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

Ricardo B. Grando[1], **Paulo L. J. Drews-Jr**[1]

[1]Programa de Pós-Graduação em Computação/PPGCOMP
Universidade Federal do Rio Grande/FURG

ricardo.bedin@utec.edu.uy,paulodrews@furg.br

Abstract. *The search for the development of new technologies drives great challenges. An example of this refers to the development of tasks related to hybrid mobile robots. 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). OThe proposed approach uses only information from distance sensors and information related to the location of the vehicle to perform navigation. Results of our approach show that it is possible to perform navigation without a map from start to finish, without the need to use any type of manual operation, only using Deep-RL-based agents. For that, the navigation of the trained agents is compared with the navigation without a map performed by an algorithm BUG2, a modern version of a standard algorithm without learning for the problem. The proposed methods are based on two state-of-the-art approaches to map-less navigation of land robots: Deep Deterministic Policy Gradient (DDPG) and Soft Actor Critic (SAC).*

Resumo. *A busca pelo desenvolvimento de novas tecnologias impulsa grandes desafios. Exemplo disto refere-se ao desenvolvimento de tarefas correlatas aos robôs móveis híbridos. In order to study and overcome these challenges, the present work seeks to establish an approach based on Deep Reinforcement Learning (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 da nossa abordagem mostram que é possível realizar navegação sem mapa do início ao fim, sem que para isso fosse necessário utilizar nenhum tipo de operação manual, somente os agentes baseados em Deep-RL. Para tanto, a navegação dos agentes treinados é comparada com a navegação sem mapa realizada por um algoritmo BUG2, uma implementação moderna de algoritmo clássico para o problema de navegação sem mapas que não utiliza aprendizado. Os métodos propostos são baseados em duas abordagens 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**

[1]https://argo.furg.br/?BDTD12911

(FURG) em maio de 2021 sobre a orientação do Prof. Dr. Paulo L. J. Drews-Jr. Resumo submetido ao XXXV Concurso de Teses e Dissertações.

1. Introdução

Nos últimos anos tem aumentado o interesse no desenvolvimento de veículos híbridos não tripulados, denominado de HUVs (do inglês, *Hybrid Unmanned Vehicles*). Estes veículos buscam unir e explorar as vantagens e desvantagens existentes entre veículos aéreos, terrestres e subaquáticos. Nesse sentido, estudos acerca de um veículo híbrido capaz de atuar no meio aéreo e subaquático - HUAUV (do inglês, *Hybrid Unmanned Aerial Underwater Vehicle*) - em despontado como um grande desafio nas computação e nas engenharias. Tal fato pode ser atribuído 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 para o seu funcionamento correto. Assim, tornar estes veículos autônomos firma-se como um importante desafio de computação, dado que as informações sensoriais coletadas podem mudar bastante de meio para meio, além dos desafios em relação a tomada de decisões e autonomia, estabelecendo-se, assim, a navegação como um problema central.

Vários modelos de HUAUVs vêm sendo propostos ao longo dos anos. Pode-se citar como exemplo o trabalho seminal de [Drews et al. 2014] e o veículo apresentado por [Mercado et al. 2019], mostrando um aumento do interesse por esse tipo de veículo na última década. Nesse sentido, o foco de pesquisas no desenvolvimento de HUAUV focado em uma estrutura do tipo quadrotor. Ademais, pode-se afirmar que 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.

Porém, esse tipo de veículo ainda se encontra em fase de conceito e prototipação, com vários aspectos em aberto no desenvolvimento mecânico, de *hardware* e *software* que ainda precisam ser definidos e/ou abordados. Dentre os problemas mais primitivos e que talvez se detenham em termos de desafio computacional, re fere-se ao conceito de navegação para esse tipo de veículo. Perguntas que se apresentam são por exemplo como esse veículo navega de uma posição para outra no espaço, que estratégia computacional utilizar e quais sensores que seriam utilizados para tal. Diante disso, o presente trabalho propõe algoritmos necessários para lidar com o problema de navegação de HUAUVs sem mapas. Parte-se de uma abordagem de aprendizado conhecida como 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. O veículo é simulado realisticamente a partir de dados e modelos obtidos de um veículo real nos ambientes aéreo e subaquático.

Até o momento, o presente trabalho possui **três artigos científicos publicados** como primeiro autor. O primeiro foi publicado no Simpósio Latino Americano de Robótica 2020 (IEEE LARS 2020), trabalho que envolveu a parte de navegação aérea 2D e que ficou entre os 10 melhores do evento, referência latinoamericana na área e qualificado como **Qualis-B1**. O segundo artigo científico foi publicado na Conferência Internacional de Robótica e Automação, a mais importante do mundo na área (IEEE ICRA 2021, **Qualis A1**), envolvendo a navegação híbrida. O terceiro artigo publicado foi a versão

estendida no *Special Issue* do LARS 2020 considerando agora a navegação em 3D no Journal of Intelligent and Robotic Systems (JINT), revista qualificada como **Qualis-A2**, e uma das principais no que se refere a métodos computacionais para a Robótica. A metodologia final, sobre *Deep-RL* com RNNs, foi submetida recentemente na *Conferência Internacional de Robôs e Sistemas Inteligentes* - (IEEE IROS 2022, Qualis A1) - e está sob avaliação. Além disso, cita-se outros artigos publicados durante a execução da tese por impacto indireto da tese, focados no desenvolvimento de algoritmos de Deep-RL e veículos. Assim, obteve-se mais dois artigo no JINT (**Qualis-A2**), um sobre planejamento de trajetórias para HUAUV em 2022, e outro sobre Deep-RL aplicado a robôs móveis em 2021, e dois artigos no IEEE LARS 2022 sobre desenvolvimento de um AUV e HAUV (ambos **Qualis-B1**). Assim, obtendo um total de 3 artigos em periódicos **Qualis-A2**, 1 artigo em conferência **Qualis-A1** e 3 artigos em conferências **Qualis-B1**.

2. Trabalhos Relacionados

[Drews et al. 2014] apresenta 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 este aspecto, sendo que a maioria foca no desenvolvimento mecânico e estrutura do veículo, havendo, dessa forma, pouco progresso no que diz respeito aos desafios de autonomia e computacional para o veículo e ao problema da transição de meio [Mercado et al. 2019].

Trabalhos relacionados à navegação sem mapa para robôs terrestres, aéreos e subaquáticos são mais comumente encontrados na literatura. Os trabalhos mais relevantes acerca de HUAUVs partem de uma abordagem utilizando uma estrutura de quadrotor [Drews et al. 2014, Mercado et al. 2019, Horn et al. 2020]. O HUAUV descrito em simulação no presente 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. 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. 2017] é um dos mais citados da área e serviu de base para o desenvolvimento do presente trabalho. O problema da navegação sem mapa para UAVs e UUVs já foi abordado em vários trabalhos [Zhu and Zhang 2021].

Este trabalho diferencia-se dos demais trabalhos discutidos, pois busca abordar o problema da navegação sem mapa para HUAUV, trazendo contribuições no que tange ao desenvolvimento de novos algoritmos e alcançando avanços também no estado da arte de veículos apenas aéreos. Além disso, busca-se ainda observar a capacidade de evitar colisão com obstáculos, levando-se em conta somente informação de sensores de distância descritos a partir de sensores de distância simulados. A partir disso, duas classes de técnicas baseadas no estado da arte de *Deep-RL* (DDPG [Lillicrap et al. 2015] e SAC [Haarnoja et al. 2018]) são abordadas e analisadas, diferentemente da maioria dos trabalhos que abordam navegação sem mapa utilizando somente um tipo de abordagem. Para além, também apresentada-se um estudo de validação, objetivando-se analisar com mais robustez as propostas apresentadas e comparar com outras relacionadas, estudo não visto em outros trabalhos. Por último, este trabalho também é o primeiro a abordar o problema da transição do meio de modo autônomo para HUAUVs.

3. Metodologia

Em um primeiro momento, buscou-se descrever veículo e ambientes de simulação. Após isso, os agentes foram estruturados e desenvolvidos. A abordagem baseada algoritmo determinístico foi nomeado NDRL-D (do inglês *Navigation Deep Reinforcement Learning Deterministic*), enquanto que a abordagem baseada em algoritmos estocásticos foi nomeada NDRL-S (do inglês *Navigation Deep Reinforcement Learning Stochastic*).

3.1. Desenvolvimento do Veículo simulado

O veículo descrito no simulador realístico Gazebo foi baseado em um veículo real chamado Hydrone [Horn et al. 2019]. Um sensor de distância do tipo Lidar baseado no *plugin ray* do Gazebo foi desenvolvido para uso no ar. O sensor possui resolução angular de $0,25°$, mostrando 1080 leituras de distância em um alcance de $270°$. Entretanto, somente 20 amostras de distância igualmente espaçadas em $13,5°$ 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 de busca e, portanto, consiga convergir na detecção de objetos mais rapidamente, além de gerar um menor tamanho de rede que implica em uma diminuição da demanda de capacidade de processamento e tempo médio de treino por episódio. Um sensor de distância para o ambiente subaquático também foi acoplado ao veículo. Utilizou-se um sonar do tipo FLS baseado no simulador proposto em [Cerqueira et al. 2016]. Para tanto, foram utilizadas 20 amostras igualmente espaçadas entre os *beams* do sonar 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. Sendo estes, um simulador de física bastante realista e amplamente adotado dentro da área de robótica e sistemas autônomos, 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. Na parte positiva do eixo Z, ocorre a simulação aérea. O oceano possui corrente de água e, no ambiente aéreo, a influência de ventos se fez presente, com vistas a tornar a simulação ainda mais realística.

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*. 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. Nesse segundo 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.

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 entre cada abordagem determinística e estocástica. A Figura 1 mostra a estrutura para o contexto 3D e híbrido.

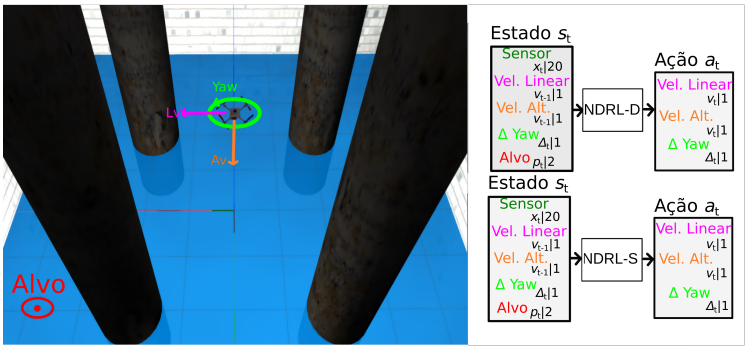


Figura 1. Veículo treinando em um dos cenários (esquerda) e o modelo básico de estrutura com entradas e saídas utilizando NDRL (direita) para o contexto 3D.

A rede do agente possui ao total 26 entradas e 3 saídas. Dos 26 valores de entrada, 20 são 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 também foi projetada para ser o mais semelhante entre elas o possível e foi inspirada no trabalho do [Tai et al. 2017]. Busca-se analisar as diferenças que cada abordagem de Deep-RL em si proporciona à problemática proposta. Para a estrutura da 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*) similar a [Tai et al. 2017]. 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 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 $0,25\ m/s$.

3.5. Função de Reforço

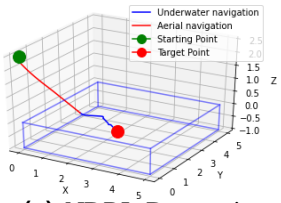
Um reforço positivo $r_{chegada}$ é dado caso o agente alcance o alvo dentro de uma margem de c_d metros. Um reforço negativo $r_{colisao}$ é dado para o episódio caso o veículo colida com a parede ou algum dos obstáculos do cenário. A verificação é feita baseada na distância mínima da leitura do sensor de distância. Caso a distância seja menor que uma distância de c_o metros, uma colisão é detectada. Os valores para os hiper-parâmetros utilizados foram: $r_{chegada}$ 100, c_d $0,5m$, $r_{colisao}$ -10, c_o $0,6m$.

4. Resultados

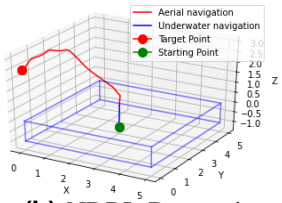
Dentre os diversos resultados obtidos, destaca-se o resultado da navegação híbrida. No primeiro cenário, o treinamento foi realizado por 1000 episódios, enquanto que no segundo cenário o agente evoluiu por um total de 5000 episódios. O número de episódios foi empiricamente definido a partir da convergência e a capacidade do agente em aprender a navegar. Os algoritmos desenvolvidos foram comparados com uma adaptação recente do algoritmo *BUG2* [Marino et al. 2016], tratando-se de um *baseline* importante que não utiliza aprendizado na sua estratégia.

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

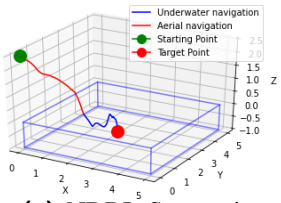
Cenário	Abordagem	Tempo Médio Ar (s)	Tempo Médio Água (s)	Total Sucesso
1	Ar-água-NDRL-D	15.29 ± 2.40	33.05 ± 10.91	96 %
1	Ar-água-NDRL-S	39.81 ± 23.16	23.64 ± 19.63	72 %
1	Ar-água-BUG2	31.963 ± 1.65	21.05 ± 0.199	100 %
1	Água-Ar-NDRL-D	18.78 ± 1.21	6.10 ± 0.17	97 %
1	Água-Ar-NDRL-S	33.98 ± 26.15	17.36 ± 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 %



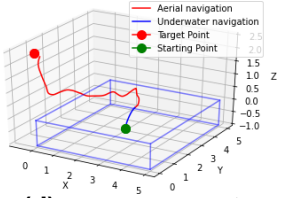
(a) NDRL-D no primeiro cenário do ar para a àgua.



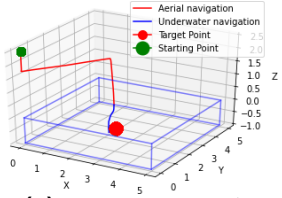
(b) NDRL-D no primeiro cenário da água para o ar.



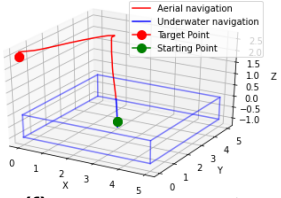
(c) NDRL-S no primeiro cenário do ar para a àgua.



(d) NDRL-S no primeiro cenário da água para o ar.

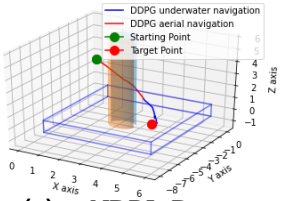


(e) *BUG2* no primeiro cenário do ar para a água.

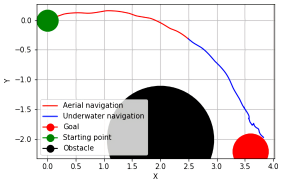


(f) *BUG2* no primeiro cenário da água para o ar.

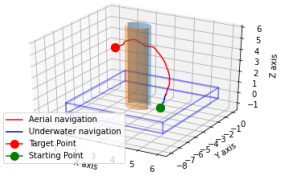
Figura 2. Caminho realizado durante 1 das 100 tentativas no contexto 3D híbrido no primeiro cenário.



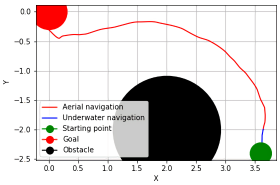
(a) NDRL-D no segundo cenário do ar para a àgua.



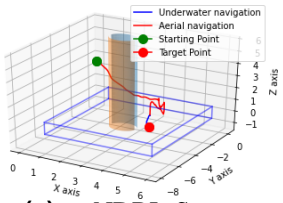
(b) NDRL-D no segundo cenário do ar para a água visto de cima.



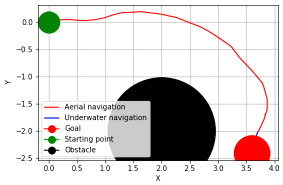
(c) NDRL-D no segundo cenário da água para o ar.



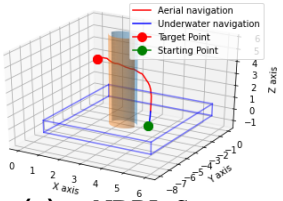
(d) NDRL-D no segundo cenário da água para o ar visto de cima.



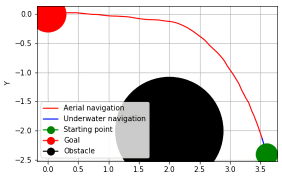
(e) NDRL-S no segundo cenário do ar para a àgua.



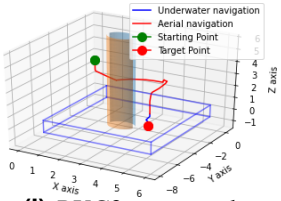
(f) NDRL-S no segundo cenário do ar para a àgua visto de cima.



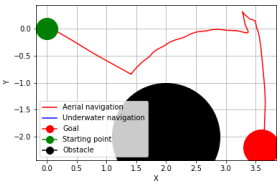
(g) NDRL-S no segundo cenário da água para o ar.



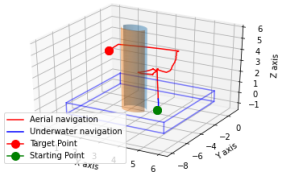
(h) NDRL-S no segundo cenário da água para o ar visto de cima.



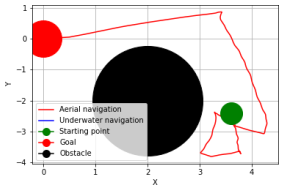
(i) *BUG2* no segundo cenário do ar para a àgua.



(j) *BUG2* no segundo cenário do ar para a àgua 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 3. Caminho realizado durante 1 das 100 tentativas no contexto 3D híbrido no segundo cenário.

O desempenho dos agentes em cada tarefa está apresentado na sequência. A 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 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 2 é possível observar 1 das 100 tentativas de navegação realizada por ambos agentes e pelo algoritmo *BUG2*, no primeiro cenário. Figura 3 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. 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.

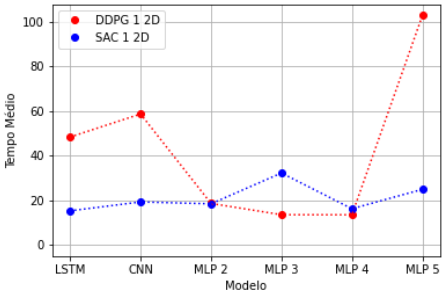
5. Estudo de Validação (*Ablation Study*)

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 (*ablation study*). 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 a abordagem estocástica se comporta melhor em ambientes com obstáculos.

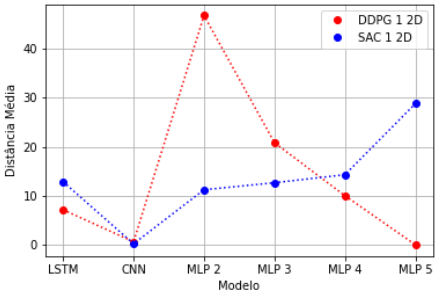
Pode-se observar que quanto maior e mais complexa a rede, maior tende a ser o reforço médio, como por exemplo para os modelos com LSTM e CNN. É importante ressaltar que isso se deve ao fato de um número maior de navegações ser realizado a cada episódio e não que os modelos são melhores ou piores. Todos 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 4b 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 4b 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 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 4b pode-se observar também a distância média para a 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.



(a) Tempo médio.



(b) Distancia média.

Figura 4. Comparação do tempo e da distância média para o contexto 2D no primeiro cenário (mais estável)

O limite para a complexidade para a abordagem proposta, entretanto, parece ser próximo ao modelo convolucional proposto. Como pode-se observar na Figura 4b 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.

6. Considerações Finais

Diante de todo o exposto, tem-se que neste trabalho, foram propostas novas 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 simulação aérea e subaquática realística. Abordagens determinísticas e estocásticas foram desenvolvidas e validadas em 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 autonomamente realizar a 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, realizando com sucesso a transição de meio em ambas as tarefas. Sem utilizar uma abordagem baseada em informação visual, que demanda um poder computacional maior que é geralmente difícil de ser embarcado em HUAUVs de pequeno porte pelo peso e consumo energético, a abordagem proposta, mesmo com poucos estados, mostrou-se ser capaz realizar a navegação do veículo e fazê-lo ser autônomo o suficiente para evitar colisões com objetos.

Referências

Cerqueira, 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 *SIBGRAPI*.

Drews, P. L., Neto, A. A., and Campos, M. F. (2014). Hybrid unmanned aerial underwater vehicle: Modeling and simulation. In *IEEE/RSJ IROS*, pages 4637–4642.

Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *arXiv preprint arXiv:1801.01290*.

Horn, A. C., Pinheiro, P. M., Grando, R. B., da Silva, C. B., Neto, A. A., and Drews-Jr, P. L. (2020). A novel concept for hybrid unmanned aerial underwater vehicles focused on aquatic performance. In *IEEE LARS/SBR*, pages 1–6.

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 *ICAR*, pages 173–178.

Li, Y. (2017). Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274*.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. (2015). Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.

Marino, R., Mastrogiovanni, F., Sgorbissa, A., and Zaccaria, R. (2016). A minimalistic quadrotor navigation strategy for indoor multi-floor scenarios. In *Intelligent Autonomous Systems 13*, pages 1561–1570. Springer.

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.

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 *IEEE/RSJ IROS*, pages 31–36.

Zhu, K. and Zhang, T. (2021). Deep reinforcement learning based mobile robot navigation: A review. *Tsinghua Science and Technology*, 26(5):674–691.