Optimizing Underwater Robotic Technology Using Artificial Intelligence

by

© *Shahbad Alam, B. Eng*

A thesis submitted to the

School of Graduate Studies

in partial fulfilment of the

requirements for the degree of

Master of Engineering

Faculty of Engineering & Applied Science

Memorial University of Newfoundland

St. John's, NL.

May, 2025

Abstract

This thesis proposes to optimize the navigation of autonomous underwater vehicles by enhancing underwater acoustic communication and target sensing in complex oceanic environments. Current deep learning techniques in underwater acoustic communication often fail to account for domain knowledge, such as first-principles signal processing and acoustic propagation laws, which govern chaotic underwater environments. The main focus is on the integration of domain knowledge of the underwater environment in training neural networks. Such a context-aware deep learning model employs theory-trained neural network to accurately learn a non-linear map between input and output data. Details of the theory-trained neural network largely remains unexplored, with many open questions.

This research introduces a context-aware methodology for regularizing neural networks. We explore three approaches to deep learning-based communication among AUVs. First, we advance long short-term memory (LSTM) neural networks for accurate underwater target detection that moves stealthily underwater. Second, we embed communication theory within a convolutional neural network (CNN) model in a supervised learning framework. Third, we test a minimally viable intelligent system that enables AUVs to communicate effectively in a hostile underwater environment. Through these venues, we provide a foundation for a more reliable and efficient underwater communication.

Lay Summary

Underwater robotics, particularly artificial intelligence in autonomous underwater vehicles (AUVs), is crucial for remotely sensing the wealth of ocean valued over $24 trillion. Thus, integration of deep learning in AUVs is an extremely demanding research topic to enhance the global underwater robotic market. Training a deep learning architecture to accurately identify a non-linear map using the theory-trained neural network approach remains in its infancy, needing to answer many open questions. Adopting deep learning techniques in underwater acoustic communication cannot account for the vast amount of domain knowledge such as the first-principle signal processing laws and remote sensing with sounds that govern the acoustic propagation in highly chaotic and turbulent underwater environment. Incorporating contextual information can regularize the neural networks reducing the dimensionality of admissible solution space to a size that is suitable for any communication scenario.

This research has focused on addressing the incomplete understanding of the theory-trained deep learning approach to ensure that a swarm of intelligent underwater robots can utilize contextual information to properly communicate with each other. The context-aware methodology of regularization in theory-trained neural network was discussed in the broader deep learning literature but not yet clearly outlined. In this study, we explored three venues for deep learning-based optimal communication between underwater autonomous vehicles. First, we advanced the long-short term memory neural networks for accurate detection of underwater targets. We tested this approach to detect stealthily operating underwater targets. Second, we investigated the performance of theory-trained neural network, where we incorporated communication theory into convolutional neural network in the context of supervised learning. Third, we have tested the design of a minimally-viable intelligent technique so that AUVs can communicate with each other by bypassing the effects of hostile underwater environments.

Co-authorship Statement

This thesis has been organized in the manuscript format, which presents three technical chapters as independent research articles. These chapters focus on the development of efficient and reliable techniques in artificial intellgience (AI) to enhance decision-making techniques in an underwater network of autonomous underwater vehicles (AUVs). The main objective is to advance deep learning technqiues for optimal wireless communication in an underwater environment. The problem of dealing with the hostile underwater environment has been reviewed and presented in Chapter 1. Tackling this problem involves addressing the underwater challenges such as multipath propagation, time-variability of the channel, and Doppler shift. To overcome these issues, three different projects were completed:

I, Shahbad Alam, have the principle authorship status for all manuscripts included in this thesis. The outcomes of this research have been demonstrated via publications and presentations in two conferences, two manuscripts for journal publications, one conference presentations, and one research poster. The list of articles, presentations, and posters is given below:

1. Alam, S., Song, Y., Li, C., Venkatesan (Venky), R. (2025), *Assessment of Underwater Acoustic Target Classification using LSTM Network and Fast Fourier Transform*, IEEE International Conference on Communications (ICC), To be submitted. Conference at June 8-12, 2025, Montreal, ON.

2. Alam, S., Li, C., Venkatesan (Venky), R. (2025), *CNN-based Context-Aware Deep Learning of Underwater Acoustic Channels* , IEEE Access, To be submitted.

3. Alam, S., Li, C., Venkatesan (Venky), R. (2025), *LSTM-based Symbol Detection for Underwater Acoustic Receiver*, IEEE Transactions on Wireless Communications, Submitted.

4. Alam, S., Li, C., Venkatesan (Venky), R. (2024), *Deep learning-based Wireless Control Systems for Underwater Robotic Communication*, Learning and control for robotics

systems in Canadian Applied and Industrial Mathematics Society (CAIMS 2024), June 24-27, 2024, Kingston University, ON, Presented a research talk.

5. Alam, S., Li, C. (2022). *Robust Data Driven Methods for Underwater Object Localization*, The 31st Annual Newfoundland Electrical and Computer Engineering Conference (NECEC), Nov. 15, 2022, St. John's, NL, Presented a research talk and published a paper.

6. Alam, S., Li, C. (2021). *Using Artificial Intelligence for Underwater Localization*. Research Day 2021, MUN, Presented a research poster.

For each chapter, as listed in Table 1, I have contributed independently in coding/data collection, simulations, research, literature review, and writing. I have accommodated whenever I received feedback from my supervisors. For Chapter 2, I have discussed with my colleague, Yuhui Song, received his comments, and included him as a co-author. The co-authorship contributions are listed in Table 1.

Table 1: The contributions of co-authors as indicated by the author's initials: Shahbad Alam (SA), Yuhui Song (YS), Cheng Li (CL), Ramachandran (Venky) Venkatesan (RVV)

	Supervision	Simulations	Research	Revision	Writing	Status
Chap. 1	CL, RVV		SA		SA	Introduction
Chap. 2	CL, RVV	SA	SA	SA, YS, CL, RVV	SA	Conference
Chap. 3	CL, RVV	SA	SA	SA, CL, RVV	SA	Journal
Chap. 4	CL, RVV	SA	SA	SA, CL, RVV	SA	Journal
Chap. 5	CL, RVV		SA		SA	Conclusion

Acknowledgement

I would like to express my sincere gratitude towards my supervisor, Dr. Cheng Li, and co-supervisor, Dr. Ramachandran (Venky) Venkatesan for their advice and support through my graduate studies. I had a long productive journey at Memorial University of Newfoundland where I have also completed my undergraduate studies. I had the opportunity to meet, interact, and learn from Drs. Li and Venkatesan through various courses and other academic activities during my studies at MUN. Without the support of my supervisors, I would not have been able to complete my thesis.

I was interested in gaining knowledge about underwater acoustic communication, artificial intelligence, and deep learning. I am fortunate that Dr. Li opened the door to this fascinating world of academic journey investigating underwater resources, where it is not easy for humans to physically explore. More recently, the 2023 OceanGate accident encouraged me to learn more about underwater robotics.

I am grateful to my colleagues and friends, including Zijun Gong, Yuhui Song, and Xiao-hang Zhao. A big thanks must go to my Mom and Dad for their unconditional love, financial support, and caring for my education and research.

I acknowledge financial support from the School of Graduate Studies fellowship, from the research grants of Dr. Li, and the teaching assistantship.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADAM Adaptive Moment Estimation

ADMM Alternating Direction Method of Multiplier

AI Artificial Intelligence

ANN Artificial Neural Network

AUV Autonomous Underwater Vehicles

CNN Convolutional Neural Network

CS Compressed Sensing

CSI Channel State Information

DFCE Decision Feedback Channel Estimation

DL Deep Learning

DNN Deep Neural Network

FFT Fast Fourier Transform

FrFT Fractional Fourier Transform

LASSO Least Absolute Shrinkage and Selection Operator

IoUT Internet of Underwater Things

LS Least Squares

LSTM Long Short Term Memory

MMSE Minimum Mean Squared Error

MSE Mean Squared Error

NN Neural Network

OFDM Orthogonal Frequency Division Multiplexing

PINN Physics Informed Neural Network

PSD Power Spectral Density

Q-Q Quantile-quantile

QAM Quadrature Amplitude Modulation

QPSK Quadrature Phase Shift Keying

RNN Recurrent Neural Network

SER Symbol Error Rate

SNR Signal to Noise Ratio

UWAC Underwater Acoustic Channel

UWSN Underwater Wireless Sensor Network

Chapter 1

Introduction

1.1 Background

More than 70% of the Earth's surface is covered by water. Oceans are our natural sources of wealth. The asset value of the ocean has been estimated to be $24 trillion [1]. However, 95% of the ocean remains unexplored. There has been an increased reliance on Underwater Robotics so that Autonomous Underwater Vehicles (AUVs) become more intelligent to perform various tasks in the marine environment. The global underwater robotics market value was estimated as $4.49 billion in 2022. Recently, the integration of deep learning techniques in underwater acoustic technology has greatly enhanced these vehicles. Nevertheless, the development of efficient and reliable artificial intelligence (AI) technology has become crucial for enhancing decision-making techniques in a network of AUVs for underwater communication. Figure 1.1 presents a typical structure of underwater acoustic communication networks. Figure 1.1 shows a network consisting of stationary nodes, AUVs, and surface buoys to gather environmental data for exploring the wealth of oceans.

Deep learning is a specialized method in artificial intelligence. In underwater acoustic communication, the deep learning techniques extract critical features from the observed signals to create a realizable AI decision-making system that aims to replicate the neural processing mechanism of the human brain [2]. Deep learning excels in handling complex data, employing

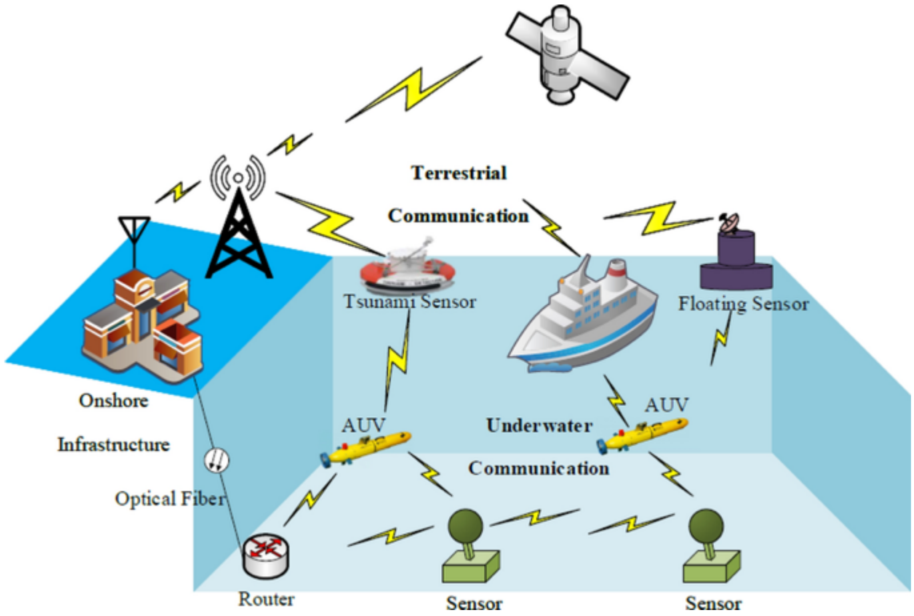


Figure 1.1: An example of underwater acoustic communication network. The stationary sensors, AUVs, and floating sensors can communicate with each other using acoustic links (The image was adapted from the Google Image database).

sophisticated feature detection algorithms, and thus often achieves superior performance. When sufficiently large training data is available, deep learning methods achieve high accuracy in mapping input data to output data. This mapping leads to a universal function approximation, which is the primary building block of deep learning algorithms. Integrating deep learning approach with underwater acoustic communication has the potential to optimize underwater robotic technology.

In this thesis, the primary objective is to develop a deep learning-based underwater robotic technology leveraging acoustic communication. This is a multidisciplinary engineering problem that integrates wireless communication with artificial intelligence. Despite many recent advancement of robotics within mechanical and marine engineering, underwater robotics still faces several communication challenges, such as multipath propagation, time-variability of the channel, and Doppler shift. In underwater robotics, accurate detection of acoustic signals is hindered by these attenuation effects. For instance, signal detection is crucial for networks of AUVs to effectively determine each other's location and perform other data acquisition tasks reliably. Current underwater wireless communication methods require further developments to address the challenges posed by time-variability, multipath interference, and Doppler effects. Therefore, the thesis focuses on the fundamentals of deep learning to mitigate the attenuation effects in underwater signal detection and provide a robust acoustic communication technique for underwater robotics. This includes evaluating the performance of deep learning algorithms in underwater sound classification, channel estimation, and signal recovery in unpredictable underwater conditions.

The underwater wireless communication research community developed sophisticated algorithms for determining location information of underwater objects [3]. Among other applications, such algorithms have the potential for unique subsea intelligence solutions [4]. The ability to precisely detect acoustic signals is crucial for such operations because robots can simultaneously listen to many existing underwater sounds. Such research on underwater wireless communication systems has great importance in various other industrial fields [5] [6] [7], including underwater search-and-rescue [8], marine military operations [9], environmental monitoring [10], etc [11].

In this context, the advancement of deep learning in orthogonal frequency division multiplexing (OFDM) systems holds immense potential for optimizing the field of underwater robotics. In particular, efficient OFDM symbol detection is critical for ensuring reliable communication between AUVs, as it maximizes spectral efficiency and mitigates signal distortion caused by multipath interference in underwater environments. Deep learning-based OFDM symbol detection offers unprecedented opportunities, whereas traditional methods often struggle to accurately localize AUVs and other underwater targets. Developing deep learning techniques to address the distortion of transmitted OFDM symbols in underwater environments is essential for improving sonar imaging, cooperative robotic systems, object tracking, and other key aspects of underwater wireless communication.

1.2 Motivation

The integration of artificial intelligence in optimizing underwater robotics seeks to accurately estimate the communication channel between wireless sensors in order to accurately estimate the locations of AUVs. Conventional techniques, including OFDM and synthetic aperture sonar, face significant challenge due to the distortion of transmitted signals in multipath underwater environments [12]. The development of deep learning techniques can mitigate signal distortion, which is crucial for enhancing synthetic aperture sonar technology [13] and cooperative robotic systems [14], among other underwater wireless communication applications. By leveraging the power of deep learning, we can significantly improve the accuracy and reliability of underwater communications, leading to better navigation, object tracking, and data transmission in challenging underwater conditions. This advancement not only promises to improve existing technologies, but also opens up new possibilities for innovative applications in marine exploration and surveillance.

In classical underwater communication, the advantage of OFDM is that one can address signal distortion in an underwater environment efficiently. OFDM is a robust and versatile modulation technique which achieves high data rates and spectral efficiency. This technique attracted significant research interest due to its potential for enhanced energy and spectral efficiences while mitigating intersymbol interference (ISI) [15]. This thesis aims to incorporate OFDM techniques in three deep learning directions in order to develop cost-effective underwater robots. The first direction focuses on long short-term memory (LSTM) for proper classification of acoustic targets. The second direction focuses on convolutional neural networks (CNNs) for accurate channel estimation in OFDM receivers. The third direction focuses on OFDM symbol detection in the receiver without requiring the statistical knowledge of the channel. The outcome of these three investigations brings impressive capabilities of acoustic communication and artificial intelligence in underwater robots. The overall development of the thesis addresses the channel modelling problems to deal with turbulent characteristics of the ocean environment and loss of coherence due to Doppler spreads [12], surface scattering [16] [17] [18] [19], and randomly moving obstacles [20] [21]. Such characteristics are difficult to resolve because there are no universal channel models that effectively represent underwater acoustic systems.

1.3 Research Contributions and Novelty

Autonomous navigation of any robots must deal with air currents or underwater chaotic conditions. This thesis has focused on the design, modelling, and analysis of deep learning-based communication for underwater robotics. One approach is to utilize the vast majority of available data to train a neural network and implement it for autonomous navigation of robots. This approach is not very efficient for the task of sensing, detecting, and monitoring objects that are farther away from the AUV. This is because the communication channel is drastically affected by the environmental conditions. However, there is no simple way of incorporating such environmental variation into deep learning methods. Context-aware neural network is a recently developed deep learning technique which aims to incorporate domain knowledge or physical context of the system into neural networks. Such a context-aware deep learning technique has the potential to fill in the research gaps and mitigate the demand for navigating a swarm of robots and individual AUVs through underwater sensor networks. This thesis is based on the results of investigation in three projects, which present a methodology of incorporating contextual information of the underwater environment in supervised learning for optimal communications between underwater robots.

The originality and novelty of this research contribution is as follows:

- **Project A**: Many underwater targets operate stealthily, and thus detecting them with classical sonar technique is a challenging endeavour. Underwater target tracking aims to identify whether a received underwater acoustic signal belongs to a certain category of sounds. Deep learning provides powerful classification algorithms that are optimized on the training dataset obtained from certain underwater environments. Project A investigated an underwater sound classification methodology using deep neural network. The long short-term memory network (LSTM) is a deep neural network that effectively learns temporal dependencies in data to address the time-variability of the underwater environment so we can detect stealthily operating vehicles. The study found that learning temporal pattern of sound signals is not sufficient for LSTM to classify stealthily operating vehicles. However, applying the LSTM network to learn spectral energy distribution of the signals can significantly improve the LSTM-based classification. The primary outcome of project A suggest that incorporating contextual information into deep neural networks has the potential to significantly improve deep neural networks to detect stealthily-operating underwater robots. Project B has further investigated context-aware deep neural network for accurately estimating underwater channels by robots.

- **Project B**: In underwater object localization and robot communication, inaccurate channel estimation can lead to false positive object detection. Many recent works suggest that CNN-based deep learning technique is a promising candidate for acoustic communication in highly time-varying underwater channels. However, detailed literature review suggests that a truly fair comparison with the widely used pilot-assisted channel estimation using orthogonal frequency division multiplexing modulation has not yet been considered for underwater acoustic communication.

  Utilizing pilot subcarriers, a sparse channel can be estimated by the least absolute shrinkage and selection operator (LASSO) algorithm. Consider a system under the OFDM modulation with n subcarriers and m symbols. Let $\boldsymbol{h}_{sp} \in \mathbb{C}^{nm}$, $\boldsymbol{x}_p \in \mathbb{C}^{nm}$, and $\boldsymbol{y}_p \in \mathbb{C}^{nm}$ denote the vectorized representation of channel path gains, transmitted symbols, and received symbols at pilot subcarriers, respectively. Here, $\boldsymbol{h}_{sp}$ is a sparse channel vector. Following [22] about sparse channel estimation, the LASSO algorithm minimizes the L_1 norm $||\boldsymbol{h}_{sp}||_1$ using a regularization parameter λ and a sensing matrix $\mathcal{A}$, where

$$\hat{\boldsymbol{h}_{sp}} = \text{argmin}||\boldsymbol{y}_p - \mathcal{A}\boldsymbol{x}_p||_2 + \lambda||\boldsymbol{h}_{sp}||_1. \tag{1.1}$$

In most applications, the number of subcarriers n is much larger than the number of symbols m. Since the complexity of solving Equation 1.1 is $\mathcal{O}(n^3m^3)$ and $\mathcal{A}$ is a sensing matrix with nm rows and nm columns. The choice of LASSO is not feasible for estimating underwater channels in chaotic environments.

This work has addressed such an underwater channel estimation problem by considering a context-aware CNN architecture, where the channel estimation algorithm is adapted to underwater channels in the Doppler-delay domain. We target the optimization of the pilot overhead that exists in classicical OFDM techniques. Although the LASSO algorithm can approximate the optimal solution in a statistical sense, it requires high computational power. In practical underwater conditions, such methods are not reliable to accommodate the multipath delay spreads of an underwater channel.

The CNN-based methodology compensates the computational burden of LASSO through the offline training process. This approach is promising because spreading, absorption, and scattering losses of underwater acoustic channels is sparse and high-dimensional in the Doppler-delay domain. In this context, exploiting channel sparsity with CNN exhibited very good trade-offs between pilot placement, algorithmic complexity, and estimation error. Statistical estimate of reliability and validity has been considered to evaluate the accuracy of the CNN-based channel estimation technique.

- **Project C**: In order to optimize underwater robotic technology using artificial intelligence, traditional approaches require channel estimation in order to accurately detect the OFDM symbols for target tracking. As discussed in **Project B**, the computational cost of underwater channel estimation can be optimized by the CNN algorithm. However, in **Project C**, our goal is to detect the transmitted symbol without performing the channel estimation. This project has thus developed a deep learning-based symbol detection technique for time-varying underwater acoustic channels which does not require any prior knowledge of channel statistics. This work has considered long short-term memory (LSTM) neural networks to classify the signal constellation of the received OFDM symbols. This approach trains the neural network on input-output pairs of pilot OFDM symbols. The trained network detects the transmitted symbol directly from the received symbol without considering the channel state information. The LSTM-based channel estimation result has been compared with that of the LS and MMSE methods. The maximum likelihood decoding technique has been considered to assess the reliability and validity of the LSTM-based channel estimation approach.

1.4 Thesis Outline

Through the outcomes of three research projects, this thesis explores a unified principle for developing context-aware deep learning models to optimize underwater robots. In underwater communication, context refers to the metadata that describes the environments of actual information in relation to a specific task. Traditionally, deep learning methods aim to extract critical features from training data while assuming everything is a similar type of information. Existing literature has recognized that feature extraction omits semantic differences, leading to interpretability issues. This thesis formulates the problem of context-awareness in the design of network architecture for underwater robots. Three research projects investigated the role of signal processing knowledge in developing a context-aware deep learning method for underwater robotics. The thesis consists of five chapters. Chapter 1 has reviewed the relevant work and summarized the contribution of this thesis. Chapters 2-4 have summarized the research findings in three directions, which are distinct but essential for improving the underwater robotic technology. These three chapters have been organized in manuscript format, with the results submitted for publication. Chapter 5 has summarized present research and discussed potential future direction of advancing the deep learning approach for exploring the wealth from oceans.

Chapter 2

Assesment of Underwater Acoustic Target Classification using LSTM Network and Fast Fourier Transform

The classification of underwater acoustic targets can optimize underwater target localization tasks. This study investigates the performance of long short-term memory network combined with fast fourier transforms in classifying underwater sounds. This study investigated the deep learning-based signal classification to design an optimal underwater target localization system. The numerical experiments compared the temporal learning performance of memory-enhanced neural networks with sequence learning performance. We observed that decoupling the time history of sound with the Fourier transform results in a sequence extracting the spectral content of the sound. In particular, for three test signals, the performance accuracy is increased by 51.53%, 62.22%, and 66.91% when the neural network model is trained on the frequency domain instead of the time domain.

2.1 Introduction

In various ocean exploration projects, automated unmanned vehicles (AUVs) descend up to a depth of $4\,000$ m [7, 9]. Some AUVs operate stealthily and gather information about their surroundings. However, the cavitation sound created by the engine forms an acoustic signature of these targets. In a reliable acoustic communication, extracting contextual information from the time and frequency domains is necessary to ensure that sensor nodes in the network can identify the sound source to optimally localize such silent targets. If a target were operating stealthily to gather information about its surroundings, probe signal-based echo analysis technique may not be an optimal strategy for such target localization. Effective AUV communication would classify whether the sound came from the target before extracting the location information from the received acoustic data. For example, the OceanGate submersible went missing in the Atlantic Ocean in June 2023 while descending towards the wreck of the Titanic. In such a search operation, sound classification is crucial to optimize the cost of localizing the missing submersible because there were many other potential sources of underwater sound. Acoustic classification techniques can determine whether the sound came from a target, which in the case of the OceanGate incident was whether the recorded banging noise came from the missing submersible. Recent studies on deep learning (e.g. [23, 24, 25, 26, 27, 28, 3, 15]) have opened a new approach with long short-term memory (LSTM) networks, which is potentially promising to optimize underwater signal classification and underwater target detection [29, 30].

In acoustic signal classification, maximum likelihood methods estimate parameters of a likelihood function, which calculates the probability that a received signal belongs to a specific class, such as environmental noise, marine life sounds, or submarine activity. In contrast, deep learning techniques take a data-driven approach, extracting critical features from received signals and often provide relatively effective decision boundaries for acoustic classification. In underwater scenarios where AUVs move silently, the deep learning approach would be relatively useful for classification of passive acoustic signals like cavitation noises from a target.

LSTM networks can capture long-term temporal dependencies, and thus suitable for dealing with temporal variability typical in acoustic signals. Note that LSTM networks have already demonstrated success in various time-varying data classification problems, such as ECG signal analysis and early stress prediction from wearable sensor data. Such a capability of LSTMs suggest their potential for acoustic signal classification in complex underwater environments.

Several studies demonstrated promising results in underwater acoustic classification by leveraging neural networks to map input signals directly to output labels/classes [31, 32]. In fact, the application of neural networks in underwater acoustics dates back to the early 1990s [33]. More recently, CNN-based deep learning methods have gained popularity in underwater communication particularly due to the ability to extract accurate features from the undersampled spectra [34, 35, 36, 37]. Works such as Gong et al. [3] presented a CNN-based classification model to take advantage of the reduced complexity offered by the discrete fractional Fourier transform (FrFT). However, such a CNN-based technique overlooked the time-varying multipath effects in a turbulent underwater channel, including changes in temperature, pressure, and the Doppler shift [38]. To address these limitations, this study evaluated the classification performance of the LSTM networks, which are well suited for the highly dynamic nature of underwater channels.

For relatively complex deep learning models, recent studies demonstrated a strong generalizability by progressively improving system performance through task-oriented learning and capturing hidden information from data [39, 26, 27, 28, 3, 15]. On the other hand, the law of parsimony suggests that simpler models, such as LSTM networks, that achieve similar level of accuracy to complex models, are better suited to unseen data. Underwater environments demand for developing generalizable network architecture for acoustic source localization and detection, where sound waves reflect off the seabed, sea surface, and other underwater objects. Considering these factors, simpler deep learning models may exhibit higher generalizability performance, if such models account for a set of characteristics such as time-varying multipath propagation, complex aquatic noise, and Doppler shift [26].

The following Sections have briefly reviewed recent progresses in deep learning-based sound source localization techniques. The basic concepts of classification via neural networks has been outlined. The Fourier domain of the signals have also been discussed with regards to their quality as training data into the LSTM model. This discussion includes simulation results and the significance of these findings.

2.2 Deep Learning for Underwater Sound Classification

2.2.1 Problem Definition and Proposed Solution

Let $x(t)$ denote the acoustic signature emitted from an underwater target. As sound propagates through the ocean, the emitted signal is significantly affected by the multipath environmental effects such as reflection off the seabed, sea surface, and other objects. The intercepted acoustic signal, $y(t)$, is a convolution of $x(t)$ and the chaotic underwater environment. Thus, channel estimation becomes crucial for accurate classification of $x(t)$ and detecting the underlying targets. The channel estimation adds additional computational complexity to the classification of $x(t)$. The conventional sound classification framework depends heavily on critical features of $x(t)$, and thus the performance can deteriorate significantly if the classifier focuses on features of $y(t)$ to avoid the cost of channel estimation.

2.2.1.1 Sound Propagation

There exists a map between $x(t)$ and $y(t)$. The map is given by a superposition of the convolution between the channel state information $h(t, \tau)$ and the delayed and attenuated copies of $x(t)$, with the intercepted signal represented as

$$y(t) = \sum_{p=1}^{P} h_p x(t - \tau_p) e^{j2\pi\nu_p t} + \varepsilon. \tag{2.1}$$

Here, the time-varying channel of path p is represented by h_p, the total number of paths is represented by P, the acoustic signal transmitted from the underwater target is represented by

$x(t)$, and the amount of noise in the underwater channel is represented by ε. The Doppler shift is represented by $\nu_p = v \cos \phi_p f_c / v_c$, where v denotes the relative speed between the sound source and the receiver, v_c denotes the sound speed through water, f_c denotes the carrier frequency, and ϕ_p denotes the angle of arrival for path p. The multipath delay and attenuation effects are represented by the channel impulse response

$$h(t, \tau) = \sum_{p=1}^{P} h_p \delta(t - \tau_p(t)), \tag{2.2}$$

where the propagation delay of path p is represented by $\tau_p(t) = \tau_p - \nu_p t$ [3].

2.2.1.2 Feature Learning

We propose a sonar system to automatically discover a representation of $x(t)$ satisfying Equation (2.1), which is employed to learn the necessary features of $x(t)$. It is mathematically and computationally convenient to assign an appropriate label to $y(t)$, where $x(t)$ is unknown. This approach does not require channel equalization. A device can be trained in any real underwater condition with respect to some known sound sources.

2.2.2 Signal Classification with Neural Networks

According to the convolution theorem, the Fourier Transform converts the convolutional Equation (2.1) into a product of the Fourier Transforms of $x(t)$ and $h(t)$, respectively. Thus, we have

$$\mathcal{H}^{-1}\mathcal{Y} = \mathcal{X}. \tag{2.3}$$

Let $W \approx \mathcal{H}^{-1}$ denote an approximation of the inverse of the channel matrix. In classification theory, our goal is to predict a class label for $x(t)$. Thus, the problem is broken into two stages. The inference stage employs a set of training data $\{\mathcal{Y}_i\}$ to estimate W which maximizes the probability of $W\mathcal{Y}$ to fall into a desired category. The subsequent decision stage employs posteriori probabilities to make optimal class assignments. In the present study, we considered the highest probability of representing certain features of $\mathcal{X}$ as the decision criteria. These features are represented by the class label, and thus, the classification framework reduces the computational complexity bypassing the classical feature extraction process. In other words, the tasks of channel estimation and feature extraction are replaced with the estimation of W.

In this research, we employed the long short-term memory (LSTM) neural network architecture, which is a special type of recurrent neural networks (RNNs). Our goal is to estimate the network weights W using a set of measurements of $y(t)$ as training data. LSTM can learn an optimal relationship between input vector $\boldsymbol{x} \in \mathbb{R}^m$ and output vector $\boldsymbol{y} \in \mathbb{R}^n$. Here, the output is a categorical variable $y \in \mathcal{C}$, where C is a set of possible target labels. LSTM consists of six steps:

- Forget gate: $f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$
- Input gate: $i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$
- Candidate cell gate: $\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$
- Cell state update: $c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$
- Output gate: $o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$
- Hidden state update: $h_t = o_t \odot \tanh(c_t)$
- LSTM Output: $y_t = \text{softmax}(W_y h_t + b_y)$

We see that a linear transformation converts the input vector $\boldsymbol{x}$ through multiple hidden states. A non-linear function of the form $\boldsymbol{y} = \sigma(\cdot)$ maps the hidden vector to the output vector $\boldsymbol{y}_t$. In other words, the LSTM network provides the best recurrence relation of a signal from one time step to the next in order to accurately predict future states of the signal. The above LSTM architecture is demonstrated schematically in figure 2.1 to indicate the methodology of accurately learning the recurrence of a signal in the time-domain so that one can predict the future state of the signal.

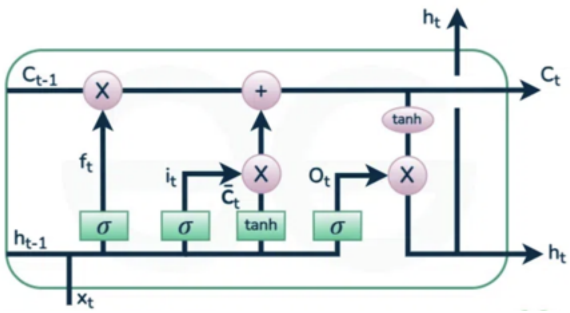


Figure 2.1: A schematic illustration of the LSTM network.

2.2.3 Classification Algorithm

In the proposed LSTM-based signal classification method, the network produces an output $y_t = \text{softmax}((W_y h_t + b_y))$, which maps the observed signal sequence x to a label $y \in \mathcal{C}$. The network is trained by minimizing the mean squared error between the predicted output and the true label:

$$E = \frac{1}{2}||\text{softmax}(\boldsymbol{W}_y \boldsymbol{h}_t + \boldsymbol{b}_y) - \boldsymbol{y}_t||^2, \tag{2.4}$$

where h_t is the hidden state at step t in the frequency domain, and y_t is the corresponding target label. For the results presented in this paper, the weights and biases were optimized through backpropagation using stochastic gradient descent. This optimization process allows the LSTM network to learn the optimal weights and biases for recognizing the power spectral density of discrete-time signals.

In real-time underwater localization systems, the computational load is handled through offline training via the backward propagation algorithm. During the detection stage, the only computational cost is due to the forward propagation. Apart from the computational gain, the LSTM method provides an efficient technology for the identification of underwater targets, particularly those that operate stealthily. An intuitive idea is to inject frequency-domain information, thereby enhancing the performance of the LSTM-based signal classification.

2.3 Simulations and Results

We considered a numerical study for underwater sound classification in order to evaluate the deep learning-based underwater acoustic communication methodology. In this article, our goal is to verify that LSTM network improves the classification accuracy by utilizing the power spectral density of the received sound. During this study, we did not have access to a real-world dataset of underwater sound. Any real-world data, such as acoustic signals generated by a ship, is behind a paywall or otherwise confidential. Thus, we instead considered a dataset obtained from `MATLAB datastore` which consists of six classes of underwater sounds. The dataset include 1000 sound signals for each class of underwater sound. Each signal has a duration of 4 s with a sample frequency of 8.192 kHz. In this experiment, we verify that the LSTM network accurately classifies underwater sound using the power spectral density.

2.3.1 Temporal vs. Sequence Learning

Figure 2.2 shows sound signals from each of the six classes. We see that class 2 and class 5 do no have an apparent pattern that can efficiently be learned by a memory-enhanced neural network such as LSTM. In other words, temporal learning with LSTM has a high chance of misclassification. Conversely, the characteristic features of the sound emitted from a target are hidden behind the time-domain pattern, which is caused by a convolutional kernel. Figure 2.3 shows the power spectral density of the aforementioned signals in each class. According to the convolution theorem, the Fourier transformed signal accounts for the channel state information as a multiplication with the original sound. These plots indicate that the training of LSTM in the Fourier domain, which is a sequence learning approach, is less likely to misclassification.

In the following Section, we compared the performance of sequence learning over the temporal learning cases. A goal of this study is to show a design a decision strategy for an underwater surveillance system, which employs the sound classification technique to optimize the efficiency of target localization.

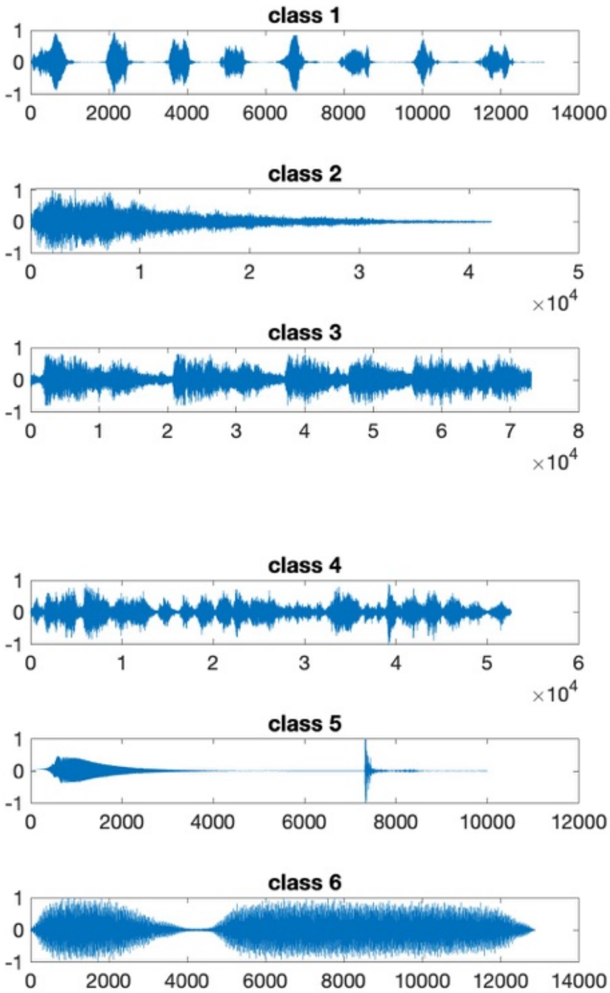


Figure 2.2: The time-domain representation of sound signals $y(t)$ from each of the six classes, where the underlying pattern was convolved with an acoustic channel. The training data consists of 1000 such signals per class.

2.3.2 Training and Validation Performance

The neural network considered for this experiment is trained using an Apple M1 Pro chip with 16GB of memory and implemented in MATLAB R2024a. This neural network considers the ADAM algorithm for optimization with a maximum of 50 epochs and a batch size of 200 data instances. The network parameters are updated during training with a learning rate of 1×10^{-3} and a weight decay of 1×10^{-4}.

In this experiment, we demonstrate the classification performance of a neural network with

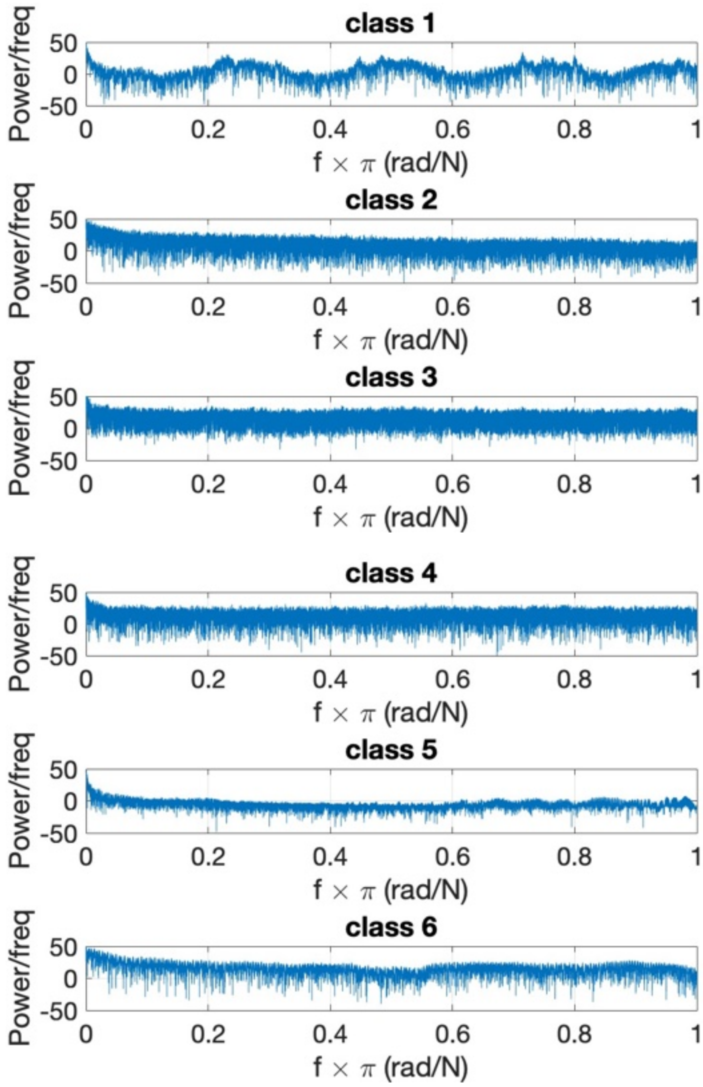
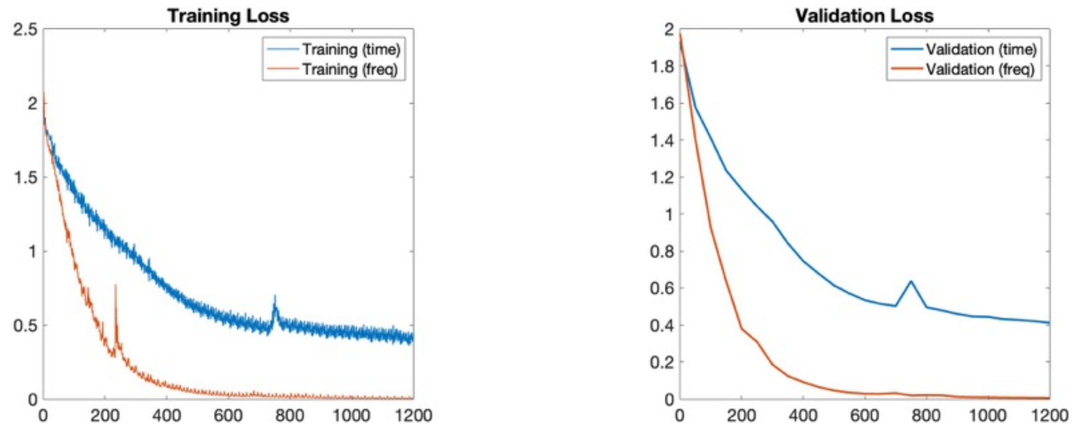


Figure 2.3: The fourier-domain representation of the signals shown in Figure 2.2.

respect to the training data that is passed into it. To demonstrate the effect of the training data, we consider the underwater signal classification in both the time-domain and the frequency domain. To quantify the benefits of considering the signals in the Fourier domain, we perform acoustic signal type classification. Table 2.1 compares the accuracy of the signal classification for data in the time- and frequency-domain, respectively. For each type of training data, we have the corresponding classification confidence scores of the predicted class on the test dataset. We can also see the effect of the type of training data (i.e. time-domain vs. frequency-domain) on the training performance of the neural network, as shown in Figure 2.4. For each class of acoustic signals, the confusion matrix in Table 2.1 shows higher classification confidence scores for training data in the frequency-domain than in the time-domain.



(a) (b)

Figure 2.4: A comparison for the convergence rate of the ADAM optimizer. (a) Training loss vs. number of iterations. (b) Validation loss vs. sampling index. The convergence rate of time-domain learning (blue curves in a and b) is relatively high compared to spectral feature learning (orange curve in a and b).

2.3.3 Discussion

An observation of Table 2.1 shows a difference in confidence scores for certain test signals when classified under a neural network trained on data in the time-domain and frequency-domain. This Table clearly shows that time-domain signals are generally less likely to have accurate predictions, compared to frequency-domain signal. This is exemplified by *Signal 2*. Though this signal has been correctly classified in the time-domain, the neural network had a mere 33.05% confidence that the assigned class label is true. In contrast, the same signal in the frequency-domain is assigned the correct class label by the neural network with a 99.96% confidence. The same logic applies for *Signals 4* and *6*, of which their time-domain form is assigned the correct class label with a mere 37.63% and 48.41% confidence, respectively. In contrast, their frequency-domain form is assigned the correct class label with a 99.85%

Table 2.1: A Comparison of Underwater Signal Classification Accuracy. (A) LSTM network was employed to learn the temporal history of the training data in the time-domain, where the signals were convolved with the channel. (B) The classification outcome of the LSTM architecture, where the network was employed to learn the spectral history of the training data in the frequency-domain, where the channel effect was decoupled from the sound signals. The diagonal elements state the classification accuracy.

A - Time-domain training						
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
S1	**0.9687**	0.24	0.008	0.010	3.35E-5	0.0043
S2	0.008	**0.3305**	0.053	0.016	6.51E-5	0.0099
S3	0.006	0.41	**0.9349**	0.0039	2.03E-5	0.0018
S4	0.015	0.011	0.002	**0.3763**	6.06E-6	0.50
S5	2.09E-4	3.9E-4	2.62E-4	4.47E-5	**0.9999**	7.32E-5
S6	0.002	0.010	0.0016	0.59	4.43E-6	**0.4841**
B - Frequency-domain training						
S1	**0.9984**	3.15E-7	1.59E-5	3.61E-7	0.0015	4.62E-5
S2	7.29E-6	**0.9996**	5.27E-5	2.90E-4	1.12E-5	5.15E-6
S3	1.87E-4	1.30E-4	**0.9995**	2.067E-4	7.63E-6	1.28E-4
S4	5.48E-6	1.65E-4	8.97E-5	**0.9985**	1.55E-4	2.08E-4
S5	3.64E-4	1.29E-6	2.39E-6	1.33E-4	**0.9975**	1.85E-4
S6	0.0010	1.25E-4	3.42E-4	8.69E-4	7.83E-4	**0.9994**

and 99.94% confidence, respectively.

A comparison made within Table 2.1 illustrates the importance of the training data passed into the neural network. It is clearly shown that the LSTM network is trained more easily on frequency-domain networks than on time-domain networks. This discrepancy in performance is also illustrated in Figure 2.4, which shows the training and validation curves to be more optimal for neural networks trained on frequency-domain signals than on time-domain signals.

2.4 Conclusion and Future Direction

2.4.1 Summary

This study investigated the potential of deep learning for signal classification in a harsh underwater environment. We employed LSTM networks for the training of our deep learning model because underwater signals rely on a strong temporal history due to multipath Doppler effects. However, the convolution between the hostile underwater channel and the desired signal hinders LSTM networks from efficiently learning signal features. Past studies have demonstrated frequency-domain classification using LSTM networks in applications such as clinical diagnostics [40], renewable energy [41], music classification [42], and time-series analysis [43]. The present study aims to account for the channel effects, where we considered the Fourier transform as contextual information and incorporated that into the training data. In other words, the consideration of frequency-domain signal features results in the time history of the training signals to be lost [39]. Nevertheless, our results show that training in the frequency domain significantly improved the classification accuracy compared to training in the time-domain. These results suggest two novel aspects of deep learning-based underwater target localization. First, channel equalization and capturing the critical signal patterns are essential for accurately classifying underwater sounds. Second, the LSTM networks are more efficient in sequence learning if the signals are not convolved with secondary effects such as an underwater channel.

2.4.2 Future Research Directions

In this research, we have not investigated the performance of deep learning-based underwater target localization methodology. Many underwater targets typically move stealthily without making a sound, except for the noise from their engines. Such targets listen to the acoustic signatures of surveillance systems. The strategy of analyzing echos of probe signals is not appropriate for localizing such targets. If we can accurately classify such stealthily operating targets via their acoustic signatures, we can apply the time-reversal mirror technique [44] to obtain the desired location information. Such a target localization technique employs the acoustic wave Equation with a point sound source that can be estimated by the time-reversal method. It is worth mentioning that the time-reversal method can accurately recognize the target using wave Equation if the target is located 1500-2000m away.

Chapter 3

CNN-based Context-Aware Deep Learning of Underwater Acoustic Channels

Recent works have studied the performance of convolutional neural networks for underwater acoustic channel estimation and target tracking. However, a full understanding of convolutional neural networks for optimizing the signal processing techniques in the widely used orthogonal frequency division multiplexing modulation requires further investigation. In this research, we address the problem of channel estimation by considering pilot schemes as incomplete measurements of channels. In this context, we present a comparison between deep learning-based channel estimation and decision feedback equalization of channel matrices. The error estimation shows that the proposed methodology performs better than the decision feedback equalizer that uses the full knowledge of the channel statistics. The results also indicate that our method can be used efficiently in localization and tracking of underwater targets. This work provides a foundation to extend convolutional neural networks in underwater acoustic communication.

3.1 Introduction

In various underwater applications, accurate channel estimation plays a significant role to overcome the challenges of detecting underwater targets using wireless sensor networks [3]. Underwater acoustic channel is extremely hostile due to multipath attenuation, strong Doppler effects, and large ambient noise. Although orthogonal frequency division multiplexing (OFDM) is highly efficient in terrestrial communication, it can be challenging to fully exploit the potential of the OFDM technique in underwater surveillance systems. There has been extensive research on the conventional pilot-based estimation techniques such as least squares (LS) and minimum mean squared error (MMSE). These methods are not efficient to directly apply to underwater OFDM communication systems because they do not account for the sparsity and the random multipath channel variations. Exploiting the sparsity and the Bayesian probability of the underwater channel, the decision feedback channel estimation (DFCE) technique can recover the channel with a sampling rate lower than the Nyquist sampling rate. Apart from these approaches, the deep learning-based methods have recently emerged as effective underwater channel estimation techniques.

A major challenge for the deep learning-based underwater acoustic communication is to integrate domain-specific knowledge or physical laws into the training process and the generalizability of trained models to unseen underwater channels for any values of the signal-to-noise ratio (SNR). The lack of training data due to limited measurements of real underwater channels and the generalizability of the machine learning model hindered the applications of deep learning-based techniques in underwater acoustic communication. To address these challenges, recent works developed theory-trained neural networks, which is similar to physics-informed neural networks [45]. This approach eliminates the need for large amounts of training data and exhibited promising attributes such as generalizability, interpretability, and robustness. In this study, we extend the application of theory-trained neural networks in underwater acoustic communication. The fundamental idea is to train the network using pilot symbols and incorporate domain-specific theoretical knowledge using a mathematical model of underwater channels.

An important contextual-expert knowledge is that the underwater acoustic communication channel is influenced by the intermittently time-varying underwater environment, where the channel matrix $\mathcal{H}$ is highly sparse in the Doppler-delay domain. Therefore, exploiting a missing data recovery technique is most likely an optimal approach to estimate $\mathcal{H}$, particularly for the task of underwater target localization through large-scale wireless sensor networks [46]. There have also been extensive work on using channel autocorrelation matrix into the minimum mean squared error (MMSE) method and sparsity into the compressed sensing (CS) method. Apart from these classical approaches of incorporating contextual information, deep learning-based methods have also considered incorporating prior distribution in the channel estimation problems.

In this research, we have developed a novel deep learning architecture to recover the channel matrix $\mathcal{H}$ by considering the pilot scheme as partial measurements of the ocean acoustic channel impulse response. In localization and object tracking problems, such a critical channel state information is essential to perform coherent detection. Generally, pilots are arranged by following a pattern that is known to both transmitter and receiver (e.g. Gaudio et al. [22]). In any communication scenario, both the pilot pattern and the channel estimation algorithm have to be optimized. To achieve a high communication rate, there is a need to optimize the amount of useful information sent in a block of pilot symbols. The optimization of this tradeoff is generally non-trivial. Existing techniques require a full knowledge of the channel statistics, which is difficult for underwater acoustic communications. The proposed deep learning methodology employs the convolutional neural network (CNN) to extract critical features of the CSI. The training of the CNN architecture requires known channel matrices $\mathcal{H}$ to find optimal values of the neural network parameters. In other words, the proposed methodology builds upon the work of Gaudio et al. [22], which introduced the integration of feature extraction with regression of missing elements in $\mathcal{H}$. Unlike their approach, our method incorporates CNN-based sensing of a sparse channel $\mathcal{H}$ in the Doppler-delay domain.

An important aspect of deep learning is to view the estimation of the channel time-frequency response as two step process. The first step involves training a neural network model for the desired task. The training step is an offline process which requires a relevant dataset. This is a computationally intensive process, and it can be performed in such a way that the resulting neural network to map the transmitted signal with the received signal. Note that the performance of such a deep learning-based channel estimation depends on the availability and suitability of the training data [27]. The second step involves using the trained model to execute the appropriate task (i.e. estimating the channel, symbols, or target locations). Several recent works have discussed the importance of understanding the fundamentals of deep learning-based prediction for underwater wireless communication. There are known challenges that can possibly lead to failure of such predictions. Nevertheless, by incorporating the domain knowledge of wireless communication, we can address any forth-coming challenges.

3.1.1 Related Work

Over the years, various applications in wireless sensor networks developed promising techniques to restore incomplete and noisy matrices [47] [48]. Robust principal component analysis is one of the powerful techniques, which is devoted to the noise removal and imputation of missing values of any noisy convex set [49]. Mean imputation is another cost effective approach that uses mean of the non-missing entries to replace the missing values [50]. The Alternating Direction Method of Multiplier (ADMM) algorithm exhibits the best performance in missing data recovery by jointly exploiting sparsity and low-rank properties of the data matrix [51]. Recently, algorithms based on convolutional neural networks have also achieved high performance in matrix completion problems [27] [52].

Acquiring accurate CSI with conventional channel estimation techniques, such as least squares (LS) and minimum mean square error (MMSE) is challenging due to high-dimensionality of channels [53] [54]. Methods that adopt the concept of compressed sensing and channel sparsity in the time-frequency domain have also been considered [55] [56] [57] [22]. Recent studies have shown that we can optimize traditional channel estimation methods by incorporating deep learning techniques [26] [58]. Several works have applied the deep learning approach for identifying signal constellations [59], detecting symbols [60], estimating channels [61], and isolating moving objects [3]. More specifically, the deep learning method can recover the optimal CSI while minimizing the effects of channel distortion [26], intersymbol interference [62], and noise [63].

An end-to-end deep learning architecture for signal transmission/reception was proposed by Ye et al. [26]. In this approach, all functionalities of a wireless communication link are treated implicitly with a trained deep learning framework, thereby treating the entire communication system as a black box. Thus, such a deep learning method aims to avoid the channel time-frequency response, which is otherwise important for applications such as object localization [3]. He et al. [61] and Soltani et al. [27] considered the channel matrix as an image and applied a CNN-based image processing technique for channel estimation.

3.1.2 Our Contributions

In this research, the main idea is to achieve a robust channel estimation by exploiting matrix completion methods as follows:

$$\mathcal{H} = f_\sigma(\tilde{\mathcal{H}}). \tag{3.1}$$

Here, $\tilde{\mathcal{H}}$ belongs to a latent space and $f_\sigma(\cdot)$ is a nonlinear function. The proposed algorithm consists of two steps. The first step learns the latent representation $\tilde{\mathcal{H}}$ by employing the pilot symbols. The second step learns the nonlinear function $f_\sigma(\cdot)$ by employing a deep CNN architecture. Thus, we have developed a novel methodology using a deep CNN architecture to recover the channel state information from pilots by completing a matrix in an end-to-end manner.

In particular, to develop a CNN-based framework for channel estimation in OFDM time-frequency grid, we assume that the pilot symbols provide an incomplete channel (i.e. an undersampled matrix). However, completing the channel with any efficient matrix completion algorithm may not be sufficient to capture the environmental conditions necessary for accurate underwater communication. We propose to incorporate such contextual information by employing an encoding technique, i.e. unsupervised learning of the critical features in the latent space of the incomplete channel matrix, (i.e. $\tilde{\mathcal{H}}$). Then, we consider a supervised learning, which decodes $\tilde{\mathcal{H}}$ through convolutional layers, thereby reconstructing the clean channel matrix $\hat{H}$. We have shown that the performance of the proposed CNN-based channel estimation algorithm is better than that of conventional methods such as decision feedback channel equalizer (DFCE).

We have surveyed the existing literature to show that our contributions have addressed a fundamental research gap. Some of the recent works (e.g. Ye et al. [26] and Mohammed et al. [58]) have conducted deep learning-based channel estimation and OFDM symbol detection. These approaches are not directly applicable to object localization and detection problems where the entire time-frequency spectrum of the channel is required (see Gong et al. [3]). Moreover, converting the frequency-domain channel response into the time-domain by inverse Fourier transform (e.g. the OFDM modulation) is computationally infeasible for practical underwater communication, such as object detection and localization. Our methodology has the potential to address this research gap. The rationale behind the proposed CNN-based methodology is that exploiting the sparsity of the Doppler-delay underwater channel leads to a context-aware methodology that would exhibit very good tradeoffs between the pilot overhead, complexity, and estimation error [22].

3.2 Underwater Acoustic OFDM Systems

This Section introduces the principles of the algorithm to incorporate the expert knowledge of wireless communication in the training of neural networks.

3.2.1 Underwater Communication in Doppler-Delay Domain

In pilot-aided underwater communication, an acoustic channel appears as an incomplete data with a large number of missing values. One of the best ways to incorporate expert knowledge is to extract critical channel features from pilot symbols. Consider an underwater acoustic channel for a relatively short period of time in the absence of the environmental shadowing effects [64, 3, 65]. The impulse response of an underwater acoustic channel in the time-domain is given by

$$h(t) = \sum_{p=1}^{P} h_p \delta(t - \tau_p + \rho_p t), \tag{3.2}$$

where the multipath scattering component will have individual path loss h_p, Doppler spread ρ_p, and the time delay τ_p for each path p. The total number of paths is represented by P. Since underwater communication channels are characterized by long delay spreads and significant Doppler effects due to the marine environmental conditions, the multipath scattering components are usually sparse in the Doppler-delay plane [22]. In the OFDM time-frequency domain, the underwater channel represented by Equation (3.2) is

$$H(t, f) = \sum_{p=1}^{P} h_p e^{j2\pi\rho_p t} e^{-j2\pi\tau_p f}. \tag{3.3}$$

Let us consider n_d symbols and n_s subcarriers. Now, we discretize the time-axis t as $n\Delta t$ for $n = 0, \ldots, n_d - 1$, and the frequency-axis f as $m\Delta f$ for $m = 0, \ldots n_s - 1$. Considering the discretized form of Equation (3.3), the Doppler-delay representation of a received signal at the m-th subcarrier and the n-th data symbol satisfies the following Equation [22]:

$$\mathcal{Y}_{m,n} = \mathcal{H}_{m,n} \mathcal{X}_{m,n} + \varepsilon_{m,n} \tag{3.4}$$

where $\mathcal{Y} \in \mathbb{R}^{n_s \times n_d}$ is the received signal, $\mathcal{H} \in \mathbb{R}^{n_s \times n_d}$ is the channel, $\mathcal{X} \in \mathbb{R}^{n_s \times n_d}$ is the transmitted signal, and $\varepsilon \in \mathbb{R}^{n_s \times n_d}$ is the noise.

According to Soltani et al. [27] and Gaudio et al. [22], the channel is sparse in the Doppler-delay domain. Thus, a suitable approach to incorporate expert knowledge of wireless communication in the training of neural networks is to transmit pilot symbols and estimate a few entries of the matrix $\mathcal{H}$. These measurements would be repeated to provide a set of sufficient training data, leading to an incomplete estimate of the channel matrix. Following Gaudio et al. [22], we have developed a method that inserted pilots at subcarriers of a fixed spacing such as $3\Delta f$ for as few as only two symbols. This leads to encoding a sparse and incomplete representation of $\mathcal{H}$ in the latent space. Following this unsupervised learning of contextual information in the latent space, supervised learning facilitates the neural network to complete the matrix.

In contrast, conventional methods using such an arrangement, pilot symbols are intercalated along two pilot columns of $\mathcal{H}$. The channel matrix $\mathcal{H}$ is then completed by interpolating row-wise, which estimates each symbols at every subcarrier. Advanced methods such as compressed sensing are more efficient to estimate the complete channel matrix. However, a major challenge of such conventional methods is to accurately capture three channel parameters: the amplitude h_p, time delay τ_p, and Doppler rate ρ_p [3, 66]. This is because the conventional matrix completion algorithms do not have a mechanism to recover channel variation due to environmental effects. The optimization of this tradeoff is not trivial and depends on the sparsity of the channel in the Doppler-delay domain. For example, Gong et al. [3] suggested that accurate representation of environmental effects in the estimated channel parameters is crucial for estimating the position and velocity of any moving target in a surveillance area.

Since the underwater channel is affected by various environmental factors, encoding channel state information into a latent space will exploit the sparsity of $H(t, f)$, which in turn would incorporate expert knowledge in the training of convolutional neural networks. In this context, the proposed work has addressed the problem of accounting for channel sparsity in the Doppler-delay domain, targeting the optimization of underwater surveillance tasks by considering artificial intelligence techniques.

In an underwater channel, the environmental dynamics are important contributors to dynamic variation of sparsity in the Doppler-delay domain [17, 67]. Thus, a truly fair comparison of the proposed CNN-based channel estimation technique with its direct concurrent and conventionally used decision feedback channel estimation (DFCE) provides a foundation to incorporate artificial intelligence in underwater search and rescue operations.

3.2.2 Incorporating Expert Knowledge into Neural Networks

We know that deep learning algorithms may be trained to reproduce the training data with high accuracy. However, the resulting network will most likely not generalize well to previously unseen underwater channels. To improve the generalization ability, the training of neural networks typically considers a validation stage using additional data not included into the training set. Thus, deep learning can provide an accurate model with respect to a given geometry and signal frequency associated with the training data. Nevertheless, no mathematical proof exists to guarantee that the network will perform with unseen data as accurately as it performed with the training data.

In contrast, leveraging disciplinary expert knowledge in the training process can significantly improve the generalization ability of deep learning algorithms. The incorporation of disciplinary knowledge in deep learning is referred to as physics-informed or theory-trained neural networks [45]. This framework has gained significant attention from different domains of research, indicating the generalization performance on independent unseen data. In wireless communication problems, a mathematical framework of the physical layer of communications (i.e. the system model) can be injected as expert knowledge to the neural networks.

In the present work, we propose to consider the decision feedback channel equalization (DFCE) algorithm to assess how accurately the network has learned the expert knowledge of the channel. Our approach minimizes the combined error of learning the data and learning the model.

Let us represent the $n_s \times n_d$ channel matrix $\mathcal{H}$ by stacking all elements into a one-dimensional vector in the Doppler-delay grid. We denote $\boldsymbol{y}_d \in \mathbb{R}^{n_s n_d \times 1}$, $\boldsymbol{x}_d \in \mathbb{R}^{n_s n_d \times 1}$ and $\boldsymbol{h}_d \in \mathbb{R}^{n_s n_d \times 1}$ to represent the system on the entire resource grid. Considering the pilot symbols $\boldsymbol{x}_p$, we get the incomplete channel state information at pilots (denoted by p) as

$$\boldsymbol{h}_p^{\text{LS}} = \boldsymbol{y}_p / \boldsymbol{x}_p. \tag{3.5}$$

To complete the matrix, we take the feedback of the pilots to fill in the missing elements of the entire channel state information $\boldsymbol{h}_d$. This is done by multiplying $\boldsymbol{h}_p^{\text{LS}}$ with a suitable matrix $\mathcal{A}^{\text{MMSE}}$ such that

$$\hat{\boldsymbol{h}}_d = \mathcal{A}^{\text{MMSE}} \boldsymbol{h}_p^{\text{LS}}.$$

Let us define the matrix in the following form:

$$\mathcal{A}^{\text{MMSE}} = R_{\boldsymbol{h}_d \boldsymbol{h}_p} \left(R_{\boldsymbol{h}_p \boldsymbol{h}_p} + \sigma_n^2 \left(\boldsymbol{x}_d \boldsymbol{x}_d^H \right)^{-1} \right)^{-1} \tag{3.6}$$

where $R_{\boldsymbol{h}_d \boldsymbol{h}_p} = \mathbb{E}\{\boldsymbol{h}_d \boldsymbol{h}_p^H\}$ is the correlation matrix between the CSI of the entire time-frequency grid ($\boldsymbol{h}_d$) and the CSI extracted at pilot positions ($\boldsymbol{h}_p^{\text{LS}}$). The filter matrix ($\mathcal{A}^{\text{MMSE}}$) given by Equation (3.6) incorporates contextual channel state information by solving the following optimization problem:

$$\text{minimize} \quad ||\boldsymbol{h}_d - \mathcal{A}^{\text{MMSE}} \boldsymbol{h}_p^{\text{LS}}||_2 \quad \text{s.t.} \quad \boldsymbol{y}_p = \boldsymbol{h}_p^{\text{LS}} \boldsymbol{x}_p. \tag{3.7}$$

It should be noted that the filter matrix $\mathcal{A}^{\text{MMSE}}$ requires the full knowledge of the correlation matrices $R_{\boldsymbol{h}_d \boldsymbol{h}_p}$ and $R_{\boldsymbol{h}_d \boldsymbol{h}_p}$ [53].

3.3 Proposed CNN-Based Channel Estimation

3.3.1 Generalizable Learning Framework

The convolutional neural network (CNN) is illustrated in Figure 3.1. CNN is a deep learning architecture which can detect objects by extracting critical features to recognize patterns. CNN slides a kernel through the target data and use the data that falls inside the kernel (i.e. the receptive field) to extract the desired patterns within the target data. The basic idea of the proposed approach is to consider a prior estimate of missing data and combine it with a CNN-based data imputation technique.

This Section briefly presents the proposed CNN-based technique, which predicts the time-frequency channel matrix using the neural network as $\hat{\mathcal{H}} = f_\sigma(\hat{W}, \boldsymbol{h}_p^{LS})$. The proposed algorithm consists of three steps. First, we estimate the elements of $\hat{\mathcal{H}}$ corresponding to pilot

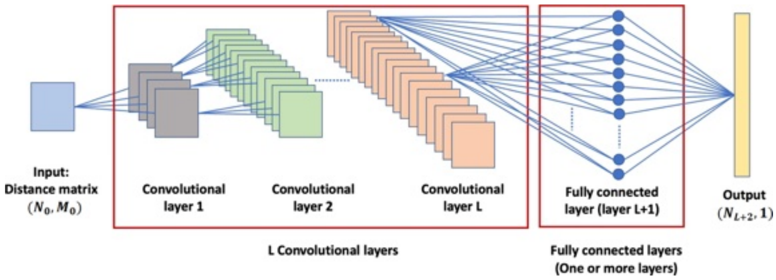


Figure 3.1: An illustration of the CNN model

symbols, $\boldsymbol{h}_p^{LS}$, and consider the remaining elements as missing data. Second, we determine some weights $\hat{W}$ to employ a data imputation operator denoted by $f_\sigma(\cdot)$. It extracts critical features of the channel by using time-frequency analysis and data imputation technique. Third, it trains a CNN architecture to learn the correlation between the missing probabilities and the critical features of the channel using training data.

The imputation is a projection operator [68],

$$\mathbb{P}(\hat{\mathcal{H}}) = \begin{cases} h_{ij} & (i,j) \in \{\text{pilots}\} \\ \tilde{h}_{ij} & (i,j) \notin \{\text{pilots}\} \end{cases},$$

which provides a set of all matrices that agree upon the non-missing entries at pilot locations. We can prove that the set of all possible completions of such an incomplete matrix forms a convex set. The above imputation operator thus projects missing values assuming $\tilde{h}_{ij}$ belongs to a convex set.

The estimated values of the channel at pilot locations will be noisy. The CNN-based deep learning algorithm will compare the imputed channel matrix $\tilde{\mathcal{H}}$ with the true channel $\mathcal{H}$ or a theory that implicitly represents the true channel. Since matrices with partial data belong to convex sets, the imputation approach guarantees that the reception field of CNN will capture local features [47, 49, 55]. Thus, one expects that i) the learning of the channel by CNNs is guaranteed; ii) incorporation of expert knowledge through theory-trained neural networks provides a CNN model that is generalizable to any unseen underwater environment, where the "theory" used for training is applicable.

In order to impute missing values of high-dimensional OFDM channels, a matrix-to-matrix mapping model with the ability to retain local time-dependent features is a good choice. In this way, the CNN kernel learns the true features within the receptive field. Without the second step, a classical missing data algorithm would introduce a substantial amount of bias, thereby making predictions more arduous. The proposed method can solve this problem because CNN is powerful to capture the discrete spectrum with much lower sampling rate.

3.3.2 Approximation and Learning

Artificial neural networks (ANN) are computational architectures, which optimizes the weights and biases to relate input $\tilde{\mathcal{H}}$ (i.e. incomplete channel matrix) with target $\mathcal{H}$ (i.e. true channel) [62]. In deep neural network (DNN), the input goes through a finite number of linear transformations called hidden layers. At each layer, an activation function accounts for the underlying non-linearity of the target. The neural network approximation at the l-th layer is

$$\mathcal{H}^l = \sigma\left(\boldsymbol{W}^l \tilde{\mathcal{H}}^{l-1} + \boldsymbol{b}^l\right), \quad l = 1 \ldots L. \tag{3.8}$$

Here, $\boldsymbol{W}^l$ denotes the weights of the linear transformation between the $(l-1)$-th layer and the l-th layer; $\mathcal{H}^l$ denotes the output vector of the l-th layer; and $\boldsymbol{b}^l$ contains the biases for the l-th layer [69]. After a finite number of transformation, $l = L$, we have $\hat{\mathcal{H}} \equiv \mathcal{H}^L$, which approximates the true channel matrix $\mathcal{H}$. In order to approximate $\mathcal{H}$ by the neural network output $\hat{\mathcal{H}}$, the back-propagation algorithm estimates the optimal values of $\boldsymbol{W}^l$ and $\boldsymbol{b}^l$.

More specifically, deep learning is a convex optimization problem, which minimizes an appropriate Lagrangian or cost function [59] [26]. Typically, the following cost function is considered:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} ||\hat{\mathcal{H}}_i - \mathcal{H}_i||^2, \tag{3.9}$$

where N is the number of observations. In practice, DNNs can discover any channel matrix $\mathcal{H}$ if there are a large number of observed channel matrices a large number of weights. However, these requirements are not suitable in wireless communication because we need to transfer a large amount of data in real time.

CNN uses the following three steps to significantly minimize the number of weights with respect to DNNs. These steps are: local receptive fields, shared weights and biases, and activation process. A local receptive field involves a certain number of neurons in one layer being connected with only a few neurons in the following layer, which results in the total number of parameters fewer than that of the ANN. We can mathematically represent a local receptive field as a $k \times k$ kernel of the matrix $\mathcal{H}$ connected to a single neuron in the hidden layer. This kernel is convolved with the original input matrix such that [70]

$$\mathcal{H}_{mn}^{l+1} = \sum_{i=0}^{k} \sum_{j=0}^{k} w_{mn}^{l} \mathcal{H}_{m+i,n+j}^{l} \tag{3.10}$$

where $\mathcal{H}_{mn}^{l}$ and $\beth_{mn}^{l}$ respectively denotes the m^{th} row and n^{th} column of channel matrix H and weight matrix w obtained at layer l of the neural network. The weight matrix $\beth_{mn}^{l}$ at layer l is convolved with $\mathcal{H}_{mn}^{l}$ via the $k \times k$ kernel. Thus, Equation (3.10) extracts the features of the input $\mathcal{H}^{0}$ through L hidden layers, and maps the L-th layer $\mathcal{H}^{L}$ to the actual channel response $\mathcal{H}$.

3.3.3 Training Underwater Sensor Nodes

A major challenge in CNN-based data recovery technique for underwater sensor nodes is to collect offline training data. One possibility is to consider a statistical channel model to generate synthetic data for training the underwater sensor nodes. We use 80% of this data for the training process. The remaining 20% is used for the testing process. The key benefit of CNN-based missing data recovery for channel estimation is to use the known pilots to find the remaining unknown elements of the channel state information. Pilot elements of $\hat{\mathcal{H}}$ can be obtained through the zero-forcing approach given by Equation (3.5). According to the universal approximation theorem, the offline training of sensor nodes not only estimates the missing elements of $\hat{\mathcal{H}}$, but also improves the pilot values estimated by the zero-forcing method [71]. The accuracy of the method is guaranteed by the hypothesis that any feed-forward neural network can approximate a continuous function if the number of neurons and the number of hidden layers are sufficiently large.

To test such an approach for supervised training of sensor nodes, this study has considered a channel model implemented in MATLAB to generate an underwater channel given by Equation (3.2). Then, we applied the OFDM demodulation of Equation (3.2) to construct the true channel impulse response $\mathcal{H} \in \mathbb{R}^{n_s \times n_d}$ in the time-frequency domain. Considering reflection and refraction in the p-th path of an underwater channel, the path gain (h_p) and the delay-Doppler pair (ρ_p, τ_p) were evaluated to generate a set of N OFDM channel matrices $\{\mathcal{H}_i\}_{i=1}^{N}$. Note that the channel model accounts for the random variations of an underwater acoustic channel. These channel matrices were passed as training labels for the CNN architecture [12] [17].

An alternative approach for offline training of sensor nodes is to consider an existing channel estimation scheme. For instance, the training data can be generated by the pilot-based compressed sensing scheme presented by Gaudio et al. [22]. In other words, the sensor nodes can use a conventional method to estimate a finite number of multipath variations of reflection and refraction in an underwater channel, thereby providing training data for network parameters of a universal function approximator. A potential advantage is that a trained network will provide good predictions of new channel variations without needing further channel estimation. Currently, there exists massive computational resources and sufficient academic experiences on deep learning [72] [4]. Thus, by incorporating the existing knowledge of wireless communication in the training of neural networks, we can further extend the proposed CNN-based channel estimation scheme.

3.4 Results and Analysis

This study presents a deep learning model for underwater acoustic communication channels, which accounts for environmental aspects of acoustic propagation and the effects of random channel variation. The proposed methodology is based on missing data recovery techniques by convolutional neural network. To evaluate the accuracy of the proposed CNN-based channel estimation technique, underwater acoustic channels were simulated to provide necessary data for training and validation. The present study follows previous works such as [64, 65, 66, 73]. The channel data generated from this simulation was considered to train the CNN framework.

3.4.1 Simulation Setup

In this study, we considered a statistical approach to account for random channel variations in a multipath underwater channel between a transmitter and a receiver. The channel consists of a line-of-sight path, and a fixed number of reflections from sea surface and the seafloor. Using the ray theory, we simulated a relatively accurate picture of the multipath environmental conditions, incorporating Doppler effects caused by the receiver's relative motion. The delay spread was randomly chosen between 0 ns and 300 ns, while the maximum Doppler shift was set to 50 kHz. The study utilized MATLAB's *5G toolbox* to simulate a transponder of an ROV, which periodically sends n_p pings toward an array of n_t transducers. Thus, for each of $n_p \times n_t$ pings, we simulated an individual realization of the channel as a matrix in the Doppler-delay domain. These channel matrices were considered to train a CNN-based underwater acoustic communication channel.

3.4.2 Recovery of Contextual Information in Underwater Channels

We investigated the performance of the DFCE algorithm to leverage contextual information in estimating underwater acoustic channels. We incorporated contextual information using the OFDM technique, while representing a class of underwater channels as $n_s \times n_d$ matrices. Due to the small-scale scattering and large-scale location uncertainty of underwater environmental conditions, the channel path gain is expected to have a normal distribution if the Doppler shift uncertainty is characterized as Gaussian. Thus, a context-aware estimation of the channel matrix $\mathcal{H}$ in the Doppler-delay domain would accurately capture the overall statistical characteristics of an underwater channel. To investigate whether the method captures contextual information, the pilots were inserted into as few as only two columns of the OFDM channel $\mathcal{H}$, where each of the 14 time slots consists of 612 subcarriers. For each of the two pilot time slots, every third subcarrier was pilot symbols. All non-pilot data elements were treated as missing values. Thus, the contextual feedback from 408 pilots, arranged in two columns, was taken to recover missing channel data, which resulted in an estimated channel $\hat{\mathcal{H}}$ of size 612×14. More specifically, we considered a pilot arrangement in which about 95% elements of the channel matrix were treated as missing values. We assessed the accuracy of the DFCE algorithm by varying the values of SNR of the received signal (e.g. Equation 3.4) in the range of 1dB to 25dB.

Figure 3.2(a) shows the true channel in the time-frequency domain. Figures 3.2(b,c,d) show the channel $\hat{\mathcal{H}}$ estimated by the DFCE algorithm for SNR values of 1dB, 10dB, 15dB. The error for each of the estimated channel $\hat{\mathcal{H}}$ was obtained with respect to the true channel $\mathcal{H}$ shown in figure 3.2(a). Figures 3.2(e,f,g) show the corresponding error. The results suggest that the DFCE algorithm has qualitatively captured the underlying log-normal distribution of the path gains because the estimation error has decreased as the SNR of the received signal increased. Specifically, the results show that only 5% elements of the channel matrix $\mathcal{H}$ in two time slots are sufficient to capture the channel for a class of underwater acoustic signals.

We investigated the performance of the CNN-based methodology while leveraging the contextual information via the DFCE algorithm. During the training stage, we considered the received pilots for an environment with SNR = 15dB and employed a linear interpolation method to input the missing elements of the channel. For each training sample, the path gains and Doppler shifts were varied randomly to account for turbulent underwater conditions. The MSE loss was defined as $||\text{NN}(w, \tilde{\mathcal{H}}) - \mathcal{H}|| + ||\text{NN}(w, \tilde{\mathcal{H}}) - \mathcal{H}^{\text{DFCE}}||$. Note that we have injected the contextual information via DFCE while training the CNN architecture. Thus, our method finds the optimal correlation between the neural network estimation, the training data $\mathcal{H}$, and a model-prediction $\mathcal{H}^{\text{DFCE}}$. In this way, the CNN kernel can accurately learn the channel and the corresponding distribution of the channel parameters.

We compared the accuracy of the CNN-based channel estimation with that of the DFCE

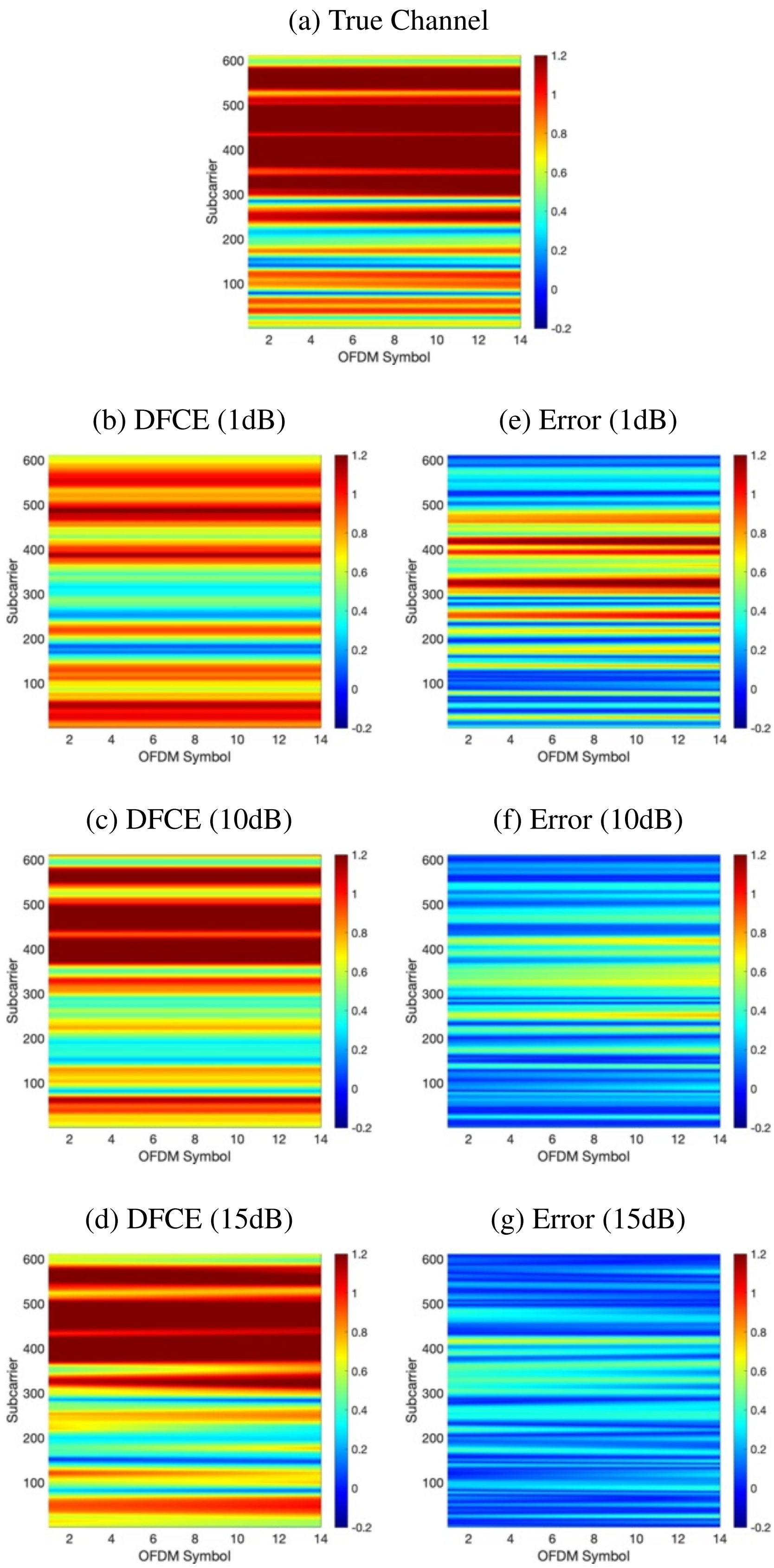
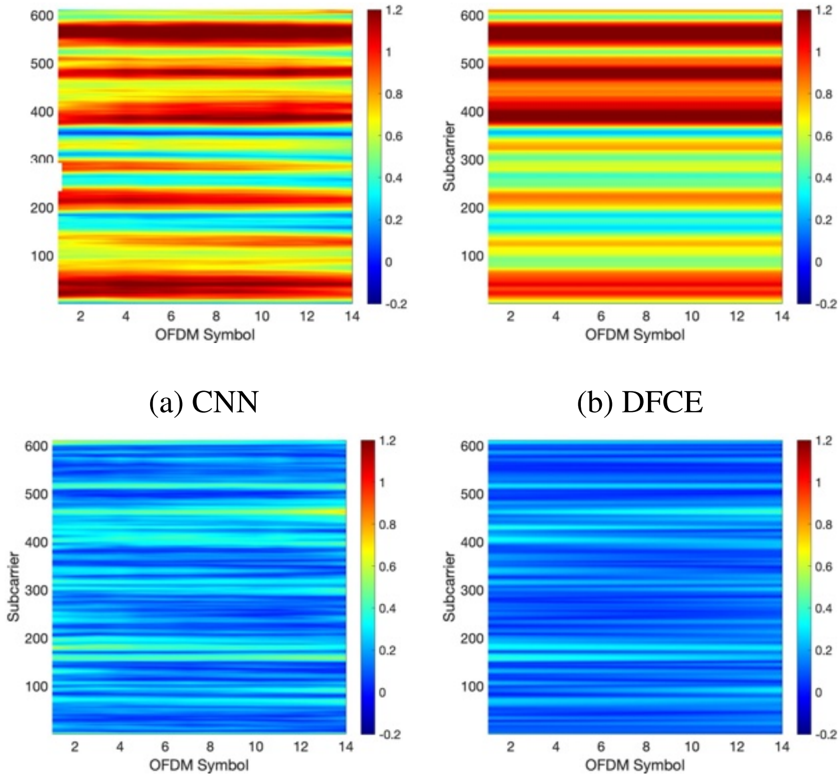


Figure 3.2: (a) An exact channel matrix in the time-frequency domain, which was generated by a statistical channel model. (b,c,d) The estimations of the true channel shown in (a) by using DFCE. (e,f,g) The errors in the DFCE estimation method with respect to the true channel.

algorithm for SNR = 5dB. Figures 3.3(a,b) show the channels estimated by the CNN and the DFCE algorithms. Figures 3.3(c,d) show the error of the CNN-based estimation (∼7%) and the error of the DFCE (∼8%). The results indicate that the accuracy of the CNN-based channel estimation is approximately the same as that of the DFCE algorithm, even for significantly different underwater environment with much lower SNR (5dB) than the training environment of high SNR (15dB).



(a) CNN (b) DFCE

(c) Error (∼ 7.04%) (d) Error (∼ 8.84%)

Figure 3.3: OFDM modulated channel state information. (a) Channel state information estimated by neural network; (b) Channel state information estimated by decision feedback channel estimation; (c) relative error of neural network; (d) relative error of decision feedback channel estimation

3.4.3 Estimation of Statistical Characteristics of Underwater Channels by Deep Learning Techniques

Statistical modelling of underwater channel is a subject of ongoing research, and no consensus exist on the fading type [74]. Several past studies demonstrated the histograms of path gains, which provide valuable insight about statistical characterization of underwater channels. To this end, we have investigated the statistical properties of underwater channels using the proposed CNN-based channel estimation technique. Our methodology is based on the assumption that the missing data recovery technique will detect the relevant distribution of the random channel variation. We assume that the fast variation of the instantaneous channel response due to scattering and motion-induced Doppler shifting would remain the same with turbulent variation of the environmental conditions. Thus, the classical maximum likelihood method is appropriate to assess the accuracy of the CNN-based channel estimation. In other words, we need to estimate the parameters of an assumed prior distribution. However, in the present work, we have considered the quantile regression methodology to measure the similarity between the distributions of the estimated channel and the true channel that acts as the training data.

Quantile regression is unbiased with sample size and can be an alternative to the binning approach that is usually biased with the sample size [75]. To make this presentation self-contained, let us briefly outline the quantile method. Consider the flattened elements h_i, $i = 1, \ldots, N$ of a channel matrix $\mathcal{H}$, and denote the expected distribution as $F(h)$. Suppose that the samples h_i were sorted into a non-decreasing sequence, where $h_i \leq h_j$ for any $i \leq j$. The quantiles of the samples are defined as:

$$F^{-1}(q_i) = \inf\{h_i : F(h_i) \geq q_i\} \tag{3.11}$$

where $q_i = \frac{i-0.5}{N}, \quad i = 1, \ldots, N$. If the distribution of the true and estimated data match closely, then

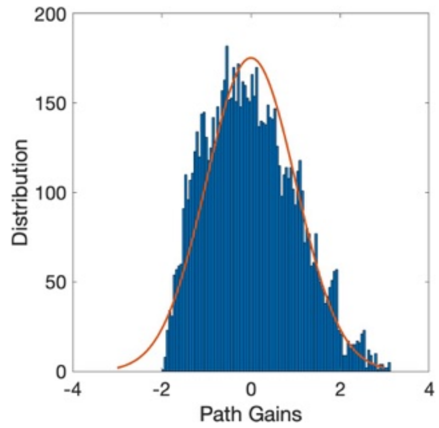
$$h_i \approx F^{-1}(q_i). \tag{3.12}$$

Thus, the relation between h_i and the corresponding quantiles of the expected distribution is approximately linear. The present study utilized MATLAB for the quantile-quantile analysis discussed as follows.

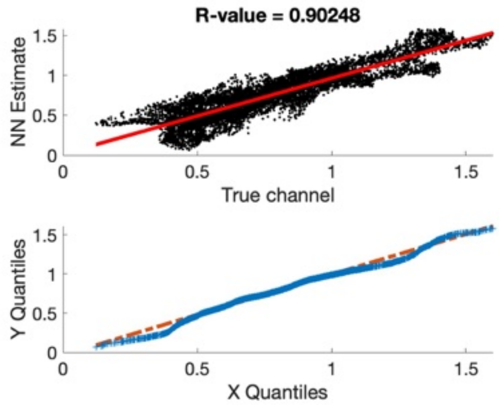
The Q-Q technique has been applied to assess the ability of learning the distribution of random channel variations. The underlying distribution of the channel parameters correspond to important contextual information that may not be directly observed. A goal of learning the underlying distribution is to ensure that the proposed CNN-based methodology can estimate the channel in a new underwater environment, which was not considered in training the network architecture. Since the neural network weights are obtained by minimizing the MSE, it corresponds to the maximum likelihood solution of the underlying distribution. More specifically, a trained neural network architecture provides a statistical channel model, which can account for the random channel variation due to various environmental effects.

Numerical experiments were considered using a range of SNR values and random variation of channel parameters. Figure 3.4(a) demonstrates the probability distribution of the channel matrices used as training labels. The skewness and the kurtosis of the corresponding distribution were 0.2426 and 2.4863, respectively, which closely matches with the skewness (0.0) and kurtosis (3.0) of the normal distribution. This means that the normalized path gains are approximately symmetrical. We considered two ways to assess the learning ability of the neural network architecture. First, the Pearson's correlation coefficient was computed to measure how accurately the estimated channel $\hat{\mathcal{H}}$ correlates with the true channel $\mathcal{H}$. Second, the quantile-quantile plot was considered to measure the difference between the distribution of the exact and the estimated channel matrices.

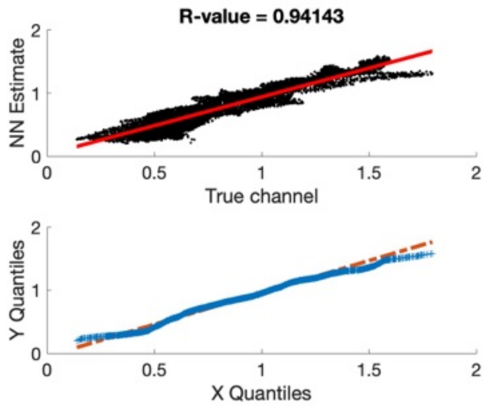
On the top panels of Figs 3.4(b,c,d), we see the scatter plot between $\mathcal{H}$ and $\hat{\mathcal{H}}$ for an SNR of 5dB, 15dB, and 25dB, respectively. The correlation coefficient increases as the SNR increases. Note that the network was trained for 15dB. The bottom panel of these Figures compares the distribution of $\mathcal{H}$ and $\hat{\mathcal{H}}$. The Q-Q plots suggest that a deep learning method is robust in learning the underlying distribution.



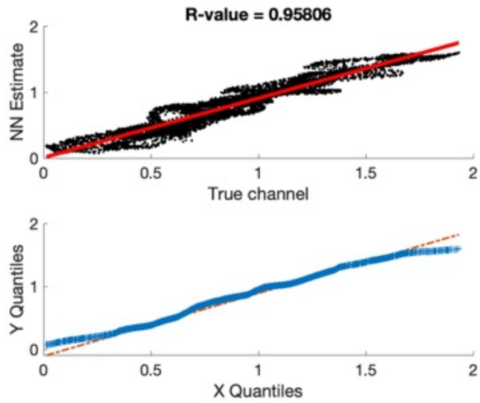
(a) True channel distribution



(b) Q-Q plot (5dB)



(c) Q-Q plot (15dB)



(d) Q-Q plot (25dB)

Figure 3.4: (a) The distribution of the true channel. (b)-(d) Scattered plots showing correlation between true channel and its prediction by CNN (top panel). Quantile-quantile plot comparing the distribution of the true channel and the distribution of the CNN prediction (bottom panel). Three values of the SNR were considered: 5dB (b), 15dB (c), 25dB (d).

3.4.4 Effect of Environmental Noise in the Training Data

In this Section, we focused on training the neural network architecture in a variety of environmental conditions that take into account random channel variations. In particular, we want to understand the effectiveness of the CNN architecture in imputing the missing values of the channel matrix in the Doppler-delay domain. To illustrate the performance of CNN-based channel estimation in various noisy underwater environments, we considered three sets of training data with SNR of 5dB, 15dB, and 25dB. Each of the trained model was tested for six values of SNR within the range of 0dB and 25dB. For each of the tested channels, we calculated the mean squared error (MSE) between the estimated and true channels.

Figure 3.5 has summarized the results. It demonstrates that if the test data has relatively high levels of noise, (i.e. low SNR values in the range of SNR $\leq$ 12dB), then the CNN performs relatively better in an underwater environment with relatively high noise (i.e. SNR $\leq$ 5dB). Additionally, it is observed that for testing data of high SNR values (i.e SNR $\geq$ 12dB), the performance of CNN in a training environment of high SNR (i.e. 15dB) is relatively better than the performance of CNN in a training environment of low SNR (i.e. 5dB). Previous studies [27] have also attempted to address the aforementioned phenomenon by dividing the channel estimation model into a low-SNR network (i.e. SNR $\leq$ 12.5), and a high-SNR network (i.e. SNR $\geq$ 12.5). Hence, the result in Figure 3.5 is consistent with past findings using other form of neural network architectures. More specifically, Figure 3.5 suggests that the matrix recovery approach is an efficient channel estimation technique.

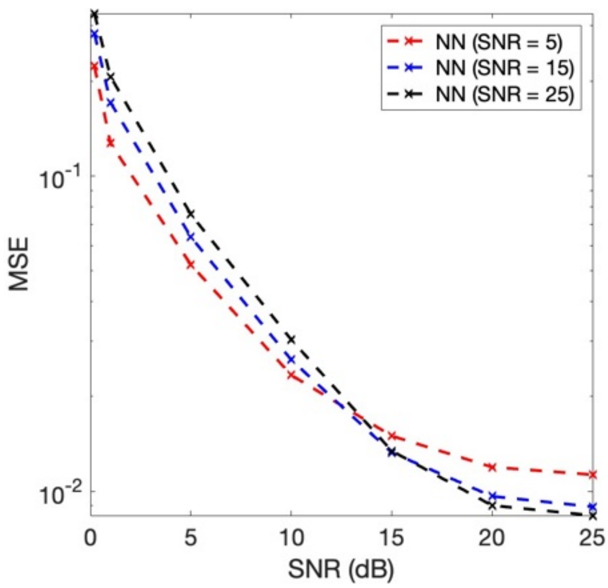
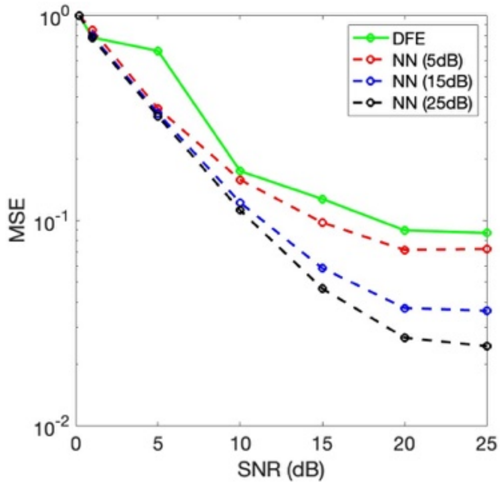


Figure 3.5: A line plot that demonstrates that the optimization of a CNN model depends on the amount of noise of both the training and testing data. The accuracy of the CNN model is optimal when the noise within training data and the testing data are equal.

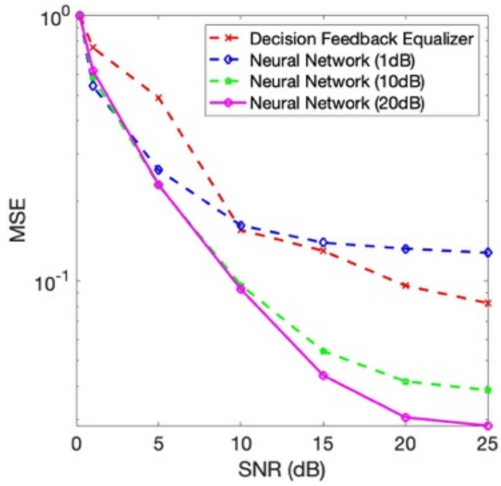
In other words, the offline training of sensor nodes with such a CNN architecture can efficiently estimate the Doppler, delay, and path gains of an OFDM channel matrix [22]. It is easy to see that the CNN-based methodology is efficient to obtain good communication rate (i.e. transmitted information within a block of symbols). The proposed methodology is most likely to address the tradeoff between the number of pilots and data symbols [76]. In underwater acoustic channels, the optimization of this tradeoff is affected by the chaotic channel propagation characteristics [17].

For the conventional channel estimation of an underwater wireless system, the estimation accuracy depends on accurate representation of the statistics of the type of channel being estimated. In contrast, the accuracy of deep learning-based technique may depend on optimizing the hyperparameters of the neural networks. However, the universal approximation theorem suggest that the effect of hyperparameters may not be a significant factor if the training epochs are sufficiently large. Considering this assumption, we have tested the effect of modulation scheme. This is shown in Figure 3.6, which compares the accuracy of CNN-based method with that of the DFCE method. This test investigated the effect of training environments (characterized by three SNR values) and OFDM modulation schemes on the performance of the deep neural network. The OFDM modulation is employed to represent contextual information regarding multipath interference and Doppler effects. Considering the training data for three environmental conditions characterized by three SNR values, we tested the performance with SNR (dB) values varying between 1dB - 20dB.

With this experiment, the configured independent variable are the channel properties, including the delay spread, maximum Doppler shift, sampling rate, etc. Overall, the CNN-based approach seems to outperform the conventional methods. Although we have not tested conventional LS and MMSE, we generally expect that the DFCE performs about the same as MMSE. Moreover, MMSE is usually better than the LS method.



(a)



(b)

Figure 3.6: Eight simulations were conducted to study the effects of SNR (dB) on training data and that of OFDM modulation scheme on the performance of the CNN-based channel estimation. The MSE was obtained with respect to the true channel corresponding to each of the three training SNR values while the SNR values of test data varied between 0dB and 25dB. (a) Comparison for QPSK. (b) Comparison for 64QAM.

3.5 Conclusion and Future Direction

In this research, we have proposed a novel methodology for context-specific characterization of underwater acoustic channels using the convolutional neural network architecture. We borrowed the concept from data analytic theory, where context is any information that characterizes the situation of different entities. In an underwater environment, many factors such as temperature, salinity, and turbulent fluctuation of water pressure, etc may influence the performance of underwater acoustic communication. We have found that one way of dealing with this problem is to incorporate relevant context in the training of neural network. Similar approaches adopted in material science are called physics-informed or theory-trained neural networks. Instead of considering a black box neural network model, we have considered the similarity between the feedforward neural network and the wireless communication system model. We observed that the neural network weight matrix is similar to the channel matrix in wireless communication. An important aspect of considering this similarity is to incorporate the domain knowledge of wireless communication and context-specific information into the deep learning-based architecture.

The underlying premise of the proposed development is that pilot-based wireless communication using the OFDM framework is equivalent to a data recovery problem. Algorithms for classical data-recovery methods are not usually cost-effective for underwater acoustic communication. However, deep learning method shifts almost all of the computational burden to offline training process, which is needed only once. Although it is a challenging endeavour to experimentally acquire accurate training data for underwater environment, we have shown that considering the domain knowledge such as channel modelling and conventional channel estimation, it is possible to generate sufficient training data. More specifically, the comparison between CNN-based channel estimation and the DFCE algorithm suggests that DFCE can be considered to produce synthetic training data.

We can extend the work of this research in underwater source localization and object detection. CNN is basically an auto-encoder. This means that the OFDM channel matrix from the time-frequency domain can be encoded to a low-dimensional feature space. When the features from a CNN architecture is decoded back to the time-frequency domain, any underlying anomalies or outliers are efficiently identified. This approach is known to be a powerful technique in data science. By incorporating signal processing knowledge with a variational auto-encoder, it is most likely to be an efficient technique to design underwater sensor nodes for source localization.

Chapter 4

Assessment of LSTM-based Symbol Detection for Underwater Acoustic Receiver

OFDM is an emerging technology for symbol detection in underwater wireless communication. This article has combined the long short term memory (LSTM) network with the feature classification technique to develop a deep learning- based OFDM symbol detection methodology for acoustic signal processing. We have investigated the applications of neural networks for channel estimation in an underwater environment. Conventional channel estimation techniques are limited by the time variability, narrow bandwidth, multipath, frequency selective fading, and the Doppler effect. To incorproate artificial intelligence for underwater acoustic communication, we turn to reccurence neural network models. More specifically, this chapter focuses on Long Short Term Memory-based neural network architectures to account for the large volume of data collected by numerous autonomous vehicles and sensors that have been deployed in various parts of the ocean. The performance of the neural network model implemented in the OFDM receiver has been assessed with respect to the theoretical error bound and compared with traditional methods.

4.1 Introduction

Orthogonal frequency division multiplexing (OFDM) is a promising modulation technique that could enhance underwater acoustic communication systems for autonomous unmanned vehicles (AUVs), particularly in target tracking and data acquisition applications [77, 78, 79, 3, 17, 80]. The OFDM technique can mitigate multipath effects, thereby improving signal clarity and transmission of large volumes of data [17]. However, signal recovery remains challenging due to the limited ability of existing channel models to accurately characterize the unpredictable nature of underwater environments. Hence, developing efficient communication systems for AUVs without heavily relying on channel estimation is a critical research area [79, 81]. More specifically, the integration of artificial intelligence techniques offers promising underwater acoustic communication between AUVs, which can bypass or minimize the traditional reliance on estimating multipath interference, Doppler shift, frequency-selective fading, and other time-varying factors [12, 17]. Overcoming these classical underwater challenges underscores the need for direct symbol detection techniques that bypass channel estimation, and thereby paving the way for more robust underwater tracking and data acquisition.

Recently, many researchers have leveraged the experimental data to advance deep learning techniques for underwater channel estimation [26, 59, 82, 83]. They observed that convolutional (CNNs) [84] and recurrent neural networks (RNNs) [85] have the potential to accurately predict Doppler shifts in underwater signals. Such data-driven methods can also mitigate the background noise and multipath interference. Long short-term memory networks (LSTM) is another data-driven approach to offer an accurate channel equalization technique. Thus, deep learning is crucial for localization and communication processes [69, 62]. Deep neural networks (DNNs) are usually trained offline, which offset the channel estimation complexity during an underwater operation. Since the deep learning approach is a data-driven model, it does not incorporate the physical characteristics of underwater channels, and thus it is not a context-aware methodology. Context-awareness is crucial for AUVs to take intelligent decisions as they are employed to track objects in large marine areas with low-illumination and high-noise situations [86]. Incorporating such contextual information in DNN-based techniques is an open research challenge.

This research has considered the OFDM technique as a domain expertise that acts as a regularization agent for incorporating contextual information into the neural network architecture. We have leveraged the deep learning algorithm to accurately identify a nonlinear map that detects a desired symbol constellation from a few OFDM data. The aim of this research is to exploit OFDM techniques for advancing the recently developed theory-trained neural networks in underwater acoustic communication. This simple, yet powerful construction, allows us to take advantage of the OFDM technique while employing the LSTM network in detecting the constellation points of received OFDM symbols. The underlying premise of the proposed LSTM network is to estimate the highest probability that a received symbol belongs to one of the OFDM constellation point. Unlike the recent progresses of deep learning in channel estimation, the proposed LSTM network detects the OFDM symbols in an end-to-end manner that neither adopts the regression of transmitted symbols nor the explicit estimation of the CSI. It is worth mentioning that the LSTM approach accounts for the time-variability of the underwater environment in such a way that a relatively low number of neurons can achieve high spectral efficiency and transmission rates in underwater acoustic communication. Consequently, the proposed LSTM-based symbol detection becomes aware of channel variability (e.g. multipath and Doppler effects), which is an important context-based characteristic for underwater acoustic communication.

To validate the proposed LSTM framework, we considered training data that incorporates the multipath and Doppler effects. The trained network then finds the optimal functional relationship between the received symbols and the corresponding constellation points such that the symbols at the OFDM receivers are properly detected. According to the universal approximation theorem, our context-aware LSTM network is guaranteed to learn the channel parameters for a wide variety of conditions. In other words, the LSTM-based symbol detection technique becomes spectrally efficient, which means that the amount of data transfer is maximized while maintaining a high level of accuracy.

We considered two approaches for assessing the accuracy of the proposed LSTM neural network for joint channel estimation and symbol detection. First, we assessed the symbol error rate of the trained LSTM neural network by considering the Bayesian statistics. In other words, if the transmitted symbols had a prior probability distribution, then the performance of the MMSE detector is not necessarily superior to that of the LSTM detector. Second, we considered perfect channel state information of an underwater channel with multipath and Doppler effects. The SER performance with respect to the perfect CSI and the MMSE estimator was considered to assess the symbol error rate predicted by the LSTM network. The comparison of such classical estimators with an LSTM-based symbol detector indicating the performance difference suggests that the deep learning model is superior to the conventional approaches.

4.1.1 Prior Studies

DNN and the universal approximation theorem of neural networks were introduced as a powerful signal processing methodology in the early 1990's [87, 88, 89, 82, 20, 90]. Table 4.1 provides some representative past contributions of neural networks in signal processing. However, deep learning techniques have not been fully investigated for underwater signal detection. A majority of recent works investigated deep learning techniques for channel estimation in the field of terrestrial wireless communication [62, 33, 91, 92, 93, 36, 94].

Existing deep learning techniques for signal processing can be categorized into data-driven and context-aware approaches. Data-driven approaches, as the name suggests, make strategic decisions by training a black-box network architecture using large amounts of training data without considering any domain knowledge of the underlying wireless communication problems [72, 95]. In the data-driven approach, the neural networks have to be trained to capture the long-term dependencies in the data, as well as the underlying sparsity of channel state information, which are typical in underwater acoustic channels due to environmental effects [96, 97, 98, 99, 12, 79, 78].

The context-aware deep learning methods incorporate the domain knowledge so that the training process simultaneously learns the features of the training data and a mathematical model that can predict these features. Several context-aware deep learning techniques employed the Bayesian estimation framework to optimize the neural network architecture by incorporating the existing expert knowledge in wireless communication [59, 100, 101]. Such a context-aware deep learning approach is relatively new in wireless communication, and its benefits were not fully exploited for underwater problems. Some past work demonstrated the robustness and superiority of context-aware deep learning over that of the purely data-driven deep learning technique for wireless communication [101, 102].

Ref. [33] reviewed the effectiveness of artificial intelligence in underwater acoustic communication. Ref. [26] observed that channel estimation performance of the data-driven deep learning method is comparable to that of the classical MMSE estimator. Ref. [27] proposed to replace the entire channel estimation step through the data-driven training of a convolutional neural network (CNN) architecture. Ref. [103, 104, 105, 62] indicated the potential for data-driven methods in signal processing, channel estimation, and underwater acoustic communication.

The aforementioned investigations were motivated by the complexity and limitations of conventional techniques in underwater communication. The performance of these techniques is influenced by statistical channel models, which is challenging due to fast dynamic changes in the aquatic environment and the complexity of the associated time-space-frequency variations [20]. Recent progresses indicate that statistical channel models may not be a detrimental factor for channel estimation by data-driven deep learning techniques if sufficient amount of training data is available [58].

Table 4.1: A Summary of Past Research Work on Deep Learning for Some Representative Topics in Underwater Acoustic Communication

Architecture	Contribution	Remark
DNN	Channel Estimation [62, 106, 107, 108], OFDM communication [62, 106, 108], Object detection [109], Doppler Estimation [110]	Combined deep learning with domain knowledge
CNN	Channel Estimation [27, 36, 111, 37], OFDM communication [35, 37, 112], Object Detection [113], Doppler Estimation [114]	Extended super-resolution technique
LSTM	Channel Estimation [94, 115, 116], OFDM communication [117, 116], Object detection [118], Symbol detection [85], Doppler Estimation [117]	The RNN model is trained by extracting sequence feature information of the OFDM received signal from the big data

4.1.2 Contributions

Many recent deep learning-based investigations focused on channel estimation accuracy [26, 82, 83, 69, 104]. Since underwater channels are affected by uncertain sparsity and other complicated factors from marine environments, the proposed deep learning method bypasses the channel estimation, thereby offering a generalized symbol detection scheme. This study has thus complimented prior investigations by advancing the context-aware deep learning technique, where the multipath and Doppler effects in underwater channels have been considered as critical contextual information. The results suggest that the proposed integration of OFDM with LSTM networks offers key benefits such as improved Doppler shift robustness, enhanced generalization across channel variations, reduced training complexity, and increased accuracy in symbol detection. Unlike conventional deep learning techniques that treat the underwater channel entirely as a black-box DNN model (e.g. Table 4.1), the proposed context-aware deep learning approach integrating the OFDM technique with neural networks facilitates real-time implementation on AUVs.

Three case studies were considered to verify the performance of the proposed LSTM-based context-aware technique. The first study assessed the symbol detection accuracy considering a training data of an idealized channel in which the number of path was fixed; however, the path gain varied randomly. The second study provided a hyperparameter analysis for the context-awareness of the neural network architecture using training data of an underwater channel, where signals experienced multipath reflections with varying delays and Doppler shifts. The third study investigated the performance of AUVs employing the proposed OFDM receivers for underwater activities. Thus, this study has assessed the proposed LSTM network, which has evaluated the learning performance of the proposed OFDM receiver and its sensitivity to the effect of Doppler shifts.

4.2 Neural Networks for Symbol Detection

In this Section, we introduce the underwater acoustic communication system, and the context-aware deep learning for symbol detection in hostile underwater channels experiencing uncertain sparsity.

4.2.1 Underwater Acoustics System

Considering the time-variability of channel parameters, the delay-Doppler spread function of an L-tap underwater acoustic channel is

$$h(t,\tau) = \sum_{i=1}^{L} h_i(t)\delta(t - \tau_i(t)). \tag{4.1}$$

In eq (4.1), L is the total number of taps and τ_i characterizes the propagation of the i-th path. According to ref. [64], one representation of the delay-Doppler spread function (4.1), where different delays move at different speeds, is given by

$$h_i(t) = a_i \exp[j(\phi_i + 2\pi t f_{\max} \cos\beta_i)]. \tag{4.2}$$

Here, a_i denotes the loss over the i-th path. Note that the maximum Doppler frequency $f_{\max} = f_c v_{\max}/c$ depends on the carrier frequency f_c, maximum relative velocity $v_{\max}$ of the AUVs, and the speed of sound c in water. The channel distortion is represented by the phase shift ϕ_i and angle of arrival β_i. According to Equations (4.1-4.2), the channel autocorrelation matrix would depend on the channel coefficients a_i, the spread of time delays τ_i, and the maximum Doppler frequency $f_{\max}$ [72, 64]. At the receiver end, the received signal $y(t)$ is given by

$$y(t) = \int_{-\infty}^{\infty} h(t,\tau)x(t-\tau)d\tau + w(t), \tag{4.3}$$

where $h(t,\tau)$ represents the delay-Doppler spread function and $w(t)$ represents the noise present in the underwater channel. Based on the universal approximation theorem, Equation (4.3) suggest that a black-box neural network would accurately model the underwater communication channel if sufficient amount of training data $\{x(t), y(t)\}$ is available. Moreover, the scattering from the entire water body, as well as the seafloor and sea surface enters into the DNN model implicitly through a large amount of training data [69, 62, 88, 89].

Applying Fourier transform, the convolutional integral, Equation (4.3), is expressed in the frequency domain as

$$Y_k = H_k X_k + W_k. \tag{4.4}$$

Equation (4.4) provides the pilot-assisted channel estimation using the OFDM modulation, which is optimal for spectral efficiency [76, 22, 74]. Knowing that the OFDM technique is highly efficient in dealing with sparse multipath channels affected by Doppler shifts, a context-aware deep learning approach that integrates OFDM with LSTM networks can add innovation to underwater acoustic communication.

4.2.2 Context-Aware Deep Learning for Symbol Detection

This Section presents a context-aware deep learning technique for joint channel estimation and signal detection. Figure 4.1(a) shows the OFDM link between two underwater AUVs, where AUV_1 sends a binary data stream to AUV_2. At both AUVs, the binary data stream is encoded into symbols according to a modulation scheme, where cyclic prefix is considered to account for the multipath propagation and Doppler spread (i.e. intersymbol interference). The entire OFDM link is robust to represent the contextual information of the complex underwater channel.

The contextual information can be lost in classical pilot-assisted channel estimation. Figure 4.1(b-c) shows two types of pilot arrangements. Each filled circle (•) denotes a pilot subcarrier, where the constellation of the transmitted symbols are known to both AUVs [76, 119]. In this situation, pilot-assisted channel estimation not only loses the contextual information, but also suffers from trade-offs between pilot overhead, complexity, and estimation error.

The proposed deep learning method provides a nonlinear map between a block of received symbol $\mathcal{Y}$ and a particular constellation point $\mathcal{C}$ that corresponds to the received symbol Y_k. This procedure is schematically illustrated in Figure 4.2. The detection of constellation points of received symbols at the k-th subcarrier Y_k ($k = 1, \ldots, n_s$) involves an entire OFDM block, which naturally contains rich contextual information. Thus, our first innovative idea has combined the contextual information of wireless channel with the neural network framework (e.g. [100]).

Note also that the proposed method does not directly estimate the channel parameters and becomes aware of the environmental variability such as multipath and Doppler effects because neural networks account for the variability of symbols across an entire OFDM block. As shown in the next Section, this approach will not have the same pilot overhead as in classical OFDM techniques [22].

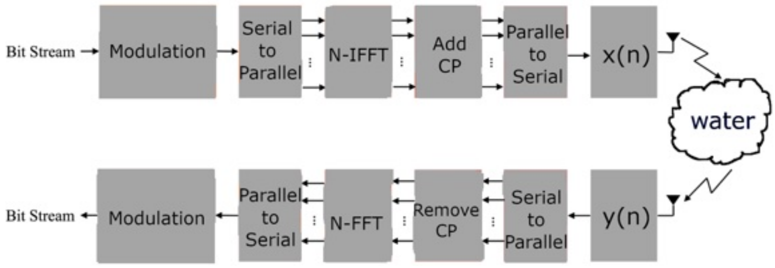
Thus, the second innovative idea has resolved the problem of pilot overhead. To achieve this, we have employed the classification technique using an LSTM network that learns the sequence of symbols of an OFDM block and assigns the network output to a corresponding constellation point. In other words, the assigned constellation point becomes the class of the OFDM block, where the LSTM framework does not depend on the pilots.

We note that learning the contextual information involves extracting complex, high-dimensional patterns from training data. According to Equation 4.4, each received symbol is associated with an OFDM constellation point. Incorporating contextual information in the LSTM networks leads to effectively recovering the constellation point of the received symbol Y_k. It is crucial to learn the inherent pattern of all subcarriers with an OFDM block. While feedforward neural networks appear to be appropriate in achieving a black-box channel model (e.g. Equation 4.3), Equation (4.4) suggest that LSTM is suitable to detect the contextual information.

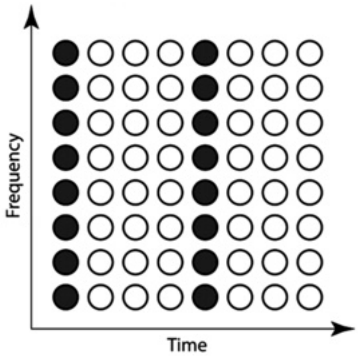
4.3 The Proposed Design of the LSTM Neural Network

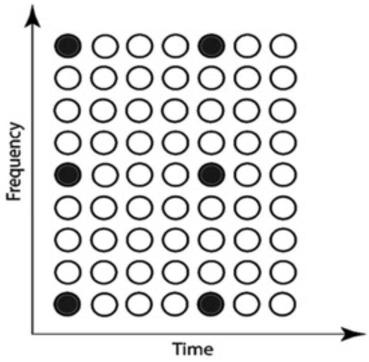
4.3.1 LSTM-based Receiver Design

The underlying premise of the proposed underwater acoustic receiver is to integrate OFDM communications in deep learning by using LSTM neural networks. Although deep learning has recently been very successful in wireless OFDM communications, their applications in underwater environment is not fully matured. The proposed LSTM-based receiver has been designed by considering that a transmitted OFDM packet through neural networks (see figure 4.2) consists of only two blocks: $\boldsymbol{x}_p$ and $\boldsymbol{x}_d$. Here, $\boldsymbol{x}_p$ contains pilot symbols and $\boldsymbol{x}_d$ contains data symbols. Pilot and data symbols are respectively denoted by a block with $\bullet$ and a block with $\circ$,



(a)





(b) (c)

Figure 4.1: (a) A block diagram showing the transmission and reception of a bit stream through water. An illustration of (b) block-type and (c) lattice type pilot arrangements, where • represents the pilot symbols, and ◦ represents the data symbols.

as indicated in figure 4.1 (b-c). Let $\mathcal{Y}$ denote the received OFDM packet. The idea is to extract the critical features through the pilot block and utilize the data block to enhance the learning capability. In the LSTM framework, the OFDM symbols are treated sequentially, and thus we have considered pilot block and data block to inject contextual information during the training process. While the design is fundamentally different from what has been considered for underwater communication, the process would train a neural network that learns the following map:

$$f_{w,b} : \mathcal{Y} \to \mathcal{C} \tag{4.5}$$

where $\mathcal{C}$ is the set of phase-shift keys. It is important to note that the constellation of pilot symbols (•) is not known to the receiver during the training of the neural network. Instead, a constellation is assigned to packet $\mathcal{Y}$ as the label for supervised learning.

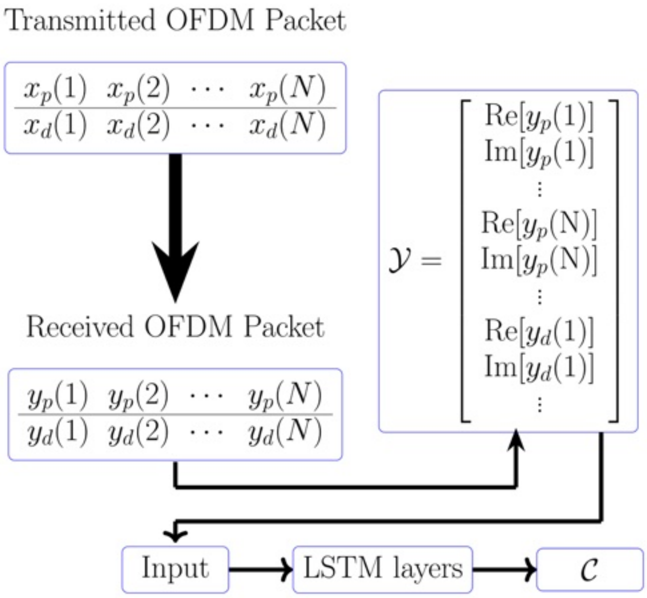


Figure 4.2: Signal detection process for a deep learning-based neural network model. Subscripts *p* and *d* represent the pilot and data symbols, respectively. The network learns the class $\mathcal{C}$ from the training data $\mathcal{Y}$.

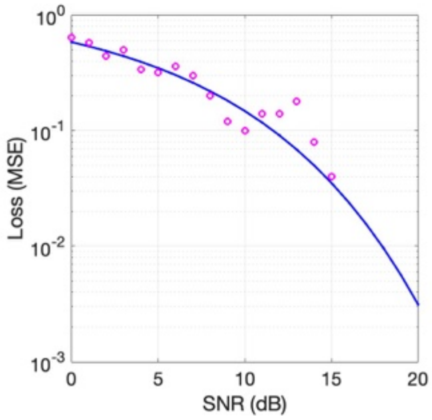
4.3.2 Verification of LSTM-based symbol error rate

Although deep learning-based joint channel estimation and signal detection have recently received a growing popularity, very little is known about the decoding performance of this approach. To address this research gap, this Section presents exact error rate characterizations under the assumption of gaussian channel and additive white noise.

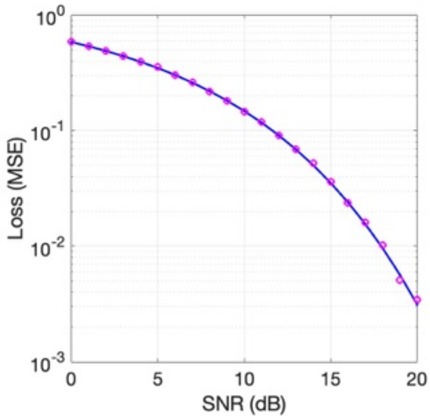
Consider an LSTM-based decision function $\mathcal{D}_A\{\cdot\}$ applied to the symbol constellation, where the symbol error rate (SER) is

$$\text{SER} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{\boldsymbol{x}^* \neq \hat{\boldsymbol{x}}_i\}} \quad \hat{\boldsymbol{x}}_i = \mathcal{D}_A\{H^\dagger \boldsymbol{y}_i\}. \tag{4.6}$$

Here, $\hat{\boldsymbol{x}}_i$ denotes the LSTM decision of the i-th received packet, $\boldsymbol{x}^*$ denotes the true symbol, and $\mathbb{1}_{\{\boldsymbol{x}^* \neq \hat{\boldsymbol{x}}_i\}}$ denotes the indicator function, which equals 1 when the detected symbol is incorrect.



(a)



(b)

Figure 4.3: A plot of the verification curve of the channel estimation technique. (a) $N = 10^2$ packets. (b) $N = 10^5$ packets

4.3.3 Symbol Detection Error

In the QPSK (quadrature phase shift keying) modulation scheme, two binary bits are modulated simultaneously such that the carrier phase of each pair of bits are shifted by either 0^o, 90^o, 180^o, or 270^o [74]. In this analysis, assume a random noise with mean $\mu = 0$ and variance $N_0/2$, where the estimated symbol $\hat{x}$ follows a Gaussian probability distribution [120, 74]:

$$p(\hat{x}) = \frac{1}{\sqrt{\pi N_0}} e^{-\frac{\hat{x}^2}{N_0}}. \tag{4.7}$$

Since the neural network architecture learns the optimal mapping between the received noisy input and the true transmitted symbols, we considered the Monte Carlo method for estimating the uncertainty of symbol detection. Thus, the probability that the symbol s_1 (e.g. $0°$ shift) has been correctly detected is

$$p(\hat{x}|s_1) = 1 - \text{erfc}\left(\sqrt{\frac{E_s}{2N_0}}\right) + \frac{1}{4}\text{erfc}^2\left(\sqrt{\frac{E_s}{2N_0}}\right). \tag{4.8}$$

To get the total symbol error probability for a neural network architecture, we assume that at least one of the symbols were decoded incorrectly (see ref. [74], chapter 3). Hence, the probability of symbol error rate $\text{SER} = 1 - p(\hat{x}|s_1)$ is given as:

$$\text{SER} = \text{erfc}\left(\sqrt{\frac{E_s}{2N_0}}\right) + \frac{1}{4}\text{erfc}^2\left(\sqrt{\frac{E_s}{2N_0}}\right). \tag{4.9}$$

where E_s/N_0 represents the signal-to-noise ratio (SNR) of the underwater environment. If sufficient data is available, then the deep learning technique has the ability to correctly recognize the pattern from any noisy measurement. Figure 4.3 shows that the SER performance follows Equation (4.9) if the number of data increases. In particular, when neural networks learn the nonlinear map between the input and output data, they may perform better than some of the conventional methods. This is a potential advantage, which attracted many researchers for adopting deep learning techniques in wireless communication [26, 121]. In Section 4.4, we present simulation results that compares the SER of the LSTM network with conventional techniques.

4.4 Results and Discussion

The SER performance of deep learning-based symbol detection in underwater channels depends on the complexity and the environmental conditions. Estimating a lower-bound for the SER performance of a deep learning-based symbol detection is not straightforward because neural networks are non-linear and highly adaptive. However, we can utilize the perfect channel state information and an information-theoretic approach to provide insights into the performance of LSTM networks in symbol detection. In this work, we assessed the LSTM-based symbol detection performance by training the neural network on a multipath underwater acoustic channel. The SER performance of the LSTM approach was compared with that of the conventional MMSE approach and the Monte-Carlo simulation.

4.4.1 Training LSTM Network

Table (4.2) shows the parameters for the training data, where acoustic communication occured between two AUVs in an underwater environment. In the present study, a packet of the training data contains two blocks of OFDM symbols (i.e. pilots x_k^p and data x_k^d) representing bit streams $\{00, 01, 11, 10\}$ of four constellations. For each received symbol $y_k \in \mathcal{Y}$, the trained LSTM determines the corresponding QAM constellation point while considering the contextual information of the multipath underwater environment. In other words, the transmitting AUV sends a message bit stream that is detected by the receiver AUV.

Table 4.2: A List of Parameters for Channel Estimation

Parameters	Specifications
number of subcarriers	64
number of pilots	8, 16, 64
Modulation constellation scheme	QPSK
number of channel paths	20
Length of cyclic prefix	16
Carrier frequency	5000 MHz
Number of packets N	10000
LSTM network	with 16 hidden units

At the receiver AUV, each pair of OFDM blocks is represented as a sequence of $2n_s$ complex numbers. These sequences carry contextual information that we want to extract through the training of our LSTM network. The real and imaginary part of each complex number were arranged to form a feature vector $\mathcal{Y}$ of $4n_s$ real numbers. The LSTM network, trained on this data, identifies the pattern of each pair to maximize the probability of correctly mapping them to the corresponding constellation points. In contrast to the classical methods, the LSTM network knows no information of the channel and pilot symbols. However, in the context of underwater communication, the proposed arrangement of the training data aims to detect the contextual information underlying the transmitted symbols without prior knowledge of channel statistics and pilots.

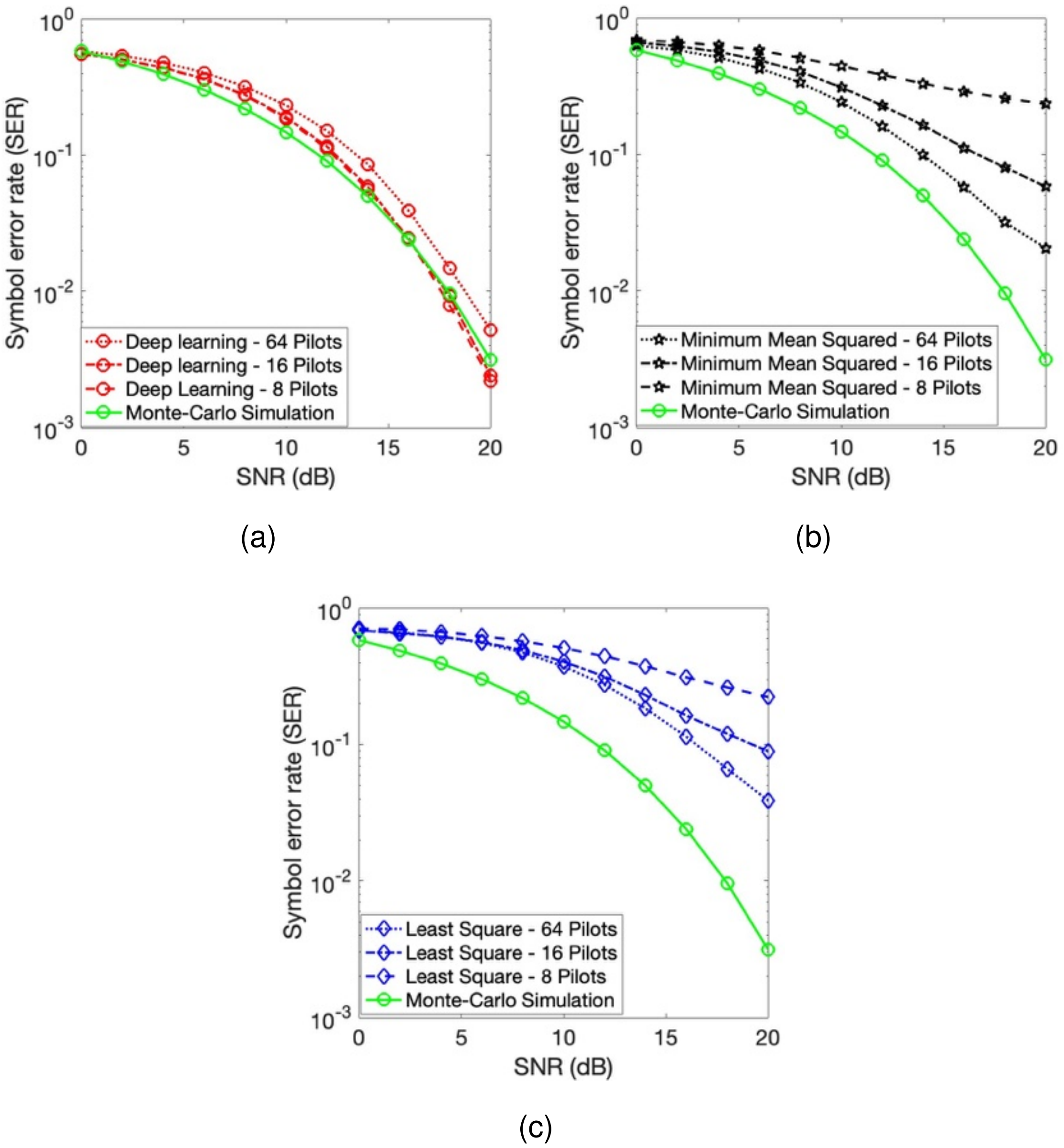


Figure 4.4: The effect of pilot overhead on the performance of (a) deep learning, (b) miminum mean square error, (c) least squares

4.4.2 Comparison between Deep Learning and Conventional Approaches

To provide a fair comparison with classical MMSE techniques, we wish to incorporate pilot symbols in training the LSTM network. We do this by considering the effect of pilot symbol arrangement on the LSTM-based OFDM receiver and comparing it with conventional methods.

To provide a fair comparison with classical MMSE technique, we need to incorporate the effects of pilot symbols in training the LSTM network. We compared the performance of the

LSTM-based OFDM receiver with that of conventional methods by considering the effects of pilot symbol arrangements. Pilot arrangement is critical to the performance of conventional methods, which must be optimized for individual communication scenario. In this Section, we have presented numerical test results showing that the performance of LSTM-based symbol detection can be insensitive to pilot arrangment for any communication scenario.

Figure 4.4 compares the SER performance of the LSTM network with respect to conventional methods. We trained the network offline using $10,000$ OFDM packets. The performance of the LSTM network was tested using another set of $10,000$ packets, which were not seen by the network during the offline training phase. The training was done in an underwater environment with a fixed SNR of 40dB. A high SNR was considered to ensure that the learning is not biased with the environmental noise. To test the learning performance of the LSTM network, the test data was taken from a noisy underwater environment with SNR values within a range between 0dB and 20dB. The LSTM results were compared with the LS and MMSE results. The effects of the number of pilots (in LS and MMSE results) were investigated by considering 8, 16, and 64 pilots, where the total number of subcarriers per OFDM block is 64.

The results indicate that the SER values gradually decrease as the SNR increases. To assess the rate of convergence, the SER profile of each of the three methods (LS, MMSE, and LSTM) was also compared with that of the Monte-Carlo simulation (denoted by the "Symbol Error Probability" in Figure 4.4). For the block-type pilot arrangement, each subcarrier of the pilot block is a pilot symbol, which were used by the LS and MMSE methods for channel estimation. For the lattice-type arrangement, pilots were interleaved with information symbols in the pilot block.

In the present test, the performance of the LSTM network is almost insensitive to the change of pilot arrangement. Moreover, the LSTM performance is better than that of the LS and the MMSE for each values of SNR. The SER values of the LSTM estimator were not affected when the number of pilots was reduced. In contrast, the SER values did not decrease as SNR increases for both the LS and MMSE when the number of pilots were 8 and 16 in block-type arrangements. An important and interesting observation is that the performance of the LSTM estimator is not sensitive to the number of pilots for the cases tested by this study.

4.4.3 Impact of LSTM Network Architecture

For a successful application of deep learning techniques in underwater acoustic communications, only a few recent works investigated how to generalize a neural network architecture to the underlying communication problems [122, 59, 89]. The Doppler effect is a potential source of communication error. Here, we considered that one of the nodes moves at a speed of 5m/s. The resulting Doppler spread is approximately 20Hz, where the carrier frequency is 6×10^3Hz, and the sound speed is 1.5×10^3m/s [38]. Another source of uncertainty is due to the variation of the sound speed with the temperature, salinity, water depth, and sediment layer, which may contribute to an additional $20 - 40\%$ noise into the receiver (see [62]).

This Section investigates the performance of three neural network architectures considering Doppler effects in a complex underwater environment. The first architecture (referred to as LSTM) is a recurrence neural network that consists of an input gate, memory cell, forget gate, and output gate. In the second architecture (LSTM1), one additional fully-connected layer was inserted. In the third architecture (LSTM2), a bidirectional LSTM framework was considered. We show that by performing an end-to-end training of a LSTM network, one can optimize the estimation accuracy of transmitted symbols, thereby providing significant system level improvements [58]. In the following analysis, the networks were trained by gradually increasing the number of training data up to $10\,000$ packets, as well as varying the SNR values of the underwater environment between 5dB and 40dB. For each of these three architectures, this study also tested the sensitivity of the number of neurons. In other words, the present hyperparameter tuning study has considered most of the possible ways that an LSTM-based supervised learning can optimize the weights and bias values during the training process. Note that the SNR value of the underwater environment was fixed during the offline training stage. However, the trained network was tested for an online underwater environment in which the

SNR values varied between 0dB and 20dB.

Figures 4.5, 4.6, and 4.7 compare the effect of LSTM, LSTM1, and LSTM2 architectures, respectively, on the underwater OFDM receivers. For each of these networks, we repeated the training in four underwater environments, characterized by the SNR: 5dB, 10dB, 20dB, and 40dB. For each of the environments, we compared the symbol error rate by varying the SNR of test data between 0dB and 20dB. The LSTM results were compared with LS and MMSE results, as well as the symbol error probability of Monte-Carlo simulation. The symbol error rate was compared for each of the training environment individually for better clarity. The results presented in these Figures (4.5-4.7) indicate that an LSTM-based OFDM receiver can be robust and relatively insensitive to a new underwater environment. As we can see, the deep learning method outperformed the LS and MMSE for the considered Doppler spread. According to the universal approximation theorem, the performance of a neural network-based estimator would increase if more training data becomes available. Since the ocean has numerous autonomous vehicles and sensors, the large volume of data collected by such sensors can be utilized in the future for developing relatively robust LSTM-based receivers that may outperform in a complex underwater environment.

4.4.4 Doppler Shift

To integrate artificial intelligence in the AUV-aided underwater search-and-rescue operation, this Section has evaluated the effect of Doppler shifts in the LSTM-based symbol detection technique. The Doppler estimation approach can occur in two stages. The first stage can employ the LSTM-based technique to recover the transmitted symbols. The second stage can estimate the Doppler rate using the estimated symbols and other informations obtained during the first stage.

As we know, acoustic signals transmitted by an AUV generally experience severe multipath effects. Thus, we investigated the effects of multipath propagation and Doppler frequency on the SER performance of the LSTM equalizer. The Doppler effect due to the relative motion

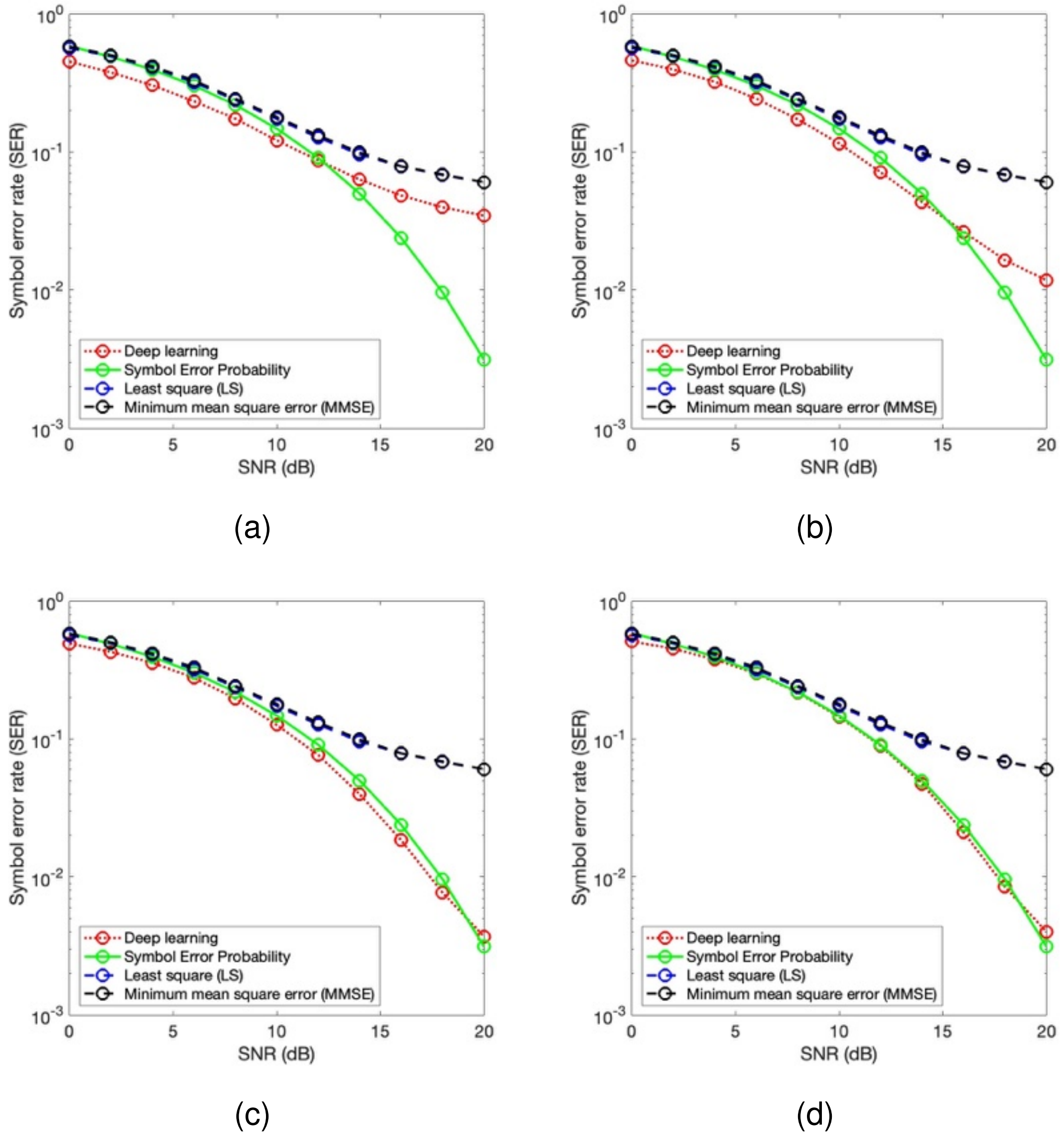


Figure 4.5: Single-layer unidirectional architecture, LSTM, and hyperparameter tuning. A comparison of the SER performance between the deep learning results (using the single-layer network) and the other methods for four training environments, which are characterized by SNR values: (a) 5dB, (b) 10dB, (c) 20dB, (d) 40dB.

of the nodes was considered, which stretches or compresses the multipath underwater signals and reduces system reliability. The SER performance of the LSTM network was assessed by comparing the LSTM results with that of the MMSE and the perfect CSI criteria.

The LSTM network was trained with signals having a bandwidth between 20 and 30kHz in a shallow water environment with a depth of 20m. The training data incorporated the Doppler

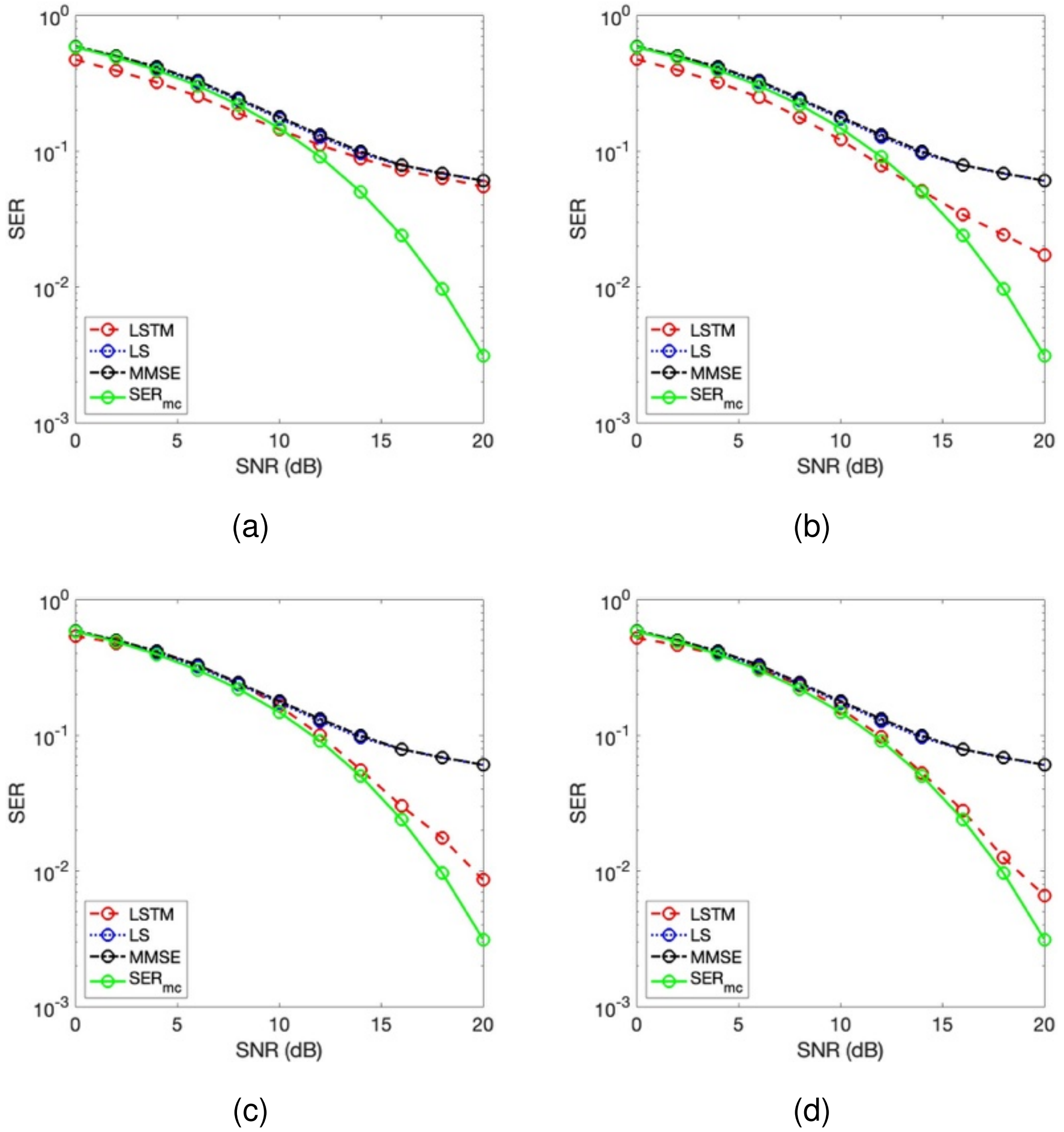


Figure 4.6: Multi-layer unidirectional architecture, LSTM1, and hyperparameter tuning. A comparison of the SER performance between the deep learning results (using the multi-layer network) and the other methods for four training environments, which are characterized by SNR values: (a) 5dB, (b) 10dB, (c) 20dB, (d) 40dB.

effect with respect to the relative node speed between 5 and 10m/s. The receiver was placed 5m deeper than the transmitter. The angle of arrival was varied to alter the Doppler rate in training data. The SNR of the training data was 40dB. Table 4.3 shows relevant parameters of the corresponding shallow water acoustic channel.

In an underwater environment, multiple reflections from the sea surface and the seabed

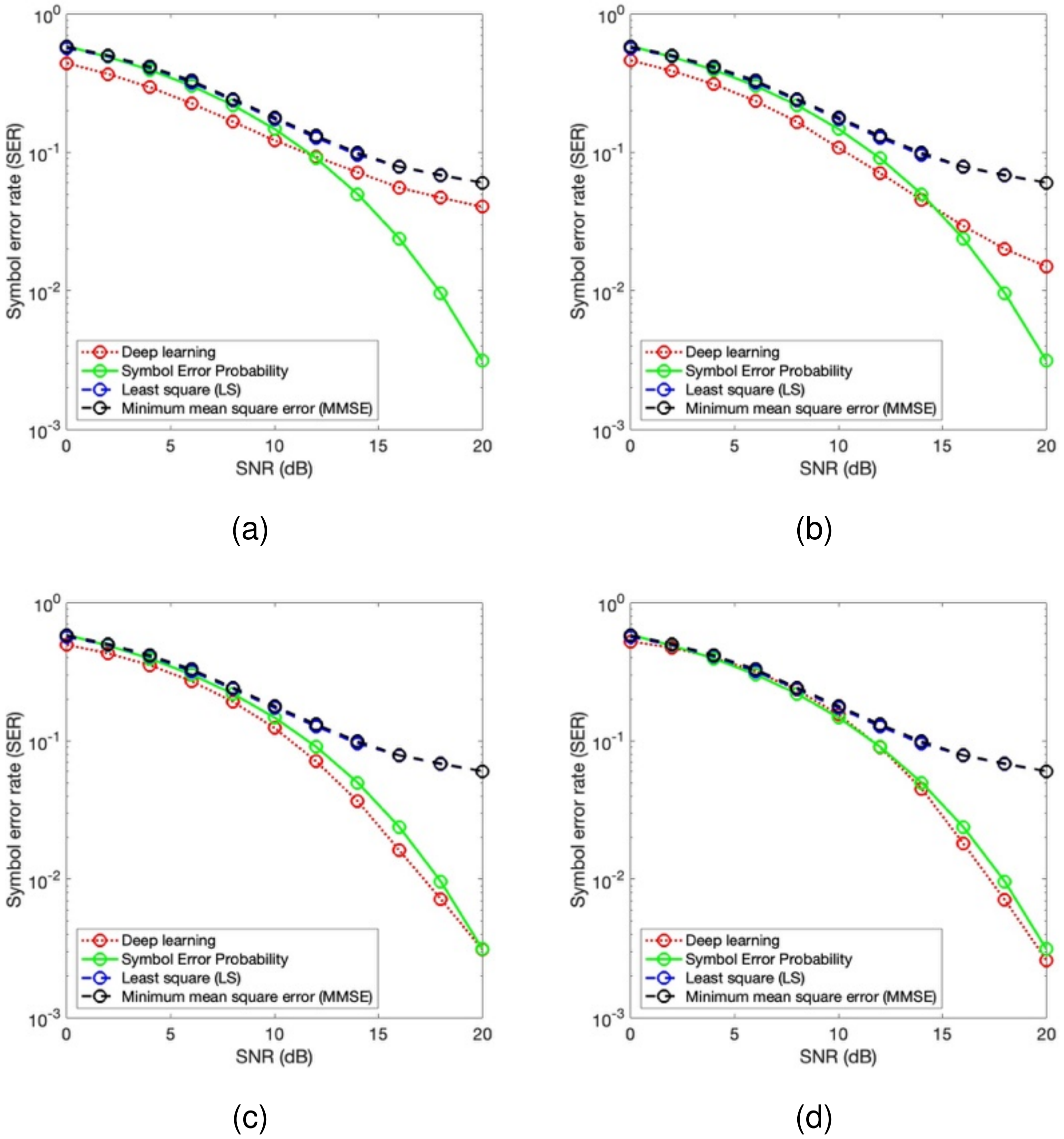


Figure 4.7: Multi-layer bidirectional architecture, LSTM2, and hyperparameter tuning. A comparison of the SER performance between the deep learning results (using the bi-directional multilayer network) and the other methods for four training environments, which are characterized by SNR values: (a) 5dB, (b) 10dB, (c) 20dB, (d) 40dB.

lead to multipath propagation, which produces severe Doppler effect. The combined effect of the linear transformation and the non-linear activation of neural networks allows for easy and effective reduction of multipath interference. Figure 4.8 shows the SER performance of the LSTM-based neural network technology when the delay spreads and Doppler shifts belonging to each paths is different. The result shows that the neural networks learn the frequency offset of

Table 4.3: A list of parameters for an underwater acoustic channel considering multipath and Doppler effects.

Parameter	Value
Communication distance	1000m
Depth of transmitter	10 m
Depth of receiver	15 m
Underwater sound speed	1538 m/s
Absorption coefficient	2dB/km
Water depth	20 m
Center frequency	25kHz
Number of paths	20

received OFDM signals due to Doppler effects. Such a frequency offset represents a distortion and rotation of transmitted signals within the QPSK mapping vectors. In other words, the proposed deep learning technique is efficient to determine the Doppler shift by estimating the constellation point of the received OFDM symbols and the corresponding rotated angles. More specifically, such an AI technology is suitable for underwater target localization.

4.5 Conclusion

This paper has presented a deep learning-based algorithm for OFDM symbol detection, aimed at optimizing communication between underwater AUVs. We have proposed an LSTM-based deep learning technique to account for contextual information in underwater communication. We carried out a fair comparison between the LSTM training algorithm and the MMSE technique for symbol detection. The effectiveness of both approaches was evaluated in terms of maximum achievable symbol error rate.

The results demonstrate that the LSTM-based technique bypasses the channel estimation and detects the symbols directly based on offline training. The LSTM technique is able to

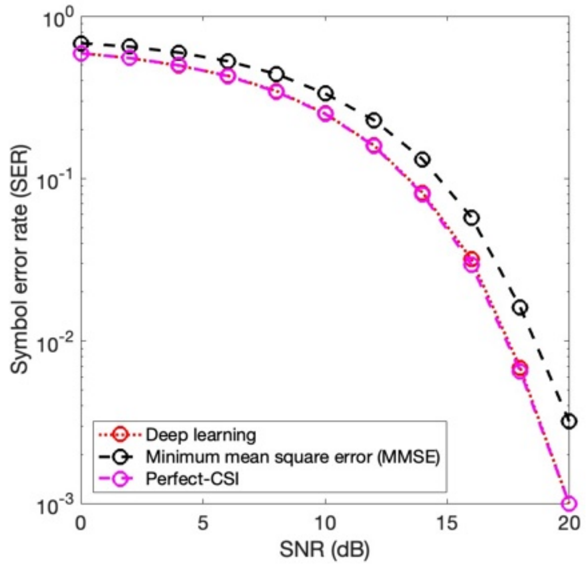


Figure 4.8: A comparison between the LSTM and the MMSE symbol recovery technique with respect to the theoretical bound when given an underwater channel that underwent the Doppler effect.

achieve an acceptable performance under any time-varying communication channel with multipath effects and Doppler shifts. This conclusion is supported by numerical results and theoretical benchmark [26].

The generalizability of the proposed LSTM-based method to any time-varying multipath channel is attributed to its integration of contextual information through OFDM modulation. This, of course, comes at the cost of cyclic prefix in OFDM. However, the cost is mitigated through the offline training. A breakthrough contribution of this research is the demonstration of the mathematically rigorous integration between the OFDM technique and the LSTM network. Moreover, our research lays the idea of theory-trained neural networks, where the theoretical knowledge of signal processing is utilized to improve the predictive capabilities of neural networks.

Furthermore, this study investigated the concept of theory-trained neural network, where the foundational signal processing knowledge is utilized to enhance the integration of artificial intelligence in communication and signal processing. Unlike traditional black-box neural networks, the theory-trained approach is transparent and interpretable. This is a key advancement for future development of deep learning in communication engineering and signal processing.

Chapter 5

Conclusions and Future Research Directions

The thesis provides a comprehensive overview of context-aware deep learning for optimizing underwater robots. Specifically, it investigates underwater acoustic communication systems in the context of optimizing robotic operations. We summarize state-of-the-art acoustic communication techniques, highlighting their applicability in fully autonomous actions and interactions of robots in the deep sea environment. The primary investigations and outcomes offer a thorough review of the context-aware deep learning techniques for developing autonomy in underwater robots. Additionally, three key technical questions regarding the applications of deep neural networks were explored. Below is a summary of the outcomes of three research projects. For each project, future extensions have been discussed. Future extensions of this research are also highlighted.

5.1 Summary of Investigations

- **Deep Learning-Based Signal Classification for Underwater Object Tracking.**

  The first question deals with how the classification algorithm can be incorporated into underwater signal processing despite the time-varying effects of the underwater environment. This research proposed a signal classification methodology to distinguish the underwater cavitation signals from the environmental cues. Since signals are characterized by temporal histories due to multipath Doppler effects, Chapter 2 investigated recurrent neural network (RNN) for signal classification.

This research took advantage of LSTM as a recurrent neural network, which learns signal properties by directly processing the discrete-time history of signals without requiring any feature extraction methods. This project is aimed at technological developments of accurately sensing stealthily operating AUVs, which means that some targets do not send active sonar signals as they intend to operate stealthily without exposing their locations. The cavitation sound from their engine becomes the acoustic signature of such targets. However, the hostile underwater channel hinders LSTM from efficiently learning signal features. To bypass the interference from the environmental cues, we apply Fourier transform to incorporate a spectrum-sensing method with the LSTM network. Numerical experiments were considered for both temporal learning and spectral learning. In other words, the LSTM model was trained on signals in the time-domain and the frequency-domain, with the results compared to analyze their performance. Experimental results show that the LSTM network is more accurate at learning spectral information compared to direct application in the time-domain. These results suggest two novel aspects of deep learning-based localization. First, channel equalization and feature extraction are essential for optimal classification. Second, the LSTM network is more efficient in sequence learning if the signals are not convolved with secondary effects such as the turbulent underwater channel.

Future directions: Although this research focuses on the classification of underwater acoustics for target localization, it does not directly address target localization itself. An potential direction for future work is to extend the acoustic classification algorithm by integrating it with localization techniques, such as the time-reversal mirror method, to track the underwater objects.

- **Context-Aware Neural Network for Underwater Multipath Underwater Channels with Doppler Shifts.**

  The second question was how to deal with limited availability of true underwater data, and how to incorporate contextual information in deep learning. Such contextual factors include temperature, salinity, and turbulent fluctuation of water pressures. Chapter 3 developed a context-aware convolutional neural network (CNN) technique to ensure optimal wireless communication in a time-varying underwater environment by utilizing classical OFDM techniques. The premise behind the proposed technique is that pilot-based wireless communication via OFDM is equivalent to the data recovery problem. For classic data-recovery techniques, they are not usually cost-effective for underwater acoustic communication. In contrast, deep learning methods shift the computational burden to offline training process, which only needs to be executed once. We observe that the context-aware neural network is different than the physics-informed neural network (PINN), which is a specialized deep learning technique that takes into consideration the physical laws of the underwater environment. Nevertheless, both approaches are functionally similar, but applied in different disciplines. Numerical experiments were considered to compare the proposed neural network with classical data recovery techniques such as decision feedback channel equalization (DFCE).

  Future directions: The research on context-aware neural network can be further extended by applying this domain-specific deep learning model into a network of underwater sensors. With this extension, we are provided with an opportunity for the context-aware CNN to detect any anomalies in the information propagated within the Internet of Underwater Things (IoUT). This is because the CNN can take in the OFDM channel matrix as its input data and encode the matrix into a low-dimensional feature space. Due to this encoding, any outliers are more easily evident.

- **LSTM-based OFDM Receiver for Underwater Acoustic Communication**

  The third question was how to optimally detect symbols transmitted underwater. Conventional techniques for symbol detection are limited by time variability, narrow bandwidth, multipath effect, frequency selective fading, Doppler effect, etc. Thus, we turn to deep learning models, in particular the LSTM. This project aims to improve the underwater surveillance and object tracking tasks that bypasses the complexity of channel estimation. The outcome of this project has the potential to be a minimally viable commercial product that recovers the symbols transmitted underwater. Numerical experiments were considered to assess the information recovery model in terms of its robustness to pilot symbol overhead and Doppler shift, as well as compare the system with conventional techniques such as minimum mean squared error (MMSE).

  Future directions: An extension of this research involve the application of LSTM-based symbol detection for a network of AUVs in various industrial work, including sub-sea exploration, oil and mineral extraction, submarine detection in warfare, etc. Using LSTM-based direct symbol detection can significantly enhance the capabilities of AUVs by improving communication, navigation, and data processing. However, it is necessary to utilize real-world data for assessing the performance of the proposed LSTM methodology. In addition, the performance of the proposed symbol detection methodology for underwater target tracking needs to be assessed. This work is currently underway.

5.2 Discussion and Final Thoughts

The thesis focuses on improving underwater communication and navigation of AUVs using deep learning techniques integrated with domain knowledge from underwater acoustics. Underwater robotics is a rapidly evolving and important area of research with acoustic communication playing a crucial role in its advancements. Recent studies suggest that combining deep learning with underwater communication is reshaping the foundation of underwater robotics, driving autonomy, collaboration, and mission success in the challenging and dynamic underwater environment.

One of the primary challenges in this field is the limited availability of real-world data for training deep learning models. Context-aware deep learning has emerged as a promising approach to address this limitation. Such a technique has the promise to improve the efficiency and robustness of underwater networks despite the high cost and difficulty of acquiring underwater measurements. However, effectively incorporating domain knowledge into neural networks remain an open research challenge. This thesis explores three key aspects of context-awareness, highlighting both its potential advancements and inherent limitations. The thesis has outlined future developments for optimizing underwater robotics by incorporating deep learning-enhanced acoustic communication

Bibliography

[1] D. Cressey, “Oceans are’worth us $24 trillion’,” *Nature*, 2015. [Online]. Available: https://doi.org/10.1038/nature.2015.17394

[2] O. Onasami, D. Adesina, and L. Qian, “Underwater acoustic communication channel modeling using deep learning,” in *Proceedings of the 15th International Conference on Underwater Networks & Systems*, no. 9, 2022, pp. 1–8.

[3] Z. Gong, C. Li, and F. Jiang, “A machine learning-based approach for auto-detection and localization of targets in underwater acoustic array networks,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15 857–15 866, 2020.

[4] J. Zhang, G. Han, J. Sha, Y. Qian, and J. Liu, “AUV-Assisted subsea exploration method in 6G enabled deep ocean based on a cooperative pac-men mechanism,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 1649–1660, 2022.

[5] Z. Zhao, R. Zhao, J. Xia, X. Lei, D. Li, C. Yuen, and L. Fan, “A novel framework of three-hierarchical offloading optimization for mec in industrial iot networks,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5424–5434, 2020.

[6] U. Behrje, C. Isokeit, B. Meyer, and E. Maehle, “A robust acoustic-based communication principle for the navigation of an underwater robot swarm,” in *2018 OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO)*, 2018, pp. 1–5.

[7] Y. Yokota and T. Matsuda, “Underwater communication using UAVs to realize high-speed AUV deployment,” *Remote Sensing*, vol. 13, no. 20, p. 4173, 2021.

[8] W. Zhao, Z. Tang, Y. Yang, L. Wang, S. Lan *et al.*, "Cooperative search and rescue with artificial fishes based on fish-swarm algorithm for underwater wireless sensor networks," *The Scientific World Journal*, vol. 2014, 2014.

[9] J. R. Lowell, "Military applications of localization, tracking, and targeting," *IEEE Wireless Communications*, vol. 18, no. 2, pp. 60–65, 2011.

[10] H. Messer, A. Zinevich, and P. Alpert, "Environmental monitoring by wireless communication networks," *Science*, vol. 312, no. 5774, pp. 713–713, 2006.

[11] I. F. Akyildiz and M. C. Vuran, *Wireless sensor networks*. John Wiley & Sons, 2010.

[12] J. Preisig, "Acoustic propagation considerations for underwater acoustic communications network development," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, no. 4, pp. 2–10, 2007.

[13] Aug 2023. [Online]. Available: https://krakenrobotics.com/products/synthetic-aperture-sonar/

[14] J. Kim and S.-C. Yu, "Convolutional neural network-based real-time rov detection using forward-looking sonar image," in *2016 IEEE/OES Autonomous Underwater Vehicles (AUV)*, 2016, pp. 396–400.

[15] M. H. Essai Ali, "Deep learning-based pilot-assisted channel state estimator for OFDM systems," *IET Communications*, vol. 15, no. 2, pp. 257–264, 2021.

[16] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*, 2nd ed. Springer Publishing Company, Incorporated, 2011.

[17] G. Qiao, Z. Babar, L. Ma, S. Liu, and J. Wu, "MIMO-OFDM underwater acoustic communication systems—a review," *Physical Communication*, vol. 23, pp. 56–64, 2017.

[18] P. Dahl, "High-frequency forward scattering from the sea surface: the characteristic scales of time and angle spreading," *IEEE Journal of Oceanic Engineering*, vol. 26, no. 1, pp. 141–151, 2001.

[19] T. Yang and T. Yates, "Acoustic inversion of bottom reflectivity and bottom sound-speed profile," *IEEE Journal of Oceanic Engineering*, vol. 21, no. 4, pp. 367–376, 1996.

[20] L. Huang, Y. Wang, Q. Zhang, J. Han, W. Tan, and Z. Tian, "Machine learning for underwater acoustic communications," *IEEE Wireless Communications*, vol. 29, no. 3, pp. 102–108, 2022.

[21] C. S. Clay and H. Medwin, *Acoustical oceanography: principles and applications*. Wiley 1st edition, 1977.

[22] L. Gaudio, G. Colavolpe, and G. Caire, "OTFS vs. OFDM in the presence of sparsity: A fair comparison," *IEEE Transactions on Wireless Communications*, vol. 21, no. 6, pp. 4410–4423, 2022.

[23] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis *et al.*, "Deep learning for computer vision: A brief review," *Computational intelligence and neuroscience*, vol. 2018, 2018.

[24] D. W. Otter, J. R. Medina, and J. K. Kalita, "A survey of the usages of deep learning for natural language processing," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 2, pp. 604–624, 2020.

[25] L. Deng and J. Platt, "Ensemble deep learning for speech recognition," in *Proc. Interspeech 2014*, 2014, pp. 1915–1919.

[26] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, 2018.

[27] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep learning-based channel estimation," *IEEE Communications Letters*, vol. 23, no. 4, pp. 652–655, 2019.

[28] X. Gao, S. Jin, C.-K. Wen, and G. Y. Li, "Comnet: Combination of deep learning and expert knowledge in OFDM receivers," *IEEE Communications Letters*, vol. 22, no. 12, pp. 2627–2630, 2018.

[29] M. Chen, "Underwater acoustic signal recognition based on salient feature," 2024. [Online]. Available: https://arxiv.org/abs/2312.13143

[30] F. Ahmad, M. Z. Ansari, R. Anwar, B. Shahzad, and A. Ikram, "Deep learning based classification of underwater acoustic signals," *Procedia Computer Science*, vol. 235, pp. 1115–1124, 2024.

[31] H. Niu, E. Ozanich, and P. Gerstoft, "Ship localization in santa barbara channel using machine learning classifiers," *The journal of the acoustical society of America*, vol. 142, no. 5, pp. EL455–EL460, 2017.

[32] R. Diamant, H.-P. Tan, and L. Lampe, "LOS and NLOS classification for underwater acoustic localization," *IEEE Transactions on Mobile Computing*, vol. 13, no. 2, pp. 311–323, 2014.

[33] H. Niu, X. Li, Y. Zhang, and J. Xu, "Advances and applications of machine learning in underwater acoustics," *Intelligent Marine Technology and Systems*, vol. 1, no. 1, p. 8, 2023.

[34] Q. Zheng, M. Yang, J. Yang, Q. Zhang, and X. Zhang, "Improvement of generalization ability of deep CNN via implicit regularization in two-stage training process," *IEEE Access*, vol. 6, pp. 15 844–15 869, 2018.

[35] B. Wang, H. Dai, K. Xu, Y. Sun, Y. Zhang, and P. Li, "A signal processing method of

OFDM communication receiver based on CNN," *Physical Communication*, vol. 59, p. 102055, 2023.

[36] P. Dong, H. Zhang, G. Y. Li, I. S. Gaspar, and N. NaderiAlizadeh, "Deep CNN-based channel estimation for mmwave massive MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 5, pp. 989–1000, 2019.

[37] A. Pradhan, S. Das, and D. Dayalan, "A two-stage CNN based channel estimation for OFDM system," in *2021 Advanced Communication Technologies and Signal Processing (ACTS)*, 2021, pp. 1–4.

[38] K. Suzuki, S. Sakamoto, S. Yoshizawa, T. Tsukui, and T. Saito, "Multipath and doppler shift modeling under highly reflective environment for underwater acoustic communication," in *OCEANS 2019 MTS/IEEE SEATTLE*, 2019, pp. 1–4.

[39] A. Mitrović, S. Djukanović, and M. Radonjić, "Implementation of signal classification using frequency spectrum features on the raspberry pi platform," in *2017 25th Telecommunication Forum (TELFOR)*, 2017, pp. 1–4.

[40] T. D. Pham, "Time–frequency time–space lstm for robust classification of physiological signals," *Scientific reports*, vol. 11, no. 1, p. 6936, 2021.

[41] S. Cen, D. O. Kim, and C. G. Lim, "A fused CNN-LSTM model using fft with application to real-time power quality disturbances recognition," *Energy Science & Engineering*, vol. 11, no. 7, pp. 2267–2280, 2023.

[42] Y. Zhang, Y. Zhang, and Y. Zhang, "Music genre classification with lstm based on time and frequency domain features," in *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, vol. 5. IEEE, 2021, pp. 1856–1860.

[43] Y.-X. Lu, X.-B. Jin, J. Chen, D.-J. Liu, and G.-G. Geng, "F-SE-LSTM: A time series anomaly detection method with frequency domain information," 2024. [Online]. Available: https://arxiv.org/abs/2412.02474

[44] D. P. S. Parihar, A. Agarwal, and M. Agrawal, "Time reversal mirror: Temporal and spatial focusing tool," in *OCEANS'10 IEEE SYDNEY*, 2010, pp. 1–7.

[45] Z. Noor, S. Hina, F. Hayat, and G. A. Shah, "An intelligent context-aware threat detection and response model for smart cyber-physical systems," *Internet of Things*, vol. 23, p. 100843, 2023.

[46] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3039–3071, 2019.

[47] N. Jain, A. Gupta, and V. A. Bohara, "PCI-MDR: Missing data recovery in wireless sensor networks using partial canonical identity matrix," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 673–676, 2019.

[48] Z. Gao, W. Cheng, X. Qiu, and L. Meng, "A missing sensor data estimation algorithm based on temporal and spatial correlation," *International Journal of Distributed Sensor Networks*, vol. 11, no. 10, p. 435391, 2015.

[49] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, pp. 1–37, 2011.

[50] M. Jamshidian and M. Mata, "Advances in analysis of mean and covariance structure when data are incomplete," in *Handbook of latent variable and related models*. Elsevier, 2007, pp. 21–44.

[51] M. Liao, D. Shi, Z. Yu, Z. Yi, Z. Wang, and Y. Xiang, "An alternating direction method

of multipliers based approach for pmu data recovery," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4554–4565, 2019.

[52] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[53] O. Edfors, M. Sandell, J.-J. van de Beek, S. Wilson, and P. Ola Borjesson, "OFDM channel estimation by singular value decomposition," in *Proceedings of Vehicular Technology Conference - VTC*, vol. 2, 1996, pp. 923–927.

[54] J.-J. van de Beek, O. Edfors, M. Sandell, S. Wilson, and P. Borjesson, "On channel estimation in OFDM systems," in *1995 IEEE 45th Vehicular Technology Conference. Countdown to the Wireless Twenty-First Century*, vol. 2, 1995, pp. 815–819.

[55] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: Useful tips and tricks," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1527–1550, 2017.

[56] K. Ardah, B. Sokal, A. L. F. de Almeida, and M. Haardt, "Compressed sensing based channel estimation and open-loop training design for hybrid analog-digital massive MIMO systems," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 4597–4601.

[57] F. Gómez-Cuba, "Compressed sensing channel estimation for OTFS modulation in non-integer delay-Doppler domain," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 1–6.

[58] A. S. M. Mohammed, A. I. A. Taman, A. M. Hassan, and A. Zekry, "Deep learning channel estimation for OFDM 5g systems with different channel models," *Wireless Personal Communications*, vol. 128, no. 4, pp. 2891–2912, 2023.

[59] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[60] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2017, pp. 1–5.

[61] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 852–855, 2018.

[62] R. Jiang, X. Wang, S. Cao, J. Zhao, and X. Li, "Deep neural networks for channel estimation in underwater acoustic OFDM systems," *IEEE Access*, vol. 7, pp. 23 579–23 594, 2019.

[63] K. Thakkar, A. Goyal, and B. Bhattacharyya, "Emergence of deep learning as a potential solution for detection, recovery and de-noising of signals in communication systems," *International Journal of Intelligent Networks*, vol. 1, pp. 119–127, 2020.

[64] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1708–1721, 2010.

[65] Y. Zhao, H. Yu, G. Wei, F. Ji, and F. Chen, "Parameter estimation of wideband underwater acoustic multipath channels based on fractional fourier transform," *IEEE Transactions on Signal Processing*, vol. 64, no. 20, pp. 5396–5408, 2016.

[66] B. Li, S. Zhou, M. Stojanovic, L. Freitag, and P. Willett, "Multicarrier communication over underwater acoustic channels with nonuniform doppler shifts," *IEEE Journal of Oceanic Engineering*, vol. 33, no. 2, pp. 198–209, 2008.

[67] R. Jiang, S. Cao, C. Xue, and L. Tang, "Modeling and analyzing of underwater acoustic channels with curvilinear boundaries in shallow ocean," in *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2017, pp. 1–6.

[68] U. Pillai, V. Murthy, and I. Selesnick, "Missing data recovery using low rank matrix completion methods," in *2012 IEEE Radar Conference*. IEEE, 2012, pp. 0101–0106.

[69] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[70] S. L. Brunton and J. N. Kutz, *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2022.

[71] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural networks*, vol. 4, no. 2, pp. 251–257, 1991.

[72] X. Cheng, D. Liu, C. Wang, S. Yan, and Z. Zhu, "Deep learning-based channel estimation and equalization scheme for fbmc/oqam systems," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 881–884, 2019.

[73] T. Xu, Z. Tang, G. Leus, and U. Mitra, "Multi-rate block transmission over wideband multi-scale multi-lag channels," *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 964–979, 2013.

[74] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[75] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[76] J.-W. Choi and Y.-H. Lee, "Optimum pilot pattern for channel estimation in OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2083–2088, 2005.

[77] M. Chitre, S. Shahabudeen, L. Freitag, and M. Stojanovic, "Recent advances in underwater acoustic communications & networking," in *OCEANS 2008*, 2008, pp. 1–10.

[78] R. Diamant and L. Lampe, "Low probability of detection for underwater acoustic communication: A review," *IEEE Access*, vol. 6, pp. 19 099–19 112, 2018.

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

[80] Q. Wu, Y. Zhang, Z. Yang, and M. R. Shikh-Bahaei, "Deep learning for secure uav swarm communication under malicious attacks," *IEEE Transactions on Wireless Communications*, vol. 23, no. 10, pp. 14 879–14 894, 2024.

[81] N. F. Josso, C. Ioana, J. I. Mars, and C. Gervaise, "Source motion detection, estimation, and compensation for underwater acoustics inversion by wideband ambiguity lag-doppler filtering," *The Journal of the Acoustical Society of America*, vol. 128, no. 6, pp. 3416–3425, 2010.

[82] S. Dörner, S. Cammerer, J. Hoydis, and S. Ten Brink, "Deep learning based communication over the air," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 132–143, 2017.

[83] L. Dai, R. Jiao, F. Adachi, H. V. Poor, and L. Hanzo, "Deep learning for wireless communications: An emerging interdisciplinary paradigm," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 133–139, 2020.

[84] X. Chen, Q. Jiang, N. Su, B. Chen, and J. Guan, "LFM signal detection and estimation based on deep convolutional neural network," in *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2019, pp. 753–758.

[85] S. Wang, R. Yao, T. A. Tsiftsis, N. I. Miridakis, and N. Qi, "Signal detection in uplink time-varying OFDM systems using RNN with bidirectional LSTM," *IEEE Wireless Communications Letters*, vol. 9, no. 11, pp. 1947–1951, 2020.

[86] R. Su, Z. Gong, C. Li, and S. Han, "High accuracy AUV-aided underwater localization: Far-field information fusion perspective," *IEEE Transactions on Signal Processing*, vol. 72, pp. 1877–1891, 2024.

[87] A. Cochocki and R. Unbehauen, *Neural networks for optimization and signal processing*. John Wiley & Sons, Inc., 1992.

[88] A. Yang, P. Sun, T. Rakesh, B. Sun, and F. Qin, "Deep learning based OFDM channel estimation using frequency-time division and attention mechanism," in *2021 IEEE Globecom Workshops (GC Wkshps)*, 2021, pp. 1–6.

[89] Q. Hu, F. Gao, H. Zhang, S. Jin, and G. Y. Li, "Deep learning for channel estimation: Interpretation, performance, and comparison," *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2398–2412, 2021.

[90] Y. Wang and H. Peng, "Underwater acoustic source localization using generalized regression neural network," *The Journal of the Acoustical Society of America*, vol. 143, no. 4, pp. 2321–2331, 2018.

[91] S. Kim, J. Son, and B. Shim, "Energy-efficient ultra-dense network using lstm-based deep neural networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4702–4715, 2021.

[92] M. Honkala, D. Korpi, and J. M. Huttunen, "Deeprx: Fully convolutional deep learning receiver," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3925–3940, 2021.

[93] N. Athreya, V. Raj, and S. Kalyani, “Beyond 5g: Leveraging cell free tdd massive MIMO using cascaded deep learning,” *IEEE Wireless Communications Letters*, vol. 9, no. 9, pp. 1533–1537, 2020.

[94] Z. Wang, F. Pu, X. Yang, N. Chen, Y. Shuai, and R. Yang, “Online LSTM-based channel estimation for HF MIMO SC-FDE system,” *IEEE Access*, vol. 8, pp. 131 005–131 020, 2020.

[95] H. Zhao, F. Ji, M. Wen, H. Yu, and Q. Guan, “Multi-task learning based underwater acoustic OFDM communications,” in *2021 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. IEEE, 2021, pp. 1–5.

[96] W. Samek, S. Stanczak, and T. Wiegand, “The convergence of machine learning and communications,” 2017. [Online]. Available: https://arxiv.org/abs/1708.08299

[97] C. Çiflikli, A. T. Özşahin, and A. Ç. Yapici, “Artificial neural network channel estimation based on levenberg-marquardt for OFDM systems,” *Wireless personal communications*, vol. 51, pp. 221–229, 2009.

[98] S. J. Nawaz, S. Mohsin, and A. A. Ikaram, “Neural network based MIMO-OFDM channel equalizer using comb-type pilot arrangement,” in *2009 International Conference on Future Computer and Communication*, 2009, pp. 36–41.

[99] C.-H. Cheng, Y.-H. Huang, and H.-C. Chen, “Channel estimation in OFDM systems using neural network technology combined with a genetic algorithm,” *Soft Computing*, vol. 20, pp. 4139–4148, 2016.

[100] X. Gao, S. Jin, C.-K. Wen, and G. Y. Li, “Comnet: Combination of deep learning and expert knowledge in OFDM receivers,” *IEEE Communications Letters*, vol. 22, no. 12, pp. 2627–2630, 2018.

[101] X. Ma, Z. Gao, F. Gao, and M. Di Renzo, "Model-driven deep learning based channel estimation and feedback for millimeter-wave massive hybrid MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2388–2406, 2021.

[102] Y. Zhang, S. Zhang, Y. Wang, Q. Liu, and X. Li, "Model-driven deep-learning-based underwater acoustic otfs channel estimation," *Journal of Marine Science and Engineering*, vol. 11, no. 8, p. 1537, 2023.

[103] S. Alam, C. Li, and R. Venkatesan, "Underwater acoustic target classification using LSTM network," in *IEEE International Conference on Communications*, 2025, submitted.

[104] ——, "CNN-based context-aware deep learning of underwater acoustic channels," Unpublished manuscript, 2025.

[105] S. Alam and C. Li, "Robust data driven methods for underwater object localization," in *31st NECEC Conference*, 2022.

[106] S. A. U. Haq, A. K. Gizzini, S. Shrey, S. J. Darak, S. Saurabh, and M. Chafii, "Deep neural network augmented wireless channel estimation for preamble-based OFDM phy on zynq system on chip," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 31, no. 7, pp. 1026–1038, 2023.

[107] Q. Li, Y. Gong, F. Meng, L. Han, and Z. Xu, "A novel channel estimation method based on deep neural network for OTFS system," in *2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2022, pp. 1–6.

[108] Y. Bao, Z. Tan, H. Sun, and Z. Jiang, "SimNet: Simplified deep neural networks for OFDM channel estimation," in *2020 IEEE 3rd International Conference on Information Communication and Signal Processing (ICICSP)*, 2020, pp. 348–352.

[109] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," *Advances in neural information processing systems*, vol. 26, pp. 1–9, 2013.

[110] T. Ngo, B. T. Kelley, and P. Rad, "Deep learning based prediction of Doppler shift for mobile communications," in *2021 Telecoms Conference (ConfTELE)*, 2021, pp. 1–6.

[111] H. He, J.-H. Wang, S. Kojima, K. Maruta, and C.-J. Ahn, "Regression CNN based fast fading channel tracking using decision feedback channel estimation," *Journal of Signal Processing*, vol. 27, no. 3, pp. 49–57, 2023.

[112] W. S. Costa, J. L. Samatelo, H. R. Rocha, M. E. Segatto, and J. A. Silva, "Direct equalization with convolutional neural networks in OFDM based VLC systems," in *2019 IEEE Latin-American Conference on Communications (LATINCOM)*, 2019, pp. 1–6.

[113] A. Ilioudi, A. Dabiri, B. J. Wolf, and B. De Schutter, "Deep learning for object detection and segmentation in videos: Toward an integration with domain knowledge," *IEEE Access*, vol. 10, pp. 34 562–34 576, 2022.

[114] M. Delamou, A. Bazzi, M. Chafii, and E. M. Amhoud, "Deep learning-based estimation for multitarget radar detection," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*. IEEE, 2023, pp. 1–5.

[115] A. K. Gizzini, M. Chafii, S. Ehsanfar, and R. M. Shubair, "Temporal averaging LSTM-based channel estimation scheme for IEEE 802.11p standard," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 1–7.

[116] R. Shankar, "Bi-directional LSTM based channel estimation in 5G massive MIMO OFDM systems over TDL-C model with rayleigh fading distribution," *International Journal of Communication Systems*, vol. 36, no. 16, p. e5585, 2023.

[117] S. Hassan, P. Chen, Y. Rong, and K. Y. Chan, "Doppler shift compensation using an

LSTM-based deep neural network in underwater acoustic communication systems," in *OCEANS 2023-Limerick*. IEEE, 2023, pp. 1–7.

[118] R. Huang, W. Zhang, A. Kundu, C. Pantofaru, D. A. Ross, T. Funkhouser, and A. Fathi, "An LSTM approach to temporal 3d object detection in lidar point clouds," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*. Springer, 2020, pp. 266–282.

[119] Y. Liao, Y. Hua, and Y. Cai, "Deep learning based channel estimation algorithm for fast time-varying MIMO-OFDM systems," *IEEE Communications Letters*, vol. 24, no. 3, pp. 572–576, 2019.

[120] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital communication*. Springer Science & Business Media, 2012.

[121] T. Wang, C.-K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," *China Communications*, vol. 14, no. 11, pp. 92–111, 2017.

[122] Y. Yang, F. Gao, X. Ma, and S. Zhang, "Deep learning-based channel estimation for doubly selective fading channels," *IEEE Access*, vol. 7, pp. 36 579–36 589, 2019.