

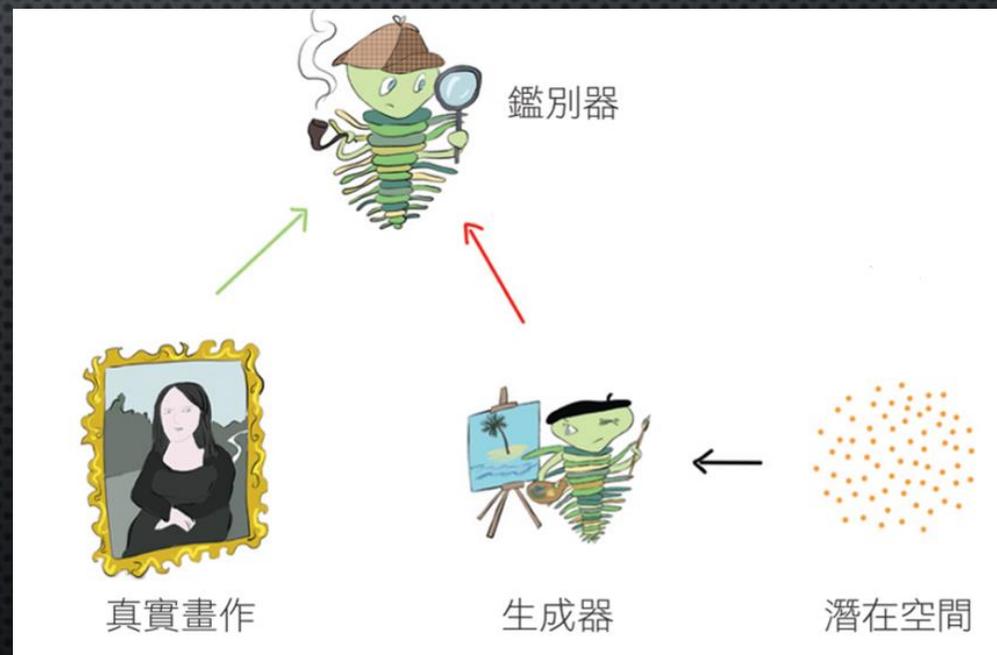
AI人工智慧— 藝術生成—Dall-e與Stable Diffusion



目前最流行的是GAN?是
Diffusion model吧?

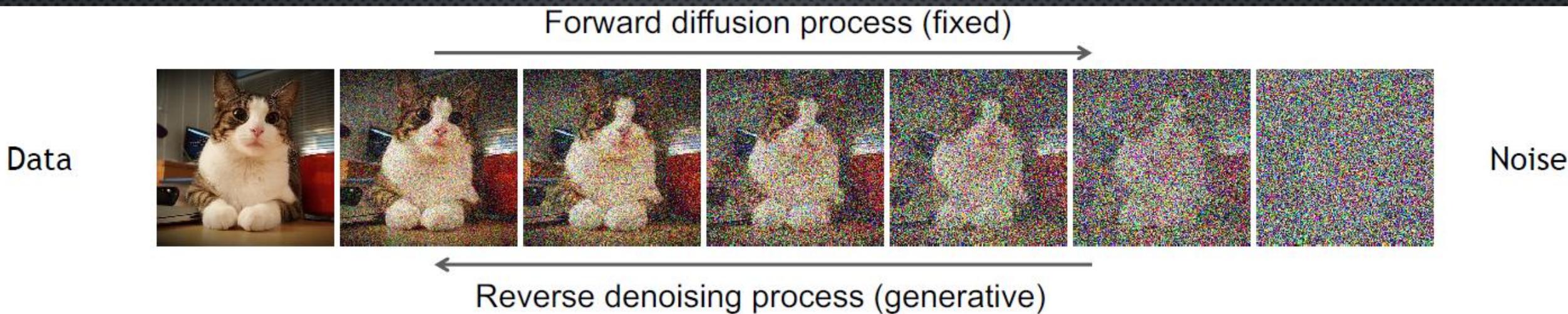
von anwendeng

GAN的基本要件 G, D Latent SPACE



Denoising Diffusion Models (DDMs)從左到右，圖像逐漸添加噪聲，直到變成完全隨機的噪聲。然後，模型學習逆轉這個過程，逐漸從圖像中去除噪聲，直到其恢復到原始狀態

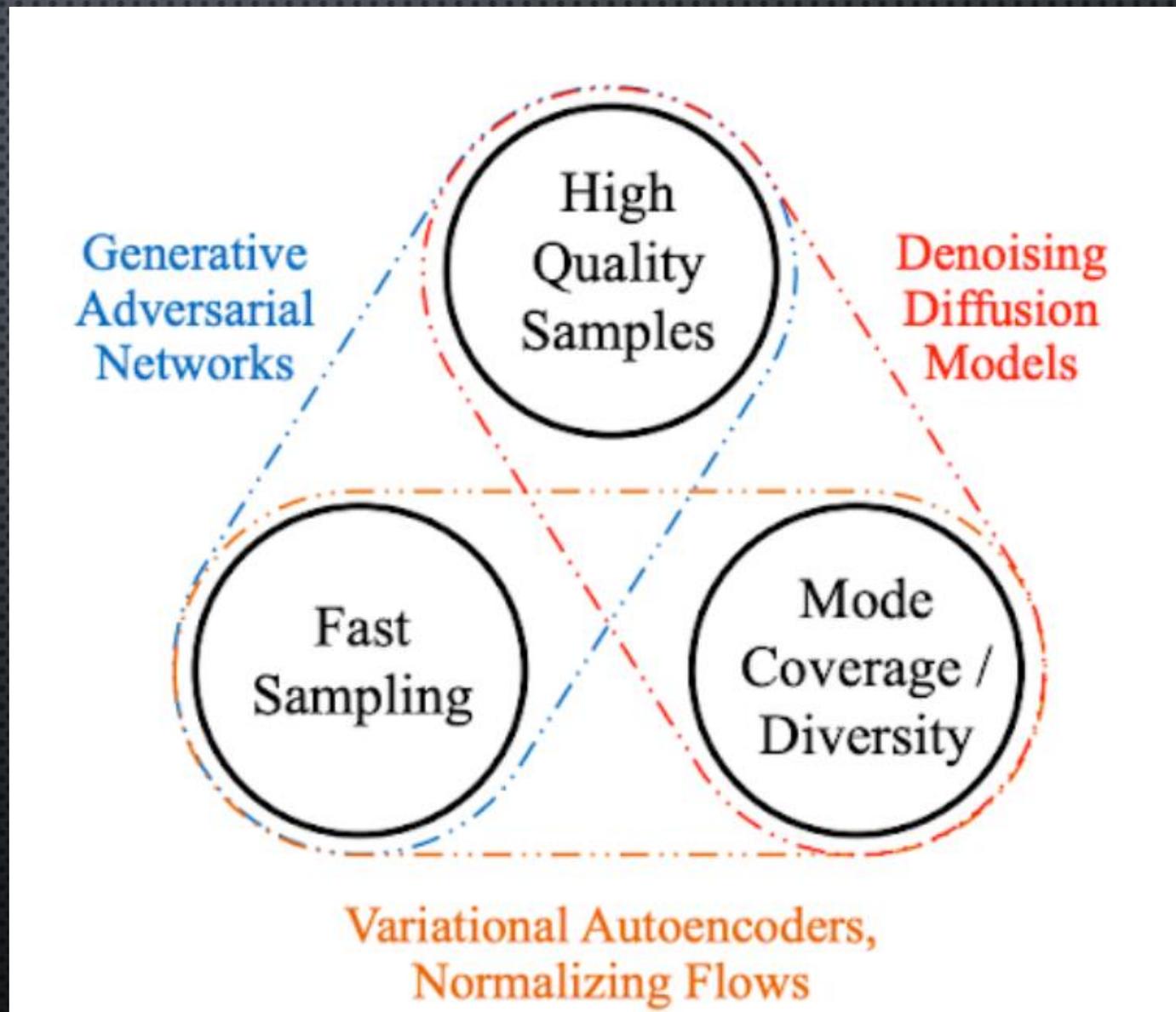
<https://cvpr2022-tutorial-diffusion-models.github.io/>



特徵	Denoising Diffusion Models (DDMs)	Generative Adversarial Networks (GANs)
工作原理	逐漸向圖像中添加噪聲，然後學習逆轉這個過程	使用兩個模型：生成器和判別器，生成器負責生成新的圖像，判別器負責區分真實圖像和生成圖像
優點	訓練更容易，不太容易過擬合，可以生成高品質的圖像，尤其是在高解析度下，可以用來執行各種圖像處理任務	可以生成高品質的圖像，尤其是在創意方面，可以用來生成多樣化的圖像
缺點	生成圖像的速度可能較慢，在一些情況下，可能會生成不自然或不逼真的圖像	訓練可能很困難，容易過擬合，在一些情況下，可能會生成模糊或不清晰的圖像
應用	圖像生成、圖像修復、超解析度、去噪	圖像生成、創意內容生成
適合應用	需要高品質和逼真度的應用	需要創意性的應用

Improving Diffusion Models as an Alternative To GANs, Part 1 (Apr 26, 2022 By Arash Vahdat and Karsten Kreis)

<https://developer.nvidia.com/zh-cn/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>



- **Stable Diffusion** 和 **DALL-E** 都是 **Denoising Diffusion Models (DDMs)**。DDMs 是一種生成模型，它可以用來生成圖像、修復圖像，以及執行其他圖像處理任務。DDMs 的工作原理是逐漸向圖像中添加噪聲，直到其變成完全隨機的噪聲。然後，模型學習逆轉這個過程，逐漸從圖像中去除噪聲，直到其恢復到原始狀態

一個簡化的 DDM 工作原理的例子：

- 從一個清晰的圖像開始。
- 向圖像中添加噪聲，直到其變成完全隨機的噪聲。
- 訓練一個模型來逆轉這個過程，逐漸從圖像中去除噪聲，直到其恢復到原始狀態。
- 要生成一個新的圖像，從一個隨機噪聲圖像開始，並使用模型逐漸去除噪聲，直到生成所需的圖像。
- DDMS 仍在開發中，但它們有可能徹底改變我們處理和生成圖像的方式。



StableDiffusionV2



```
[6] 1 from keras_cv.models import StableDiffusionV2  
    2 from PIL import Image
```

```
[7] 1 model = StableDiffusionV2(img_height=512, img_width=512 ,jit_compile=True)
```

By using this model checkpoint, you acknowledge that its usage is subject to the term

Stable Diffusion vs DALL-E

特徵	Stable Diffusion	DALL-E
類型	Denoising Diffusion Model	Denoising Diffusion Model
開源	是	否
神經網絡	UNet	Transformer
應用	圖像生成、圖像修復、超解析度等	圖像生成

“a propaganda poster depicting a black shiba inu dressed as Chinese emperor holding a piece of cheese”

by
<https://www.bing.com/images/create/Dall-E3>



Stable Diffusion 主要論文, 文中自稱 'latent diffusion models'

High-Resolution Image Synthesis with Latent Diffusion Models

Robin Rombach¹ * Andreas Blattmann¹ * Dominik Lorenz¹ Patrick Esser^ℜ Björn Ommer¹

¹Ludwig Maximilian University of Munich & IWR, Heidelberg University, Germany ℜRunway ML

<https://github.com/CompVis/latent-diffusion>

Abstract

By decomposing the image formation process into a sequential application of denoising autoencoders, diffusion models (DMs) achieve state-of-the-art synthesis results on image data and beyond. Additionally, their formulation allows for a guiding mechanism to control the image generation process without retraining. However, since these models typically operate directly in pixel space, optimization of powerful DMs often consumes hundreds of GPU days and inference is expensive due to sequential evaluations. To enable DM training on limited computational resources while retaining their quality and flexibility, we apply them in the latent space of powerful pretrained autoencoders. In contrast to previous work, training diffusion models on such a representation allows for the first time



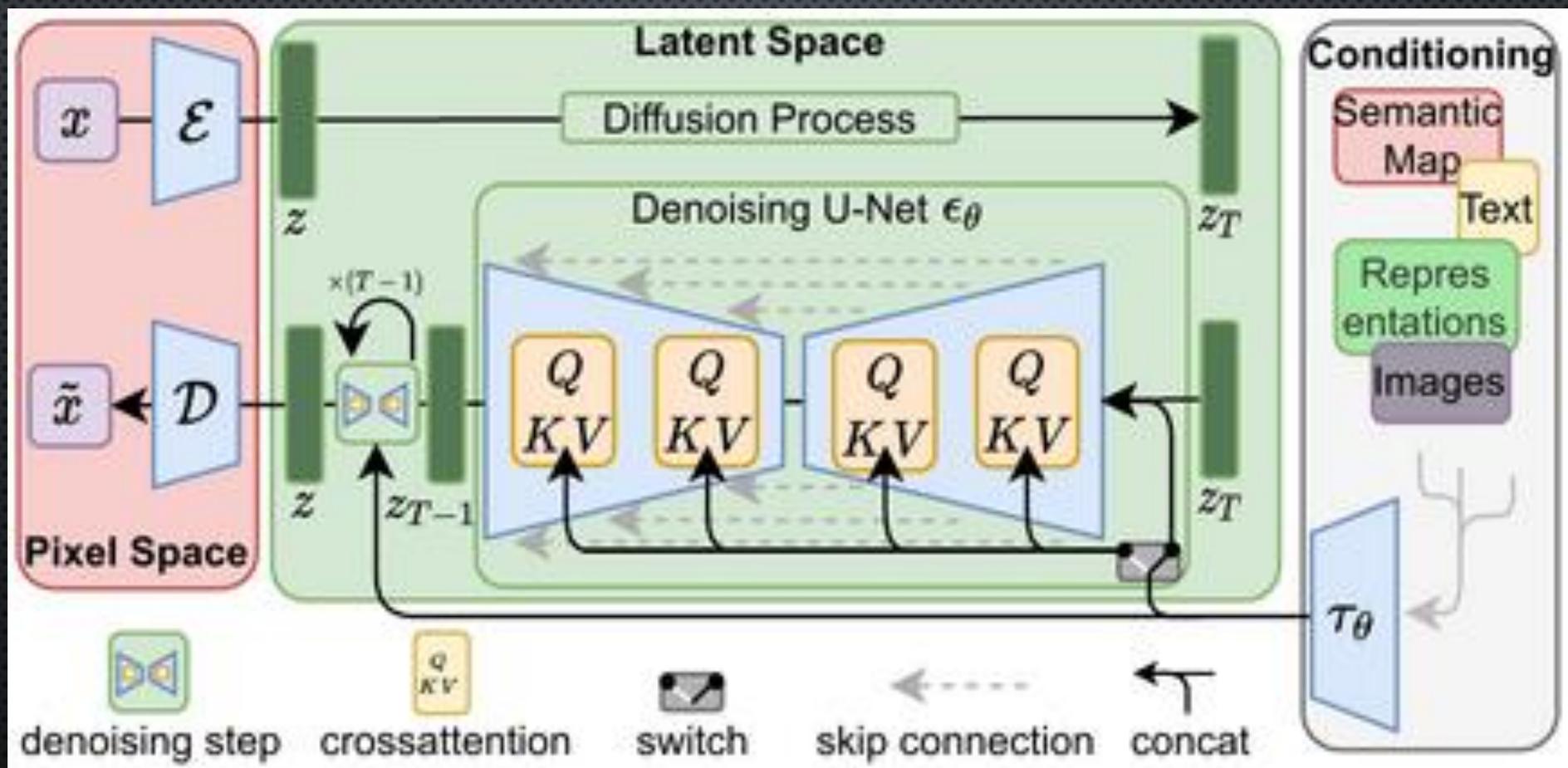
Figure 1. Boosting the upper bound on achievable quality with less aggressive downsampling. Since diffusion models offer excellent inductive biases for spatial data, we do not need the heavy spatial downsampling of related generative models in latent space, but can still greatly reduce the dimensionality of the data via suitable

cs.CV] 13 Apr 2022

- 透過將圖像生成過程分解為一系列去噪自編碼器的順序應用，擴散模型 (DMs) 在圖像數據和其他數據上實現了最先進的合成結果。此外，它們的公式允許使用引導機制來控制圖像生成過程，而無需重新訓練。然而，由於這些模型通常直接在像素空間中操作，因此強大 DM 的優化通常需要數百個 GPU 日，而且由於順序評估，推理成本也很高。
- 為了在有限的計算資源上實現 DM 訓練，同時保持其品質和靈活性，我們將它們應用於功能強大的預訓練自編碼器的潛在空間中。
- 與之前的工作相比，在這樣的表示上訓練擴散模型首次允許在複雜度降低和細節保留之間達到接近最佳的點，極大地提高了視覺保真度。
- 透過將交叉注意層引入模型架構，我們將擴散模型轉變為用于通用調理輸入（例如文本或邊界框）的強大而靈活的生成器，並以卷積方式實現高分辨率合成。我們的潛在擴散模型 (LDMs) 在圖像修復和類別條件圖像合成方面取得了新的最先進分數，並在各種任務上實現了高度競爭的性能，包括文本到圖像合成、無條件圖像生成和超分辨率，同時與基于像素的 DM 相比，計算要求大大降低。
- 簡而言之，潛在擴散模型 (LDMs) 是一種新的擴散模型，它在功能強大的預訓練自編碼器的潛在空間中進行訓練。這使得 LDMs 可以在有限的計算資源上進行訓練，同時保持擴散模型的品質和靈活性。LDMs 在各種任務上取得了新的最先進的成績，包括圖像修復、類別條件圖像合成、文本到圖像合成、無條件圖像生成和超分辨率。

使用 diffusion models 進行高分辨率圖像合成的優勢。diffusion models 是一種新的生成模型，可以產生逼真的圖像，而不需要像其他生成模型那樣進行 heavy spatial downsampling。這意味著 diffusion models 可以產生更高品質的圖像。





Stable Diffusion test-to-Image

Text-to-Image Synthesis on LAION. 1.45B Model.

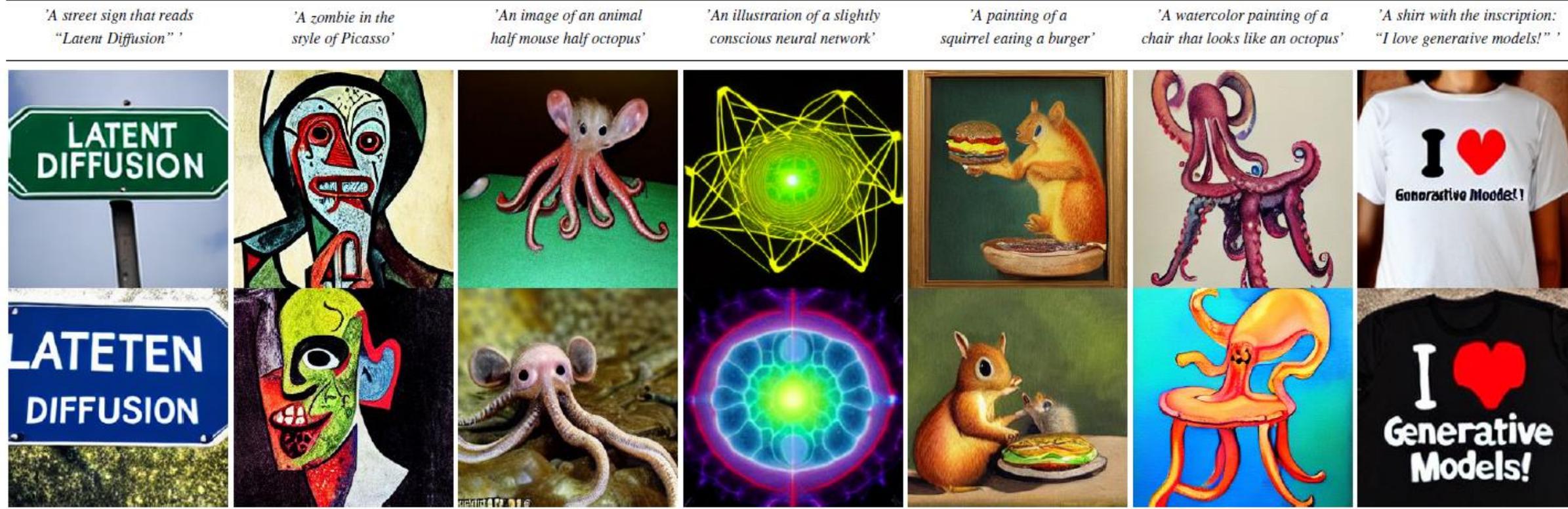


Figure 5. Samples for user-defined text prompts from our model for text-to-image synthesis, *LDM-8 (KL)*, which was trained on the LAION [78] database. Samples generated with 200 DDIM steps and $\eta = 1.0$. We use unconditional guidance [32] with $s = 10.0$.

bicubic：一種經典的圖像插值算法

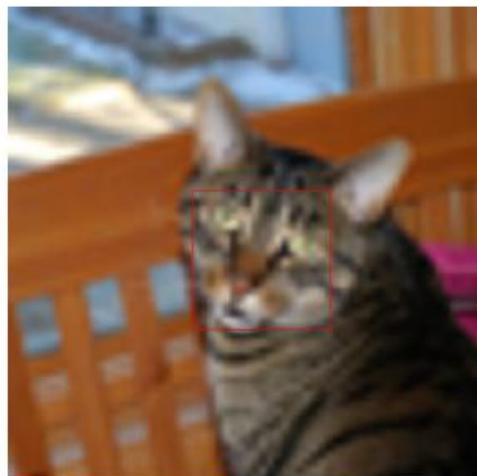
LDM-SR：一種基於潛在擴散模型的超解析度算法

SR3：一種基於 GAN 的超解析度算法

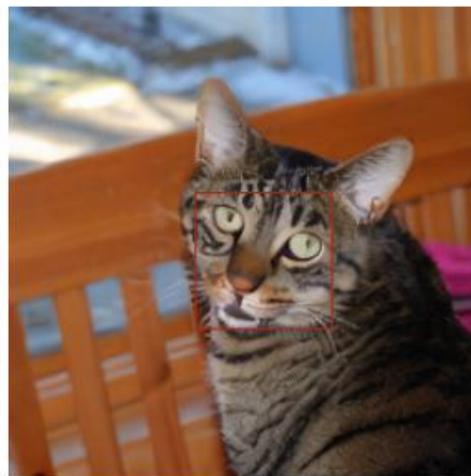
LDM-SR 算法在保留細節的同時能夠有效地提高圖像的解析度。

SR3 算法在生成更細膩的紋理方面表現更好，但在保留細節方面略遜於 LDM-SR。

bicubic



LDM-SR



SR3

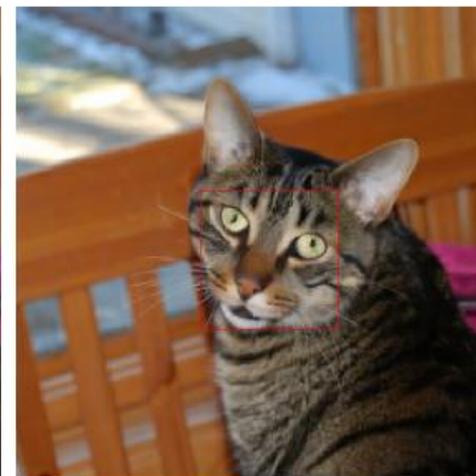


Figure 10. ImageNet 64→256 super-resolution on ImageNet-Val. *LDM-SR* has advantages at rendering realistic textures but SR3 can synthesize more coherent fine structures. See appendix for additional samples and cropouts. SR3 results from [72].

Stable Diffusion 2.1

Stable Diffusion 2.1 Demo

Stable Diffusion 2.1 is the latest text-to-image model from StabilityAI. [Access Stable Diffusion 1 Space here](#)

For faster generation and API access you can try [DreamStudio Beta](#).

a propaganda poster depicting a black shiba inu dressed as Chinese emperor holding a piece of cake

low quality

Generate image



Waifu Diffusion 模型專門用來生成高品質二次元動漫角色圖

→ ↻ 🏠 huggingface.co/hakurei/waifu-diffusion?text=masterpiece%2C+best+quality%2C+1girl%2C+blonde+hair%2C+black+leathe

⚡ **Hosted inference API** ⓘ

🔗 Text-to-Image

masterpiece, best quality, 1girl, blonde hair, black leather jack, looking at viewer, upper body, outdoors, oilcolor, night, turtlene

Computation time on gpu: 5.826 s



感謝觀賞

Herzlichen Dank für die
Aufmerksamkeit

von anwendeng