IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

Similar Fashion Finder using Reverse Image Search

S.R. Katasani[1*], T. Rachepalli[2], N. Kamat[3], M. Jadhav[4]

[1] Department of Computer Science, Ramaiah Institute of Technology, Bangalore, India
[2] Department of Computer Science, Ramaiah Institute of Technology, Bangalore, India
[3] Department of Computer Science, Ramaiah Institute of Technology, Bangalore, India
[4] Department of Computer Science, Ramaiah Institute of Technology, Bangalore, India

[*]*Corresponding Author: katasanisairupa@gmail.com, Tel.: +91-9632149451*

DOI: https://doi.org/10.26438/ijcse/v7i5.11901195 | Available online at: www.ijcseonline.org

Accepted: 23/May/2019, Published: 31/May/2019

***Abstract*—**Articulation of features of fashion into keywords is an arduous task. Description of fashion seen on other people is insufficient, not to mention, inaccurate for a conventional search engine that takes keywords as queries. To overcome this shortcoming, this paper outlines a model for a search engine that takes a fashion image as a query and returns five most similar images from its database. The model consists of CNN classification model that classifies the query image into one of the five classes and a Convolutional Autoencoder that returns five images with most similar features from that class of images. Similarity between images is found by calculating the similarity of the encoded vector of the query image with the encoded vectors of the images in the class predicted by the classifier. Since the encoding and decoding is done by the Autoencoder based on the nature of the images given for training, the model returns images based on features that prove important enough to be encoded based on the training images. In other words, the features that are necessary to ensure as little loss as possible when decoding the encoded vector. This forms the basis for using similarity between encoded vectors to find similar images to the given query image. The model is trained using fashion images to find similar fashion to query image.

***Keywords*—**Classification, Similar images, CNN, Autoencoder, Clothes

I. INTRODUCTION

E-Commerce business has become extremely popular these days and has been growing exponentially. A recent study by Forrester Analytics shows that E-Commerce accounts for 27% of total fashion sales in 2018, compared to 15% of total retail sales in 2018. With the majority of percentage of E-Commerce business being driven by fashion sales, it becomes extremely important for these E-commerce websites to give appropriate and accurate results for the user. Traditional way of shopping online for fashion items is to search for the desired item by describing it in the form of textual words. But often search results obtained by such queries are not very accurate, because describing a fashion item with a specific pattern or a unique design is hard to articulate in words. For this reason, in this paper, we propose a model for a visual search engine which takes the fashion image as a query and returns five most relevant images to the query image from a database of images.

In this paper, we propose a model which uses Convolutional Neural Network (CNN) and Convolutional Autoencoder to give similar images for a given query image. CNN is used to identify the class of the query image and a separate Autoencoder is trained on images of each class to be able to return the five most similar images of clothes from the class, determined by the CNN classifier.

In previous research, CNN or an Autoencoder separately were used for the retrieval of images with similar clothes [1][2][3][4]. Although, that approach gave satisfactory results, this paper proposes to use both a CNN and an Autoencoder because, it is the belief of the authors that, training the autoencoder on only one class of clothing images, makes it possible to detect the similarity in design, cut, colour and other features that otherwise would have gone undetected when trained on images of all classes. This is decided based on the assumption that a query image of the shirt would need only shirts with similar features to be returned as results. This is especially the case when training the model with a small dataset. Hence, separation of the classifier from the autoencoder, allows the autoencoder to focus on similarity of features of clothes rather than just being able to categorize them into one of the five classes of clothing images.

Rest of the paper is organized as follows, Section I contains the introduction of the need for Reverse Image Search and the model proposed in the paper, Section II contain the related work of the model proposed, Section III contains the methodology used to build the model for similar clothing images retrieval and all the components involved in the model , Section IV contains the results and discussion with the query images and their results, graphs showing the training and validation loss for the five classes of images and a table containing the least training and validation losses for the five classes, section V concludes the paper and describes the future scope.

II. RELATED WORK

Lad Ami D, Mori Shilpa I, Raulji Urvashi K, Shaikh Saliha H [5] proposed a content-based image retrieval system for retrieving the similar images from the database with respect to query image by considering certain features such as statistics, human pose estimation, cloth parsing, texture, colour, shape and pattern recognition can be considered to retrieve the relevant images.

M. M. Singh, C. David, P. M. Rathod [6], addressed an efficient technique for locating and finding the relevant information from the image using colour and shape-based image segmentation in their paper. In the proposed CBIR system, an effective and efficient retrieval methods are based on colour (spectral) information with spatial (position/distribution) information followed by shape-based image segmentation. The results are computed qualitatively (visually) as well as quantitatively using quality measures.

Kaur M, Sohi N [7] proposed a content-based image retrieval system using three steps, namely, feature extraction, feature matching and retrieval system design.

S. Ota, H. Takenouchi and M. Tokumaru [1], proposed that Reverse Image Search systems are an improvement of the Content Based Image Retrieval (CBIR) Systems. They propose using features that are extracted from a deep neural network (DNN). The DNN is optimized by a stacked denoising autoencoder (SdA). The accuracy of the proposed method was superior to the previous method by 30.43%.

Z.-Q. Cheng, X. Wu, Y. Liu, and X.-S. Hua [2] have proposed a novel deep neural network called AsymNet to match clothes appearing in the videos to the exact same items in online shops. Faster-RCNN and Kernelized Correlation Filters (KCF) are adopted in this paper as the clothing detector and clothing tracker.

Xi Wang, Zhenfeng Sun, Wenqiang Zhang, Yu Zhou and Yu-Gang Jiang [4] have proposed a method to match a real-world product photo to the same item in online shopping sites using a Siamese network architecture that provides a unique capability that can rank the similarity between input image pairs by a contrastive loss function.

Chiao-Meng Huang, Chia-Po Wei and Yu-Chiang Frank Wang [8] have presented an active learning-based clothing image retrieval framework, which is able to implicitly learn the preferences from the user during retrieval processes, so that the user specific recommendation can be achieved.

S Jhansi rani and V. Valli Kumari [9], have proposed an efficient CBIR system based on IHCBS technique to retrieve relevant images from an image database for a given query image. In this method, when an image is queried, the system segments the image and extracts the feature from the image and then computes the similarity measure between the extensive features of the query image and the feature existing in the feature database based on the KLD method.

Kota Yamaguchi, M. Hadi Kiapour, Luis E. Ortiz, and Tamara L. Berg [10] presented about the retrieval of similar images with correct cloth fit from the database by parsing the query image. The parsing approach consists of two major steps one is to retrieve similar images from the parsed database and the second one is to use retrieved images and tags to parse the query. Post estimator is used for determining the correct fitness of the person.

III. METHODOLOGY

In this paper the proposed method for similar cloth retrieval can be looked as a conjunction of three sub tasks, i.e., Category Recognition, Feature Vector Extraction, Retrieval of Similar Images using Extracted Feature Vectors.

A. Dataset

The dataset used has 4,500 images, which is created using the Bing Image Search API, a restful web service [11]. The user can easily call this API by using any programming language that can make HTTP requests and parse JSON. The query to be searched is sent as a parameter.

The dataset is divided into five categories shirts, pants, dress, shoes and shorts. Each category has 900 images. For each category 700 images are used for training and remaining 200 images are used for testing.

B. Category Recognition

Our proposed method requires, identification of the class that a query image belongs to, as the first step to find similar images. This is done by using a Convolutional Neural Network (CNN).

A CNN is a type of neural network, which is widely used for image recognition and image classification. CNN's, like any other neural networks, are made up of neurons with learnable weights and biases. Each neuron receives input from several neurons, takes a weighted sum over them, pass it through an activation function and responds with an output. Activation function chosen to be used in our model is SoftMax activation function, since our proposed method requires that the output of a classification model should be a probability vector representing all classes of images, by using SoftMax function, for a given image, output of CNN will be a vector of length five, wherein each element belonging to the output vector represents the probability that a given query image belongs to that respective class.

Since the dataset acquired by us is less, instead of training a CNN with the acquired dataset, we propose to use the concept of transfer learning. Transfer learning is a machine learning method, in which a neural network, which is pre-trained on a very large dataset, is used for a similar but different problem. This way all the low-level attributes such as colour, boundaries and shape of an image can be easily identified. Our model proposes to use a CNN, which is pretrained on ImageNet dataset with Inceptionv3 architecture [12][13]. Since this architecture has a low computational cost, we chose to use this architecture. Inceptionv3 has incorporated all upgrades stated for Inception v2, and in addition to that RMSProp Optimizer, Factorized 7x7 convolutions, Batch Norm in the Auxiliary classifiers and label Smoothing have been used.

In our proposed methodology, the CNN model takes a query image as input and outputs a vector of probabilities of a query image with respect to each class. The final class of a query image is determined by the class which shows maximum probability.

C. Feature Vector Extraction

After the classification of the query image has been done by the CNN model mentioned in the earlier section, the feature vector or the encoded vector of the query image is obtained by using a trained Convolutional Denoising Autoencoder (CDAE).

An autoencoder neural network is an unsupervised Machine learning algorithm which has three layers, an input or the encoding layer, an output or the decoding layer and a hidden layer. It is trained to attempt to copy its input to its output. In other words, it is learning an approximation of the identity function. Approximation is the key word, because an identity function wouldn't be of much use, but the neural network needs to identify the most important features and include only them in the encoding vector, while ensuring that there is as little loss as possible when decoding the encoding vector in the output layer. Over the course of training the autoencoder on a dataset of images, it learns which features are the most important. Hence, dimensionality reduction is one of the most important uses of autoencoders [14].

In the traditional architecture of autoencoders, it does not consider the fact that a signal can be a sum of other signals. Convolutional Autoencoders (CAE), on the other way, use the convolution operator to accommodate this observation [15]. This in turn is very conductive to image processing. That is the reason, a Convolutional Autoencoder was chosen.

As mentioned above, the autoencoder is not of much use, if it is only an identity function. To avoid this, one of the regularization techniques used often is introducing noise in the training data, i.e., corrupting the data on purpose to encourage the autoencoder to have other properties other than just being an identity function. Therefore, a CDAE to create encoded vectors that contain the most important features of the training images. During training, the encoded vectors or latent space representation of all training images are created. Keras is used to build the CDAE model.

D. Retrieval of Similar Images

The distances between the encoded vector of the query image and the encoded vectors of the images used for training the CDAE model are calculated using the numpy.linalg.norm function. The above calculates the distance between the encoded vectors by Frobenius norm formula [16]. The distances are then sorted and the encoded vectors of the training images that have the five shortest distance from the encoded vector of the query image are identified. The corresponding training images are given as results. These are the five most similar images from the training dataset to the query image.

IV. RESULTS AND DISCUSSION

The optimizer used in the CDAE model is Adadelta and the metric of loss is Binary Cross-Entropy. The plots which display the training and validation losses during training of the CDAE model of each of the five classes are illustrated in Fig. 1-5 respectively. The lowest training and validation loss for each class is depicted in Table 1. It can be observed that, dresses have the highest training loss, i.e., 0.4403, whereas shoes have the least, i.e., 0.2962. All the classes are trained on CDAEs which are identical and have the same parameters. The difference in loss can only be attributed to the differences in the training datasets among classes. On manual inspection, it was observed that the dress training dataset had a greater number of images that had humans modelling the dresses, background noise, etc., compared to the number of such images in shoes dataset. The training loss of dresses being greater than shoes is because, the CDAE considered the differences in background and in the human models in the images in dress class as important and hence, led to a lossy decoded image from the encoded vector. This can be dealt with by normalizing things like background noise and drawing bounding boxes for the clothes in the images. That is because, when using an autoencoder to identify the important features in images and include them in the encoded vectors, it is imperative that all the details in the image are normalized to ensure that the model only identifies the similarities in the clothes rather than similarities in facial features of humans modelling the clothes, the pose the human model is in or the background noise.

The optimizer used in the CNN model is Adam and the metric of loss is Categorical Cross-Entropy. The CNN model shows an accuracy of 94% and a plot of training and validation loss is as shown in Fig.6.

Table.2 illustrates the Denoised, Noisy, Query and Retrieved similar images for one query image from each class. The noisy image is obtained by adding noise to the query image by a noise factor of 0.2. The denoised image is the output obtained from the CDAE model after decoding of the encoded vector. The retrieved images are the images from the training dataset of the class of the query image as determined by the CNN classifier in the Category Recognition task, that are most alike to the query image. The five most similar images are displayed. It can be noted that the clothes in the retrieved images have similarities with the query image in terms various features such as colour, cut, design, pattern etc.

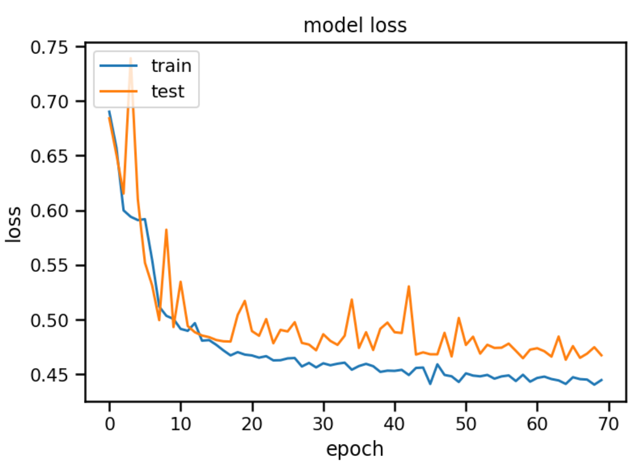

Figure 1. Training and Validation loss plot for Dresses

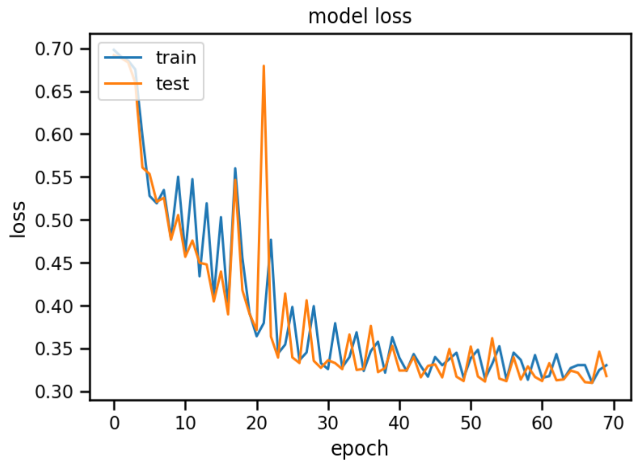

Figure 2. Training and Validation loss plot for Pants

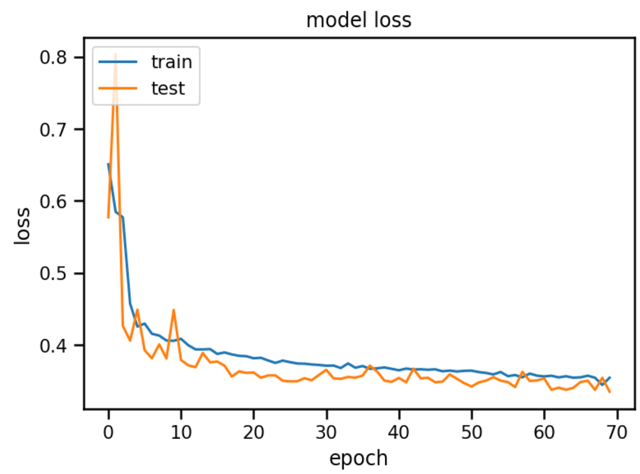

Figure 3. Training and Validation loss plot for Shirts

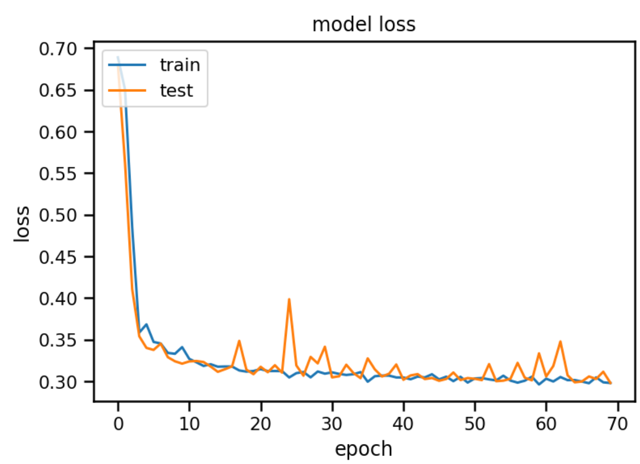

Figure 4. Training and Validation loss plot for Shoes

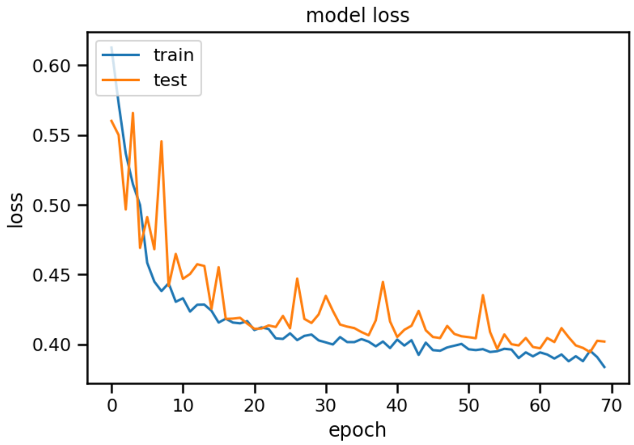

Figure 5. Training and Validation loss plot for Shorts

Table 1. Training and Validation Losses of Five Classes of Clothing Images

Class	Shirts	Shoes	Shorts	Pants	Dresses
Training Loss	0.3444	0.2962	0.3836	0.3097	0.4403
Validation Loss	0.3350	0.2981	0.3966	0.3096	0.4631

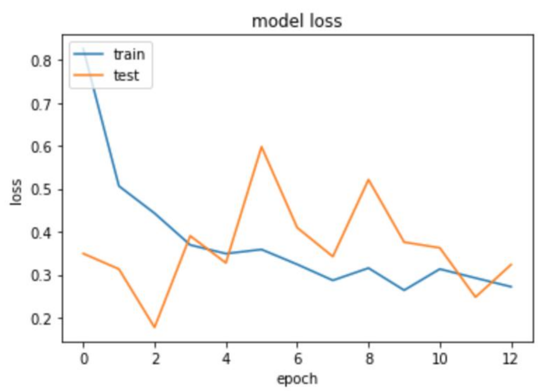

Figure 6. Training and Validation loss plot for CNN

Table 2. The Result Images for Five Classes

Class	Query Image	Retrieved Images
Shirts		
Shoes		
Shorts		
Pants		
Dress		

V. Conclusion and Future Scope

In this paper, we have proposed a neural network-based approach to find similar images for a given image. We propose that visual search is better than using textual key-words to describe a fashion item. This includes Category Recognition, Feature Extraction and comparison of the extracted feature vectors to find similar images. We also show that our methodology can be applied to various commercial domain applications, easily adapting to new E-commerce datasets by exploiting the product images and their associated categories.

We would like to expand our work to facilitate search based on shape and style, search by words, by tagging clothing images with words. And expand the work to other multimodal items and accessories. As mentioned in the previous section, we would like to focus on normalizing background noise and adding bounding boxes for the clothes in the images to reduce the loss of the CDAE model and ensure lesser lossy decoded images. We also plan to obtain the vectors of images right before the output layer during the training of the CNN model. Then these vectors will be used to find similar images to the query image by using the same method described in Section III, part D, of the paper, i.e., finding the vectors with the shortest distances to the vector of the query image and retrieving the images corresponding to those vectors. We plan to compare the results of this method, with the results of this paper.

Acknowledgment

The authors gratefully acknowledge the contributions of the authors of the refereed papers for their work and Faculty of Engineering, Ramaiah Institute of Technology for their constant support towards research.

REFERENCES

[1] S. Ota, H. Takenouchi and M. Tokumaru, "*Kansei clothing retrieval system using features extracted by autoencoder*", 1-7. 10.1109/SSCI.2017.8285320 (2017).

[2] Z.-Q. Cheng, X. Wu, Y. Liu, and X.-S. Hua. "Video2shop: Exact matching clothes in videos to online shopping images", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4048–4056, (2017)

[3] A. Y. Ivanov, G. I. Borzunov and K. Kogos, "Recognition and identification of the clothes in the photo or video using neural networks", *2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, Moscow, 2018, pp. 1513-1516.

[4] X.Wang, Z.Sun,W.Zhang, Y.Zhou,and Y.Jiang. "*Matching user photos to online products with robust deep features*", In ICMR, pages 7–14, (2016).

[5] L. Ami D, M. Shilpa I, R. Urvashi K., S. Saliha H. "*Reverse Image Search Image retrieval using various techniques*", International Journal of Advance Engineering and Research Development (IJAERD) Volume 4, Issue 4, April-2017, e-ISSN: 2348 - 4470, print-ISSN: 2348-6406.

[6] M. M. Singh, C. David, P. M. Rathod, "*Smart Image Search Android application using Reverse Image Search*", IJREAM, vol. 2, no. 1, pp. 1-4, (2016).

[7] Kaur, M., & Sohi, N. "A novel technique for content-based image retrieval using colour, texture and edge features", In Proceedings of the international conference on communication and electronics systems (ICCES), (2016)

[8] Huang, C., Wei, C., Wang, Y. "Active learning-based clothing image recommendation with implicit user preferences", In ICME Workshops, pp. 1–4 (2013)

[9] J. Rani, S & Valli Kumari, V. "Improved Hill Climbing Based Segmentation (IHCBS) technique for CBIR system", 1-10. 10.1109/ICECCT.2015.7226069 (2015).

[10] K. Yamaguchi, M. H. Kiapour, and T. L. Berg. "Paper doll parsing: Retrieving similar styles to parse clothing items". In ICCV, (2013).

[11] Adrian Rosebrock, "How to (quickly) build a deep learning image dataset", pyimagesearch.com, https://www.pyimagesearch.com/2018/04/09/how-to-quickly-build-a-deep-learning-image-dataset/ (accessed Apr.27, 2019)

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. "Rethinking the inception architecture for computer vision", In CVPR, 2016.

[13] M. Yang and K. Yu. "Real-time clothing recognition in surveillance videos", In ICIP, (2011).

[14] I. Goodfellow, Y. Bengio, A. Courville, "Deep Learning", Cambridge, MA, USA, MIT Press, ch. 14 ,2016.

[15] V. Jayawardana, "Autoencoders—Bits and Bytes of Deep Learning", towardsdatascience.com, https://towardsdatascience.com/autoencoders-bits-and-bytes-of-deep-learning-eaba376f23ad (accessed Apr. 27, 2019).

[16] The SciPy community, "numpy.linalg.norm", docs.scipy.org, https://docs.scipy.org/doc/numpy-1.14.0/reference/generated/numpy.linalg.norm.html#numpy-linalg-norm (accessed Apr. 27, 2019).

[17] M. H. Kiapour, X. Han, S. Lazebnik, A. C. Berg, and T. L. Berg, "Where to buy it: Matching street clothing photos in online shops", In ICCV, (2015)

[18] A. Y. Ivanov, G. I. Borzunov and K. Kogos, "Recognition and identification of the clothes in the photo or video using neural networks," 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), Moscow, 2018, pp. 1513-1516.

[19] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang. "Deepfashion: Powering robust clothes recognition and retrieval with rich annotations", In CVPR, pages 1096–1104, 2016.

[20] S. Thompson, M. Reilly, "*A picture is worth a thousand words: Reverse image lookup and digital library assessment*", J. Assoc. Inf. Sci. Technol., vol. 68, pp. 2264-2266, 2017, DOI: 10.1002/asi.

[21] P. Nieuwenhuysen, "*Search by image through the Internet: applications and limitations*", Libraries in the Transition Era: New Space –New Services – New Experience: Proc. of the 7th Shanghai International Library Forum, pp. 145-155, (2014), [online] Available: http://www.vub.ac.be/BIBLIO/nieuwenhuysen/presentations/.

[22] Punam Mahesh Ingale, "*The importance of Digital Image Processing and its applications*", International Journal of Scientific Research in Computer Science and Engineering, Vol.06, Issue.01, pp.31-32, 2018.

[23] N.S. Lele , "*Image Classification Using Convolutional Neural Network*", International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.3, pp.22-26, 2018.

[24] John Bosco.P., S.K.V. Jayakumar, "*A Study on Web Based Image Search by Re-Ranking Techniques*", International Journal of Scientific Research in Network Security and Communication, Vol.5, Issue.3, pp.19-26, 2017.

[25] Mahajan J.R., C.S. Rawat, "*Object Detection and Tracking using Congnitive Approach*", International Journal of Scientific Research in Network Security and Communication, Vol.5, Issue.3, pp.136-140, 2017.

Authors Profile

S.R. Katasani is currently pursuing Bachelor of Engineering with a specialization in Computer Science and Engineering from Ramaiah Institute of Technology (Affiliated to VTU), India in year 2019.

T. Rachepalli is currently pursuing Bachelor of Engineering with a specialization in Computer Science and Engineering from Ramaiah Institute of Technology (Affiliated to VTU), India in year 2019.

N. Kamat is currently pursuing Bachelor of Engineering with a specialization in Computer Science and Engineering from Ramaiah Institute of Technology (Affiliated to VTU), India in year 2019.

M. Jadhav pursed Bachelor of Engineering from Bangalore University, India in 1996 and Master of Technology from VTU in the year 2007. She is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Computer Science, Ramaiah Institute of technology, India since 2007. Her main research work focuses on computer networks, database management systems and data mining. She has 20 years of teaching experience.