



# A spatially constrained asymmetric Gaussian mixture model for image segmentation

Yunjie Chen <sup>a,\*</sup>, Ning Cheng <sup>a</sup>, Mao Cai <sup>a</sup>, Chunzheng Cao <sup>a</sup>, Jianwei Yang <sup>a</sup>, Zhichao Zhang <sup>a,b</sup>

<sup>a</sup> School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China

<sup>b</sup> Faculty of Information Technology, Macau University of Science and Technology, Macau 999078, China



## ARTICLE INFO

### Article history:

Received 25 September 2020

Received in revised form 28 April 2021

Accepted 13 June 2021

Available online 16 June 2021

### Keywords:

Skew normal distribution  
Anisotropic spatial information  
Nonlocal spatial information  
EM algorithm

## ABSTRACT

The finite Gaussian mixture model (GMM) is a flexible and powerful tool for addressing many computer vision and pattern recognition problems. The Gaussian distribution is a probability distribution that is symmetric with respect to the mean. However, in many segmentation applications, the observed data obey an asymmetric distribution. Furthermore, the GMM is sensitive to imaging noise. To alleviate these issues, a new finite anisotropic asymmetric normal mixture model is presented in this paper. Note that GMM is a degraded case of our proposed model. First, the proposed model employs anisotropic spatial information to reduce the effect of imaging noise while preserving the details, such as edges, corners and slim structure objects. Second, the anisotropic spatial information is coupled into the skew normal distribution to fit the observed data obeying an asymmetric distribution. Then the modeling and estimation of the object intensity probability density function are proposed by using the anisotropic skew normal mixture model. The proposed model not only has the capability to fit the observed data obeying a non-symmetric distribution, but also can reduce the effect of noise while preserving the objects details. Finally, expectation maximization (EM) algorithm is adopted to estimate the model parameters in order to maximize the log-likelihood function. The experiment results on synthetic images and natural grayscale images demonstrate the superior performance of the proposed model compared with other state-of-the-art segmentation methods.

© 2021 Elsevier Inc. All rights reserved.

## 1. Introduction

As one of the most important and difficult tasks in image analysis and computer vision, the definition of image segmentation is the partitioning of an image into a set of non-overlapping and consistent regions that have common characteristics. Various features from the images, such as histograms, boundaries, and local patterns can be used to guide the segmentation. However, it is hard to find automated segmentation methods, due to overlapping intensities and noise.

Many image segmentation algorithms, such as thresholding algorithms [1], fuzzy clustering algorithms [2,3], finite mixture models(FMM) [4,5], have been proposed. Among them, the FMM provides a natural way to cluster the observed data based on the components of the mixture that generates the intensity distribution of the image and has been widely used for image segmentation. Among FMM-based methods, the Gaussian mixture model(GMM) [6,7] is a well-known method for image segmentation. One advantage of GMM is that it needs to learn only a small number of parameters, which can

\* Corresponding author.

E-mail address: [priestcyj@nuist.edu.cn](mailto:priestcyj@nuist.edu.cn) (Y. Chen).

be efficiently estimated by the expectation maximization(EM) algorithm [8]. However, the GMM is not flexible enough to fit an asymmetrical distribution. Moreover, the GMM uses only the intensity information, so that it is sensitive to noise without any spatial information.

In order to improve the robustness of GMM, the asymmetric distribution has been proposed in [9–11]. Allili et al. [9] used two prior probabilities of within-cluster and between-cluster to fit the asymmetric distribution. Nguyen et al. [10] proposed a generative asymmetric Gaussian mixture model, which used the prior probabilities of within-cluster to estimate the asymmetric distribution. Both methods are asymmetric, however, it is hard to determine the number of inner-class for different images [11]. Sun et al. proposed a shifted asymmetric Laplace mixture model(SALMM) [12] to describe the distribution of the observed data with an asymmetric form. Azzalini et al. [13] proposed the Skew Normal Distribution, which has a skew parameter to describe the asymmetry of the observed data. Cao et al. [14] have proposed some kinds of measurement error models based on skew normal distribution. Although the results obtained by the asymmetric methods are much more accurate than those obtained by GMM, they are sensitive to imaging noise due to ignorance of spatial information.

To reduce the effect of noise, the Markov random field (MRF) model [15,16] has been proposed to characterize spatial interaction among neighboring pixels. Diplaros et al. [17] proposed a novel mixture model and an EM algorithm for image segmentation (SCGM-EM), which assumes that the hidden class label of each pixel was generated by prior distributions of its neighboring pixels and uses a pseudo-likelihood quantity to couple neighbor priors by means of entropic quantity based on Kullback–Leibler (KL) divergence. Although the SCGM-EM can reduce the effect of noise, the Maximization-step in the EM algorithm cannot be directly applied to estimate the model parameters from the observations. In order to deal with this problem, Nguyen and Wu [18] proposed a fast and robust spatially constrained GMM (FRSCGMM) by incorporating spatial relationships among neighboring pixels in a simpler metric and directly optimized through the EM algorithm. Ji et al. [19] proposed a rough set bounded spatially constrained asymmetric Gaussian Mixture Model (SCAGMM) by incorporating the spatial information amongst neighborhood pixels to construct a spatial factor, which combines the posterior probabilities, prior probabilities and the spatial direction. SCAGMM can reduce the effect of noise; however, it uses only the information of pixels on one direction of four directions in neighbors, and each pixel on the direction has same weights, so the method is hard to preserve details of tissues when segmenting objects with weak edges. Can H et al. [4] introduced a Bayesian statistical method and integrated the spatially constrained in finite Dirichlet mixture model (SC-DMM) for image segmentation. Although these models can reduce the effect of noise, the spatial information used in these methods is isotropic, preserving detail information difficulty. Moreover, deep learning has been extensively studied in image segmentation and has produced very good results. U-net [20] based technique also has been suggested to segment images, due to the higher representation capability and calculated effectively.

Motivated by the aforementioned observations, we propose an anisotropic skew normal mixture model for image segmentation. We first apply the structure information of the image to construct an anisotropic MRF, which is isotropic in homogeneous regions and anisotropic in edge regions. We then utilize the anisotropic MRF to improve the skew normal mixture model with the purpose of reducing the effect of noise. Also, we provide EM-type algorithms for iteratively computing maximum likelihood estimation. The experiments on synthetic images and natural grayscale images illustrate the superior performance of the proposed algorithm compared with other state-of-the-art segmentation methods.

The main contributions of our work can be summarized as follows: (1) We extend the Gaussian mixture model to the Skew Gaussian mixture model, which can fit the intensity distributions of the images with asymmetric forms. (2) We propose a novel anisotropic spatial factor, which is based on the structure tensor and the nonlocal information. The factor is isotropic in inner regions and anisotropic in edge regions. We coupled the factor to the Skew Gaussian mixture model, which makes the method can preserve more details. (3) The posterior probabilities and prior probabilities have been taken into account to improve the robustness of the model. Extensive experiments on brain MR images and natural images validate the superior capabilities of the proposed spatially asymmetric Gaussian mixture model for image segmentation. The code can be downloaded at <https://github.com/priestcyj/Yunjie-Chen>.

The paper is organized as follows. Section 2 provides a brief introduction of the related works. Section 3 describes the proposed method and also provides the parameter estimation. Section 4 demonstrates numerical experiments on synthetic and natural images. Moreover, the important aspects and limitations of the proposed method are also discussed in this Section. Finally, Section 5 summarizes our findings.

## 2. Related work

### 2.1. Finite mixture model

Let  $x_i$ , with a dimension  $\mathcal{D}$ ,  $i = (1, 2, \dots, N)$ , be an observation at the  $i$  th pixel of a to-be-segmented image  $I$ . The objective is to segment the image  $I$  into  $K$  non-overlapping regions denoted by  $(\Omega_1, \Omega_2, \dots, \Omega_k)$ . Then, the density function is denoted as

$$f(\mathbf{x}_i|\boldsymbol{\Pi}, \boldsymbol{\Theta}) = \sum_{k=1}^K \pi_k p(\mathbf{x}_i|\boldsymbol{\Theta}_k) \quad (1)$$

where  $\boldsymbol{\Pi} = \{\pi_1, \dots, \pi_K\}$  is the set of prior probabilities for the pixel  $x_i$  and satisfies:  $0 < \pi_k < 1$  and  $\sum_{k=1}^K \pi_k = 1$ .

When  $p(\mathbf{x}_i|\Theta_k)$  is a normal distribution, with a small number of parameters  $\mu_k, \Sigma_k$  in the learning process, the model is called Gaussian mixture model, which is widely used in image segmentation. However, it cannot perform well when the distribution of the data is not asymmetric. Besides, it is sensitive to noise without considering any spatial information.

## 2.2. Spatial information based on MRF

In order to reduce the effect of noise, the Markov random field (MRF) has been widely utilized into FMM. Compared to the traditional FMM, the MRF based methods use a set of prior distributions to model the probability of each pixel. Since each pixel is modeled as statistically independent, the joint conditional density over the whole image can be modeled as:

$$f(\mathbf{x}|\Pi, \Theta) = \prod_{i=1}^N \sum_{k=1}^K \pi_{ik} p(\mathbf{x}_i|\Theta_k) \quad (2)$$

where  $\pi_{ik}$  satisfies:  $0 < \pi_{ik} < 1$  and  $\sum_{k=1}^K \pi_{ik} = 1$ . In order to improve the robustness to the noise, the MRF distribution incorporates the spatial relationship of  $x_i$  amongst its neighborhood region  $\partial_i$ . Based on Hamersley–Clifford theory, MRF can be expressed by Gibbs random field[19]:

$$P(\Pi) = \frac{\exp\{-U(\Pi)\}}{Z} \quad (3)$$

where  $U(\Pi)$  is the smoothing prior and  $Z$  is a normalized constant.

There are many types of  $P(\Pi)$  that have been proposed to utilize the spatial information. Hao et al. [12] introduced shifted asymmetric Laplace distribution (SALMM) to calculate the priori probability:

$$\pi_{ik} = \frac{w_k(\mathbf{x}_i)}{\sum_{j=1}^K w_j(\mathbf{x}_i)}, w_k(\mathbf{x}_i) = \left[ \exp \left( \sum_{m \in \partial_i} SAL(\mathbf{x}_m|\Theta_k) \right) \right]^\beta \quad (4)$$

where  $\beta$  is a smoothing parameter. It can be found that the weight factor used the neighbor information about the shifted asymmetric laplace (SAL) distribution in the mixture model to reduce the effect noise. However, the neighbor region is isotropic, which makes the method hard to preserve detail information.

Ji et al. [11] proposed a spatially constrained generative asymmetric Gaussian mixture model (SCGAGMM) for image segmentation. The SCGAGMM used the Kullback–Leibler (KL) divergence to define  $P(\Pi)$ :

$$P(\Pi) = \prod_{i=1}^N p(\pi_i|\pi_{\partial_i}) \quad (5)$$

where  $\pi_{\partial_i}$  is the weighted sum of prior probability of the neighboring pixels around  $x_i$ :

$$\pi_{\partial_i} = \sum_{j \in \partial_i, j \neq i} \alpha_{ij} \pi_j \quad (6)$$

and  $\alpha_{ij}$  is the weight of neighboring pixels

$$\alpha_{ij} = \frac{\exp\left(-\sum_{m \in \partial_i, m \neq i, j} |\mathbf{x}_j - \mathbf{x}_m| / (N_{\partial_i} \times (N_{\partial_i} - 1))\right)}{\sum_{h \in \partial_i} \exp\left(-\sum_{m \in \partial_i, m \neq i, h} |\mathbf{x}_h - \mathbf{x}_m| / (N_{\partial_i} \times (N_{\partial_i} - 1))\right)} \quad (7)$$

where  $\partial_i$  is the square neighborhood region centered at pixel  $x_i$ , with the radius is  $R$ . Then  $N_{\partial_i}$  is the number of pixels in  $\partial_i$ . The value of  $\alpha_{ij}$  is small when the intensity of the neighboring pixel  $j$  is quite different with other neighboring pixels in  $\partial_i$ , indicating that a noise pixel has less effect with a tiny weight. However, the neighborhood region is isotropic, which makes that when the neighboring object pixels are located in the region with slim structures, these neighboring object pixels have smaller weights and makes the method hard to preserve the detail information. For the expression of  $p(\pi_i|\pi_{\partial_i})$ , the  $\log p(\pi_i|\pi_{\partial_i})$  is defined as:

$$\log p(\pi_i|\pi_{\partial_i}) = -\beta [K(\pi_i|\pi_{\partial_i}) + H(\pi_i)] \quad (8)$$

where  $K(\pi_i|\pi_{\partial_i})$  is the KL divergence:  $K(\pi_i|\pi_{\partial_i}) = \sum_{k=1}^K \pi_{ik} \log \pi_{ik} - \sum_{k=1}^K \pi_{ik} \log \pi_{\partial_ik}$ , which is always nonnegative and becomes zero when  $\pi_{ik} = \pi_{\partial_ik}$ .  $H(\cdot)$  is the entropy function:  $H(\pi_i) = -\sum_{k=1}^K \pi_{ik} \log \pi_{ik}$ . By using the spatial information, the proposed weight in SCGAGMM can reduce the effect of noise; however, the spatial weight only considered the intensity distance without any spatial distance so that the neighborhood region is still isotropic and then it is hard to preserve the details.

Hu et al. [4] also considered the effect of noise in finite Dirichlet mixture model (SC-DMM) and proposed a priori probability:

$$\pi_{ik} = \frac{Z_{ik} + a_k^2 \bar{Z}_{ik}^b}{\sum_{k=1}^K (Z_{ik} + a_k^2 \bar{Z}_{ik}^b)}, \bar{Z}_{ik} = \frac{\sum_{m \in \partial_i} Z_{mk}}{N_{\partial_i}} \quad (9)$$

where  $Z_{ik} = \frac{\exp \left( \ln \pi_{ik} + \ln \frac{\Gamma \left( \sum_{d=1}^D z_{kd} \right)}{\prod_{d=1}^D \Gamma(z_{kd})} + \sum_{d=1}^D (\alpha_{kd}-1) \ln X_{id} \right)}{\sum_{k=1}^K \exp \left( \ln \pi_{jk} + \ln \frac{\Gamma \left( \sum_{d=1}^D z_{kd} \right)}{\prod_{d=1}^D \Gamma(z_{kd})} + \sum_{d=1}^D (\alpha_{kd}-1) \ln X_{id} \right)}$  is the posterior probability estimated by variational Bayes algorithm.

$a_k$  and  $b$  are constants.  $\alpha_{kd}$  is an estimated parameter belongs to the  $k$ -th class and  $d$ -th dimension. The neighborhood region  $\partial_i$  is a  $3 \times 3$  square window [4].  $N_{\partial_i}$  is the number of pixels in  $\partial_i$ . From Eq. (9), we can find that  $\pi_{ik}$  is defined based on the posterior probability  $Z_{ik}$  and the average posterior probability  $\bar{Z}_{ik}$  to reduce the noise. However, the weights of neighboring pixels are still isotropic, which is hard to preserve the detail information of the regions with corners or slim structures.

Fig. 1 shows the performances of different spatial information mentioned above. Fig. 1(a) is a  $5 \times 5$  square window of an observed image with slim structures. The digit of each pixel is the normalized gray value. Fig. 1(b)–(d) show the weight factors in a  $3 \times 3$  neighborhood region centered at the pixel  $A$  by using Eq. (4) [12], Eq. (7) [11] and Eq. (9) [4], respectively. Pixels  $A$ ,  $C1$  and  $C2$  belong to the object with a slim structure. Pixels  $B1, B2, B3$  and  $D1, D2, D3$  belong to the background. The weights of  $C1$  and  $C2$  should be larger than those of the background points, otherwise, the pixel  $A$  will be more affected by the background points. The results of Eq. (4) shows that the weights of the objects  $C$  are same with the background points in the same column. The weights of the object points  $C$  in the result of Eq. (7) are only little bit larger than those of the background points. In the result of Eq. (9), the weight of  $C2$  is even smaller than those of the background points. All these methods are based on isotropic neighborhood information, which makes the weights of these methods are hard to preserve detail information. The result of our anisotropic spatial information, which will be illustrated in Section 3.1, is shown in Fig. 1(e). The weights of object points  $C$  are larger than those of background points.

### 2.3. Asymmetric mixture models

In many applications, the observed data obey an asymmetric distribution. The symmetric mixture models, such as GMM, are hard to fit this asymmetric distribution. In order to alleviate this shortcoming, Nguyen et al. [10] improved the Gaussian mixture model by using between-cluster  $\pi_{ik}$  and within-cluster  $\phi_{ikl}$ . The probability density function (pdf) is given as:

$$f(\mathbf{x}_i | \Pi, \Psi, \Theta) = \sum_{k=1}^K \pi_{ik} \left( \sum_{l=1}^L \phi_{ikl} \Phi(\mathbf{x}_i | \boldsymbol{\mu}_{kl}, \boldsymbol{\Sigma}_{kl}) \right) \quad (10)$$

where  $K$  is the number of between-clusters and  $L$  is the number of within-clusters.  $\Phi$  is the Gaussian pdf.  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are the mean and covariance, respectively. The hierarchical density function is asymmetric, however, choosing a favorable number of within-cluster in these kinds of models is always difficult for different images. The similar asymmetric distribution can be found in [11].

Sun et al. [12] proposed a shifted asymmetric Laplace mixture model (SALMM). The pdf of the mixture model is defined as:

$$p(x_i | \theta_k) = \sqrt{\frac{2}{\pi \sigma_k}} \exp \left( \frac{\alpha_k (x_i - \mu_k)}{\sigma_k} \right) \times \left( \frac{(x_i - \mu_k)^2}{2\sigma_k + \alpha_k^2} \right)^{0.25} K_{0.5} \left( \sqrt{\frac{(2\sigma_k + \alpha_k^2)(x_i - \mu_k)^2}{\sigma_k^2}} \right) \quad (11)$$

where  $K_{0.5}(\cdot)$  is the modified Bessel function with an index 0.5.  $\alpha_k$  can be regarded as a skewness parameter. When  $\alpha_k > 0$ , the distribution is skewed to left, and when  $\alpha_k < 0$ , the distribution is skewed to right. When  $\alpha_k = 0$ , the distribution reduces to

0.1	0.1	0.1	0.1	0.1	B1 0.0953	B2 0.1482	B3 0.1482	B1 0.1222	B2 0.1222	B3 0.1222	B1 0.1429	B2 0.1429	B3 0.1428	B1 0.00066	B2 0.00019	B3 0.00019
0.1	B1 0.1	B2 0.1	B3 0.1	0.1	C1 0.0953	A	C2 0.1482	C1 0.1334	A	C2 0.1334	C1 0.1429	A	C2 0.0001	C1 0.2223	A	C2 0.77569
0.1	C1 0.8	A 0.8	C2 0.8	0.8	D1 0.0953	D2 0.1482	D3 0.1482	D1 0.1222	D2 0.1222	D3 0.1222	D1 0.1429	D2 0.1429	D3 0.1428	D1 0.00066	D2 0.00019	D3 0.00019
0.1	D1 0.1	D2 0.1	D3 0.1	0.1												
0.1	0.1	0.1	0.1	0.1												

(a)

(b)

(c)

(d)

(e)

Fig. 1. Different weight factors based on different spatial information. (a) Observed data, (b) Weight factors of Eq. (4)  $\beta = 40$ , (c) Weight factors of Eq. (7), (d) Weight factors of Eq. (9)  $a = 50, b = 1.5$ , (e) Weight factors of our method.

symmetric Laplace distribution. The proposed pdf can achieve satisfactory results when the images have asymmetric distributions. However, the pdf is not smooth and non-convex so that the expected values in the EM algorithm hard to find accurate results [21].

The spatially constrained Dirichlet mixture model introduced in [4] also has the asymmetric form, in which the pdf is defined by:

$$f(\mathbf{X}_i|\pi_i, \alpha_k) = \sum_{k=1}^K \pi_{ik} \left( \frac{\Gamma(\sum_{d=1}^D)}{\prod_{d=1}^D \Gamma(\alpha_{kd})} \prod_{d=1}^D X_{id}^{\alpha_{kd}-1} \right) \quad (12)$$

where  $\pi_{ik}$  is the priori probability,  $\mathbf{X}_i$  is a D-dimensional random variable, and  $\alpha_{kd}$  is the asymmetric parameter. Note that different component values in  $\alpha_k = [\alpha_{k1}, \alpha_{k2}, \dots, \alpha_{kD}]$  will lead to asymmetric form for the probability density function in Eq. (12). The variational Bayes algorithm is applied for parameter estimation in this model to alleviate the shortcomings of the traditional EM algorithms, which often get trapped into a local optimum. Especially, in the presence of image noise, the hyperparameters are sensitive to the initialization.

As proposed by Azzalini [22],  $\mathbf{x}_i$  follows a skew normal distribution with location vector  $\boldsymbol{\mu} \in R^D$ , dispersion matrix  $\Sigma \in R^{D \times D}$  and skewness vector  $\lambda \in R^D$ , if its probability density function is given as:

$$p(\mathbf{x}|\boldsymbol{\mu}, \Sigma) = 2\varphi_n(\mathbf{x}|\boldsymbol{\mu}, \Sigma)\Phi_1(\lambda^T \Sigma^{-1/2}(\mathbf{x} - \boldsymbol{\mu})) \quad (13)$$

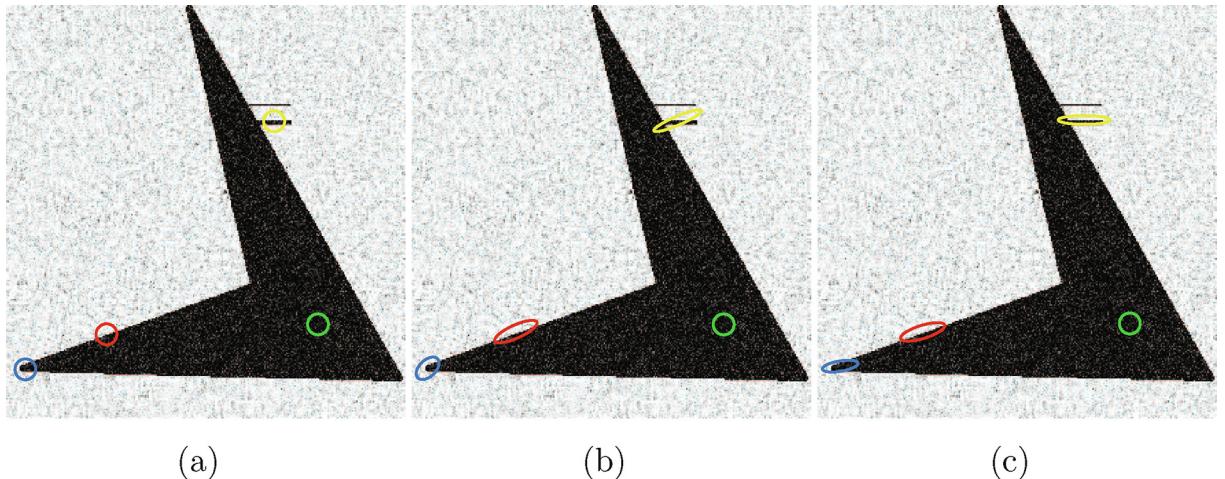
where  $\varphi_n(\mathbf{x}|\boldsymbol{\mu}, \Sigma)$  and  $\Phi_1(\lambda^T \Sigma^{-1/2}(\mathbf{x} - \boldsymbol{\mu}))$  are the pdf and the cumulative distribution function (cdf), respectively. When  $\lambda > 0$ , the distribution is skewed to right, and when  $\lambda < 0$ , the distribution is skewed to left. When  $\lambda = 0$ , the skew normal distribution reduces to the Gaussian distribution. Hence the skew normal distribution is more powerful and flexible for probabilistic data clustering. Then, the corresponding mixture model density function is given as:

$$f(\mathbf{x}_i|\Pi, \Theta) = \sum_{k=1}^K \pi_k p(\mathbf{x}_i|\boldsymbol{\mu}_k, \Sigma_k, \lambda_k) \quad (14)$$

### 3. Proposed model

#### 3.1. Anisotropic spatial information

As we analyzed above, the neighborhood region is usually a circular region or square region, which is isotropic so that the spatial information is difficult to preserve details. Fig. 2 illustrates the effect of isotropic neighborhood region and anisotropic neighborhood region. Fig. 2(a) shows an image with noise. The red region, blue region and green region are centered at edge point, corner point and inner point, respectively. The yellow region is a neighborhood region centered at a pixel located at the object region with a slim structure. In the red region, the object and background contain a similar proportion of pixels. In



**Fig. 2.** The isotropic and anisotropic neighborhood regions. (a) The isotropic neighborhood region. (b) The anisotropic neighborhood region without considering nonlocal information. (c) our anisotropic neighborhood region.

the blue region and yellow region, the neighboring pixels are dominated by the background. When the intensities of object pixels are similar to those of background, the isotropic neighborhood region cannot preserve the object details.

In order to preserve more detail information, we proposed a robust spatial factor  $\alpha_{ij}$ , which is based on structure tensor and nonlocal patch information:

$$\alpha_{ij} = \frac{G(\mathbf{x}_i, \mathbf{x}_j)}{\sum_{k=1}^{N_i} G(\mathbf{x}_i, \mathbf{x}_k)} \quad (15)$$

where  $N_i$  is the number of neighborhood pixels around  $\mathbf{x}_i$  and  $G(\cdot, \cdot)$  is an anisotropic Gaussian kernel function.

The structure tensor [24] has been proposed to describe the structure of the object. The traditional structure tensor is defined as:  $\mathbf{T} = \nabla I \nabla I^T$ . For a gray value image, the gradient  $\nabla I = [I_X, I_Y]^T$  and the structure tensor is written as:

$$\mathbf{T} = \begin{bmatrix} I_X^2 & I_X I_Y \\ I_Y I_X & I_Y^2 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{12} & T_{22} \end{bmatrix} \quad (16)$$

It can be found that  $\mathbf{T}$  is a positive definite matrix with the rank as one. The eigenvectors  $\mu^+$  and  $\mu^-$  of the tensor  $\mathbf{T}$  are directed along the local maximum and minimum variations of the image at each point, namely, the normal direction and tangent direction in local region. The eigenvalues  $\lambda^+$  and  $\lambda^-$  measure the variations along  $\mu^+$  and  $\mu^-$  and can be calculated as:

$$\begin{aligned} \lambda^+ &= \frac{1}{2} \left( T_{11} + T_{22} + \sqrt{(T_{11} - T_{22})^2 + 4T_{12}^2} \right) \\ \lambda^- &= \frac{1}{2} \left( T_{11} + T_{22} - \sqrt{(T_{11} - T_{22})^2 + 4T_{12}^2} \right) \end{aligned} \quad (17)$$

the corresponding eigenvector are:

$$\mu^+ = [\cos\theta \quad \sin\theta]^T; \mu^- = [-\sin\theta \quad \cos\theta]^T \quad (18)$$

where  $\cos\theta = 2T_{12}/T_{norm}$ ,  $\sin\theta = (T_{22} - T_{11} + \sqrt{(T_{11} - T_{22})^2 + 4T_{12}^2})/T_{norm}$ ,  $T_{norm} = \sqrt{(2T_{12})^2 + (T_{22} - T_{11} + \sqrt{(T_{11} - T_{22})^2 + 4T_{12}^2})^2}$ .

The rank of  $\mathbf{T}$  is one, which means that the eigenvalue in the tangent direction is zero. In order to deal with this problem, some methods use a smoothed version to retrieve a more coherent geometry. The smoothed version is defined as:  $\mathbf{T}_\sigma = G_\sigma * \mathbf{T}$ .

Once the local geometry  $\mathbf{T}_\sigma$  has been determined, Weickert [24] proposed a method to design a particular diffusion tensor  $D$ , which depends on the eigenvalues of  $\mathbf{T}_\sigma$  and their corresponding eigenvectors:

$$D = c_1 \mu^+(\mu^+)^T + c_2 \mu^-(\mu^-)^T \quad (19)$$

Basically,  $c_1$  and  $c_2$  are designed to control the strength along the respective directions  $\mu^+$  and  $\mu^-$ . The structure  $D$  gives an ellipse, whose major axis is the same as the tangent direction of the current point and the minor axis is the same with the normal direction. In the anisotropic region, the background pixels have little influence on the object pixels. However, the diffusion tensor is still based on gradient information of each pixel, which makes the method sensitive to noise.

Following the idea shown in [24], we proposed a novel nonlocal information based structure tensor framework to reduce the effect of noise. The improved structure tensor is defined as:

$$D = f_1(\hat{\lambda}^+, \hat{\lambda}^-) \hat{\mu}^+(\hat{\mu}^+)^T + f_2(\hat{\lambda}^+, \hat{\lambda}^-) \hat{\mu}^-(\hat{\mu}^-)^T \quad (20)$$

For a pixel  $i$ , the estimated eigenvector is computed as a weighted average of the eigenvectors of pixels in the neighborhood region centered at  $i$ .

$$\hat{\mu}^+ = \sum_{j=1}^{\|N_i\|} w_{ij} \mu_j^+, \quad \hat{\mu}^- \perp \hat{\mu}^+ \quad (21)$$

where  $w_{ij}$  is the weight, which depends on the similarity between the pixels  $i$  and  $j$ , and satisfy the usual conditions  $0 \leq w(i,j) \leq 1$  and  $\sum_j w(i,j) = 1$ . The similarity between two pixels  $i$  and  $j$  depends on the similarity of  $N_i$  and  $N_j$ , where  $N_k$  denotes a square neighborhood of fixed size and centered at a pixel  $k$ . Then, the weights are defined as:

$$w_{ij} = \exp\left(-\frac{\|N_i - N_k\|_F^2}{h}\right) / \sum_{k=1}^{\|N_i\|} \exp\left(-\frac{\|N_i - N_k\|_F^2}{h}\right). \quad (22)$$

where  $\|N_i\|$  is the number of pixels in  $N_i$  and  $h$  is a constant, which depends on the standard deviation of the noise. Similarly, the estimated eigenvalues are computed as:

$$\widehat{\lambda}^+ = \sum_{j=1}^{\|N_i\|} w_{ij} \lambda_j^+, \quad \widehat{\lambda}^- = \sum_{j=1}^{\|N_i\|} w_{ij} \lambda_j^-, \quad (23)$$

The pixels with similar patch information in the neighborhood region around  $i$  have higher weights than those with different patch information. Then the adapted eigenvectors are less sensitive to the noise and contain more accurate structure information.

The structure should be isotropic in inner regions and anisotropic in edge regions. Then, for an edge point,  $f_1$  and  $f_2$  should satisfy:  $f_1(\widehat{\lambda}^+, \widehat{\lambda}^-) \ll f_2(\widehat{\lambda}^+, \widehat{\lambda}^-)$ . For a point in homogeneous region, we have  $f_1(\widehat{\lambda}^+, \widehat{\lambda}^-) \approx f_2(\widehat{\lambda}^+, \widehat{\lambda}^-)$ . In this paper, we set  $f_1$  and  $f_2$  as:

$$f_1(\widehat{\lambda}^+, \widehat{\lambda}^-) = 1/(\gamma + \widehat{\lambda}^+ + \widehat{\lambda}^-)^2, \quad f_2(\widehat{\lambda}^+, \widehat{\lambda}^-) = 1/\gamma^2 \quad (24)$$

where  $\gamma$  is a constant. Then,  $G(\mathbf{x}_i, \mathbf{x}_j)$  in Eq. (15) can be written as:

$$G(\mathbf{x}_i, \mathbf{x}_j) = \exp(-(X_i - X_j)^T D_i(X_i - X_j)) \times w_{ij} \quad (25)$$

When  $D_i$  is a unit diagonal matrix and  $w_{ij} \equiv 1$ , Eq. (21) is an isotropic distance measurement.

The diffusion tensor can describe the local edge information easily, however, it is hard to represent corner region information. To deal with this problem, we use  $S = (\widehat{\lambda}^+ \times \widehat{\lambda}^-)/(\widehat{\lambda}^+ + \widehat{\lambda}^-)$  to determine whether it is located in corner regions. When the pixel is located in the corner region,  $S$  is much larger than that of the edge region and we set  $G(\mathbf{x}_i, \mathbf{x}_j) = w_{ij}$ . It can be also found that the single points have much larger or less intensity values than those of the neighborhood points, and the end point had at least one or two neighboring pixels with similar intensity values. Based on this assumption, we regard  $\mathbf{x}_i$  as a single point when  $1/\left(\int_{j \in \{N_i\}} \exp(-|\mathbf{x}_i - \mathbf{x}_j|^2/\zeta) dy\right) > T$ .  $T$  is a non-negative parameter and set as 3/4 in this paper and  $\zeta$  is a positive constant, which is depended on the standard deviation of the noise [25]. In this case, we set  $w_{ii} = 0$ .

Fig. 2(b) shows the results by using the traditional structure tensor. From the result, we can find that the eigenvectors of the traditional structure tensor are inaccurate in corner regions and slim structure regions. Compared with the traditional structure tensor, we used the nonlocal information to reduce the effect of noise and find the exact tangent and normal directions. Based on the exact directions, we propose a framework to construct a novel structure tensor, which is isotropic in inner objects and anisotropic in edge regions. Then, based on the improved structure tensor, the proposed spatial factor can preserve more detail information. Fig. 2(c) shows the neighborhood regions by using the proposed method. Our method performs various ellipses in edges regions, flat area regions, corners regions and slim structure regions, proving that our spatial information is anisotropic and robust to preserve more detail information.

### 3.2. Anisotropic skew normal mixture model

According to the analysis of asymmetric mixture models in Section 2.3, the skew normal mixture model can better fit the intensity distributions of images with asymmetric forms. The skew normal mixture model dose not need any additional pre-defined parameters, such as the number of within-clusters. Thus, we extend the Normal Mixture model to the Skew Normal Mixture model. According to the Bayesian rules, the posterior probability density function can be written as:

$$P(\boldsymbol{\Pi}, \boldsymbol{\Theta} | \mathbf{X}) \propto P(\mathbf{X}, \boldsymbol{\Pi} | \boldsymbol{\Theta}) P(\boldsymbol{\Pi}) \quad (26)$$

Then, the log density function can be derived as:

$$L(\boldsymbol{\Pi}, \boldsymbol{\Theta} | \mathbf{X}) = \log(P(\mathbf{X}, \boldsymbol{\Pi} | \boldsymbol{\Theta})) = \sum_{i=1}^N \log\left(\sum_{k=1}^K \pi_{ik} p(\mathbf{x}_i | \boldsymbol{\theta}_k)\right) + \log(p(\boldsymbol{\Pi})) \quad (27)$$

where  $P(\boldsymbol{\Pi}) = \prod_{i=1}^N p(\pi_i | \pi_{\partial_i})$ . Following the idea in [17], we introduce an approximation that uses an auxiliary set of distribution  $s_i$  to facilitate the optimization:

$$\log(p(\pi_i | \pi_{\partial_i}, s_i)) = -\beta[K(s_i | \pi_i) + K(s_i | \pi_{\partial_i}) + H(s_i)] \quad (28)$$

where  $K(s_i | \pi_i) = \sum_{k=1}^K s_{ik} \log s_{ik} - \sum_{k=1}^K s_{ik} \log \pi_{ik}$  denotes the KL divergence between  $s_i$  and  $\pi_{ik}$ .  $K(s_i | \pi_{\partial_i}) = \sum_{k=1}^K s_{ik} \log s_{ik} - \sum_{k=1}^K s_{ik} \log \pi_{\partial_i}$  is the KL divergence between  $s_i$  and  $\pi_{\partial_i}$ , which is always nonnegative and becomes zero when  $s_i = \pi_{ik}$  and  $s_i = \pi_{\partial_i}$ .  $H(s_i) = -\sum_{k=1}^K s_{ik} \log s_{ik}$  is the entropy of the distribution  $s_i$ . When  $s_i = \pi_i$ , the above approximation equals to Eq. (8). Eq. (26) only considers the prior information without any information from the observed data. To consider the information from the observed data, we use an additional penalty based on posterior distribution:

$$\log(p(\mathbf{Z})) = -\frac{1}{2} [K(q_i | z_i) + K(q_i | z_{\partial_i}) + H(q_i)] \quad (29)$$

where the coefficient in the penalty term is set as 1/4 because it allows a tractable M-step [17].  $q_i$  is an auxiliary class distribution for a pixel  $i$ . The posterior class distribution  $z_i$  is defined as:

$$z_{ik} = \frac{\pi_{ik} p(\mathbf{x}_i | \boldsymbol{\theta}_k)}{\sum_{k=1}^K \pi_{ik} p(\mathbf{x}_i | \boldsymbol{\theta}_k)} \quad (30)$$

The total log-likelihood function is written as:

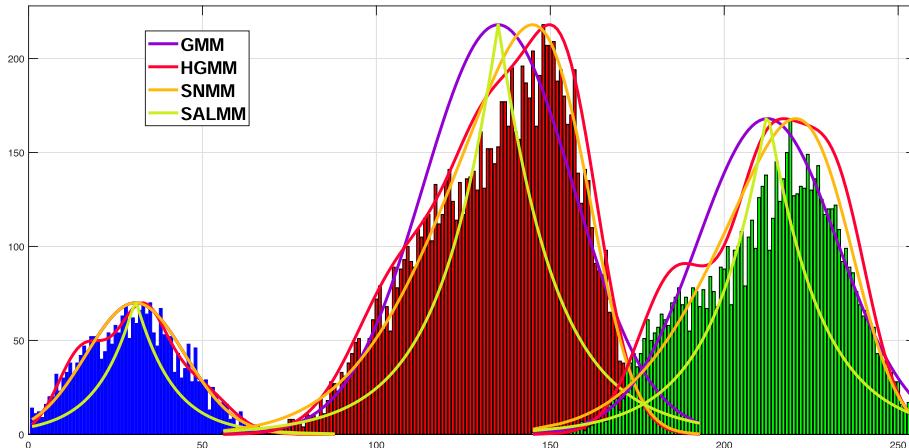
$$L(\boldsymbol{\Pi}, \boldsymbol{\Theta} | \mathbf{X}) = \sum_{i=1}^N \log \left( \sum_{k=1}^K \pi_{ik} p(\mathbf{x}_i | \boldsymbol{\theta}_k) \right) - \beta [K(s_i | \pi_i) + K(s_i | \pi_{\partial_i}) + H(s_i)] - \kappa [K(q_i | z_i) + K(q_i | z_{\partial_i}) + H(q_i)] \quad (31)$$

where the coefficient  $\kappa$  in the penalty term is set as 1/2. Then, we utilize EM algorithm to maximize the energy  $L(\boldsymbol{\Pi}, \boldsymbol{\Theta} | \mathbf{X})$ . In the E-step, we fix  $\boldsymbol{\Pi}$  and  $\boldsymbol{\Theta}$  to maximize  $L(\boldsymbol{\Pi}, \boldsymbol{\Theta} | \mathbf{X})$  over  $s$ . And in the M-step, we fix  $s$  to maximize  $L(\boldsymbol{\Pi}, \boldsymbol{\Theta} | \mathbf{X})$  over  $\boldsymbol{\Pi}$  and  $\boldsymbol{\Theta}$ .

We use a brain MR image data set, which originates from Internet Brain Segmentation Repository (IBSR2\_02), to illustrate the performance of the above Asymmetric Mixture Models. The data set has three labels ( $K = 3$ ) for white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF). From the distributions of the data set shown in Fig. 3, we can find that the distributions have asymmetric forms for WM and GM. We use the goodness-of-fit statistic value  $\chi^2$  [23] to measure the fitting accuracy of each model. The lower values indicate better results. The SALMM [12] is based on the Laplace distribution, and then it is hard to find a satisfactory result. The Gaussian mixture model can obtain a better result; however, it is hard to fit the distribution with asymmetric form. The hierarchical GMM (HGMM) [10,11] performs better by using hierarchical Gaussian distribution. The Skew Normal mixture model (SNMM) can obtain the best result. The distribution of CSF has a symmetric form, the SNMM has the same result with that of GMM. The HGMM is poor when estimating the distributions with symmetric forms.

**Remark 1.** The Spatial factor in SCGAGMM [11], SC-DMM [4], SALMM [12] are defined in isotropic region which is hard to preserve detail information. In our method, we propose a novel factor  $\alpha_{ij}$  based on the improved structure tensor and nonlocal information. Compared with the traditional structure tensor, our proposed method uses the nonlocal information to reduce the effect of noise and find the exact tangent and normal directions. Based on the exact directions, we propose a framework to construct the novel structure tensor, which is isotropic in inner regions and anisotropic in edge regions and can preserve more detail information.

**Remark 2.** The SCGA-EM [17] and FRSCGMM [18] are both based on Gaussian Mixture Model. Although Gaussian Mixture Model needs to learn only a small number of parameters, it is not suitable enough to fit asymmetric distributions. The SCAGMM [19] and SCGAGMM [11] are based on Hierarchical Gaussian Mixture models [10], which can fit asymmetric distributions better. However, the number of inner-class is hard to choose for different images. Our method is improved from these two aspects. We extend the Normal Mixture model to Skew Normal Mixture model, which fits asymmetric distributions by using a skew parameter  $\lambda$ . The skew parameter can be calculated by using EM method shown in Section 3.3 automatically.



**Fig. 3.** Estimated distributions. GMM ( $\chi^2 = 1603.0$ ). HGMM ( $\chi^2 = 968.3235$ ). SALMM ( $\chi^2 = 9627.8$ ). SNMM ( $\chi^2 = 869.1624$ ).

**Remark 3.** We coupled the anisotropic spatial factor into skew Normal Mixture model to preserve more detail information and reduce the effect of noise. Furthermore, the posterior probabilities and prior probabilities have been taken into account to improve robustness of the proposed model.

### 3.3. Parameter learning

As shown in [26], if  $X$  follows the skew normal distribution, it can be represented as:

$$\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta} |T_0| + \boldsymbol{\Sigma}^{1/2} (\mathbf{I}_n - \boldsymbol{\delta} \boldsymbol{\delta}^T)^{1/2} \mathbf{T}_1 \quad (32)$$

where  $\boldsymbol{\delta} = \lambda / \sqrt{1 + \lambda^T \lambda}$ ,  $T_0 \sim N_1(0, 1)$  that are independent of  $\mathbf{T}_1 \sim N_n(\mathbf{0}, \mathbf{I}_n)$ . Let  $\tau = |T_0|$ , we can find that  $\tau$  follows a standard half-normal distribution, denoted by  $\tau \sim HN(0, 1)$ . Then  $X$  can be written as a two-level hierarchical representation:

$$\begin{aligned} \mathbf{X} | \tau &\sim N_n(\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta} \tau, \boldsymbol{\Sigma}^{1/2} (\mathbf{I}_n - \boldsymbol{\delta} \boldsymbol{\delta}^T)^{1/2} \boldsymbol{\Sigma}^{1/2}) \\ \tau &\sim HN_1(0, 1) \end{aligned} \quad (33)$$

Then, the first term in Eq. (31) can be written as:

$$\begin{aligned} &\sum_{i=1}^N \left\{ \log \sum_{k=1}^K \pi_{ik} p(\mathbf{x}_i | \theta_k) \right\} \\ &= \sum_{i=1}^N \left\{ \log \sum_{k=1}^K \left\{ \pi_{ik} \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}_k^{1/2} (\mathbf{I}_n - \boldsymbol{\delta}_k \boldsymbol{\delta}_k^T) \boldsymbol{\Sigma}_k^{1/2}|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_i - (\boldsymbol{\mu}_k + \boldsymbol{\Sigma}_k^{1/2} \boldsymbol{\delta}_k t_i))^T (\boldsymbol{\Sigma}_k^{1/2} (\mathbf{I}_n - \boldsymbol{\delta}_k \boldsymbol{\delta}_k^T) \boldsymbol{\Sigma}_k^{1/2})^{-1} (\mathbf{x}_i - (\boldsymbol{\mu}_k + \boldsymbol{\Sigma}_k^{1/2} \boldsymbol{\delta}_k t_i))) \right\} \right\} \right\} \\ &= \sum_{i=1}^N \left\{ \log \sum_{k=1}^K \left\{ \pi_{ik} \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Gamma}_k|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_k - \boldsymbol{\Delta}_k t_i)^T \boldsymbol{\Gamma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k - \boldsymbol{\Delta}_k t_i) \right\} \right\} \right\} \end{aligned} \quad (34)$$

where  $\boldsymbol{\Delta}_k = \boldsymbol{\Sigma}_k^{1/2} \boldsymbol{\delta}_k$  and  $\boldsymbol{\Gamma}_k = \boldsymbol{\Sigma}_k - \boldsymbol{\Delta}_k \boldsymbol{\Delta}_k^T$ . By letting  $\hat{s}_{i,k} = E[t_i | \mathbf{x}_i, \theta_k = \hat{\theta}_k]$  and  $\hat{s}_{i,k}^2 = E[t_i^2 | \mathbf{x}_i, \theta_k = \hat{\theta}_k]$ , we obtain, using the moments of the truncated normal distribution, that:

$$\hat{s}_{i,k} = \hat{\mu}_{t_{ik}} + W_{\Phi_1} \left( \frac{\hat{\mu}_{t_{ik}}}{\hat{M}_{t_{ik}}} \right) \hat{M}_{t_{ik}} \quad (35)$$

$$\hat{s}_{i,k}^2 = \hat{\mu}_{t_{ik}}^2 + \hat{M}_{t_{ik}}^2 W_{\Phi_1} \left( \frac{\hat{\mu}_{t_{ik}}}{\hat{M}_{t_{ik}}} \right) \hat{M}_{t_{ik}} \hat{\mu}_{t_{ik}} \quad (36)$$

where  $W_{\Phi_1}(u) = \frac{\phi_1(u)}{\Phi_1(u)}$ ,  $\hat{M}_{t_{ik}}^2 = \frac{1}{1 + \hat{\Delta}_k^T \hat{\Gamma}_k^{-1} \hat{\Gamma}_k}$  and  $\hat{\mu}_{t_{ik}} = \frac{\hat{\Delta}_k^T \hat{\Gamma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k)}{1 + \hat{\Delta}_k^T \hat{\Gamma}_k^{-1} \hat{\Gamma}_k}$ . Then the Q function of Eq. (34) can be written as:

$$\begin{aligned} Q(\boldsymbol{\Theta} | \hat{\boldsymbol{\Theta}}) &= E[L(\boldsymbol{\Pi}, \boldsymbol{\Theta} | \mathbf{X}) | \mathbf{X}, \hat{\boldsymbol{\Theta}}] \\ &= \sum_{i=1}^N \left\{ \log \sum_{k=1}^K \left\{ \pi_{ik} \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Gamma}_k|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_k - \boldsymbol{\Delta}_k \hat{s}_{i,k})^T \boldsymbol{\Gamma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k - \boldsymbol{\Delta}_k \hat{s}_{i,k}) - \frac{1}{2} (\hat{s}_{i,k}^2 - (\hat{s}_{i,k})^2) \boldsymbol{\Delta}_k^T \boldsymbol{\Gamma}_k^{-1} \boldsymbol{\Delta}_k - \frac{1}{2} \hat{s}_{i,k}^2 \right\} \right\} \right\} \\ &- \beta [K(s_i | \pi_i) + K(s_i | \pi_{\partial_i}) + H(s_i)] - \frac{1}{2} [K(q_i | z_i) + K(q_i | z_{\partial_i}) + H(q_i)] \end{aligned} \quad (37)$$

**E-step:** Given  $\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}$ , compute  $\hat{s}_{i,k}$  and  $\hat{s}_{i,k}^2$  by using Eqs. (35) and (36). By fixing  $\boldsymbol{\Theta}$  and  $\pi$ , we can optimize  $s_i$  and  $q_i$  for each pixel as follows:

$$s_i \propto \pi_{ik} \pi_{\partial_i k} \quad (38)$$

$$q_i \propto z_{ik} z_{\partial_i k} \quad (39)$$

Where  $\pi_{\partial_i k} = \sum_{j \in \partial_i, j \neq i} \alpha_{ij} \pi_{jk}$ , and  $z_{\partial_i k} = \sum_{j \in \partial_i, j \neq i} \alpha_{ij} z_{jk}$ .  $z_{ik} = \frac{\pi_{ik} p(x_i | \theta_k)}{\sum_{k=1}^K \pi_{ik} p(x_i | \theta_k)}$ .

**M-step:** By fixing  $s$  and  $q$ , we can maximize  $Q(\boldsymbol{\Theta} | \hat{\boldsymbol{\Theta}})$  over  $\boldsymbol{\Theta}$  and  $\boldsymbol{\Pi}$  as follows:

$$\hat{\pi}_{ik} = \frac{1}{1 + 2\beta} \left( \frac{1}{2} (q_{ik} + q_{\partial_i k}) + \beta (s_{ik} + s_{\partial_i k}) \right) \quad (40)$$

$$\hat{\boldsymbol{\mu}}_k = \frac{\sum_{i=1}^N (q_{ik} + q_{\partial_i k}) (\mathbf{x}_i - \Delta_k \hat{s}_{i,k})}{\sum_{i=1}^N (q_{ik} + q_{\partial_i k})} \quad (41)$$

$$\hat{\Delta}_k = \frac{\sum_{i=1}^N (q_{ik} + q_{\partial_i k}) \hat{s}_{i,k} (\mathbf{x}_i - \boldsymbol{\mu}_k)}{\sum_{i=1}^N (q_{ik} + q_{\partial_i k}) \hat{s}_{i,k}^2} \quad (42)$$

$$\hat{\Gamma}_k = \frac{\sum_{i=1}^N (q_{ik} + q_{\partial_i k}) ((\mathbf{x}_i - \boldsymbol{\mu}_k - \Delta_k \hat{s}_{i,k}) (\mathbf{x}_i - \boldsymbol{\mu}_k - \Delta_k \hat{s}_{i,k})^T + (\hat{s}_{i,k}^2 - (\hat{s}_{ik})^2) \Delta_k \Delta_k^T)}{\sum_{i=1}^N (q_{ik} + q_{\partial_i k})} \quad (43)$$

Where  $s_{\partial_i k} = \sum_{j \in \partial_i, j \neq i} \alpha_{ij} s_{jk}$ ,  $q_{\partial_i k} = \sum_{j \in \partial_i, j \neq i} \alpha_{ij} q_{jk}$  and  $\Sigma_k = \Gamma_k + \Delta_k \Delta_k^T$ .  $\lambda_k = \frac{\Sigma_k^{-1/2} \Delta_k}{\sqrt{1 - \Delta_k^T \Sigma_k^{-1} \Delta_k}}$  Please refer to Appendix A for the detailed derivation of the updating functions.

Consequently, the complete algorithm can be summarized as follows:

Step1: Initialize the parameters  $\Theta = (\boldsymbol{\mu}_k, \lambda_k, \Sigma_k, \delta_k)$  using K-means. Initialize the priors  $\pi_{ik} = 1/K$ .

Step2: Calculate the weight factor of neighborhood pixels  $\alpha_{ij}$  by using Eq. (15).

Step3: Compute  $\hat{s}_{ik}, \hat{s}_{ik}^2, z_{ik}, z_{\partial_i k}$  by using Eq. (35), (36), (30).

Step4: Compute and normalize  $s_{ik}, s_{\partial_i k}, q_{ik}, q_{\partial_i k}$  by using Eq. (38), (39).

Step5: Update  $\hat{\pi}_{ik}, \hat{\boldsymbol{\mu}}_k, \hat{\Delta}_k, \hat{\Gamma}_k$  by using Eq. (40)–(43).

Step6: Check for the convergence of either the objective function or the parameter values. If the convergence criterion is satisfied, stop the iteration; otherwise, go to Step 3.

#### 4. Experimental results

In this section, we demonstrate the performance of the proposed method through various experiments. We compare our method with six state-of-the-art algorithms: a spatially constrained generative model (SCGA-EM) [17], a fast and robust spatially constrained Gaussian mixture model (FRSCGMM) [18], a spatially constrained asymmetric Gaussian mixture model (SCAGMM) [19], a spatially constrained generative asymmetric Gaussian mixture model (SCGAGMM) [11], a spatially constrained in finite Dirichlet mixture models (SC-DMM) [4], and a shifted asymmetric Laplace mixture model (SALMM) [12].

In this paper, we use three qualitative indices, including the misclassification ratio (MCR) [18], the Js value [25] and the Probabilistic Rand Index (PRI) [27], to evaluated the performance of segmentation results.

The misclassification ratio (MCR) [18] is defined as:

$$MCR = \frac{N_s}{N} \quad (44)$$

where,  $N_s$  is the number of misclassified pixels and  $N$  is the total number of pixels in the image. The lower MCR value indicates a better result.

The Js value [25] is defined as the ratio between intersection and union of two sets:

$$Js(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} \quad (45)$$

Where  $S_1$  is the segmentation result and  $S_2$  denotes the ground truth. The value of  $Js$  is between 0 to 1. The higher values of  $Js$  represent better segmentation results.

The Probabilistic Rand Index (PRI) [27] is defined as the ratio between the segmentation result and a set of manually segmented results (ground truths):

$$PRI(S_1, \{S_2\}) = \frac{2}{N(N-1)/2} \sum_{i,j,i < j} \left[ p_{ij} I(l_i^{S_1} = l_j^{S_1}) + (1 - p_{ij}) I(l_i^{S_1} \neq l_j^{S_1}) \right] \quad (46)$$

where  $\{S_2\} = \{S_{21}, S_{22}, \dots, S_{2K}\}$  is a set of manually segmented results.  $l_i^{S_1}$  is the label of point  $i$  in the segmentation result  $S_1$  and  $l_i^{S_{2k}}$  is the corresponding label in the ground truth  $S_{2k}$ .  $I(\cdot)$  is an indicate function.  $p_{ij} = \frac{1}{K} \sum_{k=1}^K I(l_i^{S_{2k}} = l_j^{S_{2k}})$ . The value of  $PRI$  is between 0 to 1. The higher values of  $PRI$  represent better segmentation results.

Unless otherwise specified, the parameters of the proposed model are set as follows: the size of neighborhood patch is  $3 \times 3$ , the positive constant  $h$  is set as 0.9,  $\gamma$  is set as 1 and  $\beta$  is set as 0.3. All methods are initialized by using K-means method. The parameters of other methods are set as default values shown in the corresponding papers.

#### 4.1. Experiments on synthetic data

In the first experiment, we compare the performance of our method with SCGA-EM, FRSCGMM, SCAGMM, SCGAGMM, SC-DMM and SALMM on a simulated image. The first column of Fig. 4 shows an original image with Gaussian noise (mean: 0, variance: 0.005). The second column shows the ground truth. The third column to the last column shows the results of SCGA-EM, FRSCGMM, SCAGMM, SCGAGMM, SC-DMM, SALMM and our method, respectively. The SCGA-EM uses fixed positive neighbor weights for each pixel, and then it is hard to preserve the details. The FRSCGMM uses the average priori and posteriori probabilities to reduce the effect of noise, but unfortunately, it will smooth the edge regions. The SCAGMM uses the local best direction information to construct the neighborhood information, and the local best direction contributes to preserve more details in image. However, the best direction is chosen by using the distance between the intensity of current pixel and the mean value of each class, and then it is still sensitive to high noise. The SCGAGMM uses only the intensity-based similarity measurement in the isotropic neighborhood region. When the pixels, belonging to different class, in the local region have similar intensity values, the method is hard to achieve satisfactory results. The SC-DMM regards the pixels in the neighborhood region with the same weight, at the expansion of smoothing the edge. The SALMM uses the priori probabilities of the neighborhood region to calculate the weight of each pixel belonging to each class. The neighborhood region used in this method is also isotropic, and then the method is hard to preserve details especially when the image has severe noise. Our method uses the priori and posteriori probabilities information in the anisotropic neighborhood region to reduce the effect of noise. From the MCR values, we can find that our method achieves the best results. The zoomed views of the segmentation results are shown in the second row of Fig. 4.

In order to show the robustness to the noise, we compare the performance of our method with other six methods on simulated brain MR images with different noise level in the second experiment. The simulated brain MR image is generated from Brain Web (<http://www.bic.mni.mcgill.ca/brainweb/>). BrainWeb provides full 3-dimensional data volumes with a variety of slice thicknesses, noise levels and levels of intensity inhomogeneity. In this paper, we select T1-weighted images with 1 mm slice thickness. Fig. 5 shows the segmentation results on the images with noise level 5% (1st row) and 7% (3rd row). The third column to the last column shows the segmentation results of SCGA-EM, FRSCGMM, SCAGMM, SCGAGMM, SC-DMM, SALMM and our method, respectively. In brain MR images, some tissues have a slim structure, such as in brainstem, so that the isotropic neighborhood information-based methods are hard to preserve details. Our method uses anisotropic neighborhood information and can preserve more detail information, which can be found in the zoomed views. In order to show the robustness to noise, we test all these seven methods on 140 simulated brain MR images. The segmentation accuracies of these methods are measured by JS and the results can be found in Table 1. This result demonstrates that our method achieves the best segmentation performance (with the highest means of JS values) and has the best stability (with the smallest standard deviations). The JS values for CSF also prove that our method can preserve more details.

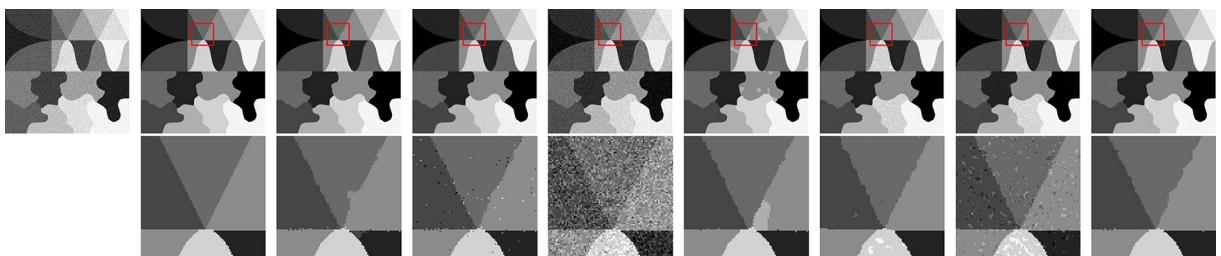
#### 4.2. Segmentation of natural images

Natural images are known to be scale-invariant with structures at all length scales. The main characteristic of the natural images is varying object sizes: from small objects with a few dozen pixels to large ones that occupy the entire image. For our experiments, we used the Berkeley segmentation data sets, which includes a collection of 500 natural images with ground truth segmentations.

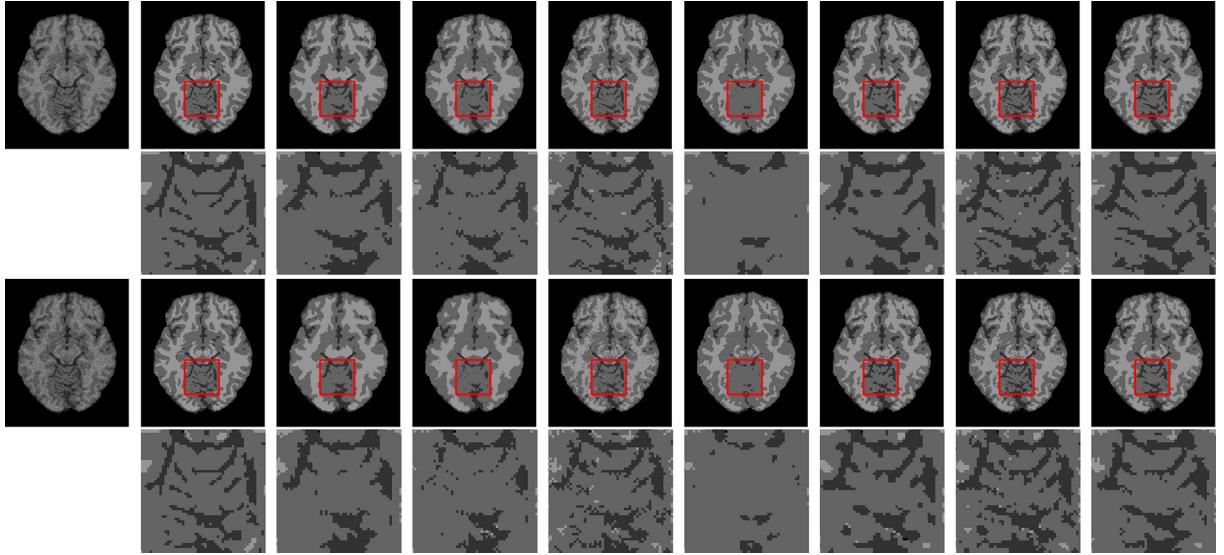
In order to determine the number of classes in a natural image, we use Marginal Mixture Information Criterion (MMIC) [28] to find an appropriate class number. The MMIC is defined as:

$$MMIC(K) = 2 \log p_{MLE}(Y|K) - d_K \log(N) \quad (47)$$

where  $K$  is the number of classes in an image,  $\log p_{MLE}(Y|K)$  is estimated by  $\sum_{i=1}^N \log \left( \sum_{k=1}^K P_k \Phi(y_i | \hat{\theta}_k) \right)$ ,  $y_i$  is the  $i$  th pixel in the image,  $\Phi(\cdot)$  is the pdf of Gaussian distribution.  $P_k$  denotes the mixture proportion of the  $k$  th component.  $\hat{\theta}_k$  is the optimal



**Fig. 4.** Classification results on a synthetic image. The first column and second column are the initial images with Gaussian noise (mean: 0, variance: 0.005) and the ground truth, respectively. From the third to right column are the segmentation results of SCGA-EM (MCR = 0.69%), FRSCGMM (MCR = 1.71%), SCAGMM (MCR = 42.58%), SCGAGMM (MCR = 2.10%), SC-DMM (MCR = 1.92%), SALMM (MCR = 7.61%) and Our method (MCR = 0.63%), respectively. The second row shows the details.



**Fig. 5.** The segmentation results on simulated brain MIR images. The first row and third row are the segmentation results on images with noise level 5% and 7%, respectively. The second row and fourth row show the details of the corresponding red rectangular areas. The first and second columns show the initial images and the ground truths. The third column to right column shows the segmentation results of SCGA-EM, FRSCGMM, SCAGMM, SCGAGMM, SC-DMM, SALMM, and our method, respectively.

**Table 1**

The average JS values and MCR values of the segmentation results on simulated brain MR images.

		SCGM-EM	FRSCGMM	SCAGMM	SCGAGMM	SCDMM	SALMM	Proposed
3%	JS_WM	0.8916±0.011	0.8919±0.009	0.8949±0.028	0.8984±0.011	0.8899±0.024	0.9071±0.036	0.9094±0.006
	JS_GM	0.8572±0.025	0.8733±0.006	0.8791±0.006	0.8363±0.021	0.8588±0.012	0.8731±0.003	0.8800±0.004
	JS_CSF	0.8258±0.012	0.8645±0.060	0.8784±0.008	0.6503±0.005	0.8252±0.007	0.8439±0.007	0.8829±0.030
	MCR	0.0716±0.005	0.0636±0.098	0.0644±0.017	0.0885±0.007	0.0681±0.003	0.0615±0.006	0.0610±0.004
5%	JS_WM	0.8511±0.015	0.8407±0.010	0.8273±0.028	0.8659±0.018	0.8681±0.032	0.8550±0.035	0.8705±0.010
	JS_GM	0.8103±0.030	0.8141±0.012	0.8064±0.010	0.8035±0.024	0.8344±0.009	0.8154±0.007	0.8375±0.005
	JS_CSF	0.7871±0.014	0.8033±0.060	0.8214±0.010	0.6146±0.006	0.7870±0.010	0.8048±0.010	0.8253±0.028
	MCR	0.0967±0.007	0.0969±0.007	0.1004±0.003	0.1079±0.006	0.1040±0.007	0.0910±0.005	0.0819±0.004
7%	JS_WM	0.8102±0.019	0.7954±0.009	0.7466±0.031	0.8203±0.009	0.8194±0.055	0.8205±0.038	0.8367±0.011
	JS_GM	0.7621±0.042	0.7623±0.014	0.7279±0.012	0.7781±0.032	0.7743±0.017	0.7790±0.007	0.7987±0.006
	JS_CSF	0.7330±0.013	0.7288±0.052	0.7491±0.012	0.5878±0.007	0.6926±0.018	0.7676±0.010	0.7789±0.031
	MCR	0.1252±0.007	0.1308±0.010	0.1502±0.004	0.1286±0.008	0.1678±0.029	0.1477±0.009	0.1021±0.005

parameter by maximizing the loglikelihood function  $\sum_{i=1}^N \log(\sum_{k=1}^K P_k \Phi(y_i|\theta_k))$ ,  $d_k$  is number of parameters in the mixture model with  $K$  components. A larger value of MMIC( $K$ ) represents a better segmentation result. More details could be referred in [28].

**Table 2** lists the MMIC results with the  $K$  varying from 2 to 8 on eight natural images. It is clear that the MMIC obtains the largest values when the  $K$  is 3,3,3,6,3,3, for images: #8143, #134035, #22090, #55067, #22023, #198035, #118035 respectively. For image # 238011, when the  $K$  is set as 8, the MMIC obtains the best result, however, the image only has three clusters: the tree, the sky and the moon. Thus, we set  $K = 3$  for image #238011.

In the next experiment, we compare our method with SCGM-EM, FRSCGMM, SCAGMM, SCGAGMM, SC-DMM, SALMM on a collection of natural images with backgrounds are virtually indistinguishable from foregrounds. Our method significantly improves background representations in Fig. 6 by constructing regions of uniform pixel values, with clear contrasts between forest backgrounds and animals posed in the front.

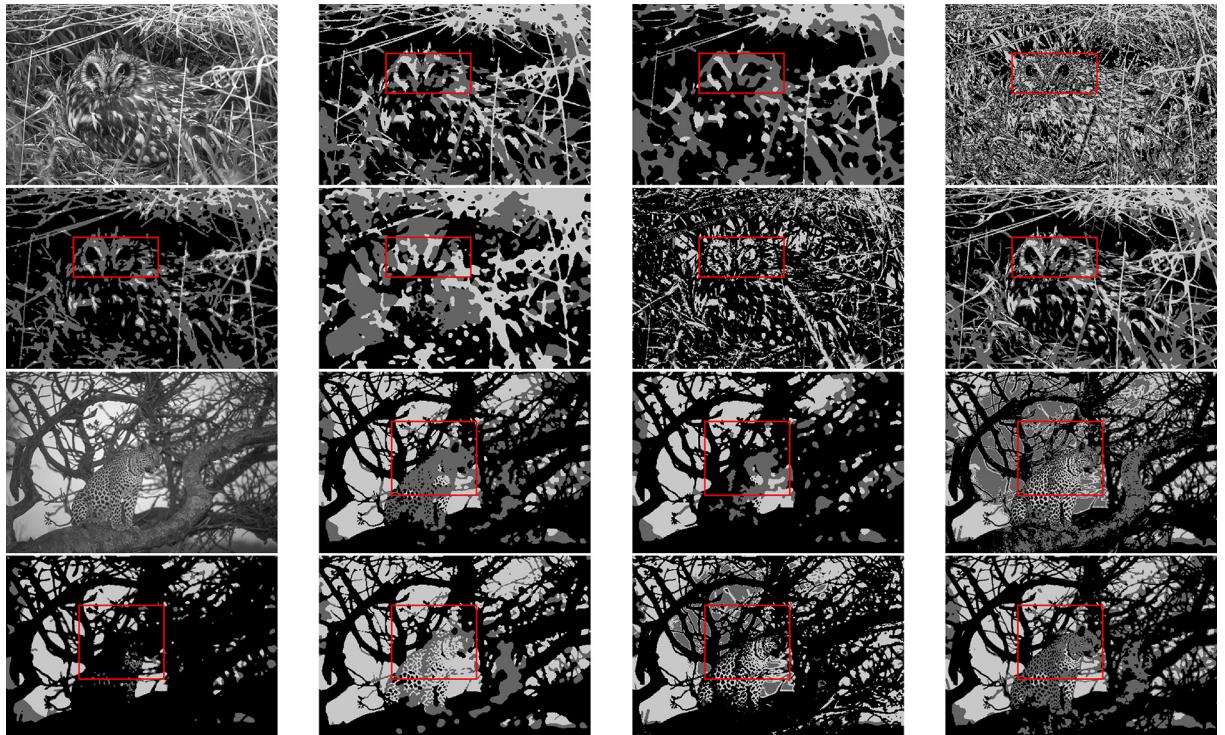
Fig. 7 shows the segmentation results on the images #22090 and #55067, whose distributions have asymmetric forms. Our method finds precise outlines of the bridge in the image #22090, while providing weak edges regarding mountains in the image #55067. Increased details are attributed to anisotropic neighborhood information and the skew distributions. However, all these methods failed to find the accurate segmentation results of the reflections of the bridge from the water in the image #22090, which should be overcome in the future work. As the results shown in Fig. 8, our method preserves more detail information of the handrail in the image #22013, while providing details of texture regions in the image

**Table 2**

The MMIC values for natural images with different K.

	K							
	2	3	4	5	6	7	8	
#8143	−189015.95	<b>−158770.19</b>	−168230.88	−197528.16	−194036.55	−194552.12	−206335.83	
#134035	−210265.69	<b>−160766.25</b>	−184234.59	−206678.33	−199802.34	−210556.47	−217702.63	
#22090	−200173.28	<b>−145404.35</b>	−154875.32	−419046.69	−365669.91	−390747.92	−395082.81	
#55067	−437013.35	−495592.07	<b>−470183.16</b>	−414504.74	<b>−355404.05</b>	−474000.92	−414836.90	
#22013	−654142.75	<b>−620270.43</b>	−630871.28	−626741.01	−625089.43	−624092.39	−625384.91	
#198023	−716567.94	<b>−650090.71</b>	−656410.87	−654274.71	−654371.88	−654106.76	−652812.41	
#118035	−191630.35	<b>−126534.76</b>	−138894.47	−136611.89	−137452.26	−158232.76	−159663.69	
#238011	−276302.041	<b>−175950.22</b>	−210336.68	−179585.35	−172763.81	−166717.48	−162700.54	

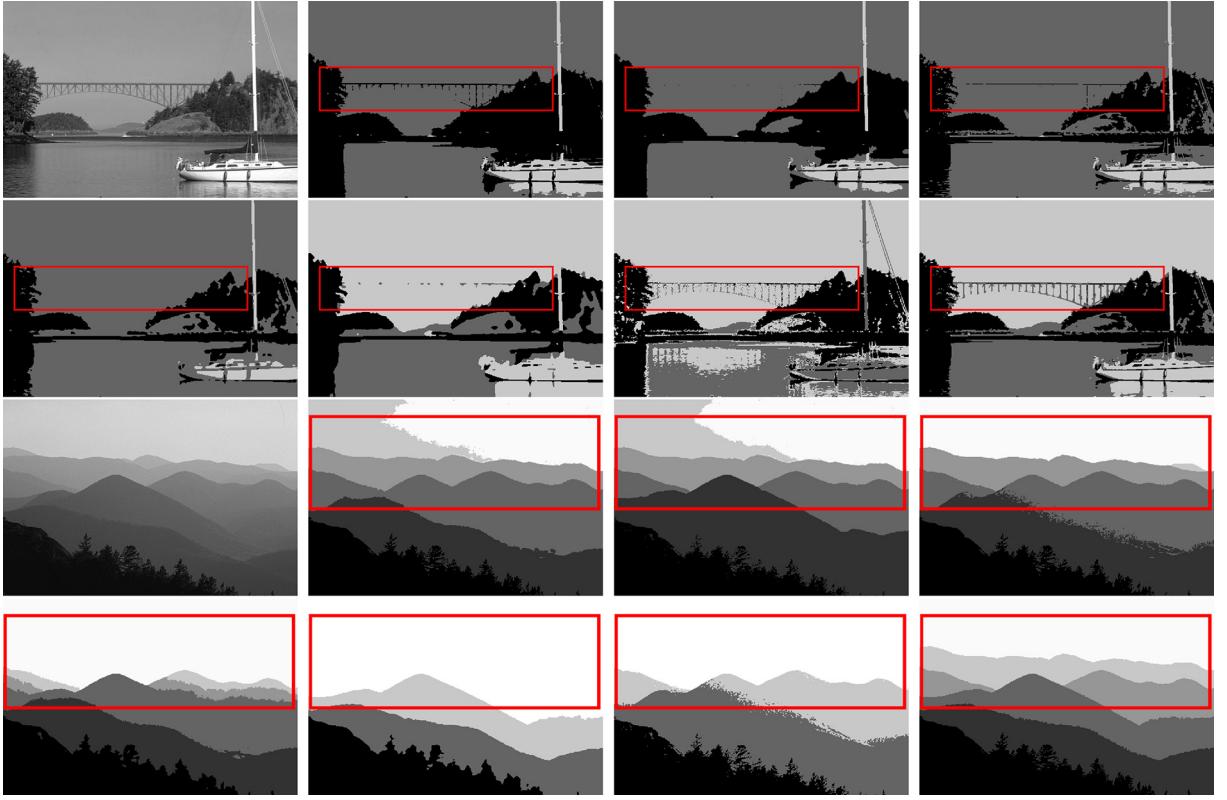
The bold font indicates that the value is the best.

**Fig. 6.** Image segmentation results on natural images (#8143, #134035). The first row and second row are the original image (#8143), segmentation results of SCGM-EM, FRSCGMM, SCAGMM, SCGAGMM, SC-DMM, SALMM and our method, respectively. The third row and fourth row are the results for the image (#134035). The number of classes in these two images is 3.

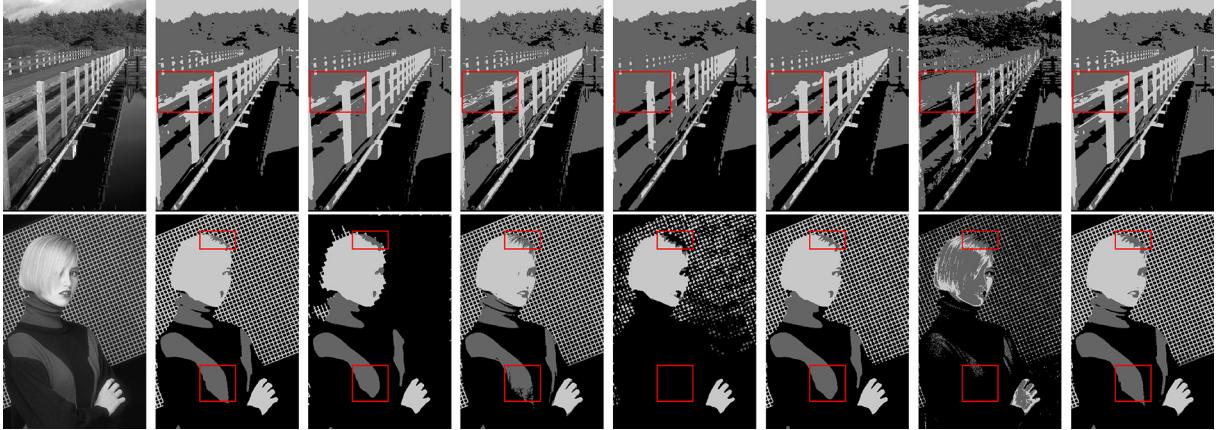
#198023. The isotropic neighborhood information-based methods easily lose details information. The analysis above implies asymmetric properties are crucial for describing contours in natural images.

In order to show the robustness to noise, the next experiment is tested on natural images (#118035, #238011) with noise. The noise is Gaussian noise with parameter: mean = 0 and variance = 0.005. Both of these images have three classes. From the results shown in Fig. 9, we can find that in the image (#118035), SCGM-EM and FRSCGMM cannot find accurate edges of the roof. SCAGMM, SC-DMM, SALMM can obtain more accurate edges of the roof, however, they are still sensitive to noise. SCGAGMM incorrectly segments some pixels belonging to the roof as the wall. In the image (#238011), SCGM-EM, FRSCGMM and SC-DMM fail to find the edge of the moon. SCAGMM and SALMM are sensitive to noise. SCGAGMM incorrectly segments some pixels belonging to the tree as the sky. Our method equipped the anisotropic neighborhood information into the skew normal distribution, and therefore, it can reduce the effect of noise and lead to the best accurate results.

In order to show the robustness of our method, we show the Box-plot of the Js values for 21 simulated brain MR images and PRI values for 20 natural images in Fig. 10. The simulated images have noise levels: 3%, 5% and 7%, respectively. From the results, we can find that the median values of our method are much larger than those of the other six methods when the noise level of the simulated brain images increases and the interquartile ranges of our method are shorter than most of the other methods. When the noise level is 3%, the median value of our method is not the best one, but the interquartile range of



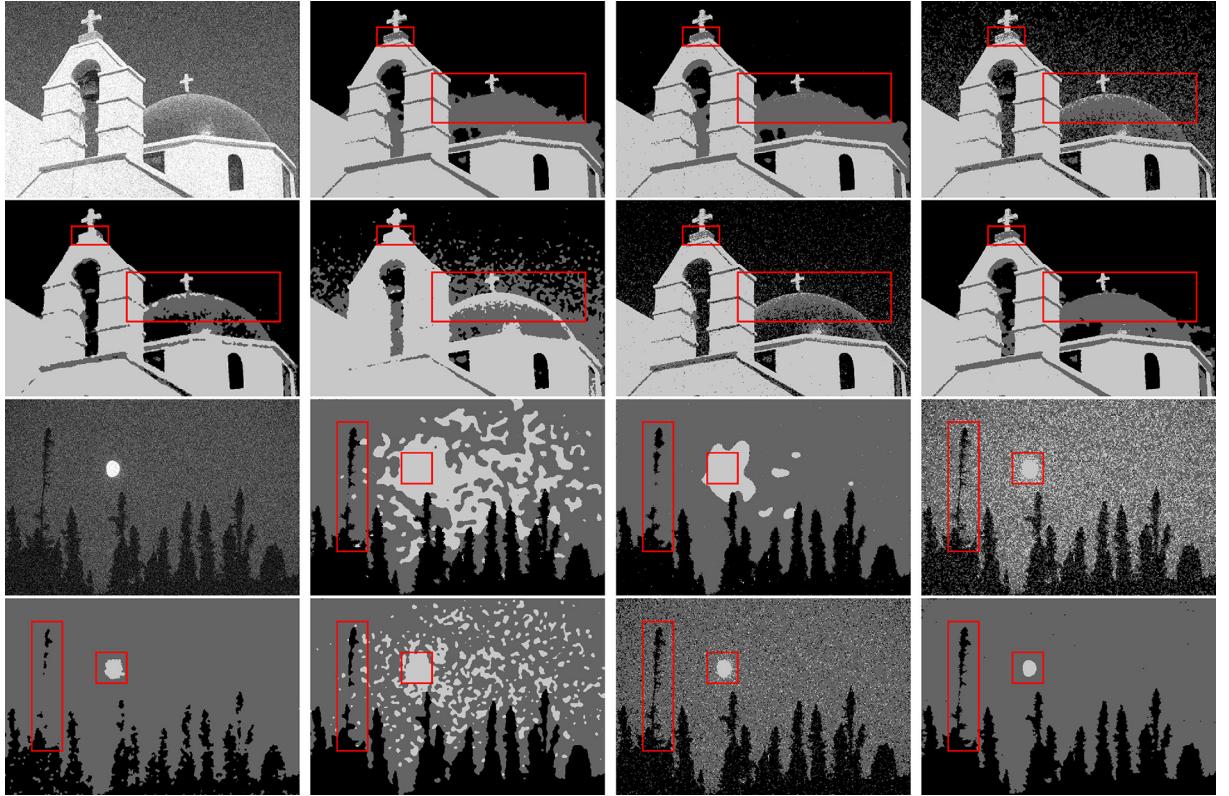
**Fig. 7.** Image segmentation results on natural images (#22090, #55067). The first row and second row are the original image (#22090), segmentation results of SCGM-EM, FRSCGMM, SCAGMM, SCGAGMM, SC-DMM, SALMM and our method, respectively. The third row and fourth row are the results of the image (#55067). The numbers of classes in these two images are 3 and 6, respectively.



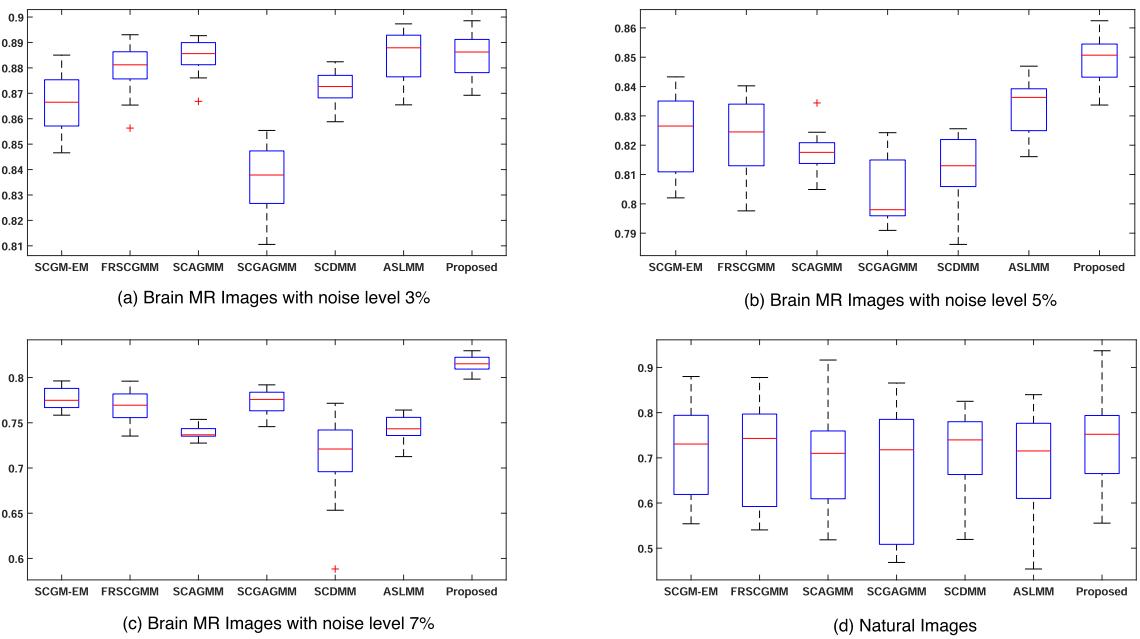
**Fig. 8.** Image segmentation results on natural images (#22013, #198023). The first row and second row are the original image (#22013), segmentation of SCGM-EM, FRSCGMM, SCAGMM, SCGAGMM, SC-DMM, SALMM and our method, respectively. The third row and fourth row are the results of the image (#198023). The number of classes in these two images is 3.

our method if shorter than that of ASLMM. The PRI values of the natural images illustrated that our method can obtain the best results.

In order to show the improvement of our method, we use the Friedman test and the Iman-Davenport test [29] to compare the difference between the segmentation results of our method and the other six methods. All these compared methods are tested on the same images in Fig. 10 and the results can be found in Table 3. The p values of the Friedman test and the Iman-Davenport test are far less than 0.05, which means that there are obvious differences between our method and the



**Fig. 9.** Image segmentation results on noise natural images (#118035, #238011). The first and second rows are the original image (#118035), segmentation of SCGM-EM, FRSCGMM, SCAGMM, SCGAGMM, SC-DMM, SALMM and our method, respectively. The third and fourth rows are the results for the images (#238011). The number of classes in these two images is 3.



**Fig. 10.** The Box-plot of the JS values for Brain MR images and PRI values for natural images.

**Table 3**

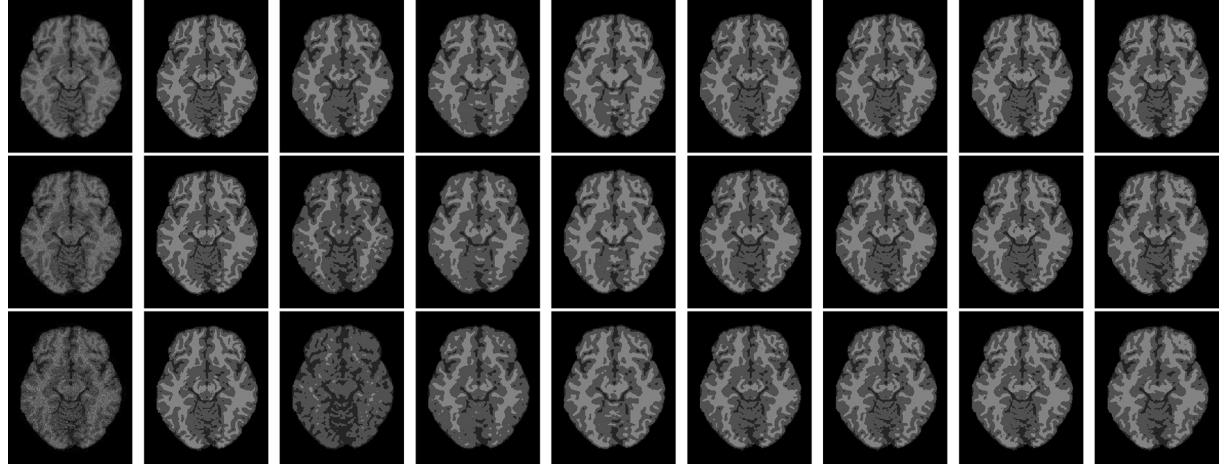
The result of the Friedman test and Iman-Davenport test.

	Friedman		Iman-Davenport	
	Statistical value	p-value	Statistical value	p-value
Brain MR image	362.541	3.1*10-75	79.419	2.2*10-16
Natural image	33.473	8.0*10-10	7.482	1.0*10-06

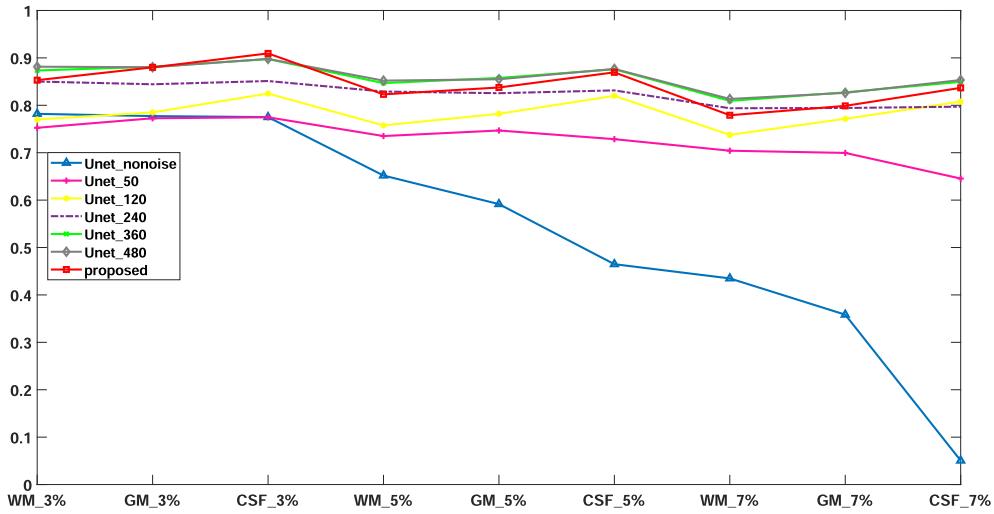
**Table 4**

The Average rank values of the Friedman test.

	SCGM-EM	FRSCGMM	SCAGMM	SCGAGMM	SCDMM	SALMM	Proposed
Average rank	4.894	4.421	3.105	2.526	4.263	3.000	5.789



**Fig. 11.** The segmentation results on the 160-th simulated MIR brain images. The first row to third row are the segmentation results on images with noise level 3%, 5% and 7%, respectively. The first column is the original images, the second column are the ground truth. The third column to right column shows the segmentation results of Unet\_nonoise, Unet\_50, Unet\_120, Unet\_240, Unet\_360 and Unet\_480, and our method, respectively.



**Fig. 12.** The JS values of the Unet methods with different sizes of the training data sets and our method.

other six methods. Furthermore, it can be found from the average rank values of the Friedman test shown in [Table 4](#) that our method obtains the biggest average rank, which means that our method has a significant improvement in segmentation accuracy.

**Table 5**

Computational complexity, convergence time, number of iterations and per iteration time by applying seven methods on simulated brain image and natural image (in the second).

Image size	Methods	SCGM-EM	FRSCGMM	SCAGMM	SCGAGMM	SC-DMM	SALMM	Our method
109 × 91	Computational complexity	$O(TNKD^2N_{\partial_i}^2)$	$O(TNKD^2N_{\partial_i}^2)$	$O(TNKLD^2N_{\partial_i}^2)$	$O(TNKLD^2N_{\partial_i}^2)$	$O(TNKD^2N_{\partial_i}^2)$	$O(TNKD^2N_{\partial_i}^2)$	$O(TNKD^2N_{\partial_i}^2)$
	Converging time	0.034123 ± 0.012064	0.040556 ± 0.028792	0.484289 ± 0.168336	0.377125 ± 0.107054	0.483215 ± 0.021023	0.252728 ± 0.021494	0.161318 ± 0.052366
	Number of iterations	7.5 ± 3.4	14.4 ± 14.5	19.8 ± 7.2	19.0 ± 6.7	26.7 ± 0.88	10.3 ± 1.49	7.95 ± 3.69
	Per iteration time	0.004727 ± 0.0013	0.003622 ± 0.001227	0.024438 ± 0.000885	0.020246 ± 0.001604	0.018088 ± 0.000524	0.024549 ± 0.00178	0.021356 ± 0.002616
217 × 181	Converging time	0.07918 ± 0.010818	0.0602 ± 0.009406	1.950382 ± 0.878773	1.006153 ± 0.199483	1.657485 ± 0.348554	0.760659 ± 0.027723	0.835782 ± 0.106822
	Number of iterations	7.5 ± 1.3	9.5 ± 1.6	27.4 ± 13	12.3 ± 2.5	26.2 ± 5.8	11.4 ± 0.49	9.19 ± 1.29
	Per iteration time	0.010619 ± 0.000856	0.006354 ± 0.0007	0.071722 ± 0.001736	0.081737 ± 0.001307	0.063222 ± 0.001519	0.066587 ± 0.001187	0.091173 ± 0.00298
161 × 241	Converging time	0.07076 ± 0.027707	0.072429 ± 0.03999	1.342088 ± 0.99159	3.137252 ± 2.661235	1.335574 ± 0.839432	0.67399 ± 0.519391	0.489126 ± 0.164408
	Number of iterations	6.9 ± 3.7	13.3 ± 10.37	18.4 ± 11.7	34 ± 13.6	36.9 ± 23.7	11.8 ± 5.8	5.4 ± 3.2
	Per iteration time	0.011455 ± 0.004092	0.006976 ± 0.002688	0.071538 ± 0.02155	0.084916 ± 0.03098	0.039749 ± 0.023438	0.054532 ± 0.011533	0.10349 ± 0.038084
321 × 481	Converging time	0.0481307 ± 0.164702	0.412781 ± 0.20773	13.27207 ± 7.26396	26.97781 ± 23.08297	8.2928 ± 10.61092	2.551299 ± 0.66429	2.848639 ± 1.254791
	Number of iterations	7.5 ± 3.1	12.7 ± 9.8	27.2 ± 15.8	34.7 ± 15.4	35.4 ± 22.3	13.1 ± 4.2	6.9 ± 3.7
	Per iteration time	0.069344 ± 0.021772	0.03703 ± 0.011105	0.49165 ± 0.149459	0.715256 ± 0.211515	0.194834 ± 0.086483	0.203542 ± 0.053036	0.455991 ± 0.13346

#### 4.3. Compared with deep learning methods

In recent years, deep learning methods have been widely used in image segmentation and good results have been obtained. One of the most famous image segmentation model is U-net [20], which is committed to segmenting biomedical images based on convolutional neural network. Compared with our proposed model, U-net is a supervised method and depends on a larger number of training data sets. Our model is an unsupervised method and does not need any training data.

In order to compare the difference performance between the U-net model and our method, we experiment on 144 simulated brain MR images without noise and  $144 \times 3$  images with the noise levels: 3%, 5% and 7%, respectively. We choose different kinds of training data sets to illustrate the effect of noise and size of training data to the U-net model. Unet\_50, Unet\_120, Unet\_240, Unet\_360 and Unet\_480 are the training data sets randomly selected from the total images with size 50, 120, 240, 360 and 480, respectively. The training data of Unet\_nonoise only has all the non-noise images.

[Fig. 11](#) shows the segmentation results of the 160th slice of brain MR images with the noise levels of 3%, 5%, 7%. [Fig. 12](#) shows the average JS values of WM, GM and CSF for segmentation results by using six kind of U-net methods and our method. From the results, we can find that the Unet\_nonoise is sensitive to noise. The accuracy of the U-net method increases with the number of training images. The size of the training data affects the accuracy of the U-net method. From the results, we can find that the U-net method cannot extract enough feature information to keep the target details in small samples. The results in the fourth and fifth column in [Fig. 11](#) reveal this shortcoming. The results of Unet\_360 and Unet\_480 present the superiority of deep learning approach for studying great quality training data. It is undeniable that they performance better than our model. However, the average JS values of CSF on images with noise 3% prove that our method can preserve more detail information. Furthermore, our model can obtain more accurate results without training samples or small samples.

#### 4.4. Comparison of computational complexity

To evaluate and compare the computational complexity and convergence time, the seven comparison algorithms were tested on images with four different sizes:  $109 \times 91$ ,  $217 \times 181$ ,  $161 \times 141$  and  $321 \times 481$ . Each kind of data sets has 200 images. The first two kinds of images are obtained from Brain Web and the other two kinds of images are natural images. In this experiment, the stopping criterion is set as  $\varepsilon = 10^{-4}$ . All algorithms were performed using the Matlab on a PC (Intel Core i7-5500 CPU, 2.4 GHz, 8 GB RAM, and 64-bit Windows 10).

[Table 5](#) presents the details about the computational complexity, the converging time, number of iterations and per iteration time for all comparison methods. The computational complexities of the compared seven methods are listed in the second row, where T is the number of iterations, N is the number of total pixels, D is the dimension of each pixel, K is the class number, L is the between-cluster mentioned in [11,19],  $N_{\partial_i}$  is the number of pixels for neighbor patch around  $\mathbf{x}_i$ . In our method, the spatial factor  $\alpha_{ij}$  should be pre-calculated before the EM iteration process, therefore, the total computational complexity should be added by  $NDN_{\partial_i}^4$ , and [Table 6](#) shows the time consuming of the spatial factor  $\alpha_{ij}$  in our method for different sizes of images.

From the results in [Table 5](#), we can find that the total convergence time of our algorithm is slightly higher than SCGM-EM, FRSCGMM and SALMM. However, our method can obtain more accurate results. Furthermore, when segmenting natural images with asymmetric form, our method converges faster, even when the images have large sizes. Although the nonlocal spatial factor  $\alpha_{ij}$  needs to be pre-calculated, the whole converging time of our methods is less than SCAGMM, SCGAGMM, SC-DMM when segmenting natural images with big sizes.

#### 4.5. Discussion

The stopping criterion affects the convergence of the proposed method. In order to show the effects of the stopping criteria, we compared all comparison methods on 200 simulated brain images with size  $217 \times 181$ . [Table 7](#) presents the details about the JS values, number of iterations and the converging time. It can be found that all these methods have similar JS values when the stopping criterion is set as  $10^{-4}$  or  $10^{-5}$ . However, the number of iterations and converging times of  $10^{-4}$  are much less than those of  $10^{-5}$ . Based on these findings, we set the stopping criteria as  $10^{-4}$ .

Although our method has achieved a good performance on different images, there are still some limitations. One limitation of the work is the nonnegative constant  $h$ , which plays the role of adjusting the weight of the neighboring pixels. A large  $h$  will lead to over-smoothness and cannot preserve detail information. While with a small  $h$ , the method will be sensitive to noise. The corresponding experiments on 200 simulated MRI brain images and 200 natural images with different noise levels

**Table 6**

Time consuming of spatial factor  $\alpha_{ij}$  on different sizes of images (unit: second).

$109 \times 91$	$217 \times 181$	$161 \times 241$	$321 \times 481$
$0.165503 \pm 0.00907$	$0.64749 \pm 0.00557$	$0.668529 \pm 0.011658$	$1.638497 \pm 0.12537$

**Table 7**

JS values, number of iterations and convergence time by applying seven methods on simulated brain image and natural image (in the second).

criteria		SCGM-EM	FRSCGMM	SCAGMM	SCGAGMM	SCDMM	SALMM	Proposed
$10^{-3}$	JS	78.8109±1.2977	75.1199±0.8397	75.4614±0.7765	76.8938±1.2786	75.6019±1.5524	71.7010±1.9086	79.1649±1.2626
	Number of iterations	2±0	1±0	1±0	7.85±1.08	3±0	1±0	2±0
	Converging time	0.0348±0.0012	0.0101±0.0015	0.1007±0.0041	0.6659±0.0686	0.3473±0.0337	0.1721±0.0050	0.2022±0.0039
$10^{-4}$	JS	83.3380±0.9524	81.4296±1.2631	81.6738±0.6588	78.6779±1.3998	77.8083±1.2950	83.1270±1.0598	84.4112±0.8109
	Number of iterations	7.5±1.3	9.5±1.6	27.4±13	12.3±2.5	26.2±5.8	11.4±0.49	9.19±1.29
	Converging time	0.0791±0.0108	0.0602±0.0094	1.9504±0.8788	1.0062±0.1995	1.6575±0.3486	0.7607±0.02723	0.8357±0.1068
$10^{-5}$	JS	83.3390±1.1662	81.4315±0.876	81.6745±0.6339	78.6782±1.6187	77.8086±1.341	83.1278±1.055	84.4114±0.8206
	Number of iterations	27±4.24	28±2.75	42±0.4	10±0.25	10±0.27	10±0.24	22.6±6.1
	Converging time	0.2879±0.1522	0.1580±0.039	3.3213±0.6901	6.5060±0.1396	3.0084±0.071	4.6139±0.3357	7.0228±1.8606

(3% 5%, and 7%, respectively) are shown in Fig. 13. The results illustrate that when  $h$  is set as 0.9, our method can obtain promising results.

Another limitation of the work is the constant  $\gamma$ , which determines the scope of the Gaussian kernel function and also affects the neighbor weight  $w_{ij}$ . The value range of  $\gamma$  is based on the neighborhood pixels coordinate distance. Fig. 14 shows the MCR values on 200 simulated MRI brain images and 200 natural images. All these images have different noise levels: 3%, 5%, and 7%, respectively. The results illustrate that when  $\gamma$  is set between 0.9 to 2, our method can obtain more accurate results.

The temperature value  $\beta$  should be chosen as small enough to preserve its sharpness and details information. However, a large  $\beta$  can reduce the effect of noise [18]. Chatzis et al. [30] provided a complete derivation method based on the specific selection of the energy function of the MRF prior. The derivation method needs to solve another optimal problem by maximizing an energy function based on the temperature value  $\beta$ , and therefore, the method seems more complexity and the results are unstable. We test on 200 brain images and 200 natural images with noise levels: 3%, 5% and 7%, respectively. The mean MCR values are shown in Fig. 15. The results indicate that when  $\beta$  is set as 0.3, the proposed method can obtain more accurate results. In the future, we will continue to investigate how to automatically optimize  $\beta$ .

The scale of the nonlocal patch used in the weight factor Eq. (15) may affect the accuracy of the results. Therefore, we tested different scales of the nonlocal patch on 100 simulated images. The average MCR and convergence times are shown in Table 8. Although the MCR values are all small in different sizes of the nonlocal patch, it is obvious that the patch with size  $3 \times 3$  has the smallest converging time. Hence, we select the size of the patch as  $3 \times 3$  in our model.

The global parameters always influence the performance of the system. Table 9 lists the number of global parameters in seven models, respectively. Our method has three global parameters, however, based on the experiments shown in Figs. 13–15, it can be found our method is less affected by the parameters  $\gamma$  and  $h$ . Although SCGM-EM has only one global parameter and there is no global parameter in FRSCGMM, their results are not better than ours.

On the other hand, the objective function of our method is nonconvex. Consequently, our method could be trapped into a local optimum. The global optimization methods can be used to find global optima. Furthermore, some images have heavy tails form, and then the Gaussian distribution is hard to find accurate results. Our future work will focus on improving the robustness to initialization and choosing more robust statistic distributions.

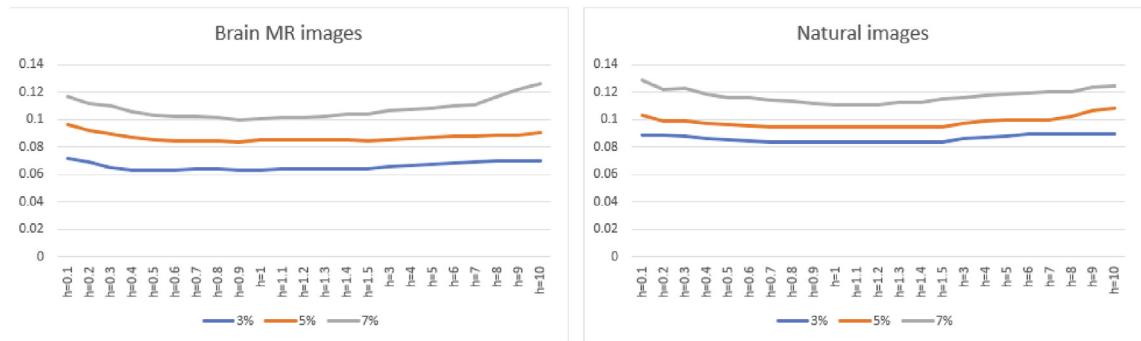


Fig. 13. The value of MCR for  $h$  on 200 brain MR images and 200 natural images with different noise levels: 3%, 5% and 7%, respectively.

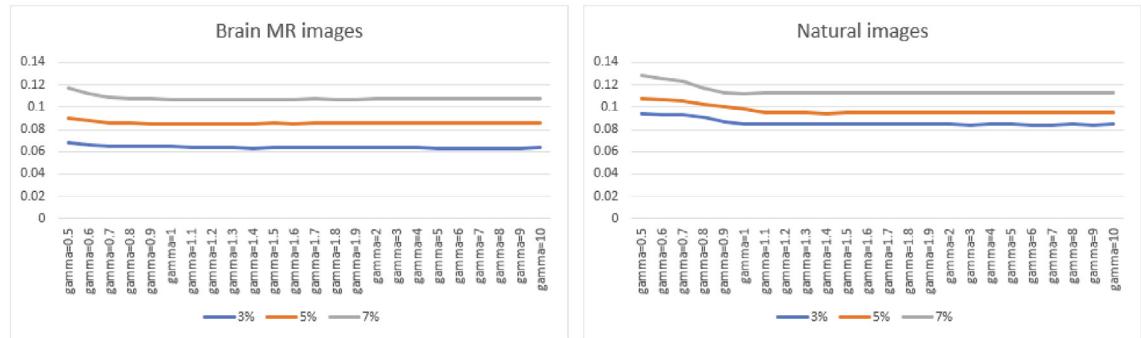
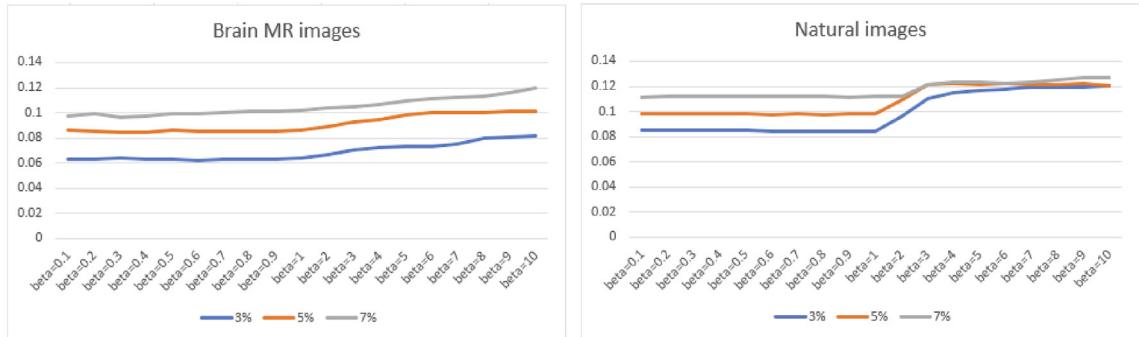


Fig. 14. The value of MCR for  $\gamma$  on 200 brain MR images and 200 natural images with different noise levels: 3%, 5%, and 7%, respectively.



**Fig. 15.** The value of MCR for  $\beta$  on 200 brain MR images and 200 natural images with different noise levels: 3%, 5%, and 7%, respectively.

**Table 8**  
MCR, Convergence time, iterations by applying three sizes of patch on noise image.

Size of nonlocal patch	$3 \times 3$	$5 \times 5$	$7 \times 7$
MCR	0.0014307%	0.0033250%	0.0042822%
Converging time	3.720318(s)	9.546815(s)	12.690129(s)

**Table 9**  
The number of global parameters.

The global parameters in the model	
SCGM-EM	$\beta$
FRSCGMM	none
SCAGMM	$\alpha, \beta$
SCGAGMM	$\alpha, \beta$
SC-DMM	$a, b, u_{kd}, v_{kd}$
SALMM	$\beta, w_k$
Proposed	$\beta, \gamma, h$

## 5. Conclusions

To further improve the segmentation accuracy for GMM based method, we have proposed a spatially constrained asymmetric Gaussian mixture model for image segmentation. This algorithm has the flexibility to fit different shapes of the distributions of the observed data, and successfully overcomes the drawbacks of existing EM-type mixture models. The results on synthetic images, brain MR images and natural images proved that the proposed algorithm is capable of producing more accurate segmentation results than several state-of-the-art algorithms.

Skew normal mixture models are competitive for pattern recognition because they possess real-valued skew parameter that can characterize different forms of the distributions in the underlying system. The nonlocal information is used to find accurate tangent and normal directions, even when the image has strong noise. Then, based on the accurate directions, we proposed a novel structure tensor framework, which is isotropic in inner regions and anisotropic in edge regions. The skew normal mixture model is improved by using the anisotropic spatial factor, which is based on the improved structure tensor and the nonlocal information. The experiments indicate superior capabilities for image segmentation.

For future work, one can extend other finite mixture models that come from the heavy-tailed distribution family (e.g., Students t distribution) to the skew context. One can also combine anisotropic skew mixture models with other machine learning approaches (e.g., deep learning) to achieve state-of-the-art results for image segmentation.

## CRediT authorship contribution statement

**Yunjie Chen:** Conceptualization, Methodology. **Ning Cheng:** Software, Writing - original draft. **Mao Cai:** Software, Data curation, Visualization. **Chunzheng Cao:** Supervision. **Jianwei Yang:** Investigation, Writing - review & editing. **Zhichao Zhang:** Validation, Writing - review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper. This work was supported in part by the National Natural Science Foundation of China 61672291, 61976241, 61572015, 61901223, in part by the Six Talent Peaks Project in Jiangsu Province SWYY-034, and in part by the Macau Young Scholars Program AM2020015.

## Appendix A. Derivation of the proposed method

As in Eq. (37), the complete log-likelihood function  $Q$  has the form:

$$\begin{aligned} Q(\Theta|\hat{\Theta}) &= E\left[L(\Pi, \Theta|X)|X, \hat{\Theta}\right] \\ &= \sum_{i=1}^N \left\{ \log \sum_{k=1}^K \left\{ \pi_{ik} \frac{1}{\sqrt{(2\pi)^n |\Gamma_k|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_k - \Delta_k \hat{s}_{i,k})^T \Gamma_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k - \Delta_k \hat{s}_{i,k}) - \frac{1}{2} (\hat{s}_{i,k}^2 - (\hat{s}_{i,k})^2) \right\} \right\} \right\} \\ &\quad - \beta [K(s_i|\pi_i) + K(s_i|\pi_{\partial_i}) + H(s_i)] \\ &\quad - \frac{1}{2} [K(q_i|z_i) + K(q_i|z_{\partial_i}) + H(q_i)] \end{aligned} \quad (48)$$

Then we utilize EM algorithm to maximize the energy  $Q$ . In the E-step, we fix  $\Theta$  and  $\Pi$  to maximize  $Q$  over  $s$  and  $q$ . In the M-step, we fix  $s$  and  $q$  to maximize  $Q$  over  $\Theta$  and  $\Pi$ .

E-step: By fixing  $\Theta$  and  $\Pi$ , we can optimize over  $s_i$ . The terms involving  $s_i$  in  $Q$  are:

$$\begin{aligned} &K(s_i|\pi_i) + K(s_i|\pi_{\partial_i}) + H(s_i) \\ &= -\sum_{k=1}^K s_{ik} \log s_{ik} + \sum_{k=1}^K s_{ik} \log \pi_{ik} - \sum_{k=1}^K s_{ik} \log \pi_{\partial_i k} + \sum_{k=1}^K s_{ik} \log s_{ik} \\ &= -\sum_{k=1}^K s_{ik} \log s_{ik} + \sum_{k=1}^K s_{ik} \log (\pi_{ik} \pi_{\partial_i k}) \end{aligned} \quad (49)$$

The above formula is a negative KL-divergence which becomes zero when [11]:

$$s_i \propto \pi_{ik} \pi_{\partial_i k} \quad (50)$$

By applying the similar derivation holds for  $q_i$ , we can get

$$q_i \propto z_{ik} z_{\partial_i k} \quad (51)$$

Therefore, we can get the updating functions for  $s_i$  and  $q_i$  as Eq. (38) and (39) in our method.

M-step: By fixing  $s$  and  $q$ , we can maximize  $Q$  over  $\Theta$  and  $\Pi$ . The terms involving  $\Pi$  and  $\Theta$  are:

$$\begin{aligned} &\sum_{i=1}^N \left\{ \log \sum_{k=1}^K \{ \pi_{ik} p(\mathbf{x}_i|\Theta) \} \right\} \\ &- \beta \left[ K(s_i|\pi_i) + \sum_{j \in \partial_i} K(s_j|\pi_{\partial_j}) \right] - \frac{1}{2} \left[ K(q_i|z_i) + \sum_{j \in \partial_i} K(q_j|z_{\partial_j}) \right] \end{aligned} \quad (52)$$

First of all, let us consider the derivation over  $z_i$ , then the terms involving only  $z_i$  are:

$$\begin{aligned} &-\frac{1}{2} \left[ K(q_i|z_i) + \sum_{j \in \partial_i} K(q_j|z_{\partial_j}) \right] \\ &= -\frac{1}{2} \left[ \sum_{k=1}^K q_{ik} \log q_{ik} - \sum_{k=1}^K q_{ik} \log z_{ik} + \sum_{j \in \partial_i, j \neq i} \left( \sum_{k=1}^K q_{ik} \log q_{jk} - \sum_{k=1}^K q_{ik} \log z_{\partial_j k} \right) \right] \end{aligned} \quad (53)$$

By ignoring the terms independent of  $z_{ik}$ , we get

$$-\frac{1}{2} \left[ -\sum_{k=1}^K q_{ik} \log z_{ik} - \sum_{j \in \partial_i, j \neq i} \sum_{k=1}^K q_{jk} \log z_{jk} \right] \quad (54)$$

Where

$$z_{\partial_j} = \sum_{m \in \partial_j, m \neq j} \alpha_{jm} z_m = \alpha_{jm} z_i + \sum_{m \in \partial_j, m \neq i, j} \alpha_{jm} z_m \quad (55)$$

To make the M-step tractable, we use Jensen's inequality to bound terms in Eq. (55):

$$\log z_{\partial_j k} = \log \sum_{m \in \partial_j, m \neq j} \alpha_{jm} z_{mk} \geq \alpha_{jm} \log z_{ik} + \log \sum_{m \in \partial_j, m \neq i, j} \alpha_{jm} z_m \quad (56)$$

Since  $\alpha_{ji} = \alpha_{ij}$ , by combining Eq. (54)–(56), we obtain:

$$\begin{aligned} & \frac{1}{2} \left[ -\sum_{k=1}^K q_{ik} \log z_{ik} - \sum_{j \in \partial_i, j \neq i} \sum_{k=1}^K q_{jk} \log z_{\partial_j k} \right] \\ &= \frac{1}{2} \left[ -\sum_{k=1}^K q_{ik} \log z_{ik} - \sum_{j \in \partial_i, j \neq i} \sum_{k=1}^K q_{jk} \log \left( \sum_{m \in \partial_j, m \neq j} \alpha_{jm} z_{mk} \right) \right] \\ &\geq \frac{1}{2} \left[ \sum_{k=1}^K q_{ik} \log z_{ik} + \sum_{j \in \partial_i, j \neq i} \left( \sum_{k=1}^K q_{jk} (\alpha_{ji} \log z_{ik}) + \sum_{m \in \partial_j, m \neq j} \alpha_{jm} \log z_{mk} \right) \right] \\ &= \frac{1}{2} \left[ \sum_{k=1}^K q_{ik} \log z_{ik} + \sum_{k=1}^K \sum_{j \in \partial_i, j \neq i} q_{jk} (\alpha_{ji} \log z_{ik}) + \sum_{j \in \partial_i, j \neq i} \sum_{m \in \partial_j, m \neq j} \alpha_{jm} \log z_{mk} \right] \end{aligned} \quad (57)$$

By only preserving the terms involving  $q_i$ , then the remaining terms of formula (63) are:

$$\begin{aligned} & \frac{1}{2} \left[ -\sum_{k=1}^K q_{ik} \log z_{ik} - \sum_{k=1}^K \sum_{j \in \partial_i, j \neq i} q_{jk} \log q_{jk} (\alpha_{ji} \log z_{ik}) \right] \\ &= \frac{1}{2} \left[ -\sum_{k=1}^K q_{ik} \log z_{ik} - \sum_{k=1}^K \sum_{j \in \partial_i, j \neq i} \alpha_{ji} q_{jk} \log z_{ik} \right] \\ &= \frac{1}{2} \left[ \sum_{k=1}^K \left( q_{ik} + \sum_{j \in \partial_i, j \neq i} \alpha_{ji} q_{jk} \right) \log z_{ik} \right] \\ &\Rightarrow \frac{1}{2} \sum_{k=1}^K \left( q_{ik} + q_{\partial_i k} \right) \log z_{ik} \end{aligned} \quad (58)$$

Where the distribution  $q_{\partial_i}$  is

$$q_{\partial_i} = \sum_{j \in \partial_i, j \neq i} \alpha_{ij} q_j \quad (59)$$

By applying the similar derivation holds for  $\pi_i$ , we can get:

$$\beta \sum_{k=1}^K (s_{ik} + s_{\partial_i k}) \log \pi_{ik} \quad (60)$$

Where the distribution  $s_{\partial_i}$  is:

$$s_{\partial_i} = \sum_{j \in \partial_i, j \neq i} \alpha_{ij} s_j \quad (61)$$

Consequently, the lower bound of complete log-likelihood function  $Q$  involving the posterior  $z_i$  and prior  $\pi_i$  becomes:

$$\log \sum_{k=1}^K \{\pi_{ik} p(\mathbf{x}_i | \Theta)\} + \beta \sum_{k=1}^K (s_{ik} + s_{\partial_i k}) \log \pi_{ik} + \frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) \log z_{ik} \quad (62)$$

In Eq. (62),  $\frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) = \frac{1}{2} \sum_{k=1}^K q_{ik} + \frac{1}{2} \sum_{k=1}^K q_{\partial_i k} = 1$ . By expanding the posterior  $z_{ik}$ , we find that maximizing

$$\begin{aligned} & \log \sum_{k=1}^K \{\pi_{ik} p(\mathbf{x}_i | \Theta)\} + \beta \sum_{k=1}^K (s_{ik} + s_{\partial_i k}) \log \pi_{ik} + \frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) \log z_{ik} \\ = & \log \sum_{k=1}^K \{\pi_{ik} p(\mathbf{x}_i | \Theta)\} + \frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) \log z_{ik} + \beta \sum_{k=1}^K (s_{ik} + s_{\partial_i k}) \log \pi_{ik} \\ = & \log \sum_{k=1}^K \{\pi_{ik} p(\mathbf{x}_i | \Theta)\} + \frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) \log \left\{ \frac{\pi_{ik} p(\mathbf{x}_i | \Theta)}{\sum_{k=1}^K \pi_{ik} p(\mathbf{x}_i | \Theta)} \right\} + \beta \sum_{k=1}^K (s_{ik} + s_{\partial_i k}) \log \pi_{ik} \\ = & \frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) \log (\pi_{ik} p(\mathbf{x}_i | \Theta)) + \beta \sum_{k=1}^K (s_{ik} + s_{\partial_i k}) \log \pi_{ik} \end{aligned} \quad (63)$$

is equivalent to maximizing:

$$\frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) \log p(\mathbf{x}_i | \Theta) + \sum_{k=1}^K \left( \frac{1}{2} (q_{ik} + q_{\partial_i k}) + \beta (s_{ik} + s_{\partial_i k}) \right) \log \pi_{ik} \quad (64)$$

Then, the Lagranges multiplier is used to enforce the constraint  $\sum_{k=1}^K \pi_{ik} = 1$  for each data point, and we can easily get the updating function for the prior  $\pi_{ik}$ :

$$\hat{\pi}_{ik} = \frac{1}{1 + 2\beta} \left( \frac{1}{2} (q_{ik} + q_{\partial_i k}) + \beta (s_{ik} + s_{\partial_i k}) \right) \quad (65)$$

Similarly, we obtain the following update equations for  $\mu$ ,  $\Gamma$  and  $\Delta$ , and. The energy function can be rewritten as:

$$\begin{aligned} Q^* &= \frac{1}{2} \sum_{K=1}^K (q_{ik} + q_{\partial_i k}) \log p(\mathbf{x}_i | \Theta) + \sum_{K=1}^K \left( \frac{1}{2} (q_{ik} + q_{\partial_i k}) + \beta (s_{ik} + s_{\partial_i k}) \right) \log \pi_{ik} \\ &= \frac{1}{2} \sum_{K=1}^K (q_{ik} + q_{\partial_i k}) \log \left\{ \frac{1}{\sqrt{(2\pi)^n |\Gamma_k|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_k - \Delta_k \hat{s}_{i,k})^T \Gamma_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k - \Delta_k \hat{s}_{i,k}) - \frac{1}{2} (\hat{s}_{i,k}^2 - (\hat{s}_{i,k})^2) \Delta_k^T \Gamma_k^{-1} \Delta_k - \frac{1}{2} \hat{s}_{i,k}^2 \right\} \right\} \\ &\quad + \sum_{K=1}^K \left( \frac{1}{2} (q_{ik} + q_{\partial_i k}) + \beta (s_{ik} + s_{\partial_i k}) \right) \log \pi_{ik} \end{aligned} \quad (66)$$

Let  $\frac{\partial Q^*}{\partial \boldsymbol{\mu}_k} = 0$ ,  $\frac{\partial Q^*}{\partial \Gamma_k^{-1}} = 0$  and  $\frac{\partial Q^*}{\partial \Delta_k} = 0$ , we can obtain:

$$\hat{\boldsymbol{\mu}}_k = \frac{\sum_{i=1}^N (q_{ik} + q_{\partial_i k}) (\mathbf{x}_i - \Delta_k \hat{s}_{i,k})}{\sum_{i=1}^N (q_{ik} + q_{\partial_i k})} \quad (67)$$

$$\hat{\Delta}_k = \frac{\sum_{i=1}^N (q_{ik} + q_{\partial_i k}) \hat{s}_{i,k} (\mathbf{x}_i - \boldsymbol{\mu}_k)}{\sum_{i=1}^N (q_{ik} + q_{\partial_i k}) \hat{s}_{i,k}^2} \quad (68)$$

$$\hat{\Gamma}_k = \frac{\sum_{i=1}^N (q_{ik} + q_{\partial_i k}) ((\mathbf{x}_i - \boldsymbol{\mu}_k - \Delta_k \hat{s}_{i,k}) (\mathbf{x}_i - \boldsymbol{\mu}_k - \Delta_k \hat{s}_{i,k})^T + (\hat{s}_{i,k}^2 - (\hat{s}_{i,k})^2) \Delta_k \Delta_k^T)}{\sum_{i=1}^N (q_{ik} + q_{\partial_i k})} \quad (69)$$

Where

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Gamma}_k + \Delta_k \Delta_k^T \quad (70)$$

$$\lambda_k = \frac{\boldsymbol{\Sigma}_k^{-1/2} \Delta_k}{\sqrt{1 - \Delta_k^T \boldsymbol{\Gamma}_k^{-1} \Delta_k}} \quad (71)$$

## References

- [1] Maciel Zortea, Eliezer Flores, Jacob Scharcanski, A simple weighted thresholding method for the segmentation of pigmented skin lesions in macroscopic images, *Pattern Recognition* 64 (2017) 92–104.
- [2] Siu Kai Choy, Shu Yan Lam, Kwok Wai Yu, Wing Yan Lee, and King Tai Leung, Fuzzy model-based clustering and its application in image segmentation, *Pattern Recognition*, 68 (2017) 141–157..
- [3] Nabanita Mahata, Sayan Kahali, Sudip Kumar Adhikari, Jamuna Kanta Sing, Local contextual information and gaussian function induced fuzzy clustering algorithm for brain mr image segmentation and intensity inhomogeneity estimation, *Applied Soft Computing* 68 (2018) 586–596.
- [4] Hu, Can, Wentao Fan, Du. Jixiang, Yuchen Zeng, Model-based segmentation of image data using spatially constrained mixture models, *Neurocomputing* 283 (2018) 214–227.
- [5] Dusan Stosic, Darko Stosic, Teresa Bernarda Ludermir, Tsang Ing Ren, Natural image segmentation with non-extensive mixture models, *Journal of Visual Communication and Image Representation* 63 (2019) 102598.
- [6] A.K. Jain, R.P.W. Duin, Jianchang Mao, Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1) (2000) 4–37..
- [7] Zexuan Ji, Yubo Huang, Yong Xia, Yuhui Zheng, A robust modified gaussian mixture model with rough set for image segmentation, *Neurocomputing* 266 (2017) 550–565.
- [8] D.M. Titterington, A.F.M. Smith, U.E. Makov, *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York, 1985.
- [9] M.S. Allili, D. Ziou, N. Bouguila, S. Boutemedjet, Image and video segmentation by combining unsupervised generalized gaussian mixture modeling and feature selection, *IEEE Transactions on Circuits and Systems for Video Technology* 20 (10) (Oct 2010) 1373–1377.
- [10] T.M. Nguyen, Q.M.J. Wu, D. Mukherjee, H. Zhang, A bayesian bounded asymmetric mixture model with segmentation application, *IEEE Journal of Biomedical and Health Informatics* 18 (1) (2014) 109–119.
- [11] Zexuan Ji, Yubo Huang, Quansen Sun, Guo Cao, A spatially constrained generative asymmetric gaussian mixture model for image segmentation, *Journal of Visual Communication and Image Representation* 40 (2016) 611–626.
- [12] Hao Sun, Xianqiang Yang, Huijun Gao, A spatially constrained shifted asymmetric laplace mixture model for the grayscale image segmentation, *Neurocomputing* 331 (2019) 50–57.
- [13] Adelchi Azzalini, Skew-Normal Family of Distributions, American Cancer Society (2015) 1–8.
- [14] Chunzheng Cao, Mengqian Chen, Yuqian Ren, Xu. Yue, Robust replicated heteroscedastic measurement error model using heavy-tailed distribution, *Communications in Statistics – Simulation and Computation* 47 (6) (2018) 1771–1784.
- [15] F. Forbes, N. Peyrard, Hidden markov random field model selection criteria based on mean field-like approximations, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (9) (2003) 1089–1101.
- [16] Chi Liu, Hengchao Li, Kun Fu, Fan Zhang, Mihai Datcu, William J. Emery, Bayesian estimation of generalized gamma mixture model based on variational em algorithm, *Pattern Recognition*, 87 (2019) 269–284..
- [17] A. Diplaros, Nikos Vlassis, Theo Gevers, A spatially constrained generative model and an em algorithm for image segmentation, *IEEE Transactions on Neural Networks* 18 (3) (2007) 798–808.
- [18] Thanh Minh Nguyen, Q.M. Jonathan Wu, Fast and robust spatially constrained gaussian mixture model for image segmentation, *IEEE Transactions on Circuits and Systems for Video Technology* 23 (4) (2013) 621–635.
- [19] Zexuan Ji, Jinyao Liu, Hengdong Yuan, Yubo Huang, Quansen Sun, A spatially constrained asymmetric gaussian mixture model for image segmentation, in: Pacific-Rim Symposium on Image and Video Technology, 2015..
- [20] Sotirios Ronneberger, Olaf and Fischer, Philipp, Brox, Thomas, U-Net: Convolutional Networks for Biomedical Image Segmentation International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, 234–241..
- [21] B.C. Franczak, R.P. Browne, P.D. McNicholas, Mixtures of shifted asymmetriclaplace distributions, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36 (6) (2014) 1149–1157.
- [22] Adelchi Azzalini, The skew-normal distribution and related multivariate families, *Scandinavian Journal of Statistics* 32 (2) (2005) 159–188.
- [23] M.S. Allili, Wavelet modeling using finite mixtures of generalized gaussian distributions: Application to texture discrimination and retrieval, *IEEE Transactions on Image Processing* 21 (4) (2012) 1452–1464.
- [24] J. Weickert, B.M.T.H. Romeny, M.A. Viergever, Efficient and reliable schemes for nonlinear diffusion filtering, *IEEE Transactions on Image Processing* 7 (3) (1998) 398–410.
- [25] Yunjie Chen, Hui Zhang, Yuhui Zheng, Byeungwoo Jeon, Q.M. Jonathan Wu, An improved anisotropic hierarchical fuzzy c-means method based on multivariate student t-distribution for brain mri segmentation, *Pattern Recognition*, 60 (2016) 778–792..
- [26] Sharon X Lee, Geoffrey J Mclachlan, On mixtures of skew normal and skew t-distributions, *Advanced Data Analysis and Classification* 7 (3) (2013) 241–266.
- [27] Sotirios Monteiro, C. Fernando, Aurélio C. Campilho, Performance Evaluation of Image Segmentation, in: Springer on International Conference Image Analysis and Recognition, 2006, pp. 248–259.
- [28] D.C. Sotirios Stanford, A.E. Raftery, Approximate Bayes Factors for Image Segmentation: The Pseudolikelihood Information Criterion (PLIC), *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 24 (11) (2002) 1517–1520.
- [29] Sotirios Demiraj, Janez, Schuurmans, D., Statistical comparisons of classifiers over multiple data sets, *Journal of Machine Learning Research*, 7(1) (2006) 1–30..
- [30] Sotirios P Chatzis, Theodora Varvarigou, A fuzzy clustering approach toward hidden markov random field models for enhanced spatially constrained image segmentation, *IEEE Transactions on Fuzzy Systems* 16 (5) (2008) 1351–1361.