An Efficient Technique for Apparel Retrieval using Transfer Learning and K-Means Clustering

Eshwar S G, Gautham Ganesh Prabhu J, A V Rishikesh, Charan N A
Undergraduate Students
{ eshwarsg24, gautham.prabhuj, avrishikesh2, charan1994 }@gmail.com

Umadevi V
Associate Professor, Department of Computer Science and Engineering
B. M. S. College of Engineering, Bengaluru, India
umadevi.cse@bmsce.ac.in

***Abstract* — Apparel retrieval involves fetching similar clothings to the query image. This is one of many applications of image based search. In our work, we have optimized the process of finding similar apparels. We make use of clustering within a class to broadly classify images and find similar images within the cluster. We compare the query image from a class with the images within a cluster to which the query image belongs to, rather than the entire dataset. Hence this technique reduces the time required to find similar images.**

Keywords—CNN, Transfer Learning, ImageNet Inception model, Similarity measure, K-means, Silhouette Coefficient.

I. INTRODUCTION

Apparel retrieval is very useful to help the users to efficiently search the apparel they want. Currently, the mainstream clothing retrieval methods are attribute semantics based. The problem with this approach is a lot of time and effort goes into this process of categorisation. Even after spending time and effort to develop efficient categories the users still do not a have convenient way to search for the required item without going through this complicated hierarchical attribute structure. This process is cumbersome and general users even after spending a reasonable amount of time would not be able to find the item they are looking for. With the improvement in image processing technologies, the natural alternative would be to use them to ease the process of apparel retrieval.

Our focus lies on the task of similar apparel retrieval given a query image. We make use of a pre-trained convolutional neural network (CNN) architecture [1] and retrain the final layer of this network to meet our requirements using transfer learning. This is responsible for the classification of the image. Then we employ kNN algorithm to find similar images to the query image as implemented in our previous work [2]. Now we make use of K-means clustering algorithm to form clusters within each class. Hence during retrieval, similar apparels are suggested from the cluster which is nearest to the query image within the predicted class. With his approach the user will be able to find the required item easily and efficiently.

II. RELATED WORK

In the past considerable amount of work has been done in the area of clustering of images. Thomas Deselaers, Daniel Keysers and Hermann Ney [3] worked on clustering of visually similar images to improve the the results of image search engines, for this task they used features invariant against translation and rotation to represent the content of the images and k-means and LGB cluster algorithms to present the images in a convenient manner to the user. Guoping Qiu [4] worked on image and feature co-clustering for this task the author makes use of a computational energy function suitable for co-clustering images and their features and then optimizes this function using hopfield model based stochastic algorithm. Brenden J Frey and Delbert Dueck [5] designed a method called affinity propagation. In this method similarity measure between a pair of data points is taken as input and then messages are exchanged between data points to form high quality clusters. Giridharan Iyengar and Andrew B Lippman [6] worked on clustering images for efficient retrieval using relative entropy. They make an assumption that visual features are represented by probability densities and based on this assumption they design a clustering algorithm for probability densities. VSVS Murthy, E Vamshidhar, JNVR Swarup Kumar and P Shankara Rao[7] extract color feature from images and combine hierarchical clustering algorithm with K-means to cluster images to implement content based image retrieval.

III. Retrieval Time Analysis

Under normal circumstances the retrieval time is the sum of time taken to calculate the similarity between the query image and the entire corpus of images (T_{sim}) and time to rank the results by sorting (O(nlogn)). Therefore the total search time is:

$$T_{normal} = nT_{sim} + O(n\log n) \qquad (1)$$

where n in the total number of images in the dataset

When the images in the database are classified into categories, the retrieval time is the sum of the time taken to classify the the image ($T_{classify}$), time to compute similarity (T_{sim}), and time to rank the results. Therefore the total search time is:

$$T_{categorized} = T_{classify} + cT_{sim} + O(c\log c) \qquad (2)$$

where c is the number of images in each category

When the images in the categories are further clustered, the retrieval time is the sum of time taken to classify, time to identify the nearest cluster, time to compute similarity within the nearest cluster and the time to rank the images. Therefore the total search time is:

$$T_{cluster} = T_{classify} + (k+m)T_{sim} + O(m\log m) \qquad (3)$$

Where k is the number of clusters; m is the number of images in the cluster closest to the query.

As $k+m < c < n$ and $2 <= k <= 5$ in our clusters from Eq. (1) (2) and (3)

$$T_{cluster} < T_{categorized} < T_{normal}$$

Our approach follows Eq. (3), we first make use of a convolutional neural network [1] which has been pre-trained and retrained using transfer learning using our dataset to extract features and then we make use of softmax regression for classification and finally we make use of k-means clustering algorithm for clustering.

IV. Methodology And Implementation

The block diagram of our system used for retrieving apparels is shown in Fig. 1. The different blocks in the system are:

(1) **Query Image:** The query image is accepted as input
(2) **Vectorize Image:** The query image is vectorized by passing it through the layers of the CNN and by making use of the pool:3 feature tensor. The resulting vector is of dimension 2048
(3) **Classifier:** The final layer in the CNN is retrained to meet our requirements and Softmax Regression is used to predict the probabilities of the classes in the order of confidence
(4) **Identifying Nearest Cluster:** Once classified, the nearest cluster which the image belongs to is identified
(5) **Similarity:** Euclidean distance is used as a similarity measure to linearly search for similar images in the nearest cluster
(6) **Suggestions:** The distances obtained are sorted and the top K suggestions are displayed as the search results

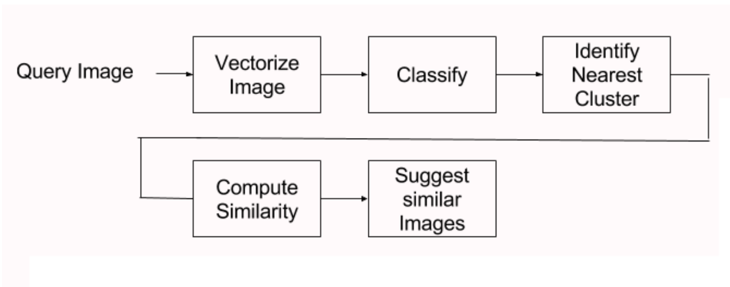

Fig 1. Block diagram of the system used for suggesting similar images

V. Convolutional Neural Network And Transfer Learning

As our dataset is small it is not good enough to train the whole CNN as it might overfit the data. Since the data is similar to original data, higher level features in the CNN can be relevant to this dataset as well. Training the linear classifier on CNN features is a good option. Hence we make use of transfer learning to retrain the final layer of a pre trained convolutional neural network model.

The GoogLeNet architecture [8] which we make use of is a 22 layer deep neural network and was initially trained on the ImageNet dataset[9] which consists of more than a one million images ranging over thousand classes. Hence we use the weights from the GoogLeNet which is trained on the ImageNet dataset. We remove the final layer of the Inception v3 GoogLeNet model which is shown in Fig. 1. and train a new final layer using our apparel dataset.

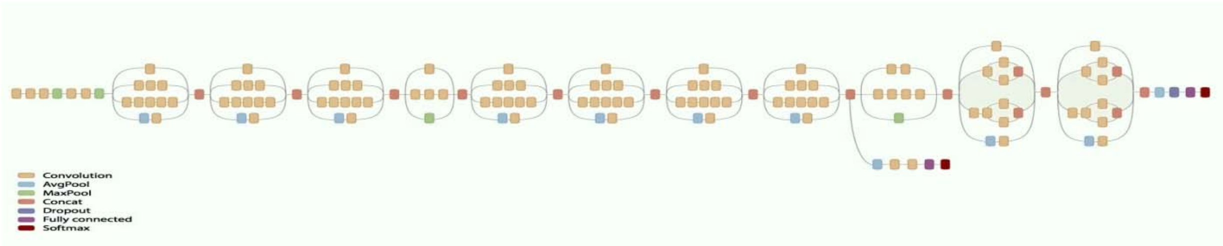

Fig 2. Diagram of the Inception V3 network

VI. Clustering

Clustering is a process of mutually exclusive partitioning of the feature space in a way which is meaningful to the application context. Nearest neighbor search can be performed in an efficient manner within the clusters. The unique aspect of this system is, the utilization of CNN as a feature extractor and k-means clustering techniques. Here we make use of a pre-trained CNN to obtain the feature vectors of the image. This is followed by retraining the final layer of the CNN by making use of the softmax activation function. The final layer is responsible for the actual classification. K-means is used to cluster the images within a given class. This results in better image search results. Hence in our system, firstly the query image is classified. After classification, the nearest cluster within the identified class is found. Finally, similarity search within the given cluster using Euclidean Distance as a metric is carried out. This drastically reduces the time taken to find the search results.

The k-means algorithm takes an input parameter K the number of clusters to be formed, and partitions a set of n data points into K clusters so that the resulting intra-cluster dissimilarity is low. Euclidean distance is used as a metric to assign a data point to its nearest cluster. After assigning this new point, the cluster centroid is updated. This process continues until the criterion function converges.

One of the challenges associated with K-means is deciding upon the optimal value of k.

We make use of Silhouette analysis to find out the optimal number of clusters K within a given class.

The Silhouette Coefficient (S) is calculated using the mean intra-cluster distance *a* and the mean nearest-cluster distance *b* for each sample as shown in Eq. (4).

$$s = \frac{b-a}{\max(b,a)} \qquad (4)$$

Hence a large value for silhouette coefficient implies that b >> a which indicates that the dissimilarity between the points in the same cluster is less. Hence the clusters formed are better.

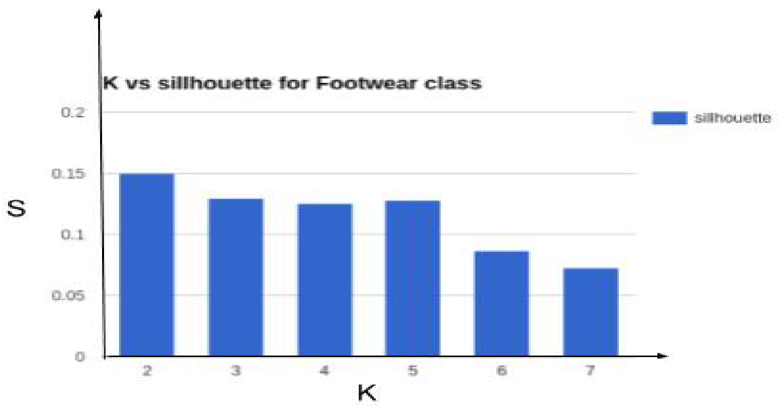

Fig 3. Results of K vs Average Silhouette score for the footwear class

We iterate through different values of K and measure the silhouette coefficient and then select the value of K which is maximum and in agreement with the nature of data in our dataset. For example, the Fig. 3. above shows a plot of K vs Silhouette coefficient for the footwear class. We can see that the maximum value of Silhouette coefficient occurs at K=2, but in our footwear class there exist different categories of footwear namely casual, formal, ladies footwear, kids footwear etc. Hence we decide upon the value of K=5 as it is consistent and in agreement with our dataset. The clustering within other classes are carried out in a similar manner.

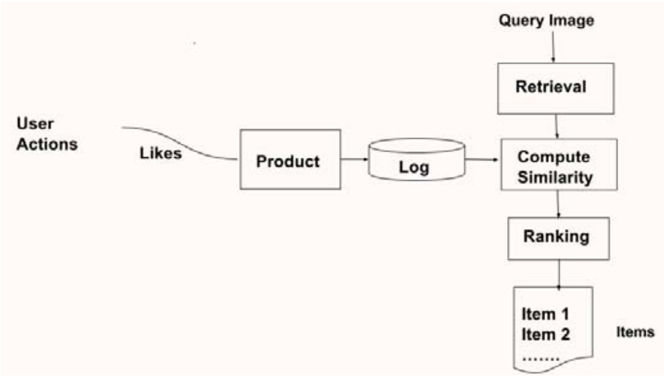

Fig 4. Recommendation scheme based on user preference

In addition to the above mentioned functionality, we implemented a feature which saves the likes and preferences of users. The block diagram of the system is shown above in Fig. 4. The user actions on the applications such as his likes are recorded into a database. This information in turn is used to identify a user's affinity to a product from a given cluster and recommend products to the user.

VII. RESULTS

Here are some of the results of our implementation. Table I. indicates the total number of classes and the number of clusters formed within each class.The results obtained by image search are shown in Fig 5,6. The first image is the query image and the next three are the top three results pertaining to the query. The screenshots of the user interface developed are shown in Fig. 7,8,9.

TABLE I. CLASSES OF APPARELS

Class ID	Apparel Classes/Types	Number of Images	Number of Clusters
1	T-Shirt	841	2
2	Pants	859	2
3	Saree	865	2
4	Ladies Kurta	894	2
5	Footwear	789	5

Fig 5. Search results for Footwear (Flip-flops)

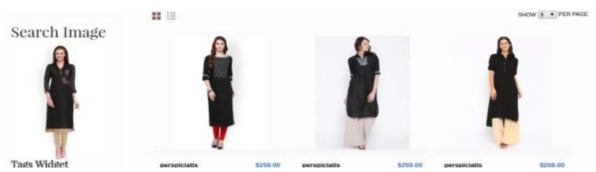

Fig 6. Search results for Ladies Kurta

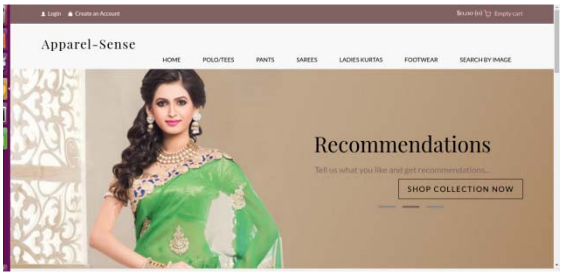

Fig 7. Home Page

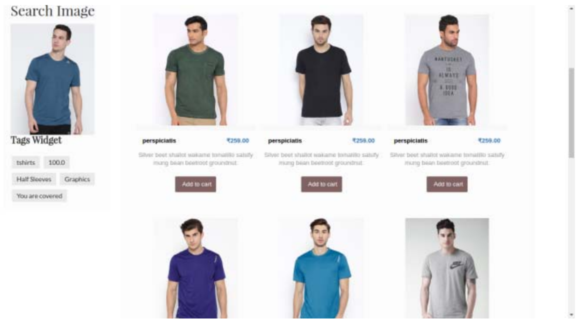

Fig 8. Search results Page

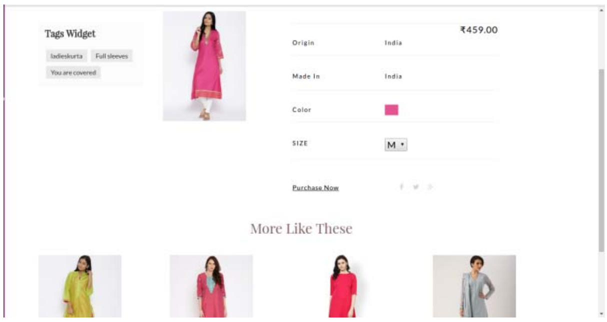

Fig 9. Product View page

VIII. Conclusion And Future Work

In our approach we make use of high dimensional vectors which increases the time taken to compute similarity. Hence as a task to be taken up in the future we intend to make use of dimensionality reduction techniques such as Deep AutoEncoder Neural Networks to reduce the dimensionality without hindering the performance and results of our system and thereby making it faster.

Acknowledgement

We are grateful to BMS College of Engineering for having provided us with the facilities needed for the successful completion of this paper. The work reported in this paper is supported by the college through the TECHNICAL EDUCATION QUALITY IMPROVEMENT PROGRAMME [TEQIP-II] of the MHRD, Government of India.

References

[1] Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. "Rethinking the inception architecture for computer vision." arXiv preprint arXiv:1512.00567 (2015).
[2] Eshwar, S. G., A. V. Rishikesh, N. A. Charan, and V. Umadevi. "Apparel classification using Convolutional Neural Networks." In ICT in Business Industry & Government (ICTBIG), International Conference on, pp. 1-5. IEEE, 2016.
[3] Deselaers, Thomas, Daniel Keysers, and Hermann Ney. "Clustering visually similar images to improve image search engines." hist 1, no. 1 (2003): 1.
[4] Qiu, Guoping. "Image and feature co-clustering." In Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, vol. 4, pp. 991-994. IEEE, 2004.
[5] Frey, Brendan J., and Delbert Dueck. "Clustering by passing messages between data points." science 315, no. 5814 (2007): 972-976.
[6] Iyengar, Giridharan, and Andrew B. Lippman. "Distributional clustering for efficient content-based retrieval of images and video." In Image Processing, 2000. Proceedings. 2000 International Conference on, vol. 1, pp. 81-84. IEEE, 2000.
[7] Murthy, V. S. V. S., E. Vamsidhar, JNVR Swarup Kumar, and P. Sankara Rao. "Content based image retrieval using Hierarchical and K-means clustering techniques." International Journal of Engineering Science and Technology 2, no. 3 (2010): 209-212.
[8] Wang, Xianwang, and Tong Zhang. "Clothes search in consumer photos via color matching and attribute learning." In Proceedings of the 19th ACM international conference on Multimedia, pp. 1353-1356. ACM, 2011.

[9] Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-9. 2015.

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In CVPR09, 2009.

[11] Liu, Si, Zheng Song, Guangcan Liu, Changsheng Xu, Hanqing Lu, and Shuicheng Yan. "Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set." In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 3330-3337. IEEE, 2012.

[12] Hara, Kota, Vignesh Jagadeesh, and Robinson Piramuthu. "Fashion apparel detection: the role of deep convolutional neural network and pose-dependent priors." In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1-9. IEEE, 2016.

[13] Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. "Rethinking the inception architecture for computer vision." arXiv preprint arXiv:1512.00567 (2015).

[14] Myntra: Online Shopping Site URL: www.myntra.com [Last Accessed: Sep 2016]

[15] Revolutions: Online blog on data analysis URL: www.blog.revolutionanalytics.com/2016/08/deep-learning-part-2.html [Last Accessed: Sep 2016]

[16] Stanford Opencoursewere URL: www.cs231n.github.io/transfer-learning/ [Last Accessed: Sep 2016]