Collaborative Research: SLES:

Foundations of Qualitative and Quantitative Safety Assessment of Learning-enabled Systems

Grant no. 2331938

Collaborate with 2331937
(Hoang-Dung Tran, UNL, Lead)

Weiming Xiang

School of Computer and Cyber Sciences
Augusta University

Project Overview

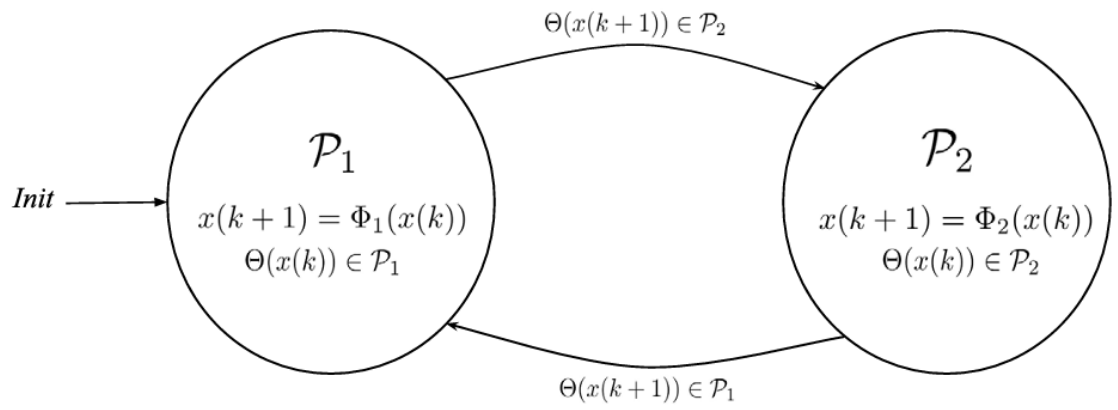
- Develop a New Specification Language that Supports Qualitative and Quantitative (Q2) Safety Reasoning
- **Develop Scalable, Memory-Efficient DNN Q2 Safety Verification Methods**
- Develop System-Level Q2 Safety Assessment Methods

Annual Progress

- Develop Scalable, Memory-Efficient DNN Q2 Safety Verification Methods
  - Develop **computational efficient and verification friendly** learning models
    - Small NNs + Transition (Hybrid Learning Structure)
    - Trustworthy NN Compression

Annual Progress

- Develop **computational efficient and verification friendly** learning models
  - Small NN + Transition (Hybrid Learning Structure)



$$\mathcal{H} \triangleq \langle \mathcal{P}, x, init, \mathcal{E}, g, \mathcal{G}, inv, \Phi \rangle$$

Neural hybrid automaton: Partitions; State Variables; Initial Conditions; Transitions; Guard Functions; Guards; Invariants; and **Neural Networks.**

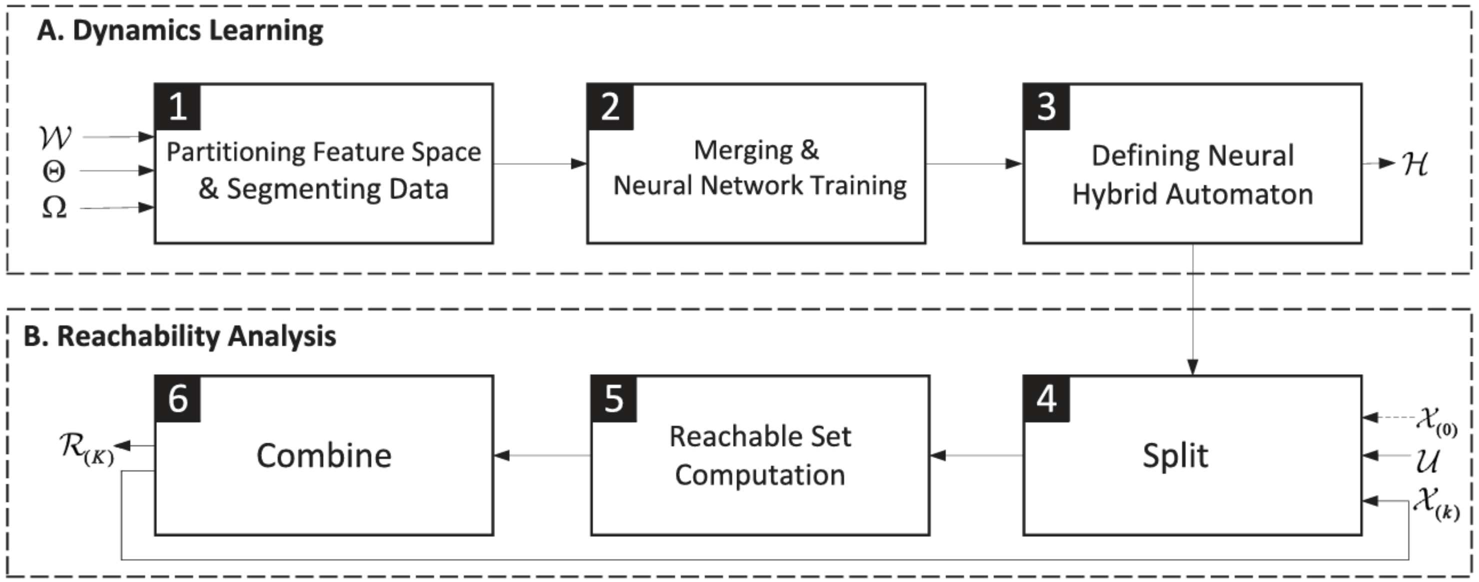
⬇

Small-size NN (Extreme Learning Machine)

Annual Progress

– Develop **computational efficient and verification friendly** learning models
  - Small NNs + Transition (Hybrid Learning Structure)

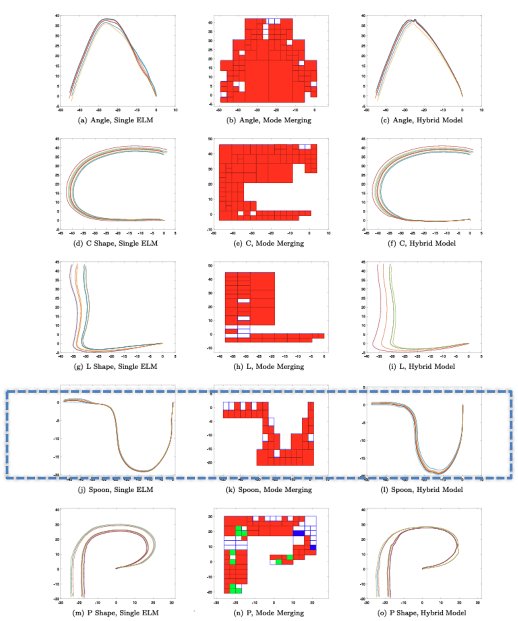
Learning and Verification Framework

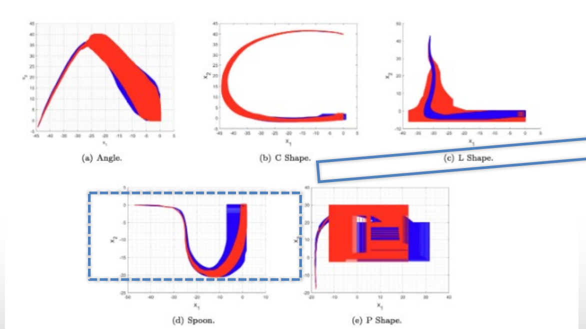


Annual Progress

- **Evaluation**
  - **Learning human handwriting motions on LASA dataset**

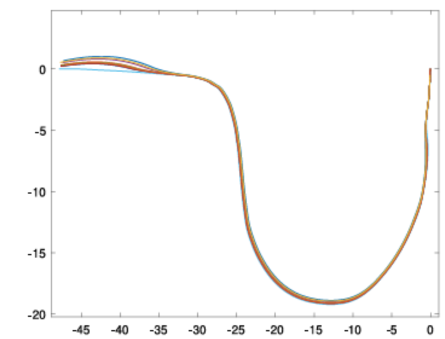
Test on diverse motions





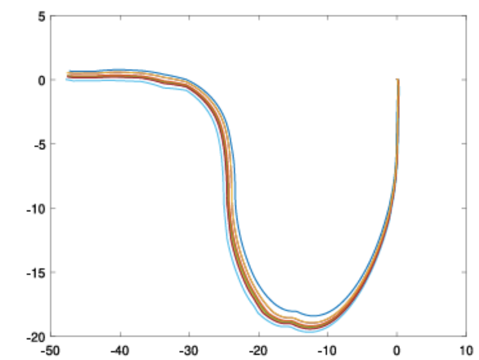
Spoon shape

Single NN



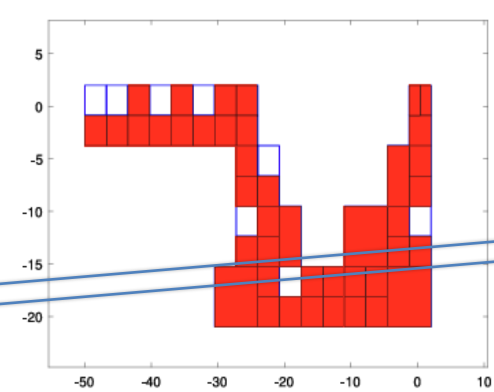
(j) Spoon, Single ELM

Hybrid Learning



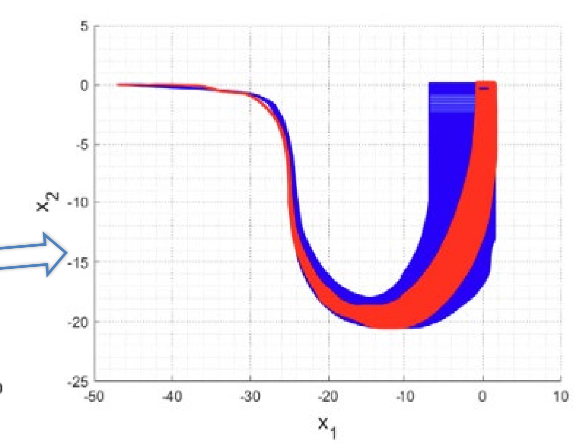
(l) Spoon, Hybrid Model

Partition (after merge)



(k) Spoon, Mode Merging

Reachable Set



(d) Spoon.

Red box
Hybrid (tighter)

Annual Progress

- **Evaluation**
  - **Learning human handwriting motions on LASA dataset**

MSE and computation time of single neural network model.

Data set	MSE	Training	Reachable set
Angle	2.739×10^{-4}	4.50×10^{-2} s	3.6466×10^{4} s
C Shape	2.375×10^{-4}	4.79×10^{-2} s	9.0068×10^{4} s
L Shape	1.972×10^{-4}	4.37×10^{-2} s	9.7783×10^{4} s
Spoon	1.767×10^{-4}	4.30×10^{-2} s	8.4465×10^{4} s
P Shape	2.301×10^{-4}	4.50×10^{-2} s	8.5201×10^{4} s

Single neural network
Single NN with 200 hidden neurons

MSE and computation time of neural hybrid automaton.

Data set	MSE	Training	Reachable set
Angle	4.787×10^{-4}	8.60×10^{-3} s	503.0428 s
C Shape	8.323×10^{-4}	1.25×10^{-2} s	3308.2473 s
L Shape	4.175×10^{-4}	7.05×10^{-3} s	1513.7453 s
Spoon	4.643×10^{-4}	7.57×10^{-3} s	69.9705 s
P Shape	9.484×10^{-4}	4.71×10^{-3} s	127.3636 s

Data set	Time
Angle	1.3%
C Shape	3.6%
L Shape	1.5%
Spoon	0.08%
P Shape	0.14%

Computation efficient and verification friendly:
Computation time is reduced.

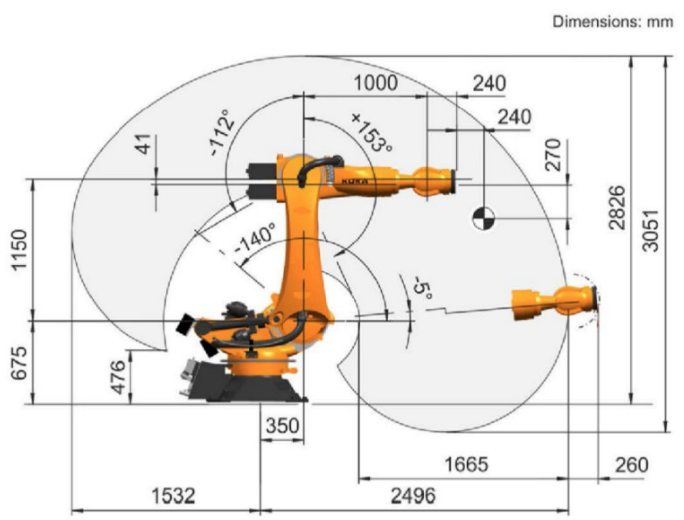
Hybrid structure
Multiple NNs, each NN with 20 neurons

Annual Progress

- **Evaluation**
  - 6 Joint Industrial Robot, high-dimensional dynamics

$$x(k+1) = f(\tau(k), u(k)) \qquad [x(k)^T, \ldots, x(k-23)^T]^T \in \mathbb{R}^{144} \qquad u(k) \in \mathbb{R}^{144}$$

Use previous 2.4 s to predict the positions in the next 0.1 s， with a sample time of 0.1 s



MSE Performance

Joint	Prediction Mode		Simulation Mode		
	Φ	$\mathcal{H}$	$\hat{f}$	Φ	$\mathcal{H}$
$x_{(1)}$	**0.1501**	0.2357	0.638	1.0009	**0.6162**
$x_{(2)}$	**0.1752**	0.3038	0.829	1.5169	**0.6466**
$x_{(3)}$	**0.1568**	0.2852	0.876	1.3408	**0.6871**
$x_{(4)}$	**0.1800**	0.3004	0.894	1.5775	**0.8145**
$x_{(5)}$	**0.1547**	0.3027	0.869	1.9444	**0.9944**
$x_{(6)}$	**0.1506**	0.2348	0.845	1.5491	**0.7885**
all joint	**0.1612**	0.2771	0.8252	1.4897	**0.7579**

Bonus: Best accuracy

J. Weigand, J. Götz, J. Ulmen, and M. Ruskowski. (2023). Dataset and Baseline for an Industrial Robot Identification Benchmark. [Online].

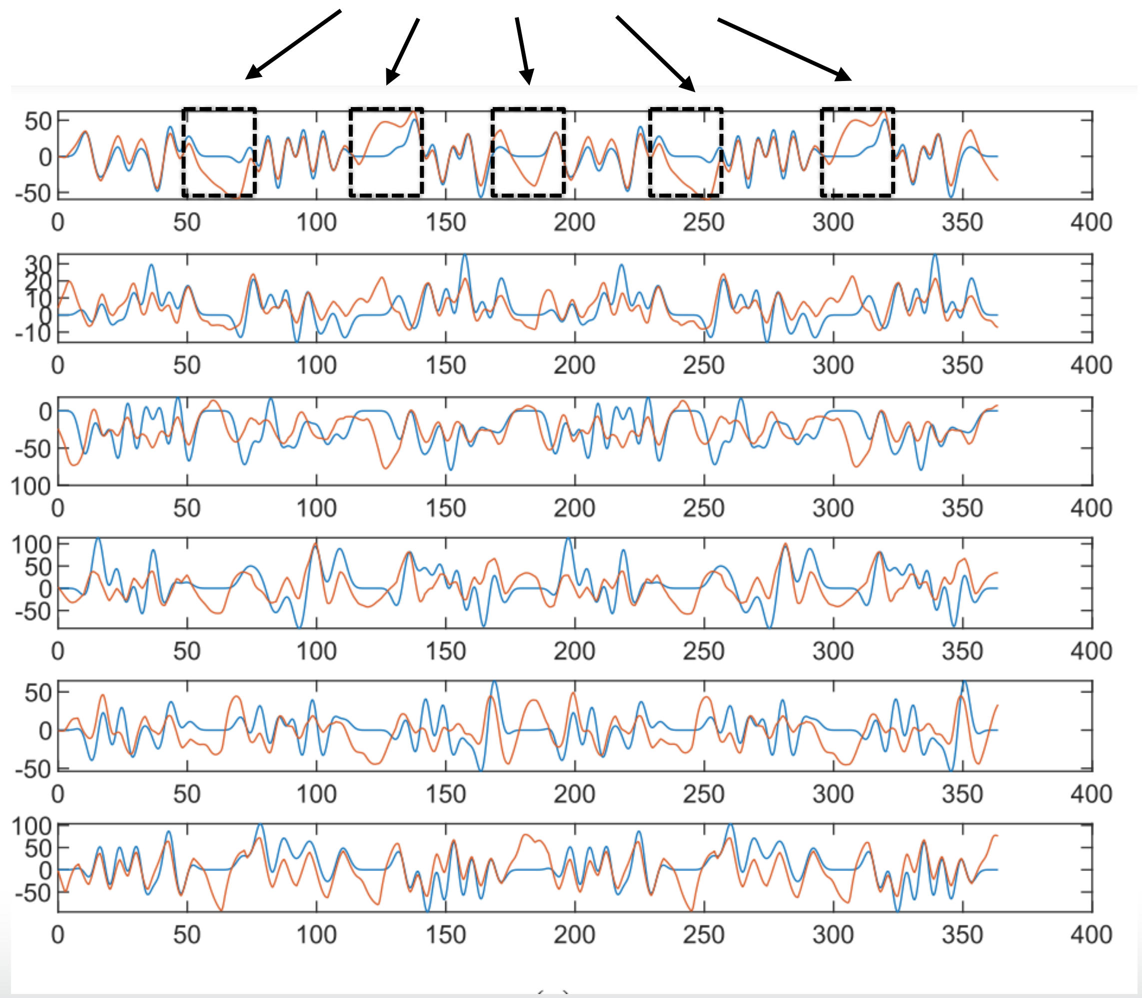
Annual Progress

- **Evaluation**

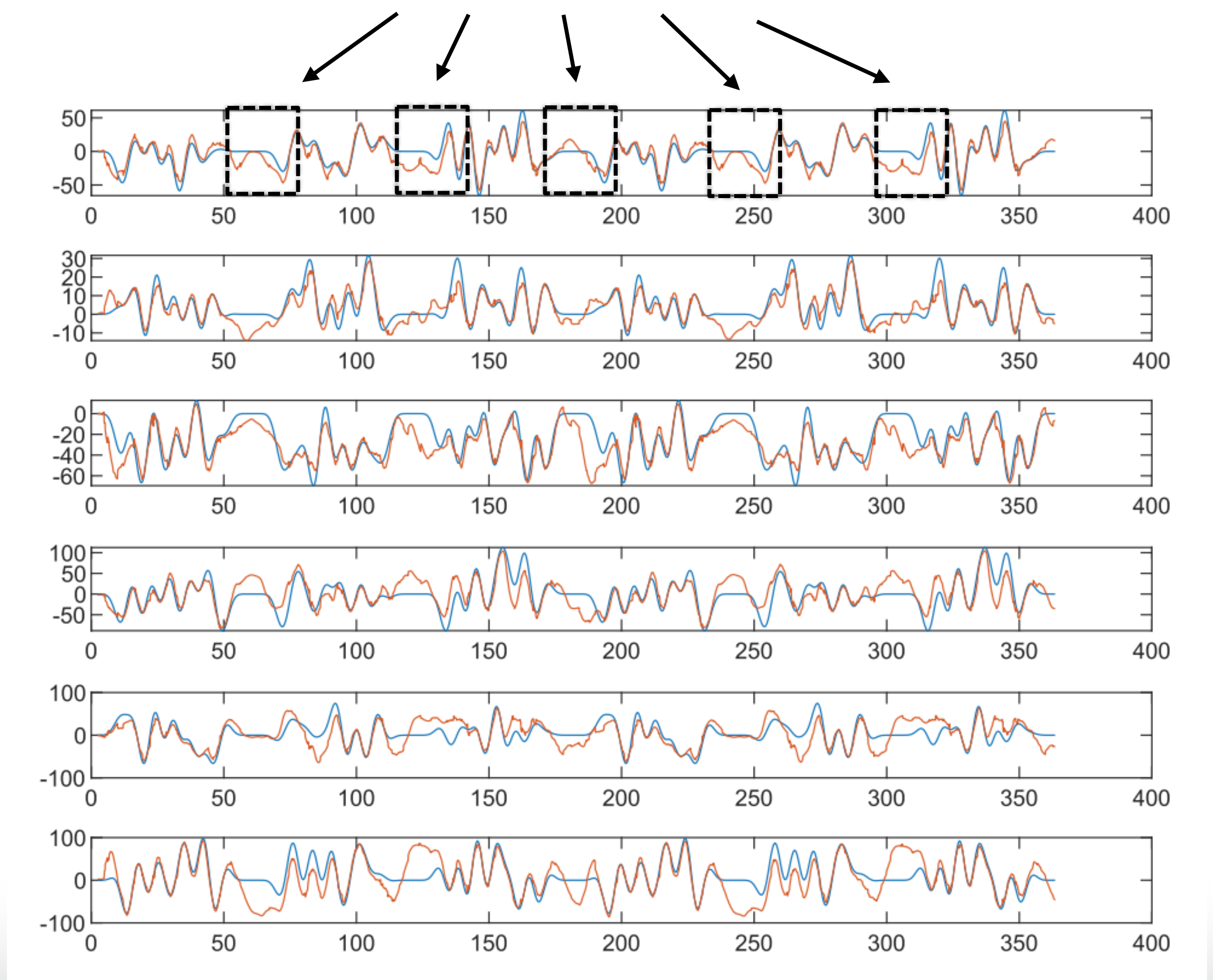
__Hybrid learning better adapts to different operational phases.__

  - **Comparison**

low-velocity phases (single model, hard to model)



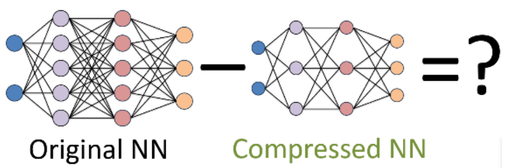
low-velocity phases (hybrid, improved accuracy)



J. Weigand, J. Götz, J. Ulmen, and M. Ruskowski. (2023). Dataset and Baseline for an Industrial Robot Identification Benchmark. [Online].

Annual Progress

- Trustworthy NN Compression

We need to consider **<u>the worst case</u>** for 

Maximal Discrepancy: $\rho = \max_{x \in X} ||f_1(x) - f_2(x)||$

where X is the input constraint set, including **<u>all possible (*infinite number of*) inputs</u>**.

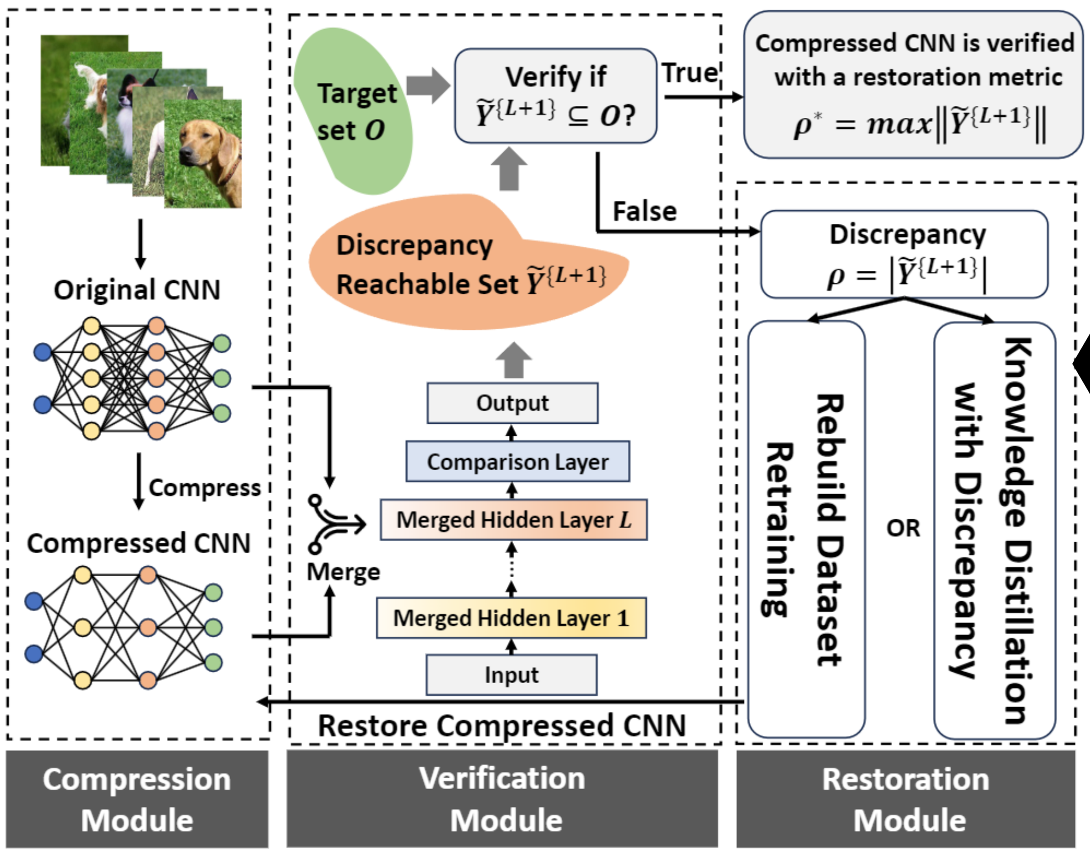
A **<u>smaller Maximal Discrepancy</u>** means a **<u>better property restoration capability</u>** to maintain the property of the original network.

Annual Progress

- Trustworthy NN Compression

Question: How to compute, verify, and restore Maximal Discrepancy?

Neural network compression
Pruning,
Quantization,
Distillation
…



Neural network restoration
Retraining
Knowledge distillation
…

Neural Network verification: reachability, bound propagation, …

Annual Progress

- Trustworthy NN Compression

Evaluation： 8 compression methods embedded in PyTorch.

MNIST: CNN contains 2 convolution layers, 1 pooling layer,and 2 linear layers, with ReLU activation.
CIFAR-10: VGG16 model has 16 layers with trainable parameters.

Compression Methods:

- 4 Quantization methods.
- 4 Pruning methods

Discrepancy Computation Methods:

- Star-set based reachability
- Bound propagation

Table 1: Comparison among compression methods for CNN with MNIST dataset and CIFAR10 dataset

Dataset	Network	Parameters	Size	Sparsity	Accuracy (Similar ➡)	Discrepancy ρ^*	Time (⬅ Different)
MNIST	Original network	1.2 M	4690 KB	0%	98%	-	-
	Eager QAT network	1.2 M	1184 KB	0%	99%	2.6147	11.5470 s
	FX QAT network	1.2 M	1179 KB	0%	99%	7.3433	5.7949 s
	Eager Static network	1.2 M	1184 KB	0%	96%	>10.0937	-
	FX Static network	1.2 M	1179 KB	0%	94%	3.4256	322.9563 s
	L-Unstru network	1.2 M	4690 KB	20%	98%	0.6061	11.3286 s
	G-Unstru network	1.2 M	4690 KB	20%	98%	**0.0604**	17.3008 s
	L-Stru network	1.2 M	4690 KB	19.97%	98%	0.6670	18.2180 s
	R-Stru network	1.2 M	4690 KB	19.97%	97%	10.6880	20.3732 s
CIFAR10	Original network	7.8 M	30791 KB	0%	80%	-	-
	Eager QAT network	7.8 M	7781 KB	0%	80%	16.3547	527.7462 s
	FX QAT network	7.8 M	7725 KB	0%	80%	17.0159	677.9000 s
	Eager Static network	7.8 M	7781 KB	0%	28%	>28.1005	-
	FX Static network	7.8 M	7725 KB	0%	28%	>27.3428	-
	L-Unstru network	7.8 M	30791 KB	20%	80%	1.9602	482.8901 s
	G-Unstru network	7.8 M	30791 KB	20%	80%	**0.9487**	494.5455 s
	L-Stru network	7.8 M	30791 KB	19.95%	32%	32.5631	427.7225 s
	R-Stru network	7.8 M	30791 KB	19.95%	26%	33.1896	364.9997 s

Table 1: Comparison between reachability method and BaB method on MNIST and CIFAR10

Dataset	Network 1		Network 2		Noise	Reachability		BaB	
	Model	Accuracy	Model	Accuracy		Discrepancy	Time	Discrepancy	Time
MNIST	FNN4	97%	FNN4	97%	3*3	4.3068	0.03s	4.2713	12s
	CNN4	90%	CNN4	89%	3*3	3.1278	0.03s	3.1229	20s
CIFAR10	VGG	75%	VGG	73%	2*1*3	18.6372	288s	18.6284	346s
	VGG	75%	VGG	73%	32*32*1	-	-	388.3810	3967s

Annual Progress

- Trustworthy NN Compression

Evaluation： 8 compression methods embedded in PyTorch

Restoration Results

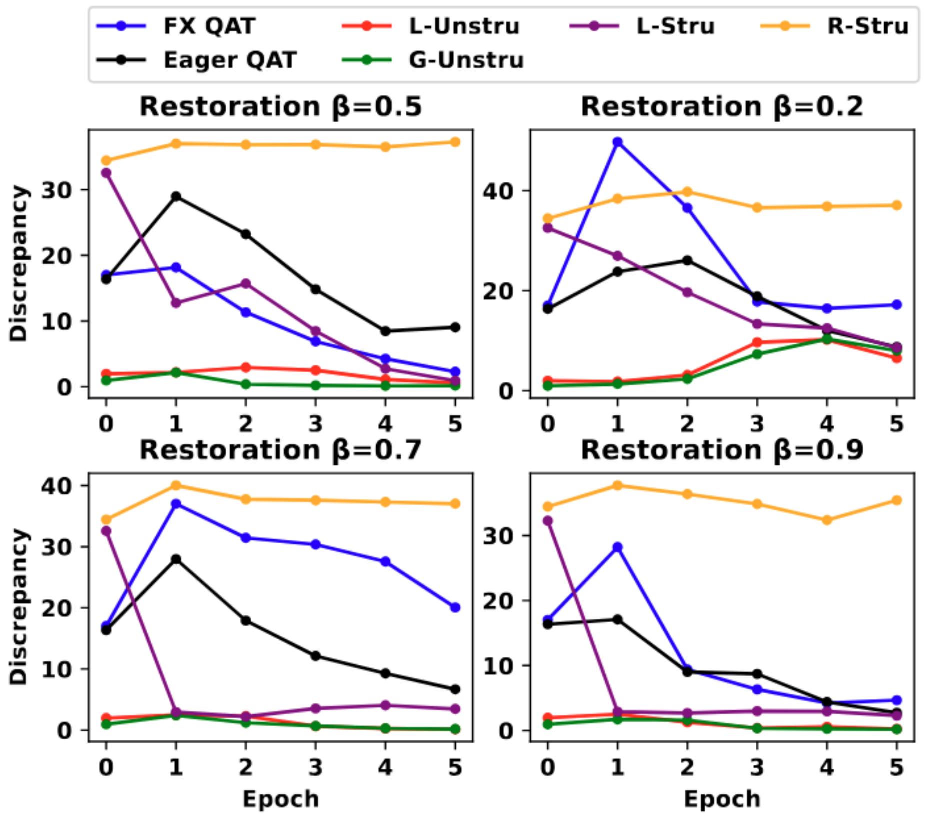


Figure 2: Restoration performance for different compression methods with different β.

Table 2: Retraining restoration performance

Methods	Ori.	$\beta = 0.5$	$\beta = 0.2$	$\beta = 0.7$	$\beta = 0.9$
FX QAT	17.0159	**2.2669**	17.1840	20.054	4.6673
Eager QAT	16.3547	9.0541	8.7409	6.6716	**2.7320**
L-Unstru	1.9602	0.5915	6.4987	**0.1254**	0.2525
G-Unstru	0.9487	**0.1633**	7.9780	0.1648	0.1836
L-Stru	32.5631	**0.9242**	8.5676	3.4599	2.2905
R-Stru	34.4265	37.2667	37.0809	37.0177	35.4302

Summary

- Develop Scalable, Memory-Efficient DNN Q2 Safety Verification Methods
  - Develop **computational efficient and verification friendly** learning models
  - Annual Progress
    - Small NN + Transition (Hybrid Learning Structure)
    - Trustworthy NN Compression
  - **Next Year**
    - Work with collaborator Dr. Tran at UNL to integrate them to ProStar framework in learning-enable CPS.