

|                  |                                                                                                                                                                                                                                                                    |
|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title            | 大規模言語モデルを用いたマルチモーダル検索モデルに基づく生活支援ロボットによる物体操作                                                                                                                                                                                                                        |
| Sub Title        |                                                                                                                                                                                                                                                                    |
| Author           | 是方, 諒介(Korekata, Ryōsuke)                                                                                                                                                                                                                                          |
| Publisher        | 慶應義塾大学AI・高度プログラミングコンソーシアム                                                                                                                                                                                                                                          |
| Publication year | 2024                                                                                                                                                                                                                                                               |
| Jtitle           | AICカンファレンス予稿集 (2024. ) ,p.10- 10                                                                                                                                                                                                                                   |
| JaLC DOI         |                                                                                                                                                                                                                                                                    |
| Abstract         | 本研究では，open-vocabularyな指示文に基づき，生活支援ロボットが環境中の画像群から対象物体および配置目標を検索することで，日常物体を指定された家具へ運搬するタスクを扱う．本論文では，大規模言語モデルを用いて埋め込み空間を切り替えることにより，単一モデルで対象物体および配置目標を検索可能なモデルを提案する．大規模環境における実画像データセットを用いた実験の結果，提案手法は標準的な評価指標において既存手法を上回った．さらに，標準化された屋内環境において実機実験を行い80%以上のタスク成功率を達成した． |
| Notes            | 会議名：AICカンファレンス2024<br>開催地：慶應義塾大学協生館AICラウンジ、および多目的教室<br>日時：2024年3月7日<br>ポスター発表要旨                                                                                                                                                                                    |
| Genre            | Conference Paper                                                                                                                                                                                                                                                   |
| URL              | <a href="https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO11003001-20240307-0010">https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO11003001-20240307-0010</a>                                                  |

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

# 大規模言語モデルを用いたマルチモーダル検索モデルに基づく生活支援ロボットによる物体操作

是方諒介

慶應義塾大学大学院 理工学研究科 開放環境科学専攻

**Abstract:** 本研究では、open-vocabulary な指示文に基づき、生活支援ロボットが環境中の画像群から対象物体および配置目標を検索することで、日常物体を指定された家具へ運搬するタスクを扱う。本論文では、大規模言語モデルを用いて埋め込み空間を切り替えることにより、単一モデルで対象物体および配置目標を検索可能なモデルを提案する。大規模環境における実画像データセットを用いた実験の結果、提案手法は標準的な評価指標において既存手法を上回った。さらに、標準化された屋内環境において実機実験を行い 80% 以上のタスク成功率を達成した。

**Keywords:** Domestic Service Robot, Object Retrieval, Large Language Model, Multimodal Foundation Model

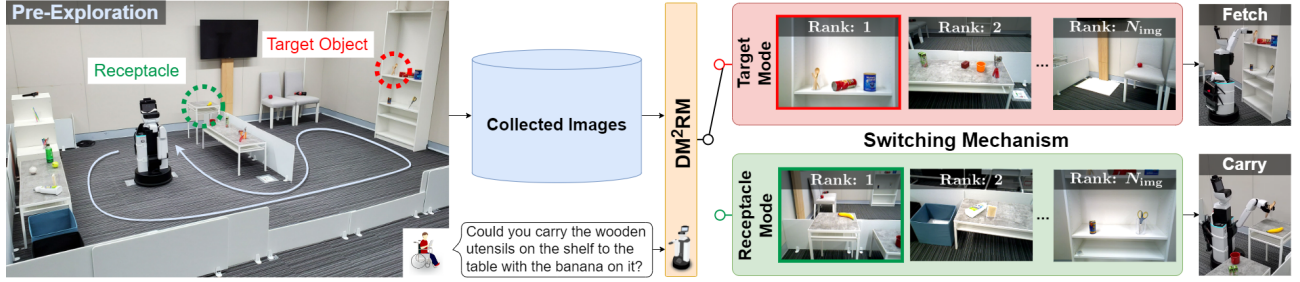


図1 提案手法の概要

## 1. 研究背景・目的

高齢化が進行する現代社会において、在宅介護者不足の解決策の一つとして生活支援ロボットが注目されている [6]。ロボットに対して自然言語で家事タスクを指示可能であれば利便性が向上するが、ロボットが人間の指示文を理解する能力はいまだ不十分である。本研究では、open-vocabulary なユーザ指示文に基づき、事前に収集された環境中の画像群から対象物体および配置目標を検索することで、日常物体を指定された家具へ運搬するタスクに焦点を当てる。

## 2. 方法

本論文では、対象物体および配置目標に関する予測を単一モデルで個別に行う方法でタスクを解くことが可能な Dual-Mode Multimodal Ranking Model (DM<sup>2</sup>RM) を提案する。図1に、提案手法の概要を示す。既存手法 (e.g., [1]) と異なる点は、マルチモーダル基盤モデル [3, 4] を用いて単一モデルで予測対象に応じたモード切り替えを行う Switching 機構を導入する点である。モードトークンおよび大規模言語モデルによる表現特定を活用することによりモードに応じて埋め込み空間を切り替えることで、各モードにおいて指示文と正解画像との類似度を高めることが期待される。

## 3. 結果

シミュレーション環境において新たに収集したデータセットおよび実機による評価を行った。前者は建物レベルの環境を数百規模で含む Habitat-Matterport 3D (HM3D) データセット [5] を基に構築し、後者は World Robot Summit [2] において標準化された家庭環境およびロボット [6] を用いた。

データセットを用いた実験の結果、提案手法は既存手法 (e.g., [1]) を画像検索タスクにおける標準的な評価指標である mean reciprocal rank (MRR) および Recall@K で上回った。表1に、実機実験の定量的結果を示す。open-vocabulary な指示文に基づく物体操作を行い、ゼロショット転移条件のもと 82% のタスク成功率 (SR) を達成した。

表1 実機実験における定量的結果

| MRR↑ [%] | Recall@10↑ [%] | SR ↑ [%]      |
|----------|----------------|---------------|
| 39       | 96             | 82 (82 / 100) |

## 4. 結論

本研究の貢献を以下に示す。

- モードトークンおよび LLM を用いた表現特定により予測対象に応じて埋め込み空間を切り替える Switching 機構を導入し、単一モデルで対象物体および配置目標の両方を個別に検索可能にする DM<sup>2</sup>RM を提案した。
- HM3D データセット [5] に基づき構築したデータセットにおいて、提案手法が既存手法を上回る結果を得た。
- 実機実験の結果、標準環境においてゼロショット転移という条件のもと 80% 以上のタスク成功率を達成した。

## 参考文献

- [1] Kanta Kaneda et al. Learning-To-Rank Approach for Identifying Everyday Objects Using a Physical-World Search Engine. *IEEE RA-L*, 9(3):2088–2095, 2024.
- [2] Hiroyuki Okada et al. What Competitions were Conducted in the Service Categories of the World Robot Summit? *Advanced Robotics*, 33(17):900–910, 2019.
- [3] OpenAI. GPT3-5Turbo, 2024. Accessed: Feb. 2024.
- [4] Alec Radford, Jong Kim, Chris Hallacy, et al. Learning Transferable Visual Models From Natural Language Supervision. In *ICML*, pages 8748–8763, 2021.
- [5] Santhosh Ramakrishnan et al. Habitat-Matterport 3D Dataset (HM3D): 1000 Large-scale 3D Environments for Embodied AI. In *NeurIPS*, 2021.
- [6] Takashi Yamamoto, Koji Terada, Akiyoshi Ochiai, et al. Development of Human Support Robot as the research platform of a domestic mobile manipulator. *ROBOMECH Journal*, 6(1):1–15, 2019.