Validation and Verification of Learning-Enabled State Estimation System for Robotics

Youcheng Sun
Queen's University Belfast, UK
02 July 2020

In collaboration with Xiaowei Huang, Wei Huang, Yifan Zhou, Simon Maskell (University of Liverpool, UK)
James Sharp, Alec Banks (Defence Science and Technology Laboratory (DstI), UK)
and Jie Meng (Loughborough University, UK)

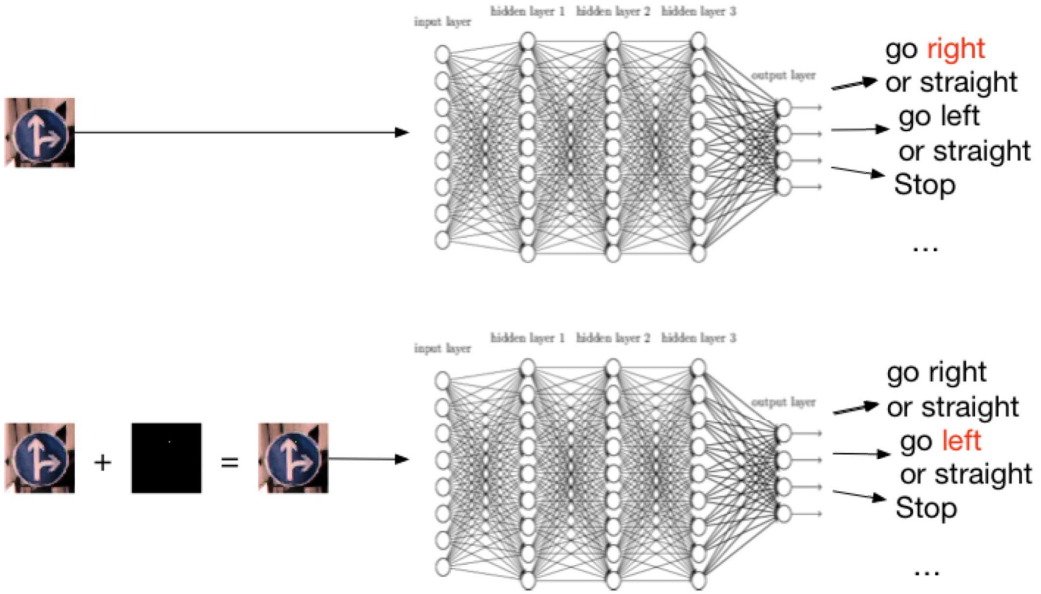
Learning Enabled Autonomous Systems

- Increased complexity in capability is driving a move towards increasing levels of autonomy
  - ⇝ safety related consequences / require higher levels of integrity

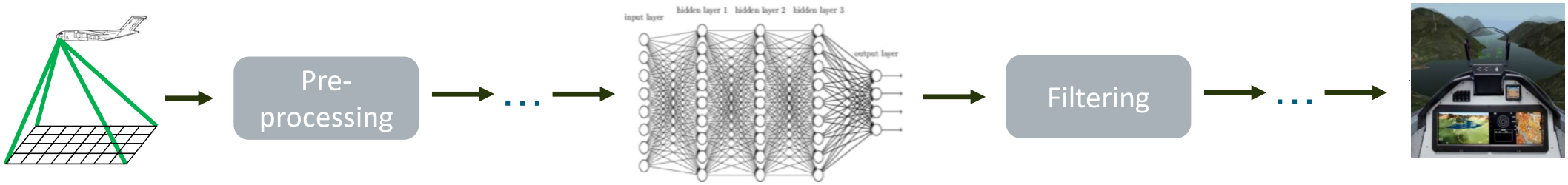
- AI systems that use Convolutional Neural Networks (CNNs)

Adversarial Examples

- An adversarial example is an input that is mislabeled by a neural network *after a minor, perhaps imperceptible, perturbation*



Neural Network as A Functional Component

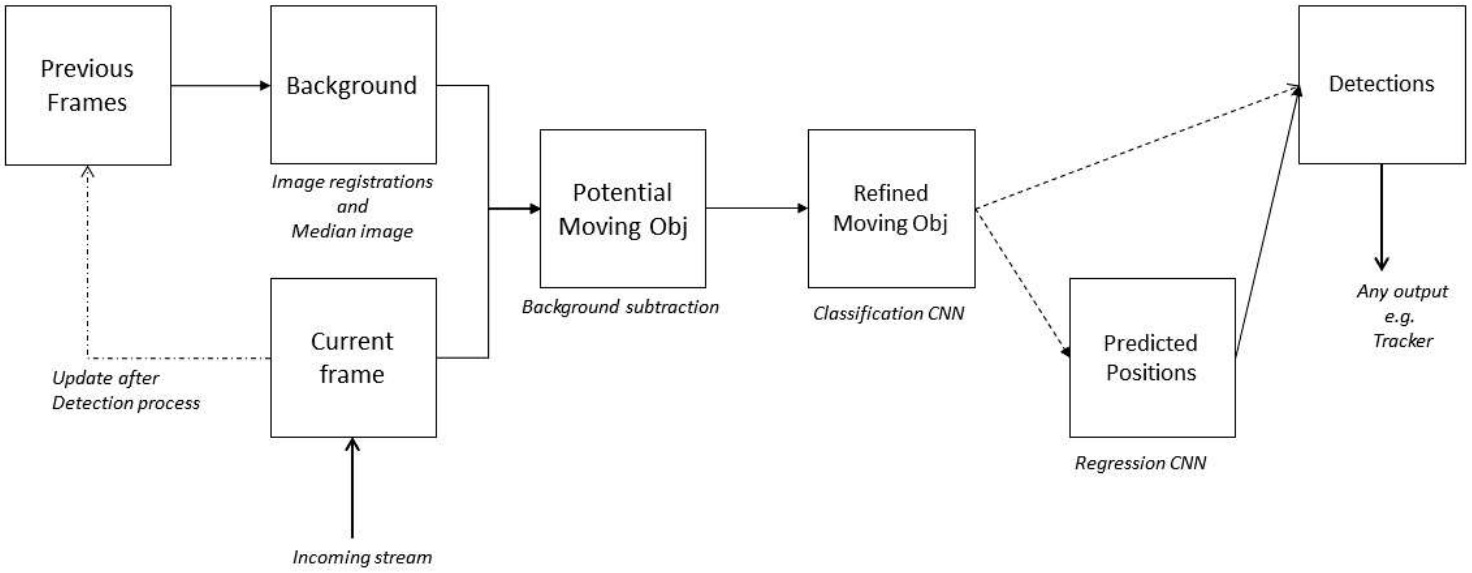


(RQ1) Can the system as a whole be resilient against the deficits discovered over the learning components?

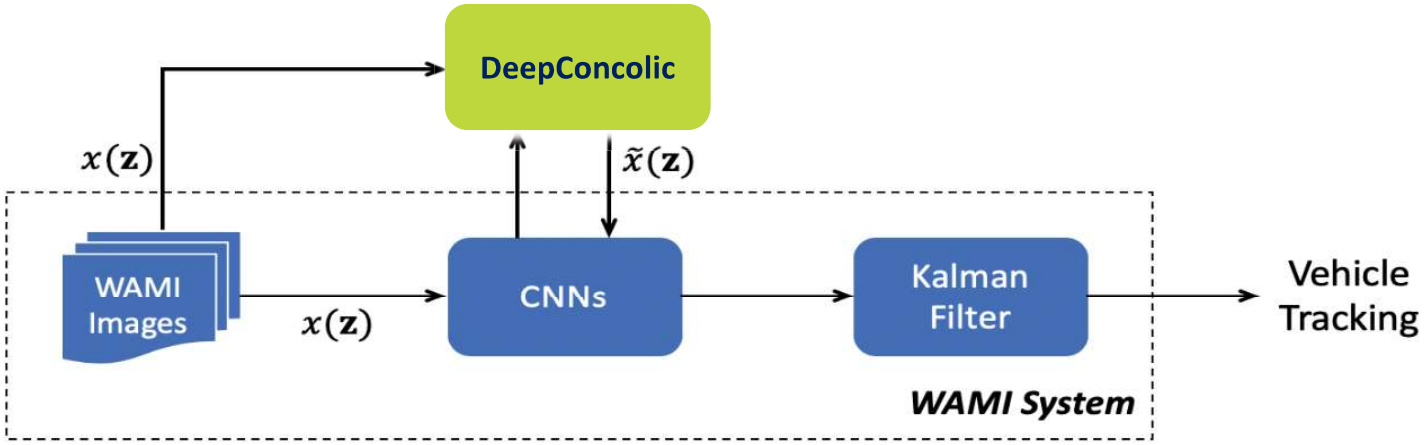
(RQ2) Is there new uncertainty needed to be considered in terms of the interaction between learning and non-learning components?

A Real-World Vehicle Tracking System

- The tracking system includes detecting multiple ground vehicles over the high-resolution Wide Area Motion Imagery (WAMI)
- Architecture of the vehicle detector



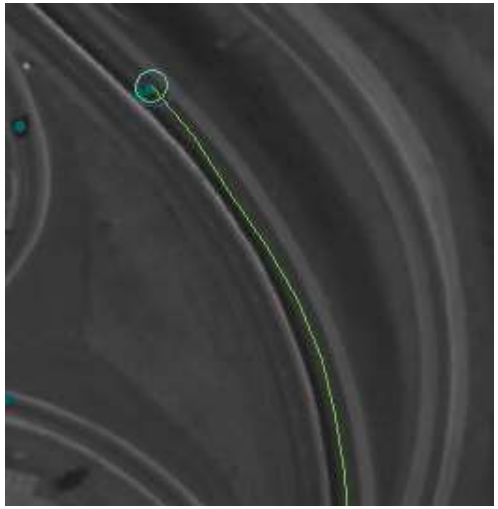
Reliability Testing Framework

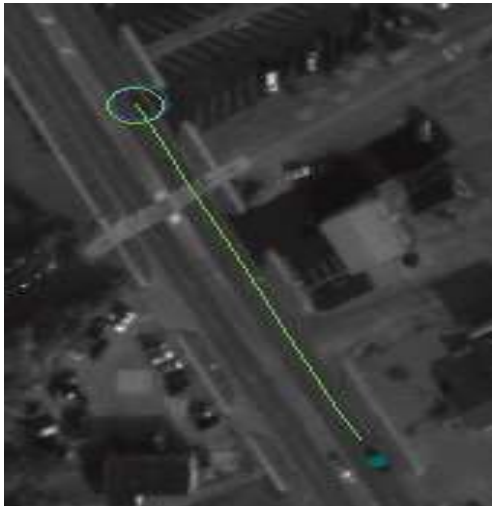


- We use the DeepConcolic to generates test cases and adversarial examples for CNNs following the MC/DC test conditions.

Examples

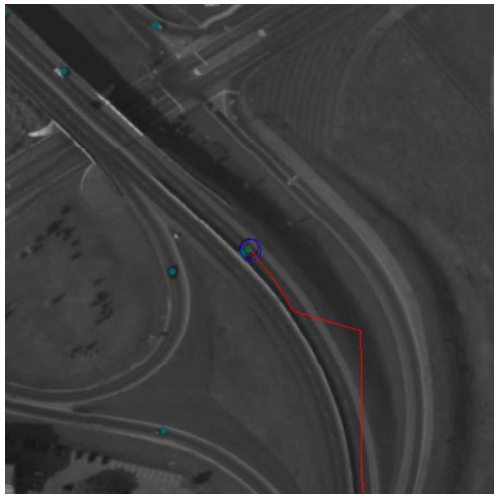
- Original detected tracks from the tracking system

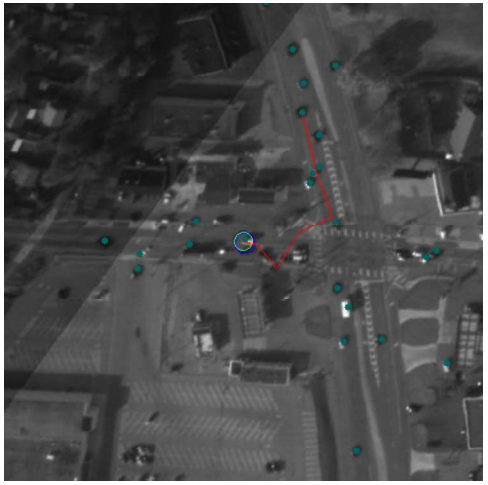


Examples

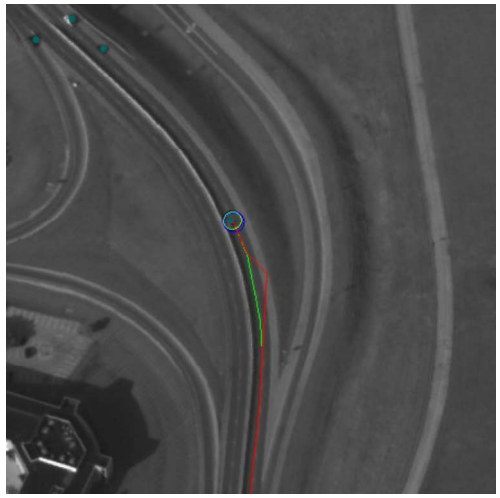
- Distorted tracks found by DeepConcolic testing



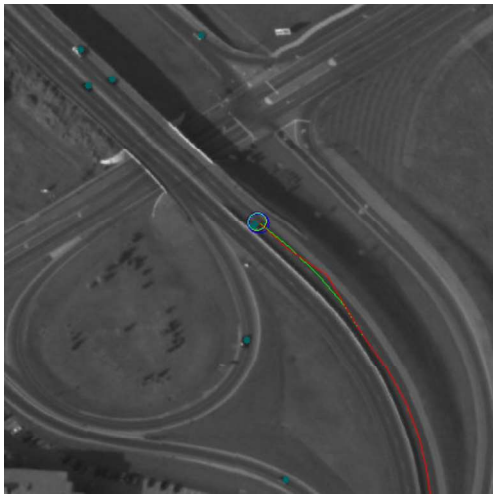



RQ1

- By only testing the deep learning component it may not be sufficient to mislead the overall tracking system.
- Adversarial tracks (red) after testing different parts of the original track (green)



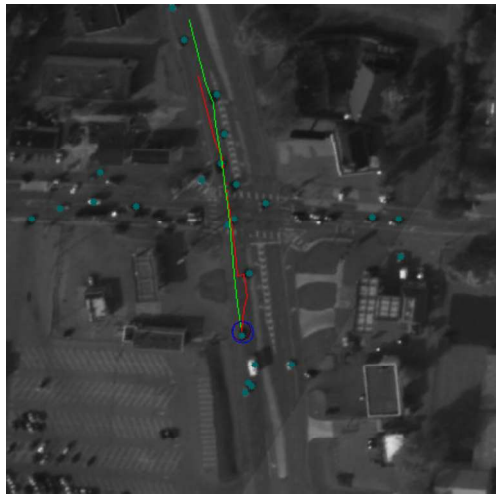
Test(5,2)



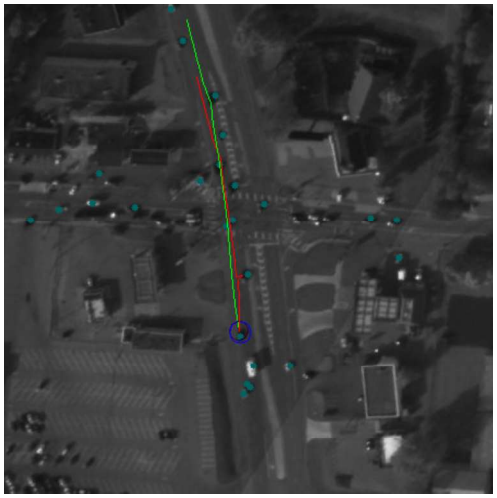
Test(11,2)

RQ2

- Changing more frames does not necessarily result in larger deviation from the original track.



Test(9,2)



Test(9,3)

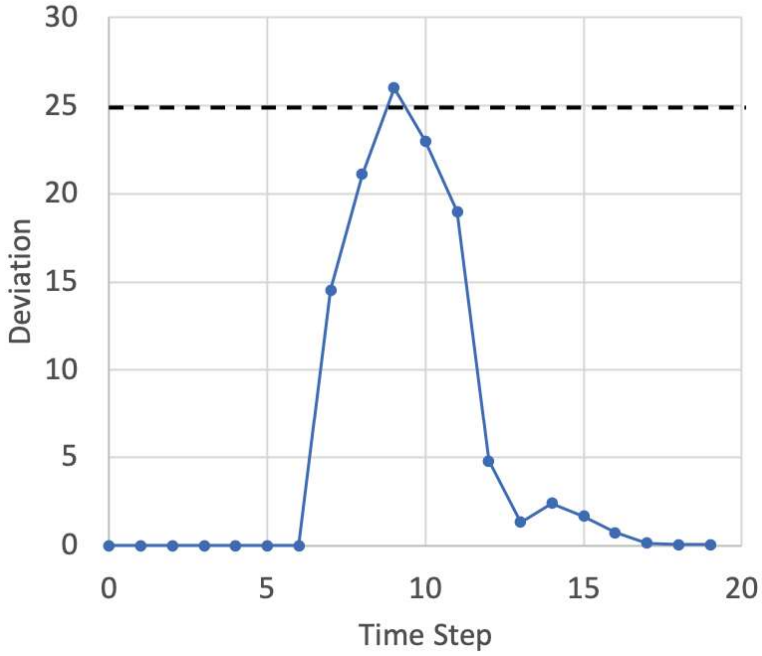
- YES!

Robustness vs Resilience

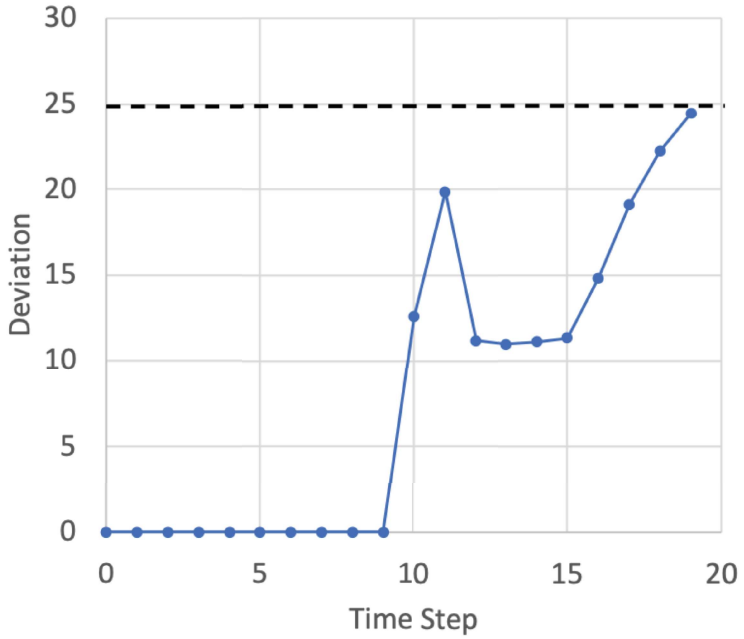
- Robustness is an enforced measure to represent a system's ability to consistently deliver its expected functionality by accommodating disturbances to the input.

- Resilience indicates an innate capability to recover sufficient functionality in the face of challenging conditions against risk or uncertainty, while keeping a certain level of vitality and prosperity.

Resilient But Not Robust



Loss of Resilience



Formalism

- Robustness

  min adversary payoff

  s.t. diff (original_path, adversarial_path) > $\varepsilon_{robustness}$

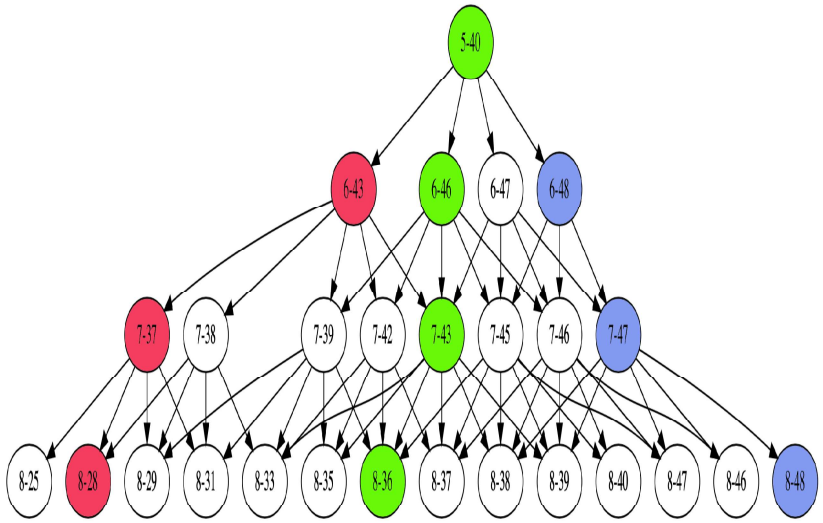
- Resilience

  min distance (original_path, adversarial_path)

  s.t. diff (original_destination, adversarial_destination) > $\varepsilon_{resilience}$

14

WAMI Tracking as A Labelled Transition System



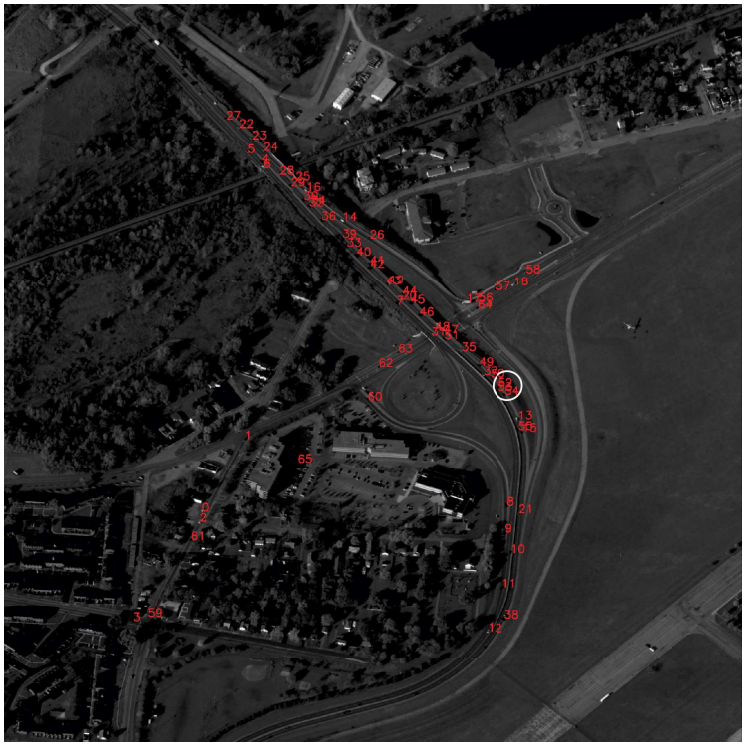
- The root node on top represents the initial state
- Each layer comprises all possible states of at step k of WAMI tracking
- Each transition connects a state at step to another state at step k

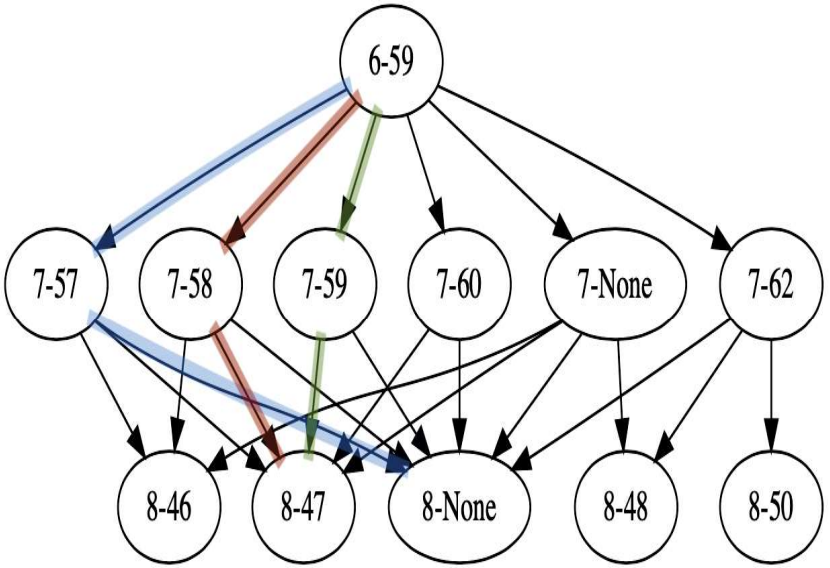
Solutions

- Verification
  - Exhaustive search for all possible tracks
- Heuristic
  - At each step, select the most distant adversarial state

Results: Verification

- Attack the original track from time step k = 6 to k = 8





Results: Verification vs Heuristic

- Sample 100 tracks of length 20
- Attack 1 to 4 WAMI frames

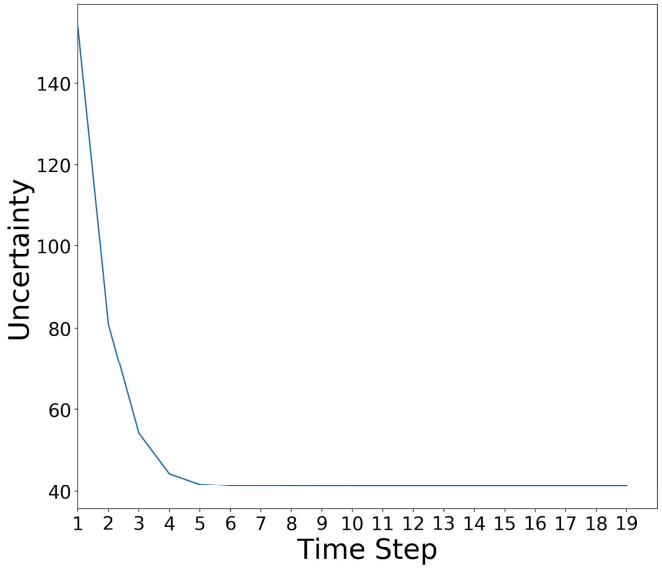
Time in seconds

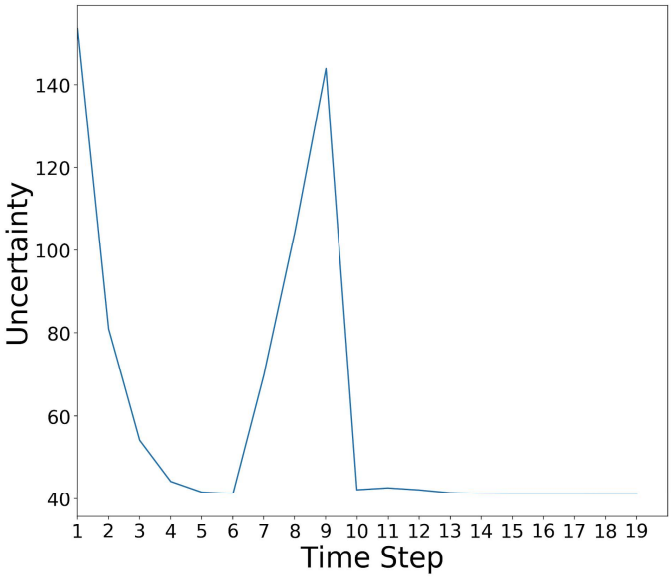
Algorithm	ε_{avg}	T	$dist$	Probability of Finding Best Adv. Track
heuristic search	0.63	78	93	80 %
verification	0.65	3465	117	100 %

The heuristic finds optimal solutions in 80% cases

Defence

1. Uncertainty monitoring





2. Joint Kalman filters

THANK YOU!

- References

1. Y. Sun, Y. Zhou, S. Maskell, J. Sharp and X. Huang, “Reliability Validation of Learning Enabled Vehicle Tracking”, ICRA 2020.

2. W. Huang, Y. Zhou, Y. Sun, Y. Zhou, J. Sharp, S. Maskell and X. Huang, “Practical Verification of Neural Network Enabled State Estimation System for Robotics”, IROS 2020.

3. Y. Sun, X. Huang, D. Kroening, J. Sharp, M. Hill and R. Ashmore, “DeepConcolic: Testing and Debugging Deep Neural Networks”, ICSE 2019.

4. Y. Sun, M. Wu, W. Ruan, X. Huang, M. Kwiatkowska and D. Kroening, “Concolic Testing for Deep Neural Networks”, ASE 2018.

- Materials for this work are publicly available https://github.com/havelhuang/wami_detector_resilience_verification