CAREER: Decision Procedures for High-Assurance AI-controlled CPS

PI: Yasser Shoukry – University of California, Irvine

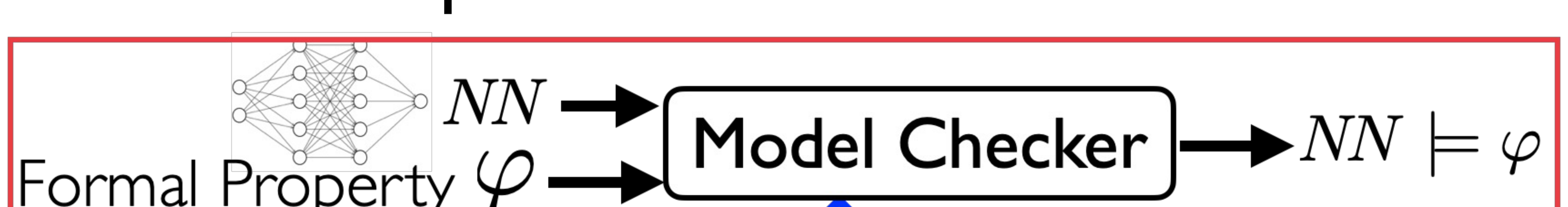
Award ID#: 2002405

Objectives:

- Develop scalable formal methods to reason about the safety and reliability of Learning Enabled CPS.
- Characterize the environments for which LE-CPS are not safe to operate.
- Train NNs with provable guarantees in terms of performance, robustness, and safety.

NN Design-for-Verifiability:

- Formal verification of NNs is NP-hard.
- Can we find NNs with special structure or semantics that lead to “fast” verification?
- Can we replace the ReLU activation non-linearity with one that is amenable to “fast” verification?
- **Result:** Formal verification of NNs with millions of parameters in few seconds.



Verify **“easier” ReLU-NN architectures** (NN structure/semantics)

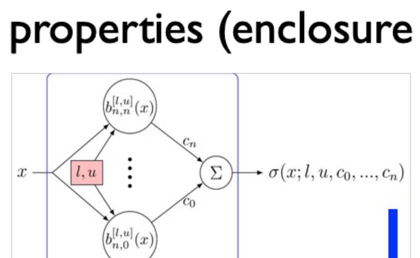
Verify NNs with **“easier” activation units** (use nice properties of other nonlinear functions)

Two-Level Lattice (TLL) NNs are **verifiable in polynomial time***

(* in the number of neurons)

Verify **“structured” properties** (use NN structure/semantics)

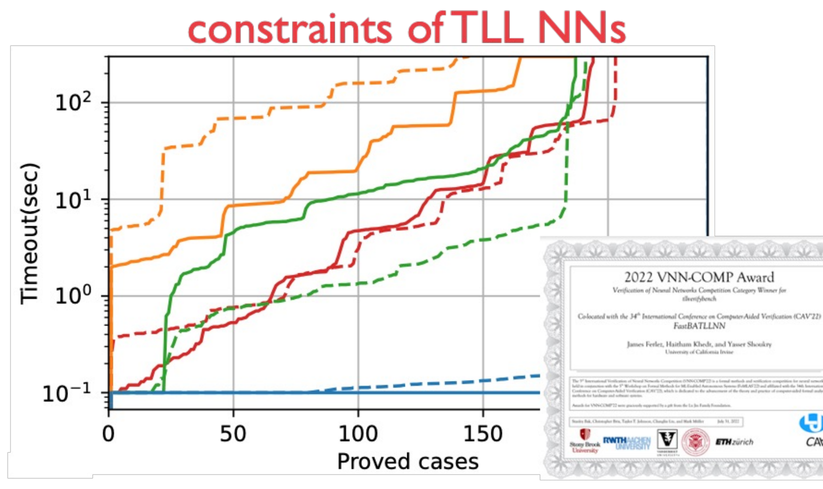
Bernstein Polynomials enjoy several “nice” properties (enclosure of range and subdivision)



$$\sigma_n^{[l,u]}(x) = \sum_{k=0}^{n} c_k b_{n,k}^{[l,u]}(x), \quad x \in [l,u],$$

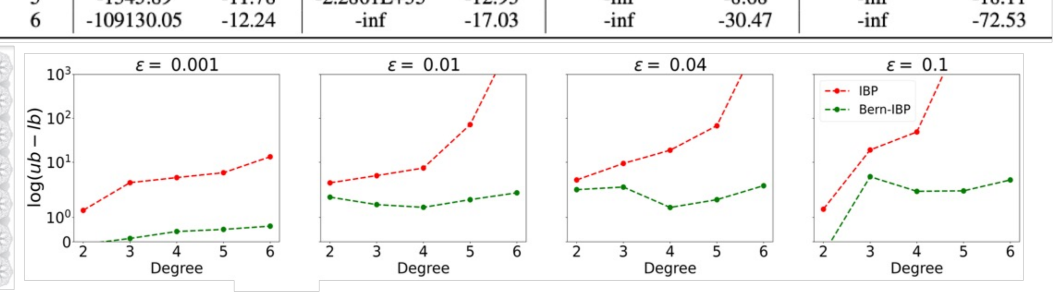
$$b_{n,k}^{[l,u]}(x) = \frac{\binom{n}{k}}{(u-l)^n}(x-l)^k(u-x)^{n-k}$$

FastBATLLNNN: Fast Box-like constraints of TLL NNs



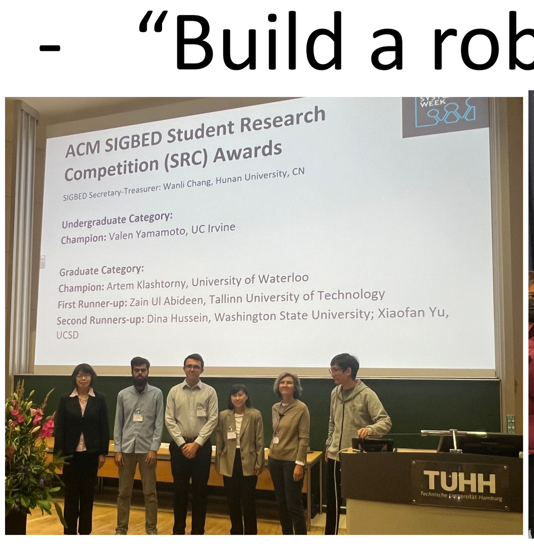
Deep Bern-Nets = Precise Bound Propagation

Order	$\epsilon = 0.001$ IBP	$\epsilon = 0.001$ Bern-IBP	$\epsilon = 0.01$ IBP	$\epsilon = 0.01$ Bern-IBP	$\epsilon = 0.04$ IBP	$\epsilon = 0.04$ Bern-IBP	$\epsilon = 0.1$ IBP	$\epsilon = 0.1$ Bern-IBP
2	-20.16	-16.63	-42.72	-16.56	-83.7	-22.22	-71.33	-8.25
3	-96.55	-12.16	-205.09	-14.02	-34962.84	-22.91	-2302369792	-137.07
4	-3550.07	-10.15	-56758.56	-13.72	-1.09065E+15	-9.23	-8.24695E+24	-23.03
5	-1345.89	-11.78	-2.2861E+35	-12.93	-inf	-8.68	-inf	-18.11
6	-109130.05	-12.24	-inf	-17.03	-inf	-30.47	-inf	-72.53

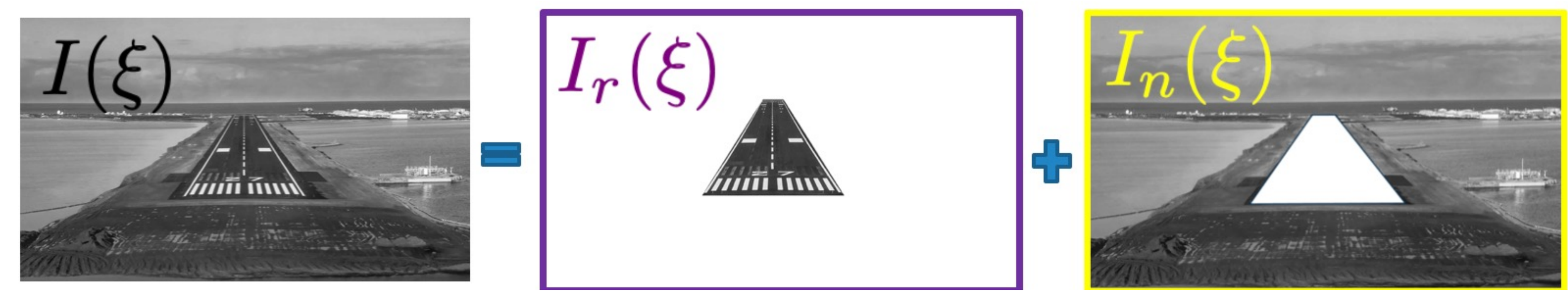


Outreach and Education:

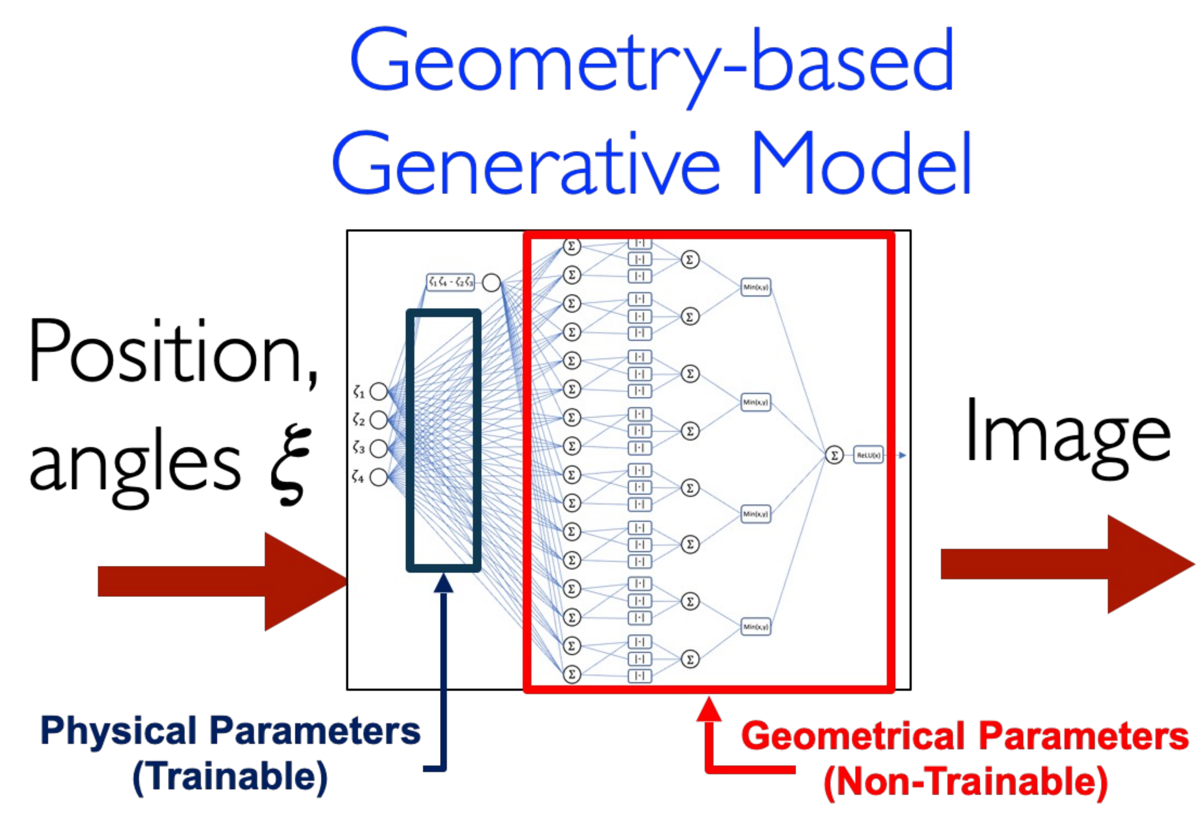
- Undergraduate student (Valen Yamamoto) wins the ACM SIGBED Student Research Competition.
- Lead elementary/middle school teams to win the regional-/state-level Robotics competitions.
- “Build a robot in a weekend” K-4 workshops.

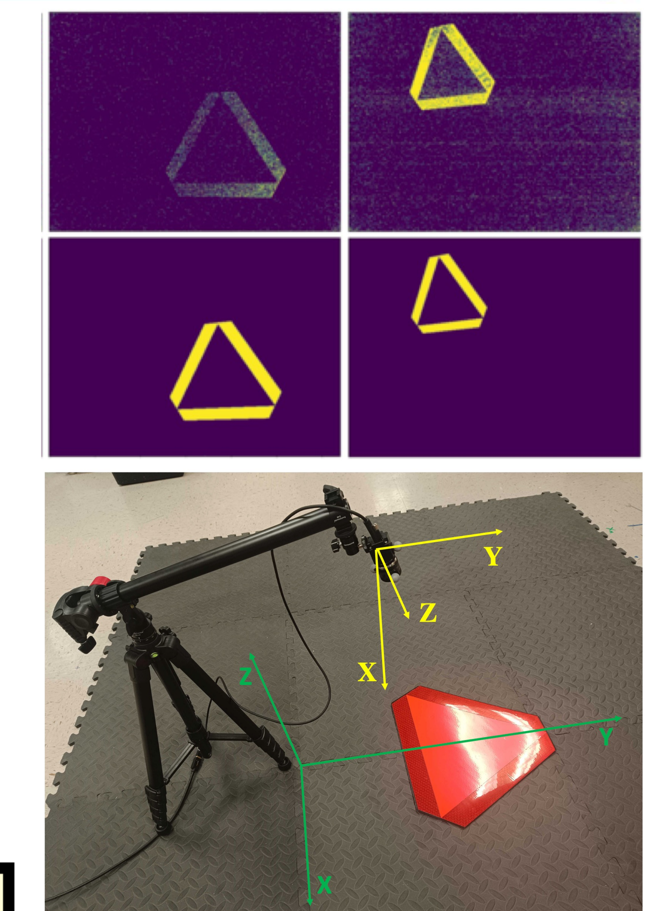


Assured NN Perception using Geometry-based Generative Models (GGMs):



Given: A camera image $I(\xi) = I_r(\xi) + I_n(\xi)$
Given: User defined error $\epsilon > 0$
Design: NN Estimator $\hat{\xi} = \mathcal{NN}(I)$ such that $||\xi - \hat{\xi}|| \leq \epsilon$

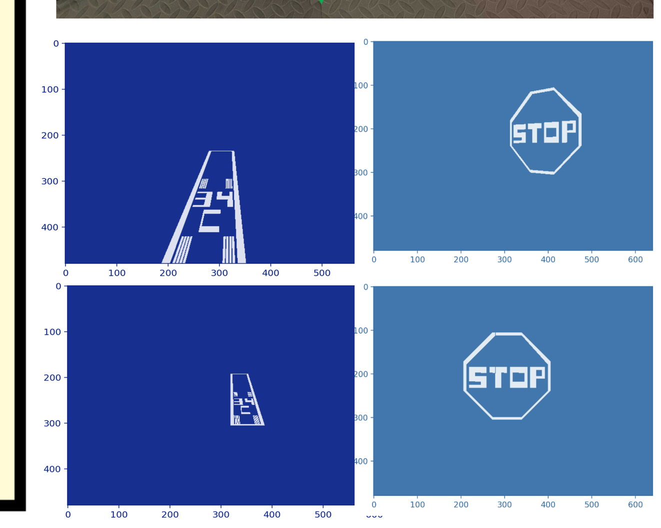


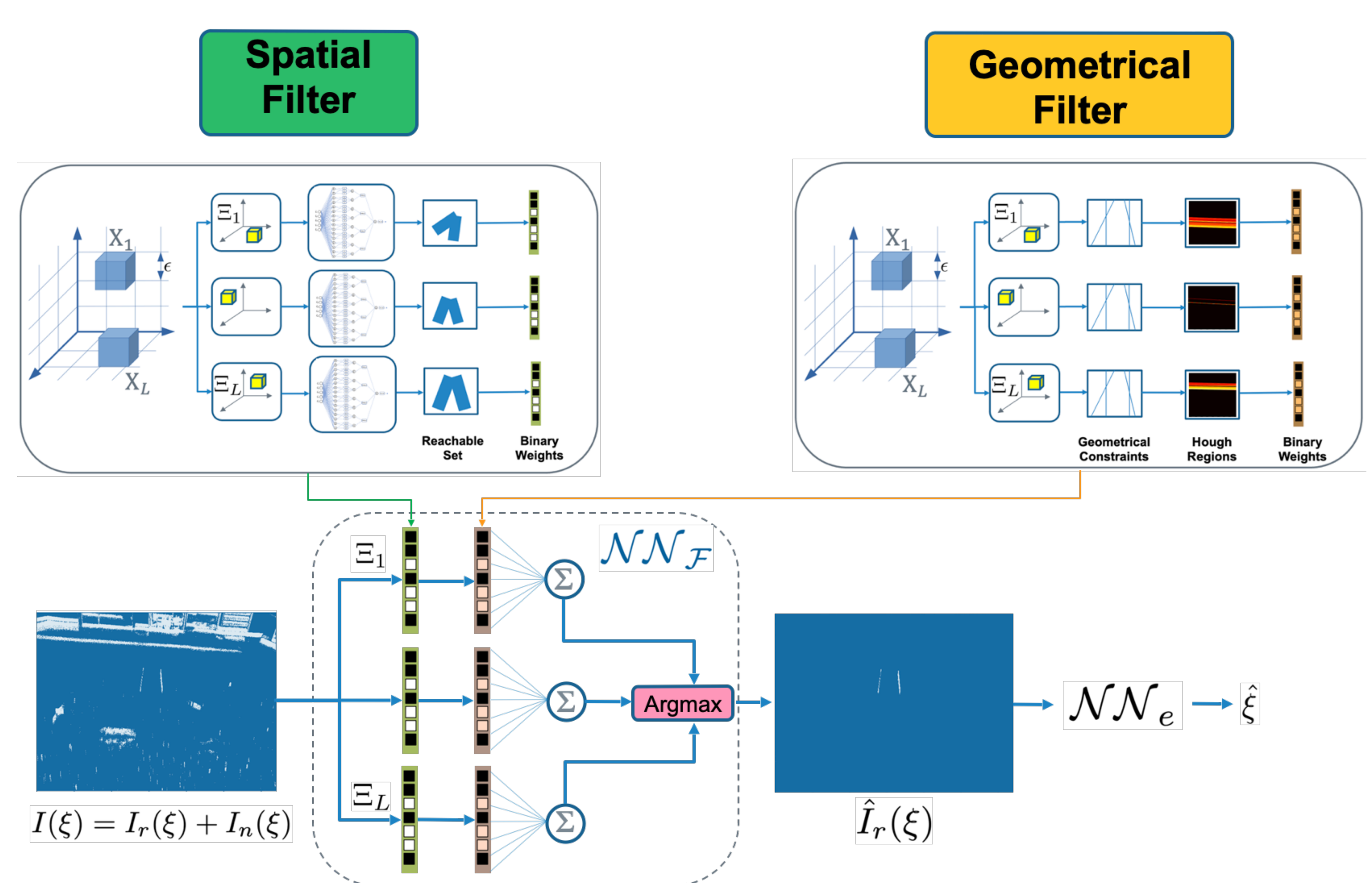


Theorem (Informal Version)

For any 2D object that can be formed as unions and intersection of polytopes, then the Geometry-based Generative Model (GGM) Neural Network is equivalent to the Pin-hole camera model, i.e.,

$$I_r(\xi) = GGM_r(\xi)$$





Theorem (Informal Version)

Given:

- A camera image: $I(\xi) = I_r(\xi) + I_n(\xi)$
- Partitioning of the state space: $\Xi_1, \ldots, \Xi_l$

Under the following assumptions:

(i) $I_n(\xi) \notin \{\mathcal{NN}_r(\xi) | \xi \in \Xi\}$

(ii) $\forall \xi \in \Xi^*. [I_n(\xi) \otimes \mathcal{NN}_r(\xi) = \mathbf{0}_{a,b}]$

The following holds:

$$\hat{\Xi} = \Xi^*$$

$$\hat{I}_r = I_r(\xi)$$

$$||\xi - \hat{\xi}|| \leq 4 L_h \delta$$

Where: $(\hat{\Xi}, \hat{I}_r) = \mathcal{NN}_F(I(\xi))$

Other objects can not be generated by the same geometric generative model, i.e., **other objects do not look like the target object.**

Other objects do not appear in the neighborhood of the target object.

NN output:

- The partition where the state belongs
- Filtered image estimate.

Bound:

L_h Lipschitz constant of Generative Model

δ Radius of the infinity ball used to partition the state space