A 0.8 V Intelligent Vision Sensor With Tiny Convolutional Neural Network and Programmable Weights Using Mixed-Mode Processing-in-Sensor Technique for Image Classification

Tzu-Hsiang Hsu, Guan-Cheng Chen, *Member, IEEE*, Yi-Ren Chen, Ren-Shuo Liu, *Member, IEEE*, Chung-Chuan Lo, Kea-Tiong Tang, *Senior Member, IEEE*, Meng-Fan Chang, *Fellow, IEEE*, and Chih-Cheng Hsieh, *Senior Member, IEEE*

***Abstract*— This article presents an intelligent vision sensor (IVS) with embedded tiny convolutional neural network (CNN) model and programmable processing-in-sensor (PIS) circuit for real-time inference applications of low-power edge devices. The proposed imager realizes the full computing functions of a customized three-layers tiny network, which includes a 3 × 3 convolution layer (stride = 3) with activation function of rectified linear unit (ReLU), a 2 × 2 maximum pooling (MP) layer (stride = 2), and a 1 × 1 fully connected (FC) layer for inference. A 0.8 V 128 × 128 IVS prototype was fabricated and verified in TSMC 0.18 μm standard CMOS technology. In normal image mode, it consumed 76.4 μW with full-resolution (126 × 126 active resolution) image output at 125 f/s. In CNN mode, it consumed 134.5 μW at 250 f/s and an achieved iFoMs of 33.8 pJ/pixel·frame. Using the proposed mixed-mode PIS circuits, the prototype is configured to demonstrate a "human face or not detection" task with an achieved accuracy of 93.6%.**

***Index Terms*— Artificial intelligent (AI), convolutional neural network (CNN) CMOS image sensor (CIS), face detection (FD), feature extraction, intelligent vision sensor (IVS), processing-in-sensor (PIS).**

I. Introduction

RECENTLY, the research of low-power and energy-efficient CMOS image sensor (CIS) has been developed and widely used in edge devices. While more advanced applications, such as dynamic vision sensor [1], computational CIS [2], [3], [4], and object detection and classification sensor [5], [6], [7], [8], [9], [10], [11], [12], are in growing demand due to the blooming development and success in artificial intelligent (AI). However, the CIS plus dedicated AI accelerator solution [8] suffers from the burdens of power and latency caused by the raw image data traffic between the imager and the companion signal processor with a neural network accelerator, making it unsuitable for the real-time inference in low-power edge devices. Consequently, imagers with near- or in-sensor processing capability has been recently developed to improve the system efficiency for specific applications. In [4], a convolutional CIS with near-sensor analog multiply-accumulate (MAC) operations was reported for assisting with the first layer computations of a convolutional neural network (CNN). However, the convolutional CIS is inadequate for some tasks, due limits on the numbers of layers/kernels, and needs a companion digital accelerator for the required operations such as rectified linear unit (ReLU), maximum-pooling (MP), fully connected (FC) layer, etc., of a complete CNN model. In [7], an analog convolutional CIS is reported with a five-layer network for CNN implementation. However, the analog MAC operations using charge sharing with a capacitor array leads to gain loss, low weight resolution, and limited accuracy. Moreover, the ReLU with MP operation using a static winner-take-all circuit is power hungry. In [9], [10], and [11], the near-sensor Haar-like filtering operations are implemented in imagers to realize face detection (FD). However, unlike using CNNs with programmable weights for different tasks, the implemented features of such prior works are limited and not configurable. To address these issues, we present an intelligent vision sensor (IVS) with an embedded tiny CNN model and programmable weights to achieve configurable feature extraction and on-chip image classification using a mixed-mode processing-in-sensor (PIS) technique. This article presents the detailed implementation of the proposed IVS operating with a 0.8 V ultralow supply voltage in 0.18 μm standard process. The proposed pixel circuit adopts the threshold-

Manuscript received 6 June 2022; revised 15 January 2023 and 24 March 2023; accepted 10 June 2023. This article was approved by Associate Editor David Stoppa. This work was supported in part by Qualcomm through the Taiwan University Research Collaboration Project; in part by the NOVATEK Fellowship; in part by the Taiwan Semiconductor Research Institute (TSRI); and in part by the Ministry of Science and Technology (MOST), Taiwan, under Contract 108-2218-E-007-022, Contract 108-2622-8-007-017, and Contract 109-2218-E-007-020. *(Corresponding author: Chih-Cheng Hsieh.)*

Tzu-Hsiang Hsu is with Mediatek Inc., Hsinchu 30078, Taiwan.

Guan-Cheng Chen, Ren-Shuo Liu, Kea-Tiong Tang, Meng-Fan Chang, and Chih-Cheng Hsieh are with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan (e-mail: cchsieh@ee.nthu.edu.tw).

Yi-Ren Chen is with Phison Electronics Corporation, Taoyuan 35059, Taiwan.

Chung-Chuan Lo is with the Institute of Systems Neuroscience, National Tsing Hua University, Hsinchu 30013, Taiwan.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSSC.2023.3285734.

Digital Object Identifier 10.1109/JSSC.2023.3285734

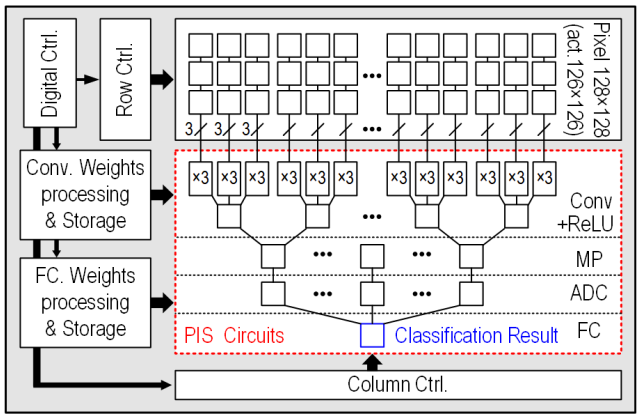


Fig. 1. Block diagram of the proposed IVS.

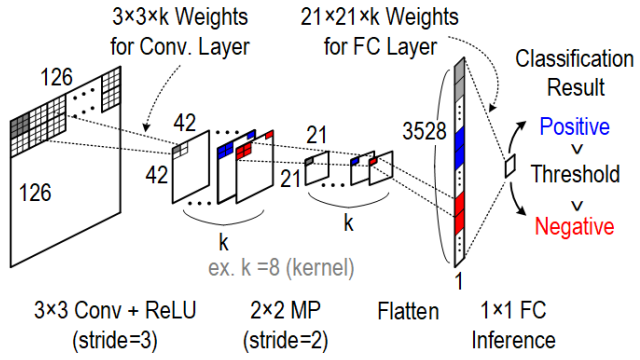


Fig. 2. Customized three-layer tiny CNN.

voltage-canceling (TVC) pulsewidth-modulation (PWM) pixel structure [13] to alleviate the threshold voltage variation issue. The mixed-mode PIS circuit, including switch-current and integration (SCI) and polarity judging circuit, realizes an all-in-one operation for convolution with ReLU and MP computation. Moreover, taking advantage of the time-domain characteristic of the intermediate MP output signal, the digital multiply-accumulation (MAC) operation of the FC layer can be efficiently realized by the customized logic operation without needing any digital multiplier and adder.

The rest of this article is organized as follows: Section II describes the overall system architecture. The detailed circuit implementation and operation of each block are explained in Section III. Section IV presents the experimental results of the prototype IVS in different operation modes. Finally, the conclusion is given in Section V.

II. System Architecture

Fig. 1 shows the block diagram of the proposed IVS. The key building blocks consist of a 126 × 126 PWM pixel array and the column-parallel PIS circuit for model computation. The peripheral supporting circuitry includes processing and storage circuit for convolution and FC weights as well as the row- and column-wise control. Fig. 2 provides the customized three-layer tiny network. It consists of a 3 × 3 convolution (Conv) and ReLU layer (stride = 3) with adjustable kernel number for feature map (FM) computation, a 2 × 2 MP layer (stride = 2) for down sampling, and a 1 × 1 FC layer

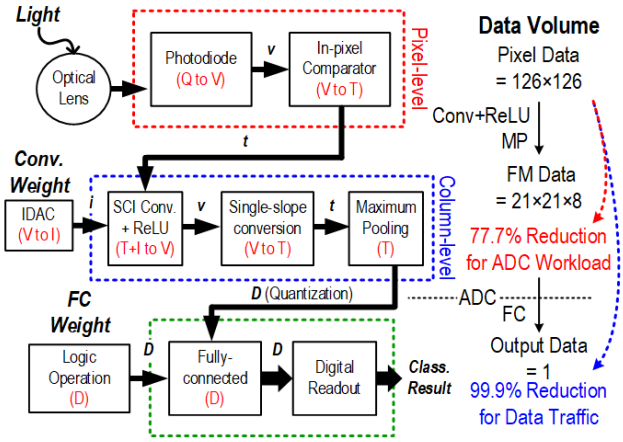


Fig. 3. Data processing flow and corresponding data volume.

for inference. The data processing can be accomplished by proposed mixed-mode PIS circuits, which can support the trainable parameters, including ±3-bit (−8 to 8) convolution weight and 1.5-bit (−1, 0, 1) ternary FC weight. Using only eight kernels, the network can execute a binary classification task, human FD, with trained accuracy of 97.26%.

Fig. 3 illustrates the data processing flow and the corresponding data volume in each stage. In the customized network for binary classification, the kernel number is designed to be $k = 8$. To evaluate the data reduction ratio from pixel to output level, we can assume 126×126 of pixel data is shrunk to $21 \times 21 \times 8$ of MP data through executing the Conv with ReLU and MP operation before ADC. Accordingly, the data reduction ratio of 77.7% can be achieved and depends on the stride value of Conv and MP operation and the number of kernels. The FM data are quantized for the digital FC layer to output a 1-bit classification result judging by a trained threshold. Also, the configurable Conv and FC weights are realized by global current-mode digital-to-analog converter (IDAC) and logic operation, respectively. Taking advantage of the PIS technique, the analog computation can effectively reduce 77.7% of the ADC workload for power saving. Moreover, compared to the raw image output, 99.9% of data traffic is reduced for energy and bandwidth saving. Consequently, the proposed IVS solution provides a configurable, low-power, and low-latency solution for a wide variety of CNN-based image classification tasks.

III. Circuit Implementation and Operation

A. Analog PIS Circuit

Fig. 4 shows the structure and basic operation of the adopted 4T PWM pixel [4], [13], [14] and which is characterized by its energy efficiency and signal robustness for PIS operation. Once the pixel is being selected by $\mathrm{PA}_{\mathrm{ROW}}\langle X \rangle$, 8 μs for ramping VR is designed for the voltage-to-pulsewidth conversion which is defined as T_{CONV}. The signal-dependent pulsewidth T_{PW} is then readout to the column circuit through the MRD and which is represented by $\mathrm{PW}\langle m \rangle$. Fig. 5 shows the column-parallel analog PIS circuit for the Conv, ReLU, and MP layers. By adopting the PWM pixel and SCI

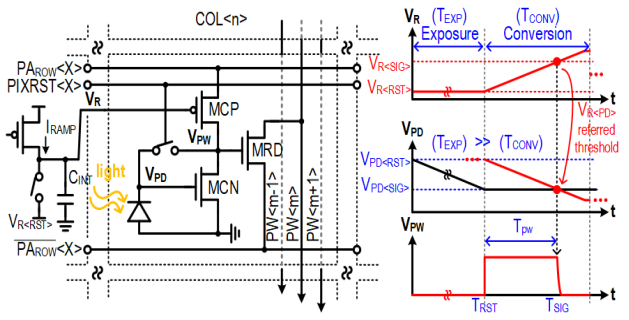


Fig. 4. Structure and basic operation of adopted PWM pixel [4].

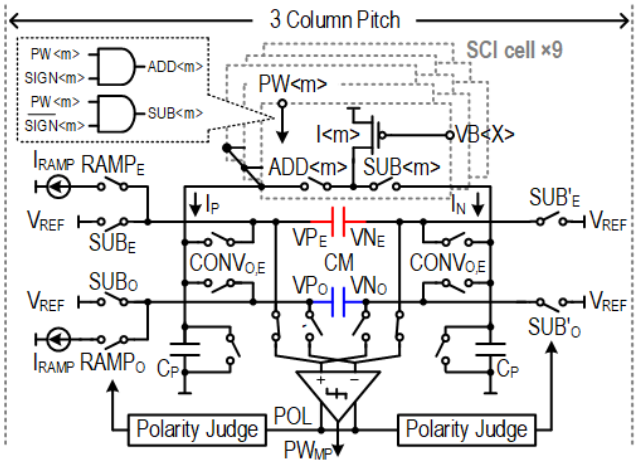


Fig. 5. Column-parallel analog PIS circuit for Conv, ReLU, and MP.

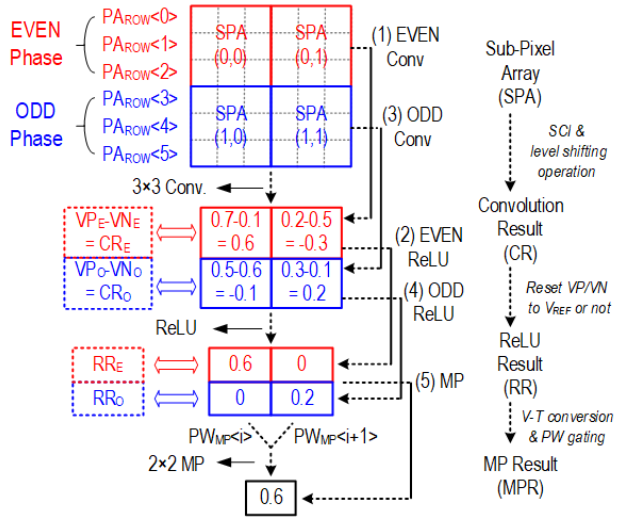


Fig. 6. Example of data computation flow.

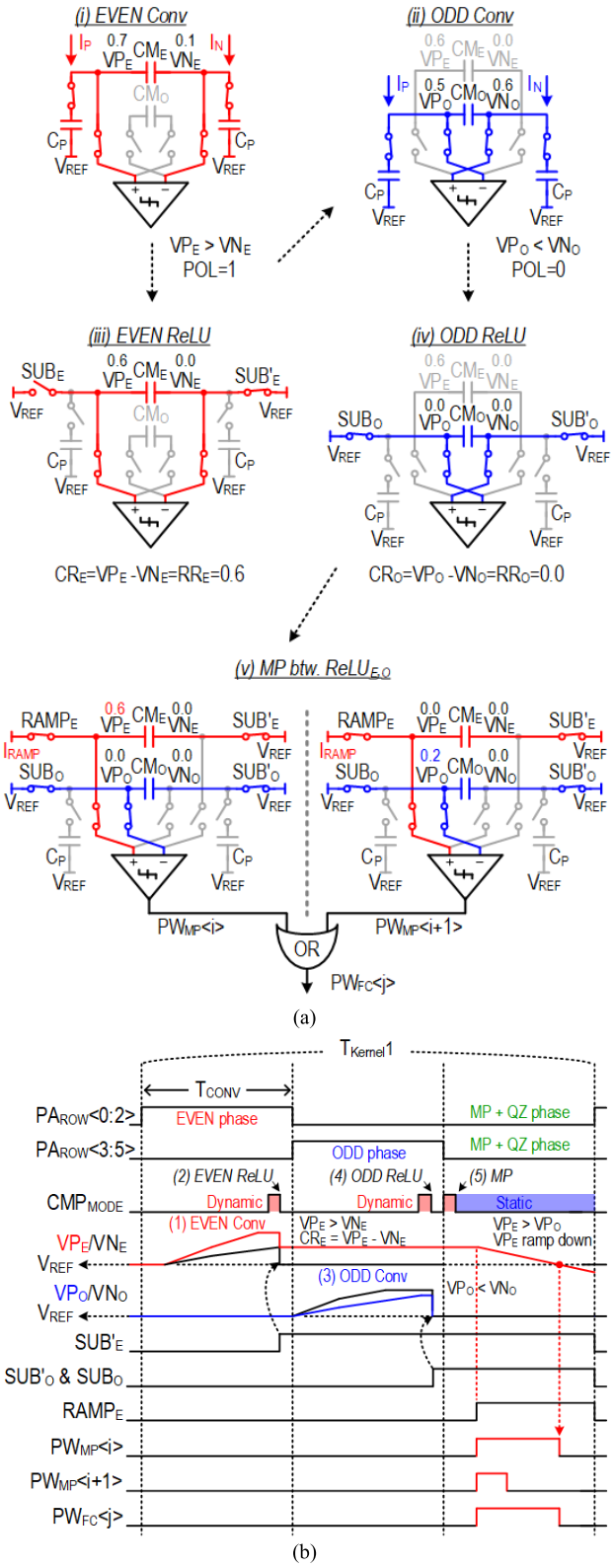


Fig. 7. Example of (a) operating sequence and (b) timing diagram for the Conv with ReLU and MP.

concept [4], [15], the Conv operation is realized using the signal-dependent pulsewidth (PW$\langle m\rangle$) and weight-dependent current level ($I\langle m\rangle$). The PW$\langle m\rangle$ is first gated to be the signal ADD$\langle m\rangle$ or SUB$\langle m\rangle$ according to the positive or negative weight value, SIGN$\langle m\rangle$, and which will enable the current $I\langle m\rangle$. The signal $I\langle m\rangle$ is generated by the current mirror and its bias VB$\langle X\rangle$ is from a 3-bit IDAC. To simultaneously provide nine different weight-dependent currents for a 3×3 kernel, an overall nine sets of IDAC are implemented. When executing the SCI convolution, positive and negative weighted current I_P and I_N will integrate on VP and VN, respectively. The polarity judge circuit is realized by

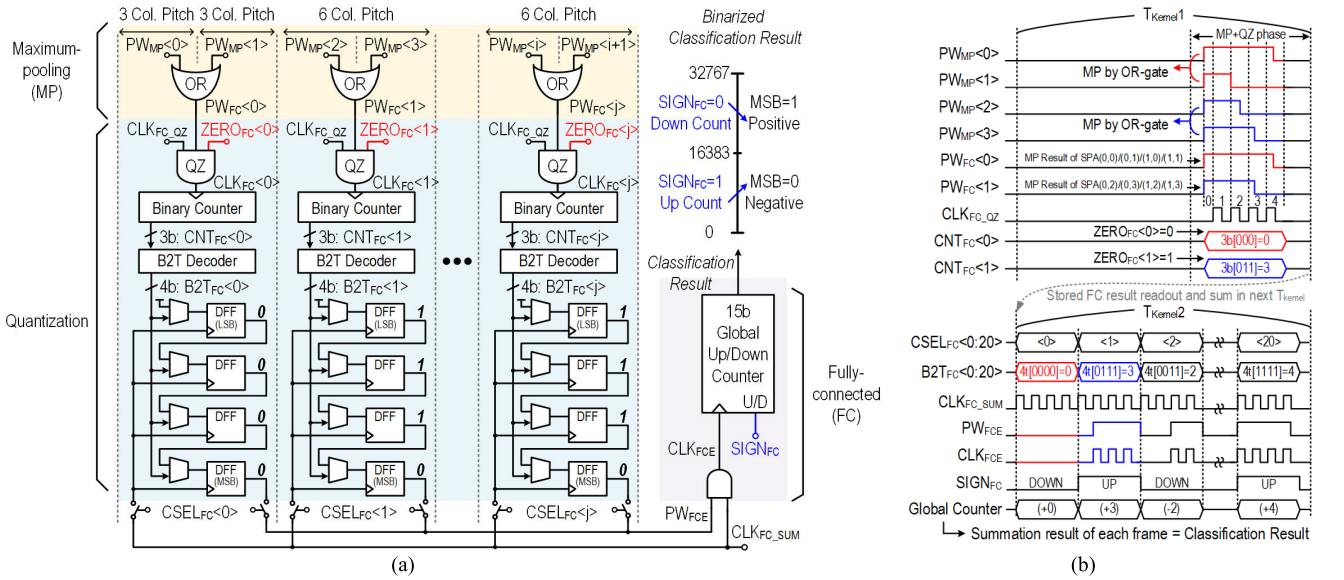


Fig. 8. (a) Digital PIS implementation of the second step of the MP operation and the FC layer. (b) Corresponding timing diagram.

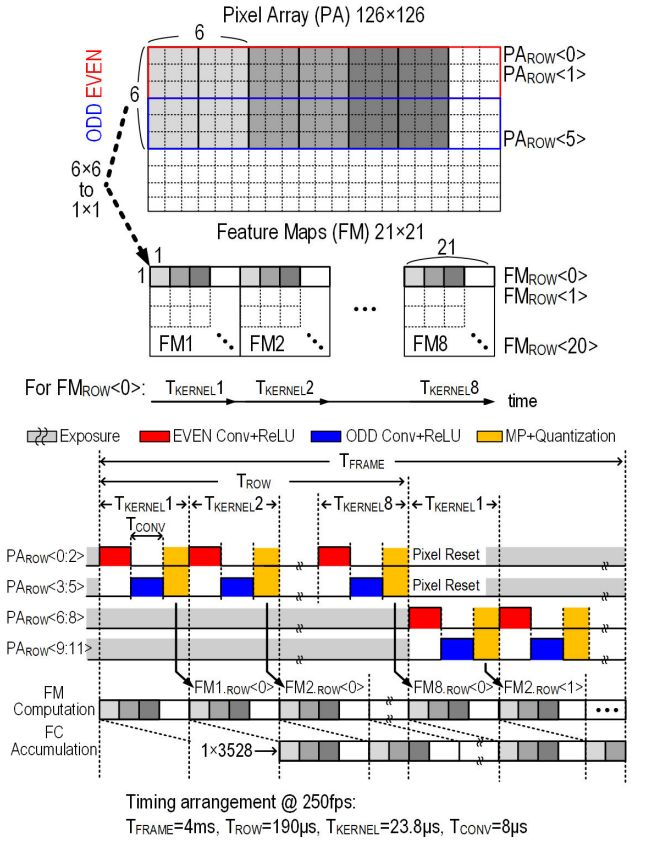


Fig. 9. Timing diagram of the implemented PIS operations for multi-kernel CNN processing.

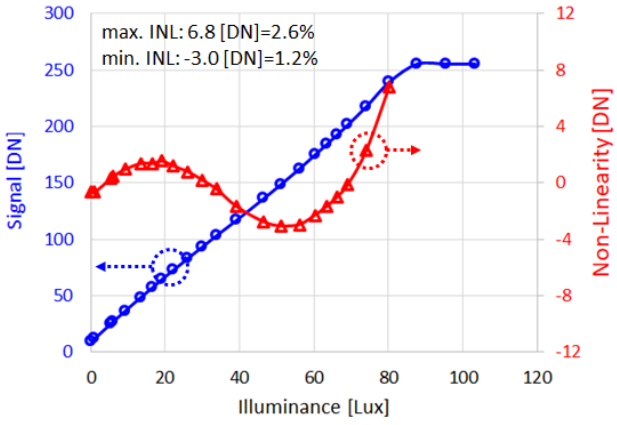


Fig. 10. Signal transfer curve and corresponding nonlinearity.

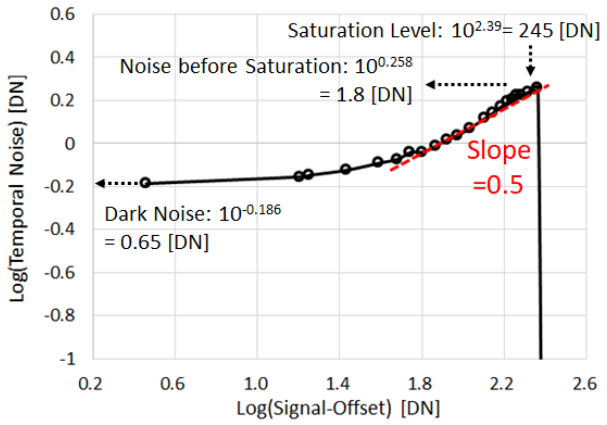


Fig. 11. Noise performance in normal image mode.

combinational logic and flip-flop, which checks the comparator output (POL) and delivers specific control signals which include $\mathrm{RAMP}_{E/O}$ and $\mathrm{SUB}_{E/O}$.

An example of operation flow is illustrated in Fig. 6. To obtain the neighboring SPA convolution results (CRs) for MP operation with low latency and without using memory, the even/odd phase is utilized for accessing $\mathrm{PA}_{\mathrm{ROW}}\langle n+2{:}n\rangle$ and $\mathrm{PA}_{\mathrm{ROW}}\langle n+5{:}n+3\rangle$, respectively. As a result, the 3×3 Conv and the following 2×2 MP can be executed efficiently. For a 3×3 sub-pixel array (SPA), the CR is realized by checking the comparison result between current-integrated signals (VP and VN), which are SCI result of SPA within

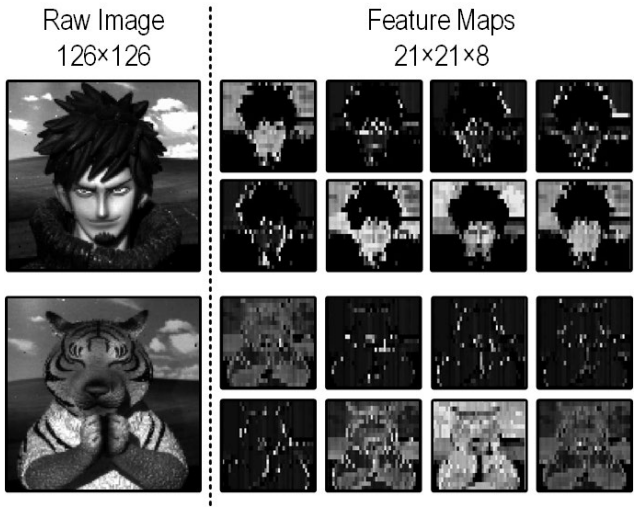


Fig. 12. Captured (a) raw images and (b) FMs computed from eight different kernels.

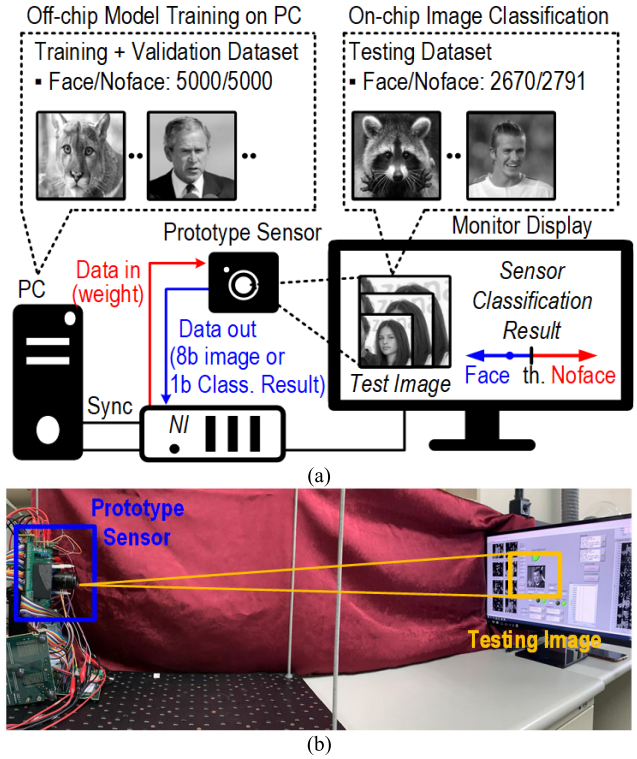


Fig. 13. (a) Experimental setup for (b) evaluating 5400 images in testing dataset.

three successive row of pixel array (PA_{ROW}) with its own positive and negative weight, and followed by the corresponding level shifting operation on capacitor CM for signal subtraction (CR = VP − VN). Then, the ReLU result (RR) can be realized by connecting VP and VN to V_{REF} to set CR = 0 when POL = 0 (VP < VN) or keeping the level shifting result when POL = 1 (VP > VN). Afterward, the MP result (MPR) is realized by the V-T conversion and pulsewidth gating operation.

A step description of the Conv with ReLU and MP operations is shown in Fig. 7(a). In the (i) EVEN Conv

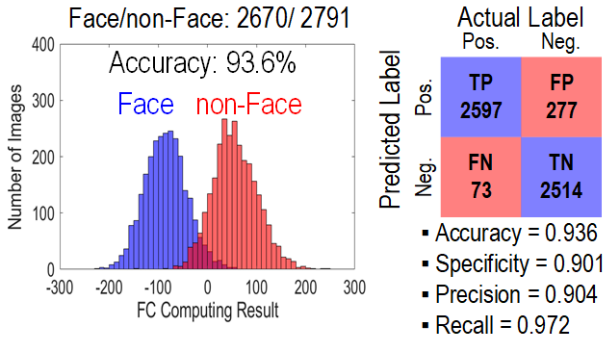


Fig. 14. Statistic distribution of classification result.

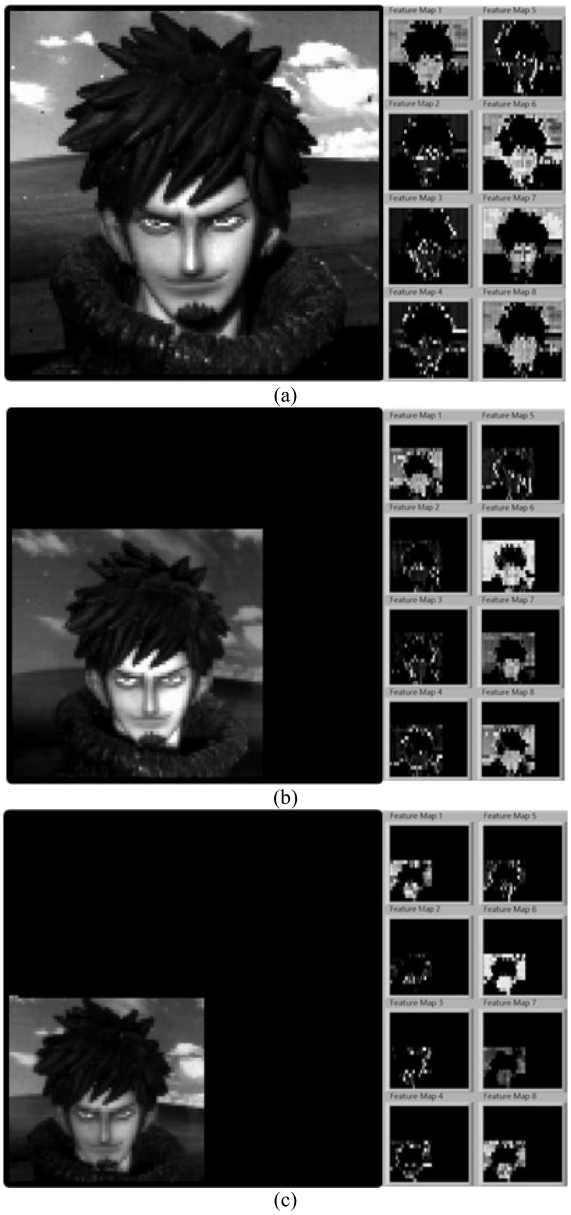


Fig. 15. Captured image and FMs with the detection windows of (a) 126 × 126, (b) 84 × 84, and (c) 66 × 66.

phase, the positive/negative SCI results of SPA (0, 0) are assumed to be $VP_E = 0.7$ and $VN_E = 0.1$ with the polarity POL = 1. Then in the (ii) EVEN ReLU phase, VN_E is

TABLE I
MEASURED PERFORMANCE IN DIFFERENT DETECTION WINDOW

Detection Window	126×126	84×84	66×66
Number of FC Parameters	21×21×8	14×14×8	11×11×8
Meas. Accuracy (%)	93.6	91.6	90.1
Max. Frame Rate (fps)	250	372	473
Power (μW)	134.5	100.6	88.5
FoM (pJ/pix·frame)	33.8	38.3	42.9

connected to the initial voltage of integration (V_{REF}) as a level shift to get the CONV and the RR ($\mathrm{CR}_E = \mathrm{VP}_E – \mathrm{VN}_E = \mathrm{RR}_E = 0.6$) on node VP_E. Conversely, in the (iii) ODD CONV phase, SCI results of SPA (1, 0) are assumed to be $\mathrm{VP}_O = 0.5$ and $\mathrm{VN}_O = 0.6$ with the polarity $\mathrm{POL} = 0$. Then in the (iv) ODD ReLU phase, the CR_O is reset to zero by switching both nodes ($\mathrm{VP}_O/\mathrm{VN}_O$) to $\mathrm{V}_{\mathrm{REF}}$ to get the $\mathrm{RR}_O = 0$ (on node VP_O). The MP function of 2×2 SPAs is realized using a two-step operation. The (v) first step of MP is to find the maximum RR value between SPA (0, 0) and (1, 0) by again checking the POL result of RR_E and RR_O using the same comparator. In this case with $\mathrm{RR}_E = 0.6$ and $\mathrm{RR}_O = 0$, the node (VP_E) with a higher level will be converted to a signal-dependent pulsewidth $\mathrm{PW}_{\mathrm{MP}}$ (V-to-T conversion) by ramping it down using a current source I_{RAMP}. The corresponding timing diagram is illustrated in Fig. 7(b). With the column-parallel architecture, the maximum values of SPA (0, 0)/(1, 0) and SPA (0, 1)/(1, 1) are converted to two pulses ($\mathrm{PW}_{\mathrm{MP}}\langle 0\rangle/\mathrm{PW}_{\mathrm{MP}}\langle 1\rangle$) simultaneously. Afterward, the maximum value among SPA (0, 0), (0, 1), (1, 0), and (1, 1) can be easily obtained because the maximum pulsewidth ($\mathrm{PW}_{\mathrm{FC}}\langle j\rangle$) between $\mathrm{PW}_{\mathrm{MP}}\langle i\rangle$ and $\mathrm{PW}_{\mathrm{MP}}\langle +1\rangle$ can be obtained using a two-input OR gate. Compared to the reported analog Conv approaches [4], [7], the proposed PIS technique in the time domain pulsewidth for Conv, ReLU, and MP operations are efficient and robust with a minimal area and power overhead.

B. Digital PIS Circuit

Fig. 8 shows the digital PIS implementation of the FC layer. For the digital MAC operations of the FC layer, each input value will be multiplied by a trained FC weight (+1, 0 or −1) and then accumulated. The time-domain MP output pulse ($\mathrm{PW}_{\mathrm{FC}}\langle j\rangle$) will be first gated out by $\mathrm{ZERO}_{\mathrm{FC}}\langle j\rangle = 0$ when FC weight = 0, or passed through by $\mathrm{ZERO}_{\mathrm{FC}}\langle j\rangle = 1$ when FC weight = ±1 to realize the multiplication function using a three-input AND gate (QZ). With an applied quantization clock ($\mathrm{CLK}_{\mathrm{FC\_QZ}}$), the output of AND gate is the quantized result and counted by a binary counter with a tunable bit depth (3-bit: ±4 in this example) and temporarily stored in the following DFF in thermometer code format. Finally, the stored code is accessed serially by column select ($\mathrm{CSEL}_{\mathrm{FC}}\langle j\rangle$) and gated with the counting clock ($\mathrm{CLK}_{\mathrm{FC\_SUM}}$) for the global up/down counter controlled by the sign bit of FC weight ($\mathrm{SIGN}_{\mathrm{FC}}$) to implement the accumulation function. By comparing the counting code, which accumulates over one frame, with a predefined threshold (e.g., half of full range, 16383), a 1-bit binary code (MSB of global counter) is output as the final classification result. Thanks to the time-domain MP output and clock-gating operation, the digital MAC operation of the FC layer with ternary weights can be efficiently implemented using a simple gating operation for multiplication and an up/down counter for accumulation without needing a multiplier and adder.

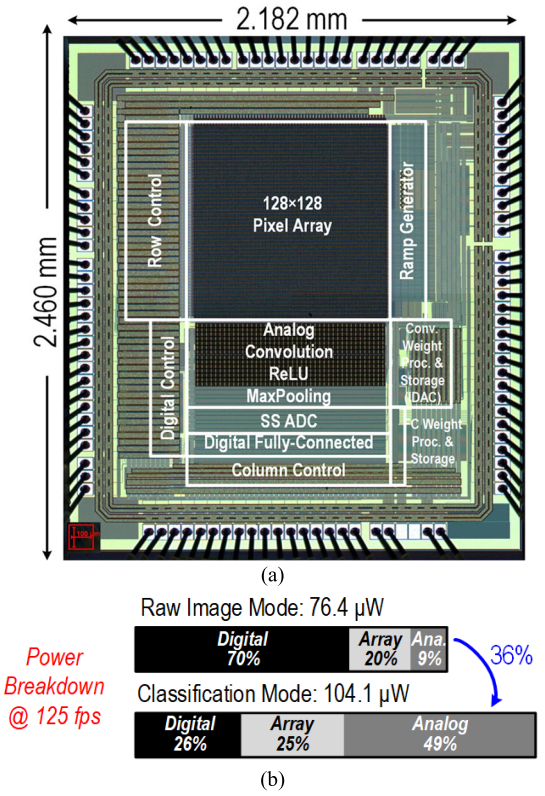


Fig. 16. (a) Chip micrograph and (b) power breakdown at 125 f/s.

C. Timing of the Proposed PIS Operations for CNN Processing

Fig. 9 shows the timing diagram of the implemented PIS operations for multi-kernel CNN processing. After the operations of 3×3 Conv + ReLU with different kernels (eight in this example) and 2×2 MP, the 126×126 raw image data is down-scaled to eight FMs (FM1–FM8) with each size of 21×21. For example, the first row ($\mathrm{FM}_{\mathrm{ROW}}\langle 0\rangle$) in FM1 is calculated and stored from the six-row data ($\mathrm{PA}_{\mathrm{ROW}}\langle 0:5\rangle$) in the pixel array after the sequential EVEN/ODD Conv+ReLU and MP+Quantization operations in $T_{\mathrm{KERNEL}}1$. Then, the stored FM1 code is accessed serially for FC accumulation during the next operation cycle ($T_{\mathrm{KERNEL}}2$). After eight operation cycles ($T_{\mathrm{ROW}} = 8 \times T_{\mathrm{KERNEL}}$) for eight kernels, the image pixel can be reset for next exposure as a rolling shutter operation. After a frame time ($T_{\mathrm{FRAME}} = 21 \times T_{\mathrm{ROW}}$), the MAC operation of the FC layer with a total of 3528 elements in FMs ($21 \times 21 \times 8$) are accumulated in the global counter and represents the classification result directly.

TABLE II
COMPARISON TABLE

	[9] 2021 ISSCC	[10] 2021 VLSI	[11] 2017 ESSCIRC	[7] 2020 Sensors	This Work
Process	65nm CMOS	180nm CIS	65nm Logic CMOS	110nm CIS	180nm CMOS
Supply	0.8V ~ 1.2V	(Analog) 2.5V (Digital) 1.8V	(Analog) 2.5V (Digital) 0.5~0.8V	(Analog) 3.3V (Digital) 1.5V	(Analog) 0.8V (Digital) 0.8V
Die Area	2 mm x 2 mm	5.2 mm x 4.1 mm	3.3 mm x 3.3 mm	5.9 mm x 5.2 mm	2.46 mm x 2.18 mm
Pixel Size	9 μm	9.8 μm	7 μm	N.A.	7.6 μm
Pixel Type	40T log(I)-V Pixel + 1 MIMCAP	14T+4C pinned-PD	N.A.	4T APS	4T PWM
Fill Factor	12.9%	20.1%	N.A.	N.A.	36%
Array Size	160x128	240x240	320x240	160x120	126x126
Frame Rate	24~268 fps	120 fps (Global shutter)	1fps	120 fps	250 fps
In-sensor-processing Tasks	Haar-like filtering	Log-Haar-like filtering + Classifiers	Haar-like filtering (Integrate Digital Vision Processor on-chip)	Convolution, ReLU, MaxPooling, Fully-connected	Convolution, ReLU, MaxPooling, Fully-connected
Processing Algorithm	Viola-Jones 2-stage cascade classifier	25 Machine Learning feature classifier	Viola-Jones (3Ana.+20Dig.)-stage cascade classifier	5-layers CNN	3-layers CNN
Weight Precision	6 scales kernel 1.5b (+1,0,-1)	Multiscale kernel 1.5b (+1,0,-1)	3 scales kernel 1.5b (+1,0,-1)	(Conv.) 2x2 kernel (FC) 5b	(Conv.) 3x3 kernel, ±3b (FC) 1.5b (+1,0,-1)
FD Task Accuracy/Precision/Recall	> 80% / N.A. / N.A. window rejection (Need digital backend)	98% / N.A. / N.A. window rejection (Need digital backend)	> 90% / N.A. / N.A. (w/ on-chip digital processor)	89.3% / 94.7% / 72% (w/o digital backend)	93% / 90.4% / 97.2% (w/o digital backend)
Dataset	KODAK	FERET	N.A.	N.A.	LFW / Kaggle Oregon Wildlife
Power	42~206 μW	2.9 mW @ 120 fps	60 μW @1 fps (averaged)	0.96 mW @ 60 fps 1.12 mW @ 120 fps	80.4 μW @ 50 fps 104.1 μW @ 125 fps 134.5 μW @ 250 fps
iFoM	2.5~103.9 pJ/pix·fps	419 pJ/pix·fps	781.2 pJ/pix·fps	833 pJ/pix·fps @ 60fps 486 pJ/pix·fps @ 120fps	101.2 pJ/pix·fps @ 50fps 52.4 pJ/pix·fps @ 125fps 33.8 pJ/pix·fps @ 250fps

IV. EXPERIMENT RESULT

A 0.8 V 128 × 128 IVS was fabricated in 0.18 μm standard CMOS. Fig. 10 shows the measured signal transfer curve of the adopted PWM pixel and the corresponding non-linearity of $+2.6\%/-1.2\%$ in normal imaging mode. Fig. 11 shows the measured photon transfer curve (PTC) of pixel performance which excluding the CNN computing circuit. The signal and noise levels were measured in raw image mode with sensor illuminated under a uniform light source at different lux condition. In each lux, 100 images were captured and calculated the mean and standard deviation over pixels to define the signal and noise level in DN, respectively. The achievable dynamic range (DR) and the peak signal-to-noise ratio (PSNR) are 47.78 and 42.6 dB, respectively. Fig. 12 shows the captured raw image and the FMs, which were computed from eight different kernels, with a resolution of 126 × 126 and 21 × 21, respectively.

Fig. 13(a) shows the experimental setup, where the prototype sensor is synchronously controlled through PC and NI chassis. The tiny three-layer CNN model training is executed using 10 000 images selected from the "Labeled Faces in the Wild (LFW) dataset" [16] and "Oregon wildlife images dataset from Kaggle" [17] for the FD binary classification. Fig. 13(b) shows the experimental setup for evaluating 5400 images in the testing dataset, where images are displayed on the monitor screen and captured using the prototype. Fig. 14 shows the statistic distribution of classification result and confusion matrix for comparison of actual and predicted labels which yield an accuracy of 93.6%. Taking advantage of the CNN processing architecture, multiple-scale image classification can be achieved using the same Conv weights with a hardware windowing operation. Fig. 15 shows the captured images and FMs with the detection windows of 126 × 126, 84 × 84, and 66 × 66. Table I provides the measured classification results are 93.6%, 91.6%, and 90.1%, respectively, where around 3% accuracy drop is due to fewer FC parameters (from $21\times21\times8$ to $14\times14\times8$ and $11\times11\times8$) as detection window is shrinking. With respect to the maximum frame rate, since the PIS circuit is realized in column-parallel architecture, the utilization of the column-wise computing unit will reduce as the detection window is getting smaller. As a result, as detection window is changed from 126 × 126 to 84 × 84 and 66 × 66, the number of T_{FRAME} will reduce from 21× to 14× and 11× of T_{ROW} and give rise to a 1.5× and 1.9× of maximum frame rate improvement.

Fig. 16 shows the die micrograph and the power breakdown when using full array (126 × 126) as detection window and operating at 125 f/s. Since the digital power is mainly consumed by the ADC and the controlling signals. In raw image mode, 8-bit single-slope ADC requires 256 clock cycles while only four cycles are used for PW_{MP} quantization in classification mode. Moreover, the amount of pixel data for quantization is reduced from 126 × 126 to 21 × 21 × 8. Consequently, the digital circuit consumes a larger ratio of power in raw image mode.

Using the mixed-mode PIS technique, the IVS in classification mode with on-chip customized CNN model operation

consumes only 36% more power compared to the raw image mode without any off-chip data processing. Table II summarizes the measured performance and comparison with other works implementing FD tasks. By realizing the PIS circuit in column-parallel topology, pixel size is kept at 7.6 μm with a fill factor of 36%. Comparing with other works which using Viola-Jones haar-like filtering, this work realizes FD tasks without backend digital processing. Moreover, the 0.8 V supply prototype is configured to realize a multiscale FD task. In classification mode, it consumes 80.4 μW at 50 f/s, 104.1 μW at 125 f/s, and 134.5 μW at 250 f/s with the resulting iFoMs of 101.2, 52.4, and 33.8 pJ/pixel·f/s, respectively.

V. Conclusion

This article presents an IVS with customized CNN computing circuit for classification task. The achievable detection rate and power efficiency are 250 f/s and 33.8 pJ/pixel·f/s, respectively, and the sensor can operate even without the support of digital backend processing. By taking advantage of 0.8 V low-voltage-operated PWM pixel and high-efficient PIS circuits, the proposed sensor demonstrates a CNN-based binary classification for FD with above 90% of accuracy. The highly parallel processing circuit is realized in a cost-efficient die area of 2.18 $\times$ 2.46 mm^2 and enables the real-time classification without sacrificing frame rate and using extra frame memory. Also, the proposed mixed-mode PIS architecture can reduce 77.7% of ADC workload and 99.9% data traffic comparing to conventional CIS plus CNN accelerator approach. Moreover, with the programmability of 3 $\times$ 3 convolution kernel and FC weights, the proposed single-chip solution for CNN computation can be applicable to more different binary classification tasks in always-on edge application devices.

Acknowledgment

The authors would like to thank Signal Sensing and Application Laboratory (SiSAL), National Tsing Hua University (NTHU), Hsinchu, Taiwan. The authors also acknowledge the support of the Taiwan Semiconductor Research Institute (TSRI), Tainan, Taiwan, and Taiwan Semiconductor Manufacturing Company (TSMC), Hsinchu, China, for the fabrication of the test chip.

References

[1] T.-H. Hsu et al., "A 0.8 V multimode vision sensor for motion and saliency detection with ping-pong PWM pixel," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2020, pp. 110–112.

[2] C. Yin, C. Chiu, and C. Hsieh, "A 0.5 V, 14.28-kframes/s, 96.7-dB smart image sensor with array-level image signal processing for IoT applications," *IEEE Trans. Electron Devices*, vol. 63, no. 3, pp. 1134–1140, Mar. 2016.

[3] X. Zhong, Q. Yu, A. Bermak, C. Tsui, and M. Law, "A 2PJ/pixel/direction MIMO processing based CMOS image sensor for omnidirectional local binary pattern extraction and edge detection," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2018, pp. 247–248.

[4] T. Hsu et al., "A 0.5-V real-time computational CMOS image sensor with programmable kernel for feature extraction," *IEEE J. Solid-State Circuits*, vol. 56, no. 5, pp. 1588–1596, May 2021.

[5] K. Bong, S. Choi, C. Kim, D. Han, and H. Yoo, "A low-power convolutional neural network face recognition processor and a CIS integrated with always-on face detector," *IEEE J. Solid-State Circuits*, vol. 53, no. 1, pp. 115–123, Jan. 2018.

[6] J. Kim, C. Kim, K. Kim, and H. Yoo, "An ultra-low-power analog-digital hybrid CNN face recognition processor integrated with a CIS for always-on mobile devices," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2019, pp. 1–5.

[7] J. Choi, S. Lee, Y. Son, and S. Y. Kim, "Design of an always-on image sensor using an analog lightweight convolutional neural network," *Sensors*, vol. 20, no. 11, p. 3101, May 2020.

[8] R. Eki et al., "A 1/2.3inch 12.3 Mpixel with on-chip 4.97 TOPS/W CNN processor back-illuminated stacked CMOS image sensor," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, vol. 64, Feb. 2021, pp. 154–156.

[9] M. Lefebvre, L. Moreau, R. Dekimpe, and D. Bol, "A 0.2-to-3.6 TOPS/W programmable convolutional imager SoC with in-sensor current-domain ternary-weighted MAC operations for feature extraction and region-of-interest detection," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, vol. 64, Feb. 2021, pp. 118–120.

[10] H. Song, S. Oh, J. Salinas, S. Park, and E. Yoon, "A 5.1 ms low-latency face detection imager with in-memory charge-domain computing of machine-learning classifiers," in *Proc. Symp. VLSI Circuits*, Jun. 2021, pp. 1–2.

[11] C. Kim, K. Bong, I. Hong, K. Lee, S. Choi, and H. Yoo, "An ultra-low-power and mixed-mode event-driven face detection SoC for always-on mobile applications," in *Proc. 43rd IEEE Eur. Solid State Circuits Conf. (ESSCIRC)*, Sep. 2017, pp. 255–258.

[12] T. Hsu et al., "A 0.8 V intelligent vision sensor with tiny convolutional neural network and programmable weights using mixed-mode processing-in-sensor technique for image classification," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, vol. 65, Feb. 2022, pp. 1–3.

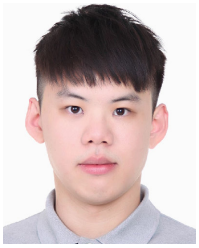
[13] A. Y. Chiou and C. Hsieh, "An ULV PWM CMOS imager with adaptive-multiple-sampling linear response, HDR imaging, and energy harvesting," *IEEE J. Solid-State Circuits*, vol. 54, no. 1, pp. 298–306, Jan. 2019.

[14] A. Kitchen, A. Bermak, and A. Bouzerdoum, "PWM digital pixel sensor based on asynchronous self-resetting scheme," *IEEE Electron Device Lett.*, vol. 25, no. 7, pp. 471–473, Jul. 2004.

[15] K. Korekado, T. Morie, O. Nomura, T. Nakano, M. Matsugu, and A. Iwata, "An image filtering processor for face/object recognition using merged/mixed analog-digital architecture," in *Proc. IEEE Sympo VLSI Circuits*, Kyoto, Japan, Jun. 2005, pp. 220–223.

[16] B. G. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, Tech. Rep., pp. 7–49, Oct. 2007.

[17] Kaggle. *Oregon Wildlife Image Collection*. Accessed: Dec. 22, 2019. [Online]. Available: https://www.kaggle.com/virtualdvid/oregon-wildlife



Tzu-Hsiang Hsu received the B.S. and Ph.D. degrees from the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, in 2015 and 2021, respectively.

He is currently a Senior Engineer with Mediatek Inc., Hsinchu. His research interests include CMOS image sensor, mixed-mode integrated circuit (IC) development for artificial intelligence (AI), and customized applications.

Guan-Cheng Chen (Member, IEEE) received the B.S. degree from National Tsing Hua University, Hsinchu, Taiwan, in 2018, where he is currently pursuing the Ph.D. degree in electrical engineering.

His research interests include CMOS time-of-flight (ToF) depth image sensor and mixed-mode integrated circuit (IC) design.

Yi-Ren Chen received the B.S. and M.S. degrees from National Tsing Hua University, Hsinchu, Taiwan, in 2018 and 2021, respectively.

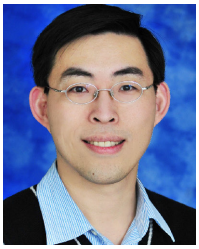
His research interests include deep learning and model compression algorithm.

Ren-Shuo Liu (Member, IEEE) received the B.S. and M.S. degrees in electronic engineering (EE) and the Ph.D. degree in computer science (CS) from National Taiwan University, Taipei, Taiwan, in 2001, 2003, and 2014, respectively.

From 2003 to 2008, he was with Acer Laboratories Inc. (ALi), Taipei. He is currently an Associate Professor with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan. His research interests include computer architecture, nonvolatile memory (NVM) and solid-state disk (SSD) systems, and hardware–software (HW–SW) co-design for artificial intelligence (AI) accelerators.

Chung-Chuan Lo received the Ph.D. degree in physics from Boston University, Boston, MA, USA, in 2004.

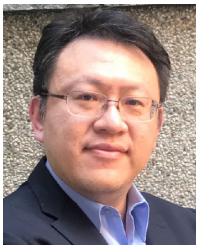
He was with Dr. Xiao-Jing Wang's Laboratory at Brandeis University, Waltham, MA, USA, from 2004 to 2006, and Yale University, New Haven, CT, USA, from 2006 to 2008, to conduct his post-doctoral research on computational neuroscience. In 2008, he holds a faculty position with National Tsing Hua University, Hsinchu, Taiwan, and gradually shifted his attention from plasticity and balance in cortical circuit models to visuomotor functions of fruit fly brains. In the past few years, he collaborated with researchers in electrical engineering and applied the knowledge he gained from the insect nervous systems to the design of neuromorphic chips.



Kea-Tiong Tang (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the California Institute of Technology, Pasadena, CA, USA, in 2001.

His research interests include bio-inspired learning chip, in-memory computing based deep learning accelerator, miniature electronic nose system, and biomedical implantable prosthetic device.

Dr. Tang was a recipient of numerous awards, including the Outstanding Young Scholar Award, the Wu Ta-You Memorial Award, the National Innovation Award, and the Outstanding Electrical Engineering Professor Award. He was the IEEE CAS Chapter Chair of Taipei Section from 2017 to 2018. He is the Chair of IEEE Taipei Section. He is serving as the Board of Governor (BoG) of CAS Society. He is also the Associate Editor-in-Chief of IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS (TBioCAS) and an Associate Editor of IEEE SENSORS JOURNAL.



Meng-Fan Chang (Fellow, IEEE) received the M.S. degree from The Pennsylvania State University, State College, PA, USA, and the Ph.D. degree from National Chiao Tung University, Hsinchu, Taiwan.

Prior to 2006, he worked in the industry for over ten years. This included the design of memory compilers (Mentor Graphics Wilsonville, OR, USA, from 1996 to 1997) and the design of embedded SRAM and flash macros (Design Service Division, Taiwan Semiconductor Manufacturing Company (TSMC), Hsinchu, from 1997 to 2001). In 2001, he co-founded IPLib, Hinshu, where he developed embedded SRAM and ROM compilers, flash macros, and flat-cell ROM products until 2006. He is currently a Distinguished Professor with National Tsing Hua University (NTHU), Hinshu, and the Director of Corporate Research at TSMC. His research interests include circuit design for volatile and nonvolatile memory, ultralow-voltage systems, 3-D-memory, circuit-device interactions, spintronic circuits, memristor logics for neuromorphic computing, and computing-in-memory for artificial intelligence.

Dr. Chang was a recipient of several prestigious national-level awards in Taiwan, including the Outstanding Research Award of MOST-Taiwan, the Outstanding Electrical Engineering Professor Award, the Academia Sinica Junior Research Investigator Award, and the Ta-You Wu Memorial Award. He is a Distinguished Lecturer of the IEEE Solid-State Circuits Society (SSCS) and the Circuits and Systems Society (CASS), the Chair of the Nano-Giga Technical Committee of CASS, an Administrative Committee (AdCom) Member of the IEEE Nanotechnology Council, and the Chair of the IEEE Taipei Section. He has been serving as an Associate Editor for IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPERS, and IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS and a Guest Editor for IEEE JOURNAL OF SOLID-STATE CIRCUITS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: EXPRESS BRIEFS, and IEEE JOURNAL ON EMERGING AND SELECTED TOPICS IN CIRCUITS AND SYSTEMS. He has also been serving on the Executive Committee for IEDM and as the Subcommittee Chair for ISSCC, IEDM, DAC, ISCAS, VLSI-DAT, and ASP-DAC. He was the Program Director for the Micro-Electronics Program at the Ministry of Science and Technology, Taiwan, and the Associate Executive Director of the Taiwan's National Program of Intelligent Electronics (NPIE) and the NPIE Bridge Program.

Chih-Cheng Hsieh (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the Department of Electronics Engineering, National Chiao Tung University, Hsinchu, Taiwan, in 1990, 1991, and 1997, respectively.

From 1999 to 2007, he was with Pixart Imaging Inc., Hsinchu, an IC design house. He led the Mixed-Mode IC Department, as a Senior Manager and was involved in the development of CMOS image sensor ICs for PC, consumer, and mobile phone applications. In 2007, he joined the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, where he is currently a Full Professor. His research interests include low-voltage low-power smart CMOS image sensor IC, ADC, and mixed-mode IC development for artificial intelligence (AI), the Internet of Things (IoT), biomedical, space, robot, and customized applications.

Dr. Hsieh serves as a TPC Member for ISSCC and A-SSCC. He was a recipient of several prestigious national-level awards in Taiwan, including the National Innovation Award, the Outstanding Electrical Engineering Professor Award, and the Outstanding Research Award of MOST-Taiwan. He was the SSCS Taipei Chapter Chair and the Student Branch Counselor of NTHU, Taiwan. He is an Associate Editor of IEEE SOLID-STATE CIRCUIT LETTERS (SSC-L) and *IEEE Circuits and Systems Magazine*(CASM).