A project on Similar Images Recommender System for eye frames

Kartikay Kumar, Neeti Asthana, Rishu Khurana

***Abstract*-** Retail Analytics space is evolving everyday towards providing a better customer experience every day. Image based product search is one such area of work. Similarly, this project explores the idea in the online eye frames product space. The project is about comparison of image features and selecting the top 10 similar images to a given eye frame image.

INTRODUCTION (MOTIVATION)

There are many times when we find it difficult to recall the exact name of a product or cannot describe it clearly for search or would love to reorder what we saw someone else wearing. In all these cases and many more, "visual product search" is what comes to the rescue. Instead of describing, we could just take a picture of the product and search for it. This work is about developing a model which would find most similar images to a particular product image. An article on mycustomer.com explains how Google has revolutionized the way we search. Consumers now demand image-based search functionalities when shopping online (not just text-based search). In a survey of 1,000 consumers, it was found that three quarters of consumers (74%) said traditional text-based keyword queries are inefficient in helping them find the right items online. Another 40% would like their online shopping experience to be more visual, image-based and intuitive. Potential customers are window shopping "by image" on search engines like Google. Retailers whose search engine optimize their product images and listings can gain a competitive advantage with consumers who prefer to shop this way.

RELATED WORK AND ARTICLES

Below are the related research papers and articles for reference about the project idea:

A. *S. Umer, Partha Pratim Mohanta, Ranjeet Kumar Rout, Hari Mohan Pandey: Machine learning method for cosmetic product recognition: a visual searching approach*

A cosmetic product recognition system is proposed in this paper. For this recognition system, a cosmetic product database has been processed that contains image samples of forty different cosmetic items. The purpose of this recognition system is to recognize Cosmetic products with their types, brands and retailers such that to analyze a customer experience what kind of products and brands they need. This system has various applications in such as brand recognition, product recognition and also the availability of the products to the vendors. The implementation of the proposed system is divided into three components: preprocessing, feature extraction and classification. During preprocessing the color images were scaled and transformed into gray-scaled images to speed up the process. During feature extraction, several different feature representation schemes: transformed, structural and statistical texture analysis approaches have been employed and investigated by employing the global and local feature representation schemes. Various machine learning supervised classification methods such as Logistic Regression, Linear Support Vector Machine, Adaptive k-Nearest Neighbor, Artificial Neural Network and Decision Tree classifiers have been employed to perform the classification tasks.

B. *Image Classification for E-Commerce — Part I-* *https://towardsdatascience.com/product-image-classification-with-deep - learning-part-i-5bc4e8dccf41*

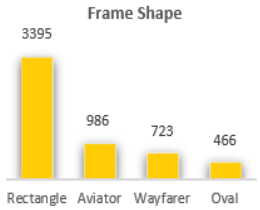
In this article, it is explained how images are used to solve one of the most popular business problems i.e. classification of products. A giant online marketplace like Indiamart has thousands of macro categories for listing various products. A product must get mapped under the most appropriate micro category on the platform. The goal of this post is to build intuition and understanding of how neural networks can be trained to identify the micro category of a product using its images.

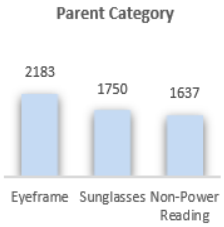
C. *Building a Reverse Image Search Engine: Understanding Embeddings- https://www.oreilly.com/library/view/practical-deep-learning/9781492034858/ch04.html*

This page contains the process of building reverse image search engine i.e. it consists of steps like performing feature extraction and similarity search on Caltech101 and Caltech256 datasets, learning how to scale to large datasets (up to billions of images), making the system more accurate and optimized, analyzing case studies to see how these concepts are used in mainstream products. The page has information of locating similar images with the help of embeddings. It contains work on a level further by exploring how to scale searches from a few thousand to a few billion documents with the help of ANN algorithms and libraries including Annoy, NGT, and Faiss. Also, the page has process of fine tuning the model to your dataset can improve the accuracy and representative power of embeddings in a supervised setting. To top it all off, it has work on how to use Siamese networks, which use the power of embeddings to do one-shot learning, such as for face verification systems.

DATASET DESCRIPTION AND PREPROCESSING

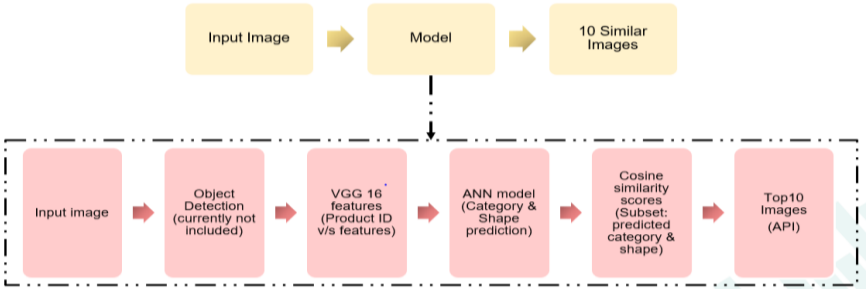
The dataset for the problem is a collection of 5570 eye frames. The file contains product name, product ids, frame shape, parent category and URLs of the eye frame images. The parent category feature has 3 classes: Eye frame, Sunglasses and Non-Power Reading while the frame shape feature has 4 classes: Rectangle, Wayfarer, Aviator and Oval. The data was preprocessed so that it could be fed to the model for training. A dictionary was prepared which contained the array values of the images from URLs and values against product ids as the key. This image data was fed to pretrained VGG16 for feature generation. The features generated were appended to the main data frame against their respective product ids. Each image is resized to (224,224,3) and size of the features generated is (7,7,512).





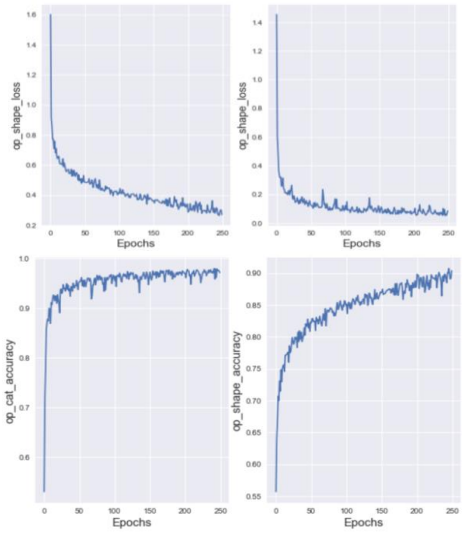
METHODOLOGY

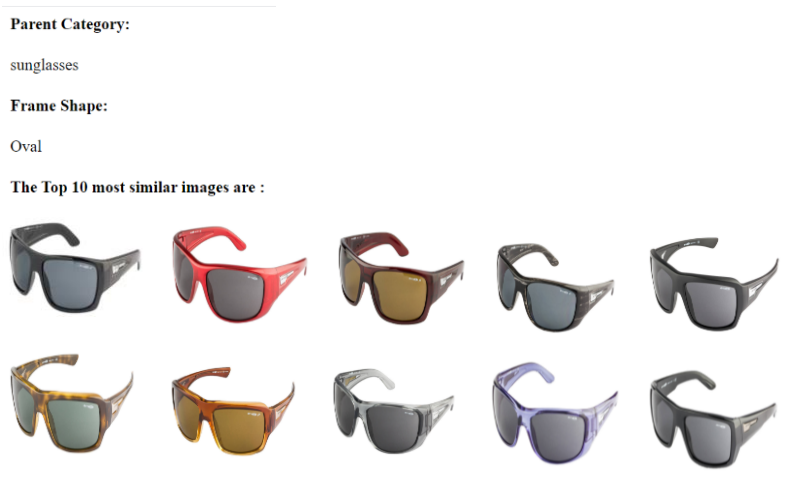
After the features from VGG16 are generated and stored for 5570 images in the dataset. A new data frame is prepared which contains product ids, parent category and frame shape. An artificial neural network (ANN) is trained on the dataset through functional API with inputs as the features and prediction of two outputs as Frame shape and Frame category. A new image when uploaded by user is first processed through VGG16 for feature generation, The features generated are made as the input to the trained ANN for its frame shape and category prediction. After the frame shape and category prediction of the submitted image, the feature array of the image is compared to the features of the subset of frames of the predicted shape and category from the main data frame of features. This approach speeds up the process of similarity score calculation since only a subset is considered for calculation and not the complete dataset. Pairwise Cosine similarity is used as a metric for calculating the similarity of features. The similarity score thus generated are sorted in a descending order and URLs for top 10 similarity scores are selected for output.



RESULTS

The ANN has Adam (learning rate=0.01) optimizer and Sparse categorical Entropy loss with Accuracy as metric of evaluation and was trained for 250 epochs with a batch size of 128 and validation split of 20%. After training, the model achieved 97.2 % accuracy, 0.0845 loss for category prediction and 90.4% accuracy, 0.2700 loss for shape prediction. Below are the plots for training accuracy and training loss. The evaluation of final results i.e., if the recommended eye frames are similar to the eye frame image uploaded, displayed depends on user evaluation. After this, a flask API was developed for improving the user interface for model testing.





INDIVIDUAL TASKS

S.No.	Tasks	Responsibility (Resource Allocation)
1	Project Proposal	Kartikay Kumar, Neeti Asthana, Rishu Khurana
2	Project Study & Scope Finalization	Kartikay Kumar, Neeti Asthana, Rishu Khurana
3	Model Architecture Conceptualization	Kartikay Kumar, Neeti Asthana, Rishu Khurana
4	Data Gathering & Storage	Rishu Khurana
5	Exploratory Data Analysis	Neeti Asthana
6	ANN Model for shape & category prediction	Kartikay Kumar, Neeti Asthana, Rishu Khurana
7	Mid Term Presentation	Kartikay Kumar, Neeti Asthana, Rishu Khurana
8	Object Detection Model	Kartikay Kumar, Neeti Asthana
9	Feature Extraction through pre-trained network	Kartikay Kumar, Neeti Asthana, Rishu Khurana
10	Image similarity Matrix & Selection of top 10 similar images	Neeti Asthana
11	Model results analysis and retraining	Kartikay Kumar, Neeti Asthana, Rishu Khurana
12	API development	Rishu Khurana
13	HTML page development	Rishu Khurana
14	Final Submission Presentation	Kartikay Kumar, Neeti Asthana, Rishu Khurana

EXPECTED FINAL OUTCOME

In this project, ANN for frame shape and category prediction and object detection with VGG would be adopted to build a model for recommending top 10 eye frames similar to the eye frame uploaded by user. In further study, we will try to conduct experiments on larger data sets or try to tune the model so as to achieve the state-of-art performance of the model.

ACKNOWLEDGMENT

This work would be supported by faculties from the Indraprastha Institute of Technology, Delhi as a part of curriculum in the Post Graduate Diploma in Data Science and AI course

REFERENCES

[1] Lowe, David G. "Distinctive image features from scale-invariant keypoints." International journal of computer vision 60.2 (2004): 91- 110.

[2] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.

[3] Ilaria Bartolini, Paolo Ciaccia, and Marco Patella. WINDSURF: A region-based image retrieval system. Technical Report CSITE-011-00, CSITE–CNR,2000. Available at URL http://www-db.deis.unibo.it/MMDBGroup/TRs.html.

[4] Zagoruyko, Sergey, and Nikos Komodakis. "Learning to compare image patches via convolutional neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[5] Agrawal, Pulkit, Ross Girshick, and Jitendra Malik. "Analyzing the performance of multilayer neural networks for object recognition." European Conference on Computer Vision. Springer International Publishing, 2014.

[6] J. Wang, “Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication),” *IEEE J. Quantum Electron.*, submitted for publication.

[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Regionbased convolutional networks for accurate object detection and segmentation. IEEE transactions on pattern analysis and machine intelligence, 38(1):142–158, 2016.

AUTHORS

First Author – Kartikay Kumar, kartikay.opd20006@iiitd.ac.in, B.E., Netaji Subhas Institute of Technology, New Delhi, India
Second Author – Neeti Asthana, neeti.opd20005@iiitd.ac.in, M.Sc. Applied Mathematics.
Third Author – Rishu Khurana, rishu.opd20036@iiitd.ac.in, B.Tech., Manav Rachna College of Engineering, Faridabad, Haryana.