Efficient Audio-based Convolutional Neural Networks via Filter Pruning

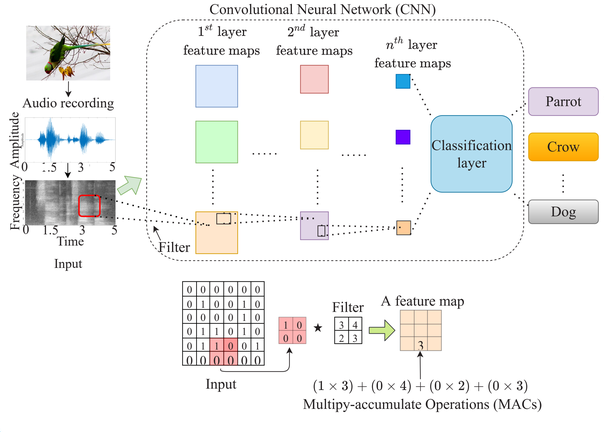
Arshdeep Singh[1] & Mark D. Plumbley

Center for Vision, Speech and Signal Processing (CVSSP),

Faculty of Engineering and Physical Sciences, University of Surrey, UK.

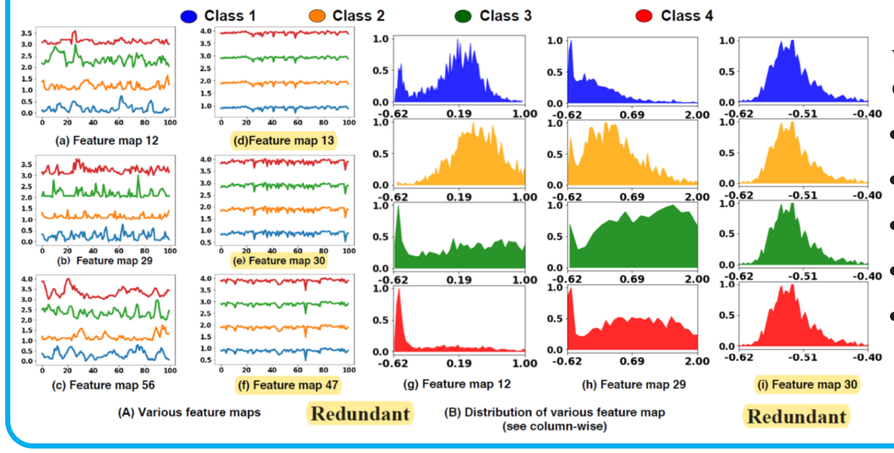
[1]Research Fellow A, Email: arshdeep.singh@surrey.ac.uk, ORCID: https://orcid.org/0000-0003-3465-0952

Introduction 1



- Convolutional neural networks (CNNs) have the capability to learn from examples or experience like humans.
- CNNs have shown state-of-the-art performance in audio classification [1], image scene classification [2] etc.
- However, CNNs are resource hungry due to their large size and heavy computations.
- This makes a bottleneck to deploy CNNs on resource-constrained devices such as smart phones or IoTs.
- Moreover, CNNs may have redundancy in their parameters or feature maps [3] (See Slide 2).
- Training CNNs for more time generates more CO_2, e.g. running GTX 1080 Ti hardware for 2 days generates CO_2 = 5.18 kg ≡ Driving an average car for 20Km.

Illustration of redundancy in CNNs 2

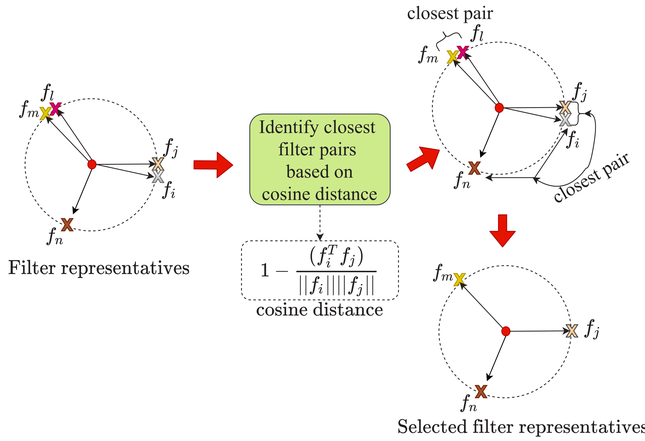


What do we achieve after eliminating redundancy???

- Less parameters.
- Low memory size.
- Low computations.
- Speed-up in inference.
- Reduced training time and less CO_2 emission.

Obtaining an Efficient CNN by eliminating Filters 3

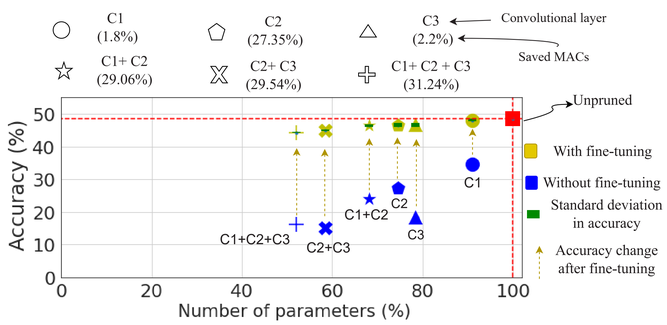
Hypothesis: Similar filters produce similar output and hence, mostly contribute to redundancy and can be eliminated.

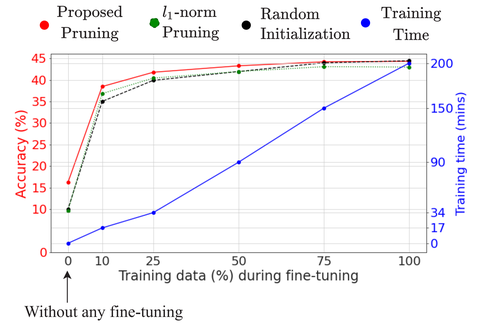


Steps: Unpruned CNN → Eliminate one of the similar filters → Pruned network → Fine-tuning.

Performance Analysis 4

Experiments are performed on DCASE 2021 Task1A baseline network for audio scene classification (ASC) [4].





Open Research Practices 5

[a] Openly available tool is used to estimate CO_2 emission. Link: https://mlco2.github.io/impact/##compute

[b] Experiments are performed on openly available ASC dataset and CNN [4].

[c] Proposed code is openly available at Gitlab. Link:https://gitlab.surrey.ac.uk/as0150/passive-pruning

Conclusions 6

- Reduced 25% parameters and 27% MACs, with less than 1% drop in accuracy.
- Fine-tuning using few training examples improves the performance of the pruned network significantly.
- Designing better similarity measure and reducing complexity in fine-tuning is a future goal.

References

[1] Q. Kong et al., "PANNs: Large-scale pretrained audio neural networks for audio pattern recognition," IEEE/ACM TALSP, vol. 28, pp. 2880–2894, 2020.

[2] Q. Wang et al., "Looking closer at the scene: Multiscale representation learning for remote sensing image scene classification ," IEEE Transactions on NNLS (in press), pp. 1–15, 2020.

[3] K. Kahatapitiya and R. Rodrigo, "Exploiting the redundancy in convolutional filters for parameter reduction," proceedings of the IEEE/CVF WACV, pp. 1410–1420, 2021.

[4] Martín-Morató, Irene, et al., "Low- complexity acoustic scene classification for multi-device audio: Analysis of DCASE 2021 challenge systems," DCAEE workshop 85–89, 2021.

Acknowledgement

This work was supported by Engineering and Physical Sciences Research Council (EPSRC) Grant EP/T019751/1 AI for Sound. Thanks to other AI4S team members: Prof. Mark D. Plumbley, Dr. Helen Cooper, Dr Emily Corrigan-Kavanagh and Andres Fernandez for their suggestions and support. AI4S project link: https://ai4s.surrey.ac.uk/index.