Battery-Less Face Recognition at the Extreme Edge

Petar Jokic*†, Stephane Emery*, Luca Benini†
**CSEM SA*, Zurich, Switzerland; †*ETH Zurich*, Zurich, Switzerland; Email: petar.jokic@csem.ch

2021 19th IEEE International New Circuits and Systems Conference (NEWCAS) | 978-1-6654-2429-5/20/$31.00 ©2021 IEEE | DOI: 10.1109/NEWCAS50681.2021.9462787

***Abstract*—Machine learning-based face recognition systems are commonly used in mobile platforms to assist the camera systems, unlock the device, or analyze the facial expressions. The computational complexity of the underlying algorithms as well as the power consumption of the entire imaging and processing system largely limit the deployment to powerful mobile processing systems with large rechargeable batteries. However, these computer vision capabilities would also be useful in miniaturized low power applications with stringent battery-size limitations. We assess the feasibility of such a computer vision edge processing system on a battery-less credit card-sized demonstrator using an ultra-low power image sensor and a machine learning system-on-chip, achieving self-sustainable operation using solar energy harvesting with a small on-board solar cell. The tested system enables continuous 1 frame-per-second battery-less imaging and face recognition in indoor lighting conditions.**

Keywords—* Edge processing, machine learning acceleration, neural network, energy harvesting, self-sustainable

I. Introduction

Machine learning (ML) algorithms are employed in a rapidly growing number of applications, supporting tasks like visual object detection [1], audio key-word spotting [2], radio signal analysis [3] and many others. Most of these algorithms are based on neural networks, achieving unprecedented accuracies while requiring processing systems with high computational throughput and large memory resources. The majority of ML applications are used in environments with virtually unlimited power access (e.g. in a server). However, battery-powered mobile platforms, often used in internet-of-things (IoT) applications [4], are becoming smarter and thus interesting targets for implementing ML algorithms. Due to the limited computational capacity in IoT devices, cloud-offloading was introduced [4], transmitting the raw data through a communication network for processing in a cloud server. High interface power consumptions and improved efficiencies of ML accelerators (e.g. [5]) led to so-called edge computing devices [6], processing and analyzing sensor input data directly on-board, at the edge (of the network).

Truly mobile applications, with battery lifetimes of more than a month, require very low power edge ML platforms. Using a common CR2032 coin cell battery, a 1-month lifetime would restrict the power consumption to 1mW [7]. This energy-limitation can be mitigated by employing on-board energy harvesting, constantly extracting power from the environment. Various approaches have been evaluated, exploiting solar [8], kinetic [9], vibration [10], or even radio frequency power [11]. The source is continuously tracked to extract the maximum power, as shown in [12] and [13].

Mobile ML implementations for computer vision (CV) applications have been shown in various previous works and industrial devices. However, they all feature a very limited battery lifetime. Xiaomi's 2-megapixel AI doorbell [14] features face identification and movement detection with up to 60 days of operation using a large 3000mAh secondary battery. The Google Clips camera [15] claims to recognize people in an always-on operation, automatically deciding when to capture photos and of whom. Its battery allows only 3 hours of operation. Orcam MyEye 2 [16] is a smart camera for blind people that can recognize people and read text out loud. It features a 13 megapixel image sensor and a 350mAh secondary battery for up to 2 hours of operation. EdgeEye [17] presents an end-to-end people counting system using a 185MOP convolutional neural network (CNN). Image sampling and processing consume 17.5mJ with an idle power consumption of 430uW. In [18], a combination of a custom 320x240 pixel image sensor and a CNN processing chip is used to perform face detection and recognition, consuming an average of 0.62mW core power, however excluding the power for image transfer and external components. The computation and memory-intensive CV algorithms challenge the tight power budget of these battery-powered edge processing applications. Thus, they either use low frame rates or accept short battery lifetimes. For embedding real-time CNN-based user identification in a miniaturized "extreme edge" device as in Fig. 1, both powerful CNN processing capabilities and a long lifetime are required.

This work presents a credit card-sized battery-less computer vision platform (Fig. 2) capable of performing edge computing for face recognition using an on-board image sensor and an ultra-low power (ULP) machine learning system-on-chip (SoC) while being powered solely by a small solar panel. With an average power consumption of 0.68mW for acquiring images and running face recognition at 1 frame per second (FPS), 1klux indoor lighting provides sufficient harvesting power to enable self-sustainable operation.

The paper is structured as follows: Section II describes the

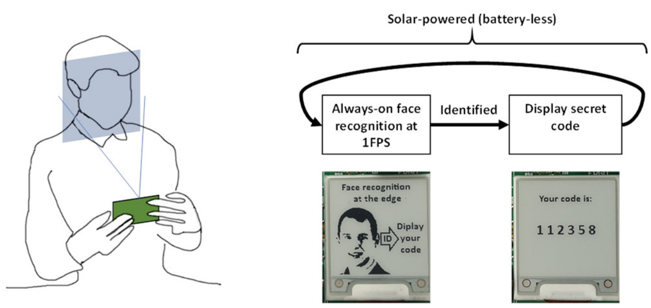

Fig. 1: Credit card-sized identification using face recognition at the edge.

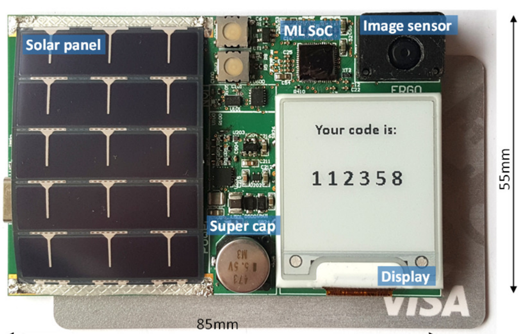

Fig. 2: Computer vision platform with solar panel (left), display (bottom right), and image sensor with flat optics (top right) in front of a credit card.

This work was funded in part by the Swiss National Science Foundation (SNSF) BRIDGE under grant number 40B2-0_181010.

system components, followed by the description of the face recognition application in Section III. Experimental results, including power measurements, are reported in Section IV.

II. System Implementation

Fig. 3 shows the block diagram of the edge processing platform including the communication interfaces between the sub-systems. It consists of four functional blocks: a) Imaging, b) control and ML-based image processing, c) displaying, as well as d) power management and energy harvesting. The platform, illustrated in Fig. 2, is implemented on a printed circuit board with the size of a credit card (55x85mm). Solar harvesting and image processing match well as both only operate in lit environments, where image sensing is possible.

A. Image acquisition

Operating on a limited power budget requires sufficient image quality in every acquired frame. Iterative exposure time approximation by sampling at different settings should therefore be avoided. Thus, we use an ERGO320 image sensor [19], featuring a high dynamic range of 120dB, enabling high contrast images under strong illumination variations. A miniaturized lens with 2.8mm focal length enables a slim demonstrator (<7mm thick) while providing a wide field-of-view (107°). The sensor requires a 1.8V supply voltage and communicates through a slave SPI interface.

B. Control and ML processing

The heart of the system is an ML SoC [20] that orchestrates the sub-systems and provides ML processing capabilities. It features a RISC-V microcontroller with various sensor interfaces, a 1MB static random-access memory (SRAM), and an efficient multi-precision CNN accelerator, supporting 1bit and 16bit weights with 16bit activations. It automatically boots from an external SPI memory, as shown in Fig. 3.

The SoC directly interfaces the image sensor, the display, and the Flash memory through SPI. The non-volatile 2MB Flash memory (MX25R1635) can additionally be accessed through USB (via a USB/SPI converter), allowing to update the program code and the ML algorithm parameters from a connected PC. Digital outputs are used to control multiple power switches to power-gate unused sub-systems (image sensor, Flash memory, display), minimizing idle currents.

CNN algorithms for object detection require millions of operations to be performed for analyzing a single image [1]. The dominating multiply-and-accumulate (MAC) operation processes an activation and a weight, then accumulates the result (and writes it back to memory), implying multiple memory accesses. Considering the energy required to process and move data [21], these operations must be minimized. Especially accessing data from external memories is very costly, as the energy for booting shows. Thus, we selected an ML SoC that provides sufficient on-chip memory for face recognition algorithms. Its CNN accelerator with 16 parallel MAC units runs independently of the microcontroller and supports 16bit and 1bit weights (layer-wise configurable), reducing the memory needs for parameters by up to 16x. The trained (and quantized using the method of [22]) network is compiled through the SoC tool-flow, mapping all parameters and layer settings to the SRAM, which is loaded from Flash memory during the booting process. To start the CNN execution, the microcontroller sets the start flag in the CNN register and gets notified by an interrupt upon completion.

C. Display

A 152x152 pixel E2154CS electronic paper display (EPD) is used to display application results. EPDs feature high contrast, making them well readable even in bright outdoor environments. Other display types, like LCD or LED matrix, require a continuous power supply during operation, while EPDs retain the last image displayed and only require power for updating the content. Fig. 4 shows the measured power consumption for a display update. After powering the EPD, its on-board controller is configured, and the new image data is transferred via SPI (busy signal high). Then, the DC/DC converter of the display is started, driving the busy signal low until the power supply is stable. Upon sending the image update command, the EPD starts its internal refresh process, updating the pixels, followed by switching off the DC/DC converter again. The entire update lasts 2625ms and requires 22.8mJ per image refresh. Comparable low power LCDs [23] consume up to 1000x less power but require constant power supply and a power-intensive backlight for similar readability.

D. Power management and energy harvesting

We utilize a BQ25570 solar energy harvester to power the platform. It performs maximum power point (MPP) tracking to efficiently operate the solar cell at its light intensity-dependent MPP. The extracted power is buffered in a capacitor (>3V) and output as a stable 1.8V supply voltage. We use a miniaturized, 5.5mm high, 47mF super-cap instead of a battery, allowing to bridge short energy peaks (e.g. for booting from Flash memory). The estimated buffered energy in the capacitor can be computed using the formula of the stored energy (1), resulting in 211.5mJ, roughly 10x more than a display update requires.

$$E = \frac{1}{2} \cdot C \cdot U_C^2 = \frac{1}{2} \cdot 47mF \cdot (3V)^2 = 211.5mJ \quad (1)$$

A flexible PowerFilm SP3-37 solar panel is used but could be replaced by other types (>100mV, <400mW). Fig. 5 shows its indoor and outdoor characteristics, achieving a maximum output power of 1.05mW and 54.49mW, respectively.

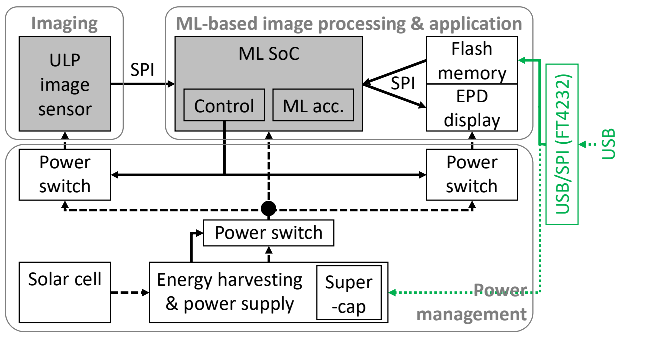

Fig. 3: System block diagram with power (dashed) and data (solid) interfaces. USB-related circuits (green) are only active during programming.

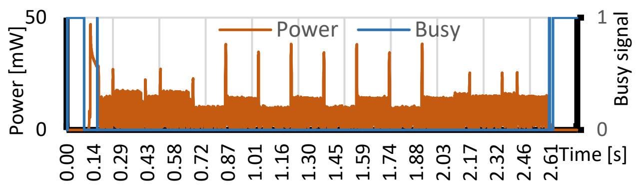

Fig. 4: Measured EPD display power consumption during a display update.

III. APPLICATION

We demonstrate end-to-end ML edge processing of an always-on face recognition application using solar energy harvesting. The credit card-sized application aims at identifying a specific user to display a secret code (e.g. pin code of the card) upon positive recognition as shown in Fig. 1. In the future, the code could be transmitted through RFID.

A. ML algorithm

We employ a 6-layer CNN with binary weights for recognizing a face on the input image. Table I shows its architecture, inspired by [24], totaling 106MMAC operations and 392kB parameters. It takes a 128x128 pixel gray-scale image as input and produces a 512-element output vector using kernel sizes of 3x3-5x5 and 48-256 channels. The face similarity is determined by the Euclidean distance between the output vector and the programmed reference vector of the face to be identified. We define a positive identification as a distance below a specific threshold. Evaluated on the LFW dataset [25], the CNN achieves 96% accuracy in simulation.

During layer-wise CNN inference, all network parameters must be stored in memory along with intermediate layer results. To minimize the memory required for buffering activations, we employ the mapping strategy of [26]. It partially overlaps the memory space of each layer's input and output activations, while precluding premature overwriting of input activations. This reduces the activation memory by 32.2% with respect to standard double buffering as shown in Fig. 6.

B. Task scheduling

Self-sustainable operation imposes two power constraints that always must be fulfilled: 1) the average power of all executed operations must be lower than the average harvested power and 2) any peak surpassing the average harvested power must be compensated by the available capacitive load.

The flow chart in Fig. 7 illustrates the main operations of the application along with the independent energy harvesting. Its timing ensures that both power constraints are fulfilled. When the 1.8V supply voltage is stable, the harvester enables the main power switch. This starts the power supplies and the 32kHz clock generator while keeping the SoC reset during the oscillator startup-time. Releasing the reset automatically starts the booting process in the SoC, loading the program and CNN parameters into the SRAM. The microcontroller then enables the internal 180MHz clock and starts the application.

Always-on face recognition is implemented as periodic image sampling with subsequent CNN analysis. Each cycle starts by triggering an image in the sensor through the 20MHz SPI interface. A sensor interrupt starts the SPI image transfer to the SoC SRAM via direct memory access. The CNN accelerator then processes the network layer-by-layer starting from the down-scaled image. An internal interrupt indicates the inference end, waking up the sleeping microcontroller to analyze the output vector. If the face is identified, the display is updated, followed by the idle screen, 60 seconds later, as shown in Fig. 1. Otherwise, the periodic task restarts at 1 FPS.

IV. EXPERIMENTS AND RESULTS

We measure the solar harvesting and power consumption of the entire system in all operation phases. Fig. 8 shows the break-down of energy per operation across the main subsystems. Idle energy is measured over 1 second for reference.

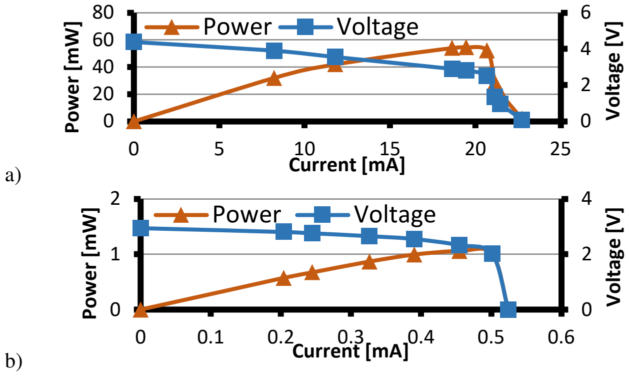

Fig. 5: SP3-37 characteristics for a) outdoor (20klux) and b) indoor (1klux).

TABLE I
CNN ARCHITECTURE

Layer	Input size	Input ch.	Output ch.	Kernel size	Stride	#MAC	Param. [kB]
CONV1	128	1	48	5	2	4.9M	0.2
CONV2	64	48	96	3	2	42.5M	5.2
CONV3	32	96	128	3	2	28.3M	13.8
CONV4	16	128	256	3	2	18.9M	36.9
CONV5	8	256	256	3	2	9.4M	73.8
CONV6	4	256	512	4	4	2.1M	262.2
Total						**106.1M**	**392.1**

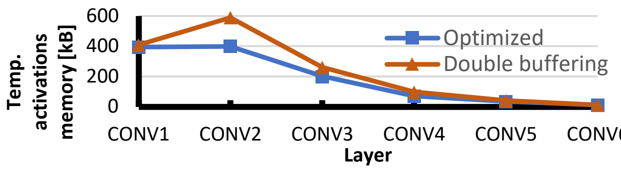

Fig. 6: Layer-wise activations memory space for both mapping strategies.

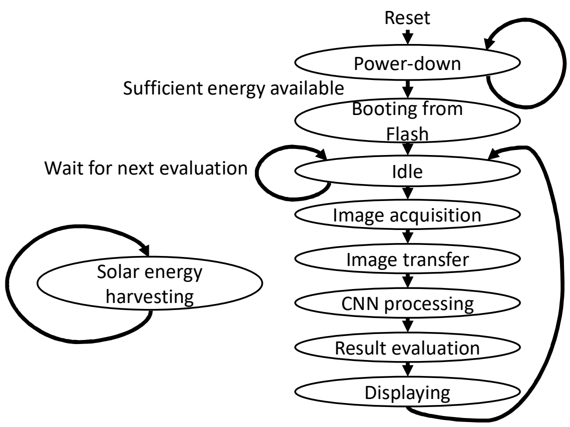

Fig. 7: Operation flow chart with continuous solar energy harvesting.

Image transfer and CNN processing have a similar energy cost while acquiring an image is roughly 3x cheaper. Booting from the Flash memory and updating the display are very intensive, consuming 7.45mJ and 23.87mJ, respectively. The ML SoC consumes 70% of the power during operation.

Fig. 10 compares the harvested indoor solar power of 0.94mW (90% harvesting efficiency) and the 0.68mW power consumption at 1 FPS, verifying self-sustainable operation. Table II summarizes the operation phases, highlighting the average power consumption. Note the theoretical values for the operation at the maximum frame rates. Booting and display updating exceed the average harvested power and thus are duty-cycled by appending an idle state phase of 7.5 seconds and 24 seconds, respectively, resulting in an average power consumption below the harvested power of 0.94mW.

Outdoor lighting delivers >50x higher harvesting power,

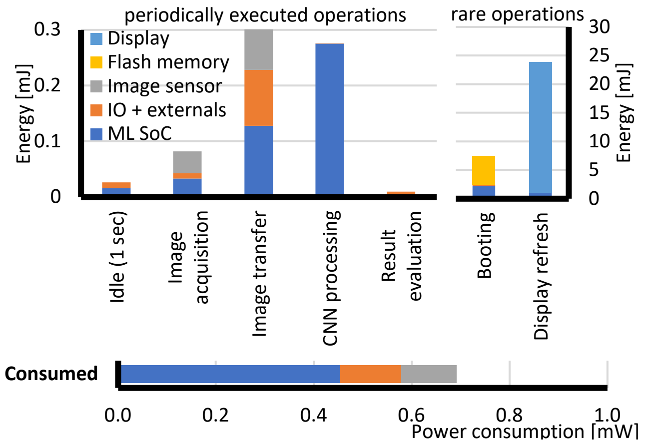

Fig. 8: System energy per operation (top) and power consumption (bot.) for 1 FPS imaging and CNN processing with break-down across sub-systems.

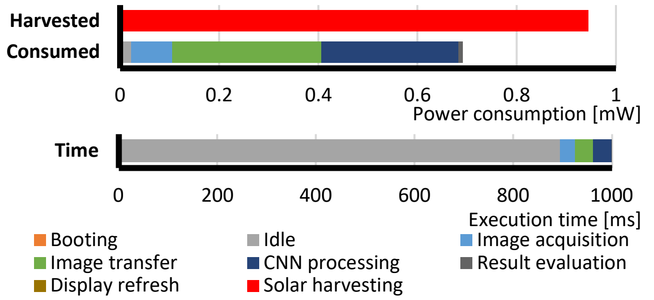

Fig. 9: System power and execution time break-down across operations for 1 FPS imaging and CNN processing using indoor solar harvesting.

TABLE II
OPERATION PHASES

Scenario	Booting + 7.5s idle	Imaging + CNN	Displaying + 24s idle	Max FPS* Imaging + CNN	Max FPS* Imaging + CNN + Display.
FPS [1/s]	N/A	1	N/A	9.6	0.4
Exec. time [ms]	8’365.3	1000.0	26’625.4	1’000.0	1’000.0
Idle time	89.7%	89.6%	90.1%	0%	0%
Av power [mW]	0.91	0.68	0.92	6.33	9.08

* Theoretical scenario to show maximum frame rates possible.

which would allow maximum frame rates of 9.6 FPS with a static display and 0.4 FPS with per-frame display update as shown in Table II. However, this requires monitoring the harvested energy and adapting the frame rate accordingly, which the current version of the platform does not support.

V. CONCLUSION

We presented a credit card-sized battery-less vision platform for end-to-end CV processing at the edge using solar power harvesting. The platform enables to acquire images and perform CNN-based face recognition at 1 FPS while consuming 0.68mW. This allows self-sustainable operation in indoor lighting conditions.

REFERENCES

[1] Z. Zhao, P. Zheng, S. Xu and X. Wu, "Object Detection With Deep Learning: A Review," *IEEE Trans. Neural Networks and Learning Systems,* vol. 30, pp. 3212-3232, 2019.

[2] Y. Zhang, N. Suda, L. Lai and V. Chandra, "Hello Edge: Keyword Spotting on Microcontrollers," *CoRR,* vol. abs/1711.07128, 2017.

[3] H. Tamura, K. Yanagisawa, A. Shirane and K. Okada, "Wireless Devices Identification with Light-Weight Convolutional Neural Network Operating on Quadrant IQ Transition Image," in *IEEE International New Circuits and Systems Conference*, 2020.

[4] J. Gubbi, R. Buyya, S. Marusic and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems,* vol. 29, pp. 1645-1660, 2013.

[5] M. Ahmadi, S. Vakili and J. M. P. Langlois, "An Energy-Efficient Accelerator Architecture with Serial Accumulation Dataflow for Deep CNNs," in *IEEE International New Circuits and Systems Conference*, 2020.

[6] W. Shi, J. Cao, Q. Zhang, Y. Li and L. Xu, "Edge Computing: Vision and Challenges," *IEEE Internet of Things Journal,* vol. 3, pp. 637-646, 10 2016.

[7] Renata, "CR2032 datasheet," 2019. [Online]. Available: https://www.renata.com/fileadmin/downloads/productsheets/lithium/3V_lithium/CR2032_MFR.pdf. [Accessed 08 01 2021].

[8] Y. Lee et al., "A Modular 1 mm³Die-Stacked Sensing Platform With Low Power I²C Inter-Die Communication and Multi-Modal Energy Harvesting," *IEEE Journal of Solid-State Circuits,* vol. 48, pp. 229-243, 1 2013.

[9] P. Mayer, M. Magno and L. Benini, "A Low Power and Smart Power Unit for Kinetic Self-Sustainable Wearable Devices," in *IEEE International Conference on Electronics, Circuits and Systems*, 2020.

[10] M. Bedier and D. Galayko, "Multiple energy-shot load interface for electrostatic vibrational energy harvesters," in *2016 IEEE International New Circuits and Systems Conference*, 2016.

[11] N. Jose, N. John, P. Jain, P. Raja, T. V. Prabhakar and K. J. Vinoy, "RF powered integrated system for IoT applications," in *IEEE International New Circuits and Systems Conference*, 2015.

[12] P. Mayer, M. Magno and L. Benini, "Smart Power Unit—mW-to-nW Power Management and Control for Self-Sustainable IoT Devices," *IEEE Transactions on Power Electronics,* vol. 36, pp. 5700-5710, 2021.

[13] J. Deng et al., "Energy-Autonomous MCU Operating in sub-VT Regime with Tightly-Integrated Energy-Harvester," in *IEEE/ACM Int. Symposium on Low Power Electronics and Design*, 2019.

[14] Xiaomi, "Xiaomo AI door bell overview," 2019. [Online]. Available: https://www.xiaomitoday.com/2019/10/29/xiaomi-xiaomo-mijia-ai-face-identifcation-1080p-door-bell/. [Accessed 09 01 2021].

[15] Google, "Google Clips specifications," 2017. [Online]. Available: https://support.google.com/googleclips/answer/7545447?hl=en. [Accessed 21 05 2019].

[16] Orcam, "Orcam MyEye 2 specifications," 2020. [Online]. Available: https://www.orcam.com/en/myeye2/specification/. [Accessed 09 02 2021].

[17] S. Benninger, M. Magno, A. Gomez and L. Benini, "EdgeEye: A Long-Range Energy-Efficient Vision Node For Long-Term Edge Computing," in *International Green and Sustainable Computing Conference*, 2019.

[18] K. Bong, S. Choi, C. Kim, D. Han and H. J. Yoo, "A Low-Power Convolutional Neural Network Face Recognition Processor and a CIS Integrated With Always-on Face Detector," *IEEE Journal of Solid-State Circuits,* vol. 53, pp. 115-123, 1 2018.

[19] CSEM, "ERGO 320 specifications," 2017. [Online]. Available: https://www.csem.ch/Doc.aspx?id=51766. [Accessed 20 05 2019].

[20] CSEM, "Machine Learning SoC," 2021. [Online]. Available: https://www.csem.ch/pdf/159793. [Accessed 07 01 2021].

[21] M. Horowitz, "1.1 Computing's energy problem (and what we can do about it)," in *IEEE International Solid-State Circuits Conference*, 2014.

[22] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv and Y. Bengio, "Quantized Neural Networks: Training Neural Networks with Low Precision Weights and Activations," *CoRR,* vol. abs/1609.07061, 2016.

[23] Sharp, "LS013B7DH05 specifications," 2020. [Online]. Available: https://www.sharpsma.com/products?sharpCategory=Memory%20LCD&p_p_parallel=0&sharpProductRecordId=1725803. [Accessed 10 02 2021].

[24] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj and L. Song, "SphereFace: Deep Hypersphere Embedding for Face Recognition," *CoRR,* vol. abs/1704.08063, 2017.

[25] G. B. Huang, M. Mattar, T. Berg and E. Learned-Miller, "Labeled Faces in the Wild," in *Workshop on Faces in 'Real-Life'*, Marseille, 2008.

[26] P. Jokic, S. Emery and L. Benini, "Improving Memory Utilization in Convolutional Neural Network Accelerators," *IEEE Embedded Systems Letters,* pp. 1-1, 2020.