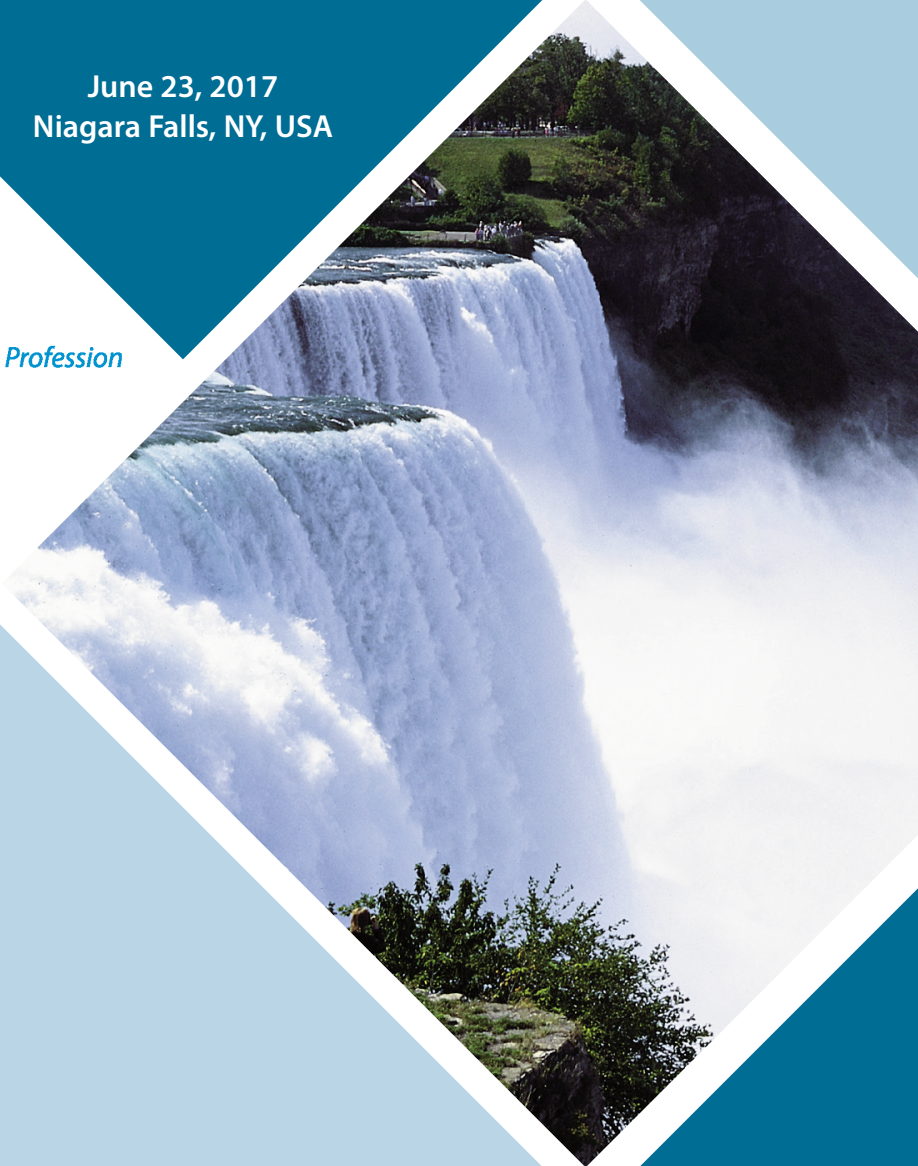


June 23, 2017  
Niagara Falls, NY, USA



Association for  
Computing Machinery

*Advancing Computing as a Science & Profession*



# EMDL'17

Proceedings of the 1st International Workshop on  
Deep Learning for Mobile Systems  
and Applications

*Sponsored by:*

**ACM SIGMOBILE**

*In-cooperation with:*

**ACM SIGOPS**

*Co-located with:*

**MobiSys 2017**



**Association for  
Computing Machinery**

*Advancing Computing as a Science & Profession*

**The Association for Computing Machinery**  
**2 Penn Plaza, Suite 701**  
**New York, New York 10121-0701**

Copyright © 2017 by the Association for Computing Machinery, Inc. (ACM). Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyright for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permission to republish from: [permissions@acm.org](mailto:permissions@acm.org) or Fax +1 (212) 869-0481.

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through [www.copyright.com](http://www.copyright.com).

**ISBN:** 978-1-4503-4962-8

Additional copies may be ordered prepaid from:

**ACM Order Department**  
PO Box 30777  
New York, NY 10087-0777, USA

Phone: 1-800-342-6626 (USA and Canada)  
+1-212-626-0500 (Global)  
Fax: +1-212-944-1318  
E-mail: [acmhelp@acm.org](mailto:acmhelp@acm.org)  
Hours of Operation: 8:30 am – 4:30 pm ET

Printed in the USA.

# 1<sup>st</sup> International Workshop on Embedded and Mobile Deep Learning

## Workshop Foreword

It is with true pleasure that we welcome you to the *First International Workshop on Embedded and Mobile Deep Learning*. This is a first-of-its-kind workshop for the mobile computing and systems community, and aims to offer a forum for both academics and practitioners alike to explore, discuss, and debate early findings and new vectors of research within this exciting area of machine learning as it applies to embedded and mobile systems.

This workshop is largely motivated by the breakthroughs seen in the field of deep learning that are transforming how sensor data (e.g., images, audio, and even accelerometers and GPS) is modeled during the extraction of high-level information needed by the latest in sensor-driven systems like smartphone apps and wearable devices. Today, the state-of-the-art in computational models that, for example, recognize a face, track user emotions, or monitor physical activities are increasingly based on deep learning principles and algorithms. Unfortunately, deep models typically exert severe demands on local device resources and this conventionally limits their adoption within mobile and embedded platforms. As a result, in far too many cases existing systems process sensor data with machine learning methods that have been superseded by deep learning years ago.

As the robustness and quality of sensory perception and reasoning is critical to mobile computing, it is important for this community to begin the careful study of two core technical questions. First, how should deep learning learning principles and algorithms be applied to sensor inference problems that are central to this class of computing? This includes a combination of applications of learning some of which are familiar to other domains (such as the processing image and audio), in addition to those more uniquely tied to wearable and mobile systems (e.g., activity recognition). Second, what is required for current -- and future -- deep learning advances to be either simplified or efficiently integrated into a variety of mobile resource-constrained systems? The scope of this workshop at MobiSys 2017 on embedded and mobile deep learning spans these two broad themes.

Putting together this workshop has been fundamentally a team endeavor. We would like to begin by thanking all the authors for contributing the technical heart of this meeting. We are also greatly appreciative of the invited keynote speakers bringing in particular insights and perspectives on this topic from outside the mobile systems domain (such as, novel deep architectures and hardware innovations). We are similarly grateful to the program committee, who took the time to review the papers and provide feedback to authors. Finally, we thank ACM SIGMOBILE for their generous support.

**Nic Lane**

*co-Program Chair  
University College London  
and Nokia Bell Labs, UK*

**Pete Warden**

*co-Program Chair  
Google Brain, USA*

# Table of Contents

<b>1st International Workshop on Embedded and Mobile Deep Learning – Organization .....</b>	<b>vi</b>
---	-----------

## **Session: Optimizations for Constrained Computing**

• <b>MobiRNN: Efficient Recurrent Neural Network Execution on Mobile GPU .....</b>	<b>1</b>
Qingqing Cao, Niranjan Balasubramanian, Aruna Balasubramanian ( <i>Stony Brook University</i> )	
• <b>RSTensorFlow: GPU Enabled TensorFlow for Deep Learning on Commodity Android Devices .....</b>	<b>7</b>
Moustafa Alzantot, Yingnan Wang, Zhengshuang Ren, Mani B. Srivastava ( <i>University of California, Los Angeles</i> )	
• <b>Pruning Filters and Classes: Towards On-Device Customization of Convolutional Neural Networks .....</b>	<b>13</b>
Jia Guo, Miodrag Potkonjak ( <i>University of California, Los Angeles</i> )	

## **Session: Novel Applications and End-to-end Systems**

• <b>Practical Processing of Mobile Sensor Data for Continual Deep Learning Predictions .....</b>	<b>19</b>
Kleomenis Katevas ( <i>Queen Mary University of London &amp; Telefónica Research</i> ), Ilias Leontiadis, Martin Pielot, Joan Serra ( <i>Telefónica Research</i> )	
• <b>Design and Implementation of the Vehicular Camera System using Deep Neural Network Compression .....</b>	<b>25</b>
Beomjun Kim, Yongsu Jeon, Heejin Park, Dongheon Han, Yunju Baek ( <i>Pusan National Univeristy</i> )	

<b>Author Index .....</b>	<b>31</b>
---------------------------	-----------

# 1<sup>st</sup> International Workshop on Embedded and Mobile Deep Learning

## Workshop Organizers

**Program Chairs:** Nic Lane (*University College London and Nokia Bell Labs, UK*)  
Pete Warden (*Google Brain, USA*)

**Program Committee:** Sourav Bhattacharya (*Nokia Bell Labs*)  
Petko Georgiev (*Google DeepMind*)  
Song Han (*Stanford University*)  
Samir Kumar (*Microsoft Ventures*)  
Robert LiKamWa (*Arizona State University*)  
Youngki Lee (*Singapore Management University*)  
Erran Li (*Uber*)  
Laurens van der Maaten (*Facebook AI Research*)  
Matthai Phillipose (*Microsoft Research*)  
Thomas Ploetz (*Georgia Tech*)  
Heather Zheng (*UC Santa Barbara*)  
Lin Zhong (*Rice University*)

**Sponsor:**



Association for  
Computing Machinery