E-PANNs: An Efficient version of PANNs for Audio Tagging

Arshdeep Singh[1], Haohe Liu, Mark D. Plumbley
Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, UK.
[1]Research Fellow A, Email: arshdeep.singh@surrey.ac.uk.

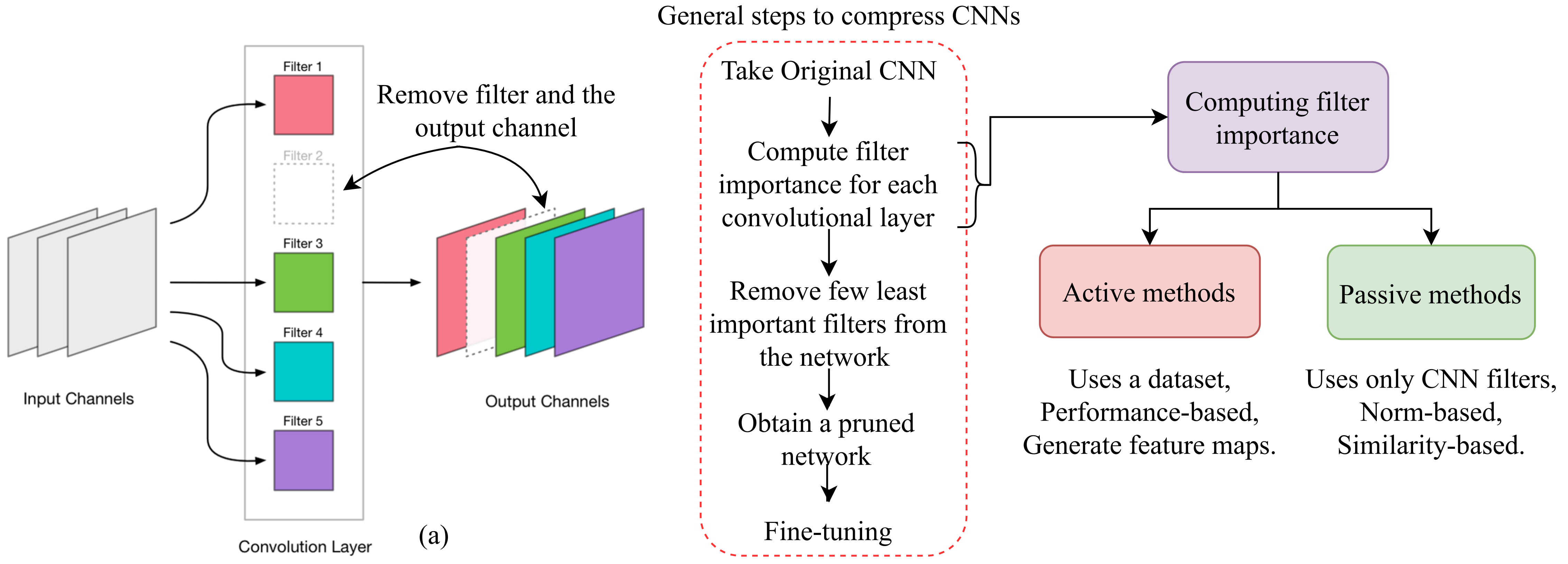
Speech and Audio in the Northeast (SANE) Workshop 2023, New York University (NYU), Brooklyn, New York.

Introduction 1

- Pre-trained audio neural networks (PANNs) [1] are a family of convolutional neural networks (CNNs) designed for audio classification tasks.
- PANNs have shown state-of-the-art performance in audio tagging and have been widely employed for many downstream tasks including audio scene classification and several DCASE [2] related tasks as a feature extractor or end-to-end classifier.
- However, the best performing PANNs require 21G computations* (multiply-accumulate operations, MACs) for inference and requires 81M parameters for storage.
- Large size and computations demand more resources and slow down the inference speed.
- This work contribute to reduce computations and memory storage requirement of PANNs while maintaining the performance as given by the original PANNs.

Filter pruning to obtain E-PANNs 3

- **Filter Pruning [3]** involves "removing" some of the filters and their connected feature maps/channels from convolutional neural networks (CNNs) to compress them (See Figure 2 (a)).
- **Benefits of Pruning:** Reduces computational cost (multiply-accumulate operations, MACs) and memory storage of CNNs while providing similar performance, hence CNNs become more efficient.



Remove filter and the output channel

General steps to compress CNNs

Take Original CNN

Compute filter importance for each convolutional layer

Remove few least important filters from the network

Obtain a pruned network

Fine-tuning

Computing filter importance

Active methods

Uses a dataset, Performance-based, Generate feature maps.

Passive methods

Uses only CNN filters, Norm-based, Similarity-based.

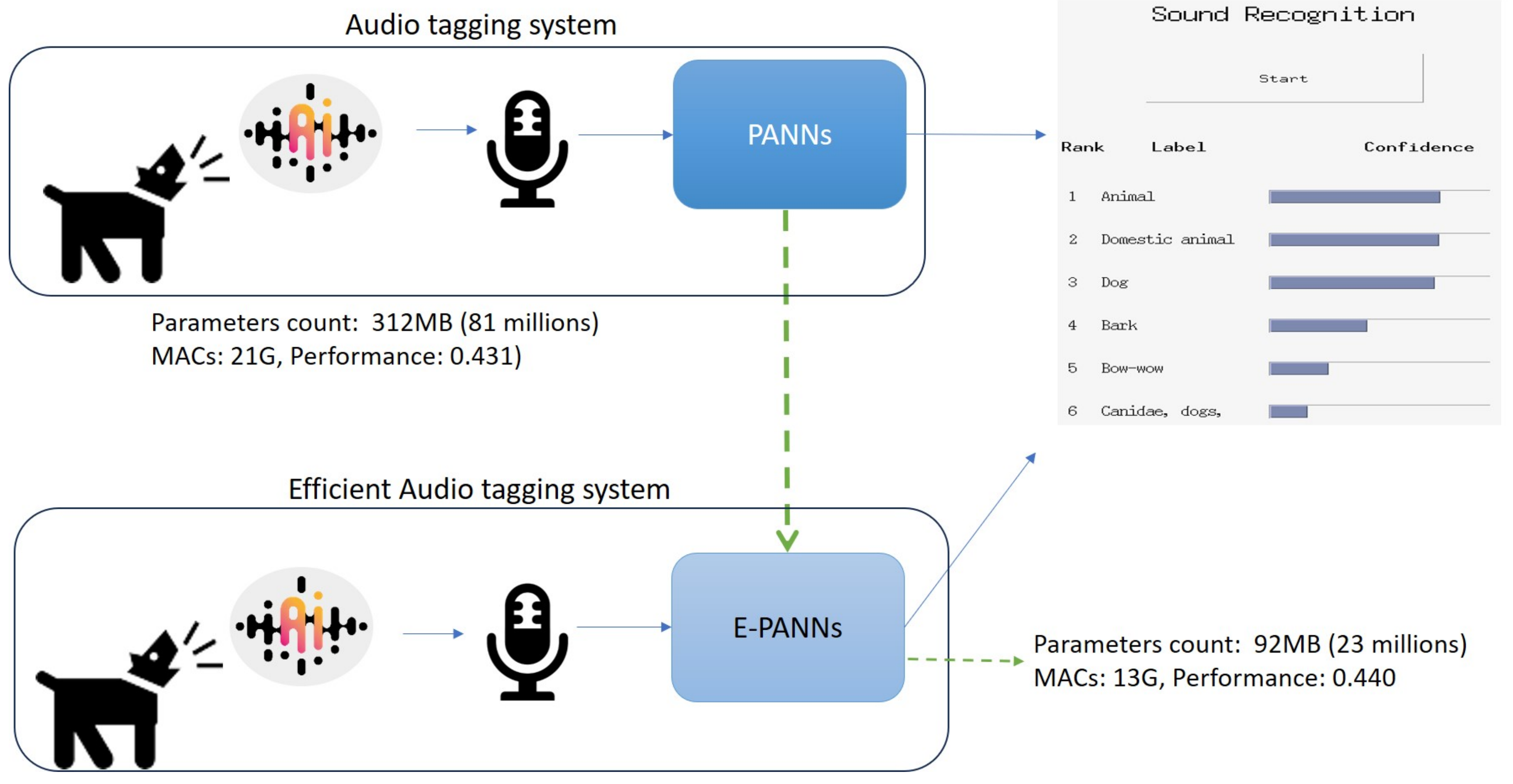
Figure 2

We opt a passive filter pruning approach to compute filter importance.A filter producing significant output, as measured using operator norm of the filter [4,5], is considered more important than others.

References

[1] Qiuqiang Kong et al., "PANNs: Large-scale pretrained audio neural networks for audio pattern recognition", IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28:2880–2894, 2020.
[2] Detection and Classification of Acoustic Scenes and Events (DCASE) challenge (https://dcase.community/)
[3] Luo, Jian-Hao, et al. "ThiNet: pruning CNN filters for a thinner net", IEEE transactions on pattern analysis and machine intelligence 41.10 (2018): 2525-2538.
[4] Arshdeep Singh , H Liu, and Mark D. Plumbley (2023) "E-PANNs: Sound Recognition Using Efficient Pre-trained Audio Neural Networks", 52nd International Congress and Exposition on Noise Control Engineering (Inter-Noise 2023), Chiba, Greater Tokyo, Japan, 20-23 August 2023.
[5] Singh, A, and Mark D. Plumbley. "Efficient CNNs via Passive Filter Pruning", arXiv preprint arXiv:2304.02319 (2023).
[6] Our Live Sound recognition demo available at: https://github.com/Arshdeep-Singh-Boparai/E-PANNs

Our Contribution 2



Parameters count: 312MB (81 millions)
MACs: 21G, Performance: 0.431)

Parameters count: 92MB (23 millions)
MACs: 13G, Performance: 0.440

Figure 1

Performance Analysis & Summary 4

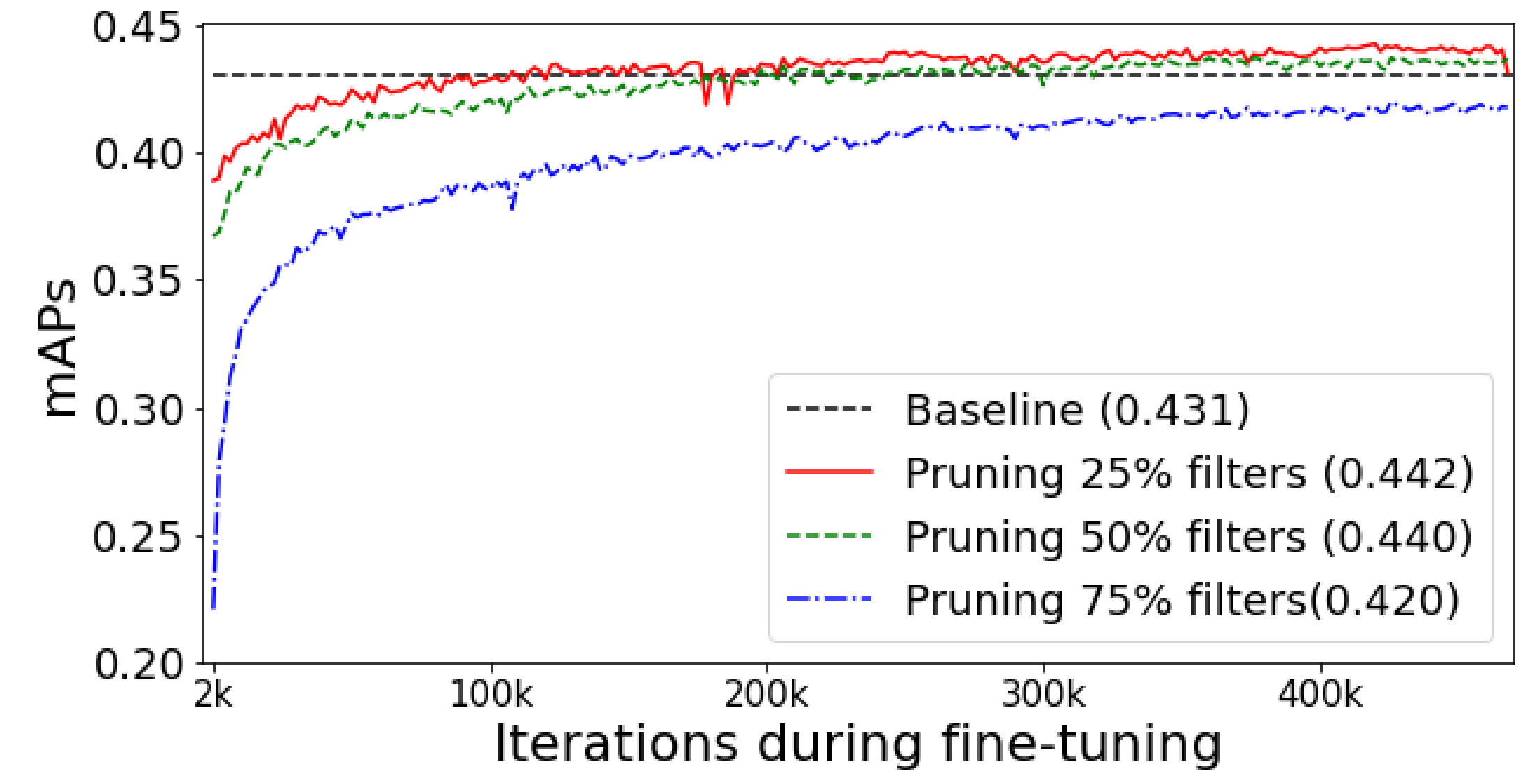


Figure 3

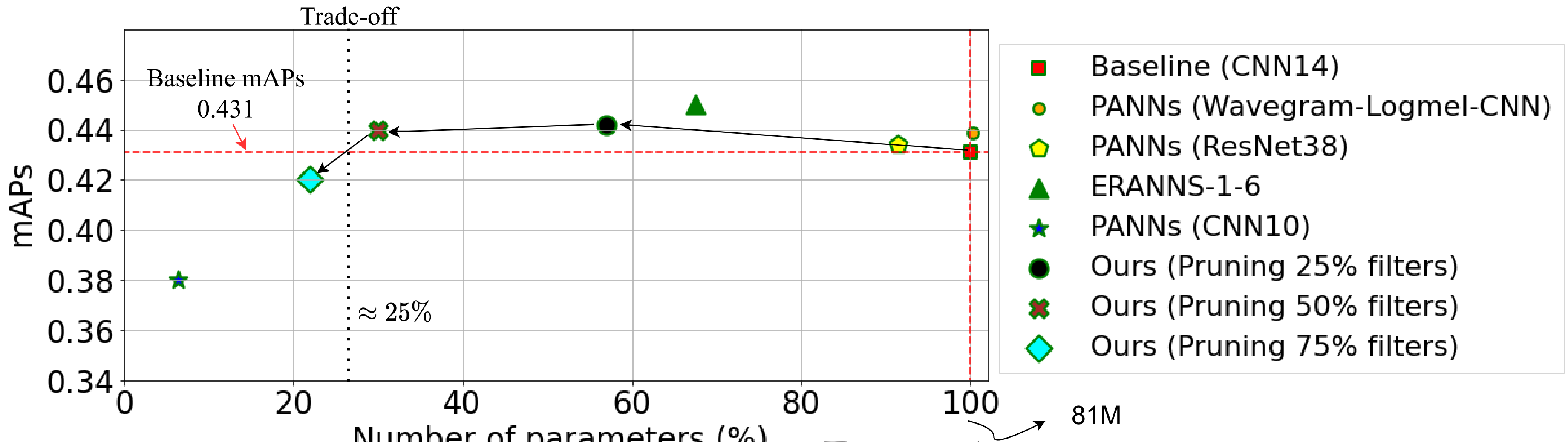


Figure 4

Summary

- A pre-trained network is made efficient by removing unimportant filters followed by a fine-tuning process (with 20-50% efforts as used in training).
- Reduced 30% computations, 70% memory storage with an improved performance for audio tagging (sound recognition) task.
- Live sound recognition demo is available at [6].

Acknowledgments

This work was partly supported by Engineering and Physical Sciences Research Council (EPSRC) Grant EP/T019751/1 "AI for Sound (AI4S)". For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.
Personal profile: https://www.surrey.ac.uk/people/arshdeep-singh