SYNTHESIZING STYLE-SIMILAR RESIDENTIAL FACADE FROM SEMANTIC LABELING ACCORDING TO THE USER-PROVIDED EXAMPLE

JIAXIN ZHANG[1, 2]; TOMOHIRO FUKUDA[2]; NOBUYOSHI YABUKI[2]; and YUNQIN LI[1]
[1] *Architecture and design college, Nanchang University, No. 999, Xuefu Avenue, Honggutan New District, Nanchang 330031, China*
[1] *{jiaxin.arch|liyunqin}@ncu.edu.cn, 0000-0002-6330-6723, 0000-0002-1886-0477*
[2] *Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, Osaka 5650871, Japan*
[2] *fukuda.tomohiro.see.eng@osaka-u.ac.jp, 0000-0002-4271-4445, yabuki@see.eng.osaka-u.ac.jp, 0000-0002-2944-4540*

Abstract. Example-guided facade synthesis aims to synthesize realistic facade images from semantic labels drawn by architects and example images of user preferences. The automated synthesis approach allows for the efficient generation of facade solutions that will facilitate effective communication between stakeholders and creative inspiration for architects. This study proposes a conditional generation adversarial network with style consistency to solve the problem of example-guided image synthesis. Specifically, the synthesis model is divided into two stages: first, the domain of the semantic label map is transferred to the domain of the realistic image using the pix2pixHD framework to ensure that the synthesized facade in the intermediate stage can be semantically consistent with the designed facade; Second, we use the Deep Photo Style Transfer (DPST) framework to faithfully move the implied features of the realistic facade image synthesized in the previous step to the domain of the provided example to ensure consistency of style. In summary, the proposed method can constrain the synthesis of new residential facades from the semantic labels and example styles. The synthesized residential facades can be consistent with the example styles provided by the client while matching the semantic labels of the facade created by the designer, producing satisfyingly realistic transitions in various cases.

Keywords. Residential Facades, Style Transfer, Image Synthesis, Generative Adversarial Networks, Building Facade Design

HUMAN-CENTRIC, Proceedings of the 28th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2023, Volume 1, 139-148.

1. Introduction

In architectural design projects, the client often has a preferred building facade style, while the designer must consider the functional and formal requirements of the proposal from a professional perspective. Suppose at the pre-proposal stage, both the user-provided examples and the designer-created sketches with semantic labels can be used as guides to synthesize the building facade images automatically. In that case, it will effectively facilitate the communication efficiency between stakeholders and save the designer's conceptual effort in the draft. Figure 1 illustrates how the proposed methodology can generate a preliminary facade plan to meet the expectations of the architect and the client in their communications.

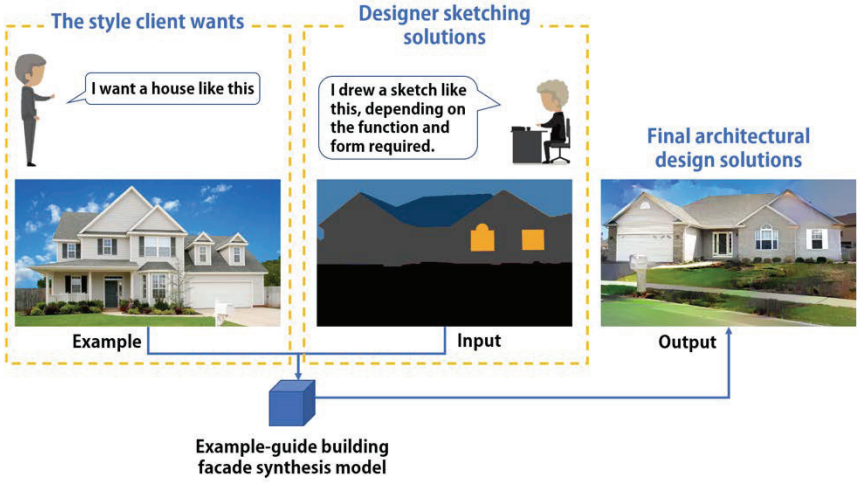

Figure 1. Automatic synthesizing facade images from user-provided examples and designer-created sketches as guides

A priori knowledge-based facade generation method can effectively solve the synthesis problem for limited style samples relying on effective knowledge discovery by designers for specific facade styles (Tang et al., 2019). Recently, data-driven computer vision (CV) methods have proven to be robust in automatically synthesizing building facade images with less reliance on expert know-how guidance (Sun et al., 2022), providing architects with creative solutions. Architects have tried to combine deep learning approaches to the facade synthesis issues. For example, Sun et al. (2022) used pix2pix to generate building facades for historical urban renovation automatically. Zhang et al. (2021) developed a method based on adversarial neural networks to automatically fill missing parts of building facades. Meng (2022) used the StyleGAN2 model to generate images of building facades by analyzing latent spaces without conditional inputs. However, all these methods have challenges in example-guided style-consistent facade image synthesis. The translation of content differences between architectural sketches and reference images is a challenge and may lead to undesirable generation results between unrelated content. The semantic accuracy and transfer faithfulness of each element of the facade is also challenging for synthesizing real-

world images. The synthesized facade appearance should match visual experiences. Furthermore, supervised learning-based approaches suffer from a lack of datasets dedicated to diverse building facade styles.

This study attempts to develop a lightweight, example-guided residential architectural facade synthesis tool to address the above issues. The tool can synthesize an image of a residential facade with the tendency of the client's preferences from semantic labeled maps and exemplar images indicating styles. We use the term "style" in this context to refer to the implied characteristics of a facade image, for instance: from an ethnic-geographical perspective, Western, Eastern, From a historical trend, classical architectural style, and modernist style. In these cases, semantic labeling maps denote the segmentation of elements such as windows, doors, roofs, and facades of residences. We propose a two-stage approach and a dataset. We first use an image translation-based framework to fix the semantics of the facade label drawings drawn by the architects. The image translation model generates an intermediate stage facade that visually matches the appearance of real-world buildings. Secondly, we introduce a style transfer model that faithfully transfers the reference style and is used to transfer the real-world domain facade image from the previous stage to the user-provided case style. Furthermore, a paired residential facade dataset containing building facade images (from different styles, periods, and stories) and corresponding facade semantic labels is proposed.

2. Related works

2.1. BUILDING FACADE GENERATION

Typically, facade design requires considering the location of windows and doors in the plan, user-preferred materials, and harmony with the surrounding environment. As a result, building facade design is a laborious task for architects. Several automated facade design studies have recently been proposed to contribute significantly to the conventional design workflow (Cao et al., 2017), freeing the architects' labor. Established studies include both a priori knowledge-based and data-driven approaches. These works show that mechanical architectural design independent of architects' intuition can significantly contribute to the conventional design workflow. Data-driven neural network-based facade generation methods have yielded promising results, including the fidelity of the generated images and the feasibility of style transfer.

2.2. IMAGE SYNTHESIS VIA GAN

A generative adversarial network (GAN) is a neural network structure consisting of a generator and a discriminator. It learns the training set's feature distribution and generates new images. Many GAN variants have been developed to cover different application scenarios. This paper deals with GAN-based image generation tasks, including image translation and style transfer.

Isola et al. (2018) proposed a conditional GAN framework for image-to-image translation tasks with paired images as supervision. We aim to synthesize photographs of real domains using semantic label maps. The pix2pixHD (Wang et al., 2018) is an upgraded version of pix2pix and is a robust framework for image synthesis and

interactive manipulation for the generation of large-size facades.

The purpose of style transfer is to transfer the style of the source image to the target image or domain. Established studies include single-example-based and holistic sense-based style transfer. Luan et al. (2017) proposed the Deep Photo Style Transfer (DPST) model, which is locally affine in the color space by constraining the transformation from input to output. The model can relocate the facade style of the original domain to the target domain, where the user provides an example.

3. Methods and datasets

3.1. SYNTHESIS OF STYLE-GUIDED BUILDING FACADE IMAGE

Figure 2 shows the overview of the proposed method. The E provided by the user can guide the synthesis of the real-world image domain y. We aim to synthesize a final facade image y from the semantics labeled map x and the exemplar E: $(x, E) \rightarrow y$. The role of x is to anchor the semantics of the synthetic image. The role of E is to provide stylistic constraints for image synthesis, and the output image y must be consistent with the style of the exemplar E. The two particular requirements we face are: (1) While the semantics of the synthetic facade image is consistent with the semantics of the labeled facade image x, it should satisfy the human visual perception of the realistic facade. We use an image translation model (pix2pixHD) to solve this supervised problem to learn the semantic transformation between paired images (semantic label maps and real-world images). Ensures that the semantics of the generated facade drawings of the real-world domain are consistent with the input label map. (2) Given an input labeled map x, the absence of the ground truth of the bootstrap style exemplar $\{E\}$ should not affect the synthesis result. To solve this weakly supervised problem, we introduce a model of DPST. It constrains the local transformation of the input to the output in color space and expresses this constraint as a fully differentiable energy term.

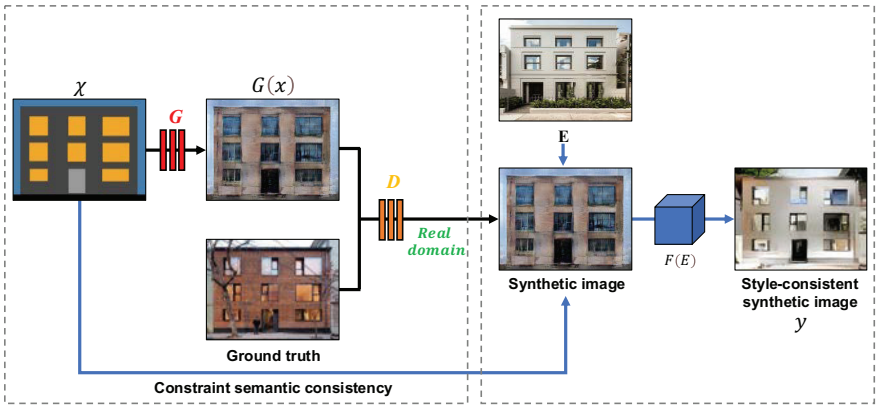

Figure 2. The overview of the proposed method.

The proposed approach builds on the single-scale version of pix2pixHD and DPST, containing (1) a generator G to semantic map x, ground truth image, and a standard discriminator D to distinguish the generated graph as real or fake. (2) A staged synthetic image of the real-world domain, a style-guided example E, and a style-consistent synthetic result. (3) A style-consistent synthesizer DPST, which is used to discover whether the synthetic image and the guide image E are style-compatible. Here, $F(E)$ is an operational procedure that produces a set of synthetic images y that are stylistically consistent with the real-world domain images. Our objective function contains two losses: a standard adversarial loss for synthesizing semantically consistent images of the real domain and an adversarial style consistency loss.

3.2. RESIDENTIAL FACADE DATASET FOR SEMANTICALLY CONSISTENT SYNTHESIS

To synthesize semantically consistent real-world images of residential facades based on image translation, we built a dataset of paired semantically labeled data with real-world residential facade images, which we call Residential Facade for Semantically Consistent Synthesis (RFSCS). RFSCS has 612 residential facades from different countries and eras with corresponding semantic labels. The semantic labels include walls, roofs, windows, the sky, greenery, and road surface in the foreground. Figure 3 shows examples of the paired semantic labels and building facades of RFSCS. Making paired datasets allows designers to draw facade semantic maps that correspond well in abstraction to real-world facade images. The unsupervised approach does not require paired data and is suitable for image transformation between objects where sample correspondence is complex. In contrast to unsupervised learning, supervised learning image translation models trained with RFSCS can accurately and faithfully synthesize the facade elements designed by the architects in line with the semantics.

Figure 3. Paired residential facade dataset for semantically consistent synthesis.

4. Experimental results

4.1. SYNTHESIS OF REAL-WORLD DOMAINS OF BUILDING FACADE BASED ON IMAGE TRANSLATION

The training set consists of 612 paired sets of building facades, and we resize all collected images to 1000 × 500. We employ two image processing methods based on the original dataset to extend the number of datasets. As shown in Figure 4, the data augmentation methods are horizontal flip and 6◦ counterclockwise rotation. The final extended dataset has 1836 pairs of images, of which 1560 pairs are set as the training set and the remaining 276 pairs (only 92 pairs for the original data and the remaining 184 pairs as the enhanced data) as the test set. The training was performed on two Nvidia Geforce 1080 Ti Graphics Cards with batch size set to 4 and total epoch to 300. During the training process, we recorded the loss values for the generator and the discriminator. Figure 5 shows the losses for both models. Since the two neural networks compete with each other, the loss of the discriminator increases as the loss of the generator decreases. The performance of the entire training model gradually improves as the generator and discriminator ebb and flow.

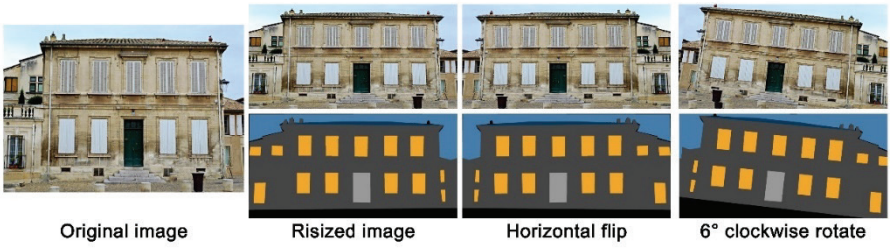

Figure 4. Data processing and augmentation.

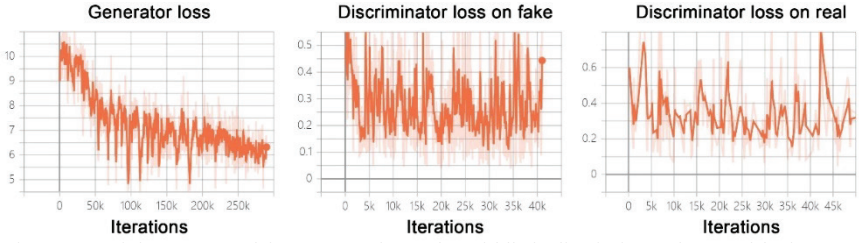

Figure 5. Training Loss. Left is generator loss. The middle is discriminator loss on fake images. Right is discriminator loss on real images.

We used the designer's manual drawings of facade labels, including walls, windows, doors, roads, greenery, and sky, as input for our testing. We experimented with flat roofs, sloped roofs, single-story villas, and multi-storey apartments. The characteristics of residential facades include having similar components and structures, their relatively fixed semantics, and a wide variety of styles. The simple semantic composition of residence facilitates experts to draw semantic labels, while non-experts

can also try various preferred styles. In generating real-world domains, we take the residential building as the main object of study.

Figure 6 shows the building facades generated by the pre-trained model pix2pixHD, and the semantics of each facade of the generated results are consistent with the appearance of real-world facades. The process focuses on faithfully transferring the semantic labels of facades to the domain of real-world facades with the help of an image translation model, locating the latent space of residential facades as photorealistic as possible.

Figure 6. Building facades generated by pix2pixHD.

4.2. SYNTHESIS RESULTS OF STYLE-SIMILAR FACADES BASED ON THE PROVIDED EXAMPLES

Figure 7 shows the results of the generated synthetic facades in four different styles. A semantic labeled map drawn by the designer is used as the scheme of the facade. Four examples provided by the client are used as the target style domains. The synthesis process is based on the initial design of a small villa with cases of western and eastern,

as well as wood and brick walls. As seen from the results in Figure 7, the color transformation with its same element semantics is localized, and the proposed method can handle context-sensitive color changes. The proposed method applies the exact semantic color mapping to match the color statistics between the input and style-guided images. As a result, the synthesized facade can faithfully respond to the color and material information of the provided example.

Figure 7. Facade synthesis results from semantic labeling and provided examples.

We also try to reveal the image translation process from semantic to real-world domain generation. Figure 8 shows the semantic labeling from the input facade to the target domain facade and then synthesizing a stylistically consistent facade image based on the provided examples. The results show that the first stage of real-world domain facade synthesis affects the final synthesis. When semantic synthesis can be generated with high-quality and seamless borders, the photorealistic facade images can

be synthesized according to the style provided in the bootstrap example.

Figure 8. Synthesize stylistically consistent facades using semantic labels and provided examples.

5. Discussion and Conclusion

The proposed method synthesizes facades that meet the designer's requirements for facade semantics, size, and form. Simultaneously, the synthesized facades can also meet the user's requirements for the solution style. The proposed model uses a multi-scale architecture that can synthesise high-resolution images (up to 2048 × 1024 resolution). Our contributions can be summarized in the following three points. (1) The proposed workflow can successfully synthesize facades consistent with the style of the provided examples, with minimal impact on the faithfulness of the transformation. (2) The proposed method solves the problem of the content difference between the input image and the reference image. For example, in the synthesis of the building facade, the result of the synthesis of each element is consistent with people's empirical judgments of the real world. The strategy we adopt is to customize a dataset of facade image translations for supervised learning. The semantic labels of the facades are matched to real-world facades to minimize the chance of inaccurate transfers. (3) When the transfer occurs between semantically equivalent sub-regions, the semantic labels of the input and stylized images are incorporated into the transfer procedure. Besides, the mapping is nearly uniform as the transfer occurs in each sub-region, which helps preserve the richness of synthetic styles.

While the merits are appreciated, several limitations deserve further study.

Synthesizing non-orthogonal projections of building facades can be unsatisfactory because of the limited number of angled datasets. In terms of metrics for evaluating the quality of synthesized facade images, common metrics are not always indicative of the perceived realism of an image, and a human evaluation may provide a more comprehensive understanding of the synthesized image quality. We have demonstrated the feasibility of synthesizing new facades by changing the semantic labels of the old facades. This application will hopefully be used in building facade renewal, where architects redesign the semantic labels of old facades, non-experts provide examples of interest, and the framework will generate new facade solutions that meet the stakeholders' requirements. In addition, the recent excellent performance of the Diffusion Model on image translation (Saharia et al., 2022) will also inspire this work. In future work, we will try to adopt the Diffusion Model framework to realize the example-guided facade style synthesis.

References

Cao, J., Metzmacher, H., O'Donnell, J., Frisch, J., Bazjanac, V., Kobbelt, L., & van Treeck, C. (2017). Facade geometry generation from low-resolution aerial photographs for building energy modeling. *Building and Environment*, 123, 601–624. https://doi.org/10.1016/j.buildenv.2017.07.018

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2018). Image-to-Image Translation with Conditional Adversarial Networks (arXiv:1611.07004). arXiv. https://doi.org/10.48550/arXiv.1611.07004.

Luan, F., Paris, S., Shechtman, E., & Bala, K. (2017). Deep Photo Style Transfer. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6997–7005. https://doi.org/10.1109/CVPR.2017.740

Meng, S. (2022). Exploring in the Latent Space of Design: A Method of Plausible Building Facades Images Generation, Properties Control and Model Explanation Base on StyleGAN2. In P. F. Yuan, H. Chai, C. Yan, & N. Leach (Eds.), *Proceedings of the 2021 DigitalFUTURES (pp. 55–68). Springer Singapore.* https://doi.org/10.1007/978-981-16-5983-6_6

Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., & Norouzi, M. (2022). Palette: Image-to-Image Diffusion Models. *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings*, 1–10. https://doi.org/10.1145/3528233.3530757

Sun, C., Zhou, Y., & Han, Y. (2022). Automatic generation of architecture facade for historical urban renovation using generative adversarial network. *Building and Environment*, 212, 108781. https://doi.org/10.1016/j.buildenv.2022.108781

Tang, P., Wang, X., & Shi, X. (2019). Generative design method of the facade of traditional architecture and settlement based on knowledge discovery and digital generation: A case study of Gunanjie Street in China. *International Journal of Architectural Heritage*, 13(5), 679–690. https://doi.org/10.1080/15583058.2018.1463415

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., & Catanzaro, B. (2018). High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8798–8807. https://doi.org/10.1109/CVPR.2018.00917

Zhang, J., Fukuda, T., & Yabuki, N. (2021). Automatic Object Removal With Obstructed Facades Completion Using Semantic Segmentation and Generative Adversarial Inpainting. *IEEE Access*, 9, 117486–117495. https://doi.org/10.1109/ACCESS.2021.3106124