Enhanced Automatic Modulation Classification using Deep Convolutional Latent Space Pooling

Clayton A. Harper
Darwin Deason Institute for Cyber Security
Dallas, TX

Lauren Lyons
Darwin Deason Institute for Cyber Security
Dallas, TX

Mitchell A. Thornton
Darwin Deason Institute for Cyber Security
Dallas, TX

Eric C. Larson
Darwin Deason Institute for Cyber Security
Dallas, TX

***Abstract*—In our work, we investigate alternative forms of automatic modulation classification with deep learning and statistical methods. With a growing number of devices communicating through wireless transmission mediums, automatic modulation classification plays a critical role in reading an observed signal. We compare our proposed method with the current state of the art and show that traditional convolutional neural networks can outperform residual neural networks for the task of modulation classification. Using an approach inspired by research from the speaker verification community, we show that the modulation method used to transmit a signal can be classified into one of 24 candidate modulation types with greater than 98% accuracy for signals with high signal to noise ratios.**

I. Introduction

Recognizing the type of signal modulation used to transmit a signal is an important, open research topic in modern communication systems. Real-time classification of modulation types can be applied to "spectrum interference monitoring, radio fault detection, dynamic spectrum access, opportunistic mesh networking, and numerous regulatory and defense applications" [1]. Once a modulated signal has been obtained, the signal must be demodulated in order to understand the transmitted message. Demodulation is the first stage after a signal has been received by a software-defined radio. Systems where the demodulation stage has to be done quickly require that the modulation type is known, and improved modulation classification can increase the throughput of these systems if the modulation type can be identified with a higher degree of accuracy. Therefore, automatic modulation classification (AMC) is currently an important research topic in the fields of machine learning and communication systems, specifically for software-defined radios.

Corgan et al. [2] illustrates that deep convolutional neural networks are able to achieve high classification performance particularly at low signal to noise ratios (SNRs) on a dataset comprising 11 different types of modulation. In [1], Oshea et al. expanded the dataset to include 24 different modulation types and achieved high classification performance using convolutional neural networks—specifically using residual connections within the network (ResNet). With respect to the expanded dataset, the ResNet seen in Table I attained approximately 95% classification accuracy at high SNR values. While [1] found that ResNets outperformed traditional CNNs for this task (see Table II), [3] demonstrates the use of spectrograms and IQ constellation plots as input features to a traditional CNN performs in nearly an equivalent manner as compared to the results obtained by the baseline CNN network in [1]. Further, [4]–[6] also utilized IQ constellations as an input feature into their machine learning models on a smaller scale of 4 or 8 modulation types. Other features have been used in AMC – [7], [8] utilized statistical features and support vector machines while [9], [10] used fusion methods in CNN classifiers.

TABLE I
ResNet Architecture in [1]

Layer	Output Dimensions
Input	2 x 1024
Residual Stack	32 x 512
Residual Stack	32 x 256
Residual Stack	32 x 128
Residual Stack	32 x 64
Residual Stack	32 x 32
Residual Stack	32 x 16
FC/SeLU	128
FC/SeLU	128
FC/Softmax	24

TABLE II
CNN Architecture in [1]

Layer	Output Dimensions
Input	2 x 1024
Conv 1D	64 x 1024
Max Pool	64 x 512
Conv 1D	64 x 512
Max Pool	64 x 256
Conv 1D	64 x 256
Max Pool	64 x 128
Conv 1D	64 x 128
Max Pool	64 x 64
Conv 1D	64 x 64
Max Pool	64 x 32
Conv 1D	64 x 32
Max Pool	64 x 16
Conv 1D	64 x 16
Max Pool	64 x 8
FC/SeLU	128
FC/SeLU	128
FC/Softmax	24

All these previous works focus on enhancing classification performance, but they do not directly explore the required time to classify a signal or network throughput. Tridgell, in his dissertation [11], builds upon these works by investigating these architectures when deployed on resource-limited Field Programmable Gate Arrays (FGPAs). His work stresses the importance of reducing the number of parameters for modulation classifiers because they are typically deployed in resource-constrained embedded systems.

In this work, we explore alternative methods to classify modulation types on radio signals by utilizing the in-phase (I) and quadrature (Q) components of signals using deep learning. No additional feature extraction is applied. Through an approach inspired by X-Vectors [12], we show that the modulation method used to transmit a signal can be classified with a high degree of accuracy while maintaining a compact neural network architecture.

II. Dataset

To evaluate different machine learning architectures, we chose the RadioML 2018.01A dataset that is comprised of the same 24 different modulation types used in [1]. There are a total of 2.56 million labeled signals each consisting of 1024 time domain digitized samples of in-phase (I) and quadrature (Q) signal components. The 24 modulation types are listed as follows: OOK, 4ASK, 8ASK, BPSK, QPSK, 8PSK, 16PSK, 32PSK, 16APSK, 32APSK, 64APSK, 128APSK, 16QAM, 32QAM, 64QAM, 128QAM, 256QAM, AM-SSB-WC, AM-SSB-SC, AM-DSB-WC, AM-DSB-SC, FM, GMSK, and OQPSK. Each modulation type includes a total of 106,496 observations ranging from -20dB to +30dB SNR in 2dB steps for a total of 26 different SNR values.

Also discussed in [1], short radio bursts are likely in many real-world applications due to high scanning antennas, so a classifier must be able to determine the modulation type with relatively few data points. Therefore, we also evaluate the performance of our method with smaller sets of signal sample points representing shorter signal lengths in time.

To evaluate the performance of our method and baseline techniques, we divided the dataset into 1 million different training observations and 1.5 million testing observations under a random shuffle split, stratified across modulation type and SNR. Because of this balance, the expected performance for a random chance classifier is 1/24 or 4.2%. With varying SNR levels across the dataset, it is expected that the classifier performs with a higher degree of accuracy as the SNR value is increased.

III. Proposed Method

We use a convolutional neural network architecture inspired by X-Vectors, first described in [12]. The CNN architecture uses approximately 30% fewer parameters than the ResNet as shown in Table III. Our approach makes use of global mean and variance pooling across convolutional filters. X-Vectors are one method for pooling a latent space temporally using statistical aggregations of the location and spread of the transformed signal. The pooled statistics of the filters are concatenated together and passed through a dense layer to produce the X-Vector. Intuitively, these statistics help to characterize how a filter representation evolves over the sequence including the average filter response and the deviation from the average—this may provide more information based on global attributes of the signal that relate to the modulation type.

Additionally, the proposed architecture maintains a fully-convolutional structure enabling variable size inputs into the network. Using statistical aggregations allows for this property to be exploited. When using statistical aggregations, the input to the first dense layer is dependent upon the number of filters in the final convolutional layer. The number of filters hyperparameter is independent of the length in time of the input into the neural network. Without the statistical aggregations, the output of the preceding convolutional layers is input into the first dense layer. The convolutional outputs are dependent on the length of the signal and are not pooled into a fixed-length. Inputs into the neural network would need to be reshaped to a fixed length in time such that there is not a size mismatch with the final convolutional output and the first dense layer.

Currently, one of the best performing networks is the ResNet shown in Table I employed by [1]. In [1], it was found that a ResNet architecture outperformed a traditional CNN when applied to modulation type classification. Our results indicate an improvement compared to the ResNet approach in terms of classification accuracy at higher SNR values by employing X-Vectors in conjunction with a traditional CNN model.

TABLE III
Proposed CNN Architecture

Layer	Output Dimensions
Input	2 x 1024
Conv 1D (ReLU)	64 x 1024
Conv 1D (ReLU)	64 x 1024
Conv 1D (ReLU)	64 x 1024
Conv 1D (ReLU)	64 x 1024
Conv 1D (ReLU)	64 x 1024
Conv 1D (ReLU)	64 x 1024
Conv 1D	64 x 1024
Average Pooling 1D	64
Variance Pooling 1D	64
Concatenate	128
FC/SeLU	128
FC/SeLU	128
FC/Softmax	24

IV. Results

Using our X-Vector inspired architecture, we were able to achieve a maximum accuracy of 98% at high SNR values. We also replicated the ResNet results from [1], achieving 93.7% accuracy at high SNR values on the same validation dataset. We note that this is slightly less than the reported 95% accuracy reported in [1], likely due to the differences in training and test separation.

Fig.1 is a plot of classification accuracy versus SNR that compares our method and a reproduced ResNet architecture for the same set of data with random chance denoted as the black dotted line. Both architectures follow a similar trend in terms of results; however, the X-Vector approach begins to outpace the ResNet model beginning around a 6dB SNR value. Due to the large size of the dataset, each additional percent of classification accuracy means thousands more correctly labeled modulation types.

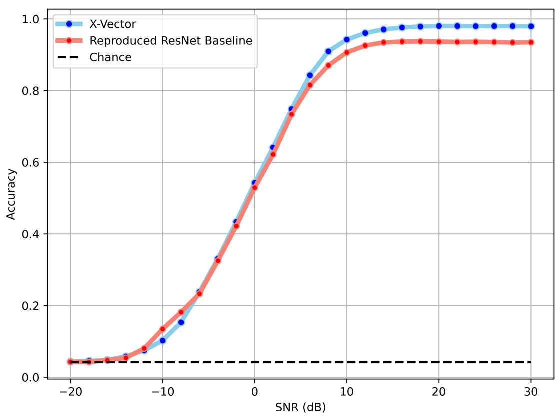

Fig. 1. An overview of the proposed CNN accuracy values for each SNR value in the dataset.

As expected, both classifiers perform better as the SNR value is increased. In signals with a low SNR value, noise becomes more dominant and the signal is harder to distinguish. We do note that in software-defined radio applications a high SNR value is not always a given. With this being said, we still see a significant improvement to random chance even at low SNR values. In systems where the modulation type must be classified quickly, this could become crucially important as fewer demodulator schemes would need to be applied.

Figures 2 and 3 show the resulting confusion matrices for the ResNet architecture and the X-Vector architecture for signals with at least 0dB SNR. We observe a similar structure where the confusion metrics are largest among classes with clusters around the QAM modulation types; however, the X-Vector architecture distinguishes modulation types with a higher degree of precision. Generally, the X-Vector approach has higher values along the diagonal indicating that the predicted label matched the true label which is desired. The ResNet architecture had difficulty distinguishing the difference between 16PSK and 32PSK at the specified SNR range. This is similar to the result found in [1] where short signal lengths are expected to have more error particularly for higher order modulation types "due to lack of information and similar symbol structure using this or any other known prior method."

TABLE IV
MAXIMUM ACCURACY ACROSS SIGNAL LENGTH

Signal Length	X-Vector	ResNet
1024	98.0%	93.7%
768	96.3%	94.7%
512	94.1%	95.1%
128	86.5%	85.0%

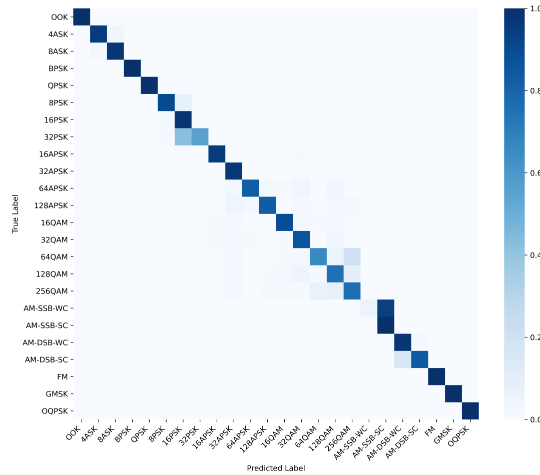

Fig. 2. Confusion matrix across all modulation types on the synthetic dataset at or above 0dB SNR using the ResNet architecture.

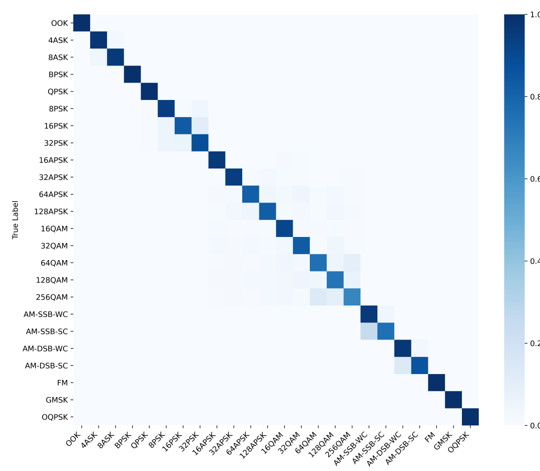

Fig. 3. Confusion matrix across all modulation types on the synthetic dataset at or above 0dB SNR with the proposed X-Vector inspired CNN architecture.

Some operational scenarios, including the use of rapidly scanning antennas on software-defined radios, are at risk of obtaining short bursts of signals due to limited sampling and may be subject to a large variance in the duration of obtained signals. Therefore, we also investigate the performance of the networks using reduced length signals, as shown in Table IV. We observe similar results to those reported in [1] where classification performance significantly degrades for signal lengths of 128 or fewer.

The X-Vector approach's performance degraded in an approximately linear fashion; however, the ResNet approach improved with shorter signal lengths initially. This behavior could be due to the training and test separation. Both approaches maintain fair performance, but the X-Vector approach has the benefit that it handles variable sized inputs organically.

V. Discussion

Utilizing an X-Vector based architecture, we are able to accomplish state-of-the-art classification performance with a traditional CNN. While other works have investigated additional features such as IQ constellations and spectrograms to boost the performance of CNNs, we have shown that traditional CNNs are able to outperform ResNets using raw time-sampled in-phase and quadrature components of a signal while having a smaller network structure.

Our method is able to handle variable sized signal lengths inherently. We are able to maintain a fully-convolutional architecture that enables variable length inputs into the network. This is achieved through the use of aggregated statistics pooling to characterize the latent space of the model. This characteristic has made the X-Vectors proposed by [12] particularly appealing in speech systems where there is high variability in the duration of speech utterances. Using this same recipe in the method proposed in this paper, additional complexities of padding, downsampling, and other input reshaping is not needed (provided that the signal is long enough for the convolutional filters). In an environment where signal bursts are unreliable, a model that handles these imperfections by its nature is beneficial to avoid required data augmentation.

Statistical aggregations in embedding layers, additionally prevents the need of fully re-training additional models for varying length signals. Because the size of the first dense layer is dependent upon the number of filters in the last convolutional layer, the same trained dense layers can be used for signals of varying length in time. If desired, the pre-trained X-Vector model can be fine-tuned for different signal lengths; however, the entire model can be used to initialize the weights of the fine-tuned model. In other approaches discussed, the dense layers are dependent upon the length in time of the input into the network, so dense layers would have to be trained from scratch for signals with different lengths.

VI. Conclusion

Through our research, we show that using an X-Vector approach with a CNN classifier can achieve up to 98% accuracy at high SNR values that additionally yields a 30% smaller model than the state-of-the-art ResNet architecture. Our X-Vector architecture provides improvements in comparison with the ResNet approach while using a significantly reduced number of parameters. In addition to achieving improved accuracy performance, our reduced model size is advantageous for deployment in resource limited devices.

References

[1] T.J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-Air Deep Learning Based Radio Signal Classification," in *IEEE Journal of Selected Topics in Signal Processing*. IEEE, 2018, vol. 12, pp. 168–179.

[2] T.J. O'Shea, J. Corgan, and T.C. Clancy, "Convolutional Radio Modulation Recognition Networks," in *International Conference on Engineering Applications of Neural Networks.*, 2016, p. 213–226.

[3] A. J. Uppal, M. Hegarty, W. Haftel, P. A. Sallee, H. B. Cribbs, and H. H. Huang, "High-Performance Deep Learning Classification for Radio Signals," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019.

[4] S. Peng, H. Jiang, H. Wang, H. Alwageed, and Y. Yao, "Modulation classification using Convolutional Neural Network Based Deep Learning Model," in *2017 26th Wireless and Optical Communication Conference (WOCC)*, 2017, pp. 1–5.

[5] S. Peng, H. Jiang, H. Wang, H. Alwageed, Y. Zhou, M. M. Sebdani, and Y. Yao, "Modulation Classification Based on Signal Constellation Diagrams and Deep Learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 3, pp. 718–727, 2019.

[6] F. Wang, Y. Wang, and X. Chen, "Graphic Constellations and DBN Based Automatic Modulation Classification," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, 2017, pp. 1–5.

[7] C. Park, J. Choi, S. Nah, W. Jang, and D. Y. Kim, "Automatic Modulation Recognition of Digital Signals using Wavelet Features and SVM," in *2008 10th International Conference on Advanced Communication Technology*, 2008, vol. 1, pp. 387–390.

[8] X. Teng, P. Tian, and H. Yu, "Modulation Classification Based on Spectral Correlation and SVM," in *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, 2008, pp. 1–4.

[9] Z. Zhang, C. Wang, C. Gan, S. Sun, and M. Wang, "Automatic Modulation Classification Using Convolutional Neural Network With Features Fusion of SPWVD and BJD," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 5, no. 3, pp. 469–478, 2019.

[10] S. Zheng, P. Qi, S. Chen, and X. Yang, "Fusion Methods for CNN-Based Automatic Modulation Classification," *IEEE Access*, vol. 7, pp. 66496–66504, 2019.

[11] S. Tridgell, "Low Latency Machine Learning on FPGAs," The University of Sydney Australia, 2019.

[12] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur, "X-vectors: Robust DNN Embeddings for Speaker Recognition," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 5329–5333.