Laakso, Mikko; Wichman, Risto

Near-field localization using machine learning: An empirical study

Published in:
2021 IEEE 93rd Vehicular Technology Conference, VTC 2021-Spring - Proceedings

DOI:
10.1109/VTC2021-Spring51267.2021.9449002

Published: 01/04/2021

Document Version
Peer reviewed version

Please cite the original version:
Laakso, M., & Wichman, R. (2021). Near-field localization using machine learning: An empirical study. In *2021 IEEE 93rd Vehicular Technology Conference, VTC 2021-Spring - Proceedings* [9449002] (IEEE Vehicular Technology Conference; Vol. 2021-April). IEEE. https://doi.org/10.1109/VTC2021-Spring51267.2021.9449002

Near-field localization using machine learning: an empirical study

Mikko Laakso* and Risto Wichman*
* *Department of Signal Processing and Acoustics*
Aalto University
Espoo, Finland
{mikko.t.laakso, risto.wichman}@aalto.fi

***Abstract*—Estimation methods for passive near-field localization have been studied to an appreciable extent in signal processing research. Such localization methods find use in various applications, for instance in medical imaging. However, methods based on the standard near-field signal model can be inaccurate in real-world applications, due to deficiencies of the model itself and hardware imperfections. It is expected that deep neural network (DNN) based estimation methods trained on the nonideal sensor array signals could outperform the model-driven alternatives. In this work, a DNN based estimator is trained and validated on a set of real world measured data. The series of measurements was conducted with an inexpensive custom built multichannel software-defined radio (SDR) receiver, which makes the nonidealities more prominent. The results show that a DNN based localization estimator clearly outperforms the compared model-driven method.**

***Index Terms*—near-field localization, deep learning, SDR**

I. Introduction

The problem of passive source localization has been widely studied in array signal processing research. The spherical wavefront in the near-field enables the joint estimation of signal emitter range and direction, using only passive sensing. Such localization estimators are valid in the Fresnel region, before the field is considered a plane wave in the far-field, and field curvature can no longer be discerned. Direction of arrival (DOA) estimation is obviously closely related to the localization problem. From the viewpoint of many estimator algorithms, the only difference is the choice between far- and near-field signal models. Analytically derived estimators for localization or plain DOA estimation include for instance MUSIC and ESPRIT, which belong to super-resolution methods [1], [2].

However, the standard near-field signal model is derived from a purely geometrical perspective and solved with a second-order Taylor series approximation. Friedlander [3] points out, that the commonly used signal-model ignores the signal amplitude and is inaccurate for an electromagnetic field. Furthermore, in real world scenarios, the estimation accuracy is deteriorated by nonidealities of the physical antenna array and receiver hardware, e.g. unequal channel responses, mutual coupling of antennas and uneven physical dimensions of the antenna array. Such imperfections are difficult, if not impossible to capture in a model-driven estimator [4], [5].

For the above reasons, it is sensible to consider a data-driven estimator, based on machine learning. The idea to apply machine learning algorithms in near-field localization, or the closely related DOA estimation in the far-field is not new, but it became computationally tractable in the 1990s [6]. In addition to deep neural network (DNN) based solutions, successful estimators have also been constructed with support vector machines (SVM) [7]. Previous research on the subject has often concentrated on training and validating the systems using simulated data, generated with the standard near-field model, with simulated impairments. Furthermore, some of the prior research on DNN direction of arrival estimation has also utilized classification models, resulting in reduced resolution, because the output is quantized into predefined regions [4]. In this work, we use a DNN based regression model to estimate the source location. Instead of simulations, the model is trained and evaluated on a set of physical measurements. An inexpensive custom built coherent multichannel SDR, called *Coherent-RTLSDR* was used as the array receiver, accompanied by a handcrafted uniform linear antenna array (ULA). Obviously, the nonideatilies of such system are much more pronounced than they would be with costly professional multichannel SDR hardware and proper antennas. Despite of this, the DNN based estimator predicts the source location with low statistical error.

An M element array can simultaneously resolve $M-1$ separate signal sources, unless sparse array processing methods are used to virtually augment the array [8]. While DNN based localization estimators can naturally be developed and trained to identify multiple sources [4], this study considers only one source signal. The main contribution of this paper is in verifying experimentally, that a machine learning based estimator can substantially improve the accuracy of near-field localization in the presence of hardware and modeling imperfections. For the scope of this paper, the developed DNN model is called near-field location predictor network, NFLOPnet.

II. Model-Driven method

This section introduces the standard near-field signal model. In addition, one super-resolution method, the well-known MUSIC [1] is also briefly described. For localization, the signal-model is considered valid in the Fresnel region of the near-field, starting from the end of the reactive near-field at

distance $0.62\sqrt{D^3/\lambda}$ and ending at $2D^2/\lambda$, where λ is the wavelength and D the aperture of the antenna array. [3], [4]

A. Signal Model

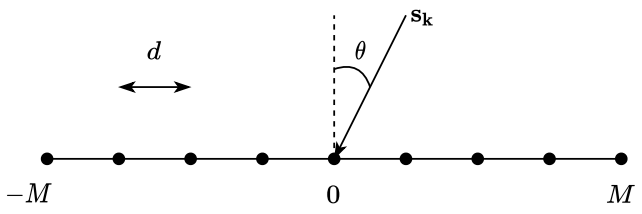

Fig. 1: Array model

Consider an M element uniform linear array (ULA) with an inter-element distance d, as shown in Fig. 1. Assuming narrowband signals, the standard model for the received signal at m-th antenna in near-field is given by [3]

$$x_m(t) = \sum_{k=1}^{K} s_k(t) e^{j(\alpha_k m + \beta_k m^2)} + w_m(t) \tag{1}$$

where $\alpha_k = -2\pi d \sin\theta_k/\lambda$, $\beta_k = \pi d^2 \cos^2\theta_k/(\lambda r_k)$, $m = 0, 1, \ldots, M-1$ is the element index and $w_m(t)$ denotes the i.i.d Gaussian noise at time instant t. The near-field signal sources s_k are located at azimuth angle θ_k and range r_k, $k = 1 \ldots K$. This is expressed more concisely in matrix form by

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{w} \tag{2}$$

where $\mathbf{A}$ is the steering matrix whose kth row is formed $[e^{j(-M\alpha_k + M^2\beta_k)}, e^{j((1-M)\alpha_k + (1-M)^2\beta_k)}, \ldots, e^{j(M\alpha_k + M^2\beta_k)}]$. The time index t has been omitted and $\mathbf{w}$ represents a i.i.d. Gaussian noise vector. As noted by Friedlander [3], model (1) is not entirely accurate from the perspective of electromagnetic field theory. For instance, the standard signal model borrows the range dependency of field strength from far-field equations of electromagnetic field theory. Also, in its most common form as it is written in (1), the model relies solely on phase information and disregards the amplitude differences between array elements. [3]

Theoretical performance limits of the standard near-field signal model have been analyzed, for instance in [9]. Cramér-Rao bound (CRB) is a regularly used tool to evaluate the theoretical minimum variance any possible unbiased estimator can obtain. The analysis on CRB for near-field localization derived from signal model (1) indicates that the minimum variance estimator for the direction $\hat{\theta}$ does not depend on the true distance r_k. However, the minimum variance of the range estimate $\hat{r}$ increases with distance. In other words, the closer the source the more accurate the estimate can be [9].

B. MUSIC

Multiple signal classification (MUSIC) was the first of the computationally feasible subspace methods, which exploit a change of basis via the Eigendecomposition to estimate signal spectrum. The key idea is to project the covariance matrix $\mathbf{R}$ into signal $\mathbf{Q_s}$ and noise subspaces $\mathbf{Q_n}$, assuming knowledge of the number of incident signals, K. To elaborate, the matrix of eigenvectors $\mathbf{Q}$ and eigenvalues Λ are viewed as $\mathbf{R} = \mathbf{Q}\Lambda\mathbf{Q}^\dagger = \mathbf{Q_s}\Lambda_\mathrm{s}\mathbf{Q_s}^\dagger + \mathbf{Q_n}\Lambda_\mathrm{n}\mathbf{Q_n}^\dagger$. The estimated signal power is then expressed by [1]

$$P_{MUSIC}(\mathbf{A}) = \frac{1}{||\mathbf{Q_n}^\dagger \mathbf{A}||^2} \tag{3}$$

where $\mathbf{A}$ is the steering matrix, orthogonal to noise eigenvectors. Evaluated in a naive way, the power spectrum over a range of directions is obtained with a grid search over the parameter space of $\mathbf{A}$. Thus, in a near-field localization problem, a two-dimensional spectral search is needed to locate the peaks of the spectrum. However, various schemes mitigating the computational complexity have been proposed, for instance in [10].

III. Proposed Data-Driven Method

When designing a machine learning model, one of the first questions is what are the input variables, or features in machine learning jargon. Signal covariance matrix is an obvious candidate, because it preserves all the relevant information for DOA estimation. Furthermore, many practical DOA estimators are defined, or can be defined via the signal covariance matrix. For this reason, the input for the DNN estimator was chosen to be derived from the covariance matrix estimate $\hat{\mathbf{R}}$. The estimate is averaged over several snapshots for two reasons: improved signal-to-noise ratio (SNR) and data reduction. In practice, $\hat{\mathbf{R}}$ is estimated by averaging over N snapshots from received zero mean $M \times N$ signal matrix $\mathbf{X}_N$ as $\hat{\mathbf{R}} = \frac{1}{N}\mathbf{X}_N\mathbf{X}_N^\dagger$, where $\dagger$ denotes the Hermitian transpose [10].

Standard DNN frameworks are currently designed to only process real valued data, therefore the complex valued input needs to be converted to a real valued representation. Noting the redundancy resulting from the fact that any covariance matrix is Hermitian, only above or below diagonal entries are really relevant for machine learning algorithms to find the nonlinear relation between input and output. Thus, the data reduction can be viewed as a map $\mathcal{M} : \mathbb{C}^{M\times M} \to \mathbb{R}^{M(M-1)}$. In practice, vectorized covariance input $\mathbf{r}_v$ is formed from the estimated covariance $\hat{\mathbf{R}}$ by concatenating the vectorized real and imaginary entries above diagonal, expressed by [5]

$$\begin{aligned} \mathbf{r} &= [R_{1,2}, R_{1,3}, \ldots, R_{1,M}, R_{2,3}, \ldots, R_{2,M}, \ldots, R_{M-1,M}] \\ \mathbf{r}_v &= [\Re\{\mathbf{r}\}, \Im\{\mathbf{r}\}]^T \end{aligned} \tag{4}$$

where $R_{j,k}$ are the entries of $\hat{\mathbf{R}}$, $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real and imaginary parts. The diagonal entries are omitted, since they only convey information of received signal power and noise variance, which is unknown. It is worth noting, that the machine learning algorithm used here is indifferent to the order in which $\mathcal{M}(\mathbf{R})$ maps the entries in $\mathbf{r}_v$, i.e. it is not a spatial representation. The topology of the proposed DNN is outlined in Fig. 2. The NFLOPnet is designed to consist only of fully connected layers. The input for such layer is one dimensional, excluding the observation (batch) axis. Essentially, a fully connected layer is simply a product

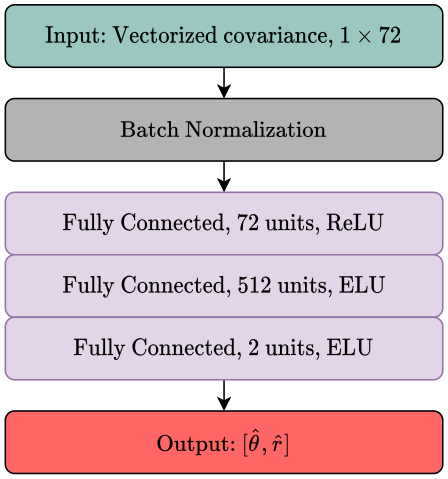

Fig. 2: DNN topology

$f(\mathbf{W}\mathbf{x} + \mathbf{b})$, where $f(\cdot)$ is the nonlinear activation function, $\mathbf{W}$ is a weight matrix and $\mathbf{b}$ is a bias vector.

Some studies have proposed utilizing 2D convolution in the first layers [4], [11]. In [4], $\hat{\mathbf{R}}$ is modified such that the real and imaginary parts are on the opposing sides of the diagonal, omitting the conjugate entries, and thus keeping the dimensions $M \times M$ [4]. The described approach is inspired by the success of DNNs in image recognition problems, and furthermore, there is the added benefit of proven and tested standard DNN topologies in the field of image processing. A convolutional neural network (CNN) topology could perform better in multiple input signal scenarios, even when the number of sources is unknown and must be estimated from data [11]. However, the CNN approach was rejected here, because it would increase complexity and the number of trainable parameters, requiring more measurements for training the network.

The NFLOPnet design uses the popular rectified linear unit (ReLU) as an activation function for the input layer. ReLU has gained popularity, because it avoids the vanishing gradient problem and does not saturate [12]. It is also computationally inexpensive to evaluate. Moreover, it is linear for the positive values and because of this retains qualities that make linear optimization straightforward, facilitating faster learning. The disadvantages of the popular activation function are the *dying ReLU problem* and issues with shifting bias. [12], [13, pp. 174–175]

The actual implementation of the DNN localization estimator was done in Keras, with a TensorFlow backend. Mean squared error (MSE) was chosen as the cost function and the model was trained for 1024 epochs, with the Adam (adaptive moments) optimizer. After the training phase, we have the weight matrices $\mathbf{W}_1 \in \mathbb{R}^{72\times72}$, $\mathbf{W}_2 \in \mathbb{R}^{72\times512}$ and $\mathbf{W}_3 \in \mathbb{R}^{512\times2}$, which correspond to the fully connected layers in Fig. 2, in a top-down order. The estimated location is computed from these by

$$[\hat{\theta}, \hat{r}] = f_2(\mathbf{W}_3 f_2[\mathbf{W}_2 f_1(\mathbf{W}_1 \mathbf{r}_\mathbf{v} + \mathbf{b}_1) + \mathbf{b}_2] + \mathbf{b}_3) \quad (5)$$

where $\mathbf{b}_1 \in \mathbb{R}^{1\times72}$, $\mathbf{b}_2 \in \mathbb{R}^{1\times512}$ and $\mathbf{b}_3 \in \mathbb{R}^{1\times2}$ are the bias vectors of the layers. The element-wise nonlinear activation functions used are defined as ReLU: $f_1(z) = \max(0, z)$ and ELU:

$$f_2(z) = \begin{cases} z, & \text{if } z > 0 \\ \alpha \exp(1) - 1, & \text{if } z \leq 0 \end{cases} \quad (6)$$

where α is a fixed scaling parameter. Unlike ReLU, ELU allows the propagation of negative values and was chosen for the two bottom layers as this was noted to decrease loss and improve accuracy in the training phase. This improvement over ReLU is likely caused by avoiding the *dying ReLU problem*.

IV. The Experiment

This section describes the experimental hardware utilized and the measurement setup.

A. Coherent receiver

The multichannel SDR receiver used in the experiments was the low-cost *Coherent-RTLSDR*[1], introduced in [14], [15]. It is a modular and expandable multichannel SDR, suitable for low budget research. The receiver system is built from the inexpensive and easily available RTL-SDR, which was originally designed as an USB DVB-T tuner dongle. Phase-coherent operation of the system is based on common clock and reference noise signals, with the help of which the receiver software is able to attain time and phase synchronization. The discrete receivers are assembled in coupler modules, each containing 7 RTL-SDR and a standard USB hub, as shown in Fig. 3. Switchable reference noise is distributed via directional coupling strips on the coupler module circuit board to the antenna input of each signal receiver. Because the receiver is assembled from self-contained receivers, each channel synthesizes the tuner down-mixing oscillator separately. Differences in the phase-locked loop frequency synthesis can cause the relative phases of the channels to drift slowly. Therefore, during the measurements, the receiver phase was re-calibrated against the reference noise every 10 minutes.

The maximum sampling rate of the system is 2.56 MHz, but the higher sample rates cause considerably more strain on both the USB bus and CPU. In the performed measurements, sampling rate was limited to 1 MHz, because the receiver software was running on resource constrained credit card sized computer, Rock Pi 4, which is a derivative of the well-known Raspberry Pi.

B. Measurement setup

The measurements were conducted outdoors on the 23 centimeter amateur radio band (1240 MHz), with a valid license. Transmitter positions were set on a grid of 11×13 locations, i.e. 11 azimuth angles θ and 13 distances r, resulting in the total of 143 collected signal matrices. The measurement grid was confined between the angles $[-30, 30]$ degrees, measured against the broadside of the array, and ranges $[1, 5.5]$ meters

[1]hardware description and software source code available at: https://github.com/mlaaks/coherent-rtlsdr

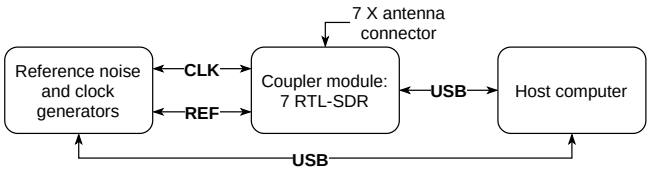

Fig. 3: Simplified block diagram of the coherent receiver system. [15]

from the array center element. Altogether 81920 samples were recorded for each location, thus the size of the raw dataset becomes approximately 11.7 million snapshots.

The receiving antenna array was a primitive hand-built 9-element uniform linear array (ULA), consisting of monopole wire elements, with inter-element separation of 125 mm, which is equivalent to $\lambda/2$ at 1.2 GHz. The Fresnel zone of the 9-element array spans the range from 1.25 m to 7.75 m. A portable signal source was transmitting a random QPSK modulated sequence at a symbol rate of 125 ksym/s. Transmitter gain was kept constant throughout the measurements. The received SNR per channel varied between 10 dB to 25 dB. depending on the distance between the source and the receiver.

V. Results and discussion

In this section the accuracy of NFLOPnet is evaluated and compared to the performance of MUSIC (3) with identical input. Number of snapshots used for estimating $\hat{\mathbf{R}}$ was decided to be $N = 256$. The dataset was randomly split into training and test sets, containing 90% and 10% of the observations, respectively. In absolute terms, this resulted in 4576 covariance estimates for the test set. A separate validation set was not generated for DNN hyperparameter tuning.

Two-dimensional MUSIC and DNN predictions were evaluated on the same test set. Figure 4 shows examples of the computed MUSIC spectrum overlaid with the true and DNN predicted locations. The ambiguity of the distance parameter generally increases when r grows, which is pronounced on the right MUSIC spectrums of Fig. 4 as a taller tail on the peak. Direction of the signal source observed from MUSIC generally agreed much more closely with the true location throughout the measured range.

The root-mean-square error (RMSE) and mean absolute error (MAE) were computed against the ground truth locations, and are shown in Table I. For MUSIC, the predicted location is taken as the maximum of the computed 2D spectrum, and is evaluated by exhaustive grid search. As noted, due to the blurred range in the spectrum, MUSIC predicts the distance rather poorly in this setup. Therefore, to obtain a somewhat reasonable comparison over the distance parameter, the same metrics were also computed for a limited range, ignoring predictions on test set cases where the distance is greater than 8λ, or approximately 2 meters. As mentioned in section II, minimum variance of r for an estimator based on model (1) increases with the true range.

TABLE I: Estimation error comparison of NFLOPnet DNN against MUSIC, root-mean-square error (RMSE) and mean absolute error (MAE).

	$\tilde{\theta}_{MUSIC}$ [deg]	$\tilde{\theta}_{DNN}$ [deg]	$\tilde{r}_{MUSIC}$ [λ]	$\tilde{r}_{DNN}$ [λ]
RMSE	5.2	1.19	6.0	0.13
MAE	4.12	0.98	4.3	0.10
RMSE $< 8\lambda$	5.1	0.99	1.4	0.16
MAE $< 8\lambda$	4.1	0.75	0.92	0.11

DNN based localization significantly outperforms the model based estimator accuracy in both direction and range. As to which one dominates, signal model or hardware deficiencies, is not analyzed in the scope of this paper. However, the proposition is that the poor performance of MUSIC is mostly explained by the quality of the measurement hardware. The monopole groundplane antenna array is not exactly equispaced, neither are the responses of array elements accurately matched. Adding to this, the polarisations of the elements made of wire were not strictly parallel to each other. Custom receiver built from inexpensive hardware adds another layer of uncertainty, even though care has been taken that the matching of phase is as accurate as possible.

The deep neural network NFLOPnet framework consists only of fully connected layers and is rather shallow, to keep the ratio of trainable parameters to available training data feasible. During development, there also arose a tradeoff between the volume of training data and the number of snapshots for covariance estimates: increasing the snapshots improves the perceived SNR, but simultaneously decreases the number of observations available. In simulations when the dataset is computer generated and thus practically unlimited, this would not be in issue. Also it was noted, that a DNN of modest proportions is able to learn the mapping from simulated data fairly accurately, even when impaired by noise. However, with the measured data, increasing the model size easily leads to overfitting due to the limited dataset size versus model complexity.

Finally, NFLOPnet is trained and tested in a scenario where the number of signal sources is one. Thus, its behaviour is uncertain if there were multiple sources present. As a workable method to extend such a model for multiple sources, a transfer learning scheme could be considered. Such a method would first train the estimator on simulated data and adapt the DNN weights to a small set of measured signals.

VI. Conclusion

This study developed and validated a machine learning based near-field localization estimator on real world measured radio frequency signals. The estimator outperforms MUSIC in the test scenario, where the signals are corrupted by receiving array nonidealities. Results suggest, that a DNN based estimator can be accurately adapted for use with highly nonideal

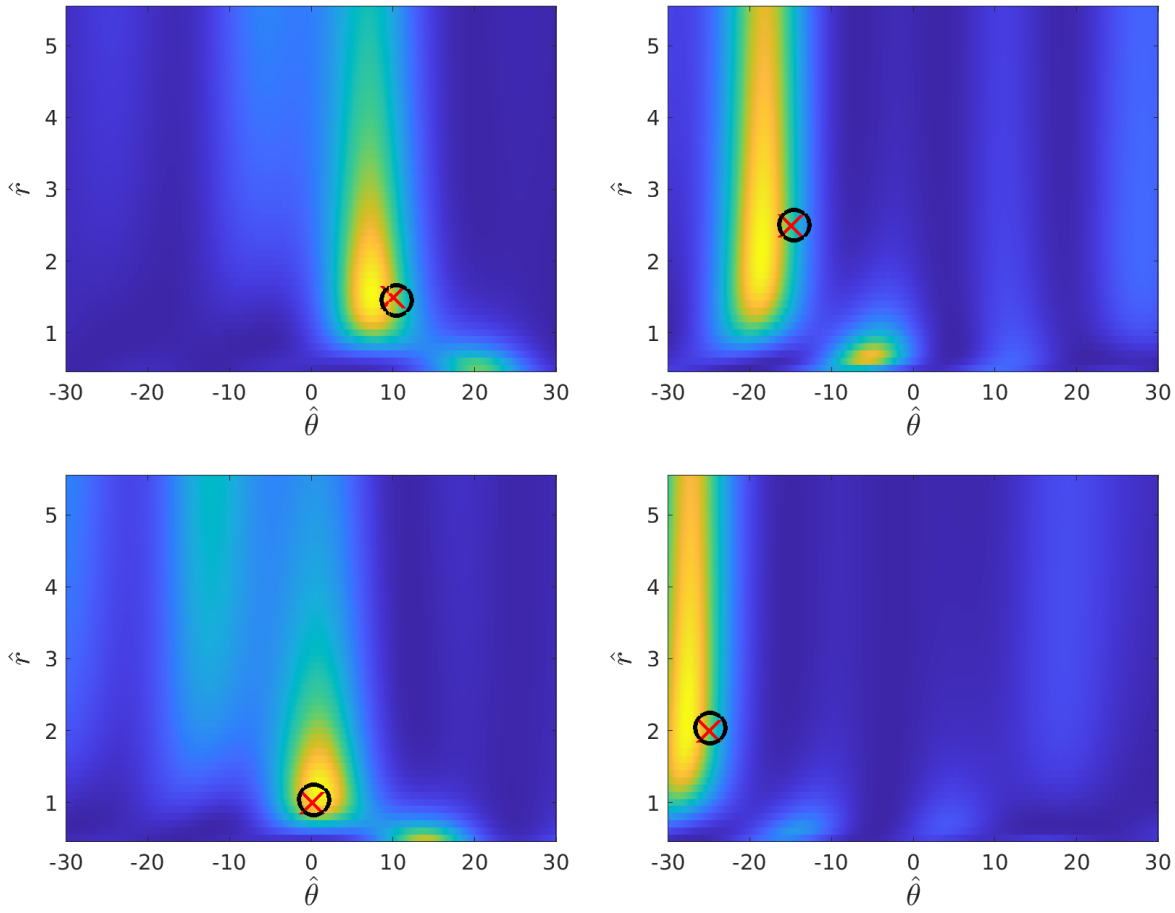

Fig. 4: Examples of MUSIC pseudospectra. The units are [m] for range r on the vertical axis and degrees for θ on the horizontal axis. Red crosses mark the true location and black circles the NFLOPnet estimated location.

hardware. For the sake of open science, we have made the dataset and the Python code of the algorithm available online. These can be found at https://github.com/mlaaks/NFLOPnet.

REFERENCES

[1] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation,* vol.34, no.3, pp.276-280, Mar. 1986.

[2] R. Roy, A. Paulraj and T. Kailath, "ESPRIT–A subspace rotation approach to estimation of parameters of cisoids in noise," in *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 34, no. 5, pp. 1340-1342, Oct. 1986.

[3] B. Friedlander, "Localization of Signals in the Near-Field of an Antenna Array," in *IEEE Transactions on Signal Processing,* vol. 67, no. 15, pp. 3885-3893, 1 Aug.1, 2019.

[4] W. Liu, J. Xin, W. Zuo, J. Li, N. Zheng and A. Sano, "Deep Learning Based Localization of Near-Field Sources with Exact Spherical Wavefront Model," *2019 27th European Signal Processing Conference (EUSIPCO),* A Coruna, Spain, 2019, pp. 1-5.

[5] Z. Liu, C. Zhang and P. S. Yu, "Direction-of-Arrival Estimation Based on Deep Neural Networks With Robustness to Array Imperfections," in *IEEE Transactions on Antennas and Propagation,* vol. 66, no. 12, pp. 7315-7327, Dec. 2018.

[6] A. H. El Zooghby, C. G. Christodoulou and M. Georgiopoulos, "Performance of radial-basis function networks for direction of arrival estimation with antenna arrays," in *IEEE Transactions on Antennas and Propagation,*, vol. 45, no. 11, pp. 1611-1617, Nov. 1997.

[7] A. Randazzo, M. A. Abou-Khousa, M. Pastorino and R. Zoughi, "Direction of Arrival Estimation Based on Support Vector Regression: Experimental Validation and Comparison With MUSIC," in *IEEE Antennas and Wireless Propagation Letters,* vol. 6, pp. 379-382, 2007.

[8] R. Rajamäki and V. Koivunen, "Comparison of Sparse Sensor Array Configurations with Constrained Aperture for Passive Sensing," *2017 IEEE Radar Conference (RadarConf),* Seattle, WA, 2017, pp. 0797-0802.

[9] M. N. E. Korso, R. Boyer, A. Renaux and S. Marcos, "Conditional and Unconditional Cramér–Rao Bounds for Near-Field Source Localization," in *IEEE Transactions on Signal Processing,* vol. 58, no. 5, pp. 2901-2907, May 2010.

[10] X. Zhang, W. Chen, W. Zheng, Z. Xia and Y. Wang, "Localization of Near-Field Sources: A Reduced-Dimension MUSIC Algorithm," in *IEEE Communications Letters,* vol. 22, no. 7, pp. 1422-1425, July 2018.

[11] S. Adavanne, A. Politis and T. Virtanen, "Direction of Arrival Estimation for Multiple Sound Sources Using Convolutional Recurrent Neural Network," *2018 26th European Signal Processing Conference (EUSIPCO),* Rome, 2018, pp. 1462-1466.

[12] D. Kim, J. Kim and J. Kim, "Elastic exponential linear units for convolutional neural networks", *Neurocomputing,* vol. 406, pp. 253-266, 2020.

[13] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning.* Cambridge, MA, USA: MIT Press, 2016.

[14] M. Laakso, "Multichannel coherent receiver on the RTL-SDR," M.Sc. Thesis, Sch. of Electrical Engineering, Aalto Univ., Espoo, 2019. Available: https://aaltodoc.aalto.fi/handle/123456789/37163

[15] M. Laakso, R. Rajamäki, R. Wichman, V. Koivunen, "Phase-coherent multichannel SDR - Sparse array beamforming", accepted for *2020 28th European Signal Processing Conference (EUSIPCO)*