Image Search Engine

CS6384 - Computer Vision

Vatsal Kishorbhai Mavani
Lavanya Gopal
Bhargav Narasimha Reddy Pulla

INTRODUCTION

- In today's world, image search engines have become an essential part of our online experience, providing us with the ability to quickly and effortlessly search for images.

- Traditional tools allow us to find images based on keywords, tags, etc. But first these images should be tagged manually.
- To reduce the workload and increase the accuracy of searching results we have employed the Artificial Intelligence.

- Also, when it comes to searching for images stored locally on a device, the process can become challenging and time-consuming.

- The Image Search Engine project addresses this issue by providing an intuitive and efficient solution for finding images stored on a local device.

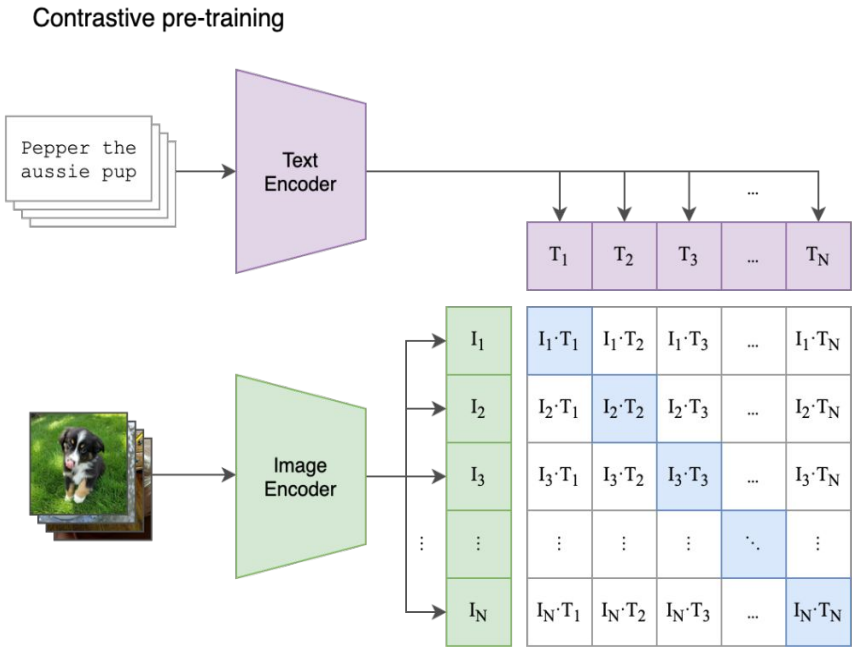
- With the Neural Image Search Engine, users can enter text queries in the form of phrases, and the engine will quickly return images related to the entered phrase, providing a seamless and convenient search experience.

Technologies used

- Python
- CLIP - *Contrastive Language–Image Pre-training*
- PyTorch
- FastAPI
- Next.JS

CLIP: Language-Image Model

- CLIP(*Contrastive Language–Image Pre-training*) is a neural network trained on about 400 million text and image pairs.
- Training uses a contrastive learning approach that aims to unify text and images, allowing tasks like image classification, image captioning to be done without any training.
- CLIP uses Visual Image Transformers(ViT-B/32) for image processing and a masked self-attention transformer for computing text embeddings.
- These encoders are trained to maximize the similarity of image-text pairs via a contrastive loss and Recall Metric has been used for performance evaluation.

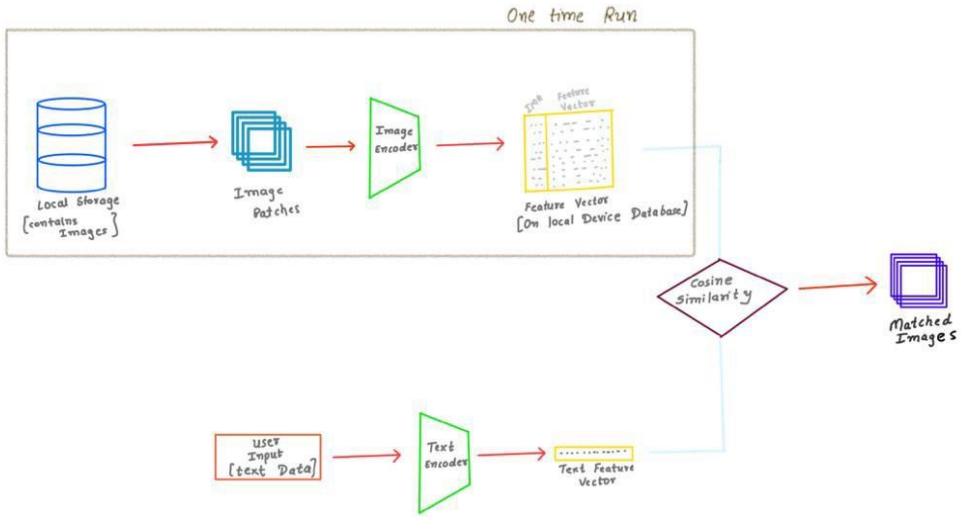


APPROACH

- First approach was with different Convolution based Image models for image processing and Language models for text processing.
- But the feature embedding distribution were not similar, thus giving very poor results.
- Next, we employed CLIP as it was particularly trained for Image-Text pairing thus fitting well for our task.
- To make the image-searching program accessible to non-technical users, the entire system features a graphical user interface.

Image Search Engine Workflow

- The first phase of the solution is to retrieve all images available on the computer or given path.
- The next module involves pre-processing all images to extract useful information using the Visual Transformer Neural Network from CLIP.
- Computed feature vectors are then stored to minimize redundant computations during subsequent searches.
- The provided text data is first processed by the Large Language Model with Transformer Architecture, which produces a feature vector.
- The comparison module then finds the similarity between the text feature vector and the image feature vectors



Un Splash

Unsplashed

The feeling when your program finally works

Search

Local Storage

Unsplashed

/content/imagenette2

dogs in the water

Search

Thank you!