LETTER

SCUT-AutoALP: A Diverse Benchmark Dataset for Automatic Architectural Layout Parsing

Yubo LIU[†], **Yangting LAI**[†], **Jianyong CHEN**[††], ***Nonmembers***, **Lingyu LIANG**[††a)], ***Member***, ***and*** **Qiaoming DENG**[†b)], ***Nonmember***

SUMMARY Computer aided design (CAD) technology is widely used for architectural design, but current CAD tools still require high-level design specifications from human. It would be significant to construct an intelligent CAD system allowing automatic architectural layout parsing (AutoALP), which generates candidate designs or predicts architectural attributes without much user intervention. To tackle these problems, many learning-based methods were proposed, and benchmark dataset become one of the essential elements for the data-driven AutoALP. This paper proposes a new dataset called SCUT-AutoALP for multi-paradigm applications. It contains two subsets: 1) Subset-I is for floor plan design containing 300 residential floor plan images with layout, boundary and attribute labels; 2) Subset-II is for urban plan design containing 302 campus plan images with layout, boundary and attribute labels. We analyzed the samples and labels statistically, and evaluated SCUT-AutoALP for different layout parsing tasks of floor plan/urban plan based on conditional generative adversarial networks (cGAN) models. The results verify the effectiveness and indicate the potential applications of SCUT-AutoALP. The dataset is available at https://github.com/designfuturelab702/SCUT-AutoALP-Database-Release.

key words: *floor plan, urban plan, image parsing, GAN*

1. Introduction

Computer aided design (CAD) technology is indispensable for architectural design, but current CAD tools require high-level design specifications from architects. It would be fantastic to construct an intelligent CAD system performing automatic architectural layout parsing (AutoALP), which allows to generate candidate designs or predict architectural attributes without much user intervention. It would not only inspire creation of architects, but also reduce tedious adjustment. Recently, AutoALP has attracted ever-growing research interests [1]–[8].

From a computational perspective, AutoALP has two subtasks. One is for prediction task, whose goal is to predict some architectural attributes [1]–[3]; the other is for generation task, which aims to generate candidate designs for the users [4]–[8], [10].

AutoALP is a challenging problem involved with the formulation of visual representation, generation and prediction for architectural layout. Previous studies regard AutoALP as a pattern computing problem and many methods were introduced from the fields of computer vision, graphics or machine learning [8]–[10]. Most recently, the reviving of neural networks advance the studies by deep learning, which allows to build an end-to-end AutoALP system with trainable representation, generator [4]–[7] or predictor [1]–[3].

Table 1 Representative datasets for AutoALP, where "Num." is for sample numbers, "App." for application, "Pre." for prediction task, "Gen." for layout generation task, "Pub." for publicly released; "FP" is for floor plan, "UP" for urban plan, "LA" for layout labels, "AT" for attribute labels; "#" means it is built by modification or expansion of previous dataset.

Dataset	Num.	App.	Label	Pre.	Gen.	Pub.
2015 [9]	122	FP	LA	-	√	√
2016 [10]	80	UP/FP	LA/AT	-	√	-
2017 [8]	870#	FP	LA	-	√	-
2017 [11]	500	FP	LA	√	√	-
2018 [5]	1.1K#	FP	LA	-	√	-
2018 [6]	200	FP	LA	-	√	-
2018 [7]	115	FP	LA	-	√	-
2018 [3]	#	UP	AT	√	-	√
2019 [4]	>80K	FP	LA	-	√	√
2019 [2]	5K#	FP	AT	√	-	√
2020 [1]	150K	FP	AT	√	-	-
Ours	602	UP/FP	LA/AT	√	√	√

Since most of current learning-based AutoALP models were data-driven, benchmark dataset become one of the essential problems of AutoALP. However, only few benchmark datasets were released publicly and the datasets were built for a specific task or with specific computation constrains, as listed in Table 1.

For urban plan parsing, Feng et al. [10] proposed a optimization-based method with human crowd properties to synthesize the paths and sites in a given input layout domain (with a few floor plan parsing); Zhang et al. [3] measured human perception of a place in a large-scale urban region by diverse properties, where the datasets for training the predictor are from the subsets of MIT places pulse dataset or collected by Tencent Street View service. However, the datasets of these works are either not publicly available [10] or based on previous dataset [3].

For floor plan parsing, most of current studies concentrated on generation task [5]–[11], and only few were for prediction task [1], [2]. Some of these works were based on existing datasets, which may be limited by the labels or applications. For example, CubiCasa5K [2] col-

Manuscript received May 28, 2020.
Manuscript revised August 9, 2020.
Manuscript publicized September 3, 2020.

[†]The authors are with the School of Architecture, South China University of Technology, China.
[††]The authors are with the School of Electronic and Information Engineering, South China University of Technology, China.
a) E-mail: lianglysky@gmail.com
b) E-mail: dengqm@scut.edu.cn

DOI: 10.1587/transinf.2020EDL8076

lected and reviewed 5000 floorplans from a larger dataset of Finnish floorplan images; [5] use the SUNCG dataset, a large database of virtual 3D scenes created by users of the online Planner5d interior design tool, to train the deep convolutional model. Many datasets have not been released publicly yet, which makes the comparison difficult. Recently, Wu et al. [4] constructed a large-scale dataset containing over 80K real floor plans with dense layout anotations, and proposed a two-stage method with deep neural networks to generate floor plans for residential buildings with given boundaries. Kato et al. [1] built a predictor for user preference of real estate properties based on a dataset containing 1.5K samples with 10 classes. Both of these works would greatly advance the studies of AutoALP.

Table 1 indicates that AutoALP is intrinsically a multi-paradigm problem with diverse tasks (prediction/geneartion), labels (layout/attribute) and applications (floor plan/urban plan). Building a dataset under specific paradigm would limit the performance and flexibility of the computational model trained on the dataset, and it would be difficult to make comparison with the models trained by the dataset with specific constraints. Therefore, we argue that AutoALP is a multi-paradigm computational problem, and propose a new diverse benchmark dataset, called SCUT-AutoALP, to achieve multi-paradigm ALP.

SCUT-AutoALP contains two subsets, where Subset-I is for floor plan design containing 300 residential floor plan layouts with boundary, layout and attribute labels (apartment area); Subset-II is for urban plan design containing 302 campus plan layouts of primary school with boundary, layout and attribute labels (like building area and building density), as shown in Fig. 1. With different combinations of samples and labels, SCUT-AutoALP is available for different applications. For example, boundary label (Fig. 1 (b)) and layout label (Fig. 1 (c)) allow interior plan generation as in [4]; samples (Fig. 1 (a)) and layout labels (Fig. 1 (c)) allow image-to-image layout recognition as in [7]. We analyzed the samples and labels statistically, and evaluated SCUT-AutoALP for different image parsing tasks, including boundary-to-layout (B2L) generation, layout recognition and layout generation. The results illustrate the effectiveness and indicate the potential applications of SCUT-AutoALP.

The main contributions of this article can be summarized as follows:

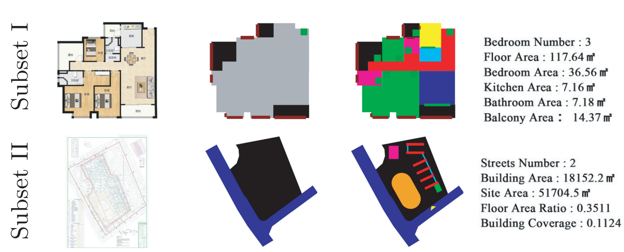
1. **Dataset:** We propose a new benchmark dataset SCUT-AutoALP containing 602 samples and dense labels, which is flexible for multi-paradigm AutoALP with diverse tasks (prediction/geneartion), labels (layout/boundary/attribute) and applications (floor plan/urban plan).
2. **Benchmark Analysis:** We analyze the samples and labels of SCUT-AutoALP statistically, and visualize some properties of SCUT-AutoALP.
3. **Layout Parsing:** Evaluation experiments were conducted for different layout parsing tasks for floor plan or urban plan, including boundary-to-layout (B2L) generation and layout recognition/generation. The results indicate the effectiveness and potential applications of SCUT-AutoALP, and can be used as baselines for the following research.

2. Construction of SCUT-AutoALP

SCUT-AutoALP Dataset contains totally 602 samples with dense labels (annotated by architectural designers), which can be divided into two subsets.

Subset-I is for floor plan design, which contains 300 residential floor plan images. The original samples were collected from "LianJia" website, and that with obvious drawing errors has been eliminated. All the cleaned samples were resized by 1: 100, and the final samples were placed on $20cm \times 20cm$ white background images. For layout parsing, we use different RGB color blocks to denote different regions for boundary labels and layout labels, as shown in Fig. 2 (a). Since all the drawing images were resized with the same scale, each of the region was labelled with the corresponding area for attribute prediction.

Subset-II is for urban plan design, which contains 302 urban plan images of primary school campus. The original samples were collected from the Internet, like websites of architectural design or government. Urban plan design is influenced by many factors, like climate or topography. To reduce the unwanted variances, we selected the samples with similar climatic and topographical conditions. All the cleaned samples were resized by 1: 1800, and the final samples were placed on $24cm \times 24cm$ white background images. For layout parsing, we use different RGB color blocks to denote different regions for boundary label and layout label,

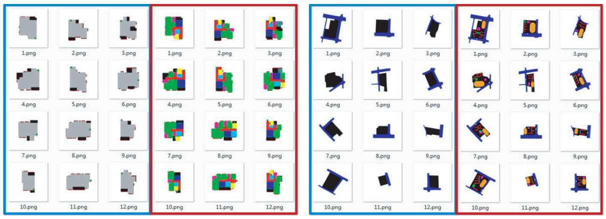


(a) Samples (b) Boundary (c) Layout (d) Attribute

Fig. 1 Illustration of samples, boundary label, layout label and attribute label in SCUT-AutoALP, where subset I is for floor plan and subset II is for urban plan.

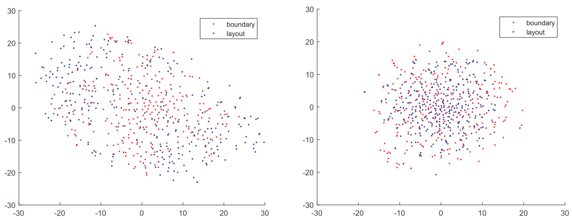
(a) Floor Plan (b) Urban Plan

Fig. 2 Denotation of boundary label and layout label with different RGB color blocks.



(a) Floor Plan (Subset-I) (b) Urban Plan (Subset-II)

Fig. 3 Samples of floor plan and urban plan with different boundary labels (blue box) and layout labels (red box).



(a) Floor Plan (Subset-I) (b) Urban Plan (Subset-II)

Fig. 4 Visualization of boundary and layout labels using t-SNE, which projects the extracted LBP-based features of the labels into a 2D space.

as shown in Fig. 2 (b). For attribute prediction, we annotate each region with building area, and estimate building density and floor area ratio for attribute prediction.

3. Benchmark Analysis

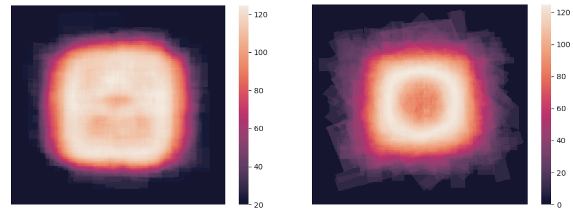
We analyzed the visual variances of different samples and labels in SCUT-AutoALP.

Figure 3 shows boundary labels and layout labels for floor plan and urban plan. It shows that these labels cover diverse visual properties with variances of boundary shape, region size, region location and orientation.

To illustrate the sample variances, we tried to visualize the variances of the label images. Since the visual diversity of different regions in the labels are encoded in the corresponding images with different texture properties, we can obtain the visual feature by local binary patterns (LBP) [14], which is one of the most powerful texture descriptors for textual representation. Then, we projected the extracted high-dimensional LBP features of these labels into a 2D space by t-SNE [15], as shown in Fig. 4. It can be observed that the data of SCUT-AutoALP have good visual diversity for both floor plan and urban plan, and it seems that the visual variances of Subset-I is larger than Subset-II.

For more direct illustration of the variances, we visualized the pixel-wise standard deviation of the layout labels for floor plan and urban plan, as shown in Fig. 5. The layout labels were first normalized, and then standard deviation of each pixel in R, G, B channel is computed, respectively. Larger value of standard deviation would obtain higher intensity value. It can be seen that the labels of floor plan have more pixels with higher standard deviation, which further indicates larger visual variances of Subset-I than Subset-II.

Since each layout sample consists of many smaller color blocks and corresponding attribute labels (as in Fig. 1), we also analyze the distribution of different block types and region area for layout labels, as shown in Table 2 and Fig. 6 respectively. It shows that some kinds of block types (like living room) exist in most of the samples, while some are not (like study room), which is consistent to practical architectural design logic of floor plan or urban plan.

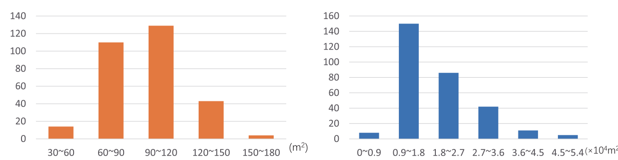


(a) Floor Plan (Subset-I) (b) Urban Plan (Subset-II)

Fig. 5 Visualization of pixel-wise standard deviation of layout labels, where larger value of standard deviation would obtain higher intensity value.

Table 2 Distribution of block types of layout labels (LA) for Subset-I (floor plan) and Subset-II (urban plan).

Subset-I Block of LA	Subset-I Block Num.	Subset-II Block of LA	Subset-II Block Num.
walkway	300	classroom	302
balcony	298	football field	301
kitchen	299	indoor stadium	281
dining room	296	library	127
bedroom	300		
living room	300		
bathroom	300		
study room	48		
indoor garden	35		



(a) RA Dist. in Subset-I (b) RA Dist. in Subset-II

Fig. 6 Distribution (Dist.) of region areas (RA) for region classes listed in Table 2 for Subset-I and Subset-II.

4. Layout Parsing of SCUT-AutoALP

4.1 cGAN-Based Model for Layout Parsing

Based on SCUT-AutoALP, different combination of samples or labels can be used for different layout parsing tasks. Since all the inputs and outputs of these tasks are images, we regard layout parsing as an image-to-image translation problem. We formulate AutoALP tasks using the cGAN-based pix2pix model [12], whose generator has a U-Net structure as shown in Fig. 7.

In the experiments of this paper, all the images were resized to 256×256. For Subset-I and Subset-II, we used 290 pairs for training and the rest for testing, respectively.

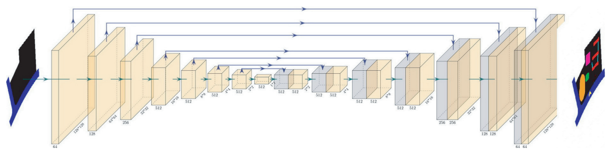
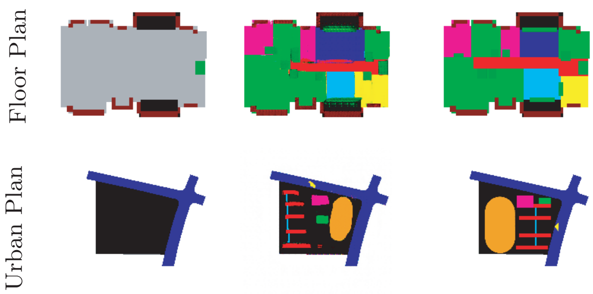


Fig. 7 The U-Net structure of the generator of the cGAN-based pix2pix model for AutoALP, which is based on [12].

(a) Result (b) Result' (c) Ground-truth

Fig. 8 Comparison of result produced by original pix2pix model, result' with horizontal flipping and ground-truth.



(a) Input (b) Output (c) Ground-truth

Fig. 9 Evaluation of boundary-to-layout (B2L) generation for floor plan and urban plan, respectively. More results are shown online [16].

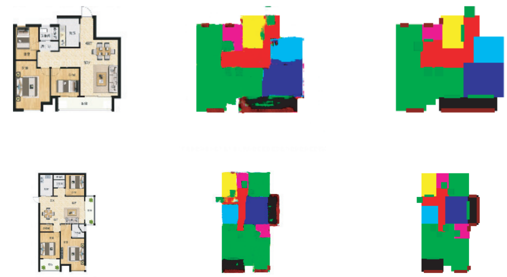
The original pix2pix model is trained with data augmentation that uses cropping+resizing [12]. However, B2L generation and layout generation in AutoALP require global constraint on boundary, directly applying the original pix2pix model for AutoALP would cause region missing or isolation artifacts, as shown in Fig. 8 (a). To tackle this problem, we use horizontal flipping for data augmentation, which obtain better generation results in Fig. 8 (b).

4.2 Evaluation and Analysis

We evaluated SCUT-AutoALP for different layout parsing tasks, including boundary-to-layout (B2L) generation, layout recognition and layout generation.

Figure 9 shows the results of B2L generation for floor/urban plan, which takes layout label as output ground-truth, boundary label as input condition, just as the problem setting of interior plan generation in [4].

Figure 10 shows the results of layout recognition for floor plan, which takes layout label as output ground-truth and sample as input condition, just like the architecture drawing recognition settings in [7] that marking rooms with different colors. Figure 11 shows the results of layout generation for floor plan, which takes sample as output ground-truth and layout label as input condition, just like the architecture drawing generation setting in [7] that generating apartment plans with the input layout blocks. More results can be found in the website of the dataset [16].



(a) Input (b) Output (c) Ground-truth

Fig. 10 Evaluation of layout recognition for floor plan with diverse input condition. More results are shown online [16].

(a) Input (b) Output (c) Ground-truth

Fig. 11 Evaluation of layout generation for floor plan with diverse input condition. More results are shown online [16].

The evaluation illustrates that SCUT-AutoALP is flexible for multi-paradigm layout parsing tasks with different combination of labels and samples. It can be observed that the results produced by our modified pix2pix model are globally consistent to the ground-truth for B2L generation and layout recognition/generation, which indicates that it is suitable to formulate AutoALP as an image-to-image translation problem. The results also indicate that the pix2pix model can be a good baseline for AutoALP, and more sophisticated prior can be integrated into the model to obtain results with better local consistency.

5. Conclusion

This paper proposes a SCUT-AutoALP dataset for multi-paradigm applications, which has totally 602 samples and labels with diverse boundary, layout and attributes. It allows to achieve different architecture layout parsing tasks for floor plan or urban plan. We analyzed the samples and labels statistically, and made evaluation of B2L generation, layout recognition and layout generation for floor plan/urban plan based on the pix2pix model. The results verify the effectiveness and indicate the potential applications of SCUT-AutoALP.

In our future works, SCUT-AutoALP can be further extended in some directions. Firstly, more types, labels and amounts of sample could be added in the dataset. Secondly, more sophisticated models could be used with SCUT-AutoALP, like GNN-based models for image analysis and editing [13].

Acknowledgements

Yubo Liu, Yangting Lai and Qiaoming Deng were supported by National Nature Science Foundation of China (No. 51978268, No. 51978269) and Undergraduate Education Reform Program (No. x2jz-Y1190701); Lingyu Liang and Jianyong Chen were supported by National Nature Science Foundation of China (No. 61872151), Natural Science Foundation of Guangdong Province (No. 2019A1515011045, No. 2020A1515011128), and Fundamental Research Funds for the Central Universities (No. 2019MS023).

References

[1] N. Kato, T. Yamasaki, K. Aizawa, and T. Ohama, "Users' preference prediction of real estate properties based on floor plan analysis," IEICE Transactions on Information and Systems, vol.E103-D, no.2, pp.398–405, 2020.
[2] A. Kalervo, J. Ylioinas, M. HÄikiÖ, A. Karhu, and J. Kannala, "Cubicasa5k: A dataset and an improved multi-task model for floorplan image analysis," Scandinavian Conference on Image Analysis, pp.28–40, Springer, 2019.
[3] F. Zhang, B. Zhou, L. Liu, Y. Liu, H.H. Fung, H. Lin, and C. Ratti, "Measuring human perceptions of a large-scale urban region using machine learning," Landscape and Urban Planning, vol.180, pp.148–160, 2018.
[4] W. Wu, X.-M. Fu, R. Tang, Y. Wang, Y.-H. Qi, and L. Liu, "Data-driven interior plan generation for residential buildings," ACM Trans. Graph., vol.38, no.6, pp.1–12, 2019.
[5] K. Wang, M. Savva, A.X. Chang, and D. Ritchie, "Deep convolutional priors for indoor scene synthesis," ACM Trans. Graph., vol.37, no.4, pp.1–14, 2018.
[6] W. Wu, L. Fan, L. Liu, and P. Wonka, "MIQP-based layout design for building interiors," Computer Graphics Forum, vol.37, no.2, pp.511–521, 2018.
[7] W. Huang and H. Zheng, "Architectural drawings recognition and generation through machine learning," Proc. Computer Aided Design in Architecture, pp.18–20, 2018.
[8] C. Liu, J. Wu, P. Kohli, and Y. Furukawa, "Raster-to-vector: Revisiting floorplan transformation," Proc. ICCV, pp.2195–2203, 2017.
[9] L.-P. de las Heras, O.R. Terrades, S. Robles, and G. Sánchez, "CVC-FP and SGT: A new database for structural floor plan analysis and its groundtruthing tool," IJDAR, vol.18, no.1, pp.15–30, 2015.
[10] T. Feng, L.-F. Yu, S.-K. Yeung, K. Yin, and K. Zhou, "Crowd-driven mid-scale layout design," ACM Trans. Graph., vol.35, no.4, 132, 2016.
[11] S. Dodge, J. Xu, and B. Stenger, "Parsing floor plan images," Proc. MVA, pp.358–361, 2017.
[12] J.-Y. Zhu, T. Park, P. Isola, and A.A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," Proc. ICCV, pp.2223–2232, 2017.
[13] L. Liang, L. Jin, and Y. Xu, "Adaptive GNN for image analysis and editing," Proc. NeurIPS, pp.3638–3649, 2019.
[14] S. Brahnam, L.C. Jain, L. Nanni, and A. Lumini, "Local Binary Patterns: New Variants and Applications," Springer, 2014.
[15] L. Maatenand and G. Hinton, "Visualizing data using t-SNE," J. Mach. Learn. Res., vol.9, pp.2579–2605, 2008.
[16] SCUT-AutoALP Dataset, URL: https://github.com/designfuturelab702/SCUT-AutoALP-Database-Release