Anoop Rehman, Andrew Gordienko, Efe Tascioglu, Joaquin Arcilla, Jiajing Chen, Tatsuya Hirano, Seoyoung Kwon, Hannah Lila, Grace Liu, Carol Meng, Irwin Ngo, Liam Sheils, Aakash Vaithyanathan Matthew Guan, Matthew Tamura, Andrew Magnuson
Reach at *firstname.lastname@mail.utoronto.ca*

Introduction

The evolution of virtual creatures is a testament to the remarkable advancements achieved in artificial life and evolutionary computation. Inspired by the seminal work of Karl Sims in 1994, and making use of modern reinforcement learning algorithms, we present a novel approach aimed at generalizing coordination and collaboration across diverse sets of creatures in their ability to play 2v2 soccer.

Implementation Details

Evolving Baseline Creatures

Physical creatures were first created capable of moving towards a specific way-point by following the genetic algorithm proposed by Karl Sims for Virtual Creatures[1]. The primary goal of the algorithm is to maximize the reward policy of walking by iterating and refining the morphologies and genotypes of creatures through a series of key steps:

Selection - The top 20% of creatures with favorable traits for reaching a target goal were chosen for further mutation and training.

Pruning - Creatures that lack joints required for any form of movement or those that excessively stray away from the target get pruned during the early stages of training episodes.

Mutation - Selected creatures are mutated to introduce new behaviors and diversity. Mutations add/remove new joint segments, add/remove/modify the location of sensors from a segment, or modify the scale of the new segment spawned for the creature.

Each creature is also equipped with neurons and sensors. Following the Karl Sims paper, each creature can have: *joint angle* sensors, *collision* sensors, and *photosensors*. Each segment of the creature can also contain several neurons that aid in computation and propagating data through the neural network.

The *photosensor* neuron is a mandatory part of the root segment of any creature that senses the position of the targets, as it is used as a key observation to create the action space output.

The reward policy used penalized creatures for their root segment being further away from the target position. The formula for this is as follows:

$$R_t = \frac{1}{(\text{distance\_from\_target})^2}$$

After about 10 generations of training and mutations, we were able to achieve diverse creatures some of which excelled in accelerating towards the goal with less accuracy, while others prioritized staying near its target source using their extended limb segments.

Training Policy Network For Low Level Motor Function

Using Proximal Policy Optimization (PPO), we crafted a universal walking policy adaptable to various creature morphologies, inspired by Karl Sims's designs, for precise navigation. Utilizing the Mujoco physics library, creatures were abstracted and then detailed in XML, featuring rectangular torsos with one to four multi-segment legs. To refine the policy, nine varied creatures were simultaneously trained, employing curriculum learning to extend target distances as objectives were met, thereby collecting ample data for model training.

Training Policy Network To Play Soccer

In the third step, a policy was developed to coordinate players to play soccer. The field is split into a grid, where each player takes up a single grid spot.

Transformer Neural Networks learned two heat maps, a player heat map, and a ball heat map. These maps were learned through random policy roll out. Keeping track of what positions led to a goal by team one, or team two, represented by +1 and -1 respectively. The player heat map shows the neighboring grid positions that increase scoring chances from a given location. The ball heat map guides the optimal direction to move the ball from any coordinate for an advantage.

Upon creating the starting weights for these two networks, they were fine-tuned using Monte Carlo Tree Search (MCTS) and the AlphaZero methodology.

End-to-end pipeline

In our simulation, we first created virtual creatures and encoded them in XML for Mujoco compatibility. These entities were then integrated into a Mujoco soccer environment. Strategy and coordination were achieved by feeding a grid representation of the field into the AlphaZero algorithm, which generated target coordinates for each player. These coordinates were translated into actions using a trained general policy, enabling the simulation of a realistic soccer match within the Mujoco framework.

Results

The first phase revolved around getting creatures capable of ambulating in all directions. Each creature is equipped with a combination of joints allowing it to move in desired directions. More specifically, a segment of the creature could be a Fixed Joint or Hinge Joint with an axis of rotation in the x, y, or z-axis. The following figures show the morphologies of different creatures that evolved through the mutation process.

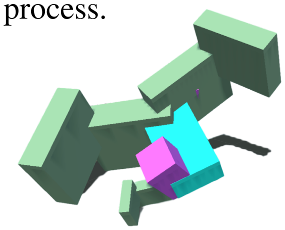

Fig. 1: Two Arm Rower Creature evolved to rotate and propel towards the target.

The figure illustrates the training of random creatures using reward beacons to develop a walking policy. The x-axis represents the episode, while the y-axis shows the distance from the spawn point for each of the nine creatures. Initially set at 1 unit, the spawn point moves further as it is sufficiently reached.

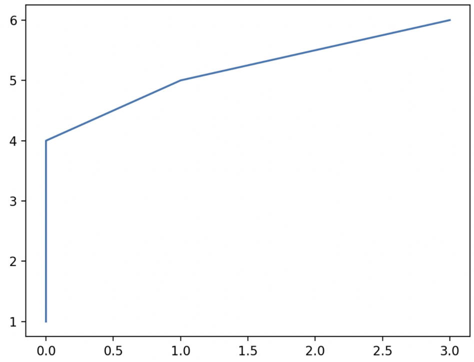

Fig. 2: Radius of beacons from creature.

The following figure shows an example of the heat maps trained before being fine-tuned with AlphaZero.

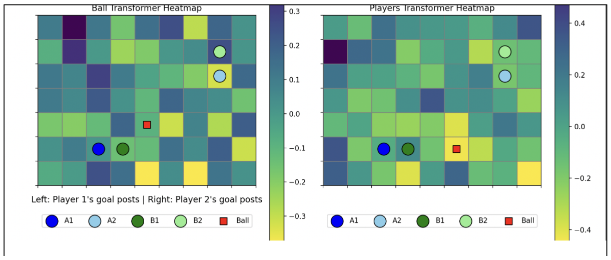

Fig. 3: Heat maps generated for players and balls.

Figure 4 shows the soccer environment with virtual creatures upon initialization.

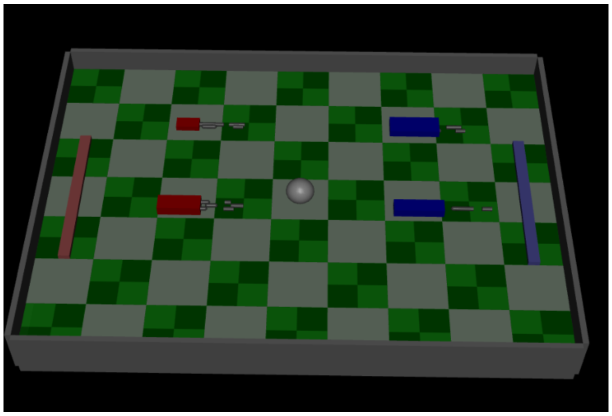

Fig. 4: Heat maps generated for players and balls.