Robustness Verification for Traffic Sign Classification System

Mohammad Shirmohammadi, mohoshirmo@gmail.com
Arcadia High School, Class of 2023
USC Viterbi Department of Electrical and Computer Engineering, SHINE 2022

Introduction

In contemporary society, existing artificial intelligence (AI) architectures are being increasingly integrated in cyber-physical systems (CPS) and, assembled on artificial neural networks (NNs), have proven to develop human-like performance on decision making tasks. However, well-trained NNs have been proven to be sensitive to disturbances and noise than can lead to incorrect decisions with large confidence. This poses a safety risk for drivers relying on their Driver Assistance System that utilize traffic sign detection methodologies. To solve this challenge, we investigate the creation and optimization of NNs to increase test accuracy. Then, we apply an satisfiability modulo convex programming (SMC) solver to verify the trained models.

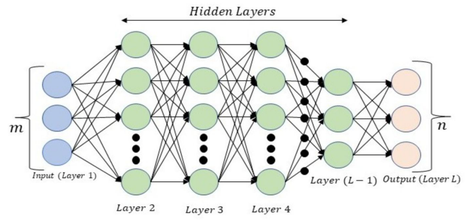

Fig. 1: Neural network diagram. PC: Nikhil Naik

Skills Learned

- Linux Terminal to install python packages
- How to navigate through conda
- How to graph using Matplotlib python package
- How to apply the mathematical chain rule when performing back propagation
- How to avoid an overfitting problem by studying Matplotlib graphs
- How to adjust hyperparameters to optimize NNs
- How to use the Z3 solver

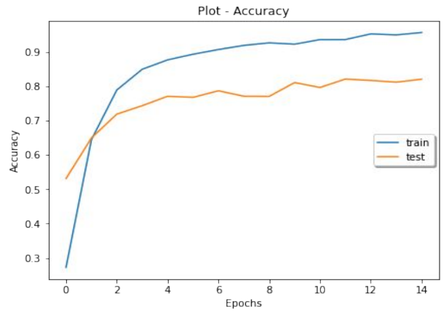

Fig. 2: GTSRB NN Matplotlib Line Graph. PC: Mohammad Shirmohammadi

Research Process

Traffic Sign Classification

- Forge a fully connected neural network and adjust its architecture
- Train the neural network using the German Traffic Sign Recognition Benchmark (GTSRB) dataset: 68% for training; 32% for testing
- Manipulate the hyperparameters of neural networks to improve performance
  - Increase testing accuracy and reduce the loss function
  - Observe how overfitting can be encountered and how to address it
- Output the accuracy and loss graphs per model

Fig. 3: GTSRB Dataset. PC: Ling Yu

No. (#)	Architecture	Epoch (#)	LR	Batches (#)	Acc. (%)
1	[2352, 50, 43]	15	0.001	200	78.93
2	[2352, 512, 43]	15	0.001	200	80.51
3	[2352, 512, 512, 43]	15	0.001	200	***80.68***
4	[2352, 1024, 512, 512, 43]	15	0.001	200	79.01

Tabel 1: Batch size, learning rate, and best accuracy for each training attempt. Created in Google Sheets.

Robustness Verification

- Train and load the model
- Retrieve input image with disturbances for verification
- Apply SMC solver for the NN model using mathematical methods with constraints
- Verify model safety via the negation of the safe property
  - Safe if unsat
  - Unsafe if sat and find a counterexample
- Test the framework with 50 randomly selected images.
- Robustness decreases with increased model complexity and disturbances

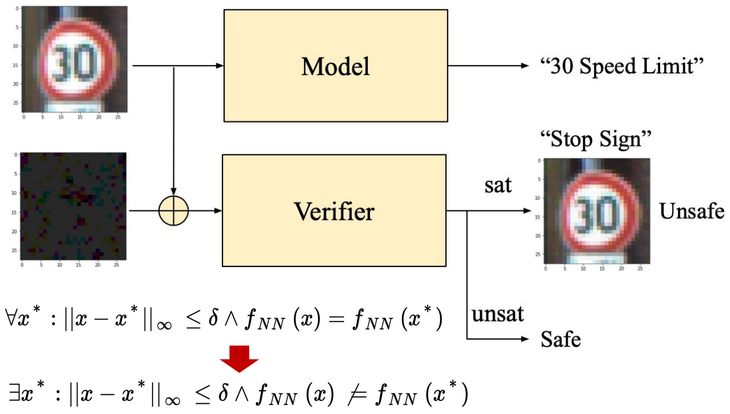

Fig. 3: Verification Framework. PC: Mohammad Shirmohammadi

Fig. 4: Research essay on the pros and cons of AI in society. PC: Mohammad Shirmohammadi

δ	Architecture	safe cases / proved cases	time(s)
0.005	[2352, 24, 43]	32/50	27.51
	[2352, 50, 43]	32/50	29.45
	[2352, 24, 24, 43]	24/50	37.7
0.01	[2352, 24, 43]	22/50	44.1
	[2352, 50, 43]	20/50	46.05
	[2352, 24, 24, 43]	16/50	47.09
0.05	[2352, 24, 43]	5/50	102.57
	[2352, 50, 43]	1/50	108.63
	[2352, 24, 24, 43]	0/50	127.53

Table 2: Safe cases over proved cases with the corresponding delta. Verification process for a GTSRB NN.

Relevance to STEM Coursework

SHINE has amplified the way I understand Computer Science and how AI can be utilized to solve societal problems such as driving fatalities and homelessness. I hope to share my research and experience in AI with students with cognitive disabilities whom I teach in my 501c3 nonprofit organization, Code Can Bridge.

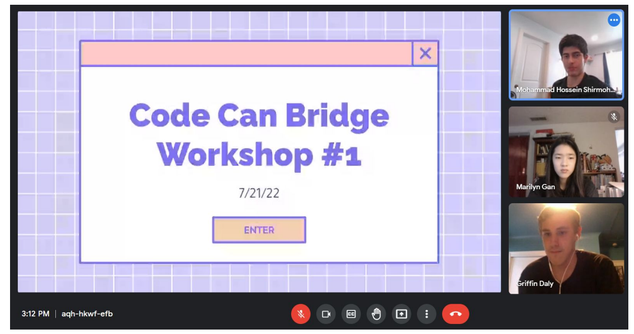

Fig. 5: Coding Workshop. PC: Code Can Bridge

Citation

[1] Saadna, Y., & Behloul, A. (2017). An overview of traffic sign detection and classification methods. *International journal of multimedia information retrieval*.

[2] Naik, N., & Nuzzo, P. (2020). Robustness Contracts for Scalable Verification of Neural Network-Enabled Cyber-Physical Systems. In *18th ACM-IEEE International Conference on Formal Methods and Models for System Design (MEMOCODE)*.

Acknowledgments

Thank you Professor Nuzzo for making this research experience possible. Thank you to my Ph.D. student mentor Yifeng Xiao for teaching me, centor mentor Marcus Gutierrez for guiding me, the SHINE team for supporting me, former SHINE researchers Rachel Loh and Branden Leong for inspiring me to join SHINE, and most of all, Dr. Mills for always taking the time to meet with me and offer her valuable personal advice.