2020 5th International Conference on Smart Grid and Electrical Automation (ICSGEA)

BatNet++: A Robust Deep Learning-based Predicting Models for Calls Recognition

Jian Hu*, Wenzai Huang, Yongdong Su, Yuting Liu, Peng Xiao

Information Center, Yunnan Power Grid Company Limited, Yunnan, 650091, China
*corresponding author's email: hujian92@126.com

Abstract—**Biodiversity is the most complex feature of our planet and it is the most vital. Without biodiversity, there is no future for humanity. People are gradually realizing the significance of maintaining biodiversity. Since the lack of effective biodiversity detection tool, human perform inefficiency in biological detection and protection. In this paper we proposed a novel method, called BatNet++, in order to monitor biodiversity of bats by identifying bat calls. Experimental results show that BatNet++ yields effective improvement in bat calls recognition. A large amount of experiments shows that our method can be harnessed for given taxa which make consistent sounds that may be used as vocal barcodes to identify species (e.g., amphibians and birds).**

Keywords- Fingerprinting; Indoor Localization; Location Search

I. INTRODUCTION

Biodiversity is closely related to human survival and development and has significant practical value [1]. Among various species, the bat is one of the most important ecological indicator animals and plays an important role in the diverse ecosystem. For instance, cave-type bats mainly feed on other small arthropods. Bats are the major food providers and species spreaders in the karst cave ecosystem, and also they can control pests. In order to maintain the biodiversity of the bat, monitoring bat species and recording bat survival, the number of ethnic groups is particularly critical.

Yunnan Province is located in the core area of continuous karst landforms in southwestern China. Karst landforms are widely distributed and diverse, forming a relatively stable cave system that can provide a good habitat for bats which provides favorable conditions for monitoring bat activity and quantity. Current methods for monitoring bats in Yunnan Province are artificially set up fog nets and harp nets. Since the uneven distribution of bats in Yunnan, this method has many uncertainties. AI approaches, for instance deep learning, enable ecologists to detect bio-sonar at heretofore ineffable resolutions. These approaches can be used to recognize numerous species based on their bio-sonar (calls), such as mosquitos [4], birds [3], as well as bats [2].

Inspiring by the deep learning techniques, in this paper, we proposed BarNet++ which adapt ResNet 50 [5] into bat calls recognition. Concretely, to make many standard deep learning methods available for bat recognition tasks, a large amount of calls wave are collected and converted into a picture. In order to speed up training and to minimize the impact taken by data skew, instead of using random initialization for NNs model, we pre-trained BatNet++ with the image-net dataset. Next, after parameter transferring, BatNet is fine-tuning on the target domain. To this end, a trained model is used in recognition bat calls, the result shows that BatNet++ can overcome previous bat call recognition methods by up to \$1%\$.

The main contributions of this paper are as follows:

1. We collected the calls data from 38 tropical bat species in South-Eastern-Asia area by manual methods.
2. We introduce a effective supervised learning model for bat specifics recognition.
3. Experiments shows that BatNet++ outperforms all of the modern works on bat calls recognition methods. BatNet++ improves the state-of-art best efficiency by up to 2%.

The rest of the paper is organized as follows: BatNet++ approach is presented in Sec II. The experimental evaluation results are reported in Sec III. We conclude the paper in Sec IV.

II. BATNET++

In this section we introduce the design of BatNet++, including supervised pre-training on the large-scale ImageNet dataset, fine-tuning the network with the latent layer to simultaneously learn domain-specific feature representation and how to prepossess the calls data, as well as how this network model is trained.

A. Pre Training

Directly using the row data to train the NNs model has the following defects: i) It takes a long time to train a new model from random initialization parameters. ii) Data used for training may be unbalanced. iii) The training model directly from random weights is not conducive to the training of a high accuracy model. To settle these problems, we use the pre-trained ResNet model proposed by Kaiming He and et al. from the Tensorflow library[6], which is trained on the large-scale ImageNet dataset which contains more than 1.2 million images categorized into 1000 object classes. The initial experimental results show that pre-training can take about \$2\\%\$ improving to model accuracy.

B. Fine-turning on Target Domain

To achieve domain adaptation, we fine tune our network model on the target domain dataset via backpropagation. 38 kinds of bat calls, across Thailand and Xishuangbanna in China, are collected by monitoring staffs. Moreover, since audio frequency is a high dimensions data which are hard processed by exist deep learning model, all the audio-files were transform to spectral images.

To begin with, the signal region of the audio file only for the target species are extract by manually. Next we assign a

978-1-7281-8278-0/20/$31.00 ©2020 IEEE
DOI 10.1109/ICSGEA51094.2020.00062

Authorized licensed use limited to: Thueringer Universitaets. Downloaded on November 16,2023 at 09:49:35 UTC from IEEE Xplore. Restrictions apply.

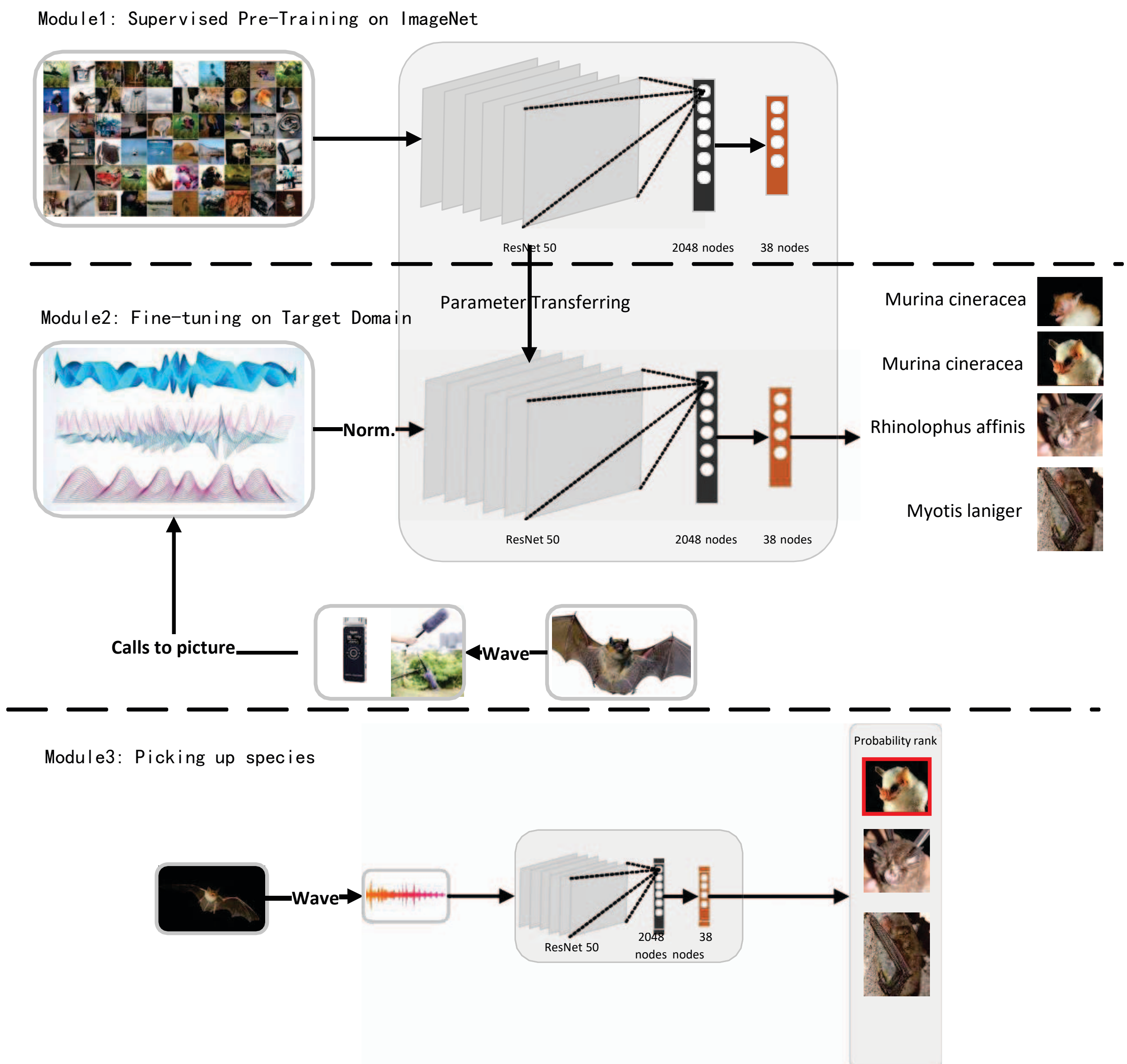

Figure 1. The framework of BatNet++. BatNet++ consists of three phase among which is the supervised pre-training phase of a neural network on the ImageNet to learn rich mid level representations and the phase of fine-tuning on the target domain dataset as well as the phase of returning recognition results based on probability rank.

unique number to selected audio files, for convenience of downstream processing. And then, in order to guarantee equal sampling independent of recording equipment, we up-sampling the training data by extracted 10363 samples from the 384001 records generated for each second of recording.

C. Picking up species

Given the bat wave a and the candidate pool $P \in R^{38}$, we first extract the outputs of the latent layer as the image signature which is denoted by $O \in R^{38}$. And then O is fitted into the softmax function, here we get the probability list $Y = \{y_1, y_2, ..., y_{38}\}$ of which species the wave a belongs to. Finally, we pick up the max probability from .

$$y_i = \frac{o}{\sum o_i} \tag{2}$$

III. EXPERIMENTS

A. Setup

Authorized licensed use limited to: Thueringer Universitaets. Downloaded on November 16,2023 at 09:49:35 UTC from IEEE Xplore. Restrictions apply.

In this sub section, we introduce the experimental set up, including the hardware and the data sets and also measurements metrics.

1) Dataset: As described before, both training and predicting data are consist of 818 bat calls audio files from 618 individuals of 38 bat category accross Thailand to Xishuangbanna in Yunnna province. In these bat calls audio files, 352 bat calls audio files were come from limestone forest in Xishuangbanna, rainforest, cave, and urban areas and 503 from 468 individuals of 23 species from forest and karsts in Thai. Note that, all the calls audio files are tansfer to figure files.

2) Baselines: The following representative NNs models are selected:

- CNN[7]: CNN made its mark in 2012, and Alex Krizhevsky won the ImageNet Challenge of that year with them. He reduced the classification error record from 26\% to 15\% and shocked the world at that time.
- VGG[8]:The VGG model is the second place in the 2014 ILSVRC competition. The VGG model performs well in multiple transfer learning tasks.
- ResNet[5]The residual network is proposed by 4 scholars from Microsoft Research. It obtained the superiority of image classification and object recognition in the ILSVRC 2015. The characteristics of the residual network are easy to optimize and can increase the accuracy by adding considerable depth. The internal residual block uses skip connections to alleviate the problem of gradient disappearance caused by increasing depth in deep neural networks. In our design, "ResNet 50" means there is 50 hidden layers in the NNs model.
- BatNet[2] BatNet is introduced in and is a novel NNs specifically for the bats call recognition (called BatNet). The BatNet has 22 convolutional layers for extracting useful acoustic features.

3) Matrix: The effectiveness of a Bat calls recognition methods is measured by the Table 1 and Eq2- Eq 5.

4)

Table 1 Evaluation of identification methods under unbalanced training data

		Predict	
		Positive	Negative
Real	Positive	Ture-Positive	False-Negative
	Negative	False-Negative	Ture-Negative

- Accuracy: Proportion of all the judged results of the classification model to the total observations.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

- Precision: In all results where model prediction is Positive, the proportion of model prediction correct

$$\Pr ecision = \frac{TP}{TP + FP} \quad (3)$$

- Recall: In all results where the true value is $Positive$, the model predicts the correct proportion

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (4)$$

- F1-score: The F1-Score indicator combines the results of Precision and Recall outputs. The value of F1-Score ranges from 0 to 1, where 1 represents the best model output, and 0 represents the worst model output.

$$\text{F1-score} = \frac{2 \bullet \text{Precision} \bullet \text{recall}}{\text{Precision+Recall}} \quad (5)$$

B. Analysis

In this sub-section, we compare the effectiveness of BatNet++ and the other NNs model in bat calls recognition.

As shown in Table 2, BatNet++ with ResNet shows a significant performance benefits over others NNs under bat calls recognition task. In terms of the Accuracy, BatNet++ with ResNet 50 is 0.956 which is over about 0.5 than BatNet++ with VGG and 0.02 than BatNet. As to F1-score, compare to BatNet which is the sta-of-art the best bat calls recognition method, BatNet++ with ResNet is also over 0.26 than BatNet.

Table 2 Compare the effective

	Accuracy	Precision	Recall	F1-score
BatNet	0.954	0.958	0.986	0.9736
BatNet with ResNet 50	0.956	0.961	0.988	0.975
BatNet with ResNet 18	0.946	0.956	0.987	0.972
BatNet with ResNet CNN	0.625	0.630	0.829	0.717
BatNet with VGG	0.907	0.898	0.946	0.923

IV. Conclusions

It is the variety of life on Earth, in all its forms and all its interactions. If that sounds bewilderingly broad, that's because it is. Biodiversity is the most complex feature of our planet and it is the most vital. Without biodiversity, there is no future for humanity. If money is a measure, the services provided by ecosystems are estimated to be worth trillions of dollars – double the world's GDP. Biodiversity loss in Europe alone costs the continent about 3\% of its GDP, or €450m (£400m), a year. Hence, it is important for us to monitoring biodiversity and maintain biodiversity. Bat is one of the most important ecological indicator animals, and it plays an important role in the ecosystem at the same time, and it is a key research object of biodiversity-protection.

In this paper, we introduce a novel deep learning model to monitoribats and recognize bats species. There are quite a few future related directions to explore, one is to try different model to improve bat calls recognition performance.

References

Authorized licensed use limited to: Thueringer Universitaets. Downloaded on November 16,2023 at 09:49:35 UTC from IEEE Xplore. Restrictions apply.

[1] Brian O'Connor, Stephan Bojinski, Claudia R¨o¨osli, and Michael E. Schaepman. Monitoring global changes learning approaches. Biological Conservation, online(1):1{10, 2019.

[2] Zhou W; Hughes AC. Automatic standardized processing and identification of tropical bat calls using deep

[3] Yi Lin and Yue Liu. Research on recognition and mobile learning of birds base on network under the condition of human-machine collaboration. In Edutainment Technologies. Educational Games and Virtual Reality/Augmented Reality Applications - 6th International Conference on E-learning and Games, Edutainment 2011, Taipei, Taiwan, September 2011. Proceedings, pages 115{122, 2011.

[4] Antonio Rodriguez, Frederic Bartumeus, and Ricard Gavald`a. Machine learning assists the classification of reports by citizens on disease-carrying mosquitoes. In Proceedings of the First Workshop on Data Science for Social Good co-located with European Conference on Machine Learning and Principles and Practice of Knowledge Dicovery in Databases, SoGood@ECML-PKDD 2016, Riva del Garda, Italy, September 19, 2016, 2016.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, pages 770{778, 2016.

[6] Guillem Ramirez-Gargallo, Marta Garcia-Gasulla, and Filippo Mantovani. Tensorflow on state-of-the-art HPC clusters: A machine learning use case. In 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGRID 2019, Larnaca, Cyprus, May 14-17, 2019, pages 526{533, 2019.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. Commun. ACM, 60(6):84{90, 2017.

[8] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.

Authorized licensed use limited to: Thueringer Universitaets. Downloaded on November 16,2023 at 09:49:35 UTC from IEEE Xplore. Restrictions apply.