第 19 卷 第 2 期 太赫兹科学与电子信息学报 Vo1.19，No.2
2021 年 4 月 Journal of Terahertz Science and Electronic Information Technology Apr.，2021

文章编号：2095-4980(2021)02-0235-09

A novel efficient automatic modulation classification algorithm using deep LSTM aided convolutional networks

WU Nan，GU Wanbo，WANG Xudong

(School of Information Science and Technology，Dalian Maritime University，Dalian Liaoning 116023，China)

Abstract：Automatic modulation classifications would play an essential part in wireless spectrum anomaly detection and radio environment awareness. With the breakthrough in deep learning algorithms, this issue can be solved with unprecedented precision and effectiveness by using neural networks. Therefore, a novel neural network termed as Long short-term Convolutional Deep Neural Network(LCDNN) is proposed, which creatively combines the complimentary merits of Long Short-Term Memory(LSTM), Convolutional Neural Network(CNN) and deep network architectures. This model directly learns from raw time domain amplitude and phase samples in training dataset without requiring human engineered features. Simulation results show that the proposed model yields a classification accuracy of 93.5% at high SNRs. Further, the noise sensitivity of the proposed LCDNN model is examined and it is showed that LCDNN can outperform existing baseline models across a range of SNRs. Finally, in order to reduce the computational complexity of the LCDNN model, a 'compact' LCDNN model is proposed, which achieves the state-of-the-art classification performance with only 0.6% parameters of the original LCDNN model.

Keywords：modulation classification；deep learning；long short-term Convolutional Deep Neural Network；convolutional neural network

CLC number：TN975 **Document code**：A **doi**：10.11805/TKYDA2020034

1 Introduction

Modulation Classification(MC) is a technique to recognize the modulation type of wireless signals without prior information at the receiver. For example, for military communications, desired signals should be transmitted securely and received rightly, whereas potentially unwanted signals must be identified and jammed. Naturally, MC is the key enabling technology in this realm. On the other hand, for commercial communications, especially for Software Defined Radio(SDR), where coping with a large variety of communication requirements are essential, blind and fast recognition of the types of the received signals is the building block of the intelligent system[1]. Hence, MC usually happens before the signal demodulations.

Traditionally, Likelihood-based(LB) and Feature-Based(FB) approaches are the main modulation recognition algorithms in the literatures [2–6]. The former approaches compare the likelihood ratio of each possible hypothesis against a threshold, which is derived from the probability density function of the observed signal. The LB methods can approach the optimal solution in theory with the drawback of excessive computational complexity. On the other hand, the FB approaches are designed to extract signal features to facilitate modulation classifications. Although FB approaches perform well in specific environments, they rely heavily on experts' knowledge. Besides, FB methods cannot be generalized to other signal settings. In order to solve these problems, a number of Machine Learning(ML) algorithms were introduced in the literatures [7–12]. Despite some performance gains having been achieved, these algorithms remain relying on expert experiences.

Only recently, deep learning techniques applied to image processing[13–14] and voice recognition[15] have achieved ground breaking results, where features are learned from raw data. In the field of MC, the authors in [16] used Convolutional Neural Networks (CNNs) at the receiver as learned matched filters of the transmitted signals, where no expert feature

Received date：2020-01-17；**Revised date**：2020-05-26

About the author：WU Nan(1979–)，Male, Doctor of Engineering., Associate Professor. Current research involves communication systems based on deep learning and visible-light. email:wu.nan@dlmu.edu.cn

extraction was needed. Its remarkable results demonstrated that blind temporal learning on large and densely encoded time series using CNNs was viable. Unfortunately, CNNs in [16] have difficulty to distinguish Quadratic Amplitude Modulations(QAM), such as 16-QAM and 64-QAM. Since then, the quest of learning more intricate signal features from raw data drove researchers to explore better deep neural network architectures[17]. For example, Oxford's Visual Geometry Group(VGG) proposed VGG models in [18], which were soon outperformed by the milestone Residual Neural Network (ResNet) models [19–20] and Densely Connected Convolutional Networks(DenseNet)[21–22]. However, these convolution-based neural networks fail to explore the inherited connections between adjacent time domain signals.

Furthermore, Recurrent Neural Networks(RNNs) have been proved to be powerful tools for time domain data processing tasks due to their unparallel ability to connect previous state information to the present. Specifically, Long Short-Term Memory(LSTM)[23] network is a RNN architecture designed to be excelled at storing and accessing information from time series signals. Naturally, RNNs can be employed to facilitate MC tasks. For example, authors in [24] introduced an LSTM model using samples in polar form as input. And [25] took advantage of the complementarity of CNNs, LSTMs and Deep Neural Networks(DNNs) by combining them into a unified architecture termed as CLDNN. Meanwhile, [22] proposed a deep ResNet model with excessive layers, while fine tuning LSTM's parameters. Simulation results were provided[22] in order to demonstrate that both of these fine-tuned models achieved high classification accuracy in the high SNR region, compared with a simple CNN approach. However, both LSTM and CLDNN networks require a large number of trainable parameters, which lead to high network computational complexity.

Another branch of revolutionary deep learning algorithm, namely Generative Adversarial Networks(GANs), can also be used in MC for data argumentation[26–28]. For example, [26] proposed an Auxiliary Classifier Generative Adversarial Network(ACGAN) and transformed the GAN augmented data into colored constellation diagram, based on which the classification decisions were made.

In this paper, aiming for improving deep learning based MC model's accuracy, sensitivity and computational complexity, a novel neural network termed as Long short-term Convolutional Deep Neural Network(LCDNN) is proposed, which creatively capitalizes on the complimentary power of LSTM, CNN and DNN. Firstly, a group of LSTM layers are utilized to extract time-domain relationship of the raw sampled data. Then, specially designed CNNs are employed to learn the intricate signal features from the LSTM-extracted data, followed by DNNs to enable the classification task. It will be demonstrated that the time-domain relationship of the raw data holds the key to distinguish various types of modulation signals.

2 Problem formulation

In this section, after formulating the MC problem, the open-source dataset used for MC in this paper is introduced.

2.1 Mathematical model

The MC task can be considered as an N-class decision problem, where the received signal can be written as complex base-band formats as [16].

$$r(t) = s(t;\psi_i) + n(t) \tag{1}$$

where

$$s(t;\psi_i) = A_i \mathrm{e}^{\mathrm{j}\theta(t)} \sum_k^K \mathrm{e}^{\mathrm{j}\varphi_k} x_k^{(i)} g(t-(k-1)T-\varepsilon T) \tag{2}$$

$n(t)$ is the Additive White Gaussian Noise(AWGN) term. More explicitly, $\theta(t) = 2\pi\Delta f t + \theta$, where Δf is the carrier frequency offset and θ is the constant carrier phase propagation delay. φ_k is the phase jitter. $\left\{x_k^{(i)}, 1 \leqslant k \leqslant K\right\}$ represents number of complex symbols extracted from the i^{th} modulation type. $g(t)$ is the result of the convolution of $h(t)$ and $p(t)$, where $h(t)$ is the channel impulse response and $p(t)$ is the transmit pulse shaping. T stands for the symbol duration and ε is the normalized epoch for time offset between the transmitter and receiver. $\psi_i = \left\{A_i, \theta(t), x_k^{(i)}, \varphi_k, \varepsilon\right\}$ is the multidimensional vector that includes the deterministic unknown signal or channel parameters for the i^{th} modulation. Therefore, the MC task can be stated as: estimate the modulation class of $s(t)$, when given $r(t)$ of equation (1).

2.2 Dataset

The synthetic dataset named RadioML2016.10b generated using GNU Radio[29] is used in this paper, which includes two analog and eight digital modulations. For analog modulations, a continuous voice signal is used as input data, which consists primarily of acoustic voice speech with some interludes and off times. For digital modulations, the entire Gutenberg works of Shakespeare in ASCII is used, with whitening randomizers applied to ensure equiprobable symbols and bits.

The GNU Radio Dynamic Channel Model hierarchical block[29] is used for channel simulation, which includes a number of channel effects, such as center frequency offset, sample rate offset, AWGN, multi-path, and fading, as detailed in Table1. In the stage of packaging data, the output stream of each simulation is randomly segmented into vectors with a sample rate of 1M samples per second. Similar to the windowing method of voice signal in speech recognition task, 128 samples are extracted by a sliding window and 64 samples are shifted to form the dataset. The detailed modulation and channel parameters are listed in Table1.

2.3 Signal representations

Radio domain operations are typically in complex base-band representations, which are not currently well supported in existing machine learning frameworks such as Keras and Tensorflow. Therefore, it is necessary to transform the signals into a new representation that is suitable for deep learning. This paper considers two simple data representations. One is the real-valued equivalent of the raw complex data, which is the default form in RadioML2016.10b. The other is the polar signal format. Besides, different signal formats affect the achievable classification accuracy[30].

2.3.1 IQ format

The IQ format is the mapping of the complex base-band $r(t)$ of equation (1) into two sets of real-valued data, one represents the in-phase samples and the other represents the quadrature component, as follows:

$$f: C^N \to \mathbb{R}^{2\times N} \tag{3}$$

$$r(t) \to r^{\text{polar}} = \begin{bmatrix} r_i^{\mathrm{T}} \\ r_q^{\mathrm{T}} \end{bmatrix} \tag{4}$$

2.3.2 Polar format

The polar format is the mapping from complex base-band $r(t)$ of equation (1) into two real-valued vectors, one carries its magnitude and the other represents its phase. Furthermore, the magnitude signal is processed using L2 normalization, and the phase in radians is normalized to −1 and 1.

Table1 Parameters of dataset RadioML2016.10b

	parameter	value
signal	modulations	8PSK, BPSK, CPFSK, GFSK, QPSK, PAM4, QAM16, QAM64, WBFM, AM-DSB
	sample length	128
	samples per symbol	8
	SNR range/dB	-20 to 18
	samples per SNR per modulation	6 000
channel	sampling frequency/kHz	200
	maximum sampling rate offset/Hz	50
	maximum carrier frequency offset/Hz	500
	fading model	Rician
	Rician K factor	4
	fractional sample delays	[0,0.9,1.7]
	magnitudes corresponding to each delay time	[1.0,0.8,0.3]

$$f: C^N \to \mathbb{R}^{2\times N} \tag{5}$$

$$r(t) \to r^{\text{polar}} = \begin{bmatrix} r_A^{\mathrm{T}} \\ r_\phi^{\mathrm{T}} \end{bmatrix} \tag{6}$$

3 Deep neural network architecture

In this section, the design philosophy and the performance of four milestone deep learning architectures used for MC are reviewed firstly, which will be used as the baseline models. Then, the novel LCDNN architecture is proposed.

3.1 CNNs

Convolutional neural networks are simply neural networks that use convolution in place of general matrix multiplications[31]. Convolution layers are the essential part of CNNs that carry out feature extractions. Parameters sharing and sparse connectivity are the two important characteristics of the convolutional layers. CNNs achieve sparse connectivity between input and output units by making the kernel size smaller than the input size, which also improves the statistical efficiency and reduces the computational memory requirements.

Authors in [16] proposed a CNN architecture for MC with four convolution layers as shown in Fig.1. We re-implement this architecture and find that it achieves 87.5% MC accuracy on the dataset of RadioML2016.10b at high SNR.

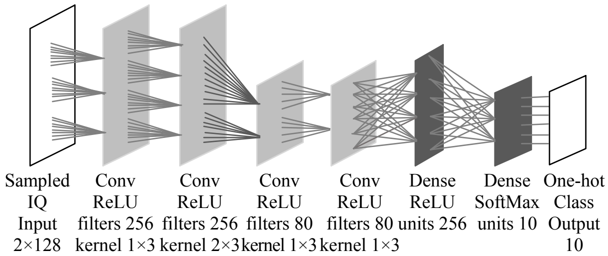

Fig.1 CNN architecture of [21] used for MC

3.2 ResNet

The deep Residual Network(ResNet) architecture was introduced in ImageNet and COCO 2015 competitions[19] to address the problem of building deeper networks to learn more complex functions and hierarchical feature relationship.

Traditional convolution feed-forward networks connect x_l , the output of the l^{th} layer, as the input to the next layer, which gives rise to the following layer $x_l = H(x_{l-1})$ where $H(\bullet)$ denotes a composite function of operations. However, ResNet adds a skip-connection that bypasses the non-linear transformations with an identity function. Thus, the identity function and the output of H_l are combined by summation, $x_l = H(x_{l-1}) + x_{l-1}$. The advantage of ResNet is that the gradients can flow directly through the identity function from later layers to the earlier layers[21].

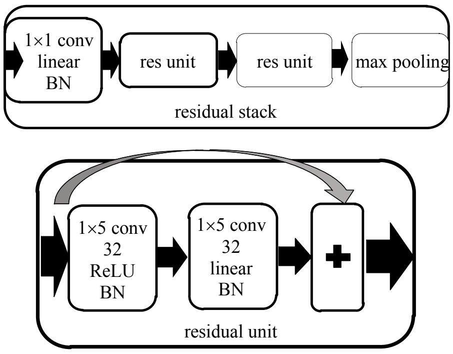

Fig.2 Hierarchical layers used in ResNet

Table2 ResNet architecture

layer	output shape
input/ IQ	2×128
residual stack×5	32×4
FC/SeLU	128
FC/SeLU	128
FC/SoftMax	10

Inspired by [20], a ResNet model for MC was proposed in [22], where five residual stacks were followed by three fully connected layers. Meanwhile, the network takes the signals in IQ representation of Equation (4) as input. As seen in Fig.2, each residual stack consists of one convolutional layer, two residual units, and one max-pooling layer. Each convolutional layer is followed by a batch normalization layer to prevent overfitting. For each residual unit, a shortcut connection is created by adding the input of the residual unit with the output of the second convolutional layer of the residual unit. Each convolutional layer in the residual unit uses a filter size of 1×5. The network employs self-normalizing neural networks[32] in the Fully Connected(FC) layers and uses Scaled exponential Linear Unit(SeLU) activation functions with mean-response scaled initialization and Alpha Dropout. The residual unit and stack of residual units used are shown in Fig.2, while the whole network architecture is shown in Table2. This ResNet architecture of Table2 on the dataset of RadioML2016.10b is tested and a classification accuracy of 92% is achieved at high SNR.

3.3 Long Short-Term Memory(LSTM)

The LSTM[23] architecture consists of a set of recurrently connected subnets, known as memory blocks. Each block contains one or more self-connected memory cells and three multiplicative units, namely the input, output and forget gates, that provide continuous analogues of write, read and reset operations for the memory cells[33]. Based on the previous state and current input data, the cells can learn the gate weights through training. The outputs from the current cell to the rest of the network emanate from the output gate multiplications. The gating mechanism helps LSTM to store history valuable information over long period so as to support persistent feature learning.

A LSTM model[22] used for modulation classification is shown in Fig.3. The first two layers are comprised of LSTM with 128 units. The output from the second LSTM layer, namely a vector of dimension 128, is fed to the following dense layers. The final dense layer with SoftMax activations maps the learned features to one of the modulation classes. The intuition to use LSTM for MC is based on the fact that different modulation schemes exhibit different amplitude and phase characteristics and the LSTM model can learn these temporal dependencies effectively[24]. This model of Fig.3 is capable of achieving a classification accuracy of 92% on the dataset of RadioML2016.10b at high SNR.

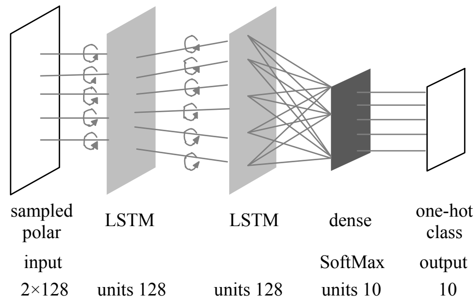

Fig.3 LSTM architecture of [21] used for MC

3.4 CLDNN

Since CNNs, LSTMs and DNNs are individually excelled in different capacities, the authors in [25] took advantage of the complementarity of their abilities by combining them into a unified architecture, termed as CLDNN. The CLDNN model firstly feeds input features, surrounded by temporal context, into a few CNN layers to reduce spectral variation. Then, the output of the CNN layers is fed into LSTM layers to reduce temporal variations. Finally, the output of the last LSTM layer is fed to a number of fully connected DNN layers, which transform the extracted features into a high dimensional space that makes that output easier to classify.

As plotted in Fig.4, the CLDNN model[22] is re-implemented and its parameters are further tuned, which has four one-dimensional convolutional (Conv1D) layers with the kernel size of 3 and filter size of 256. In addition, the polar form signal representations of Equation (6) are used as input for the model for better accuracy. This model achieves 89% accuracy at high SNR region and gets significant improvement at low SNR region than the original CLDNN model in [22], when testing on the dataset of RadioML2016.10b.

3.5 The proposed LCDNN architecture

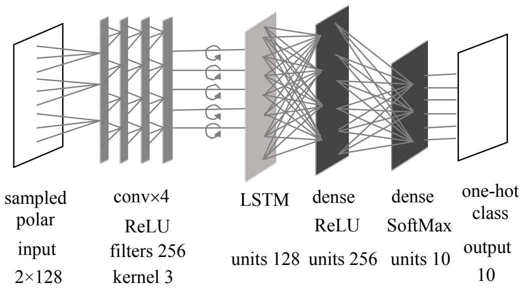

Fig.4 CLDNN architecture used for MC

Although the CLDNN of Fig.4 has the advantage of fusing different types of neural networks, the middle LSTM layers fail to fully explore the connections between the adjacent time domain signals, because the previous convolutional layers' locality connections potentially break signal dependency.

Therefore, we propose a Long short-term Convolutional Deep Neural Network (LCDNN) to capitalize further on signal's temporal dependency. The resultant network structure is shown in Fig.5. The idea is to use the front LSTM layer of Fig.5 to process the input signals, one time step at a time, to extract the feature sequences from the input vector[34]. Then, the following two LSTM layers are essentially an RNN language model except that they are conditioned on the input sequence. Both of them have 128 units. The convolutional layers of Fig.5 consist of 256 stacked filters having kernel size of 3, which perform one-dimensional convolutions on the LSTM processed signal features. Rather than using batch normalization at each CNN layer, the dropout with $p = 0.6$ is utilized to reduce overfitting. The outputs of the convolutional layers are followed by fully connected layers having 256 neurons. The output layer has 10 neurons representing ten modulation types. In addition, the LCDNN is fed with signals in polar format of Equation (6). It will be demonstrated in the next section that the proposed network is capable of outperforming the above-mentioned milestone networks in terms of classification accuracy.

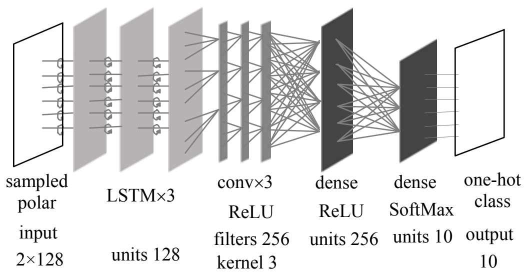

Fig.5 Proposed LCDNN architecture used for MC

4 Simulation results and discussion

In this section, the highest achievable classification accuracy is first pursued by using the proposed LCDNN model, when compared with the baseline models. Furthermore, the LCDNN's noise sensitivity measured by the performance at low SNR range is discussed. Later, reducing LCDNN's computational complexity is highlighted without significantly sacrificing the achievable performance. In addition, all the source codes in this paper will be available publicly upon publication.

4.1 Classification accuracy

In order to evaluate the performance of the proposed LCDNN classifier against the milestone models of Section 3, all the above-mentioned models are tested on dataset RadioML2016-10b.

The dataset is divided into training, validation and test subsets on a scale of 0.375:0.125:0.5. Furthermore, categorical cross-entropy is used as the loss function, while Adam optimizer is employed with a default learning rate $\alpha = 0.001$.

In Fig.6, the aggregated modulation classification accuracies of 10 modulations against SNRs using CNN, ResNet, LSTM, CLDNN and the proposed LCDNN are plotted, in addition the number of trainable parameters of each model is also given. When SNR exceeds 0 dB, our LCDNN architecture can achieve an accuracy of 93.5%, where the ResNet and LSTM counterparts have an accuracy of 92%. More importantly, when operating at low SNRs, especially between 0 dB and −10 dB, our LCDNN scheme demonstrates an improved capacity in signal recognition under noisy environment, compared with other models. This phenomenon further proves that time dependency features are the key to recognize wireless signals. On the other hand, classification using short samples is necessary, because the decision processes have to function in real time. This is the case in many real-world systems, when dealing with short observations. With short observation examples under various channel impairments, it will not be expected that 100% classification rates can be achieved on the full dataset[19]. Therefore, LCDNN's 93.5% accuracy is a non-trivial achievement.

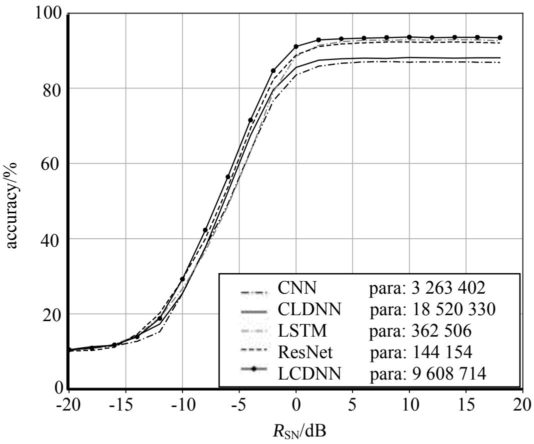

Fig.6 Total modulation classification accuracy against SNR of CNN, ResNet, LSTM, CLDNN and the proposed LCDNN network of Fig.5 using the dataset RadioML2016.10b

4.2 Noise sensitivity

The LCDNN model is trained by using data from all the SNRs under various channel conditions, because the proposed classifier is intended to work with any SNR, rather than a single SNR point. That means the learned features have to not only accommodate various channel impairments of Table1, but also to overlook the impact of noise. However, if the noise power is sufficiently large, the random noise could overshadow the signal's features, hence prevent the LCDNN from exacting useful feature from raw data. In other words, noise's feature would dominate the learned signal representations.

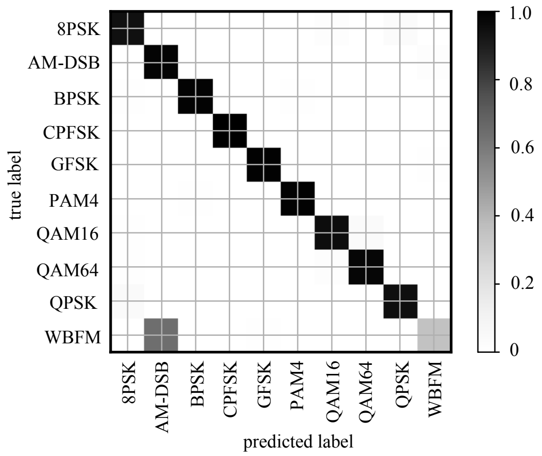

Fig.7 Confusion matrix of the LCDNN model at R_{SN} =0 dB

Although Fig.6 has demonstrated LCDNN's superior performance across the whole SNR range, compared with the baseline models, Fig.7 explicitly plots the confusion matrix using the proposed LCDNN at R_{SN} = 0 dB. Note that all the modulation types are identified almost perfectly, except for Wide-Band Frequency Modulation(WBFM) signals. Fig.7 further shows that the LCDNN mistakes WBFM for AM−DSB signals. That is because the source analog voice signals in dataset RadioML2016.10b contain periods of silence, where only a carrier tone is present[16]. Hence, the same carrier signal makes WBFM and AM−DSB indiscernible.

More importantly, Fig.7 shows that even when the noise power is significant at 0 dB, the proposed LCDNN classifier remains capable of looking through the thick 'curtain' of noise and learning the signal representations.

4.3 Model complexity and improvement

Although the proposed LCDNN model of Fig.5 achieves the best classification accuracy as seen in Fig.6, its computational complexity is excessive, measured by the number of trainable parameters. More explicitly, LCDNN of Fig.5 has around 9.6M trainable parameters, whereas the ResNet model of Table2 only uses 0.144M parameters, as detailed in Fig.6.

In order to reduce the complexity of the proposed LCDNN model, the network architecture is further simplified by reducing the number of layers as well as units of LSTM, the number of layers and filters of convolutional layers and the number of layers of dense layer. The resultant ‘compact’ LCDNN is shown in Fig.8, which consists of two LSTM layers with units of 32, one Conv1D layer with filters of 32 and only one dense layer with 10 neurons. In comparison, the LCDNN model of Fig.5 has three LSTM layers with units of 128, three CNN layers with filters of 256 and two dense layers having 256 units.

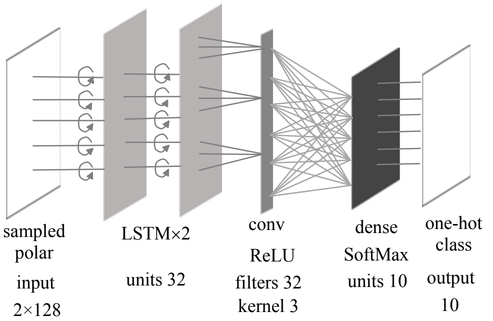

Fig.8 Proposed 'compact' LCDNN mode

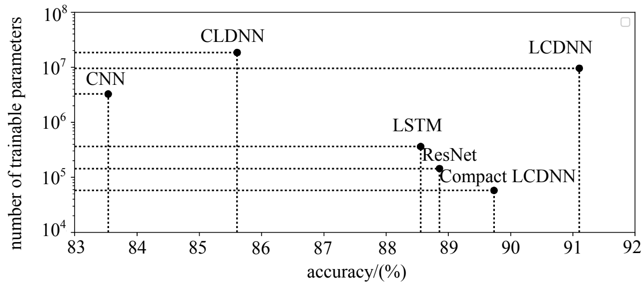

Fig.9 Number of trainable parameters of all the models, when comparing classification accuracy at R_{SN} =0 dB

Fig.9 explicitly plots the number of trainable parameters of LCDNN, compact LCDNN and all the baseline models with the corresponding classification performance at SNR of 0 dB. In Fig.9, it is observed that the compact LCDNN achieves an accuracy of 89.7%, which is slightly worse than 91% achieved by the LCDNN of Fig.5. However, the compact LCDNN still manages to outperform ResNet, LSTM, CLDNN and CNN models. On the other hand, the number of trainable parameters of the compact LCDNN is 57 770, which is only 0.6% of the original LCDNN model and 40.1% of the ResNet model. Furthermore, it can be seen in Fig.10 that the compact LCDNN model’s MC performance beats the baseline ResNet model through the whole SNR range with a much lower computational complexity. This confirms that the classification power of the LCDNN model mainly comes from its LSTM extracted CNN aided feature learning architecture.

5 Conclusion

Since deep learning has demonstrated enormous potential in improving radio signal identification’s accuracy and sensitivity, in this paper, a novel LCDNN network is proposed that further capitalizes on the power of LSTM, CNN and deep neural networks to explore the temporal relationship of wireless signals. In terms of classification accuracy, the proposed LCDNN is shown to achieve 93.5% accuracy in the high SNR region, which outperforms existing CNN, LSTM, ResNet and CLDNN baseline models under the same conditions. Secondly, the proposed LCDNN’s classification performance under noise also beat baseline models, which means that our model is more resilient to noise corruption. Next, a compact LCDNN is proposed by using 0.6% trainable parameters of the original LCDNN model, which processes the lowest computational complexity, while still achieving a slightly better classification accuracy than existing baseline models. Finally, it is believed that deep learning oriented classifiers will evolve to become the main ingredient of smart spectrum monitoring system. Since the proposed LCDNN offers high classification accuracy, better insensitivity to noise and potentially low computational complexity, it could play an important role in realizing this vision.

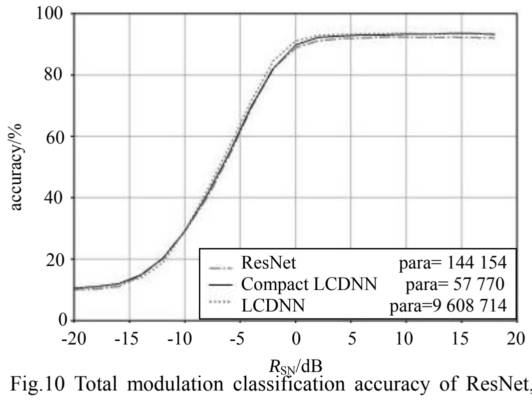

Fig.10 Total modulation classification accuracy of ResNet, the proposed LCDNN and compact LCDNN network using the dataset of RadioML2016.10b

Reference:

[1] DOBRE O A,ABDI A,BAR-NESS Y,et al. Survey of automatic modulation classification techniques:classical approaches and new trends[J]. IET Communications, 2007,1(2):137-156.

[2] LI X,DONG F,ZHANG S,et al. A survey on deep learning techniques in wireless signal recognition[J]. Wireless Communications & Mobile Computing, 2019:1-12.

[3] CHAVALI V G, SILVA C R C M D. Maximum-likelihood classification of digital amplitude-phase modulated signals in flat fading non-Gaussian channels[J]. IEEE Transactions on Communications, 2011,59(8):2051-2056.

[4] BROWN R A,LAUZON M L,FRAYNE R. A general description of linear time-frequency transforms and formulation of a fast, invertible transform that samples the continuous s-transform spectrum nonredundantly[J]. IEEE Transactions on Signal Processing, 2010,58(1):281-290.

[5] AZZOUZ E E,NANDI A K. Procedure for automatic recognition of analogue and digital modulations[J]. IEE Proceedings - Communications, 1996,143(5):259-266.

[6] DOBRE O A,HAMEED F. Likelihood-based algorithms for linear digital modulation classification in fading channels[C]// 2006 Canadian Conference on Electrical and Computer Engineering. Ottawa,Ontario,Canada:IEEE, 2006.

[7] MULLER F C B F,CARDOSO C,KLAUTAU A. A front end for discriminative learning in automatic modulation classification[J]. IEEE Communications Letters, 2011,15(4):443-445.

[8] ABDELMUTALAB A,ASSALEH K,EL-TARHUNI M. Automatic modulation classification using polynomial classifiers[C]// IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC). Washington, DC,USA:IEEE, 2014:806-810.

[9] AUBRY A,BAZZONI A,CAROTENUTO V,et al. Cumulants-based radar specific emitter identification[C]// IEEE International Workshop on Information Forensics and Security. Iguacu Falls,Brazil:IEEE, 2011:1-6.

[10] KIM K,AKBAR I A,BAE K K,et al. Cyclostationary approaches to signal detection and classification in cognitive radio[C]// 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks. Dublin,Ireland:IEEE, 2007:212-215.

[11] DOBRE O A,BAR-NESS Y,WEI S. Higher-order cyclic cumulants for high order modulation classification[C]// IEEE Military Communications Conference. Boston,MA,USA:IEEE, 2003:112-117.

[12] XU Wei,YU Jianyu,CHEN Mao. Intra-pulse modulation recognition of radar signal based on multi-dimensional features[J]. Journal of Terahertz Science and Electronic Information Technology, 2018,16(1): 81-86.

[13] REDMON J,DIVVALA S,GIRSHICK R,et al. You only look once: unified, real-time object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas,NV,USA:IEEE, 2016:779-788.

[14] ZHU Jun-Yan,PARK Taesung,ISOLA Phillip,et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]// IEEE International Conference on Computer Vision (ICCV). Venice,Italy:IEEE, 2017:1-18.

[15] ABDEL-HAMID O,MOHAMED A,JIANG H,et al. Convolutional neural networks for speech recognition[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014,22(10):1533-1545.

[16] O'SHEA T J,CORGAN J,CLANCY T C. Convolutional radio modulation recognition networks[C]// International Conference on Engineering Applications of Neural Networks. Aberdeen,UK: Springer, 2016:213-226.

[17] BA Lei Jimmy,CARUANA Rich. Do deep nets really need to be deep?[C]// Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge,MA,USA:MIT Press, 2014:2654-2662.

[18] SIMONYAN Karen,ZISSERMA Andrew. Very deep convolutional networks for large-scale image recognition[C]// ICLR 2015. San Diego,CA,USA:[s.n.], 2015:1-14.

[19] HE K M,ZHANG X Y,REN S Q,et al. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas,NV,USA:IEEE, 2016:770-778.

[20] O'SHEA T J,ROY T,CLANCY T C. Over-the-air deep learning based radio signal classification[J]. IEEE Journal of Selected Topics in Signal Processing, 2018,12(1):168-179.

[21] HUANG G,LIU Z,MAATEN L V D,et al. Densely connected convolutional networks[C]// IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu,HI,USA:IEEE, 2017:2261-2269.

[22] RAMJEE Sharan,JU Shengtai,YANG Diyu,et al. Fast deep learning for automatic modulation classification[J/OL]. arXiv, 2019:1901.05850v1.

[23] HOCHREITER S,SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997,9(8):1735-1780.

[24] RAJENDRAN S,MEERT W,GIUSTINIANO D,et al. Deep learning models for wireless signal classification with distributed low-cost spectrum sensors[J]. IEEE Transactions on Cognitive Communications and Networking, 2018,4(3):433-445.

[25] SAINATH T N,VINYALS O,SENIOR A,et al. Convolutional,long short-term memory,fully connected deep neural networks[C]// IEEE International Conference on Acoustics,Speech and Signal Processing. Brisbane,QLD,Australia:IEEE, 2015:4580-4584.

[26] TANG B,TU Y,ZHANG Z Y,et al. Digital signal modulation classification with data augmentation using generative adversarial nets in cognitive radio networks[J]. IEEE Access, 2018(6):15713-15722.
[27] TU Y,LIN Y,WANG J,et al. Semi-supervised learning with generative adversarial networks on digital signal modulation classification[J]. Computers, Materials and Continua, 2018,55(2):242-254.
[28] LI M,LIU G,LI S,et al. Radio classify generative adversarial networks: a semi-supervised method for modulation recognition[C]// 18th International Conference on Communication Technology. Chongqing,China:IEEE, 2018:669-672.
[29] O'SHEA T,WEST N. Radio machine learning dataset generation with GNU radio[C]// Proceedings of the GNU Radio Conference. Boulder,Colorado,USA:[s.n.], 2016.
[30] KULIN M,KAZAZ T,MOERMAN I,et al. End-to-end learning from spectrum data: a deep learning approach for wireless signal identification in spectrum monitoring applications[J]. IEEE Access, 2018(6):18484-18501.
[31] GOODFELLOW I,BENGIO Y,COURVILLE A. Deep learning[M]. Cambridge,MA,USA:MIT Press, 2017:330-371.
[32] KLAMBAUER G,UNTERTHINER T,MAYR A,et al. Self-normalizing neural networks[C]// Proceedings of 31st Conference on Neural Information Processing Systems(NIPS 2017). Long Beach,CA,USA:Springer, 2017:971-980.
[33] GRAVES Alex. Supervised sequence labelling with recurrent neural networks[M]. Berlin,Heidelberg:Springer, 2012:31-38.
[34] SUTSKEVER I,VINYALS O,LE Q V. Sequence to sequence learning with neural networks[C]// Proceedings of the 27th International Conference on Neural Information Processing System(NIPS 2014). Cambridge,MA,USA:MIT Press, 2014: 3104-3112.

基于深度 LSTM 辅助卷积网络的新型自动调制分类

吴　楠，谷万博，王旭东

(大连海事大学 信息科学技术学院，辽宁 大连 116023)

摘　要：自动调制分类在无线频谱异常检测和无线电环境感知中将发挥重要作用。随着深度学习算法的突破，调制分类任务可利用神经网络达到前所未有的高分类精确度。文中提出了一种新颖的神经网络，称为长短期卷积深度神经网络(LCDNN)。该网络创造性地结合了长短期记忆网络(LSTM)、卷积神经网络(CNN)和深度网络体系结构的优点。该模型无需人工提取特征，并可直接从训练数据集中的原始时域幅度和相位样本中学习。仿真结果表明，该模型在高信噪比下的分类精确度达到 93.5%。此外，对模型的噪声敏感性进行了研究，并证明在一定信噪比范围内，LCDNN 模型较现有的基础模型在噪声破坏方面更具弹性。最后，为降低模型计算的复杂度，提出了一种结构更为简洁的 LCDNN 模型，该模型仅使用原始 LCDNN 模型的 0.6%的参数即可实现高精确度的分类准确率。

关键词：调制分类；深度学习；长短期卷积深度神经网络；卷积神经网络

中图分类号：TN975　　**文献标志码：**A　　**doi：**10.11805/TKYDA2020034