Skysong Innovations
Advancing the Arizona State University Knowledge Enterprise

1475 N. Scottsdale Road, Suite 200
Scottsdale, AZ 85287-3538
Phone: 480 884 1996 Fax: 480 884 1984

Case ID:M21-253P
Published: 8/1/2022

Inventors

Deliang Fan
Adnan Siraj Rakin
Li Yang
Chaitali Chakrabarti
Yu Cao

Contact

Shen Yan
shen.yan@skysonginnovations.com

Binary Neural Network for Improved Accuracy and Defense Against Bit-Flip Attacks

-Recently, Deep Neural Networks (DNNs) have been deployed in many safety-critical applications. The security of DNN models can be compromised by adversarial input examples, where the adversary maliciously crafts and adds input noise to fool a DNN model. The perturbation of model parameters (e.g., weight) is another security concern, one that relates to the robustness of the DNN model itself. An adversarial weight attack occurs when an attacker perturbs target DNN model parameters in computing hardware to achieve malicious goals. Among the prevalent adversarial weight attacks is bit-flip attack (BFA), which has been proven to be highly successful in hijacking DNN functionality (e.g., degrading accuracy to as low as random guessing) by flipping an extremely small number (e.g., tens out of millions) of weight memory bits. Current methods to combat BFAs include binary weight neural networks. However, a challenging optimization problem exists for binary neural networks where a lower bit-width network comes with improved robustness but at the cost of lower accuracy. Thus, there is a need for a robust and accurate binary neural network (with both binary weight and activation) to simultaneously defend BFAs and improve clean model accuracy.

Researchers at Arizona State University (ASU) have developed a neural network that significantly improves robustness against bit-flip attack while also increasing clean accuracy (i.e. accuracy with no attack). First, the ASU network takes advantage of binary neural network (BNN) capability in providing improved resistance against BFA, through completely binarizing both activations and weights of every DNN layer. Second, to address the clean model accuracy loss, the ASU network selectively grows the output channels of every BNN layer to recover accuracy loss. Moreover, apart from recovering accuracy, increasing the channel size can also help to resist BFA attack.

Related publication: RA-BNN: Constructing Robust & Accurate Binary Neural Network to Simultaneously Defend Adversarial Bit-Flip Attack and Improve Accuracy

Potential Applications:

- Safety-critical deep neural networks
- Autonomous vehicles
- Cybersecurity
- Artificial Intelligence

Benefits and Advantages:

- Over 125× improvement in resistance to bit-flip attack
- 2-8% improvement of clean accuracy (i.e., with no adversarial perturbation)