

Case ID:M17-140P

Published: 3/21/2018

Inventors

Shimeng Yu

Rui Liu

Contact

Shen Yan
shen.yan@skysonginnovations.
com

Flexible and Efficient XNOR Circuit Architecture for Neural Network Based Deep Learning

Background

Neuro-inspired machine deep learning have demonstrated that they can perform complex tasks such as image and speech classification, object recognition, and real-time decision-making. The state-of-the-art deep learning algorithms employ neural networks which have millions of parameters. Deploying such deep neural networks onto mobile platforms is challenging and may be limited due to power requirements that are available on chip memory resources. In such context, there is a need for a novel design that prunes and compresses the neural network such that the number parameters may be decreased, while maintaining system performance.

Invention Description

Researchers at ASU have developed a new circuit architecture based upon a special case of a binary neural network, XNOR-Networks. The new architecture demonstrates comparable recognition accuracy when compared to deep neural networks, while still providing high precision image recognition. This circuit architecture is capable of computing in an analog and parallel fashion. Therefore, it can advantageously greatly reduce both latency and power consumption. The nature of the circuit architecture is such that it is flexible and may be deployed across a wide array of applications. The design may allow for the implementation of deep learning applications in smaller devices such as phones or tablets where power may be limited. This would greatly increase the number of possible applications, as well as increase the data inputs available, resulting in stronger and more accurate neural networks.

Potential Applications

- Electronics for Deep Learning / Artificial Intelligence
- Mobile/Handheld Applications of AI

Benefits and Advantages

- Efficient- The new circuit architecture decreases power consumption by integrating the XNOR function into the circuitry. Additionally, by reducing the

power consumption these SRAM architectures

could be utilized in handheld devices.

- Adaptable- The proposed circuit architecture is flexible and can be adapted for multiple purposes.

[Professor Yu's Directory Webpage](#)