用於 DNN 推論應用的面積與能量高效率 Base-2 Softmax 設計

Area- and Energy-Efficient Base-2 Softmax Design for DNN in Inference Applications

組別：晶片系統組　　專題生：林筱容、詹婷詒、林旻鈺　　指導教授：王進賢教授

摘要

本研究提出一個針對深度神經網路 (DNN) 推論應用所設計的高面積與能效的 Softmax 模組。透過採用 base-2 運算，並利用一次線性近似取代指數運算，本設計的 16 位元定點 Softmax 模組不需乘法器、除法器與大型查找表 (LUT)，即可在 CIFAR-10 與 ImageNet 等基準測試中達到與標準 Softmax 相同的推論準確率，同時大幅提升面積與能量效率。為了確保不同設計方法在不同技術節點間能公平比較，本研究採用 DeepScaleTool 進行多項設計參數的等效化分析。 Post-layout 後模擬結果顯示，所提出的 22 nm 設計可在 1.00 GHz 的運作頻率下，達到較現有最佳設計高出 1.95 倍的綜合效率。

簡介

隨著機器學習的快速發展，DNN 已被廣泛應用於影像 [1]、語音 [2] 與語言 [3] 等多種處理任務中。在分類任務中，DNN 通常在最後一層加入 Softmax 層，以 Softmax 函數能將特徵向量轉換為機率分佈，提升模型在訓練與推論階段的準確率。Softmax 函數的數學定義為:

$$softmax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \quad (1)$$

然而，Softmax 涉及的指數與除法運算容易造成溢位與精度問題，且在邊緣裝置中會帶來高硬體成本與低能效的挑戰。因此，利用近似運算實現高效 Softmax 硬體成為近年研究重點 [4]–[13] 現有 Softmax 架構可分為訓練導向 [4]–[8] 與推論導向 [9]–[13]。

推論導向 Softmax IC 設計的比較

本研究首先建立一套製程無關且具統一效率的比較方法，用以公平評估各推論導向 Softmax 架構：

1. DeepScaleTool [15] 製程節點調整：將不同製程下的面積、功耗與吞吐量統一換算至相同製程節點，消除製程技術差異造成的偏差。
2. 效率正規化 (Unified Efficiency)：建立統一的效能指標，綜合評估設計在運算速度、能耗與硬體成本間的平衡，公式如下。

$$E \equiv normalization\left(\frac{scaled\ Throughput}{scaled\ area \times scaled\ power}\right) \quad (2)$$

本研究針對推論導向的 Softmax 架構進行比較分析，歸納出以下重點:

1. LUT 為基礎的設計雖具有高吞吐量，但功耗較高。
2. CORDIC 為基礎的設計可降低面積成本，但犧牲了運算速度。
3. Base-2 為基礎的設計則在精度、功率與資源使用間達到最佳平衡，展現最高的整體效率，適合應用於 DNN 推論。

TABLE I.　推論導向 Softmax IC 設計的硬體指標效能比較

	[9], 2018	[10], 2020	[11], 2020	[12], 2021	[13], 2023
Approximation method (Base)	LUT (e)	CORDIC (e)	Base-2	Base-2	Base-2
Technology	28 nm	90 nm	28 nm	90 nm	28 nm
Frequency (GHz)	2.8	0.3291	2.78	0.31	1
Area (μm²)	15000	41913	10081	46816	80200
Power (mW)	51.6	n.a.	n.a.	2.769	8.93
Throughput (G/s)	22.4	0.3291	22.24	3.1	n.a.
MSE	n.a.	n.a.	1.06e-10	2.71e-04	n.a.
Scaled area (μm²)	15000	3722	10081	4158	80200
Scaled power (mW)	51.6	n.a.	n.a.	1.35	8.93
Scaled Throughput (G/s)	22.4	0.645	22.24	6.078	n.a.
Unified efficiency (E)	0.0267	n.a.	n.a.	1	n.a.

TABLE II.　推論導向 Softmax IC 設計的 Top-1 準確率比較

	[11], 2020	[12], 2021	[13], 2023
Data set	Enwik8	ImageNet1000	CIFAR-10
Accuracy	MobileNet: 67.06% (ideal: 67.06%) VGG16: 72.33% (ideal: 72.36%) InceptionV3: 79.45% (ideal: 79.45%)	MobileNet: 56.43% (ideal: 67.71%) VGG16: 69.90% (ideal: 72.52%) InceptionV3: 75.66% (ideal: 80.06%)	ResNet: 90.73% (ideal: 90.73%)

硬體架構設計

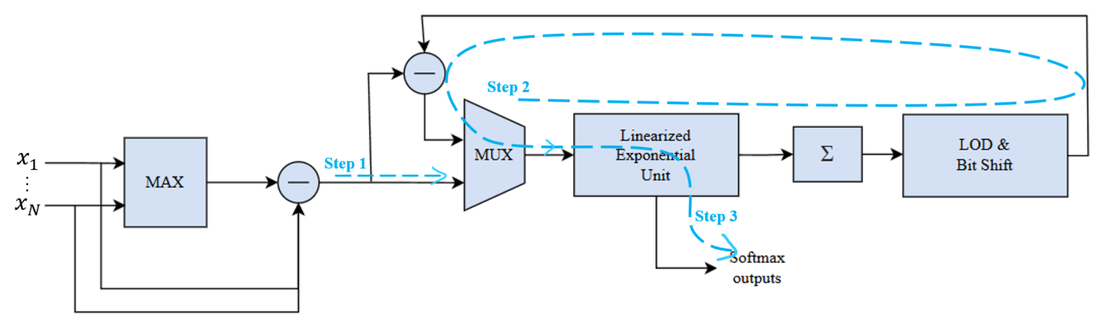

Fig. 1. Softmax 電路架構設計圖

實作結果

TABLE III.　推論導向 Softmax 設計硬體指標scaling 至 28nm 製程節點

	[12], 2021	Ours
Technology	90 nm	22 nm
Frequency (GHz)	0.31	1.00
Area (μm²)	46816	2872.87
Power (mW)	2.769	0.824
Throughput (G/s)	3.1	10.0
Scaled area (μm²)	4158	4567
Scaled power (mW)	1.35	0.96
Scaled Throughput (G/s)	6.078	9.25
Unified efficiency (E)	0.51	1

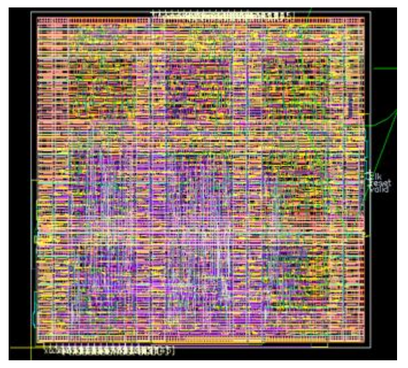

Fig. 2. 22nm Softmax 電路佈局

結論

本研究提出一個針對深度神經網路 (DNN) 推論應用所設計的高面積與能效的 Softmax 模組。透過採用以 2 為底的運算方式與一次線性近似來實現指數運算，所提出的 22 nm CMOS 製程、16 位元定點運算之 Softmax 模組，在 CIFAR-10 與 ImageNet 基準測試中能維持與標準 Softmax 相同的推論準確率，同時在綜合效能效率上較現有最佳設計提升 1.95 倍。

參考文獻

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” Commun. ACM, vol. 60, no. 6,pp. 84–90, May 2017.
[2] P. Tzirakis, J. Zhang, and B. W. Schuller, “End-to-End Speech Emotion Recognition Using Deep Neural Networks,” in IEEE Int. Conf. Acoust.,Speech Signal Process. (ICASSP), pp. 5089–5093 , Apr. 2018.
[3] Z. Liu, G. Li, and J. Cheng, “Hardware acceleration of fully quantizedBERT for efficient natural language processing,” in Design, Autom., TestEur. Conf., Exhib. (DATE), pp. 513–516, Feb. 2021.
[4] P. Nilsson, A. U. R. Shaik, R. Gangarajaiah and E. Hertz, "Hardware implementation of the exponential function using Taylor series," NORCHIP, pp. 1-4, 2014.
[5] Q. Sun et al., "A high speed SoftMax VLSI architecture based on basic-split," IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT), pp. 1-3, 2018.
[6] J. Kim, S. Kim, K. Choi and I. -C. Park, "Hardware-efficient Softmax architecture with bit-wise exponentiation and reciprocal calculation," in IEEE Trans. on Circuits and Systems (I), vol. 71, no. 10, pp. 4574-4585, Oct. 2024.
[7] Y. Zhang et al., "Base-2 Softmax function: Suitability for training and efficient hardware implementation," IEEE Trans. on Circuits and Systems (I), vol. 69, no. 9, pp. 3605-3618, Sept. 2022.
[8] Y. Zhang et al. "High-precision method and architecture for Base-2 Softmax function in DNN training," in IEEE Trans. on Circuits and Systems (I), vol. 70, no. 8, pp. 3268-3279, Aug. 2023.
[9] M. Wang, S. Lu, D. Zhu, J. Lin and Z. Wang, "A high-speed and low-complexity architecture for Softmax function in deep learning," IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), pp. 223-226, 2018.
[10] A. Kagalkar and S. Raghuram, "CORDIC based implementation of the Softmax activation function," International Symposium on VLSI Design and Test (VDAT), pp. 1-4, 2020.
[11] D. Zhu, S. Lu, M. Wang, J. Lin and Z. Wang, "Efficient precision-adjustable architecture for Softmax function in deep learning," in IEEE Trans. on Circuits and Systems (II), vol. 67, no. 12, pp. 3382-3386, Dec. 2020.
[12] Cardarilli et al. “A pseudo-Softmax function for hardware-based high speed image classification,” Sci. Rep., vol. 11, no. 1, p. 15307, Dec. 2021.
[13] D. Thabit, B. Alabassy, M. W. El-Kharashi, M. Safar and M. Dessouky, "DNN acceleration: A high-accuracy implementation for Base-2 Softmax layer," International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), pp. 1-6, 2023.
[14] M.-Y. Lin, T.-Y. Chan, H.-J. Lin, J.-S. Wang, and S.-C. Chang, “Comparative Study of Softmax IC Designs Using Process-Independent Unified Efficiency Index,” Proc. of IEEE International SoC Design Conference (ISOCC), 2025.
[15] S. Sarangi and B. Baas, "DeepScaleTool: A tool for the accurate estimation of technology scaling in the deep-submicron era," IEEE International Symposium on Circuits and Systems, pp. 1-5, 2021.