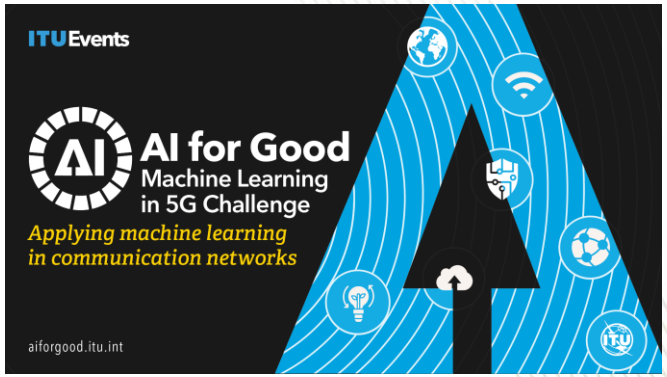


Lightning-Fast Modulation Classification with Hardware-Efficient Neural Networks

10/27/2021

The A(MC) Team:

Jakob Krzyston

PhD Student @ GT, Research Engineer, GTRI
jakobk@gatech.edu

Dr. Rajib Bhattacharjea

Principal Engineer, DeepSig Inc
raj@deepsig.ai

Dr. Andrew Stark

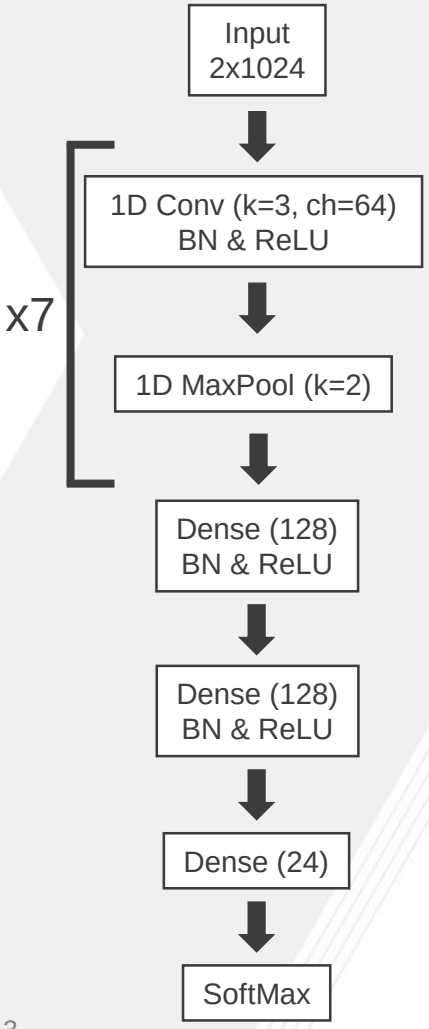
Senior Research Engineer, GTRI
andy.stark@gtri.gatech.edu

Possible Approaches

- ~~**Modify the provided architecture [Speed]**~~
- **Reduce the quantization with Brevitas [Speed]**
  - Four bits for both weights and activations
- **Prune weights [Speed]**
  - L1 unstructured Iterative Magnitude Pruning (IMP)
  - Prune when accuracy threshold reached
- **Adjust training paradigm [Accuracy]**
  - Learning Rate Scheduler → Reduce LR on Plateau

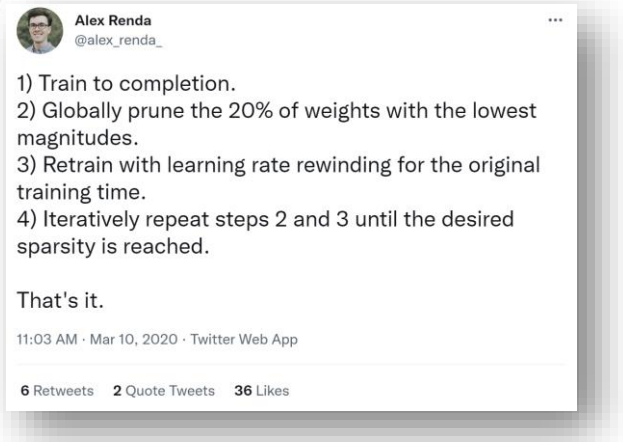
Methods

Architecture



IMP (Simplified)

```
for num_prune_iterations:
    for num_epochs:
        train model
        test model
        if model_accuracy > 0.56:
            save model weights
            prune 20% of weights
            break
        else:
            lr_scheduler.step
```



Compression Summary

Quantity	Original*	Final
Bit Ops	807,699,904	24,436,576
Weight Bits	1,244,936	68,072
Compression	1x	9.313x
Sparsity	0%	89.26%

*Values from provided code

Notes

- Did not reset LR scheduler
  - May have reduced our end performance
- Sparsity % = 1 – (0.8 ^ 10)
- Compression = 1 / (0.8 ^ 10)

Georgia Tech Research Institute

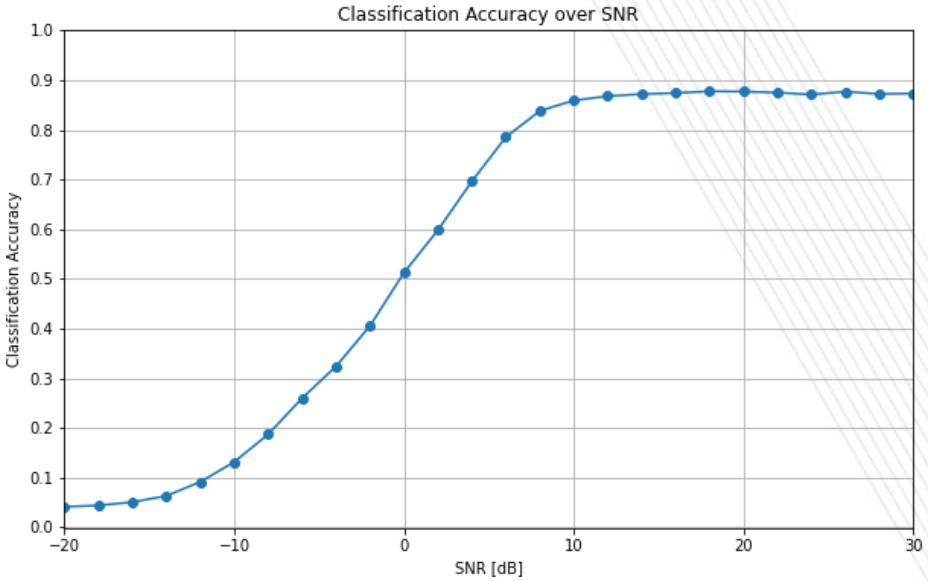
Final Results

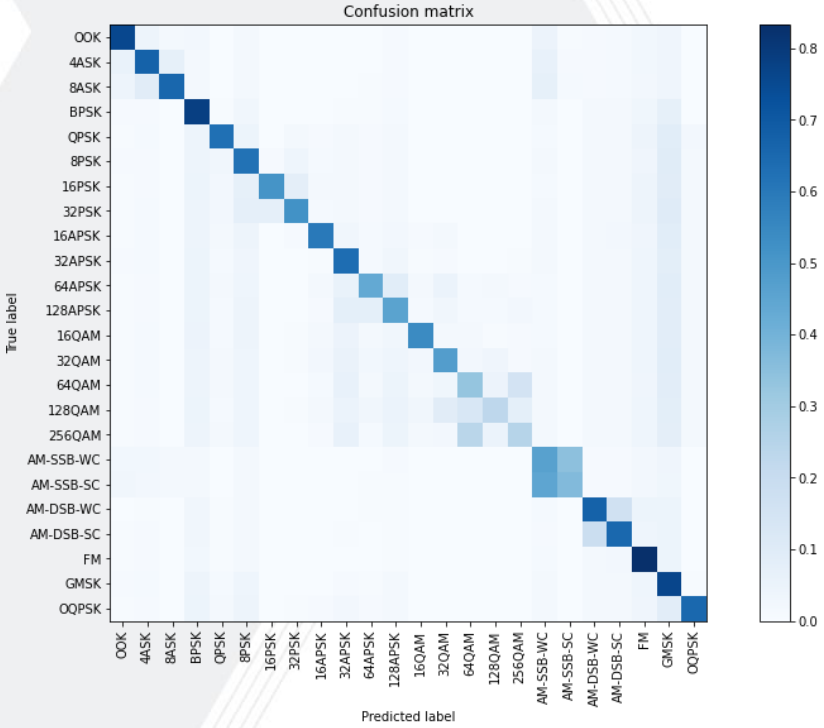
Inference Cost Score:

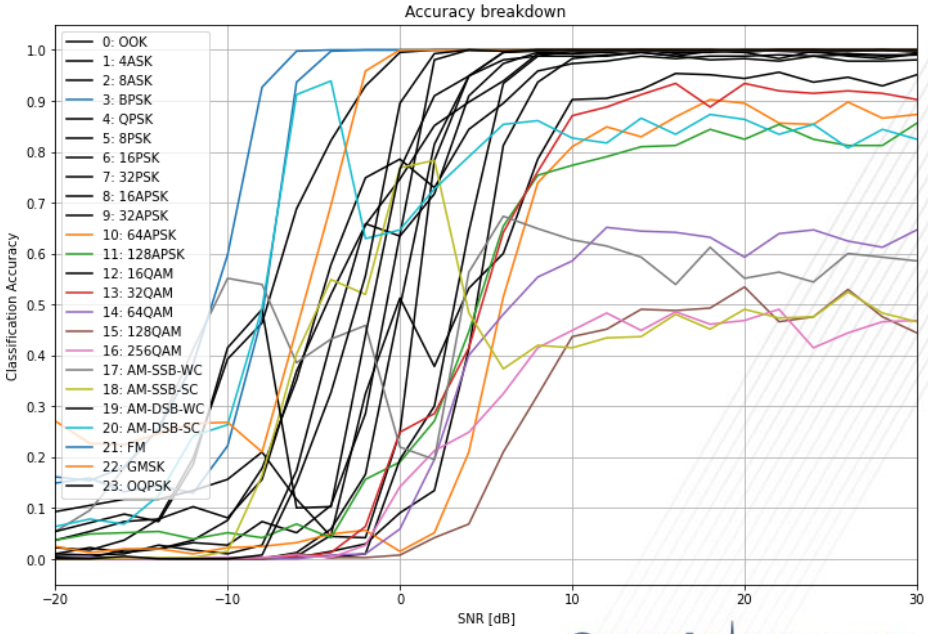
- 0.042467

Overall Test Accuracy:

- 0.5625







Georgia Tech Research Institute