ONE SHOT LEARNING FOR ACOUSTICS CLASSIFICATION OF MALAYSIA BIRD SPECIES

KOAY XIAN HONG

UNIVERSITI TEKNOLOGI MALAYSIA

ONE SHOT LEARNING FOR ACOUSTICS CLASSIFICATION OF MALAYSIA BIRD SPECIES

KOAY XIAN HONG

A project report submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Engineering (Computer & Microelectronic Systems)

School of Electrical Engineering
Faculty of Engineering
Universiti Teknologi Malaysia

JULY 2022

DEDICATION

This thesis is dedicated to my family. Whom have supported me throughout my journey in life

ACKNOWLEDGEMENT

During the preparation of this thesis, I have received help from people, indirectly or otherwise. Some have helped me go on in hard times, others improved my understanding and thoughts of the subject matter. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Prof. Madya Muhammad Mun'im Bin Ahmad Zabidi, for then encouragement, guidance, critics and friendship. Additionally, I would like to thank Dr Puan Chong Leong from Universiti Putra Malaysia and his team for providing the dataset required for this thesis to work. Without their help and contribution, this thesis would be impossible to achieve with the time constraint given.

My friends from postgraduate aid and support are very much appreciated. My sincere gratitude also extends to all my colleagues and others who have aided me at various occasions. I am grateful to all my family members for caring throughout the writing of this thesis.

ABSTRACT

Malaysia is famed for its beautiful bio-diverse forest and its bird species, some of it is still understudied. Using acoustic detection, we can study these bird as current advanced in machine learning application have resulted in cutting edge performance for acoustic classification application. However, most of these applications need large amount of data for prediction to have acceptable accuracy. We, however, do not have this kind of resources. This situation is experienced by large demographic of this country, which provide importance to our studies. Adding to that problem, one common issue that come with studying of species is that we can only know the status of species in a habitat up to certain point. Such problem needs to be solve using methodologies that can cope with the fluidity of information. As such, we propose neural network framework that able to notice any changes in the class categories and learn new classes on the fly. To solve our issue, we seek to design a Siamese Style convolutional Neural Network for one-shot-learning architecture. Additionally, we would train it using base convolutional neural network with low complexity so that it can be realistically implement in hardware of low computing power. We evaluated and benchmarked our framework, showing promising results as the Network is able to classify trained bird species with accuracy of 90% or higher with only around 100 sound clips per bird species. Additionally, it is able to detect new bird species on the fly and add it to its class successfully, however, it still needs some work as the accuracy is around 50% on this part. All of these is achieved using base Convolutional Neural Networks of low complexity with only 4 layers of conv layers and 2 layers of fully connected layers. From this thesis, It is shown that this neural network can work and I am optimistic that this work can be further improved, which can be done by using a higher variety of dataset and transfer learning and a further tweaking of the base neural network architecture.

ABSTRAK

Malaysia terkenal dengan hutan bio-pelbagai yang indah dan spesies burungnya, sebahagian daripadanya masih belum dipelajari. Menggunakan pengesanan akustik, kami boleh mengkaji burung ini kerana terkini dalam aplikasi pembelajaran mesin telah menghasilkan prestasi canggih untuk aplikasi klasifikasi akustik. Walau bagaimanapun, kebanyakan aplikasi ini memerlukan sejumlah besar data untuk ramalan mempunyai ketepatan yang boleh diterima. Kami, bagaimanapun, tidak mempunyai sumber seperti ini. Keadaan ini dialami oleh demografi besar negara ini, yang memberikan kepentingan kepada kajian kita. Menambah kepada masalah itu, satu isu biasa yang datang dengan mengkaji spesies ialah kita hanya boleh mengetahui status spesies dalam habitat sehingga tahap tertentu. Masalah sedemikian perlu diselesaikan menggunakan metodologi yang boleh mengatasi kecairan maklumat. Oleh itu, kami mencadangkan rangka kerja rangkaian saraf yang dapat melihat sebarang perubahan dalam kategori kelas dan mempelajari kelas baharu dengan cepat. Untuk menyelesaikan isu kami, kami berusaha untuk mereka bentuk Rangkaian Neural Konvolusi Gaya Siam untuk seni bina pembelajaran sekali sahaja. Selain itu, kami akan melatihnya menggunakan rangkaian neural convolutional asas dengan kerumitan rendah supaya ia boleh dilaksanakan secara realistik dalam perkakasan kuasa pengkomputeran rendah. Kami menilai dan menanda aras rangka kerja kami, menunjukkan hasil yang menjanjikan kerana Rangkaian dapat mengklasifikasikan spesies burung terlatih dengan ketepatan 90% atau lebih tinggi dengan hanya sekitar 100 klip bunyi bagi setiap spesies burung. Selain itu, ia dapat mengesan spesies burung baharu dengan cepat dan berjaya menambahkannya ke kelasnya, namun, ia masih memerlukan sedikit usaha kerana ketepatannya adalah sekitar 50% pada bahagian ini. Semua ini dicapai menggunakan Rangkaian Neural Konvolusi asas dengan kerumitan rendah dengan hanya 4 lapisan lapisan penukaran dan 2 lapisan lapisan bersambung sepenuhnya. Daripada tesis ini, Ia menunjukkan bahawa rangkaian saraf ini boleh berfungsi dan saya optimis bahawa kerja ini boleh dipertingkatkan lagi, yang boleh dilakukan dengan menggunakan pelbagai set data dan pembelajaran pemindahan yang lebih tinggi dan pengubahsuaian lanjut seni bina rangkaian neural asas.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CNN	-	Convolutional Neural Network
ReLU	-	Rectified Linear activation Unit
LCSED	-	Low Complexity Sound Event Detection
VGG	-	Visual Geometry Group
SNN	-	Siamese Neural Network
SCNN	-	Siamese Convolutional Neural Network
NN	-	Neural Network
FC	-	Fully connected
ARU	-	Automated Recording unit

LIST OF SYMBOLS

p	-	Prediction vector
σ	-	sigmoidal activation function
w	-	weight
λ	-	regularization

CHAPTER 1

INTRODUCTION

1.1 Overview

Malaysia is widely recognised for its biodiverse rainforest ecosystem, which is home to a plethora of uncommon bird species. Because of the difficulty in monitoring these species, they are frequently understudied. Nonetheless, researching these avian behaviours is critical for numerous possible uses in conservation, ecology, and archiving. This necessitates the usage and development of remote monitoring sound recording instruments and acoustic analysis software.

Any type of automated classification of bird sound data gained from rainforest recordings would be of great use in the area of ornithology. Avian communication through sounds is frequently the quickest and most important means for ornithologists to discover birds. Bird calls are one of the most important forms of communication for birds since they may be used for a variety of goals such as intimidation, reproduction, or making their presence known in general. Recordings acquired from the environment frequently contain considerable background noise that must be removed. Furthermore, manual identification of bird species is dependent on the competence level of individual specialists and, as a result, might be unreliable.

As a result of the widespread availability of automated recording units (ARUs), research into automatic categorization of bird species by bioacoustics has grown in popularity (Digby et al. 2013, Shonfield et al. 2017). These technologies provide rapid, efficient, and non-intrusive monitoring of rainforest bird species. However, when it comes to conducting an effective bird acoustic survey in the field, we still face a few challenges. The problems were as follows: (a) inability to recognise new species, (b) vast bird sounds database required, and (c) automatic bird categorization software installed in ARUs may cause slowdown or severe battery drain.

To that purpose, we propose looking into the Siamese neural network framework's potential to do one-shot learning. Utilizing One-shot Learning, neural network frameworks may "learn" to recognise new classes by merely viewing a single sample and using it to make accurate inference to that specific class (Lungu et al. 2020, Koch et al. 2015). This approach is well-studied in the realm of image classification but understudied in the domain of audio classification. The goal behind this proposal is to modify the one-shot learning approach to address the underlying issue of detecting new species and the requirement for a big dataset for training. The primary goal of this project is to (a) create a framework that can detect and adapt to changes in the class dictionary, allowing the framework to essentially learn on the fly, (b) achieve acceptable performance while training with a small amount of data, and (c) do all of this without compromising performance when installed in an ARU.

1.2 Problem Background

Malaysia is well-known for its magnificent bio-diverse forests and uncommon bird species. However, due to the variety of Malaysia's forest, numerous bird species remain unstudied. For example, there are around seventeen species of owl in Malaysia that are currently understudied (Puan et al. 2016), and our knowledge and comprehension of the ecology of these owl species is limited. This makes maintaining and safeguarding the habitats of bird species challenging, posing a threat to their existence.

Birds are highly sensitive to their surroundings, therefore studying them may reveal valuable information about their ecosystem. There are several methods for tracking birds in the jungle. One of them is keeping an eye on them via their calls. Birds are very loud creatures, and their calls serve a variety of functions, including attracting mates, intimidating predators, and communicating. Point counting is a popular approach for monitoring and assessing birds. This approach is done by having ornithologists record any observed species on the field. The difficulty is that ornithologists cannot perform this 24/7, and it is extremely dependent on the expertise of individual specialists (Scott et al 2008).

Monitoring with Autonomous Recording Units (ARUs), which are embedded devices built specifically to capture sound, is the next best thing to manual approaches (Wilhite et al. 2020). These ARUs are planted at various points around the forest, and bird calls captured this way may be processed and analysed later in the lab. Kaleidoscope is the industry standard software for sound data analysis; it is a highly specialised instrument for processing and identifying recognised species. However, this programme has a few drawbacks. Like with every specialised instrument, it must be operated by a professional. Furthermore, using this programme is not only time-consuming, but also difficult to use when recognising new species (Knight et al. 2017, Dema et al. 2017, Jean at al. 2021). As a result, any innovative ways of analysing sound data are encouraged.

Several methods have been developed to automatically analyse bird sound recordings (Stowell et al. 2016, Digby et al. 2013). In recent years, neural networks and deep learning models have grown in popularity for detecting and classifying auditory events (Mesaros et al. 2010, Parascandolo et al. 2016). As a result, a growing number of automated bird sound detection systems employ this concept (Chandu et al. 2020, Bedoya et al. 2021, Rong et al. 2012, Xie et al. 2019). However, there are certain challenges to solve when using the Neural Network architecture for efficient bird sound monitoring in the forest. For example, when it comes to unknown species, normal techniques would struggle to identify it (Stowell et al. 2014). Furthermore, deep learning approaches require a big sound database in order to work correctly (Bedoya et al. 2021). Another impediment is the ability to use on-board deep learning in systems with limited computing capability, such as ARUs (Acconcjaioco and Ntalampiras 2020). For example, when it comes to identifying animals, we frequently presume that we already have a substantial quantity of data about the species we are classifying. There has been very little research on developing a framework for classifying unknowns.

In conclusion, techniques for detecting new classes of bird sounds when vast amounts of data are absent are mostly unexplored.

1.3 Problem statement

We need a neural network framework that can (a) train with a small dataset, (b) detect changes in the class category list and learn new classes on the fly, and (c) do both without sacrificing performance during inference.

1.4 Aim

To propose a neural network framework that can: (a) train with a small dataset, (b) detect changes in the class category list and learn new classes on the fly, and (c) do so without sacrificing performance during inference.

1.5 Research Objective

The objective of this project is:

1. To review and investigate recent approaches in bird sound detection and identify their shortcomings.

2. To find a deep learning approach that can be trained under a relatively small dataset, can detect new class categories on the fly and can be inference quickly without compromising performance, especially in an environment with low computing power.

3. To design/implement the solution to meet the performance objectives

4. To benchmark and evaluate the proposed solution by comparing the performance to existing work

1.6 Scopes of Study

We will be training our Neural Network using a dataset collected by Puan Chong Leong and his team from Universiti Putra Malaysia (UPM). The Dataset consists of bird species from Malaysia. Classification of species from other countries will not be considered in our project.

We will only be evaluating our Neural Network performance only after training, we will not consider training time as part of performance review. In addition to that, we will only be evaluating our using our local CPU, we will not be evaluating the framework in a remote environment.

We will only focus on evaluating and improving the Neural Network framework. The pre-processing methodology, which is present in any form of neural network architecture, will not be considered when reviewing. We will also be only focusing on reviewing one type of Neural network framework rather reviewing multiple neural network framework

REFERENCES

A. Mesaros, T. Heittola, A. Eronen and T. Virtanen, "Acoustic event detection in real life recordings," 2010 18th European Signal Processing Conference, 2010, pp. 1267-1271.

B. Carol & M. Laura. (2021). Acoustic Censusing and Individual Identification of Birds in the Wild. 10.1101/2021.10.29.466450

B. Chandu, A. Munikoti, K. S. Murthy, G. Murthy V. and C. Nagaraj, "Automated Bird Species Identification using Audio Signal Processing and Neural Networks," 2020 International Conference on Artificial Intelligence and Signal Processing (AISP), 2020, pp. 1-5, doi: 10.1109/AISP48273.2020.9073584.

C. Max. (2019). Polyphonic Bird Sound Event Detection With Convolutional Recurrent Neural Networks. 10.13140/RG.2.2.11943.09126.

Chicco D. (2021) Siamese Neural Networks: An Overview. In: Cartwright H. (eds) Artificial Neural Networks. Methods in Molecular Biology, vol 2190. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-0826-5_3

D. Stowell, M. Wood, Y. Stylianou and H. Glotin, "Bird detection in audio: A survey and a challenge," 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), 2016, pp. 1-6, doi: 10.1109/MLSP.2016.7738875.

D. Sujoy & R. Partha & Ali, Amin & Amin, Md Ashraful. (2016). Identification of bird species from their singing. 182-186. 10.1109/ICIEV.2016.7759992.

Digby, A., Towsey, M., Bell, B.D. and Teal, P.D. (2013), A practical comparison of manual and autonomous methods for acoustic monitoring. Methods Ecol Evol, 4: 675-683. https://doi.org/10.1111/2041-210X.12060

Fan, J., Nichols, E., Tompkins, D.C., Méndez, A.E., Elizalde, B., & Pasquier, P. (2020). Multi-Label Sound Event Retrieval Using A Deep Learning-Based Siamese Structure With A Pairwise Presence Matrix. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 3482-3486.

G. Parascandolo, H. Huttunen and T. Virtanen, "Recurrent neural networks for polyphonic sound event detection in real life recordings," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 6440-6444, doi: 10.1109/ICASSP.2016.7472917.

Gabrielli, L., Ambrosini, L., Vesperini, F., Bruschi, V., Squartini, S., & Cattani, L. (2019). Processing Acoustic Data with Siamese Neural Networks for Enhanced Road Roughness Classification. 2019 International Joint Conference on Neural Networks (IJCNN), 1-7.

Guzhov, A., Raue, F., Hees, J., & Dengel, A. (2021). Audioclip: Extending clip to image, text and audio. arXiv preprint arXiv:2106.13043.

I. A. Lungu, Y. Hu and S. -C. Liu, "Multi-Resolution Siamese Networks for One-Shot Learning," 2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), 2020, pp. 183-187, doi: 10.1109/AICAS48895.2020.9073996.

J. Marchal, François Fabianek & Yves Aubry (2021) Software performance for the automated identification of bird vocalisations: the case of two closely related species, Bioacoustics, DOI: 10.1080/09524622.2021.1945952

J. Xie, K. Hu, M. Zhu, J. Yu and Q. Zhu, "Investigation of Different CNN-Based Models for Improved Bird Sound Classification," in IEEE Access, vol. 7, pp. 175353-175361, 2019, doi: 10.1109/ACCESS.2019.2957572.

Knight, E. C., K. C. Hannah, G. Foley, C. Scott, R. Mark Brigham, and E. Bayne. 2017. Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. Avian Conservation and Ecology 12(2):14. https://doi.org/10.5751/ACE-01114-120214

Koch, G.; Zemel, R. & Salakhutdinov, R. (2015), Siamese Neural Networks for One-shot Image Recognition, in .

L. Torres, N. Monteiro, J. Oliveira, J. Arrais and B. Ribeiro, "Exploring a Siamese Neural Network Architecture for One-Shot Drug Discovery," 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), 2020, pp. 168-175, doi: 10.1109/BIBE50027.2020.00035.

Lasseck M. (2018). Audio-based Bird Species Identification with Deep Convolutional Neural Networks.

Liu, Y., Liu, A., Su, Y., Schiele, B., & Sun, Q. (2020). Mnemonics Training: Multi-Class Incremental Learning Without Forgetting. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 12242-12251.

M. Acconcjaioco and S. Ntalampiras, "One-shot learning for acoustic identification of bird species in non-stationary environments," 2020 25th International Conference on Pattern Recognition (ICPR), 2021, pp. 755-762, doi: 10.1109/ICPR48806.2021.9412005.

M. E. Hossain, A. Islam and M. S. Islam, "A Proficient Model to Classify Bangladeshi Bank Notes for Automatic Vending Machine Using a Tıny Dataset with One-Shot Learning & Siamese Networks," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-4, doi: 10.1109/ICCCNT49239.2020.9225405.

Masana, M., Liu, X., Twardowski, B., Menta, M., Bagdanov, A.D., & Weijer, J.V. (2020). Class-incremental learning: survey and performance evaluation on image classification.

Morfi, V., Bas, Y., Pamuła, H., Glotin, H., & Stowell, D. (2019). NIPS4Bplus: a richly annotated birdsong audio dataset. PeerJ Computer Science, 5, e223.

N. I. Ahmad Sabri and S. Setumin, "One-Shot Learning for Facial Sketch Recognition using the Siamese Convolutional Neural Network," 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), 2021, pp. 307-312, doi: 10.1109/ISCAIE51753.2021.9431773.

P. Jordi & Serrà, Joan & Serra, Xavier. (2018). Training neural audio classifiers with few data.

P. Manocha, R. Badlani, A. Kumar, A. Shah, B. Elizalde and B. Raj, "Content-Based Representations of Audio Using Siamese Neural Networks," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 3136-3140, doi: 10.1109/ICASSP.2018.8461524.

R. Sun and N. Mita, "Nocturnal wild bird species identification by sound information using wavelet," 2012 International Conference on Wavelet Active Media Technology and Information Processing (ICWAMTIP), 2012, pp. 1-3, doi: 10.1109/ICWAMTIP.2012.6413425.

S. Brandes, T. (2008). Automated sound recording and analysis techniques for bird surveys and conservation. Bird Conservation International, 18(S1), S163-S173. doi:10.1017/S0959270908000415

S. S. R. Phaye, E. Benetos and Y. Wang, "SubSpectralNet – Using Sub-spectrogram Based Convolutional Neural Networks for Acoustic Scene Classification," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 825-829, doi: 10.1109/ICASSP.2019.8683288.

Salamon J, Bello JP, Farnsworth A, Robbins M, Keen S, et al. (2016) Towards the Automatic Classification of Avian Flight Calls for Bioacoustic Monitoring. PLOS ONE 11(11): e0166866. https://doi.org/10.1371/journal.pone.0166866

Shonfield, J., and E. M. Bayne. 2017. Autonomous recording units in avian ecological research: current use and future applications. Avian Conservation and Ecology 12(1):14. https://doi.org/10.5751/ACE-00974-120114

Stowell D. & P. Mark. (2014). Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. PeerJ. 2. e488. 10.7717/peerj.488.

Stowell, D., Wood, M. D., Pamuła, H., Stylianou,d Y., & Glotin, H. (2019). Automatic acoustic detection of birds through deep learning: the first Bird Audio Detection challenge. Methods in Ecology and Evolution, 10(3), 368-380.

T. Dema et al., "An Investigation into Acoustic Analysis Methods for Endangered Species Monitoring: A Case of Monitoring the Critically Endangered White-Bellied Heron in Bhutan," 2017 IEEE 13th International Conference on e-Science (e-Science), 2017, pp. 177-186, doi: 10.1109/eScience.2017.30.

T. Grill and J. Schlüter, "Two convolutional neural networks for bird detection in audio signals," 2017 25th European Signal Processing Conference (EUSIPCO), 2017, pp. 1764-1768, doi: 10.23919/EUSIPCO.2017.8081512.

Wilhite, N.G., Howell, P.E. and Martin, J.A. (2020), Evaluation of Acoustic Recording Devices to Survey Northern Bobwhite Populations. Wildl. Soc. Bull., 44: 200-207. https://doi.org/10.1002/wsb.1061

Y. Mingxue, P. Lujie, Li Liu, Yujiang Wang, Zhenyuan Zhang, Zhengxi Yuan, Jun Zhou, “LCSED: A low complexity CNN based SED model for IoT devices”, Neurocomputing, 2021,, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2021.02.104.

Y. R. Pandeya, B. Bhattarai and J. Lee, "Sound Event Detection in Cowshed using Synthetic Data and Convolutional Neural Network," 2020 International Conference on Information and Communication Technology Convergence (ICTC), 2020, pp. 273-276, doi: 10.1109/ICTC49870.2020.9289545.

Y. Zhang and Z. Duan, "IMINET: Convolutional semi-siamese networks for sound search by vocal imitation," 2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2017, pp. 304-308, doi: 10.1109/WASPAA.2017.8170044.

Y. Zhang and Z. Duan, "Visualization and Interpretation of Siamese Style Convolutional Neural Networks for Sound Search by Vocal Imitation," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 2406-2410, doi: 10.1109/ICASSP.2018.8461729.

Y. Zhang, B. Pardo and Z. Duan, "Siamese Style Convolutional Neural Networks for Sound Search by Vocal Imitation," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 27, no. 2, pp. 429-441, Feb. 2019, doi: 10.1109/TASLP.2018.2868428.

Y. Siew & C.L. Puan & Chang, Phooi & Azhar, Badrul. (2016). Vocal Individuality of Sunda Scops-Owl (Otus lempiji) in Peninsular Malaysia. Journal of Raptor Research. 50. 379-390. 10.3356/JRR-15-76.1.

Zhang, J., Zhang, J., Ghosh, S., Li, D., Tasci, S., Heck, L., Zhang, H., & Kuo, C.J. (2020). Class-incremental Learning via Deep Model Consolidation. 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), 1120-1129.