Mobile Sensing Through Deep Learning

Xiao Zeng
Michigan State University
zengxia6@msu.edu

ABSTRACT

Today, mobile devices are equipped with powerful processors along with various on-device sensors. Over the past few years, deep learning has become the dominant approach in the field of machine learning due to its impressive performance. We envision that in the near future, powered by deep learning, mobile devices will become more intelligent and revolutionize a wide range of applications. In this paper, we discuss the challenges of enabling deep learning on mobile platforms. Our work is to propose a deep learning framework that achieves state-of-the-art performance with low overhead on resource-limited mobile platforms. Our preliminary results show that deep learning can efficiently solve object recognition problem under noisy real world environment.

Keywords

Deep Neural Networks; Model Compression; Mobile Sensing

1. INTRODUCTION

Modern mobile devices such as smartphones and wearables are equipped with various on-device sensors. These devices serve as powerful tools for people to retrieve information from and interact with the physical world. With the development of system-on-chip, smartphones are now able to process computation tasks. The integration of sensors and processor on a uniform platform makes mobile devices a perfect tool to solve many problems, from activity recognition to speech recognition. For instance, the Google Translate mobile app can accurately translate text on books, road signs, and menus from one language into another.

In the past few years, deep learning has surpassed traditional methods and has become the dominant approach in machine learning due to its impressive capability of handling large-scale learning problems. Deep learning-based approaches have achieved state-of-the-art performance in solving a variety of computer vision problems such as object recognition, face recognition and video classification. More recently, some interesting applications have been developed based on deep learning, such as audio generation [8] and style transferring [3].

MobiSys'17 PhD Forum, June 19, 2017, Niagara Falls, NY, USA.

ACM ISBN 978-1-4503-4957-4/17/06..

DOI: http://dx.doi.org/10.1145/3086467.3086476

Implementing these deep learning-based approaches on mobile devices could bring tremendous benefits to human life, such as language translation and video surveillance, leading to a more intelligent society. However, deep learning-based approaches are very expensive in terms of computation, memory, and energy consumption compared to traditional methods such as support vector machine (SVM). As such, given the resource constraints of mobile devices, most of the existing mobile systems that require a lot of computational resources or memory offload the workload to the cloud. While offloading is a practically feasible solution, it suffers from network latency and privacy issues. A reasonable alternative is to optimize an existing large model for reducing computational and memory cost and obtain a concise mobilized model in order to achieve on-device mobile deep learning. In my dissertation research, I will explore and investigate the feasibility of applying deep learning-based approach to solve existing problems on mobile platforms. Specifically, I will focus on mobile image and audio processing where deep learning has been proven to outperform traditional methods.

2. CHALLENGES

The challenges of implementing deep leaning approaches on mobile devices mainly come from three aspects.

Data Asymmetry: Deep Learning is data-driven. The performance of deep learning modeld relies on the training data which is supposed to be similar to test data. Unfortunately, there is a significant gap between training data and data to be processed in practice. As a result, it suffers from a significant performance drop when deep learning is applied in real world. This problem is exacerbated when deep learning is applied on mobile devices. This discrepancy is caused by variations in acquisition devices and acquisition environments. For example, in mobile environment, the illumination is changing vastly and rapidly from indoor to outdoor. Data collected by accelerometer and gyroscope is affected by the deployment position and sampling precision of the device.

Multi-Modality Sensing: Even though deep learning achieves unprecedented success in computer vision, natural language processing, these methods do not perform particularly well when dealing with heterogeneous sensor data. For example, in many machine learning regression and classification problems, the feature vector contains not only categorical attributes but also continuous values. Typical sensors on smartphones include GPS, gyroscope, light sensor, microphone and cameras. Data obtained by these sensors is heterogeneous and different in both sampling frequency and scale. How to integrate these data and effectively use them as input in deep learning for machine learning tasks remains a challenging problem to mobile deep learning.

Resource Constraints of Mobile Devices: Many mobile systems only perform data sensing tasks and offload computation to the cloud. Due to network vulnerability and privacy issues, cloud-based solution may not be applicable in many scenarios where non-cloud-based system is the only choice. However, deep learning that achieves breakthrough performance requires high computational and memory cost. Although current smartphones are equipped with powerful computing resources, it still cannot support such computationally intensive methods. Hence, the gap between high computational requirement and the limited computing resources is another challenge for mobile deep learning.

3. RELATED WORK

Existing works that are focused on mobile deep learning can be classified into two groups. This first group attempt to developing a framework that uniformly leverage the sensing data for deep learning. Radu *et al.* proposed a multi-modal deep learning system that uses Restricted Boltzmann Machine for activity recognition [6]. Bhattacharya *et al.* also developed a smartwatch-based activity recognition system based on deep learning [1] . Yao *et al.* tried to integrate convolutional and recurrent neural networks to exploit local interactions within each mobile sensor and extract temporal relationships [9]. Chang *et al.* came up with a deep learning framework that aims at creating a multi-resolution deep embedding function for data mining tasks [2]. The second group aims at optimizing and compressing deep learning models. Model compression for deep networks has become a popular research area in recent years due to the significant demand on running deep learning models on resource-limited platforms. Rastegari *et al.* developed XNOR-Networks that approximate convolutions using binary operations [7]. Zhou *et al.* even used low bitwidth for weight, activation and gradient representation [10]. Han *et al.* proposed the network pruning method that converts the weight matrices into sparse matrices by replacing parameters that are a threshold with zeros [4]. Lane *et al.* exploited the redundancy within weights to derive approximations using Singular Vector Decomposition or Sparse Coding that significantly reduce the model size [5].

4. INITIAL RESULT

To address the aforementioned challenges, we propose to develop a mobile deep learning framework, as shown in Figure 1. Specifically, this framework is on top of mobile hardware, involving three important modules. The bottom module is Data Provider, which is responsible for managing and providing data. The middle module is Model Provider, which is responsible for providing optimized deep learning models. The top module is Task Provider, which defines the training and inference details. It is also responsible for task management. Our proposed approach distinguishes from other works in that it is a complete mobile deep learning framework that deals with data acquisition, data management, model management and deep learning task scheduling.

Our initial effort includes a training scheme that aims at solving data asymmetry problem and a model compression technique called Teacher-Student Optimization Framework that compresses large deep learning model. The training scheme operates in Task Provider module and the compression technique operates in Model Provider Module. Experimental results show that the training scheme successfully mitigates data asymmetric problem and the compression technique can lower the computational and storage cost of deep learning model.

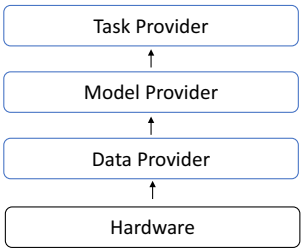

Figure 1: Illustration of the proposed framework.

5. REFERENCES

[1] S. Bhattacharya and N. D. Lane. From smart to deep: Robust activity recognition on smartwatches using deep learning. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2016 IEEE International Conference on*, pages 1–6. IEEE, 2016.

[2] S. Chang, W. Han, J. Tang, G.-J. Qi, C. C. Aggarwal, and T. S. Huang. Heterogeneous network embedding via deep architectures. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 119–128. ACM, 2015.

[3] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2414–2423, 2016.

[4] S. Han, J. Pool, J. Tran, and W. Dally. Learning both weights and connections for efficient neural network. In *Advances in Neural Information Processing Systems*, pages 1135–1143, 2015.

[5] N. D. Lane, S. Bhattacharya, P. Georgiev, C. Forlivesi, L. Jiao, L. Qendro, and F. Kawsar. Deepx: A software accelerator for low-power deep learning inference on mobile devices. In *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pages 1–12. IEEE, 2016.

[6] V. Radu, N. D. Lane, S. Bhattacharya, C. Mascolo, M. K. Marina, and F. Kawsar. Towards multimodal deep learning for activity recognition on mobile devices. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 185–188. ACM, 2016.

[7] M. Rastegari, V. Ordonez, J. Redmon, and A. Farhadi. Xnor-net: Imagenet classification using binary convolutional neural networks. *arXiv preprint arXiv:1603.05279*, 2016.

[8] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. *CoRR abs/1609.03499*, 2016.

[9] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. F. Abdelzaher. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. *CoRR*, abs/1611.01942, 2016.

[10] S. Zhou, Y. Wu, Z. Ni, X. Zhou, H. Wen, and Y. Zou. Dorefa-net: Training low bitwidth convolutional neural networks with low bitwidth gradients. *arXiv preprint arXiv:1606.06160*, 2016.