Indonesian Journal of Electrical Engineering and Computer Science
Vol. 31, No. 1, July 2023, pp. 77~87
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v31.i1.pp77-87

Content-based image retrieval using integrated dual deep convolutional neural network

Feroza D. Mirajkar[1], Ruksar Fatima[2], Shaik A. Qadeer[3]
[1]Department of Electronics and Communication Engineering, KBNCE, Kalaburagi, India
[2]Department of Computer Science and Engineering, HOD, KBNCE, Kalaburagi, India
[3]Department of Electrical and Electronics Engineering, MJCET, Hyderabad, India

Article Info

Article history:

Received Sep 22, 2022
Revised Feb 6, 2023
Accepted Feb 18, 2023

Keywords:

Content-based image retrieval
Convolutional neural networks
Image retrieval
Neural networks
Deep learning

ABSTRACT

The image retrieval focuses on finding images that are similar from a dataset that is of a large scale against an image of a query. Earlier, different hand feature descriptor designs are researched based on cues that are visual such as their shape, colour, and texture. used to represent these images. Although, deep learning technologies have widely been applied as an alternative to designing engineering that is dominant for over a decade. The features are automatically learnt through the data. This research work proposes integrated dual deep-convolutional neural networks (IDD-CNN), IDD-CNN comprises two distinctive CNN, first CNN exploits the features and further custom CNN is designed for exploiting the custom features. Moreover, a novel directed graph is designed that comprises the two blocks i.e., learning block and memory block which helps in finding the similarity among images; since this research considers the large dataset, an optimal strategy is introduced for compact features. Moreover, IDD-CNN is evaluated considering the two distinctive benchmark datasets of Paris and the oxford dataset considering metrics; also, image retrieval and re-ranking is carried out against the given query. Comparative analysis of various difficulty levels against the different CNN models suggests that IDD-CNN simply outperforms the existing model.

Corresponding Author:

Feroza D. Mirajkar
Department of Electronics and Communication Engineering, KBNCE
Kalaburagi, India
Email: mmferoza@gmail.com

1. INTRODUCTION

The retrieval of images is thoroughly researched for image matching for which all similar images have been retrieved through the database considering a query image that is given [1], [2]. The images of a database, as well as query, have similarities that are utilized for the images in the database to be ranked according to reducing similarity order [3]. Therefore, the image retrieval performance methodologies are dependent on the computation similarity among the images. The method of similarity computational score between both images must be robust, discriminative as well as efficient. The issue of searching for images that are similar or matched semantically within a huge gallery of images is termed content-based image retrieval aka (CBIR), this is performed by analysis of visual content, while provided with an image query that specifies the needs of the user. This topic has been well researched over a while in the domain of computer vision as well as community multimedia [1], [2]. Presently, the data count of videos as well as images increase exponentially, the information systems have been developed to effectively manage these huge collections of images of high importance along with searching images is an indispensable method. Therefore, the CBIR applications have extreme potential that includes remote sensing [3], re-identification of person [4], an image search for medical domain [5] and online shopping recommendations [6]. The method of CBIR have been widely classified depending on the retrieval level, which includes the levels of both categories as well for instance. Considering the instance level of image retrieval, a specific object of an image for query or a scene such as the Eiffel Tower has been given which aims to find the images that have similar scenes or objects, which could have varying conditions [7]-[10]. Figure 1 presents the state-of-art technique replaced by convolutional neural network (CNN). For feature engineering, which is before the deep learning era, this field was more dominant by feature descriptors that are hand-engineered that include features of scale invariance transform (SIFT) [11].

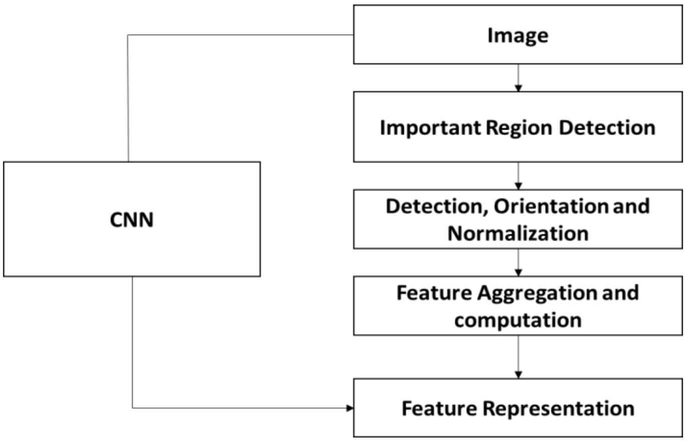


Image
Important Region Detection
Detection, Orientation and Normalization
Feature Aggregation and computation
Feature Representation
CNN

Figure 1. State-of-art technique replaced by CNN

During the stage of feature learning, which is the era of deep learning, it initializes with artificial neural networks, specifically deep CNN as well as ImageNet. Ever since deep learning has had an enormous impact on various fields of research, deep CNN is capable of learning powerful feature representations in which the data can be abstracted at various levels. The techniques of deep learning have attracted a huge amount of attention and had breakthroughs for many tasks of computer vision that include classification of images [11], detection of objects as well as retrieval of images [10]. Although, some major challenges arise concerning: i) semantic gap reduction, ii) improvising scalability of retrieval, and iii) balance of efficiency as well as the accuracy of retrieval. Moreover, above discussed challenges possesses major issues in relevant image retrieval; thus there is a requirement for efficient CBIR. Furthermore, a developed model should be able to retrieve images based on their relevancy; moreover, the existing deep learning-based mechanism as discussed in related work requires a huge number of data for training; also, it fails to rerank the image based on their relevancy. Thus, this research proposes integrated dual deep CNN (IDD-CNN) for CBIR; further contribution of research work is given as follows: i) this research work develops a dual deep CNN for CBIR; moreover, dual deep CNN integrates two distinctive CNN. First CNN extracts the deep features and this is given to the second CNN that holds the characteristics of exploiting the semantic features concerning the dataset. ii) Furthermore, the novel directed graph is designed with two distinctive nodes known as memory node and learning block node; in the case of a large dataset, a novel strategy is introduced for generating the efficient feature. iii) IDD-CNN is evaluated considering the image retrieval and metrics evaluation; image retrieval is carried out based on the given query. Metrics evaluation is carried out on the two distinctive datasets i.e., Oxford and Paris datasets.

2. RELATED WORK

Over the years, a huge number of visual contents are generated as well as shared from different domains that include platforms of social media, images of a medical domain as well as robotics. This has created new limitations and challenges that arise. Specifically, similar contents are searched through databases that are CBIR, this is a field of research that has been well established and is more accurate as well as efficient techniques that are required for retrieval in real-time. CBIR has been rapidly expanding in artificial intelligence and has improvised the intelligent search process. In this section consisting of the survey, the recently researched and experimented works of CBIR are organized as well as reviews, which have been developed based on techniques, and algorithms in deep learning that are given in the recent papers.

Normally, a completely connected layer for feature extraction is studied. Considering dimensionality reduction of principal components analysis (PCA) along with normalization [11] shows the measurement of the similarity among images. Only a completely connected layer has limitations towards the total accuracy of retrieval [12], this combines the features of various completely connected layers indicating the direct connection being made between the initial layers that is fully connected as well as the last layer that results in improvements. It has been observed that a completely connected layer has a receptive field that is global for which every neuron is connected to all other neurons in the prior layer. This characteristic gives rise to two challenges to image retrieval, which is, a lack of invariance of local geometry and spatial information. Feature fusion of various layers focuses on the combination of multiple properties of features with the extraction of the feature. The fusion of different completely connected layers is a possibility in the deep network. Considering [13], we explore various methodologies that combine different activations of various completely connected layers and start the best strategy performed for Pi-fusion that extends features along with various weights for balancing, that also built differently completely layers parallelly that exists on the backbone of ResNet. After which it is concatenated with global features by the layers that obtain the combination of these global features.

The global features are those features that are from the completely connected layers and the local features are the features that arise from the convolutional layers, these features are complementary to one another while their similarity index is being measured and to a little extent, guarantees the performance of retrieval. These two types of features could be directly combined [14], [15]. Before concatenation, the features of convolution that are mapped have to be filtered while sliding windows or region nets proposal. Methods on the bases of pooling are applied for the fusion of features. An approach of orderless fusion of multilayers (MOF) is proposed that has been inspired by an orderless pooling of multilayers (MOP) [16], [17] utilized for the retrieval of images. A solution is introduced that emphasizes varying multiple instances of a graph (VMIG) for which a constant semantic space is studied to save its query semantics that is diverse [18]. The retrieving task has been formulated with different instances of studying the problems for connecting the diverse features to the modalities. Particularly, a variation autoencoder that is query guided is used for modelling the constant semantic space rather than studying the single point of embedding. The DNA is utilized in the CBIR methodology that is proposed where the images are initially stored in sequences of DNA after which the amino acid that is corresponding to it is extracted; this is utilized as feature vectors [19]. Here, the dimensionality reduction for feature vectors is achieved as well as the required information is preserved. Pradhan *et al.*, [20], an image retrieval system that is supervised weakly termed a class agnostic method is proposed based on CNN. The images in the database have been pre-processed to split the background from the foreground and these foregrounds are stored as clusters. Yang *et al.*, [21], the focus of this paper was to decrease the calculation of the count of data, which is performed on the online stage, and avoid any mismatches that occur by mixing of backgrounds.

Above discussed related work suggests that in image retrieval, designing graph plays an important along in deep learning. However, the existing model fails to exploit the features in the large datasets as described in [22]-[24]. Also, another issue with the existing model [24] is it requires the huge amount of data to train the model. Hence, this research work proposes integrated dual deep CNN, which is discussed in the next section of the research.

3. PROPOSED METHOD

CBIR is a widely used technique for retrieving images from huge and unlabeled image databases. However, users are not satisfied with the traditional information retrieval techniques. Moreover, the emergence of web development and transmission networks and the number of images, which are available to users, continue to grow. Therefore, permanent and considerable digital image production in many areas takes place. Hence, the rapid access to these huge collections of images and the retrieving of a similar image of a given image (query) from this large collection of images presents major challenges and requires efficient techniques. The performance of a CBIR system crucially depends on the feature representation and similarity measurement. Figure 2 shows the proposed model workflow; Input image is given to the first deep-CNN to exploit the features and generated feature is given to the second deep custom CNN that helps in exploiting more features. Moreover, a novel directed graph is designed along with two distinctive blocks i.e., memory block and learning block.

3.1. Custom convolution

Considering the M parameter in F dimensional node features denoted as Z belongs to $T^{m\times F}$; forward propagation of the designed custom convolutional layer is computed as (1).

$$J^{(n+1)} = \sigma\left(F^{-1/2}\, CF^{-1/2} J^{(n)} Y^{(n)}\right) \quad (1)$$

In the (1), $J^{(n)}$ indicates the designed output layer of custom-CNN given as Z as the input which can be given as $J^{(q)} = Z$. Furthermore, C indicates an adjacent matrix for the given graph-structured data. In general, the adjacency matrix with the self-connection is defined through $\mathrm{C} = \mathrm{C} + \mathrm{IdM}$ with IdM as the identity matrix. F Indicates the diagonal degree matrix along with its element $\mathrm{D}_{\mathrm{i,j}} = \sum_{\mathrm{l}} \mathrm{C}_{\mathrm{k,l}}$. Further, $\mathrm{Y}^{(\mathrm{n})}$ indicates the weight matrix in FCN-layer; also σ indicates the activation function that indicates non-linearity in given convolutional layers. Moreover, in (1) can be further fragmented into the (2).

$$\mathrm{J}^{(\mathrm{n}+1)} = \sigma\left(\mathrm{B}^{(\mathrm{n})}\,\mathrm{Y}^{(\mathrm{n})}\right) \tag{2}$$

The (2) forms the similar forward propagation as described in the earlier of FC-layer; furthermore, nth layer output is recognized through a given weight matrix also known as the normalized adjacency matrix R. $\mathrm{J}^{(\mathrm{n})}$ indicates the node feature vector that can be updated with the adjacency node feature along with the matrix weight shown in (3).

$$\mathrm{B}^{(\mathrm{n})} = \mathrm{R}\mathrm{B}^{(\mathrm{n})} \tag{3}$$

R Indicates normalized adjacency matrix and given in the (4).

$$\mathrm{R} = \mathrm{F}^{-1/2}\,\mathrm{C}\mathrm{F}^{-1/2} \tag{4}$$

Moreover, considering the previous work on manifold mapping, this research paper develops an integrated custom-CNN, which tends to learn the novel feature representation and features are updated through the corresponding neighboring node in a given database.

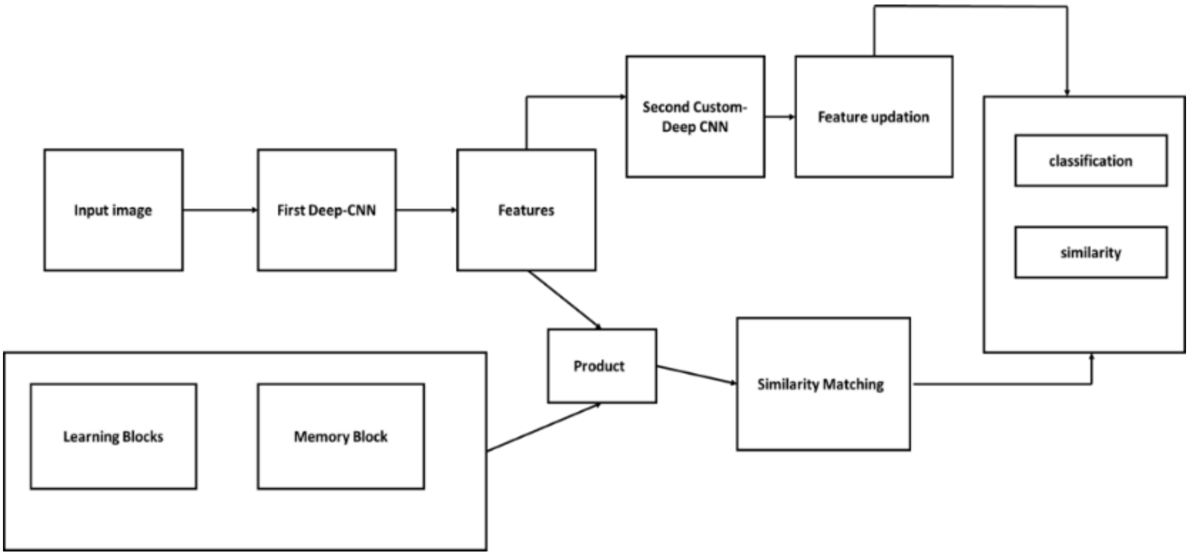


Figure 2. Proposed workflow

3.2. Custom-CNN

The proposed integrated custom CNN combines two different CNN and makes it a single CNN architecture; moreover, this aims to extract the global information for novel feature learning. To model the manifold learning from previous research, this research maps into a designed graph. Moreover, along with the adjacent graph data points, the proposed model learns and updates the feature representations. Furthermore, this is aggregated into novel blocks. The proposed framework takes the mini-batch as an image; at first, it is given in general CNN that exploits the feature representation, this research develops a custom CNN that has three fully connected layers along with five convolutional layers. Moreover, this custom CNN removes the last layer of the FC layer and a novel FC layer is added to minimize the feature dimension. Furthermore, a normalization layer is added to make the output feature. Existing CNN architecture aims to exploit the pairwise relations in mini-batch, which possess the disadvantage of being inconsistent in optimization. Hence, this research introduces a novel directed graph that landmarks the node which monitors the update of mini-batch data. Moreover, this directed graph is designed with two distinctive parameters one is a learning block and another is randomly selected samples. The learning block is designed with the k means algorithm on the designed feature for exploiting the real data distribution. This learning block further constructs semi-blocks that subdivide to save the image feature. Considering the feature of a novel-directed graph as E, learning blocks are designed as E_d belongs to $T^{m\times F}$ along with memory block as E_o belongs to $T^{m\times F}$ which can be further represented as $E = [E_d, E_o]$belongs to $T^{(m+o)\times F}$. Further, a directed graph is constructed based on semantic labels and feature representations; thus, a novel directed graph is constructed as (5).

$$C_n = (EE^v) \odot U \tag{5}$$

The (5) $\odot$ indicates multiplication along with feature dimension as F; further, we design a semantic similarity matrix as given in the (6).

$$U = \begin{cases} 0 & \text{otherwise} \\ 1 & \text{if } f_k \text{ and } f_l \text{ are semantically similar} \end{cases} \tag{6}$$

While training images in mini-batch, features are extracted through CNN and a graph is constructed. Further, the adjacency matrix C_k of given mini-batch is computed through the (7).

$$C_k = (Z_k Z_l^V) \odot U_k \tag{7}$$

In (7), C_k belongs to $T^{p\times F}$ is feature matrix and U_k indicates the similarity matrix.

Moreover, utilizing the manifold learning approach, a novel directed graph and mini-batch graph are fused and designed into another optimal graph. Thus, the fused graph is formulated as given (8).

$$C = \begin{bmatrix} C_k & D_k \\ D_k^V & C_m \end{bmatrix} \tag{8}$$

In (8), $D_k = (Z_k E^V) \odot U_k$ is mini-batch image similarity and directed graph similarity.

3.3. IDD-CNN optimization

To exploit the discriminative feature, the proposed architecture is optimized by minimizing the classification; in the dataset, each image comprises a single level. Moreover, the objective is to reduce the loss and can be formulated as shown in (9).

$$N_e(Y) = -\frac{1}{p}\sum_{j=1}^{P}\sum_{l=1}^{e} 1(A_k = l)\log r\,(A_k = l|q_k, Y) \tag{9}$$

In (9), $r(A_k = l|, Y)$ is considered as the predicted probability of defined classj; also Y indicates the weight matrix of our architecture. q_k belngs to Q indicates the output feature of a given image alongside with a_k as the label. Furthermore, e indicates the class number along with P indicating the training sample. 1(.) indicates the indicator function where 0 indicates the false 1 otherwise. In the case of the dataset, where multiple annotations are created for a particular image, the aim is to reduce the cross-entropy loss and computed as shown in (10).

$$N_e(Y) == -\frac{1}{p}\sum_{k=1}^{P}\sum_{l=1}^{e}\{1(A_k = l)\log r\,(A_k = k|, Y) + [1 - 1(A_k = l)\log[\,r\,(a_k = l|q_k, Y)]\} \tag{10}$$

In (10), $r\left(A_{k,l} = 1\middle|Q_k\ Y\right)$ indicates the predicted label. In case of multiple labels A_k belongs to $\{0,1\}^e$ indicates the label vector. Moreover, the output feature is required to be consistent considering the underlying database; we utilize the fusion graph to achieve the similarity objective and given as shown in (11).

$$N_e(Y) = -\frac{1}{o}\sum_{d=1}^{D}\sum_{k=1}^{p+o+m}\sum_{l=1}^{p+o+m} C^{(d)}\,(k, l) \tag{11}$$

In (11), $C^{(d)}\,(k, l)$ indicates the adjacency matrix of a designed integrated graph is designed mini-batch; thus, whole objective function to optimize the problem is computed as follows along with λ as the hyperparameter. Hyper-parameter is used for controlling the significant feature as shown in (12).

$$N(Y) = N_e(Y) + \lambda N_I(Y) \tag{12}$$

In the case of a large dataset, a novel strategy is introduced where k learning blocks are selected i several times and a novel directed graph is constructed with $(m + o)^i$ nodes. Hence, for a large-scale database, a novel objective function is formulated as given (13).

$$N_k(Y) = \sum_{k=0}^{i-1}\sum_{l=0}^{i-1}\left\|Z'^{(l)V}Z'^{(l)} - K(k,l)\right\|_2^2 \quad (13)$$

In (14), Z' indicates the convolutional layer of designed CNN, also presents the updated feature; (k) indicates the updation in feature. Moreover, $K(k,l)$ indicates the element in an identity matrix which is denoted by $T^{h\times h}$. Moreover, this objective is embedded into the objective function and the final objective function is designed as (14).

$$N(Y) = N_E(Y) + \rho N_K(Y) + \varrho N_k(Y) \quad (14)$$

Moreover, considering the i directed graph, matrices are computed along with mini-batch; computed integrated graphs have a similar parameter in convolutional layers. Furthermore, feature representation updated through various graphs are combined and given the input to the classification layer. In the case of a large dataset, an input image is given to the adjacent learning block in a directed graph and their indices are stored as the custom feature representation. Furthermore, this is represented in Algorithm 1.

Algorithm 1. IDD-CNN

```
Input as memory block with its size $o$, learning block $M$, novel directed graph $i$ and training image set $K_k$
The expected output is learning block $C$ and network $W$
Step1:    Parameter initialization
Step2:    The exploitation of feature generation
Step3:    Initiating the learning block $E_d = [E_d^{(0)}, E_d^{(1)}, \dots\dots, E_d^{(i-1)}$ through random process mechanism
Step4:    Designing of the memory block $E_o = [E_o^{(0)}, E_o^{(1)}, \dots\dots, E_o^{(i-1)}$
Step5:    While iter is less than MIN do
Step6:    Considering the mini-batch as the input along with CNN backbone as the output
Step7:    Designing novel-directed graph C= $[C^{(0)}, C^{(1)}, \dots, C^{(i-1)}]$
Step8:    Considering each directed graph, other CNN outputs the updated feature representations.
Step9:    Network updation with backpropagation with designed objective
Step 10:  Learning block updation with algorithm considering the feature representation
Step11:   Updation of memory block with optimized feature representation
Step12:   End while loop
Step13:   Return optimal network and learning blocks
```

3.4. Image retrieval

Each image in the dataset is given as the input to the proposed architecture and it is summarized to the adjacent block in a directed graph. After optimization, the given image is represented through i integer indices that further cost in bits and computed as $i \times \log_2(m)$. In a given query, at first feature representation is observed with various graphs and similarity index. To retrieve the relevant images along with the query, the distance among the query is computed; distance includes asymmetric and symmetric. Asymmetric computation includes the distance between an original feature of query and the corresponding feature of a given database image, which is computed as shown in (15).

$$f_c(K_s, K_k) = \sum_{k=0}^{i-1}\sum_{l=0}^{i-1}\left\|h_r^{(l)} - e_k^{(l)}\right\|_2^2 \quad (15)$$

In (15), K_k indicates the image in the dataset along with K_s denoting the query image; further e_k denotes the custom feature of database image denoted as K_k. Furthermore, h_r indicates output feature through updating the novel directed graph. Moreover, l subscript indicates the feature from a graph. Asymmetric measurement leads to optimal loss; hence, this research considers asymmetric measurement; after computation of distance among the query and images are ranked.

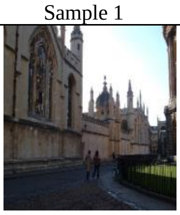
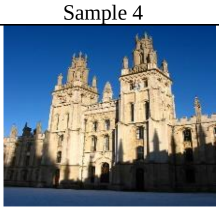
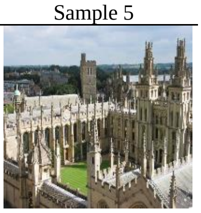
4. PERFORMANCE EVALUATION

CBIR is a system, which uses visual contents to retrieve images from an image database. This system has now become indispensable because it can effectively overcome the problems written above. In CBIR, visual contents are extracted by several techniques: histogram, and segmentation. Also are described by multidimensional feature vectors. The feature vectors and similarity measures mainly influence the retrieval performance of the CBIR system. Always a semantic difference exists between low-level image pixels captured by machines and the high-level semantics perceived by humans. The recent successes of deep learning techniques especially CNN in solving the problem of computer vision applications have inspired us to tackle this issue to improve the performance of CBIR. CNN has provided an attractive solution for CBIR; however, the traditional CNN possesses several disadvantages. Thus, this research work develops IDD-CNN for optimal similarity measurement. This section of the research presents the evaluation of the IDD-CNN with image retrieval and metrics evaluation; also, comparative analysis is carried out with various ResNet models and VGGNet models along with existing models [24].

4.1. Dataset details

We evaluate our method on five image retrieval benchmark datasets namely revisited Oxford5k (ROxford5k), and Revised Paris6k (RParis6k). The ROxford5k and RParis6k datasets are the revisited version of the original Oxford5k and Paris6k datasets [25]. The ROxford5k and RParis6k datasets contain the images of the Oxford and Paris buildings and they both have 70 queries and contain 4,993 and 6,322 images, respectively. We use the mean average precision (mAP) to evaluate the performance of each method. The two datasets have three evaluation setups of different difficulty: Easy, Medium and Hard. We use the medium and hard setups of ROxford5k and RParis6k. Table 1 shows the sample image of the oxford dataset and Paris dataset; the first row shows the five samples of the oxford dataset and the second row shows the five sample images of the oxford dataset.

Table 1. Sample of Roxford and Rparis

Parameter	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Roxford dataset					
Rparis dataset					

4.2. Image retrieval

IDD-CNN is designed not only for image retrieval but also for re-ranking based on a given query. This section of the research evaluates the retrieved image based on a given query. Figure 3 shows the query image from the oxford dataset. Table 2 shows the first 10 images based on the relevancy; this table presents the 10 images of the relevance query. Table 3 shows the first ten images retrieved of a given query from the oxford dataset.

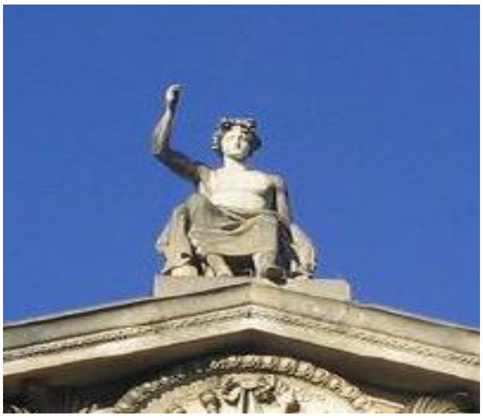


Figure 3. Query image from the oxford dataset

Table 2. Retrieved images

Parameter	Images based on the relevancy	Parameter	Images based on the relevancy
Sample 1		Sample 6	
Sample 2		Sample 7	
Sample 3		Sample 8	
Sample 4		Sample 9	
Sample 5		Sample 10	

Table 3. Retrieved images

	Parameter				
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Images retrieved from the given query					
	Sample 6	Sample 7	Sample 8	Sample 9	Sample 10
Images retrieved from the given query					

4.3. Metrics evaluation and comparison

This section of the research performs a comparative analysis with the existing mechanism; moreover, IDD-CNN is compared through image retrieval re-ranking mechanism query expansion (QE) [26], deep spatial matching (DSM) [27] and collaborative approach [24]. Moreover, a comparison is carried out considering the mAP. The mAP or sometimes simply just referred to as AP is a popular metric used to measure the performance of models doing document/information retrieval and object detection tasks.

$$\mathrm{mAP} = \mathrm{S}\left(\sum_{\mathrm{s}=1}^{\mathrm{S}} \mathrm{average\_precision}(\mathrm{s})\right)^{-1} \tag{16}$$

In (16), S indicates the defined set for queries and $\mathrm{average\_precision}(\mathrm{s})$ indicates the average precision for given querys. mAP Metrics is one of the important metrics where for given query average precision is computed; later mean of all these average precision gives the single number known as mAP that shows the performance of the model at a given query. To prove the IDD-CNN efficiency, CNN architecture ResNet and VggNet and their different variant are considered. Moreover, both architecture along with its variant is given in Table 3. The first column presents the architecture second column presents the mAP value observed. Figure 4 shows the graphical comparison of ResNet architecture with the proposed IDD-CNN model with difficulty levels as medium and hard on Roxf dataset. Figure 5 shows the graphical comparison of various ResNet models considering the difficulty level as medium and hard on the Roxf dataset. Similarly, Figure 6 and Figure 7 shows the comparison of VGGNet model medium and hard level on Roxf dataset and Rparis dataset.

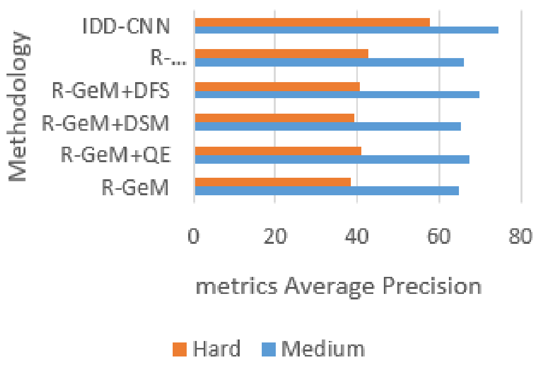


Figure 4. ResNet models comparison considering medium and hard level on Roxf dataset

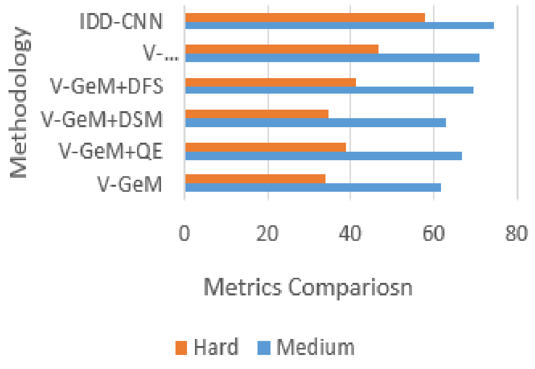


Figure 5. VggNet models comparison considering medium and hard level on Roxf dataset

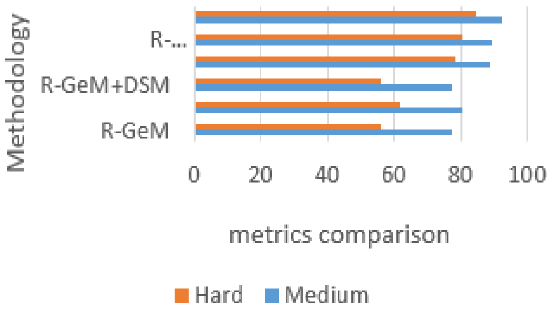


Figure 6. ResNet models comparison considering medium and hard level on Rparis dataset

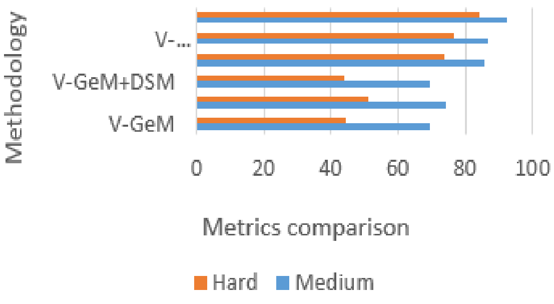


Figure 7. VggNet models comparison considering medium and hard level on Rparis dataset

4.4. Comparative analysis and discussion

This section highlights the performance enhancement over the existing model; it is observed that the most successful model among the existing model as presented in the graph and table is the collaborative model combined with the diffusion and query expansion approach. In the case of difficulty level medium for Roxf dataset; IDD-CNN observes improvisation of 4.59% in terms of mAp value and for Rparis dataset with same difficulty level; IDD-CNN observes improvisation of 6.39%. Similarly, for difficulty level hard, IDD-CNN observes improvisation of 24.08% for the Roxf dataset and 10.43% improvisation is observed concerning the Rparis dataset. Comparative analysis indicates the marginal improvisation not only collaborative approach but another existing model as well.

5. CONCLUSION

Recent research trend in CBIR suggests that CNN based model is driving progress; recent development with the customization of CNN have increased the exploitation of features to measure the similarity matching. This research work proposes IDD-CNN i.e., integrated dual deep CNN to extract the features, First CNN extracts the deep features and this is given to the second CNN that holds the characteristics of exploiting the semantic features concerning the dataset. IDD-CNN is evaluated considering the revisited dataset of oxford and Paris; evaluation is carried out in two stage i.e., image retrieval and metrics evaluation. At first, for the given query, first ten images are retrieved and re-ranked based on their relevancy; Furthermore, metrics are evaluated considering the mean average precision on difficulty level medium and hard. Moreover, in order to prove the model efficiency, IDD-CNN is compared with the CNN based ReSNet and VGGNEt architecture and its different variant. Comparative analysis indicates the IDD-CNN improvises the improvisation of 6.39%. Similarly, for difficulty level hard, IDD-CNN observes improvisation of 24.08% for Roxf dataset and 10.43% improvisation is observed concerning Rparis dataset. Future research direction in CBIR includes exploration of improvised deep learning with three distinctive aspects, which include fast image search, robustness and discriminative ability.

REFERENCES

[1] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-based image retrieval at the end of the early years,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349-1380, 2000, doi: 10.1109/34.895972.

[2] S. Yang, L. Ma, and X. Tan, “Weakly supervised class-agnostic image similarity search based on convolutional neural network,” *in IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 4, pp. 1789-1798, doi: 10.1109/TETC.2022.3157851.

[3] Z. Zeng, Z. Wang, F. Yang, and S. Satoh, “Geo-localization via ground-to-satellite cross-view image retrieval,” in IEEE *Transactions on Multimedia,* 2022, doi: 10.1109/TMM.2022.3144066.

[4] A. P. Byju, B. Demir, and L. Bruzzone, “A progressive content-based image retrieval in JPEG 2000 compressed remote sensing archives,” in *IEEE Transactions on Geoscience and Remote Sensing,* vol. 58, no. 8, pp. 5739-5751, Aug. 2020, doi: 10.1109/TGRS.2020.2969374.

[5] O. E. Dai, B. Demir, B. Sankur, and L. Bruzzone, “A novel system for content-based retrieval of single and multi-label high-dimensional remote sensing images,” in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 11, no. 7, pp. 2473-2490, July 2018, doi: 10.1109/JSTARS.2018.2832985.

[6] J. Xu, C. Wang, C. Qi, C. Shi, and B. Xiao, “Iterative manifold embedding layer learned by incomplete data for large-scale image retrieval,” *IEEE Transactions Multimedia,* vol. 21, no. 6, pp. 1551-1562, 2018, doi: 10.1109/TMM.2018.2883860.

[7] A. Raza, H. Dawood, S. Shabbir, R. Mehboob, and A. Banjar, “Correlated primary visual texton histogram features for content base image retrieval,” in *IEEE Access,* vol. 6, pp. 46595-46616, 2018, doi: 10.1109/ACCESS.2018.2866091.

[8] U. Chaudhuri, B. Banerjee, and A. Bhattacharya, “Siamese graph convolutional network for content based remote sensing image retrieval,” *Computer Vision and Image Understanding,* vol. 184, pp. 22-30, 2019, doi: 10.1016/j.cviu.2019.04.004.

[9] L. R. Nair, K. Subramaniam, and G. Prasannavenkatesan, “A review on multiple approaches to medical image retrieval system,” in *Intelligent Computing in Engineering,* vol. 1125, pp. 501-509, 2020, doi: 10.1007/978-981-15-2780-7-55.

[10] C.-Q. Huang, S.-M. Yang, Y. Pan, and H.-J. Lai, “Object-locationaware hashing for multi-label image retrieval via automatic mask learning,” *IEEE Transactions on Image Processing,* vol. 27, no. 9, pp. 4490-4502, 2018, doi: 10.1109/TIP.2018.2839522.

[11] H. Jun, B. Ko, Y. Kim, I. Kim, and J. Kim, “Combination of multiple global descriptors for image retrieval,” *arXiv preprint arXiv:1903.10663,* 2019.

[12] J. Song, Q. Yu, Y.-Z. Song, T. Xiang, and T. M. Hospedales, “Deep spatial-semantic attention for fine-grained sketch-based image retrieval,” *In Proceedings of the IEEE International Conference on Computer Vision,* 2017, pp. 5552-5561, doi: 10.1109/ICCV.2017.592.

[13] D. Yu, Y. Liu, Y. Pang, Z. Li, and H. Li, “A multi-layer deep fusion convolutional neural network for sketch based image retrieval,” *Neurocomputing,* vol. 296, pp. 23-32, 2018, doi: 10.1016/j.neucom.2018.03.031.

[14] W. Yu, K. Yang, H. Yao, X. Sun, and P. Xu, “Exploiting the complementary strengths of multi-layer CNN features for image retrieval,” *Neurocomputing,* vol. 237, pp. 235-241, 2017, doi: 10.1016/j.neucom.2016.12.002.

[15] B. Cao, A. Araujo, and J. Sim, “Unifying deep local and global features for efficient image search,” *In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16,* 2020, pp. 726-743, doi: 10.1007/978-3-030-58565-5_43.

[16] Y. Li, X. Kong, L. Zheng, and Q. Tian, “Exploiting hierarchical activations of neural network for image retrieval,” *In Proceedings of the 24th ACM International Conference on Multimedia,* 2016, pp. 132-136, doi: 10.1145/2964284.2967197.

[17] Y. Gong, L. Wang, R. Guo, and S. Lazebnik, “Multi-scale orderless pooling of deep convolutional activation features,” *In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland,* Springer International Publishing, 2014, pp. 392-407, doi: 10.1007/978-3-319-10584-0-26.

[18] Z. Zhang, L. Wang, Y. Wang, L. Zhou, J. Zhang, and F. Chen, “Dataset-driven unsupervised object discovery for region-based instance image retrieval,” *in IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 45, no. 1, pp. 247-263, 2022, doi: 10.1109/TPAMI.2022.3141433.

[19] Y. Zeng *et al.,* “Keyword-based diverse image retrieval with variational multiple instance graph,” in *IEEE Transactions on Neural Networks and Learning Systems,* 2022, doi: 10.1109/TNNLS.2022.3168431.

[20] J. Pradhan, C. Bhaya, A. K. Pal, and A. Dhuriya, “Content-based image retrieval using DNA transcription and translation,” in *IEEE Transactions on NanoBioscience,* 2022, doi: 10.1109/TNB.2022.3169701.

[21] S. Yang, L. Ma, and X. Tan, “Weakly supervised class-agnostic image similarity search based on convolutional neural network,” *in IEEE Transactions on Emerging Topics in Computing,* vol. 10, no. 4, pp. 1789-1798, 2022, doi: 10.1109/TETC.2022.3157851.

[22] F. Radenovi´c, G. Tolias, and O. Chum, “Fine-tuning cnn image retrieval with no human annotation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 41, no. 7, pp. 1655-1668, 2018, doi: 10.1109/TPAMI.2018.2846566.

[23] O. Sim´eoni, Y. Avrithis, and O. Chum, “Local features and visual words emerge in activations,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2019, doi: 10.1109/CVPR.2019.01192.

[24] J. Ouyang, W. Zhou, M. Wang, Q. Tian, and H. Li, “Collaborative image relevance learning for visual re-ranking,” in *IEEE Transactions on Multimedia,* vol. 23, pp. 3646-3656, 2021, doi: 10.1109/TMM.2020.3029886.

[25] F. Radenović, A. Iscen, G. Tolias, Y. Avrithis, and O. Chum, “Revisiting oxford and paris: Large-scale image retrieval benchmarking,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2018, pp. 5706-5715, doi: 10.1109/CVPR.2018.00598.

[26] W.-J. Li, S. Wang, and W.-C. Kang, “Feature learning based deep supervised hashing with pairwise labels,” *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16),* 2016, pp. 1711-1717.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems,* vol. 60, no. 6, pp. 84-90, 2012.

BIOGRAPHIES OF AUTHORS

Prof. Feroza D. Mirajkar is working as Assistant Professor in Department of E&CE, in Khaja Banda Nawaz College of Engineering, Kalaburagi, Karnataka, India since 2013. Completed Master of Technology from Basaveshwar College of Engineering, Bagalkot, Karnataka, India. She has attended many workshops such as “Innovative research techniques” a National workshop, CUK, Kalaburagi, participitated in a Two-week ISTE STTP on “Pedagogy for Effective use of ICT in Engineering Education” conducted by Indian institute of Institute of Technology Bombay at NK Orchid college of Engineering and Technology, Solapur and many more. She has 12 publications which include both National and international journal and conferences with one IEEE. She can be contacted at email: mmferoza@gmail.com.

Dr. Ruksar Fatima is presently working as Dean Faculty of Engineering and Technology, Khaja Bandanawaz University, Gulbarga with an experience of 18 years as a dedicated, resourceful education professional. She has received 4 awards, Award for Best Scientific publication by VGST in 2018, RURLA award for Distinguished Scientist in 2018, Chairman for IETE Gulbarga sub-center and best senior researcher (Female) by international academic and research excellence awards 2019. She has 31 International publications in reputed journals and technical member and reviewer of many famous international Journals. She can be contacted at email: ruksarf@gmail.com.

Dr. Shaik A. Qadeer is currently working as Professor in the Department of Electrical and Electronics Engineering at MJ College of Engineering, Hyderabad, Telangana since May 2015. He is the active member of the various Professional Societies and awardee for academic excellence during his bachelor and master’s degree studies including 5th rank holder in the University. He is having 18 Scopus indexed publications, 8 Web of Science publications and one Indian patent. The author having 22 years of teaching experience and his interest include industrial automation, cyber physical systems, signal processing and machine learning. He is a member of IEEE. He can be contacted at email: haqbei@gmail.com.