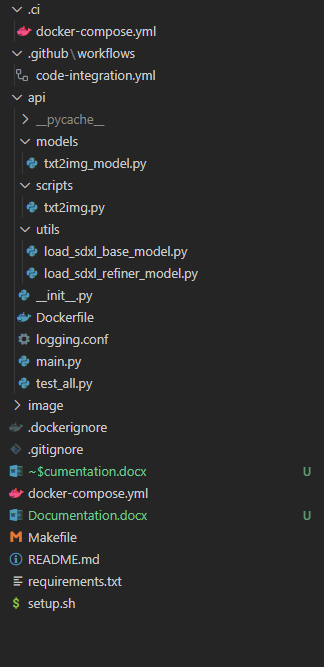
**Stable Diffusion XL API - OridosAI**

**Code Documentation**

## Project Structure



## Code description

### 2.1. .ci

2.1.1 docker-compose.yml: This file is used to define and run multi-container Docker applications. It contains the configuration for the services that make up your application, and allows you to define the relationships between them.

### 2.2. .github

#### 2.2.1 workflows:

This folder contains GitHub Actions workflows configuration files.

##### 2.2.1.1 code-integration.yml:

This file contains the configuration for a GitHub Actions workflow named "code-integration". It includes steps for continuous integration, such as running tests and checks on code changes.

### 2.3. api

#### 2.3.1 models:

This subfolder contains machine learning models used in the API.

##### 2.3.1.1 txt2img\_model.py:

This file contains the definition and implementation of a text-to-image conversion model.

##### 2.3.1.2 img2img\_model.py:

This file contains the definition and implementation of a image-to-image conversion model.

#### 2.3.2 scripts:

This subfolder contains scripts used in the API.

##### 2.3.2.1 txt2img.py:

This file contains the main functionality for converting text to an image using the model.

##### 2.3.2.2 img2img.py:

This file contains the main functionality for converting image to an image using the model.

#### 2.3.3 utils:

This subfolder contains utility functions used in the API.

##### 2.3.3.1 load\_sdxl\_base\_model.py:

This file contains code to load a sdxl base model for the API.

##### 2.3.3.2 load\_sdxl\_refiner\_model.py:

This file contains code to load a sdxl refiner model for the API.

##### 2.3.3.3 load\_scheduler.py:

This file contains code to load a diffuser scheduler for the API.

Implemented four schedulers.

dpmpp\_sde\_k, dpmpp\_2m\_k, unipc, ddim

If you want a newly, decent quality that is fast and converges, use dpmpp\_2m\_k or unipc.

In that case, steps 20 to 30 are recommended.

If you want good quality, use dpmpp\_sde\_k or ddim.

In that case, steps 10 to 15 are recommended.

#### 2.3.4 \_\_init\_\_.py:

This file indicates that the `api` directory should be treated as a Python package.

#### 2.3.5 Dockerfile:

This file contains instructions for building a Docker image for the API, specifying the environment and dependencies required to run the application.

#### 2.3.6 logging.conf:

This file contains configuration settings for logging in the API.

#### 2.3.7 main.py:

This file contains the main entry point for running the API.

#### 2.3.8 test\_all.py:

This file contains tests for the API.

### 2.4. image

This subfolder contains images used in the README.rd.

### ### Files:

#### - .dockerignore:

This file specifies patterns to exclude from the Docker build context, similar to .gitignore for Git.

#### - .gitignore:

This file specifies intentionally untracked files to ignore when using Git.

#### - docker-compose.yml:

This file is used to define the services, networks, and volumes for a Docker application. It specifies the infrastructure of the application.

#### - Makefile:

This file likely contains directives for building or compiling the project, enabling users to run build tasks with simple commands.

#### - Readme.md:

This file likely contains information about the project, including how to set it up, use it, and contribute to it.

#### - requirements.txt:

This file contains a list of Python dependencies required for running the project, often used with tools like pip for package management.

#### - setup.sh:

This file contains setup scripts for configuring and initializing the project.

## Parameters for using the txt2img function

### 3.2.1 Meaning of parameters

#### prompt

The prompt or prompts to guide the image generation. If not defined, one has to pass prompt\_embeds instead.

#### prompt\_2

The prompt or prompts to be sent to the tokenizer\_2 and text\_encoder\_2. If not defined, prompt is used in both text-encoders

#### Height

The height in pixels of the generated image. This is set to 1024 by default for the best results. Anything below 512 pixels won’t work well for stabilityai/stable-diffusion-xl-base-1.0 and checkpoints that are not specifically fine-tuned on low resolutions.

#### Width

The width in pixels of the generated image. This is set to 1024 by default for the best results. Anything below 512 pixels won’t work well for stabilityai/stable-diffusion-xl-base-1.0 and checkpoints that are not specifically fine-tuned on low resolutions.

#### num\_inference\_steps

The number of denoising steps. More denoising steps usually lead to a higher quality image at the expense of slower inference.

#### denoising\_end

When specified, determines the fraction (between 0.0 and 1.0) of the total denoising process to be completed before it is intentionally prematurely terminated. As a result, the returned sample will still retain a substantial amount of noise as determined by the discrete timesteps selected by the scheduler. The denoising\_end parameter should ideally be utilized when this pipeline forms a part of a “Mixture of Denoisers” multi-pipeline setup, as elaborated in Refining the Image Output

#### guidance\_scale

Guidance scale as defined in Classifier-Free Diffusion Guidance. guidance\_scale is defined as w of equation 2. of Imagen Paper. Guidance scale is enabled by setting guidance\_scale > 1. Higher guidance scale encourages to generate images that are closely linked to the text prompt, usually at the expense of lower image quality.

#### negative\_prompt

The prompt or prompts not to guide the image generation. If not defined, one has to pass negative\_prompt\_embeds instead. Ignored when not using guidance (i.e., ignored if guidance\_scale is less than 1).

#### negative\_prompt\_2

The prompt or prompts not to guide the image generation to be sent to tokenizer\_2 and text\_encoder\_2. If not defined, negative\_prompt is used in both text-encoders

#### num\_images\_per\_prompt

The number of images to generate per prompt.

#### eta

Corresponds to parameter eta (η) in the DDIM paper: https://arxiv.org/abs/2010.02502. Only applies to schedulers.DDIMScheduler, will be ignored for others.

#### Generator

One or a list of torch generator(s) to make generation deterministic.

#### Latents

Pre-generated noisy latents, sampled from a Gaussian distribution, to be used as inputs for image generation. Can be used to tweak the same generation with different prompts. If not provided, a latents tensor will ge generated by sampling using the supplied random generator.

#### prompt\_embeds

Pre-generated text embeddings. Can be used to easily tweak text inputs, e.g. prompt weighting. If not provided, text embeddings will be generated from prompt input argument.

#### negative\_prompt\_embeds

Pre-generated negative text embeddings. Can be used to easily tweak text inputs, e.g. prompt weighting. If not provided, negative\_prompt\_embeds will be generated from negative\_prompt input argument.

#### pooled\_prompt\_embeds

Pre-generated pooled text embeddings. Can be used to easily tweak text inputs, e.g. prompt weighting. If not provided, pooled text embeddings will be generated from prompt input argument.

#### negative\_pooled\_prompt\_embeds

Pre-generated negative pooled text embeddings. Can be used to easily tweak text inputs, e.g. prompt weighting. If not provided, pooled negative\_prompt\_embeds will be generated from negative\_prompt input argument.

#### output\_type

The output format of the generate image. Choose between PIL: PIL.Image.Image or np.array.

#### return\_dict

Whether or not to return a ~pipelines.stable\_diffusion\_xl.StableDiffusionXLPipelineOutput instead of a plain tuple.

#### cross\_attention\_kwargs

A kwargs dictionary that if specified is passed along to the AttentionProcessor as defined under self.processor in diffusers.models.attention\_processor.

#### guidance\_rescale

Guidance rescale factor proposed by Common Diffusion Noise Schedules and Sample Steps are Flawed guidance\_scale is defined as φ in equation 16. of Common Diffusion Noise Schedules and Sample Steps are Flawed. Guidance rescale factor should fix overexposure when using zero terminal SNR.

#### original\_size

If original\_size is not the same as target\_size the image will appear to be down- or upsampled. original\_size defaults to (height, width) if not specified. Part of SDXL’s micro-conditioning as explained in section 2.2 of https://huggingface.co/papers/2307.01952.

#### crops\_coords\_top\_left

crops\_coords\_top\_left can be used to generate an image that appears to be “cropped” from the position crops\_coords\_top\_left downwards. Favorable, well-centered images are usually achieved by setting crops\_coords\_top\_left to (0, 0). Part of SDXL’s micro-conditioning as explained in section 2.2 of https://huggingface.co/papers/2307.01952.

#### target\_size

For most cases, target\_size should be set to the desired height and width of the generated image. If not specified it will default to (height, width). Part of SDXL’s micro-conditioning as explained in section 2.2 of https://huggingface.co/papers/2307.01952.

#### negative\_original\_size

To negatively condition the generation process based on a specific image resolution. Part of SDXL’s micro-conditioning as explained in section 2.2 of https://huggingface.co/papers/2307.01952. For more information, refer to this issue thread: https://github.com/huggingface/diffusers/issues/4208.

#### negative\_crops\_coords\_top\_left

To negatively condition the generation process based on a specific crop coordinates. Part of SDXL’s micro-conditioning as explained in section 2.2 of https://huggingface.co/papers/2307.01952. For more information, refer to this issue thread: https://github.com/huggingface/diffusers/issues/4208.

#### negative\_target\_size

To negatively condition the generation process based on a target image resolution. It should be as same as the target\_size for most cases. Part of SDXL’s micro-conditioning as explained in section 2.2 of https://huggingface.co/papers/2307.01952. For more information, refer to this issue thread: https://github.com/huggingface/diffusers/issues/4208.

#### callback\_on\_step\_end

A function that calls at the end of each denoising steps during the inference. The function is called with the following arguments: callback\_on\_step\_end(self: DiffusionPipeline, step: int, timestep: int, callback\_kwargs: Dict). callback\_kwargs will include a list of all tensors as specified by callback\_on\_step\_end\_tensor\_inputs.

#### callback\_on\_step\_end\_tensor\_inputs

The list of tensor inputs for the callback\_on\_step\_end function. The tensors specified in the list will be passed as callback\_kwargs argument. You will only be able to include variables listed in the .\_callback\_tensor\_inputs attribute of your pipeine class.

### 3.2.2 List of parameters available for user adjustment

prompt

prompt2

height

width

num\_inference\_steps

denoising\_end

guidance\_scale

negative\_prompt

negative\_prompt\_2

num\_images\_per\_prompt

eta

output\_type

return\_dict

guidance\_rescale

original\_size

crops\_coords\_top\_left

target\_size

negative\_original\_size

negative\_crops\_coords\_top\_left

negative\_target\_size

### 3.2.3 List of parameters requiring adjustment by the programmer

generator

latents

prompt\_embeds

negative\_prompt\_embeds

pooled\_prompt\_embeds

negative\_pooled\_prompt\_embeds

cross\_attention\_kwargs

callback\_on\_step\_end

callback\_on\_step\_end\_tensor\_inputs

## Parameters for using the txt2img function

### 4.2.1 Meaning of parameters

#### prompt

The prompt or prompts to guide the image generation. If not defined, one has to pass prompt\_embeds instead.

#### prompt\_2

The prompt or prompts to be sent to the tokenizer\_2 and text\_encoder\_2. If not defined, prompt is used in both text-encoders

#### image

The image(s) to modify with the pipeline.

#### strength

Conceptually, indicates how much to transform the reference image. Must be between 0 and 1.

#### num\_inference\_steps

The number of denoising steps. More denoising steps usually lead to a higher quality image at the expense of slower inference.

#### denoising\_start

When specified, indicates the fraction (between 0.0 and 1.0) of the total denoising process to be bypassed before it is initiated.

#### denoising\_end

When specified, determines the fraction (between 0.0 and 1.0) of the total denoising process to be completed before it is intentionally prematurely terminated.

#### guidance\_scale

Guidance scale as defined in Classifier-Free Diffusion Guidance.

#### negative\_prompt

The prompt or prompts not to guide the image generation.

#### negative\_prompt\_2

The prompt or prompts not to guide the image generation to be sent to tokenizer\_2 and text\_encoder\_2.

#### num\_images\_per\_prompt

The number of images to generate per prompt.

#### eta

Corresponds to parameter eta (η) in the DDIM paper.

#### generator

One or a list of torch generator(s) to make generation deterministic.

#### latents

Pre-generated noisy latents, sampled from a Gaussian distribution, to be used as inputs for image generation.

#### prompt\_embeds

Pre-generated text embeddings. Can be used to easily tweak text inputs, e.g. prompt weighting.

#### negative\_prompt\_embeds

Pre-generated negative text embeddings.

#### pooled\_prompt\_embeds

Pre-generated pooled text embeddings. Can be used to easily tweak text inputs, e.g. prompt weighting.

#### negative\_pooled\_prompt\_embeds

Pre-generated negative pooled text embeddings.

#### output\_type

The output format of the generate image.

#### return\_dict

Whether or not to return a ~pipelines.stable\_diffusion.StableDiffusionXLPipelineOutput instead of a plain tuple.

#### cross\_attention\_kwargs

A kwargs dictionary that if specified is passed along to the AttentionProcessor as defined under self.processor in diffusers.models.attention\_processor.

#### guidance\_rescale

Guidance rescale factor proposed by Common Diffusion Noise Schedules and Sample Steps are Flawed guidance\_scale is defined as φ in equation 16.

#### original\_size

If original\_size is not the same as target\_size the image will appear to be down- or upsampled.

#### crops\_coords\_top\_left

crops\_coords\_top\_left can be used to generate an image that appears to be “cropped” from the position crops\_coords\_top\_left downwards.

#### target\_size

For most cases, target\_size should be set to the desired height and width of the generated image.

#### negative\_original\_size

To negatively condition the generation process based on a specific image resolution.

#### negative\_crops\_coords\_top\_left

To negatively condition the generation process based on a specific crop coordinates.

#### negative\_target\_size

To negatively condition the generation process based on a target image resolution.

#### aesthetic\_score

Used to simulate an aesthetic score of the generated image by influencing the positive text condition.

#### negative\_aesthetic\_score

Part of SDXL’s micro-conditioning as explained in section2.2of https://huggingface.co/papers/2307.01952.

#### clip\_skip

Number of layers to be skipped from CLIP while computing the prompt embeddings.

#### callback\_on\_step\_end

A function that calls at the end of each denoising steps during the inference.

#### callback\_on\_step\_end\_tensor\_inputs

The list of tensor inputs for the callback\_on\_step\_end function.

### 4.2.2 List of parameters available for user adjustment

prompt

prompt2

Image

strength

num\_inference\_steps

denoising\_start

denoising\_end

guidance\_scale

negative\_prompt

negative\_prompt\_2

num\_images\_per\_prompt

eta

output\_type

return\_dict

guidance\_rescale

original\_size

crops\_coords\_top\_left

target\_size

negative\_original\_size

negative\_crops\_coords\_top\_left

negative\_target\_size

aesthetic\_target\_size

negative\_aesthetic\_score

clip\_skip

### 4.2.3 List of parameters requiring adjustment by the programmer

generator

latents

prompt\_embeds

negative\_prompt\_embeds

pooled\_prompt\_embeds

negative\_pooled\_prompt\_embeds

cross\_attention\_kwargs

callback\_on\_step\_end

callback\_on\_step\_end\_tensor\_inputs