CAREER: Enabling Trustworthy Upgrades of Machine-Learning Intensive Cyber-Physical Systems

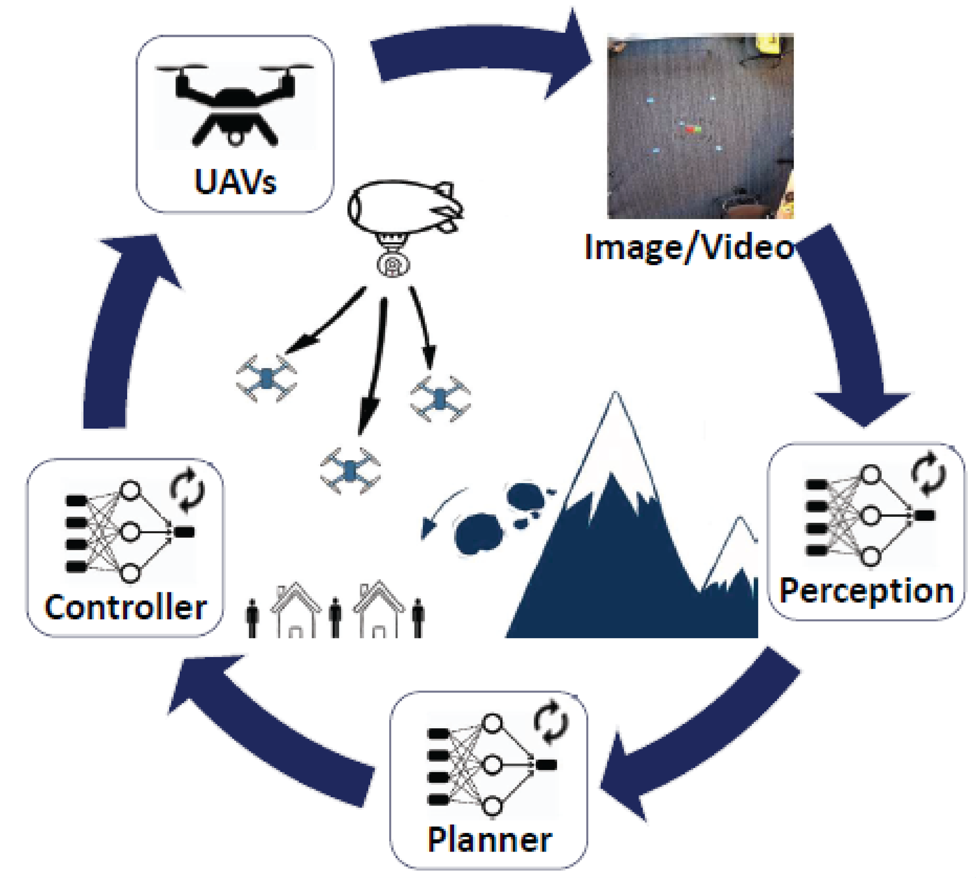
Weiming Xiang, School of Computer and Cyber Sciences, Augusta University
https://www.nsf.gov/awardsearch/showAward?AWD_ID=2143351

AUGUSTA UNIVERSITY

Goal: Develop verification and upgrade procedures to provide **formal safety guarantees for ML-intensive CPS throughout life cycles.**

Challenges

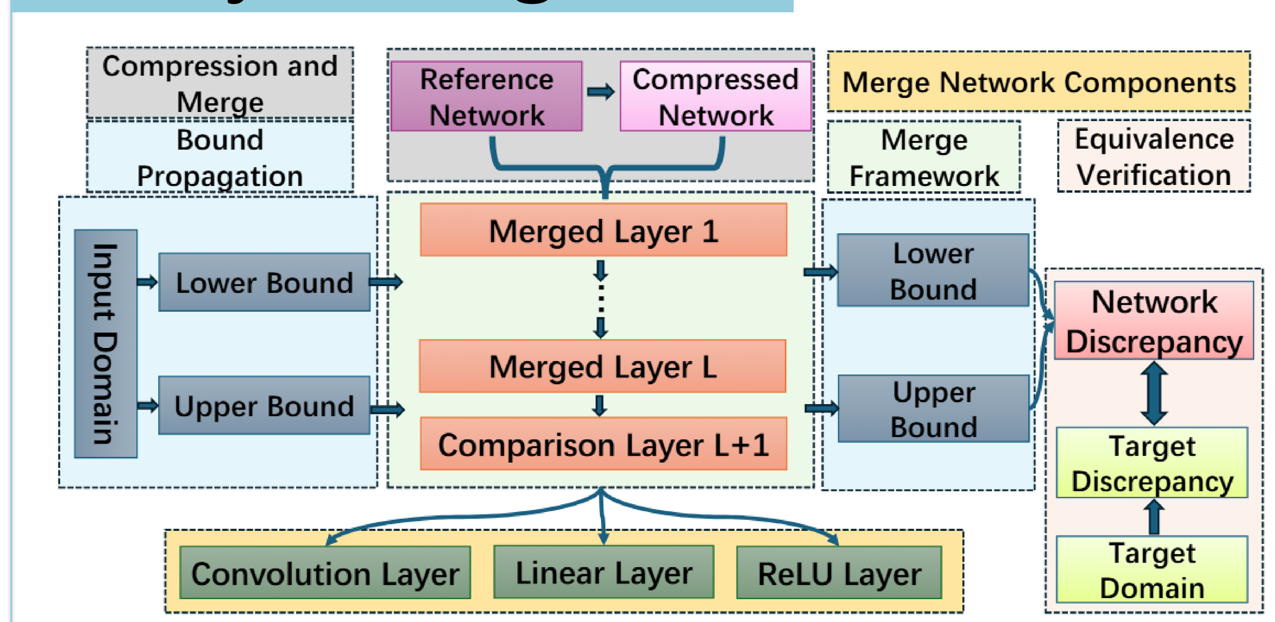
- **Vulnerabilities of ML Components**
  How to fully identify the incompatibilities caused by the ML upgrade, and formally verify upgrades of ML-intensive CPS?
- **Unique Upgrade Procedures of ML Components**
  How to develop safety-assured ML upgrade for ML-intensive CPS?



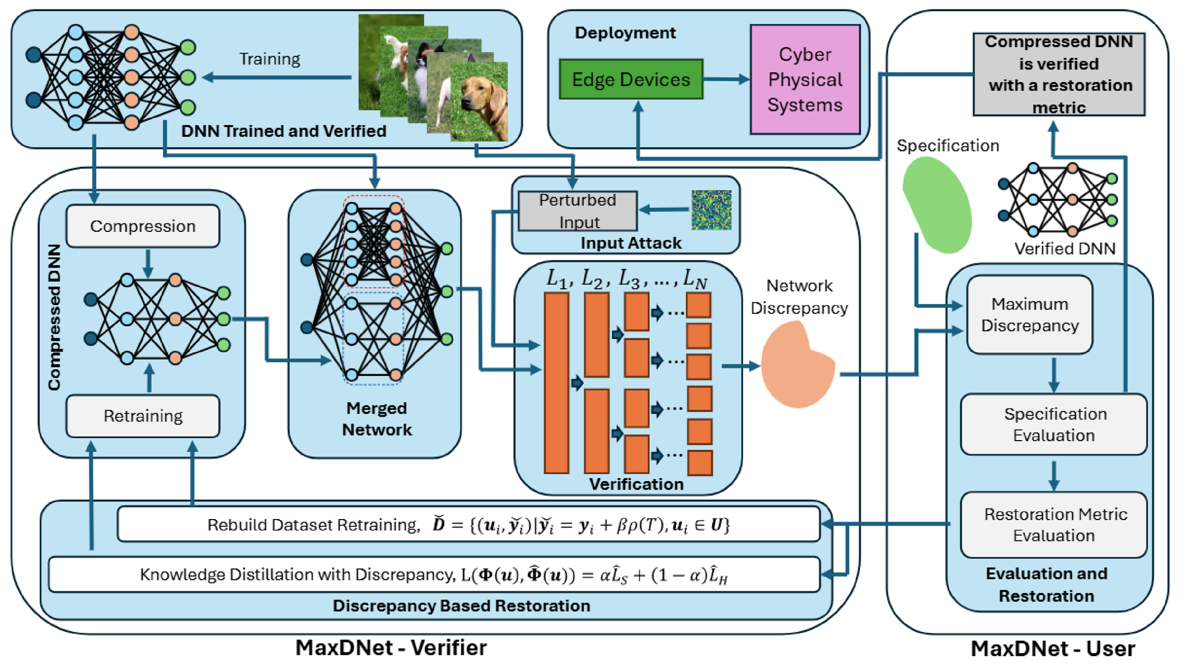
Scientific Impacts

- ***Safe Upgraded Model:*** Safety Verification and Monitoring of ML-Intensive CPS Upgrades
- ***Safe Upgrade Procedure**:* Safety-Assured Upgrades for ML-Intensive CPS
- ***Safe Upgrade Application:*** Safe Upgradable ML-Intensive Autonomy

Project Progress



EqBaB: Efficient Equivalence Verification for Compressed DNNs with Bound Propagation



MaxDNet: A Formal Framework for Verifying and Restoring Compressed Deep Neural Networks

Results

Performance restoration (MaxDNet)

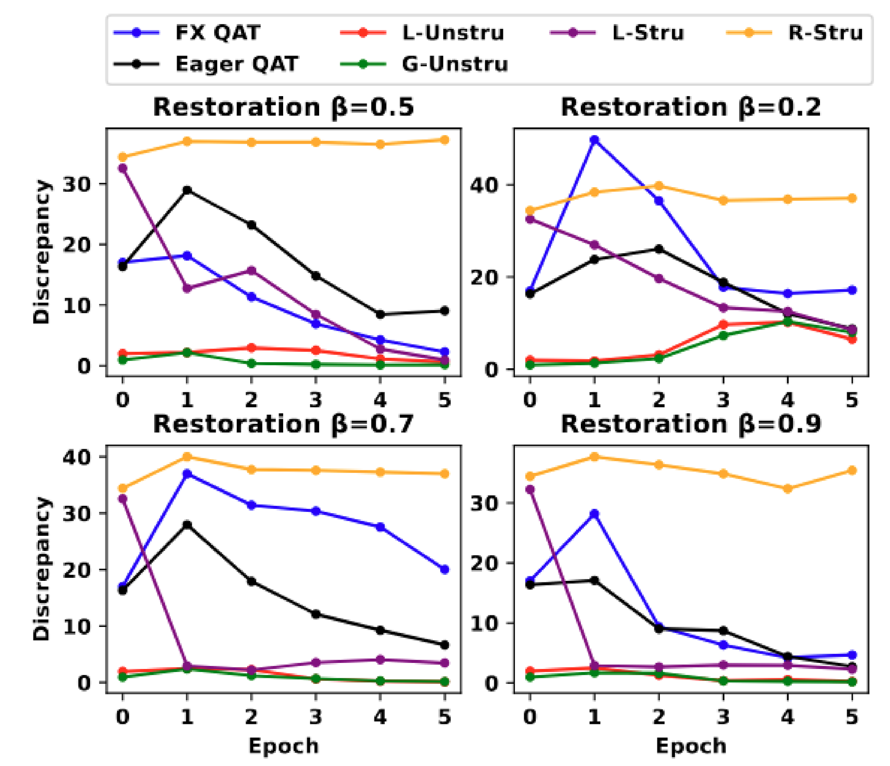


Fig. 2: Restoration performance for different compression methods with different β.

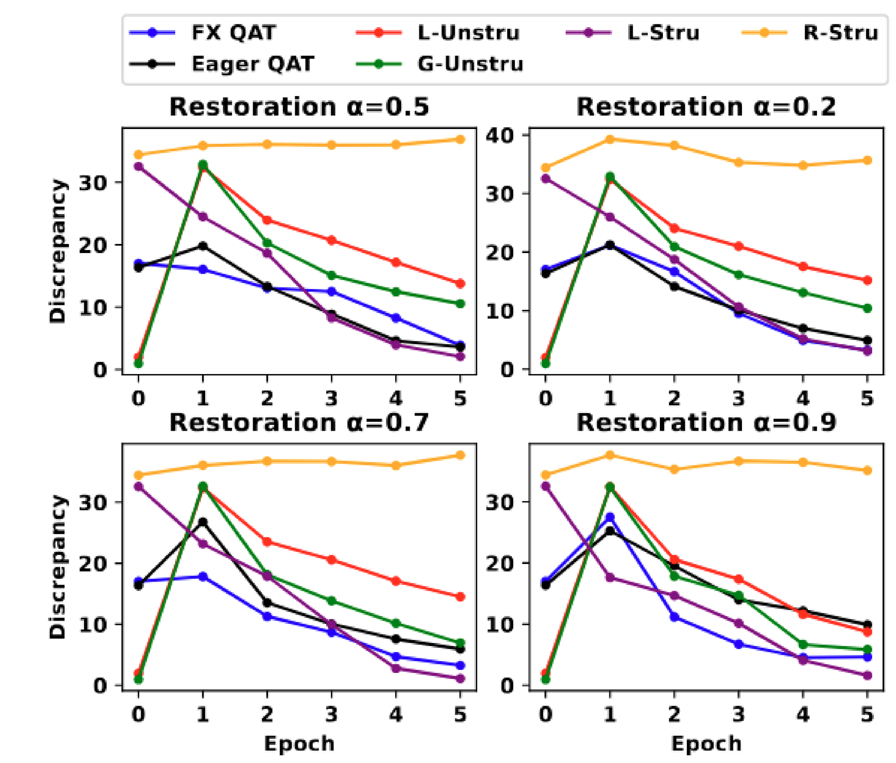


Fig. 3: Restoration performance for different compression methods with different α.

TABLE II: 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

TABLE III: Knowledge distillation restoration performance

Methods	Ori.	$\alpha = 0.5$	$\alpha = 0.2$	$\alpha = 0.7$	$\alpha = 0.9$
FX QAT	17.0159	3.8791	**3.2217**	3.2668	4.6501
Eager QAT	16.3547	**3.6152**	4.9157	5.9510	9.9261
L-Unstru	1.9602	13.7829	15.1636	14.5191	8.7380
G-Unstru	0.9487	10.5672	10.4162	6.9580	5.8600
L-Stru	32.5631	2.0776	3.1321	**1.1076**	1.6181
R-Stru	34.4265	36.9128	35.6922	37.7178	35.1513

Comparison (MaxDNet and EqBaB)

TABLE I: Comparison between reachability method and EqBaB on MNIST and CIFAR10

Dataset	Network 1		Network 2		Noise	Reachability		EqBaB	
	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

Broader Impacts

Impact to Society

- The techniques and tools will benefit CPS and ML applications to provide lifetime safety assurance.

Education and Outreach

- CPS workforce training and education, one student won DoD scholarship.
- Develop a new CPS course at AU.
- Engage in K-12 outreach activities, GenCyber Camp, High School Spotlight Event, etc.