Supplementary of "Intelligent Home 3D: Automatic 3D-House Design from Linguistic Descriptions Only"

Qi Chen[1,2*] Qi Wu[3*] Rui Tang[4] Yuhan Wang[4] Shuai Wang[4] Mingkui Tan[1†]
[1]South China University of Technology, [2]Guangzhou Laboratory, Guangzhou, China
[3]Australian Centre for Robotic Vision, University of Adelaide, Australia, [4]Kujiale, Inc.

sechenqi@mail.scut.edu.cn qi.wu01@adelaide.edu.au
{ati,daishu,luorui}@qunhemail.com mingkuitan@scut.edu.cn

1. Details of Scene Graph Parser

Given a series of linguistic requirements, we extract the corpus from dataset and then construct the scene graphs based on the corresponding expressions like [1]. First, we distribute the words in corpus into three categories: object O, relation R and attribute A. Given an input sentence S, we convert this sentence to a scene graph $G = (V, E)$. $V = \{v_1, v_2, \cdots, v_n\}$ is the objects with attributes, which have been mentioned in sentence S. Specifically, v_i consists of o_i and A_i (*i.e.*, $v_i = (o_i, A_i)$), where $o_i \in O$ denotes the object of v_i while $A_i \subseteq A$ is the attribute of v_i. $E \subseteq V \times R \times V$ denotes the set of relations between two objects. Each relation $e_i = (o_j, r_i, o_k)$, where $r_i \in R$.

Scene graph of each room. For example, for a given room (*e.g.*, "livingroom1"), we have the linguistic descriptions S_1 ="*livingroom1 is in center with 21 square meters*" and S_2 ="*livingroom1 wall is Earth_color Wall_Cloth while uses Black Log for floor*". We first transform S_1 to an object node $v_1 = (\textit{livingroom1}, \{\textit{center}, \textit{21 square meters}\})$ and the relation $e_1 = (v_1, \textit{is}, \emptyset)$. For the sentence S_2, we extract the objects $v_2 = (\textit{livingroom1}, \emptyset)$, $v_3 = (\textit{wall}, \{\textit{Earth\_color}, \textit{Wall\_Cloth}\})$ and $v_4 = (\textit{floor}, \{\textit{Black}, \textit{Log}\})$. The corresponding relations are $e_2 = (v_2, \textit{have}, v_3)$ and $e_3 = (v_2, \textit{have}, v_4)$. Since v_1 and v_2 have the same object (*i.e.*, *"livingroom1"*), we merge them together and obtain $v_5 = (\textit{livingroom1}, \{\textit{center}, \textit{21 square meters}\})$. Thus, we finally get the objective scene graph (shown in Figure A).

Scene graph of adjacency between rooms. In addition, the descriptions on our dataset also contain the adjacent information between the rooms, such as *"livingroom1 is adjacent to washroom1, bedroom1, study1"* or *"bedroom1 is next to study1"*. In order to make use of these messages, we construct another scene graph, which focuses on the relations among the rooms mentioned in given sentence. For example, given the sentences S_3 = *"livingroom1 is adjacent to washroom1, bedroom1, study1"* and S_4 = *"bedroom1 is next to study1"*, we first transform S_3 according to the aforementioned rules and obtain the objects and relations: $v_{10} = (\textit{livingroom1}, \emptyset)$, $v_{11} = (\textit{washroom1}, \emptyset)$, $v_{12} = (\textit{bedroom1}, \emptyset)$ and $v_{13} = (\textit{study1}, \emptyset)$; $e_{10} = (v_{10}, \textit{is adjacent}, v_{11})$, $e_{11} = (v_{10}, \textit{is adjacent}, v_{12})$ and $e_{12} = (v_{10}, \textit{is adjacent}, v_{13})$. For sentence S_4, the objects are $v_{14} = (\textit{bedroom1}, \emptyset)$ and $v_{15} = (\textit{study1}, \emptyset)$ while the relation is $e_{13} = (v_{14}, \textit{is next}, v_{15})$. Due to $v_{12} = v_{14}$ and $v_{13} = v_{15}$, we replace v_{14} and v_{15} by v_{12} and v_{13}, respectively. Therefore, e_{13} can be reformulated as $e_{13} = (v_{12}, \textit{is next}, v_{13})$. We exhibit the visualised scene graph of these expressions in Figure B.

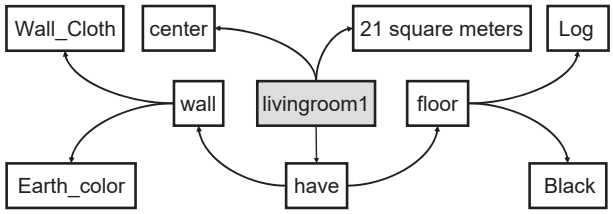


Figure A. Scene graph of *"livingroom1"* according to the sentences S_1 and S_2.

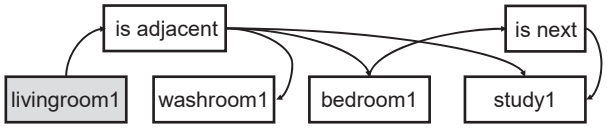


Figure B. Scene graph of adjacency between rooms according to the sentences S_3 and S_4.

*Authors contributed equally.
†Corresponding author.

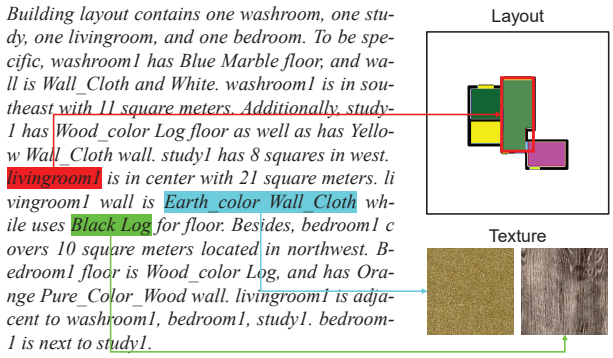


Figure C. An example from Text–to–3D House Model dataset. The given sentence contains the descriptions of both building layout and textures. For example, *"livingroom1"* involves the messages about position (*"center"* and size (*"21 square meters"*) in linguistic expressions while a coarse bounding box covering the outline of room. The wall and floor have their descriptions (*i.e.*, *"Earth_color Wall_Cloth"* and *"Black Log"*) and the corresponding ground-truth texture images.

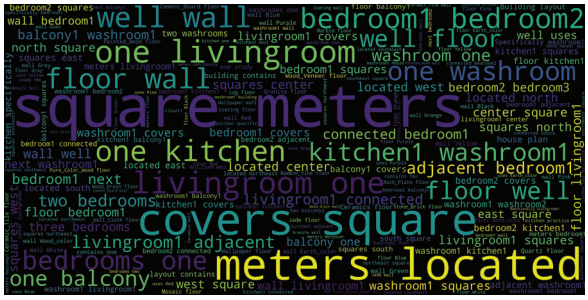


Figure D. The word cloud of the texts in our dataset.

2. Dataset Analysis

To generate 3D building models from natural language descriptions, we collect a new Text–to–3D House Model dataset, which contains $2,000$ houses, $13,478$ rooms and 873[1] texture images with corresponding natural language descriptions. These descriptions are firstly generated from some pre-defined templates and then refined by human workers. The average length of the description is 173.73 and there are 193 unique words. All the building layouts are designed on the canvas with the pixel size of 512×512, which represents 18×18 square meters in the real world. We take an example from our proposed dataset and show in Figure C. Moreover, we also provide the word cloud of our dataset and the visualised results are shown in Figure D.

[1]Some rooms have same textures so this number is smaller than the total number of rooms.

3. Contents of "Text1" and "Text2"

In this section, we give the specific descriptions of "Text1" and "Text2", which have been mentioned in Figures 6 and 7 of the submitted manuscript. For convenience, in Figure E, we exhibit the visual results of 2D floor plan corresponding to "Text1" and "Text2", respectively.

"Text1": *The building contains one washroom, one bedroom, one livingroom, and one kitchen. Specifically, washroom1 has 5 squares in northeast. bedroom1 has 14 square meters in east. Besides, livingroom1 covers 25 square meters located in center. kitchen1 has 12 squares in west. bedroom1, kitchen1, washroom1 and livingroom1 are connected. bedroom1 is next to washroom1.*

"Text2": *The house has three bedrooms, one washroom, one balcony, one livingroom, and one kitchen. In practice, bedroom1 has 13 squares in south. bedroom2 has 9 squares in north. bedroom3 covers 5 square meters located in west. washroom1 has 4 squares in west. balcony1 is in south with 6 square meters. livingroom1 covers 30 square meters located in center. kitchen1 is in north with 6 square meters. livingroom1 is adjacent to bedroom1, bedroom2, balcony1, kitchen1, bedroom3, washroom1. balcony1, bedroom3 and bedroom1 are connected. bedroom2 is next to kitchen1, washroom1. bedroom3 is adjacent to washroom1.*

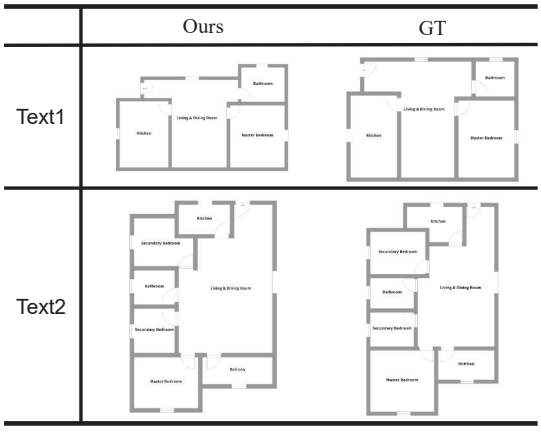


Figure E. Qualitative results of GC-LPN and ground-truth.

4. Demonstration of Failure Cases

1) Layout generation: When the desired rooms are relatively small and close to each other, our method may not handle such extremely exact layout design well. Some rooms of the generated layouts may be overlapped (See the two pink boxes in Figure F (a)). 2) Texture synthesis: some generated textures are lack of fine-grained details, such as Figure F (b). Currently, we focus more on generating controllable textures while notice less on the photo-realistic de-

tails. We will consider more about the detailed texture generation in the future. 3) 3D Building: due to the imperfect layouts and textures, some rooms of the generated 3D buildings are abrupt, *e.g.,* the green room of our generated building in Figure F (c).

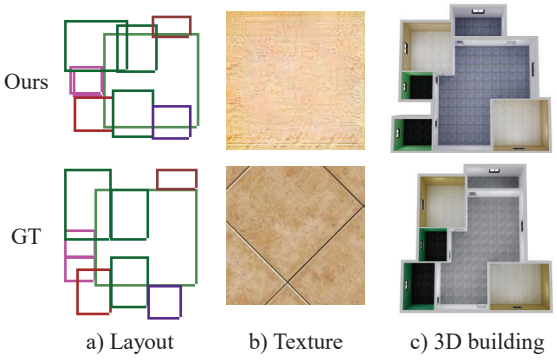


Figure F. Failure cases of a) layouts, b) texture and c) 3D building.

5. Details of Generator G

In this section, we provide more details of the generator G in our proposed LCT-GAN and show the detailed architecture in Table A.

6. More Qualitative Results

In this section, we will provide more qualitative results of our proposed LCT-GAN and baseline methods, which have been mentioned in the paper. From Figure G, the results show that our method is able to produce neater and sharper textures than the baseline methods. Besides, the generated images are more consistent with the given semantic expressions that other baselines.

7. More Qualitative Results of 3D House Plan

In this section, we report more visual results of 2D and 3D house plans corresponding to the given linguistic requirements, and show these results in Figure H.

References

[1] Sebastian Schuster, Ranjay Krishna, Angel Chang, Li Fei-Fei, and Christopher D Manning. Generating semantically precise scene graphs from textual descriptions for improved image retrieval. In *Proceedings of the fourth workshop on vision and language*, pages 70–80, 2015. 1

Module	Module details	Input shape	Output shape
Upsample	2× Upsampling	$(d_1 + d_2 + d_3$, h, w)	$(d_1 + d_2 + d_3$, 2h, 2w)
Conv2d	kernel=(5, 5), stride=(1, 1), padding=(2, 2)	$(d_1 + d_2 + d_3$, 2h, 2w)	(8F, 2h, 2w)
BN+ReLU	–	(8F, 2h, 2w)	(8F, 2h, 2w)
Upsample	2× Upsampling	(8F, 2h, 2w)	(8F, 4h, 4w)
Conv2d	kernel=(5, 5), stride=(1, 1), padding=(2, 2)	(8F, 4h, 4w)	(4F, 4h, 4w)
BN+ReLU	–	(4F, 4h, 4w)	(4F, 4h, 4w)
Upsample	2× Upsampling	(4F, 4h, 4w)	(4F, 8h, 8w)
Conv2d	kernel=(5, 5), stride=(1, 1), padding=(2, 2)	(4F, 8h, 8w)	(2F, 8h, 8w)
BN+ReLU	–	(2F, 8h, 8w)	(2F, 8h, 8w)
Upsample	2× Upsampling	(2F, 8h, 8w)	(2F, 16h, 16w)
Conv2d	kernel=(5, 5), stride=(1, 1), padding=(2, 2)	(2F, 16h, 16w)	(F, 16h, 16w)
BN+ReLU	–	(F, 16h, 16w)	(F, 16h, 16w)
Upsample	2× Upsampling	(F, 16h, 16w)	(F, 32h, 32w)
Conv2d	kernel=(5, 5), stride=(1, 1), padding=(2, 2)	(F, 32h, 32w)	(3, 32h, 32w)
Tanh	–	(3, 32h, 32w)	(3, 32h, 32w)

Table A. Detailed model design of the generator G of our LCT-GAN. "F" refers to the basic dimension of the intermediate features. "h" and "w" denote the height and width of the input, respectively.

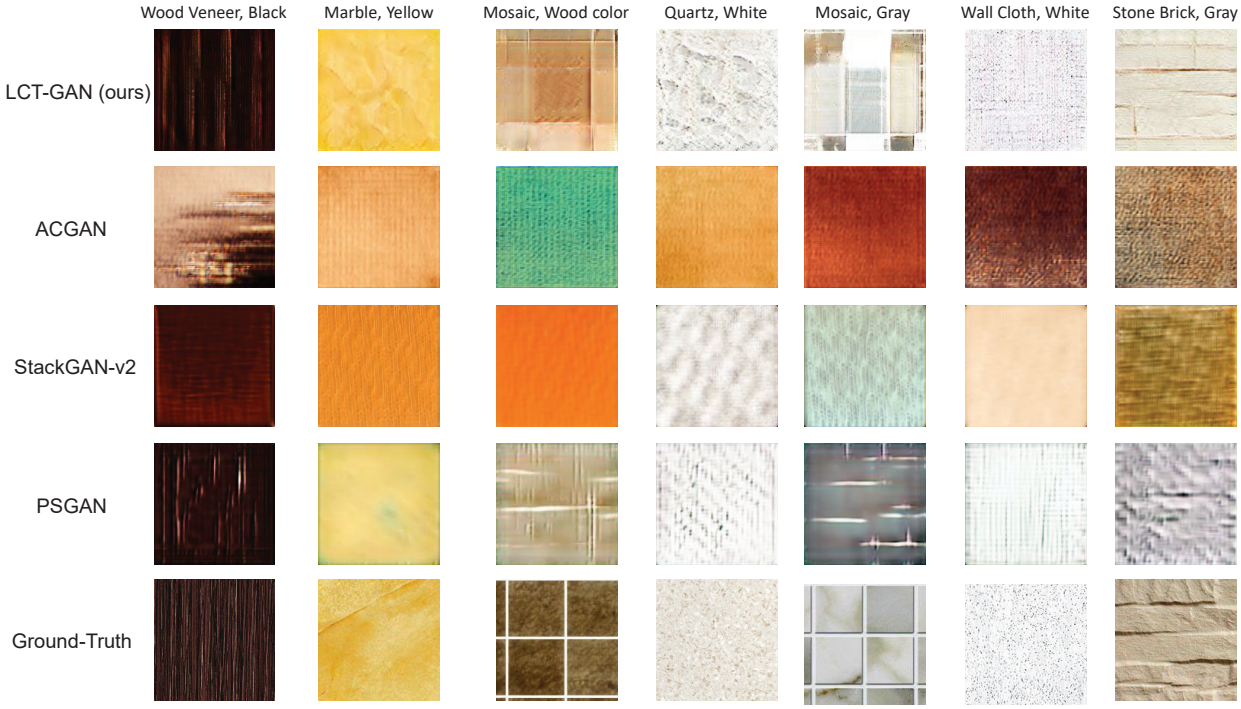


Figure G. More qualitative results of our proposed LCT-GAN and the mentioned baseline methods.

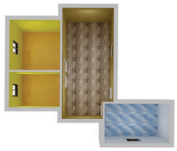
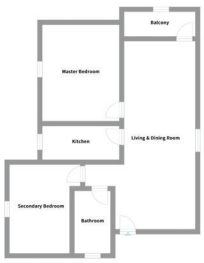
Linguistic Requirements	Ours		Ground-truth	
	2D Floor Plan	3D House Plan	2D Floor Plan	3D House Plan
The building layout contains one washroom, one study, one livingroom, and one bedroom. To be specific, washroom1 has Blue Marble floor, and wall is Wall_Cloth and White. washroom1 is in southeast with 11 square meters. Additionally, study1 has Wood_color Log floor as well as has Yellow Wall_Cloth wall. study1 has 8 squares in west. livingroom1 is in center with 21 square meters. livingroom1 wall is Earth_color Wall_Cloth while uses Black Log for floor. Besides, bedroom1 covers 10 square meters located in northwest. bedroom1 floor is Wood_color Log, and has Orange Pure_Color_Wood wall. livingroom1 is adjacent to washroom1, bedroom1, study1. bedroom1 is next to study1.				
The house has one washroom, one livingroom, one storage, two bedrooms, one kitchen, and one balcony. In practice, washroom1 is in west with 5 square meters. wall of washroom1 is Coating and Yellow, and has White Wood_Veneer floor. Moreover, livingroom1 has Yellow Marble floor as well as wall is Wall_Cloth and Black. livingroom1 is in center with 26 square meters. Besides, storage1 is in northwest with 9 square meters. storage1 has Wood_color Wood_Grain floor while wall is White Wall_Cloth. bedroom1 has 13 squares in southwest. bedroom1 uses Black Log for floor, and wall is White Wall_Cloth. bedroom2 uses Black Log for floor while has White Wall_Cloth wall. bedroom2 has 7 squares in northeast. Moreover, kitchen1 uses White Wood Veneer for floor, and wall is Coating and Yellow. kitchen1 has 5 squares in north. balcony1 has 5 squares in south. balcony1 has Black Wall_Cloth wall as well as has Yellow Marble floor. livingroom1 is adjacent to bedroom1, storage1, bedroom2, washroom1, balcony1, kitchen1. washroom1, balcony1 and bedroom1 are connected. storage1 is next to washroom1, kitchen1. bedroom2 is adjacent to kitchen1.				
The house plan has two bedrooms, one washroom, one balcony, one livingroom, and one kitchen. More specifically, bedroom1 has Pink Wall_Cloth wall, and floor is Wood_Veneer and White. bedroom1 covers 13 square meters located in north. In addition, bedroom2 has 11 squares in southwest. bedroom2 floor is White Wood_Veneer as well as wall is Wall_Cloth and Pink. Moreover, floor of washroom1 is Jade and Blue, and wall is White Wall_Cloth. washroom1 covers 5 square meters located in south. Additionally, balcony1 has 4 squares in northeast. balcony1 uses Wood_color Wood_Grain for floor as well as wall is Wall_Cloth and Earth_color. In addition, livingroom1 covers 28 square meters located in east. floor of livingroom1 is Wood_Grain and Wood_color, and wall is Wall_Cloth and Earth_color. Additionally, kitchen1 has 5 squares in center. kitchen1 has Blue Jade floor as well as wall is White Wall_Cloth. livingroom1 is adjacent to bedroom1, bedroom2, kitchen1, washroom1, balcony1. kitchen1 and bedroom1 are connected. bedroom2 is adjacent to kitchen1, washroom1.				

Figure H. More results of 3D house plans corresponding to the given input linguistic requirements.