Open-Source Large Scale Radio Frequency Machine Learning Dataset, Toolkit, & Models

Technical POC:
Luke Boegner
luke.boegner@peratonlabs.com

Team Members
Luke Boegner, Garrett Vanhoy, Manbir Gulati, Phil Vallance
Bradley Comar, Craig Lennon, Rob Miller, Silvija Kokalj-Filipovic, Dresden Feitzinger

Agenda

- Narrowband RFML
  - Sig53 Dataset
  - TorchSig
  - Signal Classification Results
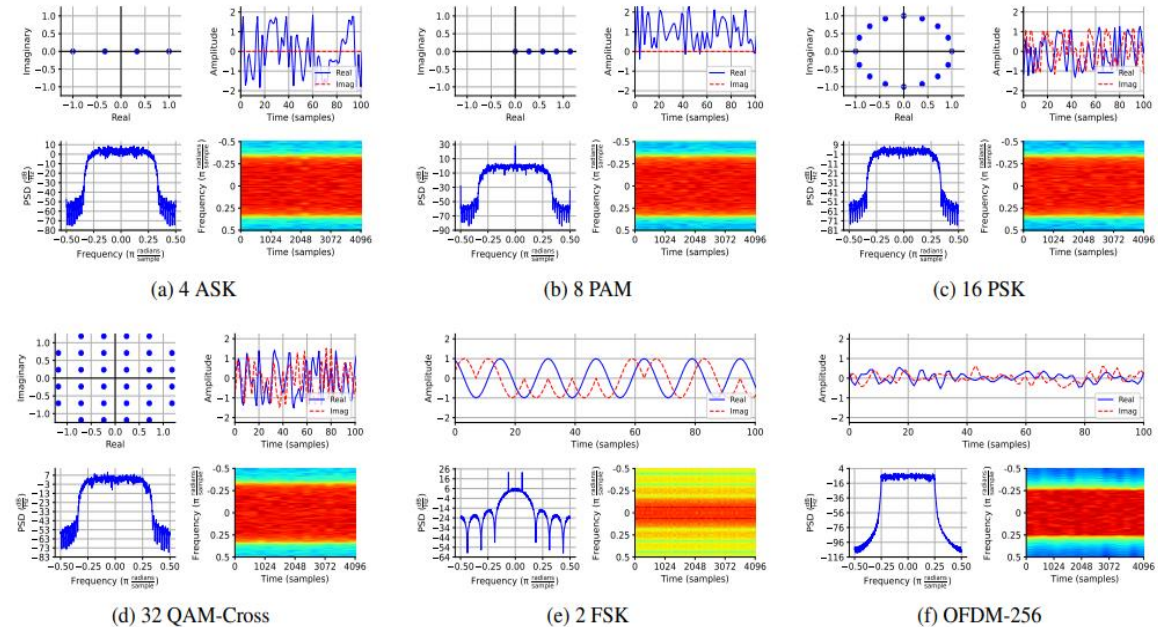
Peraton | LABS

LABORATORY FOR
TELECOMMUNICATION
SCIENCES

Narrowband RFML

Sig53 Dataset

- RadioML introduced in 2016 & 2018
  - Seminal RFML dataset
  - ML model research outpacing dataset capacity

- Introducing the Sig53 Dataset
  - Over 5M examples, split into
    - *Clean Training*
    - *Clean Validation*
    - *Impaired Training*
    - *Impaired Validation*
  - 53 signal classes over 6 modulation families
  - Increased impairment diversity
  - Open-source generation code makes Sig53 easily reproducible, reviewable, and customizable

(a) 4 ASK   (b) 8 PAM   (c) 16 PSK

(d) 32 QAM-Cross   (e) 2 FSK   (f) OFDM-256

LABORATORY FOR TELECOMMUNICATION SCIENCES
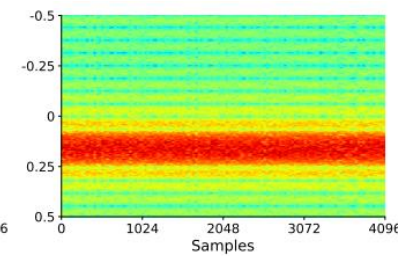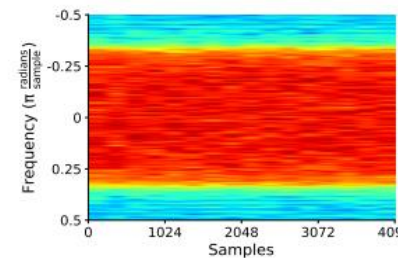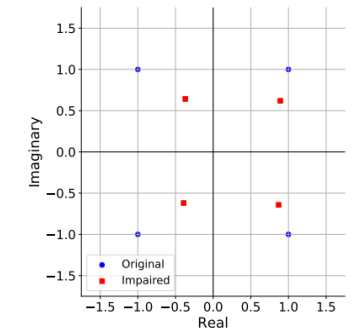
Peraton | LABS

TorchSig

- Open-source signal processing machine learning toolkit

- Python & PyTorch-based toolkit, mirroring best practices in the vision and speech domains

- Capable of generating the Sig53 dataset, wrapping the RadioML datasets, and building custom RF datasets

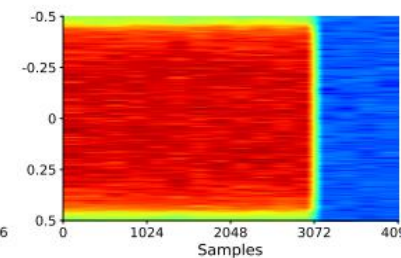- Over 50 domain-tailored augmentations and transformations
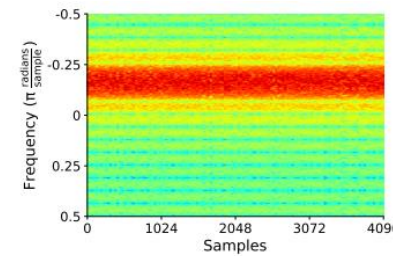
Signal Classification Results
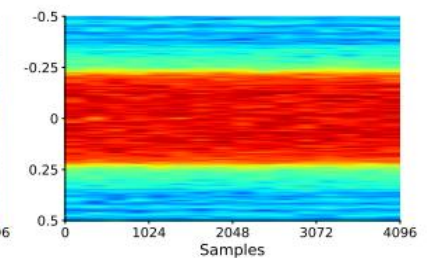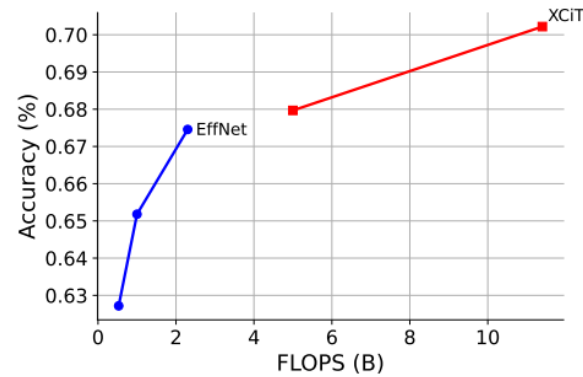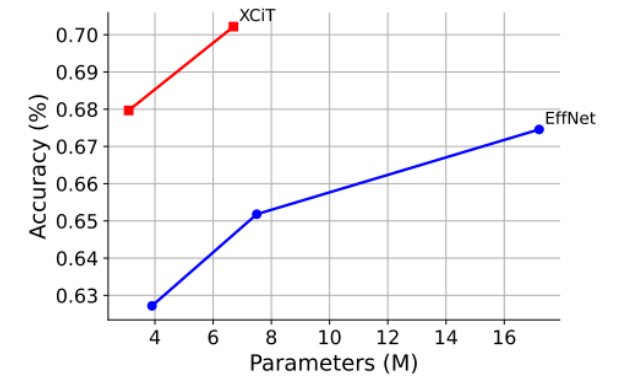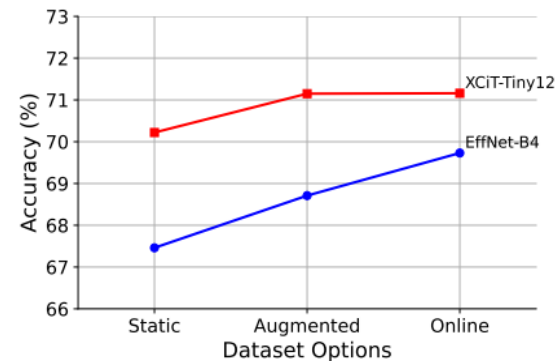
- Trained EfficientNet and XCiT on Sig53, resulting in roughly 60-70% accuracies

- Using TorchSig, demonstrated accuracy improvements from augmentations as well as online data generation
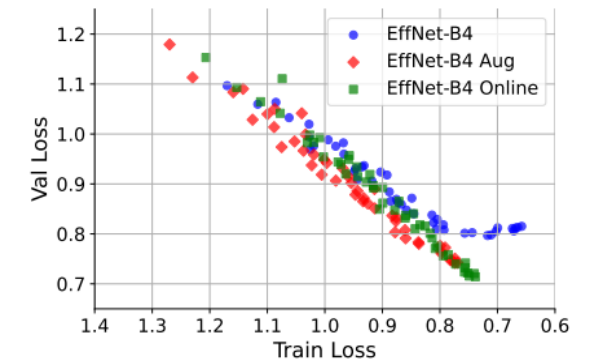

(a) Accuracy vs FLOPS

(b) Accuracy vs Number of Parameters

(a) Accuracies for static, augmented, and online data

(b) EfficientNet static, augmented, and online loss

LABORATORY FOR TELECOMMUNICATION SCIENCES

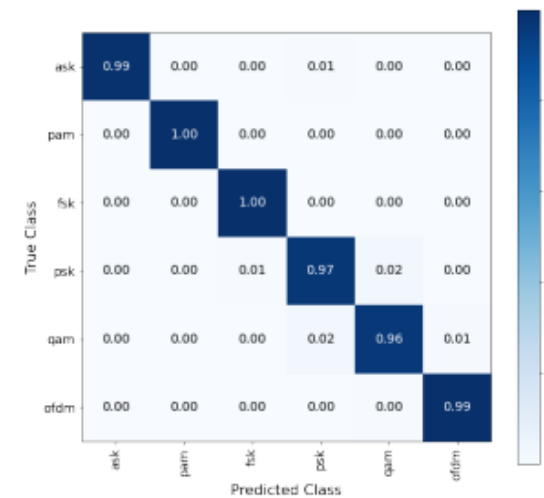Peraton LABS

Signal Classification Results

- Accuracy vs SNR reveals model scaling improvements happen primarily at higher SNRs, meaning differentiating signal classes in Sig53 is sufficiently challenging

- Accuracies across modulation families reveal FSK signals are the easiest signals to classify, while QAM is the hardest

- We also observe very high accuracy across modulation families, meaning most errors occur from intra-family confusion
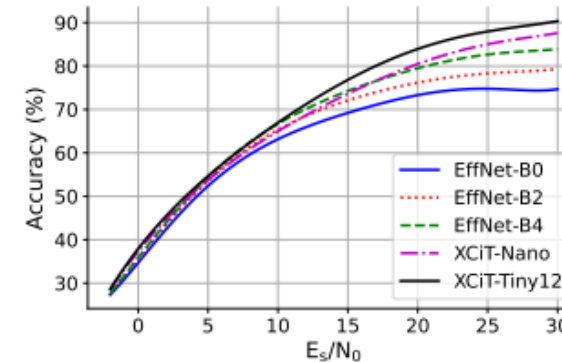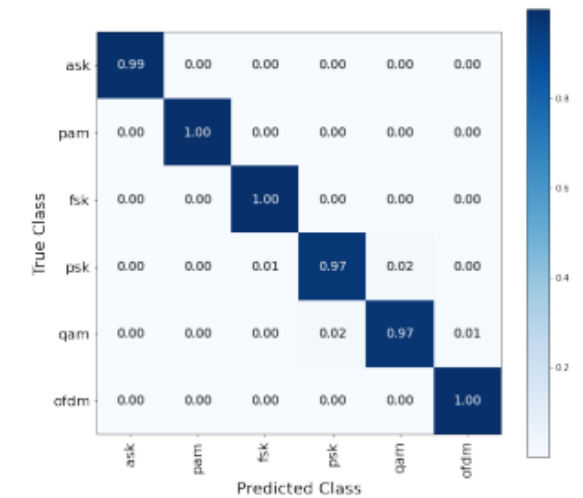

(a) Accuracy versus SNR
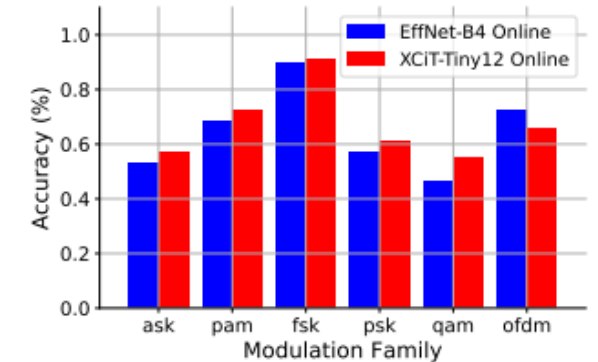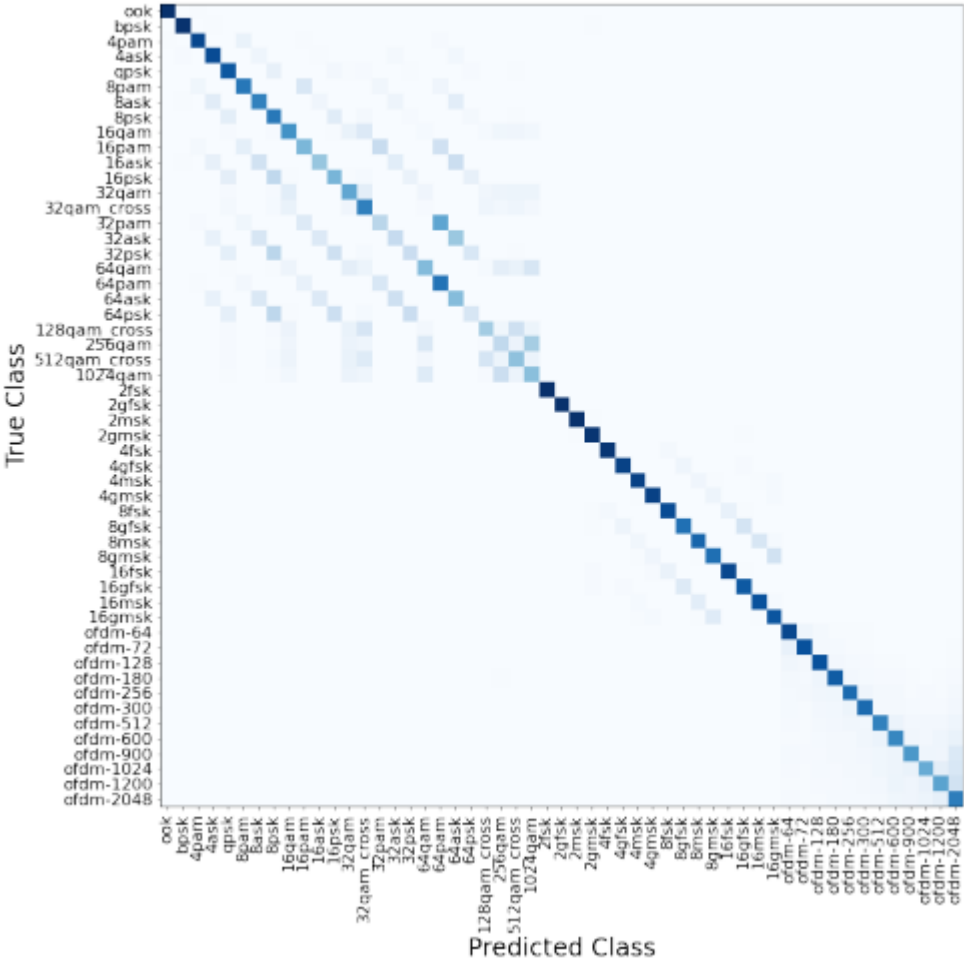

(b) EfficientNet and XCiT modulation family accuracies
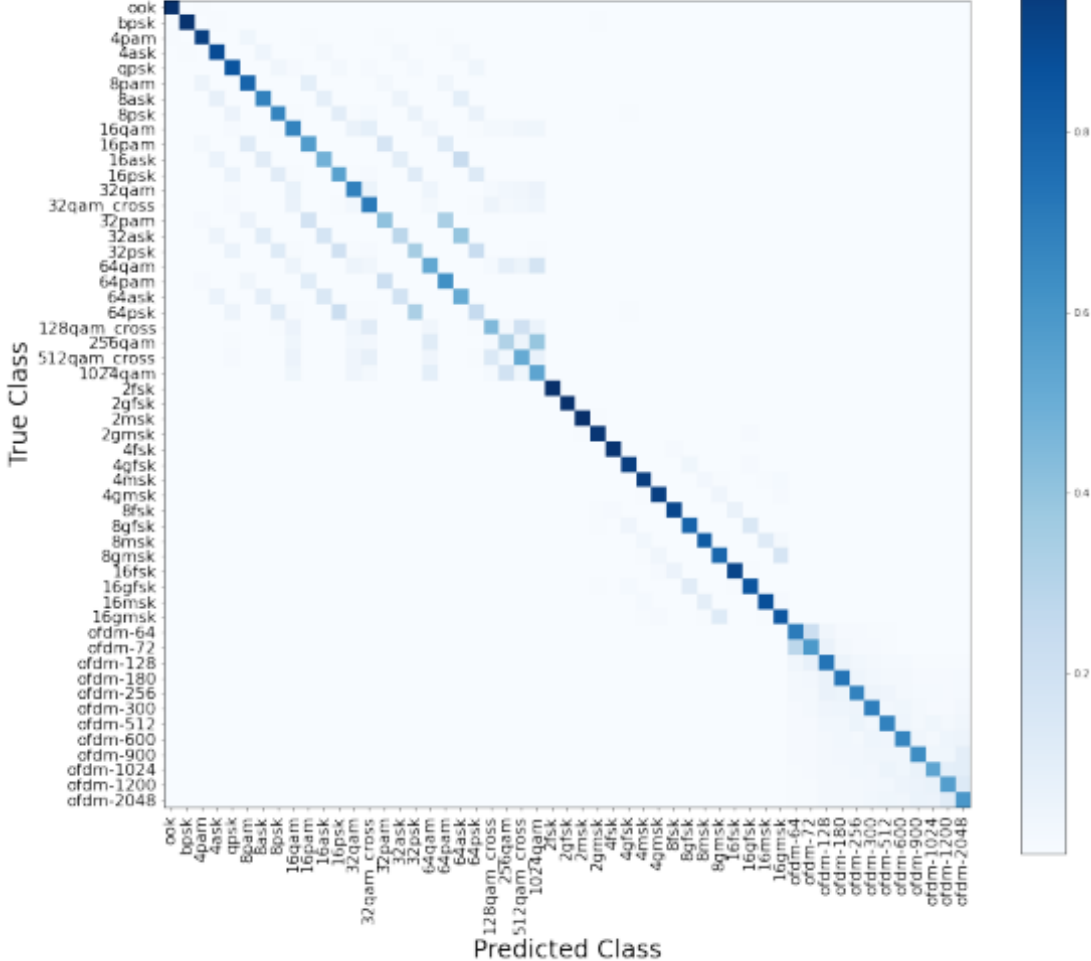

(c) EfficientNet-B4 modulation family confusion matrix

(d) XCiT-Tiny12 modulation family confusion matrix

LABORATORY FOR TELECOMMUNICATION SCIENCES

Peraton | LABS

Signal Classification Results

EfficientNet-B4

XCiT-Tiny12

Conclusions

Conclusions

- Community Challenges:
    - Beat our performance on Sig53 classification
    - Share results back with community

- Find out more at our TorchSig Workshop on Wednesday
    - Additional information also available at: torchsig.com

- We're hiring!
    - Come chat with any of our team to find out more!

- Acknowledgements
    - This work was funded by the Laboratory for Telecommunication Sciences

Peraton | LABS

Questions?