Deep Spectrum Prediction in High Frequency Communication Based on Temporal-Spectral Residual Network

Ling Yu[1], Jin Chen[1,*], Yuming Zhang[1], Huaji Zhou[2], Jiachen Sun[1]

[1] College of Communications Engineering, Army Engineering University of PLA, Nanjing 210000, China
[2] Science and Technology on Communication Information Security Control Laboratory, Jiaxing 314033, China
* The corresponding author, email: chenjin99@263.net

Abstract: High frequency (HF) communication is widely spread due to some merits like easy deployment and wide communication coverage. Spectrum prediction is a promising technique to facilitate the working frequency selection and enhance the function of automatic link establishment. Most of the existing spectrum prediction algorithms focus on predicting spectrum values in a slot-by-slot manner and therefore are lack of timeliness. Deep learning based spectrum prediction is developed in this paper by simultaneously predicting multi-slot ahead states of multiple spectrum points within a period of time. Specifically, we first employ supervised learning and construct samples depending on long-term and short-term HF spectrum data. Then, advanced residual units are introduced to build multiple residual network modules to respectively capture characteristics in these data with diverse time scales. Further, convolution neural network fuses the outputs of residual network modules above for temporal-spectral prediction, which is combined with residual network modules to construct the deep temporal-spectral residual network. Experiments have demonstrated that the approach proposed in this paper has a significant advantage over the benchmark schemes.

Keywords: HF communication; deep learning; spectrum prediction; temporal-spectral residual network.

I. INTRODUCTION

1.1 Background and motivation

High frequency (HF) communication whose frequency range is from 3 MHz to 30 MHz is capable of realizing over-the-horizon communication depending on ionosphere reflecting signals. Due to its appealing merits such as fast network reconstruction and strong maneuverability, HF communication is widely used in military, emergency and transoceanic communications [1, 2]. The ionospheric characteristics are influenced by various factors such as time, season, weather, geographical location and solar activity [2, 3], thus available communication frequency window and channel quality also correspondingly vary. Therefore, it is a vital but difficult task to choose the appropriate working frequency in HF communication. Spectrum prediction, one of critical technology, combined with channel selection and automatic link establishment is a promising solution, which is expected to improve the success rate and reliability of link

Received: May 15, 2018
Revised: Jul. 13, 2018
Editor: Yuhua Xu

Deep learning based spectrum prediction is developed in this paper by simultaneously predicting multi-slot ahead states of multiple spectrum points within a period of time.

establishment. For instance, spectrum prediction can be used in either HF radio transmitters or receivers to preliminarily select a batch of candidate channels with high quality, which avoids the blindness of frequency allocation and makes preparations for channel sounding. The actual working frequency can be further determined by channel sounding on these predicted candidate channels which can reduce time consumption and then link is automatically established.

For any user in HF communication, both long-term and short-term predictions are of great significance [4, 5]. However, most of prediction algorithms only consider making short-term prediction in a slot-by-slot manner. In this paper, we investigate long-term prediction of HF communication, which is also referred as deep spectrum prediction because the proposed spectrum model makes predictions of longer time span than traditional models and mines inner relationship of spectrum data more deeply than other benchmark schedules. Spectrum data are modeled as an image series and an idea of spectrum data visualization is introduced from computer vision, which is inspired by the work in reference [6]. The application of deep learning in computer vision field has gained wide attention in recent years [7, 8, 9]. Compared to tremendous effort of hand-crafted feature required in machine learning, deep learning can automatically extract and learn the features of samples better, bringing a substantial improvement in performance [10]. Thus, its impressive abilities to extract feature and solve problem motivated us to apply deep learning in temporal-spectral deep prediction. The future characteristics of HF communication are wideband, intelligent and integrated. There is no doubt that the application of emerging deep learning in spectrum prediction can make HF communication more intelligent. Based on observations above, we construct the spectrum images and reasonably design the HF spectrum dataset based on different time scales. The residual neural network and convolution neural network in deep learning are tailored to make deep prediction of HF spectrum points. To our best knowledge, this is the first time to apply deep learning in spectrum prediction from an image perspective.

1.2 Related work

Many studies have been carried out in spectrum prediction which can be found in the survey [11]. These studies mainly apply methods like hidden markov model (HMM) [12], partial periodic pattern [13] and artificial neural network [14] to predict binary series of spectrum occupancy, or make predictions of specific spectrum values based on support vector regression (SVR) [15]. Recently, long short term memory (LSTM) neural network has been applied in spectrum prediction [16]. Majority of existing prediction algorithms belong to short-term prediction and their datasets are constructed by sliding window, where spectrum states in next time slot are predicted by mining historical data within fixed time slots.

Studies of effective prediction algorithm in HF band are relatively few. It is very difficult to predict specific spectrum values because HF band is complex and influenced by many external factors. Reference [17] improves the accuracy of ionosphere prediction model by introducing the International Reference Ionosphere model and adding a parameter amending module to so-called ITS model. Reference [18] is one of the first studies to apply cognitive radio in HF communication, and make short-term predictions of the sojourn time of a primary user in the band based on HMM model. Reference [19] employs neural network to successfully capture the 24-hour, seasonal and long-term trend in the variability of congestion of HF band for broadcast users based on several years data.

1.3 Contributions

The main contributions of this paper are summarized as follows:

- We propose a spectrum data visualization approach to make deep spectrum prediction from an image inference perspective. Different from the existing spectrum prediction in a slot-by-slot manner, the proposed approach is more efficient which can predict multi-slot ahead spectrum states of multiple spectrum points simultaneously.

- We combine multiple neural network modules to construct the proposed deep temporal-spectral residual network. Each module consisting of residual units is designed to capture the internal relationship of HF spectrum data with different time scales.
- We validate the superiority of the constructed temporal-spectral residual network for deep spectrum prediction. Real-world HF spectrum data are used in experiments and it is observed that prediction performance of the proposed approach is better than the benchmark algorithms.

The remainder of this paper is organized as follows. Section II compares the difference between the proposed model and traditional prediction model, and analyzes the correlation of HF spectrum data in the time and frequency domain. Section III presents the structure of the deep learning neural network and construction of dataset. Section IV shows the experimental results of deep prediction for each HF spectrum point. The last part provides conclusions for this paper.

II. System Model and Problem Formulation

In HF radio communication, spectrum prediction can effectively reduce the time and energy consumption of frequency selection. Either the prediction of single spectrum point or prediction of multiple spectrum points is based on sliding window in most related studies. As shown in figure 1(a), traditional prediction model learns the inherent relationship from N consecutive column vectors $\{\boldsymbol{x}_{t-N}, \boldsymbol{x}_{t-N+1}, ..., \boldsymbol{x}_{t-2}, \boldsymbol{x}_{t-1}\}$ to predict $\boldsymbol{x}_t$ and each column vector represents spectrum data of F spectrum points. Window with fixed length moves forward slot-by-slot over time and states or qualities of spectrum points in different time slots can be predicted. Considering the actual demand and forecast timeliness, the prediction model needs to simultaneously predict the values with the acceptable error within a relatively long period of time. Therefore, we consider deep prediction where spectrum data in a fixed time period are considered as an image. Taking figure 1(b) as an example, spectrum data from the tth slot to the $(t+T')$ th slot of F spectrum points are reshaped to be an image which is denoted as $\mathbf{X}_t$. The height of the image represents the number of slots while the width of the image represents the number of spectrum points. Then all achieved HF spectrum data can be formed into an image series $\mathcal{X} = \{..., \boldsymbol{X}_t, \boldsymbol{X}_{T'+t}, \boldsymbol{X}_{2T'+t}, ...\}$. Time span T' can be adjusted depending on practical needs. Likewise, the proposed prediction model learns relationship from the several spectrum images in the previous time periods and predicts the spectrum values of multiple frequency points within multi-slot ahead.

The main differences between traditional and proposed spectrum models are listed as follows:

- Spectrum data within a relatively long time period are modeled as an image in the proposed model. Multiple consecutive images are concatenated into a three-dimensional tensor as one sample. While in traditional prediction model, spectrum data in several continuous time slots are constructed as one

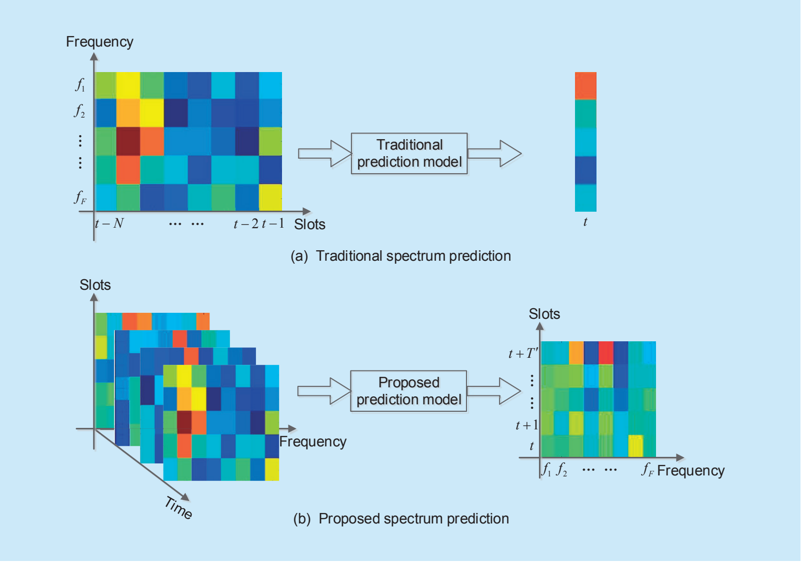


Fig. 1. *The comparison of prediction models.*

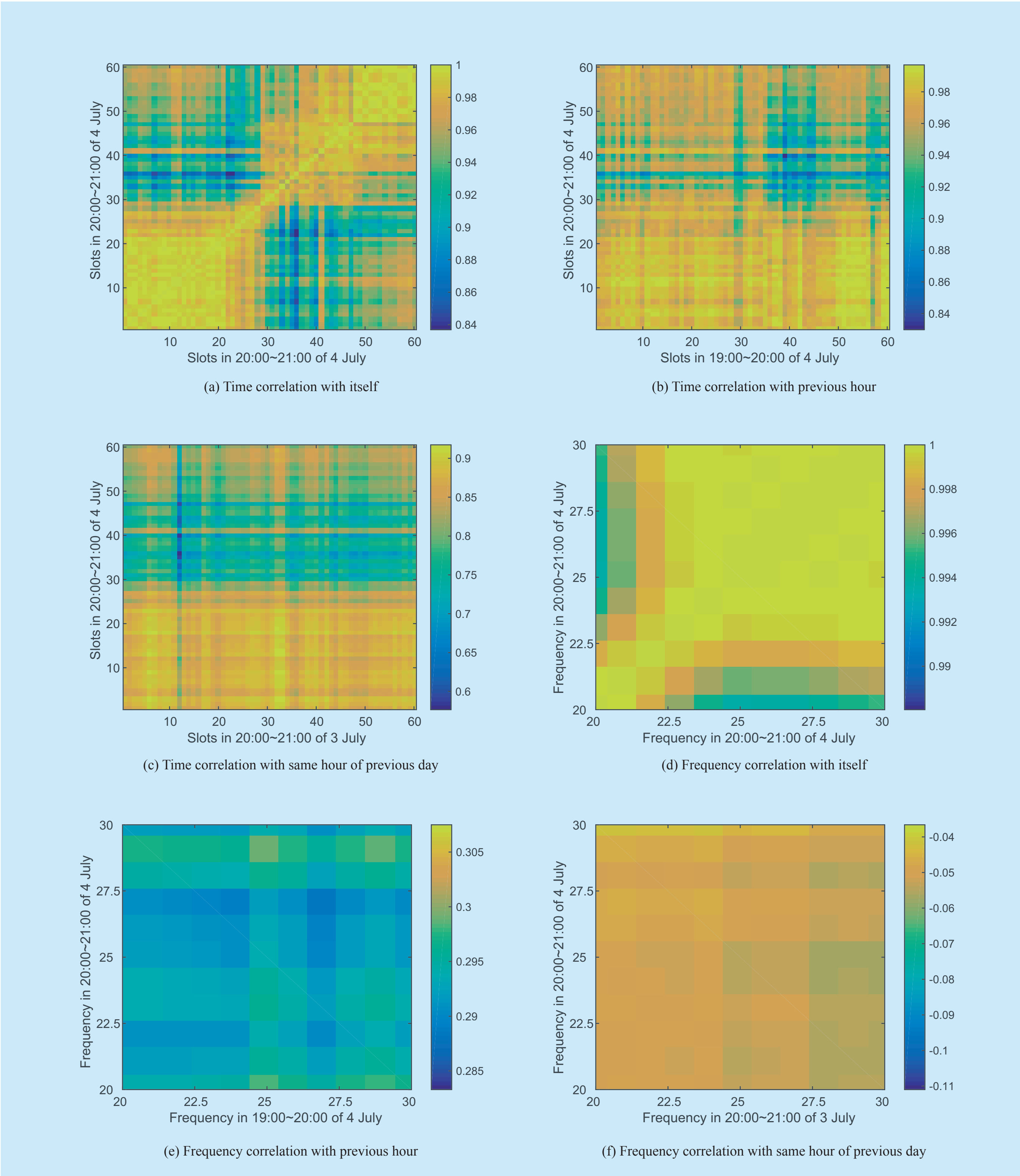


Fig. 2. *The correlations of spectrum data in 20:00~21:00 of 4 July with data in other three time intervals.*

sample whose size is two-dimensional.

- The prediction efficiency of the proposed model is higher than the traditional one. It can simultaneously predict values of multiple spectrum points within a relatively long time period. But traditional model has to carry out several predictions to achieve similar results. The longer time period is, the more times traditional model makes prediction.

Before spectrum data visualization, we first analyze the correlation of these data in the time domain and frequency domain. The raw spectrum data are organized as matrix with size $F \times T$, where $\boldsymbol{x}_{\cdot,t}, t \in \{1,...,T\}$ represents values of all spectrum points in the tth slot and $\boldsymbol{x}_{f,\cdot}, f \in \{1,2,...F\}$ represents evolution trajectories of one spectrum point. The correlation in the time domain and frequency domain is calculated by Pearson's correlation coefficient [20], as equations (1) and (2) show, respectively:

$$\rho_{t_1,t_2} = \frac{\mathrm{cov}(\boldsymbol{x}_{\cdot,t_1}, \boldsymbol{x}_{\cdot,t_2})}{\sigma_{\boldsymbol{x}_{\cdot,t_1}} \sigma_{\boldsymbol{x}_{\cdot,t_2}}}, \qquad t_1, t_2 \in \{1,...,T\} \quad (1)$$

$$\rho_{f_1,f_2} = \frac{\mathrm{cov}(\boldsymbol{x}_{f_1,\cdot}, \boldsymbol{x}_{f_2,\cdot})}{\sigma_{\boldsymbol{x}_{f_1,\cdot}} \sigma_{\boldsymbol{x}_{f_2,\cdot}}}, \qquad f_1, f_2 \in \{1,...,F\} \quad (2)$$

Taking the matrix of a spectrum image from 20:00 to 21:00 in July 4th, 2007 as an example, figure 2(a), (b) and (c) show correlation coefficient of time domain with itself, data in previous hour and the same period of previous day, respectively. We can see that the correlations in time domain are quite strong and decrease with time gap of two images increases. Figure 2(d), (e) and (f) represent correlation coefficient of frequency domain with itself, data in previous hour and the same period of previous day, respectively. It is observed that there exist some frequency band blocks with similar features such as band from 25MHz to 27.5MHz. To conclude, spectrum data of different spectrum points in different time periods has an indistinct connection. One of the main advantages of deep learning is its ability to automatically extract features and capture inherent relationships in data. The details of the proposed prediction model using deep learning will be shown in next section.

III. Deep Temporal-Spectral Residual Network

Figure 3 presents the deep temporal-spectral residual network (DTS-Resnet) for HF deep spectrum prediction. It consists of four modules which are minute scale module, hour scale module, day scale module and fusion module. Residual unit and convolution unit are basic units of these modules.

3.1 Basic units

One basic unit is convolution unit, and its calculation is shown in equation (3):

$$\mathbf{X}^{(1)} = f(\mathbf{X}^{(0)} \otimes \mathbf{W}^{(1)} + \mathbf{b}^{(1)}), \quad (3)$$

where $\mathbf{W}^{(1)}$ represent convolution kernel, whose elements are adjustable parameters of the convolution unit as well as $\mathbf{b}^{(1)}$. $\otimes$ means convolution of the input tensor with the convolution kernel while $f(.)$ represents activation function. Relu function $f(x) = \max(0, x)$ is one of commonly used activation functions, which can save the training time of deep learning neural network [22].

Another basic unit is residual unit, whose structure is shown in figure 4 [21]. When the input of the residual unit is denoted as $\mathbf{X}^{(l)}$, its output $\mathbf{X}^{(l+1)}$ is as follows:

$$X^{(l+1)} = X^{(l)} + \mathcal{F}(X^{(l)}) \quad (4)$$

where $F(.)$ is the residual function. It means two repetitive operations of relu activation

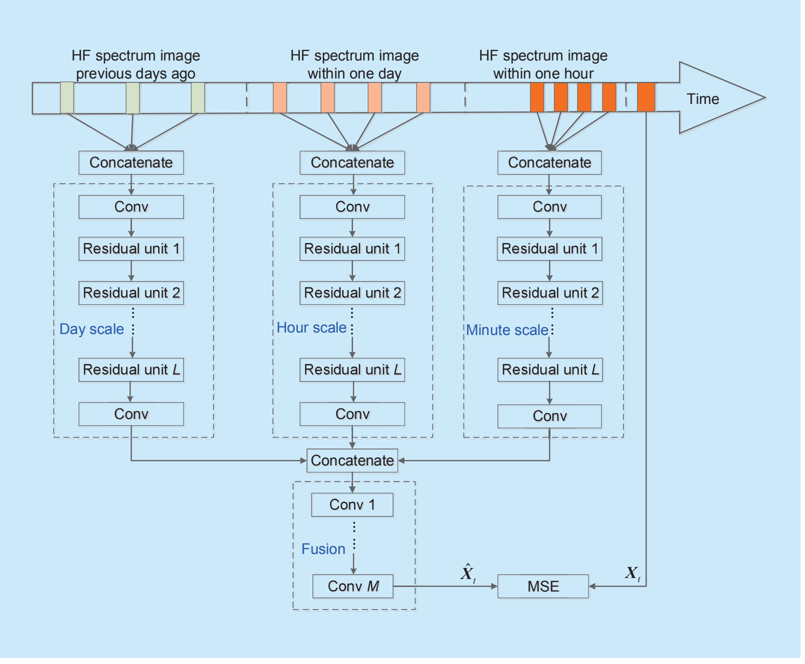


Fig. 3. *The structure of proposed network.*

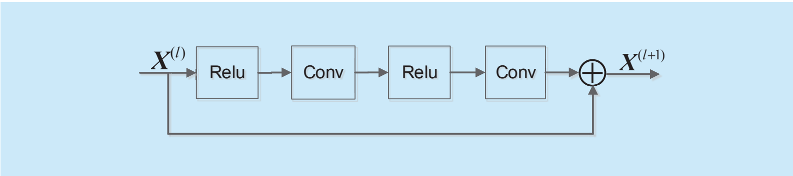


Fig. 4. *The structure of residual unit.*

followed by convolution unit in this paper. The residual unit combines the input with its high-level feature as the output, resulting in better propagation of potential features. It has been demonstrated to be very effective for training super deep neural networks in [23].

3.2 Whole structure of proposed network

As mentioned above, spectrum data in a fixed time period can be constructed as an image with time span as image height and the number of spectrum points as the image width. Suppose that the time span is set to be T' slots and F spectrum points are measured, the spectrum image in time span $[t, t+T')$ is denoted as a tensor $\mathbf{X}_t \in \mathbb{R}^{1\times T' \times F}$. HF spectrum data are collected over time, so all spectrum images can be represented as a set $\mathcal{X} = \{..., \boldsymbol{X}_t, \boldsymbol{X}_{T'+t}, \boldsymbol{X}_{2T'+t}, ...\}$. To conclude, spectrum data are constructed as many one-channel image-like tensors according to the given time span and tensors are arranged in chronological order.

For the tth time interval, all image-like tensors previous to that interval are divided into three categories: i) tensors within one hour; ii) tensors within one day and iii) tensors one day ago. When samples are constructed, the periodicity on 24-hour scale and closeness within several hours are both taken into consideration. Part of tensors in each category are fed into corresponding modules of DTS-Resnet. Three modules of different time scales share the same architecture. They all starts from one convolutional layer, then multiple layers of residual unit are stacked and finally one convolutional layer. The fusion module is composed of multiple convolutional layers and achieves prediction results by fusing outputs of three modules. The design of the proposed network is to capture the regulation of HF spectrum points in different time scales by residual learning. The mean square error between predictions and real values is computed to help the adjustment of network parameters.

Here we take day scale module as an example to detail the composition of time modules. At first l_d tensors in the third tensor category $[\mathbf{X}_{t-l_d*24*60}, \mathbf{X}_{t-(l_d-1)*24*60}, ..., \mathbf{X}_{t-24*60}]$ are concatenated along with the first axis as one tensor $\mathbf{X}_d^{(0)} \in \mathbb{R}^{l_d \times T' \times F}$. This tensor is fed into the first convolutional layer and output tensor is $\mathbf{X}_d^{(1)}$. After residual units of L layers and one convolutional layer, the final output tensor is $\mathbf{X}_d^{(L+2)}$. Similarly, l_m tensors in the first tensor category $[\mathbf{X}_{t-l_m*T'}, \mathbf{X}_{t-(l_m-1)*T'}, ..., \mathbf{X}_{t-T'}]$ are concatenated along with the first axis as one tensor $\mathbf{X}_m^{(0)} \in \mathbb{R}^{l_m \times T' \times F}$. The output of minute scale module is denoted as $\mathbf{X}_m^{(L+2)}$. l_h tensors in the second tensor category $[\mathbf{X}_{t-l_h*60}, \mathbf{X}_{t-(l_h-1)*60}, ..., \mathbf{X}_{t-60}]$ can be transformed to a tensor $\mathbf{X}_h^{(L+2)}$ after the same operation. When T' is no less than one hour, minute scale module will be deleted and only outputs of the other two modules are fed into the fusion module.

Before tensor fusion, the outputs of three modules should be concatenated along with the first axis. The fusion module is only composed of M convolutional layers without any pooling layer. The reason for this design is to guarantee that the shape of output tensor is consistent with that of input tensor. The final output is the prediction results $\hat{\mathbf{X}}_t$ while real values in the tth time interval are denoted as $\mathbf{X}_t$. DTS-Resnet is trained by back propagation algorithm to minimize the prediction error, as equation (5) is shown:

$$\mathcal{L}(w,b) = \arg\min_{w,b} \left\| \hat{\mathbf{X}}_t - \mathbf{X}_t \right\|_2^2, \qquad (6)$$

where w, b mean trainable paremeters in DTS-Resnet. The process of DTS-Resnet for HF deep prediction is listed in Algorithm 1.

IV. EXPERIMENT EVALUATION

4.1 Data preprocessing

The HF spectrum data in the paper comes from the RWTH Aachen University spectrum measurement campaign [24][1] and HF band in this dataset ranges from 20MHz to 30MHz. All data represents power spectral density (PSD) in dBm and is preprocessed using the Min-Max normalization method to scale the data into the range $[-1,1]$. The timestamps for data collection within one minute are not constant as well as data missing in original dataset of this paper, which put limitations on image-like tensor construction. Therefore, time scale in the paper is only detailed to minutes. Deep spectrum prediction on smaller time scale can be investigated if a more complete and better HF spectrum dataset can be achieved in future. For each spectrum point, data in each minute is about 34 measured values and their average is regarded as the final measured spectrum value in that minute. After finishing the construction of the dataset $\mathcal{D}$, eighty percent of the samples is divided into $\mathcal{D}_{train}$ and the remains belong to $\mathcal{D}_{test}$.

4.2 Hyper-parameters setting

For three modules with different time scales, the number of residual units is set to 3. The convolutions in the first and last convolutional layer and residual units all have 64 filters whose size is all 3×3. To keep the shape of input and output tensor of convolution operation consistent, border-mode is used when padding. For the fusion module, three convolutional layers are stacked and the number of filters of size 3×3 is 64, 32 and 1, respectively. Adam algorithm [25] is employed to find the optimal network parameters because its rate of convergence is faster. In this paper, time span is set to be 60 minutes and therefore the minute scale module is omitted. We predict values of multiple spectrum points within one hour based on four image-like tensors. Two are in previous two hours for hour scale module and the other two are in the same time interval of previous two days for day scale module.

4.3 Baseline schemes

The models used to be the baselines are as follows, the input information of which is all the same:

- **SVR**: It is developed from support vector machine, usually used in regression and prediction. The kernel employed here is radial basis function, which is non-linear.
- **CNN**: It is short for convolution neural network, including only multiple convolutional layers. The filters of each convolutional layer except the last layer are 64 and size is 3×3.
- **FC_LSTM**: In this neural network, both modules with day scale and hour scale are replaced by fully connected feedforward network. The fusion module is undertaken

[1] In the original datasets [24], the resolution bandwidth of each individual spectrum band is 200 kHz and the inter-sample time is 1.8 seconds, which results in 48000 samples one day [20].

Algorithm 1. DTS-Resnet algorithm.

Input: Historical observations: $\{\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_m\}$; Number of tensors fed into three modules: l_m, l_h, l_d; Time span: T';
Output: Trained DTS-Resnet model;
// Construct the dataset;
construct image-like spectrum tensors $\mathbf{X}_0, \mathbf{X}_{T'}, ..., \mathbf{X}_{nT'}$ from historical observations
$\mathcal{D} \leftarrow \varnothing$
For all available spectrum tensor t ($T' \le t \le nT'$) **do**
    $\mathcal{S}_m = [\mathbf{X}_{t-l_m*T'}, \mathbf{X}_{t-(l_m-1)*T'}, ..., \mathbf{X}_{t-T'}]$
    $\mathcal{S}_h = [\mathbf{X}_{t-l_h*60}, \mathbf{X}_{t-(l_h-1)*60}, ..., \mathbf{X}_{t-60}]$
    $\mathcal{S}_d = [\mathbf{X}_{t-l_d*24*60}, \mathbf{X}_{t-(l_d-1)*24*60}, ..., \mathbf{X}_{t-24*60}]$
    // $\mathbf{X}_t$ is the target within time $[t, t+T')$;
    put a sample $(\{\mathcal{S}_m, \mathcal{S}_h, \mathcal{S}_d\}, \mathbf{X}_t)$ into $\mathcal{D}$ \\
end for
divide set $\mathcal{D}$ into trainset $\mathcal{D}_{train}$ and testset $\mathcal{D}_{test}$
// Train the model;
initialize all learnable parameters w, b in DTS-ResNet
Repeat
    randomly select a batch of instances from $\mathcal{D}_{train}$
    find w, b by minimizing the objective (5)
Until the training epochs are met
// Test the model;
For each sample in $\mathcal{D}_{test}$ **do**
    fed into the trained DTS-Resnet model
    output the prediction results of that sample
end for

by multiple LSTM layers.

- **DTS-Resnet-nofusion**: It is quite similar to the proposed model except that there is no fusion module and the last convolutional layer has one filter of size 3×3. The output tensors of two modules are merged by element-wise addition.

4.4 Results and discussions

Figure 5 shows the prediction performance of the proposed model and other benchmark models for each spectrum point. Here the root of mean square error (RMSE) is defined as the evaluation metric. It corresponds to the estimation error of the spectrum of the next time slot whose time scale is minute. It can be seen that the prediction error between any two frequency points is very different, which depends on the regularity of the evolution trajectories of spectrum states. The prediction errors for spectrum points around 27.32MHz are much larger than others, which means those spectrum points are more difficult to make prediction. It is found that weak correlation for spectrum points around 27.32MHz results in poor prediction performance while strong correlation for other spectrum points brings about better performance, as shown in figure 2. The DTS-Resnet model performs better than others on majority of spectrum points other than the range from 24MHz to 26MHz where it is a little worse than CNN. The reason may be the difference of initialization or the spectrum points features. Figure 6 shows the curve of cumulative distribution function obtained from figure 5, which also illustrates that the proposed model has absolute advantage in a statistical sense. For about ninety percent of spectrum points, the RMSE of prediction is less than 0.29. Compared with DTS-Resnet-nofusion model, we can find that the fusion module of DTS-Resnet plays an important role for deep HF spectrum prediction.

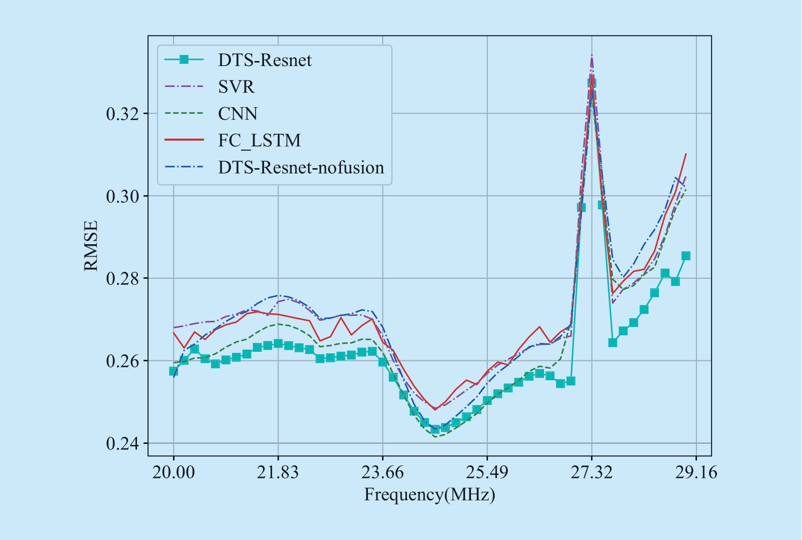


Fig. 5. *The RMSE of prediction for each spectrum point.*

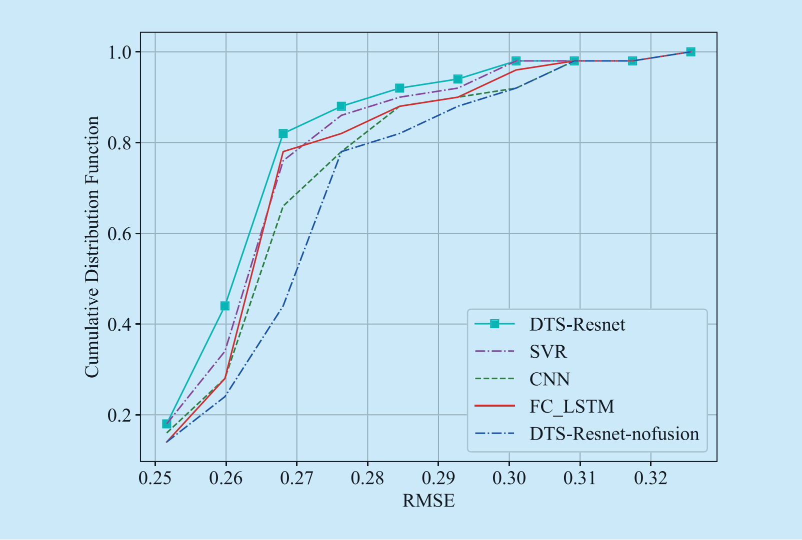


Fig. 6. *The cumulative distribution function of RMSE for all spectrum points.*

V. Conclusions

The prediction of HF spectrum has always been a tricky but urgent problem because spectrum data in this frequency band have a non-explicit complex relationship. In this paper, we explore an advanced and intelligent deep learning model to predict PSD values in given time period at the same time. The dataset is reasonably constructed based on the implied periodicity and closeness of the HF spectrum points. The proposed model, named DTS-Resnet, combines residual units and convolutional layers to perform different functions. The experimental results demonstrate that DTS-Resnet has better prediction performance than other prediction models. In spite of this, the prediction only considers mining the features from measured spectrum values without involving the factors such as time, geographical location, weather and sunspot activities. This is one research direction of future HF spectrum prediction. Moreover, spectrum prediction is such an example which is proposed for dynamic spectrum access in cognitive radio networks (CRN) and then is applied to HF communication. Just like Koski and Furman proposed in [26], studies on power and energy transfer [27] or spectrum sharing strategy [28, 29, 30] in CRN can also be transferred into HF communication, which is regarded as another research direction.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (Grants No. 61501510 and No. 61631020), Natural Science Foundation of Jiangsu Province (Grant No. BK20150717), China Postdoctoral Science Foundation Funded Project (Grant No. 2016M590398 and No.2018T110426), Jiangsu Planned Projects for Postdoctoral Research Funds (Grant No. 1501009A), and Natural Science Foundation for Distinguished Young Scholars of Jiangsu Province (Grant No. BK20160034).

References

[1] J. Goodman, *HF Communications: Science and Technology*, Van Nostrand Reinhold, 1991.

[2] N. Maslin, *HF Communications: A Systems Approach*, Pitman, 1987.

[3] E. Johnson, E. Koski, W. Furman, M. Jorgenson, and J. Nieto, *Third-generation and Wideband HF Radio Communications*, Artech House, 2012.

[4] B. Zolesi, A. Belehaki, I. Tsagouri, and L. R. Cander, "Real-time updating of the simplified ionospheric regional model for operational applications," *Radio Science*, vol. 39, no. 2, 2004, pp. 1-12.

[5] W. Zeng and X. Zhang, "Predictions of HF communication MUF in the region of south China sea," *IEEE Antennas and Propagation Magazine*, vol. 41, no. 4, 1999, pp. 35-38.

[6] Y. Tu, Y. Lin, J. Wang, and Jeong-Uk Kim, "Semi-supervised learning with generative adversarial networks on digital signal modulation classification," *Computers Materials & Continua (CMC)*, vol. 55, no. 2, 2018, pp. 243-254.

[7] A. Ioannidou, E. Chatzilari, S. Nikolopoulos, and I. Kompatsiaris, "Deep learning advances in computer vision with 3D data: A survey," *ACM Computing Surveys (CSUR)*, vol. 50, no. 2, 2017, pp. 38.

[8] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu, "Advanced deep-learning techniques for salient and category-specific object detection: A survey," *IEEE Signal Processing Magazine*, vol. 35, no. 1, 2018, pp. 84-100.

[9] B. Tang, Y. Tu, Z. Zhang, and Y. Lin, "Digital signal modulation classification with data augmentation using generative adversarial nets in cognitive radio networks," *IEEE Access*, vol. 6, 2018, pp. 15713-15722.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, 2015, pp. 436-444.

[11] G. Ding, Y. Jiao, J. Wang, Y. Zou, Q. Wu, Y. D. Yao, and L. Hanzo, "Spectrum inference in cognitive radio networks: Algorithms and applications," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 1, 2018, pp. 150-182.

[12] X. Xing, T. Jing, Y. Huo, H. Li, and X. Cheng, "Channel quality prediction based on Bayesian inference in cognitive radio networks," *Proc. IEEE INFOCOM*, 2013, pp. 1465-1473.

[13] P. Huang, C. J. Liu, X. Yang, L. Xiao, and J. Chen, "Wireless spectrum occupancy prediction based on partial periodic pattern mining," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 7, 2014, pp. 1925-1934.

[14] V. K. Tumuluru, P. Wang, and D. Niyato, "Channel status prediction for cognitive radio networks," *Wireless Communications and Mobile Computing*, vol. 12, no. 10, 2012, pp. 862-874.

[15] C. J. Yu, Y. Y. He, and T. F. Quan, "Frequency spectrum prediction method based on EMD and SVR," *Proc. 8th International Conference on Intelligent Systems Design and Applications*, 2008, pp. 39-44.

[16] L. Yu, Q. Wang, Y. Guo, and P. Li, "Spectrum availability prediction in cognitive aerospace communications: A deep learning perspective" *Proc. IEEE Cognitive Communications for Aerospace Applications Workshop (CCAA)*, 2017, pp. 1-4.

[17] Z. Yan, L. Zhang, T. Rahman, and D. Su, "Prediction of the HF ionospheric channel stability based on the modified ITS model," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 6, 2016, pp. 3321-3333.

[18] L. Melián-Gutiérrez, S. Zazo, J.L. Blanco-Murillo, I. Pérez-Álvarez, A. García-Rodríguez, and B. Pérez-Díaz, "HF spectrum activity prediction model based on HMM for cognitive radio applications," *Physical Communication*, vol. 9, 2013, pp. 199-211.

[19] H. Haralambous and H. Papadopoulos, "24-hour neural network congestion models for high-frequency broadcast users," *IEEE Transactions on Broadcasting*, vol. 55, no. 1, 2009, pp. 145-154.

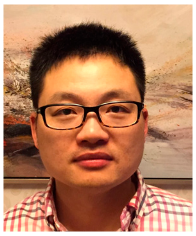
[20] G. Ding, F. Wu, Q. Wu, S. Tang, F. Song, A. V. Vasilakos, and T. A. Tsiftsis, "Robust online spec-

trum prediction with incomplete and corrupted historical observations," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, 2017, pp. 8022-8036.
[21] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," *Proc. AAAI*, 2017.
[22] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Proc. 25th International Conference on Neural Information Processing Systems*, 2012, pp. 1097-1105.
[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. Computer Vision and Pattern Recognition*, 2016, pp. 770-778.
[24] M. Wellens, "Empirical modelling of spectrum use and evaluation of adaptive spectrum sensing in dynamic spectrum access networks," *Ph.D.Dissertation*, RWTH Aachen University, May. 2010.
[25] D. P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint*, arXiv:1412.6980, 2014.
[26] E. Koski and W. Furman, "Applying cognitive radio concepts to HF communications," *Proc. 11th International Conference on Ionospheric radio Systems and Techniques (IRST 2009)*, 2009, pp. 1-6.
[27] C. Zhong, H. Suraweera, G. Zheng, I. Krikidis, and Z. Zhang, "Wireless information and power transfer with full duplex relaying," *IEEE Trans. Commun.*, vol. 62, no. 10, 2014, pp. 3447-3461.
[28] Y. Chen, G. Yu, Z. Zhang, H. Chen, and P. Qiu, "On cognitive radio networks with opportunistic power control strategies in fading channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, 2008, pp. 2752-2761.
[29] Z. Feng, C. Qiu, Z. Feng, Z. Wei, W. Li, and P. Zhang, "An effective approach to 5G: Wireless network virtualization," *IEEE Commun. Magazine*, vol. 53, no. 12, 2015, pp. 53-59.
[30] G. Zhang, W. Heng, T. Liang, C. Meng, and J. Hu, "A novel two-stage dynamic spectrum sharing scheme in cognitive radio networks," *China Communications*, vol. 13, no. 6, 2016, pp. 236-248.

Biographies

Ling Yu, received her B.S. degree in information engineering from Southeast University, Nanjing, China, in 2016. She is currently pursuing her M.S. degree at College of Communications Engineering, Army Engineering University of PLA, Nanjing, China. Her research interests include deep learning, wireless communications and cognitive radio networks.

Jin Chen, received the B.S. degree in wireless communication, the M.S. degree and Ph.D. degree in communications engineering and information system from the Institute of Communications Engineering, Nanjing, China, in 1993, 1996, and 1999, respectively. She is currently a professor in the Institute of Communications Engineering, Army Engineering University of PLA, Nanjing, China. Her current research interests include wireless communications networks and cognitive radio Networks.



Yuming Zhang, received the B.S. degree in wireless communication, the M.S. degree and Ph.D. degree in communication and information systems from the Institute of Communications Engineering, Nanjing, China, in 2002, 2005, 2008, respectively. He is currently an associate professor in the Institute of Communications Engineering, Army Engineering University of PLA, Nanjing, China. His current research interests include signal processing and digital communications.

Huaji Zhou, received his M.S. degree in electrical engineering from Xidian University, Xi'an, China, in 2013. Since 2013, he has been an engineer in 36th Research Institute of China Electronic Technology Group Corporation. Since 2016, he has been a member of Academician Yang Xiaoniu's Innovation team. He has hosted or participated in several Defense projects of the science and Technology Commission of Military Commission. His research interests include cognitive electronic warfare, machine learning, and cognitive radio networks.

Jiachen Sun, received her B.S. degree in information engineering from Southeast University, Nanjing, China, in 2016. She is currently pursuing her M. S. degree at College of Communications Engineering, Army Engineering University of PLA, Nanjing, China. Her research interests include data analytics, wireless communications and cognitive radio networks.