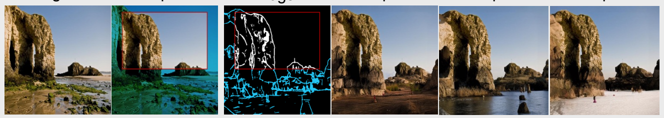
A Structure-Guided Diffusion Model for Large-Hole Image Completion

Daichi Horita[1], Jiaolong Yang[2], Dong Chen[2], Yuki Koyama[3], Kiyoharu Aizawa[1] , Nicu Sebe[4]
[1]The University of Tokyo, [2]Microsoft Research Asia, [3]AIST, [4]University of Trento

Introduction

Objective

- We aim to enable **an edge-guided large-hole image completion** with **diffusion models (DMs)**.



Motivation

- Large-hole completion remains challenging due to **limited structural information**.
- Current limitations of previous methods
  - GAN methods often fail to generate a rational structure.
  - DM methods tends to produce irrelevant contents

Approach

- We address this problem by **integrating explicit structure guidance as edges into diffusion-based image completion** forming **our structure-guided diffusion model (SGDM)**.

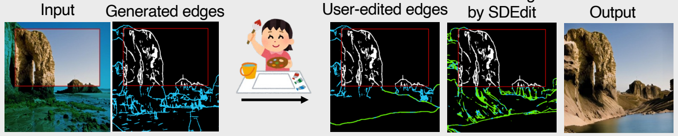
Preliminaries

- A forward step of DDPM is defined as:
  $q(x_t|x_0) = \mathcal{N}\left(x_t|\sqrt{\alpha_t}x_0, (1-\alpha_t)\mathbf{I}\right)$, α_t: pre-defined noise scale.
- Given a Gaussian noise $z \sim \mathcal{N}(z;\mu,\Sigma)$, Tweedie's formula, which is known as *optimal Bayesian denoising (OBd)*, perform a denoising in a single step: $\mathbb{E}[\mu|z] = z + \Sigma\nabla_z \log p(z)$.
- A single-step denoising operation of DDPM can be defined as:

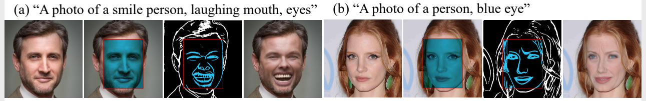
$$\mathrm{F}(x_t) := \hat{x}_0^t = \frac{x_t + (1-\alpha_t)\nabla_{x_t}\log p(x_t)}{\sqrt{\alpha_t}}.$$

Applications

- Sketch-guided image completion



- Language-guided image completion

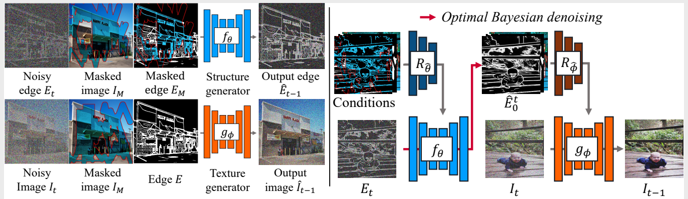


Structure-Guided Diffusion Model

- SGDM consists of **two cascaded diffusion models**. ***structure generator*** f_θ and ***texture generator*** g_ϕ.
- For conditions such as a hole and edges, we employ ControlNet R.

Training

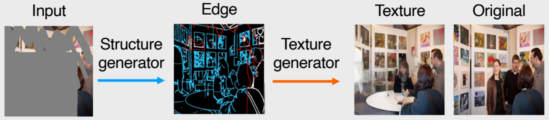
- Individual training
  - Both generators are trained independently.
- Joint fine-tuning after the individual training
  - We propose **a novel joint training strategy using *Tweedie's formula*** to enable an end-to-end training.
    - *Tweedie's formula denoises a noisy edges into noiseless edges using the output of the structure generator.*



Individual Training
Joint Training

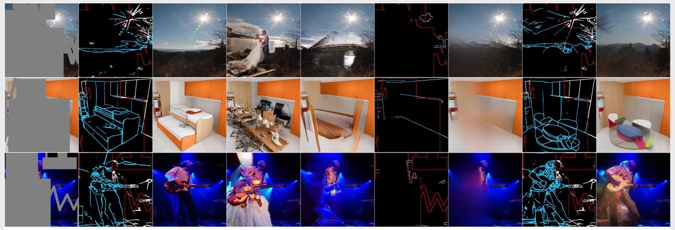
Inference

- Given an input image with missing regions, SGDM generates edges and then textures sequentially



Experiments

Qualitative comparison



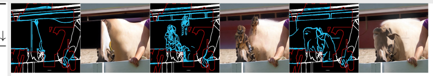
Input | Original edges | Original image | Stable diffusion | MAT | ZITS | SGDM (Ours)

Quantitative comparison

Method	Modeling	Places (512 × 512)						CelebA-HQ (512 × 512)					
		Small mask			Large mask			Small mask			Large mask		
		FID↓	P-IDS↑	U-IDS↑	FID↓	P-IDS↑	U-IDS↑	FID↓	P-IDS↑	U-IDS↑	FID↓	P-IDS↑	U-IDS↑
SGDM (ours)	DM	3.85	25.54	38.53	6.96	18.12	31.78	2.58	22.01	33.56	4.72	13.99	24.97
Stable Diffusion [2]	DM	5.36	16.32	32.05	7.21	15.34	30.80	-	-	-	-	-	-
LDM [40]	DM	5.64	13.42	30.66	8.74	11.7	27.00	-	-	-	-	-	-
MAT [19]	GAN	4.10	25.56	37.73	7.11	18.40	32.46	2.81	19.24	31.33	5.04	11.42	24.13

Ablation study

Places		Large mask	
Model	Samples	FID ↓	LPIPS ↓
(a) Indiv. only	25M	32.28	0.188
(b) + Joint w/o OBd	0.1M	28.68	0.175
(c) + Joint w OBd	0.1M	27.47	0.170
	1M	27.81	0.168



(a) Indivual training
(b) + Joint w/o optimal Bayesian denosinig
(c) + Joint with optimal Bayesian denosinig (Full)

Conclusion

- We have presented a structure-guided diffusion model (SGDM), which uses structural guidance in image completion.
- We have proposed a novel training strategy to enable effective end-to-end training.
- Incorporating structural guidance has not only improved the visual quality but also enabled user-guided image editing.

Project page can be available!!