17th International Conference on Innovative Computing, Information and Control (ICICIC2023)
Kumamoto, Japan, August 29-31, 2023

Paper ID: ICICIC2023-SS07-02

Improving Certified Robustness of Sensing-Reasoning Models via Lipschitz Neural Networks

Yicong Li, Kuanjiu Zhou*, Usman Arshad, Shangzhao Zhai and Mingyu Fan

School of Software, Dalian University of Technology, Dalian, P. R. China
*Corresponding author: zhoukj@dlut.edu.cn

Introduction

Sensing-reasoning models require the participation of power-hungry Gaussian noise training to achieve random smoothing of each sensing-CNN in exchange for certifiable robustness of the components. Moreover, the robustness of this certification approach to l_∞ perturbations is deficient. To avoid the sacrifice of this high certification cost and the adaptation defect of l_∞ perturbations, this paper proposes an optimization model based on Lipschitz property, which bypasses the perturbation radius by introducing Lipschitz neural networks as sensing components and solving the perturbation radius directly with norm-bounded affine transformations and order statistics property. The random smoothing training of CNN is used to trade-off the certification overhead and classification performance. Our model also obtains excellent l_∞ perturbations certified accuracy and exhibits stable defense against adversarial attacks.

Methodologies

Our model consists of a perception module, which contains binary classifiers constructed by a Lipschitz neural network, and an inference component, which is constructed by a probabilistic graphical model, and an inference module, which we apply the Markov logic network (MLN).

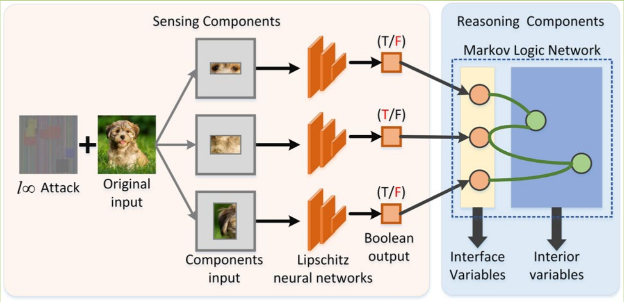


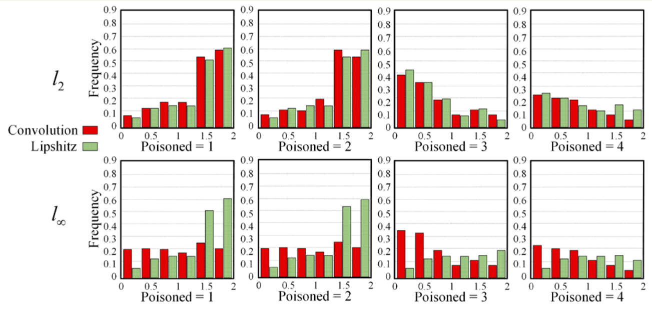
Figure 1. The sensing-reasoning model based on Lipschitz neural network

Figure 1 shows the overall architecture of our sensing-reasoning model based on Lipschitz's neural network. For each Lipschitz neural network, a unified framework is employed. This framework is essentially a fully connected neural network with H layers. The framework consists of three components: (i) parametric bounded affine transformation; (ii) Lipschitz unitary activation function; and (iii) sequential statistics.

Experiments

We trained 8 Lipschitz neural networks, and poisoned a certain size of training samples against attacks on 1 to 4 sensing components. We use the difference between the predicted probability minimum of the ground truth label and the probability maximum of the second candidate label as a metric to measure the defense capability of the model. To make the metric constant > 0, we add 1 to the difference.

We test the training time and certification time of the model on MNIST dataset and CIFAR-10 dataset using l_2 and l_∞ perturbation, respectively.



Our model significantly outperforms the CNN-based one in terms of training and validation time, thanks to the fact that Lipschitz neural network validation does not require propagation of output bounds as in IBP, and does not require adding perturbation samples to the training set, which saves 28.9% of training time, and 36.2% of validation time in CIFAR-10, and 28.9% of training time, and 78.2% of validation time in MNIST.

Conclusion

We propose a novel sensing-reasoning model based on the Lipschitz property, which builds on the original framework by 1) introducing a Lipschitz neural network as the perceptual component, bypassing the random smoothing-based CNN Gaussian noise training, and optimizing the computational complexity of robustness verification of the perceptual pipeline with better perturbation robustness at l_∞ perturbation, and 2) we provide robustness verification of the pipeline as a whole for the sensing-reasoning model.