Filter–Guided Diffusion for Controllable Image Generation

*Zeqi Gu **   *Ethan Yang **   *Abe Davis*

SIGGRAPH ASIA 2024 TOKYO

Introduction

➢ In recent years, different text-conditioned diffusion models have became popular for performing Image-to-Image Translation tasks

➢ These I2I models are mostly white-box methods, meaning they depend heavily on internal architecture of the image, which makes them more time-consuming and memory-inefficient

➢ To address this limitation, FGD introduces a novel black-box approach, i.e architecture independent, which makes the task of translating between images lightweight and efficient making it more adaptable for real time applications
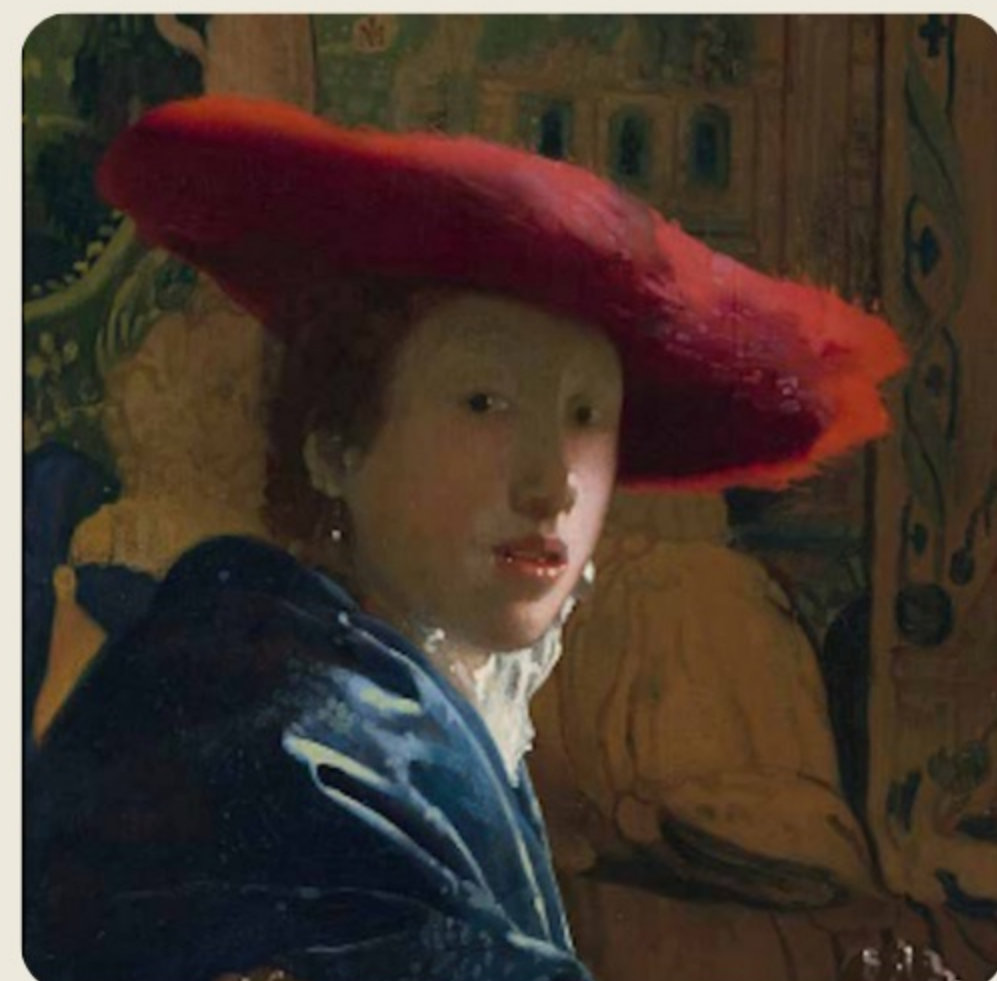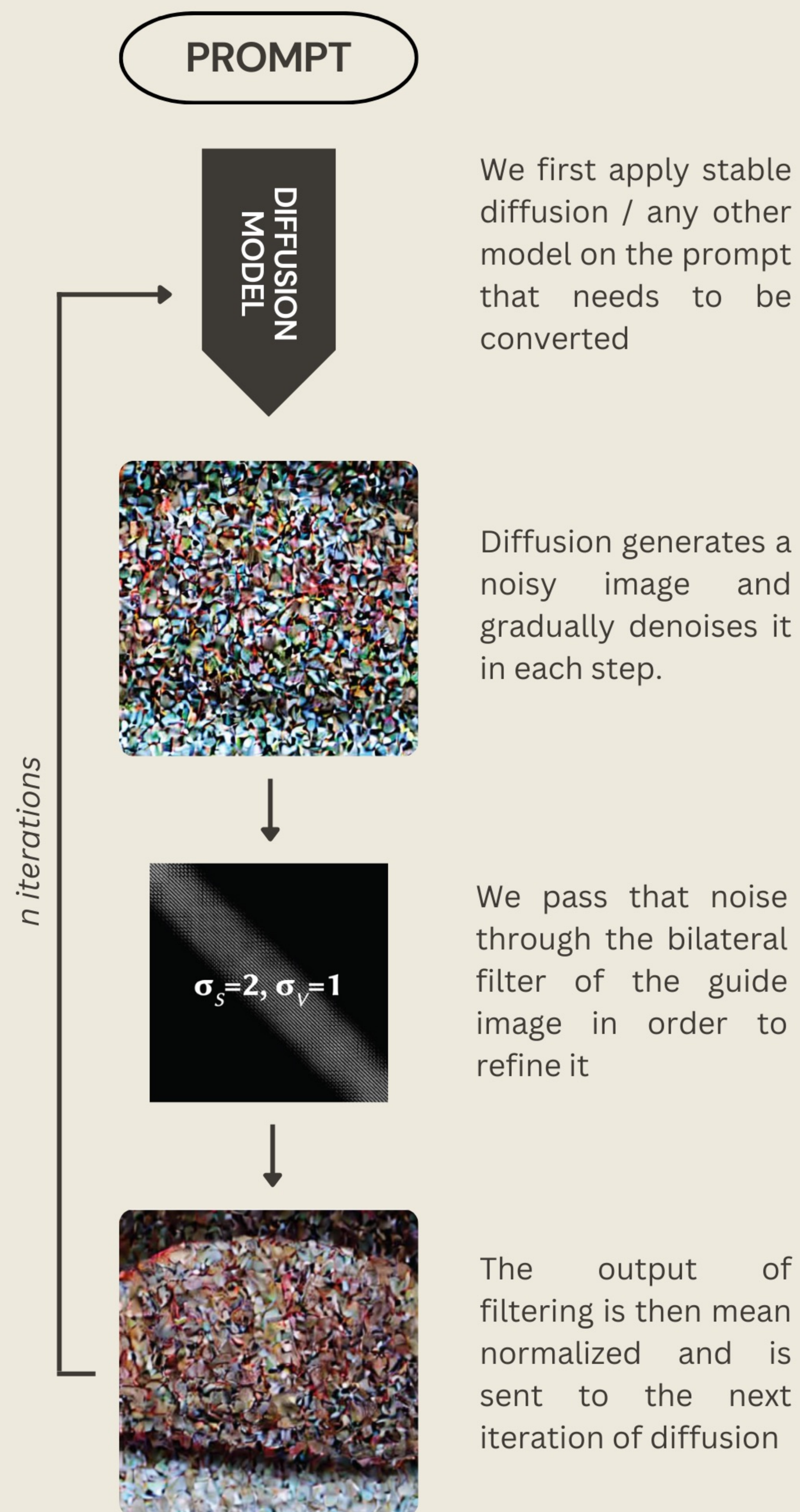

a painting of a dog in red hat

Concept

➢ The key idea for the black box approach here is to enhance the traditional diffusion model steps by introducing various filtering steps related to our guide image in order to allow more control over specific features such as structure and style

➢ While deciding upon the filter to use, we came across vairous old image processing techniques which involved filters like laplacian blending and bilateral filtering

➢ Below is the mathematical update equation for the idea that is described above -
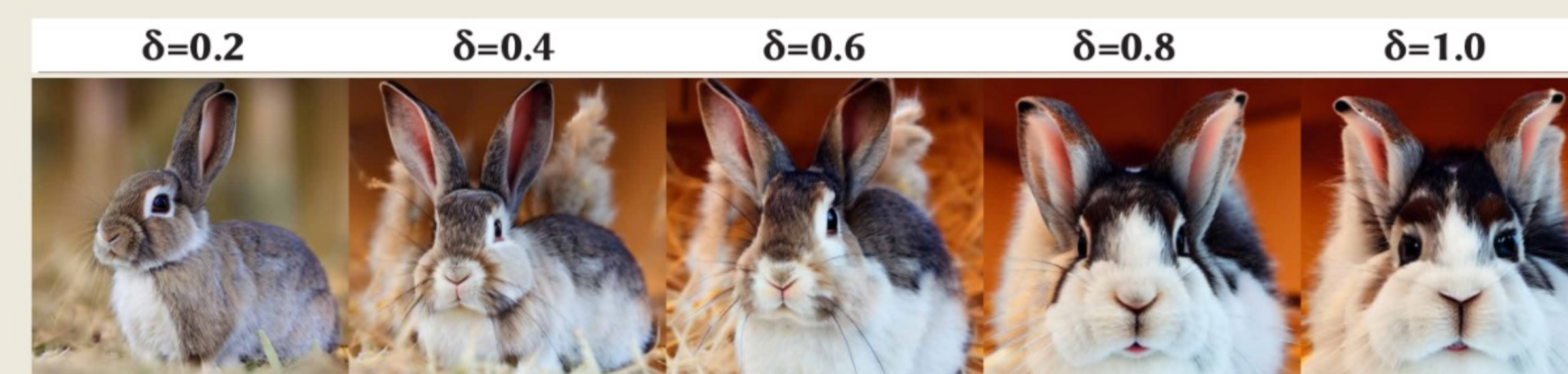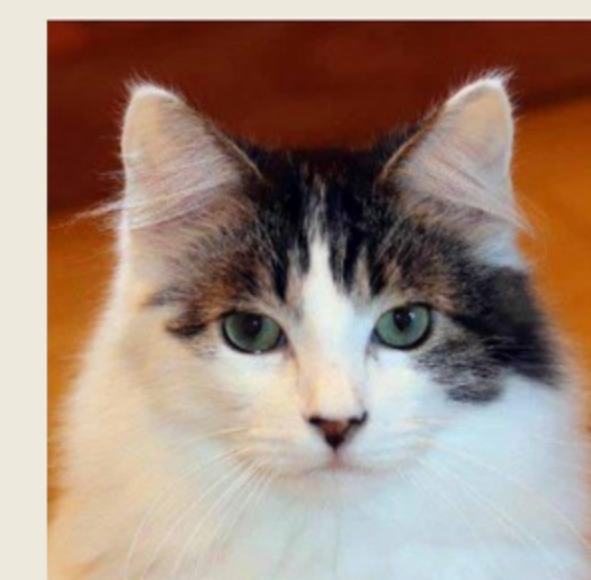
$$x_{t-1} = x'_{t-1} + \delta \times \left(f(x_g) - f(x'_{t-1}) \right)$$

Methodology

PROMPT

DIFFUSION MODEL

n iterations

We first apply stable diffusion / any other model on the prompt that needs to be converted

Diffusion generates a noisy image and gradually denoises it in each step.

$\sigma_s = 2, \sigma_v = 1$

We pass that noise through the bilateral filter of the guide image in order to refine it

The output of filtering is then mean normalized and is sent to the next iteration of diffusion

Results

➢ Variation of output image with Guidance Strength -

a photo of a rabbit

$\delta=0.2$	$\delta=0.4$	$\delta=0.6$	$\delta=0.8$	$\delta=1.0$

➢ Variation of output image with Diffusion iterations and comparison with other methods -

a photo of a desert

Threshold=50	40	30	20	10
PnP				
ILVR				
Ours				