Generative AI (Spring-2025)

Assignment-2

Instructor

Dr. Hajra Waheed PHD

Submission Guidelines:

- Submit your assignment on Google Classroom in the format "20XX.ipynb".
- The deadline is Mar 10, 2025, at 11:59 PM. No extensions will be granted.

Declarations:

- Late submissions will incur penalties: 25% deduction on the first day, 50% on the second day, and zero marks thereafter.
- Plagiarism will result in zero marks for the assignment.
- This is an individual assignment; collaboration or group work is strictly prohibited.
- Please ensure that you submit your own original work.

VIVA Policy:

- A VIVA (oral examination) will be conducted to assess your understanding of the assignment.
- The VIVA will be scheduled separately, and you will be notified of the date and time.
- Failure to attend the VIVA will result in zero marks for the assignment.

Academic Integrity:

- Plagiarism, collusion, and academic dishonesty will not be tolerated.
- Any instances of academic misconduct will be reported to the authorities and may result in severe penalties.

Objective

This assignment focuses on leveraging state-of-the-art **Vision Transformer (ViT)-based models, CLIP, and Stable Diffusion** for various computer vision tasks. The goal is to explore their effectiveness in **spoof detection, image retrieval, style transfer, and product mockup generation** by implementing and evaluating their capabilities.

1. **Vision Transformers (ViT)**
2. **CLIP**
3. **Stable Diffusion**

Datasets

We will use the following dataset:

- Download the COCO dataset's **validation images**
- nguyenkhoa/celeba-spoof-for-face-antispoofing-test (HugginFace)

Part 1 (Spoof Detection) 25%

Financial institutions and government agencies use AI-powered face authentication systems for secure access. Attackers often try to bypass these systems using spoofed images or videos (e.g., printed photos or deep fake videos). A ViT-based spoof detection model helps identify and block such fraudulent attempts

Tasks:

- **Dataset Preparation:** Load a dataset containing real and spoofed images.
    - nguyenkhoa/celeba-spoof-for-face-antispoofing-test
- **Model Training:** Fine-tune a **ViT model** for binary classification (Real vs. Spoof).
- **Evaluation:** Compute **Accuracy, Precision, Recall, and F1-score** to assess performance.
- Add **your Real Photo** to test results
- Add **your spoofed photo** (From Mobile screen – Replay attack)

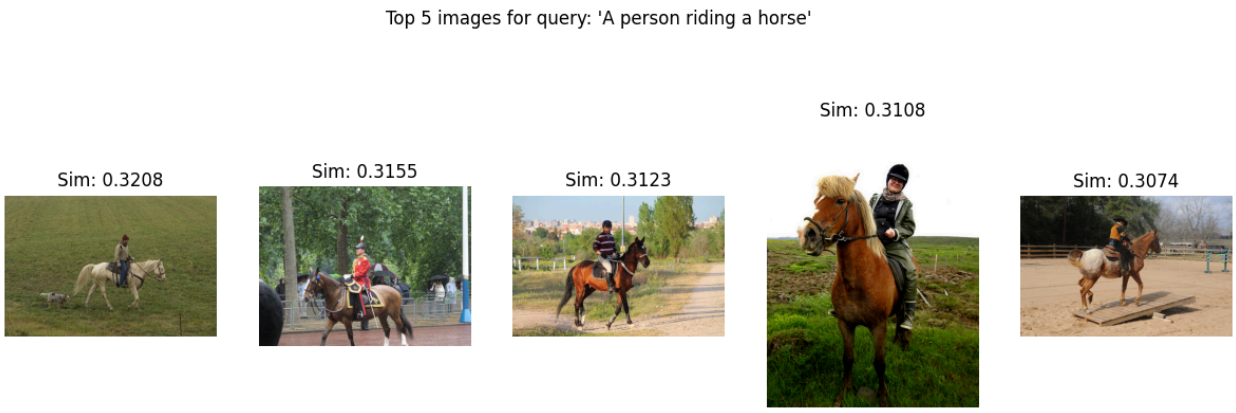
Part 2: AI-Powered Visual Search (35%)

The goal of this task is to work with OpenAI's **CLIP (Contrastive Language–Image Pretraining) model** using a Vision Transformer (ViT) backbone. You will use CLIP to encode images and text, perform similarity calculations, and implement a basic image retrieval system.

Tasks:

- Download the COCO dataset's **validation images** (`val2017.zip`) and **annotations**(`annotations_trainval2017.zip`).
- Load the `CLIP-ViT` model (`openai/clip-vit-base-patch32`)
- Retrieve the **top 5 most similar images** for a given text query.
- Display the retrieved images along with their similarity scores and Query.

Example output



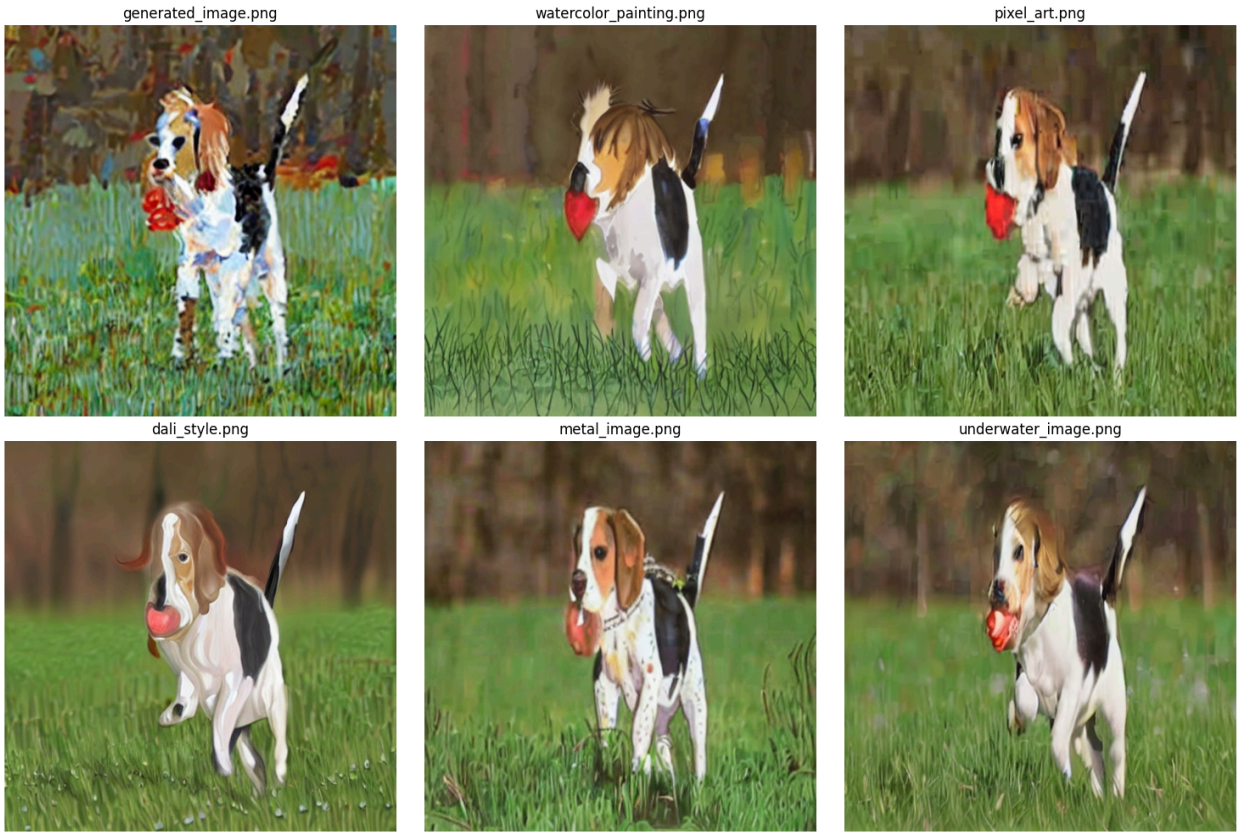
.

.

Part 3: Stable Diffusion (20%)

Tasks:

- **Load** a pre-trained Stable Diffusion model.
- **Implement** a Python script that takes an input image and generates variations based on text prompts.
- **Experiment** with strength, guidance_scale, and num_inference_steps for a single prompt.
    - Generate at least three examples with different parameter values.
    - Document how each change affects the output in terms of detail, realism, and coherence.
- **Prompt Engineering & Analysis**
    - Create at least five different prompts, such as:
        - "`watercolor painting`."
        - "`a pixel art`
        - “`in the style of Salvador Dalí`"
        - "`in the style of Van goh`"
    - Generate images for each prompt using a consistent parameter set.
    - Compare how well Stable Diffusion interprets different prompts.



Part 4: Mockup Generation (20%)

Stable Diffusion

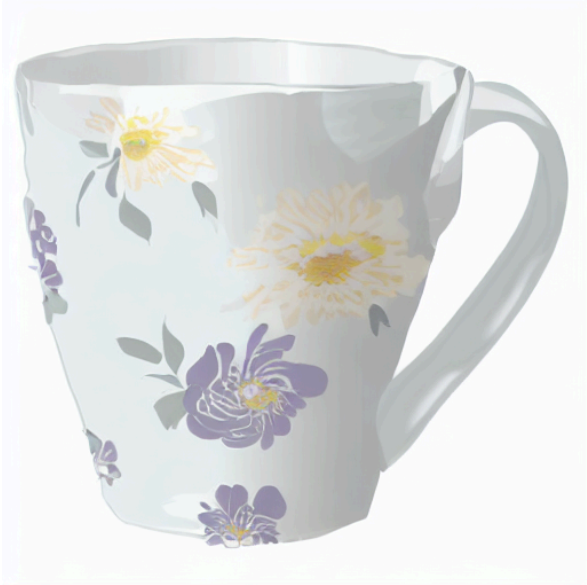
Imagine you are working for an e-commerce platform that allows users to design customized products like mugs, T-shirts, or phone cases. Using Stable Diffusion, you can generate realistic product mockups by overlaying designs onto base images, providing customers with previews before purchasing.

Tasks:

- **Basic Mockup Generation**
  - Implement a Python script to generate a product mockup from a plain base image.
  - Use a simple object (e.g., a plain white coffee mug) as the base image.
  - Generate a mockup using the prompt: "vintage floral pattern."
- **Parameter Tuning for Style Customization**

    - Experiment with different strength and guidance_scale values while keeping the prompt "minimalist geometric design."
    - Provide at least three examples showing how parameter adjustments affect the style's intensity and detail.
- **Style Diversity**
    - Create a list of five diverse styles, such as:
        - "Floral Art."

Example Output



Deliverables:

- **Notebook** with code implementations.
- **Detailed Report with screenshot attached for each Part (Properly formatted)** .
    - Discuss Parameters and their Effects on your results in Stable Diffusion