Multimedia WebDatabases Project Phase 1

Group Members

Name	Email	Student ID
Kesudh Giri	kgiri1@asu.edu	1225259867
Sai Parimala Vaishnavi Bhattaru	sbhatt86@asu.edu	1225574441
Naijoe Srinivasan	nsrini23@asu.edu	1224946138
Nikhil Venkat Ramanan	nvenka22@asu.edu	1225095677
Prashant Gaurav	pgaurav3@asu.edu	1229582226
Geeth Nischal Gottimukkala	ggottimu@asu.edu	1219532899

Abstract:

Keywords: ResNet50, Color Moments, Histogram of Gradients(HOG), Feature Descriptors, Euclidean Distance, Cosine Similarity, Vector Space Search, Caltech101.

In this project, we investigate various methods for extracting features from images, including Color Moments, Histograms of Oriented Gradients (HOG), and the activation layers from a pre-trained ResNet neural network. These methods help us create descriptions of what makes each image unique by capturing its visual characteristics and patterns. Color Moments help us understand the colors in the images, HOG tells us about local patterns and shapes, and ResNet helps us dive deep into the image's details. Each of these methods gives us a different way to analyze and represent images.

Similarity measures such as Euclidean distance and inner product are used to check the efficacy of these feature descriptors. Given a query image, and a number k, we seek to find and retrieve the k most similar images to the query image. This is also achieved through the comparisons of the feature descriptors These measurements allow us to compare the feature descriptions of a query image with those of a database of images. Then, we can find and retrieve the top k images that are most similar to the query image, which helps us understand image similarity and ranking.

Overall, our project explores various techniques for extracting image features and measuring image similarity. This research contributes to the improvement of image search and recommendation systems across different fields.

Introduction:

Terminology:
Residual Neural Network (ResNet): It is a positive feed forward neural network architecture employed in Computer Vision based applications.
Histogram Oriented Gradients (HOG): It is a feature extraction technique designed to capture local gradient or edge information in an image.
Color Moments: These are a set of statistical measures used to capture and describe color information in an image. The three main color moments are Mean, Standard Deviation, and Skewness
Euclidean distance: Euclidean distance is a common measure of similarity used in image processing and computer vision to compare the similarity between two feature vectors representing images.
Cosine similarity: It quantifies the cosine of the angle between two vectors in a high-dimensional space. In the context of image feature vectors, it is widely used as a similarity measure in comparing image feature vectors.

Goal Description:
In the world of multimedia, managing and searching through vast databases of images is becoming increasingly challenging owing to the sheer volume of data being generated on a daily basis. For this project, we seek to implement functionality to extract feature descriptors such as Histogram of Oriented Gradients, Color Moments, and ResNet layers and to store the images and their corresponding feature descriptors in a database. Upon being queried an image and a number ‘k’, the feature descriptors are used to calculate similarity measures between the query image and other images in the database and outputs the ‘k’ most similar images.

Based on the requirements specified, the task can be divided into 4 subtasks:
Task 1 - Load the image dataset and ResNet model
Task 2 - Display Feature descriptors when supplied an ImageID
Task3 - Store feature descriptors for all images in the dataset
Task4 - Display k most similar images using each feature descriptor when supplied with an ImageID and k

Assumptions:

- We assume each image is of the ‘**RGB**’ color schematic or can be converted into one. I.e., we assume that each image has either 2 channels or 3 channels

- We assume the system running the application has sufficient heap memory to run a MongoDB server and perform memory crucial tasks.

Description:

Our solution to this problem consists of an amalgamation of Machine Learning techniques and the use of a No-SQL database, MongoDB in this instance.

Foremost, we prepared the Caltech101 dataset by saving it in our local system in a location accessible by the MongoDB database. In order to enable this, the user need be have a MongoDB Community Server running on the system with a connection to the default localhost port 21017 or any other free port. After preprocessing the dataset as required for the computation of every feature descriptor, we calculate the feature descriptor and store it in our Mongo database along with the corresponding images to which the feature descriptors pertain.

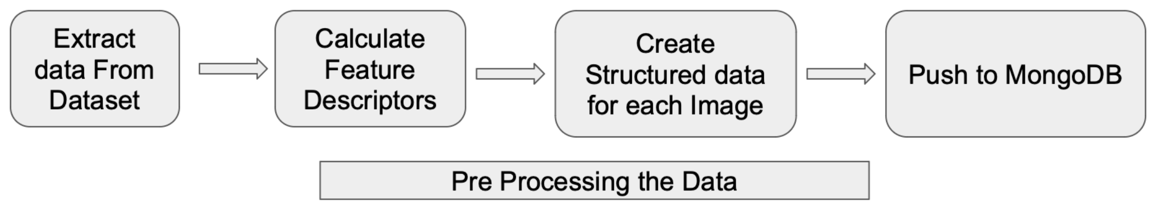
In order to tackle the latter part of the problem statement, which seeks to display ‘k’ similar images to a query image, we designed a simple GUI which asks the user to enter an imageID and the value k. Behind the scenes, we retrieve the query image and find the similarity scores against all other images in the database w.r.t the given feature descriptors. Following this, we rank the top k similar images and display those to the user.

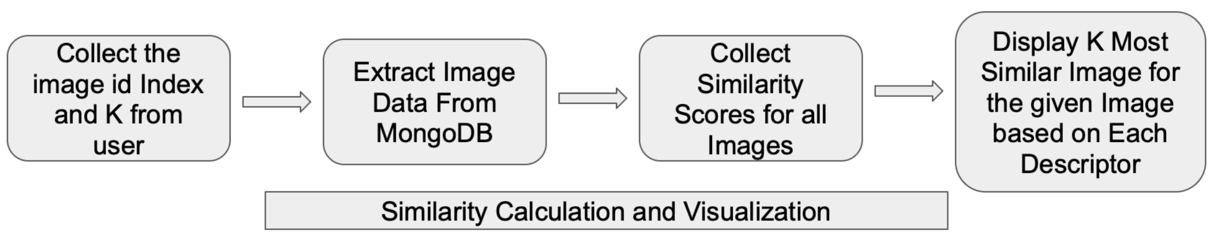
System Requirements:

- A system with a minimum of 8GB of RAM.
- MongoDB server connected to a port on the local host.
- Python 3.xx.

- Jupyter or similar notebooks capable of running .ipynb file.(should run on the local host.)

Interface specifications:





Related work:

Histogram of Oriented Gradients(HOG) is a novel method introduced in the paper titled "Histograms of Oriented Gradients for Human Detection" by Navneet Dala and Bill Triggs, and published in the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) [1]. The main objective of this paper is to develop an effective and robust technique for detecting human figures in images, particularly in the context of computer visions and surveillance applications. The authors propose a new feature representation to tackle this problem called HOG which divides a given image into small cells and computes the HOG within each cell. These histograms encode the local edge or gradient information in different directions.

Building on to the above, color moments are measures which characterize the color distribution in an image. The paper titled "Evaluating Color Descriptors for Object and Scene Recognition" by Koen van de Sande, et al., published in the IEEE Transactions and Pattern Analysis and Machine Intelligence [2] focuses on various color descriptors including color moments for the tasks of object and scene recognition. The paper evaluates these color descriptors and concludes that color moments can be considered effective descriptors when the color distribution in an image is essential for the task. Such tasks include color-based image retrieval, and certain image classification tasks (e.g.. Flower species recognition). Additionally, the paper summarizes that color moments are more suitable for real-time or resource-constrained applications due to the computational efficiency of computing color moments compared to other color descriptors.

Furthermore, papers such as "A Study on CNN Transfer Learning for Image Classification" by M. Hussain, et al., presented at the 18th UK Workshop on Computational Intelligence [5], and "Deep Residual Learning for Image Recognition" by Kaiming He, et al., presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [7] explore the application of Convolutional Neural Network transfer learning for image classification. Transfer learning is a machine learning technique where pre-trained neural network models, often trained on large datasets for different tasks, are fine-tuned or used as feature extractors for specific image classification tasks. Kaiming He, et al., propose a novel architecture called ResNet short for "Residual Network", which is a deep neural network architecture that is considered a significant advancement in the field of Computer Vision and Deep Learning. The key innovation in ResNet is the usage of residual connections to address the problem of vanishing gradients in very deep networks, which can make training extremely deep networks challenging.

Conclusions:

- There is significant variance in the accuracy of the different feature descriptors in extracting similar images.
- Color Moments and Histograms of Oriented Gradients (HOG) exhibit limitations in capturing key image characteristics. Mismatches occur in similarity calculations for specific images. The issue with Color Moments may arise from its simplified calculation, focusing on Mean, Standard Deviation, and Skew within a (30x10) cell. This simplification may result in the loss of precise color placement information, leading to inconsistent outcomes.
- Typically, Histograms of Oriented Gradients are computed for smaller portions of an image, like (8x8) or (16x16) cells. However, the specified size, (300x100), would only allow for 100 gradients to be drawn, which might not accurately represent the image's details and edges.
- Among all the feature descriptors we tested, the ones using pre-trained ResNet weights, along with image normalization based on the average and standard deviation values across the entire Caltech101 dataset, gave us the best results.
- Color Moments and HOG deliver accurate results for most of the dataset, even though there are some outliers.
- Using MongoDB to store and fetch image data along with feature descriptors made it easy and efficient to access data for individual images when the program is running
- The pre-trained ResNet50 model delivers precise results on the Caltech101 dataset, demonstrating its versatility for various applications, despite being initially trained on the ImageNet dataset

Bibliography:

[1] Dalal, N. and Triggs, B., 2005, June. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). Ieee.

[2] Van De Sande, K., Gevers, T. and Snoek, C., 2009. Evaluating color descriptors for object and scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, *32*(9), pp.1582-1596.

[3] Danielsson, P.E., 1980. Euclidean distance mapping. *Computer Graphics and image processing*, *14*(3), pp.227-248.

[4] Dehak, N., Dehak, R., Glass, J.R., Reynolds, D.A. and Kenny, P., 2010, June. Cosine similarity scoring without score normalization techniques. In *Odyssey* (p. 15).

[5] Hussain, M., Bird, J.J. and Faria, D.R., 2019. A study on cnn transfer learning for image classification. In *Advances in Computational Intelligence Systems: Contributions Presented at the 18th UK Workshop on Computational Intelligence, September 5-7, 2018, Nottingham, UK* (pp. 191-202). Springer International Publishing.

[6] Shu, C., Ding, X. and Fang, C., 2011. Histogram of the oriented gradient for face recognition. *Tsinghua Science and Technology*, *16*(2), pp.216-224.

[7] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[8] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H. and He, Q., 2020. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, *109*(1), pp.43-76.

[9] Bansal, M., Kumar, M., Sachdeva, M. and Mittal, A., 2021. Transfer learning for image classification using VGG19: Caltech-101 image data set. *Journal of ambient intelligence and humanized computing*, pp.1-12.

[10] Ba, J.L., Kiros, J.R. and Hinton, G.E., 2016. Layer normalization. *arXiv preprint arXiv:1607.06450*.

Appendix:

Our project's achievements are a testament to our team collaboration. We worked cohesively, leveraging each team member's strengths and expertise. Regular meetings and open communication were pivotal in maintaining our project's momentum.
Out of the concerned work tasks in line with the problem statement, we worked on the MongoDB Connectivity, DB operations, Visualizing as a group, and the rest of the pending tasks were evenly divided within the group as shown below:

- Kesudh Giri: Worked on ResNet-AvgPool 1024 and calculated the similarity scores using Cosine similarity for ResNet-Avg-pool Descriptor. In the report, he worked on the Conclusion.

- Sai Parimala Vaishnavi Bhattaru: Worked on ResNet-Layer3 1024 and calculated the similarity scores using Euclidean distance for ResNet-Layer3 Descriptor. In the report, he worked on the System Requirements and Interface Specification.

- Naijoe Srinivasan: Worked on ResNet FC 1000 and calculation of the similarity scores using Cosine similarities for ResNet-FC Descriptor. In the report, he worked on the Related Works section.

- Nikhil Venkat Ramanan:Worked on Color moments and Calculation of the similarity scores using Euclidean distance for Color moments Descriptor.In the report, he worked on the Description.

- Prashant Gaurav:Worked on HOG(Histogram of Oriented Gradients) and Calculation of the similarity scores using Cosine similarities for HOG Descriptor.In the report, he worked on the Abstract and Bibliography.

- Geeth Nischal Gottimukkala:Worked on HOG(Histogram of Oriented Gradients) and Calculation of the similarity scores using Cosine similarities for HogDescriptor. In the report, he worked on the Introduction.