International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 880-885 www.ijaem.net ISSN: 2395-5252

Sketch Based Image Retrieval

Somashekhar B M[1] , Nishath Kousar[2] , Sana Arshad[3] , Surabhi K C[4] , Rashmitha V R[5]

[1] *Asst.Professor, Dept of ISE, Maharaja Institute of Technology Mysore, India.*
[2,3,4,5] *Engineering student, Dept of ISE, MITM*

Date of Submission: 22-06-2020 Date of Acceptance: 11-07-2020

ABSTRACT: Sketch based image retrieval (SBIR) has received a lot of attention in the recent years. In this paper we propose an efficient SBIR system. The approach that has been employed here in SBIR system focuses on the sketches of the objects. Here, the user interface asks for a sketch image as an input from the user, on which the search has to be performed. Sketch based image retrieval is an approach where the system fetches images from the trained database, based on the query sketch received from the user through an interface. The images have embeddings which have to be extracted in order to build the data before processing further. We are making use of linear support vector machine (SVM). SVM is an algorithm that provides analysis of data for classification. We are using RESNET-50 which is a CNN (convolutional neural network) model which is constructed using pre-trained images which is taken from the pickle model. A pre-trained VGG network is used as a input model and it finds the similarity between the sketch and image database. Tkinter is a graphical user interface used. Basically when we upload the image it will load to trained dataset first and then it will process it with the original images. With the help of pickle model it picks all the similar images present in the database. The system displays at most nine images as output, matching the input sketch.

KEYWORDS: Sketch based image retrieval, RESNET-50, Pre-trained VGG16, CNN

I. INTRODUCTION

In our daily lives, we come across a lot of situations where we are looking for some product and we are not familiar with the product name. so, we just have picture of a product in our mind and it would be a boon if we actually could draw down the product image and we could get the product name or entire product. This is exactly what our system does when we give hand drawn sketch image as an input it will retrieve multiple similar images corresponding to the sketch.

In particular, drawing can help users to successfully express or deliver their own intentions regardless of their language and can be adopted by medical science to observe the behavior of autistic children. It can also be used by deaf and dumb people to communicate with other people. For example the sketch of a banana is a banana for all over the world. Thus, banana could be an effective query for a system. Likewise, we can give any query which we use day to day lives. For retrieval of a system it is mainly based on how well the user draws the sketch. Furthermore, the variations of user drawing are considerable which can affect the performance of the retrieval system. This work aims at assisting a user to draw a sketch query image easily and system will predict the user's drawing intension and provide all the set of relevant images from a database. In the past years, a rapid development and improvement has been gained in the field of artificial neural network. Neural network is gaining much more attention in recent years. convolutional neural network (CNN) is making a big success in large image processing based applications. Many of the researchers have used neural networks and obtain a great success in their field. VGG [1], GoogLeNet [2] and ResNet [3] are excellent CNN models in the past few years and these models achieved a magnificent result in image processing based applications, such as image recognition and classification tasks, object detection etc. Inspired by this we have come up with sketch based image retrieval system where it uses a support vector machine (SVM) for classification and pre-trained VGG16 network which is a input model and finds the similarity between sketch and image database and retrieve using R-CNN(region based convolutional neural network).

II. PROBLEM STATEMENT

❖ When a person travels from one place to another place, it is very essential to know the official language of that place in order to communicate. If a person is not familiar with

official language of that place, then it leads to a communication problem.

- A deaf and dumb person faces a many problems while expressing their needs with others. The person who has skills of understanding their language can understand their needs and feelings. But for an ordinary people it is very difficult to understand their feelings and needs.
- Nowadays online shopping sites become more popular, since we can buy anything from anywhere and anytime, and it is easy to search for what we want if we are familiar with the product name. If a person is not familiar with a product name but he has some image of that product on his mind, then it very difficult to search for product what he wants to buy.
- It is a difficult task to analyze the feelings and needs of a child suffering from autism spectrum disorder because of the trouble they face with social interaction, limited communication and their repetitive behavior. So there is a need of a learning tool which can be used to study their behavior better wherein they would draw things down and the system recognizes what they are asking for.

III. EXISTING SYSTEM

Sketch based image retrieval was implemented using transfer learning approach [4]. in order to learn the features of the sketch. VGG19 a pre-trained network was used instead of building a CNN from scratch. The pre-trained network was capable of processing both input sketch and related photos and based on cosine similarity function it checked the similarity between the sketch image and colored image further based on similarity rate all colored images was retrieved. An efficient fruit and vegetables classification system was proposed using image saliency [5] to draw the object regions and convolutional neural network (CNN) model to extract image features. Image saliency is utilized to select main saliency regions according to saliency map. A VGG model was chosen in order to train fruit and vegetables. An intelligent image retrieval system [6] based on a method called database revision (DR) was proposed. In this system for retrieval of image from the database, image feature extraction was used in order to extract the features like shape, color and texture. The database gets rewritten by the result of feature similarity comparison between the database images and the query image. The intelligence of system lies in the fact that the database is continuously revised during each search. Furthermore, Google lens fails to identify the sketch. When you upload hand drawn sketch image, it just terms it as sketch rather than trying to figure out what the sketch really is. Suppose, if u draw the sketch of car the Google lens outputs as sketch and not car. And it displays all irrelevant sketches which are not even close to the hand drawn sketch. Sketch identification system only considers the outline of the sketch and doesn't not consider the inner features resulting in incorrect output. It gives a lot of garbage images which also takes a lot of time to process the output.

IV. PROPOSED SYSTEM

All the existing system are independent but to improvise the search efficiency we have come up with a effective way so that the users can access, browse and retrieve a set of relevant images in real time applications.

In the proposed system all the relevant images which match to the hand drawn input sketch are retrieved. At the user interface the user has to upload a hand drawn sketch image as an input. The input image has to be in jpg format. To the input image the product images present in the database which are matching to the input image are displayed as output. This system basically contains two datasets, the trained image dataset and the test image datasets. The trained datasets contains all the product images of varying categories. Whereas the test dataset has the hand drawn sketch images. The hand drawn sketch image which is present in the test dataset is compared and matched with the trained dataset in order to give precise results. The trained images have to be extracted using the data embeddings. before the data is processed it has to be build. The images are extracted and encodings are created for the datasets which intern creates the embedding models .once the trained model is run in the output folder pickle files are created. Here Tkinter is used as a Graphical user interface in which we are using RESNET-50 which is a CNN (convolutional neural network) model which is constructed using pre-trained images which is taken from the pickle model. A pre-trained VGG network is used to find the similarity between the sketch and image database. VGG-16 network architecture is used where in the input model consists of one input layer, four hidden layers and one output layer which is further processed by ReLU network.

V. METHODOLOGY

Before building any application we require a dataset so, the first task is collecting the sketch images and the product images that is collecting the images for train dataset and test dataset. So basically we are having two datasets. One is train dataset and second one is test dataset. In train dataset the features of the images are converted into weights. In

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 880-885 www.ijaem.net ISSN: 2395-5252

test dataset the region of interest are extracted from sketch images using grey scaling.

In our project we are making use of pre-trained model called ResNet-50. In general, ResNet-50 is nothing but a pre-trained deep learning model for classification of convolutional neural network. ResNet-50 has 23 million trainable parameters which mainly useful for image recognition. Compared to other pre-trained model such has Alex Net, Google Net, ResNet has better performance with less error rate on recognition task. Hence ResNet-50 is more commonly used tool for image classification. ResNet-50 is 50 layers deep and which is trained on imageNet database which is having million images of different categories, in our project we make use of different images of different categories like fruits, clock, ball etc.., this network takes the image in two dimensional plane. We have also make use of a pre-trained VGG16 (visual geometry group) network for the input model which contains one input layer, four hidden layers, and one output layer. It consists of twelve convolutional layers, some of which are followed by max pooling layers and four fully connected layers and finally a softmax classifier. The input is 224*224 RGB image basically this image is passed through a stack of covolutional layers max-pooling is performed over a 2*2 pixel window with stride 2 the configuration of network is same as shown in a network architecture.(Fig 1) All the hidden layers are equipped with rectification (ReLU) non-linearity. Finally a softmax layer outputs the class probabilities.

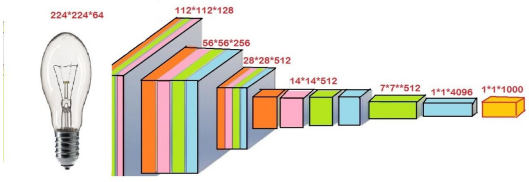


Fig 1: VGG16 network architecture

For classification we have used SVM which is nothing but support vector machine which is a supervised learning model

We are making use of linear SVM. SVM is an algorithm that provides analysis of data for classification. If we carry out plotting in the n-dimensional space. (Fig 2) Value of each feature is also the value of the specific coordinate. Then, we find the ideal hyper plane that differentiates between the two classes. Basic principle behind the working of Support vector machines is it is a hyper plane that separates the dataset into classes.

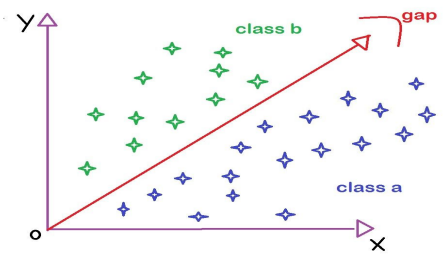


Fig 2: support vector machine (SVM)

For retrieval we are using R-CNN (Region-based Convolutional Neural Networks). The region proposals are fed into CNN so that it can produce feature vector (weights) as output. Here the weights of test sketch database and trained image datasets are compared and based on similarity all the relevant images are retrieved.

VI. IMPLEMENTATION

The hand drawn sketch image is given as an input to the system through a user interface. Tkinter is a graphical user interface which is used here. The input hand drawn image is uploaded to the system which has to be in jpg format. The system basically contains two datasets, the trained dataset and the test dataset. The trained dataset contains all the colored or the product images of varying categories for example apple, grapes, dice, ball, clock, bucket, banana, shirt, etc. In each category we have around ten to twenty varying images with different angles. Each image is supposed to be stored in jpg format. The other dataset is test dataset, where we store all the hand drawn sketch images. Whenever we draw the hand drawn sketch image it needs to be stored in the test dataset folder .Before processing the images the features of the trained images are converted to weights and the region of interest of sketch images in the test datasets are extracted using grayscaling. Basically when a user gives a input sketch as a query to the system, the system tries to extract the features of the sketch image. The features of the sketch could be extracted using some simple an effective methods like grayscaling, threshoulding and then the contour edges are detected and comparision is done in order to find the similarity between the sketch and image database. And then prediction is done. The predicted output is displayed in the user interface. So that a person can easily express his feeling or what he is trying to say. The complete flow is explained using the diagram (fig 3).

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 880-885 www.ijaem.net ISSN: 2395-5252

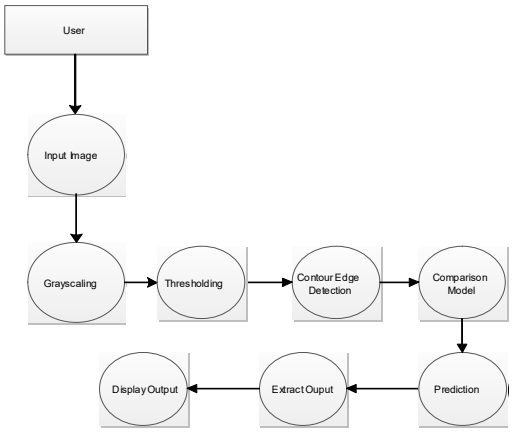


Fig 3: Data flow diagram

A. Module 1: Collecting Sketch image dataset
In this module, we collect the sketch image dataset and load into the model.
B. Module 2: Greyscaling
Grayscale is a range of monochromatic shades from black to white. Many image editing programs allow you to convert a color image to black and white, or grayscale. This process removes all color information, leaving only the luminance of each pixel.
C. Module 3: Thresholding
Image thresholding is a simple, yet effective way of partitioning an image into a foreground and background. This image analysis technique is a type of image segmentation that isolates objects by converting grayscale images into binary images.
D. Module 4: Contour Edge Detection.
Contours detection is a process can be explained simply as a curve joining all the continuous points (along with the boundary), having same color or intensity. This means that you can request outer borders of objects detected in your images.
E. Module 5: Comparison Model.
This model helps us to find the similarity between the sketch query image and colored image. From sketch query image the region of interest features are extracted and compared with the trained dataset i.e. Colored images features.
F. Module 6: Prediction.
In this module, it will extract the output by comparing the original image with the trained images. The display the output to the user.

Basically first we have to create a model before building the data. To create a model we have to build a data before processing which is done by extracting data embeddings and specified as embedding model. This model will extract the image and it will create the encoding for all the dataset we have. Once the model is created after running the trained model in output pickle file is created. Further after running the Tkinter.py which contains the RESNET model i.e. Resnet-50. This will construct the model with pre-trained images. The pre-trained images come from pickle component. Pickle will help us to fetch the information from ReLU network. RESNET50 network takes the image in two dimensional planes where we have used four kernel sizes to process the pickle model. We have used VGG16 for the input model which contains one input layer, four hidden layers, and one output layer. This is shown in below implementation code.

```
def __init__(self, number_classes = 2000, model_path="model.pkl"):
    super(VGG, self).__init__()
    self.conv11 = nn.Conv2d(3,64,kernel_size=3,stride=1,padding=1)
    self.conv12 = nn.Conv2d(64,64,kernel_size=3,stride=1,padding=1)

    self.conv21 = nn.Conv2d(64,128,kernel_size=3,stride=1,padding=1)
    self.conv22 = nn.Conv2d(128,128,kernel_size=3,stride=1,padding=1)

    self.conv31 = nn.Conv2d(128,256,kernel_size=3,stride=1,padding=1)
    self.conv32 = nn.Conv2d(256,256,kernel_size=3,stride=1,padding=1)
    self.conv33 = nn.Conv2d(256,256,kernel_size=3,stride=1,padding=1)

    self.conv41 = nn.Conv2d(256,512,kernel_size=3,stride=1,padding=1)
    self.conv42 = nn.Conv2d(512,512,kernel_size=3,stride=1,padding=1)
    self.conv43 = nn.Conv2d(512,512,kernel_size=3,stride=1,padding=1)

    self.conv51 = nn.Conv2d(512,512,kernel_size=3,stride=1,padding=1)
    self.conv52 = nn.Conv2d(512,512,kernel_size=3,stride=1,padding=1)
    self.conv53 = nn.Conv2d(512,512,kernel_size=3,stride=1,padding=1)

    self.relu = nn.ReLU(inplace = True)
    self.maxpool = nn.MaxPool2d(2, stride = 2)
    self.fc1 = nn.Linear(512*7*7, 1024)
    self.fc2 = nn.Linear(1024, number_classes)
```

Fig 4: configuration of layers.

```
def forward(self, x):
    x = self.conv11(x)
    x = self.relu(x)
    x = self.conv12(x)
    x = self.relu(x)
    x = self.maxpool(x)

    x = self.conv21(x)
    x = self.relu(x)
    x = self.conv22(x)
    x = self.relu(x)
    x = self.maxpool(x)
```

Fig 5: It shows processing of convolutional network. it is also called as ReLU and softmax.

Classification is done using linear SVM. The weights of the trained image and test image are compared and retrieved using R-CNN. Therefore once if we run Tkinter.py run module it displays a user interface so that we can upload the hand drawn sketch image we have provided with buttons whenever we upload the image to our system it will be in test dataset folder once we upload the image it shows a preview of uploaded image and then if we click on load dataset button it will check with

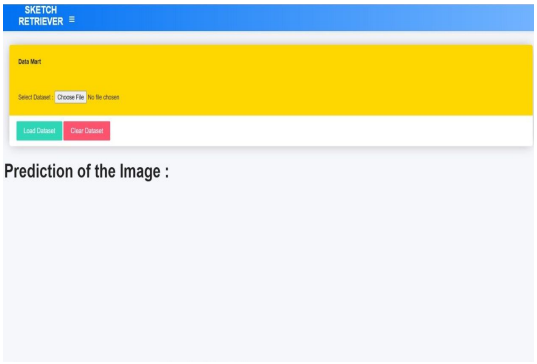
International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 880-885 www.ijaem.net ISSN: 2395-5252

trained dataset and identifies all the relevant image that are available. By identifying the contour edges, similarity pattern and with the help of pickle model it will pick all the relevant products from the database.

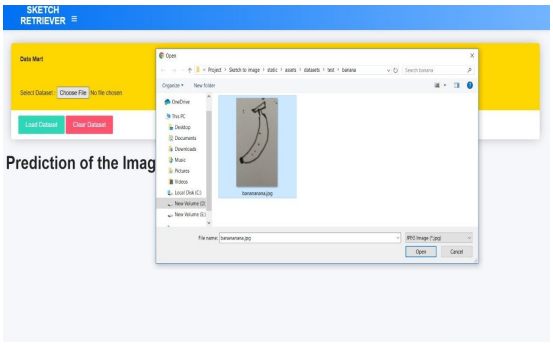
VII. EXPERIMENTAL RESULTS

In our system the user needs to interact with the user interface in which they have to choose the input file that is hand drawn sketch file which is in jpg format. Then the system fetches the colored images from trained database based on the input sketch. Here we have some snapshots where the user has to give input and how the predicted outputs will be displayed.

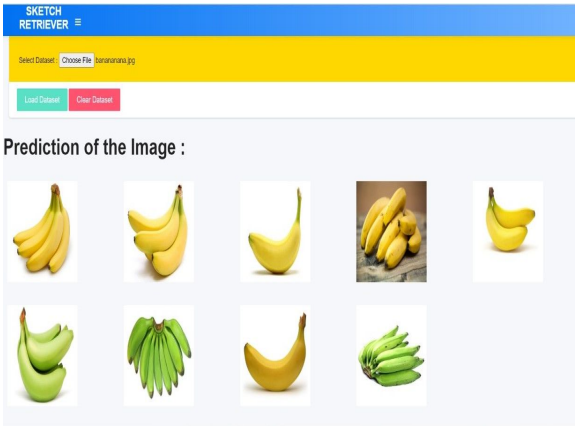
Step 1: A user interface provided to the users to choose the file. Here the user need to choose the file which should be in jpg format (for example we have choose banana.jpg file)



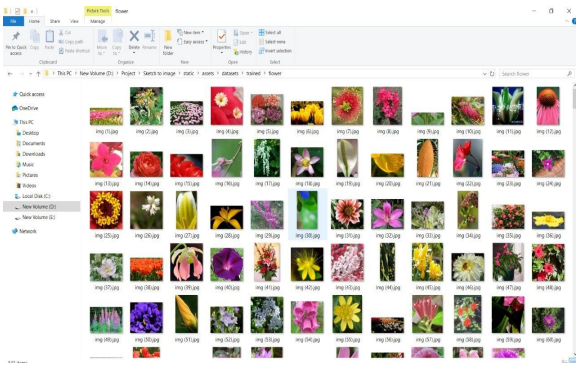
Step 2: in this it will display the file which we have chosen (here it is showing banana.jpg)



Step 3: these are the outputs for the input hand drawn sketch. The maximum limit of the output image is 9.



This is the trained dataset for the flower



VIII. CONCLUSION

Sketch based image retrieval is a challenging image processing task as they are abstract, having less information and also produce variance because the same sketches can be drawn differently by different users. Sketch-based image retrieval has been a field of interest since the early 1990s and has drawn more attention recently. SBIR system can be used to provide continuous day to day applications in different fields, replacing the conventional methods in those fields. For example, preschoolers and autistic kids can be taught in a new and efficient way using a sketch-based image retrieval system. The end user draws a sketch on the interface and similar images will be retrieved. In this system a good user interface is provided to the user to upload a sketch input image and receive related images. After analyzing the given input sketch, the system will check the trained dataset for similar images. Additionally based on the result and application that has been developed, it can be concluded that our sketch based image retrieval is a boon for the digital world which is on the verge of being exploded with new users every day and an

SBIR system will play an important role in fields like e-commerce.

FUTURE ENHANCEMENT

In future, our present system can be improved by using better algorithms so that it can give much more precise results. The storage capacity could also be increased for the scenarios where the database is very huge and the input model can also be change rather than giving the input from the file or document we can give the input without phones. The flexibility and the quality of the input image should not be a problem.

REFERENCES

[1]. Karen Simonyan and Andrew Zisserman, "very deep convolutional networks for large-scale image recognition" ICLR 2015.
[2]. Christian Szegedy et al. "Going deeper with convolutions." Computer vision and pattern recognition (2015)
[3]. Kaiming He et al. "Deep Residual Learning for Image Recognition."computer vision and pattern recognition (2015),
[4]. nikul devis, nirmal joshy pattara, sherin shoni "sketch based image retrieval using transfer learning" 2019 IEEE
[5]. guoxiangzeng "fruit and vegetables classification system using image saliency and convolutional neural network" 2017 IEEE
[6]. S. Sreedevi and S. Sebastian "Content based image retrieval based on Database revision" 2012 International Conference on Machine Vision and Image Processing (MVIP)
[7]. shuwang and zhenjiangmiao "sketch-based image retrieval using sketches tokens" 2017 IEEE
[8]. *yonggang qi yi-zhe song_ honggangzhang† junliu†"* sketch-based image retrieval via Siamese convolutional neural network" 2016 IEEE
[9]. rei narita "sketch-based manga retrieval using deep features" 2017IEEE
[10]. heydim´endez-v´azquez "local deep features for composite face sketch recognition" 2019 IEEE
[11]. sounakdey, anjandutta, suman k. Ghosh, ernestvalveny, "learning cross-modal deep embeddings for multi-object image retrieval using text and sketch" 2018 IEEE
[12]. *tianbi jiang*1, *gui-song xia*1, *qikai lu*2 "sketch-based aerial image retrieval" 2017 IEEE
[13]. sreyasee das bhattacharjee, junsong yuan, yichenghuangy, jingjingmeng, lingyuduany "query adaptive multi-view object instance search and localization using sketches" 2018 IEEE
[14]. S. Polsley, J. Ray and T. Hammond "SketchSeeker : Finding Similar Sketches" IEEE 2017.
[15]. Mathias Eitz, James Hays, and Marc Alexa " How do humans sketch objects?" ACM Trans. July 2012
[16]. Li, Yi & Song, Yi-Zhe & Gong, Shaogang " Sketch Recognition by Ensemble Matching of Structured Features" BMVC 2013
[17]. qi qi 1, (member, IEEE), qiming huo1, jingyu wang1, (member, IEEE), Haifeng sun 1, yufei cao2, and jianxin liao1, (member, IEEE) "personalized sketch-based image retrieval by convolutional neural network and deep transfer learning", 2019 IEEE

International Journal of Advances in Engineering and Management
ISSN: 2395-5252

IJAEM

Volume: 02 Issue: 01 DOI: 10.35629/5252

www.ijaem.net
Email id: ijaem.paper@gmail.com