2D CONVOLUTIONAL NEURAL NETWORK FOR THE DETECTION OF ASIAN KOEL (EUDYNAMYS SCOLOPACEUS) VOCALIZATIONS

NURFAIZAH BINTI MUSA

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

School of Electrical Engineering
Faculty of Engineering
Universiti Teknologi Malaysia

JULY 2020

DEDICATION

This project report is specially dedicated to my father, Musa bin Md Yasin, who has been my source of strength and inspiration when I thought of giving up. It is also dedicated to my mother, Kalsom bte Mohd Amin, who has taught me the meaning of patience in facing hardship. Not to forget, to my family members, Faiz, Faris, Farihah, Fatimah and Faliq for all their overflowing supports. I thank the Almighty Allah for all His blessings, for the supportive and kind people He surrounds me.

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious, and the Most Merciful.

Alhamdulillah, all praise to Allah SWT for granting me the health and strength in completing this final year project. My deepest appreciation goes to my supervisor, PM Muhammad Mun’im bin Ahmad Zabidi for all his unconditional guidance and supervision throughout this project. His constructive comments and thoughtful ideas had contributed a lot of positive outcomes to this project.

I would like to express my deepest gratitude to my beloved parents and family who had always supported and encouraged me through this tough journey in UTM. Their endless love and motivation have brought me the strength to complete this program.

Also, to my fellow friends and colleagues, I thank them for their sincere kindness and supportive thoughts throughout this project. Their opinions, thoughts and assistance are very helpful. Unfortunately, it is not possible to list all the people who had contributed in completing my final year project in this very limited space. Thank you very much.

ABSTRACT

Acoustic activity detection plays a vital role for automatic wildlife monitoring which includes the study of ecology, populations and habitats assessments. Birds are one of the few wildlife species to be monitored as their population and distribution are expected to change due to climate changes in order to conserve the ecosystem, diversity and seasonal population changes. Monitoring animals based on sound (bioacoustics) monitoring involves continuous observation to capture rare events. Several existing bird sound classification devices records sounds at point reading and processed the data off-line that involves complex Convolution Neural Network (CNN) architecture which takes longer time in the processing stages as data needs to be acquired before being processed. This approach is not applicable on the real-time monitoring. Therefore, this project investigates the best architecture that can be implemented to lower the complexity of algorithm for a bird sound classification. Data training with bird sound from all over the world and non-bird sounds will be done in optimizing the algorithm. In precise, this project focuses on a bird sound classification with low resource CNN to classify an Eudynamys Scolopaceus bird species. The bird sound detection will be assessed on the Xeno-Canto dataset which is a dataset containing bird vocalization samples and Urban8k that is shared openly are used for training and testing. Data segmentation is done on each of the samples with 16kHz sampling frequency of 25% overlapping to avoid data loss. Segmented samples are then converted into spectrograms and fed into MobileNet CNN and Bulbul CNN Architecture for training and testing. A set of testing samples were used to predict the accuracy of each model and prediction results were presented in a confusion matrix. Results from both comparisons showed that MobileNet has a higher accuracy of 80% than Bulbul CNN with 64%. Further development and optimization of model architecture with the use of more training samples can be done in the future towards achieving a higher accuracy in classifying the bird sound.

ABSTRAK

Pengesanan aktiviti akustik memainkan peranan penting dalam pemantauan hidupan liar secara automatik yang merangkumi kajian ekologi, populasi dan penilaian habitat. Burung adalah salah satu dari spesies hidupan liar yang akan dipantau oleh kerana populasi dan taburannya yang dijangkakan akan berubah akibat perubahan iklim untuk melestarikan ekosistem, kepelbagaian dan perubahan populasi bermusim. Memantau haiwan berdasarkan pemantauan bunyi (bioakustik) melibatkan pemerhatian berterusan untuk menangkap kejadian yang jarang berlaku. Beberapa alat klasifikasi bunyi burung yang merekodkan suara ketika membaca dan memproses data secara luar talian yang melibatkan seni bina Rangkaian Neural Konvolusi (CNN) yang kompleks yang memerlukan masa lebih lama dalam pemprosesan kerana data perlu diperoleh sebelum diproses. Pendekatan ini tidak sesuai digunakan untuk pemantauan masa nyata. Oleh itu, projek ini menyiasat seni bina terbaik yang dapat dilaksanakan untuk menurunkan kerumitan algoritma untuk klasifikasi bunyi burung. Latihan data dengan suara burung dari seluruh dunia dan suara bukan burung akan dilakukan dalam mengoptimumkan algoritma. Tepatnya, projek ini memfokuskan pada klasifikasi suara burung dengan CNN sumber rendah untuk mengklasifikasikan spesies burung Eudynamys Scolopaceus. Pengesanan bunyi burung akan dinilai pada dataset Xeno-Canto yang merupakan dataset yang berisi sampel vokalisasi burung dan Urban8k yang dikongsi secara terbuka untuk tujuan latihan dan ujian. Segmentasi data dilakukan pada setiap sampel dengan 16kHz frekuensi persampelan 25% bertindih untuk mengelakkan kehilangan data. Sampel yang tersegmentasi kemudian diubah menjadi spektrogram dan dimasukkan ke dalam senibina MobileNet CNN dan Bulbul CNN untuk latihan dan ujian. Satu set sampel ujian digunakan untuk meramalkan ketepatan setiap model dan hasil ramalan diganbarkan dalam matriks konfusi. Hasil dari kedua perbandingan tersebut menunjukkan bahawa MobileNet mempunyai ketepatan yang lebih tinggi iaitu 80% daripada Bulbul CNN dengan 64%. Pengembangan lebih lanjut dan pengoptimuman seni bina model dengan penggunaan lebih banyak sampel latihan dapat dilakukan di masa depan untuk mencapai ketepatan yang lebih tinggi dalam mengklasifikasikan suara burung.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ARU	-	Autonomous Recording Units
CNN	-	Convolutional Neural Network
DNN	-	Deep Neural Network
HMM	-	Hidden Markov Models
MAC	-	Multiples and Accumulative
MIML	-	Multi-instance Multi-label
MIR	-	Music Information Retrieval
MFCC	-	Mel-Frequency Cepstral Coefficients
MMSE	-	Minimum Mean Square Error
RELU	-	Rectified Linear Unit
RWCP	-	Real-World Computing Partnership
SISL	-	Single-instance Single-label
STFT	-	Short Time Fourier Transform
SVM	-	Support Vector Machine

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Research Background

Birds are a vital part of the ecosystems as they perform important tasks in balancing such as controlling crop pest, pollinate and disperse seed of many plants including many crops important to human and some birds like vultures also helps with the decomposition of organic materials [1]. *Eudynamys Scolopaceus,* commonly known as Asian Koel, is one of the bird species that contributes to seed dispersion around Asia as they inhabit many parts in Asia. This species migrates due to climate changes where it spends the summer in plains of Pakistan and migrates towards India during winter [2].

The number of birds all around the world is becoming to worrying as a hundred bird species have vanished even since 1600. This is due to the loss of habitat resulting from over exploitation by human beings and overhunting. Other factors contributing to the decreasing number of bird species also includes introduced predators in their surroundings. As a result, highly threatened species is reaching extinction and bird species and groups is declining fast [3][4]. Conservation of bird population has been done by monitoring the bird's population [5]. This can help to quantify the impact of a certain land use, study of ecology and the biodiversity of the bird's habitat. Nowadays, many methods have been introduced from the most traditional method to the current conventional method. One of the methods for bird monitoring bird is by point count, where an expert tallies the birds by sight and sound from a fixed position for a set period of time. This method is quite tedious and cannot be done during night-time [6][7].

As technology advances, autonomous recording units (ARU) was introduced to counter the drawback of manual birdwatching. In this method, bird sounds are recorded round-the-clock and data is collected and brought back to the laboratory to be analysed by experts [8]. Today, a species recognition by point count has also been enhanced where bird sounds were recorded, and deep learning is used to automate the recognition process [9].

1.2 Problem Statement

Many methods to automate bird sound detection has been introduced. The most widely used today is autonomous recording units. In this method, bird sounds were recorded at a reading point and deep learning is used to do the detection without having manual expertise help [10]. However, this method typically requires convolutional neural network (CNN), a deep learning algorithm in data analysis which is very complex and takes longer time in the processing stages as data needs to be acquired before being processed [9].

In addition, data recorded at point reading produces large volumes of data for online processing. This technique requires additional setup to solve the storage and data processing timing issues. Therefore, this project investigates the best architecture that can be implemented to lower the complexity of algorithm for the detection of Asian Koel (*Eudynamys Scolopaceus)* bird species. This species is chosen in this project as the species is quite common and produces loud and distinguishable vocalization calls. Data training with bird sound from all over the world and non-bird sounds will be done in optimizing the algorithm. In precise, this project focuses on a bird sound detection with low resource CNN to detect Eudynamys *Scolopaceus* vocalizations.

1.3 Research Objective

The objectives of this project are listed as below:

(a) To investigate the best architecture for a lightweight CNN algorithm in detecting *Eudynamys Scolopaceus* bird species

(b) To implement the low-complexity CNN model in a bird detection algorithm using Bulbul CNN.

(c) To evaluate and analyze the performance algorithm in detecting *Eudynamys Scolopaceus* bird species

1.4 Scope of Project

The scope of this study includes the investigation on the best and less complex CNN algorithm that is suitable for a bird sound detection. In order to achieve the objectives of the project, few scopes are included in this project. The scopes are listed as below:

(a) Bird sound data collection from Xeno-Canto website using Rstudio

(b) Data segmentation and augmentation using Matlab

(c) Data training and testing using GoogleColab platform

1.5 Thesis Outline

This thesis consists of five chapters. Chapter 1 discusses the project introduction, problem statement, objectives and scopes of this project. The main objective of this project is to investigate the best architecture for a lightweight CNN algorithm in detecting *Eudynamys Scolopaceus* bird species by comparing MobileNet CNN with Bulbul CNN.

In Chapter 2, discussion on the acoustic scene classification and literature review along with the previous work studies on the bird sound detection particularly the different approaches used in classifying the sound. A review on CNN architectures are also discussed in this chapter.

In Chapter 3, the techniques and methodology throughout the project is discussed. Methods in dataset preparation that include data segmentation is explained in detail. MobileNet CNN and Bulbul were built, and prepared dataset was trained on built models. Lastly, the benchmarking of investigated algorithm is also discussed in this chapter.

All results and discussion for this project will be presented in the next chapter, Chapter 4. Problems faced solutions to overcome the problems will also be discussed in this chapter. The novelty of the results and findings will be mentioned in this chapter as well. Lastly, Chapter 5 will brief on the expected outcome of this project within the time allocated.

REFERENCES

[1] M. Sankupellay and D. Konovalov, "Bird call recognition using deep convolutional neural network, ResNet-50," *Aust. Acoust. Soc. Annu. Conf. AAS 2018*, no. October, pp. 200–207, 2019, doi: 10.13140/RG.2.2.31865.31847.

[2] A. A. Khan and I. Z. Qureshi, "Vocalizations of adult male Asian koels (Eudynamys scolopacea) in the breeding season," *PLoS One*, vol. 12, no. 10, pp. 1–13, 2017, doi: 10.1371/journal.pone.0186604.

[3] "New study: conservation action has reduced bird extinction rates by 40% | BirdLife." https://www.birdlife.org/worldwide/news/new-study-conservation-action-has-reduced-bird-extinction-rates-40 (accessed Jul. 09, 2020).

[4] M. J. Monroe, S. H. M. Butchart, A. O. Mooers, and F. Bokma, "The dynamics underlying avian extinction trajectories forecast a wave of extinctions," *Biol. Lett.*, vol. 15, no. 12, Dec. 2019, doi: 10.1098/rsbl.2019.0633.

[5] S. Kahl *et al.*, "Overview of BIRDCLEF 2019: Large-scale bird recognition in soundscapes," *CEUR Workshop Proc.*, vol. 2380, pp. 9–12, 2019.

[6] L. Nanni, Y. M. G. Costa, D. R. Lucio, C. N. Silla, and S. Brahnam, "Combining Visual and Acoustic Features for Bird Species Classification," no. October 2017, pp. 396–401, 2017, doi: 10.1109/ictai.2016.0067.

[7] V. M. Trifa, A. N. G. Kirschel, C. E. Taylor, and E. E. Vallejo, "Automated species recognition of antbirds in a Mexican rainforest using hidden Markov models," *J. Acoust. Soc. Am.*, vol. 123, no. 4, pp. 2424–2431, 2008, doi: 10.1121/1.2839017.

[8] K. Darras, P. Batáry, B. Furnas, I. Fitriawan, Y. Mulyani, and T. Tscharntke, "Autonomous bird sound recording outperforms direct human observation: Synthesis and new evidence," *bioRxiv*, p. 117119, 2017, doi: 10.1101/117119.

[9] J. R. Aist and C. J. Bayles, *Ultrastructural basis of mitosis in the fungus Nectria haematococca (sexual stage of Fusarium solani) - II. Spindles*, vol. 161, no. 2–3. 1991.

[10] S. Kahl *et al.*, "Large-scale bird sound classification using convolutional neural networks," *CEUR Workshop Proc.*, vol. 1866, 2017.

[11] D. Stowell, M. Wood, Y. Stylianou, and H. Glotin, “Bird detection in audio: A survey and a challenge,” *IEEE Int. Work. Mach. Learn. Signal Process. MLSP*, vol. 2016-Novem, 2016, doi: 10.1109/MLSP.2016.7738875.

[12] J. Cai, D. Ee, B. Pham, P. Roe, and J. Zhang, “Sensor network for the monitoring of ecosystem: bird species recognition,” *Proc. 2007 Int. Conf. Intell. Sensors, Sens. Networks Inf. Process. ISSNIP*, pp. 293–298, 2007, doi: 10.1109/ISSNIP.2007.4496859.

[13] M. Towsey, B. Planitz, A. Nantes, J. Wimmer, and P. Roe, “A toolbox for animal call recognition,” *Bioacoustics*, vol. 21, no. 2, pp. 107–125, 2012, doi: 10.1080/09524622.2011.648753.

[14] I. I. Conference and S. Processing, “IMPROVED MUSIC FEATURE LEARNING WITH DEEP NEURAL NETWORKS Siddharth Sigtia , Simon Dixon Queen Mary University of London Centre for Digital Music Mile End Road , London E1 4NS , UK,” pp. 7009–7013, 2014.

[15] S. Dieleman and B. Schrauwen, “End-to-end learning for music audio,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 6964–6968, 2014, doi: 10.1109/ICASSP.2014.6854950.

[16] D. Barchiesi, D. D. Giannoulis, D. Stowell, and M. D. Plumbley, “Acoustic Scene Classification: Classifying environments from the sounds they produce,” *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 16–34, 2015, doi: 10.1109/MSP.2014.2326181.

[17] J. Cho, S. Yun, H. Park, J. Eum, and K. Hwang, “Acoustic Scene Classification Based on a Large-margin Factorized CNN,” no. October, pp. 45–49, 2019, doi: 10.33682/8xh4-jm46.

[18] S. Abdoli, P. Cardinal, and A. Lameiras Koerich, “End-to-end environmental sound classification using a 1D convolutional neural network,” *Expert Syst. Appl.*, vol. 136, pp. 252–263, 2019, doi: 10.1016/j.eswa.2019.06.040.

[19] R. V. Sharan and T. J. Moir, “Robust acoustic event classification using deep neural networks,” *Information Sciences*, vol. 396. pp. 24–32, 2017, doi: 10.1016/j.ins.2017.02.013.

[20] O. Gencoglu, T. Virtanen, and H. Huttunen, “Oguzhan Gencoglu , Tuomas Virtanen , Heikki Huttunen Department of Signal Processing , Tampere University of Technology , 33720 Tampere , Finland,” *2014 22nd Eur. Signal Process. Conf.*, pp.

506–510, 2014.

[21] P. Khunarsal, C. Lursinsap, and T. Raicharoen, "Very short time environmental sound classification based on spectrogram pattern matching," *Inf. Sci. (Ny).*, vol. 243, pp. 57–74, 2013, doi: 10.1016/j.ins.2013.04.014.

[22] S. Furui, L. Deng, M. Gales, H. Ney, and K. Tokuda, "Fundamental technologies in modern speech recognition," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 16–17, 2012, doi: 10.1109/MSP.2012.2209906.

[23] G. F. Ball and J. Balthazart, "Birdsong learning: Evolutionary, behavioral, and hormonal issues," *Curated Ref. Collect. Neurosci. Biobehav. Psychol.*, pp. 241–246, 2016, doi: 10.1016/B978-0-12-809324-5.01853-8.

[24] M. D. Beecher and E. A. Brenowitz, "Functional aspects of song learning in songbirds," *Trends Ecol. Evol.*, vol. 20, no. 3, pp. 143–149, 2005, doi: 10.1016/j.tree.2005.01.004.

[25] R. Mooney, "Birdsong: The Neurobiology of Avian Vocal Learning," *Encycl. Neurosci.*, pp. 247–251, 2009, doi: 10.1016/B978-008045046-9.01942-2.

[26] H. Barmatz, D. Klein, Y. Vortman, S. Toledo, and Y. Lavner, "A Method for Automatic Segmentation and Parameter Estimation of Bird Vocalizations," *Int. Conf. Syst. Signals, Image Process.*, vol. 2019-June, pp. 211–216, 2019, doi: 10.1109/IWSSIP.2019.8787282.

[27] H. Seth, R. Bhatia, and P. Rajan, "Feature Learning for Bird Call Clustering," *2018 13th Int. Conf. Ind. Inf. Syst. ICIIS 2018 - Proc.*, no. 978, pp. 72–76, 2018, doi: 10.1109/ICIINFS.2018.8721418.

[28] R. H. D. Zottesso, Y. M. G. Costa, D. Bertolini, and L. E. S. Oliveira, "Bird species identification using spectrogram and dissimilarity approach," *Ecol. Inform.*, vol. 48, no. July, pp. 187–197, 2018, doi: 10.1016/j.ecoinf.2018.08.007.

[29] I. Potamitis, S. Ntalampiras, O. Jahn, and K. Riede, "Automatic bird sound detection in long real-field recordings: Applications and tools," *Appl. Acoust.*, vol. 80, pp. 1–9, 2014, doi: 10.1016/j.apacoust.2014.01.001.

[30] E. Sprengel, M. Jaggi, Y. Kilcher, and T. Hofmann, "Audio based bird species identification using deep learning techniques," *CEUR Workshop Proc.*, vol. 1609, pp. 547–559, 2016.

[31] F. Briggs *et al.*, “Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach,” *J. Acoust. Soc. Am.*, vol. 131, no. 6, pp. 4640–4650, 2012, doi: 10.1121/1.4707424.

[32] S. Hershey *et al.*, “CNN architectures for large-scale audio classification,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 131–135, 2017, doi: 10.1109/ICASSP.2017.7952132.

[33] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology,” *Insights Imaging*, vol. 9, no. 4, pp. 611–629, 2018, doi: 10.1007/s13244-018-0639-9.

[34] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, “Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach,” *Procedia Comput. Sci.*, vol. 132, pp. 679–688, 2018, doi: 10.1016/j.procs.2018.05.069.

[35] H. Caracalla and A. Roebel, “Sound texture synthesis using convolutional neural networks,” pp. 1–6, 2019, [Online]. Available: http://arxiv.org/abs/1905.03637.

[36] X. Han, Y. Zhong, L. Cao, and L. Zhang, “Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification,” *Remote Sens.*, vol. 9, no. 8, 2017, doi: 10.3390/rs9080848.

[37] J. Xu, T. Lin, T. Yu, T. Tai, and P. Chang, “Acoustic Scene Classification Using Reduced MobileNet Architecture,” *2018 IEEE Int. Symp. Multimed.*, pp. 267–270, 2018, doi: 10.1109/ISM.2018.00038.

[38] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 2818–2826, 2016, doi: 10.1109/CVPR.2016.308.

[39] A. Gholami *et al.*, “SqueezeNext : Hardware-Aware Neural Network Design,” 2011.

[40] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, 2018, doi: 10.1109/CVPR.2018.00474.

[41] M. Satyanarayanan, “The emergence of edge computing,” *Computer (Long. Beach. Calif).*, vol. 50, no. 1, pp. 30–39, 2017, doi: 10.1109/MC.2017.9.

[42] “xeno-canto :: Sharing bird sounds from around the world.” https://www.xeno-

canto.org/ (accessed Jul. 09, 2020).

[43] “UrbanSound8K - Urban Sound Datasets.” https://urbansounddataset.weebly.com/urbansound8k.html (accessed Jul. 09, 2020).

[44] M. Araya-Salas and G. Smith-Vidaurre, “warbleR: an <scp>r</scp> package to streamline analysis of animal acoustic signals,” *Methods Ecol. Evol.*, vol. 8, no. 2, pp. 184–191, Feb. 2017, doi: 10.1111/2041-210X.12624.

[45] S. Hargreaves, A. Klapuri, and M. Sandler, “Structural segmentation of multitrack audio,” *IEEE Trans. Audio, Speech Lang. Process.*, vol. 20, no. 10, pp. 2637–2647, 2012, doi: 10.1109/TASL.2012.2209419.

[46] S. Chaudhury and I. Pal, “A new approach for image segmentation with MATLAB,” in *2010 International Conference on Computer and Communication Technology, ICCCT-2010*, 2010, pp. 194–197, doi: 10.1109/ICCCT.2010.5640531.

[47] P. Agrawal, S. K. Shriwastava, and S. S. Limaye, “MATLAB implementation of image segmentation algorithms,” in *Proceedings - 2010 3rd IEEE International Conference on Computer Science and Information Technology, ICCSIT 2010*, 2010, vol. 3, pp. 427–431, doi: 10.1109/ICCSIT.2010.5564131.

[48] J. B. Allen, “Applications of the short time Fourier transform to speech processing and spectral analysis,” in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 1982, vol. 1982-May, pp. 1012–1015, doi: 10.1109/ICASSP.1982.1171703.

[49] A. Djebbari and F. B. Reguig, “Short-time fourier transform analysis of the phonocardiogram signal,” in *Proceedings of the IEEE International Conference on Electronics, Circuits, and Systems*, 2000, vol. 2, pp. 844–847, doi: 10.1109/ICECS.2000.913008.

[50] B. McFee *et al.*, “Python LibROSA Module.” 2019, doi: 10.5281/zenodo.591533.

[51] “Python For Beginners | Python.org.” https://www.python.org/about/gettingstarted/ (accessed Jul. 09, 2020).

[52] H. Y. Chen and C. Y. Su, “An Enhanced Hybrid MobileNet,” in *2018 9th International Conference on Awareness Science and Technology, iCAST 2018*, Oct. 2018, pp. 308–312, doi: 10.1109/ICAwST.2018.8517177.

[53] “Keras Applications.” https://keras.io/api/applications/ (accessed Jul. 09, 2020).

[54] “Transfer Learning using Mobilenet and Keras | by Ferhat Culfaz | Towards Data Science.” https://towardsdatascience.com/transfer-learning-using-mobilenet-and-keras-c75daf7ff299 (accessed Jul. 09, 2020).

[55] F. Berger, W. Freillinger, P. Primus, W. Reisinger, and J. Kepler, “Bird Audio Detection-DCASE 2018 Special Topics-Machine Learning and Audio: a challenge.” Accessed: Jul. 09, 2020. [Online]. Available: http://machine-listening.eecs.qmul.ac.uk/bird-audio-detection-.

[56] T. Grill and J. Schluter, “Two convolutional neural networks for bird detection in audio signals,” *25th Eur. Signal Process. Conf. EUSIPCO 2017*, vol. 2017-Janua, pp. 1764–1768, 2017, doi: 10.23919/EUSIPCO.2017.8081512.

[57] S. Han *et al.*, “Optimizing Filter Size in Convolutional Neural Networks for Facial Action Unit Recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 5070–5078, 2018, doi: 10.1109/CVPR.2018.00532.

[58] N. D. Marom, L. Rokach, and A. Shmilovici, “Using the confusion matrix for improving ensemble classifiers,” in *2010 IEEE 26th Convention of Electrical and Electronics Engineers in Israel, IEEEI 2010*, 2010, pp. 555–559, doi: 10.1109/EEEI.2010.5662159.

[59] D. B. Walther, “Using confusion matrices to estimate mutual information between two categorical measurements,” in *Proceedings - 2013 3rd International Workshop on Pattern Recognition in Neuroimaging, PRNI 2013*, 2013, pp. 220–224, doi: 10.1109/PRNI.2013.63.