Reinforcement Learning Based Path Planning for Injury-Aware Mass Casualty Monitoring

Garrick Jensen[1], Robert Sato[1], Riley Smith[1], Will Gushurst[1]

[1]Areté Associates, Northridge, CA

Areté

DISCOVER. DEVELOP. DELIVER.

Objectives

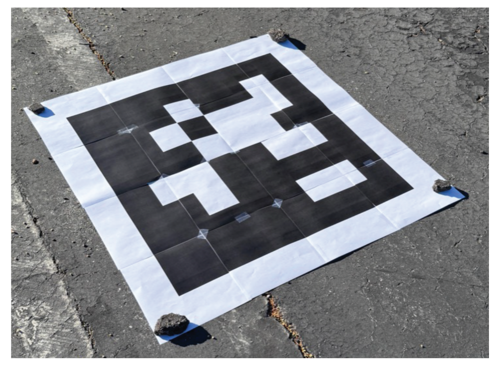
- Demonstrate drones can be autonomously directed by Reinforcement Learning (RL) agents to revisit patients with various wound severity
- Better understand how RL can flexibly handle dynamic and unknown settings

Background

- Remote vitals monitoring via airborne sensors empowers state-of-the-art mass casualty medical response
- Drones now appear in nearly every step of the care continuum, including patient identification, vitals monitoring, and up-to-date status reporting
- However, manually operating drones requires trained personnel and scales poorly
- Routing drones between patients for optimal injury-aware care is complex, and traditional algorithms are rigid and expensive to recompute

Hardware

The ModalAI Starling 2 Max Drone: Single camera mounted on drone



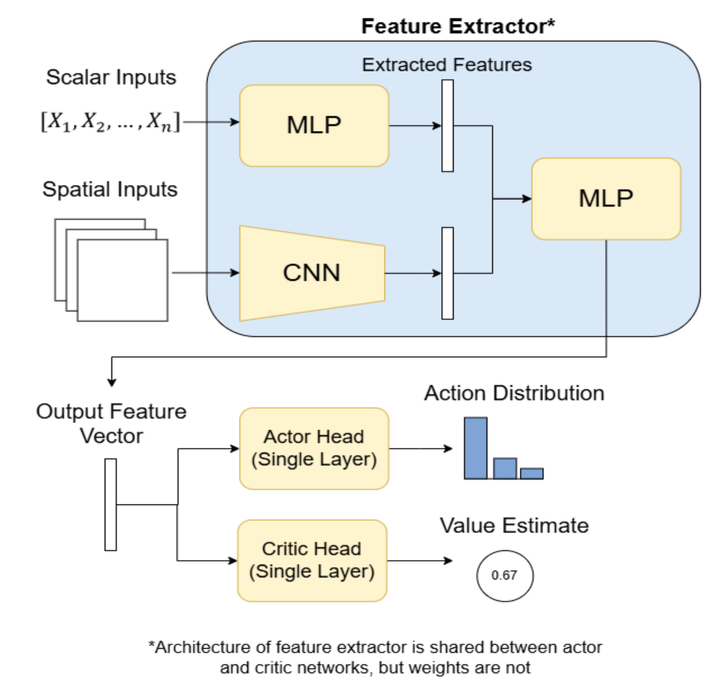
April Tag: Used in place of casualty dummies, each tag encodes a different injury severity value

Ground Station: Drones sends images and GPS back to station and RL agents send control commands

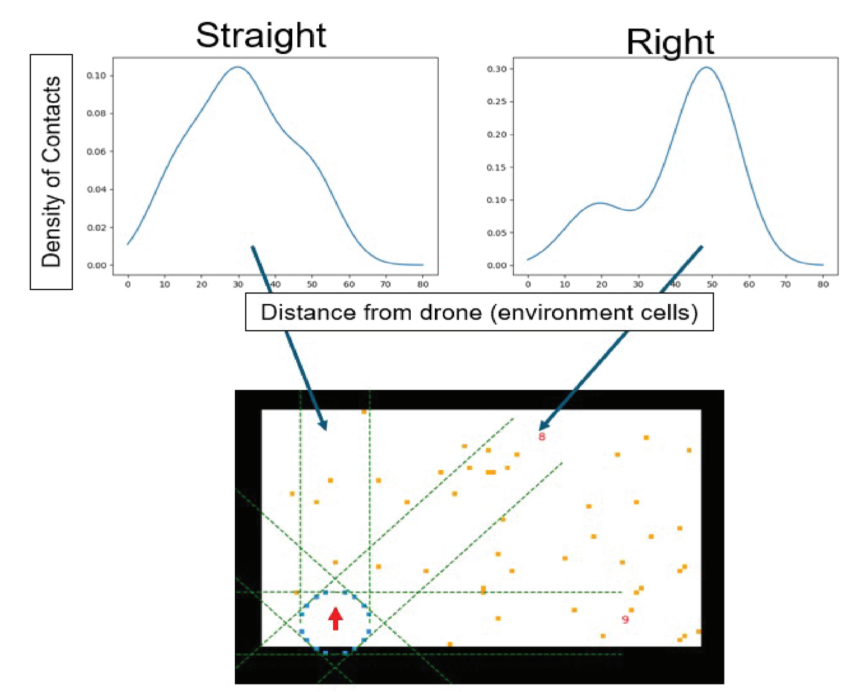
Methods

Scalar and Spatial Observations and Inputs Outputs:

- The agent model is fed both scalar and spatial (2D) values describing the setting and its current and previous states
- The model then outputs an action distribution that includes how to turn and when to observe or hover
- As well as a critic value for actor-critic training



RL Agent Architecture for Scalar and Spatial In



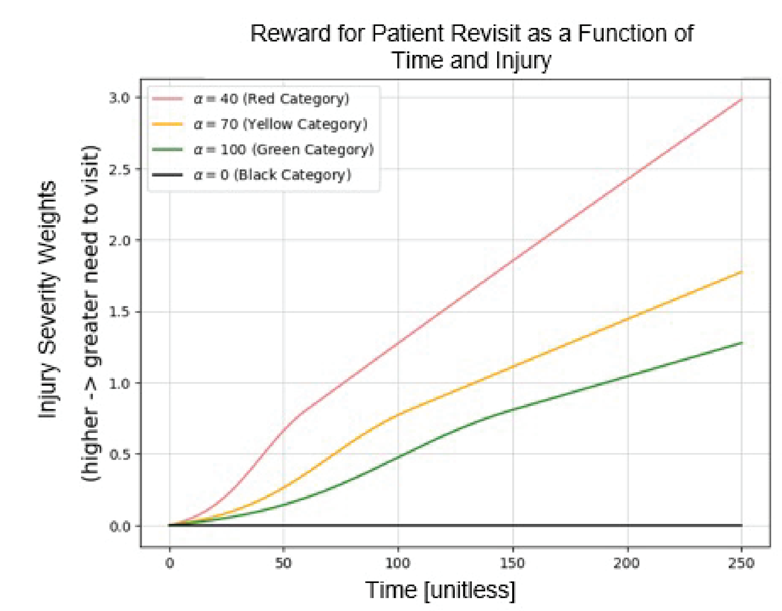
RL Agent Architecture for Scalar and Spatial Input

Contact Density Map:

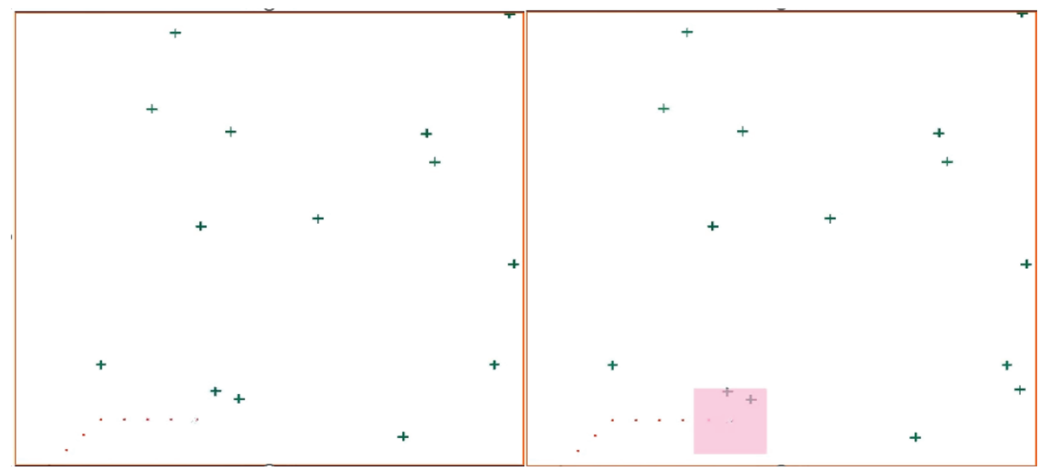
- One of the most informative spatial observations relays the location of patients in the scene and how much reward may be gained from visiting them
- The map notes how much to turn and how long until the agent reaches awaiting patients
- The angles are aligned with the turn angles in the action distribution
- The density distributions are appended together as rows to create a spatial input

Injury Severity Weighting Over Time:

- The need for care for each patient is set as an increasing function of time since last visit
- The rate at which patient priority increases is a surrogate for injury severity and is related to RL agent reward
- The weighting for each injury severity increases over time since the previous visit
- Red injuries are meant to represent patients with head wounds, amputation, or excessive bleeding
- Green represents patients that are moving and alert
- Black lines represent deceased patients that are not served by revisit
- The most severe injuries are weighted the highest and grow the fastest



Difference in how each injury severity score is weighted



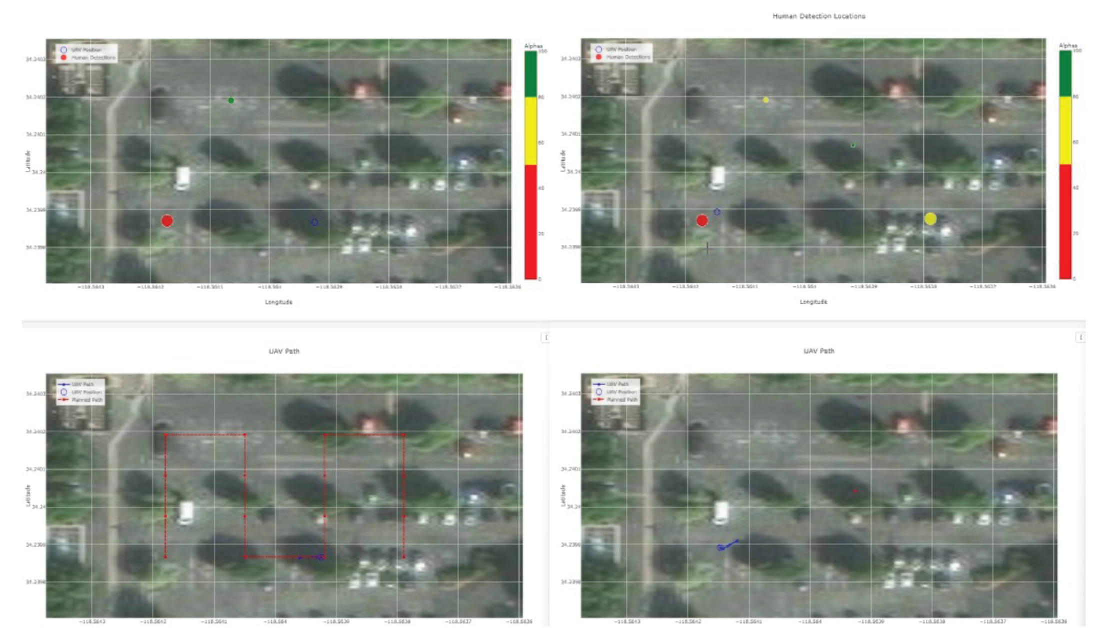
RL Agent Choosing to Use its Observe Action to Reset the Last Visited Time. It's sensor is approximated as a square as seen in Red. The left image shows before the observation occurs, the right image is during observation.

The Observe/Hover Action:

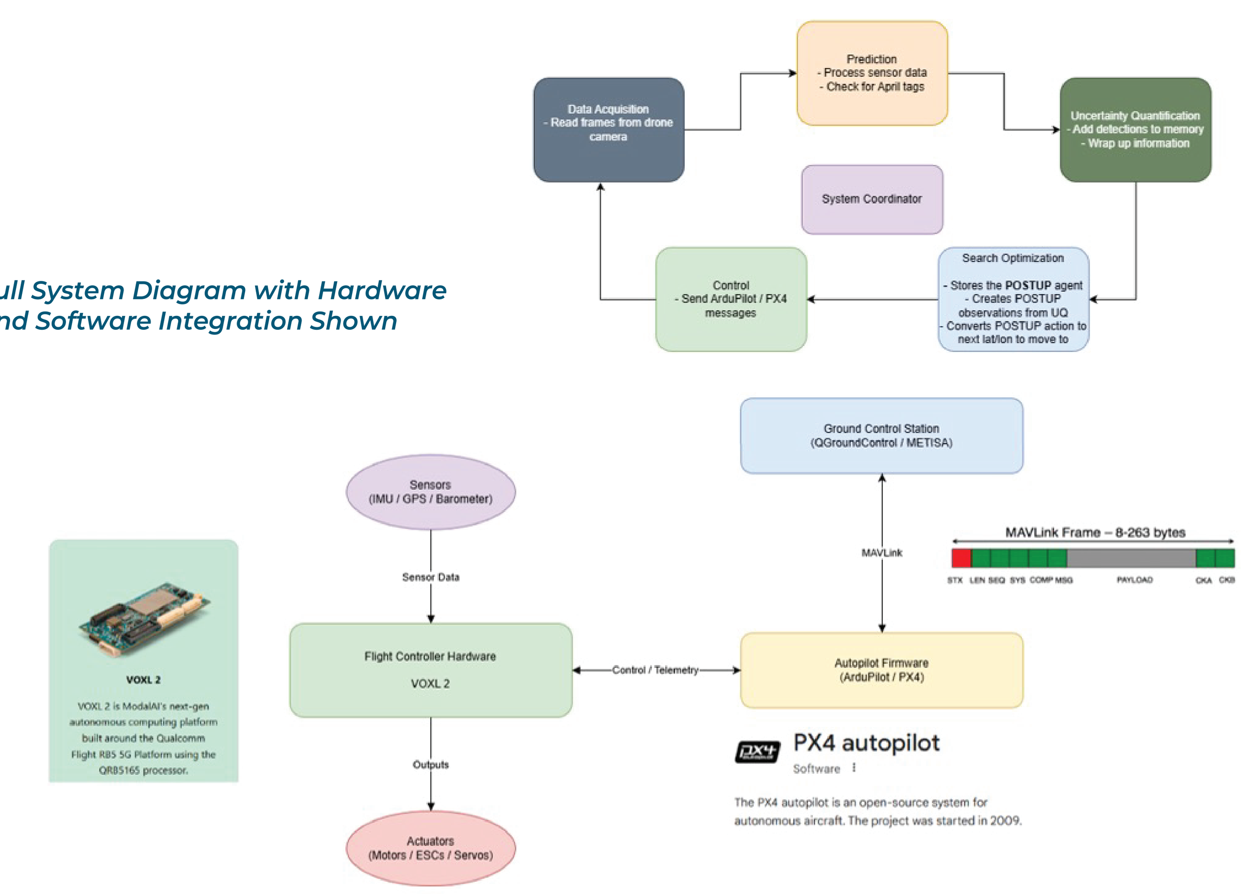
- The agent also receives a scalar value as to how much reward is gained a given timestep if the agent chooses to observe
- The reward gained depends on what patients are in range of the sensor, and thus is sensor specific
- When the agent chooses to observe, their motion is paused temporarily, the timer for each patient is restarted
- There is a small refractory period for observations to encourage visiting other patients

Results

- The agent routed the drone efficiently between casualties and adapted with simulated injury evolution
- The agent learned to visit nearby tags sequentially, to prioritize those with high injury severity, and to never re-visit casualties that had died
- The order at which casualties were revisited was a holistic mix between distance to traverse, injury severity, and how long since visit



Left Column is the Drone Doing a Precomputer Search of the space, identifying patients as colored coded by severity. Right Column is after the search is complete and the Agent is directing the drone



Full System Diagram with Hardware and Software Integration Shown

Conclusion

We have shown that RL agents can be trained to efficiently route between patients based on distance from drone, injury severity, and time since last visited, creating flexible paths that consider many changing factors about the environment.

For more information please go to www.Arete.com or contact Garrick Jensen gjensen@arete.com