AaronNet: Light and Fast CNN for Modulation Classification

Mohammad Chegini[1], Pouya Shiri[2] and Amirali Baniasadi[2]

University of Victoria, Victoria, BC, Canada {pouyashiri,amiralib}@uvic.ca

1 Introduction

Deep Neural Networks (DNNs) have become a commonly used method to analyze raw data in different domains for prediction and classification. DNNs are used frequently in the RF domain for different tasks such as modulation classification , RF interference identification and RF Fingerprint detection . DNNs have several advantages over other methods, including the high generalization ability leading to high accuracy and automatic feature extraction. However, they have their own drawbacks. Designing DNNs is usually a complicated and time-consuming process. In addition, DNNs are computationally expensive and resource demanding.

In this work, we develop AaronNet, a Deep Neural Network (DNN) specialized for Radio Frequency (RF) applications. The proposed network has a very low inference cost and is highly energy-efficient. We test AaronNet on RadioML2018dataset, and achieve a high accuracy. We further prune the network to reduce the inference cost.

2 DNN for Modulation Classification

In the RF domain, some transmitters can freely choose the modulation type of the transmitter signals. However, for the receiver to demodulate the signal, knowledge of the modulation type is necessary. In such systems, Automatic Modulation Classification (AMC), which identifies the modulation type of a received signal, is a way to identify the modulation type based on the incoming signal. Some applications of AMC include spectrum interference monitoring, radio fault detection, dynamic spectrum access and numerous regulatory and defense applications. AMC requires high-throughput and low latency, because the classifier needs the hardware implementation and it is important to reduce the computation cost of the network. This is contradictory with the nature of DNNs, so DNNs have to be further optimized to support the RF domain.

There are various methods for optimizing DNNs. The first type of optimizations include modifying the number of layers or filters used in the network. The second type of optimization consists of the network compression techniques. These methods include layer-wise quantization and pruning unstructued pruning and structured pruning.

3 Problem Statement

The initial problem we consider to test AaronNet, is the problem of reducing the inference cost of classifying the modulation of the RadioML2018 dataset, while maintaining the network accuracy. The minimum allowed accuracy is 56% and the baseline network used as the reference point for reducing the inference cost, is the VGG network. To tackle this problem we applied different network compression techniques, including quantization, unstructured and structured pruning and knowledge distillation. Here we report the results regarding the unstructured pruning applied to AaronNet.

4 AaronNet

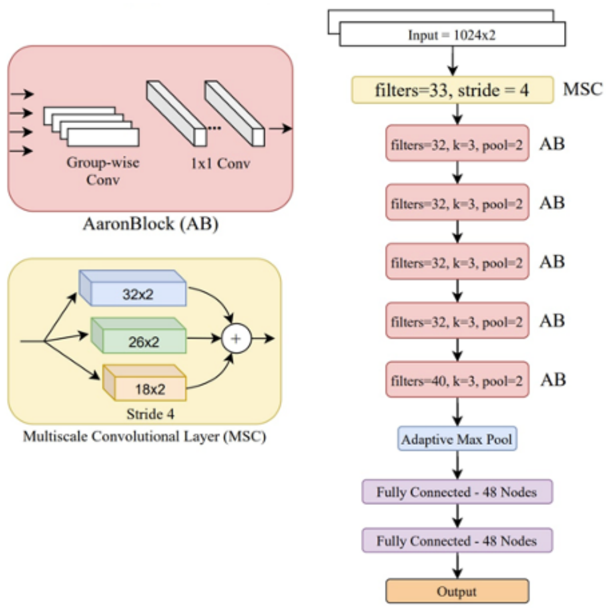

Fig. 1. AaronNet architecture.

In this work we propose AaronNet as a light and fast CNN for the RF domain. Figure 1 shows the architecture of AaronNet. As the figure shows, this network consists of point-wise and depth-wise convolutional layers to reduce the computation cost and to increase the processing power. In addition, we introduce the Multi-Scale Convolutional (MSC) layer to extract features from the RF input signal. The combination of these two, ensures a obtaining a light network with a high generalization ability. Finally, we use an adaptive max pooling to reduce the number of parameters even more before feeding the extracted features to the fully-connected layers.

We propose two configurations for AaronNet: AaronNet32 and AaronNet48, including 32 and 48 filters in each layer. In addition, to further optimize AaronNet, we propose AaronNet+ and AaronNet++ as the compressed variant of AaronNet. AaronNet+ is slightly pruned version of AaronNet while having same or higher accuracy and AaronNet++ is a highly pruned version of AaronNet to ensure obtaining a minimal inference cost.

5 Dataset

To evaluate the performance of AaronNet, we test it on RadioML2018 dataset. This dataset includes over-the-air recordings and simulated channel effects for 24 different analog and digital types. The input signal is 2×1024 in size, that consists of the quadrature and in-phase components of the input RF signal.

6 Results

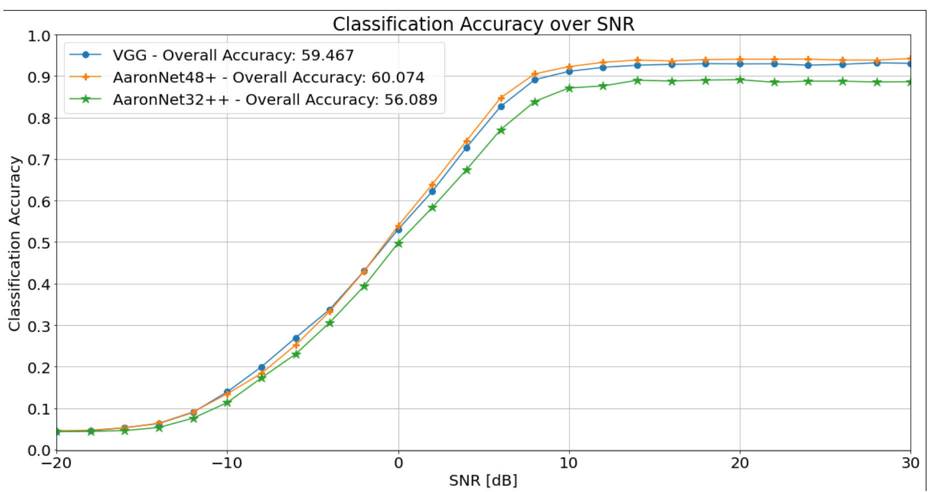

Fig. 2. Classification Accuracy for different SNRs for different combinations of AaronNet, compared to the baseline network (VGG).

Table 1 shows the accuracy and the compression ratio for different variations of AaronNet. The compression ratio denotes the amount of reduction in the inference cost copmpred to the baseline VGG network. As the table shows, with 63.67x, AaronNet32++ obtains the lowest inference cost (highest compression ratio) among other variants. In addition, AaronNet48+ obtains the significant classification accuracy of 60.07%, which is even higher than the baseline accuracy for the VGG network (59.46%).

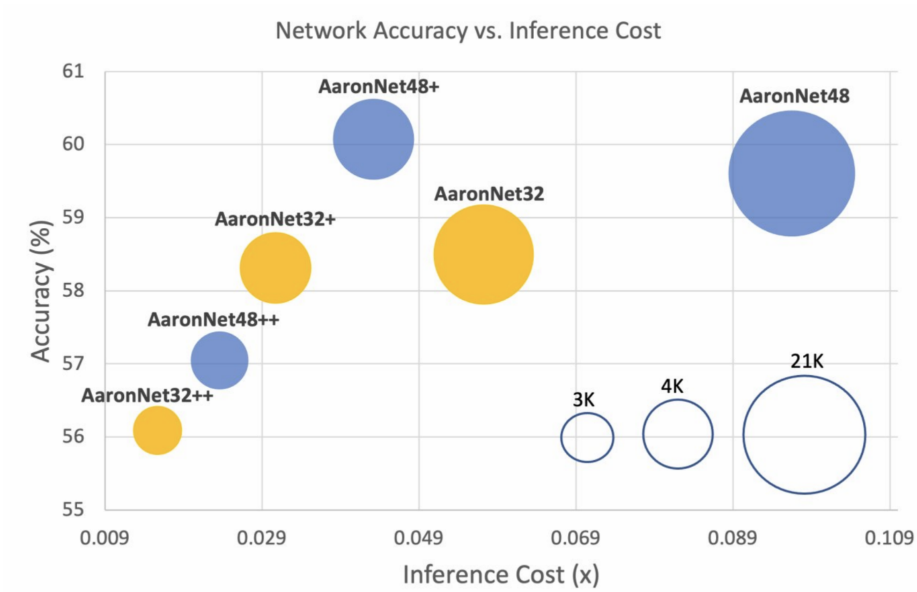

Fig. 3. Classification Accuracy of different combinations of AaronNet with respect to the network inference cost.

Figure 2 compares the network classification accuracy of AaronNet32++, AaronNet48+ and VGG for different Signal to Noise Ratios (SNRs). As the figure shows, AaronNet48+ achieves a higher classification accuracy for a wide range of SNRs, despite having 23.14x smaller inference cost. On the other hand, AaronNet32++ obtains a slightly lower accuracy among the other networks, while having 63.67x smaller inference cost.

Figure 3 shows the classification accuracy of different combinations of AaronNet with respect to the network inference cost. The size of the circles denote the number of parameters used in the network. As the figure shows, with 3K parameters, AaronNet32++ obtains the lowest inference cost among others while maintaing the accuracy over the 56% threshold. On the other hand, AaronNet48+ obtains the highest network accuracy with a very low inference cost.

Figure 4 shows the network compression rate for the different combinations of AaronNet. As the figure shows, all variants of AaronNet provide a high compression ratio. AaronNet32++ improves the inference cost by a factor of 64x and obtains the highest compression ratio among other variants.

Table 1. Accuracy and Inference of different AaronNet combinations.

Network	**Compression Ratio**	**Accuracy**
AaronNet32	17.47x	58.48%
AaronNet32+	32.56x	58.31%
AaronNet48	10.36x	59.60%
AaronNet48+	23.14x	**60.07%**
AaronNet48++	42.37x	57.04%
AaronNet32++	**63.67x**	**56.07%**

7 Conclusion

In this work, AaronNet is designed as a light and accurate network for the RF modulation classification problem. The network is tested on the RadioML2018 dataset, and compared to the VGG network, it has a significantly lower inference cost while obtaining a higher accuracy. AaronNet is a highly potential network for deep learning applications in the RF domain.

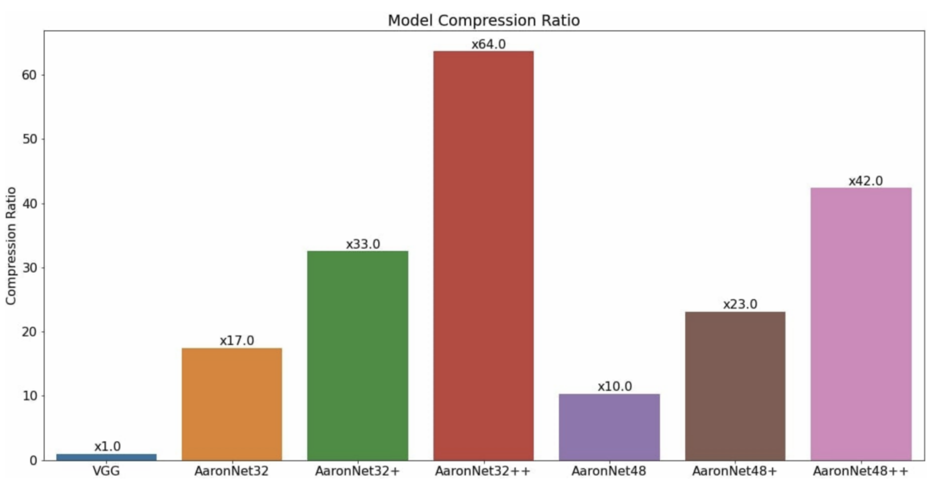

Fig. 4. Network compression ratio for the different combinations of AaronNet.