Journal of Computer Science

Original Research Paper

A Semantic Image Retrieval Technique Through Concept Co-occurrence Based Database Organization and DeepLab Segmentation

[1]R. Jayadevan and [2]V.S. Sheeba

[1]Department of Electrical Engineering, Government Engineering College Thrissur, University of Calicut, Kerala, India
[2]Government Engineering College Thrissur, Kerala, India

Article history
Received: 30-10-2019
Revised: 30-12-2019
Accepted: 09-01-2020

Corresponding Author:
R. Jayadevan
Department of Electrical Engineering, Government Engineering College Thrissur, University of Calicut, Kerala, India
Email: rnath.jayadevan@gmail.com

Abstract: In this paper, a semantic image retrieval technique that efficiently depicts users' perspective is proposed. It primarily aims in the representation of contextual diversity of the user through a high level semantic segmentation technique called DeepLab-V3+. An online user interactive step is also included during the retrieval process. The significance of intra-concept variation in image retrieval is clearly presented in this paper. An efficient database organization, which forms the essence of the retrieval methodology, based on concept co-occurrence and inter-concept distance is also proposed. ResNet-101 CNN features extracted from the regions are utilized in classification and retrieval tasks. The simulation results and performance analysis conducted on PASCAL VOC2012 and SUN '09 datasets depict the superiority of the proposed technique over other approaches.

Keywords: Semantic Segmentation, Concept Co-occurrence, Intra-concept Variation, Database Organization, Contextual Diversity, Set Formation, Subset Formation

Introduction

The scope of image retrieval has now become more diversified with its inevitable contribution in various fields like world wide web (www), medical imaging, journalism, advertising, education, entertainment, weather forecasting, forensic and crime prevention, censoring of images and videos etc. Due to the unpredictable and massive growth rate of digital image database, an efficient organization and indexing procedure in the database management is necessary. This will definitely play a crucial role in the development of a smart and robust image retrieval system. The history of image retrieval reveals the significance of providing an appropriate input for the query purpose. This is because it has a great impact on meeting the user satisfaction and preserving the user interest.

The inceptive works in the field of image retrieval employs textual query inputs which later found to be impractical for large databases. This calls for a new approach called Content Based Image Retrieval (CBIR) which focuses on conducting a similarity search in the database solely based on the image contents like color, texture, shape etc. The retrieval techniques based on this approach were being prominent in the late 20th century. Unlike early CBIR (Zhou and Huang, 2000), which relies only on low-level features, retrieval approaches in the recent past aim to perceive semantic concepts in high-level manner (Liua *et al*., 2007). This leads to remarkable contributions in the field of image retrieval during the last two decades. In order to localize the query process onto a specific area in the input image, different image segmentation algorithms were employed in the retrieval works for ROI extraction. Such an approach also helps in imparting more flexibility to the query process. Another contribution was the provision to modify the search process based on an online Relevance Feedback (RF) by the user on intermediate retrieved results. The application of machine learning (supervised and unsupervised) algorithms in the retrieval process helped in reducing the semantic gap significantly. In order to accommodate the user diversity in the retrieval, generation of a navigation pattern for each user was a key concept. The notion of exploiting the tendency of concept co-occurrence especially in natural scene

databases was noteworthy. The representation of images using high level CNN features was a breakthrough in the image retrieval domain.

However, these retrieval approaches failed to accommodate the contextual diversity of a particular user, efficiently handle multiple semantics in an image and capture visual variations within a concept. The contextual diversity refers to the varying query interest of a particular user from time to time. Most often, in the search processes involving natural scene database, the user may be interested in query inputs which carry multiple semantics. In a retrieval task, the exact determination of query concepts shall not put an end to the query process. This is because the internal variations like size, shape, position, view point etc. may cause significant changes in the feature distribution of a typical concept. Such changes can be referred to as Intra-concept Variation (IV). The aforementioned retrieval concepts are detailed in the next section.

The organization for the rest of the paper is as follows. The background study is briefly discussed in the Literature Review section. The Methodology section explains the workflow and algorithm of the proposed technique. Results and Discussion section involves an analysis of the simulation results and performance evaluation of the proposed work. The paper is summarized in the Conclusion and Future Recommendations section where some scopes for developing the research further are also suggested.

Literature Review

The development of two segmentation algorithms in the early 2000, *JSEG* (Deng and Manjunath, 2001) and *Blobworld segmentation* (Carson *et al*., 2002) has created a strong influence to adopt the practice of using ROI as query input rather than the whole image (Feng and Chua, 2003; Jing *et al*., 2003; Liu *et al*., 2004; Shi *et al*., 2004).

An attempt to incorporate the user suggestion online during the retrieval process was made by Guo *et al*. (2002; Mezaris *et al*., 2003; Rui *et al*., 1998). A provision for the *Relevance Feedback (RF)* by the user on intermediate retrieved results was included in it. But the way in which the query proceeds and the final retrieval are directly influenced by the initial retrieved results. A solution to this issue was later proposed by Su *et al*. (2011) and Lu *et al*. (2013). They developed a *navigation pattern* for each user by analyzing the user query log and user feedback history.

The key aspect in handling the *users' contextual diversity* during retrieval lies in providing an appropriate query input which precisely depicts the real semantic concept of the user. It is important to generate relevant information with very low redundancy in the initial retrieval. Hinami *et al*. (2017) has proposed a method in which user has a provision to *interactively draw a bounding box for extracting the ROI*. The extracted ROI (query input) containing single or multiple semantic concepts are then represented using multitask CNN features.

One of the vital attribute which must be considered in the retrieval tasks associated with natural scene database is the *co-occurrence information of concepts*. Linan and Bhanu (2012; 2016) utilized the co-occurrence information of concepts along with semantic visual concept relatedness to perform image annotation and retrieval tasks. The application of co-occurrence information is limited to pairs of concepts and did not extend to higher levels. An efficient organization of the database is considered to be the essence of any image retrieval system. It eases the implementation of retrieval task and has a direct influence in characterizing the performance of the system. The organization of a natural scene database for the retrieval purpose should therefore consider the co-occurrence tendency of concepts.

Recently, the state of the art retrieval works have been focusing on deep learning techniques as it tends to provide a closer human perception by reducing the semantic gap significantly. Ji *et al*. (2014) presented the utilization of deep learning for CBIR. Later, deep convolutional neural network combined with L1 regularization and PRelu activation function has been applied for image retrieval by Wei and Wang (2017). This prevents the problem of overfitting in traditional convolutional networks and improves accuracy. The application of deep learning for region retrieval is presented by several authors (Gordo *et al*, 2016; 2017; Liu *et al*. 2017; Yanti Idaya Aspura and Mohd Noah, 2017) where *CNN features* are used for ROI representation. The relevance and drawbacks of the above mentioned retrieval works are listed out in Table 1 for a quick understanding.

Based on the literature review, the drawbacks which are yet to be tackled in semantic image retrieval are listed below:

1. Requirement of an accurate and automated ROI extraction step to accommodate the users' contextual diversity
2. Lack of an efficient image database organization in order to handle the retrieval involving more than two semantic concepts
3. Retrieval approaches are deprived of an explicit depiction of Intra-concept Variation (IV) in image database

Table 1: Details of major retrieval approaches in the literature

SI No	Relevant retrieval approaches	Advantage	Drawbacks
1	Retrieval using *JSEG* and *Blobworld* segmentation for ROI extraction.	• Provides ROI as query rather than the whole image. • It helps in imparting a closer human perception as the user generally perceives information from an image locally.	• Failed to precisely capture the ROI when the image contains multiple semantics. • These low-level segmentation techniques may lead to either over or under segmentation.
2	*Relevance Feedback (RF)* by the user on intermediate retrieved results.	• Facilitates the refinement of search results based on user interest.	• Initial retrieved results have an impact on the convergence of the process. • Multiple iterations may be required before generating the final retrieved results.
3	Generation of a *navigation pattern* model for each user based on query log.	• Returns initial retrieved results that adhere to user's interest. • Reduces the probability of requiring more number of iterations, especially greater than two. • Aids in capturing the user diversity during a query process to a greater extent	• Didn't address the contextual diversity of a particular user. • Contextual variation of a user interest adversely affects the initial set of retrieved results which in turn influence the convergence of the retrieval process.
4	*Interactive capturing of ROI* to accommodate the *contextual diversity of a user*.	• Efficient capturing of ROI involving multiple semantics. • Contextual variation of user's interest could be addressed using interactive query input	• Noisy information other than the desired concept within the bounding box may lead to irrelevant retrieval. • Tedious effort will be required in interactive ROI extraction for huge database.
5	Utilization of *concept co-occurrence information* between pairs of concepts.	• Facilitates an efficient organization of database; especially natural images. • Provision for expanding the query based on the co-occurrence tendency of concepts.	• Less efficient in handling the retrieval task involving images with more than two concepts.
6	ROI representation using *CNN features*	• Provides an accurate and a high level semantic representation of ROI. • Results in a more accurate retrieval compared to conventional low-level feature representations.	• Failed to incorporate Intra-concept Variation (IV) in images, which is crucial in a highly accurate image retrieval problem.

In this context, the authors have proposed a semantic image retrieval technique which has the following contributions to address the above mentioned issues:

1. A convenient platform for the user to input semantically meaningful ROI queries through an automated high level semantic segmentation technique. This is supported with a one-time online interaction step during the retrieval process in order refine the query process
2. Extending the concept co-occurrence information to higher levels for an efficient database organization and thereby handling retrieval with multiple semantics
3. Incorporation of Intra-concept Variation (IV) in the image retrieval in order to exactly meet the user query perspective. This is achieved by clustering each database concept (node) into sub-nodes

The notion of automation in segmentation avoids the burden of user involvement in the region extraction. Meanwhile the word 'high level' focuses on the high degree of correlation between the region representation and true semantics of the image. The tendency of co-occurrence of concepts can be quantified in terms of the co-occurrence frequency of different concept combinations. The determination of appropriate sub-node for capturing IV will restrict the search space that makes the retrieval process more semantically meaningful and faster. The details of these contributions are discussed in the methodology section.

The performance of the proposed technique is analyzed and compared with other relevant retrieval approaches using a quantitative measure called Mean Average Precision (MAP). The comparison clearly shows that the proposed methodology outperforms the other approaches, while the sample retrieved results underscore its effectiveness in identifying the appropriate concepts.

Methodology

The methodology of the proposed technique consists of various steps like ROI extraction using an accurate high level semantic segmentation, representation of ROI using CNN feature descriptor, efficient database organization etc. The detection of appropriate concepts (nodes) and sub-nodes in the query region, ranking of images in the detected sub-node using CBIR are also included. Figure 1. depicts the workflow of the proposed retrieval technique. The retrieval process is initiated by the extraction of ResNet-101 CNN features (Kaiming *et al*., 2016) of the input query region. This feature descriptor is utilized by a trained classifier for identification of concept nodes. It then returns an initial set of concepts which is expected to be present in the query region. This is followed by an online user interaction step which aims

to modify the initial returned list of concepts. The modification involves the selection of desired concepts and addition of new concepts based on the co-occurrence frequency. The ROI extraction and user interaction steps call for the incorporation of users' contextual diversity in a reliable manner. After identifying the suitable concept combination node, appropriate sub-node is detected in order to capture the IV. Finally the images belonging to the detected sub-node are ranked based on the similarity using CBIR to form the retrieved result. A detailed discussion of the proposed retrieval algorithm is provided later in this section. In order to facilitate the implementation of the above mentioned retrieval approach, an efficient database organization method is proposed in this paper. The various stages involved in it are shown as a block diagram in Fig. 2. A detailed discussion of different steps in database organization is carried out below.

Semantic Image Segmentation

The requirement of high level semantic image segmentation for an accurate and meaningful extraction of ROI is a crucial stage in the proposed work. The segmentation technique employed here is DeepLab-V3+ (Chen *et al*. 2018) which is a recent variant of the semantic segmentation technique called DeepLab (Chen *et al*. 2015). DeepLab performs the semantic segmentation task effectively by employing Deep Convolutional Neural Networks (DCNN). The original architecture of DCNN is modified in order to have feature maps with high spatial resolution and the ability to segment ROIs at multiple scales in a robust manner. The modification is achieved by incorporating atrous convolution and atrous spatial pyramid pooling (Chen *et al*. 2017) respectively. Also the basic form of DCNN fails to localize the ROI boundaries accurately during segmentation.

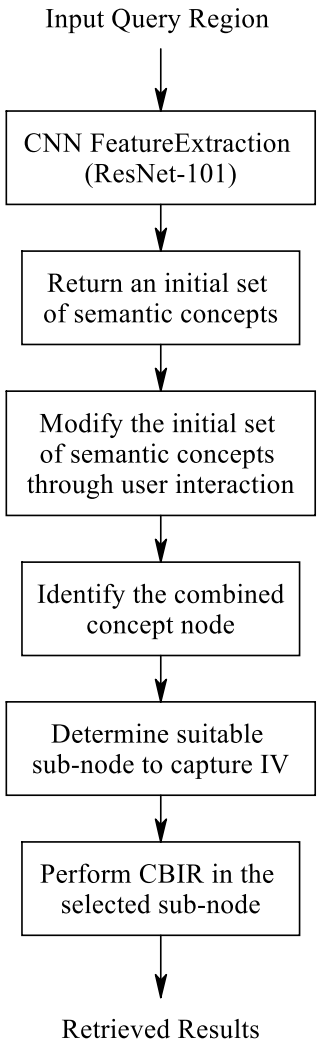


Fig. 1: Workflow of the proposed retrieval technique

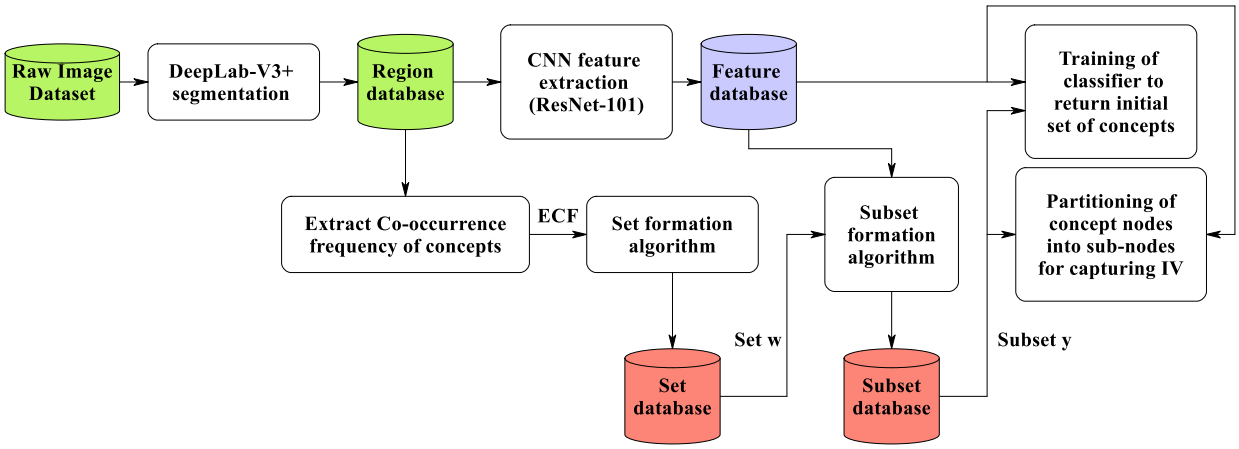


Fig. 2: Block diagram showing database organization

Fig. 3: Semantic segmentation result using DeepLab-V3+ on a sample image from PASCAL VOC2012(a) Original image (b), (c) and (d) Segmented concepts person, sofa and tv respectively, (e) Segmented regions combined

This has been tackled in DeepLab by including a fully connected Conditional Random Field (CRF) so as to combine the final DCNN layer responses with it. DeepLab-V3 (Chen *et al*. 2017) is developed from DeepLab by adding parallel or cascade atrous convolution modules with multiple atrous rates and augmenting the already existing atrous spatial pyramid pooling module. DeepLab-V3+, is an extended version of DeepLab-V3 with the addition of an effective decoder module. It fine tunes the segmentation process by precisely recovering the object boundaries. Figure 3 shows a sample segmentation using DeepLab-V3+.

Co-occurrence Frequency Calculation

The analysis of natural scene image databases reveals that certain concepts have a tendency to co-occur together in different combinations. After the region database has been formed, different concept combinations can be arranged in various levels. For an image with n concepts, there are nC_1, nC_2,…. and nC_n concept combinations in level 1, level 2,…. and level n respectively. This information enables us to calculate the co-occurrence frequency of different concept combinations from the region database. First we define a simple co-occurrence measure called Actual Concept Co-occurrence Frequency (*ACF*). *ACF* at level n for the concept combination, $x = \{c_1,c_2,\ldots c_n\}$ is defined as:

$$ACF_n(x) = \frac{b}{P} \tag{1}$$

where, b denotes the number of images in which the concept combination x occurs, P denotes the total number of images in the database and x is the combination of n concepts $c_1,c_2,\ldots c_n$.

The analysis of *ACF* in different levels for PASCAL VOC2012 and SUN '09 datasets show that the tendency of concept co-occurrence decreases considerably as the level increases. However, in the organization of database, due significance shall be provided for the higher level concept combinations although they are less in number. This leads to the requirement of employing another co-occurrence frequency measure called Effective Concept Co-occurrence Frequency (*ECF*) in the database organization. *ECF* at level n for the concept combination, $x = \{c_1,c_2,\ldots c_n\}$ is defined as:

$$ECF_n(x) = a_n * ACF_n(x) \tag{2}$$

where, a_n is the weight associated with level n and is defined as:

$$a_n = \ln\left(\frac{1}{n-1}\left[\frac{ACF_{avg,n-1}}{ACF_{avg,n}} + \frac{ACF_{avg,n-2}}{ACF_{avg,n}} + \ldots \frac{ACF_{avg,1}}{ACF_{avg,n}}\right]\right) \tag{3}$$

where, $n \geq 2$ and $ACF_{avg,n}$ is the average of the ACF_n values corresponding to all available concept combinations, x, in level n as given in Equation 4. For $n = 1$, $a_n = 1$:

$$ACF_{avg,n} = \frac{1}{NAC_n}\sum_{\forall x} ACF_n(x) \tag{4}$$

where, NAC_n is the total number of available concept combinations in level n. The purpose of using natural

logarithm (ln) in Equation 3 is to scale down the sudden increase of the weighting factor as the level goes up.

Set Formation Algorithm

The set formation is one of the key steps in database organization in which the concepts available from the region database are grouped into different sets based on their *ECF* values. The algorithm groups concepts having higher co-occurrence tendency into one set. Assume that we have N concepts in the database. The aim of the algorithm is to form sets each having K concepts. The steps involved in set formation algorithm are listed in Table 2. In the algorithm, let N_{rc} be the number of remaining concepts, t be the number of concepts in a growing set and w be the index of a growing set at any stage during the set formation.

The parameter R_{wi} is mathematically defined as:

$$R_{wi} = \sum_{n=2}^{t+1} \sum_{x_i} ECF_n(x_i), i = 1,2,\ldots N_{rc} \quad (5)$$

Here x_i at level n is a combination of (n-1) arbitrary concepts of set w and a concept c_i from N_{rc} concepts. $ECF_n(x_i)$ is the effective concept co-occurrence frequency of the concept c_i with the (n-1) arbitrary concepts of set w.

This is then summed over all x_i at level n to get the overall *ECF* of all possible combinations of the concept c_i with (n-1) concepts of set w. R_{wi} is then obtained by adding the overall *ECF* over the levels n = 2 to $n = t + 1$ as given in Equation 5. After forming the set w (Step 6), the pool of remaining N_{rc} concepts is expanded by adding those concepts, which have more tendency to co-occur with the current N_{rc} concepts. These are added from the previous set (w-g), where g = 1 (Step 10). This involves the selection of those r concepts $\{c_j, j = 1,2,\ldots r\}$ from the set (w-g), where g = 1, which satisfy the condition in Equation 6:

$$P_{(w-g),j} \geq \lambda \quad (6)$$

$$P_{(w-g),j} = \frac{1}{NAL_j} \sum_{n=2}^{N_{rc}+1} \left(\frac{1}{NAC_{n,j}} \sum_{x_j} ECF_n(x_j) \right) \quad (7)$$

$$\lambda = \frac{1}{NAL} \sum_{n=2}^{N_{rc}} \left(\frac{1}{NAC_n} \sum_{x} ECF_n(x) \right) \quad (8)$$

where, g = 1 in Equation 6 and 7. If the constraint $N_{rc} \geq K$ is not satisfied even after adding those r concepts, which satisfy Equation 6, to the pool of remaining N_{rc} concepts, then Equation 6 and Equation 7 shall be evaluated for $g > 1$ and the process repeats as given in algorithm. In Equation 7, ECF_n (x_j) is the effective concept co-occurrence frequency of the concept c_j in set (w-g) with the (n-1) arbitrary concepts among the remaining N_{rc} concepts. The overall ECF is determined by adding ECF_n (x_j) over all possible combinations x_j at level n. This value is then divided by $NAC_{n,j}$ which is the total number of possible combinations at level n. The obtained value is then averaged over the levels n = 2 to $n = N_{rc} + 1$ to obtain the score $P_{(w-g),j}$ as given in Eq. 7. Here NAL_j is the number of levels that contain different combinations of the concept c_j with N_{rc} concepts. $P_{(w-g),j}$ is thus the average co-occurrence score of a particular concept c_j in set (w-g) with the remaining N_{rc} concepts. The parameter λ denotes a threshold value which is calculated in Equation 8. In Equation 8, ECF_n (x) is the effective concept co-occurrence frequency of a combination, x, of n arbitrary concepts among the N_{rc} concepts. The overall *ECF* of all such concept combinations at level n is determined and then divided it by NAC_n which is the total number of possible combinations at level n.

Table 2: Steps in set formation algorithm

Step 1:	Start.
Step 2:	Initialize the parameters; N,K.
Step 3:	Assign $N_{rc} = N$.
Step 4:	Initialize the variables; w = 1, g = 1.
Step 5:	Select a combination of q concepts from N_{rc} concepts having highest *ECF* to initiate the formation of set w. Now $t = q$, $N_{rc} = N_{rc}$-t.
Step 6:	If $t < K$, go to Step 7, else go to Step 9 since set w formation is completed now.
Step 7:	Select a concept c_i from N_{rc} concepts having largest R_{wi}. (R_{wi} is defined in Eq. 5)
Step 8:	Add concept c_i to set w and remove c_i from N_{rc}. Now $t = t + 1$, N_{rc}-1. Go to step 6.
Step 9:	$w = w + 1$, g = 1.
Step 10:	Select r concepts, $\{c_j, j = 1,2,\ldots r\}$, which satisfy the constraint in Eq. 6, from set (w-g), where $0 \leq r \leq K$.
Step 11:	Add these r concepts to N_{rc} concepts. Now $N_{rc} = N_{rc} + r$.
Step 12:	If $N_{rc} > K$, go to Step 5, else go to Step 13.
Step 13:	If $N_{rc} = K$, go to Step 16, else go to Step 14.
Step 14:	$g = g + 1$.
Step 15:	If (w-g) = 0 go to Step 16, else go to Step 10.
Step 16:	These N_{rc} concepts form set w (last set).
Step 17:	Stop.

The obtained value is then averaged over the levels $n = 2$ to $n = N_{rc}$ to obtain the parameter λ as given in Equation 8. Here *NAL* is the total number of levels that contain different concept combinations of N_{rc} concepts ($NAL \geq 2$). The parameter λ is also termed as the co-occurrence affinity of remaining N_{rc} concepts among themselves. Each set thus formed contains K concepts except the last set which contains $K_1 = N_{rc} \leq K$ concepts. The value of K_1 depends on the number of concepts added from the previous sets. The selection of the value for the parameter K is explained in Results and Discussion Section.

Subset Formation Algorithm

The input query region might contain multiple concepts whose identification requires a classification task in various levels of co-occurrence. Such a classification process in a set has to be carried out among a large number of concept combination nodes, which greatly increases with the level of co-occurrence. This may adversely affect the classifier performance due to large number of visually similar concept combination nodes. Thus a subset formation algorithm, which categorizes each set into U subsets has been proposed. The subsets are formed in such a way that each subset contains V visually distinct unique concepts. The two parameters involved in the algorithm are defined as follows:

(a) *Cumulative ECF* (*CECF*): *CECF* of an arbitrary group p of V concepts is the sum of *ECF* values of all combinations of concepts from level 2 to V and is given by:

$$CECF_p = \sum_{n=2}^{V} \sum_{x} ECF_{n,p}(x) \tag{9}$$

where, $ECF_{n,p}(x)$ is the *ECF* value of a particular combination x of n concepts in group p.

(b) *Average Cumulative Inter-concept Distance* (*CID_{avg}*): CID_{avg} of an arbitrary group p of V concepts is defined as:

$$CID_{avg,p} = \frac{1}{VC_2}\left[\sum_{i=1}^{V-1}\sum_{j>i}^{V} ID_p(c_i, c_j)\right] \tag{10}$$

This is the mean of the Inter-concept Distance (ID_p) value between every pair of concepts in group p. There are a total of VC_2 pairs of concepts for each group containing V concepts. The Inter-concept Distance (*ID*) between two concepts c_i and c_j is used as an objective measure to visually distinguish them from each other. It is defined as:

$$ID(c_i, c_j) = \frac{1}{2}\left[\begin{array}{l} \frac{1}{N_j}\sum_{k=1}^{N_j} RED(\bar{V}_i, V_{kj}) + \\ \frac{1}{N_i}\sum_{k=1}^{N_i} RED(\bar{V}_j, V_{ki}) \end{array}\right] \tag{11}$$

where, $RED(A,B)$ is the Relative Euclidean Distance (*RED*) between the vectors A and B. It is given by:

$$RED(A,B) = \sqrt{\sum_{k=1}^{m}\left[\left(\frac{A_k}{\sqrt{\sum_{k=1}^{m} A_k^2}}\right) - \left(\frac{B_k}{\sqrt{\sum_{k=1}^{m} B_k^2}}\right)\right]^2} \tag{12}$$

where, $\bar{V}_i$ and $\bar{V}_j$ are the mean feature vectors of concept nodes c_i and c_j respectively. Similarly, N_i and N_j are the corresponding number of regions in c_i and c_j. Here A_k and B_k are kth elements of the vectors A and B which contain m elements.

The steps involved in subset formation algorithm which is applied to each set w containing K concepts are listed in Table 3.

A larger value of α_p and β_p corresponds to more frequently occurring and visually distinct concepts respectively. This facilitates the inclusion of concepts which are visually distinct and frequently occurring together in the same subset, which in turn yields better classification accuracy and retrieval result. Each subset thus formed has a maximum of VC_n concept combination nodes in the nth level of co-occurrence.

Subset formation is followed by training of one-vs-one SVM (using ResNet-101 CNN features) in each level of every subset. This is done in order to facilitate the classification of different concept combination nodes. The determination of values for the parameters U and V are given in Results and Discussion section.

Capturing Intra-concept Variation

The main idea behind the capturing of IV lies in the representation of concept diversities as different sub-nodes. This is achieved by partitioning the concept combination nodes of all the levels in each subset into clusters called sub-nodes. The diversity associated with a concept combination node is quantified in terms of Intra-Concept Visual Variability (*ICV*). *ICV* of a node S can be defined as:

$$ICV(S) = \frac{1}{N_S}\sum_{k=1}^{N_S} RED(\bar{V}_S, V_{ks}) \tag{13}$$

This is the average *RED* between the mean feature vector ($\bar{V}_S$) of node S and all of its region feature vectors V_{ks}. Here N_S is the number of regions which comprises node S. The node S is divided into two sub-nodes S_1 and S_2 if the following two criteria are satisfied:

(a) The total number of regions in that node is greater than 0.5% of the total regions in the database:

(b) $$0.7 * ID(S_1, S_2) \geq \frac{1}{2}\left[ICV(S_1) + ICV(S_2)\right] \tag{14}$$

Table 3: Steps in subset formation algorithm

Step 1:	Start
Step 2:	Input set w and initialize the parameters U and V with suitable values; $p = 1$.
Step 3:	Form $G = KC_V$ groups of V concepts from set w.
Step 4:	Compute $CECF_p$ for group p (Eq. 9).
Step 5:	Compute $CID_{avg,p}$ for group p (Eq. 10).
Step 6:	$p = p + 1$.
Step 7:	If $p > G$, go to Step 8, else go to Step 4.
Step 8:	Normalize the $CECF_p$ value of each group p and indicated as α_p. Normalization is done with respect to the L_2 squared norm of $CECF$ values of all the G groups.
Step 9:	Normalize the $CID_{avg,p}$ value of each group p and indicated as β_p. Normalization is done with respect to the L_2 squared norm of CID_{avg} values of all the G groups.
Step 10:	Compute $\gamma_p = 0.5\alpha_p + 0.5\beta_p$.
Step 11:	Arrange all groups based on the descending order of γ_p.
Step 12:	First U groups form the U subsets corresponding to set w.
Step 13:	Stop

The newly formed sub-nodes are also subjected to partitioning based on the above two criteria. A node or a sub-node will not be partitioned if any one of the above mentioned criterion is violated. The algorithm used for partitioning is hierarchical k-means clustering. The identification of a suitable sub-node during the query process reduces the search space significantly and helps in achieving a faster retrieval. The weight factor in Eq. 14 is determined as 0.7 by trial and error method in such a way that meaningful variations in the concepts are captured.

Retrieval Process

The retrieval process aims in identifying semantically similar images from the organized database in response to a query input region. The various steps involved are discussed below:

(a) Input a query region
(b) ResNet-101 CNN features are extracted from the input region
(c) The extracted feature vector V_R of size 1×2048 is applied as an input to the trained one-vs-one SVM. This results in the classification of the feature vector V_R into appropriate concept combination nodes in each level of every subset
(d) Since all the selected nodes do not lead to the desired search space, an elimination of unwanted nodes is essential at this point. A node S into which the V_R is classified can be eliminated if:

$$DN\left(V_R,\bar{V}_S\right) > \frac{ICV(S) + DN_{\max}\left(\bar{V}_S,V_i\right)}{2} \quad (15)$$

where, $DN\left(V_R,\bar{V}_S\right)$ is the *RED* between the input region feature vector V_R and the mean feature vector $\bar{V}_S$ of the node S; $ICV(S)$ is the intra-concept visual variability of node S; and $DN_{\max}\left(\bar{V}_S,V_i\right)$ is the maximum distance between $\bar{V}_S$ and the farthest feature vector V_i in that node. Thus only those remaining nodes (desired nodes) which have semantic and visual similarity with input query region shall be considered in the further retrieval process

(e) All the unique concepts are identified from the desired nodes and return this initial list of semantic concepts to the user
(f) The user shall select only required concepts form this list through a user interaction step. If the user is not satisfied with the initial returned list, probable concepts can be further added from a supplementary list. The list is prepared by including those concepts which has higher *ECF* values with the initially returned concepts. This interactive step calls for the incorporation of user's high-level semantic perception into the retrieval process. It also limits the search space to only that node which contains the selected concepts. The final list of concepts selected by the user corresponds to a particular node in level Q, where Q is the number of concepts selected by the user. If the selected concept combination is not available in the database, then the search process will be restricted to those concept combination nodes which encompass the maximum possible number of concepts among the selected Q concepts
(g) This step calls for the capturing of IV by determining an appropriate sub-node corresponding to the selected concept combination node in step (f). This is done by classifying the input region vector into an appropriate sub-node of the selected node through hierarchical k-means clustering. This also reduces the search space in the subsequent step to the selected sub-node, which helps in reducing the search time
(h) The last step of the retrieval process is the application of CBIR to the selected sub-node. It involves a ranking of regions associated with the selected sub-node based on the similarity with the input query region. The similarity measure used here is the *RED* between the input region feature vector and region feature vectors in the sub-node. The

images corresponding to the ranked regions are considered as the final retrieved result. It should be noted that the similarity matching is performed in the space where only the desired semantics and contexts are contained

The retrieval performance of the proposed technique is evaluated using a standard measure called Mean Average Precision (MAP). In image retrieval, precision is defined as the fraction of relevant retrieved images. Average Precision (AP) is calculated by averaging the precision values from rank positions where the relevant image was retrieved. MAP is obtained by averaging AP over multiple queries. It is defined as:

$$MAP = \frac{\sum_{l=1}^{L} \mathrm{Pr}_{avg}(l)}{L} \tag{16}$$

where, $Pr_{avg(l)}$ is the AP of lth query, L is the number of queries.

Results and Discussion

This section presents simulation results and performance analysis of the proposed method. A comparative study with other relevant retrieval techniques is also discussed. The simulation and the analysis are performed on PASCAL VOC2012 dataset (16,057 images and 20 concepts) (VOC, 2012) and a database (3,000 images) which consists of 20 concepts selected from SUN'09 dataset. The concepts in PASCAL VOC2012 are Aeroplane, Bicycle, Bird, Boat, Bottle, Bus, Car, Cat, Chair, Cow, Dining Table, Dog, Horse, Motorbike, Person, Potted Plant, Sheep, Sofa, Train and TV. The selected concepts from SUN'09 dataset are Beach, Building, Cat, Desert, Dog, Field, Flower, Ground, Horse, House, Lake, Mountain, Plants, Road, Rock, Sea, Shrubs, Sky, Snowfield and Trees. The former dataset is subjected to DeepLab-V3+ segmentation to yield 31,000 regions while an interactive segmentation (Liu *et al*., 2010) is performed on the latter one to form 6,500 regions. Since the number of regions associated with each concept is less in SUN' 09 dataset, the performance of interactive segmentation is found to be superior to that of DeepLab V3+. This justifies the application of a user interactive segmentation for SUN'09 dataset. Moreover, the accuracy of segmentation has a great impact on the retrieval performance. The details of concept combinations in each co-occurrence level are tabulated in Table 4. The details of sets and subsets formed after database organization are listed in Tables 5 and 6. The value of parameter N is taken as 20 since we have considered 20 concepts from both datasets.

The nodes of each subset in all levels are then subdivided into sub-nodes for capturing intra-concept variation. For example, the concepts bicycle, bus and horse have 5, 3 and 4 sub-nodes respectively for the PASCAL VOC2012 dataset. In SUN'09 dataset, as an example, the concepts sky, ground and road have 3, 2 and 3 sub-nodes respectively. The simulation is done using MATLAB 2018a and Google Co-laboratory which uses python programming.

Selection of Set and Subset Sizes

The size of each set, K, (number of concepts in one set) is fixed to be 12 which is determined by conducting an analysis on the database. We have considered sets of sizes 4, 6, 8, 10, 12, 14, 16, 18 and 20. Even values of K are preferred for the convenience of dividing it further into subsets. For each K, an analysis is performed on the set to determine the percentage of images in which all the concepts of an image belong to the same set. This percentage measure is referred to as Concepts Inclusion Capability of the Set (CICS). The CICS value is plotted for sets having sizes 4, 6, 8, 10, 12, 14, 16, 18 and 20 in Fig. 4. The graph shows that for K greater than or equal to 12, almost all the concepts of an image can be included in one set. This percentage measure is referred to as Concepts Inclusion Capability of the Set (CICS). The CICS value is plotted for sets having sizes 4, 6, 8, 10, 12, 14, 16, 18 and 20 in Fig. 4. The graph shows that for K greater than or equal to 12, almost all the concepts of an image can be included in one set.

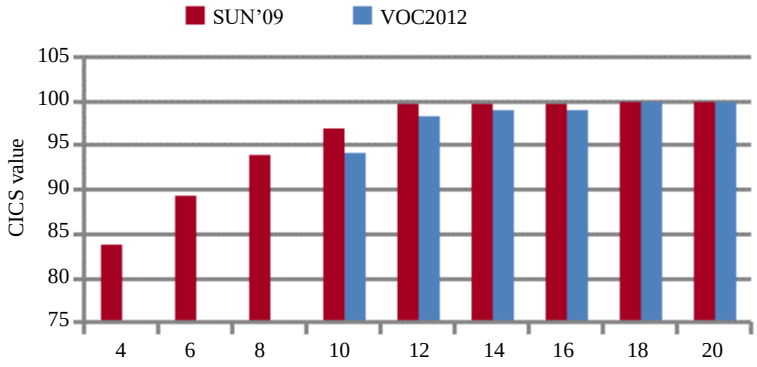


Fig. 4: CICS values for different set sizes

Table 4: Concept combination details in different levels

Levels of co-occurrence	Maximum possible number of concept combinations in each level	Actual number of concept combinations in each level	
		PASCAL VOC2012 dataset	SUN'09 dataset
Level 1	$20C_1$	20	20
Level 2	$20C_2$	180	121
Level 3	$20C_3$	499	207
Level 4	$20C_4$	465	168
Level 5	$20C_5$	159	60
Level 6	$20C_6$	23	7
Level 7	$20C_7$	1	--

Table 5: Sets and subsets in PASCAL VOC2012

	Subset No.		
Set No.	Subset 1	Subset 2	Subset 3
Set 1	Bottle	Cat	Bicycle
	Chair	Dog	Car
	Dining table	Potted plant	Motorbike
	Person	Sofa	TV
Set 2	Bicycle	Bird	Aeroplane
	Bus	Cow	Boat
	Car	Dog	Sheep
	Person	Horse	Train

Table 6: Sets and subsets in SUN '09

	Subset No.		
Set No.	Subset 1	Subset 2	Subset 3
Set 1	Ground	Field	Building
	House	Mountain	Lake
	Sky	Road	Plants
	Trees	Shrubs	Rock
Set 2	Ground	Beach	Cat
	Mountain	Horse	Desert
	Sky	Rock	Dog
	Trees	Sea	Snowfield
Set 3	Ground	Dog	
	Plants	Flower	
	Sky	Plants	
	Trees	Sea	

Table 7: Subset combinations with the highest average k-fold cross-validation classification accuracies of different set sizes in PASCAL VOC2012 dataset

Set size (K)	Subset combination of each set with the highest average k-fold cross-validation accuracy	Cross-validation accuracy
12	3 subsets with 4 concepts each	92.18%
14	4 subsets with 3 concepts each and 1 subset with 2 concepts	87.46%
16	4 subsets with 4 concepts each	85.97%
18	6 subsets with 3 concepts each	82.53%
20	5 subsets with 4 concepts each	79.31%

For PASCAL VOC2012 dataset, the CICS is computed for the sets with K greater than or equal to 10 as the set formation algorithm does not converge for lower values of K. The value of set size (K), number of subsets in each set (U) and subset size (V) are selected by performing a cross-validation classification accuracy analysis. For each value of K, subset combinations of different sizes are formed and an average k-fold (averaged over k = 10, 20, 30) cross-validation classification accuracy is computed for each subset size. The subset combinations with the highest cross-validation accuracy corresponding to each K are tabulated in Tables 7 and 8. This justifies the selection of values K = 12, U = 3, V = 4 for both the datasets. This analysis is considered as a preprocessing step for fixing appropriate values for the parameters K, U and V for a given database. In order to check the robustness of the set and subset formation algorithms, the preprocessing step has been conducted for different dataset sizes of 60%, 70%, 80%, 90% and 100%. It has been noted that the values of the parameters K, U and V and the concepts included in the sets and subsets for a given database remain unchanged for these different dataset sizes.

Performance Analysis

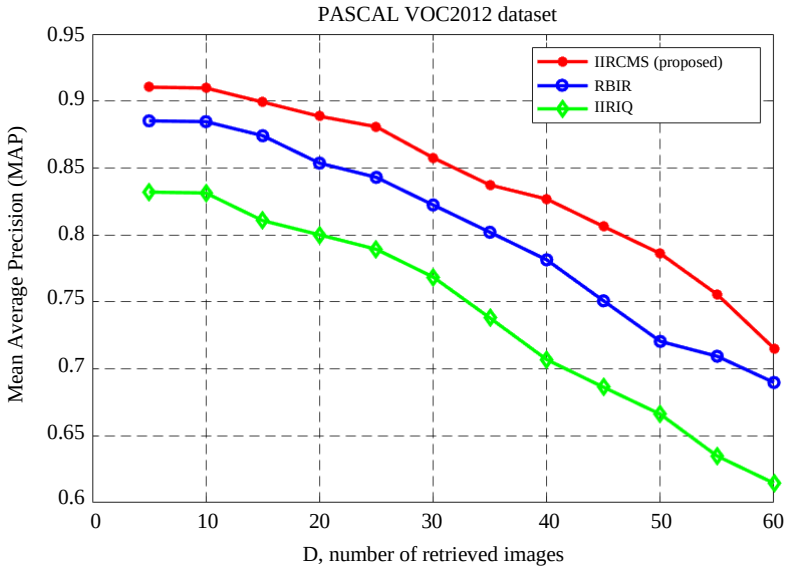
The performance of the proposed method has been analyzed in terms of MAP value of top-D retrieved images with D varying from 10 to 60. The significance of an accurate high level ROI extraction step in the proposed retrieval process is evident from Fig. 5a and 5b. In these figures, the performance of the proposed technique, Interactive Image Retrieval for Capturing Multiple Semantics (IIRCMS), for PASCAL VOC2012 dataset, is compared with that of a state of the art interactive retrieval technique using bounding box query (RBIR, Hinami *et al*. 2017). The comparison is also carried out with a retrieval approach, Interactive Image Retrieval using Image Query (IIRIQ), in which entire image is used as the query input rather than the ROI. Figure 5b also shows the same comparison, but for SUN'09 dataset. In both Fig. 5a and 5b, the proposed method consistently outperforms the other two techniques with a reasonably high MAP value.

Figure 5c and 5d shows the significance of user interaction and intra-concept variation in the proposed retrieval process. This done by comparing IIRCMS

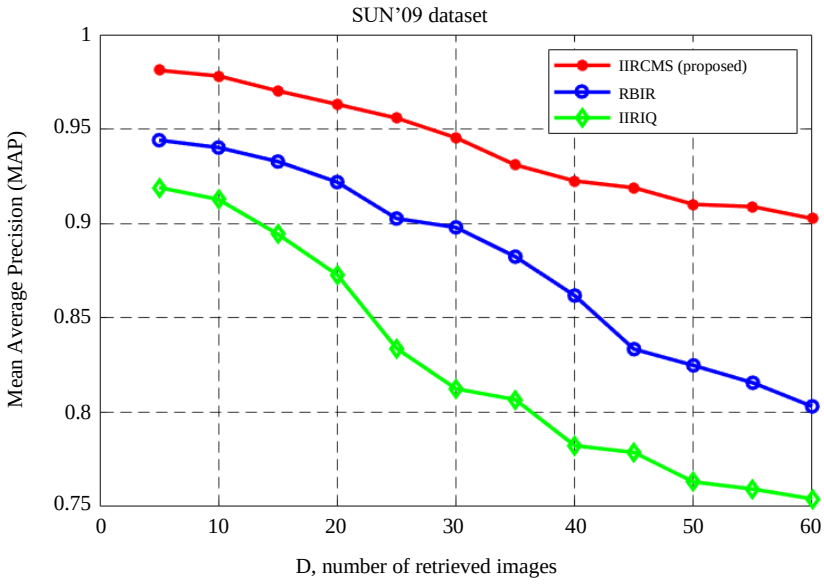
with retrieval approaches without user interaction and without intra-concept variation (without IV), in terms of MAP value. The graph only shows a slight performance improvement in IIRCMS at lower values of D. But the retrieval without IV and without user interaction degrades significantly for higher values of D. This justifies the incorporation of user interaction and IV in IIRCMS.

Table 8: Subset combinations with the highest average k-fold cross-validation classification accuracies of different set sizes in SUN '09 dataset

Set Size (*K*)	Subset combination of each set with the highest average k-fold cross-validation accuracy	Cross-validation accuracy
12	3 subsets with 4 concepts each	98.62%
14	3 subsets with 4 concepts each and 1 subset with 2 concepts	97.91%
16	4 subsets with 4 concepts each	98.06%
18	4 subsets with 4 concepts each and 1 subset with 2 concepts	96.48%
20	5 subsets with 4 concepts each	96.73%



(a)



(b)

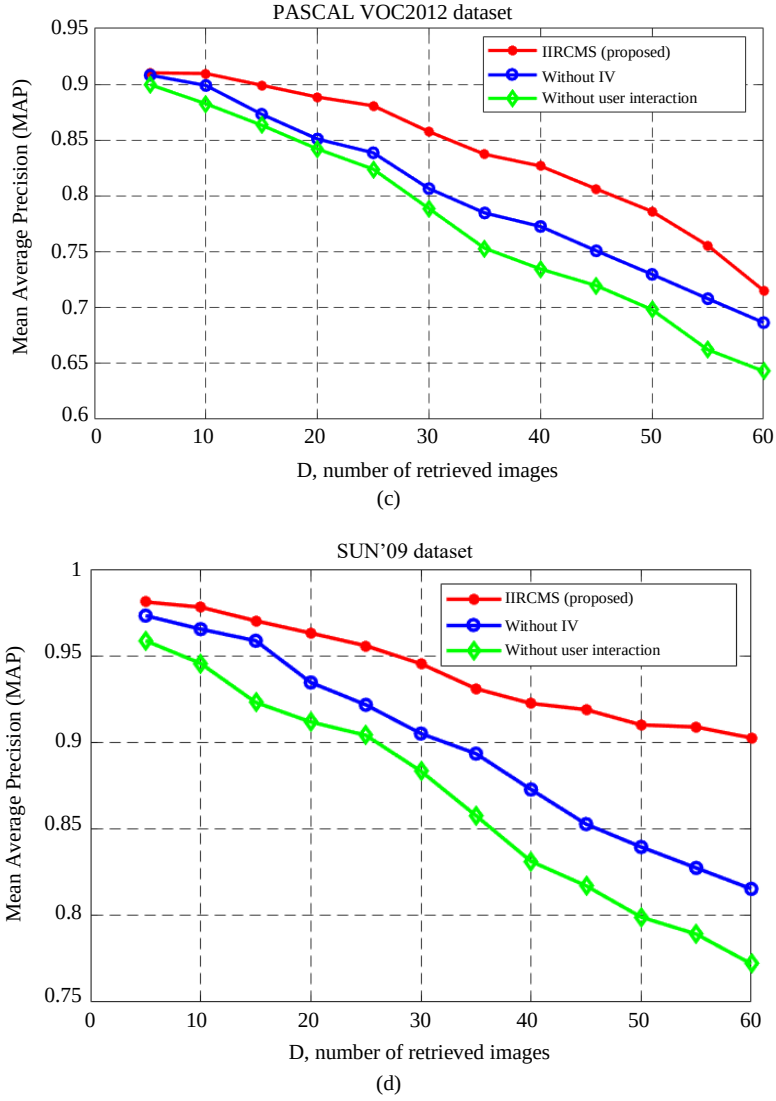


Fig. 5: Performance analysis. (a) and (b): MAP value of top-D retrieved images using IIRCMS, RBIR and IIRIQ techniques for 80% training set sizes in PASCAL VOC2012 and SUN'09 datasets respectively. (c) and (d): MAP value of top-D retrieved images using IIRCMS, retrieval without IV and retrieval without user interaction techniques for 80% training set sizes in PASCAL VOC2012 and SUN'09 datasets respectively

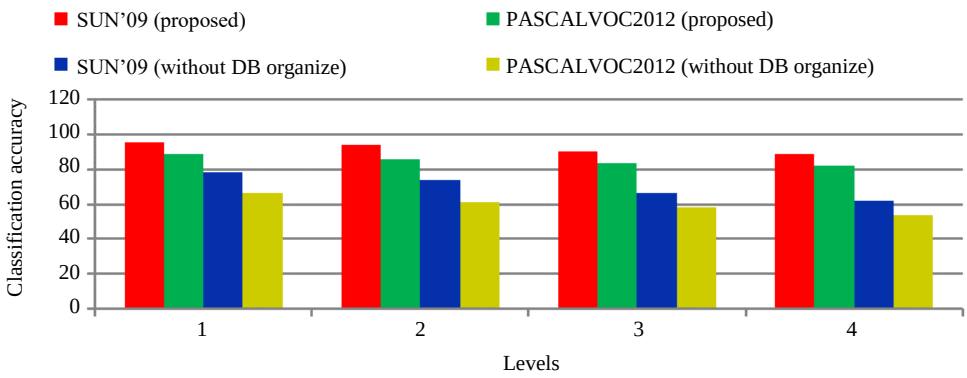


Fig. 6: Classification accuracy at different levels using with and without database organization approaches for 80% training set sizes in PASCAL VOC2012 and SUN'09 datasets

Figure 6 shows the plot of classification accuracy at different levels of co-occurrence (Level 1 to 4) with (proposed) and without database organization (without DB organize) for 80% training set sizes in VOC2012 and SUN'09 datasets. The number of levels is limited to 4 in the analysis, as the number of ROIs for each concept combination is not enough to be considered for classification in levels greater than 4. The comparison in terms of classification accuracy reveals the significance of the proposed contribution; organizing the database into sets and subsets using concept co-occurrence frequency. In the situation, where the database has not been organized into sets and subsets (without DB organize), the classification has to be performed among all the concepts available in an arbitrary level. On the other hand, in the proposed work (IIRCMS), the classification will be among the concepts in an arbitrary subset only. It should be noted that the concepts in a subset have sufficient separation in terms of the distance measure CID_{avg} in order to have a meaningful and accurate classification. Table 9 shows MAP values of top 10 retrieved images for training set sizes of 80%, 70% and 60% in PASCAL VOC2012 and SUN'09 datasets using different retrieval approaches. In this table the proposed work (IIRCMS) is compared to the retrieval approaches with bounding box query (RBIR), with image query (IIRIQ), retrieval without user interaction and retrieval without capturing IV. This observation highlights the significance of the two proposed contributions; (a) Application of an accurate high level semantic segmentation technique and an online user interaction step in order to capture contextual diversity (b) Incorporation of intra-concept variation (IV) in retrieval process.

The proposed retrieval technique (IIRCMS) is also compared with other state of the art retrieval works mentioned in the literature. The comparison has been done with (a) Retrieval utilizing concept co-occurrence information between pairs of concepts (RCCI, Linan and Bhanu 2016), (b) Retrieval technique using bounding box query (RBIR, Hinami *et al.* 2017) and (c) Retrieval using fused deep convolutional features (RFDC, Liu *et al.* 2017). Figure 7 shows the corresponding performance comparison in terms of MAP value of top-10 retrieved images for PASCAL VOC2012 and SUN' 09 datasets.

The proposed work IIRCMS when compared with RCCI has significant improvement. This is because of the fact that the latter utilized the co-occurrence information only for pairs of concepts while the former extended this to higher combination levels. Another factor which makes IIRCMS superior is the usage of CNN features against the middle level features in RCCI. The precise extraction of ROI using DeepLab-V3+ in IIRCMS leads to a more accurate retrieval compared to RBIR which relies on bounding box ROI extraction. From Fig. 7, it is evident that the retrieval performance of IIRCMS is better than that of RFDC. This highlights the incorporation of intra-concept variation (IV) in the proposed work in addition to the semantic ROI representation of concepts using CNN features. Figure 8 displays the top 10 sample retrieved results of IIRCMS (proposed), RBIR and IIRIQ techniques for a query image from PASCAL VOC2012 and SUN' 09 datasets. Here Fig. 8a and 8e are the query images respectively from PASACL VOC2012 and SUN' 09 datasets. In IIRCMS, ROI which is extracted from the query image through segmentation is the final input. The region selected through the bounding box will be the input in RBIR and query image itself makes the input in IIRIQ. The main concepts of the query image in Fig. 8a are person and sofa (among the 20 concepts selected in PASCALVOC 2012 dataset). The concept person is missing in 5th and 6th retrieved images of Fig. 8c (RBIR). In Fig. 8d (IIRIQ), the missing of the same concept happens in 5th retrieved image. Meanwhile in Fig. 8b (IIRCMS), all the 10 retrieved images contain both the concepts person and sofa. In the sample retrieval from SUN '09 dataset (Fig. 8f to 8h)), all the three techniques succeed in identifying the basic semantics in the query as 'house' and 'trees'. But a closer depiction of the query image incorporating the IV is achieved only through the proposed method (IIRCMS). The simulated results shown as example in Fig. 8 underscore the effectiveness of the proposed retrieval algorithm in capturing multiple semantics for indoor and outdoor scene images.

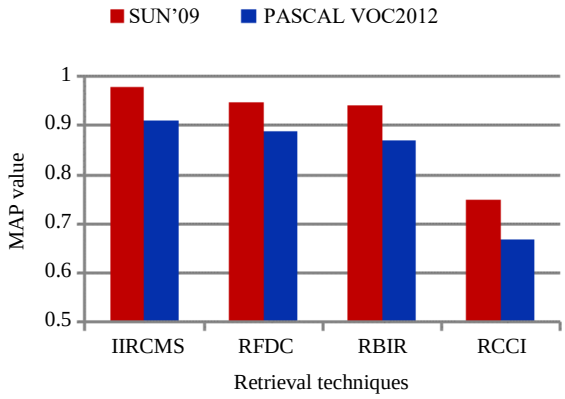


Fig. 7: Performance comparison of the proposed retrieval technique (IIRCMS) with other state of the art techniques in terms of MAP value of top-10 retrieved images for PASCAL VOC2012 and SUN' 09 datasets

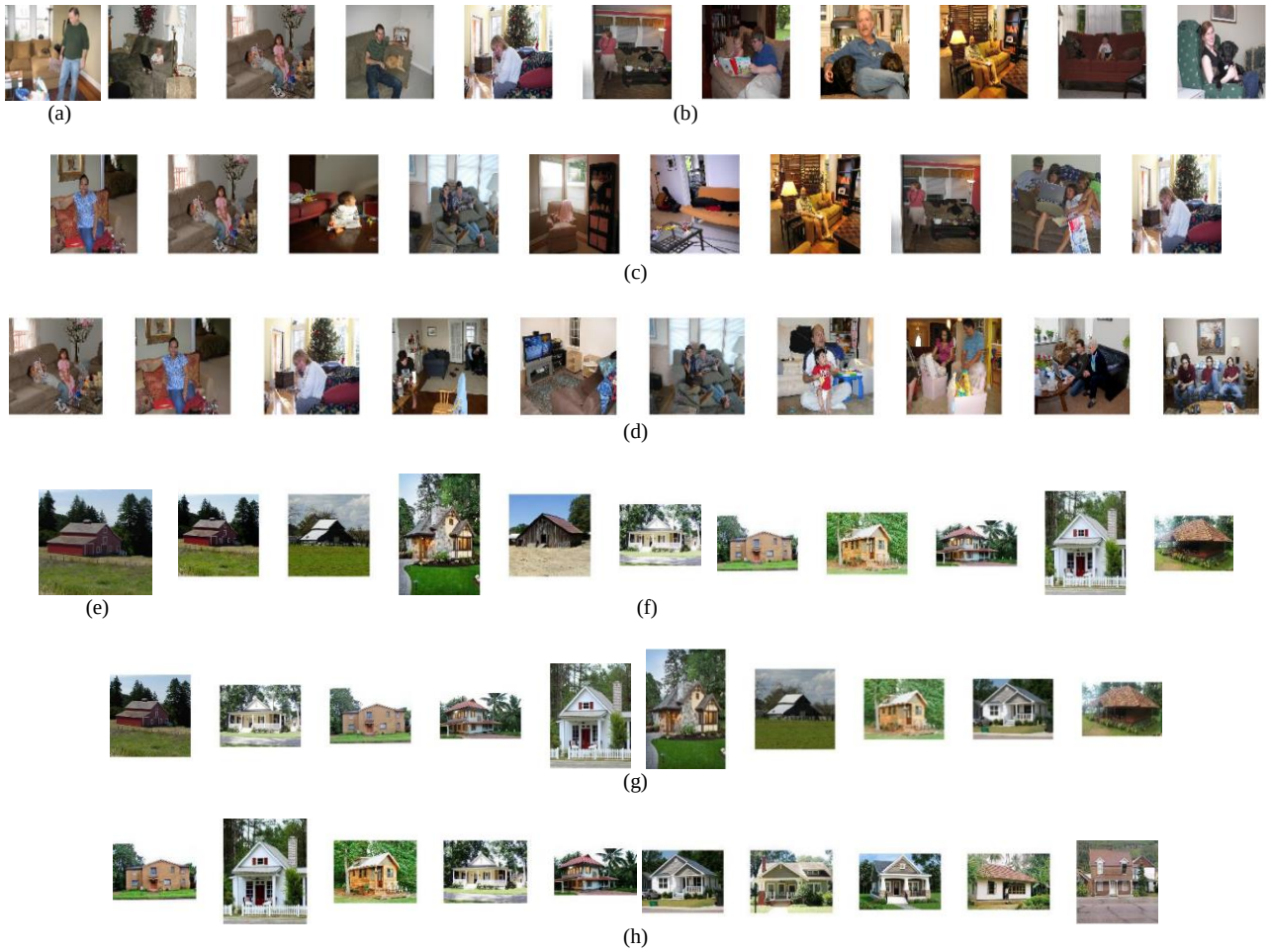


Fig. 8: Sample retrieved results of IIRCMS, RBIR and IIRIQ techniques. (a) and (e) are the query images from PASCAL VOC2012 and SUN'09 datasets respectively. (b), (c) and (d) are the top 10 retrieved images for IIRCMS, RBIR and IIRIQ techniques respectively in PASCAL VOC2012 dataset. (f), (g) and (h) are the top 10 retrieved images for IIRCMS, RBIR and IIRIQ techniques respectively in SUN' 09 dataset

Table 9: Performance analysis of the proposed retrieval technique in terms of MAP value of top 10 retrieved images to highlight the significance of the contributions

	Top-10 MAP score					
	PASCALVOC 2012 Dataset Training set size (%)			SUN'09 Dataset Training set size (%)		
Retrieval approaches	80%	70%	60%	80%	70%	60%
RBIR	0.8846	0.8614	0.8316	0.9402	0.9122	0.8867
IIRIQ	0.8313	0.8089	0.7741	0.9127	0.8840	0.8516
Without IV	0.8990	0.8706	0.8477	0.9654	0.9481	0.9134
Without user interaction	0.8826	0.8647	0.8325	0.9426	0.9312	0.8947
IIRCMS (Proposed)	0.9097	0.8964	0.8735	0.9781	0.9634	0.9340

Conclusion and Future Recommendations

The image retrieval technique proposed in this paper addresses three main challenging aspects during the retrieval. The first one is the contextual diversity in ROI of a particular user, while representation of intra-concept variation in this ROI forms the second concern. The efficient handling of retrieval with multiple semantics is the third aspect. The solutions to these challenges are the major contributions proposed in this paper. The capturing of contextual diversity in ROI selection is facilitated by a meaningful ROI extraction using DeepLab-V3+

segmentation and user interaction during retrieval process. The requirement of appropriate IV representation in the selected ROI is met by subdividing the concept nodes of each subset into sub-nodes. Also an efficient database organization methodology has been proposed by utilizing the concept co-occurrence information in the higher levels of co-occurrence. This in turn facilitates the retrieval of images containing multiple semantics in an intelligible manner. The evaluation of the proposed method under varying test conditions clearly depicts its efficiency and robustness compared to other approaches. The performance has been compared using a popularly known image retrieval measure, the MAP value of top-D retrieved images. The simulation results showing sample retrieved images justifies the inferences drawn from the performance comparisons. The application of the proposed work involves World Wide Web image search engines, law enforcement and surveillance, searches related to defense database, online shopping based on region query etc.

Because of the inclusion of a one-time online user interaction step for the query refinement, this retrieval approach cannot be viewed as a fully automated one. Hence it is recommended in future to adopt a self learning strategy to rectify the search results if necessary and thereby making the entire retrieval process more intelligible. In this paper, the number of concepts in each subset is considered as 4 which is enough for a normal database. But for a database with significant number of concept combinations at higher levels of co-occurrence, more number of concepts are needed in each subset for achieving an efficient retrieval. This may complicate and degrade the classifier performance at higher levels. The development of a suitable solution to handle the increased number of subset concepts can be carried out as another future expansion of the proposed method.

Acknowledgment

We would like to acknowledge the editors of this journal and the three anonymous reviewers for reviewing and evaluating this manuscript and thereby providing suitable feedbacks for the improvement of our research work and the article.

Author's Contributions

R. Jayadevan: Problem identification, methodology development, simulation, testing and validation, writing the manuscript.

V.S. Sheeba: Problem identification, methodology development, proof reading.

Ethics

The research work and the paper are original and contain unpublished material. The corresponding author assures that the co-author has read and approved the article and no ethical issues involved.

References

Carson, C., S. Belongie, H. Greenspan and J. Malik, 2002. Blobworld: Image segmentation using expectation-maximization and its application to image querying. IEEE Trans. Patt. Anal. Machine Intell., 24: 1026-1038. DOI: 10.1109/TPAMI.2002.1023800

Chen, L.C., G. Papandreou, I. Kokkinos, K. Murphy and A.L. Yuille, 2015. Semantic image segmentation with deep convolutional nets and fully connected CRFs. Proc. ICLR. arXiv:1412.7062v4.

Chen, L.C., Y. Zhu, G. Papandreou, F. Schroff and H. Adam, 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European Conference on Computer Vision, (CCV' 18), Springer, Cham, pp: 833-851. DOI: 10.1007/978-3-030-01234-2_49

Chen, L.C., G. Papandreou, F. Schroff and H. Adam, 2017. Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587v3.

Deng, Y. and B.S. Manjunath, 2001. Unsupervised segmentation of color-texture regions in images and video. IEEE Trans. Patt. Anal. Machine Learn., 23: 800-810. DOI: 10.1109/34.946985

Feng, H.M. and T.S. Chua, 2003. A boostrapping approach to annotating large image collection. Proceedings of the Workshop on Multimedia Information Retrieval, Nov. 7-7, ACM, Berkeley, California, pp: 55-62. DOI: 10.1145/973264.973274

Gordo, A., J. Almazan, J. Revaud and D. Larlus, 2016. Deep image retrieval: Learning global representations for image search. Proceedings of the 14th European Conference on Computer Vision, Oct. 11-14, Amsterdam, The Netherlands, pp: 241-257. DOI: 10.1007/978-3-319-46466-4_15

Gordo, A., J. Almazan, J. Revaud and D. Larlus, 2017. End-to-end learning of deep visual representations for image retrieval. Int. J. Comput. Vis., 124: 237-254. DOI: 10.1007/s11263-017-1016-8

Guo, G.D., A.K. Jain, W.Y. Ma and H.J. Zhang, 2002. Learning similarity measure for natural image retrieval with relevance feedback. IEEE Trans. Neural Netw., 13: 811-820. DOI: 10.1109/TNN.2002.1021882

Hinami, R., Y. Matsui and S. Satoh, 2017. Region-based image retrieval revisited. Proceedings of the 25th ACM International Conference on Multimedia, Oct. 23-27, ACM, Mountain View, California, USA, pp: 528-536. DOI: 10.1145/3123266.3123312

Ji, W., D. Wang, S.C.H. Hoi, P. Wu and J. Zhu *et al*., 2014. Deep learning for content-based image retrieval: A comprehensive study. Proceedings of the 22nd ACM International Conference on Multimedia, Nov, 3-7, ACM, Orlando, pp: 157-166. DOI: 10.1145/ 2647868.2654948

Jing, F., L. Mingjing, Z. Lei, Z. Hong-Jiang and Z. Bo, 2003. Learning in region based image retrieval. Proceedings of the 2nd International Conference on Image and Video Retrieval, Jul. 24-25, Urbana-Champaign, IL, USA, pp: 206-215. DOI: 10.1007/3-540-45113-7_21

Kaiming, H., Z. Xiangyu, S. Ren and J. Sun, 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Jun. 27-30, IEEE Xplore Press, Las Vegas, NV, USA, pp: 770-778. DOI: 10.1109/CVPR.2016.90

Linan, F. and B. Bhanu, 2012. Semantic-visual concept relatedness and co-occurrences for image retrieval. Proceedings of the IEEE International Conference on Image Processing, Sept. 30-Oct. 3, IEEE Xplore Press, Orlando, FL, USA. DOI: 10.1109/ICIP.2012.6467388

Linan, F. and B. Bhanu, 2016. Semantic concept co-occurrence patterns for image annotation and retrieval. IEEE Trans. Patt. Anal. Machine Intell., 38: 785-799. DOI: 10.1109/TPAMI.2015.2469281

Liu, D., K. Pulli and L.G. Shapiro, 2010. Fast interactive image segmentation by discriminative clustering. Proceedings of the ACM Multimedia Workshop on Mobile Cloud Media Computing, Oct. 29-29, ACM, Firenze, Italy, pp: 47-52. DOI: 10.1145/1877953.1877967.

Liu, H., B. Li, X. Lv and Y. Huang, 2017. Image retrieval using fused deep convolutional features. Proc. Comput. Sci., 107: 749-754. DOI: 10.1016/ j.procs.2017.03.159

Liu, Y., D. Zhang, G. Lu and W.Y. Ma, 2004. Region-based image retrieval with perceptual colors. Proceedings of the 5th Pacific Rim Conference on Advances in Multimedia Information Processing, Nov. 30-Dec. 3, Tokyo, Japan. DOI: 10.1007/978-3-540-30542-2_115

Liua, Y., D. Zhang and W.Y. Ma, 2007. A survey of content-based image retrieval with high-level semantics. Patt. Recognit., 40: 262-282. DOI: 10.1016/j.patcog.2006.04.045

Lu, H., X. Huang, Y. Lifang and M. Liu, 2013. A novel long-term learning algorithm for relevance feedback in content-based image retrieval. Telecommun. Syst., 54: 265-275. DOI: 10.1007/s11235-013-9732-z

Mezaris, V., I. Kompatsiaris and M.G. Strintzis, 2003. An ontology approach to object-based image retrieval. Proc. ICIP, 2: 511-514. DOI: 10.1109/ICIP.2003.1246729

Rui, Y., T.S. Huang, M. Ortega and S. Mehrotra, 1998. Relevance feedback: A power tool for interactive content-based image retrieval. IEEE Trans. Circuits Video Technol., 8: 644-655. DOI: 10.1109/76.718510

Shi, R., F. Huamin, C. Tat-Seng and L. Chin-Hui, 2004. An adaptive image content representation and segmentation approach to automatic image annotation. Proceedings of the 3rd International Conference Image and Video Retrieval, Jul. 21-23, Dublin, Ireland, pp: 545-554. DOI: 10.1007/978-3-540-27814-6_64

Su, J.H., W.J. Huang, P.S. Yu and V.S. Tseng, 2011. Efficient relevance feedback for content-based image retrieval by mining user navigation pattern. IEEE Trans. Knowl. Data Eng., 23: 360-372. DOI: 10.1109/TKDE.2010.124

VOC, 2012. Pattern Analysis, Statistical Modelling and Computational Learning Visual Object Classes (PASCALVOC).

Wei, Q.J. and W.B. Wang, 2017. Research on image retrieval using deep convolutional neural network combining L1 regularization and PRelu activation function. IOP Conf. Series: Earth Environ. Sci. DOI: 10.1088/1755-1315/69/1/012156

Yanti Idaya Aspura, M.K and S.A. Mohd Noah, 2017. Semantic text-based image retrieval with multi-modality ontology and DBpedia. Electronic Library, 35: 1191-1214. DOI: 10.1108/EL-06-2016-0127

Zhou, X.S. and T.S. Huang, 2000. CBIR: From low-level features to high-level semantics. Proceedings of the SPIE, Image and Video Communication and Processing, (VCP' 00), San Jose, CA, pp: 426-431. DOI: 10.1117/12.382975