Research Article

Adaptive Localizing Region-Based Level Set for Segmentation of Maxillary Sinus Based on Convolutional Neural Networks

Xianglong Qi [ID],[1] **Jie Zhong,**[2] **and Shengjia Cui**[3]

[1]*Liaoning Huading Technology Co., Ltd., Shenyang, Liaoning 110167, China*
[2]*JiangSu PangPu Network Technology Co., Ltd., JiangSu, China*
[3]*Baidu.com Times Technology (Beijing) Co., Ltd., Beijing, China*

Correspondence should be addressed to Xianglong Qi; 2010513@stu.neu.edu.cn

In this paper, we propose a novel method, an adaptive localizing region-based level set using convolutional neural network, for improving performance of maxillary sinus segmentation. The healthy sinus without lesion inside is easy for conventional algorithms. However, in practice, most of the cases are filled with lesions of great heterogeneity which lead to lower accuracy. Therefore, we provide a strategy to avoid active contour from being trapped into a nontarget area. First, features of lesion and maxillary sinus are studied using a convolutional neural network (CNN) with two convolutional and three fully connected layers in architecture. In addition, outputs of CNN are devised to evaluate possibilities of zero level set location close to lesion or not. Finally, the method estimates stable points on the contour by an interactive process. If it locates in the lesion, the point needs to be paid a certain speed compensation based on the value of possibility via CNN, assisting itself to escape from the local minima. If not, the point preserves current status till convergence. Capabilities of our method have been demonstrated on a dataset of 200 CT images with possible lesions. To illustrate the strength of our method, we evaluated it against state-of-the-art methods, FLS and CRF-FCN. For all cases, our method, as assessed by Dice similarity coefficients, performed significantly better compared with currently available methods and obtained a significant Dice improvement, 0.25 than FLS and 0.12 than CRF-FCN, respectively, on an average.

1. Introduction

Nasal diseases are growing common for individuals with serious impacts on daily life. In America, there are about 16% adults suffering from this trouble, and there are 100 million in China [1]. Acute sinusitis is simple for management with antibiotic drugs [2]. However, for chronic one, functional endonasal sinus surgery (FESS) may be the only solution of relief. FESS contains some processes: maxillary sinus fenestration, ethmoid sinus resection, etc. [3]. It is found that a lot of risks may appear in surgery due to great variations of nasal anatomy, where optical nerves and carotids have higher possibility to be affected [4]. Segmentation and the subsequent quantitative assessment of lesions in medical images provide valuable information for the analysis of diseases, and quantitative imaging can reveal clues about lesion characteristic and anatomical structure. For example, preoperative computed tomography (CT) affords radiologists the opportunity to prospectively identify anatomic variant that predispose patients to major surgical complications [5]. A real-time surgical navigation has joined in FESS [6]. To make FESS safer, surgeons use navigation systems that register a patient to his/her CT scan and track the position of tools inside the patient [7].

There are growing evidences that the qualification of sinuses increases insights into functional outcomes and requires accurate segmentation which is a challenging task for a number of reasons. The heterogeneous appearance of lesions including large variability in location, size, shape, and frequency makes it difficult to design effective segmentation rules. Although the most accurate segmentation results can be obtained through manual delineation by an experienced

expert, it costs plenty of tedious time, nearly, 8–10 hours per case of one patient [8]. Besides, the expertise decides whether a particular region is a part of the segmentation object. In order to understand the complexity of maxillary sinus, it is necessary to conduct wide studies to gain statistical pattern for drawing conclusions. Therefore, a more accurate, automatic segmentation algorithm has been required as soon as possible.

Figure 1 illustrates some of the potential challenges when proposing a computational approach for the task of automatic maxillary sinus segmentation. The figure demonstrates intensity statistics of great difference and shows examples of lesions in the cavity of maxillary sinus. Lesions can be found at multiple sites, with different shapes and sizes which misguide curve evolution as a result of high gradient noisy areas. Meanwhile, it is generally difficult to derive statistical prior knowledge to set up a similar description for shape of maxillary sinus due to common anatomical abnormality [9]. Ideally, an acceptable solution is able to adjust itself to foreground by learning from potential features in a few of examples.

Level set is popular as a curve evolution application, especially for medical image segmentation [10]. Compared with active contour methods based on points' motion, it has the ability to handle image noise, intensity heterogeneity, and discontinuous object boundaries. Level set methods include models based on edge [10–12] or region [13–15]. Edge-based models are sensitive to noise and objects with incomplete boundaries or low contrast texture. Region-based ones study spatial statistics of interesting regions to find global minima of energy functional but lack competence in details. While, in many examples of great heterogeneous foreground and background, a local region model has a better performance than the global one [16]. Hybrid models are superior, defining an energy functional with local and global constraints to obtain a more robust segmentation with little sensitive initialization.

In level set works, the choice of functional control parameters is puzzled for a long time. Li et al. [21] give proof that an inappropriate definition may lead to an inferior segmentation regardless of initialization. As Lankton et al. described [16], parameters have close relationships with direction and speed of active contour according to specific texture feature. A slight change in the group of parameters even produces entirely different results. Therefore, the fixed parameters cannot survive under different spatial distributions of intensity in maxillary sinus with lesion. Some researches tend to test an optional value over a series of training set for the entire database of images, but usually, new images may require additional experiments to find the best-fitted parameters. As a result, choosing a fixed set of parameters by trial is a time-consuming and laborious process. In addition, most users do not have enough experience to tune a large number of parameters optimally. For this reason, an adaptive strategy is absolutely desirable. If, in the process of segmentation, parameters could be adjusted dynamically based on specific texture, the approach has a higher chance of providing a more accurate segmentation.

Some papers present an algorithm to estimate parameters for energy functional before segmentation. Li et al. [21]

introduce a method to adjust parameters of the level set model according to classification of pixels with intensity. Unfortunately, the value is initialized once at the beginning of work and remains constant all the time, which causes a poor segmentation especially for the object with great variant lesions. Oliviera et al. [22] provide a mechanism to evaluate parameters from analysis of the training set which also can be reused in new images. It is obviously inappropriate for sinus cavity with highly diverse lesions. Baillard et al. [23] devise a solution for the problem of spatial heterogeneity and take full considerations for every point position in every iteration. That is, a point on the contour at first determines its status. If it belongs to the object, it should locally extend outwards. And, if not, it should move in the opposite direction. This classification depends on Gaussian and the shifted Rayleigh statistical distribution models via training set by maximizing the posterior probability. However, because of the diversity of lesion, its prediction accuracy of position would be doubted and discussed depending on the conventional machine learning method. Consequently, [21–23] are likely to perform unacceptably for highly diverse datasets and need more optimizations.

At the same time, deep learning techniques have appeared as a powerful alternative for supervised learning applications such as classification, detection, and other areas [24–27]. Convolutional neural network (CNN) [28, 29] can learn highly discriminative features and have been widespread with predominance on a variety of problems including medical imaging. Ciresan et al. [30] adopt classical CNN for the segmentation of the neural membrane. Obviously, this strategy has two drawbacks. At first, it costs so much computing time since the network must be run separately for each patch, and there is a lot of redundancy due to overlapping patches. In addition, there is a trade-off between localization accuracy and the use of context. For more accurate localization, Ronneberger et al. give a more elegant architecture U-Net derived from fully convolutional network (FCN) [31], which consists of a contracting path to capture context and a symmetric expanding path. Although the algorithm gains a certain achievement, the result of the object specifically in edge is too coarse to be accepted for medical image segmentation. Kamnitsas et al. [32] incorporate conditional random field (CRF) into FCN. Details of segmentation are refined by fuzzy classification based on relations between position and class of pixels in future. It is a pity that this method suffers from noises seriously. Assaf et al. [33] propose to predict the position of active contour with CNN. As this method focuses on entire contour status on average, it cannot deal with the irregular shape of object particularly. These methods are proved successful in computer vision applications on medical images. The idea of CNN strengthens knowledge of significant features on large scale of the training set, while a big challenge for appearance or texture with great distribution in space has come up with maxillary sinus segmentation. It is impossible to set up a class-balance training set to describe all possible cases with lesion.

In this paper, we present a novel improvement of the level set segmentation using the convolutional neural
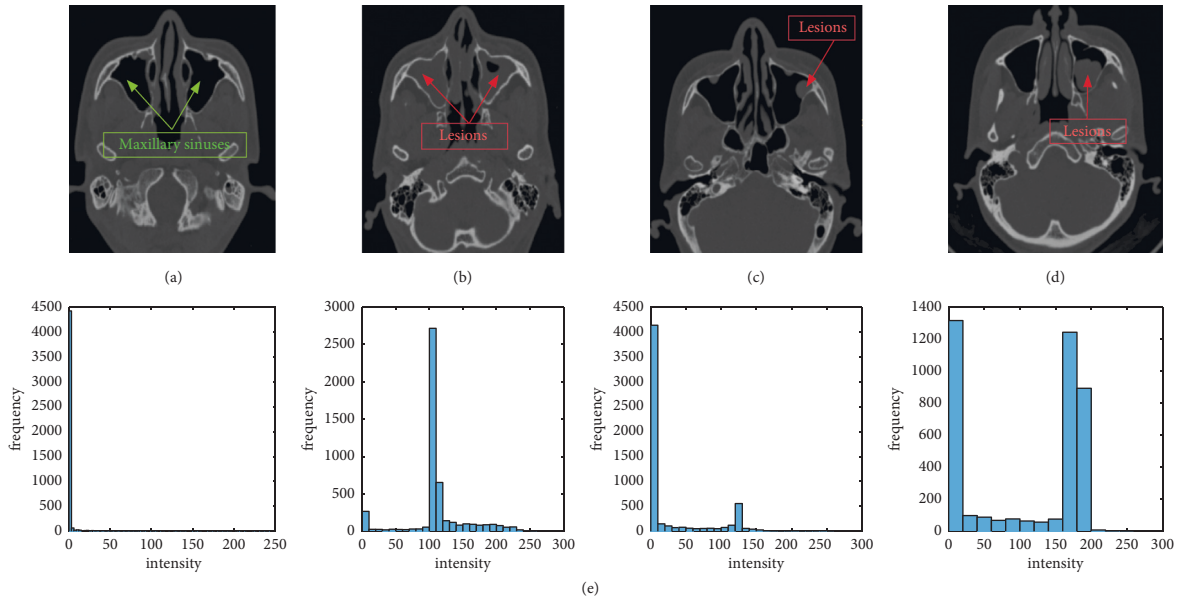
FIGURE 1: Inhomogeneous appearance of maxillary sinus shows challenges in research of segmentation. (a) A pair of healthy cavity and (b–d) lesions, sizes, locations, and shapes which are diffused and reveal the wide spatial distribution. Failed segmentation is more likely caused by incorrect noisy information. (e) The average of the normalized intensity histograms of corresponding CT images in maxillary sinus. Different distributions can be observed easily and demonstrate difficulty for most popular methods.

network. Our method is a multistage process. A convolutional neural network is used to identify the location of a point on the zero level set contour. If the point has high probability close to lesion area, its speed can be compensated to a certain extent based on output probabilities of CNN. If not, we believe it has arrived in the target area. This interactive procedure happens when minimizing the cost functional over iterations of the level set. It is unnecessary for a more accurate initial contour at the beginning of segmentation. Contrary to current hybrid level set frameworks, our method has little dependence on special initialization and does not include any assumptions about sinuses and lesion characteristics. Therefore, our proposed algorithm has the strong adaptiveness in diverse datasets of maxillary sinus with heterogeneous lesion that include plenty of noise and abnormal anatomical structure.

To the best of our knowledge, this is the first use of CNN to identify location of points on the zero level set with some magnitude of speed compensation, avoiding energy functional from being trapped into local minima and resulting in a generalized segmentation solution than methods available to date.

2. Materials and Methods

2.1. Energy Models. We used two different energy models to extensively evaluate our proposed adaptive localizing region-based segmentation algorithm: uniform modeling energy (UM) and mean separation energy (MS). In this paper, the innovative method is optimized from local version of the level set [16], and energy is given as follows:

$$E(\phi) = \int_{\Omega_x} \delta\phi(x) \int_{\Omega_y} B(x, y) \cdot F(I(y), \phi(y)) \mathrm{d}y \mathrm{d}x + \mu \int_{\Omega_y} \delta\phi(y) |\nabla\phi(y)| \mathrm{d}y, \tag{1}$$

where x and y denote the position of specific pixel located in the coordinate of current image, $B(x, y)$ represents local regions centering at every point on the contour, such as a rectangular or circle, and the function F is a generic internal energy measure used to describe local adherence to a given model at each point along the contour. In order to keep the

curve smooth, we add a regularization term as is commonly done. Therefore, the UM or MS model is selected as possible candidates for F to be implemented in our proposed method. By taking the first variation of (1) with respect to ϕ, we obtain the following evolution equation:

$$\frac{\partial \phi}{\partial t}(x) = \delta\phi(x) \int_{\Omega_y} B(x, y) \cdot \nabla_{\phi(y)} F(I(y), \phi(y)) \mathrm{d}y + \mu \delta\phi(x) \mathrm{div}\left(\frac{\nabla\phi(x)}{|\nabla\phi(x)|}\right). \tag{2}$$

2.1.1. Uniform Modeling Energy. A well-known example of an energy that uses a constant intensity model is the Chan-Vese energy [13], which we will refer to as the uniform modeling energy described as

$$F_{UM} = \int_{\Omega_y} \lambda_1 H_{\phi(y)} (I(y) - u)^2 + \lambda_2 \left(1 - H_{\phi(y)}\right)(I(y) - v)^2 \mathrm{d}y. \tag{3}$$

The energy model divides region of interest (ROI) into foreground and background. u and v represent their intensity means, respectively. Set Ω as a bounded subset in R^2 and $I(y)$ as the coordinated of a point on image I. Let $\phi(y)$ be a signed distance map and $H_{\phi(y)}$ be the foreground region. A local version of the UM model can be used by replacing u and v with their local versions, u_x and v_x, to represent the local means of a region divided into exterior and interior surrounding each contour point, according to equation (4). λ_1 and λ_2 are parameters that have impactions on speed direction and magnitude which have close relationship with the texture of ROI. The related derivation process could be provided in [13]:

$$F_{UM} = \mu \int_{\Omega_y} \delta\phi(y)|\nabla\phi(y)|\mathrm{d}y + \lambda_1 \int_{\Omega_y} |I(y) - u_x|^2 H_\phi(y)\mathrm{d}y + \lambda_2 \int_{\Omega_y} |I(y) - v_x|^2 \left(1 - H_\phi(y)\right)\mathrm{d}y. \tag{4}$$

2.1.2. Mean Separation Model. The mean separation model is first proposed by Yezzi et al. [34] which we refer to as mean separation energy:

$$F_{MS} = \int_{\Omega_y} (u - v)^2. \tag{5}$$

This energy functional arrives in minima when foreground and background regions have maximally separate mean intensities. There is a strong assumption that the object and its background have the largest difference of intensity. There is no restriction on how well regions are modeled by u and v. In our method, we design the local version of F as

$$F_{MS} = (u_x - v_x)^2. \tag{6}$$

By substituting the derivative of F_{MS} into (2), we obtain the local region based on speed flow. This allows the MS model to find image edges effectively without considering the uniformity of global internal or external regions.

2.2. The Proposed Method. The proposed method contains an interactive hybrid of the localizing level set and CNN, illustrated in Figure 2, which estimates the probabilities of points' location and helps the energy functional escape from local minima effectively. From borrowed spirits from previous work [35], we initialize the level set model with a contour surrounding the segmentation object. In general, the contour should move to the target as the required speed as level set functional. However, as the result of noisy disturb, the moving contour may be trapped in the local noises. Therefore, we predict the real target possible location and suggest new speed direction and value to accelerate the moving contour moving state evolving more reasonably.

2.2.1. Speed Compensation. As introduced in Section 1, difficulties in the segmentation of maxillary sinus are lesions with uncertain positions, shapes, and intensities. Certainly, a healthy one can be dealt with readily. Holding higher gray value, lesions are easy to frustrate evolution of the initial contour that hardly arrives in the target. In our optimization, we propose a solution of speed compensation to resolve this problem. That is, if speed of a point on the contour is detected close to zero, a devised convolutional neural network is required to identify features of its local region. For more accurate result, we train it on a large scale of the training set. About details, we will discuss it later. The CNN outputs each of two classes: boundary of sinus (p_1) or lesion (p_2). In the interactive step, we use the probability value of two classes to give speed compensation, which is calculated using the following equation based on the MS model:

$$\frac{\partial\phi}{\partial t}(x) = \delta\phi(x)\int_{\Omega_y} B(x,y)\delta\phi(y) \cdot \left(\lambda 1 \frac{(I(y)-v_x)^2}{A_v} - \lambda 2 \frac{(I(y)-u_x)^2}{A_u} + \exp\left(|\lambda_1 + \lambda_2|^{1/2}\right)\left(\frac{1+p_2}{1+p_1} - \frac{1}{2}\right)\right)\mathrm{d}y + \mu\delta\phi(x)\mathrm{div}\left(\frac{\nabla\phi(x)}{|\nabla\phi(x)|}\right), \tag{7}$$

where λ_1 impacts the weight of moving the curve outward along its normal and, alternatively, λ_2 tries to move the curve inward. Both relative items interact on each other and decide ultimate magnitude of speed. $\exp(|\lambda_1 + \lambda_2|^{(1/2)})(((1 + p_2)/(1 + p_1)) - (1/2))$ is the term of speed compensation. Our research relies on an important assumption that any initial
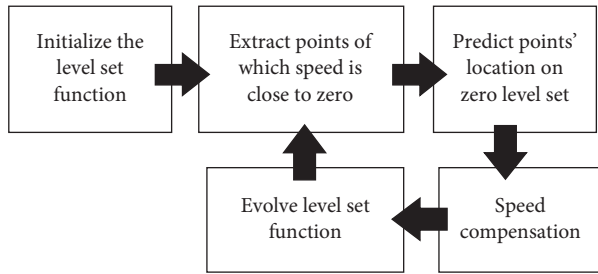
FIGURE 2: Pipeline of the proposed method with speed compensation.

contour is given inside sinus cavity, since there are plenty of other sinus components outside bringing with impossibility to automatic segmentation. For a stable point, if $p_2 \gg p_1$, it can be paid for a great of velocity and tend to move far away. If it locates in the region of the maxillary sinus boundary, $p_2 \ll p_1$, the compensation is nearly zero, maintaining current status till convergence. We drop the confused condition, $|p_1 - p_2| \leq 0.2$, without considerations in order to make sure of stability, which is found in very few cases in validation experiment.

2.2.2. CNN Architecture. For the research of medical image segmentation, the most unsatisfied results are caused by heterogeneous texture feature that leaves the energy functional in local noisy minima. Our method generalizes this problem into a machine learning process with a common architecture for the CNN.

The proposed architecture consists of two convolutional layers followed by three fully connected layers with outputs of two classes (Figure 3). The input of the image is sampled by a point's local region, and its size is fixed on 32×32. There are two convolutional layers at the beginning of our proposed CNN architecture, which includes a 5×5 filter and a 2×2 max pooling filter, respectively. The difference between two layers is the depth and stride. All hyperparameters we prefer have been tested to perform better than others. A nonliner activation function is applied to the outputs of the convolutional layer. In this paper, we choose Leaky ReLU [36] for network to give a sparse quality decreasing computing complexity and keeping backward propagation running smooth. To avoid overfitting and increase accuracy of outputs, we design a batch normalization (BN) filter [37] before ReLU configuration, which also resolves the problem of different distributions in training set or prediction. Each convolutional block of our CNN is illustrated in Figure 4. We give this light model as considerations of computation speed and the weight of our proposed program. A more complicated CNN should be offered, but it will affect the efficiency of the level set model that accounts for more computation costs.

2.2.3. Training the CNN. In the project, the cost function of CNN is selected as cross-entropy loss, which evaluates the performance of the network after each batch:

$$L_{\text{loss}}\left(t, f\left(\omega, x\right)\right) = -\frac{1}{N} \sum_{n=1}^{N} \sum_{m=1}^{M} T_{m,n} \log\left(p_{m,n}\right) + \frac{\eta}{2} \|\omega\|^2, \quad (8)$$

where t represents the true label for every training example and $f(\omega, x)$ is the prediction function. When the n_{th} example is classified into m_{th}, $T_{m,n}$ equals 1. Otherwise, $T_{m,n}$ equals 0. $p_{m,n}$ is the probability of the n_{th} example being classified into the m_{th} class. In order to prevent overfitting, L2-regularization is introduced to penalize the size of the weights in $f(\omega, x)$, where $\eta = 1.0$ is the coefficient of regularization. In the model parameters' learning, we adopt the strategy of moving average, including BN filter and stochastic gradient descent (SGD), and decay is set as 0.9997. With the same decay factor, for searching global minima effectively, every 10000 training steps, we adjust the learning rate exponentially based on the number of steps. And, in fully connected layers, we set dropout and keep probability 0.6 to make sure a sparse network for an acceptable result. To the initialization of model parameters, truncated Gaussian weight distribution is used for generation and standard deviation is 0.1. The CNN is trained with minibatch stochastic gradient descent with a batch size of 32 images. Toolkit of experiments depends on SLIM, a high level encapsulation of Tensorflow.

2.3. Implementation Details

2.3.1. Without Reinitialization of Level Set and Narrow Band. Conventional level set has a boring drawback that the necessary signed distance function (SDF) cannot maintain special characteristic after some iterations. Common solution relies on reinitialization recurrently which costs plenty of computation time and produces numerical errors frequently. Li et al. come up with a regularization idea to resolve this problem [38] effectively. By appending a SDF regularization term on objective energy functional, the evolutionary process of zero level set contour is able to keep quality of SDF without reinitialization so that the complexity of algorithm has been reduced significantly and the final result is prominent among the peer. To save more computation time, the proposed method calculates the energy functional only for grid points located within a narrow band of the distance map, since localizing region-based level set [16] focuses on the local interior and exterior of points on the zero level set contour. Besides, through experiments, we found that the window size of these points outperforms nearly less than 10, a really 'narrow' band ROI.

2.3.2. Training Set and Image Preprocessing. The training set of two classes is divided into lesion or not. As we use CNN to analyze the static point's local region feature, it certainly stays at the edge of sinus cavity or lesion. Consequently, in detail, when constructing the training set, we sampled points on both conditions in average. The size of example is 10×10 and amplified to 32 by bilinear interpolation. During experiments, we tried different sizes and this group performed outstandingly. The number of training patches with two
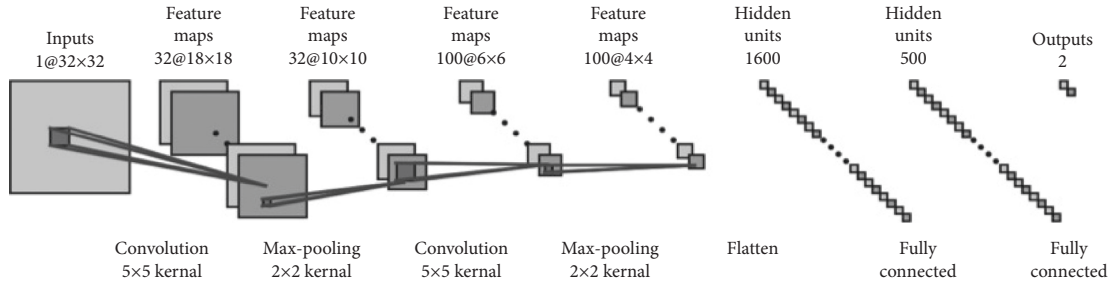
FIGURE 3: CNN architecture. The input is an 32×32 image of a point's local region. Two convolutional components include a 5×5 filter and 2×2 max pooling. Three fully connected layers contain 1600, 500, and 2 nodes, respectively.
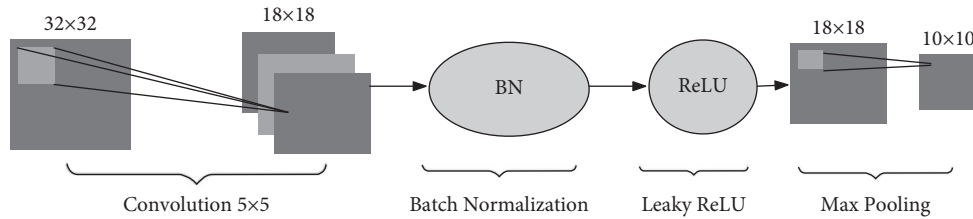

FIGURE 4: An instance of 1st convolutional block with batch normalization, Leaky ReLU, and max pooling.

classes is collected as 50 thousands, respectively. For inputs of neural network, normalization of gray values is necessary for speed and accuracy. Common strategies are often influenced by noises greatly and insensitive to significant feature. We apply contrast-limited adaptive histogram equalization (CLAHE) that enhances two classes' different texture feature expressions and reduces interference of trivial noises. Figure 5 shows a subset of the training set and corresponding preprocessed feature maps.

2.3.3. Data Augmentation. For enlarging the limited training set, the augment dataset was created by applying a combination of elastic and affine distortions. We created elastic distortions by generating stochastic displacement field with values within the range of $[-1, 1]$, convolving these fields with a range of Gaussian filters, and multiplying the resulting matrices by a range of constant factors, controlling the intensity of the deformation. We also tried to adopt rotation to augment training batches. The result showed this approach is limited in improvement of training quality. Therefore, we did not introduce this design at last.

3. Results and Discussion

3.1. Dataset. In our study, approved by an institutional review board for restricted domain in our project, we used 50 CT volume scans (12.13 GB) by SOMATOM definition AS + SIEMENS containing maxillary sinus to evaluate the proposed multitask network BE-FNet. All of them have the same 512×512 in-plane resolution but with a different number of axial slices. The spacing between pixels along ZYX axes of the acquired dataset falls within from $0.5 \times 0.35 \times 0.35$ mm to $0.625 \times 0.39 \times 0.39$ mm. We analyzed 660 CT images of maxillary sinus with lesions as training and validation sets. Another 200 images were

selected as the testing set with all possible cases. The following image acquisition parameters were used: manufactured by SIEMENS, SOMATOM definition AS+; 120 KVP; 500 mm data collection diameter; pixel spacing of 0.390625 mm; thickness of 0.6 mm; 15.2784 CTDvol. All CT images have isotropic pixels.

In the Section 1, we demonstrated that the size, location, and heterogeneity of lesion have high diversity of characteristics. A wide extent was found in the full set of images. Therefore, using a fixed group of parameters in energy functional for all cases of maxillary sinus is not feasible in practice. In other words, a novel solution should be supplied to stop points from falling into local minima. A wide range of lesion's intensity distribution illustrates the importance of and need for speed compensation based on convolutional neural network that evaluates feature of regions and achieves global optimum.

For evaluation, two radiologist supplied annotations of manual segmentation as ground truth with more than five years of experience. The final result of our proposed method was quantitatively compared with the average of the two radiologists' marking. For the research on sensitivity of segmentation to varied initializations, in every image, we gave five radii 3, 5, 7, 9, and 11 pixels. All initial contours are located inside cavity of sinus. Some were close to the center and others near the sinus boundary. This broad range of initializations allowed us to evaluate robustness of our algorithm in any case.

3.2. Segmentation Performance. Energy functional of parameters $\mu_1 = 0.05$, $\lambda_1 = 1.0$, and $\lambda_2 = 1.0$ was used, as we tested them with the best performance for all 660 CT images without speed compensation. Figures 6(c), 6(f), 6(i), 6(l), and 6(o) show some examples of segmentation for different cases. Segmentation performance was assessed using the
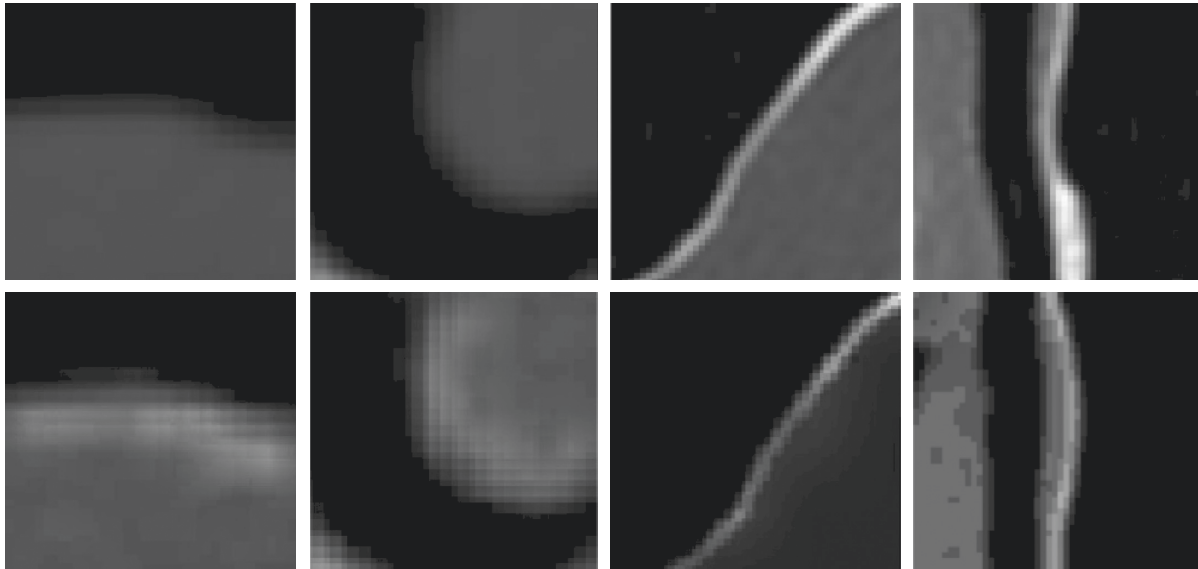
FIGURE 5: Examples of two classes and preprocessed inputs. The two columns on the left show images that were obtained on lesion edge and related preprocessed ones. The two columns on the right are from the bone of sinus cavity. All sampled points locate in the center of images with 32 pixels. By CLAHE analyzing, inputs of different classes stress their obvious features, assisting the CNN network a better performance of learning and predication.

Dice similarity coefficient (Table 1). The Dice coefficient was calculated relative to each radiologist's manual marking, and then, an average Dice score was estimated. Our proposed segmentation method has high agreement with the manual markings for different local energy models. Results were better than the overlap that was measured between the complete manual annotations of both radiologists, thus demonstrating the strength of our proposed method.

3.3. Comparison with Fixed Contour Parameter Method.

We compared our method with a state-of-the-art work energy model of localizing region-based level set based on fixed parameters [16] (FLS). In this experiment, we chose $\mu_1 = 0.05$, $\lambda_1 = 1.0$, and $\lambda_2 = 1.0$ as parameters of energy functional, since they had the best performance in all training set without speed compensation. Figure 6 shows some cases of different maxillary sinus, initial contour, and final segmentation of the object, using both our proposed method and FLS. The left column supplies a series of CT images with lesions that have uncertain sizes, intensities, and locations obviously. Therefore, FLS has higher probabilities to induce the initial contour trapped into edge of lesions. The right column proves our method's advantage.

For quantitative evaluation two method, we acquired another 200 CT images as testing set. And, the Dice coefficients were averaged on five initial contours. The statistics shows that FLS has average Dice coefficients 0.60 ± 0.12 and 0.63 ± 0.10 for the UM and MS model, respectively. These Dice coefficients were significantly lower than the proposed method. Figure 7 clearly shows that our method outperforms the state-of-the-art FLS.

Figure 8 demonstrates the convergence of FLS and our proposed algorithm on iterations of energy functional. At the beginning of running, both of them had a tendency to increase a little and FLS's extent more great. As expected, energy decreased with increasing iterations, converging on a single value; this implies minimization of the energy functional. For both metrics, substantial convergence was obtained after more than 60 iterations. Obviously, our proposed method had a faster speed of convergence. If adjusted to a proper larger length of the step, both of energy functional needed fewer times of iterations.

3.4. Comparison to Other State-of-the-Art Methods.

We compared ours to another state-of-the-art approach proposed by Kamnitsas et al. [32]. The author presented a novel model integrating fully conventional network (FCN) and conditional random field (CRF) for segmentation of the medical image. This paper devised an cost function which joined CRF item to describe more accurate classes of pixels in neighborhood, removing false positive effectively. In addition, they employed a dual pathway architecture that processes the input images at multiple scales simultaneously for accurate localization of the object. Their innovative works improved segmentation skill based on full deep learning and won growing attentions.

We tested Kamnitsas' method (CRF-FCN) of 2D mode on the same training or testing datasets. As CRF-FCN is fully automatic, it did not need any initial contours. On configuration of CRF-FCN hyperparameters, we followed the authors' idea including 50 trees and maximum depth of 30 in random forest baseline. CRF-FCN supplied 90% confidence interval of 0.75 ± 0.12 with a large distribution of Dice coefficients. These results enhance the strength of our proposed method, which is significantly better than both FLS and CRF-FCN.

Figure 9 illustrates results of the CRF-FCN method, performances of which had completely different accuracies
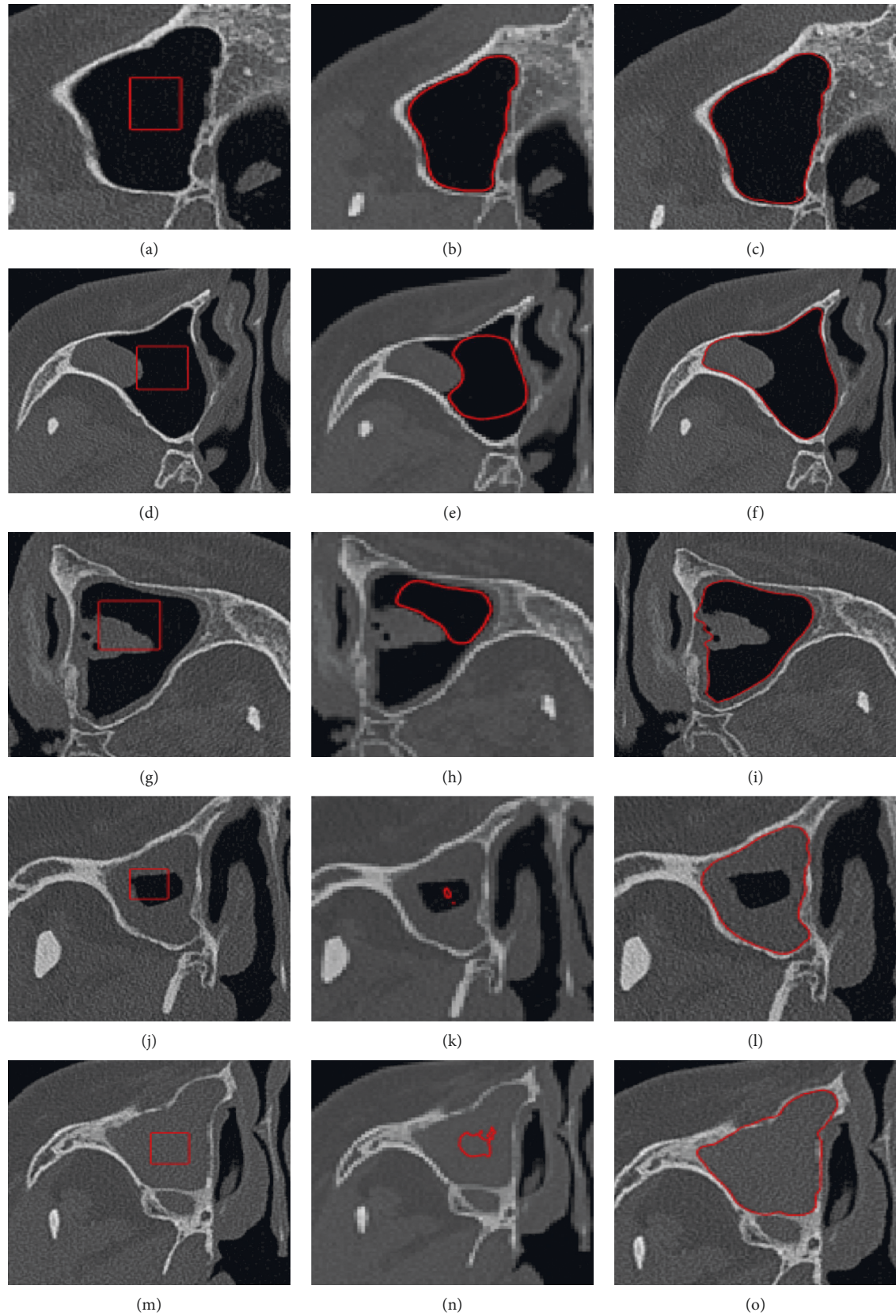
Figure 6: Comparison of FLS with our proposed method. (a, d, g, j, m) Some cases of maxillary sinus with different cases of lesion and the same initial contours. (b, e, h, k, n) Unacceptable results with FLS. (c, f, i, l, o) Predominant outcomes by our proposed method.

TABLE 1: Average dice coefficient and 90% confidence interval (CI) for the maxillary sinus segmentation using our proposed method based on CNN estimation.

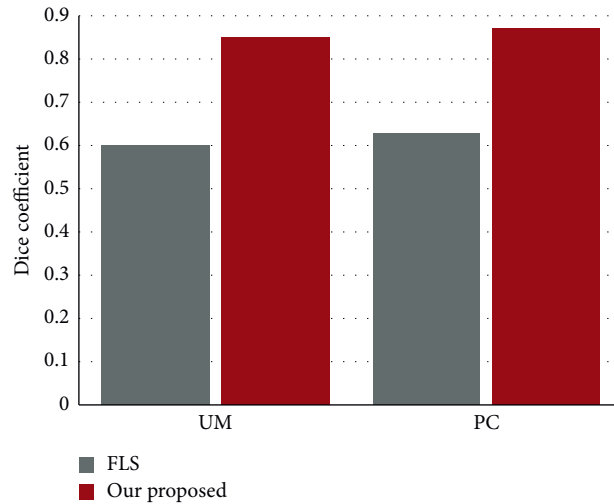|	Our proposed method vs. ground truth
UM model	0.85 ± 0.05
MS model	0.87 ± 0.04

FIGURE 7: Average dice coefficient for two methods and 90% confidence interval for maxillary sinus segmentation over five contour initializations.
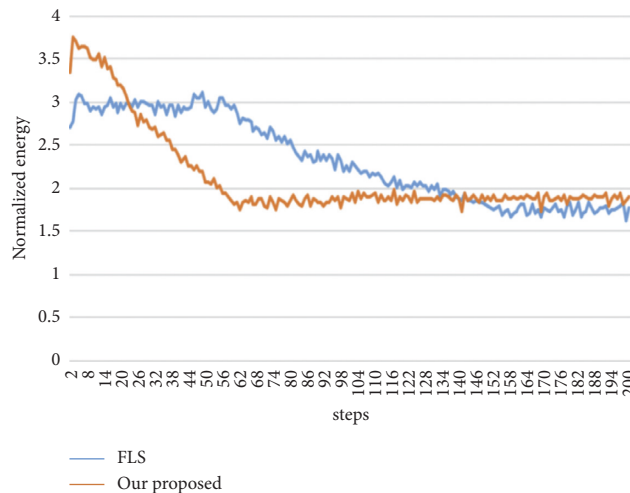

FIGURE 8: Convergence of energy functional on FLS and our proposed method.

in a large extent. To sinus with little lesion, it seemed qualified for segmentation. In the third row, a case filled with plenty of lesions failed to predict segmentation of dense pixels. In the research, we found that the performance of CRF-FCN or similar deep learning skills depends on whether the class distribution of the training set is balanced. If cases of specific pattern are seldom or redundant, they definitely lead the CNN network to stress or ignore corresponding features. Consequently, setting up a class-balanced training set is crucial but a big challenge for practice because

of the large extent of space distribution for maxillary sinuses. Moreover, at the end of CRF-FCN, there is a layer of the CRF model for refinement. It takes more considerations on a pair of pixels with similar position and intensity into the same class. Most of lesions hold approximated features with neighbor regions outside sinus or boundary of cavity, which also has a negative influence on segmentation. For the testing set, completely different distributions of class in maxillary sinus such as size, abnormal structure, and imaging level definitely causes a relative lower accuracy. So, if
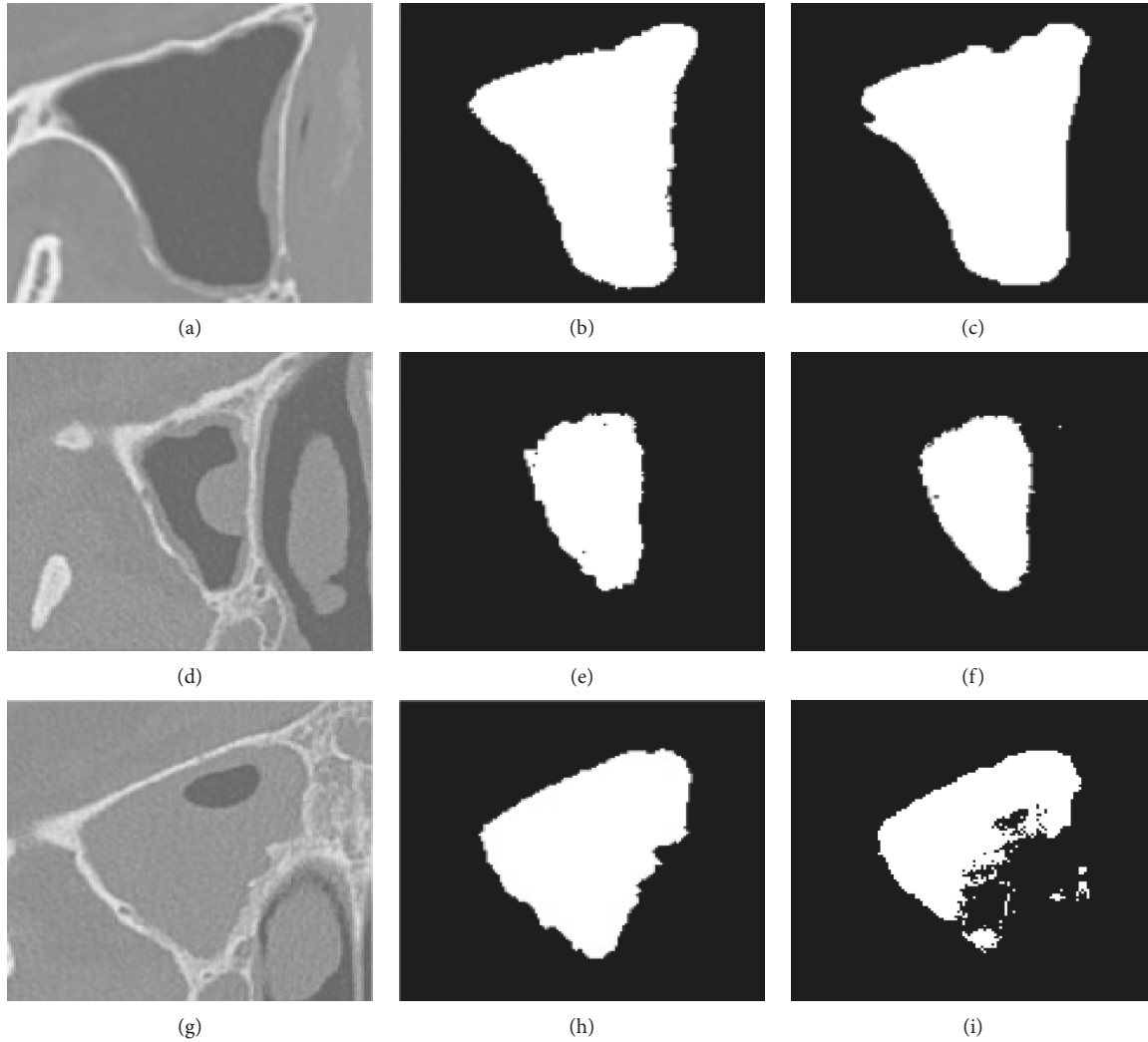
FIGURE 9: Examples of CRF-FCN methods. For explicit expression, masks had been introduced. The left column represents original CT images with lesions of different cases. Images of the middle column are ground truth of manual segmentation. The right column is the result processed by CRF-FCN.

TABLE 2: Average dice coefficient and 90% confidence interval (CI) for the maxillary sinus segmentation, comparison of our proposed method and FLS.

Our proposed method vs. FLS		
	Our proposed	FLS
UM model	0.85 ± 0.05	0.60 ± 0.12
MS model	0.87 ± 0.04	0.63 ± 0.10

there is another sufficient testing database, it should have a chance of improvement in evaluation. In a word, independent identical distribution (IID) of training or testing plays an important role especially in segmentation based on the neural network model.

3.5. Computational Time. We examined the computational time required by the proposed method to analyze segmentation. All data were trained and predicted using Tensorflow 1.2, and the process of level set happened in

MATLAB R2017b 64 bit on Mac OS. Compared with FLS, our method needs an additional procedure of estimation for points till convergence of energy. In addition, the duration of one prediction for stable points on the zero level set by CNN was around 0.2s. In statistics, our proposed method costs 30% longer than FLS on average.

3.6. Sensitivity to Initialization. Different initializations of contours have great impacts on FLS [39, 40]. In the experiment, we evaluated FLS and ours on five different initial contours on which total Dice coefficients were averaged (Table 2). Our proposed method showed better agreement with the manual marking and smaller changes in the segmentation performance when it was applied using different energy models, significantly better than FLS. The result revealed that despite with noisy lesions, our model can deal with substantial deviations of the location of the initial contour, holding great robustness.

4. Conclusions

Segmentation of maxillary sinus with lesion faces a problem that size, location, and heterogeneity are seriously irregular so that the state-of-the-art method cannot survive. Most of them are based on gray gradient which result in the active contour falling in local minima of lesion edge. How to escape and arrive the object is the key of our research.

In this paper, we present a novel method of an adaptive localizing region based of the level set using convolutional neural network. The algorithm automatically analyzes the region feature of a stable point on the active contour in every iteration. If it locates in the lesion, its speed could be compensated based on the probability of outputs by CNN. If not, it stays still till the end of iteration. The proposed mechanism makes sure the trapped point could evolve outward. Therefore, our method is adaptive to cases of sinuses with possible lesions. Our proposed method shows high agreement with expert manual marking for a diverse dataset of CT images. The variety of spatial texture characteristics in our datasets emphasized the strength of our adaptive method, which performed well with inhomogeneous lesions and with noisy backgrounds.

We compared our results to uniform modeling (UM) and mean separation (MS) models of local version [16]. With the testing set of 200 images, Figure 6 demonstrates that fixed parameters of the localizing region-based level set cannot resolve problems, and most contours stop in local minima, especially illustrated in the cavities filled with more lesions. Moreover, our method has obvious predominance over FLS and is confirmed by statistics on quantitative evaluation (Table 2).

Kamnitsas' method (CRF-FCN) [32] is very popular recently and derived from fully convolutional network and conditional random field ideas. It has been proved an efficient and effective algorithm compared with the other state-of-the-art method. With contrast experiment, our proposed method also outperformed significantly and drawbacks of CRF-FCN had been stated. Great variations of appearance in maxillary sinus denote CRF-FCN's frustration furthermore.

Besides, we discussed the convergence of FLS and our method. Figure 8 shows that ours has a faster speed of convergence although it costs 30% more longer computation time on an average. Meanwhile, experiments also gave the confidence of insensitivity to initialization of contours. It implies a great potential of our method in full automatic segmentation. Consequently, our combination of deep learning and level set captures the benefits of both approaches and overcomes their limitations, to achieve significantly better results than either method alone.

The presented work has some limitations. First, whether μ_1, λ_1, and λ_2 of energy functional are sensitive to our method is not discussed in the work. In addition, 3D segmentation based should be a future direction, as well as incorporation of full automatic segmentation, without any user input. In summary, the method presented shows more advanced skills compared with the state-of-the-art level set methods. It performed better in heterogeneous texture between foreground and background, providing a new area of research.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] K.-J. Cheng, Y.-Y. Xu, M.-L. Zhou, S.-H. Zhou, and S.-Q. Wang, "Role of local allergic inflammation and staphylococcus aureus enterotoxins in Chinese patients with chronic rhinosinusitis with nasal polyps," *The Journal of Laryngology & Otology*, vol. 131, no. 8, pp. 707–713, 2017.

[2] A. Ahovuo-Saloranta, U.-M. Rautakorpi, O. V. Borisenko, H. Liira, J. W. Williams Jr., and M. Mäkelä, "Antibiotics for acute maxillary sinusitis in adults," *Cochrane Database of Systematic Reviews*, vol. 2, no. 11, 2014.

[3] D.-Y. Cho and P. H. Hwang, "Results of endoscopic maxillary mega-antrostomy in recalcitrant maxillary sinusitis," *American Journal of Rhinology*, vol. 22, no. 6, pp. 658–662, 2008.

[4] M. Kazkayasi, Y. Karadeniz, and O. K. Arikan, "Anatomic variations of the sphenoid sinus on computed tomography," *Rhinology*, vol. 43, no. 2, pp. 109–114, 2005.

[5] W. T. O. Brien Sr., S. Hamelin, and E. K. Weitzel, "The preoperative sinus ct: avoiding a ?close? call with surgical complications," *Radiology*, vol. 281, no. 1, pp. 10–21, 2016.

[6] W. Koele, H. Stammberger, A. Lackner, and P. Reittner, "Image guided surgery of paranasal sinuses and anterior skull base-five years experience with the instatrak®-system," *Rhinology*, vol. 40, no. 1, pp. 1–9, 2002.

[7] S. Leonard, A. Reiter, A. Sinha, M. Ishii, R. H. Taylor, and G. D. Hager, "Image-based navigation for functional endoscopic sinus surgery using structure from motion," in *Medical Imaging 2016: Image Processing*vol. 9784, Bellingham, WA, USA, International Society for Optics and Photonics, Article ID 97840V, 2016.

[8] S. Pirner, K. Tingelhoff, I. Wagner et al., "Ct-based manual segmentation and evaluation of paranasal sinuses," *European Archives of Oto-Rhino-Laryngology*, vol. 266, no. 4, pp. 507–518, 2009.

[9] W. J. Taylor, R. D. Hayward, P. Lasjaunias et al., "Enigma of raised intracranial pressure in patients with complex craniosynostosis: the role of abnormal intracranial venous drainage," *Journal of Neurosurgery*, vol. 94, no. 3, pp. 377–385, 2001.

[10] S. Osher, J. A. Sethian, and A. James, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, no. 1, pp. 12–49, 1988.

[11] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997.

[12] C. Li, C. Xu, C. Gui, and M. D. Fox, "Level set evolution without re-initialization: a new variational formulation,"vol. 1, pp. 430–436, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, vol. 1, pp. 430–436, IEEE, San Diego, CA, USA, June 2005.

[13] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[14] R. Ronfard, "Region-based strategies for active contour models," *International Journal of Computer Vision*, vol. 13, no. 2, pp. 229–251, 1994.

[15] A. Tsai, A. Yezzi, and A. S. Willsky, "Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation, and magnification," *IEEE Transactions on Image Processing*, vol. 10, no. 8, pp. 1169–1186, 2001.

[16] S. Lankton and A. Tannenbaum, "Localizing region-based active contours," *IEEE Transactions on Image Processing*, vol. 17, no. 11, pp. 2029–2039, 2008.

[17] C. Chunming Li, C.-Y. Chiu-Yen Kao, J. C. Gore, and Z. Zhaohua Ding, "Minimization of region-scalable fitting energy for image segmentation," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1940–1949, 2008.

[18] Q. Zheng, Z. Lu, W. Yang, M. Zhang, Q. Feng, and W. Chen, "A robust medical image segmentation method using kl distance and local neighborhood information," *Computers in Biology and Medicine*, vol. 43, no. 5, pp. 459–470, 2013.

[19] K. Zhang, H. Song, and L. Zhang, "Active contours driven by local image fitting energy," *Pattern Recognition*, vol. 43, no. 4, pp. 1199–1206, 2010.

[20] X. Wang and N. D. Georganas, "Glcm texture based fractal method for evaluating fabric surface roughness," in *Proceedings of the CCECE'09. Canadian Conference on Electrical and Computer Engineering, 2009*, pp. 104–107, IEEE, Edmonton, Canada, May 2009.

[21] B. N. Li, C. Kong Chui, S. H. Ong, and S. Chang, "Integrating fcm and level sets for liver tumor segmentation," in *Proceedings of the 13th International Conference on Biomedical Engineering*, pp. 202–205, Springer, Singapore, December 2009.

[22] D. . A. Borges Oliveira, R. Q. Feitosa, and M. M. Correia, "Liver segmentation using level sets and genetic algorithms," *VISAPP*, vol. 2, pp. 154–159, 2009.

[23] C. Baillard, C. Barillot, and P. Bouthemy, "Robust adaptive segmentation of 3D medical images with level sets," Ph.D. thesis, INRIA, Le Chesnay-Rocquencourt, France, 1999.

[24] H. Farhat, G. E. Sakr, and R. Kilany, "Deep learning applications in pulmonary medical imaging: recent updates and insights on covid-19," *Machine Vision and Applications*, vol. 31, no. 6, pp. 53–42, 2020.

[25] A. M. Ozbayoglu, M. U. Gudelek, and O. B. Sezer, "Deep learning for financial applications: a survey," *Applied Soft Computing*, vol. 93, p. 106384, 2020.

[26] Z. Deng, B. Wang, Y. Xu, T. Xu, C. Liu, and Z. Zhu, "Multi-scale convolutional neural network with time-cognition for multi-step short-term load forecasting," *IEEE Access*, vol. 7, pp. 88058–88071, 2019.

[27] Z. Deng, C. Liu, and Z. Zhu, "Inter-hours rolling scheduling of behind-the-meter storage operating systems using electricity price forecasting based on deep convolutional neural network," *International Journal of Electrical Power & Energy Systems*, vol. 125, Article ID 106499, 2021.

[28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.

[30] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in Neural Information Processing Systems*, pp. 2843–2851, 2012.

[31] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, Boston, MA, USA, June 2015.

[32] K. Kamnitsas, C. Ledig, V. F. J. Newcombe et al., "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

[33] A. Hoogi, A. Subramaniam, R. Veerapaneni, and D. Rubin, "Adaptive estimation of active contour parameters using convolutional neural networks and texture analysis," *IEEE Transactions on Medical Imaging*, vol. 36, no. 3, pp. 781–791, 2016.

[34] Y. Anthony, A. Tsai, and A. Willsky, "A fully global approach to image segmentation via coupled curve evolution equations," *Journal of Visual Communication and Image Representation*, vol. 13, no. 1-2, pp. 195–216, 2002.

[35] Z. Deng, Q. Guo, and Z. Zhu, "Dynamic regulation of level set parameters using 3d convolutional neural network for liver tumor segmentation," *Journal of Healthcare Engineering*, vol. 2019, Article ID 4321645, 17 pages, 2019.

[36] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.

[37] T. Cooijmans, N. Ballas, Laurent, and Aaron, "Recurrent batch normalization," 2016, https://arxiv.org/abs/1603.09025.

[38] C. Chunming Li, C. Chenyang Xu, C. Changfeng Gui, and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3243–3254, 2010.

[39] W. Neuenschwander, F. Pascal, G. Székely, and O. Kübler, "Initializing snakes," in *Proceedings of the Conference on Computer Vision and Pattern Recognition, Number CVLAB-CONF-1994-002*, Seattle, WA, USA, June 1994.

[40] L. D. Cohen and R. Kimmel, "Global minimum for active contour models: a minimal path approach," *International Journal of Computer Vision*, vol. 24, no. 1, pp. 57–78, 1997.