UNDERWATER TARGET RECOGNITION BASED ON LOFAR SPECTRUM ENHANCEMENT: A DEEP LEARNING APPROACH

Sai Swaroop Kottapalli
School of Computer Science and Engineering
Digital University of Kerala
Thiruvananthapuram, Kerala, India
Kottapalli.cs22@duk.ac.in

Sumit Datta
School of Electronic Systems and Automation
Digital University of Kerala
Thiruvananthapuram, Kerala, India
sumit.datta@duk.ac.in

***Abstract*—The ocean containing mineral, marine and chemical resources has attracted much interest in the world because of its economic potential and consequent activities like seabed exploration ,oil platform surveillance and economical fish location and so forth. It is crucial to identify the objects in the underwater environment as either safe marine creatures or dangerous/ potentiative obstacles with the help of underwater acoustic target recognition technology for improving maritime safety. This thesis is focused on how the technique of deep Learning may be used for underwater target recognition with an emphasis on its superiority in feature extraction, noise immerse and identification speed. Developing a deep learning method for underwater target recognition is a new approach to the problem since it identifies features from raw signals without the necessity of direct human intervention, decreases the size of the feature space, fits the target maps, reduces the noise influence, and enhances the model's ability to generalize. The study addresses four main areas: the difficulties of data acquisition and generation and the denoise of audio with the help of optimization method FISTA [2]; the architecture of deep neural network AutoEncoder; and preparation of data to reduce the impact of environmental noise. The above mentioned can be partially mitigated with traditional noise reduction techniques because of dynamic changes in the environment and nonlinear characteristics of underwater signals. In this research, an innovation process is proposed to better complete the spectrogram extraction, including but not limited to wavelet decomposition+low-frequency analysis, the encoder-decoder full convolution for the deep network [1], and the mapping optimization for feature extraction and timing correlation. The denoised signals are then distinguished utilizing ResNet18 which indicate huge enhancement in the underwater acoustic signal processing and recognition. This detailed work showcases the capability of Deep Learning in enhancing the process of target identification beneath water and this feature would definitely enhance safety measures as well as efficiency of underwater activities.**

***Index Terms*—Underwater Target Recognition, LOFAR Spectrum, Deep Learning, Maritime Safety, FISTA, Optimization.**

I. INTRODUCTION

Underwater Acoustic Signal Processing plays important role in many marine operations such as navigation and communication and study of marine life. Underwater communication environment is characterized by challenging factors such as signal fading, signal bouncing, and different levels of background noise which makes challenging application such as signal clean-up and target detection.

Research domain of underwater acoustic signal processing is of paramount importance in several aspects. Acoounting for over seventy percent of the earths surface, underwater environments are critically important to both the natural function of the earth and man productivity. Understanding and monitoring these environments are essential for several reasons:Understanding and monitoring these environments are essential for several reasons:

a) Maritime Navigation and Safety: Underwater acoustic signal processing is required for certification of the path for submarines and identification of threats to different sea-going vessels which includes naval as well as commercial vessels.

b) Environmental Monitoring: They are applied to monitor changes in the behavior of marine animals, investigate violations of the law and fishing Gillnets, as well as geophysical formations. It helps in supporting the conservation programs as well as the natural resource management of seas and oceans.

c) Underwater Communication: Radio waves are significantly attenuated into water which means that there is need to enhance the extent of signal handling especially for underwater communication, naval submersible, and stations for explorations.

d) Resource Exploration: From this case, acoustic surveys assist in determining resources in water, particularly the presence of oil and gas fields. Signal processing helps distinguish these resources and their positions for retrieval and use with a great deal of precision.

e) Economic and Safety Considerations: There is major economic importance of marine resources such as mineral, living, and chemical resources. It is crucial to have a view over the seabed and the oil platforms because of economical benefits, but also for safety issues for the fish populations. The goal of the emerging underwater acoustic target recognition applying requirements to distinguish between standard marine life and actual threats to the safety of shipping vessels is obvious.

The difficulty in implementing USAP are presented here in two sub-sections as follows: The environment deep in the water is complicated and signals are plentiful, which result in an increase of difficulties in signal processing. Environmental features like temperature, salinity, pressure, movements of marine organisms, and man-made noises alter the acoustic signals which in turn makes the signal interpretation challenging.

f) Signal Attenuation: Sounds in water are damped, especially the high-frequency ones, which creates difficulties in the decision of whether the sounds are of a distant signal and its interpretation.

g) Multipath Propagation: Signals bounce on the sea surface and seabed and hence there are multiple signals getting reflected back to the receiver making the signal analysis a little complex.

h) Background Noise: The sounds from dolphins, tankers, wind, and waves are always present in natural environment and distort the waves while eradicating the raw quality of signals received in the process of data interpretation.

A. Motivation

Separation of noise in underwater acoustics is the basic and crucial step to enhance the signal quality as it directly impacts the SNR of further application. It is therefore very hard for them to work well as the under water environment is noisy and keeps on varying, unlike the methods such as spectral subtraction and Wiener filtering. These methods employ the predetermined noise models and are likely to miss the normal fluctuation of underwater acoustic environment. To overcome such limitations, there are the recent advancements of deep learning that may present the solutions.

In this research, the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) for the initial noise reduction of underwater acoustic signals is recommended. FISTA is a reliable optimisation technique in attaining the approximate solution to optimisation problems through enhancement of the rate of convergence of the iterative thresholding that aid in filtering of noise in the signals [3].

After that, we apply the FISTA algorithm for denoising followed by Applying Fully Convolutional Encoder-Decoder Network (FCEDN) for enhancement. Thus, the FCEDN architecture facilitates learning of mappings from noisy signals to clean signals through the use of richer training data. It consist of an encoder that is capable of finding several levels of abstraction in the input signal and a decoder that maps this encoded signal to a noiseless one. This approach not only increases the SNR but also reserves much information of the acoustic signals for the later stage.

The next crucial issue is target recognition that means choosing objects from the numerous variants, for instance, large vessels from small ones or whales from dolphins. The type of communication signals, those that are transmitted and received include different forms and this involves different environments and these are interfered with by noise making the task very complicated. Other previous methods of machine learning that have been used for feature extraction and/or classification include the Support Vector Machines (SVM) and the Principal Component Analysis (PCA). However, these methods generally yield less than ideal feature representations which further call for dose feature engineering.

In order to discharge such difficulties, this research adopts deep learning, specifically, the Convolutional Neural Network (CNN) framework, helpful in residual learning denoted as ResNet18. ResNet18 also introduces a cure for the vanishing gradient problem in deep networks and hence allows the training of very deep networks. The particularly effective structure composed of multiple layers is beneficial for image and signal classification because it is capable of hierarchal features acquisition. In the present research, ResNet18 is used for underwater acoustic signal classification, where the ShipsEar dataset, which contains recorded ship noises in actual settings, is used. The proposed SpecAugment for data augmentation and center loss contribute towards improving the discriminative capability of features learned by the network and thus high recognition accuracy of the target compared to conventional and baselines other networks.

II. Methodology

A. Data Collection

This test sample was obtained through recordings carried out with hydrophones mounted on docks and consisting of sounds produced by different vessel speeds, as well as cavitation sounds related to the docks and undocks procedures [8]. These recordings contain true vessel sounds in and environment that contains other human and natural noise and other marine mammals. The dataset includes 90 recordings in the format of . only in the wav format, 5 major classes, where each major class consists of one or more subclass. They are recorded as five-second phrases or shorter, with an option of connecting them into segments of 15 seconds up to 10 minutes long. Thus, for the analysis of the results, figure 6 highlighted the presence of different ships. Table 2 details the class divisions: Class A comprises the dredgers, fishing vessels, mussel ships, trawlers, and tug ships; Class B is for the motorboats, pilot ships, and sailing vessels; Class C for the passenger vessels; Class D for ocean liners and RORO vessels; and Class E as the natural noise which is mingled with the first four classes to forge noise-containing targets. To test the denoising capability of the model a signal with added interference consisting of two signals combined together with overlapping sounds was generated. The acoustic signal data was split into fixed epochs of 5 seconds duration and all of them were labeled; finally, 2292 sound samples were used for analysis. Random samples and target samples of the noise class were taken randomly and added to it making signals of S/N of 0dB. The dataset was then divided into validation, testing, and training sets in a ratio of 1:Thus, they selected a signal-to-noise ratio of 1:8 to evaluate the model's noise reduction capabilities.

B. LOFAR Spectrum Overview

The LOFAR (Low-Frequency Analysis and Recording) spectrum is a crucial tool in passive sonar ship target recognition. The LOFAR spectrum is used to transform signals received by passive sonar from the time domain to the time-frequency domain using the Short-Time Fourier Transform (STFT). This transformation helps in analyzing the distribution of signal frequency components over time, which is essential for detecting and recognizing underwater targets such as ships.

Calculation of LOFAR Spectrum

- **Short-Time Fourier Transform (STFT)** The STFT is performed on the signal $s(t)$ to convert it into the time-frequency domain. The formula for STFT is:

$$\text{STFT}\{s(t)\} = \int_{-\infty}^{\infty} s(t)w(t-\tau)e^{-j\omega t}dt$$

  Here, $w(t)$ is the window function that segments the signal into short, manageable frames.
- **Steps Involved**
  1) **Framing and Windowing:** The signal is divided into multiple frames, each containing N sampling points, with possible overlap between frames to ensure continuity and reduce spectral leakage.
  2) **Normalization and Decentralization:** Each frame is normalized and decentralized to make the power uniform and the mean zero.
  3) **Fourier Transform:** The Fourier transform is performed on each frame, and the spectra are arranged in the time domain to obtain the LOFAR spectrum.
- **Application and Benefits**
  1) **Line Spectrum Observation** In the case of the LOFAR spectrum reported here, one can observe line spectra that indicate general targets. This is very useful in identification and tracking of ships is especially in the process.
  2) **Handling Noise** Hence, to LOFAR spectrum can have breakpoints owing to presence of environmental noise, which is a disruption that can have impact on the target recognition. Better algorithms, including multi-step decision algorithms and deep learning, will improve the necessary LOFAR spectrum and decrease the noise, improving the recognition of the targets.

     Buffach & Plant (1984) also suggested a Sliding Window Line Spectrum Extraction technique which will be discussed in the next section.

     A sliding window technique is applied in this paper to extract the line spectra from the LOFAR spectrum. This algorithm moves along the frequency axis only but covers the entire time axis to determine the best path in the spectrum. It assists in separating and retrieving more than one line spectra that could exist and/or interfere with one another in the LOFAR spectrum.

     According to our proposed method using the LOFAR spectrum, it is possible to improve the detection/ recognition rate of passive sonar systems mainly in the noisy underwater environment.

C. Working Model

1) **Denoising using FISTA Algorithm**:
   - Apply the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) directly to the audio signals for initial denoising.
   - FISTA is used to reduce noise by solving a sparse reconstruction problem, enhancing the signal quality.
2) **Preprocessing Steps**:
   - Convert the denoised audio signal to a spectrogram (LOFAR Spectrum) using Short-Time Fourier Transform (STFT).
   - Ensure all audio samples have consistent parameters (window size, overlap, etc.).
3) **Denoising with Auto Encoder**:
   - Design and train an Auto Encoder for further spectrogram denoising.
     * **Encoder**: Layers compress the input spectrogram into a latent space representation.
     * **Decoder**: Layers reconstruct the spectrogram from the latent representation.
   - **Training**:
     * Use decomposed audio and original spectrograms for training.
     * Minimize reconstruction error.
   - **Metrics Used**:
     * The metric used for training the AutoEncoder is accuracy.
4) **Convert Denoised Spectrogram back to Audio**:
   - Apply inverse STFT to convert the denoised spectrogram back to the audio domain.
5) **Classification with ResNet18**:
   - Adapt ResNet18 to accept spectrogram inputs.
   - Train ResNet18 using the denoised spectrograms.
   - **Training**:
     * Use a labeled dataset with corresponding classes.
     * Monitor performance.
     * Avoid overfitting using techniques such as early stopping and learning rate schedules.
6) **Evaluation**:
   - Split the dataset into training, validation, and test sets.
   - Use metrics such as accuracy, precision, recall, and F1-score to evaluate the model.

D. ResNet-18

ResNet 18 is a variant of the ResNet family of models, and by this, it is meant to be a Residual Network. This deep convolution neural network handles the degradation issue in deep neural networks. This problem occurs in a scenario in which augmenting the depth of a network would still result in higher training error rates rather than the recognized fact that deeper nets should perform impressively. The novel architecture of ResNet- 18 introduces residual learning of the shortcut connections in the network, which learns residual functions concerning the input on other forms of function in the network.

- **Key Features**:
  - **Architecture**:
    - 18 layers in ResNet-18.
    - Includes one convolutional layer, several residual blocks, and one fully connected layer.
    - Residual blocks have shortcuts that enable the blocks to bypass one or many layers.

Description: The proposed methodology is divided into several key steps as outlined below:

E. FISTA Algorithm

The FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) is designed to solve convex optimization problems of the form [3]:

$$\min_{x \in \mathbb{R}^n} \left\{ F(x) = f(x) + g(x) \right\},$$

where f is a smooth convex function with a Lipschitz continuous gradient, and g is a convex function; FISTA makes the convergence rate of ISTA faster.

Algorithm

Initialization:

- Start with an initial point $x_0 \in \mathbb{R}^n$.
- Set $y_1 = x_0$ and $t_1 = 1$.

Iteration: For $k \geq 1$, compute the next iterate as follows:

$$x_k = \text{prox}_{\frac{1}{L}g}\left(y_k - \frac{1}{L}\nabla f(y_k) \right),$$

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2},$$

$$y_{k+1} = x_k + \left(\frac{t_k - 1}{t_{k+1}} \right)(x_k - x_{k-1}).$$

Mathematical Components of FISTA

Proximal Operator: The proximal operator for a function g is defined as:The proximal operator for a function g is defined as:

The proximal mapping with respect to a function g and a parameter λ of a point v is given by

$$\text{prox}_{\lambda g}(v) = \arg\min_{x} \left\{ \frac{1}{2\lambda}\|x - v\|^2 + g(x) \right\}.$$

This operator handles the non-smooth part of the objective function which is given by g.

Update Step: The update step for x_k is to first get the gradient of the smooth part f at y_k and then take 'prox' operation. It can be regarded as gradient descent step, subsequently proximal mapping which ensures that x_k possesses the structure determined by g.

Acceleration: Namely, a number t_k from the number sequence is changed, so that the convergence accelerates. The update of y_{k+1} depends on mixture of the previous iterates, x_k and x_{k-1} which add momentum into the iterations.

Convergence Analysis In theory FISTA increases the convergence rate of ISTA from $O(1/k)$ to $O(1/k^2)$. That is, it is shown that FISTA can get to a certain level of accuracy much faster than the ISTA. The core of this acceleration is the selection of the sequence t_k and the auxiliary one, y_k, allowing the convergence to be significantly accelerated while only slightly increasing the number of arithmetic operations in each iteration.

Example Algorithm

The FISTA algorithm with a constant stepsize can be summarized as:The FISTA algorithm with a constant stepsize can be summarized as:

Input: Lipschitz constant L of ∇f, the initial point x_0.

Step 0: Now let $y_1 = x_0$ and $t_1 = 1$.

For $k \geq 1$**:**s into one of the following categories depending on how it will be used: Text a product search for sale, Text b the type of advertising commonly seen in newspapers and magazines, Text c something that will be published and physically distributed, e.g., an article in a journal or booklet, and Text d another type of advertising, which is seen increasingly in Web sites and e-mail, i.e., pop-up ads and e-mails with attachments.

$$x_k = \text{prox}_{\frac{1}{L}g}\left(y_k - \frac{1}{L}\nabla f(y_k^{\star}) \right), \tag{1}$$

$$t_{k+1} = \frac{t_k + \sqrt{t_k^2 + 1}}{2} \tag{2}$$

These two recurrence formulas are the same and, on the right-hand side, give the values of the next term, t_{k+1}, through the current term t_k.

$$y_{k+1} = x_k + \left(\frac{1 - t_k}{t_{k+1}} \right) \Delta x_k \tag{3}$$

F. Algorithm Steps

Initialization:

- Choose an initial point $x_0 \in \mathbb{R}^n$.
- Set $y_1 = x_0$.
- Set $t_1 = 1$.

Iteration: For $k \geq 1$, perform the following steps:

1) Compute the proximal gradient step:

$$x_k = \text{prox}_{\frac{1}{L}g}\left(y_k - \frac{1}{L}\nabla f(y_k) \right)$$

2) Update the momentum parameter:

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$

3) Update the auxiliary variable:

$$y_{k+1} = x_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(x_k - x_{k-1})$$

This method takes advantage of the structure of the problem formulation in order to enhance the rate of convergence over the basic iterative method known as ISTA.

1) Wavelet Decomposition:

– **Objective:** Utilize wavelet transform to different the raw audio signal into many frequency segments.

– **Steps:** Following is the implementation step of the proposed model: Individual constructive interference occur in specific frequencies and hence, the signal should be divided into bands of various frequency ranges that correspond to distinct constructive interference It is possible to apply a threshold on coefficients and convert it back into an audio signal. Given an underwater acoustic signal sequence X, where x_k is a value in the sequence, and k is a time node:With reference to the notation, let X =$\{x_1, x_2, \ldots, x_N\}$, x_k is a value in the sequence and k is termed as a time node.

$$X = \{x_k\}; \tag{4}$$

with the restriction that k can only be an integer and x only being a real number.

A subsequence of X is selected and divided into even and odd segments, X_e and X_o, respectively:Sequence X is selected and a subsequence is chosen from it Elements are divided into two parts: one has even indices and the other has odd ones:

$$X_e = \{x_{2k}\} : \text{which is defined over the set of integers} \tag{5}$$

$$\{k\} \text{ and real numbers } \{x\} \tag{6}$$

$$X_o = \{x_{2k+1}\} : k \in \mathbb{Z}, \text{ x is an irrational number.} \tag{7}$$

The even sequence X_e is used to estimate the odd sequence $\hat{X}_o$: Fig 8 Odd sequence X_o is estimated the even sequence $\hat{X}_e$

$$\hat{X}_o = P(X_e). \tag{8}$$

The equation that describes the dependence of the predictor of p(.) is as follows:The equation that arises in the description of the dependence of the predictor of p(.) is given by:)

$$p(x) = x_k \tag{9}$$

$$P(X_{2k}) = \frac{X_{2k-2} + X_{2k+2}}{2} \tag{10}$$

The difference between the estimated value $\hat{X}_o$ and the actual X_o is defined as the detail signal d:The difference between the estimated value $\hat{X}_o$ and the actual X_o is defined as the detail signal d:

$$d = X_o - \hat{X}_o = X_o - P(X_e). \tag{11}$$

There are two modifications of $p(\cdot)$: a predictor $p(\cdot)$ and an updater $U(\cdot)$ They are used to keep X_e proximate to X_a and to define the update method. The second and third steps of FBP are performed where the signal is reconstructed then followed by the acquisition of the signal features from the spectrogram. Various wavelet bases are chosen for the purpose of getting the details of the signal and low frequency analysis recording method is used to extract the spectrogram characteristics of the signal. Thus, during a certain change, the manners in which corresponding frequencies occur in signal X_e are kept constant. Also, a new component referred as updater symbolized with $U(\cdot)$ is also incorporated. Thus, Eq. (3. 8) holds

$$C = X_e + U(d) \tag{12}$$

The update method can be selected from the following two functions:The update method can be selected from the following two functions:

$$U(x) = \frac{(X_{k-1} + X_{k+1})}{2} + \frac{1}{4}\text{or}$$
$$U(x) = \frac{x_k}{2}.$$

On the obtained signal, the value of threshold is imposed. As for the coefficients, the threshold Th is defined for the given problem.

$$Th = \sigma(2 \;\ln\; L)^{0.5} \tag{13}$$

$$\sigma = median(d)/0.675 \tag{14}$$

Where L is the data length of the detail signal sequence d.

$$d(i), \quad (i = 1, 2, 3, \cdots, L)$$

The following is the processing method of the threshold Th.

$$d_{Th}(i) = \begin{cases} 0, & \text{if } d(i) < Th \\ d(i) & \text{if } d(i) \geq Th \end{cases}$$

Where $d_{Th}(i)$ is the detail signal after threshold processing and reconstructed at the j^{th} level.

$$X_e = c – U(d) \tag{15}$$

$$X_o = d + P(X_e) \tag{16}$$

Place back the signals based on the positions to form original signal size.

2) Convert Decomposed Audio to Spectrogram:

- **Objective:** Transform the wavelet-decomposed audio signal into a spectrogram for visual representation.

- **Steps:**
  1) Use Short-Time Fourier Transform (STFT) to convert the decomposed audio signal into a spectrogram.
  2) Ensure all audio samples have consistent parameters (window size, overlap, etc.).

3) Denoising with Auto Encoder:

- **Objective:** Remove noise from the spectrogram using an Auto Encoder.
  Steps:
  1) Design and train an Auto Encoder for spectrogram denoising.
     Auto Encoder Structure:
     * **Encoder:** Layers to compress the input spectrogram into a latent space representation.
     * **Decoder:** Layers to reconstruct the spectrogram from the latent representation.

     This study uses a type of neural network, a fully convolutional encoder-decoder, as the basis for suppressing underwater acoustic signals. By changing the network architecture and adjusting various settings (hyperparameters), the noise-free performance can be optimized. Each layer of the network serves a specific purpose in the noise reduction process. Convolutional layers with adjustable kernel size are used to extract salient features from spectrogram data that are location independent. The encoder-decoder architecture helps highlight important data points and combine isolated local features.

     In the initial phase of the denoising process, spectrogram features obtained from wavelet analysis at low frequencies are entered into the model. Next, the encoder stage uses preset one-dimensional convolutional networks to extract enhanced features from the signal. Next, these features are fully exported using a convolutional mapping framework. Understanding the complex relationship between noise and the desired signal is essential for learning, so this structure plays an important role. Finally, the obtained features are transformed back into a time series vector through a transposed convolution process. This vector can be used to reconstruct a noise-free audio signal.
  2) **Training:** Use Decomposed Audio and original spectrograms for training and minimize the reconstruction error.
  3) **Metrics used:** Metric used for training AutoEncoder is Accuracy.

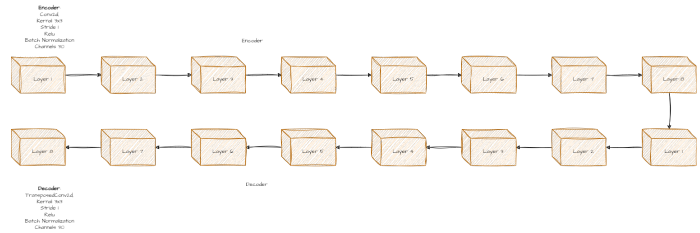

Fig. 1. Encoder Decoder Architecture for Denoise

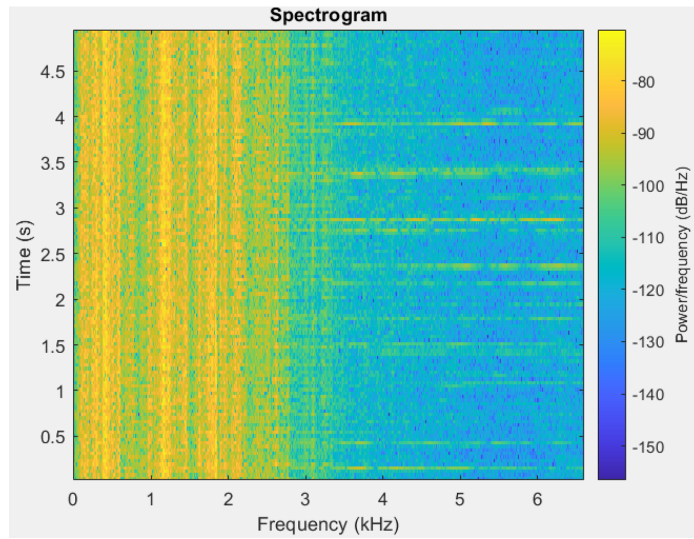

Fig. 2. LOFAR spectrum of the original signal

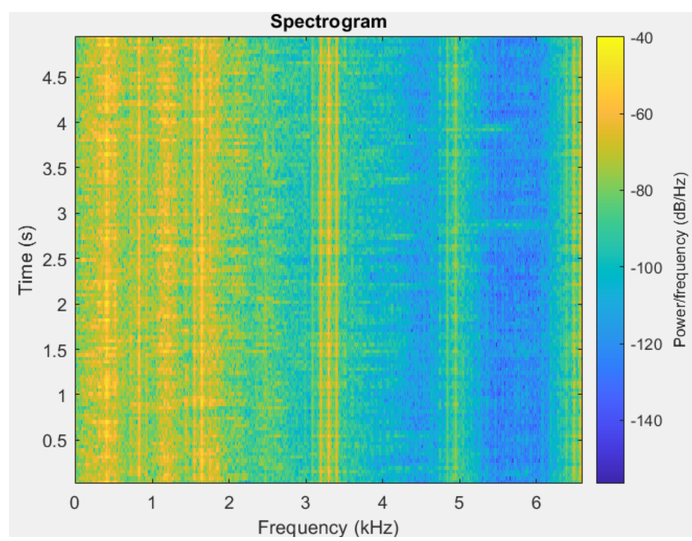

Fig. 3. LOFAR spectrum of the Denoised signal

III. ResNet-18

ResNet 18 is a variant of the ResNet family of models, and by this, it is meant to be a Residual Network. This deep convolution neural network handles the degradation issue in deep neural networks. This problem occurs in a

ENCODER	DECODER
Conv2d (Kernel 3x3, Stride 1, ReLU, BatchNorm, Input Ch 1, Out Ch 30)	TransposeConv2d (Kernel 3x3, Stride 1, ReLU, BatchNorm, Out Ch 30, Input Ch 1)
Conv2d (Kernel 3x3, Stride 1, ReLU, BatchNorm, Input Ch 30, Out Ch 30)	TransposeConv2d (Kernel 3x3, Stride 1, ReLU, BatchNorm, Out Ch 30, Input Ch 30)
Conv2d (Kernel 3x3, Stride 1, ReLU, BatchNorm, Input Ch 30, Out Ch 30)	TransposeConv2d (Kernel 3x3, Stride 1, ReLU, BatchNorm, Out Ch 30, Input Ch 30)
Conv2d (Kernel 3x3, Stride 1, ReLU, BatchNorm, Input Ch 30, Out Ch 30)	TransposeConv2d (Kernel 3x3, Stride 1, ReLU, BatchNorm, Out Ch 30, Input Ch 30)
Conv2d (Kernel 3x3, Stride 1, ReLU, BatchNorm, Input Ch 30, Out Ch 30)	TransposeConv2d (Kernel 3x3, Stride 1, ReLU, BatchNorm, Out Ch 30, Input Ch 30)
Conv2d (Kernel 3x3, Stride 1, ReLU, BatchNorm, Input Ch 30, Out Ch 30)	TransposeConv2d (Kernel 3x3, Stride 1, ReLU, BatchNorm, Out Ch 30, Input Ch 30)
Conv2d (Kernel 3x3, Stride 1, ReLU, BatchNorm, Input Ch 30, Out Ch 1)	TransposeConv2d (Kernel 3x3, Stride 1, ReLU, BatchNorm, Out Ch 1, Input Ch 30)

TABLE I
CONFIGURATION OF AUTOENCODER

scenario in which augmenting the depth of a network would still result in higher training error rates rather than the recognized fact that deeper nets should perform impressively. The novel architecture of ResNet-18 introduces residual learning of the shortcut connections in the network, which learns residual functions concerning the input on other forms of function in the network.

A. *Key Features of ResNet- 18:*

1) **Architecture:**
The details reveal that there are 18 layers in ResNet-18, sliced into several groups; that is, the model begins with one convolutional layer, contains several residual blocks, and ends with one fully connected layer. The latter enhances the system's performance if applied simultaneously in a sequence. The residual blocks in this architecture have shortcuts that enable the blocks to bypass one or many layers. Here is an outline of how the process goes: for the first layer, there is a seven by seven layer with sixty-four filters that is followed by max pooling. After that, two categories of residual blocks are applied. Each group has a couple of 3×3 convolutional layers: 2 blocks form the first group with 64 filters, the second group by 2 blocks of 128 filters, the third group by 2 blocks with 256 filters, and finally, the last group is made up of 2 blocks with 512 filters. All the downsampling here is done with stride-2 convolutions on these blocks.
2) **Residual Learning:**
The above structure explains the structure of ResNet and here we can regard the residual blocks as a basic module. Each block here rises one or more levels at least with the usage of the identity shortcut link. Thus, the network learns the mapping of residual instead of the mapping from input to the desired output, even though the latter is the goal of the function approximation problem; this is done because the former is easier to optimize than the latter.The formulation for a building block can be given in the form:In general, the language used to express a building block can be formulated as follows: Where $y = F(x, \{W_i\}) + x$ Wherein y is output defined by $F(x, \{W_i\})$ and wherein x is input and $\{W_i\}$ refers to the weights of the resided layers.
3) **Benefits:**
   - **Overcomes the Degradation Problem:** In connection with degradation, the idea behind ResNet-18 is that residual connections should be used in such a way that allows for the training of much deeper networks without the degradation of their quality.
   - **Improved Gradient Flow:** This type of connection enhances gradient flow during backpropagation, allowing for the training of deeper networks.
   - **Ease of Optimization:** Residual learning helps optimize deeper networks because, during the training process, the residual module creates a situation in which the network needs to learn only that kind of mapping being carried out by it—something fast to learn.

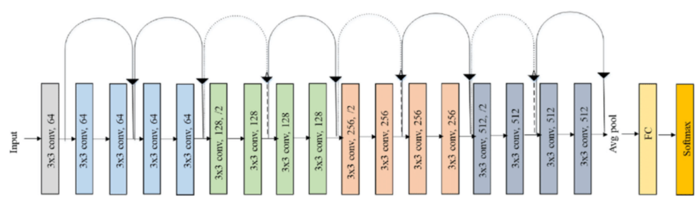

Fig. 4. Resnet18 Architecture

4) **Performance:**
Altogether, ResNet-18 appears to have enhanced accuracy with kinds of objects' recognition with the structure complexity slightly lower than ResNet-50 or ResNet-101 models, which gain more accuracy of 97.38% compared to a range of models. The network architecture is very successful in fine, detailed features within an image because of depth and residual connections. Thus, ResNet-18 is an engineering deep learning model that breaks the limitation of network depth with residual learning for handling degradation problems in deep learning and shows strong abilities in solving image recognition tasks. Its use was therefore disseminated based on design characteristics and utility for myriad types of computer vision.

IV. RESULTS

The work done with the recently introduced approach of object recognition and underwater acoustic signal denoising is presented in this section.

This section presents the findings of the work done using the introduced method of denoising the underwater acoustic signal and the object detection model in ResNet18. If classification of data samples has been done, the performance of the model can be given by classification accuracy.

Class	Classification Accuracy (%)
Average	97.38%
Class A	100.00%
Class B	92.41%
Class C	95.82%
Class D	96.54%
Class E	100.00%

TABLE II
PERFORMANCE METRICS FOR THE PROPOSED METHOD.

Metrics	Values
Accuracy	97.38%
Precision	1.0
Recall	1.0

TABLE III
PERFORMANCE METRICS FOR THE PROPOSED METHOD.

V. Discussion

The analysis and implications section of the discussion interprets the outcomes of the study and identifies their significance. Here are some key points to consider:Here are some key points to consider:

– **Performance Comparison:**It can be seen from the above power analysis that the proposed method has better advantages in terms of classification accuracy.
– **Strengths and Limitations:**The advantages of the employed approach of the proposed method are better noise reduction, and accurate object detecting. But there are also disadvantages such as computational complexity and, in certain cases, risk of overfitting should also be noted.
– **Contributing Factors:**The improvement and success of the method can be attributed to; residual learning as it creates deep neural networks that allowed an enhancement on the feature extraction of the input signal and noise suppression.
– **Real-world Applications:**The proposed method has the following scopes of interest in real life application such as Underwater navigation, Underwater biology, Underwater surveillance systems.
– **Future Research Directions:** Denoising can be conducted using Optimization techniques and they include; ISTA, FISTA, SALSA and among others. Few other deep learning techniques such as Dual Path networks and among others can be used.Then, when dealing with Signal, applying AutoEncoder to its improvement can be discussed as the possible further development.

VI. Conclusion

Deep learning techniques were introduced in this thesis with an emphasis in the enhancement of LOFAR spectrum in the aspect of underwater target recognition. The primary conclusions drawn from this research are as follows:The primary conclusions drawn from this research are as follows:

– **Deep Learning for Feature Extraction:** Several approaches can be implemented, and the most effective one is the convolutional neural networks such as ResNet18 as far as the current level of knowledge is concerned. These methods are far more effective than the conventional methods as far as the accuracy of their results is concerned.
– **Noise Reduction:** The use of wavelet decomposition for the low-frequency part introduced a more accurate denoising of the image while the fully convolutional encoder-decoder framework made the encoding-decoding process more effective. This approach minimized the effect of noise from the environment and hence an improvement in target recognition.
– **Optimization Techniques:** To this end, a mapping-based optimization technique improved the extraction of signal features and timing correlation while expanding the models' generalization of architectures.
– **Classification Performance:** By applying denoising to the signals available to the classifiers, the proposed approach of utilising ResNet18 for classifying the denoised signals clearly exhibited significant enhancements in identifying songs out of the underwater acoustic targets. The integreated feature learning superiority of the model was instrumental in attaining a high accuracy in target classification.

Acknowledgments

I would like to personally thank all those people and institutions who contributed to the completion of this thesis.

First and foremost, I express my gratitude to my advisor, Dr. Sumit Datta, for his invaluable comments, support, encouragement, and recommendations. He provided crucial guidance on refining my research question, choosing appropriate methodologies, and favored research strategies, as well as on the general approach to the research process.

Special appreciation goes to my friends and counterparts in the School of Electronics, whose support, encouragement, and productive interactions made this study proactive and stimulating. I am especially grateful to Riya Rani SS for volunteering her time, helping, and providing steadfast support.

I extend my sincere gratitude to all the administrative and technical staff of Digital University, Kerala, for their

encouragement and for providing essential assistance and technical support needed for this research.
I thank my family and friends for being with me, supporting me, and tolerating me through all these years. Your support has been the backbone of my success.
Finally, I acknowledge my indebtedness to all those who, in any manner, helped me with my thesis work. Your input in this project has been immensely helpful, and I deeply appreciate it.

REFERENCES

[1] Y. Song, Q. Chu, F. Liu, T. Wang, and T. Shen, "Underwater Acoustic Signal Noise Reduction Based on a Fully Convolutional Encoder-Decoder Neural Network," J. Ocean Univ. China, vol. 22, no. 6, pp. 1487-1496, 2023.
[2] K. Siedenburg and M. Dörfler, "Audio denoising by generalized time-frequency thresholding," in Audio Engineering Society Conference: 45th International Conference: Applications of Time-Frequency Processing in Audio, Audio Engineering Society, 2012.
[3] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imaging Sci., vol. 2, no. 1, pp. 183-202, 2009.
[4] J. Chen, B. Han, X. Ma, and J. Zhang, "Underwater target recognition based on multi-decision LOFAR spectrum enhancement: A deep-learning approach," Future Internet, vol. 13, no. 10, p. 265, 2021.
[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778.
[6] V.-S. Doan, T. Huynh-The, and D.-S. Kim, "Underwater acoustic target classification based on dense convolutional neural network," IEEE Geosci. Remote Sens. Lett., vol. 19, pp. 1-5, 2020.
[7] F. Hong, C. Liu, L. Guo, F. Chen, and H. Feng, "Underwater acoustic target recognition with ResNet18 on shipsear dataset," in 2021 IEEE 4th International Conference on Electronics Technology (ICET), 2021, pp. 1240-1244.
[8] D. Santos-Domínguez, S. Torres-Guijarro, A. Cardenal-López, and A. Pena-Gimenez, "ShipsEar: An underwater vessel noise database," *Applied Acoustics*, vol. 113, pp. 64-69, 2016.
[9] M. Irfan, Z. Jiangbin, S. Ali, M. Iqbal, Z. Masood, and U. Hamid, "DeepShip: An underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification," *Expert Systems with Applications*, vol. 183, p. 115270, 2021.