Journal of Intelligent & Fuzzy Systems 38 (2020) 115–126
DOI:10.3233/JIFS-179386
IOS Press

An image retrieval method of mammary cancer based on convolutional neural network

Dan Wang[a,b], Hongwei Zhao[a,c,e,*] and Qingliang Li[d,e]
[a]*Department of Computer Science and Technology, Jilin University, Changchun, China*
[b]*College of Information Technology and Media, Beihua University, Jilin, China*
[c]*State Key Laboratory of Applied Optics, Changchun, China*
[d]*Changchun University of Science and Technology, Changchun, China*
[e]*Department of Symbolic Computing and Knowledge Engineering, Key Laboratory of the Ministry of Education, Jilin University, Changchun, China*

Abstract. This paper designs a brand-new image retrieval method of mammary cancer based on convolution neural network. This method simulates VLAD layer in CNN network structure, designs a trainable universal VLAD layer-NET VLAD layer, reduces dimensions and optimizes VLAD descriptors, applies structure from motion algorithm to automatically label samples, and obtains the minimum loss function value by a new training program of weakly supervised ranking loss. Experiments show that this method has improved retrieval performance compared with similar retrieval methods and non-network structure retrieval methods.

Keywords: Convolutional neural network, VLAD, loss function, medical image retrieval

1. Introduction

Along with the rapid development of pattern recognition, machine learning and other technologies in the field of artificial intelligence, all levels of society and economy expect the human world and computers to establish a deep level of communication and interaction. The medical industry urgently needs computers that can functionally approximate the human visual perception mechanism to intelligently carry out scientific analysis and management of massive medical image data of various modes. Medical image processing and analysis emerge as the times require and meet the needs of the development of the medical industry. Currently, medical image processing and analysis technology is widely used in the medical industry. The research and development of Computer aided detection (CADe) system and Computer aided diagnosis (CADx) system around medical images have achieved remarkable results [1]. Medical image retrieval technology can quickly and efficiently complete query tasks in large-scale medical image data and provide diagnostic advice for medical personnel to diagnose lesions and follow up the progress of conditions, which is an important function in Computer-aided diagnosis system. In clinical work, for diseases that are difficult to be diagnosed, similar cases in the medical image database can be combined to assist diagnosis, thus eliminating the heavy and complicated manual inquiry process and improving the efficiency and accuracy of medical diagnosis. Mammary cancer is one of the most common tumor

*Corresponding author. Hongwei Zhao, Department of computer science and technology, Jilin University, Changchun, China. E-mail: wangdanjl_dx@163.com.

pathologies among women. More than 1 million women around the world suffer from mammary cancer within one year, and half of them die from the symptom [2]. The morbidity of female mammary cancer in China also keeps pace with that in developed western countries. In the past 10 years, the morbidity of mammary cancer has increased by more than 30% in major cities in China. The morbidity of mammary cancer among women is higher in urban than in rural areas [3]. How to reduce the workload of clinicians and successfully assist doctors in the diagnosis of mammary cancer is a difficult problem that needs to be solved urgently.

This paper is divided into three parts. The first part introduces the research and application status of image retrieval field of mammary cancer. The second part is the construction process of medical image retrieval method of mammary cancer based on convolutional neural network. The third part is to verify the effectiveness of medical image retrieval method of mammary cancer based on convolutional neural network, as well as design the training process of the method. Experiments show that the proposed method improves the retrieval performance of mammary cancer images to certain extent.

2. Related work

So far, lots of scholars and researchers have made many outstanding contributions in mammary cancer retrieval. Pohlman, et al. used growth method in adaptive region to detect the tumor region, and then extracted various morphological features in the region to achieve the retrieval task [4]. Hadjiiskid, et al. used linear discriminant to analyse dimension reduction method and unsupervised classification method to classify masses [5]. Guliato, et al. and Rangayyan, et al. firstly manually segmented the contour of mammary image tumor, and then realized tumor retrieval based on the segmented contour information to judge whether it is benign or malignant. In which geometric angle information is extracted as feature in contour and texture information is extracted based on gray level co-occurrence matrix [6, 7]. Sahiner, et al. took the tumor region as a rectangle, then extracted texture features based on gray level co-occurrence matrix, and finally, adopted the unsupervised classification method to complete the tumor retrieval [8]. Sahiner, et al. proposed a fully automatic segmentation method for masses, and then completed the retrieval [9]. Cheng, et al. described the automatic detection of masses and the automatic judgment of benign and malignant tumors [10]. At present, there are not many mammary searches in China. Xu used neural network structure to detect the benign and malignant tumor region and its morphological information, but the tumor itself is complicated, and the effects of various diagnostic methods are not ideal [11]. At present, the methods are roughly classified: one is fully automatic mode [12], and the other is completely manual labeling [13]. In the full-automatic algorithm, there has not been a robust algorithm that can accurately determine the complex tumor region. In the completely manual labeling method, even if the accuracy is significantly improved, the workload of doctors is greatly increased, which will cause doctors to make mistakes in the critical diagnosis process. Therefore, it is urgent to reduce the workload of doctors' manual labeling and improve the accuracy of diagnosis.

Since the acquisition technology of medical images is increasingly mature, it is very easy to acquire medical images, but it is difficult to diagnose images. Generally, self-built medical databases contain various image information of which labeled tumor information accounts for a small proportion of unlabeled tumor information. If these unlabeled tumor information images can be fully utilized, it will bring great help to medical computer-aided diagnosis, but there are not lots of research results in this area. This paper mainly studies the semi-automatic segmentation of mammary cancer region by doctors and combines with neural network architecture to make full use of unlabeled mammary cancer images as well as reduce the manual workload of doctors and improve the diagnosis of benign and malignant tumors in large-scale medical image database of mammary cancer.

3. The image retrieval method of mammary cancer based on NetVLAD

3.1. Method

According to the professional characteristics of current medical mammary cancer images, this paper constructs a retrieval method framework for mammary cancer medical images based on convolution neural network theory, designs a NetVLAD structure, and realizes the optimized loss function. In the traditional VLAD retrieval framework, the end-to-end learning method of convolution neural network

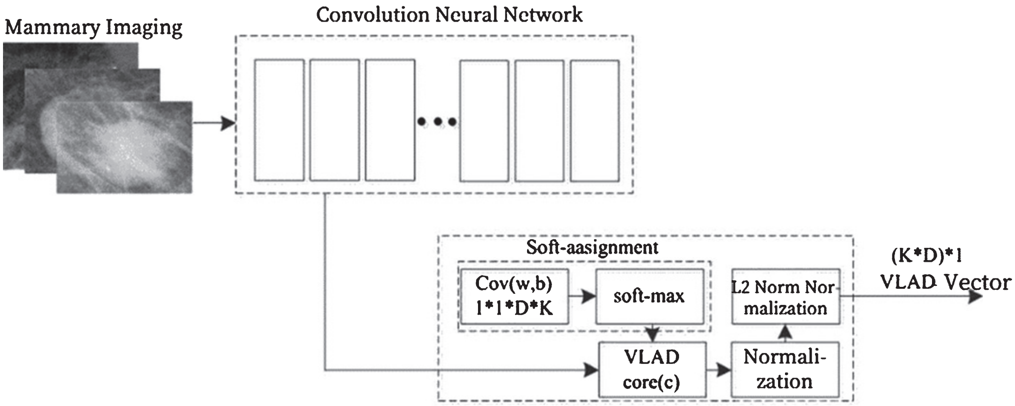

Fig. 1. Setting for document template.

structure is realized, which enhances the retrieval performance of mammary cancer images, as shown in Fig. 1.

The main research contents of this paper mainly contain the design of the new NetVLAD layer, automatic calibration of samples, dimension reduction of feature descriptors and optimization of loss function. And then, the main links of the method designed in this paper will be described respectively.

3.2. VLAD algorithm

VLAD descriptor is a brand-new image vector representation method proposed by Jegou et al., which realizes the aggregation of feature descriptors in feature space [14]. VLAD descriptor can be understood as a feature vector with a fixed coding length in a local feature descriptor subset, and the local feature descriptor subset mentioned here consists of N D-dimensional local feature descriptors. The basic workflow is as follows: firstly, the training process of visual words with the size of k is completed in the feature space through the K-means clustering method and then each feature descriptor x is matched to the visual word closest to it, with the following calculation formula:

$$\mathrm{q} : \Re^d \to \mathrm{C} \subset \Re^d \tag{1}$$

$$\mathrm{x} \mapsto \mathrm{q(x)} = \arg\min_{\mu = C} \|x - \mu\|^2 \tag{2}$$

$\|\cdot\|$ in the formula (2) represents a normalization operation method, and the norm is L2.

The main working principle of VLAD descriptors is as follows: First, for any visual word μ_i containing several local feature descriptors, the residual vector $x - \mu_i$ of these descriptors, i.e. quantization error, is calculated first, and then these residual vectors are accumulated. The distribution of local feature descriptors corresponding to the visual word space can be described through this process.

Here, if the vector dimension of the local feature descriptor is d, then the vector dimension of the VLAD descriptor should be $D = \mathrm{k} \times \mathrm{d}$. In the following formula, vi is used to represent VLAD, and i is used to represent the components of visual words. The value range of i is $i = 1, 2, \ldots, \mathrm{k}$. Here, the component value of descriptors is determined by the vector difference generated by visual words and aggregated local feature descriptors. The calculation method is shown in formula (3):

$$v_i = \sum_{x:q(x)=\mu_i} x - \mu_i \tag{3}$$

Then VLAD $V = [\mathrm{v1}, \ldots, \mathrm{vk}]$ represents last dimension that is the descriptor of d.

Then, the VLAD descriptors constructed above still need to be normalized by two methods. The first method is number normalization, and the calculation formula is $v_i := |v_i|^\alpha \times sign(v_i)$. Also, the parameter setting in the formula need satisfy $\alpha \leqslant 1$. The second method is L2 norm normalization, and the calculation formula is $v := v / \|v\|$.

Figure 2 below shows VLAD descriptors for five images. From the contents of the figure, it can be seen that there are obvious differences in appearance between the original image and the matching result image, such as changes in scale, color and rotation. However, from the perspective of VLAD, the original image and the matching result image have

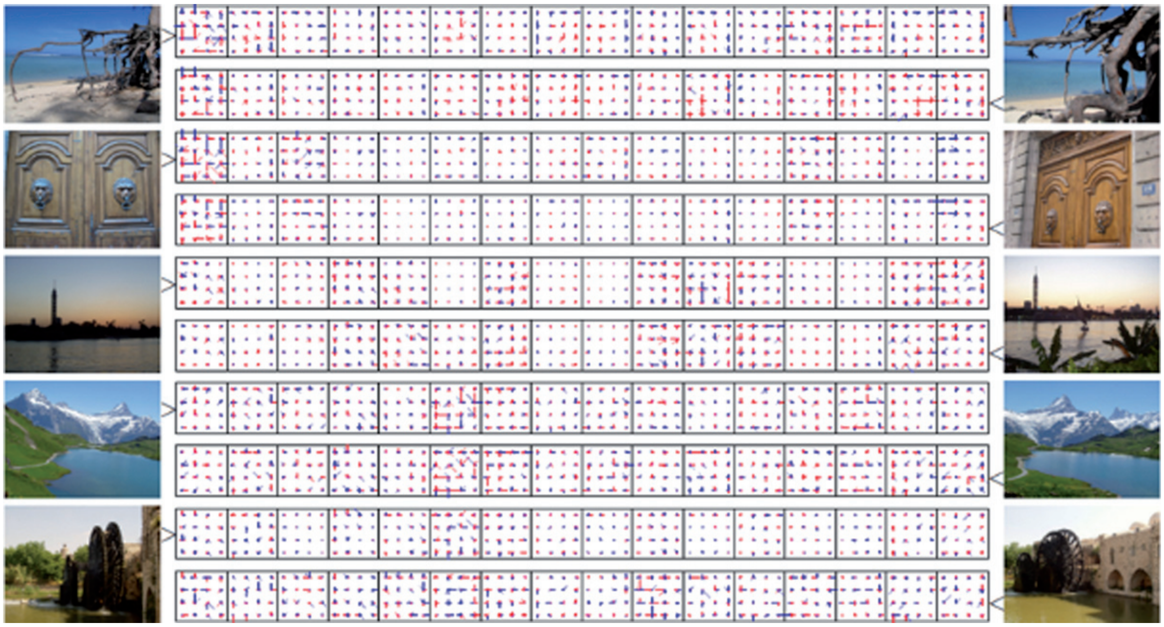

Fig. 2. VLAD of five images (This figure cited from [26]).

good similarity. There are very few components with larger values, which indicate obvious sparseness. The higher values of most descriptors are concentrated on one visual word, which shows that the structure is quite special. The above characteristics are very beneficial to the application of principal component analysis to dimension reduction of image feature descriptors.

Visual word -Fisher Vector uses Mixture Gaussian Mode to calculate image visual features [15]. Different from visual word method, this method introduces the distance between feature vectors and clustering centers in the process of modeling, but this method still loses relevant information in the feature description stage. In the VLAD feature frame similar to the visual word method, the residual vector mainly takes into account the residual of the feature vectors and the clustering center in the process of extracting feature vectors to preserve the local feature information of the image as much as possible. And VLAD retrieval framework is better than Fisher Vector and visual word method.

3.3. The architecture theory of netVLAD network layer

It is assumed that the convolutional neural network architecture includes D H×W dimensional feature maps, which can be regarded as a set of multidimensional feature vectors. In this section, the dimension expressed as feature vectors is H×W, while the dimension information of the third layer of feature vectors is D dimension.

Assuming that there are N D-dimensional clustering centers $c = \{ck\}$, the NetVLAD network layer mainly refers to the idea of VLAD feature vectors in a certain layer of the network structure to describe the image I. The reason why the network layer of the neural network structure can adopt the idea of VLAD is to make the NetVLAD network layer derivable. After the neural network structure is optimized to the NetVLAD structure, each parameter of the network structure needs to be optimized through reverse calculation

For any characteristic $xi \in x$, $ak(x_i) \in \{0, 1\}$, the values of $ak(x_i)$ are discrete distributions of 0 and 1 respectively, making the matrix $V(j, k)$ nondifferentiable. In order to change this phenomenon and make the values of $ak(x_i)$ continuous, the following adjustments are made:

$$\overline{a_k}(X_i) = \frac{e^{-a\|x_i - c_k\|^2}}{\sum_{k'} {}^{-a\|x_i - c_k\|^2}} \tag{4}$$

The above formula shows that x_i is an element of matrix $V(j, k)$, and the description of matrix $V(j, k)$ by $\overline{a_k}(x_i)$ is proportional to the distance between x_i and the center c_k of the matrix.

The term $e^{-a\|x_i\|^2}$ can be removed from formula (4) to obtain

$$\overline{a_k}(X_i) = \frac{e^{w_k^T x_i + b_k}}{\sum_{k'} e^{w_k^T x_i + b_k}} \tag{5}$$

In the formula, $W_k = 2ac_k, b_k = -a \parallel c_k \parallel^2$.

In the formula, $\{W_k\}$, $\{b_k\}$, and $\{c_k\}$, the three parameters can be optimized through inverse calculation. The original VLAD feature calculation process only contains $\{c_k\}$ and cannot be optimized. Compared with the original VLAD feature descriptor, the VLAD feature integrated into the neural network structure has three independent parameters$\{W_k\}$,$\{b_k\}$, and $\{c_k\}$. As this network structure optimizes each parameter in an end-to-end manner, the retrieval performance of the VLAD feature integrated into the neural network structure is better than that of the original VLAD feature descriptor.

3.4. Network structure of CNN-VLAD

The network structure of CNN-VLAD is shown in Fig. 1 above, and consists of basic neural network structure and network structure of NetVLAD.

In general, given N groups of D-dimensional local image descriptors X as inputs, k clustering centers ("visual words") C_k as VLAD parameters, and the output VLAD descriptors as image descriptions, where V is the feature descriptor of the k×d dimension. For convenience, we take V as a k×d dimension, but this matrix is converted into a vector, which is normalized and used as an image representation. Matrix V(j,k) is calculated as follows:

$$V(j,k) = \sum_{i=1}^{N} a_k(x_i)(x_i(j) - c_k(j)) \tag{6}$$

Thereinto $x_i(j)$ and $c_k(j)$ are the j-th descriptor and the I-th dimension of the k-th cluster center, respectively. $a_k(x_i)$ represents an element from the descriptor x_i to the k-th visual word, i.e., if the cluster center c_k is the cluster center closest to the descriptor x_i, the value is 1. It is obvious that the k-th column of each D-dimensional matrix of V records the residual sum$(x_i - c_k)$of descriptors assigned to c_k. Then, the matrix V is normalized by L2 norm (internal normalization) according to columns, and a group of vectors are converted. Finally, the features normalized by L2 norm are used as feature description information.

In order to benefit from previous image retrieval tasks, here we simulate VLAD layer in CNN network structure and design a trainable universal VLAD layer called NetVLAD. The result is that there is a robust image representation on the target task (this paper mainly focuses on the mass detection of mammary cancer images), which can be used for end-to-end training. In order to construct a layer that can be easily trained by back propagation, the operation of this layer is required to be differentiable in all its parameters and inputs. Therefore, the key challenge is to make VLAD layer distinguishable.

Layer k of the convolutional neural network contains D H×W feature maps, which are considered as H×W D dimension feature descriptors in the network, which are used as input of NetVLAD. The traditional neural network can calculate the set of layer k feature maps of N database images through the forward propagation order. The K-means clustering algorithm in VLAD can calculate k D-dimensional clustering centers. This input can be regarded as the initialization parameter of NetVLAD network structure. In Fig. 1, the dotted line in the NetVLAD network is used to calculate the corresponding weight. In this part, the matrix $V(j,k)$ calculated by the VLAD descriptor needs L2 norm normalization. The normalized matrix is expanded into K×D×1-dimensional neighbors, and the expanded vector is subjected to L2 norm normalization to obtain the final VLAD descriptor.

3.4.1. Feature dimension reduction and optimization of VLAD

This method is similar to the deep network layer of VGG-16, as shown in Fig. 3. The figure consists of 5 convolutional structures, 3 full connection structures and one output layer. The five convolutional structures are separated by a pooling layer.

Here, the feature map of conv5_3 is used as the input of NetVLAD network layer in Fig. 1. The input is for 512 feature maps, i.e. the input feature is 512 dimensions. When $K = 50$, the dimension of VLAD feature vector is 20K. Therefore, when VLAD is integrated into the convolutional neural network architecture, VLAD descriptors have too high feature dimensions and very low computational efficiency, which requires dimension reduction.

Assuming that$I = \{I_k\}$ is a set of mammary images, and f_θ represents the calculation process of the VLAD feature descriptor proposed by the algorithm in this paper, then analysis method of the principal component is used to reduce the dimension of the high-dimensional VLAD descriptor. The calculation process is as follows:

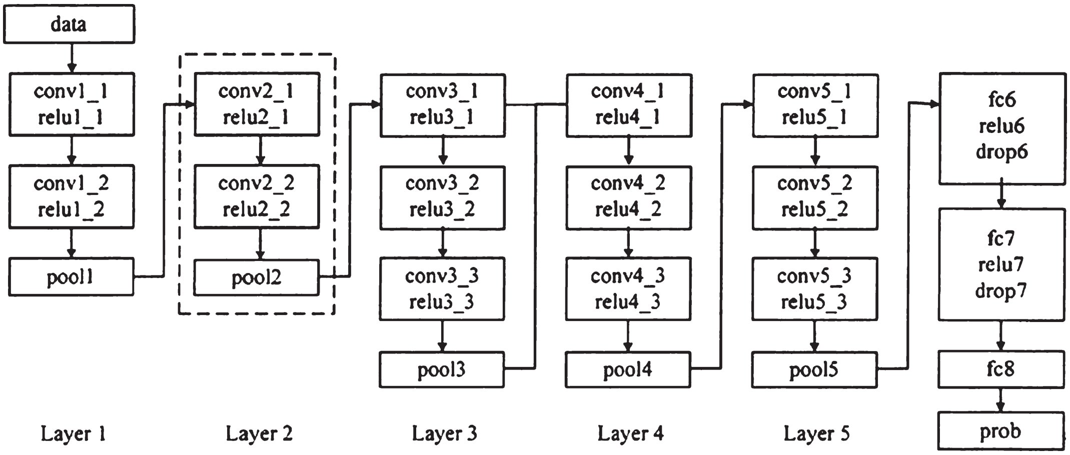

Fig. 3. VGG-16 network structure diagram.

Firstly, VLAD is integrated into a convolutional neural network to obtain a VLAD feature descriptor of K×D×1:

$$S_k = f_\theta(I_k) \tag{7}$$

Based on the conversion of VLAD descriptor S_k into VLAD eigenvector matrix $S^{M\times N}$ in (1):

$$S^{N\times M} = [S_1; S_2 ...; S_N] \tag{8}$$

In the formula, M = K×D, N represents the total number of database images, and S_k represents the VLAD descriptor of imageI_k.

Calculate covariance matrix $C_{i,j}$ for VLAD eigenvector matrix$S^{M\times N}$:

$$C_{i,j} = \frac{\sum_{i=1}^{N}(s_i - \overline{s_i})(s_j - \overline{s_j})}{N-1} \tag{9}$$

(4) Calculate the eigenvalues γ_i and eigenvectors $x_i$of covariance matrix$C_{i,j}$, reorder the eigenvalues, and calculate eigenvalue matrix $L^{M\times N}$ and eigenvector matrix $H^{M\times N}$ from small to large.

(5) Introduce noise processing

$$H^{M\times M} = H^{M\times M} * \left(1 \Big/ \sqrt{L^{M\times M}}\right) \tag{10}$$

(6) Take the matrix $H^{M\times P}$ as the submatrix of $H^{M\times M}$, and meanwhile take the matrix S' as the dimension reduction matrix of $S^{M\times N}$:

$$S' = [S'_1, S'_2, ..., S'_N] = S^{N\times M} \cdot H^{M\times P} = S^{N\times P} \tag{11}$$

In the formula, S'_K is the final dimension reduction feature.

3.4.2. *Sample calibration and loss function of CNN-VLAD*

Structure from Motion algorithm is usually widely used in three-dimensional reconstruction tasks. This paper applies it to mammary-affected mass retrieval in an attempt to increase the discrimination ability of mammary image features.

The calculation flow of the Structure from Motion algorithm is as follows:

(1) Firstly, image features are detected, and for any image $I_i \in I$, a set $\{x_i^k\}$ of feature point position relations and a subset $\{dec_i^k\}$ of feature descriptions are calculated based on SIFT feature extraction algorithm.
(2) Matching between feature points: for any two images, the basic matrix and homography matrix of the image are calculated according to the feature matching and geometric constraint RANSAC method, and the 3-dimensional point cloud of the image is recovered according to the external parameters obtained from the image. Generally, the number of matching features is selected to be acquired in at least 100 pairs of images
(3) 3D point cloud reconstruction: it is assumed that a set of matching feature sets store all matching feature pairs. After selecting an image, the matching relation between the image and the reconstructed image and the external parameters of the image are used to recover the 3-dimensional information of the image, and the 3-dimensional point cloud data is added into the Structure from Motion algorithm.

For the query image, the former K images similar to the query image are first calculated in the database. Meanwhile, the matching relationship between the 2-dimensional information and 3-dimensional information of the K images is calculated based on the Structure from Motion algorithm. Given two images I_k and I_j, if one of the images I_k is related to the image I_j then I_k is an image similar to the image, or vice versa. Assuming that the query image is defined as q, the similar image set of its positive sample is $\{p_i^q\}$ and the non-similar image set of its negative sample is $\{n_j^q\}$ so the ternary set of all training images in the database can be described according to $\left\{q, \{p_i^q\}, \{n_j^q\}\right\}$. Here, the method of calculating features in this section is described by f_θ. For $\left\{q, \{p_i^q\}, \{n_j^q\}\right\}$, VLAD feature descriptors shall conform to the following formula:

$$d_\theta(q, p_i^q) < d_\theta(q, p_j^q) \forall i, j \tag{12}$$

Thereinto

$$d_\theta\left(q, p_i^q\right) = \| f_\theta(q) - f_\theta(p_i^q) \|^2 \tag{13}$$

$$d_\theta\left(q, n_j^q\right) = \| f_\theta(q) - f_\theta(n_j^q) \|^2 \tag{14}$$

In general, the Euclidean distance between the queries image q and the positive sample image is preferably less than the Euclidean distance between the query image q and the negative sample image.

It is necessary to use convolutional neural network in large-scale mammary image data collection, but it will consume a lot of manpower and financial resources to use artificial pre-judgment for mammary image tumors. Therefore, this paper introduces a Structure from Motion algorithm to avoid the manual annotation process.

Assuming that the image set is I_i, the method structure of Structure from Motion algorithm is adopted. The point cloud information of $= \{P_i\}$ Structure from Motion algorithm usually needs to be extracted from the original data of mammary images. At present, most medical image retrievals are based on 2-dimensional graphic information, such as CT images and MRI images, which are often transformed from the original 3-dimensional images by hierarchical mapping. $\{P_i\}$ represents the 3-dimensional point cloud index of image I_i. For any influence, $I_i \in I$, $I_j \in I$, $P_{i,j} = P_{I_1} \cap P_{I_2}$.

In the definition of loss function, Triplrt Loss [17] is adopted and calculated as follows:

$$L_\theta = \sum\nolimits_{i,j} l(d_\theta(q, p_i^j) + m - d_\theta(q, n_j^q)) \forall i, j \tag{15}$$

Thereinto m is constant, $1(x) = \max(x, 0)$, and the triplets in the image set can be defined as $\{q, \{p_i^q\}, \{n_j^q\}\}$. However, due to $L_\theta = 0$, the training efficiency and speed of the network structure are greatly improved.

The automatic calibration of the Structure from Motion algorithm is different from the manual calibration. Due to the subjective influence of long-term manual calibration, the calibrated positive sample p_i^q will repeat with the image Q. The automatic labeling of the Structure from Motion algorithm will also bring many errors.

For formula:

$$p_{i*}^q = \arg\min_{p_i^q} d\theta(q, p_i^q) \tag{16}$$

The ideal case $d\theta(q, p_i^q)$ is less than q and any negative sample distance, i.e.:

$$d\theta(q, p_{i*}^q) < d\theta(q, n_j^q) \forall j \tag{17}$$

The loss function is defined as follows:

$$L_\theta = \sum\nolimits_j l(\min_i d_\theta^2(q, p_i^q) + m - d_\theta^2(q, n_j^q)) \tag{18}$$

Thereinto m is a positive number, l (x)=max (x, 0).

For triplets $\{q, \{p_i^q\}, \{n_j^q\}\}$, assuming that the Euclidean distance between image q and positive samples is less than the Euclidean distance between image q and negative samples, the difference is greater than m, and the value of loss function is greater than 0. When optimizing other parameters of the network structure, the parameters are optimized by the gradient descent method to make loss function reach the minimum value [18].

4. Experimental design and analysis

4.1. Image database of medical mammary cancer

In the experimental part of this paper, the database of the paper is selected for testing. The data is 11,617 mammary X-ray images randomly selected from DDSM database to verify the effectiveness of the algorithm [19]. Among them, 1273 mammary X-ray images are malignant tumors and 1127 mammary X-ray images are benign tumors. Meanwhile,

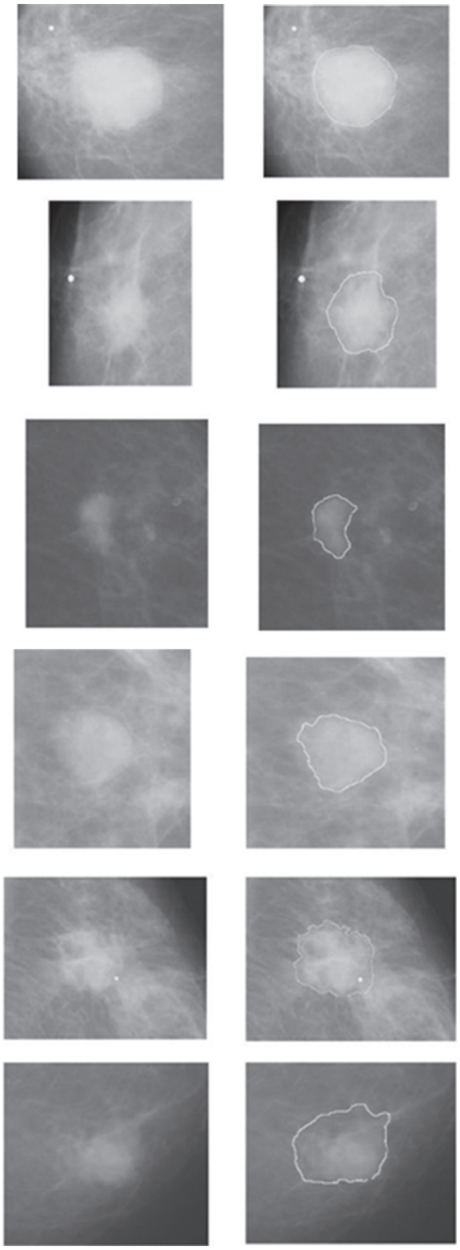

Fig. 4. VGG-16 network structure diagram.

9217 mammary X-ray images are normal mammary regions marked by radiologists. This database is all the regions of interest of lumps in mammary images. The density and size of lesion areas in mammary images are different, which indicates that the lesion degree is also different. The size and location of malignant and benign lumps in mammary images are randomly distributed. The segmentation display of mammary image lump and lump in the database is shown in Fig. 4.

Table 1
VGG-16 Training Parameters

Margin	Epoch	Momentum
0.8	50	0.8
Weight	Learning Rate	Batch size
0.005	0.0005	5

4.2. Experimental process and relevant parameter setting

The algorithm in this paper is tested in MATLAB environment with a full-automatic processing process as well as no user participating in manual operation.

Firstly, the network structure of this paper needs to be initialized, because VLAD-CNN network structure mainly consists of two layers of networks. The traditional convolutional neural network layer is tested with VGG-16 and AlexNet proposed above. The features calculated for con_5_3 of VGG-16 network layer are used as input information of NetVLAD layer. Similarly, AlexNet's con_5 layer feature map serves as input information for NetVLAD layer.

The experimental design process of VGG-16 network layer in generating cluster center $\{c_k\}$ is as follows:

(1) Randomly select n images in the mammary image database for training $T = \{T_i\}$.
(2) The data in the training set adopts the characteristics of VGG-16 network structure disclosed in [45] on ImageNet. Furthermore, the features given at the con_5_3 layer generating a clustering center $\{c_k\}$ based on the k-means clustering algorithm.

In VLAD feature descriptor extraction, $\bar{a}_k(x_i) \in \{0, 1\}$, and $\bar{a}_k(x_i)$ values range from 0 to 1, a$\longrightarrow \infty$, $\bar{a}_k(x_i)$ is equivalent to $a_k(x_i)$I, and usually a larger value is set for a, which can make $\bar{a}_k(x_i)$ more sparse.

The selection of training parameters for VGG-16 network structure is shown in Table 1.

The consumption of VGG-16 network structure and GPU is much larger than AlexNet, so it is unrealistic to optimize all parameters of VGG-16 network structure. Since this paper only uses the con_5_3 network layer of VGG-16 as the input feature of the NetVLAD network layer, the optimization parameters are also limited to the con_5_3 network layer of VGG-16. Specifically, the characteristic map of the con_5_3 network layer of VGG-16 is used as the input characteristic of the NetVLAD network layer. In reverse propagation, only the ownership value information of the con_5_3 network layer, con_5_2 network layer, con_5_1 network layer and c_k in the cluster are optimized.

For triplets $\{q, \{p_i^q\}, \{n_j^q\}\}$, images of normal mammary tissue structure are selected from the set $\{n_i^q\}$, and features are extracted by using the network structure proposed in this paper. Finally, 50 groups of images which are close to the Euclidean distance of VLAD feature descriptors are selected. According to the above structure, any triplet $\{q, \{p_i^q\}, \{n_j^q\}\}$ needs to extract thousands of VLAD feature descriptors, which leads to a very slow running speed of the network structure. Therefore, it can be defined that VLAD feature descriptors are not updated step by step, but the update frequency is set according to the size of the mammography library, and the performance and training speed of the network structure proposed in this paper under different frequencies are measured. The frequency is roughly set to 500-1000 cycles to update the VLAD feature descriptor once.

4.3. Experimental results and analysis

In this paper, the performance of the algorithm in DDSM database is analyzed from three aspects: the change of different cycle times and loss function values, the accuracy of mammary image retrieval and the actual effect of mammary image retrieval.

First of all, regarding the numerical change of loss function, for triples $\{q, \{p_i^q\}, \{n_j^q\}\}$ and their loss functions:

$$\bar{L}_\theta = \sum_j l(\min_i d_\theta^2(q, p_i^q) + m - d_\theta^2(q, n_j^q))/\#(n_j^q) \tag{19}$$

The experimental screenshots of different cycles and loss function values are shown in Fig. 5.

In Fig. 5, the loss function value decreases as the number of iterations increases, and the loss function value tends to be stable after 10 iterations. Experimental results show that training network structure can effectively reduce the Euclidean distance between the image and negative samples, and increase the Euclidean distance between the image and positive samples. Mammary image data and the network structure of this section correspond to each other and can be well trained.

In this section, after obtaining the optimal number of iterations, the retrieval performance of mammary images is tested by two measurement methods: retrieval recall rate of mammary images and retrieval mAP.

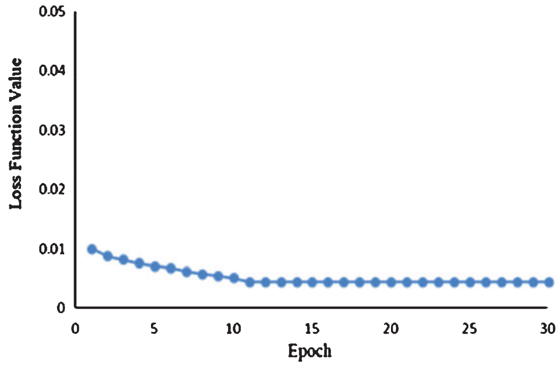

Fig. 5. Variation curve of loss function.

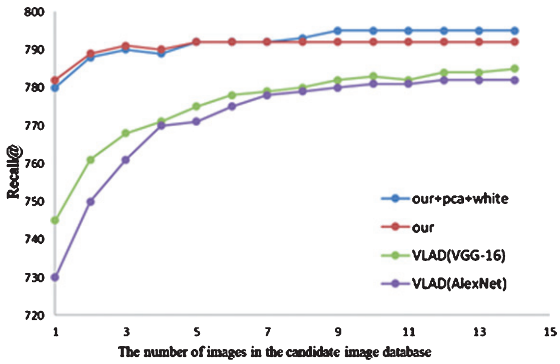

Fig. 6. The retrieval recall rate of mammary images.

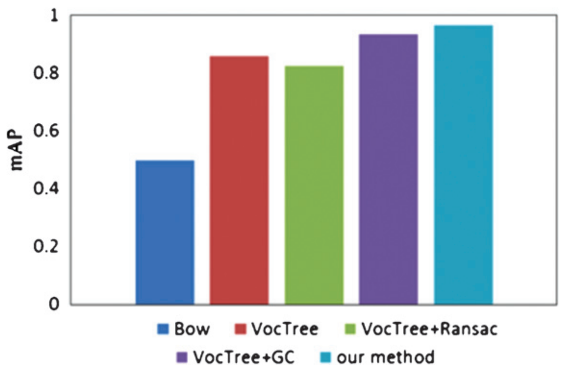

Fig. 7. Retrieval mAP of mammary images.

Firstly, VGG-16 is used in the structure layer of traditional neural network. At the same time, in order to prove the advantages of the network structure proposed in this paper, it is analyzed separately with other network structures that are popular at present.

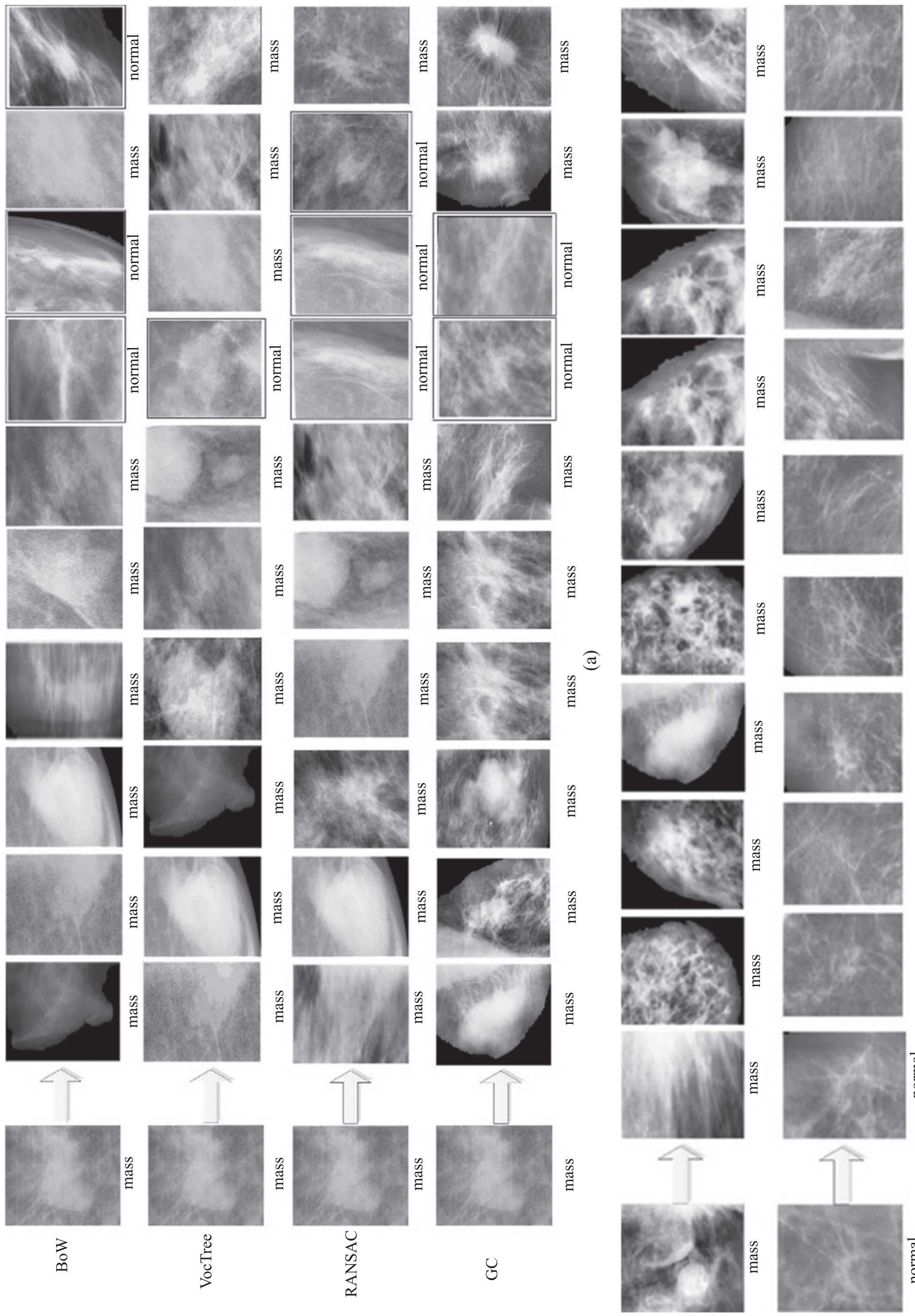

Fig. 8. Example of mammary image retrieval.

The experimental analysis in Fig. 6 shows that:

(1) The performance of the network architecture proposed in this paper is better than other methods, especially when the number of images in the candidate image database is less than five, and the performance comparison is more significant
(2) The algorithm in this paper integrates PCA dimensionality reduction algorithm and optimization algorithm to reduce the dimensionality of high-dimensional features, which greatly advances the calculation speed on the premise that the accurate retrieval obtained by high-dimensional features is equivalent.
(3) The VLAD-CNN network based on VGG-16 network structure has significantly better retrieval accuracy than AlexNet's VLAD-CNN network, because the internal structure of VGG-16 network structure is significantly more complex than AlexNet's internal structure in comparison of network structures.

Compared with other network structures, the algorithm proposed in this paper is determined to be more effective. At the same time, in the experimental stage, the algorithm proposed in this paper is tested by the mAP measurement method of mammary images, and compared with the traditional non-network structure mammary image retrieval method, namely the traditional visual word bag method [20], the method of integrating lexical tree into mammary image retrieval method [21], the method of integrating RANSAC matching feature removal algorithm into mammary image retrieval method [22], and the method of integrating geometric verification algorithm into mammary image retrieval method [23–25].

According to the above experimental analysis, as shown in Fig. 7, the algorithm proposed in this paper has higher retrieval performance than other mammary image retrieval methods without network structure.

Finally, some actual retrieval examples of mammary images are analyzed. The experimental results are shown in Fig. 8. Thereinto, Fig. 8(a) stands for an example of mammary retrieval of other algorithms compared, in which the traditional visual word bag method, the method of integrating word tree into mammary image retrieval, the method of integrating RANSAC matching feature removal algorithm into mammary image retrieval, and the method of integrating geometric verification algorithm into mammary image retrieval, the first 10 retrieval results returned by these four methods in mammary image database have error diagnosis. Figure 8(b) represents the retrieval results proposed by the algorithm in this paper, in which the mammary images queried not only wrap up the mass information, but also the mammary images of normal tissues are tested in this paper, and the first 10 retrieval results returned are all correct.

5. Conclusion

In this paper, a mammary cancer image retrieval algorithm based on NetVLAD is proposed, which is integrated into the neural network architecture by analyzing the advantages of traditional VLAD. The aim is to study a new algorithm of computer-aided mammary retrieval for large-scale mammary image database, and try to solve the problems existing in the current algorithm in order to provide an effective retrieval algorithm for mammary cancer diagnosis system. Firstly, convolutional neural network (CNN) architecture is designed, which can directly train tumor recognition tasks in an end-to-end manner. NetVLAD, the main component of the architecture, is a new general VLAD layer. Its inspiration comes from the "vector of local aggregation descriptors" image representation commonly used in image retrieval. This layer can be easily inserted into any CNN architecture and trained through back propagation. Secondly, we have developed a training program based on a new weak monitoring ranking loss to learn the parameters of the architecture in an end-to-end manner. Finally, we indicate that the proposed architecture is significantly better than the unlearned image representation and the ready-made CNN descriptor in mammary cancer medical image retrieval, and improves the current most advanced compact image representation on the standard image retrieval benchmark.

Finally, through the experimental design and analysis, it can be seen that the algorithm proposed in this paper not only promotes the retrieval performance compared with the current relatively flow network structure, but also compared with the non-network structure mammary image retrieval technology to a certain extent, which provides strong support for computer-aided diagnosis of mammary tumors. A future direction of overall integration of the system is possible [23].

Acknowledgments

This project is supported by The Education Department of Jilin Province, "13th Five-Year" science and technology project "based on behavior recognition method adaptive multiple features fusion research" (Project No. JJKH20170037kj); The National Natural Science Foundation of China Youth Project (61101155); Jilin provincial development and Reform Commission industrial innovation special project (2016C035); The Key Laboratory of symbolic computation and Knowledge Engineering Ministry of education of Jilin University in 2017.

References

[1] G.Y. Zheng, X.B. Liu and G.H. Han, A review of computer-aided detection and diagnosis systems for medical imaging, *Journal of Software* **29**(05) (2018), 1471–1514.

[2] L.Y. Ding and G.F. Shen, Case control study of mammary cancer, *Chinese Journal of Prevention and Control of Chronic Non-Communicable Diseases* (6) (1998), 283–285.

[3] J.L. Wang, S.Q. Zhou, C.P. Xu, et al., Zhao Wenhe: Don't let mammary cancer come to young life, *Healthy Life* (2) (2015), 24–25.

[4] R.M. Nishikawa, Comment on "Quantitative classification of breast tumors in digitized mammograms", *Medical Physics* **23**(8) (1997), 1337–1345.

[5] L. Hadjiiski, B. Sahiner, H.P. Chan, et al., Classification of malignant and benign masses based on hybrid ART2LDA approach, *IEEE Transactions on Medical Imaging* **18**(12) (1999), 1178–1187.

[6] D. Guliato, J.D.D. Carvalho, R.M. Rangayyan, et al., Feature Extraction from a Signature Based on the Turning Angle Function for the Classification of Breast Tumors, *Journal of Digital Imaging* **21**(2) (2008), 129–144.

[7] R.M. Rangayyan, N.R. Mudigonda and J.E.L. Desautels, Boundary modelling and shape analysis methods for classification of mammographic masses, *Medical & Biological Engineering & Computing* **38**(5) (2000), 487–496.

[8] Dr. S. Radhimeenakshi and K. Latha, Prediction of Stock Market Using Data Mining Algorithm Based on Historical Prices, *International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)* **5**(10) 2017.

[9] B. Sahiner, H.P. Chan, N. Petrick, et al., Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis, *Medical Physics* **25**(4) (1998), 516.

[10] B. Sahiner, H.P. Chan, N. Petrick, et al., Improvement of mammographic mass characterization using spiculation meausures and morphological features, *Medical Physics* **28**(7) (2001), 1455–1465.

[11] H.D. Cheng, X.J. Shi and R. Min, et al., Approaches for automated detection and classification of masses in mammograms, *Pattern Recognition* **39**(4) (2006), 646–668.

[12] W.D. Xu, *Study on computer assisted diagnosis of mammography*, Zhejiang University 2006.

[13] J. Tang and X. Liu, Classification of Breast Mass in Mammography with an Improved Level Set Segmentation by Combining Morphological Features and Texture Features, *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*. 2011.

[14] B. Sahiner, N. Petrick, H.P. Chan, et al., Computer-aided characterization of mammographic masses: Accuracy of mass segmentation and its effects on characterization, *IEEE Transactions on Medical Imaging* **20**(12) (2002), 1275–1284.

[15] H. Jégou, M. Douze, C. Schmid and P. Pérez, Aggregating local descriptors into a compact image representation, In Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). San Francisco: CA, (2010), 3304–3311.

[16] R. Ji, L-Y. Duan, J. Chen, H. Yao and W. Gao, A lowbit rate vocabulary coding scheme for mobile landmark search, In Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). Prague (2011), 2316–2319.

[17] M. Mayilvaganan and R. Manickam, Scaling Factor Analysis of Snowfall Region in Aerial View Based On DBscan Algorithm or Data Clustering Algorithm, *International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)* **5**(9), 2017.

[18] D. Deng, R. Wang, H. Wu, et al., Learning Deep Similarity Models with Focus Ranking for Fabric Image Retrieval, *Image and Vision Computing* (2017), S0262885617301919.

[19] K.Q. Weinberger and L.K. Saul, Distance Metric Learning for Large Margin Nearest Neighbor Classification, 2009.

[20] Dr. S.K. Jayanthi and C. Lalitha, Betel Leaf Disease Detection Using Histogram of Oriented Gradients and Multiclass SVM, *International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)* **5**(8) 2017. Léon Bottou. Large-Scale Machine Learning with Stochastic Gradient Descent, 2010.

[21] M. Jiang, S. Zhang, H. Li and D.N. Metaxas, Computer-aided diagnosis of mammographic masses using scalable image retrieval, *IEEE Transactions on Biomedical Engineering* **62** (2015), 783–792.

[22] D. Nistér and H. Stewénius, Scalable Recognition with a Vocabulary Tree, Computer Vision and Pattern Recognition, *2006 IEEE Computer Society Conference on. IEEE*, 2006.

[23] M. Jiang, S. Zhang, J. Liu, et al., Computer-aided diagnosis of mammographic masses using vocabulary tree-based image retrieval, *Society of Photo-optical Instrumentation Engineers* 2017.

[24] Q.L. Li, et al., Computer-aided diagnosis of mammographic masses using geometric verification-based image retrieval. *Society of Photo-optical Instrumentation Engineers Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 2017.

[25] A. Paul, A. Ahmad, M.M. Rathore, S. Jabbar, Smartbuddy: Defining human behaviors using big data analytics in social internet of things, *IEEE Wireless Communications* **23**(5) (2016), 68–74.