|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Academic Year: | 2022/2023 | Term: | Spring 2023 |  |
| Course Code: | ELC 4028 | Course Title: | Artificial Neural Networks and its Applications |

Cairo University

Faculty of Engineering

Electronics and Communications Engineering Department – 4th Year

Neural Networks Applications

- Text-to-Image synthesis using Stable Diffusion -

*Submitted to: Dr. Mona Riad*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | BN | Sec | ID | رقم الجلوس |
| احمد محمود حسيني عطية | 22 | 1 | 9180178 | 34022 |
| علي ماهر عبدالسلام نبيه | 5 | 3 | 9190067 | 34117 |
| محمد احمد طه السيد | 30 | 3 | 9191043 | 34142 |
| محمد حسام عثمان يسن | 35 | 3 | 9191083 | 34147 |
| محمد عاطف ربيع | 43 | 3 | 9190924 | 34155 |

Table of Contents

[Abstract 1](#_Toc132165948)

[1. Introduction 2](#_Toc132165949)

[2.Architecture 2](#_Toc132165950)

[References 4](#_Toc132165951)

**List of Figures**

Figure 1: text to image examples by Lexica 2

**List of Abbreviations**

# Abstract

Text to image synthesis is a process of generating an image from a given text description. It is a form of natural language processing that involves the use of deep learning algorithms to generate an image from a text description. The goal of text to image synthesis is to create an image that accurately reflects the content of the text. In this report, we will discuss a paper called “High-Resolution Image Synthesis with Latent Diffusion Models” [1] which uses Stable Diffusion (SD) also called Latent Diffusion Model (LDM) to achieve new state-of-the-art scores for image inpainting and class-conditional image synthesis and highly competitive performance on various tasks, including text-to-image synthesis, unconditional image generation and super-resolution, while signiﬁcantly reducing computational requirements compared to the previous pixel-based Diffusion Models (DM).

# 1. Introduction

Artificial Intelligence (AI) art has become extremely popular in recent years as technology advances and more people became interested in exploring the possibilities of using AI to create artwork. The Public release of SD in August 22, 2022 [2] has massively impacted this field as it’s not only a high-performance model competitive with present AI image generation models such as DALL-E by OpenAI or Imagen by Google but also its model weights and source code are fully open to anyone which allows anyone to download the model and tinker with it and adjust the internal parameters in a way that they can’t do with the closed solutions as DALL-E and Imagen. [3]. In addition, as we will discuss in this report, the latent space approach allowed relatively low resources and memory requirements [1].

The most common way of generation AI art from text is by typing a prompt to the model and it will generate the image, The best platform for finding examples and the prompts used to generate images is Lexica [4], which archives over 10 million sample artworks. Each artwork includes its full prompt and the seed number, some examples of images and their prompt are shown below

A realistic cute adorable baby owl made of crystal ball with low poly eye's surrounded by glowing aura highly detailed intricated concept art with vivid beautiful colors trending artstation 8k

A cute wizard elven, blue long hair, golden eyes, in a forest, simple dress, intricate, elegant, highly detailed, digital painting, hyper realistic, fantasy, dungeons and dragons, art by artgerm and greg rutkowski



streets of 1980s pripyat with neon lights in a foggy rainy night



Shabby Chic Teal and Pink Flowers with Fluffy Realistic Rabbit

Figure : text to image examples by Lexica

In addition, painting images from text isn’t the only feature of SD but also image to image and object removal is also possible.

# 2.Architecture

Stable Diffusion uses a kind of diffusion model (DM), called a latent diffusion model (LDM) developed by the CompVis group at LMU Munich. Diffusion models are trained with the objective of removing successive applications of Gaussian noise on training images which can be thought of as a sequence of denoising autoencoders. Stable Diffusion consists of 3 parts: the variational autoencoder (VAE), U-Net, and an optional text encoder. The VAE encoder compresses the image from pixel space to a smaller dimensional latent space, capturing a more fundamental semantic meaning of the image. Gaussian noise is iteratively applied to the compressed latent representation during forward diffusion. The U-Net block, composed of a ResNet backbone, denoises the output from forward diffusion backwards to obtain a latent representation. Finally, the VAE decoder generates the final image by converting the representation back into pixel space. The denoising step can be flexibly conditioned on a string of text, an image, or another modality. The encoded conditioning data is exposed to denoising U-Nets via a cross-attention mechanism. For conditioning on text, the fixed, pretrained CLIP ViT-L/14 text encoder is used to transform text prompts to an embedding space. Researchers point to increased computational efficiency for training and generation as an advantage of LDMs.

# References

|  |  |
| --- | --- |
| [1] | R. Rombach, A. Blattmann, D. Lorenz, P. Esser and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models | Papers With Code," 2022. [Online]. Available: https://paperswithcode.com/paper/high-resolution-image-synthesis-with-latent. [Accessed 2023]. |
| [2] | Stability AI, "Stable Diffusion Public Release — Stability AI," [Online]. Available: https://stability.ai/blog/stable-diffusion-public-release. [Accessed 2023]. |
| [3] | Two Minute Papers, "Stable Diffusion: DALL-E 2 For Free, For Everyone! - YouTube," 6 Septemper 2022. [Online]. Available: https://youtu.be/nVhmFski3vg. [Accessed 2023]. |
| [4] | "Lexica," [Online]. Available: https://lexica.art/. [Accessed 2023]. |
| [5] | J. Alammar, "The Illustrated Stable Diffusion," 2022. [Online]. Available: https://jalammar.github.io/illustrated-stable-diffusion/. [Accessed 2023]. |
| [6] | CompVis, "High-Resolution Image Synthesis with Latent Diffusion Models - Computer Vision & Learning Group," [Online]. Available: https://ommer-lab.com/research/latent-diffusion-models/. [Accessed 2023]. |
| [7] | Stability AI, "Stable Diffusion v2.1 and DreamStudio Updates 7-Dec 22 — Stability AI," [Online]. Available: https://stability.ai/blog/stablediffusion2-1-release7-dec-2022. [Accessed 2023]. |
| [8] | Stability AI, "Stable Diffusion 2-1 - a Hugging Face Space by stabilityai," [Online]. Available: https://huggingface.co/spaces/stabilityai/stable-diffusion. [Accessed 2023]. |
| [9] | CompVis, "CompVis/latent-diffusion: High-Resolution Image Synthesis with Latent Diffusion Models," [Online]. Available: https://github.com/CompVis/latent-diffusion.git. [Accessed 2023]. |

[10] Wikipedia Contributors, “Stable Diffusion,” *Wikipedia*, Apr. 11, 2023. [Online]. Available: https://en.wikipedia.org/wiki/Stable\_Diffusion. [Accessed: Apr. 12, 2023].

‌