Modulation Classification Using Convolutional Neural Network Based Deep Learning Model

Shengliang Peng*†, Hanyu Jiang†, Huaxia Wang†, Hathal Alwageed†, and Yu-Dong Yao†
* College of Information Science and Engineering, Huaqiao University, Xiamen 361021, P.R. China
† Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA
Email: peng.shengliang@gmail.com, {hjiang5, hwang38, halwagee, Yu-Dong.Yao}@stevens.edu

***Abstract*—Deep learning (DL) is a powerful classification technique that has great success in many application domains. However, its usage in communication systems has not been well explored. In this paper, we address the issue of using DL in communication systems, especially for modulation classification. Convolutional neural network (CNN) is utilized to complete the classification task. We convert the raw modulated signals into images that have a grid-like topology and feed them to CNN for network training. Two existing approaches, including cumulant and support vector machine (SVM) based classification algorithms, are involved for performance comparison. Simulation results indicate that the proposed CNN based modulation classification approach achieves comparable classification accuracy without the necessity of manual feature selection.**

I. Introduction

Deep learning (DL) is a branch of machine learning (ML) that has state-of-the-art capability for classification [1]. Recently, it has attracted great attention and been applied in various fields. During the competition of ImageNet Large Scale Visual Recognition Challenge (ILSVRC), many research teams submitted various DL algorithms for object detection and image classification, and the latest top-5 accuracy of identifying objects from 1000 categories reaches 95% [2]. In [3], the authors show that DL network can be trained to implement accurate, inexpensive and scalable economics survey with satellite imagery in developing countries. Moreover, bioinformatics can also benefit from DL. Splice junctions can be discovered from DNA sequences, finger joints can be recognized from X-ray images, lapses can be detected from electroencephalography signals, and so on [4].

Although DL is flourishing everywhere, communications field seems to be an exception. It is noticed that some traditional ML algorithms, e.g., Supported Vector Machine (SVM) and K-Nearest Neighbor (KNN), have been utilized for media access control (MAC) protocol identification [5] and modulation classification [6]. In this paper, we focus on the issue of using DL in communications systems, especially for modulation classification.

The use of DL in communications systems has multiple advantages. Firstly, because of the huge amount of communications devices and the high communications data rate, massive data, which is required by DL, are available in communications systems. Secondly, DL is able to extract features autonomously and avoids the challenging task of manual feature selection. Thirdly, since DL is evolving rapidly, there will be considerable potential to other communications applications besides modulation classification.

Modulation classification is usually a major communications problem with both civilian and military applications, such as spectrum management, signal identification, electronic warfare, threat analysis, etc. Previously, it was handled either in either traditional signal processing [7] or ML approach [6]. In this article, we aim to solve it by use of DL. One of the most prevalent DL architectures, convolutional Neural Network (CNN) [8], is considered for modulation classification. Modulated signals are converted into grid-like topological data, e.g., images of constellation diagrams, to feed CNN. AlexNet [9], a famous CNN model, is adopted and modified for network training and accuracy testing. The whole modulation classification work is implemented based on the Caffe framework [10].

The rest of this article is organized as follows. Section II describes the signal model and traditional algorithms for modulation classification. Section III gives an overview of DL, CNN, and AlexNet. Section IV shows how to use DL for modulation classification. Simulation results are provided in Section V. Finally, Section VI concludes this paper.

II. Problem Formulation

A. Signal Model

Assume that we are operating in a coherent, synchronous environment with single-tone signaling and that carrier, timing, and waveform recovery have been accomplished. we then obtain a baseband sequence composed of samples of the complex envelope [7],

$$y(n) = Ae^{j(2\pi f_0 nT+\theta_n)} \sum_{l=-\infty}^{\infty} x(l)\, h(nT - lT + \epsilon_T T) + g(n), \tag{1}$$

where $x(l)$ represents the symbol sequence, A is an unknown amplitude factor, f_0 denotes the carrier frequency offset, θ_n is the phase jitter, T represents the symbol spacing, $h(\cdot)$ denotes the residual channel effect, ϵ_T is the time error, and $g(n)$ represents the additive Gaussian noise sequence.

Our modulation classification task is to decide which modulation scheme has been utilized with the knowledge of the N sample received vector $\mathbf{y} = [y(1), y(2), \cdots .y(N)]^T$.

This paper considers four different modulation types, including quadrature phase-shift keying (QPSK), 8 phase-shift keying (8PSK), 16 quadrature amplitude modulation (16QAM),

978-1-5090-4909-7/17/$31.00 ©2017 IEEE

and 64 quadrature amplitude modulation (64QAM). Conventional approaches that use cumulants [7] and support vector machine (SVM) techniques [11] are firstly introduced to classify/identify these modulation types.

B. Modulation Classification Using Cumulants

In [7], a digital modulation classification method has been proposed based on elementary fourth-order cumulants. For a complex-valued received signal $y(n)$, the N sample estimation of its second-order moments are given by

$$\begin{aligned}\hat{C}_{21} &= \frac{1}{N}\sum_{n=1}^{N}|y(n)|^2,\\ \hat{C}_{20} &= \frac{1}{N}\sum_{n=1}^{N}y^2(n),\end{aligned} \tag{2}$$

and the estimated fourth-order cumulants can be written as

$$\begin{aligned}\hat{C}_{40} &= \frac{1}{N}\sum_{n=1}^{N}y^4(n) - 3\hat{C}_{20}^2,\\ \hat{C}_{41} &= \frac{1}{N}\sum_{n=1}^{N}y^3(n)\,y^*(n) - 3\hat{C}_{20}\hat{C}_{21},\\ \hat{C}_{42} &= \frac{1}{N}\sum_{n=1}^{N}|y(n)|^4 - |\hat{C}_{20}|^2 - 2\hat{C}_{21}^2.\end{aligned} \tag{3}$$

The modulated signals are normalized to have unit energy, which implies that $C_{21} = 1$. Thus, the normalized fourth-order cumulants estimation can be expressed as

$$\widetilde{C}_{4k} = \hat{C}_{4k}/\hat{C}_{21}^2, \quad k = 0, 1, 2. \tag{4}$$

Theoretical values of the statistics C_{40} ($N \to \infty$) for four modulation types are presented in TABLE I. Subtracting C_{40} from the real part of $\widetilde{C}_{40}$, the lowest absolute value of subtraction results indicates which modulation type has been utilized.

TABLE I. Theoretical value of statistics C_{40} for four modulation types

Types	QPSK	8PSK	16QAM	64 QAM
C_{40}	1.0000	0.0000	-0.6800	-0.6191

C. Modulation Classification Using Support Vector Machine

SVM, a supervised ML algorithm, has been widely used in classification problems. The main ideas and steps of the SVM algorithm can be summarized as follows.

1) Use a canonical equation to define an optimal hyperplane so that the data from two different classes has the maximum margin widths.
2) Considering a non-linearly separable problem, introduce a slack variable [12] which can be used to relax the constraints of the canonical equation. The goal is to find a hyperplane with a minimum misclassification rate.
3) Map data to a higher dimensional space using kernel based learning where it is easier to classify with linear decision surfaces. Therefore, reformulate the problem so that the data is mapped explicitly to this space.

Specifying features is needed for SVM based classification. For our modulation classification task, the features of the fourth and sixth order cumulants are selected [13],

$$\begin{aligned}
C_{40} &= \mathrm{cum}\,(y(n), y(n), y(n), y(n))\\
&= M_{40} - 3M_{20}^2\\
C_{41} &= \mathrm{cum}\,(y(n), y(n), y(n), y^*(n))\\
&= M_{41} - 3M_{20}M_{21}\\
C_{42} &= \mathrm{cum}\,(y(n), y(n), y^*(n), y^*(n))\\
&= M_{42} - |M_{20}|^2 - 2M_{21}^2\\
C_{60} &= \mathrm{cum}\,(y(n), y(n), y(n), y(n), y(n), y(n))\\
&= M_{60} - 15M_{20}M_{40} + 3M_{20}^3\\
C_{61} &= \mathrm{cum}\,(y(n), y(n), y(n), y(n), y(n), y^*(n))\\
&= M_{61} - 5M_{21}M_{40} - 10M_{20}M_{41} + 30M_{20}^2M_{21}\\
C_{62} &= \mathrm{cum}\,(y(n), y(n), y(n), y(n), y^*(n), y^*(n))\\
&= M_{62} - 6M_{20}M_{42} - 8M_{21}M_{41} - M_{22}M_{40}\\
&\quad + 6M_{20}^2M_{22} + 24M_{21}^2M_{20}\\
C_{63} &= \mathrm{cum}\,(y(n), y(n), y(n), y^*(n), y^*(n), y^*(n))\\
&= M_{63} - 9M_{21}M_{42} + 12M_{21}^3 - 3M_{20}M_{43}\\
&\quad - 3M_{22}M_{41} + 18M_{20}M_{21}M_{22}
\end{aligned}$$

where $M_{pq} = \mathrm{E}[y(k)^{p-q}\,(y^*(k))^q]$ represents the moment of a signal.

In the experiments, SVM based modulation classification is implemented with the help of LIBSVM [14].

III. Deep Learning

DL, a cutting-edge research field with rapid growth, has outperformed in many scientific domains, such as computer vision, natural language processing, bioinformatics and etc, where the traditional ML methods encounter bottlenecks or some experience-based knowledge/skills are heavily relied on. The most attractive advantage of DL is the elimination of manual feature engineering. Instead, it is able to extract/learn features from data autonomously. A DL model typically contains many layers with nonlinear processing units, and each layer transforms inputs from the previous layer, which eventually forms a hierarchical data representation: high-level features are more general and abstract than low-level features. In this Section, we give a brief description of DL, CNNs, and AlexNet with their highlight features.

A. Convolutional Neural Network

Compared with ordinary neural networks, the CNNs have two types of layers with constrained connectivity patterns: convolutional layers and pooling layers. A convolutional layer takes feature maps of the previous layer as inputs and make two-dimensional convolution operations between those inputs and a set of learnable filters. Then a stack of new feature maps are produced and fed into next layer. Mathematically, output feature maps of each layer can be presented as below:

$$X_n = f\Big(\sum_m \mathbf{W}_n^m * X^m + b_n\Big), \tag{5}$$

where X_n represents nth feature maps, $\mathbf{W}_n^m$ represents filters which are involved into the convolution operations $*$, and the b_n indicates bias value that corresponds to each feature map. The sparse connectivity of convolutional layer only allow neurons to connect with a local region of input volume, which significantly reduces the number of parameters in the model. This connectivity pattern enables CNNs to accept inputs with larger dimensionality which are computational infeasible to ordinary neural networks.

The pooling layer is typically inserted after a convolutional layer to reduce the dimensionality of feature maps and hence the number of parameters also. It is a non-linear down-sampling operation that aggregates a small patch of units within a feature map, commonly maximizes values of a 2x2 region with a shifting stride of 2, which makes the model invariant to small translation of inputs [15]. Given the input data with grid-like topologies (i.e., images), a stack of convolution and pooling layers are able to abstract fine-grained (i.e., points, edges) and coarse-grained representations (i.e., shapes) from data, which is also the reason that CNNs can achieve big success in image-based applications such as face recognition, object detection/localization, video analysis and etc.

Besides, there are typically multiple fully-connected (FC) layers (also known as dense layers) at the end of CNN based models. Those layers are just same as layers in ordinary neural networks where all neurons are fully connected to every activation of the previous layer and implement a matrix multiplication with them.

B. AlexNet Model

AlexNet is a large CNN based DL model that consists of 650 thousand neurons and 60 million parameters. It is designed to classify 1.2 million images into 1000 categories, as the requirement of ILSVRC-2012 contest [9], and hence needs a large learning capacity. AlexNet model mainly has 8 layers: 5 convolutional layers and 3 fully-connected layers. Some of the convolutional layers are followed by normalization and max-pooling layers, and the last FC layer is connected to a 1000-way softmax that corresponds to the number of classification categories.

In addition to assemble multiple layers together, AlexNet also employs several novel features to improve the performance on both classification accuracy and training efficiency: A non-saturating activation function, Rectified Linear Units (ReLUs) [16], is used in place of traditional saturating functions such as $f(x) = tanh(x)$ or $f(x) = (1 + e^{-x})^{-1}$. This results in a much faster training procedure than before. Another highlight feature, "dropout" [17] as a regularization method, is introduced to prevent overfitting by reducing co-adaptation of neurons. Technically, it sets outputs of hidden neuron to zero with a probability of 0.5 that forced the network to learn more robust features than usual.

IV. Modulation Classification using CNN

A. Data Conversion

As mentioned above, AlexNet is designed for image classification tasks. Considering the modulation classification problem in communication systems where only complex sample points of modulated signals are available, the data conversion is necessary to bridge the gap between two types of data. Therefore, we propose a method that converts complex sample points into a constellation diagram for the utilization of AlexNet as well as other CNN based DL models.

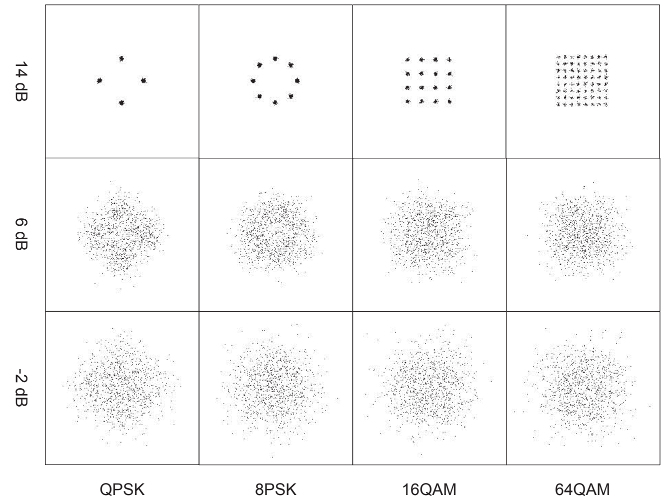

Fig. 1. constellation diagrams for four modulation categories with different SNRs.

The constellation diagram has been widely used as a two-dimensional representation of a modulated signal by mapping signal samples into scatter points on the complex plane. Note that the complex plane is infinite while the scope of an image is limited. Thus, it is necessary to select centric region of the complex plane to generate a constellation diagram. If the selected region is too small, some sample points severely polluted by the noise will be out of range and hence discarded. On the contrary, large selected region results in high resolution of images, which leads to rapid growth of computational complexity when training the deep network. To achieve trade-off between classification performance and computing costs, this paper selects part of the complex plane whose real and imaginary axes both range from -3.5 to 3.5. This kind of constellation diagram is then output an image with JPEG format. Examples of constellation diagrams for four modulation categories with different signal-to-noise ratios (SNRs) are shown in Fig. 1.

B. Network Configuration

In order to facilitate CNN based modulation classification, we adopt the *BLVC reference CaffeNet* model (a minor variation of AlexNet model within the Caffe toolkit) and slightly modify it for better performance and faster learning speed. The number of outputs in layer #8 is changed to 4 as only four modulation types are investigated in our case, and the size of layer #7 is shrinked to 256 because the default size of 4096 always leads to difficulties in convergence in training. In addition, several parameters of solver configuration, such as learning rate (0.01→0.0005), *gamma* (0.1→0.5), and *stepsize* (100,000→10,000), are also adjusted for better classification

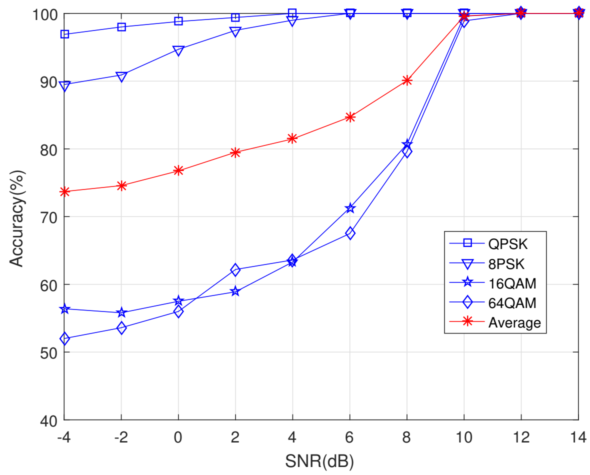

Fig. 2. Classification accuracy of four modulation types and average accuracy versus SNR.

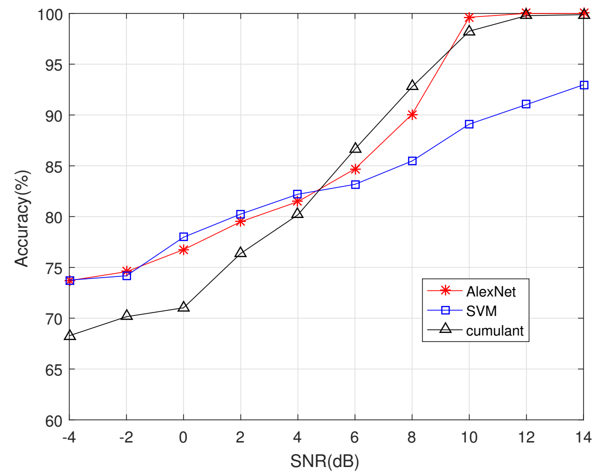

Fig. 3. Modulation classification accuracy comparison.

performance and training efficiency. We also fix the image resolution to be 227x227 as the default receptive field in AlexNet and hence avoid the random cropping operations. Finally, considering the available computing resources on graphics processing unit (GPU), we select to *mini-batch* gradient descent with a batch size of 64 in the first convolutional layer.

C. Implementation

We generate 10,000 images and 1000 constellation diagram images per modulation type for training and testing, respectively. Each image is generated based on 1000 samples of a modulated signal. The maximum number of training iteration is set to be 100,000. All models are trained on a single K40 NVIDIA GPU card with the support of GPU-accelerated libraries.

V. Simulation Results

In order to illustrate the performance of CNN based modulation classification, Fig. 1 presents the classification accuracy of each modulation type as well as the average accuracy with different SNRs range from -2dB to 14dB. For each modulation type, 1000 tests are implemented for performance evaluation. The average accuracy is obtained by averaging the classification accuracy of four modulation types. As shown in Fig. 1, in the low SNR region, the tasks of identifying QPSK and 8PSK are relatively easy (accuracy>90%), while that of identifying 16QAM and 64QAM are rather difficult (accuracy is about 60%). The classification accuracy greatly increases along with the growth of SNRs, especially for 16QAM and 64QAM. In the high SNR region (e.g., SNR¿10dB), all modulation types approach 100% accuracy, which means the proposed approach is able to achieve modulation classification without any errors.

Furthermore, TABLE II gives two confusion matrices with SNR at 4dB and 8dB to explain the detailed classification performance of each modulation type. As shown in this table, more classification errors occur when modulation order is higher. Note that it is hard to identify 16QAM and 64QAM and these two types are usually confused with each other. The reason is that they have the similar square constellation pattern as shown in Fig. 1.

TABLE II. Confusion matrices of four modulation types with SNRs at 4 and 8dB.

SNR	Types	QPSK	8PSK	16QAM	64 QAM	Accuracy
4dB	QPSK	1000	0	0	0	100%
	8PSK	0	990	6	4	99.0%
	16QAM	0	8	633	359	63.3%
	64QAM	0	2	362	636	63.6%
8dB	QPSK	1000	0	0	0	100%
	8PSK	0	1000	0	0	100%
	16QAM	0	0	807	193	80.7%
	64QAM	0	0	204	796	79.6%

Finally, we compare the proposed CNN (AlexNet) based modulation classification approach with traditional cumulant and SVM based algorithms, which have been discussed in Sec. II. Fig. 3 presents the average classification accuracy of these three algorithms versus SNR. It can be seen from this figure that, for cumulant and SVM based algorithms, the former is better when SNR is high (SNR>5dB), while the latter is superior when SNR is low. Obviously, the proposed CNN based approach achieves comparable accuracy with cumulant based algorithm in the high SNR region and is as good as SVM based algorithm in the low SNR region. Note that it obtains this classification performance without manual feature selection.

VI. Conclusion and Discussion

This paper presents the idea of using CNN to classify modulation types in communication systems. In our methodology, constellation diagrams are exploited to represent modulated signals for CNN. Alexnet model is adopted for training and testing. Compared with traditional cumulant and SVM based algorithms, the proposed CNN based approach not only avoids the challenging task of manual feature selection but also achieves comparable performance on modulation classification regardless of different SNR regions.

Although the current CNN based approach may not always outperform existing works, there is still plenty of room for improvement. For example, the data conversion procedure from complex samples to images indeed incurs information loss due to the limited resolution of images. Any enhanced data conversion method that preserves more original information is expected to be helpful. In addition, since the architecture of neural networks has great impact on classification performance, it is worthy to investigate more advanced DL models for modulation classification in the future. Finally, a larger amount of data for training is also beneficial for performance improvement.

ACKNOWLEDGMENT

The authors would like to thank NVIDIA academic partnership program for their generous donation and support.

REFERENCES

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. Cambridge, MA, USA: MIT Press, 2016.

[2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al., "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, pp. 211-252, 2015.

[3] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and ML to predict poverty," *Science*, vol. 353, pp. 790-794, 2016.

[4] S. Min, B. Lee, and S. Yoon, "Deep learning in Bioinformatics," *arXiv: 1603.06430*, 2016.

[5] S. Hu, Y. D. Yao, and Z. Yang, "MAC protocol identification using support vector machines for cognitive radio networks," *IEEE Wireless Commun.*, vol. 21, pp. 52-60, 2014.

[6] M. W. Aslam, Z. Zhu, and A. K. Nandi, "Automatic Modulation Classification Using Combination of Genetic Programming and KNN," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 2742-2750, 2012.

[7] A. Swami and B. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 416-429, Mar. 2000.

[8] Y. LeCun, "Generalization and network design strategies," *University of Toronto Technical Report CRG-TR-89-4*, 1989.

[9] A. Krizhevsky, I. Sutskever and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", in *Advances in Neural Information Processing Systems 25*, pp. 1097-1105, 2012.

[10] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, et al., "Caffe: Convolutional Architecture for Fast Feature Embedding," arXiv:1408.5093, 2014.

[11] C. Burges, "A tutorial on support vector machines for pattern recognition," *Knowledge Discovery and Data Mining*, vol. 2, pp. 1-43, 1998.

[12] K. Miiller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Networks*, vol. 12, pp. 181-201, Mar. 2001.

[13] M. Mirarab and M. Sobhani, "Robust modulation classification for PSK/QAM/ASK using higher order cumulants," in *Proc. 2007 International Conference on Information, Communications and Signal Processing*, pp. 1-4.

[14] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, issue 3, pp. 27:1–27:27, 2011.

[15] Y.-L. Boureau, J. Ponce, Y. LeCun, "A Theoretical Analysis of Feature Pooling in Visual Recognition", in *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp 111118, 2010.

[16] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines", in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp 807-814, 2010.

[17] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R.R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors", *arXiv:1207.0580*, 2012.