A Spatiotemporal Multi-Channel Learning Framework for Automatic Modulation Recognition

Jialang Xu, Chunbo Luo, *Member, IEEE*, Gerard Parr, *Member, IEEE*, and Yang Luo

***Abstract*—Automatic modulation recognition (AMR) plays a vital role in modern communication systems. This letter proposes a novel three-stream deep learning framework to extract the features from individual and combined in-phase/quadrature (I/Q) symbols of the modulated data. The proposed framework integrates one-dimensional (1D) convolutional, two-dimensional (2D) convolutional and long short-term memory (LSTM) layers to extract features more effectively from a time and space perspective. Experiments on the benchmark dataset show the proposed framework has efficient convergence speed and achieves improved recognition accuracy, especially for the signals modulated by higher dimensional schemes such as 16 quadrature amplitude modulation (16-QAM) and 64-QAM.**

***Index Terms*—Automatic modulation recognition, deep learning, multi-channel.**

I. INTRODUCTION

AUTOMATIC modulation recognition identifies the modulation pattern of communication signals received from wireless or wired networks to provide the premise guarantee for information extraction. As a key step between signal detection and demodulation, AMR has been widely used in various fields. It can be generally grouped into two categories: likelihood-based (LB) methods and feature-based (FB) methods. LB methods [1], [2] are optimal in the sense of Bayesian estimation but rely heavily on prior knowledge and parameter estimation. FB methods [3], [4] perform feature extraction and classification. For the feature extraction process, various manual features, such as phase features, instantaneous features, and constellation shape features, are carefully extracted by expert systems. Then the robust classification algorithms including artificial neural architecture, support vector machine, and decision tree are applied in the classification process to improve recognition performance.

In the past few years, deep learning frameworks, such as convolutional neural networks (CNNs) [5], recurrent neural networks (RNNs) [6], and convolutional long short-term deep neural networks (CLDNNs) [7], have been applied to AMR and shown to outperform than FB methods. However, these frameworks, which are directly borrowed from the image, speech recognition or natural language processing areas, are not specifically designed for AMR. More recently, a few deep learning approaches began to explore the specific characteristics of modulated signals for AMR. For instance, spectrograms are generated by the short-time discrete Fourier transform as the input to the CNN classifier in [8]. In [9], the amplitude and phase series are trained separately based on a parallel fusion method of 1D-CNN. In [10], a correction module is co-trained with CNN to reduce the effect of random frequency and phase noise.

To improve the performance of AMR, one intuitive idea is to utilize multiple types of the data sources, such as spectrograms, amplitude/phase series, and constellation graphs. It is reasonable because they provide different views of the modulation information for a signal and offer new classification basis for AMR models. However, obtaining multiple forms of a signal is a complex task in practical applications and different input forms might affect the performance of the algorithm. We thus propose an efficient multi-stream structure to process the I/Q multi-channel data of the original modulation signals, supplemented by the independent I/Q channel inputs to fully extract features. It is feasible because of the following two aspects. First, because the existence of amplitude and phase unbalance deteriorates the orthogonal relationship between I and Q channel, resulting in intrinsic differences of the two channels, the features extracted from I channel, Q channel and I/Q multi-channel data will be complementary. Second, it is easy to get independent I/Q channel inputs by splitting the multi-channel data, and the forms of these three input data are consistent.

The main contribution of this letter includes a novel framework to exploits the complementary information from I/Q multi-channel, I channel, and Q channel data and utilize spatial and temporal properties existing in the signal for AMR. Simulation results confirm the proposed framework achieves better recognition performance than the State-of-the-Art (SoA) deep learning frameworks. We further explore the rationality of our proposed framework through controlled groups of experiments and analyze the aspects of the confusion matrix, model complexity, and learned representation.

II. SIGNAL MODEL AND PROPOSED FRAMEWORK

This letter considers a single-input single-output communication system and the received signal $r(t)$ can be represented by

$$r(t) = s(t) * h(t) + n(t) \tag{1}$$

Manuscript received March 23, 2020; revised May 23, 2020; accepted May 30, 2020. Date of publication June 2, 2020; date of current version October 7, 2020. This work was supported by the National Science Foundation (NSFC) under Grant 61871096. The associate editor coordinating the review of this article and approving it for publication was W. Wang. *(Corresponding authors: Chunbo Luo; Yang Luo.)*

Jialang Xu and Yang Luo are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: 201921010235@std.uestc.edu.cn; luoyang@uestc.edu.cn.).

Chunbo Luo is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China, and also with the Department of Computer Science, University of Exeter, Exeter EX4 4RN, U.K. (e-mail: c.luo@uestc.edu.cn).

Gerard Parr is with the School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, U.K. (e-mail: g.parr@uea.ac.uk).

Digital Object Identifier 10.1109/LWC.2020.2999453

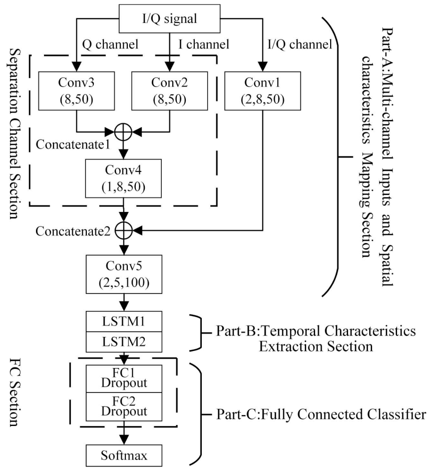


Fig. 1. The structure of the proposed framework MCLDNN.

where $s(t)$ is the modulated signal from the transmitter, $h(t)$ is the channel impulse response and $n(t)$ denotes additive white Gaussian noise (AWGN). The received signal $r(t)$ is sampled n times at a rate $f_s = \frac{1}{T_s}$ by the analog to digital converter, which generates the discrete-time observed signal $r(n)$.

The proposed framework named multi-channel convolutional long short-term deep neural network (MCLDNN) is shown in Fig. 1. MCLDNN integrates CNNs, LSTMs, and fully connected (FC) deep neural networks in a unified structure to utilize the complementarity and synergy of them for spatiotemporal feature extraction and classification. Specially, it has three functional parts: multi-channel inputs and spatial characteristics mapping, temporal characteristics extraction, and fully connected classification.

A. Multi-Channel Feature Extraction of I/Q Signals

Part-A composed of two 1D convolutional layers (Conv2 and Conv3) and three 2D convolutional layers (Conv1, Conv4, and Conv5) to provide superior features for the LSTM by reducing noise variance and feeding higher-level abstraction of the input data. Firstly, the received I/Q multi-channel signals are separated into independent I channel and Q channel data. Then these three streams of input data are fed to Conv1, Conv2, and Conv3 respectively and learn the multi-channel and individual channel features of I/Q signals. Especially, Conv2 and Conv3 use 'casual' padding to make sure the framework cannot violate the ordering in which we model the separated channel data. Later, they are fused in Concatenate2 and fed to Conv5 to plumb spatial correlations. This multi-channel input and processing structure of Part-A captures features about the input representation at different scales, which leverages the complementary information of I channel, Q channel, and I/Q multi-channel data effectively.

B. Temporal Characteristics Extraction and Fully Connected Classification

Inspired by the structure proposed in [6], Part-B has two LSTM layers with 128 cells to effectively process sequential data for extracting the temporal correlations.

For mapping features to a more separable space, we add two FC layers with 128 neurons and scaled exponential linear units (SeLu) activation to deepen our network. The dropout algorithm is adopted to prevent the framework from overfitting. The output layer uses Softmax and has 11 neurons, one for each modulation scheme.

The flexible number of input layers at the input of the proposed framework provides high extendibility, facilitating the input of other modulation information such as constellation, amplitude, and phase for more accurate classification. For example, other constellation data can be merged directly in Concatenate2 as longs as its dimension is unified.

III. Experiments Background

A. Data

These experiments adopt two benchmark open-source datasets RadioML2016.10a and RadioML2016.10b [11]. RadioML2016.10a dataset includes 220000 modulated signals with signal-to-noise ratios (SNRs) varying from -20dB to $+18$dB and considers 11 commonly used modulations: WBFM, AM-DSB, AM-SSB, BPSK, CPFSK, GFSK, 4-PAM, 16-QAM, 64-QAM, QPSK, and 8PSK. RadioML2016.10b dataset is an extended version of RadioML2016.10a containing 1200000 signals and 10 types of modulated signals except AM-SSB. Each signal in these two datasets has 128 complex floating-point time I/Q samples and is generated in harsh simulated propagation environments, corrupted by AWGN, multi-path fading, sampling rate offset, and center frequency offset, to resemble the practical environments. We divide the datasets by a ratio of 6:2:2. Specially, for each modulation type per SNR, we randomly select 600 (3600) signals as training data, 200 (1200) signals as validation data and 200 (1200) signals as test data.

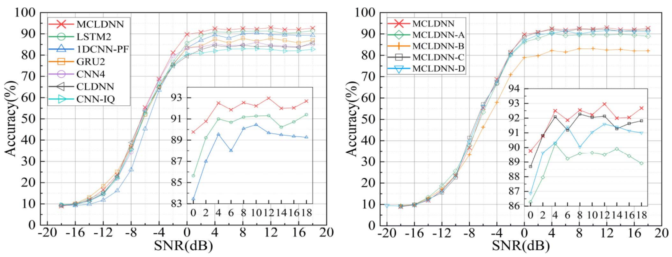
B. Implementation Details

We set two experiments to evaluate the MCLDNN. In the first experiment, we compared the MCLDNN against six SoA algorithms from [5], [6], [7], [9], [12], [13] named as CNN-IQ, LSTM2, CLDNN, 1DCNN-PF, CNN4, and GRU2, respectively. LSTM2 and 1DCNN-PF models use the normalized amplitude/phase series transformed from I/Q signals as input while the other models utilize I/Q signals directly. The second experiment explores the impact of different modules combination on the performance of the proposed framework. We set up four controlled groups of frameworks, referring to as MCLDNN-A (the separation channel section is removed from MCLDNN), MCLDNN-B (the temporal characteristics extraction section is removed), MCLDNN-C (the FC section is removed) and MCLDNN-D (the Conv1 and Concatenate2 are removed).

We use categorical cross-entropy as a loss function and Adam optimizer for all frameworks. The initial learning rate starts at 0.001 and is multiplied by a factor of 0.8 if the verification loss does not decrease within 5 epochs to improve the training efficiency. The batch size of the gradient update is 400 to avoid local optima. We stop the training process when the validation loss does not improve for 60 epochs and use the trained model with the minimum validation loss to predict the modulation type of each test signal. All experiments are conducted in TensorFlow using Keras library, supported by NVIDIA CUDA with a GeForce GTX 1080Ti GPU.

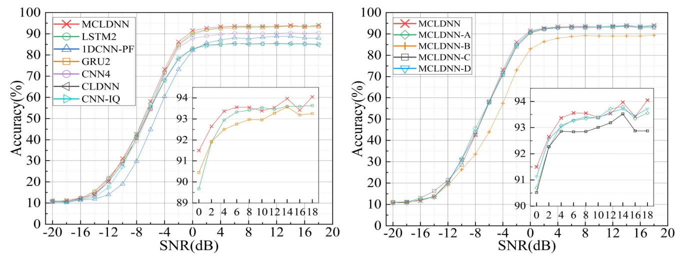
TABLE I
RECOGNITION ACCURACY ON RADIOML2016.10A AT 0DB SNR

	MCLDNN	MCLDNN-A	MCLDNN-B	MCLDNN-C	MCLDNN-D	LSTM2	1DCNN-PF	GRU2	CNN4	CLDNN	CNN-IQ
8PSK	**94.0%**	87.5%	85.5%	91.0%	92.5%	77.0%	69.0%	86.0%	87.0%	82.0%	87.0%
AM-DSB	91.0%	95.5%	56.5%	92.5%	60.0%	67.5%	**99.0%**	86.5%	98.0%	90.5%	79.0%
AM-SSB	94.5%	93.5%	81.5%	92.0%	93.0%	**97.5%**	96.5%	92.0%	96.5%	94.5%	94.0%
BPSK	**99.5%**	99.0%	**99.5%**	**99.5%**	99.0%	93.0%	99.0%	**99.5%**	**99.5%**	97.0%	98.5%
CPFSK	**100%**	**100%**	**100%**	99.5%	99.5%	99.5%	99.5%	99.0%	**100%**	97.0%	**100%**
GFSK	96.0%	**97.5%**	93.5%	**97.5%**	93.0%	94.5%	88.0%	**97.5%**	97.0%	95.0%	96.0%
PAM4	**98.5%**	**98.5%**	97.5%	**98.5%**	**98.5%**	**98.5%**	**98.5%**	98.0%	98.0%	98.0%	**98.5%**
16-QAM	**92.0%**	72.5%	50.0%	85.0%	84.5%	85.0%	84.0%	62.5%	62.5%	38.5%	33.0%
64-QAM	**88.0%**	80.0%	54.0%	87.5%	85.5%	81.5%	72.0%	71.5%	48.0%	72.5%	70.0%
QPSK	97.0%	93.5%	95.5%	95.5%	95.0%	92.5%	91.5%	87.5%	**98.0%**	72.0%	85.5%
WBFM	35.5%	31.5%	**55.0%**	37.0%	**55.0%**	55.5%	21.0%	38.5%	32.0%	37.5%	45.5%



(a) (b)

Fig. 2. Recognition accuracy comparison on RadioML2016.10a dataset between MCLDNN and (a) the other frameworks, (b) its varieties.



(a) (b)

Fig. 3. Recognition accuracy comparison on RadioML2016.10b dataset between MCLDNN and (a) the other frameworks, (b) its varieties.

IV. RESULT AND DISSUASION

A. Recognition Accuracy

Fig. 2(a) shows the recognition accuracy of the frameworks specified in the first experiment. The MCLDNN performs clearly better than other frameworks above -4dB SNRs and reaches a maximum accuracy of 92.95% when the SNR is 12dB. The average recognition rate of MCLDNN from 0 to 18dB is 92%, which is improved by 2% to 10% comparing with the others. Fig. 2(b) plots the recognition accuracy of the frameworks mentioned in the second experiment. The best recognition performance is delivered by our proposed model MCLDNN, which shows that the advantages of the modules are complementary and their union leads to a superior model. The significantly degraded recognition performance of MCLDNN-B illustrates the importance of temporal modeling of input data.

Table I shows the recognition accuracy of each modulation type when the SNR is 0dB. The MCLDNN has advantages in 8PSK, BPSK, CPFSK, and 4-PAM recognitions and excels at 16-QAM/64-QAM confusion problem. Comparing the recognition accuracy of MCLDNN with its varieties MCLDNN-A, MCLDNN-B, MCLDNN-C, and MCLDNN-D, we can conclude that adding individual I/Q features with temporal modeling improves the recognition accuracy of 16-QAM and 64-QAM significantly.

To show the robustness of our proposed model, we re-simulate all the algorithms using another large dataset (RML2016.10b). Fig. 3 illustrates the recognition rates of our MCLDNN still outperform other SoA frameworks.

B. Confusion Matrix

Fig. 4 demonstrates a series of confusion matrices at -2 dB SNR. For a confusion matrix, each row represents the real modulation type and each column represents the predicted modulation type. The following two main factors are affecting recognition accuracy at the low SNR range.

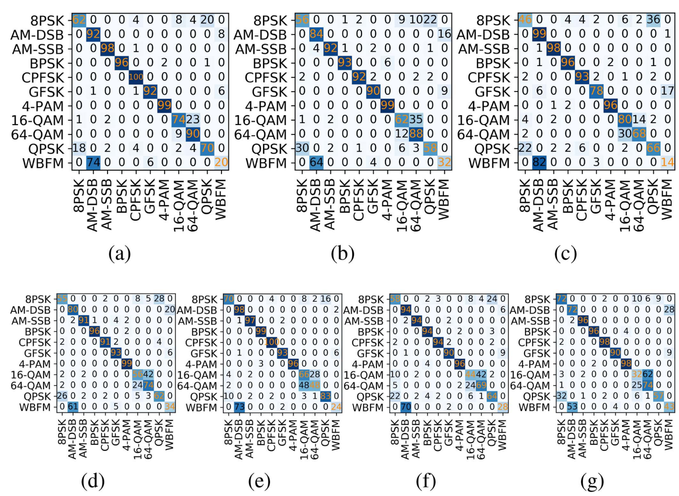


Fig. 4. Confusion matrix of (a) MCLDNN (b) LSTM2 (c) 1DCNN-PF (d) GRU2 (e) CNN4 (f) CLDNN (g) CNN-IQ on RadioML2016.10a dataset at -2dB SNR.

The first factor is the confusion problem between WBFM and AM-DSB. Both of them belong to continuous modulation, which means the distinct features between them are very weak on the complex panel. Besides, the WBFM and AM-DSB data in the dataset were generated by sampling analog audio signals, in which silent periods exist, making the situation worse. The second factor is the confusion problem between 16-QAM and 64-QAM. The reason for high classification error is because they have overlapped constellation points in the digital domain. Though such a problem still exists in the MCLDNN, it is greatly improved compared with the other SoA deep learning frameworks.

C. Computational Complexity

The computational complexity is evaluated by the number of learned parameters, average training time, training epochs,

TABLE II
MODEL COMPLEXITY ON RADIOML2016.10A DATASET

Framework	Learned parameters	Training time (second/epoch)	Traning epochs
MCLDNN	406199	17	113
MCLDNN-A	335249	12	110
MCLDNN-B	325943	8	108
MCLDNN-C	373175	17	**92**
MCLDNN-D	355349	12	141
LSTM2	201099	11	114
1DCNN-PF	174923	12	376
GRU2	**151179**	10	110
CNN4	858123	21	283
CLDNN	164233	18	309
CNN-IQ	1592383	**5**	131

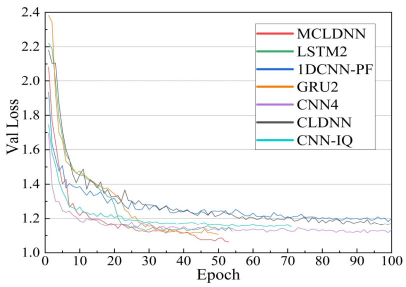


Fig. 5. Training process on RadioML2016.10a dataset.

(a) (b)

Fig. 6. Weights of (a) Conv2 (b) Conv3.

and convergences speed. As shown in Table II, the number of learned parameters of the MCLDNN is smaller than CNN-IQ and CNN4 but larger than the other models. With the improved recognition accuracy and relatively few training epochs, the training time of the MCLDNN is still acceptable.

Fig. 5 gives the training process including validation loss within 100 epochs. It indicates that the validation loss of the GRU2 keeps the fastest velocity of falling in the early stage of training (epoch<7) while the MCLDNN has better convergence speed and needs the second least number of epochs. Compared with those which based on 2D-CNN algorithms (CNN4 and CNN-IQ), the slow convergence speed is a main weakness of the frameworks based on RNN and 1D-CNN methods (LSTM2, GRU2, CLDNN, and 1DCNN-PF). Therefore, LSTM2 and 1DCNN-PF need an additional signal preprocessing to extract the amplitude-phase features of the signal manually in order to achieve promising performance and reduce the difficulty of convergence, which requires extra computation. On the contrary, even though RNN and 1D-CNN algorithms are used in our proposed framework, MCLDNN still obtains excellent convergence performance like 2D-CNN algorithms have and achieves the lowest loss value without reformatting data. This is because the higher-level modeling of data by Part-A can help to disentangle underlying factors of variation within the input, which makes LSTM layers easier to learn temporal structure between successive time steps.

D. Learned Representations

Similar to [14], we visualize the layers on RadioML2016.10a dataset to provide the insight of the proposed model. The weights of all filters of Conv2 and Conv3 are presented in Fig. 6. It can be observed that

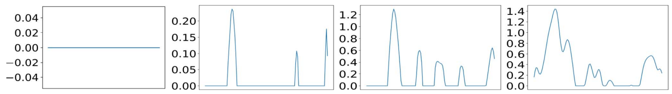


Fig. 7. Visualization of the partial outputs of Part-A.

these filters have similar functions as gradient detectors and convolve the single-dimensional channel data from different perspectives. Fig. 7 visualizes part of the outputs of Part-A. We can find the signals still have a strong temporal correlation after passing through Part-A, which can help LSTM layers to extract the features more efficiently. Since some of the outputs of Part-A are 0, we can reduce the model complexity by removing these irrelevant outputs.

V. CONCLUSION

In this letter, we propose a novel multi-channel input model to extract features from I/Q multi-channel, I channel and Q channel data for AMR. The recognition performance and convergence rate of our model demonstrate that using effective model structures and extracting features from a time and space perspective have significant potential in communication systems.

REFERENCES

[1] A. Abdi, O. A. Dobre, R. Choudhry, Y. Bar-Ness, and W. Su, "Modulation classification in fading channels using antenna arrays," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, vol. 1, 2004, pp. 211–217.
[2] B. Dulek, "Online hybrid likelihood based modulation classification using multiple sensors," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 4984–5000, Aug. 2017.
[3] D. Chang and P.-K. Shih, "Cumulants-based modulation classification technique in multipath fading channels," *IET Commun.*, vol. 9, no. 6, pp. 828–835, Apr. 2015.
[4] S. Huang, Y. Yao, Z. Wei, Z. Feng, and P. Zhang, "Automatic modulation classification of overlapped sources using multiple cumulants," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6089–6101, Jul. 2017.
[5] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, vol. 629, 2016, pp. 213–226.
[6] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, Sep. 2018.
[7] N. E. West and T. O'Shea, "Deep architectures for modulation recognition," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, 2017, pp. 1–6.
[8] Y. Zeng, M. Zhang, F. Han, Y. Gong, and J. Zhang, "Spectrum analysis and convolutional neural network for automatic modulation recognition," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 929–932, Jun. 2019.
[9] E. Perenda, S. Rajendran, and S. Pollin, "Automatic modulation classification using parallel fusion of convolutional neural networks," in *Proc. 3rd IEEE Int. Balkan Conf. Commun. Netw. (BalkanCom)*, Skopje, North Macedonia, Jun. 2019.
[10] D. Hong, Z. Zhang, and X. Xu, "Automatic modulation classification using recurrent neural networks," in *Proc. 3rd IEEE Int. Conf. Comput. Commun. (ICCC)*, Dec. 2017, pp. 695–700.
[11] T. J. O'shea and N. West, "Radio machine learning dataset generation with gnu radio," in *Proc. GNU Radio Conf.*, vol. 1, 2016, pp. 1–6.
[12] K. Tekbıyık, A. R. Ekti, A. Görçin, G. K. Kurt, and C. Keçeci, "Robust and fast automatic modulation classification with CNN under multipath fading channels," 2019. [Online]. Available: arXiv:1911.04970.
[13] D. Hong, Z. Zhang, and X. Xu, "Automatic modulation classification using recurrent neural networks," in *Proc. 3rd IEEE Int. Conf. Comput. Commun. (ICCC)*, Dec. 2017, pp. 695–700.
[14] Y. Wu, X. Li, and J. Fang, "A deep learning approach for modulation recognition via exploiting temporal correlations," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.