

# A novel large-scale multimedia image data classification algorithm based on mapping assisted deep neural network

Jun Li<sup>1</sup> · Rishav Singh<sup>2</sup> · Ritika Singh<sup>3</sup>

Received: 5 October 2016 / Revised: 15 December 2016 / Accepted: 6 January 2017 /

Published online: 20 January 2017

© Springer Science+Business Media New York 2017

**Abstract** With the increasing number of the images, how to effectively manage and use these images becomes an urgent problem to be solved. The classification of the images is one of the effective ways to manage and retrieve images. In this paper, we propose a novel large-scale multimedia image data classification algorithm based on deep learning. We firstly select the image characteristics to represent the flag for retrieval, which represents the color, texture and shape characteristics respectively. A feature of color is the most basic image data, mainly including the average brightness, color histogram and dominant color, etc. What the texture refers to is the image data in the anomalous, macroscopic as well as orderly one key character that on partial has. The contour feature extraction of image data needs to rely on the edge detection, edge of the detected edge through the connection or grouping to form a meaningful image event. Secondly, we revise the convolutional neural network model based on the pooling operation optimization, the pooling is in the process of the convolution operation to extract the image characteristics of the different locations to gather statistics. Furthermore, we integrate the parallel and could storage strategy to enhance the efficiency of the proposed methodology. The performance of the algorithm is verified, compared with the other state-of-the-art approaches, the proposed one obtains the better efficiency and accuracy.

**Keywords** Large-scale system · Multimedia architecture · Image data classification · Information retrieval · Deep learning · Convolutional neural network

---

✉ Jun Li  
lijuncas@163.com

<sup>1</sup> Quanzhou Institute of Equipment Manufacturing, Haixi Institutes, Chinese Academy of Sciences, Quanzhou 362200, China

<sup>2</sup> Infosys Limited, Bhubaneswar, India

<sup>3</sup> Amity University, Noida, India

## 1 Introduction

Along with the development of electronics technology and computer technology, the DIP (digital image processing) enters the high-speed period of development. Now we can gain the massive images from the image equipment and the image database or the Internet and thus carries on the effective classification and retrieval to the massive images also becomes the research subjects of many scientific researchers [25, 26, 38]. With the increasing number of the images, how to effectively manage and use these images becomes an urgent problem to be solved. The classification of the images is one of the effective ways to manage and retrieve images. However, most of the image classification is in the manual labeling stage. Because manual classification is not only time-consuming and laborious, but also prone to errors, so the automatic classification of the images has become the focus of current research [44, 46].

Starting from the previous research [5, 20, 23, 57], numerous researchers from the different angle analysis and solution labelling issue, expected that can find the good retrieval and the labelling method. And refer to the literatures of the [17, 19, 21, 28, 54], the deep learning has been a proper selection for implement some of the application scenarios. These methods analyze from the feature expression mechanism of image which can be divided into two kinds approximately.

- 1) Firstly, the image is divided into several homogeneous regions or image sub-blocks, then the image is analyzed with the semantic annotation. This kind of method uses image segmentation algorithm, tries to be divided into certain semantic object units the image effectively, through seeking for the labelling keyword and region semantics object or the view picture image itself within the corresponding relationships realizes automatic image labelling [3, 9].
- 2) Use images of the global visual information using the method of semantic annotation for image scene, the method to image characteristics and the label text word for the complete separation, and compare the image similarity on pure visual hierarchy which is a supervised learning method [12, 18].

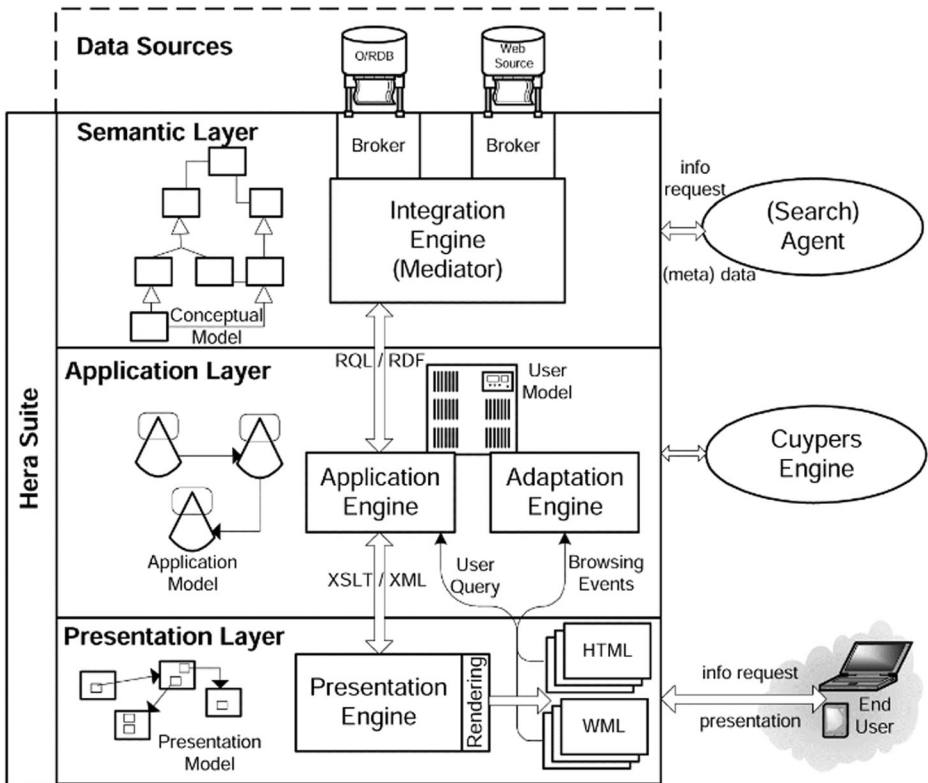
Among all the methodologies, the feature selection and extraction step is the essential one that needs in-depth analysis. From the literature review, we summarize the state-of-the-art characteristics selection approaches into the following categories.

- 1) The global feature. Image characteristics including high-level semantic feature and low-level feature. As a result of “semantic gap” existence, the high-level semantic feature is hard to gain through the computer directly, by the semantic key retrieval image was still an arduous duty. The problem of the semantic gap comes from the research of content-based retrieval, but the problem also exists in the field of text search, web search and video retrieval, for the problem of how to cross the semantic gap, the methods of the relevance feedback and automatic annotation. The former uses the interaction between the system and the user to obtain the mapping relation between the high-level semantics and the underlying characteristics of the multimedia data. The latter uses the keywords to mark the multimedia data. This technology is widely used in the image field. But using the similarity matching retrieval image of image low-level feature is another important image retrieval method, but this method relies on the validity of image characteristics.

- 2) Color characteristics. Commonly used in the image retrieval color space HSV color space, CIELAB space and CIELUV space, because they are similar to the human eye and subjective visual. The color histogram can represent the distribution of the color space value of all the pixels in an image [32, 55]. The color moment is a simple and effective color feature. Since color distribution information is mainly concentrated in the low order moment, the first order moment, the second moment and the third moment of the color are enough to express the color distribution of the image. The advantage of this approach over color histograms is that it does not need to vectorize the characteristics, thus speeding up the processing speed [35].
- 3) Texture feature. Scholars summarized six associated with human visual perception of the texture feature and its calculation method: roughness, contrast, directivity, line, regularity, and a rough degree of similarity. The first three characteristics play an important role in image retrieval and classification which reflects the degree of the coarse or fine image texture roughness while contrast to reflect the degree of a texture clear. If there is a rule in the direction of the directional reflection texture.

While implementing the automatic image classification system, the efficiency and the hardware effectiveness should be taken into consideration. With this target, the processing paradigm for the large-scale multimedia system should be finalized. Multimedia cloud computing platform through a number of servers to complete the transfer of large-scale data flow, data flow between servers in parallel processing, prone to resource conflict, resulting in waste of server resources and data flow scheduling lag. Under the multi-server scheduling data flow between the rational allocation of server resources that can be separated into the three aspects. (1) Control level: the management of scheduling platform of large-scale data flow hardware resources for each data flow to allocate the corresponding hardware resources, the data flow into a scheduling platform of large-scale data flow for real-time operation. (2) Data level: according to topology of management level establishment, regulates the server in large-scale data stream dispatch platform, the data of receive control level feedback, guarantees the large-scale data stream live transmission. (3) Management level: the deployment of the scheduling platform of large-scale data flow in the data-level topology, while the multimedia cloud computing environment, the allocation of hardware resources to regulate.

Scheduling platform of large-scale data flow also includes the timer which need to pass the packet to the timer, the timer in the packet arrival time after the transmission of data packets to the output module. In the Figs. 1 and 2, we show the large-scale multimedia system and the data interaction flow and the initialized code for the system. The configuration control module initializes scheduling platform of large-scale data flow and regulates the program flow resources. The packet input module collects the data in the multimedia cloud computing environment and classifies and filters the data, and then transfers the processed data to the streaming application. The packet input module receives the RTSP and IP packet data fed back by the streaming application and sends the data to the grouping operation module [36, 41, 48]. Under this paradigm, efficiency from the hardware perspective will be enhanced and the robustness will be improved. A distributed multimedia system, multimedia information is the object of system transmission and processing, network is the material basis for the transmission of information and security. Through the network transmission, multimedia information from the source site to reach the destination site, and multimedia information and the network there is a close and inevitable link between. Model to a communication channel for network abstraction, eliminates the complexity of network topology and various forms of research work



**Fig. 1** The large-scale multimedia system and the data interaction flow

to bring difficulties, while retaining the basic characteristics of the network that is to simplify the multimedia communication network is reasonable and effective which can be reflected from the Fig. 1.

As shown in the Fig. 2, the initialization code block for the large-scale multimedia system is demonstrated. The operating system bootloader provides a set of the software environments that are required before the operating system loads. Through the execution of this program, the embedded system can initialize the mapping between the hardware device and the memory space, and finally bring the system into the appropriate hardware and software environment. After an embedded hardware device is powered up or reset, the processor typically fetches instructions from an address pre-arranged by the manufacturer.

In this paper, to provide the more efficient image classification algorithm, we conduct research on the large-scale multimedia image data classification algorithm based on the deep learning. The theoretical details will be introduced in the later sections.

## 2 Convolutional neural network and the optimization strategy

The depth study is new domain in a machine learning research, the depth study through combining the low level characteristics, forms the higher abstract level feature expression property category, to discover that data distributional characteristics, and its characteristics

**Fig. 2** The initialization code block for the large-scale multimedia system

```

while (1)
{
  In_data=get_stream_data;
  If in_data!=NULL
  {
    handle_data (i_data) :
    insert_data (in_data) :
  }
  Flow_stack=Flow_stack+1;
  fo: i-1toHMD_num
  {
    Paek_data=get_Flow_data (i, Flow_stack) :
    If Flow_data!=NULL
    {
      Flowing_data (Flow_data) :
      insert_data (Flow_data) :
    }
    send_stack=send_stack+1;
    for l=1 to HMD_num
    {
      send_data=get_send_data (i, send_stack) :
      if send_data!=NULL
      {i
      f send_data
      {
        sending_data (send_data) :
      }}}
  }
}

```

include multi-layer perceptron structure of hidden-layer that has parallelism of information processing, the distributivity of information processing and interconnection of information processing unit, the plasticity of structure, high non-linearity and with good fault tolerance, learning control, self-organization and from compatibility and other characteristics [7, 27, 31, 43, 51]. CNN is proposed as a depth learning architecture in order to minimize the data preprocessing requirements. Convolution neural network (CNN) is essentially the multi-layer perceptron which is a variety of structures that can contain multiple nerves, input layer, hidden layer, the output layer [16, 37]. The convolutional neural network is essentially a variant of multilayer perceptron, which consists of an input layer, a hidden layer and an output layer. The hidden layer can have a plurality of layers, each layer is composed of a plurality of two-dimensional planes, and each plane is composed of a plurality of independent neurons. Theoretically, the essential part of the CNN can be separate into the S and the C layers. C levels are mainly to use the convolution kernel extraction characteristics and realize carries on the effect of the filtration and strengthening to the characteristics. In each convolution level, it firstly carries on the convolution operation the characterization diagram and convolution kernel of the previous output, then through activation function and then the characterization diagram of output this level [22, 30, 56].

$$y_j^t = f \left( \sum_{i \in P_j} k_{i,j}^t y_i^{t-1} + b_j^t \right) \quad (1)$$

S mainly by the sampling C layer characteristic dimension of the each size of n in the pool for average “pool” or “pool” largest operation that can be expressed as formula 2. Where the

$down\_sampling(y_j^{t-1} + b_j^t)$  represents the down sampling function and the  $y_j^{t-1}$ ,  $y_j^t$  represent the signal before and after the operation, respectively.

$$y_j^t = f\left(down\_sampling\left(y_j^{t-1} + b_j^t\right)\right) \quad (2)$$

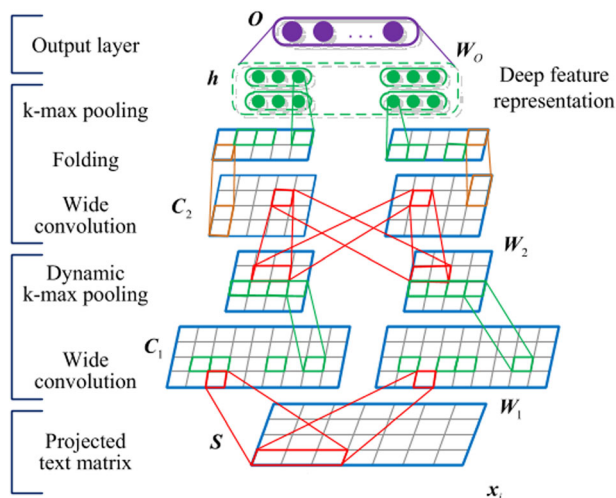
In the above formulas, the  $f$  represents activation function,  $t$  represents the layer numbers, the  $k_{i,j}^t$  denotes the kernel [14, 53],  $down\_sampling$  is the sampling function. Traditional convolution neural network is by the convolution level and depth structure of sub-sampling level alternately constitution, this depth structure can reduce the computing time effectively and establishes invariability in the spatial structure. The input picture carries on the network layer upon layer maps, and ultimately obtains the various levels regarding the image different expression forms, realizes the image the depth to indicate that convolution kernel as well as sub-sampling the mapping way of way direct decision image. In the following Fig. 3, we show the architecture of the CNN.

In training, its forward propagation is used for feature extraction, and convolution and the down-sampling are used to obtain image characteristics. The back propagation is used for the error correction. The traditional Back Propagation mechanism is used to propagate the error layer by layer, and the chain convolution rule is used to update the convolution kernel. The output function can be demoted as the formula 3.

$$z_{output} = g_L(X_L; w_L) \quad (3)$$

The  $z_{output}$  represents the output function, the  $g_L$  is the exchange function, the  $X_L$  is the value of the prior-layer output, the  $w_L$  represents the weight matrix. As shown in the Fig. 3, the convolution feature mapping phase generates a total of  $K$  feature maps, each feature map has its own automatic encoder following to the  $k$ -th feature map that

**Fig. 3** The architecture of the CNN



corresponding to the  $k$ -th automatic encoder as an example, In this paper, an automatic encoder with only the single hidden layer is used. The hidden layer feature can be formulated as Eq. 4.

$$Hiden_k = g(a_k \cdot p_k + b) \quad (4)$$

The matric of the  $a_k$  and  $b_k$  can be denoted as formula 5~6.

$$a_k = [a_{k1}, a_{k2}, \dots, a_{kL}] \quad (5)$$

$$b_k = [b_{k1}, b_{k2}, \dots, b_{kL}] \quad (6)$$

At level  $L$ , the loss function can be expressed as the follows [4, 15].

$$\gamma(x_i, y_i) = -\frac{1}{n} \sum_{i=1}^n (y_i - \ln(g_L(x_i; w_i))) + \lambda \sum_{k=1}^L \sum_{j=1}^k \|w_j\|^2 \quad (7)$$

Where the parameter of the  $\lambda$  is the essential term we should take into consideration. Overall, at present bases on the depth CNN model to be mostly centralized in the depth exploitation, in the fitting control and practical application, during the training needed the mass data support. In the case of a small sample, its performance is even less than the traditional feature extraction method, but in other large data sets on the case of pre-training model, the effect is far more than the existing manual feature model, the current model design and most applications are based on pre-training and implementation. For this concern, the general network output issue can be formulated as the following equation.

$$Net_{pj}^L = \sum_{\substack{i \\ w_{ji} \geq 0}} w_{ji} Y_{pi}^L + \sum_{\substack{i \\ w_{ji} < 0}} w_{ji} Y_{pi}^U + \theta_j \quad (8)$$

$$Net_{pj}^U = \sum_{\substack{i \\ w_{ji} \geq 0}} w_{ji} Y_{pi}^U + \sum_{\substack{i \\ w_{ji} < 0}} w_{ji} Y_{pi}^L + \theta_j \quad (9)$$

And the optimization issue can be transferred into the formula 10.

$$E_{optimized} = \max \left\{ \frac{1}{2} (t_{pj} - Y_{pj})^2, Y_{pj} \in \mathbf{Y}_p \right\} \quad (10)$$

Under the conditions of the formula 11.

$$Y_{pj} = [Y_{pj}^L, Y_{pj}^U] = [f(\text{Net}_{pj}^L), f(\text{Net}_{pj}^U)] \quad (11)$$

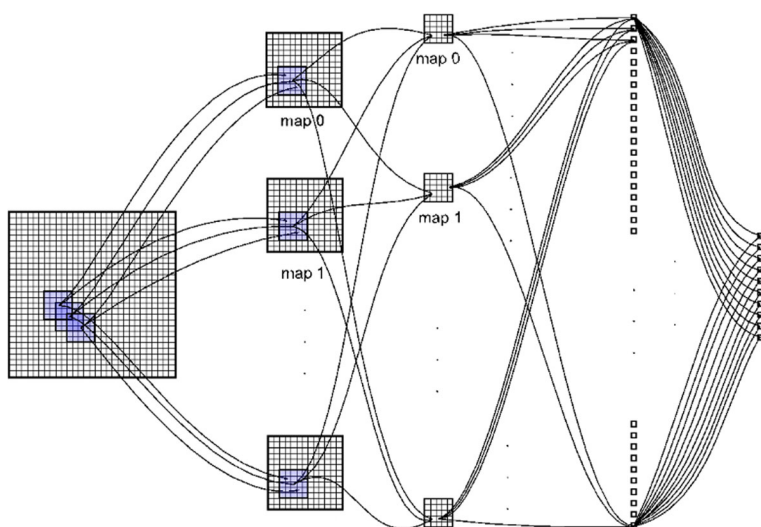
In addition to its own depth structure, the maximum difference between the convolutional neural network and the general BP neural network is to reduce the

network parameters by means of local receptive field and weight sharing method. The so-called localized receptive field means that each convolution kernel is connected only to a specific region in the image, that is, each convolution kernel only convolves a portion of the image, and in the subsequent layers these local convolutions feature, which is consistent with the spatial correlation of the image pixels and reduces number of convolution parameters. Accordingly, the task function can be expressed as the formula 12.

$$\Delta w_{ji}(t+1) = \eta \left( -\frac{\partial E_p}{\partial w_{ji}} \right) + \alpha \Delta w_{ji}(t) \quad (12)$$

The  $\partial E_p / \partial w_{ji}$  is the primary parameter we should consider. For this situation, we need the mapping optimization shown as the Fig. 4. The system input is one with the time related process, its value of exports both relies on the input function spatial polymerization, and time build-up effect close correlation that must solve this kind of problem, the traditional method needs to establish and to solve the more complex mathematical model generally. But these systems often affect factor many nonlinear systems. The main purpose of the self-organizing feature map neural network is to convert the input signal pattern of any dimension into one-dimensional or two-dimensional discrete mapping, which reflects the memory pattern of the nerve cells and excitability rules of nerve cells stimulated. The characteristics of the nervous system and this network correspond to a group of the unit neurons, rather than a neuron corresponding to a model.

The crucial technique is the transferring matrix, after the error back propagation, the error function of each network layer is obtained, then the network weights are modified by the stochastic gradient descent method, and then the next iteration is carried out until the convergence condition is reached. Note that due to the difference between the



**Fig. 4** The mapping optimization for the CNN



size of the layer and the layer, in the error transfer need to be sampled before and after the two layers of the same size, and then the error transfer. In the two cross connection layer, the weight updates also uses the chain derivative rule can be expressed as follows.

$$\begin{cases} \frac{\partial l}{\partial w_{L-1}^A} = \frac{\partial l}{\partial g_L} \frac{\partial g_L}{\partial g_{L-1}^A} \frac{\partial g_{L-1}^A}{\partial w_{L-1}^A} \\ \frac{\partial l}{\partial w_{L-1}^B} = \frac{\partial l}{\partial g_L} \frac{\partial g_L}{\partial g_{L-1}^B} \frac{\partial g_{L-1}^B}{\partial w_{L-1}^B} \end{cases} \quad (13)$$

When the convolution of a layer of sampling and at this time from the lower convolution layer error of the drop sampling error, we need to re-calculate the term as follows.

$$\begin{cases} \frac{\partial l}{\partial w_{L-2}^A} = \frac{\partial l}{\partial g_L} \frac{\partial g_L}{\partial g_{L-1}^A} \frac{\partial g_{L-1}^A}{\partial w_{L-2}^A} + \frac{\partial l}{\partial g_L} \frac{\partial g_L}{\partial g_{L-2}^A} \frac{\partial g_{L-2}^A}{\partial w_{L-2}^A} \\ \frac{\partial l}{\partial w_{L-2}^B} = \frac{\partial l}{\partial g_L} \frac{\partial g_L}{\partial g_{L-1}^B} \frac{\partial g_{L-1}^B}{\partial w_{L-2}^B} + \frac{\partial l}{\partial g_L} \frac{\partial g_L}{\partial g_{L-2}^B} \frac{\partial g_{L-2}^B}{\partial w_{L-2}^B} \end{cases} \quad (14)$$

In the process of calculating the optimal solution of the objective function, through continuous iteration, to eventually achieve convergence error state. The universal can be expressed as the formula 15.

$$\begin{cases} \frac{\partial l}{\partial w_{L-m}^A} = \frac{\partial l}{\partial g_L} \frac{\partial g_L}{\partial g_{L-1}^A} \frac{\partial g_{L-1}^A}{\partial w_{L-m}^A} + \frac{\partial l}{\partial g_L} \frac{\partial g_L}{\partial g_{L-2}^A} \frac{\partial g_{L-2}^A}{\partial w_{L-m}^A} + \dots + \frac{\partial l}{\partial g_L} \frac{\partial g_L}{\partial g_{L-m}^A} \frac{\partial g_{L-m}^A}{\partial w_{L-m}^A} \\ \frac{\partial l}{\partial w_{L-m}^B} = \frac{\partial l}{\partial g_L} \frac{\partial g_L}{\partial g_{L-1}^B} \frac{\partial g_{L-1}^B}{\partial w_{L-m}^B} + \frac{\partial l}{\partial g_L} \frac{\partial g_L}{\partial g_{L-2}^B} \frac{\partial g_{L-2}^B}{\partial w_{L-m}^B} + \dots + \frac{\partial l}{\partial g_L} \frac{\partial g_L}{\partial g_{L-m}^B} \frac{\partial g_{L-m}^B}{\partial w_{L-m}^B} \end{cases} \quad (15)$$

### 3 The fundamental basis for the proposed methodology

#### 3.1 The large-scale multimedia data properties

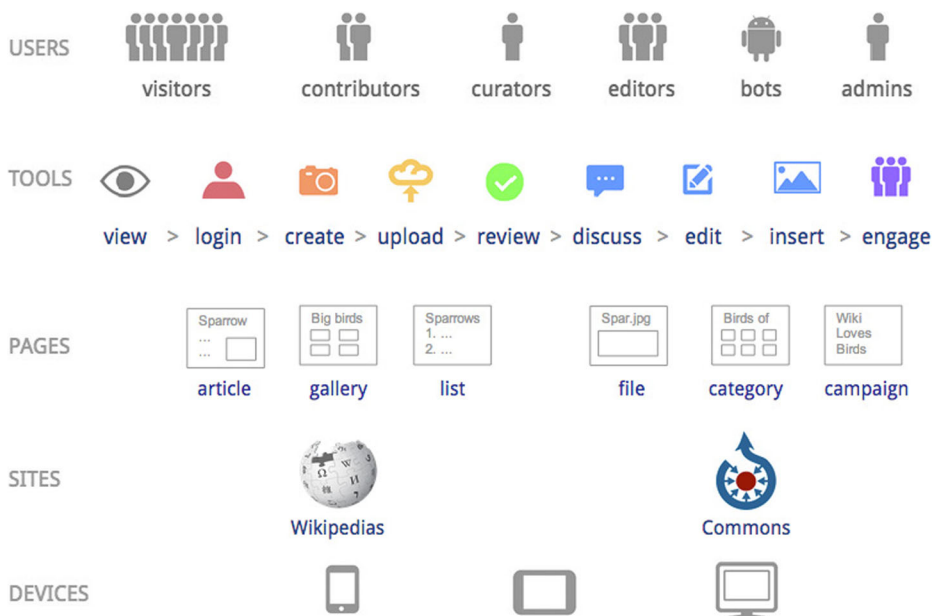
Under the cloud of multimedia data flow has the characteristics of real-time, randomness, number, etc., and multimedia application of cloud computing environment required to support multimedia services can provide high quality service, to achieve the service quality is the key to effective process data congestion problems, and thus to seek effective mass data stream scheduling method. According to the development of the network transmission technology, large-scale streaming media system has experienced three stages as follows.

- 1) IP multicast protocol to carry streaming media transmission, IP layer multicast due to congestion control, scalability, availability and other aspects of a series of problems, so the

- IP multicast-based video communications applications have not been widely used in Internet usage.
- 2) As the one special realization mode of peer-to-peer network P2P mode of basic transmission, statement of application-layer multicast agreement, especially bases on the application-layer multicast agreement the use of the end system multicast system SIGCOMM conference, symbolizes that the streaming media broadcast plan was in the 2nd stage of development [10, 33, 58].
  - 3) In the multi-sender multi-receiver transmission mode, each node can receive the data from multiple nodes or send data to multiple nodes. The data topology between nodes forms a network structure, which greatly improves the system expansion feature.

In the Fig. 5, we show the architecture of the large-scale multimedia data processing system, as shown, the data stream content extraction module through the data import module gain high-speed multimedia data class, with degree of correlation between the multi-channel degrees of the correlation computation module computation different data circulation, the multi-channel information extraction module is responsible to the channel data stream carries on the information extraction according to the modality as well as the system, vector extraction module is responsible to the processed information carries on the feature extraction, the characteristic vector that will then extract sends in the filtration computation module.

To maintain the satisfactory processing efficiency, we adopt the multi-channel connections architecture. Whether the data flows of different channels are federated or not is based on whether the channels are correlated. A filter rule consists of the conditional expressions and action expressions. The conditional expression consists of a general predicate and a relational operator  $\wedge$ , and the action expression consists of an action unit



**Fig. 5** The large-scale multimedia data processing system architecture

and a relational operator  $\wedge$ . Simply put, a filter rule means that if the conditional expression is true, then implementation of the action, or not the implementation of action. If the conditional expression is empty, the action is executed unconditionally, which means that all the multimedia streams are executed. If in a filtration rule includes the filtration demands between the filtration demands or the different channels of different modality, then this rule for fusion filtration rule. Suppose we have a sequence of  $S$  denoted as the formula 16.

$$S = \{S_1, \dots, S_p\} \quad (16)$$

Defines interrelatedness between the two channels to have the difference in the different practical application systems and below uses in our validation assemblies the definition example about interrelatedness.

$$a(S_i, S_j) = E_p * R_p(S_i, S_j) + E_c * R_c(S_i, S_j) \quad (17)$$

In our verification system, the definition of “key content dictionary” and “filter content dictionary” to calculate the channel data stream in the text content relevance that can basically reflect the channel itself semantic information relevance. We consider the textual information in the channel data stream as a “word flow” and then build a temporary “word frequency table” with our “Key Content Dictionary” and the “Filter Content Dictionary”. The standard for our methodology can be defined as the formula 18.

$$R_c(S_i, S_j) = \frac{\sum_k X_k * Y_k / N - U_i * U_j}{\sigma_i * \sigma_j} \quad (18)$$

From the above mentioned analysis, we can reach the listed conclusion.

- (1) The correlation content related to  $N$ .
- (2) The content correlation is associated with the expression of channel.
- (3) The content of correlation degree between the channel and sharing degree.

### 3.2 The image retrieval and classification paradigms

Image retrieval technology has a lot of kinds, most of which are based on the shape of the image, such as color, texture characteristics for image retrieval. The primary application of the image classification and retrieval is the automatic annotation. Automatic image annotation is to make the computer automatically able to any unsigned paintings that reflect the meaning of the image content keywords. It by using the annotated image set or other available information automatically learn the semantic concept space and the relationship between visual feature space model, and use this model with unknown semantic image, i.e., it is trying to image the high-level semantic information and to establish a mapping relationship between low-level characteristics, thus to some extent, can solve the problem of “semantic gap”.

Existing majority automatic image labelling algorithm, attempts to realize labelling of semantic keyword in the image rank directly, namely the algorithm does not need between the region and keyword of image establishes the mapping relationships of correspondences. But also some work try to solve the labelling problem from the technical angle of object recognition that is each region of image entrusts with the keyword. In the Fig. 6, we show the sample illustration of the retrieval and annotation result.

Based on the literature review, the annotation methods can be generally summarized as the following aspects. (1) Classification-based automatic image labelling algorithm. The mentality of quite direct-viewing automatic image labelling, labelling the issue regards as is the image retrieval classification issue. If regards as each semantic keyword is a category marks, then the image labelling issue transforms as the image classification issue and therefore can definitely solve the labelling problem from the angle of image classification. (2) Probability context modeling-based automatic image labelling algorithm. While the image annotation algorithm based on the probability association model is based on probabilistic statistical model, and that analyzes the symbiosis probability relation between image region feature and semantic keywords, and uses this as the annotation of the image to be annotated. Intuitively, if two images have high visual similarity, the higher the probability of the two keywords are similar. (3) Based on the learning algorithm of automatic image annotation. Graph based learning algorithm is a semi supervised learning algorithm, which is known to be involved in the learning process of the algorithm. With the traditional supervised learning and unsupervised learning, semi supervised learning can use more information in the learning stage, such as the distribution of data characteristics and it applies to a large amount of data that has been labeled a relatively small amount of data [6, 8, 34, 52].

The existing potential optimization approaches for the classification can be organized as the following aspect. (1) Hash algorithm integration. Through the HASH algorithm can under the rapid localization centralized certain probability with the data of search data correlation, coordinate Hamming space similarity measure the rapidity and characteristics of index result easily further expansion that can greatly enhance the efficiency of index and retrieval, thus Hash technology is regarded as most has the similar reconnaissance method of potential. (2) Based on the Search algorithm of the automatic image annotation [1, 42, 47]. Search based annotation method avoids the complicated parameter learning process. Moreover, since the relevant image is found by retrieval, the method is not limited



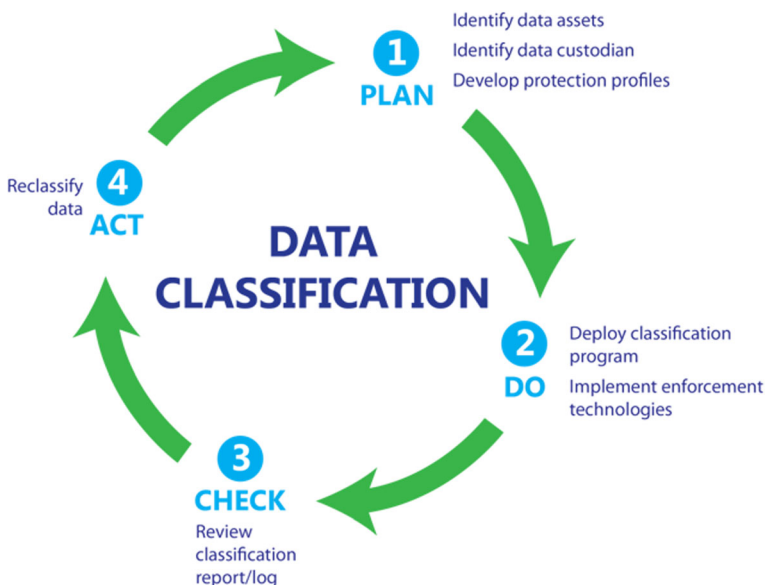
**Fig. 6** The image retrieval and annotation result demonstration

by the training set or the set of the annotated words. (3) In the candidate labelling information that the basic image labelling stage obtains possibly is incomplete, or contained some and image not related labelling information. This is mainly because the existing labelling algorithm analyzes each semantic keyword alone, has not considered the semantic connection between keywords. But under normal conditions, the semantic link between glossary and glossary is very close, usually among glossaries including hierarchical relation and relevant information. (4) What character description to express the semantic information of the images, what the image segmentation algorithm based on semantic that can effectively characterize the user perception of image content, is an important means to improve the performance of image automatic classification. In the Fig. 7, we show the procedures [39].

## 4 The proposed framework

### 4.1 The CNN based classification and retrieval framework

This article uses the depth study of the convolutional neural network, it is a kind of the feedforward neural networks which is mainly composed of the multiple convolution and full connection layer. The weight of some connections between neurons in the same layer of is shared. A feedforward neural network can be considered to be the combination of a series of function [2, 11]. The basic theories are introduced previously, in this section, we focus on the optimization and the modification operation [24, 40, 50]. Pooling is in the process of the convolution operation to extract the image characteristics of the different locations to gather statistics. Pooling can reduce the



**Fig. 7** The illustration of the classification procedures

dimension of convolution characteristics, at the same time also can prevent data fitting, and the Fig. 8 reflects the feature.

We formulize the pooling procedure as the Eq. 19.

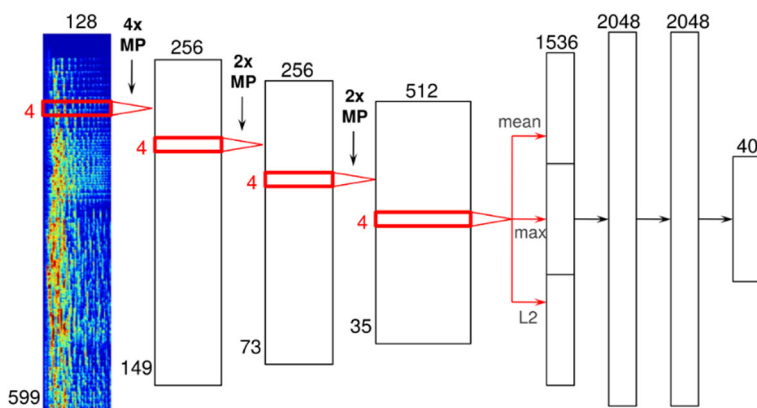
$$y_{ijk} = \max \left\{ y_{i'j'k'} : i \leq i' < i + p, j \leq j' < j + p \right\} \quad (19)$$

Under normal conditions, the image collection of target task with pre-training image collection the category quantity or the image style have very big difference, in the retrieval duty of picture target collection, often is directly hard to achieve the optimal performance with pre-training the visual feature of CNN model extraction image. Therefore, to cause the CNN model parameter of pre-training well the feature extraction for picture target collection, to the CNN model parameter of pre-training carries on the trimming using the image of the picture target collection. The entire trimming process is as follows.

- 1) **Step One.** Each image from picture target storehouse first was adjusted to the 256\*256, then withdraws this chart stochastically the sub-block or its mirror image takes the input of CNN.
- 2) **Step Two.** Regarding 1st to 7th, we use in the parameter that in the pre-training process obtains initializes it. In trimming process, with pre-training step two similar parameter set-ups based on the Eq. 20.

$$J_{fuction} = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^c \left( p_{ik} - \bar{p}_{ik} \right)^2 \quad (20)$$

- 3) **Step Three.** For the first 7 layers to set a smaller learning rate, we can ensure that the parameters obtained through the pre-training CNN model that is not destroyed during the trimming process. For the last layer to set a higher learning rate, and the whole network can be quickly converged to the new optimal point on the target image set of the general pool.



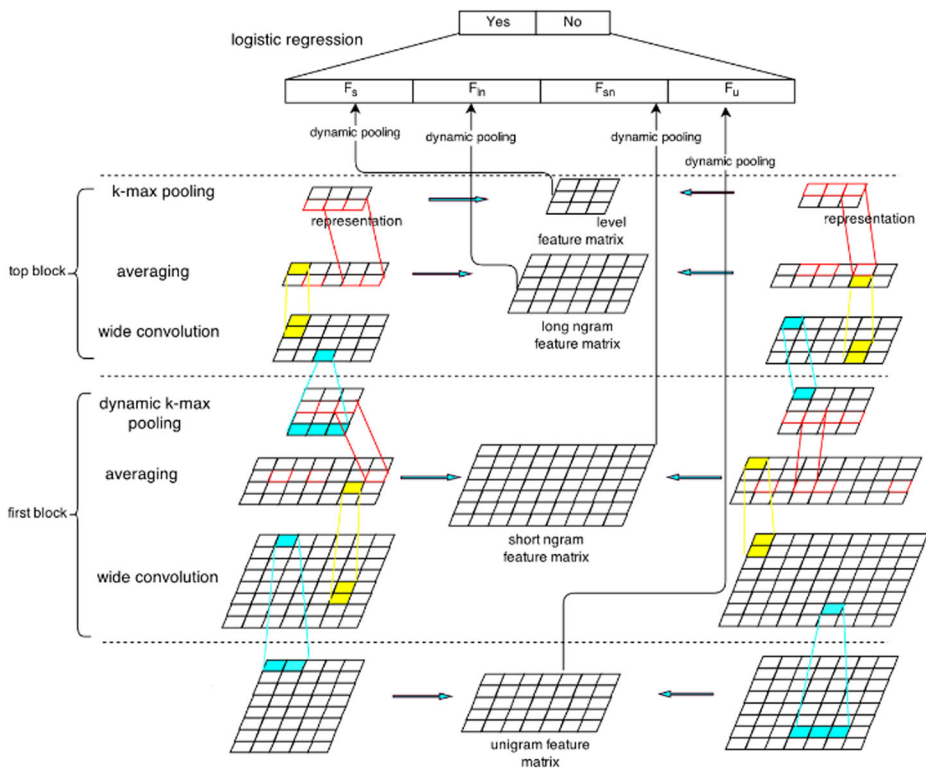
**Fig. 8** The pooling operation and the flowchart demonstration

The similar measure method of traditional image has the cosine to be away from, the Euclidean distance, among the distance through characteristic vectors judges the image the similar degree. Because the distance between the sole vectors cannot accurately reflect the similar degree between images, therefore this article used manifold distance-based ranking on manifolds method to measure the similarity between images [13, 45, 49]. The ranking function can be them formulized as the Eq. 21.

$$r^* = (I_n - \alpha S)^{-1} y \quad (21)$$

Where the  $I_n$  represents the sample matrix and in the Fig. 9, we show the modified CNN systematic architecture [29, 56].

As the further step, to test the effectiveness of the algorithm, we should define the function for the performance evaluation. The 22 ~ 24 defines the adopted standard. Image processing system appraisal effect main consideration accuracy and retrieval speed two aspect factors. Before the accuracy is decided in withdraws the image characteristic separating capacity and the match algorithm validity, the retrieval speed is decided by the image characteristic order of complexity and match algorithm order of complexity and the image database organization form. Refinement refers to the ratio of effective images in the returned result set, and is used to exclude the system from irrelevant images. The recall rate refers to the ratio of the number of



**Fig. 9** The modified CNN systematic architecture



valid images in the returned set to the number of all similar images in the database and is used to measure the ability of the system to retrieve the relevant image.

$$precision = (a/b) \times 100\% \quad (22)$$

$$recall = (a/c) \times 100\% \quad (23)$$

$$MAP(Quantity) = \frac{1}{|Quantity|} \sum_{j=1}^{|Quantity|} \frac{1}{m_j} \sum_{k=1}^{m_j} Precision \quad (24)$$



**Fig. 10** The databases adopted by the experiment



## 4.2 Classification referred multimedia system efficiency enhancement model

KD tree as a data structure of k-dimensional data space division, that mainly used in multi-dimensional space key data search. In order to facilitate the processing of image data, a large number of eigenvectors representing image characteristics are used to refer to the whole image data. This feature vector accords with the characteristics of the high-dimensional key-value, and is suitable for indexing by using the KD tree to speed up image retrieval process. The KD tree is each node is a binary tree of K vector. Each non-leaf node can regard as a planoid, but this planoid the multi-dimensional space division is the two sub-planes. Was divided into the left subtree at this planoid left point, the right point was divided into the right subtree.

There are many ways to determine the hyperplane of this partitioned subspace, so there are many ways to construct a KD tree, which calculates the variance of each dimension of all points in the sub-plane and also chooses the dimension with the largest variance as the perpendicular to the partitioned subspace. The direction of the hyperplane is one of the most popular methods of building KD trees today. This paper chooses the following architectures as the primary characteristics and reference approach.

- **Polynuclear construction:** A polynuclear construction usually is one contains two or more independent achievement to carry out the unit basically the computation chip of CPU nucleus. These CPU have the respective independent buffer and bigger shared buffer. The communication network on these nucleus usual execution different thread and data exchange between processors through shared buffer or chip.
- **The nuclear architecture:** relative to the multi-core structure, there are more and more intense in the nuclear architecture components are integrated into one chip. And that among all the nuclear architecture, the GPGPU is one of the most popular architecture. Taking into account the entire hardware storage system, each SM has a separate contains thousands of registers and dozens of KB of the shared memory on the chip memory SM is shared by several SP and is not visible to the outside. Data transfer is through the slice under the global memory. Through the above design, GPGPU in the overall cost of the basic and global memory consistent with the case to support the register and shared memory register-level access speed.

**Table 1** Performance of the methodologies: Experiment set one, data size level: original level

	Method 1	Method 2	Method 3	Method 4	Ours
People	0.211	0.615	0.668	0.415	0.711
Dog	0.415	0.511	0.612	0.662	0.715
Cat	0.211	0.615	0.411	0.436	0.781
Building	0.411	0.456	0.498	0.611	0.714
Food	0.335	0.462	0.475	0.514	0.611
Flower	0.214	0.234	0.295	0.423	0.446
River	0.612	0.662	0.615	0.658	0.663
Sea	0.256	0.299	0.652	0.578	0.774
Flight	0.452	0.511	0.615	0.663	0.654
Horse	0.241	0.221	0.235	0.452	0.569
Shoe	0.226	0.345	0.395	0.496	0.579
Ring	0.315	0.336	0.495	0.562	0.675
Chicken	0.335	0.476	0.499	0.568	0.668
Car	0.415	0.511	0.516	0.532	0.574
Table	0.235	0.345	0.339	0.561	0.642

**Table 2** Performance of the methodologies: Experiment set 2, data size level: 2\*original level

	Method 1	Method 2	Method 3	Method 4	Ours
People	0.201	0.315	0.511	0.611	0.711
Dog	0.215	0.336	0.503	0.612	0.706
Cat	0.220	0.412	0.511	0.609	0.708
Building	0.201	0.411	0.502	0.621	0.713
Food	0.209	0.423	0.526	0.621	0.714
Flower	0.265	0.453	0.511	0.620	0.725
River	0.285	0.366	0.510	0.608	0.722
Sea	0.266	0.398	0.523	0.612	0.721
Flight	0.217	0.395	0.533	0.621	0.724
Horse	0.239	0.411	0.524	0.619	0.723
Shoe	0.200	0.419	0.514	0.603	0.712
Ring	0.198	0.425	0.522	0.622	0.786
Chicken	0.193	0.431	0.512	0.621	0.755
Car	0.207	0.422	0.511	0.635	0.752
Table	0.206	0.412	0.502	0.628	0.751

## 5 Experiment and simulation

In this section, we conduct experimental simulation on the proposed methodology. In the Fig. 10, the databased used for testing is shown. The generated database is formed of by the listed ones: natural scene image library and NUS-WIDE image library. The detailed information descriptions of these two image storehouses are as follows: Natural scene image storehouse contains 2000 images, all these images contain following 5 labelling: deserts, mountains, sea, setting sun and trees. In addition, we also add the animal and building images for testing. Image library contains 30,000 kinds of the images as these images are marked with the boats, cars, flags, horses, sky, sun, towers, aircraft, zebra, including the 31 kinds of labels.

To test the effectiveness of the proposed methodology, we compare it with the other algorithms, the Method 1 [53], 2 [1], 3 [42], 4 [13], respectively. The method 1 is similarity function based content-based image retrieval, the method 2 is the multiclass associative

**Table 3** Performance of the methodologies: Experiment set 3, data size level: 3\*original level

	Method 1	Method 2	Method 3	Method 4	Ours
People	0.152	0.221	0.411	0.533	0.688
Dog	0.166	0.253	0.412	0.524	0.645
Cat	0.123	0.223	0.411	0.562	0.625
Building	0.175	0.242	0.435	0.579	0.632
Food	0.187	0.249	0.421	0.544	0.712
Flower	0.177	0.261	0.408	0.532	0.711
River	0.168	0.254	0.411	0.542	0.698
Sea	0.195	0.277	0.426	0.511	0.685
Flight	0.178	0.289	0.417	0.518	0.686
Horse	0.166	0.277	0.416	0.542	0.695
Shoe	0.157	0.286	0.411	0.526	0.714
Ring	0.185	0.288	0.423	0.548	0.711
Chicken	0.139	0.265	0.425	0.573	0.721
Car	0.174	0.264	0.422	0.523	0.689
Table	0.188	0.263	0.435	0.532	0.668

**Table 4** Performance based on the MAP standard

	Method 1	Method 2	Method 3	Method 4	Ours
10 classes	41.3	44.3	45.6	47.2	51.9
20 classes	41.1	43.2	44.6	46.6	50.3
30 classes	40.5	43.0	44.1	45.7	49.8
40 classes	40.0	42.6	43.2	45.2	49.3
50 classes	39.8	41.9	42.5	44.9	48.6
70 classes	39.2	41.3	42.2	44.3	48.1
90 classes	38.6	39.5	41.9	42.1	47.2
125 classes	38.4	38.8	40.2	41.9	47.0
200 classes	37.5	37.6	39.7	40.8	46.5
256 classes	36.4	37.3	38.6	39.2	45.7

classification algorithm, the method 3 is the sparse representations of morphological attribute profiles based image retrieval and the method 4 is the ABACOC algorithm.

The Tables 1, 2, 3, 4 and 5 shows the statistical data of the comparison experiment, the Fig. 11 is the visualized performance of the experiment algorithms, these are the visualized result of the Tables 1, 2 and 3, he Fig. 12 represents the time consuming test based on the different data center sizes. In the experiment, we use 700 images and the training set and another 700 as the prediction one. The effectiveness and efficiency of the proposed method is well proved compared with other state-of-the-art approaches. In our experiment, the proposed algorithm got the average advancement of 17.5% compared with the other ones, the recall rate is much higher.

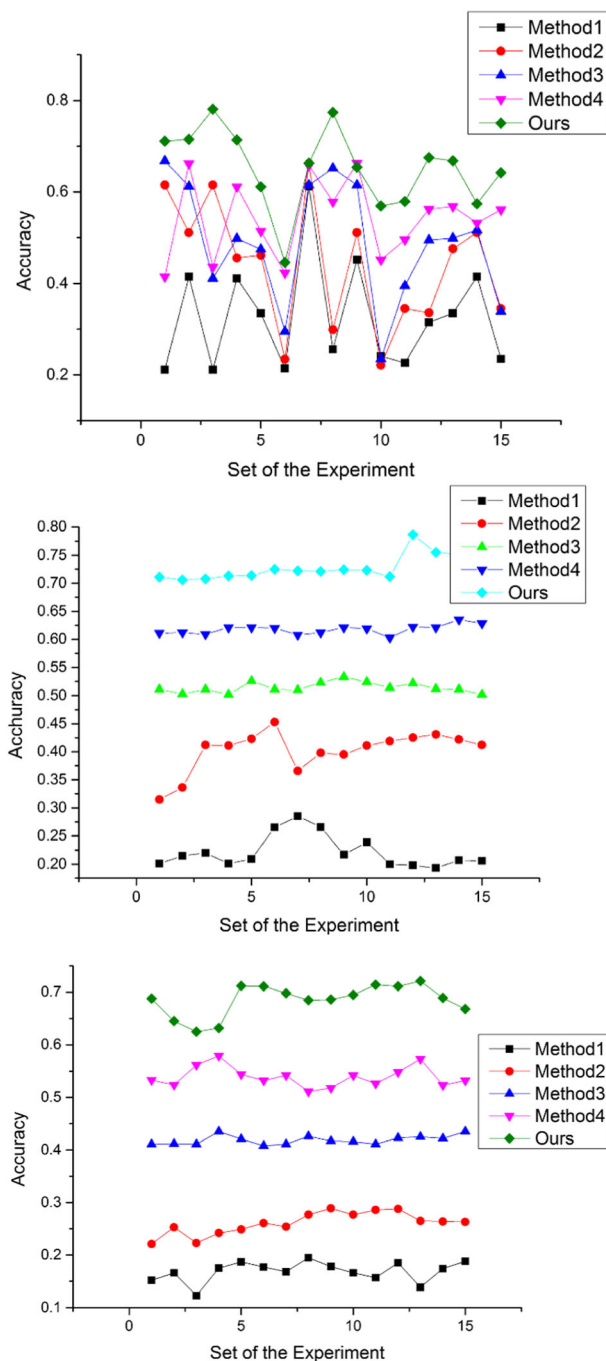
## 6 Conclusion and future work

In this paper, we propose a novel large-scale multimedia image data classification algorithm based on deep learning. This method in the label loses under the erroneous minimum significance to fill the omission label, constructs the semantic balanced neighborhood and then in the semantic uniformity of sample through the neighborhood measure study guarantee neighborhood of the multi-label information inserting, obtains some relevance between images through the sparse representation, and constructs the semantic consistent neighborhood and ensure neighborhood sample has the overall situation similarity, some relevance and semantic uniformities. To optimize the CNN, we integrate the manifold learning for systematic

**Table 5** Performance based on the recall rate

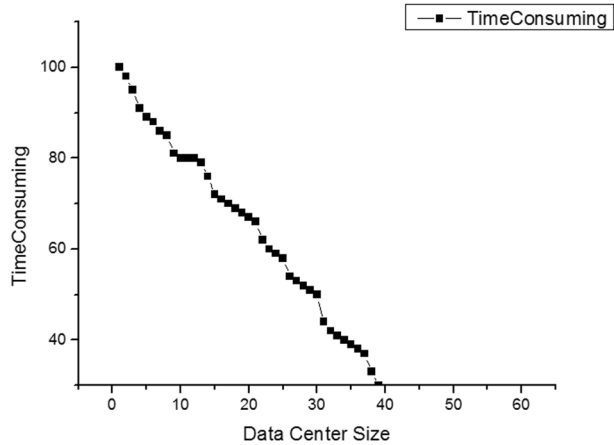
Images	Method 1	Method 2	Method 3	Method 4	Ours
10	0.09	0.11	0.13	0.16	0.19
20	0.15	0.17	0.19	0.21	0.23
30	0.23	0.28	0.28	0.27	0.29
40	0.32	0.33	0.28	0.35	0.36
50	0.41	0.44	0.46	0.51	0.52
60	0.49	0.51	0.51	0.55	0.58
70	0.51	0.53	0.55	0.61	0.63
80	0.55	0.60	0.55	0.58	0.66
90	0.62	0.64	0.66	0.62	0.69
100	0.64	0.68	0.68	0.69	0.77

**Fig. 11** The visualized performance of the experiment algorithms



modification. The experiment proves the robustness of the proposed method. In the future, we will focus on the CNN structural optimization and large scale multimedia data compression for building the more efficient classification paradigm.

**Fig. 12** The time consuming test based on different data center sizes

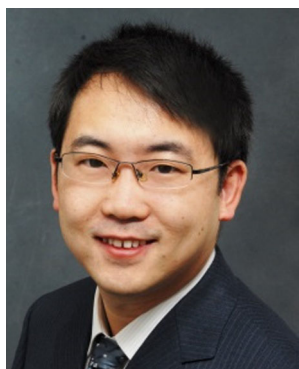


## References

1. Abdelhamid N et al (2012) MAC: a multiclass associative classification algorithm. *J Inf Knowl Manag* 11(02):1250011
2. Bae C et al (2014) Effective audio classification algorithm using swarm-based optimization. *Int J Innov Comput Inf Control* 10(1):151–167
3. Bian J, Gao B, Liu, TY (2014) Knowledge-powered deep learning for word embedding. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 132–148). Springer Berlin Heidelberg.
4. Blake K et al (2015) Use of mobile devices and the internet for multimedia informed consent delivery and data entry in a pediatric asthma trial: Study design and rationale. *Contemp Clin Trials* 42:105–118
5. Chen B et al (2013) Deep learning with hierarchical convolutional factor analysis. *IEEE Trans Pattern Anal Mach Intell* 35(8):1887–1901
6. Chen C et al (2013) A multiresolution hierarchical classification algorithm for filtering airborne LiDAR data. *ISPRS J Photogramm Remote Sens* 82:1–9
7. Chen GW, et al. (2015) A Novel Recognition Method of Multimedia Data for Social Network. *Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence (ACIT-CSI), 2015 3rd International Conference on. IEEE*
8. Chou YC, et al. (2014) A reversible data hiding method using inverse S-scan order and histogram shifting. *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2014 Tenth International Conference on. IEEE*
9. Cui H, Ganger GR, Gibbons PB (2015) Scalable deep learning on distributed GPUs with a GPU-specialized parameter server. *CMU PDL Technical Report (CMU-PDL-15-107)*
10. Dal Mas M. (2015) Layered ontological image for intelligent interaction to extend user capabilities on multimedia systems in a folksonomy driven environment. *Intelligent Interactive Multimedia Systems and Services in Practice. Springer International Publishing*, 103–122
11. Das R, Saha S (2015) Microarray gene expression data classification using modified differential evolution based algorithm. *2015 Annual IEEE India Conference (INDICON). IEEE*
12. de Freitas N (2016) Learning to Learn and Compositionality with Deep Recurrent Neural Networks: Learning to Learn and Compositionality. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 3–3). ACM
13. De Rosa R, Orabona F, Cesa-Bianchi N (2015) The ABACOC algorithm: a novel approach for nonparametric classification of data streams. *Data Mining (ICDM), 2015 I.E. International Conference on. IEEE*
14. Dharani T, Aroquiaraj IL (2013) A survey on content based image retrieval. *Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on. IEEE*
15. Fang Y, Li H, Li X (2014) Lifetime enhancement techniques for PCM-based image buffer in multimedia applications. *IEEE Trans Very Large Scale Integr (VLSI) Syst* 22(6):1450–1455

16. Gelman A, Loken E (2014) The statistical crisis in science data-dependent analysis—a “garden of forking paths”—explains why many statistically significant comparisons don’t hold up. *Am Sci* 102(6):460
17. Giesecke K, Sirignano J, Sadhwani A (2016) Deep Learning for Mortgage Risk. Working Paper, Stanford University
18. Hannagan T, Ziegler JC, Dufau S, Fagot J, Grainger J (2014) Deep learning of orthographic representations in baboons. *PLoS ONE* 9(1):e84843
19. Howell C, Stein F, Kordella S, Booker L, Rockower E, Motahari H, ... Spohrer J (2016) Cognitive Assistance in Government and Public Sector Applications. *KI-Künstliche Intelligenz* 1–2
20. Islamoglu H, Branch RM (2013) Promoting interest, engagement, and deep learning approach in online higher education settings. In: 36 th Annual Proceedings Presented at the Annual Convention of the Association for Educational Communications and Technology (pp. 444–451)
21. Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S (2016) Combining satellite imagery and machine learning to predict poverty. *Science* 353(6301):790–794
22. Jeong Y-S et al (2015) Guest editorial: advanced technologies and services for multimedia Big data processing. *Multimedia Tools Appl* 74(10):3413–3418
23. Jiu M et al (2014) Human body part estimation from depth images via spatially-constrained deep learning. *Pattern Recogn Lett* 50:122–129
24. Krizhevsky A, Hinton GE (2011) Using very deep autoencoders for content-based image retrieval. In: *ESANN*
25. Le QV (2013) Building high-level features using large scale unsupervised learning. In: 2013 I.E. international conference on acoustics, speech and signal processing (pp. 8595–8598). *IEEE*
26. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
27. Lee C-F, Chen S-T, Shen J-J (2015) Reversible dual-image data embedding on pixel differences using histogram modification shifting and cross magic matrix. 2015 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP). *IEEE*
28. Liu Y, Sun C, Lin L, Wang X (2016) Learning Natural Language Inference using Bidirectional LSTM model and Inner-Attention. *arXiv preprint arXiv:1605.09090*
29. Madadi Y, Mohammad ESA, Madadi Y (2013) An accurate classification algorithm with genetic algorithm approach. *Int J Comput Inf Technol* 1(3):198–210
30. Miller K, Morreale P (2014) Finding the needle in the image stack: performance metrics for big data image analysis. *IEEE MultiMedia* 21(1):84–89
31. Naaman M (2012) Social multimedia: highlighting opportunities for search and mining of multimedia data in social media applications. *Multimed Tools Appl* 56(1):9–34
32. Noda K, Yamaguchi Y, Nakadai K, Okuno HG, Ogata T (2015) Audio-visual speech recognition using deep learning. *Appl Intell* 42(4):722–737
33. Norouzi B et al (2014) A novel image encryption based on hash function with only two-round diffusion process. *Multimedia Systems* 20(1):45–64
34. Nouaouria N, Boukadoum M (2014) Improved global-best particle swarm optimization algorithm with mixed-attribute data classification capability. *Appl Soft Comput* 21:554–567
35. Papernot N, McDaniel P, Jha S, Fredrikson M, Celik ZB, Swami A (2016) The limitations of deep learning in adversarial settings. In: 2016 I.E. European Symposium on Security and Privacy (EuroS&P) (pp. 372–387). *IEEE*
36. Santhanam A, Min Y, Beron P, Agazaryan N, Kupelian P, Low D (2016) SU-D-201-05: on the automatic recognition of patient safety hazards in a radiotherapy setup using a novel 3D camera system and a deep learning framework. *Med Phys* 43(6):3334–3335
37. Sarisaray-Boluk P, Akkaya K (2015) Performance comparison of data reduction techniques for wireless multimedia sensor network applications. *Int J Distrib Sensor Netw* 2015:160
38. Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117
39. Schmidler MAR, Borrey R, Sarah A (2014) Data classification using machine learning techniques. U.S. Patent No. 8,719,197. 6 May
40. Sklan JE, Plassard AJ, Fabbri D, Landman BA (2015) Toward content based image retrieval with deep convolutional neural networks. In: *SPIE Medical Imaging* (pp. 94172C-94172C). International Society for Optics and Photonics
41. Song J, Gao L, Zou F, Yan Y, Sebe N (2016) Deep and fast: deep learning hashing with semi-supervised graph construction. *Image Vis Comput*
42. Song B et al (2014) Remotely sensed image classification using sparse representations of morphological attribute profiles. *IEEE Trans Geosci Remote Sens* 52(8):5122–5136
43. Suk HI, Wee CY, Lee SW, Shen D (2016) State-space model with deep learning for functional dynamics estimation in resting-state fMRI. *NeuroImage* 129:292–307

44. Sun Y, et al. (2014) Deep learning face representation by joint identification-verification. *Adv Neural Inf Process Syst*
45. Taneja S, et al. (2014) An enhanced k-nearest neighbor algorithm using information gain and clustering. 2014 Fourth International Conference on Advanced Computing & Communication Technologies. IEEE
46. Tian Y, et al. (2015) Pedestrian detection aided by deep learning semantic tasks. *Proc IEEE Conf Comput Vis Pattern Recognit*
47. Triguero I et al (2015) ROSEFW-RF: the winner algorithm for the ECBDL'14 big data competition: an extremely imbalanced big data bioinformatics problem. *Knowl-Based Syst* 87:69–79
48. Tseng PH, Paolozza A, Munoz DP, Reynolds JN, Itti L (2013) Deep learning on natural viewing behaviors to differentiate children with fetal alcohol spectrum disorder. In: *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 178–185). Springer Berlin Heidelberg
49. Visco C et al (2012) Comprehensive gene expression profiling and immunohistochemical studies support application of immunophenotypic algorithm for molecular subtype classification in diffuse large B-cell lymphoma: a report from the International DLBCL Rituximab-CHOP Consortium Program Study. *Leukemia* 26(9):2103–2113
50. Wan J, Wang D, Hoi SCH, Wu P, Zhu J, Zhang Y, Li J (2014). Deep learning for content-based image retrieval: A comprehensive study. In: *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 157–166). ACM
51. Wang H, Wang J (2014) An effective image representation method using kernel classification. 2014 I.E. 26th International Conference on Tools with Artificial Intelligence. IEEE
52. Wang J, et al. (2015) Multiple kernel multivariate performance learning using cutting plane algorithm. *Systems, Man, and Cybernetics (SMC)*, 2015 I.E. International Conference on. IEEE
53. Wang J, et al. (2016) Optimizing top precision performance measure of content-based image retrieval by learning similarity function. *arXiv preprint arXiv:1604.06620*
54. Xiao SUN, Chengcheng LI, Fuji REN (2016) Sentiment analysis for chinese microblog based on deep neural networks with convolutional extension features. *Neurocomputing*
55. Yang S, Luo P, Loy CC, Tang X (2015) From facial parts responses to face detection: A deep learning approach. In: *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3676–3684)
56. Yang X, Zhang T, Changsheng X (2015) Cross-domain feature learning in multimedia. *IEEE Trans Multimedia* 17(1):64–78
57. Zeng X, et al. (2014) Deep learning of scene-specific classifier for pedestrian detection. *European Conference on Computer Vision*. Springer International Publishing
58. Zhang L et al (2014) Feature correlation hypergraph: exploiting high-order potentials for multimodal recognition. *IEEE Trans Cybern* 44(8):1408–1419



**Jun Li** received B.S and M.S degrees in computer science from Julius Maximilians University, Wuerzburg, Germany, in 2008, and the Ph.D. degree in cognitive system modeling and robotics technology from Ludwig Maximilian University, Germany, in 2012. He finished his Postdoc from Philipps University Marburg, Germany, in 2014. He works currently in the department artificial intelligent and robotics technology of the Haixi Institutes, Chinese Academy of Sciences. His research interests include artificial intelligent, image processing, cognitive System Modeling, Human-Computer Interface and adaptive control in robotics.



**Rishav Singh** completed his B.Tech from CVRCE, Orissa, India, in 2009. He has completed his M.Tech from JUIT, Solan, India, in 2011. He is working as Corporate Trainer in Infosys Limited. His research interest includes Big Data Storage & Analytics, Machine Learning, Biometrics and Image Processing.



**Ritika Singh** completed her B.Tech. in 2006 from JIIT, Noida, India and the MBA from IMI, Delhi, India, in 2010. She completed her Ph.D. in Industrial Engineering and Management from ISM, Dhanbad, India, in 2015. She has worked at Infosys Ltd. as a software engineer from 2006 to 2008 in India. Presently she is Faculty, Amity University, Noida Her research interest includes Big Data Analytics and Information Systems.