Problem Statement

Given a query image, we have to fetch the most similar image from a given database.

Solution Intuition

Any two similar images will lie closer to each other in the latent space whereas dissimilar images will lie far away. This is the basic governing rule with which we trained our model.
After this, the retrieval part simply scours the latent space to pick up the closest image in the latent space given the representation of the query image.

Algorithm

Our methodology is divided into two parts:

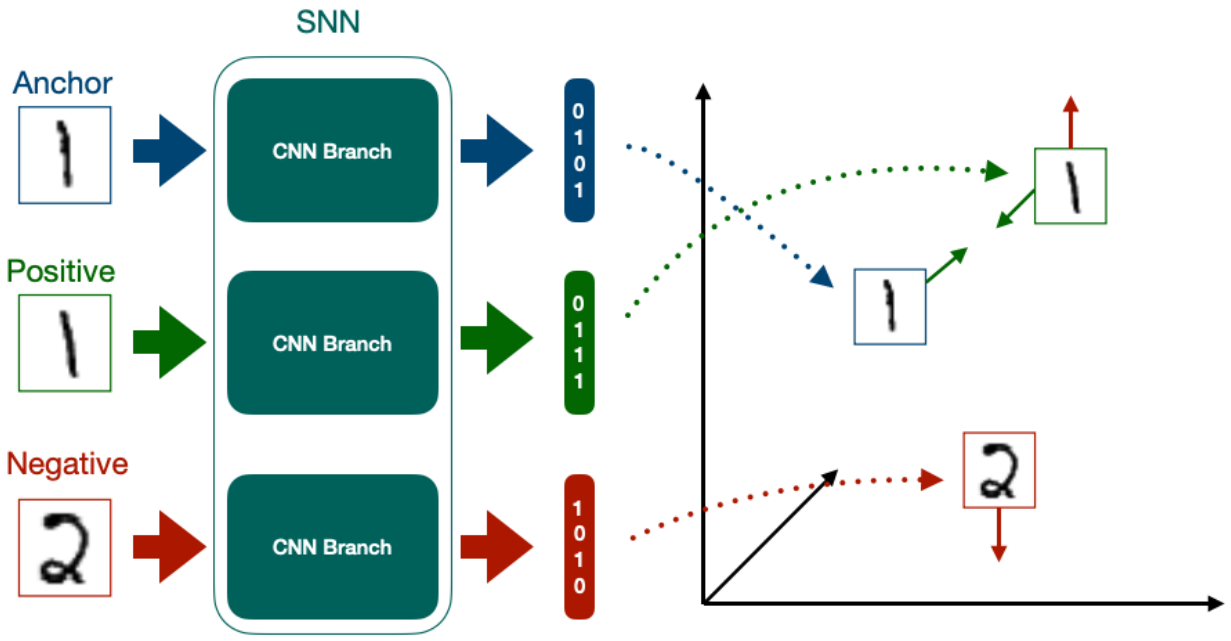
1. Image Representation in latent space
2. Search

Image Representation in the latent space

For this part we will be using Siamese Models for a similarity based representation. The basic idea behind Siamese Models is to use any state-of-the-art CNN models to train using Triplet Loss.
The basic component of this idea are triplets.

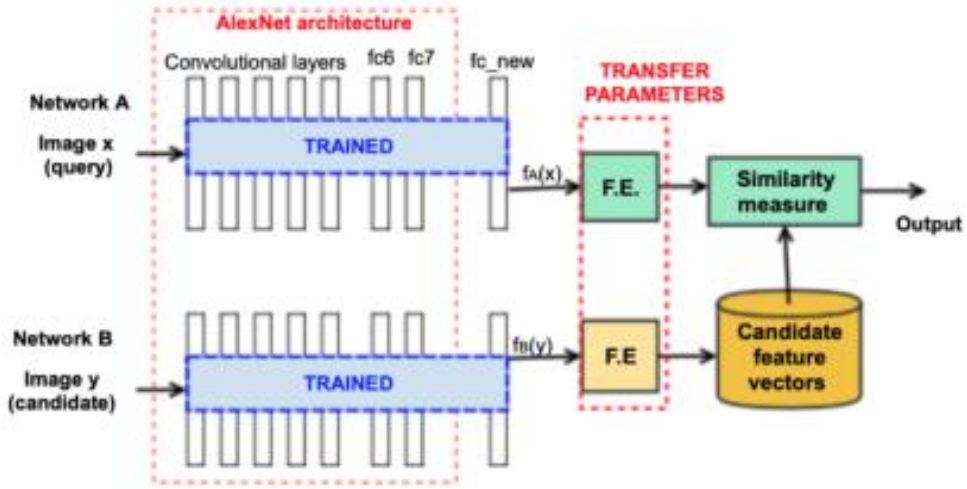
Triplets are 3 separate samples of data, let's say A (anchor),B (positive) and C (negative); where A and B are similar or have similar traits (same class maybe) while C is dissimilar to both A and B. The 3 samples altogether form one unit of the training data — the triplet.

The goal of the Siamese Model is to learn the representation of images and create a latent space where similar images (A and B) lie closer and dissimilar image(C) lie farther.



Search

After our model is trained to create a latent space, we can move forward with the search part.
Here, given a query image we will search the latent space where all the images of the database lies.
Using our trained siamese model, we will project the query image into the latent space along with all the images in the database.
After this, an exhaustive search based on euclidean distance is used to rank with the Top-k candidates. Out of which the top ranked candidate is retrieved.

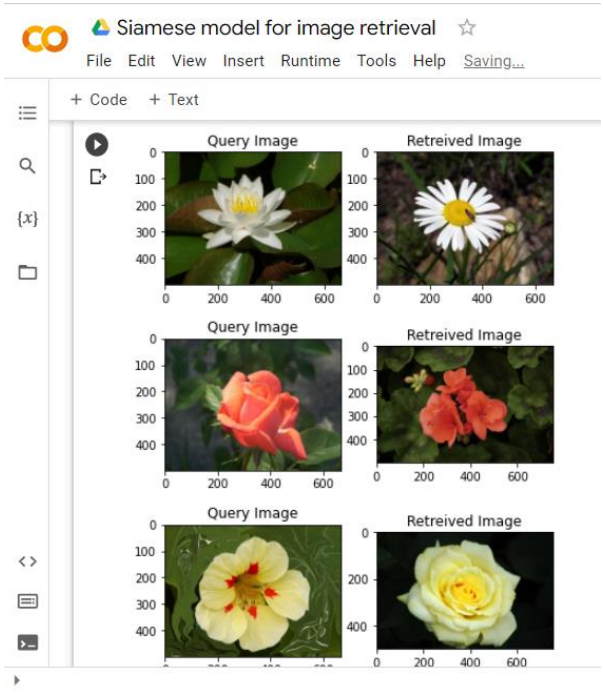


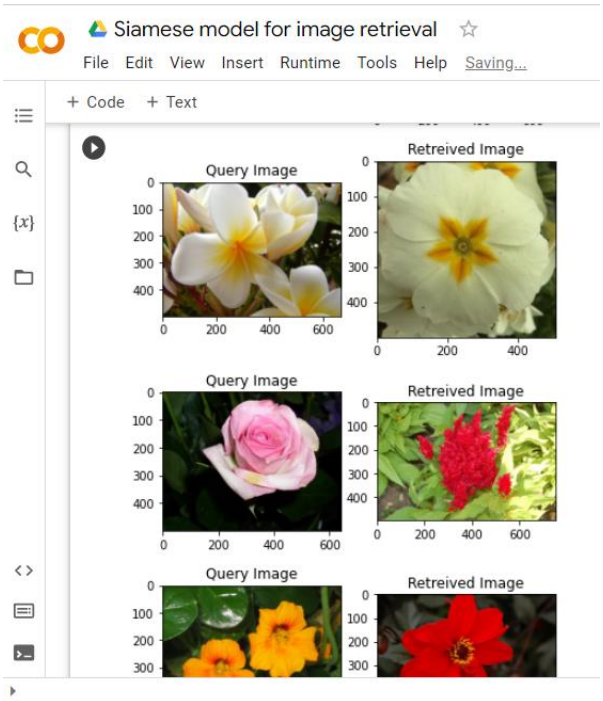
For our solution, we'll be using Faiss by Facebook Research. This is much faster than nearest neighbor and this difference in speed becomes much more prominent if we have a large number of images.

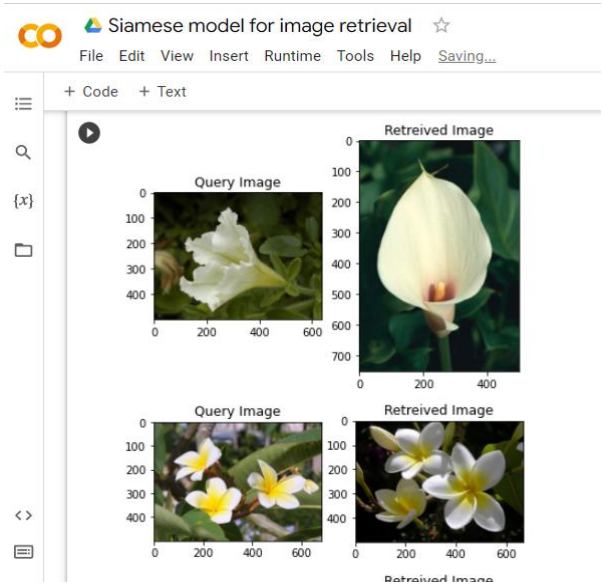
Time taken for each retrieval - 4.09 ms

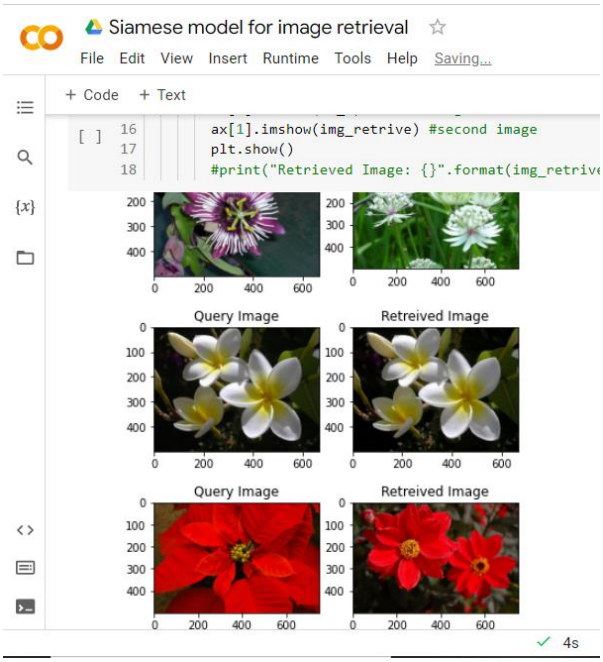
We have trained the model on the **Oxford 102 Flowers dataset,** which is a flower dataset having 102 categories and approx. 9,000 images for training.

Output after the initial training









Reference paper - https://arxiv.org/ftp/arxiv/papers/1906/1906.09513.pdf