CLAUSTHAL UNIVERSITY OF TECHNOLOGY

Master's Thesis

Deep Learning Approach to Identify Locations from Speakers Emitting the Same Ultrasound Signal in an Indoor Positioning System

Author: Ankit Sharma
First Examiner: Prof. Dr. Benjamin Leiding
Second Examiner: Prof. Dr. Benjamin Säfken
External Guidance: Prof. Dr. Sascha Spors

August 21, 2023

Abstract

An Indoor Positioning System (IPS) utilising ultrasound technology over a single channel based Fingerprinting method poses significant challenges to achieve high level of accuracy. In recent years, the deep learning techniques for IPS has achieved considerable performance over the traditional machine learning techniques as per the research community. The thesis explores the accuracy of deep learning methods for an ultrasound based Indoor Positions System developed by the DEJ Technology GmbH which is also known as Koopango IPS. The research focuses on the answers to the question: How deep learning techniques can improve accuracy in predicting location coordinates in an indoor environment using ultrasound based Fingerprinting Indoor Positioning System?

This paper provides an in-depth functionality of the Koopango IPS, an overview of how data is collected and processed exploiting the concepts of audio processing, and examines deep learning methods to improve accuracy of IPS for localization of (x,y) coordinates in an indoor space using a Design Science research methodology. The comprehensive data-sets consists of audio signal recordings with corresponding location coordinates in two distinct indoor environments are used. The first data-set is used for training and evaluation of the model, whereas the second data-set evaluates the behaviour of the same model in a different environment.

The solution obtained provides a significant performance of the deep learning model with a mean squared error of 2.67 meter square and 0.55 meter square in the predicting the location (x,y) coordinates. The result represents a significant improvement of accuracy over the existing Koopango IPS using single channel playing same ultrasound signal. The thesis is concluded with the observations and the possibility of incorporating factors like reflection and reverberation in the future research.

Acknowledgements

I am grateful to all the people who contributed to the accomplishment of this thesis. I am profoundly thankful to my highly thought thesis first supervisor, Prof. Dr. Benjamin Leiding, for his extremely valuable direction, constant encouragement, and limitless know-how throughout this research work. Prof. Leiding's motivation and constructive criticism have been instrumental in moulding this thesis and improving my comprehension of the topic.

Many thanks to Prof. Dr.-Ing. Sascha Spors, for supporting and guiding me externally from the University of Rostock and bringing valuable feedback on the concepts of digital signals and machine learning. I would like to thank Prof. Dr. Benjamin Säfken as supervisor and examiner of this thesis for his support and help.

I am profoundly thankful to my colleague Bisrat Aweno from DEJ Technologys for guiding me throughout the thesis, providing me with different approaches and always being there for his support. It was a pleasure working with him. Furthermore, I would like to thank Dr. Jonas Flint, who is the founder of DEJ Technologys, for coming up with the hypothesis of this thesis and approaching me to work on it. Dr. Jonas constantly supported and motivated me throughout this thesis.

Last but not least, I desire to additionally show my genuine appreciation to my friends and parents individuals who have remained a never-ending source of urge and inspiration.

Contents

List of Abbreviations

AAG .... Abreviations Are Good

DSR .... Design Science Research

IPS .... Indoor Positioning System

LBS .... Location Based Services

GPS .... Global Positioning System

RSS .... Received Signal Strength

IR .... Infrared

RF .... Radio frequency

AOA .... Angle of Arrival

TOA .... Time of Arrival

RSSI .... Received Signal Strength Indicator

UWB .... Ultra-wideband

WSN .... Wireless Sensor Network

WLAN .... Wireless Local Area Network

RFID .... Radio-Frequency Identification

NFC .... Near Field Communication

AR .... Augmented Reality

STFT .... Short-time Fourier Transform

Chapter 1

Introduction

Indoor positioning systems (IPS) have grown in popularity in recent years, with applications including finding medical equipment or patients at the hospital, locating products in malls, real-time location information of goods or workers in logistics, interacting with museums based on location, and so on. As the demand for localisation based services (LBS) increases the demand for a highly precise IPS also increases [1]. Several indoor positioning technologies are already being researched and tested, including Radio frequency, Bluetooth, Visible light, and Ultra-wide band. However, traditional IPS methodologies like WiFi, Sensor-based, or Bluetooth lack the capability due to low accuracy and signal interference [2]. Alternatively, Ultrasound technology for LBS opens the door for a precise and reliable localisation in indoor environments [3].

The Koopango Indoor Positioning System by DEJ Technology GmbH utilises the multi-channel fingerprinting technique based on ultrasound signals technology where the localisation achieved by the system is highly accurate. In a multi-channel fingerprinting IPS, every speaker in a location has the potential to play a unique ultrasound signal to create unique environments. But practical implementation of a multi-channel system suffers from expensive infrastructure costs as it is not flexible to connect every speaker with its own line in places like Malls, Airports, Museums, along with others. The speakers in such indoor places generally share one or two lines, therefore a single-channel fingerprinting IPS can be considered as it is cost-effective. However, the single channel fingerprinting Koopango IPS is inadequate in identifying the location coordinates precisely and is far from being useful in practical settings. In this thesis, we utilize the methodologies of deep learning to improve the accuracy of the single channel fingerprinting IPS.

In addition, the influence of different indoor environments on the accuracy of the deep learning model in predicting location coordinates is investigated.

1.1 Existing Body of Knowledge

The following sections provide an overview of localisation-based services (Section 1.1.1), existing indoor positioning systems (Section 1.1.2), and the aspect of machine learning in indoor localisation (Section 1.1.3). Furthermore, the detected research gap which serves as the foundation of the thesis is addressed in the final part of this section (Section 1.1.4).

1.1.1 Location Based Services

Localisation-based services (LBS) are becoming increasingly popular in today's world, whether they are used outdoors or indoors. Depending on the application, localisation is being researched in various environments where the global positioning system (GPS) is already extensively used for outdoor localisation [4]; however, using GPS for indoor localisation has several constraints such as wall penetration due to solid and thick obstacles as well as the higher power consumption making it difficult to use GPS for indoor localisation [5]. Furthermore, even though the signals are greatly attenuated and reflected by the building materials, GPS with high sensitivity could be used for indoor localisation but the accuracy would be very low [6]. Considering the scenario, IPS can be classified based on the technology, techniques, and algorithms employed [7]. As per literature Infrared (IR), Ultrasound, Radio frequency (RF), Visible light, Audible Sound, Optical and Vision, Dead Reckoning and Magnetic are the emerging technologies for indoor localisation [2], which use different techniques like Fingerprinting, Triangulation, Trilateration and Proximity where the metrics such as Angle of Arrival (AOA), Time of Arrival (TOA) and Received Signal Strength Indicator (RSSI) are taken into consideration [8] determined by use-case.

1.1.2 Existing IPS

As per the literature, the Classification of IPS comes with the challenge as it depends on various factors like techniques, algorithms, technologies, accuracy, complexity of the system, environment challenges, and future possibilities [2]. Starting with RF-based technologies where radio frequency signals are used for localisation [9], RF has the capability to penetrate through walls, and already installed infrastructure can be used again to save cost. Sakpere et al. survey RF localisation systems which can be further divided into Bluetooth, Ultra-wideband (UWB), Wireless Sensor Network (WSN), Wireless Local Area Network (WLAN), Radio-Frequency Identification (RFID), and Near Field Communication (NFC) [10] [11] exploiting the Fingerprinting, Trilateration, and Proximity as the techniques. Considering the accuracy of the survey by Sakpere et al, UWB provides high accuracy overall among other RF localisation technologies, but UWB is only available in a limited number of high-end smartphones [12] making the system complex to use in everyday life making it not so passive system [13] which is the use-case of Koopango IPS, also RF signals are limited for Radio waves and can have propagation effects [14].

Considering Audible sound technology IPS where smartphones can be used, and at low-cost high accuracy can be achievable at the room level [15]; however, to maintain high accuracy more sensors are required if the acoustic signal is not strong enough which leads to an increase in the cost of the system [14]. Optical/Vision IPS utilises methods like Augmented Reality (AR) using the capabilities of smartphones; however, the accuracy is very low because of interference from multiple effects such as bright light and motion blur resulting in poor efficiency [4]. IR positioning systems such as Active Badge [14] make use of low-cost IR sensors, but overall of cost of the system is high, accuracy is limited, and IR waves can be interfered with by the sunlight [2].

The use of Ultrasound is extensively popular in IPS, many systems like Active Bat, Cricket system, LOSNUS [16] and Dolphin system is researched and even enhanced with different techniques as per the Sakepere et al. The Active badge uses trilateration having a fixed position of at least three microphones to receive sound. Existing ultrasound signals provide good accuracy in centimetres at room level; however, implementation in larger spaces comes with high infrastructure costs. For example; Losnus IPS offers an accuracy of 1 cm [16] but the system utilises the deployment of many sensors which could be expensive at the places like airports.

Because of its high accuracy in comparison to other techniques, fingerprinting is the most common approach for localisation as it does not require AP line-of-sight measurements, is simple, and has broad applicability in the complex indoor environment [17]. However, Fingerprinting using Ultrasound is the unique system used in Koopango IPS, which is not covered in the literature and is discussed further in the thesis.

1.1.3 Machine Learning in IPS

Taking the Machine learning aspect, literature [3], [18] suggests that using ML with Fingerprinting data enhances the accuracy, and overall quality of the system, In two different indoor environments, DeepFi which utilises the Deep learning method for Fingerprinting can effectively reduce location error when compared to three existing methods [19]. ML with WiFi fingerprinting resulted in decreased localisation error by 30 per cent [20]. The literature review also suggests integrating deep learning into fingerprinting has yielded significant gains [21] in terms of different metrics like accuracy, precision, cost, and scalability which can be utilised for further research in the thesis.

1.1.4 Contribution - Detecting a Gap

As mentioned in Section 1.1.2, fingerprinting using ultrasound technology has not yet been investigated in the literature. Therefore, this thesis aims to determine the influence of the deep learning model in improving the accuracy of single-channel Koopango IPS to predict (x,y) location coordinates in an indoor space. Furthermore, the effect of distinct environments on predicting accuracy using two different data sets has been studied.

1.2 Research Methodology and Research Questions

The goal of the thesis considerably aligns with the utilisation of design science research methodology. Henver et al. [22] explains design science research results in the creation of a purposeful IT artifact to provide a solution for a real-world problem in a specific domain. In this thesis, design science methodology is deployed to create a concrete IT artifact in order to improve the accuracy of the Koopango IPS by means of the deep learning model, which is followed by the evaluation of the model.

The following Section 1.2.1 provides a brief about the theory of Design science research methodology. In Section 1.2.2, the application of the research methodology in the thesis is discussed. Finally, Section 1.2.3 states the main research question along with the sub-questions of the thesis.

1.2.1 Design Science Research - Theory

Hevner er al. [22] propose a framework for conducting design science research which includes environment, IS research and knowledge base as depicted in Figure 1.1. The artifacts are created to the relevant need of the environment which includes people, organisations and technologies. The knowledge base contributes by applying foundations and methodologies by means of existing theories and techniques and also achieves rigor for IS research. The effectiveness and quality of the artifact are evaluated using the techniques provided knowledge base. Finally, the created artifact provides application back to the appropriate environment and expands the existing research of the knowledge base.

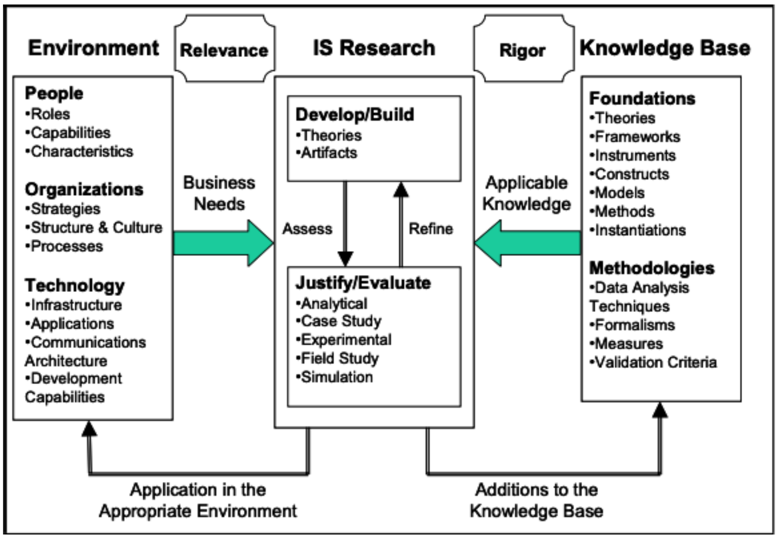

FIGURE 1.1: Design Science Research - Framework (Source: [22])

Guideline	Description
Guideline 1: Design as an Artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
Guideline 2: Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Guideline 3: Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4: Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
Guideline 5: Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
Guideline 6: Design as a Search Process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7: Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

TABLE 1.1: Design Science Research - Guidelines (Source: [22])

1.2.2 Design Science Research - Practice

Hevner et al [22] define DSR as a problem solving process and derive seven guidelines from the principles of Knowledge and understanding of a design problem where the solution is acquired by the creation and application of an artifact. The seven DSR guidelines followed in this thesis are depicted in the Table 1.1. The application of the guidelines in context of the thesis is discussed in the following sections.

1.2.2.1 Design as an Artifact

In this thesis, the design of a deep learning model tailored specially for Koopango IPS is the created artifact that aims to predict the location coordinates to improve the accuracy of Koopango IPS. The design process includes several stages such as data collection, preprocessing, model architecture selection, tuning of hyper-parameters and evaluation.

1.2.2.2 Problem Relevance

The accuracy of the indoor positioning system plays a critical role in many applications such as the ability to find objects in airports, and malls, among others. A more precise IPS also enhances the user experience and application of location-based services. Therefore, the use of deep learning techniques to improve the accuracy of IPS advances the existing body of knowledge and represents a crucial step in the direction of the final implementation.

Furthermore, predicting the location coordinates using different data-set collected in distinct environments highlights the capability of the deep learning model and opens the door to future research in the knowledge base.

1.2.2.3 Design Evaluation

Hevner et al. [22] suggests the utility, quality and efficacy of a design artifact must be thoroughly demonstrated using well-executed evaluation methods. The Table 1.2 presents DSR evaluation methods from which controlled experiment method is used to evaluate this thesis. The neural network model trained on two separate datasets from different environments is compared in terms of generalisation. Using both datasets for quantitative assessment requires determining the evaluation metrics to use, while maintaining consistent experimental conditions. A meaningful conclusion can be drawn from an evaluation that provides insights into the model's adaptability to dataset variability.

1.2.2.4 Research Contribution

This thesis contributes to the knowledge base in the field of location based services by demonstrating the ability of deep learning methods in enhancing the accuracy and reliability of ultrasound based IPS.

1.2.2.5 Research Rigor

This thesis adheres to the rigorous methodology of DSR by systematically collecting data from diverse environments with an adequate number of samples. The data is then

1. Observational	Case Study: Study artifact in depth in business environment
	Field Study: Monitor use of artifact in multiple projects
2. Analytical	Static Analysis: Examine structure of artifact for static qualities (e.g., complexity)
	Architecture Analysis: Study fit of artifact into technical IS architecture
	Optimization: Demonstrate inherent optimal properties of artifact or provide optimality bounds on artifact behavior
	Dynamic Analysis: Study artifact in use for dynamic qualities (e.g., performance)
3. Experimental	Controlled Experiment: Study artifact in controlled environment for qualities (e.g., usability)
	Simulation – Execute artifact with artificial data
4. Testing	Functional (Black Box) Testing: Execute artifact interfaces to discover failures and identify defects
	Structural (White Box) Testing: Perform coverage testing of some metric (e.g., execution paths) in the artifact implementation
5. Descriptive	Informed Argument: Use information from the knowledge base (e.g., relevant research) to build a convincing argument for the artifact's utility
	Scenarios: Construct detailed scenarios around the artifact to demonstrate its utility

Table 1.2: Design Science Research - Evaluation Methods (Source: [22])

preprocessed through normalization and STFT, followed by training of a deep learning model using appropriate techniques. The evaluation process employs proper statistical methods to ensure reliable results.

1.2.2.6 Design as a Search Process

The deep learning model specifically tailored for ultrasound-based IPS is created which is itself an iterative and dynamic process. A variety of model architectures and hyperparameters are configured to search for the optimal solution based on evaluation metrics. Feedback from the evaluation methods mentioned in Section 1.2.2.3 resulted in an enhanced deep learning model for Koopango IPS.

1.2.2.7 Communication of Research

The results of this thesis are published in an academic environment as well as to interested or involved business parties.

1.2.3 Research Questions

On the basis of previous sections, the main research question (RQ) is formulated as follows: **How to improve the accuracy of a fingerprinting indoor positioning system to find location coordinates from speakers emitting same ultrasound signal using deep learning techniques?**

In order to answer this question in a more organised way, the main research question is divided into two subquestions:

- **RQ-1: How to find the accuracy of the deep learning model of an ultrasound based fingerprinting IPS with audio signal and location coordinates data?**
- **RQ-2: How to generalize the proposed deep learning model for Koopango IPS with the fingerprinting data recorded in a different environment from the original data?**

RQ-1 is answered to designing and hypertuning the deep learning model. RQ-2 is answered to training the same model on different set of data for generlisation. Best of the result of RQ-1 and RQ-2, a well-tuned model is designed and the main RQ is answered.

1.3 Thesis Structure

The thesis is put forward as follows: In Chapter 2, we discuss the presuppositions about Koopango IPS as it currently stands, how fingerprinting is used with ultrasound and how data is collected. Chapter 3, we design a deep learning neural network as an artifact tailored made for Koopango ultrasonic IPS. Chapter 4, we generalise the model behaviour on unfamiliar data set. In Chapter 5, we evaluate the model and Chapter 6 we conclude the research and along with future work discussion.

Chapter 2

Presuppositions

The following chapter provides a brief introduction and explanation of the fundamental methods used in this thesis. Section 2.1 gives a brief about ultrasound-based indoor positioning system, Section 2.2 introduces Koopango IPS as it currently stands and how data is collected, and Section 2.3 describes the formulation of single channel fingerprinting IPS and how machine learning could be used to improve the current system.

2.1 Ultrasound Based Indoor Positioning

As discussed in Section 1.1.2, Ultrasound is one among several techniques pursued for indoor localisation renowned for having very high precision in initial setups. Most Ultrasound based positioning system employs a time of flight information to compute the position of a device [7]. A Time of Flight based ultrasound indoor positioning requires multiple speakers to allow us to clearly calculate the arrival time of signals from each speaker. The Triangulation technique can narrow down the position.

When the different Time of Arrivals (TOA) from numerous signals had been resolved, strategies like the least squares can be applied to compute the triangulation position. This method is understood to convey centimetre degree accuracy.

2.2 Koopango Indoor Positioning System

The Koopango indoor positioning system (IPS) is designed to use a fingerprinting technique based on ultrasound signals instead of triangulation, where each speaker in a location has the potential to play a unique ultrasound signal to create unique environments. With Wi-Fi and Bluetooth indoor positioning, fingerprinting is commonly used. In non-prototype environments, triangulation-based ultrasonic indoor positioning presents many challenges due to the requirement for clear line of sight and the potential for Doppler effects, multi-path propagation and interference. These reasons lead Koopango to explore ultrasound based indoor positioning based on fingerprinting.

The following Section 2.2.1 describes multichannel Koopango IPS, Section 2.2.2 gives an overview of Koopango Signals, Section 2.2.3 explains how data is recorded and equipment used, and finally Section 2.2.4 gives an introduction of Fingerprinting Koopango IPS.

2.2.1 Multichannel System overview

The Koopango indoor positioning system uses indoor positioning speakers that play unique signals arranged naturally around a building. To determine the general characteristics of the region, the area will be surveyed. All the speakers have the freedom to play their own custom ultrasonic signal in a multichannel system.

2.2.2 Koopango Signal Generation

An essential aspect of the Koopango indoor positioning system relies on the signals played by the speakers. These signals must exclusively outline the vicinity enclosing the speakers. To accomplish that objective the signals necessitate holding really low cross correlation among one another.

The small audio clips of about 0.18 second continuously repeat from each speaker uniquely assigned to them. These audio signals are made using the frequency hopping spread spectrum method [9], where time is split into multiple sections and different frequencies using sinusoidal signal are used to create the signal in each one. The audio is identified by their frequency pattern over time rather than a single frequency.

2.2.3 How data is collected?

Smartphones are used to record signals because smartphone users are the primary target audience. A WiFi access point makes data reception easier. The Raspberry Pi includes a Python server that starts the ground truth system and receives data from the smartphone via WiFi. All collected data is saved in a network storage device. A combination of a real sense tracking camera and lidar is used to determine our exact location. This location is in relation to a starting point. The starting position is set to the origin unless an initial position is specified. The ROS sensor fusion system is used to generate position (Robot Operating System).

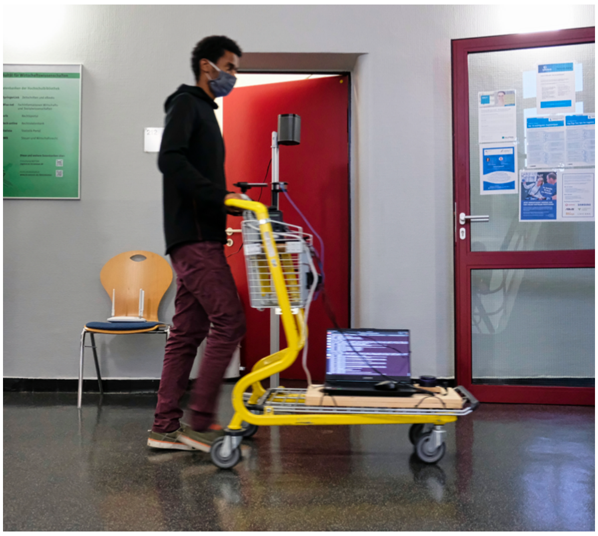

FIGURE 2.1: Experimental Setup of Koopango IPS (Source: [23])

Raw audio data in PCM16-bit format is received from the smartphone. We save the data we receive from the microphone directly without any preprocessing. Sonos speakers are mostly used to play ultrasonic music. In some of the datasets, ultrasound was played from speakers that were already built into the building. The same sound is being played by all of the speakers. The configuration of the speakers was not tracked though. mostly place the speakers based on our best judgment and experience. The majority are at least 7 meters apart. Information about the location of speakers was not important in our fingerprinting-based system because the system was independent of that information.

2.2.4 Fingerprinting IPS

The existing Koopango process employs a fingerprinting technique for localization. It depends on creating a database of mapping signal features to their exact location. This database can subsequently be utilised in a pattern-matching technique to pinpoint the precise location of a signal [13]. This two-step approach usually referred to offline and online periods of fingerprinting.

Multiple algorithms are often employed in fingerprint matching, such as Naive Bayes, k-nearest neighbours, neural networks and so on. The Koopango-based system incorporates a k-nearest neighbour pattern matching with Gaussian-based smoothing to avoid abrupt jumps.

2.3 Single Channel Problem Formulation

In many common scenarios where ultrasound localization is needed, like large shopping centres, airports, or museums, there typically isn't the adaptability to attach each audio speaker in the building to its own line. All of the speakers share one or two lines. Consequently, in these circumstances, the multichannel system would be inadequate. The following Section 2.3.1 describes the problem statement of Single Channel IPS, and Section 2.3.2 provides possible approaches for Koopango IPS.

2.3.1 Single Channel Problem

In a single-channel system, all the speakers are playing the same ultrasonic signal. However, previously in the multichannel system, each speaker had the possibility to play a unique signal, in turn creating a unique region around it. Single channel is challenging for the reason that all the regions are almost identical in their sound nature.

We are currently researching the feasibility of fingerprinting, which is based on the assumption that each x,y location is unique in terms of fingerprint data collected. However, having indistinguishable locations adds to the difficulty of repeating speaker signals. What information or techniques can we employ to maintain the requirement that each location is uniquely identified?

2.3.2 Koopango IPS Approach

Here are some methods to achieve good positioning accuracy when using a single-channel fingerprinting system.

- Movement detection towards and away from a speaker: A smartphone records audio in the same block size as the speaker signal. The position of the smartphone affects the cross-correlation maximum. Sound speed and cross-correlation shift are used in an equation for distance calculation.
- Multiple local maxima in cross correlation: Multi-local maxima provide additional information (Doppler, reflections, context) from cross-correlation with Koopango signals.
- Regional bounds: Several speaker sources in some areas. Signal sequences between specific speaker arrangements. User's travel history helps determine landing region.

- Cross correlation with shorter signal: Smartphone recordings differ from intended tracks because of distance and environment. High-frequency signals degrade over distance. Short signal cross-correlation detects spectrum changes.

- Machine Learning based Fingerprinting: As discussed in the Section 1.1.3 use of deep learning with Fingerprinting system improves the accuracy of the system [3]. Moreover, machine learning algorithms posses the capability of capturing spatial information which is often contained in the single channel data. A machine learning model can adapt to changes in the strength of the signal occurring because of the environmental factors, and non-linearity between the signal features and location coordinates are handled by machine learning algorithms with ease.

Chapter 3

Deep Learning Model for Ultrasound Based Localisation

The following chapter outlines a neural network that was constructed to find the accuracy of the ultrasound based indoor positioning system in localisation. In order to determine the input layer, data analysis was carried out, for the neural network which provides the location coordinates as output. Furthermore, the evaluation was performed by comparing the output of the neural network with the initial location coordinates.

3.1 Introduction

The goal of Chapter 3 is to answer the research question RQ-1: How to find the accuracy of the deep learning model of an ultrasound based IPS with audio signal and location coordinates data? - as mentioned in Chapter 1. In order to answer RQ-1 in a more organised and well-rounded manner, it is divided into two sub questions:

- **RQ-1.1: What deep learning model to use for predicting coordinates accuracy for an ultrasound-based indoor positioning system?**
- **RQ-1.2: What is the role of high pass filter in improving the accuracy Deep Learning model for Koopango IPS?**
- **RQ-1.3: What different hyper-parameters are required to optimize the accuracy of the model?**

Each sub question is answered independently in different sections. The first part of this chapter focuses on the introduction of experimental data and data analysis to identify the input layer as described in Section 3.2. Afterwards, Section 3.3 answers RQ- 1.1 and focuses on the architecture of the proposed neural network. Subsequently, Section 3.4 answers with RQ-1.2 and RQ-1.3 and describes impact of different hyper-parameters and high-pass filter on the neural network. Section 3.5 discuss the outcome of the final deep learning model. Finally, Section 3.6 gives the conclusion in the designing the deep learning model to improve accuracy in localisation of Koopango IPS.

3.2 Exploratory Data Analysis

In designing a deep learning model, exploratory data analysis plays an important role in understanding the artefacts and characteristics of the data. This thesis focuses on ultrasound audio data and their corresponding coordinates for localisation. Data analysis helps in determining the input layer for the neural network by providing reliable insights. The following sections introduce the IPS data available, followed by visualisation, feature selection, data preprocessing, and interpretability. These steps lead to the selection of the deep learning model.

3.2.1 Datasets Overview

In this thesis, we have used two data sets, both recorded using the same Koopango signal but at different locations. The first data set was recorded at the company's office and is labelled FpMap, while the second data set was recorded at another location and is labelled FpMap0. FpMap contains 10412 rows of fingerprints, while FpMap0 contains 6787 rows of fingerprints. Each fingerprint is a combination of audio and its corresponding position, and represents 180 milliseconds of audio. Each row also has a respective (x,y) location coordinate. It's important to note that certain fingerprints or information are only valid within their respective setup. Since the location and speaker arrangement are different in between the two setups, the fingerprints are also different.

Typically, for every one square meter area there could be more than five rows of a fingerprint. Every row contains 8192 values representing audio in a 44100Hz sampling, with 2 values indicating position (x,y). Every data set contains 8196 columns and different experiments comprised of a different number of Fingerprints rows, in our case as mentioned above Fpmap contains 10412 fingerprints and Fpmap0 contains 6787 fingerprints.

While exploring the data, it was found that the first column in each row duplicated the value from the second row, while the last column contained z-coordinate data that wasn't currently relevant for the thesis. As a result, these columns were removed, leaving

8194 columns representing first 8192 columns as audio values and last two columns as (x,y) coordinates. Figures 3.1 and 3.2 display the trajectory of (x,y) coordinates in FpMap and FpMap0 experiments, respectively. Basically, the trajectory tells us how the smartphone has moved from the original location while performing the experiment.

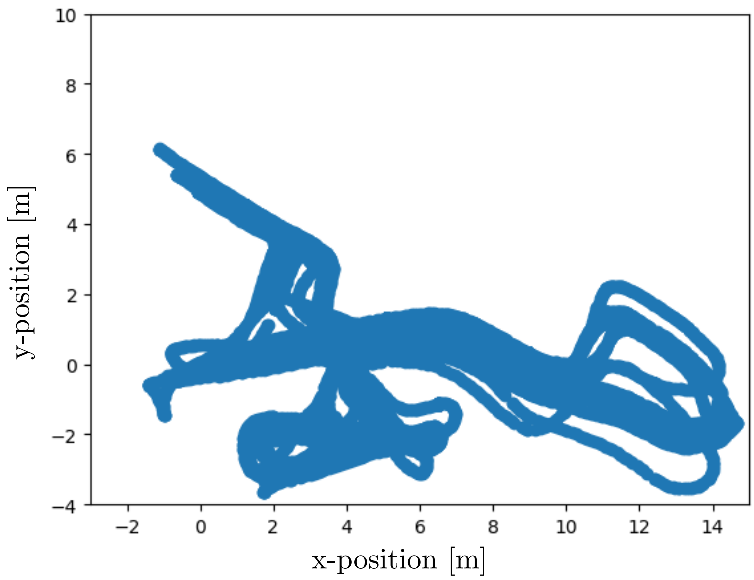

FIGURE 3.1: (x,y) coordinates trajectory of FpMap

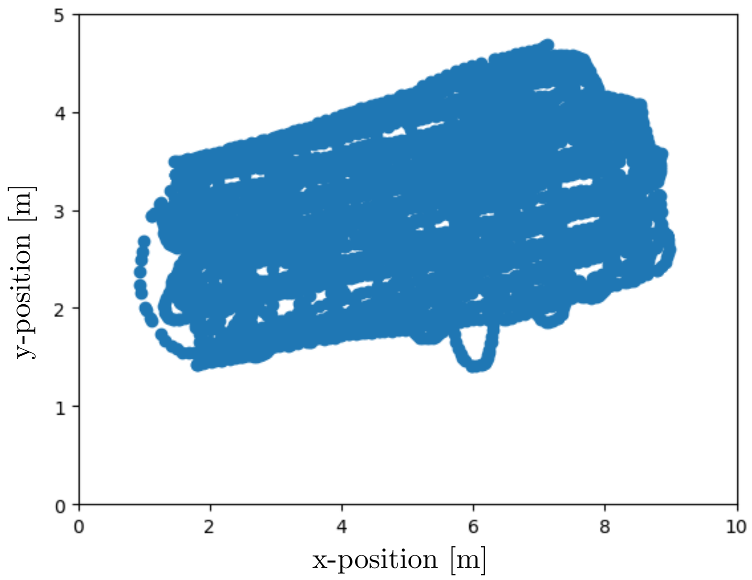

FIGURE 3.2: (x,y) coordinates trajectory of FpMap0

3.2.2 Audio Data in Time and Frequency Domain

For every 100th of the row, the audio data is analyzed separately in the time and frequency domain to identify certain artefacts. Figures 3.3 and A.1 show that data in

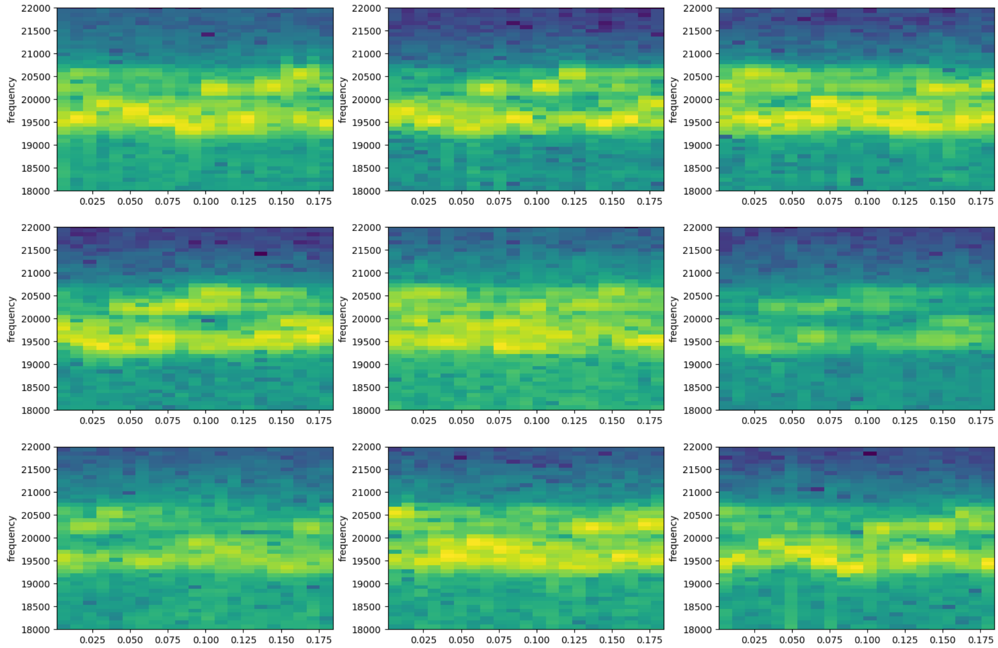

FIGURE 3.3: Audio Data in frequency domain for FpMap setup

the frequency domain of recorded audio lies in between the range of 19000 Hz and 21000 Hz, providing a first impression of the dominant frequency of the recorded Koopango Signal. However, the assumption that there could be a problem while recording the data is analysed by conducting a dominant frequency check which ensures that if recorded audio matches the signal played.

3.2.3 Dominant Frequency Check

Dominant frequency can be checked if the dominant original koopango signal has a fairly high cross-correlation value than the mean cross-correlation value of the recorded data. In Figure 3.4, we can see a normal cross-correlation plot with the first 5 Koopango signals with FpmMap0. The koopango signal played is higher than the others, but generally depending on how far or close you are from a speaker it goes up and down. It can also be interpreted on a factor of how strongly the speakers are playing sound, sometimes they are more than 7m away from the smartphone and sometimes they are placed on the ceiling and sometimes they are hanging on the wall.

3.2.4 Short-Time Fourier Transform

Short-Time Fourier Transform (STFT) gives the analysis of audio in the time-frequency domain. It provides representations that include the signal's local time and frequency content. In Figure 3.5 shows that an audio can be found in the area having a lot of energy

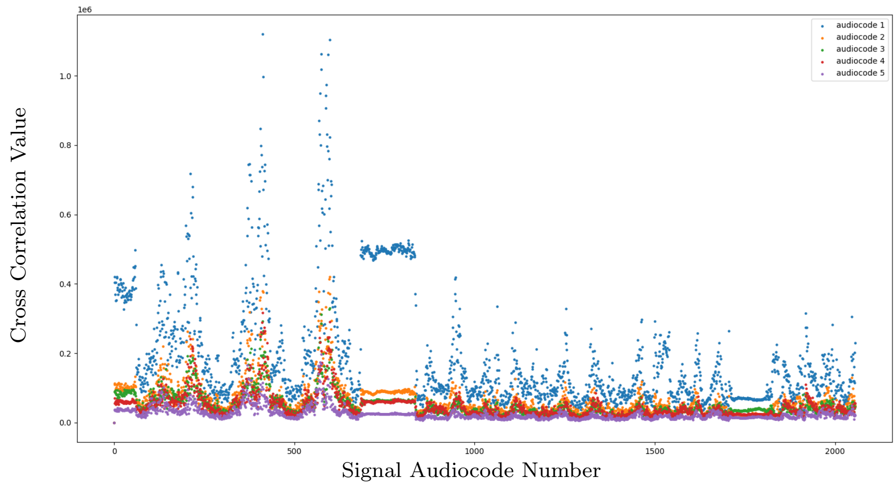

FIGURE 3.4: Dominant Frequency obtained by correlation between received signal and original audio

with high frequency in the spectrogram of STFT, where magnitude of spectrogram could be used as the input to train the CNN.

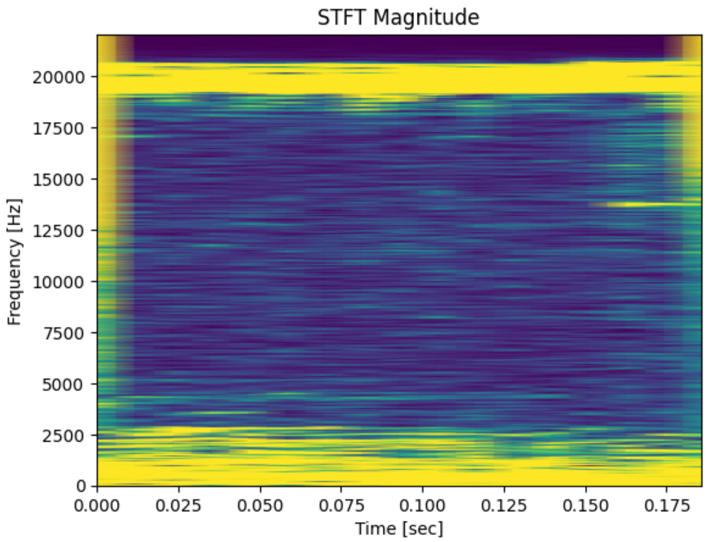

FIGURE 3.5: Short-time Fourier Transform of an audio signal of 180ms

3.2.5 Nomalization and High Pass Filter

Data Normalization is being as when the inputs are normally distributed, with a mean closer to zero and a bounded variance, neural networks tend to converge more quickly and steadily which results in the output already rather close to the desired one with the initial weights [15]. To allow a neural network to output high-dimensional objects, the minimum and maximum possible magnitude values in the spectrograms were calculated. Using an output activation function in the last layer that constrains the network's output range to the real data range can greatly aid training and also prevent invalid network predictions [24].

Applying the noise reduction in the data significant changes were not observed in the STFT spectrogram. A highpass filter applied helps in image Sharpening in the frequency domain. In Figure 3.6, less distorted boundaries can be observed and smoother high frequency is visible which can be used for the input layer.

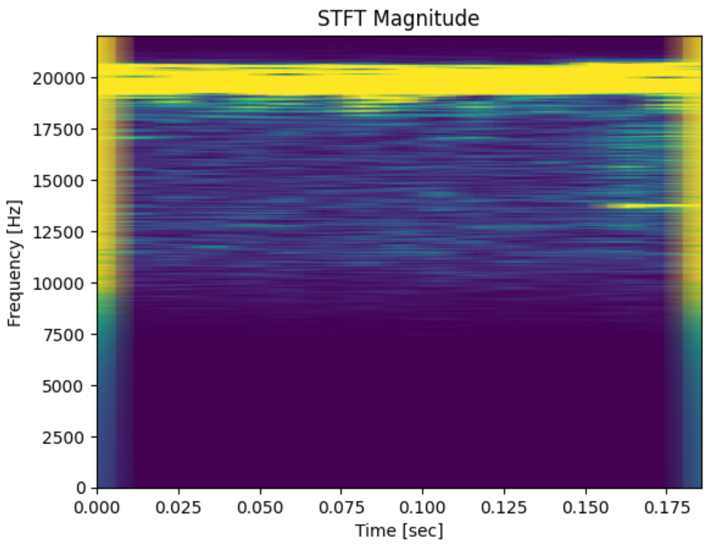

FIGURE 3.6: Short-time Fourier Transform of an audio signal after applying the high pass filter

3.3 Proposed Deep Learning Model

Deep Learning methods presents a comprehensive solution that handles feature extraction automatically, reducing the computational complexity. This is especially useful when dealing with complex data of Fingerprinting, and features obtained results in better performance of localisation [20]. In the following section, The deep learning neural network architecture is designed as an artifact to predict (x,y) coordinates to answer the question RQ-1.1 - What deep learning model to use for predicting coordinates accuracy for an ultrasound-based indoor positioning system? The different layers along with their configurations and the reasons to use them is discussed. Furthermore, the regularization methods, including dropout or L2 regularization, which help prevent over fitting and improve the model's durability, have been thoroughly examined.

3.3.1 Convolutional Neural Network

Convolutional Neural Networks (CNN) are comprised of multiple layer unique architecture which is broadly used in the field of deep learning for solving complex computational problems. The architecture has one input layer, multiple hidden layers and a single output layer where a particular neuron produces an output Y by executing some function on input X as shown in Equation 3.1 where W denotes the weight of strength of interconnection between neurons of two adjacent layers [25] .

$$F(X, W) = Y[25] \tag{3.1}$$

According to Indolia et al [25], CNN architectural components consist of a convolutional layer, pooling layers, a fully connected layer and an activation function. Convolutional layers filter the input data by performing convolutional methods to extract features and patterns, pooling layer reduces the spatial dimensions obtained from the convolutional layer which helps to decrease the computational complexity and provide translational invariance. Finally, a fully connected layer performs the classification or regression tasks using an activation function based on the extracted features.

3.3.1.1 Why CNN is used for the model?

In this thesis, CNN is used for a regression task as we are predicting the (x,y) coordinates based on the input STFT spectrogram of the recorded ultrasound audio signal. The CNN for regression basically consists of the same architecture as for classification. However, For regression, the output layer consists of a set of neurons to obtain continuous value predictions instead of predicting a class or classes. Since the time-series data has been used for the input, a 2-dimensional CNN has been employed for the model architecture as it leverages the hierarchical features extraction ability of a 2D CNN by filtering available time series samples followed by the flattening of the obtained features [26].

3.3.2 Model Architecture

The proposed neural network architecture used in this thesis incorporates a composition of convolutional layers, max-pooling layers and fully connected layers to perform regression on the input audio data to predict the location coordinates as output. Figure 3.7 depicts the architecture of the proposed deep learning model. The model initialises an input layer from the normalised training batch by incorporating the shape of input data, which is then passed through a sequence of convolutional layers by using 2D CNN

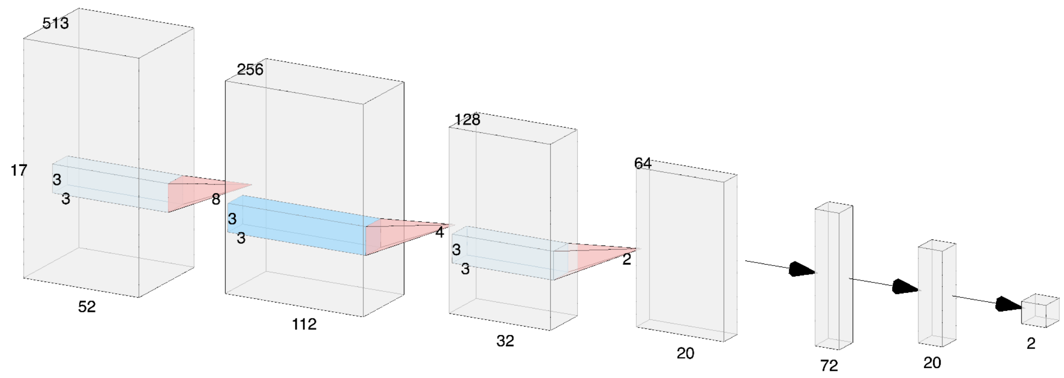

FIGURE 3.7: Proposed Deep Learning Architecture for Koopango IPS

operations with the kernel size of 3 x 3 and ReLu as an activation function to recognise spatial patterns.

The kernel size plays a crucial role to extract translationally invariant local features which is important in the case of time-series samples. Simonyan et al [27] research on the use of a smaller Kernel size of 3x3 in each convolutional layer ensures good performance for accuracy in terms of the depth of CNN architecture. Non-linear activation functions directly contribute to approximating the function capability of any neural network including non-convex functions. However, there are typically three main activation functions popularly used for artificial neural networks: Sigmoid, Hyperbolic Tangent and Rectified Linear Unit (ReLu). The ReLu has been used as an activation function for the research as it has advantages like non-linearity, sparsity and computational efficiency for improved learning performance over the other activation functions [28]. The ReLU function is mathematically defined as in Equation 3.2, where the function returns a positive input value or zero if the input value is non-positive.

$$F(x) = max(0, x) \tag{3.2}$$

The summarised algorithm for the proposed deep learning model is as follows:

Input: X_train (Training data obtained from STFT spectrogram)

- *Step 1: Create an input layer with the shape of training data.*
- *Step 2: Add Convolutional layers (Conv2D) with ReLu as an activation function, Kernel size of 3 X 3 and specified number of neurons.*
- *Step 3: Add 2D max pooling layer after each convolutional layer with a pool size of 2 X 2*

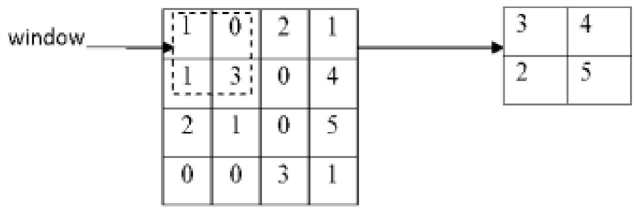

FIGURE 3.8: Pooling operation on a 2 X 2 window (Source: [25])

- *Step 4: Flatten the output of the last convolutional layer.*
- *Step 5: Add Dense layers with ReLu activation function, a specified number of neurons and kernel regularizers.*
- *Step 6: Add dropout after each dense layer.*
- *Step 7: Create an output layer with two neurons and a linear activation function to obtain coordinates as output.*
- *Step 8: Initialise and call Steps 1 to 7 for the training process.*

The model initializes by creating an input layer with the shape of training data followed by the addition of multiple convolutional layers along with the specific filters, kernel size and activation function. After implementing multiple architectures and hyper tuning the model which is discussed in Section 3.4, we proposed four two-dimensional convolutional layers (Conv2D) with the filter size of 52, 112, 32 and 20 from convolutional layer 1 to convolutional layer 4 respectively as shown in the Figure 3.7. Each convolutional layer is followed by a max pooling layer of a pool size of 2 X 2. The max pooling technique reduces the number of trainable parameters and introduces translation invariance by considerably reducing the map size [25] as shown in Figure 3.8.

In the next step, the flattening layer is called which takes the multi-dimensional output of the last convolutional layer to convert it into one-dimensional vector while keeping the desired features. The flattening layer also reduces the number of parameters and facilitates the translation from convolutional layers to fully connected layers by acting as a bridge between them. The Dense layer utilises the output of the flattened layer, in this model the first dense layer is created using 80 neurons along with ReLu activation function and the second dense layer uses 32 neurons. Every neuron in a dense layer is connected to each neuron in the flattened layer, hence also known as a fully connected layer, and the neuron computes a weighted sum of inputs to produce an output utilising the activation function [29]. It allows the dense layer to learn complex relationships and patterns within the extracted features. Both the dense layer in the model is followed by a drop-out layer to prevent overfitting and promotes generalisation by randomly deactivating a certain number of input neurons.

Finally, the last dense layer which is also an output layer contains two neurons to predict the (x,y) coordinates of Koopango IPS while utilising a linear activation function to perform the regression task.

The architecture of the proposed deep learning model for Koopango IPS is designed to employ CNN capabilities of extracting important features through the shared weights from the input audio data. The combination of a two-dimensional convolutional layer, max-pooling layer and fully connected layers effectively detect the spatial relationships among the input data, allowing it to learn the possible interactions and dependencies within the variables [29].

3.4 Training the Model

The training process of any artificial neural network includes iteratively fine-tuning the parameters to minimise the predefined loss function while finding the parameters that maximise the performance on the task. In the following sections, we answer the RQ-1.3 - What different hyper-parameters are required to optimize the accuracy of the model? We explain how the model is trained including the details of the employed loss function, the chosen optimiser, implementation of regularisation and managing the learning rate. The strategies for splitting the data set are also discussed.

3.4.1 Splitting the data

As mentioned in Section 3.2 we have used two data sets FpMap and FpMap0. For the training for the model, we use the FpMap dataset after transforming the data into an STFT spectrogram. We normalise the data and apply the high-pass filter on the data. To evaluate the influence of the high-pass filter on the performance of the deep learning model, we train the model two times by separately initialising the model with and without a high-pass filter. Furthermore, we split the data into three parts, the training data, the testing data and the validation data.

First, we split the normalized FpMap dataset having data shape of (10412,513,17,1) into the training data (X_train, y_train) and the testing data (X_test, y_train) with test size of 0.2 and random state of 42 indicating 80 percent of data as training data and 20 percent of data as testing data. Secondly, we split the remaining training data again into the final training data and the validation data using the same test size of 0.2 and random state of 42 indicating 80 percent of the final training and 20 percent of the validation data. The final data set after the split is shown in the Table 3.1. For example if shape of data set is (6663, 513, 17, 1), here 6663 are the number of sample, 513 are the number of frequency bins, 17 shows the number of time segments and 1 represents the number of channel which is single channel in our case.

The same procedure goes for high-pass filter data split and labelled as FpMap_high-pass. X_train denotes the training audio data with normalisation and Y_train denotes the training coordinate without normalisation data which is required for supervised learning to predict the output as coordinates.

Training Data	**Testing Data**	**Validation Data**
X_train	X_test	X _val
(6663,513,17,1)	(2083, 513, 17, 1)	(1666,513,17,1)
y_train	y_test	y_val
(6663,2)	(2083,2)	(1666,2)

Table 3.1: FpMap datasplit into Training, Testing and Validation Data

3.4.2 Loss Function

This thesis utilises the mean square error (MSE) as the loss function for training the neural network model as it quantifies the difference between the actual location coordinates values and the predicted coordinates output of the model. The MSE is suitable for regression tasks where the main goal is to predict the continuous value in our case the value of location coordinates by calculating the average square of the difference between the predicted and the actual values, thus yielding the measure of the capacity of model in capturing the important features in the data.

3.4.3 Optimizer

Optimizer plays an important role in improving the yield of a deep learning model. In this thesis, the Adam optimizer is selected for training and optimization of the model as it adapts the learning rate for each parameter independently and converges it quickly. Kingma et al [30] suggest that the Adam optimizer outperforms other optimization algorithms over the rate of convergence, dealing with sparse gradients and non-stationary objectives, yielding better final model performance especially when working with CNNs.

3.4.4 Hyper-Parameter Tuning

Hyper-parameter tuning enhances the performance of the model by finding the optimal values of the hyper-parameters that influence the learning process and generalisation of the deep learning model. In the following sections, we discuss the hyper-parameters that were influencing the deep learning model for Koopango IPS such as the number of neurons/filters, regularizers, learning rate, number of epochs, batch size and dropout, and the tuner employed.

3.4.4.1 Keras Tuner

For tuning the hyper-parameter, we utilise Keras Tuner as it is a hyper-parameter optimisation framework to automate the tuning process in deep learning models by employing search algorithms intelligently while maximising the performance of the model. Though we have run multiple trials using Keras Tuner to find the optimal values Table 3.2 shows the results of four trials. Considering the MSE as a loss function from other trials, we chose Trial 3 for further tuning the model manually. Moreover, The total number of trainable parameters in the model for Trial 3 came to be 138,770 however, for other trials the total number of trainable parameters was in the range of 200,000 and 600,0000. The lower number of trainable parameters tends to have less tendency to overfit the model which also makes the choice of Trial 3 better.

3.4.4.2 Number of Neurons/Filters

Each convolutional layer consists of certain number of neurons/filters to determine how well it extracts different features from the input data. Similarly, every dense layer contains neurons/units to determine dimensionality in extracting higher-level features. From Trial 3, we select conv1: 52, conv2: 112, conv3: 32, conv4: 20, dense1: 72, dense2: 20 , and dense3: 2 for the final model architecture as mentioned in Figure 3.7.

3.4.4.3 Learning Rate

During the optimization process learning rate determines the step size at iteration to stabilise the convergence for an optimal solution while managing the speed of training. From Keras tuner, the learning rate came to be 0.0003 however we trained the model manually on learning rates ranging from 0.0001 to 0.01, on increasing the learning above 0.001 the model shows overfitting. Finally, we chose the learning rate to be 0.0001 as it had better performance in the evaluation metrics MSE over the one from Keras tuner.

Hyper-parameters	**Trial 1**	**Trial 2**	**Trial 3**	**Trial 4**
Number of Filters	conv1: 52, conv2: 112, conv3: 44, conv4: 32, dense1: 80, dense2: 32	conv1: 64, conv2: 32, conv3: 48, conv4: 64, dense1: 128, dense2: 32	**conv1: 52, conv2: 112, conv3: 32, conv4: 20, dense1: 72, dense2: 20**	conv1: 56, conv2: 112, conv3: 48, conv4: 32, dense1: 80, dense2: 32
Learning Rate	0.00097	0.00051	**0.0003**	0.0001
Dropout Rate 1 and 2	0.5,0.3	0.3, 0.5	**0.1,0.1**	(0.1,0.2)
L1,L2 Regularization for Dense 1	0.0018, 0.00033	0.0011, 0.0047	**0.000013, 0.000091**	0.0011, 0.000002
L1,L2 Regularization for Dense 2	0.000051, 0.0046	0.0000015, 0.000028	**0.000029, 0.0000025**	0.0011, 0.0000065
MSE loss	3.27	3.3	**2.91**	3.2

Table 3.2: Hypertuning Trials by Keras Tuner

3.4.4.4 Regularization

Overfitting or Underfitting is one of the major concerns in deep learning models which can be handled by Regularization techniques such as L1 regularisation, L2 regularization, and dropout. L1 or Lasso regularization adds a penalty equivalent to the sum of the squares of the magnitude of coefficients whereas L2 or Ridge regularization adds the penalty equivalent to the sum of the absolute values of coefficients [31]. We use a combination of L1 and L2 in the both the Dense Layer of the model. After manually improving on the (L1,L2) values, we finalise (0.000013, 0.0003) for Dense layer 1 and (0.000018, 0.0000035) for Dense layer 2 after getting the best possible results for the model.

Dropout is another regularization technique, it controls the number of neurons to be dropped from the dense layer. It randomly turns off a certain number of neurons during the training process as per the given dropout rate to make the model more robust and less dependent on any particular neuron. For the final selected model, Dense Layer 1 and Dense Layer 2 both has a dropout rate of 0.1, which means it is deactivating 1 per cent of the neurons on both dense layers to prevent overfitting.

3.4.4.5 Batch Size

The batch size determines the size of samples used in the training process as it influences the convergence speed, the generalization, and overall performance of the model. Though we have already defined the batch size in Section 3.4.1 and Table 3.1, however, we also trained the model on splitting the data on the test size of 0.3 which means 70 per cent of training data and 30 per cent of testing data, and same for validation data. But first, it was a trade-off between better accuracy or more generalized performance. Secondly, the model was overfitting. Hence, we chose the data split as mentioned in Table 3.1

3.4.4.6 Number of Epochs

During the training process, number of epochs determines the number of times the entire training data is processed through the deep learning model to affect the convergence and prevent overfitting or underfitting. Generally, 10 per cent of the training data samples is chosen as the number of epochs. In our case 6663 training data sample makes up for 60 epochs, however, to prevent overfitting we reduced the number of epochs to 50.

3.4.5 Training with High-Pass Filter

The following section answers the research question R.Q-1.3- What is the role of a high pass filter in improving the accuracy of a Deep Learning model for Koopango IPS? We train the model separately with high-pass enabled data as mentioned in Section 3.4.1. The frequency cut-off has been set to 10,000 as mentioned in Section 3.2.5.

Figure A.2 shows the difference between training loss and validation loss of MSE on the training of the input data with the high pass filter enabled, it clearly shows the model tends to overfit the when high pass filter is enabled. The reason for this happened was that on the cutting off of the frequency, the spectrogram lose important features and results were not desirable as expected. The high-pass filters predict coordinates at the mean squared error of 4.5 meters which is not good enough for the expected model as compared with the training of data without a high-pass filter. Hence, we further dropped the hypothesis of using a high pass filter in the thesis due to time limitation. However, this could be considered for future research considering the other factors such as reflection in the room.

3.5 Results

This section presents the result final deep learning model architecture tailor-made for Koopango IPS. In Table 3.3, we show the finalised model to predict the location coordinates from the ultra-sound audio data recorded by Koopango Fingerprinting IPS.

After training the model at 50 epochs and a learning rate of 0.005, the final MSE loss resulted to be 1.9 meter square which is acceptable as the nature of data tends to overfit on increasing the learning rate and regularization. Upon considering multiple architectures and hyper-parameters, the model predicts the location coordinates at the mean square error of 2.633 meters square on FpMap0 test data which is later discussed in the evaluation section. Figure 5.1 discussed in Chapter 5 shows the improvement in performance of the model in loss value with respect to validation loss generalizing the training data with no signs of under-fitting or over-fitting. The model predicts location coordinates of Koopango IPS with the accuracy of approximately 1.6 meter in a diameter.

Hyper-parameters	**Value**
Number of Convolution Layers	4
Number of Neurons/Filters	[52,112,32,20]
Filter Size	(3 X 3)
Max-Pooling Size	(2 X 2)
Number of Dense Layers	3
Neurons of each Dense Layer	[72,20,2]
L1,L2 Regularization for Dense Layer 1	(0.000013, 0.0003)
L1,L2 Regularization for Dense Layer 2	(0.000018, 0.0000035)
Dropout rate for Dense Layer 1 and 2	(0.15, 0.1)
Learning Rate	0.0005
Total Params	138,770
Batch Size	6663
Number of Epochs	50

TABLE 3.3: Proposed Deep Learning Model after Hyper-Parameter tuning

3.6 Conclusion

We design the deep learning neural network as an artifact for Koopango Indoor positioning system on a single channel which predicts the location coordinates (x,y) from the audio data recorded in one environment. In addition, we contribute to the knowledge base on the influence of using high-pass filter on audio data. Furthermore, we discuss different hyper-parameters, their impact on the model and the final version of the CNN model after hyper-parameter tuning.

The two-dimensional convolution neural network is used to perform the regression task in predicting (x,y) location coordinates from STFT spectrogram an input of the neural network. The hypothesis of considering a high-pass filter in the audio is neglected on the basis of the loss of important features during the training process and the tendency of overfitting. We discuss the important hyper-parameters influencing the model and tune them accordingly on improving the final accuracy. We present the final model with a tuned hyper-parameter in Table 3.3 which can predict (x,y) coordinates with accuracy with an accuracy of approximately 1.6 meters in circular diameter.

For future work, certain factors like reflection, speaker volume and reverberation could be taken into the fact for location prediction which are not considered in this thesis because of time limitations. Finally, an automated ML pipeline can be created for future research to test the model in a practical environment.

Chapter 4

Model behaviour in different environments

In the following chapter the behaviour of the neural network designed for Koopango IPS is determined with the fingerprint dataset recorded in a different environment from the original dataset. The goal was to the check the generalization of the designed deep learning model and determine the difference between the accuracy of the model trained with an unfamiliar dataset.

4.1 Introduction

The intention of Chapter 4 is to answer the research question RQ-2 - How to generalize the proposed deep learning model for Koopango IPS with the fingerprinting data recorded in a different environment from the original data? - as indicated in Chapter 1. In order to answer RQ-2 efficiently, it is divided into two sub questions:

- **RQ-2.1: What accuracy does the originally trained model outputs on predicting results using a different set of fingerprinting data?**
- **RQ-2.2: What is the behavior of the model when trained on a different fingerprint dataset?**
- **RQ-2.3: What limitations can be considered in order to find accuracy from an unfamiliar dataset?**

Every sub question is answered in different sections. The first part of the Chapter 4 focus on the generalisation of the proposed deep learning model to answer RQ-2.1 and

RQ-2.2 in Section 4.2 . Afterwards, Section 4.3 discuss the limitations influencing the model's performance with unfamiliar data answering RQ-2.3, followed by a conclusion in Section 4.4.

4.2 Generalisation of the Proposed Model

This section showcase the behaviour of proposed deep learning model in 3 for Koopango IPS using a fingerprint dataset recorded in a different environment location. This section answers RQ-2.1: What accuracy does the originally trained model outputs on predicting results using a different set of fingerprinting data? and RQ-2.2: What is the behavior of the model when trained on a different fingerprint dataset? First, we discuss the second set of fingerprint data available to generalization in Section

4.2.1 Dataset Overview

In Section 3.2.1, we discussed the data analysis of two datasets labelled as FpMap and FpMap0. We leverage the FpMap dataset for the development of the deep learning model in Chapter 3, however FpMap0 remained untouched after the data analysis which is now being used for generalization purpose. As mentioned earlier, FpMap0 is recorded in the premises of Wismar university with the setting of six speakers while FpMap was recorded in the DEJ Technology's office. FpMap0 dataset consist of 6787 rows of fingerprint data recorded with the same set of equipments mentioned in the Section 2.2.3.

Figure A.1 displays the FpMap0 audio data in frequency domain, while Figure 3.2 gives the trajectory of the recorded (x,y) location coordinates with respect to the audio data. To prepare the input layer, we convert the audio data into STFT spectrogram followed by normalization of the data. The hypothesis of using high-pass filter has been dropped for the reasons mentioned in 3.4.5.

After applying STFT to FpMap0, the resulting data shape is (6787,513,17, 1).We split the data utilizing the same criteria with test size of 0.2 as described in Section 3.4.1. Table 4.1 depicts the data split into training, testing and validation data, which will be used in the next section for training of the model.

Training Data	Testing Data	Validation Data
X_train	X_test	X_val
(4343, 513, 17, 1)	(1358, 513, 17, 1)	(1086 ,513 ,17 ,1)
y_train	y_test	y_val
(4343, 2)	(1358, 2)	(1086, 2)

TABLE 4.1: FpMap0 datasplit into Training, Testing and Validation Data

4.2.2 Coordinate Predictions on the Pre-trained Model using a Different Dataset

In Chapter 1, we trained our model on the FpMap dataset to predict location coordinates. We will be using the test dataset from FpMap0, recorded in a different environment, as shown in Table 4.1. This will help us assess the model's real-time performance with unfamiliar data.

Utilizing the STFT audio data from FpMap0, the location coordinates were predicted with an MSE of 25.03 square meters. Although the final accuracy was approximately 5 meters, this falls short of the ideal accuracy required for indoor positioning systems. Nevertheless, this prediction strongly suggests that predicting coordinates from unfamiliar data doesn't yield as much accuracy as the IPS system can provide. However, we would not investigate other factors influencing the results in this research due to time limitations. The next step is to train the model on unfamiliar data, then predict coordinates..

4.2.3 Model Training on Second Dataset

In Section 3.4, we discussed the training process of the proposed model and the hyper-parameters required to tune the model for the best performance possible. For generalization of the deep learning model using unfamiliar data, we proposed the training of the model with FpMap0 dataset as it exploits the knowledge gained from the original dataset in model development.

We train the model by leveraging the configuration used in Table 3.3 with MSE as loss function. In the early trials of training, the model showed tendency to underfit. Due to the time limitation of the thesis and keeping the idea of having one model architecture, we were able to tune only two hyper-parameters, learning rate and number of epochs.

Upon the decreasing the number epochs from 50, the convergence of the curve weakened and the model was underfitting. As a result, we kept the number of epochs at 50,

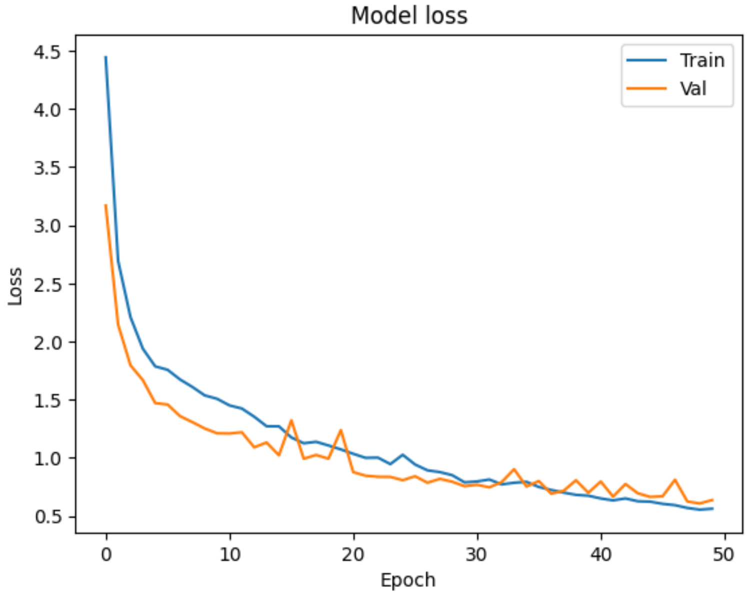

FIGURE 4.1: Training loss vs Validation Loss on FPMap0 Data

however, after increasing the learning rate from 0.0005 the curve showed convergence which leads us to final learning rate of 0.001. Finally, we achieved a MSE of 0.57 on training loss as compared to 0.60 MSE loss at validation. Figure 4.1 shows the loss value decreases gradually with the training loss versus validation loss curve, demonstrating better learning performance without overfitting or underfitting.

The model's performance is evaluated using the MSE as a performance metric to predict location coordinates which are discussed later in Chapter 5. The final MSE on predicting coordinates as compared with the actual coordinates came to be 0.57 meters square which means the accuracy of 0.75 meters is achieved by using an unfamiliar set of data. The result shows the model's capacity to generalize well in making coordinates predictions even on using the data from another location. However, there are certain other limitations that are supposed to be considered before making the final assumption on the model's generalization.

4.3 Limitations

In this section, we discuss the limitations that may have influence the model's performance with unfamiliar data thus answering the RQ-2.3: What limitations can be considered in order to find accuracy from an unfamiliar dataset?

The external factors like reflections from the objects, volume of the speakers, innate nature of speakers and number of speaker plays an important part in recording the audio signals. Since the model has been trained in a specific environment, the reflection and reverbration properties from unfamiliar environment introduces new patterns in the STFT which affects the predict accuracy. The model has been trained on specific volume level, and in different environment volume could be different introducing non-linear effects in the ultrasound signal.

Similarly, the characteristics and the combination of the speakers also generates unique patterns that model hasn't encountered yet bringing new complexities in the ultrasound STFT. The model performance with data from the new environment can also be improved by incorporating techniques like fine-tuning the model as per the requirement, and modifying the architecture to handle the encountered variations better. However, due to the time limitation we don't delve into these factors as they require testing in the real-time environment, and more data to be recorded.

In this thesis, we deal with two datasets where results shows that training with unfamiliar data predicted accuracy of 0.75 meters. However, it is important to consider that more data from different environments is required for further research in terms of generalisation.

4.4 Conclusion

We generalise the proposed deep learning model designed in Chapter 3. The coordinates are predicted on a pre-trained model from Chapter 3 using unfamiliar dataset FpMap0. For better generalisation, the designed model is trained using dataset from different environment. Furthermore, the limitations influencing the genralisation are discussed.

First, the existing model trained on FpMap data is utilised, and FpMap0 data is used to prediting the coordinates which resulted in an MSE of 0.5 meter square. Therefore, at the current stage the model will not be effective to deploy in the real-time environment. Secondly, the designed neural network is trained on FpMap0 data which predicts location coordinates with an accuracy of 0.75 meters which ensures the genelisation of the model. However, in the final step we discuss the factors like reflections, reverberation, volume and nature of speaker, where we conclude that these factors plays an important part in the generalisation. In addition, we outline the fine-tuning and modifications of model architecture can also improve the overall model performance.

The environmental factors can be considered a potential research topic for the future development of the deep learning model for ultrasound based fingerprinting IPS.

Chapter 5

Evaluation and Limitations

In the following chapter, MSE as an evaluation metrics used to analysis the performance of the deep learning model designed in Chapter 3 are discussed. In addition, we evaluate the results obtained from both the datasets used in this thesis from Chapter 3 and Chapter 4.

5.1 Introduction

The following chapter discuss the evaluation of the artifact developed in this thesis based on the design science research methodology from Chapter 1. The deep learning model designed in Chapter 3 and generalised in Chapter 4 are evaluated using controlled experiment method.

The hypothesis of evaluation using controlled experiment method is to compare the deep learning model trained on two separate datasets from two different environments. The variation of different environments generalise the overall performance of the neural network having data recorded in different conditions. The evaluation metrics are discussed to quantitatively measure the performance of the model from both datasets.The experimental conditions for both dataset used are consistent to segregate the model's performance on dataset variability. Finally, we interprate and discuss the conclusion of the evaluation.

The chapter is discussed as follows: In Section 5.2, the controlled experiment evaluation is performed on Deep learning model with two experiment datasets from Chapter 3 and Chapter 4. Limitations of this thesis is discusses in Section 5.3. We discuss the results and conclude the chapter in Section 5.4. The code source file of the evaluation is available in Appendix B.

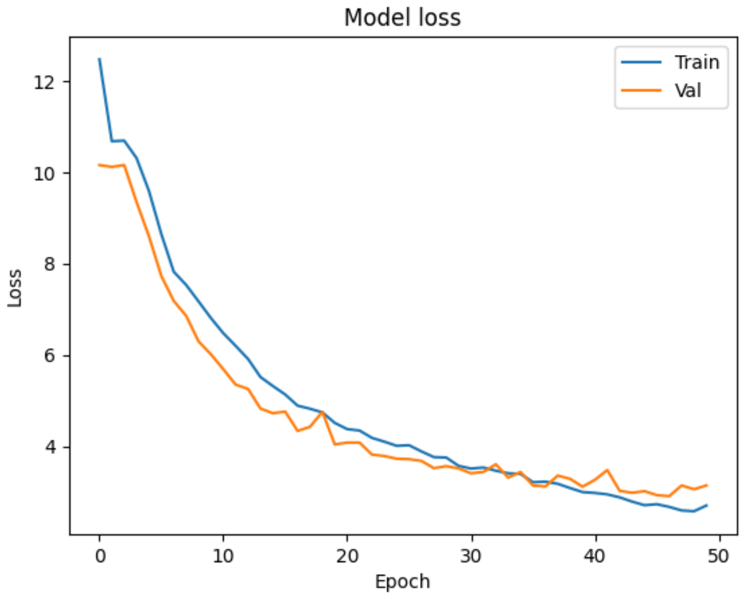

FIGURE 5.1: Training loss vs Validation Loss on FPMap Data

5.2 Evaluation of Deep Learning Model

The controlled experiment evaluation of the designed neural network is performed on FpMap dataset from Chapter 3 and FpMap0 dataset from Chapter 4. The idea behind the evaluation is to formulate the expected difference between the performance of the model in two different environments. Alhomayani et al [21] explains that there's no universal framework for evaluation of Indoor positioning systems. However, evaluation metrics like accuracy, precision and scalability could be used to evaluate the performance of the model. Considering the time limitation of the framework, we have evaluated the model based on it's accuracy.

The experimental conditions for both the dataset are kept consistent and controlled which means the same neural network architecture and hyperparameters have been used for the training, only the datasets are used for different envrionmental settings. The datasplit in the Tables 3.1 and 4.1 show that ratio of datasplit for training, testing and validation is also kept the consitent for the both experimental setup.

In the Figure 5.1 and 4.1, we can analyse the model's performance on training loss and validation loss over 50 epochs for the dataset FpMap and FpMap0 respectively. The trend for the both datasets converges at the same rate without showing signs of overfitting as both training and validation loss decrease significantly. The decrease in loss gradually indicates the increase in model's performance on making better prediction.

The Figures 5.2 and 5.3 qualitatively shows the scatter plot of the actual values scattered in blue dots and predicted values scattered in red dots for dataset FpMap and FpMap0.

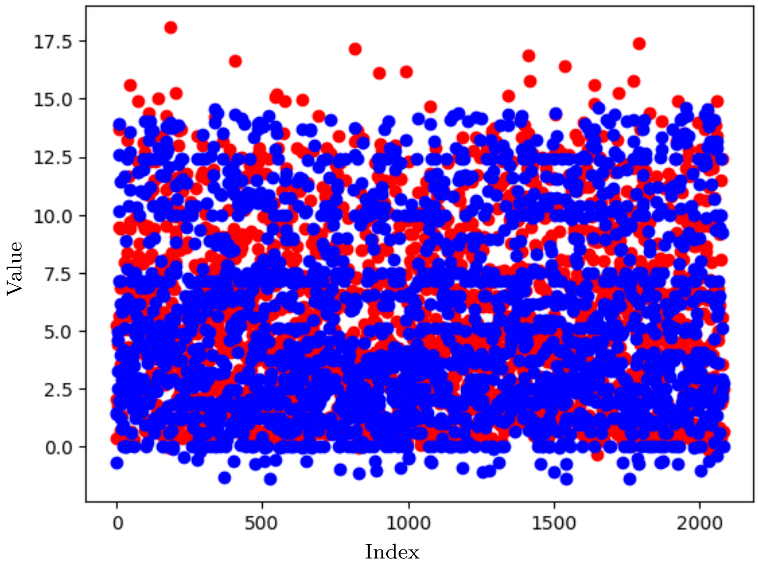

FIGURE 5.2: Scatter plot of actual values as blue dots and predicted coordinates as red dots on FPMap Dataset

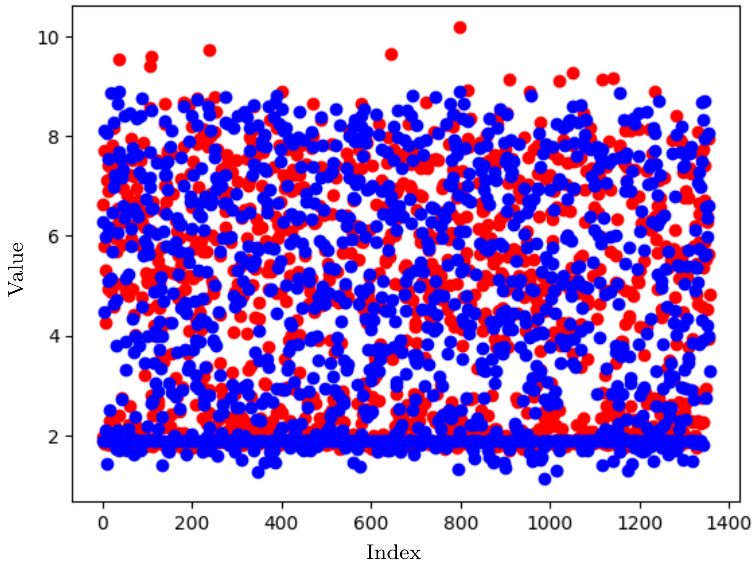

FIGURE 5.3: Scatter plot of actual values as blue dots and predicted coordinates as red dots as red on FPMap0 Dataset

In both the figures, we can see some outliers indicating the data points deviate from the actual values which means there's scope of potential limitations and bias in the model which could be through reflection or reverbration properties discusses in the section 4.3.

Accuracy have been used to quantify the model's performance in reflecting how different is the predicted outcome from the ground truth values. The mean square error is the square of Euclidean distance between ground truth and predicted locations [21]. To calculate the MSE, the y_test location coordinates dataset signifying ground truth values have been used with y_pred location coordinates signifying the predicted values. In the FpMap dataset, final MSE result between the actual and the predicted location coordinates is 2.633 meters which gives the accuracy of 1.62 meter. Similarly for FpMap0 dataset, the MSE result between the actual and the predicted location coordinates came to be 0.57 meters square which gives the accuracy of 0.75 meter.

The final accuracy of the deep learning model for both the datasets resulted better than the conventional IPS system where accuracy is in the range of 7-10 meters. However, the limitaions discussed in the Section 4.3 should be considered.

5.3 Limitations

The achieved accuracy of 1.6 meter, however promising, may still need improvement. Model architectures can be improved by incorporating contextual information, or by leveraging advanced techniques such as feature engineering. Furthermore, including diverse real-world scenarios can improve can help enhance the generalization capabilities of the model. The environmental factor are not taken in consideration for this thesis as they can impact the performance. The setup of speaker, volume of speaker, reflection of the sound can play major part in the audio data. But due to time limitation these factors are not being studied for this thesis.

5.4 Discussion and Conclusion

This chapter evaluates the performance of the designed deep learning model in predicitng (x,y) location coordinates accurately from STFT audio specgram. The Mean Squared Error (MSE) metric is used to quantify the error between the predicted coordinates and actual coordinates which relects the accuracy of the model. The obtained MSE provides insight into the performance of the model, allowing comparison with baseline models or previous approaches. The low MSE suggests that the model captures the underlying patterns and relationships in the spectrogram data, compared to the conventional Koopango IPS model. However, the training with more data from more different Koopango signals recorded is yet to be done. Considering there are many different Koopango signals discussed in Chapter 3, a separate training with different koopango signals data could provide valuable insights in improving the accuracy of the system.

In conclusion, we demonstrate the effectiveness and accuracy of the deep learning model for predicting (x, y) coordinates from audio spectrograms. With a low MSE, the model can make accurate predictions, showing its ability to perform. The accuracy of 1.6 meter from FpMap dataset amd 0.75 meter from FpMap0 dataset shows that location coordinates can be predicted in an indoor space using deep learning model, and the use of deep learning in IPS enhances the overall performance of the system. For future research, there is potential for further improvements, including investigating alternative model architectures, incorporating attention mechanisms, or exploring alternative evaluation metrics. These findings will contribute to development of Koopango Indoor Position system.

Chapter 6

Conclusion and Future Work

The following Chapter 6 concludes this thesis and answers the research questions discussed in Chapter 1. Section 6.1 gives the general conclusion of the thesis, followed by Section 6.2 which answers the main research questions of this thesis. Finally, Section 6.3 talks about the possible future research.

6.1 Conclusion

This thesis designs a deep learning neural network model tailored made for Koopango indoor positioning system to predict location coordinates from the collected fingeprint data. The goal of the thesis concludes by finding the accuracy of the neural network model in predicting coordinates from the audio input. We start by data analysis of the available where STFT spectrogram have been used for the input of the neural network. The two data-sets are used separately in separate training of the model which predicts accuracy and provides generalisation of the model. Convolutional neural network have been used to perform the regression task of predicting (x,y) location coordinates. We also dicuss the limitations of the available resources.

The result of this thesis provides an ultrasound based indoor positioning system which gives accuracy of 1.6 meters from FpMap dataset and 0.75 meter from FpMap0 dataset making it a better IPS system in single channel. Hence, saving the infrastructural cost of the overall Koopango IPS. Based on the results of this thesis, deep learning models might be a good choice for audio-based location prediction.

6.2 Answering the Research Questions

The main research question defined for this thesis: How to improve the accuracy of a fingerprinting indoor positioning system to find location coordinates from speakers emitting unique ultrasound signal using deep learning techniques? In Chapter 1 we divided the research question into two-sub questions to answer. The following sections conclude the answers of each subquestions.

6.2.1 RQ-1 - How to find the accuracy of the deep learning model of an ultrasound based fingerprinting IPS with audio signal and location coordinates data?

The two-dimensional convolutional neural netwokr is designed to perform the regression task in predicting (x,y) location coordinates from STFT audio spectrogram as an input of nueral network. Due to overfitting and the loss of important features during the training process, the hypothesis of considering an audio high-pass filter is neglected. This thesis discusses the key hyperparameters influencing the model and aims to improve its accuracy. Table 3.3 presents the final model with tuned hyperparameters that can accurately predict (x,y) coordinates with 1.6 meter in diameter on FpMap dataset and 0.7 meter on FpMap0 dataset.

6.2.2 RQ-2 - How to generalize the proposed deep learning model for Koopango IPS with the fingerprinting data recorded in a different environment from the original data?

The model is trained on two separate data-sets in order to perform generalisation. First, The unfamiliar dataset is used to predict the coordinate without training. However, the results doesn't show a better outcome thus, we conclude that model will be effective to deploy in the real-time environment. However, after the training with unfamiliar dataset the accuracy performed even better than the first dataset used during the designing process, where we can conclude that model performs good results in generalisation. Though we discuss certain environmental limitations that should be considered like reflections, volume and reverbration which can improve the overall model performance of future results.

6.3 Future Work

The accuracy of 1.6 metres indicates the pattern's competence to make use of hearing facts to correctly deduce spot info. Long term investigation initiatives ought to aim at refining the pattern additionally and checking into its employability in serious world contexts to unlock its total opportunity.

Whereas this investigation has indicated uplifting outcomes regarding anticipating (x,y) coordinates from audio spectrograms, numerous roads remain for potential examine that could additionally raise the precision and applicability of the model. Some possible instructions for potential examine comprise:

Incorporating complex structures: Finding exceptionally complicated deep training styles for example reverse neural systems (RNNs), convolutional reverse neural systems (CRNNs), or transformer-based designs may offer much better outcomes in grasping long-term connections and temporary examples in audio information. These complex designs can be analyzed to decide how acceptable they might be for position predictions.

Different datasets: Making the dataset include more real life conditions, and unique sound settings can improve how the model works for many uses. Data collected from many places, times of the day, and places with different sounds will make sure the model works well when really used.

Transfer learning: Inspecting software applying transmission grasping skills may take advantage of prepared illustrations on massive sound databases, like ImageNet, to set the commencements of the version weights. Additionally, domain adaptation tactics may be investigated to bridge the hole between training and use environments, ensuring better performance in real-world scenarios.

Deployment in real-world environment: Doing fieldwork and evaluating the model's execution in real conditions will give understandings into its sensible usability. Determining the model's execution in difficult acoustic conditions, varying separations, and different environmental settings will validate its dependability and pinpoint probable restrictions for improvement.

By handling these prospective exploration paths, we can additionally forward the correctness and sturdiness of deep learning models for presaging (x, y) coordinates from audio spectrograms. The outcomes of such research undertakings hold the capability to revolutionize location based IPS, reinforce scenario awareness, and add to various areas such as navigation.

Appendix A

Graph Plots

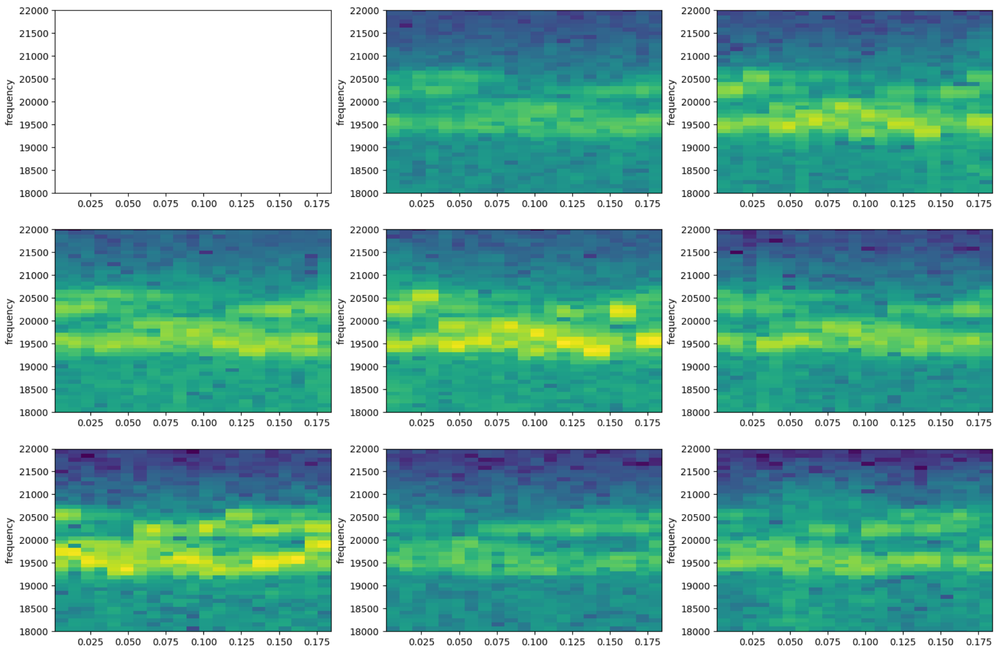

FIGURE A.1: Audio Data in frequency domain for FpMap0 setup

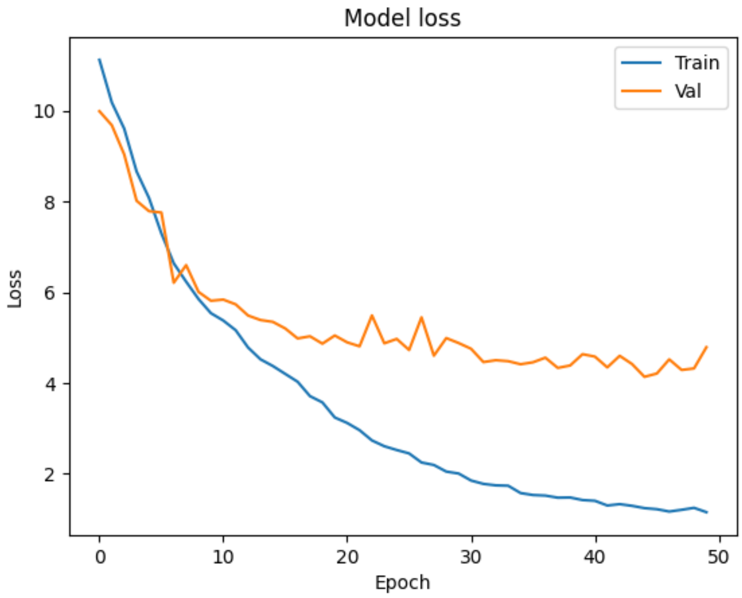

FIGURE A.2: Training loss vs Validation Loss on FPMap Data with Highpass Filter Enabled

Appendix B

Source Code

B.1 Repository

Github Respository with source code:

https://github.com/codecraze01/Deep-learning-Koopango-IPS

Bibliography

[1] Ye Cai, Weipeng Guan, Yuxiang Wu, Canyu Xie, Yirong Chen, and Liangtao Fang. Indoor high precision three-dimensional positioning system based on visible light communication using particle swarm optimization. *IEEE Photonics Journal*, 9(6): 1–20, 2017. doi: 10.1109/JPHOT.2017.2771828.

[2] W. Sakpere, M. Adeyeye-Oshin, and N. B. W. Mlitwa. *A state-of-the-art survey of indoor positioning and navigationsystems and technologies*. South African Computer Journal 29(3), 2017.

[3] C. Jain, G. V. S. Sashank, V. N, and S. Markkandan. Low-cost BLE based Indoor Localization using RSSI Fingerprinting and Machine Learning. *2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 10.:363–367, 2021.

[4] Hakan Koyuncu and Shuanghua Yang. *"A Survey of Indoor Positioning and Object Locating Systems*. 2010.

[5] P. ; Toledano-Ayala Richter and M. *Ubiquitous and Seamless Localization: Fusing GNSS Pseudoranges and WLAN Signal Strengths*. Mob. Inf. Syst, 2017.

[6] J. M. ; Yoon Kang, T. S. ; Kim, E. ; Park, and J. B. Lane-level map-matching method for vehicle localization using GPS and camera on a high-definition map. *Sensors 2020, 20*, 2166., 20.

[7] W. E. Sakpere and M. Adeyeye. Can near field communication solve the limitations in mo-bile indoor navigation? In I. *RFID technology integration for business performanceimprovement(Chap*, 3:52–79, 2015.

[8] Y. ; Ai Deng, H. ; Deng, Z. ; Gao, W. ; Shang, and J. An overview of indoor positioning and mapping technology standards. *Standards 2022*, 2, 2022.

[9] N. K. Vishwakarma and R. K. Singh. Design and implementation of FHSS (frequency hopping spread spectrum) synthesizer. *2021 7th International Conference on Signal Processing and Communication (ICSC)*, 10.:151–155, 2021.

[10] D. [Da] Zhang, F. Xia, Z. Yang, L. Yao, and W. Zhao. *Localization technologies for indoorhuman tracking.* In5th International Conference on Future Information Technology (FutureTech), Busan, 2010.

[11] Timothy Dougan and Kevin Curran. Detection of social interaction using mobile phones via device free passive localisation. *International Journal of Handheld Computing Research*, 5:20–35, 10 2014. doi: 10.4018/IJHCR.2014100102.

[12] R. F. ; García-Vázquez Brena, J. P. ; Galván-Tejada, C. E. ; Muñoz-Rodriguez, D. ; Vargas-Rosales, C. ; Fangmeyer, and J. *Evolution of Indoor Positioning Technologies: A Survey.* J. Sens, 2017.

[13] A. Behboodi. *October 24).* A Mathematical Model for Fingerprinting-based Localization Algorithms. arXiv. Org, 2016.

[14] R. Mautz. *Indoor positioning technologies(Habilitation thesis.* ETH Zurich), 2012.

[15] Y. Gu, A. Lo, and I. Niemegeers. A survey of indoor positioning systems for wireless personalnetworks. *IEEE Communications Surveys Tutorials*, 11(1), 2009.

[16] H. Schweinzer and M. Syafrudin. Losnus: An ultrasonic system enabling high accuracy and secure TDoA locating of numerous devices. *International Conference on Indoor Positioning and Indoor Navigation*, 10.:1–8, 2010.

[17] S. He and S. H. Gray Chan. "wi-Fi fingerprint-based indoor positioning: recent advances and comparisons, " IEEE Communications Surveys and Tutorials. 18(1): 466–490, 2016.

[18] L. ; Rozum Polak, S. ; Slanina, M. ; Bravenec, T. ; Fryza, T. ; Pikrakis, and A. Received Signal Strength Fingerprinting-Based Indoor Location Estimation Employing Machine Learning. *Sensors 2021, 21, 4605.*

[19] X. Wang, L. Gao, S. Mao, and S. Pandey. Deepfi: Deep learning for indoor fingerprinting using channel state information. *IEEE Wireless Communications and Networking Conference (WCNC)*, 10.:1666–1671, 2015.

[20] J. Xue, J. Liu, M. Sheng, Y. Shi, and J. Li. A WiFi fingerprint based high-adaptability indoor localization via machine learning. In *China Communications*, volume 17, pages 247–259. 10, doi, July 2020.

[21] Fahad Alhomayani and Mohammad Mahoor. (2020). *Deep learning methods for fingerprint-based indoor positioning: a review. Journal of Location Based Services. 14. 129-200*, 14.(10.):129–200, 2020.

[22] Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. Design Science in Information Systems Research. *MIS quarterly*, 28(1):75–105, 2004.

[23] DEJ Technology GmbH. Experimental setup of koopango ips, 2021. URL https://koopango.com/.

[24] DanielDiscoveries. Spectrogram input normalisation for neural networks. daniels discoveries., 2017. URL https://danielsdiscoveries.wordpress.com/2017/09/29/spectrogram-input-normalisation-for-neural-networks/.

[25] Sakshi Indolia, Anil Kumar Goswami, S.P. Mishra, and Pooja Asopa. Conceptual understanding of convolutional neural network- a deep learning approach. *Procedia Computer Science*, 132:679–688, 2018. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2018.05.069. URL https://www.sciencedirect.com/science/article/pii/S1877050918308019. International Conference on Computational Intelligence and Data Science.

[26] Antonello Rosato, Rodolfo Araneo, Amedeo Andreotti, Federico Succetti, and Massimo Panella. 2-d convolutional deep neural network for the multivariate prediction of photovoltaic time series. *Energies*, 14(9), 2021. ISSN 1996-1073. doi: 10.3390/en14092392. URL https://www.mdpi.com/1996-1073/14/9/2392.

[27] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

[28] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR. URL https://proceedings.mlr.press/v15/glorot11a.html.

[29] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, may 2017. ISSN 0001-0782. doi: 10.1145/3065386. URL https://doi.org/10.1145/3065386.

[30] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[31] Mayank Banoula. Regularization in machine learning —— simplilearn, Feb 2023. URL https://www.simplilearn.com/tutorials/machine-learning-tutorial/regularization-in-machine-learning.

Declaration of Authorship

I, Ankit Sharma, declare that this thesis titled "Deep Learning Approach to identify locations from speakers emitting the same ultrasound signal in an Indoor Positioning System" and the work presented in it are my own. I confirm that I authored this thesis independently, that I have not used any sources other than the declared sources, and that I have explicitly marked all material which has been quoted either literally or by content from external sources.

Date: ________________________________

Signed: ________________________________