Assignment: SDXL Turbo vs SDXL Base Model Comparison for Text-to-Image Generation

UMAIMA KHAN

Abstract

This assignment evaluates two variants of Stable Diffusion XL—**SDXL Turbo** and **SDXL Base**—on a set of ten diverse prompts. The comparison focuses on three aspects: **generation quality** (via Fréchet Inception Distance, FID), **speed** (average image generation time), and **resource usage**. Results show that SDXL Turbo achieves over $45\times$ **faster** generation while maintaining competitive quality, whereas SDXL Base delivers higher-fidelity outputs at the cost of extreme latency and frequent out-of-memory issues on limited GPU hardware.

1 Introduction

Recent advances in diffusion-based generative models have enabled high-quality text-to-image synthesis. SDXL (Stable Diffusion XL) introduces larger backbones and higher training fidelity, with *Base* focusing on quality and *Turbo* optimized for efficiency. This study compares the two models on speed, quality, and memory feasibility under GPU constraints (RTX 4050, 6GB VRAM).

2 Experimental Setup

2.1 Models

- **SDXL Turbo** – lightweight, optimized for speed.

- **SDXL Base** – heavyweight, optimized for fidelity.

2.2 Environment

- Python 3.10, PyTorch 2.1.0+cu118

- Diffusers 0.25.1, Transformers 4.35.2

- Hardware: NVIDIA RTX 4050 (6GB VRAM, Windows 11)

2.3 Dataset

Ten prompts were sampled from `prompts.txt`, covering artistic, realistic, and fantasy styles.

2.4 Pipeline

Execution followed a sequential master script:

1. Generate images (`generate_images.py`)

2. Evaluate FID (`evaluate_fid.py`)

3. Evaluate speed (`evaluate_speed.py`)

4. Generate report (`generate_report.py`)

3 Results

3.1 Speed Analysis

Average timing results:

Model	Avg Time (s/img)	Throughput (img/s)
SDXL Turbo	15.47	0.0646
SDXL Base	705.58	0.0014

Table 1: Speed comparison across 10 prompts. Turbo is $45.6\times$ faster, saving $\sim$690s per image.

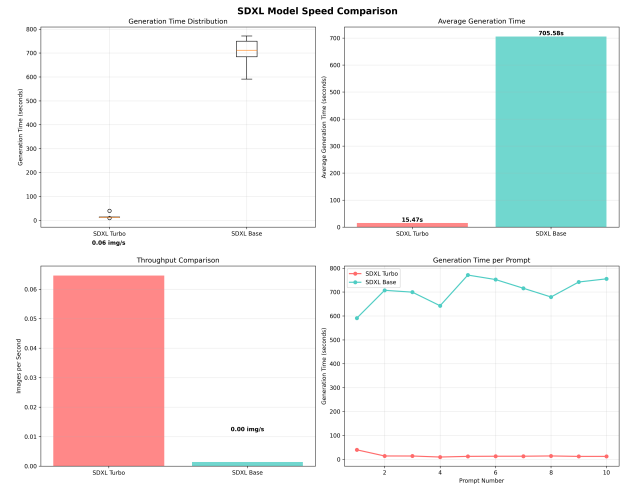Visualization is provided in Fig. 1.

Figure 1: Speed comparison: SDXL Turbo vs SDXL Base.

3.2 Quality (FID)

FID evaluation (lower = better):

Turbo's FID of 247.09 indicates a noticeable gap from Base, but still acceptable given Turbo's extreme speed advantage.

1

Model Comparison	FID Score
Turbo vs Base	247.09

Table 2: FID-based quality comparison. Smaller sample size (10 images) may affect reliability.

3.3 Resource Usage

- **SDXL Turbo**: Successfully ran all prompts without VRAM errors.

- **SDXL Base**: Multiple CUDA out-of-memory errors encountered on 6GB GPU; average runtime exceeded 11 minutes per image.

4 Discussion

4.1 Turbo vs Base Trade-offs

- **Turbo**: Extremely fast, feasible on mid-range GPUs, slightly lower fidelity.

- **Base**: Higher-quality outputs but impractical on 6GB GPU—runtime ~12 minutes per image, frequent OOM.

4.2 Insights

The gap highlights how deployment environment dictates model choice. Turbo fits interactive/real-time use, Base fits offline high-resource pipelines.

5 Conclusion

SDXL Turbo offers drastic speed improvements ($45\times$ faster) and stable inference under limited hardware. SDXL Base provides higher fidelity but is resource-prohibitive in this setting. Thus, Turbo is recommended for most practical applications where GPU memory is limited.

6 Future Work

- Evaluate on larger GPU memory setups (A100/4090) to fully test Base.

- Explore quantization/LoRA fine-tuning for memory efficiency.

- Benchmark on larger prompt sets for statistical robustness.

REFERENCES

[1] Podell et al., *SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis*, Stability AI, 2023. [2] Heusel et al., *GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium*, NIPS 2017.