Institute for Software Integrated Systems

Vanderbilt University

NNV Demo

A *N*eural *N*etwork *V*erification Tool

Design Automation for CPS and IoT (DESTION 2020) - CPS-IoT Week

April 21st, Sydney, Australia

Hoang-Dung Tran, **Diego Manzanas Lopez**, Xiaodong Yang, Patrick Musau, Luan Viet Nguyen, Xeiming Xiang, Stanley Bak, *Taylor T. Johnson*

VeriVITAL - The **Veri**fication and **V**alidation for **I**ntelligent and **T**rustworthy **A**utonomy **L**aboratory
(http://www.VeriVITAL.com)

Overview

- What is NNV?
  - Verification framework
    - Neural networks (NN)
    - Cyber physical systems (CPS)
    - CPS + NN (NNCS)
  - MATLAB
    - CORA, Hyst, and NNVMT

Agenda

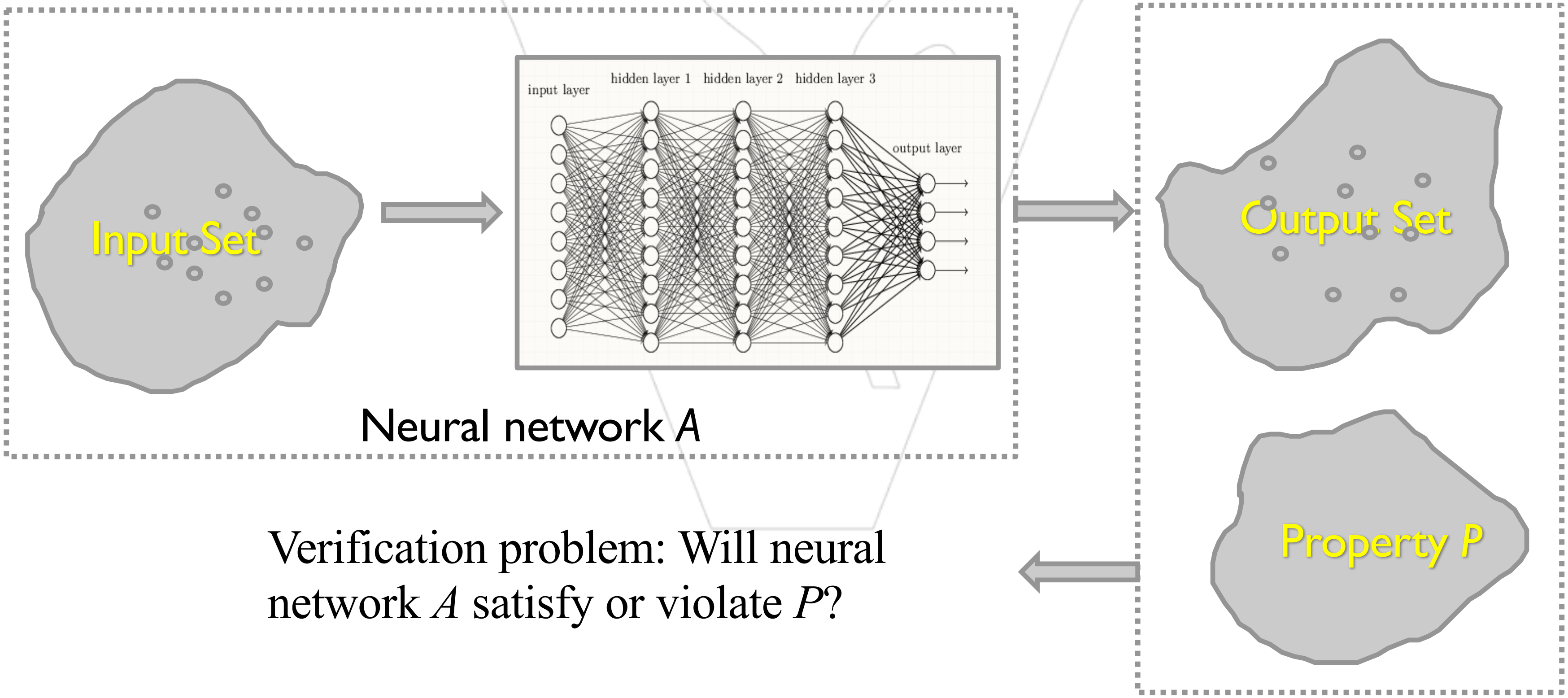
- Our approach
- NN Verification (Open-loop)
- NNCS Verification (Closed-loop)
- Highlights
- Future Work

Agenda

- **Our approach**
- NN Verification (Open-loop)
- NNCS Verification (Closed-loop)
- Highlights
- Future Work

General Approach

- Using **reachability analysis** to verify the safety and the robustness of neural networks



Input Set

Neural network *A*

Output Set

Property *P*

Verification problem: Will neural network *A* satisfy or violate *P*?

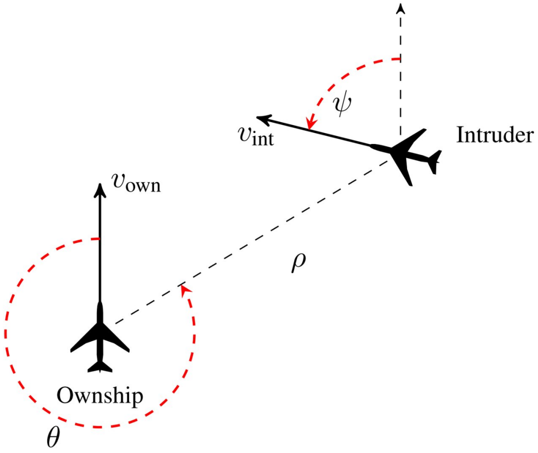
Agenda

- Our approach
- **NN Verification (Open-loop)**
- NNCS Verification (Closed-loop)
- Highlights
- Future Work

Open-loop Verification

Feedforward NN (FNN)

- Linear and Piecewise linear functions
  - Exact analysis
  - More efficient using Star set
- Also supports nonlinear activation functions (tanh, sigmoid)
  - Over-approximate analysis (only)



Julian et al., DASC 2016

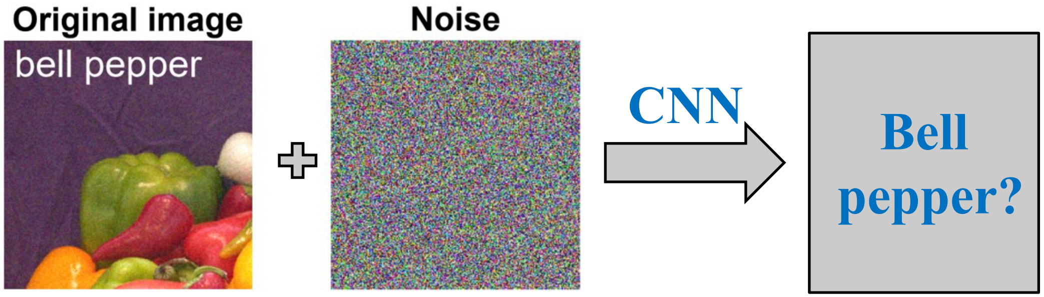
- Demo
  - AcasXu Neural Network
  - NN controller should output a safe and correct control action.

Open-loop Verification

Convolutional NN (CNN)

- Perception – Image classification
  - How robust are CNNs under different input perturbations?
    - Can the CNN classify the image correctly despite the perturbations?
  - Support
    - MaxPool2D, AveragePool2D, Relu, FullyConnected, BatchNormalization and Conv2D layers

- Demo
  - VGG – Imagenet
  - FGSM attack



Agenda

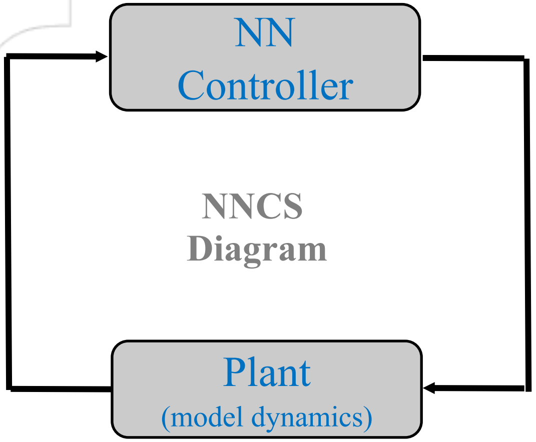
- Our approach
- NN Verification (Open-loop)
- **NNCS Verification (Closed-loop)**
- Highlights
- Future Work

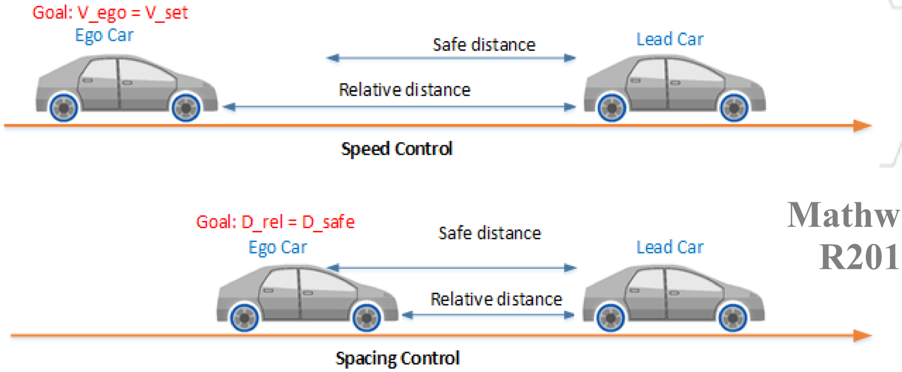
Closed-loop Verification

NNCS

- Combine FNN reachability analysis with plant reachability analysis
  - Use CORA for nonlinear dynamics
- Demo
  - Adaptive Cruise Controller (ACC)
  - Will the ego car be safe?

    Safety requirement: actual distance > safe distance





Mathworks R2018a

Agenda

- Our approach
- NN Verification (Open-loop)
- NNCS Verification (Closed-loop)
- **Highlights**
- Future Work

Highlights

- **Star-based method** for safety verification of DNNs
  - 10x-1000x runtime performance improvements vs. other state-of-the-art approaches
  - Minimizing conservatism of reachability results.
- **ImageStar-based method** , for robustness verification of deep CNNs
  - Enabling robustness analysis of networks with upwards of **100 million parameters**
- **Parallelization** of NN reachability analysis
  - Yielding 10x-1000x runtime performance improvements (versus e.g. Reluplex and other state-of-the-art approaches)
- Participate in **CPS-IoT Week ARCH-COMP'19** (NNCS) and **AAAI VNN'19** (NN) verification competitions
- **Publications**
  - **FormaLISE’19, FM’19** (Polytope, Star-set, open-loop verification)
  - **EMSOFT’19, FomLAS’19** (Star-set, closed-loop verification)
  - **VNN’19** (Simulation-based Verification for feedforward networks)
  - **ARCH’19** (Benchmarking for Neural Network Control Systems)
  - **WAAS'20** (Underwater vehicle closed-loop verification)
  - **CAV’20** (ImageStar, robustness verification)
  - **CAV’20** (Tool development)

Agenda

- Our approach
- NN Verification (Open-loop)
- NNCS Verification (Closed-loop)
- Highlights
- **Future Work**

Future Work

- Segmentation Neural Networks
- Improve CNN support
- Improve NNV user experience

Demo

- FNN Verification
  - NN architecture
  - AcasXu_1_1 example
- NNCS Verification
  - ACC nonlinear
  - ACC linear
  - Multiple runs (initial states), live plots
- CNN Verification
  - VGG robustness analysis
  - FGSM attack (vary the degree of the attack)
- Code Ocean
  - What is Code Ocean?
  - Run some experiments