

University

of Houston

Clear Lake

Learning-on-Chip: Facial Detection with Approximations of FPGA Computing

X. Yang, Y. Zhang, A. Gajjar, H. Schmoyer, N. Ly

Introduction

- ✤ Goal: to make AI a reality by offloading the complex learning applications to edge devices.
 - > On the hardware front, the majority amount of sensing data is becoming to overwhelm traditional data centers thus is expected to be performed at the network edge, due to the benefits of low-latency response, wise usage of bandwidth, as well as data security and privacy.
- > On the algorithm front, deep networks have made great strides in executing detection and recognition tasks with the computation of deep learning, requiring huge amount of hardware resource that is not affordable for edge devices such as surveillance cameras and wearable things. **Contributions:** Combining the merits of



IoT Hardware Ecosystem

Cloud service management Cloud datacenters / Amazon cloud / GizWits Challenges: Latency / Big data computation / Open APIs

FPGA Resource Cost

- **Resource cost**
 - Slice count: 32,309 for LUT and 38,130 for FF
 - > Power consumption: 714 mW Total (610 mW dynamic power and 104 mW static power)

Resource	Utilization	Available	Utilization %
LUT	32,309	63,400	50.96
LUTRAM	644	19,000	3.39
FF	38,130	126,800	30.07
BRAM	94	135	69.63
DSP	35	240	14.58
ΙΟ	41	210	19.52
MMCM	1	6	16.67

Chin D	owor							
Cillp P	OWEI							
		Dynam	nic: 0.6	0.610 W (85%)				
85%	13%	Clocks:	0.079 W	(13%)				
	22%	Signals:	0.135 W	(22%)				
	23%	Logic:	0.142 W	(23%)				
		BRAM:	0.143 W	(24%)				
	24%	DSP:	<0.001 W	(<1%)				
				0.400 \W	(470/)			

Edge/Fog Layer Traditional server / Gateway / Google Nest / TV set-top box Challenges: Open and programmable / Low latency Infrastructure Physical devices / Network function WiFi, BLE, NB-IoT / App hosting

Challenges: Wide area connection / Low energy / Secure

- Fig. 1 Cloud-edge-device computing system [1]
- hardware parallelism and the inherent error-tolerance of the learning algorithms. it offers an opportunity to minimize the energy cost corresponding to different quality constrains.
- \geq The project has a great potential to be used in some time-sensitive systems such as a facial recognition by surveillance cameras during an AMBER Alert.

Design Architecture

- ✤ A design structure of facial detection with Nexys 4 FPGA
 - > In the dotted blue box, we have developed a platform which is able to configure the OV7670 camera via the I2C controller and show images on monitor via VGA interface. The preliminary result is available on the http://sceweb.sce.uhcl.edu/xiaokun/#
 - > The Viola-Johns data-path, as depicted in the dotted orange, is workin-process by applying many approximations of FPGA computing

Fig. 4 (a) FPGA slice cost [3]



Fig. 4 (b) FPGA power consumption [3]



Fig. 4 (c) FPGA power use (by percentage) [3]

A Case Study on Self-tuning Approximation

A case study of energy-quality (E-Q) tradeoff - color to grayscale converter (12 different approximations of design)



> ×1.37 --2.16 energy savings - quality constraints (<3%) > ×2.31-2.50 energy savings - quality constraints (3% - 7.5%) Approximate Approximate **Multiplication** Addition Cout0 = (b&c) | (a&b) | (a&c); $gs = 0.2989 \times r + 0.587 \times g + 0.114 \times b;$ Sum0 = (~a&b&~c) | (a&b&c) | (a&~b&~c) | (~a&~b&c) | $gs = (1 \times r + 1 \times g)/2$; gs = (2xr + 5xg + 1xb)/8;Cout4 = a;gs = (10xr + 19xg + 3xb)/32;Sum4 = b; $gs = (77 \times r + 150 \times g + 29 \times b)/256;$

Fig. 5 (a) 12 approximations of the design on color-to-grayscale converter [2]



Fig. 2 FPGA design architecture of facial detection [1]

A Facial Detection Platform on FPGA



Fig. 3 (a) A platform of facial detection [4] Fig. 3 (b) Demonstration of Viola-Johns Algo.

Reference

[1] X. Yang and J. Andrian, "An Advanced Bus Architecture for AES-Encrypted High-Performance *Embedded Systems,*" *US20170302438A1, Oct. 2017.*

[2] X. Yang, M. Fan, Q. Han, etc., "Exploiting Energy-Quality (E-Q) Tradeoffs on Approximate FPGA Designs of Scalable Sequential Circuits," Integration, the VLSI Journal, Under Review, 2018. [3] A. Gajjar, X. Yang, etc., "An FPGA Synthesis of Face Detection Algorithm using HAAR Classifiers," Intl. Conference on Algorithms, Computing and Systems (ICACS2018), Accepted, In Press, 2018. [4] Y. Zhang, X. Yang, etc., "Exploring Slice-Energy Saving on An Video Processing FPGA Platform with Approximate Computing," Intl. Conference on Algorithms, Computing and Systems (ICACS2018), Accepted, In Press, 2018.