Blind Signal Characterization: Transformers, Triplet Losses and beyond

Srivatsan Rajagopal
srajagopal@ucsd.edu

Radhika Mathuria
rmathuria@ucsd.edu

Dinesh Bharadia
dineshb@ucsd.edu

***Abstract*—In this work, we report on progress in building a machine learning (ML) algorithm to blindly infer the signal modality of anomalous wireless signals. The system built is designed to be robust to hardware impairments like carrier frequency offset (CFO), sample frequency offset (SFO), wireless channel, sample rate changes due to radio resampling etc. The main novelty of our work is the exploration of metric learning methods for the task of blind modality/modulation classification using cyclostationary features. We describe how the ML approach evolved, with an empirical illustration of improvement in classification accuracy.**

I. Introduction

The emergence of highly affordable software-defined radios (SDRs) [1], [2], [3] combined with the softwarization of communication protocols has simplified the process of constructing bespoke communication networks [4]. These networks could potentially serve in exfiltrating confidential information from secured environments, such as private residences or governmental facilities. Furthermore, they could be employed to covertly surveil individuals, utilizing devices like cameras, microphones, and modems [5], [6]. Additionally, they enable the establishment of clandestine and unauthorized communication channels, posing a substantial risk to privacy and security. Identifying and characterizing non-protocol compliant wireless transmissions, (christened as RF anomalies in this paper), is a challenging task, because the spectral environment in which these transmissions exist is extremely busy with other standard protocol-compliant transmissions. From a practical perspective, this problem of anomalous signal characterization is important to solve for diverse spectrum sensing applications. Due to non-standard compliance we cannot use traditional decoders, to detect such RF anomalies.

In addition to non-compliant signals, anomalies can also be generated in a much more innocuous manner than just varying the signal parameters. Altered signals, for example, are protocol-compliant signals that can be decoded by standard decoders, but convey extra information by additional modifications to the signal modulation (for example, see [7]). Such transmissions can be used for covert communication and are even more difficult to intercept by conventional means, and a robust blind signal characterization system such as the one developed here, will be highly useful to intercept such transmissions.

To effectively identify such RF anomalies, we need to characterize all the emissions in the spectrum. To detect such threats, we can use SDRs to continuously scan the spectrum and we must first detect and isolate them in time and frequency, following which a characterization system can be applied to them to detect non-protocol compliance. For spectrum sensing scenarios when precise time-frequency boundaries of the signals are unknown, signal detection and isolation serve as a crucial pre-processing step. We apply a recently proposed detection system, Searchlight [8], which is a robust detector specifically designed for blind spectrum sensing scenarios. It is important to note that [8] is a signal detection system, and does not perform characterization to know if a signal is an anomaly, and therefore, is used as a pre-processing step in our work. We depict the overall flow of the classification architecture as designed in this paper in Figure 7

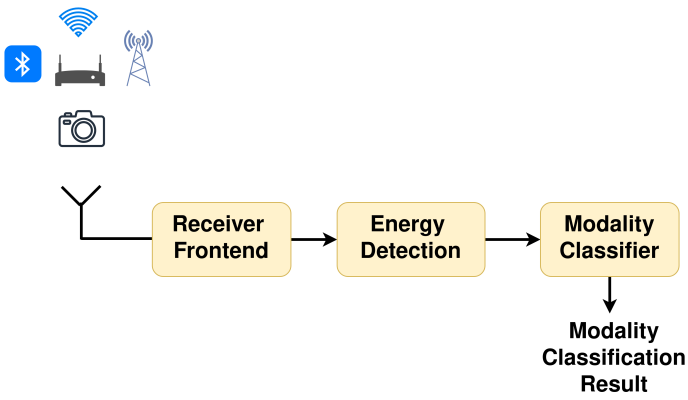

Fig. 1. Workflow of the modality neural network architecture in conjunction with capturing IQ samples and pre-processing with the energy detector

We develop techniques to build a machine learning (ML) model that can classify *signal modality*. Modality, for the purposes of this paper, is the primary information transport method in the signal, and can be one of the following three possible labels, namely,

- spread spectrum
- multi carrier
- single carrier

II. Background and Motivation

A. Prior Work

Automatic Modulation Recognition, using deep learning architectures, has a fairly large literature. See [9], [10], [11], [12] and references therein for previous attempts. In this

section, we present a comparison of our work with the previous literature.

In [9], [10], simple convolutional neural network architectures were used to perform ML-based signal characterization on data drawn from the RadioML dataset. This dataset is completely simulated, in the sense that noise and channel effects are synthetic. In addition, there is no pre-processing step similar to searchlight, that we are using in our work. In [11], a transformer-based architecture was used for automatic modulation recognition, using the RadioML dataset for training. Further, the features used were simple transformations of the input IQ samples. The lack of cyclostationary input features (see [13], [14], [15], [16] for more details) means that this approach will not be robust against sampling rate changes induced by the energy detection preprocessing step. In [12], metric learning-based losses were explored for modulation recognition, but this approach is again plagued by non-robustness due to the absence of cyclostationary features. In summary, while the problem of signal characterization has been attempted in the literature before using methods that have some overlap with the approaches of this paper, three things, in our view, complicate the exact problem we are trying to solve in this work, namely

- The signals we are concerned with have *anomalous* parameters, so that protocol alone would not be sufficient to characterize them.
- The classification is completely blind: we are presented merely with an IQ collect from an arbitrary band, with no other prior information, except for the received sampling rate.
- The energy detection pre-processing outputs IQ samples at sample rates that are different from one signal to the next.

It further needs to be stressed that, to the best of our knowledge, there exists no classification algorithm in the literature that attempts to infer modality/modulation completely blindly. We have also chosen the input features are chosen to be robust against sampling rate changes due to the detection preprocessing step. As far as the novelty of the classification architecture is concerned, while transformer neural networks, and metric learning methods have been used in the past for modulation recognition, (see [11] and [12] respectively), the current system makes use of the spectral correlation function (SCF) [14] that will be estimated exhaustively using the Strip Spectral Correlation Analyzer (SSCA) algorithm, (see [15] for an introduction). This needs to be contrasted with the previous approaches which used the raw IQ as input features.

B. Primer on Cyclostationarity and Feature Computation

The SSCA output estimates the SCF at discrete points in the bifrequency (f, α) plane. The definition of the SCF is given as

$$S_x^{\alpha}(f) = \lim_{\Delta f \to 0} \lim_{\Delta t \to \infty} \mathcal{O}(t, f, \alpha, \Delta t, \Delta f), \qquad (1)$$

where,

$$\mathcal{O}(t, f, \alpha, \Delta t, \Delta f) = \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} \Delta f X_{1/\Delta f}(t, f + \alpha/2) X_{1/\Delta f}(t, f - \alpha/2)^*.$$

Here, $X_{1/\Delta f}(t, f)$ represents the spectral content of the signal $x(t)$ in a passband centered at f with bandwidth Δf and is called the complex demodulate of $x(t)$. More precisely,

$$X_{1/\Delta f}(t, f) = \int_{t - \frac{1}{2\Delta f}}^{t + \frac{1}{2\Delta f}} x(u) e^{-2i\pi f u} du. \qquad (2)$$

Communications signals with well-defined modulation/modality exhibit peaks at characteristic cycle frequencies α. While the precise location of these peaks depends on external factors too, (e.g., if the channel effects are bad, some of the peaks may disappear), it is empirically observed that the pattern of peaks is more or less common across signals of the same modality. To estimate the SCF, we use the SSCA algorithm, whose computation proceeds as follows. Suppose that an analog signal $x(t)$ is sampled at a rate F_s, giving a discrete sequence $x[n]$. If we look at the discrete Fourier transform of the sampled signal, it is immediately clear that the sampled version has frequency components going from $-F_s/2$ to $F_s/2$. Since the SSCA acts on discretely sampled signals, it is clear from Eq.(1), that the resulting estimated SCF occurs at points (f, α) such that

$$-\frac{F_s}{2} \leq \alpha \pm 2f \leq \frac{F_s}{2}. \qquad (3)$$

The SSCA algorithm proceeds in the following fashion. We first fix two integers N, N_p such that the length L of the data record whose SCF needs to be determined satisfies (with padding if necessary) $L = N + N_p$. The larger number N is related to the resolution of the analyzer in the α-direction as $\Delta\alpha = \frac{F_s}{N}$. The smaller number N_p is related to the resolution in the f-direction as $\Delta f = \frac{F_s}{N_p}$.

- The complex demodulate of $x[n]$ is estimated as

$$X(n, f_k) = \mathrm{e}^{-2\pi j n f_k} \sum_{r=0}^{N_p - 1} a(r) x(n + r) e^{-2\pi j r f_k}. \quad (4)$$

  Here, f_k are the centered FFT frequencies of an N_p point FFT. The window $a(r)$ is a data-tapering window of length N_p.
- Next, consider $g(n)$, another data-tapering window of length N. One forms the quantity

$$S(f_l, f_k) = \sum_{n=0}^{N-1} g(n) X(n, f_k) x^*(n + N_p/2) e^{-2\pi j n f_l} \qquad (5)$$

  Here, f_l are the centered FFT frequencies of an N point FFT. This matrix $S(f_l, f_k)$ has dimensions $N \times N_p$ and represents the SCF estimates evaluated in a rotated coordinate system in the (f, α) plane.
- The mapping between (f_l, f_k) and (f, α) is given by

$$\alpha = f_l + f_k \quad f = 0.5(f_k - f_l) \qquad (6)$$

III. Our Solution

To reiterate the problem statement, we are presented with IQ samples that represent the output of the energy detector. The problem is to assign modality classification labels to these samples. The only information that is presented is the sampling rate of the samples.

A. Searchlight Processing

The searchlight system chunks the input, then convolves the input channogram with boxes of various sizes. A separate noise floor is estimated for each chunk and an energy detection is run on each convolution output. This results in time frequency bounds where a non-trivial signal is present. The complex samples corresponding to these bounds are synthesized (using the polyphase channelizer).

B. Imperfections introduced by searchlight

Searchlight does not perform perfect centering of the detected energies in the region of interest. Further, the occupied bandwidth of the energy will be much smaller than 50% of the size of the energy box. To mitigate these problems,

- We choose to center the energy by applying an appropriate phase offset.
- We increase the occupied bandwidth by performing a resampling operation (this won't result in any aliasing in the cases where we do this).

We observe that the neural network training is much more stable when these operations are done.

C. System Design

The CUDA implementation of the SSCA computation follows the algorithm given in [16]. Since many signals of interest in the dataset have information modulated in their instantaneous frequency, phase, as well as amplitude, and also in how the phase changes between the samples, the following are the six candidate features that we can compute for characterizing a signal :

- SCF estimate using the SSCA,
- Conjugate SCF using the SSCA,
- SSCA of the phase of the input,
- SSCA of the samplewise phase difference of the input,
- SSCA of the absolute value, and
- STFT of the input.

Some representative examples of features are given in Fig. 2. In a later section, we show evaluations to justify our choice of features.

D. Evaluation plan summary

We want to converge to an *optimal* (in some sense) neural network architecture that provides the best result on testing datasets that are completely different (in terms of signal parameters) from the training set (we call such a dataset an *unseen* dataset). The optimal network would then be a neural network whose performance does not degrade appreciably when evaluated on unseen datasets. We start with a barebones initial baseline model, and study the effects of various modifications to this.

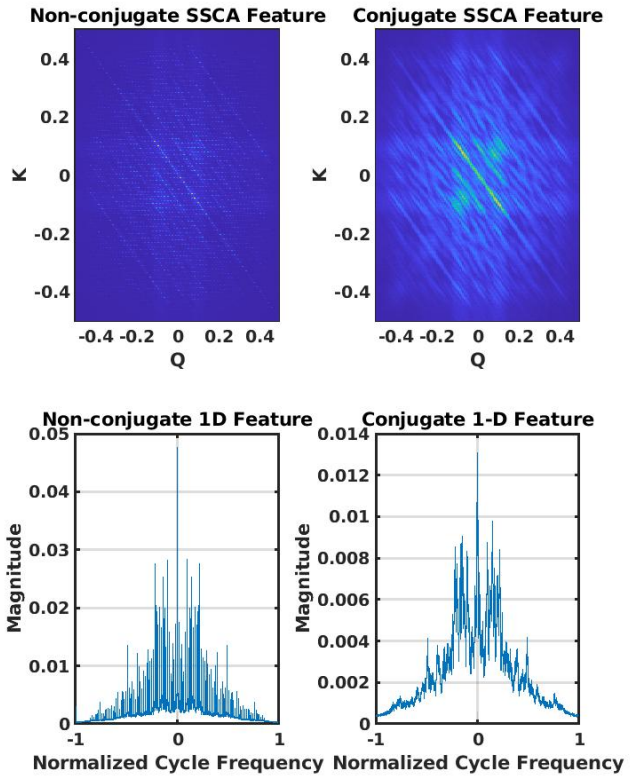

Fig. 2. Example SSCA estimated features for a single spread spectrum 16-QAM signal. The figures on the second row are the one dimensional reductions of the corresponding features in the first row.

E. Baseline Model description

Our baseline model is composed of components that are standard in the ML image classification literature, and consist of simple convolutional networks of various sizes. The input to the neural network consisted of a single SSCA output, evaluated with $N = 256, N_p = 32$. If the testing set signal parameters were very dissimilar to the ones used in the training set, we get the performance given in Figure 3. This shows the following

- The values of N and N_p are not large enough to capture all the cyclostationary signatures of the input signal.
- The neural network architecture needs to be sufficiently complex for the problem.
- Only one SSCA feature is insufficient to perform classification to the accuracy we desire.

F. Transformer architecture

The peaks in the 2d image shown in Fig. 2 exhibit a pattern that is similar for signals of the same modality. We needed some way to make a neural network learn not only the position of peaks in the SSCA image but the context in which they occurred. In machine learning, more specifically machine translation, the problem of translating text from one language to another is an important one. Here, not only is the

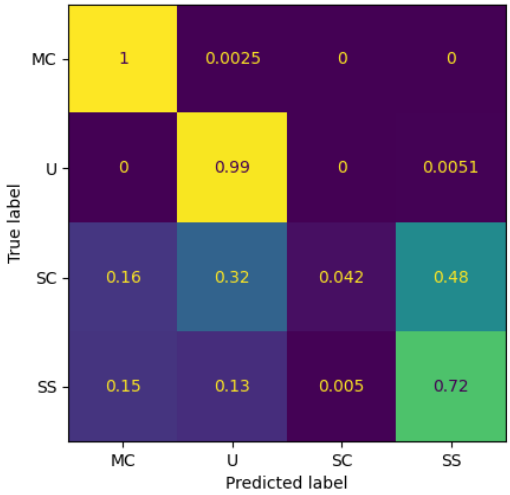

Fig. 3. Baseline Modality Neural network performance on testing set very dissimilar to training set. Single SSCA feature was used as input to a simple CNN. MC=Multi Carrier, SC=Single Carrier, U=Unknown, SS=Spread Spectrum

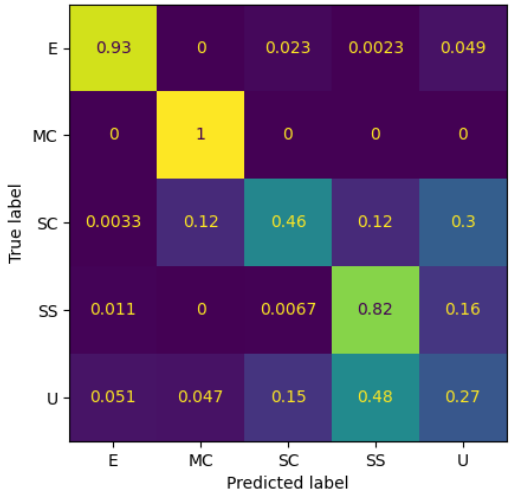

Fig. 4. Transformer Neural network performance on the testing set very dissimilar to the training set. All energies have been passed through the perfect energy detector. Three input features have been computed : SSCA, conjugate SSCA and STFT, with emanations. E=Emanations, MC=Multi Carrier, SC=Single Carrier, U=Unknown, SS=Spread Spectrum

position of each word in the input and target important but also the context in which the word occurs. Recent progress in machine learning in finding effective ways of capturing information about this context showed that the so-called attention mechanism is important for this purpose. An example neural network that automatically implements the attention mechanism is the transformer. To test the viability of this architecture for modality classification, we trained this on data processed with the perfect energy detector, and tested it on a dataset that had parameters that were very different compared to the training set. We used the following three features as input to the transformer,

- SSCA of input
- conjugate SSCA of input
- STFT of input

and we were able to get good performance. To stress the performance of the transformer architecture, emanations instances were included, which are signals collected from unintended transmissions from devices like keyboard, mouse, monitor etc. We tried training and testing the modality neural network to see how it handled this new class. The result is shown in Figure 4. The problem that we are seeing here is that te inclusion of the emanations class has destabilized the behavior of the system for the same input features. The possible reasons for this could be

- Many overt signal modulations have features difficult to distinguish from emanations.
- The three features used here are insufficient to separate the new class from the old ones.

G. Feature Engineering and Loss design

The first problem raised in the previous section, is solved by including metric learning losses as part of the training procedure to incentivize the network to discriminate neighboring classes. To solve the second problem, we decided to include all possible six features as input. The soft triplet loss incentivizes neural networks to learn efficient embeddings of input features so that samples of the same class in the training set are closer than samples of different classes. The simplest way to ensure that the condition above holds is to perform the following steps:

- Define a generalized distance between two points $\mathbf{x}$ and $\mathbf{y}$ as
$$\mathrm{d}(\mathbf{x},\mathbf{y}) = \mathbf{x}^T M \mathbf{y} \tag{7}$$
- During each epoch, minimize the distance defined in Equation 7 between points in each triplet that belong to the same class, and maximize the distance between points in each triplet that belong to different classes. This gives an update rule for the matrix entries of M after each epoch.

Some additional optimizations need to be introduced in this process. To ensure a fast way of comparing points to determine their distance one assigns centers $\mathbf{w}_i^k$, for each class. The index i labels the class, and the index k labels the center in class i. Then, the similarity of a point $\mathbf{x}$ with class j is defined to be

$$\mathcal{S}_{\mathbf{x},j} = \max_k \mathbf{x}^T \mathbf{w}_j^k \tag{8}$$

One can then show that the optimization carried out in the third point above is equivalent to minimising the following

loss

$$L = -\sum_i \log\left(\frac{\exp(\mathcal{S}_{\mathbf{x}_i,y_i} - \delta)}{\exp(\mathcal{S}_{\mathbf{x}_i,y_i} - \delta) + \sum_{j \neq i} \exp(\mathcal{S}_{\mathbf{x}_i,y_j})}\right) \quad (9)$$

We evaluated the performance of the neural network that resulted in Figure 4 with the same training and test sets, with two major differences:

- triplet loss was included in the training as discussed above
- all six features are used as input to the transformer.

When these conditions were satisfied, we get the performance recorded in Figure 5.

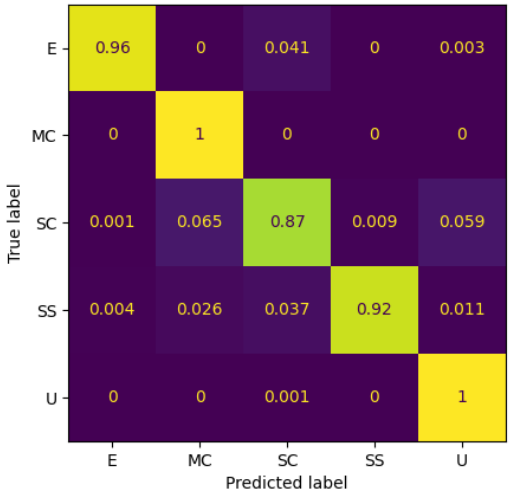

Fig. 5. Transformer Neural network performance on testing set very dissimilar to training set with triplet loss included in training. All energies have been passed through the perfect energy detector. E=Emanations, MC=Multi Carrier, SC=Single Carrier, U=Unknown, SS=Spread Spectrum

IV. Dataset generation and pre-processing

The accurate evaluation of RF ML algorithms heavily relies on the availability of representative datasets. To train and test our algorithms, the following are the key steps we took in order to ensure this:

A. Signal diversity

Each modality can have multiple possible modulations within it. For example, a Phase Shift Keying (PSK) signal can be transmitted as a single carrier, spread spectrum, or multi-carrier signal. A large number of commonly used modulations were hence considered for each modality, which are listed in Table I below.

TABLE I
Modulation Schemes Covered in the Generated Datasets

Single Carrier	PSK, QAM, FSK, CPFSK, GFSK, MSK, GMSK, AM, FM, ASK, APSK
Spread Spectrum	PSK, QAM, FSK, GFSK, MSK, GMSK
Multi carrier	OFDM PSK, OFDM QAM

B. Data diversity

ML algorithms often suffer from the problem of domain gap, and hence in order to address this we ensure we use different kinds of data - simulated as well as over-the-air for training and testing. Four kinds of data were generated:

1) Simulated data + perfect energy detector
2) Simulated data + realistic energy detector
3) OTA data + perfect energy detector
4) OTA data + realistic energy detector

At first, raw, noiseless I/Q samples are generated for each modulation in MATLAB. These are resampled to 100 Msps. After this, to generate the simulated datasets, synthetic noise, channel effects and other RF imperfections (Carrier Frequency Offset, DC Offset) are applied to this raw data. For the OTA data, the noiseless I/Q samples are transmitted and collected using USRP N320s. Hence, each noiseless I/Q file will have a simulated and OTA copy. Following this, the two kinds of energy detectors as described above are applied. Figure 6 showcases this full workflow.

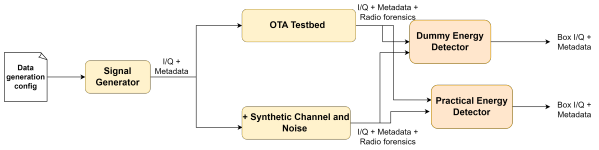

Fig. 6. Data generation and pre-processing workflow

C. Wide parameter sweeps

In order to prevent overfitting of the model to any particular kind of data, the datasets have wide sweeps over all possible tunable parameters such as symbol rate, sample rate, pulse shaping, SNR, samples per symbol etc., along with modulation-specific parameters such as chip rate, bandwidth-time product, frequency deviation etc.

D. Class balance

To ensure an unbiased evaluation of classification algorithms, it is essential to maintain class balance in the generated datasets. Hence, an equal number of files were generated for each modulation and modality in all the datasets.

V. Final results and Discussion

In conclusion, we have the following as the flow chart for the neural network classification architecture.

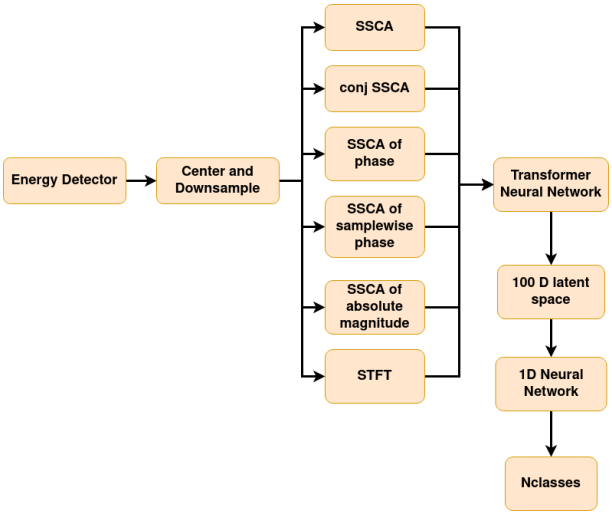

Fig. 7. Flow Diagram for the modality classification architecture

The final neural network architecture was tested in the UCSD lab environment. Signals of various modalities and modulations were transmitted over the air, processed through the Searchlight energy detector, and the IQ samples of the detected energies were given as inputs to the signal characterization block that we have been building. The training dataset was also a set transmitted over the air in a similar fashion, but the signal parameters of the training and test sets were different. When passed through this block, we get the confusion matrix as depicted in Figure 8.

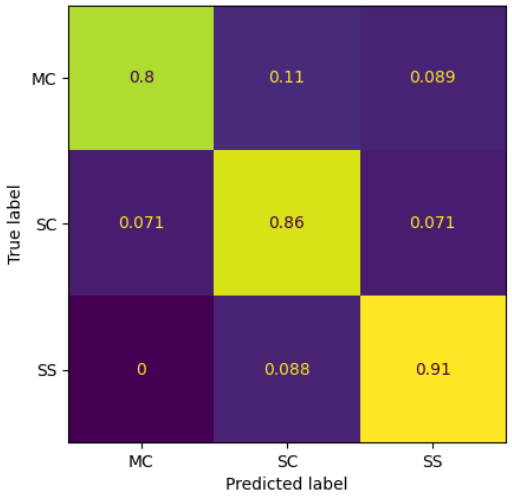

Fig. 8. Modality classification performance on data captured in UCSD lab environment. All six features were used as input, and triplet loss was included in training. MC=Multi Carrier, SC=Single Carrier, SS=Spread Spectrum

Note that this translates to an overall accuracy of $\approx 85\%$ on unseen datasets, and the final performance can be seen in Figure 8. This indicates that there is good separation between the classes, and also shows that this separation is largely insensitive to the sample rate of the output of the energy detector. The triplet loss network maps input features to a latent space. A two-dimensional reduction of these latent space vectors for each class can be seen in Figure 9. This shows a clean separation of the classes, and this behavior persists even across all modulations for a particular modality and across all SNRs. This also indicates that at some level, modality/modulation recognition is well adapted to geometrical approaches, even with channel effects and other imperfections.

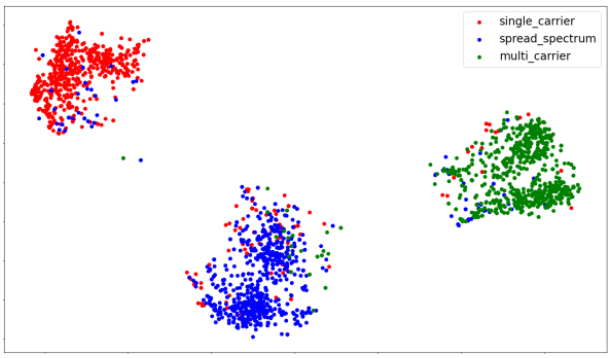

Fig. 9. Embedding diagram created by triplet loss training for the dataset on which the neural network has performance given in Figure 8. Each point here represents an energy of specific ENR and Modulation and is a two-dimensional reduction of a 100 dimensional vector. The x and y coordinates do not have any meaning.

VI. Conclusion

This work demonstrates a machine learning-based system that is capable of blind signal characterization. After the signals have been detected in arbitrary bands, their IQ is collected and analyzed by the neural network to output the modality labels of the signal under consideration. This system has shown to be insensitive to sample rate variations, as well as CFO imperfections introduced by the detector.

VII. Acknowledgements

This section is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via [2021-2106240007]. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References

[1] G. S. Gadgets, "Hackrf one," [Online; accessed 22-November-2023]. [Online]. Available: https://greatscottgadgets.com/hackrf/one/

[2] A. D. Inc., "Adalm-pluto software-defined radio active learning module," [Online; accessed 22-November-2023]. [Online]. Available: https://www.analog.com/en/design-center/evaluation-hardware-and-software/evaluation-boards-kits/adalm-pluto.htmleb-overview

[3] L. Microsystems, "Limesdr," [Online; accessed 22-November-2023]. [Online]. Available: https://limemicro.com/products/boards/limesdr/

[4] GNURadio, "Gnuradio: The free open software radio ecosystem," [Online; accessed 22-November-2023]. [Online]. Available: https://www.gnuradio.org/
[5] IARPA, "Securing compartmented information with smart radio systems (scisrs)," [Online; accessed 22-November-2023]. [Online]. Available: https://www.iarpa.gov/research-programs/scisrs
[6] B. A. Bash, D. Goeckel, D. Towsley, and S. Guha, "Hiding information in noise: fundamental limits of covert wireless communication," *IEEE Communications Magazine*, vol. 53, no. 12, pp. 26–31, 2015.
[7] R. Cogranne, P. Cao, W. Liu, G. Liu, X. Ji, J. Zhai, and Y. Dai, "A wireless covert channel based on constellation shaping modulation," *Security and Communication Networks*, 2018.
[8] R. Bell, K. Watson, T. Hu, I. Poy, fred harris, and D. Bharadia, "Searchlight: An accurate, sensitive, and fast signals detection system," *MILCOM 2023 - 2023 IEEE Military Communications Conference*, 2023.
[9] M. Zhang, Y. Zeng, Z. Han, and Y. Gong, "Automatic modulation recognition using deep learning architectures," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018, pp. 1–5.
[10] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 4074–4077, 2019.
[11] J. Cai, F. Gan, X. Cao, and W. Liu, "Signal modulation classification based on the transformer network," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 3, pp. 1348–1357, 2022.
[12] Y. Chen, X. Xu, and X. Qin, "An open-set modulation recognition scheme with deep representation learning," *IEEE Communications Letters*, vol. 27, no. 3, pp. 851–855, 2023.
[13] C. Spooner, *Chapter 18. Spectrum Sensing Based on Spectral Correlation*, 12 2009, p. 593.
[14] W. A. Gardner, *Statistical Spectral Analysis*. Prentice-Hall, 1987.
[15] W. A. Brown and H. H. Loomis, "Digital implementations of spectral correlation analyzers," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 703–720, February 1993.
[16] E. April, "On the implementation of the strip spectral correlation algorithm for cyclic spectrum estimation," Defense Research Establishment, Ottawa, Ontario, Canada, Tech. Rep. ADA289815, February 1994.