**Deep Reinforcement Learning based Path Planning with Dynamic Trust Region Optimization for Automotive Application**

**1.Introduction**

recent application of mobile robot

Framework of the proposed multi-robot systems

**2.Proposed methodology of the path planning algorithm**

obstacle avoidance

Energy efficiency

Minimum travel time

path accuracy and adaptivity environment

Deep reinforcement learning

Dynamic improvement PPO-CGA evolution strategy

Dynamic improvement TRPO-CGA evaluation strategy

DITRPO-CGA optimization problems

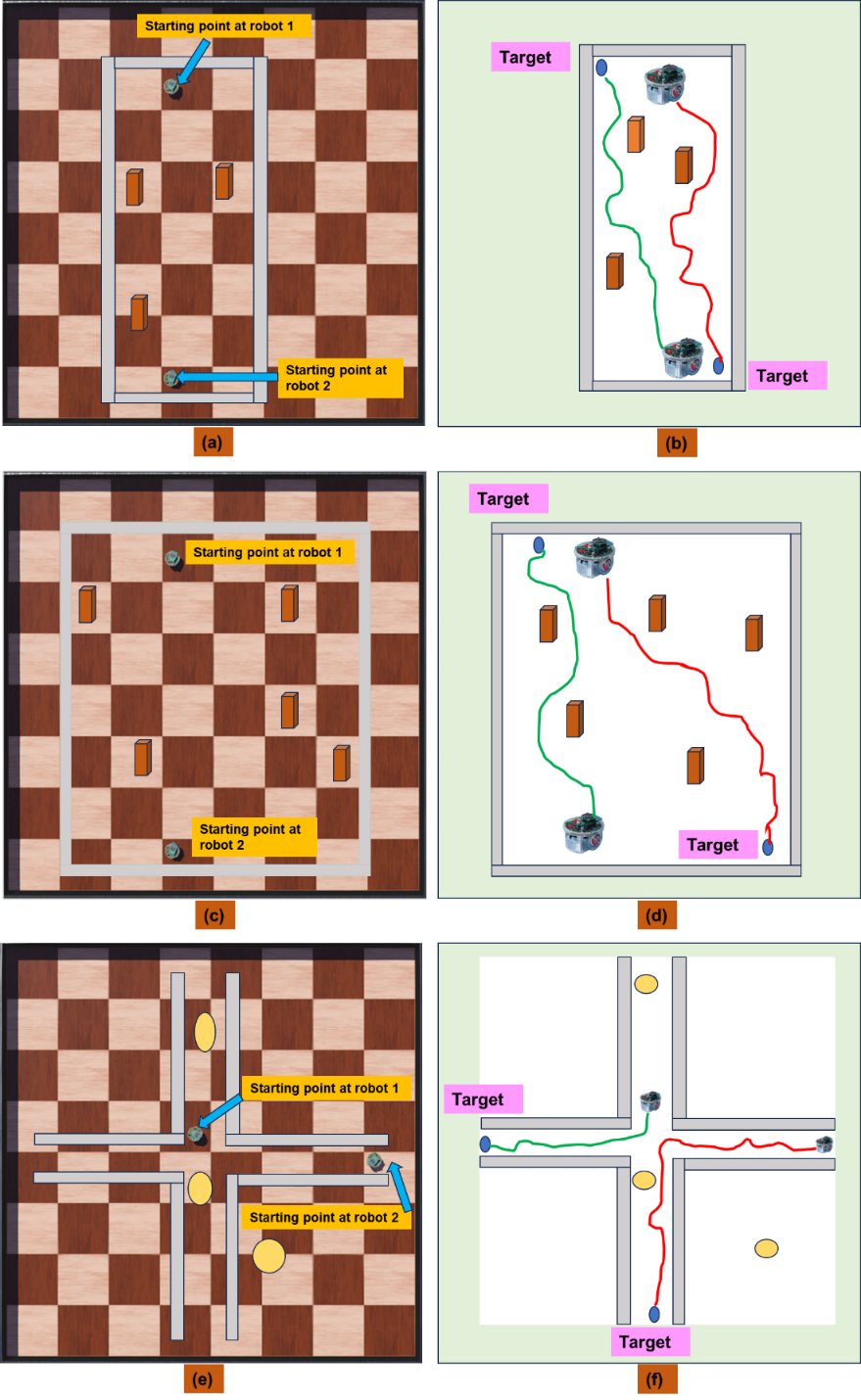
constration of the reward function and the proposed algorithm

**3. Methods**

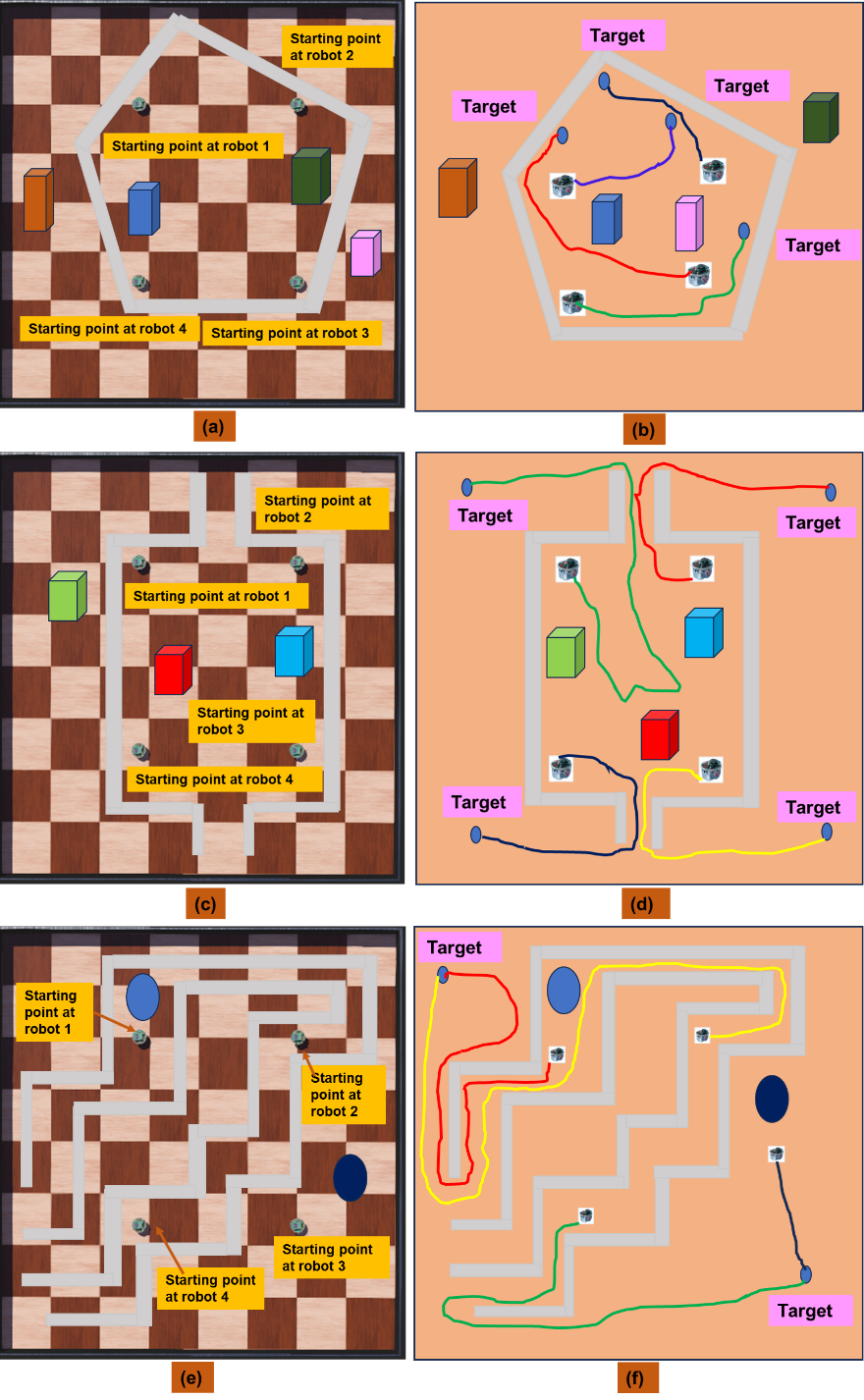
**Pseudocode for proposed algorithm (DITRPO-CGA)**

|  |
| --- |
| **Pseudocode for proposed algorithm (DITRPO-CGA)**  Step:1 Initialize the actor and critic  Actor: The policy π (S) is initialized parameters with θ  Critic: The value function V(S) is initialized with parameters ϕ  Step:2 Collect experience sequence  Generate a sequence of experiences over multiple steps of a states St, actions At, rewards Rt and the next states St+1. This sequence is collected over N steps per episode.  Step:3 Calculate temporal difference (TD) error  For each time step t in the episode  Calculate δk (TD error) using, δk = Rk + by V (St+1|ϕ) – V(St|ϕ)  This δk  represents the difference between the observed reward plus estimated future reward and the current estimated value.  Step:4 Calculate return Gt  For each time step t, compute the return Gt, which is the sum of discounted rewards and future values,  Gt =  Step:5 Compute advantage estimate Dt  Calculate Dt by subtraction the value V (St |ϕ) from the return Gt  Dt = Gt – V (St|ϕ)  Step:6 Batch update of critic  Every learning episode for select a mini-batch of size M from the trials  update the critic by minimizing the loss  Lcritic (ϕ) =  Step:7 Normalize advantage estimates  Normalize each Di in the batch using the mean and standard deviation of the batch values Ďi **=**  Step:8 Update actor, use the normalized advantage to update the actors parameters    Where, represents an entropy term, which encourages exploration by the actor  Step:9 Repeat steps 3 and 5 until the episode reaches a terminal state. |

**4. Results and discussion**

****



****