SafeAir: Safety-informed Tree Search for Robust Navigation under Distribution Shifts in General Aviation

Ricardo Gonzalo Cruz Castillo[1], Ingrid Navarro[2], Jean Oh[2]

[1] Tecnológico de Monterrey, Monterrey, NL, México [2] Carnegie Mellon University, Pittsburgh, PA, USA

Introduction

- In human-centered domains, such as **general aviation**, it is essential to develop robust algorithms that ensure safe and seamless navigation between humans and autonomous agents [1].
- However, **generating realistic, safety**-critical data to assess algorithm robustness remains an open challenge.

Idea: We propose **SafeAir**, a safety-informed framework as in [2] that seeks to improve navigation performance in safety-relevant scenarios **by leveraging existing datasets and trajectory prediction algorithms.**

Overview: SafeAir

SafeAir (Fig. 1) consists of a safety-informed **scenario characterization and scoring** approach for **general aviation** [1, 3] used for:

- Creating a **distribution shift**, where the most critical scenes are held out for testing as **out-of-distribution (OOD)** scenes.
- A **robustification** strategy to improve downstream prediction and navigation performance in **OOD.**

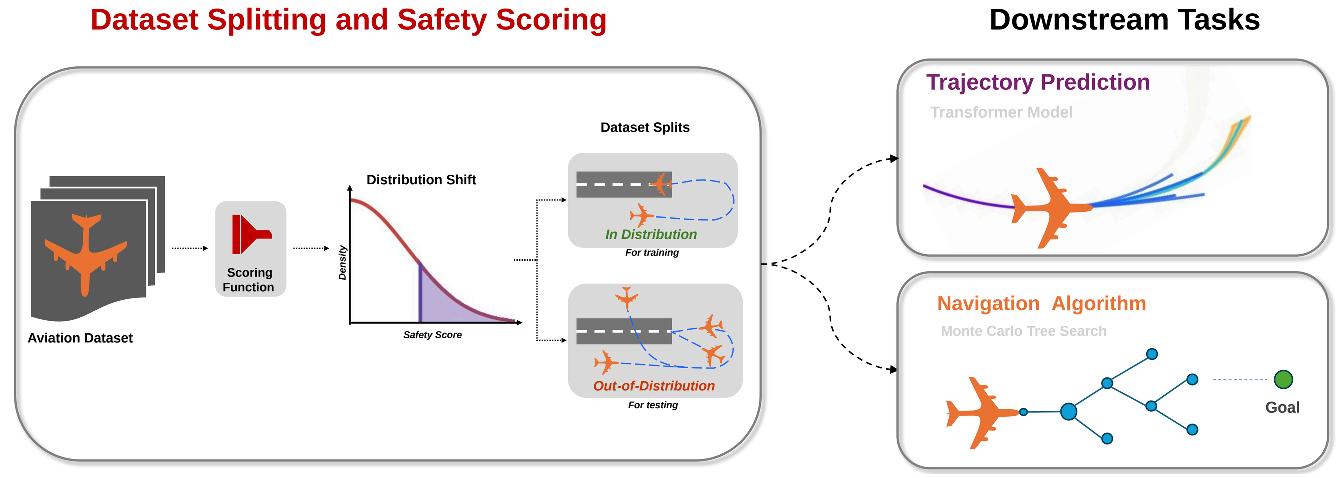

Figure 1. Overview of SafeAir

Approach: Scenario Scoring

We characterize and score ***TrajAir*** [3] scenes using ***SafeShift*** [2], where:

- For each agent in a scene, we compute **individual** and **interactive** features.
- We combine the features into a **scene score**.
- Based on the score density (Fig 2), we separate scenarios into **in-distribution** and **out-of-distribution** (Fig 3).

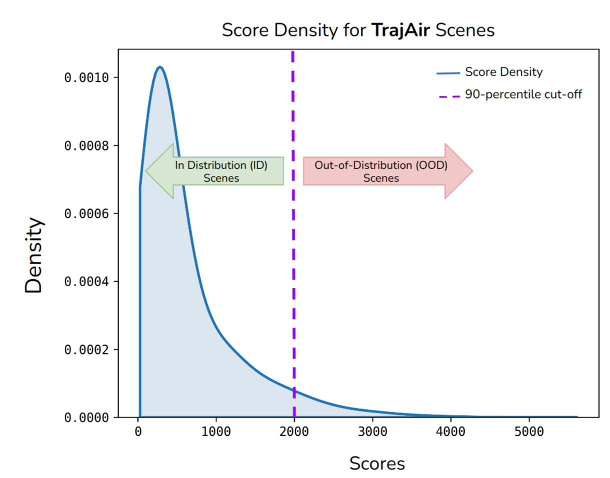

Figure 2. Scene Score Density

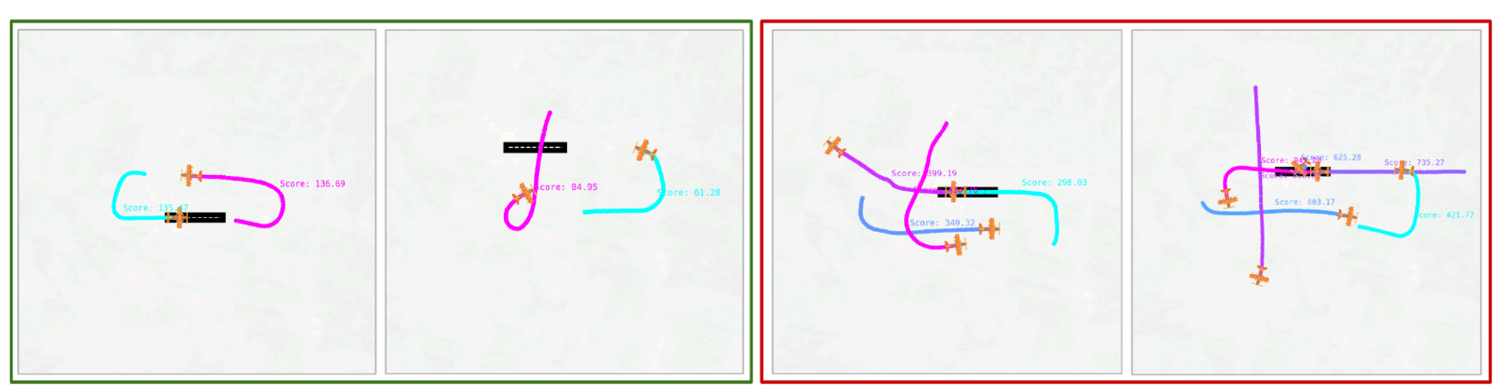

Figure 3. In Distribution (green box, left) vs Out of Distribution (red box, right) Scenarios

Approach: Robust Prediction and Planning

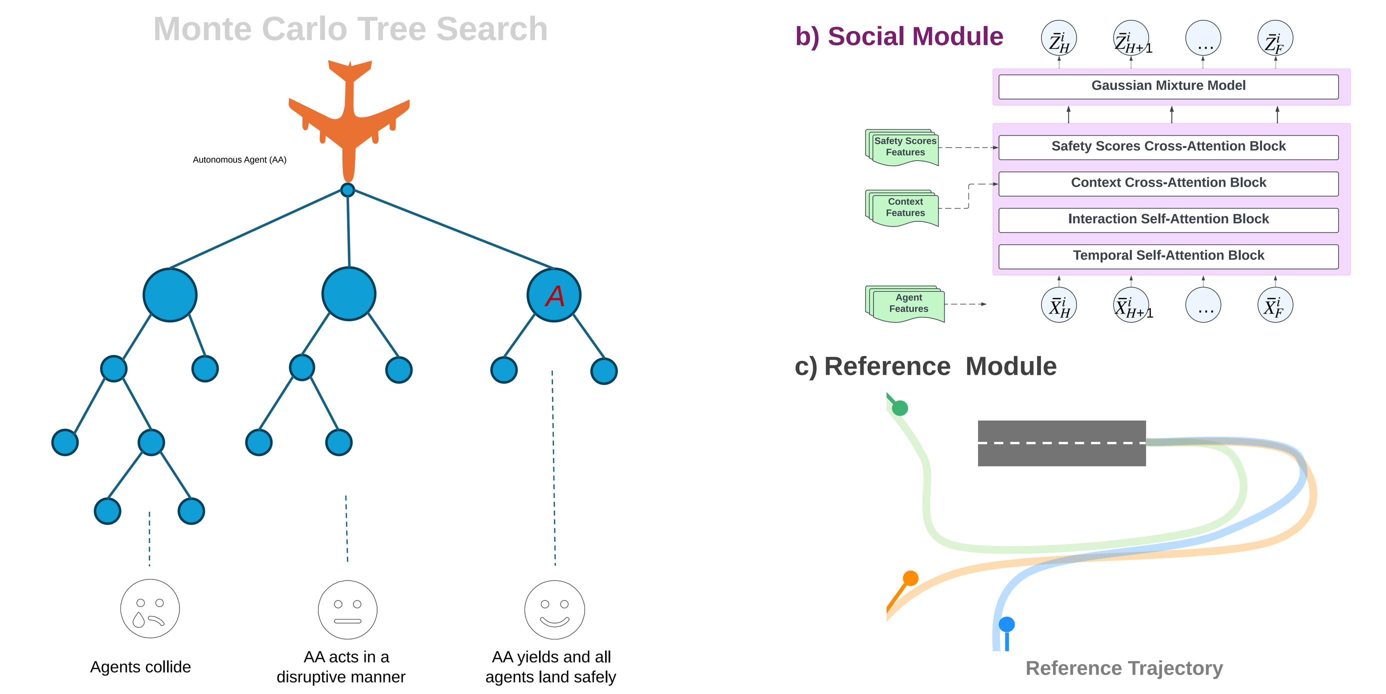

Figure 4. Navigation Algorithm Overview

Trajectory Prediction — we leverage [4] as the backbone:

- Transformer-based model that layer-wise encodes temporal, agent-to-agent, and agent-to context relationships.

Robustifying the model:

- We condition the model using **scene score** features (Fig 4, b).
- We add a **collision** avoidance component to the loss function, and loss weighing based on **scene scores** over the batch (Eq. 1).

$$\texttt{Weighted-Loss}^{(b)} = \frac{1}{B}\sum_{b}^{B}\texttt{Scene-Score}^{(b)} \cdot (\texttt{Loss}_{\texttt{reg}} + \texttt{Loss}_{\texttt{coll}})^{(b)} \qquad (1)$$

Navigation — We build upon ***SoRTS*** [1], an **MCTS**-based algorithm (Fig 4, a) biased by a **social** (Fig 4, b) and a **reference** (Fig 4 c) module, which handle short-term agent interactions and rules, respectively.

- SoRTS is originally unequipped to cope with OOD scenarios.
- Thus, we integrate our **robustified** prediction model to provide socially-aware and safety-informed decision-making.

Experiments

Trajectory Prediction — Our model is trained on ***TrajAir*** [3], under two datasets splitting methods:

- **Uniform**: random training, validation, and testing (exp 1, 2).
- **Scoring:** top 20% of scored scenarios are used for testing (exp 3, 4).
  - Score feature extraction is added to the model, and scene scores are incorporated into the loss function.

Navigation — We focus on a **landing task**, where agents are spawned around an airport and have to coordinate to land safely. We compare three ablations trained on the scoring split:

- ***SoRTS*** [1], original method.
- **SafeAir,** proposed model in a naive (exp3) vs robustified (exp4) setting.

Experiments and Results

Trajectory Prediction

- LoS **worsens** by **760%** and **120%** when using SafeShift split.
- LoS **improves** by **70%** after applying remediation, but ADE/FDE performance **drops** by **20%.**

Model	Modalities			Dataset Split	ADE / FDE	LoS GT/Pred
	Trajectory	Context	Scores			
Exp. 1	✓			Uniform	37.3 / 68.4	0.11 / 13.89
Exp. 2	✓	✓			32.2 / 57.2	0.10 / 9.24
Exp. 3	✓	✓		SafeShift	**27.2 / 48.9**	0.86 / 19.99
Exp. 4	✓	✓	✓		**32.8 / 59.9**	**0.84 / 6.16**

ADE/FDE- Average/Final Displacement Error. LoS - Loss of Separation between agent i predicted future and agent j GT/predicted future.

- Remediated model **avoids collisions** between agents (Fig 5).

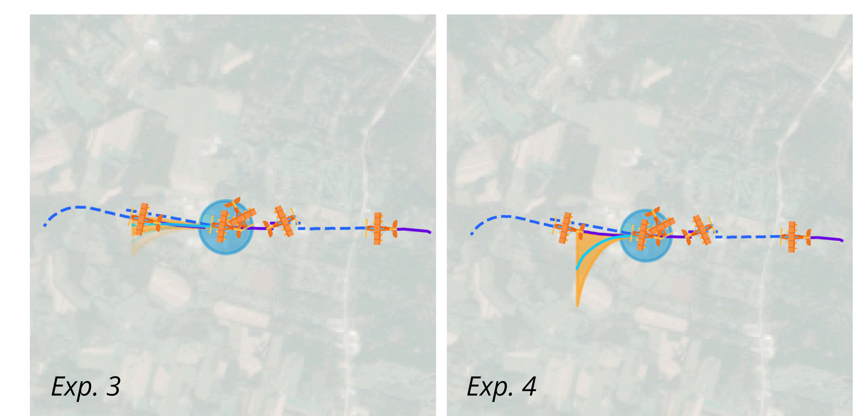

Figure 5. Prediction Results

Navigation

- SR **improves** by 4.3% and 13% in 2 and 3 agent scenarios.
- SR **decreases** by 4.7% and 17% in 4 and 5 agent scenarios, **but** LoS **decreases** by 30% and 75%.

Num. Agents	Model	Success Rate (SR)	LoS Rate	Timeout
2	SoRTS	93%	4.0%	1.0%
	SafeAir (exp. 3)	97%	1.0%	0.5%
	SafeAir (exp. 4)	-	-	-
3	SoRTS	86%	13%	0.3%
	SafeAir (exp. 3)	97%	13%	1.0%
	SafeAir (exp. 4)	-	-	-
4	SoRTS	86%	13%	0.8%
	SafeAir (exp. 3)	82%	9.0%	8.3%
	SafeAir (exp. 4)	-	-	-
5	SoRTS	82%	16%	1.6%
	SafeAir (exp. 3)	66%	4.4%	29%
	SafeAir (exp. 4)	-	-	-

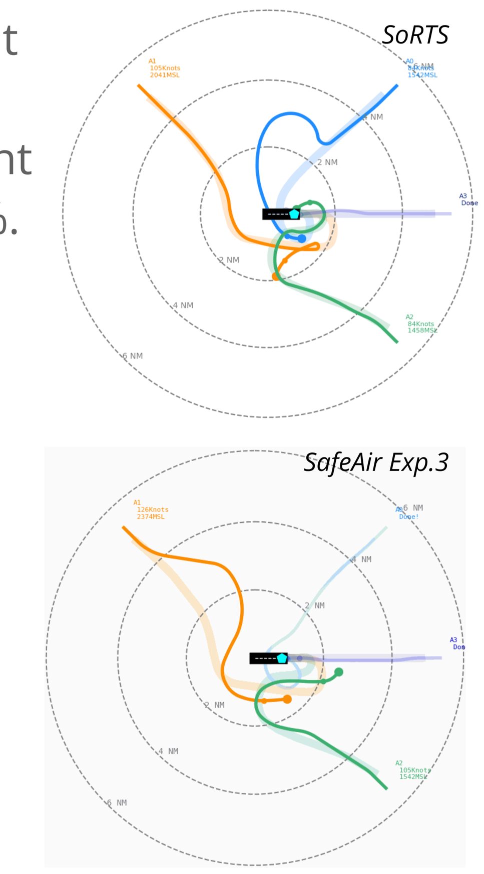

Figure 6. Navigation Results. Snapshot taken at the same timestep

Discussion and Future Work

- We aim to explore different remediation approaches to further improve the robustness of the navigation algorithm and address the performance drops in trajectory prediction.

Acknowledgement

This work was possible thanks to the constant mentoring of PhD student Ingrid Navarro. Thanks to Prof. Jean Oh and the BIG Lab for their support. Thanks to everyone who makes the RISS program possible, especially Rachel Burcin, Morgan Grimm, and John Dolan.

References

[1] Navarro, Ingrid, et al. "SoRTS: Learned Tree Search for Long Horizon Social Robot Navigation." IEEE Robotics and Automation Letters (2024).
[2] Stoler, Benjamin, Navarro, Ingrid, et al. "SafeShift: Safety-informed distribution shifts for robust trajectory prediction in autonomous driving." 2024 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2024.
[3] Patrikar, Jay, et al. "Predicting like a pilot: Dataset and method to predict socially-aware aircraft trajectories in non-towered terminal airspace." 2022 international conference on robotics and automation (icra). IEEE, 2022.
[4] Navarro, Ingrid, et al. "AmeliaTF: A Large Model and Dataset for Airport Surface Movement Forecasting." AIAA Aviation Forum and ASCEND. 2024.