Paper Title*

*Note: Sub-titles are not captured in Xplore and should not be used

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

Abstract—This document is a model and instructions for LaTeX. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.

Index Terms—component, formatting, style, styling, insert

I. Introduction

Learning in acoustic environmental noise is challenging due to its own characteristics. On the one hand, the noise waveforms of different acoustic scenes are relatively stable, and it is difficult to extract useful features. On the other hand, there are similarities in the acoustic characteristics of different environments, for example, there may be human voices in the speech data collected for several seconds from both the parks and public squares scenarios, which brings greater challenges to feature extraction.

Identify applicable funding agency here. If none, delete this.

II. Related Work

A. Acoustic scene classification

The prestigious detection and classification of acoustic scene and events (DCASE) [1] challenge covers state-of-the-art techniques for classifying acoustic environmental noise.

[2] [3] [4] both use data augmentation to expand the training set to bring larger samples for model training. [2] focuses on improving model performance on the data, demonstrating the importance of data preprocessing for embedded machine learning performance. From a data-centric perspective, [3] proves that the parameter setting of data preprocessing has a certain impact on model fairness.

In recent years, research in acoustic scene classification has focused on CNN network [5], especially ResNet [6] and DenseNet [7]. They have excellent performance in the field of image processing, but due to the characteristics of the acoustic scene, if the resnet is directly applied to them, the network performance will be greatly reduced. [8] proves this, and proposes to use 1D and 2D convolution in speech data at the same time, extending the time output to the frequency time dimension. [9] proves the effect of receptive field on generalization ability in acoustic scene classification problem.

B. Model Compression

Model compression has abundant research achievement [10]. From the perspective of model structure, [11] [12] improves the convolution kernel structure of the commonly used convolutional neural network (CNN). [13]Tensor (or matrix) operations are the basic operations of neural networks, so tensor decomposition is an effective way to shrink and speed up neural network models. [14] [15] [16]Data quantization is designed to solve the problem that most embedded devices do not support floating-point operations, and is widely used in model compression of mobile devices.

In addition, in the image domain, many lightweight networks for compressing models emerge, which greatly reduces the amount of parameters and memory overhead. The fire module of Squeezenet [17] is composed of squeeze and expand parts. The commonly used 3×3 convolution kernel is replaced with a 1×1 convolution kernel, which effectively reduces the number of parameters. In order to improve the model accuracy, a small number of 3×3 convolution kernels are spliced in the expand layer. The great thing about MobileNets [18] [19] [20] are the design of the depthwise separable convolutional structure, which reduces the complexity exponentially. These studies have achieved certain performance on images, but the compressed models are still difficult to use in low-power embedded devices.

Another perspective is the knowledge distillation method. Hinton [21] designed the teacher-student structure, that first training a huge teacher model, and then learning a relatively small model from the teacher model. Knowledge distillation is often used in acoustic scene classification problems [22].

C. Machine learning in LPWAN

In recent years, the compressed models are mostly deployed on mobile devices, which are all implemented relying on the backbone network. The implementation of AI technology in LPWAN is mainly concentrated in the field of cognitive radio [23]. [24]employs deep neural networks (DNNs) to intelligently explore data-driven test statistics to accurately characterize real-world environments. [25] proposed a cognitive C-LPWAN architecture based on an artificial intelligence cognitive engine to reduce network latency and minimum energy consumption rate, incorporating sensor selection for a battery-powered IoT-assisted cognitive radio (CR-IoT) network The strategy is applied in LPWAN to extend the life of LoRa network.

[26] [27] [28] [29] [30]implement machine learning algorithms in embedded devices. [26] designed a serial-FFT-based Mel-frequency cepstrum coefficient circuit, and used binary depthwise separable convolution to reduce power consumption. [30] jointly designed a framework for an efficient neural architecture (TinyNAS) and a lightweight inference engine (TinyEngine), and its inference speed is 1.7-3.3× faster than TF-Lite Micro and CMSIS-NN

References

[1] D. Stowell, D. Giannoulis, E. Benetos, M. Lagrange, and M. D. Plumbley, “Detection and classification of acoustic scenes and events,” IEEE Transactions on Multimedia, vol. 17, no. 10, pp. 1733–1746, 2015.

[2] D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, “Specaugment: A simple data augmentation method for automatic speech recognition,” Interspeech 2019, Sep 2019. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2019-2680

[3] W. Toussaint, A. Mathur, A. Y. Ding, and F. Kawsar, “Characterising the role of pre-processing parameters in audio-based embedded machine learning,” in Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, ser. SenSys ’21. New York, NY, USA: Association for Computing Machinery, 2021, p. 439–445. [Online]. Available: https://doi.org/10.1145/3485730.3493448

[4] A. Y. Hannun, C. Case, J. Casper, B. Catanzaro, G. F. Diamos, E. Elsen, R. J. Prenger, S. Satheesh, S. Sengupta, A. Coates, and A. Ng, “Deep speech: Scaling up end-to-end speech recognition,” ArXiv, vol. abs/1412.5567, 2014.

[5] Y. Lee, S. Lim, and I.-Y. Kwak, “Cnn-based acoustic scene classification system,” Electronics, vol. 10, no. 4, 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/4/371

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2015.

[7] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” 2018.

[8] B. Kim, S. Chang, J. Lee, and D. Sung, “Broadcasted residual learning for efficient keyword spotting,” ArXiv, vol. abs/2106.04140, 2021.

[9] K. Koutini, H. Eghbal-zadeh, M. Dorfer, and G. Widmer, “The receptive field as a regularizer in deep convolutional neural networks for acoustic scene classification,” 2019.

[10] L. Deng, G. Li, S. Han, L. Shi, and Y. Xie, “Model compression and hardware acceleration for neural networks: A comprehensive survey,” Proceedings of the IEEE, vol. 108, no. 4, pp. 485–532, 2020.

[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” 2015.

[12] B. Wu, A. Wan, X. Yue, P. Jin, S. Zhao, N. Golmant, A. Gholaminejad, J. Gonzalez, and K. Keutzer, “Shift: A zero flop, zero parameter alternative to spatial convolutions,” 2017.

[13] V. Klema and A. Laub, “The singular value decomposition: Its computation and some applications,” IEEE Transactions on Automatic Control, vol. 25, no. 2, pp. 164–176, 1980.

[14] M. Courbariaux, Y. Bengio, and J.-P. David, “Binaryconnect: Training deep neural networks with binary weights during propagations,” in Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2, ser. NIPS’15. Cambridge, MA, USA: MIT Press, 2015, p. 3123–3131.

[15] K. Hwang and W. Sung, “Fixed-point feedforward deep neural network design using weights +1, 0, and −1,” in 2014 IEEE Workshop on Signal Processing Systems (SiPS), 2014, pp. 1–6.

[16] L. Deng, P. Jiao, J. Pei, Z. Wu, and G. Li, “Gxnor-net: Training deep neural networks with ternary weights and activations without full-precision memory under a unified discretization framework,” 2018.

[17] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “Squeezenet: Alexnet-level accuracy with 50x fewer parameters and $<$0.5mb model size,” 2016.

[18] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” 2017.

[19] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” 2019.

[20] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, “Searching for mobilenetv3,” 2019.

[21] G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” 2015.
[22] J.-W. Jung, H.-S. Heo, H.-J. Shim, and H.-J. Yu, “Knowledge distillation in acoustic scene classification,” IEEE Access, vol. 8, pp. 166 870–166 879, 2020.
[23] A. J. Onumanyi, A. M. Abu-Mahfouz, and G. P. Hancke, “Towards cognitive radio in low power wide area network for industrial iot applications,” in 2019 IEEE 17th International Conference on Industrial Informatics (INDIN), vol. 1, 2019, pp. 947–950.
[24] C. Liu, J. Wang, X. Liu, and Y.-C. Liang, “Deep cm-cnn for spectrum sensing in cognitive radio,” IEEE Journal on Selected Areas in Communications, vol. 37, no. 10, pp. 2306–2321, 2019.
[25] M. Chen, Y. Miao, X. Jian, X. Wang, and I. Humar, “Cognitive-lpwan: Towards intelligent wireless services in hybrid low power wide area networks,” IEEE Transactions on Green Communications and Networking, vol. 3, no. 2, pp. 409–417, 2019.
[26] W. Shan, M. Yang, T. Wang, Y. Lu, H. Cai, L. Zhu, J. Xu, C. Wu, L. Shi, and J. Yang, “A 510-nw wake-up keyword-spotting chip using serial-fft-based mfcc and binarized depthwise separable cnn in 28-nm cmos,” IEEE Journal of Solid-State Circuits, vol. 56, no. 1, pp. 151–164, 2021.
[27] A. Andreadis, G. Giambene, and R. Zambon, “Convolutional neural networks for audio classification on ultra low power iot devices,” in 2021 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), 2021, pp. 1–6.
[28] S. Jung, C. Liao, Y. Wu, S.-M. Yuan, and C.-T. Sun, “Efficiently classifying lung sounds through depthwise separable cnn models with fused stft and mfcc features,” Diagnostics, vol. 11, no. 4, Apr. 2021, publisher Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland.
[29] G. Cerutti, R. Prasad, and E. Farella, “Convolutional neural network on embedded platform for people presence detection in low resolution thermal images,” in ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Conference Proceedings, pp. 7610–7614.
[30] J. Lin, W.-M. Chen, Y. Lin, J. Cohn, C. Gan, and S. Han, “Mcunet: Tiny deep learning on iot devices,” 2020.