

Interactive Interior Design Recommendation via Coarse-to-fine Multimodal Reinforcement Learning

He Zhang

Thrust of Artificial Intelligence, The
Hong Kong University of Science and
Technology (Guangzhou)
hzhang757@connect.hkust-gz.edu.cn

Ying Sun*

Thrust of Artificial Intelligence, The
Hong Kong University of Science and
Technology (Guangzhou)
yings@hkust-gz.edu.cn

Weiyu Guo

Thrust of Artificial Intelligence, The
Hong Kong University of Science and
Technology (Guangzhou)
guoweiyu96@gmail.com

Yafei Liu

Haonan Lu
OPPO Research Institute
liuyafei@oppo.com
luhaonan@oppo.com

Xiaodong Lin

Rutgers University
lin@business.rutgers.edu

Hui Xiong*

Thrust of Artificial Intelligence, The
Hong Kong University of Science and
Technology (Guangzhou)
Guangzhou HKUST Fok Ying Tung
Research Institute
xionghui@ust.hk

ABSTRACT

Personalized interior decoration design often incurs high labor costs. Recent efforts in developing intelligent interior design systems have focused on generating textual requirement-based decoration designs while neglecting the problem of how to mine homeowner's hidden preferences and choose the proper initial design. To fill this gap, we propose an Interactive Interior Design Recommendation System (IIDRS) based on reinforcement learning (RL). IIDRS aims to find an ideal plan by interacting with the user, who provides feedback on the gap between the recommended plan and their ideal one. To improve decision-making efficiency and effectiveness in large decoration spaces, we propose a Decoration Recommendation Coarse-to-Fine Policy Network (DecorRCFN). Additionally, to enhance generalization in online scenarios, we propose an object-aware feedback generation method that augments model training with diversified and dynamic textual feedback. Extensive experiments on a real-world dataset demonstrate our method outperforms traditional methods by a large margin in terms of recommendation accuracy. Further user studies demonstrate that our method reaches higher real-world user satisfaction than baseline methods.

CCS CONCEPTS

• Information systems → Recommender systems.

KEYWORDS

Multimodal Interaction, Interactive Recommendation, Reinforcement Learning, Interior Design

*Corresponding authors.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MM '23, October 29–November 3, 2023, Ottawa, ON, Canada.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0108-5/23/10...\$15.00
<https://doi.org/10.1145/3581783.3612420>

ACM Reference Format:

He Zhang, Ying Sun, Weiyu Guo, Yafei Liu, Haonan Lu, Xiaodong Lin, and Hui Xiong. 2023. Interactive Interior Design Recommendation via Coarse-to-fine Multimodal Reinforcement Learning. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23)*, October 29–November 3, 2023, Ottawa, ON, Canada. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3581783.3612420>

1 INTRODUCTION

As interior design becomes increasingly prevalent in our daily life, designing a decoration plan that satisfies the personalized requirements of the homeowner has become a critical challenge. Effective communication can be hindered by the homeowner's lack of professional design knowledge and a clear idea of the desired decoration effect. Consequently, a significant amount of effort is spent repeatedly redesigning and rendering the decoration plan to collect concrete feedback on the owner's preferences, resulting in extremely high labor and time costs.

In recent years, there has been an emerging trend in developing intelligent indoor design facilitation systems. Along these lines, major research efforts are being devoted to generating decoration effects based on textual descriptions [8, 28, 33, 34]. For example, Text2Room [11] generates 3D indoor scenes according to the input text, Text2Scene [26] recurrently generates objects and attributes in the current scene according to the input text. These methods can generate high-quality decoration displays when people have determined their specific requirements and can explicitly describe them. However, there is an undeniable problem: users may not be clear about what they want at the beginning. The problem of how to identify a user's hidden preferences and choose the proper initial design has been largely neglected.

Recommending the interior designs according to the inherent user preferences is a non-trivial task. First, non-expert users typically require multiple rounds of interaction to express their potential desires in terms of their instant likes and dislikes of presented decoration effects [15, 16, 37]. Learning an effective and sustainable interaction strategy for selecting optimal decoration plans from a large space that inspires and guides users to express their

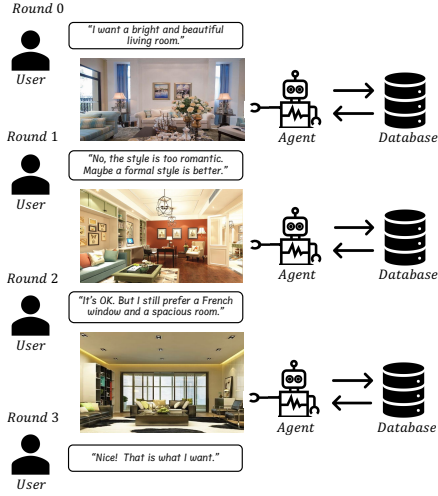


Figure 1: User-agent interaction process in IIDRS.

preferences is a challenging task. The interior decoration design interaction process involves users providing language feedback regarding visual decoration effects. Modeling hidden user preferences from complex multimodal semantic interactions [18, 19, 27, 29] is another challenging task. Third, real-world user feedback exhibits subjectivity, imprecision, uncertainty, and diversity, which can bring difficulty for model training [13, 14].

To address these challenges, we propose an Interactive Interior Design Recommendation System (IIDRS).¹ This system aims to find an ideal plan through limited rounds of interactive recommendation to a user who provides textual feedback on the gap between the recommended plan and the ideal one. Specifically, we propose a novel Decoration Recommendation Coarse-to-Fine Policy Network (DecorRCFN). Observing the multimodal interaction history, DecorRCFN employs Reinforcement Learning (RL) [36] to learn the interactive recommendation strategy that explores user preferences and fulfills their requirements. In particular, a coarse-to-fine two-policy structure is proposed to raise efficiency and effectiveness of decision-making on the large decoration action space. Furthermore, to enhance model generalization in online scenarios that are characterized by uncertain and diversified feedback, we design an object-aware feedback generation method which augments model training with dynamic and diversified textual feedback. Finally, we conduct extensive experiments on a real-world decoration dataset. The experimental results show that DecorRCFN outperforms existing methods by a large margin in terms of recommendation accuracy. Moreover, user studies demonstrate that our method achieves the highest level of user satisfaction in real-world interactions.

2 RELATED WORK

2.1 Intelligent Interior Design Systems

In the field of interior design, one direction is to learn furniture layouts for indoor scenes. For example, SceneFormer [33] uses a self-attention mechanism to learn object relations. Di *et al.* [7] use a multi-agent reinforcement learning-based scene design method to learn the optimal 3D furniture layout. With the development

of deep generative models [8, 20, 22, 23], indoor scene generation has become another important research direction. For example, Text2Room [11] generates 3D indoor scenes based on the input text. Text2Scene [26] recurrently generates objects and attributes in the current scene according to the input text. NeuRIS [31] employs a neural rendering framework to handle texture-less areas and generate high-quality 3D indoor scenes from 2D images. Different from existing works, we focus on identifying a user's hidden preferences and choosing the proper initial design. We propose an interactive recommendation system to solve this problem.

2.2 Interactive Recommendation

Interactive recommender systems aim to extract user preferences through multiple rounds of interactions, and recommend items that match their interests, including attribute-based dialogue recommendation [4, 15, 25] and generative dialogue recommendation [1, 17, 41, 42]. Traditional methods usually interact with users in a textual dialogue manner [17]. However, interactive recommendation in design scenarios depend on the user's intuitive visual experience. Therefore, visually-grounded dialog system researches have also emerged in e-commerce platforms in recent years [2, 10, 39, 40]. For example, Yuan *et al.* [39] proposes a conversational fashion image retrieval method that predicts the desired image based on text and image information in the conversation history. Different existing works, we propose a method for decoration recommendation, which contains more complex attributes and involve more diversified multi-modal interaction than single items.

3 PRELIMINARIES

3.1 Problem Formulation

The overall procedure for the interactive interior design recommendation process is shown in Figure 1. The user interacts with the agent for multiple rounds. At the t -th round, the user provides textual requirements or feedback f_t on the recommended decoration. The agent then recommends a decoration a_t from the database D . The user determines the gap between the recommended an interior design case a_t and the ideal case. Accordingly, they provide feedback f_{t+1} and an overall satisfaction score r_t on a five-level scale: *strongly dislike*, *dislike*, *neutral*, *like*, and *strongly like*. We assume that there exists an ideal decoration T that satisfies the user. The interaction ends either when the ideal decoration is recommended or when the maximum number of rounds t_{max} is reached. Intuitively, even if the user is initially unclear about the ideal decoration, their feedback and satisfaction can be regarded as generated from the differences between the recommended decoration and their ideal decoration, which guide the exploration for the ideal decoration. Our objective is to train an agent with an effective strategy to recommend a design case a_t , using historical interactions $\langle f_0, a_0, r_0, f_1, a_1, r_1, \dots, f_t \rangle$, and to reach the ideal case T within the fewest number of rounds.

3.2 Data Description

The interior design data used in this paper were gathered from a popular Chinese interior design website². The data comprises

¹The dataset and source codes are available at <https://github.com/ZhHe11/IIDRS>.

²<https://www.jiajuol.com>



整个餐厅开放式的设计，弱化了电视为中心的布局，以舒适惬意为主调，营造自在的生活氛围；整个空间以胡桃木地板铺贴出质朴，并陈列简约的木质家具，让空间自然而温润。
(The open design of entire dining room weakens the TV-centered layout and focuses on comfort to create a comfortable living atmosphere. The entire space is paved with walnut wood floors, and the simple wood-colored furniture is displayed. The space is natural and warm.)

三居室(Three bedrooms) 简约(Minimalist style) 90平米(90 square meters)

Figure 2: An interior design case from our dataset.

90,000 decoration cases covering 20 different decoration styles and 10 room types. Each instance consists of images, descriptive texts, as well as tags such as house types and room styles. The text consists of lengthy paragraphs that include multiple sentences of subjective descriptions. These descriptions are provided by expert designers and focus on the decoration of a scene, including both overall descriptions of the entire scene, such as atmosphere and color, as well as specific views of individual objects, such as French windows and pendant lights. We create an object/attribute dictionary with 300 object types to identify different aspects of these descriptions. An illustrative example is shown in Figure 2, where the red words indicate objects/attributes. Formally, we formulate each interior design instance as $(d^j, l^j, \{o_i^j\}_{i=1}^{M_j})$, where d^j denotes the decoration image, l^j denotes the description, $\{o_i^j\}_{i=1}^{M_j}$ refers to objects set appearing in the j -th instance's description, M_j refers to the number of such objects. Notably, since textual descriptions may not always be available for online interior design platforms, we develop our system for the generic case where the database $D = \{d_j\}_{j=1}^N$ only contains images, where N denotes the total number of images in the database. The textual descriptions and objects are optional to develop a synthetic environment for training augmentation and model evaluation, which are detailed in Section 4.3 and Section 5.1.2, respectively.

4 METHOD

4.1 RL Formulation

We define the interactive interior design recommendation process as a Markov Decision Process (MDP) with the formulation $(S, \mathcal{A}, \mathcal{R})$, where S , \mathcal{A} , and \mathcal{R} denote the state, action, and reward space, respectively. At round t , the state s_t is formulated as

$$s_t = \langle f_0, a_0, f_1, a_1, \dots, a_{t-1}, f_t \rangle. \quad (1)$$

In particular, $s_0 = f_0$ is the first request from the user. We use $F_t = \{f_i\}_{i=0}^t$ to represent the historical feedback from the user and $A_t = \{a_i\}_{i=0}^{t-1}$ to represent the historical recommended images selected by the agent. The action $a_t \in \mathcal{A}$ (i.e., the recommended image) is selected from action space D (i.e., the entire instance database) with a learned policy π_θ by a network parameterized θ .

To balance the exploration-exploitation trade-off during the training process, we involve a parameter ϵ valued from $[0, 1]$. When $\epsilon = 1$, the agent exploits the known optimal strategy without any exploration. When $\epsilon = 0$, the agent explores the environment fully randomly, without considering any known information. We set the reward $r_t \in \mathcal{R}$ according to the five levels of user satisfaction score, labeled as $\{-2, -1, 0, 1, 2\}$ respectively. Moreover, if the agent successfully recommends the user's ideal instance at round t , the reward r_t is assigned a value of 10 and the process is terminated. The long-term return is formulated as discounted cumulative reward:

$$v_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots = \sum_{k=0}^{t_{max}-t} \gamma^k r_{t+k}, \quad (2)$$

where γ represents the discount factor. The objective of the agent is to learn an optimal policy π_θ through interacting with the environment by maximizing v_t . Notably, different from existing works [10], our supervision signal does not come from the rank of a specific target item since the ideal instance is defined to be implicit in our problem formulation, which is unknown unless the agent succeeds in finding it in t_{max} rounds.

4.2 Decoration Recommendation Coarse-to-Fine Policy Network

In the decoration recommendation problem, the action space is large due to the many possible decorations to recommend. Learning a fine-grained policy on all possible choices with a complex model can lead to low exploration-efficiency and make model convergence difficult. Inspired by the classic preranking-ranking 2-step recommendation schema, we propose DecorRCFN, a coarse-to-fine policy network consisting of two sub-policy modules. Its architecture is shown in Figure 3. Specifically, the coarse-grained policy network learns the policy π_θ^c for selecting a set of candidate actions from the entire image set D , based on the similarity of user feedback and decoration representations. The fine-grained policy network then models multimodal state-action interaction and learns the policy π_θ^f for selecting the final action from the candidate images.

4.2.1 Coarse-grained Policy Network. Given the large action space, it is necessary to reduce the number of candidate interior design cases from the entire database D that need to be considered by a complex, fine-grained policy. Intuitively, the historical feedback provided by users can be seen as incomplete descriptions of desired decoration. Therefore, a straightforward idea is to extract a representation of the user's demands from their historical feedback and match it with representations of decoration images. We employ CLIP [21, 38], a prominent multimodal pre-trained model for image-text matching, as the encoder module for generating embeddings $F_t \in \mathbb{R}^{t \times h}$ and $D_t \in \mathbb{R}^{N \times h}$ for textual feedback F_t and decoration images in D :

$$\begin{aligned} F_t &= \{\text{Encoder}(f_i) | f_i \in F_t\}, \\ D &= \{\text{Encoder}(d_j) | d_j \in D\}, \end{aligned} \quad (3)$$

where h represents the embedding dimension. Then, we calculate a similarity matrix $M \in \mathbb{R}^{t \times N}$, the element $M_{i,j}$ represents the similarity between historical feedback $f_i \in F_t$ and $d_j \in D$ can be computed as:

$$M_{i,j} = \text{sim}(\text{Encoder}(f_i), \text{Encoder}(d_j)) \quad (4)$$

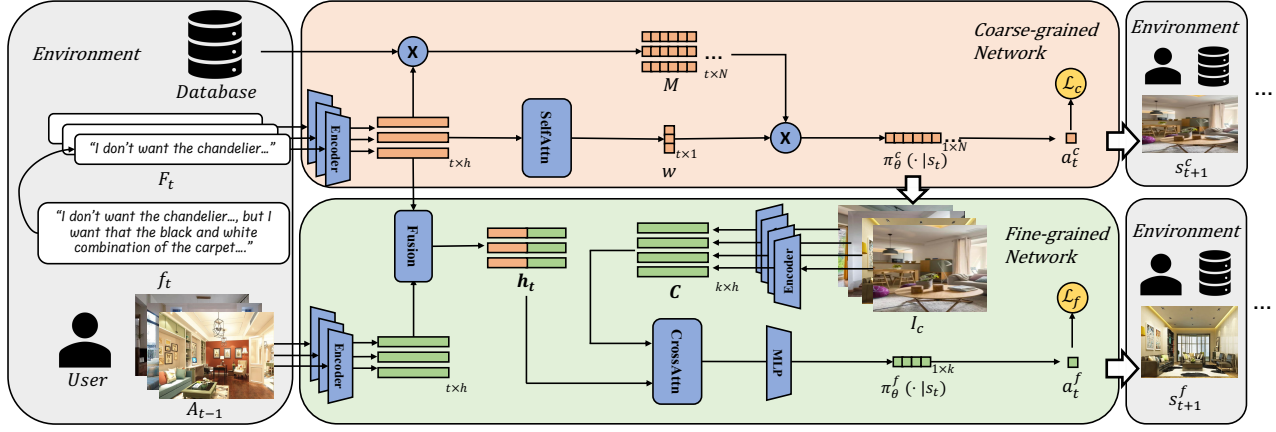


Figure 3: Architecture of DecorRCFN.

where $\text{sim}(\cdot, \cdot)$ computes the cosine similarity of two embeddings, outputting a value ranging from -1 to 1 . Based on the similarity matrix, a trivial candidate selection method is to add up the scores of all the feedback and retrieve the images with the highest scores. However, this approach may be suboptimal because different feedback contains various information and may have differing levels of importance in reflecting user preferences. Therefore, we consider candidate selection as a learnable policy, with the action space defined as the entire interior design database, i.e., $\mathcal{A}_c = D$. To model the importance of each $f_i \in F_t$, we use a multi-head self-attention module [30], formulated as

$$\text{SelfAttn}(X) = \text{softmax}\left(\frac{(XW_Q)(XW_K)^T}{\sqrt{d_k}}\right)(XW_V) \quad (5)$$

where X represents the input for the self-attention layer, W_Q , W_K , and W_V are trainable parameters, d_k is the key dimension in the attention mechanism. Based on the self-attention module, we get the attention vector $w \in \mathbb{R}^t$ as $w = \text{MLP}(\text{SelfAttn}(F_t))$, where MLP is a Multi-Layer Perceptron, and accordingly calculate the selection probability as

$$\pi_\theta^c(\cdot | s_t) = \text{softmax}(w^T M). \quad (6)$$

Since direct supervision of the proper candidate is unavailable, we train the coarse-grained policy with a reinforcement learning schema. Specifically, we employ coarse-grained policy to draw the recommendation sequence and adopt Policy Gradient (PG) [36] to train the model. For each step, among all the images in D , the one with the largest score is selected as the action, formulated as $a_t^c = \arg \max_{a^c \in \mathcal{A}_c} \pi_\theta^c(a^c | s_t^c)$. The agent aims to maximize the object function J_c , formulated as:

$$J_c = \sum_{a^c \in \mathcal{A}_c} \pi_\theta^c(a^c | s^c) \cdot v, \quad (7)$$

$$\nabla_{\theta} J_c \approx \sum_t \nabla_{\theta} \log(\pi_\theta^c(a_t^c | s_t^c)) \cdot v_t,$$

where $\pi_c(a^c | s)$ denotes the coarse-grained policy. The optimization problem can be reformulated as the minimization of the loss function \mathcal{L}_c , which is given by:

$$\mathcal{L}_c = - \sum_{t=1}^{t_{\max}} \log(\pi_\theta^c(a_t^c | s_t^c)) \cdot v_t. \quad (8)$$

Notably, the similarity calculation requires encoding images for all the images in the database everytime the parameters are updated. To reduce model complexity, we freeze the image embeddings in CLIP module so that we can pre-calculate them for quick retrieval.

4.2.2 Fine-grained Policy Network. The coarse-grained network usually prioritizes the textual modality but ignores how it interacts with the image modality in recommendation history. Indeed, since each textual feedback specifies user opinion for specific visual characteristics of the given image, looking into the informative visual signals based on the feedback can facilitate user preference modeling. Therefore, instead of directly choosing the final action with coarse-grained policy, we select the top- k decorations as candidate set I_c . Then, we train the fine-grained policy network in DecorRCFN to further determine the best choice from the candidate set. Specifically, the fine-grained policy network models the semantic interaction between A_{t-1} , F_t , and each candidate in I_c . We obtain embeddings A_{t-1} of the action history A_{t-1} and C of the candidate images I_c through the CLIP encoder module:

$$A_{t-1} = \{\text{Encoder}(a_i) | a_i \in A_{t-1}\}, \quad (9)$$

$$C = \{\text{Encoder}(c_i) | c_i \in I_c\}.$$

To capture user preferences from multiple modalities in multiple rounds, we fuse the embeddings of feedback sentences F_t and the action history A_{t-1} to form a new embedding vector h_t . In order to select the action of the fine-grained network that best matches the user preferences, we use a multi-head cross-attention module [9, 35] to compare the h_t and C . The cross-attention module is defined as:

$$\text{CrossAttn}(X_Q, X) = \text{softmax}\left(\frac{(X_Q W_Q)(X W_K)^T}{\sqrt{d_k}}\right)(X W_V), \quad (10)$$

where X_Q and X are the query and key-value matrices, respectively. We utilize h_t as the query matrix and C as the key-value matrix to obtain the attention matrix, which is then fed into a Multi-Layer Perceptron (MLP) with Softmax activation to obtain the probability of each actions:

$$\pi_\theta^f(\cdot | s_t^f) = \text{softmax}(\text{MLP}(\text{CrossAttn}(C, h_t))). \quad (11)$$

Therefore, the optimal action of the fine-grained network is represented as $a_t^f = \arg \max_{a^f \in \mathcal{A}_f} \pi_\theta^f(a^f | s_t^f)$, where $\mathcal{A}_f = I_c$ denotes

the action space (i.e., candidate pool) of the network. Similar to the coarse-grained policy, Based on PG algorithm, we define the object function and loss function of the fine-grained network as follows:

$$\mathcal{L}_f = - \sum_t \log(\pi_\theta^f(a_t^f | s_t^f)) \cdot v_t. \quad (12)$$

We combine the coarse-grained and fine-grained loss to formulate the overall Coarse-to-fine Network loss, based on sampled recommendation paths from the final policy $\pi_\theta = (\pi_\theta^c, \pi_\theta^f)$, as

$$\mathcal{L}_{cf} = - \sum_{t=1}^{t_{max}} (\alpha \cdot \log(\pi_\theta^c(a_t^c | s_t^c)) + \log(\pi_\theta^f(a_t^f | s_t^f))) \cdot v_t, \quad (13)$$

where $\alpha \in [0, 1]$ balances the importance of the two policies.

4.3 Training Augmentation with Object-aware Feedback Generation

Real-world user feedback exhibits subjectivity, imprecision, uncertainty, and diversity nature. The model needs abundant explorations before it learns effective strategies to dig the preference from the complex and uncertain user feedbacks whose patterns vary with user's personal habits. However, collecting user feedback through real interaction can be expensive. The limited data can harm model performance for easily causing overfitting problem. Therefore, data augmentation is important to enhance model generalization in complex online scenarios. To enhance model generalization in complex online scenarios, we propose augmenting our training procedure with diverse simulated feedback dynamically generated based on the differences between the recommended decoration and the target decoration. As a decoration design can be described from various aspects [3, 12, 32], our method generates difference descriptions in an objects/attributes-aware manner, containing three steps: object detection, object-aware image captioning, and difference description filtering. The overall framework is shown in Figure 4.

4.3.1 Object Detection. Our dataset includes selective descriptions of objects and attributes that appear in an image. We first train a model to identify objects or attributes that can be found in the image. Specifically, we treat the object as text and fine-tune a CLIP model with image-object pairs. This allows the model to calculate the similarity between an image d and an object o . A higher similarity score indicates a higher chance of the object existing in the image.

4.3.2 Object-aware Image Captioning. After recognizing objects in an image, we train an object-aware image captioning model to generate captions describing these objects in their corresponding images. To be specific, based on our decoration dataset, we build a training set in which each sample formulated as $\langle (d, o, l_m), y \rangle$, where d represent the image, o represents an object, y denotes the target sentence that describes object o within image d , l_m is the rest descriptions after masking the target sentence. To model these data, our model embeds the image d with a CNN backbone, and embed the text o_i and l_m with a BERT [5] encoder. The embeddings are represented as \mathbf{d} , \mathbf{o} , and \mathbf{l}_m . Then, we utilize a transformer module, which has proven to be superior in image captioning tasks [24, 30], to generate textual descriptions. However, directly feeding all types of embeddings into the transformer can lead to high computation costs due to excessive length of sequences. Therefore, we involve

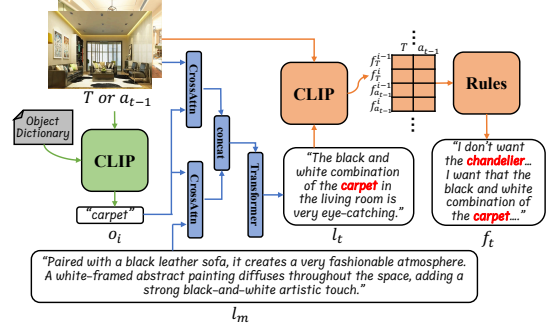


Figure 4: Framework of Object-aware Feedback Generation.

two cross-attention modules (as Eq. 10) to reduce the sequence length. The two cross-attention modules both use \mathbf{o} as Q and \mathbf{d} and \mathbf{l}_m as K and V , respectively. The features contain information of the object within the images and descriptions. Finally, we concatenate the features and feed them into the transformer module to produce an object-aware image caption \hat{y} . Following existing image captioning works, we train the model by minimizing cross-entropy loss between target description y and the generated sentence \hat{y} .

4.3.3 Difference Description Filtering. So far, a variety of sentences can be generated to describe different aspects of a decoration image. To filter the sentences that describe differences between the current image and target image, we developed a filter module with a CLIP model fine-tuned on image-sentence pairs in our dataset. Using the filter, we identify sentences that have high similarity to one image and low similarity to the other as describing their differences. Ultimately, we synthesize user behavior based on the rules introduced in Section 5.1.2. This enriches the diversity of feedback and improves the recommendation performance.

5 EXPERIMENT

5.1 Experimental Setup

5.1.1 Datasets and Settings. To ensure both the effectiveness and efficiency of reinforcement learning training in a large action space, we further selectively screen detailed and clear interior design cases from the data. Specifically, we select 5731 samples from the screened data to be used as the experimental dataset for subsequent analysis. In order to more accurately evaluate the performance and generalization of the model, we split the dataset into training and testing sets in a 1:1 ratio, which ensures that the size of the action space is the same. The CLIP models have been fine-tune with our training set. To demonstrate the effectiveness of Object-aware Feedback Generation, we then create two datasets based on the source of feedback: Real-world Sentences(RS), which consists of synthesized feedback derived from real-world descriptions, and Generation Sentences(GS), which utilizes feedback generated through the Object-aware Feedback Generation model. We select 10 objects with the highest similarity to each image. The parameters of DecorRCFN are empirically set as: $\epsilon = 0.8$ (Section 4.1), $\gamma = 0.8$ (Equation 2), $\alpha = 0.1$ (Equation 13), the number of candidates $k = 4$, and the learning rate $lr = 10^{-5}$.

5.1.2 User Simulation. For multi-round interactive recommendation systems, the environments needs real users to converse with

Model	Feedback from Generated Sentences(for test)						Feedback from Real-world Sentences(for test)					
	Max round = 10		Max round = 8		Max round = 6		Max round = 10		Max round = 8		Max round = 6	
	r@1	r@2	r@1	r@2	r@1	r@2	r@1	r@2	r@1	r@2	r@1	r@2
Random	0.43%	0.53%	0.35%	0.40%	0.31%	0.42%	0.60%	0.71%	0.35%	0.42%	0.31%	0.39%
CLIP-greedy	42.23%	47.05%	33.15%	38.86%	21.84%	27.30%	28.12%	32.78%	25.46%	28.64%	21.70%	24.43%
CLIP-random	44.11%	48.97%	32.80%	38.97%	21.27%	28.82%	32.41%	33.86%	27.80%	30.03%	22.62%	25.67%
DecorRCGN	48.12%	54.46%	38.04%	44.14%	21.66%	27.16%	32.26%	33.79%	28.33%	30.98%	21.66%	24.44%
DecorRCFN	49.82%	55.56%	38.23%	44.64%	24.53%	30.88%	33.80%	35.63%	29.40%	31.76%	23.57%	26.27%

Table 1: Performance Evaluation for IIDRS.

the agent and provide rewards, which is challenging to create for research purposes [6]. A common solution is using simulated users to interact with the agent. In IIDRS, the user simulator simulates the ideal case by randomly drawing a decoration image from the database. Accordingly, it simulates the following three behaviors: 1) providing feedback f_t according to the given image a_{t-1} and the ideal image T ; 2) providing satisfaction level (Section 4.1); 3) terminating the interaction when the target case T is recommended or the number of rounds exceeds t_{max} . To simulate feedback, we create rules that use a set of description sentences for the current and target images to construct candidate feedback. We evaluate the feedback using the CLIP model to determine whether a given sentence accurately describes the differences between the two images. An effective feedback should exhibit a significant contrast between the similarity score to T and the a_t . If a sentence f has a high similarity score to a_t but a low similarity score to T , we use the template "I don't like $\{f\}$ " to generate a sentence describing the user's dislike of the current image. Conversely, if f has a high similarity score to T but a low similarity score to a_t , we use the template "I prefer $\{f\}$ " to generate a sentence describing the user's demand for the ideal decoration. During the interaction, feedback sentences are randomly sampled from the constructed candidates based on the current and target image. For satisfaction level simulation, we design a rule based on the rank $\text{rank}(a_t, T, D)$ of a_t in terms of its embeddings' cosine similarity to T among images in D . Specifically, the satisfaction level r_t is defined as

$$r_t = \begin{cases} 2, & \text{rank}(a_t, T, D) \leq l_0 \\ 1, & l_0 < \text{rank}(a_t, T, D) \leq l_1 \\ 0, & l_1 < \text{rank}(a_t, T, D) \leq l_2 \\ -1, & l_2 < \text{rank}(a_t, T, D) \leq l_3 \\ -2, & \text{rank}(a_t, T, D) > l_3, \end{cases} \quad (14)$$

where $l_{\{0,1,2,3\}}$ denotes the thresholds for different satisfaction levels. We set $l_{\{0,1,2,3\}} = \{10, 20, 30, 50\}$ in our experiment.

5.2 Recommendation Performance Evaluation

5.2.1 Baseline Performance Analysis. We primarily evaluate the effectiveness of our recommendation model using recall@1($r@1$) and recall@2($r@2$). As online reinforcement learning for multi-round text-image recommendation represents a new task, we compared our approach with two baseline methods: Random search(Random) and CLIP models(CLIP). Specifically, for Random search, we randomly select images from the database for *Maxround* rounds. For CLIP-greedy model, we select the image that receive the highest

similarity score with all feedback during an interaction, calculated by a fine-tuned CLIP model. However, the greedy strategy is not always optimal, as user preferences are uncertain during the whole recommendation process. Therefore, we introduce a CLIP-random baseline, which is based on CLIP but randomly selects the next image from the top 4 candidates calculated by CLIP.

Table 1 presents our experimental results. Random search performed poorly, failing to hit the target image. On the other hand, the fine-tuned CLIP model achieved relatively good performance. We observed that the score of $r@2$ using CLIP-greedy outperformed $r@1$ by almost 5%, indicating that the image with the highest similarity score is not always the optimal choice. It also reflects that the user feedback is information limited, leading to the most similar image not necessarily being the most appropriate choice. Furthermore, the random strategy appears to be less stable and uncontrollable in performance compared to Greedy search and other models. Specifically, the variance of the results obtained using CLIP-random is higher than that of CLIP-greedy and other methods.

5.2.2 DecorRCGN and DecorRCFN. In Table 1, the Decoration Recommendation Coarse-Grained policy Network (DecorRCGN) and the Decoration Recommendation Coarse-to-Fine policy Network (DecorRCFN) exhibit superior performance compared to the baselines in general. DecorRCGN ranks the actions in database by recalculating the weights of different feedback sentences(Section 4.2.1). To illustrate its effect, we conduct an experiment as shown in Table 2. While all feedback sentences carry equal weights in CLIP-based models, DecorRCGN assigns different weights to each sentence based on its content. For instance, the first feedback 'White' being the shortest and providing the least information, receives the lowest weight of 0.5557. In contrast, the third line sentence, 'A white bedroom with enough space for kids to play', containing the most information, obtains the highest weight of 0.9711. After assigning weights to each feedback sentence based on their importance, the scores of similarity between each image and sentence are summed up. When selecting the next action in DecorRCGN, it is found that the third sentence, which contains the most information, has the highest weight and thus contributes the most to the next image. As shown in Table 1, DecorRCGN outperforms both CLIP-greedy and CLIP-random by 5.8% and 4.0%, respectively, in terms of $r@1$ when *Maxround* = 10. Nonetheless, DecorRCGN has two drawbacks. Firstly, it still employ the greedy strategy when selecting the next action, ignoring the fact that the image with the highest score does not necessarily represent the best choice. Secondly, when the

Feedback in English	CLIP	DecorRCGN
White	1	0.5557
A white bedroom	1	0.8720
Spacious white bedroom for kids to play.	1	0.9711

Table 2: Weights of different feedback sentences.

Maxround is set to 6, DecorRCGN fails to surpass CLIP-greedy due to insufficient feedback information. However, our proposed DecorRCFN effectively addresses these issues. The DecorRCFN model employs a more refined policy by selecting the top k candidates filtered by DecorRCGN. This helps to avoid always selecting images with the highest scores, which may not necessarily be the optimal choice. Furthermore, DecorRCFN enhances candidate comparison by utilizing information from both feedback history and images recommended in the previous round. This results in improved performance of DecorRCFN, especially achieving 24.53% in terms of $r@1$ when textual modality is limited when *Maxround* = 6, shown in Table 1. In both the test set with generated sentences(GS) and the test set with real-world sentences(RS), DecorRCFN consistently outperforms DecorRCGN and achieves the best performance.

5.2.3 Ablation Study. We conduct an ablation study to illustrate the effectiveness of the RL compared to supervised learning. In this problem formulation, the agent is unable to acquire additional information beyond feedback and rewards. Therefore, the loss of supervised learning can only incorporate rewards. In Table 3, \mathcal{L}_c and \mathcal{L}_{cf} represent the losses of the DecorRCGN and DecorRCFN, respectively. The *reward* indicates that the loss is calculated solely based on r_t , which can be regarded as reward supervised learning. The *value* means the loss is calculated according to v_t Equation 2, where $\gamma = 0.8$ to get the best results. We compare the recall@1 performance of these settings with *Maxround* = 10. And the results indicate that models based on reward supervised learning achieve much lower values of $r@1$ compared to RL models. This is because when $\gamma = 0$, the models never consider future rewards but only focus on the current choices, leading to performance similar to CLIP-greedy. Once we incorporate future rewards into the loss, the agent could choose a better action based on the entire episode.

We conduct another ablation study on the influence of object-aware feedback generation. Figure 5 displays the correlation between the number of rounds and the value of recall@1 when DecorRCFN is trained on different dataset and test on different dataset. Specifically, GS denotes the dataset using generated sentences from object-aware feedback generation, and RS denotes the dataset with real-world sentences. All of the curves show an upward trend. Specifically, the red and purple curves, which represent DecorRCFN tested on GS, exhibit increasingly steeper slopes as the number of rounds increases. This observation suggests that DecorRCFN’s performance improves with additional feedback and image information from multiple rounds. However, the models tested on RS(the yellow and blue curves) show slower growth. This is because the information from real-world sentences doesn’t increase necessarily as the rounds increasing, due to the limited number of sentences in real-world descriptions. Furthermore, the figure indicates that the models trained on GS(represented by the red and yellow curves) perform better than those trained on RS(represented by the purple and blue curves) in both test sets. This highlights the effectiveness

\mathcal{L}_c (reward)	\mathcal{L}_c (value)	\mathcal{L}_{cf} (reward)	\mathcal{L}_{cf} (value)	Max round=10 recall@1
✓				42.43%
	✓			48.26%
		✓		42.72%
			✓	49.82%

Table 3: Comparison between ours and reward supervised learning.

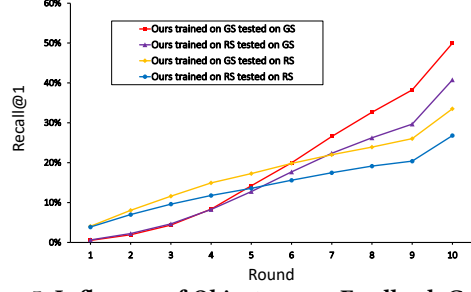


Figure 5: Influence of Object-aware Feedback Generation.

of our object-aware feedback generation model (see Section 4.3), which enriches the diversity of user feedback, augments the data in the training set, and ultimately improves the model performance.

5.3 Object Description Model Evaluation

We assess the object-aware feedback generation model using widely adopted metrics such as BLEU-1 and ROUGE-L [12], to evaluate the fluency of the generated sentences. Additionally, to demonstrate that the model could produce high-quality sentences, we compare the similarity scores, CLIP-S, of the images and generated sentences. Our generation model consists of the two cross-attention modules, called Double CrossAttn in Table 4. We compare it with the surely transformer module without cross-attention modules previously, shown in the first row of the table. Without cross-attention modules before, it joints the embeddings of d , o and l_m together, leading to an increased sequence length. Consequently, Double CrossAttn achieves comparable results to Transformer, while also requiring less time for sentence generation due to shorter sequence length. Specifically, Double CrossAttn has an average generation time of 95.5ms per sentence, compared to Transformer’s average of 218.8ms. To demonstrate the significance of the information from the masked descriptions l_m , we train a Single CrossAttn module without integrating the modality of l_m . The experimental results validate that the information from l_m cannot be disregarded. The BLEU-1 and ROUGE-L scores of Single CrossAttn are much lower than those of Double CrossAttn, indicating that the sentences generated by Single CrossAttn are less fluent as than those generated by Double CrossAttn. Meanwhile, the CLIP-S of Double CrossAttn surpasses the score of Single CrossAttn by its 22%. This implies that the sentences generated by Single CrossAttn are unable to pass the filter. We additionally involve a temperature parameter(T), which increases the probability to generate different words. Finally, the performance of Double CrossAttn(T) surpasses Double CrossAttn in terms of BLEU-1 and ROUGE-L scores, though it achieves slightly lower scores for the CLIP-S.



Figure 6: Real-world examples of interactions with our IIDRS.

Module	BLEU-1	ROUGE-L	CLIP-S
Transformer	0.4396	0.2423	27.76
Single CrossAttn	0.2721	0.0905	22.75
Double CrossAttn	0.4317	0.2506	27.85
Double CrossAttn(T)	0.4405	0.2544	27.67

Table 4: Performance Evaluation for Object Description.

5.4 User Study

To demonstrate the effectiveness of our system in interacting with humans, we recruited 20 individuals with diverse backgrounds to test our system. Each user communicated with three models: DecorRCGN, DecorRCFN, and CLIP-greedy, without being informed which model they were interacting with during the entire process. We recorded each user’s level of satisfaction after each round of interaction. Once a user was satisfied with the recommended image and decided to stop further recommendations, the model received a reward of 3 in both the current and subsequent rounds.

5.4.1 Case Study. An example interaction with our IIDRS is shown in Figure 6. The user cannot exactly describe their whole preferences at the beginning. For instance, they just ask for a room but don’t require the piano at first round. As the number of rounds increases, the interior decoration cases provided by the agent guide the user to offer more specific feedback, which is also the reason why multi-round interactive recommendation systems outperform single-round ones. In Figure 6, our agent effectively satisfies both current and historical user needs.

5.4.2 Experiments Evaluation. We compare DecorRCFN and DecorRCGN by evaluating the differences in user satisfaction between them and CLIP-greedy, calculated by subtracting the user satisfaction levels of CLIP-greedy from those of DecorRCFN and DecorRCGN. Figure 7 displays the results of the experiments. At first round, these model exhibit similar performance due to no significant difference in their policies. As the number of rounds increases, the difference in user satisfaction levels with the two models compared to CLIP-greedy increases in general, which indicates that DecorRCFN and DecorRCGN can better extract user preferences from history interaction. Moreover, the performance of DecorRCFN outperforms DecorRCGN by an increasingly large margin. The users explain that DecorRCGN tend to recommend the similar images as the previous round, which affects user satisfaction and gets lower rewards than DecorRCFN. They also comment that DecorRCGN

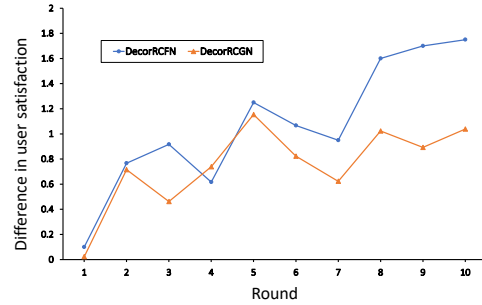


Figure 7: Differences in user satisfaction levels with different models compared to CLIP-greedy.

has the best memory of user feedback, which could be explained by a fine-grained comparison with the history feedback and images.

6 CONCLUSION

In this paper, we proposed an Interactive Interior Design Recommendation System (IIDRS) based on reinforcement learning to mine user preferences in personalized interior design recommendation. We also proposed a Decoration Recommendation Coarse-to-Fine Policy Network (DecorRCFN) to enhance decision-making the agent’s ability in large action spaces. Finally, we introduced an object-aware feedback generation model to augment the training process and optimize policy network training. Our experiments demonstrated the superiority of our method in terms of recommendation accuracy and real-world user satisfaction. We anticipate this work will have a positive impact on personalized interior decoration design by introducing an interactive approach to recommendation. We also acknowledge that current study may not meet users’ requirements when their desired interior design cases do not exist in our dataset. In future work, we will explore developing generative decoration recommendation model that can create more personalized and feasible interior design cases based on user preference.

ACKNOWLEDGMENTS

This work was supported by the City-University Joint Funding Project of Guangzhou Science and Technology Plan (No. 2023A03J0141), Foshan HKUST Projects (FSUST21-FYTRI01A, FSUST21-FYTRI02A) and OPPO Research Fund.

REFERENCES

- [1] Qibin Chen, Junyang Lin, Yichang Zhang, Ming Ding, Yukuo Cen, Hongxia Yang, and Jie Tang. 2019. Towards knowledge-based recommender dialog system. *arXiv preprint arXiv:1908.05391* (2019).
- [2] Ginger Delmas, Rafael Sampaio de Rezende, Gabriela Csurka, and Diane Larlus. 2022. ARTEMIS: Attention-based Retrieval with Text-Explicit Matching and Implicit Similarity. *arXiv preprint arXiv:2203.08101* (2022).
- [3] Jiajun Deng, Zhengyuan Yang, Tianlang Chen, Wengang Zhou, and Houqiang Li. 2021. Transvg: End-to-end visual grounding with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1769–1779.
- [4] Yang Deng, Yaliang Li, Fei Sun, Bolin Ding, and Wai Lam. 2021. Unified conversational recommendation policy learning via graph-based reinforcement learning. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1431–1441.
- [5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [6] Bhuwan Dhingra, Lihong Li, Xiujun Li, Jianfeng Gao, Yun-Nung Chen, Faisal Ahmed, and Li Deng. 2016. End-to-End Reinforcement Learning of Dialogue Agents for Information Access. *CoRR* abs/1609.00777 (2016). [arXiv:1609.00777](https://arxiv.org/abs/1609.00777)
- [7] Xinhan Di and Pengqian Yu. 2021. Multi-Agent Reinforcement Learning of 3D Furniture Layout Simulation in Indoor Graphics Scenes. *arXiv preprint arXiv:2102.09137* (2021).
- [8] Shuyang Gu, Dong Chen, Jianmin Bao, Fang Wen, Bo Zhang, Dongdong Chen, Lu Yuan, and Baining Guo. 2022. Vector quantized diffusion model for text-to-image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10696–10706.
- [9] Weiye Guo, Zhaoshuo Li, Yongkui Yang, Zheng Wang, Russell H Taylor, Mathias Unberath, Alan Yuille, and Yingwei Li. 2022. Context-Enhanced Stereo Transformer. In *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII*. Springer, 263–279.
- [10] Xiaoxiao Guo, Hui Wu, Yu Cheng, Steven Rennie, Gerald Tesauro, and Rogério Feris. 2018. Dialog-based interactive image retrieval. *Advances in neural information processing systems* 31 (2018).
- [11] Lukas Höllein, Ang Cao, Andrew Owens, Justin Johnson, and Matthias Nießner. 2023. Text2Room: Extracting Textured 3D Meshes from 2D Text-to-Image Models. *arXiv preprint arXiv:2303.11989* (2023).
- [12] MD Zakir Hossain, Ferdous Sohel, Mohd Fairuz Shiratuddin, and Hamid Laga. 2019. A comprehensive survey of deep learning for image captioning. *ACM Computing Surveys (CSUR)* 51, 6 (2019), 1–36.
- [13] Gawesh Jawaheer, Peter Weller, and Patty Kostkova. 2014. Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 4, 2 (2014), 1–26.
- [14] Denis Kotkov, Joseph A Konstan, Qian Zhao, and Jari Veijalainen. 2018. Investigating serendipity in recommender systems based on real user feedback. In *Proceedings of the 33rd annual acm symposium on applied computing*. 1341–1350.
- [15] Wenqiang Lei, Xiangnan He, Yisong Miao, Qingyun Wu, Richang Hong, Min-Yen Kan, and Tat-Seng Chua. 2020. Estimation-action-reflection: Towards deep interaction between conversational and recommender systems. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 304–312.
- [16] Wenqiang Lei, Gangyi Zhang, Xiangnan He, Yisong Miao, Xiang Wang, Liang Chen, and Tat-Seng Chua. 2020. Interactive path reasoning on graph for conversational recommendation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 2073–2083.
- [17] Raymond Li, Samira Ebrahimi Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018. Towards deep conversational recommendations. *Advances in neural information processing systems* 31 (2018).
- [18] Lizi Liao, Xiangnan He, Bo Zhao, Chong-Wah Ngo, and Tat-Seng Chua. 2018. Interpretable multimodal retrieval for fashion products. In *Proceedings of the 26th ACM international conference on Multimedia*. 1571–1579.
- [19] Mingsheng Long, Yue Cao, Jianmin Wang, and Philip S Yu. 2016. Composite correlation quantization for efficient multimodal retrieval. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 579–588.
- [20] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (2021), 99–106.
- [21] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*. PMLR, 8748–8763.
- [22] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-shot text-to-image generation. In *International Conference on Machine Learning*. PMLR, 8821–8831.
- [23] Shikhar Sharma, Dendi Suhubdy, Vincent Michalski, Samira Ebrahimi Kahou, and Yoshua Bengio. 2018. Chatpainter: Improving text to image generation using dialogue. *arXiv preprint arXiv:1802.08216* (2018).
- [24] Matteo Stefanini, Marcella Cornia, Lorenzo Baraldi, Silvia Cascianelli, Giuseppe Fiameni, and Rita Cucchiara. 2022. From show to tell: a survey on deep learning-based image captioning. *IEEE transactions on pattern analysis and machine intelligence* 45, 1 (2022), 539–559.
- [25] Yueming Sun and Yi Zhang. 2018. Conversational recommender system. In *The 41st international acm sigir conference on research & development in information retrieval*. 235–244.
- [26] Fuwen Tan, Song Feng, and Vicente Ordóñez. 2019. Text2scene: Generating compositional scenes from textual descriptions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6710–6719.
- [27] Hao Tan and Mohit Bansal. 2019. LXMERT: Learning Cross-Modality Encoder Representations from Transformers. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*.
- [28] Jiapeng Tang, Yinyu Nie, Lev Markhasin, Angela Dai, Justus Thies, and Matthias Nießner. 2023. DiffuScene: Scene Graph Denoising Diffusion Probabilistic Model for Generative Indoor Scene Synthesis. *arXiv preprint arXiv:2303.14207* (2023).
- [29] Matthew Turk. 2014. Multimodal interaction: A review. *Pattern recognition letters* 36 (2014), 189–195.
- [30] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
- [31] Jiepeng Wang, Peng Wang, Xiaoxiao Long, Christian Theobalt, Taku Komura, Lingjie Liu, and Wenping Wang. 2022. Neuris: Neural reconstruction of indoor scenes using normal priors. In *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII*. Springer, 139–155.
- [32] Peng Wang, An Yang, Rui Men, Junyang Lin, Shuai Bai, Zhikang Li, Jianxin Ma, Chang Zhou, Jingren Zhou, and Hongxia Yang. 2022. Ofa: Unifying architectures, tasks, and modalities through a simple sequence-to-sequence learning framework. In *International Conference on Machine Learning*. PMLR, 23318–23340.
- [33] Xinpeng Wang, Chandan Yeshwanth, and Matthias Nießner. 2021. Sceneformer: Indoor scene generation with transformers. In *2021 International Conference on 3D Vision (3DV)*. IEEE, 106–115.
- [34] Qihong Anna Wei, Sijie Ding, Jeong Joon Park, Rahul Sajjani, Adrien Poulenard, Srinath Sridhar, and Leonidas Guibas. 2023. Lego-net: Learning regular rearrangements of objects in rooms. *arXiv preprint arXiv:2301.09629* (2023).
- [35] Xi Wei, Tianzhu Zhang, Yan Li, Yongdong Zhang, and Feng Wu. 2020. Multi-modality cross attention network for image and sentence matching. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10941–10950.
- [36] Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Reinforcement learning* (1992), 5–32.
- [37] Kerui Xu, Jingxuan Yang, Jun Xu, Sheng Gao, Jun Guo, and Ji-Rong Wen. 2021. Adapting user preference to online feedback in multi-round conversational recommendation. In *Proceedings of the 14th ACM international conference on web search and data mining*. 364–372.
- [38] An Yang, Junshu Pan, Junyang Lin, Rui Men, Yichang Zhang, Jingren Zhou, and Chang Zhou. 2022. Chinese CLIP: Contrastive Vision-Language Pretraining in Chinese. *arXiv preprint arXiv:2211.01335* (2022).
- [39] Yifei Yuan and Wai Lam. 2021. Conversational fashion image retrieval via multi-turn natural language feedback. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 839–848.
- [40] Ruiyi Zhang, Tong Yu, Yilin Shen, Hongxia Jin, Changyou Chen, and Lawrence Carin. 2020. Reward Constrained Interactive Recommendation with Natural Language Feedback. *arXiv preprint arXiv:2005.01618* (2020).
- [41] Kun Zhou, Wayne Xin Zhao, Shuqing Bian, Yuanhang Zhou, Ji-Rong Wen, and Jingsong Yu. 2020. Improving conversational recommender systems via knowledge graph based semantic fusion. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1006–1014.
- [42] Kun Zhou, Yuanhang Zhou, Wayne Xin Zhao, Xiaoke Wang, and Ji-Rong Wen. 2020. Towards topic-guided conversational recommender system. *arXiv preprint arXiv:2010.04125* (2020).