Digital Egypt Pioneers

STYLE HOME GEN AI

Team members

Hager Hamed

Naghem Naser

Dina Mohammed

Omnya Mohammed

Under Supervision:

Eng/Mohammed Agoor

OUTLINE

1:Introduction

2:Motivation

3:Problem and solution statements

4:Related work

5:Data Set

6:Model Structure

7:Implementation

8:Technologies

INTRODUCTION

What is Generative AI?

Generative AI helps create new content, like pictures or text. One exciting use is turning written descriptions into images.

What is SDXL?

Stable Diffusion XL (SDXL) is a smart model that makes high-quality and creative images just from text.

What is this work about?

This work shows how we can use these models in art, design, and
It explains how models like Home Style GAN AI turn text into custom images, helping in art, home design, and more.

MOTIVATIONS

Motivations

The idea of this project comes from the growing need to use AI in creative areas like interior design. Traditional design can be time-consuming and requires professional skills. We wanted to make design easier and more accessible for everyone.

Importance & Impact

This project supports designers, homeowners, and anyone interested in home decor. It saves time, enhances creativity, and allows people to explore styles in a fun and smart way. Our model makes design more personal and interactive, opening new doors for using AI in everyday life. It also shows how deep learning can bring real change to industries like architecture and home styling.

Generative AI, especially SD, gave us a way to turn simple text into visual designs. This inspired us to build a system where users can describe a room and get a realistic image of it—without needing design experience.

PROBLEM AND SOLUTION

1.GANs lacked text-image accuracy.

➤ Switched to diffusion models for better text alignment.

2.Diffusion models caused memory issues on Colab.

➤ Reduced resolution and dataset size to ease memory load.

3.LoRA training failed due to SDXL model size.

➤ Found optimized LoRA guide tailored for Colab.

4.Initial SDXL outputs were average.

➤ Improved quality by increasing resolution gradually.

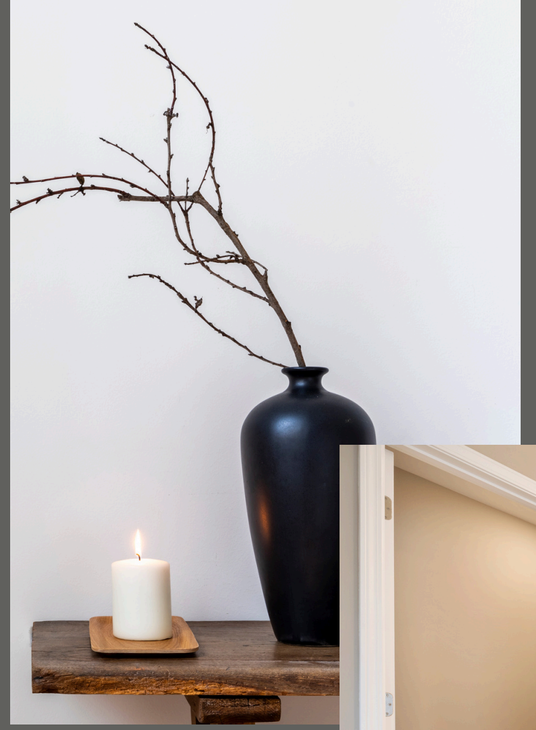
RELATED WORK

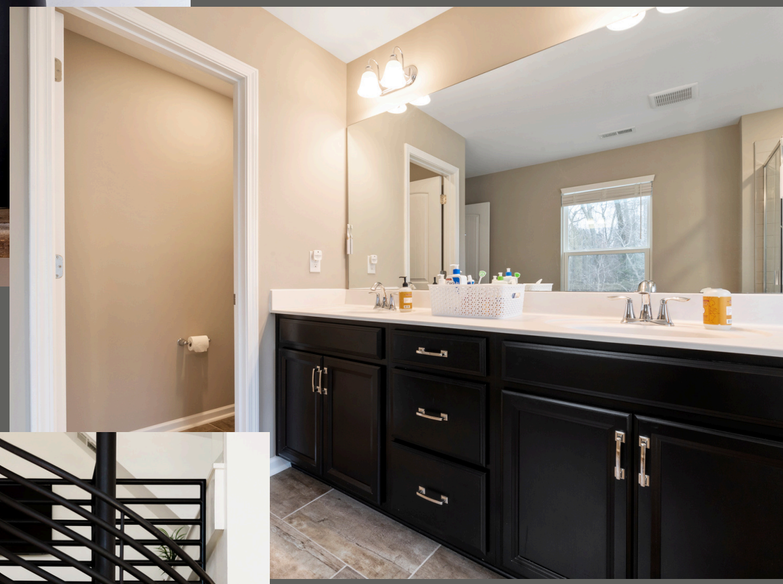
📌1. Foundational Research Papers

- DreamBooth
- Stable Diffusion

🧪 2. Related & Complementary Research

- Textual Inversion
- LoRA(LowRankAdaptation)
- Custom Diffusion





DATA SET

Dataset Overview

- Source: Images from Houzz.com
- Total Images: 18,282
- Categories: 19 interior styles (e.g., Modern, Victorian, Rustic...)
- Format: JPEG, High-Res

Data Annotation

- Used Blip2Processor for auto-captioning
- Output: (Image, Caption) pairs
- Purpose: Supervised fine-tuning for vision–text learning

Steps OF Data

1-Preprocessing

- Download and organize interior design dataset by style folders.
- Perform EDA: count images, show samples, analyze image dimensions.
- Collect image paths and labels for processing.

2-Image Captioning with BLIP2

- Used Salesforce’s BLIP2 model (2.7B version) for image captioning.
- Generates natural English captions from images.
- Why BLIP2:

* State-of-the-art zero-shot captioning.
* Higher accuracy than older models (Show & Tell, CLIPCap).

*Easy integration with Hugging Face Transformers.

3- Saving the Results

- Format: CSV file with Image ID, File path, Style, English caption, Arabic translation.
- Benefits:
- Easy data management.
- Accessible for further analysis or training.
- Simplifies multilingual captio

4- Uploading to Hugging Face

- Logged in via huggingface_hub.login()
- Uploaded dataset or models for sharing or deployment.

5- Final Outcome

- Bilingual captions for each image.
- High-quality dataset for various applications.
- Enriches Arabic content in visual domains.

Challenges and Insights

Experience with M2M100 Translation:

We initially aimed to use M2M100 by Meta AI for translating image captions into Arabic. While M2M100 is a powerful model supporting over 100 languages and doesn't require English as a pivot, it did not perform as expected due to hardware limitations.

Challenges:

- Memory and Resources: M2M100 requires high RAM, GPU, and memory resources, which led to performance issues.
- Scalability: The model struggled to handle large-scale image captioning tasks within the available hardware constraints.

CORE TECHNOLOGIES

SDXL- The Brain Behind Image Generation

Developed by Stability AI as a high-quality Text-to-Image model

Uses a larger UNet and two Text Encoders for better image fidelity

Requires powerful GPUs for training, but we used optimization techniques to fine-tune on .T4 GP

Serves as the base model for generating images from text prompt

Fine-tuned later using DreamBooth and LoRA to personalize outputs

LoRA

Implements Low-Rank Adaptation for fine-tuning large models .efficiently

Instead of updating all weights, it trains lightweight adapter layers only

Significantly reduces memory and VRAM usage, making training feasible on .GPUs like T4

Allows quick and effective customization without full model retraining

DreamBooth

Fine-tuned SDXL using 19 curated images from different parts .of a house (bathroom, kitchen, bedroom, garden

Introduced a custom Trigger Word (e.g., “TOK home”) to teach the mode .about this specific interior style

Enabled the model to generate images that reflect the visual style .and layout of the provided house designs

While the model learns these specific rooms, it may still struggle to generalize to completely new layouts not represented in the training .data

HOW DO THE THREE WORK TOGETHER?

How do the three work together?

I used a Stable Diffusion XL as a base

I added a DreamBooth to customize it with
a new modern style

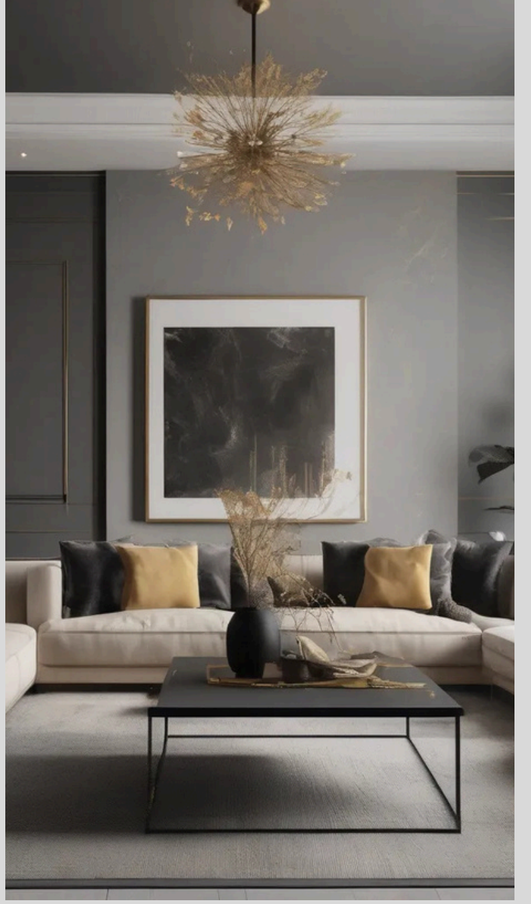
I used LoRA for training so I could train quickly and efficiently
.with limited resources

IMPLEMETATION

Implementation Steps

1. Environment Setup (GPU, libraries, login)
2. Load Pretrained Model
3. Prepare & Preprocess Dataset
4. Configure LoRA
5. Training Loop
6. Inference from Prompt
7. Save Results

Hyperparameters



pretrained_vae_model_name_or_path

Specifies the pre-trained VAE used to encode and decode images , optimized for FP16 precision.

resolution

Defines the training image size; higher resolution may improve detail but requires more memory.

train_batch_size

Number of images processed per training step; smaller values reduce VRAM usage.

Hyperparameters

gradient_checkpointing

- Saves memory during training by recomputing intermediate activations
- on the backward pass.

max_train_steps

- Sets the total number of training steps;
- more steps usually lead to better learning.

checkpointing_steps

- Saves model progress every N steps to allow
- resuming or evaluating training.

compare the results

Res 32x32 train 100

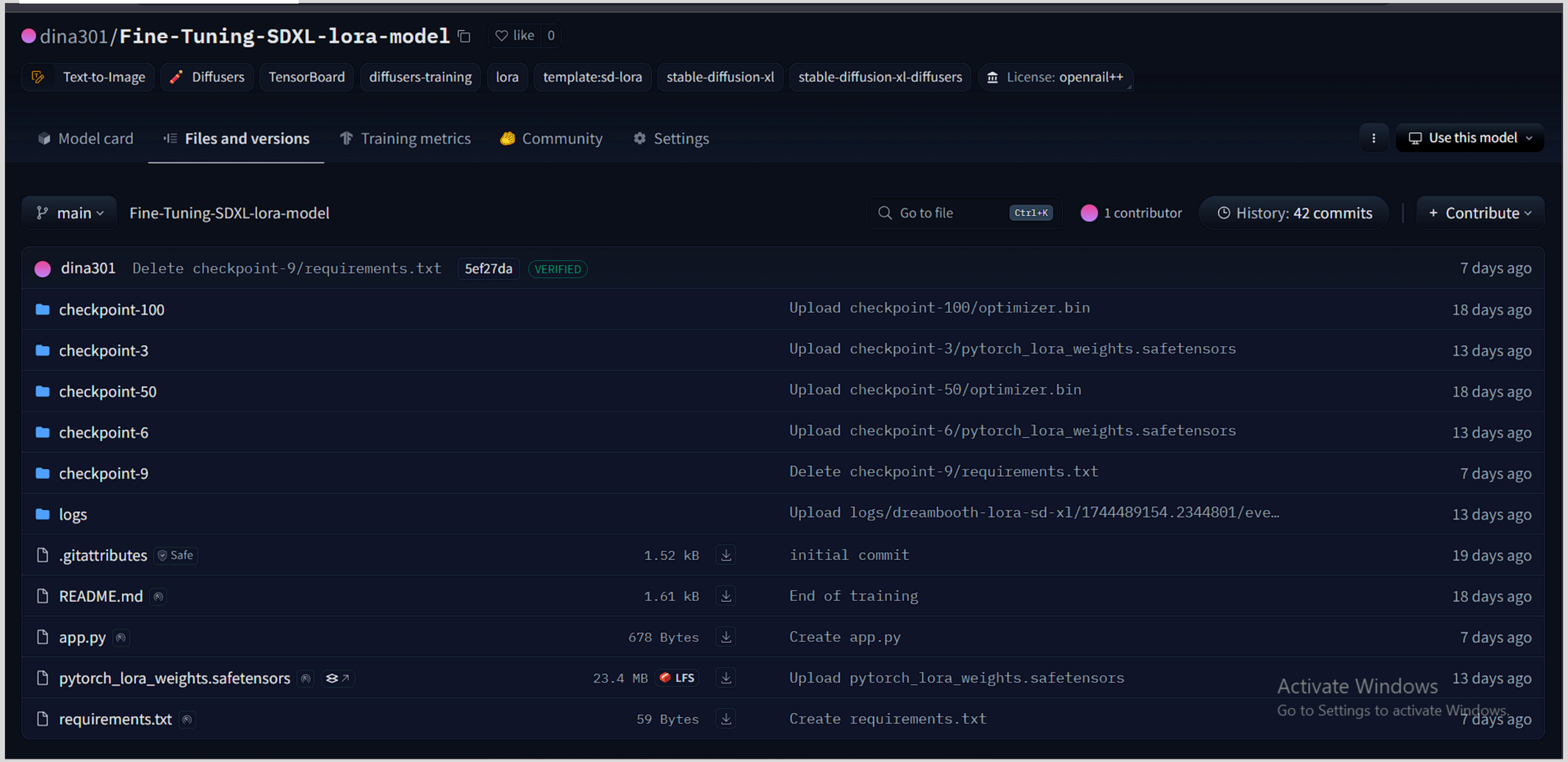
compare the results

Res 256x256 train 25

compare the results

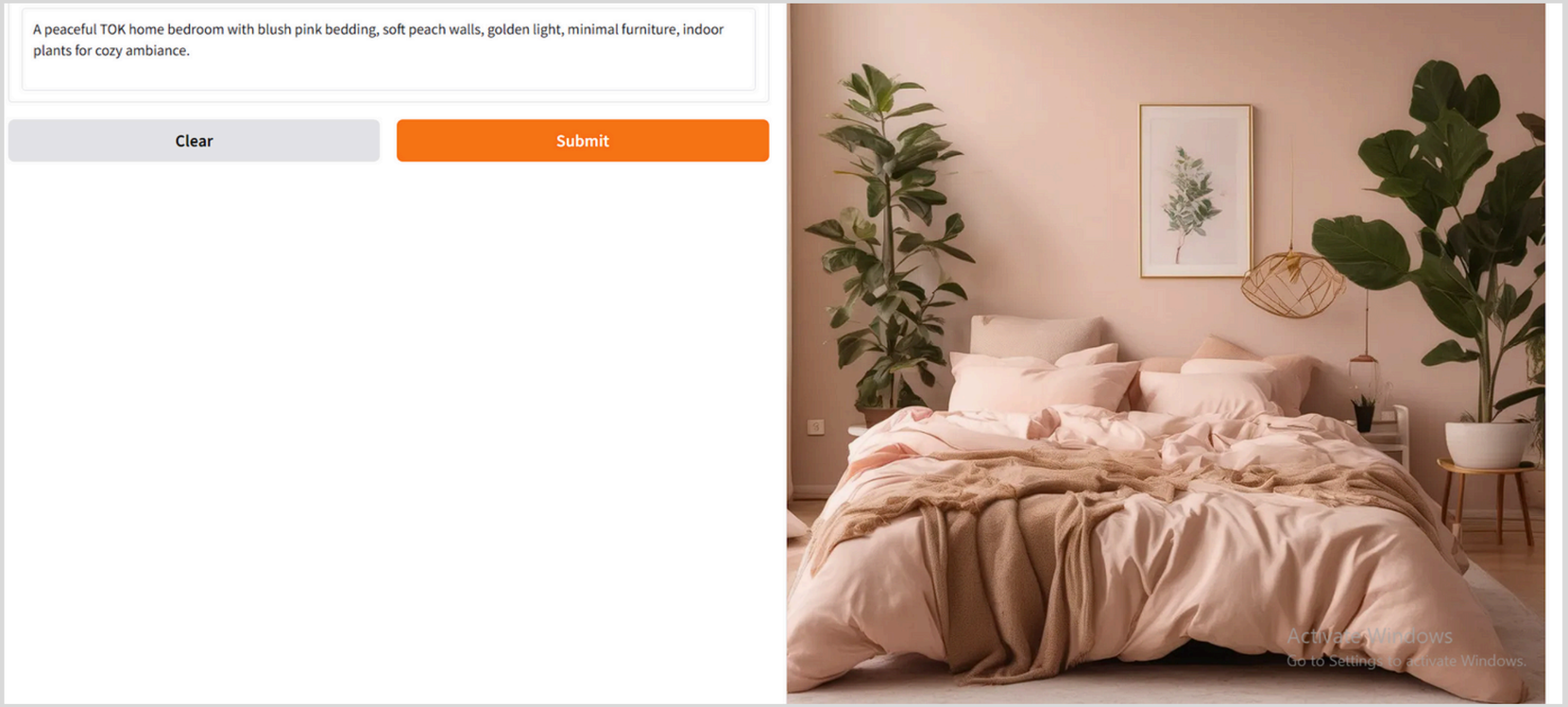
Res 512x 512 train 9

SAVING THE WEIGHTS IN HUGGING FACE



INFERENCE USING COLAB

- Uses Gradio to create a simple web interface.
- Loads a fine-tuned Stable Diffusion XL model with LoRA weights
- Utilizes FP16 precision for faster and memory-efficient inference on GPU.
- Interface shared via share=True for online access and testing.



TECHNOLOGIES


Stable Diffusion


Google Colaboratory

Conclusion

- ***Successfully trained SDXL on a custom dataset.***
- ***Generated stylistically consistent interiors.***
- ***Learned about model optimization, data pipelines, and real-world deploymen.***

ENJOY WATCHING

e2d3205fa58fef6f1a.gradio.live

Prompt

A charming TOK home garden with blooming pink flowers, soft green grass, a peach-colored seating area, warm sunset lighting.

Clear

Submit

output

processing | 13.4/46.5s

86°F
Mostly clear

Search

ENG