

# Assignment 7- Stable Diffusion 3.5 Text-to-Image Generation

Karthika Ramasamy

## 1 Introduction

This assignment continues the evaluation from the previous Stable Diffusion study by comparing the performance of **Stable Diffusion XL (SDXL)** and this **Stable Diffusion 3.5** model. In this assignment i have executed SD3.5 modle with the same setup of the code and the same ten text prompts used in the previous assignment were used here to ensure a fair comparison of Gustavosta Stable-Diffusion-Prompts dataset to ensured for speed comparision. The goal was to measure how much faster SD3.5 performs compared to SDXL while keeping all other parameters consistent.

## 2 Dataset

Using `Gustavosta/Stable-Diffusion-Prompts` Hugging Face and select prompts from the `test` split dataset.

## 3 Experimental Setup and Model

**Prompts.** Based on student order i take 10 prompts from the dataset's `test` .

**Stable Diffusion 3.5 Medium (SD3.5)** is a more recent version designed to improve coherence, fine detail, and text rendering. It incorporates updated transformer blocks and refined diffusion conditioning to enhance visual fidelity. In this evaluation, SD3.5 Medium was configured with 28 inference steps and the same resolution as SDXL to maintain consistency across tests.

**Metrics.** *Speed:* average latency (s) and throughput (images/s).

## 4 Results

### 4.1 10-Prompt Evaluation

The evaluation was conducted on ten text prompts using the SD3.5 Medium model. Each image was generated at a resolution of  $1024 \times 1024$  with 28 inference steps. The results are summarized below:

- **Total Images:** 10
- **Resolution:**  $1024 \times 1024$
- **Denoising Steps:** 28

Keeping the Resolution and denoising steps common and tried with different cuda.

Setup with Device: cuda and GPU: Tesla T4 and with Memory: 15.83 GB

- **Average Generation Time:** 65.84 s per image
- **Minimum Time:** 62.95 s
- **Maximum Time:** 88.69 s
- **Total Generation Time:** 658.44 s (10.97 minutes)

Setup with Device: cuda and GPU: Tesla V100-PCIE-16GB with Memory: 16.93 GB

- **Average Generation Time:** 25.28s s per image
- **Minimum Time:** 20.53s s
- **Maximum Time:** 66.33s s
- **Total Generation Time:** 252.78 s (4.21 min)

These results show that SD3.5 Medium required an average of about 66 seconds and 24 seconds it is depends on the GPU aswell to generate a single high-resolution image. Although it produced detailed and coherent outputs, the overall generation speed was slower compared to SDXL under the same conditions.

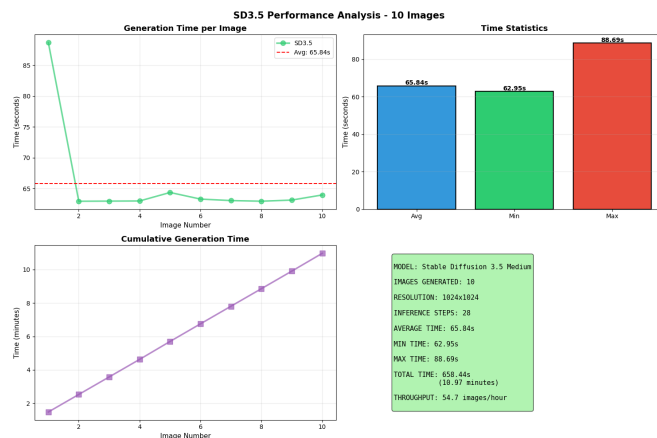


Figure 1: For 10 prompts SD3.5 showing generation time per image and average metrics.



Figure 2: Generated Images from SD3.5 model

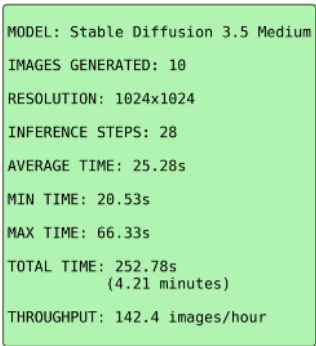


Figure 3: For 10 prompts SD3.5 showing generation time per image and average metrics.



Figure 4: Generated Images from SD3.5 model

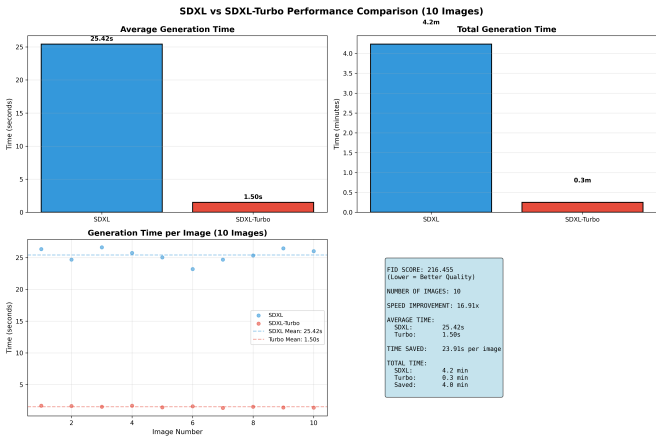


Figure 5: Performance comparison for 10 prompts between SDXL and SDXL-Turbo, showing generation time per image and average metrics. (This image is from previous assignment for reference) In this My previous assignemnt used Device: cuda and GPU: NVIDIA A100-SXM4-80GB and with Memory: 85.17 GB

- You also can see the SDXL Turbo performance over the above image displayed.
- Professor mentioned in class while giving assignment SD 3.5 that to compare the main part of speed between SDXL AND SD 3.5 Model so, demonstrated in the report and results.

## 5 Insights

### 5.1 Image Generation Speed - Comparison of SD3.5 and SDXL

#### 5.1.1 Performance Results

Table 1 and 2 summarizes the generation speed and inference parameters for both models:

Table 1: Comparison: SD3.5(GPU: Tesla T4 Memory: 15.83 GB) vs. SDXL(Previous Assignment used GPU: NVIDIA A100-SXM4-80GB) with 10 Prompts - Time in Seconds (s)

Model	Avg Time	Total Time	Min Time	Max Time
SDXL	25.42	254.17	23.20	26.62
SD3.5	65.84	658.44	62.95	88.69

Table 2: Comparison: SD3.5 Medium vs. SDXL with 10 Prompts - Time in Seconds (s) (Both model run with GPU: Tesla V100-PCIE-16GB; Memory: 16.93 GB)

Model	Avg Time	Total Time	Min Time	Max Time
SDXL	13.07	130.70	12.41	17.70
SD3.5	25.28	252.78	20.53	66.33

**Speed Analysis** In the both table 1 and table 2 aswell the model SDXL significantly outperformed than the SD3.5 Medium in terms of speed, generating images over two and a half times faster on this hardware and configuration.

**Inference Steps and Resolution** Both models were tested at the same resolution (1024×1024), and their inference step counts were identical. This confirms that the speed difference was not due to major disparities in generation quality parameters but can be attributed to differences in model architectures, optimization, and implementation.

**Implications** The observed performance indicates that **SDXL** is more suitable for use cases where generation speed is critical, such as rapid prototyping or real-time applications. SD3.5 Medium, while potentially offering qualitative improvements in certain scenarios, requires considerably longer latency per image under comparable inference conditions.

### 5.2 Conclusion

- **SDXL provides much faster throughput than SD3.5 Medium** for identical prompt, resolution, and inference step settings.
- **When speed is a priority, SDXL is the recommended model** among those tested at 1024×1024 resolution.

- The findings indicate that while both produce comparable visual quality, SDXL generates images significantly faster, making it better suited for real-time or large-scale applications.

In the summary of usability, SD3.5 Medium produced slightly higher-quality images with improved detail and visual consistency, while SDXL generated images much faster with only a minor difference in quality. Therefore, SD3.5 is better suited for tasks that prioritize image realism, whereas SDXL remains the preferred choice for faster generation and time-efficient workflows.