Learning Spatiotemporal Features by using CNN for Efficient Modulation Recognition

Godwin Brown Tunze, Gaspard Gashema, Jae-Min Lee and Dong-Seong Kim,
Networked Systems Laboratory, Dept. of IT Convergence Engineering,
Kumoh National Institute of Technology, Gumi, South Korea 39177,
(gbtunze, ggas05, ljmpaul, dskim)@kumoh.ac.kr

***Abstract*—This paper proposes a deep learning framework based on convolutional neural networks aims at extracting and processing spatialtemporal features for an efficient modulation recognition. In this architecture we integrate the strength of grouped and dilated convolutional layers to achieve the efficient recognition in terms of recognition accuracy and less complexity. To allow multilevel feature learning and model generalization we deployed skip connections. Furthermore, to verify the performance of our architecture we performed experimental analysis on RadioML 2018A open-source datasets. According to the results, our model outperforms ResNet based model with regards to recognition accuracy and parameter utilization accuracy.**

***Index Terms*—Convolutional neural network, dilated convolutional layer, spatialtemporal features.**

I. Introduction

AUtomatic modulation recognition (AMR) is the process of identifying modulation format of a received signal in a wireless communication system. Recently, AMR has gained significant attention among signal processing and communication communities due to its application in civil and military. With regards to civil applications, the recent increase in the number of communicating devices has fueled the need of efficient recognition algorithms for spectral management. According to CISCO, 75 billion gadgets are projected to be integrated into a network by 2025 [1].

In general, AMR is achieved by two methods: decision theoretic-based (DC) and feature-based methods. DC based methods regard modulation recognition as a multihypothesis test, where the likelihood of an incoming signal is compared to the threshold value by assuming the probability density function of the signal is known. FB methods were introduced to overcome the computational complexity of DC based method. Although FB solved the complexity of DC based methods, they require extensive feature engineering knowledge. To this end, nowadays deep learning (DL) has replaced convection machine learning in a broad range of applications, including computer vision, signal processing due comprehensive feature learning capability from large amount of data without feature engineering knowledge. With regard to AMR, the state-of-the-arts deep leaning based methods were intensively studied in [2]–[5]. In [2], the architectures based on XGBoost, Visual Geometry Group, and residual network (ResNet) were proposed to verify the feasibility of an open-source dataset (i.e., RadioML 2018A) . In another work [3], the convolutional neural network (CNN) based on skip-connection was proposed to leverage multilevel spatialtemporal features for efficient modulation classification. In addition, [4] proposed a CNN with cleverly adjusted the filter sizes and number besides Gaussian regularization layer for AMR beyond fifth generation communication technology. In [5], CNN trained on mixed dataset was proposed to achieve spatiotemporal feature learning by exploiting common features in the datasets to realize a more general architecture. In the forenamed literature, the advantages of grouped and dilated convolutional layers was not exploited in the design of CNN architectures. Grouped convolutional layer has been used in recent years to develop CNN models for power-constrained devices [6]. Therefore, in this paper, we integrate the strength of grouped and dilated convolutional layer to realize a model with low training cost and high recognition accuracy. So, the major contribution of this paper is to deploy a cleverly combination of standard, grouped and dilated convolutional layers to create an architecture with an excellent tradeoff between recognition accuracy and complexity.

II. Signal Model and Problem Formulation

The signal from transmitter to receiver in a wireless communication system can be represented by

$$r(s) = T(s) * h(s) + g(s), \tag{1}$$

where, $r(s)$ is the received signal, $T(s)$ is the transmitted signal, $h(s)$ represents channel impulse, $g(s)$ represents Gaussian noise with a mean of zero and the power σ^2 while $(*)$ represents convolution operation, and s represents a receiver sampling rate. Generally, modulation recognition is an m-modes classification in which the modulation models are explicitly identified based on the spatialtemporal properties of the received signal.

III. Proposed architecture

To overcome the challenges of AMR in communication systems, we have designed a CNN architecture shown in Fig. 1. Principally, it consists of multi-layers input and output modules, and three cleverly designed blocks called Lightblock. In this architecture an input is received by 3×3 dilated convolution layer that extracts generic features efficiently owing to its high receptive field. Thereafter, 2×2 average pooling strided at $(1, 2)$ is deployed to reduce the spatial component

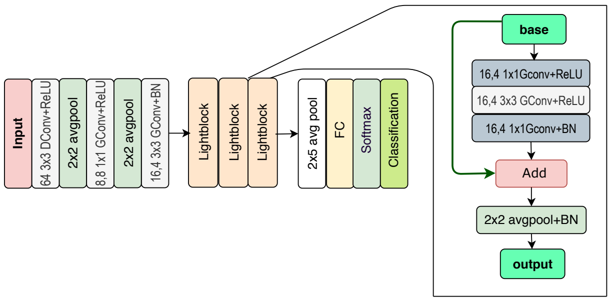

Fig. 1. The proposed CNN architecture, which comprises of multi-layers input and output module with a cascade of Lightblock embedded between input and output modules.

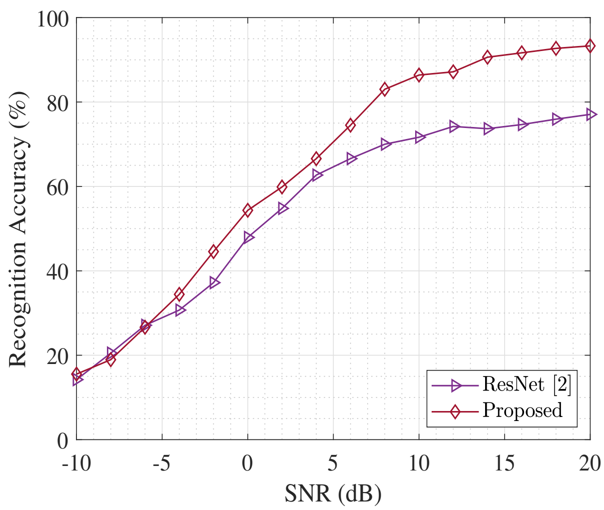

Fig. 2. Comparison of methods in 12 digital modulation schemes recognition.

horizontally. Then, a cascade of grouped point convolutional, 2×2 average pooling, and 3×3 convolutional layers is applied for further spatialtemporal feature learning and feature map down sampling. In order to extract deep feature at low cost, we have designed the Lightblock comprised of three grouped point convolutional, average pooling, and additional layers. With regard function of the layers, the combination of point convolutional and 3×3 convolutional layer learn accurate spatialtemporal features at low cost due to the computational efficiency of grouped convolutional layers. To allow the learned features to operate in multiple levels a skip connection is applied from the Lightblock input to an additional layer. After the cascade of Lightblock, average pooling layer is deployed for further feature maps down sampling. The accurate labels of modulation schemes are predicted through the combination of fully connected layer, softmax and classification layer. It is worth noting that as the number of Lightblock increases, more spatialtemporal features can be learned, but we have applied three Lightblocks to obtain an excellent tradeoff between accuracy and complexity.

IV. Results and Discussion

Investigate the effectiveness of our architecture with regards to classification accuracy and complexity, we trained the network from scratch on a reduced version of RadioML 2018A dataset (i.e., 16APSK, 32QAM, 32APSK, 8ASK, BPSK, 8PSK, 4ASK, 16PSK, 64APSK, 128QAM, 64QAM, and 16QAM). The SNR values are distributed from -10 to $+20$ dB in the interval of $+2$ dB. Also, we set 0.001 and 60 as learning rate and maximum number of epochs, respectively. To optimize learning process, we used stochastic gradient descent. Moreover, the system was designed and implemented in MATLAB 2019b. The results in Fig. 2 shows the increase in the recognition accuracy of the architectures with SNR. For example, at -10 dB the accuracy of ResNet and the proposed architecture were 14.23% and 15.53%, respectively. At SNR of $+14$ dB the accuracy increased by 59.45% for ResNet and 75.09% (proposed architecture). It can be seen that our architecture outperformed ResNet [2] in terms of recognition accuracy. Although our model has excellent recognition accuracy, its number of trainable parameters are less $(86, 802)$ compared with ResNet $(236, 344)$.

V. Conclusion and Future Work

The paper proposed a CNN architecture exploited a hybrid of dilated and grouped convolutional layer at the input to learn spatiotemporal generic features more efficiently for accurate AMR. To learn deep features, we designed a cascade comprised of three Lightblocks. Moreover, we verified our model by experiments on an open-source dataset with digital modulation schemes. The experimental results showed the superiority of our architecture over ResNet as it attained >90% at SNR of $+20$ dB. The future work regarding this work will mainly focus on additional experiments to comprehensively verify the effects of increasing Lightblocks. Also, we will perform more detailed experimental analysis on the complexity according to memory, inference time, and processing speed.

Acknowledgment

This work was supported by Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2018R1A6A1A03024003).

References

[1] L. Horwitz, "The future of IoT miniguide: The burgeoning IoT market continues," Available at https://www.cisco.com/c/en/us/solutions/internet-of-things/future-of-iot.html [Accessed July 6, 2020].

[2] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.

[3] T. Huynh-The, C. Hua, Q. Pham, and D. Kim, "MCNet: An efficient CNN architecture for robust automatic modulation classification," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 811–815, 2020.

[4] A. P. Hermawan, R. R. Ginanjar, D.-S. Kim, and J.-M. Lee, "CNN-based automatic modulation classification for beyond 5G communications," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1038–1041, 2020.

[5] T. Zhang, C. Shuai, and Y. Zhou, "Deep learning for robust automatic modulation recognition method for iot applications," *IEEE Access*, vol. 8, pp. 117 689–117 697, 2020.

[6] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, arXiv:1704.04861. [Online]. Available: https://arxiv.org/abs/1704.04861