



Deep learning aided OFDM receiver for underwater acoustic communications



Yonglin Zhang^{a,b,c}, Chao Li^{a,b,*}, Haibin Wang^{a,b}, Jun Wang^{a,b}, Fan Yang^c, Fabrice Meriaudeau^c

^aState Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China

^bUniversity of Chinese Academy of Sciences, Beijing, China

^cLaboratory ImViA, Université Bourgogne Franche-Comté, 21078 Dijon, France

ARTICLE INFO

Article history:

Received 30 April 2020

Received in revised form 1 November 2021

Accepted 2 November 2021

Available online 21 November 2021

Keywords:

Underwater acoustic communication

OFDM

CNN

Skip connections

ABSTRACT

In this study, we propose a deep learning (DL)-based orthogonal frequency division multiplexing (OFDM) receiver for underwater acoustic (UWA) communications. Compared to existing deep neural network (DNN) OFDM receivers composed of fully connected (FC) layers, our model tailors complex UWA communications with precision. To this end, it utilizes a convolutional neural network with skip connections to perform signal recovery. The stacks of convolutional layers with skip connections can effectively extract promising features from received signals and reconstruct the original transmitted symbols. Then, a multilayer perceptron is used for demodulation. To demonstrate the performance of the proposed DL-based UWA-OFDM communication system, the training and testing sets are generated using the strength of the measured-at-sea WATERMARK dataset. The experimental results show that the proposed model with skip connections can outperform the existing approaches (i.e., traditional UWA-OFDM with least squares channel estimation, and FC-DNN-based framework) in terms of both accuracy and efficiency. This is prominent in harsh UWA environments with strong multipath spread and rapid time-varying characteristics.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

Underwater acoustic (UWA) channels are known as one of the most challenging communication media [1–3]. A set of characteristics distinguishes UWA communications from general wireless communication scenarios, including severe transmission loss, time-varying multipath propagation, significant Doppler spread, and complex ocean noise. In the past few decades, the orthogonal frequency division multiplexing (OFDM) scheme has been adopted in UWA communications to overcome the long multipath spread in UWA channels [4–6]. As a multicarrier system, OFDM divides the bandwidth into several orthogonal narrowband subcarriers. The low symbol rate renders the use of a guard interval between symbols possible and affordable, enabling the system to handle time-spreading and eliminate inter-symbol interference (ISI). Because the channel suffers from frequency-selective fading and time-varying factors in certain UWA environments [7], channel estimation becomes crucial in a UWA OFDM communication system. To

this end, fixed pilot symbols are usually sent with the data subcarriers for accurate channel estimation [8].

Recently, deep learning (DL) techniques have been increasingly utilized for applications in communications. In OFDM communication scenarios where the channel is either unknown or too complex for an analytical description, DL has been proven to be particularly useful [9–11]. Some prior studies demonstrated improved performance as compared to conventional OFDM receivers based on channel estimation, equalization, and demodulation. Ye et al. [12] replaced the channel equalization and demodulation blocks of the receiver by a five-layer fully connected deep neural network (FC-DNN). Initial experiments showed that the NN-based receiver was more efficient than traditional methods. Zhang et al. [13] adopted an FC-DNN to improve the overall performance by increasing the neural density of the network. Jiang et al. [14] and Yang et al. [15] also provided similar FC-DNN schemes for OFDM communication. However, several problems remain to be solved, for example, the high susceptibility of FC-DNNs to small perturbations and distortions, which renders them unsuitable for UWA communication in complex and variable environments. Moreover, FC-DNNs require a large number of parameters, and the complexity grows exponentially with code length, which renders them

* Corresponding author at: State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China.

E-mail address: chao.li@mail.ioa.ac.cn (C. Li).

impractical for resource-constrained UWA hardware platforms. In addition, channel datasets for training and testing in [13,14] are generated using the ray-tracing software BELLHOP, which simulates static channels instead of the typical real-life time-varying UWA channels. In other words, the simulation data cannot depict the overall characteristics of a real UWA environment.

On the other hand, compared to FC-DNN-based OFDM receiver schemes, convolutional neural networks (CNNs) have strong feature extraction capabilities. A collection of neurons in the CNN only responds to a restricted area of given inputs. Because the ISI only exists between consecutive bits of the transmitted sequence and the influence of ocean noise is independent for each bit, a CNN is suitable for UWA communication systems. Xu et al. [16] proposed a joint neural network equalizer and decoder; the CNN equalizer was utilized to compensate for signal distortion. However, as the CNN-based channel equalizer had only six standard convolutional layers (sequentially connected), it was too simple to extract and transform expert knowledge in communications. In addition, the DL-based model was only tested under a wireless communication channel; such experimental results might not hold in UWA environments.

In this study, we propose a DL architecture to replace the conventional or FC-DNN-based receiver for UWA-OFDM communications. The main contributions of this study are summarized as follows:

1 We consider the OFDM signal passing through UWA channels as the interfered signal with multipath effects and ocean noise, which can be regarded as a black-box problem. Subsequently, we apply a CNN model to the physical layer to extract features of structured data in a higher-dimensional latent space and attempt to learn the underlying relationships between its inputs and outputs. In addition, the skip-connection structure provides the CNN model with maximal flexibility to achieve the best performance [17]. Furthermore, because we view the demodulation problem as a classification task, a multilayer perceptron (MLP) with three layers is cascaded after the CNN for demodulation.

2 The UWA channel database for network training and testing is developed using the time-varying channel impulse response (CIR) from the WATERMARK dataset [18,19]. WATERMARK is a benchmark for UWA communication physical-layer schemes. This allows researchers to test and compare algorithms under realistic and reproducible conditions. In this study, we also provide a criterion for channel quality evaluation under different UWA environments, which can be used as a guide for the analysis of the results. Numerical results show that the proposed model outperforms conventional OFDM communication systems as well as the FC-DNN-based model in terms of bit error rate (BER). Significant performance improvements are especially evident under harsh UWA environments.

The remainder of this paper is organized as follows. In Section 2, the conventional UWA-OFDM system model is briefly reviewed. Section 3 presents the proposed DL-based model for UWA-OFDM systems, including basic knowledge, network architectures, and network training. The experimental results and analyses are presented in Section 4. Finally, Section 5 concludes the paper.

2. UWA-OFDM communication system

A schematic of the conventional baseband UWA-OFDM system model is shown in Fig. 1.

Let $h(t, \tau)$ be the general expression of the UWA CIR, which can be represented as follows:

$$h(t, \tau) = \sum_{i=1}^I A_i(t) \delta(\tau - \tau_i(t)). \quad (1)$$

Here, we assume that the UWA channel can be well approximated by I dominant discrete paths, which is denoted in the following as a "path-based" channel model. $A_i(t)$ and $\tau_i(t)$ are the complex gain and time delay of the i -th path at time t , respectively.

The real path amplitudes change with delays. The attenuation is related to the distance traveled, as well as the physics of the propagation processes (sound refraction in the water, sound reflection at the surface, bottom, and any objects). In addition, the time variability of UWA channels is clearly induced by random effects, where fast variations are often caused by the rapid motion of the sea surface (waves) or the system itself. These variations appear randomly and thus require an additional stage of statistical modeling. However, researchers are still struggling to establish standard models that concisely describe the statistics of typical UWA channels [20]. Therefore, in this study, we utilize the feature matching strength of the DL method to recover the signal received after propagation over UWA channels.

On the transmitter side, the transmitted symbols inserted with pilots are first converted to a parallel data stream; then, the inverse discrete Fourier transform (IDFT) unit is used to modulate them on different subcarriers. Following the IDFT unit, a cyclic extension of the time length or a cyclic prefix (CP) is inserted to mitigate the ISI. Note that the length of the CP should not be shorter than the maximum delay spread of the channel.

The received signal can be expressed as

$$y(t) = x(t) \otimes h(t) + w(t), \quad (2)$$

where \otimes denotes the circular convolution, and $x(t)$ and $w(t)$ represent the transmitted signal and additive white Gaussian noise (AWGN), respectively. To simplify the notation, the UWA channel impulse response is denoted as $h(t)$, wherein a time-invariant channel within each OFDM symbol is assumed. After removing the CP and performing DFT, the received frequency-domain signal becomes

$$Y(k) = X(k)H(k) + W(k), \quad (3)$$

where $Y(k)$, $X(k)$, $H(k)$, and $W(k)$ are the DFTs of $y(t)$, $x(t)$, $h(t)$, and $w(t)$, respectively.

In the traditional UWA-OFDM system, pilot signals are extracted and utilized for channel estimation. The least squares (LS) estimation of $H(k)$ can be expressed as [21]

$$H_{LS}(k) = \frac{Y_{pilot}(k)}{X_{pilot}(k)} = H(k) + \frac{W(k)}{X_{pilot}(k)}, \quad (4)$$

and the corresponding mean-squared error (MSE) is

$$J_{LS} = \mathbb{E} \left\{ \|H(k) - H_{LS}(k)\|_2^2 \right\} \propto \frac{1}{\sigma_x^2 / \sigma_w^2}. \quad (5)$$

From Eq. 5, we can see that the MSE of the LS estimation is inversely proportional to the signal-to-noise ratio (SNR) defined as σ_x^2 / σ_w^2 .

We assume that the pilot symbols are present in the first OFDM block, while the following OFDM blocks consist of the transmitted data. Together, they form a frame. Here, the received signal consisting of one pilot block and one data block is taken as the input of the proposed DL model.

3. DL-based UWA-OFDM receiver

In this section, the design and training approaches for the DL-based OFDM receiver are introduced.

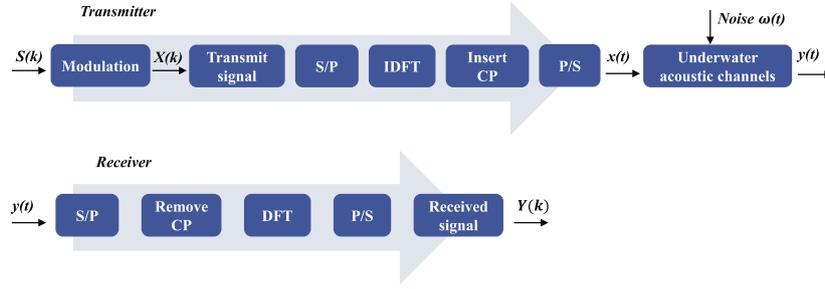


Fig. 1. Block diagram of the conventional UWA-OFDM communication system.

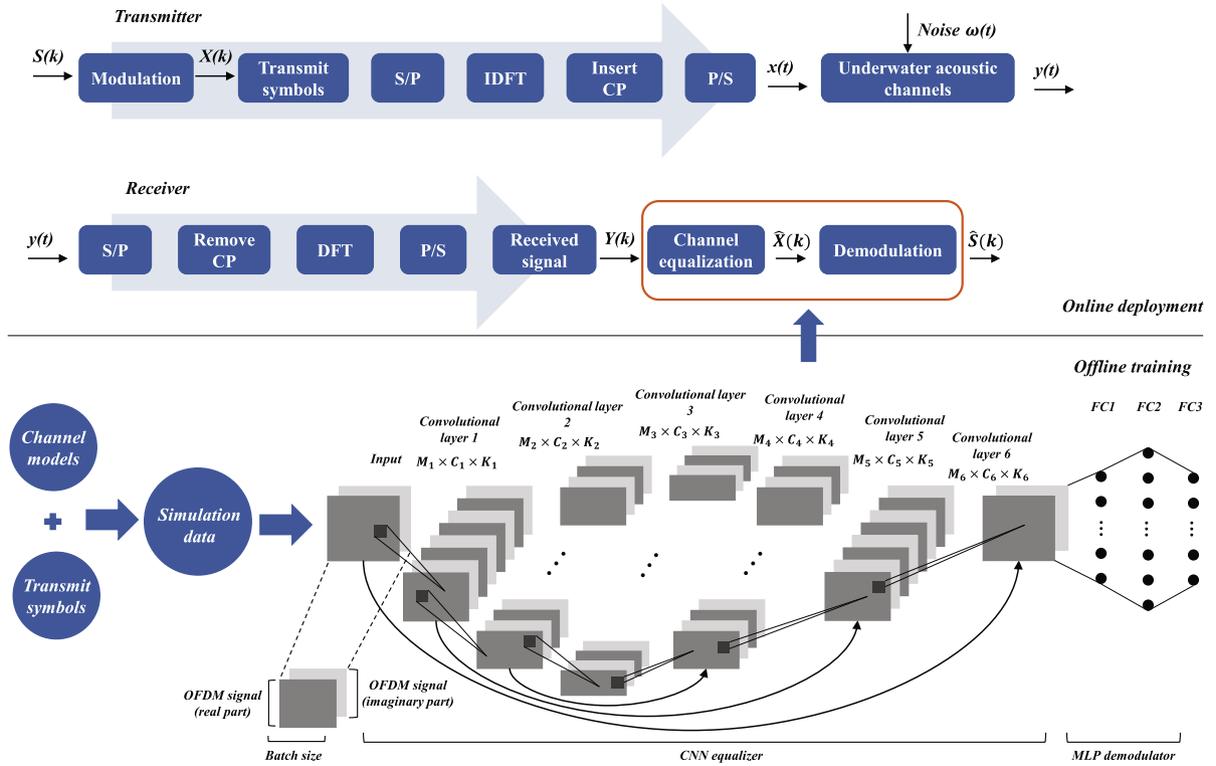


Fig. 2. DL-based UWA-OFDM system architecture.

As shown in Fig. 2, the CNN-based equalizer is utilized to compensate for the distortion of received signals. Then, the MLP is cascaded to realize demodulation.

3.1. CNN equalizer

The CNN architecture adopted in this study is shown in Fig. 3. It comprises several convolutional layers and dilated convolutional layers for the down-sampling and up-sampling, respectively. The novel structures of the residual and skip connections are added to the model as well [17].

As shown in Fig. 3, the CNN equalizer component consists of multiple convolutional layers. The input is the training data, that is, the received signal $Y(k)$. The last layer of the CNN outputs data that approach the expected value, denoted by $\hat{X}(k)$. The multiple convolutional layers are special layers where the network does not indicate the exact form or value of the output. They are only used to extract the signal features and inner connections among them. The output of the previous layer is the input to the next hidden layer. Accordingly, the output of the CNN can be expressed as

$$\hat{X}(k) = f_6(f_5(\dots f_1(Y(k)))) \quad (6)$$

where $Y(k)$ denotes the input data, $\hat{X}(k)$ denotes the output of the CNN equalizer, and function $f_n(\cdot)$ denotes the convolutional operation in each layer. Specifically, under channel equalization, the input of the CNN is a 1-D vector instead of the 2-D image, common in the field of computer vision; hence, the received complex signal $Y(k)$ must first be separated into real and imaginary parts. Then, we can rewrite the 2-D convolution with rectified linear units (ReLU) into a 1-D form for each convolutional layer $f_n(\cdot)$.

$$h_{n+1,m,j} = f_n(h_{n,c,k+j}) = \sigma \left(\sum_{c=1}^C \sum_{k=1}^K w_{n,m,c,k} h_{n,c,k+j} + b_{n,m} \right), \quad (7)$$

where $h_{n,c,k+j}$ is the input of layer n , $w \in \mathbb{R}^{N \times M \times C \times K}$ denotes the weight tensor of M filters with C channels for N layers, each containing a $1 \times K$ sized filter. In addition, $b_{n,m}$ is the m -th element of the bias vector $b \in \mathbb{R}^{N \times M}$ in layer n , and $\sigma(\cdot)$ denotes the ReLU function $\max(x, 0)$. The 1-D convolution and dilated convolution are depicted in Fig. 4.

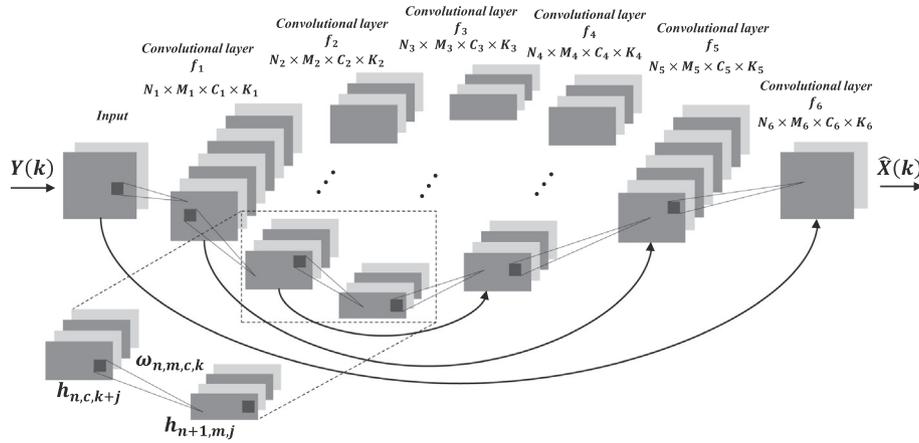


Fig. 3. CNN equalizer.

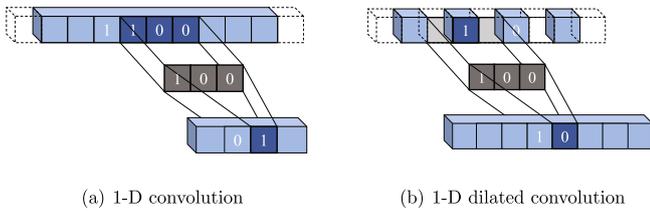


Fig. 4. Implementation of the 1-D convolution and dilated convolution.

In addition, with regard to channel equalization, there is a large amount of low-level information shared between the input and output. Thus, it is desirable to shuttle the information directly across the model. Many low-level details can be lost when reconstructing the transmitted OFDM signal if we force all information to flow through the bottleneck layer. Therefore, we add symmetrical long skip connections following the general shape of a "U-Net" [22] to transfer features from the previous feature maps. This ensures feature reusability with the same dimensionality of the earlier layers. Specifically, each skip connection simply concatenates all channels in layer i with those in layer $n - i$, where n is the total number of layers. In addition, it is easier to optimize the weights, as the gradients can flow deeper through the entire model.

3.2. MLP demodulator

MLP shares a similar feedforward structure with CNN but with neurons that are densely connected to the previous layer in a fully connected layer, as illustrated in Fig. 5. MLP converts the complex demodulated data (e.g., quadrature phase shift keying (QPSK)) into soft bits and represents each bit with two real numbers (e.g., log-Likelihoods of 0 and 1).

Here, we use MLP to demodulate the output of the CNN equalizer, which can be expressed as

$$\hat{S}(k) = g_2(g_1(\hat{X}(k))). \quad (8)$$

The computation of the single layer $g_n(\cdot)$ in MLP can be formulated as the following matrix multiplication:

$$h_{n+1,q} = g_n(h_{n,p}) = \sigma\left(\sum_{p=1}^P w_{n,p,q} h_{n,p} + b_{n,q}\right), \quad (9)$$

where $h_{n,p}$ is the p -th input of layer n , $w_{n,p,q}$ is the weight vector between the p -th input and q -th output, and $b_{n,q}$ is the bias. Note

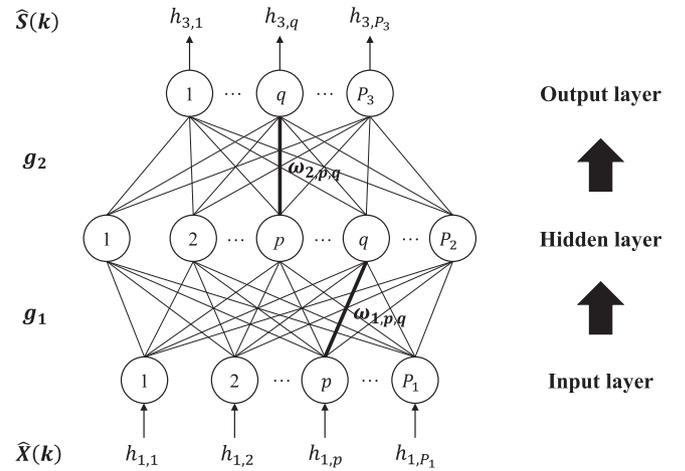


Fig. 5. MLP demodulator.

that $\sigma(\cdot)$ denotes the activation function. The sigmoid activation $(1 + e^{-x})^{-1}$ is applied to limit the output of the MLP to the range $(0, 1)$.

3.3. Training

The performance of a neural network depends greatly on the training process. First, the loss function should be carefully selected to provide an accurate measure of the distance between the model outputs and true labels. In addition, the hyperparameters related to the network structure and training determine the capabilities of neural networks. In this study, the MSE and BER are applied as the total loss functions of the model, where the BER can prevent a diminishing gradient when the MSE is very small.

$$\mathcal{L}_{total} = \mathcal{L}_{MSE} + \mathcal{L}_{BER}, \quad (10)$$

where

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_i (\hat{S}_i - S_i)^2, \quad (11)$$

and

$$\mathcal{L}_{BER} = \frac{1}{N} \sum_i (\kappa(\hat{S}_i) - S_i)^2. \quad (12)$$

\hat{S} and S represent the output soft bits of the MLP model and true information bits, respectively. $\kappa(\cdot)$ denotes the operation of a hard decision, and N is the number of transmitted bits.

Because the learning rate determines the convergence rate of the DL network, a self-adaptive learning rate algorithm, namely, the Adam algorithm, is adopted in our offline training [23]. The Adam algorithm exploits the exponential moving averages of the gradient and squared gradient to scale the learning rate. With its loss function, the Adam optimizer utilizes back-propagation to find optimal parameters that can result in minimal loss.

4. Experiments

In this section, we evaluate the performance improvement owing to the proposed model in an OFDM system. Several experiments were conducted to compare our model with other OFDM receiver schemes for UWA communication, including conventional methods and the FC-DNN model described in [12]. We first present the implementation process and then discuss the results.

4.1. Implementation

The proposed models were implemented on the advanced DL framework *Tensorflow*; the flow graph is presented in Table 1, wherein the parameters and hyperparameters for the model were chosen after optimization.

Other related parameters of the UWA-OFDM system and neural network training used in the simulation are listed in Table 2.

The SNR of the received signal is defined as

$$SNR = 10 \log \left(\frac{\sum_{i=1}^K |S_{signal}(k_i)|^2}{\sum_{i=1}^K |S_{noise}(k_i)|^2} \right), \quad (13)$$

where S_{signal} is the OFDM signal at frequency k_i , S_{noise} is the noise adjacent to the signal, and K is the number of narrowband frequencies.

To demonstrate the performance of our proposed OFDM receiver model, it was experimentally compared with two other approaches:

- [(1)] Traditional UWA-OFDM system with LS channel estimation [21]. This is a common method for pilot-based channel estimation because it offers good performance with reasonable complexity.
- [(2)] FC-DNN model. This is a typical DL-based OFDM receiver design with only fully connected layers; the number of neurons in each layer is 256, 512, 256, 120, and 128, similar to the architecture described in [12].

Table 1

Flow graph of the proposed model.

Layer	Type	Input layer	Activation	Output shape
Input	Reshape	-	-	(1000,128,2)
Conv1	Convolutional layer (8,3,2)	Input	ReLU	(1000,64,8)
Conv2	Convolutional layer (32,3,2)	Conv1	ReLU	(1000,32,32)
Conv3	Convolutional layer (128,3,2)	Conv2	ReLU	(1000,16,128)
Conv4	Deconvolutional layer (32,3,1)	Conv3	ReLU	(1000,32,32)
Conv5	Deconvolutional layer (8,3,1)	Conv4 + Conv2	ReLU	(1000,64,8)
Conv6	Deconvolutional layer (2,3,1)	Conv5 + Conv1	ReLU	(1000,128,2)
Flatten	Flatten	Conv6 + Input	-	(1000,256)
FC1	Fully connected layer n = 256	Flatten	ReLU	(1000,256)
FC2	Fully connected layer n = 512	FC1	ReLU	(1000,512)
FC3	Fully connected layer n = 128	FC2	Sigmoid	(1000,128)
Output	Reshape	FC3	-	128000

The convolutional/deconvolutional layer (C, K, stride). C: channel number, K: kernel size, stride: stride of the convolution.

Table 2

Summary of system parameters.

Parameters	Value
Optimizer	Adam
Learning rate	0.0005
Batch number	1000
Epoch number	60000
Modulation scheme	OFDM with QPSK
Training SNR	5.5:25 dB
Test SNR	0.1:10 dB
UWA channel	<i>WATERMARK</i>

Additionally, to increase the experimental reliability and accuracy, we generated UWA channels for simulations from the *WATERMARK* dataset, which is driven by at-sea measurements of the time-varying impulse response. Here, we chose three typical UWA channels measured in Norway (two sites) and Hawaii, named Norway-Oslofjord (NOF1), Norway-Continental Shelf (NCS1), and Kauai (KAU1&2). Table 3 summarizes the measurement conditions for each channel, while Fig. 6 provides schematics of the deployed platforms.

In our following experiments, the ratio of the training set to testing set was approximately 3:1. Figs. 7–9 represent the time-varying channels sampled from NOF1, NCS1, and KAU1&2, respectively.

In this study, the logarithmic BER was used as a measurement indicator to evaluate the estimation performance for SNR values in a certain interval, which is defined as follows:

$$BER = \log \left(\frac{1}{N} \sum_i (\kappa(\hat{S}_i) - S_i)^2 \right), \quad (14)$$

where \hat{S} and S represent the output soft bits of the proposed model and true information bits, respectively. $\kappa(\cdot)$ denotes the operation of a hard decision, and N is the number of transmitted bits.

To further illustrate the quality of the aforementioned *WATERMARK* UWA channels, the empirical mode decomposition (EMD) described in [24] is utilized to separate the random component from the original channel tap:

$$h_i(t) = d_i(t) + w_i(t), \quad (15)$$

where $h_i(t)$ represents the t -th sample of the i -th UWA channel tap. $d_i(t)$ is called the trend, which can be interpreted as the contribution of the pseudo-deterministic physical phenomena to channel fluctuations. In addition, $w_i(t)$ is a zero-mean wide-sense stationary uncorrelated scattering (WSS) ergodic random process, which represents the channel fluctuations attributable to the scatterers that result in fast fading [25]. It is noteworthy that the conventional EMD method is limited to real-valued time series applications; thus,

Table 3
Measurement conditions of WATERMARK channels.

Name	NOF1	NCS1	KAU1	KAU2
Environment	Fjord	Shelf	Shelf	Shelf
Time of year	June	June	July	July
Range/m	750	540	1080	3160
Water depth/m	10	80	100	100
Transmitter depl.	Bottom	Bottom	Towed	Towed
Receiver depl.	Bottom	Bottom	Suspended	Suspended

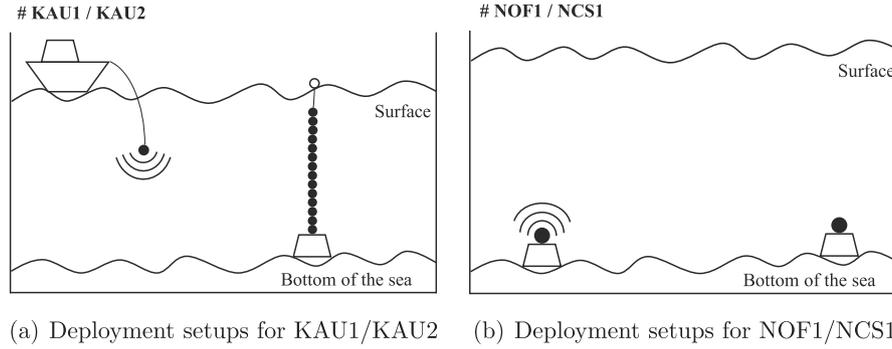


Fig. 6. Schematics of the deployment setups.

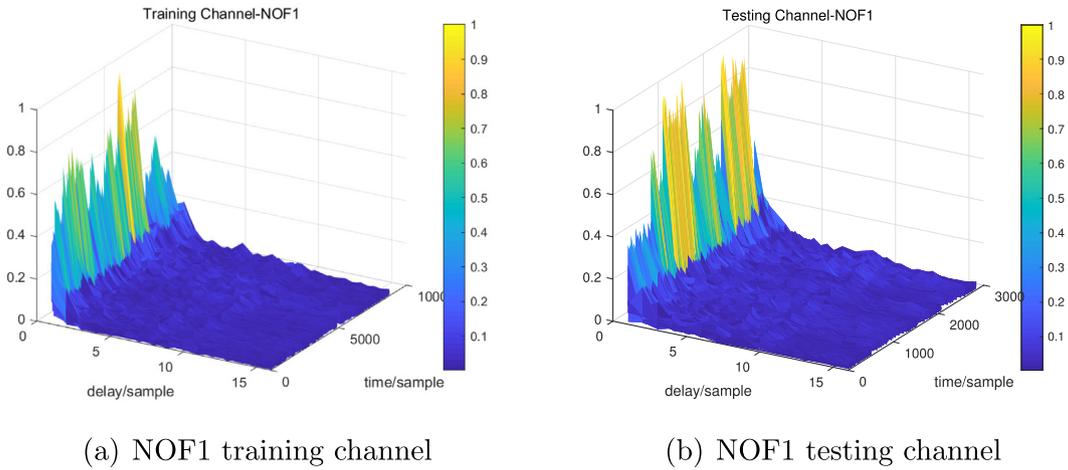


Fig. 7. NOF1 channel impulse response.

we utilize the extension to complex-valued time series proposed in [26] (the EMD-based channel separation detailed in [25]).

Owing to EMD filtering, it is possible to evaluate each UWA channel by calculating the average fade rate (AFR) as

$$AFR = 10 \log \left(\frac{Pow(w)}{Pow(h)} \right) = 10 \log \left(\frac{\sum_{t=1}^T \sum_{i=1}^I |w_i(t)|^2}{\sum_{t=1}^T \sum_{i=1}^I |h_i(t)|^2} \right). \quad (16)$$

Thus, we summarize the AFR values of each WATERMARK channel in Table 4. The NOF1 channel is of high quality because its stable paths carry most of the received signal energy. However, NCS1 and KAU1&2 with high AFR values are more challenging channels, carrying many distinct trailing paths and fluctuating arrivals.

4.2. BER performance under WATERMARK channels

In our study, cross-validation was performed to verify DL-based models for UWA-OFDM communications. The purpose of cross-validation is to choose the optimal model parameters and determine the true prediction performance of a statistical model. Here, we used 3/4 part of the total dataset to fit the model, whereas the remaining 1/4 part was used to test the model. Figs. 10–12 present the BER performance of the proposed scheme compared to the LS-based UWA-OFDM and a five-layer FC-DNN model. More accurate quantitative evaluation results are reported in Table 5. To prevent the interaction of various communication system modules from affecting the results, we did not utilize channel coding and matching filters to further improve the overall communication performance. The simulations were performed using an Intel Core i5 CPU at 1.60 GHz and 8 GB of memory storage.

As shown in Table 5, the proposed model, with a 30% reduction in parameter numbers compared to the FC-DNN model, yields the

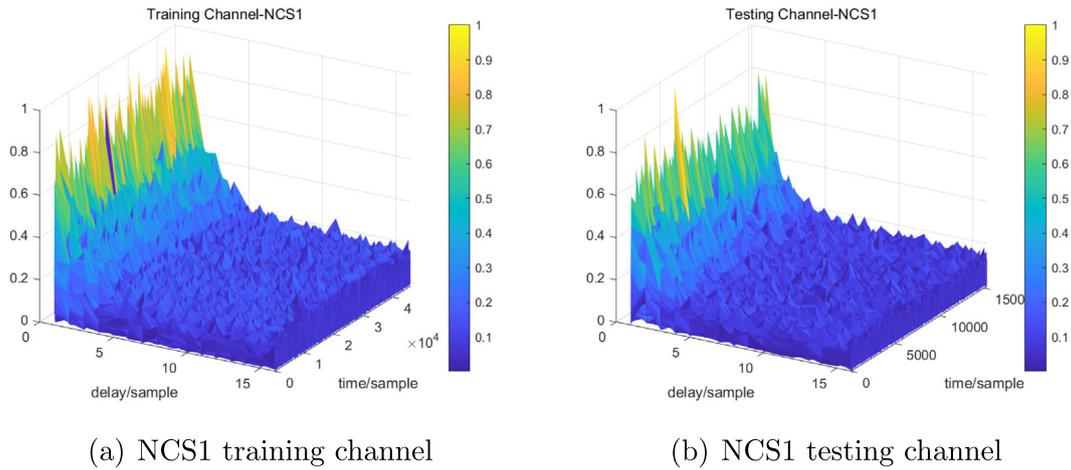


Fig. 8. NCS1 channel impulse response.

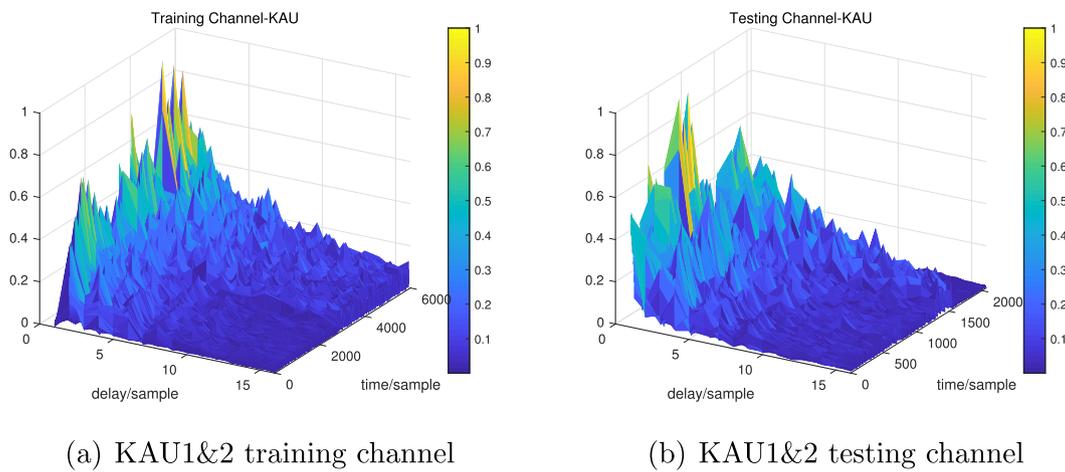


Fig. 9. KAU1&2 channel impulse response.

Table 4
Separation of WATERMARK channels.

Channels	Pow(<i>d</i>)	Pow(<i>w</i>)	Pow(<i>r</i>)	AFR
NOF1	0.0018	0.0006	0.0024	0.2581
NCS1	0.0002	0.0043	0.0045	0.9636
KAU1&2	0.0011	0.0063	0.0074	0.8521

best performance among all methods, while the FC-DNN achieves inferior performance. Traditional UWA-OFDM with LS channel estimation cannot achieve optimal performance. More specifically, in the experiment based on NOF1 channels with an SNR of 10 dB, the BER of the proposed model was 0.22 and 0.02 lower than that of the LS algorithm and FC-DNN model, respectively. In the experiment based on NCS1 channels, the BER performance improvements were 0.24 and 0.21. For the KAU1&2 channels, the improvements were 0.25 and 0.17. As the SNR value decreased, the performance of each algorithm decreased consistently. Although the DL-based algorithm was still numerically superior to the traditional algorithm, there was no better approach to recover the ideal signal from the received signal in practical applications. A series of methods, such as channel coding, must be carried out to further improve the communication performance in low SNR cases.

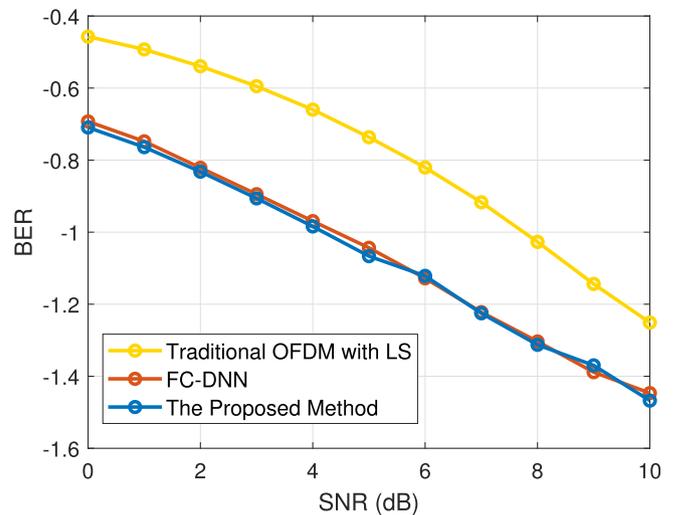


Fig. 10. BER vs. SNR for OFDM receiver systems. The channel is NOF1.

Regarding the complexity comparison of each algorithm, the DL-based algorithm includes offline training and online deployment. When offline training is completed, the model parameters

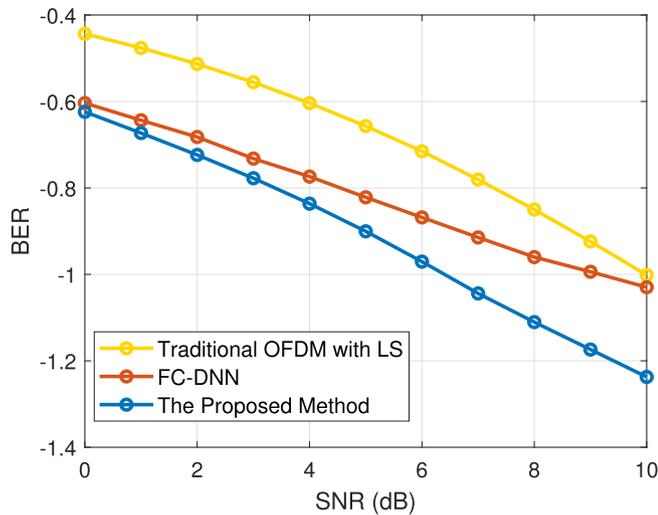


Fig. 11. BER vs. SNR for OFDM receiver systems. The channel is NCS1.

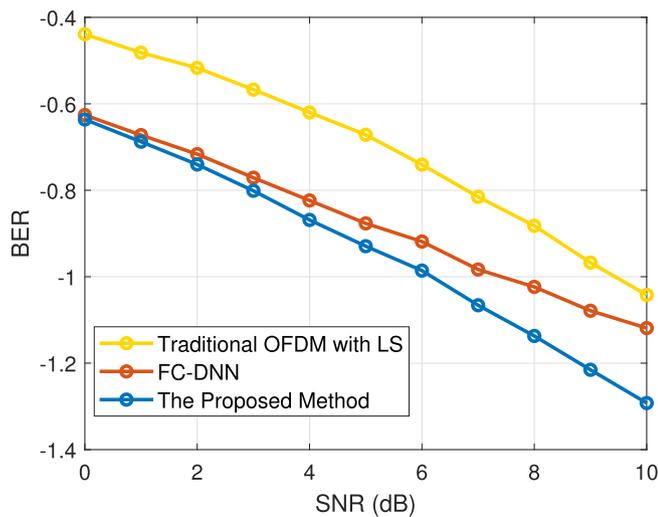


Fig. 12. BER vs. SNR for OFDM receiver systems. The channel is KAU1&2.

are fixed. In the online deployment stage, signal detection is realized by forward propagation through a well-trained DL model. Therefore, the actual calculation amount, as presented in Table 5, is not large, where each runtime on a single OFDM block has the same magnitude.

In summary, the LS method performs poorly under UWA circumstances because it is inversely proportional to the SNR, as demonstrated in Eq. 5. Although it is easy to implement, such simplicity is at the cost of relatively low accuracy. The FC-DNN model is more accurate than the traditional LS algorithm because of its

Table 5
BER and real-time performances for different approaches.

Algorithms	BER (SNR = 10 dB)			Parameter numbers	Runtime (ms)
	NOF1	NCS1	KAU1&2		
LS	-1.25	-1.00	-1.04	-	10.9
FC-DNN	-1.45	-1.03	-1.12	308224	26.9
The proposed method	-1.47	-1.24	-1.29	223152	47.8

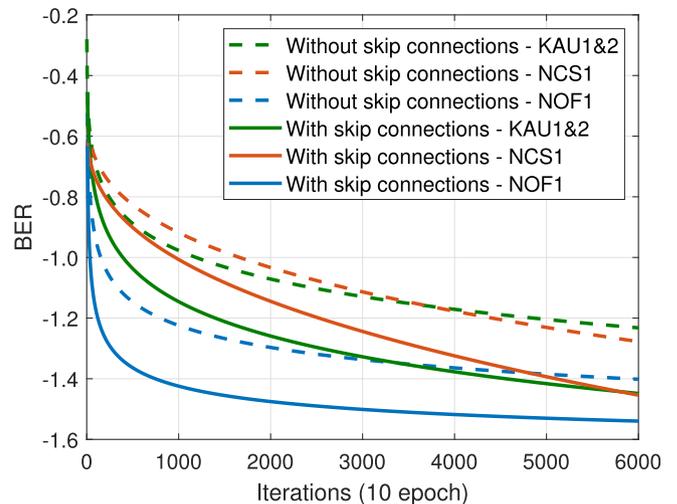


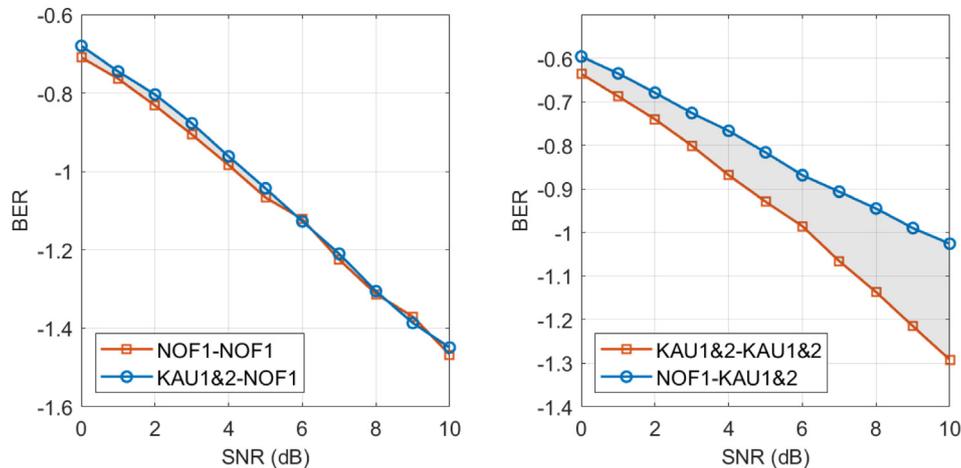
Fig. 13. Convergence performance of the proposed method with and without skip connections.

good fitting ability. However, there are also non-convex optimization and gradient disappearance problems in a fully connected structure, making it less robust when dealing with complex and variable scenarios, such as NCS1 channels and KAU1&2 channels with high AFR values. In contrast, the CNN model proposed in this study has a strong feature extraction ability, ensuring an accurate signal recovery. In addition, the proposed model can save approximately 30% of the storage resources compared to the FC-DNN model.

For real-world applications, it is important that the DL-based model possesses a good generalization ability so that it can work effectively when the online UWA environments do not exactly match the UWA channels used in the training stage. According to the experimental results, improvements are not obvious under NOF1 channels characterized by a simple multipath structure and slow time-varying dynamics. However, in the experiments on NCS1 and KAU1&2, because of the more challenging channel structure and rapid time-varying characteristics according to Table 4, the proposed model can provide significant performance benefits owing to its excellent signal recovery capability. This is in line with the above analysis. A further robustness test of the proposed model under UWA channel mismatches is presented in Section 4.4.

4.3. BER performance with or without skip connections

To further verify the effect of skip connections on the proposed model, we evaluated and compared it under the NOF1, NCS1, and KAU1&2 channels. As shown in Fig. 13, the skip connections in the proposed model can not only improve the accuracy of the fitting task but also effectively accelerate the convergence rate. The results demonstrate that the model with skip connections achieves a significant gain of 0.14, 0.22, and 0.18 under the three above-mentioned channels.



(a) BER curves, testing with NOF1 with and without mismatches (b) BER curves, testing with KAU1&2 with and without mismatches

Fig. 14. BER vs. SNR with mismatches between training and testing stages.

4.4. Robustness analysis under UWA environment mismatches

According to the simulations above, the testing stage UWA channels were sampled from the same environments used in the training stage. Substantial prior information is required to provide a priori guidance for model training. However, channel mismatches may occur between the two stages in real-world applications. Hence, it is essential to study the network behavior when the environmental conditions of the testing stage do not agree with those of the training stage. In this section, the impact of channel mismatches on the proposed model is explained. Here, the model is trained with the KAU1&2 channel and tested with the NOF1 channel. Then, for comparison, we swap the training and testing environments, that is, train the model with NOF1 and test it with KAU1&2.

As shown in Fig. 14a, when the model adapts from KAU1&2 to NOF1, the channel mismatch does not affect the performance of our model significantly; it can still achieve results close to that of environment consistent training, that is, both training and testing with NOF1. On the contrary, when the training and testing environments are exchanged, the performance degradation of the model is evident, as shown in Fig. 14b. We observe a marked decline of 0.27 when we train the model with NOF1 and test it with KAU1&2.

According to Table 4, KAU1&2 is a more complicated channel environment than NOF1. Thus, training in the more challenging channels enables the model to achieve higher robustness and better performance, which is specifically crucial for the proposed model with a better generalization capability. This is the reason why the proposed method achieves better BER performance with channel mismatch from KAU1&2 to NOF1, as shown in Fig. 14a. However, when we adapt the model trained in NOF1 to KAU1&2, the performance of the model declines significantly. This is because the information learned by the model in a channel with a relatively simple channel structure (NOF1) cannot be well adapted to a high-complexity environment (KAU1&2), thus resulting in a certain performance gap. In addition, few-shot learning can be a powerful tool to handle data-starved situations or UWA communication environment mismatches in practical applications. Related research will be conducted in the future.

5. Conclusions

In this study, we proposed a DL-based OFDM receiver for UWA communications, in which a CNN model was innovatively applied

to extract latent features from the received OFDM signals and an MLP was cascaded to demodulate the recovered sequence. The proposed model was trained offline, considering UWA OFDM communication as a black-box problem. After sufficient training, the model was applied as an end-to-end UWA-OFDM receiver. The experimental results based on the measured-at-sea *WATERMARK* dataset demonstrate that the proposed model outperforms traditional UWA-OFDM with LS channel estimation and an FC-DNN-based UWA-OFDM framework. Furthermore, our approach can save almost 30% of the storage resources compared to the FC-DNN model. In particular, the proposed model yields a BER gain of over 0.17 under harsh UWA environments with strong multipath spread and rapid time-varying characteristics. Moreover, the benefits of skip connections in the proposed model were evaluated, the results of which show that our novel structure can not only accelerate the training process but also has approximately a 0.20 BER gain under challenging UWA channels. Finally, a robustness analysis under UWA channel mismatch circumstances was presented, whose results reveal the impact of channel mismatches on the proposed model and can significantly guide practical UWA communication applications.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the China Scholarship Council, Chinese Academy of Sciences and National Natural Science Foundation of China (Grant No. 62171440).

References

- [1] Stojanovic M, Preisig J. Underwater acoustic communication channels: propagation models and statistical characterization. *IEEE Commun Mag* 2009;47:84–9.
- [2] Diamant R, Lampe L. Low probability of detection for underwater acoustic communication: A review. *IEEE Access* 2018;6:19099–112.
- [3] Huang J, Wang H, He C, Zhang Q, Jing L. Underwater acoustic communication and the general performance evaluation criteria. *Frontiers Inf Technol Electronic Eng* 2018;19:951–71.

- [4] B. Li, S. Zhou, M. Stojanovic, L. Freitag, P. Willett, Multicarrier communication over underwater acoustic channels with nonuniform doppler shifts, *IEEE J. Ocean. Eng.* 33 (2008) 198–209.
- [5] Li B, Huang J, Zhou S, Ball K, Stojanovic M, Freitag L, Willett P. Further results on high-rate mimo-ofdm underwater acoustic communications. In: *OCEANS* 2008. p. 1–6.
- [6] Li B, Huang J, Zhou S, Ball K, Stojanovic M, Freitag L, Willett P. MIMO-OFDM for high-rate underwater acoustic communications. *IEEE J Oceanic Eng* 2009;34(4):634–44.
- [7] Ghauri SA, Alam S, Sohail MF, Ali A, Saleem F. Implementation of OFDM and channel estimation using LS and MMSE estimators. *Int J Computer Electron Res* 2013;2(1):41–6.
- [8] Zakharov YV, Morozov AK. OFDM transmission without guard interval in fast-varying underwater acoustic channels. *IEEE J Oceanic Eng* 2015;40(1):144–58.
- [9] O'Shea T, Hoydis J. An introduction to deep learning for the physical layer. *IEEE Trans Cognitive Commun Network* 2017;3(4):563–75.
- [10] Qin Z, Ye H, Li GY, Juang BF. Deep learning in physical layer communications. *IEEE Wirel Commun* 2019;26(2):93–9.
- [11] Dörner S, Cammerer S, Hoydis J, Brink St. Deep learning based communication over the air. *IEEE J Selected Topics Signal Processing* 2018;12(1):132–43.
- [12] Ye H, Li GY, Juang B. Power of deep learning for channel estimation and signal detection in OFDM systems. *IEEE Wireless Commun Letters* 2018;7(1):114–7.
- [13] Zhang Y, Li J, Zakharov Y, Li X, Li J. Deep learning based underwater acoustic OFDM communications. *Appl Acoust* 2019;154:53–8.
- [14] Jiang R, Wang X, Cao S, Zhao J, Li X. Deep neural networks for channel estimation in underwater acoustic OFDM systems. *IEEE Access* 2019;7:23579–94.
- [15] Yang Y, Gao F, Ma X, Zhang S. Deep learning-based channel estimation for doubly selective fading channels. *IEEE Access* 2019;7:36579–89.
- [16] Xu W, Zhong Z, Be'ery Y, You X, Zhang C. Joint neural network equalizer and decoder. In: 2018 15th International Symposium on Wireless Communication Systems (ISWCS). IEEE; 2018. p. 1–5.
- [17] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. p. 770–8.
- [18] van Walree P, Otnes R, Jenserud T. The watermark manual and user's guide. Forsvarets Forskningsinstitut, Horten, Norway, FFI-rapport 2016;1378:2016.
- [19] van Walree PA, Socheleau F-X, Otnes R, Jenserud T. The watermark benchmark for underwater acoustic modulation schemes. *IEEE J Oceanic Eng* 2017;42(4):1007–18.
- [20] M. Stojanovic, P.-P.J. Beaujean, *Acoustic communication*, in: *Springer Handbook of Ocean Engineering*, Springer, 2016, pp. 359–386.
- [21] Scaglione A, Giannakis GB, Barbarossa S. Redundant filterbank precoders and equalizers. i. unification and optimal designs. *IEEE Trans Signal Process* 1999;47(7):1988–2006.
- [22] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.
- [23] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
- [24] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences* 454 (1971) (1998) 903–995.
- [25] Socheleau F-X, Laot C, Passerieux J-M. Stochastic replay of non-wss underwater acoustic communication channels recorded at sea. *IEEE Trans Signal Process* 2011;59(10):4838–49.
- [26] Rilling G, Flandrin P, Gonçalves P, Lilly JM. Bivariate empirical mode decomposition. *IEEE Signal Processing Letters* 2007;14(12):936–9.