DISCERN for Generalizable Robotic Contexts

Swagnik Roychoudhury
Department of Computer Science
New York University
New York, New York
sr6474@nyu.edu

Aparna S. Varde
School of Computing
Montclair State University, NJ
vardea@montclair.edu
ORCID ID: 0000-0002-3170-2510

***Abstract*—This work demonstrates DISCERN (Detection Image System with Commonsense Efficient Ranking Network), a novel generalizable task-ranking approach to improve human-robot collaboration via "discern"-ing with commonsense knowledge (CSK) derived from huge data repositories, augmented with image models and other everyday premises. It is an explainable, efficient solution useful to dynamic multipurpose robots.**

***Index Terms*—AI & Robotics, CSK, Commonsense Reasoning, Human-Robot Collaboration, Task Planning, Sustainable AI, XAI**

I. INTRODUCTION

The DISCERN algorithm (Detection Image System with Commonsense Efficient Ranking Network) was designed to improve task prioritization in human-robot collaboration by incorporating commonsense reasoning. Unlike traditional approaches that rely on pre-training or imitation learning, DISCERN operates "out of the box" by using pre-trained image models, commonsense knowledge bases (CSKBs), and human-derived priorities to determine optimal task order.

Initially applied in a household dining task [1], the originally proposed DISCERN system proved effective at prioritizing based on object attributes like size, distance, and potential danger—without the training requirements of ML-based models. We proffer further enhancements here with respect to the definition of more CSK premises, and the execution in a different yet related context.

From a big data perspective, it is to be noted that by minimizing the need for resource-intensive training and retuning, DISCERN aligns with the principles of sustainable AI, addressing both ecological and computational efficiency [2]. Moreover, it also contributes to the growing field of Explainable AI (XAI) [3], as DISCERN's rule-based prioritization provides easy interpretability, ensuring more trust in its deployment within critical human-robot collaborative tasks.

In this paper, we demonstrate the DISCERN system with its execution in an everyday context. We also explain its extension which allows flexible categorization and rule-based priority definitions, enabling it to support multiple semantic classes, adaptable attributes, and context-specific rules, making it a robust solution across various collaborative settings.

II. RELATED WORKS

Imitation tasks, such as IRIS [4], while performing well in their domain, fail to generalize in new and unseen contexts, requiring new data and training to function.

Guérin et al. [5] propose an algorithm to sort objects by visual similarity via CNN & clustering. Being an unsupervised algorithm, it does a great job of generalizing. However, it is limited only to visual similarity and doesn't account for the semantics of the objects and other details not discernable purely by visuals. DISCERN's semantic understanding of context not only covers more visual features (size, weight, shape) but also includes other useful aspects such as temperature and fragility.

Commonsense is important, not just in humans but in robots as well. Whether through static graphs or non-deterministic models, commonsense knowledge is crucial for reasoning and decision-making in AI systems [6]. Recent work towards semantic understanding in robotic environments has been focusing on Large Language Models (LLMs). For example, Ocker et al. [7] explore the use of LLMs to extract commonsense knowledge for robots, proposing their integration with traditional knowledge bases. Their experiments demonstrate that while LLMs are effective at large-scale knowledge extraction for creating ontologies, they lack the reliability needed for consistent performance. This suggests that a hybrid approach, combining LLMs with symbolic reasoning systems, is necessary for robust performance.

A recent review [8] highlights the importance of commonsense knowledge (CSK) in cognitive robotics for handling dynamic tasks, emphasizing object-focused knowledge like location and affordances in household settings. Certain machine learning applications do better when commonsense knowledge is applied to them [9]. For example, Hidalgo et al. [10] demonstrate the use of commonsense knowledge in robotics by integrating a knowledge base for task-relevant object classification, outperforming deep learning-only approaches in adaptability, consistency, and explainability in domestic settings. CSK-Detector [11] uses commonsense knowledge to enable domestic robots to understand environments for specific tasks, achieving efficient and explainable object detection without extensive image annotation.

Additionally, commonsense reasoning has been applied in human-robot collaboration (HRC) to enhance task execution and safety in manufacturing contexts. For example, Töberg et al. [12] present a system that uses CSK-based reasoning to prioritize tasks, improving efficiency and scalability in both simulations and real-world experiments.

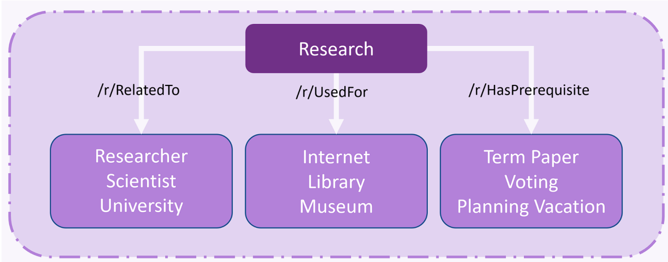

Fig. 1. Concept Net Relations

III. Approach and Implementation

We design an approach *"DISCERN" (Detection Image System with Commonsense Efficient Ranking Network) [1]* that has an architecture with flexible categorization and rule-based priorities. It uses an object detection model only for basic identification of individual objects from scenes and thereafter deploys Commonsense Knowledge (CSK) extracted from large-scale knowledge bases such as ConceptNet [13]. The ConceptNet KB is a large-scale, multilingual knowledge graph designed to provide commonsense knowledge for AI systems, enabling reasoning about everyday concepts and relationships. Nodes in the graph are connected by unidirectional relations, such as in Fig. 1 It reasons using CSK along with other commonsense premises, as discussed next.

Afterward, the classification is used to apply physical and semantic attributes to each object. Then, the algorithm sorts the objects based on these attributes depending on user-specified priorities. For example, the objects can be sorted based on their size, weight, temperature, fragility, or whether the object should be thrown out or recycled. Fig. 2 shows a bird's eye view of the whole DISCERN process.

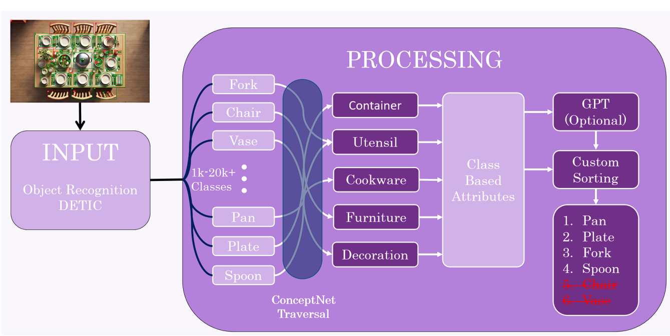

Fig. 2. A Panorama of the DISCERN Process

But how well, and more importantly, how easily, does DISCERN generalize to new contexts? To answer this question, we introduce configurable components — classes, priorities, rules — allowing it to operate in multiple CSK-based contexts. We propose 2 types of rules for class discernment: *Ordering* & *Absolute*. Ordering rules manage class conflicts, where an object can belong to multiple classes, e.g. a wine glass may be as a *decoration* or *container*; ordering rules prioritize *container* over *decoration*. Absolute rules offer associations to override flexible assignments, linking certain keywords to specific classes. They expedite class assignments, reducing ambiguity, e.g. *glass* can be mapped directly to *container*, and prioritized. Absolute rules give precise class boundary control, define the reach of classes, and manage overlapping categories. Ordering & absolute rules together can enhance task prioritization via CSK-based "discern"ment.

Fig. 3. Main Objects in Scene

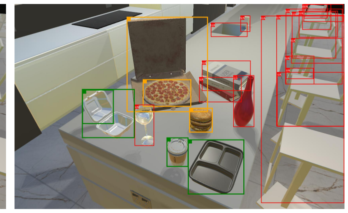

Fig. 4. DISCERN Process Visualized

DISCERN is a boon to robotics & its enormous data. Else, excessive pre-training on millions of full scenes is inefficient & non-explainable. DISCERN is efficient & explainable via its use of image models for basic object identification in scenes (only 100s vs. millions), reusable knowledge extracted one-time from CSK sources, and clearly "discern"able premises.

IV. Experimental Setup

To test this generalized variant of DISCERN, we employed it in another domestic context - recycling household items. DISCERN, being both a classification and ordering algorithm, can do either or both efficiently. In this test, we assort a range of common items that are usually trashed or recycled, as well as objects that can be found with these items but are not to be thrown out. Fig. 3. The list of items and their ground truth labels can be found in Table I.

TABLE I
Attributes of Items in Recycling Task

Recycle	Garbage	Permanent
Plastic Tray	Burgers	Vase
Takeout Box	Pizza	Books
Foam Cup	Pizza Box	Wineglass

We then ask the agents to sort these objects according to their commonsense knowledge. We compare the generalized DISCERN against a random baseline, in which the robot has no prior intuition as to where each object should go.

The human agent in the simulation has a constant pick-up time of three seconds for each object. Since there are eight total removable objects, it takes 24 seconds to trash/recycle all removable objects. We assume that the human, according to their commonsense and intuition, knows the ground truth result of each object (recycle, garbage, permanent) and thus makes not mistakes.

The robot agent also has a constant pick-up time of three seconds. However, the robot can fail to handle the object in one of two ways. Either it attempts to throw out or recycle a permanent object (i.e. a book), or misclassify a garbage item as a recyclable (and vice versa). In either case, the robot goes through a one-second delay, but the object is not removed. We equal this to the case that the human intervenes and prevents the robot from going through with the action.

We set up four agent scenarios. The first experiment is the human baseline to provide an idea the "current" time, which we want to enhance by introducing human-robot collaboration. In the second experiment, the robot agent sorts randomly, while in the third it uses DISCERN. The fourth experiment introduces the human-robot collaboration with random robot sorting. Finally, the fifth experiment simulates collaborative sorting with the robot agent using the DISCERN algorithm.

V. Results

The results of these experiments are summarized in Table II. As seen here, DISCERN outperforms other alternatives in general. In a solo scenario, the robot agent performs better than the random baseline due to less misclassification. The better performance is translated to the collaborative environment, where the DISCERN-enabled robot agent finished the collaborative task about 21% faster. Less misclassification results in the burden of the sorting shifting towards the robot rather than relying extensively on the human, as shown in the increase of removal count by the robot in last column of Table II.

TABLE II
Simulation Runtimes (seconds)

Experiment	Avg. Time (s)	Human Ct.	Robot Ct.
Human Only	24	8	0
Robot Only (Random)	40	0	6
Robot Only (DISCERN)	27	0	6
Collaborative (Random)	19	6	2
Collaborative (DISCERN)	15	4	4

VI. Conclusions and Future Work

We design a novel approach *"DISCERN"*, able to *discern* robotic settings. With minimal setup and no pre-labeled training data, it adequately *discerns* its surroundings via CSK-based classes, rules & priorities. To the best of our knowledge, DISCERN is unique in its architecture, especially its *ordering* & *absolute* rules on CSK premises, enhancing multipurpose context discernment. It is efficient & explainable, a big asset to the big data world. In the future, we will deploy DISCERN in a live robotics lab to test it in real-world settings. We also hope to test DISCERN in sorting tasks with state-of-the-art pure-ML algorithms, in order to compare compute time, performance, and generalizability to unseen contexts. Additionally, we recognize that deployment in households necessitates the model to be robust in different cultural settings. Accordingly, we hope to address cultural variations by integrating methodologies such as those in Candle [14], which extract and organize cultural commonsense knowledge (CCSK) to enhance situative adaptability and inclusivity for diverse user needs. We can delve into other advances such as machine unlearning [15] with respect to robotics to address further perspectives.

In general, in addition to demonstrating an interesting application of CSK extraction [16] in line with modern-day AI systems, the DISCERN system contributes to sustainable AI, and explainable AI, hence making broader impacts on responsible AI [17] as well.

Acknowledgements

This work is supported by the NYU Dean's Undergraduate Research Fund for Swagnik Roychoudhury. The commonsense knowledge project thrives on a research visit by Dr. Aparna Varde at Max Planck Institute for Informatics, Saarbrucken, Germany, with further work at Montclair. She also acknowledges NSF grants 2117308 and 2018575.

References

[1] S. Roychoudhury, A. S. Varde, and W. Wang, "Efficient task organization with commonsense knowledge for human-robot collaborative tasks," in *IEEE MIT URTC*, 2024.
[2] A. Van Wynsberghe, "Sustainable ai: Ai for sustainability and the sustainability of ai," *AI and Ethics*, vol. 1, no. 3, pp. 213–218, 2021.
[3] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, *et al.*, "Explainable ai (xai): Core ideas, techniques, and solutions," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–33, 2023.
[4] A. Mandlekar, F. Ramos, B. Boots, S. Savarese, L. Fei-Fei, A. Garg, and D. Fox, "Iris: Implicit reinforcement without interaction at scale for learning control from offline robot manipulation data," *arXiv preprint arXiv:1911.05321*, 2020.
[5] J. Guérin, T. Stéphane, E. Nyiri, and O. Gibaru, "Unsupervised robotic sorting : Towards autonomous decision making robots," *Intl. J. of Artificial Intelligence & Applications*, vol. 9, pp. 81–98, 03 2018.
[6] E. Davis, *Representations of commonsense knowledge*. Morgan Kaufmann, 2014.
[7] F. Ocker, J. Deigmöller, and J. Eggert, "Exploring large language models as a source of common-sense knowledge for robots," *arXiv preprint arXiv:2311.08412*, 2023.
[8] J.-P. Töberg, A.-C. Ngonga Ngomo, M. Beetz, and P. Cimiano, "Commonsense knowledge in cognitive robotics: a systematic literature review," *Frontiers in Robotics and AI*, vol. 11, 2024.
[9] N. Tandon, A. S. Varde, and G. de Melo, "Commonsense knowledge in machine intelligence," *ACM SIGMOD Record*, vol. 46, no. 4, pp. 49–52, 2017.
[10] R. Hidalgo, J. Parron, A. S. Varde, and W. Wang, "Robo-csk-organizer: Commonsense knowledge to organize detected objects for multipurpose robots," *arXiv preprint arXiv:2409.18385*, 2024.
[11] I. Chernyavsky, A. S. Varde, and S. Razniewski, "Csk-detector: Commonsense in object detection," in *2022 IEEE International Conference on Big Data (Big Data)*, pp. 6609–6612, IEEE, 2022.
[12] C. J. Conti, A. S. Varde, and W. Wang, "Human-robot collaboration with commonsense reasoning in smart manufacturing contexts," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 3, pp. 1784–1797, 2022.
[13] R. Speer, J. Chin, and C. Havasi, "Conceptnet 5.5: an open multilingual graph of general knowledge," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, p. 4444–4451, 2017.
[14] T.-P. Nguyen, S. Razniewski, A. Varde, and G. Weikum, "Extracting cultural commonsense knowledge at scale," in *Proceedings of the ACM Web Conference 2023*, pp. 1907–1917, 2023.
[15] L. Bourtoule, V. Chandrasekaran, C. A. Choquette-Choo, H. Jia, A. Travers, B. Zhang, D. Lie, and N. Papernot, "Machine unlearning," in *2021 IEEE Symposium on Security and Privacy (SP)*, pp. 141–159, IEEE, 2021.
[16] S. Razniewski, N. Tandon, and A. S. Varde, "Information to wisdom: Commonsense knowledge extraction and compilation," in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pp. 1143–1146, 2021.
[17] V. Dignum, *Responsible artificial intelligence: how to develop and use AI in a responsible way*, vol. 2156. Springer, 2019.