Proceedings of 12th IOE Graduate Conference
Peer Reviewed
ISSN: 2350-8914 (Online), 2350-8906 (Print)
Year: 2022 Month: October Volume: 12

Content Based Image Retrieval Based on Image Feature Fusion and Principal Component Analysis

Ritu Thapa [a], Sanjeeb Prasad Panday [b]

[a,b] *Department of Electronics and Computer Engineering, Pulchowk Campus, IOE, TU, Nepal*
[a] riyaazkshetri@gmail.com , [b] sanjeeb@ioe.edu.np

Abstract
This paper proposes an analysis of image feature fusion(IFF) based method with different combinations of Inception-V3, Resnet50 and EfficientnetB0 for efficient image retrieval. Principal component analysis(PCA) technique has been introduced to produce fused image feature maps with the help of dimensionality reduction concept. Experimentations have been conducted using euclidean, cityblock and correlation distance functions on popular natural image dataset Corel-1k. The results have been evaluated using mean average precision(MAP) scores. The experimental results on Corel-1k dataset show that the best mean average precision scores for top 20 and top 50 retrieved images are 96.85% and 95.92% respectively. The experimental results also show that the proposed method outperforms some methods using shallow descriptors for feature extraction purpose and the methods using fusion of low level features extracted from classical approaches.

Keywords
CBIR, CNN, DNN, IFF, QI, PCA, NR, AP, MAP

1. Introduction

It is widely accepted that due to rapidly increasing digital image databases, it is critical to have a reliable, fast and effective retrieval strategy for image retrieval. Traditionally, the image retrieval procedure necessitates a text annotation and a keyword search to describe all images. As the volume and diversity of image contents have substantially expanded, it appears that they are extremely challenging and ambiguous steps for text-based image retrieval to provide a fast and efficient search for the query image. As a result, content based image retrieval (CBIR) is a method for automatically representing and indexing images using image features [1]. The fundamental notion of CBIR is to search for relevant images using distance measurements for a query image from a database. Deep learning has emerged as a successor and the go-to tool for many machine learning tasks has been used widely in most applications and has proven to outperform other traditional machine learning approaches in the image retrieval task[2].

The proposed work draws from topics in machine learning, computer vision and image processing. An overview of basic building blocks, concepts and terminology is provided in Figure 1 to give context.

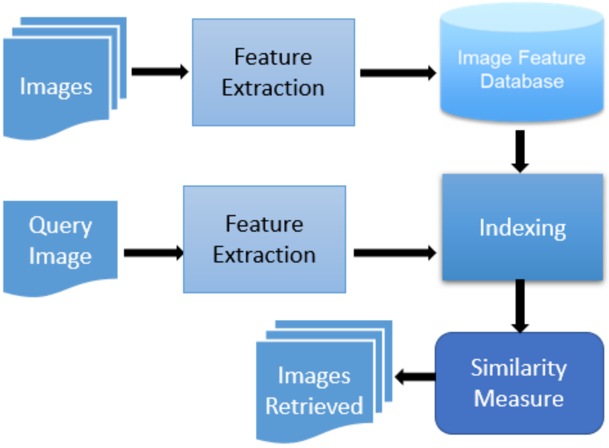

Figure 1: Basic CBIR system

Image representation by feature extraction,indexing and similarity measure are basic blocks for the system. Feature refers to color, texture, shape etc. of an image. All images are represented as feature maps when passes through the feature extraction blocks, indexing of those feature maps is provided for fast image searching task and similarity measure is done for comparison between feature maps in terms of their similarity. Deep neural networks (DNNs) using deep architecture have lately shown to be very good at complicated machine learning tasks like image classification, image retrieval and speech recognition. Many researchers have recommended CNNs as excellent feature extractor for CBIR approach [2]. Most of the CBIR approaches have utilized a single CNN architecture for feature extraction [3]. Although good performance has been achieved with the single feature extractor architecture, it can be improved using different feature fusion strategies [4]. Layer-level fusion and model-level fusion has been implemented and analysed in literature [4]. Many experimentation on layer-level fusion has been done in which features form different layers of a network are extracted and fused for efficient retrieval. For model-level fusion most of the experiment has been conducted with shallow networks for intra-model fusion [4]. But, there is a lack of subsequent experimentation on inter-model fusion [4] strategy with multiple combination of different powerful CNN architectures for image retrieval. This work tries to address this research gap.

2. Related Work

Previously, various researchers worked on evaluating the CBIR approach to retrieve images from large datasets [2]. It has been seen that the earliest research work on CBIR using deep learning was published by M. Flickner, H. Sawhney, et al., in 1995 [5]. The authors have investigated content-based retrieval strategies for the creation of the Query By Image Content(QBIC) system. Searching for images by their content in a large dataset has been one of the challenging tasks in computer vision [1]. In recent years, it has shown that CBIR using Convolution Neural Network is a rapidly evolving technology with enormous potential [6, 3]. Researchers are focusing on finding the feature of how human's vision sense works and implementing that conclusion in computers to understand images[7]. With the recent development of the CBIR system, predicting similar images is faster and more efficient due to the use of various kinds of similarity measurement scores [8].

Since the advent of deep learning, many research works on CBIR have used it as an essential tool [2, 4, 9]. Recently, many researchers were interested on feature fusion idea for CBIR. In research by Xie, G.et al. in 2020 [10], low level features extracted from two shallow descriptor dominant color descriptor and Hu moment are fused together for efficient image retrieval approach using shallow feature fusion strategy for three different benchmark datasets Corel-1K, Corel-5K, Corel-10K. In research by Jiang, D.et al.in 2021 [11] low level features are extracted using hue-saturation-value (HSV) histogram, uniform local binary patterns (LBP), Dual-Tree complex wavelet transform (DTCWT) are fused using fisher coding technique and the high level feature extraction is done by Alexnet. Again, the high level features are fused with the low level features implementing shallow feature and deep feature fusion strategy for benchmark dataset Corel-1K.

A promising work in feature fusion for CBIR is the recent experiment by Jiang,D. and Kim,J. [8]. It has been shown that shallow feature extraction and representation can be improved by using Shallow feature fusion for CBIR system [8]. Low level features are extracted using hue-saturation-value (HSV) histogram, uniform local binary patterns (ULBP), Dual-Tree complex wavelet transform (DTCWT) are fused implementing shallow feature fusion strategy [8]. Object-centric network and place-centric network are introduced for deep feature fusion using Resnet50 based on Discrete Cosine Transform (DCT). Also, improved performance using PCA instead of DCT thus by implementing shallow feature and deep feature fusion strategy for different natural image datasets Corel-1K, Corel-5K, Corel-10K, Corel-DB, Oxford-5K. Instead of fusing features extracted by fully connected layers of networks we can experiment the image feature fusion from different layers inspired by the survey [4].

3. Proposed Work

3.1 Dataset Description

The dataset utilized to evaluate the performance of CNNs for the proposed work is labeled high-quality standard dataset i.e., Corel 1K. The dataset is made up of 1000 digital images, 100 of which are for each class. The digital images are either (256x384) or (384x256) in size.

3.2 System flow diagram

The system is divided into offline feature extraction and on-line image retrieval. The system collects fused feature maps from each image and stores them in the database in the off-line stage (features of the images are extracted and represented with feature vectors). The user submits an image query to the system during the on-line stage.The proposed system flow diagram is

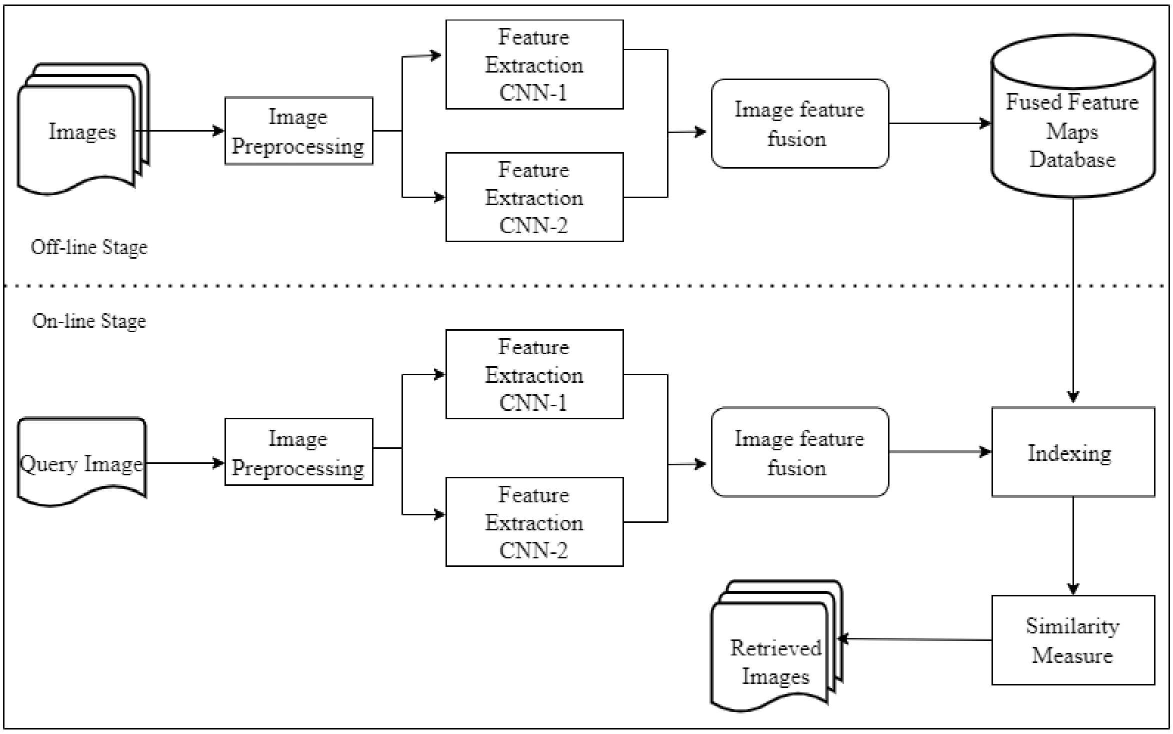

Figure 2: Proposed methodology for CBIR

explained as given below:

3.2.1 Image Preprocessing

Resizing and normalization is used to get the image data in the range required by the input layer of the neural networks. Data augmentation has also been performed as a preprocessing step in order to increase the effective number of samples fed to the neural network for training. Random resized crop and random rotation have been performed on the data during this step.

3.2.2 Feature Extraction

Inception-v3, Resnet50 and EfficientNetB0 have been trained using transfer learning from pre-trained models on imagenet dataset and they have been used for feature extraction.

3.3 Image Feature Fusion

Image feature fusion is the fusion of feature matrices to produce fused feature map. Two different learning models are used to extract feature matrices of the input images as shown in Figure 3. Considering the suitable feature dimensions, the feature dimension of different networks after discarding the fully connected layers are used for efficient feature representation. The feature dimension after concatenation of features from two networks are passed to PCA for feature dimensional reduction.

$$F_f = PCA\{f_1, f_2\} \tag{1}$$

Here the F_f is defined as fused feature map from simplest image feature fusion which is concatenation of different feature vectors and applying PCA. The terms f_1 and f_2 are D dimensional features extracted by CNN-1 and CNN-2 respectively as shown in Figure 3. The reduction of dimension of concatenated feature map is done by using principal component analysis technique due to its effective performance for finding the most important components or features from large numbers of scattered features. PCA is one of the powerful technique used in machine learning, pattern recognition and image processing as a dimensionality reduction method [12, 13]. The Image feature fusion can also be done using more than two networks for numerous experimentation with the help of concatenations of different feature vectors from different networks.

3.3.1 Similarity Measure

Let us suppose x_{QRI} and x_{DBI} referred to feature vector of query image and feature vector of database image respectively with the n dimension, and let j be the class image then Euclidean distance (D_1) , Manhattan distance/city block (D_2) and Correlation distance (D_3)

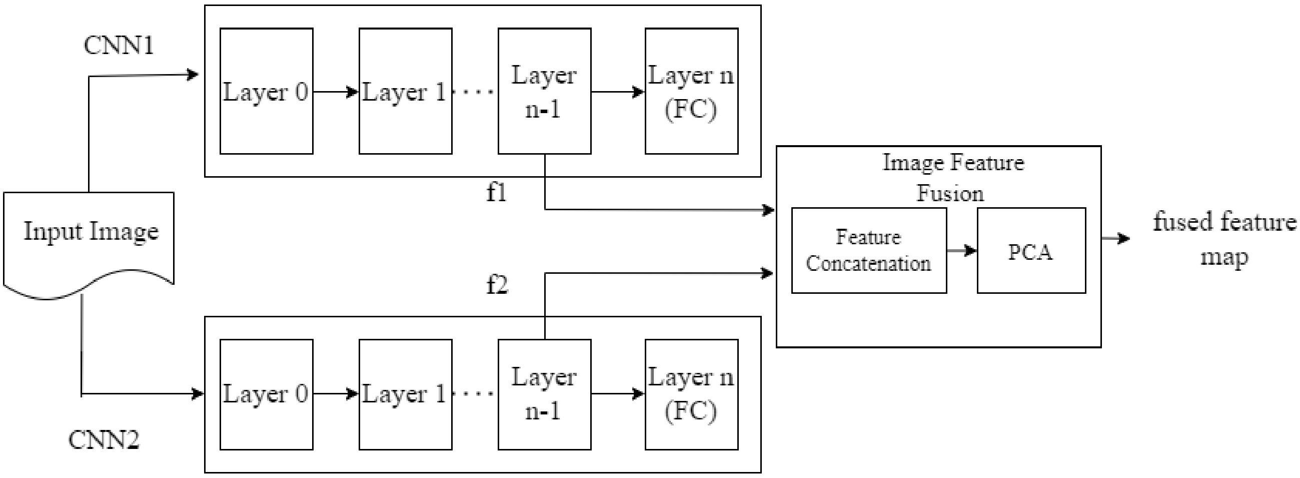

Figure 3: Basic architecture for image feature fusion approach

can be defined as follows [8]

$$D_1 = \sqrt{\sum_{j=1}^{n} (X_{QRI\,j} - X_{DBI\,j})^2} \tag{2}$$

$$D_2 = \sum_{j=1}^{n} |X_{QRI\,j} - X_{DBI\,j}| \tag{3}$$

$$D_3 = 1 - \frac{a}{\sqrt{b}\sqrt{c}} \tag{4}$$

where,

$$a = (x_{QRI} - \overline{x_{QRI}})(x_{DBI} - \overline{x_{DBI}})^T \tag{5}$$

$$b = (x_{QRI} - \overline{x_{QRI}})(x_{QRI} - \overline{x_{QRI}})^T \tag{6}$$

$$c = (x_{DBI} - \overline{x_{DBI}})(x_{DBI} - \overline{x_{DBI}})^T \tag{7}$$

and $\overline{x_{QRI}} = \frac{1}{n}\sum_j X_{QRI\,j}$ and $\overline{x_{DBI}} = \frac{1}{n}\sum_j X_{DBI\,j}$

3.4 Verification and Validation

Precision will be used as evaluation metrics for validation of the results of the proposed work.

In this proposed work precision (P) will be calculated as shown below [8].

$$P_k = \frac{IR_k}{TIR_k} \tag{8}$$

where, P_k is the image retrieval precision for query image k, IR_k is number of relevant images retrieved for query image k and TIR_k is total number of images retrieved for query image k.

$$AP_c = \frac{\sum_{k=1}^{n} P_k}{n} \tag{9}$$

where, AP_c is average precision of class c, P_k is precision for the query image k, ($k \in c$), and n is total number of images in the class c.

The mean average precision (MAP) will be calculated as follows.

$$MAP = \frac{\sum_{c=1}^{m} AP_c}{m} \tag{10}$$

where, MAP is mean average precision of image retrieval, AP_c is average precision of class c, and m is total number of classes in the database.

4. Experimental results and discussions

4.1 Quantitative Results

The AP scores and MAP scores of the CBIR system have been evaluated for different values of the system parameters. Initially, the networks had been finetuned using the corel-1k dataset with a train-test split ratio of 90:10. Due to the finetuning, accuracies on the validation set were increased to 98.00% for Inception-v3 [6] and 97.00% for Efficientnet-b0 [14] and Resnet-50 [6] models. The system was then thoroughly analysed for multitudes of values of system parameters in order to discern the optimal configuration of the system for CBIR. The system parameters used for subsequent analyses are shown in Table 1.

Table 1: CBIR System Parameters used to Generate the Results

S.N.	System Parameter	Domain
1.	Similarity Measure	{euclidean, cityblock, correlation}
2.	Model Combination	{inception-v3, resnet50}, {inception-v3, efficientnetb0}, {resnet50, efficientnetb0}
3.	PCA Components	{64, 250, 900}
4.	Number of retrieval (NR)	{20, 50}

The feature dimensions taken from the second last layer of inception-v3, resnet50 and efficientnetb0 are 2048, 2048 and 1280 respectively and features have been fused accordingly. Table 2, Table 3 and Table 4 are the examples for calculated AP scores for different classes and overall MAP scores with 64 PCA feature dimension with three different similarity measures for model combination of 'inception-v3 and resnet50', 'inception-v3 and efficientnetb0' and 'resnet50 and efficientnetb0' respectively. Similarly, the results for other PCA feature dimensions have been calculated. The experiment has been extended for combination of 'inception-v3 and resnet50 and efficientnetb0' for further analysis. The overall results of MAP scores using the multiple combinations of system parameters are depicted from Figure 4 to Figure 9.

Table 2: APs using inception-v3 and resnet50

Classes	Feature dimension of PCA = 64					
	Top20			Top50		
	D1	D2	D3	D1	D2	D3
beaches	0.81	0.78	0.80	0.78	0.71	0.78
bus	1.00	1.00	1.00	1.00	1.00	1.00
dinosaurs	1.00	1.00	1.00	1.00	0.99	1.00
elephants	1.00	1.00	1.00	1.00	1.00	1.00
flowers	1.00	1.00	1.00	1.00	1.00	1.00
foods	0.90	0.87	0.96	0.89	0.77	0.95
horses	1.00	1.00	1.00	1.00	1.00	1.00
monuments	0.99	0.98	1.00	0.97	0.80	0.98
mountains	1.00	1.00	1.00	0.99	0.97	1.00
people	0.90	0.79	0.90	0.82	0.63	0.85
MAP score	**0.96**	0.94	**0.96**	0.94	0.88	0.95

Table 3: APs using inception-v3 and efficientnetb0

Classes	Feature dimension of PCA = 64					
	Top20			Top50		
	D1	D2	D3	D1	D2	D3
beaches	0.85	0.78	0.81	0.83	0.71	0.78
bus	1.00	1.00	1.00	1.00	1.00	1.00
dinosaurs	1.00	1.00	1.00	1.00	0.95	1.00
elephants	1.00	1.00	1.00	1.00	1.00	1.00
flowers	1.00	1.00	1.00	1.00	0.97	1.00
foods	0.92	0.85	0.96	0.89	0.71	0.93
horses	1.00	1.00	1.00	1.00	0.99	1.00
monuments	0.99	0.98	1.00	0.97	0.73	0.98
mountains	1.00	0.99	1.00	0.99	0.98	0.99
people	0.83	0.76	0.91	0.77	0.53	0.86
MAP score	0.95	0.93	**0.96**	0.94	0.86	0.95

Table 4: APs from resnet50 and efficientnetb0

Classes	Feature dimension of PCA = 64					
	Top20			Top50		
	D1	D2	D3	D1	D2	D3
beaches	0.81	0.78	0.80	0.77	0.71	0.77
bus	1.00	1.00	1.00	1.00	1.00	1.00
dinosaurs	1.00	1.00	1.00	1.00	0.99	1.00
elephants	1.00	1.00	1.00	1.00	1.00	1.00
flowers	1.00	1.00	1.00	1.00	1.00	1.00
foods	0.92	0.91	0.97	0.91	0.81	0.95
horses	1.00	1.00	1.00	1.00	0.99	1.00
monuments	1.00	0.99	1.00	0.97	0.84	0.99
mountains	1.00	1.00	1.00	0.99	0.97	0.99
people	0.91	0.83	0.91	0.83	0.62	0.86
MAP score	**0.96**	0.95	**0.96**	0.94	0.89	0.95

It is found that changing the similarity measures significantly changes the performance. The distance function cityblock didn't produce stable results while changing the system parameters. The euclidean distance function produced good results for different combination of system parameters. Among all distance measures, the overall performance of correlation similarity measure is found to be the best due to it's stable performance for all kind of combinations provided. The model combination used does not significantly affect performance. The overall trend shows increasing performance on decreasing the number of PCA components considered. This is an expected result because since PCA removes the dimensions that carry less information, the similarity measures calculated will better represent the actual similarity between the images.

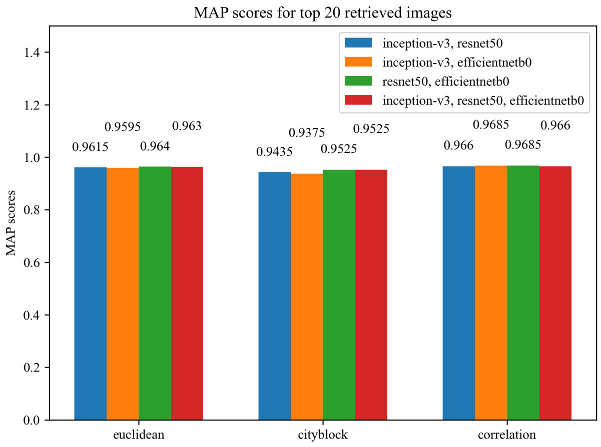

Figure 4: MAP Results for NR=20 & PCA dimension=64

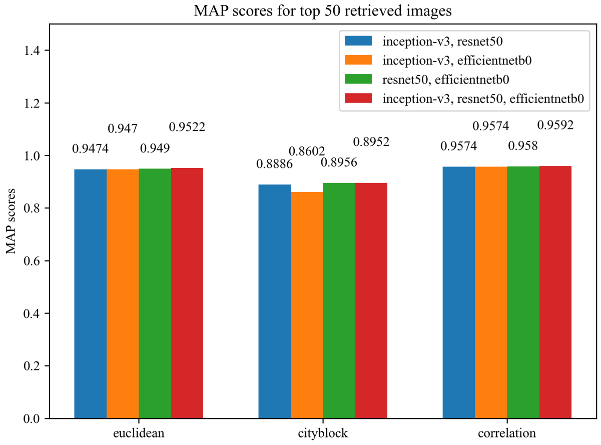

Figure 5: MAP Results for NR=50 & PCA dimension=64

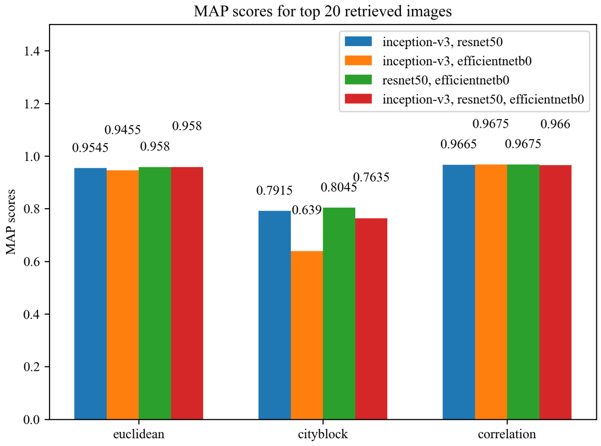

Figure 6: MAP Results for NR=20 & PCA dimension=250

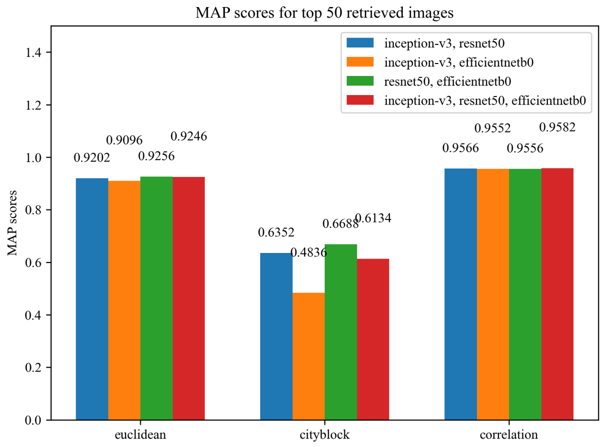

Figure 7: MAP Results for NR=50 & PCA dimension=250

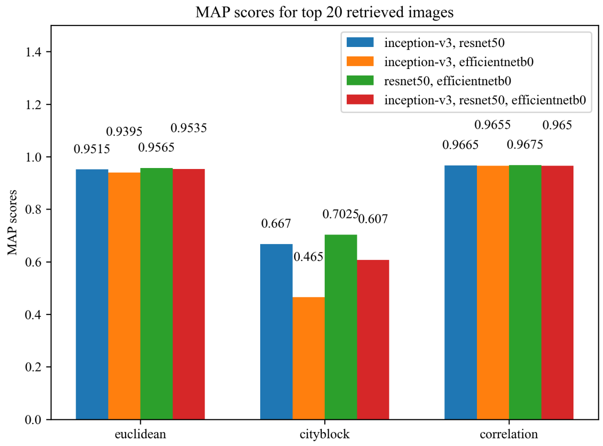

Figure 8: MAP Results for NR=20 & PCA dimension=900

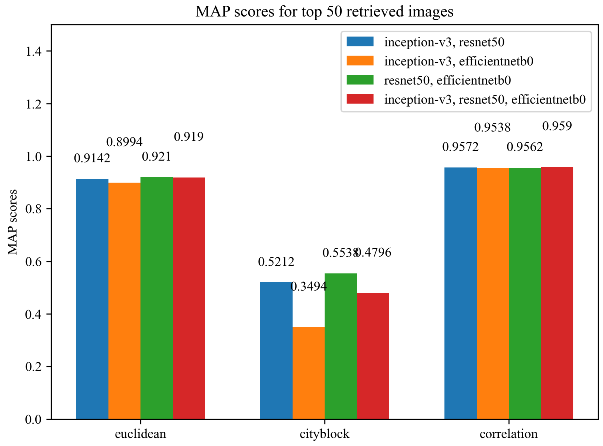

Figure 9: MAP Results for NR=50 & PCA dimension=900

A comparative analysis is done for the proposed work for various combination of provided system parameters in terms of mean average precision. It have been seen that the results for PCA feature dimension 64, number of top retrieved images 20 and similarity measure correlation produced the superior outputs for combination of 'inception-v3 and efficientnetb0' and 'resnet50 and efficientnetb0' considering the Figure 4. Results using different model combinations are not significantly changed and for PCA 250 and PCA 900 the produced results are also acceptable for top 20 retrieved images and correlation distance function. Furthermore, the best MAP score of the proposed method is found to be 96.85% and the result is compared with the results of other different methods for Corel-1K for image retrieval approach. From Table 5 it is found that the proposed method performs better than the previous works.

Table 5: Comparison of MAP scores obtained by proposed methods with other methods for Corel-1K, NR=20

Method	MAP score(%)
Ahmed, K. T. et al.[15]	83.50
Xie, G. et al.[10]	74.05
Jiang, D.[11]	91.40
Jiang, D. et al.[8]	95.43
Proposed method	**96.85**

4.2 Qualitative Result

The qualitative result shown in Figure 11 was generated with feature fusion of *inceptionv3* and *resnet50* models with *900 PCA feature dimension* and *correlation* as the similarity measure for class 'bus' of Corel 1K dataset. Similar results have been obtained for other values of the system parameters as well.

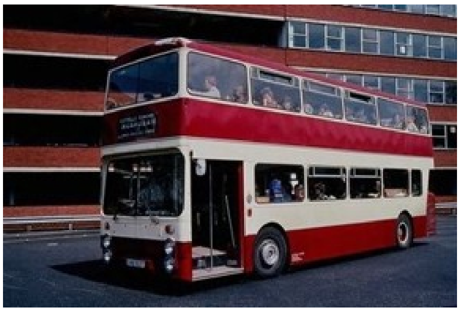

Figure 10: Query Image

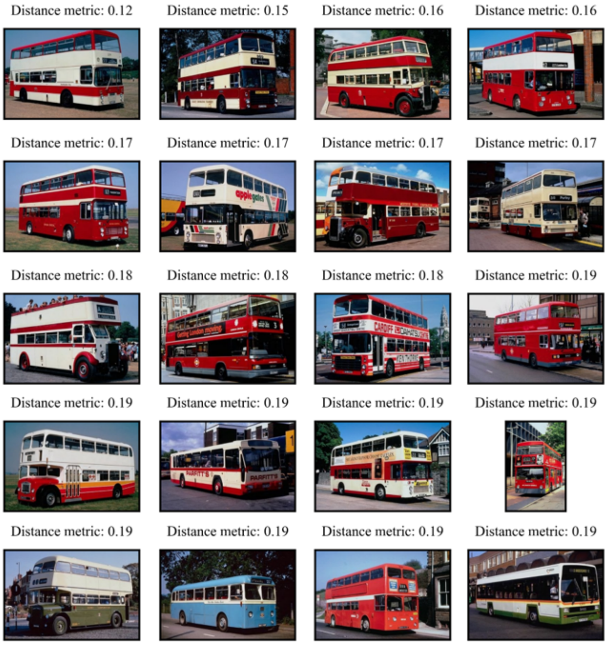

Figure 11: Top 20 Retrieved Images for the Query Image with Corresponding Similarity Measures

5. Conclusion

The proposed method for image retrieval based on the feature fusion and principal component analysis can improve the retrieval capability of the image retrieval system. The overall performance using correlation distance function was found to be the best among the three distance functions. The best results were found to be the mean average precision of scores 96.85% and 95.92% for top 20 and top 50 retrieved images respectively while reducing the feature dimension to value 64. The experimental results in Corel-1K dataset for benchmark top 20 show that the proposed method using concatenation of features extracted from efficientnetb0 with features extracted from either inceptionv3 or resnet50 implementing correlation distance function and with PCA feature dimension 64, outperforms some methods using shallow descriptors for feature extraction purpose and the methods using fusion of low level features extracted from classical approaches. For future research, we will evaluate and analyse the proposed system on a larger dataset for further validation of the method.

References

[1] Mohammed Alkhawlani, Mohammed Elmogy, and Hazem El Bakry. Text-based, content-based, and semantic-based image retrievals: a survey. *Int. J. Comput. Inf. Technol*, 4(01):58–66, 2015.

[2] Shiv Ram Dubey. A decade survey of content based image retrieval using deep learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[3] Rajiv Kapoor, Deepak Sharma, and Tarun Gulati. State of the art content based image retrieval techniques using deep learning: a survey. *Multimedia Tools and Applications*, 80(19):29561–29583, 2021.

[4] Wei Chen, Yu Liu, Weiping Wang, Erwin Bakker, Theodoros Georgiou, Paul Fieguth, Li Liu, and Michael S Lew. Deep learning for instance retrieval: A survey. *arXiv preprint arXiv:2101.11282*, 2021.

[5] Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, et al. Query by image and video content: The qbic system. *computer*, 28(9):23–32, 1995.

[6] Asifullah Khan, Anabia Sohail, Umme Zahoora, and Aqsa Saeed Qureshi. A survey of the recent architectures of deep convolutional neural networks. *Artificial intelligence review*, 53(8):5455–5516, 2020.

[7] Mutasem K Alsmadi. Content-based image retrieval using color, shape and texture descriptors and features. *Arabian Journal for Science and Engineering*, 45(4):3317–3330, 2020.

[8] DaYou Jiang and Jongweon Kim. Image retrieval method based on image feature fusion and discrete cosine transform. *Applied Sciences*, 11(12):5701, 2021.

[9] R Rani Saritha, Varghese Paul, and P Ganesh Kumar. Content based image retrieval using deep learning process. *Cluster Computing*, 22(2):4187–4200, 2019.

[10] Guangyi Xie, Baolong Guo, Zhe Huang, Yan Zheng, and Yunyi Yan. Combination of dominant color descriptor and hu moments in consistent zone for content based image retrieval. *IEEE Access*, 8:146284–146299, 2020.

[11] Dayou Jiang. Image feature fusion and fisher coding based method for cbir. In *2021 International Conference on Communications, Information System and Computer Engineering (CISCE)*, pages 503–508. IEEE, 2021.

[12] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[13] Ahmad S Tarawneh, Ceyhun Celik, Ahmad B Hassanat, and Dmitry Chetverikov. Detailed investigation of deep features with sparse representation and dimensionality reduction in cbir: A comparative study. *Intelligent Data Analysis*, 24(1):47–68, 2020.

[14] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[15] Khawaja Tehseen Ahmed, Shahida Ummesafi, and Amjad Iqbal. Content based image retrieval using image features information fusion. *Information Fusion*, 51:76–99, 2019.