

Aprendizado por Reforço Profundo para Navegação sem Mapa de um Veículo Híbrido Aéreo-Aquático

Ricardo B. Grando¹, Paulo L. J. Drews-Jr¹

¹Grupo de Automação e Robótica Inteligente – Universidade Federal do Rio Grande (FURG)

{ricardo.grando, paulodrews}@furg.br

Abstract. *Performing navigation between environments is a challenging task for hybrid mobile robots, especially in scenarios with obstacles. This work presents an approach based on deep reinforcement learning (Deep-RL) for autonomous navigation of a specific type of hybrid mobile robot: a Hybrid Unmanned Aerial Underwater Vehicle (HUAUV). Our approach uses only information from range findings sensors and the vehicle's relative localization data to perform the navigation. Results show that it is possible to perform mapless navigation from start to finish without any manual operation, only using Deep-RL-based agents. The navigation performed by the agents is compared with the mapless navigation performed by a classic algorithm - BUG2. We based our approach in two state-of-art Deep-RL techniques for mapless navigation of terrestrial robots: Deep Deterministic Policy Gradient (DDPG) and Soft Actor Critic (SAC).*

Resumo. *Realizar a navegação entre meios é uma tarefa desafiadora para robôs móveis híbridos, especialmente em cenários com obstáculos. Este trabalho apresenta uma abordagem baseada em aprendizado por reforço profundo (Deep-RL) para navegação autônoma de um tipo específico de robô móvel híbrido: um Veículo Híbrido Tipo Ar-Água (HUAUV). A abordagem proposta utiliza somente informação de sensores de distância e de informações relativas à localização do veículo para realizar a navegação. Resultados obtidos mostram que é possível realizar navegação sem mapa do início ao fim sem usar nenhum tipo de operação manual, somente utilizando os agentes baseados em Deep-RL. A navegação dos agentes treinados é comparada com a navegação sem mapa realizada por um algoritmo clássico - BUG2. A abordagem é baseada em dois métodos do estado da arte para navegação sem mapa de robôs terrestres: Política de Gradiente Determinístico Profundo (DDPG) e Soft Actor-Critic (SAC).*

Dissertação de mestrado ¹ apresentada ao Programa de Pós Graduação em Engenharia de Computação (PPGComp) da Universidade Federal de Rio Grande (FURG) publicada em maio de 2021 sobre a orientação do professor Dr. Paulo Drews-Jr. Artigo submetido ao Concurso de Teses e Dissertações em Robótica (CTDR-2021).

1. Introdução

Nos últimos anos tem aumentado o interesse no desenvolvimento de veículos híbridos não tripulados HUVs (do inglês, *Hybrid Unmanned Vehicles*). Esses veículos buscam unir e

¹<https://argo.furg.br/?BDTD12911>

explorar as vantagens e desvantagens dos veículos aéreos, terrestres e subaquáticos. O estudo acerca de um veículo híbrido capaz de atuar no meio aéreo e subaquático - HUAUV (do inglês, *Hybrid Unmanned Aerial Underwater Vehicle*) - tem despontado como um grande desafio na ciência devido ao contraste de atuação e percepção que os meios envolvidos demandam, onde muitos dos sensores e atuadores somente se aplicam a um determinado meio ou requerem adaptações mecânicas para um funcionamento correto.

Vários modelos de HUAUVs já foram propostos nas pesquisas de [Drews et al. 2014] e [Maia et al. 2015], mostrando um aumento do interesse por esse tipo de veículo na última década com foco por um HUAUV com uma estrutura do tipo quadrotor. Um veículo híbrido com esse tipo de formato propicia a realização de variadas tarefas, tais como inspeção e manutenção, mapeamento e coleta de dados, e também, vigilância, busca e resgate.

Esse tipo de veículo, entretanto, ainda se encontra em fase de conceito e prototipação, onde vários aspectos nas partes mecânica, *hardware* e *software* ainda precisam ser definidos e/ou abordados. Dentre os problemas mais primitivos e que talvez se detenha com maior foco no escopo de *software* e *hardware* é o conceito de navegação para esse tipo de veículo. Como esse veículo navegaria de uma posição para outra no espaço, que tipo de inteligência e a informação de quais sensores que seriam utilizados para tal uma vez que a estrutura do veículo estivesse mecanicamente estável são questões a serem abordadas e discutidas.

O presente trabalho busca, especificamente, realizar uma discussão a partir do escopo do problema primitivo da navegação para HUAUVs e a inteligência necessária para resolvê-lo. Parte-se de uma abordagem de Inteligência Artificial baseada em Aprendizado por Reforço Profundo ou Deep-RL (do inglês Deep Reinforcement Learning) [Li 2017] aplicado a tarefas que envolvem navegação sem mapa orientada a alvo de um HUAUV simulado a partir de um veículo real nos ambientes aéreo e subaquático.

Até o momento, o presente trabalho possui duas contribuições científicas. A primeira foi realizada no Simpósio Latino Americano de Robótica 2020 (IEEE LARS 2020) [Grando et al. 2020], trabalho que envolveu a parte de navegação aérea 2D e que ficou entre os 10 melhores do evento. Outro artigo científico foi publicado na Conferência Internacional de Robótica e Automação (IEEE ICRA 2021, Qualis A1) [Grando et al. 2021], envolvendo a navegação híbrida. Atualmente, os autores deste trabalho vem desenvolvendo uma versão estendida do artigo do LARS 2020 para o *Journal of Robotics & Intelligent Systems* (JINT) e também uma versão estendida do ICRA 2021 para o *IEEE Robotics and Automation Letters*.

2. Trabalhos Relacionados

Drews et al. [Drews et al. 2014] destaca no seu trabalho um histórico acerca de HUAUVs, resumindo as primeiras ideias e protótipos de veículos até a data na qual o trabalho foi publicado. Além desse histórico, Drews et al. [Drews et al. 2014] também realiza um estudo comparativo entre os tipos de veículos aéreos e subaquáticos já propostos e/ou desenvolvidos, dando ênfase às características que beneficiam e prejudicam cada modelo. Esse trabalho é um dos primeiros acerca de HUAUVs. Atualmente, ainda existem poucos trabalhos desenvolvidos sobre HUAUVs, sendo que a maioria foca na mecânica e estrutura do veículo, havendo, dessa forma, pouco progresso no que diz respeito ao *soft-*

ware e inteligência para o veículo e ao problema da transição de meio [Ravell et al. 2018] [Maia et al. 2017] [Mercado et al. 2019]. Trabalhos relacionados a navegação sem mapa para robôs terrestres, aéreos e subaquáticos são mais comuns. Os trabalhos mais relevantes acerca de HUAUVs partem de uma abordagem utilizando uma estrutura de quadrotor [Drews et al. 2014], [Ma et al. 2018], [Maia et al. 2015], [Mercado et al. 2019]. O HUAUV descrito em simulação neste trabalho levou em consideração essa convergência de propostas, sendo baseado também em uma estrutura de quadrotor.

No contexto de navegação sem mapa em geral, Tai et al. [Tai et al. 2017] destaca-se pela utilização de um robô móvel terrestre e Deep-RL para realizar navegação sem mapa em ambientes com obstáculos. O trabalho desenvolvido por Tai et al. [Tai et al. 2017] serviu de base para o desenvolvimento deste trabalho. O problema da navegação sem mapa para UAVs e UUVs já foi abordado em vários trabalhos. Rodriguez et al. [Rodriguez-Ramos et al. 2018] e Carlucho et al. [Carlucho et al. 2018] são os principais exemplos para cada um dos respectivos veículos.

Este trabalho diferencia-se dos demais trabalhos discutidos por buscar abordar o problema da navegação sem mapa para HUAUV, mesmo que sendo em simulação. Além disso, a capacidade de evitar colisão com obstáculos também é analisada levando em conta somente informação de sensores de distância descritos a partir de sensores de distância simulados. Duas abordagens baseadas em técnicas de Deep-RL do estado da arte (DDPG e SAC) são abordadas e analisadas, diferentemente da maioria dos trabalhos que abordam navegação sem mapa utilizando somente um tipo de abordagem. Além disso, também é apresentado um estudo de validação para analisar com mais robustez as propostas apresentadas e comparar com propostas relacionadas, estudo não visto em outros trabalhos. Por último, este trabalho também é o primeiro a abordar o problema da transição do meio autônomo para HUAUVs.

3. Metodologia

Primeiramente o veículo e os ambientes de simulação foram descritos. Após isso, os agentes foram estruturados e desenvolvidos. Para abordagem baseada em DDPG foi dada a nomenclatura NDRL-D do inglês (Navigation Deep Reinforcement Learning Deterministic), enquanto que para a abordagem baseada em SAC foi dado a nomenclatura NDRL-S do inglês (Navigation Deep Reinforcement Learning Stochastic)

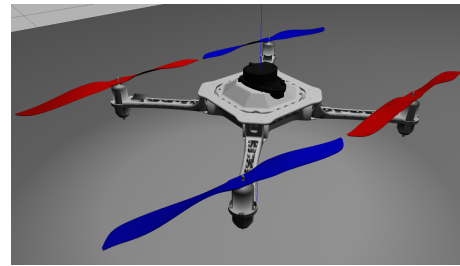
3.1. Desenvolvimento do Veículo simulado

O veículo descrito no Gazebo foi baseado em um veículo real chamado Hydrone [Horn et al. 2019]. Na Figura 1 pode-se observar o veículo real e o veículo final descrito em simulação.

Um sensor de distância do tipo Lidar baseado no plugin *ray* do Gazebo foi descrito para uso no ar. O sensor possui resolução angular de $0,25^\circ$, mostrando 1080 leituras de distância em um alcance de 270° . Entretanto, somente 20 amostras de distância igualmente espaçadas em $13,5^\circ$ foram utilizadas como entradas para a rede dos agentes de Deep-RL. Apesar de gerar uma perda de resolução na detecção do ambiente, essa simplificação permite que o agente tenha menos estados e, portanto, consiga convergir na detecção de objetos mais rapidamente, além de gerar um menor tamanho de rede que implica em um menor demanda de capacidade de processamento e tempo médio de treino por episódio.



(a) Veículo real.



(b) Veículo descrito.

Figura 1. Veículo simulado baseado no veículo real Hydroner.

Um sensor de distância para o ambiente subaquático também foi acoplado ao veículo. Foi utilizado um sonar do tipo FLS baseado na simulação de [Cerqueira et al. 2016]. 20 amostras igualmente espaçadas entre os *beams* do sonar foram usadas como entrada para o estado dos agentes, seguindo a simplificação do lidar.

3.2. Desenvolvimento dos Ambientes de simulação

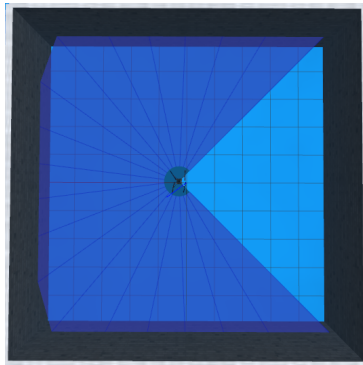
Após realizar a descrição do veículo, foram desenvolvidos ambientes de simulação no Gazebo a fim de treinar e testar as abordagens de Deep-RL propostas. A estrutura básica do mundo utilizada é composta por um oceano na parte negativa do eixo Z , onde ocorre o efeito de ambos os frameworks. Na parte positiva do eixo Z somente o framework de simulação aérea RotorS atua, simulando o ambiente aéreo. O oceano possui corrente de água e no ambiente aéreo foi utilizado a influência de ventos.

A partir dessa estrutura básica, dois cenários diferentes foram descritos utilizando modelos disponíveis no Gazebo. No primeiro cenário, um espaço com dimensões de $5 \times 5 \times 6$ metros foi descrito fazendo-se do uso do modelo *grey-wall*. Os modelos foram instanciados de forma estática de modo a criar um cenário que varia de -1 metros a 5 metros no eixo de Z , centralizado na origem para os eixos X e Y . Dessa forma o cenário representa um tanque com 1 metro de coluna de água e 5 metros de coluna de ar.

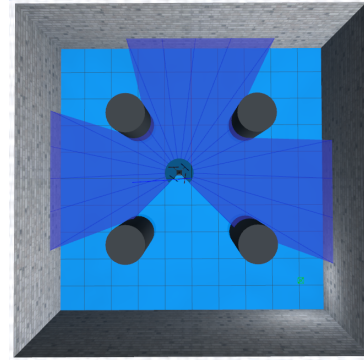
Um segundo cenário, baseado no primeiro, também foi descrito. Nesse segundo cenário, quatro cilindros com 0,6 metros de raio e 3 metros de comprimento foram adicionados à estrutura do primeiro cenário. Os quatro objetos também foram instanciados de forma estática no cenário e foram colocados em posições no cenário que os tornam possíveis obstáculos, representando dutos de perfuração de petróleo. Nesse cenário busca-se que os agentes de Deep-RL aprendam a desviar dos objetos a fim de chegar na posição de destino levando em conta um cenário mais realístico onde um veículo híbrido poderia atuar. Na Figura 2 é possível observar ambos os cenários.

3.3. Estrutura dos agentes

Além de abordar o problema da navegação sem mapa, o presente trabalho também busca desenvolver um modelo de estrutura para os agentes o mais idêntica possível, buscando evitar discrepâncias muito grandes entre cada abordagem determinística e estocástica. A Figura 3 mostra a estrutura para o contexto 3D e híbrido.



(a) Cenário 1 visto de cima.



(b) Cenário 2 visto de cima.

Figura 2. Cenários de simulação.

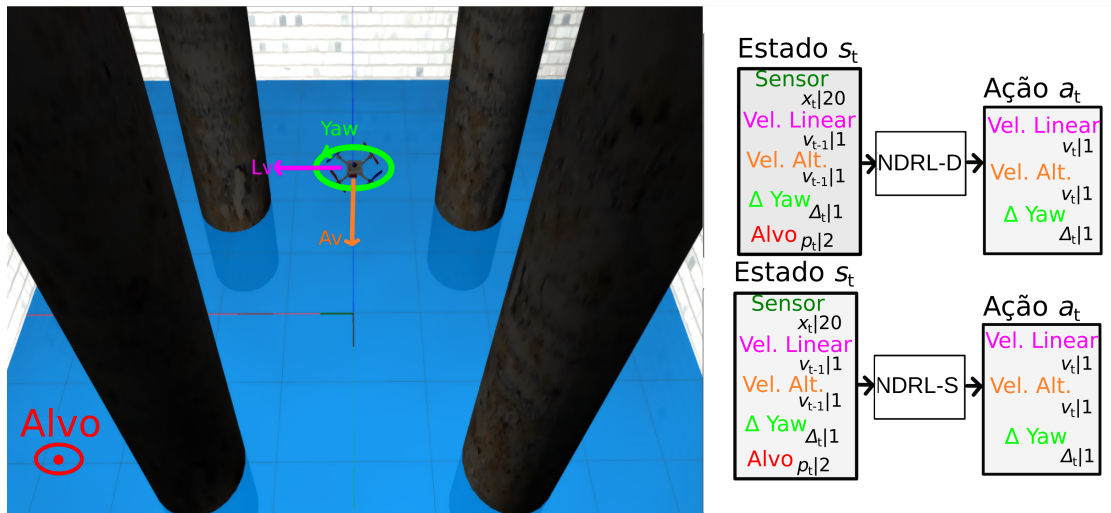


Figura 3. Veículo treinando em um dos cenários (esquerda) e o modelo básico de estrutura com entradas e saídas utilizando NDRL (direita) para o contexto 3D.

A rede do agente proposta possui ao total 26 entradas e 3 saídas. Dos 26 valores de entrada, 20 são os valores de leitura do sensor de distância simulado, 3 valores representam a velocidade linear, de altitude e a variação do ângulo de *yaw* da última ação e os 3 últimos valores representam a posição relativa do veículo, o ângulo em relação à posição alvo no plano $x - y$ e um ângulo relativo entre os planos Z e distância. O ângulo e a distância relativos foram utilizados para forçar o agente a aprender a minimizá-los. As saídas da rede representam a velocidade linear e a variação do ângulo de *yaw* a ser aplicada no veículo.

3.4. Estrutura da rede

A estrutura de rede entre cada abordagem também foi projetada para ser o mais semelhante possível e foi baseada no trabalho do [Tai et al. 2017]. Busca-se analisar as diferenças que cada abordagem de Deep-RL em si proporciona à problemática proposta. A estrutura para a rede para ambas as abordagens no contexto 3D pode ser observada na Figura 4. Para a rede do ator foram utilizadas 3 camadas totalmente conectadas de 512

neurônios cada, ativadas usando a função de ativação ReLU. A saída da terceira camada é conectada a um neurônio que representa a velocidade linear e a outro neurônio que representa a variação do ângulo de yaw δ_{yaw} . O tipo de rede usada tanto para o ator quanto para o crítico é do tipo MLP (do inglês *Multi Layer Perceptron*).

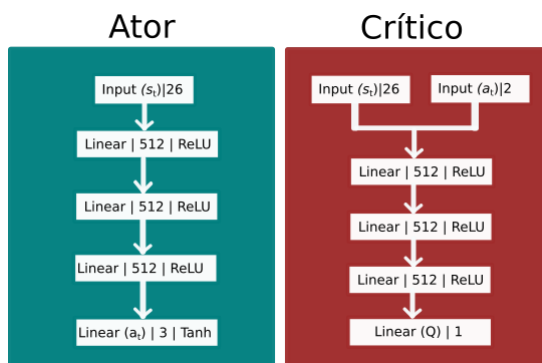


Figura 4. Estrutura da rede para o contexto 3D.

A função de ativação de tangente hiperbólica (*Tanh*) foi utilizada nos neurônios de saída da rede. O alcance da saída da ativação Tanh que varia entre -1 e 1 no primeiro neurônio é escalonado para 0 e 0,25, representando a velocidade linear mínima e máxima a ser aplicado no veículo em m/s respectivamente. A saída do segundo neurônio é escalonado entre $-0,25$ e $0,25$, representando a variação angular mínima e máxima a ser aplicado no veículo em $radianos$ respectivamente. No contexto 3D, a terceira saída que representa a velocidade de altitude é escalonada entre $-0,25$ e $0,25$ m/s . Os valores da saída são concatenados formando a ação que é enviada ao veículo. A rede crítica, por sua vez, também possui 3 camadas completamente conectadas de 512 neurônios e ativadas usando ReLU. A entrada da ação a_t é concatenada com as entradas que representam o estado do agente. A estrutura para o contexto 2D é semelhante, mas sem a velocidade de altitude e a informação relativa do ângulo no plano z – *distancia*.

4. Função de Reforço

Independentemente da abordagem de Deep-RL a ser utilizada e do contexto, é necessário definir um sistema de recompensa e penalidade ou um sistema de reforço. Um reforço positivo $r_{chegada}$ é dado caso o agente alcance o alvo dentro de uma margem de c_d metros. Um reforço negativo $r_{colisao}$ é dado para o episódio caso o veículo colida com a parede ou algum dos obstáculos do cenário. A verificação é feita baseada na distância mínima da leitura do sensor de distância. Caso a distância seja menor que uma distância de c_o metros, uma colisão é detectada. Os valores para os hiperparâmetros utilizados podem ser observados foram: $r_{chegada}$ 100, c_d 0,5m, $r_{colisao}$ -10, c_o 0,6m.

5. Resultados

Dentre diversos resultados validados, destaca-se o resultado da navegação híbrida. No primeiro cenário o treinamento foi realizado por 1000 episódios, enquanto que no segundo cenário o agente evoluiu por um total de 5000 episódios. Não houve um critério específico para a escolha do limite de episódios de treinamento, apenas a análise empírica do número aproximado episódios que os agentes foram capazes de realizar a navegação sem mapa.

Tabela 1. Estatísticas da navegação orientada a alvo para o contexto 3D híbrido.

Cenário	Abordagem	Tempo Médio Ar (s)	Tempo Médio Água (s)	Total Sucesso
1	Ar-água-NDRL-D	15.29 ± 2.40	33.05 ± 10.91	96 %
1	Ar-água-NDRL-S	39.81 ± 23.16	23.64 ± 19.63	72 %
1	Ar-água-BUG2	31.963 ± 1.65	21.05 ± 0.199	100 %
1	Água-Ar-NDRL-D	18.78 ± 1.21	6.10 ± 0.17	97 %
1	Água-Ar-NDRL-S	33.98 ± 26.15	17.36 ± 12.87	75 %
1	Água-Ar-BUG2	32.74 ± 3.33	3.86 ± 0.28	100 %
2	Ar-água-NDRL-D	20.06 ± 13.92	59.83 ± 22.49	54 %
2	Ar-água-NDRL-S	60.88 ± 30.25	17.38 ± 16.10	37 %
2	Ar-água-BUG2	48.94 ± 28.03	12.55 ± 8.26	47 %
2	Água-Ar-NDRL-D	53.56 ± 31.28	5.98 ± 1.31	57 %
2	Água-Ar-NDRL-S	29.98 ± 14.48	6.61 ± 0.82	71 %
2	Água-Ar-BUG2	117.266 ± 79.54	3.775 ± 0.31	32 %

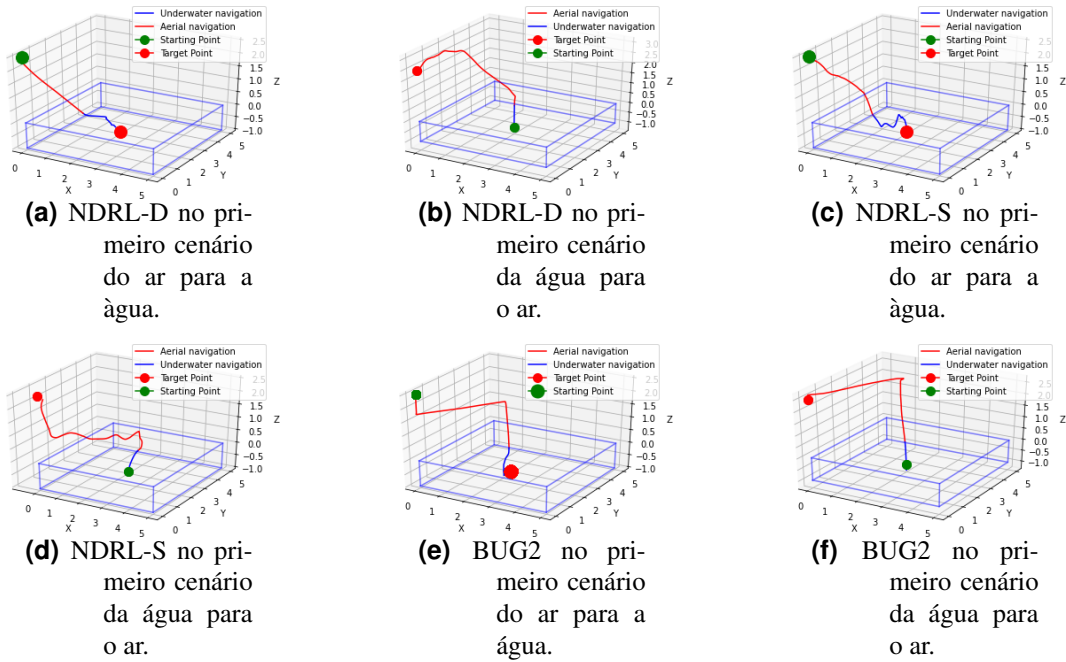
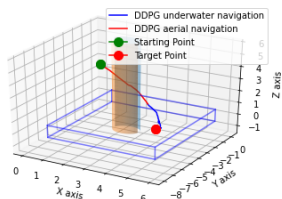
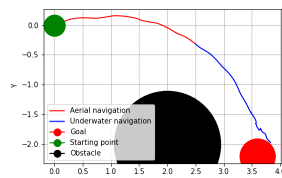


Figura 5. Caminho realizado durante 1 das 100 tentativas no contexto 3D híbrido no primeiro cenário.

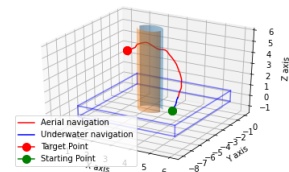
O desempenho dos agentes em cada tarefa está apresentado na sequência. É importante ressaltar que somente a primeira tarefa foi realizada para o contexto híbrido dada a maior complexidade na navegação entre meios diferentes. Entretanto, essa tarefa foi realizada partindo de pontos iniciais distintos (um aéreo e um subaquático), buscando analisar com maior robustez a transição de meio. Para o primeiro cenário, a posição aérea de partida foi definida como sendo a coordenada (0,0; 0,0; 2,5) no mundo do Gazebo, enquanto que a posição subaquática de partida foi definida como sendo (2,0; 3,0; -1,0). A



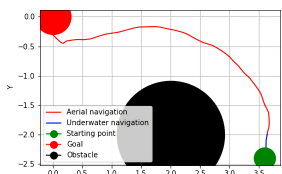
(a) NDRL-D no segundo cenário do ar para a água.



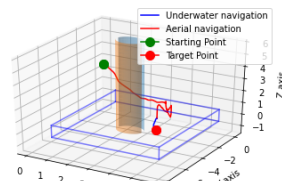
(b) NDRL-D no segundo cenário do ar para a água visto de cima.



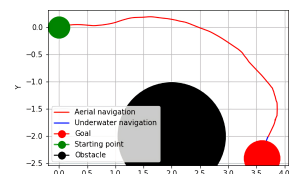
(c) NDRL-D no segundo cenário da água para o ar.



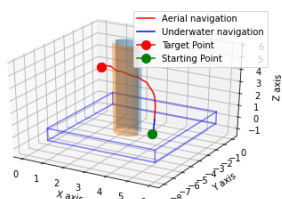
(d) NDRL-D no segundo cenário da água para o ar visto de cima.



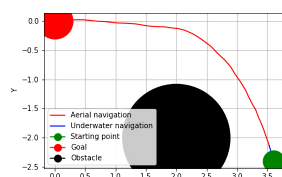
(e) NDRL-S no segundo cenário do ar para a água.



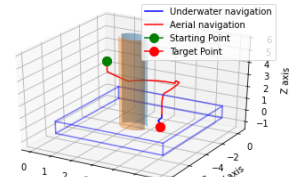
(f) NDRL-S no segundo cenário do ar para a água visto de cima.



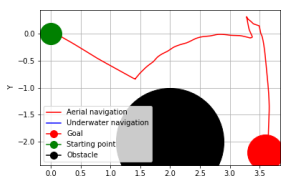
(g) NDRL-S no segundo cenário da água para o ar.



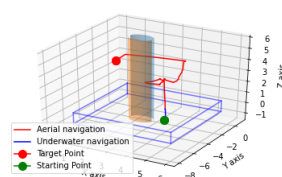
(h) NDRL-S no segundo cenário da água para o ar visto de cima.



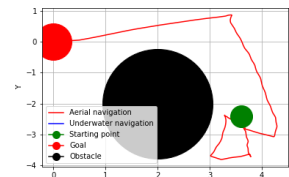
(i) BUG2 no segundo cenário do ar para a água.



(j) BUG2 no segundo cenário do ar para a água visto de cima.



(k) BUG2 no segundo cenário da água para o ar.



(l) BUG2 no segundo cenário da água para o ar visto de cima.

Figura 6. Caminho realizado durante 1 das 100 tentativas no contexto 3D híbrido no segundo cenário.

posição alvo para o ponto de partida aéreo é a posição de partida subaquática, enquanto que o inverso ocorre para o ponto de partida subaquático. Dessa forma, em ambos os testes o veículo é obrigado a sair de um meio e transitar para o alvo que está no outro meio. Para o segundo cenário o mesmo foi definido, mudando-se somente as posições nas quais

foram definidas como sendo $(0,0; 0,0; 2,5)$ e $(3,6; -2,4; -1,0)$, respectivamente. Busca-se também avaliar a capacidade de evitar colisão com obstáculos nesse cenário.

Na Tabela 1 pode-se observar o total de tentativas em que a navegação foi realizada com sucesso para cada abordagem em cada cenário. O tempo médio e o desvio padrão de realização da tarefa também são amostrados. Na Figura 5 é possível observar 1 das 100 tentativas de navegação realizada por ambos agentes e pelo algoritmo BUG2, no primeiro cenário. Figura 6 mostra o mesmo para o segundo cenário. Destaca-se a robustez dos agentes ao serem capazes de navegar do ar para a água e vice-versa com pequena variação no desempenho. Destaca-se também o total de sucesso das abordagens de Deep-RL quando comparadas com o BUG2 no segundo cenário. Assim como no contexto 2D, ambas as abordagens foram mais eficazes em média no desvio dos obstáculos para chegar ao alvo em um ambiente diferente do ponto de partida. Nesse contexto fica ainda mais claro as características de cada abordagem no que diz respeito às ações a cada passo. Pode-se observar novamente as características determinísticas e estocásticas e a semelhança no espaço de ações final.

6. Estudo de Validação

A fim de analisar com maior detalhes as abordagens de Deep-RL desenvolvidas neste trabalho, foi proposto um estudo de validação de ambas as abordagens com variadas configurações de rede. Arquiteturas de rede com duas, três, quatro e cinco camadas escondidas foram criadas, além de duas estruturas com LSTM e CNNs. De modo geral, a extensiva validação dos variados modelos criados e testados mostra que ambos os agentes são flexíveis com relação ao tipo de ANNs utilizada. Pode-se concluir que a abordagem determinística possui um melhor desempenho em um cenário sem obstáculos, enquanto que o contrário ocorre com a abordagem estocástica.

Pode-se observar que quanto maior e mais complexa a rede, maior tende a ser o reforço médio, como por exemplo para os modelos com LSTM e CNN. É importante ressaltar que isso se deve ao fato de um número maior de navegações ser realizado a cada episódio e não que os modelos são melhores ou piores. Todos os modelos com reforço médio superior a 100 podem ser considerados funcional. O número maior de navegações é devido ao passo mais lento com redes mais complexas, permitindo que a tarefa seja concluída e otimizada com um número menor de ações. Na Figura 8 é possível observar a comparação do tempo médio para a realização da primeira no contexto 2D, contexto onde as abordagens apresentaram resultados médio próximos ao máximo possível para todas as estruturas.

É interessante observar na Figura 8 como elas apresentam as características de cada abordagem com mais detalhes. É possível observar que a abordagem estocástica possui um tempo médio semelhante entre as estruturas, enquanto que a abordagem determinística varia com maior intensidade. Isso é devido a maior capacidade de generalização que o método com viés estocástico como a SAC proporciona, ao mesmo tempo que a abordagem determinística pode ser muito boa para estruturas específicas. De modo geral, com relação ao tempo é possível concluir que a abordagem estocástica tende a ser em média um pouco maior e mais previsível, enquanto que o oposto ocorre com abordagens determinísticas.

Na Figura 8 pode-se observar também a distância média para a segunda tarefa

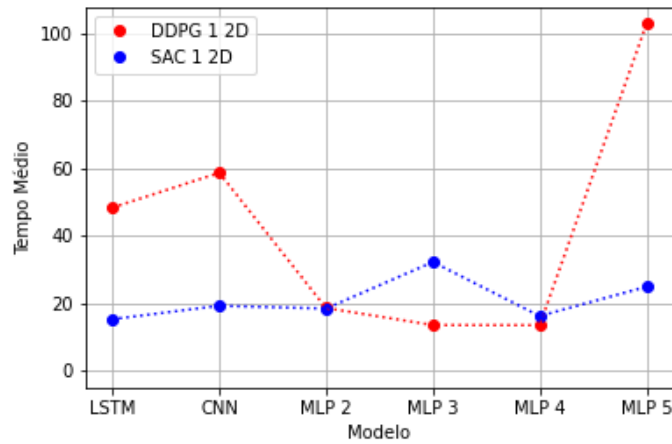


Figura 7. Comparação do tempo médio para o contexto 2D no primeiro cenário (mais estável) na primeira tarefa.

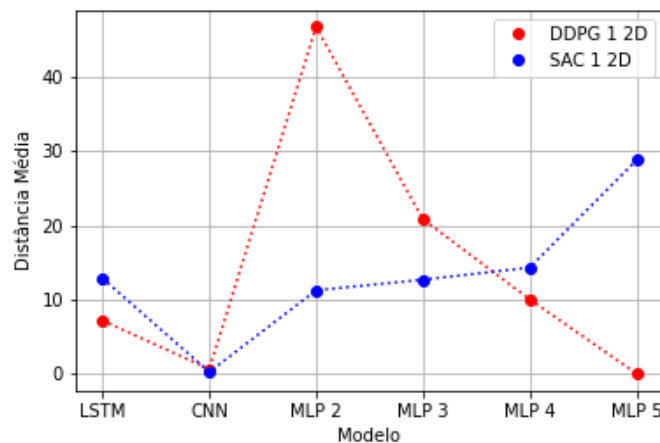


Figura 8. Comparação da distância média para o contexto 2D no primeiro cenário (mais estável) na segunda tarefa.

no segundo cenário, também no contexto 2D por ser mais de modo geral estável entre as estruturas. Dessa ilustração é interessante observar ainda com mais detalhes as características de cada abordagem. Pode-se observar como a abordagem determinística possui um melhor desempenho com 2 camadas e como o desempenho cai de forma exponencial com o aumento da complexidade da rede. Enquanto isso, a abordagem estocástica apresenta melhores resultados com estruturas de rede mais complexas, aumentando o desempenho conforme o aumento do número de camadas, por exemplo. Isso se deve a capacidade de generalização e criação de maiores gradientes que o método baseado em com viés estocástico SAC possui. De modo geral, pode-se concluir que quanto maior a rede melhor tende a ser o desempenho de agentes estocásticos, enquanto que o oposto ocorre as abordagens determinísticas.

O limite para a complexidade, entretanto, parece ser próximo ao modelo convolucional proposto. Como pode-se observar na Figura 8 e também nos resultados para o

contexto 3D, ambas as abordagens com CNN não conseguiram aprender a realizar as tarefas. A solução para isso pode ser redes contrastivas [Srinivas et al. 2020]. A utilização de redes contrastivas com Deep-RL pode ser um caminho não só para resolver esse problema com CNNs, mas também para otimizar a problemática do trabalho como um todo.

7. Considerações Finais e Perspectivas

Neste trabalho, foi proposto abordagens baseadas em Deep-RL para realizar navegação sem mapa de um veículo híbrido capaz de atuar no ar e na água e realizar a transição de meio. Um HUAUV com estrutura de quadrotor baseado em um veículo real foi descrito utilizando *plugins* de simulação aérea e subaquática disponíveis para simulação de robôs. Duas abordagens determinísticas e estocásticas foram desenvolvidas e aplicadas a duas tarefas em variados ambientes e cenários no contexto 2D, 3D e híbrido e comparadas com o algoritmo BUG2.

Com os resultados obtidos, pode-se verificar que abordagens baseadas em Deep-RL para navegação sem mapa, tradicionalmente usadas em robôs móveis terrestres, podem ser usadas para fazer a inteligência de navegação para HUAUVs. A estrutura de agente e rede proposta, utilizando somente informação de um sensor de distância e a informação da localização do veículo, mostrou ter potencial para realizar navegação sem mapa orientada a alvo e navegação sem mapa orientada a múltiplos alvos, realizando com sucesso a transição de meio em ambas as tarefas. Sem utilizar uma abordagem baseada em informação visual, que demanda um *hardware* com alta capacidade de processamento que é geralmente custoso em peso e difícil de ser embarcado em HUAUVs de pequeno porte, tal como o HUAUV Hydrone, a abordagem simples e com poucos estados proposta mostrou-se ser capaz realizar a navegação do veículo e fazê-lo ser inteligente o suficiente para evitar colisões com objetos em um contexto 2D. Apesar de promissores, os resultados obtidos até o momento possuem bastante espaço para serem aprimorados como mostrado com mais detalhes no estudo de validação.

Para trabalhos futuros propõe-se várias perspectivas. Primeiro propõe-se uma modelagem subaquática mais completa pois como mostrados nos resultados da parte subaquática do trabalho há muito espaço para isso ser aprimorado. Isso, entretanto, depende de um estudo futuro mais aprofundado. Além disso, propõe-se a criação e testes em ambientes mais complexos e com obstáculos dinâmicos, o que ainda não foi possível. Por último, propõe-se a utilização de abordagens que utilizem uma combinação de estruturas de rede com LSTM e com CNN, mas com aprendizado contrastivo. Essas parecem ser duas vertentes interessantes no futuro próximo para o problema abordado neste trabalho.

Referências

- Carlucho, I., De Paula, M., Wang, S., Menna, B. V., Petillot, Y. R., and Acosta, G. G. (2018). Auv position tracking control using end-to-end deep reinforcement learning. In *OCEANS 2018 MTS/IEEE Charleston*, pages 1–8. IEEE.
- Carqueira, R., Trocoli, T., Neves, G., Oliveira, L., Joyeux, S., Albiez, J., and Center, R. I. (2016). Custom shader and 3d rendering for computationally efficient sonar simulation. In *29th Conference on Graphics, Patterns and Images-SIBGRAPI*.

- Drews, P. L., Neto, A. A., and Campos, M. F. (2014). Hybrid unmanned aerial underwater vehicle: Modeling and simulation. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4637–4642. IEEE.
- Grando, R. B., de Jesus, J. C., and Drews-Jr, P. L. (2020). Deep reinforcement learning for mapless navigation of unmanned aerial vehicles. In *2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE)*, pages 1–6. IEEE.
- Grando, R. B., de Jesus, J. C., Kich, V. A., Kolling, A. H., Bortoluzzi, N. P., Pinheiro, P. M., Neto, A. A., and Drews Jr, P. L. J. (2021). Deep reinforcement learning for mapless navigation of a hybrid aerial underwater vehicle with medium transition. *arXiv preprint arXiv:2103.12883*.
- Horn, A. C., Pinheiro, P. M., Silva, C. B., Neto, A. A., and Drews-Jr, P. L. (2019). A study on configuration of propellers for multirotor-like hybrid aerial-aquatic vehicles. In *2019 19th International Conference on Advanced Robotics (ICAR)*, pages 173–178. IEEE.
- Li, Y. (2017). Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274*.
- Ma, Z., Feng, J., and Yang, J. (2018). Research on vertical air–water trans-media control of hybrid unmanned aerial underwater vehicles based on adaptive sliding mode dynamical surface control. *International Journal of Advanced Robotic Systems*, 15(2):1729881418770531.
- Maia, M. M., Mercado, D. A., and Diez, F. J. (2017). Design and implementation of multirotor aerial-underwater vehicles with experimental results. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 961–966. IEEE.
- Maia, M. M., Soni, P., and Diez, F. J. (2015). Demonstration of an aerial and submersible vehicle capable of flight and underwater navigation with seamless air-water transition. *arXiv preprint arXiv:1507.01932*.
- Mercado, D., Maia, M., and Diez, F. J. (2019). Aerial-underwater systems, a new paradigm in unmanned vehicles. *Journal of Intelligent & Robotic Systems*, 95(1):229–238.
- Ravell, D. A. M., Maia, M. M., and Diez, F. J. (2018). Modeling and control of unmanned aerial/underwater vehicles using hybrid control. *Control Engineering Practice*, 76:112–122.
- Rodriguez-Ramos, A., Sampedro, C., Bavle, H., Moreno, I. G., and Campoy, P. (2018). A deep reinforcement learning technique for vision-based autonomous multirotor landing on a moving platform. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1010–1017. IEEE.
- Srinivas, A., Laskin, M., and Abbeel, P. (2020). Curl: Contrastive unsupervised representations for reinforcement learning. *arXiv preprint arXiv:2004.04136*.
- Tai, L., Paolo, G., and Liu, M. (2017). Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 31–36. IEEE.