

# 確率論理とマルチモーダル場所概念の統合による 確率的プランニング

長谷川 翔一 (立命館大学)

## 1 はじめに

サービスロボットにとって、物体の配置場所の確率である場所と物体の関係を理解することは、ユーザの日常生活を支援する際に重要である [1-7]. 特に不慣れた環境では、ロボットはこの情報を素早く利用して、特定の物体に関連したタスクを実行する必要がある. 例えばサービスロボットの文脈では、ユーザが要求する物体の探索や片付けなどのタスクは、場所と物体の関係を決定するために確率推論を活用することで達成できる. しかし新しい環境にロボットが導入されたとき、場所や物体の関係を迅速に推論してユーザを効果的にサポートするのは難しいことがある. さらに、同じ物体が異なる家庭内で異なる場所に存在する可能性があるため、事前に定義された常識的知識のみに依存することは、このばらつきに対処するためには不十分である. 新しい環境では、常識的知識だけでも、現場での学習だけでも不十分であり、両者を適切に組み合わせて場所と対象物の関係を理解することが重要である.

環境地図に意味付けを行うセマンティックマッピング [1, 8] では、現場環境における場所と物体の関係を推測するために様々な手法が提案されている [2, 4, 5, 9]. しかしこの方法では、多くのデータを必要とすることがしばしば課題となっている. 例えば古田らは、ロボットのセンサデータを長期間蓄積することで、物体の位置を確率的に表現する物体配置地図を考案した [2]. この地図は、物体ラベルと占有格子地図上の位置座標を関連付けることで、ロボットが地図を参照して物体の最も可能性の高い位置を予測することを可能にした. さらに谷口らは、環境地図上に物体ラベルを接地するだけでなく、ロボットが周辺環境から収集したマルチモーダル情報に基づいて場所の語彙を獲得し、場所をカテゴリゼーションする場所概念形成モデルを提案している [4, 5, 10-12]. このようなアプローチの一例として、場所概念の獲得と地図作成を容易にする、自己位置推定と地図作成を同時に行うオンライン場所概念・語彙習得 (SpCoSLAM) がある [10, 11]. 磯部らや谷口らは、ロボットが場所と物体の関係や空間概念を獲得するために、物体検出頻度を利用したモデルを開発し [4, 5], ロボットが物体のありそうな場所に関する情報を推論することを可能にした. しかし、これらの方法を用いて、異なる家庭環境内で物体の片付けなどの家事タスクを実行するためには、ロボットはセンサデータを再取得する必要がある. そのため、新しい家庭環境に導入された場合、ロボットは場所と物体の関係を学習するために、データ量の多さに直面した.

常識的知識と膨大なオンラインのデータセットを利用することで、現場環境における場所と物体の関係の学習時間を短縮するアプローチがいくつかある [3, 6, 7, 13-16]. 常識的知識とは、個人が一般的に持っている集成的な一般知識やルールを指す. 一方で現場学習から得られる知

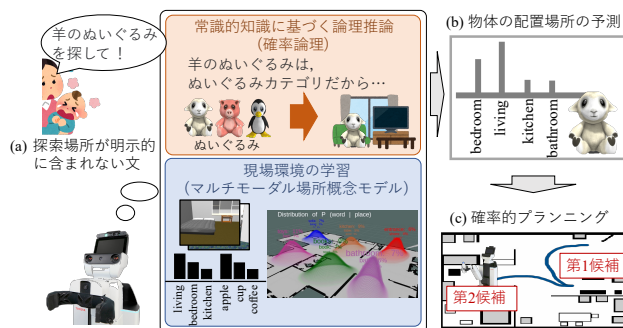


Figure 1: 我々の研究アプローチは以下のステップを取る: (a) ユーザがロボットに、羊のぬいぐるみのような配置場所が定義の物体を探すことを指示する. (b) ロボットは、ぬいぐるみカテゴリに関する常識的知識を用いて物体の配置場所を論理的に推論する. 同時に、探索対象物の配置される場所の確率を計算するために、ロボットの観測情報から構築された現場知識に基づく推論を使用する. (c) 算出された物体の配置場所の確率を用いて、ロボットは数回の探索で効率よく物体の配置場所を特定するために探索順序を計画する.

識は、ロボットが周囲の環境から収集したセンサデータの分析を通じて得た洞察に關係する. これらのアプローチの中で常識的知識を用いるには、データの収集や処理に時間と労力を要し、配置場所が未定義の物体を効果的に見つけることは依然として根強い課題である. ここで“配置場所が未定義な物体”とは、明示的に決定された場所を持たず、確率的な位置情報を持たない物体を示す. Zhang らは、オントロジーとオンラインの画像データを活用し、ロボットが効率的な物体探索タスクを行えるようにした [13]. 具体的には、画像データと物体と家具の位置関係 (例えば、 $isOn(Pillow, Sofa)$  は“枕がソファの上にある”ことを意味する) に基づいて、様々な場所にある物体の存在確率を作成した. しかしながらこのアプローチでは、データの取得と処理に多大な時間と労力を要し、オントロジー内であらかじめ定義されたオブジェクトの位置に依存していた. その結果、ロボットが配置場所が未定義の物体を特定しようとすると、困難に直面した.

配置場所が未定義な物体を特定するために、記号論理に基づく確率論理を活用するアプローチが提案されている [17, 18]. この枠組みにおける推論は、反駁による証明によって計算でき、事前知識を用いて未定義の物体の位置を特定するための事後確率を導出できる [19, 20]. 言い換えれば、ある家庭環境における羊のぬいぐるみの正確な位置や配置分布が不明な場合でも、ぬいぐるみカテゴリ内の他の物体の事前知識を活用することで、羊のぬ

いぐるみの位置を予測できる。別の研究では、Veiga らが確率論理とセマンティックマップを組み合わせた。これにより、彼らのロボットは配置場所が未定義の物体の発見に成功した [17, 18]。しかし彼らのアプローチでは、セマンティックマップは人間によって手動で提供された。

Veiga らから着想を得て、我々は確率論理とセマンティックマップを統合し、ロボットの観測に基づくセマンティックマップを構築することで、人手による地図構築の必要性を低減する。ロボットの観測結果を活用して、現場学習からセマンティックマップを学習し、常識的知識に基づく論理推論を採用して、配置場所が未定義の物体を特定することは、新しい環境における場所と物体の関係を理解するための効果的なアプローチである。しかし、配置場所が未定義の物体に関連する場所と物体の関係を学習するためには、新しい家庭環境においてロボットが必要とするセンサデータ量を考慮することが重要である。ロボットが部屋をくまなく探索し、センサデータを収集する方法は考えられるが、部屋全体のセンサデータを取得するには時間がかかる。

そこで本研究では、ロボットが場所と物体の関係を学習し、配置場所が未定義の物体を発見するために必要な訪問部屋数を決定する。本研究の目的は、新しい環境において特定の物体を探索する際に、現場学習と常識的知識に基づく論理推論を統合することで、場所と物体の関係を決定に伴う学習コストを軽減できるかを検証することである。この目的のために、我々は提案するアプローチと3つのアプローチを比較し、仮想的な家庭環境内で、配置場所が未定義なものを含む日常的な物体を探索するロボットの能力の有効性を評価した。本研究で比較した4つのアプローチは以下の通りである。1) 現場学習 (SpCoSLAM), 2) アンケート調査により獲得した常識知識 (Prior), 3) 現場学習と常識知識の組み合わせ (SpCoSLAM + Prior), 4) 現場学習と常識知識に基づく論理推論の組み合わせ (提案手法: SpCoSLAM + ProbLog)。各アプローチの選択は、ロボットの環境に対する意味理解、学習効率、配置場所が未定義な物体の推定への貢献度などの要因に基づいている。アプローチ4) は、文献 [17, 18] に関連し、確率論理のみを用いる従来手法に相当する。提案手法では、確率論理とマルチモーダル場所概念の組み合わせにより、論理推論の有効性を間接的に評価できるため、本研究では比較対象としていない。比較実験を通じて、提案手法の有効性、特に、ロボットが新しい家庭環境において、全ての物体の位置を特定するために異なる部屋を訪問する回数を減らす能力を検証する。

Figure. 1 は本研究で採用したアプローチを示している。はじめに、ロボットはユーザから、配置場所が未定義の羊のぬいぐるみなどの物体を探索するように指示される (Figure. 1 (a))。その後、ロボットは常識的知識に基づく論理的推論と現場知識に基づく推論によって物体の存在確率を計算し、得られた確率分布を統合する (Figure. 1 (b))。Figure .1 では、羊のぬいぐるみの配置は未定義であるが、ぬいぐるみカテゴリに属する他の物体に関する事前知識を用いて論理推論が行われる。これにより、配置場所が未定義の物体の存在確率を計算できる。学習により現場での知識が蓄積されると、ロボットは環境に関する知識を推論する能力を獲得する。最後に、計算された新しい環境での物体存在確率を活用することで、物体探索の反復回数を減らすことができる (Figure .1 (c))。コンピュータビジョンの分野では、Vision-and-Language

Navigation (VLN) [21, 22] や Object-Goal Navigation (ObjNav) [23, 24] として知られるタスクは、明示的な言語指示、物体ラベル、画像シーケンスを使用するエージェントを含む。我々のタスクはVLNとは異なり、検索指示は場所を特定せず、物体の名前のみを提供する。さらに、ObjNavは各物体の探索ごとに地図を初期化することで、各物体を独立に評価する。対照的に、ロボットが各物体の位置を学習した後、我々の評価は探索のために全ての物体を含み、新規環境における物体の効率的な探索に焦点を当てる。

本研究は以下の3つの貢献から構成される。\*

1. 新しい環境において、常識的知識に基づく論理的推論と現場学習に基づく推論を組み合わせることで、位置が未確定の物体を効率的に探索できることを示す。
2. ロボットが新しい環境において、位置が未確定な物体を含む複数の物体探索タスクを実行した場合、本手法は他の3つの手法と比較して学習コストを1.6倍に削減できることを示す。
3. 常識的知識に基づく論理的推論と現場学習を統合し、新たな実環境における定性的分析により、どのような場合にタスクが成功・失敗するかを示す。

本研究の構成は以下の通りである：2章では関連研究の概要を述べる。3章では、提案手法の本質的な構成要素である確率論理と場所概念モデルについて詳述する。4章では、常識的知識に基づく論理推論とマルチモーダルな場所概念を統合した提案手法を紹介する。5章では、提案手法の有効性をシミュレーション環境内で評価するために行った実験の概要を述べ、得られた結果についての考察を行う。6章では、実環境における提案手法の有効性を検証するために行った実験について、実験結果の定性的評価とともに述べる。7章では、研究の結論と今後の研究の方向性を示す。

## 2 関連研究

この章では場所と物体の関係に関する関連研究を紹介し、それらとの関係における我々の方法の立ち位置を示す。

**2.1 現場環境における学習** 物体や空間のカテゴリゼーションを含む意味情報を環境地図として帰属させるタスクは、セマンティックマッピングと呼ばれる [1, 8]。多くの研究者がセマンティックマップを生成し、物体の位置に関する知識を獲得できるサービスロボットの構築を試みている [1, 2, 4, 5, 8, 9]。例えば古田らは、ロボットのセンサデータを長期間蓄積し、物体の位置を確率的に示す確率的物体配置地図を考案した [2]。収集されたセンサデータは、多変量混合ガウス分布に適用され、分布の平均値が確率的な物体の位置を示す。環境地図内の物体カテゴリを根拠づけることに加えて、ロボットの観測から空間のカテゴリゼーションと場所の語彙獲得を同時に行う取り組みがある [4, 5]。これは記号創発ロボティクスの分野 [26, 27] で行われており、環境との相互作用か

\*本論文は長谷川翔一を筆頭著者として、立命館大学の谷口彰、萩原良信、El Hafi Lotfi、谷口忠大と共同で 2023 IEEE/SICE International Symposium on System Integrations (SII 2023) で発表した内容 [25] に追加実験を加えて再構成したものである。その他の情報は <http://www.em.ci.ritsumei.ac.jp/jp/information/detail/1314079/> を参照されたい。

ら得られる感覚入力を通じてロボットが概念と言語を獲得するプロセスをモデル化し、検証している。例えば谷口らは、場所と物体の関係、物体の位置からの場所の領域、場所を表す画像における物体の出現頻度、場所の語彙を学習する場所概念獲得モデルを開発した。このモデルは、ベイジ理論に基づく確率的生成モデルを用いて構築されており、ロボットからのセンサデータを確率分布で表される特徴量に変換してモデルに入力している [5]。このモデルのクラスタリングによって形成されたクラスタは、片付けなどのサービス指向のタスクを行う場所のカテゴリ知識として利用された。

従来のアプローチは、ロボットが特定の環境内で大量のセンサデータを収集し、物体と場所の関係性を学習することを前提としている。しかし、ロボットを新しい環境に導入する場合、データを収集し直す必要があり、時間がかかる。そこで本アプローチでは、常識的知識と現場学習を組み合わせることで、ロボットの新しい環境への適応を促進することを目指す。

**2.2 大規模データセットから構築した常識的知識** 現場環境に必要なセンサデータの量を減らすために、1つのアプローチは、オンラインでアクセス可能なデータセットから構築された常識的知識を採用することである [3, 6, 7, 13–16]。Zhang らは、オントロジーに基づく共通感覚知識を持つロボットが効率的な物体探索タスクを実行できるようにした [13]。彼らは、オントロジー内の記号表現を使用して、物体と家具の関連付けを定義した。さらに、彼らは場所を表す画像データセットを使用し、画像データ内に注釈された物体ラベルの出現頻度に基づいて計算された、各部屋に物体が存在する確率をロボットに提供した。一方で Hernandez らは、部屋のレイアウトに関する幾何学的情報と、画像データセットから得られる物体の存在確率をロボットに与えた [6]。これにより、ロボットはピクセル単位で、地図上の物体がどの視点から位置するかを予測することが可能となった。

これらの先行研究では、ロボットにあらかじめデータセットや手動で生成した場所と物体の関係性に関する知識を与えることで、効率的に知識を獲得することができた。しかしこれらの手法では、データの収集や処理に時間と手間がかかり、正確な位置が不明な環境中の物体の位置を迅速に特定することが困難であった。そこで本アプローチでは、常識的知識に基づく論理的推論と現場知識に基づく推論を統合することで、物体の存在が確定していない新たな環境において、迅速に物体の位置を特定することを目指す。

**2.3 常識的知識に基づく確率的論理推論** 配置場所が未定義な物体をロボットが特定できるようにするために、記号論理に基づいた確率的論理推論アプローチが提案された [17, 18]。Veiga らは、確率論理を用いて環境の意味マップを表現した [17, 18]。具体的には、彼らのロボットは、物体クラス、物体サイズ、各部屋に物体が存在する確率を含む常識的知識を備えていた。また、ロボットの意思決定プロセスをモデル化するために部分観測マルコフ決定過程 (POMDPs) を採用し、物体検出の出力を2値で表現し、POMDPsの枠組みで観測情報として使用した。これらの手法を組み合わせることで、位置が未定義の物体が存在する事後確率を生成し、ロボットが環境中の位置が未定義の物体を効率的に探索できるようにした。

しかし Veiga らによって提案されたアプローチでは、現場の環境を表す人手で構築されたセマンティックマップが採用されていた。これに対し我々のアプローチでは、セマンティックマップと常識的知識に基づく確率的論理推論を統合しているため、ロボットの経験データに基づいてセマンティックマップが構築される。

**2.4 Object-Goal Navigation** コンピュータビジョンの分野では、エージェントによる空間の意味理解を扱う VLN [21, 22] に関する多くの研究が行われてきた。具体的にはこれらの研究において、エージェントは、家庭環境を再現するように設計されたシミュレータ環境の部屋間を移動しながら、明示的な言語命令と画像シーケンスを使用して学習し、指示された目的地に到達した。VLNのサブカテゴリである ObjNav [23, 24] は、物体探索命令で指定された物体ラベルに対応する環境内の物体を探索するもので、ObjNav は広く研究されている [28–31]。福島らは、Object Scene Memory と呼ばれる長期的なシーンや物体の意味情報を記憶する機構と、Transformer [32] という時系列データを扱うモデルを用いて、ObjNav の性能を向上させた [28]。ObjNav に関する最近の研究では、Majumdar らが基盤モデル [33] である CLIP [34] を用いたゼロショット ObjNav を実証している。CLIP の視覚エンコーダとテキストエンコーダは、それぞれ現場環境での学習と物体探索タスクの実施に使用された。

ObjNav では各物体探索は独立して評価され、タスクごとに地図が初期化される。一方で我々の研究は、まずロボットに場所と物体の関係性を学習させ、次に探索するすべての物体を同時に評価することで、新しい環境における複数の物体の効率的な検索に焦点を当てた。

## 3 準備

本節では、提案手法の構築に用いた確率論理と場所概念獲得モデルについて説明する。

**3.1 確率論理** 確率論理とは、記号論理学における述語論理に確率値を割り当てたものである [19, 20]。例えば、寝室に羊のぬいぐるみが存在する確率が40%であることを確率論理で表現するには、述語記号と定数記号を使って述語論理式の “0.40::exist(sheep\_doll, bedroom)” で表現できる。確率論理の実用的な応用は質問に答えることであり、“羊の人形はどこにあるか” といった質問応答は述語論理式を用いれば解決できる。確率論理の述語論理式で事前知識が節に変換され、記号論理学における単一化と導出を繰り返し、反駁による証明を採用することで解が導かれる。確率的論理型プログラミング言語である Probabilistic Prolog (ProbLog) [35] や Programming in Statistical Modeling (PRISM) [36] は、確率論理を扱うためのツールである。これらの言語は、佐藤らによって提案された分布意味論 [37] に基づいて構築されており、述語論理の規則と確率分布の割り当てに基づいて、解決されたクエリを確率分布で提供する。ProbLog は分布意味論に基づき、式 (1)–(2) を用いて問い合わせに対する解の確率を計算する。ここで  $T$  は ProbLog で定義される論理プログラム、 $T = \{p_1 : c_1, \dots, p_n : c_n\}$  を表す。 $c_i$  は節を表し、 $p_i$  は  $c_i$  に割り当てられた確率を表す。 $M$  は  $T$  集合内の様々な節の組み合わせから導かれる論理プログラムの集合を示す。 $L_T$  は、 $T$  集合から導出可能な全ての実現可能な論理プログラムの集合を表し、 $M$  は  $L_T$  の部分集合である。 $\theta$  は論理導出や推論の過程で発生す

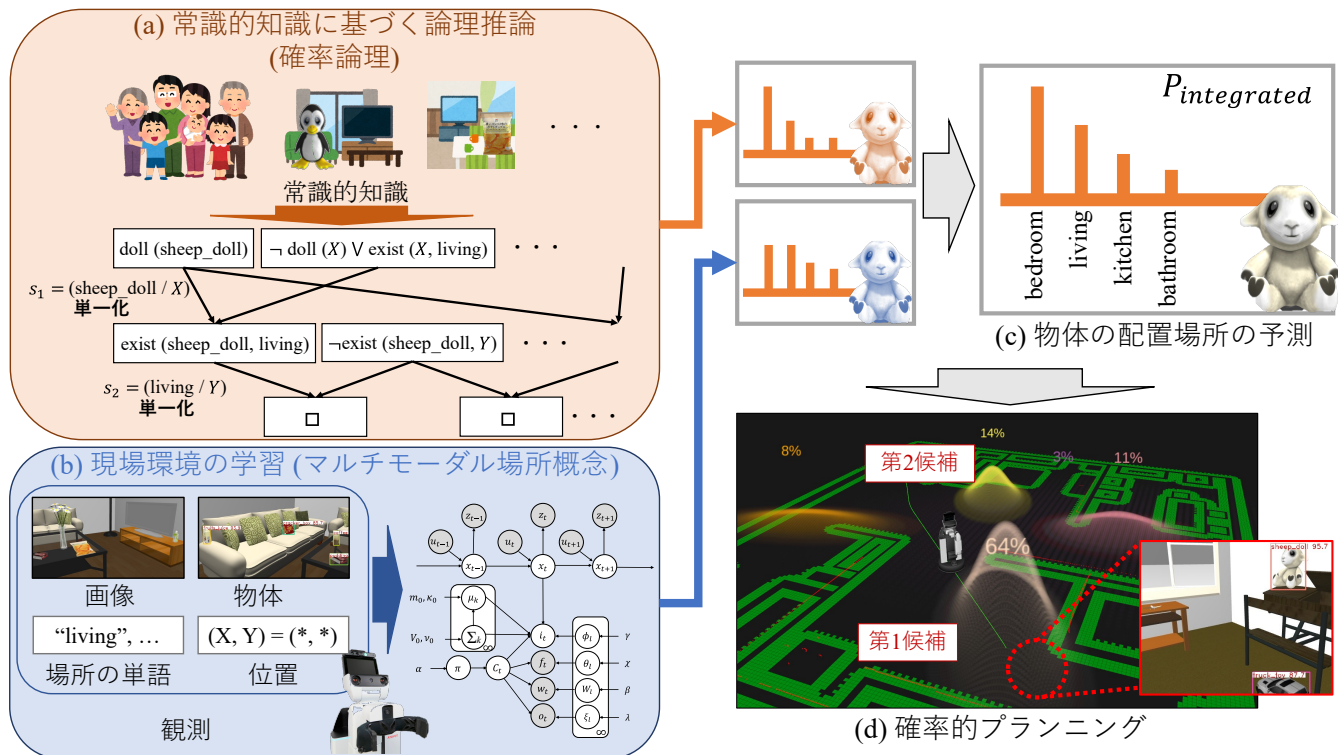


Figure 2: 提案モデルの概要: (a) 常識的知識を用いた確率的論理推論. (b) マルチモーダル場所概念を使用し, 物体ラベル  $o_t$  から場所の単語  $w_t$  のクロスモーダル推論. (c) (a) と (b) で得た確率の加重平均による物体の存在確率の計算. (d) 物体の存在確率の高い順にロボットが確率的に物体を探索.

る項を表し, 変数の割り当てを表す.  $q$  は特定の問い合わせを表す.

$$P(q|M) = \begin{cases} 1 & \exists \theta : M \models q\theta \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$P(q|T) = \sum_{M \subseteq L_T} P(q|M) \prod_{c_i \in M} p_i \prod_{c_i \in L_T \setminus M} (1 - p_i) \quad (2)$$

私たちのアプローチは, ProbLog を使用した確率論理を採用し, 家庭内の未定義の場所にある物体を効率的に見つけるものである.

**3.2 場所概念モデル** 谷口らは, 場所概念を場所に関するマルチモーダル情報に基づく教師なし学習によってロボットが形成する確率分布上のパラメータの集合と定義した. 彼らは人間の知能を理解する試みとして, バイズ理論に基づく確率的生成モデルに基づく場所概念モデルを提案している [4, 5, 10, 11]. 代表的なモデルは SpCoSLAM [10] で, 場所概念の学習, 場所の語彙獲得, 地図の作成を同時に行う. SpCoSLAM への入力情報には, 場所を説明する話し言葉, 場所を表す画像, ロボットの自己位置が含まれる. このマルチモーダル情報を活用して, SpCoSLAM は場所のカテゴリゼーションを行い, マルチモーダルな場所概念を学習する. SpCoSLAM は Rao-Blackwellized パーティクルフィルタに基づく FastSLAM2.0 [38] から構成されるため, 場所概念のモデルパラメータは地図作成中に逐次更新することができる. SpCoSLAM の生成過程とアルゴリズムの詳細については, 文献 [10] を参照

されたい. SpCoSLAM の配置推定精度を向上させるために, 谷口らは SpCoSLAM 2.0 を提案し [11], SpCoSLAM のオンライン学習アルゴリズムに固定ラグ活性化を活用した.

本アプローチでは SpCoSLAM 2.0 を採用し, 現場学習のための場所のカテゴリゼーションを行う. さらに, ロボットが場所と物体の関係性を学習するために, 場所を表す画像中の物体の検出頻度を観測情報として SpCoSLAM 2.0 に追加した. 前節で説明した確率論理と SpCoSLAM 2.0 の組み合わせにより, ロボットは少ない学習回数で, 新しい環境における空間の意味を理解することが可能となる.

## 4 提案手法

我々は常識的知識に基づく確率的論理推論と現場学習に基づく推論を統合し, 新しい環境において配置が未定義の物体を効率的に探索するモデルを提案する. Figure. 2 に提案モデルの概要を示す.

**4.1 常識的知識を用いた確率的論理推論** 本研究のロボットは, 新しい環境における物体の位置を論理的に推論するために, 物体, 場所, それらの関係性に関する常識的知識を備えている. 家庭内における物体の位置に関する情報を収集するためにアンケートを実施し, その回答から常識的知識を構築した. アンケートでは, 参加者はある部屋に一般的に配置されている複数の物体を選択するよう指示された. 常識的知識の構築に続いて, 確率的論理推論を行った.

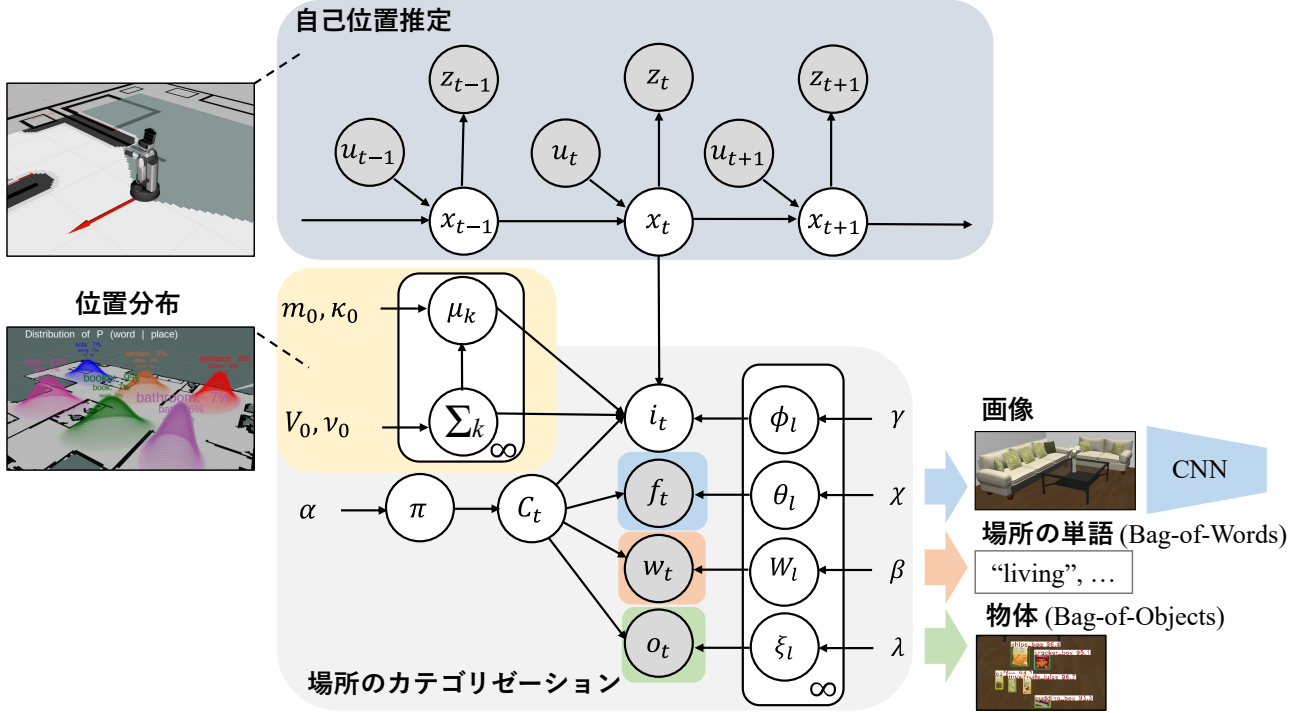


Figure 3: 本研究で使用した場所概念モデルのグラフィカルモデル.

Table 1: 場所概念モデルの変数定義.

変数	定義
$x_t$	ロボットの自己位置情報
$z_t$	距離データ
$u_t$	制御データ
$C_t$	場所概念のインデックス
$i_t$	位置分布のインデックス
$w_t$	場所の単語の特徴量
$f_t$	場所を表す画像の特徴量
$o_t$	物体検出の頻度特徴量
$\mu, \Sigma$	多次元ガウス分布のパラメータ
$\pi$	カテゴリ分布のパラメータ
$\phi_l, W_l, \theta_l, \xi_l$	多項分布のパラメータ
$\alpha, \beta, \chi, \gamma, \lambda$	事前分布のハイパーパラメータ
$m_0, \kappa_0, V_0, \nu_0$	

**4.2 マルチモーダル場所概念を用いたクロスモーダル推論** また、ロボットが環境中で収集したマルチモーダル情報に基づく場所と物体の関係性を取得するため、SpCoSLAM 2.0 に観測情報として物体の頻度データを取り込み、モデルを拡張した。本研究で用いた場所概念モデルのグラフィカルモデルと変数定義を Figure. 3 と Table. 1 に示す。さらに、SpCoSLAM 2.0 から変更した生成過程を式 (3)– (5) に記述する。ここで、Mult(-) は多項分布、Dir(-) はディリクレ分布を表す。

$$w_t \sim \text{Mult}(W_l) \quad (3)$$

$$o_t \sim \text{Mult}(\xi_l) \quad (4)$$

$$\xi_l \sim \text{Dir}(\lambda) \quad (5)$$

Table. 1 に示した  $\{W, \mu, \Sigma, \theta, \xi, \phi, \pi\}$  と  $\{C_t\}$  は場所概念モデルのパラメータである。 $\{W, \theta, \xi\}$  は観測特徴の生成を担う確率分布のパラメータ、 $\{\mu\}$  と  $\{\Sigma\}$  は場所の領域を構成する確率分布のパラメータであり、 $\{\phi\}$  と  $\{\pi\}$  は位置分布と場所概念インデックスに関連する確率分布のパラメータである。場所概念モデルに内在するこれらのパラメータは、4.3 節と 4.4 節で物体の配置場所を推論するために使用された。さらに、一旦地図が構築されると、これらのパラメータは、粒子フィルタを用いて事後分布からサンプリングされた場所概念と位置分布インデックスを活用することによって推定された。詳細はこの文献 [11] を参照されたい。

クロスモーダル推論とは、あるモダリティの情報を別のモダリティから推論することである。本研究では、マルチモーダル情報を用いてロボットが学習した場所概念モデルのパラメータから、物体の存在確率を推定するためにクロスモーダル推論を採用した。具体的な計算式は式 (6) になり、場所概念モデルの物体情報  $o_t$  から位置を示す単語情報  $w_t$  へのクロスモーダル推論を行うことで、各物体の具体的な位置を推論することが可能となる。

$$P(w_t | o_t, \xi, W, \pi) \propto \sum_{C_t} P(w_t | W_{C_t}) P(o_t | \xi_{C_t}) P(C_t | \pi) \\ = \sum_{C_t} \text{Mult}(w_t | W_{C_t}) \text{Mult}(o_t | \xi_{C_t}) \text{Mult}(C_t | \pi) \quad (6)$$

場所概念モデルの入力として、bedroom などの場所名をテキスト形式で表現したものを用いた。さらに単語情報を捉えるために、Bag-of-Words (BoW) 特徴ベクトル表現形式を採用した。画像情報には、事前学習済みの畳み込みニューラルネットワーク (CNN) モデル Places365-ResNet [39] の最終層で生成された 365 次元の特徴量を用

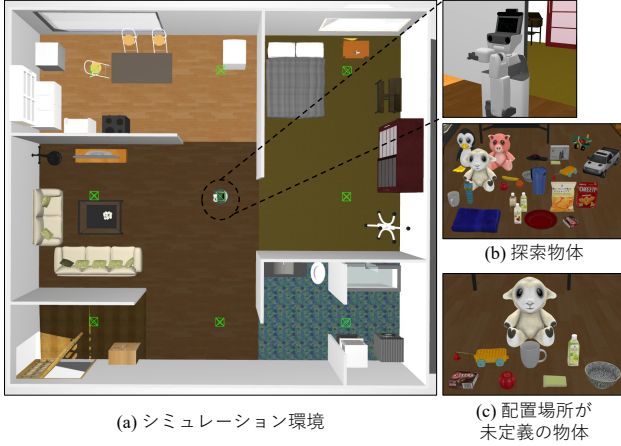


Figure 4: 実験1で使用するシミュレーション環境と物体.

いた。最後に、物体の頻度情報を表現するために、Bag-of-Objects (BoO) 特徴量を使用した。

**4.3 確率論理とマルチモーダル場所概念の統合** 我々は2つの推論の結果を加重平均することで統合した(式(7)): 常識的知識を用いた確率的論理推論と、現場知識に基づくクロスモーダル推論である。ここで、 $P_{integrated}(w_t|o_t)$  は推論結果を統合した後の物体の存在確率を表し、 $P_{logic}$  は論理推論から得られた確率を表し、 $\eta$  は加重平均に割り当てられた重みを表す。重み  $\eta$  の値は経験的に決定された。

$$P_{integrated}(w_t|o_t) = \eta P_{logic} + (1 - \eta)P(w_t|o_t, \xi, W, \pi) \quad (7)$$

**4.4 確率的プランニング** 我々のアプローチでは、ロボットは物体の存在確率を考慮した確率的なプランニングを行い、これは常識的な知識に基づく確率的論理推論と現場学習に基づく推論を統合して導き出されたものである。場所概念モデルでは混合ガウスモデルを採用し、物体の位置の空間的分散を捉えた位置分布を構築した。各位置分布にはインデックスを  $i_t$  を割り当てた。確率的プランニングには、物体の頻度情報  $o_t$  が与えられたときの位置分布インデックス  $i_t$  の条件付き確率を式(8)を用いて計算した。そして、ロボットは確率の高い順に各場所を訪問した。Figure 2 (d) に確率分布を示す。確率的プランニングに用いた確率分布は式(8)を用いて以下のように計算した：

$$\begin{aligned} P_{integrated}(i_t|o_t) &= \sum_{w_t} P(i_t|w_t)P_{integrated}(w_t|o_t) \\ &= \sum_{w_t} \left\{ \sum_{C_t} \text{Mult}(i_t|\phi_{C_t})\text{Mult}(w_t|W_{C_t})\text{Mult}(C_t|\pi) \right\} \\ &\quad \left\{ \eta P_{logic} + (1 - \eta)P(w_t|o_t, \xi, W, \pi) \right\} \quad (8) \end{aligned}$$

## 5 実験1：シミュレーション環境

実験1の目的は、配置場所が未確定の物体をロボットが探索する際に、場所の学習コストをどの程度まで削減

できるかを検証することである。比較手法には、Table 2 に示す4つの方法を用いた。4つの手法の評価では、場所の学習ごとにロボットが物体の位置を特定するために必要な部屋の訪問回数を求め、評価項目とした。

**5.1 実験環境** 実験のセットアップでは、OSにUbuntu (22.04 LTS ver) を使用したROS (Noetic ver.) [40] フレームワークに依存するロボット用ソフトウェア開発環境 [41] を採用した。さらに仮想の家庭環境を作成し、Gazebo [42] を使用してロボットをシミュレーションした (Figure 4)。ロボットはHuman Support Robot (HSR) [43] を使用した。物体のデータセットは、YCB [44]、Google Scanned Objects [45]、Unity Asset Store のライセンスフリーの物体<sup>†</sup>、そして我々が独自に作成した物体から構成されている。

24個の物体は、4つの部屋 (living, bedroom, kitchen, bathroom) のうちの1つ部屋の事前に定義された3次元の位置に配置された。Figure 5 に示すように物体の配置を決定した。さらに24個の物体を8個のカテゴリに分類した。これらのカテゴリのうち、ロボットが物体の存在確率に関する常識的知識を持たない物体として、各カテゴリから1個ずつ、計8個の物体を選択した。物体探索タスク中、ロボットはcupのような特定の物体ラベルを探索する言語指示を受けた。

**5.2 実験前準備** 物体探索タスクの開始前に、ロボットは家庭環境の占有格子地図を取得しているものと仮定した。地図作成にはHector SLAM [46] を採用した。環境内でロボットを遠隔操作し、HSRのLiDARセンサから得られた距離計測値を用いて地図作成を行った。さらにロボットには各部屋の領域の情報が与えられ、占有格子地図上にその空間領域が表現された (Figure 5)。24個の探索対象の物体が環境内に配置されていない状態で、living, bedroom, kitchen, bathroomといった場所の単語を用いて、場所概念を学習するSpCoSLAMを使用することで実現した。

物体検出器として、合計24個の物体ラベルを検出できるYou Only Look Once version 5 (YOLOv5) [47] を使用した。YOLOv5を訓練するために、訓練画像と検証画像それぞれ11,575枚、100枚からなるデータセットを使用した。バッチサイズは8であり、事前学習された重み (yolov5m<sup>‡</sup>) を用いて30エポックにわたって学習が行われた。

場所や物体の名前に関する常識的知識として、場所にはkitchen、物体にはdoll(pig-doll)といった表現を取り入れた。確率論理に記述された各場所に物体が存在する確率は、63名の参加者を対象としたアンケート調査の結果に基づいて決定された。アンケートでは、事前選択された24個の物体が存在する可能性のある部屋を回答してもらった。物体の存在確率を計算するために、ディリクレ事前分布を設定し、各場所における物体数を全参加者の回答に基づいて平滑化した。また提案手法では、カテゴリを考慮した場合の各部屋における物体の存在確率を扱う述語論理式を採用した (すなわち、“Xがぬいぐるみのカテゴリである場合、各部屋におけるその存在確率”を表す述語論理式)。この述語論理式に割り当てられる確率は、5.1節で述べたように、配置場所が未

<sup>†</sup><https://assetstore.unity.com/>

<sup>‡</sup><https://github.com/ultralytics/yolov5#pretrained-checkpoints>

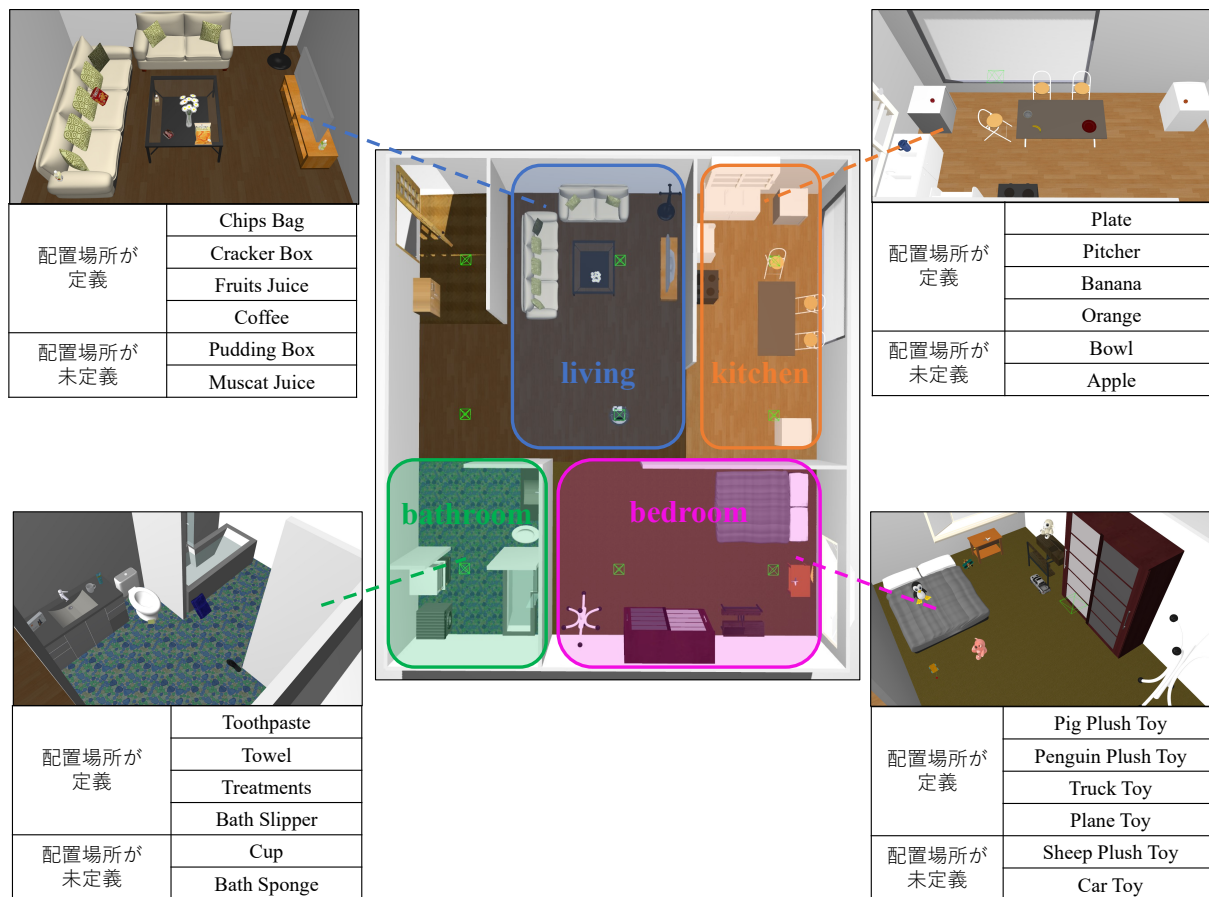


Figure 5: 各部屋における物体の配置 (実験 1). 各部屋の領域は色の付いた四角形で示されている. 各部屋に配置される物体は表の通りである. 各表の“定義”という用語は, ロボットが各部屋に物体が存在する確率に関する情報を持つことを示し, ロボットにとって既知の物体を表す. 一方で, “未定義”は配置場所が事前に定義されていない物体を示し, ロボットにとって未知の物体を表す.

定義の物体を除いたアンケート結果を用いて計算した.

**5.3 場所と物体の関係性の学習** 場所と物体の関係性を学習するために, 各部屋に 24 個の物体を配置した (Figure .5). その後, 場所概念獲得モデルは, living, bedroom, kitchen, bathroom の順に, 各部屋における物体の観測を含む場所概念を学習した. この学習過程は 5.2 節で詳述したように, 空室において獲得した場所概念の上に構築された. 場所と物体の関係性を学習するために, 4 つの部屋について 1 サイクルで合計 12 回の学習を行った. 各学習では, ロボットのカメラで撮影された 1 枚の画像, モンテカルロ位置推定から得られたロボットの 2 次元位置, 場所を表す画像内での物体検出頻度などのデータを組み合わせて観測データとして用いた.

**5.4 比較手法** 実験では, 提案手法を含む Table. 2 に示す 4 つの手法の比較を行った. 各手法はロボットに実装し, 24 物体の探索タスクの実行により評価した. 以下に各手法の詳細な説明を行う.

1. SpCoSLAM: 4.2 節で説明した SpCoSLAM に基づくマルチモーダル場所概念獲得モデル.
2. Prior: アンケート回答から得られる平滑化された常識知識のみを用いる方法.
3. SpCoSLAM + Prior: 前の 2 つの手法で得られた物体存在確率の加重平均を計算する手法であり, 重

みの値は 0.35 である.

4. SpCoSLAM + ProbLog: SpCoSLAM + Prior に確率的論理推論を統合した提案手法で, 重みの値は 0.35 である.

SpCoSLAM のハイパーパラメータは以下のように設定した:  $\alpha = 18.5$ ,  $\beta = 0.1$ ,  $\chi = 0.1$ ,  $\gamma = 0.1$ ,  $m_0 = [0, 0]^T$ ,  $\kappa = 1.0$ ,  $V_0 = \text{diag}(2, 2)$ ,  $\nu_0 = 3.0$ , and  $\lambda = 0.1$ . 場所概念モデルにおける粒子数は 30 で使用し, 固定ラグ活性化のラグ値は 10 である.

物体探索タスク中, ロボットは最も尤度の高い粒子を選択し, その粒子に関連する場所概念に基づいて物体の存在確率を推測した. 物体が配置されていそうな場所は, 4.4 節で詳述するように, 式 8 を用いて推定した. 本研究では, ロボットは場所の確率的な位置を表すガウス分布の平均値に移動すると仮定した. 目的地に到着したロボットは, 物体を検出するために頭部と基部を回転させた. 移動と物体検出の手法は, 4 つの手法とも同じものを採用した.

**5.5 評価項目** 我々は本アプローチの有効性を評価するために, 複数の評価項目を採用した. 1) 4.4 節で説明し, Figure .6 に示すように, ロボットが目標物体を見つける前に訪問した部屋の数. 2) 5.3 節で説明し, Figure .6 と Table. 3 に示すように, ロボットが場所と物体の関係

Table 2: 比較手法

比較手法	現場学習	常識的知識	論理推論
SpCoSLAM	✓	-	-
Prior	-	✓	-
SpCoSLAM + Prior	✓	✓	-
SpCoSLAM + ProbLog (proposed)	✓	✓	✓

Table 3: 評価結果. 表中の数値は, 1回の探索でロボットが発見した物体の割合を示す.

Rooms visited during learning	0	1	2	3	4	5	6	7	8	9
SpCoSLAM	0.33	0.13	0.46	0.46	0.54	0.54	0.54	0.54	0.83	0.67
Prior	0.50	<b>0.54</b>	0.50	0.50	0.50	0.42	0.46	0.46	0.50	0.50
SpCoSLAM + Prior	0.50	0.38	0.63	<b>0.71</b>	0.63	0.75	0.75	0.67	<b>0.92</b>	0.83
SpCoSLAM + ProbLog (proposed)	<b>0.54</b>	0.50	<b>0.71</b>	0.67	<b>0.75</b>	<b>0.88</b>	<b>0.79</b>	<b>0.79</b>	<b>0.92</b>	<b>0.92</b>

10	11	12
0.83	0.83	<b>0.96</b>
0.42	0.38	0.42
<b>0.92</b>	<b>0.92</b>	<b>1.00</b>
0.88	<b>0.92</b>	<b>1.00</b>

性に関する知識を得るために訪問した部屋の数. 3) 1回の探索でロボットが発見した物体の割合. ロボットは家庭環境内の24個の物体それぞれについて探索を行い, その後, 訪問・探索した部屋数の平均と標準偏差を計算した. ロボットが物体を発見した割合は,  $\frac{1}{N} \sum_i n_i$  を使用して計算した. ここで  $N$  は家庭環境内に位置する物体数を表し,  $n_i$  はロボットによって見つかった物体数を表す. 本研究では, 満足閾値として物体発見率を0.8とした.

また物体の誤認識により物体が発見された場合や, 探索タスク中に探索対象が検出されなかった場合は, タスク失敗とみなし, タスク成功率から除外した.

**5.6 実験結果** Table. 3は, 探索タスクのときに24個の物体を1部屋の訪問でロボットが発見した物体の比率を示している. ロボットが学習時に訪問した部屋数が5回の場合, 提案手法であるSpCoSLAM + ProbLogは0.88の探索精度を達成し, 24個中21個の物体の発見に成功した. SpCoSLAM + Priorは0.88を超える精度を得たが, これは学習時に8回以上部屋を訪問した場合のみであった. これらの結果は, 提案手法に確率的論理推論を組み込むことで, 他の手法と比較して, 場所と物体の関係性に関する知識を獲得するための学習コストを1.6倍以上削減できることを示している.

Figure .6は, ロボットが対象物を発見するために探索した部屋数を示している. 学習時にロボットが物体を観測しなかった場合(学習時に訪問した部屋数が0の場合), SpCoSLAMのパラメータ $\xi$ は一様分布となった. その結果, ロボットがランダムに各場所を訪問したため, 平均訪問回数は2.38回となり, 標準偏差は1.18と大きな値となった. 一方Priorでは, ロボットは場所の学習時をせずに, 場所と物体の関係性に関する常識的知識を活用した. その結果, 特に学習時に探索する部屋数が少ない場合, 対象物を見つけるのに必要な訪問回数はSpCoSLAMよりも少なかった. 特にSpCoSLAM + ProbLogでは, 配置場所が未確定な物体を常識的知識に基づく論理推論で正確に推論した. その結果, 平均訪問回数は1.5回とな

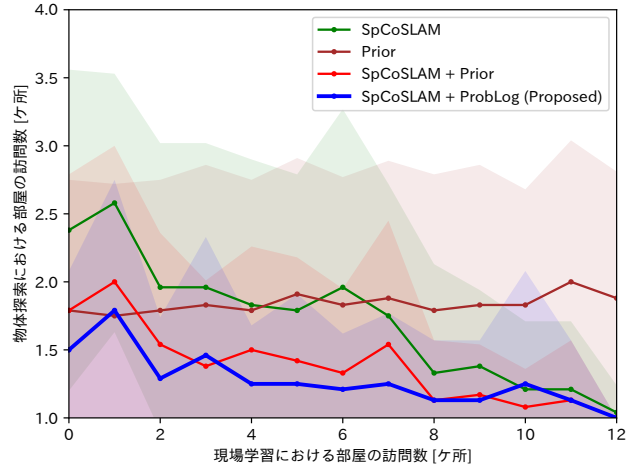


Figure 6: 各手法における24物体を探すタスクの結果: 縦軸は探索中にロボットが探索した部屋数を表す. 横軸は現場環境の学習段階でロボットが訪問した部屋数を表す. 実線は部屋の訪問数の平均値を示し, 線の色と手法の対応は, SpCoSLAM (緑), Prior (茶), SpCoSLAM + Prior (赤), SpCoSLAM + ProbLog (青)である. また色付きの透明な面は, 部屋の訪問数の標準偏差を表す.

り, 他手法と比較して少ない訪問回数で物体を発見した. SpCoSLAMに基づく手法は, 学習時における訪問した部屋数が増加するにつれて, 物体探索時に訪問した部屋数が1回に収束することをFigure .6で示している. これは, 物体検出の頻度を表すパラメータ $\xi$ が, 各部屋における物体の観測回数が増えるにつれて, 環境に適応するようになったためと考えられる.

## 6 実験2: 実環境

実験2では, ロボットが物体の配置場所を特定するために, 現場知識に基づく推論と常識的知識に基づく論理





Figure 7: 各部屋における物体の配置 (実験 2). 各部屋の領域は色の付いた四角形で示されている. 各部屋に配置される物体は表の通りである. 各表の“定義”という用語は, ロボットが各部屋に物体が存在する確率に関する情報を持つことを示し, ロボットにとって既知の物体を表す. 一方で, “未定義”は配置場所が事前に定義されていない物体を示し, ロボットにとって未知の物体を表す.

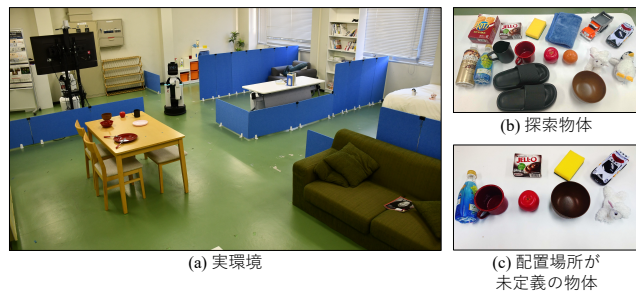


Figure 8: 実験 2 で使用する実環境と物体.

推論を統合する有効性を, 具体的な事例を通して定性的に検証した. 物体の存在確率は占有格子地図上に表示され, 各推論における物体探索の成功シナリオと失敗シナリオを定性的に分析する.

**6.1 実験環境** 物体探索タスクでは, YCB オブジェクト [44] と日常的な物体の組み合わせからなる 16 個の物体リストを準備した. ロボットは 16 個の物体のうち 8 個については, 事前に定義された配置場所を教えられなかった. しかし残りの 8 個の物体については, ロボットは各場所における物体の配置分布の情報を持っていた. これらの確率は, 5.2 節で説明したアンケート調査の結果から導き出されたものである. Figure .8(a)~(c) は, 家庭環境を模した実験環境と, 配置場所が定義された物体, 定義されていない物体を示す. これらの実験で使用したソフトウェア開発環境とロボットは, 実験 1 と同じである.

また, 探索対象の配置方法, 物体のカテゴリ数, 配置

場所が未定義の物体の決め方, 探索方法も実験 1 と同様である. Figure .7 に探索対象の配置を示す.

**6.2 実験前準備** 物体探索タスクの前に, 占有格子地図上の各部屋の領域を表す情報をロボットに与えるなどの条件は, 実験 1 と同じである. 物体検出手法は YOLOv5 [47] を使用し, ロボットは 16 個の物体クラスを検出できる. 訓練データセットと検証データセットは, それぞれ 19,710 枚, 657 枚の画像から構成された. バッチサイズは 8 を使用し, エポック数は 30 であり, 事前学習された重み (yolov5m §) で初期化された. 確率論理によって定義された常識的知識は, 実験 1 と同じである. 最後に, 場所と物体の関係性の学習方法について説明する. 実験 1 では, ロボットのカメラ画像 1 枚を観測データとして学習を行ったが, 実験 2 では, ロボットのカメラ画像は環境光の影響を受け, 色かぶりが生じていた. この色かぶりを軽減するために, 特定の場所での学習フェーズにおいて, ロボットは同じカメラ姿勢で 6 枚の画像を撮影し, これらの画像から得られた特徴量を平均化した. そして, これらの平均化された特徴量を場所概念モデルの画像特徴量として用いた. 物体検出頻度情報についても同様に, 6 枚の撮像画像に対して物体検出を行い, 各画像から得られた BoO 特徴量を平均化した. 得られた平均化特徴量を場所概念モデルの物体検出頻度情報として使用した.

**6.3 実験条件** 実験では, 提案手法 (SpCoSLAM + ProbLog) を用いた. SpCoSLAM に用いたハイパーパラメータは以下のように設定した:  $\alpha = 18.5$ ,  $\gamma = 0.01$ ,  $\beta = 0.1$ ,  $\chi = 0.1$ ,  $m_0 = [0, 0]^T$ ,  $\kappa = 0.2$ ,  $V_0 = \text{diag}(1, 1)$ ,

§ <https://github.com/ultralytics/yolov5#pretrained-checkpoints>

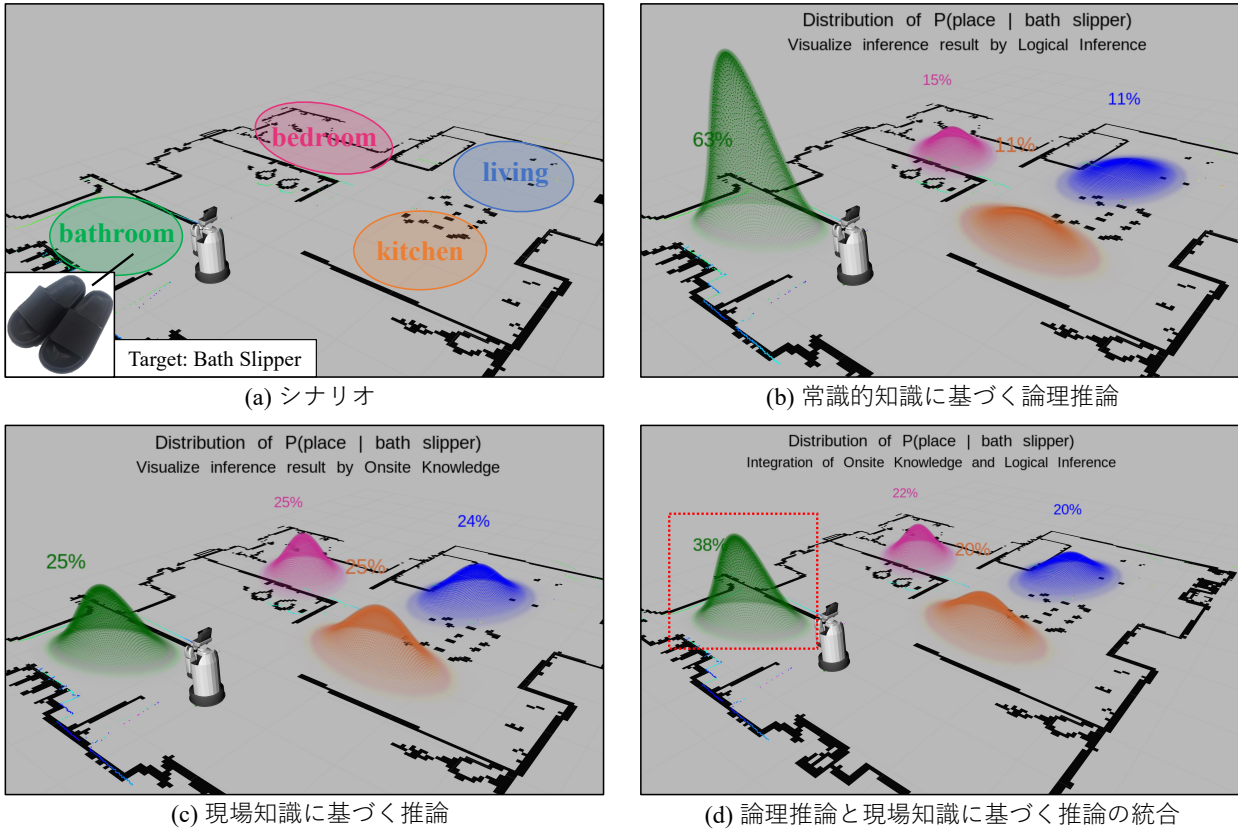


Figure 9: 提案手法 (SpCoSLAM + ProbLog) により推定された bath slipper の配置分布を可視化した様子. (常識的知識に基づく論理推論が有効な事例.) 具体的には, (b) は常識的知識による論理推論, (c) は現場知識に基づく推論, (d) は常識的知識に基づく論理推論と現場知識に基づく推論の統合を示している. 第 5.3 で述べた場所と物体の関係性の学習は 1 回だけ行われた. 緑, オレンジ, 青, ピンクのガウス分布は, それぞれ bathroom, kitchen, living, bedroom の領域を表している. bath slipper は, 実験環境 (Figure. 9(a)) の bathroom の領域内に配置された. Figure. 9(d) の赤い点線枠は, bath slipper が配置された最も可能性の高い場所を示している.

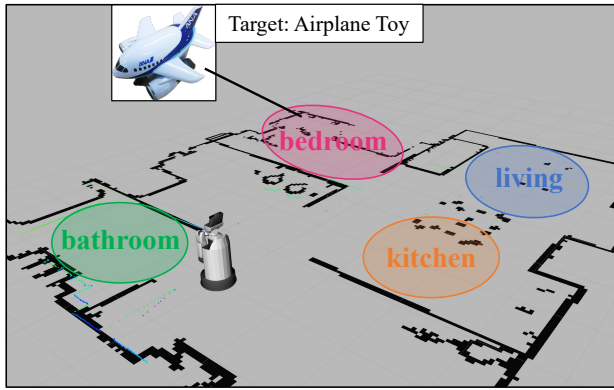
$\nu_0 = 3.0$ ,  $\lambda = 5.0$ . 場所概念モデルは 30 個のパーティクルを使用し, 固定ラグ活性化のラグ値は 10 であった. 重み  $\eta$  は 0.35 とした.

**6.4 定性的評価** 実験 2 では提案手法を使用し, 物体の配置場所を推論するために, 常識的知識に基づく論理的推論と現場知識に基づく推論の有効性を可視化し, 評価するための定性的分析を行った. はじめに, 場所の領域を表すガウス分布を実験環境の占有格子地図上にマッピングした (Figure. 9). また, 式 8 で計算された各場所の領域における物体の存在確率も地図上に表示している. 物体の存在確率に基づく多変量ガウス分布の高さを視覚的に表現するために, 視覚化の際に式 8 で計算された混合比に 3 倍を乗じた. Figure. 9~Figure. 11 に, 現場知識による推論の有効性, 常識的知識による論理的推論の有効性, 常識的知識による論理的推論が有効でない場合のシナリオを示す.

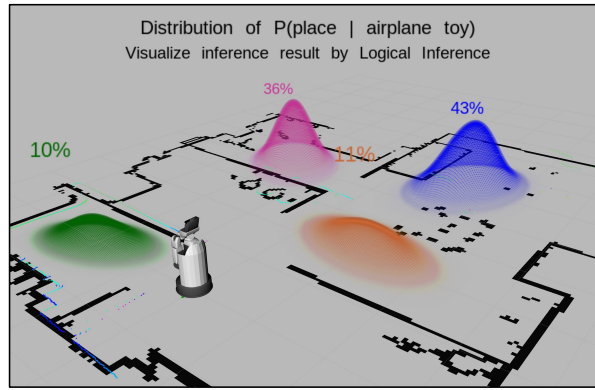
Figure. 9 は, 常識的知識に基づく論理的推論が, 物体の配置場所の推論に良い影響を与える場合を示している. Figure. 9 の (b)~(d) はそれぞれ, 常識的知識に基づく論理推論, 現場知識に基づく推論, 両者の統合による bath slipper の配置場所の推論結果を示す. Figure. 9 において, 緑, オレンジ, 青, ピンク色のガウス分布は, それぞれ bathroom, kitchen, living, bedroom の領域を表

す. 実験環境では, bathroom の領域に bath slipper を置き (Figure. 9 (a)), 5.3 節で述べたような場所と物体の関係の学習を一度行った. ロボットが bath slipper の位置を推論する際, 現場知識に基づく推論では, bathroom, kitchen, bedroom に物体が存在する確率値はほぼ等しかった (Figure. 9 (c)). 場所と物体の関係を学習する初期段階では, ロボットは主に living 内の場所について学習した. バススリッパは living には存在しないため, living 以外の場所の方が物体の存在する確率が相対的に高く, このような現象が発生した. 逆に常識的知識に基づく論理推論では, 物体が bathroom に存在する確率が最も高くなった (Figure. 9 (b)) この結果は, 5.2 節で述べたアンケート調査から得られた物体の配置分布が, 現場の環境とよく一致したために得られたものである. このように, 現場知識に基づく推論と常識的知識に基づく論理推論を統合することで (Figure. 9 (d)), 現場学習が限定的であっても, bath slipper の実際の配置を高い確率値で予測できることを示した. Figure. 9 (d) の赤い点線枠は, 最も可能性の高い配置を示す.

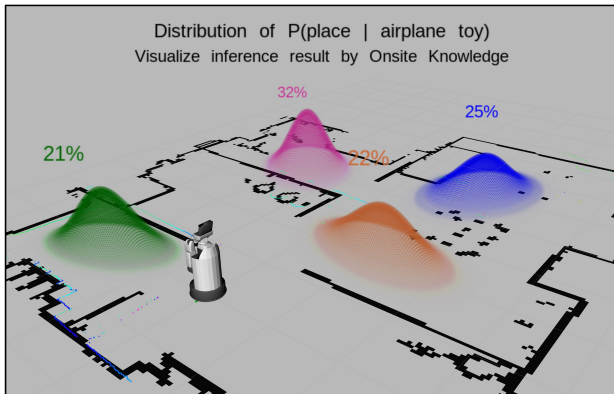
Figure. 10 は, 現場知識に基づく推論が物体の配置場所の推論に好影響を与えた事例である. 実験環境では, bedroom の領域に airplane toy を配置し (Figure. 10 (a)), 5.3 節で述べた場所と物体の関係学習を 11 回行った. ロボットが airplane toy の位置を推論



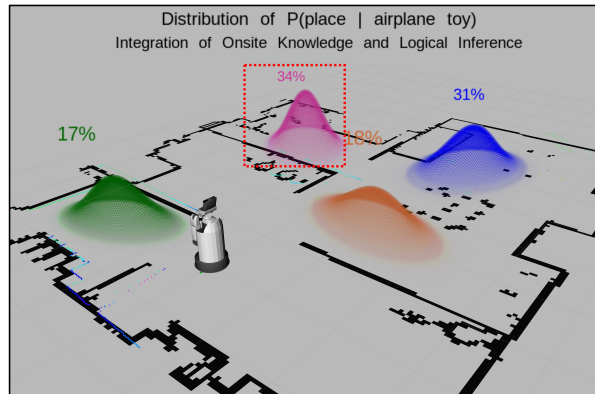
(a) シナリオ



(b) 常識的知識に基づく論理推論



(c) 現場知識に基づく推論



(d) 論理推論と現場知識に基づく推論の統合

Figure 10: 提案手法 (SpCoSLAM + ProbLog) により推定された airplane toy の配置分布を可視化した様子. (現場知識に基づく推論が有効な事例.) 5.3 節で述べた場所と物体の関係性の学習は 11 回行われた. airplane toy は実験環境の bedroom の領域に配置された (Figure. 10(a)). Figure. 10(d) の赤い点線枠は, airplane toy が配置された最も可能性の高い場所を示している.

した際、物体が存在する確率が最も高い値を living に割り当てた。常識的知識に基づく論理推論では、実際に airplane toy が置かれた bedroom が 2 番目に高い確率値を示した (Figure. 10 (b)). しかし現場知識と常識的知識に基づく推論を統合すると、常識的知識に基づく論理推論の結果は抑制され、airplane toy が実際に置かれた場所が最も高い確率で予測された (Figure. 10 (d)). これは、物体検出の頻度を表す確率分布のパラメータ  $\xi$  が、場所と物体の関係性を学習する過程で、より現場環境に適応するように更新され、推論の統合の過程で論理推論の結果が抑制されたためと考えられる。

最後に Figure. 11 は、常識的知識に基づく論理推論が、物体の配置場所の推論に悪影響を及ぼす場合を示す。実験環境では、cracker box を living に配置し (Figure. (a)), 5.3 節で述べた場所と物体の関係性の学習処理を 11 回行った。ロボットがクラッカーボックスの配置場所を推論した際、常識的知識に基づく論理推論により、物体が存在する確率値が最も高いのは kitchen であり、実際に物体が置かれた living は 2 番目に高い確率値であった (Figure. (b)). Figure. 10 とは対照的に Figure. (d) は、現場知識による推論結果を統合しても、常識的知識に基づく論理推論が抑制されず、クラッカーボックスが実際に置かれた場所を、他の場所よりも高い確率値で予測できないことを示す。提案手法では、常識的知識と現場知識による推論結果を組み合わせる際に、加重平均を

採用している。しかしながら、加重平均に用いる重みはロボットの学習回数に依存しないため、常識的知識による推論結果の影響が残っていることがわかる。したがって、物体の配置場所の学習過程に応じて、逐次的に重みを更新する仕組みを構築することが重要である。

実環境での実証実験では、周囲の光の色によってロボットの撮影画像に色かぶりが生じ、物体検出器の誤検出によって物体探索タスクが失敗することがあった。今回の実験では物体検出器として YOLOv5 を採用し、実験環境下での物体検出に特化した微調整を行った。しかしながらこの微調整には、環境光の影響は十分に考慮されていない。近年の研究では、40 億の画像テキストペアを含むデータセットを用いて事前に訓練された CLIP [34] に基づく物体検出器 [48, 49] が提案されている。このような物体検出器を用いることで、この問題を解決し、性能を向上させることができる。

## 7 おわりに

本研究の目的は、現場学習と常識的知識に基づく推論を組み合わせ、新規環境において物体の位置を推論するために、場所と物体の関係性の獲得に関する学習コストの削減の可能性を検討することである。この目的を達成するために、ロボットが仮想的家庭環境内で、配置場所が未確定の物体を含む日常的な物体の探索を行う実験を行っ

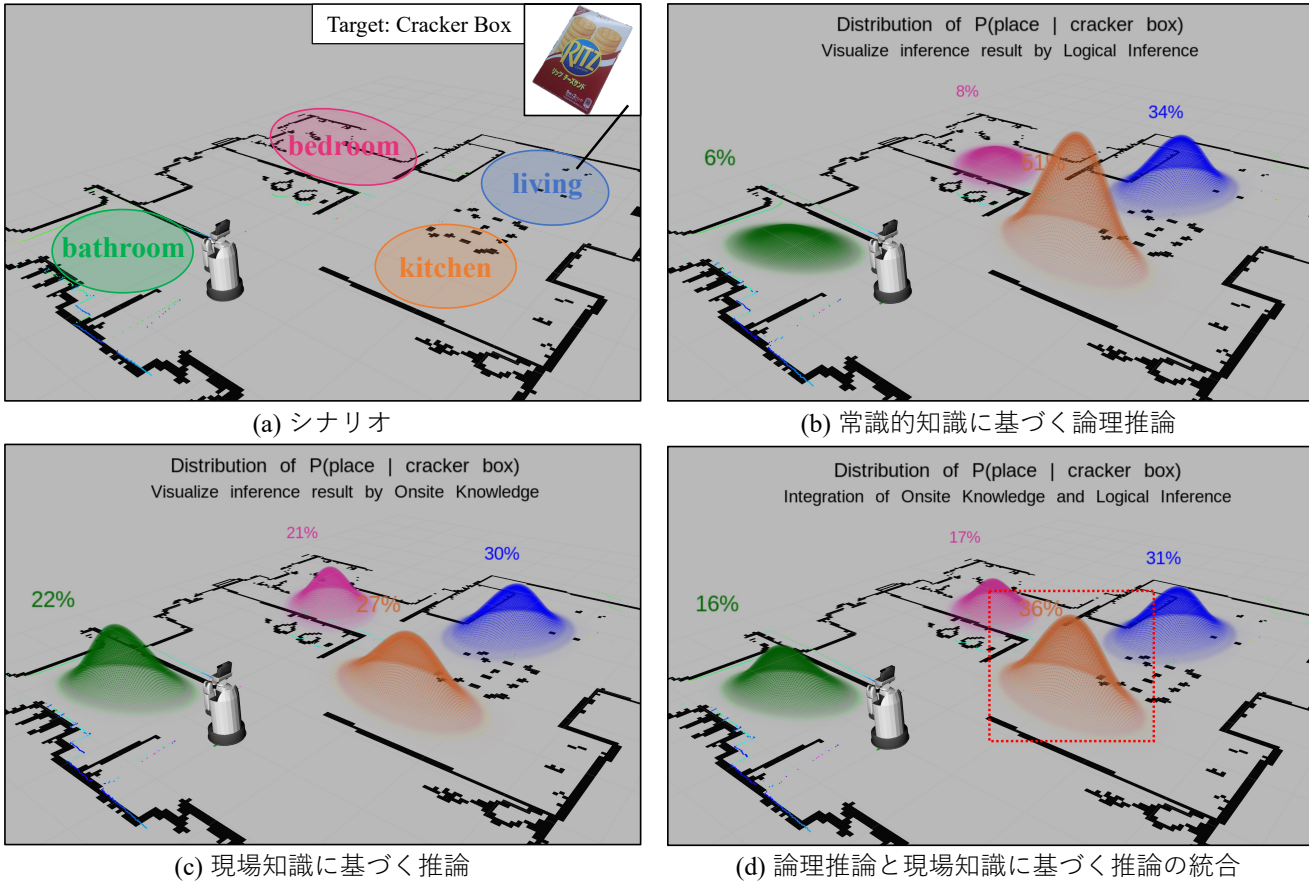


Figure 11: 提案手法 (SpCoSLAM + ProbLog) により推定された cracker box の配置分布を可視化した様子. (常識的知識に基づく論理推論がマイナスに働くケース.) 5.3 節で述べた場所と物体の関係性の学習は 11 回行われた. cracker box は実験環境の living の領域に配置された (Figure. 11 (a)). Figure. 11 (d) の赤い点線枠は, cracker box が配置された最も可能性の高い場所を示している.

た. この実験では 4 つの異なるアプローチが採用された: SpCoSLAM, Prior, SpCoSLAM + Prior, SpCoSLAM + ProbLog (提案手法) である. ロボットが環境内の全ての物体の位置を特定するために, 必要な部屋の訪問回数を比較することで, 提案手法がより効果的であることを示した.

また提案手法の学習コストは, 3 つのベースライン手法と比較して 1.6 倍削減されることを示した. この低減は, 探索時に 1 部屋の訪問で配置場所が未確定の物体を見つけることを含む 24 個の物体探索タスクにおいて観測された. 提案手法は, 学習時における部屋の訪問数の必要性が減少し, 確率的論理推論を持たないベースライン手法と比較して, 提案手法は優れた探索性能を達成した.

さらに, 現場学習に基づく推論と常識的知識に基づく論理推論がタスク成功率に与える影響を調査するために, 定性的分析を行った. 分析の結果, 二つの異なる推論方法がそれぞれ異なる効果を持つことが明らかになった. 一つ目は, 現場学習に基づく推論方法で, これは良い効果をもたらす. 二つ目は, 常識的知識を基にした論理的推論方法で, 良い効果と悪い効果をもたらす可能性が明らかになった.

しかし, 提案したアプローチにはいくつかの限界がある. ロボットが部屋を訪問して物体を探索する場合, 物体を探索するロボットの姿勢はあらかじめ決められてい

る. そのため物体の配置によって, ロボットが物体検出の際に物体を見落とす可能性がある. そこで, 本研究で用いたロボットの位置情報  $x_t$  に加えて, ロボットの姿勢  $\theta$  も観測データとすることを検討する. ロボットが物体  $o_t$  を探索する際の  $P(x_t|o_t)$  を計算し, 最も確率値の高い位置, 姿勢を選択することで, ロボットが部屋を訪れた際に検出するべき方向を決定できる.

さらに近年, 自己教師あり学習によって事前に訓練された大規模言語モデル (PaLM [50], GPT-3 [51], GPT-4 [52]) や基盤モデル [33] をロボティクスに利用することへの注目が高まっている [53–58]. ロボットが推論中にプロンプトを取り込む, 文脈内学習として知られるプロセスを経て, ロボットは行動計画を生成したり, 環境内の物体の存在確率を推定できる. さらに別の関連研究 [58] では, CLIP [34] のような基盤モデルをエンコーダとして用い, エンコーダの特徴量をニューラル表現に用いることで, 表現力の高いセマンティックマップが構築された. 本研究では, 提案アプローチのモデル化を容易にするために, 多変量混合ガウス分布を組み込んだ場所概念獲得モデルを用いた. しかし今後の研究では, 表現力を高めたセマンティックマッピング手法の開発も目指している. また, 大規模言語モデルを用いて言語指示からロボットの行動を生成する研究 [53, 59] が盛んに行われている. これらの研究は, 言語モデルを用いたロボットの

行動生成における可能性を主張している。今後はロボットの行動計画を含めた研究を進めることが求められ、言語指示とロボットの行動計画を統合することで、より高度なタスク実行や人との協調作業が可能となると期待される。

## 謝辞

本研究は JST【Moonshot R&D - MILLENNIA Program】課題番号 JPMJMS2011, JSPS KAKENHI JP20K19900, JP22K12212, JP23K16975, JST 科学技術イノベーション創出に向けた大学フェローシップ創設事業 JPMJFS2146 の一部支援を受けた。

## A 実験で使用了物体のリスト

Table. 4 と 5 に実験 1 と 2 で使用了物体リストを示す。

Table 4: 実験1で使用した物体の3Dモデルと名称のリスト. 定義された物体は, どこに置かれる可能性が高いかという存在確率が常識的知識として確率論理で定義された物体. 未定義の物体は, 存在確率が常識的知識として定義されていない物体. ()内のラベルは, 物体検出器によって検出されたときの物体ラベル.



Category	snack	bottle
Defined Object	Chips Bag(chips_bag) Cracker Box(cracker_box)	Fruits Juice(fruits_juice) Coffee(coffee)
Undefined Object	Pudding Box(pudding_box)	Muscat Juice(muscat)
	tableware	fruits
	Plate(plate)	Banana(banana)
	Pitcher(pitcher_base)	Orange(orange)
	Bowl(bowl)	Apple(apple)
		doll
		Pig Plush Toy(pig_doll)
		Penguin Plush Toy(penguin_doll)
		Sheep Plush Toy(sheep_doll)
	vehicle_toy	toiletries
	Truck Toy(truck_toy)	Toothpaste(tooth_paste)
	Plane Toy(airplane_toy)	Towel(towel)
	Car Toy(car_toy)	Cup(cup)
		bath_goods
		Treatments(treatments)
		Slipper(bath_slipper)
		Bath Sponge(sponge)

Table 5: 実験2で使用した物体と名称のリスト. 定義された物体は, どこに置かれる可能性が高いかという存在確率が常識的知識として確率論理で定義された物体. 未定義の物体は, 存在確率が常識的知識として定義されていない物体. ()内のラベルは, 物体検出器によって検出されたときの物体ラベル.



Category	snack	bottle
Defined Object	Cracker box(cracker_box)	Coffee(coffee)
Undefined Object	Pudding box(pudding_box)	Muscat juice(muscat)
	tableware	fruits
	Pitcher(pitcher_base)	Orange(orange)
	Bowl(bowl)	Apple(apple)
		doll
		Penguin plush toy(penguin_doll)
		Sheep plush toy(sheep_doll)
	vehicle_toy	toiletries
	Truck toy(truck_toy)	Towel(towel)
	Car toy(car_toy)	Cup(cup)
		bath_goods
		Bath slipper(bath_slipper)
		Bath sponge(sponge)

## B 確率論理で構築された常識的知識

提案手法は、述語論理を用いて常識的知識を取り入れ、物体カテゴリ、物体の存在確率、特定のカテゴリに属する物体が各場所に存在する確率を表現する。例えば、食器カテゴリに属する物体の例を以下に示す。

はじめに、述語記号と定数記号を用いて物体名と物体カテゴリの関係を定義する。

```
tableware(plate).  
tableware(bowl).  
tableware(pitcher_base).
```

その後、物体の存在確率の定義を以下のように定義した。

```
0.276 :: exist(plate, living).  
0.699 :: exist(plate, kitchen).  
0.001 :: exist(plate, bedroom).  
0.024 :: exist(plate, bathroom).  
  
0.280 :: exist(pitcher_base, living).  
0.550 :: exist(pitcher_base, kitchen).  
0.150 :: exist(pitcher_base, bedroom).  
0.020 :: exist(pitcher_base, bathroom).  
  
0.250 :: exist(bowl, living).  
0.250 :: exist(bowl, kitchen).  
0.250 :: exist(bowl, bedroom).  
0.250 :: exist(bowl, bathroom).
```

最後に、物体カテゴリを考慮した物体の存在確率の記述を以下に示す。

```
0.280::exist(X, living);  
0.640::exist(X, kitchen);  
0.060::exist(X, bedroom);  
0.020::exist(X, bathroom) :- tableware(X).
```

## C 場所概念モデルの生成過程

本研究で使用した場所概念モデルの生成過程は式 (9)–(22) に示す。ここで、 $\text{Mult}(\cdot)$  は多項分布、 $\text{Dir}(\cdot)$  はディリクレ分布、 $\mathcal{N}(\cdot)$  は多次元ガウス分布、 $\mathcal{IW}(\cdot)$  は逆ウィシャート分布、and  $\text{DP}(\cdot)$  はディリクレ過程である。

$$x_t \sim P(x_t|x_{t-1}, u_t) \quad (9)$$

$$z_t \sim P(z_t|x_t, m) \quad (10)$$

$$i_t \sim P(i_t|x_t, \mu, \Sigma, \phi, C_t) \quad (11)$$

$$\phi_l \sim \text{DP}(\gamma) \quad (12)$$

$$\Sigma_k \sim \mathcal{IW}(\Sigma|V_0, \nu_0) \quad (13)$$

$$\mu_k \sim \mathcal{N}(\mu|m_0, \Sigma_k/\kappa_0) \quad (14)$$

$$C_t \sim \text{Mult}(\pi) \quad (15)$$

$$\pi \sim \text{DP}(\alpha) \quad (16)$$

$$w_t \sim \text{Mult}(W_l) \quad (17)$$

$$o_t \sim \text{Mult}(\xi_l) \quad (18)$$

$$f_t \sim \text{Mult}(\theta_l) \quad (19)$$

$$W_l \sim \text{Dir}(\beta) \quad (20)$$

$$\theta_l \sim \text{Dir}(\chi) \quad (21)$$

$$\xi_l \sim \text{Dir}(\lambda) \quad (22)$$

## D アンケート

確率論理で使用する物体の配置に関する常識的知識を得るために、家庭環境内の物体の配置に関するアンケート調査を実施した。アンケートは 20 代から 60 代までの 63 名（男性 52 名、女性 11 名）を対象とし、実験に使用する各物体の部屋の配置について回答してもらった。本研究で使用したアンケートを以下に示す。

アンケートから得られた回答は多項分布に従うと仮定した。この情報を取り入れるため、ディリクレ平滑化を適用し、各物体の配置場所の確率を推定した。実験 1 ではディリクレ事前分布のハイパーパラメータを 0.1、実験 2 では 5.0 とした。

# 家庭環境における物体の配置に関するアンケート

この度はアンケートにご協力して頂き、誠にありがとうございます。

## 1. 研究目的

本研究では、家庭用ロボットが物体の配置に関する事前知識を用いることで、新規家庭環境における物体の存在する場所の知識をロボットが少ない学習回数で獲得可能な手法の開発を目的としています。

## 2. アンケート方法

「家庭環境における物体の配置」に関して3分程度のアンケートを行います。

## 3. 研究への参加と撤回について

アンケート回答後に辞退されることになっても、何も不利益を被ることはありません。その際は、それまでに収集したデータを分析対象としてよいのか、廃棄を希望されるのかをお聞かせいただければ、それに従ってデータを取り扱います。

## 4. 守秘や個人情報、研究データの取り扱いについて

調査にて得られた回答データは統計的に処理され、個人が特定されることのないよう厳重に管理し、研究以外で使用することはありません。研究が終了次第速やかに破棄させていただきますので、安心して回答ください。

その他、本アンケートに関する意見や質問等がございましたら、下記までお問い合わせください。

### <実施者>

立命館大学情報理工学研究科 博士課程前期課程 人間情報科学コース 2回生  
創発システム研究室  
長谷川 翔一  
[hasegawa.shoichi@em.ci.ritsumei.ac.jp](mailto:hasegawa.shoichi@em.ci.ritsumei.ac.jp)

### <指導教員>

立命館大学情報理工学部教授  
谷口 忠大

---

\*必須



- 1。 上記をご理解いただいた上で、研究参加の意思がある方は「同意する」にチェックをお願いします。「同意しない」をチェックする場合は、これで終了になります。\*

1つだけマークしてください。

- 同意する  
 同意しない

### 家庭環境における物体の配置に関するアンケート

- 2。 1. 氏名をご記入ください。(Please enter your full name.)\*

---

- 3。 2. 性別をお答えください。(Please tell me your gender.)\*

1つだけマークしてください。

- 男性 (Man)  
 女性 (Woman)

- 4。 3. 当てはまる年代を選択してください。(Please select your age.)\*

1つだけマークしてください。

- 0～20代 (0～20s)  
 30代～40代 (30s～40s)  
 50代～60代 (50s～60s)  
 70代～ (70s～)

- 5。 4. リビングに置いていそうな物体を選択してください。(複数選択可) (Select the objects that you are likely to put in your living. (Multiple selections possible))

\*

当てはまるものをすべて選択してください。

- 皿 (Plate)
- お椀 (Bowl)
- 水差し (Pitcher)
- バナナ (Banana)
- リンゴ (Apple)
- オレンジ (Orange)
- ビスケットクラッカー (Cracker Box)
- チョコレート (Chocolate)
- ポテトチップス (Chips Bag)
- コーヒー (Coffee)
- フルーツジュース (Fruits Juice)
- グレープジュース (Grape Juice)
- ブタのぬいぐるみ (Pig Doll)
- ヒツジのぬいぐるみ (Sheep Doll)
- ペンギンのぬいぐるみ (Penguin Doll)
- 飛行機のおもちゃ (Airplane Toy)
- トラックのおもちゃ (Truck Toy)
- 車のおもちゃ (Car Toy)
- 歯磨き粉 (Tooth Paste)
- タオル (Towel)
- コップ (Cup)
- トリートメント (Treatment)
- ボディスポンジ (Body Sponge)
- バススリッパ (Bath Slippers)

- 6。 5. キッチンに置いていそうな物体を選択してください。(複数選択可) (Select the objects that you are likely to put in your kitchen. (Multiple selections possible)) \*

当てはまるものをすべて選択してください。

- 皿 (Plate)
- お椀 (Bowl)
- 水差し (Pitcher)
- バナナ (Banana)
- リンゴ (Apple)
- オレンジ (Orange)
- ビスケットクラッカー (Cracker Box)
- チョコレート (Chocolate)
- ポテトチップス (Chips Bag)
- コーヒー (Coffee)
- フルーツジュース (Fruits Juice)
- グレープジュース (Grape Juice)
- ブタのぬいぐるみ (Pig Doll)
- ヒツジのぬいぐるみ (Sheep Doll)
- ペンギンのぬいぐるみ (Penguin Doll)
- 飛行機のおもちゃ (Airplane Toy)
- トラックのおもちゃ (Truck Toy)
- 車のおもちゃ (Car Toy)
- 歯磨き粉 (Tooth Paste)
- タオル (Towel)
- コップ (Cup)
- トリートメント (Treatment)
- ボディスポンジ (Body Sponge)
- バススリッパ (Bath Slippers)

- 7。 6. ベッドルームに置いていそうな物体を選択してください。(複数選択可)  
(Select the objects that you are likely to put in your bedroom. (Multiple selections possible)) \*

当てはまるものをすべて選択してください。

- 皿 (Plate)
- お椀 (Bowl)
- 水差し (Pitcher)
- バナナ (Banana)
- リンゴ (Apple)
- オレンジ (Orange)
- ビスケットクラッカー (Cracker Box)
- チョコレート (Chocolate)
- ポテトチップス (Chips Bag)
- コーヒー (Coffee)
- フルーツジュース (Fruits Juice)
- グレープジュース (Grape Juice)
- ブタのぬいぐるみ (Pig Doll)
- ヒツジのぬいぐるみ (Sheep Doll)
- ペンギンのぬいぐるみ (Penguin Doll)
- 飛行機のおもちゃ (Airplane Toy)
- トラックのおもちゃ (Truck Toy)
- 車のおもちゃ (Car Toy)
- 歯磨き粉 (Tooth Paste)
- タオル (Towel)
- コップ (Cup)
- トリートメント (Treatment)
- ボディスポンジ (Body Sponge)
- バススリッパ (Bath Slippers)

8. 7. バスルームに置いていそうな物体を選択してください。(複数選択可)  
(Select the objects that you are likely to put in your bathroom. (Multiple selections possible)) \*

当てはまるものをすべて選択してください。

- 皿 (Plate)
- お椀 (Bowl)
- 水差し (Pitcher)
- バナナ (Banana)
- リンゴ (Apple)
- オレンジ (Orange)
- ビスケットクラッカー (Cracker Box)
- チョコレート (Chocolate)
- ポテトチップス (Chips Bag)
- コーヒー (Coffee)
- フルーツジュース (Fruits Juice)
- グレープジュース (Grape Juice)
- ブタのぬいぐるみ (Pig Doll)
- ヒツジのぬいぐるみ (Sheep Doll)
- ペンギンのぬいぐるみ (Penguin Doll)
- 飛行機のおもちゃ (Airplane Toy)
- トラックのおもちゃ (Truck Toy)
- 車のおもちゃ (Car Toy)
- 歯磨き粉 (Tooth Paste)
- タオル (Towel)
- コップ (Cup)
- トリートメント (Treatment)
- ボディスポンジ (Body Sponge)
- バススリッパ (Bath Slippers)

---

このコンテンツは Google が作成または承認したものではありません。

Google フォーム

## 参考文献

- [1] Ioannis Kostavelis and Antonios Gasteratos. Semantic mapping for mobile robotics tasks: A survey. *Robotics and Autonomous Systems*, Vol. 66, pp. 86–103, 2015.
- [2] Yuki Furuta, Kei Okada, Yohei Kakiuchi, and Masayuki Inaba. An Everyday Robotic System that Maintains Local Rules using Semantic Map based on Long-Term Episodic Memory. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7641–7647, 2018.
- [3] David Fernandez-Chaves, Jose-Raul Ruiz-Sarmiento, Nicolai Petkov, and Javier Gonzalez-Jimenez. From Object Detection to Room Categorization in Robotics. In *International Conference on Applications of Intelligent Systems (APPIS)*, pp. 1–6, 2020.
- [4] Shota Isobe, Akira Taniguchi, Yoshinobu Hagiwara, and Tadahiro Taniguchi. Learning Relationships between Objects and Places by Multimodal Spatial Concept with Bag of Objects. In *International Conference on Social Robotics (ICSR)*, pp. 115–125, 2017.
- [5] Akira Taniguchi, Shota Isobe, Lotfi El Hafi, Yoshinobu Hagiwara, and Tadahiro Taniguchi. Autonomous Planning Based on Spatial Concepts to Tidy Up Home Environments with Service Robots. *Advanced Robotics*, Vol. 35, No. 8, pp. 471–489, April 2021.
- [6] Alejandra C Hernandez, Erik Derner, Clara Gomez, Ramon Barber, and Robert Babuška. Efficient Object Search Through Probability-Based Viewpoint Selection. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6172–6179, 2020.
- [7] Alejandra C Hernandez, Maximilian Durner, Clara Gomez, Iris Grix, Oskars Teikmanis, Zoltan-Csaba Marton, and Ramon Barber. Searching for Objects in Human Living Environments based on Relevant Inferred and Mined Priors. In *European Conference on Mobile Robots (ECMR)*, pp. 1–7, 2021.
- [8] Sourav Garg, Niko Sünderhauf, Feras Dayoub, Douglas Morrison, Akansel Cosgun, Gustavo Carneiro, Qi Wu, Tat-Jun Chin, Ian Reid, Stephen Gould, Peter Corke, and Michael Milford. Semantics for Robotic Mapping, Perception and Interaction: A Survey. *Foundations and Trends in Robotics*, Vol. 8, No. 1–2, pp. 1–224, 2020.
- [9] Clara Gomez, Alejandra C Hernandez, Erik Derner, Ramon Barber, and Robert Babuška. Object-Based Pose Graph for Dynamic Indoor Environments. *IEEE Robotics and Automation Letters (RA-L)*, Vol. 5, No. 4, pp. 5401–5408, 2020.
- [10] Akira Taniguchi, Yoshinobu Hagiwara, Tadahiro Taniguchi, and Tetsunari Inamura. Online Spatial Concept and Lexical Acquisition with Simultaneous Localization and Mapping. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 811–818, 2017.
- [11] Akira Taniguchi, Yoshinobu Hagiwara, Tadahiro Taniguchi, and Tetsunari Inamura. Improved and Scalable Online Learning of Spatial Concepts and Language Models with Mapping. *Autonomous Robots*, Vol. 44, No. 6, pp. 927–946, 2020.
- [12] Lotfi El Hafi, Shota Isobe, Yoshiki Tabuchi, Yuki Katsumata, Hitoshi Nakamura, Takaaki Fukui, Tadashi Matsuo, Gustavo Alfonso Garcia Ricardez, Masaki Yamamoto, Akira Taniguchi, Yoshinobu Hagiwara, and Tadahiro Taniguchi. System for Augmented Human-Robot Interaction through Mixed Reality and Robot Training by Non-Experts in Customer Service Environments. *Advanced Robotics*, Vol. 34, No. 3-4, pp. 157–172, 2020.
- [13] Ying Zhang, Guohui Tian, Jiaying Lu, Mengyang Zhang, and Senyan Zhang. Efficient Dynamic Object Search in Home Environment by Mobile Robot: A Priori Knowledge-based Approach. *IEEE Transactions on Vehicular Technology*, Vol. 68, No. 10, pp. 9466–9477, 2019.
- [14] Shaopeng Liu, Guohui Tian, Ying Zhang, Mengyang Zhang, and Shuo Liu. Service planning oriented efficient object search: A knowledge-based framework for home service robot. *Expert Systems With Applications*, Vol. 187, p. 115853, 2022.
- [15] Puwanan Chumtong, Yasushi Mae, Kenichi Ohara, Tomohito Takubo, and Tatsuo Arai. Object search using object co-occurrence relations derived from web content mining. *Intelligent Service Robotics*, Vol. 7, pp. 1–13, 2014.
- [16] Alejandra C Hernández, Clara Gómez, Ramón Barber, and Oscar Martínez Mozos. Object-Based Probabilistic Place Recognition for Indoor Human Environments. In *International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO)*, pp. 177–182, 2018.
- [17] Tiago S Veiga, Pedro Miraldo, Rodrigo Ventura, and Pedro U Lima. Efficient Object Search for Mobile Robots in Dynamic Environments: Semantic Map as an Input for the Decision Maker. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2745–2750, 2016.
- [18] Tiago S. Veiga, Miguel Silva, Rodrigo Ventura, and Pedro U. Lima. A Hierarchical Approach to Active Semantic Mapping using Probabilistic Logic and Information Reward POMDPs. In *International Conference on Automated Planning and Scheduling (ICAPS)*, Vol. 29, pp. 773–781, 2019.
- [19] Nils J Nilsson. Probabilistic Logic. *Artificial Intelligence*, Vol. 28, No. 1, pp. 71–87, 1986.
- [20] Raymond Ng and Venkatramanan Siva Subrahmanian. Probabilistic Logic Programming. *Information and Computation*, Vol. 101, No. 2, pp. 150–201, 1992.
- [21] Peter Anderson, Qi Wu, Damien Teney, Jake Bruce, Mark Johnson, Niko Sünderhauf, Ian

- Reid, Stephen Gould, and Anton Van Den Hengel. Vision-and-Language Navigation: Interpreting Visually-Grounded Navigation Instructions in Real Environments. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3674–3683, 2018.
- [22] Muhammad Zubair Irshad, Chih-Yao Ma, and Zsolt Kira. Hierarchical Cross-Modal Agent for Robotics Vision-and-Language Navigation. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 13238–13246, 2021.
- [23] Peter Anderson, Angel Chang, Devendra Singh Chaplot, Alexey Dosovitskiy, Saurabh Gupta, Vladlen Koltun, Jana Kosecka, Jitendra Malik, Roozbeh Mottaghi, Manolis Savva, et al. On Evaluation of Embodied Navigation Agents. *arXiv preprint arXiv:1807.06757*, 2018.
- [24] Dhruv Batra, Aaron Gokaslan, Aniruddha Kembhavi, Oleksandr Maksymets, Roozbeh Mottaghi, Manolis Savva, Alexander Toshev, and Erik Wijmans. ObjectNav Revisited: On Evaluation of Embodied Agents Navigating to Objects. *arXiv preprint arXiv:2006.13171*, 2020.
- [25] Shoichi Hasegawa, Akira Taniguchi, Yoshinobu Hagiwara, Lotfi El Hafi, and Tadahiro Taniguchi. Inferring Place-Object Relationships by Integrating Probabilistic Logic and Multimodal Spatial Concepts. In *IEEE/SICE International Symposium on System Integration (SII)*, pp. 1–8, 2023.
- [26] Tadahiro Taniguchi, Takayuki Nagai, Tomoaki Nakamura, Naoto Iwahashi, Tetsuya Ogata, and Hideki Asoh. Symbol Emergence in Robotics: A Survey. *Advanced Robotics*, Vol. 30, No. 11-12, pp. 706–728, 2016.
- [27] Tadahiro Taniguchi, Emre Ugur, Matej Hoffmann, Lorenzo Jamone, Takayuki Nagai, Benjamin Roman, Toshihiko Matsuka, Naoto Iwahashi, Erhan Oztog, Justus Piater, et al. Symbol Emergence in Cognitive Developmental Systems: A Survey. *IEEE Transactions on Cognitive and Developmental Systems (TCDS)*, Vol. 11, No. 4, pp. 494–516, 2018.
- [28] Rui Fukushima, Kei Ota, Asako Kanezaki, Yoko Sasaki, and Yusuke Yoshiyasu. Object Memory Transformer for Object Goal Navigation. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11288–11294, 2022.
- [29] Arjun Majumdar, Gunjan Aggarwal, Bhavika Devnani, Judy Hoffman, and Dhruv Batra. ZSON: Zero-Shot Object-Goal Navigation using Multimodal Goal Embeddings. *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 35, pp. 32340–32352, 2022.
- [30] Devendra Singh Chaplot, Dhiraj Prakashchand Gandhi, Abhinav Gupta, and Russ R Salakhutdinov. Object Goal Navigation using Goal-Oriented Semantic Exploration. *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 33, pp. 4247–4258, 2020.
- [31] Anwesan Pal, Yiding Qiu, and Henrik Christensen. Learning hierarchical relationships for object-goal navigation. In *Conference on Robot Learning (CoRL)*, pp. 517–528, 2021.
- [32] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is All you Need. *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 30, , 2017.
- [33] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the Opportunities and Risks of Foundation Models. *arXiv preprint arXiv:2108.07258*, 2021.
- [34] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning Transferable Visual Models From Natural Language Supervision. In *International Conference on Machine Learning (ICML)*, pp. 8748–8763, 2021.
- [35] Luc De Raedt, Angelika Kimmig, and Hannu Toivonen. ProbLog: A Probabilistic Prolog and Its Application in Link Discovery. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2468–2473, 2007.
- [36] Taisuke Sato and Yoshitaka Kameya. PRISM: A Language for Symbolic-Statistical Modeling. In *International Joint Conference on Artificial Intelligence (IJCAI)*, Vol. 97, pp. 1330–1339, 1997.
- [37] Taisuke Sato. A Statistical Learning Method for Logic Programs with Distribution Semantics. In *International Conference on Logic Programming (ICLP)*, pp. 715–729, 1995.
- [38] Giorgio Grisetti, Cyrill Stachniss, and Wolfram Burgard. Improving Grid-based SLAM with Rao-Blackwellized Particle Filters by Adaptive Proposals and Selective Resampling. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2432–2437, 2005.
- [39] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 Million Image Database for Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 40, No. 6, pp. 1452–1464, 2018.
- [40] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, Andrew Y Ng, et al. ROS: an open-source Robot Operating System. In *IEEE International Conference on Robotics and Automation (ICRA) Workshop on Open Source Software*, Vol. 3, p. 5, 2009.
- [41] Lotfi El Hafi, Gustavo Alfonso Garcia Ricardez, Felix von Drigalski, Yuki Inoue, Masaki Yamamoto, and Takashi Yamamoto. Software Development Environment for Collaborative Research

- Workflow in Robotic System Integration. *Advanced Robotics*, Vol. 36, No. 11, pp. 533–547, 2022.
- [42] Nathan Koenig and Andrew Howard. Design and Use Paradigms for Gazebo, An Open-Source Multi-Robot Simulator. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vol. 3, pp. 2149–2154, 2004.
- [43] Takashi Yamamoto, Koji Terada, Akiyoshi Ochiai, Fuminori Saito, Yoshiaki Asahara, and Kazuto Murase. Development of Human Support Robot as the Research Platform of a Domestic Mobile Manipulator. *ROBOMECH journal*, Vol. 6, No. 1, pp. 1–15, 2019.
- [44] Berk Calli, Aaron Walsman, Arjun Singh, Sidhartha Srinivasa, Pieter Abbeel, and Aaron M. Dollar. Benchmarking in Manipulation Research: Using the Yale-CMU-Berkeley Object and Model Set. *IEEE Robotics & Automation Magazine*, Vol. 22, No. 3, pp. 36–52, 2015.
- [45] Laura Downs, Anthony Francis, Nate Koenig, Brandon Kinman, Ryan Hickman, Krista Reymann, Thomas B McHugh, and Vincent Vanhoucke. Google Scanned Objects: A High-Quality Dataset of 3D Scanned Household Items. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2553–2560, 2022.
- [46] Stefan Kohlbrecher, Oskar Von Stryk, Johannes Meyer, and Uwe Klingauf. A Flexible and Scalable SLAM System with Full 3D Motion Estimation. In *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pp. 155–160, 2011.
- [47] Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, NanoCode012, Yonghye Kwon, Kalen Michael, TaoXie, Jiacong Fang, imyhxy, Lorna, Zeng Yifu, Colin Wong, Abhiram V, Diego Montes, Zhiqiang Wang, Cristi Fati, Jebastin Nadar, Laughing, UnglvKitDe, Victor Sonck, tkianai, yxNONG, Piotr Skalski, Adam Hogan, Dhruv Nair, Max Strobel, and Mrinal Jain. ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation, 2022.
- [48] Xiuye Gu, Tsung-Yi Lin, Weicheng Kuo, and Yin Cui. Open-vocabulary Object Detection via Vision and Language Knowledge Distillation. In *International Conference on Learning Representations (ICLR)*, 2022.
- [49] Xingyi Zhou, Rohit Girdhar, Armand Joulin, Philipp Krähenbühl, and Ishan Misra. Detecting Twenty-Thousand Classes Using Image-Level Supervision. In *European Conference on Computer Vision (ECCV)*, pp. 350–368, 2022.
- [50] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. PaLM: Scaling Language Modeling with Pathways. *arXiv preprint arXiv:2204.02311*, 2022.
- [51] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 33, pp. 1877–1901, 2020.
- [52] OpenAI. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774*, 2023.
- [53] Sai Vemprala, Rogerio Bonatti, Arthur Buckler, and Ashish Kapoor. Chatgpt for robotics: Design principles and model abilities. *Microsoft Auton. Syst. Robot. Res*, Vol. 2, p. 20, 2023.
- [54] Boyuan Chen, Fei Xia, Brian Ichter, Kanishka Rao, Keerthana Gopalakrishnan, Michael S Ryoo, Austin Stone, and Daniel Kappler. Open-Vocabulary Queryable Scene Representations for Real World Planning. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11509–11522, 2023.
- [55] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, et al. Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. *arXiv preprint arXiv:2204.01691*, 2022.
- [56] Chenguang Huang, Oier Mees, Andy Zeng, and Wolfram Burgard. Visual Language Maps for Robot Navigation. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 10608–10615, 2023.
- [57] Jimmy Wu, Rika Antonova, Adam Kan, Marion Lepert, Andy Zeng, Shuran Song, Jeannette Bohg, Szymon Rusinkiewicz, and Thomas Funkhouser. TidyBot: Personalized Robot Assistance with Large Language Models. *Autonomous Robots*, 2023.
- [58] Nur Muhammad Mahi Shafiullah, Chris Paxton, Lerrel Pinto, Soumith Chintala, and Arthur Szlam. Clip-Fields: Weakly Supervised Semantic Fields for Robotic Memory. In *Robotics: Science and Systems (RSS)*, 2023.
- [59] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, Daniel Ho, Jasmine Hsu, Julian Ibarz, Brian Ichter, Alex Irpan, Eric Jang, Rosario Jauregui Ruano, Kyle Jeffrey, Sally Jesmonth, Nikhil Joshi, Ryan Julian, Dmitry Kalashnikov, Yuheng Kuang, Kuang-Huei Lee, Sergey Levine, Yao Lu, Linda Luu, Carolina Parada, Peter Pastor, Jornell Quiambao, Kanishka Rao, Jarek Rettinghouse, Diego Reyes, Pierre Sermanet, Nicolas Sievers, Clayton Tan, Alexander Toshev, Vincent Vanhoucke, Fei Xia, Ted Xiao, Peng Xu, Sichun Xu, Mengyuan Yan, and Andy Zeng. Do as i can and not as i say: Grounding language in robotic affordances. In *arXiv preprint*



*arXiv:2204.01691*, 2022.