JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Improved Deep Transfer Learning Based VGG19 Network Enabled Content Based Image Retrieval Model

P. Sasikumar, S. Sathiamoorthy

Research Scholar, Dept. of Computer and Information Science, Annamalai University, Chidambaram, India

Assistant Professor, Annamalai University, PG Extension Centre, Villupuram, India.

mailtoausasikumar@gmail.com, ks_sathia@yahoo.com

Abstract—Image retrieval (IR) methods gains more popularity because of its huge accessibility of multi-media datasets. The current IR system had an incredible performance over labeled dataset. But often, data labelling is expensive and infrequently not possible. Thus, unsupervised and self-supervised learning techniques are renowned techniques. Many self or unsupervised techniques are sensitive to the several classes and cannot combine labeled data on availability. With this study, this study focuses on the design of CBIR model using VGG19 Network with Class Attention Mechanism (CBIR-VGGCAM). The CBIR-VGGCAM technique derives a class attention layer (CAL) for capturing the discriminatory class-specific feature. In addition, the Canberra distance is applied as a similarity measurement metric for determining the resemblance among the feature vector of the images. Upon providing the query image (QI), the feature vectors are derived by the VGG19 network and a similarity measurement is made with the feature vector of the image database. Eventually, the images with high resemblance are retrieved in the database. A wide-ranging experiment was take place on benchmark databases and the obtained outcomes depict the advancement of the CBIR-VGGCAM algorithm over the recent approaches.

Keywords— Content based image retrieval, Class attention layer, VGG19 networks, Deep learning, Corel10K dataset

1. Introduction

Currently, the image-oriented information has been rapidly increasing because of the explosion of social media and smartphones. Similarly, the demand and use of image-based queries are accelerating that is people search for data through the textual queries. Image similarity search is a type of image retrieval (IR) policy that search images on the basis of specified query image (QI). The image similarity is defined by several features like shape, color, structure and texture. Image-based similarity search techniques are termed as a content-based image retrieval (CBIR) system. At present, CBIR is ruling because of the heavy demand of image-oriented data information retrieval system. Hence, CBIR systems are modelled as domain-specific. Face retrieval system query for same facial image for the given QI. Medical IR system helps diagnoses to be accurate and easy. Then Product retrieval system will find out the cherished products of users from online shopping. Cloth retrieval system may aid the users to find their desirable product. By comparison with domain-specific IR system, multipurpose IR system sightsee the general relation, unbiased data. Currently, search engine inherits the expediency of CBIR system for querying similar images. CBIR system consist of 2 key elements one is finding similar images and another one is understanding image content. The overview of CBIR is given in Fig. 1.

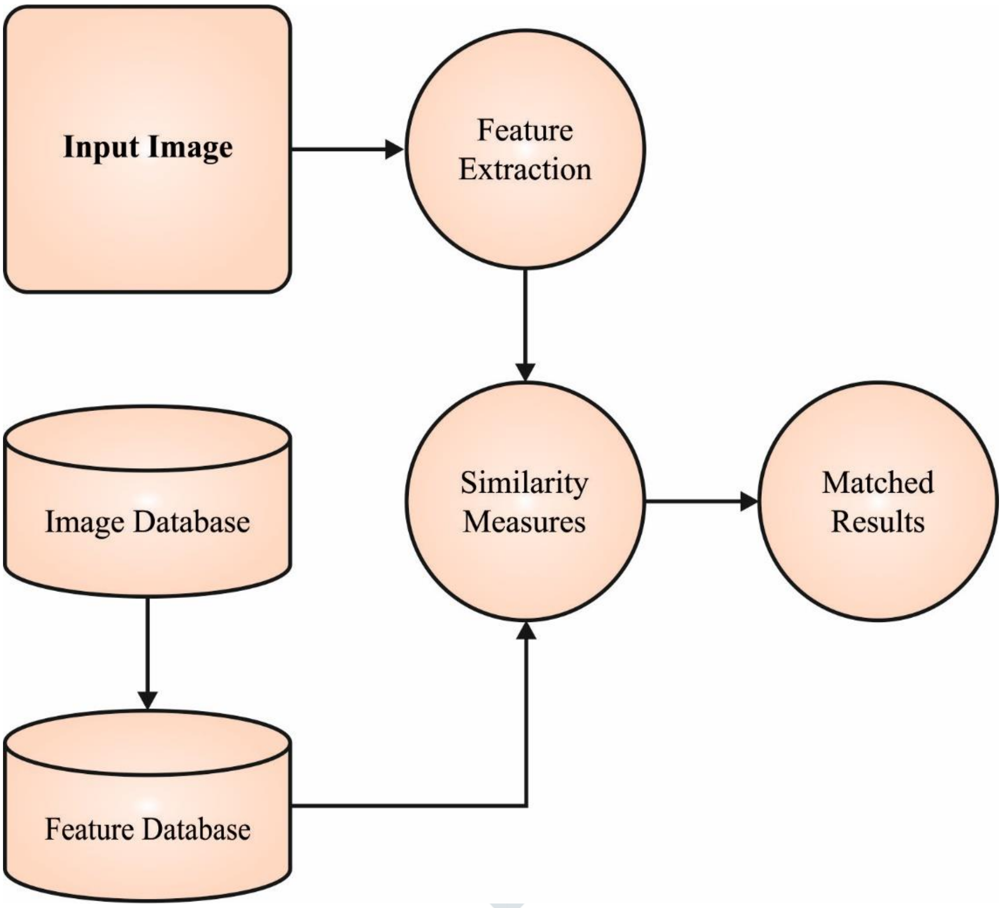

Fig. 1. Overall workflow of CBIR model

Over a year, a transition was noted in feature representation from hand engineering to learning based afterward on the appearance of deep learning (DL) methods [5]. The convolutional neural network (CNN) based feature learning replaces an advanced form of conventional hand-crafted features. The DL method is a hierarchical feature representation method for learning the abstract feature in data that is significant to that application and dataset [6]. Depending on the types of data to be treated, distinct frameworks have been established namely CNN for image data, Recurrent Neural Networks (RNN) for time sequence dataset, Multilayer Perceptron (MLP), or Artificial Neural Network (ANN) for one dimensional data [7]. Tremendous growth has been created in this year for utilizing the influence of DL method for CBIR model.Yang et al. [8] presented a DNN based IR technique in which saliency map has been resultant to procedure human gaze shifting paths with constraint metrics. Clearny, it is primary plan for a DNN-based image saliency forecast. Afterward, it can be leverage image quality assessment (IQA) technique for selecting high-quality salient region that is concatenated in order with utilizing presented constraint metrics for mimicking human visual perceptions. Passalis et al. [9] introduced a discriminatory deep metric learning technique which purposes to not only learning a representation which permits discrimination amongst distinct classes, however, it can be also able of encoded the latent generative factor distinctly to all classes, overcoming this restraint.

Tzelepi and Tefas [10] employed a deep CNN technique for accomplishing the feature representation in the activation of convolution layers utilizing Maxpooling, and afterward, it can be adapting as well as retraining the network, for producing further effectual compact image descriptors that enhance combined the retrieval efficiency and the memory necessities, rely on the existing data. Öztürk [11] examined the suitability of sparse vector from the dictionary learning technique to CBIR task. As DL generally carries out the learning procedure from unsupervised approach, it could not produce robust features to the retrieval tasks, particularly when the difficult background has been difficult. For overcoming this disadvantage, the DL technique utilizing the feature illustration power of CNN has been presented. Wu et al. [12] investigated 2 novel rotation-aware CNN-based CBRSIR techniques. The final pooling layer has been interchanged from 4 distinct angles in the trained, During the Feature Map Transformation Based Rotation-Aware Network (FMT-RAN).

This paper proposes a novel CBIR model using VGG19 Network with Class Attention Mechanism (CBIR-VGGCAM). The CBIR-VGGCAM technique derives a class attention layer (CAL) for capture the discriminatory class-specific feature. In addition, the Canberra distance is applied as a similarity measurement metric for determining the resemblance among the feature vectors of the images. The design of VGG19 network with CAL shows the novelty of the work. In order to measure the improved performance of CBIR-VGGCAM model, a comprehensive study is made using Corel 10K dataset.

2. The Proposed model

In this study, a new CBIR-VGGCAM technique is derived to effectively retrieve the set of images based on the QI. The projected CBIR-VGGCAM technique encompasses two major processes namely VGG19 with CAL based feature extraction and Canberra distance based similarity measurement. The overall working process is given in the subsequent section

2.1. Design of Feature Extraction Module

In this phase, the feature extraction of the QI and image database takes place using the VGG19 network. CNN is a kind of DL method for processing information which has a grid pattern, like images and developed to adaptively and automatically explore spatial hierarchy of features, from lower to higher level pattern. CNN is an arithmetical model is generally made up of 3 kinds of layer (or fundamental component): fully connected (FC), convolution, and pooling layers. The primary two, pooling and convolution layers, implement feature extraction, where the third, a FC layer, map the extracted feature to last output, like classification. A convolutional layer plays an important role in CNN, i.e., made up of a stack of arithmetical functions like convolutional, a special kind of linear operation [13]. The procedure of parameter tuning like kernel is known as training, i.e., implemented for minimizing the difference among ground truth labels and outputs via an optimization method named gradient descent and, backpropagation amongst others.

The VGG19 is analogous to VGG16 however these network consists of 19 layers rather than16 that has 3 FC dense layers, 16 convolution layers. The first two layer encompasses 64 filter of 3 X 3 kernel and Maxpooling layer. The second and third convolution layers includes 128 filters of 3 X 3 kernel and Maxpooling layer [22]. Numerous sets of 4 convolutional layers including 512 filters of 3 X 3 kernel and Maxpooling layers can be organized in a sequential manner. After that, this output can be provided into FC layers. There will be 3 FC dense layers having 1000, 4096, and 4096 neurons correspondingly. Excluding the last dense layer, the activation function was ReLU for every layer where softmax function has been employed.

Though the deep features derived by the VGG19 model are higher level and it could be straightway passed into the FC layer for producing the outcome, it can be impossible for discovering the high order probabilistic dependency via repeatedly feeding it to matching features. So, the extraction of different class wise feature performs as a vital part to discover the class dependency and proficiently bridge the VGG19 model for effective performance. The CAL layer includes two levels namely generation of class attention map by 1×1 convolutional layer with stride 1, and vectorization of all the classes attention map for attaining class-oriented features. In formal, for a provided feature map X, derived from the VGG19 manner with a size of $W \times W \times K$, and assume w_l denote the l-th convolution filter in the CAL layer. The attention map M_l to class l is derived using Eq. (1):

$$M_l = X * w_l, \tag{1}$$

where l lies in the range of 1 to class count and $*$ denotes the convolutional function. Assume that filter size of 1×1 convolution, M_l class attention map is inherently a linear mixture of every channel in X. The CAL layer has the ability to learn discriminative class attention maps [15]. Rather than completely linking the class attention maps to every hidden unit under the subsequent layer, a class wise connection amongst the class attention map is generated and the one-to-one hidden unit.

2.2 Similarity Measurement

The feature vector of the QI and database images is denoted as $F_eq = [F_eq(1), F_eq(2) \dots \dots . F_eq(n)]$ and $F_eVdb = [F_eVdb(1), F_eVdb(2) \dots \dots . F_eVdb(n)]$. The aim of the similarity metric is to define the optimal 'n' related images from the database which are identical to the QI. In this study, the Canberra distance is used for similarity measurement, which can be represented using Eq. (2).

$$CA_d = \sum_{f=1}^{n} \frac{\left|F_eV_{db}(f) - F_{e_q}(f)\right|}{|F_eV_{db}(f)| + |F_eV_q(f)|} \tag{2}$$

where n represents the feature vector length. F_eVq and F_eVdb represent the feature vectors of the QI and database images correspondingly. When the distance becomes low, the retrieval of images will be effective.

3. Experimental Validation

The performance of CBIR-VGGCAM approach is validated by means of Corel10K dataset [16]. The outcomes are investigated based on precision and recall. Fig. 2 depicts the few sample images of query and retrieved images.

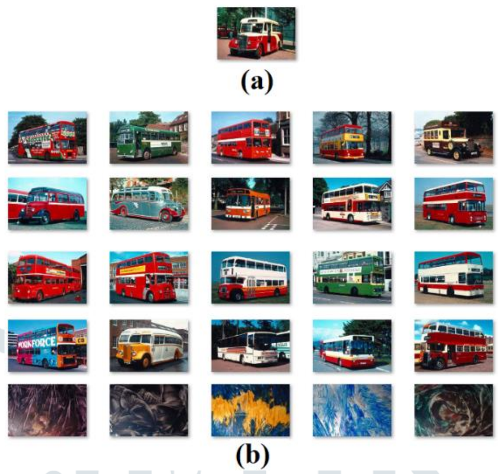

Fig. 2. Visualization of sample output (a) QI (b) Set of retrieved images

Fig. 3 demonstrates the comparative outcomes analysis of the CBIR-VGGCAM technique with recent ones with respect to precision. The results demonstrated that the IGA technique has accomplished lower performance over the compared approaches. In addition, the BR-SVM model has gained slightly improved precision values over the IGA technique. Followed by, the GAGD, MHA-CBIR, CTEF, and FIF-IRS methods have accomplished considerably precision values. However, the CBIR-VGGCAM technique has outperformed other ones with an increased precision value under distinct classes.

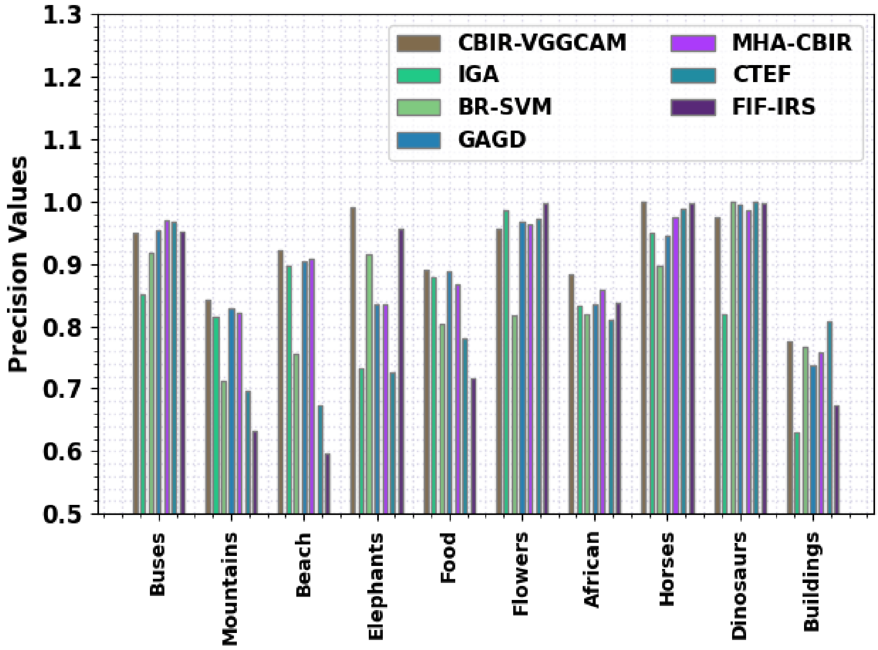

Fig. 3. Result analysis of CBIR-VGGCAM model with respect to precision

Fig. 4 showcases the comparative outcomes analysis of the CBIR-VGGCAM approach with existing ones in terms of recall. The outcomes outperformed that the IGA manner has accomplished minimum efficiency on the compared methods. Besides, the BR-SVM technique has attained somewhat enhanced recall values over the IGA technique. At the same time, the GAGD, MHA-CBIR, CTEF, and FIF-IRS methodologies have accomplished considerably recall values. Eventually, the CBIR-VGGCAM system has demonstrated enhanced efficiency with the maximal recall value under distinct classes.

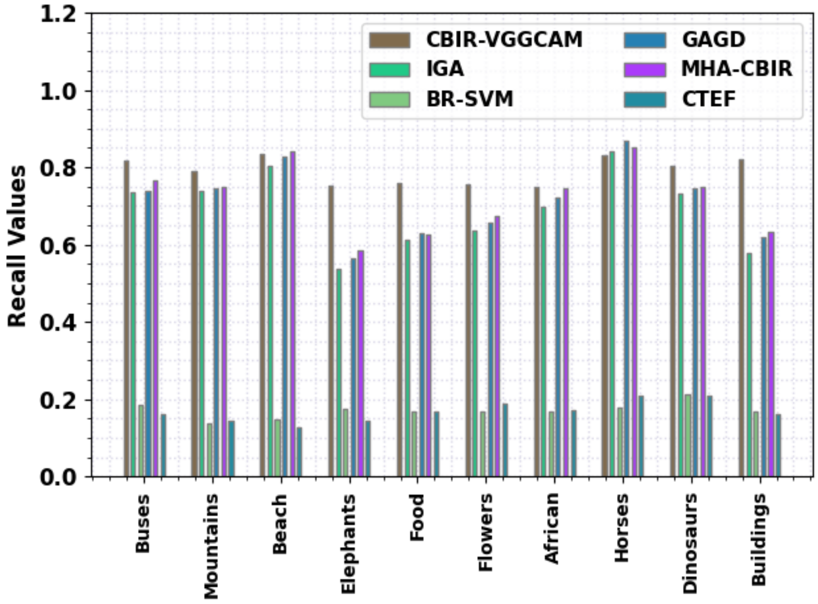

Fig. 4. Result analysis of CBIR-VGGCAM model with respect to recall

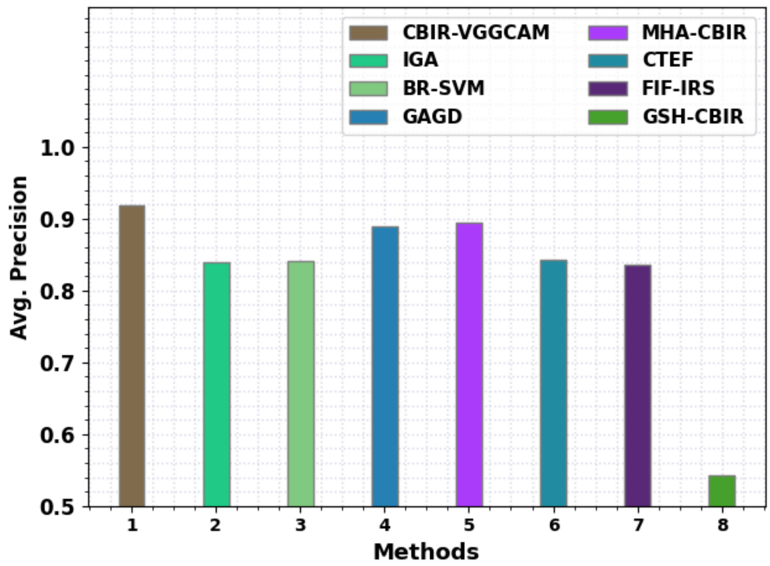

Fig. 5. APN analysis of ELD-THCR model

Finally, a comprehensive comparative results analysis of the CBIR-VGGCAM with recent techniques is in Figs. 5-6 [17-23]. The results demonstrated that the GSH-CBIR method has lower performance over the other methods with the APN and ARL of .054 and 0.54 respectively.

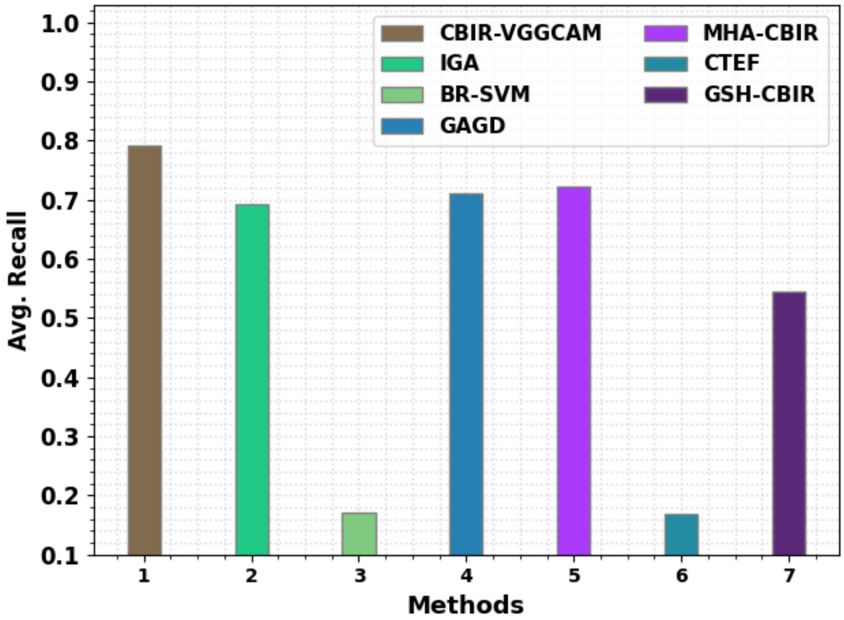

Fig. 6. ARL analysis of ELD-THCR model

At the same time, the BR-SVM, IGA, CTEF, FIF-IRS, and GAGD techniques have shown moderately closer performance. However, the proposed CBIR-VGGCAM technique has outperformed the existing techniques with the maximum APN and ARL of 0.912 and 0.789 respectively.

From the abovementioned discussion and results, it is evident that the CBIR-VGGCAM technique has the ability to proficiently retrieving the images.

4. Conclusion

In this study, a new CBIR-VGGCAM technique is derived to effectively retrieve the set of images based on the QI. The presented CBIR-VGGCAM technique encompasses two major processes namely VGG19 with CAL based feature extraction and Canberra distance-based similarity measurement. The application of CAL in the VGG19 network helps for capturing the discriminative class-specific features from the applied input images. For inspecting the improved retrieval outcomes of the CBIR-VGGCAM technique, a comprehensive comparative study is made and the experiment results highlighted the advancement of the CBIR-VGGCAM technique with the maximum APN and ARL of 0.912 and 0.789. Therefore, the CBIR-VGGCAM technique can be employed to retrieve images in real time applications. In the future, the performance of CBIR-VGGCAM technique has been improved by utilize of metaheuristic based hyperparameter optimization approaches.

References

[1]Dubey, S.R., 2021. A decade survey of content based image retrieval using deep learning. IEEE Transactions on Circuits and Systems for Video Technology.

[2]W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the earny years," IEEE TPAMI, vol. 22, no. 12, pp. 1349–1380, 2000.

[3]H. Muller, W. M ¨ uller, D. M. Squire, S. Marchand-Maillet, and T. Pun, ¨ “Performance evaluation in content-based image retrieval: overview and proposals,” Pattern Recog. Letters, vol. 22, no. 5, pp. 593–601, 2001.

[4]T. Deselaers, D. Keysers, and H. Ney, “Features for image retrieval: an experimental comparison,” Information Retrieval, vol. 11, no. 2, pp. 77–107, 2008.

[5]L. Jing and Y. Tian, “Self-supervised visual feature learning with deep neural networks: A survey,” IEEE TPAMI, 2020.

[6]Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[7]J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Gated feedback recurrent neural networks,” in ICML, 2015, pp. 2067–2075.

[8]Yang, Y., Jiao, S., He, J., Xia, B., Li, J. and Xiao, R., 2020. Image retrieval via learning content-based deep quality model towards big data. Future Generation Computer Systems, 112, pp.243-249.

[9]Passalis, N., Iosifidis, A., Gabbouj, M. and Tefas, A., 2020. Variance-preserving deep metric learning for content-based image retrieval. Pattern Recognition Letters, 131, pp.8-14.

[10] Tzelepi, M. and Tefas, A., 2018. Deep convolutional learning for content based image retrieval. Neurocomputing, 275, pp.2467-2478.

[11] Öztürk, Ş., 2021. Convolutional neural network based dictionary learning to create hash codes for content-based image retrieval. Procedia Computer Science, 183, pp.624-629.

[12] Wu, Z.Z., Zou, C., Wang, Y., Tan, M. and Weise, T., 2021. Rotation-aware representation learning for remote sensing image retrieval. Information Sciences, 572, pp.404-423.

[13] Yamashita, R., Nishio, M., Do, R.K.G. and Togashi, K., 2018. Convolutional neural networks: an overview and application in radiology. Insights into imaging, 9(4), pp.611-629.

[14] Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F. and Zhang, Y., 2018. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. Remote Sensing, 10(7), p.1119.

[15] Hua, Y., Mou, L. and Zhu, X.X., 2019. Recurrently exploring class-wise attention in a hybrid convolutional and bidirectional LSTM network for multi-label aerial image classification. ISPRS journal of photogrammetry and remote sensing, 149, pp.188-199.

[16] http://www.ci.gxnu.edu.cn/cbir/Dataset.aspx

[17] Madhavi, K.V., Tamilkodi, R. and Sudha, K.J., 2016. An innovative method for retrieving relevant images by getting the top-ranked images first using interactive genetic algorithm. Procedia Computer Science, 79, pp.254-261.

[18] Ashraf, R., Bajwa, K.B. and Mahmood, T., 2016. Content-based Image Retrieval by Exploring Bandletized Regions through Support Vector Machines. J. Inf. Sci. Eng., 32(2), pp.245-269.

[19] Alsmadi, M.K., 2017. An efficient similarity measure for content based image retrieval using memetic algorithm. Egyptian journal of basic and applied sciences, 4(2), pp.112-122.

[20] Alsmadi, M.K., 2018. Query-sensitive similarity measure for content-based image retrieval using meta-heuristic algorithm. Journal of King Saud University-Computer and Information Sciences, 30(3), pp.373-381.

[21] Pavithra, L.K. and Sharmila, T.S., 2018. An efficient framework for image retrieval using color, texture and edge features. Computers & Electrical Engineering, 70, pp.580-593.

[22] Bella, M.I.T. and Vasuki, A., 2019. An efficient image retrieval framework using fused information feature. Computers & Electrical Engineering, 75, pp.46-60.

[23] Yuan, B.H. and Liu, G.H., 2020. Image retrieval based on gradient-structures histogram. Neural Computing and Applications, pp.1-11.