Root Folder: Avatar Assignment

#Create conda environment:
conda create -n avatar python=3.10
conda install pytorch torchvision torchaudio pytorch-cuda=12.1 -c pytorch -c nvidia

#Install python dependencies using:

pip install -r requirements.txt

#Install thresstudio for stablezero123 GitHub - threestudio-project/threestudio: A unified framework for 3D content generation. GitHub - DSaurus/threestudio-mvimg-gen

cd threestudio

pip install -r requirements.txt

For tinycudann to install without issues:
Set export CUDA_HOME=/usr/lib/cuda

config.yaml:
Edit the following in config

```yaml
root_dir: "results/sofa_a50_p15_v2"  # Set the root directory for
results

segmentation:
  method: "sam"
  model: "facebook/sam-vit-base"

pose_editing:
  method: "zero123"

inpainting:
  method: "stable_diffusion"

input_image: "./sofa.jpg"
inpaint_output_image: "${root_dir}/inpaint_output_image.png"
mask_image: "${root_dir}/segmentation_mask.png"
segmented_image: "${root_dir}/segmented_rgba.png"
pose_rotated_mask: "${root_dir}/pose_rotated_object_mask.png"
pose_edited_image: "${root_dir}/pose_edited_image.png"
inpaint_mask: "${root_dir}/inpaint_mask.png"

# User input for LLM-based entity extraction
```

```
user_input: "Rotate the sofa chair by azimuth 50 degrees and polar 15
degrees."

# Optional quantization flag for LLM
quantize: False  # Set to True to enable 8-bit quantization

prompt: "Strictly restore the missing parts of white wall, floor,
borders of a single chair sofa facing sideways to complete the scene.
high resolution, smooth sharp and consistent"
negative_prompt: "artifacts,chair,sofa,furniture"
guidance_scale: 8.0
```

Run python main.py to get the results. `inpaint_output_image.png is the final output.`

Work Flow:

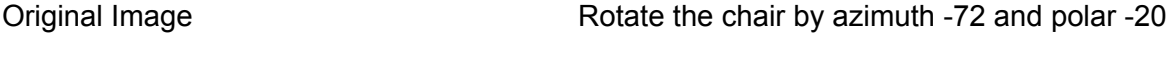
1. **Entity Extraction: Use Phi3.5 to extract the object name, azimuth angle and polar angle from the text**
2. **Segmentation: Use Grounding Dino to get the bounding box for the object using text prompt, further use this as grounding for segment anything model to get the mask.**
3. **StableZero123: use stable zero123 to get the 3d object from image and novel view synthesis.**
4. **Inpainting: Use stable diffusion 2 inpainting model with dilation, prompt, negative prompt and guidance scale. Tune these parameters for better results.**

Report:

For Chair.jpg:
Inpainting prompt used:

```
prompt="fill the mask as a continuation of wall and floor to complete
the scene. strictly no new entities or objects."
negative_prompt ="not wall, not floor,art,add objects,add partial
objects,new objects in scene,chair extended"
guidance_scale=8
```

Original Image



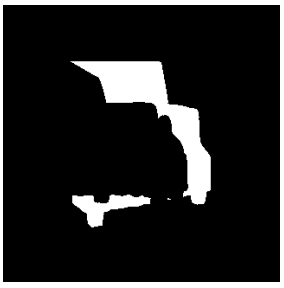
Rotate the chair by azimuth -72 and polar -20

For Sofa.jpg:

Issue: The inpainting with prompt and mask was not helping as the mask it was filling as extension of the object.

Solution: Dilated the mask by 5 x 5 convolutions so it extends to surrounding objects as well like wall and floor and use the prompt to inpaint these objects.

Original mask



Dilated mask

Inpainting prompt:

```
prompt ="Strictly restore the missing parts of white wall, floor,
borders of a single chair sofa facing sideways to complete the scene.
high resolution, smooth sharp and consistent"
negative_prompt = "artifacts,chair,sofa,furniture"
guidance_scale = 8
```

Original Image

Rotate the sofa chair by azimuth 50 and polar 15

For table.jpg:
Inpainting Prompt:

```
# prompt ="Strictly restore the missing parts of white sofa in the
background, floor and table to complete the scene."
# negative_prompt = "artifacts, table,furniture"
# guidance_scale = 6
```

Original Image

Rotate the table by azimuth -63 polar 25

Observations:

Stable zero123 is only generating 256 x 256 images(Interpolated later to match image resolution) and using the non commercial checkpoint the generated meshes or views aren't very good.

For chairs with holes the stable zero123 generation is bad, inpainting the mask was giving bad results even with dilation.

Office Chair

Rotate by polar azimuth 50 polar 15
Stable zero123 output over layed on original
Scene without chair.

Need to try better image to 3d models and novel view synthesis models. That give high resolution outputs.