Image Generation and Analysis API Documentation

This API take the requests from the streamlit frontend in the form of a JSON format and then communicates those requests to the generation and analysis models. The responses are then sent back to the frontend to be processed as per needed.

Base URL of the Flask server:

http://localhost:5000/

Endpoints:

1. **/generate (POST)**
   Generates an image based on a text prompt using Stable Diffusion and performs a basic analysis using CLIP based on the keywords provided by the user at runtime.

Request Body:

Parameter	Type	Required	Description
prompt	string	yes	The text prompt to generate the image.
labels	string	no	Comma-separated list of labels for CLIP analysis.
samples	int	no	Number of inference steps (default: 30).
cfg	float	no	Guidance scale for image generation (default: 7.5).
height	int	no	Height of the generated image in pixels (default: 512).
width	int	no	Width of the generated image in pixels (default: 512).

Response Body:

Field	Type	Description
request_id	string	Unique identifier for the request.
generated_image	string	Base64-encoded image generated using Stable Diffusion.

clip_analysis	object	Contains ‘concepts’ (list of concepts) and ‘confidence_scores’ (dict).

Example Request:

```
{
  "prompt": "A futuristic cityscape at sunset",
  "labels": "city, sunset, futuristic",
  "samples": 50,
  "cfg": 8.0,
  "height": 512,
  "width": 768
}
```

Example Response:

```
{
  "request_id": "a1b2c3d4-5678-9101-1121-314151617181",
  "generated_image": "<base64-encoded-image>",
  "clip_analysis": {
    "concepts": ["city", "sunset", "futuristic"],
    "confidence_scores": {
      "city": 0.95,
      "sunset": 0.87,
      "futuristic": 0.76
    }
  }
}
```

2. **/analyze (POST)**
   Analyse an existing image using CLIP based on the objects given in the *common_objects.txt* file and SAM for basic segmentation.

Request Body:

Parameter	Type	Required	Description

image	string	yes	Base64-encoded image to be analysed.

Response Body:

Field	Type	Description
request_id	string	Unique identifier for the request.
generated_image	string	Base64-encoded image generated using Stable Diffusion.
clip_analysis	object	Contains 'concepts' (list of concepts) and 'confidence_scores' (dict).
basic_segmentation	object	Contains 'masks' (list) and 'polygons' (list) for segmented parts of the image.

Example Request:

```
{
  "image": "<base64-encoded-image>"
}
```

Example Response:

```
{
  "request_id": "b2a3d4c5-6789-0112-1324-253647580293",
  "generated_image": "<base64-encoded-image>",
  "clip_analysis": {
    "concepts": ["landscape", "river", "mountains"],
    "confidence_scores": {
      "landscape": 0.92,
      "river": 0.85,
      "mountains": 0.80
    }
  },
  "basic_segmentation": {
    "masks": [
      [[0, 0, 1, 1], [1, 0, 1, 1]],
      [[1, 0, 1, 1], [0, 1, 1, 1]]
```

```
   ],
   "polygons": [ [ 100, 150], [ 300, 400], [ 250, 350],... ]
  }
}
```

Error Handling:

HTTP Status Code	Error Message	Description
400	"'prompt' is required and must be a string."	Missing or invalid prompt in the /generate endpoint.
400	"'image' is required and must be a string."	Missing or invalid image in the /analyze endpoint.
400	"'field' must be a positive number."	Invalid samples, cfg, height, or width value.
500	{"error": "Description of the error."}	Generic server error message.

Model Configurations

Stable diffusion: stable-diffusion-v1-5

Clip analysis: ViT-L/14

SAM checkpoint: sam_vit_l_0b3195.pth

Features implemented

- Created a pipeline where an image is generated from a prompt, CLIP analysis is done on the image, and then the image is segmented to extract ROI polygons. **(required feature)**
- Advanced segmentation visualization by plotting all the generated masks with segmentation boundaries, create separate image for each extracted polygon, and provide polygon coordinates for each segment. **(advanced feature)**