

Aprendizado por Reforço Profundo Aplicado ao Jogo DOOM

Comparação de Algoritmos Dueling DQN e PPO em Ambientes 3D Complexos

Graduando: Dayon Victor Czykailo - Sistema de informação 8º Período

Orientadora: Rosanete Grassiani dos Santos

Centro Universitário FACEAR



Resumo da Pesquisa

Desenvolvemos e avaliamos um agente de aprendizado por reforço profundo para o jogo DOOM, usando ViZDoom.

Algoritmos Comparados

Dueling Deep Q-Network (DQN) e Proximal Policy Optimization (PPO).

Objetivo

Analizar eficiência, estabilidade e adaptabilidade em cenários 3D.

Implementação

Python, PyTorch e Stable Baselines3, com monitoramento via TensorBoard.

Cenários

Basic, Defend the Center e Deadly Corridor, com recompensas e dificuldades variadas.

Objetivos da Pesquisa

1

Fundamentos da IA e DRL

Investigar teorias de Deep Reinforcement Learning para agentes autônomos em ambientes dinâmicos.

2

Estudo da Plataforma ViZDoom

Analisar e utilizar o ViZDoom como ambiente de simulação para o desenvolvimento de agentes inteligentes.

3

Implementação de Algoritmos

Aplicar algoritmos DQN e PPO para aprendizado e adaptação contínua do agente.

4

Desenvolvimento do Agente Visual

Criar um agente que percebe o ambiente de jogo visualmente, processando frames e extraíndo características.

5

Avaliação de Desempenho

Comparar DQN e PPO com métricas de jogo como taxa de acerto, sobrevivência e eliminação de inimigos.

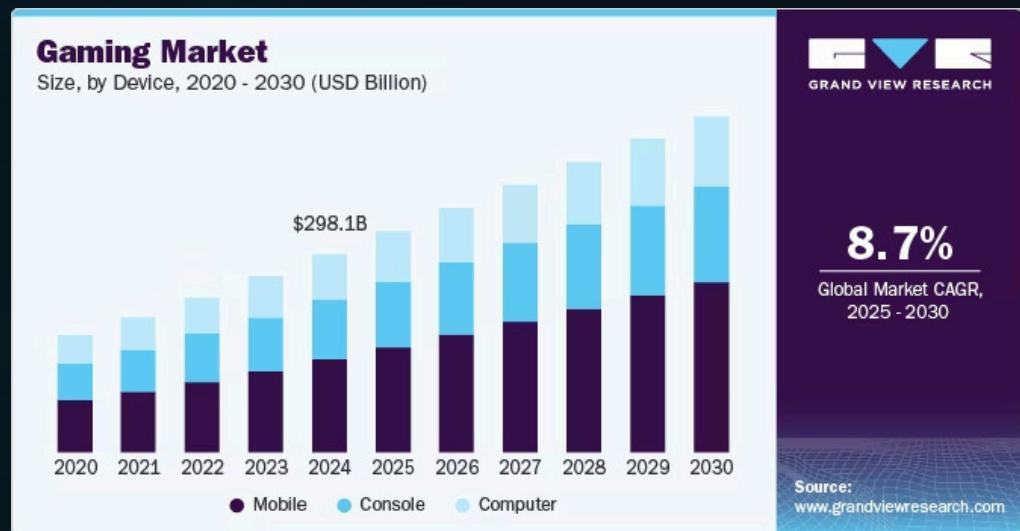
6

Identificação de Desafios e Oportunidades

Analisar limitações, propor melhorias e sugerir pesquisas futuras para agentes em jogos eletrônicos.

IA e o Mercado de Jogos

A IA impulsiona o crescimento da indústria de jogos, criando experiências imersivas e desafiadoras.



Crescimento Acelerado

Mercado de jogos cresce 8.7% ao ano, impulsionado por inovações em IA.

Bots Inteligentes

NPCs mais realistas e adaptáveis, superando habilidades humanas.

Investimento

Grandes empresas investem em IA para personalização e adaptação em tempo real.

O Desafio em Ambientes 3D

Ambientes tridimensionais complexos apresentam desafios significativos para o aprendizado por reforço profundo.



1 Alta Dimensionalidade

Espaço de estados complexo e observações visuais.



2 Recompensas Escassas

Dificuldade em obter feedback claro para o aprendizado.



3 Decisões em Tempo Real

Necessidade de ações rápidas e precisas.



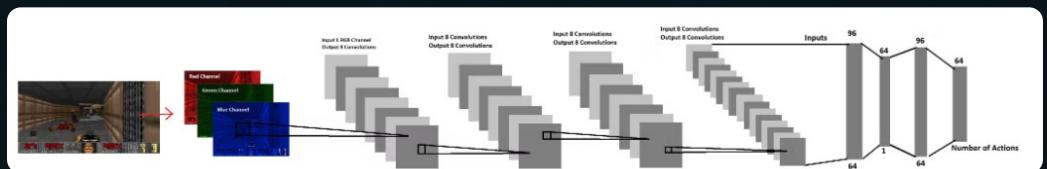
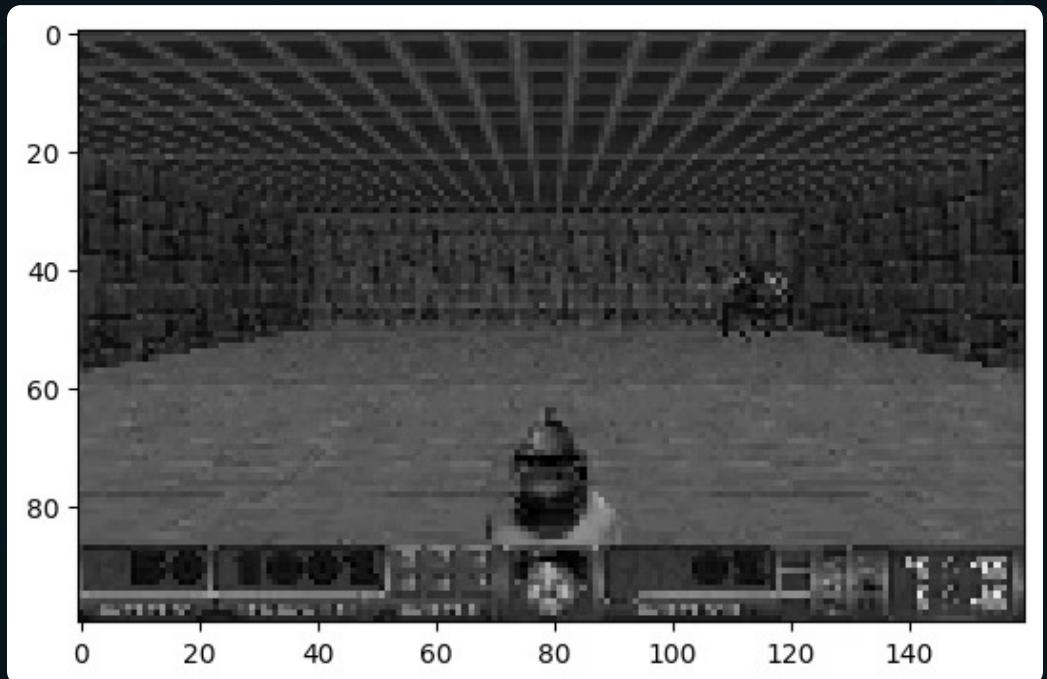
4 Eficiência Amostral

Grande número de interações para convergir.



Pré-processamento de Imagens no ViZDoom

No ViZDoom, o pré-processamento de frames é vital para otimizar o aprendizado. As imagens são redimensionadas (para 80x160 pixels), convertidas para escala de cinza e normalizadas. Isso reduz a carga computacional e foca em informações essenciais. O empilhamento de frames consecutivos (quatro, por exemplo) adiciona uma dimensão temporal, permitindo ao agente inferir movimento e dinâmica do ambiente, tornando a percepção visual mais eficaz para a tomada de decisões.



DQN vs. PPO

A comparação entre DQN e PPO permite avaliar suas diferenças em eficiência, estabilidade e desempenho.

DQN (Deep Q-Network)

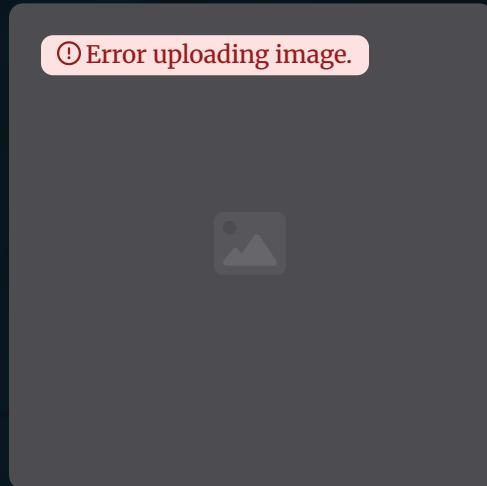
- Técnica baseada em valor (value-based).
- Combina aprendizado por reforço com redes neurais convolucionais.
- Pioneiro em superar desempenho humano em jogos Atari.
- Usa "experience replay" e "target network" para estabilidade.

PPO (Proximal Policy Optimization)

- Método baseado em política (policy-based).
- Melhora a estabilidade de aprendizado via otimização proximal.
- Mais robusto em ambientes dinâmicos e contínuos.
- Menos dependente de grandes buffers de replay.

Metodologia: ViZDoom e Cenários

ViZDoom oferece um ambiente 3D semi-realista para testar agentes de IA em cenários de dificuldade crescente.



BASIC

Eliminar um inimigo estacionário.
Foco em percepção e pontaria.



DEFEND THE CENTER

Sobreviver a ondas de inimigos.
Foco em reação e priorização.



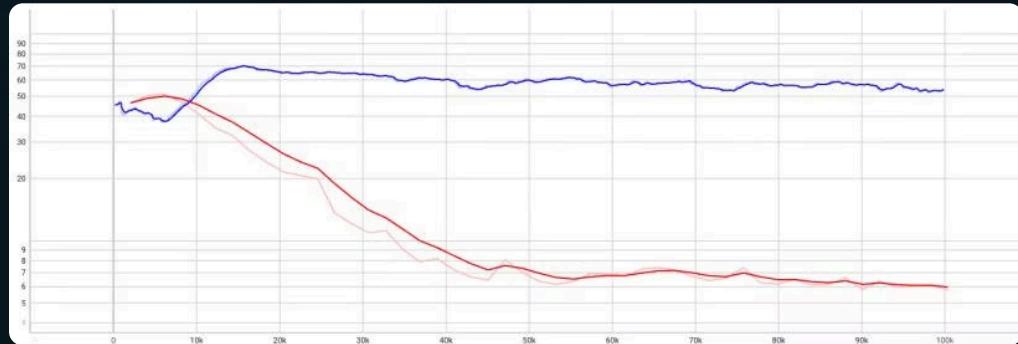
DEADLY CORRIDOR

Atravessar um corredor perigoso.
Foco em planejamento e gerenciamento de recursos.

Resultados: Cenário BASIC

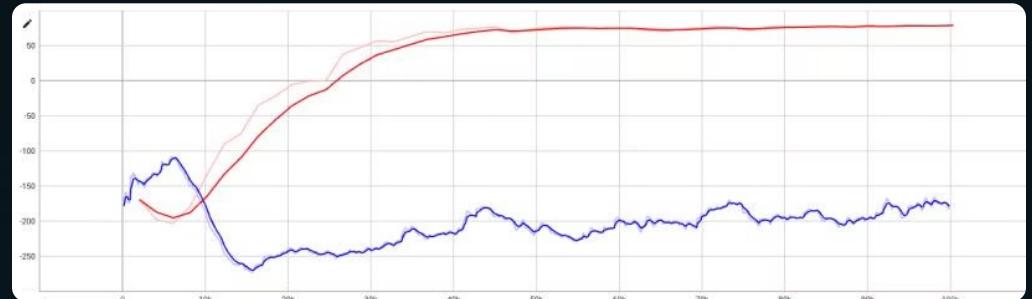
No cenário BASIC, o PPO demonstrou aprendizado superior e mais estável em comparação ao DQN.

Duração Média dos Episódios



PPO reduziu rapidamente a duração, indicando maior eficiência. DQN mostrou variação irregular.

Recompensa Média por Episódio



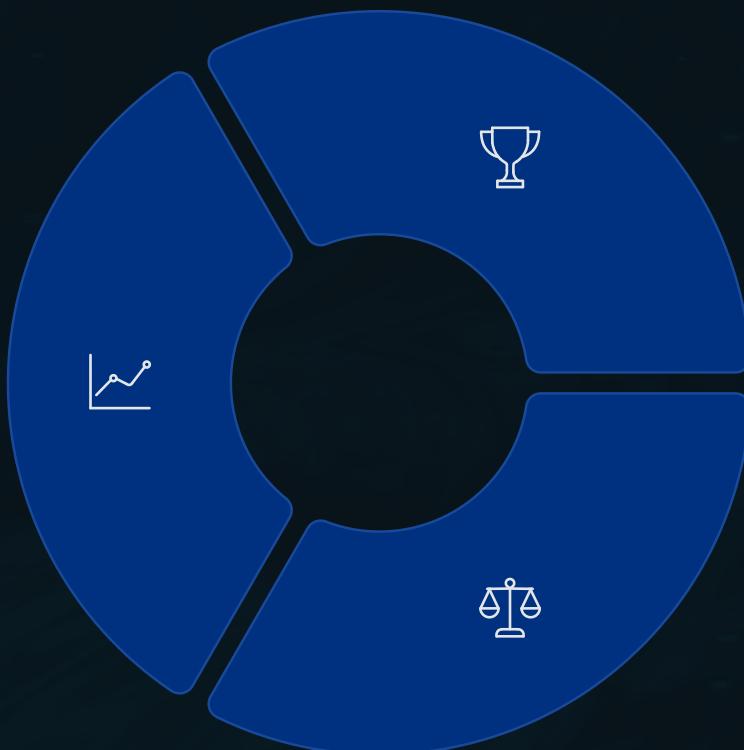
PPO obteve recompensas mais elevadas e consistentes. DQN apresentou oscilações e quedas.

Resultados: Cenário DEFEND THE CENTER

Ambos os algoritmos aprenderam, mas o PPO superou o DQN em estabilidade e qualidade da política a longo prazo.

Duração do Episódio

Ambos cresceram consistentemente, indicando aprendizado de estratégias defensivas.



Recompensa Média

DQN teve vantagem inicial, mas PPO o superou, finalizando com média superior (10,99 vs 10,22).

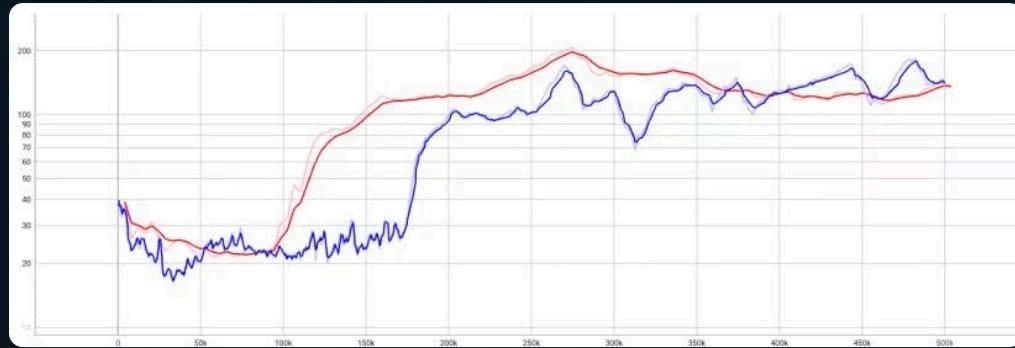
Estabilidade

PPO mais estável, refinando políticas de maior qualidade. DQN rápido, mas com flutuações.

Resultados: Cenário DEADLY CORRIDOR

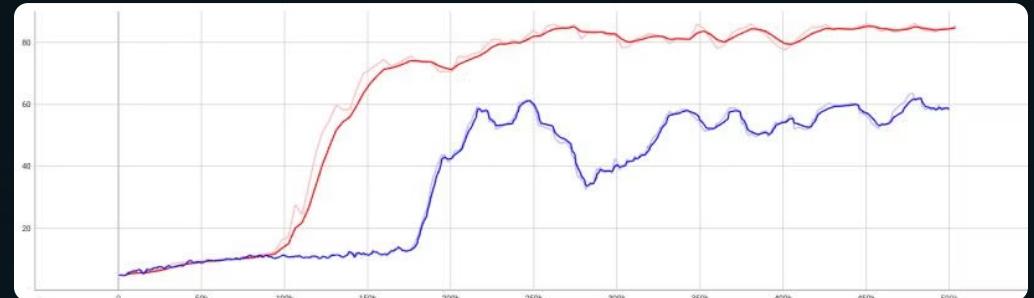
No cenário mais complexo, o PPO manteve desempenho superior e mais estável, com recompensas 20-40% maiores.

Duração Média dos Episódios



Ambos aumentaram a duração, mas PPO mostrou evolução mais regular, indicando aprendizado robusto.

Recompensa Média por Episódio



PPO superou DQN consistentemente, com 20-40% mais recompensas, devido à sua natureza on-policy e atualizações refinadas.

Conclusão e Perspectivas Futuras

O PPO se mostrou mais robusto e eficiente em ambientes complexos, enquanto o DQN oferece aprendizado inicial mais rápido.

PPO: Estabilidade e Eficiência

Ideal para ambientes complexos, com múltiplas variáveis de decisão.

DQN: Velocidade Inicial

Eficiente para aprendizado rápido em estágios iniciais, mas menos estável.

Futuro

Explorar arquiteturas híbridas e técnicas de aprendizado hierárquico para otimizar o desempenho.

Referências e Agradecimentos

Gostaríamos de expressar nossa profunda gratidão à equipe de pesquisa, aos desenvolvedores de ViZDoom e à comunidade de aprendizado por reforço pelo suporte e recursos valiosos. Este trabalho foi fundamental para explorar os limites da IA em ambientes complexos.

Referências Principais

- Wang, Z., et al. (2016). Dueling Network Architectures for Deep Reinforcement Learning.
- Schulman, J., et al. (2017). Proximal Policy Optimization Algorithms.
- Kempka, M., et al. (2016). ViZDoom: A Doom-based AI Research Platform for Reinforcement Learning.
- Mnih, V., et al. (2015). Human-level control through deep reinforcement learning.

