# On Robustness and Out-of-Distribution Generalization of a Deep Neural Architecture for Underwater Acoustic Direct Localization

Amir Weiss
*Faculty of Engineering*
*Bar-Ilan University*
Ramat-Gan, Israel
amir.weiss@biu.ac.il

***Abstract*—Neural architectures have emerged in recent years as potentially enhanced solutions to the longstanding underwater acoustic localization problem. In this unique domain, nontrivial physical phenomena, such as depth-varying speed of sound, play a key role in the environment-dependent propagation model. Consequently, extracting location-related information of acoustic emitters, encapsulated in the relevant channel response, is both analytically and computationally challenging. Thus, deep neural networks (DNNs), which can circumvent the need for exact analytical characterizations and solutions, have been considered a prospective alternative to classical approaches. However, localization systems are typically required to be robust (in several respects), a property that DNNs do *not* necessarily possess. In this work, we focus on this critical aspect and show through a diverse set of simulations that our DNN-based localizer consistently manifests such robustness. Specifically, we focus on out-of-distribution generalization for input data, and show that our model remains performant under various distributional deviations.**

***Index Terms*—Localization, robust neural network, underwater acoustics, out-of-distribution generalization, compression.**

## I. Introduction

Underwater acoustic localization (UAL) remains a significant challenge in applications of growing relevance, ranging from environmental monitoring and underwater navigation to surveillance and defense (e.g., [1]–[3]). The unique physical characteristics of the underwater environment, such as depth-varying speed of sound and rich multipath propagation, create considerable complexities in accurately estimating the location of an acoustic source. For this reason, traditional localization methods (e.g., [4]), relying on analytical tractability, often involve simplifying assumptions, which can (and do) lead to performance degradation when faced with real-world conditions that deviate from the idealized assumptions.

Recent advances, particularly in machine learning, have opened new avenues for tackling complex signal processing tasks, including those in challenging environments like underwater acoustics, e.g., [5]–[9]. Unlike classical approaches, deep neural networks (DNNs) offer the ability to model (/learn) intricate statistical relations in the data without requiring explicit analytical formulations. This data-driven approach is particularly appealing in scenarios where the environmental conditions are either unknown or too complex to model accurately. Consequently, DNNs have been explored as a promising alternative solution for UAL, offering potential improvements in computational complexity and accuracy.

However, a significant concern with DNN-based systems is their generalization ability, particularly when encountering input data that follows a different distribution from that of the training data. This issue, known as out-of-distribution (OOD) generalization [10], is crucial for practical deployment of future DNN-based localization systems. In underwater environments, where the physical properties of the medium can vary significantly across operational scenarios, the robustness of a localization system to such variations is paramount.

Motivated by the above, we addresses this critical aspect by investigating the robustness and OOD generalization capabilities of a recently proposed DNN architecture for UAL, referred to as CNN-DLOC [11]. Complementing the theoretical study in [12] to a general model mismatch, our main contribution in this paper is a comprehensive robustness analysis of CNN-DLOC. Specifically, we conduct an extensive simulation study to evaluate the performance of CNN-DLOC under various types of distributional shifts, including changes in the statistical properties of the emitted acoustic waveform (Section IV-C), perturbations in the attenuation effects (Section IV-D), and the application of different compression techniques (Section IV-E). Our results show that CNN-DLOC maintains robustness across a wide range of OOD scenarios.

## II. A Data-Driven Formulation for Localization

Consider $L$ time-synchronized receivers positioned at different and known locations, each consisting of a single omnidirectional sensor. We assume that an acoustic source, emitting an unknown waveform, is present at an unknown location, denoted by $\mathbf{p} \in \mathcal{V} \subset \mathbb{R}^3$, where $\mathcal{V}$ is a three-dimensional environment of interest (i.e., the "uninformative prior" of $\mathbf{p}$). We further assume that the source is either static or moves slowly enough relative to the receivers' sampling period to be considered approximately static during the observation time.

In a classical formulation of the UAL problem, we seek a "handcrafted" signal model based on knowledge of the relevant physics. A solution is then developed for this (simplified)

model, assuming that it is a sufficiently good approximation to the true, underlying signal model. In contrast, our approach employs a data-driven methodology that does not require a precise analytical signal model, avoiding potentially unjustified simplifications that could degrade performance.

Accordingly, in the development of our solution, we do not use knowledge of the environment and the underlying propagation model, but assume that a dataset of received signals with labeled source positions $\mathcal{D}_{\text{train}}^{(T)} \triangleq \{(\mathbf{x}_1^{(i)}, \ldots, \mathbf{x}_L^{(i)}, \mathbf{p}^{(i)})\}_{i=1}^{T}$ of size $T$ is available. Here, $\mathbf{x}_\ell^{(i)}$ denotes the $i$-th example (/dataset entry) of the received signal at the $\ell$-th receiver for an emitting source located at $\mathbf{p}^{(i)}$. We assume the existence of a statistical model $\mathbb{P}(\mathbf{x}_1, \ldots, \mathbf{x}_L; \mathbf{p})$, encapsulating the relation between $\mathbf{p}$ and $(\mathbf{x}_1, \ldots, \mathbf{x}_L)$. The assumption regarding the availability of such a dataset has become relatively common in ongoing efforts for such a solution approach (e.g., [13]), and recent technological advancements (e.g., [14]) suggest that it is soon to be an even more realistic one.

Assuming that the data, or possibly some compressed version thereof, from all receivers can be reliably transmitted to a single central terminal, our goal is to learn an approximating function, based on the available dataset $\mathcal{D}_{\text{train}}^{(T)}$, for an estimator of $\mathbf{p}$, denoted as $\widehat{\mathbf{p}}$, given the received signals $\{\mathbf{x}_\ell\}_{\ell=1}^{L}$, so as to minimize the risk $\mathbb{E}\left[\mathcal{L}(\widehat{\mathbf{p}}, \mathbf{p})\right]$ for a loss function $\mathcal{L} : \mathcal{V} \times \mathcal{V} \to \mathbb{R}_+$, where the expectation is with respect to all the randomness[1] in $\{\mathbf{x}_\ell\}_{\ell=1}^{L}$. In this work, we consider $\mathcal{L}(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|_2^2$, giving the mean-square error risk.

## III. DNN-based Localization

In a recent work [12], an upper bound on the localization MSE under a general model mismatch was derived. The simulation therein shows that the neural architecture for $3D$ UAL proposed in [11], termed "CNN-DLOC" and illustrated in Fig. 1, is stable with respect to "small" variations in the environment-characterizing physical parameters (e.g., sound speed profile, see [12] for details and the full description).

The CNN-DLOC architecture in Fig. 1, whose basis is convolutional layers, is trained in two phases. In the first phase, three "sub-models" are trained to *separately* estimate range, azimuth and inclination ($r, \theta, \phi$, respectively), with custom, specifically matched loss functions.[2] In the second phase, these models are integrated into a single model, which is then retrained to *jointly* estimate the three coordinates, namely the source position, so as to capture the statistical dependencies among them. The superior performance of this solution over relevant benchmarks has been demonstrated in a number of scenarios with different propagation models [11], [12].

While it is not the focus of this work, we briefly demonstrate the strength of DNN-based localization via the results of the in Fig. 2, which presents the localization root mean-square error (RMSE) vs. the signal-to-noise ratio (SNR). In this experiment, a small seamount and an undulating surface were considered, and the channel between each of $L = 4$ source-receiver pairs was simulated via the Bellhop simulator [16]. As seen, already at 0dB CNN-DLOC outperforms SBL [4] and the "generalized cross-correlation with phase transform" (GCC-PHAT) [17], [18] that is known for its resilience to multipath.

## IV. Robustness Analysis

We now compliment the analysis in [12] of the neural architecture from [11] for $3D$ UAL via an extensive, diverse empirical study. Specifically, we examine its robustness to different types of variations in the test set inputs caused by deviations from the training set's statistical model. To facilitate this, we introduce a general signal model that will enable us to explore various types of such deviations.

### A. A General Signal Model

In the underwater environment, the water surface and bottom, as well as other potential floating objects, serve as"reflectors" of the source waveform, and give rise to an intricate multipath channel, which can be approximately described using ray propagation. In such a model, the acoustic wave that is emitted from the source and measured at the the receiver is represented as a (possibly infinite) sum of attenuated and delays versions of the signal, which correspond to different rays, where each is propagating in a different path determined by the physical properties of the environment (e.g., sound speed, bathymetry, etc.). For simplicity, we assume a set of environmental parameters, denoted by $\mathcal{P}_{\text{env}}$, that fully characterize the propagation model. Additionally, we hereafter assume that source is located sufficiently far from all $L$ receivers to permit a planar wavefront ("far-field") approximation.

Specifically, we consider the general signal model

$$\mathsf{x}_\ell[n] = \sum_{r=1}^{R} b_{r\ell}\mathsf{s}_{r\ell}[n] + \mathsf{w}_\ell[n] \triangleq \mathbf{s}_\ell^{\mathsf{T}}[n]\mathbf{b}_\ell + \mathsf{w}_\ell[n] \in \mathbb{C}, \quad n = 1, \ldots, N,\ \forall \ell \in \{1, \ldots, L\}, \tag{1}$$

for the received baseband signal of the $\ell$-th receiver, where we have defined $\mathbf{s}_\ell[n] = [\mathsf{s}_{1\ell}[n] \cdots \mathsf{s}_{R\ell}[n]]^{\mathsf{T}} \in \mathbb{C}^R$ and $\mathbf{b}_\ell = [b_{1\ell} \cdots b_{R\ell}]^{\mathsf{T}} \in \mathbb{C}^R$, using the notation:

1) $b_{r\ell} \in \mathbb{C}$ as the unknown attenuation coefficient[3] from the source to the $\ell$-th sensor associated with the $r$-th signal component (line-of-sight (LOS) or non-LOS reflections);
2) $\mathsf{s}_{r\ell}[n] \triangleq \mathsf{s}\left(t - \tau_{r\ell}(\mathbf{p}, \mathcal{P}_{\text{env}})\right)|_{t=nT_s} \in \mathbb{C}$ as the sampled $r$-th component of the unknown (possibly random) source waveform at the $\ell$-th sensor, where $\mathsf{s}\left(t - \tau_{r\ell}(\mathbf{p}, \mathcal{P}_{\text{env}})\right)$ is the analog, continuous-time waveform delayed by $\tau_{r\ell}(\mathbf{p}, \mathcal{P}_{\text{env}})$, and $T_s$ is the sampling period; and
3) $\mathsf{w}_\ell[n] \in \mathbb{C}$ as the overall additive, ambient and internal receiver, noise at the $\ell$-th receiver.

Applying the normalized DFT[4] to (1) yields the equivalent frequency-domain representation for all $\ell \in \{1, \ldots, L\}$,

[1] To be defined explicitly in Section IV-A; see the details below (1).

[2] We use the mean cyclic error (e.g., [15]) for the angles, which are, of course, periodic quantities by definition.

[3] These coefficients are actually also a function of the source's position $\mathbf{p}$ and the environment $\mathcal{P}_{\text{env}}$, but we use $b_{r\ell}$ rather than $b_{r\ell}(\mathbf{p}, \mathcal{P}_{\text{env}})$ for brevity.

[4] $\bar{\mathbf{z}}$ denotes the normalized discrete Fourier transform (DFT) of $\mathbf{z}$.

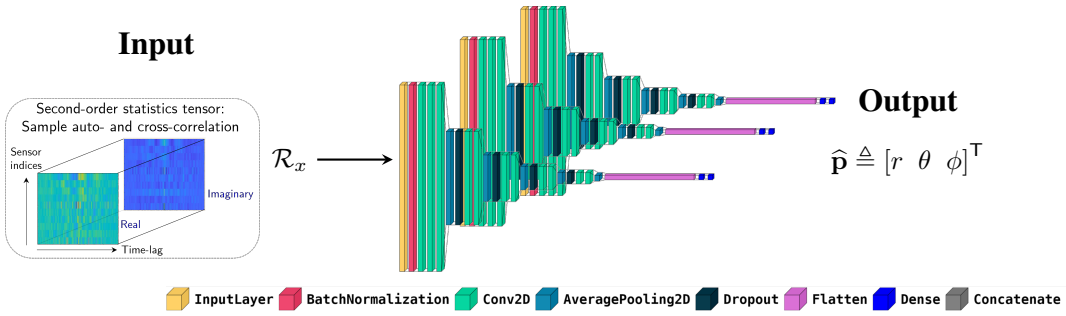
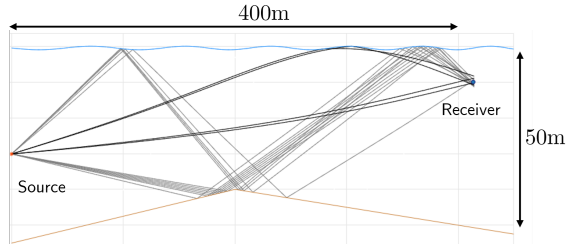
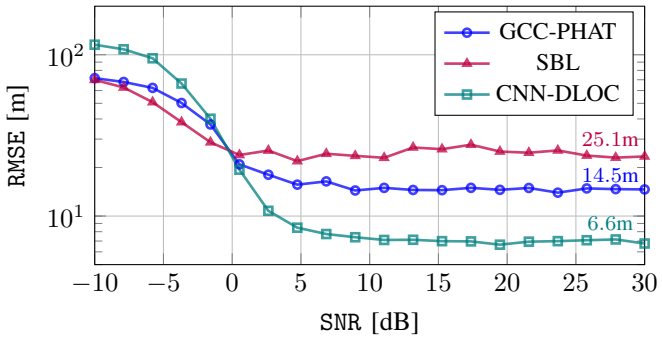



Fig. 1: Illustration of the CNN-DLOC architecture for $3D$ UAL [11]. Input: second-order statistic tensor (all the empirical auto- and cross-correlation functions between signals from all sensors), Output: position in spherical coordinates. The three sub-model (for range, azimuth and inclination) are first trained individually, and then trained jointly, with connecting layers and a customized loss function.




(a) Illustration of the simulated ray propagation model using Bellhop.




(b) Localization RMSE vs. SNR for the three different localization methods.

Fig. 2: CNN-DLOC outperforms benchmark methods for nontrivial propagation model in a simulated environment (with a small seamount), thus demonstrating the potential of data-driven methods for UAL.

$$\bar{\mathsf{x}}_\ell[k] = \sum_{r=1}^{R} b_{r\ell}\bar{\mathsf{s}}[k]e^{-\jmath\omega_k\tau_{r\ell}(\mathbf{p},\mathcal{P}_{\text{env}})} + \bar{\mathsf{w}}_\ell[k]$$
$$= \bar{\mathsf{s}}[k]\cdot \underbrace{\mathbf{d}_\ell^{\mathsf{H}}[k]\mathbf{b}_\ell}_{\substack{\text{a function of}\\ \mathbf{p} \text{ and } \mathcal{P}_{\text{env}}}} + \bar{\mathsf{w}}_\ell[k] = \bar{\mathsf{s}}[k]\cdot \underbrace{\bar{h}_\ell[k]}_{\substack{k\text{-th frequency}\\ \text{response}}} + \bar{\mathsf{w}}_\ell[k] \in \mathbb{C}, \tag{2}$$

where we have defined $\bar{h}_\ell[k] \triangleq \mathbf{d}_\ell^{\mathsf{H}}[k]\mathbf{b}_\ell \in \mathbb{C}$, and

$$\mathbf{d}_\ell[k] \triangleq [e^{-\jmath\omega_k\tau_{1\ell}(\mathbf{p},\mathcal{P}_{\text{env}})}\cdots e^{-\jmath\omega_k\tau_{R\ell}(\mathbf{p},\mathcal{P}_{\text{env}})}]^{\mathsf{H}} \in \mathbb{C}^R, \tag{3}$$

with $\{\omega_k \triangleq \frac{2\pi(k-1)}{NT_s}\}_{k=1}^N$. As shorthand, we further define

$$\bar{\mathbf{x}}_\ell \triangleq [\bar{x}_\ell[1]\cdots\bar{\mathsf{x}}_\ell[N]]^{\mathsf{T}}, \bar{\mathbf{w}}_\ell \triangleq [\bar{\mathsf{w}}_\ell[1]\cdots\bar{\mathsf{w}}_\ell[N]]^{\mathsf{T}},$$
$$\mathbf{D}_\ell \triangleq [\mathbf{d}_\ell[1]\cdots\mathbf{d}_\ell[N]]^{\mathsf{H}}, \mathbf{H}_\ell(\mathbf{p},\mathcal{P}_{\text{env}}) \triangleq \text{Diag}\left(\mathbf{D}_\ell\mathbf{b}_\ell\right),$$

as well as $\bar{\mathbf{s}} \triangleq [\bar{\mathsf{s}}[1]\cdots\bar{\mathsf{s}}[N]]^{\mathsf{T}}$, where $\text{Diag}(\cdot)$ forms a diagonal matrix from its vector argument. Note that $\mathbf{H}_\ell(\mathbf{p},\mathcal{P}_{\text{env}})$ is a nonlinear function of the unknown source position $\mathbf{p}$ and the environmental parameters $\mathcal{P}_{\text{env}}$, as seen from (3). Indeed, $\{\tau_{r\ell}(\mathbf{p},\mathcal{P}_{\text{env}})\}$ are generally nonlinear (nonidentical) functions of $\mathbf{p}$. With this notation, we may write (2) compactly as

$$\bar{\mathbf{x}}_\ell = \mathbf{H}_\ell(\mathbf{p},\mathcal{P}_{\text{env}})\bar{\mathbf{s}} + \bar{\mathbf{w}}_\ell \in \mathbb{C}^N, \;\; \forall \ell \in \{1,\ldots,L\}. \tag{4}$$

Assuming a statistical relation $\mathbb{P}(\bar{\mathbf{x}}_1,\ldots,\bar{\mathbf{x}}_L;\mathbf{p})$ between $\mathbf{p}$ and $\{\bar{\mathbf{x}}_\ell\}$, which depends (nonlinearly) on $\{\mathbf{H}_\ell(\mathbf{p},\mathcal{P}_{\text{env}})\}$, the sheer complexity of the UAL problem, which consists of estimating $\mathbf{p}$ from $\{\bar{\mathbf{x}}_\ell\}_{\ell=1}^L$, is now reflected from (4).

### *B. Robustness of the Oracle Least-Squares Estimator*

We shall set our expectations for the best achievable robustness of a DNN-based localization solution relative to an oracle solution, which has access to precise knowledge of the environment and its exact analytical characterization $\mathcal{P}_{\text{env}}$. Under relatively mild technical assumptions, the oracle least-squares (LS) estimator of $\mathbf{p}$, namely,

$$\widehat{\mathbf{p}}_{\text{OLS}} \triangleq \arg\min_{\mathbf{p}\in\mathcal{V}} \left\{ \min_{\substack{\bar{\mathbf{s}}\in\mathcal{B}\\ \{\mathbf{b}_\ell\in\mathbb{C}^R\}}} \sum_{\ell=1}^{L} \|\bar{\mathbf{x}}_\ell - \mathbf{H}_\ell(\mathbf{p},\mathcal{P}_{\text{env}})\bar{\mathbf{s}}\|_2^2 \right\}, \tag{5}$$

where $\mathcal{B} \triangleq \{\mathbf{x}\in\mathbb{C}^N : \|\mathbf{x}\|_2 = 1\}$, is given by [11]

$$\widehat{\mathbf{p}}_{\text{OLS}} = \arg\max_{\mathbf{p}\in\mathcal{V}} \lambda_{\max}\left(\mathbf{Q}(\mathbf{p},\mathcal{P}_{\text{env}})\right), \tag{6}$$

where $\lambda_{\max}\left(\mathbf{Q}(\mathbf{p},\mathcal{P}_{\text{env}})\right)$ denotes the largest eigenvalue of

$$\mathbf{Q}(\mathbf{p},\mathcal{P}_{\text{env}}) \triangleq \sum_{\ell=1}^{L} \bar{\mathbf{X}}_\ell\mathbf{D}_\ell^*\left(\mathbf{D}_\ell^{\mathsf{T}}\mathbf{D}_\ell^*\right)^{-1}\left(\bar{\mathbf{X}}_\ell\mathbf{D}_\ell^*\right)^{\mathsf{H}} \in \mathbb{C}^{N\times N} \tag{7}$$

is a data–dependent matrix, where $\{\bar{\mathbf{X}}_\ell \triangleq \text{Diag}(\bar{\mathbf{x}}_\ell)\}_{\ell=1}^L$.

The estimator (6), whose derivation is given in [11], is consistent for an arbitrary unknown waveform $\bar{\mathbf{s}}$ and an arbitrary set of attention coefficients $\{b_{r\ell}\}$. We thus conclude that, at least theoretically, and at least for the general signal model (4), there exist a robust estimator that is oblivious (in the consistency sense) to the source waveform structure and the attenuation coefficients. While this theoretical property may not be surprising, it is nevertheless nontrivial for a data-driven, *learned* solution to possess it. We next demonstrate that CNN-DLOC indeed reflects such desirable robustness.

### *C. Robustness to Source Waveform Statistical Characteristics*

We consider a scenario with $L = 4$ receivers, whose locations are given in Table I. We use the Bellhop simulator

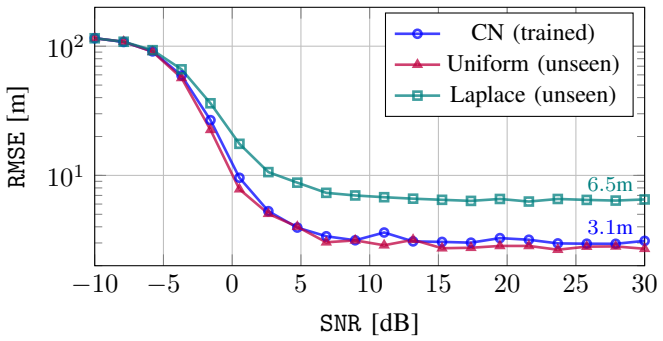



Fig. 3: $3D$ Localization RMSE vs. SNR for three different source distributions: CN (mesokurtic), Laplace (leptokurtic) and uniform (platykurtic). The trained model demonstrates resilience to the source's waveform distribution.

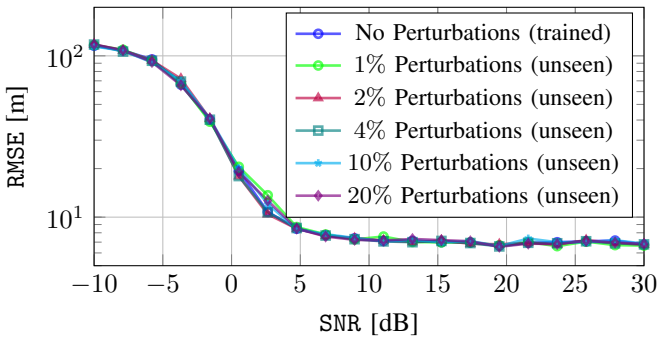



Fig. 4: $3D$ Localization RMSE vs. SNR for CNN-DLOC with different levels of perturbation in the attenuation coefficients $\{b_{r\ell}\}$.

| | $x$ [m] | $y$ [m] | $z$ [m] |
|---|---|---|---|
| **Receiver 1** | 150 | −250 | 10 |
| **Receiver 2** | 50 | −250 | 15 |
| **Receiver 3** | −50 | −250 | 20 |
| **Receiver 4** | −150 | −250 | 25 |

TABLE I: Positions of the four receivers in Cartesian coordinates.

to generate $\{\mathbf{H}_\ell(\mathbf{p}, \mathcal{P}_{\text{env}})\}$, where in each dataset entry, $\mathbf{p}$ is drawn uniformly and independently from

$$\mathcal{V} = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 \;\middle|\; \begin{array}{c} x \in (-150, 150) \text{ [m]} \\ y \in (-100, 0) \text{ [m]} \\ z \in (5, 45) \text{ [m]} \end{array} \right\} \tag{8}$$

in Cartesian coordinates, and $\mathcal{P}_{\text{env}}$ was set and fixed. For example, $\mathcal{P}_{\text{env}}$ includes the speed of sound profile, which, unlike the usual isovelocity assumption, is set to be varying as a function of depth in this case. The full description of $\mathcal{P}_{\text{env}}$ is given in [19]. The attenuation coefficients were drawn independently from the complex normal (CN) distribution, such that $\mathbb{E}\left[|b_{rl}|^2\right] = 1$, with variance $0.1^2$, and we use $N = 100$. All results are based on averaging $2 \cdot 10^5$ independent trials.

We test our trained model with random waveforms of three types of distributions: mesokurtic, leptokurtic and platykurtic. For this, we choose the CN, Laplace and uniform distributions. Figure 3 shows the RMSE vs. the SNR for the three types of waveform distributions. It is seen that the model exhibits a similar trend, converging to asymptotic RMSE values, which are different by only a few ($\sim$3) meters. As expected, the heavier-tailed Laplace-distributed waveform results in more frequent outliers, leading to inevitable degradation. Still, CNN-DLOC exhibits considerable robustness to variations in $\bar{\mathbf{s}}$.

### D. Robustness to Attenuation-Affecting Physical Variations

We next test the robustness to all $\{b_{r\ell}\}$. For this, consider five different test sets, where in each test set we introduce random perturbations to the calculated attenuation coefficients computed by Bellhop according to the underlying physical model. Specifically, we add a zero-mean uniform random variable whose support is set as a fixed factor of the average $\frac{1}{R}\sum_{\ell=1}^{L}|b_{r\ell}|$ for the $\ell$-th receiver.[5] For a more comprehensive inspection that broadens the variety of considered scenarios, we change some of the physical characteristics of the simulated environment in the set $\mathcal{P}_{\text{env}}$. In particular, we consider a nonflat bathymetry with a small seamount, as illustrated in Fig. 2a. The detailed parametric environmental description is omitted due to the limited space. Apart from these changes, the setting is identical to the one in Section IV-C, and the source waveform is standard circular CN white noise.

As seen from Fig. 4, CNN-DLOC exhibits excellent robustness to variations of this physical aspect as well. These results suggest that the architecture has the capacity to learn an estimator that—on top of learning the environmental features relevant for localization—similarly to the oracle LS estimator (5), is oblivious to physical variations that lead to perturbations in the attenuation coefficients $\{b_{r\ell}\}$.

### E. Robustness to Data Compression Techniques

We now test CNN-DLOC against a more challenging input statistical variation, and go beyond what can be expected from the oracle LS estimator (5). Our goal now is to examine the model's OOD generalization to a different statistical deviation, and specifically its behavior when applied to *compressed* versions of the data. This scenario is relevant since the single-sensor receivers are non-collocated, and therefore need to transmit the data to a central processing unit. Given that underwater acoustic channels typically support only low-bandwidth communication [20], data compression becomes essential.

The simulation setting in this section is identical to that in Section IV-C with a standard circular CN white noise source waveform. However, we now generate the second-order statistics input tensor (as illustrated in Fig. 1) after applying the following compression methods to $\{\mathbf{x}_\ell\}$:

- **Max-comp**: Motivated by [21], we zero out all the samples in each signal $\mathbf{x}_\ell$ except for the $K_{\max}$ samples with the maximal magnitudes. We evaluate the performance with this compression method for $K_{\max} \in \{1, 3\}$;
- **Onebit**: We apply the widely adopted one-bit per sample quantization (e.g., [22], [23]) to the signals $\{\mathbf{x}_\ell\}_{\ell=1}^{L}$; and
- **All-but-one Onebit**: The same as Onebit above, but only for $(L-1)$ signals out of the $L$ signals $\{\mathbf{x}_\ell\}_{\ell=1}^{L}$. This simulates the setting in which the central computing unit is collocated with one of the receivers, and all the other $(L-1)$ receivers must compress their data before transmitting it to the central computing unit.

[5] For example, a factor of $1/5$ in our terminology is "20% perturbations".

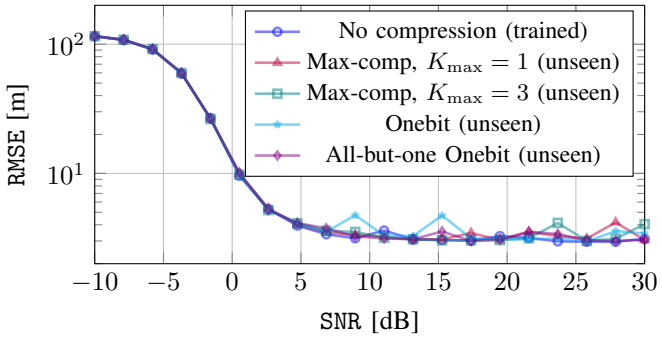



Fig. 5: $3D$ Localization RMSE vs. SNR. Evidently, CNN-DLOC is (approximately) oblivious to these forms of data compression methods.

Interestingly, and perhaps surprisingly, Fig. 5 shows that the model, which was trained for uncompressed input data, delivers almost the same performance for all four types of compression, which are approximately the same performance we obtain without compression. On one hand, this may suggest that the model could be improved for uncompressed signals.[6] On the other hand, this result shows that the model is highly robust to various types of quantization forms, which is a desirable (nontrivial) property in itself for a localization method. It follows, therefore, that the model is extracting the most relevant features for localization, which are somewhat preserved under these types of quantization methods.

We argue, without theoretical justification (in the present work), that a primary contributing factor to the inherent robustness of the solution is its specific input structure. In particular, by preprocessing the raw data and feeding the empirical auto- and cross-correlation functions of all the pairs of sensors, we—in some sense—"Gaussianize" the input, from a central limit theorem argument. Intuitively (and informally), this "Gaussianizing transformation" takes different input distributions (due to source waveform, perturbations in attenuation, compression methods) and, on top of highlighting features like time-difference-of-arrival, making them more similar to the input distribution of the training data.

## V. Concluding Remarks and Future Work

This work investigates the robustness and OOD generalization of the recently proposed DNN-based solution [11] for UAL. Through various types of relevant deviations from the input training distribution, we present a diverse simulation study that demonstrates the robustness of a DNN trained for UAL in complex environments. This further emphasizes the potential integration of DNNs in future UAL systems.

While our results constitute an important milestone in the development of data-driven UAL solutions, a plethora of yet unexplored aspects await to be addressed. Among them are natural extensions of this work, such as OOD generalization to localization outside the volume $\mathcal{V}$ for which the model was trained, perturbations in the bathymetry, sound speed profile and surface geometry, along with an accompanying theory to support the successful operation under these conditions.

[6]But this is not guaranteed: it is theoretically possible that a lossy compression optimized for distributed localization leads to negligible loss of accuracy.

## References


[1] E. R. Potokar, K. Norman, and J. G. Mangelson, "Invariant extended Kalman filtering for underwater navigation," *IEEE Robot. Autom. Lett.*, vol. 6, no. 3, pp. 5792–5799, 2021.

[2] V. P. Dubey, J. Saha, S. Bhaumik, and A. Dey, "Tracking an underwater target in a large surveillance region with sensor location uncertainty," *IEEE Access*, vol. 11, pp. 140007–140021, 2023.

[3] X. Su, I. Ullah, X. Liu, and D. Choi, "A review of underwater localization techniques, algorithms, and challenges," *J. Sens.*, vol. 2020, no. 1, pp. 1–24, 2020.

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

[5] L. Houégnigan, P. Safari, C. Nadeu, M. André, and M. van der Schaar, "Machine and deep learning approaches to localization and range estimation of underwater acoustic sources," in *2017 IEEE/OES Acoustics in Underwater Geosciences Symposium (RIO Acoustics)*, 2017, pp. 1–6.

[6] Y. Wang and H. Peng, "Underwater acoustic source localization using generalized regression neural network," *J. Acoust. Soc. Am.*, vol. 143, no. 4, pp. 2321–2331, 2018.

[7] Z. Huang, J. Xu, Z. Gong, H. Wang, and Y. Yan, "Source localization using deep neural networks in a shallow water environment," *J. Acoust. Soc. Am.*, vol. 143, no. 5, pp. 2922–2932, 2018.

[8] A. Testolin and R. Diamant, "Underwater acoustic detection and localization with a convolutional denoising autoencoder," in *Proc. of CAMSAP*, 2019, pp. 281–285.

[9] S. Sun, T. Liu, Y. Wang, G. Zhang, K. Liu, and Y. Wang, "High-rate underwater acoustic localization based on the decision tree," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2021.

[10] H. Ye, C. Xie, T. Cai, R. Li, Z. Li, and L. Wang, "Towards a theoretical framework of out-of-distribution generalization," in *Proc. Int. Conf. Neural Inf. Process. Syst*, 2021, vol. 34, pp. 23519–23531.

[11] A. Weiss, T. Arikan, and G. W. Wornell, "Direct localization in underwater acoustics via convolutional neural networks: A data-driven approach," in *Proc. IEEE 32nd Int. Workshop Mach. Learn. Signal Process. (MLSP)*, 2022, pp. 1–6.

[12] A. Weiss, A. C. Singer, and G. W. Wornell, "Towards robust data-driven underwater acoustic localization: A deep cnn solution with performance guarantees for model mismatch," in *Proc. of ICASSP*, 2023, pp. 1–5.

[13] R. Chen and H. Schmidt, "Model-based convolutional neural network approach to underwater source-range estimation," *J. Acoust. Soc. Am.*, vol. 149, no. 1, pp. 405–420, 2021.

[14] D. P. Zitterbart, A. Bocconcelli, M. Ochs, and J. Bonnel, "TOSSIT: A low-cost, hand deployable, rope-less and acoustically silent mooring for underwater passive acoustic monitoring," *HardwareX*, vol. 11, pp. e00304, 2022.

[15] S. Basu and Y. Bresler, "A global lower bound on parameter estimation error with periodic distortion functions," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 1145–1150, 2000.

[16] M. B. Porter, "The BELLHOP manual and user's guide: Preliminary draft," *Heat, Light, and Sound Research, Inc., La Jolla, CA, USA, Tech. Rep*, vol. 260, 2011.

[17] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, 1976.

[18] M. S. Brandstein and H. F. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," in *Proc. of ICASSP*, 1997, vol. 1, pp. 375–378.

[19] A. Weiss, A. C. Singer, and G. W. Wornell, "Towards robust data-driven underwater acoustic localization: A deep CNN solution with performance guarantees for model mismatch, supplementary materials," https://www.weissamir.com/project/DLOC, 2022.

[20] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 84–89, 2009.

[21] A. Weiss, Y. Kochman, and G. W. Wornell, "A joint data compression and time-delay estimation distributed systems via extremum encoding," in *Proc. of ICASSP*, 2024, pp. 9366–9370.

[22] X. Huang and B. Liao, "One-bit MUSIC," *IEEE Signal Process. Lett.*, vol. 26, no. 7, pp. 961–965, 2019.

[23] A. Weiss and G. W. Wornell, "One-bit direct position determination of narrowband Gaussian signals," in *Proc. IEEE Workshop Stat. Signal Process. (SSP)*, 2021, pp. 466–470.