms the second by 7% on Dataset 1 but underperforms by 1% on Dataset 2, while dual-path achieves the highest accuracy on both datasets. This verifies the second-path contribution to overall performance, as its extracted local features capture pulse-wise PDW patterns that identify pulse similarities, overcoming feature distribution shifts from operating mode variations in Dataset 2.

4) Effectiveness of Combined Loss Function: To validate the combined loss function (15), three experiments compare models with/without "OL": 1) first path (expID 10 versus 9); 2) second path (expID 2 versus 11); and 3) dual-path (expID 1 versus 6). Fig. 9 consistently shows higher accuracy with "OL," e.g., expID 1 gains 3%/1% over expID 6 on both datasets. This stems from "OL" clustering the same radar AFVs while separating different radar AFVs, yielding compact distributions that clarify decision boundaries. Enhanced noise robustness and uncertainty tolerance further boost stability. t-SNE visualizations in Fig. 10 confirm concentrated radar-specific features under (15), directly improving RSS performance.

E. Computational Complexity and Runtime Analysis

To evaluate practical engineering feasibility, the computational complexity and runtime efficiency of the proposed Dual-Path-K scheme are assessed against benchmarks in Table VI. The evaluation metrics include the number of trainable parameters (Params), the number of floating-point operations (FLOPs) required for a single forward pass of the neural network, the average inference latency (Latency) required to process each pulse, and the GPU memory (Memory) consumption during inference. It should be noted that since SVM and LPACD are traditional, nonneural network-based methods, we only report their Latency. While lightweight models (e.g.,

TABLE VI
COMPUTATIONAL COMPLEXITY AND RUNTIME ANALYSIS OF DIFFERENT RSS METHODS

	Params ($\times 10^3$)	FLOPs ($\times 10^3$)	Latency (ms/pulse)	Memory (MB)
FCNN	3.114	3.229	0.0025	9.22
ABMR	1203.21	8350.86	0.0046	22.46
SVM	-	-	0.121	-
ResGCN-RSS	2.63	7014.4	0.0115	34.88
LSTM-FCNN	2109.51	2129.28	0.027	120.1
RNN-FCNN	2065.3	2082.56	0.027	119.1
LPACD	-	-	42.4763	-
BLSTM	613.61	618592	0.1985	1593.55
Dual-Path-K	597.15	17140	0.6357	1384.84

FCNN) exhibit extremely low latency, their simple structures severely limit sorting accuracy under complex feature shifts; conversely, heavy recurrent models (e.g., ABMR) introduce excessive parameter overhead. Striking an optimal balance, the proposed method maintains a moderate parameter scale (597.15×10^3) and when evaluated on an NVIDIA RTX 4090 GPU, achieves an average inference latency of 0.6357 ms per pulse alongside a memory consumption of 1384.84 MB. This submillisecond latency is substantially faster than traditional algorithmic approaches like LPACD (42.47 ms per pulse) and fully satisfies the real-time signal processing constraints required in practical electronic reconnaissance. Therefore, the minor increase in computational overhead is a highly worthwhile trade-off for the significant improvements in sorting accuracy, making the proposed scheme highly feasible for actual hardware deployment.

VII. CONCLUSION

In this article, we have introduced the concept of virtual PRIs to extract PRI-related information for effective RSS. To ensure the tight connection between virtual PRI measurements and the ground-truth PRIs, we have used a fine-grained PDW-based clustering scheme with the number of clusters far larger than the number of categories. This has ensured that pulses within the same cluster are highly likely to originate from the same operating mode of a specific radar, allowing the virtual PRI to serve as a surrogate for the ground-truth PRI in signal sorting. We then design a dual-path network to extract RSS-related information from PDWs and virtual PRIs and to address the overclustering issue by assigning an emitter label to each fine-grained cluster and fusing clusters with the same label for final RSS. The effectiveness of our virtual PRI-assisted RSS scheme has been validated through extensive experiments. Results show that our scheme facilitates robust RSS under challenging conditions with complex PRI modulations and missing pulses. Furthermore, the dual-path network can maintain a certain degree of RSS capability even when presented with unseen radar modes, by adaptively leveraging the global and local features extracted from the dual paths.

REFERENCES

[1] G. Xu, "Research on clustering-based radar pulse stream sorting," M.S. thesis, School Inf. Commun. Eng., Univ. Electron. Sci. Technol. China, Chengdu, China, 2024.

[2] S. Q. Wang, G. P. Hu, Q. L. Zhang, C. Y. Gao, and T. Cai, "The background and significance of radar signal sorting research in modern warfare," *Proc. Comput. Sci.*, vol. 154, pp. 519–523, Apr. 2019.

[3] X. Wang, Y. Zhao, and Z. Huang, "A survey of deep transfer learning in automatic modulation classification," *IEEE Trans. Cognit. Commun. Netw.*, vol. 11, no. 3, pp. 1357–1381, Jun. 2025.

[4] W. Huo et al., "A rapid radar signal sorting method based on density and Gaussian mixture model data stream clustering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 61, no. 4, pp. 1–16, Aug. 2025.

[5] H. Wang, Z. Tao, J. He, and T. Shu, "Deinterleaving of intercepted radar pulse streams via temporal convolutional attention network," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 61, no. 4, pp. 9327–9343, Aug. 2025.

[6] L. Lesieur, J.-M. Le Caillec, A. Khenchaf, V. Guardia, and A. Toumi, "An overview and classification of machine learning approaches for radar signal deinterleaving," *IEEE Access*, vol. 13, pp. 28008–28028, 2025.

[7] S. Wang, D. Song, G. Min, and B. Peng, "Pulse sequence segmentation for multifunction radar using multiple reconnaissance receivers," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 61, no. 3, pp. 6873–6884, Jun. 2025.

[8] H. K. Mardia, "New techniques for the deinterleaving of repetitive sequences," *IEE Proc. F Radar Signal Process.*, vol. 136, no. 4, pp. 149–154, 1989.

[9] D. J. Milojević and B. M. Popović, "Improved algorithm for the deinterleaving of radar pulses," *IEE Proc. F Radar Signal Process.*, vol. 139, no. 1, pp. 98–104, 1992.

[10] D. J. Nelson, "Special purpose correlation functions for improved signal detection and parameter estimation," in *Proc. IEEE Int. Conf. Acoust.*, Aug. 1993, pp. 73–76.

[11] K. Nishiguchi, "A new method for estimation of pulse repetition intervals," in *Proc. Nat. Conv. Rec. IECE Jpn.*, 1983, p. 1.

[12] K. Nishiguchi and M. Kobayashi, "Improved algorithm for estimating pulse repetition intervals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 407–421, Apr. 2000.

[13] Z. Xiao and Z. Yan, "Radar emitter identification based on naive Bayesian algorithm," in *Proc. IEEE 5th Inf. Technol. Mechatronics Eng. Conf. (ITOEC)*, Jun. 2020, pp. 1242–1244.

[14] C.-S. Shieh and C.-T. Lin, "A vector neural network for emitter identification," *IEEE Trans. Antennas Propag.*, vol. 50, no. 8, pp. 1120–1127, Aug. 2002.

[15] H. Li, W. Jing, and Y. Bai, "Radar emitter recognition based on deep learning architecture," in *Proc. CIE Int. Conf. Radar (RADAR)*, Oct. 2016, pp. 1–5.

[16] X. Li, Z. Liu, Z. Huang, and W. Liu, "Radar emitter classification with attention-based multi-RNNs," *IEEE Commun. Lett.*, vol. 24, no. 9, pp. 2000–2004, Sep. 2020.

[17] L. Cain, J. Clark, E. Pauls, B. Ausdenmoore, R. Clouse, and T. Josue, "Convolutional neural networks for radar emitter classification," in *Proc. IEEE 8th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2018, pp. 79–83.

[18] X. Li, Z. Huang, F. Wang, X. Wang, and T. Liu, "Toward convolutional neural networks on pulse repetition interval modulation recognition," *IEEE Commun. Lett.*, vol. 22, no. 11, pp. 2286–2289, Nov. 2018.

[19] P. Lang et al., "A comprehensive survey of machine learning applied to radar signal processing," 2020, *arXiv:2009.13702*.

[20] S. Jinping, L. Zhen, L. Li, and L. Xiang, "Progress in radar emitter signal deinterleaving," *J. Radars*, vol. 11, no. 3, pp. 418–433, 2022.

[21] P. Lang, X. Fu, J. Dong, H. Yang, and J. Yang, "A novel radar signals sorting method via residual graph convolutional network," *IEEE Signal Process. Lett.*, vol. 30, pp. 753–757, 2023.

[22] Z. Zhou, X. Fu, J. Dong, and M. Gao, "Radar signal sorting with multiple self-attention coupling mechanism based transformer network," *IEEE Signal Process. Lett.*, vol. 31, pp. 1765–1769, 2024.

[23] T. Chen, B. Yang, and L. Guo, "Radar pulse stream clustering based on MaskRCNN instance segmentation network," *IEEE Signal Process. Lett.*, vol. 30, pp. 1022–1026, 2023.

[24] H. Xiang, F. Shen, and J. Zhao, "Deep ToA mask-based recursive radar pulse deinterleaving," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 2, pp. 989–1006, Apr. 2023.

[25] W. Chao, S. Liting, L. Zhangmeng, and H. Zhitao, "A radar signal deinterleaving method based on semantic segmentation with neural network," *IEEE Trans. Signal Process.*, vol. 70, pp. 5806–5821, 2022.

[26] K. Chi, J. Shen, Y. Li, Y. Li, and S. Wang, "Multi-function radar signal sorting based on complex network," *IEEE Signal Process. Lett.*, vol. 28, pp. 91–95, 2021.

[27] J. Zhang, M. Sun, and J. Wang, "Multi-function radar signal sorting based on complex networks and cslpa," *J. Detection Control*, vol. 45, no. 4, pp. 87–94, 2023.

[28] S. Li, J. Hu, and B. Yan, "Multi-function radar signal sorting based on improved complex network community detection," *Telecommun. Eng.*, vol. 65, pp. 1851–1858, Nov. 2025.

[29] M. A. Nuhoglu, Y. K. Alp, M. E. C. Ulusoy, and H. A. Cirpan, "Image segmentation for radar signal deinterleaving using deep learning," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 1, pp. 541–554, Feb. 2023.

[30] Z.-M. Liu and P. S. Yu, "Classification, denoising, and deinterleaving of pulse streams with recurrent neural networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 4, pp. 1624–1639, Aug. 2019.

[31] Y. Luo, W. Zhang, and W. Tian, "Radar signal sorting and individual recognition based on LSTM," in *Proc. IEEE 6th Adv. Inf. Manage., Communicates, Electron. Autom. Control Conf. (IMCEC)*, May 2024, pp. 1206–1209.

[32] X. Wang, X. Chen, Y. Zhou, Y. Chen, B. Xiao, and H. Wang, "A radar signal sorting algorithm based on improved dbscan algorithm," *J. Air Force Eng. Univ.*, vol. 22, no. 3, pp. 47–54, 2021.

[33] D. K. Barton, *Radar System Analysis and Modeling*. Norwood, MA, USA: Artech House, 2004.

[34] W. Jiang, X. Fu, and J. Chang, "Improved de-interleaving algorithm of radar pulses based on dual fuzzy vigilance ART," *J. Syst. Eng. Electron.*, vol. 31, no. 2, pp. 303–311, Apr. 2020.

[35] M. Zhu, Y. Li, and S. Wang, "Model-based time series clustering and interpulse modulation parameter estimation of multifunction radar pulse sequences," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 6, pp. 3673–3690, Dec. 2021.

[36] P. Lang, X. Fu, Z. Cui, C. Feng, and J. Chang, "Subspace decomposition based adaptive density peak clustering for radar signals sorting," *IEEE Signal Process. Lett.*, vol. 29, pp. 424–428, 2022.

[37] M. Wan, Y. Zhang, Y. Bai, Y. Sun, Q. Yu, and Q. Wang, "A real-time radar signal sorting method under Bayesian framework with dynamic cluster merging," *IEEE Sensors J.*, vol. 24, no. 17, pp. 27859–27869, Sep. 2024.

[38] X. Feng, X. Hu, and Y. Liu, "Radar signal sorting algorithm of K-means clustering based on data field," in *Proc. 3rd IEEE Int. Conf. Comput. Commun. (ICCC)*, Dec. 2017, pp. 2262–2266.

[39] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DBSCAN revisited, revisited: Why and how you should (Still) use DBSCAN," *ACM Trans. Database Syst.*, vol. 42, no. 3, pp. 1–21, Jul. 2017, doi: 10.1145/3068335.

[40] S. Li, F. Liu, L. Jiao, P. Chen, and L. Li, "Self-supervised self-organizing clustering network: A novel unsupervised representation learning method," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 1857–1871, Feb. 2024.

[41] D. Zhang and K. Luo, "Clustering algorithm based on artificial bee colony optimization," in *Proc. Adv. Eng. Res.*, vol. 12, 2015, pp. 126–131.

[42] X. Wang, X. Fu, J. Dong, and J. Jiang, "Dynamic modified chaotic particle swarm optimization for radar signal sorting," *IEEE Access*, vol. 9, pp. 88452–88466, 2021.

[43] D. van der Merwe and A. Engelbrecht, "Data clustering using particle swarm optimization," *Proc. Congr. Evol. Comput.*, vol. 1, pp. 215–220, Jun. 2003.

[44] Z. Liang, M. Tao, L. Wang, J. Su, and X. Yang, "Automatic modulation recognition based on adaptive attention mechanism and ResNeXt WSL model," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 2953–2957, Sep. 2021.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Jun. 2016, pp. 770–778.

[46] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017, pp. 6000–6010.

[47] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 4580–4584.

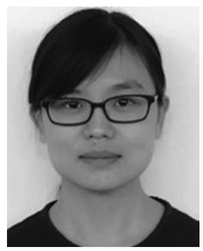
[48] Z. Yuan and R. Yuan, "Application of SVM classifier design in radar signal sorting," *Electron. Inf. Warfare Technol.*, vol. 26, no. 6, pp. 139–142, 2018.

Huajie Wu received the B.S. degree from China Agricultural University, Beijing, China, in 2023. He is currently pursuing the Ph.D. degree with the School of Information and Electronics, Beijing Institute of Technology (BIT), Beijing.

His current research interests include radar signal deinterleaving, intelligent signal processing, and artificial intelligence.

Xingyun Chen received the B.S. and M.S. degrees from Dalian University of Technology, Dalian, China, in 2020 and 2023, respectively. He is currently pursuing the Ph.D. degree with the School of Cyberspace Science and Technology, Beijing Institute of Technology (BIT), Beijing, China.

His current research interests include anti-jamming communications and deep reinforcement learning.



Ying Ma (Member, IEEE) received the Ph.D. degree from the Department of ECE, University of Florida, Gainesville, FL, USA, in 2021.

From July 2015 to July 2016, she visited the Department of ECE, University of Southampton, Southampton, U.K. She was an Assistant Professor at the Department of ECE, University of Central Florida, Orlando, FL, USA, from 2020 to 2022. She is currently an Assistant Professor with Beijing Institute of Technology (BIT), Beijing, China. Her current research interests include deep learning and signal processing.

Haichuan Ding (Member, IEEE) received the B.Eng. and M.S. degrees in electrical engineering from Beijing Institute of Technology (BIT), Beijing, China, in 2011 and 2014, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2018.

From 2012 to 2014, he was a Visiting Student with the Department of Electrical and Computer Engineering, University of Macau, Macau, China. He is currently a Professor with the School of Cyberspace Science and Technology, BIT. His current research interests include cognitive radio, intelligent signal processing, and physical layer security.

Sheng Chen (Life Fellow, IEEE) received the B.Eng. degree in control engineering from East China Petroleum Institute, Dongying, China, in 1982, the Ph.D. degree in control engineering from the City University of London, London, U.K., in 1986, and the D.Sc. degree from the University of Southampton, Southampton, U.K., in 2005.

From 1986 to 1999, he held research and academic appointments at The University of Sheffield, Sheffield, U.K.; The University of Edinburgh, Edinburgh, U.K.; and the University of Portsmouth, Portsmouth, U.K. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, where he holds the post of Professor of Intelligent Systems and Signal Processing. In 2023, he was appointed as a Distinguished Professor with the Faculty of Information Science and Engineering, Ocean University of China, Qingdao, China. He has published more than 700 research articles. He has more than 22 000 Web of Science citations with an H-index of 65 and more than 42 000 Google Scholar citations with an H-index of 87. His research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural networks, and machine learning.