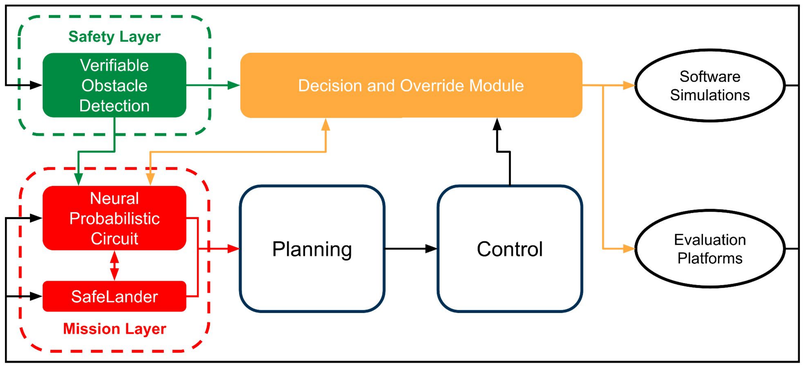
Certifiably Safe and Robust Perception for Urban Air Mobility

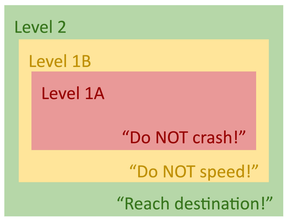
PIs: Lui Sha, Bo Li, Naira Hovakimyan

Collaborators: Hunmin Kim, Hyung-Jin Yoon, Han Zhao, Multiped LLC

Doctoral Candidates: Simon Yu, Mikael Yeghiazaryan, Jiawei Zhang

Overview and Rationale





Perception Simplex: Using simplicity to guard complexity

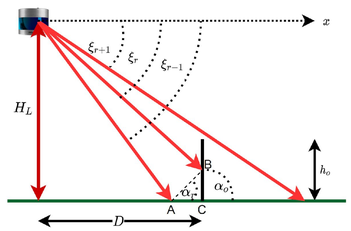
- Two-Layer Design: Separating safety assurance from mission objectives.
- **Safety Layer: Detecting** any and all surrounding objects (level 1A).
- **Mission Layer: Recognizing** any and all surrounding objects (level 1B).
- Decision Module: Safety override with mission layer feedback (level 2).

Perception Simplex Design Assurance Levels (DALs)

- **Level 1A:** Collision avoidance
  - The risk of colliding with surroundings should be avoided at all costs.
- **Level 1B:** Regulatory rules
  - Without breaching level 1A, avoid violating any regulatory rules.
- **Level 2:** Mission objectives
  - Without breaching levels 1A and 1B, complete assigned missions.

Safety Layer - Verifiable Obstacle Detection

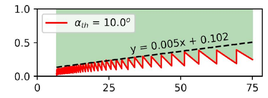
Deterministic guarantees for obstacle existence detection using LiDAR-based geometric reasoning.



Use downward-facing LiDAR to verify a clear landing path. On collision risk, decelerate and hover. Override and Control modules coordinate to maintain capability within tight bounds amid disturbances.

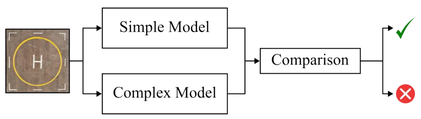
Theorem 1. The obstacle is detected at distance D, if and only if one of the following conditions are true:

1) $H_r(D) \leq h_o < H_{r+1}(D)$ AND $\alpha_{th} < atan2(H_r(D), |D - \frac{H_L}{tan(\xi_{r-1})}|)$;
2) $h_o \geq H_{r+1}(D)$.

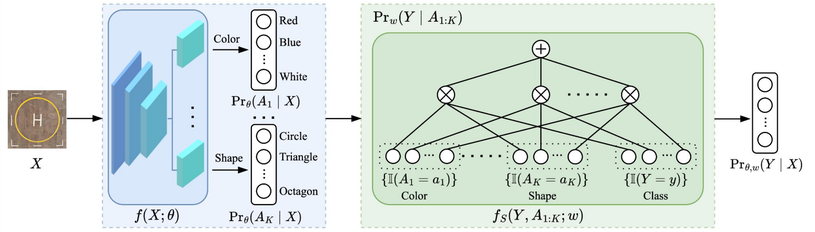


Obstacle distance (Y) vs Obstacle height (X) in meters.
Obstacles always detected in green region.

Use a simpler computer vision model to verify a more complex one's predictions. This allows us to "verify" *some* of the predictions made by the complex model.



Mission Layer - Neural Probabilistic Circuit (NPC)



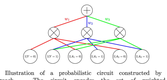
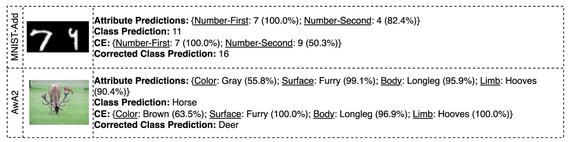


Figure 2: Illustration of a probabilistic circuit constructed by the knowledge-injected approach. The circuit encodes the set of weighted logical rules: $\{w_1 : (\mathbb{I}(A_1 = 0) \wedge \mathbb{I}(A_2 = 0) \wedge \mathbb{I}(Y = 0)),\ w_2 : (\mathbb{I}(A_1 = 1) \wedge \mathbb{I}(A_2 = 1) \wedge \mathbb{I}(Y = 1)),\ w_3 : (\mathbb{I}(A_1 = 0) \wedge \mathbb{I}(A_2 = 1) \wedge \mathbb{I}(Y = 1))\}$.

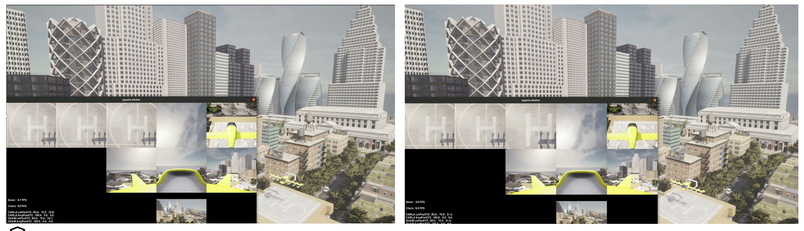
$$\Pr_{\theta,w}(Y = y \mid X = x)$$
$$= \sum_{a_{1:K}} \Pr_{\theta,w}(Y = y, A_{1:K} = a_{1:K} \mid X = x)$$
$$= \sum_{a_{1:K}} \Pr_w(Y = y \mid A_{1:K} = a_{1:K}, X = x) \cdot \Pr_\theta(A_{1:K} = a_{1:K} \mid X = x)$$
$$= \sum_{a_{1:K}} \underbrace{\Pr_w(Y = y \mid A_{1:K} = a_{1:K})}_{\text{task predictor}} \cdot \underbrace{\prod_{k=1}^{K} \Pr_\theta(A_k = a_k \mid X = x)}_{\text{attribute recognition model}}.$$

Table 2: Classification accuracy of NPCs and four baseline models on four benchmark datasets over five random seeds. "*" denotes uninterpretable models. The best results for each dataset are highlighted in bold, while the second-best results are underlined.

Model	MNIST-Add (%)	GTSRB (%)	CelebA (%)	AwA2 (%)
DNN*	99.057 ± 0.08	$\underline{99.939} \pm 0.04$	$\mathbf{36.963} \pm 0.72$	$\mathbf{93.351} \pm 0.17$
CBM	98.606 ± 0.03	99.810 ± 0.04	16.552 ± 0.87	82.286 ± 0.47
CEM*	98.740 ± 0.10	99.736 ± 0.06	25.218 ± 0.30	85.102 ± 0.27
DCR	94.597 ± 2.05	87.071 ± 6.93	7.055 ± 3.04	44.117 ± 10.05
NPC(Data)	$\underline{99.171} \pm 0.11$	99.888 ± 0.08	$\underline{33.739} \pm 0.90$	$\underline{87.281} \pm 0.39$
NPC(Knowledge)	$\mathbf{99.189} \pm 0.08$	$\mathbf{99.944} \pm 0.04$	31.727 ± 0.51	68.519 ± 3.54



Mission Layer - SafeLander



Post-Safety Verification

Predicate Extraction

Environmental Predicates:

```
Landing(x)=1, Altitude_below_10(x)=1,
Velocity_more_than_1(x)=1 ...
```

Planning Action Predicate:

```
PlanningMaintain(x)=1, PlanningStop(x)=0
PlanningDecelerate(x)=0, ...
```

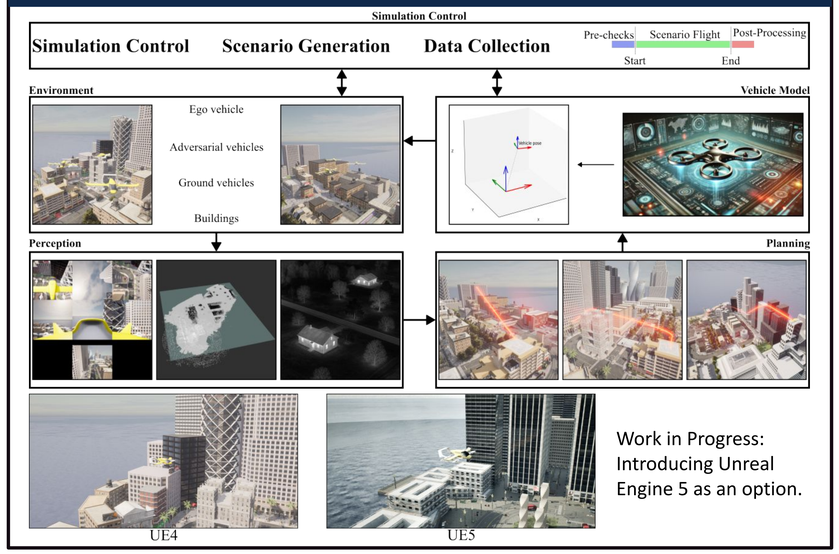
Markov Logic Network

Weight	Formula(Knowledge Rules)
10.02	Landing ∧ Altitude_below_10 ∧ Velocity_more_than_1 →
8.03	Decelerate
8.47	Cruising ∧ Altitude_below_50 → Up
8.77	Obstacles_in_1m_radius → Stop
...	Landing ∧ Velocity_more_than_2 → Decelerate

Inference

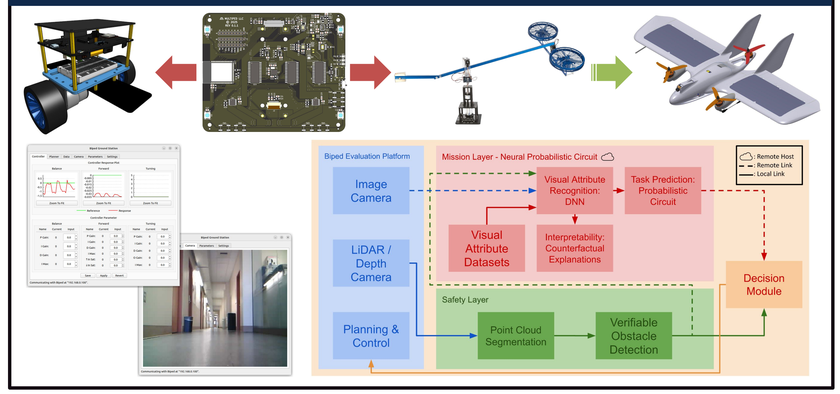
Observed Action Predicates	Probability	
Maintain(x)	0.02	*Correction!*
Decelerate(x)	0.88	Decelerate
...		

Software Simulation

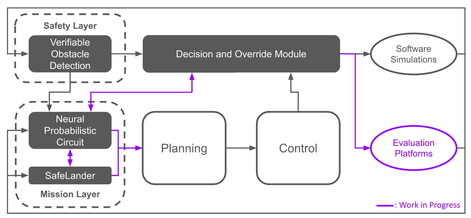


Work in Progress:
Introducing Unreal Engine 5 as an option.

Evaluation Platforms



Integration and Future Work



- Synergistic connections between the decision module and the mission layer components: online feedback and learning.
- Fusion between the mission layer components: Cooperation between the NPC and SafeLander.
- Complete development of the Biped evaluation platform and scenarios.
- Integrated evaluations of the end-to-end perception simplex pipeline with the evaluation platforms (Biped-Quanser, Minihawk).

Yu, Simon, Weixin Chen, Huajie Shao, Lui Sha, and Han Zhao. "Neural Probabilistic Circuits: Enabling Compositional and Interpretable Predictions through Logical Reasoning." arXiv preprint arXiv:2501.07021 (2025).

Bansal, Ayoosh, Hunmin Kim, Simon Yu, Bo Li, Naira Hovakimyan, Marco Caccamo, and Lui Sha. "Verifiable obstacle detection." In 2022 IEEE 33rd International Symposium on Software Reliability Engineering (ISSRE), pp. 61-72. IEEE, 2022.

Zhang, Jiawei, Xuan Yang, Taiqi Wang, Yu Yao, Aleksandr Petiushko, and Bo Li. "SafeAuto: Knowledge-Enhanced Safe Autonomous Driving with Multimodal Foundation Models." arXiv preprint arXiv:2503.00211 (2025).

Biped evaluation platform was designed and sponsored by Multiped LLC, an Illinois limited liability company.

The material presented herein is based upon work supported by NASA under Grant 80NSSC22M0070.