Bioinspiration & Biomimetics

PAPER

Detection of passageways in natural foliage using biomimetic sonar

RECEIVED
7 March 2022

REVISED
31 May 2022

ACCEPTED FOR PUBLICATION
21 June 2022

PUBLISHED
10 August 2022

Ruihao Wang[1], Yimeng Liu[2] and Rolf Müller[1,*]

[1] Department of Mechanical Engineering, Virginia Tech, Blacksburg, VA 24061, United States of America
[2] Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: rolf.mueller@vt.edu

Keywords: biosonar, passageway detection in foliage, field robotics, deep learning, transfer learning, transparent AI

Abstract

The ability of certain bat species to navigate in dense vegetation based on trains of short biosonar echoes could provide for an alternative parsimonious approach to obtaining the sensory information that is needed to achieve autonomy in complex natural environments. Although bat biosonar has much lower data rates and spatial (angular) resolution than commonly used human-made sensing systems such as LiDAR or stereo cameras, bat species that live in dense habitats have the ability to reliably detect narrow passageways in foliage. To study the sensory information that the animals may have available to accomplish this, we have used a biomimetic sonar system that was combined with a camera to record echoes and synchronized images from 10 different field sites that featured narrow passageways in foliage. The synchronized camera and sonar data allowed us to create a large data set (130 000 samples) of labeled echoes using a teacher–student approach that used class labels derived from the images to provide training data for echo-based classifiers. The performance achieved in detecting passageways based on the field data closely matched previous results obtained for gaps in an artificial foliage setup in the laboratory. With a deep feature extraction neural network (VGG16) a foliage-versus-passageway classification accuracy of 96.64% was obtained. A transparent artificial intelligence approach (class-activation mapping) indicated that the classifier network relied heavily on the initial rising flank of the echoes. This finding could be exploited with a neuromorphic echo representation that consisted of times where the echo envelope crossed a certain amplitude threshold in a given frequency channel. Whereas a single amplitude threshold was sufficient for this in the previous laboratory study, multiple thresholds were needed to achieve an accuracy of 92.23%. These findings indicate that despite many sources of variability that shape clutter echoes from natural environments, these signals contain sufficient sensory information to enable the detection of passageways in foliage.

1. Introduction

Achieving autonomous navigation in natural environments for systems such as drones or terrestrial vehicles could enable automation of outdoor tasks that are associated with applications that include precision agriculture [1], vegetation mapping [2–4], detection of wildfires [5, 6] and wildlife tracking [7, 8]. One of the most fundamental navigational tasks that such a system would have to accomplish in these applications is the ability to find passageways in foliage [9].

In terms of sensing, the state of the art for systems that are to achieve autonomy in outdoor environments is the use of optical methods such as stereo vision or LiDAR [10–14]. However, despite their established capabilities, these sensing modalities also have their drawbacks: camera vision fails under conditions of poor visibility [15] such as darkness or fog [16], and while LiDAR operates readily in the dark it has been reported to fail in fog [17]. Besides their vulnerability to weather conditions, stereo vision, and especially LiDAR systems, also generates very high data rates (e.g. over 250 000 points per revolution at five revolution per second; HDL-64E S3 LiDAR

sensor, Velodyne LiDAR, San Jose, CA [18]) which result in a high computational cost for data handling and processing [19].

Similar to LiDAR, the conventional approach to sonar-based mapping of an environment has been based on the formation of narrow beams [20], although the beams that can be achieved with ultrasound are always much wider than those of a laser. Despite its inability to replicate the angular resolution of laser or vision systems [21], conventional sonar resolves different targets by narrow beams that produce echo returns when aimed between targets that are much lower than the returns when the sonar is on target. With this approach, large beamwidths result in poor spatial (angular) resolution which precludes the resolution of closely spaced targets and hence narrow passageways between them.

Bat species hunting prey in dense forest environments, such horseshoe bats (Rhinolophidae [22]) and Old World leaf-nosed bats (Hipposideridae [23]), can navigate using biosonar as their primary source of sensory information about their surroundings [24]. This is a noteworthy ability, since bat biosonar is characterized by even wider beamwidths ($62^\circ \pm 52^\circ$ standard deviation across a set of different species [25]) than can be found in human-made sonar (typically $<10^\circ$ [25]). Since bats have only two ears that are placed fairly close together, consideration of binaural beamforming cannot be expected to close the gap in beamwidth between engineered sonar and bat biosonar. In addition to the wide beams, the unpredictable locations of the individual leaves and other reflectors in foliage result in random echo waveforms ('clutter') that do not lend themselves readily to the detection of patterns that may be related to the presence of a passageway [26]. Finally, bats operate on streams of pulsed biosonar echoes with duty cycles ranging from 1.7% to 57.1% [27]. Given that the typical flight speed of bats ranges from 2 m s^{-1} to 3 m s^{-1} [28], and with a sonar-sensing range that is either limited by the fairly high propagation losses of ultrasound in air or the confines of the spaces a bat can operate in within a densely vegetated habitat, bats will have to detect the presence of a passageway in foliage based on a small number of echoes. As a result of this situation, it can be hypothesized that bats have evolved ways to find narrow passageways in foliage based on wide beams and a small number of unpredictable echo waveforms. If this ability could be reproduced by human-made sonar it would have a transformative impact on achieving autonomy in natural environments without limitations that are imposed by the much bulkier, slower and more energy-consuming systems that are needed for the current technical options.

An earlier study has already demonstrated the ability to detect gaps in an artificial hedge set up in the laboratory [29]. These classifications were performed by a deep neural network (DNN) model that operated on spectrogram representations of the echo signals and was able to achieve a detection accuracy of up to 99% [29]. Furthermore, a transparent artificial intelligence (AI) approach has indicated that the rising flank of the echoes conveyed most of the sensory information that the DNN used to detect the passageways in the laboratory setting. While the laboratory arrangement of artificial foliage that was studied in this previous work has resulted in echo waveforms that appeared unpredictable to the human observer, it is highly likely that they were less variable than echoes from natural foliage. Potential sources of additional variability seen in natural foliage could be a greater variability in leaf size and shape and in the leaf density of the foliage, as well as more variable foliage surfaces [26, 30–33]. The latter could be of particular importance for determining the presence of a passageway, since passageways in natural foliage could be obscured by other features in foliage surfaces with similar geometries. As in the previous work, we will attempt to distinguish foliage and gap echoes based on a simple spike code [29]. This is to serve as a test of the hypothesis that timing differences in the rising flank of the echo are the salient features that allow gap detection. If successful, this would also demonstrate that finding gaps in foliage does not require the large computational expense associated with DNNs.

The goal of the current research was to investigate whether the finding of passageways in natural outdoor environments based on biomimetic sonar echoes is possible despite all these potential sources of additional variability. Like the preceding laboratory study [29], the current work makes use of deep-learning classifiers but adds a student–teacher network approach in which camera images acquired in parallel with the sonar echoes can be used to create a large labeled data set to facilitate echo analysis. With these large labeled data sets, the deep-learning approach can be put to test with outdoor data to establish whether finding narrow passageways with biomimetic sonar in foliage is possible in the real world.

2. Methods

All echo data used in the analysis were acquired with a biomimetic sonar (figure 2) [19] that consisted of one ultrasound emitter [electrostatic loudspeaker, Series 600 open-face ultrasonic transducer, diameter 38 mm (SensComp, Livonia, MI, USA), resonance frequency around 50 kHz] and two receivers [microelectromechanical system capacitive microphones consisting of Knowles FG23629 capsules with an approximately flat response from 10 to 125 kHz integrated into a Mononmic microphone (Dodotronic, Rome, Italy)]. Each microphone was equipped with a conical horn to serve as a baffle. Over a length of 10 cm, the cones transitioned from an outer diameter of 5 cm to an

Figure 1. Examples of different outdoor settings with gaps in foliage that were used for data collection.

inner diameter of 5 mm that served as an opening for the microphone. While the biomimetic sonar was equipped with two microphones, all results below are based on recordings from one of the two microphones that was selected based on it providing a better signal-to-noise ratio.

The ultrasonic emitter was used to emit a sonar pulse with a waveform that featured a carrier with a linear downward frequency modulation from 100 kHz to 20 kHz over a duration of 2 ms. A Hamming window was used as the envelope for the pulse. Digital-to-analog conversion of the pulse waveform was carried out with a conversion rate of 400 kHz and 16 bit resolution in a microcontroller (Arduino Due, Arduino, Ivrea, Italy). The same microcontroller with the same sampling rate and resolution was also used for digitizing the received echo waveforms. The returning foliage echoes were recorded over a time window with a length of 25 ms that started with the beginning of the emitted pulse. Hence, the first 2 ms of each recording contained a recording of the direct transmission of the pulse from the emitter to the receiver, whereas the remaining 23 ms of the recording contained only echoes. A small video camera (Hero 4, GoPro Inc., San Mateo, CA, USA) was mounted vertically beneath the loudspeaker at a distance of 12 cm and was aligned with the loudspeaker's pointing direction. The camera was used to collect an image (resolution 1280 $\times$ 720 pixels) alongside each of the recorded sonar echoes.

The image and echo data were collected from natural foliage of bushes and small trees that had small gaps in them (figure 1). A total of 10 sites distributed over the Virginia Tech campus and the neighboring Virginia Tech Corporate Research Center Campus in Blacksburg, VA, USA were selected for these recordings. The woody plant species that made up the foliage at these recording sites were *Cercidiphyllum japonicum*, *Quercus ilex*, *Schefflera arboricola*, *Prunus laurocerasus*, *Photinia glabra*, *Celtis australis*, *Acer negundo*, *Buxus sempervirens* and *Magnolia grandiflora*. All plants were identified based on the reference camera images taken (online plant image classification tool Pl@ntNet [34]).

The experimental sites were sorted into two different categories based on the width of the respective gap in the foliage: narrow (from 15 cm to 20 cm) and wide (from 25 cm to 30 cm). The gaps were either 'O-shaped' with foliage on all sides (at five of the experimental sites) or 'U-shaped' with foliage on both sides and the ground at the bottom (at the remaining five experimental sites). The sonar was positioned at three different distances to the frontal surface of the foliage (1.5 m, 2 m, 2.5 m). The sonar was mounted on a tripod that was positioned at points that were spaced at intervals of 5 cm along a straight line segment of 2 m in length that was oriented parallel to the foliage surface. The starting poin was chosen so that the scan line included positions where the sonar was pointed into the gap as well as points were it was pointed at foliage. At each position along the sonar scan line, the sonar was rotated in azimuth about the tripod axis over a total distance of $\pm 5^\circ$ from the direct line of sight to the foliage with a step width of 2.5°. Across

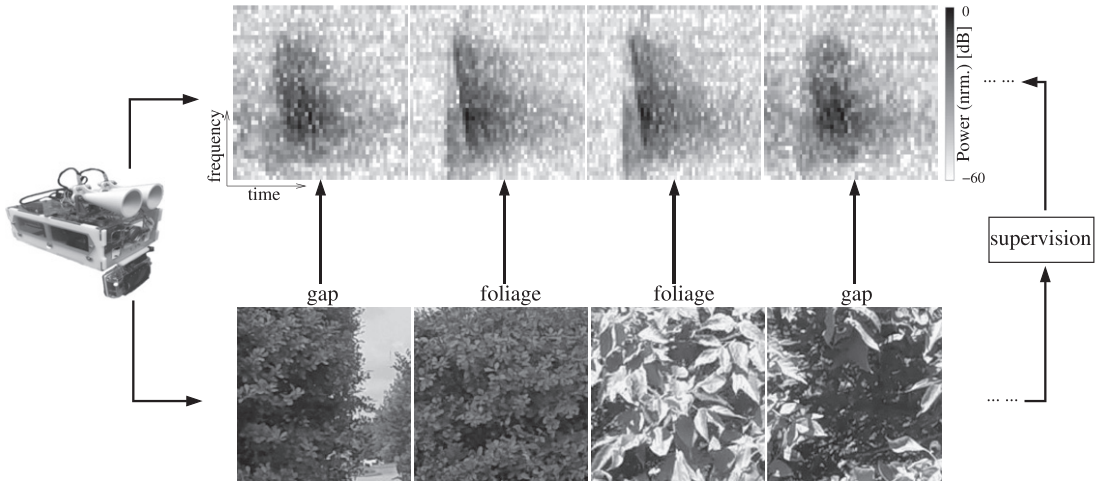

Figure 2. Data collection and teacher–student approach to training an echo classifier starting with an camera image classifier as teacher (bottom row) and the echo classifier as the student (top row).

all sites, a total of 127 390 echoes and synchronized camera images (10 sites × 41 positions × 5 rotations × 3 distances × 20 repetitions) were collected.

All camera images of foliage (with or without a gap) were cropped to a central square-shaped section that was sized to match the area illuminated by the sonar. To accomplish this, the side length of these images was determined based on the widest −6 dB beamwidth of the sonar (32° at 20 kHz). This beamwidth corresponds to a diameter of 86 cm at a distance of 1.5 m and 1.4 m at a distance of 2.5 m. Projecting these distances into the camera image plane resulted in image widths ranging from 160 to 267 pixels.

In an initial processing step, the echo recordings were cropped from their original length of 25 ms to a segment of 18 ms that included the echo but not the pulse. To determine the starting time of the echo return, the echo envelope was estimated using a sequence of local waveform maxima in sliding windows of length 0.25 ms. The maxima connected with a spline interpolation to yield the envelope estimate. As a reference for the envelope amplitude, the standard deviation of the noise was computed from the final 3 ms of each echo recording, where—at a distance of more than 4 m—no echoes were expected. The beginning of the echo was then estimated at the time when the envelope amplitude exceeded five times the standard deviation of the noise. The duration of the echo segment was chosen to cover the time of flight corresponding to the measured depth of the foliage (maximum 2.7 m, i.e. a time of flight of ∼16 ms), plus a 2 ms safety margin to include any multi-path echoes from inside the foliage.

To compute the Fourier spectrogram representations of the echo segments, different Hamming windows that varied in length from 0.5 ms (200 points) to 5 ms (2000 points) and window overlap values from 0% to 95% were tested for their potential effect on classification accuracy. The echo spectrograms were cropped to a range from 20 to 100 kHz along the frequency dimension based on the frequency content of the pulses. The classifiers either used the raw echo spectrogram amplitudes or amplitudes within a range that was clipped at −15 dB from the peak amplitude to remove background noise.

The camera image and echo data were used to train an echo-based classifier for distinguishing gaps from continuous foliage using a teacher–student model (figure 2) [35, 36]: in the first step of this process, a vision-based classifier was trained on a subset of 5000 camera images that were hand labeled and then divided into 75% training and 25% testing data. In the second step, the trained vision-based classifier was used to create a much larger set of echoes that were labeled based on the vision-based classification of the respective camera images. This vision-labelled data set contained 127 390 echoes that were—like the previous camera image data—divided into 75% training and 25% testing data.

In order to compare the current work with previous laboratory results on foliage-gap echo classification [29], two simple classifiers were tested as references for the more advanced classifiers that are being tested here for the first time in this context: the most basic classifier was based on echo energy and had been previously used as a reference for the classification of the laboratory echo data [29]. It was used here for the same purpose and made its decision based a single scalar quantity, i.e. the sum of the squares of the echo amplitude values that served as an estimate for the echo energy.

A basic convolutional neural network (CNN) was tested to provide a simple deep-learning reference for the more advanced classifiers. Like all DNNs used in the present work, the basic CNN classifier operated on echo spectrograms as its input. It consisted of five convolution layers, each followed by a max-pooling

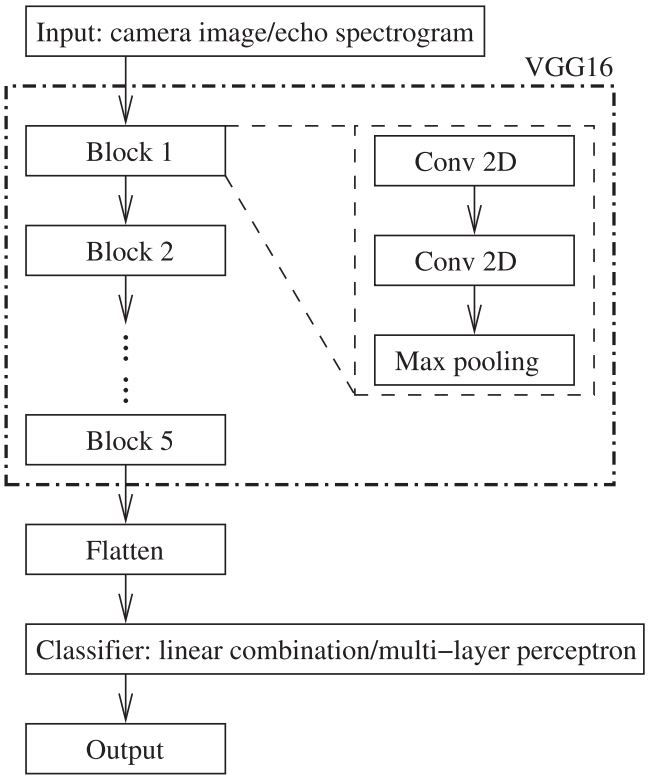

Figure 3. Transfer-learning model for camera image/echo spectrogram classification. The input of the classifier is either a camera image (for the teacher network) or an echo spectrogram (for the student network). The final stage of the network, i.e. the classifier proper, was a linear combination for the camera images and a multi-layer perceptron for the echoes spectrogram.

layer with a stride of two and a rectified linear unit (ReLU) activation function. The classification result was computed using two dense layers followed by a sigmoid activation function. The two dense layers have 1024 nodes and one node, respectively.

To assess whether a much deeper state-of-the-art CNN network could improve classification performance, a CNN based on the VGG16 architecture [37, 38] was tested. This DNN architecture was used for the classification of the camera images as well as the foliage echoes. The feature-extraction part of a VGG16 consists of a stack of five sets of convolution layers, each followed by a max-pooling layer with a stride of two along both camera image directions [37]. The number of convolution layers in each of the five sets was (in the direction of data flow) two, two, three, three and three, respectively. The convolution layers in the first set contained 64 filters each, the layers in the second set contained 128 filters each and, in the third set, there were 256 filters per layer. All subsequent convolution layers in sets four and five contained 512 filters each. After the feature extraction performed by the convolution stages, the feature maps were flattened into a vector that was used as input to three fully connected network layers with 4096, 4096 and 1000 nodes, respectively.

For camera image classification (figure 3), the weights in the feature-extraction part of the network were taken from a version of the VGG16 that had been pre-trained on camera images from the ImageNet Large Scale Visual Recognition Challenge [37, 39]. For the current work, these weights remained frozen, whereas the weights of the classifier part of the network were trained based on the hand-labelled data set of foliage images. Two different classifier parts were added to the feature-extraction part of the VGG16 network in order to map the features onto a classification decision: in the first and simplest version of the classifier part, a linear combination of weights for the features and an additive bias was used to map from the flattened feature vector onto a scalar output. The final output, i.e. an estimate for the probability of foliage, was computed by passing the output of the linear combination through a sigmoid activation function. In the second and more complex version of the classifier part, a multi-layer perceptron (MLP) [40] replaced the original fully connected layers of the original VGG16 architecture. The MLP contained three fully connected layers with 4096, 4096 and one node(s), respectively. The activation functions of the first two layers of the MLP were ReLu functions. As was the case for the simple linear-combination classification layer, the last layer of the MLP classifier used a sigmoid activation function to map the scalar output of the classifier to an estimate of the probability of the input being foliage.

In order to increase the reliability of the classifications obtained from the image-based classifiers, two-sided thresholds were applied to the probability estimates that came out of the sigmoid function. This

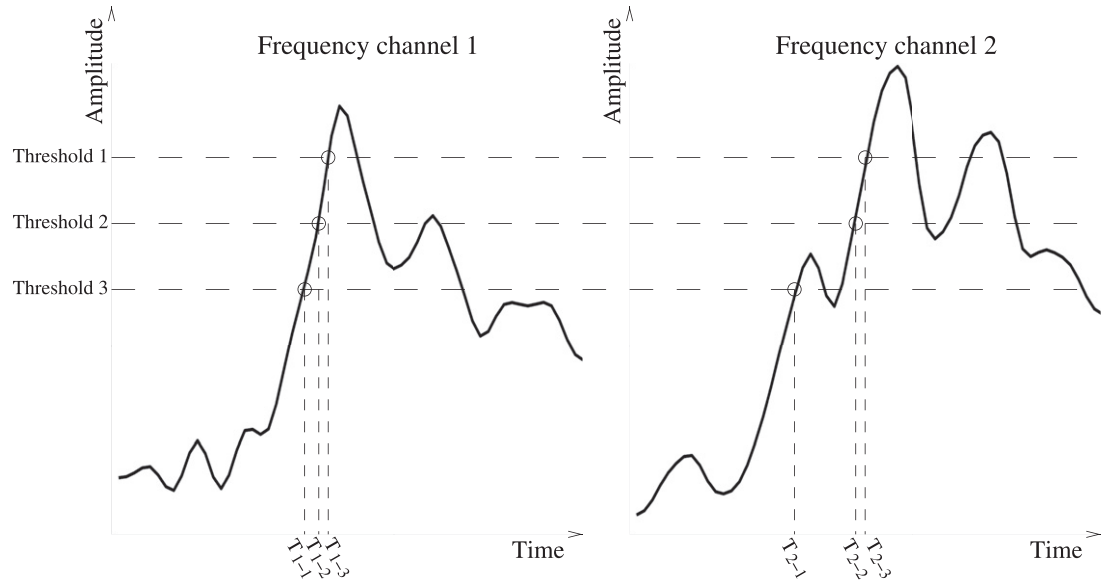

Figure 4. Neuromorphic signal representation: examples of multiple threshold-crossing times in two different frequency channels.

was done for both classifier networks tested, i.e. the linear combination as well as the MLP. Using this two-sided threshold, camera images with an estimated probability for foliage greater than or equal to 0.95 were classified as foliage and those with a probability for foliage less than 0.05 were classified as gaps. Camera images for which the estimated probability values fell between these two thresholds were discarded from the sample that was used to train the echo classifier. In this way, the echo-based classifier was only trained on data for which the image-based classifier had determined the target class with a high reliability.

For echo spectrogram classification, VGG16 feature-extraction networks were combined with the same two types of classifier parts as for the camera image classification, i.e. linear combination and MLP. However, in case of the echo spectrogram classifiers, the weights in the feature-extraction part were unfrozen and hence the weights for the feature-extraction and classification parts were being trained at the same time.

To explore which features were used by the CNNs to make the gap-versus-foliage decisions, class-activation mapping (CAM) [41] was applied to investigate whether the classifier network weighted specific regions of the echo spectrogram particularly strongly in its decision. The CAM network used for this purpose was derived from the VGG16-based echo classifier networks by replacing the classification part of the network by a global average pooling (GAP) operation [42] that produces a spatial average over the 512 feature maps that were generated in the last convolutional layer of the network. The CAM network made its target-class decisions based on the weighted sum of the averaged feature values. This sum of weighted VGG16 feature maps was projected back to the dimensions of the original echo spectrogram to indicate the relative importance of different regions of the echo spectrogram for the class decision. To facilitate this projection, the max-pooling layers were removed from the VGG16 inside the CAM network so that the final class-activation map retained the same size as the input echo spectrograms.

A simple spiking model was used to explore the suitability of a neuromorphic representation [43, 44] of the echo spectrograms for detecting passageways in foliage. The spiking model produced a one-dimensional vector that recorded the times when the echo amplitude in a certain frequency band crossed a given threshold for the first time. The threshold-crossing times were determined for 50 different frequency bands with center frequencies spaced uniformly between 20 and 100 kHz and a uniform bandwidth of 1.6 kHz. For each frequency band, the crossing times were determined for a set of 17 uniformly spaced amplitude threshold values ranging from 0.2 to 0.6 relative to the normalized output amplitude of the respective bandpass channel. The output amplitudes was normalized across the different bandpass channels into a value range between 0 and 1 by subtracting the minimum value and dividing by the difference between maximum and minimum values. The vectors consisting of the threshold-crossing times were fed into either a linear discriminant analysis or a MLP consisting of three layers with 64, 32 and one node(s), respectively, to carry out the foliage-versus-gap classification (figure 4).

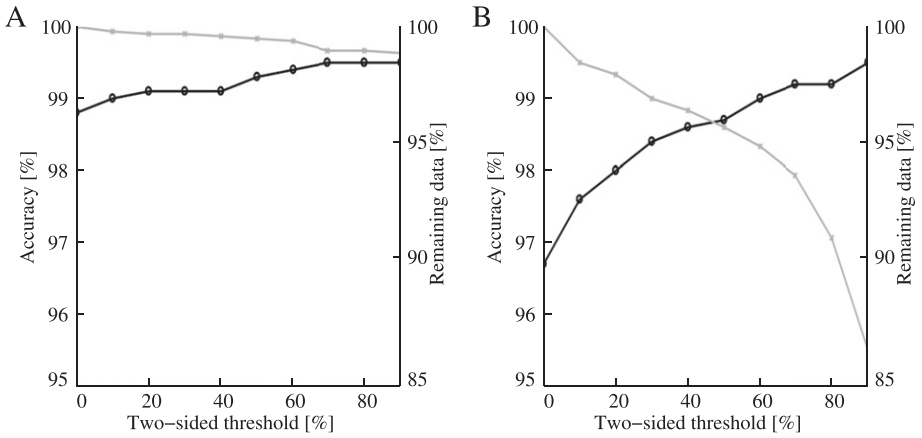

Figure 5. Relationship between the accuracy achieved by image-based gap-versus-foliage classifiers and the portion of data retained as a function of the two-sided threshold value: (a) linear-combination classifier and (b) multi-layer perceptron classifier. In each graph, the black line shows accuracy (left-hand axis) and the gray line shows the portion of data retained (right-hand axis). The location of the two-sided threshold is given by the distance between the lower and upper threshold with the center as 50% on a per cent scale.

3. Results

The accuracy achieved by the image-based classifiers and the portion of data that was retained by the two-sided-threshold mechanism depended on the threshold value and the type of classifier network used (figure 5). As the region where samples were rejected increased to 90% of the full range, the accuracy on the retained samples increased to 99.5% for the linear-combination and 99.6% for the MLP classifier. While the two classifier types yielded similar classification performance they differed in the portion of the camera image samples that were discarded due to low confidence in their classification. For the linear-combination classifier, the portion of discarded samples was 1.7% at 90% of the probability range being included by the two-sided thresholds (figure 5(a)). For the MLP classifier, this portion of rejected data was much higher, at 13.1% (figure 5(b)). Based on these results, a linear-combination classifier with a two-sided threshold enclosing 90% of the range was adopted for creating the automatically labeled data set for training the echo classifier.

The classification performance of the energy-based classifier was characterized by the area under the receiver operating characteristic (ROC) curve (AUC) (figure 6). The AUC was found to be 68.8% when tested on echo data compounded across all 10 test sites (figure 6(a)). While well above chance level (i.e. 50%), this level of performance would generate numbers of false alarms and/or misses that would be irreconcilable with highly reliable navigation. In addition, the energy-based classifier showed poor generalization across the different field sites, as evident from a wide scatter in the ROC curve estimates for the individual data sets belonging to each of the field sites (figure 6(b)). The difference between the best performance (AUC 92.8%) and the worst performance (AUC 67.0%) found across sites was 25.8% (figure 6(b)).

The simple five-layer CNN was found to have a much better performance than the energy classifier, achieving an AUC of 96.9% across all field sites. Using the VGG16 CNN resulted in yet another substantial improvement across all sites, with an AUC of 99.2%. In addition, the VGG16 classifier displayed much less variation across the different field sites (minimum AUC 89.8%, maximum 100.0%) than the energy-based classifier (figure 6(b)).

The classification accuracy achieved by the VGG16 CNN classifier was found to depend on the overlap parameter of the spectrogram representation of the echoes. The accuracy tended to increase with increasing overlap as long as the overlap was below approximately 50% of the window length. For overlap values above 50%, the classifier accuracy saturated at its maximum value (figure 7(a)). Unlike the window overlap, varying the window length from 200 points (0.5 ms) to 2000 points (5 ms) did not affect the classification performance in any systematic fashion. The slight variations observed were contained within a range of 1.3% (from 95.4% to 96.7%; figure 7(b)). Based on these findings, a 256-point Hamming window with a 243-point overlap was adopted to compute the Fourier spectrograms of the echoes that were used to compare the VGG16 CNN classifier with the other spectrogram-based classifiers (table 1).

The CAM classifier performed slightly less well than the full VGG16 CNN classifier from which it was derived, with an AUC of 96.3%. The MLP classifier based on the crossing times for multiple amplitude thresholds (17 thresholds, from 0.2 to 0.6, with a step with of 0.025) delivered a performance (AUC 97.5% over all sites) that fell between the results for the VGG16 and class-activation mapping classifiers (figure 6(a)). The results from the class-activation mapping indicated that the CNN classifiers

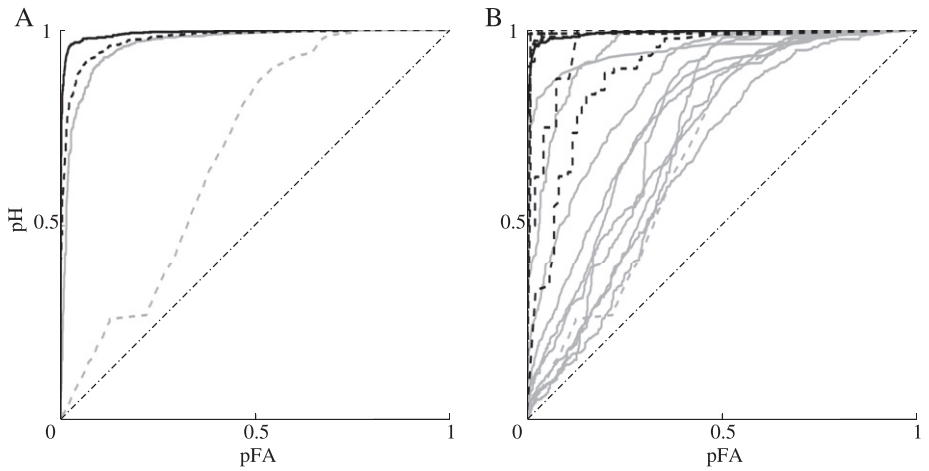

Figure 6. Receiver operating characteristic curves for the detection of gaps in foliage. (a) Classification performance across all 10 field sites with different classifiers: pre-trained VGG16 CNN (solid black), class-activation mapping (solid gray) based on the echo spectrograms, multi-layer perceptron (MLP) based on multiple threshold-crossing times (dashed black), classification based on total energy (dashed gray). (b) Classification performance for the 10 individual field sites: echo-energy classifier (field sites, solid gray lines) pre-trained VGG16 CNN validated in different sites (dashed black lines), pre-trained VGG16 CNN (across all sites, solid black line).

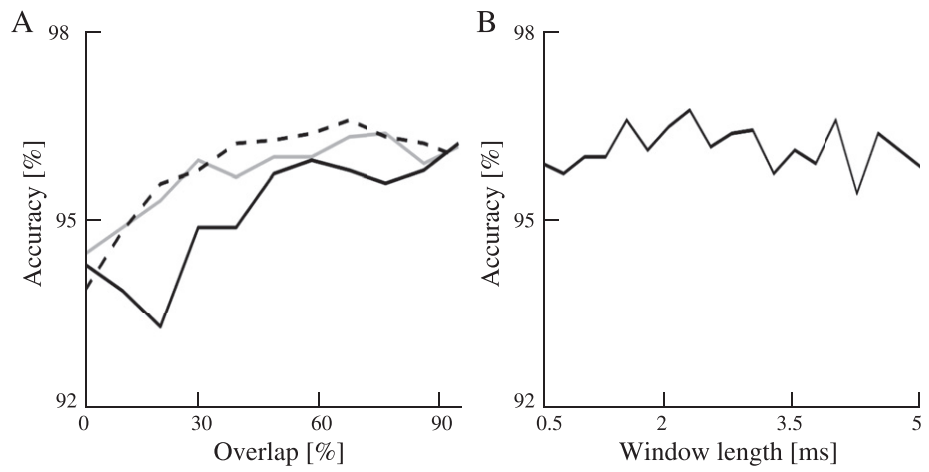

Figure 7. Relationship between classification accuracy and the spectrogram parameters window length and overlap for the VGG16 CNN classifier. (a) Classification accuracy as a function of overlap, with curves for the shortest (0.5 ms, solid gray), a medium (2.75 ms, dashed black) and the longest (5 ms, solid black) window length tested. (b) Classification accuracy as a function of window length for the highest overlap tested (95%).

relied heavily on a narrow region of the echo spectrograms that was closely aligned with the rising flank of the echo (figure 8).

For all CNN classifiers, the provision of echo spectrograms thresholded in amplitude at a level of −15 dB down from the maximum value to remove low-amplitude noise resulted in better performance than that achieved when the raw spectrograms were used as inputs. This was true across all three deep learning approaches with the difference in the mean accuracy being 1.43% on average (table 1). All three differences between raw and threshold echo spectrograms were statistically significant (t-test with Bonferroni correction, any of the t-tests has $p < 0.01$, 50 accuracy values obtained via 10 repeats and five-fold cross-validation).

For the neuromorphic classifiers, echo representations utilizing multiple threshold-crossing times were found to result in better performance than that obtained when single-threshold-crossing-time representations were fed into same classifier (i.e. LDA or MLP). The average difference in accuracy between single- and multiple-threshold representations was 6.18% (table 1). The MLP-based classifiers yielded a better performance than the LDA-based ones when using the same input representations (i.e. single-threshold or multiple-threshold). All accuracy differences seen between the different neuromorphic representations (i.e. single/multiple threshold(s)) and classifier portions (i.e. MLP/LDA) were found to be statistically highly significant (t-test with Bonferroni

Table 1. Classification accuracies (%) of difference CNN models and inputs represented by mean values and the respective standard deviations estimated based on 10 repeats and five-fold cross-validation.

Model	Spec.	Raw echo spectrogram	Thresholded echo spectrogram
CNN	Five-layer	91.53 ± 0.49	93.21 ± 0.72
	Pre-trained	95.77 ± 0.32	96.64 ± 0.44
	CAM	89.68 ± 0.41	91.42 ± 0.49
Neuromorphic	Single threshold + MLP	85.98 ± 0.86	
	Single threshold + LDA	81.23 ± 0.89	
	(Linear Discriminant Analysis)		
	Multiple thresholds + MLP	92.23 ± 0.52	
	Multiple thresholds + LDA	87.34 ± 0.74	

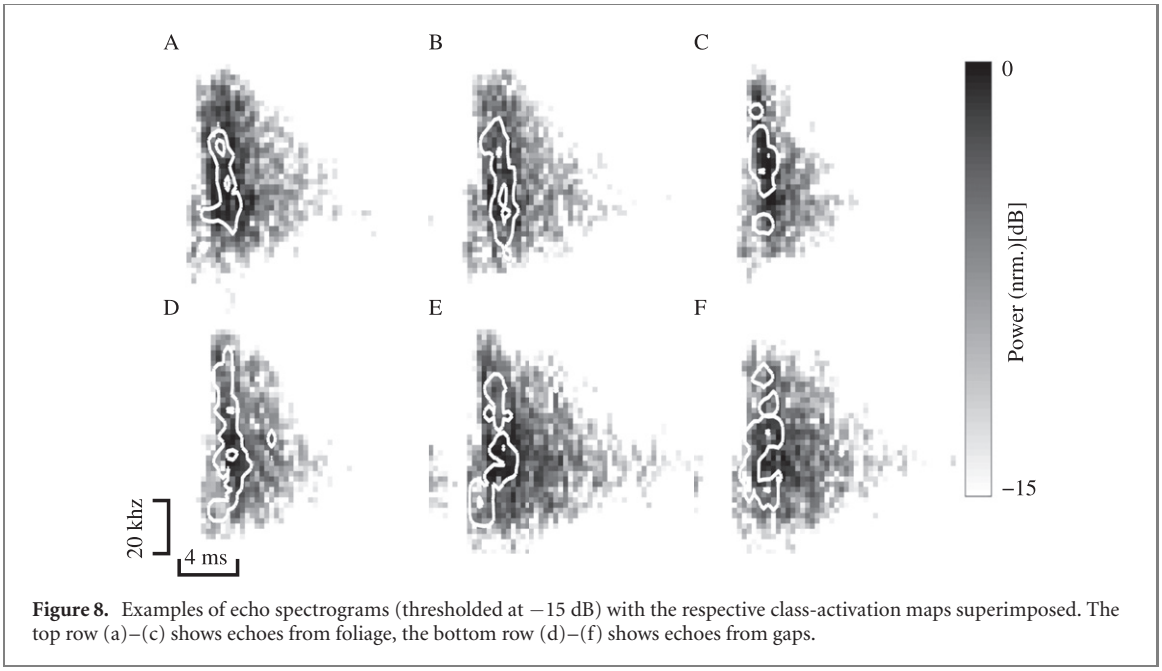

Figure 8. Examples of echo spectrograms (thresholded at −15 dB) with the respective class-activation maps superimposed. The top row (a)–(c) shows echoes from foliage, the bottom row (d)–(f) shows echoes from gaps.

correction, any of the t-tests had $p < 0.01$, 50 accuracy values obtained via 10 repeats and five-fold cross-validation).

4. Discussion

The field echoes studied here are likely to contain a much larger variability than the echoes from laboratory experiments that have been previously reported [29]. This additional variability could pose a challenge to reliably extracting the presence of a passageway from the echoes and could even render the laboratory methods useless under field conditions. However, despite this expected greater variability, the methods previously developed in a laboratory setup [9, 29] were found to be effective in the field as well. The DNN-based classifiers operating on echo spectrogram representations delivered an almost identical performance to what had previously been achieved in the laboratory (an accuracy of ∼97% in the laboratory [29] as well as in the field). However, the DNN-based classifiers had to be improved by using a deeper feature extraction portion of the network to maintain this performance. This demonstrates that the DNN methods had sufficient flexibility to deal with addition of variability in the field data. The need for a deeper network could be seen as an indication that detecting a passageway in the field data indeed poses a greater challenge than in the laboratory data.

As in the laboratory work, the transparent AI approach pointed at the rising flank of the field echoes as the main source of relevant sensory information. This can be seen as an indication that the information-bearing features of the field echoes are similar—or at least similarly positioned in time and frequency—to those in the laboratory echoes. Further circumstantial evidence regarding the nature of the informative echo features can be derived from the experiments that used different window lengths for computing the echo spectrograms. The finding that classification performance was virtually independent of window length over a the entire range of values tested (0.5–5 ms), demonstrates that the informative echo features are accessible in the time domain and the frequency domain, as well as in various mixtures of the two.

However, the informative features of the field echoes were not quite as accessible as those of the laboratory echoes for the classifiers that operated on the neuromorphic echo representation derived from a single amplitude threshold. For these classifiers, detection accuracy dropped from 95% on the laboratory data [29] down to 81% in the field. This decrease in performance could be addressed—at least partially—by extending the neuromorphic echo representation to multiple amplitude thresholds; this resulted in a best accuracy of 92%.

With respect to understanding the biosonar system of bats, the current findings demonstrate that bats flying in outdoor environments have accurate and reliable information regarding the presence of passageways that are much narrower than their biosonar beams. This suggests that beamwidth may not necessarily be a critical parameter for explaining the navigational abilities of bats in natural vegetation. However, the results of the current work can only provide experimental evidence with regard to the sensory information that bats—hypothetically—could use for navigation in foliage. They do not allow us to draw any conclusions as to if and how bats use this information. These questions can only be addressed with animal experiments that were entirely beyond the scope of the current study. For the design of autonomous drones that can operate in complex natural environments, it may hence also not be necessary to focus on sensors with high angular resolution, since obtaining at least passageway information does not seem to depend on resolution (within the tested range).

While passageway finding is an important component of navigation [9], it is not the only ability that is needed for successful navigation and goal attainment of an autonomous system. Recent work has demonstrated the presence of location-specific information in biomimetic clutter echoes that could be used to pinpoint a location on a large (kilometer) scale [45, 46] as well as on small (meter) scale [47]. This location information could very well support the constructions of maps, although this has yet to be demonstrated. Combined with the passageway-finding ability demonstrated in the current work, echo-based location identification could make for a complete navigation system that is not only capable of avoiding obstacles but can also determine which path to take to reach a destination within an area it has mapped.

Future work on how bats and biomimetic systems could use clutter echoes for navigation should be extended to a wider range of natural habitats to survey more of the variability in potentially informative echo features that these environments produce. It would also be worthwhile to study the abilities and strategies that bats use in their habitats, for example how the animals' flight trajectories are aligned with their respective microhabitats. Combined with further insights into how bats represent the information from their biosonar systems in three dimensions [48], these findings could lead to novel bioinspired navigation systems that could provide a higher performance at much lower data rates than the approaches that are currently available to engineers in the field.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Ruihao Wang https://orcid.org/0000-0002-9046-0662
Rolf Müller https://orcid.org/0000-0001-8358-4053

References

[1] Alsalam B H Y, Morton K, Campbell D and Gonzalez F 2017 Autonomous UAV with vision based on-board decision making for remote sensing and precision agriculture *2017 IEEE Aerospace Conf.* (IEEE) pp 1–12
[2] Xie Y, Sha Z and Yu M 2008 Remote sensing imagery in vegetation mapping: a review *J. Plant. Ecol.* **1** 9–23
[3] Salamí E, Barrado C and Pastor E 2014 UAV flight experiments applied to the remote sensing of vegetated areas *Remote Sens.* **6** 11051–81
[4] Anderson K and Gaston K J 2013 Lightweight unmanned aerial vehicles will revolutionize spatial ecology *Front. Ecol. Environ.* **11** 138–46
[5] Yuan C, Zhang Y and Liu Z 2015 A survey on technologies for automatic forest fire monitoring, detection, and fighting using unmanned aerial vehicles and remote sensing techniques *Can. J. For. Res.* **45** 783–92
[6] Restas A 2015 Drone applications for supporting disaster management *WJET* **3** 316
[7] Gonzalez L F, Montes G A, Puig E, Johnson S, Mengersen K and Gaston K J 2016 Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation *Sensors* **16** 97
[8] Linchant J, Lisein J, Semeki J, Lejeune P and Vermeulen C 2015 Are unmanned aircraft systems (UASs) the future of wildlife monitoring? A review of accomplishments and challenges *Mamm. Rev.* **45** 239–52
[9] Wang R, Jin X, Wang H and Müller R 2018 The problem of finding passageways in foliage with biomimetic sonar *J. Acoust. Soc. Am.* **143** 1727
[10] Campos M B, Tommaselli A M G, Honkavaara E, Prol F d S, Kaartinen H, El Issaoui A and Hakala T 2018 A backpack-mounted omnidirectional camera with off-the-shelf navigation sensors for mobile terrestrial mapping: development and forest application *Sensors* **18** 827
[11] Wallace L, Lucieer A, Watson C and Turner D 2012 Development of a UAV-LiDAR system with application to forest inventory *Remote Sens.* **4** 1519–43
[12] Merino L, Caballero F, Martínez-de Dios J R, Maza I and Ollero A 2012 An unmanned aircraft system for automatic forest fire monitoring and measurement *J. Intell. Robot. Syst.* **65** 533–48
[13] Tang L and Shao G 2015 Drone remote sensing for forestry research and practices *J. For. Res.* **26** 791–7

[14] Adão T, Hruška J, Pádua L, Bessa J, Peres E, Morais R and Sousa J 2017 Hyperspectral imaging: a review on UAV-based sensors, data processing and applications for agriculture and forestry *Remote Sens.* **9** 1110
[15] Chen C, Chen Q, Xu J and Koltun V 2018 Learning to see in the dark *Proc. IEEE Conf. on Comput. Vis. Pattern Recognit.* pp 3291–300
[16] Oakley J P and Satherley B L 1998 Improving image quality in poor visibility conditions using a physical model for contrast degradation *IEEE Trans. Image Process.* **7** 167–79
[17] Rasshofer R H, Spies M and Spies H 2011 Influences of weather phenomena on automotive laser radar systems *Adv. Radio Sci.* **9** 49–60
[18] Ashraf I, Hur S and Park Y 2017 An investigation of interpolation techniques to generate 2D intensity image from LiDAR data *IEEE Access* **5** 8250–60
[19] Wang H, Wang C, Chen C and Xie L 2021 F-loam: fast LiDAR odometry and mapping *IEEE Int. Conf. Intell. Robots Syst.* (IEEE) pp 4390–6
[20] Fazli S and Kleeman L 2006 Simultaneous landmark classification, localization and map building for an advanced sonar ring *Robotica* **25** 283–96
[21] Achtelik M, Bachrach A, He R, Prentice S and Roy N 2009 Stereo vision and laser odometry for autonomous helicopters in GPS-denied indoor environments *Proc. SPIE* **7332** 733219
[22] Neuweiler G, Metzner W, Heilmann U, Rübsamen R, Eckrich M and Costa H H 1987 Foraging behaviour and echolocation in the rufous horseshoe bat (*Rhinolophus rouxi*) of Sri Lanka *Behav. Ecol. Sociobiol.* **20** 53–67
[23] Pavey C, Grunwald J-E and Neuweiler G 2001 Foraging habitat and echolocation behaviour of Schneider's leafnosed bat, *Hipposideros speoris*, in a vegetation mosaic in Sri Lanka *Behav. Ecol. Sociobiol.* **50** 209–18
[24] Pavey C R 2021 Comparative echolocation and foraging ecology of horseshoe bats (Rhinolophidae) and old world leaf-nosed bats (Hipposideridae) *Aust. J. Zool.* **68** 382
[25] Todd B D and Müller R 2017 A comparison of the role of beamwidth in biological and engineered sonar *Bioinsp. Biomim.* **13** 016014
[26] Müller R and Kuc R 2000 Foliage echoes: a probe into the ecological acoustics of bat echolocation *J. Acoust. Soc. Am.* **108** 836–45
[27] Fenton M B, Faure P A and Ratcliffe J M 2012 Evolution of high duty cycle echolocation in bats *J. Exp. Biol.* **215** 2935–44
[28] Tian X *et al* 2006 Direct measurements of the kinematics and dynamics of bat flight *Bioinsp. Biomim.* **1** S10
[29] Wang R and Müller R 2021 Bioinspired solution to finding passageways in foliage with sonar *Bioinsp. Biomim.* **16** 066022
[30] Ming C, Gupta A K, Lu R, Zhu H and Müller R 2017 A computational model for biosonar echoes from foliage *PLoS One* **12** e0182824
[31] Kuc R 2002 Binaural sonar electronic travel aid provides vibrotactile cues for landmark, reflector motion and surface texture classification *IEEE Trans. Biomed. Eng.* **49** 1173–80
[32] McKerrow P and Harper N 2001 Plant acoustic density profile model of CTFM ultrasonic sensing *IEEE Sensors J.* **1** 245–55
[33] Harper N and McKerrow P 2001 Recognising plants with ultrasonic sensing for mobile robot navigation *Robot. Autonom. Syst.* **34** 71–82
[34] Joly A *et al* 2016 A look inside the Pl@ntNet experience *Multimed. Syst.* **22** 751–66
[35] Ba L J and Caruana R 2013 Do deep nets really need to be deep? (arXiv:1312.6184)
[36] Gupta S, Hoffman J and Malik J 2016 Cross modal distillation for supervision transfer *Proc. IEEE Conf. on Comput. Vis. Pattern Recognit.* pp 2827–36
[37] Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale image recognition (arXiv:1409.1556)
[38] Zhong M, Castellote M, Dodhia R, Lavista F J, Keogh M and Brewer A 2020 Beluga whale acoustic signal classification using deep learning neural network models *J. Acoust. Soc. Am.* **147** 1834–41
[39] Olga R *et al* 2015 ImageNet large scale visual recognition challenge *Int. J. Comput. Vis.* **115** 211–52
[40] Rosenblatt F 1961 Principles of neurodynamics. Perceptrons and the theory of brain mechanisms *Technical Report* Buffalo, NYCornell Aeronautical Lab Inc.
[41] Zhou B, Khosla A, Lapedriza A, Oliva A and Torralba A 2016 Learning deep features for discriminative localization *Proc. IEEE Conf. on Comput. Vis. Pattern Recogit.* pp 2921–9
[42] Lin M, Chen Q and Yan S 2013 Network in network (arXiv:1312.4400)
[43] Khyam M O, Alexandre D, Bhardwaj A, Wang R and Müller R 2019 Neuromorphic computing for autonomous mobility in natural environments *NICE '19: Proc. 7th Ann. Neroinspired Comput. Elements Workshop* pp 1–7
[44] Gerstner W and Kistler W M 2002 *Spiking Neuron Models: Single Neurons, Populations, Plasticity* (Cambridge: Cambridge University Press)
[45] Zhang L and Müller R 2022 Large-scale recognition of natural landmarks with deep learning based on biomimetic sonar echoes *Bioinsp. Biomim.* **17** 026011
[46] Müller R, Chakrabarti S, Eshera I M, Lagad S V, Wang R and Zhang L 2021 Autonomy, soft-robotics, deep learning, and bat biosonar *J. Acoust. Soc. Am.* **150** A325
[47] Zhang L, Farabow A, Singhal P and Müller R 2021 Deep-learning exploration of the acoustic granularity of bat habitats *J. Acoust. Soc. Am.* **150** A201
[48] Kothari N B, Wohlgemuth M J and Moss C F 2018 Dynamic representation of 3D auditory space in the midbrain of the free-flying echolocating bat *Elife* **7** e29053