gibbonNetR: an R Package for the Use of Convolutional Neural Networks and Transfer Learning on Acoustic Data

A Preprint

Dena Jane Clink
K. Lisa Yang Center for Conservation Bioacoustics
Cornell Lab of Ornithology, Cornell University, Ithaca, New York, United States

dena.clink@cornell.edu

Abdul Hamid Ahmad
Institute for Tropical Biology and Conservation
Universiti Malaysia Sabah (UMS), Kota Kinabalu, Sabah, Malaysia

August 31, 2023

Abstract

Automated detection of acoustic signals is crucial for effective monitoring of vocal animals and their habitats across large spatial and temporal scales. Recent advances in deep learning have made high performing automated detection approaches more accessible two more practitioners. However, there are few deep learning approaches that can be implemented natively in R. The 'torch for R' ecosystem has made the use of transfer learning with convolutional neural networks accessible for R users. Here we provide an R package and workflow to use transfer learning for the automated detection of acoustics signals from passive acoustic monitoring (PAM) data collected in Sabah, Malaysia. The package provides functions to create spectogram images from PAM data, compare the performance of different pre-trained CNN architectures, and deploy trained models over directories of sound files.

***K*eywords** deep learning · passive acoustic monitoring · gibbon · automated detection

1 Introduction

1.1 *Passive acoustic monitoring*

We are in a biodiversity crisis, and there is a great need for the ability to rapidly assess biodiversity in order to understand and mitigate anthropogenic impacts. One approach that can be especially effective for monitoring of vocal yet cryptic animals is the use of passive acoustic monitoring (Gibb et al. 2018), a technique that relies autonomous acoustic recording units. PAM allows researchers to monitor vocal animals and their habitats, at temporal and spatial scales that are impossible to achieve using only human observers. Interest in use of PAM in terrestrial environments has increased substantially in recent years (Sugai et al. 2019), due to reduced price of the recording units and improved battery life and data storage capabilities. However, the use of PAM often leads to the collection of terabytes of data that is time- and cost-prohibitive to analyze manually.

1.2 *Automated detection*

Some commonly used non-deep learning approaches for the automated detection of acoustic signals in terrestrial PAM data include binary point matching (Katz, Hafner, and Donovan 2016), spectrogram cross-correlation (Balantic and Donovan 2020), or the use of a band- limited energy detector and subsequent classifier, such as support vector machine (Clink et al. 2023; Kalan et al. 2015). Recent advances in deep learning have revolutionized image and speech recognition (LeCun, Bengio, and Hinton 2015), with important cross-over for the analysis of PAM data. Traditional approaches to machine learning relied heavily on feature engineering, as early machine learning algorithms required a reduced set of representative features, such as features estimated from the spectrogram. Deep learning does not require feature engineering (Stevens, Antiga, and Viehmann 2020) . Convolutional neural networks (CNNs) — one of the most effective deep learning algorithms—are useful for processing data that have a 'grid-like topology', such as image data that can be considered a 2-dimensional grid of pixels (Goodfellow, Bengio, and Courville 2016). The 'convolutional' layer learns the feature representations of the inputs; these convolutional layers consist of a set of filters which are basically two-dimensional matrices of numbers and the primary parameter is the number of filters (Gu et al. 2018). Therefore, with CNN's there is no feature engineering required. However, if training data are scarce, overfitting may occur as representations of images tend to be large with many variables (LeCun, Bengio, and others 1995).

Transfer learning is an approach wherein the architecture of a pretrained CNN (which is generally trained on a large dataset) is applied to a new classification problem. For example, CNNs trained on the ImageNet dataset of >1 million images (Deng et al. 2009)such as ResNet have been applied to automated detection/classification of primate and bird species from PAM data (Dufourq et al. 2022; Ruan et al. 2022). At the most basic level, transfer learning in computer vision applications retains the feature extraction or embedding layers, and modifies the last few classification layers to be trained for a new classification task (Dufourq et al. 2022). Transfer learning has been shown to outperform CNNs trained with random initial weights (Tan et al. 2018). Transfer learning is particularly appropriate when there is a paucity of training data (Weiss, Khoshgoftaar, and Wang 2016), such as common in PAM data.

1.3 *'torch for R' ecosystem*

The two most popular open-source programming languages are R and Python (Scavetta and Angelov 2021). Python has surpassed R in terms of overall popularity, but R remains an important language for the life sciences (Lawlor et al. 2022). 'Keras' (Chollet and others 2015), 'PyTorch' (Paszke et al. 2019) and 'Tensorflow' (Martín Abadi et al. 2015) are some of the more popular neural network libraries; these libraries were all initially developed for the Python programming language. Until recently, deep learning implementations in R relied on the 'reticulate' package which served as an interface to Python (Ushey, Allaire, and Tang 2022). However, the recent release of the 'torch for R' ecosystem provides a framework based on 'PyTorch' that runs natively in R and has no dependency on Python (Falbel 2023). Running natively in R means more straightforward installation, and higher accessibility for users of the R programming environment. Keydana (2023) provides tutorials for transfer learning in the 'torch for R' ecosystem, and the functions in 'gibbonNetR' rely heavily on these tutorials.

2 Overview

This package provides functions

3 Usage

3.1 First we create spectrogram images

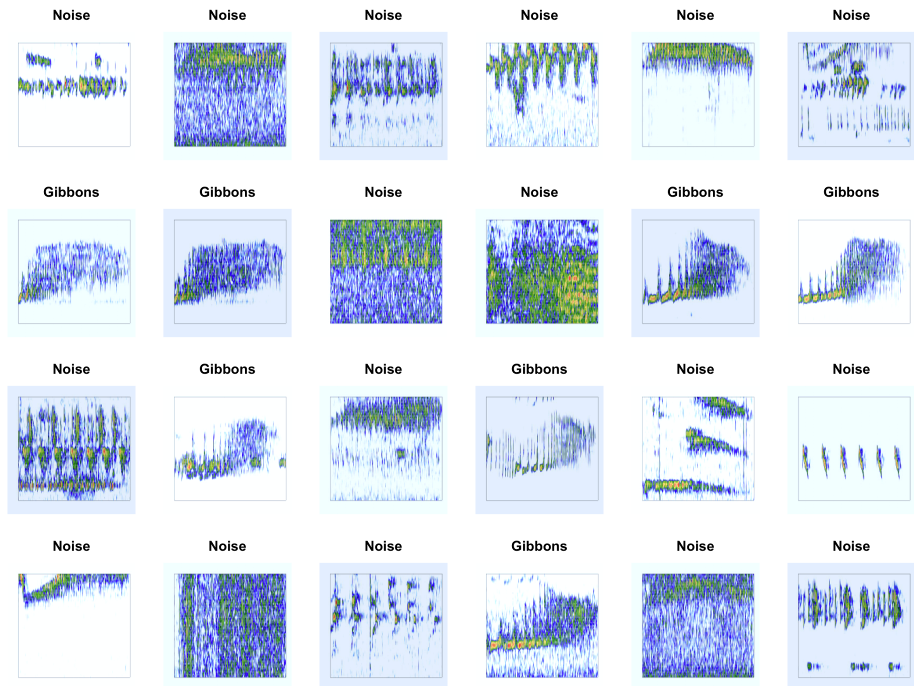

Figure 1: Spectrograms of training clips for CNNs

3.2 Then we train the model

```
gibbonNetR::train_alexNet(input.data.path=input.data.path,
                          test.data=test.data.path,
                          unfreeze = unfreeze.param,
                          epoch.iterations=epoch.iterations,
                          early.stop = "yes",
                          output.base.path = "data/",
                          trainingfolder=trainingfolder.short,
                          positive.class="Gibbons",
                          negative.class="Noise")
```

3.3 We can compare the performance of different CNN architectures

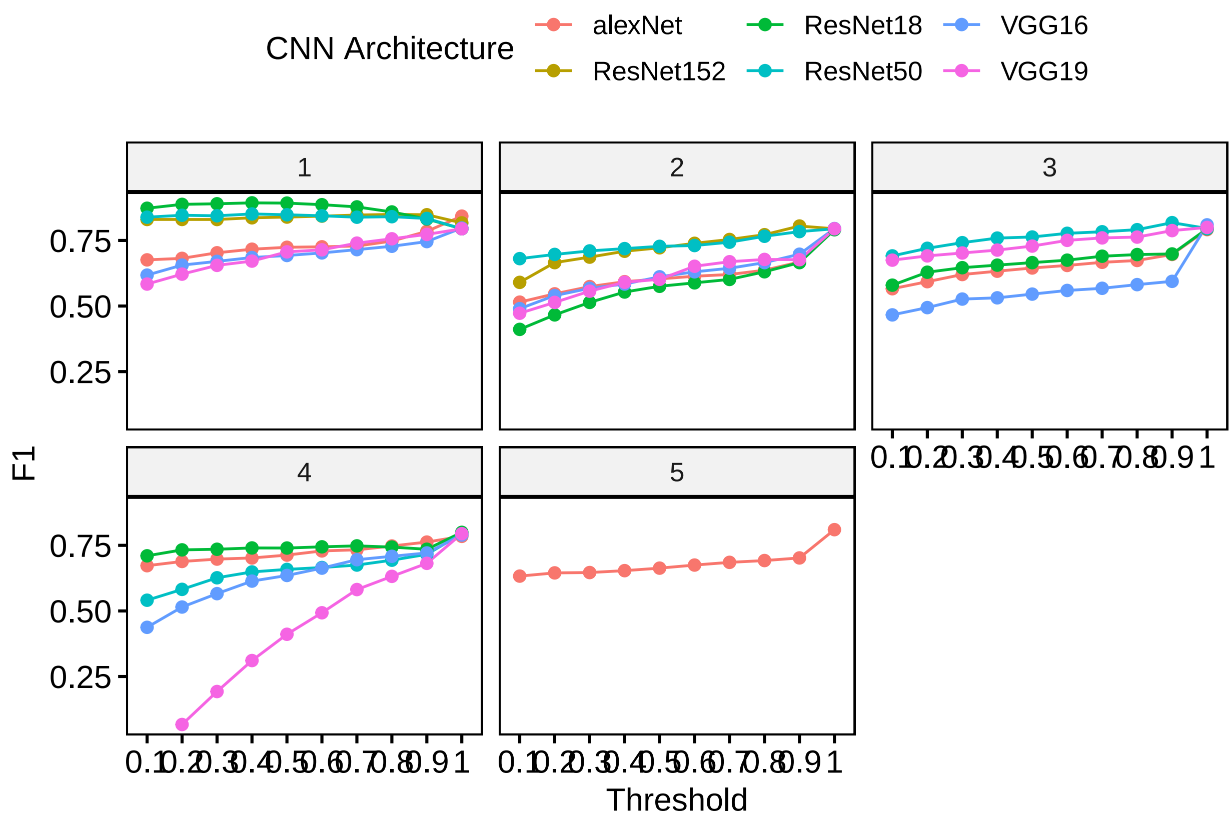

Figure 2: Evaluating performance of pretrained CNNs

4 Discussion

5 Acknowledgments

So long and thanks for all the fish.

References

Balantic, Cathleen, and Therese Donovan. 2020. "AMMonitor: Remote Monitoring of Biodiversity in an Adaptive Framework with r." *Methods in Ecology and Evolution* 11 (7): 869877.

Chollet, François, and others. 2015. "Keras." `https://keras.io`.

Clink, Dena J., Isabel Kier, Abdul Hamid Ahmad, and Holger Klinck. 2023. "A Workflow for the Automated Detection and Classification of Female Gibbon Calls from Long-Term Acoustic Recordings." *Frontiers in Ecology and Evolution* 11. `https://www.frontiersin.org/articles/10.3389/fevo.2023.1071640`.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. "Imagenet: A Large-Scale Hierarchical Image Database." In, 248255. Ieee.

Dufourq, Emmanuel, Carly Batist, Ruben Foquet, and Ian Durbach. 2022. "Passive Acoustic Monitoring of Animal Populations with Transfer Learning." *Ecological Informatics* 70: 101688. https://doi.org/`https://doi.org/10.1016/j.ecoinf.2022.101688`.

Falbel, Daniel. 2023. *Luz: Higher Level 'API' for 'Torch'*. `https://CRAN.R-project.org/package=luz`.

Gibb, Rory, Ella Browning, Paul Glover-Kapfer, and Kate E. Jones. 2018. "Emerging Opportunities and Challenges for Passive Acoustics in Ecological Assessment and Monitoring." *Methods in Ecology and Evolution*, October. https://doi.org/10.1111/2041-210X.13101.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press.

Gu, Jiuxiang, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, et al. 2018. "Recent Advances in Convolutional Neural Networks." *Pattern Recognition* 77: 354377.

Kalan, Ammie K., Roger Mundry, Oliver J J Wagner, Stefanie Heinicke, Christophe Boesch, and Hjalmar S. Kühl. 2015. "Towards the Automated Detection and Occupancy Estimation of Primates Using Passive Acoustic Monitoring." *Ecological Indicators* 54 (July 2015): 217226. https://doi.org/10.1016/j.ecolind.2015.02.023.

Katz, Jonathan, Sasha D Hafner, and Therese Donovan. 2016. "Assessment of Error Rates in Acoustic Monitoring with the r Package monitoR." *Bioacoustics* 25 (2): 177196.

Keydana, Sigrid. 2023. *Deep Learning and Scientific Computing with r Torch*. CRC Press.

Lawlor, Jake, Francis Banville, Norma-Rocio Forero-Muñoz, Katherine Hébert, Juan Andrés Martínez-Lanfranco, Pierre Rogy, and A. Andrew M. MacDonald. 2022. "Ten Simple Rules for Teaching Yourself R." *PLOS Computational Biology* 18 (9): e1010372. https://doi.org/10.1371/journal.pcbi.1010372.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–44. https://doi.org/10.1038/nature14539.

LeCun, Yann, Yoshua Bengio, and others. 1995. "Convolutional Networks for Images, Speech, and Time Series." *The Handbook of Brain Theory and Neural Networks* 3361 (10): 1995.

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, et al. 2015. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems." https://www.tensorflow.org/.

Paszke, Adam, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, et al. 2019. "PyTorch: An Imperative Style, High-Performance Deep Learning Library." In, 80248035. Curran Associates, Inc. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

Ruan, Wenda, Keyi Wu, Qingchun Chen, and Chengyun Zhang. 2022. "ResNet-Based Bio-Acoustics Presence Detection Technology of Hainan Gibbon Calls." *Applied Acoustics* 198: 108939. https://doi.org/https://doi.org/10.1016/j.apacoust.2022.108939.

Scavetta, Rick J, and Boyan Angelov. 2021. *Python and r for the Modern Data Scientist*. O'Reilly Media, Inc.

Stevens, Eli, Luca Antiga, and Thomas Viehmann. 2020. *Deep Learning with PyTorch*. Simon; Schuster.

Sugai, Larissa Sayuri Moreira, Thiago Sanna Freire Silva, José Wagner Ribeiro, and Diego Llusia. 2019. "Terrestrial Passive Acoustic Monitoring: Review and Perspectives." *BioScience* 69 (1): 1525. https://doi.org/10.1093/biosci/biy147.

Tan, Chuanqi, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. 2018. "A Survey on Deep Transfer Learning." In, 270279. Springer.

Ushey, Kevin, J. J. Allaire, and Yuan Tang. 2022. *Reticulate: Interface to 'Python'*.

Weiss, Karl, Taghi M Khoshgoftaar, and DingDing Wang. 2016. "A Survey of Transfer Learning." *Journal of Big Data* 3 (1): 140.