

PROBABILISTIC MEMORY MODEL FOR VISUAL IMAGES CATEGORIZATION

Linxia XIAO, Yanjiang WANG*, Baodi LIU, Weifeng LIU

College of Control Science and Engineering

China University of Petroleum (East China)

Qingdao, 266580, China

e-mail: b17050163@s.upc.edu.cn, yjwang, liubaodi, liuwf@upc.edu.cn

Abstract. During the past decades, numerous memory models have been proposed, which focused mainly on how spoken words are studied, whereas models on how visual images are studied are still limited. In this study, we propose a probabilistic memory model (PMM) for visual images categorization which is able to mimic the workings of the human brain during the image storage and retrieval. First, in the learning phase, the visual images are represented by the feature vectors extracted with convolutional neural network (CNN) and each feature component is assumed to conform to a Gaussian distribution and may be incompletely copied with a certain probability or randomly produced in accordance to an exponential distribution. Then, in the test phase, the likelihood ratio between the test image and each studied image is calculated based on the probabilistic inference theory, and an odd value in favor of an old item over a new one is obtained based on all likelihood values. Finally, if the odd value is above a certain threshold, the Bayesian decision rule is applied for image classification. Experimental results on two benchmark image datasets demonstrate that the proposed PMM can perform well on categorization tasks for both studied and non-studied images.

Keywords: Memory model, image categorization, probability distribution, probabilistic inference, Bayesian decision

Mathematics Subject Classification 2010: 68-Txx

* corresponding author

1 INTRODUCTION

As was known to all, computer vision research aims to make the computers easily perceive, understand, and remember the visual images effortlessly as humans do. With the rapid development of cognitive neuroscience and cognitive psychology, numerous new theories and methods have emerged [1], which constantly deepen the study of the complex mechanisms of the human brain and promote the progress of computer vision and artificial intelligence. The fact that the human may retrieve objects from their cluttered surroundings without any difficulties is closely related to the human memory mechanism. Since the time when the first psychology memory model put forward by Atkinson and Shiffrin [2], many memory models have been proposed to date, such as SAM (Search of Associative Memory) [3], BCDMEM (Bind Cue Decide Model of Episodic Memory) [4], TCM (Temporal Context Model) [5], and REM (Retrieving Effective from Memory) [6], etc. However, the majority of these memory models merely focus on the words list study, seldom, if ever, the research is conducted on the representation, storage, and retrieval of visual images.

Recently, machine learning algorithms have been found to be able to categorize visual images with high accuracy and the performances can be even, to some extent, comparable to human beings [7, 8, 9]. However, the key distinction between machine learning based image classification algorithms and human decision is that the machine learning based algorithms will classify an image from a new category into a studied category rather than report unknown like a human. In other words, machine learning based classification methods can distinguish one object from some limited studied objects, while humans can distinguish a certain category of objects from some infinite unknown categories. When one perceives an object never seen before, he will respond ‘unknown’ to it immediately instead of checking it against all the studied objects in his memory, indicating that the human brain memory works in a different way during performing classification tasks. Therefore, in this paper, we will shed light on how visual images are represented and stored as well as how the stored images are retrieved when performing categorization tasks.

Murdock et al. [10] strengthened that a reasonable memory model should address the following issues, i.e., how the information is organized, encoded and stored during learning and how the information could be retrieved or recalled when needed. These topics have gained strong concerns in research areas such as computer vision and image processing and have come a long way. However, the performances yielded are still far from human beings who can make decisions based on the experience and the knowledge learned in the past. Therefore, we can glean some insights from the human brain, especially the mechanism for making an inference and decision to enhance the robustness of the computer vision algorithms. In fact, in the early 1980s, Hinton and Sejnowski showed in their study that the human brain can be regarded as a machine which could infer and make decisions based on external uncertainties [11], suggesting that the brain represents knowledge and makes decisions in the form of probability distributions. Knill and Pouget [12] further developed the idea based on psychophysical evidence and proposed the concept of Bayesian brain which works on

Bayesian probability. In their research, they also described the neural representation mechanism of uncertainty in the human brain. Meanwhile, Lee and Mumford [13] also suggested that the hierarchical Bayesian inference might model the interactive computations in visual cortex and provided some neurophysiological evidences that support the plausibility of their study. Since then, Bayesian models were widely applied in inductive learning and reasoning [14], free recall of memory [15], visual search and recognition [16] as well as visual attention [17].

In this paper, in order to imitate the memory function of the human brain, we apply the probabilistic inference and Bayesian decision theories to the modeling of visual images storage and retrieval and put forward a probabilistic memory model (PMM) for visual images categorization. First, regarding the image representation and storage, we apply the deep convolutional neural network (CNN) to extract the image features, and each feature component is supposed to conform to a Gaussian distribution and may be incompletely copied with a certain probability or randomly produced according to an exponential distribution. The probability representation aims to mimic the sensory error as well as the memory noise of the human brain in the study phase. Then the likelihood ratio between the test image and all the studied images can be obtained, and an odd value can be computed based on all the likelihood values, which is regarded as a criterion for deciding whether the image is studied or not. For image retrieval, the Bayesian decision rule is performed. We evaluate the performance of the proposed PMM by applying it on image categorization tasks in which experiments are conducted on two benchmark image databases and the results are compared with the state-of-the-art image classification methods.

The structure of the paper is organized as follows. Section 2 elaborates the related work. Section 3 introduces the proposed PMM model using probabilistic inference and Bayesian decision. Section 4 provides the experiments and performances of PMM on visual image classification tasks. Finally, some conclusions are summarized in Section 5.

2 RELATED WORK

In this section, we will retrospect some relevant work about visual image representation, storage and recall, i.e., feature representation and the traditional memory model – REM model [6].

As for feature representation, currently a mass of methods for images feature extraction have been put forward, such as histogram of oriented gradient (HOG) [18], local binary patterns (LBP) [19], Haar-like [20] and gradient location-orientation histogram (GLOH) [21], etc. These methods can achieve good results when applied to simple images, but the effects will get worse when dealing with complex natural images. Recently, deep learning based feature extraction methods exhibit better performance when handling images with noise or cluttered background in comparison to the traditional approaches, such as autoencoder neural networks (ANN) [22], restricted Boltzmann machine (RBM) [23], recurrent neural network (RNN) [24] and

convolutional neural networks (CNN) [25], etc. Notably, the hierarchical convolutional neural network (HCNN) [26] can model the neural single-unit and population responses in the visual cortical areas of the human brain, especially, the VGG-19 model [25] is known for its deep representation of visual images.

As for the traditional memory models, the REM model is the most typical memory model for words list study. The REM model supposed that human memory consisted of a series of separate images, with each one being a vector of feature values, which are stored randomly or incompletely copied from the studied vector with error. When a word was studied, the storage probability of each feature was μ^* . It was worth noting that when any feature value was stored, it did not change any longer. When some information was stored in the feature vector, the probability of stored value consisted of two parts: the rightly copied probability was c and the randomly chosen probability was $1 - c$ which satisfied a geometric probability distribution $P(V = j) = (1 - g)^{j-1}g$, $j = 1, 2, \dots, \infty$. Given the test vector either studied or unstudied, it was matched against the studied feature vectors. The matching result $D = \{D_j\}_{j=1,2,\dots,N}$ was then computed to find the matched and non-matched locations, and neglected the locations where the feature value was zero, where D_j denoted the aligning result between the test and the j^{th} word feature vector.

Next, calculated the likelihood ratio λ_j based on the matching result D_j :

$$\lambda_j = \frac{P(D_j|S_j)}{P(D_j|N_j)} \quad (1)$$

where S_j and N_j stood for the events that the j^{th} image was an s -image and a d -image, respectively. Finally, calculated the odds in favor of an old over a new item based on the likelihood values:

$$\Phi = \frac{1}{N} \sum_{j=1}^N \lambda_j. \quad (2)$$

If $\Phi > 1$, the test word was judged as an old one; or else, the test word was identified as new.

REM model offers a memory mechanism to elucidate a series of episodic memory phenomena, such as the list-length, list-strength, word-frequency and mirror effects, etc. However, it can only provide some accounts for word list study. Due to the complex characteristics of visual images, the research about the representation, storage, and retrieval of visual images is still a challenging task.

3 THE PROPOSED PMM MODEL

In what follows, we will elaborate the proposed PMM modeling process, including image representation and storage with probability, memory recall by probability reference, and memory retrieval with Bayesian decision.

3.1 Visual Image Representation and Storage with Probability

The primary part of the memory modeling process is feature representation for the information to be stored, which is essential for the performance of the model. In words list study, the feature of a word is expressed by integers, which is very simple and easy to handle, whereas the feature of an image is much more complicated. Neural evidence shows that the visual information in human brain is processed hierarchically [27, 28]. Thus a pre-trained convolutional neural network model with 19 layers, named VGG-19 [25], is adopted to extract the primitive image features.

More specifically, given a visual image set $X = \{X_{11}, X_{12}, \dots, X_{1n_1}, X_{21}, X_{22}, \dots, X_{2n_2}, \dots, X_{C1}, X_{C2}, \dots, X_{Cn_C}\}$, where X_{ij} denotes the j^{th} image coming from the i^{th} category, C represents the number of categories in the image set and is signified by $Z = Z_1, Z_2, \dots, Z_C$, and the number of images in the i^{th} category Z_i is n_i . We first employ the VGG-19 model to extract the feature values of images during the study phase, which are represented as $F = \{F_{11}, F_{12}, \dots, F_{1n_1}, F_{21}, F_{22}, \dots, F_{2n_2}, \dots, F_{C1}, F_{C2}, \dots, F_{Cn_C}\}$, where $F_{ij} \in \mathbf{R}^{1 \times D}$ indicates the feature vector of image X_{ij} , D implies the dimension of the feature vector.

To simulate the visual sensor error, for computational simplicity and convenience, each extracted image feature is modeled using a Gaussian distribution:

$$q(F_{ij}|\mu_i) \sim N(f_{ij} : \mu_i, \sigma^2) \quad i \in [1, C] \tag{3}$$

where f_{ij} represents the observed feature values of the j^{th} image from the i^{th} category, μ_i is the mean value of the i^{th} category, and σ^2 is the precision parameter. Suppose each feature value of images of each category is independent for each other, then the probability distribution of the image features of each category takes the form:

$$q(F|\mu) = \prod_{i=1}^C q(F_{ij}|\mu_i). \tag{4}$$

To imitate the storage process of the human brain, we hypothesize that the feature vector of each studied image is stored either correctly copied or an incomplete and error prone copy, suggesting the image features extracted by CNN are incomplete, with some feature components produced randomly. Here, we assume that each stored feature value may be either correctly copied with probability ρ , satisfying a Gaussian distribution, or randomly produced with probability $1 - \rho$, conforming to an exponential distribution (allowing the probability of accidentally selecting the correct value):

$$g(F = x) = \begin{cases} \rho e^{-gx}, & x > 0, \\ 1 - \rho, & \text{otherwise,} \end{cases} \tag{5}$$

where g is the exponential distribution parameter. The reason for choosing exponential distribution is that exponential distribution can describe the time probability

distribution between the events in the Poisson process, which is broadly consistent with the neurons encoding process in the human brain [29].

In summary, the probabilistic storage process of the PMM can be described by a 4-tuple, i.e., the three parameters of the two probability distributions and one correctly copied probability ρ :

$$F \sim \{\mu, \sigma, g, \rho\}. \quad (6)$$

3.2 Visual Images Recall

Given a test image X_t , memory recall means determining whether it is studied or not. The feature vector $F_t \in \mathbf{R}^{1 \times D}$ of the test image X_t is extracted by deep CNN first and then matched with the studied feature set $F = \{F_{11}, \dots, F_{Cn_C}\}$. The matching rule between the test feature vector F_t and the learned feature vector F_{ij} is set as follows:

$$R_{ij}(d) = \begin{cases} 1, & \text{if } F_t(d) \neq 0 \text{ and } F_{ij}(d) \neq 0, \\ -1, & \text{if } F_t(d) \neq 0 \text{ and } F_{ij}(d) = 0, \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

where $1 \leq d \leq D$, in what follows, $F_t(d)$ and $F_{ij}(d)$ are abbreviated as F_d and $F_{d,ij}$ respectively. The matched result R_{ij} is used to find those positions whose values match or mismatch. When F_t and F_{ij} are matched, $R_{ij} = 1$, when F_t and F_{ij} are not matched, $R_{ij} = -1$, the other cases are ignored. The whole matching result is expressed by $R = \{R_{11}, R_{12}, \dots, R_{Cn_C}\}$.

Then the likelihood ratio L_{ij} is calculated based on each matching result R_{ij} and equals the probability that image X_{ij} is an s -image (same as the test image) over the probability being a d -image (different from the test image). Suppose the prior probability of s -image and d -image is identical, then we have:

$$L_{ij} = \frac{P(Y_{ij}|R_{ij})}{P(N_{ij}|R_{ij})} = \frac{\frac{P(Y_{ij})P(R_{ij}|Y_{ij})}{P(R_{ij})}}{\frac{P(N_{ij})P(R_{ij}|N_{ij})}{P(R_{ij})}} = \frac{P(R_{ij}|Y_{ij})}{P(R_{ij}|N_{ij})} \quad (8)$$

where Y_{ij} and N_{ij} represent the events that image X_{ij} is an s -image and a d -image, respectively. Then,

$$L_{ij} = \prod_{d=1}^D \frac{P(F_{d,ij}|Y_{ij}, F_d)}{P(F_{d,ij}|N_{ij}, F_d)} \quad (9)$$

where F_d denotes the d^{th} feature value coming from the test image, $F_{d,ij}$ is the d^{th} feature value of image X_{ij} . $P(F_{d,ij}|Y_{ij}, F_d)$ is the probability when X_{ij} is an s -image, the d^{th} feature value in the test image is F_d , and the d^{th} feature value in the feature vector of the image X_{ij} is $F_{d,ij}$. $P(F_{d,ij}|N_{ij}, F_d)$ is the probability when X_{ij} is an d -image, the d^{th} feature value in the test image is F_d , and the d^{th} feature value in

Step 1. Studied features																		
0 2.1792 0 0 0 0 0				0 0 8.8967 0 0 0 0				0 2.4122 0 0 0.0617 0										
0 0 2.2379 0 0 0 0				0 12.0695 0 0.0218 0 0 0				0 9.6202 0 0 0 0 3.9504										
Step 2. Mean μ and variance σ^2 of studied features																		
μ	0	1.0896	1.1190	0	0	0	0	6.0348	4.4484	0.0109	0	0	0	6.0162	0	0	0.0309	1.9752
σ^2	0	1.1190	1.2522	0	0	0	0	36.4182	19.7883	0.0001	0	0	0	36.1947	0	0	0.0010	3.9014
Step 3. the probe feature																		
0 0.6078 0.6281 0 0 0																		
Step 4. Matched and non-matched location																		
0 1 -1 0 0 0 0				0 -1 1 0 0 0 0				0 1 -1 0 0 0 0										
0 -1 1 0 0 0 0				0 1 -1 0 0 0 0				0 1 -1 0 0 0 0										
Step 5. Matching result																		
\ 1.0225 0.7 \ \ \				\ 0.7 0.7849 \ \ \				\ 0.7627 0.7 \ \ \										
\ 0.7 1.0183 \ \ \				\ 0.7625 0.7 \ \ \				\ 0.7627 0.7 \ \ \										
Step 6. The likelihood ratio																		
$L_{11}=0.7158$				$L_{21}=0.5494$				$L_{31}=0.5339$										
$L_{12}=0.7128$				$L_{22}=0.5338$				$L_{32}=0.5339$										
Step 7. The class likelihood ratio																		
$L_1=\sum_j L_{1j}=1.4286$				$L_2=\sum_j L_{2j}=1.0832$				$L_3=\sum_j L_{3j}=1.0678$										
Step 8. Φ (odds)																		
$\Phi=(L_1+L_2+L_3)=1.1932$																		
Step 9. Decision																		
$\Phi>1$, so the probe image is old $L_1=\max(L_1,L_2,L_3)$, so the probe image is classified as the first class.																		

Figure 1. An illustration of the storage and retrieval process for PMM

the feature vector of image X_{ij} is $F_{d,ij}$. According to the matching results in R , Equation (9) can be decomposed into

$$L_{ij} = \prod_{d \in Q} \frac{P(F_{d,ij}|Y_{ij}, F_d)}{P(F_{d,ij}|N_{ij}, F_d)} \prod_{d \in M} \frac{P(F_{d,ij}|Y_{ij}, F_d)}{P(F_{d,ij}|N_{ij}, F_d)} \tag{10}$$

where M and Q are the sets of exponents for the matched and nonmatched non-zero features, respectively. When $d \in Q$, for s -images, the d^{th} feature of X_{ij} does not match with the test image, meaning the feature is wrongly copied, i.e.,

$$P(F_{d,ij}|Y_{ij}, F_d) = (1 - \rho)P(F_{d,ij}|N_{ij}, F_d). \tag{11}$$

Let $n_{ij,Q}$ be the number of all mismatched non-zero feature values in image X_{ij} , then by substituting Equation (11) into Equation (10), we obtain:

$$L_{ij} = (1 - \rho)^{n_{ij,Q}} \prod_{d \in M} \frac{P(F_{d,ij}|Y_{ij}, F_k)}{P(F_{d,ij}|N_{ij}, V_d)}. \tag{12}$$

When $d \in M$, the d^{th} feature of X_{ij} matches with the test image, for d -images, the probability of storing value $F_{d,ij}$ satisfies an exponential distribution $g(F)$, as

shown in Equation (5). Whereas for s -images, the probability of storing value $F_{d,ij}$ consists of two parts: the correctly copied probability which satisfies the Gaussian distribution and the randomly chosen probability which is distributed exponentially, i.e.,

$$P(F_{d,ij}|Y_{ij}, F_d) = \rho q(F_{d,ij}) + (1 - \rho)g(F_{d,ij}), \tag{13}$$

$$\frac{P(F_{d,ij}|Y_{ij}, F_d)}{P(F_{d,ij}|N_{ij}, F_d)} = \frac{\rho q(F_{d,ij}) + (1 - \rho)g(F_{d,ij})}{g(F_{d,ij})} \tag{14}$$

where $q(F_{d,ij}) = \frac{1}{\sqrt{2\pi\sigma}} \exp[-\frac{1}{2\sigma^2}(F_{d,ij} - \mu_i)^2]$ and $g(F_{d,ij}) = \exp(-gF_{d,ij})$. Then, we have

$$\begin{aligned} L_{ij} &= (1 - \rho)^{n_{ij,Q}} \prod_{d \in M} \frac{\rho q(F_{d,ij}) + (1 - \rho)g(F_{d,ij})}{g(F_{d,ij})} \\ &= (1 - \rho)^{n_{ij,Q}} \prod_{d \in M} \frac{\rho q(F_{d,ij}) + (1 - \rho)ge^{-gF_{d,ij}}}{ge^{-gV_{d,ij}}}. \end{aligned} \tag{15}$$

In summary, the likelihood ratio set symbolized as $L = \{L_{ij}\}_{i=1, \dots, C, j=1, \dots, n_C}$ can be obtained by checking against the given test image X_t with all studied images $X = \{X_{ij}\}_{i=1, \dots, C, j=1, \dots, n_C}$.

With the above likelihood ratios, whether the test image X_t is studied (old) or not (new) during the learning phase can be determined.

Let Φ denote the odds, which equals the probability that the test image X_t is old over the probability being new. Then we have,

$$\Phi = \frac{P(O|R)}{P(N|R)} = \frac{\frac{P(O)P(R|O)}{P(R)}}{\frac{P(N)P(R|N)}{P(R)}} = \frac{P(R|O)}{P(R|N)} \tag{16}$$

where $P(R|O) = \sum_{i=1}^C \sum_{j=1}^{n_i} P(R|Y_{ij})P(Y_{ij})$, C is the number of image categories, n_i is the number of images of the i^{th} category Z_i .

Assume the probability of being an s -image for each class is identical and the probability of being an s -image for each test image is identical too, i.e. $P(Y_{ij}) = \frac{1}{C} \frac{1}{n_i}$. Then,

$$\Phi = \sum_{i=1}^C \sum_{j=1}^{n_i} \frac{1}{C} \frac{1}{n_i} \frac{P(R|Y_{ij})}{P(R|N)} = \frac{1}{C} \sum_{i=1}^C \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{P(R|Y_{ij})}{P(R|N)}. \tag{17}$$

When the studied X_{ij} is an s -image, i.e., event Y_{ij} happens, the other images are d -image, i.e., N event happens, then

$$P(R|Y_{ij}) = P(R_{ij}|Y_{ij}) \prod_{i \neq i', j \neq j'} P(R_{i'j'}|N_{i'j'}). \tag{18}$$



Figure 2. Exemplars in Caltech-101 dataset

Finally, we have

$$\begin{aligned}
 \Phi &= \frac{1}{C} \sum_{i=1}^C \frac{1}{n_i} \left(\frac{\sum_{j=1}^{n_i} P(R_{ij}|Y_{ij}) \prod_{i \neq i', j \neq j'} P(R_{i'j'}|N_{i'j'})}{\sum_{j=1}^{n_i} P(R_{ij}|N_{ij}) \prod_{i \neq i', j \neq j'} P(R_{i'j'}|N_{i'j'})} \right) \\
 &= \frac{1}{C} \sum_{i=1}^C \frac{1}{n_i} \left(\sum_{j=1}^{n_i} \frac{P(R_{ij}|Y_{ij})}{P(R_{ij}|N_{ij})} \right) \\
 &= \frac{1}{C} \sum_{i=1}^C \frac{1}{n_i} \sum_{j=1}^{n_i} L_{ij} \\
 &= \frac{1}{C} \sum_{i=1}^C \sum_{j=1}^{n_i} \frac{1}{n_i} L_{ij}. \tag{19}
 \end{aligned}$$

Given a threshold θ , if $\Phi > \theta$, the test image is judged as ‘studied’ (old), otherwise ‘unstudied’ (new).

3.3 Memory Retrieval with Bayesian Decision

When the test image is studied, the Bayesian decision rule can be adopted to determine its category, i.e., memory retrieval.

Given a test image X_t , assume there are two different classes Z_a and Z_b , $1 \leq a$ and $b \leq C$. Assume the posterior probability of being class Z_a and Z_b are $P(Z_a|X_t)$ and $P(Z_b|X_t)$, $1 \leq a$ and $b \leq C$, respectively, then according to Bayesian decision theory,

$$X_t \in \begin{cases} Z_a, & \text{if } P(Z_a|X_t) > P(Z_b|X_t), \\ Z_b, & \text{otherwise.} \end{cases} \tag{20}$$

Let $\{X_{aj}\}_{j=1, \dots, n_a}$ and $\{X_{bj}\}_{j=1, \dots, n_b}$ denote the studied images for the two classes Z_a and Z_b , respectively. Hence, for two-category problem, the Bayesian decision rule

takes the form,

$$I_t \in \begin{cases} Z_a, & \text{if } \frac{P(Z_a|X_t)}{P(Z_b|X_t)} > 1, \\ Z_b, & \text{otherwise.} \end{cases} \quad (21)$$

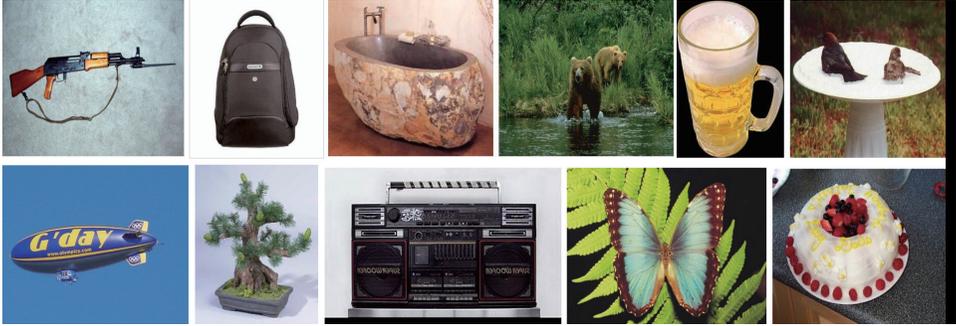


Figure 3. Exemplars in Caltech-256 dataset

$$\begin{aligned} \frac{P(Z_a|X_t)}{P(Z_b|X_t)} &\propto \frac{P(R|Y_a)}{P(R|Y_b)} \\ &= \frac{\sum_{j=1}^{n_a} P(R|Y_{aj})P(Y_{aj})}{\sum_{j=1}^{n_b} P(R|Y_{bj})P(Y_{bj})} \\ &= \frac{\frac{1}{C} \frac{1}{n_a} \sum_{j=1}^{n_a} P(R|Y_{aj})}{\frac{1}{C} \frac{1}{n_b} \sum_{j=1}^{n_b} P(R|Y_{bj})} \\ &\propto \frac{\frac{1}{n_a} \sum_{j=1}^{n_a} P(R_{aj}|Y_{aj})}{\frac{1}{n_b} \sum_{j=1}^{n_b} P(R_{bj}|Y_{bj})} \\ &\propto \frac{\frac{1}{n_a} \sum_{j=1}^{n_a} L_{aj}}{\frac{1}{n_b} \sum_{j=1}^{n_b} L_{bj}} \end{aligned} \quad (22)$$

Applying the Bayesian decision rule, we have,

$$X_t \in \begin{cases} Z_a, & \text{if } \frac{1}{n_a} \sum_{j=1}^{n_a} L_{aj} > \frac{1}{n_b} \sum_{j=1}^{n_b} L_{bj}, \\ Z_b, & \text{otherwise.} \end{cases} \quad (23)$$

Similarly, for multi-category classification problem, given the studied class set $Z = Z_1, Z_2, \dots, Z_C$, then the final Bayesian decision rule can be defined as

$$X_t \in Z_a, \text{ if } \frac{1}{n_a} \sum_{j=1}^{n_a} L_{aj} > \frac{1}{n_b} \sum_{j=1}^{n_b} L_{bj} \quad (24)$$

for all $b \neq a$, $1 \leq a, b \leq C$. The storage and retrieval process for PMM is illustrated in Figure 1, where the parameters ρ , g and θ are assumed to be 0.1, 0.01 and 1, respectively. The process includes nine steps:

The storage and retrieval process for PMM
<i>Step 1:</i> Given the studied feature vectors coming from three different classes with each consisting of two images.
<i>Step 2:</i> Estimate the corresponding means μ and variance σ^2 of the two studied feature values in each class, respectively.
<i>Step 3:</i> Present the feature vector of the test image whose category needs to be judged.
<i>Step 4:</i> Find the matched and nonmatched locations between the feature values of the studied images and the test image according to Equation (7), where 1 represents the matched features, -1 stands for the nonmatched features.
<i>Step 5:</i> Given the matching results. For the positions where the value is 0, the matching result is ignored. For nonmatched positions where the value is -1, the matching result is set as $1-\rho$, whereas for the matched positions where the value is 1, the matching result is computed according to Equation (14).
<i>Step 6:</i> Calculate the likelihood ratio L_{ij} by multiplying all the matching results for each image according to Equation (15).
<i>Step 7:</i> Calculate the class likelihood ratio based on the summation of all the ratios of the two studied images from the same category.
<i>Step 8:</i> Compute the odd value Φ based on all likelihood values according to Equation (19).
<i>Step 9:</i> Make the decision. If the odd value is larger than a certain threshold, the test image is decided as old, and the Bayesian decision rule is applied to the image classification, otherwise decided as new.

4 EXPERIMENTS

For the purpose of performance evaluation of the proposed PMM, we conduct the experiments on two benchmark image databases, Caltech-101 [30] and Caltech-256 [31]. The experiments are implemented on MATLAB R2017b. The test environment is a laptop computer with Intel (R) Core (TM) i7-4770 CPU @ 3.40 GHz, 16 GB memory, and Windows10 operating system. First, the parameters of the PMM are tuned to reach the optimal performance, and then the effects on image classification by PMM are compared with several image classification methods, including the traditional support vector machine (SVM) [32] and some state-of-the-art methods such as VGG-19 [33], ResNet-50 [34], SLRC [35], ICS-DLSR [36], WKRBM [37], Euler-SRC [38] and NRC [39]. Finally, the memory recognition performance of PMM is evaluated. All the experimental results are averaged by 10 runs.



a) Exemplars of training set



b) Exemplars of studied images for testing

c) Exemplars of unstudied images for testing

Figure 5. Exemplars of training images and test images on Caltech-256 dataset

then randomly choose 10 images of each category for hit rate evaluation, meanwhile randomly choose 20 new categories (10 images of each category) for false alarm evaluation, that is, the test images include 200 studied images and 200 new categories images. Some image exemplars are shown in Figure 4 and Figure 5. Figure 6 and Figure 7 demonstrate the comparison results of the recognition performance under varying parameters g and ρ . As shown in Figure 6 and Figure 7, for both datasets, when $g = 0.01$ and $g = 0.02$, the hit rates are higher than other g , while when $g = 0.02$, the false alarm rates are obviously lower than $g = 0.01$. When $\rho = 0.1$, the hit rates are higher and the false alarm rates are lower than other ρ . Accordingly, the optimal parameters are set as $g = 0.02$ and $\rho = 0.1$ for both datasets.

4.3 Image Classification with PMM

For further validation of the pure classification performance, we compare the PMM model with SVM [32] and some state-of-the-art methods such as VGG-19 [33], ResNet-50 [34], SLRC [35], ICS-DLSR [36], WKRBM [37], Euler-SRC [38] and NRC [39] on both datasets. In the experiment, for each dataset, we randomly select 20, 40, 80 categories and 30 images per category, including 10 images as training set and 20 images as test set or 25 images as training set and 5 images as test set, respectively. The image features are extracted with the trained convolutional neural network VGG-19 model except the ResNet-50 method, in which the image features are extracted with the trained residual neural network model. In SVM, the extracted features are fed into a linear support vector machine (LIB-SVM [32]). As

shown in Tables 1 and 2, the classification performance of PMM has the highest classification rate in comparison to all the other eight models with both datasets.

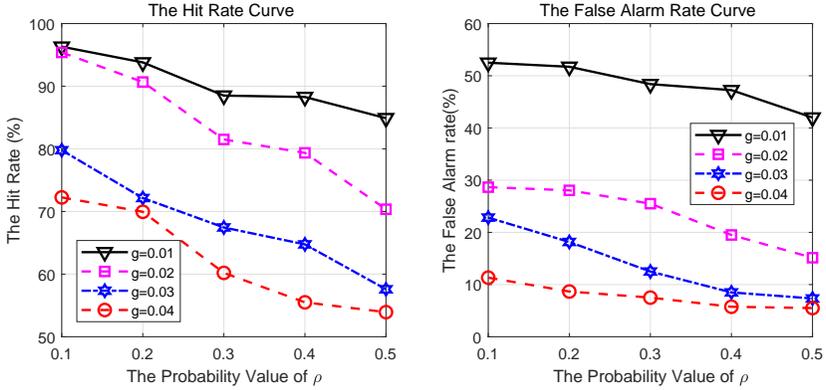


Figure 6. Comparison results of the memory recognition performance under varying parameters ρ and g on Caltech-101 dataset

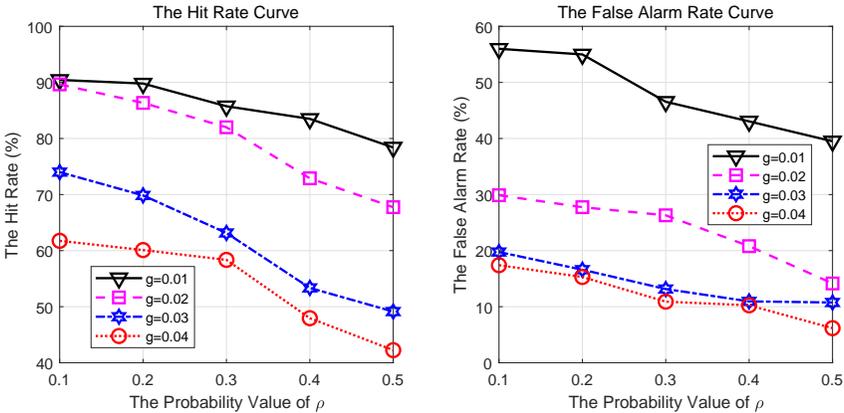


Figure 7. Comparison results of the memory recognition performance under varying parameters ρ and g on Caltech-256 dataset

4.4 Recognition Performance Evaluation

Memory recognition performance means the rate of responding ‘old’ or ‘new’ to a probe image. To evaluate the effects of recognition, we compare the PMM model with some machine learning based approaches including VGG-19 [33], ResNet-50 [34] and NRC [39] on both datasets, which also show good classification performance. In

Images	Methods	20 categories	40 categories	80 categories
10 train/20 test	PMM	95.80	94.55	91.65
	SVM	91.73	87.94	83.96
	VGG-19	92.72	91.34	91.31
	ResNet-50	95.10	91.63	91.35
	SLRC	90.24	88.97	87.60
	ICS-DLSR	89.61	87.82	86.80
	WKRBM	88.51	87.31	85.22
	Euler-SRC	88.11	87.16	86.77
	NRC	93.55	91.47	90.32
25train/5test	PMM	97.80	96.40	92.85
	SVM	92.80	91.85	88.95
	VGG-19	96.52	94.62	92.75
	ResNet-50	96.60	94.70	92.92
	SLRC	92.81	90.61	88.53
	ICS-DLSR	89.91	88.32	87.34
	WKRBM	92.51	90.43	89.53
	Euler-SRC	91.89	90.13	88.64
	NRC	94.11	93.72	91.96

Table 1. Comparison classification performance of PMM with the other eight methods on Caltech-101 dataset (%)

Images	Methods	20 categories	40 categories	80 categories
10 train/20 test	PMM	91.83	87.31	84.89
	SVM	85.72	80.45	76.65
	VGG-19	88.22	87.09	83.56
	ResNet-50	91.05	87.23	83.96
	SLRC	88.69	86.25	81.81
	ICS-DLSR	87.95	85.83	80.62
	WKRBM	88.53	84.41	81.38
	Euler-SRC	88.43	85.75	82.12
	NRC	89.25	86.33	83.31
25 train/5 test	PMM	93.78	90.44	86.44
	SVM	89.56	88.61	82.81
	VGG-19	90.01	89.72	86.15
	ResNet-50	92.40	89.95	86.50
	SLRC	90.86	87.50	84.50
	ICS-DLSR	91.31	88.83	84.08
	WKRBM	91.38	87.13	83.66
	Euler-SRC	91.85	87.65	84.25
	NRC	91.80	88.75	85.75

Table 2. Comparison classification performance of PMM with the other eight methods on Caltech-256 dataset (%)

the experiment, for each dataset, we randomly choose 20 categories and 20 images per category for training, then randomly choose 10 images of each category for testing, meanwhile randomly choose 20 new categories (10 images of each category) for testing, that is, the test images include 200 studied images and 200 new category images. Tables 3 and 4 list the hit rates and false alarm rates for each of the methods for both datasets. It can be seen that these machine learning based methods fail to decide whether a test image is studied or not, whereas the proposed PMM is able to recognize most of the unstudied images.

Methods	P(H)	P(F)
PMM	96.07	29.28
VGG-19	91.85	100
ResNet-50	93.33	100
NRC	92.25	100

Table 3. Comparison recognition results of the hit rates P(H) (%) and the false alarm rates P(F) (%) of PMM, VGG-19, ResNet-50 and NRC methods on Caltech-101 dataset

Methods	P(H)	P(F)
PMM	91.67	29.98
VGG-19	87.37	100
ResNet-50	90.33	100
NRC	88.78	100

Table 4. Comparison recognition results of the hit rates P(H) (%) and the false alarm rates P(F) (%) of PMM, VGG-19, ResNet-50 and NRC methods on Caltech-256 dataset

5 CONCLUSIONS

In this paper, we have presented a probabilistic memory model (PMM) for visual image storage and recall based on probabilistic inference and Bayesian decision. We assume that the original visual image feature values extracted by deep CNN are stored stochastically, with being correctly copied with some probability or randomly produced conforming to an exponential distribution. The probabilistic generative model can facilitate the matching process between the test image and the studied images during learning. Thus, when a new object of a new category is presented, PMM can judge it as an unknown object instead of identifying it from a studied category. Additionally, the proposed PMM can also be applied to the classification of natural multi-category images. We have conducted experiments on Caltech-101 and Caltech-256 databases to validate the effectiveness of PMM. The results manifest that PMM achieves better classification performance in image classification tasks than most state-of-the-art image classification methods. It is worth highlighting that PMM can also determine the images from an unstudied category, whereas the other methods fail.

The current PMM can imitate the human brain memory for visual images learning to some extent and offer a new way for visual images categorization, however, there are also some limitations to PMM. First, the probabilistic model parameters are chosen experimentally and might be influenced by the image datasets used. Secondly, the matching rule between the test feature vector and the learned feature vectors seems simple and might not be consistent with the way of humans memorizing visual images. These two issues will remain for further investigation in our future work.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (No. 62072468), the Shandong Provincial Natural Science Foundation, China (Nos. ZR2018MF017, ZR2019MF073), and the Fundamental Research Funds for the Central Universities, China University of Petroleum (East China) (No. 18CX060-60A).

REFERENCES

- [1] LEE, M. D.—VANPAEMEL, W.: Determining Informative Priors for Cognitive Models. *Psychonomic Bulletin and Review*, Vol. 25, 2018, No. 1, pp. 114–127, doi: 10.3758/s13423-017-1238-3.
- [2] ATKINSON, R. C.—SHIFFRIN, R. M.: Human Memory: A Proposed System and Its Control Processes. *Psychology of Learning and Motivation*, Vol. 2, 1968, No. 1, pp. 89–195, doi: 10.1016/s0079-7421(08)60422-3.
- [3] RAAIJMAKERS, J. G. W.—SHIFFRIN, R. M.: SAM: A Theory of Probabilistic Search of Associative Memory. *Psychology of Learning and Motivation*, Vol. 14, 1980, pp. 207–262, doi: 10.1016/s0079-7421(08)60162-0.
- [4] DENNIS, S.—HUMPHREYS, M. S.: A Context Noise Model of Episodic Word Recognition. *Psychological Review*, Vol. 108, 2001, pp. 452–478, doi: 10.1037/0033-295x.108.2.452.
- [5] POLYN, S. M.—KAHANA, M. J.: Memory Search and the Neural Representation of Context. *Trends in Cognitive Sciences*, Vol. 12, 2008, No. 1, pp. 24–30, doi: 10.1016/j.tics.2007.10.010.
- [6] SHIFFRIN, R. M.—STEYVERS, M.: A Model for Recognition Memory: REM – Retrieving Effectively from Memory. *Psychonomic Bulletin and Review*, Vol. 4, 1997, No. 2, pp. 145–166, doi: 10.3758/bf03209391.
- [7] LIU, B.—JING, L.—LI, J.—YU, J.—GITTENS, A.—MAHONEY, M. W.: Group Collaborative Representation for Image Set Classification. *International Journal of Computer Vision*, Vol. 127, 2019, No. 2, pp. 181–206, doi: 10.1007/s11263-018-1088-0.
- [8] WAN, W.—ZHONG, Y.—LI, T.—CHEN, J.: Rethinking Feature Distribution for Loss Functions in Image Classification. *Proceedings of 2018 IEEE/CVF Conference*

- on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, pp. 9117–9126, doi: 10.1109/cvpr.2018.00950.
- [9] LIU, B. D.—GUI, L.—WANG, Y.—WANG, Y. X.—SHEN, B.—LI, X.—WANG, Y. J.: Class Specific Centralized Dictionary Learning for Face Recognition. *Multimedia Tools and Applications*, Vol. 76, 2017, No. 3, pp. 4159–4177, doi: 10.1007/s11042-015-3042-2.
- [10] MURDOCK, B. B.: A Theory for the Storage and Retrieval of Item and Associative Information. *Psychological Review*, Vol. 89, 1982, No. 6, pp. 609–626, doi: 10.1037/0033-295x.89.6.609.
- [11] HINTON, G. E.—SEJNOWSKI, T. J.: Optimal Perceptual Inference. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Washington, D. C., June 1983, pp. 448–453.
- [12] KNILL, D. C.—POUGET, A.: The Bayesian Brain: The Role of Uncertainty in Neural Coding and Computation. *Trends in Neurosciences*, Vol. 27, 2004, No. 12, pp. 712–719, doi: 10.1016/j.tins.2004.10.007.
- [13] LEE, T. S.—MUMFORD, D.: Hierarchical Bayesian Inference in the Visual Cortex. *Journal of the Optical Society of America A – Optics Image Science and Vision*, Vol. 20, 2003, No. 7, pp. 1434–1448, doi: 10.1364/josaa.20.001434.
- [14] TENENBAUM, J. B.—GRIFFITHS, T. L.—KEMP, C.: Theory-Based Bayesian Models of Inductive Learning and Reasoning. *Trends in Cognitive Sciences*, Vol. 10, 2006, No. 7, pp. 309–318, doi: 10.1016/j.tics.2006.05.009.
- [15] SOCHER, R.—GERSHMAN, S.—SEDERBERG, P.—NORMAN, K.—PEROTTE, A.—BLEI, D.: A Bayesian Analysis of Dynamics in Free Recall. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., Culotta, A. (Eds.): *Advances in Neural Information Processing Systems 22 (NIPS 2009)*, 2009, pp. 1714–1722.
- [16] ELAZARY, L.—ITTI, L.: A Bayesian Model for Efficient Visual Search and Recognition. *Vision Research*, Vol. 50, 2010, No. 14, pp. 1338–1352, doi: 10.1016/j.visres.2010.01.002.
- [17] CHIKKERUR, S.—SERRE, T.—TAN, C.—POGGIO, T.: What and Where: A Bayesian Inference Theory of Attention. *Vision Research*, Vol. 50, 2010, No. 22, pp. 2233–2247, doi: 10.1016/j.visres.2010.05.013.
- [18] SONG, S.—XU, B.—YANG, J.: SAR Target Recognition via Supervised Discriminative Dictionary Learning and Sparse Representation of the SAR-HOG Feature. *Remote Sensing*, Vol. 8, 2016, No. 8, Art.No. 683, doi: 10.3390/rs8080683.
- [19] KAPLAN, K.—KAYA, Y.—KUNCAN, M.—MINAZ, M. R.—ERTUNÇ, H. M.: An Improved Feature Extraction Method Using Texture Analysis with LBP for Bearing Fault Diagnosis. *Applied Soft Computing*, Vol. 87, 2020, Art.No. 106019, doi: 10.1016/j.asoc.2019.106019.
- [20] AI, J.—TIAN, R.—LUO, Q.—JIN, J.—TANG, B.: Multi-Scale Rotation-Invariant Haar-Like Feature Integrated CNN-Based Ship Detection Algorithm of Multiple-Target Environment in SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 57, 2019, No. 12, pp. 10070–10087, doi: 10.1109/tgrs.2019.2931308.

- [21] MIKOLAJCZYK, K.—SCHMID, C.: A Performance Evaluation of Local Descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, 2005, No. 10, pp. 1615–1630, doi: 10.1109/tpami.2005.188.
- [22] CHICCO, D.—SADOWSKI, P.—BALDI, P.: Deep Autoencoder Neural Networks for Gene Ontology Annotation Predictions. *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB '14)*, 2014, pp. 533–540, doi: 10.1145/2649387.2649442.
- [23] LAROCHELLE, H.—MANDEL, M.—PASCANU, R.—BENGIO, Y.: Learning Algorithms for the Classification Restricted Boltzmann Machine. *The Journal of Machine Learning Research*, Vol. 13, 2012, No. 1, pp. 643–669.
- [24] GREGOR, K.—DANIHELKA, I.—GRAVES, A.—REZENDE, D.—WIERSTRA, D.: DRAW: A Recurrent Neural Network for Image Generation. *Proceedings of the 32nd International Conference on Machine Learning, PMLR*, Vol. 37, 2015, pp. 1462–1471.
- [25] SIMONYAN, K.—ZISSERMAN, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, CA, USA, 2015. arXiv:1409.1556v6, 2015.
- [26] YAMINS, D. L. K.—DICARLO J. J.: Using Goal-Driven Deep Learning Models to Understand Sensory Cortex. *Nature Neuroscience*, Vol. 19, 2016, No. 3, pp. 356–365, doi: 10.1038/nn.4244.
- [27] FUKUSHIMA, K.: Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, Vol. 36, 1980, pp. 193–202, doi: 10.1007/bf00344251.
- [28] RIESENHUBER, M.—POGGIO, T.: Hierarchical Models of Object Recognition in Cortex. *Nature Neuroscience*, Vol. 2, 1999, No. 11, pp. 1019–1025, doi: 10.1038/14819.
- [29] BECK, J. M.—MA, W. J.—KIANI, R.—HANKS, T.—CHURCHLAND, A. K.—ROITMAN, J.—SHADLEN, M. N.—LATHAM, P. E.—POUGET, A.: Probabilistic Population Codes for Bayesian Decision Making. *Neuron*, Vol. 60, 2008, No. 6, pp. 1142–1152, doi: 10.1016/j.neuron.2008.09.021.
- [30] LI, F. F.—FERGUS, R.—PERONA, P.: Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories. *Computer Vision and Image Understanding*, Vol. 106, 2007, No. 1, pp. 59–70, doi: 10.1016/j.cviu.2005.09.012.
- [31] GRIFFIN, G.—HOLUB, A.—PERONA, P.: Caltech-256 Object Category Dataset. California Institute of Technology, 2007.
- [32] CHANG, C. C.—LIN, C. J.: LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 2, 2011, No. 3, Art. No. 27, doi: 10.1145/1961189.1961199.
- [33] HE, K.—ZHANG, X.—REN, S.—SUN, J.: Deep Residual Learning for Image Recognition. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778, doi: 10.1109/cvpr.2016.90.
- [34] DENG, W.—HU, J.—GUO, J.: Face Recognition via Collaborative Representation: Its Discriminant Nature and Superposed Representation. *IEEE Transactions on Pat-*

- tern Analysis and Machine Intelligence, Vol. 40, 2018, No. 10, pp. 2513–2521, doi: 10.1109/tpami.2017.2757923.
- [35] WEN, J.—XU, Y.—LI, Z.—MA, Z.—XU, Y.: Inter-Class Sparsity Based Discriminative Least Square Regression. *Neural Networks*, Vol. 102, 2018, No. 10, pp. 36–47, doi: 10.1016/j.neunet.2018.02.002.
- [36] QIN, Y.—TIAN, C.: Weighted Feature Space Representation with Kernel for Image Classification. *Arabian Journal for Science and Engineering*, Vol. 43, 2018, No. 12, pp. 7113–7125, doi: 10.1007/s13369-017-2952-x.
- [37] SONG, Y.—LIU, Y.—GAO, Q.—GAO, X.—NIE, F.—CUI, R.: Euler Label Consistent K-SVD for Image Classification and Action Recognition. *Neurocomputing*, Vol. 310, 2018, pp. 277–286, doi: 10.1016/j.neucom.2018.05.036.
- [38] XU, J.—AN, W.—ZHANG, L.—ZHANG, D.: Sparse, Collaborative, or Nonnegative Representation: Which Helps Pattern Classification? *Pattern Recognition*, Vol. 88, 2019, pp. 679–688, doi: 10.1016/j.patcog.2018.12.023.
- [39] SEDERBERG, P. B.—NORMAN, K. A.: Learning and Memory: Computational Models. *Encyclopedia of Behavioral Neuroscience*, 2010, pp. 145–153, doi: 10.1016/b978-0-08-045396-5.00140-8.



Linxia XIAO received her Master's degree in communication and information system from Shandong University of Science and Technology in 2017. Currently, she is studying for her Ph.D. degree in control science and engineering in China University of Petroleum. Her research interests include pattern recognition, computer vision, and cognitive science.



Yanjiang WANG is currently Full Professor with the College of Control Science and Engineering, China University of Petroleum (East China), China. He received his M.Sc. degree in communication and electronic system from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1989 and the Ph.D. degree in signal and information processing from the Beijing Jiaotong University, Beijing, China, in 2001. He was Post-doctoral Researcher with the Institute of Drilling Engineering, Shengli Oilfield, Dongying, China, from 2003 to 2006. From 2013 to 2014, he was Visiting Scholar with the Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana, USA. His current research interests include bio-inspired pattern recognition, cognitive memory modeling and human brain connectivity. He has authored or co-authored more than 200 papers in top journals and prestigious conferences.



Baodi LIU received his Ph.D. degree in electronic engineering from the Tsinghua University, Beijing, China. He is currently Associate Professor with the College of Information and Control Engineering, China University of Petroleum, Qingdao, China. His research interests include computer vision and machine learning.



Weifeng LIU received his double B.Sc. degree in automation and business administration and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China in 2002 and 2007, respectively. His current research interests include pattern recognition and machine learning.