



Mana Saleh Al Aqeel

Image Search and Caption Generation System



SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority

Introduction

The project aims to develop an image search and caption generation system using pre-trained models. It leverages the CLIP model for image search and the BLIP model for image captioning.



Approach



The project is divided into two main components:

Image Search: Uses the CLIP model to find the most relevant image based on a given text query.

Image Captioning: Uses the BLIP model to generate captions for images.

Image Search

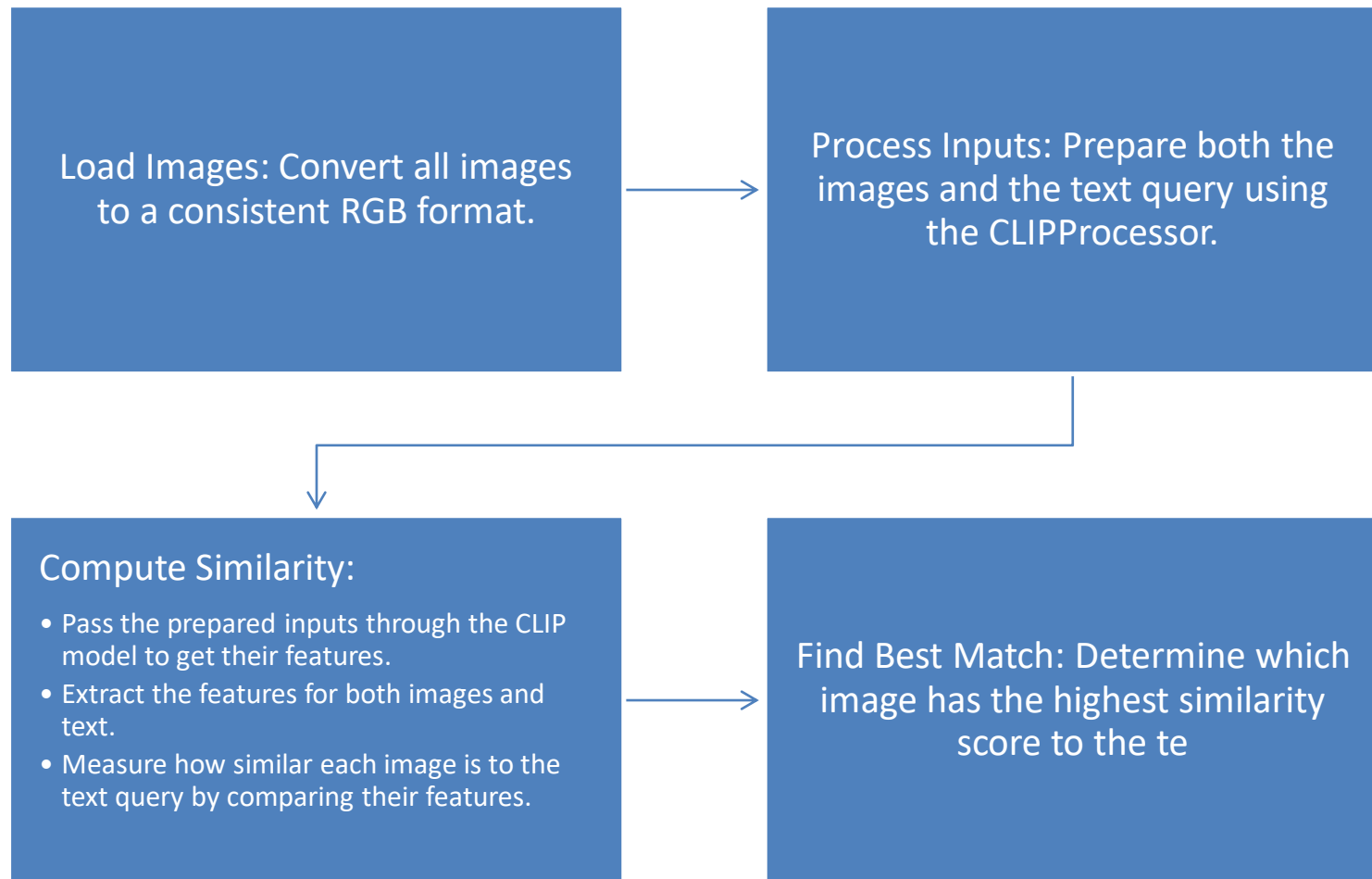


Image Captioning



Loading and
converting images to
RGB format.

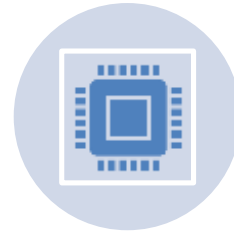


Using BLIP to
generate captions for
the images.

Challenges Faced



Model Loading and Compatibility:
Ensured model and library compatibility through careful selection of versions and regular updates.



Handling Large Image Sets: Limited the number of images processed at a time to manage memory usage and considered batch processing for future improvements.



Processing Speed: Used GPU acceleration to speed up processing and explored optimizations like model quantization for better performance.



Text and Image Preprocessing: Converted images to RGB format and used preprocessing tools to ensure consistent input formatting.

The slide features decorative curved lines in the top right and bottom left corners. These lines are composed of multiple overlapping layers in shades of light blue, grey, and orange, creating a sense of depth and movement.

Results

The CLIP model finds the most relevant image for a text query by comparing the similarity between text and image embeddings. The BLIP model generates accurate and contextually relevant captions for the images.

Search for Images by Caption

Enter a caption to search for images

football



Best Match Caption: football Similarity Score: 0.27

Results

Image Upload and Caption Generation

Choose an image file



Drag and drop file here

Limit 200MB per file • JPG, JPEG, PNG

Browse files



Screenshot from 2024-08-02 18-19-58.png 182.5KB



Uploaded Image

A SUNSET OVER THE CLOUDS

Results



Conclusion

The project showcases the effective use of pre-trained models for image search and caption generation. Despite facing some challenges, it delivers accurate and efficient results. Future improvements could include optimizing performance, managing larger datasets, and enhancing the user interface for better accessibility.