1 Data Preparation

We train our networks and other methods on the 3D-FRONT dataset [1]. To guarantee the quality of the training samples, we first reserve the rooms with regular and reasonable size, which is smaller than 8×8 meters. Then we manually modify the rooms and the object instances by following instructions: (1) delete the wrong classified objects; (2) modify the scale of objects so that it is inside the floor plan; (3) translate or delete the objects avoiding mesh collisions; (4) delete the objects floating in the air. After the mentioned improvement, we obtained and relabeled 4989 rooms of four types: bedrooms, dining rooms, libraries, and living rooms. To cooperate with our group-based scene generation framework, we re-annotate the filtered 3D-FRONT scenes as a set of object instance functional groups. We choose 12 semantic group categories that satisfy indoor space's daily requirements and cover all the presented objects in the scene dataset. The semantic group includes all relative objects to support certain daily functionality. All the semantic groups are manually annotated and re-checked by our project members. The room and group number of our annotated dataset is listed in Table 1 and Table 2, respectively. We illustrate the exemplars of the semantic groups in Figure 1. Figure 1 shows the representative groups that consists multi-object to better differ the concept of semantic group and single object. Considering the human-labor efficiency, we also develop an annotation toolkit on the Unity3D platform. We visualize a screenshot of the software to show an exemplar of the annotation process in Figure 2.

Room Types	Room Amount	Group Amount
Bedroom	1684	5776
Dining room	1037	3424
Library	821	2666
Living room	1447	4897

Table 1: Re-annotated room amounts and semantic group numbers.

Group Types	Group Amount	Group Types	Group Amount
Sleeping	1699	Book Storage	1276
Relaxing	815	Shelf Storage	1660
Bed Watching TV	360	Kithcen Storage	470
Lighting	4229	Watching TV	1965
Dressing	240	Working	940
Cloth Storage	1362	Eating	1747
Total	16763		

Table 2: Annotated semantic groups categories and numbers.

2 Model Choice Discuss

ProposeNet. The ProposeNet is designed to generate continuous and low-dimensional group descriptors. Although other model choices are possible, such as VAE or GAN, the normalizing flow is more stable and lightweight to train, and suitable to produce continuous vectors. To better understand the ProposeNet, we visualize the ground truth's group proposal position and rotation distributions and the ProposeNet outputs in Figure 3. In order to visually evaluate the learned distributions, we select four representative semantic group categories of the

bedroom and sample 150 latent features for each selected semantic group class. We record the output's center and orientation in a normalized room space. As we can see from the illustration, the ProposeNet can successfully fit the location patterns subject to the conditions of different semantic categories in 3D space. It validates that the proposed conditional normalizing flow-based ProposeNet can produce plausible proposals for the proposed selection and optimization phases.

CompleteNet. Conversely, CompleteNet needs to generate a graph seed to produce local semantic group objects. Thus, the normalizing flow is inefficient in generating a high-dimensional (64 or 128 channels) graph's global descriptor auto-regressively, both in the training and sampling phases. Due to this limit of the flow-based approach, we adopt a typical VAE architecture to generate graph global feature conditioning on the user-intent inputs efficiently.

3 Additional Visualized Results

Due to the page limit, we provide more qualitative results in order to better exhibit the results of our proposed method. We show the additional visualized results of our method on four scene types of bedroom, dining room, library, and living room in Figure 4.

References

[1] Huan Fu, Bowen Cai, Lin Gao, Lingxiao Zhang, Jiaming Wang, Cao Li, Qixun Zeng, Chengyue Sun, Rongfei Jia, Binqiang Zhao, and Hao Zhang. 3d-front: 3d furnished rooms with layouts and semantics. In *IEEE/CVF ICCV*, pages 10913–10922, 2021. doi: 10.1109/ICCV48922.2021.01075. URL https://doi.org/10.1109/ICCV48922.2021.01075.

Figure 1: Exemplar of semantic groups.

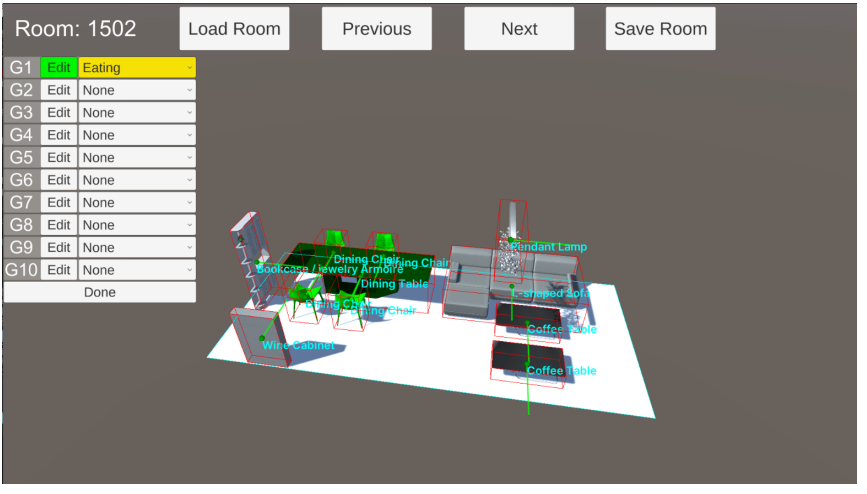

Figure 2: Screenshot of Unity3D-based semantic group annotation toolkit. The objects in green are annotated as a semantic group.

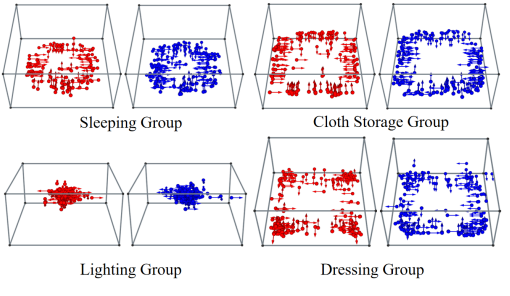

Figure 3: Qualitative evaluation of the ProposeNet performance. We illustrate the semantic group center position (point) and orientation (arrow) distributions of the ground truth (red) and the ProposeNet predictions (blue) in the normalized 3D room space.

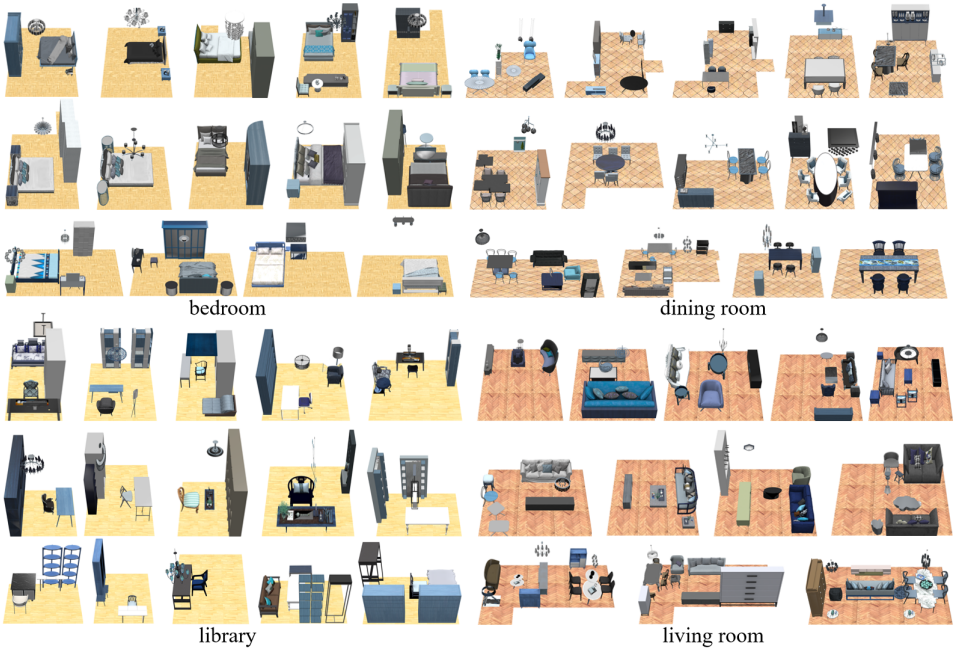

Figure 4: Exemplar of generated scenes from our method.