

DS-Fusion: Artistic Typography via Discriminated and Stylized Diffusion

Maham Tanveer, Yizhi Wang, Ali Mahdavi-Amiri, Hao Zhang
Simon Fraser University

2023/09/13 陸品潔



Contents



01

Introduction



02

Related Work



03

Method



04

Results



The background is a light cream color decorated with various watercolor-style elements. There are large, soft-edged shapes in shades of blue, pink, purple, and yellow. Smaller, more defined shapes include a cluster of red dots in the top right, a group of blue diagonal strokes in the top left, and several thin, horizontal purple strokes near the top center. The overall aesthetic is artistic and hand-drawn.

01

Introduction

Typography

- The art and technique of arranging type to make written language legible, readable, and appealing when displayed.
- Conflicting goals:
 1. Artistic stylization
 2. Legibility

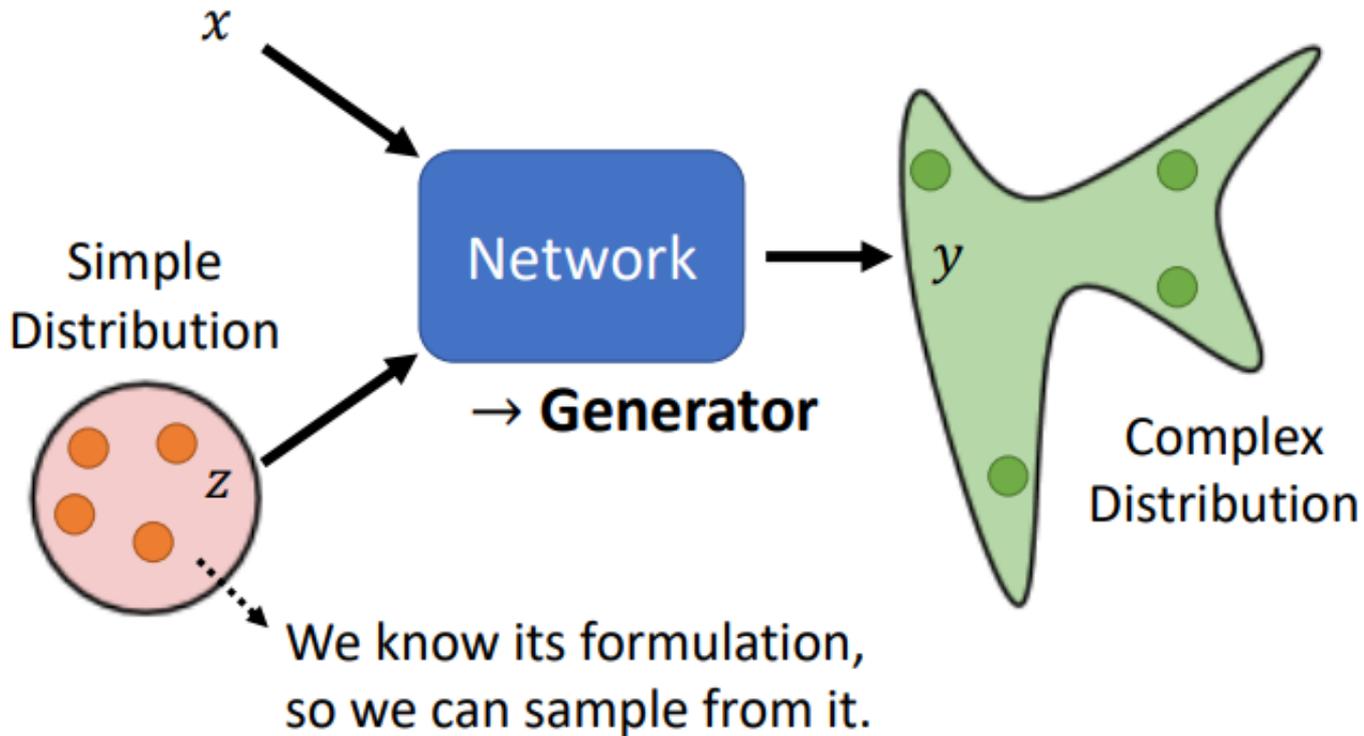


The background is a white canvas decorated with various watercolor-style elements. There are large, soft-edged washes of color in shades of blue, pink, purple, and yellow. Interspersed among these are smaller, more defined patterns: clusters of small dots in red and orange, groups of parallel lines in blue and purple, and some abstract, brush-like strokes in teal and orange. The overall aesthetic is artistic and hand-drawn.

02

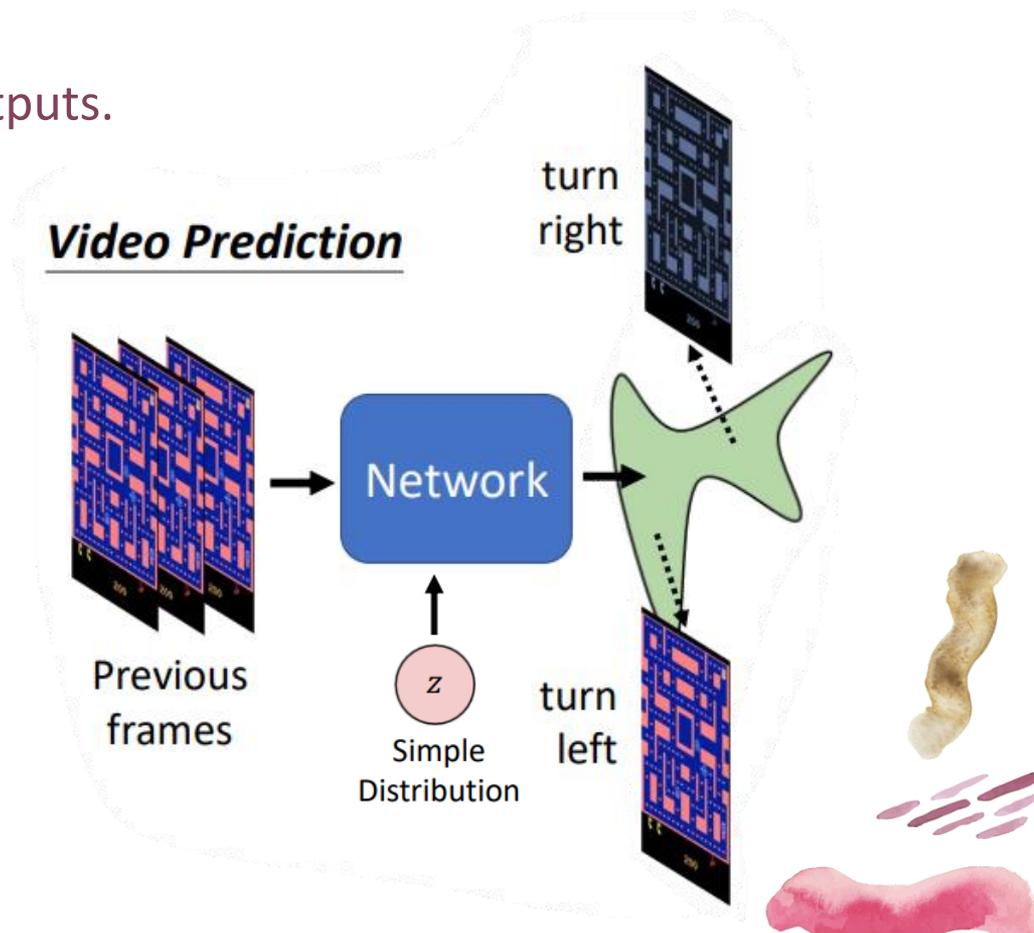
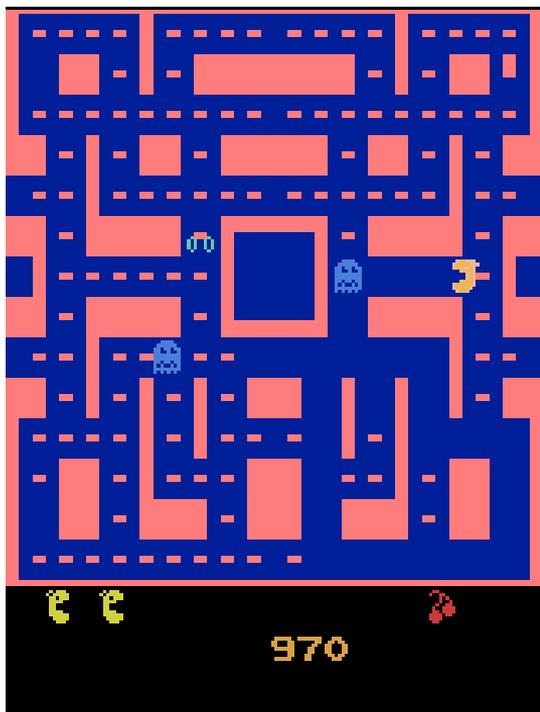
Related Work

Generation Model



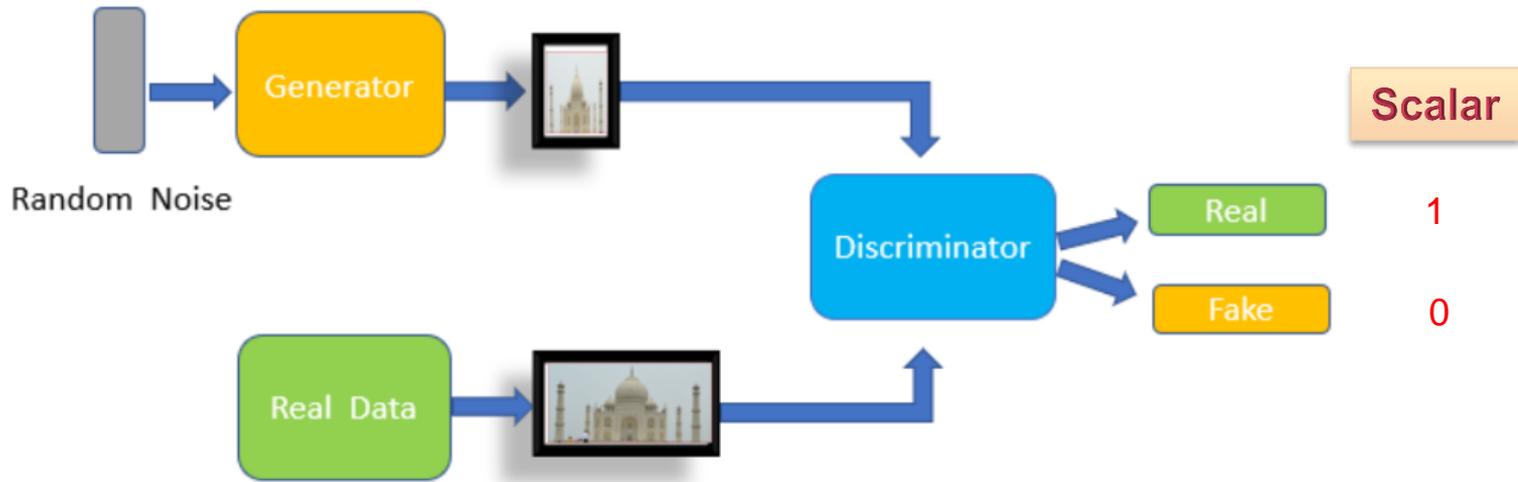
Why distribution?

The same input has different outputs.



Generative Adversarial Network (GAN)

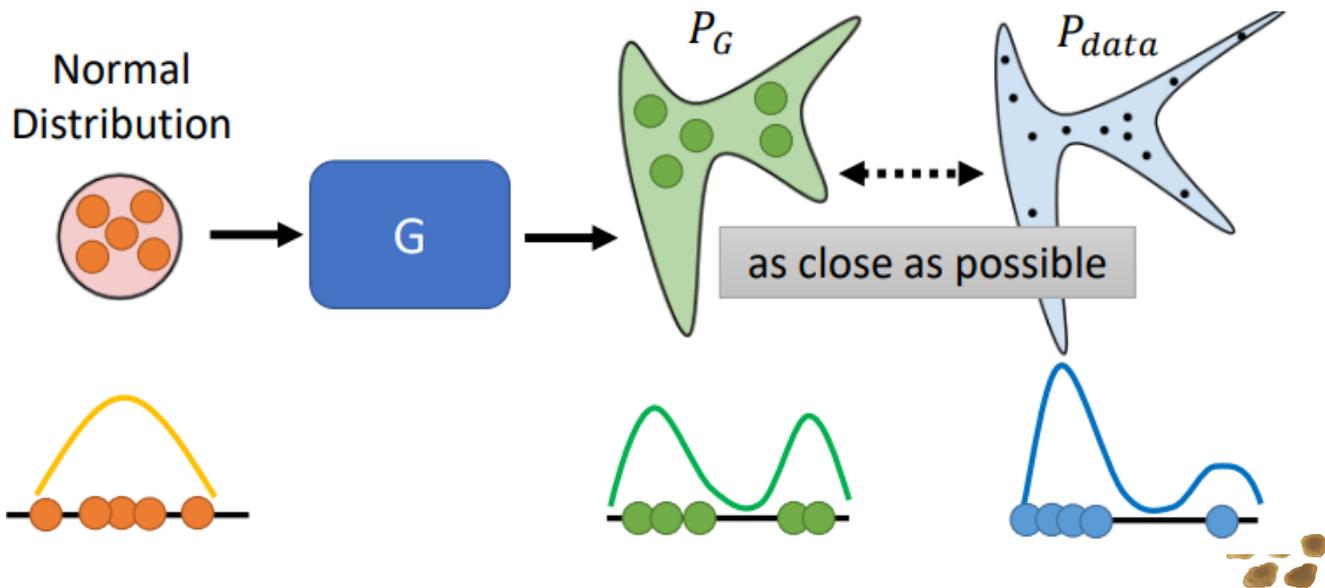
An algorithmic architecture that consists of **two neural networks**, which are in **competition with each other** (thus the “adversarial”) in order to generate new, replicated instances of data that can pass for real data.



Training Objective

Generator:

$$G^* = \arg \min_G \text{Div}(P_G, P_{data})$$

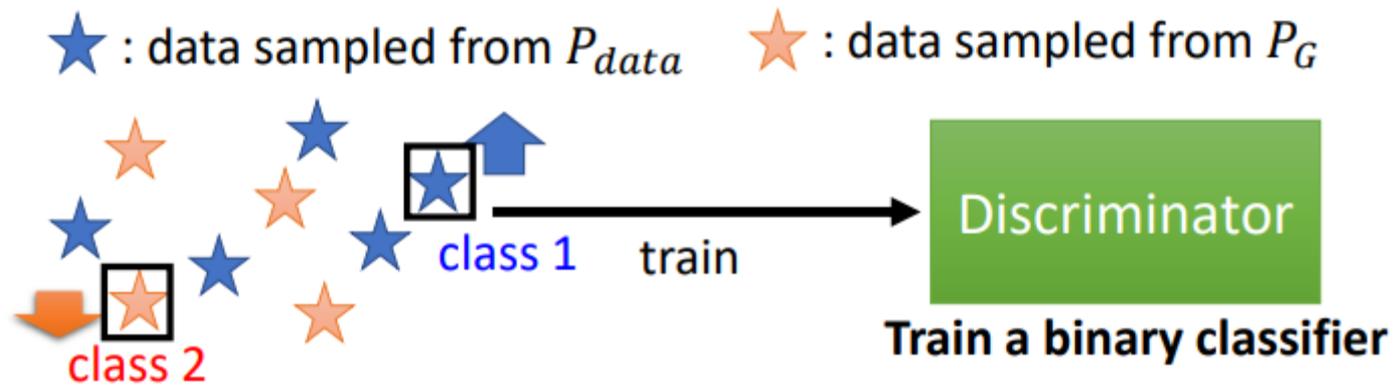


Training Objective

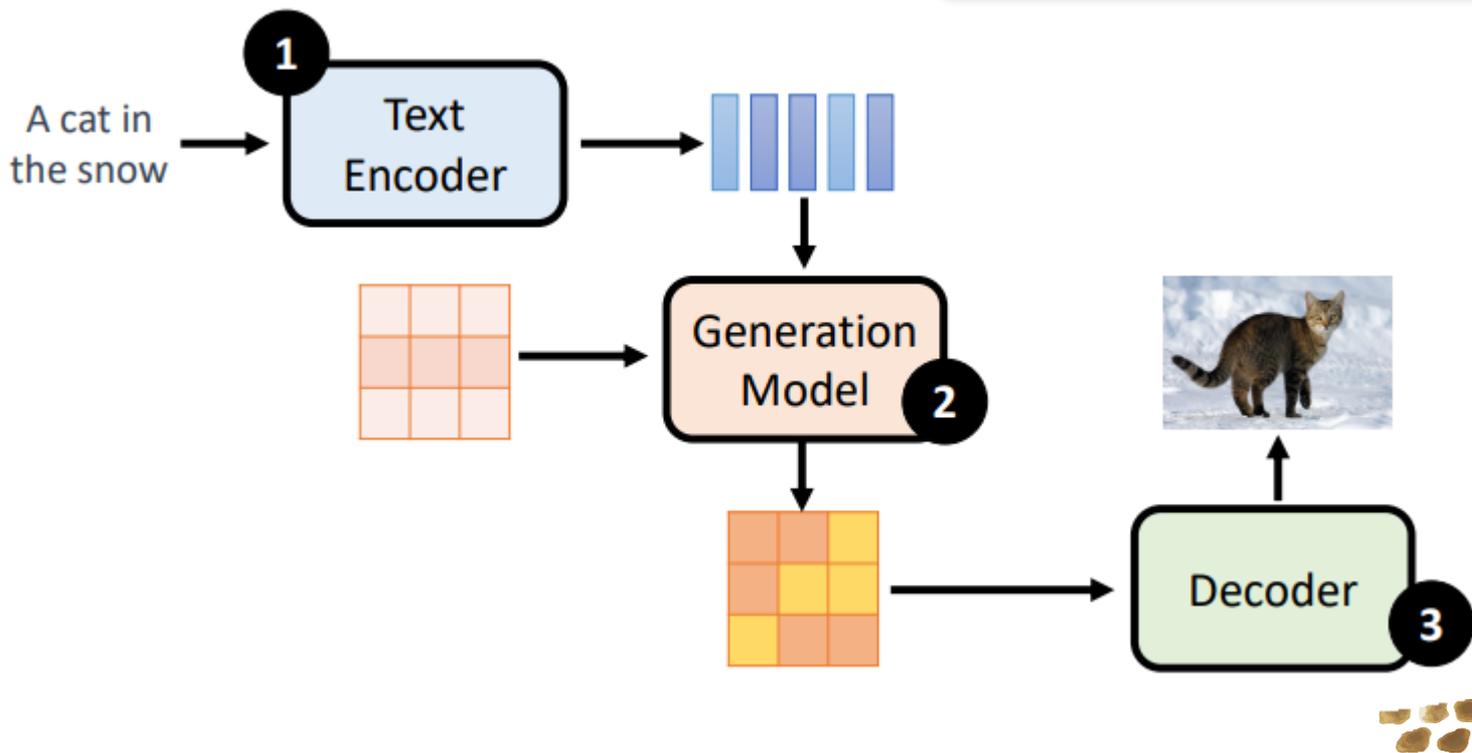
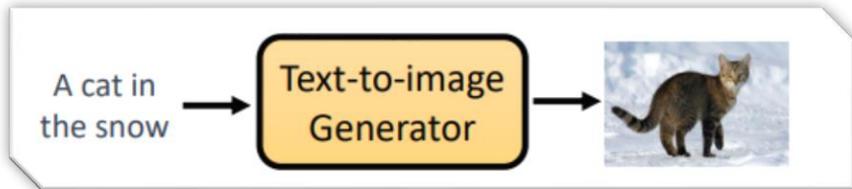
Discriminator:

$$D^* = \arg \max_D V(D, G)$$

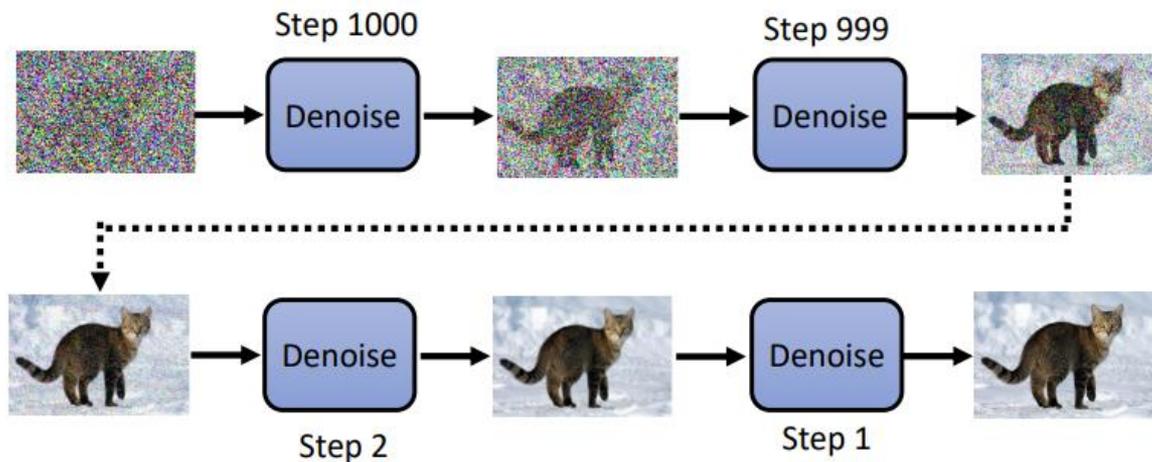
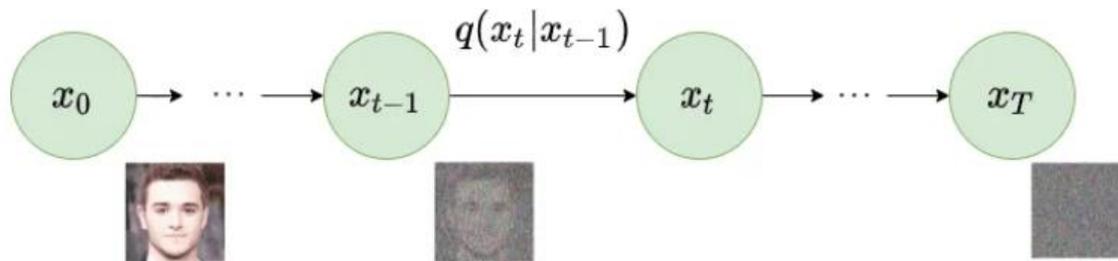
$$V(D, G) = E_{y \sim P_{data}} [\log D(y)] + E_{y \sim P_G} [\log(1 - D(y))]$$



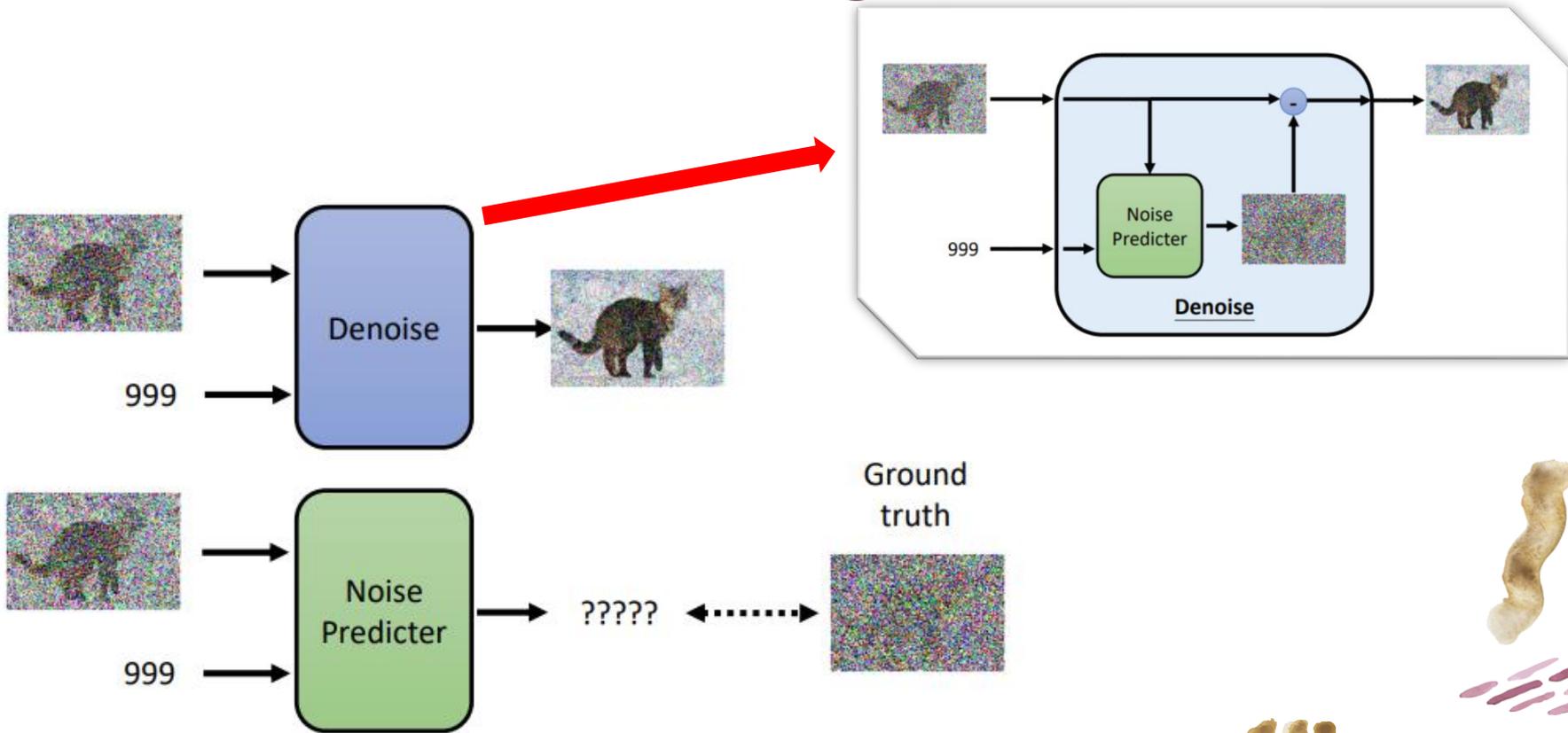
Diffusion Model



Forward & Reverse Process



Noise Predictor Training



The background is a white canvas decorated with various watercolor-style elements. There are large, soft-edged shapes in shades of blue, pink, purple, and yellow. Smaller, more defined shapes include a cluster of red dots, a group of teal dashes, and several orange pencil-like strokes. The overall aesthetic is artistic and hand-drawn.

03

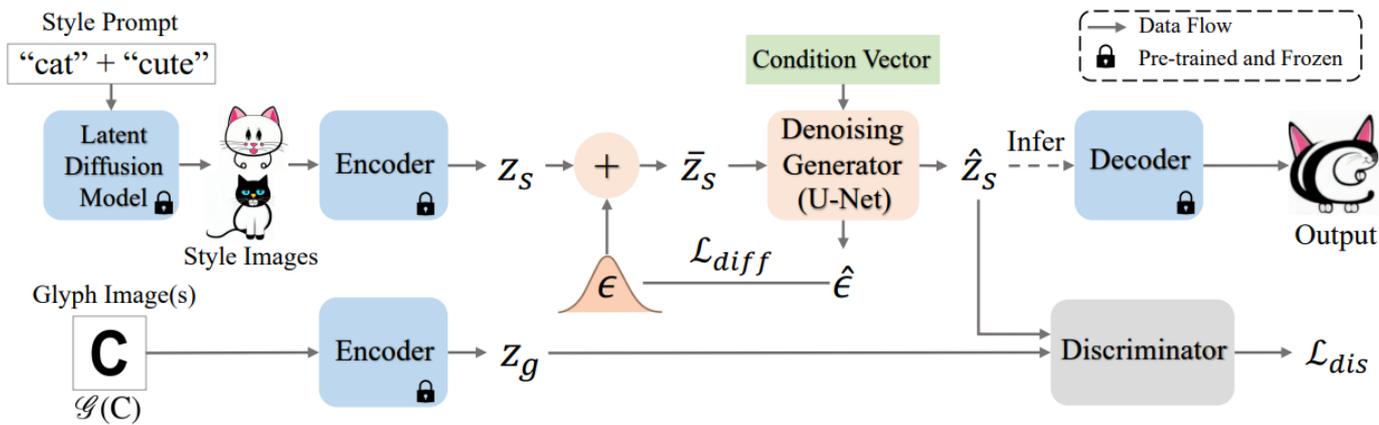
Method

Overview

Input: style prompt & glyph

Style word + (opt) Style attribute

Output: stylized version of the glyph based on the style prompt



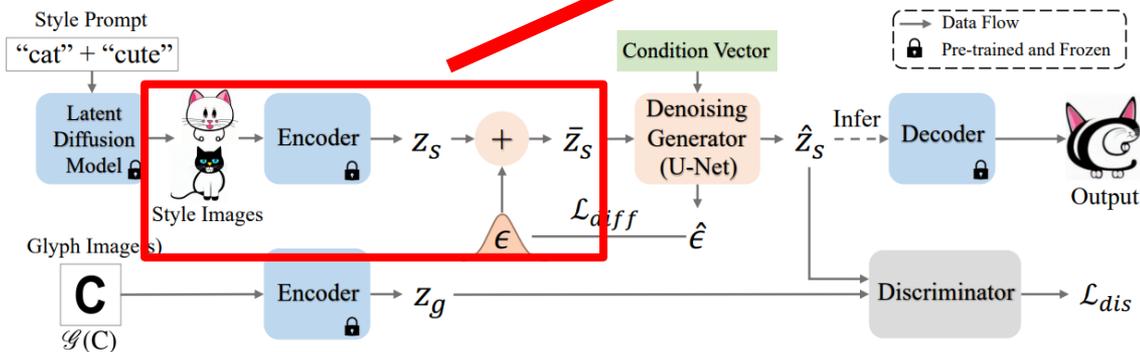
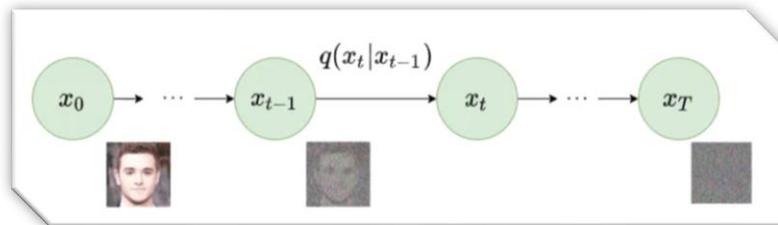
The Style Latent Space

- LDM:

- Encoder

- Input: style images
- Output: **features maps** Z_S

- Apply noise on Z_S to obtain \bar{Z}_S



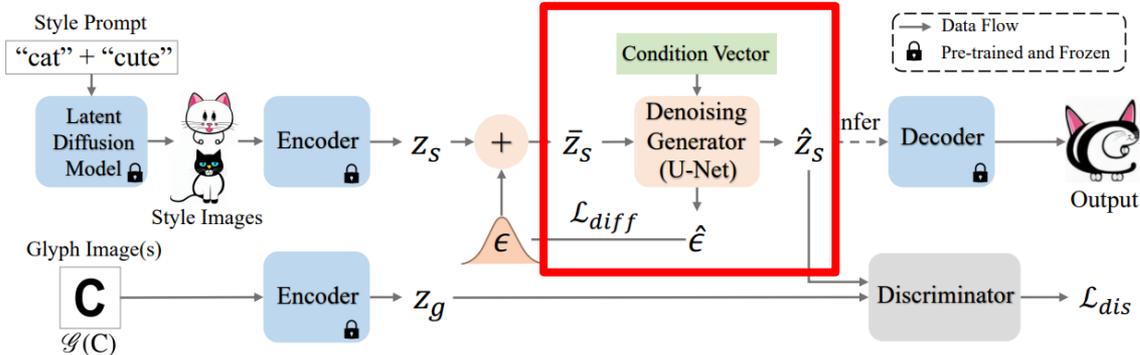
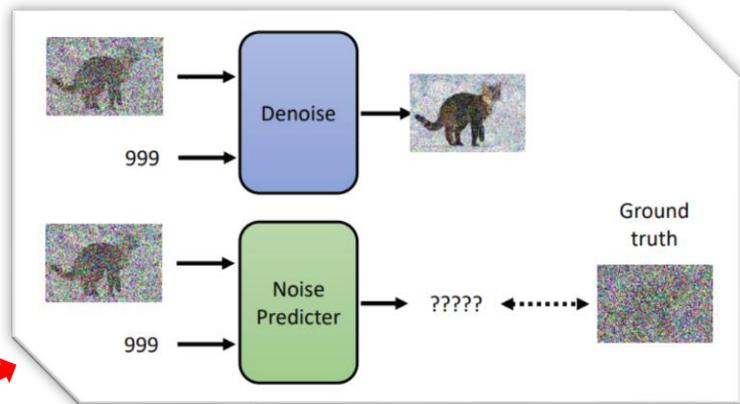
The Style Latent Space

- Denoising generator:

1. Predict the added noises $\hat{\epsilon}$ from \bar{z}_S

2. Denoise \bar{z}_S to \hat{z}_S

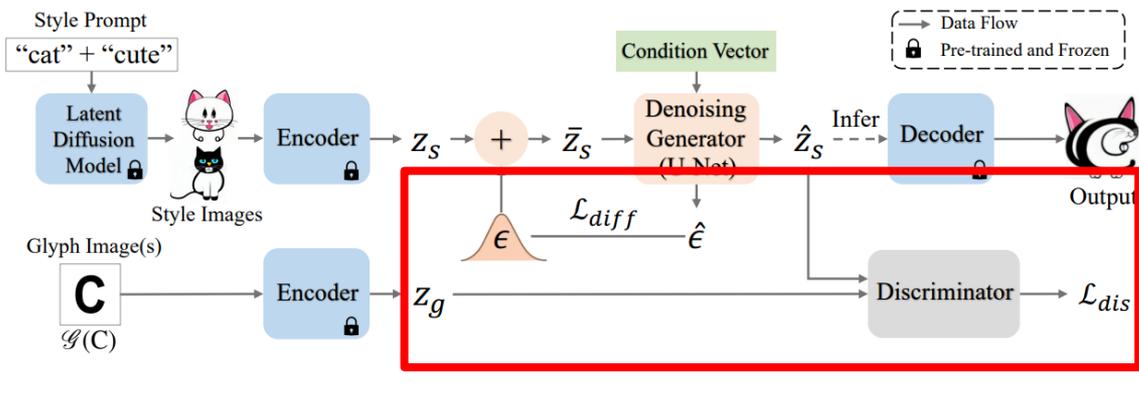
3. Diffusion loss: $L_{diff} = \|\hat{\epsilon} - \epsilon\|_2^2$



The Discriminator

- Different from vanilla GANs, the discriminator here takes input as feature maps instead of raw images.

• **Discriminator loss:** $L_{dis} = \log(D(Z_g)) + \log(1 - D(\hat{Z}_S))$



Overall Loss Function

- $\min_G \max_D (L_{diff} + \lambda L_{dis})$
- Employ **CLIP** to judge the quality of results from both stylistic and glyph preservation standards.

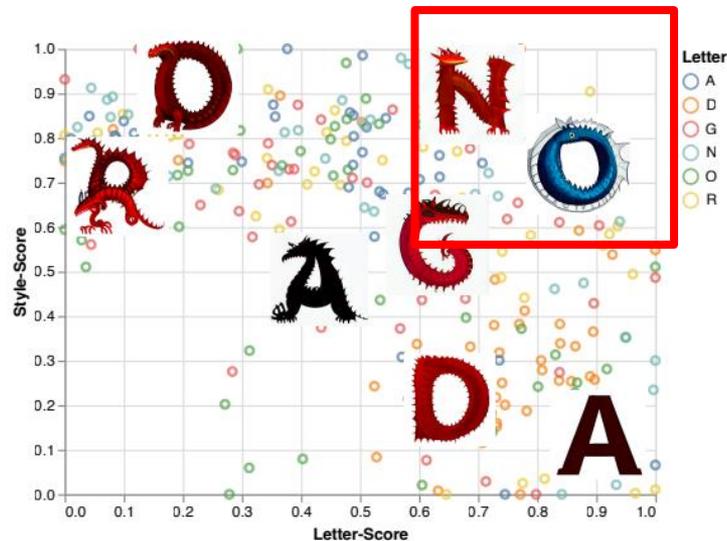


Figure 3. Ranking results. The horizontal and vertical axes respectively denote the scores of glyph and stylistic preservation.

The background is a white canvas decorated with various watercolor-style elements. There are large, soft-edged washes of color in shades of blue, pink, purple, and yellow. Interspersed among these are smaller, more defined patterns: clusters of small dots in red and orange, groups of parallel lines in blue and purple, and some abstract, brush-like strokes in teal and orange. The overall aesthetic is artistic and hand-drawn.

04

Results

Single-letter Input

UNICORN WINE CAFE LAMP VASE
ISLAND MERMAID SNAKE PEACOCK

ROSE DRAGON ROBOT PLANT
ASTRONAUT SOCKS PARROT SNAIL

Multi-letter Input



Comparisons

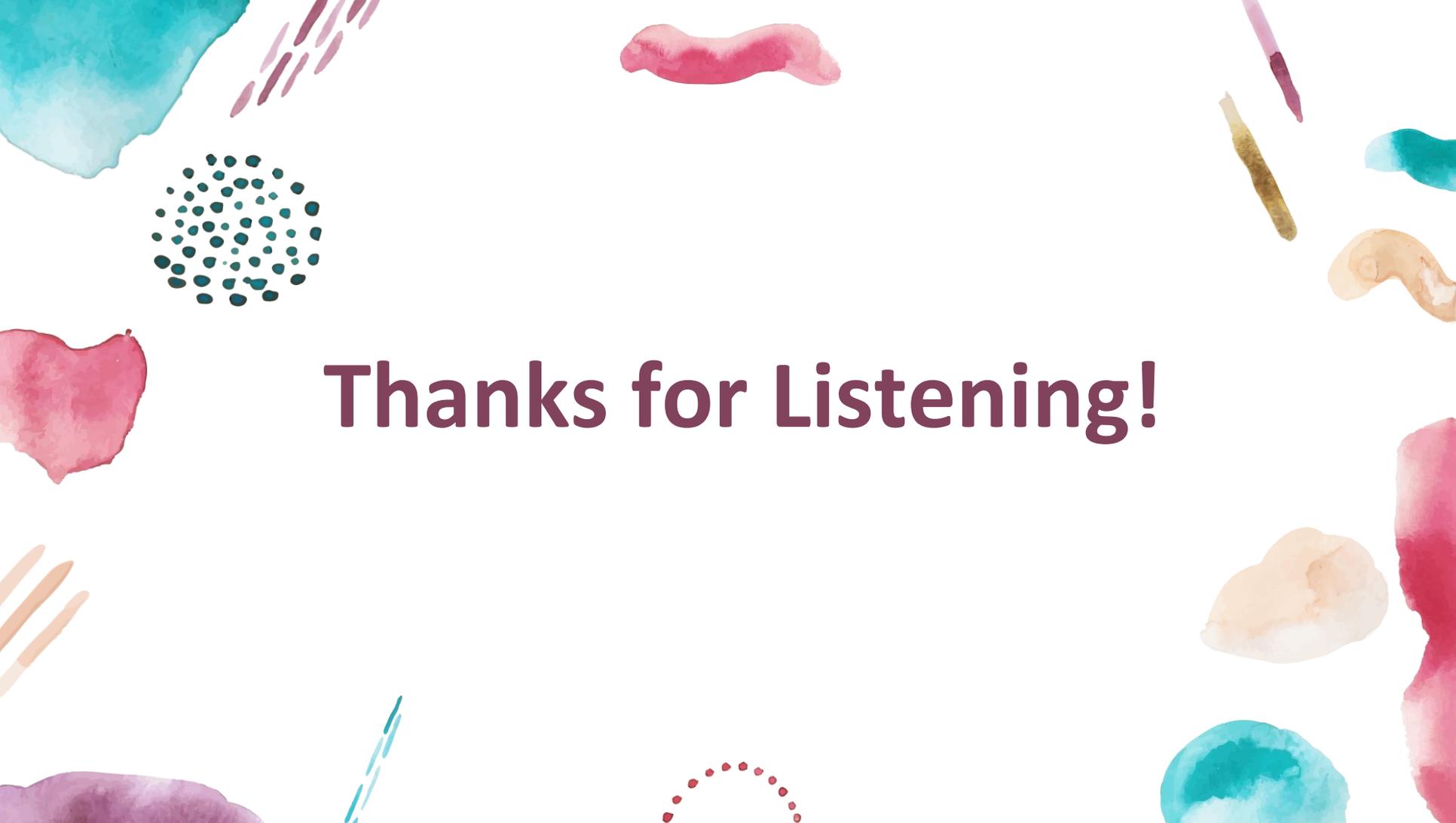
	DRAGON	PLANT	OCTOPUS	SNAIL
Google Search				
SD				
CLIPDraw				
DALL-E 2				
Ours				



Limitation & Future Work

- When dealing with multi-letter inputs, our method may struggle to generate satisfactory results if the style images and letters are too dissimilar.
- Future work could involve training a network for a particular style that can generate any letter during inference.



The background is a white canvas decorated with various watercolor-style elements. In the top left, there's a large teal wash. Below it, a cluster of small teal dots. To the right, a pinkish-red brushstroke. Further right, a purple and gold brushstroke. In the bottom left, a purple wash and a blue dashed line. At the bottom center, a semi-circle of red dots. On the right side, there are several other washes in teal, orange, and pink. The central text is a bold, dark purple color.

Thanks for Listening!