Check for updates

Sparse Representation-Based Discriminative Metric Learning for Brain MRI Image Retrieval

Guohua Zhou[1,2,3], Bing Lu[1], Xuelong Hu[3] and Tongguang Ni[2]*

[1] School of Information Engineering, Changzhou Institute of Industry Technology, Changzhou, China, [2] School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China, [3] College of Information Engineering, Yangzhou University, Yangzhou, China

Magnetic resonance imaging (MRI) can have a good diagnostic function for important organs and parts of the body. MRI technology has become a common and important disease detection technology. At the same time, medical imaging data is increasing at an explosive rate. Retrieving similar medical images from a huge database is of great significance to doctors' auxiliary diagnosis and treatment. In this paper, combining the advantages of sparse representation and metric learning, a sparse representation-based discriminative metric learning (SRDML) approach is proposed for medical image retrieval of brain MRI. The SRDML approach uses a sparse representation framework to learn robust feature representation of brain MRI, and uses metric learning to project new features into the metric space with matching discrimination. In such a metric space, the optimal similarity measure is obtained by using the local constraints of atoms and the pairwise constraints of coding coefficients, so that the distance between similar images is less than the given threshold, and the distance between dissimilar images is greater than another given threshold. The experiments are designed and tested on the brain MRI dataset created by Chang. Experimental results show that the SRDML approach can obtain satisfactory retrieval performance and achieve accurate brain MRI image retrieval.

Keywords: medical image retrieval, magnetic resonance imaging, brain images, sparse representation, metric learning

INTRODUCTION

The number of tumor patients has increased rapidly in recent years. Tumors have become one of the most common diseases in the world. The prevention, early diagnosis, and treatment of tumors have aroused widespread concern and discussion in the medical community and the public. Pathological examination is still the gold standard for tumor detection in the medical field. Medical imaging is a common method for diagnosing tumors, such as X-ray, magnetic resonance imaging (MRI), and computed tomography (CT) scans (Zhang et al., 2019; Chahal et al., 2020; Jiang et al., 2021). At present, early cancer screening mainly depends on the experience of doctors. The accuracy of decision-making largely depends on the knowledge and experience of doctors and the quality and

quantity of information available. At the same time, the development of digital imaging technology and artificial intelligence has jointly promoted the emergence of digital pathology. Digital pathology aims to use computers to automatically analyze the characteristics of digital histopathology images to serve different tasks such as detection, segmentation, and retrieval. Digital pathology can provide physicians with knowledge-supported assistive technology in their decision-making process, which is of great practical significance.

With the popularization of Internet technology, the scale of medical imaging databases is getting larger and larger, and more and more past cases can be provided to doctors. Content-based image retrieval (CBIR) has technical advantages in medical diagnosis (Müller and Unay, 2017; Zheng et al., 2018). CBIR not only provides doctors with a prediction, but also improves the efficiency and accuracy of diagnosis by retrieving similar historical cases. Therefore, CBIR is essentially consistent with the physician's diagnostic decision-making process. In addition, the use of CBIR can develop knowledge-based medical diagnosis decision support systems and case interpretation systems. Medical image retrieval is the core technology of the CBIR system, which can retrieve medical images similar to the image to be diagnosed from the medical database. We know that medical images are very different from natural images. First, the resolution of medical images is relatively high, but most of them are grayscale images. Secondly, the important information of medical images is mostly concentrated in small areas. Thirdly, the semantic content of visually similar medical images may be very different. Therefore, the retrieval effect is often unsatisfactory when the conventional natural image processing method is directly applied to medical imaging. Especially for brain tumor images, the shape, size, and texture of the same type of brain tumor are different due to the severity of the patient's condition, age, and other factors. Therefore, different pathological types may show similar appearances, which make the recognition of brain tumor images very challenging.

Medical image retrieval is to retrieve images similar to the histopathological images to be diagnosed from the database, to provide references for doctors' diagnoses. The histopathological image retrieval algorithm generally includes two stages: feature extraction and similarity measurement model construction (Zhang et al., 2015). The tumor boundary in brain tumor images is usually unclear, so the shape information cannot provide a clear tumor contour area. Therefore, it is important to determine the intensity and texture features. Huang et al. (2012) combined the spatial pyramid model and the Bow model and used the bag-of-words (BoW) histogram with spatial information to describe the characteristics of brain tumor regions, which improved the retrieval performance of brain tumor images. Huang et al. (2014) regarded the boundary of the brain tumor as the region of interest (ROI), used the region division learning method, and expressed the features with the original image pixel intensity. Finally, the authors aggregated the local features of each subregion and spatialized them, to improve the discriminative ability of image features. Cheng et al. (2016) adopted adaptive space pool and fisher vector method for brain tumor image retrieval.

In the field of medical image retrieval, similarity measurement refers to using a given medical image pair to calculate the distance between the feature vectors of medical image pairs, and it is often used to judge the type of medical image. For medical images, similarity measures include semantic correlation and visual similarity. The semantic correlation measure depends on the class labels of medical images. For example, the labels of two brain MRI images are healthy, which means that the two brain MRI images are semantically related. The visual similarity measure describes the similarity of features, which usually describes the similarity of medical images from a visual perspective. The definition of similarity measures generally considers the use of distance measures. The commonly used distance measurement methods include Euclidean distance, angle cosine distance, Mahalanobis distance, and Minkowski distance, etc., Yang et al. (2012) extracted local features based on pixel intensity information along the brain tumor boundary and used BoW to generate global features. At the same time, the authors designed a distance metric learning method to improve retrieval accuracy. Verma and Jawahar (2017) proposed a two-pass k-nearest neighbor model. Firstly, a more balanced training data set was designed for each image to reduce the label imbalance problem. Then, the large interval nearest neighbor algorithm was extended to the case of multi-label classification for metric learning, and the optimal weights of combined basis distance and features were obtained for semantic label prediction. Gu et al. (2017) proposed a distance measurement method of weighted heterogeneous values for tumor diagnosis.

Inspired by the sparse mechanism of the human visual system, sparse representation has been widely used in the field of computer vision. Sparse representation succinctly represents the image to be labeled as a linear combination of a few atoms in the dictionary. These sparse representation coefficients and corresponding atoms fully reveal the internal essence of the image and also adapted to human visual perception. Dictionary learning is a common method in sparse representation research. The dictionary learned from a large number of samples can adaptively sparse represent various feature information in the image, which greatly improves the sparse representation performance. Sparse representation methods can be divided into two categories. The first is methods based on predefined dictionaries. Srinivas et al. (2014) used the correlation between the RGB three-channels of color images to construct a dictionary and applied it to histopathological image representation and classification. However, this method based on the predefined dictionary may not be able to make full use of the discriminative information hidden in the training samples. The second type of method learns the dictionaries from training samples. For example, Gao et al. (2014) regarded the samples of each subclass as a subgroup and used the multi-layer group sparsity to classify and annotate images. Tang et al. proposed a semi-supervised learning method based on a sparse graph for image classification. By using the multi-scale and spatial information of the image, Tang et al. (2010) proposed a multi-scale representation learning algorithm for breast image classification. Zhang et al. (2015) proposed a heterogeneous feature fusion method and applied it to the image classification of pathological tissues.

The above methods mainly focus on the reconstruction of histopathological images, so their classification performances are not optimal. The dictionary learning method based on supervised learning uses robust feature representation from the original visual information through mining supervised or semi-supervised features and coding coefficients. Cao et al. (2015) used the group sparse reconstruction method to learn the semantic correlation of symbiotic labels for optimizing the dictionary. The authors used the group sparse framework to reconstruct the test image for label prediction and realized weakly supervised image annotation. Lu et al. (2015) studied the semantic sparse recording method of visual content to generate a more descriptive and robust bow representation for image classification.

This paper proposes sparse representation-based discriminative metric learning (SRDML) approach for brain MRI image retrieval. SRDML integrates sparse representation and metric learning into a discriminative model. On the one hand, the model uses the sparse representation to learn the robust feature representation of brain MRI images. On the other hand, the model learns a metric space with discriminative ability, so that similar brain MRI images are closely projected in the metric space, and dissimilar brain MRI images are separated from each other as much as possible. The advantages of SRDML are as follows: (1) the local information retention term of coding coefficients maintains the semantic correlation and visual similarity of brain MRI images in the projection space. (2) Our approach finds an appropriate metric matrix under the constraints of similarity, so that the distance between similar images' coding coefficients is less than the given threshold, and the distance between dissimilar images' coding coefficients is greater than the given threshold. Its goal is to achieve the dispersion of "maximum inter-class" and "minimum intra-class". (3) The learning of sparse representation and metric learning is optimized through alternate iterations. This learning strategy allows the model parameters to reach the optimal solution at the same time. (4) The experimental results show that SRDML has achieved good performance on the brain MRI image dataset, which shows the feasibility of our retrieval approach in the brain tumors diagnosis.

The rest of this paper is organized as follows: first, related works of metric learning and sparse representation are introduced. Second, the proposed sparse representation-based discriminative metric learning approach is formulated in section "Sparse Representation-Based Discriminative Metric Learning". Third, experiments are presented to verify our approach. Finally, a summary of our approach is presented, and its future work is discussed.

BACKGROUNDS

Metric Learning

The traditional metric learning is based on labeled samples of medical image data sets and the label information appears in the form of paired constraints of samples. Assuming that the medical image data set is expressed as $\{x_i\}_{i=1}^N \in R^d$, where x_i

is the ith sample, taking Mahalanobis distance as an example, the distance measurement between two images x_i and x_j can be written as,

$$d_M^2(x_i, x_j) = \|x_i - x_j\|_M^2 = (x_i - x_j)^T M(x_i - x_j). \tag{1}$$

The positive semi-definite matrix $\mathbf{M}$ can be decomposed into $M = W^T W$, and the matrix $W \in R^{d \times m}(m \leq d)$ is called the metric matrix. Therefore, Eq. 1 can be expressed as,

$$d_M^2(x_i, x_j) = (Wx_i - Wx_j)^T(Wx_i - Wx_j). \tag{2}$$

Therefore, the essence of the Mahalanobis metric is to learn a projection space, in which that similar image pairs output a positive value close to zero, and dissimilar image pairs output a larger value.

A type of famous metric learning algorithm is large-margin distance metric learning (Weinberger and Saul, 2009; Nguyen et al., 2020). Let $\mathbf{S}$ and $\mathbf{D}$ be two datasets of pairwise constraints: $\mathbf{S} = \{(x_i, x_j) \mid x_i \text{ and } x_j \text{ are similar}\}$, and $\mathbf{D} = \{(x_i, x_j) \mid x_i \text{ and } x_j \text{ are dissimilar}\}$. The large-margin distance metric learning takes the paired similar or dissimilar data sets as model input and constructs the metric learning as a convex optimization problem,

$$\begin{aligned} \min_M \quad & \sum_{(x_i - x_j) \in S} \|x_i - x_j\|_M^2, \\ s.t. \quad & \sum_{(x_i - x_j) \in D} \|x_i - x_j\|_M^2 \geq c, \\ & M \geq 0, \end{aligned} \tag{3}$$

where c is a positive. It can be seen from Eq. 3 that the distance between dissimilar pair samples is greater than the constant c, while the distance between similar pair samples is as small as possible.

Sparse Representation

The goal of sparse representation is to use the linear combination of dictionary atoms to represent the observation data, and to use sparsity constraints on the combination coefficients, so that each observation data is only represented by a subset of all dictionary atoms (Gu et al., 2021a,b). The traditional sparse representation framework is,

$$\begin{aligned} \min_{B, Y} \quad & \|X\text{-}BY\|_F^2 + \lambda\phi(Y), \\ s.t. \quad & \|b_i\|_2 \leq 1, \ \forall i, \end{aligned} \tag{4}$$

where $\mathbf{B}$ and $\mathbf{Y}$ are the dictionary matrix and sparse coding coefficient matrix, respectively. $\|\|_F^2$ is F-norm.

In order to avoid that the value of $\mathbf{B}$ is too large and the sparse coefficient $\mathbf{Y}$ is too small, the dictionary column vector is usually set to satisfy the constraint $\|b_i\|_2 \leq 1$. $\phi(Y)$ is the regularization term controlling the sparsity of $\mathbf{Y}$. λ is the regularization parameter. The objective function is not a joint convex function of $(\mathbf{B}, \mathbf{Y})$, but when one variable is fixed, the objective function is convex for the other variable. Therefore, Eq. 4 can be solved by a turn-by-turn optimization method.

SPARSE REPRESENTATION-BASED DISCRIMINATIVE METRIC LEARNING

In order to better explain the proposed model, first, we introduce some important terms: (1) a labeled training image set $X = [x_1, x_2, ..., x_n] \in R^{d \times n}$; (2) the testing image set contains images available for the retrieval of the target brain MRI images; (3) each image in testing image set is considered as the query image.

The Pairwise Constraint of Sparse Coding Coefficients

From the perspective of image retrieval, similar images should have similar coding coefficients, and dissimilar images should have different coding coefficients. In other words, in the learned metric space, the smaller the better the distance between similar images, at the same time, the greater the better the distance between dissimilar images, to reduce the uncertainty of misjudging the coding coefficient of dissimilar images as similar images. Based on this idea, two thresholds σ_1 and σ_2 are set for calculating the metric distances of similar images and dissimilar images in the metric space as,

$$\begin{cases} d_M^2(y_i - y_j) < \sigma_1, \ if \ \{x_i, x_j\} \in S \\ d_M^2(y_i - y_j) > \sigma_2, if \ \{x_i, x_j\} \in D \end{cases} \tag{5}$$

where $\sigma_1 < \sigma_2$.

To unify the Eq. 5 into an inequality, a common threshold is introduced, i.e., $\sigma_1 = \tau - 1$ and $\sigma_2 = \tau + 1$, then Eq. 5 can be written as,

$$\ell_{ij}(\tau - d_M^2(y_i - y_j)) > 1, \tag{6}$$

where if x_i and x_j are similar images, $\ell_{ij} = 1$; otherwise, x_i and x_j are dissimilar images, $\ell_{ij} = -1$.

With the pairwise constraint of coding coefficients, it is expected to have a large margin between each similar and dissimilar pairs in the learned metric space. By introducing the generalized logistic loss function $g(y) = \frac{1}{\theta} \log(1 + \exp(\theta y))$ with the sharpness parameter θ, the pairwise constraint of coding coefficients can be defined as,

$$\min_Y \ g\left(1 - \ell_{ij}(\tau - d_M^2(y_i - y_j))\right). \tag{7}$$

The Locality Constraint of Atoms

First, to exploit the local structure information in the sparse representation learning, the local manifold of the dictionary atoms is built on the nearest neighbor graph $\mathbf{P}$ of dictionary $\mathbf{B} \in R^{d \times K}$. The element P_{ij} of $\mathbf{P}$ is defined as,

$$P_{ij} = \begin{cases} \exp\left(-\frac{\|b_i - b_j\|_2^2}{\mu}\right), & if \ b_j \in \text{KNN}(b_i) \\ 0, & else \end{cases} \tag{8}$$

where KNN( ) is the k-nearest neighbor function, and μ is the KNN parameter.

In the learned metric space, we assume that the similar images will keep close to each other, thus it is expected the similar

dictionary atoms will keep close to each other as well. A Laplacian matrix $\mathbf{L}$ is constructed based on the matrix $\mathbf{P}$,

$$L = W - P, \tag{9}$$

where $W = diag(W_1, ..., W_K)$, and $W_i = \sum_j^K P_{ij}$. K is the number of atoms in $\mathbf{B}$.

Using the Laplacian matrix $\mathbf{L}$, the locality constraint of atoms is represented as,

$$\min_Y \ \frac{1}{2} \sum_i^K \sum_j^K (y_i - y_j)^2 P_{ij} = Tr(Y^T L Y). \tag{10}$$

The Objective Function of SRDML and Its Optimization

Now we have the two terms needed to be embedded into the sparse representation framework and compose the SRDML objective function. SRDML combines the sparse representation, pairwise constraint of coding coefficients, and locality constraint of atoms together to yield an image retrieval approach. The SRDML approach takes advantage of pairwise metric constraints and local structure knowledge preserved provided by the coding coefficients. The objective function of SRDML is,

$$\min_{M,B,Y} \ \|\text{X-BY}\|_2^2 + \alpha \, Tr(YLY) + \beta \sum_{i,j} g\left(1 - \ell_{ij}\left(\tau - d_M^2(y_i - y_j)\right)\right) + \gamma \|Y\|_2^2,$$

$$s.t. \quad M \geq 0,$$

$$\|b_i\|^2 = 1, \ \forall i \tag{11}$$

where α, β, and γ are regularization parameters.

There are three variables $\{M, B, Y\}$ needed to be tuned in the SRDML approach. An alternating optimization method is used to solve Eq. 11.

Tune $\mathbf{B}$. With fixed $\mathbf{M}$ and $\mathbf{Y}$, the objective function of $\mathbf{B}$ can be written as,

$$\min_B \ \|\text{X-BY}\|_2^2,$$

$$s.t. \quad \|b_i\|^2 = 1, \ \forall i, \tag{12}$$

Obviously, it is a least square problem with quadratic constrains. The Lagrange dual function is used, and we have,

$$h(B, \tau) = \left(\|\text{X-BY}\|_2^2 + \sum_i^K \tau_i(\|b_i\|^2 - 1)\right), \tag{13}$$

where τ is the Lagrange vector. A diagonal matrix $\Delta \in R^{K \times K}$ is introduced, where $\Delta_{ii} = \tau_i$ for all i. We can obtain the following problem,

$$h(B, \tau) = \left(\|\text{X-BY}\|_2^2 + Tr(B^T B \Delta) - Tr(\Delta)\right). \tag{14}$$

Taking the first-order partial derivatives of $\mathbf{B}$ in Eq. 14, we can obtain the close-solution of $\mathbf{B}$ as,
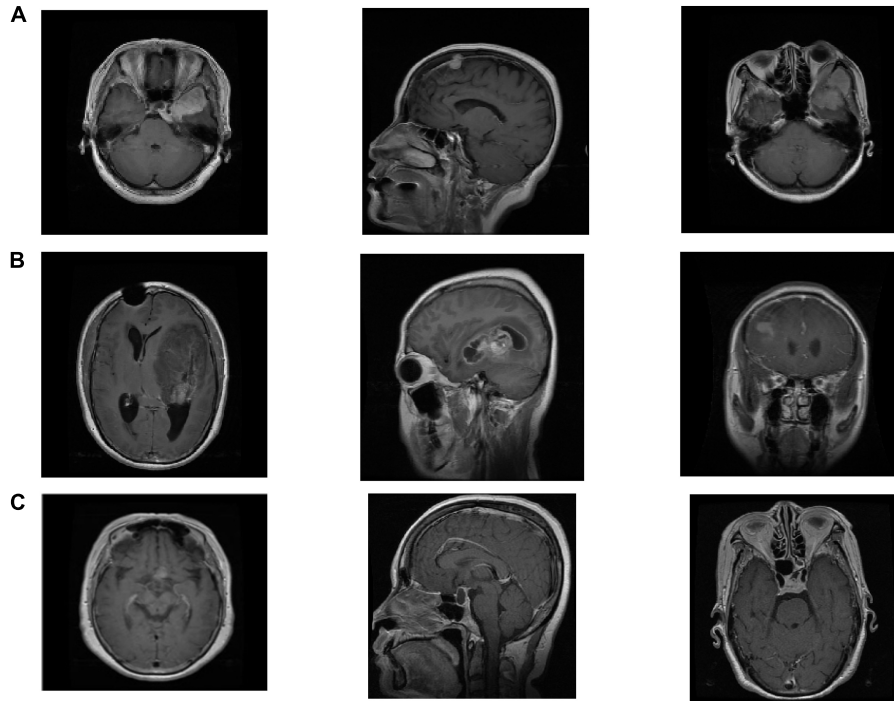
$$B^* = XY^T(YY^T + \Delta)^{-1}. \tag{15}$$

FIGURE 1 | The example samples in the Cheng dataset, **(A)** Meningioma, **(B)** Glioma, and **(C)** Pituitary.

Tune **Y**. With fixed **M** and **B**, the objective function of **Y** can be written as,

$$\min_Y \|X\text{-BY}\|_2^2 + \alpha\, Tr(YLY) + \beta \sum_{i,j} g\left(1 - \ell_{ij}\left(\tau - d_M^2(y_i - y_j)\right)\right)$$
$$+ \gamma \|Y\|_2^2, \tag{16}$$

Taking the first-order partial derivatives of y_i in Eq. 16, we can obtain the close-solution of y_i as,

$$y_i^* = (B^T B + \gamma I + \alpha L + \beta \sum_{i,j} g'(\Lambda) M)^{-1}(B^T Y$$
$$+ \beta M \sum g'(\Lambda) y_j), \tag{17}$$

where $\Lambda = 1 - \ell_{ij}\left(\tau - d_M^2(y_i - y_j)\right)$.

Tune **M**. With fixed **Y** and **B**, the objective function of **M** can be written as,

$$\min_M \beta \sum_{i,j} g\left(1 - \ell_{ij}\left(\tau - d_M^2(y_i - y_j)\right)\right),$$
$$s.t.\ M \geq 0, \tag{18}$$

TABLE 2 | Retrieval performance of SRDML for the query image Glioma on 5-fold test set.

	*m*AP	prec@10	prec@20
Fold 1	97.87	95.75	95.74
Fold 2	97.86	95.71	95.72
Fold 3	97.81	95.68	95.70
Fold 4	97.80	95.70	95.70
Fold 5	97.75	95.68	95.67
Mean (standard deviation)	97.82 (0.044)	95.70 (0.026)	95.70 (0.023)

TABLE 1 | Retrieval performance of SRDML for the query image Meningioma on 5-fold test set.

	*m*AP	prec@10	prec@20
Fold 1	92.12	89.93	89.93
Fold 2	92.13	89.95	89.93
Fold 3	92.14	89.92	89.88
Fold 4	92.19	89.97	89.96
Fold 5	92.19	89.95	89.92
Mean (standard deviation)	92.15 (0.030)	89.94 (0.017)	89.92 (0.026)

TABLE 3 | Retrieval performance of SRDML for the query image Pituitary on 5-fold test set.

	*m*AP	prec@10	prec@20
Fold 1	97.87	96.45	96.44
Fold 2	97.93	96.43	96.46
Fold 3	97.94	96.50	96.50
Fold 4	97.96	96.51	96.47
Fold 5	97.87	96.46	96.54
Mean (standard deviation)	97.92 (0.037)	96.47 (0.030)	96.48 (0.035)

The Lagrange dual function is used, and we have,

$$h(M, \lambda) = \beta \sum_{i,j} g\left(1 - \ell_{ij}\left(\tau - d_M^2(y_i - y_j)\right)\right) + \lambda\|M\text{-}I\|_2^2,$$

(19)

where λ is the Lagrange parameter.

Let $s_{ij} = (y_i - y_j)^T(y_i - y_j)$, taking the first-order partial derivatives of M in Eq. 19, we obtain,

$$\frac{\partial h}{\partial M} = \beta \sum_{\{x_i, x_j\} \in A_{set}} g'(\Lambda)C_{ij} + 2\gamma(M - I).$$

(20)

Then we can obtain M by $M_t = M_{t-1} - \xi \frac{\partial h}{\partial M_{t-1}}$, where ξ is the learning rate in gradient descent method.

The proposed SRDML approach is shown in algorithm 1.

Algorithm 1 SRDML approach
Input: Similar image pair subset and dissimilar image pair subset in **X**, parameters α, β, and γ;
Output: coding coefficient matrix **Y**, dictionary **B**, and metric matrix **M**;
1: Initializing **B** and **Y** with the K-SVD algorithm, **M** with the identity matrix;
2: Calculating graph Laplacian matrix **L** using Eqs 8 and 9;
3: while $t \leq$ maximum number of iterations T_{max}
    Calculating dictionary **B** using Eq. 15;
    Calculating graph Laplacian matrix **L** using Eqs 8 and 9;
    Calculating coding coefficient matrix **Y** using Eq. 17;
    Calculating metric matrix **M** using Eq. 20;
4: Obtaining $B^* = B^{T_{max}}$, $Y^* = Y^{T_{max}}$, and $M^* = M^{T_{max}}$.

EXPERIMENTS

Experimental Settings

We use the Cheng brain MRI dataset (Cheng et al., 2016) to evaluate the effectiveness of the proposed SRDML approach.

This brain MRI dataset contains a total of 3064 slices from 233 patients. The brain MRI images contain three types of tumor images Meningioma, Glioma, and Pituitary. The numbers of the three types of tumor images are 708, 1426, and 930, respectively. The size of each image is 512×512 pixels. 600 images of each type of brain tumor are selected in the experiment. **Figure 1** depicts the example samples of the Cheng dataset.

We use the GIST descriptor implemented by Gumaei et al. (2018, 2019) in this study. GIST descriptor extracts the image features based on the spatial envelope. In the experiment, we obtain 512 dimensional features of each image. We randomly divide brain MRI images into five subgroups of the same size. It is ensured that the different types of tumors in the five subgroups do not overlap and had equal proportions. We use 5-fold cross-validation to evaluate performance. Each image in the test data set is regarded as a query image. We adopt the mean average precision (mAP) and top-n retrieval precision (named as Prec@n) (Cheng et al., 2016). mAP and Prec@n can be calculated as

$$mAP = \frac{\sum_{i=1}^{n} \frac{i}{rank(i)}}{n},$$

(21)

where n represents the number of retrieved images of the same type of brain tumor in the dataset, $rank(i)$ represents the ranking number of the i-th retrieved image of the same type of brain tumor in the search results.

$$Prec@n = \frac{1}{s}\sum_{i=1}^{s}\sum_{j=1}^{k}\frac{j}{Position(j)},$$

(22)

where s is the number of queries, and $Position(j)$ refers to the position of the j-th relevant sample in the search results.

The retrieval performance of SRDML is compared with various matrix learning algorithms LMNN (Weinberger and Saul, 2009), RCA (Bar-Hillel et al., 2005), MPP (Yang et al., 2012), TFFHD (Sun et al., 2018), and RDML-CCPVL (Ni et al., 2018). The parameter setting of the comparison algorithm refers to
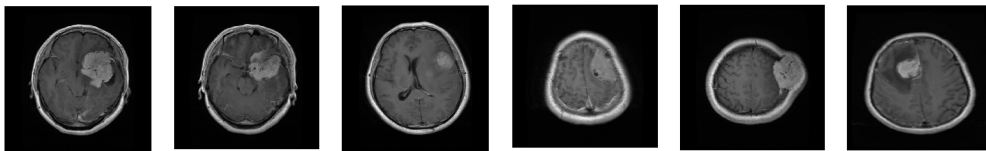
FIGURE 2 | Retrieval result examples of SRDML in Meningioma.
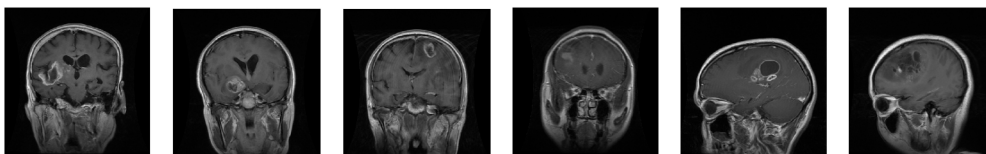

FIGURE 3 | Retrieval result examples of SRDML in Glioma.
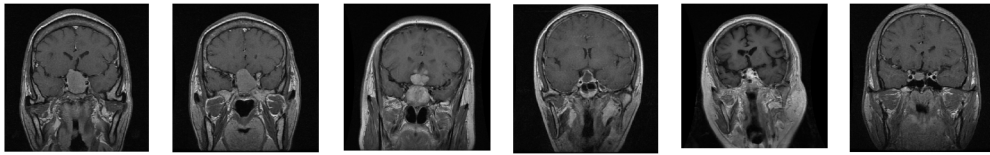
FIGURE 4 | Retrieval result examples of SRDML in Pituitary.

the setting of the original literature. The parameters of SRDML are set as follows. The size of the dictionary is equal to the number of the training set. The regularization parameters are set in the grid $\{10^{-3}, 10^{-2}, ..., 10\}$. The dimension of metric matrix is set in the grid $\{50, 100, 200, ..., 500\}$. The sharpness parameter in generalized logistic loss function is set to be 1. The KNN parameter is set in the grid $\{3, 5, ..., 13\}$. The running environment of all algorithms is CPU i7-8700k, 3.2 GHZ, 32GB RAM, and the running software is Matlab 2019.

Performance Evaluation

Tables 1–3 list the retrieval performance of SRDML for the query image Meningioma, Glioma, Pituitary, respectively. These tables show the results of the 5-fold cross validation test set. We can see that the results obtained by the SRDML approach are consistent on the 5-fold test set, which verifies that the retrieval performance of our approach is stable. We conduct further case analysis on the SRDML. **Figures 2–4** depict the retrieval result examples of SRDML in Meningioma, Glioma, Pituitary, respectively. The first column of images in **Figures 2–4** are the query images. The second to sixth images are five retrieved images. We can see that the retrieved images are all highly related to the query images. The results show that SRDML can give full play to its advantages in brain MRI sets. The SRDML approach uses sparse representation and metric learning to mine the structural features and discrimination information of the data. The local information term and similarity constraints based on coding

coefficients ensure the "maximum inter-class dispersion" and "minimum intra-class dispersion" of the extracted brain MRI features. Thus, the SRDML approach has the important practical value in assisting clinical diagnosis.

Performance Comparison

We use the SRDML approach to compare the performance with LMNN, RCA, MPP, TFFHD, and RDML-CCPVL. **Figures 5–7** show the retrieval performance of various comparison algorithms in Meningioma, Glioma, and Pituitary. The results show these methods in retrieval performance of mAP, Prec@10, as well as Prec@20. According to the results in **Figures 5–7**, the following results are obtained:

(1) In the Meningioma, the SRDML approach is compared with the second-best in mAP, Prec@10, as well as Prec@20 by 0.81, 1.28, and 2.69%, respectively. SRDML has excellent retrieval performance, especially in Prec@20. This shows that the proposed approach has high retrieval performance. In the Glioma, the SRDML approach is compared with the second-best approach in mAP, Prec@10 as well as Prec@20 increased by 1.80, 1.81, and 1.87%, respectively. In the Pituitary, the SRDML approach is compared with the second-best approach in mAP, Prec@10, as well as Prec@20 increased by 1.37, 2.20, and 2.26%, respectively.

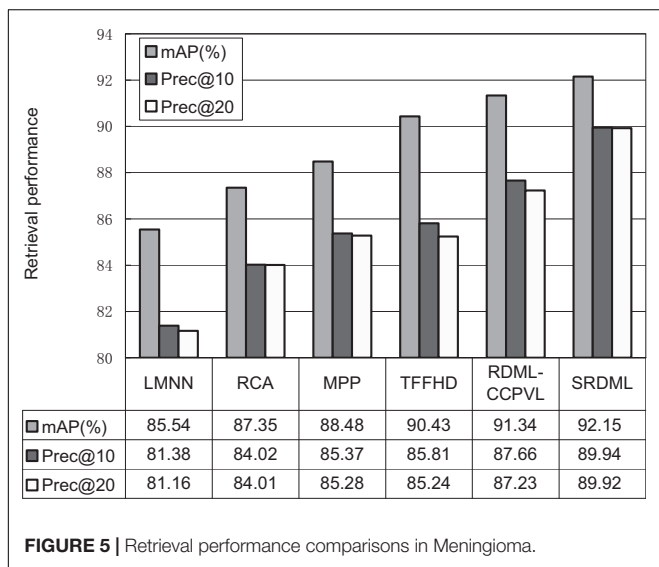(2) The SRDML approach has significantly improved the retrieval performance compared with the traditional

	LMNN	RCA	MPP	TFFHD	RDML-CCPVL	SRDML
■mAP(%)	85.54	87.35	88.48	90.43	91.34	92.15
■Prec@10	81.38	84.02	85.37	85.81	87.66	89.94
□Prec@20	81.16	84.01	85.28	85.24	87.23	89.92

FIGURE 5 | Retrieval performance comparisons in Meningioma.

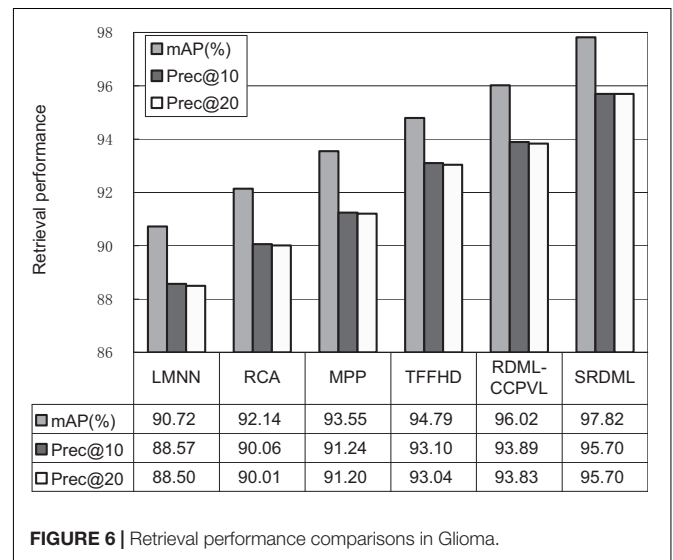|	LMNN	RCA	MPP	TFFHD	RDML-CCPVL	SRDML
■mAP(%)	90.72	92.14	93.55	94.79	96.02	97.82
■Prec@10	88.57	90.06	91.24	93.10	93.89	95.70
□Prec@20	88.50	90.01	91.20	93.04	93.83	95.70

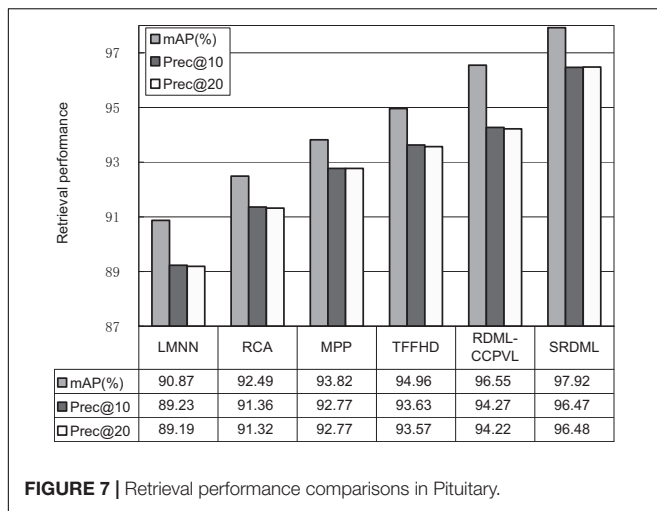**FIGURE 6 |** Retrieval performance comparisons in Glioma.

FIGURE 7 | Retrieval performance comparisons in Pituitary.

TABLE 4 | mAP(%) of the SRDML approach with different k.

	$k = 3$	$k = 5$	$k = 7$	$k = 9$	$k = 11$	$k = 13$
Meningioma	90.53	91.77	**92.15**	**92.15**	92.00	91.42
Glioma	96.50	97.08	97.69	**97.82**	97.47	96.99
Pituitary	97.09	97.51	**97.92**	97.89	97.44	97.01

The bold values represent the best results in comparison experiments.

TABLE 5 | mAP(%) of the SRDML approach with different m.

	$m = 50$	$m = 100$	$m = 200$	$m = 300$	$m = 400$	$m = 500$
Meningioma	89.98	91.43	**92.15**	92.10	92.12	91.58
Glioma	96.90	**97.82**	**97.82**	97.59	97.60	97.01
Pituitary	96.86	**97.92**	97.90	97.80	97.33	97.02

The bold values represent the best results in comparison experiments.

approaches. The SRDML approach uses the sparse representation to learn the robust feature representation of brain MRI images, and learns the metric space with discrimination ability, so that the similar brain MRI pairs are closely mapped in the metric space and the dissimilar brain MRI pairs are separated from each other as much as possible.

Parameter Analysis

The parameters of the SRDML approach are obtained by the grid optimization method. Here we mainly evaluate the influence of KNN parameter k and the dimension of metric matrix m. **Table 4** lists the mAP(%) of the SRDML approach with different k. The parameter k is set in the grid {3, 5, …, 13}. The value of k increases from 3 to 7, and the mAP value increases significantly. It is verified that the local information item is an important part in the SRDML approach. Extracting the appropriate local information can be helpful to describe the image features and improve retrieval accuracy. When the value of k increases from 9 to 13, the retrieval performance begins to decrease, indicating that too large the value of k will cause useless information of the local

information and too much useless information, which affects the retrieval performance of SRDML. **Table 5** lists the mAP(%) of the SRDML approach with different m. The parameter m is set in the grid {50, 100, 200, …, 500}. It can be seen from the results in **Table 5** that when m is set as 100 or 200, mAP(%) of the SRDML approach can reach the optimal retrieval performance. If the value of m is too large or too small, the metric space cannot properly represent the internal data structure of brain MRI images. Therefore, it is feasible to use the grid search method to determine the optimal values of k and m.

CONCLUSION

The purpose of medical image retrieval is to retrieve similar image data from a huge imaging database. The retrieval results should not only be similar in image feature measurement, but also fit in image semantics as much as possible, to help doctors retrieve images with the same pathology. This paper proposes a similarity retrieval approach for brain MRI images based on sparse representation and metric learning. The characteristics of this approach are using sparse representation to extract the robust features of brain MRI image, which can more effectively represent the information of brain MRI image. The relative similarity constraint and local preserving information are used to project the low dimensional metric matrix, which improves the retrieval accuracy. The experiments are carried out on public data sets. The experiments show the effectiveness and accuracy of our approach in brain MRI image retrieval. The future work will be considered from the following aspects: (1) while improving the retrieval accuracy of brain MRI images, ensure the retrieval speed, to further provide assistance and support for doctors in diagnosing brain tumor lesions. (2) The experimental data scale of this paper is not large enough. During the retrieval process, several images with the highest similarity are returned by comparing the feature vector similarity of the data set in turn. When dealing with large-scale data, we will consider how to combine the technologies such as hash sorting of retrieval results to realize the real-time performance of retrieval results. (3) For different images and images with different principles, the feature algorithms involved are different. Forming a more general medical image retrieval system is the next research direction.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The Cheng dataset analyzed for this study can be found in this link [https://figshare.com/articles/dataset/brain_tumor_dataset/1512427].

AUTHOR CONTRIBUTIONS

GZ and TN conceived and designed the proposed model and wrote the manuscript. BL and XH ran the experiment and analyzed the results. All authors read and approved the manuscript.

FUNDING

REFERENCES

Bar-Hillel, A., Hertz, T., Shental, N., Weinshall, D., and Ridgeway, G. (2005). Learning a Mahalanobis metric from equivalence constraints. *J. Mach. Learn. Res.* 612, 937–965. doi: 10.1115/1.1902843

Cao, X., Zhang, H., Guo, X., Liu, S., and Meng, D. (2015). SLED: semantic label embedding dictionary representation for multilabel image annotation. *IEEE Trans. Image Process.* 24, 2746–2759. doi: 10.1109/TIP.2015.2428055

Chahal, P. K., Pandey, S., and Goel, S. (2020). A survey on brain tumor detection techniques for MR images. *Multimed. Tools Appl.* 79, 21771–21814. doi: 10.1007/s11042-020-08898-3

Cheng, J., Yang, W., Huang, M., Huang, W., Jiang, J., Zhou, Y., et al. (2016). Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation. *PLoS One* 11:e0157112. doi: 10.1371/journal.pone.0157112

Gao, S., Chia, L., Tsang, I. W., and Ren, Z. (2014). Concurrent single-label image classification and annotation via efficient multi-layer group sparse coding. *IEEE Trans. Multimedia* 16, 762–771. doi: 10.1109/TMM.2014.2299516

Gu, D., Liang, C., and Zhao, H. (2017). A case-based reasoning system based on weighted heterogeneous value distance metric for breast cancer diagnosis. *Artif. Intell. Med.* 77, 31–47. doi: 10.1016/j.artmed.2017.02.003

Gu, X., Zhang, C., and Ni, T. (2021a). A hierarchical discriminative sparse representation classifier for EEG signal detection. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 1679–1687. doi: 10.1109/TCBB.2020.3006699

Gu, X., Shen, Z., Xue, J., Fan, Y., and Ni, T. (2021b). Brain tumor MR image classification using convolutional dictionary learning with local constraint. *Front. Neurosci.* 15:679847. doi: 10.3389/fnins.2021.679847

Gumaei, A., Sammouda, R., Malik, A., Al-Salman, S., and Alsanad, A. (2018). An improved multispectral palmprint recognition system using autoencoder with regularized extreme learning machine. *Comput. Intell. Neurosci.* 2018:8041609. doi: 10.1155/2018/8041609

Gumaei, A., Sammouda, R., Malik, A., Al-Salman, S., and Alsanad, A. (2019). Anti-spoofing cloud-based multi-spectral biometric identification system for enterprise security and privacy-preservation. *J. Parallel Distrib. Comput.* 124, 27–40. doi: 10.1016/j.jpdc.2018.10.005

Huang, M., Yang, W., Wu, Y., Jiang, J., Gao, Y., Chen, Y., et al. (2014). Content-based image retrieval using spatial layout information in brain tumor T1-weighted contrast-enhanced MR images. *PLoS One* 9:e102754. doi: 10.1371/journal.pone.0102754

Huang, M., Yang, W., Yu, M., Lu, Z., Feng, Q., and Chen, W. (2012). Retrieval of brain tumors with region-specific bag-of-visual-words representations in contrast-enhanced MRI images. *Comput. Math. Methods Med.* 2012:280538. doi: 10.1155/2012/280538

Jiang, Y., Gu, X., Wu, D., Hang, W., Xue, J., Qiu, S., et al. (2021). A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 40–52. doi: 10.1109/TCBB.2019.2963873

Lu, Z., Han, P., Wang, L., and Wen, J. (2015). Semantic sparse recoding of visual content for image applications. *IEEE Trans. Image Process.* 24, 176–188. doi: 10.1109/TIP.2014.2375641

Müller, H., and Unay, D. (2017). Retrieval from and understanding of large-scale multi-modal medical datasets: a review. *IEEE Trans. Multimed.* 19, 2093–2104. doi: 10.1109/TMM.2017.2729400

Nguyen, B., Morell, C., and Baets, B. D. (2020). Scalable large-margin distance metric learning using stochastic gradient descent. *IEEE Trans. Cybern.* 50, 1072–1083. doi: 10.1109/TCYB.2018.2881417

Ni, T., Ding, Z., Chen, F., and Wang, H. (2018). Relative distance metric leaning based on clustering centralization and projection vectors learning for person re-identification. *IEEE Access* 6, 11405–11411. doi: 10.1109/ACCESS.2018.2795020

Srinivas, U., Mousavi, H. S., Monga, V., Hattel, A., and Jayarao, B. (2014). Simultaneous sparsity model for histopathological image representation and classification. *IEEE Trans. Med. Imaging* 33, 1163–1179. doi: 10.1109/TMI.2014.2306173

Sun, X., Zhang, N., Wu, H., Yu, X., Wu, X., and Yu, S. (2018). Medical image retrieval approach by texture features fusion based on Hausdorff distance. *Math. Probl. Eng.* 2018:7308328. doi: 10.1155/2018/7308328

Tang, J. H., Hong, R. C., Yan, S. C., and Chua, T. S. (2010). Image annotation by kNN-sparse graph-based label propagation over noisily-tagged Web images. *ACM Trans. Intell. Syst. Technol.* 2, 1–15. doi: 10.1145/1899412.1899418

Verma, Y., and Jawahar, C. V. (2017). Image annotation by propagating labels from semantic neighbourhoods. *Int. J. Comput. Vis.* 121, 126–148. doi: 10.1007/s11263-016-0927-0

Weinberger, K. Q., and Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. *J. Mach. Learn. Res.* 10, 207–244. doi: 10.1007/s10845-008-0108-2

Yang, W., Feng, Q., Yu, M., Lu, Z., Gao, Y., Xu, Y., et al. (2012). Content-based retrieval of brain tumor in contrast-enhanced MRI images using tumor margin information and learned distance metric. *Med. Phys.* 39, 6929–6942. doi: 10.1118/1.4754305

Zhang, X., Dou, H., Ju, T., Xu, J., and Zhang, S. (2015). Fusing heterogeneous features from stacked sparse autoencoder for histopathological image analysis. *IEEE J. Biomed. Health Inform.* 20, 1377–1383. doi: 10.1109/JBHI.2015.2461671

Zhang, Y., Chen, C., Tian, Z., Feng, R., Cheng, Y., and Xu, J. (2019). The diagnostic value of MRI-based texture analysis in discrimination of tumors located in posterior fossa: a preliminary study. *Front. Neurosci.* 13:1113. doi: 10.3389/fnins.2019.01113

Zheng, Y., Jiang, Z., Zhang, H., Xie, F., Ma, Y., Shi, H., et al. (2018). Histopathological whole slide image analysis using context-based CBIR. *IEEE Trans. Med. Imaging* 37, 1641–1652. doi: 10.1109/TMI.2018.2796130