The Journal of the Acoustical Society of America

JULY 01 2024

A deep learning method for reflective boundary estimation

Toros Arikan ; Amir Weiss ; Hari Vishnu ; Grant B. Deane ; Andrew C. Singer ; Gregory W. Wornell

J. Acoust. Soc. Am. 156, 65–80 (2024)
https://doi.org/10.1121/10.0026437

JASA ARTICLE

A deep learning method for reflective boundary estimation

Toros Arikan,[1,a] Amir Weiss,[2,b] Hari Vishnu,[3,c] Grant B. Deane,[4,d] Andrew C. Singer,[5,e] and Gregory W. Wornell[2,f]

[1] *Department of Electrical and Computer Engineering, Rice University, Houston, Texas 77005, USA*
[2] *Faculty of Engineering, Bar-Ilan University, Ramat-Gan, 5290002, Israel*
[3] *Acoustic Research Laboratory, National University of Singapore, Singapore 119227, Singapore*
[4] *Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92037, USA*
[5] *Department of Electrical and Computer Engineering, State University of New York at Stony Brook, Stony Brook, New York 11794, USA*

ABSTRACT:
Environment estimation is a challenging task in reverberant settings such as the underwater and indoor acoustic domains. The locations of reflective boundaries, for example, can be estimated using acoustic echoes and leveraged for subsequent, more accurate localization and mapping. Current boundary estimation methods are constrained to high signal-to-noise ratios or are customized to specific environments. Existing methods also often require a correct assignment of echoes to boundaries, which is difficult if spurious echoes are detected. To evade these limitations, a convolutional neural network (NN) method is developed for robust two-dimensional boundary estimation, given known emitter and receiver locations. A Hough transform-inspired algorithm is leveraged to transform echo times of arrival into images, which are amenable to multi-resolution regression by NNs. The same architecture is trained on transform images of different resolutions to obtain diverse NNs, deployed sequentially for increasingly refined boundary estimation. A correct echo labeling solution is not required, and the method is robust to reverberation. The proposed method is tested in simulation and for real data from a water tank, where it outperforms state-of-the-art alternatives. These results are encouraging for the future development of data-driven three-dimensional environment estimation with high practical value in underwater acoustic detection and tracking.

© *2024 Acoustical Society of America*. https://doi.org/10.1121/10.0026437

(Received 22 December 2023; revised 7 May 2024; accepted 28 May 2024; published online 1 July 2024)

[Editor: Oleg A. Godin] Pages: 65–80

I. INTRODUCTION

Environment learning in reverberant settings is an important task in difficult domains such as the underwater acoustic (Niu *et al.*, 2017a; Niu *et al.*, 2019) and indoor acoustic channels (Wu *et al.*, 2021). Depending on the application, environment learning may be the goal and this is the case in indoor room shape estimation (Lee *et al.*, 2019) and ocean remote sensing (Ali *et al.*, 2019) or may be an intermediate step in a processing chain for enhanced accuracy in localization. For example, to passively localize an unknown emitter with a collection of receivers, line of sight (LOS) arrivals to the receivers are used for time difference of arrival (TDOA; Korhonen, 2008) or time of arrival (TOA)-based localization (Ribeiro *et al.*, 2010; Brutti *et al.*, 2010). However, if we also leverage the non-line of sight (NLOS) arrivals from an accurately learned environment, higher accuracy is attainable (Naseri and Koivunen, 2016). Thus, we envision assuming an estimated emitter position (and known receiver positions) as a known ground truth and using the NLOS arrivals to accurately estimate the reflective boundaries in the environment, after which further joint localization and environment learning can be performed (Arikan *et al.*, 2023a).

Within the general scope of environment learning, we focus on reflective boundary estimation for shallow-water underwater acoustic settings,[1] as in Fig. 1. Although there may typically be some prior knowledge on the rough position of boundaries, such as the sea surface and seafloor (and, therefore, of the number of boundaries as well), we require accurate knowledge of their positions to make use of the corresponding NLOS arrivals. Over short ranges, we can approximate boundaries as (piecewise) planar and the speed of sound as constant and model them as producing mirror images of the emitter as "virtual emitters" as per Snell's Law (Deane, 1994). Euclidean distance matrices (EDM; Dokmanic *et al.*, 2015) or other methods (Naseri *et al.*, 2014) can then be used for boundary estimation through the localization of these virtual emitters. However, in the ocean, we often have low signal-to-noise ratios (SNRs; Dardari *et al.*, 2009) and the model mismatch that arises from a dynamic environment, which these methods do not address.

An alternative boundary estimation methodology relies on the two-dimensional (2D) NLOS arrival correspondence

[a] Email: ta55@rice.edu
[b] During the work on this paper, the author was at Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. Email: amir.weiss@biu.ac.il
[c] Email: tmshv@nus.edu.sg
[d] Email: gdeane@ucsd.edu
[e] Email: andrew.c.singer@stonybrook.edu
[f] Email: gww@mit.edu

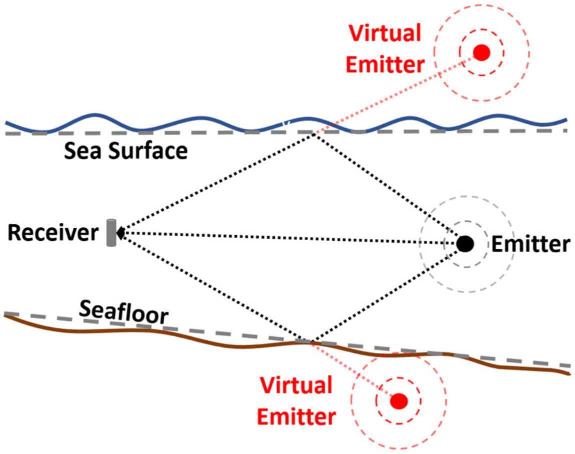


FIG. 1. (Color online) A general underwater acoustic setting, highlighting the typical NLOS arrivals and corresponding virtual emitters.

to a path distance of d_{NLOS}. This allows us to identify the locus of potential reflective boundaries encountered by the arrival as an ellipse whose foci are the emitter and receiver locations, denoted as $\mathbf{p}_{\text{e}}$ and $\mathbf{p}_{\text{r}}$, respectively, as in Fig. 2(a). By definition, points on the ellipse have a total distance of d_{NLOS} to the emitter and receiver, and the reflective boundary is a tangent line of this ellipse. With multiple receivers, multiple ellipses are defined by such NLOS arrivals, and the reflective boundary is their common tangent. Therefore, by fitting common tangents to ellipses, the boundaries can be estimated while avoiding a computationally challenging echo labeling problem in multipath environments, as illustrated in Fig. 2(b). Here, each boundary is a common tangent to a single arrival's ellipse from each receiver, highlighted by matching boundary and ellipse colors. Assigning ellipses to tangents is a problem of combinatorial complexity and error-prone for inaccurate time-delay estimates (Crocco *et al.*, 2017). Moreover, missing or spurious arrivals in the received signals complicate the echo labeling, thus, motivating a solution that bypasses this task altogether.

In light of these challenges, we propose a convolutional neural network (CNN)-based regression method for boundary estimation through supervised learning. Our data-driven method operates by parametrizing tangents to ellipses by the range (ρ) and azimuth (θ) values of their normal vectors (Naseri and Koivunen, 2016), calling this (ρ, θ) space the common tangents to spheroids (COTANS) domain. This COTANS transform maps the environment geometry and time-delay estimates to images in the (ρ, θ) space, transforming the data into an input representation that is easier for operation by CNNs. The proposed COTANS neural network (NN) method, termed Neuro-COTANS, incorporates a modified AlexNet (Krizhevsky *et al.*, 2012) architecture. It is trained on a simulated dataset to estimate the locations of reflective boundaries from unlabeled NLOS arrivals over a wide range of SNRs: Neuro-COTANS is trained only once for multiple SNRs such that the techniques become robust at any SNR. The resulting network can be used with simulated and recorded data. A key influence for our work was the successful recent use of NNs for emitter localization and environment learning, including the underwater acoustic setting (Niu *et al.*, 2017b; Niu *et al.*, 2019) and reverberant indoor environments (Wu *et al.*, 2021). Although we target the short-range shallow-water underwater acoustic setting as opposed to a general open ocean setting, we propose a general-purpose boundary estimation method for any setting where straight-ray propagation holds, which outperforms its alternatives in simulation and real-data experiments.

Our main contributions are the following:

- A robust NN method for boundary estimation that is superior to state-of-the-art alternatives and straightforward to retrain for different environments;
- a study of the performance and stability of alternative boundary estimation methods; and
- a Cramér-Rao lower bound (CRLB) for boundary range estimation, filling this gap in the literature.

In the literature, the problems of time-delay estimation and localization are often treated separately with different error/noise models. In this paper, we provide a unified framework that establishes a common setting for these tasks, allowing for the study of Neuro-COTANS and its alternatives under more realistic conditions. A short version of this work, incorporating some of our earlier simulation results, was presented in Arikan *et al.* (2023b).

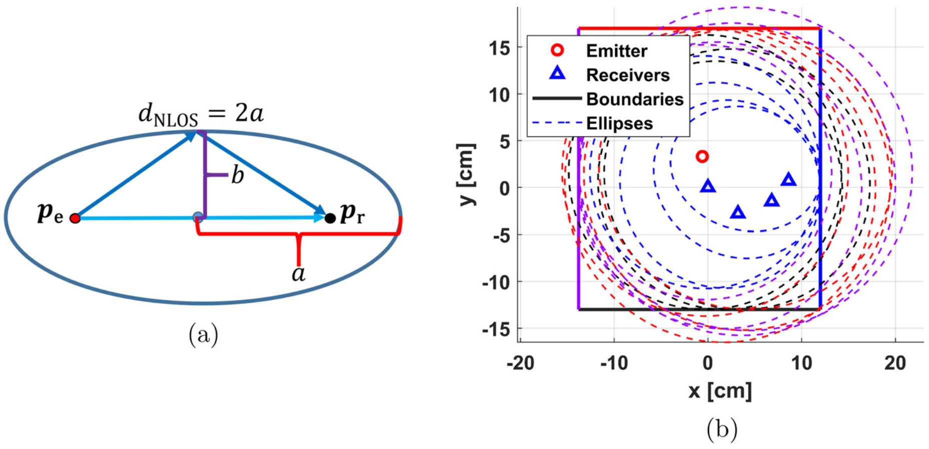


FIG. 2. (Color online) NLOS arrivals define ellipses with an emitter and receiver as their foci (a), and in a rich multipath setting, solving the echo labeling problem is difficult (b). Each boundary in (b) is a common tangent to a single ellipse due to each receiver, highlighted by matching colors.

A. Prior work

Fitting tangent planes to spheroids for boundary estimation is becoming an increasingly widespread approach for indoor settings. This methodology was first proposed in Antonacci *et al.* (2010) and Antonacci *et al.* (2012). A similar approach (Naseri and Koivunen, 2016) was used for joint localization and boundary estimation, which employed a method inspired by the Hough transform rather than leveraging an analytical cost function as in Antonacci *et al.* (2010) (which assumed only small-scale errors). In Park and Choi (2021), a Hough transform-inspired method was used to estimate an indoor environment and perform echo labeling with provisions to reject incorrectly chosen second-order echoes. Although these techniques refer to the Hough domain, they do not fit planes to point clouds as the Hough transform does. Instead, they sample points from a spheroid's surface (typically randomly), and then deduce the tangent at each point. We have previously derived a closed-form mathematical method to perform this transformation without such sampling of point clouds (Arikan *et al.*, 2023b) and use this more rigorous and reliable method in the current paper. To avoid conflating our plane-fitting method with the Hough transform, we refer to a COTANS transform and COTANS domain instead.

Plane-fitting boundary estimation methods typically apply a smoothing filter to COTANS images, followed by the extraction of maxima (Naseri and Koivunen, 2016) to estimate the boundaries when TOA estimation errors are present. However, this handcrafted filtering operation is suboptimal and parameters, such as filter sizes and kernels, are manually tuned to specific settings. Our original motivation in pursuing a NN method was to automate and combine these filtering and peak extraction tasks for different reverberant settings, thus, implementing a multi-scale filtering approach. If a NN is trained with a wide range of geometries and realistic estimation errors, it can potentially learn the optimal inference rule, which can be viewed as joint (and implicit) filtering and peak extraction. The resulting NN can then be retrained for different environments. However, the resulting NN method which we devised actually performs a high-level inference over the entire COTANS transform image, rather than solve a local peak estimation task. Furthermore, Neuro-COTANS is not constrained by the pixel resolution of the COTANS images, unlike existing plane-fitting methods. Thus, Neuro-COTANS is a fundamentally different improvement over other tangent-fitting methods instead of being a simple NN extension of this overall approach.

In addition to the above considerations, handcrafted filters can cause implementation issues because boundary estimation tasks have different physical scales or domains (e.g., indoor or underwater acoustic). It is also challenging to make a fair comparison between these various methods with different filter sizes and kernels. Neuro-COTANS can be retrained in an automated manner *without* having to modify any implementation-specific hyperparameters.

The Neuro-COTANS method that we propose is envisioned as a key component of a larger underwater acoustic localization and tracking system. In a previous publication (Arikan *et al.*, 2023a), we proposed the passive end-to-end localization (PEEL) method, which only featured a set of known and fixed receiver positions that were used for surreptitious localization and tracking of an unknown mobile pulsed emitter. The PEEL method first established synchronization with the emitter and produced a reasonable estimate of its position using TDOA localization. In this TDOA method, the received signals were cross-correlated and these cross-correlation results were used to localize the emitter at the intersection of hyperbolae of equidistance. To produce an initial estimate of the NLOS boundary positions in the environment, we used the (non-NN) COTANS transform method that was existing in the literature.

Within the context of PEEL, Neuro-COTANS provides an improved method for initial environment estimation before we attempt to track a moving emitter that can disappear behind occluding objects. Our use case is passive localization and tracking of an emitter that is not necessarily controlled or emitting in a particular location, in contrast to an echosounder or fathometer. On the contrary, this uses the emitter as a source of opportunity for boundary estimation. Thus, the "known" emitter position is actually a TDOA estimate, obtained at a position where the LOS to the receivers has not been occluded. As the known emitter position can have errors, Neuro-COTANS has to be robust to model mismatch.

We develop Neuro-COTANS, as illustrated in Fig. 3, and structure the paper as follows. In Sec. II, we formulate the problem, and in Sec. III, we revisit fundamental results regarding time-delay estimation, which are necessary for understanding the motivation of our solution approach. The COTANS methodology and our NN method are presented in Sec. IV. Alternative methods are outlined in Sec. V, and the CRLB for boundary range estimation is derived in Sec. VI. Simulation and experimental results are presented in Sec. VII, and concluding remarks are given in Sec. VIII. Throughout the paper, lowercase bold variables are vectors, and uppercase bold variables are matrices.

II. PROBLEM FORMULATION

We now present the notation, signal model, and environment geometry that frame the boundary estimation problem. We model a static 2D environment with N planar boundaries (tangent lines), where N is assumed to be known. These boundaries are described by the range $\rho \in \mathbb{R}_+$ and azimuth $\theta \in [0, 2\pi)$ of their normal vector relative to the (arbitrarily-chosen) origin. Thus, the jth boundary is parametrized as the vector $\boldsymbol{\eta}_j = [\rho_j\ \theta_j]^{\mathrm{T}}$, for all boundaries $j \in \mathcal{S}_N$, where $\mathcal{S}_K \triangleq \{1, \ldots, K\}$ for some $K \in \mathbb{N}$, denoting here the number of boundaries. We assume a single isotropic emitter in the environment at a known location, $\mathbf{p}_{\mathrm{e}} = [x_{\mathrm{e}}\ y_{\mathrm{e}}]^{\mathrm{T}}$, and M isotropic receivers at known $\mathbf{p}_{\mathrm{r},i} = [x_i\ y_i]^{\mathrm{T}}, i \in \mathcal{S}_M$, with

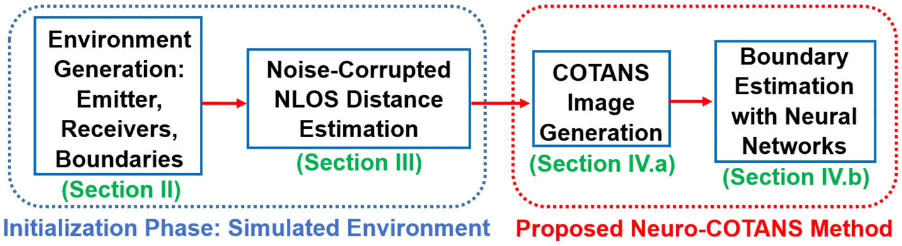

Environment Generation: Emitter, Receivers, Boundaries
(Section II)
Noise-Corrupted NLOS Distance Estimation
(Section III)
COTANS Image Generation
(Section IV.a)
Boundary Estimation with Neural Networks
(Section IV.b)
Initialization Phase: Simulated Environment
Proposed Neuro-COTANS Method

FIG. 3. (Color online) Summary of the Neuro-COTANS method's presentation.

isotropicity only needed for every boundary to produce a NLOS arrival at every receiver.[2]

The received signal at the ith receiver, $r_i(t) \in \mathbb{R}$, is modeled as the sum of the LOS and single-reflection NLOS arrivals, delayed by their respective TOAs. We assume that higher-order reflections are heavily attenuated in the underwater acoustic setting (Weiss *et al.*, 2022) as compared to first-order reflections from boundaries such as the sea surface and seafloor. In an isovelocity ocean environment, the LOS TOA, $\tau_{i,0}$, is given by

$$\tau_{i,0} = \frac{\|\mathbf{p}_{\mathrm{r},i} - \mathbf{p}_{\mathrm{e}}\|_2}{v_{\mathrm{s}}}, \quad \forall i \in \mathcal{S}_M, \tag{1}$$

where v_{s} is the speed of sound, which is approximated as a known constant.[3]

Reflections can be interpreted as producing virtual emitters, and for the jth boundary, we obtain the virtual emitter location, $\mathbf{p}_j$, by finding the corresponding reflection of $\mathbf{p}_e$ (see Fig. 1). The NLOS TOA at the ith receiver from the jth boundary, $\tau_{i,j}$, is equal to the TOA from the ith receiver ($\mathbf{p}_{\mathrm{r},i}$) to the corresponding jth virtual emitter ($\mathbf{p}_j$):

$$\tau_{i,j} = \frac{\|\mathbf{p}_{\mathrm{r},i} - \mathbf{p}_j\|_2}{v_{\mathrm{s}}} \triangleq \frac{d_{i,j}}{v_{\mathrm{s}}}, \quad \forall i \in \mathcal{S}_M, \quad \forall j \in \mathcal{S}_N. \tag{2}$$

We denote the known emitted waveform as $s(t)$, which will ultimately be used to match-filter the received signals. Merging the effects of attenuation and reflection into the equivalent attenuation coefficient, $\alpha_{i,j}$, for each path, then the received signal at the ith receiver is

$$r_i(t) = \sum_{j=0}^{N} \alpha_{i,j} s(t - \tau_{i,j}) + \xi_i(t), \tag{3}$$

where $j = 0$ corresponds to the LOS path, and $\xi_i(t)$ is a noise signal that is a realization of a spectrally flat Gaussian process. In practice, the environment can be reverberant and feature higher-order reflections and noise that may not be Gaussian (Chitre, 2007), but here we consider simpler scenarios for the analysis. Nevertheless, this does not limit the applicability of our method to signals with non-Gaussian noise. We work with a discrete-time sampled version of Eq. (3) as $r_i[n] \triangleq \{r_i(t)|_{t=nT_{\mathrm{s}}}\}_{n\in\mathbb{Z}}$, where T_{s} is the sampling period, and the sampling rate, f_{s}, is greater than twice the Nyquist rate for the signals considered.

The geometric information for boundary estimation consists of the known $\mathbf{p}_{\mathrm{e}}$ and $\{\mathbf{p}_{\mathrm{r},i}\}$ and unknown $\{\tau_{i,j}\}$. Hereafter, the NLOS TOAs are estimated using an (at least asymptotically) optimal estimator. For example, these estimates can be obtained by matched-filtering $r_i[n]$ with $s[n] \triangleq s(nT_{\mathrm{s}})$ and picking the TOAs corresponding to the N largest peaks (excluding the LOS) as $\{\hat{\tau}_{i,j}\}$. The distance estimates $\{\hat{d}_{i,j} \triangleq v_{\mathrm{s}}\hat{\tau}_{i,j}\}$ from Eq. (2) are then used to estimate the boundaries as $\{\hat{\boldsymbol{\eta}}_j\}_{j=1}^{N}$, given the environment model in Table I.

In many environment estimation methods, $\{\hat{d}_{i,j}\}$ are modeled as corrupted by Gaussian noise (Cheung *et al.*, 2004). However, it is the received signals of Eq. (3) that are instead subject to Gaussian noise, hence, we adopt a more realistic error model for $\{\hat{d}_{i,j}\}$, as follows.

III. FUNDAMENTALS OF NLOS TIME-DELAY ESTIMATION

Time-delay estimation has been extensively studied; in this section, we describe an estimation model for the time-delay estimates, $\hat{\tau}_{i,j}$. For a given value of the SNR as S, we obtain a corresponding error, $\epsilon_{i,j}(S)$, that is not necessarily Gaussian. This will serve our ultimate goal of generating a wide range of time-delay errors in our dataset, modeling operational conditions under high and low SNRs.

In the boundary estimation problem, the NLOS arrivals from each boundary will produce a received signal with multiple peaks, and we will be picking a given number of the highest peaks from the matched-filtered $\{r_i(t)\}$ to obtain the NLOS time-delays.[4] We assume that the multipath arrivals are typically sufficiently separated and the signal is of high enough bandwidth such that after matched-filtering, they do not affect each other's time-delay estimation performances through interference. If this condition is not expected to hold, more advanced TOA estimation algorithms, like SAGE, can be used (Demirli and Saniie, 2001), which are able to handle overlapping arrivals but are out of the scope of the present work.

TABLE I. Model of the boundary estimation problem.

Problem feature	Modeling assumptions
Speed of sound, v_{s}	Known and constant within the environment
Environment	Static, short-range shallow-water environment
Reflectors, $\boldsymbol{\eta}_j$	Planar, known number, and unknown positions
Emitter and receivers, $\mathbf{p}_{\mathrm{e}}$ and $\{\mathbf{p}_{\mathrm{r},i}\}$	Known positions and synchronized
Transmissions	Known pulse waveforms

Time-delay estimation performance is specific to a given emitted signal such as the following standard Gaussian pulse, which we employ throughout:

$$p(t) = e^{-2\pi t^2/\tau_p^2}, \tag{4}$$

where $\tau_p = 1/B$, and B is the 3 dB bandwidth (in Hz). This pulse has energy $E_p = \int p^2(t)\mathrm{d}t$. Whereas the infinite-length signal in Eq. (4) is truncated in practice to some finite length, τ_d, the pulse exponentially decays to negligible magnitudes, and the finite-length signal is functionally equivalent to its infinite-length formulation. We assume that we have real additive white Gaussian noise,[5] and the one-sided power spectral density of the noise $\xi_i(t)$ is equal to N_0. Thus, the SNR[6] is defined as E_p/N_0. Given a desired SNR, S, in dB, it follows that we obtain $E_p/N_0 = 10^{S/10}$. For a sampling rate, f_s, the average noise power, N_avg, is $N_\mathrm{avg} = N_0 f_\mathrm{s}$. Thus, for a desired S, the required variance, σ^2, of the sampled, discrete-time additive Gaussian noise is

$$\sigma^2 = \frac{f_\mathrm{s}E_p}{10^{S/10}}. \tag{5}$$

It is well-known that optimal time-delay estimation has a performance profile, which is characterized by a transition from a non-informative region at low SNRs, through a threshold phenomenon, to a "small-errors" regime at high SNRs (Weiss and Weinstein, 1983). At high SNR, the Gaussian noise added to $r(t)$ results in a Gaussian, small-scale perturbation of $\hat{\tau}$ and matched-filtering mean squared error (MSE) estimation performance that asymptotically coincides with the CRLB for time-delay estimation, which for a given S has variance $\sigma^2_{\mathrm{CRLB}}(S)$ (Dardari *et al.*, 2006). The resulting estimation error, $\epsilon_{i,j}$, is called a "local error," where $\epsilon_{i,j} \sim \mathcal{N}(0, \sigma^2_{\mathrm{CRLB}}(S))$. At lower SNRs, peaks of noise can have a greater magnitude than that of the true arrival. Picking one of these spurious peaks results in a "global error" that leads to a drastic performance reduction; because the noise peaks are distributed uniformly in the time interval on which the matched-filtering is performed, global errors cause $\hat{\tau}$ to be distributed uniformly on the observation time interval. Thus, examining $r(t)$ in the time interval, $[\delta, T_p + \delta]$, where δ is some time increment in seconds and T_p is the time length of our received signal observation window, we have $\hat{\tau}_{i,j} \sim \mathrm{U}(\delta, T_p + \delta)$, where $\mathrm{U}(a, b)$ denotes the uniform distribution of a random variable within $[a, b]$. As SNR is progressively reduced below a certain threshold, there is a transition to such global errors having a higher probability of occurrence, which increasingly dominates the MSE (Weinstein and Weiss, 1984).

In Fig. 4, we conduct a range estimation simulation with 10 000 realizations of the simulated noise added to $p(t)$ per SNR value considered and compare it to the CRLB for time-delay estimation. As discussed previously, below a SNR threshold, global errors eventually cause catastrophic estimation errors with higher probability. The results illustrate the need for a boundary estimation method that performs accurately when errors are small (local errors) and robustly when errors are large (global errors). Note that localization performance is fundamentally different than time-delay estimation: while time-delay estimation accuracy could be on the order of 0.1 m, for example, the localization accuracy from leveraging multiple receivers can be much more refined, as will be calculated in Sec. VI.

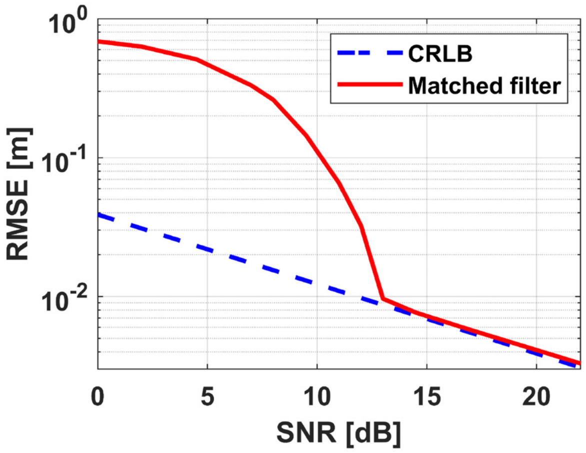


FIG. 4. (Color online) The CRLB on the range estimation root-mean squared error (RMSE) for a Gaussian pulse of 15.4 kHz bandwidth and the simulated empirical matched-filtering performance. The global error threshold for this particular signal is observed to be ~13.5 dB SNR, indicating that large errors can be encountered even at seemingly high signal strengths.

IV. NEURO-COTANS FOR BOUNDARY ESTIMATION

Having defined the boundary estimation problem in Sec. II, we present the Neuro-COTANS method. Suppose that a set of $\{\hat{\tau}_{i,j}\}$ has been estimated from $\{r_i(t)\}$ and is unlabeled with respect to the corresponding boundaries. The goal is to estimate $\{\boldsymbol{\eta}_j\}$ in a way that is robust to estimation errors in $\{\hat{\tau}_{i,j}\}$. We first discuss how the COTANS transform is used to generate images for a given geometry and set of $\{\hat{\tau}_{i,j}\}$ (Naseri and Koivunen, 2016). Then, we detail how Neuro-COTANS estimates the $\{\boldsymbol{\eta}_j\}$ from these COTANS images.

A. Generation of COTANS images

We summarize the generation of COTANS images for a given geometry and $\{\hat{\tau}_{i,j}\}$. In 2D, a boundary defined by ρ and θ can be conceptualized as a point (ρ, θ) in a COTANS transform domain; working out the (ρ, θ) expression of a line is performed by computing its COTANS transform (Borrmann *et al.*, 2011). The inputs to the COTANS transform are a receiver and emitter pair's locations in space and a specific NLOS TOA; the output of the COTANS transform is the set of all points, (ρ, θ), which represent the valid tangent planes to the ellipse defined by these inputs.

In practice, the COTANS transform generally does not have a closed-form solution, and we, therefore, apply the COTANS transform to individual tangent planes to yield a large collection of points, (ρ, θ). Here, we derive a mathematical solution for the COTANS transform which precludes the need for randomly sampling points (ρ, θ) on the surface of an ellipse[7] (Naseri and Koivunen, 2016). Thus, we avoid a heuristic of the number of sampling points to achieve a desired image resolution, allowing for simple COTANS image generation. We follow the steps outlined in Fig. 5 for ease of explanation of the COTANS transform.

Step 1: For a given θ, we use vector geometry to obtain the COTANS transform's ρ for an origin-centered ellipse as

$$\rho(\theta) = \sqrt{a^2 \cos^2\theta + b^2 \sin^2\theta}, \tag{6}$$

where

$$a = d_{\mathrm{NLOS}}/2, \quad b = \sqrt{d_{\mathrm{NLOS}}^2 - d_{\mathrm{LOS}}^2}/2 \tag{7}$$

are the standard ellipse axes calculated from the $\{\hat{\tau}_{i,j}\}$, as in Fig. 2(a).

Step 2: To move a collection of $\{(\rho, \theta)\}$ centered on the origin to $\mathbf{p}_{\mathrm{e}}$ and $\mathbf{p}_{\mathrm{r}}$, we rotate the points to match the true ellipse's orientation (adjusting θ to be some θ_{rot}) as in Fig. 5(b).

Consider the azimuth of the vector $\mathbf{p}_{\mathrm{e}} - \mathbf{p}_{\mathrm{r}}$, designating this angle θ_{rot}. To align our starting standard ellipse with the target ellipse, we replace each (ρ, θ) with $(\rho, (\theta + \theta_{\mathrm{rot}})\mathrm{mod}2\pi)$, which we term (ρ, θ''). The rotation operation leaves the ρ-value of each tangent line unchanged and only affects the azimuth, whereas the new foci are at $\mathbf{p}_{\mathrm{e}}''$ and $\mathbf{p}_{\mathrm{r}}''$.

Step 3: We now translate the points (yielding a final ρ_{COTANS} and θ_{COTANS}), as in Fig. 5(c). Here, care must be taken in how (ρ, θ) is modified in Fig. 5.

Building on step 2, we calculate a translation vector,

$$\mathbf{p}_{\mathrm{trans}} = \mathbf{p}_{\mathrm{r}} - \mathbf{p}_{\mathrm{r}}'', \tag{8}$$

which would be added to any point on the rotated standard ellipse to obtain the target ellipse. To obtain the resulting $(\rho_{\mathrm{COTANS}}, \theta_{\mathrm{COTANS}})$ pairs, we first calculate the dot product:

$$\rho_{\mathrm{proj}} = \mathbf{p}_{\mathrm{trans}} \cdot \hat{\boldsymbol{\rho}}, \tag{9}$$

where $\hat{\boldsymbol{\rho}} = [\cos\theta'', \sin\theta'']^{\mathrm{T}}$ is the unit vector pointing toward the tangent line. Thus, we project the translation vector, $\mathbf{p}_{\mathrm{trans}}$, onto $\hat{\boldsymbol{\rho}}$. If $\rho_{\mathrm{proj}} \geq 0$, then we merely advance the tangent line in the same direction without changing its azimuth, such that we replace (ρ, θ'') with $(\rho + \rho_{\mathrm{proj}}, \theta'')$. If $\rho_{\mathrm{proj}} < 0$, then we subtract the projection result from ρ, thus, replacing ρ with $|\rho - |\rho_{\mathrm{proj}}||$, which ensures that ρ is positive as per definition. If $|\rho_{\mathrm{proj}}| < \rho$ and $\rho_{\mathrm{proj}} < 0$, then we do not modify the azimuth θ''; else, because the line has been translated past the origin and the direction of the $\hat{\boldsymbol{\rho}}$-vector has been flipped, we replace θ'' with $(\theta'' + \pi)\mathrm{mod}\, 2\pi$.

Performing the set of operations in steps 1–3 for each of the original $\{(\rho, \theta)\}$ points representing tangent lines, we obtain a final transformed set of $\{(\rho_{\mathrm{COTANS}}, \theta_{\mathrm{COTANS}})\}$, rounded to a desired accuracy. This set of COTANS transform results can be conveniently illustrated as a COTANS image, where we define an array over ρ and θ with this resolution and for each rounded point, increment the corresponding array cell by one.

Generating and adding the separate COTANS images for every NLOS arrival and corresponding emitter and receiver pairs, we essentially discretize the space $\rho \times \theta$ as a matrix and increment this "accumulator" array over every candidate (ρ, θ) to yield a composite COTANS-domain image (e.g., as in Fig. 6). Here, the maxima are at the true boundaries $\{(\rho_j, \theta_j)\}$ in the absence of errors. Note that the mapping to the space $\rho \times \theta$ is not one-to-one, which is one of the fundamental reasons why multiple receivers are needed for localization (unless the search space is constrained) and why the intersecting arcs in an accumulator array are required.

When time-delay estimation errors are present (which is always the case in practice), the COTANS curves do not exactly intersect at the correct boundary locations, as seen in Fig. 7(a). This issue prevents us from simply picking the N largest local maxima of an image to estimate the boundaries as the curves do not intersect to yield such maxima. In the literature, a heuristic, handcrafted smoothing filter is typically used for local averaging of the image (Naseri and Koivunen, 2016), followed by selection of as many maxima as there are boundaries as in Fig. 7(b), where the neighborhood of each maximum is set to zero to avoid picking the same boundary multiple times.[8] This suboptimal methodology can increase the estimation errors because it distorts the original COTANS image, and it only uses the information in a small part of the image rather than exploiting other

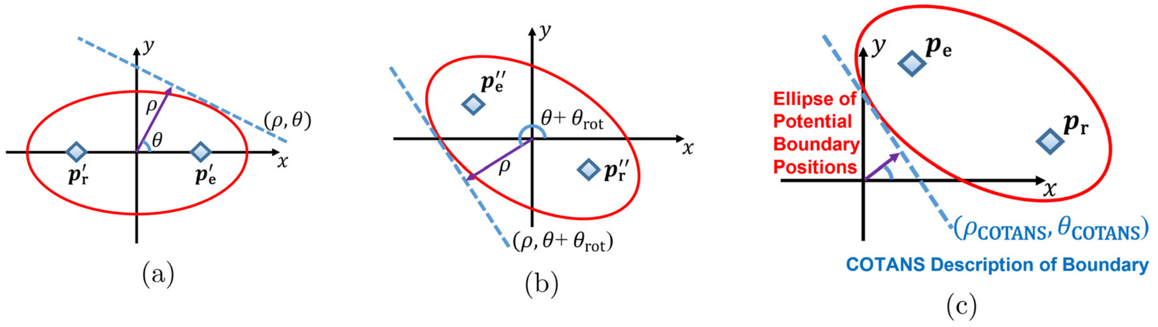
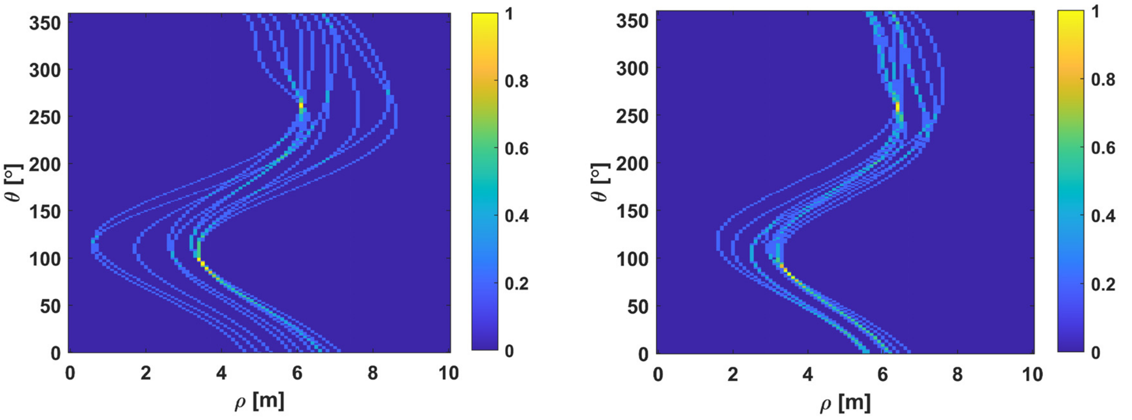

y
ρ
θ
(ρ, θ)
x
$\mathbf{p}_{\mathrm{r}}'$
$\mathbf{p}_{\mathrm{e}}'$
(a)
y
$\mathbf{p}_{\mathrm{e}}''$
$\theta + \theta_{\mathrm{rot}}$
x
ρ
$\mathbf{p}_{\mathrm{r}}''$
$(\rho, \theta + \theta_{\mathrm{rot}})$
(b)
y
$\mathbf{p}_{\mathrm{e}}$
Ellipse of Potential Boundary Positions
$\mathbf{p}_{\mathrm{r}}$
x
$(\rho_{\mathrm{COTANS}}, \theta_{\mathrm{COTANS}})$
COTANS Description of Boundary
(c)

FIG. 5. (Color online) Steps to obtain the COTANS transform of a tangent line. Description of one tangent to a standard ellipse (a), rotation of this origin-centered ellipse and its tangent (b), and the translation of this ellipse to its real position (c) are shown.



(a)

(b)

FIG. 6. (Color online) Examples of COTANS images for the transition region at 12 dB SNR (a) and high-SNR region at 20 dB SNR (b) for an environment with two boundaries at $(\rho, \theta) = (3.5{,}84)$ and (6.4,258). The images are colored for convenience; the original images are in gray scale.

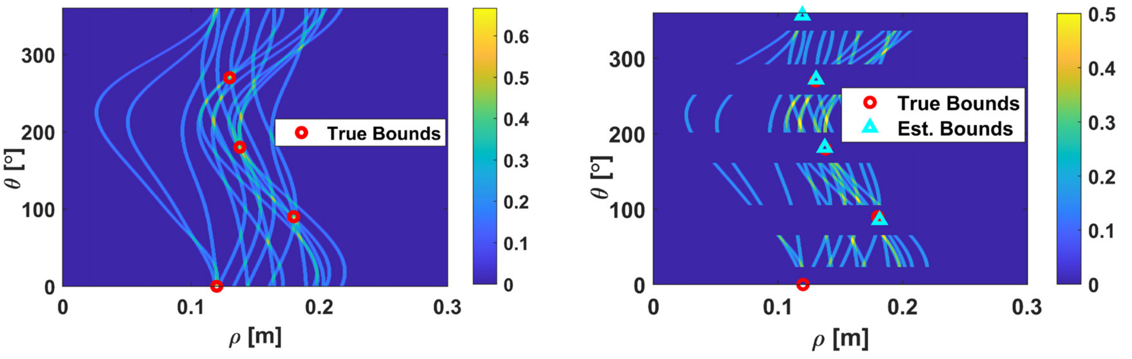
potential patterns in the full image. Furthermore, the smoothing filter's dimensions and kernel are heuristically tuned to specific environments and COTANS image resolutions, making it difficult to generalize. Therefore, rather than work with model-driven methods that may not be able to fully use the information in the image, we introduce a NN method for higher-level inference of boundary locations over entire COTANS images, which is not constrained by such limitations.

B. Neuro-COTANS method

Neuro-COTANS uses CNNs for multi-output regression from COTANS images. We repurpose the eight-layer and two-GPU AlexNet architecture (Krizhevsky *et al.*, 2012) by replacing the final classification layer with a regression layer, where MSE is used as the cost function. Here, we are guided by previous approaches that repurpose AlexNet for regression (Szegedy *et al.*, 2013). To work with color images, AlexNet has three channels; however, the COTANS images only have a single value for each pixel scaled to be within [0,1], therefore, we modify AlexNet to only have one channel. Our network inputs are COTANS images, and outputs are the boundary parameter estimates, $[\hat{\rho}_1 \cdots \hat{\rho}_N \, \hat{\theta}_1 \cdots \hat{\theta}_N]^{\mathrm{T}}$. We use the ground truth values, $[\rho_1 \cdots \rho_N \, \theta_1 \cdots \theta_N]^{\mathrm{T}}$, as the target for training the NN. Thus, our output layer size is $2N$, and the NN implements a function that projects COTANS images onto this $2N$-dimensional space.

We use AlexNet as a building block for Neuro-COTANS because it does not incorporate any specific image classification features as we have an image regression task instead. Thus, we can easily replace the final classification layer with a regression layer and modify the input image dimensions. Our training hyperparameters are given in Table II; note that ℓ_2 regularization (which compensates for image noise) is set to zero because the COTANS transform images (on which the NN operates) are not noise corrupted, although the underlying acoustic data may be noisy.

To use this architecture on COTANS images, we generate training image datasets by simulating scenarios with randomized $\mathbf{p}_{\mathrm{e}}$ and $\{\mathbf{p}_{\mathrm{r},i}\}$ and randomized boundary positions, as in Fig. 8. We train Neuro-COTANS on different SNRs in the relevant SNR range, including the transition region of global errors. We generate 50 000 training and 3000 validation images for each SNR. As COTANS curves for all receivers are summed up into a single image, Neuro-COTANS does not need modification to handle variable numbers of receivers.



(a)

(b)

FIG. 7. (Color online) The COTANS accumulator for $\{\hat{\tau}_{i,j}\}$ (a) and its boundary estimates (b). Note that the image is periodic in azimuth.

TABLE II. Training specifications for Neuro-COTANS.

Parameter	Optimizer	Number of epochs	Mini batch size	ℓ_2 regularization	Initial learn rate
Value	Adam	25	50	0	0.001

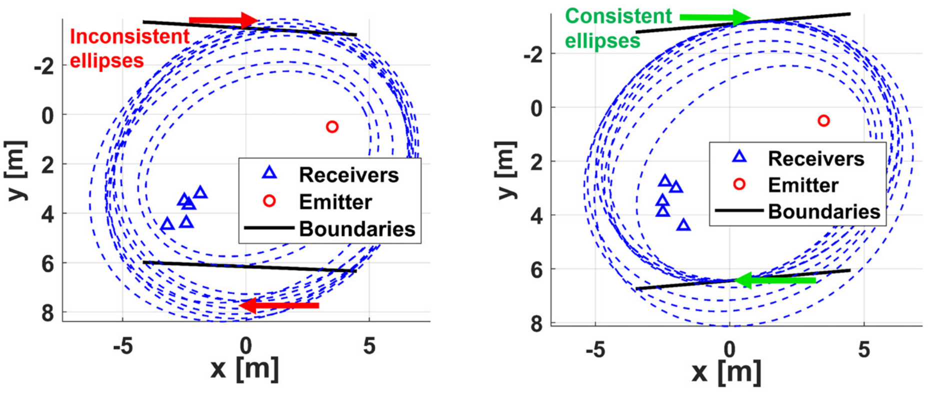
Computational resources constrain the COTANS image dimensions for the NNs, necessitating a trade-off between complexity and bin resolutions. Thus, to plot the complete COTANS curves, we adopt input dimensions of 101 pixels in ρ and 360 pixels in θ. The θ-resolution is therefore 1°, whereas we use a ρ-resolution of 0.1 m, leading to a ρ-axis of 0–10 m. This range of ρ is appropriate for our simulation and real experiment settings but can be scaled or translated to a different interval for different applications while keeping the same image dimensions. We are able to do so because the NN is trained on a scaled ρ range of $[0, 1]$ instead of a specific $[0, \rho_{\max}]$. Therefore, while our coarse-resolution COTANS images currently represent a maximum range of 10 m and a range resolution of 0.1 m, they can also represent a maximum range of 100 m and a range resolution of 1 m without any modification to the method itself. The method is more or less agnostic to the dimensions that are actually represented in the real-life setting and may only begin to break down when the problem is scaled to much larger dimensions, when other assumptions such as those on the speed of sound begin to break down.

To surpass the performance limitations imposed by the resolution constraints on COTANS images, we design *successive* stages of Neuro-COTANS with finer resolutions for refined performance. This leads to the multistage, multiresolution Neuro-COTANS method, as summarized in Fig. 9. A single NN, termed Coarse-NN, forms the first stage which is trained on coarse-resolution, complete COTANS images. Whereas this NN is a good overall first-pass estimator, its performance saturates at high SNRs, where the limited resolution can constrain performance. To overcome this limitation, we zoom into the vicinities of Coarse-NN boundary estimates on the full images and perform further stages of estimation on these high-resoluting images. At each stage of zooming, we increase the ρ- and θ-resolutions by a certain (fixed) factor, such as ten, so that the second stage in our particular implementation yields images with a resolution of $(0.01\,\text{m}, 0.1^\circ)$ in (ρ, θ). The image dimensions are retained to be the same at 101×360 pixels such that pretraining employed for the first stage can be used for successive stages as well. This procedure is illustrated in Fig. 10, where we have highlighted the vicinity of one of the two boundary estimates from Coarse-NN.

To train a NN for fine-resolution images, first, we generate a new set of 25 000 training and 1500 validation coarse-resolution COTANS images for each SNR. Then, we zoom into the $1\,\text{m} \times 3.6^\circ$ image region around the Coarse-NN estimates of each boundary and generate images within these regions to obtain a zoomed dataset of 50 000 training and 3000 validation images. Training the same NN architecture with this dataset yields a new NN, which we call Fine-NN. Using Coarse-NN and Fine-NN in sequence leads to enhanced performance.

After using Fine-NN for estimation, further stages of zooming have diminishing performance returns. These stages yield images featuring crisscrossing lines, as in Fig. 11, rather than intersecting curves. Therefore, we employ a basic weighted averaging procedure for interpolating the estimates for these stages instead of training new stages of NNs. Recall that COTANS images are scaled such that the maximum intensity pixel, corresponding to the maximum number of crossing curves, has a value of one. We heuristically threshold the image pixels to only retain those with values ≥ 0.5. Weighting every remaining pixel by its value and finding the average pixel coordinates yields a refined boundary estimate. This stage's performance gain only becomes relevant at high SNR.

Neuro-COTANS performs better when the image inputs to Coarse-NN are the echo-labeled curves from a single boundary rather than the sum of unlabeled curves from all the boundaries. Such single-boundary images can have less-distorted maxima, and the accuracy of successive stages of



(a) (b)

FIG. 8. (Color online) Random geometries and the corresponding NLOS ellipses, showing the transition region at 8 dB SNR (a) and high-SNR region at 20 dB SNR (b).

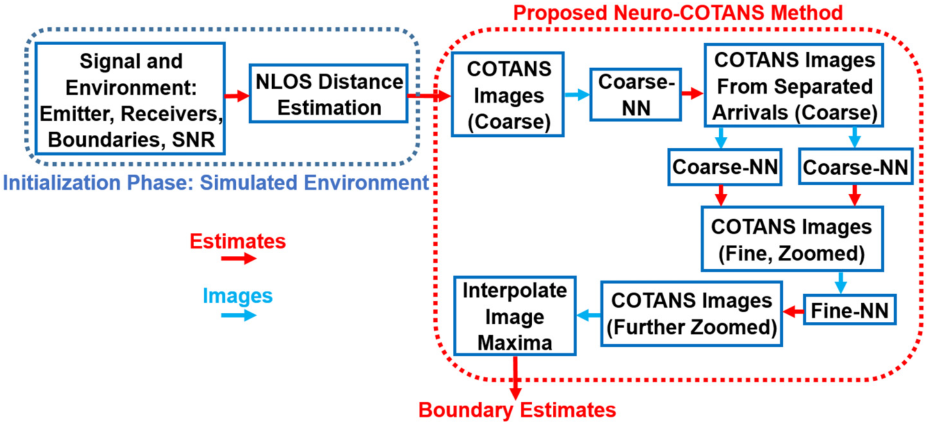

Proposed Neuro-COTANS Method
Signal and Environment: Emitter, Receivers, Boundaries, SNR
NLOS Distance Estimation
Initialization Phase: Simulated Environment
COTANS Images (Coarse)
Coarse-NN
COTANS Images From Separated Arrivals (Coarse)
Coarse-NN
Coarse-NN
COTANS Images (Fine, Zoomed)
Fine-NN
COTANS Images (Further Zoomed)
Interpolate Image Maxima
Boundary Estimates
Estimates
Images

FIG. 9. (Color online) Flowchart of the Neuro-COTANS method.

refinement is contingent on this Coarse-NN performance. However, we do not attempt to solve the complete, combinatorial echo labeling problem. Instead, we first perform boundary estimation with Coarse-NN and then use the resulting estimates to estimate the correct assignment of echoes. Each boundary estimate corresponds to a set of NLOS TOAs, $\tilde{\tau}_{i,j'}$, to each receiver, which differs from the NLOS TOAs, $\hat{\tau}_{i,j}$, that were obtained by time-delay estimation. At each receiver, i, we make the echo assignment of j to j' such that

$$\min_{\pi(\cdot)} \sum_{j=1}^{N} (\tilde{\tau}_{i,j'} - \hat{\tau}_{i,j})^2 \quad \text{subject to } \pi(j) = j', \tag{10}$$

where $\pi(\cdot)$ is a permutation mapping. We, then, generate separate new COTANS images for each j' from the sorted time-delays, $\hat{\tau}_{i,j'}$, and use Coarse-NN for estimation. Note that this procedure does not necessarily lead to a completely correct labeling when global errors have been made, but the performance of Coarse-NN is strong enough that the labeling of echoes with only local errors is generally accurate. The correct assignment of most of the echoes to separate images is found sufficient to deliver superior performance relative to other methods.

In implementation, the zooming operations that we have outlined do not involve simply generating higher-resolution complete Neuro-COTANS images and taking specific regions of these images into consideration. For a large dataset, the memory requirements of progressively higher-resolution images become prohibitive. Instead, we maintain a lookup table of the coarse-resolution azimuth transformations, $\theta \rightarrow \theta_{\text{COTANS}}$. We generate the higher-resolution images by running the COTANS transformation only on the relevant θ-interval that yields the desired θ_{COTANS}-interval.

V. PRIOR ART IN BOUNDARY ESTIMATION

In this section, we summarize the least squares (LS) (Cheung *et al.*, 2004) and EDM (Dokmanic *et al.*, 2013) algorithms, which are state-of-the-art alternatives to Neuro-COTANS. LS and EDM are used to localize emitters using LOS arrivals; they are similarly used for boundary estimation by localizing virtual emitters using NLOS arrivals. Whereas Neuro-COTANS is currently limited to 2D, LS and EDM have the advantage of being three-dimensional (3D) estimation methods. However, we will observe that they assume a small-scale error regime and require solving the

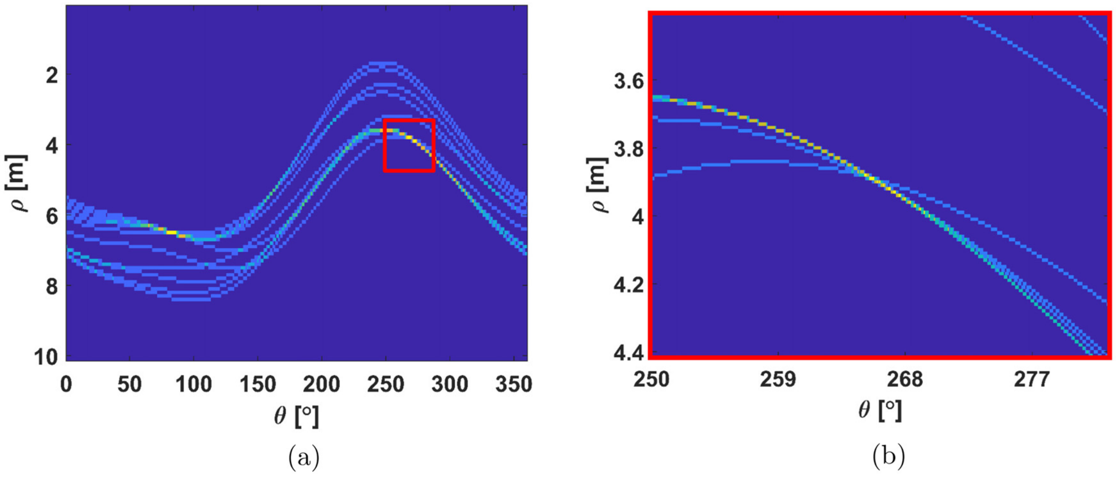


FIG. 10. (Color online) A coarse-resolution COTANS image with the region highlighted in red centered on one of the Coarse-NN boundary estimates (a) and the resulting zoomed-in image in stage 2 (b).

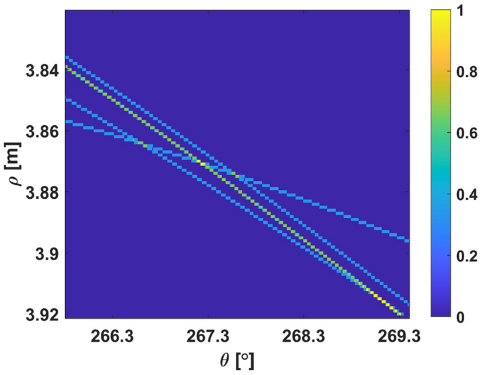


FIG. 11. (Color online) Example of a second-stage zoomed COTANS image, where finding the center of mass of the crossing lines is preferred to training a new NN.

computationally hard echo labeling problem that Neuro-COTANS bypasses.

A. The least-squares solution for boundary estimation

In LS localization, we solve for the variables, x_e and y_e, that define $\mathbf{p}_\mathrm{e}$. Assuming a high-SNR regime, we have the Gaussian noise-corrupted range estimates, $\{r_i\}_{i=1}^{M}$, as

$$r_i = d_i + n_i = \sqrt{(x_\mathrm{e} - x_i)^2 + (y_\mathrm{e} - y_i)^2} + \xi_i, \tag{11}$$

where $\xi_i \sim \mathcal{N}(0, \sigma_r^2)$. Defining $r_\mathrm{e} \triangleq \sqrt{x_\mathrm{e}^2 + y_\mathrm{e}^2}$, we solve the equations $\mathbf{Ax} = \mathbf{b}$ such that

$$\mathbf{A} \triangleq \begin{bmatrix} x_1 & y_1 & -0.5 \\ \vdots & \vdots & \vdots \\ x_M & y_M & -0.5 \end{bmatrix}, \quad \mathbf{x} \triangleq \begin{bmatrix} x_\mathrm{e} \\ y_\mathrm{e} \\ r_\mathrm{e}^2 \end{bmatrix},$$

$$\mathbf{b} \triangleq \frac{1}{2} \begin{bmatrix} x_1^2 + y_1^2 - r_1^2 \\ \vdots \\ x_M^2 + y_M^2 - r_M^2 \end{bmatrix}. \tag{12}$$

In the presence of Gaussian noise, the LS solution for Eq. (12) is

$$\hat{\mathbf{x}}_\mathrm{e} = \underset{\tilde{\mathbf{x}}}{\operatorname{argmin}} \, (\mathbf{A}\tilde{\mathbf{x}} - \mathbf{b})^\mathrm{T} (\mathbf{A}\tilde{\mathbf{x}} - \mathbf{b}), \tag{13}$$

where $\tilde{\mathbf{x}}_\mathrm{e} = [\tilde{x}_\mathrm{e} \, \tilde{y}_\mathrm{e} \, \tilde{r}_\mathrm{e}^2]^\mathrm{T}$ is the optimization variable vector. In the presence of range estimate errors, it is critical to also introduce the nonlinear constraint,

$$\tilde{x}_\mathrm{e}^2 + \tilde{y}_\mathrm{e}^2 - \tilde{r}_\mathrm{e}^2 = 0, \tag{14}$$

and solving Eq. (13) constrained by Eq. (14) yields the LS algorithm for localization. For virtual emitters, we only need to reconceptualize our optimization variable vector as $\tilde{\mathbf{x}}_\mathrm{v} = [\tilde{x}_{\mathrm{v},j} \, \tilde{y}_{\mathrm{v},j} \, \tilde{r}_{\mathrm{v},j}^2]^\mathrm{T}$ for each boundary $j = 1, \ldots N$, and use the NLOS range estimates, $r_{i,j}$. Going beyond the literature, inequality constraints on $\tilde{x}_\mathrm{e}$ or $\tilde{y}_\mathrm{e}$ can also be added, thus, confining LS solutions to a region in which the emitter is known to be present, and thereby exploiting prior information on the environment and, in particular, on the area of interest.

B. Boundary estimation with Euclidean distance matrices

EDM was designed with the insight that given matrices of squared distances between nodes (receivers and virtual emitters), the matrix with the correct permutation of entries (i.e., with proper echo labeling) will have the lowest rank. Thus, if localization is performed on each permutation of echoes, we simultaneously obtain the correct echo labeling and virtual emitters (Dokmanic *et al.*, 2013). Noise, however, makes it difficult to use this (sensitive) rank criterion, and a heuristic metric and optimization method is applied instead.

EDM uses the "*s*-stress criterion," where given the measured $\{r_{i,j}\}$, the objective function to be minimized over $\tilde{x}_{\mathrm{v},j}$ and $\tilde{y}_{\mathrm{v},j}$ is

$$s\left(\tilde{x}_{\mathrm{v},j}, \tilde{y}_{\mathrm{v},j}\right) = \sum_i \left[\left(\tilde{x}_{\mathrm{v},j} - x_i\right)^2 + \left(\tilde{y}_{\mathrm{v},j} - y_i\right)^2 - r_{i,j}^2 \right]^2. \tag{15}$$

Ideally, Eq. (15) is minimized by the correct echo labeling. In our simulations, however, we had difficulty using *s*-stress for noisy echo labeling, where the correct set of echoes does not necessarily have the smallest *s*-stress. Therefore, in our implementation of EDM, we use the correctly echo-sorted results for boundary estimation, as with LS. Thus, we give these methods an inherent advantage, here, of access to knowledge that would need to be inferred in reality.

VI. CRLB FOR BOUNDARY RANGE ESTIMATION

We now derive a theoretical benchmark for the asymptotic performances of boundary estimation algorithms given noisy range measurements, as in Eq. (11), which enables us to verify their correct implementation, as well as assess the fundamental limitation of the asymptotic accuracy. Our starting point is the CRLB for emitter localization (Cheung *et al.*, 2004), which we modify to obtain the CRLB for boundary estimation. The CRLB in Cheung *et al.* (2004) is derived for variable SNR at each receiver; as we assume the same SNR at each receiver (to present performance curves with respect to overall SNR), we note that the case of equal SNR is a special case of this CRLB formula.

The CRLB for estimating $\mathbf{p}_\mathrm{e} = [x_\mathrm{e} \, y_\mathrm{e}]^\mathrm{T}$ is obtained through the Fisher information matrix,

$$\mathbf{I}(\mathbf{p}_\mathrm{e}) \triangleq \begin{bmatrix} \mathrm{I}_{1,1} & \mathrm{I}_{1,2} \\ \mathrm{I}_{1,2} & \mathrm{I}_{2,2} \end{bmatrix}, \tag{16}$$

where

$$\mathrm{I}_{1,1} = \sum_{i=1}^{M} \frac{(x_\mathrm{e} - x_i)^2}{\sigma^2 \left((x_\mathrm{e} - x_i)^2 + (y_\mathrm{e} - y_i)^2 \right)}, \tag{17}$$

$$\mathrm{I}_{1,2} = \sum_{i=1}^{M} \frac{(x_\mathrm{e} - x_i)(y_\mathrm{e} - y_i)}{\sigma^2\Big((x_\mathrm{e} - x_i)^2 + (y_\mathrm{e} - y_i)^2\Big)}, \tag{18}$$

$$\mathrm{I}_{2,2} = \sum_{i=1}^{M} \frac{(y_\mathrm{e} - y_i)^2}{\sigma^2\Big((x_\mathrm{e} - x_i)^2 + (y_\mathrm{e} - y_i)^2\Big)}. \tag{19}$$

Inverting Eq. (16) yields a matrix with terms $\mathrm{J}_{1,1}$, $\mathrm{J}_{1,2}$, and $\mathrm{J}_{2,2}$. The (individual) CRLBs for the coordinates of $\mathbf{p}_\mathrm{e}$ are

$$\mathrm{CRLB}(\hat{x}_\mathrm{e}) = \mathrm{J}_{1,1} = \frac{\mathrm{I}_{2,2}}{\mathrm{I}_{1,1}\mathrm{I}_{2,2} - \mathrm{I}_{1,2}^2}, \tag{20}$$

$$\mathrm{CRLB}(\hat{y}_\mathrm{e}) = \mathrm{J}_{2,2} = \frac{\mathrm{I}_{1,1}}{\mathrm{I}_{1,1}\mathrm{I}_{2,2} - \mathrm{I}_{1,2}^2}. \tag{21}$$

We also have the cross term,

$$\mathrm{J}_{1,2} = -\frac{\mathrm{I}_{1,2}}{\mathrm{I}_{1,1}\mathrm{I}_{2,2} - \mathrm{I}_{1,2}^2}. \tag{22}$$

Then, the CRLB on the mean square range estimation error is $\mathrm{CRLB}(\hat{x}_\mathrm{e}) + \mathrm{CRLB}(\hat{y}_\mathrm{e})$ (Jia and Buehrer, 2008). For virtual emitters, $\mathbf{p}_\mathrm{v} = [x_\mathrm{v}\, y_\mathrm{v}]^\mathrm{T}$ is substituted into Eqs. (20) and (21).

We specify a given boundary by its ρ and θ and use the root-mean squared error (RMSE) range error as the measure of performance of boundary estimation. Thus, we transform the CRLB for $\hat{\mathbf{p}}_\mathrm{v}$ into a CRLB for boundary range estimation. Given $\mathbf{p}_\mathrm{v}$, we obtain the orthogonal vector, $\boldsymbol{\rho}_\mathrm{v}$, to the boundary, referring to Fig. 12. First, note that $\mathbf{n}_\mathrm{v}$, pointing from the emitter to the virtual emitter, has the same direction as the unit normal $\hat{\mathbf{n}}_\mathrm{v}$ from the origin to the boundary:

$$\hat{\mathbf{n}}_\mathrm{v} = \left[\frac{x_\mathrm{v} - x_\mathrm{e}}{\sqrt{(x_\mathrm{v} - x_\mathrm{e})^2 + (y_\mathrm{v} - y_\mathrm{e})^2}} \frac{y_\mathrm{v} - y_\mathrm{e}}{\sqrt{(x_\mathrm{v} - x_\mathrm{e})^2 + (y_\mathrm{v} - y_\mathrm{e})^2}}\right]^\mathrm{T}. \tag{23}$$

The vector from the origin to the intersection point of the boundary and $\mathbf{n}_\mathrm{v}$ is

$$\mathbf{b}_\mathrm{v} = \left[\frac{x_\mathrm{v} + x_\mathrm{e}}{2} \quad \frac{y_\mathrm{v} + y_\mathrm{e}}{2}\right]^\mathrm{T}. \tag{24}$$

The orthogonal vector from the origin to the boundary is then given by the orthogonal projection of $\mathbf{b}_\mathrm{v}$ onto $\hat{\mathbf{n}}_\mathrm{v}$, i.e., by $(\mathbf{b}_\mathrm{v}^\mathrm{T}\hat{\mathbf{n}}_\mathrm{v})\hat{\mathbf{n}}_\mathrm{v}$. The result is

$$\boldsymbol{\rho}_\mathrm{v} = \left[\frac{(x_\mathrm{v} - x_\mathrm{e})(x_\mathrm{v}^2 - x_\mathrm{e}^2 + y_\mathrm{v}^2 - y_\mathrm{e}^2)}{2\Big((x_\mathrm{v} - x_\mathrm{e})^2 + (y_\mathrm{v} - y_\mathrm{e})^2\Big)} \quad \frac{(y_\mathrm{v} - y_\mathrm{e})(x_\mathrm{v}^2 - x_\mathrm{e}^2 + y_\mathrm{v}^2 - y_\mathrm{e}^2)}{2\Big((x_\mathrm{v} - x_\mathrm{e})^2 + (y_\mathrm{v} - y_\mathrm{e})^2\Big)}\right]^\mathrm{T}. \tag{25}$$

Finally, the range to the boundary is given by the magnitude of $\boldsymbol{\rho}_\mathrm{v}$, yielding

$$\rho_\mathrm{v} = \frac{|x_\mathrm{v}^2 - x_\mathrm{e}^2 + y_\mathrm{v}^2 - y_\mathrm{e}^2|}{2\sqrt{(x_\mathrm{v} - x_\mathrm{e})^2 + (y_\mathrm{v} - y_\mathrm{e})^2}}. \tag{26}$$

This series of operations is the same as those used to transform the LS and EDM virtual emitter estimates into boundary estimates. Without loss of generality, we will assume that $x_\mathrm{v}^2 - x_\mathrm{e}^2 + y_\mathrm{v}^2 - y_\mathrm{e}^2 > 0$ for our subsequent derivations.

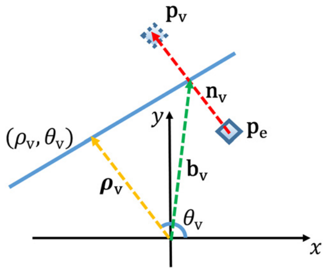


FIG. 12. (Color online) A geometric reference for the transformation of the CRLB for virtual emitter positions into the CRLB for boundary range estimation.

We have now obtained the range as a function of x_v and y_v, where x_e and y_e are known constants. The CRLBs for x_v and y_v can now be transformed into a CRLB for ρ_v. We calculate the derivatives of ρ_v with respect to x_v and y_v as

$$\frac{\partial\rho_\mathrm{v}}{\partial x_\mathrm{v}} = \frac{x_\mathrm{v}\Big((x_\mathrm{v} - x_\mathrm{e})^2 + (y_\mathrm{v} - y_\mathrm{e})^2\Big)}{\Big((x_\mathrm{v} - x_\mathrm{e})^2 + (y_\mathrm{v} - y_\mathrm{e})^2\Big)^{3/2}} - \frac{(x_\mathrm{v} - x_\mathrm{e})(x_\mathrm{v}^2 - x_\mathrm{e}^2 + y_\mathrm{v}^2 - y_\mathrm{e}^2)}{2\Big((x_\mathrm{v} - x_\mathrm{e})^2 + (y_\mathrm{v} - y_\mathrm{e})^2\Big)^{3/2}}, \tag{27}$$

$$\frac{\partial\rho_\mathrm{v}}{\partial y_\mathrm{v}} = \frac{y_\mathrm{v}\Big((x_\mathrm{v} - x_\mathrm{e})^2 + (y_\mathrm{v} - y_\mathrm{e})^2\Big)}{\Big((x_\mathrm{v} - x_\mathrm{e})^2 + (y_\mathrm{v} - y_\mathrm{e})^2\Big)^{3/2}} - \frac{(y_\mathrm{v} - y_\mathrm{e})(x_\mathrm{v}^2 - x_\mathrm{e}^2 + y_\mathrm{v}^2 - y_\mathrm{e}^2)}{2\Big((x_\mathrm{v} - x_\mathrm{e})^2 + (y_\mathrm{v} - y_\mathrm{e})^2\Big)^{3/2}}. \tag{28}$$

The resulting CRLB for ρ_{v} is obtained by the transformation of parameters as

$$\mathrm{CRLB}(\hat{\rho}_{\mathrm{v}}) = \begin{bmatrix} \dfrac{\partial\rho_{\mathrm{v}}}{\partial x_{\mathrm{v}}} & \dfrac{\partial\rho_{\mathrm{v}}}{\partial y_{\mathrm{v}}} \end{bmatrix} \begin{bmatrix} \mathrm{I}_{1,1}^{-1} & \mathrm{I}_{1,2}^{-1} \\ \mathrm{I}_{1,2}^{-1} & \mathrm{I}_{2,2}^{-1} \end{bmatrix} \begin{bmatrix} \dfrac{\partial\rho_{\mathrm{v}}}{\partial x_{\mathrm{v}}} \\ \dfrac{\partial\rho_{\mathrm{v}}}{\partial y_{\mathrm{v}}} \end{bmatrix}$$

$$= \mathrm{I}_{1,1}^{-1}\left(\frac{\partial\rho_{\mathrm{v}}}{\partial x_{\mathrm{v}}}\right)^2 + 2\mathrm{I}_{1,2}^{-1}\frac{\partial\rho_{\mathrm{v}}}{\partial x_{\mathrm{v}}}\frac{\partial\rho_{\mathrm{v}}}{\partial y_{\mathrm{v}}} + \mathrm{I}_{2,2}^{-1}\left(\frac{\partial\rho_{\mathrm{v}}}{\partial y_{\mathrm{v}}}\right)^2. \tag{29}$$

$\sqrt{\mathrm{CRLB}(\hat{\rho}_{\mathrm{v}})}$ is, therefore, the lower bound on the RMSE boundary range estimation error. It will be observed that the CRLB falls exponentially with SNR, as will be shown for a case example in Fig. 14.

VII. SIMULATION AND EXPERIMENTAL RESULTS

We study the performance of Neuro-COTANS, obtaining time-delay estimates as per Sec. III, and evaluating the NN method first in simulation to compare its performance to LS and EDM and also to the CRLB derived in Sec. VI. After retraining Neuro-COTANS, we apply it to a real-life underwater acoustic setting, where it outperforms LS. Finally, we conduct simulations that demonstrate the robustness of Neuro-COTANS to model mismatch and reduce prior knowledge of the environment.

A. Simulated performances

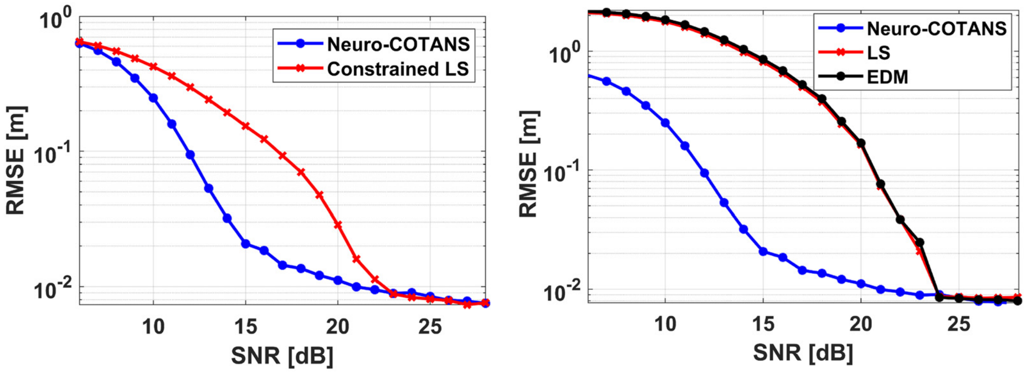
We test Neuro-COTANS on $K = 50\,000$ COTANS images per SNR value, having trained it previously on 14 different SNRs in the 10–30 dB SNR range (which covers low, medium, and high SNRs for this particular scenario). Once Neuro-COTANS has been trained on this wide range of representative SNRs, it is applied to different ranges of SNRs without needing to be retrained. One boundary has its ρ and θ parameters uniformly drawn from the intervals $[3, 3.5]$ m and $[260, 280]^{\circ}$, respectively, whereas the other has parameters in $[6, 6.5]$ m and $[80, 100]^{\circ}$. These boundaries model a sea surface and shallow bottom, as in Fig. 8. The variations in range/angles of the boundaries could arise from either surface wave motion in the case of the sea surface or bathymetric variations in the case of the seafloor. The $\mathbf{p}_{\mathrm{e}}$ and $\{\mathbf{p}_{\mathrm{r},i}\}$ are uniformly drawn from the vicinities of two fixed points, $(3.5, 0.5)$ and $(-2.5, 3.5)$, respectively. Although we have conducted simulation experiments with three boundaries as well, the resulting performance curves are qualitatively similar to the two-boundary case. Hence, we only present the two-boundary simulation results. Also, we only present the range estimates as the azimuth estimates are likewise qualitatively similar.

We first compare Neuro-COTANS to an ideal LS implementation, which is initialized at the ground truth locations of the $\{\mathbf{p}_{\mathrm{v},j}\}$ with correct echo labeling and virtual emitter solutions constrained to lie within the same parameter space that Neuro-COTANS is trained on. Our performance metric is the range RMSE (in m) over all N reflective boundaries and all K environment realizations for each SNR S, which are defined as

$$\rho_{\mathrm{RMSE}}(S) \triangleq \sqrt{\frac{\sum_{j=1}^{N}\sum_{k=1}^{K}\left(\rho_{j,k}^{(S)} - \hat{\rho}_{j,k}^{(S)}\right)^2}{NK}}. \tag{30}$$

Figure 13(a) demonstrates that Neuro-COTANS and LS performances are nearly identical for SNR greater than 23 dB, a high-SNR operating regime in which global errors are rare and the $\{\hat{\tau}_{i,j}\}$ are accurate due to small noise. Below 23 dB SNR, as global errors become increasingly common and we transition to an intermediate-SNR regime, Neuro-COTANS outperforms LS by up to 6 dB SNR. This performance advantage narrows at low SNRs, where accurate boundary estimation becomes infeasible using either method.

In Fig. 13(a), it appears that Neuro-COTANS merely outperforms LS at SNR less than 23 dB and has equivalent performance otherwise. In fact, the LS sometimes suffers failures due to global errors, which are constrained to lie within a relatively narrow parameter space. We conduct the same experiment with unconstrained LS solutions and also



(a)

(b)

FIG. 13. (Color online) Neuro-COTANS performance compared to constrained LS (a) and LS and EDM (b).

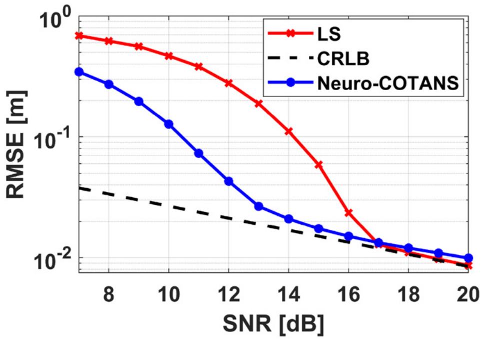


FIG. 14. (Color online) The CRLB for single boundary range estimation, calculated for a fixed scenario, and compared against the Neuro-COTANS and LS performances for the same scenario.

TABLE III. SOARS estimation error magnitudes for boundaries, with drastic errors bolded.

Parameter	ρ_1	ρ_2	θ_1	θ_2
Neuro-COTANS	0.083 m	0.019 m	0.7°	2.4°
LS	0.134 m	0.025 m	**174°**	5.7°

apply EDM to obtain Fig. 13(b). Neuro-COTANS outperforms LS and EDM by up to 9 dB SNR and marginally outperforms them in the high-SNR regime as well. LS and EDM have similar performances, which arises from how they both minimize the squared error between measured and estimated distances.

Finally, we conduct simulations with different noise realizations and a fixed environment, allowing us to compare against the CRLB in Sec. VI for a single boundary present in the environment. Because LS localization performance comes close to the CRLB for emitter localization at high SNR, LS virtual emitter localization should similarly come close to the boundary range estimation CRLB. We observe that this is the case in Fig. 14, confirming that the CRLB has been formulated correctly and is a benchmark for performance in the high-SNR regime as intended.

B. Underwater acoustic experiment results

To verify that Neuro-COTANS performs well under realistic conditions, we perform experiments in a controlled underwater acoustic setting. We use the Scripps Ocean-Atmosphere Research Simulator (SOARS) wave tank facility [Fig. 15(a)] with the top-view of the experiment geometry as in Fig. 15(b). The hydrophones are suspended at the same depth such that we have a 2D estimation problem for the side walls, located at $y = -1.235$ m and $y = 1.235$ m. The NLOS reflections from the other boundaries arrive later and, therefore, were time gated to reduce the problem to a 2D case.

We retrain the Neuro-COTANS approach of Fig. 13(b) on a dataset that is similar to the geometric scenario of SOARS. We then use the COTANS image generated from the SOARS experiment to estimate the boundaries. The results with Neuro-COTANS and LS are given in Table III. Neuro-COTANS achieves an accuracy on the order of centimeters in ρ and a few degrees in θ. LS suffers a large error for one boundary and is consistently outperformed by Neuro-COTANS.

C. Neuro-COTANS robustness analysis

We now present simulation results that study Neuro-COTANS's robustness. A common pitfall in NN design is to overtrain on a particular dataset, yielding a network that is fragile to model mismatch or one that only works with a narrow parameter space. A robust method will have a gradual performance decline under model mismatch rather than abrupt deterioration and remain functional for difficult estimation scenarios.

First, we explore the effect of model mismatch in the $\mathbf{p}_e$ assumed in generating the training data. We train Neuro-COTANS with $\mathbf{p}_e$ drawn randomly from a square 0.25 m wide, centered at $(3.5, 0.5)$. We progressively shift the center of this square by 0.25 m in x and y while continuing to use Neuro-COTANS as trained on source locations within the original square region. We thereby obtain the results in

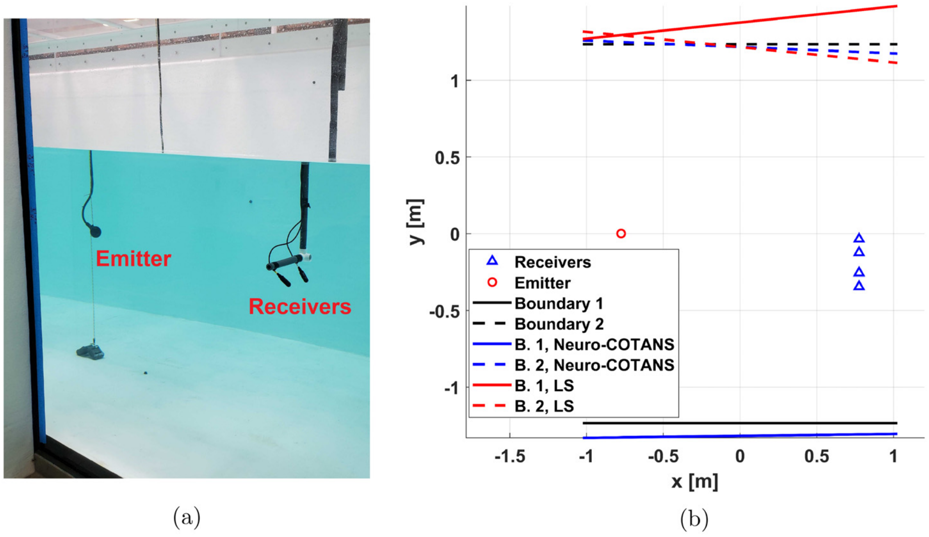


FIG. 15. (Color online) Emitter deployment in the SOARS water tank (a) and the top-view schematic of the water tank illustrating the experiment geometry and the estimation results (b).

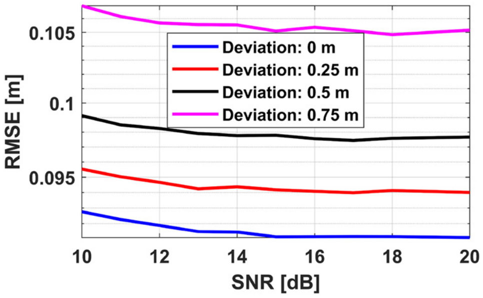


FIG. 16. (Color online) Performance of Neuro-COTANS as the average assumed emitter position increasingly deviates from the true one.

Fig. 16, where Neuro-COTANS continues to be stable despite increasingly worse performance as model mismatch creates estimation biases that are unaccounted for. In this experiment, Fine-NN does not yield a performance improvement over Coarse-NN as the error due to mismatch is dominant.

Whereas the source and receiver locations may have discrepancies with the assumed ground truth, larger errors are more likely to occur in the emitter rather than in the receiver locations. We have control over receiver deployment, and the error here is mainly the result of measurement errors under realistic dynamic sea conditions. The emitter position, however, is an estimate from a previous localization stage, which is assumed here to be the ground truth position for environment estimation. Thus, we expect the main source of model mismatch errors to be the discrepancy in the emitter position rather than in the receivers.

In a different experiment, we relax the bounds on θ that the $\boldsymbol{\eta}_j$ can have such that Neuro-COTANS handles a larger parameter space. We retrain Neuro-COTANS, originally having a ±10° θ-margin as in Sec. VII A, with ±20° and ±30° θ-margins as well. A larger parameter space requires a correspondingly larger training set, but we instead use 50 000 training images per SNR as before to assess Neuro-COTANS's robustness. Our results in Fig. 17(a) indicate that Neuro-COTANS remains stable despite being trained on harder scenarios.

To analyze the deterioration caused by a larger parameter space, we retrain Neuro-COTANS to operate on a ±20° θ-margin, and then test it on the same ±10° θ-margin dataset of the original NN. The resulting performances in Fig. 17(b) indicate that by sequentially using Neuro-COTANS on progressively smaller parameter spaces, we could achieve greater accuracy.

VIII. CONCLUDING REMARKS

In this paper, we propose the Neuro-COTANS image regression method for 2D reflective boundary estimation, exploiting the multi-scale filtering and domain adaptation capabilities of CNNs. Our method leverages prior knowledge of the environment to deliver robust performance in simulation and experimental underwater acoustic settings despite model mismatch, in part, by avoiding separated suboptimal echo labeling and filtering steps, which are fragile without high SNR. These experiments demonstrated that Neuro-COTANS was consistently accurate even when large errors were present in the time-delay estimates, outperforming alternative state-of-the-art boundary estimation methods.

The richness of deep learning techniques enables a range of potential improvements and extensions to Neuro-COTANS. Neuro-COTANS currently works in 2D, and extending it to 3D is nontrivial. Although replacing the 2D convolutional layers with 3D layers is a first step, the key difficulty is that 3D data increases computational demands dramatically (the "curse of dimensionality"). Hence, a 3D Neuro-COTANS requires a network and data structure that makes more efficient use of computational resources.

The fact that the NNs have to be retrained at all is an important limitation of the proposed Neuro-COTANS method, which will have to be addressed in future work. Ideally, we would train the NNs on a much wider variety of boundary, emitter, and receiver positions, and the method would provide a good estimate for any feasible estimation scenario without the need for retraining. The current limitation is a result of the NN architecture being used, rather than being a limitation of the overall methodology. AlexNet, which we adopted early on because of its proven track record in image regression, requires too many images per SNR level for training to generalize to a wider range of

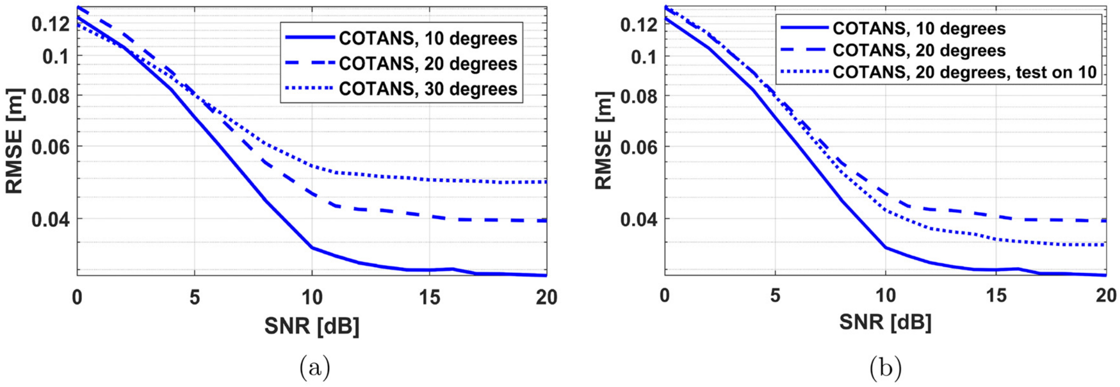


FIG. 17. (Color online) Neuro-COTANS performance on progressively larger θ-margins (a) and performance on the same margin after being trained on different margins (b).

https://doi.org/10.1121/10.0026437

problem geometries. Our main motivation in envisioning a custom NN architecture for the future is to overcome this limitation.

Neuro-COTANS currently estimates the locations of reflective boundaries but could potentially be modified to solve related problems as well. While the final layer of the NN only provides the boundary location parameters, we could expand the network's capabilities to label the arrivals, estimate the number of boundaries present, or produce a metric of confidence in the estimation results. If the NLOS TOAs come from an underwater acoustic simulator, such as Bellhop (Porter, 2011), the training dataset would be richer than our current signal model, potentially leading to better performance in ocean deployments.

This work calls for a number of studies to potentially improve Neuro-COTANS. There are alternative ways of assembling the training data such as incorporating attenuation coefficients by scaling each NLOS curve by its magnitude in the COTANS images. It may also be possible to modify the network inputs, providing this data to the NN in other formats than COTANS images, to explore a wider range of estimation methods. The MSE-based cost function is heuristic and its modification can also improve performance (Huang *et al.*, 2018). The further derivation in 3D of a CRLB for the azimuth, θ, and elevation, ϕ, for each boundary would lead to important insights. Neuro-COTANS's demonstration of the feasibility and desirability of transform-based NN boundary estimation is encouraging for these future studies.

ACKNOWLEDGMENT

This work was supported, in part, by the Office of Naval Research (ONR) under Grant Nos. N00014-19-1-2661, N00014-24-1-2004 (to A.C.S.), and N000014-23-1-2803 (to G.W.W.). We would like to thank Dr. Jim Preisig for his valuable insights and Dr. Dale Stokes for his help with the SOARS water tank experiment.

AUTHOR DECLARATIONS

Conflicts of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

[1] Therefore, in underwater acoustic settings, the speed of sound can be assumed to be constant.

[2] For the emitter and receivers' locations, we use Cartesian coordinates.

[3] This can be a valid assumption for well-mixed shallow-water settings at short ranges and high acoustic frequencies (Chitre, 2007; Too *et al.*, 2019).

[4] This is the maximum likelihood estimator for the time-delay estimation of a single arrival.

[5] If complex signals are used, some constants are modified in the following results.

[6] For the sake of deriving time-delay statistics, we assume that S is the same at each receiver such that we can define an average error versus S for assessing boundary estimation performance. In practice, S varies across receivers.

[7] We have also derived a COTANS transform in 3D for spheroids, which are outside of our current scope.

[8] Given constraints, such as a convex environment, we can also select only the single-reflections as COTANS inputs (Park and Choi, 2021).

Ali, W. H., Bhabra, M. S., Lermusiaux, P. F., March, A., Edwards, J. R., Rimpau, K., and Ryu, P. (**2019**). "Stochastic oceanographic-acoustic prediction and Bayesian inversion for wide area ocean floor mapping," in *OCEANS 2019 MTS/IEEE Seattle*, October 27–31, 2019, Seattle, WA (IEEE, New York), pp. 1–10.

Antonacci, F., Filos, J., Thomas, M. R., Habets, E. A., Sarti, A., Naylor, P. A., and Tubaro, S. (**2012**). "Inference of room geometry from acoustic impulse responses," IEEE Trans. Audio, Speech, Lang. Process. **20**(10), 2683–2695.

Antonacci, F., Sarti, A., and Tubaro, S. (**2010**). "Geometric reconstruction of the environment from its response to multiple acoustic emissions," in *ICASSP 2010—IEEE International Conference Acoustics, Speech and Signal Processing*, March 14–19, 2010, Dallas, TX (IEEE, New York), pp. 2822–2825.

Arikan, T., Weiss, A., Vishnu, H., Deane, G. B., Singer, A. C., and Wornell, G. W. (**2023a**). "An architecture for passive joint localization and structure learning in reverberant environments," J. Acoust. Soc. Am. **153**(1), 665–677.

Arikan, T., Weiss, A., Vishnu, H., Deane, G. B., Singer, A. C., and Wornell, G. W. (**2023b**). "Learning environmental structure using acoustic probes with a deep neural network," in *ICASSP 2023—IEEE International Conference Acoustics, Speech and Signal Processing*, June 4–10, 2023, Rhodes, Greece (IEEE, New York), Vol. 26, pp. 1–5.

Borrmann, D., Elseberg, J., Lingemann, K., and Nüchter, A. (**2011**). "The 3D Hough transform for plane detection in point clouds: A review and a new accumulator design," 3D Res. **2**(2), 3.

Brutti, A., Omologo, M., and Svaizer, P. (**2010**). "Multiple source localization based on acoustic map de-emphasis," EURASIP J. Audio, Speech, Music Process. **2010**, 147495.

Cheung, K. W., So, H. C., Ma, W. K., and Chan, Y. T. (**2004**). "Least squares algorithms for time-of-arrival-based mobile location," IEEE Trans. Signal Process. **52**(4), 1121–1130.

Chitre, M. (**2007**). "A high-frequency warm shallow water acoustic communications channel model and measurements," J. Acoust. Soc. Am. **122**(5), 2580–2586.

Crocco, M., Trucco, A., and Del Bue, A. (**2017**). "Uncalibrated 3D room geometry estimation from sound impulse responses," J. Franklin Inst. **354**(18), 8678–8709.

Dardari, D., Chong, C. C., and Win, M. Z. (**2006**). "Improved lower bounds on time-of-arrival estimation error in realistic UWB channels," in *2006 IEEE International Conference on Ultra-Wideband*, September 24–27, 2006, Waltham, Ma (IEEE, New York), pp. 531–537.

Dardari, D., Conti, A., Ferner, U., Giorgetti, A., and Win, M. Z. (**2009**). "Ranging with ultrawide bandwidth signals in multipath environments," Proc. IEEE **97**(2), 404–426.

Deane, G. B. (**1994**). "A three-dimensional analysis of sound propagation in facetted geometries," J. Acoust. Soc. Am. **96**(5), 2897–2907.

Demirli, R., and Saniie, J. (**2001**). "Model-based estimation of ultrasonic echoes. Part I: Analysis and algorithms," IEEE Trans. Ultrason., Ferroelect., Freq. Contr. **48**(3), 787–802.

Dokmanic, I., Parhizkar, R., Ranieri, J., and Vetterli, M. (**2015**). "Euclidean distance matrices: Essential theory, algorithms, and applications," IEEE Signal Process. Mag. **32**(6), 12–30.

Dokmanic, I., Parhizkar, R., Walther, A., Lu, Y. M., and Vetterli, M. (**2013**). "Acoustic echoes reveal room shape," Proc. Natl. Acad. Sci. **110**(30), 12186–12191.

Huang, Z., Xu, J., Gong, Z., Wang, H., and Yan, Y. (**2018**). "Source localization using deep neural networks in a shallow water environment," J. Acoust. Soc. Am. **143**(5), 2922–2932.

Jia, T., and Buehrer, R. M. (**2008**). "A new Cramer-Rao lower bound for TOA-based localization," in *MILCOM 2008—IEEE Military Communications Conference*, November 16–19, 2008, San Diego, CA (IEEE, New York), pp. 1–5.

Korhonen, T. (**2008**). "Acoustic localization using reverberation with virtual microphones," in *Proceedings of the International Workshop on*

Acoustic Echo and Noise Control (IWAENC), September 14–17, 2008, Seattle, WA (IEEE, New York), pp. 211–223.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (**2012**). "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, December 3–6, 2012, Lake Tahoe, NV (IEEE, New York), Vol. 25.

Lee, J. Y., Kim, Y., Lee, S., Cho, W., and Kim, S. C. (**2019**). "Estimation of room shape using radio propagation channel analysis," IEEE Sens. J. **19**(24), 12316–12324.

Naseri, H., Costa, M., and Koivunen, V. (**2014**). "Multipath-aided cooperative network localization using convex optimization," in *48th Asilomar Conf. Signals, Syst. Comput*., November 2–5, 2014, Pacific Grove, CA (IEEE, New York), pp. 1515–1520.

Naseri, H., and Koivunen, V. (**2016**). "Cooperative simultaneous localization and mapping by exploiting multipath propagation," IEEE Trans. Signal Process. **65**(1), 200–211.

Niu, H., Gong, Z., Ozanich, E., Gerstoft, P., Wang, H., and Li, Z. (**2019**). "Deep-learning source localization using multi-frequency magnitude-only data," J. Acoust. Soc. Am. **146**(1), 211–222.

Niu, H., Ozanich, E., and Gerstoft, P. (**2017a**). "Ship localization in Santa Barbara Channel using machine learning classifiers," J. Acoust. Soc. Am. **142**(5), EL455–EL460.

Niu, H., Reeves, E., and Gerstoft, P. (**2017b**). "Source localization in an ocean waveguide using supervised machine learning," J. Acoust. Soc. Am. **142**(3), 1176–1188.

Park, S., and Choi, J. (**2021**). "Iterative echo labeling algorithm with convex hull expansion for room geometry estimation," IEEE/ACM Trans. Audio. Speech. Lang. Process. **29**(3), 1463–1478.

Porter, M. B. (**2011**). "The BELLHOP Manual and User's Guide: Preliminary draft," pp. 1–57.

Ribeiro, F., Zhang, C., Florêncio, D. A., and Ba, D. E. (**2010**). "Using reverberation to improve range and elevation discrimination for small array sound source localization," IEEE Trans. Audio, Speech, Lang. Process. **18**(7), 1781–1792.

Szegedy, C., Toshev, A., and Erhan, D. (**2013**). "Deep neural networks for object detection," in *Advances in Neural Information Processing Systems*, December 5–10, Lake Tahoe, NV (IEEE, New York), Vol. 26.

Too, Y. M., Chitre, M., Barbastathis, G., and Pallayil, V. (**2019**). "Localizing snapping shrimp noise using a small-aperture array," IEEE J. Ocean. Eng. **44**(1), 207–219.

Weinstein, E., and Weiss, A. (**1984**). "Fundamental limitations in passive time delay estimation–Part II: Wide-band systems," IEEE Trans. Acoust., Speech, Signal Process. **32**(5), 1064–1078.

Weiss, A., Arikan, T., Vishnu, H., Deane, G. B., Singer, A. C., and Wornell, G. W. (**2022**). "A semi-blind method for localization of underwater acoustic sources," IEEE Trans. Signal Process. **70**, 3090–3106.

Weiss, A., and Weinstein, E. (**1983**). "Fundamental limitations in passive time delay estimation–Part I: Narrow-band systems," IEEE Trans. Acoust., Speech, Signal Process. **31**(2), 472–486.

Wu, Y., Ayyalasomayajula, R., Bianco, M. J., Bharadia, D., and Gerstoft, P. (**2021**). "Sslide: Sound source localization for indoors based on deep learning," in *ICASSP 2021—IEEE International Conference Acoustics, Speech and Signal Processing*, June 6–11, 2021, Toronto, Canada (IEEE, New York), pp. 4680–4684.