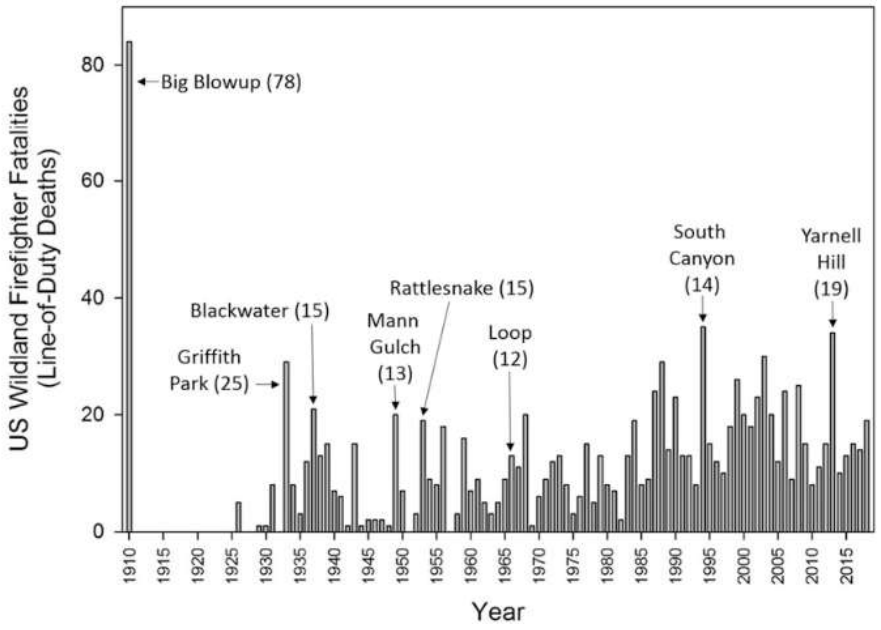
Benefit

- Reducing on-duty-personnel fatalities in forest fire disasters.



US Wildlands Firefighter fatalities [1]

Related Works

- What's Missing?
  - In [2] (2018):
    - Fire Fighting scenario – Stochastic Fire Model.
    - Wind factor is excluded – Heterogenous agents are not explored.
  - In [3] (2021):
    - Fire spread monitoring scenario – Stochastic fire model - Wind factor is included.
    - No Firefighting – Heterogenous agents are not explored.

We aim to:

- Add an extra layer of realism to a forest firefighting scenario.
- Explore heterogenous agents.
- Explore an advanced state-of-the-art MARL algorithm.

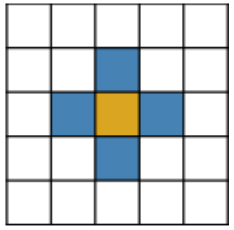
Components

- Environment
- Agents
- Reward System
- Algorithm

Environment

- 50x50 2D grid = 2499 trees and 1 refill station.

- Trees’ diagonal neighbors are included in the neighbor set.



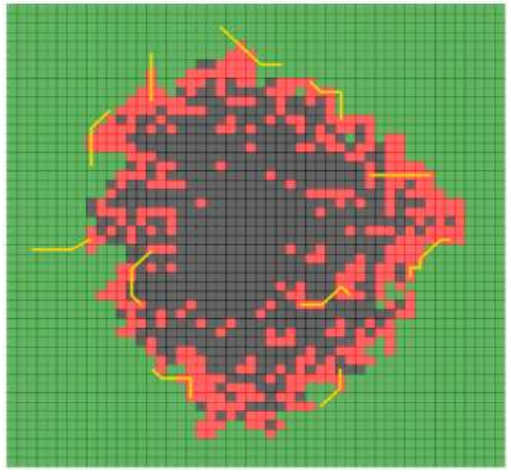
Neighbor set in [2]

Neighbor set in our work

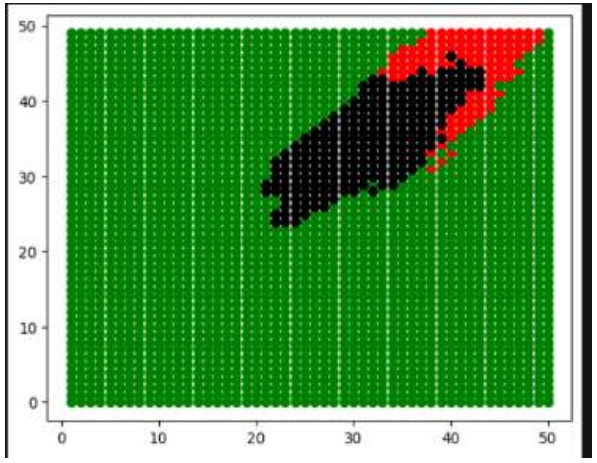
Environment

- Fire starts at the center point of the grid.

- Stochastic fire spread model with wind direction at angle 50 degrees.



Fire spread in [2]



Fire spread in our work

Environment

- Wind speed added to control the spread of the fire.
  - How many simulation runs before agents start to act.

- ε for a tree to be completely suppressed.

Agents

- Multiple heterogenous agents.

- β for each class of agents to dictate their impact on the fire.

Agents

- Autonomous Ground Vehicles (AGVs):
  - 5x5 view space
  - Step distance = 1
  - Diagonal moves
- Unmanned Aerial Vehicles (UAVs):
  - 11x11 view space
  - Step distance = 10
  - Retardant capacity = 10
- Common capabilities:
  - Memory size of 15 elements.

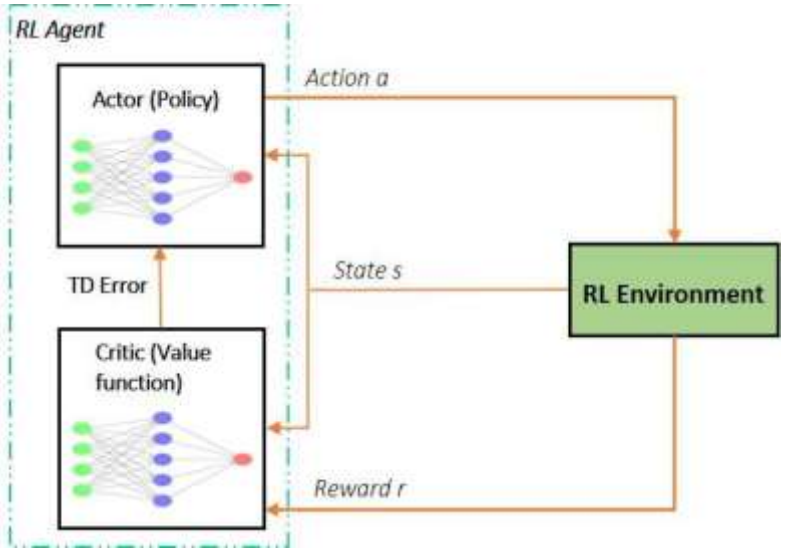
Reward System

- Penalized when:
  - A tree starts burning.
  - A tree completely burns.

- Rewarded when:
  - Observing a new burning tree.
  - Colliding with the coordinates of a burning tree.
  - Suppressing a burning tree.
  - UAVs ONLY: Refilling from the station if retardant capacity = 0

Algorithm

- Multi-Agent Deep Deterministic Policy Gradient (MADDPG).
  - Extension of the Actor-Critic architecture.



General Actor-Critic Framework [4]

Algorithm

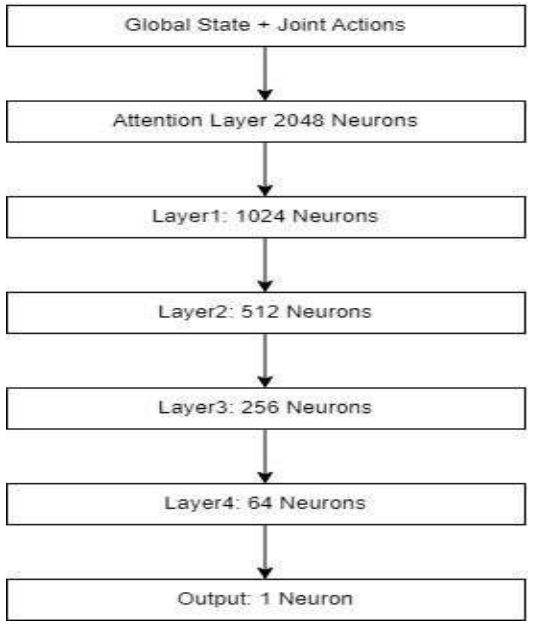
- Multi-Agent Deep Deterministic Policy Gradient (MADDPG).
  - Update rule for the actor network:
    - Obj = Q-Value1 * Action1 + Q-Value2 * Action2 + …. + Q-ValueN * ActionN
    - Obj = $\sum_{t=0}^{T} \mathrm{Q}(\mathrm{st}\ ,\ \mathrm{a_t}:\ \theta) * \mathrm{A}(\mathrm{s_t}:\ \theta\mathrm{a})$
  - Update rule for the critic network:
    - Gradient step on the MSE loss function.
    - Loss = (Target – Predicted)^2
    - Predicted = Q(current_state, current_action)
    - Target = R + ϒ*Q(next_state, next_action)

Algorithm

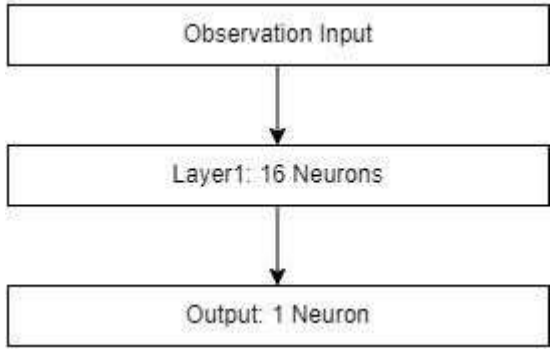
- Multi-Agent Deep Deterministic Policy Gradient (MADDPG).
  - Addition of two extra neural networks:
    - Target Actor Network (A').
    - Target Critic Network (Q').
    - Purpose: Training Stability.
  - Update rule for the target networks:
    - A' = (Tau*A) + (1-Tau)* A'
    - Q' = (Tau*Q) + (1-Tau)* Q'
  - Addition of memory buffer to store and sample experiences for training.

Algorithm

- Multi-Agent Deep Deterministic Policy Gradient (MADDPG).



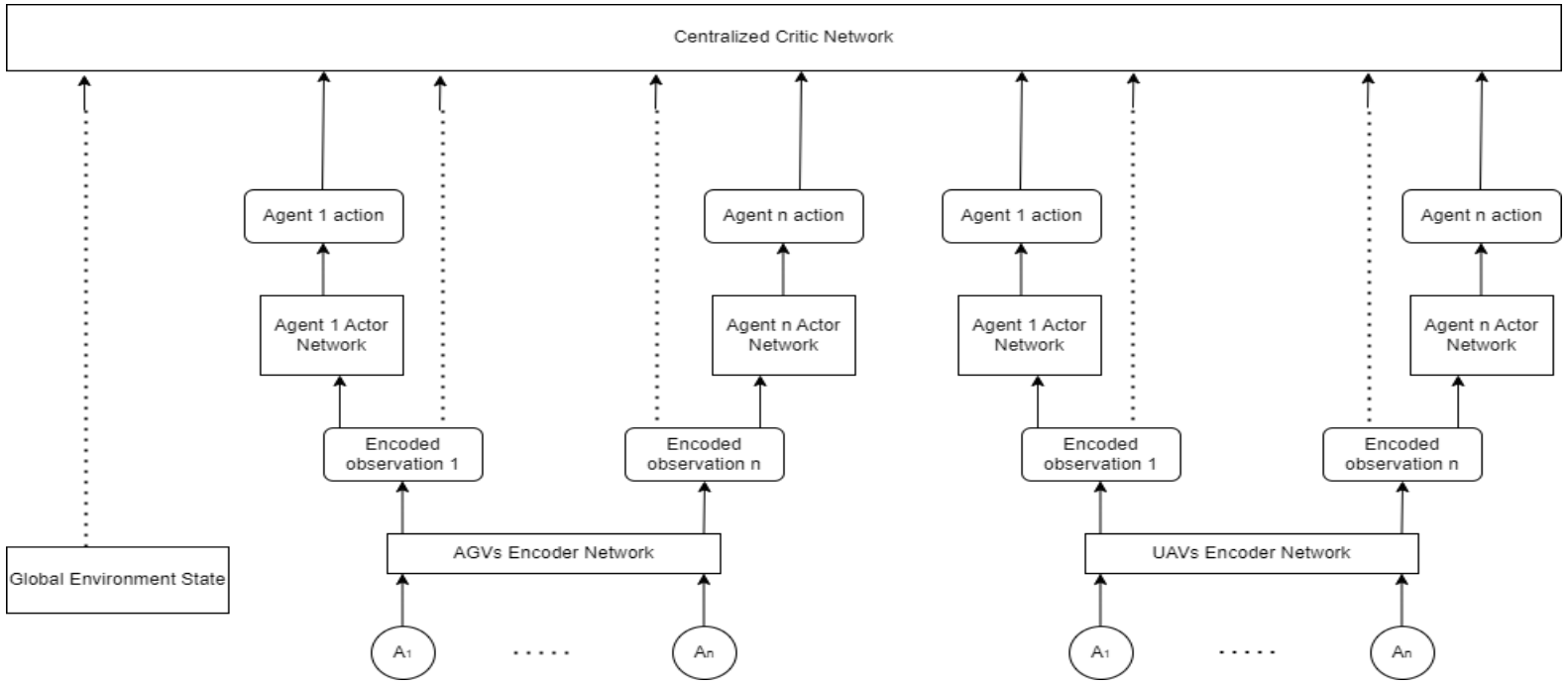
Our Attention-based Critic Network



Our Actor Network in heterogenous scenarios

Algorithm

- Multi-Agent Deep Deterministic Policy Gradient (MADDPG).



Our MADDPG Framework

Experimenting Scenarios

#	Agents	Wind Speed	Centralized Critic Architecture
Scenario 1	Heterogenous Agents	1	Standard Critic
Scenario 2	Homogenous Agents	1	Standard Critic
Scenario 3	Heterogenous Agents	2	Standard Critic
Scenario 4	Homogenous Agents	2	Standard Critic
Scenario 5	Heterogenous Agents	1	Attention-Based Critic
Scenario 6	Homogenous Agents	1	Attention-Based Critic
Scenario 7	Heterogenous Agents	2	Attention-Based Critic
Scenario 8	Homogenous Agents	2	Attention-Based Critic

- Passing the Global State to the Centralized Critic.
- Passing only the local observations to the Centralized Critic as in [5].

Fixed Training Hyperparameters

- Learning Rate = 0.01
- Tau = 0.05
- Batch Size = 64
- Gamma = 0.95
- Encoded Dimensions = 32

- Homogenous Scenarios: 12 AGVs
- Heterogenous Scenarios: 8 AGVs + 4 UAVs

Evaluation Metrics

- Total accumulated Rewards.
- Total number of trees suppressed.
- AGVs vs UAVs performance.

Training Results (Global State Passing)

	Standard Critic				Attention-based Critic			
Aspect	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 4**	**Scenario 5**	**Scenario 6**	**Scenario 7**	**Scenario 8**
Minimum Normalized Rewards	0.1874	**1**	0.79685	0.15594	0	0.55365	0.5565	0.4621
# of trees suppressed	Max: 39 Min: 0	Max: 49 Min: 0	Max: 51 Min: 0	Max: 39 Min: 0	Max: 42 Min: 0	**Max: 56** Min: 0	Max: 43 Min: 0	Max: 39 Min: 0
# of trees suppressed by AGVs	Max: 37 Min: 0	Max: 49 Min: 0	Max: 50 Min: 0	Max: 39 Min: 0	Max: 42 Min: 0	**Max: 56** Min: 0	Max: 43 Min: 0	Max: 39 Min: 0
# of trees suppressed by UAVs	Max: 6 Min: 0	-	Max: 3 Min: 0	-	**Max: 9** Min: 0	-	Max: 1 Min: 0	-

Training Results (Local Observation Passing)

	Standard Critic				Attention-based Critic			
Aspect	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 4**	**Scenario 5**	**Scenario 6**	**Scenario 7**	**Scenario 8**
Maximum Normalized Rewards	0.9186	**1**	0	0.5051	0.38983	0.738983	0.58983	0.522
# of trees suppressed	Max: 56 Min: 0	**Max: 67** Min: 0	Max: 28 Min: 0	Max: 45 Min: 0	Max: 36 Min: 0	Max: 49 Min: 0	Max: 43 Min: 0	Max: 47 Min: 0
# of trees suppressed by AGVs	Max: 54 Min: 0	**Max: 67** Min: 0	Max: 28 Min: 0	Max: 45 Min: 0	Max: 36 Min: 0	Max: 49 Min: 0	Max: 43 Min: 0	Max: 47 Min: 0
# of trees suppressed by UAVs	**Max: 9** Min: 0		Max: 2 Min: 0		Max: 2 Min: 0		Max: 3 Min: 0	