Team Name: Elusive Images

Project Title: Exploring Content-based Image Retrieval

Project Summary:

Digital images have proliferated in the recent years and image retrieval systems have become a necessity with applications in web search, e-commerce, medicine, and recommendation systems.

Images retrieval systems are broadly divided into concept or content-based systems. Concept based systems rely on human annotations while content-based systems use feature extraction and object recognition to identify image content.

For our project, we are looking to explore content-based image retrieval systems. We will look at different similarity measures and their impact on retrieval outcomes. We also plan to explore couple of neural network architectures to understand the differences and benefits associated with each.

Approach:

Our plan is to implement an end-to-end image retrieval pipeline using content-based search for similar images in the database. We plan to explore and compare different methods for feature extraction, including classical methods like color histograms and text/shape-based features as well as couple of modern CNN architectures like Siamese Networks, GoogleNet, ResNet or ViT that has shown promising results in various computer vision tasks. We also plan to explore different similarity metrics like - Triplet loss, contrastive loss on image encodings, use of specific intermediate feature maps. If time permits, we also plan to get direct user feedback from our classmates for comparative human performance results.

We plan to use the Oxford flowers database that also includes text annotations.

Resources/Related Work:

The metric generally used to evaluate image retrieval tasks is Mean Average Precision. It seems that research is generally more focused on either the ranking component or the embedding component. A highly performant ranking technique that also achieves state-of-the-art MAP is diffusion with dynamic late fusion (using ResNet for the features) [8]. For feature embeddings, ResNet networks fine-tuned on relevant data for the task tend to perform the best, but even off the shelf methods are able to perform well [9], and more recently, some researchers have found the Vision Transformer to be able to outperform ResNet on some tasks [10].

[1] “A Hands-On Introduction to Image Retrieval in Deep Learning with PyTorch”, Maukh Bhattacharyya,

[2] “Image Retrieval”, https://paperswithcode.com/task/image-retrieval

[3] Fine-tuning CNN Image Retrieval with No Human Annotation, https://arxiv.org/pdf/1711.02512.pdf

[4] CNN Image Retrieval Learns from BoW: Unsupervised Fine-Tuning with Hard Examples, https://arxiv.org/pdf/1604.02426.pdf

[5] Learning Deep Representations of Medical Images using Siamese CNNs with Application to Content-Based Image Retrieval, https://arxiv.org/pdf/1711.08490.pdf

[6] Implementing Content-Based Image Retrieval with Siamese Networks in PyTorch

[7] Oxford Flowers Image Retrieval PyTorch, https://www.kaggle.com/mayukh18/oxford-flowers-image-retrieval-pytorch

[8] Efficient Image Retrieval via Decoupling Diffusion into Online and Offline Processing https://arxiv.org/pdf/1811.10907v2.pdf

[9] Deep Learning for Instance Retrieval: A Survey https://arxiv.org/pdf/2101.11282.pdf

[10] Investigating the Vision Transformer Model for Image Retrieval Tasks https://arxiv.org/pdf/2101.03771.pdf

Datasets:

Oxford 102 Flower (102 Category Flower Dataset): https://www.tensorflow.org/datasets/catalog/oxford_flowers102

Team Members:

Lewis C Haas, Hazel J John, Matthew R Caseres, Sai Chaitanya Nandipati