Biological Conservation 241 (2020) 108269

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Automatic standardized processing and identification of tropical bat calls using deep learning approaches

Xing Chen[a,1], Jun Zhao[b,1], Yan-hua Chen[a], Wei Zhou[b,*], Alice C. Hughes[a,*]

[a] Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun 666303, China
[b] Software School, Yunnan University, Kunming 650500, China

ARTICLE INFO

Keywords:
Bats
Bioacoustics
Automated monitoring
Algorithms
Deep learning
Neural network
Automatic processing
Biodiversity metrics
Machine learning
Calls
Echolocation
Monitoring protocol

ABSTRACT

Consistent and comparable metrics to automatically monitor biodiversity across the landscape remain a gold-standard for biodiversity research, yet such approaches have frequently been limited to a very small selection of species for which visual approaches (e.g., camera traps) make continuous monitoring possible. Acoustic-based methods have been widely applied in the monitoring of bats and some other taxa across extended spatial scales, but are have yet to be applied to diverse tropical communities.

In this study, we developed a software program "Waveman" and prepared a reference library using over 880 audio-files from 36 Asian bat species. The software incorporated a novel network "BatNet" and a re-checking strategy (ReChk) to maximize accuracy. In Waveman, BatNet outperforms three other published networks: CNN_{FULL}, VggNet and ResNet_v2, with over 90% overall accuracy and 0.94 AUC on the ROC plot. The classification accuracy rates for all 36 species are at least 86% when analysed in combination. Moreover, our library preparation and ReChk greatly improved the sensitivity and reduced the false positive rate, when tested with 15 species for which more detailed and situationally diverse records were available. Finally, BatNet was successfully used to identify *Hipposideros larvatus* and *Rhinolophus siamensis* from three different environments. We hope this pipeline is useful tool to process bioacoustic data accurately, effectively and automatically, therefore allowing for greater standardization and comparability for researchers to understand bat activities across space and time and therefore provide a consistent tool for monitoring biodiversity for management and conservation.

1. Introduction

With the growing global awareness of the need to maintain biodiversity, and the development of new targets for countries to maintain their biodiversity (e.g., the Convention on Biological Diversity, the Conference of the Parties and Aichi targets, Marques et al., 2014) new mechanisms and metrics to quantify and monitor biodiversity are needed. There is also a collective acknowledgement of the need for the development of consistent standards to monitor biodiversity across space and time (Gasc et al., 2013; Walters et al., 2013; Proença et al., 2017). The lack of long-term datasets for most regions precludes measures of past biodiversity change and development of new methods to monitor current biodiversity provides a means to not only develop conservation priorities (Meyer et al., 2010; Christin et al., 2018), but also to measure the performance of different regions in maintaining biodiversity over shorter timescales, and even understanding the ecology of individual species (Hughes et al., 2012; Cardinale et al., 2018). Visual encounter methods such as the use of camera trapping has allowed continuous monitoring of a subset of species for decades (Trolle and Kéry, 2003; Rich et al., 2017), but it is taxonomically limited and time consuming because of manual processing (Rich et al., 2017) though automated techniques are relatively widely implemented (Yu et al., 2013). These limitations pose challenges to standardising visual methods for biodiversity monitoring, especially for species where predicting the direction of movement and thus best camera orientation is challenging, or which are too small or cryptic for such approaches to be applicable (Barratt et al., 1997; Newey et al., 2015). The use of bioacoustic methods can ameliorate these issues by vastly increasing the number and range of species that can be monitored, and the lack of directional recording bias in many microphones reduces random bias associated with camera traps (Gasc et al., 2013). However, until this point the application of such approaches for mainstream monitoring and management is limited. The application of such approaches has been further hampered by the challenge of identifying species in

* Corresponding authors.
E-mail addresses: zwei@ynu.edu.cn (W. Zhou), ACHughes@xtbg.cas.cn (A.C. Hughes).
[1] Contributed equally.

https://doi.org/10.1016/j.biocon.2019.108269
Received 16 May 2019; Received in revised form 21 September 2019; Accepted 27 September 2019
Available online 18 November 2019
0006-3207/

complex communities of tropical areas, such as Southeast Asia (Astaras et al., 2017). Thus, the automation of such techniques holds significant promise as a monitoring tool for management and conservation (Gasc et al., 2013).

As visual monitoring only serves such a small subset of species, additional more inclusive methods are needed to accurately identify other species to allow effective biodiversity monitoring. For species which produce species-specific calls scientists need a method to identify calls to the highest available taxonomic resolution based on call structure (Russo and Voigt, 2016). Large numbers of publications document-methodological developments and advances for identification of acoustic data of diverse selection of bat species based on call libraries (ZINGG, 2019; Parsons and Jones, 2000; Russo and Jones, 2002; Gager et al., 2016; Meagher et al., 2018). As species call structure is closely linked to species phylogeny and evolution (Altes and Titlebaum, 1970; Boonman and Schnitzler, 2005; Clement et al., 2014; Gager et al., 2016), analysis allows the identification of species based on calls alone, and using machine learning can also capture intra-specific variation (Clement et al., 2014; Russo et al., 2017). Basic methods for call based species identification are normally based on a subset of acoustic features, including frequency of maximum energy, time duration, bandwidth, and sweep rate (Rydell et al., 2017), which limits the use and value of calls as a reference in species with similar calls unless more detailed approaches are utilized, as machine learning approaches can automatically extract millions of features based on a library without filtering background noise (Mac Aodha et al., 2018).

Artificial intelligence algorithms such as deep learning, enables ecologists and evolutionary biologists to look at bio-sonar at hitherto impossible resolutions. Such a tool can be applied to identify various species based on their calls including bats (Mac Aodha et al., 2018), birds (Stowell et al., 2019), frogs (Hill et al., 2018), and mosquitos (Kiskin et al., 2018). A relatively simple network (CNN_{FULL}) has been successfully applied to localise European bats based on the presence of their calls in a large volume of data (Mac Aodha et al., 2018), significantly outperforming Random Forest algorithms. However, species identification using the network CNN_{FULL} has not previously been tested.

In this study, we developed a free open source software Waveman to build a model for automatically identifying 36 Tropical bat species from South Eastern Asia (where overall diversity for these areas is unknown due to probable cryptic species, but is estimated at over 320 species (Hughes et al., 2012). We tested models using both filtered and unfiltered data under a variety of conditions to ensure that the models could accurately be applied to natural situations (Fig. 1).

2. Materials and methods

2.1. Modelling

2.1.1. The development of the reference acoustic library

We surveyed a range of habitats and recorded 817 audio-files from 678 individuals of 35 bat species across Thailand between 2003 to 2009 and Xishuangbanna in China from 2017 to 2018 (Table 1). Among them, 353 audio-files from 218 individuals of 24 species were collected from rainforest, limestone forest, cave, and urban areas in Xishuangbanna (China); 504 from 469 individuals of 25 species from forest and karsts in Thailand (14 overlapped with species in China, Table 1). A Pettersson D-240X and two M500-384 (Pettersson Elektronik, Sweden) recording devices were used in Thailand and China as previously described (Hughes et al., 2011). In addition 21 audio-files from 21 individuals of four species collected in Malaysia, which extracted from a public bioacoustic database (Baker et al., 2015).

The calls used to make our reference library had two criteria. Firstly species identity was validated based on morphological measurements and in some cases molecular approaches. Secondly, to maximize intra-specific variation of bat calls, we recorded the calls under a variety of different conditions, including in bags, hand-held, hand-released, or free flying in a room. Hipposiderid species have a slight frequency change of about 2 kHz when they flew in a room; the Rhinolophid species call duration was highly variable when released, and was longer during flight than when stationary; vespertilionid species also has a slight frequency change when they flew in the room.

2.1.2. Converting images

All the selected audio-files were converted to spectral images using a python script we developed called “Coverting_Image.py” because frequency and sweep related information were recorded as images, to enable downstream operation of manually selection. First, we manually extracted signal region of audio-files only for target species; the files including more than one species with overlapping calls were clipped or removed. The selected audio-files were given a unique number to improve efficiency for downstream processing. Secondly, we subset audio-files and extracted 10,368 samples from the 384,000 generated for each second of recording by M500-384, or 22,100 by D-240X, this is to ensure equal sampling independent of recording device. The number was optimized to make each segment contain few call repeats for the majority of species. Time duration of audio segments was 0.027 s when generated by the device M500-384 with 384,000 sample number per second (0.027*384,000 = 10,368) and almost 0.47 s using the D-240X

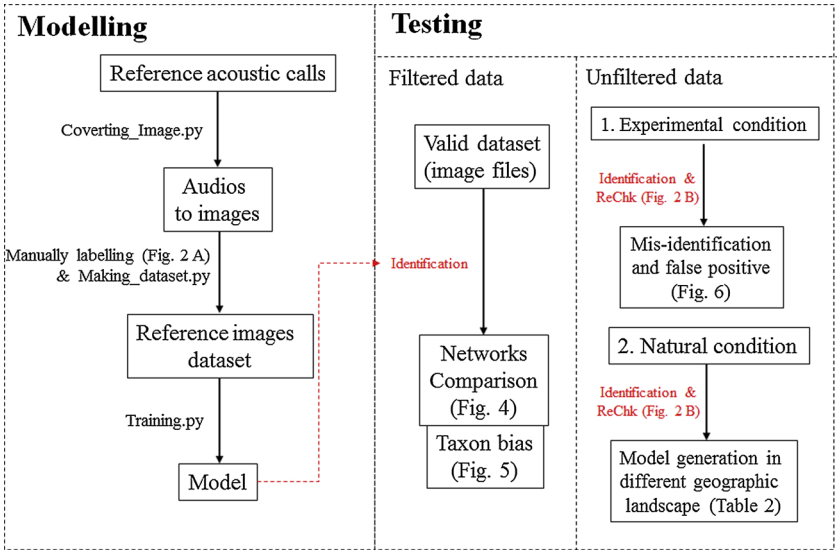


Fig. 1. Schematic diagram of pipeline. We first built model and then test the model. There are three python scripts used to for modelling, including Coverting_image.py, Making_database.py, and Training.py in a self-developed open source software Waveman. Tested process using both filtered images and unfiltered audio-files.

Table 1
Taxa summary of the 36 species and itscall information.

	Family	Species	Individual number**	No. of images***	Location	Release type
2	Hipposideridae	*Aselliscus stoliczkanus*	18(15/3)	3,709	China, Thailand	In bag, Release, Hand hold, Free fly
3	Hipposideridae	*Hipposideros armiger*	39(9/30)	3068(376)	China, Thailand	In bag, Release, Hand hold, Free fly
4	Hipposideridae	*Hipposideros bicolor*	114	4,910	Thailand	In bag, Release, Hand hold, Free fly
5	Hipposideridae	*Hipposideros cineraceus*	3	1,959(1453)	China	In bag, Release, Free fly
6	Hipposideridae	*Hipposideros diadema*	4(1/3)	1802(552)	Thailand, Malysia	Unknown
7	Hipposideridae	*Hipposideros larvatus*	40(17/21/2)	4,281	China, Thailand, Malysia	In bag, Release, Hand hold, Free fly
8	Hipposideridae	*Hipposideros lekaguli*	13	1,547(1047)	Thailand	In bag, Release, Hand hold
9	Hipposideridae	*Hipposideros pomona*	32(16/16)	3,500	China, Thailand	In bag, Release, Hand hold, Free fly
10	Hipposideridae	*Hipposideros turpis*	32	2,035(1040)	Thailand	Release, Hand hold
11	Megadermatidae	*Megaderma spasma*	22(2/20)	2,189	China, Thailand	Release, Hand hold, Free fly
12	Molossidae	*Cheiromeles torquatus*	11	2,482(1167)	Malysia	Free fly
13	Rhinolophidae	*Rhinolophus affinis*	16(1/11/5)	3010(808)	China, Thailand, Malysia	Release
14	Rhinolophidae	*Rhinolophus coelophyllus*	10	1,816(1169)	Thailand	In bag, Release
15	Rhinolophidae	*Rhinolophus lepidus*	35	2,172	Thailand	Release, Hand hold
16	Rhinolophidae	*Rhinolophus malayanus*	71(30/41)	4,790	China, Thailand	In bag, Release, Hand hold, Free fly
17	Rhinolophidae	*Rhinolophus pearsonii*	9(5/4)	3,491	China, Thailand	In bag, Release, Hand hold, Free fly
18	Rhinolophidae	*Rhinolophus pusillus*	21	2,596(1236)	Thailand	Unknown
19	Rhinolophidae	*Rhinolophus rex*	1	2,820(766)	China	In bag, Release, Hand hold
20	Rhinolophidae	*Rhinolophus robinsoni*	7	2,026(1291)	Thailand	Release, Hand hold, Free fly
21	Rhinolophidae	*Rhinolophus siamensis*	12	5,553	China	In bag, Release, Free fly
22	Rhinolophidae	*Rhinolophus sinicus*	21	4,350	China	In bag, Release, Free fly
23	Rhinolophidae	*Rhinolophus stheno*	54(19/35)	4,248	China, Thailand	In bag, Release, Hand hold, Free fly
24	Rhinolophidae	*Rhinolophus yunanensis*	3	1,933(1405)	Thailand	Release, Hand hold
25	Vespertilionidae	*Ia io*	1	2,289(1053)	China	In bag, Release
26	Vespertilionidae	*Miniopterus magnater*	5(1/4)	1447(844)	China, Thailand	In bag, Release
27	Vespertilionidae	*Hypsugo pulveratus*	3	1,715(801)	China	In bag, Release
28	Vespertilionidae	*Scotomanes ornatus*	2	2,054(953)	China	In bag, Release, Hand hold
29	Vespertilionidae	*Tylonycteris robustula*	6(3/3)	5,738	China, Thailand	In bag, Release
30	Vespertilionidae	*Tylonycteris pachypus*	1	355(328)	Thailand	unknown
31	Vespertilionidae	*Kerivoula hardwickii*	22(14/8)	3,771	China, Thailand	In bag, Release, Free fly
32	Vespertilionidae	*Phoniscus jagorii*	1	2,549(382)	China	In bag, Release
33	Vespertilionidae	*Murina tubinaris*	3	4,384	China	In bag, Release, Hand hold, Free fly
34	Vespertilionidae	*Murina cyclotis*	17(9/8)	3,448	China, Thailand	In bag, Release, Free fly
35	Vespertilionidae	*Myotis laniger*	26	2,436	China	In bag, Release
36	Vespertilionidae	*Myotis muricola*	17(13/4)	5,302	China, Thailand	In bag, Release, Free fly
37	Vespertilionidae	*Myotis siligorensis*	17	1,920(1189)	Thailand	Free fly
38		No signal*		12,537		
39		Weak*		10,535		

* except 36 species, over 10,000 images were also selected to detect no or weak signal images.
** numbers in the parentheses are individual numbers corresponding to the county in location column.
*** numbers in the parentheses are the rescaled spectral images.

with 22,100 sample number per second. These audio segmentations contained one to two signal pulses for almost all the species from Vespertilionidae, Hipposideridae and Megadermatidae. Only members of the families Rhinolophidae and *Cheiromeles torquatus* are sometimes exceptions, with pulses usually split into two parts.

Each segment was converted into a spectral image with size of 256*256 pixels using function spectrum (incorporated short-time Fourier transformation, STFT) in the Python package matplotlib (Hunter, 2007).

2.1.3. Labeling images and making image library

We defined four categories for these spectral images: strong, weak, no call and other (Fig. 2A). Strong is defined as a strong signal of the target species presented in the image; weak is a particular category for images contained a weak call structure; no call is for images which had no signal at all; other is defined as ambiguous or non-target species calls. We built a weak signal category for the reference image dataset for two reasons: 1) we directly imported audio-files for identification without manual selection of strong calls. Therefore, this saves one step of manual pulse selection for processing large volume of data; 2) as extremely weak calls are usually misidentified, therefore assigning them to another category “weak” reduces the misidentification or false positive rate.

The images were manually assigned into the four categories as above (Fig. 2A). The weak and no call categories contained 10–100 images from each audio-file of different species. Then, images with a size of 256*256 pixels were converted to 64*64 pixels to save training time and reduce high memory graphics use. The size-reduced images packaged into two datasets: training dataset with 85% of images (111,244/130,858) and validation dataset with 15% (19,614/130,858) using a Python script Making_dataset.py in Waveman. To balance the image number for all the species, we provided an upper limit for when a species had too much data (Supplementary information 1, S1). We also developed a method to increase the number of images for the rare species (S2), in which signals were rescaled exponentially, and we shifted the window slightly either side of the call to change background noise. Both these measures ensure rare species with small image numbers will not be under-represented and under-classified relative to common species (S1), as though we cannot incorporate the same level as call variation as in common species this measure rebalances the probability of identification based on relative call ratios.

2.1.4. Training models with the reference library

The network is key to optimizing the parameters in the model using the reference dataset. The higher the efficiency of extracting acoustic feature from spectrum images, the higher the accuracy. We built a new network specifically for the tropical bats (termed BatNet), which was incorporated into our software Waveman. The BatNet has 22 convolutional layers for extracting useful acoustic features (Simonyan and Zisserman, 2014) and eight shortcut connections between layers to avoid the problem of information loss as layer number increases. There are two key parameter settings of batch size equal to 64, and learning

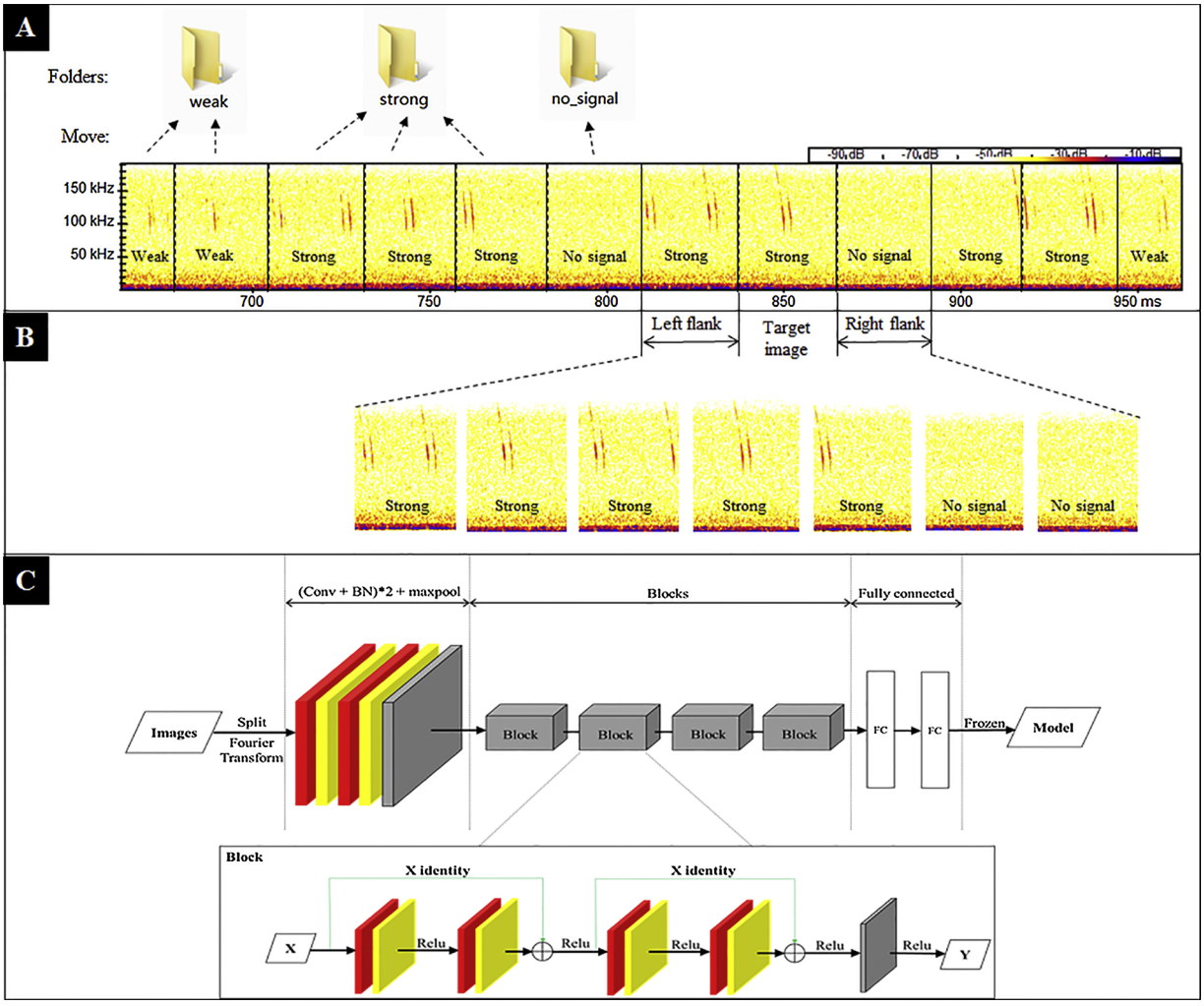


Fig. 2. Detail of pipeline in Waveman. A illustrates the image selection procedures for an audio file recorded a *Kerivoula hardwickii*. The images were generated using STFT and move them to the corresponding folders; B illustrates an example to use ReChk strategy. The target images together with two side flanks were segments to seven episodes with overlaps. There are five out of seven images were identified and passed the checking. Thus, the target segment has bat signal; C is architecture of BatNet. The structure of BatNet has three parts: convolutional, block and fully connected. The convolutional part has two convolutional layers and one max pooling layers; there are four blocks. Each block has five convolutional layers and one max pooling or average pooling layer together with two shortcuts for residual function. The full connected part has two layers. All the 22 convolutional layers have kernel with the size of 3*3. All the batch normalization layers follow behind with convolutional layers. Activation functions are Rectified Linear Unit for non-linear classification.

rate equaled 1e-3 (only a network called ResNet_v2 was set to 1e-5). We trained a model using the training dataset (incl. 111,244 images) for 50–60 times using Graphic Processing Unit (GPU, Nvidia 1080ti, US) for two hours and 10 min. After training the model contained refined parameters and a graph which were saved for the downstream analysis.

2.2. Testing and further refinements of Waveman

We used filtered data (i.e., validation dataset in which 19,614 images were selected) to test different model and taxon bias. Then we collected audio-files from a corridor for the 15 species to test how to lower the misidentification and false positives. Finally, we optimized the BatNet and collected audio from different human and natural conditions to test the generalization of the model by recording the same species under different conditions these audio-files are unfiltered and were directly imported into Waveman.

2.2.1. Comparison of BatNet and other three networks

We compared the performance of BatNet with three other networks: CNN_{FULL}, VggNet, and ResNet_v2. They have many convolutional layers for capturing features in a high-level dimensional space (discussed and explained in Simonyan and Zisserman, 2014). CNN_{FULL} (Mac Aodha et al., 2018) is the relatively simple network with two convolutional layers; VggNet has 16 convolutional layers (Simonyan and Zisserman, 2014); ResNet_v2 has 50 convolutional layers (Silberman, 2017). We input the validation dataset (incl. 19,614 images) to Waveman by changing the network settings. Models were trained in the GPU for the four networks as mentioned above. Then, using the validation dataset to evaluate the performance of the four models. Receiver operating characteristic (ROC) curve, area under the curve (AUC), overall accuracy, sensitivity, specificity and false positive rate were calculated using Python package sklearn (Pedregosa et al., 2011) and plotted using package matplotlib (Hunter, 2007).

We used a confusion matrix to evaluate the taxonomic bias of the four networks using the package sklearn (Pedregosa et al., 2011). To check whether closely genetically related taxa impact on the identification for the four networks, we also draw a phylogenetic tree to illustrate genetic relationships among the taxa. The phylogeny was reconstructed using six mitochondrial genes from the 36 species extracted from public database GenBank (see **File S1**, Benson et al., 2015). We adopted a maximum likelihood method using a software called RAxML v8.2.2 with GRT + Γ model and 200 bootstraps (Stamatakis, 2014) to

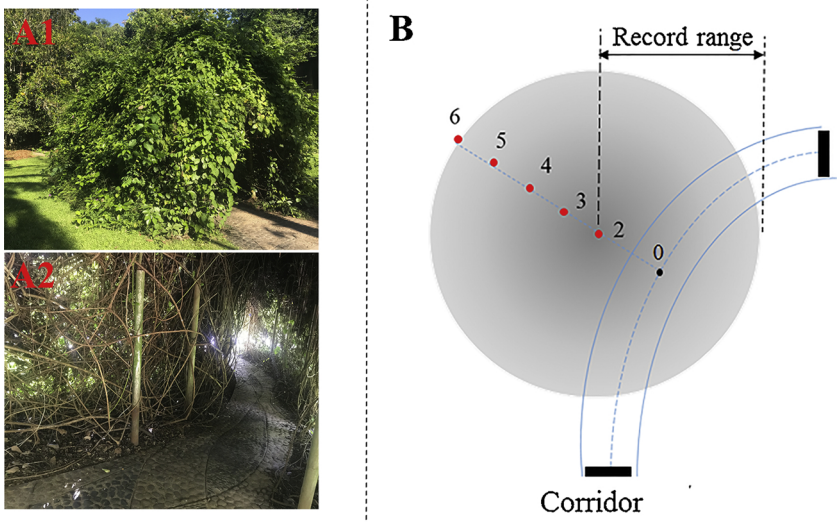


Fig. 3. Experimental design for recording bat calls from different distances. A1 photo is outside the corridor fully covered with vines *Porana racemosa*. A2 shows the inside the corridor. B is the design plot for recording bat calls from different distance. The arch-like corridor has eight meters long. We blocked the two ends of the corridor in order to release bat fly freely in the side. There are six positions labeled 0, 2 to 6 located in the dotted line which vertical with axis of the corridor. The numbers represent the distance to the axis with unit of meter. We put one Pettersson microphone at position 0 and another one at other positions outside of the corridor. which is depend on bat frequency. High frequency bat requires small distance while low frequency can be record with long distance. The gray circle illustrates the record range of Pettersson in the two meters from corridor for most of bats which frequency is between.30–120 kHz.

build the phylogeny. The acoustic traits sweep-rate and frequency traits were attached in the phylogeny with different colours to illustrate its evolutionary pattern of the two traits along the phylogeny. This sweep-rate and frequency traits were extracted (**S3**) using Sonobat V4.2.2 (Sonobat Co. Arcata, CA) and plotted using R packages picante and Geiger (Kembel et al., 2010; Team, 2013; Pennell et al., 2014).

2.2.2. Weak signal and ReChk strategy

To test the model generated by BatNet, we designed an experiment to collect bat calls with different strengths. Distance between bat and recorder greatly influences the recording quality. When recorders are close to the bats, the acoustic signal was strong and the call structure is usually comparatively intact. However, as distance increases from the calling bat, only stronger elements of the call can be recorded and other elements may distort, attenuate or disappear. The experiment was carried out in an eight-meter-long corridor covered with plants where bats could fly freely (Fig. 3). Two recorders were placed near the corridor. One was placed upto six meters away from the axis of corridor (depending on the frequency of the call); another recorder was placed inside the corridor. We released bats individually and utilized both recorders simultaneously. Fifteen bat species flew freely in the corridor alone for 30 s to five minutes, including five vespertilionid species (i.e., *Myotis muricola*, *M. laniger*, *Murina cyclotis*, *Kerivoula hardwickii,* and *Tylonycteris robustula)*, five hipposiderid species (i.e., *Hipposideros larvatus*, *H. cineraceus*, *H. pomona*, *H. armiger*, and *Aselliscus stoliczkanus*), and five rhinolophid species (i.e., *Rhinolophus malayanus*, *R. sinicus*, *R. stheno*, *R. pearsonii*, and *R. siamensis)*. Three species are relatively hard to trap using harptraps in our main site (Table 1), these include *T. robustula, H. cineraceus,* and *R. pearsonii.*

We utilized two methods to reduce the false identification rate: 1) add a weak signal category to our reference library before identification; 2) introduce a new checking strategy after identification. We built two models. one model was built using a dataset with a weak signal category (termed dataset-with-weak) and compared with another model using dataset without a weak signal category (termed dataset-without-weak). The two models were used BatNet with the GPU as above mentioned. Secondly, all these collected audio-files were sent to Waveman to identify using BatNet to get preliminary results which were checked as follows. Images classified to a certain bat species with probability higher than 80% were checked again using a new strategy for filtering false positive and misidentified results. The checking strategy includes both target regions and two sides of its flanking region to identify because bat calls echolocation are always repeat several times in a specific region. For example, when we checked the target image (in Fig. 1B) of a *Kerivoula hardwickii* and presents with over 80% probability, the image region together with two sides of flanking regions were extracted. We then segmented the regions into seven sequential pieces with some overlap. All these pieces converted to image and then were checked again to assess whether it contained bat signals with the probability greater than 80%. If more than three pieces with *K. hardwickii* (the target species in this example) were present, the target segment passes the check (term this strategy as ReChk). The strategy was incorporated in the Waveman, as an automatic process to reconfirm identification.

2.2.3. Further improvement of BatNet for the "unknown" frequency calls

We further improved Waveman by modifying BatNet and optimizing parameter setting of batch size. We add new kind of BNorm layers behind the 22 convolutional layers to prevent overfitting when we trained models using BatNet (Fig. 2C). Therefore the model "learned" to generalize from a trend in both "known" and "unknown" datasets rather than to maximize the performance on the "known" datasets (usually called training datasets, Ioffe and Szegedy, 2015). Computers can only train with small volumes of images at once as they have too little Random Access Memory or Graphic memory. Batch size was set to limit the image number. In this study, we set a large batch size equal to 128, which means training with 128 images for each iteration.

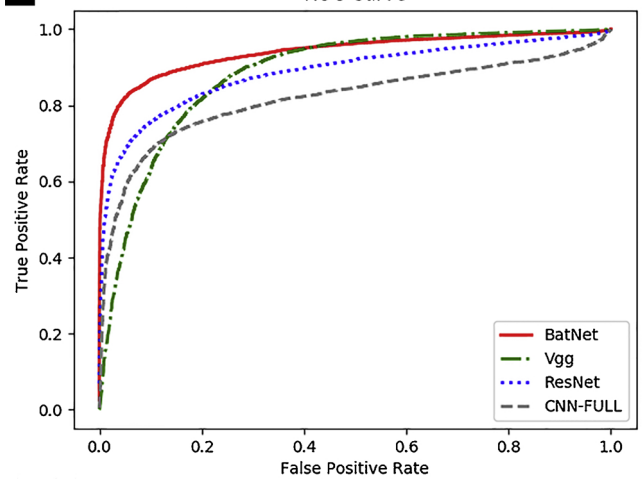
The modified Waveman was evaluated using two bat species calls from three conditions, including a densely forested area (DF), an auditorium (AU) and a cave entrance (CE) which is located over 40 km from the DF and AU in Xishuangbanna Tropical Botanical Garden. In CE, we recorded calls of *H. larvatus* and *R. siamensis*; in AU, we caught three *H. larvatus* and *R. siamensis* in a nearby rainforest and released them in AU for recording multiply times; in DF, we caught both *H. larvatus* and *R. siamensis* in limestone forest and recorded their calls during release. *R. siamensis* from all the three sites has a similar frequency at approximately 75 kHz; yet, *H. larvatus* from CE has relatively high frequency calls of approximately 94–96 kHz compared to DF and AU of approximately 88–91 kHz. All the collected audio-files were sent to the modified Waveman to identify and filter false positives using the ReChk.

3. Results

3.1. BatNet outperformed other networks

BatNet was compared to three other networks using the validated dataset as described above. It included 19,613 images from calls of 36 species. The AUC and overall accuracy for BatNet was better than other

A



B

Networks	AUC	Overall accuracy	Sensitivity	Specificity	False positive
BatNet	0.94	0.902	0.909	0.853	0.147
CNN_{FULL}	0.802	0.733	0.907	0.344	0.656
VggNet	0.88	0.884	0.880	0.908	0.092
ResNet	0.89	0.866	0.911	0.634	0.366

Fig. 4. Evaluation of four networks. A shows ROC curves. B shows five parameters of the networks. The best values were highlighted.

analyses (Fig. 4A). ResNet_v2 had the highest sensitivity and VggNet performed the best for specificity and false positives (Fig. 4B).

We also compared how the four networks performed within different taxa. The confusion matrix result for **BatNet** shows that the accuracy rates ranges from 86% to 100% across the 36 species (Fig. 5). There was no obvious bias for the closely related species. Misidentification rates of above 1% occurred for 22 pairs of species, including 14 in Vespertilionidae and eight other taxa, but all remained lower than 6%. The **ResNet_v2** confusion matrix result showed that a relatively high bias in Vespertilionidae compared to BatNet. For example, *Murina cyclotis* has the correct classification rate of only 76%; it misidentified *M. tubinaris* 11% of the time (Fig. S1). **VggNet** misidentified signal images of No Call or weak images in *Hipposideros*. For example, *H. pomona* had only 80% correct rate and 14% misidentified as weak signal. *H. cineraceus* had only 82% correct rate and 16% misidentified as No Call signal (Fig. S2). **CNN_{FULL}** bias in the Vespertilionidae was lower than 80% accuracy rate for the six pairs of species (Fig. S3). For example, *M. laniger* had only 38% accuracy rate with 50% misidentification to *Hypsugo pulveratus*.

3.2. The library with weak signal and ReChk shows significantly lower misidentification and fewer false positives

Compared to the "dataset-without-weak", "dataset-with-weak" not only has lower misidentification, but also larger numbers of correct bat identification for all the 15 species (Fig. 6, Figs. S4, and S5). Furthermore, ReChk reduces the misidentifications for the 15 species. Take *M. muricola* as an example (Fig. 6), when recorded from greater distances, the sensitivity increased from 42.86% to 57.12% after incorporating the weak category into the reference data (Fig. 62A and B). Weak signal images were detected across the entire audio-time range (gray bar in 2B in Fig. 6). But there was still a 5.77% misidentification rate. We introduced the ReChk and filtered all calls to maintain accuracy (Fig. 62C), despite decreasing sensitivity. The audio-files recorded from greater distances have very high misidentification rate (Fig. 61A), which was reduced by introducing the weak category in the reference dataset (Fig. 61B). ReChk further reduced the misidentification rate to 0 (Fig. 61C), despite the identification rate decreasing slightly as correct but weak calls were removed. The other 13 species showed similar results of low misidentification with weak and echoed signals removed (Figs. S4 and S5), only *H. larvatus* and *R. sinicus* have a slightly higher misidentification rates.

3.3. Two key optimized settings increased accuracy rate for the calls from different locations

Compared to DF, the sensitivity rate increased and misidentification rate decreased significantly (Table 2) in both AU and CE after adding BNorm layers and increasing the batch size as above described (method part). In DF some audio-files were used to make reference datasets, the sensitivity of *R. siamensis* changed little, and *H. larvatus* changed by under 10%; both species showed no misidentification. However, in both AU and CE, the sensitivity increased almost three-fold and the misidentification rate decreased at least four-fold, especially as *H. larvatus* which formerly had over 10% error misidentification rate, despite calls varying upto 6 kHz compared to the populations from DF. The sensitivity of the *R. siamensis* had moderate increased in CE.

4. Discussion

In this study, we demonstrate that our model has high enough accuracy to identify the 36 tropical bat species for both filtered and unfiltered datasets. For the selected data BatNet outperforms CNN_{FULL} by increasing the overall accuracy rate from 77% to 91% using the validated dataset, and shows greatest improvement for the 12 vespertilionid species. This improvement indicates the complicated architecture of BatNet is more suitable for the application of unfiltered audio-files recorded in tropical regions than simple networks (i.e., CNN_{FULL}). Secondly, the unfiltered data collected from different environments, including nature habitats (i.e., CE (cave entrance) and DF (densely forested) where bats usually emerge) and human living areas (i.e., AU (Auditorium) is relatively wide open). The audio-files not only contained no call regions, but only contained signal regions full of echos, background noises, and high intra-species acoustic variation (such as sweep rate, band width, and call duration, quantified, which are described in S4 and displayed in Fig. S6 A). We introduced dataset-with-weak and ReChk strategy, which greatly reduced the false positive and misidentification rate. In addition, we also tried commercial software Kaleidoscope and SonoBat to delete noise and extract signal as a comparison to our new approach. Only a few calls were left after filtering, for example we recorded a bamboo bat for almost 40 s which generated over 430 call pulse segments had call pulse using Waveman, and 70 segments were removed using ReChk; while only about 100 pulses were extracted using Kaleidoscope therefore providing much lower sensitivity. In addition, although modelling is somewhat laborious, Waveman saves one step of manual filtering weak signal pulse for large volumes of monitoring data compared to commercial software. We simply imported raw audio-files into it and generated species information and occurrence time results.

BatNet has a cascade of multiple layers of nonlinear processing units for feature extraction from low-level features and syntheses to high-level features (by converting data matrix shapes and synthesis the matrix). Low-level features are some of raw data points (i.e., pixels in images). High-level features include frequency, bandwidth, and hundreds of other combined features which may have no specific meaning in isolation. During synthesis as the data from "shallow" to "deep" layers, the algorithm weights the features iteratively according to the reference dataset in order to best describe call structure for specific bat species. However, the acoustic signal synthesis with huge number of layers usually leads to information loss to some degree (He et al., 2015). We, thus, introduced eight shortcuts among layers in order to send partial original information to the next layer (proposed by He et al., 2015), which solved the information loss issue as layer number increases in the process of bat acoustic feature extraction. Another

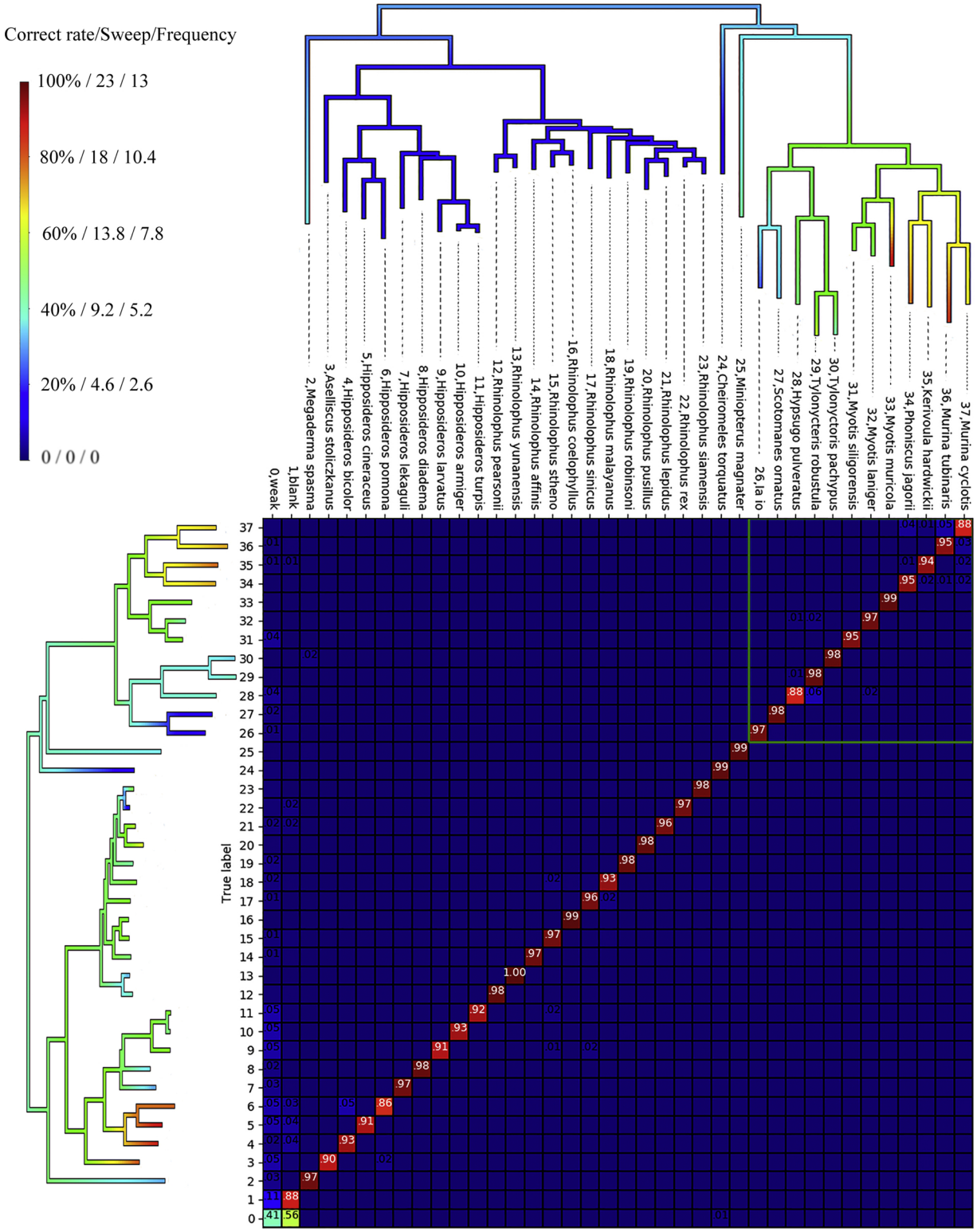


Fig. 5. Confusion matrix result using BatNet. We use valid dataset included 19,134 images to get the accuracy for the 36 species, one No Call and weak categories. The percentage in each cell is identify rate for species and the misidentifiable rate higher than 1% shows in this matrix. Vespertilionid species circled with green lines. The phylogeny shows the relationship of all the species. The color of each phylogenetic branch represents the sweep-rate values which details the frequency modulation of the call. The tree is inferred from six mitochondrial genes using maximum likelihood method, 200 replications. Our phylogeny strong support the monopoly of Miniopteridae and Vespertilionidae, which is same with the phylogeny used 16 nuclear genes (Miller-Butterworth et al., 2007) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

problem is overfitting, this can be solved by adding BNorm layers to standardise the parameters with batch images, which greatly alleviates the impact of high image variability including distortions or overlaps which can occur in recording and lead to errors without proper processing. Thus, the model after batch normalized could be used to identify data from wider regions.

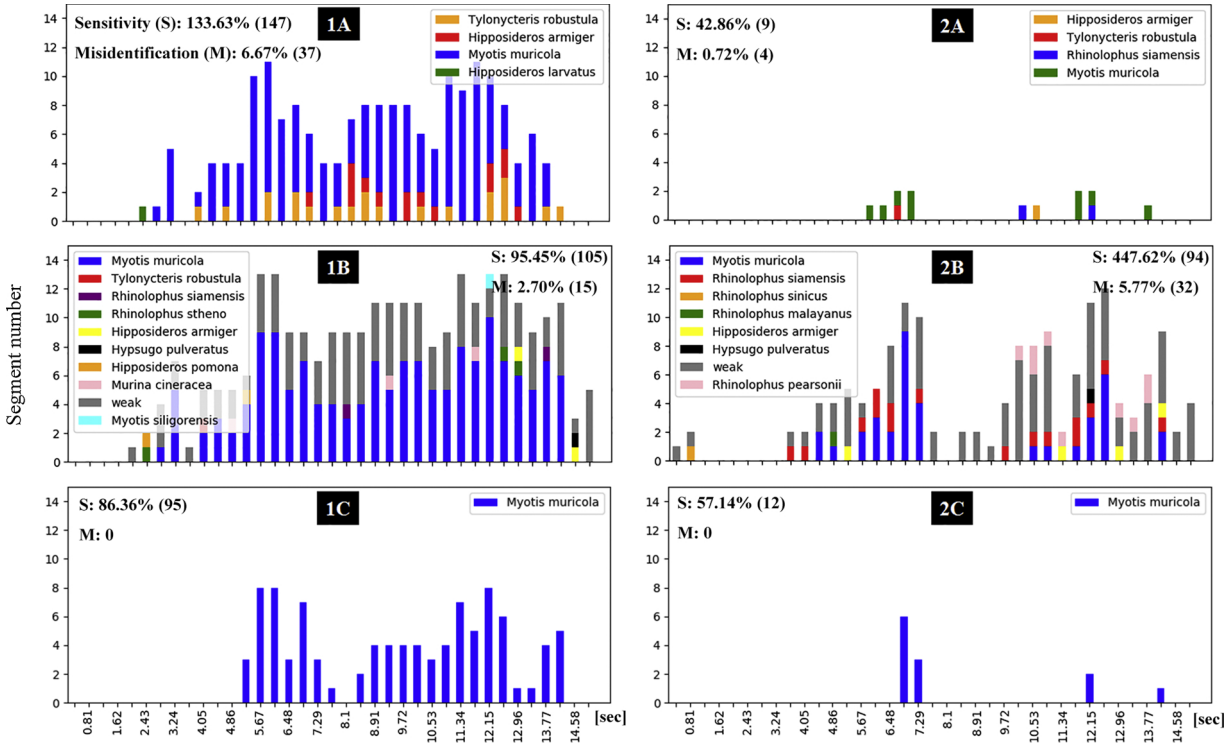


Fig. 6. Stack Histogram plot from different distances for the *Myotis muricola* using passive recorders. The left three plots 1A, 1B and 1C show the results recorded from the short distance of location 0; the right three plots 2A, 2B and 2C from the long distance of the location 4 (illustrated in Fig. 4). The top two plots 1A and 2A show the results using the dataset without weak category; the middle two plots 1B and 2B using the dataset with weak; the bottom two plots 1C and 2C show the identification results that filtered by using ReChk strategy and using the dataset with weak. All the identifications are use network BatNet. We manually counted that 110 and 21 image had pulse for the long and short distance respectively. The sensitive rate equals to (Waveman detected image number/manually counting image number) * 100%; and misidentification rate equals to (Waveman detected misidentified image number/totally image number) * 100%.

Table 2
The sensitivety rate of species signal in the audio files for two species in the three sites.

Sites	Species	without_BN	with_BN	
		BC_64	BC_64	BC_128
DF	HL	39.50%	42.70%	48.83%
	RS	43.80%	43.27%	44.33%
AU	HL	20.33% (2.94%)*	27.85% (2.23%)	53.00% (0.35%)
	RS	18.55% (0.02%)	19.58% (0.01%)	21.33% (0.01%)
CE	HL**	25.74% (11.88%)	39.60% (5.94%)	64.36% (2.97%)
	RS	9.45%	11.94%	15.92%

HL is Hipposideros larvatus; RS is Rhinolophus siamensis; DF is densely forested area; AU is auditorium; CE is cave entrance; without_BN is network not include batch normalization layers whille with_BN is including the layers; BC_64 is batch size setting to 64; BC_128 is batch size setting to 128.

* the percetage in parenthesis is misidentification rate.

** In cave entrance place, the rate static not include calls from a Ia io since it also was recorded at the same time.

Call frequency variation is influenced by geographic factors for some rhinolophid and hipposiderid bats in Afro-Eurasia (Kingston and Rossiter, 2004; Mao et al., 2013; Wilkins et al., 2013; Jacobs et al., 2017), this may be selective to allow bats to exploit different sizes of insect prey (Kingston and Rossiter, 2004) or result from drift due to geographic isolation (Thabah et al., 2006). *H. larvatus* shows huge frequency variation and may represent a species complex. We collected them from species-rich sites in Xishuangbanna of China and in Thailand, with calls showing frequency ranges between 80 kHz to 104 kHz (referred our data and other published data, (Thabah et al., 2006; Jiang et al., 2010)).We built a call library for Xishuangbanna area in China where all *H. larvatus* calls were under 91 kHz, thus local libraries and localization are also important factors. Waveman successfully captured and identified the high frequency calls with high accuracy rate from other locations (i.e., CE where the *H. larvatus* has 94–96 kHz) (Table 2). Waveman has powerful learning ability to generate models which include spatial call variation and identifies general structural features, which makes the library preparation easier than before; there is no need to capture all the possible individuals with variable calls, especially for rhinolophid and hipposiderid species.

Researchers can easily infer the call to the family level by referring call structure traits (such as sweep-rate, bandwidth, and call-duration), despite of large intra-specific variation; the species-average values of these traits are evolutionarily constrained so provide accurate information on higher levels of taxonomic identity (S4). While frequency related traits, such as frequency of maximum energy, allows identification of some bats to the species level (Hughes et al., 2012). It shows relatively low variability compared to the sweep-rate, bandwidth, and call-duration based on the phylogeny of the 36 species when considering their intra-specific variation (S4). Thus, Waveman can mine useful information related to frequency and call structure traits. The mining processing is automatically synthesizes from low level to high level to accurately identify species.

4.1. Challenges

We tried to standardise spectral images from difference devices by changing frequency axes. Since the sampling rate of M500-384 is 384,000 Hz, which corresponds to the frequency axis of spectral image with half of the rate of 192,000 Hz; whist 22,100 Hz for D-240X, and compression varies by an order of 10. However, the adjustments make difference in identification result. This may greatly impede data sharing globally since researchers cannot use the data from difference sources. Thus, to build data sharing networks, it requires standardising key

settings (such as sample rate) to allow comparability between different devices, in addition to noting regional variations.

4.2. Standardized metrics for monitoring

Standardized and comparative metrics for monitoring biodiversity across space and time remain a goal for ecology at all scales, providing a metric for management from local reserves to national and international standards. Here, by utilizing advances in computer learning algorithms we provide a mechanism to generate standardized information on bat diversity and abundance based on passive recording using any given bat detector and under a range of environmental conditions. Through extensions of the approaches these techniques could be harnessed for any given taxa which produces consistent calls which can be utilized as an acoustic barcode to recognize the species in a reliable way (e.g., amphibians and birds). This approach, and extensions based on it hold real promise for transforming the way we monitor biodiversity, allowing proactive management at all levels and understanding system ecology from species to community levels at a higher resolution than is currently possible. Furthermore, by automatically including a large number of different measures and metrics, we enable the application of the approach to diverse systems, such as tropical ecosystems, with the ability to discriminate even subtle differences in call structure we can apply these approaches to any given system once a training dataset has been developed.

As the cost of bioacoustic monitoring continues to decrease, these tools are likely to provide a major advance in enabling us to monitor ecosystems across space and time, and therefore the approach put forward here represents a major step in developing reliable and consistent metrics to quantify biodiversity change in any given system across space and time, and therefore to target efforts to manage and maintain biodiversity in these systems.

Declaration of Competing Interest

There are no conflicts of interest.

Acknowledgements

Supported by Chinese National Natural Science Foundation (Grant #: U1602265, Mapping Karst Biodiversity in Yunnan). Supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA20050202). Supported by the High-End Foreign Experts Program of Yunnan Province (Grant #: Y9YN021B01, Yunnan Bioacoustic monitoring program). Supported by the CAS 135 program (No. 2017XTBG-T03).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.biocon.2019.108269.

References

Altes, R.A., Titlebaum, E., 1970. Bat signals as optimally Doppler tolerant waveforms. J. Acoust. Soc. Am. 48, 1014–1020.
Astaras, C., Linder, J.M., Wrege, P., Orume, R.D., Macdonald, D.W., 2017. Passive acoustic monitoring as a law enforcement tool for Afrotropical rainforests. Front. Ecol. Environ. 15, 233–234.
Baker, E., Price, B.W., Rycroft, S.D., Hill, J., Smith, V.S., 2015. BioAcoustica: a free and open repository and analysis platform for bioacoustics. Database 2015.
Barratt, E., Deaville, R., Burland, T., Bruford, M.W., Jones, G., Racey, P., et al., 1997. DNA answers the call of pipistrelle bat species. Nature 387, 138.
Benson, D.A., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2015. GenBank. Nucleic Acids Res. 43, D30.
Boonman, A., Schnitzler, H.-U., 2005. Frequency modulation patterns in the echolocation signals of two vespertilionid bats. J. Comp. Physiol. A 191, 13–21.
Cardinale, B.J., Gonzalez, A., Allington, G.R., Loreau, M., 2018. Is local biodiversity declining or not? A summary of the debate over analysis of species richness time trends. Biol. Conserv. 219, 175–183.
Christin, S., Hervet, E., Lecomte, N., 2018. Applications for deep learning in ecology. bioRxiv, 334854.
Clement, M.J., Rodhouse, T.J., Ormsbee, P.C., Szewczak, J.M., Nichols, J.D., 2014. Accounting for false-positive acoustic detections of bats using occupancy models. J. Appl. Ecol. 51, 1460–1467.
Gager, Y., Tarland, E., Lieckfeldt, D., Ménage, M., Botero-Castro, F., Rossiter, S.J., et al., 2016. The value of molecular vs. Morphometric and acoustic information for species identification using sympatric molossid bats. PLoS One 11, e0150780.
Gasc, A., Sueur, J., Jiguet, F., Devictor, V., Grandcolas, P., Burrow, C., et al., 2013. Assessing biodiversity with sound: Do acoustic diversity indices reflect phylogenetic and functional diversities of bird communities? Ecol. Indic. 25, 279–287.
Hill, A.P., Prince, P., Piña Covarrubias, E., Doncaster, C.P., Snaddon, J.L., Rogers, A., 2018. AudioMoth: evaluation of a smart open acoustic device for monitoring biodiversity and the environment. Methods Ecol. Evol. 9, 1199–1211.
Hughes, A.C., Satasook, C., Bates, P.J., Bumrungsri, S., Jones, G., 2012. The projected effects of climatic and vegetation changes on the distribution and diversity of Southeast Asian bats. Glob. Change Biol. 18, 1854–1865.
Hughes, A.C., Satasook, C., Bates, P.J., Soisook, P., Sritongchuay, T., Jones, G., et al., 2011. Using echolocation calls to identify Thai bat species: vespertilionidae, Emballonuridae, Nycteridae and Megadermatidae. Acta Chiropt. 13, 447–455.
Hunter, J.D., 2007. Matplotlib: a 2D graphics environment. Comput. Sci. Eng. 9, 90–95.
Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv 1502.03167.
Jacobs, D.S., Catto, S., Mutumi, G.L., Finger, N., Webala, P.W., 2017. Testing the Sensory Drive Hypothesis: geographic variation in echolocation frequencies of Geoffroy's horseshoe bat (Rhinolophidae: rhinolophus clivosus). PLoS One 12, e0187769.
Jiang, T., Liu, R., Metzner, W., You, Y., Li, S., Liu, S., et al., 2010. Geographical and individual variation in echolocation calls of the intermediate leaf-nosed bat, Hipposideros larvatus. Ethology 116, 691–703.
Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., et al., 2010. Picante: r tools for integrating phylogenies and ecology. Bioinformatics 26, 1463–1464.
Kingston, T., Rossiter, S.J., 2004. Harmonic-hopping in Wallacea's bats. Nature 429, 654.
Kiskin, I., Zilli, D., Li, Y., Sinka, M., Willis, K., Roberts, S., 2018. Bioacoustic detection with wavelet-conditioned convolutional neural networks. Neural Comput. Appl. 1–13.
Mac Aodha, O., Gibb, R., Barlow, K.E., Browning, E., Firman, M., Freeman, R., et al., 2018. Bat detective—deep learning tools for bat acoustic signal detection. PLoS Comput. Biol. 14, e1005995.
Mao, X., He, G., Zhang, J., Rossiter, S.J., Zhang, S., 2013. Lineage divergence and historical gene flow in the Chinese horseshoe bat (Rhinolophus sinicus). PLoS One 8, e56786.
Marques, A., Pereira, H.M., Krug, C., Leadley, P.W., Visconti, P., Januchowski-Hartley, S.R., et al., 2014. A framework to identify enabling and urgent actions for the 2020 Aichi Targets. Basic Appl. Ecol. 15, 633–638.
Meagher, J., Damoulas, T., Jones, K., Girolami, M., 2018. Phylogenetic Gaussian processes for bat echolocation. Statistical Data Sci. 111.
Meyer, C.F., Aguiar, L.M., Aguirre, L.F., Baumgarten, J., Clarke, F.M., Cosson, J.-F., et al., 2010. Long-term monitoring of tropical bats for anthropogenic impact assessment: gauging the statistical power to detect population change. Biol. Conserv. 143, 2797–2807.
Newey, S., Davidson, P., Nazir, S., Fairhurst, G., Verdicchio, F., Irvine, R.J., et al., 2015. Limitations of recreational camera traps for wildlife management and conservation research: a practitioner's perspective. Ambio 44, 624–635.
Parsons, S., Jones, G., 2000. Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. J. Exp. Biol. 203, 2641–2656.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.
Pennell, M.W., Eastman, J.M., Slater, G.J., Brown, J.W., Uyeda, J.C., FitzJohn, R.G., et al., 2014. Geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 30, 2216–2218.
Proença, V., Martin, L.J., Pereira, H.M., Fernandez, M., McRae, L., Belnap, J., et al., 2017. Global biodiversity monitoring: from data sources to essential biodiversity variables. Biol. Conserv. 213, 256–263.
Rich, L.N., Davis, C.L., Farris, Z.J., Miller, D.A., Tucker, J.M., Hamel, S., et al., 2017. Assessing global patterns in mammalian carnivore occupancy and richness by integrating local camera trap surveys. Glob. Ecol. Biogeogr. 26, 918–929.
Russo, D., Jones, G., 2002. Identification of twenty-two bat species (Mammalia: chiroptera) from Italy by analysis of time-expanded recordings of echolocation calls. J. Zool. 258, 91–103.
Russo, D., Voigt, C.C., 2016. The use of automated identification of bat echolocation calls in acoustic monitoring: a cautionary note for a sound analysis. Ecol. Indic. 66, 598–602.
Russo, D., Ancillotto, L., Jones, G., 2017. Bats are still not birds in the digital era: echolocation call variation and why it matters for bat species identification. Can. J. Zool. 96 (2), 63–78.
Rydell, J., Nyman, S., Eklöf, J., Jones, G., Russo, D., 2017. Testing the performances of automated identification of bat echolocation calls: a request for prudence. Ecol. Indic. 78, 416–420.
Silberman, N., 2017. TF-Slim: A Lightweight Library for Defining, Training and Evaluating Complex Models in TensorFlow.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv 1409.1556.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis

of large phylogenies. Bioinformatics 30, 1312–1313.
Stowell, D., Wood, M.D., Pamuła, H., Stylianou, Y., Glotin, H., 2019. Automatic acoustic detection of birds through deep learning: the first Bird Audio Detection challenge. Methods Ecol. Evol. 10, 368–380.
Team, R.C, 2013. R: a Language and Environment for Statistical Computing.
Thabah, A., Rossiter, S.J., Kingston, T., Zhang, S., Parsons, S., Mya, K.M., et al., 2006. Genetic divergence and echolocation call frequency in cryptic species of Hipposideros larvatus sl.(Chiroptera: hipposideridae) from the Indo-Malayan region. Biol. J. Linn. Soc. 88, 119–130.
Trolle, M., Kéry, M., 2003. Estimation of ocelot density in the Pantanal using capture-recapture analysis of camera-trapping data. J. Mammal. 84, 607–614.
Yu, X., Wang, J., Kays, R., Jansen, P.A., Wang, T., Huang, T., 2013. Automated identification of animal species in camera trap images. EURASIP J. Image Video Process. 2013 (1), 52.
Walters, C.L., Collen, A., Lucas, T., Mroz, K., Sayer, C.A., Jones, K.E., 2013. Challenges of Using Bioacoustics to Globally Monitor Bats. Bat Evolution, Ecology, and Conservation. Springer, pp. 479–499.
Wilkins, M.R., Seddon, N., Safran, R.J., 2013. Evolutionary divergence in acoustic signals: causes and consequences. Trends Ecol. Evol. 28, 156–166.
ZINGG, P.E., 2019. Akustische Artidentifikation Von Flederrnausen (marnrnalia: chiroptera) in der Schweiz. Rev. Suisse Zool. 294 p, 263.