**Kitchen Cabinet Icon generation**

**Model Summary from Midterm Report:**

The first model result provided in the midterm report leverage prompt file generated by BLIP model as well as Home Depot icons. However, the results are not as good and require user to manually update the prompt.

The second model result provided in the midterm report was generated by manually created prompt file that follows a specific format and utilize training data that includes Home Depot icons, free icons found on the internet, and realistic product photo. The result from this model is good however the style resembles realistic product photo. Even with only including a few realistic product photos as training data, the generated icons will fully resembles / heavy skews towards realistic product photo. Thus, in the final models, I’ve removed all the realistic product photo from training data.

**Model Summary for Final Report:**

For the models used in final report, I leverage Home Depot icons and free icons found online. The final dataset contains around 42 icons.

1. Long Prompt with Class Images: The class images are 100 high-quality, style consistent Home Depot icon images, not restricted to kitchen cabinets. The training dataset for this model is the same as for the short prompt and long prompt model.
   * Since we won’t be discussing about class in our final report, I won’t generate CLIP and FID score for this model. Just including this model here in this summary for reference purposes.
2. Short Prompt: The training prompt only uses keywords, such as "thumb head screw icon" and "slotted drive screw icon."
3. Long Prompt: The training prompt uses longer descriptions of the training dataset images that follows a specific format, "a photo of TOK kitchen cabinet icon, <Style: 2D>, a <cabinet type> kitchen cabinet in <color family> with <description of the cabinet type or corner cabinet type>”. The two options for style include <Style: 2D> and <Style: 3D>. Example of a text prompt, "a photo of TOK kitchen cabinet icon, <style: 2D icon>, a glass frame material".
4. The 4th model is DALL.E 3, which we can use via the DALL.E 3 API to generate images using the same inference prompt (for both short and long prompt).

**Training Parameters:**

The parameters used for the SDXL models are as follows:

#!/usr/bin/env bash2

!accelerate launch train\_dreambooth\_lora\_sdxl.py \

--pretrained\_model\_name\_or\_path="stabilityai/stable-diffusion-xl-base-1.0" \

--pretrained\_vae\_model\_name\_or\_path="madebyollin/sdxl-vae-fp16-fix" \

--dataset\_name="/kaggle/input/kc-icons-only" \

--output\_dir="/kaggle/working/output" \

--caption\_column="prompt"\

--mixed\_precision="fp16" \

--instance\_prompt="a photo of TOK kitchen cabinet icon" \

--resolution=512 \

--train\_batch\_size=1 \

--gradient\_accumulation\_steps=3 \

--gradient\_checkpointing \

--learning\_rate=1e-4 \

--snr\_gamma=5.0 \

--lr\_scheduler="constant" \

--lr\_warmup\_steps=0 \

--mixed\_precision="fp16" \

--use\_8bit\_adam \

--max\_train\_steps=600 \

--checkpointing\_steps=150 \

--seed="0"

**Inference**:

For each model, I generated 40 inference images using 10 different prompts, each with num\_inference\_step = 100 and num\_images\_per\_prompt = 4. I saved the associated prompt used for generation and select the best image(s) per prompt and calculated the average CLIP score for each model. Additionally, I selected 3 Home Depot icons and their corresponding inference images, revised these files to have paired names, and calculated the average FID score. Since there’s not a lot of kitchen Cabinet icons from Home Depot, I can only inference on 5. (for GT and other source, I can inference on more icons as majority of my training icons are from free source)

**FID Results:**

* Short Prompt FID Score: 325.0759482684852
* Long Prompt FID Score: 325.6341972131611
* DALL.E FID (short prompt) Score: 412.7144867477459
* DALL.E FID (long prompt) Score: 329.18331850433776

**CLIP Results:**

* Short Prompt FID Score: 0.3258218228816986
* Long Prompt FID Score: 0.2941736325621605
* DALL.E FID (short prompt) Score: 0.32772575318813324
* DALL.E FID (long prompt) Score: 0.3053057983517647

For GT & open-source paper, I’ve removed all Home Depot icons, add additional free icons from the internet to make the training dataset the same size of model with HD icons (i.e. 42 icons), and train using the same model parameters and respective text prompt, below are the inference results:

**FID Results:**

* Short Prompt FID Score: 307.4057050557609
* Long Prompt FID Score: 299.93992850632713

**CLIP Results:**

* Short Prompt FID Score: 0.32190975844860076
* Long Prompt FID Score: 0.3134744793176651

**Other Findings & Learnings:**

* Importance of training data: how it helps model learn style and will skew towards realistic product photo
* CLIP score not always the best method because short prompt has better CLIP score but less possibility to customize the icon style