

Low-Complexity CNNs for Acoustic Scene Classification

DCASE 2022 Workshop Spotlight Presentation
Session II: P1

Arshdeep Singh, Mark D. Plumbley

CVSSP | Centre for Vision,
Speech and Signal
Processing



EPSRC

Engineering and Physical Sciences
Research Council



Introduction

- Convolutional neural networks (CNNs) have been widely employed in various applications [1-2].
- However, CNNs consume more power + have high latency due to their large size and heavy computations, (MACs: multiply-accumulate operations).
- **CNNs → resource-hungry and not environment friendly (emit more CO2 due to heavy computations)**
- This makes a bottleneck to deploy CNNs on resource-constrained devices (Edge computing).
- For example, smart phones/watches/headphones that may use context-aware services like audio scene classification (ASC).
- **Efficient CNNs: To utilize the underlying resources effectively.**
(Compress CNNs: As CNNs may have redundant parameters (filters/weights) [3] + Lottery-ticket hypothesis [4].

[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.

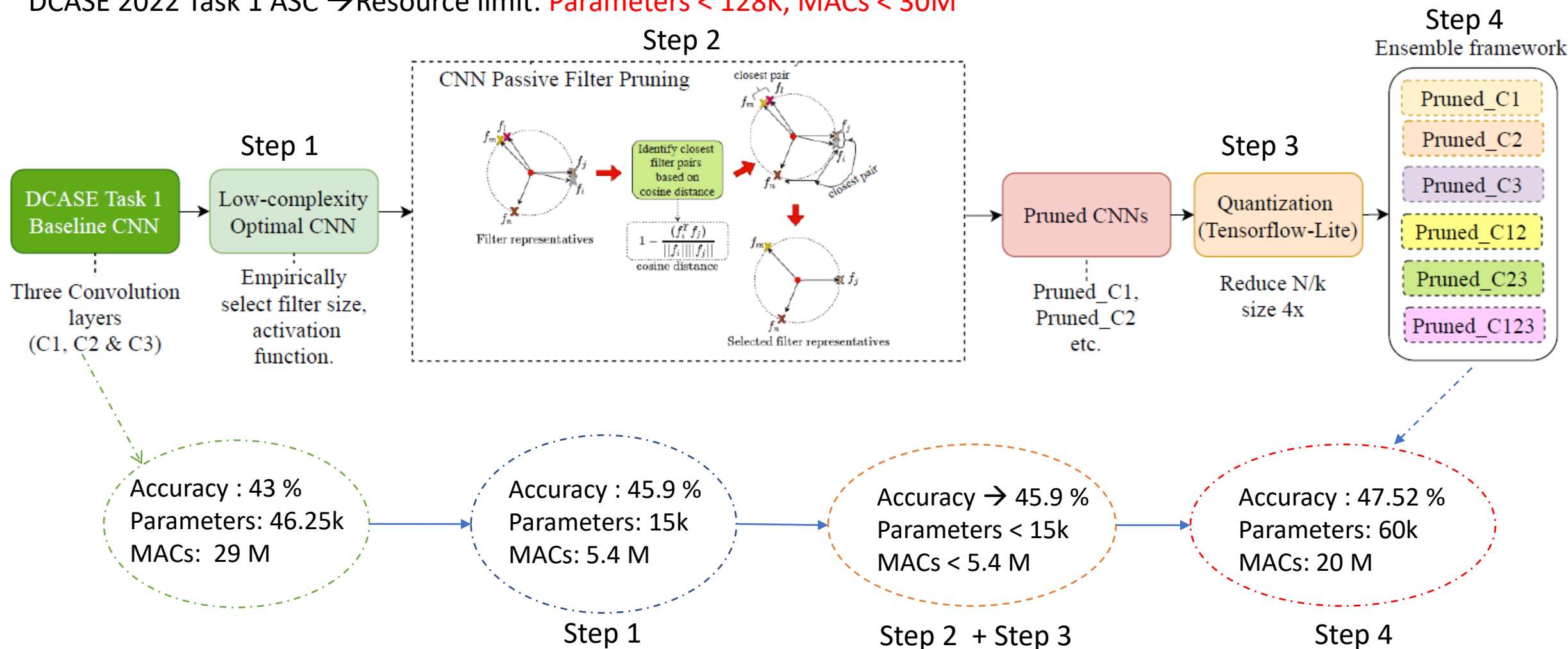
[2] Q. Wang et al., "Looking closer at the scene: Multiscale representation learning for remote sensing image scene classification," IEEE Transactions on NNLS (in press), pp. 1–15, 2020.

[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.

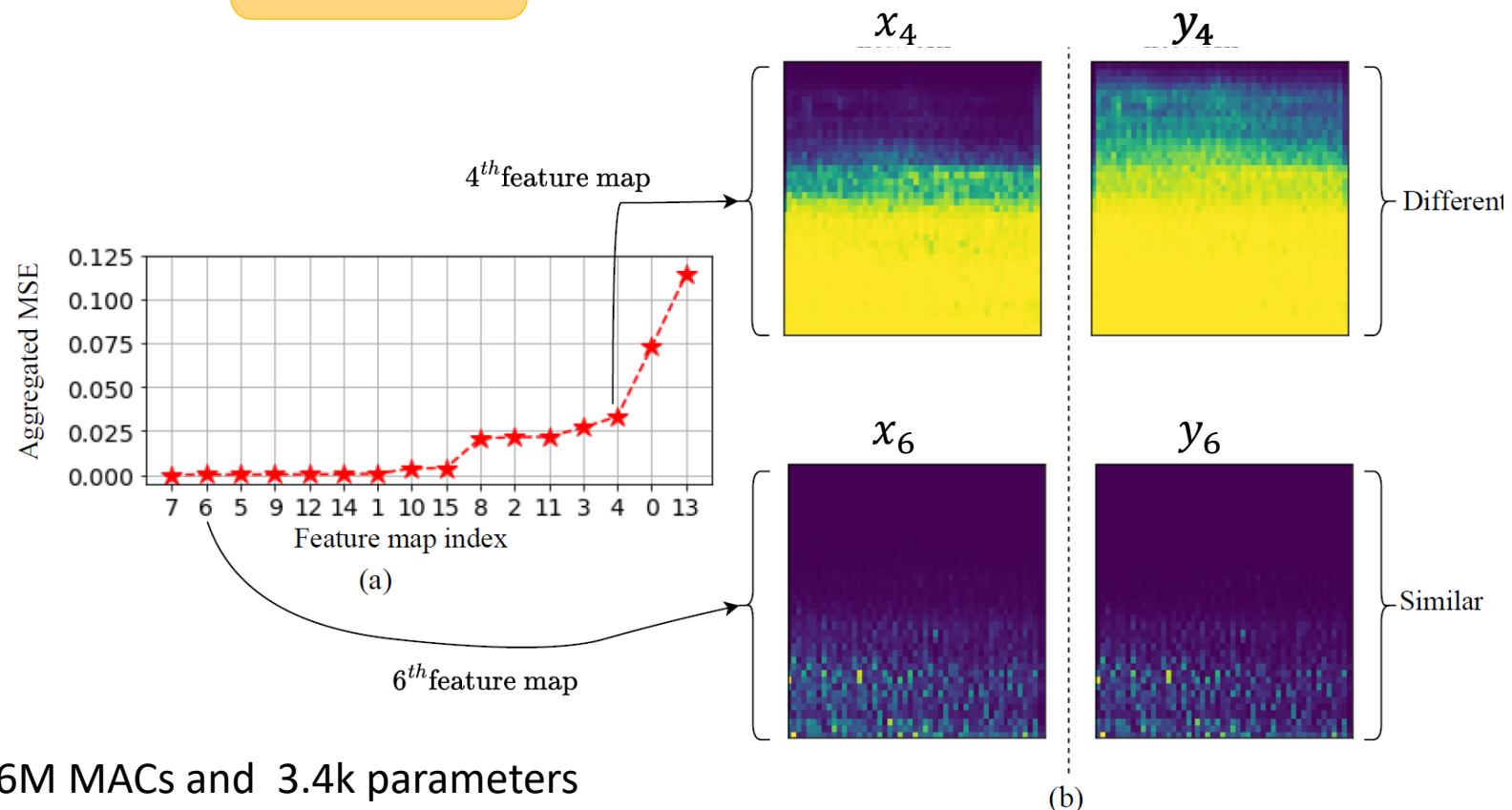
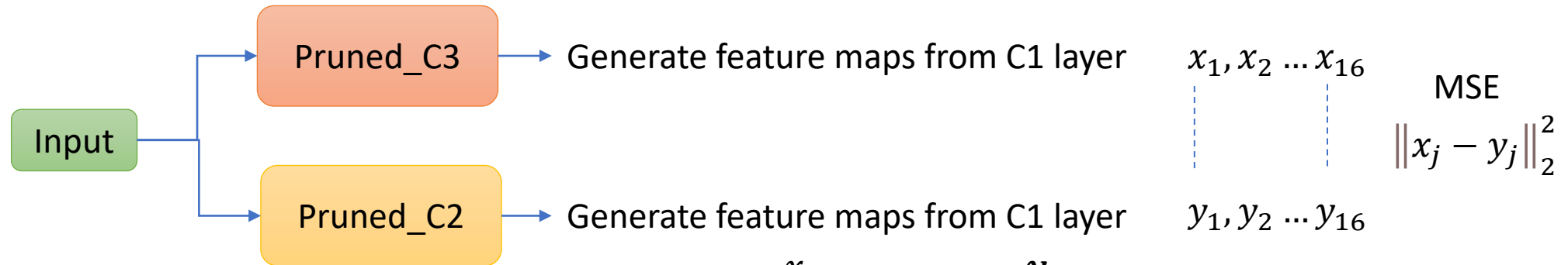
[4] Frankle et al., "The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks", ICLR, 2018.

Proposed Pipeline

DCASE 2022 Task 1 ASC → Resource limit: **Parameters < 128K, MACs < 30M**



Ensemble framework: Intra-model redundancy



→ Can reduce further 6M MACs and 3.4k parameters

Take away points

- To utilize underlying resources effectively to deploy CNNs,
 - Select optimal filter size,
 - Perform Pruning + quantization,
 - Ensemble to improve performance,
 - Eliminate intra-model redundancy in the ensemble.

Please visit my poster for suggestions and discussion: Session II: P1