Course Project Report (Jul-Dec 2020)

Sketch-Based Image Retrieval By Relevance Feedback and Feature Adaptation using Convolutional Neural Networks

Submitted By

Niteesh Kumar (171IT228)
Rashad Ahmed (171IT132)
Venkatesh B.H (171IT248)

as part of the requirements of the course

Information Retrieval (IT458)

under the guidance of

Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

undergone at

DEPARTMENT OF INFORMATION TECHNOLOGY

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, SURATHKAL

NOVEMBER 2020

Department of Information Technology
National Institute of Technology Karnataka, Surathkal

C E R T I F I C A T E

This is to certify that the Course project Work Report entitled **"Sketch-Based Image Retrieval By Relevance Feedback and Feature Adaptation using Convolutional Neural Networks"** is submitted by the group mentioned below -

Details of Project Group

Name of the Student	Register No.	Signature with Date
Niteesh Kumar	171IT228	
Rashad Ahmed	171IT132	
Venkatesh B.H	171IT248	

as the record of the work carried out by them as part of the course **Information Retrieval (IT458)** during the semester **Jul - Dec 2020**. It is accepted as the Course Project Report submission in the partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Information Technology.**

(Name and Signature of Course Instructor)
Dr. Sowmya Kamath S

D E C L A R A T I O N

We hereby declare that the project work report entitled **"Sketch-Based Image Retrieval By Relevance Feedback and Feature Adaptation using Convolutional Neural Networks"** submitted by us for the course **Information Retrieval (IT458)** during the semester **July - Dec 2020**, as part of the partial course requirements for the award of the degree of Bachelor of Technology in Information Technology at NITK Surathkal is our original work. We declare that the project has not formed the basis for the award of any degree, associateship, fellowship or any other similar titles elsewhere.

Details of Project Group

Name of the Student	Register No.	Signature with Date
1. Niteesh Kumar	171IT228	
2. Rashad Ahmed	171IT132	
3. Venkatesh B.H	171IT248	

Place: NITK, Surathkal
Date:

Sketch-Based Image Retrieval By Relevance Feedback and Feature Adaptation using Convolutional Neural Networks

Niteesh Kumar[1] and Rashad Ahmed[1] and Venkatesh B.H[1]

Abstract— **Sketch-based Image Retrieval (SBIR) is an approach where natural images are retrieved according to the given input sketch query. The main challenge involved in the implementation of such a system in the absence of semantic information in the sketch query. In our approach we tackle this issue by using image preproceesing of the natural image database through Canny-Edge detection technique. The intermediate phase involves the extraction of feature vectors using Convolutional Neural Network (CNN) models such as the VGG-16, InceptionV3 and the ResNet50 network architectures. Each of these models makes use of a certain number of convolutional layers independent of the type of input. Cosine similarity and Euclidean distance measures are adopted to generate the rank list of candidate natural images. A relevance feedback using Rocchio's method is used to adapt the query of sketch images and feature weights according to relevant images and non-relevant images.**

Keywords: Sketch-based Image Retrieval, VGG16, InceptionV3, ResNet50, Relevance Feedback, Mean Average Precision

I. INTRODUCTION

In present day scenario the amount of image data being circulated across the internet has skyrocketed. With the increase in optical technology, the quality of image content is at its peak when comparing with the previous decade. Along with traditional text-based information retrieval, image retrieval has garnered quite some attention due to its influence in domains such as medical healthcare diagnosis, spatio-temporal image sensing, e-commerce and educational research. In traditional context-based image retrieval (CBIR) techniques there is a need for textual information to be needed which introduces added difficulty in information retrieval. Thus the concept of Sketch-based image retrieval (SBIR) which requires free-hand sketches, specific to the particular candidate has amounted to ubiquitous scrutiny. Since the sketch pattern is completely dependent on the users hand motion and imaginative capabilities, sketch-based image retrieval becomes a very challenging method requiring high-level feature extraction and similarity comparison. Such a system of image retrieval that doesn't require any textual information of the same with only a vague description of the subject in the form of a sketch is quite impressive and has high importance in the present day of digitization. This also counter attacks the fact that the user need not have an exact image of the subject in hand. Direct application of CBIR approach to solve SBIR problem isn't preferred due to the fact that the sketches used in SBIR lack the presence of semantic information like color, contract and brightness.

The most basic approach to SBIR includes the use of computer vision libraries like Oreinted FAST and Rotated BRIEF (ORB) detectors or Histogram of Oriented Gradients (HoG) feature descriptor that extracted low level features from the images. These features are then compared to check similarity between the images in the database and the sketch query. By making use of a certain threshold, the final rank list of similar images is retrieved. Most recent approaches deal with the extraction of high-level features using convolutional and deep neural network models which give a better semantic comparison between the images and the sketch. A better view on various researches done on SBIR in recent years is briefed in the next section.

II. RELATED WORK

In this section we'll look into some of the techniques adopted in order to tackle the challenge posed by SBIR. (Devis et al., 2019) makes use of the VGG-19 CNN model as a means to extract features from the images. The implementation involves the interclass classification of the sketch image using the Sketch database (Sangkloy et al., 2016) that consists on approximately 75000 sketch-based images spanned over 125 class labels. Once the label of the sketch query is retrieved, corresponding intra-class checking is done on the image dataset such that final rank list of similar images is retrieved. The main drawback noticed in this work is the absence of a processing phase of the real life images since the feature map of a sketch image and a real image will differ drastically even though they point to a similar subject of reference. The model performed below par on the TU Berlin dataset (Eitz et al., 2012).

(Qi et al., 2019) is another such work done in SBIR that made use of a CNN model for feature extraction. Here a VGG-16 model was used for transfer learning pre-trained on the ImageNet dataset (Deng et al., 2009). The natural images were preprocessed to extract contour images of the same following which feature extraction was done on both the contour database and the sketch query. Cosine similarity was implemented to take note of the similarity measure between the features extracted previously. Flickr15K dataset (Schifanella et al., 2015) was used comprising of approximately 15000 images of both natural and sketch domain spanned over 33 categories. Finally user feedback is used to refine the final rank list of retrieved candidate images. It was observed that with the inclusion of relevance feedback, the performance of the model improved drastically compared to that of the general implementation but the only drawback is the ability to capture user feedback since it requires manual intervention from the user after initial retrieval.

Zero-shot learning (ZSL) is a setup where the developer encounters unique samples during testing that the respective model wasn't trained upon. (Dutta et al., 2020) makes use of ZSL by combining it with SBIR to propose a novel implementation of Zero-shot Sketch-based image retrieval (ZS-SBIR). In the process, two modules where introduced namely content-style fusion module and content-style decomposition module which outputs fake image features. A fine-tuned pre-trained VGG-16 model was adopted for training on the Sketchy dataset. Research was further carried out to improve the performance using a generalized approach with the addition of cross-domain prototype computation. A 7% increase in average precision was noticed following the addition of the generalized approach. The only drawback being an increase in retrieval time by approximately 30 milliseconds.

The retrieval of aerial images through hand-drawn free sketch has been successfully conquered in (Jiang et al., 2017). The work deals with the extraction of cross-domain feature representation using multi-scale network architecture. The base CNN model implemented in the network architecture is the AlexNet (Krizhevsky et al., 2012) system. The dataset acted upon in the research consists of natural and sketch-based images spanned over a definite set of pre-defined categories that were used to train the AlexNet network to build a preliminary model. Euclidean distance is used as a similarity measure to build the final rank list of candidate image retrievals. The use of multi-scale network architecture has increased the accuracy performance by a factor of 2%. The main downside is the requirement of labelled dataset which is quite a daunting task in SBIR.

After going through the various approached to SBIR, our method begins with a image pre-processing of the natural image database using Canny-Edge detection to get a fine contour representation of the same. The high-level feature extraction phase deals with the implementation of 3 standard CNN models namely VGG-16, InceptionV3 (Szegedy et al., 2015) and ResNet50 (He et al., 2015). The similarity measure from the feature vectors extracted is calculated using both the cosine similarity and Euclidean distance methods. Finally a relevance feedback technique using Rocchio method is applied to fine tune the rank list before calculating the different performance metrics.

The reminder of the report is partitioned as follows, Section 3 gives an insight into the proposed methodology along with the detailed system architecture which is then followed by the specification of the implementation environment in Section 4. The experimental results and analysis of the same is scripted in Section 5. Finally the conclusion part of the report is narrated in the final section.

III. METHODOLOGY

A. Image Preprocessing

Hand-drawn sketches and natural images have many differences, therefore, contour features information should be extracted using image preprocessing techniques. The aim

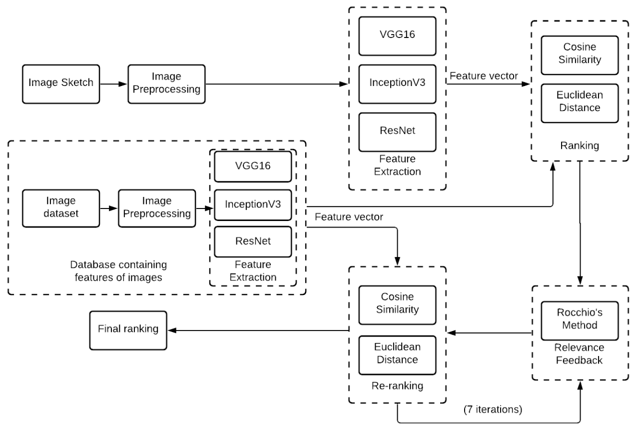

Fig. 1: System Architecture

is to achieve the unity of the real picture and the hand-drawn image on the image domain and adapting to deep-neural network techniques. Canny edge detection is used to extract the useful structural property from an image . The resultant image is processed to binary with a threshold of 127 and a final image of 0-1 is obtained. Similarly, hand-drawn images are prepossessed to remove noise and to retain useful structural property. The final binary image is fed as an input to CNN models for feature extraction.

Fig. 2: Edge Detection

B. Feature Extraction

Features are extracted using different pre-trained standard CNN models from a binary image obtained after image preprocessing. We have used VGG16, InceptionV3 and ResNet50 CNN models to extract features. 4096 , 8x8x2048 and 7x7x2048 features were extracted from each image using VGG16, InceptionV3 and ResNet50 models respectively. Features extracted from natural images are stored in the database as a vector, which are used to compute similarity score with respect to features of a hand-drawn sketch images provided by a user.

C. Image Similarity Matching

Closeness of the two vectors - a vector representing a natural image and another vector representing a sketch, is measured using similarity and distance. We have used cosine similarity and euclidean distance to measure the closeness of a sketch with respect to the images present in the database. Cosine similarity between a query sketch q and an image d_i is computed using the formula 1. Higher cosine similarity value indicates that the sketch is highly close to the image, hence this image will be ranked top for the user output. Formula 2 is used to compute the closeness between q and d_i using euclidean distance with the help of features. Lower euclidean distance between the sketch and the image indicates the closeness between the two features is high. Therefore, if euclidean distance is low, the corresponding image is ranked at top among the retrieved output images.

$$Cos - sim(q, d_i) = \frac{\sum_{j=1}^{t} f_{q,j} f_{i,j}}{\sqrt{f_{q,j}^2}\sqrt{f_{i,j}^2}} \quad (1)$$

$$d(q, d_i) = \sqrt{\sum_{j=1}^{t} (f_{q,j} - f_{i,j})^2} \quad (2)$$

D. Relevance Feedback

Query provided by user often requires to be reformulated to obtain the results of user's interest. This process is commonly referred as feedback cycle. We have used relevance feedback to perform re-ranking of the retrieved images. Relevance feedback gathers information from retrieved images and uses the information to perform re-ranking. Basically, there exists two types of feedback - explicit feedback and blind feedback. Explicit feedback requires user's feedback on the retrieved images and uses this feedback to re-rank the images, whereas, blind feedback doesn't require any user feedback.

(Portenier et al., 2017) uses explicit feedback for re-ranking. First, the model extracts low-level features and deep features from the initial results. Next, the authors have employed a clustering algorithm to cluster them with respect to features. Images from the high-score clusters are re-ranked against of those low-score clusters. The model proposed by (Matsui et al., 2017) uses user's selection of initial result or a modified query as a new query for re-ranking the retrieved results.

We have used blind feedback for re-ranking the retrieved images using Rocchio's method. Rocchio's method using formula 3, re-weights the features and also add new features to modified sktech found from initial retrieved images.

$$Q_1 = Q_0 + \frac{\beta}{n_r}\sum_{i=1}^{n_r} R_i - \frac{\gamma}{n_{nr}}\sum_{i=1}^{n_{nr}} S_i \quad (3)$$

In the formula 3, Q_0 is the feature vector of initial sketch, R_i is the feature vector for the relevant images i, S_i is the feature vector for the non-relevant images i, n_r is the number of relevant images chosen, n_{nr} is the number of non-relevant images chosen, and β and γ are for tuning the importance of relevant and non-relevant features. In this project we have selected 0.75 and 0.25 as the value for β and γ respectively.

IV. Implementation Specifics

This project was implemented using google colab. We have used TPU and 35.35GB RAM provided by google colab to extract the features from the images using pre-trained CNN models.

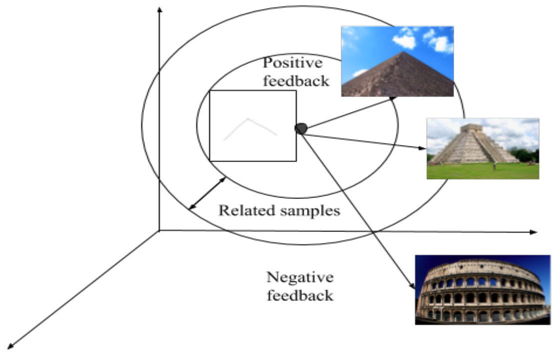

Fig. 3: Relevance feedback with positive and negative feedback for give sketch image.

V. Experimental Results and Analysis

A. System Implementation

To extract feature vectors for SBIR from the pre-trained model, we create a sketch-category mark mapping relationship based on the available dataset with the help of canny edge detection. In accordance with the image sketch input and the label information, the pre-trained models are used to extract features. The features obtained from the models are saved in the database with a mapping with original colour image. When the user provides a sketch query, the sketch is pre-processed to remove noise and feature is extracted using the models. The features are represented in a vector which is compared with the available features in the database, i.e., similarity score is calculated between features of the sketch and the images in the database. The colour images with highest similarity score is considered as a candidate result set and returned as output against the user's sketch query.

Relevance feedback is used after producing the output by the system for the sketch query based on relevant images and non-relevant natural images to optimise the results. Depending on the user's natural image relevance feedback and the input hand-drawn sketch images, a mapping relationship is created, then the new feature vectors of the sketch image information is stored in the system. To change the weights associtated with sketch's features, Rocchio's method is used. After the automatically created re-weighted terms for the sketch image, new terms are applied to the sketch images(found from the relevant natural images).

B. Image Dataset

This project is based on Yahoo's Flickr15K public dataset, which is an important image sharing platform. The real/natural images in the dataset are from the websites and in the SBIR area, this dataset is considered as a benchmark dataset. There are 33 types of details in the data collection, each containing 10 hand-drawn sketch images drawn manually and which is total with 330 sketch image dataset. Many of the traditional colour images are natural landscape images, which are more dynamic than ordinary object images in representing shape characteristics. A total of 14,501 images are categorised scientifically into 33 groups.

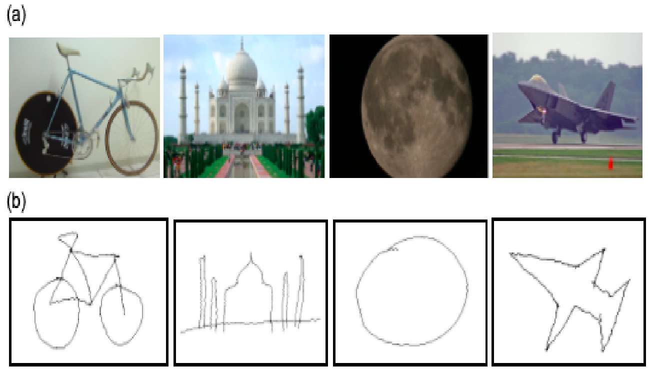

Fig. 4: Sample images from the dataset (a) Flickr15k normal images, (b) Hand-drawn sketch images.

The Sample images are shown in Fig.4. This sample contains normal images and sketch images related i.e., normal and sketch images are bicycle, Taj Mahal, moon and airoplane.

C. Experiments

TABLE I: Comparison of MAP

Methods/Model	Similarity measure	MAP
VGG16	Euclidean distance	0.3199
VGG16	Cosine similarity	0.5615
ResNet50	Euclidean distance	0.5765
ResNet50	Cosine similarity	0.6535
InceptionV3	Euclidean distance	0.5865
InceptionV3	**Cosine similarity**	**0.6739**
Personalized SBIR by CNN	Euclidean distance	0.6449

Experimental results are shown in the table 1. The results obtained for this dataset demonstrate that the InceptioinV3 model for feature vector extraction and Cosine similarity measure with relevance feedback using Rocchio's method approach is clearly better than the other methods i.e., the MAP(Mean Precision Average) for this method is 0.6739. Again the approach of Rocchio strengthened more which shows that compared to Euclidean distance similarity, the Cosine similarity is capable of producing better MAP performance. The other comparison which we have is VGG16 and ResNet50 CNN model. With VGG16 and euclidean distance similarity measure the MAP value which we got is 0.3199 which is low compared to other methods and with VGG16 cosine similarity MAP value is 0.5615. For ResNet50 CNN model with euclidean distance similarity the MAP value is 0.5765 and with cosine similarity the MAP value is 0.6535. The InceptionV3 outperformed better MAP value with cosine similarity i.e., MAP value we got is 0.6739 and with euclidean distance similarity the MAP value is 0.5865 which is less than cosine similarity measure. This all experiments done with relevance feedback using Rocchio's method and the number of iterations we considered is 7. Without relevance feedback the MAP value which we got for VGG16,ResNet and InceptionV3 are 0.9604 , 0.1324 and 0.9475 respectively and this results are very less compared to with feedback applied.

TABLE II: Comparison of MAP with 5 iterations of relevance feedback using Euclidean distance similarity

Model+Euclidean	P3	P4	P5	P6	P7
VGG16	0.4025	0.3765	0.3523	0.3199	0.3199
ResNet50	0.5592	0.5767	0.5794	0.5768	0.57266
InceptionV3	0.5474	0.5612	0.5906	0.5825	0.5865

The results on different 5 iterations for relevance feedback using Rocchio's method is shown in the table II. The results are obtained using Euclidean distance similarity based ranking. Here the iterations considered are 3,4,5,6,7. The results of MAP value for this similarity measure is not better as cosine similarity based ranking as shown in table III and the VGG16 MAP value decreased as iteration increased using Euclidean distance similarity based ranking.Corresponding Graph is shown in Fig 5.

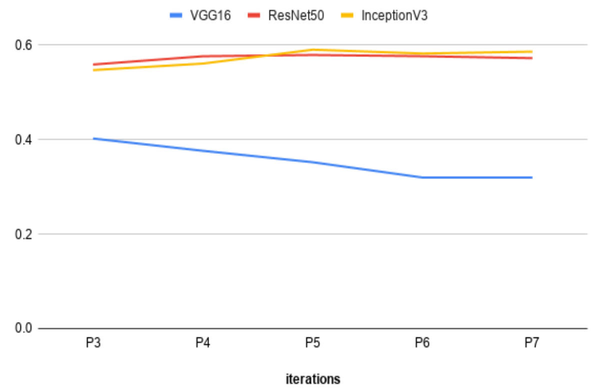

Fig. 5: Results on 5 iterations of relevance feedback using Euclidean distance similarity

TABLE III: Comparison of MAP with 5 iterations of relevance feedback using Cosine similarity

Model+Cosine	P3	P4	P5	P6	P7
VGG16	0.4964	0.5284	0.5452	0.5615	0.5615
ResNet50	0.5763	0.614	0.6339	0.6456	0.6535
InceptionV3	0.5594	0.623	0.6415	0.6529	0.6739

The results on different 5 iterations for relevance feedback using Rocchio's method is shown in the table III. The results are obtained using cosine similarity based ranking. Here the iterations considered are 3,4,5,6,7. 7th iteration outperformed well with InceptionV3 as shown in table III. Compared to Euclidean distance similarity based ranking the cosine similarity ranking showed positive results. Corresponding Graph is shown in Fig 6. The best MAP value which we got is 0.6739 for InceptionV3 with Cosine similarity based ranking for 7th iteration as shown in table III.

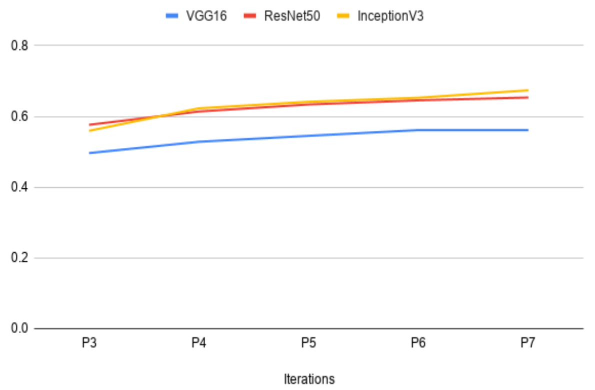

Fig. 6: Results on 5 iterations of relevance feedback using Cosine similarity

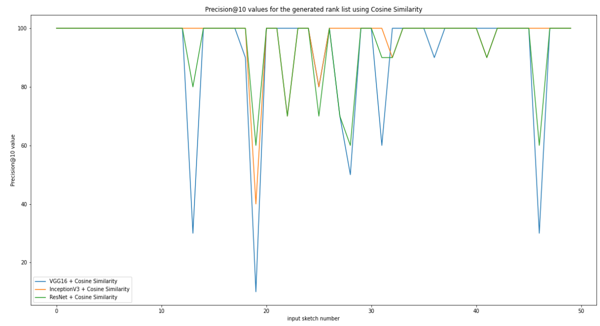

Fig. 7: Precision@10 using Cosine similarity

Graphs 12 and 8 shows that InceptionV3 outperforms other two model in SBIR system. For most of the sketches, InceptionV3 and ResNet has almost same Precision@10 values whereas VGG16 has performed least with respect to Precision@10 values. Similarly, graphs 9 and 10 shows the recall@10 values for all the three models used. We can observe that InceptionV3 outperforms other two models whereas, VGG16 performed least with respect to recall@10 values. Hence, InceptionV3 is good model for feature extraction which can be used in SBIR systems.

VI. Discussion

After effective implementation of the proposed SBIR system and evaluation of the same using various metrics, the further discussion of the same shall be elaborated in this section. Starting with the brief flow of the proposed system involving canny edge preprocessing of the natural image database to extract contour features followed by the use if pre-trained CNN models to update high-level features rather than low-level attributes. Finally the inclusion of Rocchio's method to introduce the technique of relevance feedback so as to power the performance of the system. Evaluation techniques includes Precision@K (P@K) which is the proportion of retrieved top-K images that are relevant and Recall@K (R@K) which is the proportion of relevant

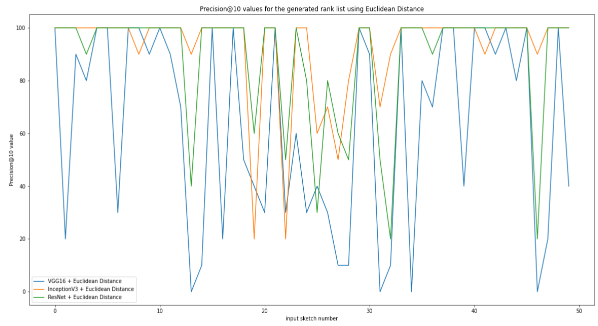

Fig. 8: Precision@10 using Euclidean Distance

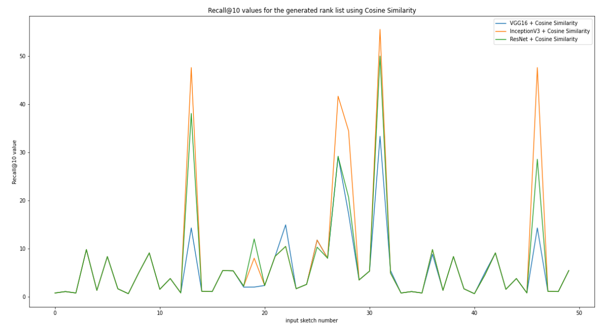

Fig. 9: Recall@10 using Cosine Similarity

images that are retrieved in the top-K. Visualization of P/R@K for a large database is quite daunting, hence the use of average-based metrics, that is, Average Precision (AP). Both cosine similarity and Euclidean distance were used as similarity measures. Figure 7 and 8 represent the plot for P@10 using cosine similarity and Euclidean distance respectively whereas Figure 9 and 10 represents the same for R@10.

AP roughly equals to the average area under the Precision-Recall curve plotted for a set of sketch queries. Unlike P@K and R@K, AP doesn't require a manual K value while

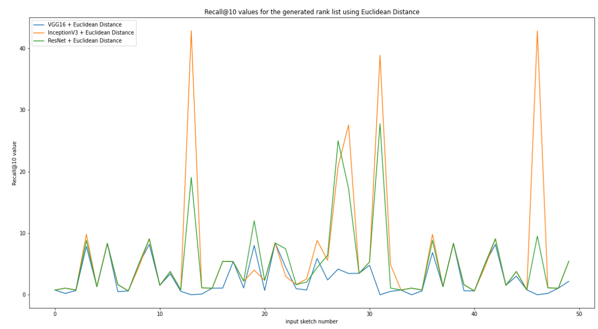

Fig. 10: Recall@10 using Euclidean Distance

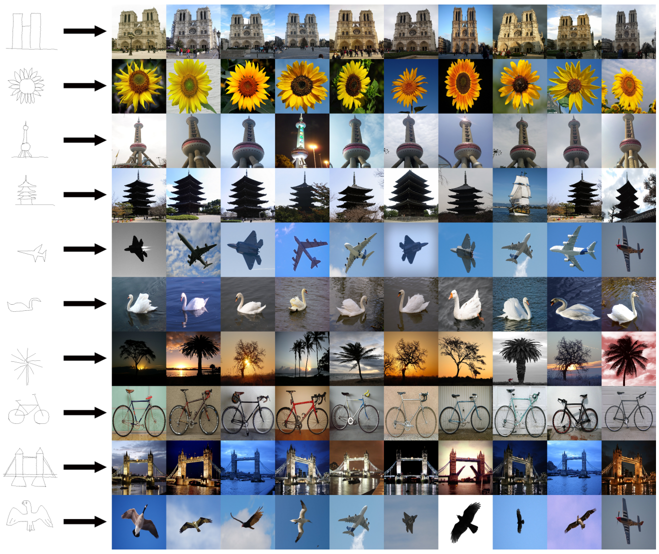

Fig. 11: Sample Result of Proposed SBIR

accounting for both precision and recall. The ranking mishits at the top of the rank list are more influential and those at the bottom are still accounted equally. Mean Average Precision (MAP) is the mean of the average precision for each sketch query. Table 1, 2 and 3 illustrates the MAP scores calculated for the respective combinations. InceptionV3 model with cosine similarity performed the best among all combinations. It is observed that the re-weighting of the sketch feature in successive iterations of the Rocchio method improved performance. This is due to the fact that the method gives weight to both the positive and negative sample based on the constants β and γ.

The comparison of contour natural image features with that of the input sketch produced some interesting observations like the category mapping between them. Since due to the lack of relevant semantic information in the sketch images. This also adds to the presence of semantic information in the natural images which is not used. The AP values for categories of images having multiple labels seem to produce lower performance. This is a challenge that requires characteristic information of the particular user as the hand drawn images depend on the imaginary and is specific to that individual. The MAP scores in table 2 and 3 follows the above discussion. The gradual increase of precision from 55.94% to 67.39 over 5 iterations proves the superiority of InceptionV3 model over the other 2. This result was quite expected due to the large feature size of the model but the only drawback being the extra computational time consumed by the same.

VII. CONCLUSIONS AND FUTURE WORK

Introducing the idea of feature adaptation, a novel sketch-based image retrieval system has been implemented. We dynamically increase the importance of relevant images i.e., adapt the query parameters and feature weights and also the collection of relevant images and non-relevant images during the retrieval process. In accordance with conventional methods, experimental findings indicate a major increase in the precision of retrieval. In comparison to the previous works done in the domain of SBIR, the proposed system performs comparatively better in categorizing the natural images based on the weighted features of the sketch query. The method can be further improved with a deep study in optimal feature substitution rules in order to achieve better adaptation.

Optimal clustering can be used to reduce computational time by taking into consideration only that part of the database which overlaps with the cluster headed by the corresponding sketch query. Finally for better user experience, an application can be built to expand the horizon of imagination to its fullest.

VIII. INDIVIDUAL CONTRIBUTION

Name	Roll No.	Contribution
Rashad Ahmed	171IT132	Detailed literature survey of 4 papers, collection of datasets, Euclidean distance similarity implementation.
Niteesh Kumar	171IT228	Detailed literature survey of 2 papers, Image preprocess, MAP results implementation.
Venkatesh B H	171IT248	Feature extraction using CNN model, Cosine similarity implementation, Rocchio's method implementation.

Fig. 12: Individual contribution

REFERENCES

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

Devis, N., Pattara, N. J., Shoni, S., Mathew, S., and Kumar, V. A. (2019). Sketch based image retrieval using transfer learning. In *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, pages 642–646.

Dutta, T., Singh, A., and Biswas, S. (2020). Styleguide: Zero-shot sketch-based image retrieval using style-guided image generation. *IEEE Transactions on Multimedia*, pages 1–1.

Eitz, M., Hays, J., and Alexa, M. (2012). How do humans sketch objects? *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(4):44:1–44:10.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition.

Jiang, T., Xia, G., and Lu, Q. (2017). Sketch-based aerial image retrieval. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3690–3694.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, page 1097–1105, Red Hook, NY, USA. Curran Associates Inc.

Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., and Aizawa, K. (2017). Sketch-based manga retrieval using manga109 dataset. In *Multimedia Tools and Applications*, volume 76, page 21811–21838.

Portenier, T., Hu, Q., Favaro, P., and Zwicker, M. (2017). Smartsketcher: Sketchbased image retrieval with dynamic semantic re-ranking. In *Proc. Symp. Sketch Based Interfaces Model*.

Qi, Q., Huo, Q., Wang, J., Sun, H., Cao, Y., and Liao, J. (2019). Personalized sketch-based image retrieval by convolutional neural network and deep transfer learning. *IEEE Access*, 7:16537–16549.

Sangkloy, P., Burnell, N., Ham, C., and Hays, J. (2016). The sketchy database: Learning to retrieve badly drawn bunnies. *ACM Trans. Graph.*, 35(4).

Schifanella, R., Redi, M., and Aiello, L. M. (2015). An image is worth more than a thousand favorites: Surfacing the hidden beauty of flickr pictures. In *ICWSM'15: Proceedings of the 9th AAAI International Conference on Weblogs and Social Media*. AAAI.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2015). Rethinking the inception architecture for computer vision.