Neural Computing and Applications (2023) 35:14535–14549
https://doi.org/10.1007/s00521-022-07054-2

S.I.: ML4BD SHS

Segmentation of ultrasound image sequences by combing a novel deep siamese network with a deformable contour model

Bo Ni[1,2] · Zhiyuan Liu[2] · Xiantao Cai[1] · Michele Nappi[3] · Shaohua Wan[4]

Received: 2 February 2021 / Accepted: 31 January 2022 / Published online: 14 March 2022

Abstract

Deformable contours are widely applied in medical image segmentation, which are usually derived from appearance cues in medical images. However, the performance of deformed contour is suppressed in ultrasonic image segmentation by the weak, misleading boundaries and the complex shapes of lesion regions. In this paper, a novel deformable contour model is proposed for segmenting ultrasound image sequences, which aims to utilize the powerful ability of deep learning network in learning of image features to help the deformable contour model resist weaknessses of ultrasound images. The deep learning network is designed as a densely connected siamese architecture. It trains a contrastive loss that serves as a boundary searching metric of a deformable contour to segment ultrasound image sequences. In this network, the densely residual blocks and the attention focused blocks are designed to make the network efficiently propagate features and focus on the lesion region, and the feature memory module stores and generates the prior features to aid the evolution of a deformable contour. Moreover, for resisting the impact of misleading or weak boundary, the shape similarity of lesion regions is used to as a shape prior and integrated into the framework of deformable contour to constrain the change of contours. The experimental results for the clinical ultrasound image sequences demonstrate that compared to the state-of-the-art methods, the proposed method can provide more accurate results in HIFU ultrasound images.

Keywords Uterine fibroids · Ultrasound image · Deformable contour · Deep Siamese network · Shape similarity

1 Introduction

High-intensity focused ultrasound (HIFU), which is a noninvasive ultrasound guided therapy, has been extensively employed for treating uterine fibroids that are a common gynaecological condition [1]. As shown in Fig. 1, accurate delineation of the boundary of the lesion region in each frame of ultrasound image sequences is an important step in constructing a preoperative plan of HIFU therapy.

However, ultrasound image segmentation often fail to obtain promising results due to the attenuation, speckle and signal dropout and some weak or misleading boundaries of lesion regions, which presents a challenge for current segmentation methods (refer to Fig. 2). Currently, the segmentation of lesion regions in HIFU ultrasound images is still performed manually or semi-automatically, which is time-consuming and boring. Thus, the development of an effective segmentation method is urgent for the treatment of uterine fibroids in HIFU. Deformable contours are widely applied to extract object boundaries in medical

✉ Shaohua Wan
shwanhust@zuel.edu.cn

Bo Ni
nb@hbpu.edu.cn

Zhiyuan Liu
68568224@qq.com

Xiantao Cai
Caixiantao@whu.edu.cn

Michele Nappi
mnappi@unisa.it

[1] School of Computer Science, Wuhan University, Wuhan 430072, People's Republic of China

[2] School of Computer Science, Hubei Polytechnic University, Huangshi 435003, People's Republic of China

[3] The University of Salerno, Fisciano, Italy

[4] Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen 518110, People's Republic of China

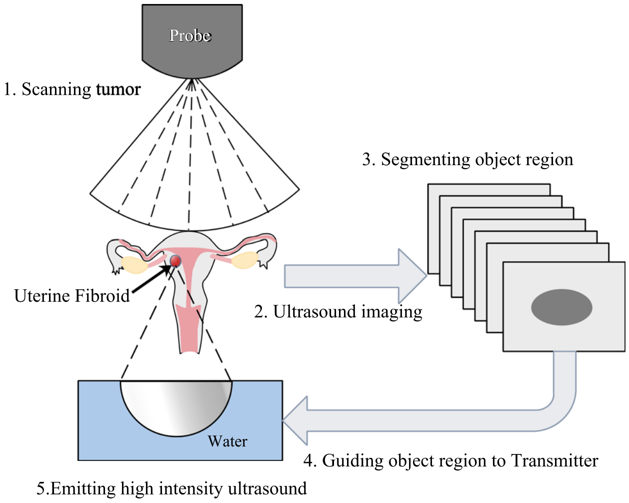

Probe
1. Scanning tumor
3. Segmenting object region
Uterine Fibroid
2. Ultrasound imaging
Water
4. Guiding object region to Transmitter
5.Emitting high intensity ultrasound

Fig. 1 Schematic of the HIFU workflow: (1) a tumour is scanned by an ultrasound scanning device; (2) slices of ultrasound images of the tumour are generated; (3) the object regions in ultrasound images are segmented; and (4) the segmented regions guide the energy transmitter of high-intensity focused ultrasound

images. These mainly due to image features and shape priors, which complement each other to guid deformable contours search the boundaries of target regions. In the past time, convolutional neural networks (CNNs) have shown powerful potential in image segmentation by learning hierarchical features of data. Some excellent networks have been proposed for medical image segmentation, such as fully convolutional network [2, 3], and U-shape networks [4–7].

Recently, the efforts [8–11] of combining the advantages of deformable contours and CNNs are proposed for segmenting images to improve the performance of the deformable contour models by the CNNs. In this paper, our aim is to develop a novel CNN to make a deformable contour extract the object boundary accurately in HIFU ultrasound image sequences. In summary, our method are described as follows:

1. We propose a novel deep siamese network to train a loss function that serves as a contrastive metric of a deformable contour for the searching boundaries of target regions. In this network, the densely residual blocks and the attention focused blocks are designed to enable the network to efficiently propagate feature maps and focus on the lesion region.
2. A feature memory module is also developed in the network to store the features learned from the training stage of the network, and then generates the object and background prior features in the process of segmentation to compute the contrastive loss between the input data and the object and background prior features. The module provides robust object and background features for a deformable contour to resist the artifacts, attenuation, speckle and signal dropout in ultrasound images.
3. To alleviate the impact of the misleading or weak boundary in HIFU images to deformable contours, the rank of matrix is applied to measure the similarity of multi-shapes and is applied to constrain the change of deformable contours. This process can be regarded as a shape prior model based on unsupervised learning.

The experimental results on the real HIFU ultrasound image sequences demonstrate that, compared to the state-of-the methods, our method can extract more accurate object boundaries in the HIFU ultrasound image sequences of different quality.

The remaining sections of this paper are organized as follows: Sect. 2 introduces related work about deformable contours and CNN-based medical image segmentation methods. Section 3 firstly introduces details of the deep siamese network and how to integrate the loss into a framework of a deformable contour model, and then using the matrix rank to measure the variations in multiple shapes is presented in Sect. 3.6. Section 4 describes the algorithm of segmentation of an ultrasound image sequence. Section 5 demonstrates the performance analysis of our method by the experimental results from the clinic ultrasound images. The conclusions and future work are given in Sect. 7.

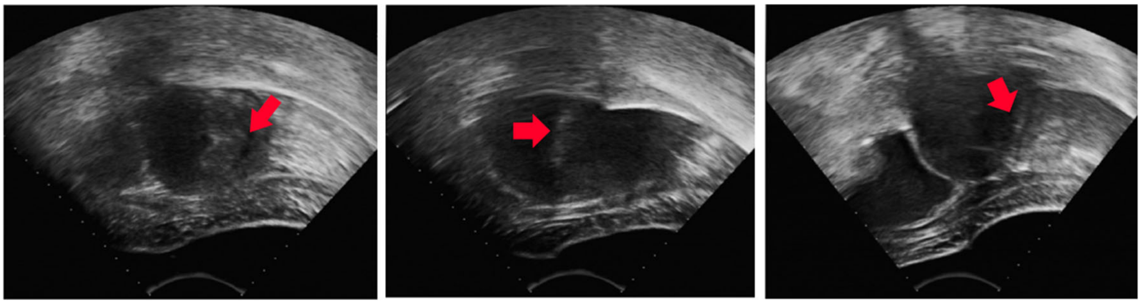


Fig. 2 Characteristics of an ultrasound image. From left to right: inhomogeneous grey distribution, misleading boundary and weak boundary

2 Related works

In this section, we give a brief review of deformable contours and convolutional neural networks for medical image segmentation respectively, and the progress and limitations are also discussed of some current efforts in this section.

2.1 A deformable contour for medical image segmentation

Deformable contour models depend mainly on detect the boundaries of the target region by image features and shape priors. The early image features usually include image gradient vectors [12, 13], saliency boundaries [14–17] and the gradient distribution of local boundaries [18–20]. Some efforts [18, 21–23] of constructing the local features of target boundaries attempt to make the deformable contours resist the disturbance of noise in images. And then, some methods [24, 25] try to consider the correlation between local and global image information to segment medical images. However, only the image cues are not enough to resist the interference of defects from ultrasound imaging to the deformable contour. Therefore, it is very important to apply the target shape prior model to constrain the evolution of deformed contours. The typical shape prior models include point distribution model [26], sparse shape composition [27–29], dynamic models [30, 31] and manifold learning [32, 33], the recovery of Low-rank matrix [34, 35]. However, obtaining a large number of annotated medical images as a training set is a difficult task, and it is often questioned whether the existing shapes in the training set are sufficient to model the shapes of objects in the new images.

Recently, the study [9] utilizes the features about the shape and the area of target region extracted by deformable contours to train a loss function of CNN for segmenting images. The approaches [10, 11] use the CNNs to provide robustive features of the interest of regions to the deformable contours. Geometrical convexity optimizations are used to be shape prior models in deformable contours to segment the interest of region [36, 37].

2.2 Convolutional neural network for medical image segmentation

CNN-based methods have achieved remarkable results in medical image segmentation. The fully convolutional network(FCN) [2] is a major milestone in medical image segmentation, which is trained end-to-end to perform pixels-to-pixels segmentation. U-net[4] and V-net[38] are the extensions based on FCN. Deep residual networks, such as a deeper CNN, are also designed to learn more discriminating features and achieve state-of-the-art segmentation performance [39–41]. However, the deeper networks cause the reduction of weak features in medical images. Some dense convolutional networks were proposed to elevate the vanishing gradient and strengthen the features propagation by adding the connections between the layers of classical CNNs, such as residual dense networks [42–47] and attention mechanism [5, 6, 48].

Another important puzzle in the field of medical image analysis is the lack of training data, which can cause the overfitting of deep learning networks. This challenge can usually be alleviated by fine-tuning a pre-trained network from another task with more labeled data. For instance, the transfer learning based methods [49–53]. Some generative adversarial networks [41, 54–59]? were employed to augment training data by synthesizing new images from real data. But, it is dubious that the quality of the synthesized images satisfies the requirements of clinical practice. The matching networks [60–63] try to train the loss function to learn the relationship between input data and the trained model, have been proposed to solve the problem of overfitting and class imbalance. Recently, Transformers [64, 65] have been proposed as a new self-attentive mechanism for medical image segmentation with satisfactory results, but these methods is extremely memory-consuming.

3 Method

In this section, firstly, we describe an overview of the proposed network and discuss the various modules of our network. Secondly, we introduce how to incorporate the contrastive loss of the network into the framework of a deformable contour as a searching metric of the object boundary. Finally, the low-rank based shape prior and the algorithm for image sequence segmentation are described.

3.1 Overview of the proposed network

We propose a novel deep siamese network with a feature memory module for training a loss function that is servered as a more discriminative searching metric of target boundaries for a deformable contour. The architecture of our network is shown in Fig. 3. The main part of the network is the deep siamese network with a pair of image patches as input data, where each branch has the same structure and parameters, in where the loss function is defined by

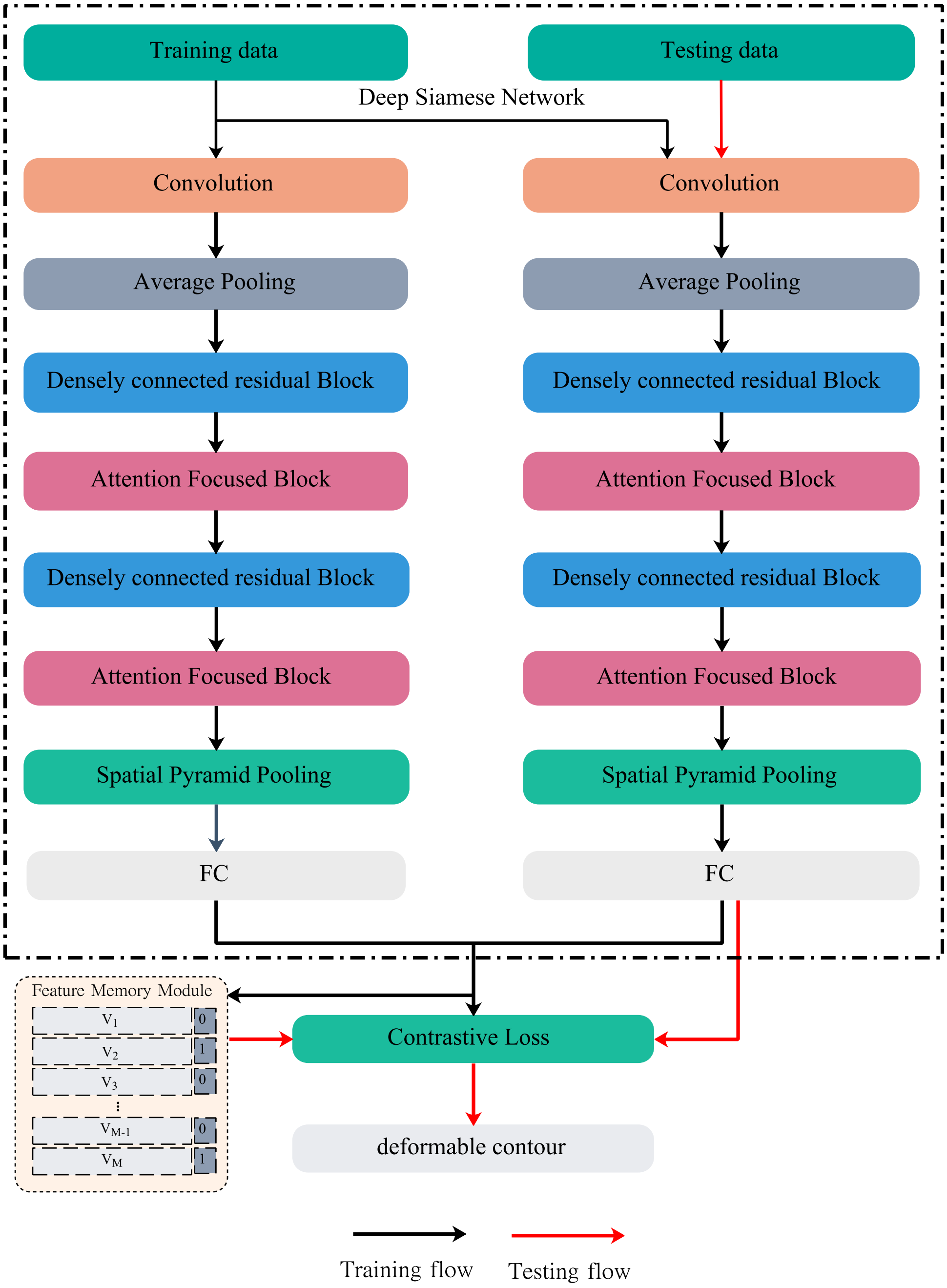


Fig. 3 The framework of the proposed method

$$Loss_t(p, a, n) = d_{a,p} + \max\left(\left(d_{a,p} - d_{a,n} + \alpha\right), 0\right), \tag{1}$$

where α denotes a training parameter, the input variables p, a, n denote a positive sample, an anchor sample and negative sample, respectively. The distance between the input varables can be calculted by

$$d_{a,b} = 1 - \frac{\mathbf{f}_a \cdot \mathbf{f}_b}{||\mathbf{f}_a||_2||\mathbf{f}_b||}, \tag{2}$$

in where a and b denote two input data and $\mathbf{f}_a$ and $\mathbf{f}_b$ are the L_2 normalized features of the two input data. In our work, the training data of the network comprise a number of image pairs and the corresponding label $y \in 0, 1$.

Each branch of the network is designed as a densely connected network with attention focused blocks. Figure 4 A shows the structure of the branch, which consists of one convolutional layer, three densely connected residual blocks (DCRBs), three attention focused blocks (AFBs), three average pooling layers, a spatial pyramid pooling (SPP) layer and a fully connected layer (FC). The average pooling layer uses a stride of two, which gradually reduces the resolution of the feature map and increases the receptive field of the convolutional layers.

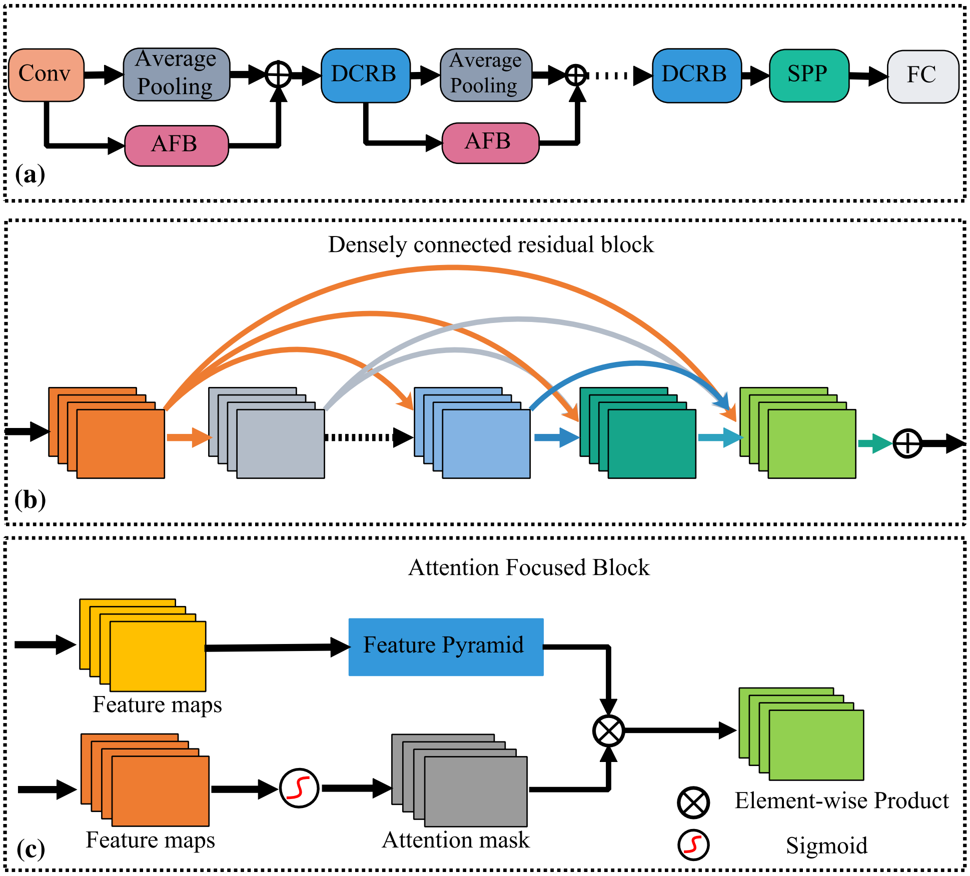


Fig. 4 **A** Structure of the densely connected network with attention focused blocks; **B** Structure of densely connected residual block; **C** Structure of attention-focused block

The DCRBs in our network are aim to alleviate the problem of the vanishing gradient by promoting information propagation within the network, in which dense connections are proposed to connect with all subsequent layers and the feature maps produced by all proceeding layers are concatenated as input for the subsequent layers. The role of the AFBs, which are on the path between the two DCRBs, are to reduce the effects of misleading factors from artificial or noise by introducing an spatial attention mechanism in the network to focus the network on object boundaries. The SPPs applied in our network , which was proposed in [66], serve to be able to ensure the input feature maps have the same size. At the end of the network, the feature of input data are generated by the FC.

Table 1 shows the details of the branch, each branch in the proposed network has more than 40 layers, including convolutional layers, pooling layers, layers in dense blocks, transitional layers and dropout layers. The densely connected residual block includes a different number of BN-ReLU-Conv(1x1) layers and BN-ReLU-Conv(3x3) layers. The transition layer is implemented using a BN-ReLU-Conv(1x1) layer. After each Conv(3x3) layer, a dropout layer with 0.3 dropout rate is added to overcome potential overfitting problem.

Table 1 The details of the branch in the proposed network

Layer & Block	Size of Kernal
Convolution 1	(3 ×3)
Average pooling	(2 ×2), stride=2
DCRB1	(1 ×1), (3 ×3), num =4
TransLayer1	(1 ×1)
Average pooling	(2 ×2), stride=2
DCRB2	(1×1), (3×3), num =6
TransLayer2	(1 ×1)
Average pooling	(2×2), stride=2
DCRB3	(1×1), (3×3), num =8
TransLayer3	(1 ×1)
Convolution 2	(1 ×1)
Spatial Pyramid Pooling	(2 ×2), (4×4)
FC	(1 × 4032)

The feature memory module in our network is a key-value structure, which stores the object-background features in the step of training the netwoke. And then, in the testing stage, the loss function takes in the features of the input data and the a priori features of the object and background generated by the module to compute the contrastive loss.

3.2 Densely connected residual block

It is proven that dense connections enable a deep network to enhance features propagation and alleviate the disappearing gradient [42]. Let x_l denote the output of the l^{th} convolutional layer, which denotes the results after applying the operation H_l, which is defined as a nonlinear transformation followed by batch normalization and a rectified linear unit (ReLU) in the l^{th} layer. For a classical CNN layer with a straightforward connection, x_l can be modelled as

$$x_l = H_l(x_{l-1}) \tag{3}$$

where x_{l-1} is the output of the $l-1$th layer. However, when a network goes deeper, the network may suffer from the vanishing gradient or explode, which produces large training errors and prevents convergence of the network training. Here, we make the lth layer receive all feature maps produced by $[0, 1, \ldots, l-1]$layers as inputs. In addition, to reduce the number of features and fuse the features from the dense layers, a transition layer, which consists of a 1 convolution layer, a batch-normalization and an ReLU, is added at the end of each DCRB. Thus, the output of the lth layer can be defined as

$$x_l = H_t(H_l[x_0, x_1, \cdots, x_{l-1}]) \tag{4}$$

where H_t is a nonlinear transformation of the transition layer. To further promote information propagation and make the network easier to optimize, we also employ a residual connection into our block.

3.3 Attention-focused block

The attention-focused block in our proposed network is employed to make the network focus on the lesion region rather than the noise in the images. The block consists of a spatial pyramid layer, a sigmoid layer and an element-wise multiplication layer. The spatial pyramid layer is applied in the block to ensure that the input feature maps have the same size. The output of the attention-focused block is the element-wise multiplication of input feature maps and attention masks. The attention masks are produced by the sigmoid layer:

$$M_t(x) = f(H_t(x)) \tag{5}$$

$$f(x) = \frac{1}{1+e^{-x}} \tag{6}$$

where $M_{t(x)}$ denotes the attention mask, whose values range from [0, 1], and $H_{t(x)}$ denotes the feature map from a long connection.

3.4 Feature memory module

In our method, the feature memory module is the object and background feature spaces and designed as a key-value structure, which stores the object and background features and their corresponding labels in the slots. The object and background features in the module are given by the FC layer of the network, and all of them are L^2-normalized.

In our work, the corresponding keys of the object and the background features are set to 0 and 1, respectively. Given the training samples including N_t pairs, the module contains N_t slots. When the module are full, it can be updated by (7):

$$v_i = \alpha * v_i + (1-\alpha) * \mathrm{PCA}(M[v(:), k_i]), k \in \{0, 1\}, \tag{7}$$

where $M[v(:), k_i)]$ denotes the vectors composed of the features of the label k_i, PCA(.) denotes the operator of the principal component analysis [67], and the hyper-parameter $\alpha \in [0, 1]$ controls the updating rate.

In the process of the segmenting images, the object and background prior features generated by the module and the feature learned by the proposed network are sent to the loss to discriminate whether the learned features close to the object region or away it. In particular, we separately apply PCA to the object and the background vectors, and the most significant eigenvalues of the covariance matrix of the two features set (accounts of 98% of the total variation).

3.5 Integrating the contrastive loss into the framework of the deformable contour

We consider the results of (1) to estimate whether the landmarks on the deformable contour are similar to the representation feature of the object boundary or deviate from it, i.e. the larger value given by (1) denotes the farther distance from the marked point from the object region, and vice versa. Here, we adopt the classical model in [68] as the framework of the deformable contour model and integrate the contrast loss into its energy function as

$$\begin{aligned} E(\mathcal{C}) = \lambda \int_0^{\mathrm{Length}(\mathcal{C})} \mathrm{Loss}_t(P_{I(x,y)}, P_o, P_b)ds + \\ \int_\Omega (c_1 - I(x,y))dxdy + \int_{\frac{\Omega}{\Omega c}} (c_2 - I(x,y))dxdy, \end{aligned} \tag{8}$$

where ds is the Euclidean distance between two landmarks on the curve $\mathcal{C}$; $Loss_t(P_{I(x,y)}, P_o, P_b)$ presents the contrastive loss of the image patch centred on landmark (x, y) to the object and background prior features P_o, P_b, respectively; the first term of Eq. (8) is the length of the curve, $I(x,\ y)$ is the image to be segmented; and Ω_C is the closed domain of the curve in the image domain Ω. The mean values inside and outside Ω_C are c_1 and c_2, respectively. λ is a fixed weight that controls the smoothness of $\mathcal{C}$.

3.6 Shape similarity measurement

As shown in Fig. 1, an image sequence provided by the ultrasound probe associated with HIFU equipment is composed by a serise of slices. We observe that the shapes of the lesion regions in the each slices is similar. It is desired that this similarity is utilized to constrain the change of the multi-shapes of the lesion regions. Therefore, it is very important to choose an appropriate similarity measure. In [35], the similarity of multiple contours is measured by the rank of matrix, and the low rank of the matrix is proved in detail.

Specific to the work in this paper, the contour of lesion region in the sliece is parameterized to a closed curve $\mathcal{C} = [x_1, \ldots, x_n, y_1, \ldots, y_n]^T \in \mathbb{R}^{2n}$, in where (x_i, y_i) indicates a mark point on the contour. The lesion regions to be segmented in the image sequence can be represented as a matrix $\mathbf{X} = [\mathcal{C}_1, \ldots, \mathcal{C}_m]$, the matrix can be regarded as the variation space of the target shapes in a HIFU image sequence It is assumed that any contour $\mathcal{C}_i$ can be generated from the other contour $\mathcal{C}_\tau$ in the matrix by an affine transformation, i.e,

$$\begin{bmatrix} \mathcal{C}_i^x \\ \mathcal{C}_i^y \end{bmatrix} = \begin{bmatrix} \mathcal{C}_\tau^x & \mathbf{0} & \mathcal{C}_\tau^y & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathcal{C}_\tau^x & \mathbf{0} & \mathcal{C}_\tau^y & \mathbf{0} & \mathbf{1} \end{bmatrix} \Phi_\tau, \tag{9}$$

where $\Phi_\tau = [\mathbf{w}_\tau^{11}, \mathbf{w}_\tau^{12}, \mathbf{w}_\tau^{21}, \mathbf{w}_\tau^{22}, \mathbf{t}_\tau^1, \mathbf{t}_\tau^2]^T$ is an affine transformation matrix. Since the dimention of

$$\begin{bmatrix} \mathcal{C}_\tau^x & \mathbf{0} & \mathcal{C}_\tau^y & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathcal{C}_\tau^x & \mathbf{0} & \mathcal{C}_\tau^y & \mathbf{0} & \mathbf{1} \end{bmatrix}$$

only depends on $\mathcal{C}_\tau$ is at most 6, the rank $(\mathbf{X}) \leq 6$. Thus, the low-rank attribute of the matrix $\mathbf{X}$ is employed to constrain the change of the contours such as translation, scaling, rotation and the local variation caused by image defects in the process of segmenting the image sequence.

4 Algorithm of segmenting an ultrasound image sequence

To apply the deformable contour (8) to segment a sequence of images and keep the contours similar with each other, we propose the objective function by

$$E(\mathbf{X}) = \min_{\mathbf{X}} F(\mathbf{X}) + \beta * \mathrm{Rank}(\mathbf{X}). \tag{10}$$

where $F(\mathbf{X}) = \sum_{i=1}^N E(\mathcal{C}_i)$ denotes the closed curves computed by (8), and $\mathrm{Rank}(\mathbf{X})$ is the operator of the calculating rank of the matrix, which is employed as a shape prior penalty term. In [69], the nuclear norm $\|\mathbf{X}\|_*$ is used as a tight convex surrogate of the rank operator for solving $\mathrm{Rank}(\mathbf{X})$, and the small perturbation in the curves may cause a large increase in $\mathrm{Rank}(\mathbf{X})$. In our work, the Proximal Gradient method in [70] is applied to solve (10), which makes Eq. (10) converge to a stationary point by

$$\mathbf{X}^{i+1} = \arg\min_{\mathbf{X}} \frac{1}{2} \left\| \mathbf{X} - \left[\mathbf{X}^i - \frac{1}{\mu} \nabla F\left(\mathbf{X}^i\right) \right] \right\|_F^2 + \lambda \|\mathbf{X}\|_*. \tag{11}$$

Here,$\nabla F\left(\mathbf{X}^i\right) = \left[\nabla E\left(\mathcal{C}_1^i\right), \ldots, \nabla E\left(\mathcal{C}_N^i\right)\right]$, in [68], $\nabla E(\mathcal{C})$ can be solved by

$$\begin{aligned} &\nabla E(\mathcal{C}) \\ &= \textstyle\sum_{i=1}^n \Big\{ Loss_t(p_i) \left[(c_1 - I(p_i))^2 - (c_2 - I(p_i))^2 \right] \mathbf{N}_{p_i} \\ &\quad + \omega k_{p_i} \mathbf{N}_{p_i} \Big\}, \end{aligned} \tag{12}$$

where p_i denotes a landmark on the contour, and $\mathbf{N}_{p_i}$ and k_{p_i} are the normal vector and the curvature at p_i, respectively, and $\|\mathbf{X}\|_*$ is solved by (13), which refers to a singular value thresholding algorithm. The algorithm is described in [71].

$$\|\mathbf{X}\|_* = \sum_{i=1}^{\min(m,n)} (\sigma_i - \alpha)_+ \mathbf{u}_i \mathbf{v}_i^T, \tag{13}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the left and right singular vectors of $\mathbf{X}$ and $(\cdot)_+ = \max(\cdot, 0)$.

In the process of image sequence segmentation, we initialize the proposed deformable contour models in the image sequence to be segmented, implement the proposed deformable contours to search the object boundaries and implement the low-rank attribute to constrain the evolution of the deformable contours. The details is summarized in Algorithm 1.

Algorithm 1 The solution of (10) with the Proximal Gradient method

```
Initialization: $t_0 = t_{-1} = 1$, $\mathbf{X}^k = \mathbf{X}^{k-1}$
for $i = 0 \to iterations_{max}$ do
    $\mathbf{Y}^i = \mathbf{X}^i + \frac{t^i - 1}{t^i}(\mathbf{X}^i - \mathbf{X}^{i-1})$
    for $j = 1 \to N$ do
        $\mathcal{C}_j^i \leftarrow -\frac{1}{\mu}\nabla E(\mathcal{C}_j^i)$
    end for
    Using $\mathbf{Y}^i$ to update $\mathbf{X}^{i+1}$ by Eq.(11)
    $t^{i+1} = \frac{t^i + t^{i-1}}{t^i}$
    if $\|X^{i+1} - X^i\|_2 < Threshold$ then
        retrun
    end if
end for
```

5 Experiments

In this section, the details about data set and evaluating metrics and the implements of the proposed network are firstly discussed respectively. And then, the performance comparsion between our method and the state-of-the-art methods are elaborated.

5.1 Dataset

The 78 uterine fibroid ultrasound sequences acquired from different patients were applied to our experiments to intuitively evaluate the performance of our method, which were obtained by a Philips ultrasonic scanner, in which each sequence consists of 53–60 slices, each slice has 800×600 pixels, and the pixel size is $0.15\,\mathrm{mm} \times 0.15\,\mathrm{mm}$.

In the experiments, the mean values from the manual segmentation results given by the experienced radiologists were available as the gold standard for comparison. The radiologists removed some images without uterine fibroids from the data set and croped 3559 pairs of patches from the 54 image sequences to be training set and 1450 pairs to be test set. The size of patches is specified in the range from 45×45 pixels to 112×112 pixels. As shown in Fig. 5, we cropped the object and background patches to construct our sample data. Specifically, the red curve presents the boundary of the lesion region. The most obvious characteristic of the object boundary in the ultrasound images is the change in the intensity inside and outside the boundary. The object samples were clipped by centring on the object marked points, and the background samples were clipped on the regions inside and outside the boundary.

The radiologists used the remaining 24 image sequences to validate the performance of the comparison methods and classified the validation set into three groups: high, medium, and low, based on image quality. The high level group contains 7 image sequences, while the medium level and low level groups contain 12 and 5 image sequences, respectively, each containing approximately 20 to 35 slices.

5.1.1 Evaluation

We choose the Husdorff distance (HD) and Dice coefficient (Dice) as the metrics to quantitatively evaluate the comparison between segmentation results and manual segmentation results. HD and Dice are defined as follows:

$$\mathrm{HD}(\mathcal{C}_a, \mathcal{C}_m) = \max\left\{\sup_{\mathcal{P}_a \in \mathcal{C}_a} d(\mathcal{P}_a, \mathcal{C}_m), \sup_{\mathcal{P}_m \in \mathcal{C}_m} d(\mathcal{P}_m, \mathcal{C}_a)\right\} \tag{14}$$

$$\mathrm{Dice}(\mathcal{C}_a, \mathcal{C}_m) = \frac{2|\Omega_{\mathcal{C}_a} \cap \Omega_{\mathcal{C}_m}|}{|\Omega_{\mathcal{C}_a}| + |\Omega_{\mathcal{C}_m}|}. \tag{15}$$

Here, $\mathcal{C}_a$ and $\mathcal{C}_m$ denote the results given by the proposed segmentation method and the results obtained following manual segmentation, respectively. $\mathcal{P}_a$ and $\mathcal{P}_m$ indicate the marked points of $\mathcal{C}_a$ and $\mathcal{C}_m$, respectively, and $d(\mathcal{P}_a, \mathcal{C}_a)$ indicate the marked points of $\mathcal{P}_a$ to contour $\mathcal{C}_a$. $|\Omega_{\mathcal{C}}|$ is the

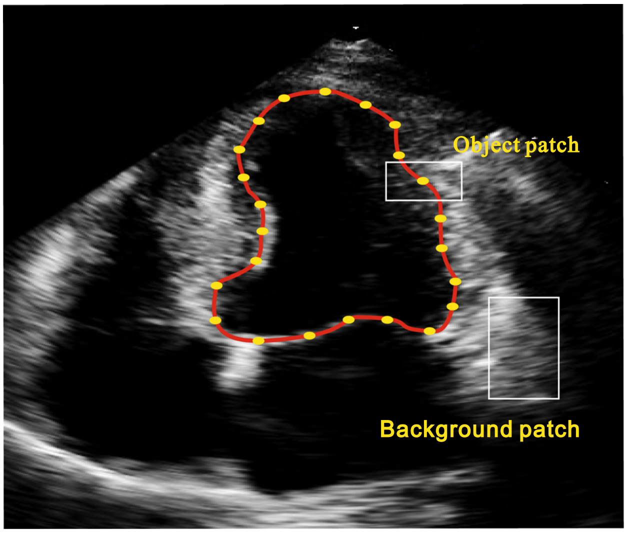


Fig. 5 Illustration of defining object-background patches

total number of pixels inside the region. A smaller HD or a larger Dice coefficient indicates a more accurate segmentation.

5.2 Implementation details of the proposed network

To overcome the overfitting of the model, we performed data augmenting operations with the training samples, such as horizontal and vertical flips, and rotated them to 90°, 180°, 270°. All experiments in this paper were implemented on the deep learning open source library Pytorch. Furthermore, to improve the performance and accelerate the convergence of the proposed network, hard sample mining [72] was exploited to train the network. The mining entails selecting positive samples with a large feature vector distance and negative samples with a smaller (very similar) feature vector distance to form one training batch. In the training stage of our network, the size of the training batch was specified as 24, and the stochastic gradient descent was adopted. The learning rate was initialized to 0.001, the learning rate was decreased by the weight of 10e-6, and the momentum was set to 0.9. The experiments are carried out on GeForce GTX1080Ti GPU with 11GB memory.

Throughout the implementations, the contours are firstly initialized as a series of rough ellipses, and the initialized contours are manually placed near the target regions, and the number of sampled points for each contour was n=12. The coefficient for controlling the updating rate in (7), α was =0.3. β in (10) and λ in (11) usually took the empirical values between 0 and 1 for the best performance.

5.3 Comparison with state-of-the-art methods

Attention Unet [5], Cr-unet [7], Unet++ [6] and Learning-AC [9] were chosen to compare with our method on the experimental data. The Attention Unet, the Cr-unet and the Unet++ are the extensions of U-net [4]. In the Attention Unet, a novel attention gate, which can automatically focus on target structures of varying shapes and sizes, is integrated into a standard U-net architecture. The Cr-unet is aim to store prior representations of images by the spatial recurrent neural network to improve the performance of the Unet. The Unet++ realizes a novel feature fusion solution that aggregates the representation features on different scales by the decoder and the redesigned skip connections in our network. The Learning-AC is also a deep learning-based active contour model that designs a novel loss function to learn the area and size features of object regions and constrain the change of an active contour during each iteration.

We firstly compared the segmentation results of the compared methods on all test data quantitatively(see Fig. 6). It can be observed that the Learning-AC and our method obtain better scores on the mean values of both HD and Dice metrics. It indicates that the combing deformable contours and CNNs can achieve better results than the other methods using only CNNs in our experiments. In addition, compared to the other methods, the medians of both HD and Dice computed by the results of our method are more close to the mean values. This finding shows that our method provides more robust results for images of differing quality.

In addition, the segmentation results of all the methods on five randomly selected image sequences were compared quantitatively. Table 2 shows the mean and standard deviation (SD) of HD and Dice of the segmentation results on the three groups. In the high and mediam-level groups, the segmentation results obtained by all of compared methods were quite satisfactory, even from the mean values of HD and Dice, and the Learning-AC obtained better results than our method in the high-level group. However, it is worth noting that in the low-level group, our method clearly achieves better segmentation results than the other methods, and our method gives lower SD values than the other methods in both metrics, which means that our method gives more stable segmentation results at different image qualities.

Figure 7 shows the example of segmenting one image sequence. For the sequence, seven slices are chosen to show the results of the compared methods. The red curve in the iamges denotes the results of the manual segmentation. The columens 2 to 3 show the segmented examples computed by the Attention Unet, the Cr-unet and the Unet++. In the fourth column and fifth column, the green curves show the results computed by the Learning-AC and our method, respectively. It is observed that the shapes of the lesion regions in the sequence resemble an ellipse, but there are artifacts inside or outside of the lesion regions. The results of the Attention Unet , the Cr-unet and the Unet++ were interfered by buzzy or leaking boundaries. The results computed by our method can effectively resist the image defects.

6 Discussion

In this section, the effect of some components in the proposed method on performance is discussed in detail.

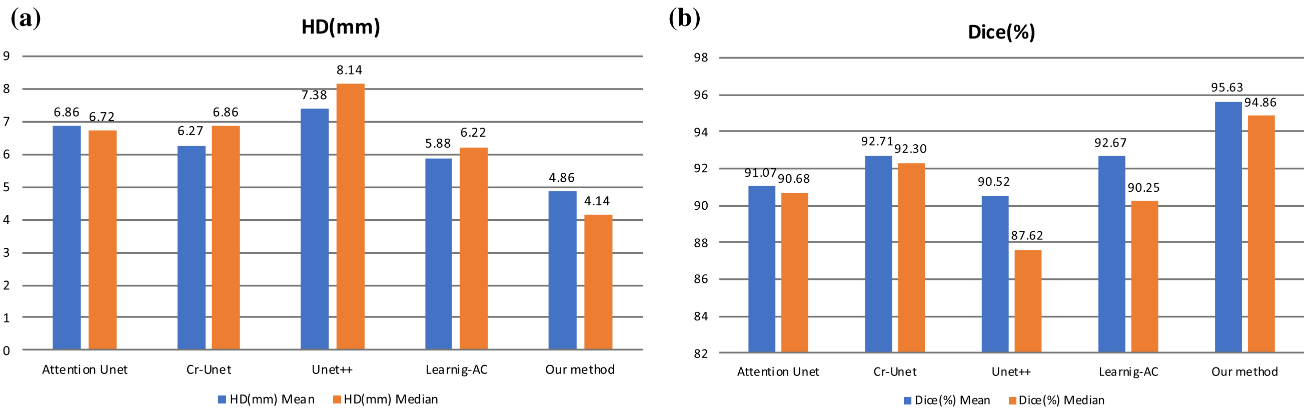


Fig. 6 Statistics of HD and Dice values of segmenting ultrasonic image sequences with the selected methods

Table 2 Performance statistics of our method versus other methods to segment results on the uterine fibroid ultrasound image sequences of different quality

High-level group		
Method	HD (mm)	Dice (%)
Attention unet ([5])	6.62 ± 3.69	91.68 ± 6.98
Cr-unet ([7])	6.32 ± 4.33	92.22 ± 6.12
Unet++ ([6])	7.62 ± 4.37	91.13 ± 6.54
Learning-AC ([9])	5.44 ± 3.59	94.52 ± 5.98
Our method	5.88 ± 2.91	92.12 ± 4.54
Medium-Level Group		
Attention unet ([5])	6.78 ± 3.69	85.68 ± 9.54
Cr-unet ([7])	6.02 ± 4.87	88.31 ± 9.48
Unet++ ([6])	7.05± 4.16	89.12 ± 8.05
Learning-AC ([9])	5.86 ± 3.94	91.71 ± 6.87
Our method	5.16 ± 3.34	92.22 ± 6.23
Low-Level Group		
Attention Unet ([5])	12.26 ± 6.12	82.33 ± 9.54
Cr-unet ([7])	9.02 ± 7.87	83.31 ± 9.48
Unet++ ([6])	10.05± 6.16	87.12 ± 9.05
Learning-AC ([9])	9.66 ± 6.94	89.10 ± 8.87
Our method	8.84 ± 4.76	90.82 ± 7.03

6.1 The impact of size of feature memory module

Intuitively, the larger the size of the feature memory module, the more discriminative the priori features it provides, but the computational cost will increase accordingly. Figure 8a shows that the average running-time of our method is longer than that of the other merhods. The main reason is that the computing efficiency of PCA operation in the feature memory modul is low. We tried to reduce the size of feature memory module and analysed the impact to the segmention results. Figure 8b shows that when the size of the module decreases exponentially, the decrease of segmentation performance deteriorates relatively smoothly.

6.2 Ablation study of network structure

To evaluate the effectiveness of the dense connections and attention focused blocks in our model, we performed a set of ablation experiments to analyse the performance of our model. First, we designed two configurations of our model, i.e. using only DCRFs(refers to DC-Net) and using only AFBs (refers to AF-Net), to analyse the learning behaviours of the network. Figure 9 and 10 show the training loss and validation loss of the different models, respectively. It can be observed that the proposed model converges faster and achieve lower validation loss than the network at the other configurations. Figure 10 further shows that the DCRFs can accelerate the convergence speed on the limited training data and the AFBs can alleviate the risk of noise in ultrasound images.

6.3 Impact of initializing contours

The influence of the initialized contour on the final result is mainly in two aspects, the placement and the number of sampling points. In general, the closer the initialized contour is to the target area, the smaller the range of the contour search and the smaller the number of iterations. Conversely, the farther away from the target area, the more iterations. The number of sampling points is also a factor affecting the segmentation accuracy. A higher number means that the more intensive sampling of the target edge features, the more accurate the segmentation results, but it also affects the computational efficiency of the method. And vice versa. In conclusion, the balance between the position of the initialized contour and the number of

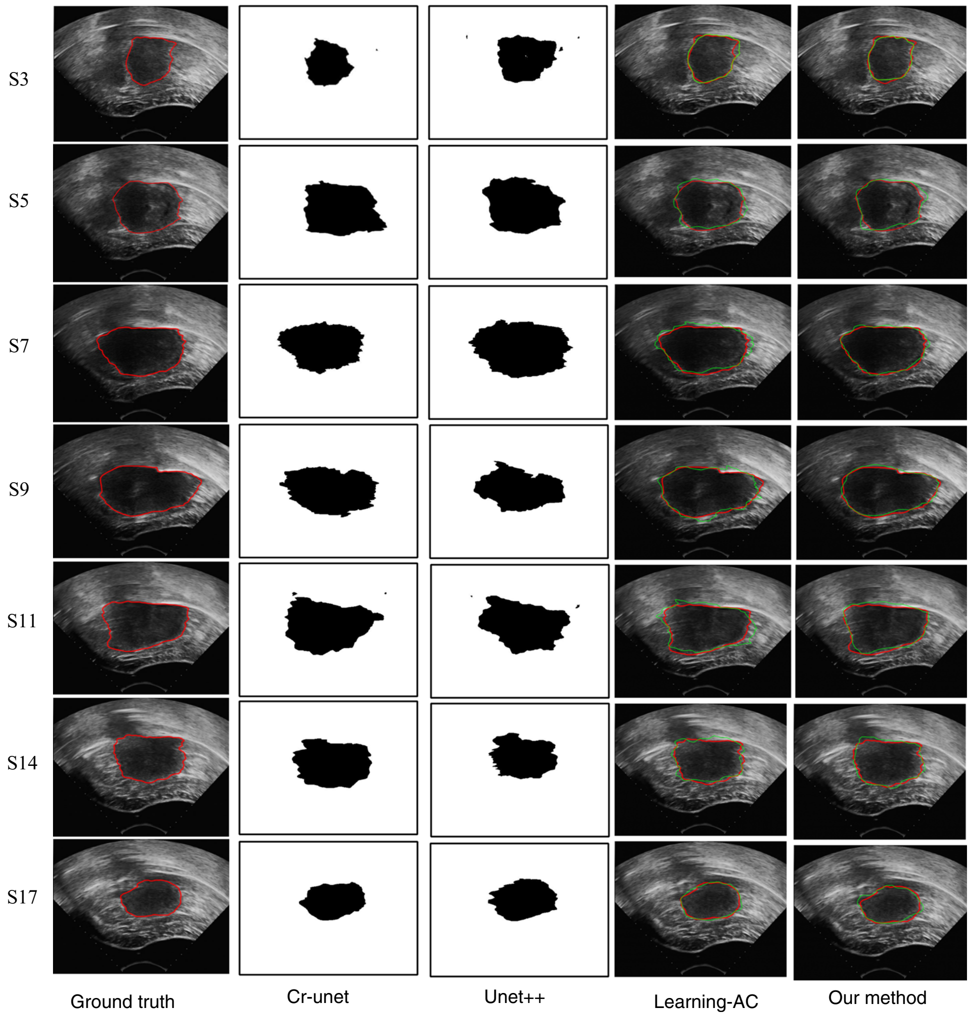


Fig. 7 Example of segmenting one HIFU uterine fibroid ultrasound image sequence with the compared methods

sampling points and the computational efficiency is an empirical.

6.4 Impact of α

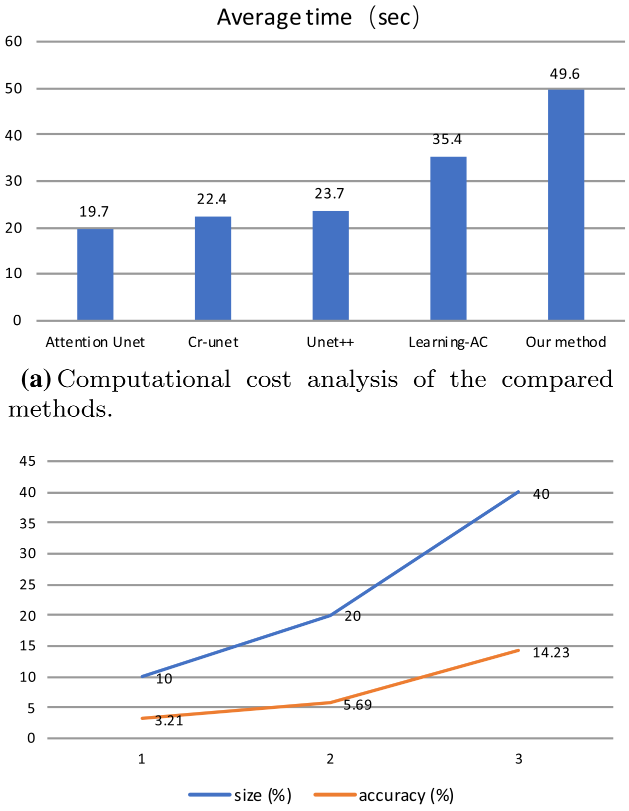
The parameter α in (7) is an important parameter of controlling the updating of the feature memory module, which has an important effect on the contrastive loss. Here, we selected α empirically and applied the same value to all experiments. Intuitively, the higher the update rate means that the more features are stored in the memory module, and the value of the loss function should be smaller. Figure 11 shows the changes of our loss with different α values in the testing data. It can be observed that the value of $\alpha > 0.45$ and the Loss values become stable. It is possible that PCA operator plays an important role in our analysis.

6.4.1 Impact of β

The parameter β in (10) is an important parameter of controlling the shape similarity between the deformable contours during segmentation of HIFU image sequences. Thus, we specified β empirically in our experiments. Figure 12 shows the influence to the average values of HD and Dice with the change in β in our experiments. We observe that the segmentation accuracy increases, while the β is specified within the range. This finding implies that the proposed shape similarity constrain is important for improving the performance of the segment ultrasound image sequences. However, when β is not in this range, the accuracy decreases as excessive regularization produces a large bias in the shape constrain.

6.5 Size of foreground / background patches

Intuitively, the performance of the proposed network is impacted by the size of the foreground and background patches. Since the significant grayscale variation near the target boundaries in ultrasound images, when the size of the patches are set large, the foreground patches may contain some non-target features. Similarly, if the size is small, the patches does not contain enough information



(a) Computational cost analysis of the compared methods.

(b) The impact of the different size of the module on segmentation performance.

Fig. 8 The impact of the feature memory module

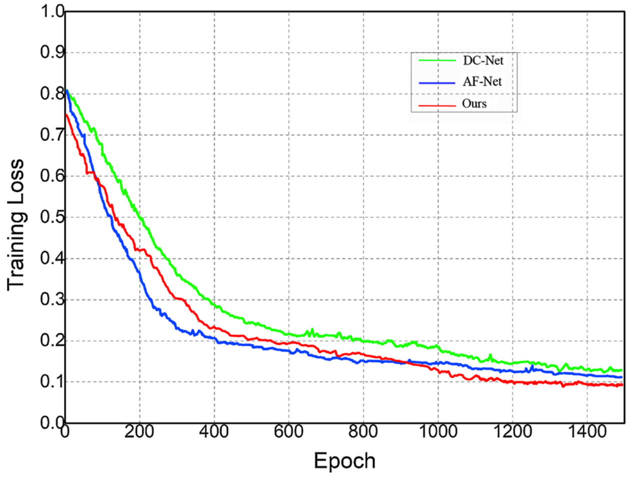


Fig. 9 Training loss of the network with different structures

about the foreground or the background, which also reduces the ability of the network to learn the discriminative features. Therefore, in practice, we usually choose some empirical values based on the image quality.

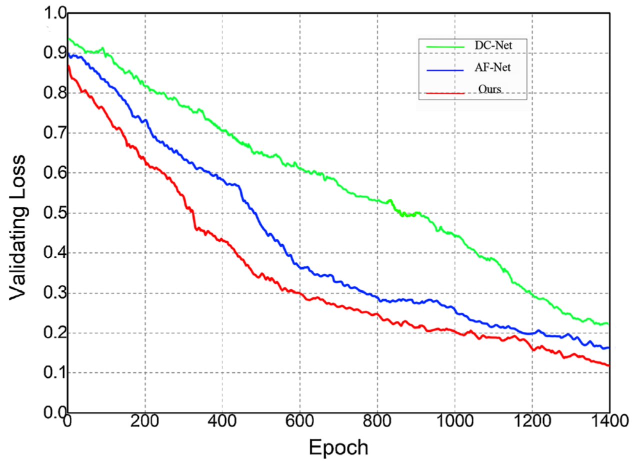


Fig. 10 Validation loss of the network with different structures

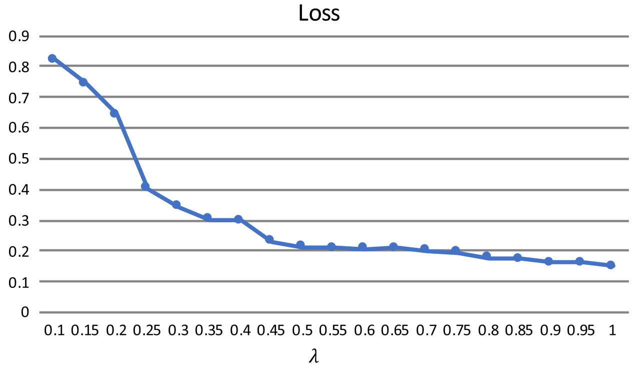


Fig. 11 Impact of the different α values on the Loss results

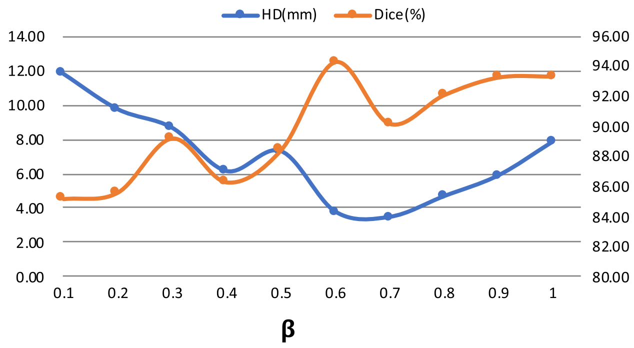


Fig. 12 Impact of the different β values on the segmenting results

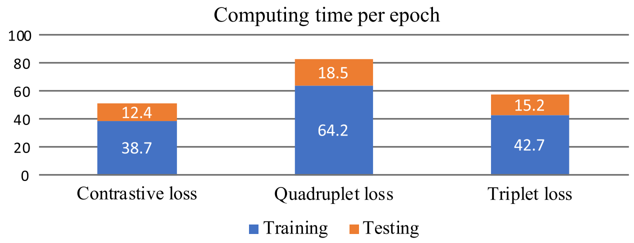


Fig. 13 Running time per epoch for the different loss functions with the proposed network

Table 3 Performance comparison of the proposed method with the two loss functions

Loss function	HD ± SD(mm)	Dice ±SD(%)
The constrastive loss	3.69 ± 1.87	95.23 ± 3.84
The quadruplet loss	4.12 ± 2.59	93.81 ± 3.67
The triplet loss	3.31 ± 1.98	95.22 ± 1.98

6.6 Effect of different contrastive loss functions

For evaluating the validity of the triplet loss in our method, we compared the impacts of the quadruplet loss on the performance of the method with the contrastive loss. These metric function are described detailly in [72]. Figure 13 shows that the computing time per epoch for the contrastive loss is 38.7s, which is shorter than the triplet loss, during training for the proposed network. For testing, the constrastive loss and the triplet loss take almost the same time when the proposed network is used. Table 3 shows the performance comparision of our method though the two loss functions. It shows that the network with the triplet loss can achieve better HD and Dice scores than the network with another loss functions.

6.7 The limitation of method

According to the above discussion, there are three limitations in our method:

- The premise of using the matrix rank as the measurement of shape similarity to constrain the evolution of the deformable contours is that the change of the target shape in the image sequence satisfies the linear change. When the shapes of tissues are complex, this prior model cannot precisely reflect the details of the shape.
- The balance between the performance and the memory consumption and computational efficiency of our method is a dilemma. To be specific, the larger size of the memory feature module, the object and background feature spaces are more robust, but it also consumes the memory of computer, and the computational efficiency of the operation PCA applied to the module is low.
- The cost of running memory and computation is still high, which is a bottleneck of this method.

7 Conclusion and futhure work

In this paper, we propose a novel deep siamese network to improve the performance of deformation contour model in segmentation of uterine fibroid ultrasound image sequence. The experimental results enable us to reach the following conclusions:

- Compared with end-to-end deep learning segmentation methods, combining the respective advantages of deformable contours model and deep learning networks can be seen as a viable way to effectively overcome the artefact and noise in ultrasound images and the lack of training data.
- The powerful ability of deep siamese networks in learning image features can provide more discriminative edge search clues for deformable contours.
- The dense connections and attention mechanism enable the deep siamese network to enhance the propagation of feature maps and the gradient information of lesion region in ultrasound images in forward and backward directions and simultaneously address the vanishing gradient issues, which is significant to preserve the weak boundary feature of the lesion region and prevent the interference of noise in ultrasound image segmentation.

In further research, we will introduce new attention mechanism into deep learning networks for medical image segmentation, such as Transformer model [65], and combine 2D and 3D convolution to segment image sequences.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China (No.62172438), the fundamental research funds for the central universities (31412111303, 31512111310) and by the open project from the State Key Laboratory for Novel Software Technology, Nanjing University, under Grant No.KFKT2019B17.

Declaration

Conflict of interest The authors declare that there is no conflict of interests regarding the publication of this article.

References

1. Kim J, Choi W, Park EY, Kang Y, Lee KJ, Kim HH, Kim WJ, Kim C (2019) Real-time photoacoustic thermometry combined with clinical ultrasound imaging and high-intensity focused ultrasound. IEEE Trans Biomed Eng 66(12):3330–3338
2. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp 3431–3440
3. Hu Y, Soltoggio A, Lock R, Carter S (2019) A fully convolutional two-stream fusion network for interactive image segmentation. Neural Netw:Off J Int Neural Netw Soc 109:31–42

4. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing & Computer-Assisted Intervention
5. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, Rueckert D (2018) Attention u-net: Learning where to look for the pancreas. arXiv:1804.03999
6. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2019) Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Transactions on Medical Imaging
7. Li H, Fang J, Liu S, Liang X, Yang X, Mai Z, Van MT, Wang T, Chen Z, Ni D (2019) Cr-unet: A composite network for ovary and follicle segmentation in ultrasound images. IEEE J Biomed Health Inform 24:974–983
8. Marcos D, Tuia D, Kellenberger B, Zhang L, Bai M, Liao R, Urtasun R (2018) Learning deep structured active contours end-to-end. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition pp 8877–8885
9. Chen X, Williams BM, Vallabhaneni SR, Czanner G, Williams RS, Zheng Y (2019) Learning active contour models for medical image segmentation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp 11624–11632
10. Fang Z, Qiao M, Guo Y, Wang Y, Li J, Zhou S, Chang C (2019) Combining a fully convolutional network and an active contour model for automatic 2d breast tumor segmentation from ultrasound images. J Med Imaging Health Informatics 9:1510–1515
11. Hu Y, Guo Y, Wang Y, Yu J, Li J, Zhou S, Chang C (2019) Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model. Med Phys 46:215–228
12. Xu C, Prince JL (1998) Snakes, shapes, and gradient vector flow. IEEE Trans Image Process 7(3):359–69
13. Paragios N, Deriche R (2000) Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Trans Pattern Anal Mach Intell 22(3):266–280
14. Guo Y, Şengür A, Tian JW (2016) A novel breast ultrasound image segmentation algorithm based on neutrosophic similarity score and level set. Comput Methods Programs Biomed 123:43–53
15. Zhao Y, Zhao J, Yang J, Liu Y, Zhao Y, Zheng Y, Xia L, Wang Y (2017) Saliency driven vasculature segmentation with infinite perimeter active contour model. Neurocomputing 259:201–209
16. Radoglou-Grammatikis P, Robolos K, Sarigiannidis P, Argyriou V, Lagkas T, Sarigiannidis A, Goudos SK, Wan S (2021) Modelling, detecting and mitigating threats against industrial healthcare systems: a combined sdn and reinforcement learning approach. IEEE Transactions on Industrial Informatics
17. Wan S, Xia Y, Qi L, Yang YH, Atiquzzaman M (2020) Automated colorization of a grayscale image with seed points propagation. IEEE Transactions on Multimedia
18. Zhou S, Wang J, Zhang S, Liang Y, Gong Y (2016) Active contour model based on local and global intensity information for medical image segmentation. Syst Eng Electron 186((C)):107–118
19. Yu H, He F, Pan Y (2018) A novel region-based active contour model via local patch similarity measure for image segmentation. Multimedia Tools Appl 3:1–23
20. Wang, L., Li, M., Fang, X., Nappi, M., & Wan, S. (2022). Improving random walker segmentation using a nonlocal bipartite graph. Biomedical Signal Processing and Control, 71, 103154.
21. Gao Y, Bouix S, Shenton ME, Tannenbaum AR (2013) Sparse texture active contour. IEEE Trans Image Process 22:3866–3878
22. Yu J, Tan M, Zhang H, Tao D, Rui Y (2019) Hierarchical deep click feature prediction for fine-grained image recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence
23. Zhao Y, Li H, Wan S, Sekuboyina A, Hu X, Tetteh G, Piraud M, Menze B (2019) Knowledge-aided convolutional neural network for small organ segmentation. IEEE J Biomed Health Inform 23(4):1363–1373
24. Fang J, Liu H, Liu J, Zhou H, Liu H (2021) Fuzzy region-based active contour driven by global and local fitting energy for image segmentation. Appl Soft Comput 100:106982
25. Subudhi P, Mukhopadhyay S (2021) A statistical active contour model for interactive clutter image segmentation using graph cut optimization. Signal Process 4:108056
26. Van Ginneken B, Frangi AF, Staal JJ, Haar TBM, Romeny VMA (2002) Active shape model segmentation with optimal features. IEEE Trans Med Imaging 21(8):924–933
27. Zhang S, Zhan Y, Dewan M, Huang J, Metaxas DN, Zhou XS (2012) Towards robust and effective shape modeling: Sparse shape composition. Med Image Anal 16(1):265–277
28. Huang X, Dione DP, Compas CB, Papademetris X, Lin BA, Bregasi A, Sinusas AJ, Staib LH, Duncan JS (2014) Contour tracking in echocardiographic sequences via sparse representation and dictionary learning. Med Image Anal 18(2):253–271
29. Korez R, Ibragimov B, Likar B, Pernuš F, Vrtovec T (2015) A framework for automated spine and vertebrae interpolation-based detection and model-based segmentation. IEEE Trans Med Imaging 34(8):1649–1662
30. Ni B, He F, Yuan ZY (2015) Segmentation of uterine fibroid ultrasound images using a dynamic statistical shape model in hifu therapy. Comput Med Imaging Graph 46(3):302–314
31. Wang, H., Zhang, D., Ding, S. et al. Rib segmentation algorithm for X-ray image based on unpaired sample augmentation and multi-scale network. Neural Comput & Applic (2021). https://doi.org/10.1007/s00521-021-06546-x
32. Wachinger C, Yigitsoy M, Rijkhorst EJ, Navab N (2012) Manifold learning for image-based breathing gating in ultrasound and mri. Med Image Anal 16(4):806–818
33. Gao Z, Li Y, Wan S (2020) Exploring deep learning for view-based 3d model retrieval. ACM Trans Multimed Comput Commun Appl 16(1):1–21
34. Zhou X, Huang X, Duncan JS, Yu W (2013) Active contours with group similarity. 2013 IEEE Conference on Computer Vision and Pattern Recognition pp 2969–2976
35. Ni B, He F, teng Pan Y, Yuan Z (2016) Using shapes correlation for active contour segmentation of uterine fibroid ultrasound images in computer-aided therapy. Appl Math-A J Chin Univ 31:37–52
36. Gamarra M (2020) Convexity shape constraints for retinal blood vessel segmentation and foveal avascular zone detection. Computers in Biology and Medicine 127
37. Yan S, Tai XC, Liu J, Huang HY (2020) Convexity shape prior for level set based image segmentation method. IEEE Transactions on Image Processing PP(99):1–1
38. Milletari F, Navab N, Ahmadi SA (2016) V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 Fourth International Conference on 3D Vision (3DV) pp 565–571
39. Alom MZ, Yakopcic C, Hasan M, Taha T, Asari V (2019) Recurrent residual u-net for medical image segmentation. J Med Imaging 6:014006–014006
40. Yu J, Li J, Yu Z, Huang Q (2019) Multimodal transformer with multi-view visual representation for image captioning. IEEE Transactions on Circuits and Systems for Video Technology PP(99):1–1
41. Ding S, Qu S, Xi Y, Wan S (2019) Stimulus-driven and concept-driven analysis for image caption generation. Neurocomputing

42. Huang G, Liu Z, Weinberger KQ (2017) Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp 2261–2269
43. Yang W, Zhou Q, Lu J, Wu X, Zhang S, Latecki L (2018) Dense deconvolutional network for semantic segmentation. 2018 25th IEEE International Conference on Image Processing (ICIP) pp 1573–1577
44. Li H, He X, Zhou F, Yu Z, Ni D, Chen S, Wang T, Lei B (2019) Dense deconvolutional network for skin lesion segmentation. IEEE J Biomed Health Inform 23:527–537
45. Park B, Yu S, Jeong J (2019) Densely connected hierarchical network for image denoising. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) pp 2104–2113
46. Yuan Y, Qin W, Guo X, Buyyounouski M, Hancock SH, Han B, Xing L (2019) Prostate segmentation with encoder-decoder densely connected convolutional network (ed-densenet). 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) pp 434–437
47. Minnema J, Wolff J, Koivisto J, Lucka F, Batenburg KJ, Forouzanfar T, Eijnatten M (2021) Comparison of convolutional neural network training strategies for cone-beam ct image segmentation. Comput Methods Prog Biomed 207:106192
48. Fu J, Liu J, Tian H, Fang Z, Lu H (2019) Dual attention network for scene segmentation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp 3141–3149
49. Wang G, Li W, Zuluaga MA, Pratt R, Patel PA, Aertsen M, Doel T, David AL, Deprest J, Ourselin S et al (2018) Interactive medical image segmentation using deep learning with image-specific fine tuning. IEEE Trans Med Imaging 37(7):1562–1573
50. Shan H, Zhang Y, Yang Q, Kruger U, Kalra MK, Sun L, Cong W, Wang G (2018) 3-d convolutional encoder-decoder network for low-dose ct via transfer learning from a 2-d trained network. IEEE Trans Med Imaging 37(6):1522
51. Opbroek AV, Achterberg HC, Vernooij MW, Bruijne MD (2018) Transfer learning for image segmentation by combining image weighting and kernel learning. IEEE Transactions on Medical Imaging PP(99):1–1
52. Han D, Liu Q, Fan W (2018) A new image classification method using cnn transfer learning and web data augmentation. Expert Syst Appl 95:43–56
53. Yu J, Zhu C, Zhang J, Huang Q, Tao D (2019) Spatial pyramid-enhanced netvlad with weighted triplet loss for place recognition. IEEE Transactions on Neural Networks and Learning Systems
54. Dong N, Trullo R, Lian J, Li W, Petitjean C, Su R, Qian W, Shen D (2018) Medical image synthesis with deep convolutional adversarial networks. IEEE Transactions on Biomedical Engineering PP(99):1–1
55. Wolterink JM, Kamnitsas K, Ledig C, Išgum I (2018) Generative adversarial networks and adversarial methods in biomedical image analysis. arXiv preprint arXiv:1810.10352
56. Yu B, Zhou L, Wang L, Shi Y, Fripp J, Bourgeat P (2019) Ea-gans: Edge-aware generative adversarial networks for cross-modality mr image synthesis. IEEE Transactions on Medical Imaging PP(99):1–1
57. Li Y, Chen Y, Shi Y (2020) Brain tumor segmentation using 3d generative adversarial networks. International Journal of Pattern Recognition and Artificial Intelligence
58. Lei B, Xia Z, Jiang F, Jiang X, Wang S (2020) Skin lesion segmentation via generative adversarial networks with dual discriminators. Med Image Anal 64:101716
59. Kugelman J, Alonso-Caneiro D, Read SA, Vincent SJ, Collins MJ (2021) Data augmentation for patch-based oct chorio-retinal segmentation using generative adversarial networks. Neural Computing and Applications (4)
60. Zagoruyko S, Komodakis N (2015) Learning to compare image patches via convolutional neural networks. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp 4353–4361
61. Cai Q, Pan Y, Yao T, Yan CC, Mei T (2018) Memory matching networks for one-shot image recognition. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition pp 4080–4088
62. Li Z, Kamnitsas K, Glocker B (2021) Analyzing overfitting under class imbalance in neural networks for image segmentation. IEEE Trans Med Imaging 40(3):1065–1077. https://doi.org/10.1109/TMI.2020.3046692
63. Kamran SA, Hossain KF, Tavakkoli A, Zuckerbrod SL, Sanders KM, Baker SA (2021) Rv-gan : Retinal vessel segmentation from fundus images using multi-scale generative adversarial networks
64. Petit O, Thome N, Rambour C, Soler L (2021) U-net transformer: Self and cross attention for medical image segmentation
65. Valanarasu JMJ, Oza P, Hacihaliloglu I, Patel VM (2021) Medical transformer: Gated axial-attention for medical image segmentation. arXiv:2102.10662
66. He K, Zhang X, Ren S, Sun J (2014) Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell 37(9):1904–16
67. Wold S, Esbensen KH, Geladi P (1987) Principal component analysis
68. Ye Y, He C (2012) Adaptive active contours without edges. Math Comput Model 55(5–6):1705–1721
69. Huang Y, Yan HY, Wen YW, Yang X (2018) Rank minimization with applications to image noise removal. Inf Sci 429:147–163
70. Beck A, Teboulle M (2009) A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J Imaging Sci 2:183–202
71. Cai J, Candès E, Shen Z (2010) A singular value thresholding algorithm for matrix completion. SIAM J Optim 20:1956–1982
72. Hermans A, Beyer L, Leibe B (2017) In defense of the triplet loss for person re-identification. arXiv:1703.07737

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.