**The Stable Artist**

**Paper Summary :**

-Use text prompts to guide latent diffusion model space

-No training involved

**Links :**

<https://github.com/ml-research/semantic-image-editing/tree/StableArtist>

<https://github.com/huggingface/diffusers>

<https://huggingface.co/stable-diffusion-v1-5/stable-diffusion-v1-5>

**Inference Check :**

Errors

1. Upgrade pillow
2. Install diffusers using : !pip install --upgrade diffusers[torch]
3. Fix stable diffusers : In semdiffusers\pipeline\_latent\_edit\_diffusion.py

=> diffusers.pipelines.pipeline\_utils

**Result :**

13 mins runtime on 4GB RTX 3050 for base image generation



     editing\_prompt=[

                    'oil painting, drawing',

                    'modern bridge',

                    'ship on a river, ship']

30 mins runtime



**What we can do :**

1. Train audio encoder and text encoder to get the editing prompts from audio
2. Use audio and text data for training
3. Use TPOS audio and text encoder module

* We could also change the way the text encoding moves the stable diffusion process in latent space