4.3 資源使用量比較

表 4 列出四者模型於推論時之資源使用量，當中的數據以科學記號方式表達，同時對實數位小數點第三位以下的數進行無條件捨去，其中，VGGSKCCT 因爲使用 3 個架構相同的模型做 fusion ，因此其參數量與計算量以單一模型之數據的 3 倍作爲實驗結果數據。首先比較參數量，RepVGGRNN 與 baseline 、VGGSKCCT 與 Ebbers UPB task 4_4 相比皆爲其中最少者，總參數量約爲 49.6 萬個，僅使用 baseline 參數量約 111.2 萬之 44.6%，顯示了 RepVGGRNN 透過整體架構的縮減仍可以相對較少的參數量達到接近 baseline 系統的效能，與 VGGSKCCT 系統相比，僅約其 748.5 萬參數量之 6.6%，且是 Ebbers UPB task 4_4 系統 1.34 億參數量之 0.3% 。除了空間上的占用量外，運算量亦爲輕量化模型所需縮減的目標之一，RepVGGRNN 之運算量爲三個模型中最少者，其處理單一筆資料共需約 5.279 億次浮點運算，爲 baseline 9.309 億次運算之 56.7%，且爲 VGGSKCCT 所需 107 億次運算之 8.7%，可見 RepVGGRNN 模型在重參數化與縮減模型層數後，其在資源使用上具有相當的優勢。

5 結論

近年來隨著移動式裝置的普及，結合深度學習的移動端應用亦隨之而發展，除了網路模型本身的效能外，硬體資源使用的情形如記憶體使用量、續航力與運算需求亦是模型部屬所考量的方向，透過我們的實驗結果可見，RepVGGRNN 在驗證集中以 PSDS-1 , PSDS-2 分別爲 0.408% , 0.677% 皆高於 baseline 系統所達到的 0.344%, 0.572% ，且在資源使率中其參數量僅使用約 49.6 萬個，少於 baseline 系統所具有的 111.2 萬個參數，顯示了相比於 baseline 系統，其兼具了高準確性及輕量化的特色。在未來，希望能持續增進系統並在移動端裝置實踐聲音事件偵測之相關應用。

References

Çağdaş Bilen, Giacomo Ferroni, Francesco Tuveri, Juan Azcarreta, and Sacha Krstulović. 2020. A framework for the robust evaluation of sound event detection. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 61–65. IEEE.

Minjun Chen, Tian Wang, Jun Shao, Yiqi Tang, Yangyang Liu, Bo Peng, Jie Chen, and Xi Shao. 2022. Dcase 2022 challenge task4 technical report. Technical report, DCASE2022 Challenge.

Xiaohan Ding, Xiangyu Zhang, Ningning Ma, Jungong Han, Guiguang Ding, and Jian Sun. 2021. Repvgg: Making vgg-style convnets great again. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13733–13742.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Geoffrey Hinton, Oriol Vinyals, Jeff Dean, et al. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2(7).

Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Changmin Kim and Siyoung Yang. 2022. Sound event detection system using fixmatch for dcase 2022 challenge task 4. Technical report, DCASE2022 Challenge.

Xiang Li, Wenhai Wang, Xiaolin Hu, and Jian Yang. 2019. Selective kernel networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 510–519.

Koichi Miyazaki, Tatsuya Komatsu, Tomoki Hayashi, Shinji Watanabe, Tomoki Toda, and Kazuya Takeda. 2020. Convolution-augmented transformer for semi-supervised sound event detection. Technical report, DCASE2020 Challenge.

Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Antti Tarvainen and Harri Valpola. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30.

Nicolas Turpault, Romain Serizel, Justin Salamon, and Ankit Parag Shah. 2019. Sound event detection in domestic environments with weakly labeled data and soundscape synthesis.

Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. 2017. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*.

Xu Zheng, Han Chen, and Yan Song. 2021. Zheng ustc team's submission for dcase2021 task4 — semi-supervised sound event detection. Technical report, DCASE2021 Challenge.