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

Typography

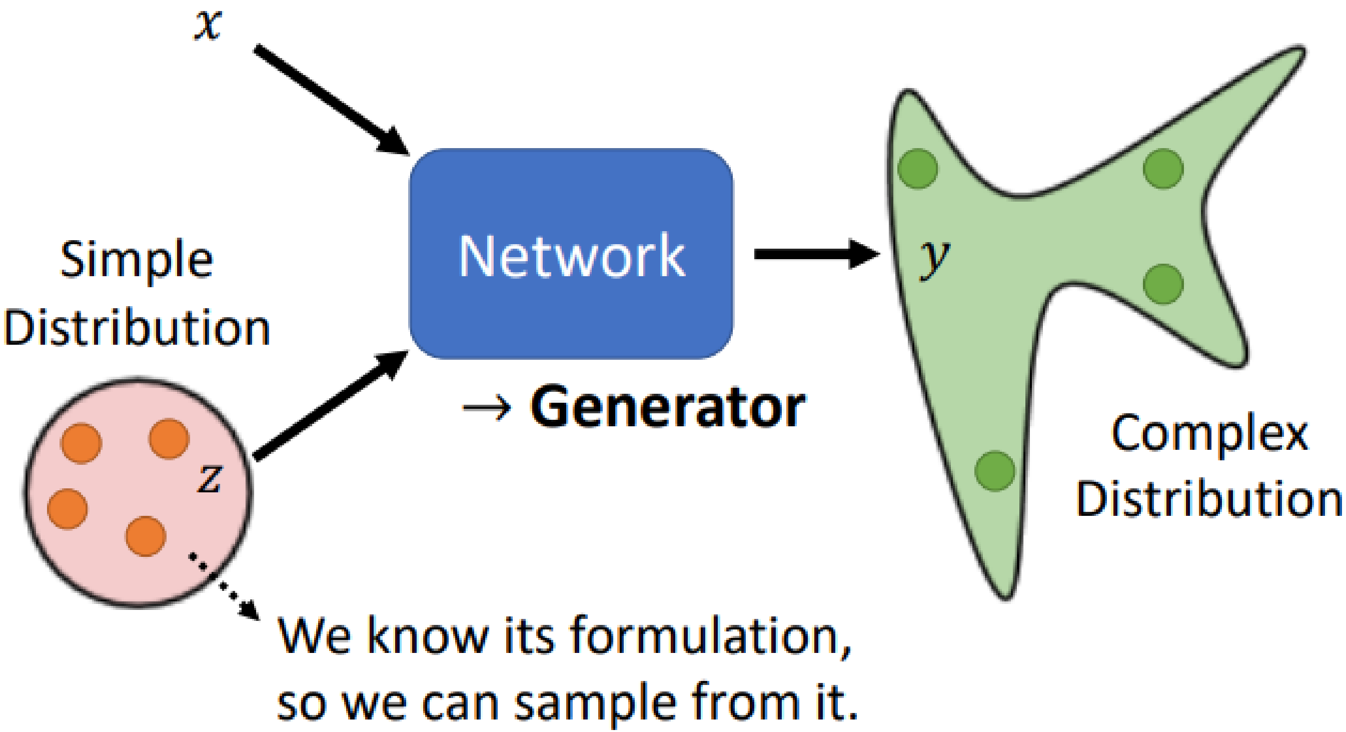
- The art and technique of arranging type to make written language legible, readable, and appealing when displayed.

- Conflicting goals:
    1. Artistic stylization
    2. Legibility

02

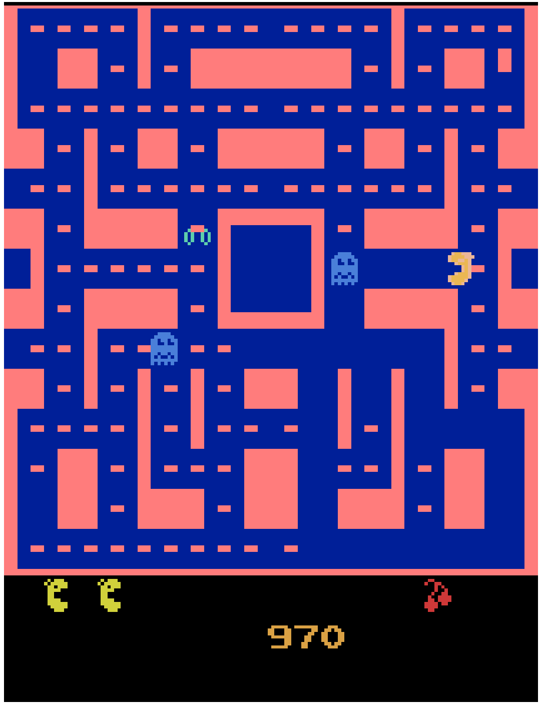
Related Work

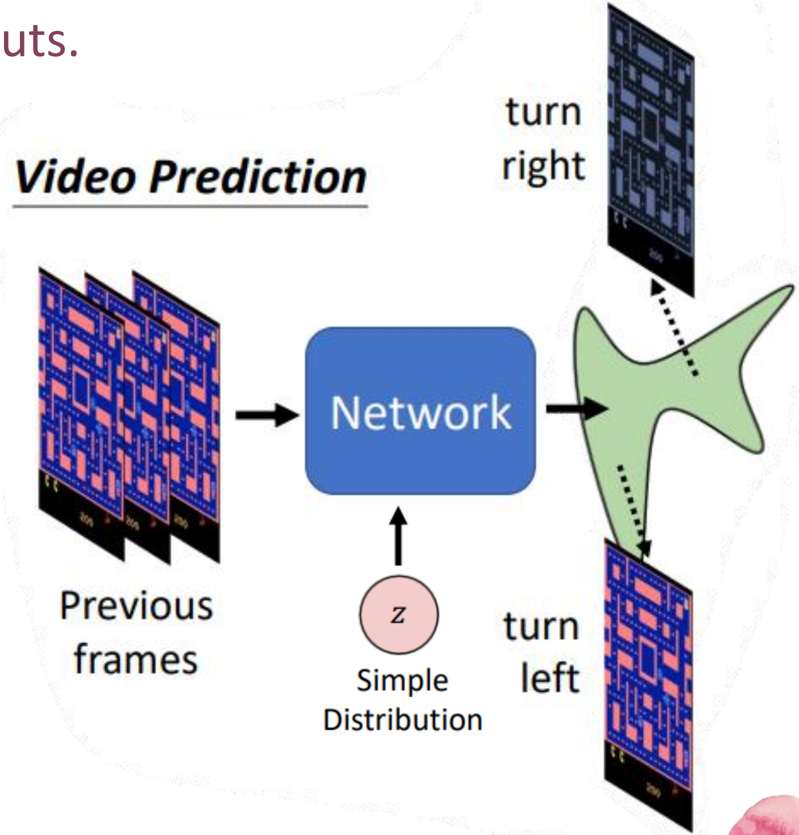
Generation Model



Why distribution?

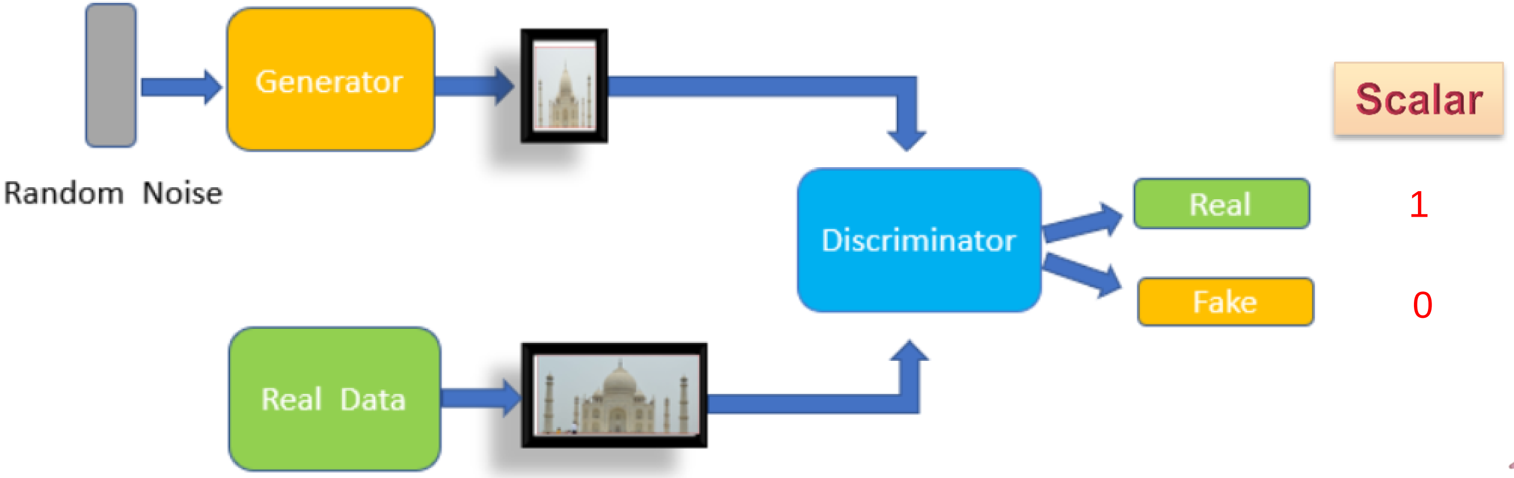
The same input has different outputs.





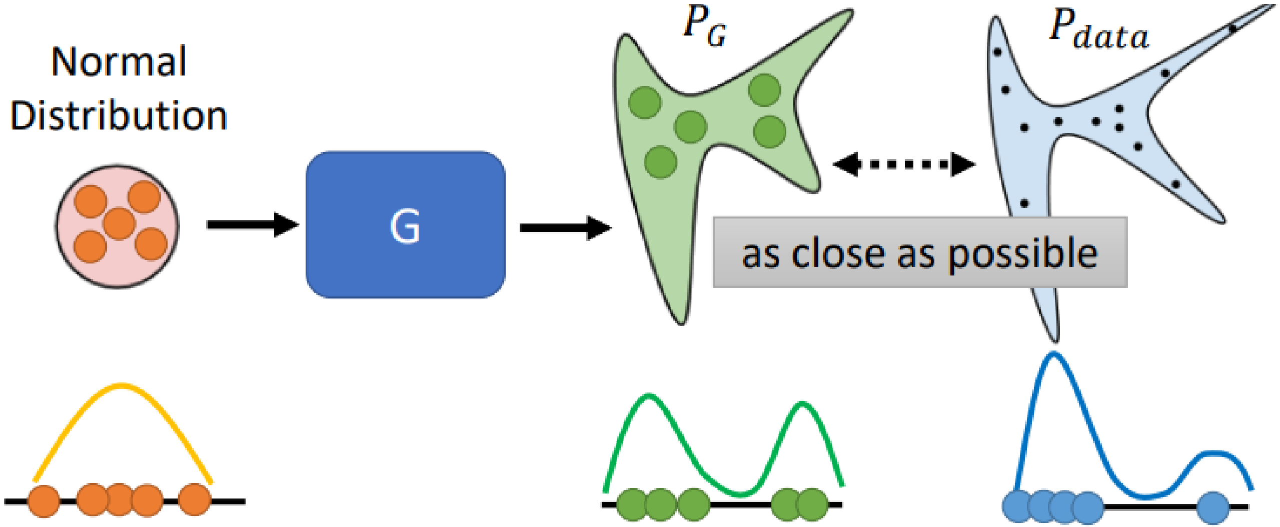
Generative Adversarial Network (GAN)

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

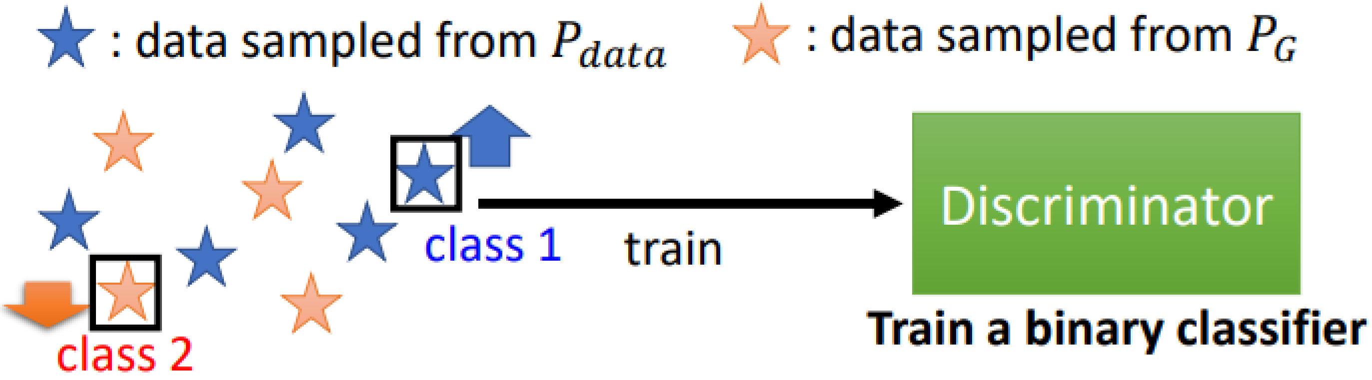
Generator: $G^* = \arg\min_G Div(P_G, P_{data})$



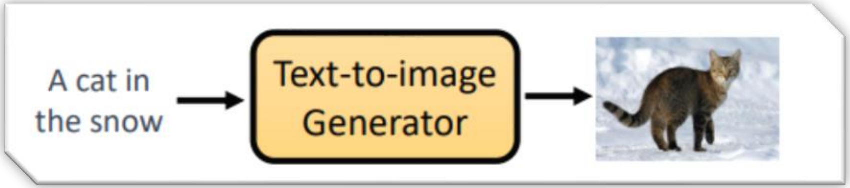
Training Objective

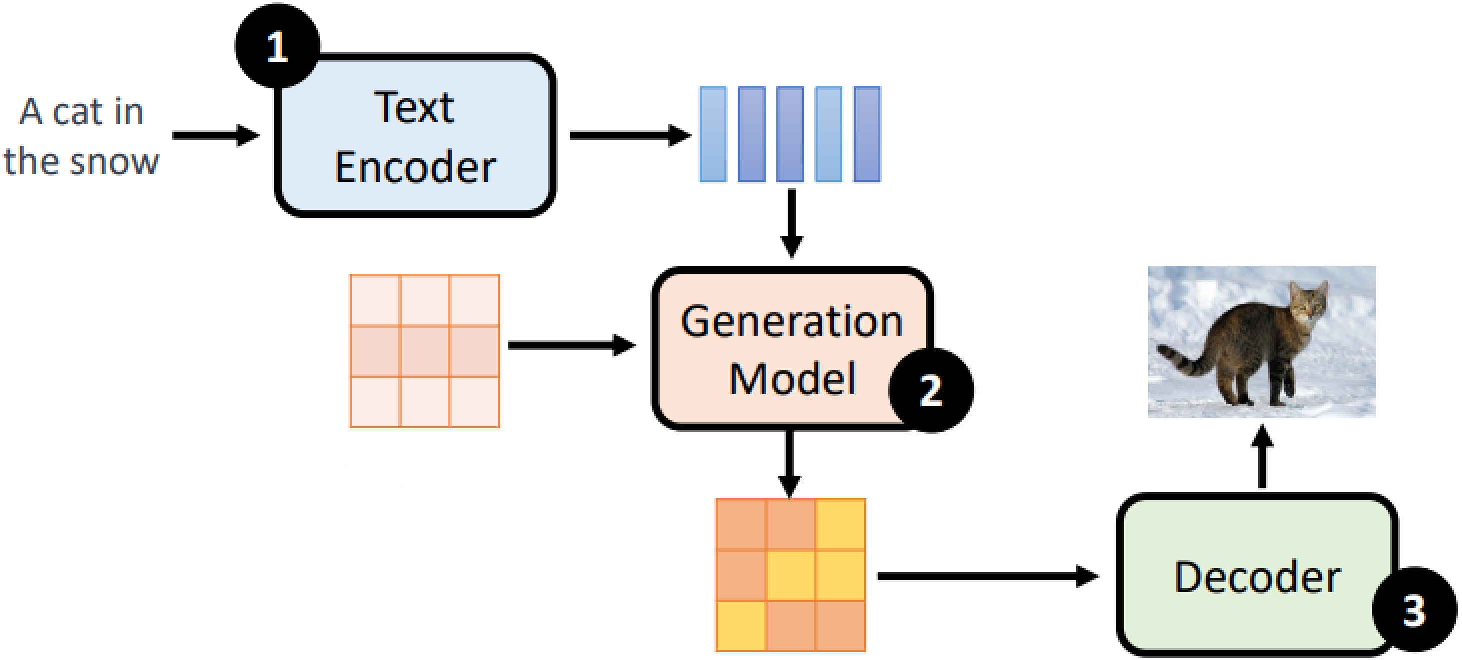
Discriminator: $D^* = \arg\max_D V(D,G)$

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

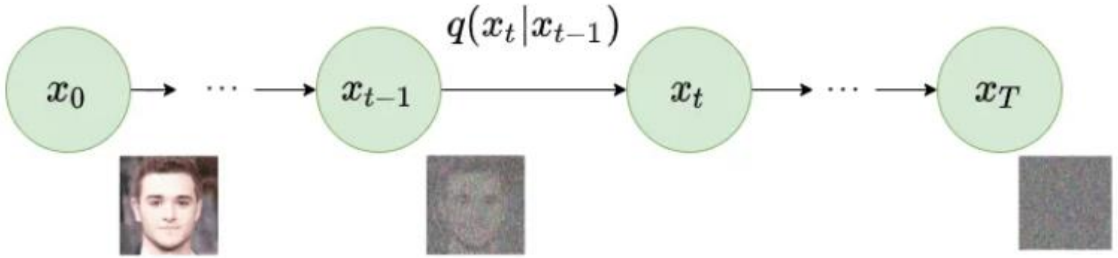


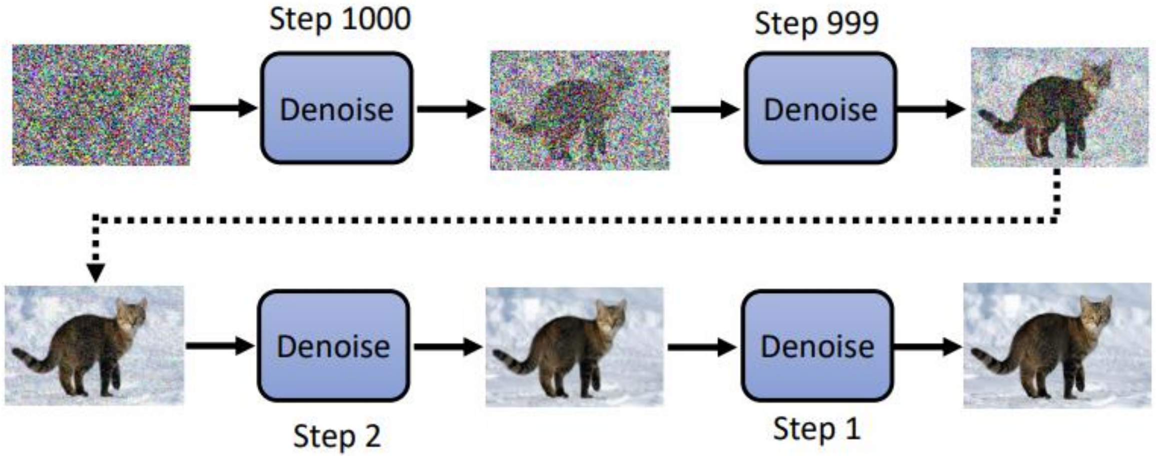
Diffusion Model



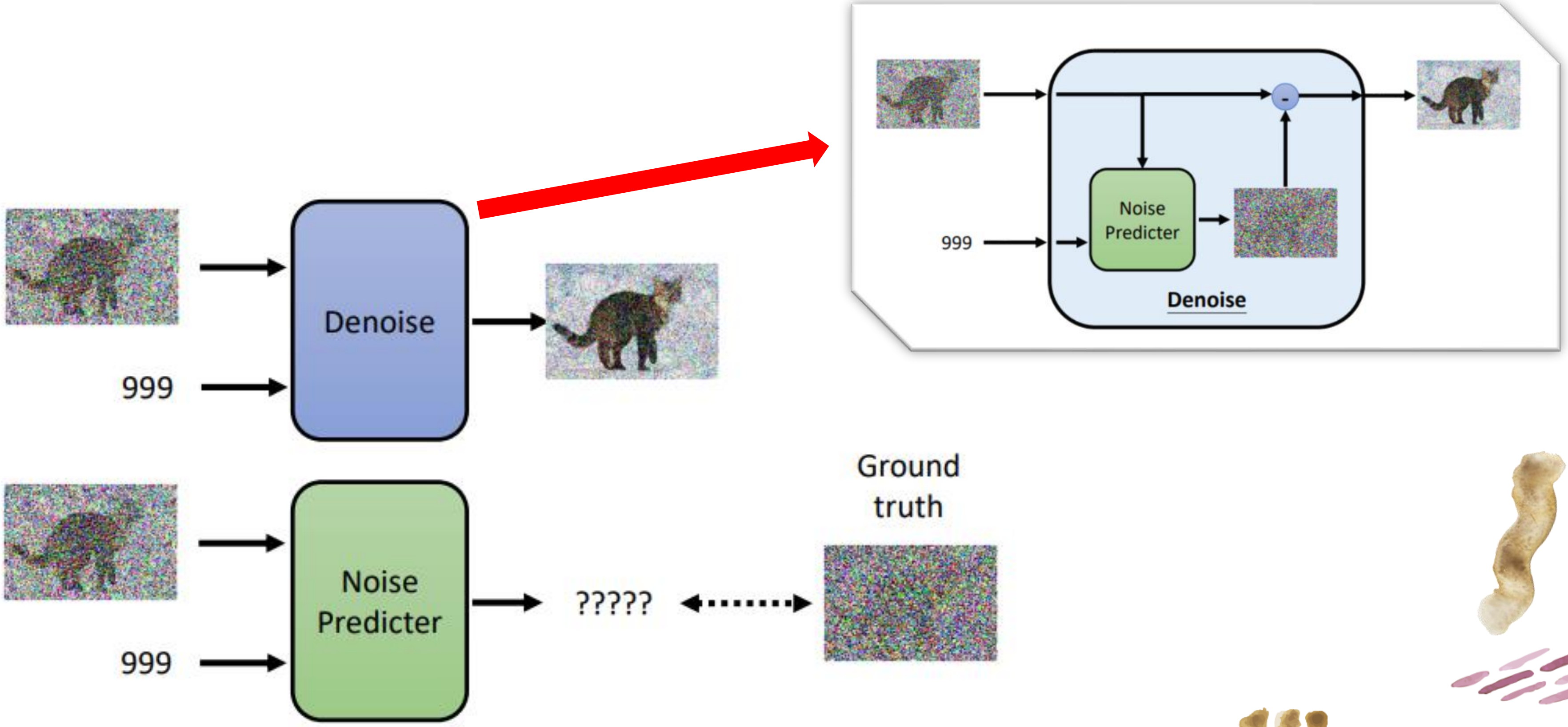


Forward & Reverse Process





Noise Predictor Training



03

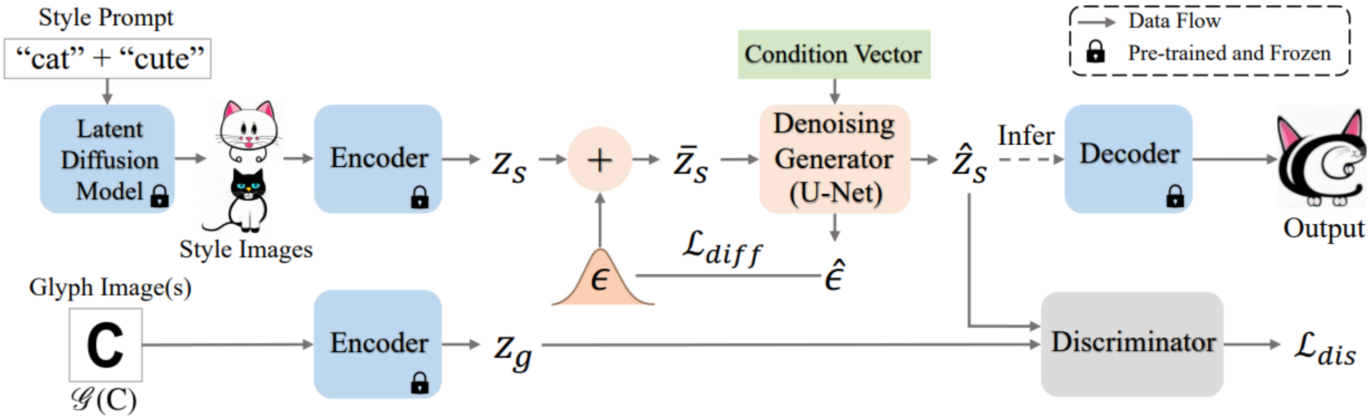
Method

Overview

Input: Style prompt & glyph

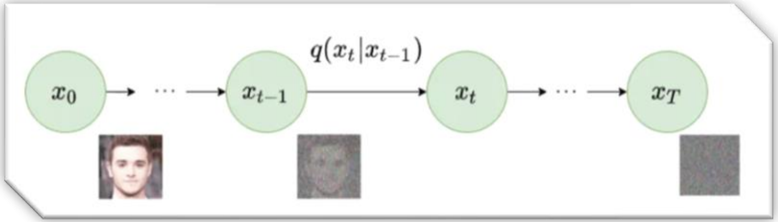
Style word + (opt) Style attribute

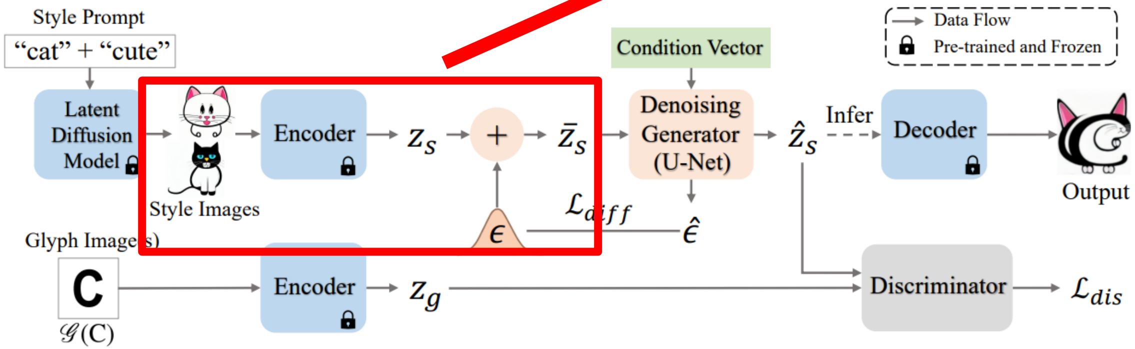
Output: stylized version of the glyph based on the style prompt



The Style Latent Space

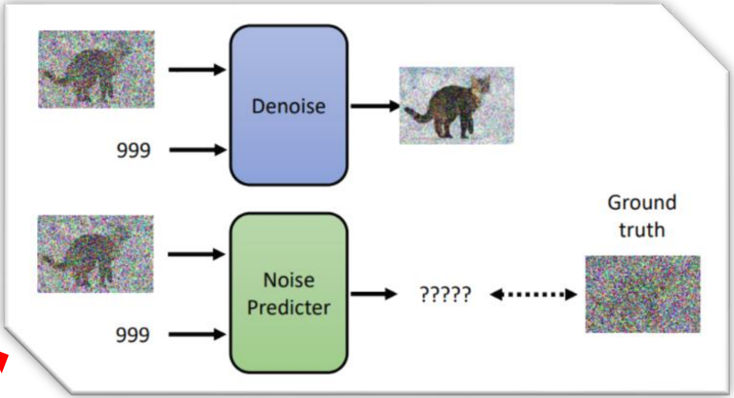
- LDM:
  1. Encoder
     - ➢ Input: style images
     - ➢ Output: **features maps** z_s
  2. Apply noise on z_s to obtain $\bar{z}_s$

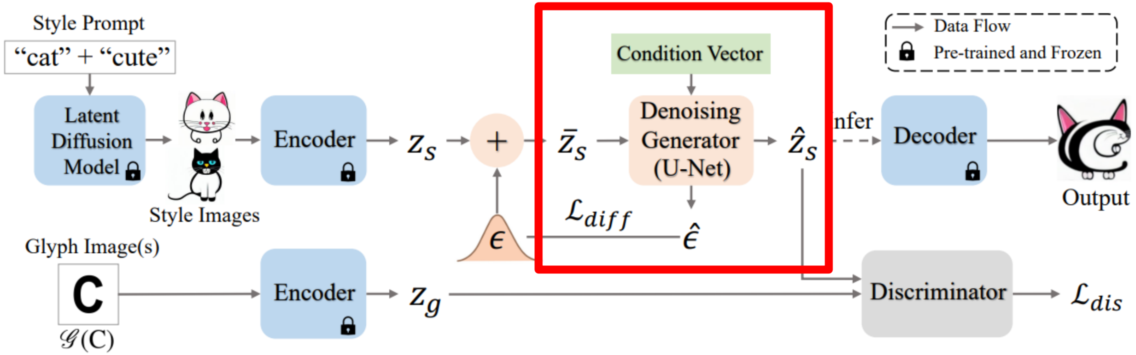




The Style Latent Space

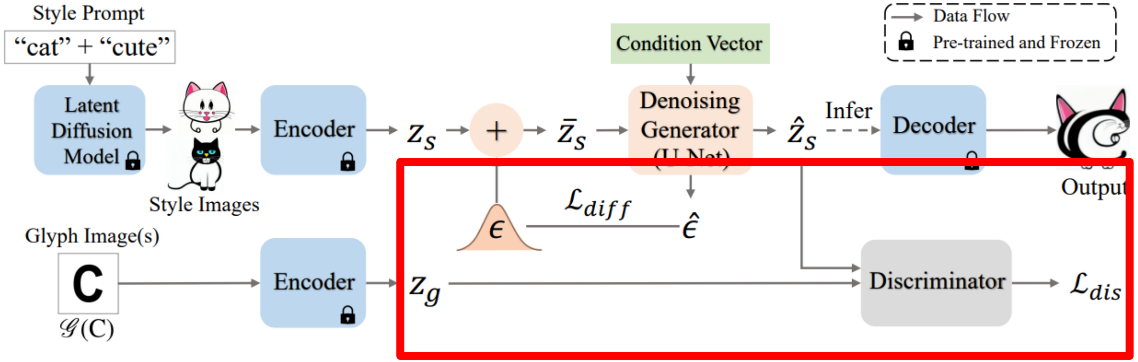
- Denoising generator:
  1. Predict the added noises $\hat{\epsilon}$ from $\bar{Z}_S$
  2. Denoise $\bar{Z}_S$ to $\widehat{Z}_S$
  3. **Diffusion loss**: $L_{diff} = || \hat{\varepsilon} - \varepsilon ||_2^2$





The Discriminator

- Different from vanilla GANs, the discriminator here takes input as feature maps instead of raw images.
- **Discriminator loss:** $L_{dis} = \log\left(D(Z_g)\right) + \log(1 - D(\widehat{Z_S}))$



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.

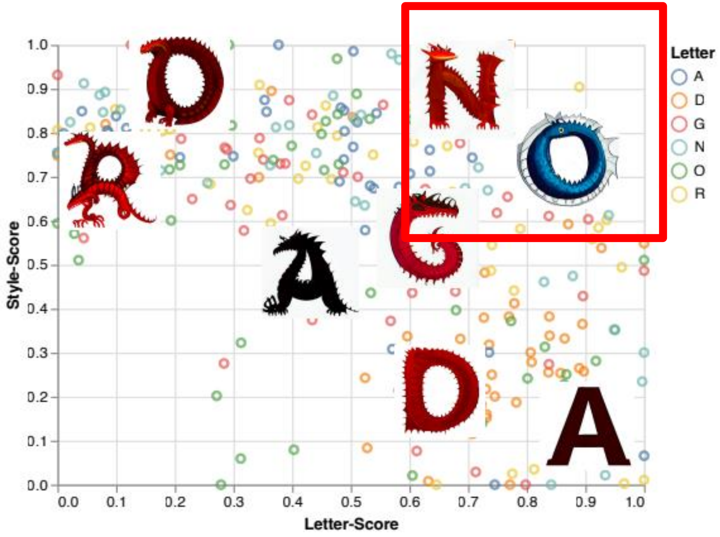


Figure 3. Ranking results. The horizontal and vertical axes respectively denote the scores of glyph and stylistic preservation.

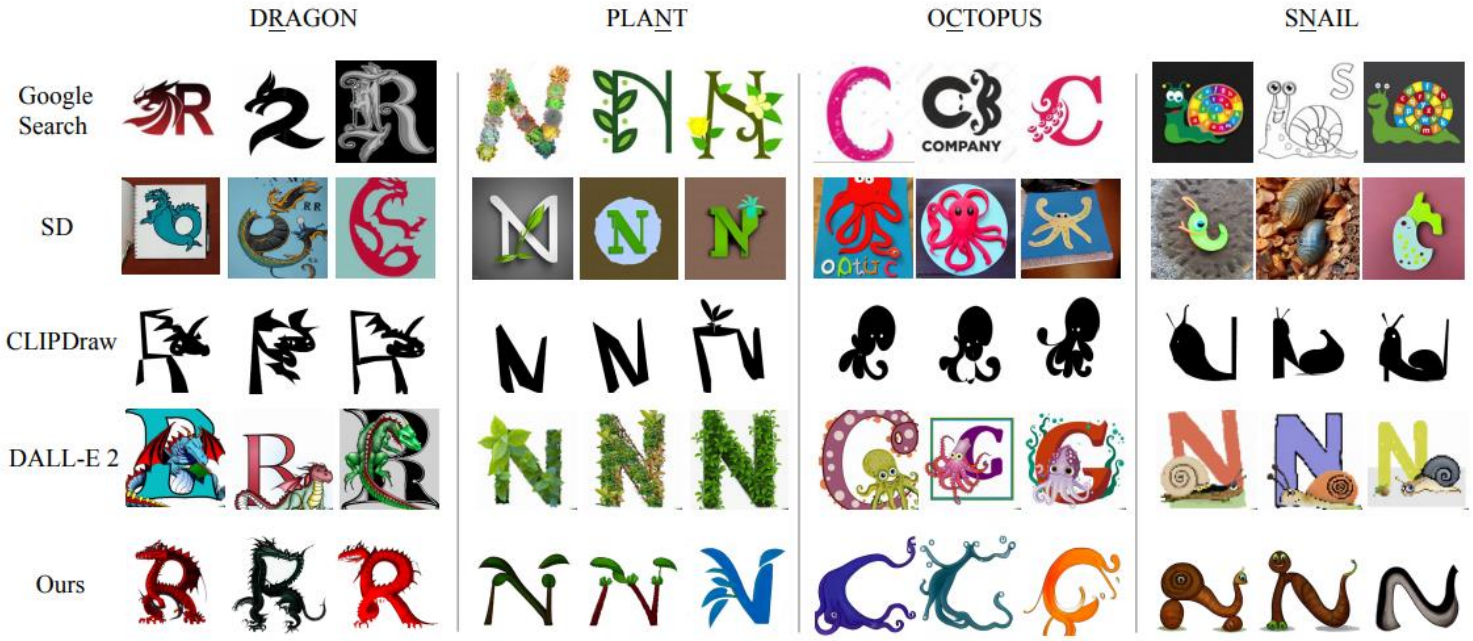
04

Results

Single-letter Input

Multi-letter Input

Comparisons



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.

Thanks for Listening!