ORIGINAL ARTICLE

Local bit-plane decoded convolutional neural network features for biomedical image retrieval

Shiv Ram Dubey[1] · Swalpa Kumar Roy[2] · Soumendu Chakraborty[3] · Snehasis Mukherjee[1] · Bidyut Baran Chaudhuri[2]

Abstract

Biomedical image retrieval is a challenging problem due to the varying contrast and size of structures in the images. The approaches for biomedical image retrieval generally rely on the feature descriptors to characterize the images. The feature descriptor of query image is compared with the descriptors of images from the database, to find the best matches. Several hand-crafted feature descriptors have been proposed so far for biomedical image retrieval by exploiting the local relationship of neighboring image pixels. It is observed in the literature that the local bit-plane decoded features are well suited for this retrieval task. Moreover, in recent past, it is also observed that the convolutional neural network-based features such as AlexNet, Vgg16, GoogleNet and ResNet perform well in many computer vision-related tasks. Motivated by the success of the deep learning-based approaches, this paper proposes a local bit-plane decoding-based AlexNet descriptor (LBpDAD) for biomedical image retrieval. The proposed LBpDAD is computed by max-fusing the ReLU operated feature maps of pre-trained AlexNet at a particular layer, obtained from the original and local bit-plane decoded images. The proposed approach is also compared with Vgg16, GoogleNet and ResNet models. The experiments on the proposed method over three benchmark biomedical databases of different modalities such as MRI, CT and microscopic show the efficacy of the proposed descriptor.

Keywords Biomedical images · Convolutional neural networks · AlexNet · Image retrieval · Local bit-plane decoding

✉ Shiv Ram Dubey
  srdubey@iiits.in

  Swalpa Kumar Roy
  swalparoy@gmail.com

  Soumendu Chakraborty
  soum.uit@gmail.com

  Snehasis Mukherjee
  snehasis.mukherjee@iiits.in

  Bidyut Baran Chaudhuri
  bbc@isical.ac.in

[1] Computer Vision Group, Indian Institute of Information Technology, Sri City, Chittoor, Andhra Pradesh 517646, India

[2] Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata 700108, India

[3] Computer Science and Engineering Department, Indian Institute of Information Technology, Lucknow, Uttar Pradesh, India

1 Introduction

Computer-based biomedical image analysis techniques facilitate the medical experts and technicians to improve their diagnosis of diseases, based on the crucial inputs suggested by the computer system [49, 53]. The biomedical image retrieval is one of the fundamental and very challenging problem for medical and health informatics [28]. In image retrieval, the best matching image along with its descriptions is identified from a database against a query image, based on the content similarity between the query and database images [19, 52]. In order to measure the similarity between images, the feature representation plays an important role [38, 47, 56].

In the past, the local binary pattern (LBP) was very popular for image representation [35]. Numerous LBP variants were proposed for addressing challenges in image retrieval in the past decades due to the huge success and simplicity of LBP [38]. Some notable LBP variants are

local ternary pattern (LTP) [55], local derivative pattern (LDP) [62], local gradient hexa pattern (LGHP) [3], local directional gradient pattern (LDGP) [4] and local directional order pattern (LDOP) [11] for face recognition/retrieval purposes; local tetra pattern (LTrP) [29] and multi-channel decoded LBP (mdLBP) [16] for image retrieval; local intensity order pattern (LIOP) [58] and interleaved intensity order-based local descriptor (IOLD) [12] for local image matching; and complete dual-cross pattern (CDCP) [44], local directional ZigZag pattern (LDZP) [46] and local jet pattern (LJP) [45] for texture classification. The LBP-based approaches are also widely used in biomedical image analysis such as pulmonary emphysema analysis [51], cell phenotype classification [33], biomedical image classification [34] and stem cell classification [36]. The latest developments over the LBP variant descriptors for biomedical image retrieval include: local mesh pattern (LMeP) [31], local ternary co-occurrence pattern (LTCoP) [30], local diagonal extrema pattern (LDEP) [13], local bit-plane dissimilarity pattern (LBDISP) [17], local bit-plane decoded pattern (LBDP) [15] and local wavelet pattern (LWP) [14]. Lan and Zhou have used the compressed scattering coefficients for medical image retrieval [24]. It is observed from the literature that the bit-plane decoding-based descriptor is more suitable for the biomedical image retrieval task [15]. Thus, in this work, we utilize the bit-plane decoded information with convolutional neural network (CNN) framework.

During the past few years, the CNN-based methods have emerged very rapidly. The CNN-based approaches show better efficacy compared to the hand-designed feature-based classifications. The first revolutionary work in this direction was AlexNet architecture by Krizhevsky et al. [23] for image classification. After AlexNet, various deep architectures have been proposed such as Vgg16 with deep network [48], GoogleNet with inception module [54] and ResNet with residual module [21] for image classification. The CNN has also proven for other problems such as Faster R-CNN [43] for object detection, Mask R-CNN [20] for semantic segmentation, image fusion [22], CNN-ranker [61] for retrieval and Cross-CNN [59] for multiple modality data representation. The CNN-based methods are also proven to be efficient for biomedical image analysis such as colon cancer recognition [50], cervical cell classification [63], pneumonia detection [41], multispectral MR images segmentation [5] and medical image registration [57].

In order to train the deep CNNs, a huge number of images are required which may not be collected in many real-life scenarios. This issue is generally dealt with by applying the transfer learning with pre-trained models, trained over some big databases. Researchers have used the pre-trained CNN models for applications such as content-

based image retrieval [25], remote sensing image retrieval [18], face retrieval [10], military object recognition [60] and dumpsters recognition [42]. The CNN models pre-trained by ImageNet database [9] are also successfully applied in medical image applications such as mammogram analysis [2], bioimage classification [32] and domain transfer for biomedical images [37].

Some attempts have been made to utilize the CNN for biomedical image retrieval. Qayyum et al. [39] used a eight-layer CNN architecture similar to AlexNet for medical image retrieval. They trained the network over a database of 7200 images obtained from different sources and gained a mean average precision of 0.69. Due to lack of sufficient training images, they could not get very high performance. Qiu et al. [40] have used the hash code over 'FC6' and 'FC7' AlexNet features for medical image retrieval. The retrieval time is reduced in [40] due to binary feature, but at the cost of degraded performance. Chung et al. [7] used a deep Siamese CNN (SCNN) for diabetic retinopathy fundus image retrieval. The retrieval performance of SCNN last layer proposed in [7] is quite similar to the CNN softmax layer. Chowdhury et al. [6] used the CNN and edge histogram descriptor for radiographic image retrieval. This approach works in two steps. First, the relevant database classes are computed for a query image using CNN and then, the hand-crafted edge histogram descriptor is used to retrieve the images only from the relevant classes. This approach has combined CNN with hand-crafted descriptor in sequential fashion. However, in our proposed approach, the CNN features are computed over hand-designed feature map and fused with the original CNN features (i.e., parallel fusion).

Motivated by the suitability of bit-plane decoding mechanism for biomedical images, the success of CNN in various challenging problems and the re-usability of the pre-trained models, we propose local bit-plane decoded CNN descriptors for biomedical image retrieval. The main contributions of this paper can be summarized as follows:

- The local bit-plane decoding mechanism is used for image transformation similar to LBDP [15].
- The pre-trained CNN models such as AlexNet [23], Vgg16 [48], GoogleNet [54] and ResNet50 [21] are employed to generate the features.
- The CNN features are generated over raw input image as well as bit-plane decoded image and combined at the last representation layers using different fusion strategies.

The rest of the paper is structured as follows: Section 2 proposes the local bit-plane decoded CNN descriptor; Sect. 3 presents the experimental setup including retrieval framework, databases and evaluation criteria; Sect. 4

reports the experimental results and analysis; and Sect. 5 concludes the paper.

2 Proposed local bit-plane decoded CNN descriptor

This section illustrates the proposed local bit-plane decoded AlexNet descriptor (*LBpDAD*) obtained by integrating the trained AlexNet [23] with local bit-plane decoding mechanism [15]. The trained weights of AlexNet model[1] are used in this paper which is computed over a large-scale ImageNet database [9]. The proposed method for biomedical image retrieval is illustrated in Fig. 1. The input image I of dimension $m \times n \times 3$ is passed through the local bit-plane decoding mechanism proposed in [15] to generate the local bit-plane decoded map I_M as follows:

$$I_M^{i,j,k} = \sum_{b=1}^{8} \text{sign}\left(I^{i,j,k}, B_D^{i,j,k,b}\right) \times 2^{b-1} \tag{1}$$

where $i = 2, 3, \ldots, m-1$, $j = 2, 3, \ldots, n-1$, $k = 1, 2, 3$ represents the kth channel, $b = 1, 2, \ldots, 8$ represents the bth bit-plane, $I^{i,j,k}$ is the value at position (i, j, k) in input image, $I_M^{i,j,k}$ is the value at position (i, j, k) in the output image map of local bit-plane decoding, $B_D^{i,j,k,b}$ is the local bit-plane decoded decimal value in bth bit-plane for the center pixel (i, j) in kth channel and the $\text{sign}(\alpha, \beta)$ is given as,

$$\text{sign}(\alpha, \beta) = \begin{cases} 1, & \text{if } \alpha \geq \beta \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

The $B_D^{i,j,k,b}$ is computed as,

$$B_D^{i,j,k,b} = \sum_{n=1}^{8} B_n^{i,j,k,b} \times 2^{n-1} \tag{3}$$

where $B_n^{i,j,k,b}$ is the binary bit in bth bit-plane of kth channel corresponding to the nth neighbor of $I^{i,j,k}$ at unit distance in the direction of $(n-1) \times 45°$ from positive x axis.

Now, the input image I and local bit-plane decoded image map I_M are converted into I_A and I_{MA}, respectively, to satisfy the dimension required from the input image for the pre-trained AlexNet. The I_A and I_{MA} are computed as,

$$I_A = \tau(I, [227, 227]) \tag{4}$$

$$I_{MA} = \tau(I_M, [227, 227]) \tag{5}$$

where $\tau(\Gamma, [\xi, \xi])$ is a function to resize any 3D volume Γ of dimension $\varrho \times \upsilon \times \psi$ into the dimension of $\xi \times \xi \times \psi$.

The 227×227 denotes the spatial resolution needed from the input for AlexNet.

Define *Alex* as a function of combinations of convolutional, ReLU, max-pooling and fully connected layers, which returns the features at a particular layer of pretrained AlexNet for an input image of dimension $227 \times 227 \times 3$. The AlexNet features *AlexNet* and *LBpD_Alex* are computed for input images I_A and I_{MA}, respectively, at class score layer ('cs') as,

$$AlexNet = ReLU(Alex(I_A, cs)) \tag{6}$$

$$LBpD\_Alex = ReLU(Alex(I_{MA}, cs)) \tag{7}$$

where *ReLU* [23] is a function defined as,

$$ReLU(\phi_v) = \begin{cases} \phi_v, & \text{if } \phi_v \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

$\forall v = 1, 2, \ldots, D(\phi)$, where ϕ represents a feature vector and $D(\phi)$ represents the length of feature vector ϕ. The ReLU operator is basically used in CNN framework to introduce nonlinearity into convolved features by filtering the negative values. Note that the ReLU operator over feature vector is required to remove the negative values as only nonnegative values are useful in most distance measures.

The *Maximum* ('Max') fusion technique is used to combine the *AlexNet* and *LBpD_Alex* feature vectors into final *LBpDAD* descriptor as,

$$LBpDAD_v = M(AlexNet_v, LBpD\_Alex_v) \tag{9}$$

where $LBpDAD_v$, $AlexNet_v$ and $LBpD\_Alex_v$ are the vth elements of *LBpDAD*, *AlexNet* and *LBpD_Alex* feature vectors, respectively, $v = 1, 2, \ldots, D(AlexNet)$ with $D(AlexNet) = D(LBpD\_Alex)$ and M is a 'Max' operator defined as,

$$M(\alpha, \beta) = \begin{cases} \alpha, & \text{if } \alpha \geq \beta \\ \beta, & \text{otherwise.} \end{cases} \tag{10}$$

The $LBpDAD^{fc7}$ (i.e., final fused feature vector at 'fc7' layer) is computed as,

$$LBpDAD_v^{fc7} = M(AlexNet_v^{fc7}, LBpD\_Alex_v^{fc7}) \tag{11}$$

where $LBpDAD_v^{fc7}$, $AlexNet_v^{fc7}$ and $LBpD\_Alex_v^{fc7}$ are the vth elements of $LBpDAD^{fc7}$, $AlexNet^{fc7}$ and $LBpD\_Alex^{fc7}$ feature vectors, respectively. The $AlexNet^{fc7}$ and $LBpD\_Alex^{fc7}$ are the feature vectors computed at 'fc7' layer for the input images I_A and I_{MA}, respectively, as,

$$AlexNet^{fc7} = ReLU(Alex(I_A, fc7)) \tag{12}$$

---

[1] The trained AlexNet weights available in the MATLAB are considered.
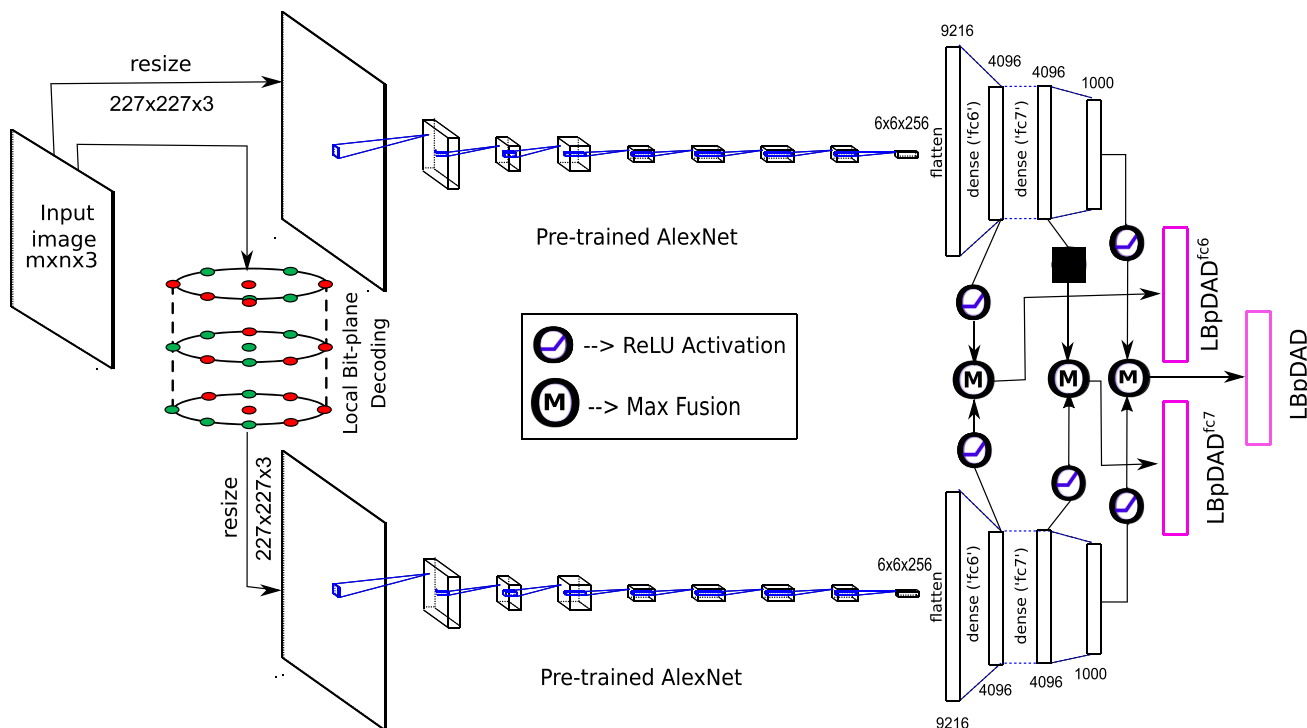
Fig. 1 Proposed local bit-plane decoded AlexNet descriptor (LBpDAD) by fusing the original AlexNet features with local bit-plane decoded AlexNet features

$$LBpD\_Alex^{fc7} = ReLU(Alex(I_{MA}, fc7)) \quad (13)$$

Similarly, the feature fused feature vector at 'fc6' layer can be computed as,

$$LBpDAD_v^{fc6} = M(AlexNet_v^{fc6}, LBpD\_Alex_v^{fc6}) \quad (14)$$

where $LBpDAD_v^{fc6}$, $AlexNet_v^{fc6}$ and $LBpD\_Alex_v^{fc6}$ are the vth elements of $LBpDAD^{fc6}$, $AlexNet^{fc6}$ and $LBpD\_Alex^{fc6}$ feature vectors, respectively. The $AlexNet^{fc6}$ and $LBpD\_Alex^{fc6}$ are the feature vectors computed at 'fc6' layer for the input images I_A and I_{MA} respectively as,

$$AlexNet^{fc6} = ReLU(Alex(I_A, fc6)) \quad (15)$$

$$LBpD\_Alex^{fc6} = ReLU(Alex(I_{MA}, fc6)). \quad (16)$$

Note that all the feature descriptors are normalized to make the unit sum using following formula,

$$\phi_v = \frac{\phi_v}{\sum_{i=1}^{D(\phi)} \phi_i} \quad (17)$$

where ϕ is any feature vector of dimension $D(\phi)$. This normalization makes the descriptors robust against image resolution variations.

3 Experimental setup

In this section, at first, we present the biomedical image retrieval framework using proposed descriptor. Then biomedical databases used for the experiments and finally the evaluation measures are described.

3.1 Proposed biomedical image retrieval framework

The biomedical image retrieval framework using proposed local bit-plane decoded AlexNet descriptor (LBpDAD) is portrayed in Fig. 2. The feature extraction steps are the same for both query image and database images. The image is passed through the pre-trained AlexNet to generate the direct features. The input image is also converted into a local bit-plane decoded map which is then passed through the pre-trained AlexNet to generate the local bit-plane decoded features. Finally, the direct Alex features and local bit-plane decoded Alex features are combined using 'Max' fusion strategy to generate the final LBpDAD descriptor. As the biomedical images are gray scale and AlexNet requires three-channel input, the same gray scale channel of our image is copied three times to create the three-channel input. Once the descriptors are computed for all images including query and database, the feature matching is performed by computing the distances between

Fig. 2 The biomedical image retrieval framework using proposed *LBpDAD* features
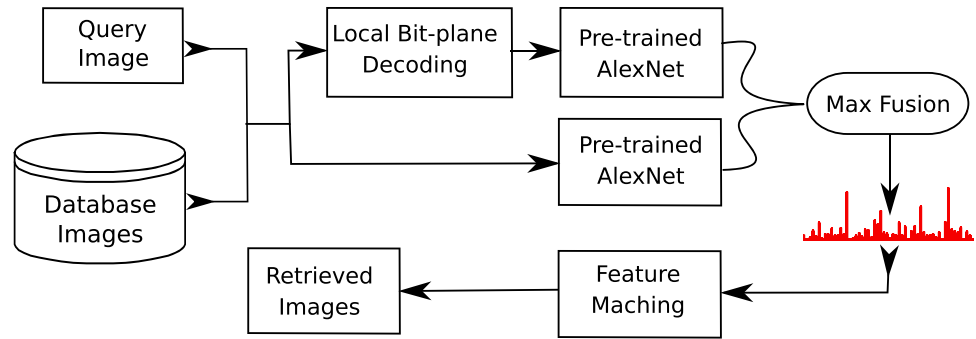
descriptors of query image and database images. Based on the distances, the top matching images are retrieved from the database against the given query image. The 'Chi-square' distance measure is adapted in this paper as it has shown better performance for state-of-the-art descriptors [14, 15]. However, the performance of proposed *LBpDAD* descriptor is also analyzed with other distances such as 'Euclidean,' 'Manhattan,' 'Cosine' and 'Canberra' in the Experiment Section.

3.2 Biomedical databases used

Three biomedical databases of different modalities including OASIS-MRI [27], TCIA-CT [8] and HeLa-Microscopic [1] are used in this paper to justify the improved performance of proposed *LBpDAD* descriptor in image retrieval framework. The Open Access Series of Imaging Studies has released a magnetic resonance imaging database (OASIS-MRI) in public domain for research and analysis [27]. This database is based on the 421 subjects from the age-group between 18 and 96 years. The OASIS-MRI database contains the 176×208 resolution cross-sectional images. The database is divided into four categories similar to [15] having 106, 89, 102 and 124 images. The different categories of this database represent varying ventricular shape inside the images. The cancer image archive (TCIA) is a storage for various cancer location images in Digital Imaging and Communications in Medicine (DICOM) image format [8]. These images are publicly accessible for research. We have used the same TCIA-CT database which is used in [14]. This database has 604 Colo_prone 1.0B30f CT images of the DICOM series number 1.3.6.1.4.1.9328.50.4.2 of study instance UID 1.3.6.1.4.1.9328.50.4.1 for subject 1.3.6.1.4.1.9328.50.4.0001. The database is divided into eight categories having 75, 50, 58, 140, 70, 92, 78 and 41 images as per the size and structure of Colo_prone. The original image size in TCIA-CT database is 512×512 pixels. We have also used fluorescence microscope images for the experiment taken from the 2D HeLa database [1]. This database contains total 862 images of HeLa cells from

ten different categories corresponding to 10 different subcellular patterns using fluorescence microscopy.

3.3 Evaluation criteria

The average retrieval precision (ARP), average retrieval rate (ARR), *F*-Score and average normalized modified retrieval rank (ANMRR) are used for the performance measurement similar to [13–15, 17, 30, 31]. The ARP and ARR are computed as,

$$ARP = \frac{1}{C} \sum_{c=1}^{C} MP_c \qquad (18)$$

$$ARR = \frac{1}{C} \sum_{c=1}^{C} MR_c \qquad (19)$$

where C is the number of classes in a database, MP_c and MR_c are the mean precision and mean recall for cth class and defined as,

$$MP_c = \frac{1}{n_c} \sum_{i=1}^{n_c} \frac{\#CR_i}{\#TR} \qquad (20)$$

$$MR_c = \frac{1}{n_c} \sum_{i=1}^{n_c} \frac{\#CR_i}{\#TG_c} \qquad (21)$$

where n_c is the number of images in cth class, $\#CR_i$ is the number of correctly retrieved images, $\#TR$ is the total number of retrieved images, and $\#TG_c$ is the number of ground truth images in cth class. The *F*-Score is calculated from the ARP and ARR as,

$$F\text{-}Score = 2 \times \frac{ARP \times ARR}{ARP + ARR}. \qquad (22)$$

The ANMRR is calculated by following the steps provided in [26]. Higher values of ARP, ARR and *F*-Score and lower value of ANMRR represent better performance.
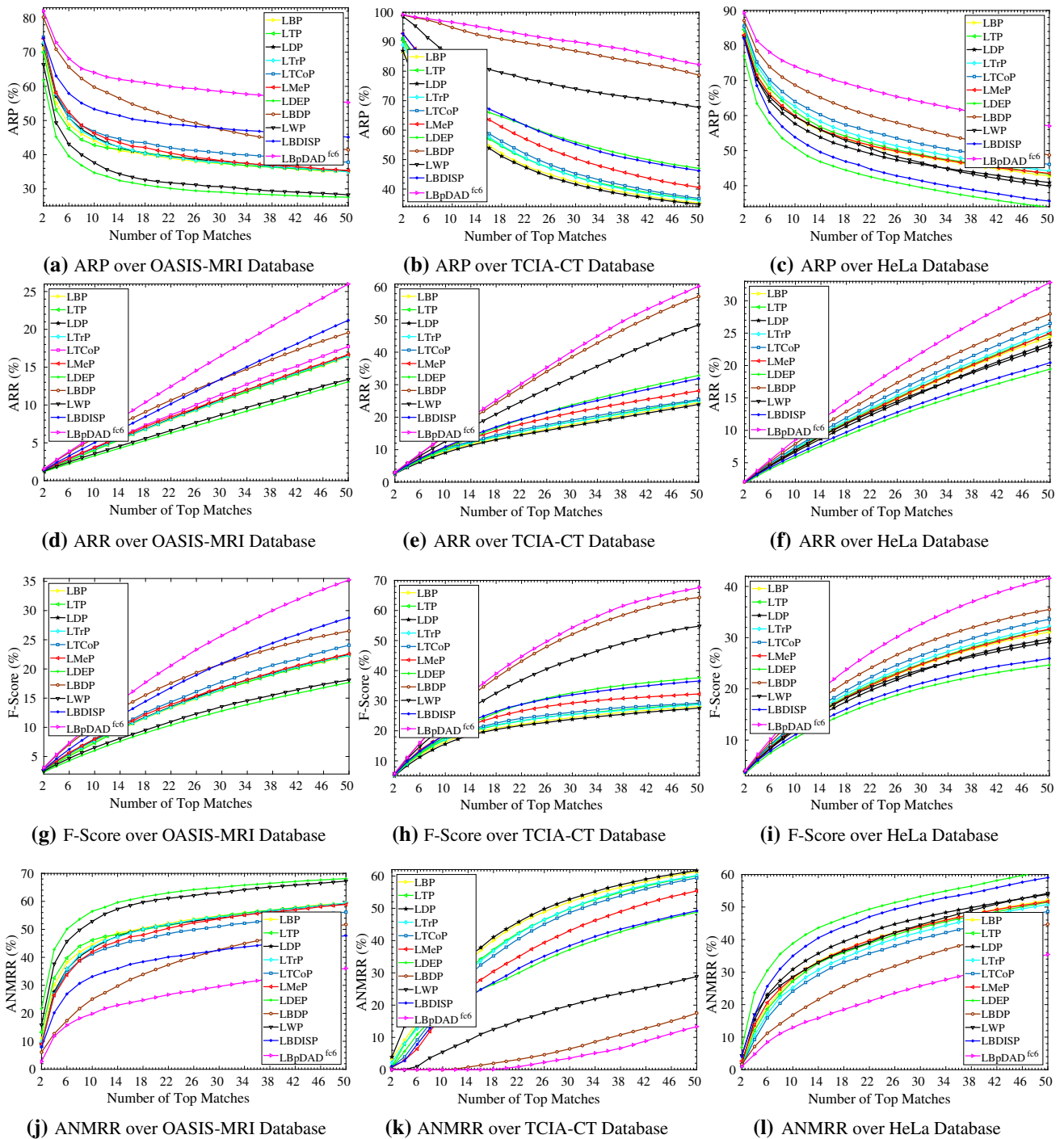
(a) ARP over OASIS-MRI Database

(b) ARP over TCIA-CT Database

(c) ARP over HeLa Database

(d) ARR over OASIS-MRI Database

(e) ARR over TCIA-CT Database

(f) ARR over HeLa Database

(g) F-Score over OASIS-MRI Database

(h) F-Score over TCIA-CT Database

(i) F-Score over HeLa Database

(j) ANMRR over OASIS-MRI Database

(k) ANMRR over TCIA-CT Database

(l) ANMRR over HeLa Database

Fig. 3 The retrieval results comparison over OASIS-MRI, TCIA-CT and HeLa databases

4 Results and analysis

This section presents the experimental results, comparison between the methods and analysis. First, the results of proposed model are compared with the state-of-the-art methods and then its performance is analyzed for different layers, fusion strategies, distance measures and CNN models.

4.1 Results comparison

In order to express the improved performance of proposed model, the $LBpDAD^{fc6}$ results are compared with the
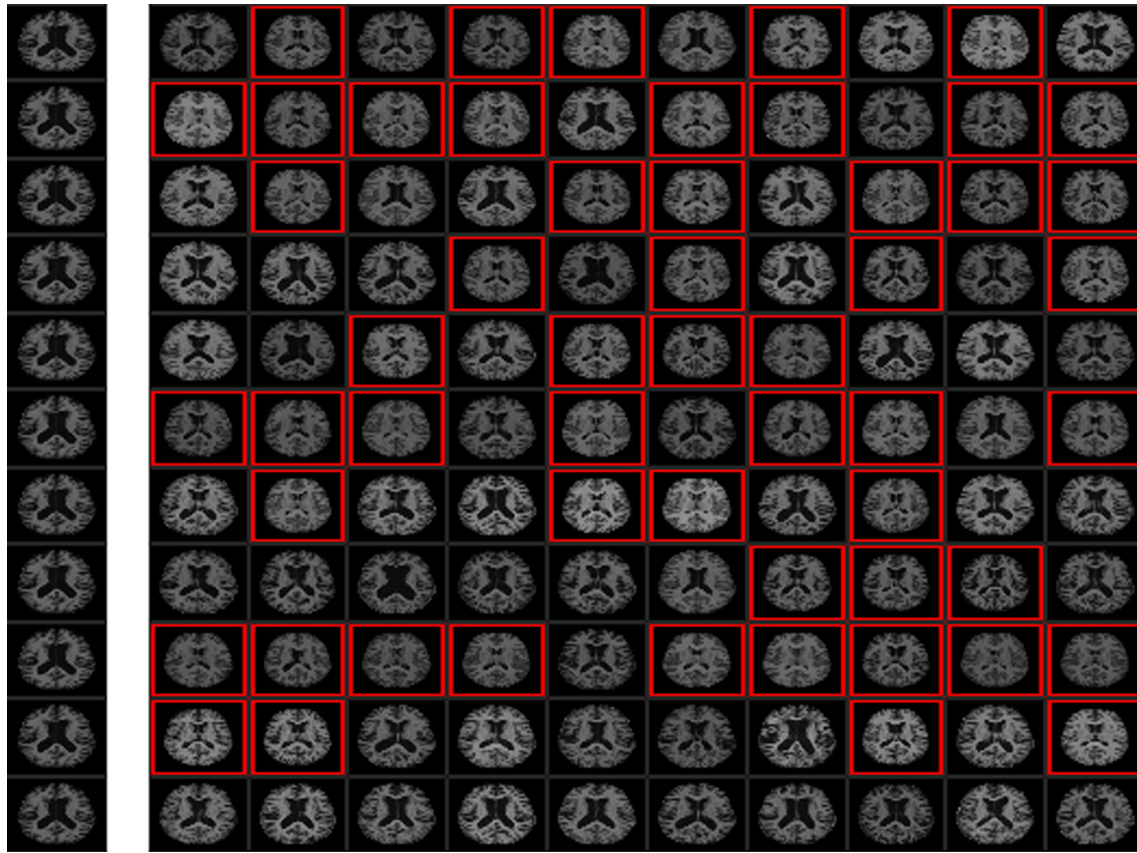
Fig. 4 The retrieval results from OASIS-MRI database. The first column represents the query image. The third to last columns represent the top ten retrieved images in decreasing order of similarity against the query image in first column. The results in first to 11th rows are corresponding to *LBP* [35], *LTP* [55], *LDP* [62], *LTrP* [29], *LTCoP* [30], *LMeP* [31], *LDEP* [13], *LBDP* [15], *LWP* [14], *LBDISP* [17] and proposed $LBpDAD^{fc6}$ descriptors, respectively. The false positive retrieved images are highlighted in red color rectangles

results of state-of-the-art descriptors such as *LBP* [35], *LTP* [55], *LDP* [62], *LTrP* [29], *LTCoP* [30], *LMeP* [31], *LDEP* [13], *LBDP* [15], *LWP* [14] and *LBDISP* [17]. Note that $LBpDAD^{fc6}$ is used here for comparison, whereas the comparison between $LBpDAD^{fc6}$, $LBpDAD^{fc7}$ and *LBpDAD* descriptors is carried out in the next subsection. The image retrieval results in terms of the ARP (%), ARR (%), *F*-Score (%) and ANMRR (%) by varying the number of retrieved images are presented in Fig. 3. The first, second, third and 4th rows contain the ARP, ARR, *F*-Score and ANMRR plots, respectively. The first, second and third columns are dedicated to the results over OASIS-MRI, TCIA-CT and HeLa databases, respectively. The Chi-square distance is used for feature matching.

It is observed from the results of Fig. 3a, d, g, j that the proposed $LBpDAD^{fc6}$ descriptor outperforms the state-of-the-art descriptors with a big margin. The $LBpDAD^{fc6}$ descriptor is also succeeded on TCIA-CT database and just beats the best performing *LBDP* descriptor in terms of all the evaluation measures (see Fig. 3b, e, h k). Similar improvement in the results is also observed for HeLa

database as shown in Figs. 3c, f, i, l using the proposed descriptor as compared to the existing descriptors. The improved performance of the proposed descriptor may be due to the following three reasons: (1) The used CNN features are more discriminative when trained over big ImageNet database, (2) the local bit-plane decoding mechanism is better suited for biomedical images, and (3) the fusion of raw CNN feature and local bit-plane decoded CNN feature further improves the discriminative power of the resultant descriptor.

The retrieved images using different methods for the example query image of OASIS-MRI, TCIA-CT and HeLa database are shown in Figs. 4, 5 and 6, respectively. In these figures, the results in first to 11th rows are corresponding to *LBP* [35], *LTP* [55], *LDP* [62], *LTrP* [29], *LTCoP* [30], *LMeP* [31], *LDEP* [13], *LBDP* [15], *LWP* [14], *LBDISP* [17] and proposed $LBpDAD^{fc6}$ descriptors, respectively. The first column represents the query image. The third to last columns represent the top ten retrieved images in decreasing order of similarity against the query image in first column. The false positive retrieved images are highlighted in red color rectangles. It can be observed

Fig. 5 The retrieval results from TCIA-CT database. The first column represents the query image. The third to last columns represent the top ten retrieved images in decreasing order of similarity against the query image in first column. The results in first to 11th rows are corresponding to *LBP* [35], *LTP* [55], *LDP* [62], *LTrP* [29], *LTCoP* [30], *LMeP* [31], *LDEP* [13], *LBDP* [15], *LWP* [14], *LBDISP* [17] and proposed $LBpDAD^{fc6}$ descriptors, respectively. The false positive retrieved images are highlighted in red color rectangles

from these results that the proposed method (last row) outperforms other methods. The $LBpDAD^{fc6}$ is able to gain the 100%, 90% and 100% precision over OASIS-MRI (Fig. 4), TCIA-CT (Fig. 5) and HeLa (Fig. 6) databases, respectively.

4.2 Performance analysis over different layers

The previous subsection presented a comparative result of $LBpDAD^{fc6}$ descriptor with the existing descriptors. In this experiment, the results of proposed descriptor are analyzed at different layers, i.e., *LBpDAD* for 'class score layer,' $LBpDAD^{fc7}$ for 'fc7' layer and $LBpDAD^{fc6}$ for 'fc6' layer (see Fig. 7). Moreover, the results of original AlexNet (i.e., *AlexNet*, $AlexNet^{fc7}$ and $AlexNet^{fc6}$ for 'class score,' 'fc7' and 'fc6' layers, respectively) as well as the results of local bit-plane decoded AlexNet without fusion (i.e., $LBpD\_Alex$, $LBpD\_Alex^{fc7}$ and $LBpD\_Alex^{fc6}$ for 'class score,' 'fc7' and 'fc6' layers, respectively) are also compared in Fig. 7. The results are shown for ARP (first row) and ANMRR (second row) evaluation metrics over

OASIS-MRI (first column), TCIA-CT (second column) and HeLa (third column) databases in Fig. 7. It is perceived across the plots of Fig. 7 that in general, the performance of fused local bit-plane decoded AlexNet descriptors (i.e., *LBpDAD*, $LBpDAD^{fc7}$ and $LBpDAD^{fc6}$) is better than the local bit-plane decoded AlexNet descriptors without fusion (i.e., $LBpD\_Alex$, $LBpD\_Alex^{fc7}$ and $LBpD\_Alex^{fc6}$), respectively, which in turn better than the original AlexNet descriptors (i.e., *AlexNet*, $AlexNet^{fc7}$ and $AlexNet^{fc6}$), respectively. Moreover, the performance gain due to 'Max' fusion is very prominent over HeLa database. This observation also supports that the CNN features extracted over local bit-plane decoded image are more discriminative compared to raw CNN features. This is so because the local bit-plane decoded image is rich with the local relationship at each bit-plane, whereas both CNN features have the complementary information due to different input modalities (i.e., raw input image and local bit-plane decoded input image). It is also discovered from this experiment that the features of $LBpDAD^{fc6}$ at 'fc6' layer are more discriminative than the features of $LBpDAD^{fc7}$ at 'fc7' and
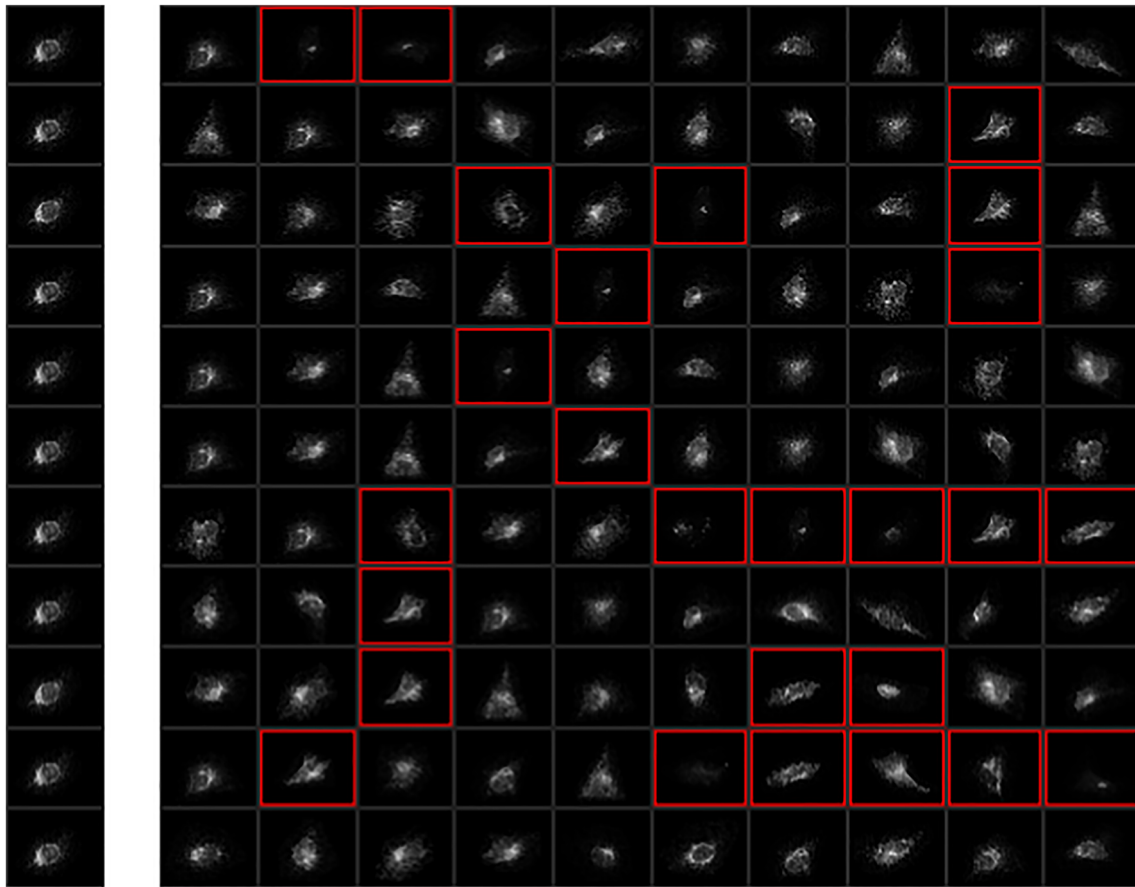
Fig. 6 The retrieval results from HeLa database. The first column represents the query image. The third to last columns represent the top ten retrieved images in decreasing order of similarity against the query image in first column. The results in first to 11th rows are corresponding to *LBP* [35], *LTP* [55], *LDP* [62], *LTrP* [29], *LTCoP* [30], *LMeP* [31], *LDEP* [13], *LBDP* [15], *LWP* [14], *LBDISP* [17] and proposed *LBpDAD*^{fc6} descriptors, respectively. The false positive retrieved images are highlighted in red color rectangles

LBpDAD at 'class score' layers for the OASIS-MRI and TCIA-CT databases because the later 'fc7' and 'class score' layer features are more fitted toward the training database as compared to the earlier 'fc6' layer features. However, the *LBpDAD*^{fc7} descriptor at 'fc7' layer is best the performing one on HeLa database due to the presence of more homogeneous regions in the images.

4.3 Performance analysis using different fusion strategies

This experiment is done to analyze the effects of different fusion strategies for combining the features of original AlexNet and local bit-plane decoded AlexNet. The ARP (%) values using *Maximum* (Max), *Addition* (Add), *Product* (Prod), *Absolute Difference* (Diff) and *Division* (Div) fusion strategies are summarized in Table 1. Note that all features are passed through the ReLU operator before fusion. The *LBpDAD*, *LBpDAD*^{fc7} and *LBpDAD*^{fc6} descriptors are used over OASIS-MRI, TCIA-CT and HeLa

databases to validate the results in this experiment. The number of retrieved images is set to 5 and Chi-square distance is used in this experiment. The main objective of the proposed method is to study the effect of feature-level fusion of hand-crafted and CNN features. There could be many possible fusion strategies. In experiments, we have explored some of them. Even though *Product* 'Prod' fusion technique has performed better in many instances (as *Product* of two nonnegative feature vectors is more sparse, which decreases the effect of inter-class variability over the final feature vector), it introduces additional computational overheads. Hence, we have opted for the 'Max' fusion strategy in the remaining experiments.

In order to observe the statistical difference between the results of different fusion strategies, we conduct the *t test* over the results of each pair of fusion strategy. Note that the higher value of absolute *t test* represents high variability between two distributions and vice versa. Moreover, the positive sign represents the greater values for the corresponding distribution against other distribution. We
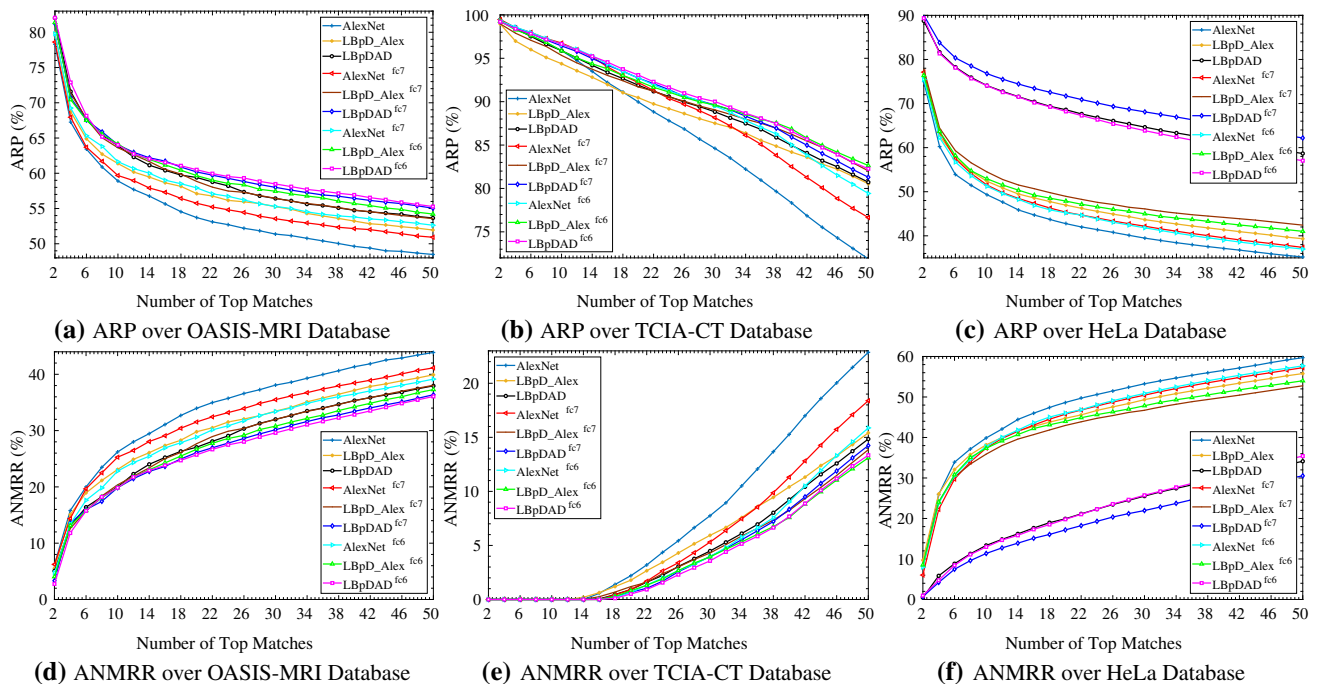
(a) ARP over OASIS-MRI Database

(b) ARP over TCIA-CT Database

(c) ARP over HeLa Database

(d) ANMRR over OASIS-MRI Database

(e) ANMRR over TCIA-CT Database

(f) ANMRR over HeLa Database

Fig. 7 The comparison between *AlexNet*, *LBpD_Alex* and *LBpDAD* features taken from 'Softmax,' 'FC7' and 'FC6' layers over OASIS-MRI, TCIA-CT and HeLa databases using ARP and ANMRR evaluation measures. Here, *AlexNet* refers to the features computed over raw image, *LBpD_Alex* represents the *AlexNet* features computed over local bit-plane decoded image instead of original image, and *LBpDAD* depicts the features obtained after fusing *AlexNet* and *LBpD_Alex* using 'Max' fusion strategy

summarize the *t* test values for the results of Table 1 in Table 2. It is clear from this table that the overall performance using *Max* fusion is better, as it has positive *t* test values as compared to all other fusion strategies. The *t* test analysis confirms the choice of using *Max* fusion strategy in proposed method. It is also observed that statistically, the {*Max*, *Add*} fusion approaches and {*Prod*, *Diff*} fusion approaches are very similar.

4.4 Performance analysis using different distance measures

The Chi-square distance measure is used in the previous results to find the dissimilarity between two images. This experiment is conducted to analyze the effect of distance measures over the performance of proposed descriptors. The *Euclidean* (Eucld), *Manhattan* (L1), *Cosine* (Cosn), *Canberra* (Canb) and *Chi-square* (Chisq) distance measures are considered for this experiment. The results with different distance measures in terms of the ARP (%) for 5 retrieved images using *LBpDAD*, *LBpDAD^{fc7}* and *LBpDAD^{fc6}* descriptors are illustrated in Table 3. The Chi-square distance is generally used with many hand-crafted descriptors, as it works well with histograms, whereas the feature vector of the proposed descriptor is not in the form of histogram. Though Canberra distance is better to find the distance between two vectors (not histogram), for fair

Table 1 The results comparison in between *Maximum* (Max), *Addition* (Add), *Product* (Prod), *Absolute Difference* (Diff) and *Division* (Div) fusion strategies in terms of the ARP values for 5 number of retrieved images

Method	Dataset	Max	Add	Prod	Diff	Div
LBpDAD	OASIS	68.67	69.53	**69.72**	66.12	63.45
	TCIA	98.00	97.88	97.93	**98.03**	97.77
	HeLa	79.43	78.74	78.63	**79.48**	77.26
LBpDAD^{FC7}	OASIS	68.95	68.73	**69.37**	68.12	66.59
	TCIA	**98.27**	98.25	98.19	98.15	98.23
	HeLa	82.21	81.77	80.16	82.14	**82.94**
LBpDAD^{FC6}	OASIS	69.80	69.95	**70.10**	69.50	67.86
	TCIA	98.13	98.16	**98.29**	98.11	98.12
	HeLa	79.62	79.26	77.74	79.91	**83.14**

The best results are highlighted in bold face

Table 2 The *t* test computed over the results of Table 1

Distance	Add	Prod	Diff	Div
Max	0.02	0.08	0.09	0.19
Add		0.06	0.07	0.17
Prod			0.01	0.12
Diff				0.1

Table 3 The comparison among *Euclidean* (Eucld), *Manhattan* (L1), *Cosine* (Cosn), *Canberra* (Canb) and *Chi-square* (Chisq) distance measures in terms of the ARP values for 5 number of retrieved images

Method	Dataset	Eucld	L1	Cosn	Canb	Chisq
LBpDAD	OASIS	68.23	68.63	67.96	**68.91**	68.67
	TCIA	97.69	97.87	97.71	97.98	**98.00**
	HeLa	78.56	78.91	78.59	**79.63**	79.43
LBpDAD^{FC7}	OASIS	**69.48**	68.65	68.69	67.02	68.95
	TCIA	98.19	98.19	98.18	**98.30**	98.27
	HeLa	81.62	82.18	81.04	80.75	**82.21**
LBpDAD^{FC6}	OASIS	69.20	69.34	69.68	68.68	**69.80**
	TCIA	98.09	98.15	98.06	**98.28**	98.13
	HeLa	74.75	79.78	77.16	**83.29**	79.62

The best results are highlighted by bold style

Table 4 The *t* test computed over the results of Table 3

Distance	L1	Cosn	Canb	Chisq
Eucld	− 0.15	− 0.03	− 0.18	− 0.19
L1		0.12	− 0.03	− 0.04
Cosn			− 0.15	− 0.16
Canb				− 0.01

comparison with the state-of-the-art hand-crafted descriptors (i.e., histogram-based features), we have used *Chi-square* 'Chisq' distance measure in rest of the results of this paper.

The *t* test values for the results of Table 3 are shown in Table 4. It can be seen that the Chi-square distance has the maximum *t* test value as compared to all other distances. The performance of Canberra distance is also close to Chi-square, as suggested by the smallest *t* test value between them. This experiment stamps the choice of Chi-square

distance for the proposed biomedical image retrieval framework.

4.5 Performance analysis using other CNN models

In this experiment, we analyze the suitability of proposed approach with other widely adapted convolutional neural network (CNN) models such as 'Vgg16' [48], 'GoogleNet' [54] and 'ResNet50' [21]. The pre-trained weights of these models available in MATLAB are used. The 'class score' features of these models are considered in this experiment. Similar to *AlexNet*, the original features of these models are referred as *Vgg16*, *GoogleNet* and *ResNet50*. Similar to *LBpDAD*, the local bit-plane decoded CNN descriptors for 'Vgg16', 'GoogleNet' and 'ResNet50' models are denoted by *LBpDVD*, *LBpDGD* and *LBpDRD*, respectively. The retrieval results using ARP (%) versus number of retrieved images are displayed in Fig. 8 for proposed *LBpDAD*, *LBpDVD*, *LBpDGD* and *LBpDRD* descriptors corresponding to 'AlexNet', 'Vgg16', 'GoogleNet' and 'ResNet50' models, respectively. Note that the feature dimension is 1000 in all these descriptors. The results of local bit-plane decoded CNN descriptors fused at 'class score' layer are compared with the original CNN features obtained at 'class score' layer in Fig. 8. All the features are passed through the ReLU operator before use. It is observed through this experiment that the proposed approach is well suited for 'AlexNet', 'Vgg16', 'GoogleNet' and 'ResNet' models over OASIS-MRI and TCIA-CT databases. In case of HeLa database, the performance of *LBpDAD* and *LBpDGD* features is better than the *AlexNet* and *GoogleNet* features. In general, the 'ResNet50' is more discriminative than 'AlexNet', 'Vgg16' and 'GoogleNet', because the last layer features of 'ResNet50' are generated through deep hierarchical transformations.

(a) ARP over OASIS-MRI Database   **(b)** ARP over TCIA-CT Database   **(c)** ARP over HeLa Database
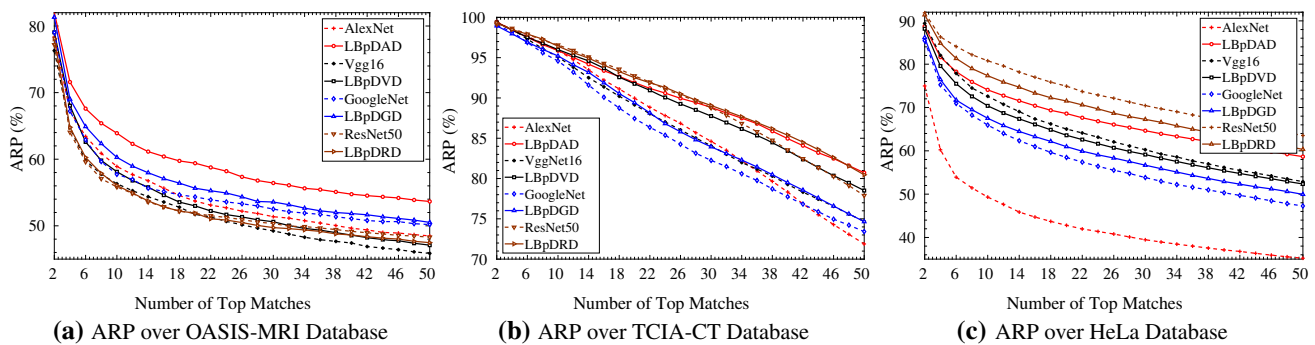
Fig. 8 The results in terms of the ARP versus number of retrieved images by applying the proposed architecture over *AlexNet* [23], *VGG*16 [48], *GoogleNet* [54] and *ResNet*50 [21] models. Here, *AlexNet*, *VGG*16, *GoogleNet* and *ResNet*50 represent the features obtained by applying ReLU over 'softmax' layer. The *LBpDAD*, *LBpDVD*, *LBpDGD* and *LBpDRD* refer to the features obtained by applying ReLU over 'softmax' layer in the proposed architecture corresponding to the *AlexNet*, *VGG*16, *GoogleNet* and *ResNet*50 models

5 Conclusion

A local bit-plane decoding and convolutional neural network-based (CNN) architecture is proposed to produce the image descriptors in this paper. The introduced approach fuses the features at a particular layer of CNN using maximum fusion strategy. The two feature vectors are computed from the raw image and local bit-plane decoded map image. All features are operated by ReLU operator before fusion. The proposed *LBpDAD* descriptor corresponding to the 'AlexNet' model is tested in image retrieval framework over three biomedical databases of different modalities. It is noted that the proposed descriptor outperforms the state-of-the-art biomedical image descriptors. It is also investigated that the performance at 'FC6' layer is generally better than 'FC7' and 'class score' layers. Moreover, the performance of fused features is better than the individual features. Another observation points out that the 'Product'-based fusion strategy is more suitable in the proposed architecture. As per experimental results using different distances, the 'Canberra' distance measure is more appropriate. A favorable observation is made with respect to the proposed architecture with different CNNs such as 'AlexNet,' 'Vgg16,' 'GoogleNet' and 'ResNet50'. The experiments and analysis support the proposed local bit-plane decoding-based CNN descriptor in terms of improved retrieval performance over biomedical images.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Boland MV, Murphy RF (2001) A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of hela cells. Bioinformatics 17(12):1213–1223
2. Carneiro G, Nascimento J, Bradley AP (2015) Unregistered multiview mammogram analysis with pre-trained deep learning models. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin, pp 652–660
3. Chakraborty S, Singh S, Chakraborty P (2016) Local gradient hexa pattern: a descriptor for face recognition and retrieval. IEEE Trans Circuits Syst Video Technol 28:171–180
4. Chakraborty S, Singh SK, Chakraborty P (2017) Local directional gradient pattern: a local descriptor for face recognition. Multimed Tools Appl 76(1):1201–1216
5. Chang YY, Tai SC, Lin JS (2012) Segmentation of multispectral mr images through an annealed rough neural network. Neural Comput Appl 21(5):911–919
6. Chowdhury M, Bulo SR, Moreno R, Kundu MK, Smedby Ö (2016) An efficient radiographic image retrieval system using convolutional neural network. In: 2016 23rd international conference on pattern recognition (ICPR). IEEE, New York, pp 3134–3139
7. Chung YA, Weng WH (2017) Learning deep representations of medical images using siamese CNNs with application to content-based image retrieval. Preprint. arXiv:1711.08490
8. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M et al (2013) The cancer imaging archive (TCIA): maintaining and operating a public information repository. J Dig Imaging 26(6):1045–1057
9. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database. In: IEEE conference on computer vision and pattern recognition, CVPR 2009. IEEE, New York, pp 248–255
10. Dubey SR, Chakraborty S (2018) Average biased ReLU based CNN descriptor for improved face retrieval. Preprint. arXiv:1804.02051
11. Dubey SR, Mukherjee S (2018) Ldop: local directional order pattern for robust face retrieval. Preprint. arXiv:1803.07441
12. Dubey SR, Singh SK, Singh RK (2014) Rotation and illumination invariant interleaved intensity order-based local descriptor. IEEE Trans Image Process 23(12):5323–5333
13. Dubey SR, Singh SK, Singh RK (2015) Local diagonal extrema pattern: a new and efficient feature descriptor for CT image retrieval. IEEE Signal Process Lett 22(9):1215–1219
14. Dubey SR, Singh SK, Singh RK (2015) Local wavelet pattern: a new feature descriptor for image retrieval in medical CT databases. IEEE Trans Image Process 24(12):5892–5903
15. Dubey SR, Singh SK, Singh RK (2016) Local bit-plane decoded pattern: a novel feature descriptor for biomedical image retrieval. IEEE J Biomed Health Inf 20(4):1139–1147
16. Dubey SR, Singh SK, Singh RK (2016) Multichannel decoded local binary patterns for content-based image retrieval. IEEE Trans Image Process 25(9):4018–4032
17. Dubey SR, Singh SK, Singh RK (2016) Novel local bit-plane dissimilarity pattern for computed tomography image retrieval. Electron Lett 52(15):1290–1292
18. Ge Y, Jiang S, Xu Q, Jiang C, Ye F (2017) Exploiting representations from pre-trained convolutional neural networks for high-resolution remote sensing image retrieval. Multimed Appl 77:1–27
19. Guo K, Duan G (2014) 3D image retrieval based on differential geometry and co-occurrence matrix. Neural Comput Appl 24(3–4):715–721
20. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: 2017 IEEE international conference on computer vision (ICCV). IEEE, New York, pp 2980–2988
21. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
22. Hermessi H, Mourali O, Zagrouba E (2018) Convolutional neural network-based multimodal image fusion via similarity learning in the shearlet domain. Neural Comput Appl 30:1–17
23. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems conference, pp 1097–1105
24. Lan R, Zhou Y (2017) Medical image retrieval via histogram of compressed scattering coefficients. IEEE J Biomed Health Inf 21(5):1338–1346

25. Liu P, Guo JM, Wu CY, Cai D (2017) Fusion of deep learning and compressed domain features for content-based image retrieval. IEEE Trans Image Process 26(12):5706–5717

26. Lu K, He N, Xue J, Dong J, Shao L (2015) Learning view-model joint relevance for 3D object retrieval. IEEE Trans Image Process 24(5):1449–1459

27. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL (2007) Open access series of imaging studies (OASIS): cross-sectional mri data in young, middle aged, non-demented, and demented older adults. J Cogn Neurosci 19(9):1498–1507

28. Müller H, Michoux N, Bandon D, Geissbuhler A (2004) A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. Int J Med Inf 73(1):1–23

29. Murala S, Maheshwari R, Balasubramanian R (2012) Local tetra patterns: a new feature descriptor for content-based image retrieval. IEEE Trans Image Process 21(5):2874–2886

30. Murala S, Wu QJ (2013) Local ternary co-occurrence patterns: a new feature descriptor for MRI and CT image retrieval. Neurocomputing 119:399–412

31. Murala S, Wu QJ (2014) Local mesh patterns versus local binary patterns: biomedical image indexing and retrieval. IEEE J Biomed Health Inf 18(3):929–938

32. Nanni L, Brahnam S, Ghidoni S, Lumini A (2018) Bioimage classification with handcrafted and learned features. IEEE/ACM Trans Comput Biol Bioinform. https://doi.org/10.1109/TCBB.2018.2821127

33. Nanni L, Lumini A (2008) A reliable method for cell phenotype image classification. Artif Intell Med 43(2):87–97

34. Nanni L, Lumini A, Brahnam S (2010) Local binary patterns variants as texture descriptors for medical image analysis. Artif Intell Med 49(2):117–125

35. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987

36. Paci M, Nanni L, Lahti A, Aalto-Setala K, Hyttinen J, Severi S (2013) Non-binary coding for texture descriptors in sub-cellular and stem cell image classification. Curr Bioinform 8(2):208–219

37. Pang S, Yu Z, Orgun MA (2017) A novel end-to-end classifier using domain transferred deep convolutional neural networks for biomedical images. Comput Methods Programs Biomed 140:283–293

38. Pietikäinen M, Hadid A, Zhao G, Ahonen T (2011) Computer vision using local binary patterns, vol 40. Springer, Berlin

39. Qayyum A, Anwar SM, Awais M, Majid M (2017) Medical image retrieval using deep convolutional neural network. Neurocomputing 266:8–20

40. Qiu C, Cai Y, Gao X, Cui Y (2017) Medical image retrieval based on the deep convolution network and hash coding. In: 2017 10th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI). IEEE, New York, pp 1–6

41. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K et al (2017) Chexnet: radiologist-level pneumonia detection on chest X-rays with deep learning. Preprint. arXiv:1711.05225

42. Ramírez I, Cuesta-Infante A, Pantrigo JJ, Montemayor AS, Moreno JL, Alonso V, Anguita G, Palombarani L (2018) Convolutional neural networks for computer vision-based detection and recognition of dumpsters. Neural Comput Appl. https://doi.org/10.1007/s00521-018-3390-8

43. Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems conference, pp 91–99

44. Roy SK, Chanda B, Chaudhuri B, Ghosh DK, Dubey SR (2017) A complete dual-cross pattern for unconstrained texture classification. In: 4th IAPR Asian conference on pattern recognition (ACPR 2017), Nanjing, pp 741–746

45. Roy SK, Chanda B, Chaudhuri BB, Banerjee S, Ghosh DK, Dubey SR (2017) Local jet pattern: a robust descriptor for texture classification. Preprint. arXiv:1711.10921

46. Roy SK, Chanda B, Chaudhuri BB, Banerjee S, Ghosh DK, Dubey SR (2018) Local directional ZigZag pattern: a rotation invariant descriptor for texture classification. Pattern Recognit Lett 108:23–30

47. Sezer A, Sezer HB, Albayrak S (2017) Hermite-based texture feature extraction for classification of humeral head in proton density-weighted MR images. Neural Comput Appl 28(10):3021–3033

48. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. Preprint. arXiv:1409.1556

49. Singh GAP, Gupta P (2018) Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. Neural Comput Appl. https://doi.org/10.1007/s00521-018-3518-x

50. Sirinukunwattana K, Raza SEA, Tsang YW, Snead DR, Cree IA, Rajpoot NM (2016) Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Trans Med Imaging 35(5):1196–1206

51. Sorensen L, Shaker SB, De Bruijne M (2010) Quantitative analysis of pulmonary emphysema using local binary patterns. IEEE Trans Med Imaging 29(2):559–569

52. Srivastava D, Rajitha B, Agarwal S et al (2018) Pattern-based image retrieval using GLCM. Neural Comput Appl. https://doi.org/10.1007/s00521-018-3611-1

53. Suri JS, Wilson D, Laxminarayan S (2005) Handbook of biomedical image analysis, vol 2. Springer, Berlin

54. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A et al (2015) Going deeper with convolutions. In: CVPR

55. Tan X, Triggs B (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans Image Process 19(6):1635–1650

56. Wang DH, Conilione P (2012) Machine learning approach for face image retrieval. Neural Comput Appl 21(4):683–694

57. Wang G, Xu X, Jiang X, Ding S (2016) Medical image registration based on self-adapting pulse-coupled neural networks and mutual information. Neural Comput Appl 27(7):1917–1926

58. Wang Z, Fan B, Wu F (2011) Local intensity order pattern for feature description. In: 2011 IEEE international conference on computer vision (ICCV). IEEE, New York, pp 603–610

59. Wu Y, Wang L, Cui F, Zhai H, Dong B, Wang JY (2016) Cross-model convolutional neural network for multiple modality data representation. Neural Comput Appl 30:1–11

60. Yang Z, Yu W, Liang P, Guo H, Xia L, Zhang F, Ma Y, Ma J (2018) Deep transfer learning for military object recognition under small training set condition. Neural Comput Appl. https://doi.org/10.1007/s00521-018-3468-3

61. Yao J, Liu F, Geng Y (2017) Query-specific optimal convolutional neural ranker. Neural Comput Appl. https://doi.org/10.1007/s00521-017-3257-4

62. Zhang B, Gao Y, Zhao S, Liu J (2010) Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor. IEEE Trans Image Process 19(2):533–544

63. Zhang L, Lu L, Nogues I, Summers RM, Liu S, Yao J (2017) Deeppap: deep convolutional networks for cervical cell classification. IEEE J Biomed Health Inf 21(6):1633–1643