AIMS African Institute for Mathematical Sciences SOUTH AFRICA

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

BADEN-WÜRTTEMBERG STIPENDIUM®

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

EVOLUTIONARY SPECTROGRAM OPTIMIZATION

for Passive Accoustic Monitoring

Ufuk Çakır, [1]
Lorène Jeantet, [2,4]
Joël Lontsi, [2,4]
Emmanuel Dufourq [2,3,4]

GitHub

1 Interdisciplinary Center for Scientific Computing, Heidelberg University, Germany
2 African Institute for Mathematical Sciences, South Africa
3 African Institute for Mathematical Sciences, Research and Innovation Centre, Rwanda
4 Department of Mathematical Sciences, Stellenbosch University, South Africa

MOTIVATION

Passive Acoustic Monitoring (PAM) provides an important tool for wildlife monitoring by using acoustic recorders deployed in natural environments. PAM field surveys are conducted over several months, and the resulting datasets contain thousands of hours of audio, corresponding to Terabytes of data that needs to be analysed. Deep learning classifiers are typically used to automate the processing of these large datasets and are trained to detect animals vocalisations. However, these classifiers are often computationally expensive and cannot be easily deployed on low resource devices. Novel approaches are required to facilate real-time monitoring on low-resource deivces. Hardware and software considerations need to be factored together to facilitate real-time monitoring systems. Thus, there is a need for an approach that can reduce the inference time and model complexity, while maintaining high classifier performance.

ESO ALGORITHM

We propose ESO, a genetic algorithm [1] based approach to optimise the input spectrograms to the deep neural network. ESO is tasked to select rectangular bands from the spectrogram. Typically, a full spectrogram would be used for classification, however this is often not required as the species of interest usually only vocalises within specific frequency ranges. In Figure 1 we compare a classical approach to the creation of a CNN for animal vocalisations using full spectrograms (top), and our proposed ESO approach (bottom). We applied ESO on an acoustic dataset containing vocalisations of the Hainan gibbon, the world's rarest primate.

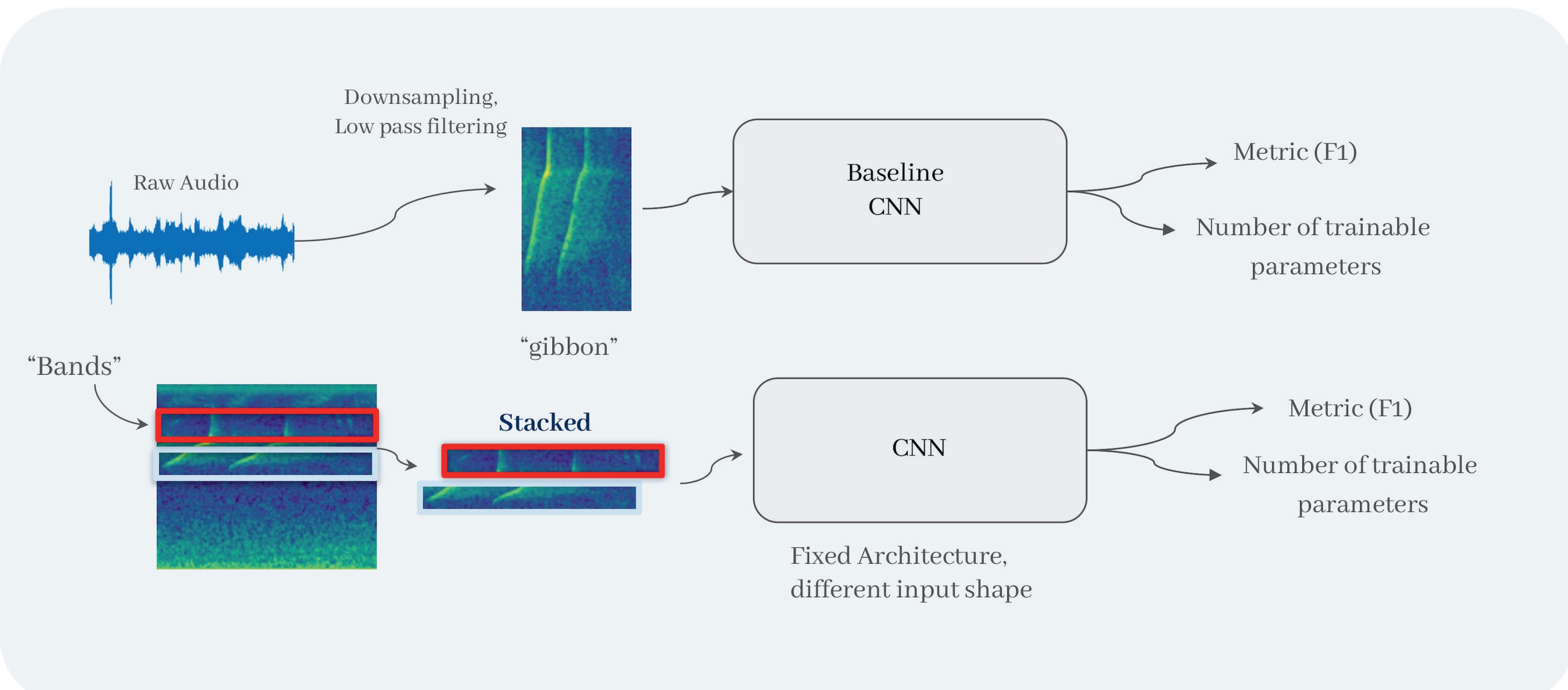

Figure 1: Comparing a typical approach to our proposed ESOalgorithm.

In ESO, a gene extracts one band from the spectrogram. A chromosome contains various genes, and thus can extract multiple bands. The bands are then stacks to form a new compressed image representation, and input to the CNN (Figure 2). The original spectrogram had a shape of 75 × 128 (9,600 pixels), and the ESO chromosome produced an image of size 2 × 75 × 5 (750 pixels).

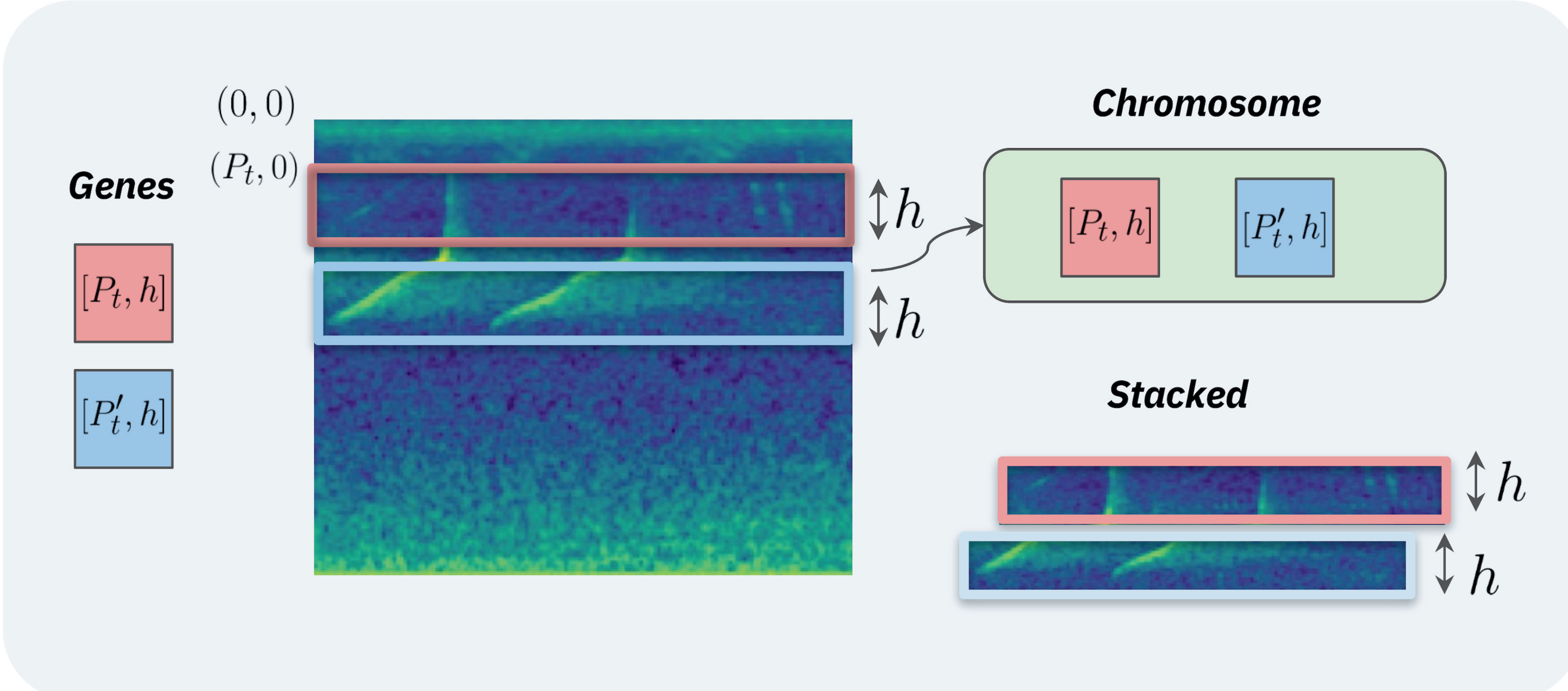

Figure 2: ESO Gene and Chromosome : Each gene defines one band in the spectogram.

ESO optimises for the best chromosomes by maximizing its fitness defined as

$$\text{Fitness} = -\lambda_1 \frac{\text{F1}_{\text{base}} - \text{F1}_{\text{chromosome}}}{\text{F1}_{\text{base}}} + \lambda_2 \frac{p_{\text{base}} - p_{\text{chromosome}}}{p_{\text{base}}}$$

where F1 is the F1-Score and p is the number of trainable parameters of the CNN model. This fitness function is defined such that a high fitness means a better tradeoff between model complexity and model performance.

RESULTS

The table below shows the average performance metric of Baseline and ESO models over 10 runs with a population size of 20. (ESO took 1hr to execute on GeForce GTX 1050 Ti, 4 GB). A comparison between the metrics obtained for baseline CNN and the ESO model shows a 91 % reduction in the number of trainable parameters and a 71% decrease in inference time.

Metric	Baseline	ESO	Difference (%)
Accuracy	0.96	$0.94_{\pm 0.00}$	$-2.01_{\pm 0.00}$
Sensitivity	0.95	$0.92_{\pm 0.01}$	$-3.16_{\pm 1.05}$
Specificity	0.97	$0.95_{\pm 0.01}$	$-2.06_{\pm 1.03}$
Precision	0.92	$0.88_{\pm 0.02}$	$-4.35_{\pm 2.17}$
F1-Score	0.94	$0.90_{\pm 0.01}$	$\mathbf{-4.26}_{\pm 1.06}$
Parameters	132, 234	$11,863_{\pm 247}$	$\mathbf{-91.03}_{\pm 0.19}$
Image Size	9, 728	$4,817_{\pm 800}$	$\mathbf{-50.48}_{\pm 8.22}$
Inference time [s]	376.1	106.9	$\mathbf{-71.58}$

CONCLUSION

ESO can significantly reduce the spectrogram input to CNN classifiers by identifying suitable compress bands. Inference time is reduced, thus saving costs to process large acoustic datasets. ESO represents a novel approach to facilitating real-time monitoring systems.

REFERENCES

[1] Back, T. (1996). Evolutionary algorithms in theory and practice: Evolution strategies, evolutionary programming, genetic algorithms. Oxford University Press.

UC was supported by the HeiAIMS partnership with support from Baden-Württemberg-STIPENDIUM and Baden-Württemberg Stiftung. **ED** was funded by a grant from the Carnegie Corporation of New York. **LJ** was funded by a grant from the International Development Research Centre, Ottawa, Canada, www.idrc.ca, and with financial support from the Government of Canada, provided through Global Affairs Canada (GAC), www.international.gc.ca. The statements made and views expressed are solely the responsibility of the author(s). **JL** was funded by a grant from the Department of Science and Innovation.