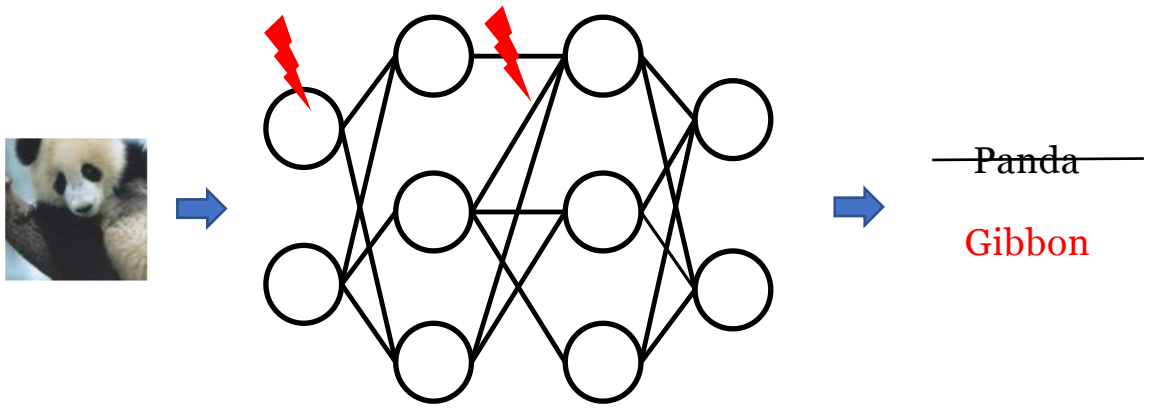
On the Modular Redundancy of Deep Neural Networks

Yu Li, Yannan Liu, Luo Bo, Ye Tian, Min Li, and Qiang Xu,
Computer Science and Engineering,
The Chinese University of Hong Kong

Presented by: Jinhong LI
07/12/2018

Fault Injection Attacks (FIA) on DNN

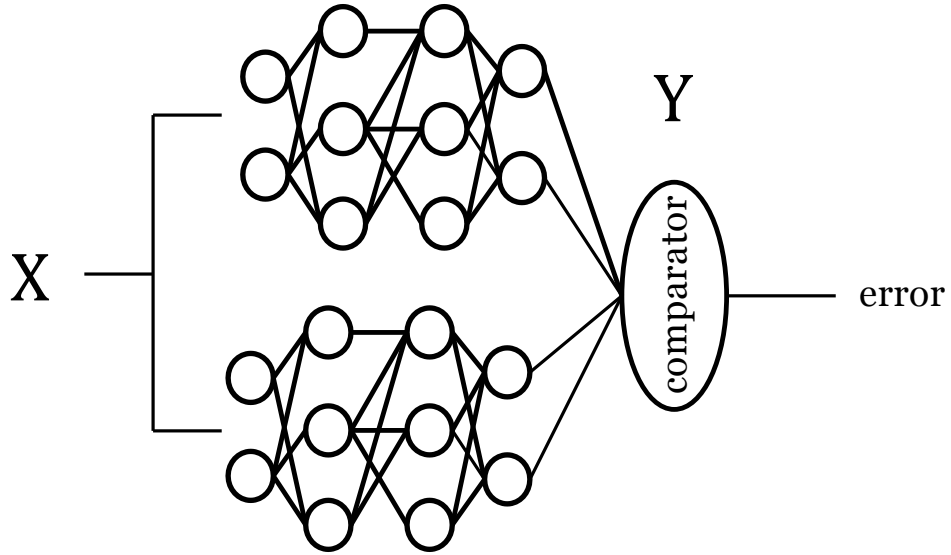


- FIA on **inputs** (e.g., adversarial example attacks [1])
- FIA on network **model** (e.g., SBA [2])

Runtime Integrity Checking

[1] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," arXiv preprint arXiv:1412.6572, 2014
[2] Y. Liu, L. Wei, B. Luo, and Q. Xu. Fault injection attack on deep neural network. In *2017 IEEE/ACM International Conference on Computer- Aided Design (ICCAD)*, pages 131–138, Nov 2017.

Dual Modular Redundancy (DMR) for DNN

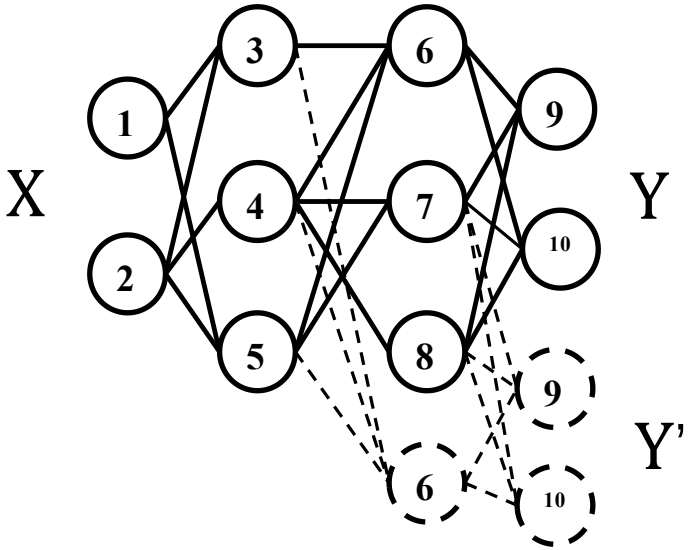


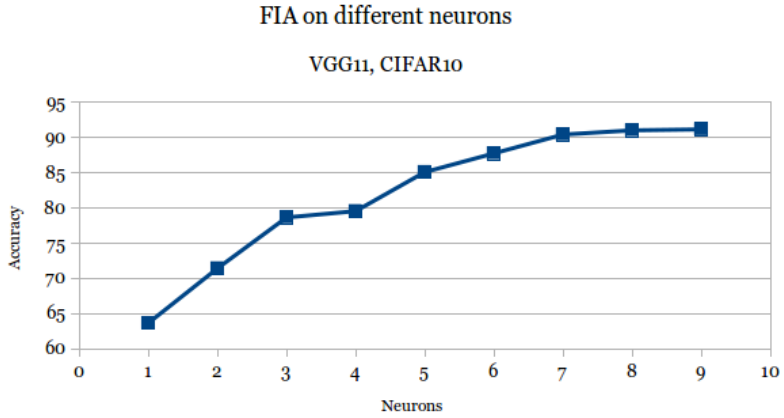
- No FIAs on one model will be missed
- power consumption & chip area ×**2**

What if the chip resources are limited?

Fine-grained DMR Net

- Redundancy at the neuron-level instead of network-level
- Motivation
  - DNN inherently tolerates some injected faults
  - Complete DMR is not always necessary.





More vulnerable neurons have higher priority to be protected!

* Neuron 9 and 10 are required to be duplicated to protect neuron 6

Fine-grained DMR Net (Cont.)

- Given maximum N redundant neurons, we do
  - Critical analysis
    - Evaluate and assign each neuron with a vulnerability value *vul*
    - e.g., Neuron with higher weight sum has higher *vul*
  - DMR Net topology construction
    - Select N neurons with higher *vul* than others
    - Handle connections among original neuron and the introduced dummy neurons.
      - Selected neurons
      - Unselected neruons
      - Dummy neurons
  - DMR Net parameter fine-tuning
    - To protect the unselected neurons
    - We try to differentiate the influence of unselected neurons to the two networks

Preliminary Results

We lunch FIA on parameters:

- FIA is on one random bit of one random parameter
- *dup_ratio*: portion of duplicated neurons
- We exam the portion of FIAs that have been tolerated, detected, and missed
- $miss\ ratio = \frac{Missed\ FIAs}{(Missed+Detected)\ FIAs}$

- Conclusions
  - We are able to increase the DNN's security levels using limited resources
  - We do not need 100% duplication to achieve near zero miss ratio

VGG11, CIFAR10				
dup_ratio	tolerated	detected	missed	miss_ratio
1.0	99.67	0.33	0	0.00
0.9	99.69	0.3	0.02	0.06
0.8	99.72	0.24	0.04	0.14
0.7	99.76	0.14	0.11	0.44
0.6	99.78	0.08	0.13	0.62
0.5	99.8	0.05	0.16	0.76
0.4	99.8	0.02	0.18	0.90
0.3	99.8	0.01	0.19	0.95
0.2	99.8	0	0.19	1.00
0.1	99.8	0	0.2	1.00
0.0	99.67	0	0.33	1.00

Future Work…

- To reduce the miss ratio
  - Critical analysis algorithms
    - Correct and accurate
  - Neuron selection strategy
    - How many neurons should be selected in each layer?
  - Retraining objective
    - How to improve the detection ability of unprotected neurons while keeping the protection for vulnerable neurons?
- To validate our results on different network structures and datasets

Thank you!

Q & A

If you have any questions, please contact:

yuli@cse.cuhk.edu.hk