2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

Rearrange Indoor Scenes for Human-Robot Co-Activity

Weiqi Wang[1,⋆] Zihang Zhao[2,3,⋆] Ziyuan Jiao[1,2,⋆] Yixin Zhu[4,†] Song-Chun Zhu[2,4] Hangxin Liu[2,†]

https://sites.google.com/view/coactivity

***Abstract*— We present an optimization-based framework for rearranging indoor furniture to accommodate human-robot co-activities better. The rearrangement aims to afford sufficient accessible space for robot activities without compromising everyday human activities. To retain human activities, our algorithm preserves the functional relations among furniture by integrating spatial and semantic co-occurrence extracted from SUNCG and ConceptNet, respectively. By defining the robot's accessible space by the amount of open space it can traverse and the number of objects it can reach, we formulate the rearrangement for human-robot co-activity as an optimization problem, solved by adaptive simulated annealing (ASA) and covariance matrix adaptation evolution strategy (CMA-ES). Our experiments on the SUNCG dataset quantitatively show that rearranged scenes provide a robot with 14% more accessible space and 30% more objects to interact with on average. The quality of the rearranged scenes is qualitatively validated by a human study, indicating the efficacy of the proposed strategy.**

I. Introduction

Service robots are gaining popularity in domestic settings, where they are expected to perform various complex household tasks. Typically, indoor scenes are designed and organized in accordance with human needs, often too confined and clustered for conventional service robots to navigate and interact. To tackle these challenges, researchers have developed several effective planning algorithms, designed to (i) retrieve objects in confined and cluttered spaces [1–3], (ii) coordinate whole-body motions for articulate objects [4–6], (iii) coordinate foot-arm via virtual mechanisms for sequential manipulation tasks [7, 8], and (iv) integrate robot perception and task planning for household environments [9–11]. Nevertheless, indoor scenes designed purely for humans may fundamentally prohibit robot activities due to their different morphology and movement patterns; robots with bulky embodiments have difficulty in imitating human motions and require larger open spaces.

Fig. 1a illustrates the above problem. In this example, a person may easily approach and interact with the nightstand, whereas a robot cannot due to its bulky mobile base that is larger than the space between the bed and the wall. Similarly, a robot cannot reach the bookshelf due to the narrow passageway between the chair and the bed. These prevalent restrictions of service robots significantly restrict the robot's capabilities in household environments. Notably, better planning algorithms cannot solve this problem; the room layout must be optimized for human-robot co-activity.

(a) human activity only (b) human-robot co-activity

Fig. 1: **Rearrange indoor scenes for human-robot co-activity.** (a) A person may pass through the narrow passages at A and B for daily activities, whereas a robot cannot due to its larger footprint. As a result, the robot's activities are limited in household environments designed purely for human activities. (b) A rearranged scene, optimized for human-robot co-activity, provides sufficient open space for robot activities while preserving human preferences.

We argue that the essence of scene rearrangement for human-robot co-activity is to preserve human preference for indoor activities while affording more robot activities in terms of accessible space and interactions with objects. This dual objective requires simultaneously modeling both **human preference** and **robot preference** in the indoor scenes.

Scene synthesis has primarily modeled human preference and automatically generates scene layout from scratch subjects to constraints. Classic methods exploit simple heuristics to construct these constraints, such as interior design guidelines [12] or user-provided positive examples [13]. Modern treatments adopt learning-based approach, such as learning object-object relations [14, 15], modeling human activities with objects [16, 17], mining topological relationships among object groups [18], and capturing latent information via generative models [19–21]. In particular, the common assumption in scene synthesis is that large datasets capture the statistics (*i.e.*, scene layouts) for downstream tasks. This assumption no longer holds for human-robot co-activity as existing datasets only capture human preference, lacking statistics to model their robot counterpart.

Scene rearrangement has modeled either human or robot preference individually, such as to (i) reduce the risk of patient falls [22], (ii) improve robot navigation efficiency [23], (iii) boost workspace task performance [24], and (iv) promote collaboration [25]. The lack of joint modeling of both human and robot preferences calls for alternative approaches.

⋆ W. Wang, Z. Zhao, and Z. Jiao contributed equally to this work. † Corresponding authors. Emails: yixin.zhu@pku.edu.cn, liuhx@bigai.ai.

[1] UCLA Center for Vision, Cognition, Learning, and Autonomy (VCLA) [2] National Key Laboratory of General Artificial Intelligence, Beijing Institute for General Artificial Intelligence (BIGAI) [3] College of Engineering, Peking University [4] Institute for Artificial Intelligence, Peking University

2023 IEEE International Conference on Robotics and Automation (ICRA) | 979-8-3503-2365-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICRA48891.2023.10160303

To model **human preference**, we encode it implicitly by functional groups of objects (*i.e.*, furniture) [26, 27]. Intuitively, we place a nightstand beside a bed and a chair alongside a table. More broadly, objects within a functional group should be (re)arranged together, whereas their relative poses can vary within a smaller range. As such, a straightforward idea to model functional groups is based on spatial co-occurrence, *i.e.*, the frequency with which two objects appear together and in close proximity [13, 14, 28, 29]. Nonetheless, such a statistical perspective might be deceiving; the proximity of two objects does not necessarily suggest that they belong to the same functional group, especially in cluttered indoor scenes. In Fig. 1a, the chair and the bed, as well as the nightstand and the bed, are pretty close to each other. However, the bed and the chair are not in the same functional group. To overcome the ambiguity stemming from spatial co-occurrence, we employ ConceptNet [30], a sizeable open-source knowledge graph database, to refine functional relations based on additional semantic co-occurrence of objects.

To model **robot preference**, we utilize *accessible space*: the amount of open space a robot can traverse and the number of objects it can reach. Intuitively, a scene must afford sufficient open space for a robot to explore while performing given tasks. Objects must also be properly oriented for successful manipulation; for instance, a desk can be placed against a wall, but cabinets and drawers must face outside. To effectively encode and compute various possible interactions between a robot and a scene, we introduce a signed distance field (SDF) to represent (i) the scene's navigable area, given robot footprint and object placements, and (ii) the interaction affordance defined on the object boundary, akin to "dark matters" [31, 32] that attract or repel possible interactions.

Computationally, we design an optimization framework of scene rearrangement for human-robot co-activity that takes as input the above robot and human preferences. Fig. 1b illustrates an exemplar result, in which robot and human preferences are co-optimized: (i) the robot functions more efficiently due to larger open space and potentially more interactions with objects, and (ii) the objects within functional groups remain close to satisfying human preference. In the experiment, we evaluate our method using the SUNCG dataset [33]. Not only do the rearranged scene layouts afford an average of 30% more robot activities and 14% more open space, but also keep the *Naturalness* based on a human study.

Our **contributions** are threefold: (i) We develop a new method to capture human preferences via functional relations among objects by combining the spatial and semantic co-occurrence. (ii) We model robot preferences by its accessible space, represented by an SDF for efficient computation. (iii) We devise an optimization-based framework to balance human and robot preferences when rearranging scene layouts.

The remainder of this paper is organized as follows. Sec. II formally describes our modeling of human and robot preferences, subsequently formulated as an optimization framework presented in Sec. III. Experimental results are presented in Sec. IV. We conclude the paper in Sec. V.

II. Human and Robot Preferences

In this section, we describe how we model (i) human preference by combining objects' semantic and spatial co-occurrences to determine functional object groups and (ii) robot preference by its accessible space in the scene.

A. Human Preference

We model human preference implicitly by functional object groups. Discovering functional groups that emerge from objects in cluttered scenes is challenging due to the inherited ambiguity of objects' functional relations. Two nearby objects with close proximity do not necessarily indicate that they are functionally related (*e.g.*, the example of the bed and chair shown in Fig. 2a), and objects with similar semantics (*e.g.*, a desk and a coffee table) also do not imply that they are within a functional group. Hence, relying solely on spatial or semantics alone cannot correctly uncover the functional groups. In this work, we seek an integrated solution.

Objective: We employ a weighted scene graph $\mathcal{G} = (V, S)$ to represent the relations among objects within a scene. Specifically, an object is represented by a node $v_i = \langle o_i, B_i, p_i \rangle \in V$ in the scene graph, including object label o_i, oriented 3D bounding box B_i, and the object planar pose p_i. Each edge $s_{i,j} = \langle v_i, v_j, w_{ij} \rangle \in S$ is a tuple that encodes the relation between two objects v_i and v_j. The edge weight w_{ij} indicates how likely v_i and v_j are within a functional group:

$$w_{ij} = P(v_i, v_j|\mathcal{G}) = \frac{1}{Z} P_{sem}(o_i, o_j|\mathcal{G}) P_{spa}(o_i, o_j|\mathcal{G}), \quad (1)$$

where Z is a normalizing factor obtained by summing over all the products of the pairwise semantic/spatial probabilities:

$$Z = \sum_{o_i, o_j, i \neq j} P_{sem}(o_i, o_j|\mathcal{G}) P_{spa}(o_i, o_j|\mathcal{G}), \quad (2)$$

and $P_{sem}(\cdot)$ and $P_{spa}(\cdot)$ are the probabilities reflecting the semantic and spatial correlations of two objects, respectively.

Semantic Relation: To obtain two object's semantic relation $P_{sem}(o_i, o_j|\mathcal{G})$, we utilize ConceptNet [30], an open-source knowledge graph database that characterizes the strength of the semantic relation between two object labels o_i and o_j by $h_{ij} \in [0, 1]$; a greater h_{ij} suggests a stronger semantic relation. Unfortunately, ConceptNet would assign a large h_{ij} to two synonyms (*e.g.*, a chair and an armchair, a desk and a table), resulting in wrong functional relations between similar objects. To tackle this issue, when o_i is synonymous with o_j, we replace h_{ij} returned from ConceptNet by h averaged over all other pairs of objects in the scene to indicate a neutral relation:

$$h^*_{ij} = \begin{cases} h_{ij}, & \neg\ \texttt{IsA}(o_i, o_j) \\ \texttt{Avg}(\{h_{mn} | \neg\ \texttt{IsA}(o_m, o_n)\}), & \text{otherwise} \end{cases}, \quad (3)$$

where IsA is an edge type, indicating o_i is synonymous with o_j. Notably, we do not set h between two synonyms to 0 to avoid rearranging these two objects apart. Taken together, the semantic relation in Equation (1) is given by:

$$P_{sem}(o_i, o_j|\mathcal{G}) = \frac{h^*_{ij}}{\sum_{s \in S} h}. \quad (4)$$

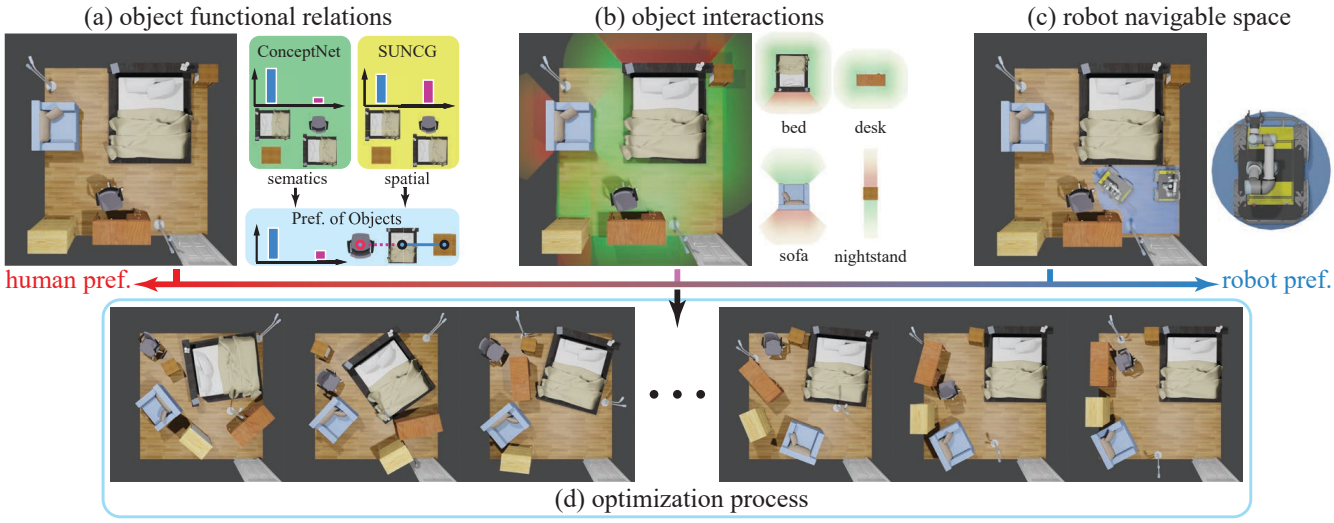


Fig. 2: **Essential factors to rearrange scenes for human-robot co-activity.** (a) To preserve human preference encoded by object relations, we model object co-occurrence using semantic and spatial relations extracted from ConceptNet and SUNCG, respectively. (b) Robots can only interact with certain furniture from specific directions, such as approaching a cabinet or drawer from the front. Hence, we devise a pseudo-interaction function to indicate the desired furniture orientation. (c) After fitting an inflated circular base to the robot's footprint, the robot's accessible space in the scene (shaded blue) can be determined by an SDF. (d) These three factors are formalized into an optimization framework to rearrange scenes for human-robot co-activity.

Spatial Relation: To determine how likely objects o_i and o_j are related based on their spatial distance d_{ij}, we query their co-occurrence in the SUNCG dataset:

$$P_{spa}(o_i, o_j|\mathcal{G}) \propto P_d(d_{ij}|o_i, o_j) P_{co}(o_i, o_j), \tag{5}$$

where $P_d(\cdot|o_i, o_j)$ is the distribution of relative poses between o_i and o_j in the SUNCG dataset. In practice, $P_d(\cdot|o_i, o_j)$ is biased if o_i and o_j rarely co-occur in the same scene. Hence, we add a term P_{co} for bias correction:

$$P_{co}(o_i, o_j) = \frac{N_{ij}}{\min\left(\sum_{k\in O} N_{jk}, \sum_{k\in O} N_{ik}\right)}, \tag{6}$$

where N_{ij} is the number of co-occurrences of o_i and o_j in the dataset, and O is the set of all semantic labels.

B. Robot Preference

Due to the existence of narrow passages and obstacles, a human-made environment is unfit for the functioning of a service robot. By rearranging the scene, we seek to expand the open space for robot activity. Specifically, we model the robot preference for a scene based on its accessible space $\mathcal{R}$, which has a closed boundary. The accessible space of a robot consists of two components: (i) the open space it can traverse and (ii) the number of objects it can reach.

Size of Open Space: The open space that a service robot can traverse can be effectively computed by:

$$f_{\mathcal{R}}(q) = f_{\mathcal{B}}(q) - r_b, \tag{7}$$

where $f_{\mathcal{B}} : \mathbb{R}^2 \to \mathbb{R}$ is an SDF that measures the shortest signed Euclidean distance from a query point q to the bounding boxes of objects in the scene. When the robot is outside, on the boundary, or inside (*i.e.*, invalid robot pose), the SDF's value is larger than, equal to, or less than zero, respectively. After imposing a circular base inflated with a radius of r_b as a safety margin, $f_{\mathcal{R}} : \mathbb{R}^2 \to \mathbb{R}$ denotes the entire open space of the robot.

Number of Interacting Objects: Although the size of the open space reveals a robot's capability to traverse the environment, it does not necessarily reflect a robot's ability to interact with objects. For instance, the cabinet door should not face the wall, as a robot cannot interact with it otherwise. Fig. 2b provides other examples; the green shaded areas around different objects indicate how a robot may approach and interact with these objects, whereas the red shaded areas imply the opposite. Computationally, we further introduce a pseudo-interaction function:

$$f_{\mathcal{I}_i}(q) = \text{sgn}(q) \cdot \max(0, -\frac{f^+_{\mathcal{B}_i}(q)}{d^{\max}} + 1), \tag{8}$$

where $f^+_{\mathcal{B}_i}(q)$ is an SDF that only returns a positive distance from a query point q (*e.g.*, a robot pose) to the bounding box of the object v_i, $d^{\max}$ is the normalizing factor set to the maximum distance that the robot arm can reach, and $\text{sgn}(\cdot)$ assigns positive weights to the green shaded areas in Fig. 2b (*i.e.*, the area where a robot would interact with the object) and negative to the red shaded areas.

III. Optimization

After building both human and robot preferences, we devise an optimization framework that balances these preferences and rearranges scenes accordingly. The resulting scene supports better human-robot co-activity.

A. Objectives

Human Term: Recall that human preference is modeled as a weighted scene graph $\mathcal{G}$ defined in Equation (1). We first cluster the edges in $\mathcal{G}$ w.r.t. the edge weight by Gaussian mixture models (GMM). Next, assuming that edges within the same functional group have large weights, we prune cluster edges that have small weights, resulting in a

filtered scene graph $\mathcal{G}^\star$. As a result, each connected subgraph $\mathcal{G}^\star$ is a functional group. Finally, the human term is defined as:

$$H_{s_{i,j}} = 1 - \frac{P_d\left(d|o_i, o_j; s^\star_{ij} \in \mathcal{G}^\star\right)}{\sup ||P_d\left(\cdot|o_i, o_j; s^\star_{ij} \in \mathcal{G}^\star\right)||}, \tag{9}$$

where $s^\star_{ij}$ is the edge connecting o_i and o_j in $\mathcal{G}^\star$.

Robot Term: The robot term is the combination of Equations (7) and (8):

$$I_i = -\int_{q \in \mathcal{I}_i \cap \mathcal{R}} f_{\mathcal{I}_i}(q) + \alpha f_{\mathcal{R}}(q) dq, \tag{10}$$

where α is an empirically set balancing constant. A smaller I_i is preferred for more open space for the robot.

Objective: Let $\psi = \{p_i | v_i \in V\}$ denote the scene layout. We formulate the problem of rearranging scenes as an optimization problem that minimizes the human and robot terms introduced in Equations (9) and (10).

$$\begin{aligned} \min_{\psi} \quad & \sum_{s^\star \in \mathcal{G}^\star} H_s + \beta \sum_{v_i \in V} I_i, \\ \text{s.t.} \quad & d(v_i, v_j) > 0, \quad \forall v_i, v_j \in V, i \neq j, \end{aligned} \tag{11}$$

where β is an empirically set balancing constant and the constraint d forces the minimum distance between object pairs to be positive; *i.e.*, each object pair is collision-free.

B. Optimization

We adopt the adaptive simulated annealing (ASA) [34] to solve the above optimization problem. ASA searches for a global minimum in an energy landscape defined by the objective function. Specifically, a candidate scene layout is sampled from a uniform distribution and is accepted based on the Metropolis criterion. ASA adjusts step size after several optimization iterations to maintain an approximately equal number of accepted and rejected samples for each variable. In practice, we find that ASA excels in escaping local minima but struggles to converge due to the enormous search space.

Search Space: We hierarchically optimize the layout to reduce the search space. The optimization is decomposed into two steps given a native human-centric scene. First, each functional group of the scene is treated as a sub-scene and is optimized independently w.r.t. Equation (11). Second, an optimized functional group is treated as a single object, and the scene layout is optimized over functional groups. Fig. 2d shows an optimization process with intermediate layouts.

Convergence: We adopt covariance matrix adaptation evolution strategy (CMA-ES) to expedite convergence. CMA-ES is a derivative-free stochastic method for numerical optimization, which repeatedly applies the survival of the fittest process to its population and rapidly converges to a nearby local minimum. At the beginning of each iteration, CMA-ES draws samples from a multivariate normal distribution: $\psi' \sim \mathcal{N}(m, \sigma^2 C)$, where m is the weighted mean of the most promising layouts of previous samples, σ is the overall standard deviation or step size, and C is the estimated covariance matrix. At the end of each iteration, the algorithm updates these parameters according to the performance of the population, shifting the expected variance in the same direction as the estimated gradient.

IV. Experiments

We test our method on SUNCG [33]. We randomly select 90% of the scenes for learning and the remaining for testing. Seven scenes were chosen for human evaluation.

A. Qualitative Results

Fig. 3 qualitatively show 10 scenes. The blue shaded area depicts the robot's accessible space, whereas the red objects are out of reach. Comparing optimization based solely on robot preference (Fig. 3b) with optimization based on both human and robot preferences (Fig. 3c), we qualitatively demonstrate that our method (i) simultaneously increases the robot's accessible space and affords more robot interactions with objects, and (ii) properly maintain the human preference encoded by the functional objects; *e.g.*, the desk–chair and bed–nightstand functional object pairs are together.

B. Quantitative Results

Robot Preference: We devise two evaluation criteria: (i) a simple heuristic based on the number of reachable objects, and (ii) the robot term defined in Equation (10). Fig. 4 summarizes quantitative results aggregated from 154 scenes randomly selected from the test set. Specifically, it plots the change of these two criteria. We note that most of the rearranged scenes are located in the first quadrant, indicating greater support for robot activity. Some scenes are located in the fourth quadrant because interaction spaces were sacrificed in exchange for more accessible objects.

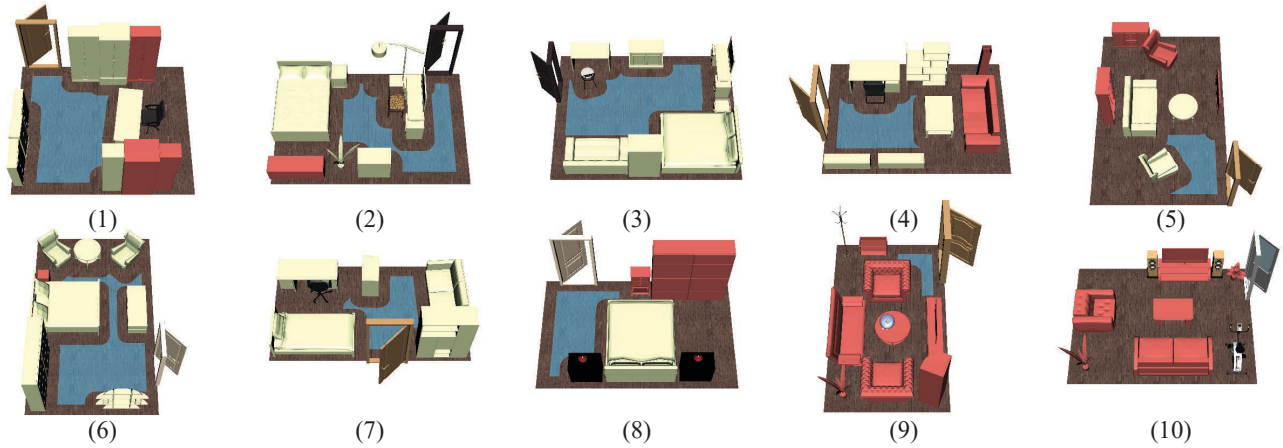
Human Preference: We conducted a human study to validate if the rearranged scenes preserve human preference. 11 participants were recruited to evaluate 7 pairs (Fig. 3(1)–(7)) of original (Fig. 3a) and rearranged (Fig. 3c) scenes:

- The *Naturalness* (How natural does the scene look?);
- The *Functionality* (How well do you think you can perform your daily activities in the scene?).

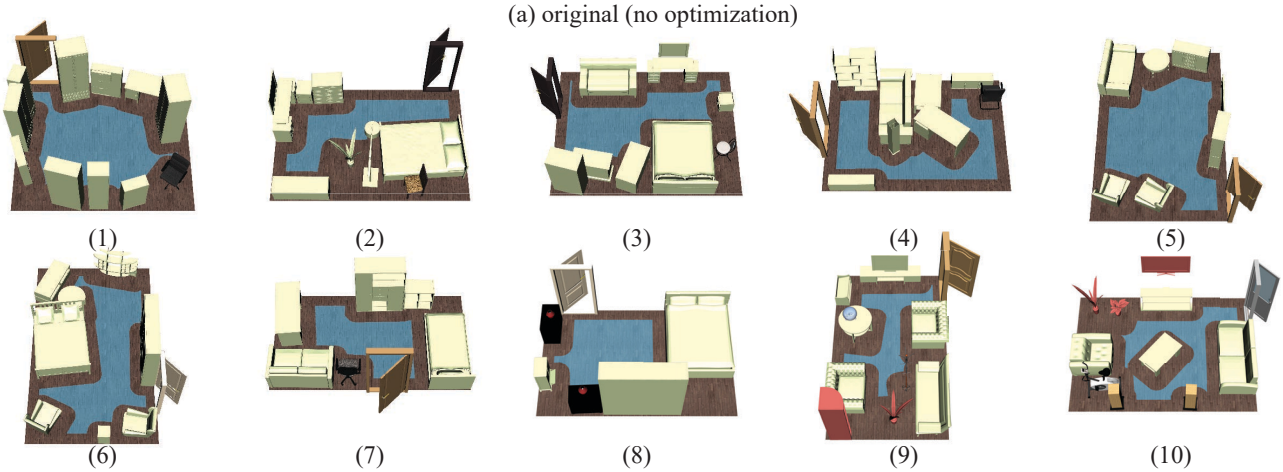
They were asked to provide ratings to the above questions on a scale of 1 to 5, with 1 being *not at all* and 5 *very much*.

The collected responses were analyzed using an independent samples t-test with a significance level of 0.05. Five of seven scenes (*i.e.*, Fig. 3(1)–(5)) lack statistical significance for either *Naturalness* or *Functionality*. Although the rearranged scenes tend to obtain slightly lower human ratings than their original layouts, the insignificance indicates that the rearranged scenes do not significantly diminish human values, while improving the robots' accessible space and the number of supported activities.

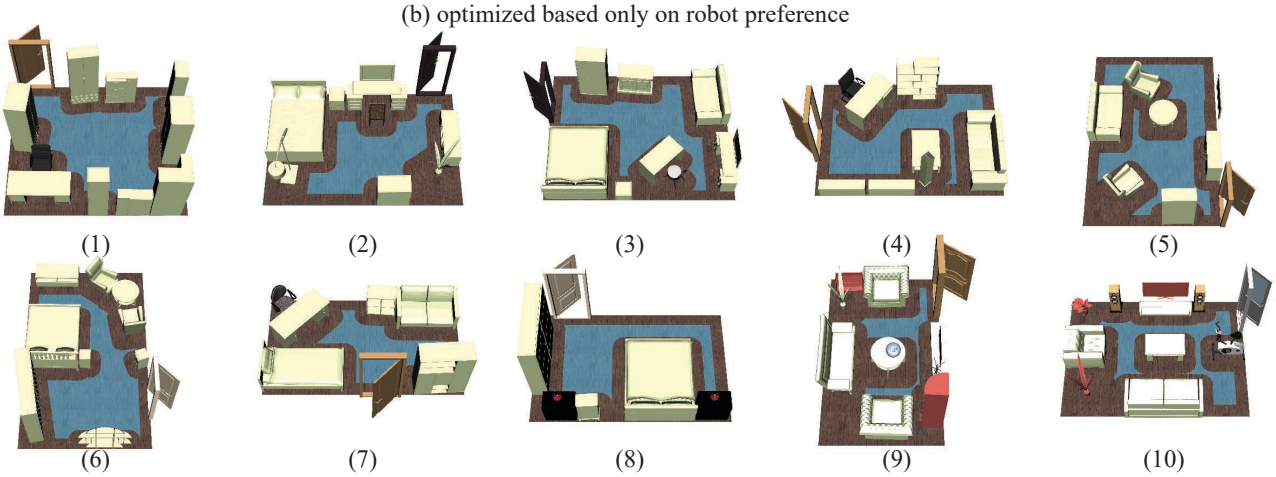
The differences in human rating for scene 6 and scene 7 are statistically significant, with $t(10) = 1.0, p = 0.0029$ and $t(10) = 1.0, p = 0.0023$, respectively. Although the rearranged scenes are more friendly to robots, human participants appear to be quite sensitive to unusual object arrangements. After rearrangement, they can quickly identify the undesired orientation of the bed in Fig. 3(6) and the chair blocked by the desk in Fig. 3(7), resulting in much lower ratings for these two scenes.



(a) original (no optimization)



(b) optimized based only on robot preference



(c) optimized based on both human and robot preferences

Fig. 3: **Examples of rearranged scenes.** Blue area denotes the accessible space, whereas the red objects are out of reach.

C. Ablations

Robot Factors: To further study how robot specifications may impact the results of scene rearrangement, we use three living areas presented in Fig. 6a to illustrate how our method performs differently for two distinct robots: Husky with UR5, which is larger, and Dingo with UR3, which is smaller. The scenes in Fig. 6(1) and Fig. 6(2) depict the scenes optimized for Dingo and Husky robots, respectively. Despite the final layouts for both robots (Fig. 6b and Fig. 6c) are similar, the layout in Fig. 6(3) is still not optimal for the Husky robot due to its larger size. These results suggest that if the robot's dimension exceeds a particular threshold, it may be impossible to design a layout that accommodates the robot's activities. Similarly, the improvements offered by the scene rearrangement based on a specific type of robot are only partially transferable to another, possibly requiring an entirely new scene rearrangement.

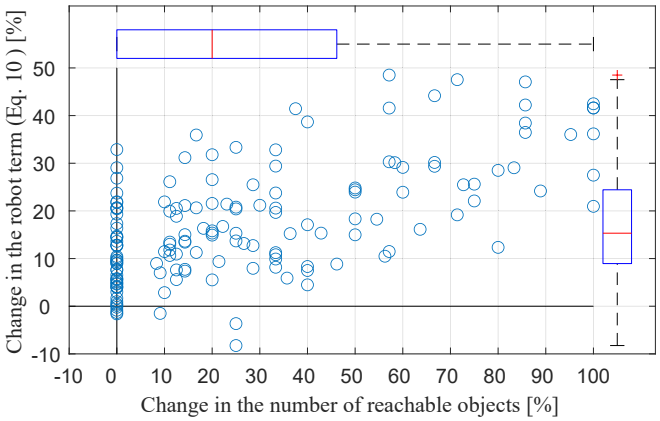


Fig. 4: **Quantitative results evaluated on 154 scenes in SUNCG dataset.** The scatter plot depicts the improvement of each scene in terms of robot preference following scene rearrangement. The boxplots depict the distribution of improvements along each axis, with the red horizontal lines representing the median improvements, and the bottom and top edges of the blue boxes representing the 25th and 75th percentiles, respectively.

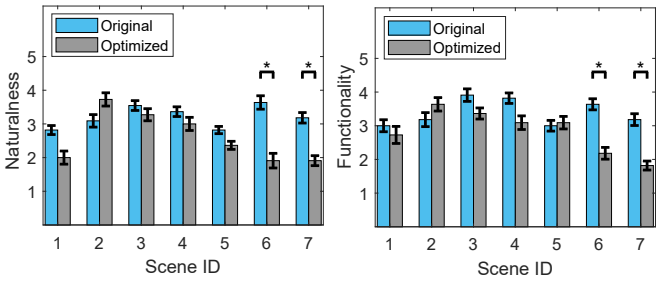


Fig. 5: **Averaged human rating scores regarding whether or not rearranged scenes maintain preferences in terms of *Naturalness* and *Functionality*.** The black error bars represent a 95% confidence interval. The first five pairs of scenes following the order in Fig. 3 are not statistically significant, whereas the last two are.

Task-Specific Factors: In this ablation, we demonstrate the applicability of our method to task-specific scenarios by defining *Robot Motion Cost* $M_T = \sum_{t \in T} \mathcal{L}(p(q_i^t, q_j^t))$ as the robot's total navigation distance of a set of activities T. Adding this cost term to the optimization objective defined in Equation (11) can rearrange scenes to reduce the robot's motion effort further, hence improving its task efficiency. In an office environment depicted in Fig. 7a, which contains three types of functional groups—working, relaxing, and dining—the robot has been assigned three tasks involving numerous activities. The "clean" task (red paths, Fig. 7b) is performed primarily in the working functional group, the "restocking" task (yellow paths, Fig. 7c) requires the robot to visit all cabinets and shelves in the scene, and the "distribution" task (green paths, Fig. 7d) requires the robot to traverse among all three functional groups. These scenes are rearranged according to the Dingo robot's specifications.

V. Conclusion

This paper presented an optimization-based framework for rearranging indoor scenes according to both human and robot preferences. Specifically, the human preference was captured by uncovering the functional relations among objects governing their arrangements, modeled by their

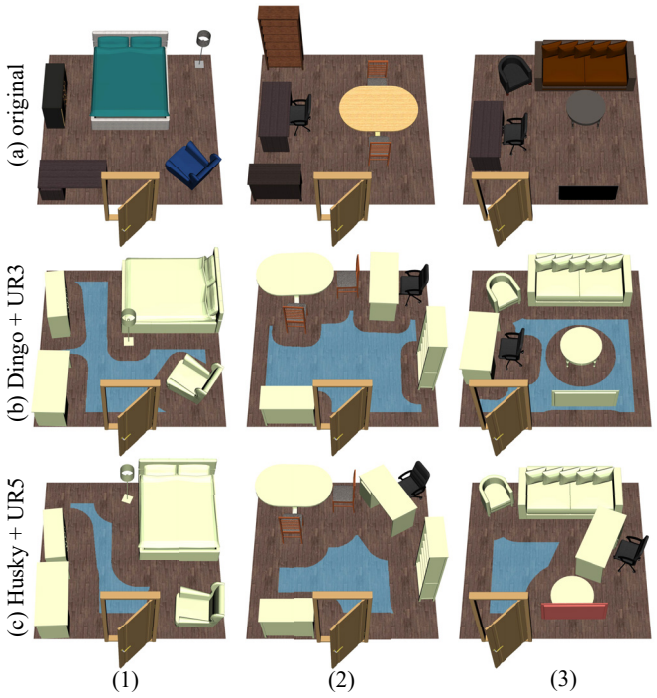


Fig. 6: **Scenes rearranged for Dingo and Husky.**

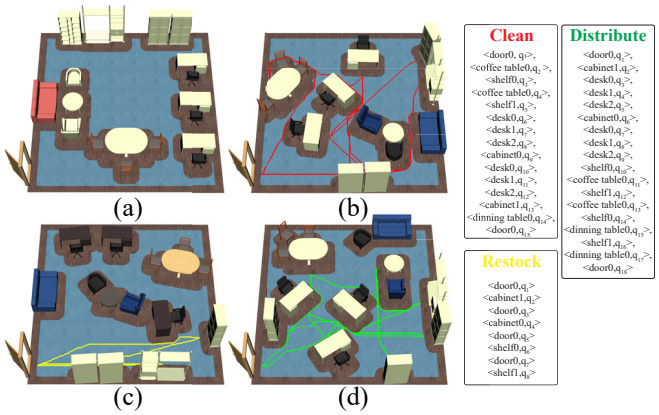


Fig. 7: **Scenes optimized for human-robot co-activity given various activities.**

semantic and spatial co-occurrences from the ConceptNet and SUNCG datasets. The robot preference was represented by its accessible space. We formulated these factors into an optimization problem that rearranges a given scene by optimizing furniture layouts. Experimental results showed that the proposed method expands the open space, increases the number of reachable objects, and minimizes traveling effort in robot activities. Moreover, our human study revealed that most of the rearranged scenes remained natural and acceptable to humans, as the ratings were statistically insignificant compared to the original layouts. These findings signified that rearranges produced by the proposed method effectively promote human-robot co-activities.

Acknowledgments: This work is supported in part by the National Key R&D Program of China (2022ZD0114900), the Beijing Municipal Science & Technology Commission (Z221100003422004), and the Beijing Nova Program.

REFERENCES

[1] D. Berenson, J. Kuffner, and H. Choset, "An optimization approach to planning for mobile manipulation," in *International Conference on Robotics and Automation (ICRA)*, 2008.

[2] S. H. Cheong, B. Y. Cho, J. Lee, C. Kim, and C. Nam, "Where to relocate?: Object rearrangement inside cluttered and confined environments for robotic manipulation," in *International Conference on Robotics and Automation (ICRA)*, 2020.

[3] Z. Han, J. Allspaw, G. LeMasurier, J. Parrillo, D. Giger, S. R. Ahmadzadeh, and H. A. Yanco, "Towards mobile multi-task manipulation in a confined and integrated environment with irregular objects," in *International Conference on Robotics and Automation (ICRA)*, 2020.

[4] K. Shankar, *Kinematics and Local Motion Planning for Quasi-static Whole-body Mobile Manipulation*. PhD thesis, California Institute of Technology, 2016.

[5] D. M. Bodily, T. F. Allen, and M. D. Killpack, "Motion planning for mobile robots using inverse kinematics branching," in *International Conference on Robotics and Automation (ICRA)*, 2017.

[6] S. Chitta, B. Cohen, and M. Likhachev, "Planning for autonomous door opening with a mobile manipulator," in *International Conference on Robotics and Automation (ICRA)*, 2010.

[7] Z. Jiao, Z. Zhang, X. Jiang, D. Han, S.-C. Zhu, Y. Zhu, and H. Liu, "A virtual kinematic chain perspective for manipulation in household environments," in *International Conference on Intelligent Robots and Systems (IROS)*, 2021.

[8] Z. Jiao, Z. Zhang, W. Wang, D. Han, S.-C. Zhu, Y. Zhu, and H. Liu, "Efficient task planning for mobile manipulation: a virtual kinematic chain perspective," in *International Conference on Intelligent Robots and Systems (IROS)*, 2021.

[9] M. Han, Z. Zhang, Z. Jiao, X. Xie, Y. Zhu, S.-C. Zhu, and H. Liu, "Reconstructing interactive 3d scenes by panoptic mapping and cad model alignments," in *International Conference on Robotics and Automation (ICRA)*, 2021.

[10] Z. Jiao, Y. Niu, Z. Zhang, S.-C. Zhu, Y. Zhu, and H. Liu, "Sequential manipulation planning on scene graph," in *International Conference on Intelligent Robots and Systems (IROS)*, 2022.

[11] M. Han, Z. Zhang, Z. Jiao, X. Xie, Y. Zhu, S.-C. Zhu, and H. Liu, "Scene reconstruction with functional objects for robot autonomy," *International Journal of Computer Vision (IJCV)*, vol. 130, no. 12, pp. 2940–2961, 2022.

[12] P. Merrell, E. Schkufza, Z. Li, M. Agrawala, and V. Koltun, "Interactive furniture layout using interior design guidelines," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4, pp. 1–10, 2011.

[13] M. Fisher, D. Ritchie, M. Savva, T. Funkhouser, and P. Hanrahan, "Example-based synthesis of 3d object arrangements," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 6, pp. 1–11, 2012.

[14] L. F. Yu, S. K. Yeung, C. K. Tang, D. Terzopoulos, T. F. Chan, and S. J. Osher, "Make it home: automatic optimization of furniture arrangement," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4, 2011.

[15] Z. S. Kermani, Z. Liao, P. Tan, and H. Zhang, "Learning 3d scene synthesis from annotated rgb-d images," *Computer Graphics Forum (CGF)*, vol. 35, no. 5, pp. 197–206, 2016.

[16] S. Qi, Y. Zhu, S. Huang, C. Jiang, and S.-C. Zhu, "Human-centric indoor scene synthesis using stochastic grammar," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[17] C. Jiang, S. Qi, Y. Zhu, S. Huang, J. Lin, L.-F. Yu, D. Terzopoulos, and S.-C. Zhu, "Configurable 3d scene synthesis and 2d image rendering with per-pixel ground truth using stochastic grammars," *International Journal of Computer Vision (IJCV)*, vol. 126, no. 9, pp. 920–941, 2018.

[18] M. Keshavarzi, A. Parikh, X. Zhai, M. Mao, L. Caldas, and A. Y. Yang, "Scenegen: Generative contextual scene augmentation using scene graph priors," *arXiv preprint arXiv:2009.12395*, 2020.

[19] D. Ritchie, K. Wang, and Y.-a. Lin, "Fast and flexible indoor scene synthesis via deep convolutional generative models," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[20] Z. Zhang, Z. Yang, C. Ma, L. Luo, A. Huth, E. Vouga, and Q. Huang, "Deep generative modeling for scene synthesis via hybrid representations," *ACM Transactions on Graphics (TOG)*, vol. 39, no. 2, pp. 1–21, 2020.

[21] N. Nauata, S. Hosseini, K.-H. Chang, H. Chu, C.-Y. Cheng, and Y. Furukawa, "House-gan++: Generative adversarial layout refinement network towards intelligent computational agent for professional architects," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[22] S. Chaeibakhsh, R. S. Novin, T. Hermans, A. Merryweather, and A. Kuntz, "Optimizing hospital room layout to reduce the risk of patient falls," in *International Conference on Operations Research and Enterprise Systems (ICORES)*, 2021.

[23] J. Zhi, L.-F. Yu, and J.-M. Lien, "Designing human-robot coexistence space," *IEEE Robotics and Automation Letters (RA-L)*, vol. 6, no. 4, pp. 7161–7168, 2021.

[24] W. Liang, J. Liu, Y. Lang, B. Ning, and L.-F. Yu, "Functional workspace optimization via learning personal preferences from virtual experiences," *IEEE Transactions on Visualization and Computer Graph (TVCG)*, vol. 25, no. 5, pp. 1836–1845, 2019.

[25] Y. Zhang, H. Huang, E. Plaku, and L.-F. Yu, "Joint computational design of workspaces and workplans," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 6, pp. 1–16, 2021.

[26] Y. Zhao and S.-C. Zhu, "Image parsing with stochastic scene grammar," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2011.

[27] Y. Chen, S. Huang, T. Yuan, S. Qi, Y. Zhu, and S.-C. Zhu, "Holistic++ scene understanding: Single-view 3d holistic scene parsing and human pose estimation with human-object interaction and physical commonsense," in *International Conference on Computer Vision (ICCV)*, 2019.

[28] K. Wang, Y.-A. Lin, B. Weissmann, M. Savva, A. X. Chang, and D. Ritchie, "Planit: Planning and instantiating indoor scenes with relation graph and spatial prior networks," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 1–15, 2019.

[29] S.-H. Zhang, S.-K. Zhang, W.-Y. Xie, C.-Y. Luo, Y.-L. Yang, and H. Fu, "Fast 3d indoor scene synthesis by learning spatial relation priors of objects," *IEEE Transactions on Visualization and Computer Graph (TVCG)*, vol. 28, no. 9, pp. 3082–3092, 2021.

[30] R. Speer, J. Chin, and C. Havasi, "Conceptnet 5.5: An open multilingual graph of general knowledge," in *AAAI Conference on Artificial Intelligence (AAAI)*, 2017.

[31] D. Xie, T. Shu, S. Todorovic, and S.-C. Zhu, "Learning and inferring "dark matter" and predicting human intents and trajectories in videos," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 40, no. 7, pp. 1639–1652, 2017.

[32] Y. Zhu, T. Gao, L. Fan, S. Huang, M. Edmonds, H. Liu, F. Gao, C. Zhang, S. Qi, Y. N. Wu, J. Tenenbaum, and S.-C. Zhu, "Dark, beyond deep: A paradigm shift to cognitive ai with humanlike common sense," *Engineering*, vol. 6, no. 3, pp. 310–345, 2020.

[33] S. Song, F. Yu, A. Zeng, A. X. Chang, M. Savva, and T. Funkhouser, "Semantic scene completion from a single depth image," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[34] L. Ingber *et al.*, "Adaptive simulated annealing (asa)," *Global optimization C-code, Caltech Alumni Association, Pasadena, CA*, 1993.