Combining Denoising Autoencoders and Dynamic Programming for Acoustic Detection and Tracking of Underwater Moving Targets

Alberto Testolin

Department of Information Engineering, University of Padova, Via Gradenigo 6/B, Padova 35141, Italy
alberto.testolin@unipd.it

Roee Diamant

Hatter Department of Marine Technologies, University of Haifa, Israel
roee.d@univ.haifa.ac.il

Abstract

Accurate detection and tracking of moving targets in underwater environments pose significant challenges, because noise in acoustic measurements (e.g., SONAR) makes the signal highly stochastic. In continuous marine monitoring a further challenge is related to the computational complexity of the signal processing pipeline: due to energy constraints, in off-shore monitoring platforms algorithms should operate in real time with limited power consumption. In this paper we present an innovative method that allows to accurately detect and track underwater moving targets from the reflections of an active acoustic emitter. Our system is based on a computationally- and energy-efficient pre-processing stage carried out using a deep convolutional denoising autoencoder (CDA), whose output is then fed to a probabilistic tracking method based on the Viterbi algorithm. The CDA is trained on a large database of more than 20,000 reflection patterns collected during

50 designated sea experiments. System performance is then evaluated on a controlled dataset, for which ground truth information is known, as well as on recordings collected during different sea experiments. Results show that, compared to the benchmark, our method achieves a favorable trade-off between detection and false alarm rate, as well as improved tracking accuracy.

Keywords: Underwater signal detection; Deep learning; Viterbi algorithm; Marine monitoring; Acoustic detection; SONAR; Track before detect.

1. Introduction

Underwater detection and localization of moving targets is a key enabling technology for both ecological and security applications. Marine ecology research and well as fishery legislation decisions rely heavily on abundance indications: in this context, the ability to remotely identify marine mammals, pelagic species, or other animals like sea turtles is a game changer in environmental research and management, where abundance is mostly derived from biased, fishery-dependent data [1]. Since acoustic waves propagate well underwater and are known by standards to be safe for marine animals [2], harnessing active SONAR technology that detects targets by acoustic reflections shows great potential to obtain reliable, fishery-independent biomass aggregation during ecological surveys. For example, fish biomass and size spectra can be quantified using directional acoustic methods (echo sounders), which may reflect also perturbation of the entire ecosystem [3]. Active SONAR technology is also already widespread for military applications such as detection of submerged vessels and scuba divers [4]. However, due to the low reflection signature of such targets, the signal-to-clutter ratio (SCR) is usu-

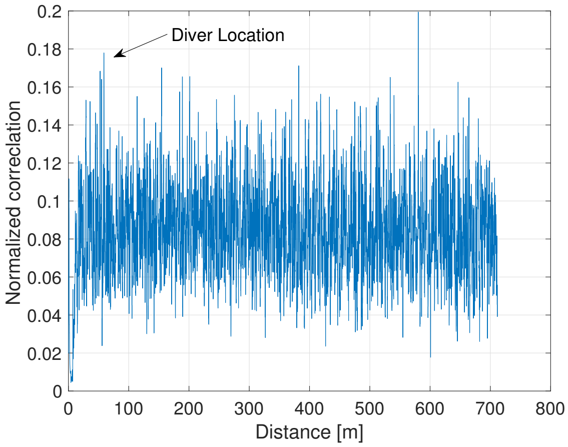


Figure 1: A single reflection pattern from a scuba diver with a low reflection closed-circuit re-breather.

ally very small, and progress in the detection of submerged mobile targets through active acoustics is thus still a major challenge.

In this paper, we describe a general detection and tracking framework based on a single omni-directional acoustic transceiver, which transmits and receives over a wideband frequency band. This allows flexibility in deployment as well as energy efficiency, such that long-term detection efforts can by made even from small buoys. Analyzing the reflections obtained from the single omni-directional receiver, the main challenge considered here is to detect the target's-based reflection within the clutter noise. The latter includes stationary reflections (e.g., from rocks or chains) as well as reflections from waves or volume scatters. An example of a reflection response from a scuba diver is shown in Fig. 1: note that the diver's reflection is almost invisible within the clutter.

To detect targets in high clutter, the track-before-detect (TBD) approach has been widely adopted [5]. This approach aims to increase the clutter-to-noise ratio by performing detection over a sequence of observations. The method applies tracking by maximum-likelihood probabilistic data association (ML-PDA) [6], filtering [7], dynamic programming tracking by Markov chain representation [8], and probabilistic multi-hypothesis tracking [9]. Yet, tracking assumes an underline dynamics for the tracked target [10], which may be hard to model for the case of marine animals whose motion tends to be of rapid orientation changes. Considering this, we have recently introduced a probabilistic approach for the case of tracking a single target [11], which allows to detect the target's reflections within the clutter by using the Viterbi algorithm to identify structured patterns within a time-distance (TD) matrix formed by concatenating matched filter's outputs sequentially.

However, the main limitation of TBD approaches is the time consuming analysis of sometimes tens of thousands of samples for each reverberation response. Moreover, besides issues of computational complexity the above methods are mostly applied for the detection of a single target, while general solutions should also support efficient detection of multiple targets (see [12, 13] for recent approaches tackling this issue). A promising emerging technology to cope with these limitations is represented by *deep learning* [14], which has recently achieved state-of-the-art performance in a variety of difficult pattern recognition tasks, ranging from image classification [15] to speech recognition [16, 17], without requiring domain-specific expert knowledge about the signal characteristics.

In this work we describe a novel method, referred to as *CDA-TBD*, that combines deep learning with dynamic programming: the former is used as an efficient denoising filter, which removes clutter and highlight target-based reflections in real-time from the time-distance (TD) matrix, while the latter further identifies unique targets and precisely tracks their trajectories. More specifically, we use a convolutional denoising autoencoder (CDA) [18] to highlight potential lines within the TD. Differently from clutter (whose structure is random), such lines likely represent reflections from moving targets. We then apply the forward-backward algorithm, whose states are target's ID and observations are the values provided by the denoised TD matrix. The latter are considered as emission probabilities, while the state transitions are set by limitations over the motion of the tracked target.

A critical aspect that should be considered for improving the performance of deep learning models is the careful definition of a training data set, which should contain a representative sample of the statistical distribution of the target signals that will be observed during system testing. Since the underwater acoustic reverberation channel is difficult to model analytically, both in terms of the channel impulse response and in terms of the target and clutter reflection patterns, in our work we rely on a large set of real measurements collected from a multitude of more than 50 sea experiments. Each experiment includes both verified clutter and target (fish) reflections, which are systematically combined to create a large-scale training set. To evaluate the generalization capability of the proposed system, we then test it on separate sea experiments carried out with different moving targets (i.e., scuba divers).

To the best of our knowledge, our CDA-TBD approach constitutes the first attempt to combine deep learning with dynamic programming for identifying targets within a reflected acoustic signal. Our contribution is thus threefold:

1. Develop a convolutional denosing autoencoder architecture for the detection of curved lines within a reflection (TD) image.
2. Implement a computationally efficient method that combines deep learning pre-processing with a probabilistic algorithm applied over a track-before-detect approach.
3. Create a statistically large-enough database containing clutter and reflections of acoustic patterns, which we freely share with the community for reproducibility and further research.

Our results show that even in low SCR, where the reflection pattern from the target is weak, our method yields a favorable trade-off between precision and recall, which exceeds the performance of fully probabilistic approaches (i.e., the Viterbi algorithm) at a much lower computational complexity, and also allows for a more accurate fine-grained tracking of the target path. Further, the results show that our approach easily scales to scenarios featuring multiple targets.

The paper is organized as follows. In Section 2 we discuss the state of the art in probabilistic tracking and ML-based detection. Our system's model and objectives are outlined in Section 3, along with a description of the sea experiments that allowed to create the large-scale dataset of real measurements used for training and testing our system. Our CDA-TBD methodology is explained in Section 4, and system performance is analyzed in Section 5. Conclusions are drawn in Section 6. Preliminary results, which did not include the TBD method and only explored CDA performance on simulated data, have been recently presented as a conference paper [19].

2. Related Work

Detection of targets using active acoustic transmission is nowadays performed by continuous active SONAR (CAS) or by transmission of separated pulses [8]. The former involves multiple narrowband transmissions across the band to detect Doppler components that indicate motion [20]. However, this may induce significant energy consumption and may also harm the bio fauna in the surveyed environment. Further, it may not fit the detection of slowly moving targets. We thus focus on the latter approach. To identify a target within heavy clutter, detection based on a single reflection pattern may fail, and the available literature turns to detection according to a sequence of reverberation patterns. Based on the transmission of consecutive wideband signals of high processing gain such as chirps, single reverberation patterns are analyzed and concatenated to form a TD matrix [21]. This analysis can be performed by a matched filter (MF) [22], a channel equalization such as orthogonal matching pursuit [23], or detectors based on a local estimation of the noise distribution [24]. Once the TD matrix is formed, the detection is performed through tracking.

Tracking over the TD matrix is possible through filtering, for example by exploiting variants of the Kalman filter [10] or using blind tracking to handle non-Gaussian clutter [21]. Alternatively, clutter could be classified using a mixture of distributions [25], such that detection is matched to local clutter patterns within the reflected pattern. A more common approach uses tracking by a particle filter that learns the clutter's and target's probabilistic model though statistically sampling the discrete grid of the state-space. This can also be applied for multi-target scenario of active SONAR [7]. Yet, such filtering directs the solution by a most probable grid search and may thus fail to detect targets in high clutter.

For cases of low SCR, the TBD approach has proved useful (a comparison between TBD approaches can be found in [26, 5]). Instead of detection per-transmission or by relying on a motion model, tracking is performed probabilistically. A common TBD approach is maximum-likelihood probabilistic data association (ML-PDA), which applies a likelihood ratio test to parameters observed from the TD matrix [27], and shows good tracking in low SCRs with applications of SONAR detection [6, 28]. Alternatively, TBD can use Bayesian tracking, where dynamic programming is applied on a hidden Markov model, as shown in [8] and in our recent work [11]. Another approach is probabilistic multi-hypothesis tracking (PMHT), which tracks possible targets by explicitly separating target and clutter components [29]. The PMHT approach can be extended to combine features such as the intensity distribution of observations [5, 30] or the spatial information obtained by arrays of acoustic receivers [31]. This method is also flexible enough to handle fluctuations of the clutter and target distribution within the TD matrix [32], and [33] even offered an indicative metric to determine the conditions in which tracking is feasible. Yet, while TBD approaches often achieve good results, their main disadvantages are the sensitivity to different targets dynamic types (especially those unknown a-priori) and the high algorithmic complexity, which prevents their application in realistic scenarios (where the TD matrix might contain hundreds of thousands of elements).

These limitations call for the adoption of innovative computational methods. A promising approach is offered by machine learning, which allows to effectively recognize recurring patterns in high-dimensional data by extracting high-order statistical features from a set of training examples. In particular, deep learning methods are particularly effective in pattern recognition tasks where domain knowledge is limited, because they can automatically learn intricate statistical structure from the data by exploiting multiple levels of representation [34]. Furthermore, once trained deep networks are computationally very efficient, since signal processing can be carried out in parallel hardware using basic algebraic operations [35, 36].

Deep learning is being widely used in many engineering fields, ranging from compressed sensing [37] and telecommunications [38] to fault diagnosis [39] and video surveillance [40]. Deep learning detection methods achieve impressive performance even when the signal is corrupted by high levels of noise [41], suggesting it can be successfully applied also in underwater monitoring. Preliminary work exploiting deep learning for the analysis of passive SONAR has been recently proposed [42, 43], highlighting the superiority of deep learning over traditional methods based on Mel frequency cepstral coefficients and Hilbert-Huang transform [44].

Considering the surveyed literature, we can identify three main gaps. The first is the applicability of modern deep learning frameworks for the task of efficiently detecting moving targets within a TD matrix. Specifically, we are interested in exploring the denoising performance of stacked autoencoders when the input image contains high levels of environmental noise. Second, the current statistical and probabilistic approaches cannot provide accurate tracking performance in low complexity when the SCR is low and when the target's characteristics and dynamics are unknown. Third, the application of active SONAR target tracking is currently performed using large arrays, whose deployment is complex and expensive. A much preferred solution would be a single transceiver, preferably of low energy, that can be deployed from small buoys over long periods of time and provide real time detection capability. Confronting these challenges, in the following we present our CDA-TBD approach that combines denoising autoencoders with a TBD solution.

3. Problem Formulation

3.1. System Model

Our CDA-TBD system comprises a single transceiver emitting a sequence of wideband signals of short duration. In our experiments, we use 20 chirp signals of duration 10ms and frequency range 7kHz-17kHz. The signals are spaced by a 0.7s guard interval to suppress reflections from previous emissions, which corresponds to detection for distances of roughly 530m. The transceiver is omni-directional both in transmission and in reception, such that reflections from all directions are received. We make two assumptions on the target: 1) an upper bound, w, on the size of the target, and 2) an upper bound, v, on the speed of the target relative to the receiver. While the first bound can be set loose since the explored area is large, the second bound should fit well the target's expected motion to avoid false detection in low signal-to-clutter (SCR) scenarios.

Without prior knowledge of the target's reflection pattern, we estimate the reflection pattern by the matched filter

$$\mathrm{MF}(\tau, r(t)) = \frac{\int_0^{T_s} s(t) r(t-\tau) dt}{\sqrt{\int_0^{T_s} s^2(t) dt \int_0^{T_s} r^2(t-\tau) dt}} \quad 0 < \tau < T_{\text{guard}} \ , \tag{1}$$

where $s(t), 0 < t < T_s$ and $r(t), 0 < t < T_{\text{guard}}$ are the transmitted signal and received reflections of duration T_s and T_{guard}, respectively. We use a normalized matched filter to provide an initial detection threshold of the direct path based on our previous work [45]. This allows the alignment of each received signal's reflection, without the need for time-synchronization. The aligned reflections are then stored in a TD matrix representing the time and distance for each reflection.

Being a representative of the time-varying reflection pattern, the TD matrix includes reflections of clutter or either clutter or target. Formally, for the ith emission and at distance j, entry (i, j) of the TD matrix is

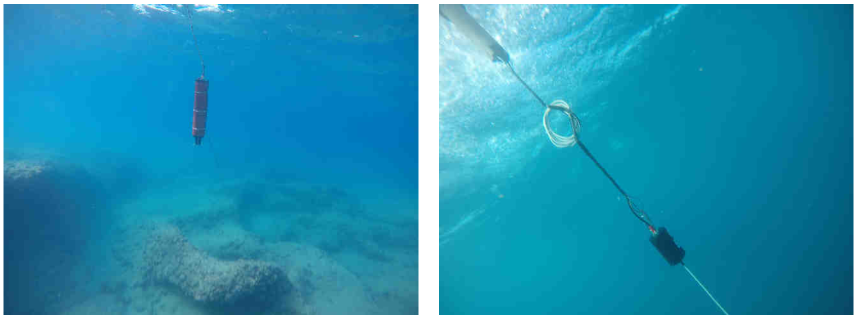
$$\mathrm{TD}(i, j) = \begin{cases} \mathrm{MF}(j, \bar{n}(i)) & \text{clutter} \\ \mathrm{MF}(j, \bar{y}(i)) & \text{target} \end{cases} \ , \tag{2}$$

where $\bar{n}(i), \bar{y}(i)$ are the sampled vector of the clutter and target reflections, respectively, and the ratio $\mathrm{MF}(j, \bar{y}(i))/\mathrm{MF}(j, \bar{n}(i))$ is the SCR. As the expression in (2) hints, a moving target will show as a curved line in the TD matrix, whereas, due to its random nature, clutter will show as random points. We also note that the TD matrix can represent reflections from static targets such as rocks and anchors, which being stationary, will show on the TD matrix as nearly straight lines. In this work we assume these lines are already discarded, e.g., by the process described in [11]. Our goal in this work is thus focused on the identification and tracking of curved lines within the TD matrix.

3.2. Data description

Our deep-learning-based solution requires the availability of a large database of TD matrices, annotated with the corresponding ground truth information about the locations of the target reflections (if any). Producing annotated samples in underwater scenarios is challenging. One option to circumvent this issue could be to train the deep network with a simulative model, and perform data augmentation to generate a large set of annotated images (e.g., [46, 47, 48]). However, we argue that this approach would fail for the considered task mainly because of two factors: 1) The TD matrix is a representation of a time-varying impulse response of the underwater acoustic reverberation channel. This channel is hard to model, especially due to the highly non-linear reflection pattern within the target's body, but also even for a simple clutter reflection from the non-homogeneous sea surface. 2) The creation of the TD matrix based on the normalization in (1) is a non-linear operation that is hard to simulate. For example, in the proximity of a strong reflection the normalization factor would decrease the matched filter result, leading to a shadowing effect that highly depends on the SCR. In light of this, we opted for the creation of a database based on *real* measurements.

To obtain our database we performed more than 50 sea experiments. Each experiment included a single transceiver deployed from a buoy or a small vessel. As shown in Fig. 2, data was obtained from two system configurations: a standalone Subnero M25M acoustic modem, which analyzed the data on the fly emitting signals at frequency range 20kHz-30kHz, and a remotely operated EvoLogics LF acoustic modem that emitted 10ms duration chip signals at the range of 7kHz-17kHz. In both cases, recording of raw acoustic



(a) Standalone system.

(b) Remotely controlled system.

Figure 2: Pictures of the two configurations of transceiver system from two of the sea experiments.

measurements was done in full duplex, allowing capturing of the direct path. Emissions were done at a period of 0.7s, allowing reception of targets located up to roughly 530m. As such, this is a mono-static acoustic system. In each of the 50 experiments, we recorded at least five hours of data. We then analyzed the recordings offline, in order to identify TD matrices of 20 rows (i.e., 20 reflection patterns) including targets, and TD matrices including clutter-only. The identification was based on the sophisticated procedure discussed in [11], which allows accurate tracking of single targets. This information was sufficient for the aim of offline training the CDA. The experiments were conducted in four different sea environments: 1) at the Red Sea near the shores of Eilat, Israel, at water depth of 30m and a seabed including a complex reef environment; 2) at the Mediterranean Sea across the shores of Haifa, Israel, at water depth of 15m with seabed of rocks; 3) at the Mediterranean Sea across the shores of Hedera, Israel, at water depth of 20-10m with seabed of sand; and 4) at the Mediterranean Sea 11km west of the northern shores of

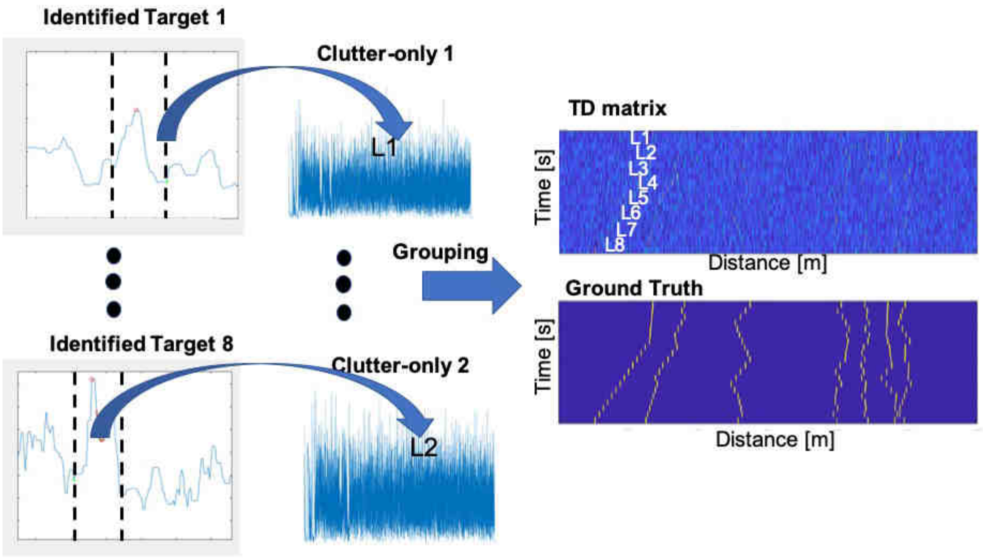


Figure 3: Illustration of the process of data augmentation to create TD matrices. Targets identified by the process described in [11].

Israel, at water depth of 160m with seabed of clay. To verify the reliability of our tagging system, during the experiments we also included targets with verified ground truth information, such as divers dragging buoys with GPS receivers or sharks and tuna fish released after capture for tagging purposes. Further, among others, we identified opportunistic targets such as a dolphin, mackerel, and parrot fish. Our dataset is made freely available through the Open Science Framework[1].

Overall, our experiments yielded roughly 1000 different target-based TD matrices, and more than 20,000 clutter-based TD matrices. To balance the database and improve the generalization ability of the CDA, we implemented a data-mixing approach. More specifically, referring to the block diagram in

[1] https://osf.io/h79mt/

Fig. 3, we augmented our database to create a larger number of 10,000 target-based TD metrics by slicing buffers of normalized matched filter outputs around identified target's reflection, and inserting those over clutter-based TD matrices according to a desired SCR value. Using this methodology, we could also generate different types of reflection lines to reflect various target's dynamics. This was performed by a smoothed *drunken step* of an auto-regression model of the TD matrix column number where the target is inserted. These locations where then served as the ground truth information for the training and testing of the CDA. The result is a balanced database of 20,000 clutter- and target-based TD matrices: an example of such a formed target-based TD matrix is shown in Fig. 3.

4. Detection and Tracking Methodology

4.1. Convolutional Denoising Autoencoder (CDA)

The TD matrix is initially filtered using a deep Convolutional Denoising Autoencoder (CDA), which receives as input the noisy image representing the TD matrix and returns as output a denoised version of the same image (see Fig. 4 for a graphical representation). The denoised matrix is then given as input to the TBD algorithm for further processing (see next section).

The autoencoder is composed by four convolutional layers containing, respectively, 24, 48, 72 and 96 kernels of size 4×4, 6×6, 8×8 and 12×16. In order to reduce the dimensionality of the input, the first, second and third layers are followed by pooling layers, with with pool size 1×2 and stride 1×2. Pooling and stride are applied only column-wise because the TD matrix usually contains few rows but tens of thousands of columns. Con-

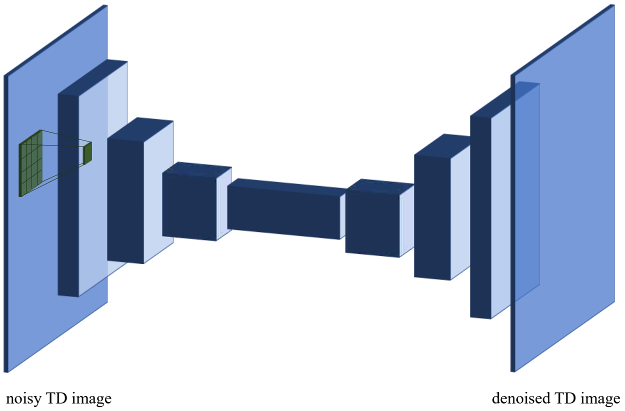


Figure 4: Diagram of the Convolutional Denoising Autoencoder. The noisy TD image is given as input and processed by a stack of convolutional layers, which detect increasingly more complex features in the signal that are then used by the decoder to produce a denoised TD matrix.

volutional (encoding) layers are followed by four deconvolutional (decoding) layers of the same size, which used nearest neighbor as upsampling function. Rectified linear units (ReLUs) are used as activation functions in all layers. A final layer using logistic units is added as a final step to produce output values ranging between zero and one. The CDA architecture and hyperparameters were optimized over a separate validation set using a random search procedure. The CDA was entirely implemented in TensorFlow [49].

To monitor overfitting, the complete data set is split into separate training (60%), validation (20%) and test (20%) sets. The CDA is trained with error backpropagation, using weighted cross-entropy as loss function[2]. Learning

[2] The positive class weight is set to 100 in order to counterbalance the sparsity of the signal (target detections). Extensive simulations showed that the CDA training is robust to variations in this hyperparameter.

occurs over mini-batches of 100 images, and continues until the validation loss starts to increase (early stopping).

4.2. Detection Through Dynamic Programming

In [11] we offered a track-before-detect approach to follow sequences of observations using the Viterbi algorithm constrained to upper bounds on the motion of the target. Specifically, we considered the distance domain of the TD matrix as problem states while the matrix's rows reflected observations. Each entry of the matrix served as an emission indication, while transition probabilities were chosen by an uniform distribution bounded by the maximum states the target can pass between consecutive observations. While this approach yielded acceptable results also in low SCR, it fits the tracking of only a single target. Further, its computational complexity is extremely high. In this work, building on top of the CDA denoised matrix, we modify the above approach to solve both challenges.

4.2.1. Tracking

As illustrated in the block diagram in Fig. 5, we start by setting a threshold, Th over the CDA matrix activation output, $a_{i,j}$, for each transmission/row $i = 0, \dots, T-1$ and each distance slot/column $j = 0, \dots, D-1$. Setting the sigmoid

$$\bar{a}_{i,j} = \begin{cases} 0 & a_{i,j} < \mathrm{Th} \\ 1/\left(1 + e^{-a_{i,j}}\right) & \text{else} \end{cases} , \tag{3}$$

we transform each matrix entry to a measure of probability. This threshold is determined during the CDA training phase and can be set loose since it lies at the beginning of the detection chain.

Next, considering our upper bound on the size of the target's reflecting surface, w, we identify unique line detections. Specifically, denoting c as the sound speed and F_s as the sampling frequency, for each row j, we unify non-zero entries $\bar{a}_{i,j},\ j = 0, \dots, D-1$ that are spaced less than $w/c \cdot F_s$ entries away, to a merged entry whose value is the average of the unified entries while zero forcing its surrounding. The result is a sparse matrix, $\tilde{\boldsymbol{A}}$, of non-zero entries, each reflecting a unique line detection concentrated in one column. On the next step, utilizing our expectation of the target's maximum speed, v, we further compress the smoothed matrix and create a lattice the size of T observations and K possible targets. These targets are identified by vectors $\boldsymbol{t}_k, \boldsymbol{p}_k,\ k = 0, \dots, K-1$, whose entries $t_{i,k}$ and $p_{i,k}$ contains a non-zero value $\tilde{a}_{i,j}$ from $\tilde{\boldsymbol{A}}$ and its location j, respectively, such that $p_{i,k}$ is spaced no more than $v/c \cdot \Delta T F_s$ from location $p_{i-1,k}$, where ΔT is the guard time between each transmission (in our setting 0.7s). The result is a merge of the CDA matrix into filtered identified target lines.

Lattice $\boldsymbol{t}_k$ as much smaller dimension of $K \times T$ compared to the original TD matrix. As a result, we can now apply dynamic programming while still maintaining real-time capability. To that end, we consider the identified targets $k = 0, \dots, K-1$ as the problem states, the values $t_{i,k},\ i = 0, \dots, T-1$ as observations, and the transition probability between targets k, q is set by the average of $\boldsymbol{p}_k$ and $\boldsymbol{p}_q$. Running a dynamic programming like the Viterbi algorithm over the lattice yields the most probable path of a single target. This path reflects the target's line whose probability entries are the highest and their position in the original TD matrix are the most homogeneous such that the least number of *leaks* to other targets is found. Then, more targets

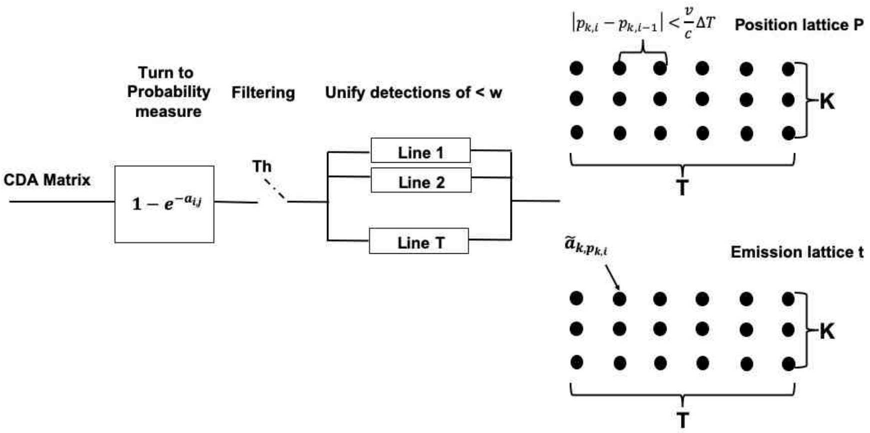


Figure 5: Block diagram for the processing of the CDA matrix before dynamic programming-based tracking.

can be found by discarding the already found targets from the lattice and running the dynamic programming again. Finally, to filter detections, assuming the target should exist throughout most of the observation window T, we only regard targets k whose tracked path by the dynamic programming's solution is stable throughout at least $\rho \cdot T$ of the lattice, where ρ is a user parameter.

4.2.2. Detection

Once the tracking of several targets is achieved, we turn to make a binary detection regarding the existence of a target. Our detection approach compares the likelihood ratio between the elements of the chosen path to non-identified paths, i.e., clutter noise. To that end, for the chosen path $\hat{k}$ and a reference path j, we place a threshold, T_L, over the log-likelihood ratio

$$\mathrm{LLR}_j = \log\left(t_{1,\hat{k}} t_{2,\hat{k}} \ldots \cdot t_{T,\hat{k}}\right) - \log\left(\bar{a}_{1,j} \bar{a}_{j_2,j} \ldots \cdot \bar{a}_{T,j}\right) \ , \tag{4}$$

which compares the likelihood of the chosen path with that of an arbitrary path j across the denoised matrix. Then, identifying a set $\boldsymbol{j} = \{j_1, j_2, \ldots\}$ of arbitrary paths, none of which belong to lattice $\boldsymbol{t}_k,\ k = 0, \ldots, K$, we test

$$\text{Detect} = \begin{cases} 0 & \exists j_m \in \boldsymbol{j} : \text{LLR}_j < T_L \\ 1 & \text{else} \end{cases} . \tag{5}$$

4.3. Computational complexity

Consider the emission of T consecutive signals whose reflections are recorded to yield a $D \times T$ TD matrix. The complexity of a direct track-before-detect run over this matrix using the Viterbi algorithm is $\mathcal{O}(TD^2)$ (cf. [11]) which, since D can be on the order of 10^4 samples, is rather high. Instead, in the CDA-TBD method the denoised TD matrix produces a a lattice of $K \times T$ entrees, where K is the maximum number of possible targets. Regarding the CDA pre-processing, the computational cost of a forward pass through a 2D convolution is $\mathcal{O}(F_I MNmnF_O)$, where F_I and F_O are the number of input and output channels, $M \times N$ is the size of the feature map, and $m \times n$ is the size of the kernel. This bound can be further reduced in the case of deep architectures with square kernels and increasing number of filters [50], as in our case, leading to $\mathcal{O}(pF_I F_O)$, where p is the largest kernel size.

Overall, the computational complexity of our CDA-TBD method is thus in the order of $\mathcal{O}(pF_I F_O) + \mathcal{O}(TK^2)$, with $p, F_I, F_O, K < 10^2$.

5. Results

For the case of single target images, the performance of our CDA-TBD method is validated against two alternative approaches. The first benchmark method, denoted *CDA-Max*, is derived by considering the output provided by the CDA alone. For the tracking task, the target position is estimated by considering the maximum CDA activation at each row: elements should be 1 only in correspondence to the target positions (i.e., the center of the line in the TD image) and 0 elsewhere. For the detection task, in CDA-Max we compare the number of logistic activation along the best path that passes a desired probability P_{act} to threshold $\rho \cdot T$. This detection strategy checks that the number of valid reflections along the identified path is significant. Formally, let $\hat{k}$ be the chosen path and $t_{i,\hat{k}}$, $i = 0, \ldots, T-1$ its related activation. Then, we set the detection flag

$$\text{Detect} = \begin{cases} 1 & \exists \boldsymbol{i} = \{i_1, i_2, \ldots\} : t_{i_j,\hat{k}} > P_{\text{act}}, |\boldsymbol{i}| > \rho \cdot T \\ 0 & \text{else} \end{cases}, \qquad (6)$$

which yields a per-TD matrix detection hard decision. The second benchmark method, denoted *Viterbi*, is the "pure TBD strategy" reported in [11], which was shown to outperform other TBD approaches surveyed above. This method compares the emission probability accumulated throughout the chosen path by the Viterbi algorithm to a number of random paths (excluding the chosen path) along the columns of the TD matrix.

As quality metrics, we consider both detection and tracking performance. The former is measured in terms of the receiver operating characteristics (ROC) to explore the trade off between detection and false alarm probability. Tracking error is measured as the average Euclidean distance between predicted position and ground truth. In the following, we show that our CDA-TBD approach clearly outperforms both benchmark solutions. However, we should note that CDA-TBD holds the disadvantage of setting threshold T_L by e.g., training, whereas, in the CDA-Max approach, both P_{act} and ρ can be set by some knowledge about the motion and shape of the expected target. This observation emphasizes the need for a sufficiently large dataset, such as the one we share.

Results are first qualitatively shown in terms of representative examples of TD matrices and their denoised version, over which we highlight the target path detected by our CDA-TBD approach. Samples might contain either a single target or multiple targets, and refer to different sea experiments. We then present quantitative measures referring to average errors and ROC curves computed over the entire dataset of more than 20,000 clutter-based TD matrices, separately grouped according to SCR level.

5.1. Representative Results and Sea Trial Demonstration

A representative set of TD matrices, their denoised version, and the final tracking result is shown in Fig. 6 for three different levels of SCR (ground truth target position is reported in the bottom panels), and for cases of a single target and of multiple targets. The leftmost columns demonstrates detection in the case of multiple targets: all target positions are accurately tracked over time. Successive columns show reflections from single targets, where the target's motion varies between the TD matrices. We observe that in all cases our CDA-TBD method can accurately track the target, even in the presence of very noisy input (e.g., SCR of 4dB), as indicated by the close match between the tracked line and the ground truth position. Note how the CDA output provides a denoised version of the TD matrix, where the most likely target positions are highlighted. We observe how, after the denoising operation, the position of the target is much better identified than over the original TD matrix.

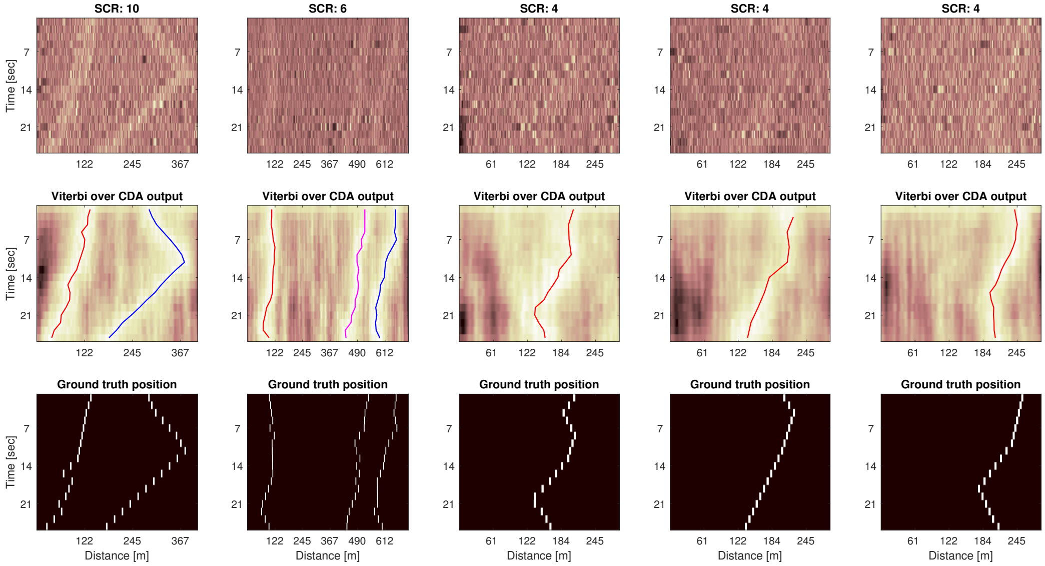


Figure 6: Tracking examples for several TD matrices featuring multiple and single moving targets, at different levels of Signal-to-Clutter ratios. The top row shows the input (noisy) images. The middle row shows the CDA (denoised) images, with the tracked path discovered by our CDA-TBD algorithm superimposed as a red curve. Bottom panels show the corresponding ground truth position of the targets.

While the above examples show results for TD matrix created by combining real recordings of clutter and of target's reflections, our solution should be readily applied also in realistic scenarios where the TD matrix includes both reflection types. Such is the case in Fig. 7, where we show the original TD matrix, its denoised version, and the tracking result for two sea experiments including scuba divers. These particular experiments were performed in the Mediterranean Sea, across the shores of Northern Israel, at water depth of roughly 90m. The sea level was “2” with waves exceeding 0.5m height. The seabed was a combination of rocks and clay, and the sound speed was roughly

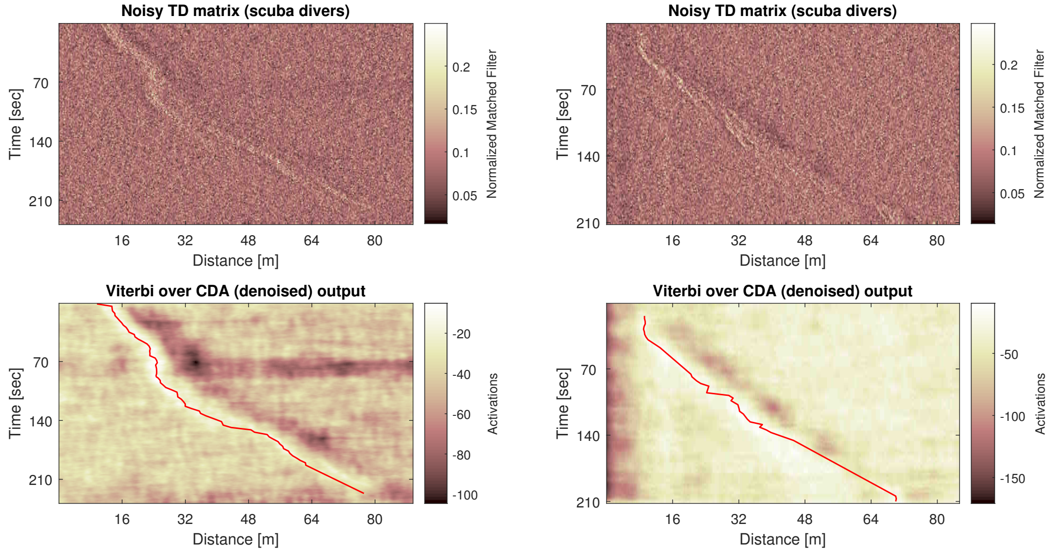


Figure 7: Application of the proposed methodology to TD matrices recorded from the movement of scuba divers.

steady at 1530-1525m/s at the top 25m and decreasing linearly to 1510m/s near the bottom. The target were two scuba divers swimming closely. The divers used closed re-breather systems with neoprene-covered air tanks, which made their target strength particularly low. Observing the original TD matrix we note that, while the divers' path is visible, per acoustic emission, the reflection pattern is very low and compares to the clutter (i.e., low SCR). This motivates the need for pattern-based detection. The denoised matrices shown in Fig. 7 emphasize the divers' path, making it easier to track the target. As clearly observed, the chosen track matches the motion of the target divers.

5.2. Statistical Analysis

5.2.1. Detection Performance

Fig. 8 shows the ROC performance of all the methods considered, where the different detection and false alarm rates are obtained by changing the detection threshold for each method. We test performance for two relatively low SCR of 4dB and 6dB. We observe that, without the denoising step provided by the CDA, performance of the Viterbi algorithm is poor. This is because at low SCR, while the Viterbi approach may catch the right track, the combined probability of the reference tracks are similar to that of the chosen path, thus the likelihood ratio is low. Instead, thanks to the denoising step, in the CDA-TBD case the probability of the reference tracks is low compared to the best path, and the ratio (4) is high even at low SCR. This insensitivity to the clutter noise is also the reason why performance of CDA-TBD is better than CDA-Max. That is, while the latter is making detection decision based on a single denoised observation, the former combines tracks before making a hard decision.

5.2.2. Tracking Performance

Next, we explore the tracking capability of the three approaches. The tracks are obtained by separately setting the detection thresholds for the Viterbi, CDA-Max and CDA-TBD using the ROC curves in Fig. 8, according to a desired false alarm rate of 10^{-4}. Results are shown in Fig. 9 as a function of the SCR. We observe that, already at SCR of 6 dB, tracking capability of CDA-Max is low. This happens because taking the maximum value only considers the instantaneous reflection, whereas the other methods observe a global pattern in the denoised matrix. Still, considering a single reflection

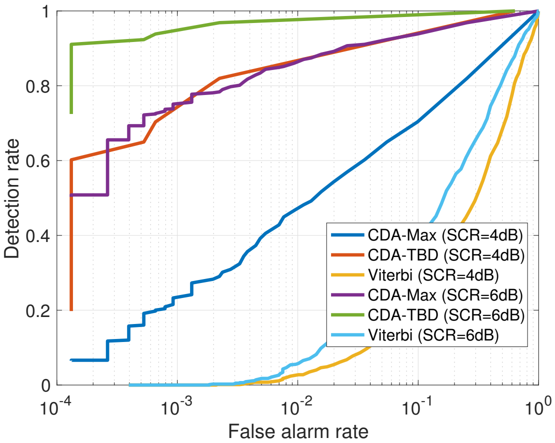


Figure 8: Receiver operating characteristics (ROC) for the three compared methods for SCR=4 dB and 6 dB. Results shows a favourable trade-off between detection and false alarm rates for CDA-TDB.

holds the advantage of independence of the motion of the target. Thus, CDA-Max outperforms the Viterbi approach at high SCR, for which the denoising step is able to filter out much of the clutter. However, the best performance is always given by the CDA-TBD approach, which considers a much lower number of possible targets and it is thus able to produce very accurate results even at low SCR levels. In particular, a sub-meter accuracy is still obtained for low SCR of 8dB.

6. Conclusions

In this paper, we presented an innovative CDA-TBD approach for the efficient detection of multiple mobile submerged targets by active acoustics. Our method takes as input a time-distance (TD) matrix, which concatenates reflections from a train of emitted signals. Motivated by the curved-like pat-

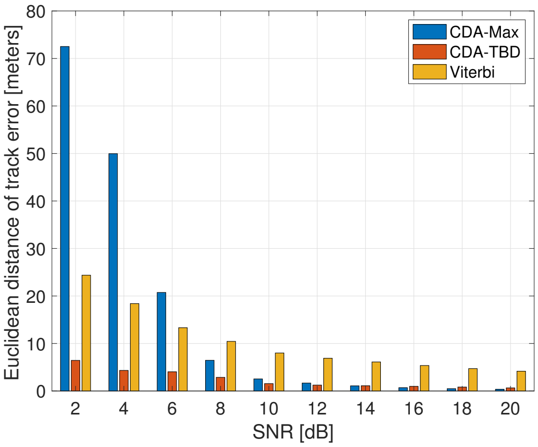


Figure 9: Average tracking error for the three detection approaches as a function of the SCR. Results show resilience of CDA-TBD to high clutter.

tern created by the target along the time domain, the TD matrix is then filtered through a convolutional denoising autoencoder (CDA) in order to highlight potential patterns in the images. The CDA is trained by an augmented database collected during 50 designated sea experiments, performed under a variety of sea environments. The denoised image is further processed by a probabilistic track-before-detect (TBD) approach to choose paths that fits user-defined expectations about the targets' maximum size and velocity. This is performed through dynamic programming such that, rather than exploring single reflections, all reflections are considered, thereby allowing detection and tracking even at very low signal-to-clutter ratios. Notably, combining dynamic programming with deep learning allows to cut down computational complexity, which makes the proposed approach a perfect candidate for low-power marine monitoring devices. Moreover, results over the collected dataset of sea experiments show a favourable detection-false alarm trade-off and far better tracking performance over two benchmark schemes. In order to promote further developments, we freely share our dataset with the community. As a promising research direction, in future work we will explore how detection performance might be improved by training the CDA in a completely unsupervised way, for example by implementing an anomaly detection scheme where a change in the structure of the clutter could be interpreted as the presence of a potential target.

Acknowledgements

This work was funded by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 773753 (SYMBIOSIS), and by the NATO Science for Peace and Security Programme under grant G5293. A.T. gratefully acknowledges the support of the NVIDIA Corporation for the donation of a Titan Xp GPU used for this research.

Part of this work has been presented in the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Dec., 2019.

The authors have no competing interests to declare.

References

[1] A. Bertrand, E. Josse, Acoustic estimation of longline tuna abundance, ICES Journal of Marine Science 57 (2000) 919–926.

[2] D. Ketten, Experimental measures of blast and acoustic trauma in marine mammals, ONR Final Report: N000149711030, 2006.

[3] M. J. Parsons, I. Parnum, K. Allen, R. McCauley, C. Erbe, Detection of sharks with the Gemini imaging SONAR, Acoustics Australia 42 (2014) 0.

[4] N. O. Bakir, A brief analysis of threats and vulnerabilities in the maritime domain, in: Managing critical infrastructure risks, Springer, 2007, pp. 17–49.

[5] S. Davey, M. Rutten, B. Cheung, A comparison of detection performance for several track-before-detect algorithms, in: International Conference on Information Fusion, 2008. doi:10.1155/2008/428036.

[6] P. Willett, S. Coraluppi, Application of the MLPDA to bistatic sonar, in: IEEE Aerospace Conference, 2005, pp. 2063–2073. doi:10.1109/AERO.2005.1559498.

[7] C. Jing, Z. Lin, J. Li, Detection and tracking of an underwater target using the combination of a particle filter and track-before-detect, in: IEEE OCEANS, 2016, pp. 1–5.

[8] M. Wei, B. Shi, C. Hao, S. Yan, A novel weak target detection strategy for moving active SONAR, in: IEEE OCEANS, 2018.

[9] S. Schoenecker, P. Willett, Y. Bar-Shalom, Resolution limits for tracking closely-spaced targets, IEEE Transactions on Aerospace and Electronic Systems (2018) 1–1. doi:10.1109/TAES.2018.2832939.

[10] J. Wang, A. von Trojan, S. Lourey, Active sonar target tracking for anti-submarine warfare applications, in: IEEE OCEANS, 2010.

[11] R. Diamant, D. Kipnis, E. Bigal, A. Scheinin, D. Tchernov, A. Pinchasi, An active acoustic track-before-detect approach for finding underwater mobile targets, IEEE Journal of Selected Topics in Signal Processing 13 (2019) 104–119.

[12] W. Yi, M. R. Morelande, L. Kong, J. Yang, An efficient multi-frame track-before-detect algorithm for multi-target tracking, IEEE Journal of Selected Topics in Signal Processing 7 (2013) 421–434.

[13] H. Jiang, W. Yi, T. Kirubarajan, L. Kong, X. Yang, Multiframe radar detection of fluctuating targets using phase information, IEEE Transactions on Aerospace and Electronic Systems 53 (2017) 736–749.

[14] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (2015) 436.

[15] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[16] G. Hinton, L. Deng, D. Yu, G. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, B. Kingsbury, et al., Deep neural networks for acoustic modeling in speech recognition, IEEE Signal processing magazine 29 (2012).

[17] S. Takamichi, Y. Saito, N. Takamune, D. Kitamura, H. Saruwatari, Phase reconstruction from amplitude spectrograms based on directional-statistics deep neural networks, Signal Processing 169 (2020) 107368.

[18] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, Stacked

denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, Journal of machine learning research 11 (2010) 3371–3408.

[19] A. Testolin, R. Diamant, Underwater acoustic detection and localization with a convolutional denoising autoencoder, in: IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), IEEE, Guadeloupe, West Indians, 2019.

[20] J. R. Bates, D. Grimmett, G. Canepa, A. Tesei, Towards doppler estimation and false alarm rejection for continuous active sonar, The Journal of the Acoustical Society of America 143 (2018) 1972–1972.

[21] J. Renard, L. Lampe, F. Horlin, Sequential likelihood ratio test for cognitive radios., IEEE Transaction on Signal Processing 64 (2016) 6627–6639.

[22] K. J. Sangston, K. R. Gerlach, Coherent detection of radar targets in a non-Gaussian background, IEEE Transactions on Aerospace and Electronic Systems 30 (1994) 330–340. doi:10.1109/7.272258.

[23] J. Tropp, A. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Transactions on information theory 53 (2007) 4655–4666.

[24] M. Desai, R. Mangoubi, Robust Gaussian and non-Gaussian matched subspace detection, IEEE Transactions on Signal Processing 51 (2003) 3115–3127.

[25] K. Lau, M. Salibian-Barrera, L. Lampe, Modulation recognition in the 868 Mhz band using classification trees and random forests, AEU-International Journal of Electronics and Communications 70 (2016) 1321–1328.

[26] S. Schoenecker, P. Willett, Y. Bar-Shalom, ML-PDA and ML-PMHT: Comparing multistatic sonar trackers for VLO targets using a new multi-target implementation, IEEE Journal of Oceanic Engineering 39 (2014) 303–317. doi:10.1109/JOE.2013.2248534.

[27] C. Jauffret, Y. Bar-Shalom, Track formation with bearing and frequency measurements in clutter, IEEE Transactions on Aerospace and Electronic Systems 26 (1990) 999–1010. doi:10.1109/7.62252.

[28] W. Blanding, P. Willett, S. Coraluppi, Sequential ML for multistatic sonar tracking, in: OCEANS, 2007, pp. 1–6. doi:10.1109/OCEANSE.2007.4302356.

[29] R. Streit, T. Luginbuhl, Maximum likelihood method for probabilistic multihypothesis tracking, in: Signal and Data Processing of Small Targets, volume 2235, International Society for Optics and Photonics, 1994, pp. 394–406.

[30] S. Davey, M. Wieneke, H. Vu, Histogram-PMHT unfettered, IEEE Journal of Selected Topics in Signal Processing 7 (2013) 435–447.

[31] H. Vu, S. Davey, F. Fetcher, S.Arulampalam, R. Ellem, C. Lim, Track-before-detect for an active towed array SONAR, in: Proceedings of Acoustics, 2013.

[32] H. Gaetjens, S. Davey, S.Arulampalam, F. Fletcher, C. Lim, Histogram-PMHT for fluctuating target models, IET Radar, Sonar Navigation 11 (2017) 1292–1301.

[33] S. Schoenecker, P. Willett, Y. Bar-Shalom, The effect of K-distributed clutter on trackability, IEEE Transactions on Signal Processing 64 (2016) 475–484. doi:10.1109/TSP.2015.2478745.

[34] D. Yu, L. Deng, Deep learning and its applications to signal and information processing [exploratory dsp], IEEE Signal Processing Magazine 28 (2010) 145–154.

[35] A. Testolin, I. Stoianov, M. De Filippo De Grazia, M. Zorzi, Deep unsupervised learning on a desktop pc: a primer for cognitive scientists, Frontiers in psychology 4 (2013) 251.

[36] H. Lee, R. Grosse, R. Ranganath, A. Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, in: Proceedings of the 26th annual international conference on machine learning, ACM, 2009, pp. 609–616.

[37] H. Palangi, R. Ward, L. Deng, Convolutional deep stacking networks for distributed compressive sensing, Signal Processing 131 (2017) 181–189.

[38] M. Zorzi, A. Zanella, A. Testolin, M. D. F. De Grazia, M. Zorzi, Cognition-based networks: A new perspective on network optimization using learning and distributed intelligence, IEEE Access 3 (2015) 1512–1530.

[39] C. Lu, Z.-Y. Wang, W.-L. Qin, J. Ma, Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification, Signal Processing 130 (2017) 377–388.

[40] W. Huang, H. Ding, G. Chen, A novel deep multi-channel residual networks-based metric learning method for moving human localization in video surveillance, Signal Processing 142 (2018) 104–113.

[41] P. Baldi, P. Sadowski, D. Whiteson, Searching for exotic particles in high-energy physics with deep learning, Nature communications 5 (2014) 4308.

[42] E. L. Ferguson, R. Ramakrishnan, S. B. Williams, C. T. Jin, Convolutional neural networks for passive monitoring of a shallow water environment using a single sensor, in: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2017, pp. 2657–2661.

[43] S. Kamal, S. K. Mohammed, P. S. Pillai, M. Supriya, Deep learning architectures for underwater target recognition, in: 2013 Ocean Electronics (SYMPOL), IEEE, 2013, pp. 48–54.

[44] G. Hu, K. Wang, Y. Peng, M. Qiu, J. Shi, L. Liu, Deep learning methods for underwater target feature extraction and recognition, Computational intelligence and neuroscience 2018 (2018).

[45] R. Diamant, Closed form analysis of the normalized matched filter with a test case for detection of underwater acoustic signals, IEEE Access 4 (2016) 8225–8235.

[46] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Boochoon, S. Birchfield, Training deep networks with synthetic data: Bridging the reality gap by domain randomization, arXiv preprint arXiv:1804.06516 (2018).

[47] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, R. Webb, Learning from simulated and unsupervised images through adversarial training, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2107–2116.

[48] J. Salamon, J. P. Bello, Deep convolutional neural networks and data augmentation for environmental sound classification, IEEE Signal Processing Letters 24 (2017) 279–283.

[49] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., Tensorflow: A system for large-scale machine learning, in: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), 2016, pp. 265–283.

[50] P. Maji, R. Mullins, On the reduction of computational complexity of deep convolutional neural networks, Entropy 20 (2018) 305.