**هدایت ماهواره به‌صورت یادگیری تقویتی مقاوم مبتنی بر بازی دیفرانسیلی در محیط‌های پویای چند جسمی با پیشر‌ان‌کم**

**Robust Reinforcement Learning Differential Game Guidance of Low-Thrust Satellite in Multi-Body Dynamical Environments**

**معرفی موضوع:**

در سال‌های اخیر، پیشرفت‌های فناوری در زمینه‌های مختلف، از جمله کنترل پرواز، پردازش سیگنال و هوش مصنوعی، به افزایش کاربردهای ماهواره با پیشران‌کم در محیط‌های چند جسمی multi-body)) مانند منظومه زمین‐ماه کمک کرده است. ماهواره با پیشران‌کم می‌تواند برای تعقیب ماهواره‌ها، انتقال مداری و استقرار ماهواره‌ها استفاده شود. روش‌های هدایت بهینه قدیمی جهت کنترل ماهواره‌ها اغلب نیازمند فرضیات ساده کننده، منابع محاسباتی فراوان و شرایط اولیه مناسب هستند. الگوریتم‌های مبتنی بر یادگیری تقویتی این توانایی را دارند که بدون مشکلات اشاره‌شده هدایت ماهواره را انجام دهند. به همین دلیل، این الگوریتم‌ها می‌توانند امکان محاسبات درونی (On-board Computing) را فراهم کنند.

**اهمیت موضوع:**

رویکردهای قدیمی برای هدایت فضاپیما در محیط‌های پیچیده چندجسمی اغلب به ایستگاه‌های کنترل زمینی متکی هستند که می‌توانند در معرض خرابی ارتباطات، تاخیرهای زمانی و محدودیت‌های انتقال داده باشند. لذا، ارائه روش‌های هدایت بهینه و مقاوم که امکان محاسبه فرامین هدایت را درون ماهواره فراهم کند، از اهمیت بالایی برخوردار است.

**کاربردها:**

الگوریتم‌های هدایت مقاوم و بهینه که محاسبات آن‌ها به‌صورت درونی انجام می‌شود، در ماموریت‌های فضایی مختلفی مانند سفر به ماه کاربرد دارند. این الگوریتم‌ها دارای مزایایی مانند استقلال از ایستگاه زمینی، سرعت تصمیم‌گیری بالا و بهبود ایمنی هستند. از کاربردهای این نوع الگوریتم‌ها می‌توان به کنترل مدار و وضعیت فضاپیماها در طول سفر به ماه، کنترل مدار و وضعیت فضاپیماها در مدارهای نزدیک به زمین، تعمیر و نگهداری ماموریت‌های فضایی، تعقیب اهداف در مدارهای مختلف و انجام مانور‌های پیچیده اشاره کرد.

**تعریف دقیق مسئله:**

در پژوهش حاضر، مسأله هدایت فضاپیما برای سیستم زمین-ماه با استفاده از مدل دینامیکی سه‌جسم محدود دایره‌ای Circular Restricted Three-Body Problem (CR3BP) مطرح می‌شود. در این مسأله، یک فضاپیما با سیستم پیشران‌کم به یک مسیر مرجع در سیستم CR3BP منتقل می‌شود. شرایط اولیه فضاپیما از مسیر مرجع انحراف‌های تصادفی دارند. هدف از این پژوهش، توسعه یک الگوریتم هدایت حلقه‌بسته با استفاده از اصول یادگیری تقویتی بر مبنای بازی دیفرانسیلی است که به فضاپیما اجازه می‌دهد با فرض بدترین اغتشاش به مسیر مرجع بازگردد، مسیر مرجع را دنبال کرده و به مدار مقصد برسد. بازی دیفرانسیلی موجب می‌شود الگوریتم هدایت نه تنها بهینه، بلکه مقاوم نیز باشد.

**فرضیات مسئله:**

* مدل دینامیکی مسأله، مسأله سه ‌جسم محدود دایره‌ای سیستم زمین-ماه است. این فرضیه یک سیستم دو جسمی ساده است که فضاپیما را به‌عنوان جسم سوم بسیار کوچک فرض می‌کند.
* حرکت فضاپیما تنها در صفحه اتفاق می‌افتد و تنها نیروهای گرانشی و اغتشاش در‌نظر گرفته می‌شوند.
* فضاپیما به یک سیستم پیشران‌کم مجهز است که شعاع تاثیر مشخصی دارد. میزان و جهت پیشران به‌صورت مستمر تغییر کرده و قابل تنظیم است.
* مسأله هدایت شامل انتقال بین مدارهای دوره‌ای (مدارهای لیاپانوف) در مدل CR3BP زمین-ماه است.
* محیط یادگیری تقویتی ویژگی‌های مارکوف را دارد؛ به این معنا که حالت فعلی تمام اطلاعات لازم برای پیش‌بینی حالت‌های آینده را فراهم می‌کند.
* فضاپیما در مسیر نزدیک به مرجع با فرض بدترین اغتشاش می‌ماند تا به مدار مقصد برسد.

**روش انجام کار:**

**مرحله 1: مدل‌سازی و شبیه‌سازی محیط دینامیکی**

پیاده‌سازی مدل دینامیکی محیط مورد نظر انجام می‌شود. این مدل باید شامل تمام جوانب مهم محیط از جمله قوانین حرکت، ابعاد فضایی و حالت اولیه باشد. مدل دینامیکی محیط به‌صورتی که اطلاعات مورد نیاز برای یادگیری تقویتی (مانند حالت فعلی) را تامین کند، ایجاد می‌شود. یادگیری الگوریتم هدایت در حضور اغتشاشاتی مانند تشعشع خورشیدی انجام می‌شود. اغتشاش در این مرحله مدل‌سازی می‌شود و بدترین اغتشاش بر اساس مدل بدست آمده اعمال می‌شود.

مرحله 2: ایجاد و آموزش بازیکن هدایت با استفاده از بازی دیفرانسیلی

در این مرحله، بازیگری برای انجام هدایت سفینه در محیط ایجاد می‌شود. این بازیگر مسئول انتخاب اعمال (اعمال کم‌پیشران) بر اساس حالت فعلی محیط است. بر اساس بازی دیفرانسیلی حل شدن مسلئه min-max خود به خود باعث می‌شود که بهترین جواب در حضور بدترین اغتشاش به‌دست آید. یادگیری به‌صورت آفلاین انجام می‌شود و سپس در هنگام حرکت نیز به صورت آنلاین نیز یادگیری انجام می‌شود تا هدایت به‌طور مناسب انجام شود.

از الگوریتم‌های یادگیری تقویتی مختلف مانند Q-Learning، Deep Q-Networks (DQN) یا Proximal Policy Optimization (PPO) برای ایجاد بازیگر هدایت استفاده می‌شود. انتخاب الگوریتم بستگی به محیط و اهداف آموزش دارد.

**یشینه‌ی موضوع:**

پژوهش‌های خارجی:

هدایت فضاپیماها معمولاً با استفاده از ایستگاه‌های زمینی انجام می‌شود. با این حال، این تکنیک‌ها دارای محدودیت‌هایی از جمله حساسیت به قطع ارتباطات، تاخیرهای زمانی، و محدودیت‌های منابع محاسباتی هستند. الگوریتم‌های یادگیری تقویتی و بازی‌های دیفرانسیلی می‌توانند برای بهبود قابلیت‌های هدایت فضاپیماها، از جمله مقاومت در برابر تغییرات محیطی، کاهش تاخیرهای ناشی از ارتباطات زمینی، و افزایش کارایی محاسباتی، مورد استفاده قرار گیرند.

هدایت فضاپیماها معمولاً پیش از پرواز انجام می‌شود. این روش‌ها می‌توانند از تکنیک‌های بهینه‌سازی فراگیر [1] یا برنامه‌نویسی غیرخطی برای تولید مسیرها و فرمان‌های کنترلی بهینه استفاده کنند. با این حال، این روش‌ها معمولا حجم محاسباتی زیادی دارند و برای استفاده درون‌سفینه نامناسب هستند [2]. یادگیری ماشین می‌تواند برای بهبود قابلیت‌های هدایت فضاپیماها استفاده شود. کنترل‌کننده شبکه عصبی حلقه‌‌بسته می‌تواند برای محاسبه سریع و خودکار تاریخچه کنترل استفاده شود. یادگیری تقویتی نیز می‌تواند برای یادگیری رفتارهای هدایت بهینه استفاده شود.

روش‌های هدایت و بهینه‌سازی مسیر فضاپیماها به‌طور کلی به راه‌حل‌های اولیه مناسب نیاز دارند. در مسائل چند جسمی، طراحان مسیر اغلب حدس‌های اولیه کم‌هزینه‌ای برای انتقال‌ها با استفاده از نظریه سیستم‌های دینامیکی و منیفولدهای ثابت [3, 4] ایجاد می‌کنند.

شبکه‌های عصبی ویژگی‌های جذابی برای فعال‌سازی هدایت در فضاپیما دارند. به‌عنوان مثال، شبکه‌های عصبی می‌توانند به‌طور مستقیم از تخمین‌های وضعیت به دستورهای پیشران کنترلی که با محدودیت‌های مأموریت سازگار است، برسند. عملکرد هدایت شبکه‌های عصبی در مطالعاتی مانند فرود بر سیارات [23] ، عملیات نزدیکی به سیارات [24]و کنترل فضاپیما با پیشران ازدست‌رفته [12] نشان داده شده است. تازه‌ترین پیشرفت‌های تکنیک‌های یادگیری ماشین در مسائل خودکارسازی درونی به‌طور گسترده‌ای مورد مطالعه قرار گرفته‌اند؛ از پژوهش‌های اولیه تا توانایی‌های پیاده‌سازی. به‌عنوان مثال، الگوریتم‌های یادگیری ماشین ابتدایی در فضاپیماهای مریخی نبرد برای کمک به شناسایی ویژگی‌های زمین‌شناسی تعبیه شده‌اند. الگوریتم AEGIS توانایی انتخاب خودکار هدف توسط یک دوربین در داخل فضاپیماهای Spirit، Opportunity و Curiosity را فعال دارد .[5] در کامپیوتر پرواز اصلی، فرآیند دقت افزایی (Refinement Process) نیاز به 94 تا 96 ثانیه دارد [6]، که به طور قابل توجهی کمتر از زمان مورد نیاز برای ارسال تصاویر به زمین و انتظار برای انتخاب دستی توسط دانشمندان است. برنامه‌های آینده برای کاربردهای یادگیری ماشین درون‌سفینه شامل توانایی‌های رباتیکی درون‌سفینه برای فضاپیمای Perseverance [8, 7] و شناسایی عیب برای Europa Clipper [6] می‌شود. الگوریتم‌های یادگیری ماشین پتانسیلی برای سهم مهمی در مأموریت‌های اتوماسیون آینده دارند. علاوه بر رباتیک سیاره‌ای، پژوهش‌های مختلفی به استفاده از تکنیک‌های مختلف یادگیری ماشین در مسائل نجومی پرداخته‌اند. در طراحی مسیر عملکرد رگرسیون معمولاً مؤثرتر هست. به عنوان مثال، از یک شبکه عصبی (NN) در بهینه‌سازی مسیرهای رانشگر کم‌پیشران استفاده شده است.[9] پژوهش‌های جدید شامل شناسایی انتقال‌های هتروکلینیک [10]، اصلاح مسیر رانشگر کم‌پیشران [11] و تجزیه و تحلیل مشکلات ازدست‌رفتن رانشگر [12]می‌شود.تکنیک‌های یادگیری نظارتی می‌توانند نتایج مطلوبی تولید کنند؛ اما، دارای محدودیت‌های قابل توجهی هستند. یکی از این محدودیت‌ها این است که این رویکردها بر وجود دانش پیش از فرآیند تصمیم‌گیری متکی هستند. این امر مستلزم دقیق‌بودن داده‌های تولید‌شده توسط کاربر برای نتایج مطلوب و همچنین وجود تکنیک‌های موجود برای حل مشکل کنونی و تولید داده است.

در سال‌های اخیر، قابلیت یادگیری تقویتی (RL) در دستیابی به عملکرد بهینه در دامنه‌هایی با ابهام محیطی قابل توجه، به اثبات رسیده است .[14, 13] هدایت انجام‌شده توسط RL را می‌توان به‌صورت گسترده بر اساس فاز پرواز دسته‌بندی کرد. مسائل فرود [16, 15] و عملیات در نزدیکی اجسام کوچک [24, 23]، از حوزه‌های پژوهشی هستند که از RL استفاده می‌کنند. تحقیقات دیگر شامل مواجهه تداخل خارجی جوی [17]، نگهداری ایستگاهی [19, 18] و هدایت به‌صورت جلوگیری از شناسایی [20] است. مطالعاتی که فضاپیماهای رانشگر کم‌پیشران را در یک چارچوب دینامیکی چند بدنی با استفاده از RL انجام‌شده‌است، شامل طراحی انتقال با استفاده از Q-learning [21]، Proximal Policy Optimization [22] و هدایت نزدیکی مدار [23] است.

در مرجع [25]، از یادگیری تقویتی متا برای هدایت خودکار فضاپیما استفاده می‌کند. سیاست کنترل (فرمان کنترل) با یک شبکه عصبی پیچشی-بازگشتی جایگزین می شود. نتایج نشان می‌دهد که این روش می‌تواند فضاپیما را به درستی به سمت نقطه نهایی هدایت کند.

در مرجع [26]، **یک کنترل‌کننده مبتنی بر یادگیری تقویتی را بر روی سخت‌افزار NVIDIA Jetson Nano پیاده‌سازی کرده است و نشان می‌دهد که این سخت‌افزار کنترل‌کننده مناسب پرواز است.**

در مرجع [27]، از یادگیری تقویتی برای حل مشکلات ایستایی (stationkeeping) ماهواره کم‌پیشران در مدارهای نزدیک به ماه استفاده می‌کند. این روش به فضاپیما کمک می‌کند تا مدار خود را بدون نیاز به دانش قبلی از مانور حفظ کند. نتایج نشان می‌دهد که این روش می‌تواند به طور موثری زمان مانورها و مصرف سوخت را کاهش دهد و در عین حال اطمینان از موفقیت ماموریت را تضمین کند.

در مرجع **[28]، یک رویکرد برای شناسایی همگرایی بازیابی مسیر در فضاپیماهای فرارکننده (**runaway**) استفاده می‌کند.** **مطالعه نشان می‌دهد که الگوریتم پیشنهادی در مقایسه با روش‌های حفظ مدار، بازیابی را پس از گذشت زمان بیشتری امکان‌پذیر می‌کند و نتایج بهتری در شناسایی زمان و کنترل مانور ارائه می‌دهد.**

**در مرجع [29]، یادگیری تقویتی چندهدفه برای کشف یک زیرمجموعه از فضای راه‌حل چندهدفه برای انتقال کم‌پیشران بین دو مدار L2 جنوبی در مسئله سه‌جسم زمین-ماه استفاده می‌شود. سیاست‌های متعددی در این سناریو آموزش داده می‌شوند تا انتقال‌ها را انجام دهند. این سیاست پاداش‌ها را که بر اساس وزن‌دهی‌های نسبی متمایز دو هدف متفاوت هستند، به حداکثر می‌رسانند. این اهداف شامل به حداقل رساندن زمان پرواز و مصرف سوخت. ارزیابی این سیاست‌ها به طور موفقیت‌آمیزی انتقال‌هایی بین دو مدار با اشکال مختلف را نشان می‌دهد. این نتایج همچنین با راه‌حل‌های نزدیک (**nearby solutions**) به یک مسئله بهینه‌سازی مقید مقایسه شده است.**

پژوهش‌های داخلی:

در زمینه هدایت و کنترل فضاپیماها و ماهواره‌ها، پیشرفت‌های زیادی در سال‌های اخیر حاصل شده است. این پیشرفت‌ها عمدتاً ناشی از توسعه روش‌های جدیدی مانند یادگیری تقویتی، کنترل‌های تطبیقی و بینایی کامپیوتر است. این پیشرفت‌ها به بهبود ایمنی و کارایی ماموریت‌های فضایی کمک کرده است. در مرجع [30]، یادگیری تقویتی برای مانور اتصال دو فضاپیما انجام شده است. در مرجع [31]، کنترل‌های تطبیقی برای بازپیکربندی آرایش پروازی ماهواره‌ها پیاده‌سازی شده است. در مرجع [32]، بینایی کامپیوتر برای ناوبری و کنترل بازوی رباتیک ماهواره اجرا شده است.

در مرجع [33] ، یک ساختار جدید از کنترل‌کننده هوشمند مبتنی بر یادگیری عاطفی در مغز و شبکه‌های عصبی معرفی شده است. این ساختار برای کنترل وضعیت یک ماهواره زمین‌گرا با رفتار دینامیکی بسیار پیچیده در مدار پایین (LEO) مورد استفاده قرار گرفته است. نتایج شبیه‌سازی نشان می‌دهد که این ساختار کنترلی می‌تواند به‌طور موثری وضعیت دینامیکی پیچیده این ماهواره را کنترل کند.

در مرجع [34] ، یک کنترل‌کننده تطبیقی بهینه بدون مدل برای بازپیکربندی آرایش پروازی ماهواره‌ها معرفی شده است. این کنترل‌کننده با استفاده از یادگیری تقویتی، به‌صورت برخط و بدون نیاز به مدل ریاضی دقیق، می‌تواند مسئله ردیابی بازپیکربندی را حل کند.

**اهداف پژوهش:**

* طراحی یک بازیکن جهت هدایت فضاپیما در محیط‌های پویا چند جسمی با استفاده از یادگیری تقویتی
* طراحی یک الگوریتم هدایت مقاوم هدایت فضاپیما در محیط‌های پویا چند جسمی به کمک بازی دیفرانسیلی
* بهبود عملکرد هدایت فضاپیما در محیط سه جسم
* رفع نیاز به ایستگاه زمینی جهت محاسبه و صدور فرمان هدایت فضاپیما در فاصله‌های دور
* فراهم کردن امکان پیاده‌سازی الگوریتم هدایت مقاوم درون‌سفینه‌ای در محیط سه جسم پویا شبیه‌سازی‌شده

**تابستان سال اول:** انجام مطالعه و مروری جامع در زمینه بازی دیفرانسیلی و یادگیری تقویتی. بررسی و تعریف دقیق مسأله یادگیری تقویتی و الگوریتم هدایت به وسیله یادگیری تقویتی در مسئله سه‌جسم. طراحی ساختار محیط سه‌جسم و بازیکن یادگیری تقویتی و ارتباط بین محیط و بازیکن.

**نیم‌سال اول سال دوم**

طراحی مسیر انتقال بهینه بین دو مدار. شبیه‌سازی دقیق محیط مسأله سه‌جسم جهت آموزش بازیکن یادگیری تقویتی. طراحی و توسعه سیستم یادگیری تقویتی جهت محاسبه دستور

**نیم‌سال دوم سال دوم**

پیاده‌سازی هدایت فضاپیما با استفاده از الگوریتم‌های کلاسیک. آموزش الگوریتم‌های یادگیری تقویتی در محیطهای شبیه‌سازی شده

**تابستان سال دوم**

اصلاح و بهبود عملکرد یادگیری تقویتی در هدایت فضاپیما. نتیجه‌گیری از تحقیقات و ارائه گزارش نهایی شامل معرفی مسئله، روش‌های استفاده‌شده، نتایج حاصل و پیشنهادات برای تحقیقات آتی

**نتایج مورد انتظار:**

• افزایش ایمنی ماموریت به‌دلیل استفاده از بازی دیفرانسیلی

• کاهش مصرف سوخت و هزینه به‌دلیل بهینه‌بودن الگوریتم هدایت

• هدایت فضاپیما بدون نیاز به ایستگاه زمینی

• کاهش هزینه ساخت به‌دلیل عدم نیاز به سیستم‌های ارتباطی قوی

• الگوریتم در حضور سایر اغتشاشات مدل نشده در آموزش کارایی لازم را داشته باشد.

•از انجایی که هدف مسئله هدایت (کنترل مسیر) است، انتظار می‌رود که الگوریتم هدایت که با استفاده از مدل جرم نقطه‌ای (دو درجه آزادی در صفحه) آموزش دیده است، کارایی لازم را در هدایت سه درجه آزادی (در صفحه) داشته باشد.

**روش صحت‌سنجی نتایج:**

• مقایسه با سایر روشهای معتبر

• مدل‌سازی و شبیه‌سازی سیستم محیط و بازیکن

**گلوگاه‌های پیش‌بینی‌شده:**

• آموزش شبکه: به‌علت پیچیده‌بودن شبکه و محیطی که طراحی شده است، محاسبات سنگین می‌شود.

• پایداری و کنترل: به‌علت غیرخطی‌بودن دینامیک سامانه و ناپایداری ذاتی سیستم، کنترل و پایداری آن مشکل است.

**نوآوری‌ها:**

پیش از این، از ترکیب یادگیری تقویتی و بازی دیفرانسیلی در یک مسئله چند جسمی پویا استفاده نشده است. به‌علاوه، منطق هدایت ساخته‌شده از ترکیب دو روش، یک هدایت مقاوم و در عین حال بهینه است، که تاکنون در این مسئله انجام نشده است.

.

[1] Vavrina, Matthew A, Jacob A Englander, Sean M Phillips, and Kyle M Hughes. ”Global, multi-objective trajectory opti- mization with parametric spreading.” In AAS AIAA Astrodynamics Specialist Conference 2017, Tech. No. GSFC-E-DAA- TN45282, 2017.

[2] Ocampo, Cesar. ”Finite Burn Maneuver Modeling for a Generalized Spacecraft Trajectory Design and Optimization Sys- tem.” Annals of the New York Academy of Sciences 1017 (2004): 210-233. doi: 10.1196/annals.1311.013.

[3] Marchand, B. G. and Scarritt, S. K. and Pavlak, T. A. and Howell, K. C. “A dynamical approach to precision entry in multi-body regimes: Dispersion manifolds.” Acta Astronautica 89 (2013): 107-120. doi: 10.1016/j.actaastro.2013.02.015.

[4] T.A.Estlin,B.J.Bornstein,D.M.Gaines,R.C.Anderson,D.R.Thompson,M.Burl,R.Castaño,M.Judd.”Aegisautomated science targeting for the MER Opportunity rover.” ACM Trans. Intell. Syst. Technol. (TIST), 3, 1-19, 2012.

[5] R. Francis, T. Estlin, G. Doran, S. Johnstone, D. Gaines, V. Verma, M. Burl, J. Frydenvang, S. Montano, R. Wiens, S. Schaffer, O. Gasnault, L. Deflores, D. Blaney, B. Bornstein. ”Aegis autonomous targeting for ChemCam on Mars Science Laboratory: Deployment and results of initial science team use.” Science Robotics, 2, 2017.

[6] K.L. Wagstaff, G. Doran, A. Davies, S. Anwar, S. Chakraborty, M. Cameron, I. Daubar, C. Phillips. ”Enabling onboard detection of events of scientific interest for the Europa Clipper spacecraft,” in: 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Association for Computing Machinery, Anchorage, Alaska, 2019, pp. 2191–2201, doi: 10.1145/3292500.3330656.

[7] S. Higa, Y. Iwashita, K. Otsu, M. Ono, O. Lamarre, A. Didier, M. Hoffmann. ”Vision-based estimation of driving energy for planetary rovers using deep learning and terramechanics,” IEEE Robot. Autom. Lett. 4 (2019) 3876–3883.

[8] B. Rothrock, J. Papon, R. Kennedy, M. Ono, M. Heverly, C. Cunningham. ”Spoc: Deep learning-based terrain classification for Mars rover missions,” in: AIAA Space and Astronautics Forum and Exposition, SPACE 2016, American Institute of Aeronautics and Astronautics Inc, AIAA, 2016, pp. 1–12.

[9] B. Dachwald. ”Evolutionary neurocontrol: A smart method for global optimization of low-thrust trajectories,” in: AIAA/AAS Astrodynamics Specialist Conference and Exhibit, Providence, Rhode Island, 2004, pp. 1–16.

[10] S. De Smet, D.J. Scheeres. ”Identifying heteroclinic connections using artificial neural networks,” Acta Astronaut. 161 .199–192 )2019(

[11] N.L.O. Parrish. ”Low Thrust Trajectory Optimization in Cislunar and Translunar Space (Ph.D. thesis),” University of Colorado Boulder, 2018.

[12] A. Rubinsztejn, R. Sood, F.E. Laipert. ”Neural network optimal control in astrodynamics: Application to the missed thrust problem,” Acta Astronaut. 176 (2020) 192–203.

[13] N.Heess,D.TB,S.Sriram,J.Lemmon,J.Merel,G.Wayne,Y.Tassa,T.Erez,Z.Wang,S.M.A.Eslami,M.A.Riedmiller, D. Silver. ”Emergence of locomotion behaviours in rich environments,” 2017. CoRR abs/1707.02286. arXiv:1707.02286.

[14] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, D. Hassabis. ”Mastering the game of Go without human knowledge,” Nature 550 (2017) Article.

[15] R. Furfaro, A. Scorsoglio, R. Linares, M. Massari, ”Adaptive generalized ZEM-ZEV feedback guidance for planetary landing via a deep reinforcement learning approach,” Acta Astronaut. 171 (2020) 156–171.

[16] B.Gaudet,R.Linares,R.Furfaro,”Deepreinforcementlearningforsixdegreesoffreedomplanetarylanding,”Adv.Space Res. 65 (2020) 1723–1741.

[17] J. Broida, R. Linares, ”Spacecraft rendezvous guidance in cluttered environments via reinforcement learning,” in: *29th AAS/AIAA Space Flight Mechanics Meeting*, American Astronautical Society, Ka’anapali, Hawaii, 2019, pp. 1–15.

[18] B. Gaudet, R. Furfaro, R. Linares, ”Reinforcement learning for angle-only intercept guidance of maneuvering targets,” *Aerosp. Sci. Technol.* 99 (2020).

[19] D. Guzzetti, ”Reinforcement learning and topology of orbit manifolds for station-keeping of unstable symmetric periodic orbits,” in: *AAS/AIAA Astrodynamics Specialist Conference*, American Astronautical Society, Portland, Maine, 2019, pp. .20–1

[20] J.A. Reiter, D.B. Spencer, ”Augmenting spacecraft maneuver strategy optimization for detection avoidance with compet- itive coevolution,” in: *20th AIAA Scitech Forum*, AIAA, Orlando, Florida, 2020, pp. 1–11.

[21] A. Das-Stuart, K.C. Howell, D.C. Folta, ”Rapid trajectory design in complex environments enabled by reinforcement learning and graph search strategies,” *Acta Astronaut.* 171 (2020) 172–195.

[22] D.Miller,R.Linares,”Low-thrustoptimalcontrolviareinforcementlearning,”in:*29thAAS/AIAASpaceFlightMechanics Meeting*, American Astronautical Society, Kaanapali, Hawaii, 2019, pp. 1–18.

[23] C.J. Sullivan, N. Bosanac, ”Using reinforcement learning to design a low-thrust approach into a periodic orbit in a multi- body system,” in: *20th AIAA Scitech Forum*, AIAA, Orlando, Florida, 2020, pp. 1–19.

[24] B. Gaudet, R. Linares, R. Furfaro, ”Terminal adaptive guidance via reinforcement meta-learning: Applications to au- tonomous asteroid close-proximity operations,” *Acta Astronaut.* 171 (2020) 1–13.

[25] [Image-Based Meta-Reinforcement Learning for Autonomous Guidance of an Asteroid Impactor](https://arc.aiaa.org/doi/abs/10.2514/1.G006832), Lorenzo Federici, Andrea Scorsoglio, Luca Ghilardi, Andrea D’Ambrosio, Boris, Benedikter, Alessandro Zavoli, and Roberto Furfaro, Journal of Guidance, Control, and Dynamics 2022 45:11, 2013-2028

[26] Wilson, C.; Riccardi, A. Enabling Intelligent Onboard Guidance, Navigation, and Control Using Reinforcement Learning on near-Term Flight Hardware. Acta Astronautica 2022, 199, 374–385. <https://doi.org/https://doi.org/10.1016/j.actaastro.2022.07.013>.

[27] Nicholas B. LaFarge, Kathleen C. Howell and David C. Folta. ["An Autonomous Stationkeeping Strategy for Multi-Body Orbits Leveraging Reinforcement Learning,"](https://arc.aiaa.org/doi/10.2514/6.2022-1764)AIAA 2022-1764. *AIAA SCITECH 2022 Forum*. January 2022.

[28] LaFarge, N. B.; Howell, K. C.; Folta, D. C. Adaptive Closed-Loop Maneuver Planning for Low-Thrust Spacecraft Using Reinforcement Learning. Acta Astronautica 2023, 211, 142–154. <https://doi.org/https://doi.org/10.1016/j.actaastro.2023.06.004>.

[29] [Designing Low-Thrust Transfers near Earth–Moon L2 via Multi-Objective Reinforcement Learning](https://arc.aiaa.org/doi/abs/10.2514/1.A35463), Christopher J. Sullivan, Natasha Bosanac, and Rodney L. Anderson, Journal of Spacecraft and Rockets 2023 60:2, 634-647

[30] صیامی عراقی, محمد و شفیعی نژاد, ایمان,1399,طراحی مسیر اتصال میان دو ماهواره با الگوریتم یادگیری تقویتی Q,پنجمین کنفرانس ملی مهندسی مکانیک و هوافضا,تهران,,,https://civilica.com/doc/1039632

[31]کنکاشور, محمدرسول, بلندی, حسین, مزینی, ناصر. (1401). 'طراحی کنترل‌کننده تطبیقی بهینه بدون مدل برای باز پیکربندی آرایش پروازی ماهواره‌ها با یادگیری تقویتی', دانش و فناوری هوافضا, 11(1), pp. 41-60.

[32] موسوی، فرزان، ناوبری و کنترل بازوی رباتیک ماهواره پرواز ـ آزاد حین جا به‌ جایی و قراردادن محموله در ایستگاه‌ فضایی.، صنعتی شریف، 1401، مهندسی هوافضا، کارشناسی ارشد،

[33] کامزن, میلاد و غنی فر, مانا و تشنه لب, محمد و روشنی یان, جعفر و نیکخواه, امیرعلی,1401,کنترل مانور وضعیت یک ماهواره زمین گرا مدار پایین بر پایه روش کنترل تقویتی مبتنی بر یادگیری عاطفی در مغز,بیست و یکمین کنفرانس بین المللی انجمن هوا فضای ایران,تهران,,,https://civilica.com/doc/1668247

[34] کنکاشور, م.; بلندی, ح.; مزینی, ن. طراحی کنترل کننده تطبیقی بهینه بدون مدل برای باز پیکربندی آرایش پروازی ماهواره ها با یادگیری تقویتی. دوفصلنامه دانش و فناوری هوافضا 1401,