1 Appendix

1.1 Hyperparameters

Table 1: Hyperparameters and Training Settings for Visual Navigation

Parameter	**Value**
Actor Learning Rate	1e-5
Actor Update Interval	1
Critic Learning Rate	1e-4
Cost Critic Learning Rate	1e-4
Distance Critic Bins	20
Cost Critic Bins	40
Targets Update Interval	5
Polyak Update Coefficient	0.05
Initial Lagrange Multiplier	0
Lagrange Learning Rate	0.035
Optimizer	Adam
Visual Inputs Dimensions	(4, 32, 32, 4)
Replay Buffer Size	100000
Batch Size	64
Initial Collect Steps	1000
Training Iterations	600000
Neural Network Architecture	Conv(16, 8, 4) + Conv(32, 4, 4) + FC(256)
Maximum Episode Steps	20

1.2 Limitations

One key limitation of our approach is the absence of strict risk-bounded guarantees on the cumulative risk incurred by all agents. In real-world scenarios, an ideal system would accept user input regarding risk averseness and automatically adjust its behavior accordingly. Although there are existing methods for multi-objective optimization, to our knowledge, none guarantee bounded execution risk while also maximizing reward. Additionally, the constrained low-level policy was trained on a single agent in a fixed environment, so its applicability in dynamic settings has not been empirically validated—even though SoRB has demonstrated effectiveness in diverse environments. In fast-changing, dynamic settings, the safer behaviors provided by our approach may be less effective.

1.3 Real-World Connections

To validate the approach in real-world settings, it is essential to ensure that agents adhere to the timings of each waypoint determined by the high-level CBS search in order to avoid collisions with other agents. If deviations from the nominal trajectories occur,

low-level agents should utilize individual collision avoidance strategies by communicating their positions with one another, ensuring that when agents come too close, one can yield until the other has passed. Moreover, for smoother deployment, the operating height of the agents must be considered, as it defines the obstacle boundaries relevant to the safety function. Incorporating the height (z-coordinate) could lead to more natural plans, such as enabling drones to avoid obstacles by flying over them.