The Brain Mimicking Visual Attention Engine: An 80x60 Digital Cellular Neural Network for Rapid Global Feature Extraction

Seungjin Lee, Kwanho Kim, Minsu Kim, Joo-Young Kim, and Hoi-Jun Yoo

Division of Electrical Engineering, School of Electrical Engineering and Computer Science, KAIST
Guseong-dong, Yuseong-gu, Daejeon, 305-701, Korea
E-mail: seungjin@eeinfo.kaist.ac.kr

Abstract

The Visual Attention Engine(VAE), an 80x60 digital Cellular Neural Network, rapidly extracts global features used as attentional cues to streamline detailed object recognition. A peak performance of 24GOPS is achieved by 120 processing elements (PE) shared by the cells. 2D Shift register based data transactions enable 93% PE utilization. Integrated within an object recognition SoC, the 4.5mm^2 VAE running at 200MHz improves object recognition frame rate by 83% while consuming just 84mW.

Introduction

Visual attention is a critical component of object recognition in the human brain that filters useful parts from the vast amount of incoming visual stimuli. Similarly, machine object recognition is evolving toward a two stage approach [1] in which attention is used to select regions of interest in an image for further detailed processing with local feature based methods, as shown in Fig. 1. Cellular Neural Networks (CNN) [2] are very suitable for the global feature extraction that is involved in attention, in contrast to local feature oriented multi-processor based vision chips [3] that suffer from high data access overhead when operating on global data. In this work, an 80x60 cell digital CNN for extracting global attentional cues is presented. Named the Visual Attention Engine (VAE), it is unique in that it is integrated with the parallel processors in a full object recognition SoC [4]. By taking on the role of the attentional "filter" depicted in Fig. 1, the VAE reduces the complexity of operations carried out by the parallel processors. In order to realize small area and high energy efficiency, shared processing elements (PE) are used and custom dynamic circuits are applied to the cells. Meanwhile, high performance is achieved through pipelined PE operation and shift register based data transactions.

VAE Architecture

CNN hardware implementations can avoid performance degradation by having at least the same number of cells as the input image. However, due to the large size of digital arithmetic blocks, previous digital CNN implementations [5] integrate only a small number of cells and require virtualization to handle larger image sizes. In this work, the VAE integrates 4,800 cells that each correspond to one pixel in the 80x60 input image while maintaining small area. This is possible by outsourcing the processing functions of the cells to a smaller number of shared PEs. Fig. 2 shows the VAE composed of 4 20x60 cell arrays placed around 2 linear arrays of 60 PEs. The cells only perform storage and inter-cellular communication functions to minimize area. The PEs are responsible for processing the cells' data, and each PE controls 2 read buses and 1 write bus that connect it to 40 cells. The VAE's controller generates signals for sequencing the operation of the cells and PEs and facilitates loading and reading of the cell contents through an on-chip network interface.

VAE Cell

Each VAE cell consists of a 6T SRAM based 4x8bit register file and one 8bit shift register as shown in the lower left part of Fig. 2. The register files store intermediate data and result data of CNN operation. Data in the shift register is initially loaded from the register file of the same cell, after which it can be shifted to adjacent cells. A shift operation on the entire chip requires only 1 cycle to complete.

It is critical that the cell area is as small as possible since the cells account for more than 60% of the entire area. Since all cells always shift in the same direction, one bidirectional wire per bit can be used between cells to save routing channels. This is accomplished by a dynamic NMOS pass-transistor based MUX/DEMUX scheme shown in Fig.3. In this scheme, the value of dynamic node D, which is precharged to VDD, is evaluated through one of many possible paths selected by the signals 'N_en', 'E_en', 'S_en', 'W_en', and 'load_en' before being latched by the pulsed latch. Compared to a static MUX/DEMUX design, this reduces cell area by 40%.

VAE Processing Element and CNN Operation

The PEs execute the functions required for CNN operation such as MAC, MUL, and ADDI, as wells as functions such as ABS, and MIN/MAX required for general image manipulation. Each PE is shared by a group of 40 cells through 2 read buses, which are single-ended to save routing resources, and 1 write bus, which is double-ended to reliably write to the SRAM based register files. The read buses are split into left and right segments at the PE to reduce wire capacitance. Fig. 4 illustrates the pipelined execution of the PEs. Pipelined operation necessitates read and write of register files in the same cycle. An asynchronous control circuit allocates the first half of each cycle for reading and the second half for writing. Thanks to the pipelined operation, 1 op/cycle throughput is achieved and it takes 42 cycles for the PEs to execute one instruction on the entire cell array.

The most time consuming process of digital CNN operation is calculating the weighted sum of neighborhood cell values. Fig. 5 visualizes an optimal method for finding the weighted sum that involves a spiraling shift sequence that can be straightforwardly extended to neighborhoods larger than the 3x3 neighborhood shown. Thanks to the efficient shift pattern and single cycle shift operations, data communication overhead is only 2.4% and 93% utilization of the PE pipelines is realized. For a complete iteration of a 3x3 CNN template, 858 cycles or 4.3 μs is required.

Implementation Results

The 4.5mm^2 VAE was fabricated as part of a 36 mm^2 object recognition SoC using a 0.13μm 8 metal logic CMOS technology. The cell arrays are custom designed to minimize area while the PE and controller blocks are synthesized. Power consumption is 84mW when running at 200MHz. The 120 PEs have a peak performance of 24GOPS and show utilization rate of 93% during CNN operation. Fig.7 shows the result of applying VAE to an object recognition SoC designed for real-time (>15fps) operation on 320x240 video input. The VAE rapidly performs a saliency based attention algorithm on the 80x60 reduced image and outputs a pixel map marking regions of interest. This map is used later by the parallel processors to filter the input image and reduce complexity of detailed local feature based object recognition. For object images with background clutter, the average number of local features is drastically reduced, increasing frame rate by 83% and reducing energy per frame by 45% without degrading recognition rate. The energy overhead of the VAE itself is only 0.2 mJ/frame thanks to the short processing time of 2.4ms/frame.

References

[1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", *IEEE Trans. PAMI*, vol. 20, pp. 1254-1259, 1998.

[2] L. O. Chua and L. Yang, "Cellular Neural Networks: Theory," *IEEE Trans. Circuits Syst.*, vol. 35, pp. 1257-1271, 1998.

[3] D. Kim, K. Kim, J.-Y. Kim, S. Lee and H.-J. Yoo, ``An 81.6 GOPS object recognition processor based on NoC and visual image processing memory," IEEE CICC, 2007, pp. 443-446.

[4] K. Kim et al., "A 125GOPS 583mW network-on-chip based parallel processor with bio-inspired visual attention engine," IEEE ISSCC 2008 session 16.2.

[5] P. Keresztes et. al, "An emulated digital CNN implementation," *Journal of VLSI Signal Processing*, vol. 23, pp.291-303.

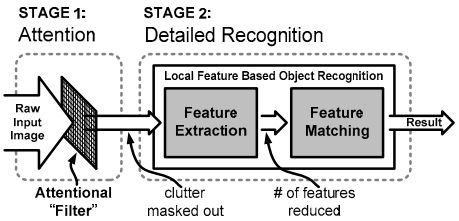


Fig. 1 Two stage object recognition

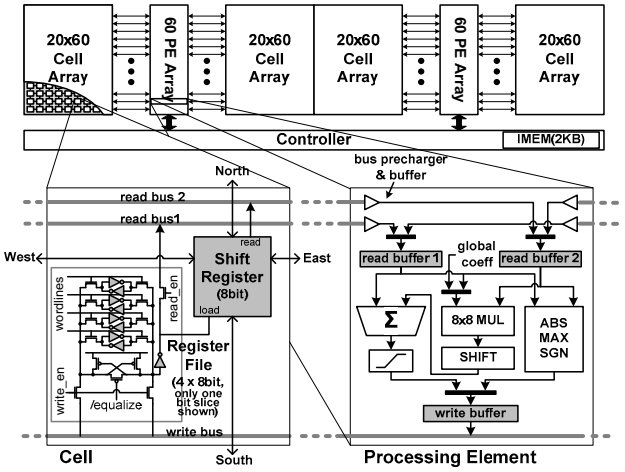


Fig. 2 Visual Attention Engine architecture

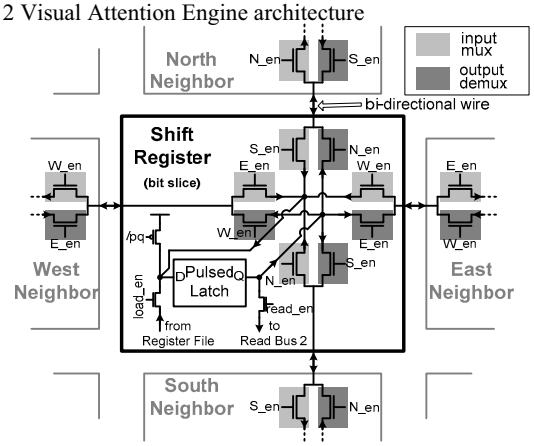


Fig. 3 Shift register dynamic multiplexer/demultiplexer

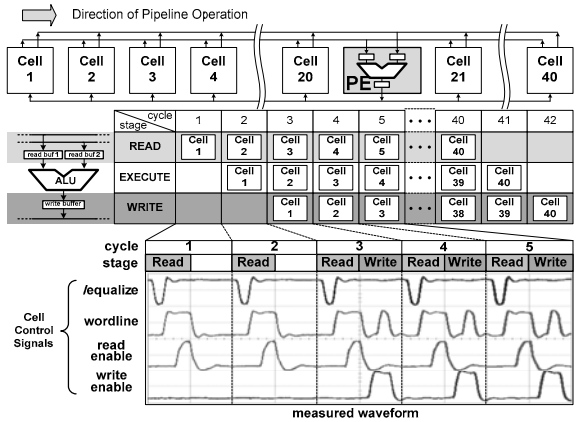


Fig. 4 Pipelined operation of processing element

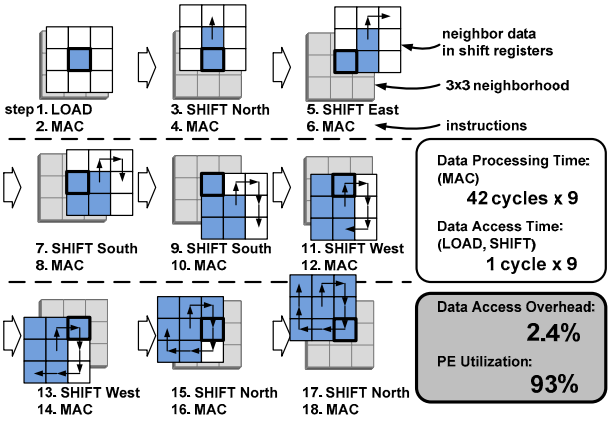
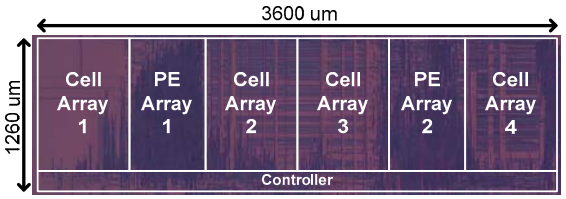
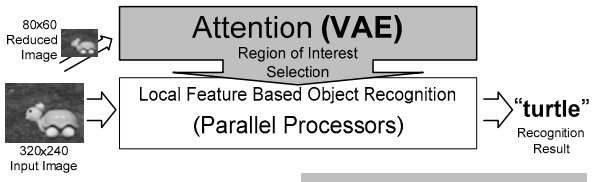


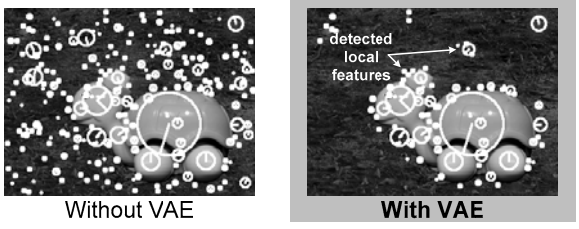
Fig. 5 Spiral shift sequence for CNN operation (weighted sum)



Process Technology	0.13μm CMOS
Area	4.5mm^2
Cell Dimensions	80x60
Number of PEs	120
Operating Frequency	200 MHz
Peak Performance	24 GOPS
Active Power Consumption	84 mW

Fig. 6 Implementation results





	Without VAE	With VAE	Improvement
Number of Extracted Local Features	279	**100**	**64%** ↓
Recognition Frame Rate	12	**22**	**83%** ↑
Energy per Frame	42mJ	**23mJ**	**45%** ↓

Fig. 7 Performance evaluation