3D-FRONT: 3D Furnished Rooms with layOuts and semaNTics
— Supplementary Material —

Figure 1: **Metrics in "Validation and Assessment".** "1-N" here means 1-7 because there are seven required objects in the room. When recommending a single sofa, the designer selected the *16th* sofa from the retrieval sequence. Thus, the single sofa ranks 16 in this recommendation step. Hit@K refers to TopK recall accuracy.

1. How to perform recommendation?

We have explained how we create room suites through our reconmender system driven by FCS-GAE in Sec. 3.1 in the main paper. Here, we take a specific example to show the recommendation process. As show in Figure 1 (the middle part), given a room suite that contains a double sofa and a stool, the goal is to add a single sofa to current suite. We first utilize the trained GAE to perform primary-ranking, then employs the trained GBDT-LR model to re-rank the selected candidates. GBDT-LR is a module of FSC. Specifically, if we have n single sofas in the furniture pool, we can obtain n room suite candidates. For each candidate, we adopt GAE to compute its compatibility score. Then we re-compute the compatibility scores of the Top50 candidates via GBDT-LR and re-rank them. As stated in the main paper, GAE takes visual feature vectors (extracted from VEN of FSC) of objects, GBDT-LR additionally considers objects' attributes.

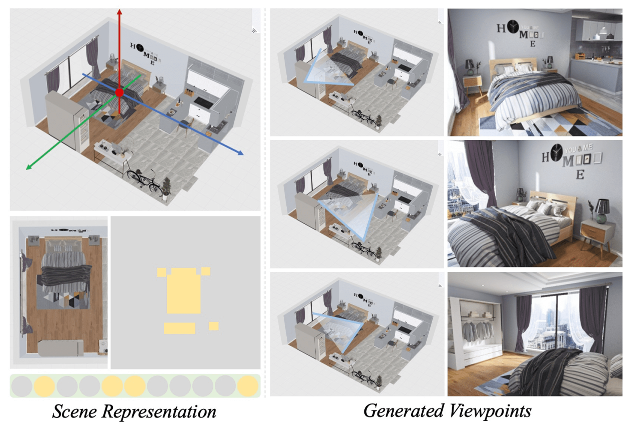


Figure 2: **Viewpoint Generation.** Each scene is associated with several natural camera views to facilitate rendering.

2. Metrics in "Validation and Assessment"

We have briefly explained some metrics in Sec. 4 (Validation and Assessment) in the main paper. Here, we present an example in Figure 1 to make them more clear.

3. Viewpoint Generation

The viewpoint generation stage aims to assign several cameras to each scene (or room), and ensures most of the cameras have practical viewpoints. We are provided with many excellent scenes (or templates) with suggested cam-

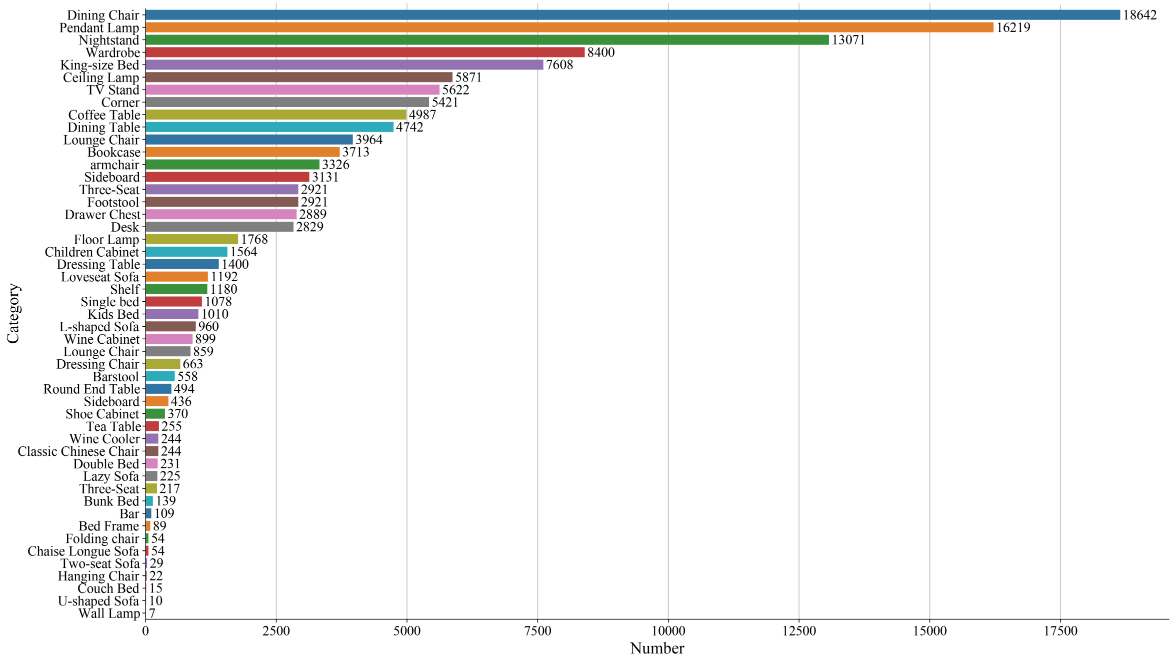


Figure 3: **Distribution of annotated instances** for 3D-FRONT's scenes corresponded to 3D-FUTURE's model categories.

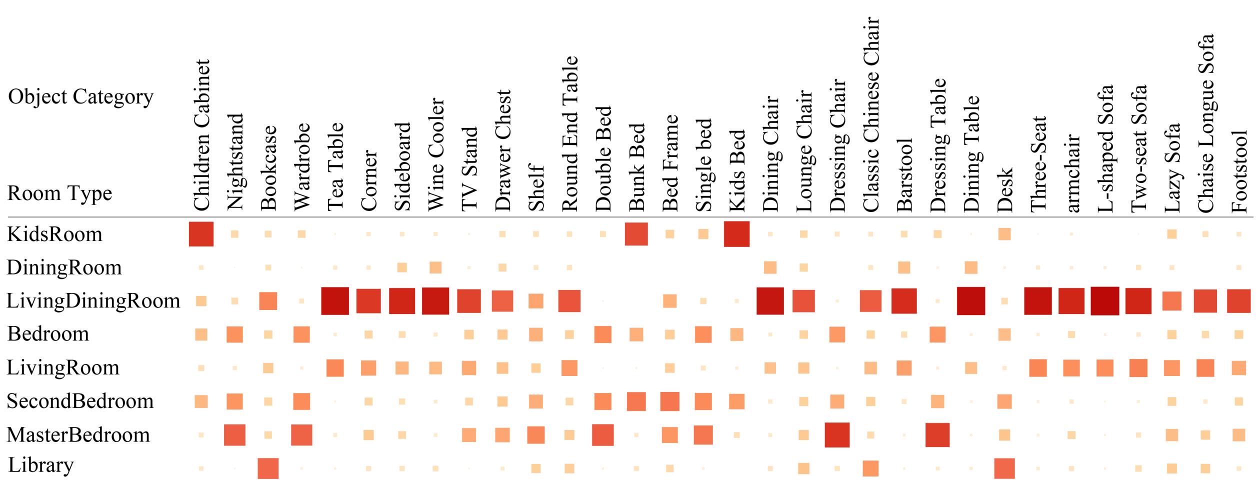


Figure 4: **Distribution of object categories conditioned on different room types.**

era viewpoints by expert designers (about 5K). We devised a retrieval algorithm to "propagate" these viewpoints to 3D-FRONT's scenes. The first thing is to represent a scene as an feature vector. As shown in Figure 2, we choose a "center" object as the origin to build the world coordinate. A "center" object could be any object in the scene. Choosing different "center" objects will give different 2D representations of the scene, thus would increase the diversity of the generated viewpoints for the scene. Here in the figure, we take the bed as a "center" object. We compute the normalized distances to define an object group (like a graph), and convert its mask projection as a feature vector to represent the scene structure. With this method, we can compute cosine similarity to perform scene retrieval from our tem-

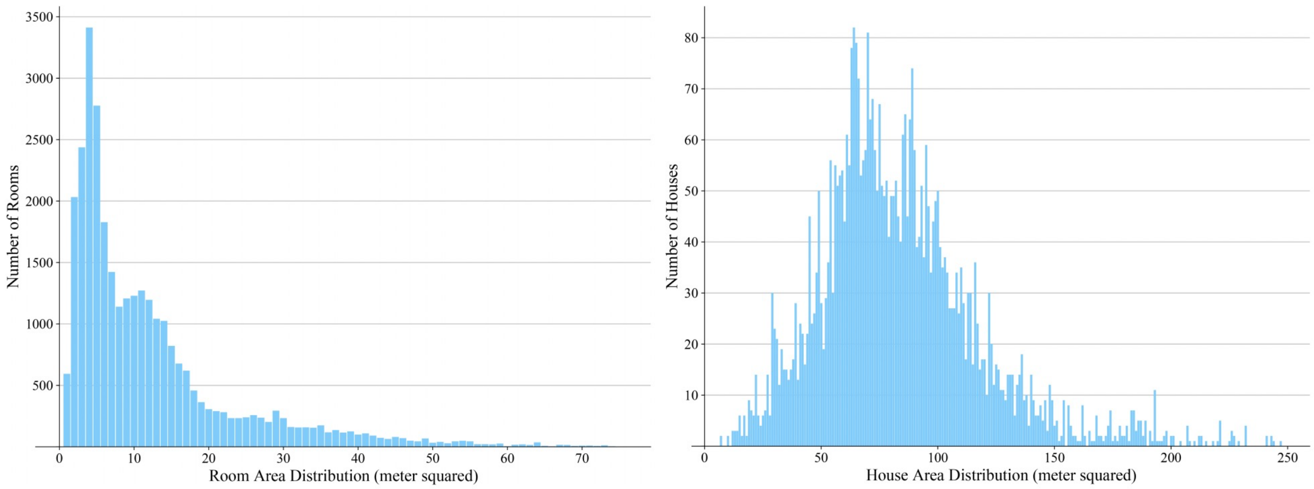


Figure 5: **Distribution of physical sizes** (in meters^2) per room (Left) and house (Right) of the 3D-FRONT dataset.

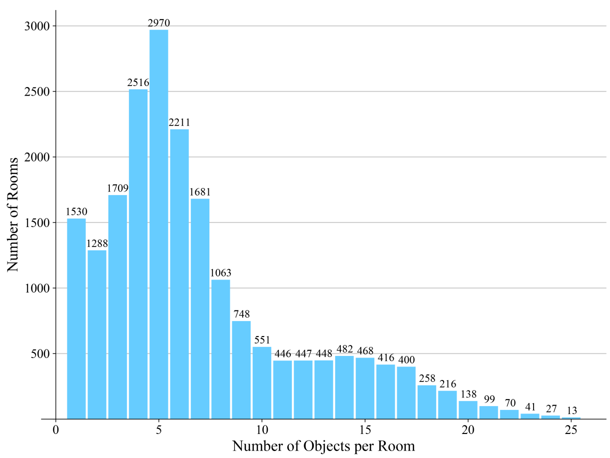


Figure 6: **Distribution of number of objects per room.**

plates, thus generate practical camera viewpoints for new scenes.

4. Other Statistics

In Figure 3, we report the distribution of annotated object labels for 3D-FRONT's scenes corresponded to the 3D-FUTURE [2] 3D CAD model categories.

In Figure 4, we present the physical sizes over the rooms and houses. The 3D-FRONT dataset are measured in real-world spatial dimensions (units are in meters).

In Figure 5, we show the distribution of object categories conditioned on different room types. The area of the square denotes the frequency of a given object category that appears in a certain room type. The frequency is normalized for each object category. It strongly implies the relationships between objects and rooms. For example, categories such as children cabinet, bunk bed, and kids bed are more likely to appear in a kid room, while bookcase and desk are more likely to appear in a study room. We can learn rich design knowledge from the distribution.

In Figure 6, we show the distribution of number of objects per room. At this time, 3D-FRONT's rooms are furnished by the functional furniture.

5. User Studies

We perform user studies to show the quality of 3D-FRONT using Amazon Mechanical Turk (AMT). The compared datasets are 3D-FRONT, SUNCG [3], and ShapeNet [1]. The questions and the scores are shown in Figure 7. The tasks are explained below.

Dataset: *Scene Quality*. We have 90 pairs of scenes randomly sampled from SUNCG and 3D-FRONT based on scene types (LivingRoom, DiningRoom, and Bed Room) in our questionnaire. Each scene type contains 30 pairs. We study layout plausibility, design quality, texture quality, and style compatibility in this task. We have collected 20 questionnaires from master-level annotators in AMT. That means each scene pair has been labeled by 20 annotators. Thus, the final scores are calculated using 1,800 feedback.

Dataset: *Model Quality*. We have 30 pairs of furniture models randomly sampled from SUNCG and 3D-FRONT based on categories in our questionnaire. We study texture quality and model's visual quality in this task. We have collected 20 questionnaires from master-level annotators in AMT. Thus, the final scores are calculated using 600 feedback.

Application: *Layout Synthesis*. We randomly sampled 60 rooms from 3D-FRONT (36) and SUNCG (24). For each room, we synthesizing a pair of layouts, with one

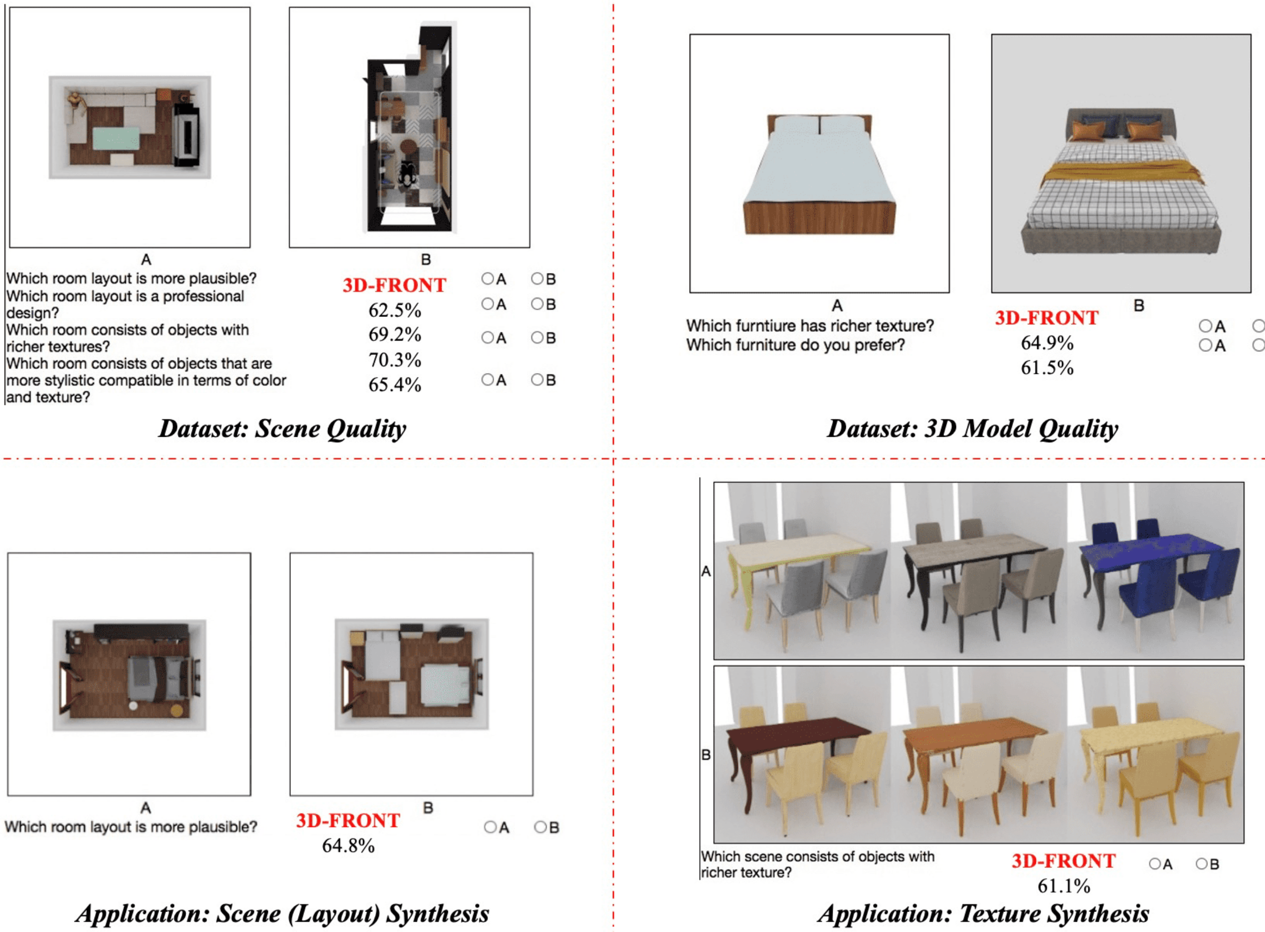


Figure 7: **User Studies.** We perform the listed user studies using Amazon Mechanical Turk (AMT). 3D-FRONT's scores are reported in the figure. From the dataset comparisons, we see that for each quality criterion assessed, the majority of Turkers (between 60% and 70%), preferred data presented by 3D-FRONT. See Sec. 4 for the experimental settings.

produced by the model trained on 3D-FRONT and the other generated by the model trained on SUNCG. We study layout plausibility in this task. We have collected 20 questionnaires from master-level annotators in AMT. Thus, the final score is calculated using 1,200 feedback.

Application: ***Texture Synthesis*.** We have selected 5 DiningRoom corners from 3D-FRONT and textured their chairs and tables using the learned texture synthesis models (3D-FRONT vs. ShapeNet). For each corner, we have perform texture synthesis three times with random noises. We study texture diversity in this task. We have collected 20 questionnaires from master-level annotators in AMT. Thus, the final score is calculated using 100 feedback.

6. More House & Room Examples

In Figure 8, Figure 9, Figure 10, Figure 11, and Figure 12, we present more house and room examples to demonstrate the quality of 3D-FRONT.

References

[1] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. ShapeNet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 3

[2] Huan Fu, Rongfei Jia, Lin Gao, Mingming Gong, Binqiang Zhao, Steve Maybank, and Dacheng Tao. 3d-future: 3d furniture shape with texture. *arXiv preprint arXiv:2009.09633*, 2020. 3

[3] Shuran Song, Fisher Yu, Andy Zeng, Angel X Chang, Manolis Savva, and Thomas Funkhouser. Semantic scene completion from a single depth image. In *Proceedings of 30th IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 3

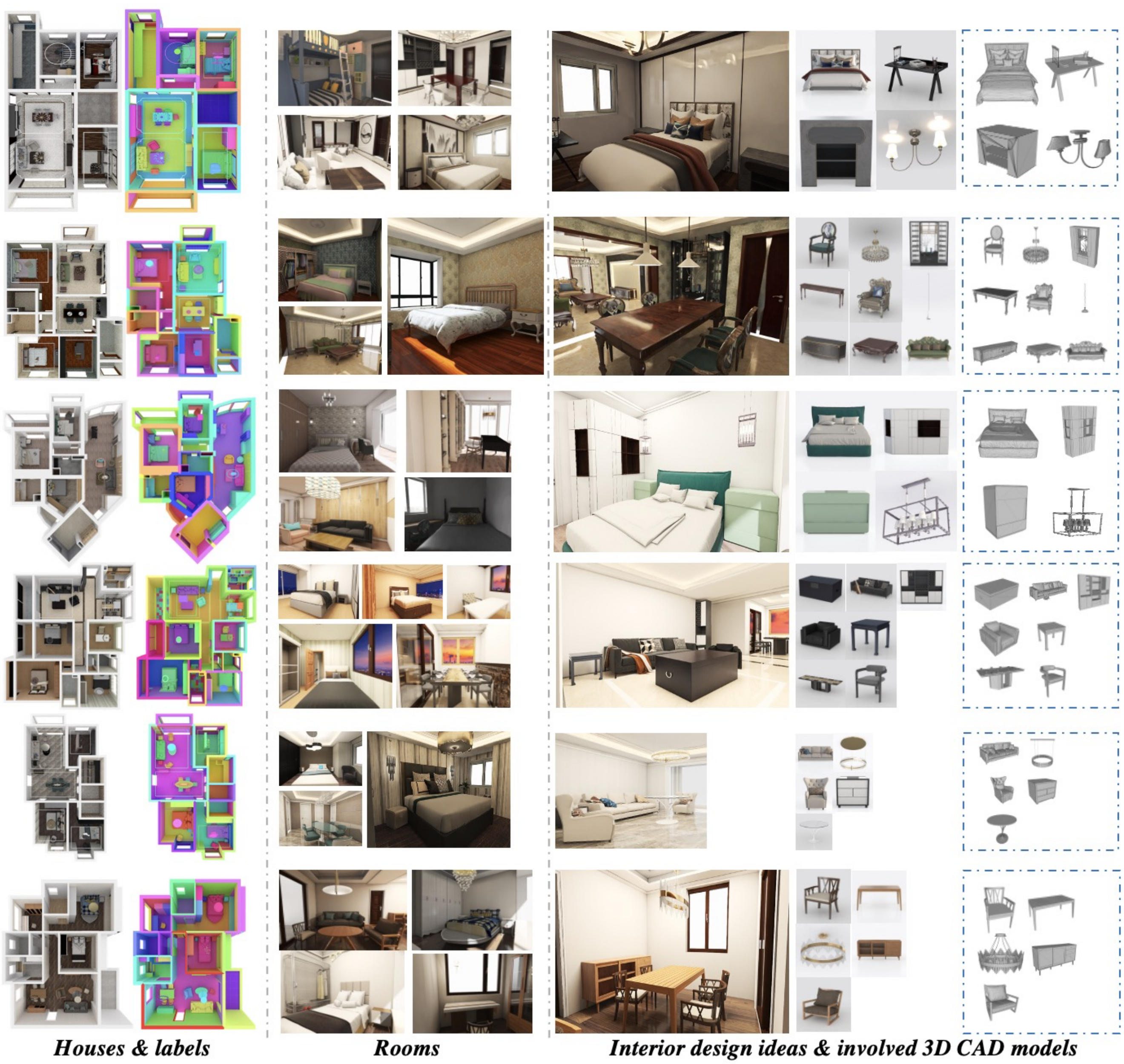


Figure 8: **House Examples - Part 1.** Zoom in for better view.

Figure 9: **House Examples - Part 2.** Zoom in for better view.

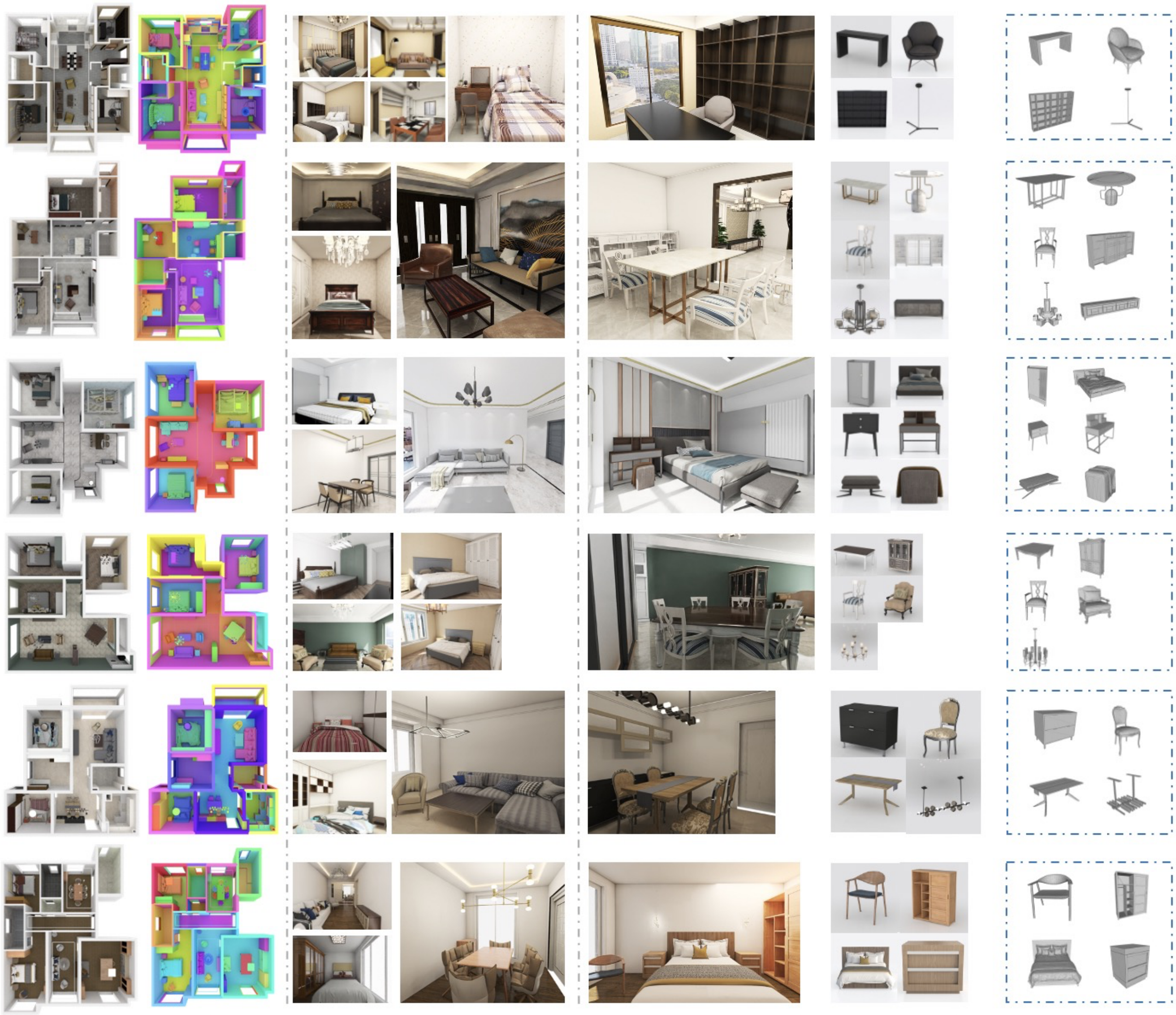
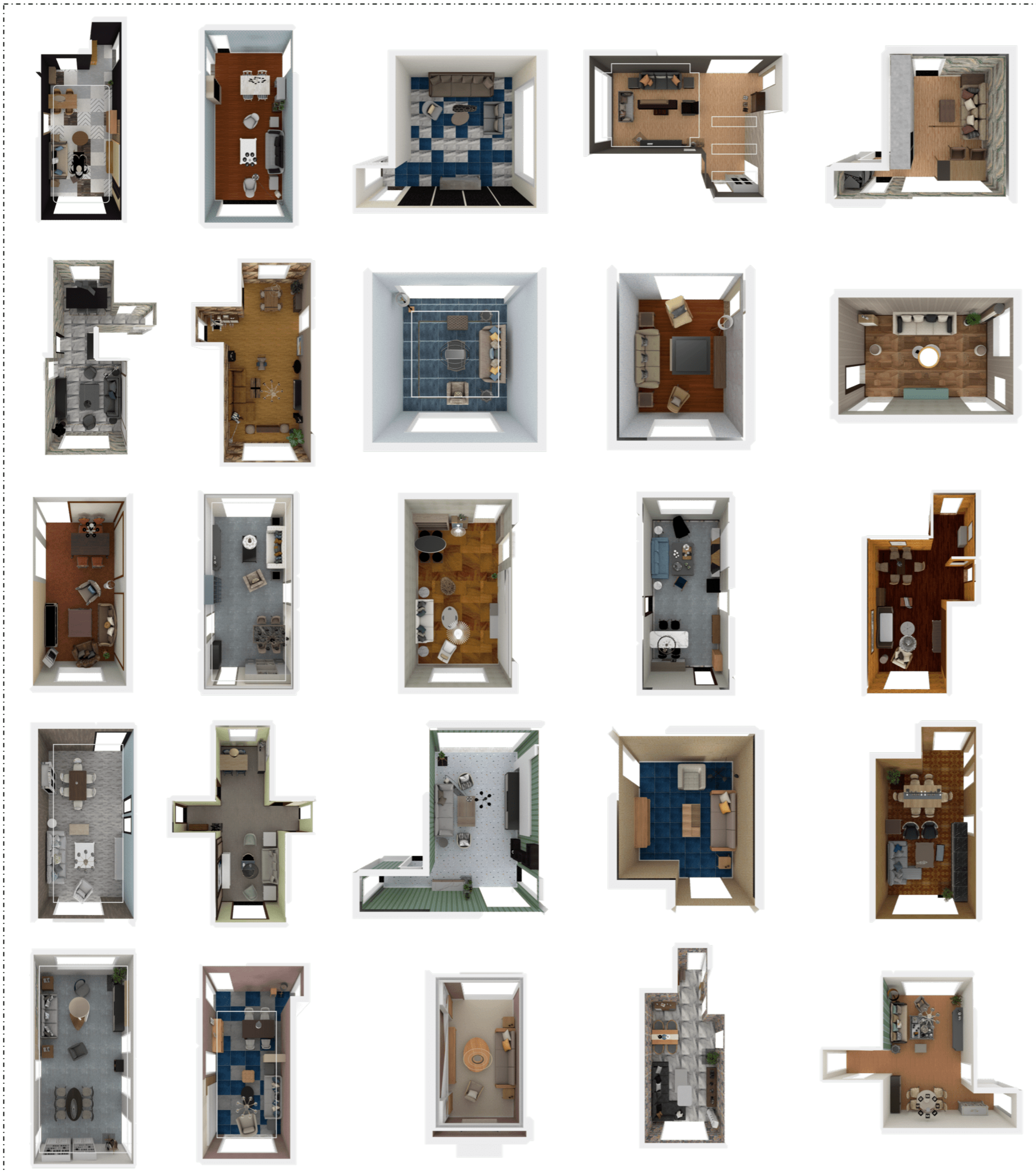


Figure 10: **House Examples - Part 3.** Zoom in for better view.



Living Room

Figure 11: **Room Examples - Part 1.** Zoom in for better view.

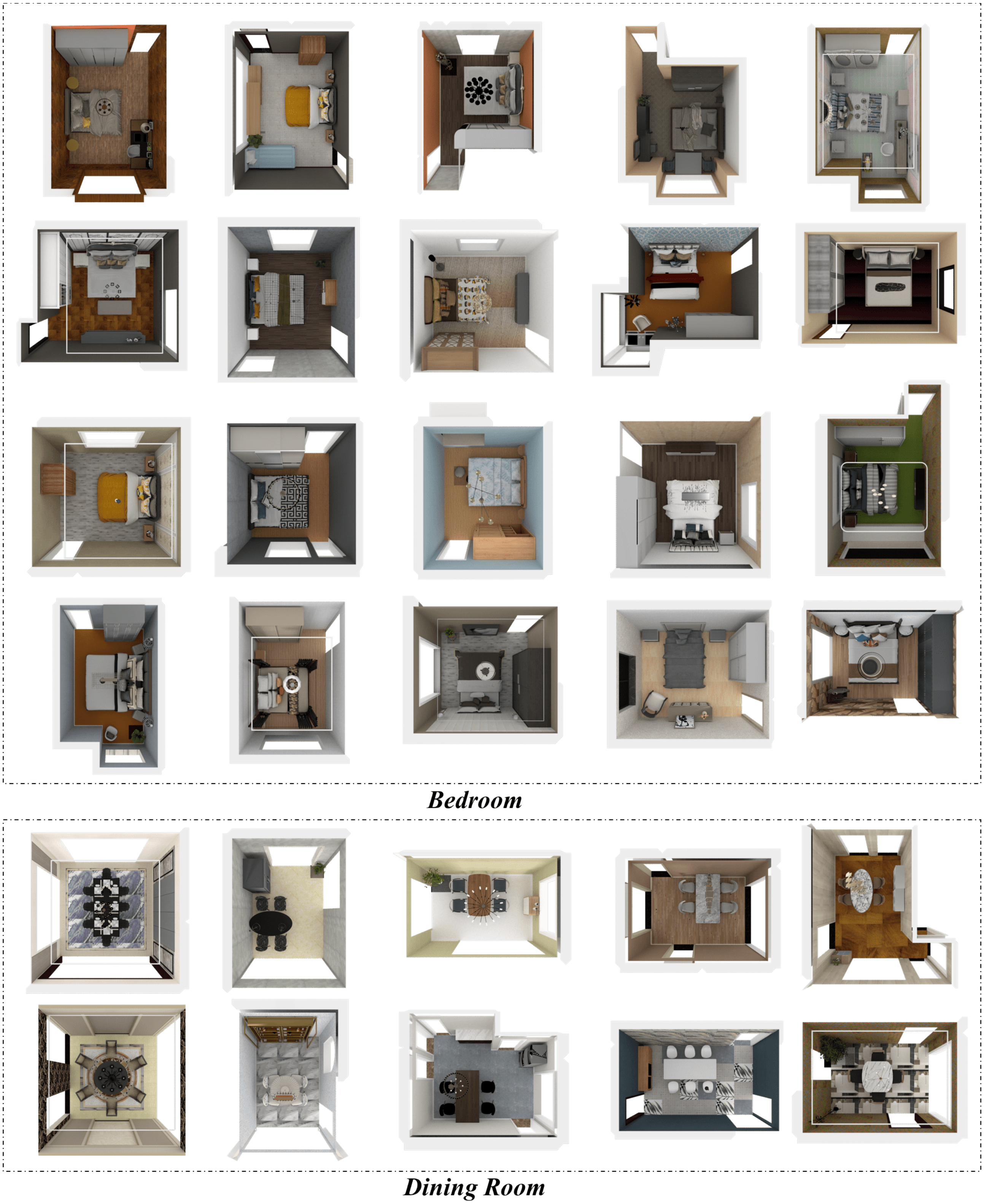


Figure 12: **Room Examples - Part 2.** Zoom in for better view.