DESAIN PERILAKU AKSI MOBIL PADA GAME SIMULASI MENGGUNAKAN REINFORCEMENT LEARNING ACTOR-CRITIC

Bakti Yudha Wiradhika Utama

Universitas Jenderal Achmad Yani, Cimahi, Indonesia
E-mail : baktiyudha91@gmail.com

ABSTRAK

Dalam game simulator mobil, pengambilan keputusan yang tepat dan optimal dalam mengendalikan mobil merupakan tantangan penting. penelitian sebelumnya tentang penggunaan simulator torcs, yang merupakan simulator balap mobil open source, telah berhasil mengembangkan agent AI untuk mobil otonom dengan menerapkan Reinforcement Learning (RL) namun tentu dalam pelatihan agen yang dilakukan oleh algoritma tersebut masih perlu dilakukan penelitian lebih lanjut sehingga Dalam penelitian ini menggunakan metode Reinforcement Learning (RL) Actor-Critic dengan metode Advantage Actor-Critic (A2C) untuk meningkatkan aksi yang dihasilkan oleh agen virtual dalam game simulasi mobil, terutama dalam hal pengaturan kecepatan mobil. Melalui pengumpulan data lingkungan dari simulasi mobil, Reinforcement Learning (RL) Actor-Critic akan dilatih untuk menentukan kecepatan yang sesuai berdasarkan keadaan lingkungan seperti jarak antara mobil, kecepatan saat ini, dan batasan kecepatan pada lintasan. Tahapan penelitian ini mencakup pengumpulan data lingkungan simulasi mobil, preprocessing data untuk mempersiapkannya untuk pelatihan Reinforcement Learning (RL) Actor-Critic, dan implementasi Reinforcement Learning (RL) Actor-Critic dengan menggunakan metode Advantage Actor-Critic (A2C). Selanjutnya, Reinforcement Learning (RL) Actor akan bertanggung jawab untuk memilih tindakan kecepatan yang optimal, sedangkan Critic akan mengevaluasi kinerja dan memberikan umpan balik yang diperlukan.Penelitian ini diharapkan dapat menghasilkan agen virtual dalam car simulator yang mampu mengambil keputusan yang lebih tepat dan optimal dalam pengaturan kecepatan Mobil. Selain itu, penelitian ini juga dapat memberikan kontribusi penting dalam penerapan Reinforcement Learning (RL) Actor-Critic dengan metode Advantage Actor-Critic (A2C) dalam konteks game seperti simulasi mobil.

Kata Kunci : *Reinforcement Learing, Actor-Critic, Game Simulator, Multi-Agent , Multi-Agent Reinforcement Learning, Advantage Actor-Critic*

ABSTRACT

In car simulator games, making accurate and optimal decisions in controlling trains is a crucial challenge. Previous research has successfully developed AI agents for autonomous cars using Reinforcement Learning (RL) by utilizing the open-source racing car simulator torcs. However, further investigation is needed regarding the training process of these agents. In this study, we employ the Reinforcement Learning (RL) Actor-Critic method with the Advantage Actor-Critic (A2C) approach to enhance the actions generated by virtual agents in train simulator games, particularly in terms of speed regulation. By collecting environmental data from the train simulation, the Reinforcement Learning (RL) Actor-Critic model is trained to determine appropriate speeds based on environmental factors such as train distance, current speed, and track speed limits. The research process involves data collection from car Simulator environments, data preprocessing for Reinforcement Learning (RL) Actor-Critic training, and the implementation of Reinforcement Learning (RL) Actor-Critic using the Advantage Actor-Critic (A2C) method. The Reinforcement Learning (RL) Actor is responsible for selecting optimal speed actions, while the Critic evaluates performance and provides necessary feedback.We expect this research to yield a virtual agent within car Simulator capable of making more accurate and optimal decisions in train speed regulation. Additionally, this study contributes to the effective application of Reinforcement Learning (RL) Actor-Critic with the A2C method in car simulation games.

Keywords : *Reinforcement Learing, Actor-Critic, Game Simulator, Multi-Agent , Multi-Agent Reinforcement Learning, Advantage Actor-Critic*

PENDAHULUAN

Simulator mobil menciptakan pengalaman realistis dalam mengoperasikan sebuah mobil, ini mencakup sejumlah fenomena untuk memberikan pelatihan yang efektif kepada pengemudi. Fenomena ini mencakup navigasi kompleks, yang melibatkan kemampuan operator atau agen dalam menangani berbagai kondisi trek Mobil , termasuk tikungan, stasiun, dan persimpangan. Kepatuhan terhadap aturan lintas juga menjadi aspek penting, di mana pengemuni belajar untuk memprioritaskan jalur benar, mengikuti sinyal, dan menanggapi tanda-tanda jalan.

Metode reinforcement learning (RL) dalam simulator mobil membawa pelatihan pengemudi mendekati pengalaman simulasi pada kondisi dunia nyata. Dalam simulasi mobil ini, agent menggunakan Reinforcement learning untuk belajar merespons terhadap fenomena umum dalam operasi mobil. Dengan memanfaatkan teknologi ini, operator Mobil dapat melatih keterampilan navigasi, mematuhi aturan lintas dengan responsif, dan mengelola kecepatan kendaraan sesuai dengan kondisi trek dan situasi lalu lintas yang berubah-ubah.

Penelitian yang dilakukan ini membuat sebuah game simulasi mobil yang akan memanfaatkan Reinforcement learning actor critic menggunakan metode A2C di dalamnya sehingga membuat agent dari mobil aksi didalamnya lebih optimal dan efisien.

Potensi penggunaan AI pada game simulator sebagai peningkat kualitas game bisa mengembangkan agen cerdas yang terdapat dalam game simulator tersebut meningkat menggunakan Reinforcement Learning yang masuk kedalam kategori machine learning. machine learning memiliki potensi dan peran yang sangat besar dalam perancangan dan pengembangan game[1]. Adapun penelitian sebelumnya tentang penggunaan simulator torcs, yang merupakan simulator balap mobil open source, telah berhasil mengembangkan agent AI untuk mobil otonom dengan menerapkan Reinforcement Learning. Hal ini menunjukkan keberhasilan dalam mengoptimalkan perilaku agen untuk mencapai kinerja yang optimal dalam mengemudi mobil otonom[2]. Adapun beberapa agen dapat berinteraksi satu sama lain dan dengan lingkungan mereka menggunakan kombinasi jaringan actor critic, memungkinkan agen-agen untuk belajar dari interaksi mereka dengan lingkungan[3].

Penelitian sebelumnya penggunaan reinforcement learning adalah pendekatan yang menjanjikan untuk mendapatkan perilaku otonom NPC dalam permainan video. AI4U (Artificial Intelligence for Unity) membantu mengatasi hambatan dalam mengintegrasikan algoritma reinforcement learning terkini ke dalam lingkungan permainan dengan mempromosikan spesifikasi visual fungsi reward, lingkungan, dan agen, serta generasi kode otomatis kemudian hasil yang didapatkannya merupakan Pendekatan pembelajaran dianggap sebagai pendekatan yang menjanjikan untuk mendapatkan perilaku otonom NPC dalam video game. Salah satu hambatan dalam penelitian pendekatan baru untuk reinforcement learning yang diterapkan dalam permainan adalah kompleksitas yang ditambahkan untuk mengintegrasikan algoritma RL terkini ke dalam lingkungan permainan. AI4U (Artificial Intelligence for Unity) membantu mengatasi hambatan ini dengan mempromosikan spesifikasi visual fungsi reward, lingkungan, dan agen, serta generasi kode otomatis[4].

Penelitian sebelumnya menjelaskan tentang Reinforcement learning memungkinkan AI untuk belajar dari interaksinya dengan lingkungan, Algoritma reinforcement learning khusus yang telah berhasil dalam melatih AI untuk video game pertarungan termasuk Proximal Policy Optimization (PPO),yang merupakan metode kebijakan gradien dalam reinforcement learning yang telah digunakan untuk melatih agen permainan dengan deep reinforcement learning

kemudian untuk hasil dari penelitian ini implementasi algoritma pembelajaran penguatan untuk bot AI dalam permainan video pertarungan telah berhasil dilakukan lalu Konvergensi pembelajaran agen terjadi ketika nilai total hadiah kumulatif berada dalam rentang 140 hingga 120 artinya agen telah belajar dengan baik dan mampu mengambil keputusan yang optimal berdasarkan pengalaman yang diperoleh dari lingkungan[5].

Penelitian yang telah dilakukan ini menggunakan vektor sebagai representasikan berbagai aspek dalam operasi mobil. Vektor diisi oleh variabel-variabel seperti Posisi Mobil, Kecepatan mobil, Jarak ke Berikutnya, Waktu, Status Sinyal, dan Kondisi Jalur. Dengan teknologi ini, kita dapat memodelkan dan mensimulasikan operasi Mobil secara lebih baik. Hal ini akan meningkatkan daya tarik game dan menarik lebih banyak pemain.

Penelitian yang dilakukan ini akan membuat sebuah simulasi mobil yang akan memanfaatkan Reinforcement learning actor critic menggunakan metode A2C di dalamnya sehingga membuat agent dari mobil aksi didalamnya lebih optimal dan efisien.

METODE

Di Dalam penelitian ini melewati beberapa tahap seperti :

Tahap Perencanaan:

a) Disini menentukan tujuan,ruang lingkup, dan metode dari penelitian seperti tujuan dari penelitian ini membuat agen yang bisa mengeluarkan aksi sesuai dengan kondisi kondisi sesuai dengan lingkungan yang berjalan menggunakan metode reinforcement learning actor critic, input state yang dimasukkan kedalam metode ini berupa posisi, jarak mobil , sudut mobil, kecepatan mobil , titik jarak berikutnya

Tahap Pra Proses:

a) Mempersiapkan lingkungan simulasi car Simulator, lingkungan ini akan menggunakan pygame sebagai basenya kemudian pyton sebagai Bahasa pemprograman yang digunakan
b) Mengumpulkan data pelatihan. Seperti kecepatan mobil,jalur titik mobil, total reward, dan penalty point

Tahap Perancangan Sistem:

a) Mempersiapkan desain model state dan action space input state akan diambil 5 input kedalamnya kemudian akan dinormalisasi untuk pemproses Reinforcement learning dengan skala yang lebih kecil bertujuan untuk masuknya data kedalam metode
b) Membangun arsitektur actor-critic dengan pendekatan Advantage Actor-Critic (A2C). masing masing actor dan critic masing masing memiliki neural network dengan 4 lapisan, yang terdiri dari input layer , hidden layer , hidden layer dan output
c) Menentukan fungsi reward
d) Menentukan Inisialisasi agen
e) Menentukan Pelatihan agen

Tahap Evaluasi:

a) Evaluasi kinerja agen dengan menggunakan based reward evaluation yang mana ini menghitung comulative reward dan moving avarange

Tahap Analisis dan Kesimpulan:

a) Menganalisis hasil

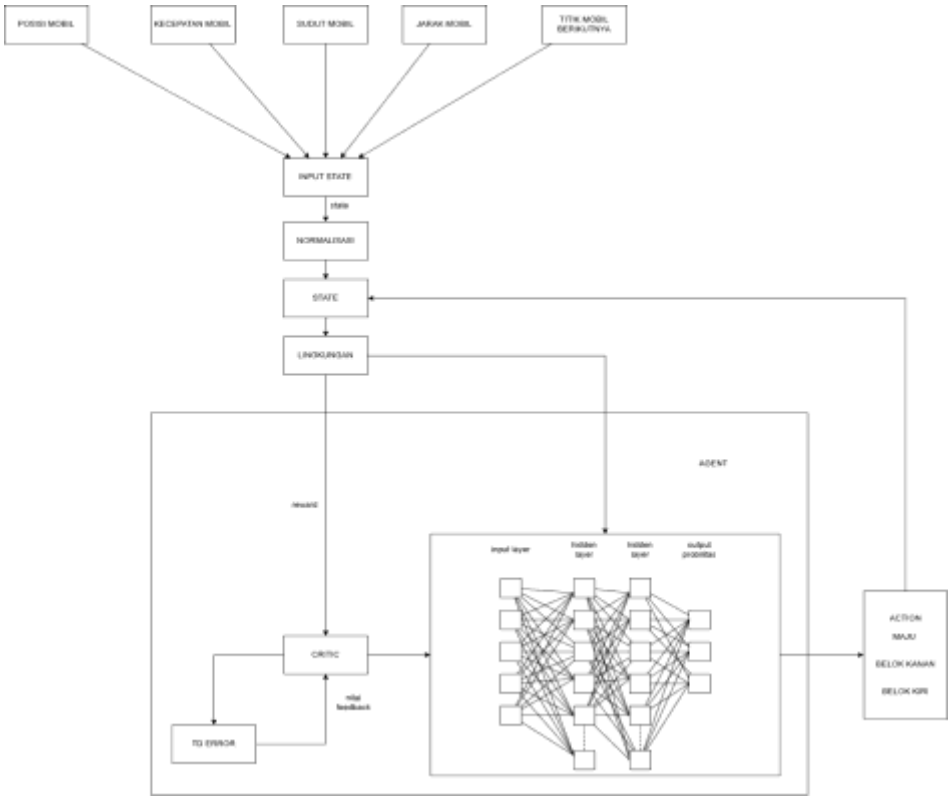
Perancangan sistem

Berikut merupakan analisis perancangan dari sistem yang ada didalam penelitian yang dibuat:

a) Jumlah Episode
   Parameter ini menentukan berapa banyak episode yang akan dilakukan dalam pelatihan agen. Jumlah episode yang sudah dilakukan sebanyak 1000 episode dalam setiap percobaan.
b) Reward Function
   Terdapat beberapa reward function didalamnya seperti:
   1. Jarak terdekat:
      Deskripsi: jarak terdekat mobil dari titik jalur yang benar.
   2. Jalur benar:
      Deskripsi: posisi agen apakah dia berada di jalur yang benar atau tidak.
   3. Penalty tabrakan:
      Deskripsi: penalti jika agen menabrak pembatas jalur.
c) Perolehan data
   Perolehan data skor yang diakumulasi reward dan actor critic loss selama 1000 episode didapatkan melalui jalannya simulasi sebanyak 1000 kali, setiap simulasi menghasilkan skor yang mencerminkan kinerja agen self driving yang dilatih menggunakan metode actor critic
d) Sudut mobil
   Sudut mobil yang digunakan ini menjadi pusat sudut yang akan diproses nanti jadi sudut mobil yang akan masuk perhitungan terhadap lingkungan dimulai dengan sudut 0^o

Perancangan agent

Pada bagian ini akan membahas tentang perancangan sistem agent Mobil



Gambar 1. Perancangan Sistem

Bisa dilihat seperti diagram gambar 1 agent menerima 5 fitur yang menjadi input state awal yaitu posisi mobil meliputi koordinat x dan y, kecepatan mobil v , sudut mobil dengan sudut defaultnya 0*, jarak mobil , dan titik mobil berikutnya ini meliputi jarak mobil ketujuan. Setelah actor menerima input state melalui lingkungan actor akan melakukan aksi awal sehingga nantinya akan dievaluasi oleh critic yang feedback nya ini akan diterima Kembali oleh actor untuk aksi yang lebih baik, proses ini terus terjadi berulang ulang sampai goals agent tercapai.

Perancangan Perangkat lunak

Membahas tentang proses perancangan perangkat lunak dalam penelitian meliputi Use case diagram yang merupakan interaksi antara pengguna sistem dengan sistem. Deskripsi Usecase yang menjelaskan beberapa yang terjadi didalam usecase diagram Use case scenario terdiri dari mobil bergerak kedepan, memperlambat kecepatan, berhenti, menjaga jarak aman, activity diagram , sequence diagram dan lain lain

Output

Agen dalam penelitian ini menghasilkan sebuah keluaran berupa aksi maju kedepan, belok kekiri, dan belok kekanan. aksi agen mobil yang akan dipilih oleh agen akan berpengaruh dari masukan variable penting didalam membangun agen ini diantaranya, menginisialisasi Posisi Mobil, Kecepatan Mobil, sudut arah agen yang akan membangun aksi yang optimal dan efisien bisa diberikan oleh agen.

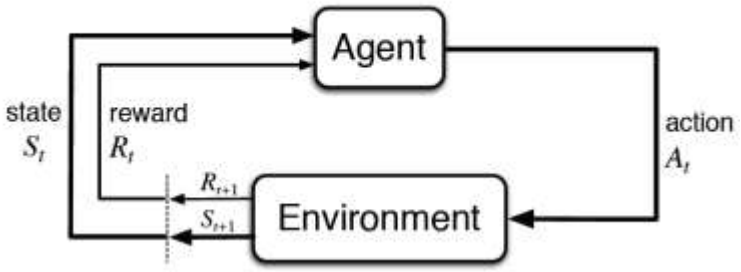
HASIL PENELITIAN DAN PEMBAHASAN

Berdasarkan beberapa informasi yang telah didapatkan ada beberapa pembahasan pembahasan yang berkaitan dengan penelitian

Reinforcement Learning

Ilmu pengambilan keputusan yang mempelajari perilaku yang optimal dalam lingkungan interaktif. Dalam RL, agen belajar berperilaku dengan melakukan tindakan dan melihat hasil dari tindakan tersebut. Setiap tindakan yang baik memberikan umpan balik positif, sedangkan tindakan yang buruk menghasilkan umpan balik negatif atau penalty

Ketika agen berinteraksi dengan lingkungan pada saat tertentu t, akan mempersepsikan suatu keadaan st+1 dan melakukan tindakan at. Keadaan st dan tindakan at secara unik menentukan keadaan berikutnya s_{t+1}. Untuk setiap tindakan at, agen menerima imbalan rt ∈ ℝ. Siklus persepsi, tindakan, dan imbalan ini berlangsung seiring waktu[4].



Gambar 2. Arsitecture Reinforcement learning

Keterangan:
P: Fungsi transisi
γ: Diskon faktor

Actor-Critic

Sebuah sistem Actor-Critic merupakan kombinasi dari reinforcement learning berbasis nilai (value-based) dan reinforcement learning berbasis kebijakan (policy-based). Dalam sistem ini, terdapat dua jaringan yang memiliki parameter berbeda:

Critic: Jaringan ini memperkirakan fungsi nilai (seperti pada metode reinforcement learning berbasis nilai). Fungsi nilai ini mengestimasi nilai sebenarnya dari suatu keadaan tertentu atau pasangan keadaan-tindakan.

Actor: Jaringan ini memperbarui kebijakan (policy) berdasarkan arahan yang disarankan oleh Critic (seperti pada metode reinforcement learning berbasis kebijakan).

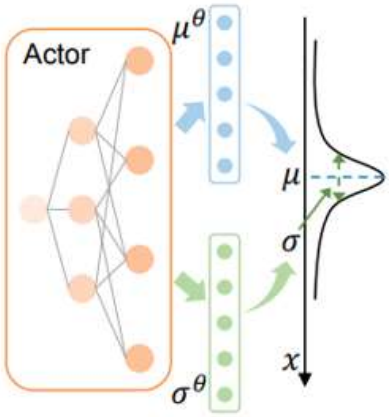
$$critic_{loss} = \frac{1}{n}\sum\left(V(s) - \left(r + \gamma V(s')\right)^2\right)$$

Persamaan 1. Persamaan Critic

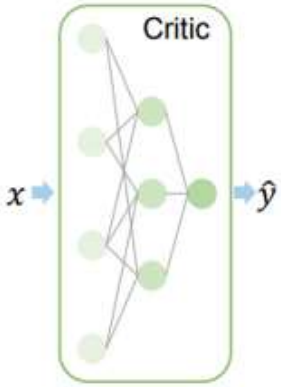
$$aktorloss = -log(\pi(a \mid s)) \cdot advantage$$

Persamaan 2. Persamaan Actor

Keterangan:
n: Jumlah sampel dalam batch.
π(a|s): Probabilitas memilih aksi a pada state



Gambar 3. Actor



Gambar 4. Critic

Gambar 3 menggambarkan struktur jaringan untuk ruang desain kontinu. Dalam ruang desain kontinu, tindakan yang dihasilkan oleh aktor adalah bilangan riil. Struktur jaringan ini menggunakan fungsi softmax untuk menghasilkan distribusi probabilitas dari tindakan yang mungkin. Dalam hal ini, setiap tindakan memiliki probabilitas tertentu, dan tindakan yang dipilih adalah tindakan dengan probabilitas tertinggi[6].

Gambar 4 menunjukkan struktur jaringan untuk kritikus (critic) dalam metode Actor-Critic. Kritikus bertanggung jawab untuk mengevaluasi kinerja aktor dengan memberikan umpan balik (feedback) tentang keuntungan (advantage) dari tindakan yang diambil oleh aktor. Struktur jaringan ini menggunakan fungsi regresi linier untuk memperkirakan keuntungan dari setiap tindakan yang diambil oleh aktor[6].

Implementasi sistem

Berdasarkan perancangan sebelumnya hasil dari penelitian ini merupakan sebuah perangkat lunak yang digunakan untuk mengembangkan agen mobil menggunakan reinforcement learning metode actor, implementasi perangkat lunak ini menggunakan Bahasa pyton dengan memanfaatkan beberapa library pygame untuk base gamenya serta beberapa lagi seperti torch, Numpy, Tensorflow, dan matplotlib.

Model actor critic ini melibatkan Tensorflow, model ini terdiri dari 4 lapisan network masing masing antara actor sebagai pengambil probilitas Keputusan dan critic sebagai feedback

untuk nilai fungsi yang diambil oleh actor. Dalam pelatihan ini agen menerima beberapa reward untuk aksinya tetap sesuai ketujuannya diantaranya reward jarak titik selanjutnya,kesesuaian dalam jalur, tidak menabrak , mencapai tujuan.

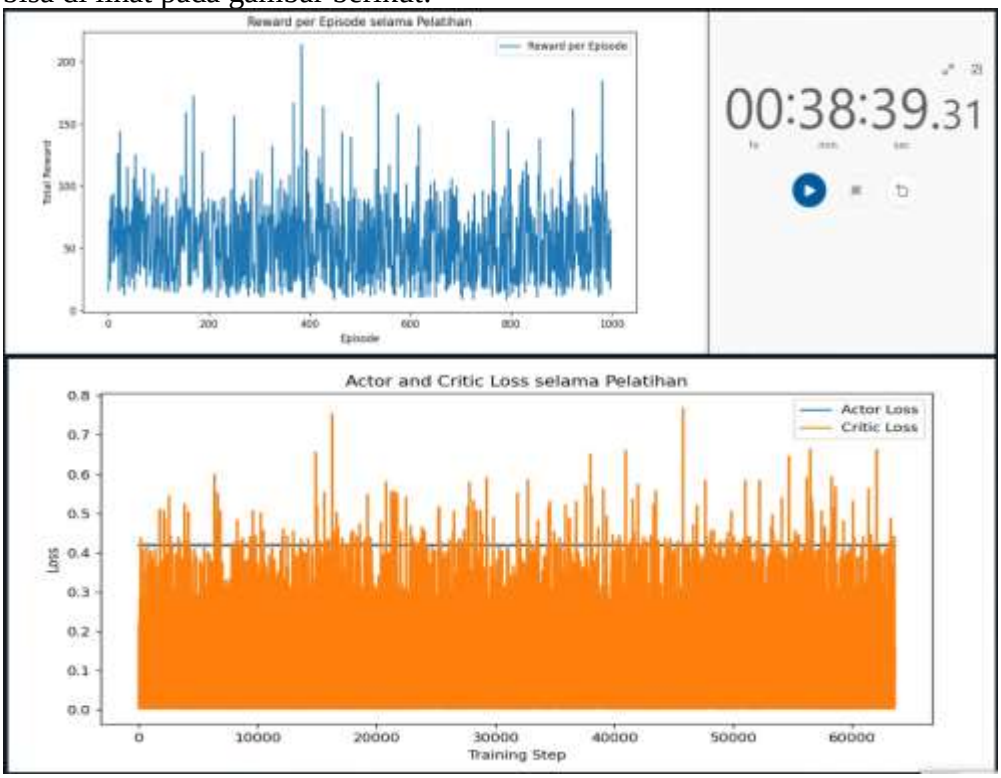
Lingkungan dalam simulasi ini memanfaatkan pygame didalamnya Dimana lingkungannya memiliki elemen yang diimplementasikan diantaranya agen,lintasan dan score reward. Proses implementasi dimulai dengan membuat dan melatih model actor critic yang dijalankan langsung agen akan berinteraksi dengan lingkungannya
untuk mencapai titik tujuan. Reward akan di berikan kepada agent melalui car acara seperti pada jalurnya,tidak menabrak, dan sampai tujuan.

Pengujian kinerja agent

Pengujian model actor critic ini kinerja agen dilatih beberapa pengujian disini agen dilakukan beberapa runing dengan total episode 1000 per runing

a) Percobaban pertama

percobaan pertama dilakukan yang menghasilkan agen masih belum mencapai tujuannya dengan baik akan tetapi reward yang dihasilkan per episode masih terbilang rangenya kecil, kemudian actor loss dan critic loss untuk actor dia cenderung stabil akan tetapi critic loss masih berfluktuasi besar menandakan masih melakukan eksplorasi terus menerus grafik bisa di lihat pada gambar berikut:



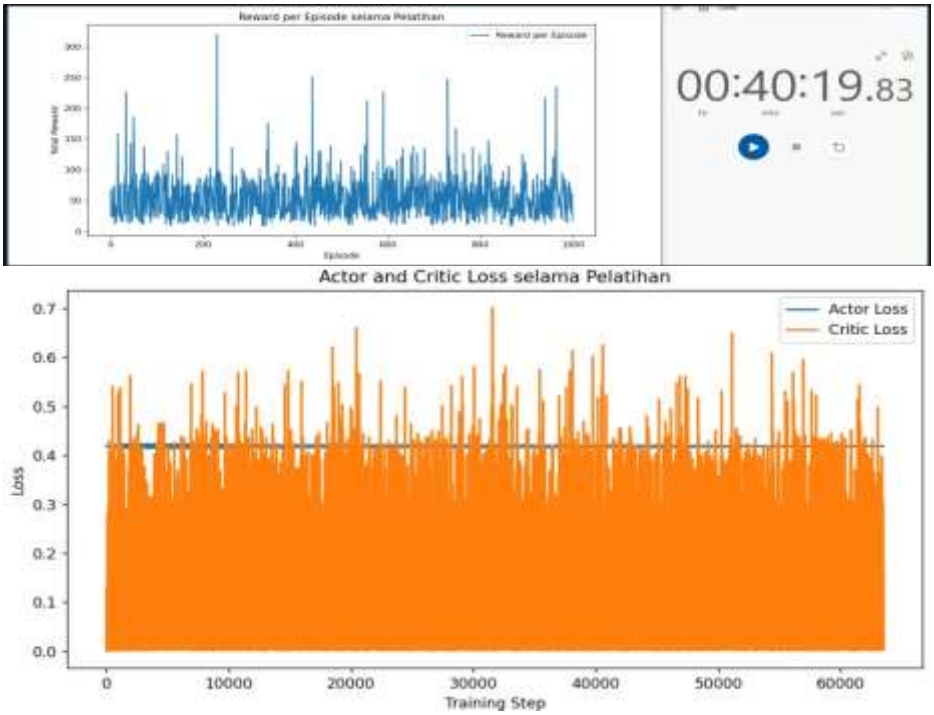
Gambar 5. Percobaan pertama

Dapat dilihat untuk percobaan awal mencapai 1000 episode dibutuhkan waktu kurang lebih 38 menit.

b) Percobaan kedua

Percobaan kedua mekanismenya sama seperti percobaan sebelumnya terlihat reward sudah sedikit menurun dibanding percobaan pertama akan tetapi ada beberapa episode yang mencapai

reward tertinggi dari percobaan pertama dan agen belum bisa mencapai tujuan grafik bisa dilihat Digambar berikut:

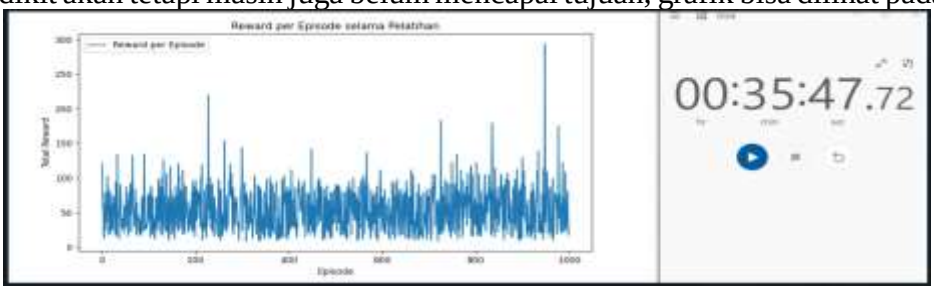


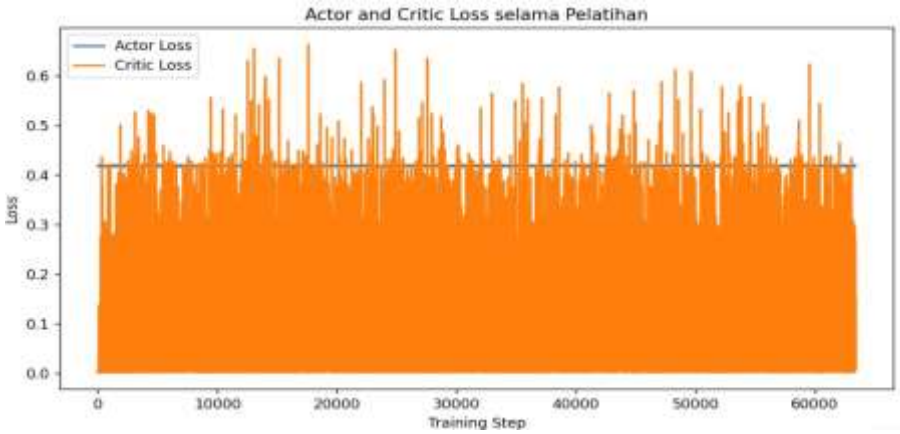
Gambar 6. Percobaan kedua

Dapat dilihat pada grafik percobaan kedua ini memerlukan waktu kurang lebih 40 menit untuk mencapai 1000 episode dan nilai tertinggi rewardnya berada di range antara episode 200 – 300 yaitu 300

c) Percobaan ketiga

Percobaan ketiga ini mekanismenya masih seperti sebelumnya terlihat reward grafiknya ini stabil akan tetapi cenderung di nilai terendah diantara 50 – 100 rata rata disini sudah mengikuti jalur sedikit akan tetapi masih juga belum mencapai tujuan, grafik bisa dilihat pada gambar berikut:



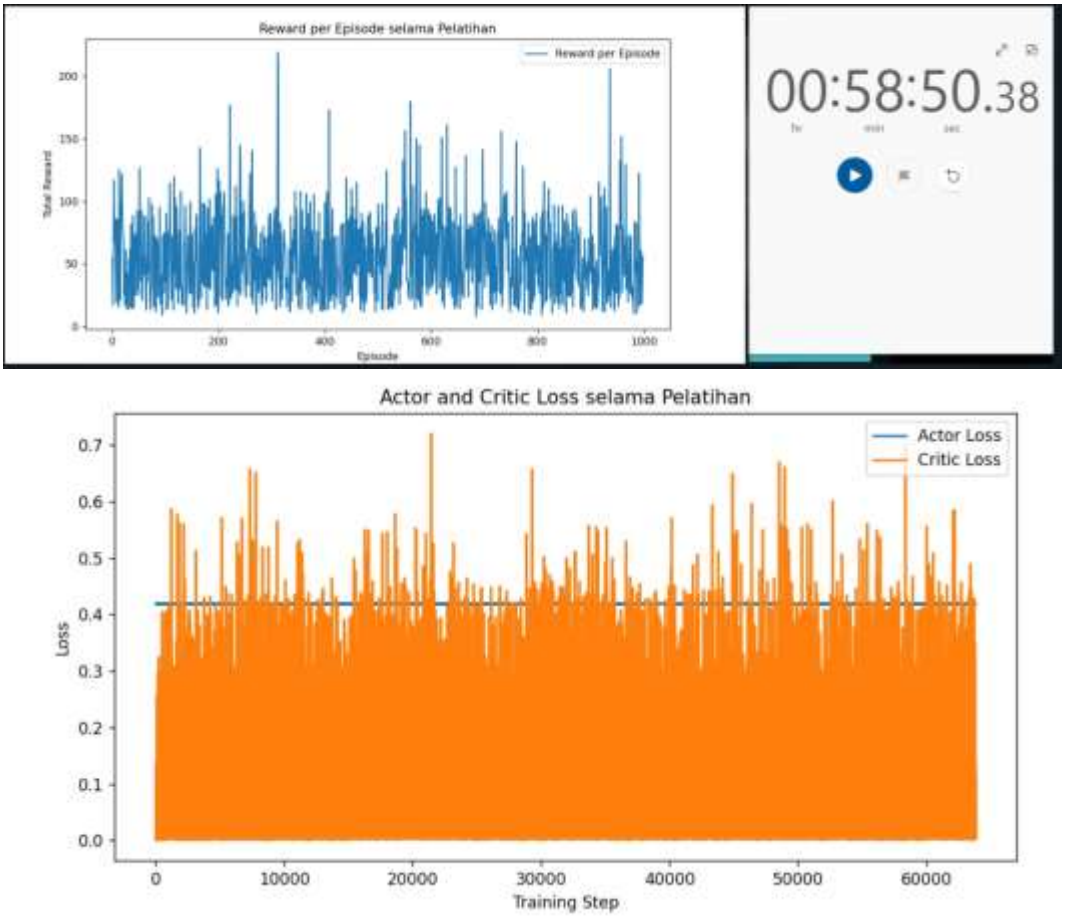


Gambar 7. Percobaan ketiga

Percobaan ketiga ini memerlukan waktu yang lebih cepat dari ke2 percobaan sebelumnya kurang lebih 35 menit

d) Percobaan keempat

Percobaan keempat ini mengalami peningkatan reward score diatas 50 sering terjadi di percobaan ini akan tetapi waktu pelatihan ini juga meningkat grafiknya bisa dilihat Digambar berikut:

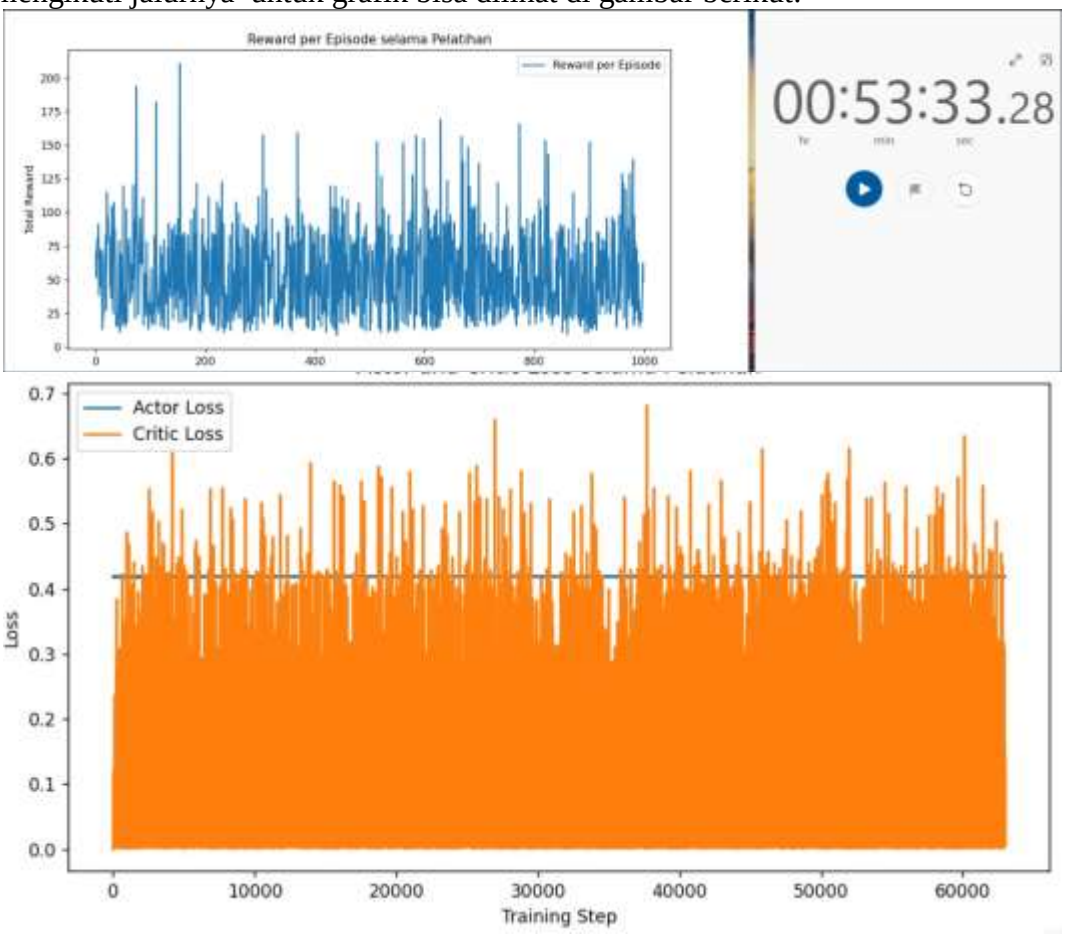


Gambar 8. Percobaan keempat

Percobaan keempat ini bisa dilihat waktu pelatihan memakan waktu kurang lebih 58 menit reward cenderung lebih besar per episode tetapi nilai tertingginya ada di 200

e) Percobaan kelima

Pada percobaan kelima ini nilainya tidak beda jauh dengan percobaan keempat agen lebih sering mengikuti jalurnya untuk grafik bisa dilihat di gambar berikut:



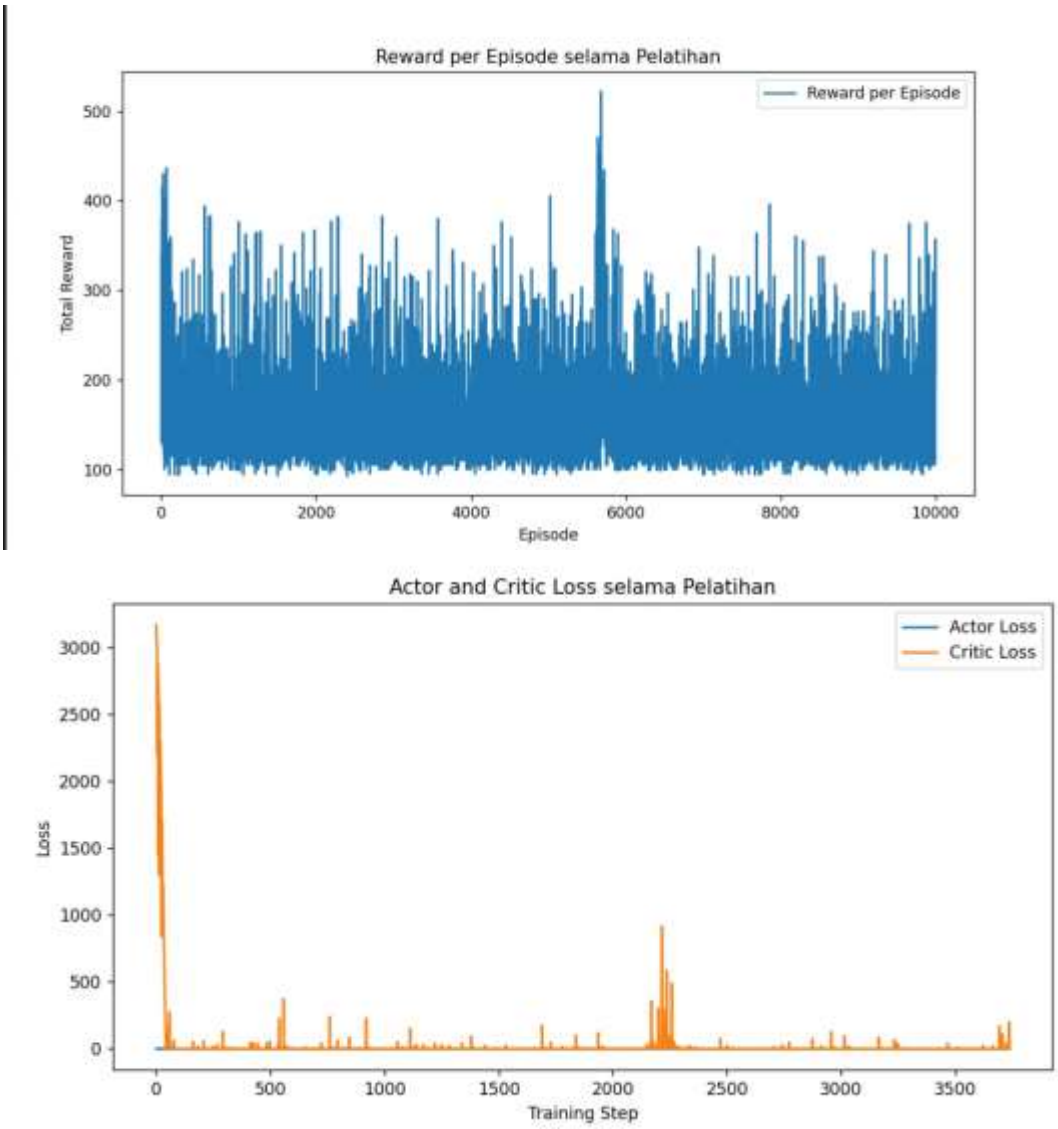
Gambar 9. Percobaan kelima

Percobaan kelima ini bida di lihat waktu pelatihannya kurang lebih 53 menit reward cenderung menyentuh diatas rata rata beberapa kali dan score reward tertingginya ada di 200

Percobaan untuk melatih model actor critic telah diambil 5 percobaan awal tampak terlihat grafik reward mengalami kenaikan dari percobaan pertama ke percobaan selanjutnya akan tetapi agen memang lebih sering mengikuti jalurnya tapi belum sampai ketujuan yang ditentukan diatas sehingga harus dilakukan percobaan agar bisa lebih optimal.

f) Percobaan keenam

Percobaan ini merupakan percobaan terakhir dengan cara melakukan running 10.000 Episode kemudian dihasilkan grafik seperti ini:



Gambar 10. Percobaan keenam

Merupakan reward yang dialami agen selama pelatihan perepisodenya dapat terlihat diantara 5500 – 6000 episode agen mengalami kenaikat tertinggi reward di angka 500, akan tetapi agen masih saja belum mencapai goalnya

Grafik untuk menunjukkan critic loss dan actor loss diawal pelatihan mengalami fluktuasi akan tetapi seiring pelatihan mengalami penurunan ini artinya critic masih sedang mencoba untuk menentukan kebijakan yang optimal Uji coba 10000 episode saat menjalankan program memang menghasilkan keluaran reward yang meningkan akan tetapi agen masih tetapi belum mencapai tujuan kemudian untuk optimal action ratenya ada di 59%.

Hasil pengujian

Percobaan sudah dilakukan dengan cara melakukan running program sebanyak 1000-episode setiap percobaan didapatkan agen memang sudah melakukan jalannya simulasi menggunakan metode reinforcement learning actor critic namun setelah percobaan dilakukan memang masih ada beberapa kesalahan. Melakukan percobaan dengan 10000-episode juga masih

tetap memiliki kekurangan dalam mencapai goalsnya akan tetapi action rate dalam melakukan jalur yang tepat ada di nilai 59% akan tetapi agen masih belum mencapai goals. faktor faktor yang mempengaruhi ini menurut pendapat saya masih ada masih harus dilakukan pencocokan antara hadiah dan hukuman yang dialami oleh agen agar menemukan opsi aksi yang lebih baik,

KESIMPULAN

Penelitian ini agen menggunakan metode Reinforcement Learning (RL) untuk mempelajari bagaimana bergerak dalam jalur yang telah ditentukan. Percobaan melalui 10000 episode menghasilkan rata rata reward berada kurang lebih 175 score dengan agen berada dijalur yang tepat 59% Meskipun hasil yang diperoleh belum cukup baik, terdapat beberapa pencapaian yang didapatkan yaitu agen mampu menavigasi jalur dengan beberapa kali keberhasilan, meskipun tingkat keberhasilan secara keseluruhan masih rendah. agen sudah dapat mengambil aksi berdasarkan state yang diberikan, termasuk mempertimbangkan posisi, kecepatan, dan sudut. Kemudian implementasi metode Reinforcement learning menggunakan pendekatan Actor-Critic memberikan dasar yang kuat untuk pengembangan lebih lanjut Namun, terdapat beberapa kelemahan yang memengaruhi performa agen tingkat keberhasilan agen masih rendah karena agen sering keluar jalur atau mengambil tindakan suboptimal. Sehingga sistem reward masih perlu dioptimalkan agar agen lebih termotivasi untuk mengikuti jalur dengan konsistensi yang lebih tinggi.

Berdasarkan hasil penelitian ini, beberapa saran untuk pengembangan lebih lanjut seperti tingkatkan bobot pada progress reward agar agen lebih termotivasi untuk menyelesaikan jalur, reward ini penting agar agen bisa lebih memilih aksi dari kebijakan yang di nilai oleh reward, lakukanlah Pengujian di Lingkungan Berbeda Uji agen di berbagai jalur atau kondisi lingkungan untuk meningkatkan kemampuan generalisasi agen. Terakhir Perbaikan pada Model RL Pertimbangkan untuk menggunakan algoritma RL yang lebih canggih, seperti Proximal Policy Optimization (PPO) atau Deep Q-Learning (DQN), untuk meningkatkan performa agen.

DAFTAR PUSTAKA

[1] A. Khan, M. Naeem, M. Z. Asghar, A. Ud Din, and A. Khan, “Playing first-person shooter games with machine learning techniques and methods using the VizDoom Game-AI research platform,” *Entertain. Comput.*, vol. 34, May 2020, doi: 10.1016/j.entcom.2020.100357.

[2] N. Özkurt, T. Yıldırım, Yaşar Üniversitesi, Institute of Electrical and Electronics Engineers. Turkey Section., and Institute of Electrical and Electronics Engineers, *2019 Innovations in Intelligent Systems and Applications Conference (ASYU) : proceedings : 31 October-2 November 2019, Izmir, Turkey*.

[3] G. Paczolay and I. Harmati, “A New Advantage Actor-Critic Algorithm for Multi-Agent Environments,” in *2020 23rd IEEE International Symposium on Measurement and Control in Robotics, ISMCR 2020*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020. doi: 10.1109/ISMCR51255.2020.9263738.

[4] G. Gomes, C. A. Vidal, J. B. Cavalcante-Neto, and Y. L. B. Nogueira, “A modeling environment for reinforcement learning in games,” *Entertain. Comput.*, vol. 43, Aug. 2022, doi: 10.1016/j.entcom.2022.100516.

[5] A. A. Bin Ramlan, A. M. Ali, N. H. Abdul Hamid, and R. Osman, "The Implementation of Reinforcement Learning Algorithm for AI Bot in Fighting Video Game," in *Proceedings - ISAMSR 2021: 4th International Symposium on Agents, Multi-Agents Systems and Robotics*, Institute of Electrical and Electronics Engineers Inc., Sep. 2021, pp. 96–100. doi: 10.1109/ISAMSR53229.2021.9567749.
[6] K. Li, Q. S. Jia, and J. Yan, "An Actor-Critic Method for Simulation-Based Optimization," in *IFAC-PapersOnLine*, Elsevier B.V., Jul. 2022, pp. 7–12. doi: 10.1016/j.ifacol.2022.08.040.