|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 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](#_Toc133512614)

[1. Introduction 1](#_Toc133512615)

[1.1. Motivation for Stable Diffusion 2](#_Toc133512616)

[2. Overview of Stable Diffusion Main Components 2](#_Toc133512617)

[2.1. Text Encoder 2](#_Toc133512618)

[2.2. Image Information Creator 2](#_Toc133512619)

[2.2.1. Diffusion 3](#_Toc133512620)

[2.3. Image Decoder 4](#_Toc133512621)

[3. Diffusion Model Training 4](#_Toc133512622)

[4. Diffusion Model Testing 4](#_Toc133512623)

[5. Why Latent Space Instead of Pixel Space? 5](#_Toc133512624)

[6. The Text Encoder 6](#_Toc133512625)

[6.1. CLIP Training 6](#_Toc133512626)

[6.2. Merging Text with Diffusion Process 7](#_Toc133512627)

[7. Results 8](#_Toc133512628)

[7.1. Compression Trade-off 8](#_Toc133512629)

[7.2. Image Generation with Latent Diffusion 8](#_Toc133512630)

[7.3. Conditional Latent Diffusion 9](#_Toc133512631)

[7.4. Super-Resolution with Latent Diffusion 10](#_Toc133512632)

[7.5. Inpainting with Latent Diffusion 11](#_Toc133512633)

[8. Social Impact of AI Art 12](#_Toc133512634)

[9. Conclusion 12](#_Toc133512635)

[10. References 13](#_Toc133512636)

**List of Figures**

Figure 1: text to image examples by Lexica 1

Figure 2: internal architecture of SD 2

Figure 3: diffusion example from a noise pattern then slowly morphs into a desired output 3

Figure 4: visualizing outputs by a sweep over steps parameter 3

Figure 5: creating a noise predictor model from dataset 4

Figure 6: denoising or testing process of diffusion model 5

Figure 7: block diagram of full SD process 5

Figure 8: CLIP dataset 6

Figure 9: CLIP training cycle 7

Figure 10: U-Net with cross attention to include text 7

Figure 11: samples of datasets used for LDMs training 8

Figure 12: sample quality as a function of training progress for 2M steps on the ImageNet dataset 8

Figure 13: Super Resolution experiment 10

Figure 14: inpainting examples 11

**List of Tables**

[Table 1: summary of Stable Diffusion main components 4](#_Toc133512651)

[Table 2: summary of training of unconditional models 9](file:///D:\Ali\College\Neural%20Networks\ELC-4028-NN\Presentation%20dr-mona\txt2img-SD-report.docx#_Toc133512652)

[Table 3: Evaluation of text-conditional image synthesis on the 256x256-sized MS-COCO dataset 9](file:///D:\Ali\College\Neural%20Networks\ELC-4028-NN\Presentation%20dr-mona\txt2img-SD-report.docx#_Toc133512653)

[Table 4: additional comparisons with recent state-of-the art models with LDMs trained on COCO and OpenImages finetuned on COCO 10](file:///D:\Ali\College\Neural%20Networks\ELC-4028-NN\Presentation%20dr-mona\txt2img-SD-report.docx#_Toc133512654)

[Table 5: FID and LPIP scores of LDM and LaMa 11](file:///D:\Ali\College\Neural%20Networks\ELC-4028-NN\Presentation%20dr-mona\txt2img-SD-report.docx#_Toc133512655)

[Table 6: User Study for SR and Inpainting 11](file:///D:\Ali\College\Neural%20Networks\ELC-4028-NN\Presentation%20dr-mona\txt2img-SD-report.docx#_Toc133512656)

# Abstract

# 1. Introduction