

2018 1st Workshop on Energy Efficient Machine Learning and Cognitive Computing for Embedded Applications (EMC2 2018)

**Williamsburg, Virginia, USA
25 March 2018**



**IEEE Catalog Number: CFP18Q97-POD
ISBN: 978-1-5386-7368-3**

**Copyright © 2018 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP18Q97-POD
ISBN (Print-On-Demand):	978-1-5386-7368-3
ISBN (Online):	978-1-5386-7367-6

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2018 1st Workshop on Energy Efficient Machine Learning and Cognitive Computing for Embedded Applications **EMC2 2018**

Table of Contents

EMC2 2018 Preface	vii
--------------------------	------------

Technical Papers

Keynote Abstract: Safety and Security at the Heart of Autonomous Driving	1
<i>Kamal Khouri (NXP Semiconductors)</i>	
Invited Talk Abstract: Challenges and Solutions for Embedding Vision AI	2
<i>Charles Qi (Cadence)</i>	
Invited Talk Abstract: Introducing ReQuEST: An Open Platform for Reproducible and Quality-Efficient Systems-ML Tournaments	3
<i>Grigori Fursin (dividiti and cTuning Foundation)</i>	
A Case for Dynamic Activation Quantization in CNNs	4
<i>Karl Taht (University of Utah), Surya Narayanan (University of Utah), and Rajeev Balasubramonian (University of Utah)</i>	
A High Efficiency Accelerator for Deep Neural Networks	9
<i>Aliasger Zaidy (FWDNXT Inc.), Andre Xian Ming Chang (FWDNXT Inc.), Vinayak Gokhale (Google Inc.), and Eugenio Culurciello (FWDNXT Inc.)</i>	
A Quantization-Friendly Separable Convolution for MobileNets	14
<i>Tao Sheng (Qualcomm Canada Inc.), Chen Feng (Qualcomm Canada Inc.), Shaojie Zhuo (Qualcomm Canada Inc.), Xiaopeng Zhang (Qualcomm Canada Inc.), Liang Shen (Qualcomm Canada Inc.), and Mickey Aleksic (Qualcomm Technologies Inc.)</i>	
Deep Learning Inference on Embedded Devices: Fixed-Point vs Posit	19
<i>Seyed Hamed Fatemi Langroudi (Rochester Institute of Technology), Tej Pandit (Rochester Institute of Technology), and Dhireesha Kudithipudi (Rochester Institute of Technology)</i>	
Efficient Compiler Code Generation for Deep Learning Snowflake Co-Processor	24
<i>Andre Xian Ming Chang (FWDNXT Inc.), Aliasger Zaidy (FWDNXT Inc.), and Eugenio Culurciello (FWDNXT Inc.)</i>	

Event Prediction in Processors Using Deep Temporal Models .29.....	
<i>Tharindu Mathew (SRI International), Aswin Raghavan (SRI International), and Sek Chai (SRI International)</i>	
Moving CNN Accelerator Computations Closer to Data .34.....	
<i>Sumanth Gudaparthi (University of Utah), Surya Narayanan (University of Utah), and Rajeev Balasubramonian (University of Utah)</i>	
Author Index .39	