Analysis of DNN Verification Techniques and Approaches

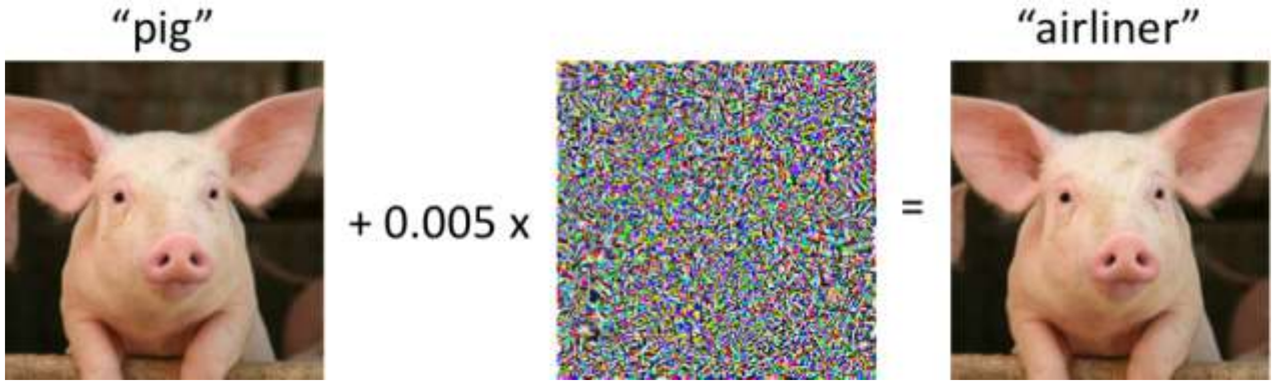
컴퓨터공학과 20180462 채승현

연구 목적

Rise in use of DNN (Deep Neural Network)
in safety-critical fields
(e.g. autonomous driving cars)

DNN weak against adversarial attacks

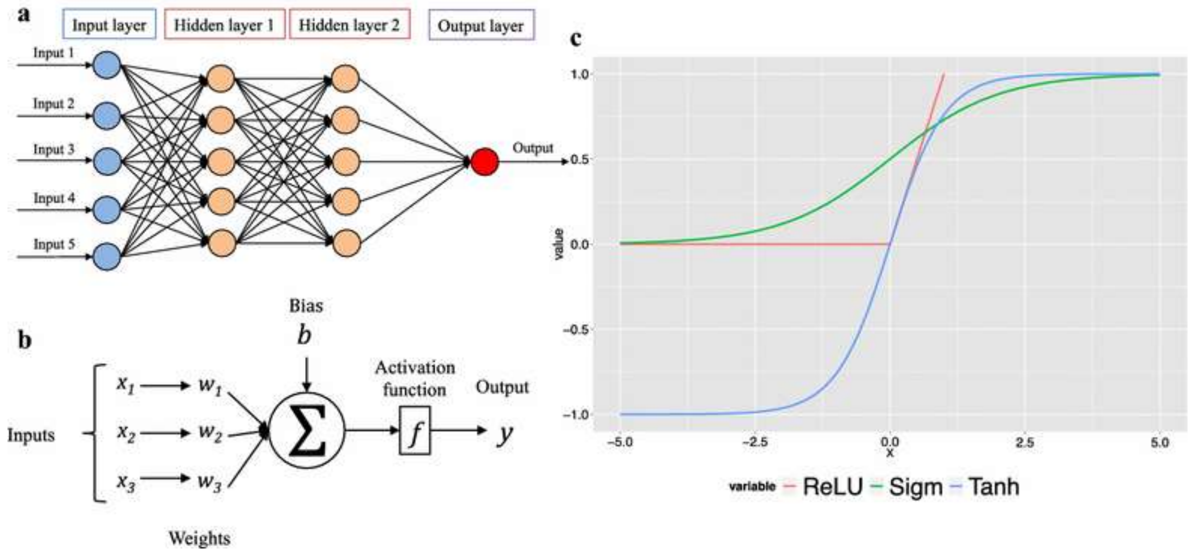
→ Need for DNN verification techniques on the rise



Adverserial attack

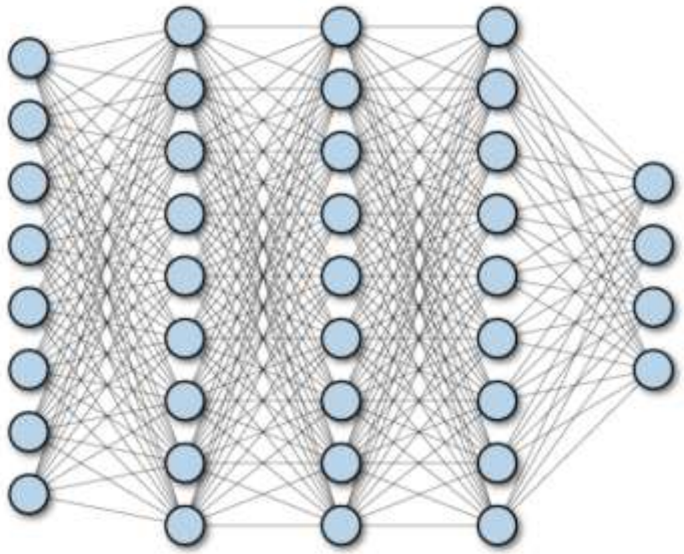
→ Analysis of the various techniques and approaches needed

연구 배경



Difficulty in dealing with NN due to non-linearity providing activation functions

(e.g. sigmoid, ReLU)



Current techniques only deals with limited structure & size of NN

→ Lack of comprehensive and standardized framework for verifying properties of NN

연구 방법

Algorithms for Verifying Deep Neural Networks

Changliu Liu
Carnegie Mellon University
cliu6@andrew.cmu.edu

Tomer Arnon
Stanford University
tarnon@stanford.edu

Christopher Lazarus
Stanford University
clazarus@stanford.edu

Christopher Strong
Stanford University
castrong@stanford.edu

Clark Barrett
Stanford University
barrett@cs.stanford.edu

Mykel J. Kochenderfer
Stanford University
mykel@stanford.edu

October 11, 2020

Abstract

Deep neural networks are widely used for nonlinear function approximation, with applications ranging from computer vision to control. Although these networks involve the composition of simple arithmetic operations, it can be very challenging to verify whether a particular network satisfies certain input-output properties. This article surveys methods that have emerged recently for soundly verifying such properties. These methods borrow insights from reachability analysis, optimization, and search. We discuss fundamental differences and connections between existing algorithms. In addition, we provide pedagogical implementations of existing methods and compare them on a set of benchmark problems.

1 Introduction

Neural networks [26] have been widely used in many applications, such as image classification and understanding [28], language processing [42], and control of autonomous systems [44]. These networks represent functions that map inputs to outputs through a sequence of layers. At each layer, the input to that layer undergoes an affine transformation followed by a simple nonlinear transformation before being passed to the next layer. These nonlinear transformations are often called *activation functions*, and a common example is the *rectified linear unit* (ReLU), which transforms the input by setting any negative values to zero. Although the computation involved in a neural network is quite simple, these networks can represent complex nonlinear functions by appropriately choosing the matrices that define the affine transformations. The matrices are often learned from data using stochastic gradient descent.

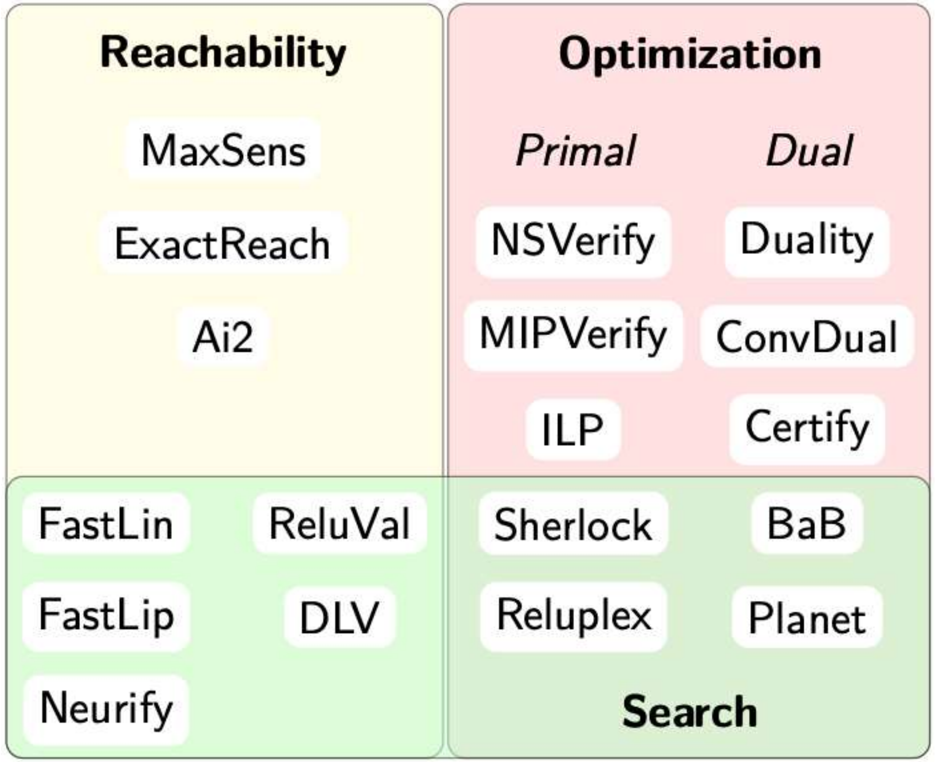
Neural networks are being used for increasingly important tasks, and in some cases, incorrect outputs can lead to costly consequences. Traditionally, validation of neural networks

1

Based on "Algorithms for Verifying Deep Neural Networks", categorized DNN verification methods according to utilized analysis approach

Selected the most frequently used, and representative technique for each categories for research

연구 결과 – 1 Analysis Approaches



1.Reachability

Utilizing layer-by-layer reachability analysis to compute output reachable set

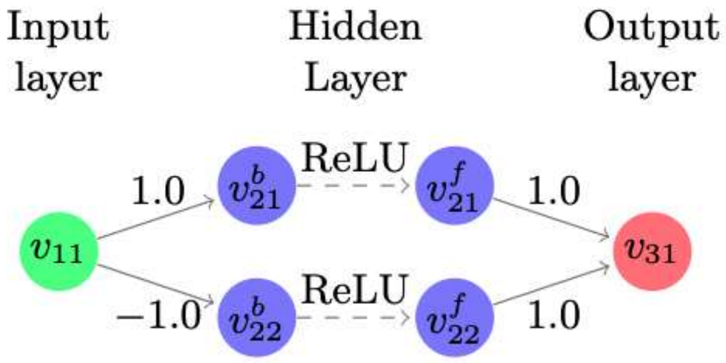
2. Optimization

Considering the neural network itself as a constraint in the optimization process

3. Search

Search for a case to falsify the assertion

연구 결과 – 2 Reluplex, Marabou



Reluplex: apply simplex algorithm to ReLU activated NN

Searches for a variable assignment that simultaneously satisfies the query's linear and non-linear constraints

1. Encode ReLU neuron into a weighted sum variable, and an ReLU activation function variable
2. Repeatedly correct a violated linear constraints or a violated non-linear constraint

↓

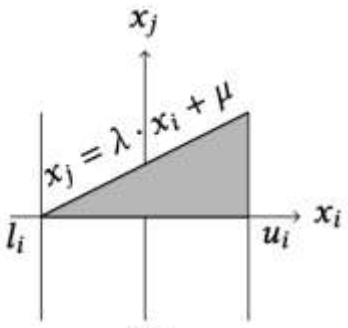
Marabou

upgrade version of Reluplex deals with piecewise-linear activation functions

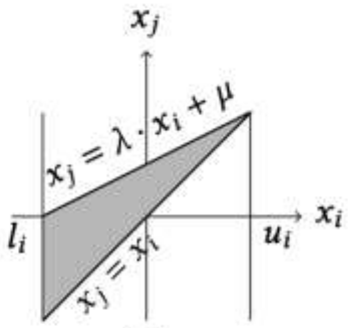
Support network-level reasoning and deduction based on network topology

→ Transform non-linear constraints into linear constraints

연구 결과 – 3 DeepPoly



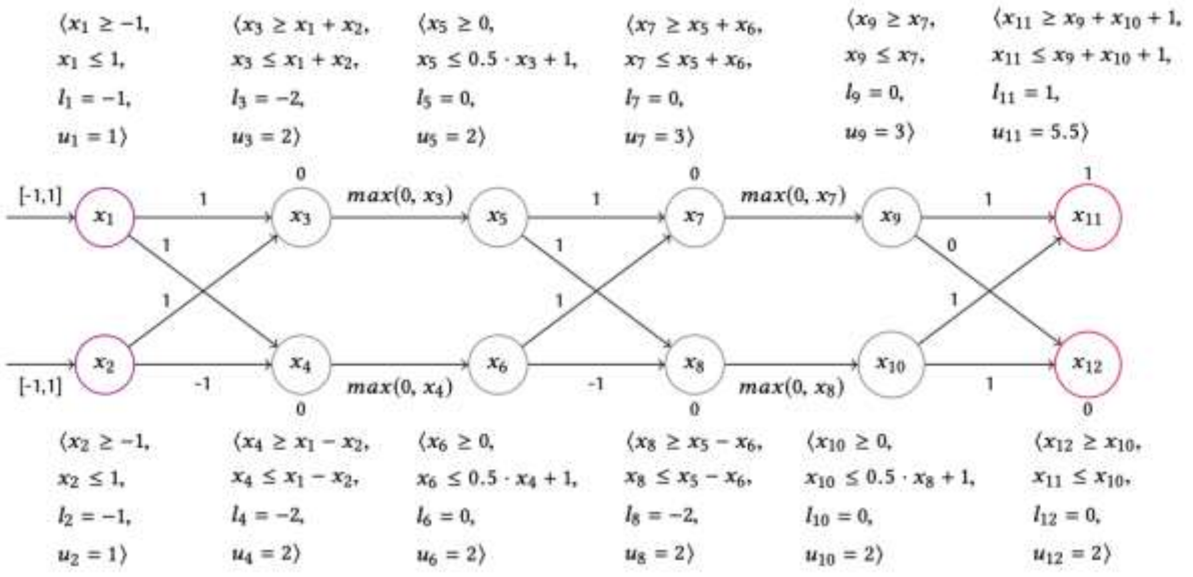
(b)



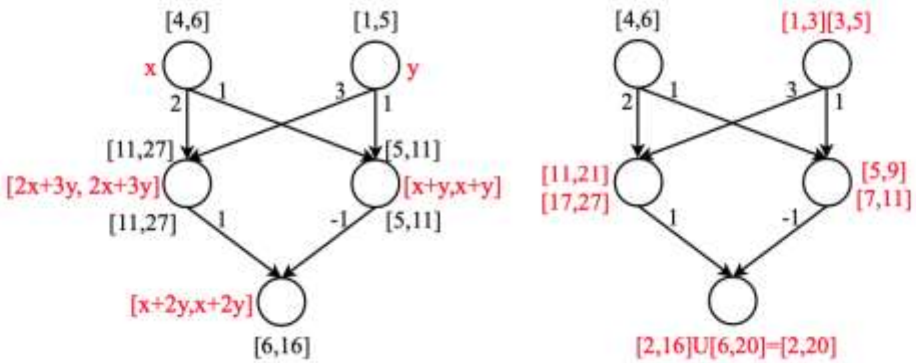
(c)

DeepPoly: abstractor transformers to calculate reachable set

1. Expand neuron into affine transformation and activation node
2. Apply abstract transformers to transform into relational polyhedral constraints and concrete constraints
3. Use back substitution and analysis to compute reachable set



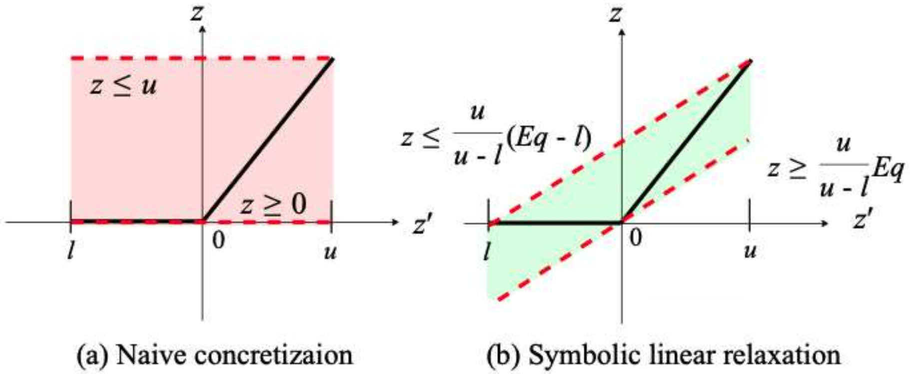
연구 결과 – 4 Neurify



Neurify: interval analysis to compute output set

Upgrade version of ReluVal

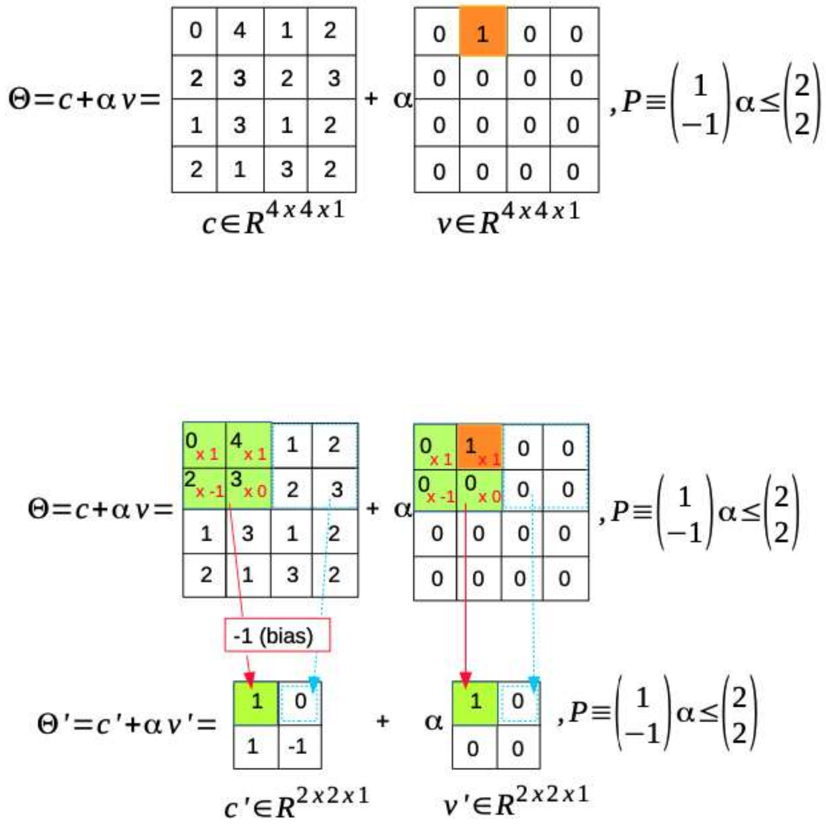
1. Use symbolic interval propagation (ReluVal)
2. Iterative refinement to reduce overestimation (ReluVal)



Enhancements

1. Symbolic linear relaxation
2. Directed constraint refinement

연구 결과 – 5 ImageStar



ImageStar:

Analysis through exact, over-approximate reachability algorithm

1. Represent input set as an ImageStar, a star set generalization

2. Use exact and over-approximate reachability algorithms to construct reachable sets

토론 및 전망

Lack of comprehensive and standardized framework for verifying properties of NN

Current approaches suffers from scalability problem yet unable to deal with realistic-sized neural networks

Still from Reluplex, Marabou, DeepPoly, Neurify, to ImageStar

→ Techniques getting more powerful

토론 및 전망

[1] C.Liu, T.Arnon, C.Lazarus, C.Barrett, and M.J.Kochenderfer, "Algorithms for Verifying Deep Neural Networks," in *Technical Report http://arxiv.org/abs/1903.06758*

[2] G.Singh, T.Gehr, M.Puschel, and M.Vechev, "An Abstract Domain for Certifying Neural Networks," in *ACM Symposium on Principles of Programming Languages*, 2019

[3] G.Katz, C.Barrett, D.Dill, K.Julian, and M.Kochenderfer, "Reluplex: An Efficient SMT Solver for Verifying Deep Neural Networks," in *Proc. 29th Int. Conf. On Computer Aided Verification (CAV)*, 2017

[4] G.Katz, D.Huang, D.Ibeling, K.Julian, C.Lazarus, R.Lin, R.Shah, S.Thakoor, H.Wu, A.Zeljic, D.Dill, M.Kochenderfer, and C.Barrett, "The Marabou Framework for Verification and Analysis of Deep Neural Networks,", in *Proc. 31st Int. Conf. On Computer Aided Verification (CAV)*, 2019

[5] Tran, H.-D., et al., "NNV: The Neural Network Verification Tool for Deep Neural Networks and Learning-Enabled Cyber-Physical Systems, in *32nd Int. Conf. On Computer Aided Verification (CAV), 2020*

[6] Wang, S., Pei, K., Whitehouse, J., Yang, J., and Jana, S., "Efficient Formal Safety Analysis of Neural Networks", in *Advances in Neural Information Processing Systems 31, 2018*

[7] Tran, H., Bak, S., Xiang, W., Johnson, T.T, "Verification of Deep Convolutional Neural Networks Using ImageStars", in *32nd Int. Conf. On Computer Aided Verification (CAV), 2020*

[8] R.Sebastiani, "Lazy Satisfiability Modulo Theories", *Journal on Satisfiability, Boolean Modeling and Computation*, 2007

[9] Zhang, J. ve Li, J., "Testing and verification of neural-network-based safety-critical control software: A systematic literature review", *in Information and Software Technology, 106296*

감사합니다