Brain Tumor Segmentation Based on Local Independent Projection-based Classification

Meiyan Huang, Wei Yang, Yao Wu, Jun Jiang, Wufan Chen, Senior Member, IEEE, and Qianjin Feng*, Member, IEEE

Abstract—Brain tumor segmentation is an important procedure for early tumor diagnosis and radiotherapy planning. Although numerous brain tumor segmentation methods have been presented, enhancing tumor segmentation methods is still challenging because brain tumor MRI images exhibit complex characteristics, such as high diversity in tumor appearance and ambiguous tumor boundaries. To address this problem, we propose a novel automatic tumor segmentation method for MRI images. This method treats tumor segmentation as a classification problem. Additionally, the local independent projection-based classification (LIPC) method is used to classify each voxel into different classes. A novel classification framework is derived by introducing the local independent projection into the classical classification model. Locality is important in the calculation of local independent projections for LIPC. Locality is also considered in determining whether local anchor embedding is more applicable in solving linear projection weights compared with other coding methods. Moreover, LIPC considers the data distribution of different classes by learning a softmax regression model, which can further improve classification performance. In this study, 80 brain tumor MRI images with ground truth data are used as training data and 40 images without ground truth data are used as testing data. The segmentation results of testing data are evaluated by an online evaluation tool. The average dice similarities of the proposed method for segmenting complete tumor, tumor core, and contrast-enhancing tumor on real patient data are 0.84, 0.685, and 0.585, respectively. These results are comparable to other state-of-the-art methods.

Index Terms—Brain tumor segmentation, local independent projection-based classification, local anchor embedding, softmax regression

I. INTRODUCTION

BRAIN tumor segmentation is one of the crucial procedures in surgical and treatment planning. However, at present, brain tumor segmentation in brain tumor images is mostly performed manually in clinical practice. Apart from being time consuming, manual brain tumor delineation is difficult and depends on the individual operator. Currently, multimodal MRI images are used simultaneously by radiologists in segmenting brain tumor images because multimodal MRI images can provide various data on tumors. In glioma, the tumor area is usually divided into necrosis, contrast-enhancing tumor, non-enhancing tumor, and edema [1]. Different MRI image modalities can reveal different parts in the tumor area. For instance, T1C (T1-weighted images with contrast enhancement) highlights contrast-enhancing regions, whereas T2 highlights edema regions (Fig. 1). Although multimodal MRI images can provide complementary information in the tumor area, brain tumor segmentation is still a challenging and difficult task. Brain tumors can have various sizes and shapes and may appear at different locations. In addition to tumor heterogeneity, tumor edges can be complex and visually vague (Fig. 1). Moreover, some tumors may deform surrounding structures in the brain because of the mass effect or edema (Fig. 1). Additionally, artifacts and noise in brain tumor images increase the difficulty when segmenting tumors. Thus, designing of a semi-automatic or automatic brain tumor segmentation approach is necessary to provide an acceptable performance.

Numerous algorithms have been developed to perform brain tumor detection and segmentation. These methods include thresholding and morphological techniques [2-4], watershed method [5], region growing approach [4, 6], asymmetry analysis [7-10], atlas-based method [11-14], contour/surface evolution method [15-17], interactive algorithm [18-20], and supervised [21-29] and unsupervised [6, 30-35] learning methods.

A. Intensity-Based Method

Low-level operations, such as thresholding, edge detection,
and morphological techniques [2-4], are fast and can be easily adjusted. However, the tumor segmentation performance of these methods highly depends on evident difference in the intensities between tumor and non-tumor regions. Watershed [5] and region growing [4, 6] approaches are simple and consistently produce complete boundaries. However, these two methods are sensitive to noise, which is a common problem in the intensity-based method. Moreover, most intensity-based methods tend to oversegment tumors because of the weak and diffused edges caused by edema [15].

B. Asymmetry Analysis

The healthy human brain is largely symmetric across the mid-sagittal plane. The asymmetric analysis method [7-10] for tumor segmentation is based on the principle that tumors, which appear in one of the cerebral hemispheres, can cause asymmetry between the left and right cerebral hemispheres. This asymmetry can be detected, and tumors can be roughly located in the corresponding cerebral hemisphere. The asymmetric analysis method can hasten the tumor detection and segmentation process because tumor segmentation is implemented in one of the cerebral hemispheres. However, accurately finding the mid-sagittal plane is a challenging and time-consuming task [7]. More importantly, asymmetry analysis may not be useful when a tumor is located across the mid-sagittal plane [9].

C. Atlas-Based Method

Atlas-based segmentation methods have been extensively investigated [36, 37]. Brain atlases can provide important data prior to tumor segmentation enhancement by measuring the difference between abnormal and normal brains. However, the deformable registration of the brain atlas to brain images with tumor is an extremely challenging task because of the intensity variations around the tumor caused by edema and the deformations of healthy tissue morphology caused by the tumor mass effect [38]. In a previous study [11], affine registration is used to align the atlas to the tumor image data. When a large brain structure deformation appears, the misalignment issues are noted on the aligned atlas, which may significantly decrease segmentation accuracy [11].

D. Contour/Surface Evolution Method

The contour/surface evolution method has been widely used for the tumor segmentation of 2D/3D data [15-17]. This method can be represented implicitly as a level set function or explicitly as an active contour model/snake function. Compared with the parametric active contour model, the level set method can represent contours with complex topology and handle topological changes, such as splitting and merging in a natural and efficient way [19]. Furthermore, the extension of the level set method to 3D is straightforward and does not require additional machinery [16]. However, the contour/surface evolution method does not easily determine the initializations and tune the parameters even when 3D level set surfaces are used.

E. Interactive Algorithm

Graph-based seeded segmentation framework is one of the popular methods among interactive algorithms. Graph-based seeded segmentation is a global optimization approach, which showed outstanding performance for tumor segmentation in our previous studies [19, 20]. However, this method needs manual seed selection in different tissues, and distinguishing different tissues in the tumor is difficult during the selection of initial seeds for different tissues. In a previous study [18], a cellular automata-based seeded method, called tumor-cut, has been presented for brain tumor segmentation. This method only requires the user to draw a line over the largest visible tumor diameter. Although this initial seed selection strategy can reduce manual interaction and decrease the sensitivity of the method to initialization, this procedure may not include all tumor areas within the volume of interest along the depth direction [18], thus leading to tumor undersegmentation.

F. Supervised and Unsupervised Learning Methods

Unsupervised learning method, such as k-means and fuzzy clustering, has become popular for brain tumor segmentation in recent years [6, 30-35]. The fuzzy method considers that medical images are inherently fuzzy, so it is a very strong tool for medical image processing. Furthermore, the fuzzy method can capture pixel proximity in the same objective region without a training step. However, most fuzzy methods work well only for tumors that present hyper-intensity and exhibit poor performance on segmenting non-enhanced tumors [7]. These conditions are due to the fact that these fuzzy methods typically use intensity-based method, such as thresholding and morphological operations, as pre- or post-processing.

Supervised classification learning method is widely used in tumors segmentation [21-29]. Well-trained classifiers can extract discriminative information from the training data and estimate the label of each voxel in a testing volume. However, the traditional classification methods classify each voxel into different classes without considering the spatial correlation between current and nearby voxels. This method may not obtain a global optimized result. To address this problem, a classification method is generally combined with a regularization step. The regularization step can be implemented by modeling the boundary [25] or by applying a variant of a
random field spatial prior (MRF/CRF) [23, 24, 26, 29]. In the previous studies [21, 22], context-aware spatial features and the probabilities obtained by tissue-specific Gaussian mixture models are used as inputs for classifiers, and satisfied segmentation results are achieved without using post-hoc regularization. Similar to the previously described methods [21, 22], we propose a novel classification method, named local independent projection-based classification (LIPC), for brain tumor segmentation without using explicit regularization.

The contribution of this study can be summarized as follows.

First, an LIPC-based method is introduced to solve the tumor segmentation problem. Based on recent studies on sparse representation [39-41], the sparse representation-based classification (SRC) method [42] was adopted in medical image segmentation, thereby producing robust results. In the SRC method, a sample can be sparsely represented in a specific dictionary, which contains samples from all classes. Contrary to SRC, the proposed LIPC assumes that the training samples from different classes are located on different submanifolds. Then, the training samples are divided into different groups and subsequently used to construct different dictionaries. The testing sample is independently projected on these different dictionaries using the local anchor embedding (LAE) method [43].

Second, a softmax model is used to determine the relationship between data distribution and reconstruction error norm. When the data distribution is uniform and the noise is low, classification may be performed well. However, data distribution is complex in brain tumor MRI images. In addition, the data distribution of different classes (i.e., tumor, edema, and brain tissue) may vary widely. Therefore, the data distribution of each class should be considered when segmenting brain tumors. Our evaluation on synthetic data and public available brain tumor image data demonstrates that considering the data distribution of different classes can further improve the segmentation performance.

This study is organized as follows. Section II introduces methodology of the proposed algorithm. Section III presents the experimental analysis and results. Section IV provides several discussions and conclusions on the proposed method. A preliminary work of this study appeared in another study [44].

TABLE I

<table>
<thead>
<tr>
<th>DETAILS OF THE BRAIN TUMOR IMAGE DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real patient data</td>
</tr>
<tr>
<td>High-grade</td>
</tr>
<tr>
<td>Training data</td>
</tr>
<tr>
<td>Testing data</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Fig. 2. Flowchart of the proposed method

In the ground truth data, the complete tumor region is subdivided into necrosis, contrast-enhancing tumor, non-enhancing tumor, and edema parts for real patient data, whereas the complete tumor region is subdivided into tumor core and edema parts for synthetic data. The remaining 40 images are testing data with publicly unavailable ground truth data. T1, T2, FLAIR, and post Gadolinium T1 MRI images are available for each subject. All volumes are linearly co-registered to the T1 contrast image, skull stripped, and interpolated to 1 mm isotropic resolution.

In addition to the brain tumor image data, a set of interleaving spirals was also used in elucidating the mechanism of the proposed LIPC method.

B. Overview of the Proposed Method

The proposed method consists of four major steps, i.e., preprocessing, feature extraction, tumor segmentation using the LIPC method, and postprocessing. To reduce computational costs, we embedded the proposed method in a multi-resolution framework. The flowchart of the proposed method is illustrated in Fig. 2.

C. The basic principle of LIPC

Brain tumor segmentation can be considered as a multiclass classification problem. To solve this problem, a one-versus-all (OvA) strategy can be used. In the OvA approach, a classifier is trained per class to distinguish a class from all other classes. Therefore, N classifiers $f = \{f_i\}_{i=1}^N$ have to be determined in this study, where $N$ represents the number of classes. Given a testing sample $x \in \mathbb{R}^M$, $N$ real classification scores $y = \{y_i\}_{i=1}^N$ are computed using the learned classifiers $f(x)$; where the sample $x$ stands for the image feature in the current study (details in Section II. F) and $y_i \in [0,1]$ stands for the probabilities that the sample belongs to the $i$th class. The label of sample $x$ can be defined as follows.
l = \arg \max_{i=1,\ldots,N} f_i(x) = \arg \max_{i=1,\ldots,N} y_i. \quad (1)

Before the proposed LIPC was introduced, the following assumption was considered as the base for LIPC:
Assumption I: Samples from different classes are located on different non-linear submanifolds, and a sample can be approximately represented as a linear combination of several nearest neighbors from its corresponding submanifold.

For N-class classification, this assumption indicates that the samples are found on a manifold \( \{ M_i \}_{i=1}^N \), which consists of N submanifolds; \( M_i \) represents the submanifold associated with the ith class. For a dictionary \( D = \{ D_i \}_{i=1}^N \), which consists of N subdictionaries, \( D_i = [d_i^1, d_i^2, \ldots, d_i^N] \) consists of \( N_i \) typical samples from the ith submanifold.

Considering Assumption I, a testing sample \( x \) can be projected into each submanifold using the following linear representation:
\[
x = D_i a_i + \varepsilon_i = \sum_{j=1}^{N_i} a_i^j d_i^j + \varepsilon_i
s.t. \quad \|\varepsilon_i\| < \tau, \forall d_i^j \not\in N_i^c(k), a_i^j = 0
\]
where \( N_i^c(k) \) denotes the set of \( k \) nearest neighbors of \( x \) in dictionary \( D_i \), \( \varepsilon_i \) is the reconstruction error, \( a_i = [a_i^1, a_i^2, \ldots, a_i^N]^T \) is the weight coefficient vector of the linear combination, and \( \tau \) is a small positive real number that ensures the reconstruction accuracy. Many (at least \( N_i-k \) elements of \( a_i \) are zeroes for an arbitrary sample \( x \), which indicates that the described linear representation is sparse. The L2-norm of \( \varepsilon_i \), noted as \( \| \varepsilon_i \|_2 \), can be used in measuring the proximity between \( x \) and its projection on \( M_i \). Furthermore, a smaller \( \| \varepsilon_i \|_2 \) is associated with increased probability that \( x \) belongs to the ith class. Therefore, the classification scores of \( x \) have a negative correlation with \( \| \varepsilon_i \|_2 \) and can be formulated as
\[
y_i = f_i(x) = f_i\left( \sum_{j=1}^{N_i} a_i^j d_i^j + \varepsilon_i \right)
= g_i(\varepsilon_i) = g_i\left( x - \sum_{j=1}^{N_i} a_i^j d_i^j \right) \quad (3)
\]
where \( g_i(\cdot) \) is an inverse proportional function. Based on Eq. (3), the label prediction in Eq. (1) can be rewritten as
\[
l = \arg \max_{i=1,\ldots,N} f_i(x) = \arg \max_{i=1,\ldots,N} y_i = \arg \max_{i=1,\ldots,N} g_i(\varepsilon_i). \quad (4)
\]

D. LIPC Implementation

We performed dictionary construction, locally linear representation, and classification score computation to implement LIPC.

1) Dictionary construction

The manually labeled original samples in a training set are used to construct \( D \) [41]. However, numerous original training samples possibly produce a large \( D \), which dramatically increases computational and memory costs. In the present study, more than half a million samples for each class are available for training. Thus, subsequent processes are impractical when conducted traditionally. Applying a dictionary learning method is necessary in learning a compact representation of the original training samples. The \( k \)-means [46] method can obtain the typical structures of the original sample space; and thus, this method is used in the current study to learn a compact representation of the original training samples of each class.

2) Locally linear representation

Several methods, such as sparse coding (SC) [41], locality-constrained linear coding (LLC) [47], and LAE, have been proposed to approximately represent a sample that is linearly based on training samples. Sparse representation emphasizes the sparsity of the representation and attempts to use the smallest number of training samples in reconstructing the testing sample with minimal error. LLC and LAE belong to the special cases of sparse representation. These two approaches focus on locality rather than sparsity by limiting the linear coding within a local neighborhood. In particular, LAE forces the weight coefficients to maintain non-negativity and a sum that is equal to one, which ensures that the reconstructed sample is a convex combination of its closest neighbors. Therefore, the reconstructed sample is located in a convex region on a hyperplane spanned by its closest neighbors. Actually, this condition is a tough constraint of the locality. The locality of the representation is highly preferable for classification tasks [43, 47, 48]. Thus, LAE is appropriate in solving the linear representation in the current study. For completeness, we briefly introduce LAE.

Considering the concrete task in this paper, we formally re-formulate the cost function of LAE as
\[
a^* = \arg \min_a \left\| x - \sum_{j=1}^{N} a_j d_j \right\|^2 \quad (5)
s.t. \quad \forall d_j \not\in N_i^c(k), a_j = 0, \sum_{j=1}^{N} a_j = 1, a_j \geq 0
\]

Three steps were performed to obtain the solution of LAE. First, the \( k \) nearest neighbors of the testing sample \( x \) was selected from \( D \) and \( N_i^c(k) \) was constructed. Second, for \( d_s \) that do not belong to \( N_i^c(k) \), the associated \( a_j \)s were set to 0.

Third, for the remaining \( d_s \) that belong to \( N_i^c(k) \), their corresponding \( a_j \)s was calculated using the projected gradient method [43]. The updating rule in the projected gradient method can formally be expressed as the following iterative formula:
where \(t\) is the number of iterations, \(\eta > 0\) is the step size, \(\nabla h(a)\) is the gradient of function \(h\) at \(a\), and \(P_x(a)\) is the simplex projection operator on any \(a \in \mathbb{R}^p\). The simplex projection operator can be formulated as

\[
P_x(a) = \arg \min_{a' \in \mathcal{X}} \|a' - a\|\]

As introduced in [43], Eq. (6) uses the simplex projection operator in Eq. (7), is implemented efficiently in \(O(s \log s)\), and can be accelerated using Nesterov’s method [49].

3) Classification score computation

As described in Section II. C, the classification scores of \(x\) have a negative correlation with \(\|e\|\). Furthermore, to implement the classification within a multi-resolution framework (details in Section II. G), the classification score can be easily represented as a probability form:

\[
y_i = g(e_i) = \exp(-\|e_i\|)/\sum_{i=1}^N \exp(-\|e_i\|)
\]

However, Eq. (8) considers only the relationship between the reconstruction error norm and the classification score, which ignores the relationship between the reconstruction error norm and the data distribution on each submanifold. In general, the data distributions on different submanifolds are different. Therefore, the reconstruction error norms of the samples cover a wide range and may sometimes violate the negative correlation as listed in Eq. (8) [Fig. 3 (a)]. To address this problem, we propose to establish a softmax regression model using the reconstruction error norms of the training data to achieve the classification score [Fig. 3 (b)]. Thus, the classification score can be defined as

\[
y_i = \exp(w_i^T \|e_i\|)/\sum_{i=1}^N \exp(w_i^T \|e_i\|)
\]

Given the original training set \(T = \{T_i\}_{i=1}^N = \{x_i, l_i\}_{i=1}^N\) and the learned dictionary \(D = \{D_i\}_{i=1}^N\), the weight coefficient vector \(a^{(i)}\) that corresponds to each sample in \(T\) can be calculated using Eq. (5). Then, the reconstruction error norm \(\|e_i\| = \left\| \left[ \|e_i^{(1)}\|, \ldots, \|e_i^{(N)}\| \right]^T \right\|\) of all training samples in \(T\) can be computed using Eq. (2). With \(\|e\|\), the cost function of the softmax regression model can be defined as

\[
J(w) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N 1\{l(i) = j\} \log \frac{\exp(w^T \|e_{ij}\|)}{\sum_{j=1}^N \exp(w^T \|e_{ij}\|)}
\]

where \(1\{\cdot\}\) is an indicator function. Given that no closed-form technique can be used to solve for the minimum of \(J(w)\), we employ the gradient descent algorithm to iteratively optimize Eq. (10).

E. Summary of LI PC

We provide the pseudo-code of the proposed method in Algorithm 1.

F. Feature Extraction

Before extracting image features to represent a sample, image inhomogeneity correction and intensity normalization should be performed because the image intensities in MRI images do not have a fixed meaning and widely vary within or between subjects. In this study, all MRI image modalities are processed as follows. First, the N3 algorithm [50] is applied to remove the bias field artifacts from the images. Second, intensity values at the 1% and 99% quantiles are computed for the brain region (including tumors, edema, and brain tissues), and then these two values are used to linearly scale the voxel intensities to the range [0,100].

In the present paper, a patch-based technique was used in extracting the image feature. The intensity values in a patch around a voxel \(v\) were obtained and rearranged as a feature vector. Four modalities of MRI images were used, so the feature of \(v\) that used a cubic patch with a size of \(w \times w \times w\) was \((w^3 \times 4)\)-dimensional.

G. Multi-resolution Framework

To improve robustness and reduce computational cost, a multi-resolution framework was adopted in the proposed method. For a multi-resolution framework with \(P\) levels, classification originated from the coarsest \(L_{P-1}\) to the finest level \(L_0\) (original resolution) to classify the voxels into different classes. The classification scores for all voxels in coarser levels were up-sampled to initialize the classification for a finer level using the trilinear interpolation method. At each level, a confidence threshold \(\alpha\) was defined to determine the specific voxels that can be directly labeled using the propagated initial classification scores and the specific voxels that required further processing. In this study, the voxels with initial classification scores \(y_i > 1 - \alpha\) were labeled as the \(i\)th class.

The remaining voxels with \(y_i \in [0,1-\alpha]\) were fed into the proposed classifier for an accurate classification. This procedure was repeated until the resolution of the original level.
ALGORITHM I

Input: Training set \( T = \{T_i\}_{i=1}^{N} = \{(x_i, y_i)\}_{i=1}^{N} \); A testing sample \( x \).

Output: The classification scores \( y = \{y_i\}_{i=1}^{N} \) and the label \( l \) of \( x \).

Stage 1: Construction of sub dictionary for each class
Partition \( \{x_i, T_i\}_{i=1}^{N} \) into \( N \) sub-sets and calculate the cluster centers denoted as \( D_i = [d_i^1, \ldots, d_i^K] \) using k-means methods.

Stage 2: Calculation of locally linear representation coefficients
Reconstruct each \( x_i \) in \( T \) based on dictionary \( D \) using the LAE method and calculate the corresponding coefficient vector \( \{a_i^k\}_{k=1}^{K} \).

Stage 3: Softmax regression model determination
Calculate reconstruction error vector \( g_i \) of all training samples for each class based on dictionary \( D = \{D_i\}_{i=1}^{N} \) and coefficient vectors \( \{a_i^k\}_{k=1}^{K} \).

Use the calculated reconstruction error norms to determine a softmax regression model by minimizing Eq. (10).

Stage 4: Core of the proposed method
Reconstruct the input sample \( x \) based on dictionary \( D \) using the LAE method and calculate the coefficient vectors \( \{a_{i,l}\}_{l=1}^{L} \).

Calculate the reconstruction error vector \( g \) of \( x \) for each class according to Eq. (2).

Calculate the classification scores \( y = \{y_{i,l}\}_{l=1}^{L} \) of \( x \) based on Eq. (9).

Achieve the label \( l \) of \( x \) based on Eq. (1).

When \( L_0 \) was reached. At level \( L_0 \), we directly used Eq. (1) to achieve the final label.

Three levels (\( P = 3 \)) were used with corresponding voxel sizes of 1 mm × 1 mm × 1 mm, 2 mm × 2 mm × 2 mm, and 4 mm × 4 mm × 4 mm. The voxel number for each class in each training image was set to 8,000 to construct the training set at each level, and the threshold value \( \alpha \) was empirically set to 0.2 at \( L_1 \) and \( L_2 \). In the coarser levels, the voxels in the tumor or edema regions were very few. To increase the number of training samples, we randomly transformed the MRI images by smoothing and adding noise to the images at each level.

H. Post-processing with Spatial Constraint

In this study, we used the assumption that each edema region is located near tumor core regions to post-process the classified edema region as introduced in [11]. According to this assumption, each classified edema region must have a voxel near the classified tumor regions within some small distance. Therefore, the connected component algorithm and mathematical morphology can be used to refine the classified edema regions. First, a binary image representing the classified edema regions is formed. Second, the binary image is used as an input for the connected component algorithm, and then some individual edema regions are generated. Third, each individual edema region is dilated with a small structuring element and compared against the classified tumor regions. Finally, the dilated edema regions that share at least a voxel with the classified tumor regions are considered valid. The edema regions from these valid regions are retained as the final edema classification results, whereas the other edema regions are discarded.

III. EXPERIMENTAL RESULTS

A. Experiments on Synthetic Data

A two-spiral structure was generated to evaluate the classification performance of the proposed LIPC method. A comparison was also made between the proposed method and two other popular classification methods [SRC and support vector machine (SVM) [51]] using the generated two-spiral structures. As described in Section II, C, SRC uses the whole dictionary to represent the testing sample and classifies \( x \) based on the performance of the coefficients associated with the training samples. In the proposed LIPC method, each subdictionary was used to represent \( x \), and the closest subspace was determined to accomplish the classification task. Moreover, a softmax model was learned in LIPC to capture the information of data distribution of different classes. SVM attempts to minimize the empirical classification error and maximize the geometric margin simultaneously in the training set, which leads to high generalization ability in the new samples.

In this section, the parameters in the two-spiral structures...
were designed as follows. The number of training samples on one spiral was fixed to 500. Then, the number of training samples on other spiral was set to 50, 250, and 500. The number of testing samples on each spiral was set to 1000. Noises with different levels (0.5 to 3) were also added to the generated interleaving spirals. The coordinates of each point in the spirals were the features to be fed to different classification methods. For LIPC and SRC, the training samples were used as dictionary because the sample number was small. To evaluate the LIPC effectiveness, both SRC and LIPC used LAE as the coding method. To make a fair comparison, we carefully designed and selected the parameters with optimal performance in each classification method. The classification results of different methods on different data are shown in Fig. 4. When the number of training samples on the two spirals was the same (e.g. 500 and 500) and the noise was small, the difference in the classification accuracy of the different methods was nonsignificant. However, the classification accuracy of the different methods varied significantly when the training samples on one spiral were much sparser than the training samples on the other spiral (e.g., 50 and 500). A higher noise level led to a large variety of classification accuracy under the different methods. In general, the classification accuracy decreased when the data distribution was non-uniform and the noise level was high. However, the proposed method with a learned softmax regression model achieved the lowest error rate when a serious non-uniformity of data distribution and high noise were observed in the training samples. The higher classification accuracy of the proposed method contributes in considering the data distribution of different classes. The classification performance of the proposed method without the learned softmax regression model was comparable to that of SVM. The proposed method with or without the learned softmax regression model also outperformed the SRC method, which indicates that the proposed LIPC was more effective than SRC in this classification task. The decision curves of the proposed method are shown in Fig. 5. As shown in Figs. 5 (b) and (c), the decision curves of the proposed method with and without the learned softmax regression model were similar when the non-uniformity of the data distribution was not obvious. However, the decision curve of the proposed method without the learned softmax regression model was worse than that of the proposed method with the learned softmax regression model when the training samples on one spiral were much sparser than those on the other spiral [Figs. 5 (e) and (f)].

**Table II** Summary of the parameter settings used in the proposed method

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k)</td>
<td>Number of the nearest neighbors of LAE</td>
<td>10</td>
</tr>
<tr>
<td>(N)</td>
<td>Dictionary size</td>
<td>40,000</td>
</tr>
<tr>
<td>(w)</td>
<td>Patch size ((w \times w \times w))</td>
<td>5</td>
</tr>
<tr>
<td>(P)</td>
<td>Levels of Multi-resolution</td>
<td>3</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>Confidence threshold described in Section</td>
<td>2.5 to 0.2 at levels</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{L}_0 & \quad \text{Tumor} \\
\text{L}_0 & \quad \text{Edema} \\
\text{L}_1 & \quad \text{Background} \\
\end{align*}
\]

Fig. 5. (a) The number of training samples in each of the two classes was 500, and the noise level was 2.9. The decision curves learned by LIPC (b) without and (c) with the learned softmax regression model on the testing samples. (d) The number of training samples in the two classes was 50 and 500, and the noise level was 2.9. The decision curves learned by LIPC (e) without and (f) with the learned softmax regression model on the testing samples.

Fig. 6. Effect of different numbers of \(k\) nearest neighbors on the reconstruction error.

B. **Experiments on Brain Tumor Image Data**

A series of experiments were performed on the training and testing brain tumor image data. For the training data, a complete tumor is subdivided into tumor core (necrosis, contrast-enhancing tumor, and non-enhancing tumor in real patient data) and edema parts. The proposed method was evaluated using a five-fold cross-validation fashion. All experiments were repeated five times, and the final results were reported as the mean and standard deviation of the results from the individual runs. For each run, a total of 64 images were used in training and 16 images were used for testing. Meanwhile, no joint set existed between the training and testing datasets in all experiments. To evaluate the proposed method for the testing data, the training data was used to train classifiers for different tissues. The segmentation results of the testing data were...
The classification results are defined as the number of correctly classified samples divided by the total number of samples. These two similarities are defined as follows:

\[
DS = 2\frac{|A \cap B|}{|A| + |B|}, \quad JS = \frac{|A \cap B|}{|A \cup B|}
\]

where \( A \) and \( B \) are the voxel sets of the result and ground truth, respectively. The false positive rate (FPR) and false negative rate (FNR) were also adopted to quantify the over-segmentation and under-segmentation. These two metrics are computed as follows:

\[
FPR = \frac{|A \setminus B|}{|A|}, \quad FNR = \frac{|A \cap B|}{|A|}
\]

### Parameter Optimization

The parameter settings of the proposed method were carefully considered to obtain the optimum performance during our experiments. A summary of the parameter settings used in the proposed method is listed in Table II.

In this study, we employed LAE method to achieve the linear representation coefficients to linearly reconstruct a testing sample and ensure the minimum reconstruction error. To obtain the solution of LAE, we first selected the \( k \) nearest neighbors of the testing sample from the dictionary. In this experiment, we varied \( k \) from 5 to 100 to evaluate the effects of the different numbers of \( k \) nearest neighbors on the reconstruction error. We first randomly selected 10,000 samples from the training data for each class at each level. For each class, we then computed the reconstruction error norms of the selected samples using the corresponding dictionary. We fixed the dictionary and patch sizes of each class at each level to be 40,000 and 5, respectively. The effect of the varied numbers of \( k \) nearest neighbors on the reconstruction error is shown in Fig. 6. We found that the lowest value for the reconstruction error was always obtained when \( k \) was equal to 10 for each class at each level. This result is significant in rapidly obtaining the solution of LAE.

The dictionary size \( N \) was a crucial parameter in the proposed LIPC. Therefore, we performed the proposed method uploaded to the online evaluation platform and were evaluated automatically using the online evaluation tool (http://virtualskeleton.ch/).

#### Quantitative evaluation metrics

The dice similarity (DS) and jaccard similarity (JS) were used in this study to quantitatively assess the performance of the proposed method. These two similarities are defined as follows:

\[
DS = 2\frac{|A \cap B|}{|A| + |B|}, \quad JS = \frac{|A \cap B|}{|A \cup B|}
\]

where \( A \) and \( B \) are the voxel sets of the result and ground truth, respectively. The false positive rate (FPR) and false negative rate (FNR) were also adopted to quantify the over-segmentation and under-segmentation. These two metrics are computed as follows:

\[
FPR = \frac{|A \setminus B|}{|A|}, \quad FNR = \frac{|A \cap B|}{|A|}
\]
on the brain tumor image data to compare the effect of different dictionary sizes \( N \) on the classification performance. We set \( N \) for different classes to be of the same values at different levels, and the dictionary size was denoted as \( N \). We varied \( N \) from 5,000 to 40,000. \( k \) was also set to 10 for each class at each level for different values of \( N \). The patch size for each class at each level was fixed to 5. As shown in Fig. 7, increasing \( N \) improved the classification accuracy on the different image data groups. Meanwhile, the standard deviation decreased with increasing \( N \). In general, large dictionaries contain numerous discriminative data, which result in a more accurate classification compared with small dictionaries. However, to obtain the trade-off between memory and computational costs and accuracy, we selected \( N = 40,000 \) for the following experiments.

The patch size \( w \) is also an important parameter that should be carefully determined. We set the same \( w \) for the different classes at different levels and fixed \( w \) to 3, 5, and 7. Parameters \( k \) and \( N \) were set to 10 and 40,000, respectively, for each class at each level. The classification accuracy was improved by increasing \( w \) from 3 to 5 (except for the tumor classification accuracy on real patient data with high-grade gliomas and synthetic data with high-grade gliomas), in which the average \( DS \) increased from 0.742 to 0.782 (Fig. 8). The average \( DS \) of \( w = 5 \) (0.782) was also higher than that of \( w = 7 \) (0.756). Therefore, we selected \( w = 5 \) for the following experiments.

3) Qualitative results

The qualitative results of the proposed method with the learned softmax regression model on different data groups are shown in Fig. 9. Parameters \( k, N, \) and \( w \) were set to 10, 40,000, and 5, respectively, for each class at each level. The tumor boundaries of the real patient data were more blurry than those of the synthetic data [Fig. 9 (a)]. Therefore, the tumor classification performance was better in the synthetic data than that in the real patient data [Fig. 9 (c)]. The edema boundaries of both real patient data and synthetic data were quite blurry [Fig. 9 (a)], which led to more inaccurately classified voxels in the edema regions than those in the tumor regions [Fig. 9 (b)]. Two other typical segmentation results using the proposed method are shown in Fig. 10. Fig. 10 shows that the tumor boundary of low-grade real patient data was more blurry than that of high-grade real patient data. The edema region was also blurry in low-grade real patient data. Thus, the classification accuracy of the tumor and edema in the low-grade patient data was lower than those in the high-grade patient data.

4) Effectiveness of LAE method

Numerous coding methods can be used to solve the coefficients of the dictionary. In this section, we show the comparative results of the tumor segmentation with several coding methods. The SC, LLC, and LAE coding methods were used for comparison. SC focuses on the sparse property, and LLC concentrates on the local property. In this study, LAE was used as the coding method, which emphasizes the local and

<table>
<thead>
<tr>
<th>Table III</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effect of Learning a Softmax Regression Model Using the Reconstruction Error Norm. “With” and “Without” Represent the Proposed Method With and Without Using the Learned Softmax Regression Model, Respectively.</strong></td>
</tr>
<tr>
<td><strong>High-grade (real)</strong></td>
</tr>
<tr>
<td><strong>Edema</strong></td>
</tr>
<tr>
<td><strong>DS</strong> Without</td>
</tr>
<tr>
<td>(%)</td>
</tr>
<tr>
<td><strong>JS</strong> Without</td>
</tr>
<tr>
<td>(%)</td>
</tr>
<tr>
<td><strong>FPR</strong> Without</td>
</tr>
<tr>
<td>(%)</td>
</tr>
<tr>
<td><strong>FNR</strong> Without</td>
</tr>
<tr>
<td>(%)</td>
</tr>
</tbody>
</table>

Fig. 11. Tumor segmentation results of LLC, SC, and LAE.

Fig. 12. Comparison of LIPC and SRC
non-negative properties. To obtain a fair comparison, these three coding methods were performed in the LIPC framework, and the classification scores were computed using Eq. (8). Parameters $N$ and $w$ were set to 40,000, and 5, respectively, for each class at each level. For LAE, parameter $k$ was set to 10 for each class at each level. For LLC and SC, we first conducted an experiment, which was similar to the experiment described in Section III. B. 2), to determine the number of nonzero values. We found that when the numbers of nonzero values were 90 and 15 for LLC and SC, respectively, the corresponding reconstruction error of the randomly selected samples reached the minimum value. The tumor segmentation results with different coding methods are shown in Fig. 11. LAE exhibited better robustness compared with SC and LLC. The mean $DS$ of LAE was 2.6% and 3.1% higher than that of LLC and SC, respectively. This result indicates that LAE was more applicable in addressing difficulties in brain tumor segmentation compared with SC and LLC.

5) Comparison between LIPC and SRC

To evaluate the effectiveness of LIPC, both SRC and LIPC used LAE as the coding method. Moreover, the classification scores were computed using Eq. (8). For LIPC, parameters $k$, $N$, and $w$ were set to 10, 40,000, and 5, respectively, for each class at each level. For SRC, a dictionary containing samples from three classes was constructed for each level. This dictionary consisted of three subdictionaries and each subdictionary corresponded to a class. For a fair comparison, the size of each subdictionary was set to 40,000. Therefore, the size of the dictionary for SRC was 120,000 at each level. The number of nonzero values in LAE for SRC was determined as follows. We first randomly selected 10,000 samples from the training data for each class at each level and computed the reconstruction error norms of all the selected samples using the dictionary. The number of nonzero values in LAE was varied from 5 to 1,200. Finally, the minimum reconstruction error was found when the number of nonzero values was set to 1000. After we investigated the results of different data groups, the mean $DS$ of LIPC was 5.3% higher than that of SRC. The classification results with LIPC and SRC on different data groups are displayed in Fig. 12, which shows that the proposed LIPC could be effectively used in tumor segmentation.

6) Effectiveness of softmax regression model learning

To assess the effect of learning a softmax regression model using the reconstruction error norm on the classification performance, we compared the classification results obtained from Eqs. (8) and (9). Parameters $k$, $N$, and $w$ were set to 10, 40,000, and 5, respectively, for each class at each level. As listed in Table III, the accuracy of the proposed method with the learned softmax regression model was higher than that of the proposed method without using the learned softmax regression model. The case without the learned softmax regression model had a higher $FPR$ than that with the learned softmax regression model, which indicates the former tended to retain extra background.

7) Comparison with other methods

The segmentation results of the testing data were evaluated by the online evaluation tool in the evaluation platform. This evaluation platform contains an archive of all uploaded results, which enables segmentation methods to be objectively benchmarked and compared with each other. The tumor region in the real patient data was subdivided into four classes (necrosis, contrast-enhancing tumor, non-enhancing tumor, and edema), and the evaluation was performed for three different tumor subregions: complete tumor (necrosis, contrast-enhancing tumor, non-enhancing tumor, and edema), tumor core (necrosis, contrast-enhancing tumor, and non-enhancing tumor), and contrast-enhancing tumor. The tumor region in the synthetic data was subdivided into two classes (tumor core and edema). For the real patient data, five dictionaries and five classifiers were learned from the real patient training data. For the synthetic data, three dictionaries and three classifiers were learned from the synthetic training data. For all testing data, parameters $N$, $w$, and $k$ were set to 40,000, 5, and 10 for each class at each level. Table IV shows the tumor segmentation results with different methods evaluated by the online evaluation tool. The top five results are listed. At the time of writing of this study, the proposed method was in the second and third places in the 2012 and 2013 real patient data, respectively. Moreover, the proposed method is in the first place in 2012 synthetic data. These results indicate that the proposed method is comparable to other state-of-the-art methods.

8) Computation time

In this study, all experiments were implemented on a standard PC using a single thread on an Intel Core i5-2400 processor at 3.10 GHz. In the testing step, the processing time for tumor segmentation of a subject with $N = 40,000$ was approximately 26 min, at which 2 min was used for intensity normalization and 24 min was used for segmentation. In the training step, the construction of a dictionary with $N = 40,000$
from a training dataset with 64 images took 5 h.

IV. DISCUSSION AND CONCLUSION

An automatic method is proposed for brain tumor segmentation in MRI images. An LIPC-based method was introduced to solve the tumor segmentation problem. The proposed LIPC used local independent projection into the classical classification model, and a novel classification framework was derived. Compared with other coding approaches, the LAE method was more suitable in solving the linear reconstruction weights under the locality constraint. The data distribution in each submanifold was important for the classification, and we used a softmax model to learn the relationship between the data distribution and reconstruction error norm. We evaluated the proposed method using both synthetic data and public available brain tumor image data. In both problems, our method outperformed competing methods.

A. Comparison with SRC Method

The proposed LIPC method applied each subdictionary to represent the testing sample instead of using the whole dictionary to represent the testing sample in the SRC method. The classification strategy used in LIPC was more applicable than that used in SRC, especially when the data distribution is nonuniform and the noise level is high (Fig. 4). When the data distribution of one class is sparser than that of other classes, SRC may select few nearest neighbors in the sparse class. This selection strategy may result in the wrong classification if the testing sample belongs to the sparse class. To prevent this problem, SRC tended to select more nearest neighbors than LIPC (1000 neighbors in SRC vs. 10 neighbors in LIPC as mentioned in Section III. B. 5)), which leads to increased computation costs.

B. Comparison with the State-of-the-Art in Brain Tumor Segmentation

The brain tumor segmentation problem has been widely investigated; putting our results in context helps to reveal their significance. The proposed method was compared with other methods by uploading the segmentation results to the online evaluation tool. In the proceedings of the BRATS2012, Zikic et al. [21] and Bauer et al. [23] proposed an automatic classification method, respectively, and their methods performed best on the image data. Zikic et al. [21] achieved average DS values of 0.75, 0.47, and 0.41 for complete tumor, tumor core, and contrast-enhancing tumor for real patient testing data, respectively, and 0.91 and 0.86 for complete tumor and tumor core for synthetic testing data, respectively. Bauer et al. [23] achieved average DS values of 0.75, 0.54, and 0.37 for complete tumor, tumor core, and contrast-enhancing tumor for real patient testing data, respectively, and 0.87 and 0.81 for complete tumor and tumor core for synthetic testing data, respectively. Our results show average DS values of 0.82, 0.63, and 0.44 for complete tumor, tumor core, and contrast-enhancing tumor for real patient testing data, respectively, and 0.92 and 0.88 for complete tumor and tumor core for synthetic testing data, respectively, the average of which was 6% higher than those obtained by Zikic et al. [21] and Bauer et al. [23]. In the proceedings of BRATS2013, Cordier et al. [52] proposed a tumor segmentation method by choosing similar patches from the training data and combining labels of chosen patches to be the labels of testing data. The patches selection step in [52] is similar to the locally linear representation step in our proposed method. However, our proposed method focuses not only on similarity and sparsity, but also on locality. The average DS values of [52] reported in [45] were 0.84, 0.68, and 0.65 for complete tumor, tumor core, and contrast-enhancing tumor, respectively. Our results show the average DS values of 0.86, 0.74, and 0.73 for complete tumor, tumor core, and contrast-enhancing tumor, respectively, the average of which was 5.4% higher than those obtained by Cordier et al. [52].

C. Computational Complexity

Compared with the classical classification model, our proposed method requires an additional step of constructing a dictionary for each class at each level in the training stage. This dictionary construction step is an off-line procedure; thus this step is performed only once and is used for all testing images. In the testing stage, the $k$ nearest neighbor searching and coding algorithms, which were the most time-consuming steps, were run in a single thread. However, the voxel-based implementation in this study can be parallelized and implemented to exploit the multi-core CPUs, thereby significantly decreasing processing time.

D. Estimation of the Proposed Method

In this study, we learned a softmax regression model using the reconstruction error norm to achieve the classification scores. The experimental results displayed an improvement on the classification performance using the learned softmax regression model (Figs. 4 and 5 and Table III). As shown in Table III, the classification accuracy of the proposed method with the learned softmax regression model was much higher than that of the proposed method without the learned softmax regression model for the real data groups ($p < 0.002$). However,
the classification accuracy of the proposed method with the learned softmax regression model was similar to that of the proposed method without the learned softmax regression model for the synthetic data groups ($p > 0.6$). This finding may be attributed to the more complex intensity distribution in the real data groups than that in the synthetic data groups (Fig. 13). The intensities of the three classes could be easily separated in synthetic data with high-grade gliomas, whereas the intensity distribution of the three classes largely overlapped with one another in the real data groups (Fig. 13). We used a patch-based feature, so the image intensity distribution had a strong correlation with the sample distribution. The distribution of the training data in each submanifold is an important clue for the classification task and can bring discriminative information when classifying a testing sample. Thus, the proposed method with a learned softmax regression model is more applicable for data with complex distribution than data without the learned softmax regression model.

The proposed classification method requires no explicit regularization because the patch feature contains the contextual information of a voxel in the image. The proposed method leads to a natural smoothness to the segmentation results without explicit regularization by using this contextual information. However, the patch feature may be insufficient to discriminate the brain tumor segmentation task because of the complex characteristics of brain MRI images. Other contextual features may be added in future studies to further improve the classification accuracy obtained in this study.

REFERENCES


