IEICE TRANSACTIONS

on Fundamentals of Electronics, Communications and Computer Sciences

DOI:10.1587/transfun.2024VLP0013

Publicized:2024/09/05

This advance publication article will be replaced by the finalized version after proofreading.

A PUBLICATION OF THE ENGINEERING SCIENCES SOCIETY

The Institute of Electronics, Information and Communication Engineers

Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

PAPER

RGB-Event Multi-modal NV-CiM to Detect Object by Mapping-Oriented Enhanced-Feature Pyramid Network with Mapping-Aware Group Convolution

Yuya Ichikawa†a), Naoko Misawa†, Chihiro Matsui†, Nonmembers and Ken Takeuchi†, Member

SUMMARY To overcome the excessive memory capacity of non-volatile CiM (NV-CiM) for multi-modal AI, this paper proposes Mapping-oriented enhanced-FPN (Feature Pyramid Network) fusion (More-FPN) as a RGB-event fusion object detection model. More-FPN includes three proposals. First proposal, Mapping-aware Group Convolution (MAGC), reduces the required NV-CiM capacity by suppressing the number of subarrays in NV-CiM at a fixed subarray size. In MAGC, the number of groups is optimized with no inference accuracy degradation. By adopting MAGC to FPN fusion of an RGB-event fusion object detection model, 54.7% subarrays are reduced. The second proposal, Separable Bridge (SepBridge), further reduces the number of subarrays by 26.1% from MAGC-adopted FPN fusion. Third proposal, Top-down path trainable BiFPN (TDT-BiFPN), achieves accuracy improvement with a slight subarray increase by adding bottom-up path and making top-down path trainable. By combining three proposals, More-FPN achieves both the reduction in subarrays by 61% and the accuracy improvement by 4.6%, compared with conventional FPN fusion CiM.

key words: Computation-in-Memory, group convolution, subarray separation, multi-modal AI, non-volatile memory

1. Introduction

Computation-in-Memory (CiM) is the promising accelerator for edge computing due to high-speed and low-power Multiply-Accumulate (MAC) calculation. By adopting emerging non-volatile memories (NVM) to CiM (NV-CiM), energy reduction is achieved because NVM does not require a power supply to maintain its information [1, 2]. In NV-CiM, weights of neural network are stored in conductance of NVM cells. With Kirchoff's current law, NV-CiM operates MAC by applying input voltage to the word-lines, and the MAC result is obtained as the bit-line current.

Multi-modal processing is performed to increase accuracy for autonomous driving, drone control, and audio-visual speech recognition [3, 4, 5, 6] In particular, fusing event sensor data [7, 8, 9] with RGB data has attracted much attention for object detection. For example, Feature Pyramid Network fusion (FPN fusion) [6] has achieved high object detection accuracy by combining RGB data and event sensor data. However, in multi-modal processing, the number of required weight parameters becomes large, which leads to the excessive memory capacity in NV-CiM implementation.

†The authors are with Dept. of Electrical Engineering and Information Systems, The University of Tokyo, 113-8656 Japan.
a) E-mail: ichikawa@co-design.t.u-tokyo.ac.jp

Assuming the limited memory capacity of NV-CiM [2, 10, 11], implementing multi-modal AI on NV-CiM is a big challenge. In [12], this challenge has been addressed with memory capacity-efficient RGB-event fusion, but the memory capacity is not directly reduced.

When mapping weights on CiM, partitioning into subarrays with fixed subarray size is usually performed [13]. To maintain high utilization rate of CiM, subarray size should be small. However, finely divided subarrays increase CiM area due to increased peripheral circuits and interconnections. Also, assuming the fixed subarray size, the number of subarrays is proportional to the memory capacity. Therefore, subarray reduction is important to reduce the memory capacity of NV-CiM and overhead of peripheral circuit, and to realize multi-modal AI on NV-CiM.

In this paper, Mapping-oriented enhanced-FPN fusion (More-FPN), an RGB-event fusion model is proposed to realize multi-modal NV-CiM (Fig. 1). More-FPN involves three proposals. The first proposal, Mapping-aware Group Convolution (MAGC), reduces the number of subarrays in NV-CiM by utilizing group convolution [14], which leads to the memory capacity reduction (Fig. 2). In MAGC, first, the search space for the number of groups of group convolution is narrowed by using three conditions. By narrowing the search space, the number of groups is determined for subarray reduction with no accuracy degradation. The second proposal, Separable Bridge (SepBridge) also reduces the number of subarrays. By combining proposed MAGC and SepBridge, significant CiM subarray reduction is achieved. The third proposal, Top-down path-trainable bi-directional FPN (TDT-BiFPN), overcomes the deficiencies in the FPN structure [15]. By adding bottom-up path and making top-down path trainable, accuracy is improved with a slight subarray increase. The objectives of each proposal 1-3 in More-FPN and overall proposed More-FPN model are shown in Fig. 1(b) and Fig. 1(c), respectively. By combining three proposals, More-FPN achieves both subarray reduction and accuracy improvement to realize the multi-modal AI on NV-CiM.

In addition, in this paper, two major issues of NV-CiM are investigated. The first issue is the trade-off between accuracy and area/energy due to the bit-precision of weight memory cells and DAC/ADC [16, 17, 18]. Appropriate clipping ranges for weights and activations are investigated for the reduction in memory capacity and ADC energy. The

Fig. 1 Proposed Mapping-oriented enhanced FPN (feature pyramid network) fusion (MoreFPN). (a) Overall proposed MoreFPN, an RGB-event fusion object detection model for CiM implementation, with MAGC (Proposal 1), SepBridge (Proposal 2), and TDT-BiFPN (Proposal 3). Objective of (b) proposed modules and (c) MoreFPN. MAGC and SepBridge reduce number of subarrays and memory cells. TDT-BiFPN improves accuracy of object detection.

second issue is non-idealities of NV, such as write variation [19] and conductance shift by data-retention [20, 21]. In this paper, the tolerance against these errors is also verified.

The remainder of this paper is organized as follows. In Chapter 2, methods of each proposal in MoreFPN (Proposal 1: MAGC, Proposal 2: SepBridge, Proposal 3: TDT-BiFPN) are described. In Chapter 3, firstly the configuration of proposed MAGC adopted for MoreFPN is determined with the method to narrow the search space of the number of groups. Then, under the determined MAGC configuration, the effectiveness of each proposal on subarray reduction and accuracy improvement is investigated. In Chapter 4, quantization & clipping (Q&C) of activation and weight in proposed MoreFPN are investigated for the reduction in memory capacity and ADC energy. Additionally, the impact of write variation and data retention error in NV-CiM is investigated.

2. Methods of proposals in MoreFPN

To realize NV-CiM of multi-modal AI, MoreFPN, an RGB-event fusion object detection model, is proposed (Fig. 1). In MoreFPN, MAGC (Section 2.1) and SepBridge (Section 2.2) are adopted for subarray reduction, and TDT-BiFPN (Section 2.3) is adopted for mAP improvement.

2.1 Proposal 1: Mapping-aware Group Convolution (MAGC)

In this section, Mapping-aware Group Convolution (MAGC) is proposed as a subarray-reduction method to reduce the memory capacity in NV-CiM. MAGC utilizes group convolution to reduce the number of weight

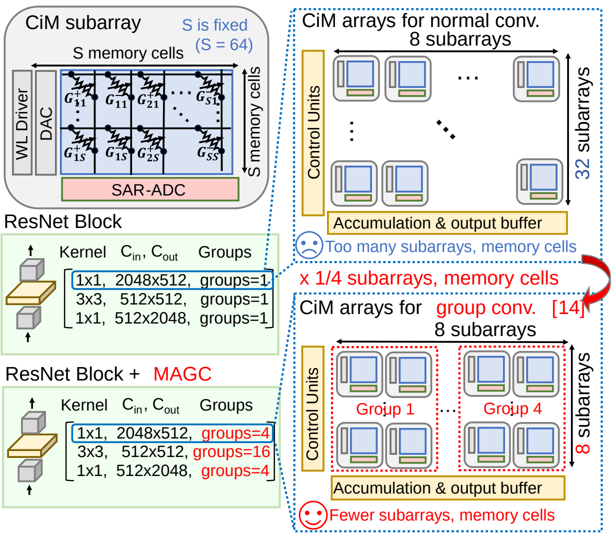


Fig. 2 Impact of group convolution on subarray reduction. Because subarray size (S) is fixed, smaller number of subarrays leads to smaller memory capacity of CiM.

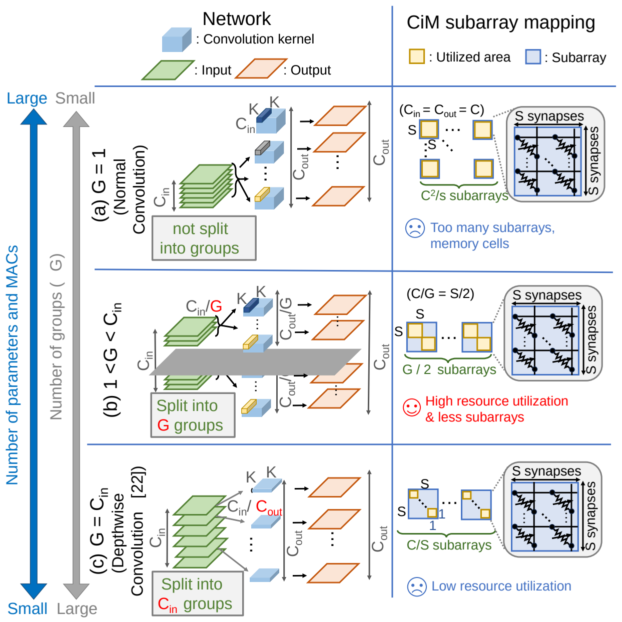


Fig. 3 Network diagram and CiM array mapping of group convolution with (a) G = 1 (Normal convolution), (b) 1 < G < C_{in}, and (c) G = C_s (Depthwise convolution). By utilizing group convolution with G groups, number of subarrays is reduced.

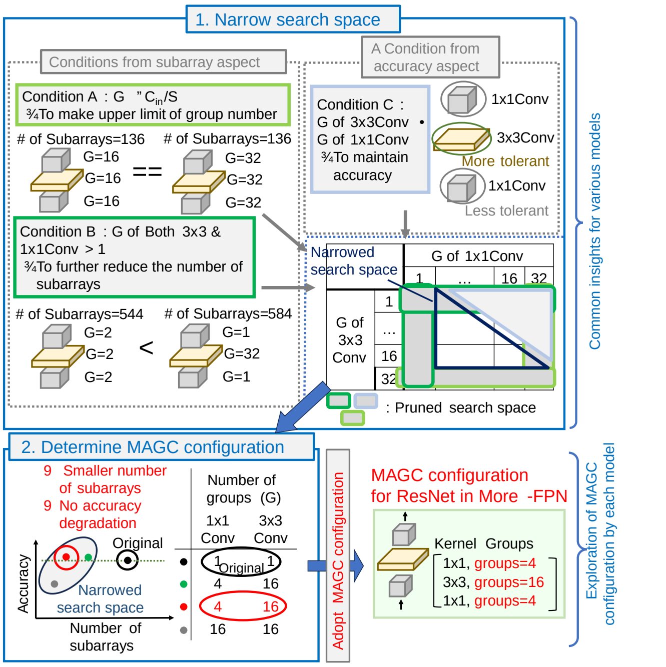


Fig. 4 Proposal 1: MAGC for subarray reduction. Search space of group numbers (G) is narrowed by three conditions (Condition A, B, C). MAGC configuration (i.e., G of 1x1Conv and 3x3Conv) to achieve smaller number of subarrays is determined.

parameters and MACs in the convolutional layer [14, 22, 23, 24]. Fig. 3 shows the network model and the CiM mapping of group convolution. In group convolution, input channels are split into G groups, and convolution calculation is adopted to each group. Group convolution can be treated a normal convolution when G=1 (Fig. 3(a)), and Depthwise convolution [22] when G = C_{in} (Fig. 3(c)). By applying group convolution (Fig. 3(b)), the number of parameters and the number of multiply-accumulate operations (MACs) are reduced by a factor of G. By utilizing group convolution, the number of subarrays in NN-CiM is reduced [14], which leads to the memory capacity reduction (Fig. 2).

There is a trade-off between the group of number and the accuracy. To reduce the number of subarray and memory capacity of CiM, it is desired to increase the number of groups. However, the increase in the number of groups leads to the accuracy degradation. In [23] and [24], it is reported that 3x3 convolutional layer (3x3Conv) is more tolerant to grouping than 1x1 convolutional layers (1x1Conv). Therefore, in proposed MAGC, the number of groups of 1x1Conv and 3x3Conv is investigated separately to make the number of groups large while maintaining accuracy.

Fig. 4 shows the method of proposed MAGC. The configuration about the group number (G) of MAGC is determined according to the following sequence. First, the search space of the group number of 1x1Conv and 3x3Conv is narrowed by Condition A, B, and C. Condition A means that too large group number does not lead to the subarray reduction. With Condition A, meaningless search space for subarray reduction is pruned. Condition B means that both

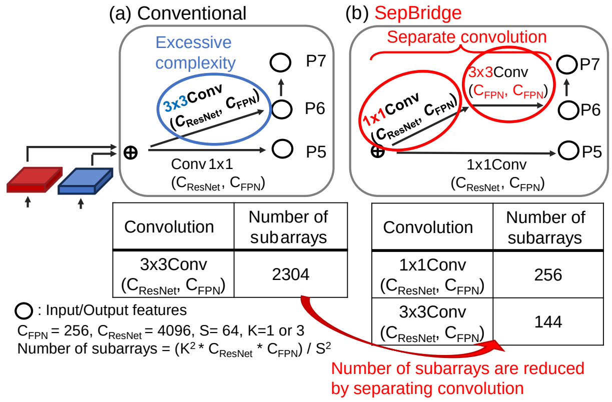


Fig. 5 (a) Conventional convolution layer between ResNet and FPN. (b) SepBridge (Proposal 2). By dividing large 3x3Conv into 1x1Conv and 3x3Conv, number of required subarrays is reduced.

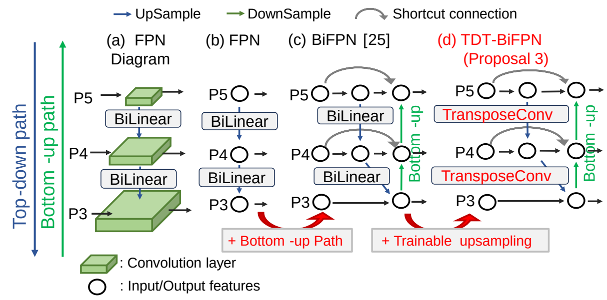


Fig. 6 (a) FPN diagram. Simplified diagrams of (b) conventional FPN, (c) conventional BiFPN, and (d) TDT-BiFPN (Proposal 3). By adopting bottom-up path and trainable transpose convolution (TransposeConv), TDT-BiFPN improves object detection accuracy (i.e., mAP).

the group number of 1x1Conv and 3x3Conv should be larger than 1 for further subarray reduction. With Condition B, search space with less decrease in the number of subarrays is pruned. Condition C means that the group number of 3x3Conv should be larger than that of 1x1Conv. With Condition C, the search space where the accuracy decreases greatly is pruned. With the three conditions, search space is narrowed and the optimal combination of the group number of 1x1Conv and 3x3Conv can be explored at minimal cost of training and inference. Section 3.1 describes how to acquire these conditions.

Second, the optimal MAGC configuration for subarray reduction is selected in the narrowed search space. In the following experiments, the subarray size is fixed to 64 to maintain high utilization rate of CiM while to reducing CiM area.

2.2 Proposal 2: Separable Bridge (SepBridge)

The diagram of Separable Bridge (SepBridge) is shown in Fig. 5. In the proposed More-FPN, the number of channels of ResNet output (C_{ResNet}) is larger than that of Feature Pyramid Network (FPN) input (C_{FPN}). In SepBridge, large 3x3Conv is separated into large 1x1Conv and small 3x3Conv, impressed by separable convolution [22]. With this separation of convolutional layers, the required number of weight parameters for 3x3Conv is significantly reduced. With this separation, the total number of subarrays in CiM is reduced.

2.3 Proposal 3: TDT-BiFPN for accuracy improvement

The diagram of Top-down path trainable BiFPN (TDT-BiFPN, Proposal 3) is shown in Fig. 6. Conventional FPN has several flaws. First, the deep features in conventional FPN are not enhanced because FPN has only top-down path. Therefore, BiFPN [25] is applied to incorporate bottom-up paths and enhance deep features. Second, the semantic gaps between each level of ResNet module are not considered in FPN. To solve this problem, transpose convolution (TransposeConv) is adopted to make top-down paths trainable and narrow the semantic gaps.

3. Evaluation Results of More-FPN

In this chapter, the configuration of MAGC (Proposal 1) is determined first. Then, the subarray reduction is investigated by applying MAGC (Proposal 1) and SepBridge (Proposal 2). Then, the accuracy improvement is evaluated with TDT-BiFPN (Proposal 3).

3.1 Evaluation setup

In this paper, the configuration of datasets is determined with reference to [6, 12]. As a dataset, DSEC is utilized. The object detection labels provided in [6] are utilized and the labels of Car and Pedestrian are used. These labels are automatically annotated by YOLOv5 [26]. Average Precision (AP), with setting the threshold of Intersection over Union (IoU) to 50%, is used as the accuracy of object detection. Same as in [12], mean AP (mAP) indicates the average of the APs of each label. Preprocessing [12] is adopted to RGB frame and event voxel grid to improve the object detection accuracy, mAP.

FPN fusion [6] is utilized as a base model. The backbone of FPN fusion is ResNet-50 [27]. To avoid falling into the local minima and stabilize the training, a warmup cosine learning rate (LR) scheduler with 0.005 of max learning rate is utilized. The epoch number is set to 50.

3.2 Determination configuration and subarray reduction by MAGC (Proposal 1)

To narrow the search space about the number of groups (G) of 1x1Conv and 3x3Conv, the impact of group convolution on accuracy and the number of groups is investigated.

First, with the equation to calculate the number of subarrays (Fig. 7(a)), the correlation between the number of subarrays and G in one layer is investigated as Condition A (Fig. 7(b)). C and S mean the number of channels and subarray size, respectively. In all the assumed cases, the number of subarrays does not decrease when G ≥ C/S. From this insight, Condition A (i.e., G ≤ C/S) is acquired.

Second, the number of subarrays in FPN fusion and mAP are investigated when group convolution is adopted in Fig. 8(a) and Fig. 8(b), respectively. As shown in Fig. 8(a), the number of subarrays becomes smaller when group convolution is applied to both 1x1Conv and 3x3Conv than when group convolution is applied only to 1x1Conv or 3x3Conv. From this insight, Condition B (G of 3x3Conv > 1 and G of 1x1Conv > 1) is acquired to reduce the number of subarrays. As shown in Fig. 8(b), 3x3Conv (red line) keeps higher mAP accuracy than 1x1Conv (black line) with large G. From this insight, Condition C (G of 1x1Conv ≤ G of 3x3Conv) for maintaining the accuracy is acquired. In

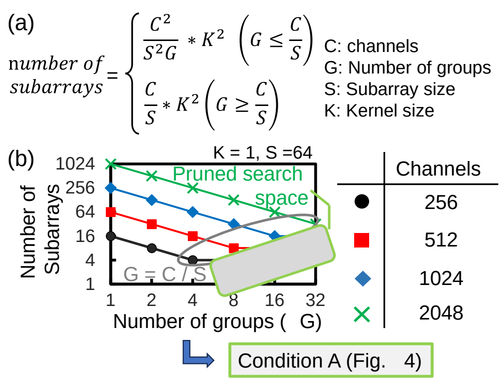


Fig. 7 (a) Equation for calculating number of required subarrays in one layer. (b) Correlation between number of groups (G) and number of subarrays in one layer. Number of subarrays does not decrease when G ≥ C/S (Condition A).

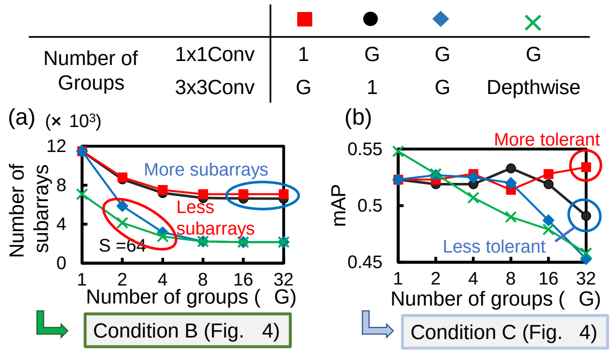
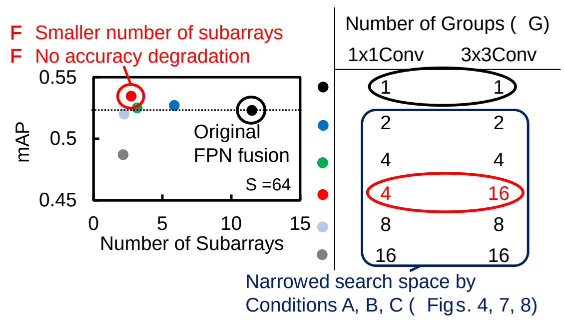


Fig. 8 Correlation between G and (a) number of subarrays and (b) mAP. Applying group convolution to both 3x3Conv and 1x1Conv reduces more subarrays (Condition B). 3x3Conv is more tolerant against group convolution than 1x1Conv (Condition C).



Number of Groups (G) 1x1Conv	3x3Conv
1	1
2	2
4	4
4	16
8	8
16	16

Fig. 9 Number of subarrays and mAP when each combination of 1x1Conv and 3x3Conv is applied to FPN fusion. Search space is narrowed by Conditions A, B, and C. G = 4 and 16 for 1x1Conv and 3x3Conv achieve smaller number of subarrays and no accuracy degradation in mAP.

Table I Reduction in number of parameters, MACs, and CiM subarray by MAGC (Proposal 1)

Model	mAP	Params			MACs[1]			CiM Subarrays[2]		
		All (M)	ResNet (M)	ResNet / All (%)	All (G)	ResNet (G)	ResNet / All (%)	All (K)	ResNet (K)	ResNet / All (%)
FPN fusion	0.523	65.5	46.9	71.7 %	88.3	50.5	57.2 %	16.1	11.5	71.7 %
FPN fusion + Depthwise 3x3Conv	0.548	42.8	24.3	56.0 %	65.8	28.0	42.6%	11.6	7.08	60.9 %
FPN fusion + MAGC (Prop.1)	0.534	26.1	7.50	28.9 %	48.2	9.83	20.8 %	7.29	2.74	37.6 %

-54.7 %

1: MACs = {parameters * (height of output feature) * (width of output feature)}
2: Subarray size (S) of 64 is assumed.

addition, when group convolution with G is adopted to both 3x3Conv and 1x1Conv (blue line and green line), the accuracy degrades more than when group convolution is adopted only to 1x1Conv (black line). This result indicates that adopting group convolution with large G to both 3x3Conv and 1x1Conv is not appropriate for maintaining accuracy. In other words, the appropriate number of groups for 1x1Conv and 3x3Conv should be explored separately to maintain accuracy.

As a result, the search space of groups G is narrowed to satisfy all Conditions A, B, and C (Fig. 4). The requirements to reduce subarrays while maintaining accuracy is to apply group convolution to both 3x3Conv and 1x1Conv, while satisfying that group number of 3x3Conv is larger than that of 1x1Conv.

The optimal number of groups for subarray reduction in FPN fusion is investigated from the narrowed search space (Fig. 9). To achieve the smaller number of subarrays with no mAP accuracy degradation, the optimal choice is found as 3x3Conv with G = 16 and 1x1Conv with G = 4. By adopting proposed MAGC (Proposal 1) with these configurations to FPN fusion, the number of subarrays in the FPN fusion is reduced by 54.7% (Table I). By considering the impact of group convolution on accuracy and subarray reduction, MAGC overcomes the memory capacity issue of multi-modal AI for NV-CiM implementation. Note that Condition A, B, and C in MAGC method can be utilized for various models to narrow the search space and to reduce the number of subarrays.

Table II Reduction in subarrays by MAGC (Proposal 1) and SepBridge (Proposal 2). Base model is FPN fusion.

	MAGC (Prop. 1)	SepBridge (Prop. 2)	mAP	Param (M)	MACs (G)	Number of Subarrays (K)
FPN fusion			0.523	65.5	88.9	16.1
	F		0.534	26.1	48.2	7.29
		F	0.516	57.7	88.5	14.2
	F	F	0.523	17.0	45.0	5.39

-66.5% -54.7% -26.1%

MACs = {params * (height of output feature) * (width of output feature)}

Table III AP improvement and increase in parameters, MACs, and subarrays by BiFPN and TransposeConv in TDT-BiFPN (Proposal 3). Base model is FPN fusion.

	BiFPN	Transpose Conv	AP (Car)	AP (Pedest.)	mAP	Params (M)	MACs (G)	Number of subarrays (K)
FPN fusion			0.688	0.358	0.523	65.4	88.3	16.1
FPN fusion w/ Prop. 3	F		0.732	0.347	0.540	68.4	89.3	16.8
		F	0.703	0.373	0.538	65.9	89.9	16.2
	F	F	0.743	0.389	0.566	69.1	90.9	17.0

+4.3% +5.7%

Table IV mAP improvement and reduction in parameters, MACs, and subarrays by Proposals 1, 2, and 3 in proposed More-FPN.

	MAGC (Prop. 1) SepBridge (Prop. 2),	TDT-BiFPN (Prop.3)	mAP	Params (M)	MACs (G)	Number of subarrays (K)
FPN fusion			0.523	65.4	88.3	16.1
Proposed More-FPN	F		0.523	18.2	47.3	5.39
		F	0.566	69.1	90.93	17.0
	F	F	0.558	21.9	49.8	6.30

+3.5% -66.5% -43.6% -60.9%

3.3 Subarray reduction with MAGC (Proposal 1) and SepBridge (Proposal 2)

In Table II, the impact of MAGC and SepBridge on the subarray reduction is investigated. MAGC reduces the number of subarrays by 54.7% without accuracy degradation. SepBridge further reduces the number of subarrays by 26.1% with a slight accuracy decrease in mAP. As a result, the number of subarrays is reduced by 66.5% in total with MAGC and SepBridge.

3.4 Accuracy improvement with TDT-BiFPN (Proposal 3)

In Table III, the impact of proposed TDT-BiFPN on mAP accuracy improvement is investigated. Both BiFPN and TransposeConv in the proposed TDT-BiFPN effectively improve mAP. By combining BiFPN and TransposeConv, TDT-BiFPN achieves 4.3% mAP improvement with only 5.7% increase in subarrays.

3.5 Subarray reduction and accuracy improvement with Proposals 1-3

Finally, mAP improvement and the reduction in the number of parameters, MACs, and subarrays by proposed modules (Proposals 1, 2, and 3) are investigated (Table IV). The combination of Proposals 1, 2, and 3 reduces parameters, MACs, and subarrays by >66.5%, >43.6%, and >60.9%, respectively, while improving mAP by 3.5%.

4. Evaluation of More-FPN CiM

In this chapter, quantization & clipping (Q&C) of activation and weight in proposed More-FPN are investigated for the reduction in memory capacity and ADC energy of NV-CiM [16, 17, 18]. Additionally, the impact of write variation and data-retention error in NV-CiM is investigated.

4.1 Quantization & Clipping and error injection in NV-CiM

Fig. 10 illustrates Q&C and error injection schemes. The percentile clipping range of activation is predetermined, considering the predetermined upper and lower bounds of ADC/DAC of CiM [28] (Fig. 10(b)). For weight value, zero-centered symmetrical Q&C is applied (Fig. 10(c)), assuming that differential pairs are used to represent weight value (Fig. 10(a)) [28]. Write variation is reproduced by gaussian errors with a standard deviation (σ_{wv}) (Fig. 10(d)), while conductance shift (β) due to data-retention error is replicated by adding a constant value (Fig. 10(e)). The baseline mAP is set to 0.550, which is 0.8% lower than the mAP achieved by More-FPN with 32-bit precision (Table V).

4.2 Evaluation results of More-FPN CiM

Fig. 11 shows the bit-precision sensitivity of activation and weight when different clip range is applied to the proposed More-FPN. As for activation, 0.01% clipping and 8-bit quantization is optimal. As for weights, 3σ clipping and 4-bit quantization is optimal. Fig. 12 shows the error tolerance of the proposed More-FPN and conventional FPN fusion. In this evaluation, weights are quantized to 8-bit with 3σ clipping, not to degrade mAP by quantization. The unit of error size "n.s." stands for normalized step, meaning the relative size to weights normalized between -1 and 1. The results show that the proposed More-FPN (red line) tolerates up to 0.03 n.s. gaussian error (Fig. 12(a)) and 0.003 n.s. shift error (Fig. 12(b)) to keep mAP 0.55. The conventional FPN fusion (blue line) does not achieve the baseline mAP = 0.55 even without errors. According to [19] and [28], the gaussian error with write verify is 0.03 n.s. when the weight is stored in the differential pair of NV-CiM. Because shift error to weights affects the inference result more than gaussian error [29], the error tolerance for shift error is lower than gaussian error. Therefore, the tolerance against write variance of proposed More-FPN is demonstrated.

Table V shows the summary of this paper. In the proposed More-FPN CiM, the number of subarrays is reduced by 61%, compared with FPN fusion CiM. As a result, memory cells and ADC energy are reduced by 61% and 49%, respectively. TDT-BiFPN increases the number of subarrays by 5.7% (Table III); while MAGC and SepBridge decreases the number of subarrays by 66.5% (Table IV). As a result, the three proposals together reduce the number of subarrays by 61%. When write variation with 0.03 n.s. is injected to weights, More-FPN CiM achieves 4.6% higher mAP than FPN fusion CiM. By incorporating Proposal 1, 2, and 3, More-FPN achieves both mAP improvement and reduction of memory cells and energy. This result shows the possibility of realizing multi-modal AI on NV-CiM.

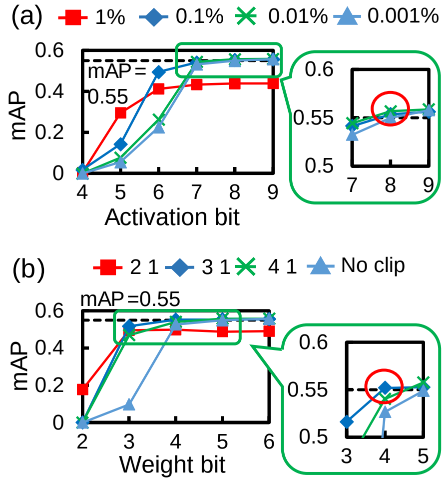


Fig. 11 Low-bit quantization sensitivity of (a) activation and (b) weight with each clip range. 0.01% clipping and 8-bit quantization is optimal for activation. 3σ clipping and 4-bit quantization is optimal for weight.

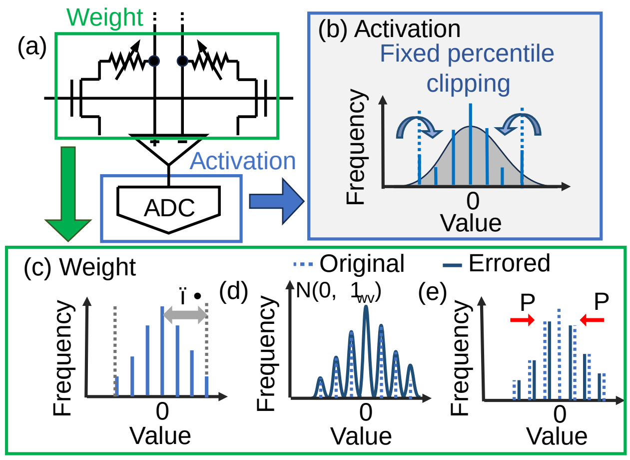


Fig. 10 Quantization and error injection scheme. (a) Weight cell and ADC in CiM circuit. (b) Activation quantization and clipping (Q&C). (c) Weight Q&C. (d) Gaussian error and (e) shift error injection to weight values, respectively.

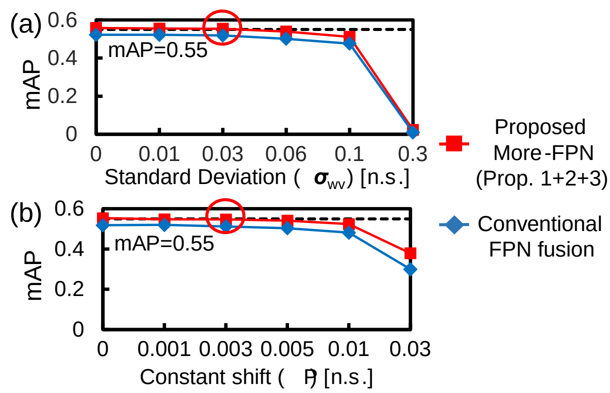


Fig. 12 Error-tolerance when (a) gaussian or (b) shift errors are injected to proposed More-FPN and FPN fusion. More-FPN tolerates up to 0.03 n.s. gaussian error and 0.003 n.s. constant shift.

5. Conclusion

To reduce memory capacity of feature extraction modules in multi-modal AI and realize it on NV-CiM, this paper proposes More-FPN. In More-FPN, three proposals (MAGC, SepBridge, and TDT-BiFPN) are adopted. MAGC (Proposal 1) is a subarray reduction algorithm to reduce the memory capacity in NV-CiM. By adopting MAGC to FPN fusion, a 54.7% reduction in required number of subarrays is achieved. SepBridge (Proposal 2) achieves further 26.1% subarray reduction from MAGC-adopted FPN fusion. With MAGC and SepBridge (Proposal 1 and 2), the memory cells and ADC energy are reduced by 61% and 49% compared with conventional FPN fusion CiM, respectively. Moreover, TDT-BiFPN (Proposal 3) in More-FPN achieves a 4.6% improvement in mAP when considering write variation. These results show the possibility of realizing RGB-event fusion multi-modal AI on edge NV-CiM. Proposed method in this paper is a basic study on using multi-modal data with Large Language Model (LLM) and Transformer.

Acknowledgments

This paper is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

Table V Comparison between CiMs of each model

		FPN fusion CiM [6]	More-FPN CiM (proposed)	
CiM Configuration	Weight bit -precision	4-bit (Fig. 11)		
	Weight clipping	1) L11)		
	I/O bit-precision	8-bit (Fig. 11)		
	I/O clipping	0.01% (Fig. 11)		
	Subarray size (S)	64		
mAP	w/o error	0.504	0.547	+4.6%
	w/ Write variation (gauss σ_{wv} = 0.03 n.s.)	0.500	0.546	
	w/ Write variation & Data retention error (Const Δ = 0.003 n.s.)	0.491	0.540	
CiM Performance	Number of subarrays	16,098	6298	
	Number of memory cells considering subarrays [1]	131M	51.6M	-61%
	ADC Area (Normalized)	1	0.391	
	ADC Energy [2] (Normalized)	1	0.564	-49%

1: Number of memory cell = $2S^2 \times (number\ of\ subarrays)$
2: ADC Energy ∝ Number of MACs

References

[1] N. Verma et al., "In-Memory Computing: Advances and Prospects," IEEE Solid-State Circuits Magazine (SSCM), vol. 11, no. 3, pp. 43-55, 2019.
[2] N. Lepri et al., "In-memory computing for machine learning and deep learning," IEEE Journal of the Electron Devices Society, vol. 11, pp. 587-601, 2023.
[3] J. Lin and F. Zhang, "R3LIVE: A Robust, Real-time, RGB-colored, LiDAR-Inertial-Visual tightly-coupled state Estimation and mapping package," IEEE International Conference on Robotics and Automation (ICRA), 2022, pp. 10672-10678.
[4] Z. Wu et al., "Robust RGB-D Fusion for Saliency Detection," International Conference on 3D Vision (3DV), 2022, pp. 403-413.
[5] Z. Zhou et al., "RGB-Event Fusion for Moving Object Detection in Autonomous Driving," IEEE International Conference on Robotics and Automation (ICRA), 2023, pp. 7808-7815.
[6] A. Tomy et al., "Fusing Event-based and RGB camera for Robust Object Detection in Adverse Conditions," IEEE International Conference on Robotics and Automation (ICRA), 2022, pp. 933-939.
[7] T. Finateu et al., "A 1280×720 Back-Illuminated Stacked Temporal Contrast Event-Based Vision Sensor with 4.86µm Pixels, 1.066GEPS Readout, Programmable Event-Rate Controller and Compressive Data-Formatting Pipeline," IEEE International Solid-State Circuits Conference (ISSCC), 2020, pp. 112-114.
[8] G. Gallego et al., "Event-Based Vision: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 44, no. 1, pp. 154-180, 2022.
[9] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128 X 128 120db 30mw asynchronous vision sensor that responds to relative intensity change", IEEE International Solid-State Circuits Conference (ISSCC), 2006, pp. 2060-2069.
[10] J. Han et al., "ERA-LSTM: An Efficient ReRAM-Based Architecture for Long Short-Term Memory", IEEE Transactions on Parallel and Distributed Systems, vol. 31, no. 6, pp. 1328-1342, 2020.
[11] C.-X. Xue et al., "A 22nm 2Mb ReRAM Compute-in-Memory Macro with 121-28TOPS/W for Multibit MAC Computing for Tiny AI Edge Devices", IEEE International Solid-State Circuits Conference (ISSCC), 2020, pp. 244-246.
[12] Y. Ichikawa, A. Yamada, N. Misawa, C. Matsui, K. Takeuchi, "REM-CiM: Attentional RGB-Event Fusion Multi-modal Analog CiM for Area/Energy-efficient Edge Object Detection during both Day and Night", IEICE Transactions on Electronics, 2024 (to be published).
[13] X. Peng, R. Liu, and S. Yu, "Optimizing Weight Mapping and Data Flow for Convolutional Neural Networks on Processing-in-Memory Architectures", IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 67, no. 4, pp. 1333-1343, 2020
[14] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Networks", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5987-5995.
[15] C. Guo, B. Fan, Q. Zhang, S. Xiang, and C. Pan, "AugFPN: Improving Multi-Scale Feature Learning for Object Detection", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 12592-12601.
[16] P. Chen et al., "RIMAC: An Array-level ADC/DAC-Free ReRAM-Based In-Memory DNN Processor with Analog Cache and Computation," 28th Asia and South Pacific Design Automation Conference (ASP-DAC), 2023, pp. 228-233.
[17] S. Yu et al., "Compute-in-Memory with Emerging Nonvolatile-Memories: Challenges and Prospects," IEEE Custom Integrated Circuits Conference (CICC), 2020, pp. 1-4.
[18] H. Jiang et al., "A 40nm Analog-Input ADC-Free Compute-in-Memory RRAM Macro with Pulse-Width Modulation between Sub-arrays," IEEE Symposium on VLSI Technology and Circuits, 2022, pp. 266-267.

[19] R. Mochida et al., "A 4M Synapses integrated Analog ReRAM based 66.5 TOPS/W Neural-Network Processor with Cell Current Controlled Writing and Flexible Network Architecture," IEEE Symposium on VLSI Technology, 2018, pp. 175-176.
[20] S. Fukuyama et al., "Comprehensive Analysis of Data Retention and Endurance Trade-Off of 40nm TaOx-based ReRAM," IEEE International Reliability Physics Symposium (IRPS), 2019, pp. 1-6.
[21] Y.-H. Lin et al., "Performance Impacts of Analog ReRAM Non-ideality on Neuromorphic Computing," IEEE Transactions on Electron Devices, vol. 66, no. 3, pp. 1289-1295, 2019.
[22] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017
[23] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 6848-6856.
[24] G. Huang, S. Liu, L. v. d. Maaten, and K. Q. Weinberger, "CondenseNet: An Efficient DenseNet Using Learned Group Convolution," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 2752-2761.
[25] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10778-10787.
[26] Ultralytics, Yolov5. [Online]. Available: https://github.com/ultralytics/yolov5
[27] K. He et al., "Deep Residual Learning for Image Recognition," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.
[28] A. Yamada, N. Misawa, C. Matsui, and K. Takeuchi, "LIORAT: NN Layer I/O Range Training for Area/Energy Efficient Low-Bit A/D Conversion System Design in Error Tolerant Computation-in-Memory," IEEE/ACM International Conference on Computer Aided Design (ICCAD), 2023, pp. 1-9.
[29] K. Higuchi, C. Matsui, N. Misawa, and Ken Takeuchi, "Comprehensive Computation-in-Memory Simulation Platform with Non-volatile Memory Non-Ideality Consideration for Deep Learning Applications," International Conference on Solid State Devices and Materials (SSDM), 2021, pp. 1-2.

Yuya Ichikawa Received the B.S degree in Information and Communication Engineering from the University of Tokyo in 2022. He is now a master course student in Takeuch Laboratory in the department of Electrica Engineering and Information Systems, the University of Tokyo. His current resea interests include RGB-event fusion multimodal AI and Computation-in-Memory system.

Naoko Misawa Received the M.S. degr from Imperial College London in 2012. Sh currently an academic staff in Takeuch Laboratory in the department of Electr Engineering and Information Syste Graduate School of The University of Tok Her research interests include emerging- volatile memories, neuromorphic comput and Vision Transformer.

Chihiro Matsui is currently a Proje Associate Professor in the Departmen Electronics Engineering and Informat Systems, Graduate School of Engineerin The University of Tokyo. Her research inte includes system, circuit, and device design with emerging non-volatile memories fo enterprise applications. She earned her and M.S. degrees in Physics from Ochano University, Tokyo, Japan, in 2003 and 20 respectively, and her Ph.D. degree Information Security Sciences from Chuo University, Tokyo, Japan, in 2018. She wa Project Assistant Professor of Research Development Initiative at Chuo Univers from 2018 to 2020 and a Project Assis Professor in the Department of Electro Engineering and Information System Graduate School of Engineering of University of Tokyo from 2020 to 2023

Ken Takeuchi is currently a Professor Department of Electrical Engineering a Information Systems, Graduate School Engineering of The University of Tokyo. He now working on data-centric computing su as computation in memory, approxim computing, data-center scale computing, A chip design and brain-inspired memory. H received the B.S. and M.S. degrees in Ap Physics and the Ph.D. degree in Ele Engineering from The University of Tokyo 1991, 1993 and 2006, respectively. In 2003 also received the M.B.A. degree from Stan University. Since he joined Toshiba in 1993 had been leading Toshiba's NAND fl memory circuit design for fourteen years. was an Associate Professor at Departme Electrical Engineering and Informati Systems, Graduate School of Engineerin The University of Tokyo from 2007 till 201 He was a Professor at Department of Elect Electronic and Communication Engineer Faculty of Science and Engineering of C University from 2012 till 2020. In 2020, h rejoined The University of Tokyo. He design six world's highest density NAND fla memory products such as 0.7um 16M 0.4um 64Mbit, 0.25um 256Mbit, 0.16u 1Gbit, 0.13um 2Gbit and 56nm 8Gbit NAN flash memories. He holds 228 pat worldwide including 124 U.S. pater Especially, with his invention, "multipage c architecture", presented at Symposium VLSI Circuits in 1997, he successfu commercialized world's first multi-level cel NAND flash memory in 2001. He has autho numerous technical papers, one of which the Takuo Sugano Award for Outstand Paper at ISSCC 2007. He is serving as program chair of Asian Solid-State Circuit Conference (A-SSCC) in 2023. He served the symposium chair/co-chair of Symposium on VLSI Circuits in 2021/2020. He served the program chair/co-chair of Symposium o VLSI Circuits in 2019/2018. He has a served on the program committee membe International Solid-State Circuits Conferen (ISSCC), Custom Integrated Circu Conference (CICC), Asian Solid-State Circuit Conference (A-SSCC), International Memo Workshop (IMW), International Conference Solid State Devices and Materials (SSDM) Non-Volatile Memory Technology Symposiu (NVMTS). He served as a tutorial speake ISSCC 2008, forum speaker at ISSCC 2 SSD forum organizer at ISSCC 2009, 3D-I forum organizer at ISSCC 2010, Ultra-low voltage LSI forum organizer at ISSCC 2 and Robust VLSI System forum organize ISSCC 2012.