

Collaborative Learning based Spectrum Sensing under Partial Observations

Weishan Zhang, *Student Member, IEEE*, Yue Wang, *Senior Member, IEEE*, Xiang Chen, *Member, IEEE*, Lingjia Liu, *Senior Member, IEEE*, and Zhi Tian, *Fellow, IEEE*

Abstract—To deal with the complex wireless cognitive radios, data-driven learning technologies have been advocated for spectrum sensing. While the existing learning-based methods are designed for basic single-band circumstances, they may not work well in practical wideband regimes. Due to the limited sensing capability and hardware constraints of practical secondary users (SUs) devices, individual SUs can only collect limited training data to observe a narrowband part of the entire wideband spectrum pool. It is known as the issue of partial observations, which leads to a heterogeneous multi-task learning problem. To overcome these challenges, this work proposes a novel framework of cooperative spectrum sensing via collaborative learning among distributed SUs. Capitalizing on the hierarchical nature of neurons of deep neural networks (DNN) in heterogeneous feature extraction, we propose a novel multi-task DNN architecture to detect wideband spectrum occupancy accurately and efficiently. By decoupling the large multi-band DNN into smaller band-specific sub-networks, these sub-networks can be jointly trained among distributed SUs even with heterogeneous local data. Simulation results indicate that our proposed method outperforms existing benchmarks in small-data regimes by achieving higher learning accuracy with less model complexity and computational cost.

Index Terms—Cognitive radio, spectrum sensing, collaborative learning, partial observation, deep neural network, decoupling.

I. INTRODUCTION

Cognitive Radio (CR) has been widely recognized as an enabling technology for spectrum-efficient wireless communications [1]. Spectrum sensing, as the key technique in CR, aims to accurately and efficiently detect the spectrum opportunities in terms of idle spectrum resources, and has been intensively studied in the CR literature [2]–[5]. While traditional spectrum sensing methods, e.g., energy detection, matched filtering, and cyclostationary detection work well in the ideal settings with perfectly known signal and channel models, they unfortunately

This work was supported in part by the National Science Foundation grants #1939553, #2003211, #2128594, #2128596, #2231209 and #2413622, and the Virginia Research Investment CCI grant #223996. Part of this work was presented at the IEEE International Conference on Communications Workshop, Seoul, South Korea, May 2022.

Weishan Zhang is with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA 22030, USA (e-mail: wzhang23@gmu.edu).

Yue Wang is with the Department of Computer Science, Georgia State University, Atlanta, GA 30303, USA (e-mail: ywang182@gsu.edu).

Xiang Chen is with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA 22030, USA (e-mail: xchen26@gmu.edu).

Lingjia Liu is with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061, USA (e-mail: ljliu@vt.edu).

Zhi Tian is with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA 22030, USA (e-mail: ztian1@gmu.edu).

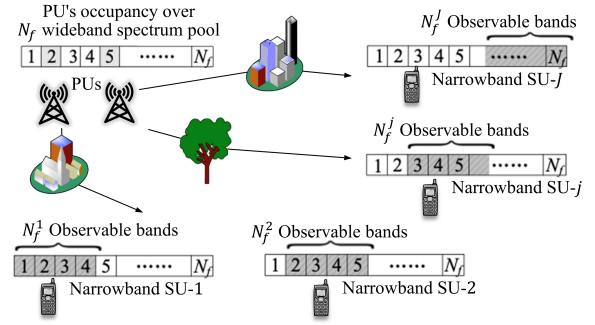


Fig. 1. Cooperative spectrum sensing under partial observations.

become vulnerable to model mismatch issues such as channel uncertainty and/or noise uncertainty [6]–[9].

To cope with such limitations of the traditional model-based techniques, deep learning (DL) based methods are recently advocated for single-band spectrum sensing [10], [10]–[17] and the multi-band case [18], [19]. Such methods utilize the strong capability of DL in learning the underlying representation of complicated models from training data in physical layer communications [20]. Despite the success of these pioneer works in homogeneous settings and idealized observation conditions, the goal in realistic CR practice is to find out as many as spectrum opportunities from wideband spectrum under the constraint of locally partial observations [21]. As shown in Fig. 1, given the limited sensing capability and hardware constraints of individual secondary users (SUs), each SU can only collect limited data to observe a small narrowband part of the entire wideband spectrum pool. As a result, the small dataset collected by each SU merely reflects a few locally observable bands, which means that the standalone DL-based methods are incapable for wideband spectrum sensing under partial observations. To build collaboration among distributed SUs, straightforward learning-based cooperative spectrum sensing methods have been proposed via collecting the measurements from all SUs and aggregating the data at a data center to make a one-shot decision on spectrum occupancy [16], [22], [23]. Unfortunately, such data aggregations lead to high communication overhead, huge computation overload, and unwilling measurement data privacy exposure, which are all unacceptable for practical CR networks.

To overcome the aforementioned drawbacks of the existing methods while utilizing the cooperation benefits among distributed SUs, this paper aims to develop novel deep neural network (DNN) based cooperative spectrum sensing framework and training techniques. Through a holistic integration of the design of low-cost DNN architecture with the development

of efficient DL solutions, we develop effective collaboration among distributed SUs under partial observations. To the best of our knowledge, this has not been studied in the literature of DL-based wideband spectrum sensing regimes. Over the years, parameter server (PS) [24] and federated learning (FL) [25] have been proposed as two main techniques for the implementation of collaborative learning. However, neither PS nor FL can be directly applied for CR wideband spectrum sensing under partial observations. This is because they both hinge on a prerequisite that all local participants share a homogeneous DNN structure for a common learning task. But, spectrum data collected through partial observations are heterogeneous at different SUs, which leads to a more challenging multi-task learning problem.

To achieve accurate and efficient wideband spectrum sensing under partial observations, we propose a novel DNN-based cooperative spectrum sensing framework via collaborative DL across distributed SUs whose local training data observe partially overlapping bands. Considering the data characteristics in wideband spectrum pools and the demand to detect each band distinctively, we design a multi-task DNN for multi-band spectrum sensing, where a flexible DNN decoupling and reconfiguration scheme enables efficient collaboration among SUs with heterogeneous data and sensing tasks. The proposed multi-task learning DNN can dramatically save the computation and storage costs thanks to the smaller model size that is locally trained at individual SUs than the global model. Meanwhile, the proposed collaborative training method enhances the performance of band-wise spectrum sensing with small datasets on SUs under partial observations. To the best of our knowledge, this is the first work using collaborative learning for cooperative spectrum sensing under partial observations. The contributions of this paper are summarized below.

- For the first time, we formulate the problem of learning-based wideband spectrum sensing with distributed users under partial observations. To address the major challenges of heterogeneity of both learning tasks and training data, we propose a collaborative multi-task learning framework to accommodate the heterogeneous DNNs on different users. The key idea is to efficiently utilize a critical connection between the neuron activation preference in deep hidden layers of DNNs and the band-specific spectrum occupancy characteristics in wideband regimes.
- Exploiting the hierarchical nature of DNN neurons in feature extraction of CR wideband spectrum, we reconfigure the original dense DNN structure into multiple learning paths, by reducing unnecessary connections between the data flows for detecting different bands. This enables the band-wise parameter sharing among the multi-task sensing models to fit the heterogeneously observable bands at individual SUs, which is otherwise unavailable in existing standard DNN methods.
- In addition, we develop an efficient training process that is customized for the implementation of the proposed collaborative learning based cooperative spectrum sensing. It not only jointly optimizes the heterogeneous DNNs among all distributed SUs with small training datasets,

but also quickly obtains the global detection of channel occupancy over entire wideband spectrum pool.

- In view of the scarcity of available training data of wideband spectrum measurements, we also present a data generation approach to produce synthetic wideband spectrum datasets under partial observations for various modulation types and channel conditions.

Simulations results show that our proposed method outperforms the benchmarks of existing DL-based methods in terms of better spectrum sensing performance with small datasets under partial observations, more robustness to noise uncertainty and higher computation efficiency thanks to less DNN parameters to train at individual SUs.

II. RELATED WORKS

To improve spectrum sensing performance in complex wireless environments, DNNs have been introduced to CR systems in detecting spectrum occupancy from the signals of interest based on their temporal and spectrum features.

A. Deep learning for single-band spectrum sensing

Recently, various DNN-based methods have been proposed for single-band spectrum sensing by learning the underlying features of primary signals. In the basic single-SU case, DL-based algorithms are proposed for occupancy detection [14] and modulation recognition [26]. Among them, Xie *et al.* design a CNN-LSTM model that features a cascade structure containing a CNN for the correlation feature extraction and then an LSTM to estimate PU activity pattern [10]. Xing *et al.* develop a composite DNN which incorporates convolutional, LSTM, and self-attention to extract local features and global correlations from the time series data [11]. Mei *et al.* design a learning-based sparse signal reconstruction method by leveraging deep unfolding and CNN, to conduct wideband spectrum sensing with sub-Nyquist sampled data [12]. Chae *et al.* develop a CNN-based spectrum sensing model with multiple antennas, to enhance sensing performance by using cross-correlation features between multiple antennas of a receiver [13]. All these learning methods only concern centralized implementation.

To overcome the sensing performance degradation due to wireless fading and shadowing impacts, DNN techniques have been integrated into cooperative spectrum sensing recently in the simple single-band scenarios [10]–[17]. Lee *et al.* develop a deep cooperative sensing framework with learnable decision fusion, which is a CNN-based cooperative sensing method [16]. Janu *et al.* design a LSTM-based cooperative spectrum sensing method to learn the PU activity pattern hierarchically at both the SU and group levels [27]. While these DNN-based cooperative spectrum sensing techniques improve the sensing accuracy in the presence of model uncertainties by learning the latent features of the primary signal, most of them are designed for single-band signal or single detection objective. Further, exiting DNN-based cooperative spectrum sensing methods involve data fusion, which means SUs have to share sampled spectrum. As a result, undesired privacy leakage and communication consumption inevitably happen

in practical applications. In summary, the existing works on learning-based methods for single-band scenarios limit their efficiency and quality in wideband CR environments.

B. Deep learning for multi-band spectrum sensing

Recently, a few research efforts have been found to design DNN models for single-user multi-band spectrum sensing. Zhang *et al.* solve the sub-band occupancy pattern identification problem in a multi-band environment which can accommodate a few occupancy statuses [18]. Ambika *et al.* propose another DNN model to detect occupancy statuses from a dataset that contains the spectrograms of multiple channels [19]. These multi-band occupancy detectors are designed based on classifiers to differentiate all possible occupancy statuses. Specifically, in the design of their output layers, softmax and cross-entropy loss function are adopted, which is however not suitable for band-wise spectrum sensing problems in wideband scenarios. This is because the problem size for classification goes exponential in the number of bands. As a result, these existing works can only identify limited categories of all the possible occupancy statuses. Moreover, these methods rely on large training datasets to learn spectrum features, which is impractical due to the dynamic nature of wireless environments.

C. Data parallel distributed learning

As a major branch of distributed learning, data parallel methods leverage the parallel computation power at distributed nodes of a multi-agent system to learn general knowledge about a global task from distributed data [28]. Along this line, PS [24] and FL [25] have been proposed to implement distributed training. While these training methods result in high communication efficiency and privacy provisioning, they usually hinge on an assumption that all agents share a homogeneous DNN structure and independent and identically distributed (IID) data to achieve a common learning task. This requires an idealized CR scenario where all SUs need to have the global observation over the entire wideband spectrum pool. However, this is not the case for real CR sensing systems. Under the circumstance of partial observations, local spectrum data and corresponding local model structure become heterogeneous. To the best of our knowledge, there is no existing work on collaborative learning based wideband spectrum sensing.

III. PROBLEM STATEMENT

A. Signal model

In this section, we first formulate the signal model of the power spectrum density (PSD) based measurement data collected by distributed SUs under partial observations. Suppose a wideband spectrum pool is uniformly divided into N_f bands as in Fig. 1, where each band is assigned to a certain primary user (PU) who may occupy the band or not. For a CR system with J SUs, the time series sampled at the SU- j , $\forall j \in [1, J]$, are regarded as measurements of received signals in the time domain. Each sample is given by a summation of

every primary signal sequence reaching SU- j plus the noise vector $\mathbf{w}^j$:

$$\mathbf{x}^j = \sum_{n=1}^{N_f} y_n h_n^j \mathbf{x}_n + \mathbf{w}^j, \quad (1)$$

where $\mathbf{x}_n$ is the modulation signal transmitted on band- n if it is occupied by a certain PU, y_n is the n -th entry of a Boolean vector $\mathbf{y} = [y_1, y_2, \dots, y_{N_f}] \in \mathbb{B}^{N_f}$ that stores the ground-truth occupancy conditions of all bands with values equal to either 1 (occupied) or 0 (unoccupied), coefficient h_n^j denotes the channel gain of band- n between PU and SU- j . In the channel model, we consider the impacts of path-loss and shadowing. The channel gain is formulated as $h_n^j = \sqrt{\beta \left(\frac{d_0}{d_n^j}\right)^\alpha} 10^{\frac{-\psi_n^j}{10}} [29]$, where β is a constant related to the antenna characteristics and average attenuation, α is the path-loss exponent, d_n^j is the distance between SU- j and the PU on band- n , d_0 is the reference distance, and ψ_n^j is a Gaussian-distributed random variable with mean zero and variance $\sigma_{\psi_n^j}^2$ that measures the shadow fading of the channel over band- n between PU and SU- j .

To extract the spectrum features in the frequency domain for learning process, we observe $\mathbf{x}^j$ over multiple snapshots and calculate their second-order statistics in terms of their autocorrelation. Then, by applying Fourier transform on autocorrelation, the wideband PSD at SU- j is obtained as:

$$\mathbf{s}^j = \text{FT}(\text{Corr}(\mathbf{x}^j)) \in \mathcal{N}^{N_w N_f \times 1}, \quad (2)$$

where FT denotes the Fourier transform, $\text{Corr}(\cdot)$ computes the signal autocorrelation, and N_w is the spectrum resolution of each band. Considering that all bands share a common bandwidth, the wideband PSD $\mathbf{s}^j$ in (2) can be uniformly segmented into N_f band-specific PSD vectors as $\mathbf{s}^j = [\mathbf{s}_1^j; \dots; \mathbf{s}_n^j; \dots; \mathbf{s}_{N_f}^j]$, where $\mathbf{s}_n^j \in \mathcal{N}^{N_w \times 1}$, $n = 1, 2, \dots, N_f$. Due to power leakage [30], channel aggregation [31], and other factors, there exist inherent correlations of primary signals between different bands in realistic wideband environments. To unveil and capture such inter-band correlations via deep learning-based methods, we rewrite $\mathbf{s}^j$ into an $N_w \times N_f$ matrix as the input training data:

$$\mathbf{S}^j = \text{vec}^{-1}(\mathbf{s}^j) = [\mathbf{s}_1^j, \dots, \mathbf{s}_{N_f}^j], \quad (3)$$

where $\text{vec}^{-1}(\cdot)$ is the inverse vectorization operation.

B. Deep learning-based wideband spectrum sensing

Now, we discuss how spectrum sensing is formulated into a deep learning problem. Given the input training data (3), the most basic CR spectrum sensing problem at SU- j for multi-band settings can be expressed as a binary hypothesis testing problem in $\mathbf{H}_0$ when band- n is vacant; or in $\mathbf{H}_1$ when band- n is occupied. For a basic single-band model with input data $\mathbf{s}_n$, output label $\{0, 1\}$ and DNN parameter set $\mathcal{W}$, the spectrum sensing problem can be expressed as a function, $f(\mathbf{s}_n, \mathcal{W}) \in [0, 1]$. In this sense, the task of deep learning-based spectrum sensing at a specific SU- j for a single band n is to find the optimal parameter set $\mathcal{W}^*$ that generates the correct hypothesis based on the received band-specific PSD $\mathbf{s}_n^j$:

$$\begin{cases} f(s_n^j | \mathbf{H}_1, \mathcal{W}^*) \geq 0.5; \\ f(s_n^j | \mathbf{H}_0, \mathcal{W}^*) < 0.5. \end{cases} \quad (4)$$

Thanks to the representation power of DNN even in the lack of expert knowledge of underlying signal and channel models, the learning-based single-band detectors can be automatically trained with sufficient labeled data. The objective of training can be formulated as:

$$\min_{\mathcal{W}} \sum_{\{s_n^j, y_n\} \in \mathcal{D}} \text{Loss}^b(f(s_n^j, \mathcal{W}), y_n) \quad (5)$$

where $\mathcal{D}$ is the dataset including the PSD and the labeled occupancy of the target single band, the ground truth occupancy $y_n = \{0, 1\}$ is used as the label, and Loss^b denotes the binary cross-entropy loss function of two probability-based confidence values defined as [32]

$$\text{Loss}^b(p, q) = p \log q + (1 - p) \log (1 - q). \quad (6)$$

When extending the single-band case to the multi-band case, Eq. (6) plays a key difference from the regular cross-entropy loss used in the existing classifiers [18], [19]. Specifically, multi-band sensing models are developed to distinguish one category out of a total number of $N_c = K^{N_f}$ classes where K denotes the number of occupancy status for each band. However, in order to encode all N_c occupancy categories of N_f bands, the size of softmax output layers for classifiers grows exponentially. In fact, the sensing technique in [18], [19] only handles limited occupancy conditions, which leaves most situations unattended. For high computational efficiency, the sensing model is expected to detect all possible occupancy conditions of the spectrum pool with the least number of output channels. To this end, we next design a novel multi-class predictor based DNN structure with smaller model size and reduced computational complexity.

IV. COOPERATIVE SPECTRUM SENSING

In this section, we aim to develop a novel band-wise cooperative spectrum sensing framework, which utilizes collaborative training of multi-band DNNs among distributed SUs. It is designed to detect a wideband spectrum pool that turns to be heterogeneous over partially observable bands at different SUs.

A. Band-specific DNN structure reconfiguration

To implement collaborative learning for wideband spectrum sensing under partial observations, we first design an efficient DNN reconfiguration scheme to support band-wise parameter training and sharing. The key idea is to cut the unnecessary connections between neurons among multi-task data flows for detecting different bands in heterogeneous DNNs. It is implemented by utilizing the band-specific neuron sensitivity in DNN and then decoupling their corresponding sub-networks, which will be discussed in detail next.

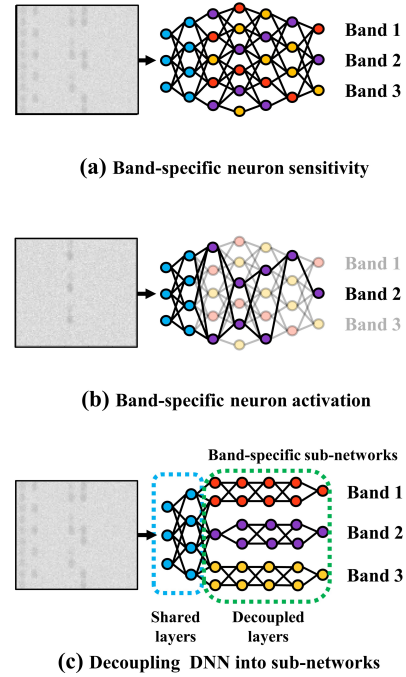


Fig. 2. Band-specific neuron separating and the proposed CNN structure.

1) *Hierarchical band-specific neuron sensitivity*: In DNNs, the neurons play as fundamental feature extractors by generating an effectively large output, once they are activated by certain features from their inputs [33]. We observe that neurons on different layers of a DNN exhibit a hierarchical nature in feature extraction. Specifically, in the multi-task learning model for wideband spectrum sensing as illustrated by Fig. 2.(a), there exist two facts: (1) the neurons in the shallow layers (in blue color) are sensitive to the common features that are widely presented by the PSD sample data collected from all bands; (2) the neurons in the deep layers corresponding to the multiple tasks highlighted in red, purple and yellow colors, respectively, are sensitive to the task-specific features of different bands.

For the neurons in the deep layers of the densely connected multi-task DNNs, while their outputs are calculated with all outputs from their prior layer, only those neurons that contribute large values can play a dominant role in decision making. In this sense, when a PSD data is fed into the multi-task DNN, the neurons of the shallow layers are fully activated while only a part of the neurons in individual deep layers are activated as shown in Fig. 2.(b). Accordingly, the data flow of the multi-task DNN for spectrum sensing passes all neurons of the shallow layers, but then only goes through some activated neurons of the deep layers selectively.

2) *Decoupling multi-band DNN into band-specific sub-networks*: Given the hierarchical nature of DNN neuron's sensitivity to band-specific input features, we identify the links through which the band-specific data flows pass within the multi-task DNN model. This insight motivates us to keep the corresponding links that carry each data flow related to the sensing task for a specific band. Meanwhile, we remove the

other unnecessary links between the neurons that belong to the different data flows. The key idea is that the original dense DNN structure can be compressed into a couple of compact sub-networks with (much) less training parameters without sacrificing learning capability. In this way, we decouple the original large multi-task DNN into multiple sub-networks. The proposed decoupling scheme is illustrated as shown in Fig. 2.(c). Since such sub-networks corresponding to individual spectrum sensing tasks have fewer parameters than the original dense DNN, they can be trained in a more efficient manner.

Note that our proposed reconfiguration scheme is different from the existing DNN pruning methods in the following two perspectives. (1) Our decoupling scheme is motivated by the hierarchical neuron sensitivity to different sensing tasks and is actually applied before training in a proactive manner. On the other hand, the existing pruning techniques are usually conducted to remove the insignificant weights from a well-trained model [34]. (2) Fig. 2.(a) and (c) indicate that our decoupling method neither changes the depth of DNN nor alters the number of its total neurons. However, the existing pruning method may lead to a decrease of the number of neurons (a.k.a., filter pruning), when all the input links connected to some neurons have trivial weights [35]. Furthermore, such variation in neuron numbers between different SUs may unfortunately lead to complicated coordination among these SUs for local model averaging in collaborative learning. Thus, our proposed architecture retains model capacity for multi-band spectrum learning at reduced model complexity.

B. Model design under partial observations

Different from the DNN classifiers that produce the multiple softmax outputs in [18], [19], we apply a sigmoid function to activate each output channel of our proposed multi-task DNN. Thus, each output digit independently represents the probability-based confidence value of the occupancy of each band and its value is restricted between 0 and 1. Accordingly, we adopt the binary cross-entropy (6) to calculate the loss function value of the output digit of each band, instead of the regular cross-entropy loss for classifiers in [18], [19]. Considering the 2-D nature of our PSD-based spectrum data $\mathbf{S}_j$ in (3), we choose convolutional neurons as the feature extractor [36], [37], to take advantage of the correlation in the 2-D input spectrum measurement data via the shift invariance of CNN¹.

Ideally, if the wideband spectrum pool is globally observable to all SUs, then the learning-based cooperative spectrum sensing boils down to a standard multi-band detection problem for training a homogeneous model with IID training data. However, for realistic CR systems under partial observations, the multi-task DNNs trained on different SUs turn to be heterogeneous. This is because practical CR systems usually focus on wideband spectrum sensing over a large physical area with the sensing-capability constrained SUs as illustrated

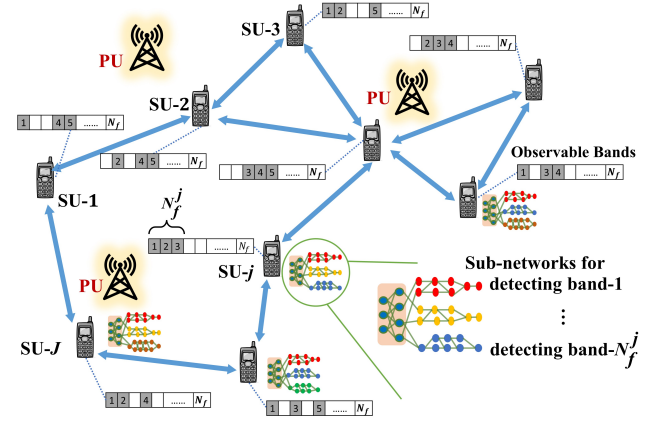


Fig. 3. Collaborative training system with partial observers.

in Fig. 1. Meanwhile, the path-loss related term h_n^j in (1) is inversely proportional to the distance of signal propagation. For example, if the PU on band- n is far away from SU- j , then the PSD of the received signal on this band becomes (much) smaller than that caused by noise. Then, the spectrum occupancy characteristics in the local PSD on this band, s_n^j , cannot be captured by the DNN detector. As a result, these kinds of bands become unobservable to SU- j . In this sense, the multi-task DNN at SU- j generates only N_f^j ($< N_f$) effective output logits, and each of them corresponds to one band that is observable locally at SU- j .

In order to obtain the global detection of the entire wideband spectrum, decision fusion is necessary to combine the local decisions from distributed SUs. However, under partial observations, if the multi-task DNN detector on each SU is independently trained in a standalone manner, then it is prone to over-fit the local model to the data collected under location-dependent wireless conditions, which makes the trained model infeasible for dynamic CR systems and insufficient local data. Considering that different SUs share partially overlapping bands, it is motivated to exchange the learned knowledge about the common spectrum occupancy on these overlapping bands across different SUs. However, conventional data parallel collaborative learning like FL hinges on homogeneous learning tasks and IID training data across all SUs, which is unrealistic under partial observations. To collaboratively learn the knowledge about wideband spectrum occupancy through band-specific parameter sharing, we next develop efficient band-wise collaborative training among DNNs for heterogeneous tasks with non-IID data.

C. Collaborative training among partial observers

Given the band-specific sub-networks decoupled via our DNN reconfiguration scheme proposed in Section IV-A, we now aim at developing collaborative training solutions to achieve effective cooperation among heterogeneous SUs in detecting their overlapping bands.

1) *Collaborative training system:* To detect the entire wideband spectrum pool with N_f bands under partial observations, we setup a learning-based cooperative spectrum

¹Note that the proposed neural network architecture and the corresponding methodology developed in this work can be extended to other DNN models as well.

sensing system that is composed of J multi-task DNNs. The j -th multi-task DNN is reconfigured into N_f^j sub-networks corresponding to the N_f^j observable bands at SU- j , as shown in Fig. 3. Our collaborative learning system can be deployed in a centralized topology. There exists a centralized fusion center to collect the locally updated model for parameter averaging and then broadcast the averaged model parameters back to the SUs for their next round local update [38]. Meanwhile, our proposed method can be applied in a decentralized manner as well, where the distributed SUs directly exchange their local updates with their neighboring SUs. This enhances the system robustness to node/communication failures [39]. For the implementation of the band-wise collaborative learning among SUs on their overlapping bands, the reconfigured sub-networks for the detection on the same bands should have the homogeneous network structures. The training scheme for such collaborative learning based cooperative spectrum sensing will be discussed next.

2) *Collaborative training scheme*: The training scheme for our collaborative learning based spectrum sensing consists of two alternative stages: local training and parameter averaging.

Local training: At SU- j , the parameters of the local multi-task DNN, denoted by $\mathcal{W}^j$, are optimized through stochastic gradient descent (SGD) with its locally available data. Here, these local model parameters $\mathcal{W}^j$ are updated by minimizing the binary cross-entropy loss function in a batch form as (5). Such a local training process can be conducted over multiple sub-networks simultaneously, thanks to the separable nature of these multi-task sub-networks, which is enabled by the proposed reconfiguration scheme via decoupling the original dense DNN structure. Further, compared with the training process for an occupancy-status classifier [18], [19], our training solution is actually optimized over a smaller searching space whose dimension is reduced by $2^{N_f^j - N_f^j}$.

Parameter averaging: After the above local training at SUs is completed, the learned knowledge in terms of the updated local model parameters $\mathcal{W}^j$ needs to be generalized through parameter averaging and then shared among SUs in collaborative learning. Considering the heterogeneous property between SUs due to their different partial observations of the entire wideband spectrum pool, the parameter averaging of $\{\mathcal{W}^j\}, j = 1, \dots, J$ should be conducted in a hierarchical way.

Since the shallow layers are common for all bands at all SUs, their parameter averaging is conducted as

$$\bar{\mathcal{W}} = \frac{1}{J} \sum_{j=1}^J (\bar{\mathcal{W}}^j), \quad (7)$$

where $\bar{\mathcal{W}}^j$ is the local parameters of the shallow layers at SU- j . For the deep layers, their parameter averaging is done in a band-wise manner. For band- n , the deep layer parameter averaging is operated as

$$\tilde{\mathcal{W}}_n = \frac{1}{|\mathcal{J}_n|} \sum_{j \in \mathcal{J}_n} \tilde{\mathcal{W}}_n^j, \quad (8)$$

where $\mathcal{J}_n$ is the set including the indices of SUs which can observe band- n , $|\mathcal{J}_n|$ is the cardinality of $\mathcal{J}_n$, and $\tilde{\mathcal{W}}_n^j$ denotes

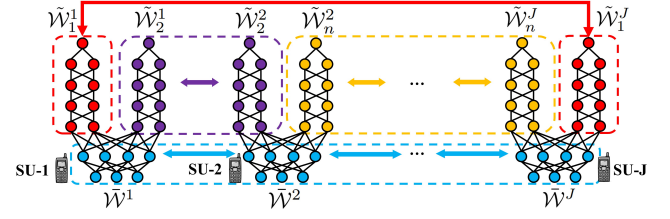


Fig. 4. Band-specific model averaging for collaboration.

the deep layer parameters at SU- j for detecting band- n . The collaborative training scheme is illustrated in Fig. 4, whose pseudo-code implementation is listed in Algorithm 1.

Algorithm 1 Collaborative training of shallow and deep layers.

- 1: Initialize $\mathcal{W}^j, j = 1, \dots, J$
 - 2: **for** each round $i = 1, 2, \dots, I$ **do**
 - 3: **for** each SU- $j, j = 1, \dots, J$ in parallel **do**
 - 4: $\mathcal{W}^j \leftarrow \text{Local training via SGD } (i, \mathcal{W}^j, \mathcal{D})$
 - 5: **end for**
 - 6: *Parameter averaging:*
 - 7: Shallow-layer averaging via (7)
 - 8: Deep-layer averaging via (8) for $n = 1 \dots N_f$
 - 9: **end for**
-

Remark 1 (Reduced complexity): The total number of trainable parameters of our decoupled multi-task DNN is greatly reduced compared with that of the original dense DNN. This benefits from the band-specific DNN structure reconfiguration via sub-network decoupling and removing the unnecessary links between the neurons belonging to different sub-networks. This can then greatly save the computation, storage, and communication loads consumed by individual SUs during the collaborative learning process.

Remark 2 (Increased adaptability): While our DNN reconfiguration and collaborative training are proposed to enhance the sensing performance and computation efficiency, the trained multi-task model also has good adaptability to new testing cases during its deployment in practice. The trained sub-networks can be reassembled to form a new DNN model without extra training. It is able to detect the spectrum occupancy of new multi-band combinations which have never been trained, when new SUs participate for new sensing tasks. This merit results from the separable nature of these multi-task sub-networks, which are well represented in our decoupling-based reconfiguration and efficiently trained through our hierarchical collaborative training.

V. SIMULATION RESULTS

In this section, we present the simulation results to evaluate the performance of the proposed collaborative learning based cooperative spectrum sensing², compared with the benchmark

²For implementation details, please refer to our online repository available at <https://github.com/FrancisZWS/PartialObservation>.

methods of standalone learning³, federated learning⁴, as well as the traditional cooperative spectrum sensing based on energy detection method.

A. Data generation

Considering the scarcity of wideband sensing datasets collected by distributed SUs, we first explain how to generate the synthetic wideband PSD data, which reflects the heterogeneity caused by partial observations. For the inherent correlations between different bands, we consider power leakage and channel aggregation effects which widely exist in realistic wideband environments.

In a wideband spectrum pool containing N_f bands, when band- n with $n \in \{1, \dots, N_f\}$ is occupied by a certain PU, we generate the original single-band PU signal $\mathbf{x}_n$ in (1) by modulating a random message sequence through a predefined modulation scheme used by the PU. This modulation scheme results in a specific inner-band feature in terms of the unique PSD waveform which is different to other modulation types.

To depict the power leakage issue between contiguous bands, we calculate the PSD of $\mathbf{x}_n$ in the form of an overflow $\mathbf{s}'_n \in \mathbb{R}^{3N_w}$ which includes PSD of its side lobes leaking to its adjacent bands:

$$\mathbf{s}'_n = [\mathbf{s}'_{n,L}, \mathbf{s}'_{n,M}, \mathbf{s}'_{n,R}], \quad (9)$$

where $\mathbf{s}'_{n,M} \in \mathbb{R}^{N_w}$ is the PSD main lobe on band- n , and $\mathbf{s}'_{n,L} \in \mathbb{R}^{N_w}$ and $\mathbf{s}'_{n,R} \in \mathbb{R}^{N_w}$ denote the left and right PSD sidelobes of $\mathbf{x}_n$ leaked to band- $(n-1)$ and band- $(n+1)$, respectively. Then, the band-wise PSD $\mathbf{s}_n$ collected by SU- j under the impact of potential power leakage from band- $(n-1)$ and band- $(n+1)$ can be generated as:

$$\mathbf{s}_n^j = y_n |h_n^j|^2 \mathbf{s}'_{n,M} + y_{n-1} |h_{n-1}^j|^2 \mathbf{s}'_{n-1,R} + y_{n+1} |h_{n+1}^j|^2 \mathbf{s}'_{n+1,L} + \mathbf{w}_n, \quad (10)$$

where $\mathbf{w}_n \in \mathbb{R}^{N_w}$ represents the AWGN noise over this band, y_n and h_n^j as defined in (1) indicate the ground truth occupancy condition and channel gain on band- n , respectively. The wideband PSD matrix $\mathbf{S}^j$ for the input of DNN on SU- j can be generated by repeating (10) over $n = 1, \dots, N_f$ and then stacking $\{\mathbf{s}_n^j\}_{n=1}^{N_f}$ according to (3). In this way, the band-wise PSD $\mathbf{s}_n^j$ is also affected by the occupancy of its contiguous bands and their sidelobes' waveforms. According to our discussion about partial observation in section IV-B, when the PU on band- n is far away from SU- j , the channel gain h_n^j become small correspondingly. When this band is occupied, the single-band PSD $\mathbf{s}_n^j$ we generated for SU- j has:

$$y_n |h_n^j|^2 \mathbf{s}'_{n,M} \ll \mathbf{s}_n^j \quad (11)$$

Consequently, the occupancy of band- n cannot be learned from our $\mathbf{s}_n^j$ because: $\mathbf{s}_n^j|_{y_n=1} \approx \mathbf{s}_n^j|_{y_n=0}$. In this way, the partial observation issue is reflected in our synthetic dataset.

Considering the factor of channel aggregation by the same PU over multiple bands, the occupancy conditions of certain

non-contiguous bands are strongly correlated. To reflect such spectrum dependencies, let $\mathbf{B}_n$ contain indices of bands aggregated with band- n . When $\forall n' \in \mathbf{B}_n \neq \phi$, we have $y_n = y_{n'}$. In this way, band- n shares the same spectrum occupancy condition as band- n' , when $n' \in \mathbf{B}_n$. For a certain SU, the partial observation to the spectrum pool can be generalized to the partial observation to different PUs.

B. Experimental settings and results

1) *Simulation settings*: In our simulation, a multi-band CR system monitors $N_f = 20$ spectrum bands where $N_w = 64$ frequency points are sampled for the PSD measurement of individual bands. Thus, the data dimensionality of one input PSD sample $\mathbf{S}^j$ to SU- j becomes (64×20) . The CR system contains 10 SUs and 10 PUs whose locations are shown in Fig. 5. Each PU is assigned to use one or several non-contiguous aggregated bands, which are specified in Table I. For wireless channels, we simulate the impact of path-loss with $\alpha = 3.71$, $\beta = 10^{-3.154}$, and the log-normal shadow fading as a Gaussian-distributed random variable with mean zero and standard deviation $\sigma_{\psi_n^j} = 3.65$ dB. The reference distance is set as $d_0 = 1$ m. Due to the partial observation issue, the input PSD samples collected by a certain SU can only reflect the occupancy of spectrum bands utilized by its adjacent 3 PUs, which are specified in Table II. The sensing performances of different methods are evaluated under a range of signal-to-noise ratios (SNRs) between $[-16, -2]$ dB.

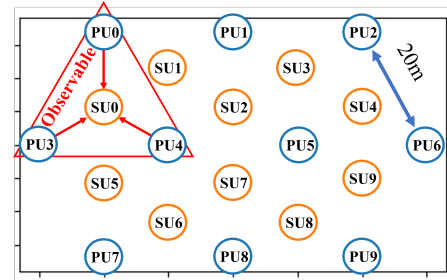


Fig. 5. Locations of PUs and SUs in the CR system.

TABLE I
BAND UTILIZATION OF PUS

PU index	Assigned bands	Modulation scheme
0	0	BPSK
1	1, 10	MSK
2	2, 11, 14	2FSK
3	3	16PSK
4	4, 19	4FSK
5	5, 13	16QAM
6	6, 15, 17	BPSK
7	7, 12, 18	MSK
8	8, 16	2FSK
9	9	16PSK

2) *Training and testing*: We let the 10 distributed SUs to conduct training for multi-band spectrum sensing given partial observations on their local data. The final decision on

³For standalone learning, each SU trains its own local model and only collaborates with other SUs in detection decision fusion.

⁴In federated learning, all SUs work on a homogeneous densely-connected CNN.

TABLE II
OBSERVABLE BANDS OF SUS

SU index	Observable PU	Observable bands
0	0, 3, 4	0, 3, 4, 19
1	0, 1, 4	0, 1, 4, 10, 19
2	1, 4, 5	1, 4, 5, 10, 13, 19
3	1, 2, 5	1, 2, 5, 10, 11, 13, 14
4	2, 5, 6	2, 5, 6, 11, 13, 14, 15, 17
5	3, 4, 7	3, 4, 7, 12, 18, 19
6	4, 7, 8	4, 7, 8, 12, 16, 18, 19
7	4, 5, 8	4, 5, 8, 13, 16, 19
8	5, 8, 9	5, 8, 9, 13, 16
9	5, 6, 9	5, 6, 9, 13, 15, 17

the occupancy of each specific band is obtained by doing a majority vote decision fusion among all SUs who can observe that band. Our collaborative learning method is compared with standalone learning and federated learning in five different cases with limited training data. For Case 1, we generate 20 local training samples on individual SUs for each of the total $2^{N_{\text{PU}}} = 1024$ occupancy patterns. Based on Case 1, we further evaluate the performance and generalization capability of our proposed model and method in another four different scenarios by modifying the data generation in training and testing stages. For Case 2, there are only 10 samples per occupancy pattern on individual SUs, which aggravates the insufficiency of training data. For Case 3, we test our proposed method across different channel conditions by changing the path-loss exponent of (1) from $\alpha = 3.71$ in training to $\alpha = 5.0$ in testing. For Case 4, we evaluate the capability of different methods in dynamic spectrum access scenarios, where PUs may change their modulation types from training to testing stages. Specifically, we randomly select the modulation type for all active PUs when generating input samples for training and testing. For Case 5, we testify the generalization power across different partial observation conditions by changing SU locations, which exposes SUs with different observable bands in training and testing. Different learning-based spectrum sensing methods are conducted with the same training scheduling scheme. The stochastic gradient descent is chosen as the optimizer and the training batch sizes are fixed to 50. The initial learning rates are equally set to 0.05 and it will be reduced by multiplying 0.2 every 20 training epochs. For our collaborative learning method and federated learning, the model averaging operations are conducted under the same frequency, which is once per training epoch.

3) *Our multi-task DNN*: The multi-task DNN model on each SU contains 4 convolutional layers followed by 1 fully connected output layer. The first convolutional layer is regarded as the shared shallow layer and it has 40 convolutional filters (*i.e.*, neurons). The following 3 convolutional layers are decoupled along the band-specific data flow, where 8 convolutional filters are allocated on each layer of one sub-network. Finally, the output layer is also band-wise decoupled so that each neuron accepts activation only from the last convolutional layer of its sub-network to generate a sigmoid output. Other configuration details for the DNN on SU- j 's are specified in Table III. The local detection outcome of

each band is calculated by comparing the sigmoid output of the matching sub-network with a pre-defined threshold, (*e.g.*, equal to 0.5). Our collaborative sensing method as well as other spectrum sensing methods involved in the simulation is implemented with PyTorch 2.0.1.

4) *Benchmark DNNs*: To the best of our knowledge, there is no existing work on learning-based wideband cooperative spectrum sensing among distributed narrowband SUs, given the challenges of heterogeneous learning tasks at partial observers with their non-IID data. Thus, we simulate the following methods as the benchmarks for our proposed solution. Energy detection followed by band-wise majority-vote decision fusion is tested as the conventional model-based spectrum sensing method [21]. To compare our solution with the off-the-shelf learning-based techniques that can be reasonably used under partial observations, we apply the standalone learning [33] and federated learning [25] for multi-band spectrum sensing. In this standalone learning method, each SU trains a heterogeneous densely connected CNN to detect its locally observable bands with its own dataset. To make a fair comparison, we let the dense standalone CNN on certain SU have the same number of neurons per layer as in our reconfigured multi-task DNN, but the latter has much fewer links between the neurons in the deep layers than the former after DNN reconfiguration and decoupling. In the federated learning method, each SU trains a homogeneous dense CNN that is the 20-band version of the standalone CNN, *i.e.*, it has 160 neurons for the 2nd, 3rd, and 4th convolutional layers. In this method, all DNNs use $\mathbf{y}$, which stores the occupancy of all bands, as the training label while the same majority-vote decision fusion strategy is applied in the testing phase to detect each band.

TABLE III
MULTI-TASK DNN ARCHITECTURE FOR SU- j

Layers	In_ch	Out_ch	Kernel	Groups
Conv1(ReLu) BatchNorm	1	40	(3×3)	1
Conv2(ReLu) BatchNorm	40	8	(3×3)	N_f^j
Maxpool1			(4×1)	
Conv3(ReLu) BatchNorm	8	8	(3×3)	N_f^j
Conv4(ReLu) BatchNorm	8	8	(3×3)	N_f^j
AvgPool			(4×5)	
FC(ReLu) Sigmoid	32	1		N_f^j

In_ch: the input channel number (per sub-network);
Out_ch: the output channel number (per sub-network);
Groups: the number of sub-networks;
Strides and paddings of convolutional filters are set to 1

5) *Complexity evaluation*: we compare the complexity of our method with that of standalone learning and federated learning in terms of model parameter size and computation cost. For the former, we compare the total number of trainable parameters of the DNNs on 10 SUs. For the latter, the entire Multiply-Accumulate Operations (MACs) of 10 SUs each processing one partially-observed input PSD sample is calculated as their computation costs. According to Table IV, the model

size of our collaborative learning method is only 5% of the federated learning method with homogeneous local models, while our computation cost is only 14% of it. Compared with the standalone learning method, our parameter volume is 40% of it while our computation cost is 77% of it. In this sense, our proposed method can significantly reduce the storage consumption and computation load on SUs devices.

TABLE IV
PARAMETER SIZES AND COMPUTATION COSTS

Method	Our method	Standalone	FL
Complexity			
Parameter size	253020	622428	5331400
MACs	245M	317M	1676M

C. Performance evaluation

In this part, we conduct simulations in different cases to compare our method with the benchmarks. First, we evaluate their training convergence under certain SNRs, then we plot the receiver operating characteristic (ROC) of different methods. We also compare their probability of detection (PD) under a fixed probability of false alarm (PFA) of 5% over all bands.

1) *Case 1 Sufficient training data:* As shown in Fig. 6, our proposed method converges to the highest sensing accuracy under SNRs of -4dB , -10dB , and -16dB . Compared with federated learning, our method achieves a higher convergence speed. This is because our model-decoupling operation can effectively separate the DNN parameters trained to detect different bands and coordinate model averaging accordingly. Due to the over-fitting issue caused by limited training data volumes, the final accuracy of each learning-based sensing method is lower than its highest accuracy in the training process. For the three SNR values in Fig. 6, the largest gap between our highest sensing accuracy and final sensing accuracy is only 1.0%, while for the standalone method and the federated learning method they are 5.0% and 5.4%, respectively. This means that our method is more resistant to the performance reduction caused by the over-fitting issue.

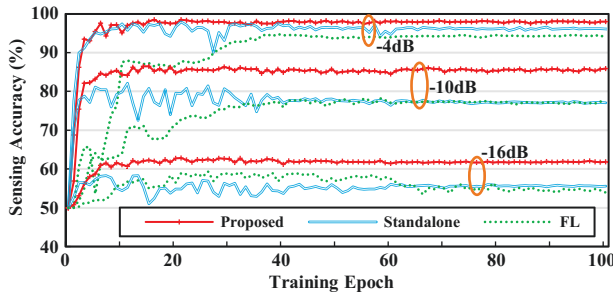


Fig. 6. Training convergence of learning-based methods in Case 1.

We also run the ROC curve of the best models of different methods under these SNRs. As shown in Fig. 7, our proposed collaborative learning method achieves the best ROC performance under all SNRs. Thanks to the power of deep learning, our method and the standalone learning method outperform the

energy detection method which is denoted as ED in the plot. In comparison, the learning capability of federated learning is inconsistent, especially under $\text{SNR} = -4\text{dB}$ when its ROC is the worst among these methods. Further, We compare the PD of different methods under a range of SNRs between $[-16, 2]\text{dB}$. As plotted in Fig. 8, our method achieves the best performance under all SNRs. The standalone learning method outperforms the energy detection method while the federated learning method achieves the worst performance under high SNRs. That is because the homogeneous DNN structure and the trivial model averaging scheme of federated learning are unsuitable for collaborative learning of SUs which detect heterogeneous bands due to partial observation.

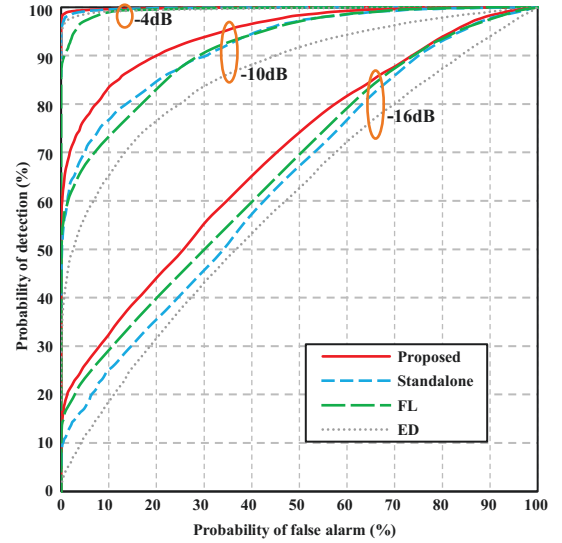


Fig. 7. The ROC of different methods in Case 1.

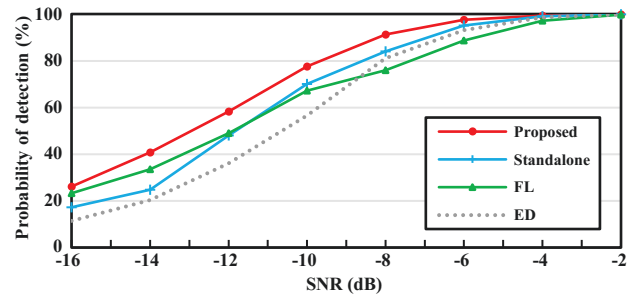


Fig. 8. The PD of different models given PFA = 5% in Case 1 with sufficient training data.

2) *Case 2 Small training data:* We evaluated the performance of these methods in Case 2 where the smaller training datasets aggravate the over-fitting issue. As shown in Fig. 9, our proposed method still achieves the best sensing accuracy among these methods during the training process. Meanwhile, we also achieve the highest convergence speed in this case. With the smaller training data volume in this case, the highest accuracy of the federated learning method under each SNR is apparently lower than those in Case 1. This indicates that the federated learning method with homogeneous local DNNs

is more vulnerable to the over-fitting issue caused by data insufficiency.

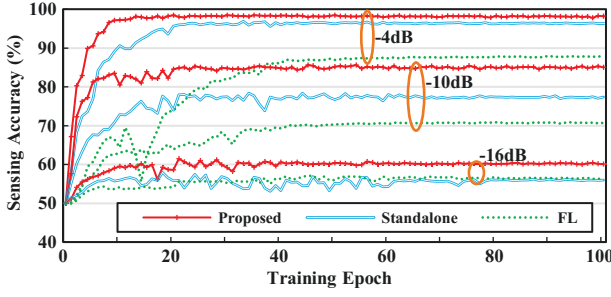


Fig. 9. Training convergence of learning-based methods in Case 2.

We also run the ROC curve of different methods under the same selected SNRs, which is shown in Fig. 10. While the ROC performances of the standalone learning and the federated learning methods degrade significantly due to the reduced training data volume in this case, the ROC of our method is still the best, which remains nearly unchanged as in Case 1. We also run the PD of these methods under the same range of SNRs covered in Case 1. As shown in Fig. 11, our method still achieves the highest PD under all SNRs. Due to the smaller training dataset, the standalone learning method works worse than energy detection under high SNRs in $[-8, -2]$ dB. The performance of federated learning is also degraded that it works worse than energy detection under most SNRs. By comparison, our collaborative learning method is the only learning-based method that consistently outperforms energy detection under all SNRs in this case.

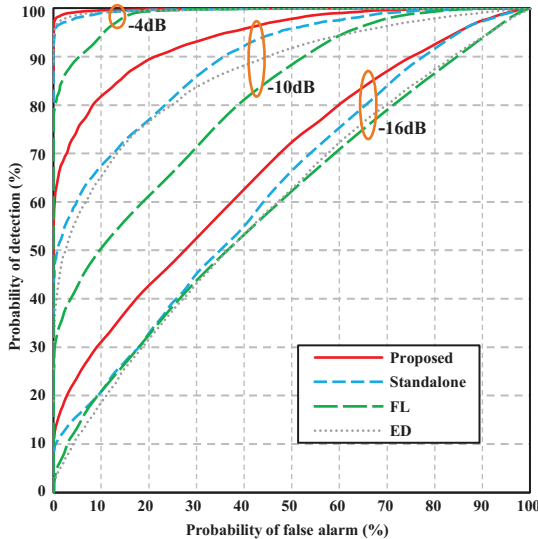


Fig. 10. The ROC of different methods in Case 2.

3) *Case 3 Varying channel condition:* We evaluate these methods by comparing their performances on the testing data with $\alpha = 5.0$ after their models are trained with $\alpha = 3.71$. As shown in the simulation results in Fig. 12, for a given PFA, the PD of our proposed method is apparently higher than the benchmarks. This demonstrates that the proposed method can

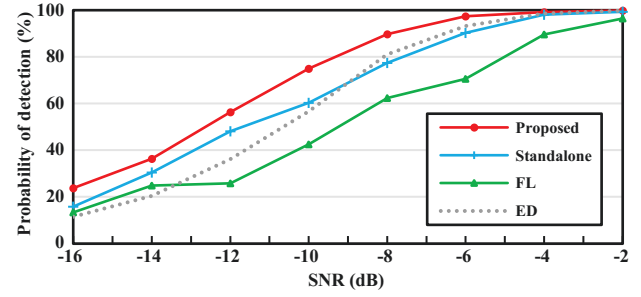


Fig. 11. The PD of different models given PFA = 5% with small-volume training data.

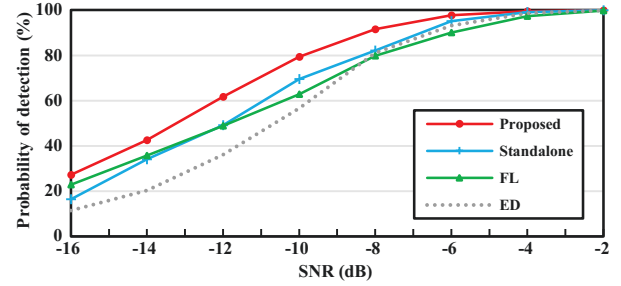


Fig. 12. PD of different models given PFA = 5% with different channel conditions ($\alpha = 3.71$ in training and then $\alpha = 5$ in testing).

generalize across different path-loss conditions and achieve better sensing accuracy than the baselines.

4) *Case 4 Dynamic modulation type:* The training data in this case is applicable for a wide range of scenarios where the exact modulation types used by PUs are varying and unknown. As shown in Fig. 13, the proposed method always outperforms the standalone learning and federated learning. This is because our proposed model and method can take advantage of band-specific collaborative learning, via hierarchical parameter sharing between heterogeneous models at distributed SUs, under partial observations. Our proposed method retains the capability of collaborative deep learning by enabling band-specific parameter sharing between heterogeneous models to overcome the partial observation issue.

5) *Case 5 Varying partial observation condition:* We evaluate the generalization power of our method to varying partial observation conditions from training to testing. For our method, when SU location changes after the model is trained, we can reconfigure the DNN on SU to match the observable bands in testing as discussed in Remark 2. As shown in Fig. 14, our method still works better than other benchmark methods in this case. Comparing with Fig. 8 where the partial observation condition in testing aligns with that in training, our method remains its performance in PD. This means that the knowledge learned under the historical partial observation condition is successfully generalized to detect the new observable bands at SUs. In contrast, the standalone learning method fails to work, as it suffers from the model mismatch issue seriously.

We also run the ROC curves of different methods in Cases 3, 4, and 5, when SNR = -8 dB, in Fig. 15. The ROC performances of the standalone learning and the federated

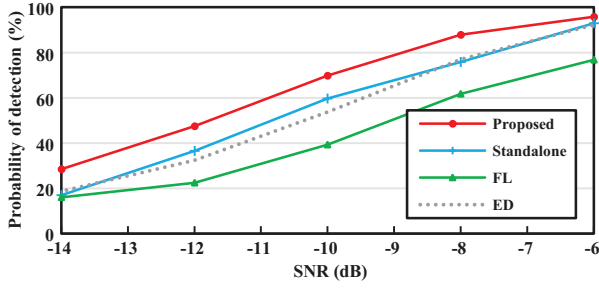


Fig. 13. PD of different models given PFA = 5% when PUs randomly use modulation types in training and testing.

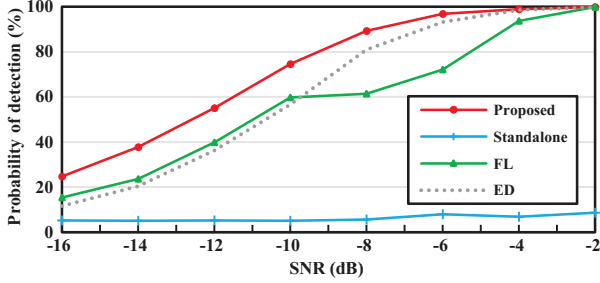


Fig. 14. PD of different models given PFA = 5% when SUs have different locations from training to testing.

learning methods degrade significantly in Case 4 and Case 5, which are even worse than the conventional energy detection method. This means that the local CNNs that focus on static observable bands in the training of the standalone learning are ineffective in generalizing across dynamic modulation types or changing partial observation conditions. The federated learning method, on the other hand, suffers from parameter mismatch problems, when the homogeneous CNNs trained corresponding to heterogeneous observable bands are averaged for partial observers. The ROC of our method significantly outperforms all the baselines, while only showing very tiny variations across Cases 3, 4, and 5. This means that our method is not only effective for dynamic spectrum environments but also highly transferable to generalize across channel conditions and partial observation conditions different from those in the training stage.

6) *Runtime evaluation*: we compare the execution time of our method with that of the federated learning, by running their training process on our desktop computer with an Intel Core i9-12900KF CPU, an Nvidia RTX3090 GPU, and 128GB-RAM. Specifically, we record their wall-clock runtime taken to train their corresponding models in one iteration, i.e., calculating the gradient with a single mini-batch of data, and then plot the trend of their runtime versus varying batch sizes. The model adopted for federated learning is a densely-connected CNN with $N_f = 20$ sigmoid outputs, whose specifications have been provided in Section V.B.4). For our decoupled CNN models, we test their runtime with three different model configurations. In Configuration-1, all 20 sub-networks of our model are activated to detect all 20 bands. In Configuration-2 or Configuration-3, only 8 or 4 sub-networks of our model are activated to detect 8 or 4 bands, corresponding to the

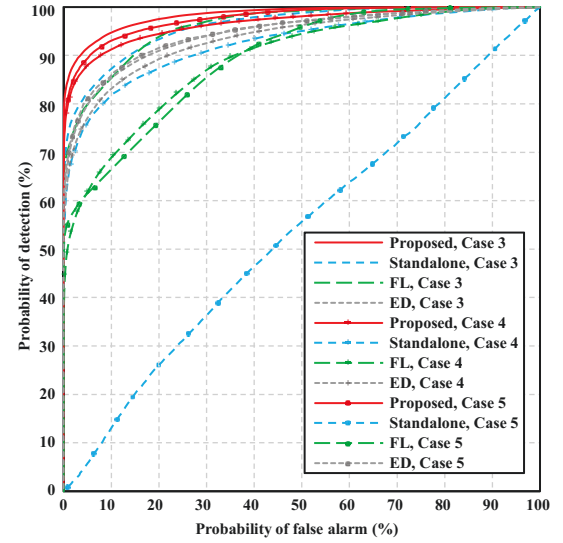


Fig. 15. The ROC of different models in Cases 3, 4, and 5, when SNR = -8dB.

partial observation scenarios such as SU-4 or SU-0 specified in Table II, respectively. As shown in Fig. 16, our decoupled DNN always runs faster than the densely-connected CNN, even when all 20 sub-networks are activated as in the FL benchmark, thanks to the band-wise decoupling architecture. Moreover, as the number of observable bands decreases, the runtime can be further reduced at the cost of reduced sensing capability per SU.

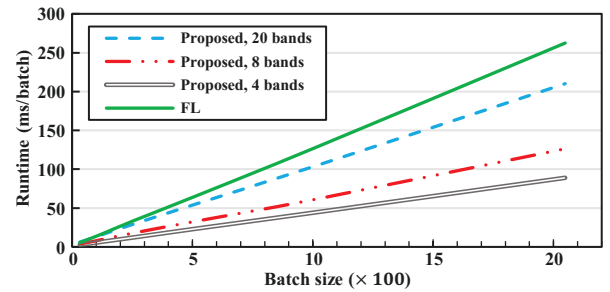


Fig. 16. Runtime of different models per batch for varying batch size.

In summary, all these results demonstrate that our collaborative learning method is suitable for wideband sensing scenarios under partial observations and data insufficiency. Considering the relatively low model complexity and computation costs, our method achieves a desired trade-off between wideband sensing performance and model training complexity. As a result, our method not only improves model efficiency through DNN reconfiguration but also enhances learning capability under small data, through collaborative training of shared tasks between heterogeneous models on different SUs.

VI. CONCLUSIONS

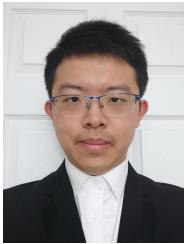
This paper develops a novel collaborative learning framework with distributed partial observers to conduct wideband cooperative spectrum sensing. Capitalizing on the hierarchical neuron sensitivity of deep neural networks to band-specific

features, our proposed technique decouples the original large deep neural network into smaller heterogeneous sub-networks, which are collaboratively trained at distributed secondary users detecting the overlapping bands. The simulation results verify that our method achieves higher learning accuracy and computation efficiency with faster convergence speed and presents better robustness to noise effect than the existing benchmarks.

REFERENCES

- [1] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," *IEEE personal communications*, vol. 6, no. 4, pp. 13–18, 1999.
- [2] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing - ICASSP '07*, vol. 4, 2007, pp. IV–1357–IV–1360.
- [3] Y. Wang, Z. Tian, and C. Feng, "Sparsity order estimation and its application in compressive spectrum sensing for cognitive radios," *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 2116–2125, 2012.
- [4] Z. Tian, Y. Tafesse, and B. M. Sadler, "Cyclic feature detection with sub-Nyquist sampling for wideband spectrum sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 1, pp. 58–69, 2012.
- [5] Z. Qin, X. Zhou, L. Zhang, Y. Gao, Y.-C. Liang, and G. Y. Li, "20 years of evolution from cognitive to intelligent communications," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 1, pp. 6–20, 2019.
- [6] A. Sonnenschein and P. M. Fishman, "Radiometric detection of spread-spectrum signals in noise of uncertain power," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, pp. 654–660, 1992.
- [7] C. Jiang, Y. Li, W. Bai, Y. Yang, and J. Hu, "Statistical matched filter based robust spectrum sensing in noise uncertainty environment," in *2012 IEEE 14th International Conference on Communication Technology*. IEEE, 2012, pp. 1209–1213.
- [8] J. Lundén, V. Koivunen, A. Huttunen, and H. V. Poor, "Spectrum sensing in cognitive radios based on multiple cyclic frequencies," in *2007 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications*. IEEE, 2007, pp. 37–43.
- [9] Y. Arjoun and N. Kaabouch, "A comprehensive survey on spectrum sensing in cognitive radio networks: Recent advances, new challenges, and future research directions," *Sensors*, vol. 19, no. 1, p. 126, 2019.
- [10] J. Xie, J. Fang, C. Liu, and X. Li, "Deep learning-based spectrum sensing in cognitive radio: A cnn-lstm approach," *IEEE Communications Letters*, vol. 24, no. 10, pp. 2196–2200, 2020.
- [11] H. Xing, H. Qin, S. Luo, P. Dai, L. Xu, and X. Cheng, "Spectrum sensing in cognitive radio: A deep learning based model," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 1, p. e4388, 2022.
- [12] R. Mei and Z. Wang, "Compressed spectrum sensing of sparse wideband signals based on deep learning," *IEEE Transactions on Vehicular Technology*, 2024, Early access.
- [13] K. Chae and Y. Kim, "Ds2ma: A deep learning-based spectrum sensing scheme for a multi-antenna receiver," *IEEE Wireless Communications Letters*, 2023, Early access.
- [14] W. M. Lees, A. Wunderlich, P. J. Jeavons, P. D. Hale, and M. R. Souryal, "Deep learning classification of 3.5-ghz band spectrograms with applications to spectrum sensing," *IEEE transactions on cognitive communications and networking*, vol. 5, no. 2, pp. 224–236, 2019.
- [15] P. Shachi, K. Sudhindra, and M. Suma, "Convolutional neural network for cooperative spectrum sensing with spatio-temporal dataset," in *2020 International Conference on Artificial Intelligence and Signal Processing (AISP)*. IEEE, 2020, pp. 1–5.
- [16] W. Lee, M. Kim, and D.-H. Cho, "Deep cooperative sensing: Cooperative spectrum sensing based on convolutional neural networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 3005–3009, 2019.
- [17] J. Gao, X. Yi, C. Zhong, X. Chen, and Z. Zhang, "Deep learning for spectrum sensing," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1727–1730, 2019.
- [18] Y. Zhang, B. Shen, J. Wang, and T. Yan, "Cnn based wideband spectrum occupancy status identification for cognitive radios," in *2020 International Conference on Wireless Communications and Signal Processing (WCSP)*. IEEE, 2020, pp. 569–574.
- [19] N. Ambika, K. Muthumeenakshi, and S. Radha, "Classification of primary user occupancy using deep learning technique in cognitive radio," in *Advances in Automation, Signal Processing, Instrumentation, and Control*. Springer, 2021, pp. 1795–1804.
- [20] Z. Qin, H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep learning in physical layer communications," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 93–99, 2019.
- [21] Y. Wang, Z. Tian, and C. Feng, "Collecting detection diversity and complexity gains in cooperative spectrum sensing," *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2876–2883, 2012.
- [22] J. Ma, G. Zhao, and Y. Li, "Soft combination and detection for cooperative spectrum sensing in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4502–4507, 2008.
- [23] J. Unnikrishnan and V. V. Veeravalli, "Cooperative sensing for primary detection in cognitive radio," *IEEE Journal of selected topics in signal processing*, vol. 2, no. 1, pp. 18–27, 2008.
- [24] M. Li, D. G. Andersen, J. W. Park, A. J. Smola, A. Ahmed, V. Josifovski, J. Long, E. J. Shekita, and B.-Y. Su, "Scaling distributed machine learning with the parameter server," in *11th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 14)*, 2014, pp. 583–598.
- [25] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 1273–1282.
- [26] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *International conference on engineering applications of neural networks*. Springer, 2016, pp. 213–226.
- [27] D. Janu, K. Singh, S. Kumar, and S. Mandia, "Hierarchical cooperative lstm-based spectrum sensing," *IEEE Communications Letters*, vol. 27, no. 3, pp. 866–870, 2023.
- [28] J. Verbraeken, M. Wolting, J. Katzy, J. Kloppenburg, T. Verbelen, and J. S. Rellermeyer, "A survey on distributed machine learning," *ACM Computing Surveys (CSUR)*, vol. 53, no. 2, pp. 1–33, 2020.
- [29] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
- [30] T.-H. Yu, S. Rodriguez-Parera, D. Markovi, and D. abri, "Cognitive radio wideband spectrum sensing using multitar windowing and power detection with threshold adaptation," in *2010 IEEE International Conference on Communications*, 2010, pp. 1–6.
- [31] Nidhi, A. Mihovska, and R. Prasad, "Overview of 5g new radio and carrier aggregation: 5g and beyond networks," in *2020 23rd International Symposium on Wireless Personal Multimedia Communications (WPMC)*, 2020, pp. 1–6.
- [32] K. P. Murphy, *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press, 2012.
- [33] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [34] Z. Qin, F. Yu, C. Liu, and X. Chen, "Captor: A class adaptive filter pruning framework for convolutional neural networks in mobile applications," in *Proceedings of the 24th Asia and South Pacific Design Automation Conference*, 2019, pp. 444–449.
- [35] J.-H. Luo, J. Wu, and W. Lin, "Thinet: A filter level pruning method for deep neural network compression," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017, pp. 5068–5076.
- [36] Z. Qin, F. Yu, C. Liu, and X. Chen, "How convolutional neural network see the world-a survey of convolutional neural network visualization methods," *arXiv preprint arXiv:1804.11191*, 2018.
- [37] —, "Functionality-oriented convolutional filter pruning," *arXiv preprint arXiv:1810.07322*, 2018.
- [38] P. Xu, Z. Tian, Z. Zhang, and Y. Wang, "Coke: Communication-censored kernel learning via random features," in *Proc. IEEE Data Science Workshop (DSW)*, 2019, pp. 32–36.
- [39] P. Xu, Y. Wang, X. Chen, and T. Zhi, "COKE: communication-censored kernel learning for decentralized non-parametric learning," *Journal of Machine Learning Research*, vol. 22, 2021. [Online]. Available: <https://www.jmlr.org/papers/volume22/20-070/20-070.pdf>

BIOGRAPHIES



Weishan Zhang is currently a PhD student with the Department of Electrical and Computer Engineering at the George Mason University. He received his B.S. degree in Electronic Science and Technology from University of Science and Technology of China, Hefei, China, in 2018. His current research focuses on deep learning for wideband spectrum sensing.



Zhi Tian is currently a Professor with the ECE Department of George Mason University since 2015. Prior to that, she was on the faculty of Michigan Technological University from 2000 to 2014. She served as a program director with the U.S. National Science Foundation from 2012 to 2014. Her general interests are in the areas of signal processing, communications, detection and estimation. Her current research focuses on decentralized optimization and learning over networks, statistical inference from distributed data, compressed sensing for random processes, cognitive radios, and millimeter-wave MIMO communications. She served on the Board of Governors of the IEEE Signal Processing Society from 2019 to 2021. She was Chair of the IEEE Signal Processing Society Big Data Special Interest Group, and a member of the IEEE Signal Processing for Communications and Networking Technical Committee. She was a Distinguished Lecturer of both the IEEE Communications Society and the IEEE Vehicular Technology Society. She is Editor-in-Chief for IEEE Transactions on Signal Processing, and served as Associate Editor for IEEE Transactions on Wireless Communications and IEEE Transactions on Signal Processing. She received the IEEE Communications Society TCCN Publication Award in 2018. She is a Fellow of the IEEE.



Yue Wang received the Ph.D. degree in communication and information system from the School of Information and Communication Engineering, the Beijing University of Posts and Telecommunications, Beijing, China, in 2011. He is currently an Assistant Professor in the Department of Computer Science at the Georgia State University, Atlanta, GA, since August 2023. Prior to that, he was a Research Assistant Professor in the Electrical and Computer Engineering Department at the George Mason University, Fairfax, VA. His general interests

include the interdisciplinary areas of machine learning, signal processing, wireless communications, and their applications in cyber physical systems. His specific research focuses on distributed optimization and machine learning, sparse signal processing, massive MIMO, mmWave communications, cognitive radios, spectrum sensing, Internet of Things, direction of arrival estimation, and high-dimensional data analysis. He is a Senior Member of the IEEE.



Xiang Chen received the Ph.D. degree in computer engineering from the University of Pittsburgh, Pittsburgh, PA, USA, in 2016. After that, he joined George Mason University, Fairfax, VA, USA, and founded the Intelligence Fusion Laboratory. His research works focus on high-performance computing, artificial intelligence, large-scale systems, and various mobile and edge applications. He also received the NSF CAREER Award, the Best Paper Award in DATE, and the several other competition awards.

With close collaboration with industrial labs and vast universities, he is currently leading multiple research projects funded by NSF and Air Force Research Lab (AFRL).



Lingjia Liu received the B.S. degree in electronic engineering from the Shanghai Jiao Tong University and the Ph.D. degree in electrical and computer engineering from the Texas A& M University. Currently, he is a Professor and the Bradley Senior Faculty Fellow with the ECE Department, Virginia Tech (VT), and also the Director of the WirelessVirginia Tech, a center focusing on wireless technology. He spent more than four years with the Mitsubishi Electric Research Laboratory (MERL) and the Standards and Mobility Innovation Laboratory, Samsung

Research America (SRA), where he received the Global Samsung Best Paper Award in 2008 and 2010. He was leading Samsungs efforts on multiuser MIMO, CoMP, and HetNets in 3GPP LTE/LTE-advanced standards. His general research interests include enabling technologies for 5G-advanced/6G networks, including incorporating domain knowledge into machine learning for wireless networks, massive MIMO, massive MTC communications, and mm-wave communications. His research received eight best paper awards. In 2021, he received the VT College of Engineering Deans Award for Excellence in Research.