Safe and Generalizable Autonomy under Uncertainty

Esen Yel

Postdoctoral Scholar, Stanford University

esenyel@stanford.edu

Motivations:

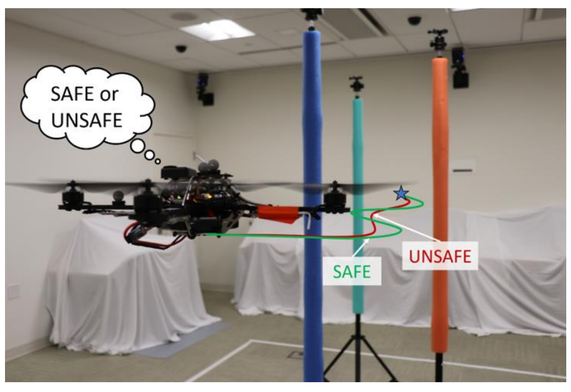
- Autonomous systems are used in safety-critical applications
- They face various types of uncertainty (e.g., disturbances, environmental changes, etc.) risking safety

Research Goal:

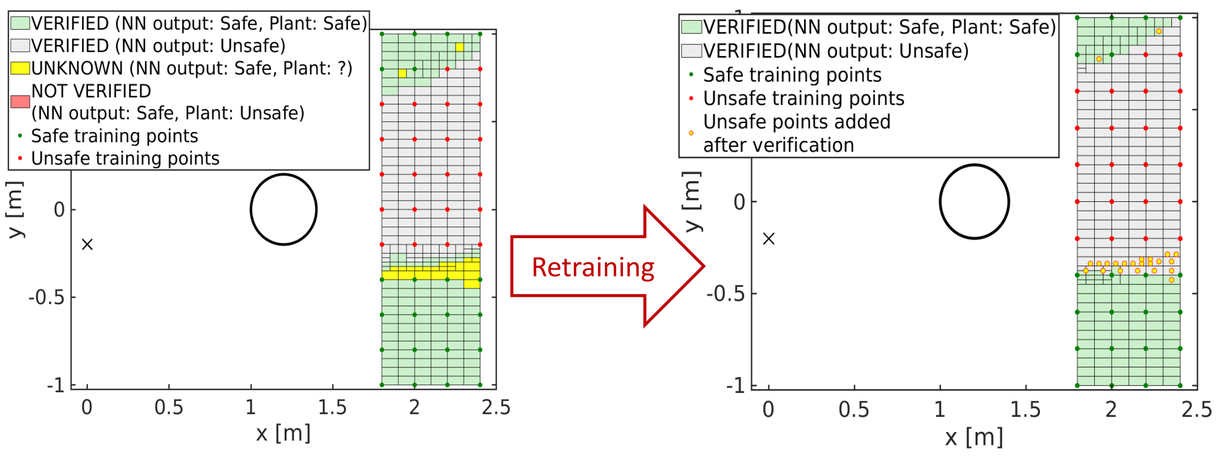
- Achieve safe, trustworthy, and generalizable autonomy for systems under uncertainty

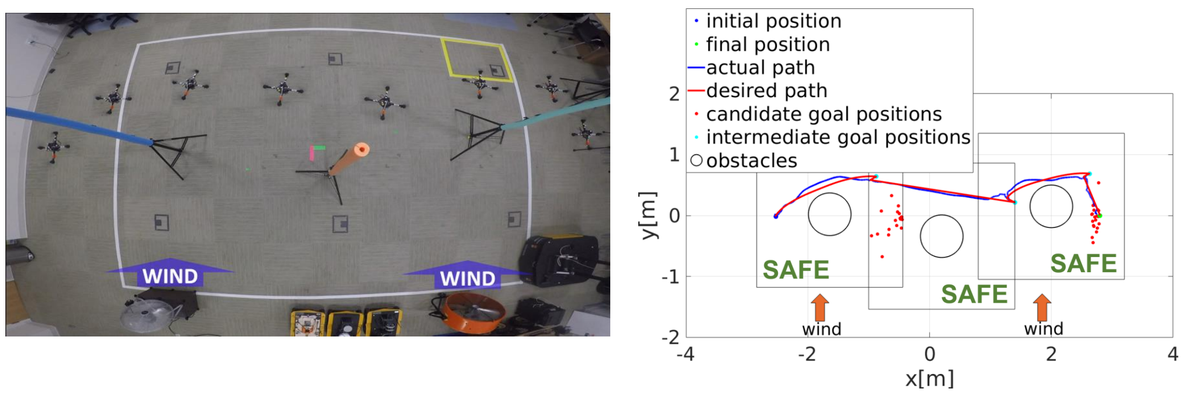
Assured Runtime Monitoring and Replanning

Reachability analysis (RA) can be used to predict the future states of the system under uncertainty, but it can get computationally expensive



Key idea: Neural networks to imitate the safety decisions of RA, and verification tools to ensure safety

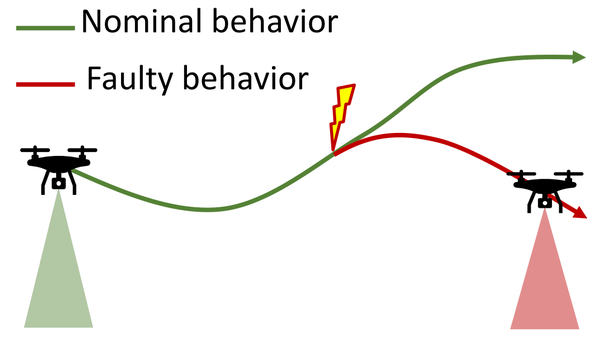




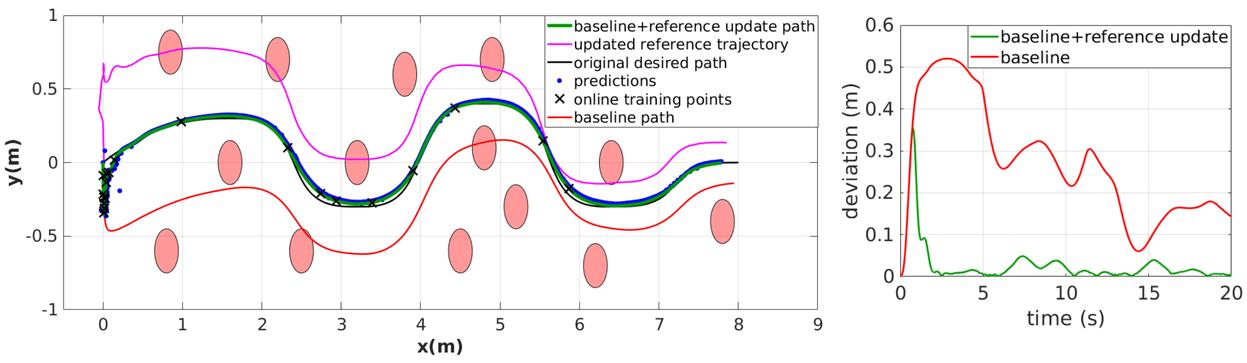
E. Yel, T. Carpenter, C. di Franco, R. Ivanov, Y. Kantaros, I. Lee, J. Weimer, N. Bezzo, *"Assured Runtime Monitoring and Planning: Towards Verification of Neural Networks for Safe Autonomous Operations"*, IEEE Robotics & Automation Magazine, vol. 27, no. 2, pp. 102-116, 2020

Online Adaptation to New Conditions

It is critical to adapt to new conditions (e.g., faulty behaviors, new traffic conditions) for long term autonomy

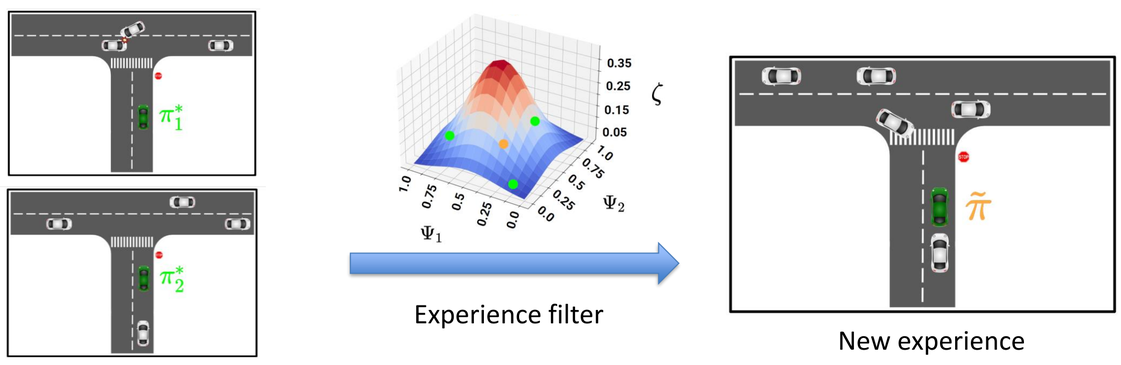


Key idea: Use meta-learning to efficiently fine-tune predictions models and replan for new situations



E. Yel, N. Bezzo, *"A Meta-Learning-based Trajectory Tracking Framework for UAVs under Degraded Conditions"*, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2021

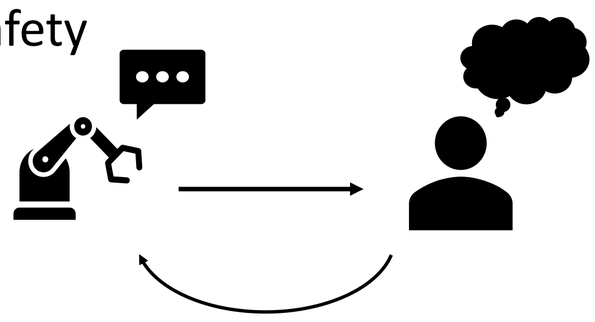
Key idea: Transfer experiences from the past to new conditions through experience filter



A. Yildiz, **E. Yel**, A. L. Corso, K. H. Wray, S. J. Witwicki, and M. J. Kochenderfer, *"Experience filter: Transferring past experiences to unseen tasks or environment"*, IEEE Intelligent Vehicles Symposium 2023 (accepted)

Future Directions

- Safety/efficiency trade-off
- Learning task representations and relations for better adaptation
- Trustworthiness beyond safety
  - Explainability
  - Alignment
  - Communication



Broader Impacts

- Safety-critical applications
  - Healthcare
  - Transportation (remote communities)
  - Human-robot interaction (elderly care)
- Improved technological literacy through explainable autonomy
- Knowledge transfer through collaborations