Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2021 Page No. 944-957
Article History: Received: 12 May 2021 Revised: 25 June 2021 Accepted: 22 July 2021 Publication: 31 August 2021

Deep Multimodal Neuroimaging Retrieval Based on Adaptive Hash Semantics

Wenyin Tao, Feng An*
School of Artificial Intelligence, Suzhou Industrial Park Institute of Service Outsourcing, Suzhou, Jiangsu 215123, China
*Corresponding Author.

Abstract:

Neuroimaging has been widely used in computer-assisted clinical diagnosis and treatment. In particular, multimodal neuroimaging retrieval technology, as an auxiliary technical means, can effectively improve the efficiency and accuracy of medical decision-making. However, the rapid increase of neuroimaging libraries has brought huge challenges to the rapid and efficient retrieval of neuroimaging. Existing image retrieval algorithms, on the other hand, frequently fail when applied directly to multimodal neuroimaging databases, because they typically use triplet loss functions to capture high-order semantic associations between samples. Triplet loss usually can only capture the local semantic similarity between neuroimaging samples. However, neuroimaging usually has complex semantic distribution, such as small inter-class differences and large inter-modal differences, which results in poor effects of the existing method. In order to solve these problems, this paper proposes a deep multimodal neuroimaging retrieval method based on adaptive hash semantics. Specifically, by directly learning the semantic space of neuroimage semantic tags, the hash network directly learns the Hamming semantic space distribution of each neuroimage from the semantic tags, thus avoiding the disadvantage of triplet loss. Meanwhile, the method directly uses category semantic tags for learning, which can achieve great learning effect. Widespread experimental results show that our method can generate effective hash codes and enable the most advanced multimodal neuroimaging retrieval performance.

Keywords*: Dultimodal neuroimaging, Hash learning, Medical image retrieval, Adaptive, Auxiliary diagnosis.*

I. INTRODUCTION

Neuroimaging analysis is of great help to modern clinical analysis and automatic diagnosis research[1]. Existing neuroimaging captures and records human medical data in digital image formats, such as magnetic resonance imaging (sMRI) and positron emission computed

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2021 Page No. 944-957
Article History: Received: 12 May 2021 Revised: 25 June 2021 Accepted: 22 July 2021 Publication: 31 August 2021

tomography (PET). At present, neuroimaging has been widely used in computer-aided clinical diagnosis and treatment[2]. However, interpreting neuroimaging data typically necessitates a great deal of hands-on expertise and professional knowledge, and differences among observers can influence the diagnosis and treatment outcome. Previous cases, that is, visually identical scans and accompanying treatment data, are presented to doctors in clinical practice to help clinical decision-making. This makes clinical case reasoning and evidence-based medicine decision-making easier. In particular, multimodal neuroimaging retrieval technology, as an auxiliary technical means, can effectively improve the efficiency and accuracy of medical decision-making[3-5]. In order to prove the assistance of neuroimaging retrieval to clinical diagnosis through examples, as introduced in the literature[5], a comparative experiment was conducted on newly recruited radiologists. As shown in Fig 1, there are two images related to nodules. Observe whether these doctors accurately determine benign and malignant nodules by referring to the neuroimaging retrieval system. Through experimental comparison, it is found that relying on the neuroimaging retrieval system can significantly improve the judgment accuracy. Content-based neuroimaging retrieval is a kind of instance-level image retrieval, which belongs to long-term research of neuromedical images. In the practice of observer research, the benefits of instance-level picture retrieval for medical image screening and diagnosis may be confirmed.

(A) Benign nodules on X-ray chest film

(b) Malignant nodules on X-ray chest film

Fig 1: samples of benign and malignant nodules

The neuroimaging retrieval system demands good scalability and efficient retrieval performance. Considering the reasonable balance between search effect and computing performance, hash learning methods have attracted more and more attention. Hash learning converts content-based neuroimaging search into hash code-based retrieval by encoding neuroimaging into a hash code in the Hamming space. Existing hash learning approaches are largely classified into two types according on whether supervised information is incorporated in

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2021 Page No. 944-957
Article History: Received: 12 May 2021 Revised: 25 June 2021 Accepted: 22 July 2021 Publication: 31 August 2021

the learning stage: unsupervised hash learning[6] and supervised hash learning[7]. Unsupervised hash learning method is usually used to learn the hash mapping function from the original feature space to the Hamming space by using topological information, original data structure, and data distribution. In contrast, supervised hash learning improves the learning quality of non-linear hash functions and improves the semantic discrimination of hash codes by using raw data and semantic label information[8]. To achieve a good balance between retrieval effect and processing cost, we focus on hash-based multimodal neuroimaging retrieval in this study.

However, due to the following reasons, when the existing methods are directly applied to multimodal neuroimaging retrieval, usually no good results are achieved. The main reason is that, on the one hand, compared with natural images, neuroimaging usually contains complex tissue textures and anatomical structures. Minor lesions in local areas of the brain can significantly affect the diagnosis result. The reliability is high, which means that neuroimaging may show small inter-class variations. Distinct neuroimaging technologies, on the other hand, can produce different visual representations for the same item (for example, a pair of sMRI and PET scans from the same object), resulting in significant inter-modal variances. The existing methods use triplet loss for hash learning. Triplet loss usually only captures the local semantic similarity between neuroimaging samples, which cannot well solve the semantic distribution problem of complex neuroimaging. Therefore, it is very necessary to develop an advanced hash learning technology to solve the problems of small interclass changes and large inter-modal differences in the process of neuroimaging image retrieval, thereby effectively increasing the neuroimaging retrieval effect.

In order to handle inter-class and intra-class differences of neuroimaging and effectively solve the problem of triplet loss, we propose a deep multimodal neuroimaging retrieval method based on adaptive hash semantic learning. This method uses convolutional neural networks to learn semantic information behind images. At the same time, a network structure is designed to perform semantic hash coding on the semantic tags of all neuroimages for learning by the hash network. Finally, based on Bayesian learning framework, semantic distribution of neuroimaging is learnt, so that the generated hash code can effectively distinguish neuroimaging of different types.

The rest of this paper is arranged as follows. The second part reviews related hash learning methods in detail, the third part details the deep multimodal neuroimaging retrieval method based on adaptive hash semantics, and the fourth part is the analysis of experimental results for detailed description of the experimental effects of our method in some neuroimaging data sets. The fifth part is the conclusion, which summarizes, analyzes and prospects the deep

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2021 Page No. 944-957
Article History: Received: 12 May 2021 Revised: 25 June 2021 Accepted: 22 July 2021 Publication: 31 August 2021

multimodal neuroimaging retrieval method.

II. RELATED WORK

Data-independent and data-related approaches are two types of traditional hash learning algorithms. The data-independent method learns the nonlinear hash function from hand-made features in two stages, with the hash code learning process and the feature learning process separated, which may result in the development of a sub-optimal hash function. Data-related methods sometimes become learning-based hashing methods, which can be further divided into (1) hashing methods based on shallow learning, such as hash forests based on metrics and hashing methods based on kernel functions; (2) hash methods based on deep learning, such as compact hash code learning based on image restoration, and deep hash network methods. Data-related approaches, as opposed to data-independent methods, extract global information for hashing in an end-to-end manner and discover the best hash function by integrating a hash network layer[9-12]. The relevance of neuroimaging retrieval is mainly based on the visual similarity of neuroimage rather than the entire image, so it is necessary to effectively explore regional instances. Recently, many existing image retrieval work usually extracts visual features by using convolutional neural networks (CNN), so that the unique visual features of image instances are not lost in the global image. Early methods mainly focus on replacing traditional manual feature descriptions with fully connected layer features. The existing methods have achieved significant progress mainly by encoding the activation of the convolutional layer as a regional feature descriptor. In the image retrieval based on convolutional neural network, it is necessary to fully consider the system retrieval effect and performance. This paper mainly studies data-related methods. Next, we respectively summarize some representative works of shallow learning and deep learning methods.

2.1 Shallow Hash Learning Method

The shallow hash learning method mainly uses hand-made features to learn linear or non-linear mapping functions, so that the neuroimages of each modal are mapped and converted into binary vectors. Representative methods in this category include cross-modal similarity-sensitive hashing (CMSSH), semantic correlation maximization hashing method (SCM), cross-media hashing (IMH), cross-view hashing (CVH), latent semantic sparseness hashing (LSSH), collective matrix factorization hashing (CMFH) and semantic preserving hashing (SePH). CMSSH is a supervised hash learning method that uses feature decomposition and promotion, and designs a cross-mode to preserve the similarity within the class. SCM[11] uses label information to learn the conversion information of specific modalities and retains the maximum semantic relevance between modalities. IMH[8] is an unsupervised hashing

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2021 Page No. 944-957
Article History: Received: 12 May 2021 Revised: 25 June 2021 Accepted: 22 July 2021 Publication: 31 August 2021

approach for achieving inter-modal and intra-modal consistency by encoding data. CVH[9] proposes an unsupervised cross-modal spectral hashing method so that the hash function preserves cross-modal similarity. In the public domain, LSSH employs sparse coding and matrix decomposition to generate a unified binary representation using the latent space learning method. CMFH learns a unified binary hash code by using a latent factor model for matrix factorization in the training phase. By building an affinity matrix in the probability distribution while reducing the KL divergence, SePH[10] generates a unified binary hash code.

2.2 Deep Hash Learning Method

Many deep cross-modal hashing algorithms have recently been proposed to improve the hash learning impact and ability, owing to deep neural networks' powerful arbitrarily nonlinear representation ability. To create binary code, Deep Visual Semantic Hashing (DVSH) learns a visual semantic fusion network with cosine hinge loss, and to generate a hash function, it learns a modal-specific deep network. However, DVSH can only be used in some special cross-modal scenarios, one of which must be temporal dynamics. Deep Cross Modal Hashing (DCMH) [13] is a deep learning system that leverages negative log-likelihood loss to construct cross-modal similarity-preserving hash codes. CAH (Auto-Encoding Correlation Hash) learns the hash function by optimizing the common features and semantic correlation between distinct modalities through auto-encoder architecture. Adversarial Cross-Modal Retrieval (ACMR) uses classification and adversarial learning methods to distinguish different modalities and generate binary hash codes. Self-Supervised Adversarial Hashing (SSAH) employs two adversarial networks to simultaneously model distinct modalities and capture their semantic importance, all while generating binary hash codes under the supervision of learnt semantic features. Cross-Modal Deep Variational Hashing (CMDVH) uses a two-step framework. The method learns the uniform hash code of cross-modal pairs in the database in the first step, and uses the learned uniform hash code to learn the hash function in the second step. As a result, for CMDVH, the hash function learned in the second stage cannot provide input to guide the unified hash code improvement.

The hash learning method has been studied by many scholars. However, when directly applied to multimodal neuroimaging retrieval, these methods usually cannot achieve good results mainly because neuroimaging has small interclass variation. At the same time, neuroimaging also has great inter-modal differences. The existing methods cannot well solve the semantic distribution problem of complex neuroimaging. Therefore, this paper proposes a deep multimodal neuroimaging retrieval method based on adaptive hash semantics.

III. DEEP MULTIMODAL IMAGE RETRIEVAL BASED ON ADAPTIVE HASH

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2021 Page No. 944-957
Article History: Received: 12 May 2021 Revised: 25 June 2021 Accepted: 22 July 2021 Publication: 31 August 2021

SEMANTICS

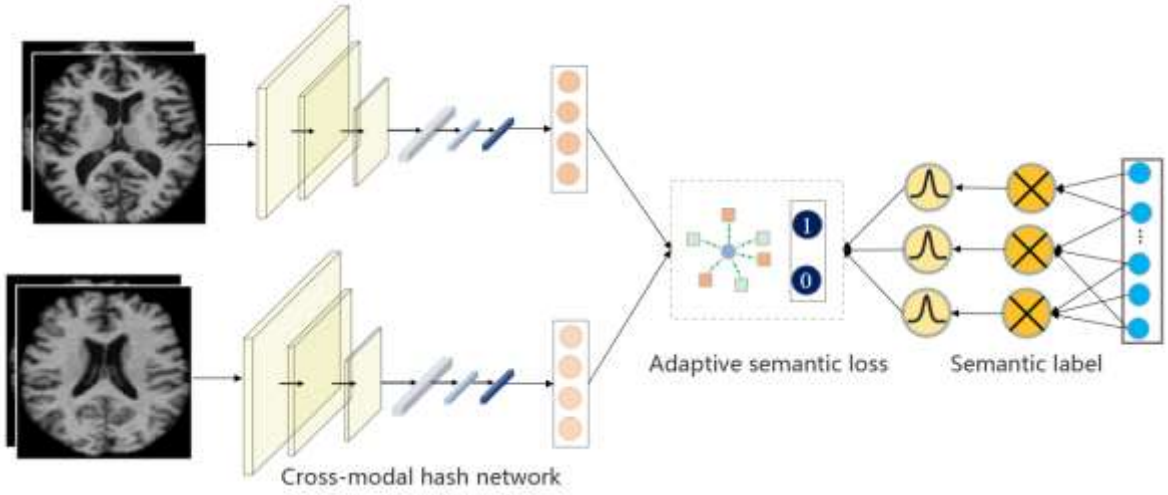
3.1 Problem Definition

Bold capital letters (for example, A) indicate the matrix, and A^T indicates the transpose of the matrix A. In addition, the sign function represents an element-by-element symbolic function, which is defined as:

$$sign(x) = \begin{cases} -1 & if\ x < 0, \\ 1 & if\ x >= 0 \end{cases} \quad (1)$$

Neuroimaging hashing is mainly to learn hash functions, so that neuroimaging images can be mapped to hash codes in Hamming space. Here, the hash code represents a binary vector, and the Hamming space contains a set of binary vectors. Suppose we have N given data image dataset $X = \{x_i\}_{i=1}^{N}$, and each image x_i is associated with a label vector l_i. We denote $B = \{b_i\}_{i=1}^{N}$ as the hash code of X, where $b_i \in \{-1, +1\}^k$ represents the binary hash code of sample x_i, k is the code length of the hash value, b_{ix} represents the xth element of b_i.

Furthermore, we must assume that all modalities of each database instance have the same hash code in order to successfully bridge the gap between them. Furthermore, the query data points' generated hash codes can maintain semantic similarity in order to learn the database instances' hash codes and hash functions. Calculate the Hamming distance between two separate hash codes, as well as the semantic relationship of the similarity matrix that goes with it, for any two hash codes.



Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2021 Page No. 944-957
Article History: Received: 12 May 2021 Revised: 25 June 2021 Accepted: 22 July 2021 Publication: 31 August 2021

Fig 2: a framework diagram of deep multimodal neuroimaging retrieval based on adaptive hash semantics

3.2 Overall Framework

The overall framework is shown in Fig 2. The overall network structure contains two modal network branches. We use a similar network structure. Inspired by AlexNet, these two branch deep convolutional neural networks consist of five convolutional layers and three fully connected layers. Each fully connected layer learns a nonlinear mapping $z_i = W_i * z_{i-1} + b_i$. Where, z_i is the output representation of the i-th hidden layer of the point x_i, W_i and b_i are the weight and bias parameter of the i-th layer, the activation function generally adopts the sigmod function. For the connection function of all hidden layers and fully connected layers, the ReLU function is generally selected. For the hash learning function, we replace the fully connected layer of the softmax classifier in the original AlexNet with a new fully connected layer of K hidden units. By using $h_i = z_i^l$, the corresponding hidden layer representation is converted into K-dimensional hash coding. Where, $l = 8$ is the total number of layers, which is the hidden representation of the fully connected layer. In order to popularize binary code as the representation of the hash layer, we first compress its output into [−1, 1] by using the hyperbolic tangent tanh activation function. In order to guarantee that the hash layer representation will be a good hash code, we must retain the similarity between a given pair of samples in the original sample similarity matrix, and control the quantization error of binarizing the hidden representation into a binary code. In addition to the specific network structure, the important innovation of this paper lies in the design of the objective function. In order to handle inter-class and intra-class differences of neuroimaging, and effectively solve the problem of triplet loss, we propose a deep multimodal neuroimaging retrieval method based on adaptive hash semantic learning. This method uses convolutional neural networks to learn semantic information behind images. At the same time, a network structure is designed to perform semantic hash coding on the semantic tags of all neuroimages for learning by the hash network. Finally, based on Bayesian learning framework, semantic distribution of neuroimaging is learnt, so that the generated hash code can effectively distinguish neuroimaging of different types.

3.3 Objective Function of Adaptive Hash Semantics

The objective function exerts an important effect on the generation of high-quality hash codes. In order to learn the semantic information behind images, our objective function includes Bayes-based semantic coding loss. At the same time, in order to preserve the similarity between the original ternary samples, our objective function also retains the original ternary

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2021 Page No. 944-957
Article History: Received: 12 May 2021 Revised: 25 June 2021 Accepted: 22 July 2021 Publication: 31 August 2021

loss function. However, in order to weaken its effect on the overall optimization process, we control it through hyperparameters so that it has a limited impact on the optimization of our overall objective function. Our overall objective function is as follows:

$$\min_{\Theta} \mathcal{L}_{\text{all}} = \mathrm{L}_{\text{class}} + \alpha T \tag{2}$$

Here, $\mathrm{L}_{\text{class}}$ represents semantic loss function based on the Bayesian framework. T represents the similarity loss of ternary samples. a is a hyperparameter used to control T loss.

3.3.1 Bayesian semantic loss function

$\mathrm{L}_{\text{class}}$ is derived from the Bayesian framework and is defined as follows: $\mathrm{L}_{\text{class}} = \log p\left(B^x, B^y | S\right) \propto log\, p(S|B^x, B^y)\, p(B^x)\, p(B^y)$.

Where, $p(B^x)$ and $p(B^y)$ are the prior distributions of the corresponding modal hash codes, and $p(S|B^x, B^y)$ is an adaptive loss function based on semantics. $p(S|B^x, B^y)$ is defined as follows:

$$p(S|B^x, B^y) = \sum_{s_{ij}\, in\, S} w_{ij} log p\left(s_{ij} \middle| b_i^x, b_j^y\right) \tag{3}$$

w_{ij} represents the weight corresponding to each sample. $p\left(s_{ij} \middle| b_i^x, b_j^y\right)$ is used to describe the influence of two modal hash codes on s_{ij}. When s_{ij} is 1, the corresponding definition is as follows:

$$p\left(s_{ij} \middle| b_i^x, b_j^y\right) = \sigma\left(I\left(b_i^x, b_j^y\right)\right) \tag{4}$$

σ is the activation function of adaptive learning, which is defined as follows:

$$\sigma = 1/\left(1 + e^{(-ax)}\right) \tag{5}$$

Similarly, when s_{ij}=0, $p\left(s_{ij} \middle| b_i^x, b_j^y\right)$ is defined as follows:

$$p\left(s_{ij} \middle| b_i^x, b_j^y\right) = 1 - \sigma\left(I\left(b_i^x, b_j^y\right)\right) \tag{6}$$

In the Bayesian framework, we adaptively control the hash semantic similarity between different modalities by using the similarity of paired samples. At the same time, each modal's image is represented by a hash code, which is an instance-invariant feature vector. In the

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2021 Page No. 944-957
Article History: Received: 12 May 2021 Revised: 25 June 2021 Accepted: 22 July 2021 Publication: 31 August 2021

training phase, we input images to different branch networks and feed them forward to the hash layer. The hash layer is the last layer of each modal branch network. The intra-class distance can be optimized and lowered in the final layer using a loss function based on adaptive hashing semantics, while the inter-class distance can be maximized, avoiding the problem of easy confusion between difficult-to-distinguish samples.

In order to allow branches of different modalities to learn the class-perceptual semantic information of pathological regions, it is used to distinguish the same manifestations of different diseases. Simultaneously, the spatial nuances of problematic regions from several branches are recorded in the hash layer in order to locate the nuances of the same disease at various phases. After the hash layer absorbs visual cues from many modalities in the training phase, feature aggregation in the testing phase allows us to produce hash codes from the learnt core nodes. The final hash code is generally implemented through the sgn(·) function. However, the sgn(·) function is not differentiable at zero, and for non-zero input, its derivation will be zero. This means that when the formula L_{class} is minimized, the parameters of the modal hash network unique to modality will not be updated using the back propagation algorithm. Therefore, we directly discard the sgn(·) function to ensure that the parameters of our hash model can be updated, and add a quantization loss so that each element of different modalities can be close to "+1" or "-1". Furthermore, the query set is sampled from the database during the training phase. As a result, the hash code created by the learnt hash function should be the same as the directly learned hash code for generating the quantitative loss. That is, if a query instance is sampled from the database, the hash code and the learnt hash code should be as similar as possible.

3.3.2 Ternary linguistic loss function

In addition to considering Bayesian semantic loss, we also need retain the ternary semantic loss of different modal samples, so that the subtle differences between different neuroimages can be learned. In order to construct the T loss function, we use the triple loss, which is defined as follows:

$$T = L_{metric} + L_{push} + L_{balancing} \tag{7}$$

L_{metric} is used to measure the similarity of different samples. L_{push} is to push each hash value to a discrete space. $L_{balancing}$ is used to balance each hash code. T loss is based on the visual semantic similarity intuition. In the learned metric space, photos with the same label should be closer to each other than images with different labels. In more detail, a sample of triples is randomly sampled from the data (anchor point, positive sample, negative sample).

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2021 Page No. 944-957
Article History: Received: 12 May 2021 Revised: 25 June 2021 Accepted: 22 July 2021 Publication: 31 August 2021

Where, the positive sample is the sample closer to the anchor point, and the negative sample is the sample not in the same category as the anchor point. In order to use the stochastic gradient descent method to train this function, we set a specific metric loss function, which is defined as:

$$L_{metric} = \sum_{i=1}^{M} \max(0, |f(a) - f(p)|_2^2 - |f(a) - f(n)|_2^2 + c) \quad (8)$$

Where, f(a), f(p) and f(n) represent anchor point sample, positive sample, and negative sample, respectively. c represents a parameter set empirically, which is used to control the distance between the positive, negative samples and the anchor point. Sometimes called the minimum margin threshold, it is mandatorily between a positive distance and a negative distance. Function f is a function that represents the various modes mentioned above. The output layer is the hash layer. In particular, we use the LeakyReLU function in the two hidden layers to allow negative gradients to flow during the back propagation, and use sigmoid activation in the last layer to limit the output activation in [0, 1]. In order to push the final real activation to the end of the sigmoid function range, we design the second loss to maximize the sum of squared errors between the output layer activation and the value 0.5, which is defined as follows:

$$L_{push} = \frac{1}{K} \sum_{i=1}^{M} |f(x) - 0.5|^2 \quad (9)$$

Where, M represents the number of batch samples. In addition, each neuron is encouraged to output the 01 hash code with a 50% probability. This means that the binary code representation of the image will have balanced 0s and 1s, so all bits of the hash code are used equally. Therefore, the balance loss is defined as follows:

$$L_{balancing} = \sum_{i=1}^{M} \big(mean\big(f(x)\big) - 0.5\big)^2 \quad (10)$$

Where, mean(f(x)) is the average value of output activation. After calculating the final loss of each mode, our final binary code is output through the sgn function. At the same time, in the retrieval process, the Hamming distance between the query image and each image in the data file is calculated, and the obtained distances are sorted in ascending order of magnitude.

IV. EXPERIMENTAL ANALYSIS

4.1 Data Set

We evaluate our method on a popular benchmark dataset: the Alzheimer's disease

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2021 Page No. 944-957
Article History: Received: 12 May 2021 Revised: 25 June 2021 Accepted: 22 July 2021 Publication: 31 August 2021

neuroimaging dataset ADNI1. Next, we introduce this data set in more detail.

ADNI1 contains 821 subject-weighted sMRI scans, of which only 397 subjects had PET images. Each subject is annotated with a category-level label, namely Alzheimer's Disease (AD), Normal Control (NC), or Mild Cognitive Impairment (MCI). These labels are determined based on standard clinical criteria, including simple mental status test scores and clinical dementia scores. Among the subjects who underwent sMRI scans in ADNI1, there were 229 NC, 393 MCI, and 199 AD subjects. For the PET data in ADNI1, there were 100 NC, 93 AD, and 204 MCI subjects.

We randomly select 10% of the photographs from each class to establish a test set, and the other images serve as the retrieval set in this data set. To improve the suggested technique, we choose 90 percent of the images from the retrieval set at random as the training set, and the remaining images as the validation set. We use standard pipelines to pre-process all sMRI and PET scans, including anterior commissure (AC)-posterior commissure (PC) correction, intensity correction, skull dissection, and cerebellar removal. Each PET image is aligned with its associated sMRI scan using affine registration.

4.2 Benchmark Method

The method proposed in this paper is first compared with three state-of-the-art cross-modal hashing methods based on traditional machine learning techniques, including CVH [9], SEPH [10] and SCM [11]. Then, it is compared with the three recently proposed modal hashing deep learning methods, including PGDH [12], DCMH [13] and CMHH [8]. Three traditional methods (i.e., CVH, SCM, and SEPH) are implemented using MATLAB, and four deep learning methods (i.e., DCMH, PGDH, CMHH, and our method) are implemented using Pytorch. All of the methods are overseen. We extract the gray matter volume from 90 ROIs as features representing sMR and PET images and use these ROI features as input for existing methods. For the deep learning method, we use the original image as input. In our method, for a specific modal hash network, all parameters are initialized randomly. We initially pre-train each modal specific hash network using a simplified optimized single modal version of the network to speed up model training. DCMH and PGDH use the same backbone network architecture and pre-training procedure as our method for a fair comparison. The experiment's parameters are all empirically determined. Adam is used to optimize the network parameters of four deep learning methods (i.e., DCMH, PGDH, CMHH and our method). Where, the initial learning rate is set to 0.01.

4.3 Experimental Analysis

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2021 Page No. 944-957
Article History: Received: 12 May 2021 Revised: 25 June 2021 Accepted: 22 July 2021 Publication: 31 August 2021

To evaluate all methodologies, this paper uses two evaluation indicators. These indicators, which include average precision average (MAP) and precision-recall, are based on the Hamming space ranking, which rates the returned data points according to the Hamming distance between them and the query. Because it initially generates a hash search table and returns data points within the given Hamming radius, the precise recall measure is based on hash search. The average query accuracy (MAP) is defined as the relationship between the query image and the database image for a particular query and a list of database sorted retrieval examples. Precision-recall indicates retrieval accuracy at various recall levels, which is a useful predictor of overall search performance.

Table I shows the MAP results obtained by all methods on the ADNI1 data set. Here, "M→P" means that the query is an sMRI scan and the database contains PET images. "P→M" means that the query is PET image and the database has sMRI scans. From the results in table, we can observe an interesting finding. In other words, the MAP value of "M→P" is usually higher than the MAP value of "P→M". This may be due to the limited number of PET images. At the same time, since deep models usually require a large amount of training data to reduce overfitting, limited PET scans may lower the generalization ability of the hash function against PET modalities. Therefore, the hash code generated by MRI query may be superior to the hash code of PET query, which may cause a performance gap between the "M→P" and "P→M" tasks. Finally, from the results, we can also conclude that our method usually outperforms other benchmark methods to a large extent.

Fig 3 shows the 16-bit, 32-bit, and 64-bit Precision-Recall curves (PR curves) on the ADNI1 data set. From these figures, we can observe that the Precision-Recall curve area of our method is larger than that of all baseline methods. It indicates that our method can effectively learn the hash semantic information behind different modal neuroimages, and has good algorithm retrieval performance.

TABLE I. Experimental results of all methods

Task	Method	ADNI1			
		16bit	32bit	64bit	128bit
M→P	CVH	0.3933	0.3842	0.4163	0.3752
	SEPH	0.4378	0.4416	0.4063	0.4320
	SCM	0.5037	0.4713	0.5109	0.5174
	CMHH	0.4738	0.4323	0.4502	0.4144
	PGDH	0.5124	0.5181	0.5276	0.5225
	DCMH	0.5562	0.5529	0.5366	0.5306

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2021 Page No. 944-957
Article History: Received: 12 May 2021 Revised: 25 June 2021 Accepted: 22 July 2021 Publication: 31 August 2021

	Ours	0.5625	0.5533	0.5431	0.5437
P→M	CVH	0.3913	0.3992	0.4079	0.4138
	SEPH	0.4585	0.3962	0.4587	0.3688
	SCM	0.4806	0.4658	0.4599	0.4659
	CMHH	0.4017	0.4098	0.4072	0.4008
	PGDH	0.5097	0.4565	0.4422	0.4592
	DCMH	0.5248	0.4601	0.4518	0.4468
	Ours	0.5312	0.4837	0.4989	0.4826

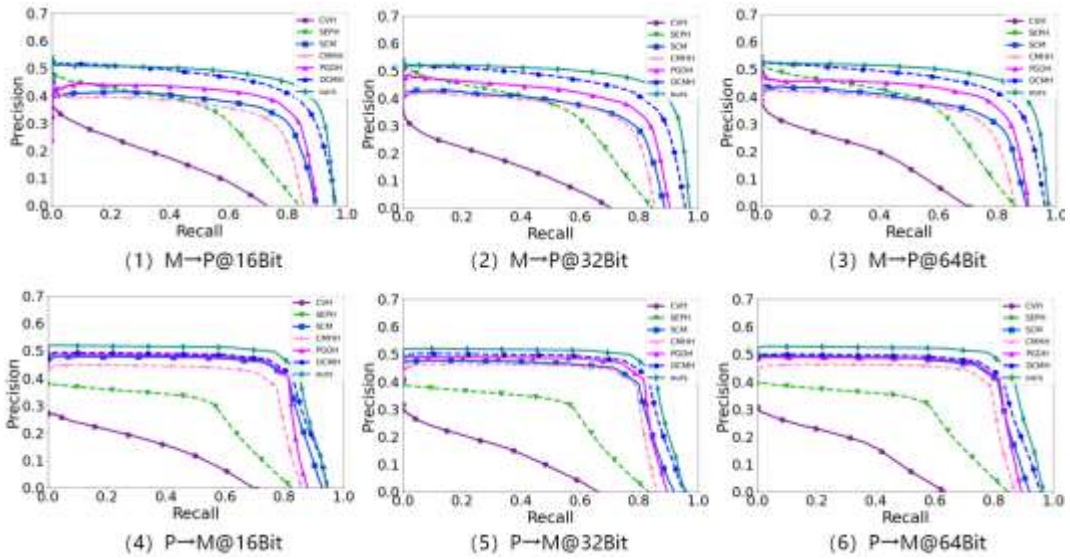


Fig 3: Precision-Recall curves of all methods

V. CONCLUSION

This paper introduces an adaptive hash semantic algorithm for multimodal neuroimage retrieval. Specifically, a deep multimodal neuroimaging retrieval method is proposed based on adaptive hash semantic learning. This method uses convolutional neural networks to learn semantic information behind images. At the same time, a network structure is designed to perform semantic hash coding on the semantic tags of all neuroimages for learning by the hash network. Finally, based on Bayesian learning framework, semantic distribution of neuroimaging is learnt, so that the generated hash code can effectively distinguish neuroimaging of different types. Comprehensive experiments show that our method has the most advanced cross-modal retrieval performance on multimodal neuroimage datasets.

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2021 Page No. 944-957
Article History: Received: 12 May 2021 Revised: 25 June 2021 Accepted: 22 July 2021 Publication: 31 August 2021

ACKNOWLEDGEMENT

1. Jiangsu Qing Lan Project ([2020] 10).

2. 2020 Jiangsu Provincial Education Science "13th Five-Year Plan" (D/2020/03/50).

3. This research was funded by Special Equipment Safety Supervision Inspection Institute of Jiangsu Province, grant number FMZ202018.

4. The Natural Science Foundation of the Jiangsu Higher Education Institutions of China (21KJB520037).

REFERENCES

[1] Tajbakhsh N, et al. (2016) Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Trans. Med. Imag. 35(5): 1299–1312

[2] Cheng CH and Liu WX (2018) Identifying degenerative brain disease using rough set classifier based on wavelet packet method. Journal of Clinical Medicine 7(6): 124

[3] Cao Y, et al. (2014) Medical image retrieval: a multimodal approach. Cancer Informatics 13: CIN–S14 053

[4] Vikram M, Anantharaman A, and BS S (2019) An approach for multimodal medical image retrieval using latent dirichlet allocation. in COMAD 44–51

[5] Fang J, Fu H, Zeng D, et al. (2021) Combating Ambiguity for Hash-code Learning in Medical Instance Retrieval. IEEE Journal of Biomedical and Health Informatics.

[6] Song J, Yang Y, Yang Y, Huang Z, and Shen H (2013) Inter-media hashing for large-scale retrieval from heterogeneous data sources. In SIGMOD 785–796

[7] Kumar S and Udupa R (2011) Learning hash functions for cross-view similarity search. In IJCAI

[8] Cao Y, Liu B, Long M, and Wang J (2018) Cross-modal hamming hashing. In ECCV 202–218

[9] Kumar S and Udupa R (2020) Learning hash functions for cross-view similarity search. In IJCAI

[10] Lin Z, Ding G, Hu M, and Wang J (2019) Semantics-preserving hashing for cross-view retrieval. In CVPR 3864–3872

[11] Zhang D and Li WJ (2017) Large-scale supervised multimodal hashing with semantic correlation maximization. In AAAI

[12] Yang E, Deng C, Liu W, Liu X, Tao D, and Gao X (2017) Pairwise relationship guided deep hashing for cross-modal retrieval. In AAAI, 1618–1625

[13] Jiang Q and Li W (2017) Deep cross-modal hashing. In CVPR, 3232–3240