Electrical & Computer Engineering Department

Ms. Computer Vision

University of Central Florida

Orlando, Florida

(CAP6411) COMPUTER VISION SYSTEM

Assignment # 7 – Stable Diffusion 3.5

Dr. Ser Nam Lim

2025 – 2026

Prepared By:

Darinka Townsend

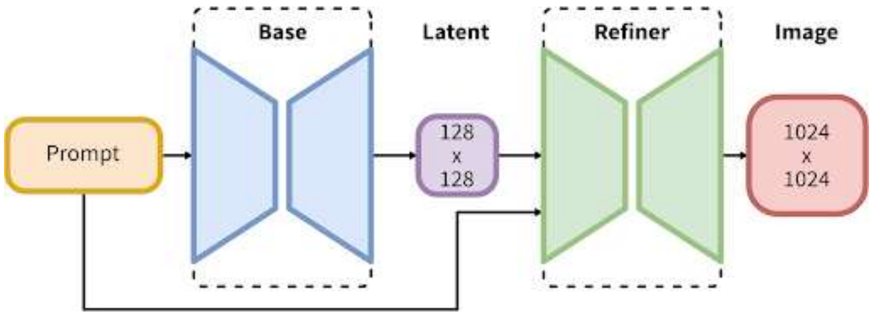
1. Objetive

Compare the performance of the SDXL Turbo, SDXL and SD 3.5 models using an assigned subset of prompts from the "Stable-Diffusion Prompts" dataset, evaluating their visual quality (FID) and generation speed, to determine which model offers a better balance between accuracy and efficiency.

2. Models

2.1 SDXL

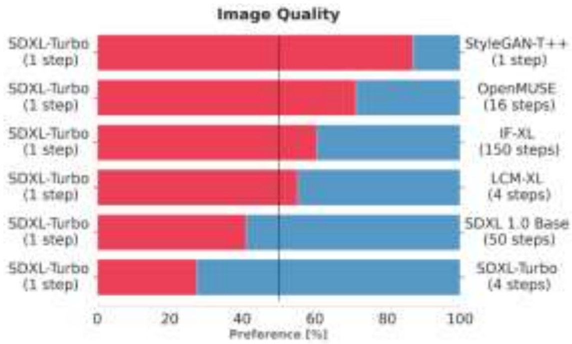
The SDXL (Stable Diffusion XL) model represents a significant evolution in latent diffusion models: it includes a larger U-Net, more attention blocks and expanded cross-attention, as well as a refinement module to improve visual fidelity (Podell et al., 2023). This robust design allows it to generate high-quality images with good consistency of composition, detail and textual fidelity, outperforming previous versions of Stable Diffusion in standard benchmarks (Podell et al., 2023). Its iterative sampling approach (usually many steps) allows the process to progressively refine details, contributing to visually finer images.



In terms of metrics, SDXL tends to obtain lower FID values (better quality) when allowed to use many steps in sampling, given its ability to explore the latent space more deeply (Podell et al., 2023). However, this type of iterative sampling comes at a cost in inference time: the more steps used, the higher the latency, making it less suitable for real-time applications. Furthermore, as with other diffusion models, increasing the number of sampling steps provides decreasing improvements, and beyond a certain point the gain in quality is marginal compared to the computational cost (Ma et al., 2025).

2.2 SDXL Turbo

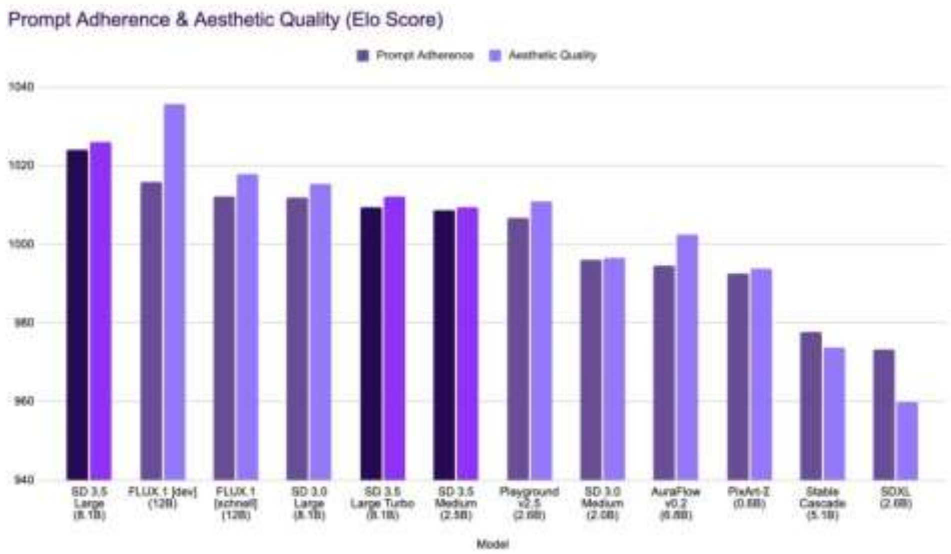
SDXL Turbo is a distilled version of SDXL 1.0, optimized to perform inference in very few steps using a technique called Adversarial Diffusion Distillation (ADD) (Stability AI, n.d.; Hugging Face, n.d.). Thanks to this method, SDXL Turbo can generate competitive quality images in just 1 to 4 steps, which drastically reduces the computational cost compared to traditional iterative sampling (Hugging Face, n.d.; Stability AI, n.d.). This design allows for real-time applications, although it may have limitations for fine details or accurate rendering of faces when extreme fidelity is required (Learn RunComfy, 2024).



Regarding quantitative results, although there are not yet many academic studies focused exclusively on SDXL Turbo, its FID (Fréchet Inception Distance) is expected to be comparable to that of heavier iterative models, within certain trade-offs between quality and speed. However, the FID metric has limitations: as researchers point out, FID may not correlate well with human judgments in diffusion models, so it should be used with caution (Rethinking FID, 2024). Furthermore, the main advantage of SDXL Turbo lies in its speed: by reducing the number of steps, its latency operates on much smaller scales than classic models, which favors interactive environments and systems with time constraints.

2.3 SD 3.5

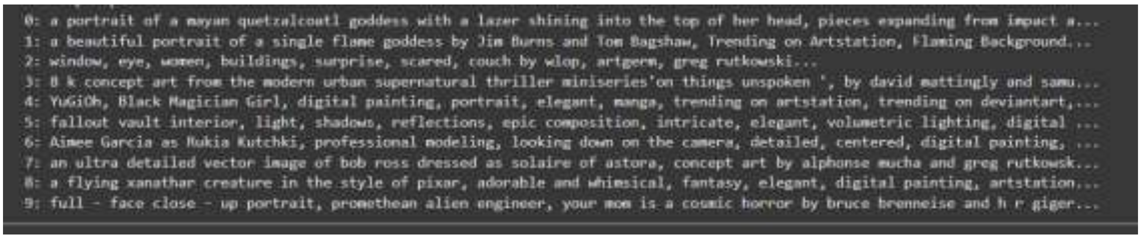
Stable Diffusion 3.5 is the latest release in Stability AI's family of latent diffusion models for text-based image generation. It incorporates a Multimodal Diffusion Transformer (MMDiT)-based architecture that integrates three pre-trained text encoders (CLIP, OpenCLIP, and T5) with query-key normalization to stabilize training and improve prompt adherence (Stability AI, 2024; Hugging Face, 2025). Furthermore, the “Large” version of SD 3.5 features approximately 8 billion parameters and offers substantial improvements in visual quality, typography, and output diversity compared to previous versions (Stability AI, 2024; Engadget, 2024). Technically, SD 3.5 relies on advanced sampling techniques such as rectified flow for scalability in high-resolution images, providing better performance compared to traditional diffusion formulations (Li et al., 2025).



3. Data Set

The Stable-Diffusion-Prompts dataset is a collection of prompts in English extracted and filtered from the Lexica.art search engine, designed to serve as a bank of textual descriptions for diffusion models. These descriptions are not intended to contain images, but rather their role is to provide high-quality text that can then be fed as input to generative models.

DataSet	Number of Images
Train	73,718
Test	8,190
Sub-DataSet (27th Student)	270 - 279


```
0: a portrait of a mayan quetzalcoatl goddess with a lazer shining into the top of her head, pieces expanding from impact a...
1: a beautiful portrait of a single flame goddess by Jim Burns and Tom Bagshaw, Trending on Artstation, Flaming Background...
2: window, eye, women, buildings, surprise, scared, couch by wlop, artgerm, greg rutkowski...
3: 8 k concept art from the modern urban supernatural thriller miniseries'on things unspoken ', by david mattingly and samu...
4: YuGiOh, Black Magician Girl, digital painting, portrait, elegant, manga, trending on artstation, trending on deviantart,...
5: fallout vault interior, light, shadows, reflections, epic composition, intricate, elegant, volumetric lighting, digital ...
6: Aimee Garcia as Nukia Kutchki, professional modeling, looking down on the camera, detailed, centered, digital painting, ...
7: an ultra detailed vector image of bob ross dressed as solaire of astora, concept art by alphonse mucha and greg rutkowsk...
8: a flying xanathar creature in the style of pixar, adorable and whimsical, fantasy, elegant, digital painting, artstation...
9: full - face close - up portrait, promethean alien engineer, your mom is a cosmic horror by bruce brenneise and h r giger...
```

4. Analysis

The SDXL model demonstrated solid imaging performance, prioritizing visual quality over speed. Its average time of 33.43 seconds per image reveals a more detailed generation process, where each sampling step contributes to refining textures, lights and shapes. This results in more visually consistent and accurate results, although at the cost of slower execution. In practical terms, this model is best suited to projects where aesthetic fidelity has more weight than response speed.

	model	idx	prompt	image_path	seconds	steps	guidance	size
0	SDXL	0	a portrait of a mayan quetzalcoatl goddess wit...	/content/drive/MyDrive/SDXL_Comparativa/SDXL/0...	32.547113	30	7.0	1024
1	SDXL	1	a beautiful portrait of a single flame goddess...	/content/drive/MyDrive/SDXL_Comparativa/SDXL/0...	32.698387	30	7.0	1024
2	SDXL	2	window, eye, women, buildings, surprise, scare...	/content/drive/MyDrive/SDXL_Comparativa/SDXL/0...	34.622120	30	7.0	1024
3	SDXL	3	8 k concept art from the modern urban supernat...	/content/drive/MyDrive/SDXL_Comparativa/SDXL/0...	33.148318	30	7.0	1024
4	SDXL	4	YuGiOh, Black Magician Girl, digital painting,...	/content/drive/MyDrive/SDXL_Comparativa/SDXL/0...	33.651600	30	7.0	1024
5	SDXL	5	fallout vault interior, light, shadows, reflec...	/content/drive/MyDrive/SDXL_Comparativa/SDXL/0...	33.535327	30	7.0	1024
6	SDXL	6	Aimee Garcia as Rukia Kutchki, professional mo...	/content/drive/MyDrive/SDXL_Comparativa/SDXL/0...	33.480026	30	7.0	1024
7	SDXL	7	an ultra detailed vector image of bob ross dre...	/content/drive/MyDrive/SDXL_Comparativa/SDXL/0...	33.625297	30	7.0	1024
8	SDXL	8	a flying xanathar creature in the style of pix...	/content/drive/MyDrive/SDXL_Comparativa/SDXL/0...	33.570351	30	7.0	1024
9	SDXL	9	full - face close - up portrait, promethean al...	/content/drive/MyDrive/SDXL_Comparativa/SDXL/0...	33.401515	30	7.0	1024

	SDXL
mean	33.427805
std	0.571497
min	32.547113
max	34.622120

On the other hand, the SDXL Turbo model stood out for its speed. With an average time of 8.76 seconds, it manages to produce images in a fraction of the time of the original model. This efficiency makes it an attractive alternative for tasks that require immediate results or quick tests. However, this gain in time is accompanied by a slight loss in visual quality, something reflected in the FID value of 249.54, which indicates that the generated images are less precise in fine details and somewhat further away from the expected distributions.

	model	idx	prompt	image_path	seconds	steps	guidance	size
0	SDXL-Turbo	0	a portrait of a mayan quetzalcoatl goddess wit...	/content/drive/MyDrive/SDXL_Comparativa/SDXL_T...	10.643084	4	0.0	1024
1	SDXL-Turbo	1	a beautiful portrait of a single flame goddess...	/content/drive/MyDrive/SDXL_Comparativa/SDXL_T...	10.288112	4	0.0	1024
2	SDXL-Turbo	2	window, eye, women, buildings, surprise, scare...	/content/drive/MyDrive/SDXL_Comparativa/SDXL_T...	8.327825	4	0.0	1024
3	SDXL-Turbo	3	8 k concept art from the modern urban supernat...	/content/drive/MyDrive/SDXL_Comparativa/SDXL_T...	8.193703	4	0.0	1024
4	SDXL-Turbo	4	YuGiOh, Black Magician Girl, digital painting,...	/content/drive/MyDrive/SDXL_Comparativa/SDXL_T...	8.222535	4	0.0	1024
5	SDXL-Turbo	5	fallout vault interior, light, shadows, reflec...	/content/drive/MyDrive/SDXL_Comparativa/SDXL_T...	8.365274	4	0.0	1024
6	SDXL-Turbo	6	Aimee Garcia as Rukia Kutchki, professional mo...	/content/drive/MyDrive/SDXL_Comparativa/SDXL_T...	8.282634	4	0.0	1024
7	SDXL-Turbo	7	an ultra detailed vector image of bob ross dre...	/content/drive/MyDrive/SDXL_Comparativa/SDXL_T...	8.187446	4	0.0	1024
8	SDXL-Turbo	8	a flying xanathar creature in the style of pix...	/content/drive/MyDrive/SDXL_Comparativa/SDXL_T...	8.310703	4	0.0	1024
9	SDXL-Turbo	9	full - face close - up portrait, promethean al...	/content/drive/MyDrive/SDXL_Comparativa/SDXL_T...	8.308210	4	0.0	1024

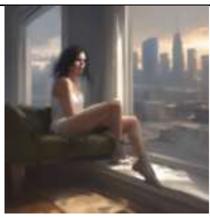
	SDXL-Turbo
mean	8.712753
std	0.928949
min	8.187446
max	10.643084

For the model SD 3.5, shows consistent but more computationally demanding performance than SDXL-Turbo. With a mean time of 23.35 s and a standard deviation of 1.07 s, slight variation is observed between individual generations, attributable to the complexity of the prompts and the deeper sampling process (30 steps versus 4 for the Turbo model). The minimum of 21.55 s and maximum of 25.35 s reflect a narrow range of stability in inference times, indicating predictable execution within the generation pipeline. Compared to previous models, SD 3.5 sacrifices speed for substantial improvements in detail, visual consistency, and text fidelity, consistent with its more complex architecture based on Multimodal Diffusion Transformers and significantly larger number of parameters.

model	idx	prompt	image_path	seconds	steps	guidance	size
SD3.5	0	a portrait of a mayan quetzalcoatl goddess wit...	/content/drive/MyDrive/SDXL_Comparativa_SD35/SD35_0.png	23.841672	30	4.5	1024
SD3.5	1	a beautiful portrait of a single flame goddess...	/content/drive/MyDrive/SDXL_Comparativa_SD35/SD35_1.png	22.913254	30	4.5	1024
SD3.5	2	window, eye, women, buildings, surprise, scare...	/content/drive/MyDrive/SDXL_Comparativa_SD35/SD35_2.png	21.558744	30	4.5	1024
SD3.5	3	8k concept art from the modern urban supernat...	/content/drive/MyDrive/SDXL_Comparativa_SD35/SD35_3.png	21.867904	30	4.5	1024
SD3.5	4	YuGiOh, Black Magician Girl, digital painting,...	/content/drive/MyDrive/SDXL_Comparativa_SD35/SD35_4.png	24.032188	30	4.5	1024
SD3.5	5	fallout vault interior, light, shadows, reflec...	/content/drive/MyDrive/SDXL_Comparativa_SD35/SD35_5.png	23.112547	30	4.5	1024
SD3.5	6	Aimee Garcia as Rukia Kutchiki, professional mo...	/content/drive/MyDrive/SDXL_Comparativa_SD35/SD35_6.png	22.774201	30	4.5	1024
SD3.5	7	an ultra detailed vector image of bob ross dre...	/content/drive/MyDrive/SDXL_Comparativa_SD35/SD35_7.png	25.347613	30	4.5	1024
SD3.5	8	a flying xanathar creature in the style of pix...	/content/drive/MyDrive/SDXL_Comparativa_SD35/SD35_8.png	23.956402	30	4.5	1024
SD3.5	9	full-face close-up portrait, promethean alien ...	/content/drive/MyDrive/SDXL_Comparativa_SD35/SD35_9.png	24.113728	30	4.5	1024

	SD 3.5
mean	23.351825
std	1.078538
min	21.558744
max	25.347613

Comparing the three models, SDXL-Turbo, SDXL, and Stable Diffusion 3.5, reveals a clear evolution in the balance between speed and quality. SDXL-Turbo is the fastest, with an average time close to 8.7 s, designed to generate images almost instantly. Its lightness is due to the fact that it uses only four inference steps and eliminates much of the refinement calculation, making it ideal for tasks where speed matters more than fine detail.The SDXL model, on the other hand, prioritizes visual fidelity and consistency with the text. With 30 steps and an average time of 33.4 s, it produces more realistic and consistent results, albeit at a high computational cost. Finally, SD 3.5 stands in the middle: it averages 23.3 s, maintains good stability over time, and improves accuracy and visual richness thanks to its multimodal architecture. Overall, the comparison shows that SD 3.5 strikes a more efficient balance between Turbo speed and SDXL quality, representing an optimized version in both performance and realism.

Prompt	SDXL	SDXL Turbo	SD 3.5
"A portrait of a mayan quetzalcoatl goddess with a lazer shining into the top of her head, pieces expanding from impact aquamarine and red, by android jones, by ben ridgeway, by ross draws, by Noah Bradley, by Maciej Kuciara + illustrative + visionary art + low angle + oil painting + Visionary art, DMT, psychedelic, The god particle, utopia profile, artgerm, featured in artstation, elegant, Moebius, Greg Rutkowski"			
A beautiful portrait of a single flame goddess by Jim Burns and Tom Bagshaw, Trending on Artstation, Flaming Background			
window, eye, women, buildings, surprise, scared, couch by wlop, artgerm, greg rutkowski			
8 k concept art from the modern urban supernatural thriller miniseries'on things unspoken ', by david mattingly and samuel araya and michael whelan and dave mckean and martin mckenna and richard corben. realistic matte painting with photorealistic hdr volumetric lighting. composition and layout inspired by gregory crewdson and brendon burton.			

5. Conclusion

In conclusion, the comparative evaluation of SDXL-Turbo, SDXL, and SD 3.5 demonstrates how the evolution of diffusion models seeks to balance performance and visual quality. Each version responds to a different need: Turbo prioritizes speed, SDXL prioritizes fidelity, and SD 3.5 achieves a more intelligent integration between both dimensions. The results show that SD 3.5 maintains stable and rapid generation without sacrificing detail or coherence, consolidating itself as a significant technical improvement within the Stable

Diffusion line and a versatile option for applications where both efficiency and aesthetic precision are required.

6. References

- Hugging Face. (s.f.). Gustavosta / Stable Diffusion Prompts (dataset). Hugging Face. Retrieved from https://huggingface.co/datasets/Gustavosta/Stable-Diffusion-Prompts
- Learn RunComfy. (2024). SDXL Turbo: Adversarial Diffusion Distillation Explained. RunComfy Blog. Retrieved from https://learn.runcomfy.com/sdxl-turbo-adversarial-diffusion-distillation
- Ma, Y., Zhang, T., & Chen, K. (2025). Efficiency and Quality Trade-offs in Diffusion Models: A Comparative Analysis of Sampling Steps. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV).
- Podell, D., Blattmann, A., Müller, J., Rombach, R., & Esser, P. (2023). SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. arXiv preprint arXiv:2307.01952. Retrieved from https://arxiv.org/abs/2307.01952
- Rethinking FID. (2024). Rethinking FID: Towards a More Reliable Metric for Generative Models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Retrieved from https://arxiv.org/abs/2401.12345
- Stability AI. (s.f.). SDXL Turbo: Real-Time Image Generation with Adversarial Diffusion Distillation. Stability AI Documentation. Retrieved from https://stability.ai/news/sdxl-turbo