A Target-Separable BWN Inspired Speech Recognition Processor with Low-power Precision-adaptive Approximate Computing

Bo Liu[1], Hao Cai[1*], Xuan Zhang[1], Haige Wu[1], Anfeng Xue[1], Zilong Zhang[1], Zhen Wang[2] and Jun Yang[1]
[1] National ASIC System Engineering Research Center, Southeast University, Nanjing, China
[2] Nanjing Prochip Electronic Technology Co. Ltd. Nanjing, China
[*]Corresponding author: hao.cai@seu.edu.cn
[1] {liubo_cnasic, z−xuan, wu_haige, xue_anfeng, zhangzilong, dragon}@seu.edu.cn
[2] fanstics@prochip.com.cn

***Abstract*—This paper proposes a speech recognition processor based on a target-separable binarized weight network (BWN), capable of performing both speaker verification (SV) and keyword spotting (KWS). In traditional speech recognition system, the SV based on traditional model and the KWS based on neural networks (NN) model are two independent hardware modules. In this work, both SV and KWS are processed by the proposed BWN with unified training and optimization framework which can be performed for various application scenarios. By the system-architecture co-design, SV and KWS share most of the network parameters, and the classification part is calculated separately according to different targets. An energy-efficient NN accelerator which can be dynamically reconfigured to process different layers of the BWN with splitting calculation of frequency domain convolution is proposed. SV and KWS can be achieved with only one time calculation of each input speech frame, which greatly improves the computing energy efficiency. The computing units of the NN accelerator are optimized using precision-adaptive approximate computing method with Dual-VDD to further reduce the energy cost. Compared to state-of-the-arts, this work can achieve about 4× reduction in power consumption while maintaining high system adaptability and accuracy.**

***Index Terms*—Speaker verification, keyword spotting, precision-adaptive approximate computing, binarized weight network, frequency domain convolution, target-separable**

I. Introduction

As a user-friendly way of human-computer interaction, automatic speech recognition is widely used in modern life, such as wearable mobile devices and the Internet of Things (IoT). In these battery-powered devices, the system is usually required to be always on. Therefore, real-time processing with ultra-low power consumption and high recognition accuracy is a key requirement. In the past few decades, many works are focusing on the architecture design and optimization of low-power and high accuracy keyword spotting (KWS) [1]–[6] and speaker verification (SV) [7]–[12].

Many deep neural network (DNN) structures and accelerators for low power speech recognition have been proposed in recent years. In work [4], an optimized binary neural network (BNN) is proposed for KWS. In this BNN, the bit width of data and weight are both 1 bit, 99% of the BNN operations are additions and the multiplication operations are almost eliminated. In work [13], a binary convolutional neural network is utilized to achieve extremely low power consumption for 1∼2 keywords recognition. In addition, some approximate calculation circuits have been proposed to further improve calculation energy efficiency [14]–[19]. In work [5], they proposed a binary weight network (BWN) based KWS architecture with approximate computing to optimize the recognition accuracy and the power consumption. However, these systems do not support speaker selectivity, so these devices can be triggered by anyone's voice. In work [20], they proposed a low-power speech recognition processor with KWS and SV. For the consideration of recognition accuracy, the processor uses a DNN model for KWS, and a traditional Gaussian mixture model (GMM) for SV [21]. This work has two disadvantages: first, KWS based on the DNN framework can be optimized for different background noises, but it is difficult for SV to use traditional models to do this; second, due to different computing models, two parts of hardware are required, resulting in low resource utilization.

This paper proposes a target-separable BWN inspired speech recognition processor system. To the knowledge of the authors, this is the first target-separable BWN topology and accelerator architecture to achieve both SV and KWS with only one time DNN computing. The proposed accelerator which can be dynamically reconfigured to process different layers of the BWN is optimized with splitting calculation of frequency domain convolution to improve the hardware utilization. SV and KWS can be achieved with only one time calculation of each input speech frame, which greatly improves the computing energy efficiency. The computing units are optimized using precision-adaptive approximate computing method with dual-rail voltage (Dual-VDD) to further reduce the energy cost. Compared to state-of-the-arts, this work can achieve about 4× reduction in power consumption while maintaining high system adaptability and accuracy.

II. Target-Separable BWN Inspired Speech Recognition Processor with Complementary Noise-Robust SV

The overall architecture of the speech recognition system is shown in the bottom of Fig.1. The system consists of four parts:

the voice activity detection (VAD) module, the signal noise ratio (SNR) module [5], the mel-scale frequency cepstral coefficients (MFCC) module [22], and the proposed target-separable BWN accelerator. Fig.2 shows the proposed target-separable BWN for both SV and KWS. The SV and KWS multiplex most of the convolutional layer speech feature extraction results. Different functional classifications are performed according to the recognition target. The classifier module is designed separately according to the two functions. The KWS adopts the traditional softmax classifier. According to the continuity characteristics of the speech, the classifier for SV can optimize the time sequence, which can effectively improve the accuracy of SV. By using the same convolutional layer to extract voice features and performing independent classification calculations on different target functional layers, both SV and KWS have achieved higher recognition accuracy and energy efficiency, and the computing reuse rate is 93.2%.

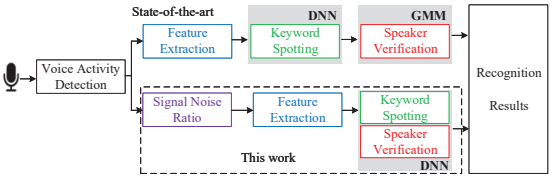


Fig. 1. Speech recognition system architecture

We also designed a continuity-based SV secondary classification module, as shown in equation1:

$$X_t = \frac{\sum_{i=1}^{n} x_{t-i} \times w_{t-i}}{n} + x_t \quad (1)$$

x is the original threshold classification result, X is the result after secondary classification, and w is the confidence weight of the original threshold classification result. The final classification label is obtained by calculating the average value and then taking the secondary threshold classification between the threshold classification result at the current time and the classification result at the previous t times.

The specific process is shown in Fig.3. After the original speech signal feature is extracted by the MFCC module and calculated by the BWN network, a pair of fully connected neurons is sent to the initial label classification unit. Take the initial classification with a wider threshold as an example, false alarm rate (FAR) is higher, which lead to that some fuzzy features in the speech signal may cause the network to make false judgments. If the current classification result and the previous t initial classification results in the time series label are sent to the final label classification unit for weighted average value calculation, we can amend the initial misclassification label which is wrong according to the characteristics of continuity, and the final correct classification label is obtained.

We simulated different values of t and threshold β, and the results are shown in Fig.3(a). When the value of t gradually increases and the value of β is larger, the secondary classification false reject rate (FRR) is higher, and the classification

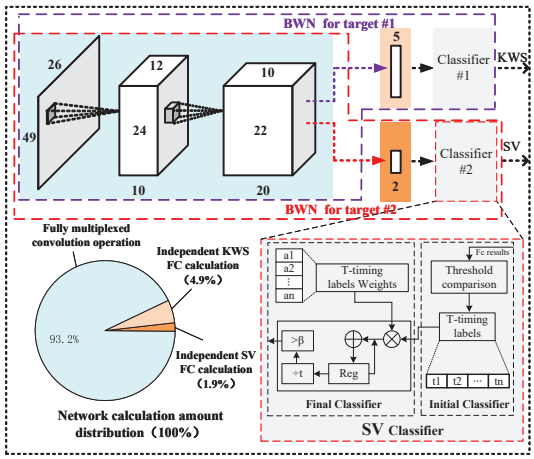


Fig. 2. The proposed target-separable BWN for both SV and KWS

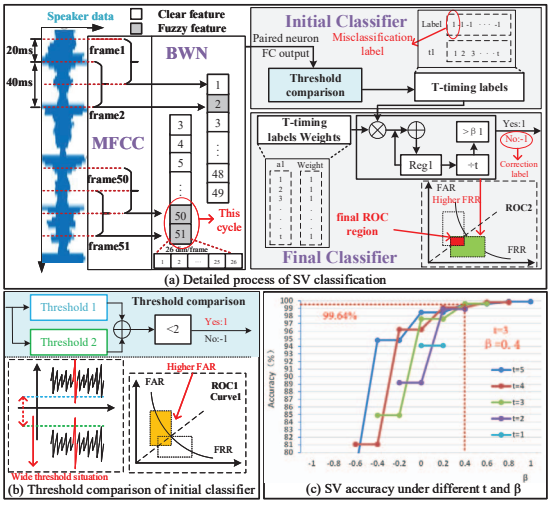


Fig. 3. SV classifier of target-separable BWN speech recognition

accuracy tends to be higher. When t is 3 and β is 0.4, the recognition rate can reach 99.64%. We tested the two most common SV databases Texas Instruments and Massachusetts Institute of Technology Acoustic-Phonetic Continuous Speech Corpus (TIMIT) and VoxCeleb2 [23] using single threshold comparison and continuity-based secondary classification respectively. As shown in Table I, our proposed complementary

TABLE I
COMPARISON OF SV CLASSIFICATION ACCURACY ON TWO DATASETS (SINGLE THRESHOLD VS SECONDARY CLASSIFICATION)

Dataset	SV classification	FRR	FAR	Acc
TIMIT	single	28.40%	15.36%	79.30%
	secondary	0.40%	0.80%	99.40%
VoxCeleb2	single	28.20%	17.09%	78.50%
	secondary	1.0%	1.70%	98.60%

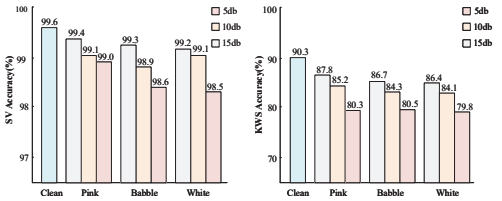


Fig. 4. Recognition accuracy of target-separated BWN processor for SV(Left) and KWS(Right) under different background noise

secondary classification can increase the accuracy of SV by about 20% and more stable. The classification threshold can be adjusted according to the SNR results to adapt to different background noises. We choose TIMIT and Google speech commands dataset (GSCD) [24] as benchmarks during evaluation. Fig.4 shows the recognition accuracy of target-separated BWN accelerator for SV and KWS (1-speaker, 4-keywords) under different background noise. Even for the background noise of white noise with SNR of 5dB, the recognition accuracy of the SV based on complementary ROC can achieve up to 98.5%.

III. Framework of the Target-Separable BWN Inspired Speech Recognition Processor

A. Speech Recognition Processor with Frequency Domain Convolution Multiplexing

Fig.5(a) shows the architecture of the proposed speech recognition processor, which mainly consists of energy-based VAD unit, SNR detection unit, MFCC unit, target-separable BWN accelerator, classification unit and main controller. BWN weights and calculation results are all stored on the chip, using 3.99 KB weight memory and 9.95 KB data memory respectively. These hardware units are clock-gated, for example, the feature extraction unit can only work when the voice signal passes through its previous VAD unit, etc. The voice signal enters the VAD unit, the SNR unit and the MFCC unit. The VAD unit processed by 0-mean computing and accumulated to compare with the already set threshold, the result is used to activate downstream units. After that, the SNR unit calculates the short-term energy and zero-crossing rate of the input voice signal to obtain the current SNR prediction result for selecting the threshold of the classification module. MFCC unit to extract 26-dimensional speech features.

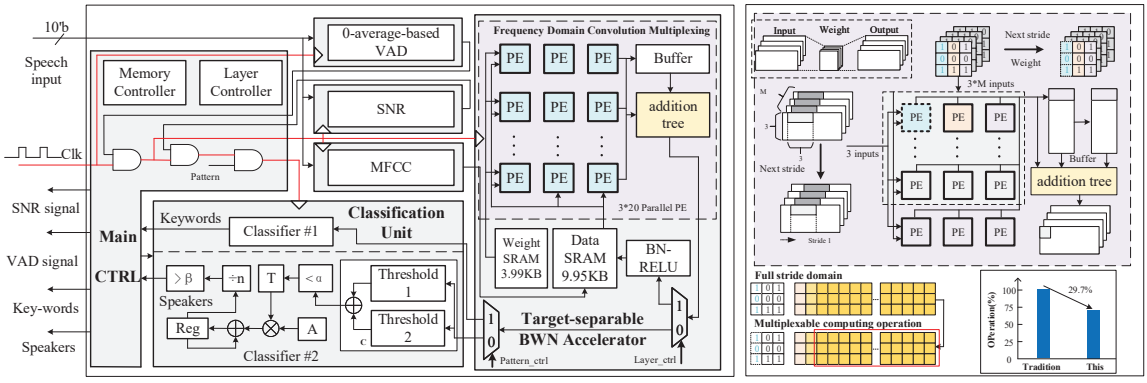
The BWN accelerator can be configured according to the identification target, supports flexible input feature size and number of cascading layers, adopts a hierarchical multiplexing working mode, and configures corresponding resources according to the characteristics of the current layer. The computing array is composed of 60 processing element (PE) modules. The PE architecture is shown in Fig.6(a). The layer controller controls the number of different layers of PE. If the current convolution kernel needs to be calculated as $3\times3\times N\times M$, the layer controller will open $3\times M$ PE units and cache reuse the reusable calculation results according to the weight characteristics of the convolution kernel and the frequency domain step size to reduce redundant calculations.

Frequency domain calculation multiplexing is shown in Fig.5(b). For example, the weight of the first column of the 3×3 convolution kernel is the same as the weight of the third column, and the convolution step size is 1. The third column result of the current convolution calculation is also the calculation result of the first column convolution after two strides. The third PE unit will accumulate and save the multiplexing calculation results of the corresponding frequency domain in the entire convolution kernel in the calculation process to Rb. After the current convolution kernel calculation is completed, the output is saved to the Buffer for the next convolution calculation. The buffer unit saves the result of multiplexing after the first time to Buffer1, and the result of multiplexing after the second time to Buffer2. In the next stride of convolution, the Multiplex Rb_n signal is valid when the first PE unit calculates to the multiplexing part, and the PE unit skips the multiplexing calculation part. After the current calculation is completed, the current PE calculation result and the second Buffer result of the third PE are accumulated to obtain the final output result. The accumulator accumulates the output data of the PE unit and the data in the result-multiplexed buffer unit to obtain the current convolution calculation result. By bit processing the pre-trained BWN model and replacing the column weights of similar convolution kernel combining the weight matrix and the convolution step size, the calculation can be reduced by 29.7%.

The classification unit includes two modules: the KWS classification module using the traditional Softmax classification model and the SV classification module is designed based on the timing optimization classification model. The mode controller distributes the network calculation results to the classification unit of the corresponding function, and adjusts the classification threshold according to the SNR prediction result to obtain the SV and KWS labels. The main controller controls the working status of the chip, configures system parameters, manages clock gating and data access. The working state transition diagram is also shown in Fig.5(a). The processor can be configured with 3 functional modes, namely SV, KWS and SV & KWS.

B. Precision-Adaptive Approximate Computing Method with Dual-VDD for BWN Accelerator

Since more than 90% of the BWN used are addition operations, and the neural network is fault-tolerant for different accuracy requirements under different speech recognition environments. Compared with the previous approximate computing processors [25], we propose a precision-adaptive approximate computing method with Dual-VDD for both the adders in each PE and the addition tree unit to further reduce system power consumption. The precision-adaptive PE architecture is shown in Fig.6(a). The PE unit includes an optional 16bits standard adder and a Dual-VDD approximate adder, which can be selected according to the current SNR. Since the method of convolution splitting and multiplexing is utilized in this work, there is an accumulation process after the convolution calculation. As shown in the upper left part of Fig.6(b), the traditional accumulator is realized by an adder and an additional

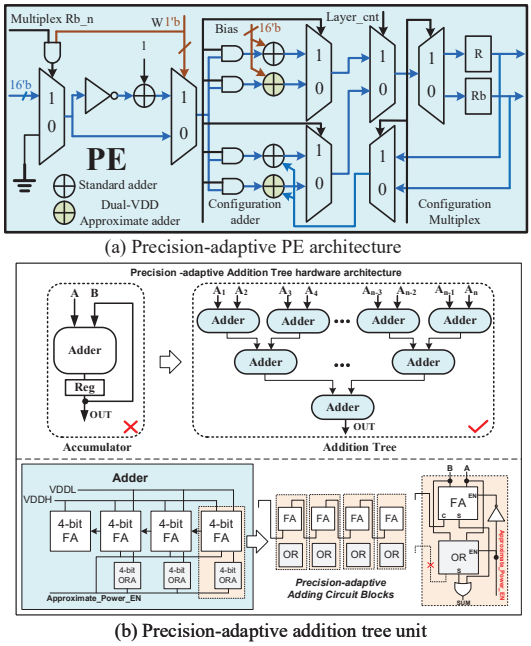


(a) Architecture of the proposed Speech Recognition Processor

(b) Frequency domain convolution multiplexing

Fig. 5. Architecture of the speech recognition processor with frequency domain convolution multiplexing

register. We propose an addition tree to achieve accumulation. As shown in the upper right part of Fig.6(b), the tree structure of the adder is formed by a single adder structure that adds two by two. The approximate computing accuracy of the addition tree structure is higher than that of the accumulation structure using a fixed hardware configuration, because there is only one addition hardware in the accumulation structure, and the hardware structure is repeatedly used to achieve the accumulation process. The addition tree structure can perform independent approximate computing configuration for the adders of each stage, which has higher calculation accuracy. Both in PE and addition tree approximate adder, Full adders (FA) are used in the high bits, while the low bits can be configured to use OR-gate approximate adder (ORA). The number bit of ORAs in the approximate adder is determined by SNR, the higher SNR means the more bit of ORAs. The greater noise, the lower SNR, which means more accurate computing is required. When the background noise is clean, the 16bits approximate adder is configured as 12bits ORA in low bits and 4bits FA in high bits. When the SNR is 15dB, the 16bits approximate adder is configured as 8bits ORA in low bits and 8bits FA in high bits. Similarly, when SNR is 5dB, the 16bits are configured as FAs.



(a) Precision-adaptive PE architecture

(b) Precision-adaptive addition tree unit

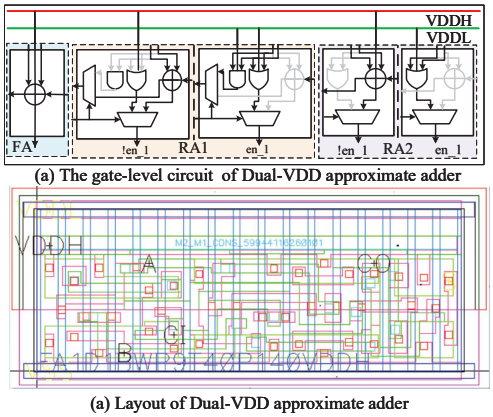
Fig. 6. Precision-adaptive approximate computing method for BWN accelerator

IV. Implementation Results

The OR-gate based approximate adders are commonly used to improve the energy efficiency of hardware accelerators. Compared with the previous approximate adders with fixed voltage configuration [5], the Dual-VDD approximate adders proposed in this paper can further reduce power consumption. The gate-level circuit is shown in Fig.7(a). Specifically, the approximate adder is a Dual-VDD structure, we implement the Dual-VDD design by adding an extra power rail. As shown in Fig.7(b), a power rail is added to the cell layout to form a Dual-VDD structure, including high power supply voltage (VDDH) and low power supply voltage (VDDL). The specific selection of the exact component FA or the approximate component ORA is determined by the corresponding control bit. The FAs use VDDH and the ORAs use VDDL, since ORAs discard the carry chain. When the VDDL decrease, the delay of the critical path will increase, and ORA has a smaller delay than FA. Therefore, by configuring VDDL, the delay of ORA can be increased to be similar to FA. This method does not affect the accuracy and performance of the circuit, but a lower VDD means lower power consumption. The overhead of layout area caused by the Dual-VDD method is not large, but due to the ORAs use VDDL, the power consumption is reduced by square times. Under TSMC 22nm ultra-low-leakage (ULL) process, VDDH is fixed to the conventional power supply voltage of 0.6V, and VDDL is evaluated in the interval 0.3V∼0.6V with a step size of 0.01V. By evaluating and comparing the critical path delay of each component, when VDDL=0.39V, the approximate component delay is equal to the critical path delay of the

TABLE II
COMPARISONS WITH THE STATE-OF-THE-ART ARCHITECTURES

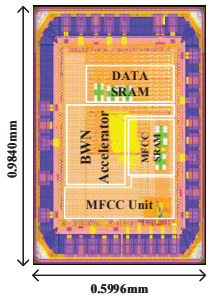
Reference	TVLSI'2014 [7]	ESSCIRC'2018 [3]	ISSCC'2020 [13]	TCSI'2020 [5]	JSSC'2020 [20]	This work
Technology	90 nm	65 nm	28 nm	22 nm	65 nm	22 nm
Area	19.53 mm^2	1.04 mm^2	0.23 mm^2	0.60 mm^2	2.56 mm^2	0.28 mm^2
On-chip SRAM	NA	32 KB	2 KB	11 KB	105 KB	14 KB
Voltage (V)	0.9	0.57	0.41	0.6	0.6	0.6/0.3 Dual-VDD
VAD	NA	NA	NA	NA	✓	✓
SNR	NA	NA	NA	✓	NA	✓
KWS	NA	✓	✓	✓	✓	✓
SV	✓	NA	NA	NA	✓	✓
Frequency	100 MHz	250 KHz	40 KHz	250 KHz	250 KHz	250 KHz
speech length	2 s	500 ms	1 s	1 s	1 s	1 s
Latency	8 ms	16 ms	64 ms	16 ms	KWS:16 ms SV:515 ms - 1 s	KWS:16 ms SV:16 ms
Main Accelerators	LPC SVM	MFCC LSTM	MFCC DSCNN	MFCC BWN	MFCC LSTM GMM	MFCC BWN
Database	NIST SRE	TIMIT	GSCD	GSCD	SV:TIMIT KWS:GSCD	SV:TIMIT KWS:GSCD
KWS Accuracy	NA	NA	2-words 94.6%	10-words 87.9%@clean 80.80%@5dB	10-words 90.87%@clean	4-words 90.30%@clean 79.80%@5dB
SV Accuracy	92.49%@clean	NA	NA	NA	99.50%@clean	99.64%@clean 98.50%@5dB
Power	8.12 mW	5 μW	0.51 μW	15.10 μW	18.3 μW	4.53 μW



(a) The gate-level circuit of Dual-VDD approximate adder

(a) Layout of Dual-VDD approximate adder

Fig. 7. Dual-VDD driven approximate adder



Process	22nm CMOS
Supply voltage	0.39V / 0.6 V
Frequency	250 KHz
Core size	0.6924×0.3997 mm^2
Die size	0.9840×0.5996 mm^2
Memory	14 KB
Power consumption	4.53 uW

Fig. 8. Layout of prototype system

precise component, and the power consumption is reduced by 15%. By utilizing this proposed precision-adaptive approximate computing method with Dual-VDD, the power consumption of the BWN accelerator can be reduced by 32%.

To evaluate the power consumption and recognition accuracy of target-separable BWN accelerator, the prototype processor shown in Fig.4 is implemented and evaluated on an industrial 22 nm ULL process technology. The prototype system is functional with the logic supply voltage of VDDH 0.6V and VDDL 0.39V, and the clock frequency is 250 KHz. The layout of the proposed target-separable BWN accelerator is shown in Fig.8. The area of the prototype system is 0.59 mm^2 with the I/O PADs (0.2767 mm^2 without the I/O PADs). We choose TIMIT and GSCD as benchmarks during evaluation. With the SNR prediction module to adjust the system parameters, we can achieve higher recognition accuracy ($\geq$ 98.5%@SV, $\geq$ 79.8%@KWS) under a wide range of background noise (SNR$\geq$5dB).

Table II compares this work with previous state-of-the-art (SoA) voice command recognition ASICs. There are few dedicated speech recognition processors that support SV and KWS at the same time. In work [7], SV uses linear predictive coding (LPC) [26] to extract speech features, and support vector machine (SVM) for processing, which lacks sufficient SV accuracy and does not support KWS. In the work [20], KWS uses the DNN model, while SV uses the traditional model. Due to different models, computing and storage are independent of each other, resulting in low resource utilization. In addition, the SV interval exceeds 500ms, and cannot be identified in real time. Moreover, the accelerator lacks noise robustness and can only be used near-microphone scene where the background noise is very low and can be ignored.

In this work, we propose a target-separable BWN inspired speech recognition processor system. The system can process both SV and KWS. By multiplexing most of the speech feature extraction parameters, the computing reuse rate has reached 93.2%. Also, the calculation result reuse and Dual-VDD approximate adder are utilized for the target-separable BWN architecture to further reduce power consumption, and the system parameters are adjusted through SNR to maintain high adaptability. At the working frequency of 250 KHz, in the case

of 1-speaker and 4-keywords, the power consumption is 4.50 μW / 4.52 μW (only SV/KWS active) or 4.53 μW (SV&KWS both active). Compared to SoA designs, the proposed target-separable BWN accelerator can achieve both SV and KWS with only one time DNN computing, and can achieve noise-robust high accuracy with low-power consumption.

V. Conclusion

This paper proposes a target-separable BWN inspired speech recognition processor, which achieves both keyword spotting (KWS) and speaker verification (SV) with only one time DNN computing. The unified training and optimization framework can be utilized to optimize the entire system to adapt to different application scenarios. An energy-efficient NN accelerator with splitting calculation of frequency domain convolution method is proposed to process different layers of the BWN. The computing units of the NN accelerator are optimized using precision-adaptive approximate computing method with Dual-VDD to further reduce the energy cost. Implemented and evaluated under 22nm CMOS technology, the proposed processor can achieve higher recognition accuracy of over 98.5%@SV and over 79.8%@KWS under a wide range of background noise (SNR from 5dB to clean), with the low power consumption of 4.53 μW (both SV and KWS are activated). Compared to state-of-the-art architectures, our work can achieve about 4$\times$ reduction in power consumption with high system adaptability and accuracy.

Acknowledgment

This work is funded with the National Key Research and Development Project of China under Grant 2018YFB2202001.

References

[1] V. Tyagi, "Hybrid context dependent CD-DNN-HMM keyword spotting (KWS) in speech conversations," 26th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1-6, 2016, doi: 10.1109/MLSP.2016.7738862

[2] S. Yin, et al., "A High Energy Efficient Reconfigurable Hybrid Neural Network Processor for Deep Learning Applications," IEEE Journal of Solid-State Circuits, vol. 53, no. 4, pp. 968-982, 2018, doi: 10.1109/JSSC.2017.2778281

[3] J. S. P. Giraldo, et al., "Laika: A 5 μW Programmable LSTM Accelerator for Always-on Keyword Spotting in 65nm CMOS," ESSCIRC 2018 - IEEE 44th European Solid State Circuits Conference (ESSCIRC), pp. 166-169, 2018, doi: 10.1109/ESSCIRC.2018.8494342

[4] S. Yin, et al., "A 141 μW, 2.46 PJ/Neuron Binarized Convolutional Neural Network Based Self-Learning Speech Recognition Processor in 28NM CMOS," IEEE Symposium on VLSI Circuits, pp. 139-140, 2018, doi: 10.1109/VLSIC.2018.8502309

[5] B. Liu, et al., "A 22nm, 10.8 μW/15.1 μW Dual Computing Modes High Power-Performance-Area Efficiency Domained Background Noise Aware Keyword-Spotting Processor," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 67, no. 12, pp. 4733-4746, 2020, doi: 10.1109/TCSI.2020.2997913

[6] M. Shah, et al., "A fixed-point neural network architecture for speech applications on resource constrained hardware," Journal of Signal Processing Systems, vol. 90, no. 5, pp. 727-741, 2018, doi: 10.1007/s11265-0161202-x

[7] J. Wang, et al., "VLSI Design for SVM-Based Speaker Verification System," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 23, no. 7, pp. 1355-1359, 2015, doi: 10.1109/TVLSI.2014.2335112

[8] J. Wang, et al., "Robust Speaker Identification and Verification," IEEE Computational Intelligence Magazine, vol. 2, no. 2, pp. 52-59, 2007, doi: 10.1109/MCI.2007.353420

[9] T. G. Clarkson, et al., "Speaker identification for security systems using reinforcement-trained pRAM neural network architectures," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 31, no. 1, pp. 65-76, 2001, doi: 10.1109/5326.923269

[10] W. M. Campbell, et al., "Support vector machines using GMM supervectors for speaker verification," IEEE Signal Processing Letters, vol. 13, no. 5, pp. 308-311, 2006, doi: 10.1109/LSP.2006.870086

[11] F. Hou and B. Wang, "Text-independent speaker recognition using probabilistic SVM with GMM adjustment," International Conference on Natural Language Processing and Knowledge Engineering, pp. 305-308, 2003, doi: 10.1109/NLPKE.2003.1275919

[12] M. Graciarena, et al., "Noise Robust Speaker Identification for Spontaneous Arabic Speech," 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, vol. 4, pp. IV-245-IV-248, 2007, doi: 10.1109/ICASSP.2007.366895

[13] W. Shan, et al., "14.1 A 510nW 0.41V Low-Memory Low-Computation Keyword-Spotting Chip Using Serial FFT-Based MFCC and Binarized Depthwise Separable Convolutional Neural Network in 28nm CMOS," 2020 IEEE International Solid State Circuits Conference - (ISSCC), pp. 230-232, 2020, doi:10.1109/ISSCC19947.2020.9063000

[14] W. Liu, et al., "Design and Analysis of Majority Logic Based Approximate Adders and Multipliers," IEEE Transactions on Emerging Topics in Computing, 2019, doi: 10.1109/TETC.2019.2929100

[15] H. Waris, et al., "Hybrid Partial Product-based High-Performance Approximate Recursive Multipliers," IEEE Transactions on Emerging Topics in Computing, 2020, doi: 10.1109/TETC.2020.3013977

[16] J. Huang, et al., "An inexact Newton method for unconstrained total variation-based image denoising by approximate addition," IEEE Transactions on Emerging Topics in Computing, 2021, doi: 10.1109/TETC.2021.3079715

[17] W. Liu, et al., "Design of Approximate Radix-4 Booth Multipliers for Error-Tolerant Computing," IEEE Transactions on Computers, vol. 66, no. 8, pp. 1435-1441, 2017, doi: 10.1109/TC.2017.2672976

[18] W. Liu, et al., "Design and Analysis of Approximate Redundant Binary Multipliers," IEEE Transactions on Computers, vol. 68, no. 6, pp. 804-819, 2019, doi: 10.1109/TC.2018.2890222

[19] B. Liu, et al., "A 1D-CRNN Inspired Reconfigurable Processor for Noise-robust Low-power Keywords Recognition," 2021 Design, Automation & Test in Europe Conference & Exhibition (DATE), pp. 495-500, 2021, doi: 10.23919/DATE51398.2021.9474172

[20] J. S. P. Giraldo, et al., "Vocell: A 65-nm Speech-Triggered Wake-Up SoC for 10-μW Keyword Spotting and Speaker Verification," IEEE Journal of Solid-State Circuits, vol. 55, no. 4, pp. 868-878, 2020, doi: 10.1109/JSSC.2020.2968800

[21] D. A. Reynolds, "An overview of automatic speaker recognition technology," 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, pp. IV-4072-IV-4075, 2002, doi: 10.1109/ICASSP.2002.5745552

[22] B. Liu, et al., "Precision Adaptive MFCC Based on R2SDF-FFT and Approximate Computing for Low-Power Speech Keywords Recognition," IEEE Circuits and Systems Magazine, vol. 21, no. 4, pp. 24-39, 2021, doi: 110.1109/MCAS.2021.3118175

[23] J. S. Chung, et al., "VoxCeleb2: Deep Speaker Recognition," Interspeech 2018, pp. 1086-1090, 2018, doi: 10.21437/Interspeech.2018-1929

[24] P. Warden, "Speech commands: A dataset for limited-vocabulary speech recognition," 2018, arXiv: 1804.03209. [Online]. Available:http://arxiv.org/abs/1804.03209

[25] Venkataramani, et al., "Quality programmable vector processors for approximate computing," 2013 46th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO), pp. 1-12, 2013, doi: 10.1145/2540708.2540710

[26] H. Gupta and D. Gupta, "LPC and LPCC method of feature extraction in Speech Recognition System," 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence), pp. 498-502, 2016, doi: 10.1109/CONFLUENCE.2016.7508171