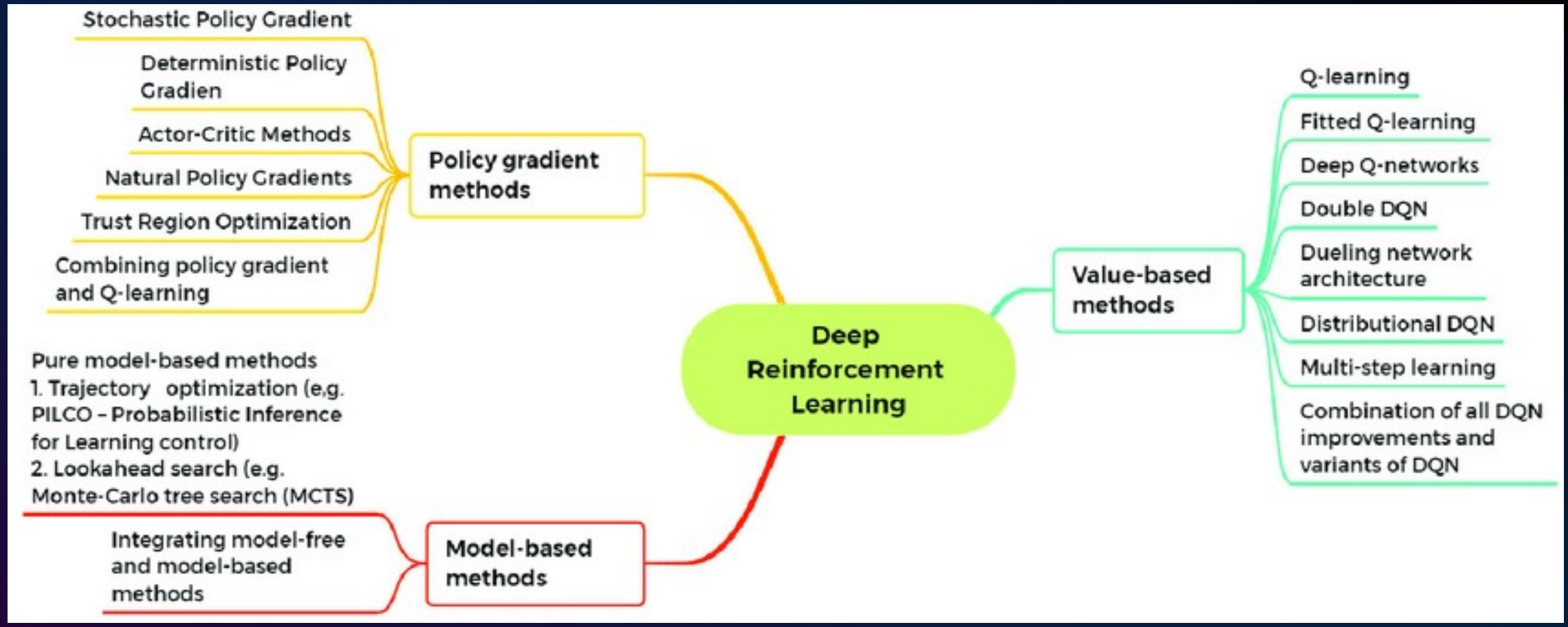
REINFORCEMENT LEARNING

Advanced RL Models

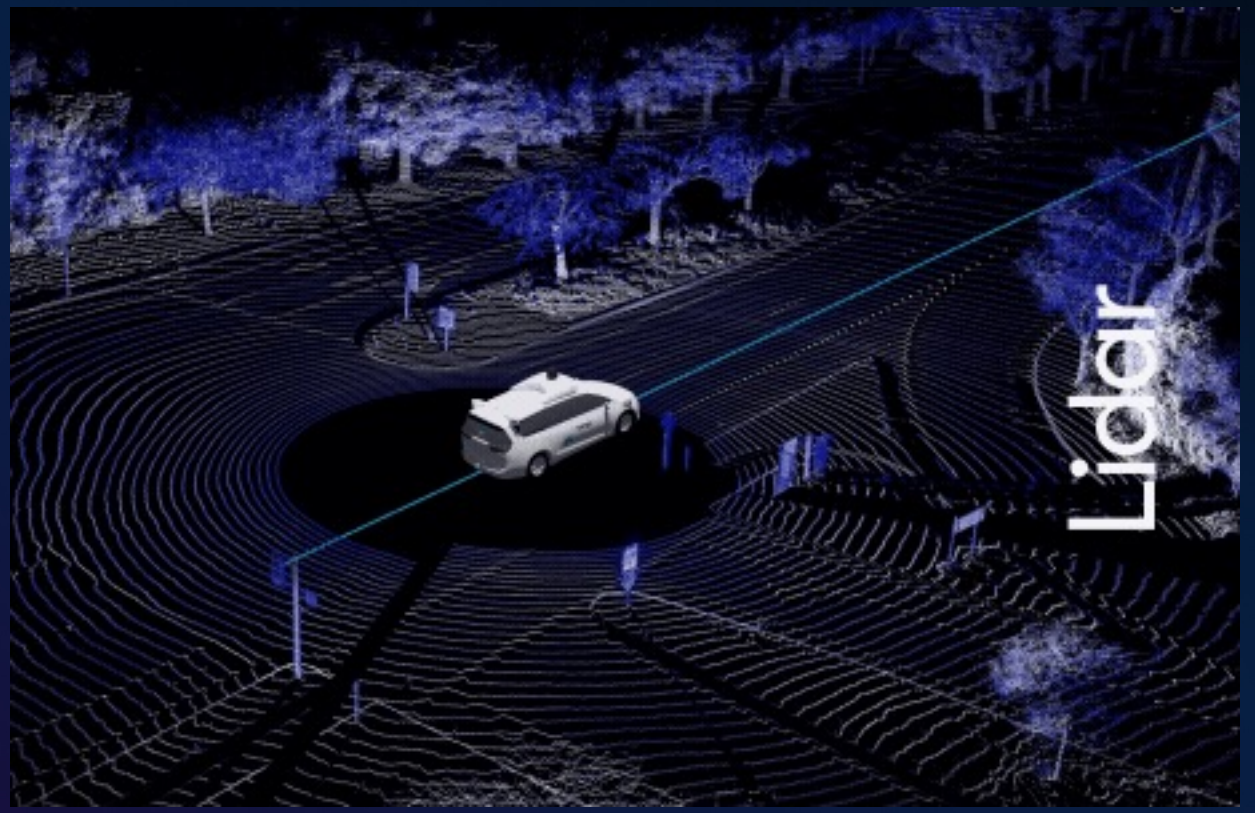
Irshad Chohan

Principal Solutions Architect
AWS India

RL – Learning algorithms



Waymo



AWS DeepRacer Student

Learn to apply machine learning (ML) skills in an autonomous racing league for a chance to win prizes and scholarships

Get started for free

2024 racing is underway

AWS DeepRacer Student helps high school and college-enrolled students around the globe develop their ML skills in a fun, hands-on autonomous racing league. Any student age 16 or older, can leverage 20 hours of ML educational material and 10 hours of monthly model training compute resources for free. Put your new found skills to the test for your chance to win prizes by becoming one of the top racers in the global student league.

Let’s go through the demo

Youtube

Let’s go through AWS Console

Tips for implementing RL algorithms

- Start simple. Implement basic Q-learning first, then incrementally increase complexity.
- Monitor training - track metrics like loss, rewards and evaluate model performance.
- Tune key hyperparameters like learning rate, discount factor and epsilon greedy exploration.
- Handle large state/action spaces with function approximation like neural networks.
- Use experience replay buffers to decorrelate training samples.
- Evaluate algorithms by comparing performance across multiple random seeds

Network architecture – Deep Learning

Network Architecture

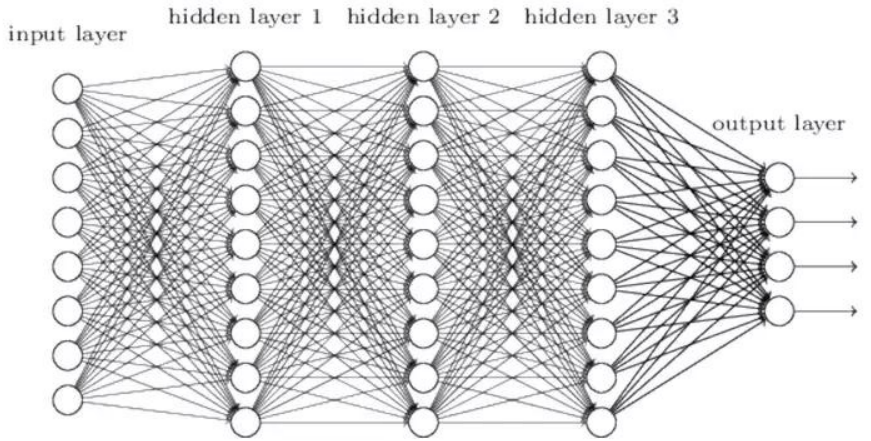
Neural architecture plays an important role in Deep Learning. In Actor-Critic algorithm, there are two kinds of network architecture:

- Separate policy network and critic network
- Two-head network

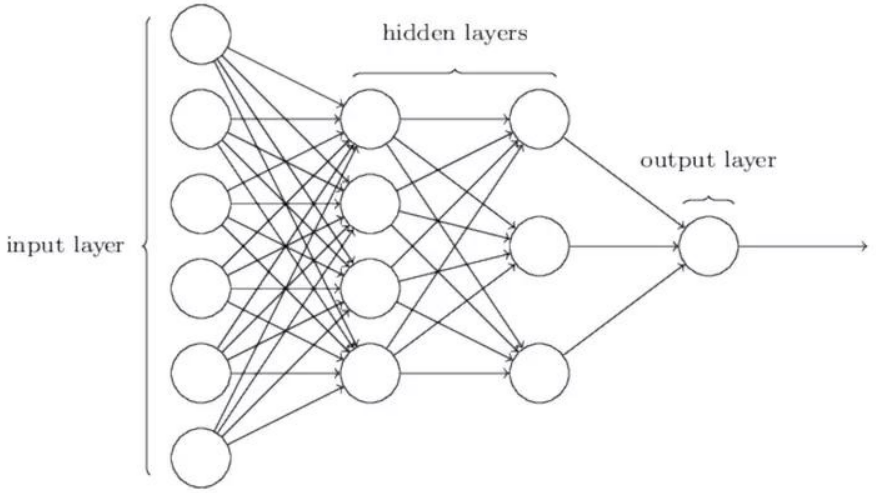
Network Architecture

Separate policy and critic network

- More parameters
- More stable in training



policy network

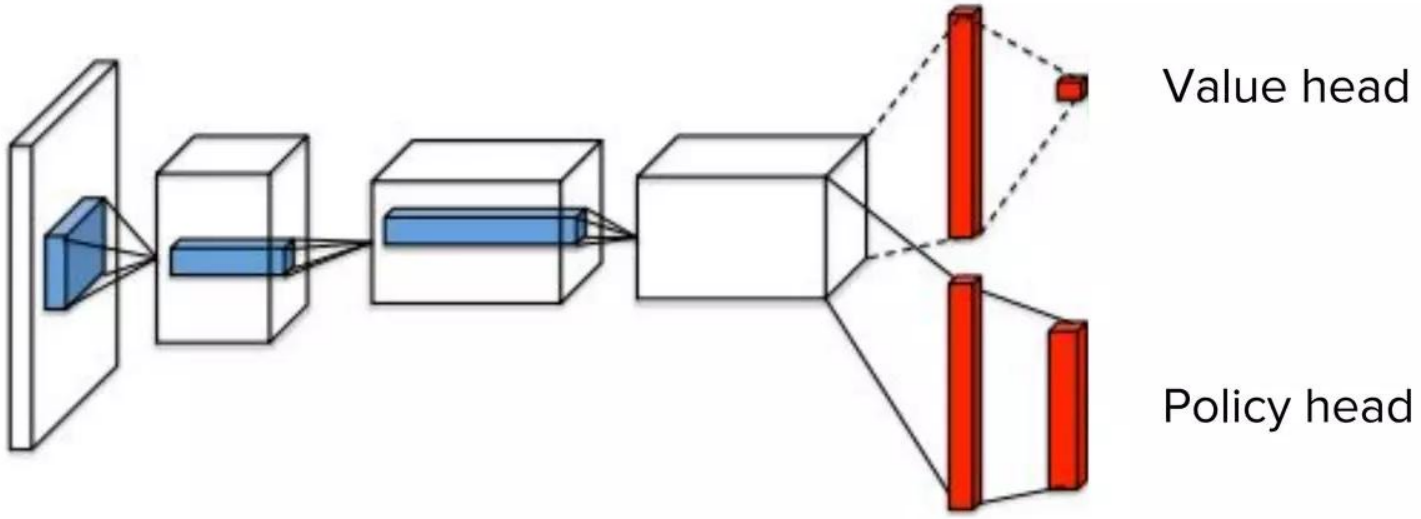


critic network

Network Architecture

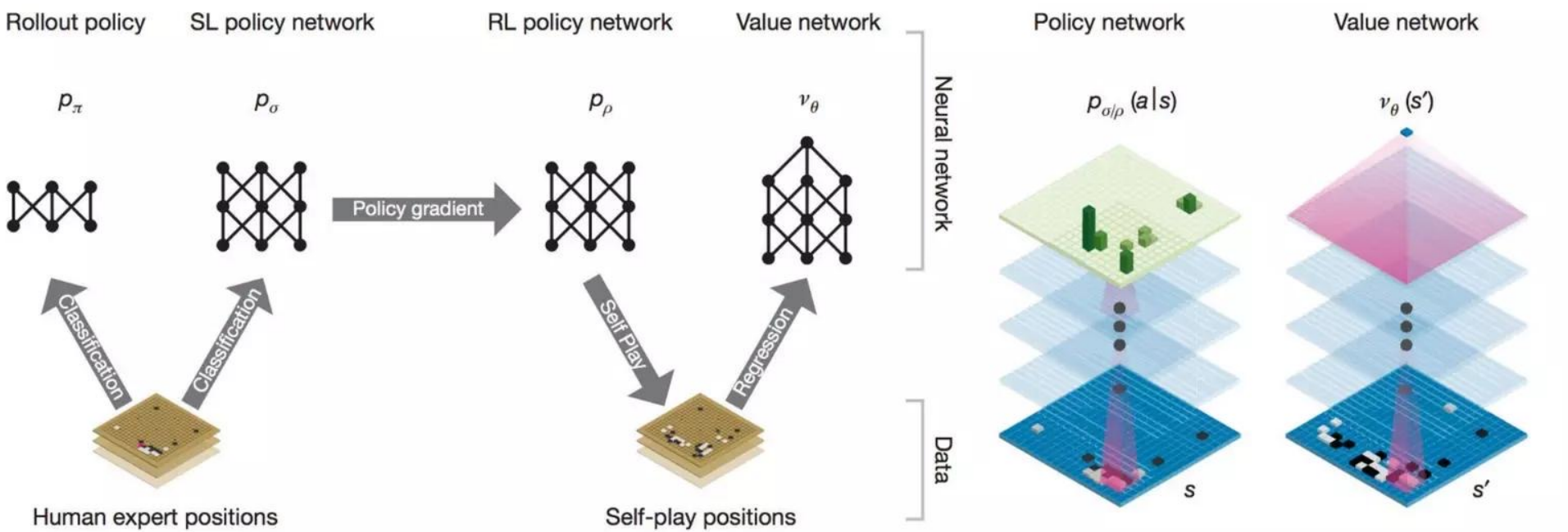
Two-head network

- Share features, less parameters
- Hard to find good coefficient to balance actor loss and critic loss



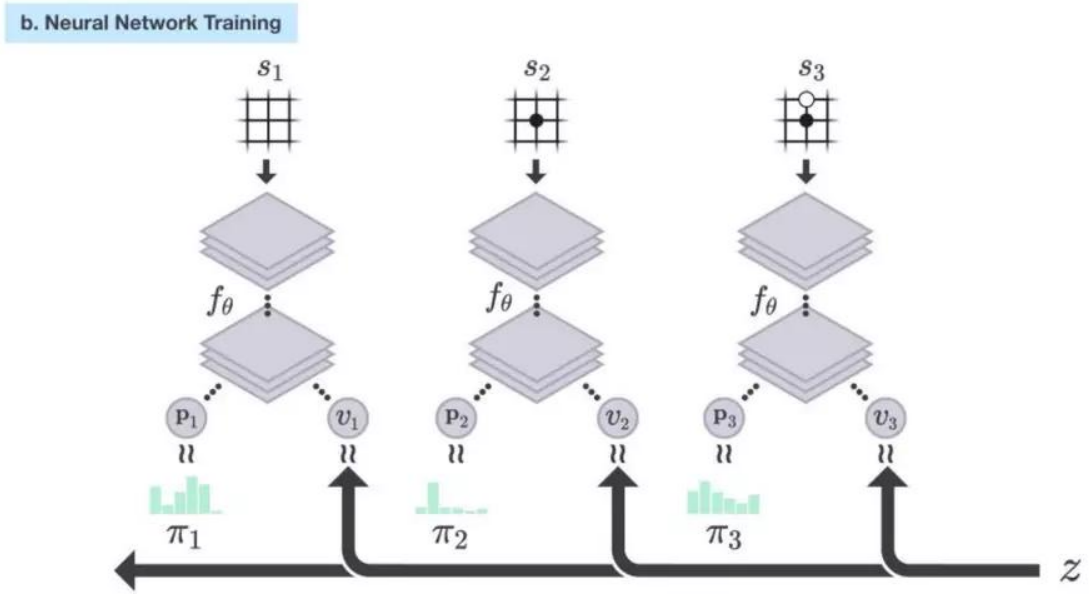
AlphaGO

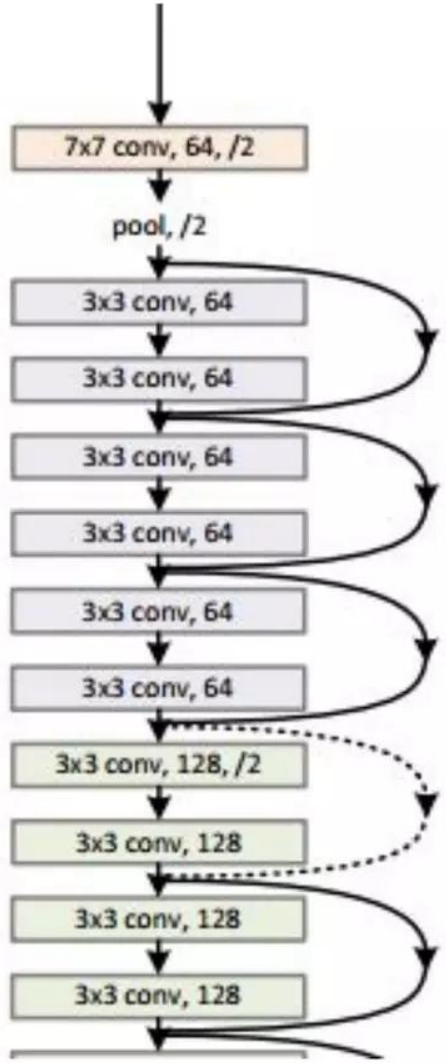
- MCTS, Actor-Critic algorithm
- Separate network architecture



AlphaGo Zero

- MCTS, Policy Iteration
- Shared ResNet





Correlation Issue

Online actor-critic algorithm:

1. Take action, get one-step experience (s, a, s', r)
2. Fit Value function

$$L(\phi) = \frac{1}{2}\sum_i \left\| \widehat{V}_\phi^\pi(s_i) - y_i \right\|^2$$

3. Evaluate advantage function

$$A^\pi(s_t, a_t) \approx r(a_t, s_t) + \gamma \widehat{V}_\phi^\pi(s_{t+1}) - \widehat{V}_\phi^\pi(s_t)$$

4. $$\nabla_\theta J(\theta) \approx \nabla_\theta \log \pi_\theta(a|s) \widehat{A}^\pi(s, a)$$

5. $$\theta \leftarrow \theta + \alpha \nabla_\theta J(\theta)$$

In online actor-critic algorithm, there still exist correlation problem.

Policy Gradient algorithm is **on-policy** algorithm so that we **cannot use replay buffer** to solve correlation problem.

Think about how to solve correlation problem in Actor-Critic.

Advance Actor-Critic algorithm

Currently, many state-of-the-art RL algorithms are developed on the basis of Actor-Critic algorithm:

- Asynchronous Advantage Actor-Critic (A3C)
- Synchronous Advantage Actor-Critic (A2C)
- Trust Region Policy Optimization (TRPO)
- Proximal Policy Optimization (PPO)
- Deep Deterministic Policy Gradient (DDPG)

Advance Actor-Critic algorithm

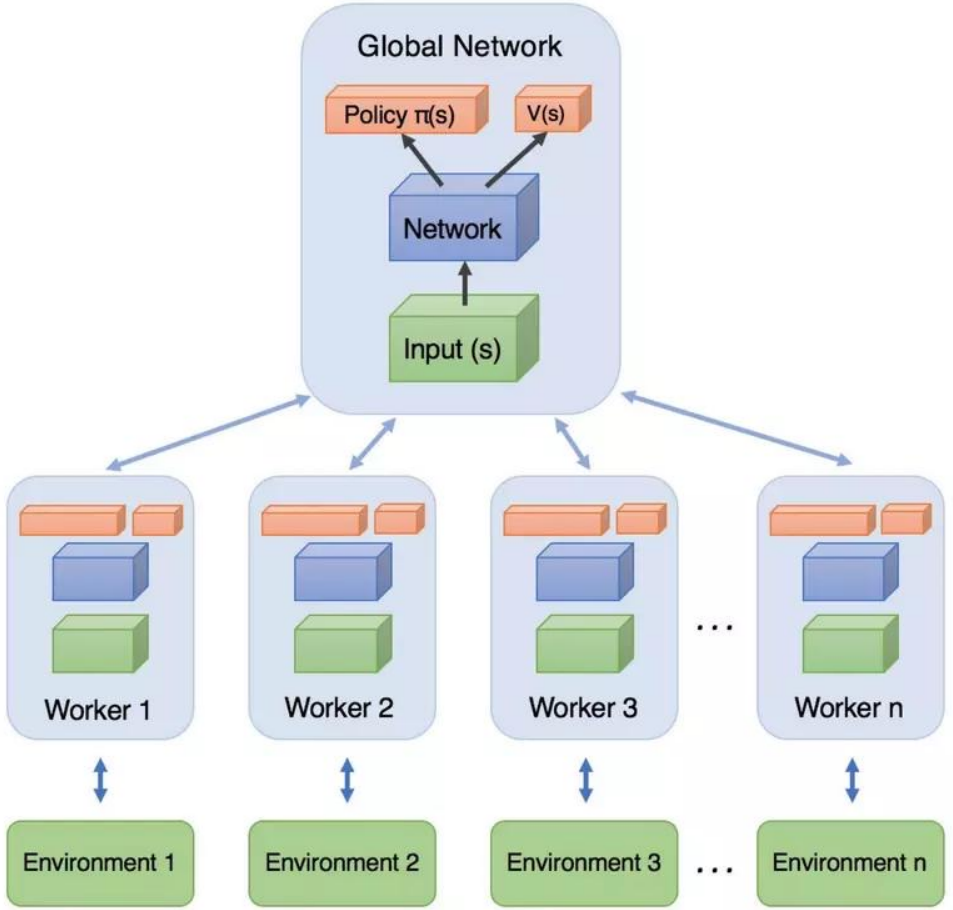
Asynchronous Advantage Actor-Critic (A3C)

V Mnih et al. (ICML 2016) proposed a parallel version of Actor-Critic algorithm which not only solves the correlation problem but also speeds up learning process. It's so-called **A3C. (Asynchronous Advantage Actor-Critic)**, A3C become the state-of-the-art baseline in 2016, 2017

Asynchronous Advantage Actor-Critic

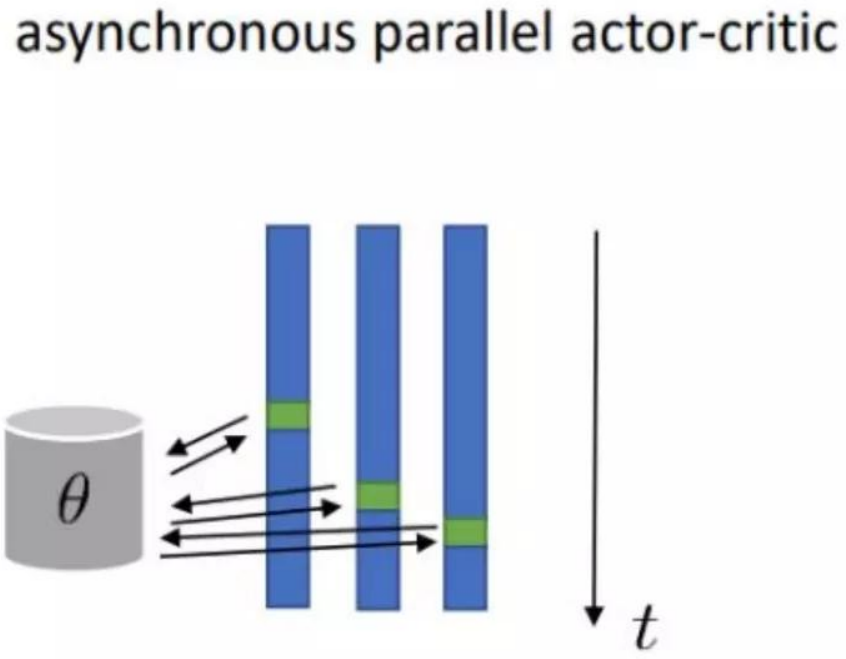
They use **multiple workers** to sample n-step experience. Each worker have **shared global network** and their **local network**.

Upon collecting enough experience, each worker computes the gradients of its local network, copies the gradients to shared global network and then does backpropagation to update global network asynchronously.



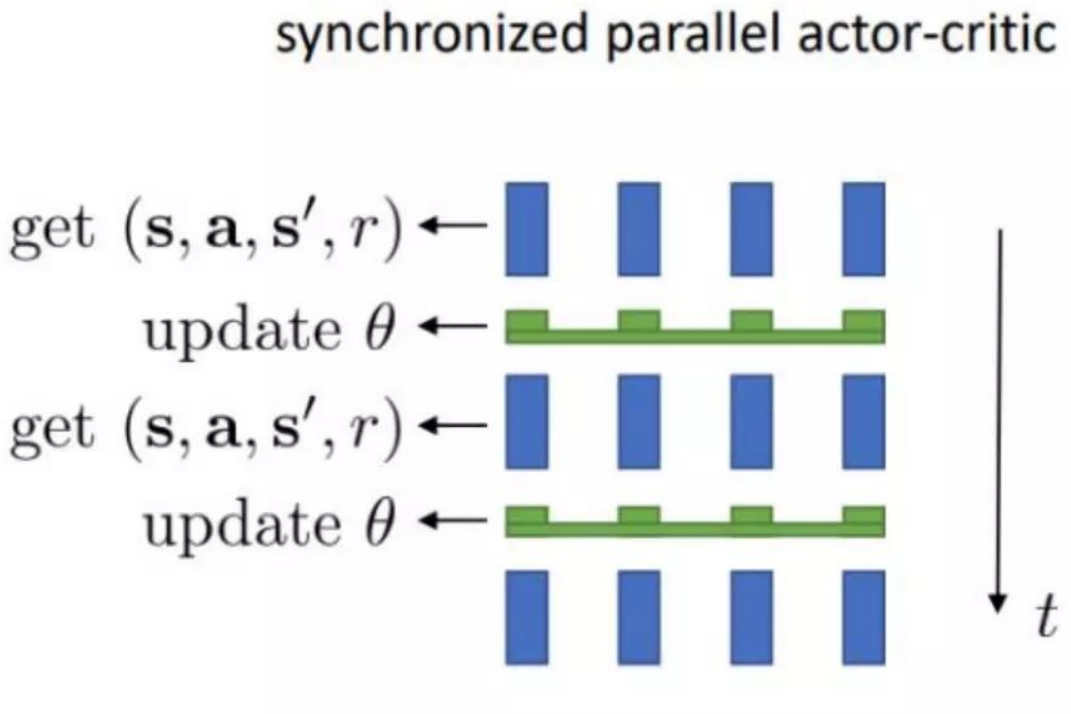
Asynchronous Advantage Actor-Critic

- Asynchronous Advantage Actor-Critic
  - proposed by DeepMind
  - use n-step bootstrapping
  - easier to implement (without lock)
  - some variability in exploration between workers
  - only use CPUs, poor GPU usage



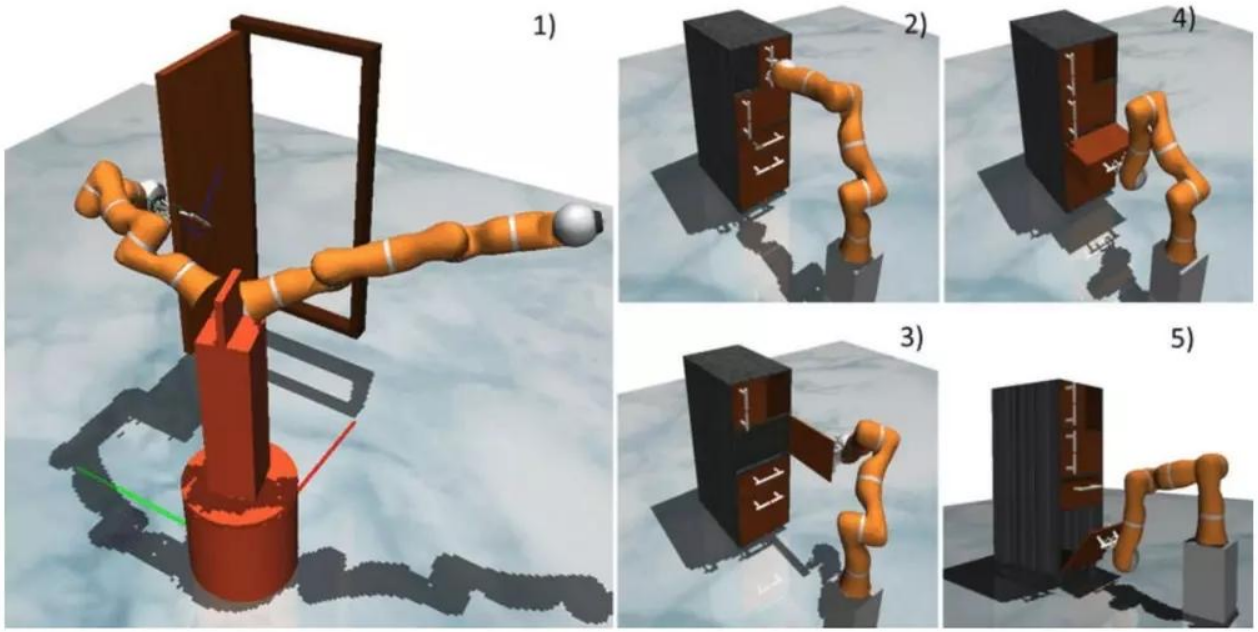
Synchronous Advantage Actor-Critic

- Synchronous Actor-Critic
  - proposed by OpenAI
  - synchronous workers
  - use n-step tricks
  - better GPU usage



The usage of Actor-Critic algorithm

Gaming, model research



Robotics, continuous control

WHAT IS AI IN GAMING?

AI in gaming refers to excellent game experiences which is more-

1. Responsive
2. Adaptive
3. Challenging

Artificial Intelligence brings revolution in player experience, cost reduction, better performance in gaming sector.

WHY DOES AI MATTER IN GAMING?

01

PLAYER EXPERIENCE

02

BETTER PERFORMANCE

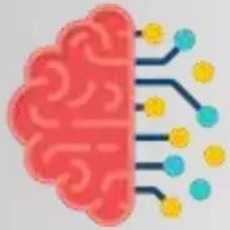
03

COMPATIBILITY

04

INTERACTIVE STORY

BENEFITS OF AI IN GAMING



The games become smarter and more realistic

Makes it easier for the user to play

Developers make Games that are Very Human-Like

BENEFITS OF AI IN GAMING

Eliminates the predictability of the game

Create New Level Gaming using Intelligence

Simplifying the Game Creation and Development Process.

02

APPLICATIONS

Typical applications of AI in games

ADDING INTELLIGENCE TO NON-PLAYING CHARACTERS

GARRY KASPAROV

vs

DEEP BLUE

(1997)

ALPHA GO
vs
LEE SEDOL
(2016)

- More complex than Chess
- Used neural networks instead of probability algorithms

AI-BASED NPCs

Not Pre-programmed

Less Predictable

Get Smarter

Imitate Top Players

AI ENEMY

Well-implemented enemy AI is characterized by the believability of opposing NPC movement, how they converse, behave, and react to any given situation created by the player.

AI COMPANION

AI companions provide friendly faces in a hostile environment. They assist in navigation, puzzle-solving, and combat. They alleviate loneliness.

HALO GAME SERIES

HALO ENEMY AI

Smart enough to retreat and take cover

Micro-decisions and short-term strategies to avoid repetitive outcome

Previous fight records used to make the enemy learn

GAME LEVEL GENERATION

AI IN GAME-LEVEL GENERATION

PROCEDURAL CONTENT GENERATION

GENERATE NEW ENVIRONMENTS

CREATE NEW ASSETS

DIFFICULTY ESTIMATION

NO MAN'S SKY

One of the primary aims of play, aside from staying alive, building stuff and acquiring goods, is to explore as many of the over 18 quintillion randomly generated planets as you like.

Open world? Try open universe.

STORY DEVELOPMENT

AI IN STORY DEVELOPMENT

PLAYER EXPERIENCE MODELING

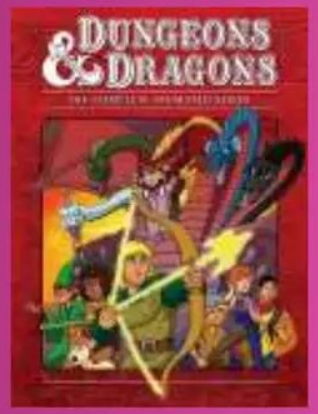
DATA MINING

INTERACTIVE NARRATIVE

Every interaction opens up a series of possible improvements.

These actions are somewhat randomized by design so the AI doesn't start to feel predictable.

AI DUNGEON



Based on
Dungeons & Dragons

Generate New
Adventure

Infinite
Possibilities

Planet Omega is a lush, verdant planet. Inhabited by many different sentient species, Omega has been a peaceful world until the arrival of the Seekers. The Seekers have been traveling from world to world, converting entire civilizations to their cause of tyranny. Once a planet has been conquered, the Seekers and their robotic armies strip the planet bare of its resources, including the lives of the inhabitants. The only hope for Planet Omega is the Resistance.

You are Mavrick, a male human gunslinger and mercenary in the service of the resistance fighting against the Seeker's rebellion on Planet Omega. You live in the Holy City. You make your living as a bounty hunter, hunting down and executing the most wanted criminals. You never thought you'd be a part of this resistance, let alone be a gunfighter. The only thing that matters right now is killing Seeker scum, and you need money to do it.

> You find a seeker, kill him, and take all his money.

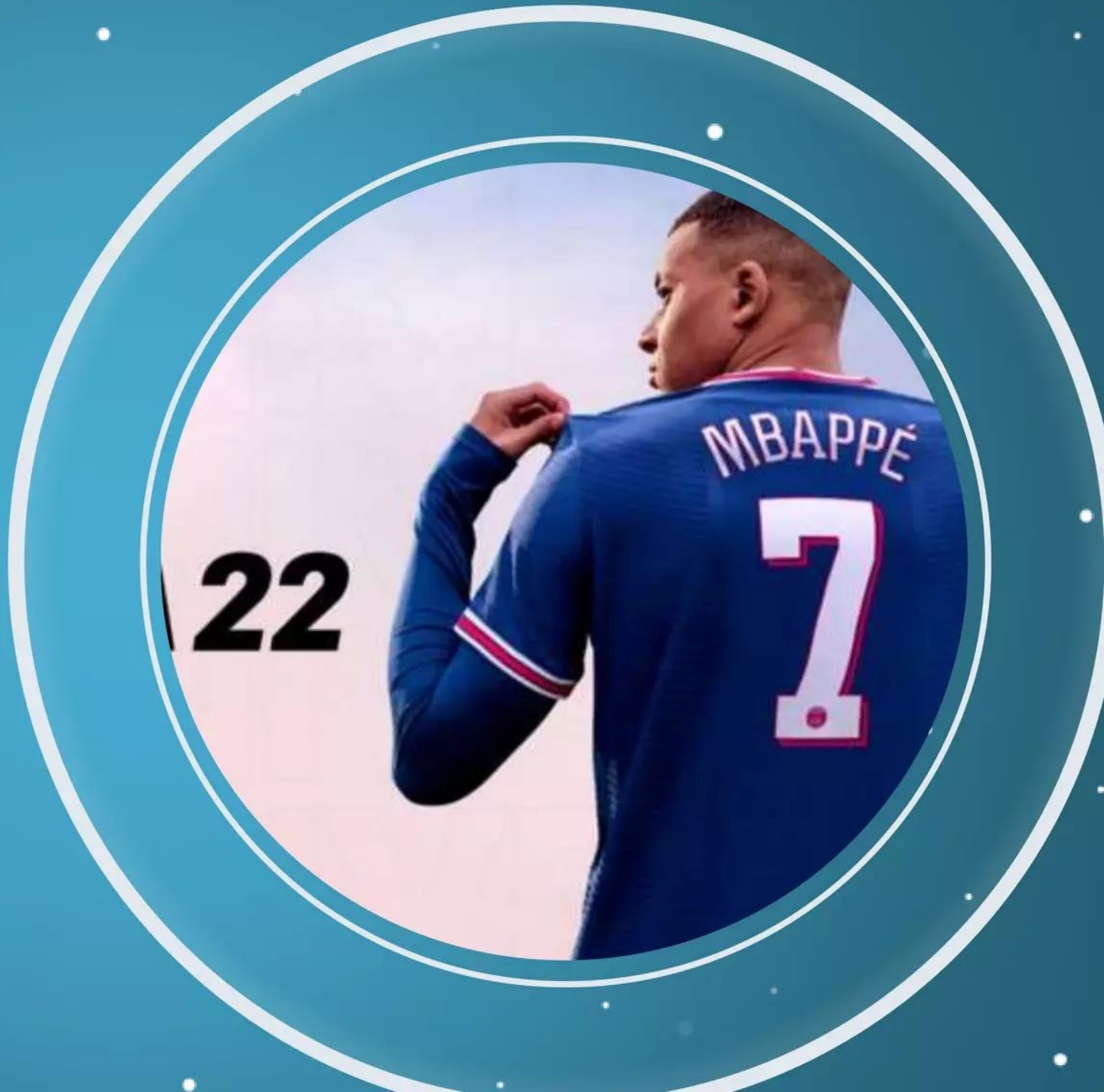
You find a lone Seeker, and shoot him in the back of the head. You quickly take all his money, and leave the Holy City before the Seeker's fellows realize what you've done. You head back to the Gunsmith, and hide the money in a secret compartment in the wall of the vault.

You'll need it to buy the supplies you need as a gunfighter, and you're pretty sure you'll be able to find a place where you can stash it.
"Well, I guess I'm a part of the resistance now," you say to yourselves, as you finish your work.

UNDO REDO RETRY EDIT RESTORE MEMO VII FEEDBACK HELP

Do What do you do?

IN-GAME BALANCING



AI IN IN-GAME BALANCING

PLAYER EXPERIENCE MODELING

OPTIMIZATIO N

TESTING

FIFA 22

ULTIMATE MODE

- Team Chemistry
- In-game Events
- Competitiveness

03

ALGORITHMS in GAMES

Some defined game-AI algorithms

BOARD GAME

- GAME TREE CALCULATION
- BOARD EVALUATION

VIDEO GAME

PATH FINDING -
STEERING BEHAVIORS -

VIDEO GAME

PATH FINDING ALGORITHMS

INFORMED SEARCH

A* SEARCH

UNINFORMED SEARCH

BFS/DFS

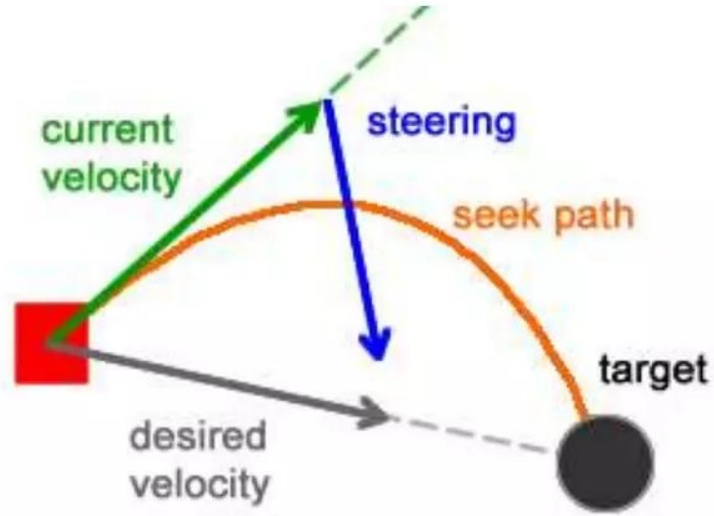
VIDEO GAME
STEERING BEHAVIOR

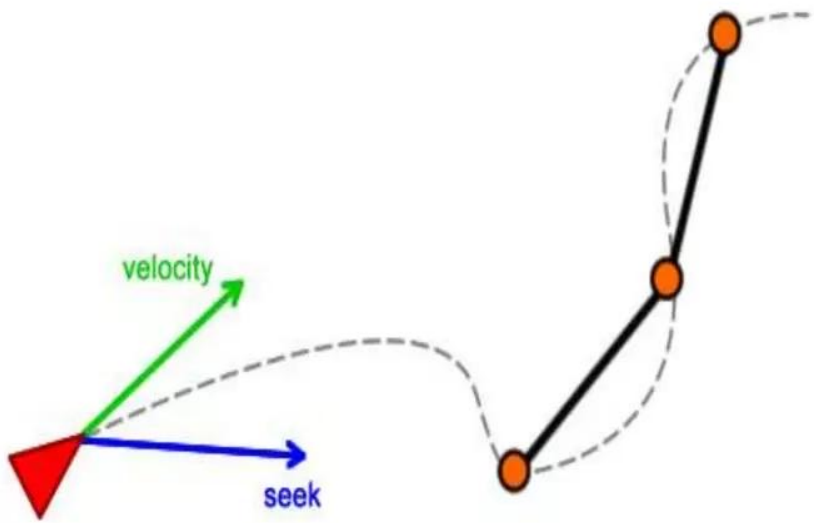
SEEK

PATH
FOLLOWING

SEEK

The addition of steering forces to the character every frame makes it smoothly adjust its velocity, avoiding sudden route changes.





PATH FOLLOWING

The path following steering refers to the character constantly adjusting its direction to catch the target.

Thank you!