Location Finding in Natural Environments with Biomimetic Sonar and Deep Learning

Liujun Zhang

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Engineering

Rolf Müller, Chair
Leonessa, Alexander
Abbott, A. Lynn
Stilwell, Daniel J.
Abaid, Nicole Teresa

September 22, 2022
Blacksburg, Virginia

Keywords: Bats, airborne sonar, biomimetic robot, machine learning, outdoor navigation, ultrasound echolocation

Location Finding in Natural Environments with Biomimetic Sonar and Deep Learning

Liujun Zhang

(ABSTRACT)

Bats are famous for their capability of navigating in dense forests for hundreds of kilometers within one night by using their sonar system. Airborne sonar hasn't been heavily used in the industrial world compared to other sensors such as lidar, radar, and cameras. In this study, we applied a biosonar robot to navigate in a dense forest with bat-like FM-CF ultrasonic signals with deep learning. The results presented show that airborne biosonar can classify different areas' plants, in addition to achieving a similar level of navigation granularity compared to GPS, which is about 6 meters of radius resolution. The time-frequency representations of echoes from the forest are used as input data to explore the biosonar navigation ability, and the state-of-the-art CNN deep network (Resnet 152) is used as the processor to do the echolocation in the dense forest. The navigation ability can be improved significantly by combining multiple 10 ms long echoes, however, the data size of the reflected waves is much smaller than the other popularly used sensors, as echo can be collected at a rate of 40 echoes per second. The results can prove that airborne sonar can be used to navigate in GPS-denied environments, and can be an important sensor used in a scenario when other sensors meet constraints, like in the sensor fusion applications.

Location Finding in Natural Environments with Biomimetic Sonar and Deep Learning

Liujun Zhang

(GENERAL AUDIENCE ABSTRACT)

The ability to identify natural landmarks could contribute to the navigation skills of echolocating bats and also advance the quest for autonomy in natural environments with man-made systems. The critical sensors used in autonomous robot navigation are camera array, radar, and lidar, airborne sonar hasn't been verified for its navigation efficiency. However, recognizing natural landmarks based on biosonar echoes has to deal with the unpredictable nature of echoes that are typically superpositions of contributions from many different reflectors with unknown properties. This dissertation intends to explore the bioinspired airborne sonar navigation ability in dense natural forests. The first part of this project is to use reflected echoes to navigate on a large scale. Data were collected from different mountains which are dozens of kilometers away from each other, and we achieved the use of one single navigator in those locations. The second part is to explore the navigation granularity of airborne sonar sensors. Data were collected from a small dense forest area, we try to classify which part of the foliage was based on the echo, and in the end, we achieved GPS accuracy for navigation. The finding in this work proves that the sonar sensor can play an important role in the sensing system, with the help of a deep neural network, with a 10 ms long echo, it can have a similar navigation ability to GPS.

Dedication

I dedicate this dissertation to my father Furong Zhang, my mother Guiyun Chen, and my Yima Xueyun Chen, for the love they give me, for the greatest memory I have, and for the strength they provide. I also would like to thank my sister Yufeng Zhang, thanks for raising me up to a better world, also my sister Gaijie Zhang, Fengjie Zhang, Yuping Zhang, and my older brother Ruqiang Zhang, thanks for the happy memories, give me a family filled with love. In addition, my wife Yajun Wu, without her company, might have finished my Ph.D. last year. Last but not least, the kids from the family, Haipan Xue, Shuang Zhang, Haixiao Xue, Yapeng Zhang, Yupeng Zhang, Dongming Zhang, Jiayue Wang, Jiayi Wang, Zimu Zhang, and Jiajun Wang, thanks for the happiness they provided.

Acknowledgments

Most importantly, I am very grateful for my advisor Rolf Müller, for the consistent support for research, for the opportunities he provided, and for the willingness to help me out when I was stuck. I learned a lot from him during the past seven years, learned how to be a professional engineer/researcher, how to solve a problem effectively with intelligence, learned how to think critically about research problems. I met Dr. Mueller when I haven't figured out my future career pass, now I am earning a Ph.D. degree in the USA, and am full of passion for the next chapter of my life. I really appreciate him for bringing me to a much better stage for me to shine. I wish I can be a person like him in the future, full of passion for work and life, full of energy for the future, and working hard toward the goal. In addition, I also would like to express my sincere gratitude to my committee members: Drs. Abbott, Leonessa, Abaid, and Stilwell who provided valuable guidance and suggestions for my research project. I would like to acknowledge the students who helped me with data collection and signal pre-processing, Andrew Farabow, Pradyumann Singhal, Boyuan Wang, and Jake Bennett. I would like to acknowledge Michael Goldsworthy helping for me understand the machine learning algorithm and Lucas Mun, Brandon Walker, and Joseph Sutlive for building a sonar head for data collection. I would also like to thank my labmate and friends, Xiaoyan Yin, Ruihao Wang, Ananya Bhardwaj, Yihao Hu, Shubham Dawda, Andrew Farley, Lucas Mun, Joseph Sutlive, and Luhui Yang, thanks for the company and fun time. Finally, I would like to acknowledge my family for all their support.

Contents

List of Figures

List of Tables

List of Abbreviations

CF Constant Frequency

CNN Convolution Neural Network

dB Decibels

DNN Deep Neural Network

FFT Fast Fourier Transform

FM Modulate Frequency

GMM Gaussian Mixture Model

MLP Multi-layer Perceptron

ReLU Rectifiel Linear Unit

ResNet50 Deep Residual Network 50 Layers

RMS Root Mean Square

SNR Signal Noise Ratio

STD Standard Deviation

Chapter 1

Introduction

1.1 Airborne sonar navigation

Autonomous vehicles and drones have been widely used in agriculture and industrial applications, such as drones in rescue and search missions [134], wildfire detection and monitoring [105], and planting seeds and shooting plant nutrients in agriculture [111]. Recognizing landmarks in natural environments is important for outdoor robot navigation [60, 139]. One of the widely used applications can be the autonomous vehicle for 3D localization mapping and landmark classification in complex environments [29, 65, 121]. A sensor that can recognize targets has many potential applications, and this ability is applied well to agriculture applications. A use case scenario of landmark classification for agriculture is the detection of weeds in crop rows, selective spraying of herbicides, cutting off branches, and harvesting fruits [8, 45]. In addition, living assistance robots can be used for indoor environment navigation, and assisting older adults living [104]. Besides the camera, radar, laser scanner, and IR sensors, airborne biosonar plays a useful role in landmark classifications, and bats are a good model for us to learn how to utilize airborne sonar.

Conventional approaches to navigation in natural environments include GPS, lidar, radar, RF, and camera arrays. However, all the sensors have their own limitations. For example, GPS has limitations: reduced accuracy under foliage, GPS-denied environments the accuracy is not reliable [53, 103]. Visual sensors [61, 100] rely on objects that can be recognized using

salient cues, and plants have complex shapes that can not be described in terms of simple geometrical primitives [20]. For laser sensors [50], processing the millions of 3D coordinates data produced by these sensors in real-time leads to high computational loads and power consumption.

Air-based sonar sensors also can be used for target detection, however, air-based sonar sensors are only applied to detect the distance [102] to the target due to their low angular resolutions compared to other sensors. Recent applications focus on wider uses such as landmark-based localization, mapping, and making position prediction more precise [132]. However, airborne sonar has its advantages compared to others: it does not need direct physical contact and the capability is not easily influenced by the material or obstacles' optical properties. Consequently, airborne sonar is well-suited for robot operation in terms of the working environment, such as navigation in unknown structure caves or low-visibility scenarios. Sonar is a very useful and cost-effective mode of sensing for mobile robots. These abilities exceed what was achievable with cutting-edge autonomous navigation applications, where autonomous robots are still challenged by obstacle avoidance and target segmentation [29, 72, 140], hence making learning from bat biosonar [89, 91] an attractive solution to meet the challenges posed by autonomous navigation in natural environments [88]. Sonar sensors were universally applied in autonomous vehicles, such as robots and driverless cars. Navigation in natural environments presents a multitude of applications opportunities in forest search and rescue [17, 59, 73], agriculture [8, 45], and surveillance [120].

In the past, different methods have been proposed for recognizing objects based on the airborne sonar approach. A few studies directly addressed simple objects [86] classifications using echolocation, for example, three types of targets (plane, edge, and corner) [6]. This is usually based on simple cues that can be easily recognized in the time-frequency representation of the echoes, such as a certain notch arrangement in the frequency domain. Template

matching has been shown to be a possible solution for target classification and place recognition [132] tasks, which are using the reflected echoes compared to the stored echoes that represent known objects or known places. Most of the previous work using airborne sonar to recognize targets is mainly on a single artificial target, or simple shapes [6, 81, 132]. Airborne sonar for classification work is not only useful for simple targets but is also useful for complex objects. Further research has attempted more complex targets such as artificial trees, different targets have different reflected spectrograms [123].

However, the reflection echoes have a high degree of complexity from the natural target [81]. Those techniques are limited to the complex natural environment, and the classification ability will face severe difficulties when the target is changing, or when the reflected echo is out of the range of stored labeled echoes. Some methods have been applied on classifying the real/natural targets on the computational side. For example, a few studies try to classify complex echoes, using a few selected parameters corresponding to the classes, such as the target from the natural environment [86, 87]. However, the previously determined parameters (energy of the echo or impulse response) have strong assumptions related to physical plausibility, and some of the key features might be overlooked. Plants have complex shapes that cannot simply be described by the geometrical primitives [93].

At the same time, using airborne sonar needs to avoid more inherent limitations. The first limitation of a biosonar is its low dimension. Compared to the camera-based sensors, biosonar does not have a lot of flexibility to change its resolutions like cameras can adjust lenses for specific environments. Biosonar echoes use a 1D array from the ear to represent the 3D targets reflections, so a short echo cannot provide enough information to describe complex natural foliage. In addition, the signal might distort during transmission between the sonar head and targets. Due to the duration of pulse emission, the reflected echo can not solely represent the reflection target linearly between distance and time, the echo at the

same time always overlaps from different parts of the 3D target. To solve those problems, we need to improve the capability of the bio-inspired robot and use more advanced methods to quantitatively distinguish the echoes, such as the machine learning approach.

1.2 Sonar sensing inspired by bats

The limitations of the traditional sensing and airborne sonar sensing approaches lead to an interest in bioinspired techniques, and bats are an ideal example to learn from of their astonishing flying abilities. Bats are highly adept at using their sonar system, traveling long distances in dense forest habitats. For example, bats can travel 100 kilometers in a single night and return to their roosts in the morning [77]. Bat's sonar navigation capabilities show that bats could build a detailed map of their environments [29, 65]. In addition, echolocating bats are capable of navigation in a wide variety of natural environments, such as dense forest vegetation, using biosonar [114]. Bats have been shown to successfully identify foraging habitats such as meadows, bushes, trees, etc. which are indicators of specific foods sources [58, 126].

Bats emit ultrasound pulses through their nostrils or mouth, using the returning echoes from foliage to navigate [43, 114]. Bats can emit around 14 ultrasonic chirps every second to navigate and hunt [62], based on the reflected echoes bats can distinguish their prey and objects. Furthermore, bats can infer the geometrical shape and texture of the object from returning echoes [31, 35, 44, 115, 117], which allows them to use those acoustical landmarks for navigation.

The ability of the bats to navigate in natural environments based on biosonar is of particular interest because of the special nature of sonar echoes from natural environments random [86, 141], i.e., unpredictable waveforms. The foliage has complex shapes that can

not be described in terms of simple geometrical primitives [93]. From an acoustical point of view, foliage can be approximated as a stochastic array of reflectors formed by plants [76]. Their ability to fly in the completely dark cave and lush foliage, and the ability to recognize landmarks in nature are important for outdoor robot navigation. The interpretation of that capability has a critical influence to understand the bat's sonar system, this insight could have important implications for the understanding of bat biology, also a better-bioinspired bat robot. Inspired by the bats' biosonar system, some researchers have utilized active sensing techniques in biomimetic robots and tried to classify different landmarks and textures using high-frequency echo signals. Biosonar-based feature detection and landmark classification was a very popular topic in the engineering area; there are plenty of applications for autonomous vehicles, delivery drones, medical devices as well as agriculture robots. Visual-based sensors have been widely used for relative applications. Here we show that airborne sonar can play an important role in object classifications. The prior art has achieved landmark findings based on a bat-like sonar system, using reflected echo with the neural network to do the binary classification: open space or obstacles [29], or distinguish different target geometries (disc, cylinder, and hollow hemisphere). [65]. However, no researcher has used the airborne sonar sensors to do target classification in a complex environment, such as natural forest landmarks, or try to navigate in GPS denied environment, one of the reasons would be a short one dimension wave is hard to interpret the 3 D space foliage detail information. In order to achieve this, we either need to have more information from the forest, like adding more visual information then use sensor fusion to recognize landmarks, which will add more latency for signal processing, and cost more money and computing power. Or use an advanced algorithm to understand the spatial information from a short pulse, such as the deep neural network.

1.3 Deep learning algorithm applied on nature environment navigation

Deep neural networks (DNN) are employed as classifiers for airborne sonar-based target classification. DNN have weaker assumptions for the distribution of input data, compared to traditional statistical methods, but performance is more robust for classification. The motivation behind the use of the DNN sonar classifier is to emulate the remarkable pattern recognition capability of humans and animals/bats [116]. Neural networks have been employed efficiently for pattern classifications in numerous applications [71]. For example, the convolutional neural network (CNN) started to play a role in classifying targets based on echoes, such as sphere objects with different diameters by using the reflected spectrogram [26] and binary obstacle classification ("plant/no plant") for robot navigation using a generic feature from audio processing [29]. The feature of the reflected echoes here is the preprocessed data and uses DNN for learning and recognizing different features [40, 46, 145, 146].

In this research, we have explored the capabilities of airborne sonar used for landmark classifications. Furthermore, after achieving the large-scale navigation, then we take the sonar head into the dense forest in our Virginia Tech stadium wood to verify the sonar robot navigation granularity. For the first step of work, we take the biomimetic sonar robot [123] into the real forest distributed in a large area to navigate, by imitating real bats flying through the real forest. A biosonar robot navigation system equipped with sonar sensors is presented for natural foliage, and big data has been collected from different locations, which supposedly cover a variety of natural landmarks. We apply the machine learning approach to classify different natural landmarks, using the reflected echoes spectrogram as the input. The expectation for this work would be that our biosonar robot would be able to distinguish different data collection locations/targets based on a single echo with the help of a neural

network and big data.

After we combined the sonar device and the deep neural network, prior work has shown that large-scale identification of different locations in natural environments based on single (15 ms) echoes is possible using deep learning to determine ten different locations that were spaced within a 50-kilometer diameter, but also neighboring walking trails at the same location, this could indicate that not only different habitats can be distinguished, but finer discrimination is possible building on these findings. Then the second part of the work sought to determine the granularity of natural habitats to sonar-based location findings i.e., how finely different locations can be distinguished based on biosonar echoes in natural habitats. In order to do this, we take our sonar head to scan the entire stadium wood forest area on the Virginia Tech campus. The stadium wood is a dense forest with a size of 150 m wide by 180 m in length, we take our sonar robot to scan the entire area. The entire area is separated into 70 rows in the horizontal direction. We start from one corner of the first row, used ultrasound to scan the edge, then move about 2 m into the forest to the second row, scan it again and keep doing it until finish scan the entire 70 rows.

Chapter 2

Rationale of the approach

2.1 Natural forest data collection

A huge amount of data was collected from the natural forest by using the biosonar robot. The data included the large-scale forest area distributed around the Virginia Tech campus with a 50 km diameter, and additionally from a small forest area which is about 150 m in width and 180 m in length located on the campus.

For the large area data collection, 10 sites (Pandapas Pond, Mountain Lake, Cascades, etc. about dozens of kilometers away from each other (Table. 3.2)) were selected which have different plants, including pine trees, apple trees, bamboo, etc. At each site, students collected data from two tracks. The tracks were about a few hundred meters apart and approximately 500 m long. Collecting data from the two tracks not only enabled us to have more data used for DL location classification, but also allowed us to classify the tracks within the same site.

During data collection at each track, students hold the sonar head (Fig. 3.1, describe more in Chapter 3) by hand and let the sonar head face the trees. Students hold the sonar head from top (around the same height as the head) toward to the trees, then move downward to the bottom (around the height of the thigh), and then walked slowly along the trail around the speed of $360\,\mathrm{m\,h^{-1}}$. Because trees are randomly located along the trail, and the sonar

head always faces the trees, the reflected echoes' direction is also randomly oriented. Here, the random case is only in a horizontal direction. No data is in a case of the sonar head [90] emitting the ultrasound face to the sky or soil. The student tried to keep the foliage at a distance from the sonar head in a range of 1 m to 1.5 m. The reason is that to keep a small distance, which would give us a good SNR. The sonar head also can not be too close, otherwise the original emission would have a big overlap with the echo, and it would be hard for us to separate those two signals.

However, because of the randomness of the trees, some data was from a distance outside of this range. For example, some areas of the foliage are very dense, so it's hard to maintain a distance of exactly 1 m, but the majority of the data is in that range. At each track, at least 2000 echoes were collected.

In natural environments such as forests, the trees are randomly distributed by different species, sizes, textures, and densities. The reflected echoes collected from such environments also proved to not be correlated with previous research [14]. Here, the echoes have been verified to contain invariant information during the one-track data collection (Fig. 3.4). A continuous path that included 50 echoes were picked out, and the correlation coefficient between different echoes was estimated at a mean of 0.13 with a standard deviation of 0.04. This means for the random forest location classification, the foliage target responds without discernible deterministic patterns.

2.2 Time - frequency representation of the echoes

In the present work, time-frequency representations, i.e., spectrograms (Fig. 2.4), were used for representing the echoes. This has been common practice in the classification of audio signals [39, 131] and is based on the hypothesis that the relevant features are particularly

accessible in the time-frequency plane, as is the case in speech, where part of the information (e.g., different vowels) exists in the frequency domain and syllables can be recognized from changes over time.

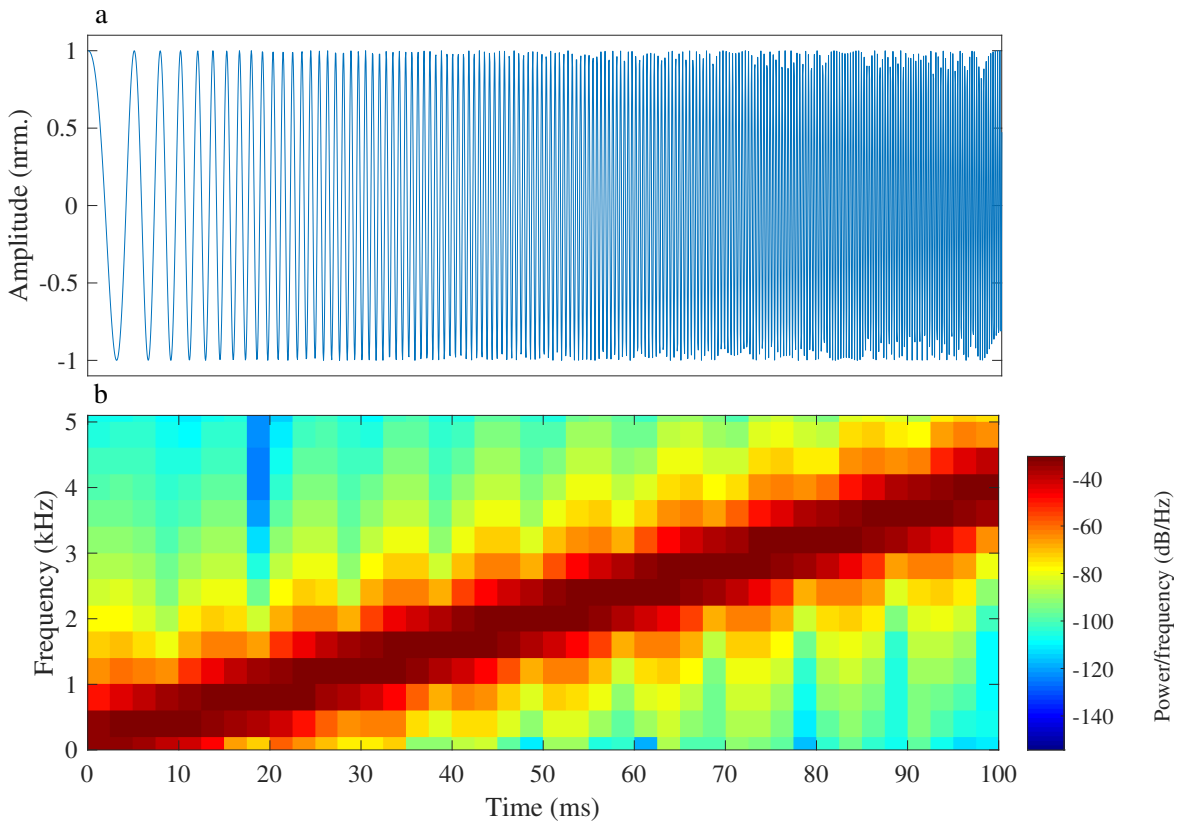

Figure 2.1: **Audio signal representation:** a) Shows the traditional sound wave representation, the sound pressure amplitude changes over time (Fig. 4.2a). b) Shows the time-frequency representation, the X dimension represents time, the Y dimension represents frequency, and the color represents the energy contained at each time-frequency bin (Fig. 4.2b).

The computation of a spectrogram has several parameters that can be used to control the properties of the representation: sampling rate, which is fixed for the hardware; FFT/window length, which decides how many samples are used to calculate the frequency information; and the overlap used to decide how many samples are reused to calculate the FFT in the

next time window.

Not all spectrograms are represented in the same way. An algorithm known as the "Fast Fourier Transform," (FFT) is used to compute the three dimension display (Fig. 2.2). Many parameters that feature a spectrogram display allow to adjust the size of the FFT, changing the FFT window size or overlap between the windows will change the way the algorithm computes the spectrogram. This causes a different time-frequency resolution. Depending on the type of audio that one is working with and visualizing, changing the FFT size may help to understand the audio signal. The FFT process (Fig. 2.2) uses an input vector of amplitude along time. The algorithm selects a piece of the signal in a window with a certain length, then calculates the spectrum by using the FFT, which gives the frequency information inside of this time window. Then, the window is moved to the next time period, and the FFT repeated until the end of the sample to have all the spectra. In the end, all of the spectra plotted against time will give us the spectrogram.

As a rule, the time bin width of the spectrogram is equal to the sampling rate divided by the FFT window size, so higher FFT window sizes give more detail in frequency direction, referred to as frequency resolution, while lower FFT window sizes give more detail in time, referred to as time resolution. The time and frequency resolutions are inverses of each other (Fig. 2.3). Here we use a Gaussian example to demonstrate the relation between the time and frequency resolutions. When a waveform is in temporal domain, its amplitude has a standard deviation of 0.1 s. When the signal is in frequency domain, its amplitude has a standard deviation of 10 Hz, which means the time resolution is the inverse of the frequency resolution. Another example of the time-frequency representation is 0.5 s STD in time domain, while the STD is 2 Hz in frequency domain. The resolution can be changed by interpolating one of the parameters while the other is fixed; however, the default setting of the time-frequency resolution is inverse as required by the Matlab spectrogram function.

This relationship between temporal and frequency resolution has been verified by applying the FFT to the same FM signal. The chirp signal sweeps from 40 kHz up to 60 kHz with a duration of 20 ms. The sampling rate is fixed at 400,000. For simplification, we set the overlap to zero, then modified the FFT window size from 600 to 400 and 200 samples, which corresponds to the time bin width which is 1.5 ms, 1 ms, and 0.5 ms. As the results show here, as the FFT window length decreases, the time resolution is worsening, while the frequency resolution is improving.

If trying to identify a piece of muddy low-frequency information, a higher FFT window size in the spectrogram settings will help. If trying to identify a high-frequency event, choose a lower FFT window size.

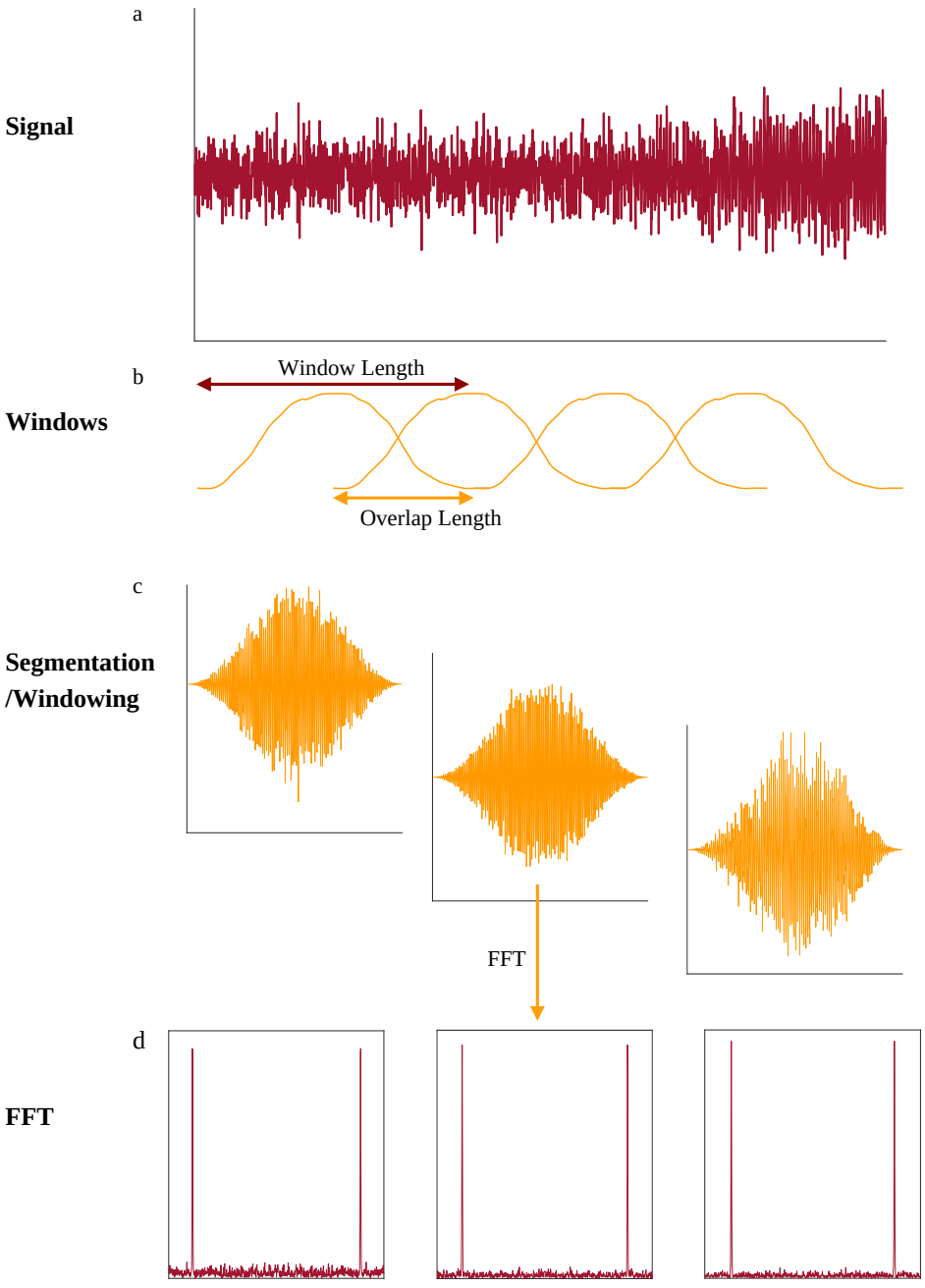

Figure 2.2: **Formation of a spectrogram representation from short-time Fourier transform (STFT).** a) Original audio waveform in time domain. b) A time segment FFT window used to calculate the contain frequencies within this time window. c) Pulse segment after window multiplication. d) Spectrum of the selected time window waveform.

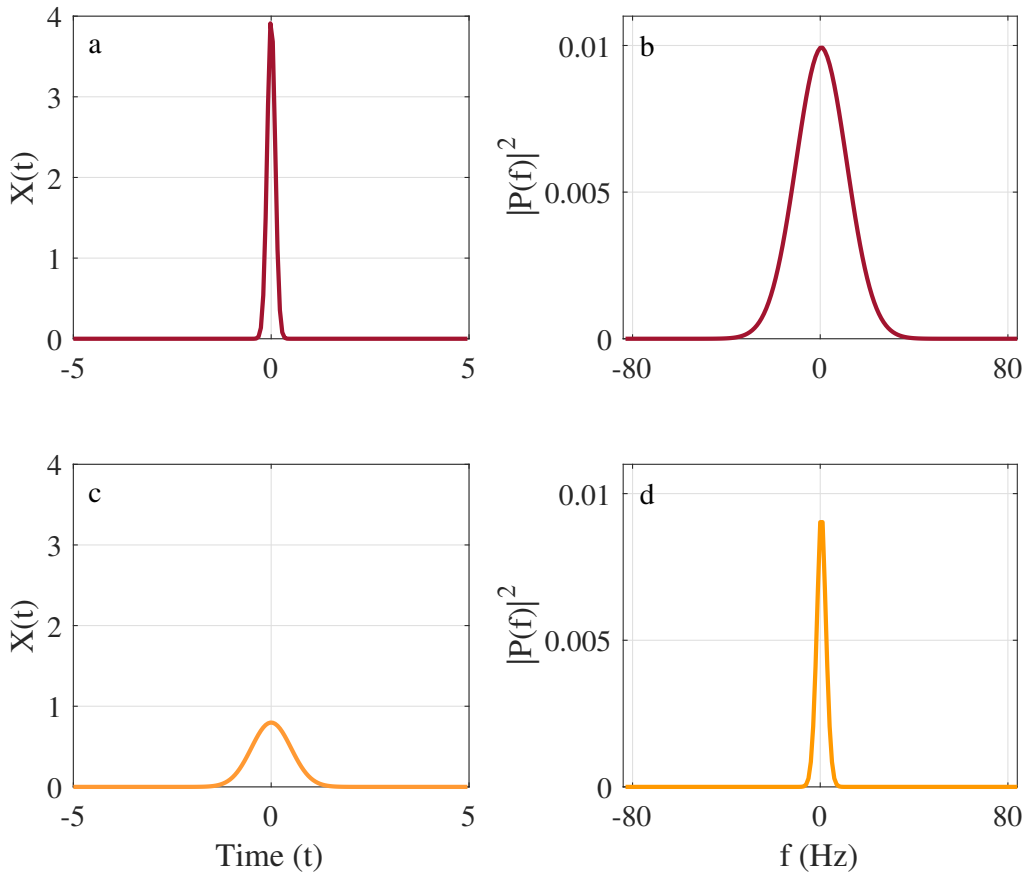

Figure 2.3: **Gaussian pulse in time domain and frequency domain.** a) Gaussian pulse in time domain with a standard deviation of 0.1. b) Gaussian pulse in frequency domain with a standard deviation of 10. c) Gaussian pulse in time domain with a standard deviation of 0.5. d) Gaussian pulse in frequency domain with a standard deviation of 2.

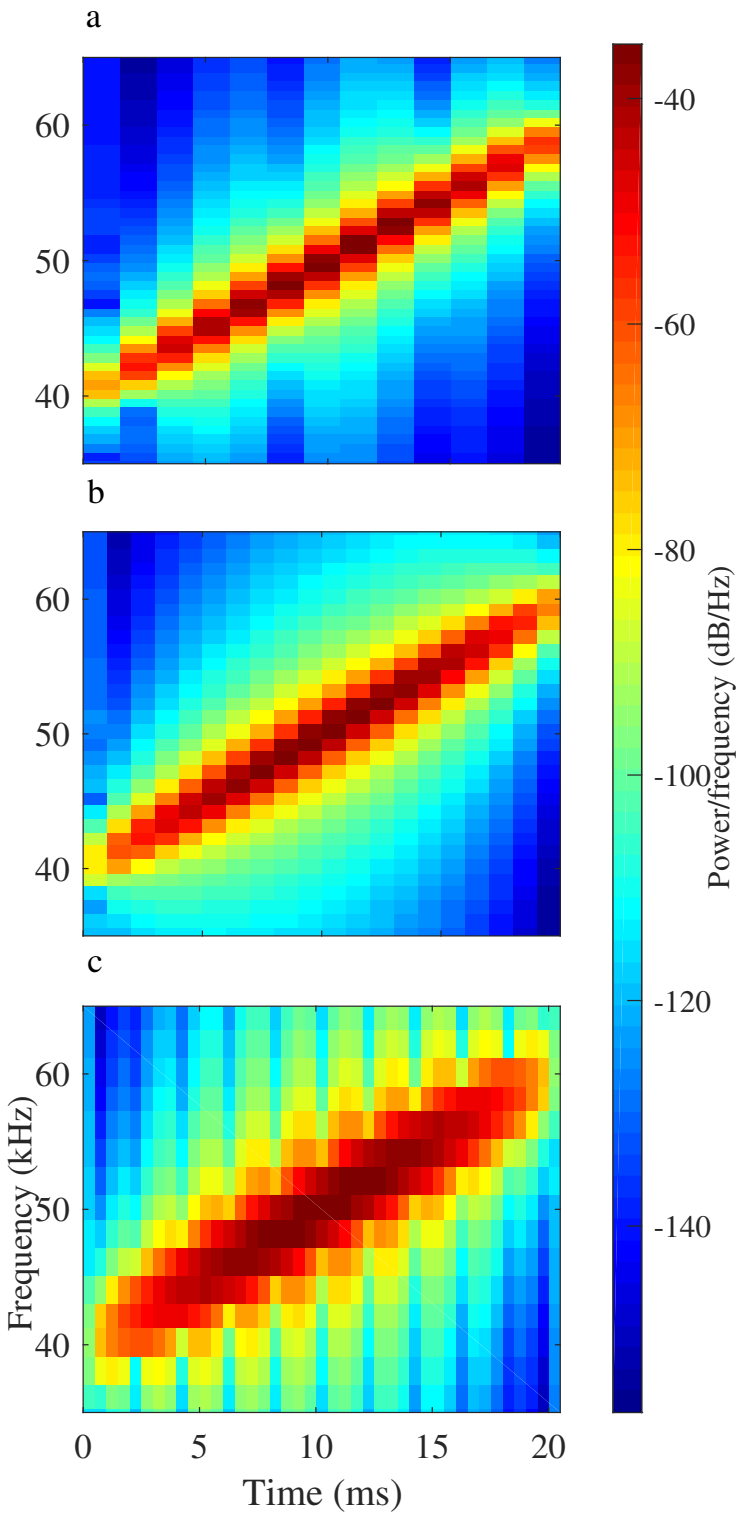

Figure 2.4: **Spectrograms with different time-frequency resolution.** The same sound wave spectrogram representation with different FFT window length. a) The FFT window length is 600, each time bin width is 1.5 ms. b) The FFT window length is 400, each time bin width is 1 ms. c) The FFT window length is 200, each time bin width is 0.5 ms.

Chapter 3

Large-scale recognition of natural landmarks

3.1 Title

Large-Scale Recognition of Natural Landmarks with Deep Learning Based on Biomimetic Sonar Echoes

3.2 Abstract

The ability to identify natural landmarks on a regional scale could contribute to the navigation skills of echolocating bats and also advance the quest for autonomy in natural environments with man-made systems. However, recognizing natural landmarks based on biosonar echoes has to deal with the unpredictable nature of echoes that are typically superpositions of contributions from many different reflectors with unknown properties. The results presented here show that a deep neural network (ResNet50) was able to classify 10 different field sites and 20 different tracks (2 at each site) distributed over an area about 40 kilometers in diameter. Based on spectrogram representations of single echoes, classification accuracies up to 99.6% for different sites and 94.7% for different tracks have been achieved. Classification

performance was found to depend on the used pulse component (constant-frequency - CF vs. frequency-modulated - FM) and the trade-off between time and frequency resolution in the spectrogram representations of the echoes. For the former, classification performance increased monotonically with better time resolution. For the latter, classification performance peaked at an intermediate trade-off point between time and frequency resolution indicating that both dimensions contained relevant information. Future work will be needed to further characterize the quality of the spatial information contained in the echoes, e.g., in terms of spatial resolution and potential ambiguities.

3.3 Introduction

The ability to recognize and map landmarks in natural environments [13, 24] is considered an important prerequisite for outdoor navigation whenever relying on GPS is not an option [9, 60, 75, 112, 139]. In addition to navigation, sensory systems capable of recognizing natural vegetation targets could support applications in the areas of environmental surveillance and precision agriculture [8, 45]. Most prior attempts to provide the necessary sensory information for navigation in natural environments have relied on sensory modalities that are commonly used in autonomous systems such as (stereo)vision [63, 119], radar [4, 15, 98], laser scanning [12, 16], infrared sensing [23, 33], and thermal imaging [113]. In contrast to these more commonly considered sensory modalities, sonar has only been investigated in a few studies that were aimed at landmark classification [29, 65]. Prior work has often focused recognizing simple geometrical shapes [65] or building maps from individually categorized natural objects [29]. It is not clear whether such approaches can work with the full complexity of natural environments and at the scales on which bats are known to navigate. However, the exceptional potential of sonar is continuously being demonstrated by

the ability of echolocating bats to navigate in complex natural environments [84, 94] based on biosonar as their primary far sense [42, 127].

Laboratory experiments have already shown a range of target classification abilities in bats: At the most basic level, discrimination between different reflector spacings [115] and simple deterministic geometries [35] has been demonstrated. Biomimetic sonar systems have been able to achieve classification abilities for targets with deterministic geometries that resemble those seen in bats [11, 26, 65, 118]. At the next higher level of complexity are classification tasks that involve more intricate patterns and greater levels of variability such as discrimination between different wing-beat patterns associated with different species of insect prey [133] and natural objects [117] that still have deterministic shapes but exhibit greater complexity and variability than the geometric primitives commonly used in classification experiments. A different class of target identification problems known to be solved by bats involves target responses without discernible deterministic patterns such as echoes from random textures [31] and signals that are realizations of random processes with simple parametric structures and differences [44]. However, none of the target classification experiments carried out with bats so far has come anywhere close to the level of complexity and variability that can be expected to be found in echoes from natural environments such as forests [86]. Recognizing a location based on echoes from natural targets has been previously attempted using templates estimated from recordings obtained for different sonar positions and orientations [132]. This approach was aimed at accomplishing short-range landmark recognition, i.e., over distances less than one meter. By comparison, certain bat species have been shown to navigate over round-trip distances of about 100 km in a single night [77] and hence must be able to recognize sites or habitat types over much larger distances. Furthermore, a neural substrate for spatial memory over a distance of 200 meters has been demonstrated in bats [30].

Navigation over large distances based on echoes from natural landmarks could pose a particularly tough challenge, because such echoes are typically superpositions of many scattering facets such as leaves in foliage. In such situations, even small changes in position and orientation of the sonar relative to facets in the targets result in alterations to these components and their relative weights in the echoes. This makes the waveforms of individual echoes exceedingly hard to predict. As a result of the unpredictable and irreproducible nature, a traditional correlation analysis of natural foliage echoes did not pick up any common patterns in echo waveforms recorded from forests beyond the pulse that elicited them [14]. These findings do not bode well for traditional pattern recognition methods that depend on deterministic templates and linear dependencies that can be picked up by correlation measures. For a bat or an autonomous drone, finding the exact location where a particular echo signature may be reproduced after many kilometers of flight is likely either impossible or at least highly impractical.

Deep neural networks (DNN, [71]) have been used extensively to emulate acoustic pattern recognition capabilities in humans, especially with respect to speech [25]. Similarly, DNNs such as convolutional neural networks (CNN, [66, 67]) have been used previously to classify sonar targets, such as sphere objects with different diameter [26] or perform binary obstacle classifications ("plant/no plant" [29]) based on echo spectrograms. However, to the best of our knowledge, DNNs have not been used to investigate the problem of landmark recognition over long distances based on biosonar or biomimetic sonar echoes. Nevertheless, the proven ability of DNNs to pick up patterns in data that have eluded traditional approaches to pattern recognition [82, 97] makes it worthwhile to investigate whether the power of these methods can also shed light on the problem of large-scale landmark-based navigation based on biosonar.

In the current work, large-scale identification of landmarks has been attempted based on a

set of echoes that had been collected with a biomimetic sonar system across vegetated field sites distributed over an area with a radius of about 20 km. If successful, this experiment would demonstrate that bats – and man-made systems mimicking them – could have access to location information based on a single echo from the respective site. This could not only lead to new hypotheses for the sensory biology of bats, but could also provide new navigation paradigms for autonomous systems that have to navigate in natural environments without GPS.

3.4 Methods

3.4.1 Biomimetic robot

Echo data was collected with a biomimetic sonar head (Fig. 3.1b,c) that was modeled after greater horseshoe bats (*Rhinolophus ferrumequinum*). In this system, ultrasonic pulses were generated by two electrostatic ultrasonic loudspeakers (600 Series, SensComp Inc., Livonia, MI, USA, Fig. 3.1a) with a peak response frequency around 50 kHz and a −6 dB passband that covered a frequency range from approximately 40 to 80 kHz. The generated pulses were conveyed via waveguides (conical horns, length 7.6 cm) that facilitated the transition from the loudspeaker (diameter 3.8 cm) to the outlets (diameter 3 mm) which represented the nostrils of the bat. The waveguide outlets were surrounded by a concave silicone baffle (height 2.1 cm, width 1.3 cm) which mimicked the noseleaf of horseshoe bats and included all major anatomical features (anterior leaf, sella, and lancet, [22]) found in these animals.

The returning echoes were received through silicone baffles designed to mimic the pinnae of horseshoe bats. These baffles were coupled to capacitive MEMS microphones (Monomic, Dodotronic, Rome, Italy, approximately flat frequency response from 2 to 125 kHz, one per

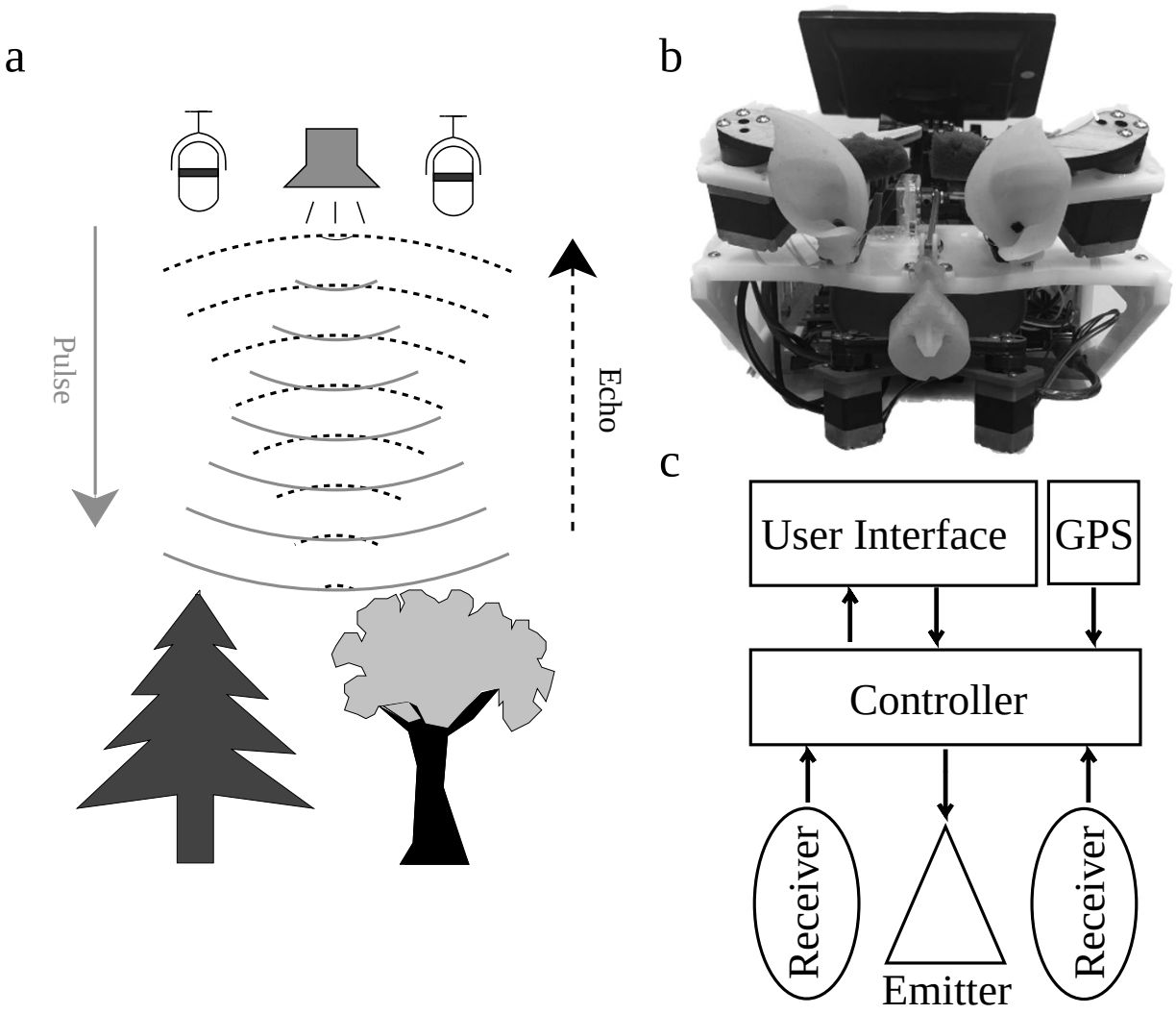

Figure 3.1: **Experimental setup: (a)** Sonar data collection paradigm for foliage echoes, **(b)** Biomimetic sonar head used for the data collection, **(c)** Block diagram with the main functional components the sonar head.

pinna) via short (length 2 mm) cylindrical pipes that served as "artificial ear canals". The sizes of the noseleaf (height 2.1 cm) and the pinnae (height 5.8 cm) were scaled up by a factor of approximately two relative to the respective structures in greater horseshoe bats.

Pulse generation and echo digitization were handled by an onboard microcontroller (Arduino Due, Arduino, Boston, MA, USA, clock frequency 84 MHz). The microcontroller digital-to-

analog conversion of the pulses was conducted with a sampling rate of 1.6 MHz and 12 bit resolution that were fixed by the hardware. For echo digitization, the sampling rate of each microphone channel was 400 kHz with a resolution of 16 bit. Since the input bandwidth was limited to 80 kHz by the emission (no other ultrasonic sound sources were present during the experiments), no anti-aliasing was necessary. A GPS module (Adafruit Ultimate GPS, Breakout 3, New York, NY, USA) recorded the positions during echo collection with an absolute error of less than 1.8 m [2]. In addition, videos were recorded along each track (HERO 3, GoPro San Mateo, CA, USA) for documentation purposes. An onboard computer (Raspberry Pi 3 Model B+, RS Components, Cambridge, UK) provided top-level control of the experiments as well as a user interface for the experimenter. The entire system was powered by a DC battery (Lithium Polymer RC Battery, 22.2 V, 4.5 Ah, Floureon, Nantou, Taiwan).

The emitted ultrasonic pulses were designed to mimic the constant frequency – frequency modulated (CF-FM) biosonar pulses that are typical of hipposiderid and rhinolophid bats [56]. The pulses consisted of a CF component that was located at 45 kHz in frequency which placed it near the resonance frequency of the transducers as well as at a bit more than half that of greater horseshoe bats (*sim* 82 kHz, [133]). Together with the scaling of the noseleaf and pinna, this resulted in ratio of noseleaf/pinna size to wavelength that was similar to that of horseshoe bats [22].

The CF component of the pulses had a duration of 8 ms (Fig. 4.1C, black solid box) and was followed by an FM component that swept down from the CF frequency to 30 kHz over a duration of 7 ms (Fig. 4.1C, white solid box). Each ultrasonic recordings was started with the beginning of the emission and lasted for a total duration of 25 ms. This means each recording contained the entire duration of the pulse (15 ms) with any overlapping echoes as well as a 10 ms-segment consisting only of echo recordings.

To establish whether CF or FM signals differed in their ability to support echo classification, these component were separated in frequency range and in time. For example, the CF signal frequency were designed at 45 kHz, and the emission time was from 0 to 8 ms; the FM signal frequency were swiped between 30 to 40 kHz, and the emission time was from 8 to 15 ms (Fig. 4.1C).

3.4.2 Field sites and data collection

The echo data was collected from vegetated areas around the campus. Ten different field sites (Fig. 4.1B) were selected to represent different regional habitat types (Fig. 4.1A). The largest distance between any two of these 10 sites was 40 km while the smallest distance was 2 km. Each site was labeled with a letter code ('a' to 'j'). At each site, data was collected along two separate tracks that were labeled as '1' and '2' for each field site. The tracks belonging to one site were selected to be qualitatively similar in terms of their vegetation cover and to allow for data collection along a path of at least 200 m in length. All tracks were located in flat terrain. During data collection, the sonar head was hand-carried along the track (Fig. 3.1b) with an approximately constant walking speed of around 0.2 m/s. The sonar head collected echoes with a rate of approximately three per second. During the entire field data acquisition, the sonar head was moved slowly up-down and left-right to scan the foliage while being moved along the walking tracks. The distance between the sonar head and the foliage was kept in a range from approximately 1 to 1.5 m (Fig. 3.1a) to minimize the overlap between the emitted signal and the returning echoes while maintaining a favorable signal-to-noise ratio for the echoes. All data collection sites have a similar distribution of the trees in terms of density and roughness, which reduced the variance of the data cased by the different targets, and also allowed a safe data collection process.

3.4.3 Data sets

In total, 41,000 echoes were collected across all 10 field sites. For each track, about 2,100 echoes (minimum 1,710 echoes) were collected, i.e., the minimum number of echoes collected per field site (two tracks each) was 3,500. Echo signals with evidence of amplitude clipping defined as saturated envelope amplitudes for at least 20 consecutive samples were eliminated from the data set. Based on this criterion, less than 5% of the original echo recordings were removed from the data set.

3.4.4 Signal processing

To reduce out-of-band noise, the recorded echo signals (Fig. 4.1C) were pre-processed with a bandpass filter that covered the entire band of the employed pulses (-3 dB corner frequencies at 25 and 50 kHz, finite impulse response (FIR) filter design based on a 256-point Hamming window. For all further processing, only the time segment of the recorded signals that contained solely echoes, i.e., a 10 ms time window starting from 15 to the end of the recording at 25 ms was retained.

An average signal-to-noise ratio (SNR) was estimated based on 1,000 randomly selected echo recordings for the band covered by the FM pulses. For each selected echo, the root-mean-square (RMS) value of the noise was estimated from the first millisecond of the recording, i.e., before the start of any echoes. Separate RMS values of the signal were obtained from the first and last millisecond of the FM echo respectively, i.e., right after the end of the pulse and at the end of the recording.

To carry out echo classification based on energy, the root-mean-square (RMS) value of each echo spectrogram was calculated, and each track was represented by the overall mean and standard deviation of the RMS values across all echoes . The deep-learning classifiers were

presented with spectrograms that were either normalized by their respective RMS value or were left in their original, i.e., unnormalized, condition.

The echo segments used in this analysis were defined differently for the CF and FM components to account for their different locations in time and frequency. The CF echoes were taken from the time window from 8 to 18 ms since the start of the pulse and the frequency band from 42 to 47 kHz whereas the FM echoes were taken from the time window 15 to 25 ms and the frequency band from 27 to 42 kHz (Fig. 4.1C). Hence, the echo segments for both components each covered a duration of 10 ms and maximum frequency range respectively.

Correlation coefficients between the recorded echo waveforms were calculated to make sure the echoes contained invariant information during the site's data collection. A confusion matrix used to show the correlation between each pair of echoes along the foliage trial recordings, 50 continued recording echo examples correlation coefficient (mean and standard deviation) were calculated along one of the trail, about 10 m. As a reference for the experimental correlation data, simulated echoes were used to calculate the correlation coefficient for random independently distributed impulse responses that were convolved with the same pulse template that was used to generate the foliage echoes in the field experiments [14]. The waveforms of the simulated echoes had the same assumed sampling rate (400 kHz) and duration (10 ms) as the physical echo data. The impulse responses were generated from a superposition of delta pulses that were taken to represent individual reflection in the echoes, e.g., from an individual leaf. In time, these delta pulses were distributed by virtue of a 200-point Poisson process. The amplitude of each delta impulse was weighted with a scalar value that was drawn from a zero-mean Gaussian distribution. Spectrograms with different trade-offs between time and frequency resolutions were used to evaluate the contribution of time and frequency domain signal features to the classification performance. For this investigation, the CF and FM echoes were represented by six different time resolutions, ranging

from 1/5 ms to 10 ms where the upper limit equals the length of the analyzed echo segments. Since the time resolution of a spectrogram is the inverse of its frequency resolution [108], the corresponding frequency resolutions ranged from 100 Hz to 5 kHz where the upper limit equals the analyzed frequency range. The spectrograms were represented by images that ranged from 15 to 25 ms pixels along the time dimension and from 25 to 50 kHz pixels along the frequency dimension. All spectrograms were computed using a Hamming window with 50% overlap. For each pixel of the spectrogram, the power spectral density in the respective time-frequency bin was represented by an eight-byte floating point number.

Combining the two different signal components (i.e., CF and FM) and the normalized and original conditions of the echo amplitudes, resulted in four different input types with different trade-offs in the time-frequency resolution being evaluated for all of them. For each of the data sets resulting from these signal transformations, a DNN was trained individually. While each DNN was trained with different inputs, they all used an identical architecture.

3.4.5 Echo classification

As a reference, the echoes were classified according to their location based solely on their respective energy content using maximum-likelihood estimation. For this purpose, each echo was represented by an RMS spectrogram amplitude value of the time-frequency region that contained (dashed boxes, Fig. 4.1C). The means and standard deviations of the RMS values from each site were used as parameters for a Gaussian model that described each site's spectrogram RMS distribution. The maximum likelihood criterion was then used to assign each echo to a location based on its RMS spectrogram amplitude value. In addition, a t-test (with Bonferroni correction) was used to determine whether the distributions of echo energy differed between the ten field sites or the 20 tracks.

The deep-learning approach used to classify echoes from different landmarks and different tracks was based on a convolutional neural network (CNN, [48, 64, 66]) operating on the echo spectrograms described above. The architecture of this network was inspired by ResNet50 [48]. The network was implemented in TensorFlow [1] via the Keras [19] interface library (version 2.4.3) and the Python programming language (version 3.7). In this implementation, a modified network architecture (Fig. 3.3) was used where the convolution kernel size was reduced from 7×7 to 3×3 to account to the smaller input data size of the spectrograms compared with the high-resolution images processed by ResNet50. Similarly, the kernel size of the max-pooling layers was reduced from 3×3 to 2×2 kernel. The central portion of the network (Fig. 3.3a) was made up of four groups each containing a convolution block (Fig. 3.3b) followed by an identity block (Fig. 3.3c). The convolution and identity blocks contained three convolution layers each. All individual convolution layers were each followed by a rectified linear unit (ReLU) activation function [38]. An average pooling layer connected the final identity block of the network to a soft-max location estimate (Fig. 3.3a). In training the network, the Adam optimization algorithm was used to update the network weights and cross entropy was served as a measure for training loss. Network performance was found to converge within 20 epochs. Training the network for 20 epochs took about 12 hours on an Intel Core i5-7200U CPU with a clock rate of 2.5 GHz and 8 GB of RAM.

Because of the size of the input data was three times smaller for the CF than the FM spectrograms along frequency dimension, the kernel size was reduced from 3×3 for FM to 2×2 for CF. In addition, the maximum pooling layer was not included because of the smaller size of the CF spectrograms.

The data set used for deep-learning classification consisted of 1,700 echoes from each track that were randomly divided into subsets for training (1,500 echoes per track) and testing (200 echoes per track). A ten-fold cross-validation was used to reduce the error in the prediction

error estimate, where the average prediction accuracy and the associated standard deviation were calculated across ten repetitions.

3.5 Results

The recorded echo data showed little correlation between different echo samples. The average correlation coefficient between different echoes was estimated at a mean of 0.13 with a standard deviation of 0.04 (Fig. 3.4). The simulated echo data that was based on statistically independent random impulse responses yielded similar values (mean correlation coefficient 0.16 ±0.11 standard deviation). All recorded echoes were clearly distinguishable from recording noise. The signal-to-noise ratio (SNR) at the beginning of the echoes was 23 ± 5.3 dB (mean ± standard deviation). At the end of the echoes, it was reduced to 15.2 ± 4.3 dB (N=1,000 echoes). The recorded echoes were found to differ significantly in their energy across the different sites (43 out of 45 pairwise site comparisons had p-values less than 10^{-5}, one less than 0.05, and one greater than 0.1, t-test with Bonferroni correction) as well as across the different tracks of all sites (185 out of 190 pairwise track comparisons with p-values less than 0.001, three less than 0.01, and the remaining two pairings showed no significant differences, t-test with Bonferroni correction, Fig. 3.5). A maximum-likelihood estimator based on the echo energy distributions associated with the different sites and tracks was able to distinguish fairly well between two site groups (sites 1 to 6 versus sites 7 to 10 and the associated tracks, Fig. 3.6a,d), but performed poorly on making distinctions within these groups. This matched the pairwise similarities between the distributions of the echo energy between the respective sites and tracks where a t-test showed the amplitude distributions between sites 1 to 6 and 7 to 10 to differ with a p-value of less than 10^{-4}. For comparisons within each of these site groups, the p-values were much larger. For example, the p-value

for t-test between the energy values for track 2 of site d and track 2 of site f was 0.955. Overall, 35.2% of the site classifications and 19.5% of the track classifications performed by the energy-based classifier were correct. While a poor performance, these results were still well above the respective chance level performances of 10% and 5% respectively.

The DNN classifier was found to be much more accurate than the energy-based classifier (Fig. 3.6a,b) for all tested echo types. However, the classifier performance did depend strongly on the type of the input signals used (Table. 3.1). Classification performance based on the FM signals was higher than for the CF signals and for both signal types, raw signals supported better classifier performance than amplitude-normalized signals. The best performance was hence achieved for raw FM-signals with an accuracy of 99.6% (standard deviation 0.7%, N=10 cross-validation runs) and the worst performance was achieved for amplitude-normalized CF-signals with an accuracy of 83.3% (standard deviation 0.5%, N=10 cross-validation runs). The differences between the classification accuracy values achieved for the different input signal types were all found to be highly significant (all p-values less than 0.001, t-test with Bonferroni correction). In addition, the type of the pulse signals used as input affected the speed of learning (Fig. 3.7, Table. 3.1). The differences in the learning speeds that were achieved with the different signal types matched the ordering obtained in terms of classification accuracy with the fasted learning occurring for raw FM signals and the slowest for amplitude-normalized CF signals.

The time-frequency resolution of the echo spectrogram was found to exert an influence on the target classification performance of the CNN (Fig. 3.8). For the FM echoes, an optimum time-frequency resolution (1 ms and 1 kHz) was found that yielded a classification accuracy of 95% (Fig. 3.8a). Above and below this resolution, the performance decreased with the lowest classification accuracy for FM echoes (71%) being recorded for a time-frequency resolution of 10 ms and 100 Hz, i.e., the lowest time resolution tested. The dependence

of classification performance on a time-frequency resolution for the CF-echoes (Fig. 3.8b) was qualitatively different from the situation in the FM-echoes. In the latter, classification performance showed a monotonic decay with decreasing time resolution, i.e., the best performance (87% correct) was achieved for the highest time resolution (1/5 ms) and the worst performance (40% correct) occurred with the lowest time resolution (10 ms).

3.6 Discussion

The results presented here demonstrate that bat biosonar and its technical mimics can produce sensory information that could support landmark-based navigation on a large scale, i.e., tens of kilometers. In situations where the classification results obtained here are representative, it should be possible for a bat or an autonomous system equipped with biomimetic sonar to determine its location based on a single recorded echo. This was not necessarily an expected outcome since the echo waveforms from all field sites were found to be profoundly unpredictable in nature and hence no location-specific patterns were obvious in the echo waveforms to the human observer. The success achieved with the DNN classifiers indicates that some patterns must exist in the echoes albeit in a way that is not accessible to the human observer. The interpretation of those information might have a critical influence to understand the bat's sonar system, this insight could have important implications for the understanding of bat biology since there are a number of bat species [52, 83, 129] that are known to travel over distances as large as – or even larger – than the distances between the field sites of the present study. Similarly, the current results could provide a sensory foundation for autonomous navigation in natural environments by man-made systems without GPS. Compare with the traditional vision-based which has the deterministic template matching method, the sonar can use for an even more complex environment with smaller

data sizes, especially camera or lidar/radar-based navigation. A smaller data size would be more applaudable in the application which needs to navigate in real-time such as autonomous vehicle driving. Here we also achieved higher landmark classification accuracy with a more challenging environment compared to some previous work. [29, 65, 121]. Finally, it is noteworthy that the echoes were collected with a hand-carried system that did not include any controls for directing the sonar beam. Hence, bats or man-made systems could utilize the same location information without the need for intricate control.

The results demonstrate that a – very modest, but above chance-level – location identification performance can be achieved based simply on echo energy. This modest performance is not surprising since echo energy can be linked to features such as overall foliage density and leave size [80, 86, 141]. However, it is also not a surprise that this approach does not yield reliable location information since there is substantial overlap between the distributions from the different sites and similar overall echo amplitudes could be achieved in different ways, e.g., by virtue of a high leaf density or large leaf sizes in case of high echo amplitude. The low reliability of the energy approach to natural landmark identification would almost certainly require compensatory measures such as integration with other information sources. This additional information can come from time-frequency structure of the echoes as demonstrated by the present results from the CNN classifier. The comparison of the classification performance achieved for different time-frequency resolutions in the underlying spectrogram signal representation casts some light on the nature of the features that have been used by the DNN classifiers. For echoes elicited by CF-pulses, all information must be encoded in the time domain since these narrowband signals leave very little room for information encoding along the frequency axis. Hence, it is not a surprise that classification performance decreased as the time resolution was reduced. The results for the FM-echoes indicate a mixed encoding of location information in the time and frequency domain since

an optimum weighting between the resolutions along the two dimensions was found to exist.

While the current result indicates that a DNN approach can find patterns in apparently unpredictable biosonar echoes from natural foliages that can convey location information, more work is needed to thoroughly evaluate the nature of these features and their potential utility for navigation in bats or man-made systems. To better understand the nature of the encoded location information, future work should hence try to uncover more information about the echo features that are being used by the DNN. The current results obtained by varying the time-frequency resolution of the echo spectrograms indicate that the location information is encoded in the time as well as in the frequency dimension of the signals. In a recent study of passageway finding based on vegetation echoes [135], it was possible to use a transparent AI method (class activation mapping [135] to narrow down the encoding of the relevant sensory information to the rising flank of the echo. It remains to be seen if a similar approach could be used for the location-finding problem studied here. For a gap in vegetation, the critical geometrical feature is the rim of the gap which is spatially localized and could hence be hypothesized to give rise to echo features that are likewise localized in the spectrogram. For the location-related echo features, no a priori reasoning in favor of such a localization is obvious. To assess the utility of the echo features that carry location information, research should be conducted to establish the spatial resolution that can be achieved with these features and also evaluate the existence of ambiguities that could arise from similar habitats occurring within the range of a bat or an autonomous drone.

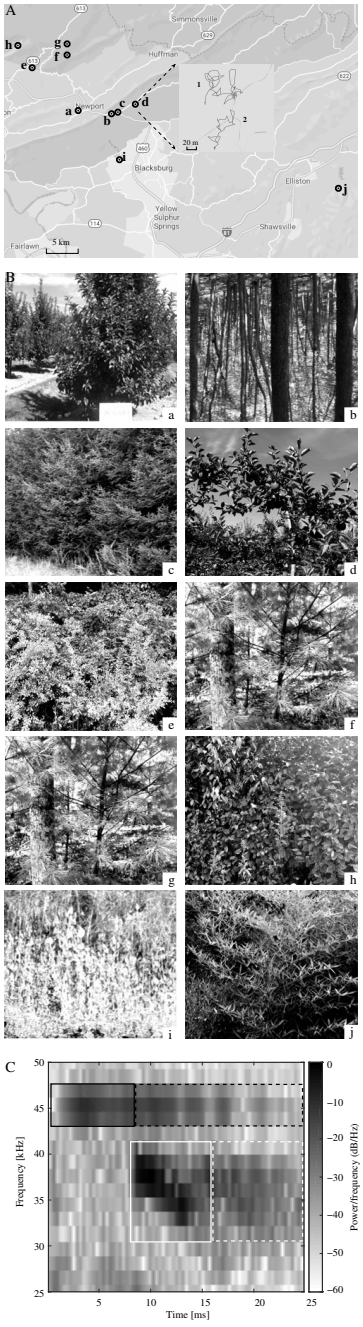

Figure 3.2: **Data collection sites and example echo: (A)** Location of the field sites in the area that surrounds the campus. The sites are indicated by dots labeled with the letter code of the respective site in the main map. At each site, echoes were sampled along two tracks (shown in the inset for site d as an example), **(B)** Example photos of vegetation at the 10 different field sites of the study (alphabetical labels correspond to **(A)**), **(C)** Example recording of a pulse-echo pair. The emitted pulse consisted of a CF part (box with black solid lines) with 45 kHz carrier frequency that occupied the first 8 ms of the pulse and an FM part with a linear frequency modulation from 40 to 30 kHz that occupied the last 7 ms (box with white solid lines). The echoes components that follow each pulse component are enclosed in boxes marked with dashed lines and with the same line color as the respective pulse components.

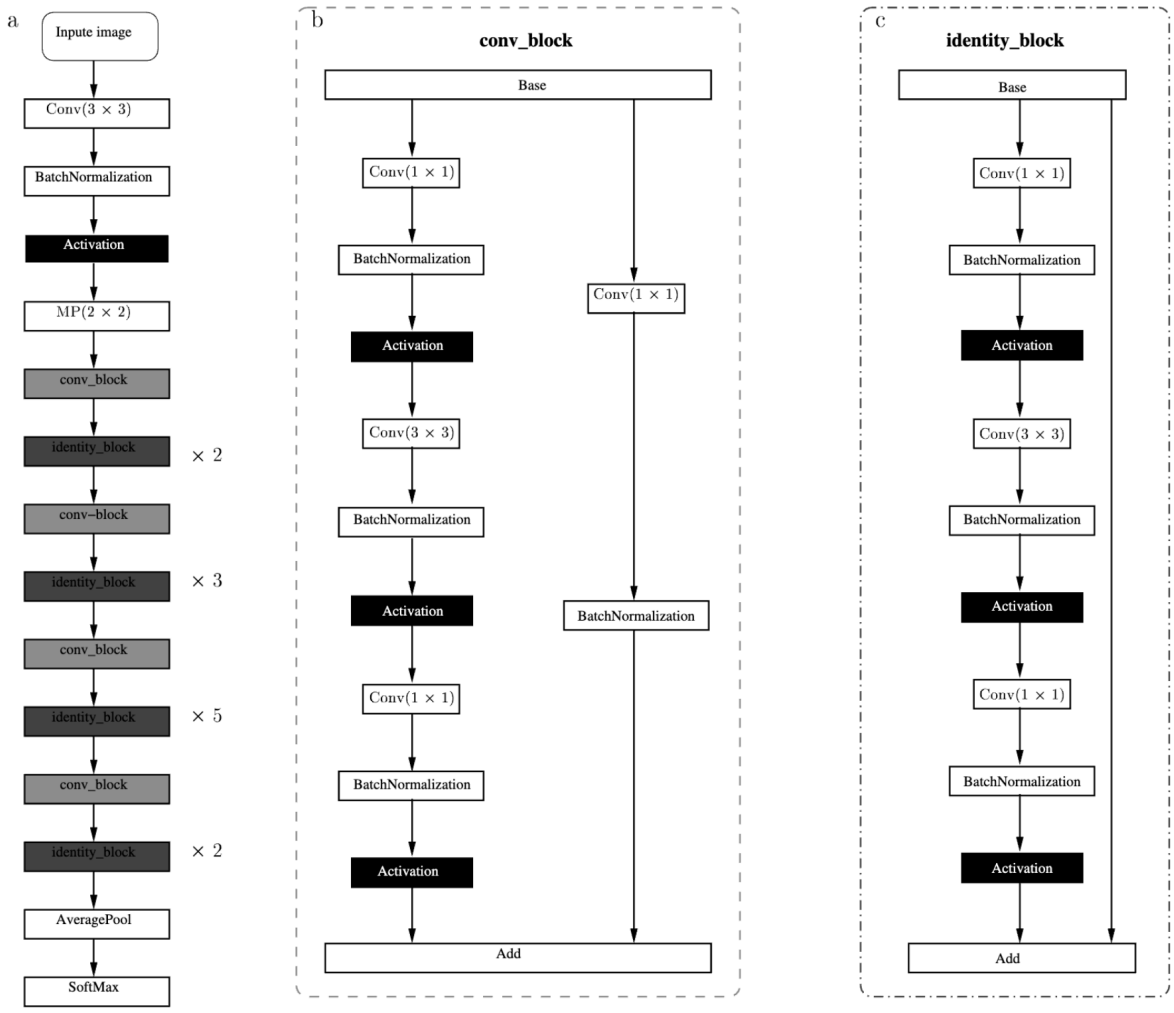

Figure 3.3: **Neural-network architecture used for landmark identification**. The architecture of this network has been inspired by ResNet50. At the center of the overall architecture (a) is a repeated sequence of convolution (b) and identity blocks (c). The 1×1 convolution layers in these blocks perform cross-channel pooling to reduce the number of channels from 256 to 64.

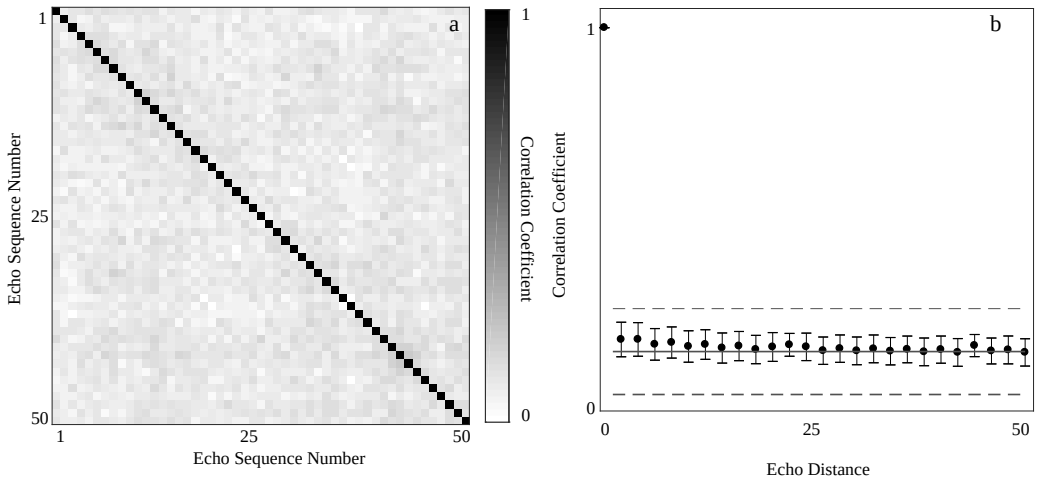

Figure 3.4: **Cross-correlation structure of the recorded echoes**. a) Correlation matrix of 50 consecutively recorded echoes. b) Correlation as a function of distance (in terms of position in the recording sequence) between the echoes (mean and standard deviation across all echo pairs with the respective distance). Black filled circles are mean values computed from the field echo recordings; the error bars indicate the standard deviations. The gray solid line represents the mean value of the simulations and the dashed lines the range spanned by the respective standard deviation.

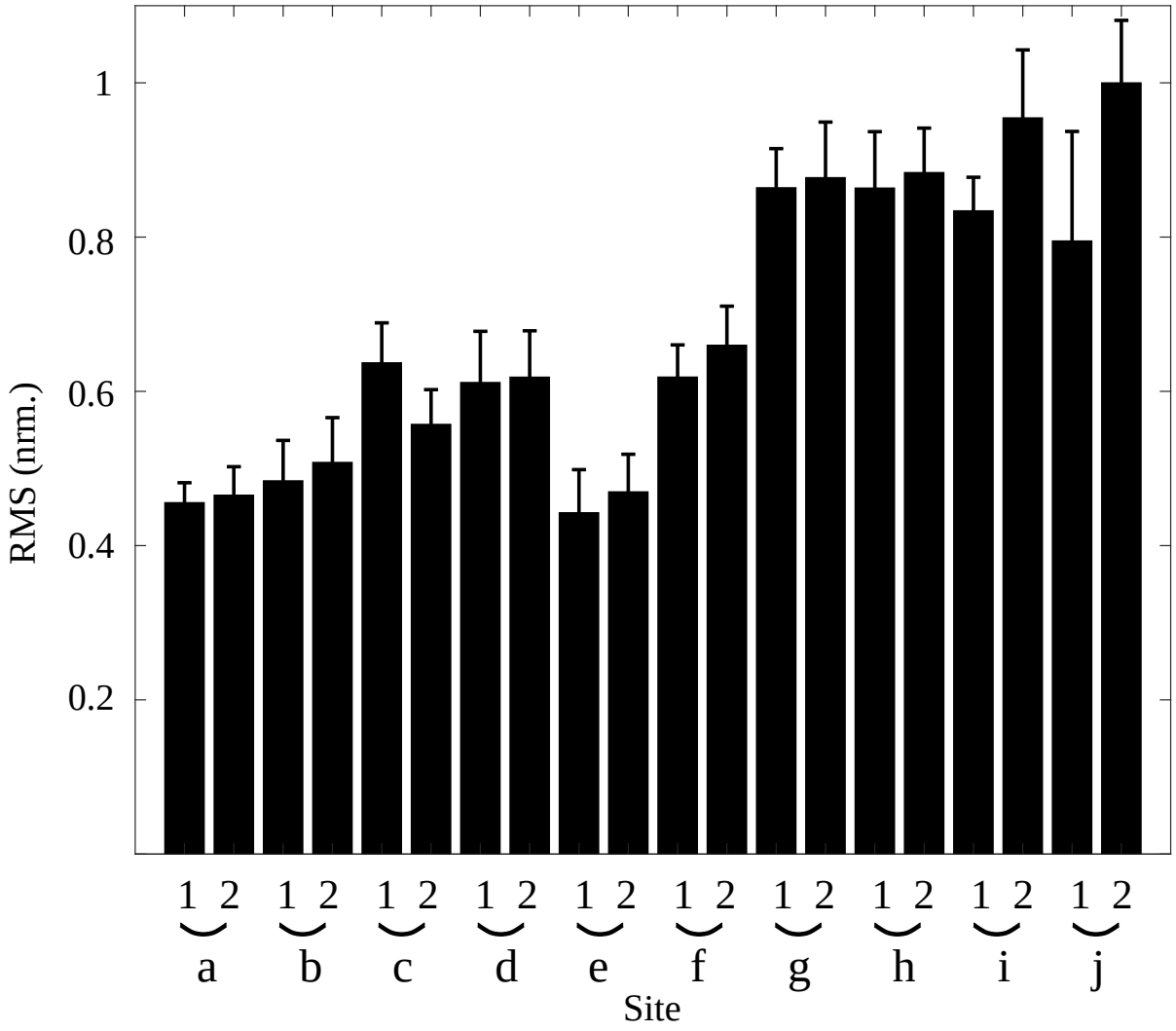

Figure 3.5: **Distribution of echo energy across the different recording sites and tracks**. The height of each bar indicates the average energy for each track/site and the error bars the respective standard deviation.

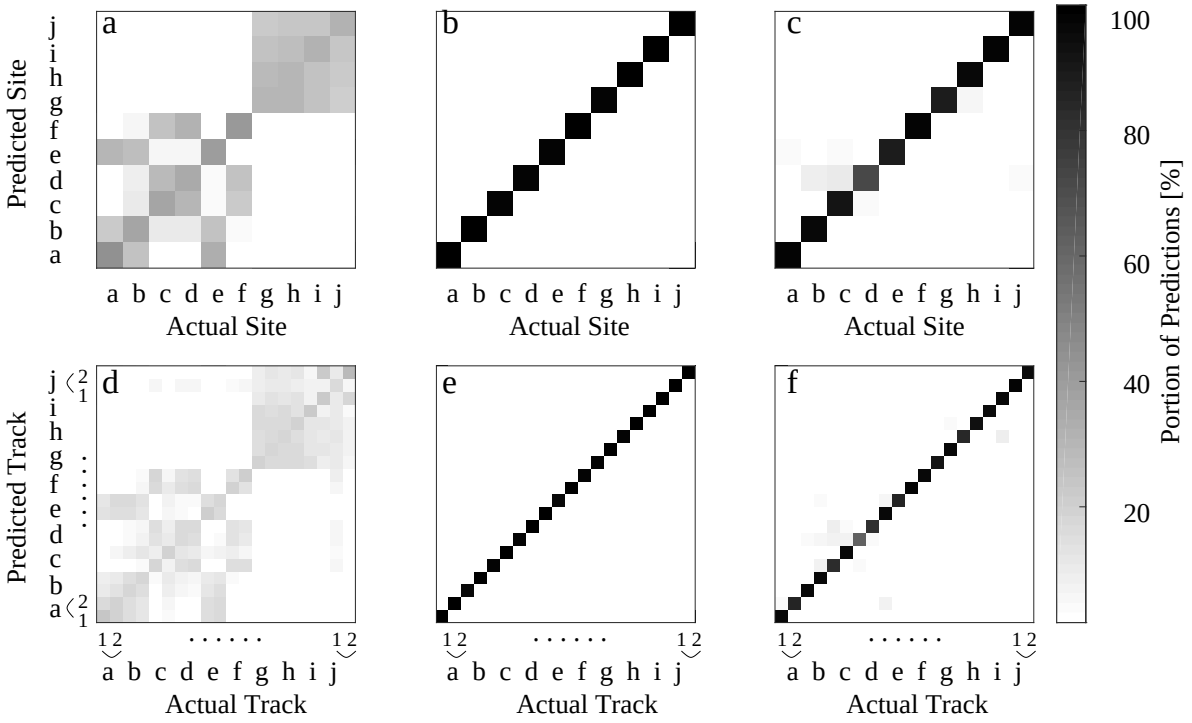

Figure 3.6: **Accuracy of the location predictions**. Confusion matrices for the classification of different sites (top row) and different tracks (bottom row). The classifier used were energy-based maximum-likelihood classifier (first column, a and d) and CNN classifiers based on raw spectrograms (central column, b and e) or amplitude-normalized spectrograms (third column, c and f). FM part of the echo's spectrograms used to generate confusion matrices.

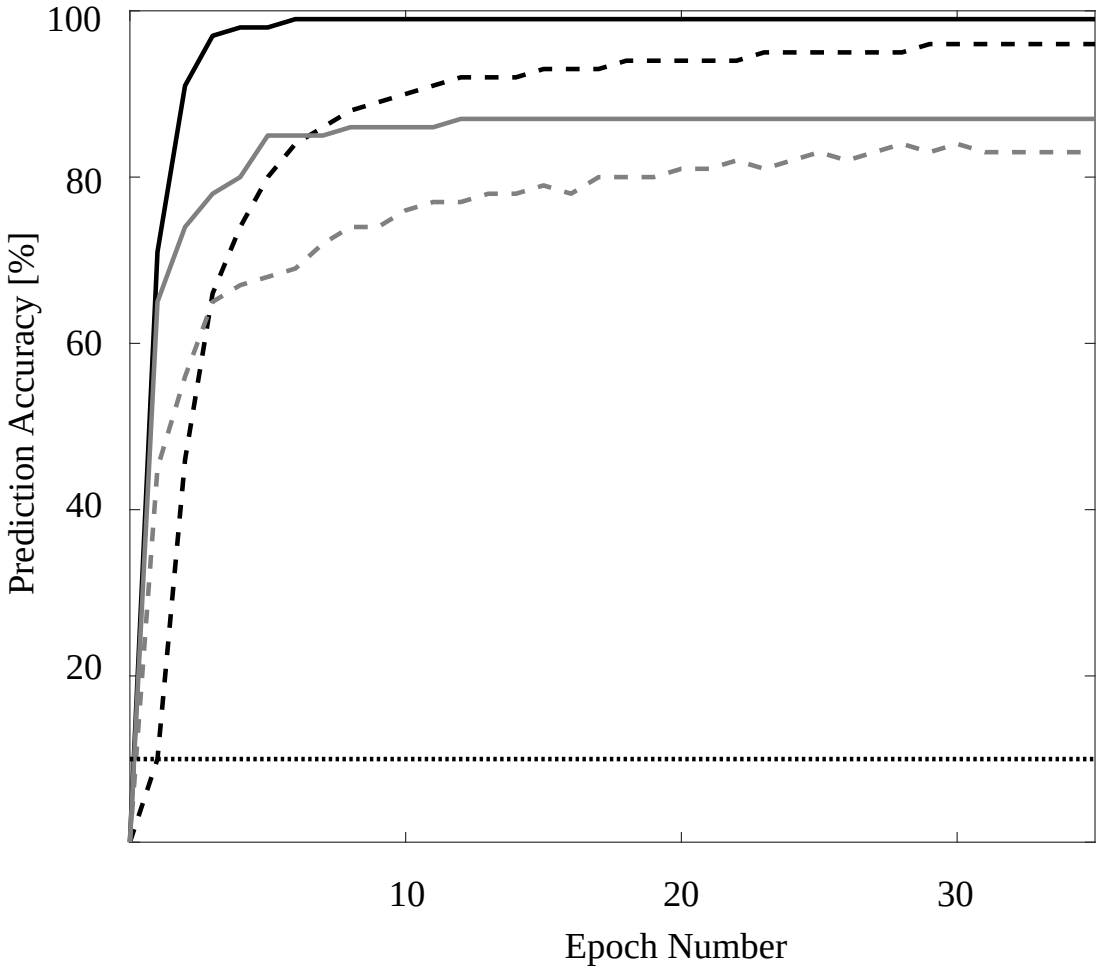

Figure 3.7: **Classifier training results for different input data**. Validation accuracy for site classification based on raw FM-echoes (solid black line), raw CF-echoes (solid gray line), normalized FM-echoes (dashed black line), and normalized CF-echoes (dashed gray line). The dash-dotted line denotes the chance level for site classification (10%).

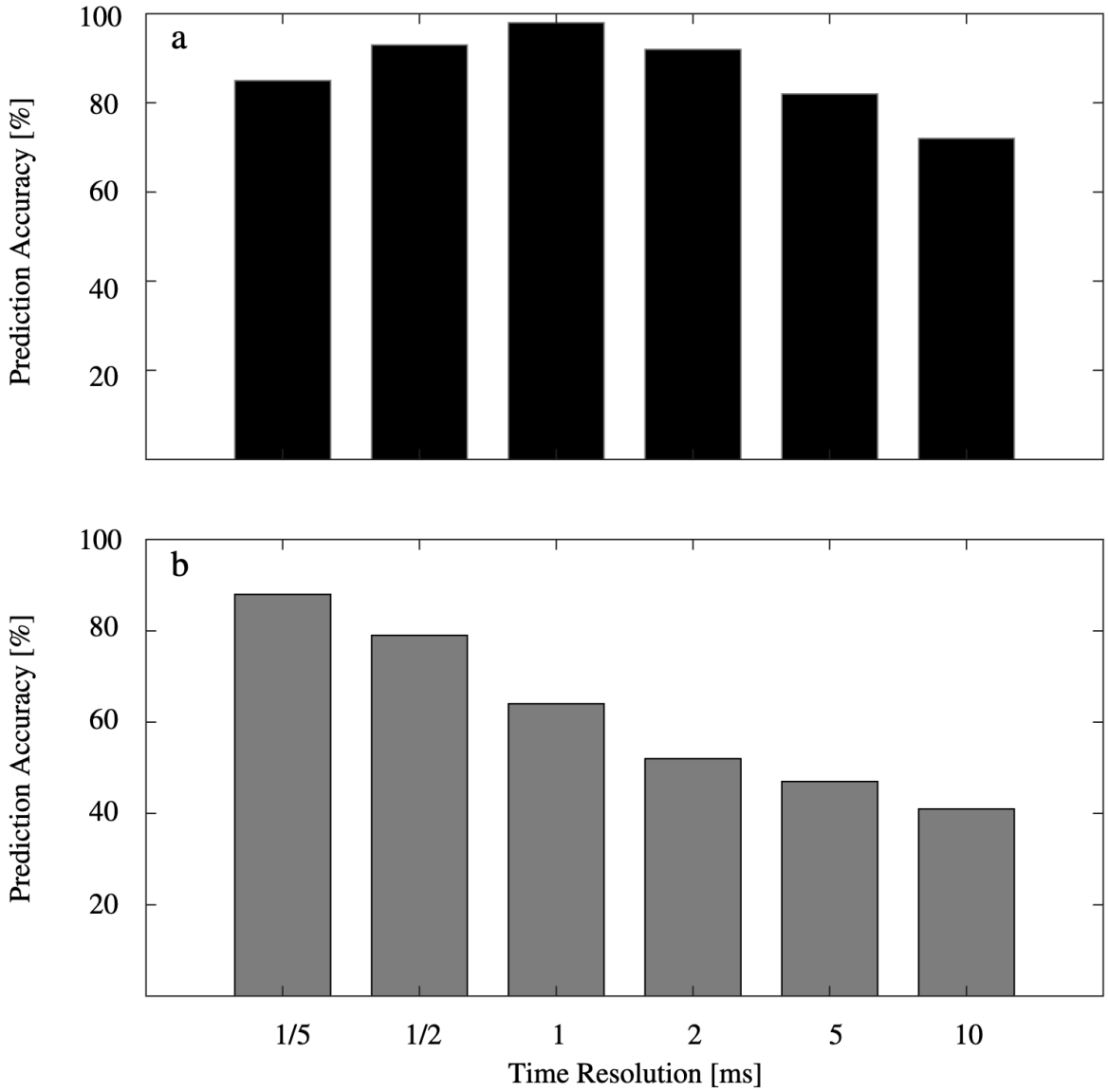

Figure 3.8: **Landmark identification accuracy as function of time-frequency resolution in the spectrogram representation of the echoes**. Classification based on a) echoes elicited by FM pulses and b) echoes elicited by CF pulses.

Table 3.1: Prediction accuracy and learning time for the different input signal types tested.

Signal	**Sites (%)**	**Tracks (%)**	**Time (nrm.)**
FM (raw)	99.6 ± 0.7	94.7 ± 0.5	1
CF (raw)	87.1 ± 0.5	81.2 ± 0.6	1.7
FM (nrm.)	97.2 ± 0.7	91.2 ± 0.5	4.1
CF (nrm.)	83.3 ± 0.5	78.6 ± 0.5	4.6

Table 3.2: GPS coordinates of the start points for all data collection sites.

Site	**Latitude (°)**	**Longitude (°)**
a	37.2833	-80.5208
b	37.2813	-80.4924
c	37.2884	-80.4722
d	37.2883	-80.4610
e	37.3266	-80.5771
f	37.3400	-80.5450
g	37.3581	-80.5345
h	37.3547	-80.5976
i	37.2684	-80.4723
j	37.2141	-80.1908

Chapter 4

Small-scale acoustic granularity of a natural environment

4.1 Title

Deep-learning exploration of the acoustic granularity of bat habitats

4.2 Introduction

The ability to navigate autonomously in natural, vegetated environments could enable a multitude of technical applications that include search and rescue [17, 59, 73], precision agriculture [8, 45], as well as environmental surveillance [120]. A fundamental ability that is needed for many of the navigation challenges posed by outdoor applications of autonomous systems is identifying a location, e.g., for the purpose of building a map of the environment where the system is supposed to operate.

The most common approach to identifying a given location is to utilize the Global Positioning System (GPS, [51]). However, GPS has its limitations: There are a number of environments where GPS is not available at all, e.g., under water [69, 125] or in caves and mines [10, 57]. In addition, GPS has been shown to be vulnerable to jamming or manipulation [41]. But even

under conditions where an unadulterated GPS signal could be accessed in principle, there can be circumstances that result in a substantially reduced accuracy, e.g., under dense foliage cover [78], in indoor environments [103], or in urban canyons with dense foliage cover [53].

An alternative to location identification with GPS is provided by vision-based landmark recognition [61, 100]. These approaches typically rely on the recognition of object shapes by virtue of matching deterministic templates. While it is easy to see how template matching would work for man-made landmarks such as buildings that have clearly recognizable shape patterns, plants in natural vegetation have complex shapes with a large amount of randomness [20] that may not be easily captured by a deterministic image template. Template matching based on three-dimensional optical recordings of an environment, e.g., using lidar [21, 37], are likely to suffer from the same problem. In addition, navigation based on laser scans requires acquiring, storing, and processing large amounts of data. A single lidar sensor, for example, can produce data rates of about $23\,\mathrm{Mbit/s}$ [7, 49]. The large computational cost and power consumption associated with handling such data rates may be difficult to satisfy in the context of small autonomous systems.

Echolocating bats are capable of navigation in a wide variety of natural habitats that include densely vegetated environments [36, 95]. In addition, bats have been shown to travel distances as large as 50 km in a single night and then return to their roosts in the morning [79, 106]. Hence, bats from these species must be able to create maps of their environments that allow them to find their ways to their feeding grounds and back to their roosts.

The ability of bats to navigate in densely vegetated environments based on biosonar is of particular interest because of the special nature of sonar echoes from natural vegetation [86, 141]. A typical foliage is composed of a multitude of leaves and other sound-reflecting elements. Each of these elements contributes to the received echoes based on its position, orientation,

and shape. Since all the specific values of the parameters that the determine the reflection from an individual leaf remain unknown, a foliage is best approximated as a stochastic array of reflectors [76, 86] that results in likewise unpredictable echo waveforms [86]. For sonar-based navigation, the implication of this randomness is that any sonar system will never see the same echo waveform again. Hence, conventional template-based methods for recognition of a location-specific pattern will not work. Despite these difficulties, the biosonar abilities of bats demonstrate that sonar-based solutions to the location-finding problem must exist and can be realized in a highly reliable and parsimonious fashion.

As a first step towards replicating the biosonar-based location-finding skills of bats, prior work by the authors [144] has shown that large-scale identification of different locations in natural environments based on single (15 ms) echoes is possible using deep learning methods based on time-frequency representations of the echoes. In this study, it was possible to not only identify ten different locations that were spaced within a 50-kilometer diameter, but also neighboring walking trails at the same habitat. The latter results have hinted at the possibility that biomimetic sonar echoes can convey location information with much finer resolution.

To examine the spatial resolution for locations that biomimetic sonar could provide in natural structure-rich habitats, the current work has characterized biomimetic echo data collected in a natural forest. This was done by combining GPS-based clustering of the echo-sampling locations and deep-learning classification of the corresponding echo waveforms to see which level of spatial granularity can be resolved based on the echoes.

4.3 Materials and methods

4.3.1 Biomimetic sonar

A biomimetic sonar head (Fig. 4.1A, [144]) was used to collect all foliage echoes for the present study. The system consisted of a sonar emitter and two receivers, a top-level controller, a set of microcontrollers used for digital-to-analog and analog-to-digital conversion, cameras to record images for documentation purposes, a GPS, as well as a power system to support all these devices.

For the sonar system proper, the main components were two electrostatic ultrasonic loudspeakers (diameter 3.8 cm, 600 Series, SensComp Inc., Livonia, MI, USA) with a peak response frequency located at ∼50 kHz and a -6 dB passband extending from ∼40 up to ∼80 kHz. Two capacitive MEMS microphones (Monomic, Dodotronic, Rome, Italy, -6 dB frequency range from ∼2 to ∼125 kHz) were used to receive the returning echoes. The loudspeaker and the microphones were each coupled to silicone baffles that were designed to mimic simplified, generic versions of the noseleaf and pinnae shapes seen in bat groups such as the horseshoe bats (Rhinolophidae) and the closely related Old World leaf-nosed bats (Hipposideridae). The silicone noseleaf had a height of 2.1 cm and a width of 1.3 cm. The silicone pinnae were about 5.8 cm tall, i.e., about two times of the length of a large hipposiderid bat's pinna (e.g., ∼30 to 35 mm in Pratt's roundleaf bat, *Hipposideros pratti* [142]). The noseleaf was coupled to two electrostatic loudspeakers (one per nostril) via conical-horn waveguides with a length of 7.6 cm. The diameter on the loudspeaker end of the waveguide matched that of the electrostatic loudspeaker (3.8 cm) and the diameter on the opposite end matched that of the nostrils outlet(diameter 3 mm). The pinnae were each coupled to the respective MEMS microphone via a short artificial ear canal (length 2 mm, diameter 3 mm).

The top-level computer of the sonar head (Raspberry Pi 3, Model B+, RS Components, Cambridge, UK) was used to control data collection via a user interface, store digital data from the microphones, and to visualize the incoming echoes in approximate real-time. A microcontroller (Arduino Due, Arduino, Somerville, MA, USA, clock frequency 84 MHz) was tasked with handling digital-to-analog and analog-to-digital conversion of the sonar pulses and the returning echoes respectively. The emitted pulses were converted to analog input signals for the ultrasonic transducer with a sampling rate of 1.6 MHz (the device maximum) and 12 bits amplitude resolution. The conversion of the microphone outputs to digital signal representations was conducted with a sampling rate of 400 kHz per channel and with an amplitude resolution of 16 bits.

The entire system was powered by a DC battery (Lithium Polymer RC Battery, 22.2 V, 4.5 Ah, Floureon, Nantou, Taiwan). A power supply unit (picoPSU-120-WI-25, 12-25V, 120 Watt, ATX Power Supply, Fremont, CA, USA) converted the battery output to the different DC voltages need by the onboard computer, the microcontroller, and the microphones (5 V each) and the user display (12 V). A separate high-voltage amplifier (APEX PA98, 1000 $^{V}/_{\mu s}$, maximum output voltage 450 V, Apex Microtechnology Corporation, Tucson, AZ, USA) converted the battery voltage to a 400 V bias voltage for the electrostatic transducers.

A GPS module (Adafruit Ultimate GPS, Breakout 3, Adafruit Industries, New York, NY, USA) recorded the geographic coordinates associated with each echo collection site. The GPS had a manufacturer-specified position error of 1.8 m [3]. To assess whether specification was valid for the recording conditions of the experiments reported here, 100 GPS data points were recorded along a straight paved road that was aligned with the edge of the vegetation in which the echoes were recorded. Under the assumption that the road edge can be described by a straight line, the root-mean-square error associated with fitting a line to this position data can be used as a measure for the accuracy of the GPS. This error was found to be about

5.6 m.

Finally, all data acquisition work was documented with a stereo-pair of two video cameras (HERO 3, GoPro, San Mateo, CA, USA) that were mounted on the biomimetic sonar head, faced in the same direction as the transducers, and recorded videos during echo data collection. None of the recorded videos were subjected to any form of quantitative analysis in this study.

4.3.2 Data collection

The echo data was collected in a natural wooded area (known as the "Stadium Woods", Fig. 4.1D) on the Virginia Tech campus in Blacksburg, Virginia. The size of the study site was approximately 180 m by 100 m. In general, the terrain of the study site was slightly rolling with a uniform vegetation cover of mature forest and substantial amounts of undergrowth (Fig. 4.1B). However, some small spots could not be walked safely due to the presence of local obstacles such as boulders or ravines and were hence left out of the data acquisition. An estimate of the area covered by the echo recordings has been derived from the measured GPS positions using a morphological closing operation which consists of a dilation followed by an erosion to determine a contiguous area covered by points [27]. The input data for this operation were the geographical positions associated with the echo recordings as determined by a reading from the GPS receiver. For the closing operation, each GPS location was represented by a point corresponding to a radius of 1.5 m in the real world. The structuring element of the morphological closing operation, i.e., the mask that is used to probe the image, was a circle with a radius that corresponded to 2.5 m.

During data collection, the biomimetic sonar was hand-carried (to approximate the natural variability in the flight paths that bats might take through the forest) at a distance of about 1

to 1.5 m from the nearest vegetation. While being moved through vegetation, the biomimetic sonar head was also rotated by hand in scanning motions that covered azimuth as well as elevation. As for the walking path, these motions were intended to approximate how a bat's biosonar might scan its surroundings. The sampling paths were traversed with a constant walking speed of about 0.2 $^{m}/_{s}$ while the data collection rate was about three echoes per second. The GPS module was used to record the location of the sonar (Fig. 4.1C) with an update rate of 0.2 Hz. A second-order polynomial fit was used to interpolate the GPS data for each instant of echo collection. The interpolated GPS locations were then attached to each recorded echo as the label for supervised learning.

The pulse waveform that was used to trigger the vegetation echoes was inspired by the biosonar pulses of constant frequency (CF) - frequency-modulated (FM) bats that combine narrow-band and frequency-modulated portions [56]. To mimic both of these signal components, the first 7 ms portion of the emitted pulses consisted of an FM chirp that sweeped from 55 kHz down to 45 kHz (Fig. 4.2, black solid box). The second part consisted of an CF signal centered at 60 kHz with a duration of 5 ms (Fig. 4.2, white solid box). Hence, the FM and CF components of the signals were hence not connected in frequency which does match the continuity in the time-frequency contours of bat biosonar pulses, but made separating the CF and FM components easier. The total length of each echo recording was 25 ms. Since each echo recording was started with the beginning of the respective pulse which including the two direct transmissions from the speaker, and also the reflections. The resulting echoes (Fig. 4.2, dashed box) were used as input for data classification.

4.3.3 Clustering of the GPS data

The study area was broken up into coherent patches based on the GPS coordinate data using a clustering approach (MiniBatch k-means, [96], implemented in the sklearn Python library [99]). The MiniBatch approach reduces the computation time from that of k-means by processing random batches of data with a fixed size small enough so that they can be stored in memory. Clustering is repeated in an iterative process where each iteration is based on a random pick of samples and the iteration stops once the convergence criterion is reached. Like in regular k-means, the objective of the algorithm is to minimize the within-cluster sum of squares. For each clustering attempt, the centroids of the clusters were randomly initialized three times, then the algorithm picked the best of the initialization as measured by the sum-of-square distances to their cluster center. To prevent premature stopping, the maximum consecutive number of mini-batches that do not yield an improvement on the figure of merit was set to 10. The size of the mini-batches was set to 256 sample points, and the maximum number of iterations to 100.

The silhouette value [109] was used as a measure of how coherent the generated spatial clusterings were. It measures how similar the elements within a cluster are to each other (cohesion) compared to how similar elements are across different clusters (separation). The silhouette value was calculated by the difference between the average within-cluster and cross-cluster distances normalized by the maximum value of these distances. Hence, the silhouette value ranges from -1 to $+1$, where a high value indicates that the elements are well matched to their respective clusters and poorly matched to neighboring clusters.

Using this approach, the GPS data were clustered into different numbers of spatial "patches" that ranged from a minimum of two up to a maximum of 100 in number (Fig. 4.3). For each desired number of patches, the clustering was repeated 100 times, and the result with the

highest silhouette value was retained for the subsequent analysis. In addition to the silhouette value, the distribution of sample locations across the different clusters was monitored to ensure that it was approximately even (Fig. 4.4).

4.3.4 Acoustical signal processing

In the first step of processing the echoes, a bandpass filter (finite impulse response, FIR, filter design based on a 256-point Hamming window with 50% overlap) was used to extract the frequency band occupied by the employed pulses from the echo recordings. For the FM pulses, this passband ranged from 40 to 58 kHz (-3 dB corner frequencies). The same bandpass filter design was used for the CF pulses, but the -3 dB passband covered the frequency range from 58 to 62 kHz in this case.

The echoes were converted into spectrogram representations (Hanning window with a length of 256 samples, FFT length 256 samples, 50% overlap). The spectrogram images served as input for classifying the echoes into the corresponding location patches. For the FM pulses, the spectrogram matrix size was 18×15 with the frequency range from 45 to 55 kHz. For CF pulses, it was 7×11, with the frequency range 58 to 62 kHz. For each pixel of the spectrogram, the power spectral density in the respective time-frequency bin was represented by an eight-byte floating-point number.

4.3.5 Deep-learning for location classification

The network used for patch classification was inspired by a state-of-the-art convolutional deep neural network for image classification (ResNet152, [47]). For the current work, the published ResNet152 architecture was modified by reducing the initial kernel size from 7×7 to 3×3 pixels, and the stride of the pooling layers from 3×3 to 2×2 pixels. The deep neural

network was implemented in TensorFlow (version 1.13.2, Google Brain Team, [1]) via the Keras interface library (version 2.3.1, F. Chollet, [19]) and the Python programming language (version 3.7).

The network started with the full input spectrogram size (18×15 pixels for FM, 7×11 pixels for CF , followed by an initial convolution layer, batch normalization [54], an activation function (rectified linear unit, ReLu, [38]), and a single maximum-pooling layer (2×2 pixels for FM only, [124]). These layers were followed by a parallel identity block and a convolution block. The identity block is the key feature of residual networks, it was used in a deep network for keeping the original feature information. Its output is appended to the output of the convolution block. The convolution block had three convolutional layers, each followed by batch normalization and ReLu. The ResNet152 repeated the parallel network 50 times in series. The final layers of the network consisted of an Average Pooling layer connected to a fully connected (fc1000) layer and the output Softmax [40] prediction layer.

Since bats have ready access to multiple echoes to determine their location, integrating the multiple inputs with the networks (Fig. 4.6) was important for the navigation task. The first network was based on ResNet152, to extract the features in the spectrogram. However, the outputs of the Softmax layers for between 2 and 10 echoes from the same patch (bats can easily emit 10 pulses within 1 s) were concatenated and fed into a multilayer perceptron (MLP) [107], which performed the aggregate supervised classification. The MLP included three layers, the first layer has a dimension of 512, then the next layer has 256, and the last layer was the SoftMax. The Arg max was used to do the patches classification based on the SoftMax layer.

The entire data set was separated so that 85% was used for training and the remaining 15% was used for testing. During the entire process, a five-fold cross-validation [122] was used for network evaluation. For the ResNet152 training and evaluation (Fig. 4.7), the prediction

accuracy curve for training and test data with a different number of epochs used to visualize the training process. Cross-entropy loss [92] was used as the loss function. The network generally converged after 20 epochs, and the overfitting problem was minimized after the network optimization.

A confusion matrix was used to compare the results from different numbers of echoes and different numbers of patches. For the former we have the results from 2 patches up to 100 patches and from for the latter 1 echo up to 10 echoes, as mentioned before. The X-axis shows the number of echoes, while Y-axis shows the number of patches, a grayscale color bar was used to represent different prediction accuracy values.

4.3.6 Saliency map for visualization and understanding the echo

To understand the inside of the network and interpret the spectrogram for feature extraction, saliency maps [5] were used to visualize the network. A saliency map is a way to measure the spatial support of a particular class in each input matrix. For example, the time or frequency domain intensity in a particular patch of the foliage might give a unique saliency map. The saliency map was built using gradients of the output over the input. This highlights the areas of the spectrograms which were relevant for the location classification. The last convolution layer before the SoftMax weight matrix was saved as the saliency, and the average saliency map from different patches was calculated (Fig. 4.8). The region was from the pure echo part (Fig. 4.2, dashed black box), and the average was from 2000 echoes in each patch. For each echo, the absolute value of the weight matrix was used to calculate the average saliency map and the average saliency map was normalized to compare the differences from each patch.

After getting the average saliency map, each patch showed a different high-weight spatial region. To know if the high weight value was more important for feature classification than

the low weight pixels, the intersection of the high saliency value from different patches was selected for the comparison, as well as the intersection of the low saliency pixels. The top 50% saliency value and the bottom 50% saliency value from 6 different patches were selected from the spectrogram. The white part showed the high-value intersection (Fig. 4.9), while the black part showed the low-value intersection. An MLP was used to do the supervised learning classification based on respect for selected regions. The process steps were as follows: first, flatten the selected region to a vector, second, do the patches classification based on the input vector, finally, compare the classification results based on high and low saliency intersection regions.

4.4 Results

During the foliage scans, a total of 37,136 echo recordings were collected from each microphone channel and 2,280 GPS points were saved. The area covered by this data collection was estimated to cover approximately 13,400 m^2, based on the morphological closing method.

The ratio between the maximum and the minimum number of echoes per cluster would change depending on the number of patches, with higher patches numbers resulting in larger ratios. For example, when classifying GPS data into 2 patches, the ratio was about 1.1, and when the number of patches was 100, the ratio was 6.8. Repeating the clustering 100 times allowed us to minimize the ratio. At the same time, a threshold of the silhouette value was set to 0.4 to optimize the classifications.

The training converged after 50 epochs. The loss decreased very sharply in the first 20 epochs and then started to level out and slowly converge with additional epochs. Training was stopped at 100 epochs due to a lack of clear improvements beyond this point for both training and validating data. The over-fitting was minimized with optimization methods,

including five folders cross-correlation and batch normalization, which resulted in an accuracy difference between training and testing after 50 epochs of less than 3 % — meaning the module was trained properly (Fig. 4.7).

The results for the two networks with a different number of patches and a different number of echoes have shown a pattern (Fig. 4.10), in which more echoes always resulted in better prediction accuracy and more patches always resulted in worse prediction accuracy. For the easiest case where the foliage was clustered into two parts, classification of one echo achieved 94.6% accuracy — 44.6% better than the chance level. With more echoes, the performance easily approached 100% accuracy. For the most challenging navigation resolution tested, which was 100 patches, classification with one echo resulted in 44% accuracy, while with ten echoes (which should be within one second), the two network modules achieved 83 % accuracy — 82 % higher than the chance level.

The comparable results showed the results from 10, 40, and 80 patches with different numbers of input echoes. More echoes gave higher classification accuracy; however, the curve converged beyond a certain number of echoes. At the same time, the results with 2, 5, and 10 echoes classify on different numbers of patches, when the patches number was the same, more echoes get better; it can be imagined, with more echoes, the results can be better. Even if the accuracy got dropped when having more patches, the network still can do better jobs, which was much higher than the chance level.

The average saliency map from nine different patches shows somewhat different patterns (Fig. 4.8); for example, the first patch had high saliency pixels more spread out than the other patches, especially in the time direction. One hypothesis to interpret the features hidden in the echo's spectrogram would be that the patch one region had more layers, which caused a long duration of reflections. Another clear difference we can see was the high saliency region was also different in frequency direction in the spectrogram. For instance,

patch four had a higher frequency than patch seven. This might be caused by the structure of the foliage, like the reflection target's shape or size. However, it's hard to prove which feature caused the echo's difference. This feature visualization method showed how the DNN does the classification based on the time-frequency intensity maps.

The results from the intersection of the saliency pixels are shown in Fig. 4.9. For example, using a three-layer MLP, five patches, and one echo input resulted in 91% accuracy when using the top 50% intersection and 62 % accuracy with the bottom 50% intersection. This showed the saliency highlights the important features for the classification, resulting in a 29 % difference in terms of the classification performance. In addition, the Resnet152 with raw input spectrogram perormed only only about 2 % better than a three-layer perceptron while taking much longer to train. The saliency not only pointed out which part of the input the network was focusing on, but also selected the most important information for the classification task and helped a lot in saving time with almost the same performance.

4.5 Discussion

Prior work has already established that "clutter echoes" from natural vegetation contain information about the targets that produce them. This is true despite the profoundly unpredictable and unrepeatable nature of the individual waveforms [86]. Early on, it was demonstrated that different tree species can be distinguished based on echoes of their foliages [85, 86]. Furthermore, prior work by the authors has already demonstrated that ten different locations in natural environments distributed over an area with a diameter of about 50 kilometers could be recognized reliably based on single clutter echoes [144]. It is possible that the differences in the echoes obtained across at least some of these sites were due to very dissimilar vegetation such as deciduous versus pine forest. In these cases, it can be expected

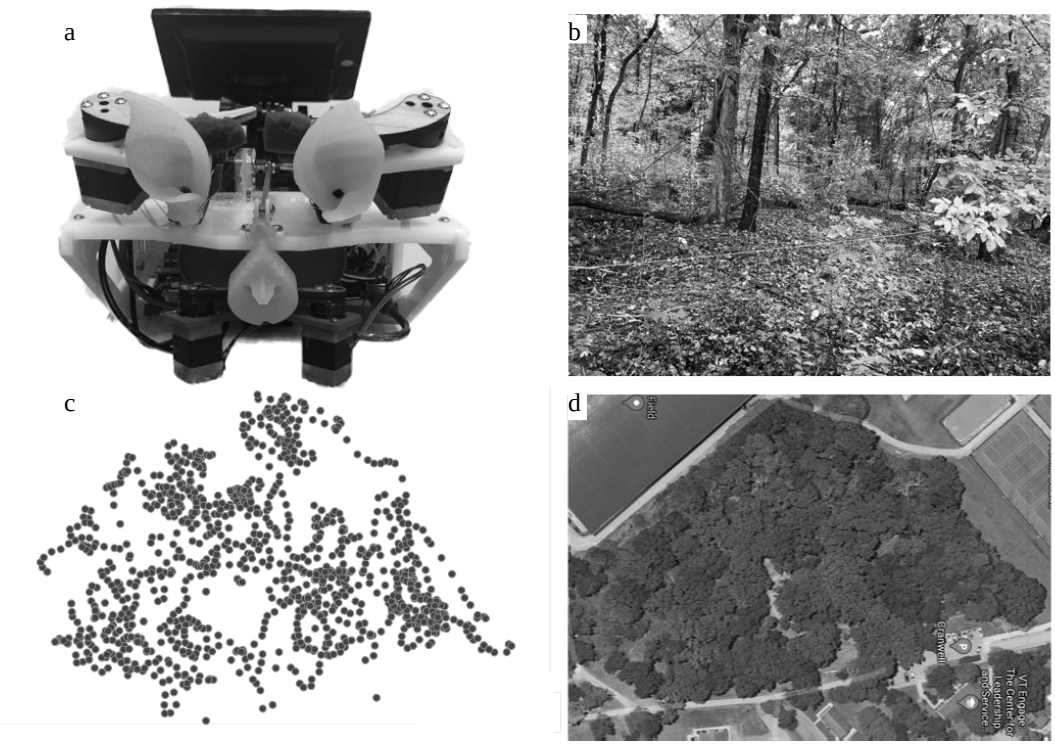

Figure 4.1: **Biomimetic sonar robot and field site used for data collection.** **(a)** Front view of the biomimetic sonar robot consisting of two ultrasonic loudspeakers to produce the pulses and two microphones mounted into the ears for echo reception, the screen in the back of the device provides the user interface. **(b)** Forest habitat at the field site. **(c)** GPS locations associated with the collected echo data set. **(d)** Satellite image of the entire data collection field site (size: 150 m by 180 m).

that foliage types with large differences in parameters such as leaf size and density also give rise to very different echo waveforms. However, the same study has also demonstrated identification of two different tracks that were walked for echo collection at each of the ten sites [144]. The latter finding could be seen as an indication that location identification based on clutter echoes could be possible on a finer scale than would be defined based on fundamentally different vegetation types. Still, the different tracks of the previous study

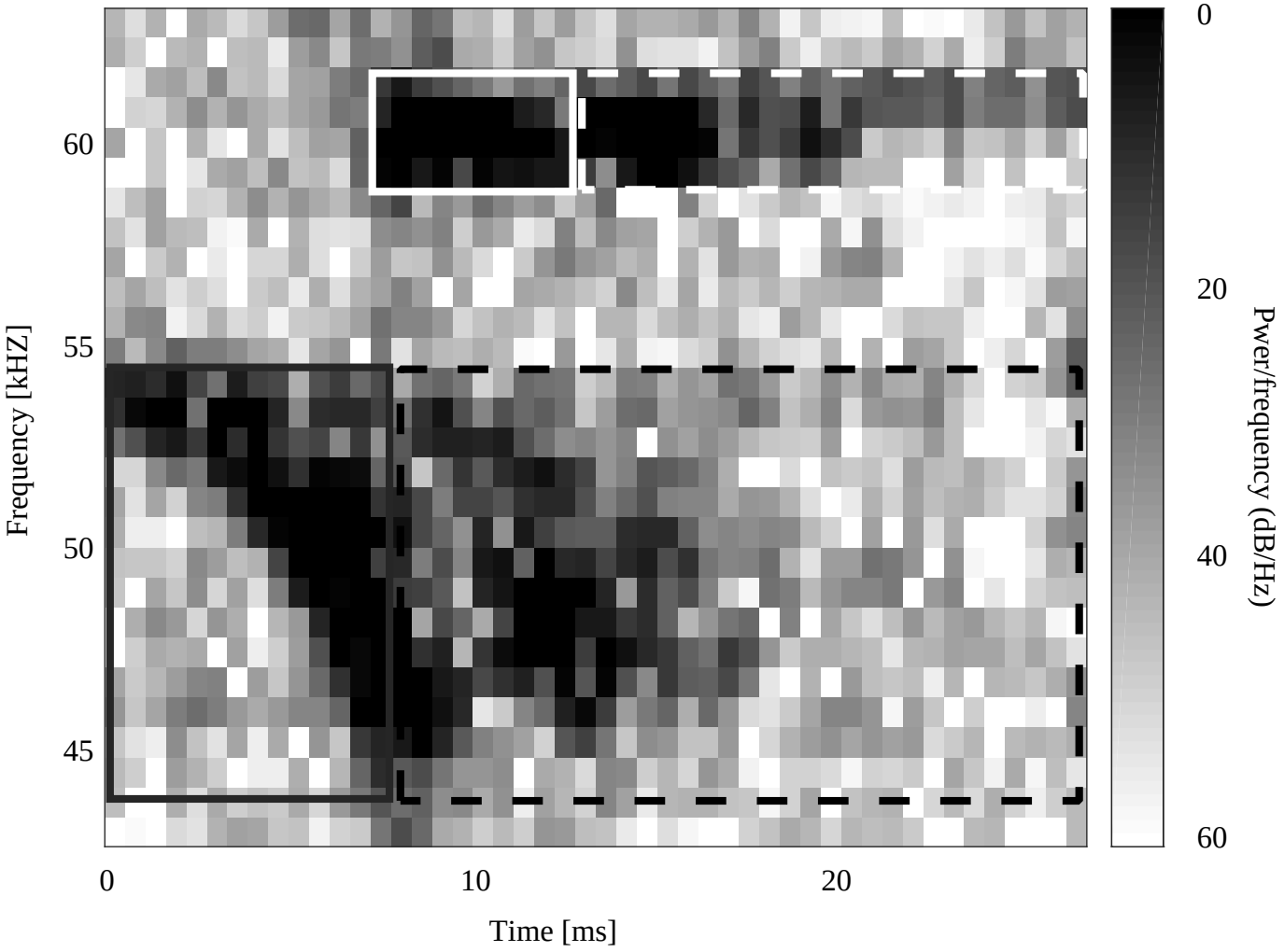

Figure 4.2: **Spectrogram of an example of the echoes that have been used for location identification.** The emitted signals consisted of a CF-FM pulse pairs where the FM pulse swept from 55 kHz down to 45 kHz over a duration of 7 ms (solid black box) and was followed by a CF pulse centered at 60 kHz and a duration of 5 ms (solid white box). The echoes to both pulses are shown in the dashed boxes (black dashes: FM echoes, white dashes: CF echoes).

could reflect somewhat large-scale changes in the vegetation, e.g., due to differences in soil, exposure, or the level of maturity of a forest.

The ability to recognize locations on a large scale, e.g., by virtue of different vegetation types,

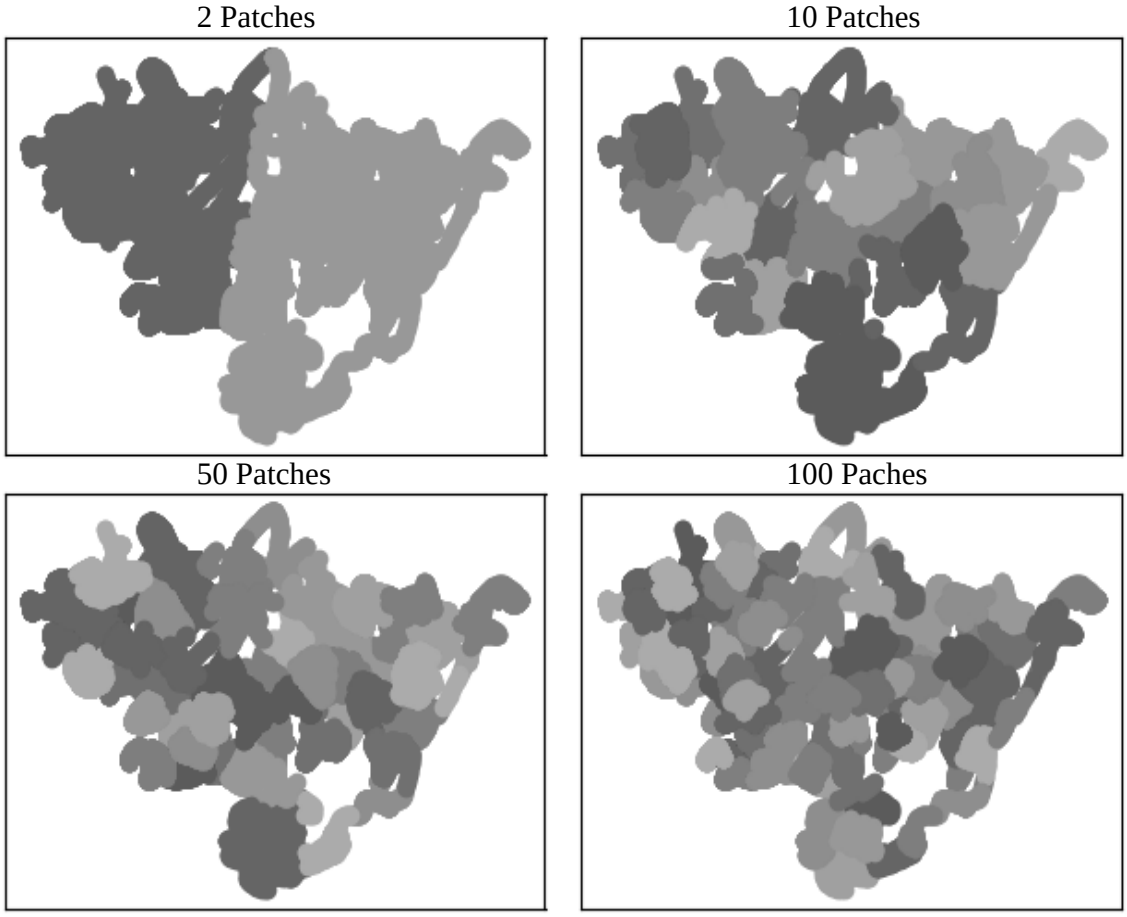

Figure 4.3: **GPS location data clustered into different numbers of spatial patches using the MiniBatch k-means method**. The clustering examples shown are 2, 10, 50, and 100 spatial patches (the locations belonging to each patch are shown in different gray levels).

is likely not enough to support an efficient navigation which should be able to chart a path to the destination in a continuous fashion and hence requires frequent, accurate, and precise updates on location. In this context, the results of the current study are significant because they demonstrate a much finer resolution that could very well support efficient navigation

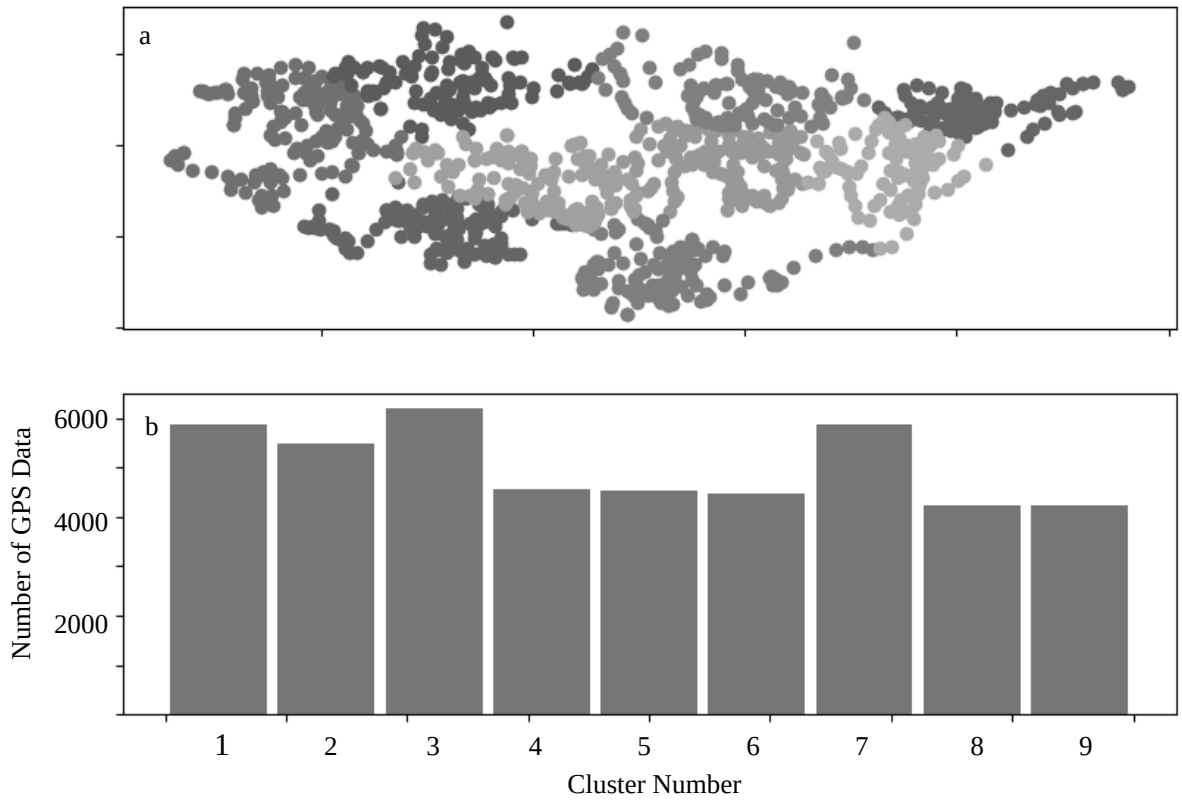

Figure 4.4: **Clustering the GPS locations into spatial patches while avoiding heavily skewed allocations across clusters**. **(a)** Allocation of the GPS locations into different clusters (nine in this example, each marked by a different gray level). **(b)** Number of the GPS locations included in each cluster showing a maximum-to-minimum ratio of 1.41 in this example.

and hence explain how bats can find their way through the forest. They could also offer an interesting solution to the problem of navigation in GPS-denied environments [53, 103] for man-made systems.

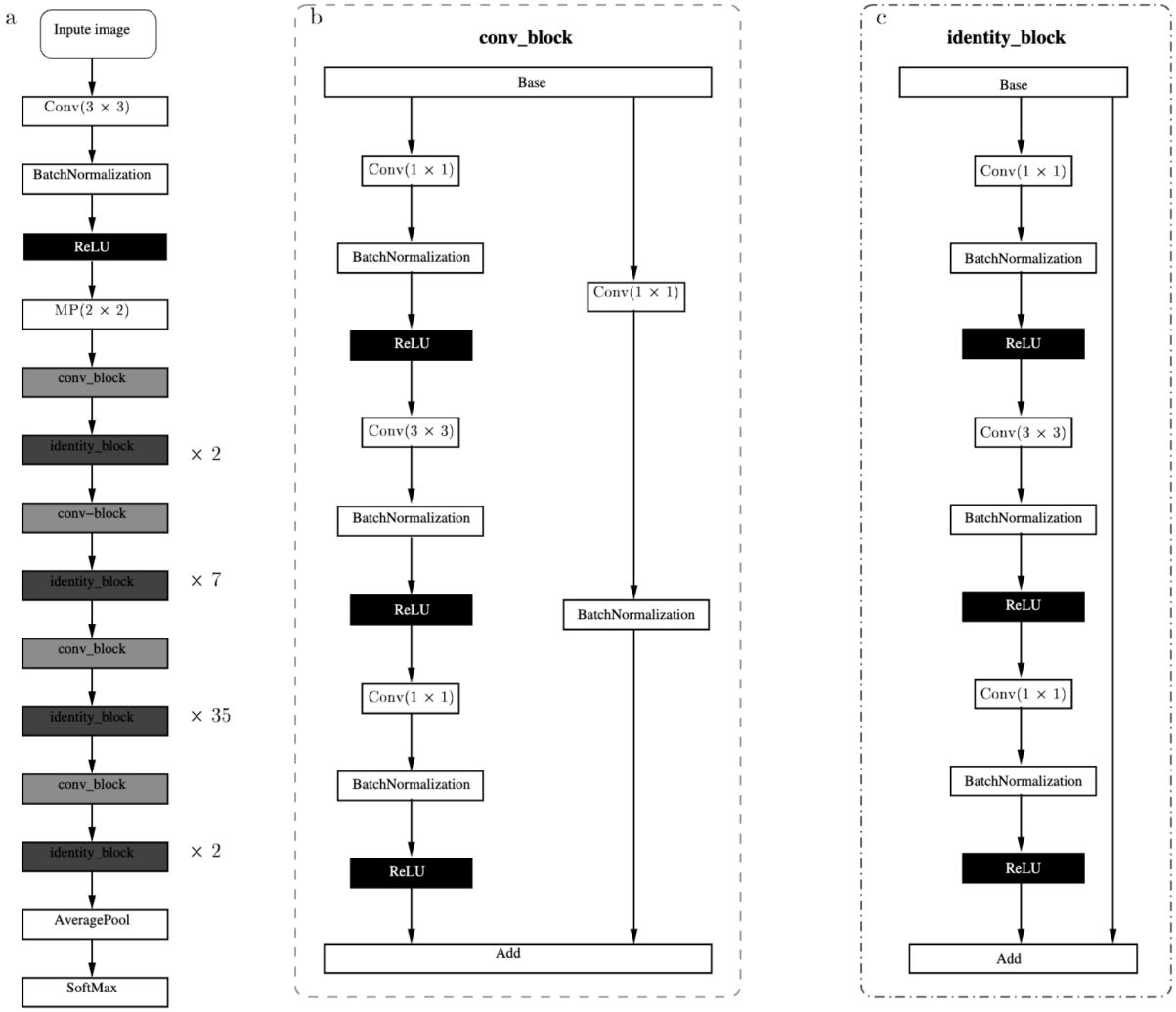

Figure 4.5: **Deep convolutional neural network architecture for classification of spatial patches based on biomimetic echoes**. **(a)** Overall architecture of the ResNet152 with four convolution blocks and 46 identity blocks), **(b)** architecture of an individual convolution block with three convolution stages and one layer convolution used to adjust the number of filters. **(c)** identity block architecture with three convolution layers and the original input propagated in parallel.

In the latter context, it is interesting to note that the spatial resolution achieved by the location patches that could be correctly identified here is not far from what has been reported for GPS operating under foliage cover: An evaluation of a recreation-grade GPS device (Suunto Ambit 3 Peak device) operated under foliage cover [78] has yielded RMS errors for

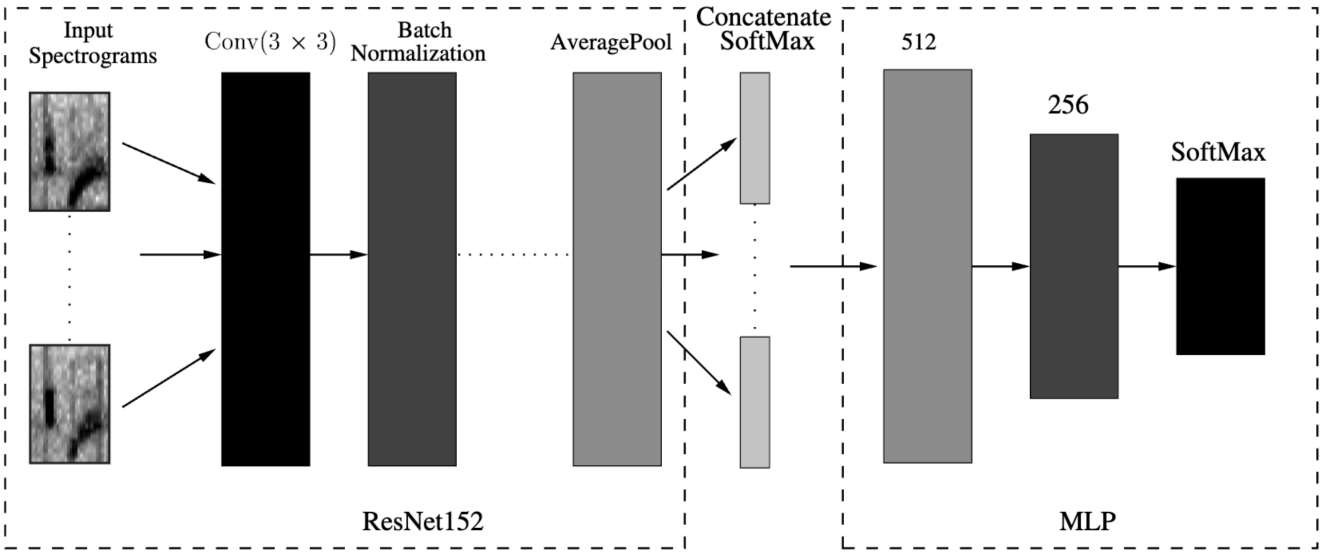

Figure 4.6: **Network architecture for the identification of spatial patches based on sets of multiple echoes**. The spectrogram representations of all echoes in the set are fed into a ResNet152 to extract time-frequency features from the entire echo set. The feature vectors derived from the output of the final SoftMax layer of the ResNet152 were concatenated into a single vector containing the feature maps for all individually echoes. The concatenated feature vector is passed into a multi-layer perceptron (MLP) to perform the supervised identification of the corresponding spatial patches.

location of 10.06 m for coniferous forest and 15.81 m for deciduous forest [68, 130]. If the total area covered in the current work is broken up into 100 equal-area spatial patches, each patch would have a radius error of about 6 m. For this scenario, an accuracy higher than 85% was achieved based on sets of ten echoes. While it is difficult to do an exact comparison of these numbers and some of the scatter in the results presented here may actually go back to errors in the GPS reference, it appears that the biomimetic sonar-based localization explored here could achieve a similar accuracy than GPS.

Besides the spatial resolution for different locations, it is also worth considering the effort that is required for dealing with the data from different sensory modalities. State-of-the-art lidar systems, for example, can generate data rates between 20 and 100 $^{\text{Mbit}}/_{\text{s}}$ (for systems

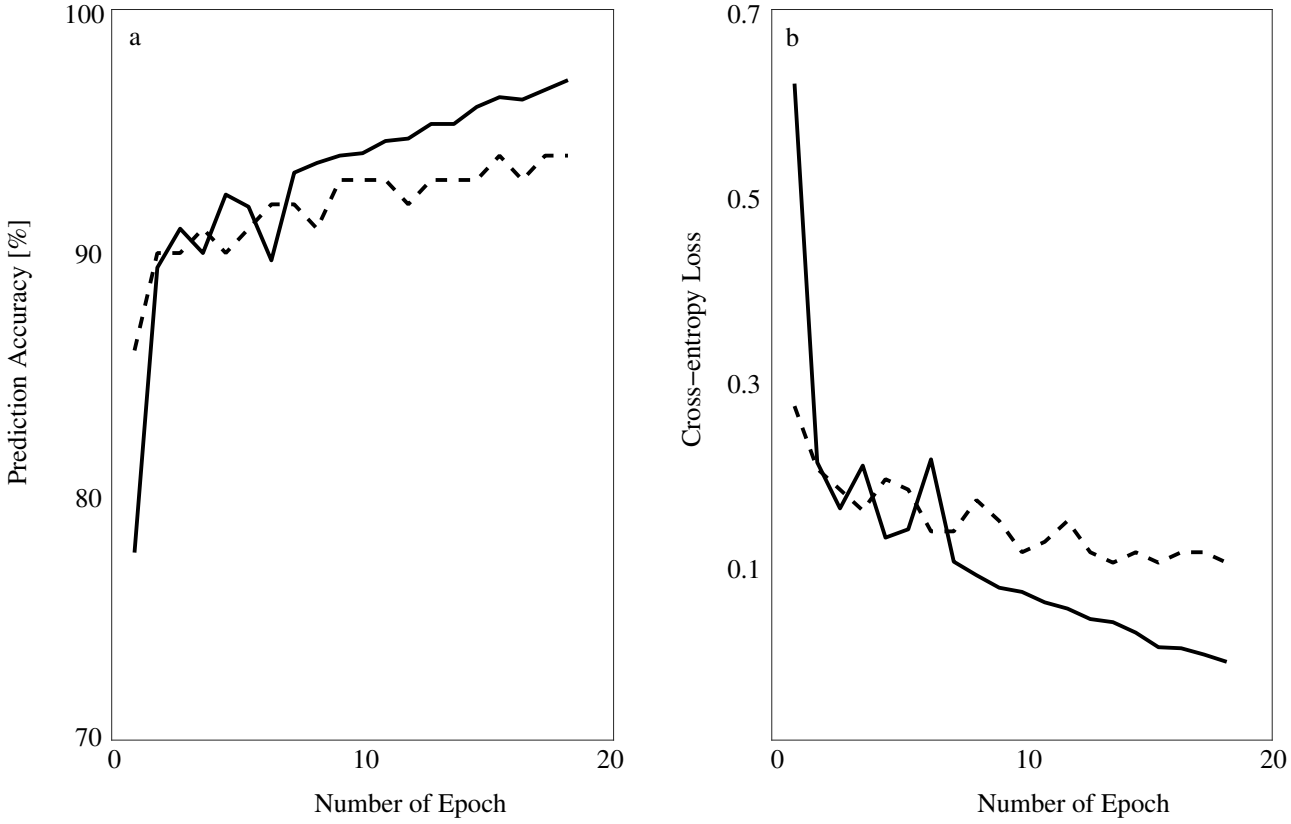

Figure 4.7: **Training (solid line) and validation (dashed line) performance of the deep neural network for location identification, one echo used to classify two patches**. **(a)** Prediction accuracy curve along the number of epochs. **(b)** Cross-entropy loss curve along the number of epochs.

with one to five sensors, [49]). Even higher data rates of 500 to 3,500 Mbits been reported for arrays of six to 12 cameras [49]. By comparison, each of the echo waveforms that were analyzed here just required 64 kbit of data (at 10 ms duration, 400 kHz sampling rate, and 16 bits resolution) to be represented without any compression. If the ten echoes that were used in the largest classification data sets in the present work were to be collected within one second, this would result in a data rate of 640 $^{\text{kbit}}/_{\text{s}}$. This would be just less than one thirtieth of the 20 $^{\text{Mbit}}/_{\text{s}}$ data rate generated by a single lidar sensor. Hence, bioinspired approaches like

the one explored here could offer much more parsimonious ways to support navigation than data-intensive sensors such as lidar. Operating on such low data rates could enable small, agile platforms that consume little power and are capable of fast reactions. In addition to the low data rates, work on a different navigation problem, passageway finding[136], has found that the computationally expensive deep-learning methods could be replaced by a simple neuromorphic approach. This approach could be implemented in analog hardware and would hence be extremely fast and power-efficient. For the current task, location identification, the feasibility of such an approach has yet to be established. However, the localized nature of the relevant information in the time-frequency plane that was evident in the saliency maps could be conducive to encoding information that exists in a certain frequency channel and in a certain time interval using a simple spike code.

If bats are able to exploit the location information contained in the clutter echoes, it would provide an explanation for how the animals are able to find their way in densely vegetated environments without the need of reconstructing any deterministic features in their surroundings. Given the small size of brains in bats is about $0.82 \pm 0.21\,\mathrm{g}$ [28] demonstrating this skill in bats would also be a strong indication that parsimonious implementations of the location estimates on clutter echoes are possible.

Future work on the ability to identify locations from clutter echoes should investigate the use of a better reference than consumer-grade GPS to better reference to evaluate resolution and accuracy that can be achieved. Ideally, such a detailed study of the resolution of the approach should be repeated across different habitats to see if some of them are better suited than others. An additional aspect that should investigated is the stability of location information over time. Since the informative echo properties do not rely on any deterministic spatial pattern, it could be hypothesized that they are very robust against changes to the positions of individual reflecting facets (e.g., leaves). However, seasonal changes to a foliage could disrupt

identification. This is very obvious when considering a deciduous forest in summer and in winter, but even much more gradual changes may eventually degrade location identification. Finally, if location identification based on biomimetic echo is found to be sufficiently accurate and stable, it could be investigated how the specific nature of these echoes could be best integrated into state-of-the-art map-building approaches such as SLAM[101].

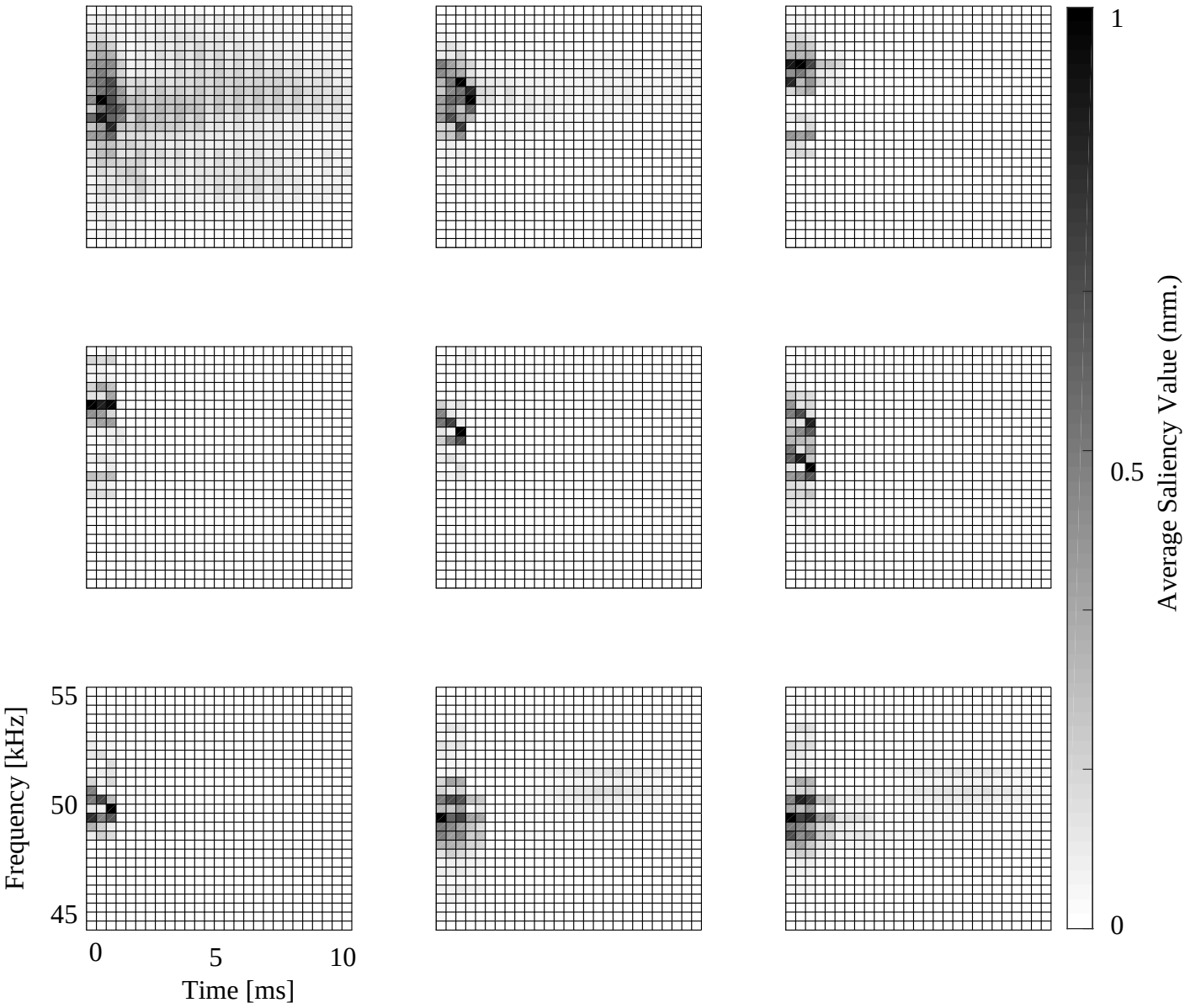

Figure 4.8: **Classification features in the time-frequency domain**. Average of 2,000 saliency maps for nine different spatial patches. The data set sizes for this figure ranged from 2,500 to 4,600 saliency maps. For data sets greater than 2,000, the averaged saliency maps were randomly picked to yield an equal sample size. Each saliency map has the same size as the input spectrogram.

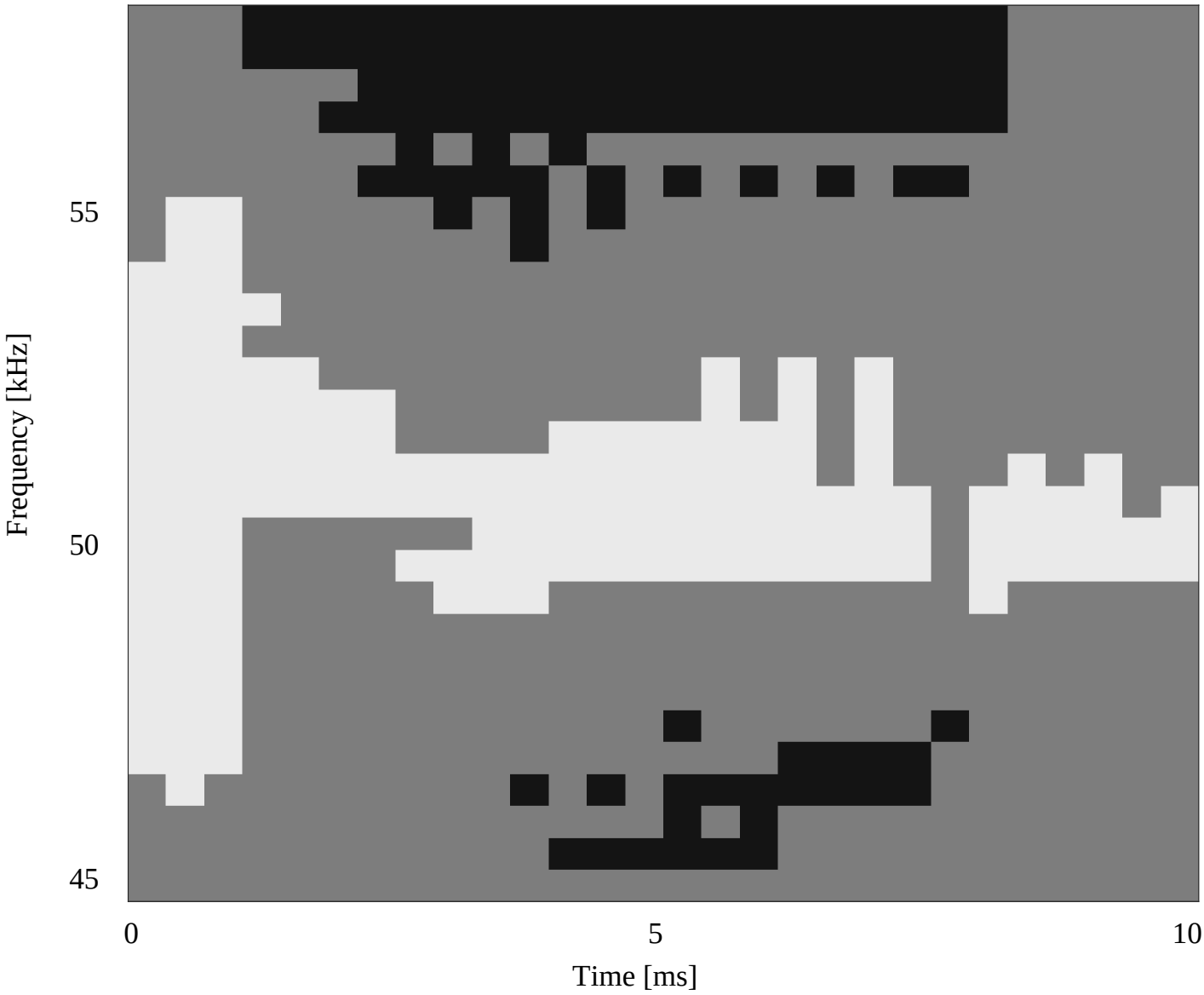

Figure 4.9: **Breakdown of the echo time-frequency plane into regions of different saliency**. Top 50% salience intersection (light gray), bottom 50% (black) . The regions were determined as the intersection of the individual saliency values, i.e., a time-frequency bin belongs to the top 50% values if the saliency values in all individual maps belong to that value range.

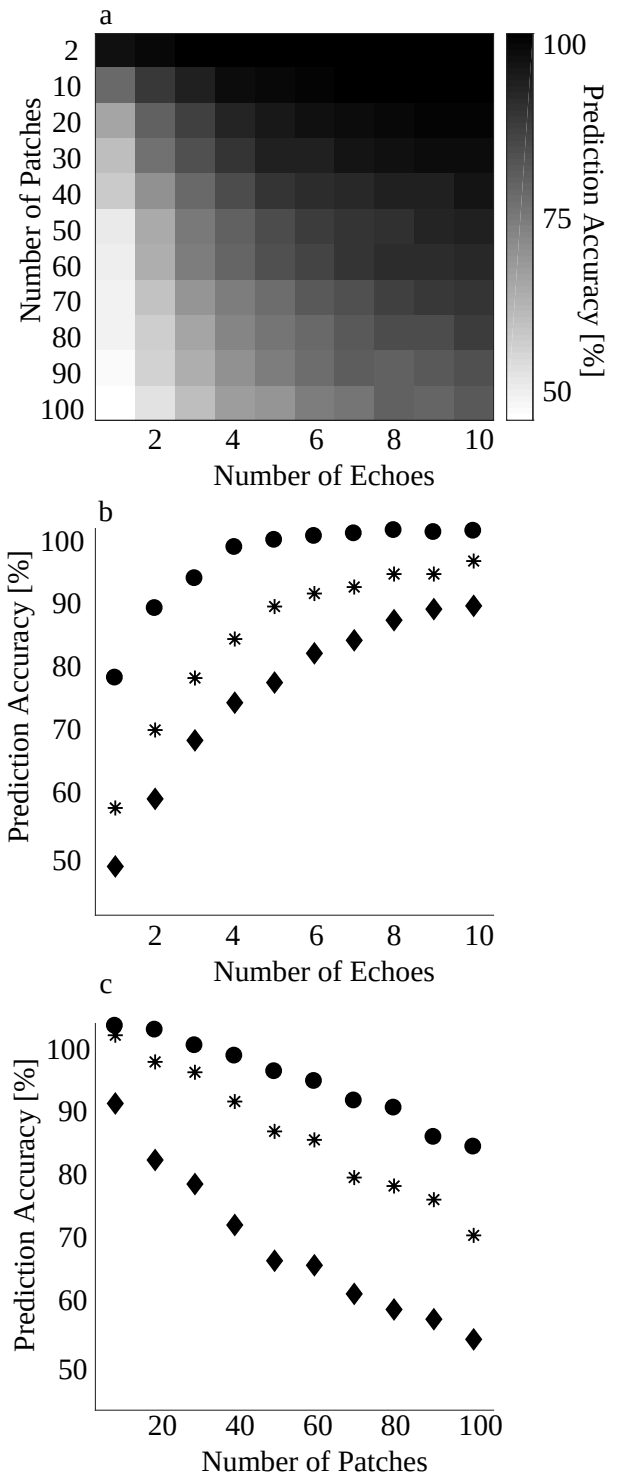

Figure 4.10: **Location identification performance for different numbers of spatial patches and echoes. (a)** Performance as a function of both variables (number of patches and echoes). **(b)** Prediction accuracy as a function echo data set size for three different number of spatial patches (circles: 10 patches, stars: 40 patches, diamond: 80 patches). **(c)** Prediction accuracy as a function of the number of spatial patches for different echo set sizes (circles: 2 echoes, stars: 5 echoes, diamonds: 10 echoes).

Chapter 5

Conclusions

5.1 Research accomplishments and findings

I have utilized a bioinspired sonar robot for localization in a natural forest. The sonar robot uses the sonar system to classify different reflection targets and distinguish the different patches of a dense forest area. The use of the deep neural network provided more than 95% accuracy for large-scale airborne sonar navigation and also achieved higher accuracy for small patches localization compared to the GPS system. The detailed findings are here:

- Pioneered deep-learning analysis of clutter echoes from natural environments. Bat's biosonar systems can achieve higher navigation ability than GPS, making bat-inspired sonar a navigation sensor choice in GPS-denied environments.
- Demonstrated large-scale identification of locations – between different locations, and also at the same location but different tracks.
- Demonstrated small-scale resolution comparable to recreation-grade GPS.
- Explored the nature of echo features with trading off time-frequency resolution, which showed that CF and FM signals encode location information differently.
- Discovered that airborne sonar achieved better navigation resolution than the GPS, which is about a 6 m radius under the dense forest.

- Developed a method for combining multiple echoes. Multiple echoes, combination leads to better performance for echolocation. Here, we found that, with 10 echoes, we can classify the area into 100 patches with more than 85% accuracy. Therefore, utilizing multiple 10 ms long echoes for biosonar navigation would not take too long, but would produce a better performance.

- Utilized the Saliency maps to visualize the different locations have different time-frequency signatures.

5.2 Identification without deterministic template

The key novel insight from the research of this thesis is that, although the vegetation echoes are random due to lack of knowledge and show no similarity in terms of correlation beyond what is introduced by the common input pulses, it is possible to extract location-specific information from them. In order to demonstrate that classification of random processes without correlation or the presence of a deterministic pattern is possible, we present a simple "toy problem" based on simulation data: In this case, we consider an environment that consists of two spatial patches (Fig. 5.1A). The echoes recorded from both patches are random processes where the individual sample amplitudes are drawn from independent and identically distributed (IID) Gaussian processes with zero mean (Fig. 5.1B). Hence, the simulated echoes from the two fictive patches only differ in their standard deviation which was one for the first patch and 1.5 for the second (Fig. 5.1C). Since the echoes were realizations of IID processes, the correlation coefficients between any two echoes within and between patches can be expected to be zero. In accordance with this expectation, the experimentally determined values of correlation coefficient were found to be 0.000036 ± 0.0158 for echoes drawn from the process of standard deviation 1 (Fig. 5.2B), 0.00009 ± 0.0158 for echoes

drawn from processes of standard deviation 1.5, and 0.00089 ± 0.0156 for echo pairs drawn from across the two different processes. In each case, 5,000 echo pairs were used for the respective estimate. The results from the simulation data is smaller with our foliage echoes (Fig. 3.4), for which the average cross correlation coefficient is 0.13 with a standard deviation of of 0.04. This means the echoes collected from foliage is not entirely random, for example, the beginning part of each recording has a higher amplitude than the end of the echo. This a pattern exists in the real foliage data; however, the correlation coefficient is still very small which is 0.13.

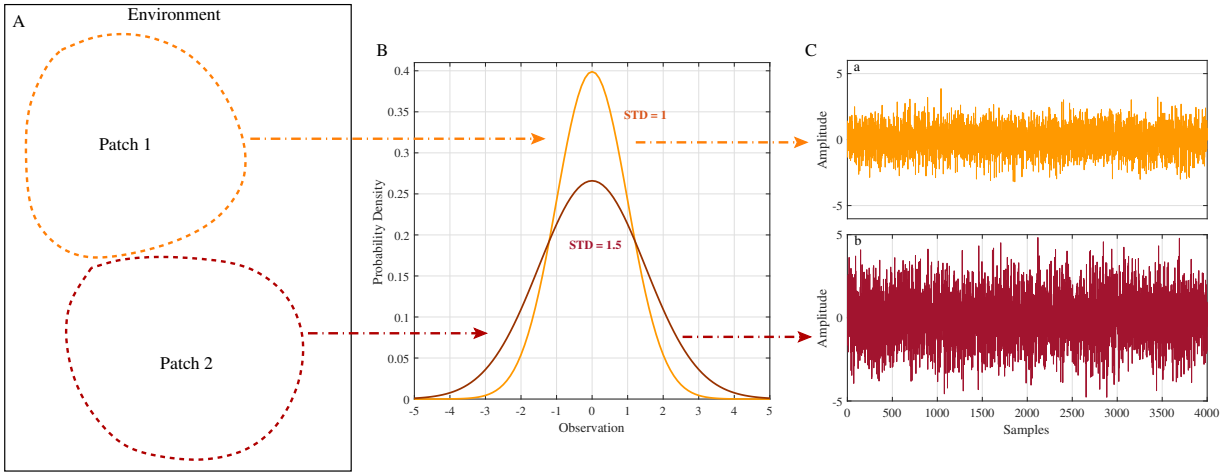

Figure 5.1: **Simulated foliage echoes.** A) Artificial environment (consisting of two patches. B) Gaussian probability distribution of the two groups of simulated echoes assigned to the two patches. C) Two simulated echoes, both from IID Gaussian processes with zero mean, but different standard deviations: a) standard deviation 1, b) standard deviation 1.5.

To demonstrate that classification based on random processes without a deterministic template is possible, a simple neural network was trained using supervised learning to classify the two fictitious patches. The network (Fig. 5.3A) consisted of an input layer with 4,000 elements to match the length of the simulated echoes followed by three fully connected layers with dimension of 128, 64, and 32 to reduce the size of the input. Finally, a Softmax

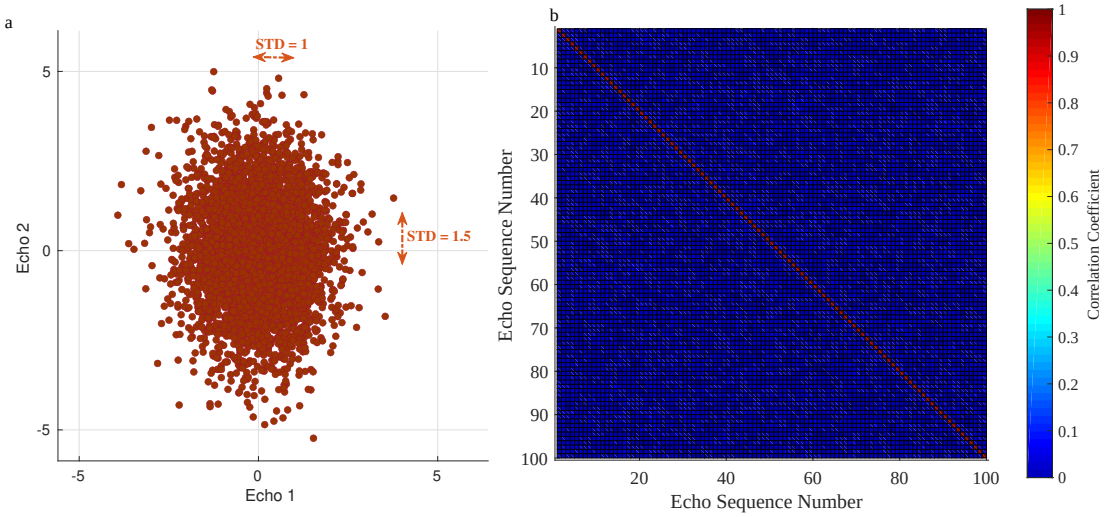

Figure 5.2: **Simulated echo classification** A) Joint amplitude distribution for two random echoe examples with an estimated correlation coefficient of 0.0028. B) Example correlation matrix for 100 simulated echoes from the Gaussian IID process with zero mean and a standard deviation of one.

function was used to convert the network output into a class estimate. To train and test the network, a total 6,000 simulated echoes, 3,000 for each of the two patches, were generated as described above. This data set was randomly split into training data (80%) and validation data (20%). The training curves (Fig. 5.3B) demonstrated that even a simple three-layer network can converge to a working estimator after six epochs. The classification accuracy was found to be 96%.

While the foliage echoes that were acquired in this research were certainly much more complicated signals than the IID Gaussian processes of this toy problem, the example problem demonstrates the results that have been achieved with these echoes do not violate any fundamental limits on classification or the capabilities of deep neural networks.

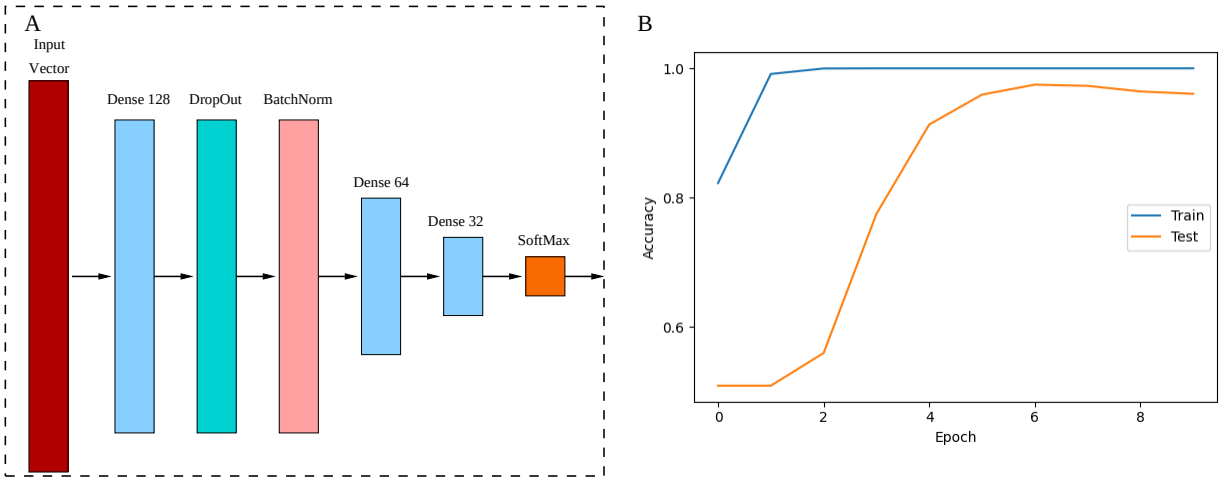

Figure 5.3: **Multilayer perceptron used to classify the model echoes created by iid Gaussian processes with different standard deviations.** A) Network architecture used to classify the different simulation echoes. B) Training and validation accuracy performance curves for the model echoes.

5.3 Discussion

Our biomimetic sonar robot demonstrates that robots can navigate in the natural forest, using a single 10 ms long ultrasonic echo. On a large scale, biosonar is able to classify different tree species, which can help bats navigate based on landmarks to fly long distances. In addition, we also explore how the airborne sonar can achieve a high navigation resolution, which is similar to GPS devices. Based on this, our sonar robot can build a map inside of a dense forest area. As a result, a large map can be built for navigation based on the small maps and the large-scale landmark differences. As with the other sensors, airborne sonar also can be used in navigation tasks with the map built in.

Experiments have been done to integrate the airborne sonar and landmark classifications, however, previous methods are limited in applying simple artificial target classifications [35], such as the different size of sphere [26], and different indoor edges and cylinders [11]. There are also a few studies done with more complex targets classifications, like using the

echo to classify artificial trees and flowers [118]. However, all of the bat-inspired landmark classifications were not as close as real bats in navigating and finding landmarks in the natural forests. Here, our data was collected from the complex natural forest which covers 20 tracks in a 25 km radius area, instead of using sonar equipment to scan a few artificial trees.

In contrast to previous approaches, we did not classify the landmarks relying on the features from biological or practical plausibility assumptions [118]. Instead, we used the raw spectrograms at a fixed frequency range, and a convolutional neural network to detect the potential features. When we tested the classifiers on a single echo from the test data set, we got a classification performance of 99.6%. This indicates that the biomimetic echoes created by a FM/CF ultrasonic sweep can be highly informative about the landmark compositions. This might one of the explanatory bases for some of the observed abilities of bats in classifying complex targets such as vegetation landmarks or food sources. The fact that the neural network was able to classify more complex targets with better classification performance indicates that the neural network must be making more effective use of the available data than the methods used before.

Although the classification task was more difficult, the performance was better than previous methods [143], making the airborne sonar sensor more able to perform real-world experiment and applications. In addition, we provide experimental evidence that passive in-air sonar landmarks can indeed be perceived in dense clutter. Airborne sonar sensors communicating with bio-inspired sonar landmarks could work as important backup systems, which would support navigation capability of visual-based sensors [18, 34, 110] as they also work under visually challenging conditions such as in glass-rich, dark, or dusty environments.

Both broadband and single-band frequency spectrograms are applied to the classification tasks, and the results show the spectrogram was a reasonable descriptor. Broadband would

be suggested for the landmark classification task as it achieves higher prediction accuracy and less time consumption. This provided a reference to use broadband as a better option for future sonar based object classification tasks. Here, we have shown the resolution of the spectrogram was an optimized parameter in terms of echo representation. For the signal processing and signal classification, the time-frequency resolution of the spectrogram can provide new dimensional information of the CF and FM frequency signals. Here, we achieved the best echolocation prediction accuracy in the middle of the resolution range for the FM spectrogram, which was a 30% improvement from the worst resolution case; for CF spectrogram the best resolution prediction accuracy is twice of the worst resolution. Therefore, there is a necessary step to optimize the spectrogram's resolutions before use.

Previous research has measured the GPS navigation errors with a Suunto Ambit 3 Peak device, and has found differences when testing it in the foliage-covered areas or in open spaces [68]. Under the seasonal forest and leaf conditions (Fig. 5.4), the average RMS error for navigation is 10.06 m in coniferous forests, 15.81 m in deciduous forests, and 15.01 m in areas with no foliage [68, 130]. In our sonar device, the results show the accuracy will be higher than 85%. With the input being 100 patches with 10 echoes (each echo is 10 ms), the navigation radius error is about 6 m.

In addition, the performance based on sonar is much better than the expectation. The one-dimensional echo superposition can encode more information with the random forest reflection. The previous work [144] has shown no single pair of echoes are the same, based on the cross-correlation. With the help of the DNN, the foliage echos can be decoded into small pieces with high accuracy. For the sonar system shown here, the navigation precision from echoes could be as precise or even more precise than GPS.

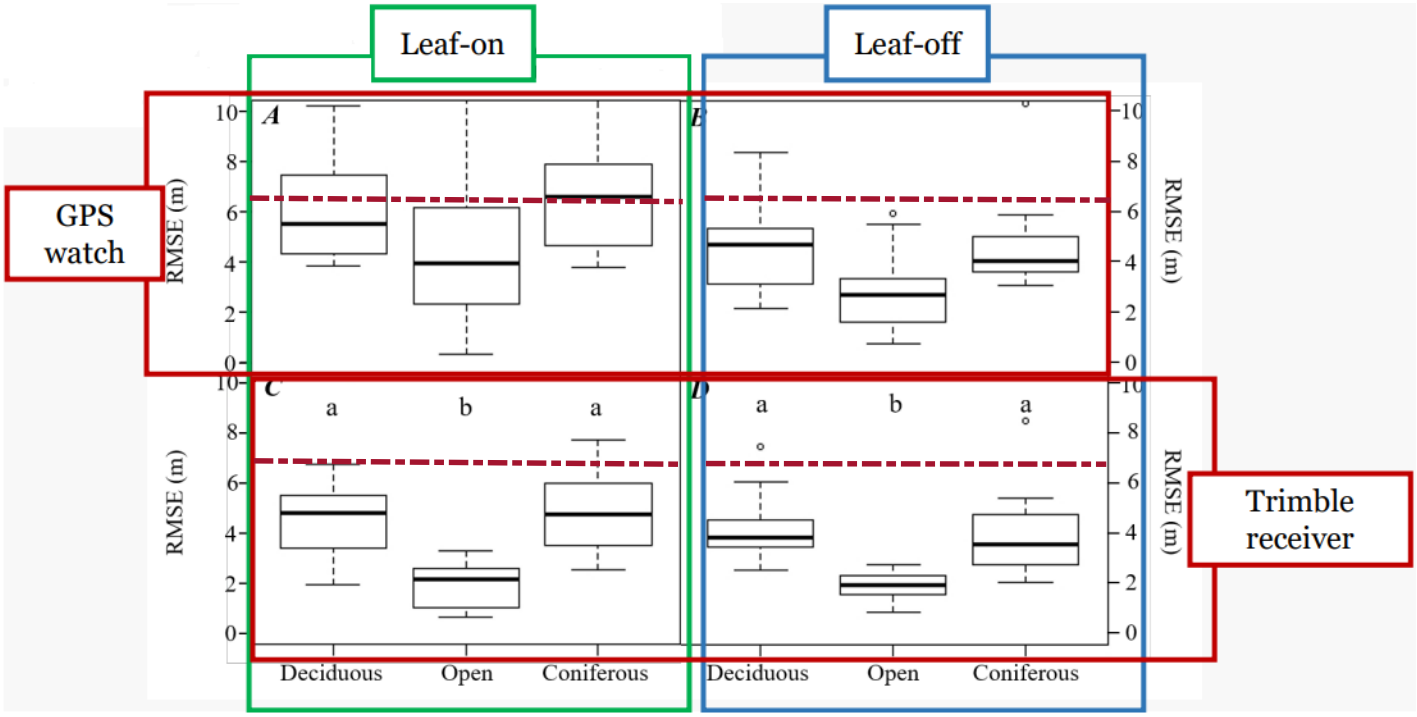

Figure 5.4: **GPS devices navigation accuracy (reprinted with permisson from the copyright holder [68])**. A) recreation-grade GPS (GPS watch) during the leaf-on season. B) Mapping-grade GNSS receiver during the leaf-on season. C) GPS watch during the leaf-off season. D) Mapping-grade GNSS receiver during the leaf-off season. The dashed horizontal maroon lines indicate the estimated accuracy achieved in the current work based on biomimetic sonar echoes.

We not only found that airborne sonar can be used for navigation, but we also proved that multiple concatenated echoes would also improve the performance. Using multiple echoes didn't increase the data size dramatically, and it is still within a few hundred kb per concatenated sample. Thus, using multiple echoes this is one way to avoid the low accuracy,

and can allow the sonar system to have reliability. With the results shown (Fig. 4.10), the performance can be much better with more echo input.

Current work on autonomous vehicles heavily relies on vision-based sensors [119] and lidar systems [12], however, cameras aren't reliable in obscured-visibility scenarios [61, 100] like dark nights and bad weather. Furthermore, cameras are also not efficient for processing huge data sets in real time, as previous papers show that cameras array (including between 6 and 12 cameras) need another 500 to 3,500 Mbit/sec [49]. While lidar makes up for some of these shortcomings, collecting thousands of 3D coordinate data points and creating huge data sets [21, 37] make real-time processing (required for navigation) very computationally expensive. For lidar sensor (including between one and five lidar components), it is 20 to 100 Mbit/sec [49]. Sonar sensors could work as an alternative/backup system for cameras and lidar navigation sensors, resistant to the challenges of low-light conditions, and limited computer resources. The sonar waveform for each echo is only 40 kb. Even though multiple echoes are needed to perform accuracy the localization, data collection can be done very fast because each echo is 10 ms long. For example, collecting 10 40 kb echoes, would take 1 s and require storing only 400 kb of data. With the help of deep networks [48], 10 ms long echo results in nearly 100% accuracy [144] at large-scale landmarks classifications.

Results indicate that the echo features carry location information, and suggest sonar sensor-based navigation is applicable to man-made robot systems, such as autonomous vehicles [55, 70]. It can help for map building and navigation in a GPS-denied environment (e.g., under dense vegetation [137], underwater [74, 138]) as long as there is sufficient texture.

The findings of the current work could also inform new directions in the development of autonomous drones that navigate in dense forests. A common approach to providing the sensory foundation for autonomous systems is to achieve high spatial/angular resolutions, e.g. by using large sonar arrays [32] or lidar scanners [137]. The current results indicate that

high spatial resolution may not be a necessity, at least for the studied task of navigating in dense forests. It could mean that sensory systems for autonomous drones do not need to be designed for high angular resolution. Without this specification, sensory systems are much smaller, lightweight, and require much less power for signal conditioning and computations.

The current work shows that, with more echoes in each patch, the localization accuracy can be higher. However, the limited data only separates the entire area into 100 patches, making the resolution about a 6 m radius. Better performance and more precision can be achieved with more data. The research takes advantage of the perceived properties of bats, landmark identification, and prey identification strategies and mechanisms. In addition, the machine learning approach achieved a better classification ability than the traditional methods, but it's hard to explain the reason mathematically. In the future, it is hoped that sonar sensors integrate with machine learning methods, such as transfer learning [128], which can help robots navigate in a natural forest in real-time.

Bibliography

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL http://tensorflow.org/. Software available from tensorflow.org.

[2] *5V, 20mA Ultimate GPS Breakout Board*. Adafruit, 2012. URL https://learn.adafruit.com/adafruit-ultimate-gps?view=all.

[3] *Adafruit Ultimate GPS DataSheet*. Adafruit Industries, 11 2013. Rev. 3.

[4] M. Adams, M. D. Adams, and E. Jose. *Robotic navigation and mapping with radar*. Artech House, Norwood, MA, 2012.

[5] Julius Adebayo, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. Sanity checks for saliency maps. *arXiv preprint arXiv:1810.03292*, 2018.

[6] Huzefa Akbarally and Lindsay Kleeman. A sonar sensor for accurate 3d target localisation and classification. In *Proceedings of 1995 IEEE International Conference on Robotics and Automation*, volume 3, pages 3003–3008. IEEE, 1995.

[7] Bhaskar Anand, Vivek Barsaiyan, Mrinal Senapati, and P Rajalakshmi. An experimental analysis of various multi-channel lidar systems. In *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, pages 644–649. IEEE, 2020.

[8] C. W. Bac, E. J. van Henten, J. Hemming, and Y. Edan. Harvesting robots for high-value crops: State-of-the-art review and challenges ahead. *J. Field Robot.*, 31(6): 888–911, 2014.

[9] Abraham Bachrach, Samuel Prentice, Ruijie He, and Nicholas Roy. Range–robust autonomous navigation in gps-denied environments. *Journal of Field Robotics*, 28(5): 644–666, 2011.

[10] Joseph Nsasi Bakambu and Vladimir Polotski. Autonomous system for navigation and surveying in underground mines. *Journal of Field Robotics*, 24(10):829–847, 2007.

[11] B. Barshan, B. Ayrulu, and S. W. Utete. Neural network-based target differentiation using sonar for robotics applications. *IEEE Trans. Robot.*, 16(4):435–442, 2000.

[12] L. C. Básaca, J. Rodríguez, O. Y. Sergiyenko, V. V. Tyrsa, W. Hernández, J. I. N. Hipólito, and O. Starostenko. 3d laser scanning vision system for autonomous robot navigation. In *2010 IEEE International Symposium on Industrial Electronics*, pages 1773–1778. IEEE, 2010.

[13] S. Betge-Brezetz, P. Hebert, R. Chatila, and M. Devy. Uncertain map making in natural environments. In *Proc. IEEE International Conference on Robotics and Automation*, volume 2, pages 1048–1053. IEEE, 1996.

[14] Ananya Bhardwaj, Mohammad Omar Khyam, and Rolf Müller. Biomimetic detection of dynamic signatures in foliage echoes. *Bioinspir. Biomim.*, 2021.

[15] JC N. Borge and C. G. Soares. Analysis of directional wave fields using x-band navigation radar. *Coast. Eng.*, 40(4):375–391, 2000.

[16] C. Brenner and B. Elias. Extracting landmarks for car navigation systems using existing gis databases and laser scanning. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS)*, 34(3/W8):131–138, 2003.

[17] Claire Burke, Paul R McWhirter, Josh Veitch-Michaelis, Owen McAree, Harry AG Pointon, Serge Wich, and Steve Longmore. Requirements and limitations of thermal drones for effective search and rescue in marine and coastal areas. *Drones*, 3(4):78, 2019.

[18] Wei Chen, Ting Qu, Yimin Zhou, Kaijian Weng, Gang Wang, and Guoqiang Fu. Door recognition and deep learning algorithm for visual based robot navigation. In *2014 ieee international conference on robotics and biomimetics (robio 2014)*, pages 1793–1798. IEEE, 2014.

[19] François Chollet et al. Keras. https://github.com/fchollet/keras, 2015.

[20] James S Cope, David Corney, Jonathan Y Clark, Paolo Remagnino, and Paul Wilkin. Plant species identification using digital morphometrics: A review. *Expert Systems with Applications*, 39(8):7562–7573, 2012.

[21] Jean-François Côté, Jean-Luc Widlowski, Richard A Fournier, and Michel M Verstraete. The structural and radiative consistency of three-dimensional tree reconstructions from terrestrial lidar. *Remote Sensing of Environment*, 113(5):1067–1081, 2009.

[22] G. Csorba, P. Ujhelyi, and N. Thomas. *Horseshoe Bats of the World: (Chiroptera: Rhinolophidae)*. Alana Books, Devon, UK, 2003. ISBN 9780953604913.

[23] M. Cutler, B. Michini, and J. P. How. Lightweight infrared sensing for relative navigation of quadrotors. In *2013 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 1156–1164. IEEE, 2013.

[24] K. Czyńska and P. Rubinowicz. Classification of cityscape areas according to landmarks visibility analysis. *Environ. Impact Assess. Rev.*, 76:47–60, 2019.

[25] Li Deng, Geoffrey Hinton, and Brian Kingsbury. New types of deep neural network learning for speech recognition and related applications: An overview. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8599–8603. IEEE, 2013.

[26] M. Dmitrieva, M. Valdenegro-Toro, K. Brown, G. Heald, and D. Lane. Object classification with convolution neural network based on the time-frequency representation of their echo. In *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE, 2017.

[27] Edward R Dougherty. An introduction to morphological image processing. *SPIE, 1992*, 1992.

[28] John F Eisenberg and Don E Wilson. Relative brain size and feeding strategies in the chiroptera. *Evolution*, pages 740–751, 1978.

[29] I. Eliakim, Z. Cohen, G. Kosa, and Y. Yovel. A fully autonomous terrestrial bat-like acoustic robot. *PLOS Comput. Biol.*, 14(9):e1006406, 2018.

[30] T. Eliav, S.R. Maimon, J. Aljadeff, M. Tsodyks, G. Ginosar, L. Las, and N. Ulanovsky. Multiscale representation of very large environments in the hippocampus of flying bats. *Science*, 372(6545), 2021.

[31] B. Falk, T. Williams, M. Aytekin, and C. F. Moss. Adaptive behavior for texture discrimination by the free-flying big brown bat, *Eptesicus fuscus*. *J. Comp. Physiol. A.*, 197(5):491–503, 2011.

[32] Saeid Fazli and Lindsay Kleeman. A real time advanced sonar ring with simultaneous firing. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 2, pages 1872–1877. IEEE, 2004.

[33] A. M. Flynn. Combining sonar and infrared sensors for mobile robot navigation. *Int. J. Robot. Res.*, 7(6):5–14, 1988.

[34] José Gaspar, Niall Winters, and José Santos-Victor. Vision-based navigation and environmental representations with an omnidirectional camera. *IEEE Transactions on robotics and automation*, 16(6):890–898, 2000.

[35] D. Genzel and L. Wiegrebe. Size does not matter: size-invariant echo-acoustic object classification. *J. Comp. Physiol. A*, 199(2):159–168, 2013.

[36] Daria Genzel, Yossi Yovel, and Michael M Yartsev. Neuroethology of bat navigation. *Current Biology*, 28(17):R997–R1004, 2018.

[37] Luis Gézero and Carlos Antunes. Automated three-dimensional linear elements extraction from mobile lidar point clouds in railway environments. *Infrastructures*, 4(3): 46, 2019.

[38] X. Glorot, A. Bordes, and Y. Bengio. Deep sparse rectifier neural networks. In *Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 315–323, Ft. Lauderdale, FL, 2011.

[39] Somya Goel, Raghav Pangasa, Suma Dawn, and Anuja Arora. Audio acoustic fea-

tures based tagging and comparative analysis of its classifications. In *2018 Eleventh International Conference on Contemporary Computing (IC3)*, pages 1–5. IEEE, 2018.

[40] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[41] Alan Grant, Paul Williams, Nick Ward, and Sally Basker. Gps jamming and the impact on maritime navigation. *The Journal of Navigation*, 62(2):173–187, 2009.

[42] D. R Griffin, Frederic A. W., and C. R. Michael. The echolocation of flying insects by bats. *Anim. Behav.*, 8(3-4):141–154, 1960.

[43] Donald R Griffin. Listening in the dark: the acoustic orientation of bats and men. 1958.

[44] J. Grunwald, S. Schornich, and L. Wiegrebe. Classification of natural textures in echolocation. *Proc. Natl. Acad. Sci. U.S.A.*, 101(15):5670–5674, 2004.

[45] I. A. Hameed. Intelligent coverage path planning for agricultural robots and autonomous machines on three-dimensional terrain. *J. Intell. Robot. Syst.*, 74(3-4): 965–983, 2014.

[46] Simon Haykin. *Neural networks and learning machines, 3/E*. Pearson Education India, 2009.

[47] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[48] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European Conference on Computer Vision*, pages 630–645. Springer, 2016.

[49] Stephan Heinrich and Lucid Motors. Flash memory in the emerging age of autonomy. *Flash Memory Summit*, pages 1–10, 2017.

[50] NE Hodge, RM Ferencz, and JM3218840 Solberg. Implementation of a thermomechanical model for the simulation of selective laser melting. *Computational Mechanics*, 54 (1):33–51, 2014.

[51] Bernhard Hofmann-Wellenhof, Herbert Lichtenegger, and James Collins. *Global positioning system: theory and practice*. Springer Science & Business Media, 2012.

[52] MA Horner, TH Fleming, and CT Sahey. Foraging behaviour and energetics of a nectar-feeding bat, leptonycteris curasoae (chiroptera: Phyllostomidae). *J. Zool.*, 244 (4):575–586, 1998.

[53] Li-Ta Hsu, Yanlei Gu, and Shunsuke Kamijo. Sensor integration of 3d map aided gnss and smartphone pdr in urban canyon with dense foliage. In *Proceedings of IEEE/ION PLANS 2016*, pages 85–90, 2016.

[54] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[55] Zhen Jia, Arjuna Balasuriya, and Subhash Challa. Autonomous vehicles navigation with visual target tracking: Technical approaches. *Algorithms*, 1(2):153–182, 2008.

[56] G. Jones and J. MV Rayner. Foraging behavior and echolocation of wild horseshoe bats *Rhinolophus ferrumequinum* and *R. hipposideros* (Chiroptera, Rhinolophidae). *Behav. Ecol. Sociobiol.*, 25(3):183–191, 1989.

[57] Himangshu Kalita, Steven Morad, Aaditya Ravindran, and Jekan Thangavelautham.

Path planning and navigation inside off-world lava tubes and caves. In *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pages 1311–1318, 2018.

[58] Elisabeth KV Kalko and MA Condon. Echolocation, olfaction and fruit display: how bats find fruit of flagellichorous cucurbits. *Functional Ecology*, 12(3):364–372, 1998.

[59] S Karma, E Zorba, GC Pallis, G Statheropoulos, I Balta, K Mikedi, J Vamvakari, A Pappa, M Chalaris, G Xanthopoulos, et al. Use of unmanned vehicles in search and rescue operations in forest fires: Advantages and limitations observed in a field trial. *International journal of disaster risk reduction*, 13:307–312, 2015.

[60] D. Kim, J. Sun, S. M. Oh, J. M. Rehg, and A. F. Bobick. Traversability classification using unsupervised on-line visual learning for outdoor robot navigation. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, pages 518–525. IEEE, 2006.

[61] Donghoon Kim, Donghwa Lee, Hyun Myung, and Hyun-Taek Choi. Artificial landmark-based underwater localization for auvs using weighted template matching. *Intelligent Service Robotics*, 7(3):175–184, 2014.

[62] Laura N Kloepper, Andrea Megela Simmons, and James A Simmons. Echolocation while drinking: pulse-timing strategies by high-and low-frequency fm bats. *Plos one*, 14(12):e0226114, 2019.

[63] D. J. Kriegman, E. Triendl, and T. O. Binford. Stereo vision and navigation in buildings for mobile robots. *IEEE Trans. Rob. Autom.*, 5(6):792–803, 1989.

[64] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*, pages 1097–1105, Lake Tahoe, NV, 2012.

[65] P. K. Kroh, R. Simon, and S. J. Rupitsch. Classification of sonar targets in air: A neural network approach. *Sensors*, 19(5):1176, 2019.

[66] S. Lawrence, C. L. Giles, Ah C. Tsoi, and A. D. Back. Face recognition: A convolutional neural-network approach. *IEEE Trans. Neural Netw. Learn. Syst.*, 8(1):98–113, 1997.

[67] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proc. IEEE*, 86(11):2278–2324, 1998.

[68] Taeyoon Lee, Pete Bettinger, Chris J Cieszewski, and Alba Rocio Gutierrez Garzon. The applicability of recreation-grade gnss receiver (gps watch, suunto ambit peak 3) in a forested and an open area compared to a mapping-grade receiver (trimble juno t41). *PLoS One*, 15(4):e0231532, 2020.

[69] John J Leonard and Alexander Bahr. Autonomous underwater vehicle navigation. *Springer handbook of ocean engineering*, pages 341–358, 2016.

[70] Qingqing Li, Jorge Peña Queralta, Tuan Nguyen Gia, Zhuo Zou, and Tomi Westerlund. Multi-sensor fusion for navigation and mapping in autonomous vehicles: Accurate localization in urban environments. *Unmanned Systems*, 8(03):229–237, 2020.

[71] R. Lippmann. An introduction to computing with neural nets. *IEEE ASSP Mag.*, 4 (2):4–22, 1987.

[72] Yuncheng Lu, Zhucun Xue, Gui-Song Xia, and Liangpei Zhang. A survey on vision-based uav navigation. *Geo-spatial information science*, 21(1):21–32, 2018.

[73] Eleftherios Lygouras, Nicholas Santavas, Anastasios Taitzoglou, Konstantinos Tarchanidis, Athanasios Mitropoulos, and Antonios Gasteratos. Unsupervised human detection with an embedded vision system on a fully autonomous uav for search and rescue operations. *Sensors*, 19(16):3542, 2019.

[74] Somajyoti Majumder, Steve Scheding, and Hugh F Durrant-Whyte. Multisensor data fusion for underwater navigation. *Robotics and Autonomous Systems*, 35(2):97–108, 2001.

[75] Frank Mascarich, Taylor Wilson, Christos Papachristos, and Kostas Alexis. Radiation source localization in gps-denied environments using aerial robots. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6537–6544. IEEE, 2018.

[76] Phillip McKerrow and Neil Harper. Plant acoustic density profile model of ctfm ultrasonic sensing. *IEEE Sensors Journal*, 1(4):245–255, 2001.

[77] R. A. Medellin, M. Rivero, A. Ibarra, J. Antonio de la Torre, Tania P. Gonzalez-Terrazas, L. Torres-Knoop, and M. Tschapka. Follow me: foraging distances of *Leptonycteris yerbabuenae* (Chiroptera: Phyllostomidae) in Sonora determined by fluorescent powder. *J. Mammal.*, 99(2):306–311, 2018.

[78] Krista Merry and Pete Bettinger. Smartphone gps accuracy study in an urban environment. *PloS one*, 14(7):e0219890, 2019.

[79] Christoph FJ Meyer, Moritz Weinbeer, and Elisabeth KV Kalko. Home-range size and spacing patterns of macrophyllum macrophyllum (phyllostomidae) foraging over water. *Journal of mammalogy*, 86(3):587–598, 2005.

[80] C. Ming, H. Zhu, and R. Müller. A simplified model of biosonar echoes from foliage and the properties of natural foliages. *PLoS One*, 12(12):e0189824, 2017.

[81] Chen Ming, Anupam Kumar Gupta, Ruijin Lu, Hongxiao Zhu, and Rolf Müller. A computational model for biosonar echoes from foliage. *PloS one*, 12(8):e0182824, 2017.

[82] A. Mohamed, G. Dahl, and G. Hinton. Deep belief networks for phone recognition. In *NIPS Workshop on Deep Learning for Speech Recognition and Related Applications*, volume 1, page 39. Vancouver, Canada, 2009.

[83] A. Moreno-Valdez, R. L. Honeycutt, and W. E. Grant. Colony dynamics of leptonycteris nivalis (mexican long-nosed bat) related to flowering agave in northern mexico. *J. Mammal.*, 85(3):453–459, 2004.

[84] J. Müller, R. Brandl, J. Buchner, H. Pretzsch, S. Seifert, C. Strätz, M. Veith, and B. Fenton. From ground to above canopy—bat activity in mature forests is driven by vegetation density and height. *For. Ecol. Manag.*, 306:179–184, 2013.

[85] R. Müller. A computational theory for the classification of natural biosonar targets based on a spike code. *Network: Comput. Neural Syst.*, 14:595–612, August 2003. doi: 10.1088/0954-898X/14/3/311.

[86] R. Müller and R. Kuc. Foliage echoes: a probe into the ecological acoustics of bat echolocation. *J. Acoust. Soc. Am.*, 108(2):836–845, 2000.

[87] Rolf Müller. A computational theory for the classification of natural biosonar targets based on a spike code. *Network: Computation in Neural Systems*, 14(3):595–612, 2003.

[88] Rolf Müller, Ru Zhang, Liujun Zhang, Peiwen Qiu, and Xiaoyan Yin. Why hipposiderid biosonar is worth studying. *The Journal of the Acoustical Society of America*, 141(5): 3484–3484, 2017.

[89] Rolf Müller, Michael Goldsworthy, Ruihao Wang, and Liujun Zhang. Machine learning challenges in bat biosonar. *The Journal of the Acoustical Society of America*, 146(4): 2982–2982, 2019.

[90] Rolf Müller, Sounak Chakrabarti, Ibrahim M Eshera, Sanmeel Vijay Lagad, Ruihao Wang, and Liujun Zhang. Autonomy, soft-robotics, deep learning, and bat biosonar. *The Journal of the Acoustical Society of America*, 150(4):A325–A325, 2021.

[91] Rolf Müller, Sounak Chakrabarti, and Liujun Zhang. Toward a computational theory for the sensory world of bat biosonar. *The Journal of the Acoustical Society of America*, 151(4):A100–A100, 2022.

[92] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[93] A Nabout, R Gerhards, B Su, HA Nour Eldin, and W Kuhbauch. Plant species identification using fuzzy set theory. In *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 48–53. IEEE, 1994.

[94] G. Neuweiler. Foraging ecology and audition in echolocating bats. *Trends. Ecol. Evol.*, 4(6):160–166, 1989.

[95] G Neuweiler, W Metzner, U Heilmann, R Rübsamen, M Eckrich, and HH Costa. Foraging behaviour and echolocation in the rufous horseshoe bat (rhinolophus rouxi) of sri lanka. *Behavioral ecology and sociobiology*, 20(1):53–67, 1987.

[96] James Newling and François Fleuret. Nested mini-batch k-means. *Advances in neural information processing systems*, 29:1352–1360, 2016.

[97] Kuniaki Noda, Yuki Yamaguchi, Kazuhiro Nakadai, Hiroshi G Okuno, and Tetsuya Ogata. Audio-visual speech recognition using deep learning. *Appl. Intell.*, 42(4):722–737, 2015.

[98] Itai Orr, Moshik Cohen, and Zeev Zalevsky. High-resolution radar road segmentation using weakly supervised learning. *Nat. Mach. Intell.*, 3(3):239–246, 2021.

[99] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[100] David Prasser and Gordon Wyeth. Probabilistic visual recognition of artificial landmarks for simultaneous localization and mapping. In *2003 IEEE International Conference on Robotics and Automation (Cat. No. 03CH37422)*, volume 1, pages 1291–1296. IEEE, 2003.

[101] A Alan B Pritsker. *Introduction to Simulation and SLAM II*. Halsted Press, 1984.

[102] Richard J Przybyla, Hao-Yen Tang, Andre Guedes, Stefon E Shelton, David A Horsley, and Bernhard E Boser. 3d ultrasonic rangefinder on a chip. *IEEE Journal of Solid-State Circuits*, 50(1):320–334, 2014.

[103] Pavel Puricer and Pavel Kovar. Technical limitations of gnss receivers in indoor positioning. In *2007 17th International Conference Radioelektronika*, pages 1–5. IEEE, 2007.

[104] Parisa Rashidi and Alex Mihailidis. A survey on ambient-assisted living tools for older adults. *IEEE journal of biomedical and health informatics*, 17(3):579–590, 2012.

[105] Agoston Restas et al. Drone applications for supporting disaster management. *World Journal of Engineering and Technology*, 3(03):316, 2015.

[106] H-U Reyer et al. Nectar intake and energy expenditure in a flower visiting bat. *Oecologia*, 63(2):178–184, 1984.

[107] Martin Riedmiller and AM Lernen. Multi layer perceptron. *Machine Learning Lab Special Lecture, University of Freiburg*, pages 7–24, 2014.

[108] Giorgio Rizzoni and James Kearns. *Fundamentals of electrical engineering*. McGraw-Hill New York, 2009.

[109] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[110] Marco Sabatini, Giovanni B Palmerini, and Paolo Gasbarri. A testbed for visual based navigation and control during space rendezvous operations. *Acta Astronautica*, 117: 184–196, 2015.

[111] Arnab Kumar Saha, Jayeeta Saha, Radhika Ray, Sachet Sircar, Subhojit Dutta, Soummyo Priyo Chattopadhyay, and Himadri Nath Saha. Iot-based drone for improvement of crop quality in agricultural field. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 612–615. IEEE, 2018.

[112] Davide Scaramuzza, Michael C Achtelik, Lefteris Doitsidis, Fraundorfer Friedrich, Elias Kosmatopoulos, Agostino Martinelli, Markus W Achtelik, Margarita Chli, Savvas Chatzichristofis, Laurent Kneip, et al. Vision-controlled micro flying robots: from system design to autonomous navigation and mapping in gps-denied environments. *IEEE Robotics & Automation Magazine*, 21(3):26–40, 2014.

[113] David C Schedl, Indrajit Kurmi, and Oliver Bimber. Search and rescue with airborne optical sectioning. *Nat. Mach. Intell.*, 2(12):783–790, 2020.

[114] Hans-Ulrich Schnitzler, Cynthia F Moss, and Annette Denzinger. From spatial orientation to food acquisition in echolocating bats. *Trends in Ecology & Evolution*, 18(8): 386–394, 2003.

[115] J. Simmons, W. Lavender, B. Lavender, C. Doroshow, S. Kiefer, R. Livingston, A. Scal-

let, and D. Crowley. Target structure and echo spectral discrimination by echolocating bats. *Science*, 186(4169):1130–1132, 1974.

[116] James A Simmons, Prestor A Saillant, Janine M Wotton, Tim Haresign, Michael J Ferragamo, and Cynthia F Moss. Composition of biosonar images for target recognition by echolocating bats. *Neural Networks*, 8(7-8):1239–1261, 1995.

[117] R. Simon, M. W. Holderied, C. U. Koch, and O. von Helversen. Floral acoustics: conspicuous echoes of a dish-shaped leaf attract bat pollinators. *Science*, 333(6042): 631–633, 2011.

[118] R. Simon, S. Rupitsch, M. Baumann, H. Wu, H. Peremans, and J. Steckel. Bioinspired sonar reflectors as guiding beacons for autonomous navigation. *Proc. Natl. Acad. Sci. U.S.A.*, 117(3):1367–1374, 2020.

[119] S. Sivaraman and M. Trivedi. Combining monocular and stereo-vision for real-time vehicle ranging and tracking on multilane highways. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1249–1254. IEEE, 2011.

[120] Guangming Song, Kaijian Yin, Yaoxin Zhou, and Xiuzhen Cheng. A surveillance robot with hopping capabilities for home security. *IEEE Transactions on Consumer Electronics*, 55(4):2034–2039, 2009.

[121] J. Steckel, A. Boen, and H. Peremans. A sonar system using a sparse broadband 3d array for robotic applications. In *2012 IEEE International Conference on Intelligent Robots and Systems*, pages 3223–3228. IEEE, 2012.

[122] Mervyn Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)*, 36(2):111–133, 1974.

[123] Joseph Sutlive and Rolf Müller. Dynamic echo signatures created by a biomimetic sonar head. *Bioinspiration & biomimetics*, 14(6):066014, 2019.

[124] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[125] Gunnar Taraldsen, Tor Arne Reinen, and Tone Berg. The underwater gps problem. In *OCEANS 2011 IEEE-Spain*, pages 1–8. IEEE, 2011.

[126] Wibke Thies, Elisabeth KV Kalko, and Hans-Ulrich Schnitzler. The roles of echolocation and olfaction in two neotropical fruit-eating bats, carollia perspicillata and c. castanea, feeding on piper. *Behavioral Ecology and Sociobiology*, 42(6):397–409, 1998.

[127] J. A. Thomas, C. F. Moss, and M. Vater. *Echolocation in bats and dolphins*. University of Chicago Press, Chicago, IL, 2004.

[128] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global, 2010.

[129] A. Tsoar, R. Nathan, Y. Bartan, A. Vyssotski, G. Dell'Omo, and N. Ulanovsky. Large-scale navigational map in a mammal. *Proc. Natl. Acad. Sci. U.S.A.*, 108(37):E718–E724, 2011.

[130] Zennure Ucar, Pete Bettinger, Steven Weaver, Krista L Merry, and Krisha Faw. Dynamic accuracy of recreation-grade gps receivers in oak-hickory forests. *Forestry: An International Journal of Forest Research*, 87(4):504–511, 2014.

[131] Karthikeyan Umapathy, Sridhar Krishnan, and Raveendra K Rao. Audio signal feature

extraction and classification using local discriminant bases. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(4):1236–1246, 2007.

[132] D. Vanderelst, J. Steckel, A. Boen, H. Peremans, and M. W. Holderied. Place recognition using batlike sonar. *eLife*, 5:e14188, 2016.

[133] G. von der Emde and H. Schnitzler. Classification of insects by echolocating greater horseshoe bats. *J. Comp. Physiol. A.*, 167(3):423–430, 1990.

[134] Sonia Waharte and Niki Trigoni. Supporting search and rescue operations with uavs. In *2010 international conference on emerging security technologies*, pages 142–147. IEEE, 2010.

[135] R. Wang and R. Müller. Biomimetic solution to finding passageways in foliage with sonar. *Bioinspir. Biomim.*, 16(6):066022, November 2021. doi: 10.1088/1748-3190/ac2aff.

[136] Ruihao Wang, Yimeng Liu, and Rolf Müller. Detection of passageways in natural foliage using biomimetic sonar. *Bioinspiration & Biomimetics*, 17(5):056009, 2022.

[137] Carl Wellington, Aaron Courville, and Anthony Stentz. A generative model of terrain for autonomous navigation in vegetation. *The International Journal of Robotics Research*, 25(12):1287–1304, 2006.

[138] Stefan B Williams, Paul Newman, Julio Rosenblatt, Gamini Dissanayake, and Hugh Durrant-Whyte. Autonomous underwater navigation and control. *Robotica*, 19(5): 481–496, 2001.

[139] K. M. Wurm, R. Kümmerle, C. Stachniss, and W. Burgard. Improving robot navigation in structured outdoor environments by identifying vegetation from laser data. In *2009*

IEEE International Conference on Intelligent Robots and Systems, pages 1217–1222. IEEE, 2009.

[140] Yasufumi Yamada, Kentaro Ito, Arie Oka, Shinichi Tateiwa, Tetsuo Ohta, Ryo Kobayashi, Shizuko Hiryu, and Yoshiaki Watanabe. Obstacle-avoidance navigation by an autonomous vehicle inspired by a bat biosonar strategy. In *Conference on Biomimetic and Biohybrid Systems*, pages 135–144. Springer, 2015.

[141] Y. Yovel, P. Stilz, M. O. Franz, A. Boonman, and H. Schnitzler. What a plant sounds like: the statistics of vegetation echoes as received by echolocating bats. *PLOS Comput. Biol.*, 5(7):e1000429, 2009.

[142] Frank E Zachos. De wilson and ra mittermeier (chief editors): Handbook of the mammals of the world. vol. 9. bats., 2020.

[143] L. Zhang and R. Müller. Large-scale recognition of natural landmarks with deep learning based on biomimetic sonar echoes. *Code Ocean https://codeocean.com/capsule/9456203/tree*, 2021.

[144] Liujun Zhang and Rolf Mueller. Large-scale recognition of natural landmarks with deep learning based on biomimetic sonar echoes. *Bioinspiration & Biomimetics*, 2022.

[145] Liujun Zhang, Ananya Bhardwaj, Michael Goldsworthy, and Rolf Müller. Deep learning of biosonar landmarks for navigation in forest environments. *The Journal of the Acoustical Society of America*, 146(4):3025–3025, 2019.

[146] Liujun Zhang, Andrew Farabow, Pradyumann Singhal, and Rolf Müller. Deep-learning exploration of the acoustic granularity of bat habitats. *The Journal of the Acoustical Society of America*, 150(4):A201–A201, 2021.