Generative AI Story & Illustration Creator
利用生成式AI創作故事與插圖

郭舒文 NTNU_010153217　　　馮意凡 NTU_R10521609

Github:
https://github.com/Evan102/Generative-AI-Story-Illustration-Creator/tree/main

Outline

- Motivation
- Literature review
- Objectives
- Story generation
- Illustration generation
- Conclusion & Platform demo

Motivation

Until 2023/6/3,
793 books co-authored with chatGPT, 78 books illustrated by Mid Journey.

Literature review

AI.R Taletorium Flowchat

Take a Selfie

Facial Attribute Mapping

Assign a Cast for User

AI Story Generation

Scene Graph

| owl | | penguin | kids |

back to — go across — by — play with

tree house — pool — butterfly

play on

swingset

Generated Story

Then the *owl* returned back to his *tree house*, and his *kids* are playing with a *butterfly*. He goes across the *pool* and he saw a *penguin* playing on a *swing-set* by the pool.

Doodler Layout Generation

AI.R Doodling Run

Reference: Mutltimodal AI Companion for Interactive Fairytale Co-creation

4

Literature review

Output

You may think that I may have an Improper view of my fellow men, but I can see only one point. A man's heart changes when a man comes to accept his object as more important than his hopes and hopes change his heart.

....

(3) Images coherency re-ranking +Style transfer

(2) Image retrieval by caption-generated text cosine similarity

Coherency | Readability | Sentiment | Diversity | Simplicity| Tale Like

(1) Per sentence Rule Based Re-ranking

Decoder
⋮
Decoder

Fine-tuned GPT-2

Input   The   Myth   of   the   Objective   \n

The system generates text while re-ranking, retrieves images from Unsplash dataset, applies **style transfer** and then re-ranks stories according to the story's visual consistency.

Reference: <u>FAIRYTAILOR: A MULTIMODAL GENERATIVE FRAMEWORK FOR STORYTELLING</u>

5

Objectives

1. Empower students to experience generative AI's 'creativity' in storytelling through adaptation and creation of stories with illustrations.
2. Leverage a large language model to design story and illustration prompts or fine-tune the model for story creation and illustration.
3. Compare the results of different models for the same prompt.

Story Generation

- Methodology
- Prompt Evolution Process

Methodology

1. Using the ChatGPT API
    a. We connected to the ChatGPT via its API, creating prompts for children's books.
2. Prompt Examples
    a. Students can generate stories directly using simple keywords or detailed descriptions.
3. Language and Token Limitation
    a. As text-to-image generation is more appropriate with English stories
    b. Shortened them to fit the model's 50-token requirement.
4. Web Application Development -Gradio

郭舒文 • 大約1分鐘

302 Padlet

老師示範   ⋮

第1組   ⋮

第2組

＋

＋

中文的故事

先把AI協作的故事貼在這裡，依照自己的想法條調整故事內容

對AI下指令的技巧：
1. 設定AI助理的角色：假設你是兒童故事的編輯，請協助我修改故事。
2. 輸入關鍵字：(故事的描述，越詳細越好)
3. 輸出的要求：請用「100字」完成一個「適合10歲兒童」閱讀的「中文故事」。
4. 如果產生結果不滿意，可以要求AI繼續調整。但必須具體說明，例如：加入新角色小強、改為快樂的結局等等。

♡ 0

英文故事繪圖

如果要使用AI繪圖功能，需要將故事翻譯成簡短的英文故事，你可以利用AI協助翻譯。

對AI下指令的技巧：
「請將以下內容翻譯成英文。
  --- (利用斷行或分隔線)
  需要翻譯的內容」

♡ 0

班級/姓名/座號 - 故事主題

小小企鵝 Cosmo，渴望探索宇宙的奧妙。他在一次研究夜空時，發現了一顆神秘的星星，聽說暗藏著一個宇宙秘密。Cosmo 心懷冒險之心，決定踏上尋找之旅。在旅途中，他結識了來自不同星系的夥伴，大家一起克服了許多難關，終於找到了神秘的星星。他們發現，這顆星星其實是一個友善的太空生物，他們一起享受了屬於青春、冒險和宇宙的美好時光。

Little penguin Cosmo longed to explore the mysteries of the universe. During one study of the night sky, he discovered a mysterious star rumored to hold a cosmic secret. With a heart full of adventure, Cosmo decided to embark on a journey to find it. Along the way, he made friends
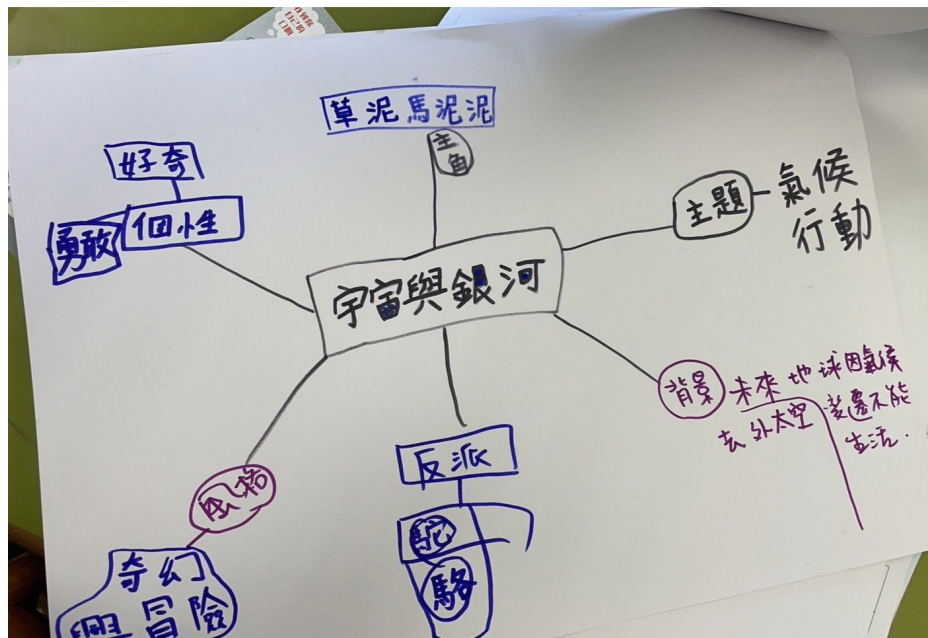
Prompt Evolution Process

Before

Step 1: Create a Chinese story

Step 2: Translate the Chinese story into English

Step 3: Extract keywords for illustration

After

Give ChatGPT a role , input keywords, and specify requirements for the output format (e.g., token limits, language).

Input: Chinese or English keywords

Output: English translation (50-word token summary)

Tip: Prompt allows for the use of mixed Chinese and English languages.

```
prompt = f"As a {role}, Please write a story of around 200 words in Traditional Chinese , Story condensed into
50 tokens in English based on the following keywords and style, style:  {style}, keywords: {',
'.join(keywords)} ."
```

Keywords, separated by commas

小貓 狗 森林

Style

○ Adventure   ● Happy   ○ Sad   ○ Warm

Model Choice

● OpenAI GPT-3.5 Turbo   ○ OpenAI text-davinci-003   ○ OpenAI text-davinci-002

Clear

Submit

output

Traditional Chinese:

從前，有一隻小貓和一隻小狗，牠們住在一個美麗的森林裡。牠們是好朋友，經常在森林裡玩耍和探險。有一天，牠們走到了一個很陌生的地方，迷路了。但是，牠們不怕，因為牠們互相幫助，最後成功地找到回家的路。從此以後，牠們更加珍惜彼此的友誼，每天都過得非常開心。

English:

Once upon a time, there was a little cat and a little dog who lived in a beautiful forest. They were good friends and often played and explored in the forest. One day, they wandered into an unfamiliar place and got lost. But they weren't afraid because they helped each other and eventually found their way home. From then on, they cherished their friendship even more and lived happily every day.

(50 tokens)

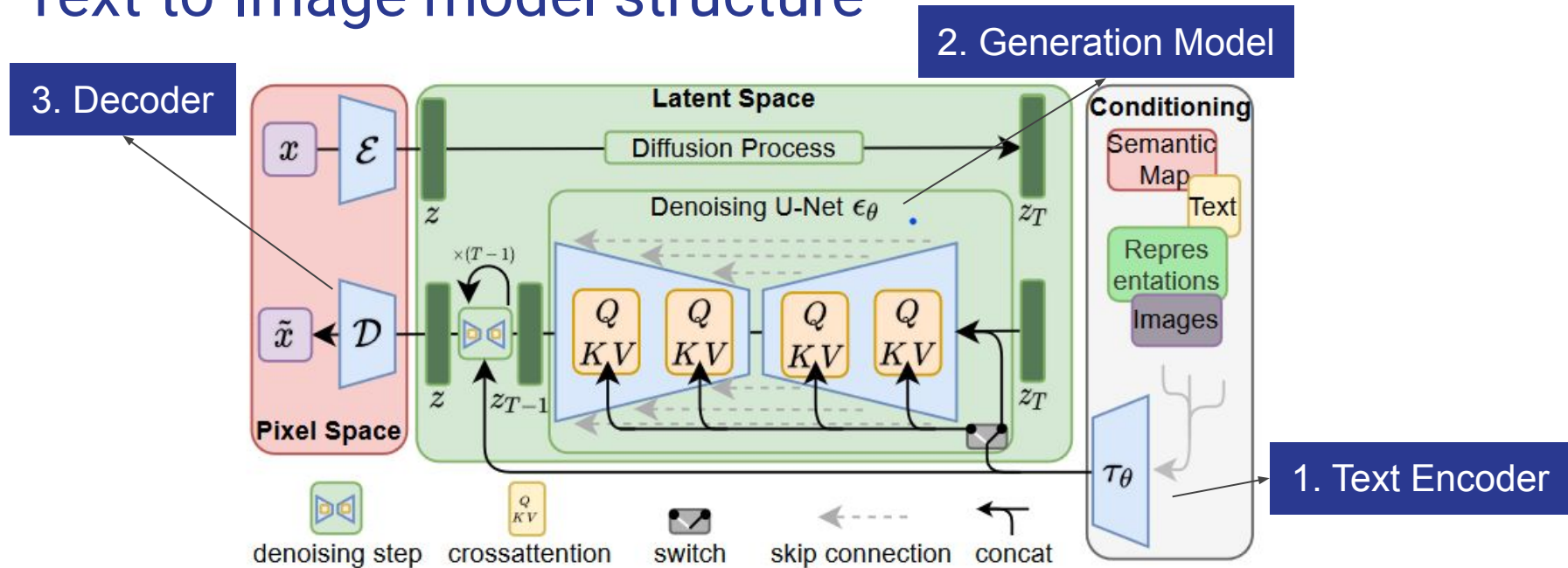Little cat and dog lost in forest, helped each other and found way home. Cherished friendship and lived happily.

AI Story Creator v3.ipynb - Colaboratory
(google.com)

Flag

Illustraion Generation

- Text-to-Image model structure
- Low-Rank Adaptation(LoRA) introduction
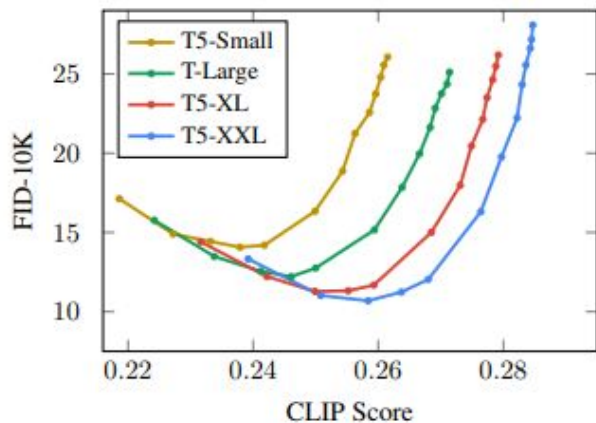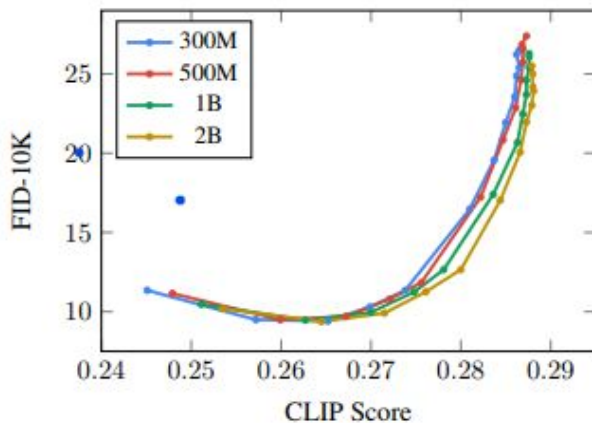- Comparison of Text-to-Image model prediction

Text-to-Image model structure

Reference: High-Resolution Image Synthesis with Latent Diffusion Models

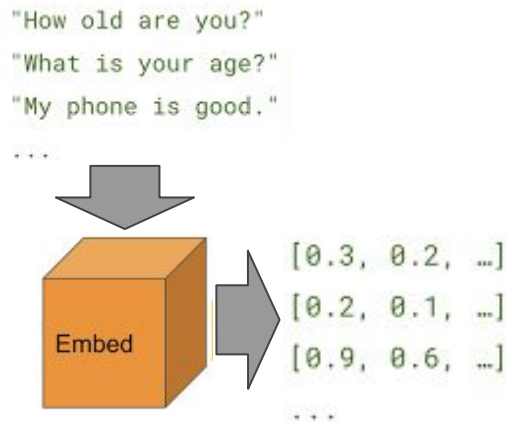# Text-to-Image model structure - 1. Text Encoder

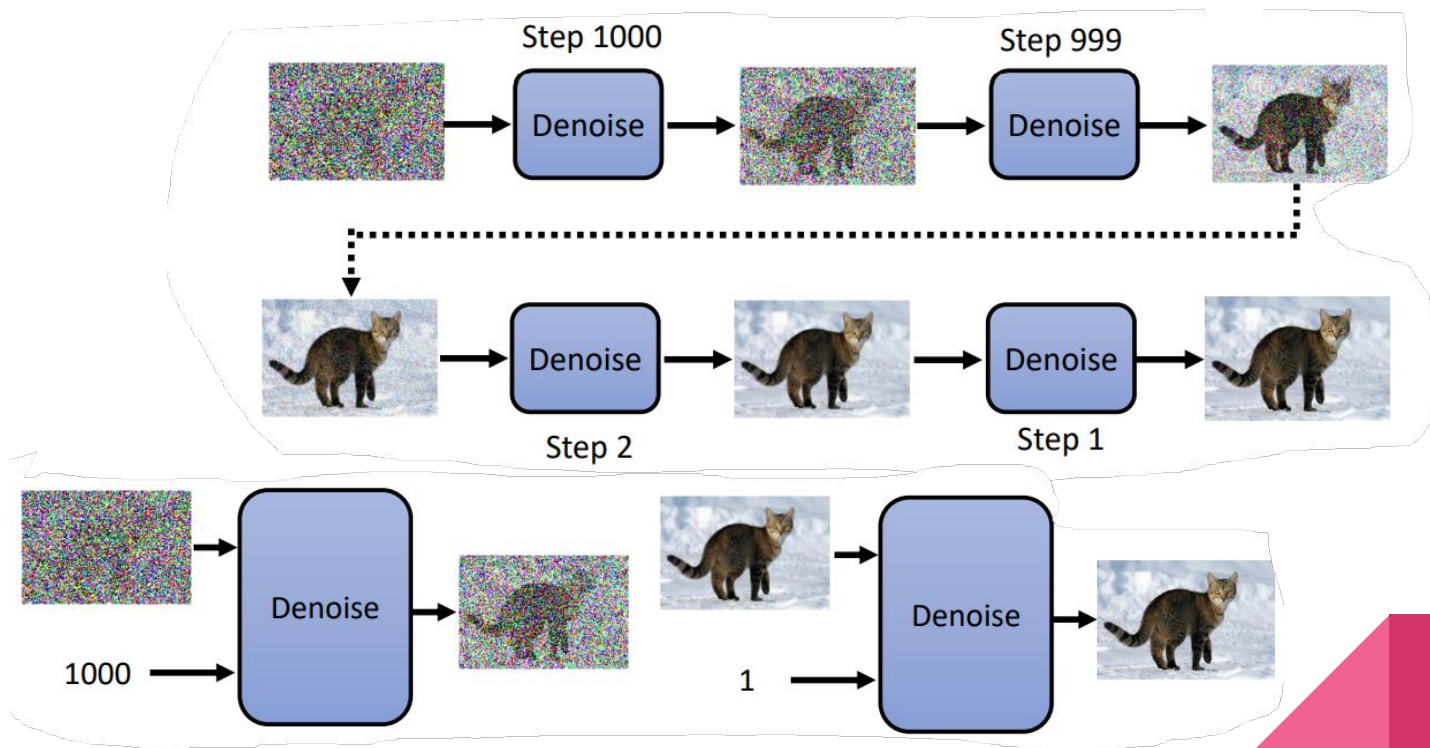ClipText (A GPT-based model) or BERT can be used as a text encoder.
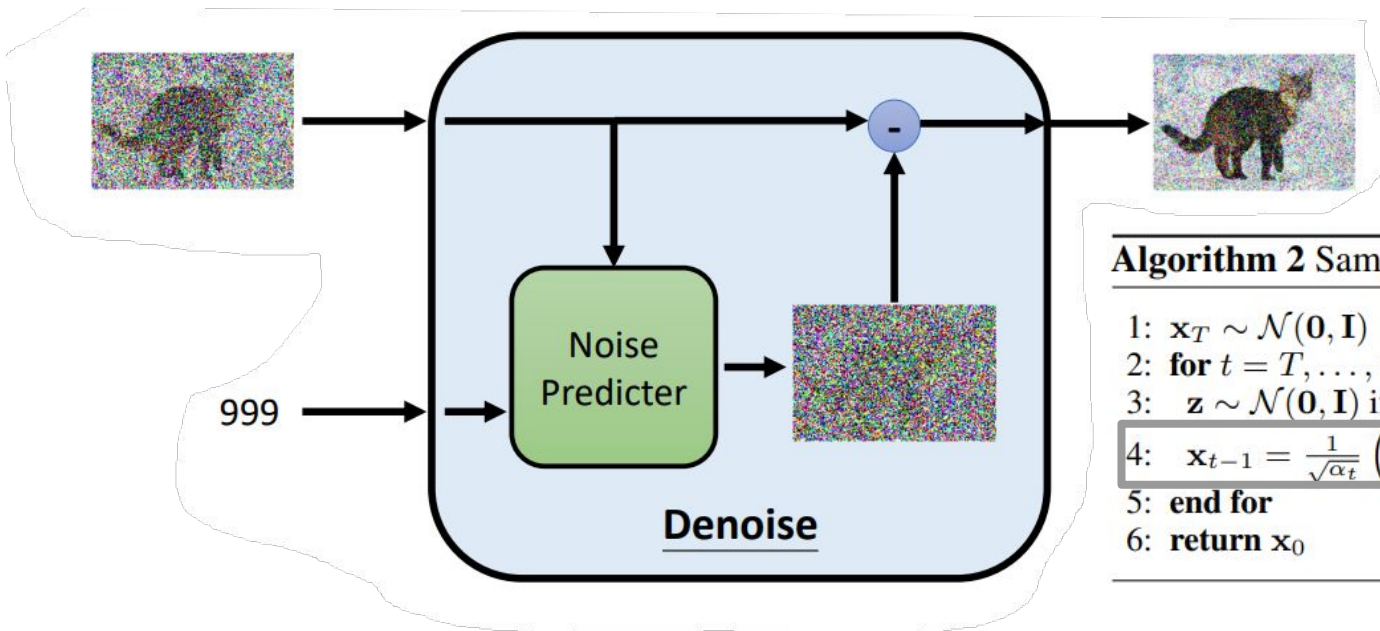

(a) Impact of encoder size.

(b) Impact of U-Net size.

Text-to-Image model structure - 2. Generation Model

Reference:

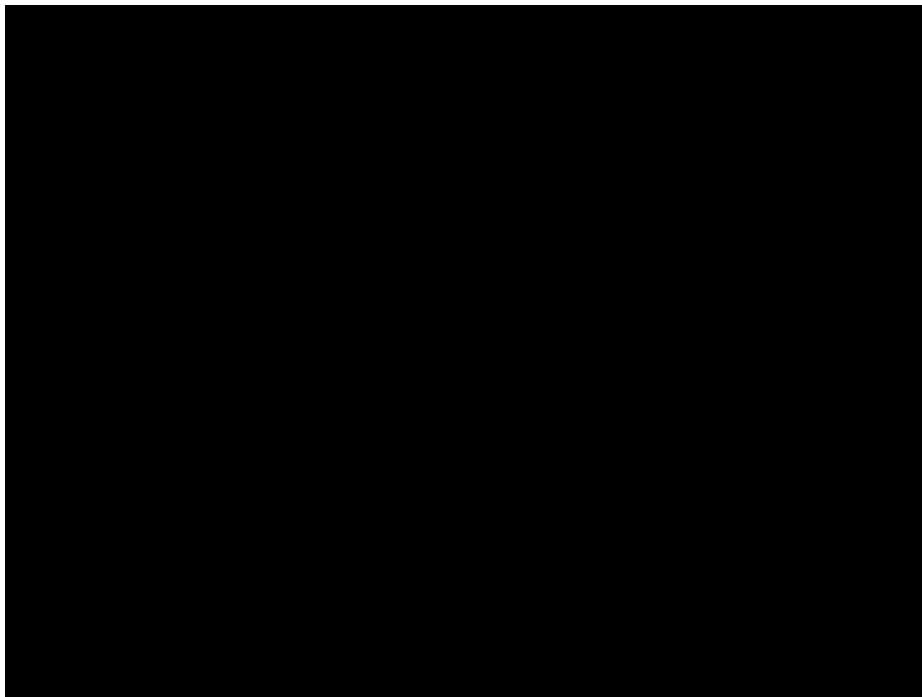# Text-to-Image model structure - 2. Generation Model

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\quad \mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

Reference:
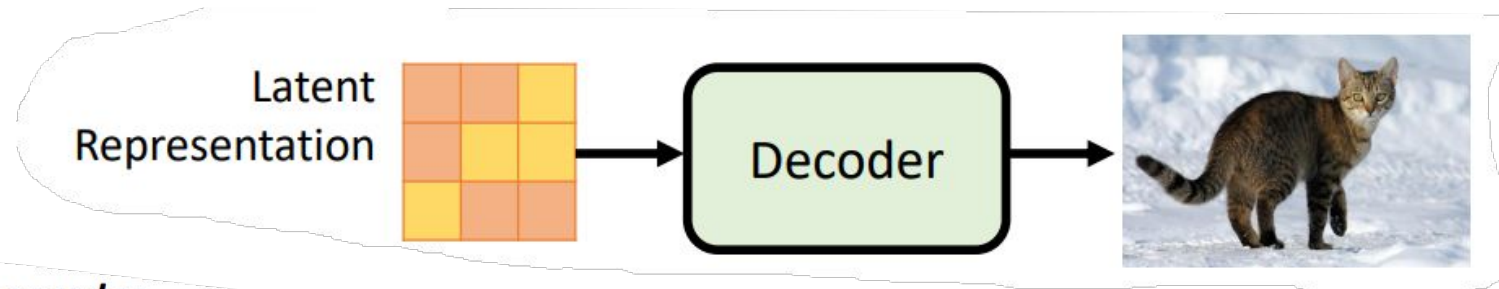[1] HUNG-YI LEE ML 2023 spring course - Diffusion Model
[2] Denoising Diffusion Probabilistic Models
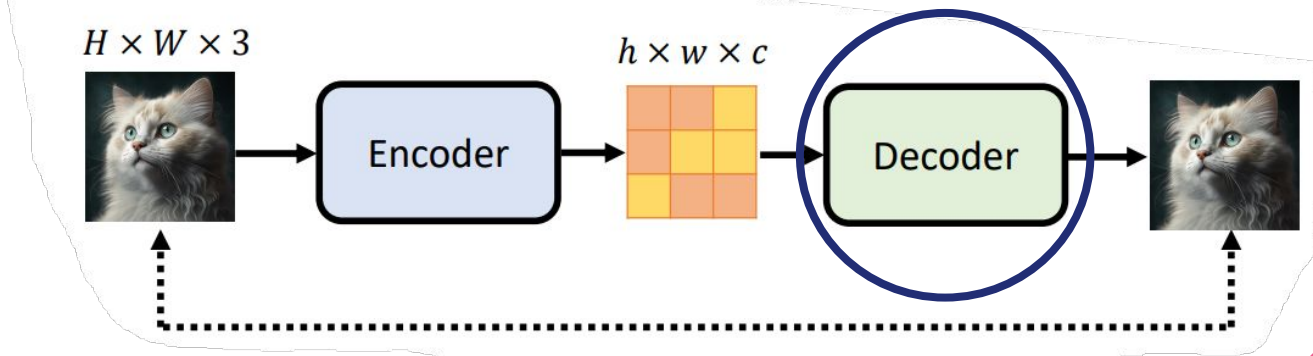
Text-to-Image model structure - 2. Generation Model

Reference: https://jalammar.github.io/illustrated-stable-diffusion/

Text-to-Image model structure - 3. Decoder

Low-Rank Adaptation(LoRA) introduction

$$h' = h + \Delta h$$

Weight size: **over 3GB**

h

Feed forward layer
from transformer layer

d_{FFW}

Feed-forward
up-project

d_{model}

Δh

d_{FFW}

r

Weight size: **over 3MB**

default: 4 neurons

r
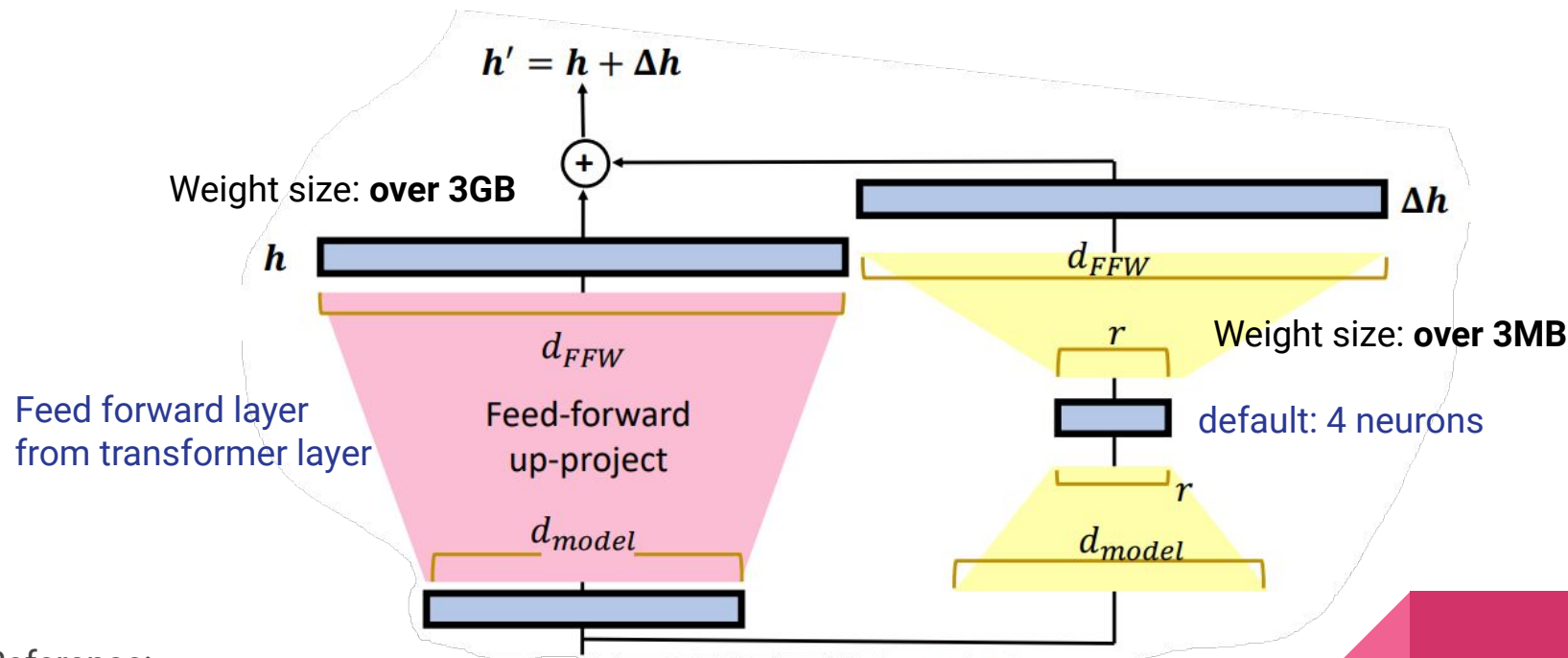
d_{model}

Reference:
[1] HUNG-YI LEE AACL 2022 Tutorial
[2] LORA: LOW-RANK ADAPTATION OF LARGE LANGUAGE MODELS

Comparison of Text-to-Image model prediction

- Story content:
  Fox Xiaohua and Turtle Aman had a race in the forest. Xiaohua ran fast but was overconfident and took a nap. Aman persisted and eventually won.
- Prompt1: '{story content}'
  Prompt2: 'manga, details line without color style: {story content}'
  Prompt3: 'manga, details line without color style: {story content} – niji'
  → Goal: the style of the illustrations from the three prompts can be identical.
- LoRA Fine-tune dataset: lambdalabs/pokemon-blip-captions(833 pokemon images with text)

image (image)	text (string)
	"a drawing of a green pokemon with red eyes"

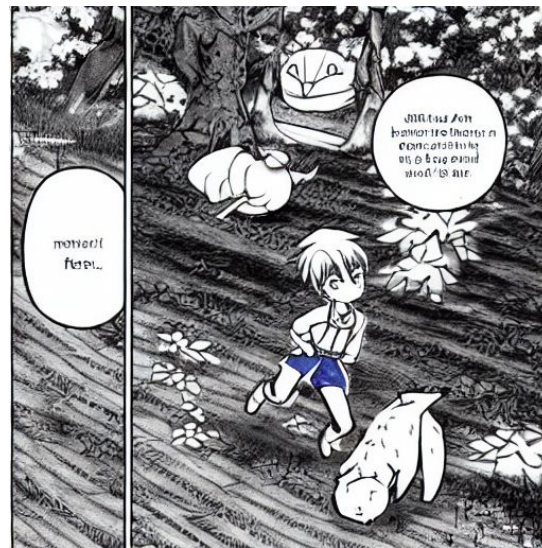# CompVis/stable-diffusion-v1-4 model prediction

Prompt1

Prompt2

Prompt3

The first prompt is highly different from the other two prompts.

[CompVis/stable-diffusion-v1-4](#) add pokemon LoRA
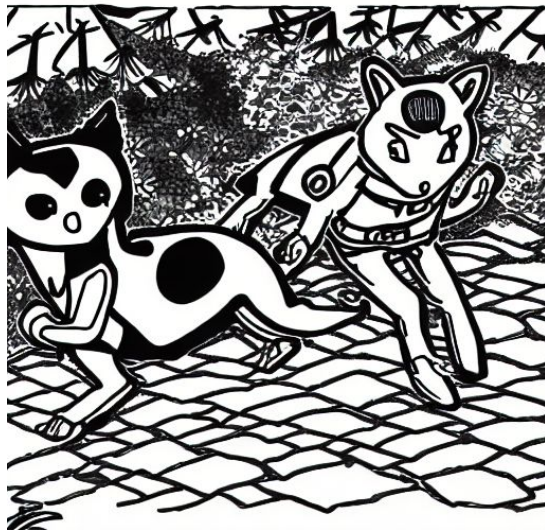
Prompt1

Prompt2

Prompt3

The style changes, but the content is weird.

CompVis/stable-diffusion-v1-4 add Children's Drawing LoRA

Prompt1

Prompt2

Prompt3

The style changes, but the content is weird.

Children's Stories model prediction

Prompt1

Prompt2

Prompt3

The style is similar, but combine human and animal.

gsdf/Counterfeit-V2.5 model prediction

Prompt1

Prompt2

Prompt3

The style is similar,
but some content combine human and animal.

gsdf/Counterfeit-V2.5 add pokemon LoRA

Prompt1

Prompt2

Prompt3

Only animals are in the three images,
and the style is quite the same!

CompVis/stable-diffusion-v1-4 pokemon LoRA training process

validation prompt="Fox Xiaohua and Turtle Aman had a race in the forest. Xiaohua ran fast but was overconfident and took a nap. Aman persisted and eventually won."

Before training   **epoch 1**   **epoch 5**

gsdf/Counterfeit-V2.5 pokemon LoRA training process

validation prompt="manga, details line without color style: Fox Xiaohua and Turtle Aman had a race in the forest. Xiaohua ran fast but was overconfident and took a nap. Aman persisted and eventually won.--niji"

Before training

epoch 1

epoch 5

Conclusion

1. When generating stories with ChatGPT, the same keywords would produce different stories each time.

2. Unlike the web version, the API doesn't remember past conversation. Thus, to continue a story, previous context is needed in the prompt. Eventually, the prompt will exceed the API token limit.

3. LoRA fine-tuning is helpful to change the generation image style from text-to-image model in just one epoch.

4. It is difficult to generate continuous illustrations or two characters for a story via one text-to-image model. Models for Inputs **including text and image** might be a solution for generating continuous illustrations.

Platform demo

Story generation

Please enter some keywords, select a style, and a model to make a story!

Keywords, separated by commas

cat, tree

Story content

Once upon a time, a cat climbed a tree. She was thrilled to see the beautiful view.

Flag

Style

○ Adventure   ● Happy   ○ Sad   ○ Warm

Model Choice

● OpenAI GPT-3.5 Turbo   ○ OpenAI text-davinci-003

○ OpenAI text-davinci-002

Clear   Submit

Use via API 🚀 · Built with Gradio 🧱

For creating story

31

For creating illustration from a story

Keywords, separated by commas

cat, tree

Style

(•) Adventure   ( ) Happy   ( ) Sad   ( ) Warm

Model Choice for Story

(•) OpenAI GPT-3.5 Turbo   ( ) OpenAI text-davinci-003

( ) OpenAI text-davinci-002

Model Choice for Illustration

(•) Counterfeit-V2.5 add pokemon LoRA

( ) stable-diffusion-v1-4 add Children's Drawing LoRA   ( ) stable-diffusion-v1-4

Story content

Once upon a time, a curious cat climbed a tall tree. As she reached the top, she saw a treasure hidden in a hole. Excited, she jumped down and ran to her adventurous friends. Together, they found the treasure and celebrated their bravery.

Story illustration

Creating story and illustration