MODELESS: MODulation rEcognition with LimitEd SuperviSion

Wei Xiong, Petko Bogdanov, Mariya Zheleva
Department of Computer Science, University at Albany—SUNY
Emails:{wxiong,pbogdanov,mzheleva}@albany.edu

Abstract*—Modulation recognition (modrec) is an essential transmitter fingerprinting task that enables future spectrum-sharing applications such as access management and enforcement. Traditional supervised modrec requires labeled training data for all target modulations, which cannot be readily met with the advent of new, customized and data-driven waveforms. Thus, a keystone question for the applicability of modrec is: *Can we perform automatic recognition of previously unobserved modulations by adapting and reusing models that were trained on different but related modulations?

To this end, we develop `MODELESS` (MODulation rEcognition with LimitEd SuperviSion) that exploits knowledge from observed modulations to classify samples from unobserved ones. Our solution is grounded in zero-shot transfer learning, which employs side information among observed and unobserved classes to transfer learned classifiers. In particular we quantify the similarity among the theoretical constellation diagrams of unobserved and observed modulations and employ them in a zero-shot transfer learning framework. Our framework is general, as it can produce predictions for arbitrary modulations as long as their theoretical constellations can be specified. We evaluate `MODELESS` on synthetic and real-world traces and in comparison with zero-shot counterparts from the literature. We demonstrate near-ideal classification accuracy in the majority of the testing cases and draw recommendations for future research into classification tasks with sub-par performance.

I. Introduction

While Dynamic Spectrum Access (DSA) is projected to become a key capability in future wireless networks, its practical success hinges on the ability to automatically measure, characterize and enforce spectrum utilization while operating with limited prior knowledge of transmitter characteristics. Modulation recognition (modrec) is one such critical spectrum analysis capability. A growing body of feature-based [6] and deep-learning based [24] modrec approaches has gained traction due to their computational efficiency. However, these methods pose stringent data collection requirements, as they assume the availability of labeled training data for all target modulations [10], [30], [7], [14]. In other words, a modrec engine (e.g. a third-party spectrum enforcer) needs to have controlled access to target transmitters, in order to collect labeled spectrum traces for supervised training.

The training data requirement presents a fundamental challenge for practical modrec deployments in emerging wireless networks. First, in following the evolution of radio front-end capabilities, we see a rapid increase in the modulation families (e.g. high-order SISO modulations). Second, companies often perform proprietary modifications of their respective modulation implementations. Finally, modulation implementations evolve from fixed to dynamic and data-driven [20], [37], embedding fine-grained modulation adaptation in response to channel conditions; with commercial products already becoming available [1]. In the face of ever-growing modulation variety, a fully-supervised modrec framework quickly becomes infeasible [35], [21], [15], [4] as it will require equally diversified labeled training data. These factors increasingly challenge traditional modrec algorithms and underpin the need for robust modulation classification with limited training. Thus, *the goal of our paper is to develop a framework for automatic recognition of previously unseen modulations.*

In this paper, we develop `MODELESS`, which *strives to determine the modulation family and order of previously unobserved modulations*. `MODELESS` assumes two types of input: *domain knowledge* and *measurement data*. In the modrec context, the domain knowledge comprises the theoretical understanding of a modulation's constellation diagram, whereas the measurement data consists of `IQ` samples from spectrum measurements. Further, we differentiate between *seen* and *unseen* modulation classes. For the *seen classes*, during training, we have both domain knowledge and measurement data, whereas for the *unseen classes*, we only have the domain knowledge. Using the domain knowledge, we create similarity representations across seen and unseen modulation classes, which are then used in the modulation recognition process. Intuitively, modulations with similar constellations will have similar representations. `MODELESS` exploits this property to make robust predictions for previously unseen classes. Our framework is inspired by max-margin learning and employs embeddings for both the domain knowledge (i.e. the similarity of theoretical constellations computed via the Earth Mover's Distance), and the features derived from observations. We train an SVM class decision function via a joint supervised feature embedding and discriminator learning. During testing, we employ the trained score function to predict the class for instances arising from both known and unknown classes.

We evaluate `MODELESS` on synthetic and real data [21] with two popular modulation families: PSK and QAM, each represented with four classes (2/4/8/16PSK and 16/32/64/128QAM). We consider all combination where two out of the eight classes (25%) were not observed in training. When the two unseen classes are from different families, `MODELESS` has near-ideal performance regardless of modula-

978-1-6654-4108-7/21/$31.00

tion complexity. With target classes from the same family, our method requires further consideration of the training mixture.

Our paper makes the following contributions:

• **Novelty:** We conceptualize modulation recognition of previously unseen signals and draw on domain knowledge to design a data-driven similarity embedding framework that exploits IQ constellations' geometry to automatically determine transferable traits between seen and unseen modulations. We employ state-of-the-art classification features and lightweight SVM classifiers for robust detection. Our framework is highly-applicable as it does not require expert input or prior knowledge of the modulation implementation.

• **Generality:** MODELESS seamlessly extends to new modulations beyond the set considered in this paper. Furthermore, it can employ alternative similarity embeddings, signal features and classifiers, allowing further expansion as the field evolves.

• **Applicability:** We study MODELESS' performance in synthetic and over-the-air (OTA) spectrum traces and outline its strengths and limitations across key criteria including the amount of training data, knowledge transferability and parameter settings. We propose guidelines for in-situ deployments.

II. Related Work

Modulation recognition has been tackled as either unsupervised (i.e. likelihood-based) [23] or supervised (i.e. feature-based) [6] classification problem. While likelihood-based approaches are optimal, they are computationally-expensive and sensitive to sensor imperfections and channel conditions [30], which limits their applicability. Feature-based approaches extract features from measured IQ samples and offer a lower-complexity alternative, which has been extensively utilized in recent work [10], [30], [7], [14], [2], [9]. In terms of features, prior work employs order statistics (OS) [10], high order cumulants (HOC) [30], [7], [9], kernel density functions [2], local sequential patterns [34] and fractal dimensions [36]. Various classification techniques have been employed ranging from support vector machines [9], [34], [36] to artificial neural networks [18], [32], [21], [26], [27]. All of these approaches require sufficient training data prior to classification, which is a practical challenge [40], [5], [37], [20], [1]. Thus, our key contribution is the design and implementation of a modrec framework for unseen modulations. In terms of feature engineering and classifier selection, our work is orthogonal to prior literature, as we focus on the design of similarity embeddings for the modrec domain. Our framework is seamlessly extensible with new features and classifiers.

A few recent studies consider the application of transfer learning within a deep neural network for modrec with limited supervision [21], [15], [4]. Albeit limited, all of these works require a certain percentage of actual observations (i.e. 5%-10%) from the target classes in order to learn a corresponding model. Our work is different as it does not require any prior observations. Closest to MODELESS is [35], which uses zero-shot transfer learning and requires expert-designed transfer attributes. Our work strives to eliminate the human in the loop by designing data-driven similarity embeddings for fully-automated recognition of previously unseen modulations.

Notation	Meaning
$\mathcal{S}$	The set of seen modulations.
$\mathcal{U}$	The set of unseen modulations.
$\boldsymbol{C}$	Modulation similarity matrix.
$\{\mathbf{X}, \mathbf{y}\}$	Training instances $\mathbf{X}$ with mod. annotations $\mathbf{y}$.
l, K	Shingle and dictionary sizes for LP features [34].
$P = \{p_1, .., p_o\}$	A theoretical constellation with o complex symbols.
$\boldsymbol{w}$	Learned classifier hyperplane.
$\boldsymbol{Z}$	Class similarity embedding.
ϕ	Class-specific feature embedding.

TABLE I: Notation used throughout the paper.

Zero-shot learning. Our goal in this paper is to enable modrec with limited supervision and no expert input. Thus, we employ a similarity-based zero-shot learning approach. Zero-shot learning has been extensively considered in the image segmentation literature, where too, the lack of sufficient labeled training data has plagued progress in robust image classification [8], [22]. Recent progress can be considered in terms of transfer features representation and classifier design. For transfer representations, prior work has used attribute-based [28], [12], [11] and similarity-based approaches [39], [16], [19]. Attribute-based work requires semantic description of key class attributes that can be employed for knowledge transfer. For example, the semantic attributes for a seen class "cats" may include "four legs", "fur" and "paws", whereas these for an unseen class "dogs" may include "fur", "sharp teeth" and "paws". Several of the semantic attributes of the seen class directly transfer to the unseen class, which underpins zero-shot classification. While intuitive, attribute-based zero-shot learning inherently requires expert input in the attribute design phase, which is prohibitive in the modulation recognition context. Similarity-based methods typically employ data-driven computation of inter-class similarities, and as such, do not require expert input. Thus, similarity-based approaches such as [16], [19] are very-well suited for our goal. A key question tackled in this paper is *how to design similarity embeddings for the modrec domain?* Beyond new representations, various zero-shot learning classifiers have been considered: from traditional SVM [33] to novel deep learning architectures [38]. MODELESS employs lightweight SVM, however, it is extensible to other classifiers.

III. Preliminaries and Notation

We first discuss modrec preliminaries and our running notation (also summarized in Table I for further reference).

Feature-based supervised modulation recognition. The raw input data for feature-based supervised methods is a sequence $(r_1, r_2, ..r_m)$ of m consecutive IQ samples, where $r_i \in \mathbb{C}$ is the i-th instantaneous complex signal sample. Following a feature extraction procedure [30], [7], [9], [10], [2], [34], [36], samples are mapped to a fixed-dimensional feature vector (instance) $\mathbf{x} \in \mathcal{R}^d$ of size d. Instances are further associated with modulation labels to form a training dataset of n instances $\{\mathbf{X}, \mathbf{y}\}$, where $\mathbf{X} \in \mathbb{R}^{n \times d}$ is the set of measured instances stacked in a matrix and and $\mathbf{y} \in \mathcal{M} = \{c_0, c_1 \dots c_{k+k'}\}^{n \times 1}$ is a class vector encoding the modulation types giving rise to

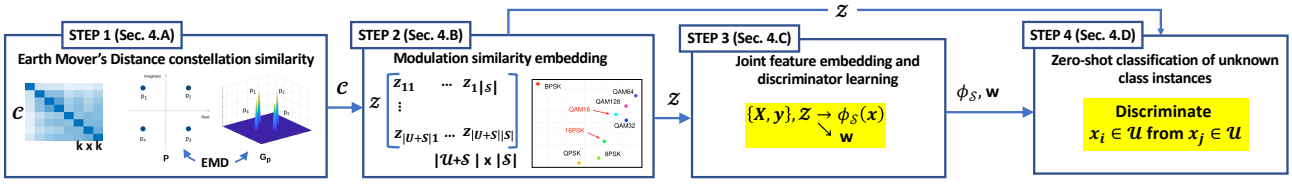


Fig. 1: An overview of the MODELESS pipeline.

each instance. While prior feature-based approaches require observations for all target classes (i.e. $\mathcal{M} = \mathcal{S}$), MODELESS assumes that $\mathcal{M}$ is comprised of a subset of seen $\mathcal{S}$ and unseen $\mathcal{U}$ modulations, and thus, it targets the recognition of the previously unseen classes $\mathcal{U}$.

Feature families. *High order cumulants (HOC)* [3], [30], [9], [7], [3] and *order statistics (OS)* [10] are two families of features that have been extensively used in the literature. In terms of HOC, subsets of the fourth- $\{C_{40}, C_{41}, C_{42}\}$ and sixth-order cumulants $\{C_{60}, C_{61}, C_{62}, C_{63}\}$ have received the most attention [30], [9], [7], [3]. OS [10] offer an alternative global summary of the IQ samples' distribution. The k-th OS of a random real sample is its k-th smallest value.

Both HOC and OS features capture the global statistical properties of IQ samples, however they ignore the local sequence of samples. A recently proposed feature representation, called *Local Pattern (LP)* [34], exploits the local order of IQ samples. The key idea is to learn representative transitions of amplitude and phase sequences. A sequence of IQ samples $(r_1, r_2, ..r_m)$ is transformed into amplitude and phase time series of the same length. A sliding window of length l is employed to obtain $m - l$ sub-sequences, called shingles of length l. A set of K representative shingle shapes (shingle dictionary) across all modulations is then learned in an unsupervised manner using a Gaussian Mixture Model (GMM) framework [34]. All observations are then encoded based on the learned dictionary to obtain the feature representation $\mathbf{X}$.

We study the applicability of the above features to MODELESS (§V). Our overall framework is flexible as any combination of features, including such proposed in the future, can be employed without any changes to the framework.

Zero-shot modrec v.s. supervised modrec. The goal of classical supervised modrec is to learn a classifier which maps measured modulation instances to one of the *known* classes $f(\mathbf{x}) \rightarrow \mathcal{S} = \{c_0, c_1 \dots c_k\}$, for which a classifier was trained. Zero-shot learning, in contrast, seeks to "adapt" a learnt classifier to predict a set of unobserved modulations $\mathcal{U}$, which were not available during training, i.e., $f_0(\mathbf{x}) \rightarrow \mathcal{U} = \{c_{k+1} \dots c_{k+k'}\}$. There are two streams of zero-shot learning approaches, which are applicable to our problem: *attribute-based* and *similarity-based*. Attribute-based methods require the explicit definition of a trait that transfers between seen and unseen classes (e.g. "is furry" or "is a mammal"), and as such, require that the target domain grants itself to a semantic description generated by humans. Similarity-based approaches, in turn, are data-driven and determine the transferable traits across domains in the form of automatically-computed similarity matrices. In the context of modrec, attribute-based approaches will require extensive expert intervention both to design the attributes and to extend them as new modulations arise. Similarity-based methods, in contrast, are more applicable as they automatically determine transferable traits based on domain-informed functions. Thus, MODELESS employs a similarity-based transfer learning framework. Our classifier closely follows the framework proposed in [39]. Our novel contribution is in the design of similarity embeddings, which exploit the theoretical shape of modulation constellations.

IV. Methodology

In this section, we describe our zero-shot similarity-based framework MODELESS for modulation recognition. Its steps are depicted in Fig. 1. We first estimate a pair-wise similarity matrix $\boldsymbol{C}$ among the theoretical constellations of both known and unknown modulations employing a sparse Earth Mover's Distance formulation based on Gaussian Mixture Models (GMMs). We further embed modulation similarity vectors via a sparse coding as mixtures of known classes resulting in an embedding matrix $\boldsymbol{Z}$. Next, we train a class decision function for instances via a joint supervised feature embedding and discriminator learning. Finally during testing, we employ the trained score function to predict the class for unlabeled instances arising from both known and unknown modulations.

A. Similarity across modulations

We approach the problem of modrec with limited supervision within a *similarity-based* zero-shot framework [39]. We take as an input the *theoretical constellations* of seen and unseen modulations and employ their pair-wise similarity in the transfer learning process. A key question for our framework is: *How to compute the pair-wise similarities among theoretical constellations with different number of symbols?* We model modulation constellations as two-dimensional distributions in the IQ space and quantify pairwise modulation similarities using the *Earth Mover's Distance (EMD)* [29]. Originally defined as an edit-distance for histograms, EMD computes the cost of an optimal transformation of one histogram (or distribution) into another, where an elementary edit operation is the transportation of a unit of mass from one histogram bin to another bin. The costs of such elementary movements is quantified by a *ground distance*, i.e. a distance among histogram bins. EMD is a true metric if the underlying ground distance is also a metric [29] and these nice theoretical properties underpin its wide adoption in image classification.

A given modulation's theoretical constellation $P = \{p_1, \cdots p_o\}, p_i \in \mathbb{C}$ is specified by a set of complex numbers corresponding to the noise-free positions of constellation symbols in IQ space, where o is the number of symbols

(also known as modulation order). We model a theoretical constellations P as *Gaussian mixture model (GMM)* G_P with the same number of components as the number of symbols in P, where all components have the same likelihood and a fixed variance σ. This model is illustrated for the QPSK constellation in the left pane of Fig. 1. Note that this model assumes that (i) all symbols are equi-probable (ii) the noise around symbols has zero covariance and (iii) the phase offset is fixed. While we demonstrate that under the above assumption we can successfully predict unknown classes, we believe that the performance for challenging cases may be improved by re-visiting the above assumptions in future research.

While the classical EMD [29] can be employed for arbitrary (even non-parametric) distributions, we employ a recent EMD formulation tailored to GMMs to quantify distances among modulations constellations [13]. Given two GMM models G_P and G_Q the goal is to quantify the cost of transforming one into the other formalized as a sparse coding problem:

$$d_{EMD}(G_P, G_Q) = min_{\mathbf{f}} \frac{1}{2}||\mathbf{y} - \mathbf{ADf}||_2^2 + \tau||\mathbf{f}||_1, \;\; s.t. \;\; \mathbf{f} \geq 0, \tag{1}$$

where $\mathbf{y}$ is a concatenated vector of the prior weights (in our case uniform) of GMM components for both modulations G_P and G_Q, $\mathbf{A}$ is an indicator $0/1$ matrix encoding the relationship between components and "transportation" paths among them, $\mathbf{D}$ is a diagonal ground distance matrix for each path between GMM components, $\mathbf{f}$ is the optimal transportation schedule and τ is a regularization parameter controlling the sparsity of the transportation schedule $\mathbf{f}$. This sparse formulation is more robust to noise and is more efficient to compute than a general (non-GMM) formulation [13]. It naturally allows for a wide range of ground distances to be incorporated. We employ the *Lie group* ground distance, as detailed in [13], in IQ space for our modrec application.

Note that while $d_{EMD}(G_P, G_Q)$ allows us to quantify the distance between any two modulations, we need to quantify a similarity between two modulations to enable similarity-based zero-shot learning. To this end, we transform the distance into a similarity by using an RBF kernel function [31] as follows:

$$c(P, Q) = e^{-\lambda d_{EMD}(G_P, G_Q)}, \tag{2}$$

where λ is a parameter controlling the rate of decrease of the similarity as a function of the EMD distance. We evaluated MODELESS's sensitivity to λ by varying it from 0.1 to 1 in increments of 0.1 and determined that it does not affect the classification performance (figure omitted in interest of space). Thus, we set $\lambda = 0.2$ and quantify the similarities among all pairs of modulations (both known and unknown). Fig. 2(left) shows the EMD-based similarity matrix $\boldsymbol{C}$ for 8 modulations (four PSK and four QAM) in medium SNR regime (10dB). Darker colors present higher similarity. We note that in-family similarity (e.g. between the PSK or QAM classes) is higher than between classes from different families. This property holds across all SNR regimes and underpins MODELESS's high performance in classification of unseen modulations.

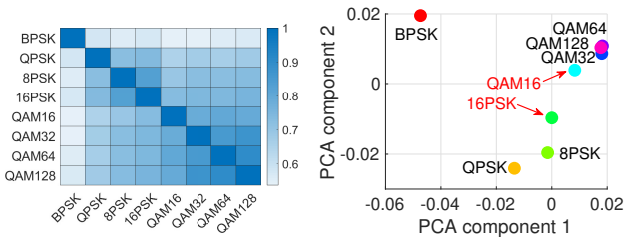


Fig. 2: (left) Class-level similarity matrix $\boldsymbol{C}$ using EMD. (right) Class-level embedding $\boldsymbol{Z}$ for target modulations 16PSK and QAM16 at SNR 10dB. PCA dim. reduction is used to visualize the 6D embeddings $\boldsymbol{Z}$ in 2D.

B. Sparse modulation similarity embedding

The similarity matrix $\boldsymbol{C}$ embeds all modulations in a "dense" similarity space purely based on the theoretical constellation and without additional semantic knowledge about the modulations. The key premise in the similarity-based zero-shot learning framework we adopt [39] is that the source domain, or all modulations in $\mathcal{S}$ and $\mathcal{U}$, can be represented as a sparse mixture of the similarity vectors of known classes. Particularly $\mathbf{c}_i \approx \sum_{j \in \mathcal{S}} \mathbf{z}_{ij} \mathbf{c}_j$, where $\mathbf{c}_j$ is the j-th row of $\boldsymbol{C}$ and $\mathbf{z}_{ij}$ is the mixture proportion of the j-th (known) modulation in the i-th modulation embedding. More importantly we can learn to "match" the source domain embedding $\mathbf{z}_i$ to a target domain (features) embedding $\phi(\mathbf{x}_u)$ obtained by embedding instances from unknown classes $\mathbf{x}_u$ during testing, i.e. when the zero-shot modrec is deployed. We next present the approach for modulation (source) embedding $\boldsymbol{Z}$, while the feature (target) embedding $\phi(\mathbf{x})$ is discussed in the following subsection.

The input for the modulation embedding function is the similarity matrix $\boldsymbol{C}$. The embedding process for modulation i is formulated as a sparse encoding optimization as follows:

$$\mathbf{z}^{(i)} = \underset{\mathbf{z}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{c}_i - \sum_{j \in \mathcal{S}} \mathbf{z}_j \mathbf{c}_j||^2 + \frac{\gamma}{2}||\mathbf{z}||^2, \;\; \text{s.t.} \sum_j \mathbf{z}_j = 1 \tag{3}$$

where the first term in the optimization promotes accurate mixture representation and the second term shrinks the small proportions in z according to a regularizer parameter γ. The embedding is constrained to be on the simplex, i.e. it can be viewed as a distribution over the known classes in $\mathcal{S}$. We learn the embedding for all modulations $\mathbf{z}^{(i)}, \forall i \in \mathcal{S} \cup \mathcal{U}$ by employing a quadratic programming solver resulting in the modulation similarity embedding matrix $\boldsymbol{Z} \in \mathcal{R}^{|\mathcal{U} \cup \mathcal{S}| \times |\mathcal{S}|}$.

Fig. 2(right) shows and example similarity embedding $\boldsymbol{Z}$ at a mid-range SNR of 10 dB for unknown modulations 16QAM and 16PSK. The remaining six modulations are treated as known. To produce this 2D visualization, we apply *principal component analysis (PCA)* on the 6-dimensional embeddings in $\boldsymbol{Z}$. Of note is that we use PCA only for this illustrative example, whereas the MODELESS framework uses all embedding dimensions. What stands out is that the unknown modulations are "clustered" with other known members from the same family with the exception of BPSK which acts as an outlier.

Our framework is general, as it allows the addition of new modulations as long as their theoretical constellation can be specified. Our evaluation (§V) demonstrates that the discriminative power of the similarity embedding persists across all

SNR regimes. The classification performance with new classes will depend on whether the trained classifier captures traits from that class's family (e.g. if a newly added modulation has nothing in common with previously observed classes, it is likely that the classification will be poor). Our discussion (§VI) considers unique benefits and further exploration.

C. Joint supervised feature embedding and classifier learning

The modulation similarity embedding $\boldsymbol{Z}$ employs solely the theoretical modulation properties, and no IQ data samples. Next, we employ this embedding along with training data $\{\mathbf{X}, \mathbf{y}\}$ from known classes $\mathbf{y}_i \in \mathcal{S}$ to learn a max-margin class scoring function $f(\mathbf{x}, y)$, which can be used to predict both known and unknown modulations, i.e. y can be an index from $\mathcal{S}$ and $\mathcal{U}$. The scoring function has the following form:

$$f(\mathbf{x}, y) = \sum_{s \in \mathcal{S}} z_{ys} \mathbf{w}^T \phi_s(\mathbf{x}), \tag{4}$$

where $\mathbf{z}_{ys}$ is the proportion of known class s in the modulation similarity embedding of candidate class y, $\mathbf{w}$ is a classifier hyperplane (akin to those employed in SVM) and $\phi_s(\mathbf{x})$ is a class-specific feature embedding of the instance $\mathbf{x}$ with respect to class s. Our objective is to learn the classifier hyperplane and class-specific embedding functions $\phi()$ such that training instances maximize the scoring function for their true class.

The zero-shot similarity approach from [39] proposes different feature embedding families ϕ. In our evaluation we employ the *rectified linear unit (ReLU)* [17], which is defined as $\phi_s(\mathbf{x}) = max(0, \mathbf{x} - \boldsymbol{v}_s)$ and "focuses" on the patterns in dimensions of $\mathbf{x}$ which exceed the class-specific thresholds which we learn in $\boldsymbol{v}_s$, while ignoring (zeroing out) other dimensions. Alternative embedding families can also be considered, however, we leave this for future exploration.

To learn the joint feature embedding functions $\phi_s()$ and classifier $\mathbf{w}$ we adopt the max-margin scheme from [39], formalized as the following constrained optimization problem:

$$\min_{\mathbf{V}, \mathbf{w}, \boldsymbol{\epsilon}, \boldsymbol{\eta}} \frac{1}{2} \|\mathbf{w}\|_2^2 + \frac{\lambda_1}{2} \|\mathbf{V}\|_F^2 + \lambda_2 \sum_{y,s \in \mathcal{S}} \boldsymbol{\epsilon}_{ys} + \lambda_3 \sum_{\substack{i=1\ldots n \\ y \in \mathcal{S}}} \boldsymbol{\eta}_{iy}, \tag{5}$$

$$\text{s.t.} \sum_{i=1}^{n} \frac{I_{y_i = y}}{n_y} (f(\mathbf{x}_i, y) - f(\mathbf{x}_i, s)) \geq \Delta_{y,s} - \boldsymbol{\epsilon}_{ys}, \forall y, s \in \mathcal{S} \tag{6}$$

$$f(\mathbf{x}_i, y_i) - f(\mathbf{x}, y) \geq \Delta_{y_i, y} - \boldsymbol{\eta}_{iy}, \forall i = 1 \ldots n, y \in \mathcal{S} \tag{7}$$

$$\epsilon_{ys} \geq 0, \boldsymbol{\eta}_{iy} \geq 0, \mathbf{v} \geq 0, \forall \mathbf{v} \in \mathbf{V}, \tag{8}$$

where λ_i are non-negative regularization parameters, $\boldsymbol{\epsilon}$ and $\boldsymbol{\eta}$ are slack variables, n_y is the number of instances of class y, I_{cond} is an indicator function with value 1 when the condition $cond$ is true and 0 otherwise, and Δ_{ys} is a structural loss between similarity vectors of the corresponding known classes defined as $\Delta_{ys} = 1 - \boldsymbol{c}_y^T \boldsymbol{c}_s$.

The optimization (Eq. 5) is a max-margin objective imposing shrinkage on the class-specific threshold vectors stacked in matrix $\boldsymbol{V}$ via a Frobenius norm and trading off margin width for training error expressed by the two sets of slack variables. The set of constraints in Eq. 6 limits the alignment loss for known class distributions, while the constraints in Eq. 7 limit the classification loss much like in standard SVMs. The overall objectives enable a balance between margin width, classification loss for instances and alignment loss for the distributions of known classes.

The constrained max-margin formulation can be optimized within an alternating optimization scheme in which the classifier $\{\mathbf{w}, \boldsymbol{\epsilon}, \boldsymbol{\eta}\}$ and the class-specific threshold vectors in $\boldsymbol{V}$ parametrizing the feature embedding functions $\phi()$ are optimized in turn [39]. The first group can be solved by a standard SVM while the second is optimized via a *concave-convex procedure* by exploiting the structure of the constraints [39]. This description is a synthesized version of the overall optimization. For detailed discussion refer to [39].

D. Classification of unknown modulation instances

Given an arbitrary instance vector $\mathbf{x}$, we predict its class based on the scoring function $f()$ from Eq. 4, which employs the modulation $\boldsymbol{Z}$ and feature $\phi()$ embeddings and the classifier $\mathbf{w}$ learned in the previous sections. Namely, the class decision function for an unlabeled instance is:

$$\hat{y} = \underset{y \in \mathcal{S} \cup \mathcal{U}}{\operatorname{argmax}} f(\mathbf{x}, y). \tag{9}$$

Note that MODELESS can provide predictions for instances arising from both known and unknown classes, i.e. without observing instances from the latter during training.

Parameter optimization. Our method employs several parameters: shingle and dictionary sizes (l, K) for local pattern features, the EMD sparsity parameter τ, the RBF kernel λ, the shrinkage regularizer γ in the modulation similarity embedding, and the regularizer parameters λ_i in the max-margin learning. An important question is how to set those parameters. We can employ one of two cross-validation techniques: instance-based and class-based. Instance-based cross-validation is the standard stratified partitioning of the training, which maximizes predictions for the known classes $\mathcal{S}$, however, it cannot help us directly optimize the zero-shot predictions for unknown classes $\mathcal{U}$. To optimize the latter, we can employ a class-based cross-validation in which we leave out known classes from $\mathcal{S}$ in turn, train MODELESS on the remainder, and quantify modrec accuracy on the left-out class.

V. Evaluation

A. Experimental setup

Data. We evaluate MODELESS on synthetic and over-the-air (OTA) real spectrum traces. We employ eight popular digital modulations from two families: 2/4/8/16PSK and 16/32/64/128QAM. Our synthetic dataset is generated with the MATLAB Communication Toolbox across ten SNR regimes (from 0 to 20dB in increments of 2) and across three realistic channel models: AWGN, Rayleigh and Rician. We also utilize an OTA dataset from [21], collected with USRB B210 radios across SNR settings (0-20dB in increments of 2).

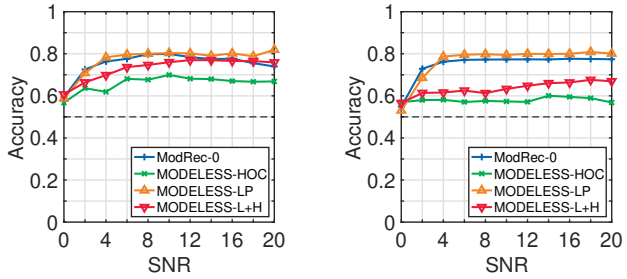
MODELESS implementation. Our framework is implemented in MATLAB. Experiments were executed on a Linux server with 256GB of RAM and 96 Intel Xenon 2.0GHz processors. We implement the similarity embeddings calculation following [13]. The local sequential pattern feature extraction is based on [34]. For the transfer learning framework (both training and prediction) we adopt the implementation from [39].

Evaluation strategy. In all experimental settings we evaluate the *accuracy* of MODELESS, defined as the fraction of correctly-predicted instances over all instances. *Evaluation criteria:* First, we assess the sensitivity of our framework to the mixture of unseen modulations in both synthetic and OTA traces. We consider two unseen mixture types: *in-family* and *cross-family*. For the in-family test case, the unseen pool is comprised of multiple modulations from the same family (e.g. several PSKs or several QAMs), whereas for the cross-family test case the unseen pool is comprised of a mix of modulations from different families. We then evaluate performance as a function of the training pool (i.e. seen modulations) size and mixture. We also study the effects of instance size on performance. Finally, we evaluate performance across varying SNR. *Training:* Unless otherwise noted, we use all eight modulation classes in each experiment. We designate the experiments in terms of the test (i.e. "unseen") classes and note that all remaining classes were used in training (i.e. were considered as "seen" classes). *Parameters:* Training and testing instances of each modulation contain 512 complex IQ samples. We train on 2000 instances of each modulation and test on another set of 2000 instances. For the LP feature extraction, we set the shingle and dictionary size to 3 IQ samples and 30 representative vocabularies. §V-D evaluates the effects of parameter selection on MODELESS' performance and informs the above parameters.

Baselines. We compare with ModRec-0 [35], which is the only prior work that tackles modrec of unseen classes. It employs an attribute-based zero-shot framework with high order cumulants (HOC), which requires expert-defined attributes for transfer. In contrast, MODELESS does not require expert knowledge as it employs similarity between known theoretical constellations. We consider variants of MODELESS employing different features: HOCs [7], local patterns (LP) [34] and the combination of HOCs and LPs. Comparison to fully-supervised [30], [7], [14], [10], [34] or partially-supervised [21], [15], [4] methods is not possible as they would require actual samples for unobserved classes, unlike MODELESS and ModRec-0.

B. MODELESS performance

We evaluate MODELESS' performance with synthetic and OTA traces. First, we explore MODELESS's performance with different state-of-the-art features: High Order Cumulants (MODELESS-HOC), Local Patterns (MODELESS-LP) and their combination (MODELESS-H+L). Second, we seek to compare MODELESS to ModRec-0[35]. Finally we test these four counterparts on real-world traces. We report results using two classes for testing and the remaining six for training.



(a) Synthetic (b) Over-the-air

Fig. 3: Average accuracy as a function of increasing SNR.

Considering our 8 modulation classes (§V-A), for two vs. all testing/training there exist a total of $C_8^2 = 28$ test combinations. 16 of them are cross-family (i.e. we seek to classify one PSK and QAM modulation) and 12 are in-family combinations (i.e. both target modulations are either from the PSK or QAM family). We employ MODELESS on each of these combinations while increasing the SNR level.

Fig. 3 presents average accuracy across all 28 combinations for synthetic data under AWGN channel (left) and OTA (right) traces. The dashed line indicates a random guess, which in our case is 0.5, as we are classifying two modulations. MODELESS-LP achieves the highest performance, followed by ModRec-0 in both synthetic and real data. MODELESS-HOC and MODELESS-L+H on synthetic data suffer an average accuracy drop of 12% and 4%, respectively, compared to MODELESS-LP. This performance drop is even more substantial with real data, whereby the average accuracy deterioration is 18% for MODELESS-HOC and 13% for MODELESS-L+H. These results point to several important insights. First, MODELESS-LP outperforms the state-of-the-art and obviates the requirement of expert input. Second, features that capture local signal properties (i.e. LPs [34]) are more conducive to zero-shot learning than global statistics (i.e. HOCs [7]).

To further understand MODELESS-LP's performance, we break down the accuracy of each test combination across three SNR regimes (Table II): 2, 10 and 20dB for synthetic and 4, 10 and 20dB for OTA data. All cross-class combinations (i.e. 1-16) achieve near-perfect classification at SNR 10 and 20dB and perform substantially better than a random guess at 2(4)dB. These trends persist across synthetic and OTA traces. However, with a few exceptions, all the in-class combinations (17-28) gain a random guess. We hypothesize that this poor performance is due to our exhaustive training strategy, which uses all seen classes in the training phase. Effectively, this leads to training data being dominated by instances from the opposite class, which has a detrimental effect on classification performance. We further explore this phenomenon in §V-C and show that indeed, balanced selection of the training mixture and feature design can boost the in-class modrec performance.

C. Less is more: effects of the training mix

In this section, we set out to evaluate the effects of the training mixture on in-family and cross-family modulation classification. We consider the following training mixtures: (i) *in-family*, whereby the training phase incorporates instances

Train		Test		Acc. Synth. (SNR, dB)			Acc. OTA (SNR, dB)		
PSK	QAM	PSK	QAM	2	10	20	4	10	20
4/8/16	32/64/128	2	16	.9	1	1	.54	.99	1
4/8/16	16/64/12	2	32	.91	1	1	.57	.99	1
4/8/16	16/32/128	2	64	.94	1	1	.55	1	1
4/8/16	16/32/64	2	128	.92	1	1	.56	.99	.99
2/8/16	32/64/128	4	16	.68	1	1	.76	.99	.99
2/8/16	16/64/128	4	32	.67	1	1	.7	.99	.99
2/8/16	16/32/128	4	64	.72	1	1	.84	.99	.99
2/8/16	16/32/64	4	128	.75	1	1	.76	.99	1
2/4/16	32/64/128	8	16	.9	1	.99	.92	.98	.96
2/4/16	16/64/128	8	32	.9	1	1	.93	.92	.85
2/4/16	16/32/128	8	64	.91	1	1	.96	.98	.91
2/4/16	16/32/64	8	128	.91	1	1	.96	.95	.95
2/4/8	32/64/128	16	16	.9	1	1	.93	.98	.98
2/4/8	16/64/128	16	32	.91	1	1	.93	.96	.96
2/4/8	16/32/128	16	64	.94	1	1	.96	.96	.97
2/4/8	16/32/64	16	128	.88	1	1	.96	.97	.99
8/16	16/32/64/128	2/4	-	.5	.5	.5	.9	.5	.5
4/16	16/32/64/128	2/8	-	.5	.5	.5	.88	.5	.5
4/8	16/32/64/128	2/16	-	.5	.5	.5	.9	.77	.96
2/16	16/32/64/128	4/8	-	.5	.99	.95	.62	.73	.60
2/8	16/32/64/128	4/16	-	.5	.5	.97	.67	.8	.75
2/4	16/32/64/128	8/16	-	.5	.5	.5	.5	.52	.51
2/4/8/16	64/128	-	16/32	.53	.5	.5	.5	.47	.54
2/4/8/16	32/128	-	16/64	.5	.5	.5	.51	.5	.5
2/4/8/16	32/64	-	16/128	.5	.5	.5	.5	.5	.5
2/4/8/16	16/128	-	32/64	.51	.5	.5	.51	.5	.49
2/4/8/16	16/64	-	32/128	.52	.5	.5	.51	.5	.5
2/4/8/16	16/32	-	64/128	.5	.5	.55	.5	.5	.5

TABLE II: Accuracy break down across testing combinations.

only from the same family as the testing classes; (ii) *balanced cross-family with exclusion*, in which we take an equal amount of classes from the PSK and QAM family by excluding some classes from the over-represented family (e.g. if the seen classes are 4/8PSK and 16/32/64/128QAM we would exclude 16/32QAM to balance the two families); (iii) *balanced cross-family without exclusion* in which we simply add more samples from the underrepresented family in order to balance the two classes (e.g. in the above example of seen classes 4/8PSK and 16/32/64/128QAM we would double the samples in QPSK and 8PSK to balance the two families); and (iv) *exhaustive* training, whereby all non-test classes are used in training. We note that for our pool of 8 classes, we can produce at most one in-family, three balanced cross-family with exclusion, one balanced cross-family without exclusion and one exhaustive training mixtures per target combination.

1) Effects of training mixture on cross-family modrec: Our results in §V-B indicate near-perfect classification accuracy for all cross-family test combinations and across a wide range of SNRs. We now evaluate the effects of the training mixture on the classification performance. We choose the 16PSK+16QAM test combination and gradually increase the training dataset from two to six classes adding one class per family at a time (i.e. 8PSK+32QAM; 4/8PSK+32/64QAM and 2/4/8PSK+32/64/128QAM). The accuracy remains near 100% in all training cases (figure omitted in interest of space), which emphasizes the robustness of cross-family classification to the training mixture and underpins `MODELESS`' applicability to real-world cross-family modrec with limited supervision.

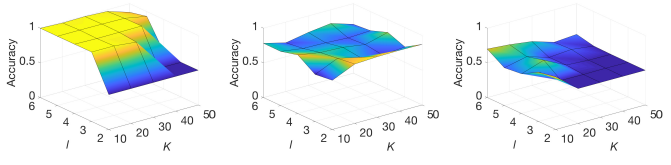
2) Effects of training mixture on in-family modrec: We also evaluate the training mixture effects on in-family classification. Our hypothesis is that employing exhaustive training for in-family classification leads to the opposite class dominating the training outcomes, which hampers the knowledge transfer with in-family test cases. We thus, evaluate the performance across all in-family combinations while controlling the mixture of training classes. Table III reports the accuracy for three combinations (16/32QAM, 2/8PSK and 4/16PSK) at SNR 20dB. For the first and second combination, in-family training gains a significant performance boost, lifting the accuracy from 0.5 (a random guess) with exhaustive training to 0.99. In the third case, we see the opposite trend: in-family training gains a random guess, while both cross-family and exhaustive training result in near-perfect classification. These results show that careful selection of the training mixture can be beneficial in some test combinations, whereas others require further investigation into training mixture and feature extraction.

Test	Train	Acc.
16/32 QAM	64/128 QAM (in-family)	0.87
	4/8 PSK, 64/128 QAM (cross-fam. w. exclusion)	0.50
	4/16 PSK, 64/128 QAM (cross-fam. w. exclusion)	0.50
	8/16 PSK, 64/128 QAM (cross-fam. w. exclusion)	0.48
	2/4/8/16 PSK, 64/128 QAM (cross-fam. w/o exclusion)	0.50
	2/4/8/16 PSK, 64/128 QAM (exhaustive)	0.50
2/8 PSK	4/16 PSK (in-family)	0.99
	4/16 PSK, 16/32 QAM (cross-fam. w. exclusion)	0.50
	4/16 PSK, 16/64 QAM (cross-fam. w. exclusion)	0.78
	4/16 PSK, 16/128 QAM (cross-fam. w. exclusion)	0.90
	4/16 PSK, 16/128 QAM (cross-fam. w/o exclusion)	0.50
	4/16 PSK, 16/32/64/128 QAM (exhaustive)	0.50
4/16 PSK	2/8 PSK (in-family)	0.50
	2/8 PSK, 16/32 QAM (cross-fam. w. exclusion)	0.88
	2/8 PSK, 16/64 QAM (cross-fam. w. exclusion)	0.96
	2/8 PSK, 16/128 QAM (cross-fam. w. exclusion)	0.97
	2/8 PSK, 16/128 QAM (cross-fam. w/o exclusion)	0.99
	2/8 PSK, 16/32/64/128 QAM (exhaustive)	0.97

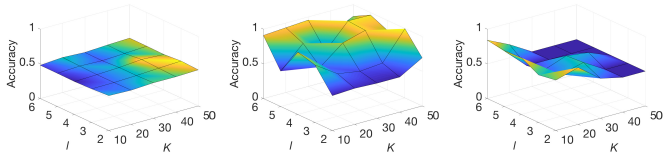
TABLE III: Effects of training mixture on in-family modrec (SNR=20dB).

D. Effects of parameter selection

In this section, we evaluate the impact of input parameters on performance. We investigate the effects of local sequential pattern extraction, specifically focusing on shingle and dictionary size. We also consider the instance size (i.e. the number



(a) Syn. 4/16PSK 20dB ET (b) OTA 4/16PSK 20dB ET (c) Syn. 8/16PSK 20dB ET



(d) OTA 8/16PSK 20dB ET (e) Syn. 8/16PSK 20dB BT (f) Syn. 8/16PSK 10dB BT

Fig. 4: Effects of feature parameter selection l, K on accuracy. Depicted are two in-family combinations from synthetic and over-the-air data. (a) - (d) use exhaustive training (ET); (e), (f) use balanced training without exclusion (BT). Overall, higher shingle and dictionary size lead to better performance, however, the optimal values vary depending on the training mixture, channel conditions (i.e. synth. vs. OTA) and SNR.

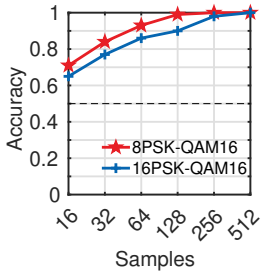


Fig. 5: Accuracy as a function of the input instance size.

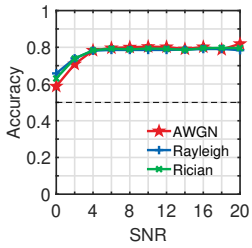


Fig. 6: Accuracy across SNR for different realistic channels.

	Test		
↓ Train	2dB	10dB	20dB
2dB	0.9	0.5	0.5
10dB	0.5	1.0	1.0
20dB	0.5	0.5	1.0
Mixed	0.5	1.0	1.0

Fig. 7: Effects of training SNR on accuracy for 16PSK+QAM16.

of IQ samples per measurement instance), which relates to the amount of data per class we need in order to perform modrec.

1) Effects of feature extraction: As detailed in §IV there are two essential parameters in the feature extraction framework [34]: shingle l and dictionary size K. While fully-supervised modrec was not sensitive to these parameters [34], MODELESS' performance is affected by their selection. We evaluate the accuracy across varying shingle and dictionary sizes with different training mixtures and SNR. Fig. 4 shows our results for two representative test cases from the PSK family: 4/16PSK and 8/16PSK with synthetic and OTA data. (4a)-(4d) use exhaustive training (ET), whereas (4e)-(4f) use balanced without exclusion (BT). Overall, higher shingle and dictionary size lead to better performance, however, the optimal l, K vary depending on the training mixture, channel conditions and SNR. Thus, further exploration is needed in feature design/tuning for zero-shot transfer learning.

2) Effects of input sample size: We now evaluate the effects of the input instance size (number of IQ samples) on classification accuracy. Fig. 5 shows the result for two cross-family test cases: 8PSK+16QAM (red) and 16PSK+16QAM (blue). The accuracy increases with the size of the input instance. In addition, the 16PSK+16QAM combination achieves maximal accuracy at instance size of 512 IQ samples, whereas the 8PSK+16QAM combination requires four times smaller instance size (128 IQ samples) to achieve the same accuracy. Based on these results we postulate that (i) higher order modulations and (ii) test cases with similar modulation orders require larger instance sizes. These findings inform the instance size of 512 IQ samples for our evaluation.

E. Effects of SNR and channel conditions

The SNR and channel conditions can significantly impact a modulation's constellation shape and in turn, the performance of our methodology. Thus, we first evaluate MODELESS under two realistic channel models: (i) Rician and (ii) Rayleigh, and in comparison with AWGN. Fig. 6 presents average accuracy across all 28 combinations for synthetically-generated data, and shows that MODELESS is robust to channel conditions.

We then evaluate MODELESS across various combinations of testing and training SNR, seeking to understand whether training should be SNR-aware (i.e. training performed on the same SNR as testing) or SNR-blind. Fig. 7 presents our result for one test combination: 16PSK+16QAM. Vertically, are the training SNRs of 2, 10, 20dB and a mix of the three. Horizontally are the testing SNR: 2, 10 and 20dB, respectively. Each result in the table presents the accuracy achieved for the particular testing/training combination. For very low SNR regimes (i.e. 2dB), SNR-aware training is necessary. For a wide range of SNR settings, however, training on mixed signals gains perfect modulation recognition, which underpins MODELESS' potential for SNR-blind modrec.

VI. Discussion

While our analysis shows promise for modrec with limited supervision, there are several avenues for further exploration.

Counteracting biases in application data distribution. As noted earlier (§V-A), some application data streams may gain non-uniform symbol distribution creating biases in the symbol representation. In addition, malicious transmitter activity may intentionally bias the application data stream to obfuscate the used modulation and hamper modrec efforts [25]. While the local sequential pattern features are robust to constellation bias [34], further investigation of the similarity embeddings is necessary to make them robust to such biases.

Data-driven training mix selection. We show that the training mixture has a clear impact on classification accuracy in some test combinations. Future work should investigate principled and data-driven approaches to training mix selection to facilitate uniform modrec performance across all test combinations.

Further exploration that includes modulation families beyond PSK and QAM would advance our understanding of MODELESS' applicability and underpin a richer set of test combinations. This will allow performance analysis with one or more than two unknown classes, shedding light on performance across various degrees of supervision. Finally, extension to unknown number of unseen classes will aid the generalizability of our approach and its applicability to real-world spectrum sensing, where the number of unknown modulations may not be apriori available.

VII. Conclusion

This paper tackles automatic recognition of previously unseen modulations using limited supervision. Our method, dubbed MODELESS employs data-driven similarity embeddings, state-of-the-art signal features and lightweight SVM classification for robust modrec. MODELESS does not require extensive expert input and is, thus, highly applicable in the face of constant waveform innovation. We evaluate MODELESS' performance on both synthetic and real over-the-air data

and demonstrate near-perfect classification accuracy in cross-family modrec across a wide range of SNR regimes (from 6 to 20dB) and in both synthetic and real traces. We also pinpoint key drawbacks in in-family classification and draw paths for future research to tackle these limitations.

Our work addresses a key disconnect between modrec training requirements and labeled data availability, which will further widen with the advent of new radio technologies. MODELESS shows promising results on a particularly challenging over-the-air dataset, which underlines its potential for robust modrec with limited supervision in the wild. Our proposed framework is highly-modular and allows for experimentation with alternative similarity embeddings, signal features and classifiers, which constitutes a solid foundation for future work in modrec with practical applications to shared-spectrum access, security and enforcement.

VIII. Acknowledgements

This work is funded by NSF CAREER grant CNS-1845858 and NSF Smart and Connected Communities (S&CC) grant CMMI-1831547.

References

[1] DeepSig Inc. https://www.deepsig.io/.

[2] H. Abuella and M. K. Ozdemir. Automatic modulation classification based on kernel density estimation. *Canadian Journal of Electrical and Computer Engineering*, 39(3):203–209, 2016.

[3] M. W. Aslam, Z. Zhu, and A. K. Nandi. Automatic modulation classification using combination of genetic programming and knn. *IEEE Trans. Wireless Communications*, 11(8):2742–2750, 2012.

[4] K. Bu, Y. He, X. Jing, and J. Han. Adversarial transfer learning for deep learning based automatic modulation classification. *IEEE Signal Processing Letters*, 2020.

[5] M. Di Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo. Spatial modulation for generalized mimo: Challenges, opportunities, and implementation. *Proceedings of the IEEE*, 102(1):56–103, 2013.

[6] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su. Survey of automatic modulation classification techniques: classical approaches and new trends. *IET communications*, 1(2):137–156, 2007.

[7] O. A. Dobre, Y. Bar-Ness, and W. Su. Higher-order cyclic cumulants for high order modulation classification. In *IEEE MILCOM*, Boston, MA, 2003.

[8] Y. Fu, T. Xiang, Y.-G. Jiang, X. Xue, L. Sigal, and S. Gong. Recent advances in zero-shot recognition: Toward data-efficient understanding of visual content. *IEEE Signal Processing Magazine*, 35(1):112–125, 2018.

[9] H. Gang, L. Jiandong, and L. Donghua. Study of modulation recognition based on HOCs and SVM. In *IEEE VTC Spring*, Milan, Italy, 2004.

[10] L. Han, F. Gao, Z. Li, and O. A. Dobre. Low complexity automatic modulation classification based on order-statistics. *IEEE Trans. Wireless Communications*, 16(1):400–411, 2017.

[11] S. J. Hwang, F. Sha, and K. Grauman. Sharing features between objects and their attributes. In *Proc. IEEE CVPR*, pages 1761–1768, 2011.

[12] C. H. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *Proc. IEEE CVPR*, pages 951–958, 2009.

[13] P. Li, Q. Wang, and L. Zhang. A novel earth mover's distance methodology for image matching with gaussian mixture models. In *Proc. IEEE ICCV*, pages 1689–1696, 2013.

[14] G. Lu, K. Zhang, S. Huang, Y. Zhang, and Z. Feng. Modulation recognition for incomplete signals through dictionary learning. In *IEEE WCNC*, pages 1–6, San Francisco, CA, 2017.

[15] F. Meng, P. Chen, L. Wu, and X. Wang. Automatic modulation classification: A deep learning enabled approach. *IEEE Transactions on Vehicular Technology*, 67(11):10760–10772, 2018.

[16] T. Mensink, E. Gavves, and C. G. Snoek. Costa: Co-occurrence statistics for zero-shot classification. In *Proc. IEEE CVPR*, pages 2441–2448, 2014.

[17] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

[18] A. K. Nandi and E. E. Azzouz. Modulation recognition using artificial neural networks. *Signal processing*, 56(2):165–175, 1997.

[19] M. Norouzi, T. Mikolov, S. Bengio, Y. Singer, J. Shlens, A. Frome, G. Corrado, and J. Dean. Zero-shot learning by convex combination of semantic embeddings. In *International Conference on Learning Representations*, 2014.

[20] T. J. O'Shea, K. Karra, and T. C. Clancy. Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention. In *2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 223–228. IEEE, 2016.

[21] T. J. O'Shea, T. Roy, and T. C. Clancy. Over-the-air deep learning based radio signal classification. *IEEE Journal of Selected Topics in Signal Processing*, 12(1):168–179, 2018.

[22] S. J. Pan, Q. Yang, et al. A survey on transfer learning. *IEEE Trans. Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

[23] P. Panagiotou, A. Anastasopoulos, and A. Polydoros. Likelihood ratio tests for modulation classification. In *Proc. IEEE MILCOM*, volume 2, pages 670–674, 2000.

[24] E. Perenda, S. Rajendran, G. Bovet, S. Pollin, and M. Zheleva. Learning the unknown: Improving modulation classification performance in unseen scenarios. In *IEEE International Conference on Computer Communications (IEEE INFOCOM 2021)*, 2021.

[25] H. Rahbari and M. Krunz. Full frame encryption and modulation obfuscation using channel-independent preamble identifier. *IEEE Trans. on Information Forensics and Security*, 11(12):2732–2747, 2016.

[26] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin. Deep learning models for wireless signal classification with distributed low-cost spectrum sensors. *IEEE Trans. on Cognitive Communications and Networking*, 4(3):433–445, 2018.

[27] F. Restuccia, S. D'Oro, A. Al-Shawabka, M. Belgiovine, L. Angioloni, S. Ioannidis, K. Chowdhury, and T. Melodia. Deepradioid: Real-time channel-resilient optimization of deep learning-based radio fingerprinting algorithms. In *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2019.

[28] B. Romera-Paredes and P. Torr. An embarrassingly simple approach to zero-shot learning. In *Proc. ICML*, pages 2152–2161, 2015.

[29] Y. Rubner, C. Tomasi, and L. J. Guibas. The Earth Mover's Distance as a metric for image retrieval. 40(2):99–121, 2000.

[30] A. Swami and B. M. Sadler. Hierarchical digital modulation classification using cumulants. *IEEE Trans. Communications*, 48(3):416–429, 2000.

[31] J.-P. Vert and K. Tsuda. A primer on kernel methods. *Kernel methods in computational biology*, 47:35–70, 2004.

[32] N. E. West and T. O'Shea. Deep architectures for modulation recognition. In *Dynamic Spectrum Access Networks (DySPAN), 2017 IEEE International Symposium on*, pages 1–6. IEEE, 2017.

[33] Y. Xian, C. H. Lampert, B. Schiele, and Z. Akata. Zero-shot learning-a comprehensive evaluation of the good, the bad and the ugly. *IEEE Trans. Pattern Analysis and Machine Intelligence*, pages 1–1, 2018.

[34] W. Xiong, P. Bogdanov, and M. Zheleva. Robust and efficient modulation recognition based on local sequential iq features. In *Proc. IEEE INFOCOM*, pages 1612–1620. IEEE, 2019.

[35] W. Xiong, P. Bogdanov, and M. Zheleva. Towards scalable zero-shot modulation recognition. *VTC-Fall*, 2020.

[36] W. Xiong, L. Zhang, M. McNeil, P. Bogdanov, and M. Zheleva. Exploiting self-similarity for under-determined mimo modulation recognition. In *Proc. IEEE INFOCOM*. IEEE, 2020.

[37] M. Yang, H. Rastegarfar, and I. B. Djordjevic. Sdn-enabled adaptive modulation and coding in hybrid c-rans. In *2019 21st International Conference on Transparent Optical Networks (ICTON)*, pages 1–4. IEEE, 2019.

[38] L. Zhang, T. Xiang, and S. Gong. Learning a deep embedding model for zero-shot learning. In *Proc. IEEE CVPR*, pages 3010–3019, 2017.

[39] Z. Zhang and V. Saligrama. Zero-shot learning via semantic similarity embedding. In *Proc. IEEE ICCV*, pages 4166–4174, 2015.

[40] Z. Zhou, B. Vucetic, M. Dohler, and Y. Li. Mimo systems with adaptive modulation. *IEEE transactions on Vehicular Technology*, 54(5):1828–1842, 2005.