

Received November 30, 2021, accepted December 16, 2021, date of publication January 4, 2022, date of current version January 10, 2022.

Digital Object Identifier 10.1109/ACCESS.2021.3140146

DeepAoANet: Learning Angle of Arrival From Software Defined Radios With Deep Neural Networks

ZHUANGZHUANG DAI¹, YUHANG HE, VU TRAN¹, NIKI TRIGONI¹, AND ANDREW MARKHAM

Department of Computer Science, University of Oxford, Oxford OX1 3QD, U.K.

Corresponding author: Zhuangzhuang Dai (zhuangzhuang.dai@cs.ox.ac.uk)

This work was supported by the National Institute of Standards and Technology (NIST) via the Pervasive, Accurate, and Reliable Location-based Services for Emergency Responders under Grant 70NANB17H185.

ABSTRACT Direction finding and positioning systems based on RF signals are significantly impacted by multipath propagation, particularly in indoor environments. Existing algorithms (e.g MUSIC) perform poorly in resolving Angle of Arrival (AoA) in the presence of multipath or when operating in a weak signal regime. We note that digitally sampled RF frontends allow for the easy analysis of signals, and their delayed components. Low-cost Software-Defined Radio (SDR) modules enable Channel State Information (CSI) extraction across a wide spectrum, motivating the design of an enhanced AoA solution. We propose a Deep Learning approach for deriving AoA from a single snapshot of the SDR multichannel data. We compare and contrast deep-learning based angle classification and regression models, to estimate up to two AoAs accurately. We have implemented the inference engines on different platforms to extract AoAs in real-time, demonstrating the computational tractability of our approach. To demonstrate the utility of our approach we have collected IQ (In-phase and Quadrature components) samples from a four-element Universal Linear Array (ULA) in various Line-of-Sight (LOS) and Non-Line-of-Sight (NLOS) environments, and published the dataset. Our proposed method demonstrates excellent reliability in determining number of impinging signals and realized mean absolute AoA errors less than 2° .

INDEX TERMS Angle-of-Arrival, deep neural networks, machine learning, signal processing, software defined radio.

I. INTRODUCTION

The modern digital world is constructed in part by the ubiquity of wireless communication, enabling information exchange without mobility constraints. Although the primary purpose is typically for communication, RF signatures in the frequency domain become valuable sources of information in their own right. These signals, also called Signals of Opportunity, have sparked numerous applications such as jammer detection [1], location fingerprinting [2], radio planning etc.

Estimating Angle-of-Arrival (AoA) is one of the most popular radiogoniometric parameters to provide positional awareness about RF signals. Based on their importance, a number of classical algorithms have been proposed to accurately estimate bearing angles, typically based on

electromagnetic propagation theory and physical models. One of the most well known is the MUSIC [3] approach for multiple source estimation, with the ability to provide super-resolution beyond the resolution of the antenna array. However, MUSIC performs poorly in Non-Line-of-Sight (NLOS) conditions, which is further compounded by operation in low Signal-to-Noise Ratio (SNR) regimes. Moreover, both super-resolution methods and iterative Maximum Likelihood methods require knowing the number of impinging signals *a-priori* [4], [5].

As an alternative, data-driven or learning-based AoA methods have drawn increasing attention [6], [7] in the past decade. However, we note that none of the existing data-driven methods utilize real-world channel data for training. Rarely have these methods, which are taught by synthetic data, been validated in real-world environments [6]. To some extent, this is due to the high cost and challenges of obtaining basebanded IQ data, as most RF front ends for AoA

The associate editor coordinating the review of this manuscript and approving it for publication was Li He¹.

estimation are dedicated to a specific use case. The high-end USRPs (Universal Software Radio Peripheral) which cover VHF (Very High Frequency), UHF (Ultra High Frequency), and SHF (Super High Frequency) bands are costly, limiting their adoption. We instead focus on investigating how low-cost Software-Defined Radio (SDR), e.g., RTL-SDR dongles, makes passive sensing of AoAs possible with a minimal setup. Moreover, state-of-the-art AoA estimator [8] which is blind of the impinging signals relies on a 16-element Universal Linear Array (ULA). Yet has it been validated in real-world measurement. To this end, we have proposed a hybrid classification and regression DNN model to predict the AoAs from SDR arrays agnostic of the number of impinging signals or the type of modulation.

We note that AoA is encapsulated through phase-differences between channels due to propagation delays [9]. Deep Learning based methods excel at extracting highly-abstract and representative features directly from a large amount of data. This makes Deep Learning an excellent candidate in attending AoA from quadrature demodulated IQ data which are noisy and prone to coupling effect. Regularities of the phase delays can be further revealed by computing correlations among channels, also known as the sample covariance matrix. This acts to reduce the impact of low SNR through correlation over a window of samples. Using the sample covariance matrix to estimate AoA also negates the effect of symbol modulation schemes, pulse shaping, or carrier frequency drift [10]. To this end, we have developed a supervised learning approach to AoA using low-cost SDRs. Given a limited number of antenna array elements, our proposed method can estimate multiple AoAs from a single-snapshot sample covariance matrix. However, this task is non-trivial. There are four main difficulties that need to be addressed:

- i Estimating AoAs in multipath conditions with low SNRs remains a long-standing problem.
- ii Classic AoA methods have a limited Field of View (FOV) [8] when using a ULA which is prevailing. ULA's (as opposed to Uniform Circular Arrays, UCAs) are commonly encountered due to their slim form-factor.
- iii A blind sensing approach in the physical layer to AoAs is confronted by agnosticism of the signal types, number of sources, or symbol modulation.
- iv Constructing a balanced dataset that allows ergodic supervised-learning is a formidable task because the sheer varieties of multipath channel conditions are infinite.

To the best of our knowledge, state-of-the-art data-driven methods of AoA estimation are all based on training with synthetic data [6], [10]–[12]. Furthermore, real-time performance validation upon multiple signals in multipath environment remains scarce. We utilize a supervised learning approach to predicting up to two AoAs in complex indoor environments. We have taken real-world tests to validate our approach. The source code of this project has been made

accessible to the public.¹ In addition, we have made our dataset accessible to foster research in this area.²

Our contributions are threefold: (1) We have published the sample covariance matrix feature vectors used for supervised AoA learning, and the corresponding ground-truth AoAs from various real-world environments; (2) We have proposed a series of Deep Neural Network models, termed DeepAoANet, which classify the number of impinging signals before predicting their AoAs accurately without prior knowledge of the signals; (3) Using KerberosSDR, we have developed a real-time AoA estimation and visualization tool to facilitate a low-cost and low-power AoA solution.

The rest of this paper is organized as follows. Section II reviews existing AoA estimation methods and advocates the motivations of our approach. The RF front end, DeepAoANet architecture, and dataset composition are expanded in details in Section III. Performance evaluation of the DeepAoANet is shown in Section IV. Section V summarizes this work.

II. BACKGROUND

Passive sensing of RF primitives, such as RSSI (Received Signal Strength Indicator), TDOA (Time Difference of Arrival), AoA etc., has seen broad applications in radar, localization, and source detection. RSSI provides a coarse measurement of signal quality versus noise. A limited positioning precision is achievable with RSSI alone [13], [14]. Localization using TDOA requires rigorous synchronization between transmitter and receivers [15], which can be expensive and challenging to achieve. AoA on the other hand is a competitive candidate for positioning and source detection in terms of accuracy and cost [16].

Listening to RF signals non-intrusively is made possible with low-power and low-cost SDRs. The most remarkable advantage of SDR is the flexibility of adapting to new RF needs based on a common hardware platform [17]. Compared to the high-end USRP, an RTL-SDR dongle, costing under \$20, is a significantly low-cost, low-power, and compact transceiver of VHF and UHF signals [18]. The latest RTL-SDR V3 unit produces a maximum sampling rate of 3.2MHz covering frequency bands from 24MHz up to 1.7GHz. This gives rise to potentially an extremely low-cost passive AoA estimator in the corresponding frequency bands.

N. BniLam *et al.* [19] proposed an AoA estimator of LoRa signals based on an SDR front end and found 2° mean error in LOS (Line-of-Sight), and 10° mean error in NLOS. However, a hand-crafted crystal oscillator has to be integrated to accomplish synchronization among channels. The emerging KerberosSDR [20] comprises four coherent RTL-SDR dongles. The shared crystal oscillator and calibrator board with internal noise source allow sample and phase synchronization with ease. The KerberosSDR makes a cost-effective AoA solution possible.

¹<https://github.com/zdai257/DeepAoANet>

²https://drive.google.com/drive/folders/1421NOSQcveTE-TpKAM6cPg3SN_vatJPQ?usp=sharing

Conventional AoA algorithms, such as Capon and MUSIC etc., suffer from accuracy degradation in the presence of multipath or a low SNR [4], [21]. Despite widely used with a desirable form factor, the ULA usually results in a limited FOV [8]. Taking MUSIC for example, the Wave Vector, \vec{k} , propagating along the x -axis is defined as

$$\vec{k} = \frac{2\pi}{\lambda} \sin(\theta) \quad (1)$$

where the wavelength is λ ; the impinging angle with respect to the antenna array is θ . The magnitude of the Wave Vector is Wave Number, $k = 2\pi/\lambda$. Given M array elements, the position of the m -th element of the array placed along y -axis can be written as;

$$\vec{y}_m = (0, y_m) \quad (2)$$

and

$$y_m = \alpha (m - 1) \lambda \quad (3)$$

given the inter-element spacing factor, α . The steering vector represents the set of phase delays a plane wave experiences evaluated at a set of array elements, which can be expressed as;

$$\mathbf{a}(\theta) = [e^{-j\vec{k}\vec{y}_1}, e^{-j\vec{k}\vec{y}_2}, \dots, e^{-j\vec{k}\vec{y}_M}]^T \quad (4)$$

Hence, the received signal, \mathbf{x} , can be expressed as;

$$\mathbf{x} = \sum_{l=1}^L \mathbf{a}(\theta_l) \cdot s_l + \mathbf{n} \quad (5)$$

where L is the number of impinging signals, s denotes the source signal, and \mathbf{n} is the zero-mean Gaussian noise vector. Given a sampling window of length N , the sample covariance matrix, \mathbf{R} , reads

$$\mathbf{R} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}\mathbf{x}^H \quad (6)$$

where \mathbf{x}^H denotes the Hermitian Transpose of the received signal.

In order to obtain AoAs of multiple sources, MUSIC takes an eigen-decomposition approach to distinguishing signals from noise which co-exist in a plane wave. The basic hypothesis lies in the orthogonality of signal and noise in the subspace. The eigenvalues and eigenvectors of the signal and noise can be written as;

$$\mathbf{R} = \mathbf{Q}_s \mathbf{D}_s \mathbf{Q}_s^H + \mathbf{Q}_n \mathbf{D}_n \mathbf{Q}_n^H \quad (7)$$

where \mathbf{Q}_s is the signal subspace formed by eigenvectors corresponding to large eigenvalues, and \mathbf{D}_s is a diagonal matrix containing eigenvalues; Likewise, \mathbf{Q}_n is the noise subspace formed by eigenvectors corresponding to small eigenvalues, and \mathbf{D}_n is the diagonal matrix.

In ideal conditions, the steering vector in the signal subspace is orthogonal to the noise subspace;

$$\mathbf{a}(\theta) \cdot \mathbf{Q}_n = 0 \quad (8)$$

Thereafter, MUSIC algorithm is implemented by minimizing the following term (we refer the readers to [3] for details of derivation);

$$\theta_{MUSIC} = \underset{\theta}{\operatorname{argmin}} \mathbf{a}^H(\theta) \mathbf{Q}_n \mathbf{Q}_n^H \mathbf{a}(\theta) \quad (9)$$

Thus, a search is launched over the steering vector candidates to identify the ones that maximize the power spectrum of the received signal;

$$P(\theta) = \frac{1}{\mathbf{a}^H(\theta) \mathbf{Q}_n \mathbf{Q}_n^H \mathbf{a}(\theta)} \quad (10)$$

However, MUSIC relies on the orthogonality of signal and noise subspace which does not hold in real-world multipath conditions. Moreover, the model-based derivation of the above power spectrum is susceptible to array positional imperfections and mutual coupling [6]. To overcome these challenges, the research community have moved towards learning-based AoA in the past few years. Some works attempt to learn AoAs from base-band IQ data directly [9], [11]. However, received data from RF front ends are subject to various channel characteristics, such as antenna gains, mutual coupling between array elements, negative SNRs etc. Only when the SNR is above 18dB with a fixed window length does the method in [11] produce acceptable results. In addition, almost all works have used synthetic data for training and validation. It is inevitable that these networks overfit to the assumptions used to simulate the IQ data, which may not hold in the real world.

Learning AoAs from the sample covariance matrix have also drawn increasing attention recently [12]. The sample covariance matrix calculates correlations among channels while masking the morphology of the signal itself. For instance, SVR (Support Vector Regression) has seen success in generating robust and accurate AoAs from the sample covariance matrix [22]. SVR is a member of the SVM (Support Vector Machine) family. It shines in mapping multi-variate non-linear sample covariance matrix features to latent phase differences. A supervised learning method by applying SVM upon PDPs (Power Delay Profile) of antenna arrays to estimate AoA has been proposed in [7].

Nevertheless, regression cannot resolve the ambiguity of unknown number of signals at different time instants [23], leading the majority of existing learning approaches to focus on estimating a single source at a time. A two-stage classification model has been proposed in [6]. In the first stage, a multi-output autoencoder extracts phase delay features from the sample covariance matrix into pre-defined angular subregions. A one-versus-all classifier then determines if a signal exists in each angular subregion. In addition, synthetic noise is introduced in the training data as a form of data augmentation, including gain-phase inconsistency, sensor position errors, and mutual antenna coupling, to increase robustness of the algorithm. However, a fine-grained angular resolution trades off system efficiency and simplicity in constructing the training data, since angular bins of 0.1° require 1201

classifiers given a FOV = 120°, as well as an immense set of ground-truth labels for training.

In addition, the vast majority of the data-driven methods utilize artificial signals [8], [11], [24], and few have been validated using real-world signals [6]. This is probably due to the sheer complexity of real-world RF propagation and signal diversity and the challenges of obtaining sufficient ground-truth data for balanced training. We note that similar issues exist in the acoustic signal processing and speech recognition community. The AoA of acoustic signals have been widely studied using temporal-spectral features from microphone arrays [25], [26]. Despite huge differences in propagating speed and wavelength, acoustic and RF AoA estimation share common challenges such as multipath, also known as reverberation, and interference of multiple sources [27]. S. Adavanne *et al.* [28] proposed a Convolutional Recurrent Neural Network to detect multiple indoor sound sources from the spectrogram of a four-element array. Y. He *et al.* proposed SoundDet [29] which is an end-to-end sound classification and direction finding system.

In summary, we observe that (a) classical learning approaches, although mature and well studied, do not perform well in high multipath and low SNR conditions, (b) existing learning approaches are predominantly only able to resolve a single source, and (c) more critically, learning approaches are largely trained and validated with simulated data which is not realistic. This leads us to design a learning based system to overcome these challenges.

III. DeepAoANet

We develop a data-driven algorithm to predict AoA's robustly and accurately in multipath conditions. In doing so, we first create a dataset of sample covariance matrix observations in various scenarios. We use relative GPS positions and a compass for direction alignment to label the impinging LoRa signals [17]. Thereafter, we generate augmented data by introducing phase shifts and multiple levels of Additive White Gaussian Noise (AWGN) to provide sufficiently dense and balanced training labels [30]. We propose DeepAoANet, a hybrid classification and regression Deep Learning model, to perform end-to-end AoA estimation from a single sample covariance matrix snapshot. It is of key importance for DeepAoANet to learn generic AoA features regardless of signal strength, number of sources, or symbol modulation. We further utilize multi-step fine-tuning techniques [31] to jointly train the classification and regression heads. In order to validate the DeepAoANet performance in real-time, real-world conditions, we have developed a Graphic User Interface (GUI) tool allowing AoA visualization with the aid of KerberosSDR. Moreover, we emphasize that our approach works well with a *single* low-cost SDR device with only 4 antennas. The device can passively sniff the channel, detect any RF transmission (within the frequency band) and estimate the AoA accurately. Such a passive AoA detector is extremely useful in many applications such as indoor localization, intruder detection, etc.

A. PROBLEM FORMULATION

Real-world AoA estimation using conventional methods faces numerous challenges. The presence of multipath undermines the assumption that the signal and noise subspace are orthogonal to one another. Coupling between array elements and slight positional imperfections act to reduce the effective FOV. Moreover, it is difficult for a simple peak search algorithm to determine the number of signals from the spatial power spectrum (Eqn. 10).

Specifically, the first challenge in AoA-based systems is the limited FOV of antenna array (for simplicity, we assume the antenna array is a ULA, although we note that any other geometric structure is applicable). In principle, a ULA is capable of estimating the AoA of arriving signal from -90° to $+90^\circ$. However, the resolution degrades significantly towards the antenna axis (near $+90^\circ$ and -90°) and thus limits the effective FOV of the antenna. Figure 1 plots the power pattern of a 4-element antenna array with an inter-element spacing ($\alpha\lambda$) of $\lambda/4$. The figure shows that the -3dB (half the power) beamwidth of the main beam significantly increases when the steering angle approaches 60° (the beamwidth increases from 54° to 132°). Because of the presence of noise, a wider beamwidth will consequently result in higher AoA error.

In real-world environments, robustness to noise is crucial. For AoA-based applications, the size of an antenna array significantly affects the accuracy of AoA estimation in the presence of noise. A smaller antenna array (fewer elements or shorter inter-element spacing) causes wider beamwidth and thus results in lower AoA resolution. Figure 2 shows the power pattern of a ULA with different number of elements (4 and 8) and inter-element spacing ($\lambda/2$ and $\lambda/4$). It is noticeable that the -3dB beamwidth increases more than twice when the number of elements decreases from 8 to 4, or when the inter-element spacing decreases from $\lambda/2$ to $\lambda/4$. Practical indoor systems cannot usually afford an extremely large antenna array for highly accurate AoA estimation. Consequently, a classic highest peak search may fail to detect the correct angle under noisy, multi-path conditions.

Fundamentally, the classic methods are based on a thresholding and/or a searching algorithm to find the most probable impinging signal(s). For example, to find the number of signals, a system can compare the eigenvalues to a threshold; the number of eigenvalues higher than a threshold is considered the number of signals [32]. However, this method is not reliable when the environment sees a large number of traveling paths and when the antenna array has few elements. As an example ([32] required 16 and [33] required 8 antennas, with spatial smoothing, to achieve an acceptable AoA accuracy of around 5°). ArrayTrack [32] also shows that the direct path is not always the most probable path (in the maximum likelihood sense) in multi-path environments, and thus needs to use multiple antenna arrays to resolve the ambiguity.

In this paper, we explore the use of data-driven approaches to resolve the limitations of classic methods. We argue that a Deep Learning method is able to estimate accurately both the number of impinging signals and the AoA of each individual

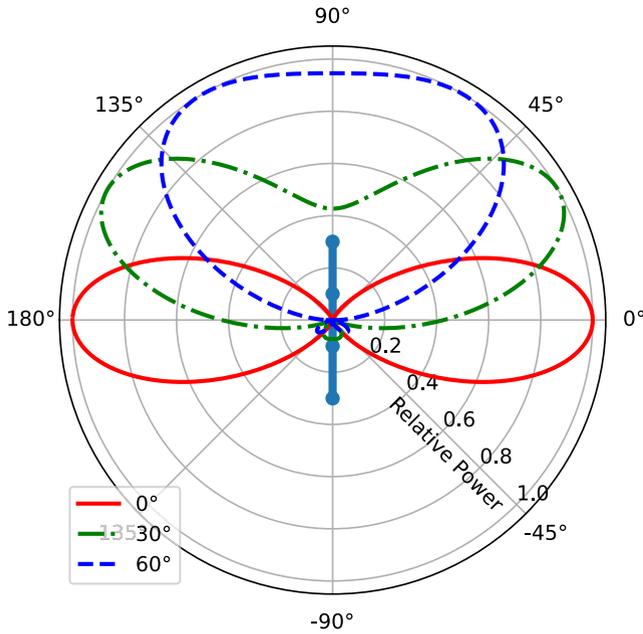


FIGURE 1. Power pattern of a 4-antenna ULA with an inter-element spacing factor (α) of 0.25 with 3 different steering angles of 0° , 30° and 60° respectively. The beamwidth of antenna array becomes larger when the steering angle moving toward $\pm 90^\circ$, and thus more susceptible to noise.

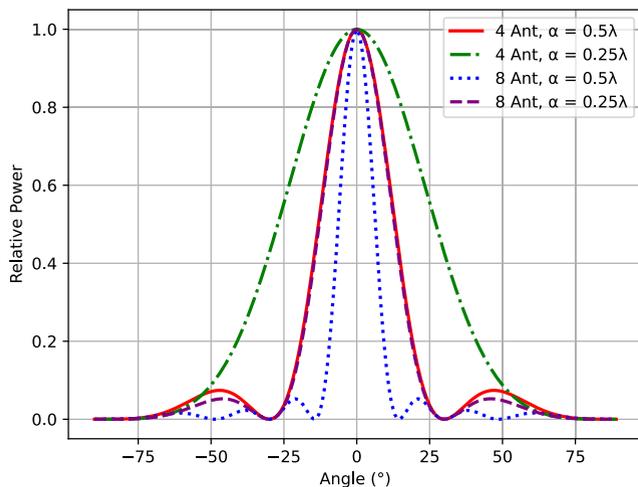


FIGURE 2. Power pattern of 4 or 8-antenna ULA configurations (steering angle = 0°). A ULA with fewer elements, or with a smaller inter-element spacing factor (α), are more susceptible to noise (because of larger beamwidth). However, it may be extremely expensive to accommodate a large number of elements or a large inter-element spacing because of form factor requirements.

signal. Given a passive RF receiver system with M synchronized antennas, arranged in some geometric, sub-wavelength spaced configuration. The system passively listens to the RF channel. Whenever the system detects a spike in RF energy, it will execute an algorithm to estimate the number of impinging signals and the corresponding AoA. Instead of relying on a predefined threshold and peak searching algorithms, we design the DeepAoANet that automatically estimates the number of signals and their respective AoA's. In doing so, we have taken a standard method to firstly

derive the sample covariance matrix following Eqn. 6. The benefits are twofold: Using the sample covariance matrix negates the need to constrain the sampling window to a fixed configuration; Computing the correlations among channels frees the AoA estimation from prior knowledge of signal type or modulation. Therefore, the task of the DeepAoANet is essentially recovering θ_l from \mathbf{R} ;

$$\theta_1, \theta_2, \dots, \theta_L = g(\mathbf{R}) \tag{11}$$

B. DeepAoANet ARCHITECTURE

An intuitive solution to derive accurate AoA's is regression. However, classic regression methods, e.g., SVR [5], cannot compute AoA's directly if the number of signals is unknown. On the other hand, state-of-the-art classification methods require a vast number of angular bins in order to achieve a high angular resolution, as they essentially discretize the angular space. For example, [6] created 1201 angular bins of 0.1° which merely covers an FOV of 120° . This not only makes the classification task complex but demands a more sophisticated dataset, e.g., C_{1201}^2 classes for labeling. We suggest that this could be a fundamental reason which precludes researchers from using measured data for training.

Instead, we argue that a hybrid method of joint classification and regression is potentially a more sensible end-to-end model. The intuition behind our approach is that decision on the number of sources is a natural classification problem i.e. choosing from a discrete set of options, whereas estimating angle of arrival is better suited to a regressor operating over a continuous domain. We developed DeepAoANet which comprises a multilayer perceptron (MLP), a classifier to estimate the number of impinging signals, and multiple regression heads to predict AoA's. We denote this model as DeepAoANet-FC. The network architecture of the DeepAoANet-FC is shown in Fig. 3. The layer details of the DeepAoANet-FC are shown in Table 1. Four fully-connected layers with 'ReLU' (Rectified Linear Unit) activation are utilized to extract latent phase difference features among the four channels [34]. A classification head predicts the number of sources. Two regression heads predict θ_1 and θ_2 , respectively.

In DeepAoANet-FC, the input feature *input* is first rearranged into a one-dimensional vector. Then, several learnable fully-connected layers with weight \mathbf{W} and bias term \mathbf{b} individually are applied to sequentially map the input feature to the output feature *output*, with nonlinear activation *ReLU* between to consecutive fully-connected layers,

$$output = dropout(ReLU(\mathbf{W} \cdot input + \mathbf{b})) \tag{12}$$

where *dropout*(\cdot) is a random dropout operator.

We notice the non-diagonal components of the sample covariance matrix are especially small whenever there is no signal. Prevailing acoustic and RF signal processing research all utilize filtering mechanism to determine the presence of a signal [29], [35]. This is not only efficient but reliable from our observation because sample covariance matrix essentially

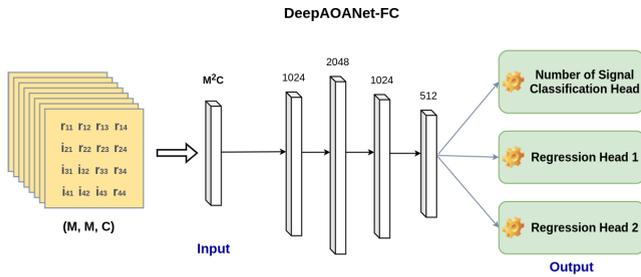


FIGURE 3. Input, output, and the network architectures of the DeepAoANet-FC.

computes the correlation of signal co-occurrences among different channels, whereas, noises are independent. Moreover, it is difficult to classify ‘noise’, as it is subject to RF frontend settings and the environment. We argue a predefined threshold can be used to determine whether signals of interest present. Thereafter, if it detects a signal the number of sources and their AoAs can be estimated using DeepAoANet.

The sample covariance matrix is naturally square-shaped in which the latent phase delays reside. Based on this observation, we propose an alternative network architecture using CNN (Convolutional Neural Network) for feature extraction, namely DeepAoANet-CNN. DeepAoANet-CNN takes the input sample covariance matrix as a $M \times M$ ‘image’, and applies a 2D convolution layer before three fully-connected layers. The input ‘image’ can be expanded to C layers by splitting the sampling windows into C smaller windows. This enriches the temporal features of the input by incorporating temporal variations. The network architecture of DeepAoANet-CNN is shown in Fig. 4.

In DeepAoANet-CNN, the input, *input*, is treated as an image-patch like feature. A 2D convolution *Conv2D* is applied to consume the image-patch like feature. The same nonlinear operator *ReLU*, as well as random dropout, is applied to the 2D convolved feature before obtaining the output feature *output*,

$$output = dropout(ReLU(Conv2D(input))) \quad (13)$$

where Conv2D-processed feature is further fed to fully-connected layers for further process.

TABLE 1. DeepAoANet-FC Neural Network Architecture Illustration. (B is batch size; M is number of array elements; C is number of sample covariance matrix channels).

Layer Name	Units	Output Size
Input	n/a	$[B, M^2 \cdot C]$
FC1	1024	$[B, 1024]$
Dropout1	rate = 0.2	$[B, 1024]$
FC2	2048	$[B, 2048]$
Dropout2	rate = 0.2	$[B, 2048]$
FC3	1024	$[B, 1024]$
Dropout3	rate = 0.2	$[B, 1024]$
FC4	512	$[B, 512]$
Dropout4	rate = 0.2	$[B, 512]$
Classification Head	Activation='sigmoid'	$[B, 1]$
Regression1 Head	Activation='sigmoid'	$[B, 1]$
Regression2 Head	Activation='sigmoid'	$[B, 1]$

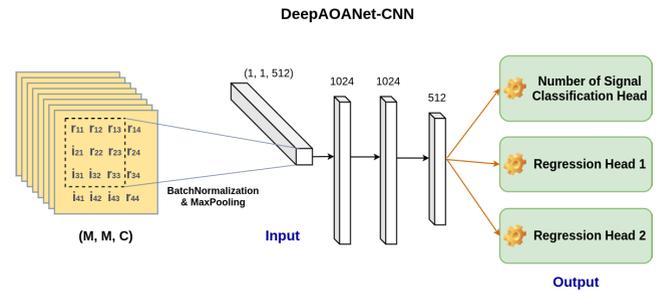


FIGURE 4. Input, output, and the network architectures of the DeepAoANet-CNN.

The network architecture of the DeepAoANet-CNN, as shown in Table 2, is similar in which only the first layer is replaced by a 2D Convolution layer with Batch Normalization (BN) and Max-Pooling. A BN layer is inserted right before the first activation function to produce better convergence.

Since the classification and regression heads use different loss metrics, namely, Binary Cross-Entropy (BCE) and Mean-Squared Error (MSE), a loss weight coordination is necessary in the joint training process. The BCE loss function for the classifier, L_c , can be written as;

$$L_c = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (14)$$

where N is total number of samples, y_i is the ground-truth label of having two sources for the i -th sample, and p_i is the corresponding probability from the ‘sigmoid’ activation. On the other hand, either regression loss, L_r , can be written as;

$$L_r(z, \hat{z}) = \frac{1}{N} \sum_{i=1}^N (z_i - \hat{z}_i)^2 \quad (15)$$

where z is the ground-truth angle and \hat{z} is the prediction. We define the joint loss function, L , as follows;

$$L = \tau \cdot [L_c, L_r(z_1, \hat{z}_1), L_r(z_2, \hat{z}_2)]^T \quad (16)$$

TABLE 2. DeepAoANet-CNN Neural Network Architecture Illustration (B indicates the batch size).

Layer Name	Units	Output Size
Input	n/a	$[B, M, M, C]$
CNN1	512, kernel=3 × 3	$[B, 2, 2, 512]$
BN	n/a	$[B, 2, 2, 512]$
Activation	Activation='ReLU'	$[B, 2, 2, 512]$
MaxPool	pool size=2 × 2	$[B, 1, 1, 512]$
FC2	1024	$[B, 1024]$
Dropout2	rate = 0.2	$[B, 1024]$
FC3	1024	$[B, 1024]$
Dropout3	rate = 0.2	$[B, 1024]$
FC4	512	$[B, 512]$
Dropout4	rate = 0.2	$[B, 512]$
Classification Head	Activation='sigmoid'	$[B, 1]$
Regression1 Head	Activation='sigmoid'	$[B, 1]$
Regression2 Head	Activation='sigmoid'	$[B, 1]$

where τ is a weight vector for the three heads. We discuss further the selection of the weight vector in Section IV.

Notice [8] only considers correctly classified numbers of signal in the mean error calculations which overestimates its realistic performance. We have defined the RMSE and MAE metrics, with penalties for incorrect classifications as follows;

$$RMSE = \begin{cases} \sqrt{\frac{\sum_{i=1}^N \left[\frac{(\hat{\theta}_{i1} + \hat{\theta}_{i2})}{2} - \theta_{i1} \right]^2}{N}} & \text{if } l = 1 \\ \sqrt{\frac{\sum_{n=1}^N \left[(\hat{\theta}_{i1} - \theta_{i1})^2 + (\hat{\theta}_{i2} - \theta_{i2})^2 \right]}{N}} & \text{if } l = 2 \end{cases} \quad (17)$$

$$MAE = \begin{cases} \frac{\sum_{i=1}^N \left\| \frac{(\hat{\theta}_{i1} + \hat{\theta}_{i2})}{2} - \theta_{i1} \right\|}{N} & \text{if } l = 1 \\ \frac{\sum_{n=1}^N \left\| \hat{\theta}_{i1} - \theta_{i1} \right\| + \left\| \hat{\theta}_{i2} - \theta_{i2} \right\|}{N} & \text{if } l = 2 \end{cases} \quad (18)$$

where l is the ground-truth number of impinging signals, $\hat{\theta}_{ik}$ and θ_{ik} are the estimated angle and ground-truth angle of the k -th regression head, respectively.

C. DATASET COMPOSITION

To conduct supervised learning upon DeepAoANet, a crucial step is to construct a balanced and representative dataset. We used intermittent LoRa beacons at 0.5Hz as the vendor signal, which transmits a narrow-band signal at 868MHz in the EU. The benefits are threefold: (1) excellent range in multipath environments; (2) highly configurable air time, transmitting power, and payload [36]; (3) free licensing in the Industrial, Scientific and Medical (ISM) band.

We used KerberosSDR as the receiver for data collection thanks to its low cost and simplified synchronization steps. An open-source driver of KerberosSDR with GUI is provided. The KerberosSDR driver enables real-time AoA visualization in several steps as shown in Fig. 5. Several conventional AoA algorithms are built-in such as Capon [37] and MUSIC. We took the following procedures to setup the KerberosSDR as a passive LoRa receiver;

- Connect KerberosSDR to a laptop through USB and launch the driver.
- Specify the center frequency as 868MHz, sampling rate (2MHz), a uniform gain for four channels (22.9dB), and bandwidth (125kHz).
- Disconnect the antennas. Activate the noise source before conducting sample and phase synchronization.
- Connect the antennas. Apply a band-pass filter, with FIR tap size 100 and decimation 4, at the center frequency.
- Specify the ULA inter-element spacing factor as 0.2. Activate AoA estimation.

We labeled the data using relative GPS positions and a compass for direction alignment. A Dragino LoRa-GPS board stacked on a Raspberry Pi 4 was used as the transmitter. The LoRa payload contains a full NMEA Message [13]. We have specified the LoRa SF (Spreading Factor) equal

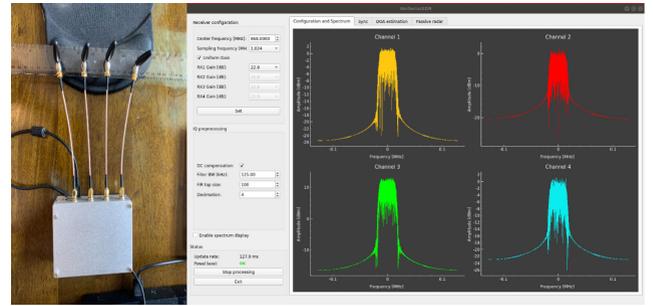


FIGURE 5. On the left is a KerberosSDR connected to four whip antennas with an inter-element spacing factor $\alpha = 0.2$. The screenshot on the right shows the spectrum of the four LoRa channels after band-pass filtering.

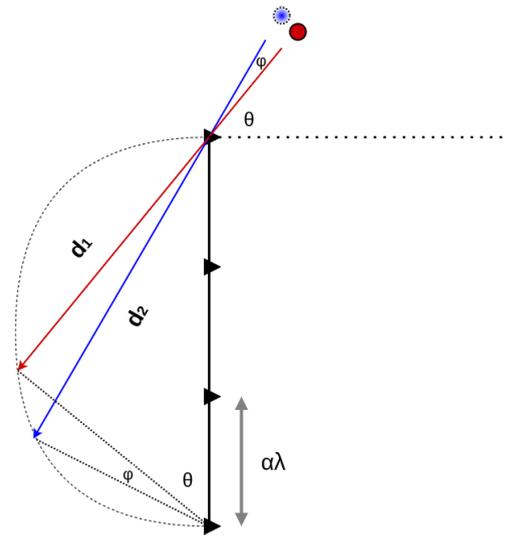


FIGURE 6. Effects of impinging angle change, φ , to the delay of the last channel with respect to the first channel. The red circle represents original source location while the red quiver represents the impinging angle, θ ; Blue quiver stands for the new impinging angle after a phase augmentation where the blue circle is the virtual source; Black triangles are the ULA elements with a spacing factor of α . The phase delay (or advancement), $\Delta\xi$, is equal to $(d_2 - d_1)/\lambda$. Note that smaller impinging angles see greater phase shifts given the same φ .

to 11. This not only increases the range but lengthen the air time to increase the efficiency of signal recording.

During data collection, we collected sample covariance matrix snapshots measured in a variety of propagation conditions. Nearly 50k snapshots of sample covariance matrices were collected in (a) open area, (b) open area with copper sheet as reflectors, (c) office, (d) corridor, and (e) interior-to-outdoor environments. In scenarios (a) and (b), there is no high-rise building or vegetation within 200m. Datasets (c) and (d) were collected in a department building with approximately 20m x 16m floor area. In (e), the transmitter was fixed indoor while the KerberosSDR conducted measurements outside on a parking lot full of cars. The composition of the training dataset is illustrated in Table 3.

Discrete ground-truth AoA labels, between $[-70^\circ, 70^\circ]$ with 10° increment, were recorded by referring to the GPS positions. We observed a threshold can be used upon R_{ij} values to differentiate signals from noise. If over 90% of

TABLE 3. Training Data Composition.

Dataset No.	Data Collection Scenario	Raw	Phase-shifted	Synthetic AWGN or Carrier Phase	Label
1	LOS	8.8k	35.5k	133.3k	$(Class_{L=1}, \theta_1, \theta_1)$
2	LOS with rotating copper reflectors	10.1k	40.2k	150.8k	
3	NLOS office environment	11.5k	46.1k	172.8k	
4	NLOS corridor environment	9.1k	36.3k	136k	
5	NLOS interior-to-outdoor	7.2k	28.8k	108.1k	
6	Combination of [2, 3]	29k	164.3k	164.3k	$(Class_{L=2}, \theta_1, \theta_2)$
7	Combination of [2, 4]	27.6k	156.5k	156.5k	
8	Combination of [2, 5]	23.3k	132.3k	132.3k	
9	Combination of [3, 4]	24.2k	137.2k	137.2k	
10	Combination of [3, 5]	19k	108k	108k	
11	Combination of [4, 5]	19.8k	112.1k	112.1k	

non-diagonal R_{ij} magnitudes are above $1e^{-4}$, signals are detected. For each category of environment, Additive White Gaussian Noises (AWGN) of zero mean and different Standard Deviations have been introduced to enrich the dataset. This emulates receiving signals at different SNRs or ranges.

To approach an enhanced angular resolution, we have introduced phase shifts to the dataset. A discrete angular increment or decrement of 2° is posed to the original data as illustrated in Fig. 6. This angular augmentation is restricted within 4° to best preserve the ratio of raw data versus synthetic and enrich the labels for regression. Therefore, the m -th array element of the receiver sees a phase shift as below

$$\hat{\mathbf{x}} = \mathbf{x} \cdot e^{j\Delta\xi} = \mathbf{x} \cdot e^{j(m-1)\alpha[\sin\theta - \sin(\theta+\varphi)]} \quad (19)$$

where θ is the ground-truth AoA and $\varphi \in \{-4^\circ, -2^\circ, 2^\circ, 4^\circ\}$. Hence, a FOV ranging from -74° to 74° has been created for a fine-grained AoA learning.

Similarly, AoA detection of multiple sources is constructed through superposition of the collected single-source data. The IQ data collected at scenarios other than LOS of different ground-truth AoAs are superimposed to create synthetic IQs. Notice the carrier signals come with stochastic initial phases wherever they interfere. A random phase difference between the carriers is inserted to simulate constructive or destructive interference. The steps are expanded as follows.

The modulated signals at carrier frequency, $F_0 = 868\text{MHz}$, can be written as;

$$\mathbf{m}_1(t) = x_1(t)e^{j\phi_1} e^{j2\pi F_0 t} = [i_1(t) + jq_1(t)] e^{j2\pi F_0 t} \quad (20)$$

$$\mathbf{m}_2(t) = x_2(t)e^{j\phi_2} e^{j2\pi F_0 t} = [i_2(t) + jq_2(t)] e^{j2\pi F_0 t} \quad (21)$$

where $x_1(t)$ and $x_2(t)$ are the base-band signals, and ϕ_1 and ϕ_2 stand for the initial phases of the carrier signals which are independent. Only in the modulated frequency band are the signals additive. As a result, the superimposed signal reads

$$\mathbf{m}_1(t) + \mathbf{m}_2(t) = [x_1(t) + x_2(t)e^{j(\phi_2 - \phi_1)}] e^{j\phi_1} e^{j2\pi F_0 t} \quad (22)$$

If we define the carrier phase difference $\Delta\phi = \phi_2 - \phi_1$, IQ samples of the superimposed signal with random original carrier phases yields

$$i_1(t) + jq_1(t) + [i_2(t) + jq_2(t)] e^{j\Delta\phi} \quad \Delta\phi \in [0, 2\pi) \quad (23)$$

in which $\Delta\phi$ conforms continuous uniform distribution within a period.

The phase differences of the SDR channels lie in the upper triangular components of the sample covariance matrix, R , as values of the upper-right and lower-left triangles are symmetric. The diagonal components are always real numbers equal to variances of the received signals. Therefore, we serialized as inputs the real parts of R in the top-right triangle, concatenated with imaginary parts of the rest;

$$\mathbf{b}_r = [r_{11}, r_{12}, \dots, r_{jk}, \dots, r_{44}]^T \quad \text{for } j \leq k \quad (24)$$

$$\mathbf{b}_i = [i_{21}, i_{31}, \dots, i_{jk}, \dots, i_{43}]^T \quad \text{for } j > k \quad (25)$$

$$\mathbf{b} = [\mathbf{b}_r; \mathbf{b}_i] \quad (26)$$

$$\text{input} = \frac{\mathbf{b}}{\|\mathbf{b}\|} \quad (27)$$

The KerberosSDR sampling window length is configured as 2^{15} . In this work, we chose a sliding window of length 2^{12} to generate an 8-window sample covariance matrix. This not only exploits temporal variations but enriches the input feature space. Since there are four array elements in KerberosSDR, each sample covariance matrix window comprises $M \times M = 16$ floating numbers. Thus, the input vector contains $16 \times 8 = 128$ elements, or a feature ‘image’ of size (4, 4, 8). We utilize standard scaling to normalize all the inputs.

As for labels, we have arranged them as $(Class_L, \theta_1, \theta_2)$, in which θ_1 is populated if only one source exists. In this work, we only consider three circumstances as $L \in \{0, 1, 2\}$. Since no-signal case can be solved effectively with a threshold-based detector, we assigned binary classification of $L = 1$ and $L = 2$ for the classifier. As ground-truth AoAs fall in $[-74^\circ, 74^\circ]$, we uniformly projected θ to be within $(0, 1)$. Note that θ_1 and θ_2 are rearranged in ascending order to avoid confusing the regression heads.

We utilized ROS (Robot Operating System) for field data collection as well as data synthesis. The total number of sample covariance matrix snapshots exceed 2.6 million. The process of data collection, labeling, and training are outlined in Fig. 7. We have further validated that our dataset contains sufficient amount of training sources as well as label diversity in Section IV.

D. AoA INFERENCE AND VISUALIZATION

In order to validate our approach, we built an inference engine of the DeepAoANet for real-time estimation. A GUI is developed to visualize the impinging signals. At runtime, two LoRa sources are configured with different SF and power settings transmitting messages of random lengths. To test the generalization capacity of the DeepAoANet, the transmitters have been placed in different indoor and outdoor environments. An example of the GUI showing two impinging signals are displayed in Fig. 8. A detailed performance analysis is presented in Section IV.

IV. EVALUATION

We have trained and tested the DeepAoANet using Keras with Tensorflow backend [38]. Adam optimizer with an initial learning rate of $1e^{-3}$ is used with a $1e^{-6}$ decaying factor. We modify the weight vector τ in the loss function (Eqn. 16) in a two-stage training process: for the first 40 epochs, τ equals [0.1, 1, 1]; for the last 10 epochs, $\tau = [0.001, 1, 1]$ is assigned. Because the BCE loss stops declining at a relatively large value with classification saturates above 99% accuracy, we suppress its weight in the loss to allow regressions proceed further. We split our dataset into 60% for training, 30% for validation, and 10% for testing, with a batch size of 512. It takes 1.5 hours and 2 hours on a Dell XPS 15 laptop with an Intel Core i7 CPU and a GeForce GTX 1650 GPU to train DeepAoANet-FC and DeepAoANet-CNN, respectively.

To validate the effectiveness of our proposed DeepAoANet, we contrast DeepAoANet's against other classic and regression methods. We have conducted tests in unseen environments with different LoRa configurations under various SNRs. We have investigated the error distributions of DeepAoANet to validate its effective FOV. We have measured the DeepAoANet's latency on different computing platforms to validate its efficiency.

1) COMPARISON TO OTHER METHODS

Since SVR can only deal with single source [22] and MUSIC performance is subject to the selected peak search algorithm, we evaluated DeepAoANet against other methods using only the single-source split with a $0dB$ SNR for a fair comparison in the Cumulative Distribution Function (CDF) plot. The CDFs of DeepAoANet-FC, DeepAoANet-CNN, MUSIC, and SVR are displayed in Fig. 9. As can be seen, both DeepAoANet candidates yield RMSEs of approximately 2.5° for 80% of the tested samples. In contrast, SVR produces a RMSE of more than 10° . Notice MUSIC fails in this benchmark due to its limited FOV when estimating large impinging angles.

2) EFFECTIVE FOV

We have investigated where the errors reside in the angular spectrum using the DeepAoANet candidates versus SVR as shown in Fig. 10. Generally, the errors are evenly distributed. To quantify this statement, we calculated the

error standard deviations when the ground-truth AoAs are within $(-37^\circ, 37^\circ)$ or outside. They are 2.101° and 2.454° using DeepAoANet-FC, or 1.954° and 2.369° using DeepAoANet-CNN, respectively. It can be seen larger impinging angles are more difficult to estimate. In contrast, SVR suffers large estimation errors with large impinging angles. The standard deviation for AoA's within $(-37^\circ, 37^\circ)$ equals 9.202° ; for AoA's outside $(-37^\circ, 37^\circ)$ it is 12.845° . Nevertheless, DeepAoANets have demonstrated a superb stability across a wider FOV in comparison to SVR.

3) ACCURACY VERSUS SNRS

It can be seen from Fig. 11 both DeepAoANet-FC and DeepAoANet-CNN exhibit great resilience dealing with negative SNRs i.e. signals below the noise floor. Either SVR or MUSIC performs poorly. Notice DeepAoANet-CNN demonstrates better reliability when the SNR is less than $-5dB$. We think this owes to the DeepAoANet-CNN's superb power in generalizing to deteriorated SNRs that are lower than what is presented during training.

We investigated the classification performance when the SNR equals $0dB$. The classification confusion matrices of the DeepAoANet-FC and DeepAoANet-CNN are shown in Table 4 and Table 5. The DeepAoANet candidates demonstrate excellent capabilities in identifying the number of impinging signals.

4) REAL-WORLD TEST

To verify DeepAoANet can handle different types of signals, we have placed the KerberosSDR and two LoRa transmitters inside a laboratory with a Vicon system to provide ground-truth device positions of a *mm*-level precision, so as to ground-truth AoAs. The laboratory is in its typical state filled with items. The KerberosSDR remains static while one LoRa source is configured with SF 9 and $17 dBm$ of transmitting power, and the other at SF 12 and $10 dBm$ of transmitting power. Both were generating payloads of random lengths. During the test, one transmitter was fixed and the other was placed at four different locations. The estimated AoAs using DeepAoANet-CNN and the ground-truths are plotted in Fig. 12. This proves the DeepAoANet generalizes well in unseen environments dealing with signals without knowing the symbol modulation.

5) RUN-TIME EFFICIENCY

Both DeepAoANet-FC and DeepAoANet-CNN have merely a few million parameters. This makes them suitable for real-time applications. We have logged the DeepAoANets', as well as MUSIC and SVR's, runtime performance on platforms with different resource constraints, including the Dell XPS15 laptop, an NVIDIA Jetson AGX Xavier with a 512-core Volta GPU, and a Raspberry Pi 4, as shown in Table 6. It can be seen using DeepAoANet inference is more efficient compared to MUSIC (based on *numpy* library). SVR is significantly faster. Nonetheless, the DeepAoANets

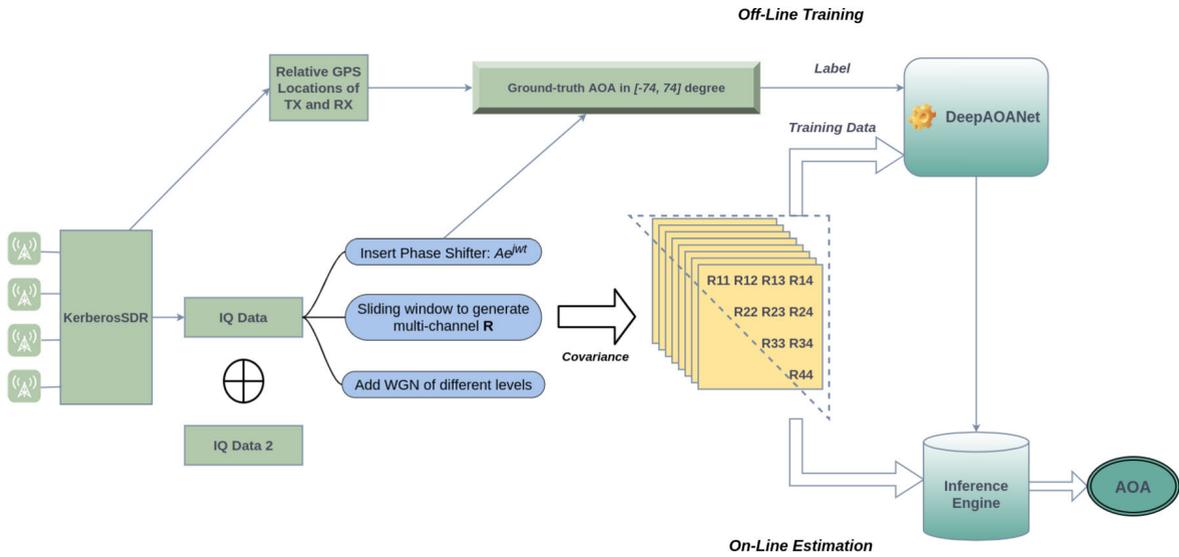


FIGURE 7. A flowchart demonstrating how the offline training and online AoA estimation work.

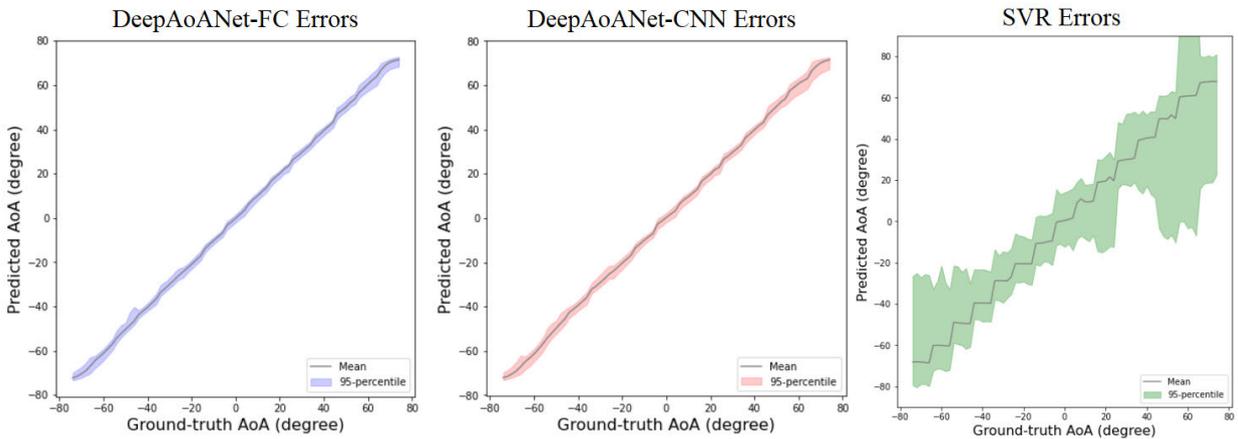


FIGURE 10. Distributions of 95-percentile most accurate AoA’s predictions using DeepAoANet-FC, DeepAoANet-CNN, and SVR.

TABLE 4. Confusion Matrix of DeepAoANet-FC Classifier.

Accuracy (%)		True Number of Signals	
		$L = 1$	$L = 2$
Prediction	$L = 1$	100.0%	0.151%
	$L = 2$	0.0	99.849%
Total Samples		96635	175764

TABLE 5. Confusion Matrix of DeepAoANet-CNN Classifier.

Accuracy (%)		True Number of Signals	
		$L = 1$	$L = 2$
Prediction	$L = 1$	100.0%	0.22%
	$L = 2$	0.0	99.78%
Total Samples		96635	175764

allow inference rates over 60 samples-per-second even with a single-board computer.

6) LIMITATIONS

There are several limitations in our current approach. The DeepAoANet can only handle up to two sources at the moment. This is because the amount of training data becomes exponentially larger with the number of sources. Given 75 angular regions for regression, there are $C_{75}^3 = 67525$ combinations of labels, each of which needs to be contaminated by different levels of noise. We plan to use semi-supervised learning rather than ergodic supervised learning to tackle more than two AoA’s in future work. Although both configurations of DeepAoANet delivered good results dealing with LoRa signals of arbitrary configurations, they failed when we shifted the transmitter center frequency by 100 kHz. The accuracy also severely degraded if we changed the antenna inter-element spacing factor at runtime. As specified in Section III-C, only when the KerberosSDR is tuned to the same center frequency, bandwidth, sampling rate, and

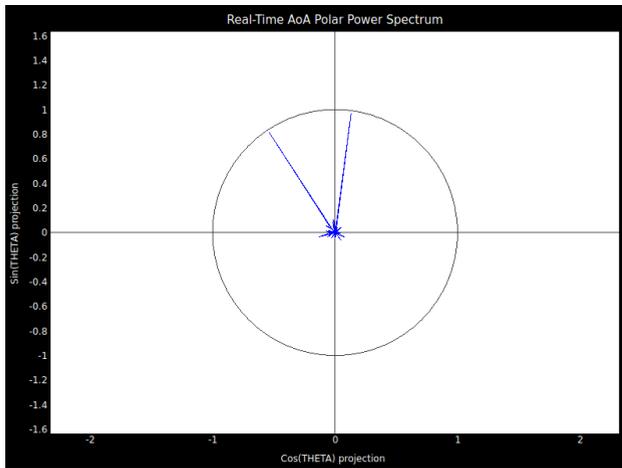


FIGURE 8. An example of visualizing two impinging signals from -30° and 10° using the developed GUI tool.

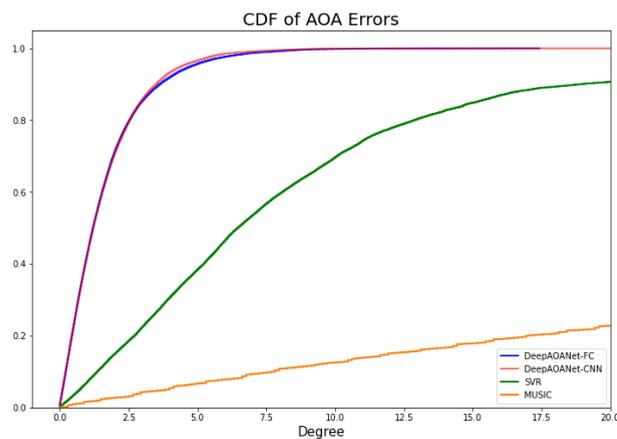


FIGURE 9. CDF of estimating single-source AoA. Notice MUSIC is severely under-performing because its reliable FOV stays between $(-45^\circ, 45^\circ)$ from our observation. Whereas, the ground-truth angles in this benchmark span across $(-74^\circ, 74^\circ)$.

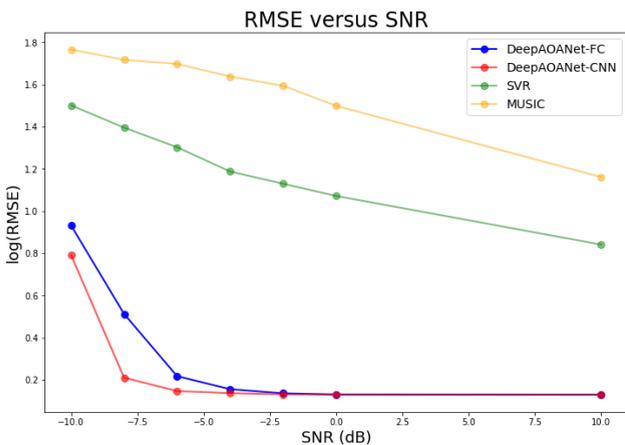


FIGURE 11. Logarithm of RMSE to base 10 of the DeepAoANets, SVR, and MUSIC against varying SNRs. Note SVR is analysed upon single-source split of the testing dataset.

decimation does the DeepAoANet find its ground for phase delay analysis. This is because the IQ data is subject to the RF

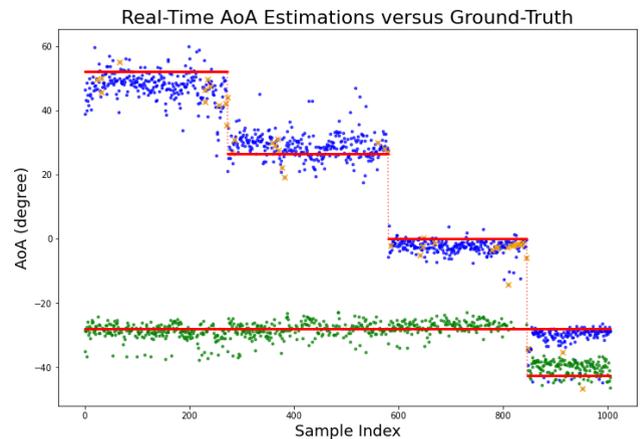


FIGURE 12. Real-time AoA estimations versus the ground-truth using DeepAoANet-CNN. The red lines mark the ground-truth AoAs of the two LoRa sources. The green dots represent the 1st regression output. The blue dots represent the 2nd regression output. The orange crosses stand for samples with wrong classification outcomes.

TABLE 6. Inference latency benchmark.

Model	Parameters	Latency (ms)		
		Laptop	Jetson	Pi 4
MUSIC	n/a	5.973	19.069	61.277
SVR	n/a	0.869	2.071	5.721
DeepAoANet-FC	4.86M	5.988	4.829	15.212
DeepAoANet-CNN	2.14M	4.436	5.116	7.984

front end setup in which the training dataset were collected. In the next step, we will collect more data under different setups in pursuit of a truly generic AoA solution independent of the RF front end or carrier frequency.

V. CONCLUSION

Extracting AoA of RF signals from passive sensing has enjoyed long-standing research interest due to its importance for a number of applications. Recent SDR front ends make a robust and low-cost AoA estimator possible. However, conventional AoA algorithms fall short in multipath environments with low SNRs. Data-driven AoA methods have emerged in the past decade showing capabilities for extracting features from noisy multichannel data. Yet these methods either require prior knowledge of the signal type, number, and modulation, or merely realize offline performance with synthetic data. We propose DeepAoANet, a pair of Deep Learning models, to determine the AoAs of up to two impinging signals. In doing so, we have collected sample covariance matrices derived from received IQ data in various multipath environments and published the dataset. We have validated the reliability of the DeepAoANet in different scenarios, and found a MAE of less than 2° . This sheds light on a pure physical layer approach to AoA acquisition regardless of the propagation environment, transmission time, or modulation of the signals.

In future work, we shall investigate a finer angular resolution for training with different antenna array setups. On top of the current DeepAoANet baseline, we are going to explore deeper network designs. We plan to generalize our approach

to signals of other frequency bands, and also AoA in both azimuth and elevation domains.

REFERENCES

- [1] Z. Dai, P. R. Shepherd, and R. J. Watson, "UAV-aided source localization in urban environments based on ray launching simulation," in *Proc. IEEE Int. Symp. Antennas Propag., USNC/URSI Nat. Radio Sci. Meeting*, Jul. 2017, pp. 595–596.
- [2] L. Wu, D. He, B. Ai, J. Wang, H. Qi, K. Guan, and Z. Zhong, "Artificial neural network based path loss prediction for wireless communication network," *IEEE Access*, vol. 8, pp. 199523–199538, 2020.
- [3] M. Kaveh and A. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 2, pp. 331–341, Apr. 1986.
- [4] K. T. Wong and M. D. Zoltowski, "Uni-vector-sensor ESPRIT for multi-source azimuth, elevation, and polarization estimation," *IEEE Trans. Antennas Propag.*, vol. 45, no. 10, pp. 1467–1474, Oct. 1997.
- [5] M. Pastorino and A. Randazzo, "A smart antenna system for direction of arrival estimation based on a support vector regression," *IEEE Trans. Antennas Propag.*, vol. 53, no. 7, pp. 2161–2168, Jul. 2005.
- [6] Z.-M. Liu, C. Zhang, and P. S. Yu, "Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections," *IEEE Trans. Antennas Propag.*, vol. 66, no. 12, pp. 7315–7327, Dec. 2018.
- [7] M. Yang, B. Ai, R. He, H. Chen, Z. Ma, and Z. Zhong, "Angle-of-arrival estimation for vehicle-to-vehicle communications based on machine learning," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2020, pp. 154–158.
- [8] O. Bialer, N. Garnett, and T. Tirer, "Performance advantages of deep neural networks for angle of arrival estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 3907–3911.
- [9] A. Khan, S. Wang, and Z. Zhu, "Angle-of-arrival estimation using an adaptive machine learning framework," *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 294–297, Feb. 2019.
- [10] J. Yu, W. W. Howard, D. Tait, and R. M. Buehrer, "Direction-of-arrival estimation with a vector sensor using deep neural networks," in *Proc. IEEE 93rd Veh. Technol. Conf. (VTC-Spring)*, Apr. 2021, pp. 1–7.
- [11] M. Gall, M. Gardill, T. Horn, and J. Fuchs, "Spectrum-based single-snapshot super-resolution direction-of-arrival estimation using deep learning," in *Proc. German Microw. Conf. (GeMiC)*, Mar. 2020, pp. 184–187.
- [12] L. Wu, Z.-M. Liu, and Z.-T. Huang, "Deep convolution network for direction of arrival estimation with sparse prior," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1688–1692, Nov. 2019.
- [13] Z. Dai and F. J. W. Podd, "A power-efficient BLE augmented GNSS approach to site-specific navigation," in *Proc. IEEE/ION Position, Location Navigat. Symp. (PLANS)*, Apr. 2020, pp. 1305–1310.
- [14] A. A. Abdallah, K. Shamaei, and Z. M. Kassas, "Indoor positioning based on LTE carrier phase measurements and an inertial measurement unit," in *Proc. 31st Int. Tech. Meeting Satell. Division Inst. Navigat. (ION GNSS)*, Oct. 2018, pp. 3374–3384.
- [15] Z. Li, T. Braun, X. Zhao, Z. Zhao, F. Hu, and H. Liang, "A narrow-band indoor positioning system by fusing time and received signal strength via ensemble learning," *IEEE Access*, vol. 6, pp. 9936–9950, 2018.
- [16] M. T. Rahman, N. Tadayon, S. Han, and S. Valaee, "LoCHunt: Angle of arrival based location estimation in harsh multipath environments," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.
- [17] P. Robyns, P. Quax, W. Lamotte, and W. Thenaers, "A multi-channel software decoder for the Lora modulation scheme," in *Proc. 3rd Int. Conf. Internet Things, Big Data Secur. (IoTBDs)*, 2018, pp. 41–51.
- [18] A. Marquet, N. Montavont, and G. Papadopoulos, "Towards an SDR implementation of LoRa," *Comput. Commun.*, vol. 153, pp. 595–605, Mar. 2020.
- [19] N. BniLam, D. Joosens, M. Aernouts, J. Steckel, and M. Weyn, "LoRay: AoA estimation system for long range communication networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 2005–2018, Mar. 2021.
- [20] Z. Li, D. Zhang, Q. Zhu, H. Gu, S. Huang, Y. Kuang, and Y. Liu, "Application research on DOA estimation based on software-defined radio receiver," in *Proc. 2nd Int. Conf. Electron. Eng. Inform.*, 2020.
- [21] H. Xiang, B. Chen, M. Yang, T. Yang, and D. Liu, "A novel phase enhancement method for low-angle estimation based on supervised DNN learning," *IEEE Access*, vol. 7, pp. 82329–82336, 2019.
- [22] H. Xiang, J. Bin, Z. Jingli, S. Yueguang, and Z. Liu, "Direction of arrival estimation based on smooth support vector regression," in *Proc. 2nd Int. Conf. Future Comput. Commun.*, May 2010, p. V1-818.
- [23] W. Zhu and M. Zhang, "A deep learning architecture for broadband DOA estimation," in *Proc. IEEE 19th Int. Conf. Commun. Technol. (ICCT)*, Oct. 2019, pp. 244–247.
- [24] M. Comiter and H. T. Kung, "Localization convolutional neural networks using angle of arrival images," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–7.
- [25] W. Q. Zheng, Y. X. Zou, and C. Ritz, "Spectral mask estimation using deep neural networks for inter-sensor data ratio model based robust DOA estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 325–329.
- [26] Y. Sun, J. Chen, C. Yuen, and S. Rahardja, "Indoor sound source localization with probabilistic neural network," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6403–6413, Aug. 2018.
- [27] M. J. Bianco, S. Gannot, and P. Gerstoft, "Semi-supervised source localization with deep generative modeling," in *Proc. IEEE 30th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2020, pp. 1–6.
- [28] S. Advanne, A. Politis, and T. Virtanen, "Direction of arrival estimation for multiple sound sources using convolutional recurrent neural network," in *Proc. IEEE Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 1462–1466.
- [29] Y. He, N. Trigoni, and A. Markham, "SoundDet: Polyphonic moving sound event detection and localization from raw waveform," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Jul. 2021.
- [30] A. M. Elbir and K. V. Mishra, "Sparse array selection across arbitrary sensor geometries with deep transfer learning," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 1, pp. 255–264, Mar. 2021.
- [31] Y. Shi, K. Davaslioglu, Y. E. Sagduyu, W. C. Headley, M. Fowler, and G. Green, "Deep learning for RF signal classification in unknown and dynamic spectrum environments," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Nov. 2019, pp. 1–10.
- [32] J. Xiong and K. Jamieson, "ArrayTrack: A fine-grained indoor location system," in *Proc. 10th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, 2013, pp. 71–84.
- [33] V. H. Tran, A. Misra, J. Xiong, and R. K. Balan, "WiWear: Wearable sensing via directional WiFi energy harvesting," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Mar. 2019, pp. 1–10.
- [34] A. H. E. Zooghy, C. G. Christodoulou, and M. Georgiopoulos, "A neural network-based smart antenna for multiple source tracking," *IEEE Trans. Antennas Propag.*, vol. 48, no. 5, pp. 768–776, May 2000.
- [35] I. V. Malygin, S. A. Belkov, M. R. Usvyatsov, and A. D. Tarasov, "Radio signal detection using machine-learning approach," in *Proc. Inf. Technol., Telecommun. Control Syst. (ITTCS)*, 2017, pp. 57–61.
- [36] M. N. Mahfoudi, G. Sivadoss, O. B. Korachi, T. Turetli, and W. Dabbous, "Joint range extension and localization for LPWAN," *Internet Technol. Lett.*, 2019.
- [37] P. Stocia, P. Händel, and T. Söderström, "Study of Capon method for array signal processing," *Circuits, Syst., Signal Process.*, vol. 14, no. 6, pp. 749–770, Nov. 1995.
- [38] F. Chollet et al., "Keras," Tech. Rep. [Online]. Available: <https://github.com/fchollet/keras>



ZHUANGZHUANG DAI received the B.Eng. degree in electrical and electronics engineering from the University of Birmingham and the M.Sc. degree in digital communications and the Ph.D. degree from the University of Bath. Before, he worked in knowledge transfer for the University of Manchester. He is currently a NIST Software Engineer with the Cyber Physical Systems Group, Department of Computer Science, Oxford University. His research interests include embedded systems, RF modeling, computer vision, and indoor and urban localization.



researchers with different backgrounds to tackle the challenges arising from various areas.

YUHANG HE received the B.Eng. degree from Wuhan University, China. He is currently pursuing the D.Phil. degree with the Department of Computer Science, University of Oxford. His research interests include spatial sound perception with deep neural networks, joint audio-visual recognition and navigation, and designing novel neural networks to process various signals in an end-to-end manner, while maintaining neural network interpretability. He also actively collaborates with



VU TRAN received the bachelor's and Master of Engineering degrees from the Ho Chi Minh University of Technology, in 2009 and 2012, respectively, and the Ph.D. degree in computer science from Singapore Management University, in 2020. He is currently a Postdoctoral Research Associate with the Cyber-Physical Systems Group, Department of Computer Science, University of Oxford. His research interests include mobile & wearable sensing, wireless communication & sensing, and indoor localization.



Machines and Systems, a program that combines machine learning, robotics, sensor systems and verification/control. She also leads the Cyber Physical Systems Group, which is focusing on intelligent and autonomous sensor systems with applications in positioning, healthcare, environmental monitoring, and smart cities.

NIKI TRIGONI received the D.Phil. degree from the University of Cambridge, in 2001. She is currently a Professor with the Department of Computer Science, Oxford University, and a fellow of the Kellogg College. She became a Postdoctoral Researcher with Cornell University, from 2002 to 2004, and a Lecturer with the Birkbeck College, from 2004 to 2007. At Oxford, she is currently the Director of the EPSRC Centre for Doctoral Training on Autonomous Intelligent



He designed novel sensors, investigated new algorithms (increasingly deep and reinforcement learning-based) and applied these innovations to solving new problems. Previously, he was an EPSRC Postdoctoral Research Fellow, working on the UnderTracker Project.

ANDREW MARKHAM received the Ph.D. degree from the University of Cape Town, South Africa, in 2008, researching the design and implementation of a wildlife tracking system, using heterogeneous wireless sensor networks. He is currently an Associate Professor working on sensing systems, with applications from wildlife tracking to indoor robotics to checking that bridges are safe. He works with the Cyber Physical Systems Group.

...