Bird Audio Detection using Convolutional Neural Networks and Binary Neural Networks

Jianan Song

Beijing University of Posts and
Telecommunications,
Beijing,100876,China
jnsong17@foxmail.com

Shengchen Li

Beijing University of Posts and
Telecommunications,
Beijing,100876,China
shengchen.li@bupt.edu.cn

ABSTRACT

For the bird audio detection task of the IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events (DCASE2017), we propose a audio classification method for bird species identification using Convolutional Neural Networks (CNNs) and Binarized Neural Networks (BNNs).Although deep learning networks is currently popular in bird audio detection[1], the complex network structure makes it difficult to design the hardware of the detection system. Therefore, after the design of the CNNs, the convolutional layer and the fully connected layer are binarized on the basis of the original network, and both network structures are tested. Finally Area Under ROC Curve (AUC) score is used as the evaluation index. The results of using CNNs and BNNs in the preview score are 88.75% and 68.60%.

***Index Terms*—** Bird audio detection, convolutional neural network, binarized neural network

1. THE NETWORK'S SYSTEM

1.1 Convolutional neural network system

The first system trained is a convolutional neural network with six conolutional layers{128, 64, 128, 512,128,512},three fully connected layers{512,128,1}, and the last layer is the sigmoid out-put layer.The input signal is logmel energies.The model was validated by a five-fold cross-validation method.

1.2 Binarized neural network system

Binarized Neural Networks (BNNs) is a kind of compression method of deep learning network, which can make the network structure simple. BNNs can play significant role in compressing the network's model and accelerating the training process. According to the scene characteristics of the audio event detection and the structure of the BNNs[2-4], this research intends to use the binarized neural networks in bird audio detection system which is based on CNNs and the network model is consistent with the previous one.

BNNs constrains both the weights and the activations to either +1 or −1. Those two values are very advantageous from a hardware perspective. is mainly achieved by the symbol function:

$$x^b = Sign(x) = \begin{cases} +1 & if \quad x \geq 0, \\ -1 & otherwise \end{cases} \qquad (1)$$

where x^b is the binarized variable (weight or activation) and x is the real-valued variable. In this way, the input of the other layers is 1 bit except the first layer of the floating-point data.

Although the parameters of BNNs and the activation values for each layer are binarized, the gradient has to be stored with higher-precision floating-point data instead of binarized data. The main reasons are as follows: firstly, the magnitude of the gradient is small; secondly, the gradient has a cumulative effect, and it has a certain amount of noise. The noise is generally considered to be as Normal Distribution. Therefore, the gradient needs more Sub-cumulative in order to make the average noise disappear. Another problem is that the gradient of Sign (x) function is zero, so we can't use gradient descent method to train the network. In order to solve this problem, we use a Hard tanh function (Figure 1) to approximate Sign (x) function.

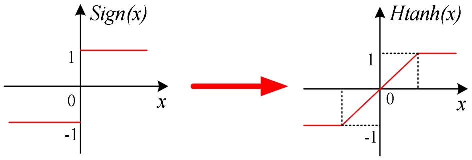

Figure 1: The picture of Sign(x) adn Htanh(x).

2. THE EXPERIMENTAL RESULT

Under two identical network structures, the AUC results of the CNN and BNN on development and challenge datasets are as follows:

Table 1. AUC scores on development set

Dataset	Method	
	CNN	BNN
Development	0.99	0.698
Evaluation	0.934	0.680

Table 2. Preview AUC Score on challenge dataset

CNN	BNN
0.8775	0.6960

3. REFERENCES

[1] EmreÇakır, Parascandolo G, et al. Convolutional Recurrent Neural Networks for Bird Audio Detection[J]. 2017.

[2] P. Smaragdis, "Bitwise Neural Networks,"2016.

[3] Y. Bengio, "Binarized Neural Networks: Training Deep Neural Networks with Weights and Activations Constrained to +1 or -1," 2016.

[4] M. Rastegari, V. Ordonez, J. Redmon, and A. Farhadi, "XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks," pp. 1–17, 2016.