**Comparative Analysis of Stable Diffusion Models**

## **Colab Link:** [Text To Image using diffusion.ipynb](https://colab.research.google.com/drive/1rr3gIZwuk-MlQWkvrlU6DL8Sdsg6GgNU?usp=sharing) **Drive Link for Results Obtained:** [results](https://drive.google.com/drive/folders/1GCMoEZzGoAddwRc77QBDgHjwTSHjGv0c?usp=sharing) **Introduction**

This report presents a comparative analysis of different Stable Diffusion models for generating photorealistic images from text prompts. The purpose of this study is to identify the most effective model and settings for achieving high-quality, photorealistic results. The models compared include Stable Diffusion 2.1, Stable Diffusion 2.0, and Stable Diffusion 1.5.

## **Methodology**

The comparison was conducted using a set of predefined text prompts representing various scenes, subjects, and styles. For each model, multiple combinations of inference steps (25, 50, 100, 150) and guidance scales (5.0, 7.5, 10.0, 12.0) were tested. The generated images were evaluated based on photorealism, match to the prompt, and visual quality. The generation time for each image was also recorded.

## **Code Explanation**

The provided code is a Python script that automates the process of generating images using different Stable Diffusion models and settings. Here's a brief overview of the code:

1. The necessary libraries are imported, including torch, diffusers, PIL, time, tabulate, os, and google.colab.
2. The Google Drive is mounted to allow saving the generated images and comparison metrics.
3. The models list defines the Stable Diffusion models to be compared, along with their respective schedulers.
4. The prompts list contains the text prompts used for generating images.
5. The num\_inference\_steps\_list and guidance\_scale\_list lists define the different settings to be experimented with.
6. The main loop iterates over each combination of model, prompt, and settings. For each combination, the model is loaded, and an image is generated using the specified prompt, number of inference steps, and guidance scale.
7. The generated images are saved on Google Drive in a structured manner, with subfolders for each model and prompt.
8. Comparison metrics, including the model name, prompt, inference steps, guidance scale, generation time, and image path, are stored in the comparison\_metrics list.
9. After generating images for all combinations of settings for a given prompt, a comparison image grid is created and saved.
10. Finally, the comparison metrics are displayed in a tabular format using the tabulate function.

## **Comparison Analysis**

Based on the generated results and the collected comparison metrics, the following observations can be made:

### **Image Quality**

* Stable Diffusion 2.1 and Stable Diffusion 2.0 consistently produced images with higher photorealism compared to Stable Diffusion 1.5. The images generated by these models exhibited more realistic textures, lighting, and overall visual coherence.
* Increasing the number of inference steps generally led to more detailed and refined images across all models. Images generated with 100 and 150 inference steps showed improved sharpness and finer details compared to those generated with 25 and 50 steps.
* Higher guidance scales resulted in images that more closely matched the text prompts. However, very high guidance scales (e.g., 12.0) sometimes introduced artifacts and reduced the overall image quality.

### **Generation Time**

* Stable Diffusion 1.5 had the fastest generation times among the compared models. It consistently generated images faster than Stable Diffusion 2.0 and Stable Diffusion 2.1.
* Increasing the number of inference steps significantly increased the generation time for all models. The relationship between the number of inference steps and generation time was nearly linear.
* The guidance scale had a minimal impact on generation time. Changing the guidance scale while keeping the number of inference steps constant did not result in significant variations in generation time.

### **Prompt Specificity**

* More specific and detailed text prompts generally led to higher-quality and more accurate images. Prompts that included information about the scene, subject, style, lighting, and camera settings tended to produce images that better matched the desired outcome.
* Prompts with vague or ambiguous descriptions often resulted in images that were less coherent or relevant to the intended subject matter.

## **Conclusion:**

Based on the comparative analysis, Stable Diffusion 2.1 and Stable Diffusion 2.0 outperformed Stable Diffusion 1.5 in terms of photorealism and overall image quality. These models consistently generated images with more realistic textures, lighting, and visual coherence.

In conclusion, Stable Diffusion 2.1 and Stable Diffusion 2.0 are recommended for generating photorealistic images from text prompts. A combination of 100-150 inference steps and a guidance scale of 7.5-10.0 seems to provide a good balance between image quality and generation time. Careful prompt engineering is essential to obtain the best results.