Enhanced Automatic Modulation Classification using Deep Convolutional Latent Space Pooling

Clayton A. Harper
Darwin Deason Institute
Dallas, TX

Lauren Lyons
Darwin Deason Institute
Dallas, TX

Mitchell A. Thornton
Darwin Deason Institute
Dallas, TX

Eric C. Larson
Darwin Deason Institute
Dallas, TX

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]. In this work, we explore alternative methods to classify modulation types on radio signals using deep learning. Using an approach inspired by X-Vectors [2], 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 high SNR signals.

II. Previous Works

There has been significant interest in automatic modulation classification in recent years. Corgan et al. [3] 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), [4] 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].

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 [5], 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. Our proposed method places an emphasis on reducing the required parameters to perform signal modulation classification through the use of X-Vectors, while simultaneously boosting accuracy.

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

III. 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.

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%.

IV. Proposed Method

We use a convolutional neural network architecture inspired by X-Vectors, first described in [2]. The CNN architecture uses approximately 30% fewer parameters than the ResNet as shown in Table III. Our approach makes use of global mean and standard deviation 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. 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 as 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

V. 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. 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.

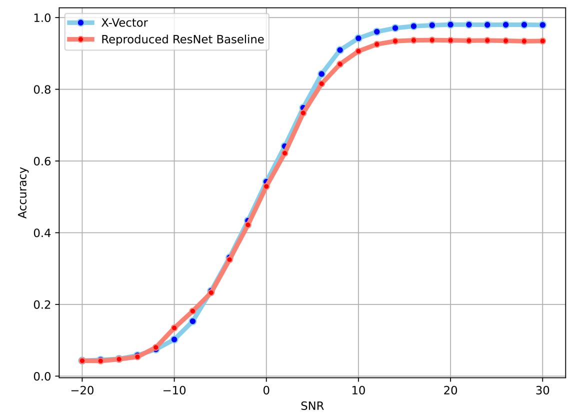

Fig. 1. An overview of the proposed CNN accuracy values for each SNR value in the dataset.

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.

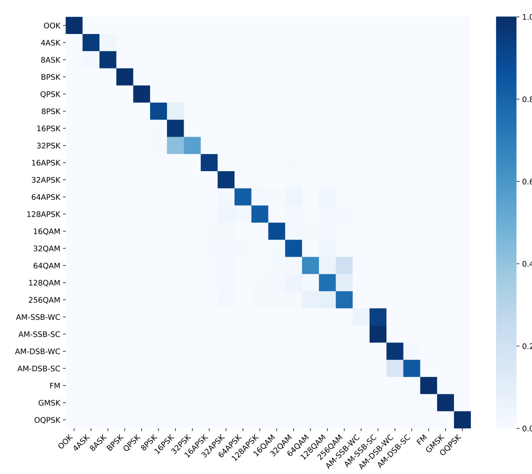

Fig. 2. Confusion matrix across all modulation types on the synthetic dataset at or above 0dB SNR using the ResNet architecture.

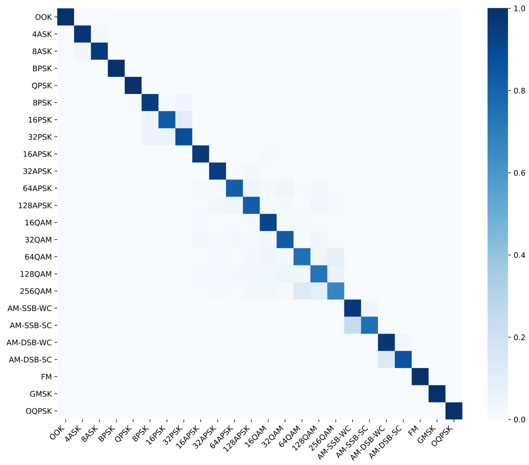

Fig. 3. Confusion matrix across all modulation types on the synthetic dataset at or above 0dB SNR with the proposed CNN architecture.

Due to the limited sampling of rapid scanning receivers, it is possible that relatively few datapoints are observed for a signal. 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.

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%

VI. CONCLUSION

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 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 as compared to [1], our reduced model size is advantageous for deployment in resource limited devices as discussed in [5].

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] 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.

[3] 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.

[4] 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.

[5] S. Tridgell, "Low latency machine learning on fpgas," The University of Sydney Australia, 2019.