Near Duplicate Image Retrieval using Multilevel Local and Global Convolutional Neural Network Features

Tejas Mehta[1]

Computer Engineering Department
Charotar University of Science and Technology
Changa, India

C. K. Bhensdadia[2]

Computer Engineering Department
Dharmsinh Desai University
Nadiad, India

Abstract—In this work, we present an approach based on multilevel local as well as global Convolutional Neural Network (CNN) feature matching to retrieve near duplicate images. CNN features are suitable for visual matching. The CNN features of entire image may not give accuracy in retrieval due to various image editing/capturing operations. Our retrieval task focuses on matching image pairs based on local and global levels. In local matching, an image is segmented into fixed size blocks followed by extracting patches by considering neighboring regions at different levels. Matching local image patches at different levels provides robustness to our retrieval model. In local patch extraction, we select blocks containing SURF feature points instead of selecting all blocks. CNN features are extracted and stored for each image patch and then followed by extraction of global CNN features. Finally, similarity between image pairs is computed by considering all extracted CNN features. Our similarity function is based on correlation and number of blocks found in matching. We implemented our proposed approach on benchmarking Holiday dataset. Retrieval results show remarkable improvement in mean average precision (*mAP*) on the dataset.

Keywords—*Near duplicate image retrieval; local CNN features; global CNN features*

I. INTRODUCTION

Recently with the growth of social media, tremendous amount of multimedia data is uploaded day to day. Majority of the contents are edited or taken from the different camera viewpoint forming the near duplicate content. Storage requirements are increasing rapidly due to duplicate / near duplicate contents. According to [1], near duplicate contents are found in two main sources, Identical near duplicates or non-identical near duplicates. Identical near duplicate images are those which are derived from the same digital source after applying some transformations, including cropping and rescaling, etc. Non identical near duplicate image source are derived from images having same scene or object with change in viewpoint, object occlusions or movement, etc. [2]. Defining near duplicates is a subjective matter. Detection of near duplicates has found many applications including copyright infringements [3], digital forgery [4], fraud detection [5], etc. Retrieving non identical images is found to be difficult. Some of the difficult cases such as different foreground object, severe zooming and change in view point, are shown in Fig. 1.

Objective of identifying near duplicates varies based on applications. In some cases, there is a need to filter out near duplicate contents to obtain novel content and also to reduce down the storage requirements. On the other hand, the objective may be to retrieve all relevant content for a given query.

In this work, our objective is to retrieve all near duplicate images from the set of images for a given query image. In order to handle various cases of matching near duplicate images for retrieval, selection and the way robust features utilized are important tasks. This motivated us to make the use the robust features. Features extracted from Convolutional Neural Network (CNN), [6] are found to be robust. Near duplicates have various cases as stated earlier. Matching full image may fail for the case when only some portion of image gets matched. This inspired us to perform local matching which is achieved through segmenting image into equal sized blocks. This raises the issue of selection of blocks having salient portions of image. To handle the issue of block selection, we utilized location of SURF features as a guiding mechanism. Selection of only local blocks does not provide robustness in matching near duplicate pairs in case of different zooming situations. To overcome this problem, we extract CNN features of current block as well as neighboring regions at different levels. Additionally, matching full image helps us to match overall content. Considering this aspect, we employ extraction of CNN features at global level also. In case when whole image is just a small portion of another image, that is the case of severe zooming, similarity computation is equally important. This motivated us to employ a novel similarity measure which computes correlation based feature similarity along with proportion of image pair matching. In order to decide proportion of matching between image pairs, we consider number of blocks for which matching is successful along with computed value of correlation.

Fig. 1.   Various Difficult Cases of Near Duplicate Image Pairs: Change in Viewpoint (Left), Object Occlusion (Middle), Extreme Zooming (Right).

In order to perform retrieval, many researchers traditionally focus on either local or global features. Our model considers both local and global features. Usually, local CNN matching ignores matching of neighboring regions unlike our approach. Majority of the researches employ either traditional features such as SURF (Speeded up Robust Feature) [7], SIFT (Scale invariant feature transform) [8] or utilize CNN features to carry retrieval task. Limitation of their research is that they obtain low mean average precision value. In our work, we utilize SURF and CNN features to satisfy our different objectives. SURF features are used to detect local points around which we extract multilevel local regions for matching at later stage. CNN features are used for matching at multilevel local patches and also matching at global level such as an entire image. In order to retrieve near duplicate images, we pre-compute local as well as global CNN features and then we match them. We do not utilize SURF (Speeded up Robust Feature) descriptor for matching. Only locations of SURF feature point are used as guiding mechanism for region selection. Matching based on current as well as neighboring regions help us to handle matching at various scales. Using these combined features we obtain high mean average precision. First row of Fig. 2 shows extracted SURF features and their selected blocks of the sample image. Second row of Fig. 2 shows one of the selected blocks along with image patches extracted at various levels.

To summarize, the contribution of our work is as follows:

1) Pre-computed multi level local and global CNN block based matching: We focused on obtaining and storing CNN features of local blocks. This avoids extracting CNN features as and when there is a match, as mentioned in our earlier work [9]. This eliminates unnecessary overhead of repeatedly extracting CNN features from the same image.

2) Improvement in our previous approach [9]: We found that retrieving images based on considering only local image regions may not always give correct results. The proposed technique extracts local CNN features as well as global CNN features from images.

3) Block selection based on local features: In order to decide selection of local image region, we have adopted SURF feature guided region extraction. However, it may not work if no local feature points are found. To overcome this limitation, we use CNN features of a full image.

The paper is divided into different sections. In previous section we gave overall introduction of our approach. Various near duplicate image retrieval techniques are mentioned in Section 2. Our approach is detailed in Section 3 followed by results of implementation of our approach. Finally, we discuss conclusion and future remarks.

II. RELATED WORK

Earlier retrieval systems were based on global features which characterize entire image. A Histogram is a simple global feature for image retrieval. However, such global features do not perform retrieval task effectively and accurately. Later on, researchers found local features to be an effective way for various computer vision tasks and are more robust than global features. In multimedia retrieval, traditional local feature descriptors like Scale invariant feature transform-SIFT [9] and speeded up robust features-SURF [7] are popular. However, these feature descriptors may give false matching. Images may contain regions with less or no local feature descriptors thereby making retrieval process difficult. Our approach performs both local and global matching to handle such a case. PCA-SIFT [10] derived from SIFT descriptor provides more compactness and distinctness representation than SIFT. BOVW (Bag of visual words) [11] is one of the popular approaches in image or video retrieval. Efficiency of BOVW is improved by encoding using Fisher Vector (FV) [12] or Vector of Locally Aggregated Descriptors (VLAD) [13]. Performance of FV is better than VLAD without dimensionality reduction of vector. However, performance of VLAD is improved by performing dimensionality reduction technique (PCA). Furthermore, efficiency of a VLAD based technique is improved by incorporating color feature [14]. In [15], more improvement in efficiency is observed by utilizing both color and gradient in order to create VLAD vectors. In [16], performance of searching visual words is improved significantly by two ways, viz. hamming embedding and weak geometric consistency. Hamming Embedding (HE) generates binary signature while Weak Geometric Consistency (WGC) filters are inconsistent descriptors in terms of angle and scale. Apart from traditional matching techniques, matching can also be performed using graph based techniques by reducing image matching problem into a graph matching problem. In [17], an Attributed Relation Graph (ARG) is constructed followed by computing the similarity between two ARGs to detect image near duplicates.

Recently, Convolution Neural Network (CNN) features have attracted many researchers in the area of computer vision. These features are found robust and efficient in various computer vision applications including image retrieval or classification. High level features can be obtained by activation of fully connected layer of CNN. Such features provide semantic representation of an input image. In [18], vectors generated from each CNN layer are aggregated for retrieval. Application of CNN on a full image may not give better retrieval accuracy. In order to improve performance, CNN features are extracted for different patches with stride of 32 pixels and concatenated [19]. Author in [20] extracted and aggregated CNN features at patch level. In order to perform

image matching, objects can be detected and CNN features are obtained for object level matching [21]. CNN features extracted at local level gives better matching than features extracted at global image level. In [22], Fusion of object, scene and point level CNN is carried out for the purpose of image retrieval. Such Integration of CNN with SIFT gives good retrieval performance. This indicates that CNN and SIFT are not alternative to each other. Although CNN is powerful, it does not always perform better than SIFT. In [23] CNN extracted at various levels are fused with SIFT descriptors. CNN based techniques discussed above perform matching at local levels. Similarity value is obtained by comparing raw image pairs globally in [24]. Above discussion motivates us to explore the use of CNN features at both local and global level.

III. PROPOSED APPROACH

Our work proposed approach retrieves image based on matching multi-level local and global CNN features. Local features are extracted to determine the region of interest. We use locations of SURF features as a guide to extract local region. However, it is not mandatory to use only SURF local features. Any local feature may be used to detect salient region of image. CNN features are extracted not only for local regions but also for surrounding regions. Each image is segmented into equal size blocks and they are numbered in row major order. These block numbers are used in marking to avoid repeated selection of a block. After obtaining SURF features, we extract locations of features under consideration. After that, multilevel patches are extracted for each corresponding block. Then, we extract CNN features for all selected blocks and their multilevel patches. VGG19 [25], a well-known pre-trained CNN model, is used to extract and store CNN features. 4096 dimensional features are extracted by activation of fully connected layer 'fc7' of VGG19. As per requirement of VGG19 Neural network model, each image patch under consideration is resized to 224x224x3 dimension. Detailed procedure of extraction of features and matching is discussed in subsequent sub-sections. Our approach comprises of two phases. First phase is offline processing during which we extract required image features. In second phase, we perform online retrieval using features obtained during the offline phase.

A. Offline Process

Features extraction is an important task in near duplicate image retrieval. In this stage, we extract necessary features and store them for matching in next phase. Image is segmented into fixed-size square blocks which results into blocks present in each row and column of an image. Then, we extract SURF local features for a given image. Co-ordinates of SURF features help us to guide selection of image blocks. Multilevel local patches are extracted for each selected block. To obtain such image patches, we obtain block co-ordinates and

corresponding block numbers. A block number is obtained using equation mentioned in (1) where (x,y), bsize and blocks_per_row represent co-ordinate of SURF feature point, size of block under consideration and total number of blocks available in each row respectively.

$$bno = \left\lceil \frac{y}{bsize} \right\rceil + \left(block\_per\_row * \left\lceil \frac{x}{bsize} \right\rceil - 1 \right) \qquad (1)$$

Extraction of an image patch, comprising of neighboring blocks at different levels helps us to perform block matching under various zooming conditions. Let *(x1,y1)* and *(x2,y2)* be the block co-ordinates for the given block *bno*. Size of patches depend on level of neighboring windows. We consider patches with only up to level 2 neighboring window. Co-ordinates of neighboring window for the given level *l* is obtained using (2) where *s1*, *s2* represent size of image and *l*∈[0,1,2] represents level under consideration. Size of patches at level 0 is same as the size of block.

$$x1_l = max(x1 - bsize * l, 1)$$

$$x2_l = min(x2 + bsize * l, s1)$$

$$y1_l = max(x1 - bsize * l, 1)$$

$$y2_l = min(x2 + bsize * l, s2) \qquad (2)$$

Subsequently, we extract CNN features of current block, level 1 (3×3 neighboring blocks) and level 2 (5×5 neighboring blocks) patches as shown in Fig. 2. Next, we mark all the blocks of level 1 patch. It gives three 4096 dimensional features for the current location. Marking of blocks facilitates block selection process by selecting a block which has not been previously selected. Marking at level 1 patch helps us to select various blocks that are not neighbor of previously marked blocks. This helps in reducing number of patches extracted and their by reduces number of CNN features. We store CNN features of all locally obtained patches for matching during online query processing phase. After having local CNN features, we perform CNN activation on entire image in order to get global CNN feature. The process is repeated for all images. As a result, we obtain set of multi-level local and global CNN features for all images. The entire process is shown in Fig. 3 and detailed algorithm is mentioned in Fig. 4.

B. Online Retrieval Process

In online query processing stage, we retrieve images by computing correlation between all CNN features of a query set and all CNN features of all images. We do pair wise comparison of features. For each image pair, a correlation matrix is computed. Using correlation matrix, we compute our similarity value with the help of (3). Correlation values which are above threshold are considered for computing similarity between image pairs. We use weighted sum of such correlations to compute similarities between image pairs.

Fig. 2. First Row Detected SURF Local Features (Middle), Selected Blocks (Right), Second Row (Left) Sample Selected Block (Level 0), Level 1 Patch (Middle), Level 2 Patch (Right).
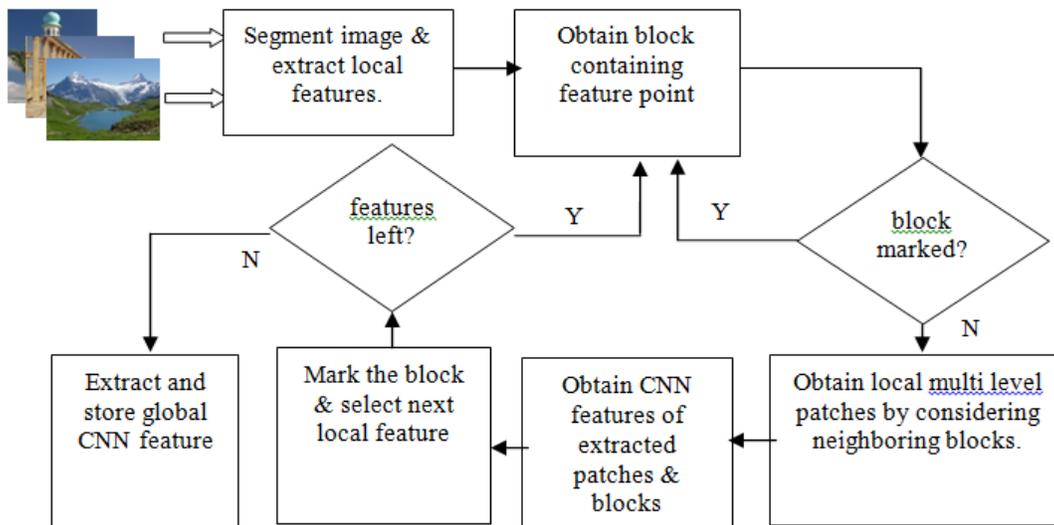

Fig. 3. Offline Extraction of Local and Global CNN Features.

Generate_cnndb(*imageset*)

$t_i \in imageset$

Segment image t_i into fixed size blocks

Obtain local feature points

For each feature points

    Obtain block for the feature point as per (1).

    If block is not marked

        Mark all blocks at level 1

        Extract CNN feature for given block and different level of neighboring window as per (2).

        Store all Extracted CNN features for given image t_i

    End

End

Extract and store global CNN feature for entire image t_i.

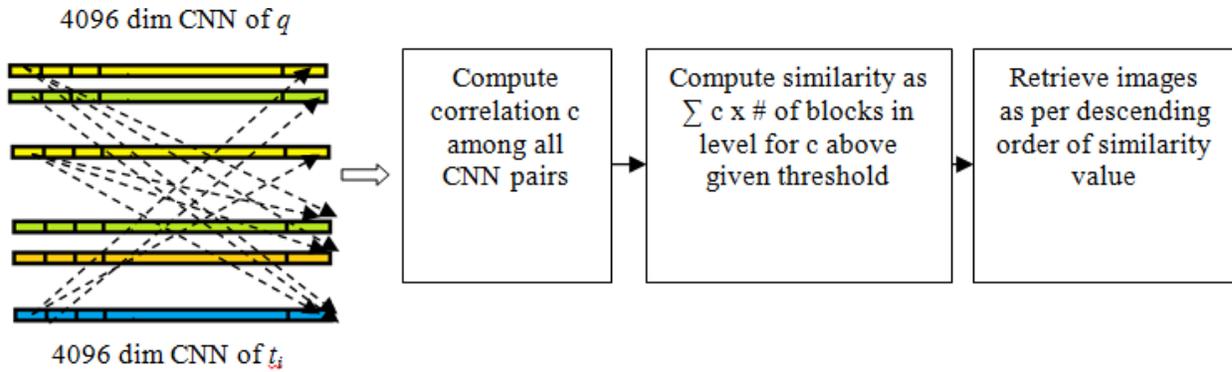Fig. 4. Algorithm to Extract and Store Local and Global CNN Features.

Fig. 5.   Online Retrieval Process.

In order to measure weighted similarity, correlation values along with number of blocks are considered. This results into higher value of similarity if higher correlation value is found with more number of blocks and vice versa. In (3), level l represents image patch with number of blocks available. A higher value of l represents an image patch with more number of neighboring blocks, as shown in Fig. 2. The similarity measure is computed using (3). A query image is compared with all images of an image set using similarity values. Then, images are retrieved based on descending values of similarity. The retrieval process for a given query image q is shown in Fig. 5.

$$Sim(q,ti) = \sum_{i=1}^{n} \sum_{j=1}^{m} c(i,j)x\,(2*l+1)^2$$

$$l \in [0,1,2], \exists c(i,j) \geq th, i \in [1..n], j \in [1..m]$$

$$q \in \{queryset\}, t_i \in \{imageset\} \qquad (3)$$

IV. EXPERIMENTAL RESULT AND DISCUSSION

Experiments are carried out on Holiday benchmarking dataset in MATLAB 2017 with VGG19 neural network toolbox model on TitanXP Nvidia GPU system. Input images are resized to 30% in multiple of block size for Holiday dataset. 56x56 block size is considered in this experiment. Smaller the block size, higher the number of CNN features and vice versa. However, in this experiment we followed the same block size mentioned in [9]. Performance improvement is found due to obtaining CNN for all selected local blocks as well as global (entire image). In order to measure performance of our model, mean average precision (*mAP*) is used and is calculated as shown in (4) where r_i, N and M represents rank of ith retrieved image, number of relevant images and total number of queries respectively.

$$mAP = \frac{\frac{1}{N}\sum_{i=0}^{N} i/r_i}{M} \qquad (4)$$

Results are obtained for Holiday dataset with 500 query images from total of 1491. For online query processing, results are obtained with threshold value set to 0.7 and 0.8. Setting threshold value 0.7 provides better retrieval accuracy than 0.8 threshold value. Setting threshold value high may miss certain good matches. Setting threshold value low may include false positive matches. In Fig. 6, we can see the degradation of retrieval performance of high threshold value that is 0.8. In

offline experimental setup, size of neighboring window is considered up to level 2. Level of neighboring window affects the number of features extracted for each block. Increasing depth of level results into increase in number of features extracted which affects the searching performance significantly.

Fig. 6 shows sample correlation values obtained by computing and matching using global CNN features only. It gives lower correlation values as it is using only global aspect of matching. First row in Fig. 6 represents correctly retrieved images due to our approach of multilevel local matching. Second row of Fig. 6 shows a partial failure case where wrong image is retrieved having nearly same visual content but differs in scene. In such situation our model faces some problems when images are having same structural similarity but in fact represent different scene or context. The performance of searching is affected while measuring similarity as correlations are computed on all 4096 dimensional vectors in brute force manner. However, we improve the search efficiency by performing pre-computation of features and their by avoiding repeated computations in every images as was the case in the model presented in [9]. Sample retrieved images for the given query image are shown in Fig. 7.

Fig. 8 compares performance of our approach with various state of art near duplicate image retrieval techniques. In [26], noticeable comparative analysis of Bag of words, Fisher vectors and VLAD representations with and without dimensionality reduction are mentioned. In general, standard fisher vector seems to give better performance than standard VLAD vectors which has scope for improvements in VLAD encoding [27]. Irrespective of various encoding mechanisms presented in [26][27][28][14][15], the mean average precision of our model is found better. Even the performance of triangular embedding [29] with descriptor of size 8024 is found to be lower than ours which has dimension size 4096. We achieve better retrieval performance as all techniques mentioned above make use of traditional features which are less robust than CNN features, one of the features that we use in our approach.

Author in [24] proposes a CNN based technique and computes global similarity between image pairs. Computing global image pair similarity significantly reduces performance. In [30], a global matching approach is presented based on using retraining and rotation on dataset. In spite of not

performing such costly operations, our approach give the better result than that technique. In [31], an approach is presented by aggregating local descriptors and CNN. However, its mean average precision is less than our approach. Our current model has found significant improvement of around 5% compared to our previous work reported in [9]. Our model outperforms

existing CNN based techniques [24][30][31][9] with the parameter mean average precision. The improvement in performance in our approach is mainly contributed to use of multilevel local CNN matching, global CNN matching and computation of similarity measure in terms of correlation and matching proportion.

0.4585   Correct sample retrieval with correlation threshold greater than 0.7

0.5634   Sample retrieval with correlation threshold greater than 0.7 with incorrect image (middle)
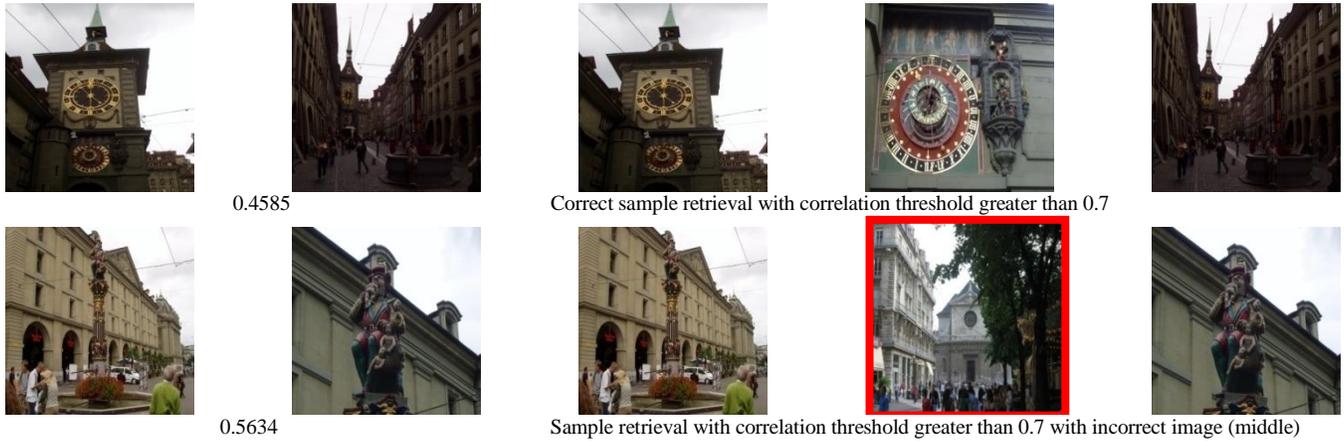
Fig. 6.   Correlation Value Found using Global CNN Only (First Two Column), ours Retrieval (Columns 3-5).
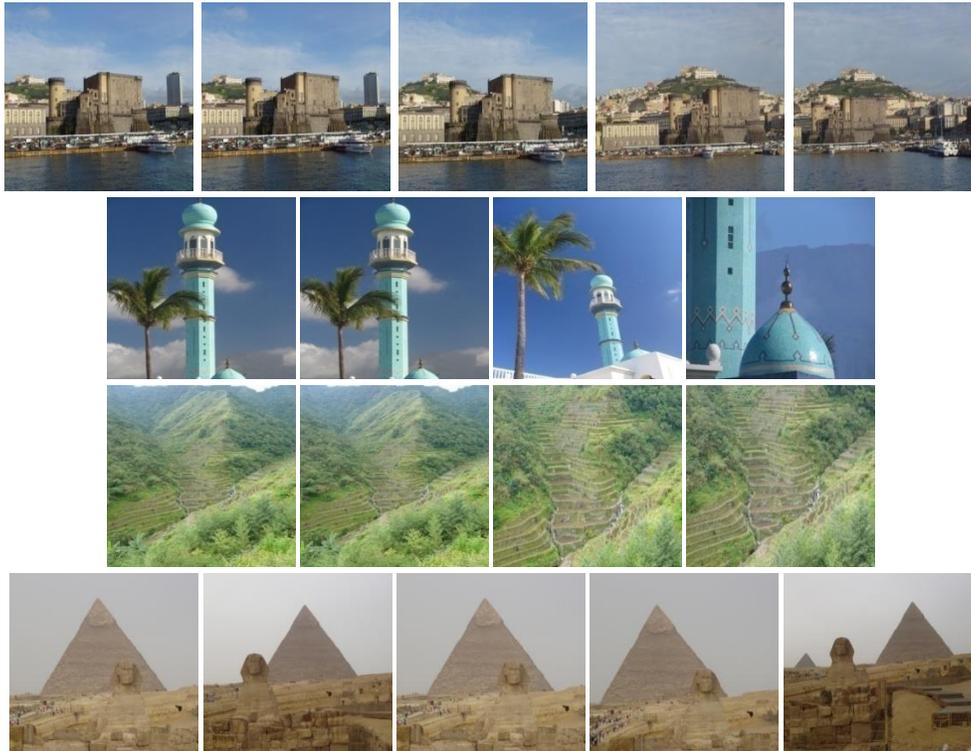

Fig. 7.   Sample Retrieved Images with Corresponding Query Image (First Column).
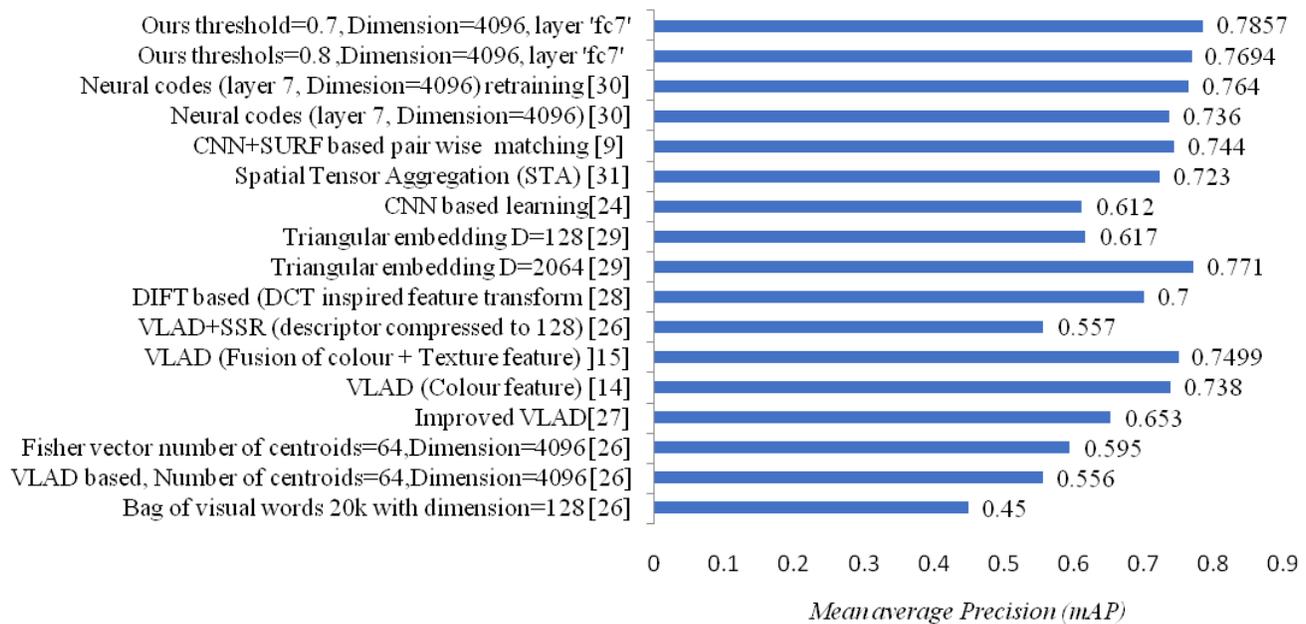
Fig. 8. Mean Average Precision Values of Various State of Art Techniques.

V. CONCLUSION

In this work, combined power of local and global CNN is presented. CNN features are found to be robust than techniques based on traditional image descriptor. Extraction of local CNN features handles difficult matching cases of near duplicate retrieval. Global CNN matching may help in certain cases of images with different viewpoints. CNN features extracted at different levels of neighboring windows help in matching images at different zooming Complementary use of local and global CNN features achieves better retrieval performance. Our model gives significant improvement in retrieval performance with mean average precision value of 0.7857. Pre-computed CNN features improve search time efficiency.

VI. FUTURE SCOPE

- Popular indexing technique such as local sensitive hashing (LSH) may be employed on pre-computed features for further improving search performance.

- Model may be extended for near duplicate video retrieval.

ACKNOWLEDGMENTS

REFERENCES

[1] J. Foo and R. Sinha, "Pruning SIFT for scalable near-duplicate image matching," in Proceedings of the Eighteenth Conference on Australasian Database, Vol. 63, pp.63-71, Darlinghust, Australia, March 2007.

[2] M. Lia, F. Lamberti, "Benchmarking unsupervised near-duplicate image detection," Expert Systems with Applications, Vol. 135, pp.313-326, November 2019.

[3] Z. Zhou, Y. Wang, Q. J. Wu, C. Yang, X. Sun, "Effective and efficient global context verification for image copy detection", IEEE Transactions on Information Forensics and Security, Vol. 12, No. 1, 2017.

[4] H. Chennamma, L. Rangarajan, M. Rao, "Robust near duplicate image matching for digital image forensics", International Journal of Digital Crime and Forensics, Vol. 1, pp.62-79, 2009.

[5] M. Cicconet, H. Elliott, D. L. Richmond, D. Wainstock, M. Walsh, "Image forensics: Detecting duplication of scientific images with manipulation-invariant image similarity", pp. 1-12, February 2018.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems, Vol. 60, No. 4, pp.1097-1105, January 2012.

[7] H. Bay, A. Ess, T. Tuytelaars, and L. G. Van, "Speeded up robust features (SURF)", Computer Vision and Image Understanding, Vol. 110, No. 3, pp.346-359, June 2008.

[8] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, Vol.60, No.2, pp.91-110, January 2004.

[9] T. Mehta, C. Bhensdadia, "Adaptive Near Duplicate Image Retrieval Using SURF and CNN Features", International Journal of Intelligent Engineering and Systems, Vol. 12, No. 5, pp. 104-115, October 2019.

[10] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors", in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.506-513, Washington, DC, USA, July 2004.

[11] D. Nister and H. Stewenius, "Scalable Recognition with a Vocabulary Tree", in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, USA, pp.2161-2168, June 2006.

[12] Perronnin, Y. LIU, J. Sánchez, and H. Poirier, "Large-scale image retrieval with compressed Fisher vectors", in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.3384-3391, San Francisco, CA, USA, June 2010.

[13] H. Jégou, M. Douze, C. Schmid, and P. Perez, "Aggregating local descriptors into a compact image representation", in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.3304-3311, San Francisco, CA, USA, June 2010.

[14] E. Spyromitros-Xioufis, S. Papadopoulos, I. Y. Kompatsiaris, G. Tsoumakas, and I. Vlahavas, "A comprehensive study over vlad and product quantization in large-scale image retrieval", IEEE Transactions on Multimedia, Vol.16, No.6, pp.1713-1728, October 2014.

[15] Y. Wang, Y. Cen, R. Zhao, S. Kan, and S. Hu, "Fusion of multiple VLAD vectors based on different features for image retrieval", in

Proceedings of IEEE 13th International Conference on Signal Processing (ICSP), pp. 742-746, Chengdu, China, November 2016.

[16] H. Jegou, M. Douze, and C. Schmid C, "Hamming Embedding and Weak Geometric Consistency for Large Scale Image Search", in Proceedings of 10th European Conference on Computer Vision, pp.304-317, Marseille, France, October 2008.

[17] D. Zhang and S. Chang, "Detecting image near-duplicate by stochastic attributed relational graph matching with learning", in Proceedings of the 12th annual ACM international conference on Multimedia, New York, USA, pp.877-884, October 2004.

[18] 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, pp.132-136, Amsterdam The Netherlands, October 2016.

[19] Y. Gong, L. Wang, R. Guo, and S. Lazebnik, "Multi-scale orderless pooling of deep convolutional activation features", Computing Research Repository, Vol. abs/1403.1840, September 2014.

[20] K. R. Mopuri and R. V. Babu, "Object level deep feature pooling for compact image representation", in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp.62-70, Boston, MA, USA, June 2015.

[21] L. Xie, R. Hong, B. Zhang, and Q. Tian, "Image Classification and Retrieval are ONE", in Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, pp.3-10, Shanghai, China, June 2015.

[22] K. Yan, Y. Wang, D. Liang, T. Huang, and Y. Tian, "CNN vs. SIFT for Image Retrieval: Alternative or Complementary?", in Proceedings of the 24th ACM international conference on Multimedia, pp.407-411, Amsterdam, The Netherlands, October 2016.

[23] L. Zheng, S. Wang, J. Wang, and Q. Tian, "Accurate Image Search with Multi-Scale Contextual Evidences", International Journal of Computer Vision, Vol.120, No.1, pp.1-13, March 2016.

[24] Y. Zhang, Y. Zhang, J. Sun, H. Li, and Y. Zhu, "Learning Near Duplicate Image Pairs using Convolution Neural Networks", International Journal of Performability Engineering, Vol.14, No.1, pp.168-177, January 2018.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", in Proceedings of International Conference on Learning Representations, Kuala Lumpur , November 2015.

[26] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid, "Aggregating Local Image Descriptors into Compact Codes", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.34, No.9, pp.1704-1716, September 2012.

[27] R. Arandjelovic and A. Zisserman, "All about VLAD", in Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pp. 1578-1585, Portland, OR, June 2013.

[28] Y. Wang, M. Shi, S. You, and C. Xu, "DCT Inspired Feature Transform for Image Retrieval and Reconstruction", IEEE Transactions on Image Processing, Vol. 25, No. 9, pp.4406-4420, July 2016.

[29] H. Jégou and A. Zisserman, "Triangulation Embedding and Democratic Aggregation for Image Search", in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.3310-3317, Columbus, OH, September, 2014.

[30] A. Babenko, A. Slesarev, A. Chigorin and V. Lempitsky, "Neural Codes for Image Retrieval", in Proceedings Of 13th European Conference on Computer Vision, Zurich, Switzerland, Lecture Notes in Computer Science, Springer, Vol. 8689, pp.584-599, Zurich, Switzerland, September 2014.

[31] D. Picard, "Preserving local spatial information in image similarity using tensor aggregation of local features", in Proceedings Of IEEE International Conference on Image Processing, pp. 201-205, Phoenix, AZ, August 2016.