DecoMind: A Generative Ai System for Personalized Interior Design Layouts

Reema Alshehri, Rawan Alotaibi, Leen Almasri, Rawan Altaweel

***Abstract*—This paper introduces a system for generating interior design layouts based on user inputs, such as room type, style, and furniture preferences. CLIP extracts relevant furniture from a dataset, and a layout that contains furniture and a prompt are fed to the Stable Diffusion with ControlNet to generate a design that incorporates the selected furniture. The design is then evaluated by classifiers to ensure alignment with the user's inputs, offering an automated solution for realistic interior design.**

I. Introduction

Interior design has traditionally relied on professional designers, which can be costly and time-consuming, especially when users seek custom layouts that align with their style and space constraints. With the rise of artificial intelligence, there is a growing opportunity to automate this process and make it accessible to everyone. Our project, DecoMind, addresses this challenge by combining powerful AI models to create personalized, realistic interior designs. It allows users to input their room dimensions, preferences, and furniture selections, and the system generates intelligent layouts that meet those needs. DecoMind bridges the gap between creativity and automation in the field of interior design.

II. Related Work

Several previous research papers have explored AI-driven interior design generators. The CLIP-Layout model [**?**] introduced a method that learns visual correspondences between layout and text, enabling style-consistent and aesthetically pleasing scenes. However, this approach primarily focused on matching visual style and lacked structured control over furniture placement.

Another approach is DeepFurniture [1], which uses object detection and re-ranking modules to identify furniture in room images and recommend compatible sets from a fixed dataset. Although useful for retrieval, it does not support generating new layouts or adapting to customized user constraints such as room size or door/window placement.

More recently, diffusion-based techniques have shown promise in generating realistic interior scenes. For example, CreativeDiffusion [2] integrates ControlNet with Stable Diffusion to guide image generation based on structural layouts. This model improves coherence between generated images and the room's physical structure, yet its focus is limited to visual realism and does not include user-specific furniture or interactive customization.

Our Contribution: Unlike previous works, our project *DecoMind* offers a fully integrated system that:

- Combines CLIP and Stable Diffusion for furniture-aware generation based on both spatial structure and user preferences.
- Allows users to define the room dimensions, door/window positions, furniture source (e.g., IKEA), and specific aesthetic styles.
- Automatically extracts suitable furniture, filters complex scenes, removes backgrounds, and generates structured layout images.
- Uses ControlNet-guided diffusion to synthesize realistic designs while preserving layout alignment.
- Finally, classifies the generated output and evaluates it against user intent, providing feedback for future tuning.

This holistic approach bridges the gap between retrieval-based systems like DeepFurniture and style-guided models like CLIP-Layout, enabling a more personalized, interactive, and layout-consistent design experience.

III. Data Description

To develop the DecoMind system, we utilized three separate datasets due to the absence of a comprehensive dataset that simultaneously includes room type, interior style, and IKEA furniture annotations. The datasets used are summarized as follows:

- IKEA Furniture Dataset [3]: This dataset contains three folders: Train, Validation, and Test. Each folder includes images of IKEA furniture items (e.g., sofas, chairs, tables, etc.). It is primarily used to guide the model in generating images that feature IKEA furniture accurately. Additionally, we performed a custom filtering process to remove low-quality or irrelevant images, ensuring a cleaner and more accurate dataset for furniture representation.
- House Rooms Image Dataset [4]: This dataset is organized into multiple folders, with each folder representing a specific room type (e.g., bedrooms, kitchens, living rooms, etc.). It enables the model to recognize and differentiate between various types of rooms.
- Interior Design Styles Dataset [5]: This dataset consists of several folders, each corresponding to a particular interior style (e.g., modern, classic, minimalist, etc.). It helps the model identify and distinguish between various interior design styles. *
  * This dataset includes copyright images from Houzz.com. Use is strictly limited to academic and research purposes only. Redistribution or commercial use is not allowed.

IV. Methodology

We have multiple stages in our study, each utilizing different methods for achieving the desired results.

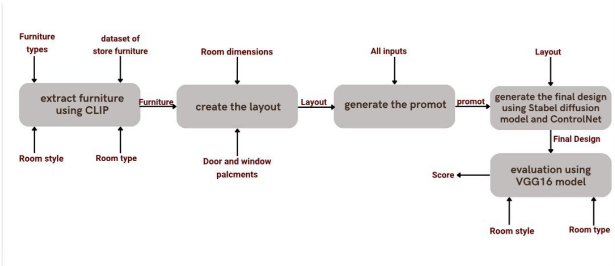

Fig. 1. Overview of the general workflow of the DecoMind system.

A. Furniture Extraction

The process begins with extracting furniture from the selected store. The extraction process is driven by user choices, which include room type, style, and furniture types. This is achieved using a trained CLIP model, which uses a dual-encoder architecture—one for processing images and the other for processing text. CLIP matches the provided text (room type, style, furniture types) with relevant images to ensure the best outputs.

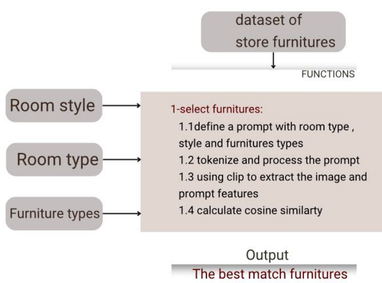

Fig. 2. Extracting furniture.

B. Image Generation Using Stable Diffusion

The next stage involves the main component of the system: the Stable Diffusion model. This model generates images from text inputs using a latent diffusion process. It has achieved impressive results in recent years but has limitations, particularly in using reference images. In our case, we wanted the model to generate interior designs with the furniture selected by the user. To overcome this limitation, we integrated ControlNet, which guides Stable Diffusion using additional input such as the layout of the room (room dimensions, window and door positions) and the extracted furniture from the previous stage.

-Stable Diffusion Architecture: Stable Diffusion uses a deep convolutional neural network with a transformer-based architecture for image generation. It operates by mapping the text prompt into a latent space, where it iteratively refines the image through a denoising process.

-ControlNet Architecture: ControlNet acts as a conditioning network, guiding the generative process by adding constraints from the layout and furniture data, ensuring that the generated design adheres to the required dimensions and object placements.

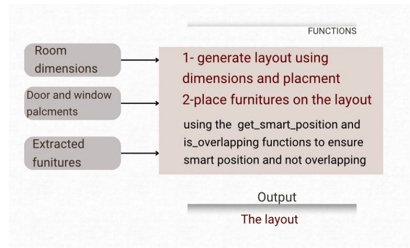

Fig. 3. Generate the layout.

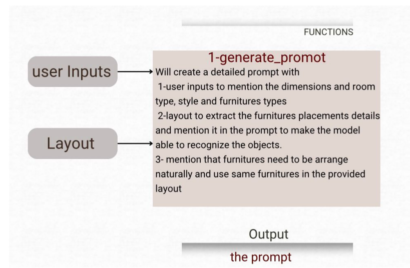

Fig. 4. Generate the prompt.

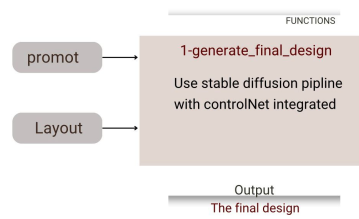

Fig. 5. Generate the final design.

C. Challenges in Dataset

We encountered some issues with the store's dataset. Many furniture items were insufficient for the model's generation process, and the dataset didn't provide full 3D views of the objects. To address these problems, we instructed the model in the prompt to "imagine and correct" the angles of the furniture objects and feel free to add more items to complete the design.

D. Fine-tuning Classifiers

To ensure better results, we fine-tuned two classifiers based on the VGG16 architecture. One classifier was fine-tuned on room types datasets, and the other on room styles datasets. These classifiers are used to score the match between the user inputs (room type and style) and the generated design. If the design matches both the room type and style, the score will be high. However, if one or both do not match, the score will decrease.

-VGG16 Architecture: The VGG16 model is a deep convolutional network with 16 layers. It is used for image classification tasks, where it maps input images to predefined classes (room type and style, in our case) based on learned features.

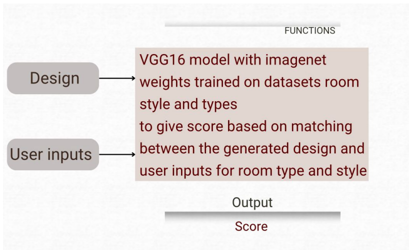

Fig. 6. Evaluation.

V. Experiments

In this section, we first introduce the experimental setup in Subsection V-A. Then, we present the results of qualitative evaluation and quantitative evaluation in Subsections V-B and V-C.

A. Experimental Setup

The DecoMind system was built using the Google Colab environment, utilizing a Tesla T4 GPU. Three different datasets were used in this project (more details are provided in the Data Description section). Several models were integrated to build the system (all implementation details and specific usages of each model are explained in the Methodology section).

To evaluate the final results in terms of room type and style accuracy, we fine-tuned two classifiers based on the VGG16 architecture (further information about the classifiers is also provided in the Methodology section).

B. Qualitative Evaluation

The DecoMind system demonstrated its effectiveness through the quality of the generated design images. The outputs closely matched the user's textual descriptions, with only minor differences, if any. Examples of the system's results are presented below in Figures 7 to 9 to illustrate its ability to align visual generation with user preferences.

Fig. 7. A generated modern living room design

Fig. 8. A generated modern bedroom design

Fig. 9. A generated minimalist bedroom design

C. Quantitative Evaluation

To quantitatively assess the performance of the DecoMind system, We used the two fine-tuned classifiers based on the VGG16 architecture to evaluate whether the generated design matched the expected category. For the case where the user described a ”modern living room with dimensions (4,5) the window to the north , the door to the west with furniture types[couch and bed] ” the results of layout and generated design below

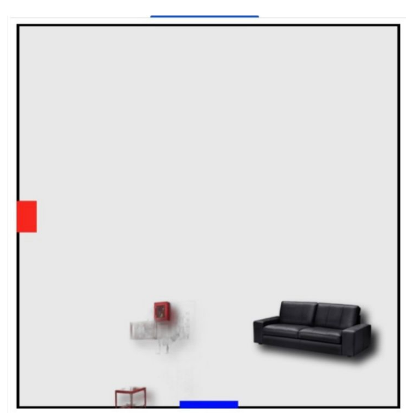

Fig. 10. The generated layout

Fig. 11. The generated image

The results of both classifiers are summarized below.

TABLE I
RESULTS OF THE TWO CLASSIFIERS EVALUATING THE GENERATED LIVING ROOM DESIGN.

Evaluation Report	
Room Type Match	Yes
Style Match	No
Final Score	0.50

VI. Discussion

In this section, we outline several proposed improvements to further enhance the DecoMind system, categorized into dataset improvements, layout improvements, model improvments, and user interaction enhancements.

A. Dataset Improvements

- Expand Datasets: Integrate datasets from multiple furniture stores and include 3D views of furniture to provide more accurate and diverse furniture options, thus improving model generalization.
- IKEA Dataset Enhancements:
  - Background Removal: Completed background removal for transparent furniture images.
  - Furniture Filtering: Completed filtering to retain only individual furniture pieces, removing full-room images.

B. Layout Improvements

- Furniture Placement: Implemented smart position generation and overlap detection mechanisms to place furniture realistically and avoid visual collisions between items.

C. Model Improvements

- Stable Diffusion Model: The Stable Diffusion model showed limitations in generating accurate furniture matches and realistic room layouts. We attempted to overcome this using the Inpainting variant, which maintained furniture consistency but produced visually unrealistic and low-quality results, indicating the need for fine-tuning.

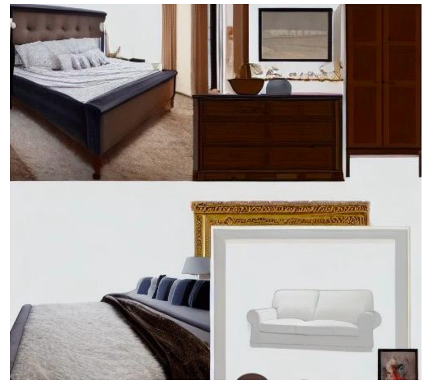

Fig. 12. Result generated by the Inpainting Stable Diffusion model.Furniture is consistent, but the design lacks realism.

  To improve spatial realism, we explored 3D simulation techniques by adjusting layouts and crafting prompts that emphasized perspective (e.g., "a view from the room corner," "showing both side walls"). We also attempted multi-view generation from different angles to mimic a 3D experience. While these strategies improved some visual cues, the model still lacked consistent depth and spatial coherence. This highlighted the core limitation: Stable Diffusion was not trained on 3D interior data. Fine-tuning on a 3D-specific dataset would likely resolve this, but was beyond our project's scope due to time and resource constraints.
- CLIP Model: The furniture extraction and description capabilities of the CLIP model require fine-tuning to improve the accuracy of matching furniture types with user requests.

D. User Interaction Enhancements

- Custom Furniture Arrangement: Allow users to manually arrange furniture to provide greater flexibility and customization within the generated room layout.
- Feedback System: Implement an easy-to-use feedback mechanism that enables users to rate and suggest improvements to the generated designs, thereby supporting iterative system enhancement.

VII. Conclusion

In conclusion, this paper demonstrates the potential of combining CLIP, Stable Diffusion, and ControlNet to automate the process of generating interior design layouts based on user specifications. By extracting furniture, generating layout designs, and evaluating the results against room type and style classifiers, the proposed system provides a powerful tool for non-designers and enthusiasts alike. Despite the promising results, challenges remain in ensuring that the generated designs closely match the user's selected furniture, and improving the performance of the CLIP model in extracting the most relevant furniture pieces. Nonetheless, the system represents a significant step forward in applying AI to the creative field of interior design, with future improvements aimed at increasing the realism and accuracy of the designs.

Appendix A
Code Repository

The complete implementation of the DecoMind system, is available at the following public Google Colab link:

- https://colab.research.google.com/drive/17H8tYgIIZx9IsUZgw5ym1UaR1LQIZYgB?usp=sharing

References

[1] B. Liu, J. Zhang, X. Zhang, W. Zhang, C. Yu, and Y. Zhou, "Furnishing your room by what you see: An end-to-end furniture set retrieval framework with rich annotated benchmark dataset," *arXiv preprint arXiv:1911.09299*, January 2020.

[2] Z. Xu, Y. Zhang, H. Luo, and Y. Wang, "Creative interior design matching the indoor structure generated through diffusion model with an improved control network," *Frontiers of Architectural Research*, vol. 14, no. 3, pp. 614–629, 2025.

[3] P. AI, "Ikea furnitures computer vision dataset," https://universe.roboflow.com/projet-ai/ikea-furnitures/dataset/2/download, accessed: May 04, 2025.

[4] R. Reni, "House rooms image dataset," https://www.kaggle.com/datasets/robinreni/house-rooms-image-dataset, accessed: May 04, 2025.

[5] S. Arullin, "Interior design styles," https://www.kaggle.com/datasets/stepanyarullin/interior-design-styles?utm_source=chatgpt.com, accessed: May 04, 2025.