TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 05/13 pp833–844
DOI: 10.26599/TST.2020.9010042
Volume 26, Number 6, December 2021

A Chan–Vese Model Based on the Markov Chain for Unsupervised Medical Image Segmentation

Quanwei Huang, Yuezhi Zhou*, Linmi Tao*, Weikang Yu, Yaoxue Zhang, Li Huo, and Zuoxiang He

Abstract: The accurate segmentation of medical images is crucial to medical care and research; however, many efficient supervised image segmentation methods require sufficient pixel level labels. Such requirement is difficult to meet in practice and even impossible in some cases, e.g., rare Pathoma images. Inspired by traditional unsupervised methods, we propose a novel Chan–Vese model based on the Markov chain for unsupervised medical image segmentation. It combines local information brought by superpixels with the global difference between the target tissue and the background. Based on the Chan–Vese model, we utilize weight maps generated by the Markov chain to model and solve the segmentation problem iteratively using the min-cut algorithm at the superpixel level. Our method exploits abundant boundary and local region information in segmentation and thus can handle images with intensity inhomogeneity and object sparsity. In our method, users gain the power of fine-tuning parameters to achieve satisfactory results for each segmentation. By contrast, the result from deep learning based methods is rigid. The performance of our method is assessed by using four Computerized Tomography (CT) datasets. Experimental results show that the proposed method outperforms traditional unsupervised segmentation techniques.

Key words: medical image; unsupervised segmentation; Markov chain

1 Introduction

As a key step in medical research, the accurate segmentation of medical images is commonly applied in preoperative planning, intraoperative navigation, and postoperative assessment[1]. Good segmentation divides the image into multiple regions with a homogeneous color (or texture), and the boundaries are "simple" and "spatially accurate"[2]. However, medical image segmentation remains challenging for the following three technical reasons:

(1) Objects in medical images are far less clear and inhomogeneous than those in natural images. Generally, most medical imaging technologies involve indirect computerized imaging. For instance, Computerized Tomography (CT) uses a series of X-rays and is combined with sophisticated mathematical techniques to obtain image representation[3]. The resulting image contains only one channel with low gradient information to facilitate segmentation. In medical images, the contrast between the target tissue and the background is shallow, and the tissue boundary with noise is ambiguous even to experienced doctors.

(2) Considerable attention has been recently focused on designing supervised deep segmentation algorithms

• Quanwei Huang, Yuezhi Zhou, Linmi Tao, and Yaoxue Zhang are with BNRist; and Key Laboratory of Pervasive Computing (Ministry of Education) and the Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China. E-mail: hqw19@mails.tsinghua.edu.cn; {zhouyz, linmi, zhangyx}@mail.tsinghua.edu.cn.
• Weikang Yu is with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China. E-mail: yuweikang@buaa.edu.cn.
• Li Huo is with the Department of Nuclear Medicine, Peking Union Medical College Hospital, Beijing 100730, China. E-mail: huoli@pumch.cn.
• Zuoxiang He is with the School of Clinical Medicine, Tsinghua University, and also with Beijing Tsinghua Changgung Hospital, Beijing 100084, China. E-mail: zuoxianghe@hotmail.com.
∗To whom correspondence should be addressed.
Manuscript received: 2020-08-17; accepted: 2020-09-23

for medical images[4, 5]. One popular recent approach is the U-net[6] architecture, a fully convolutional network that has been used to achieve impressive results in the biomedical image domain. These segmentation methods require a significant amount of pixel-wise labeled training data. Unfortunately, sufficient supervised pixel level labels are hardly obtained in many domains of medicine[7].

(3) In practical application, an accurate segmentation is necessary. Although deep segmentation networks can produce good results for general images, the segmentation result of deep learning based methods is rigid, which means that users cannot improve or adjust the segmentation error manually. Furthermore, the effectiveness of deep learning based methods is limited by the machine's inability to explain its decisions and actions to users[8].

Given the difficulty of obtaining supervised samples of medical images and the complexity of medical image imaging, many segmentation methods developed in the past decades have demonstrated that the straightforward segmentation of medical images at the pixel level achieves minor progress in terms of accuracy and cannot meet the demands of emerging precision medicine in practice. By contrast, unsupervised learning does not require annotated data. It automatically analyzes a large amount of data and learns its internal data structure. At present, unsupervised learning in deep learning is mainly divided into two categories: a deterministic self-coding method and a probabilistic method limited by a Boltzmann machine.

Given the importance of the segmentation problem in medical domains and the lack of supervised data for many problems, we revisit the problem of unsupervised image segmentation, utilizing classical ideas from the Chan–Vese model. In an unsupervised scenario, image segmentation is used to predict general labels, such as "foreground" and "background", especially without any prior information[9]. Several well-known techniques, including normalized cuts[10, 11], Markov random field based methods[12], hierarchical methods[13], mean shift[14], and active contour methods[15], are used for this problem. These algorithms aim to segment the object instance in each image automatically without any supervision.

We propose a novel method for unsupervised medical image segmentation on the basis of the Chan–Vese model[16], which is widely used for medical images. We divide the input image into nonoverlapping superpixels and construct an undirected graph equivalent to the image. Then, we use the min-cut algorithm iteratively to solve and obtain a segmentation result.

To maximize image information for processing images with intensity inhomogeneity, we fuse local and global information and consider the spatial and color similarities of pixels simultaneously. Our unsupervised method merges neighboring pixels into a superpixel by using the local information of the neighboring pixels. Then, the image is segmented at the superpixel level, supported by the global difference between the target tissue and the background. As the segmentation processing unit, the superpixel helps eliminate abnormal pixels, restrain noise and ineffective signals, and remove redundant information at the local and global levels[17].

Specifically, we set up a weight map based on the Markov chain for each superpixel, thereby allowing our method to combine the spatial and color (or texture) similarities of pixels well. Thus, our method can perform well in images with intensity inhomogeneity. We leverage the expected round trip steps of the Markov chain between this node and other nodes to establish a weight map for each superpixel node, representing the influence of other nodes on the classification (or segmentation) of this node.

In the experiments, we test our method on real 2D CT images and achieve the highest accuracy for the majority of the experiments. The details of the applied methods and the results are illustrated in the subsequent sections.

2 Related Work

Mature image segmentation methods have been introduced into the field of medical image processing. Mcinerney and Terzopoulos[18] and Pham et al.[19] summarized the segmentation methods used in medical images in the 20th century. Studies in the new century have addressed medical image segmentation intensively because of its emerging real-world applications. Deep learning based methods have been introduced to medical image segmentation in recent years, but only a few of them can produce good results. Many end-to-end deep learning models[20, 21] cannot achieve accurate segmentation due to the lack of available medical data, the difficulty of data labeling, and the absence of a fixed set of features that are known *a priori* to segment new images correctly. Moreover, even when we have a large amount of labeled data available, deep learning methods inevitably produce incorrect segmentation for some images; this error is unacceptable in practical

applications because the results cannot be corrected.

Most unsupervised image segmentation approaches involve utilizing features, such as color, brightness, or texture over local patches, and then perform pixel level clustering based on these features. Among these schemes, the most widely used methods include mean shift, normalized cuts, and graph-based methods[22, 23]. These methods classify pixels according to the global distribution of pixel values via iterative energy minimization. As a heavily nonuniform distribution of objects in medical images, this process takes too long and needs postprocessing to achieve good results.

Extensions of the Chan–Vese model have received considerable attention[16, 24–26]. The Chan–Vese model is based on the Mumford–Shah function for segmentation and is used widely in the medical imaging field, especially for the segmentation of images of the brain, heart, and trachea. The objective of the Chan–Vese algorithm is to minimize the energy functional $F(c_1, c_2, C)$, which is defined by

$$F(c_1,c_2,C)=\mu\cdot\text{Length}(C)+v\cdot\text{Area}(\text{inside}(C))+\lambda_1\int_{\text{inside}(C)}|u_0(x,y)-c_1|^2\,\mathrm{d}x\mathrm{d}y+\lambda_2\int_{\text{outside}(C)}|u_0(x,y)-c_2|^2\,\mathrm{d}x\mathrm{d}y \quad (1)$$

where $\mu, v, \lambda_1, \lambda_2$ are fixed parameters, and C is a piecewise $C^1[0,\ 1]$ parameterized curve. Moreover, c_1 denotes the average pixels' intensity inside C, and c_2 denotes the average intensity outside C. The energy function can be reformulated in the level set formulation, leading to a convenient way to solve the problem. To make the model obtain an improved effect in practical application, DRLSE[27] proposes distance regularization that eliminates the need to reinitialize level set functions in iterations. LBF[24] introduces a local binary fitting energy with a kernel function, enabling the extraction of accurate local image information. However, these methods ignore the effect of pixels with similar gray values but are far apart and unsuitable for the segmentation of inhomogeneous and complex objects.

3 Method

Our method is mainly divided into three steps, as shown in Fig. 1. In the first step, the input image is oversegmented to obtain superpixels. In the second step, an undirected graph corresponding to superpixels is constructed, weight maps based on the Markov chain are generated, and the min-cut algorithm is used for solving. Lastly, we combine the results of the two parts to obtain the final segmentation result. These steps are discussed below in detail.

3.1 Superpixel segmentation

A superpixel is a set of pixels with similar colors or gray levels, which are spatially adjacent but nonoverlapping. Compared with the pixel representation of an image, superpixel representation provides much more cues and is more consistent with the input image. Furthermore, it considerably reduces the number of image primitives and improves computational efficiency. Many superpixel segmentation algorithms[28–30] are available, and all of them work well. In our study, for increased execution efficiency, we use Simple Linear Iterative Clustering (SLIC)[28], which utilizes a K-means clustering approach to generate compact superpixels efficiently and has a

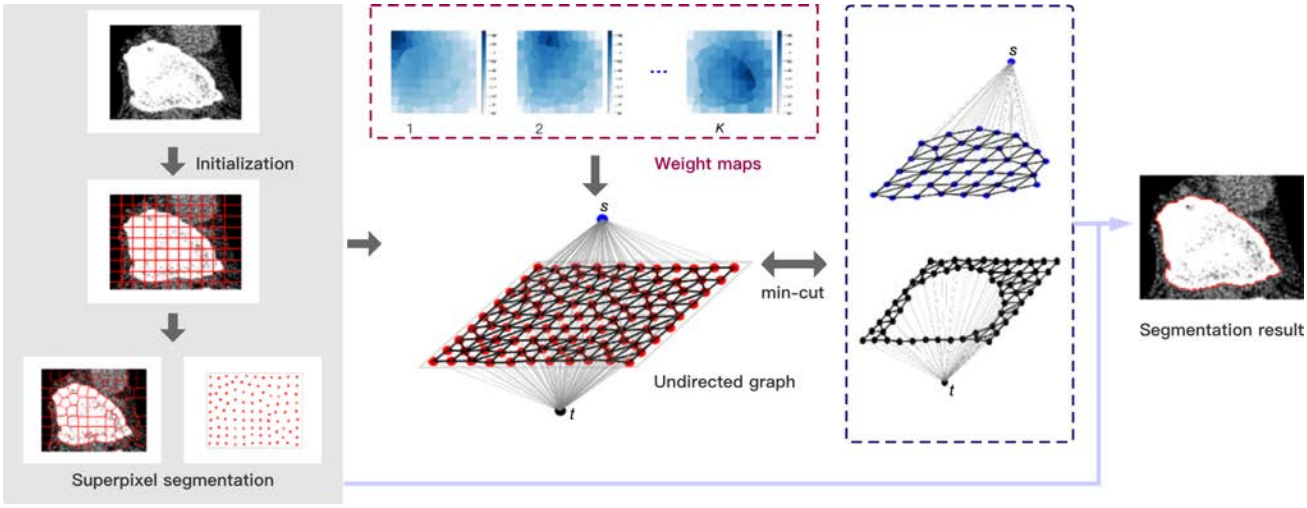


Fig. 1 Overall segmentation flow diagram. Here, s represents foreground node and t represents background node.

complexity of $O(N)$, where N is the number of pixels.

We initialize cluster centers evenly and divide the input image into K regions via SLIC. Each region contains approximately 10 pixels in our experiments. Using too many pixels may produce poor results. Notably, precise segmentation cannot be achieved at the pixel level, similar to the kernel method. We use a superpixel to integrate the information of neighboring pixels into a high-dimensional information representation wherein we can segment easily.

3.2 Undirected graph construction

Let $\mathcal{G} = (V, E)$ be an undirected graph defined on a given superpixel set, which represents a weighted graph containing a set of nodes V and edges $E \subseteq V \times V$. Then, every superpixel x_i is identified uniquely by a node vertex $v_i \in V$ in our undirected graph, where two special nodes $s, t \in V$ represent the foreground and background nodes, respectively. For convenience of description, $V \setminus \{s, t\}$ is represented by V_s, where the node value is equal to the average gray value of the corresponding superpixel. Similarly, edge E can be divided into E_1 and E_2, where $E_1 \subseteq V_s \times V_s$ and $E_2 \subseteq \{s, t\} \times V_s$. To represent the image structure by edge weights, one must define a function that maps a change in image intensities to edge weights. In this work, we use the typical Gaussian weighting function given by

$$G(x) = \exp(-\frac{x^2}{2\sigma^2}) \tag{2}$$

where σ is a parameter defined by the user.

$A = (a_1, \ldots, a_i, \ldots, a_K), a_i \in \{0, 1\}$ is a vector whose component a_i specifies assignments to node v_i in V_s. Vector A defines a segmentation. If a_i is equal to 0, then node v_i belongs to the foreground; otherwise, it belongs to the background.

Two kinds of edges exist in the graph; for $e_{ij} \in E_1$, its weight is defined as follows:

$$e_{ij} = \begin{cases} \alpha \cdot G(\|g_i - g_j\|) + l_{ij}, & \text{if } i \sim j; \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $i \sim j$ means nodes v_i and v_j are adjacent nodes; and g_i and g_j denote the image pixel values at nodes v_i and v_j, respectively. l_{ij} represents the length of the adjacent boundary of the superpixels corresponding to v_i and v_j. α is a weighting factor. $e_{ij} \in E_1$ represents the correlation between nodes; the greater the value, the greater the probability that the nodes are in the same category. $e \in E_2$ represents the similarity between node $v \in V_s$ and the foreground or background. For edges that belong to E_2, we define their weight as follows:

$$\begin{aligned} e_{si} &= \lambda_{g1}(g_i - g_t)^2 + \lambda_{l1} c_{i1}; \\ e_{it} &= \lambda_{g0}(g_i - g_s)^2 + \lambda_{l0} c_{i0} \end{aligned} \tag{4}$$

where g_s and g_t represent the pixel values of nodes v_s and v_t, whose values are equal to the average pixel values of the nodes belonging to the foreground and background, respectively. c_{i0} and c_{i1} represent the local properties of v_i. $\lambda_{g1}, \lambda_{l1}, \lambda_{g0}$, and λ_{l0} are fixed parameters.

To calculate c_{i0} and c_{i1}, for $v_i \in V_s$, the degree of each node is computed as $d_i = \sum_{j \in V_s} w_{ij}$. For all the edges that are incident on v_i, w_{ij} measures the similarity between two neighboring nodes v_i and v_j. Thus we define $w_{ij} = G(\|g_i - g_j\|)$. The transition probability matrix is obtained as follows:

$$P_{ij} = \begin{cases} w_{ij}/d_i, & \text{if } i \sim j; \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

According to spectral graph theory, the Markov chain converges to a unique stationary distribution:

$$\pi_i = d_i / \sum_{j \in V_s} d_j \tag{6}$$

Hence, we can obtain

$$Z = (I - P + W)^{-1}$$

where I is an identity matrix, P is the transition probability matrix, and W is equal to P^∞, and each of its row vectors is equal to π, where $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$. Then, we derive the expected number of steps $\mathrm{ET}_i(T_j)$ from nodes v_i to v_j:

$$\mathrm{ET}_i(T_j) = \begin{cases} \dfrac{1}{\pi_j} \times (Z_{jj} - Z_{ij}), & \text{if } i \neq j; \\ \dfrac{1}{\pi_j}, & \text{if } i = j \end{cases} \tag{7}$$

We define a matrix $T_{ij} = \mathrm{ET}_i(T_j) + \mathrm{ET}_j(T_i)$, where T_{ij} is called the expected round trip steps between nodes v_i and v_j. After row-normalizing the matrix, we obtain matrix $\mathcal{T}$. Lastly, we calculate the local properties c_{i0}, c_{i1} of node v_i as follows:

$$\begin{aligned} c_{i0} &= \frac{\sum_{a_j=0} G(\mathcal{T}_{ij})(g_i - g_j)^2}{\sum_{a_j=0} G(\mathcal{T}_{ij})}; \\ c_{i1} &= \frac{\sum_{a_j=1} G(\mathcal{T}_{ij})(g_i - g_j)^2}{\sum_{a_j=1} G(\mathcal{T}_{ij})} \end{aligned} \tag{8}$$

Each row of $G(\mathcal{T})$ represents the influence of other nodes on the segmentation of a particular node, which is called a weight map.

In summary, we can derive K weight maps, in which one superpixel node corresponds to a weight map.

In almost all our experiments, when calculating the transition probability matrix P, we fix the parameter σ of the Gaussian weighting function to 0.1. If we want to focus on the difference between the node pixel values, then we can set a smaller value of this function. The result of the weight map of a specific superpixel node under different parameters σ is shown in Fig. 2. The darker the color (the bluer), the higher the weight, that is, the greater the influence during the segmentation. In the case of $\sigma = 0.1$, even if some areas are far away but the gray value is similar to the superpixel at the red dot, their weight is also considerable. With the increase in σ, the proportion of color similarity gradually decreases, and the proportion of distance similarity increases.

Thereafter, all edges are weighted, and an undirected graph is constructed.

3.3 Segmentation

The partition problem can be converted into a process of minimizing energy. With the Chan–Vese model[16], we define two energies as fitting terms $\mathcal{E}_{\mathrm{f}}$ and regularization terms $\mathcal{E}_{\mathrm{r}}$ as follows:

$$\begin{aligned} \mathcal{E}_{\mathrm{f}} &= \sum_{a_i=1} e_{si} + \sum_{a_i=0} e_{it}; \\ \mathcal{E}_{\mathrm{r}} &= \sum_{v_i, v_j \in V_s; a_i \neq a_j} e_{ij} \end{aligned} \tag{9}$$

The fitting term $\mathcal{E}_{\mathrm{f}}$ represents the differences within the foreground and background. When the intensity of the pixels inside the foreground and background is uniform, these terms reach the minimum value. The regularization term $\mathcal{E}_{\mathrm{r}}$ represents the difference between the foreground and background. As a penalty term, the smaller the value is, the more significant the difference between the foreground and background is.

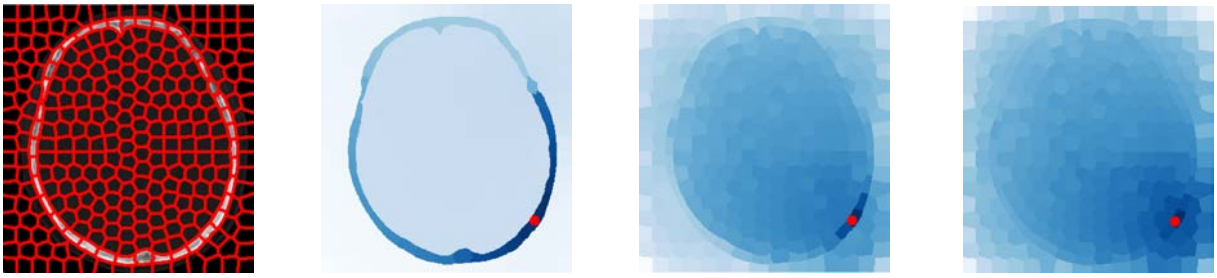
Our goal is to minimize $\mathcal{E}_{\mathrm{f}} + \mu\mathcal{E}_{\mathrm{r}}$, where the coefficient $\mu \geqslant 0$ specifies a relative importance of the fitting term $\mathcal{E}_{\mathrm{f}}$ versus the regularization term $\mathcal{E}_{\mathrm{r}}$. We initialize a segmentation vector A, in which several of the nodes that belong to our desired target are initialized as foreground nodes and calculate the edge weights e_{si} and e_{it} on the basis of vector A. Then, the min-cut algorithm is used to divide the graph into two parts; the nodes connected to node v_s belong to the foreground, and the others belong to the background. If the min-cut algorithm is not changed or almost unchanged, the iteration stops; otherwise, the min-cut algorithm updates segmentation vector A, recalculates e_{si} and e_{it}, and continues the iteration.

During the entire segmentation process, we need to fine-tune a few parameters. As mentioned above, for superpixel segmentation, parameter K is set at approximately $N/10$, and the parameter σ in the weight maps is 0.1. For parameters, $\lambda_{g1}, \lambda_{l1}$ and $\lambda_{g0}, \lambda_{l0}$ have similar effects, respectively. The difference is that if the image is relatively uniform, the values of λ_{g0} and λ_{g1} should be increased; conversely, the values of λ_{l1} and λ_{l0} should be increased. If $\lambda_{l0} > \lambda_{l1}$ ($\lambda_{g0} > \lambda_{g1}$), then the result is more consistent within the foreground, and the range is smaller.

According to the segmentation result of the superpixel and the undirected graph, we obtain a superpixel-based segmentation result. Lastly, we mark all pixels in the same superpixel with the same segmentation result. For example, the result of superpixel S_i is a_i, and the result of all pixels belonging to S_i is a_i. The complete segmentation process is specified as Algorithm 1.

4 Experiment

To evaluate our unsupervised model's performance thoroughly, we conduct preliminary experiments on four datasets of varying difficulty and compare the results with those of other methods. In addition, we demonstrate the advantages of our proposed method over the Chan–Vese model in terms of initialization and iteration times.



(a) Superpixels (b) $\sigma = 0.1$ (c) $\sigma = 0.3$ (d) $\sigma = 0.6$

Fig. 2 Results of the weight maps under different parameters σ. (a) Result of the superpixels; and (b)–(d) weight maps corresponding to the superpixel at the red dot under different parameters σ, respectively.

Algorithm 1 Complete segmentation algorithm

Input: Image $I \in \mathbb{R}^{H\times W}$; number of superpixels K;
Output: Segmentation result $L \in \mathbb{R}^{H\times W}$

1: Partition I into superpixels $S_1, \ldots, S_K$. For some desired superpixel S_i, initialize its label $a_i = 0$, and the others with 1;
2: Construct an undirected graph with $K+2$ nodes, and establish a weight map for each superpixel node;
3: **repeat**
4: Update the value of the nodes V_s and V_t:
$$g_s = \frac{\sum_{a_i=0} g_i}{\sum_{a_i=0} \mathbf{1}}, \quad g_t = \frac{\sum_{a_i=1} g_i}{\sum_{a_i=1} \mathbf{1}}$$
5: Update the edge weights e_{si}, e_{it} according to g_s, g_t, and the weight maps:
$$e_{si} = \lambda_{g1}(g_i - g_t)^2 + \lambda_{l1} c_{i1};$$
$$e_{it} = \lambda_{g0}(g_i - g_s)^2 + \lambda_{l0} c_{i0}$$
6: The min-cut algorithm is used to obtain a new segmentation result A.
7: **until** The min-cut is not changed or is almost unchanged.
8: **for** $S_i \in \{S_1, S_2, ..., S_K\}$ **do**
9: **for** pixel $I_j \in S_i$ **do**
10: Label pixel I_j as a_i.
11: **end for**
12: **end for**
13: **return** Final pixel labels $L \in \mathbb{R}^{H\times W}$;

4.1 Datasets

• **CVIP:** This dataset is from Spine Web, which is an online collaborative platform. This dataset provides a total of 349 spinal CT scans and the corresponding ground truths for five patients[31]. In some subsequent experiments on parameters, we use 19 CT slices of one patient (referred to as **MINI-CVIP**).

• **LUNG:** This dataset contains 267 lung CT images and manual masks of the corresponding lungs; the images were acquired from an open competition, Finding and Measuring Lungs in CT Data.

• **SKULL**, **SPINE:** This dataset is provided by a hospital; it contains numerous CT images of the skull and spine, as well as manual segmentation results.

4.2 Experimental settings

We conduct several experiments on real CT scans to evaluate the performance of the proposed method. Our method does not utilize any annotation information for segmentation.

Evaluation metrics: The accuracies of segmented contours are quantified using the Dice Similarity Coefficient (DSC), which is calculated as follows:

$$\mathrm{DSC} = 2\frac{|M_1 \cap M_2|}{|M_1| + |M_2|} \tag{10}$$

The Jaccard Similarity Coefficient (JSC) is also adopted as an evaluation metric to evaluate the accuracy of the proposed segmentation method; JSC is defined as follows:

$$\mathrm{JSC} = \frac{|M_1 \cap M_2|}{|M_1 \cup M_2|} \tag{11}$$

where M_1 is the area segmented by the algorithm, and M_2 is the ground-truth area manually delineated.

Implementation details: Some of the experiments compare our proposed method with the supervised method U-net; for each dataset, 70% is used as the training set, and 30% is used as the testing set. All methods conduct performance evaluation on the testing set. Moreover, the spinal column is relatively smaller than that in the original CT scan with a size of 512×512. For a transparent and fair comparison, the segmentation is limited within a bounding box that contains the ground truth. Then, the DSC and JSC are computed in the region.

4.3 Results

4.3.1 Effect of superpixels

Superpixels are useful in a wide range of vision tasks, and we can fully take advantage of boundary information for subsequent segmentation due to the superpixel. We test the results of whether or not to use superpixels in our proposed method. In the case of not using superpixels, each node of the undirected graph corresponds to a particular pixel of the input image. As shown in Fig. 3, in the images with intensity inhomogeneity, nodes with coarser granularity have more boundary information. Therefore, the superpixels provide more cues compared with a single pixel, and the segmentation results do not easily fall into the local optima. Hence, we can derive additional global results.

In our method, we adopt the standard SLIC algorithm. Given the single-channel characteristics of medical images, we apply the algorithm on the original features

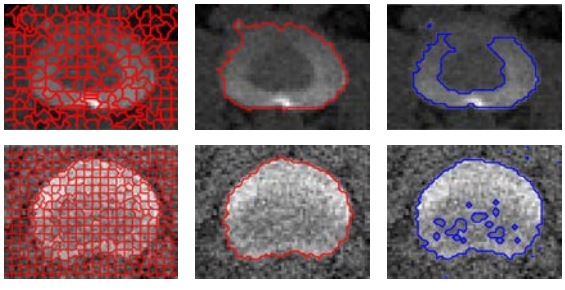


Fig. 3 Comparison of whether to use superpixels.

rather than the *XYLab* features. In the choice of parameters, we test the number of superpixels K on the dataset MINI-CVIP, and the experiment shows that as the value of K decreased (the number of pixels in each superpixel increased), the segmentation performance did not fluctuate within a certain range (as shown in Fig. 4) and then declined rapidly. The smaller K leads to more pixels in each superpixel, ultimately reducing the boundary segmentation precision. Therefore, for experimental accuracy and computational efficiency, we keep the number of pixels in each superpixel at approximately 10; that is, the value of the parameter K is approximately $N/10$.

4.3.2 Comparison of different methods

Table 1 shows the segmentation results of different methods. To demonstrate the merits of our model, we compare K-means clustering, the Chan–Vese model, the snake model, the Local Binary Fitting (LBF) model, and our method. In addition, we compare our method with the supervised method U-net on some datasets. In all the experiments, all methods are tested on the corresponding testing set. Table 1 shows that our method achieves the highest accuracy for majority of the datasets. The K-means method focuses on global losses and does not perform well on images with intensity inhomogeneity. Compared with the snake and Chan–Vese models, our method considers local information and shows better segmentation results in all experiments. LBF, an implicit active contour model driven by local binary fitting energy, performs best on the CVIP dataset. However, each image's processing time takes long due to the introduction of the convolution operation. Although U-net performs better on the SPINE dataset, this performance is mainly due to the fact that spines have roughly the same shape. Another reason is that an extensive training set with 87 training images is provided.

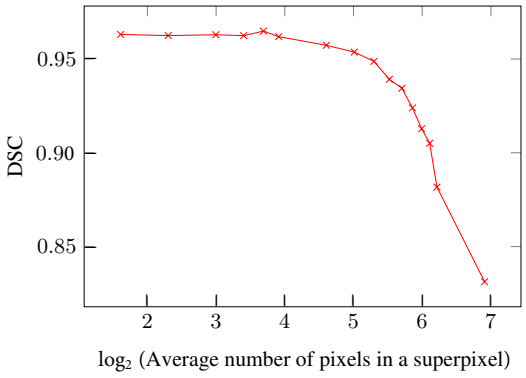


Fig. 4 Influence of different numbers of superpixels on the segmentation result.

As shown in Fig. 5, obtaining the separated bones is difficult because no clear distinction exists between the gap and bones; thus, other areas can be easily mistaken as bones. With our method, we can fit the target boundary well. In addition, the similarity between superpixels is used to enhance the local relevance to obtain a compact segmentation. To achieve a compact contour, the Chan–Vese model introduces a smoothing term, which leads to segmentation results that cannot fit the target boundary well, especially for images with intensity inhomogeneity. For the snake model, the result is often not locally smooth, and the final result may have many isolated contours.

Our proposed method has remarkable advantages in terms of initialization and iteration times. Compared with our model, the Chan–Vese model is more sensitive to initialization. In many cases, it does not converge to the target region. As shown in Fig. 6a, if the initial contour is around the target, then the algorithm easily obtains the desired segmentation; otherwise, as shown in Fig. 6b, the result is poor due to the failure to consider the remote pixel. In addition, as shown in Figs. 7 and 8, our method iterates less than ten times in all experiments compared with the Chan–Vese model, which requires dozens and even hundreds of iterations to achieve convergence. Moreover, our method has a certain iterative stop condition; that is, the min-cut of

Table 1 Segmentation accuracy of different methods on multiple datasets.

Method	CVIP		LUNG		SKULL		SPINE	
	DSC	JSC	DSC	JSC	DSC	JSC	DSC	JSC
K-means	0.918	0.902	0.817	0.811	0.901	0.841	0.921	0.918
Chan–Vese	0.965	0.934	0.836	0.821	0.919	0.851	0.970	0.941
Snake	0.954	0.924	0.823	0.801	0.898	0.837	0.956	0.947
LBF	**0.970**	**0.943**	0.865	0.841	0.921	0.849	0.975	0.954
U-net	0.958	0.929	–	–	0.912	0.843	**0.976**	**0.954**
Proposed method	0.968	0.939	**0.876**	**0.865**	**0.923**	**0.858**	0.973	0.948

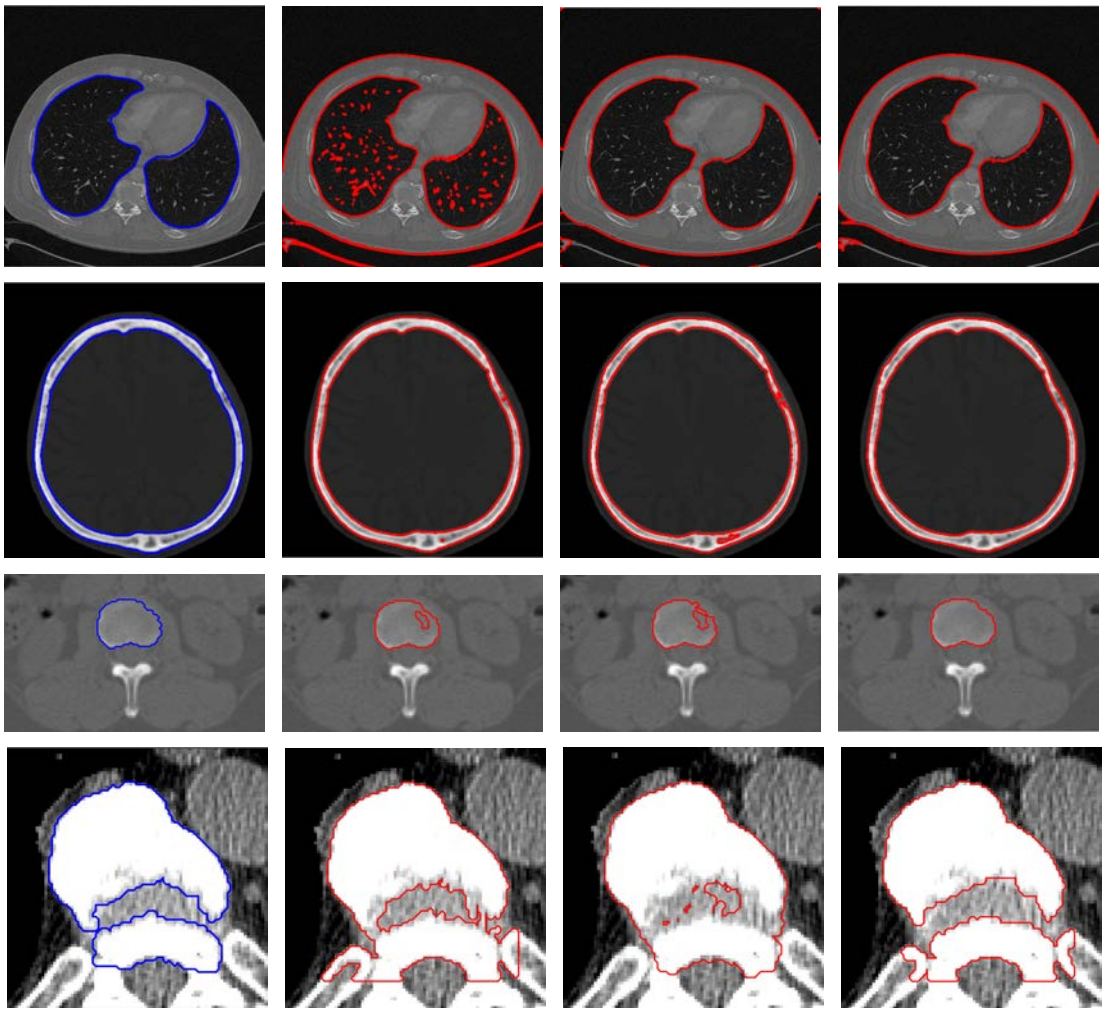


Fig. 5 Segmentation results of different methods. Each row represents a sample corresponding to the ground truth and the segmentation results under different methods.

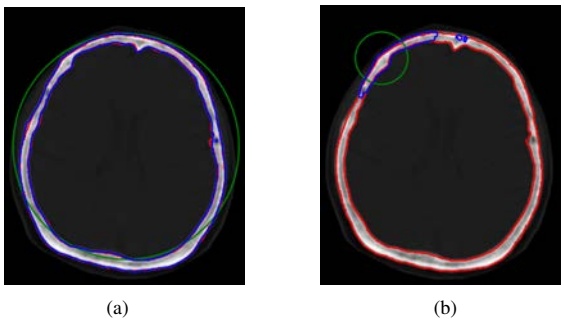


Fig. 6 Results of Chan–Vese model and our model under different initializations. The green line represents the initial contour, the blue line is the segmentation result of the Chan–Vese model, and the red line is the result of our model.

the undirected graph is no longer changed. However, for the Chan–Vese model, the evolution contour produces different results due to different iteration times and stop conditions. It is irreversible for the evolution contour to cross the target contour in the iteration process.

4.3.3 Ablation study

We further explore the parameters of our method. In practice, the segmentation results are adjusted mainly by four parameters ($\lambda_{g0}, \lambda_{g1}, \lambda_{l0}$, and λ_{l1}). λ_{l0} and λ_{g0} have the same function (λ_{l0} is mainly used for intensity inhomogeneous images) as λ_{l1} and λ_{g1} do. As shown in Fig. 9, by setting different λ_{l1} and λ_{l0}, we can find that the larger λ_{l0} is, the more uniform the interior of the foreground target becomes; conversely, the larger λ_{l1} is, the more consistent the interior of the background becomes. In addition, we conduct ablation studies, as shown in Table 2. Compared with using only global information, the weight maps we introduced result in great improvement because the weight maps mainly consider the pixels in the surrounding neighborhood. Therefore, in actual use,

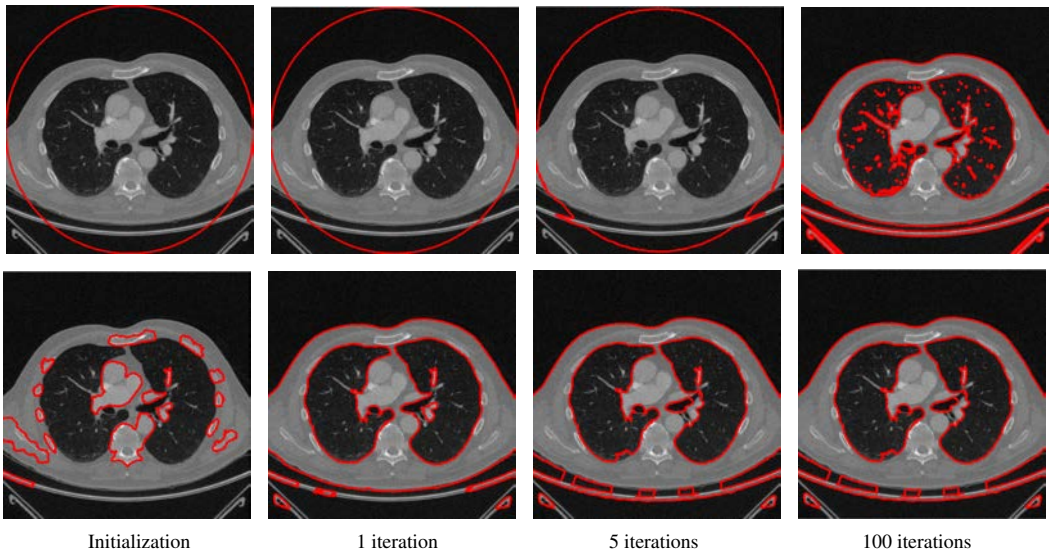


Fig. 7 Result of the Chan–Vese model and our model under different numbers of iterations.

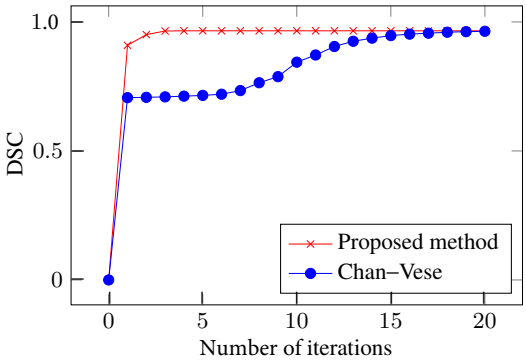


Fig. 8 Segmentation performance under different numbers of iterations (MINI-CVIP).

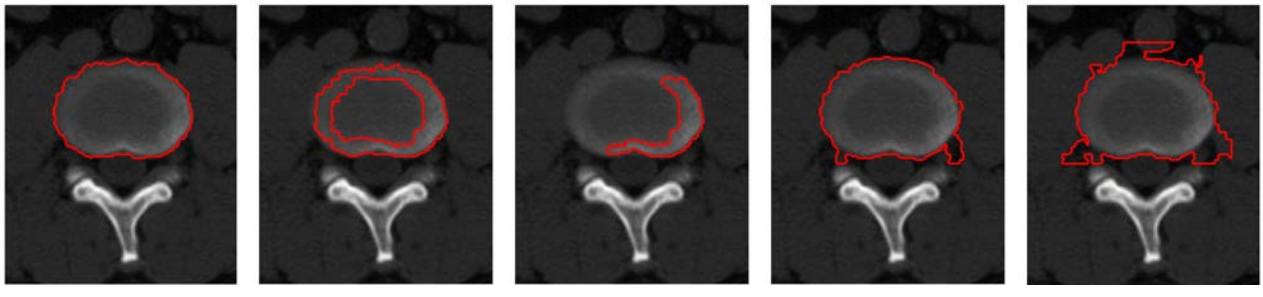
images with intensity inhomogeneity must be segmented by adjusting λ_{l1} and λ_{l0}.

5 Conclusion

A novel unsupervised method for medical image segmentation was proposed in this study. We projected the ambiguous boundary information at the pixel-level to the superpixel-level so that we could use additional clues for a robust and accurate segmentation. The spatial interaction and gray similarity of superpixels were modeled to establish weight maps via the Markov chain, which allows our method to adapt to various situations. Our method does not require any annotated data and gives the user the power of fine-tuning parameters for correcting segmentation errors and obtaining satisfactory segmentation results. The performance of the proposed method was assessed on four datasets and compared with that of other segmentation methods. Our method outperformed other segmentation algorithms in terms of DSC and JSC metrics. We will further test our method with doctors for in-field testing of medical applications.

Table 2 Ablation studies. Δ denotes the change of DSC compared with using global and local information simultaneously.

Parameter setting	DSC	Δ (%)
$\lambda_{g0}=1,\lambda_{g1}=1,\lambda_{l0}=1,\lambda_{l1}=1$	0.9656	0
$\lambda_{g0}=1,\lambda_{g1}=1,\lambda_{l0}=0,\lambda_{l1}=0$	0.9404	−2.52
$\lambda_{g0}=0,\lambda_{g1}=0,\lambda_{l0}=1,\lambda_{l1}=1$	0.9551	−1.05



(a) $\lambda_{l0}=1,\lambda_{l1}=1$ (b) $\lambda_{l0}=10,\lambda_{l1}=1$ (c) $\lambda_{l0}=20,\lambda_{l1}=1$ (d) $\lambda_{l0}=1,\lambda_{l1}=10$ (e) $\lambda_{l0}=1,\lambda_{l1}=20$

Fig. 9 Segmentation results under different parameters.

Acknowledgment

This paper was supported by the National Natural Science Foundation of China (Nos. 61672017 and 61272232), and the Key-Area Research and Development Program of Guangdong Province (No. 2019B010137005).

References

[1] C. T. Yeo, T. Ungi, P. U. Thainual, A. Lasso, R. C. McGraw, and G. Fichtinger, The effect of augmented reality training on percutaneous needle placement in spinal facet joint injections, *IEEE Trans. Bio-Med. Eng.*, vol. 58, no. 7, pp. 2031–2037, 2011.

[2] A. Y. Yang, J. Wright, Y. Ma, and S. Sastry, Unsupervised segmentation of natural images via lossy data compression, *Comput. Vis. Image Underst.*, vol. 110, no. 2, pp. 212–225, 2008.

[3] L. W. Goldman, Principles of CT and CT technology, *J. Nucl. Med. Technol.*, vol. 35, no. 3, pp. 115–128, 2007.

[4] B. Hui, Y. Liu, J. Qiu, L. Cao, L. Ji, and Z. He, Study of texture segmentation and classification for grading small hepatocellular carcinoma based on CT images, *Tsinghua Science and Technology*, vol. 26, no. 2, pp. 199–207, 2021.

[5] G. Zheng, G. Han, and N. Q. Soomro, An inception module CNN classifiers fusion method on pulmonary nodule diagnosis by signs, *Tsinghua Science and Technology*, vol. 25, no. 3, pp. 368–383, 2020.

[6] O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation, arXiv preprint arXiv: 1505.04597, 2015.

[7] X. D. Xia and B. Kulis, W-net: A deep model for fully unsupervised image segmentation, arXiv preprint arXiv: 1711.08506, 2017.

[8] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, *Inf. Fusion*, vol. 58, pp. 82–115, 2020.

[9] A. Kanezaki, Unsupervised image segmentation by backpropagation, presented at IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP), Calgary, Canada, 2018, pp. 1543–1547.

[10] T. Cour, F. Benezit, and J. B. Shi, Spectral segmentation with multiscale graph decomposition, presented at 2005 IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 1124–1131.

[11] J. B. Shi and J. Malik, Normalized cuts and image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.

[12] Y. Zhang, M. Brady, and S. M. Smith, Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm, *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, 2001.

[13] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, Contour detection and hierarchical image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, 2011.

[14] D. Comaniciu and P. Meer, Mean shift: A robust approach toward feature space analysis, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, 2002.

[15] M. Kass, A. Witkin, and D. Terzopoulos, Snakes: Active contour models, *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1988.

[16] T. F. Chan and L. A. Vese, Active contours without edges, *IEEE Trans. Image Proc.*, vol. 10, no. 2, pp. 266–277, 2001.

[17] X. Fan, M. Dai, C. Liu, F. Wu, X. Yan, Y. Feng, Y. Feng, and B. Su, Effect of image noise on the classification of skin lesions using deep convolutional neural networks, *Tsinghua Science and Technology*, vol. 25, no. 3, pp. 425–434, 2020.

[18] T. Mcinerney and D. Terzopoulos, Deformable models in medical image analysis: A survey, *Med. Image Anal.*, vol. 1, no. 2, pp. 91–108, 1996.

[19] D. L. Pham, C. Y. Xu, and J. L. Prince, A survey of current methods in medical image segmentation, *Annu. Rev. Biomed. Eng.*, vol. 2, pp. 315–337, 2000.

[20] E. Shelhamer, J. Long, and T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, 2017.

[21] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2018.

[22] Y. Y. Boykov and M. P. Jolly, Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images, in *Proc. 8^{th} IEEE Int. Conf. Computer Vision*, Vancouver, Canada, 2001, pp. 105–112.

[23] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut": Interactive foreground extraction using iterated graph cuts, *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, 2004.

[24] C. M. Li, C. Y. Kao, J. C. Gore, and Z. H. Ding, Implicit active contours driven by local binary fitting energy, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (*CVPR*), Minneapolis, MN, USA, 2007, pp. 1–7.

[25] X. F. Wang, D. S. Huang, and H. Xu, An efficient local Chan–Vese model for image segmentation, *Pattern Recognit.*, vol. 43, no. 3, pp. 603–618, 2010.

[26] S. G. Liu and Y. L. Peng, A local region-based Chan–Vese model for image segmentation, *Pattern Recognit.*, vol. 45, no. 7, pp. 2769–2779, 2012.

[27] C. M. Li, C. Y. Xu, C. F. Gui, and M. D. Fox, Distance regularized level set evolution and its application to image segmentation, *IEEE Trans. Image Proc.*, vol. 19, no. 12, pp. 3243–3254, 2010.

[28] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, 2012.

[29] O. Veksler, Y. Boykov, and P. Mehrani, Superpixels and supervoxels in an energy optimization framework, in *Computer Vision–ECCV 2010*, K. Daniilidis, P. Maragos, and N. Paragios, eds. Cham, Germany: Springer, 2010, pp. 211–224.

[30] J. B. Shen, Y. F. Du, W. G. Wang, and X. L. Li, Lazy random walks for superpixel segmentation, *IEEE Trans. Image Proc.*, vol. 23, no. 4, pp. 1451–1462, 2014.

[31] M. S. Aslan, A. Shalaby, and A. A. Farag, Clinically desired segmentation method for vertebral bodies, presented at 2013 IEEE 10th Int. Symp. Biomedical Imaging, San Francisco, CA, USA, 2013, pp. 840–843.

Quanwei Huang received the BS degree from Tsinghua University, Beijing, China in 2019. He is now pursuing the master degree at the Department of Computer Science and Technology, Tsinghua University, Beijing, China. His main research interest includes computer vision and deep learning.

Yuezhi Zhou received the PhD degree in computer science and technology from Tsinghua University, China in 2004 and is now working as a research professor at the same University. He worked as a visiting scientist at the School of Computer Science, Carnegie Mellon University in 2005. His research interests include ubiquitous/pervasive computing, cloud/edge computing, and distributed and mobile systems. He has published over 100 technical papers in international journals or conferences. He received the Best Paper Award at IEEE AINA 2007, IEEE HPCC 2014, and ICA3PP 2018. He is a senior member of the IEEE and a member of the ACM.

Weikang Yu received the BS degree from Beihang University, China in 2020. He is now pursuing the master degree at the School of Science and Engineering, Chinese University of Hong Kong, Shenzhen, China. His main research interest includes computer vision and deep learning.

Linmi Tao received the BS, MS, and PhD degrees in 1986, 1991, and 2001, respectively. He has worked at International Institute of Advance Science Studies, University of Verona, Italy, and the Department of Computer Science and Technology, Tsinghua University. Based on the cross-disciplinary background, his research covers a brand spectrum on computer vision, machine learning, and deep learning based medical image processing and understanding, which are supported by Chinese National Key Scientific Research Project, the National Natural Science Foundation of China, European FP7 Large-Scale Integrating Project, and international cooperative projects. He has published research papers over international peer reviewed journals and conferences, and he is delighted to receive "2020 Excellent Paper Award" from *Tsinghua Science and Technology*.

Yaoxue Zhang received the BS degree from Northwest Institute of Telecommunication Engineering, China in 1982, and the PhD degree in computer networking from Tohoku University, Japan in 1989. Currently, he is a full professor at the Department of Computer Science and Technology, Tsinghua University, China. His research interests include computer networks, operating systems, ubiquitous/pervasive computing, and big data. He has published over 200 technical papers in international journals and conferences, as well as 9 monographs and textbooks. He is an academician of the Chinese Academy of Engineering and is currently serving as the editor-in-chief for the *Chinese Journal of Electronics*.

Li Huo received the PhD degree in nuclear medicine from Peking Union Medical College Hospital, China in 2007. She is the director of Department of Nuclear Medicine, Peking Union Medical College Hospital. She is a chief physician specializing on nuclear medicine. As one of the best senior physicians in nuclear medicine in China for over 10 years, she is in charge of routine clinical work and many clinical trial and research works granted by National Natural Science Foundation of China and Capital government. She has published over 30 SCI papers and 10 professional books. Now she is the dean of National Center for Quality Control of Nuclear Medicine and vice chairman of Beijing Branch of the Chinese Medical Association. She is qualified in translational medicine, clinical trials, and image analysis with novel nuclear medicine techniques for clinical application.

Zuoxiang He received the MD degree from Peking Union Medical College (PUMC), China in 1989. He served as a visiting research fellow at University of Nice Medical School in France, and a postdoctoral fellow at Baylor College of Medicine in the USA. He was the professor of nuclear medicine at PUMC, and served as chairman of nuclear medicine at Fu Wai Hospital. He is currently the professor of Tsinghua University and the vice-preisident of Institute for Precision Medicine, and chairman of nuclear medicine at Beijing Tsinghua Changgung Hospital. He has extensive clinical and basic research experience in the field of nuclear medicine. He has published over 200 articles in major peer-review international and national scientific journals including *Circulation* and *Journal of Nuclear Medicine*.