# Deep transfer learning for source ranging: Deep-sea experiment results

**Article** *in* The Journal of the Acoustical Society of America · October 2019

DOI: 10.1121/1.5126923

CITATIONS
47

READS
327

**7 authors**, including:

Wenbo Wang
Chinese Academy of Sciences
**25** PUBLICATIONS **154** CITATIONS
SEE PROFILE

Haiyan Ni
Chinese Academy of Sciences
**7** PUBLICATIONS **95** CITATIONS
SEE PROFILE

su Lin
Chinese Academy of Sciences
**17** PUBLICATIONS **127** CITATIONS
SEE PROFILE

Qunyan Ren
Institute of acoustics Chinese Academy of Sciences
**83** PUBLICATIONS **311** CITATIONS
SEE PROFILE

# Deep transfer learning for source ranging: Deep-sea experiment results

Wenbo Wang, Haiyan Ni, Lin Su, Tao Hu, Qunyan Ren, Peter Gerstoft, and Li Ma

Citation: The Journal of the Acoustical Society of America **146**, EL317 (2019); doi: 10.1121/1.5126923
View online: https://doi.org/10.1121/1.5126923
View Table of Contents: https://asa.scitation.org/toc/jas/146/4
Published by the Acoustical Society of America



# Deep transfer learning for source ranging: Deep-sea experiment results

**Wenbo Wang,[1,a)] Haiyan Ni,[1,a)] Lin Su,[1] Tao Hu,[1] Qunyan Ren,[1,b)] Peter Gerstoft,[2] and Li Ma[1]**

[1]*Key Laboratory of Underwater Acoustic Environment, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China*

[2]*Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0238, USA*

*wangwenbo215@mails.ucas.edu.cn, ihaiyan@mail.ioa.ac.cn, sulin807@mail.ioa.ac.cn, hutao@mail.ioa.ac.cn, renqunyan@mail.ioa.ac.cn, pgerstoft@ucsd.edu, mali@mail.ioa.ac.cn*

**Abstract:** A deep transfer learning for underwater source ranging is proposed, which migrates the predictive ability obtained from synthetic environment (source domain) into an experimental sea area (target domain). A deep neural network is first trained on large synthetic datasets generated from historical environmental data, and then part of the neural network is refined on collected data set for source ranging. Its performance is tested on a deep-sea experiment through comparing with convolutional neural networks of different training datasets. Data processing results demonstrate that the ranging accuracy is considerably improved by the proposed method, which can be easily adapted for related areas.



[CCC]

**Date Received:** July 9, 2019 **Date Accepted:** September 3, 2019

## 1. Introduction

Passive source ranging is an area of interest that is important in underwater acoustics. Matched-field processing (MFP) has been proposed and applied for years in source ranging,[1,2] which could perform poorly in the presence of mismatch between observed and predicted field. Data-based deep learning provides a promising approach for source ranging even when the environmental model is not chosen adequately.[3,4] However, the training process for deep learning requires much data from sea trial, the collection of which is expensive and time-consuming. Alternatively, a large number of replicas from environmental models can be used to overcome the lack of measured data for deep learning.[5–7] The accuracy of this approach is correlated with the consistency between the fields from these environmental models and actual ocean waveguide, and this approach might only work in regions with similar environment as the network is trained in.

For unfamiliar areas, there are neither sufficient acoustic field data nor adequate environmental models to generate replica data to train accurate deep neural network (DNN). Transfer learning[8] has emerged as a learning framework to migrate existing knowledge into a new environment. Deep transfer learning (DTL)[9] migrates the predictive ability of a trained DNN into a new similar environment through sharing some DNN parameters and rediscovering others. This paper adapts DTL for source ranging in an uninvestigated deep-sea area, which takes full advantage of the historical environment data and limited experimental acoustic data. The feasibility and accuracy of the approach is tested on real data.

## 2. Deep transfer learning for source ranging

DTL for underwater ranging utilizes the knowledge learned from a scenario (modeled field or historical environmental field) to assist source ranging in an area similar to the scenario environment. For new areas, traditional machine learning trains environment-related models individually as shown in Fig. 1(a), which cannot be applied to other environments directly. Transfer learning techniques can transfer a pre-trained knowledge model to a new area. Here, the pre-trained knowledge model is obtained on historical data including prior information on the environment model and the standardized sound field data. Then, a small amount of experimental acoustic data is used to transfer the pre-trained knowledge model for source ranging to a new area.

---

[a)]Also at: University of Chinese Academy of Science, Beijing 100049, China.

[b)]Author to whom correspondence should be addressed.

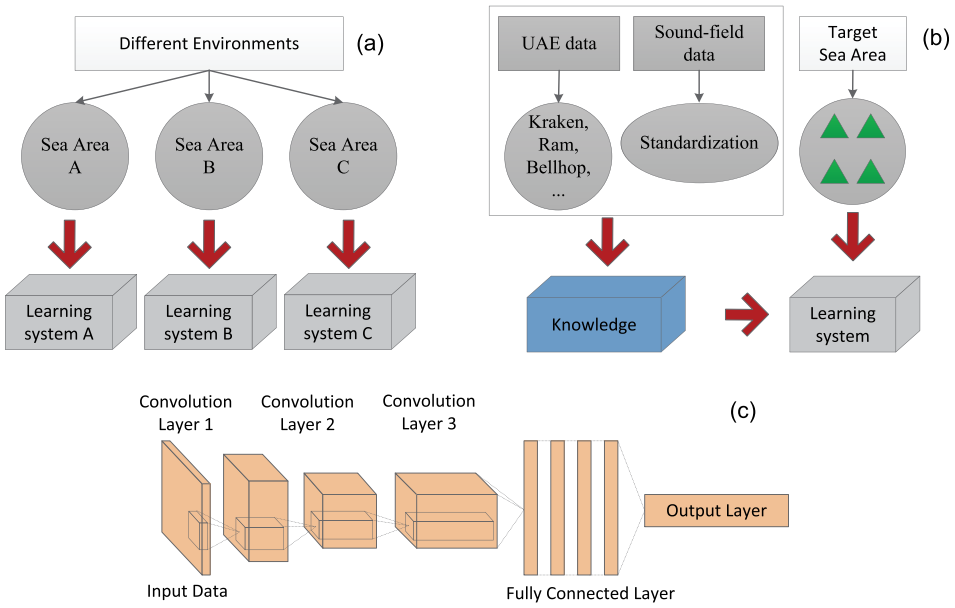


Fig. 1. (Color online) Ranging learning processes with (a) traditional machine learning and (b) transfer learning using underwater acoustic environment data and sound-field data. (c) Schematic of a typical convolutional neural network.

### *2.1 Convolutional neural network*

Convolutional neural network (CNN) is a class of DNNs and has been successfully used in acoustic target classification and localization.[5,10,11] Figure 1(c) shows the schematic of a CNN as used here. It is a DNN consisting of convolution layers, pooling layers and full connection layers. Convolution layers use convolution kernels to extract feature maps from input multidimensional arrays by convolution operations. A pooling layer performs down-sampling by dividing the input rectangular pooling region, and computing the maximum or average of each region.

### *2.2 Data preprocessing, labels, and CNN structures*

The input data, including the simulated sound field data and the actual received data, are all pre-processed to eliminate the influence of the source spectrum. To reduce the effect of the source amplitude, the complex pressure $\boldsymbol{p}(f,l)$ at frequency $f$ and element $l$ is normalized by

$$\tilde{\boldsymbol{p}}(f,l)=\frac{\boldsymbol{p}(f,l)}{\sqrt{\sum_{l=1}^{L}|\boldsymbol{p}(f,l)|^2}}, \tag{1}$$

where $L$ is the number of sensors and $\boldsymbol{p}(f,l)$ is the complex pressure obtained from a Fourier transform with a step $\Delta t$ and a time window $N_t\cdot\Delta t$. The real and imaginary parts of the normalized pressure $\tilde{\boldsymbol{p}}(f,l)$ are interleaved in the matrix of size $L\times 2F$ ($F$ is number of frequencies), which is the input of CNN.

Source ranging has been modelled as a classification problem, where the source range is discretized into bins to label the corresponding sound field data using supervised machine learning[3] assuming the target distribution is a delta function. To accommodate uncertainty in the true source range, the target distribution is here assumed Gaussian centered at the observed GPS range $r_{n,\mathrm{GPS}}$ with a standard deviation $\sigma$. The discrete target distribution $\mathbf{t}_n$, for each range bin is mapped to a $1\times K$ vector,

$$t_{nk}\propto\exp\left(-(r_{nk}-r_{n,\mathrm{GPS}})^2/(2\times\sigma^2)\right), \tag{2}$$

where $r_{nk}$ is equidistant position, $\Delta r$ is equal distance interval, $K$ is the neural number of output layer, $\sigma=0.02$ km represents the uncertainty in GPS positions at both receiver and source. These target distributions are used to train the CNN.

The CNN, see Fig. 1(c), has three convolutional pooling layers and four fully connected layers, other parameters are in Table 1. For convolutional pooling layer, the convolution kernel sizes are $5\times 5$, $3\times 3$, and $3\times 3$, respectively. The pooled kernel is

Table 1. CNN parameters.

| Items | Factors |
|---|---|
| Initial learning rate | 0.0001 |
| Learning rate drop factor | 0.1 |
| Learning rate drop period | 20 |
| Dropout factor | 0.3 and 0.5 |
| Optimization method | Adam (Ref. 13) |
| Maximum iteration steps | 20 000 |

$2 \times 2$, of which the first and second layers have 128 convolution kernels, the third layer has 256 convolution kernels. The activation function is rectified linear unit (ReLU), $f(x) = \max(0, x)$. Hereafter, the output layer is connected to four fully connected layers of length 2048. To prevent over-fitting, a dropout layer[12] (dropout factor is 0.3) is placed between the convolutional pooling layer and the fully connected layer as well as between (dropout factor is 0.5) the fully connected layer and the output layer.

### *2.3 Criteria for performance test*

To compare the performance of different source ranging methods, the probability of credible localization (PCL)[14] of the estimated values at all sample points within a certain relative error range is defined as

$$P_{\mathrm{PCL}-5\%} = \frac{\sum_{k=1}^{K} \eta(k)}{K} \quad \begin{cases} \eta(k) = 1, & \dfrac{|\tilde{r}_0(k) - r_0|}{r_0} \times 100\% \le 10\% \\ \eta(k) = 0 & \text{else}, \end{cases} \tag{3}$$

in which $K$ is the number of positionings at a fixed source target intensity, $\tilde{r}_0$ is the estimated range, $r_0$ is the GPS range. The PCL is the ratio of the number of reliable positionings at all ranges to the number of distance points, the estimated range is considered reliable when the relative error is 5%. The signal-to-noise ratio (SNR) for received signal is defined as

$$SNR = 10 \log_{10} \left( \frac{\sum_{i=1}^{F} |S_{signal}(f_i)|^2}{\sum_{i=1}^{F} |S_{noise}(f_i)|^2} \right), \tag{4}$$

where $S_{signal}$ are the received signal at frequency $f_i$ and $S_{noise}$ noise adjacent to the signal, $F$ is the number of narrow band frequencies.

## 3. Experiment and results

In November 2017, a passive source ranging experiment was conducted in the deep-sea area of the South China Sea. Figure 2(a) shows the seabed topography of the experimental area from a multibeam echo-sounder. The eight-element vertical line array (VLA) has an element spacing of 20 m and middle array depth of 1950 m. Source positions and source depths of all track lines are shown in Figs. 2(a) and 2(c). Data from track 1 A, track 1B, and track 1 C are used as the training data, these tracks are overlapped as described in Fig. 2. The data from track 2 are used as test data. Track 1 (dotted) has deeper water depth at the middle (2200 m) and shallowest at the end (minimum 1450 m). Track 2 (dashed) has a flat bottom with an average depth 2156 m. During the experiment, the source is towed at depth varying between 40 and 100 m, see Fig. 2(c), emitting eight frequencies (63, 79, 105, 126, 160, 203, 260, and 315 Hz) line-spectra signals. Each group of signals contains four segments of whose transmitting power decreases successively, and denoted as E1, E2, and E3, respectively. The duration of each group is 60 s, and each group is transmitted cyclically every 776 s.

The implementation steps of DTL are as follows.

STEP 1: A deep CNN is set up as shown in Fig. 1(c) depending on the complexity of the data. The CNN structures and factors are given in Sec. 2.2.

Kraken-C[16] is used to generate the sound-pressure-field replicas with historical environment parameters. A single-layer seabed model is used in the sound field simulation, in which the bottom acoustic parameters are as follows: density 1.5 g/cm$^3$, sound

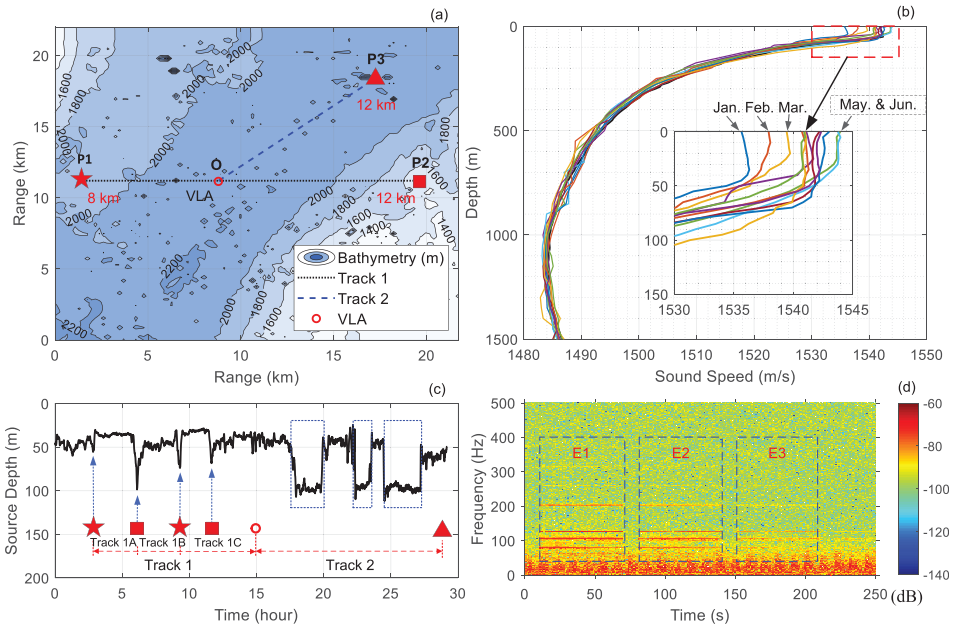


Fig. 2. (Color online) (a) Seafloor topography with VLA (circle) and track 1 A (P1-P2), track1B (P2-P1), track 1 C (P1-P2-O), and track 2 (O-P3). (b) Historical 12-month average sound speed profiles (SSPs) near experimental area (Ref. 15). (c) Source depth during the experiment and representation corresponding to track lines in (a). The sudden depth deepening (indicated by star and square) corresponds to maneuvering of source ship. The deepening (dotted box) is due to source ship stopping. (d) Spectrum of transmitted signals (dashed box) at 10.05 km from the first element of VLA and the SNRs of E1, E2, and E3 are 15.7, 12.0, and 6.3 dB, respectively.

speed can be 1500, 1600, 1700, and 1800 m/s, attenuation coefficient 0.2 dB/wavelength, and historical SSPs data are shown in Fig. 2(b). The source depths are set to 10, 20, 40, 60, 80, and 100 m, sequentially. The horizontal distance interval is 0.02 km and with the longest distance of 15 km, which gives 3-km-longer data than the experimental data processed. A total of 210 000 (12 SSPs × 4 bottom speeds × 6 source depths × 750 ranges) pressure replicas are generated by employing the above mentioned parameters. These data sets are utilized to train deep CNNs to establish a pre-trained CNN model.

For measured data, data from track 1 are divided into track 1 A (from P1 to P2), track 1B (from P2 to P1), and track 1 C (fromP1 to P2 to O) and then used as training data set for DTL. Track 2 (from O to P3) is used to form the test data set. The time-domain pressure field data of an eight-element VLA, including eight frequencies, are calculated by short-time discrete Fourier transform (SDFT) with a delay of 0.25 s and a time window of 1 s. And then, the frequency-domain pressure field data set is normalized according to Eq. (1). Thus, for single signal such as E1 in Fig. 2(d), a total of 240 (60 s divided by 0.25 s) samples are obtained. During the experiment, a total of 16, 14, 21, and 21 sets of signals have been emitted in track 1 A, track 1B, track 1 C, and track 2, respectively. Thus, there are 11 520, 10 080, and 15 120 samples in the training data set of track 1 A, track 1B, and track 1 C, respectively. The data of each emission energy in track 2 have 5040 testing samples. The ranges between source and array are calculated according to GPS data of VLA and ship.

STEP 2: The fully connected layers and output layer of pre-training model are replaced by new fully connected layers and output layer which re-initialize the parameters randomly. The sound field data from track 1 A, track 1B, and track 1 C are used to re-train the new model. A small number of replicas are supplemented to compensate the discontinuity of the measured data.

STEP 3: The sound field data (8 frequencies) from track 2 are used as the test data set for the DTL, whose results are compared to that of CNN-S, CNN-M, and CNN-S&M. Here, CNN-S and CNN-M represent CNN models trained on only simulated data and only measured data from track 1. Combinations of simulated data and all the measured field data in track 1 are used as training data set for CNN-S&M. The training data sets of CNN-S, CNN-M, and CNN-S&M are, respectively, used as replica fields for the Bartlett MFP (MFP-S, MFP-M, and MFP-S&M) for reference purpose. The PCL-5% performances of estimated distances as compared with actual GPS distances are given in Table 2.

As shown in Table 2, the ranging accuracy of MFP-S is between 80.7% and 85.5%, these moderate results are probably due to the mismatch between the actual

Table 2. Reliability analysis of ranging methods for combinations of training and test data sets. In the first column, M and S indicate the measured and simulated data sets, and combinations used in these algorithms are indicated for each algorithm. In the second column, S represent the simulated data set and T1A, T1B, and T1C measured data sets from the corresponding track.

| Ranging algorithms | Training set for CNN or replica for MFP | Number of samples | Iteration steps | Training time (min) | PCL-5% in test data sets (%) | | |
|---|---|---|---|---|---|---|---|
| | | | | | E1 | E2 | E3 |
| MFP-S | S | 216 000 | — | — | 84.2 | 85.5 | 80.7 |
| CNN-S | S | 216 000 | 20 000 | 322 | 90.9 | 89.7 | 79.8 |
| MFP-M | T1A &T1B &T1C | 36 720 | — | — | 86.3 | 82.3 | 83.7 |
| CNN-M | T1A &T1B &T1C | 36 720 | 5000 | 70 | 82.9 | 88.3 | 83.1 |
| MFP-S&M | S&T1A &T1B &T1C | 252 720 | — | — | 86.3 | 88.9 | 86.4 |
| CNN-S&M | S&T1A &T1B &T1C | 252 720 | 20 000 | 322 | 94.6 | 95.2 | 90.0 |
| DTL-1 | T1A | 11 520 | 250 | 2 | 91.7 | 90.8 | 85.2 |
| DTL-2 | T1A &T1B | 21 600 | 250 | 2 | 92.7 | 91.5 | 89.6 |
| DTL-3 | T1A & T1B &T1C | 36 720 | 250 | 2 | 96.0 | 97.8 | 91.2 |

ocean waveguide of track 2 (range-independent) and the modelled environment (range-dependent) generating the replica data set. However, the performance of MFP-S is expected to be more accurate if the ocean environment model agrees well with actual ocean waveguide. The CNN-S has better results on data sets E1 and E2 (by 4%), demonstrating the ability of CNN in extracting features from simulated data. The MFP-S slightly outperforms (by 1%) the CNN-S on data set E3. The rest discussions will focus on the comparisons between traditional CNN and DTL.

The results in Table 2 demonstrate the ranging accuracy of CNN-M is significantly lower than that of CNN-S, though they are trained on the same network structure [Fig. 1(c)]. Such results are likely due to the small amount of training data set, the CNN-M might be overfitted and has weak generalization ability on test data set.

The results of DTL demonstrate an increasing performance with the increment of migrated data, whose accuracies on these test data sets all exceed that of CNN-S and CNN-M. These results indicate that a relatively small amount of measured data can greatly improve the ranging accuracy of CNN using DTL method. It is encouraging that the DTL is also shown to be better than that of CNN-S&M, besides, the CNN-S&M needs much longer training time in practice than the DTL in CPU time (on a ThinkPad T570 laptop with 4-core i7-7500U CPU @ 2.70 GHz). This comparison demonstrates the effectiveness and reliability of DTL are better than traditional CNN.

Figure 3 shows the performance comparison of DTL method and traditional CNN method for range estimation. For traditional CNN, with the decrement of signal energy, there are increasing errors in range estimation at the distance of 4–6 km, and

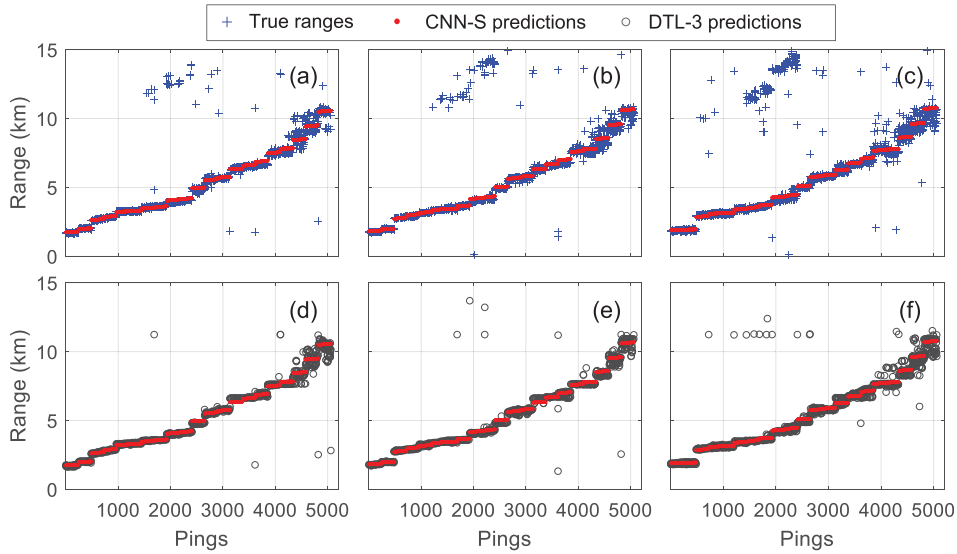


Fig. 3. (Color online) Ranging predictions for testing data set in track 2 using traditional CNN method with emission energy of (a) E1, (b) E2, and (c) E3 and DTL-3 method with emission energy of (d) E1, (e) E2, and (f) E3.

there is noticeable increment of wrong estimates at the distance of 11–15 km. For DTL results in Figs. 3(d), 3(e), and 3(f), the estimated ranges are basically consistent with the true ranges. This shows that the data generated only by the sound field calculation program will appear side lobes of distance estimation because of the difficulties in obtaining accurate environmental parameters.

## 4. Conclusion

This paper proposes a source ranging method based on deep transfer learning, which has been tested in a deep-sea environment without adequate environmental information. Replica sound field data are generated from historical environment information to train a pre-trained CNN model; then, the model parameters are adjusted by a little data as collected at sea. The source ranging results from at-sea data processing suggest that the DTL approach is an attractive method compared with the traditional MFP and deep learning methods (that only use simulated data or real data), which generally outputs more accurate ranging results with an increase in the amount of transferred data.

## Acknowledgment

Q.R. and L.S. would like thank the support of the project 11604360 and 11704396 as funded by the National Natural Science Foundation of China.

## References and links

[1]A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**, 401–424 (1993).
[2]K. L. Gemba, S. Nannuru, P. Gerstoft, and W. S. Hodgkiss, "Multi-frequency sparse Bayesian learning for robust matched field processing," J. Acoust. Soc. Am. **141**, 3411–3420 (2017).
[3]H. Niu, E. Reeves, and P. Gerstoft, "Source localization in an ocean waveguide using supervised machine learning," J. Acoust. Soc. Am. **142**, 1176–1188 (2017).
[4]H. Niu, E. Ozanich, and P. Gerstoft, "Ship localization in Santa Barbara Channel using machine learning classifiers," J. Acoust. Soc. Am. **142**(5), EL455–EL460 (2017).
[5]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. **143**(5), 2922–2932 (2018).
[6]H. Niu, Z. Gong, E. Ozanich, P. Gerstoft, H. Wang, and Z. Li, "Deep-learning source localization using multi-frequency magnitude-only data," J. Acoust. Soc. Am. **146**(1), 211–222 (2019).
[7]M. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M. Roch, S. Gannot, and C. Deledalle, "Machine learning in acoustics: Theory and applications," J. Acoust. Soc. Am. in press (2019).
[8]S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Trans. Know. Data Eng. **22**(10), 1345–1359 (2010).
[9]J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," Adv. Neural Inf. Process. Syst. **27**, 3320–3328 (2014).
[10]E. L. Ferguson, R. Ramakrishnan, S. B. Williams, and 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)*, New Orleans, LA (2017), pp. 2657–2661.
[11]H. Yang, J. Li, S. Shen, and G. Xu, "A deep convolutional neural network inspired by auditory perception for underwater acoustic target recognition," Sensors **19**, 1104 (2019).
[12]N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," J. Mach. Learn. Res. **15**, 1929–1958 (2014).
[13]D. Kingma and B. Jimmy, "Adam: A method for stochastic optimization," arXiv:1412.6980 (2014).
[14]R. Duan, K. Yang, Y. Ma, Q. Yang, and H. Li, "Moving source localization with a single hydrophone using multipath time delays in the deep ocean," J. Acoust. Soc. Am. **136**(2), EL159–EL165 (2014).
[15]*World Ocean Atlas 2013*, National Oceanographic Data Center (Last viewed September 16, 2019).
[16]M. B. Porter, "The KRAKEN normal mode program," SACLANT Undersea Research Centre, Italy (1991), Vol. 1.