RealFill: Reference-Driven Generation for Authentic Image Completion

Abstract

Recent advances in generative imagery have brought forth outpainting and inpainting models that can produce high-quality, plausible image content in unknown regions. However, the content these models hallucinate is necessarily inauthentic, since they are unaware of the true scene.RealFill is a generative inpainting model that is personalized using only a few reference images of a scene. These reference images do not have to be aligned with the target image, and can be taken with drastically varying viewpoints, lighting conditions, camera apertures, or image styles. Once personalized, RealFill is able to complete a target image with visually compelling contents that are faithful to the original scene.

Keywords: Image Completion, Diffusion Model

1 Introduction

Photographs capture ephemeral and invaluable experiences in our lives, but can sometimes fail to do these memories justice. In many cases, no single shot may have captured the perfect angle, framing, timing, or composition, and unfortunately, just as the experiences themselves cannot be revisited, these elements of the captured images are also unalterable. We show one such example in Fig.2: imagine having taken a nearly perfect photo of your daughter dancing on stage, but her unique and intricate crown is partially cut out of the frame. Of course, there are many other pictures from the performance that showcase her crown, but they all fail to capture that precise special moment: her pose mid-dance, her facial expression, and the perfect lighting. Given this collection of imperfect photos, you can certainly imagine the missing parts of this perfect shot, but actually creating a complete, shareable version of this image is much harder.

In this work. Given a few reference images (up to five) and one target image that captures roughly the same scene (but in a different arrangement or appearance), we aim to fill missing regions of the target image with high-quality image content that is faithful to the originally captured scene. Note that for the sake of practical benefit, we focus particularly on the more challenging, unconstrained setting in which the target and reference images may have very different viewpoints, environmental conditions, camera apertures, image styles, or even moving objects.

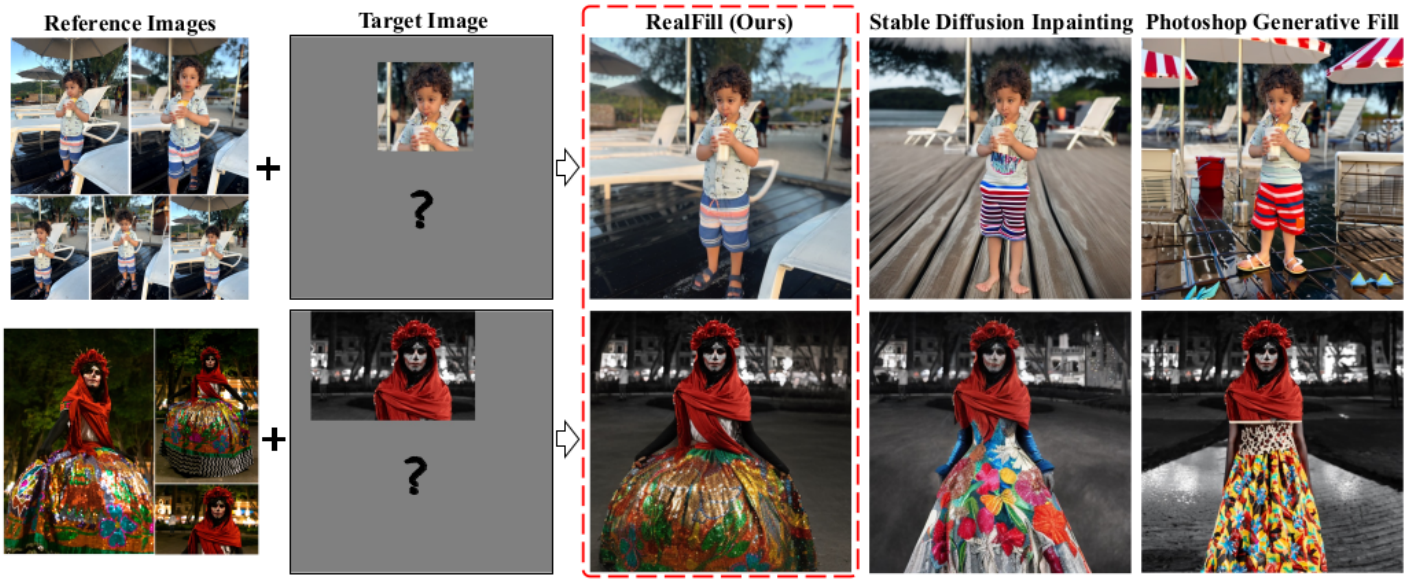

Figure 1. Given a few reference images that roughly capture the same scene, and a target image with a missing region, RealFill is able to complete the target image with image content that is faithful to the true scene. In contrast, standard prompt-based inpainting methods hallucinate plausible but inauthentic content due to their lack of knowledge of the original scene.

2 Method

2.1 Overview

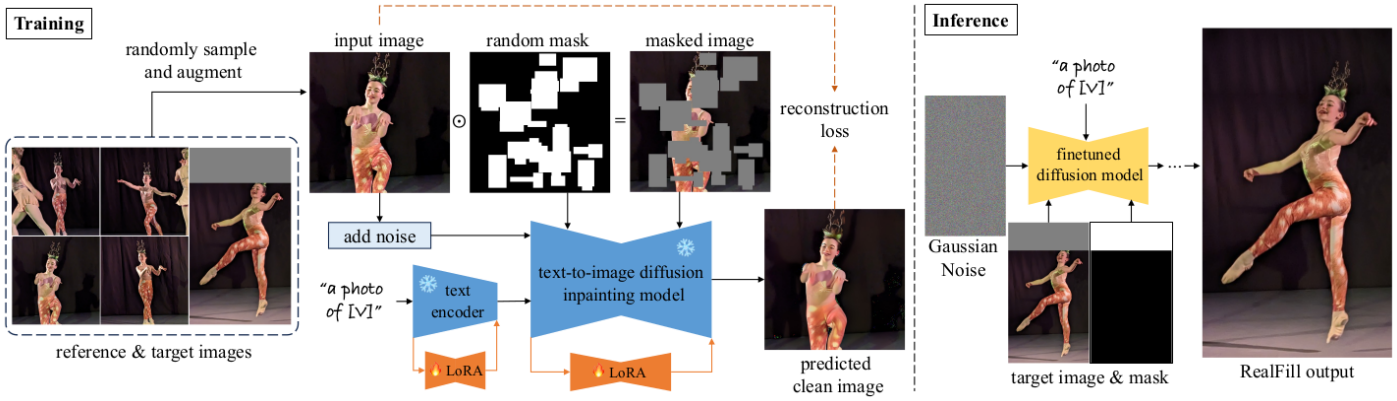

Figure 2. Overview of the method. Training and inference pipelines of RealFill.

RealFill's inputs are a target image to be filled and a few reference images of the same scene. We first finetune LoRA weights of a pretrained inpainting diffusion model on the reference and target images (with random patches masked out). Then, we use the adapted model to fill the desired region of the target image, resulting in a faithful, high-quality output. For example, the girl's crown is recovered in the target image, despite the girl being in very different poses in the reference images.

3 Implementation details

The generation of 3D cities and the completion of building facades are critical tasks in urban planning, architectural design, and digital content creation. With the rapid advancement of generative models, particularly Diffusion Models, there is a growing opportunity to address these challenges with high-quality, detailed, and realistic outputs. Diffusion Models, which operate by iteratively denoising data, have demonstrated remarkable success in image generation and are now being extended to 3D domains, offering a promising approach for generating complex urban structures and completing missing architectural details.

In the context of building facade completion, the goal is to reconstruct missing or damaged parts of a building's exterior while maintaining architectural consistency and aesthetic appeal. Traditional methods often struggle with generating plausible and diverse completions, especially when dealing with complex geometries and textures. Diffusion Models, however, excel at capturing fine-grained details and can be conditioned on partial inputs to generate coherent and realistic completions.

This work explores the application of Diffusion Models to building facade completion. We discuss the methodologies for training and inference.

3.1 Comparing with the released source codes

There is no official sourse code, but unofficial one. Source code: https://github.com/thuanz123/realfill

3.2 Experiment setup

We follow the paper setting.

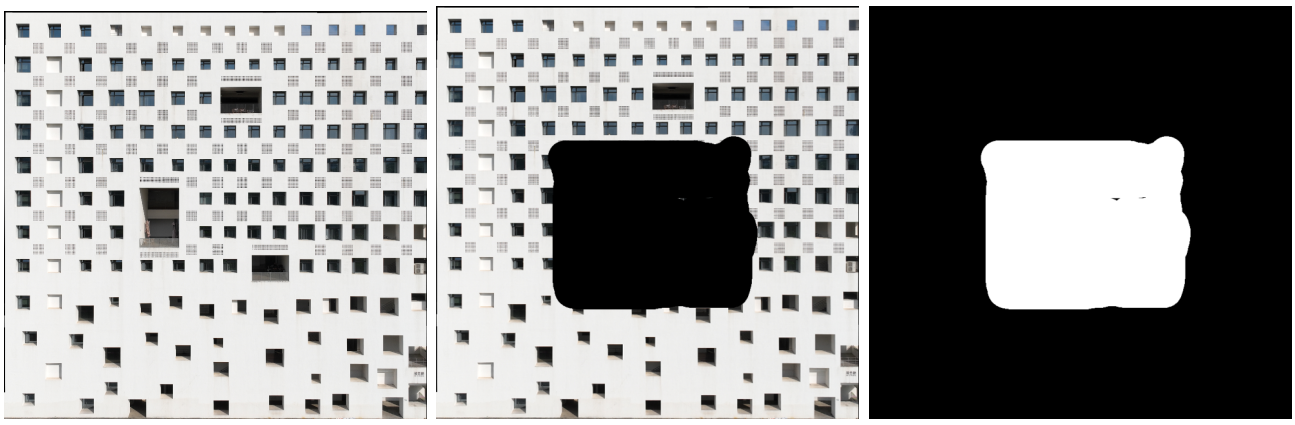

Figure 3. Left: Ground truth image, middle: target image, right: mask image

As Fig.3 shows, the building facade has many repeative patterns, it's difficult to traditional methods to fix this problem.By leveraging the capabilities of Diffusion Models(RealFill), This work explores the application of RealFill to building facade completion.

For the reference images, we use the image taken by DJI, as Fig.4 shows.

Figure 4. Reference images

4 Results and analysis

We use these reference images to finetiue the model with 1k-epoch and 2k-epoch as shows in Fig. 5 and Fig. 6 respectively.

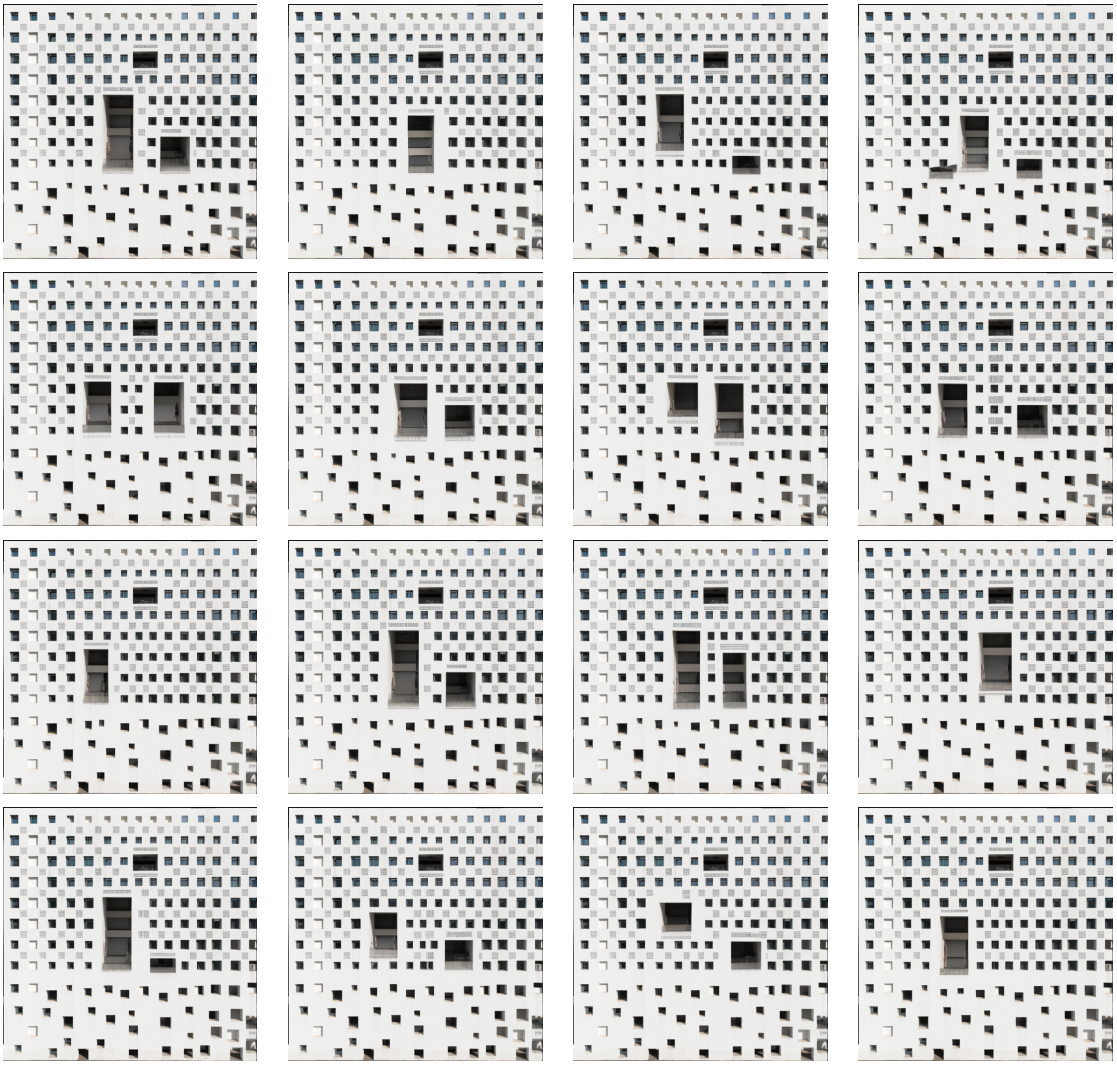

Figure 5. 1K-eopch Result

As the result shows, this inpainting model inpaint well. But it don't learn the layout of building facade of reference images.

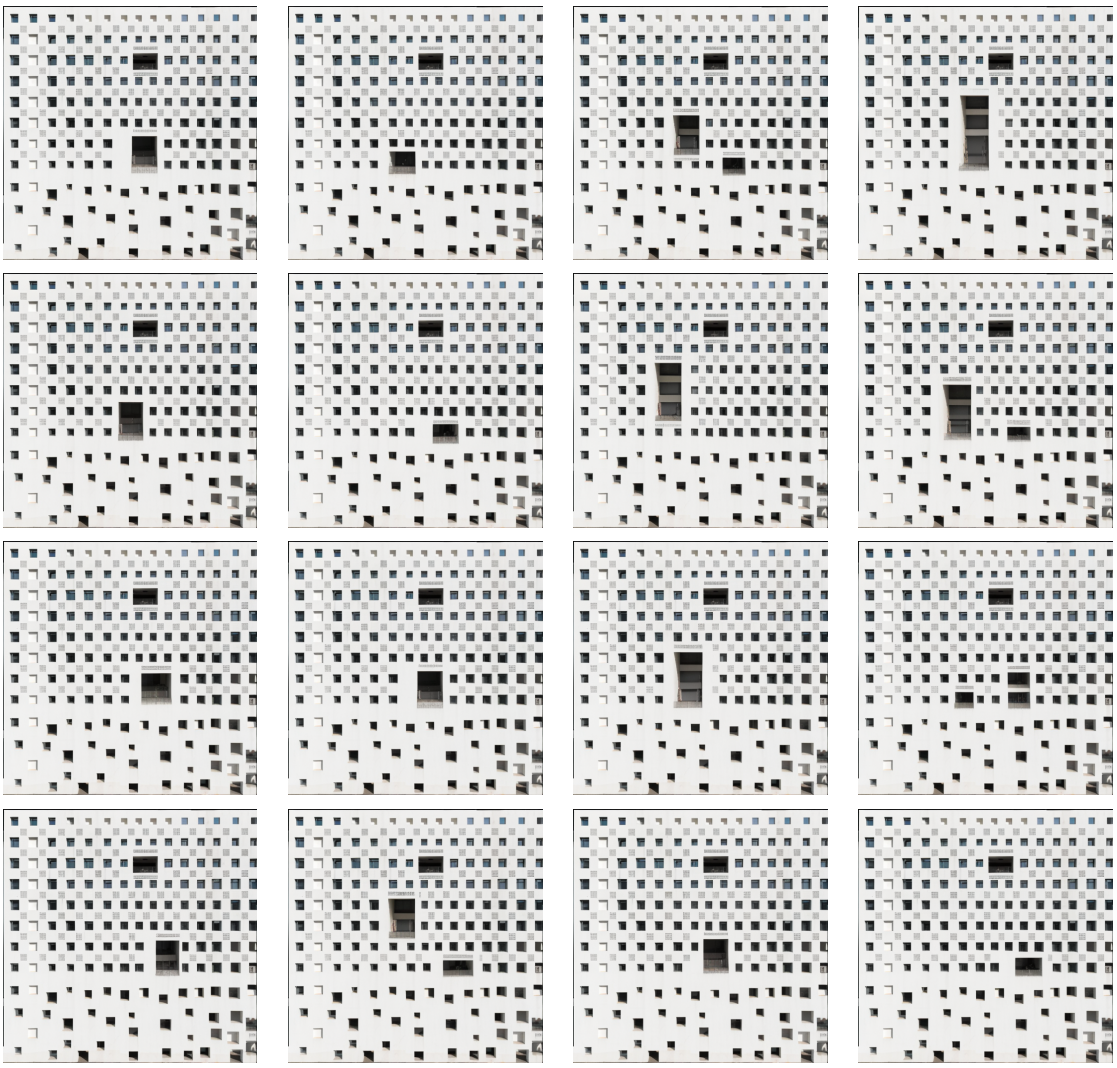

Figure 6. 2K-eopch Result

5 Conclusion and future work

RealFill produces high-quality image completions that are faithful to the content in the reference images, even when there are large differ- ences between reference and target images such as viewpoint, aperture, lighting, image style, and object pose.This work explores the application of Diffusion Models to building facade completion.

In the future, we aim to solve the 3D city generation.The generation of 3D cities and the completion of building facades are critical tasks in urban planning, architectural design, and digital content creation. With the rapid advancement of generative models, particularly Diffusion Models, there is a growing opportunity to address these challenges with high-quality, detailed, and realistic outputs. Diffusion Models, which operate by iteratively denoising data, have demonstrated remarkable success in image generation and are now being extended to 3D domains, offering a promising approach for generating complex urban structures and completing missing architectural details.