Automatic spinal cord segmentation from axial-view MRI slices using CNN with grayscale regularized active contour propagation

Xiaoran Zhang [a,c,1,2], Yan Li [b,1], Yicun Liu [a,1], Shu-Xia Tang [d], Xiaoguang Liu [b,**], Kumaradevan Punithakumar [e], Dawei Shi [a,*]

[a] School of Automation, Beijing Institute of Technology, Beijing, 100081, China
[b] Department of Orthopaedics, Peking University Third Hospital and the Engineering Research Center of Bone and Joint Precision Medicine, Ministry of Education, Beijing, China
[c] Department of Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA, 90095-1594, USA
[d] Department of Mechanical Engineering, Texas Tech University, Lubbock, TX, 79409, USA
[e] Department of Radiology and Diagnostic Imaging, University of Alberta, Edmonton, Alberta, 8440, Canada

ARTICLE INFO

ABSTRACT

Accurate positioning of the responsible segment for patients with cervical spondylotic myelopathy (CSM) is clinically important not only to the surgery but also to reduce the incidence of surgical trauma and complications. Spinal cord segmentation is a crucial step in the positioning procedure. This study proposed a fully automated approach for spinal cord segmentation from 2D axial-view MRI slices of patients with CSM. The proposed method was trained and tested using clinical data from 20 CSM patients (359 images) acquired by the Peking University Third Hospital, with ground truth labeled by professional radiologists. The accuracy of the proposed method was evaluated using quantitative measures, the reliability metric as well as visual assessment. The proposed method yielded a Dice coefficient of 87.0%, Hausdorff distance of 9.7 mm, root-mean-square error of 5.9 mm. Higher conformance with ground truth was observed for the proposed method in comparison to the state-of-the-art algorithms. The results are also statistically significant with *p*-values calculated between state-of-the-art methods and the proposed methods.

1. Introduction

Cervical spondylotic myelopathy (CSM) is a clinical manifestation of a series of neurological damage caused by degenerative changes of the cervical intervertebral disc and intervertebral joints, which leads to the compression of cervical spinal cords [4,42]. Generally, CSM with two or more segments of cervical spinal cord compression is called multilevel cervical spondylotic myelopathy (MCSM), which is a common variation of CSM. The morbidity rate of CSM is reported in the literature between 41 and 2310 in a million [32,33,41]. Surgical treatment is the primary approach for patients with clearly diagnosed CSM according to evident-based medical reports [15]. The location of the disease-causing segment, also known as the responsible segment, that requires surgical

treatment of CSM is currently highly dependent on imaging, especially on magnetic resonance imaging (MRI). However, patients who have no clinical symptoms but have asymptomatic compression on MRI images are not uncommon [21,27]. For patients with MCSM defined by imaging, surgeons currently use individual experience combined with imaging to locate the disease-causing segment, which may unnecessarily expand the scope of surgery due to no recognized and objective criteria for locating disease-causing segments. Therefore, accurate localization of the pathogenic segment of the CSM will allow for more precise surgical procedures and reduce the incidence of surgical trauma and surgical complications, reducing the patient surgical costs [14,15,45].

Segmenting the spinal cord is an important step for the diagnosis and management of CSM. Several techniques were proposed for MRI spinal
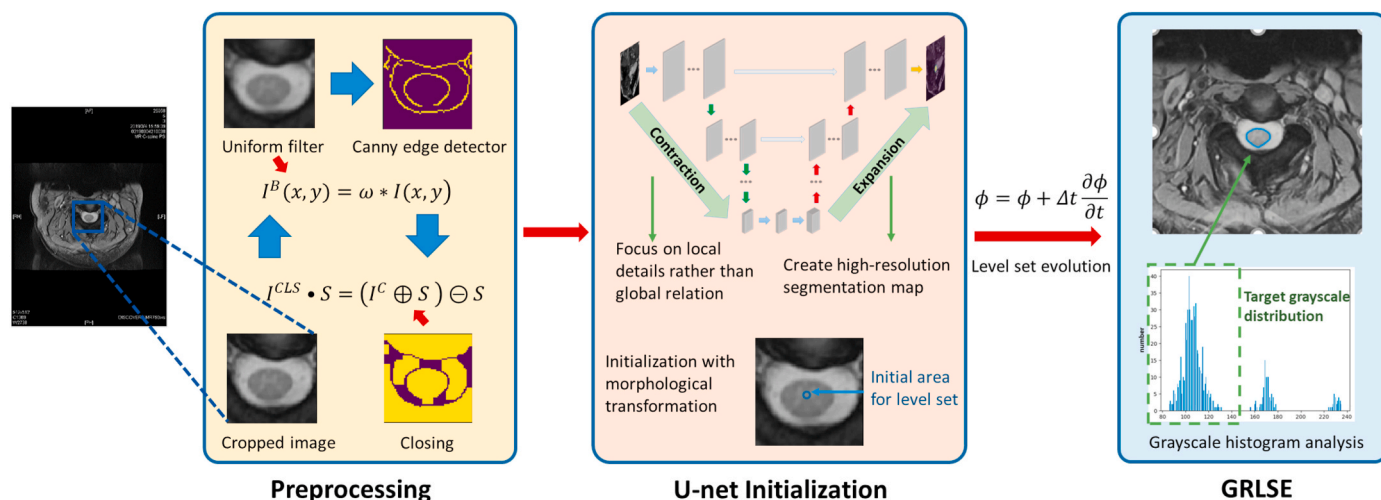

Fig. 1. The overall architecture of our proposed method.

cord segmentation, most of which are semi-automated, in the past two decades. Behrens et al. [3] proposed a robust semi-automated segmentation method of tubular structures in 3D medical images by parametric object detection and tracking. The method demonstrates impressive performance modeling the human spinal cord for segmentation, given manual empirical selection of the start point. McIntosh and Hamarneh [30] proposed a semi-automatic segmentation approach that focuses on 3D geometry to delineate the structure from 3D MRI. Mayerich and Keyser [28] proposed a hardware-accelerated segmentation tool focused on 3D volumetric data and tested on the mouse spinal cord. Another semi-automatic segmentation is proposed by Horsfield et al. [19] that requires the user to mark approximated cord center-line on a few representative slices. Mukherjee et al. [31] proposed an automatic segmentation of the spinal cord that assumes the spinal cord is symmetric in the MRI slice. Due to the degenerative changes and compression nature of CSM patients, the spinal cord pattern could be asymmetric. Chen et al. [6] proposed a topology-preserving automatic segmentation in MRI which was tested on a set of 20 images. Koh et al. [22] proposed an automatic segmentation scheme for the spinal cord based on sagittal T_2-weighted MR images instead of the axial view. There are also semi-automated approaches [17,23] focusing on spinal canal segmentation using region growing method that require human intervention for initialization. Similarly, De Leener et al. [9] proposed a semi-automated segmentation scheme for 2D MRI slices using iterative propagation of a deformable model that requires a manual starting point and orientation. An extended version of the method is proposed in Ref. [10] for the semi-automated segmentation of both spinal cord and canal that requires initialization of an approximated spinal cord position and orientation. Urban and Tanacs [43] proposed an atlas-based global and local segmentation scheme and tested on the spinal cord. There are also several studies recently [1,11,35] segmenting gray matter [8] and white matter inside the spinal cord in MRI and several on sagittal view [38].

With the advent of deep learning, a number of approaches were proposed recently for spinal cord segmentation most of which rely on computed tomography (CT) images. Vania et al. [44] proposed an automatic segmentation using convolutional neural network (CNN) from CT images and achieved a high Dice score. Inspired by the popular U-net based CNN architecture [37], Dong et al. proposed an automatic multiorgan segmentation technique and tested the approach on the spinal cord in thorax CT images using U-net GAN [12]. Jakubiecek et al. [20] proposed another deep learning-based fully automatic spine center lie detection in CT images. Although CT imaging offers increased patient comfort, faster scanning times and higher resolution images output, MRI is often prescribed by physicians due to its better contrast in pictures of

organs, soft tissues, bone and other internal body structures without the use of ionizing radiation or intravenous contrast media [34]. To the best of the authors' knowledge, McCoy et al. [29] proposed the first deep learning-based 2D MRI segmentation scheme for the spinal cord with a thorough analysis of MRI clinical images. The group achieved a high Dice score, however, the detailed CNN architecture information is not provided in the paper. Gros et al. [18] also proposed an automatic segmentation framework with CNN that relied on 3D input image volume.

In this work, we aim to develop a fully automatic segmentation algorithm for 2D MRI axial-view slices without manual initialization and improve the convergence performance for the active contour methods. Compared with existing 2D MRI segmentation approaches in the literature, our major contributions are: (i) we propose a novel fully automatic framework for axial-view 2D MRI slices segmentation from CSM patients with little manual ground truth contours for training; (ii) by utilizing a CNN architecture, manual initialization of the approximate spinal cord position and orientation is no longer required; (iii) the active contour method is strengthened with a grayscale regularizer with a better convergence in large iterations compared to the previous state-of-the-art level set approaches. We tested our proposed algorithm on clinical images provided by Peking University Third Hospital and evaluated with the dice metric, Hausdorff distance, root-mean-square error and reliability metric.

2. Methods

In typical deep learning-based approaches, larger volumes of data generally produce higher prediction accuracy. However, it is very time-consuming to generate a large volume of corresponding ground truth contours, which are usually segmented manually by clinicians. In order to mitigate the trade-off between precise deep learning-based segmentation and small data, our method integrates the advantages of strong generalization ability from the convolutional neural network with a novel level set evolution in a robust and effective fashion without the need to segment large medical ground truth for training. The overall flowchart is depicted in Fig. 1.

2.1. Preprocessing

The 2D MRI slices are firstly center cropped to 256×256 pixels and then processed by a 3×3 box blur kernel ω. For the input grayscale image I and $(x, y) \in \Omega$, we have
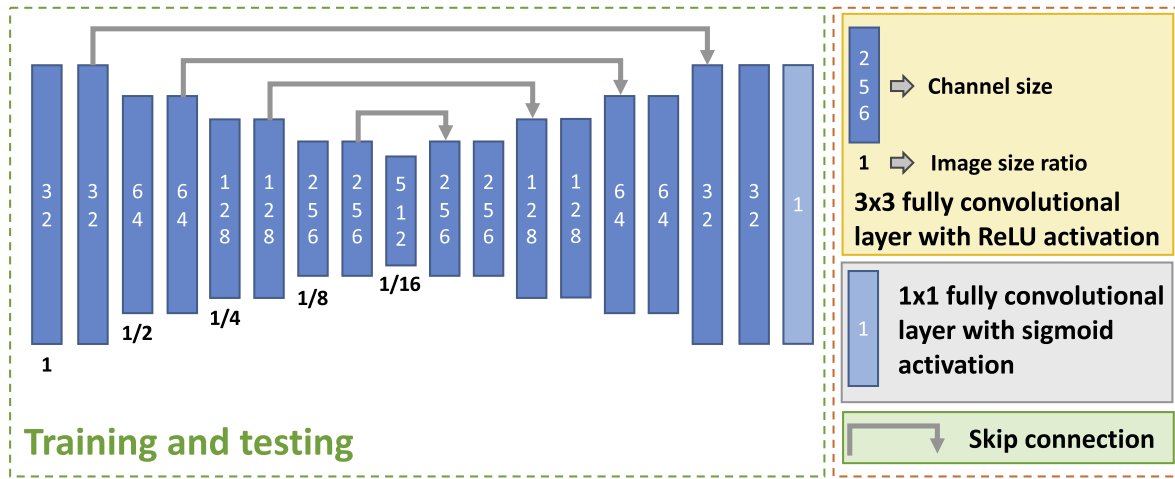
Fig. 2. U-net architecture used in our proposed method. Final output is a 256 × 256 image mask corresponding to the U-net spinal cord segmentation.

$$I^B[x,y] = \omega * I = \sum_{dx=0}^{2} \sum_{dy=0}^{2} \omega[dx,dy] \cdot I[x+dx, y+dy] \tag{1}$$

where I^B is the blurred image and $\Omega = \{0, 1, 2, ..., 255\}^2$ is the entire pixel set.

In order to enhance the boundary difference of the image, we utilize the Canny edge detector [5] for I^B. The image after Canny operator, denoted as I^C is then followed with a morphological closing operation for further preprocessing, which is implemented as:

$$I^{CLS}[x,y] = \left(I^C[x,y] \oplus S\right) \ominus S, \tag{2}$$

where I^{CLS} denotes the image after morphological closing; S is the structuring element in $\mathbb{R}^{7 \times 7}$; $\oplus$ and $\ominus$ denote the dilation and erosion

respectively. The closing operation removes the small holes inside the raw edge detected by the Canny operator and provides a pixel map for edge enhancement. The pixel on the edge of the final image I^F after preprocessing is weighted with a larger value to increase the boundary difference between the area inside the region of interest (RoI) and the area outside RoI. For $(x,y) \in \Omega_E$, where Ω_E is the edge pixel set detected by the Canny operator, we have

$$I^F[x,y] = \alpha \max_{(s,t) \in \Omega} \left\{ I^{CLS}[s,t] \right\} + (1-\alpha) I^{CLS}[x,y], \tag{3}$$

where $\alpha = 0.8$. For $(x,y) \notin \Omega_E$, we have

$$I^F[x,y] = I^B[x,y]. \tag{4}$$
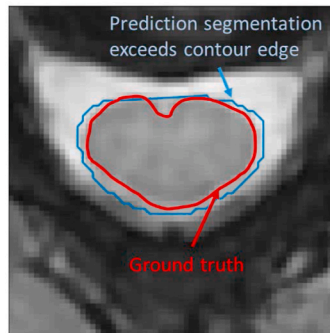
a) Example of U-net segmentation map compared to ground truth.
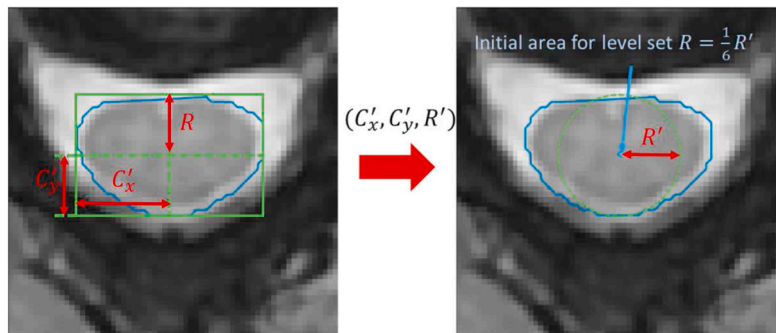
b) Morphological transformation of U-net segmentation.

Fig. 3. Example of raw U-net segmentation compared to ground truth. C_x' and C_y' denote the centroid of the contour in x and y dimension respectively. R' denotes the smallest inscribed circle inside the contour bounding box. R denotes the radius of the circle initialization of the active contour post-processing.

Patient 1   Patient 2   Patient 3

Fig. 4. Sample images from different patients performed using LSE with the same iterations (*iteration* = 500), where the first image indicates that the cu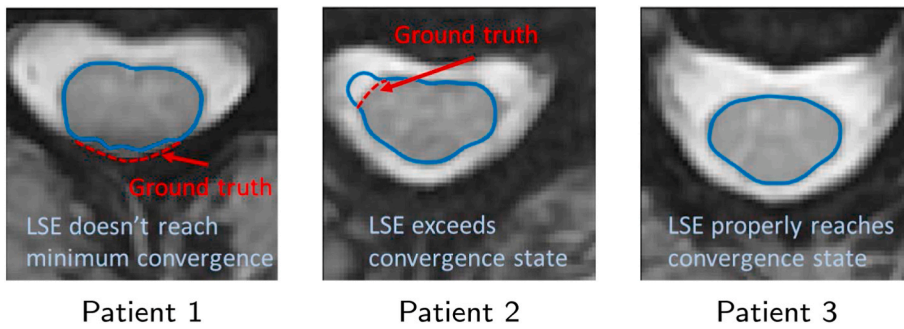rrent number of iteration is too small to converge, the second is enough but LSE doesn't properly stop at the boundary and the third properly match. The number of iteration is usually set to a large value in order to ensure active fronts fully extend to the boundary. The set of image indicates current convergence criterion needs to be strengthened as LSE cannot properly stop at the edge when the number of iteration is large.

2.2. U-net initialization

To obtain an initial location of the RoI, a common U-net architecture [37] is implemented to segment the target spinal cord. Fig. 2 showed the U-net architecture utilized in our proposed method. We utilized the negative of Dice coefficient as our loss function. Deep learning-based architectures are usually data-driven and the performance will be significantly enhanced when the dataset is larger [48]. However, when the number of training sample is limited, the prediction on the boundary is inaccurate and hard to manipulate as shown in Fig. 3a. Thus, we utilize an improved active contour method after the U-net prediction as the global location of the predicted segment is generally accurate.

In order to ensure the initial area inside the RoI, we transform the U-net segmentation by finding the rectangular bounding box shown in Fig. 3b. The initial area is a concentric circle of the smallest inscribed circle in the bounding box. We determine the center and radius of the inscribed circle by computing the geometric center and the width of the bounding rectangular, respectively. The radius of the initial area R is set to a proportion of the radius of the inscribed circle R'.

2.3. Grayscale regularized level set evolution (GRLSE)

Level-set methods are a common framework for using level sets as a tool to make numerical analysis of surfaces and shapes [47]. We assume C_φ as the parametric curve also known as fronts, which can be represented by the zero level set:

$$C_\varphi = \{(x,y)|\varphi(x,y,t) = 0\}, \ \forall t \geq 0 \tag{5}$$

where $\varphi(x,y,t)$ is the level set function (LSF). In the task of image segmentation, the boundary curve C_φ, represented by LSF φ, separates the foreground object from the background in an image. The foreground and background regions are usually denoted by $\{(x,y)|\varphi(x,y,t) < 0\}$ and $\{(x,y)|\varphi(x,y,t) > 0\}$ respectively. The LSF evolution is a powerful tool for segmentation task however it generally has two limitations. Firstly, the initial level set area needs to be defined manually, which includes human intervention making the method a semi-automatic approach instead of fully automatic. In last section, we introduce a U-net structed CNN to automatically initialize the area. Secondly, although the distance regularization effect of LSF demonstrates a significant improvement compared with conventional level set methods [25], the convergence of the edge-based active contour model is not stable as the stopping condition of the propagation is dependent on the iterations. To improve the boundary performance and automate the initialization of the active contour methods, we added a grayscale regularizer and utilized the automated U-net predictions to provide the initial estimate of the target region.

We firstly follow the standard steps proposed by Li et al. [25] by calculating the edge indicator function g

$$g = \frac{1}{1 + \left|\left|\nabla\left(G_\sigma * I^F\right)\right|\right|_2^2}, \tag{6}$$

where I_F is defined in Eq. (3) and Eq. (4). G_σ is the Gaussian kernel with

the standard deviation $\sigma = 0.8$. To solve the standard curve evolution in partial differential equation formulation

$$\frac{\partial\varphi}{\partial t} = F\left|\left|\varphi\right|\right|_2, \tag{7}$$

where F is the speed function that controls the motion of the contour, we formulate the LSF with Neumann boundary condition [13] and follow the energy functional defined in Ref. [25], which is

$$E(\varphi,g) = \mu\mathscr{R}_p(\varphi) + \lambda\mathscr{E}(\varphi,g) + \alpha\mathscr{A}(\varphi,g) \tag{8}$$

where the distance, edge, and area energy functional are defined as

$$\mathscr{R}_p(\varphi) = \int_\Omega p\left(\left|\left|\nabla\varphi\right|\right|_2\right)\mathrm{d}x\mathrm{d}y \tag{9}$$

$$\mathscr{E}(\varphi) = \int_\Omega g\delta_\varepsilon(\varphi)|\nabla\varphi|\mathrm{d}x\mathrm{d}y \tag{10}$$

$$\mathscr{A}(\varphi) = \int_\Omega gH_\varepsilon(\varphi)\mathrm{d}x\mathrm{d}y \tag{11}$$

and the double-well potential function $p(\cdot)$, Dirac delta function δ_ε, and Heaviside function H_ε are defined as

$$p(x) = \begin{cases} \frac{1}{(2\pi)^2}(1 - \cos(2\pi x)), & x \leq 1 \\ \frac{1}{2}(x-1)^2, & x > 1 \end{cases} \tag{12}$$

$$\delta_\varepsilon(x) = \begin{cases} \frac{1}{2\varepsilon}\left(1 + \cos\left(\frac{\pi x}{\varepsilon}\right)\right), & |x| \leq \varepsilon \\ 0, & |x| > \varepsilon \end{cases} \tag{13}$$

$$H_\varepsilon(x) = \begin{cases} \frac{1}{2}\left(1 + \frac{x}{\pi} + \frac{1}{\pi}\sin\left(\frac{\pi x}{\varepsilon}\right)\right), & |x| \leq \varepsilon \\ 1, & x > \varepsilon \\ 0, & x < -\varepsilon \end{cases} \tag{14}$$

The energy functional in Eq. (8) could be minimized by solving the gradient flow in the following

$$\frac{\partial\varphi}{\partial t} = \mu\mathrm{div}\left(d_p\left(\left|\left|\nabla\varphi\right|\right|_2\right)\nabla\varphi\right) + \lambda\delta_\varepsilon(\varphi)\mathrm{div}\left(g\frac{\nabla\varphi}{\left|\left|\nabla\varphi\right|\right|_2}\right) + \alpha g\delta_\varepsilon(\varphi) \tag{15}$$

given the initial LSF, φ^0 predicted by U-net. The distance regularization term is computed by

$$d_p(x) = \frac{p'(x)}{x}. \tag{16}$$

Level set methods require a sufficient large iteration number for segmentation to let the boundary curve C_φ fully extend to the edge.
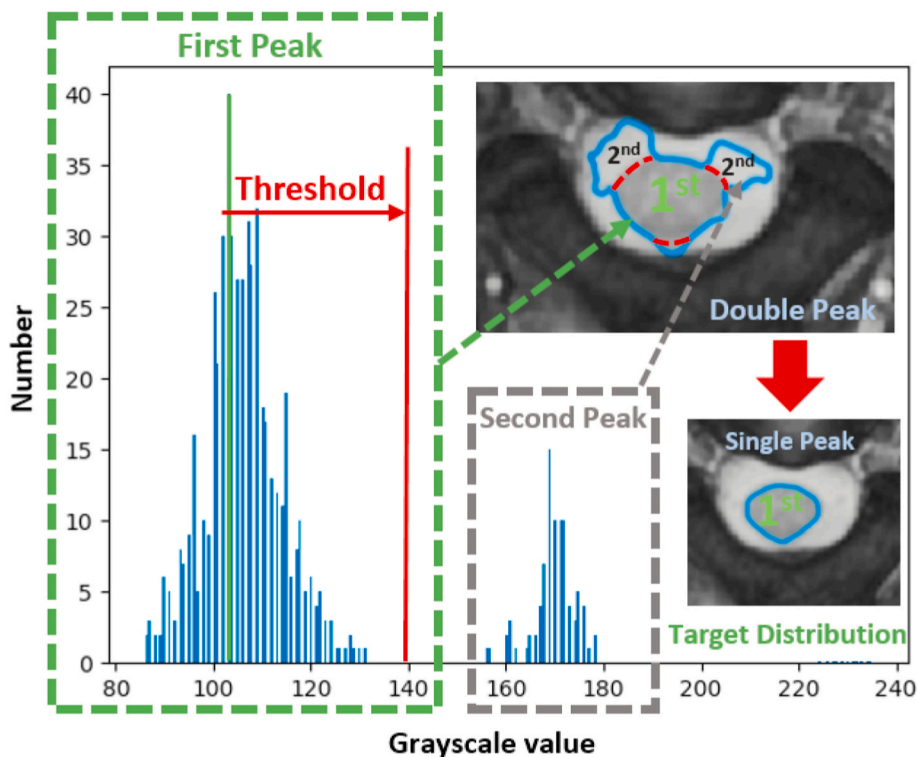
Fig. 5. Histogram of intensities after LSE where the first peak corresponds to the region of interest. We calculate the histogram after flattening the image grayscale value inside the LSE fronts and update convergence stopping criterion at each iteration.

Table 1

Details of data set used in the proposed study.

Description	Dataset
Number of subjects	20
Imaging modality	Transversal T2-weighted imaging
Patient's sex	11 Males/9 Females
Patient's age	39–75 years
Scanner protocol	Discovery MRI750 ws
Cord coverage	C3 – C7
Magnetic strength	3.0 T
Number of frames	19–25
Repetition time	4.15–5.26 ms
Echo time	12.24–12.91 ms
Image size	$(272, 192) - (288, 256)$ pixels
Field of view	$(190, 190) - (210, 210)$ mm

However, when the iteration step is large, the contours cannot always converge properly as shown in Fig. 4. Thus, we propose a grayscale regularizer based on pixel intensity level to ensure the contour propagation will properly stop when facing spinal cord boundary.

We notice that the contour performed by distance regularized level set evolution (DRLSE) in large iteration will first approximate the spinal cord boundary and then exceed, which demonstrates a clear double-peak feature in the image grayscale distribution when plotting the histogram as shown in Fig. 5. Thus, we find the trough of the first peak and

set a threshold based on the mode of the grayscale histogram, making it stop before it propagates into other regions.

We calculate the mask from the predicted contour C_φ and then compute the grayscale distribution $G : \mathbb{R}^2 \to \mathbb{R}$ of the input image I^F inside the RoI

$$G = \left\{ I^F[x, y] \,\middle|\, (x, y) \in \mathrm{Mask}(C_\varphi) \right\} \tag{17}$$

and the grayscale regularization term $\beta : \mathbb{R} \to \mathbb{R}^2$ is computed by

$$\beta(G) = \begin{cases} -1, & I^F[x, y] \geq \mathbf{Mo}(G) + \Theta \ \text{ and } \ (x, y) \in \mathrm{Mask}(C_\varphi) \\ 1, & \text{otherwise} \end{cases} \tag{18}$$

where the threshold $\Theta = 50$. The regularized LSF can be then formulated as

Table 3

Statistical comparison between the proposed methods and U-net with random rotation and horizontal flipping (indicated by †), U-net using patch-based training (indicated by ‡), Deepseg, and DRLSE.

Metric	U-net†	U-net‡	Deepseg	DRLSE
Dice coefficient (%)	0.0272	0.0040	< 0.001	0.0468
Hausdorff distance (mm)	0.0272	0.2474	0.0159	0.3393
Root-mean-square error (mm)	0.4294	0.4457	< 0.001	< 0.001

Table 2

Evaluation of automated segmentation over 12 patients' data with 207 MRI slices using the DC, HD, and RMSE. The results of the proposed method are compared with U-net with random rotation and horizontal flipping (indicated by †), U-net using patch-based training (indicated by ‡), Deepseg, and Distance Regularized Level Set Evolution (DRLSE) against expert manual ground truth produced by clinicians.

Metric	U-net†	U-net‡	Deepseg	DRLSE	Proposed
Dice coefficient (%)	83.7 ± 12.0	82.7 ± 11.9	71.2 ± 24.4	85.8 ± 18.0	**87.0 ± 18.4**
Hausdorff distance (mm)	**8.4 ± 8.5**	8.7 ± 9.6	14.9 ± 31.6	9.8 ± 7.9	9.7 ± 7.9
Root-mean-square error (mm)	6.6 ± 12.6	6.7 ± 14.7	13.3 ± 19.8	6.4 ± 7.3	**5.9 ± 7.3**

Table 4

Hyperparameter optimization with the size of morphological structuring element and alternative edge detection algorithm. The best value for each comparison is indicated with bold font.

Hyperparameter	$k = 3$	$k = 5$	$k = 7$	$k = 9$	Sobel
Dice coefficient (%)	86.1 ± 9.3	85.8 ± 9.5	$\mathbf{87.0 \pm 18.4}$	86.2 ± 9.4	86.2 ± 9.3
Hausdorff distance (mm)	12.1 ± 16.6	12.3 ± 16.5	$\mathbf{9.7 \pm 7.9}$	11.5 ± 16.6	11.8 ± 16.5
Root-mean-square error (mm)	10.3 ± 26.0	10.7 ± 25.9	$\mathbf{5.9 \pm 7.3}$	9.9 ± 25.8	10.3 ± 25.8
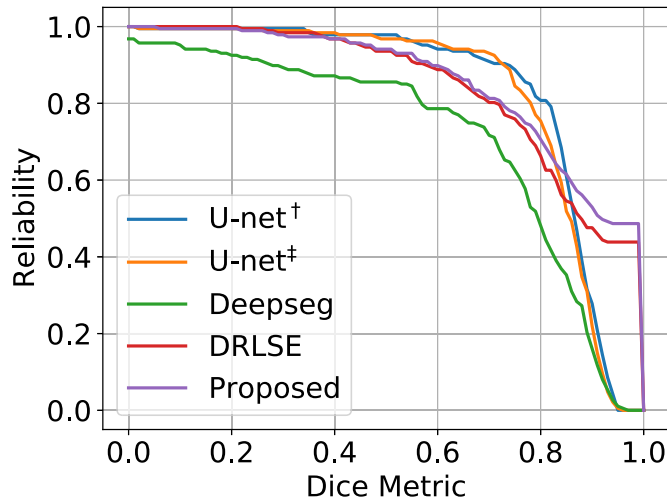
Fig. 6. The reliability metric versus Dice metric is calculated for U-net with random rotation (indicated by †), U-net with patch-based training (indicated by ‡), DRLSE, and the proposed method. The evaluations are performed over 207 MRI slices acquired from 12 CSM patients. Lines closer to the upper right corner in the plot are considered more ideal and robust according to the definition of the reliability metric.

$$\varphi = \beta(G) \odot \varphi^i, \tag{19}$$

where $\odot$ is the element-wise multiplication; and **Mo** is the mode of the image grayscale histogram. The regularization is updated iteratively until reaching the maximum iteration. The proposed GRLSE could be summarized in Alg. 1.

Algorithm 1. Grayscale regularized level set evolution (GRLSE) algorithm

3. Experiments

3.1. Clinical data

The clinical data is collected and provided by the Peking University Third Hospital. The details of the dataset used in the proposed study are listed in Table 1. The MRI scans were performed on a 3.0 T scanner (Discovery MRI750ws, GE Healthcare, Chicago, Illinois, United States) and the transversal T2-weighted imaging was performed using 2D Multiple Echo Recombined Gradient Echo (MERGE) sequence ant the disc level from C3 to C7 of each patient. The repetition and echo times were set to the range from 4.15 ms to 5.26 ms, and 12.24 ms–12.91 ms, respectively. The data is selected based on a number of excluding conditions, including:

- patients who have taken multiple cervical surgeries;
- patients with a history of cervical spine injury leading to spinal cord dysfunction;
- patients with cervical kyphosis;
- patients with spinal cord dysfunction due to iatrogenic factors, including surgical trauma or hematoma in the perioperative period;
- patients with C5 nerve root palsy or other cervical nerve root damage, causing irreversible damage to nerve function in the perioperative period.

The ground truth annotations for the 2D MRI slices used in this article are segmented manually and then corrected by an expert radiologist with 152 images from 8 CSM patients for training and 207 images from 12 CSM patients for testing. The neural network in the proposed study was trained using 114 images acquired from 6 subjects and 38 images acquired from 2 additional subjects were used as the validation set. The training and validation split is 0.75 for the U-net training.

3.2. Experiment setup

The U-net based model is implemented in Python programming

Algorithm 1: Grayscale regularized level set evolution (GRLSE) algorithm

1   **Input:** Input image I^F and initial area by U-net ϕ^0

2   Calculate edge indicator function: $g = 1/(1 + ||\nabla(G_\sigma * I^F)||_2^2)$

3   **For** $i = 1, 2, ..., M$ **do:**

4   **For** $j = 1, 2, ..., N$ **do:**

5        $\phi_j^i = \text{NeumannBoundCond}(\phi^{i-1})$

6   Calculate curvature: $\kappa_j^i = \text{div}(\nabla\phi_j^i/(||\phi_j^i||_2 + \epsilon_1))$

7   Calculate distance regularization term: $R_j^i = \text{div}(\nabla\phi_j^i d_p(||\phi_j^i||_2))$

8   Calculate edge term: $E_i^j = (\nabla\phi_j^i/(||\phi_j^i||_2 + \epsilon_1) \cdot \nabla g + g\kappa_j^i)\delta_{\epsilon_2}(\phi_j^i)$

9   Calculate area term: $A_i^j = g\delta_{\epsilon_2}(\phi_j^i)$

10   Update LSF: $\phi_{j+1}^i = \phi_j^i + \Delta t(\mu A_i^j + \lambda E_j^i + \alpha A_j^i)$

11      $\phi^i = \phi_{N+1}^i$

12   Calculate grayscale distribution: $G^i = \{I^F[x, y] | (x, y) \in \text{Mask}(C_{\phi^i})\}$

13   Calculate grayscale regularizer: $\phi^i = \beta^i(G^i) \odot \phi^i$

14   **Output:** Grayscale regularized LSF: ϕ^M

Fig. 7. Quantitative evaluation of spinal cord segmentation on a test set using 207 MRI slices from 12 CSM patients. Box plots for DC, HD, and RMSE for the proposed method are presented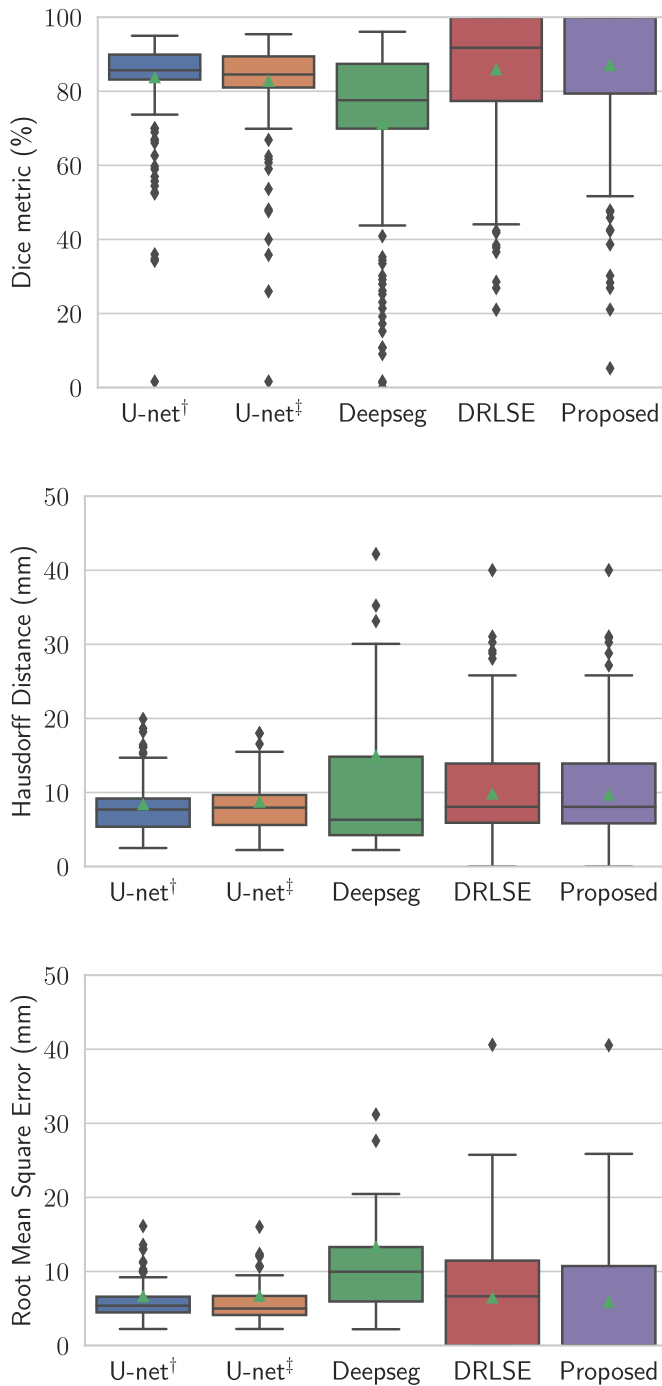 and compared with U-net with with random rotation and horizontal flipping (indicated by †), U-net with patch-based training (indicated by ‡), Deepseg and DRLSE. Boxes lower and upper edges are 25th and 75 the percentile respectively and the mean values are shown in the green triangles. Whiskers show the extreme values and data points beyond the whiskers are displayed using diamonds.

language using Keras with Tensorflow backend and NVIDIA CUDA Deep Neural Network library (cuDNN). We utilize a linear variable learning rate with Adam optimizer with $1e^{-4}$ in initial and the decay is uniform per epoch. The total number of epochs is 100. Inner and outer iterations of the level set methods are set as 5 and 25, respectively. We set $\Delta t = 1$, $\mu = 0.2, \lambda = 4, \alpha = -9, \varepsilon_1 = 1e^{-10}, \varepsilon_2 = 2$. The structuring parameters in closing operation are embedded in the OpenCV library. We also run

several simulations to optimize the size of the structuring element and test Sobel filter as an alternative edge detection algorithm. We report the results in Table 4. The results reported in the table demonstrate that the proposed method yielded better conformance compared with other hyperparameter settings. The experiment is conducted on the laptop with Intel Core i7-7700 HQ CPU with 16 GB RAM and the neural net models were trained and tested with an NVIDIA GTX 1050 GPU with 4 GB memory. The pre-processing for a single MRI slice took less than 10 ms on our laptop CPU.

3.3. Evaluation metrics

The proposed algorithm is evaluated quantitatively using the Dice coefficient (DC), Hausdorff distance (HD), root-mean-square error (RMSE) and reliability metric. We define set A as the automatic prediction region generated by predicted contour C_φ and denote M as manual segmentation ground truth.

3.3.1. Dice coefficient
DC measures the overlap between two delineated regions [39]:

$$DC = \frac{2|A \cap M|}{|A| + |M|} \tag{20}$$

3.3.2. Hausdorff distance
HD measures maximum deviation between the automatic contour and manual segmentation ground truth in terms of Euclidean distance [36]:

$$d_H(A, M) = \max \left\{ \sup_{a \in A} \inf_{m \in M} (a, m), \sup_{m \in M} \inf_{a \in A} (a, m) \right\} \tag{21}$$

where sup denotes supremum and inf infimum.

3.3.3. Root mean square error
RMSE measures the root-mean-square surface distance between predicted and ground truth contours [2]:

$$RMSE = \sqrt{\frac{\sum_{x \in A} d^2(x, M) + \sum_{y \in M} d^2(y, A)}{|A| + |M|}}. \tag{22}$$

3.3.4. Reliability metric
The reliability function measures how reliable an algorithm is when yielding an accuracy d, which is evaluated using the reliability function [2].

$$R(d) = P_r(DC > d) = \frac{\text{Images with DC higher than } d}{\text{Total images}}. \tag{23}$$

4. Results

4.1. Quantitative assessment

The quantitative assessments using DC, HD, RMSE of the 207 images from 12 CSM patients are presented in Table 2. The best performances for our testing data in terms of DC and RMSE are obtained from using our proposed method. Compared to the state-of-the-art level set method DRLSE, our method improves the quantitative result by showing a higher score in all the metrics. Although DRLSE has already demonstrated accurate segmentation accuracy in terms of Dice metric, it requires manual initialization and is less robust in boundaries. We compare our method with U-net with random rotation and flipping augmentation and state-of-the-art method, DeepSeg, specifically designed for spinal cord segmentation. As shown in Table 2, our proposed method achieves higher conformance with ground truth in all quantitative metrics. The results are also statistically significant with *p*-

| U–net† | U–net‡ | Deepseg | DRLSE | Proposed | GT |
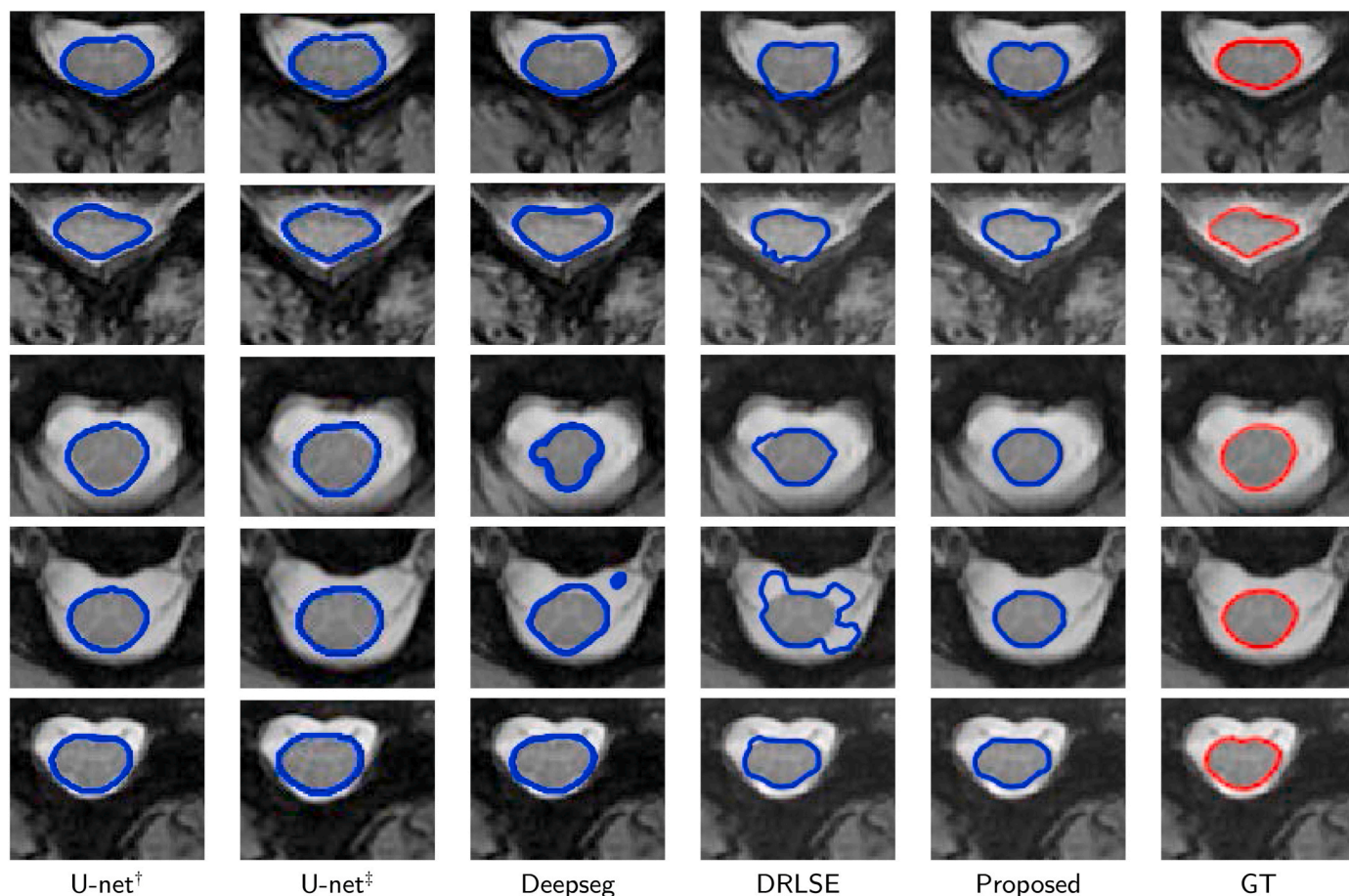
Fig. 8. Prediction examples of five frames from a patient in the testing set. The blue curves in the figure are predictions from U-net with random rotation and horizontal flipping (indicated by †), U-net with patch-based training (indicated by ‡), Deepseg, DRLSE, and the proposed method, respectively. The red curves are ground truth (GT) segmented by expert radiologists. Higher conformance was observed between expert manual segmentation and the results by the proposed fully automated approach.

Example 1   Example 2   Example 3
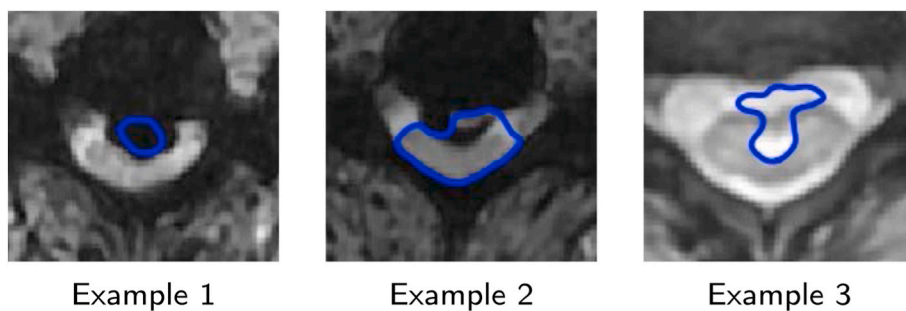
Fig. 9. Examples of the failed segmentation by our proposed method. The first example shows that the U-net fails to correctly localize the initial estimation; the second example shows that the grayscale intensity of the target spinal cord contour is similar to the background in the upper right corner and yields in the inaccurate segmentation; the third example shows that the intensity level inside the target spinal cord is distributed in a large difference and results in early evolution termination.

values calculated between state-of-the-art methods and the proposed method which are reported in Table 3. Similar performance is also observed in reliability metric shown in Fig. 6, boxplots shown in Fig. 7, and visual assessment shown in Fig. 8. Although the proposed method yields high conformance overall, U-net shows strengths in several aspects including higher conformance measured in terms of mean HD and achieved a smaller standard deviation for Dice score.

The reliability assessment is depicted in Fig. 6. According to its definition, the ideal reliability curve ought to be as close as possible to the right upper corner. Higher accuracy is observed from U-net segmentation in the Dice score range from $[0, 0.69]$ but our proposed method shows critical improvement in $(0.69, 1]$ compared with U-net. A similar improvement is also observed compared with DRLSE which

further confirms our statement in the quantitative assessment in terms of DC, HD, and RMSE.

The improvement of accuracy is also observed using box plots shown in Fig. 7. Box plots graphically display groups of numerical data through their quartiles. The highest median value in terms of DC and the lowest value in terms of RMSE are obtained from our proposed method.

Fig. 8 shows example automatic contours by our proposed method as well as U-net with data augmentation, U-net with patch-based training, Deepseg, and DRLSE in the blue curve. The data augmentation is performed in the baseline U-net model to increase the data size by 4 times and the augmentation techniques include random rotation of 90, 180, and 270° and horizontal flipping. The patch-based training was implemented by random cropping the image with size 128×128 patches.

Each image contains 4 random patches. The examples are also compared with the ground truth in the red curve. The U-net segmentation result in the example demonstrates accurate initial localization of the spinal cord but is lack of precise prediction on the boundaries compared with our proposed method. The result segmented by the DRLSE yields a higher tightness in the boundary compared with U-net. However, this alone requires manual initialization and lacks stable convergence on boundaries, where a proportion of the predicted contour extends to the nearby region with similar grayscale intensity. Our proposed method improves the result by adding a grayscale regularizer on the image intensity and demonstrates a higher convergence performance in boundaries. The integration of the U-net model with the level set approach also reduces the need of time-consuming manual initialization. The visual assessment further confirms the observation obtained in the quantitative assessment.

5. Discussion

In this study, we propose and validate a fully automated segmentation approach for 2D axial-view MRI sequences from CSM patients using convolutional neural network and an improved level set evolution. A large portion of previous studies in spinal cord segmentation are usually conducted in CT images [12,44], DTI images [16,24,26,40,46] instead of MRI. The related studies in MRI segmentation approaches are semi-automatic [9,10,19] instead of fully automatic, focusing on 3D spinal canal [17,18,23,28,30] instead of 2D spinal cord slices, sagittal view [22,38] instead of axial view, healthy patients [6,9,10,31] instead of CSM patients. To the best of the authors' knowledge, this is the first study that proposes a fully automated segmentation approach and thoroughly analyzed in terms of various evaluation metrics on the clinical 2D axial-view MRI data from CSM patients. Our focus is to develop an approach to help surgeons to localize the disease-causing segment of CSM. As designing the method with normal cases will not guarantee sufficient accuracy on pathological images, we devote our efforts to training and validating the method on pathological images directly.

Our method integrates the generalization ability of the U-net architecture with the accurate boundary prediction of the level set evolution. The deep learning-based approach usually needs large datasets to Ref. [48] achieve a high segmentation accuracy, which requires time-consuming manual ground truth segmentation and heavy computation hardware support. Level set evolution generally requires the manual initial approximation of contour location and orientation, and it is not robust enough in boundary convergence. As our proposed method introduces a grayscale regularizer in the active contour model, it achieves a better convergence performance on the contour boundary and is robust to different hyperparameter settings. Based on the literature [7, 10,22], spinal cord segmentation can be considered as accurate with the DC against manual segmentations higher than 0.85.

Although our proposed method yields an accurate prediction in various evaluation metrics, it still has several limitations. Firstly, the CNN model used for initial estimation is a supervised approach. Despite the reduction of the number of training samples, it still requires manual segmentation from clinicians. Secondly, our proposed method relies on the U-net model for contour initialization based on the clear assumption that U-net correctly localized all the initial points of the dataset. We notice that images with severely suppressed spinal cord resulted in lower Dice scores compared to other images. The diminished performance might due to the lack of accurate initialization of U-net as shown in Example 1 from Fig. 9. Thirdly, the grayscale regularizer is less effective when the grayscale intensity level of the target spinal cord is similar to the background or the intensity level inside the target spinal cord is distributed in a large difference shown in Example 2 and 3 in Fig. 9 respectively. In addition, the evaluations are performed on 207 images acquired from 12 patients and some of these images might have been similar. Future studies will be devoted to the segmentation and

evaluation of volumetric data. Although the proposed method yielded promising results when tested on 12 subjects, its generalization ability on larger datasets remains to be explored in the future. The applicability of the method in other scenarios such as multi-center context or different MR modalities also remains to be explored in the future.

6. Conclusion

This paper proposes a fully automated approach to segment the spinal cord using clinical 2D axial-view MRI slices from CSM patients. The proposed method introduces CNN to provide initialization for the improved level set-based method. The level set-based method introduces a grayscale regularizer to improve the convergence performance on contour boundaries. The proposed algorithm is evaluated with clinical data using both quantitative evaluation metrics and visual assessment. The quantitative assessment shows that our proposed framework outperforms the state-of-the-art algorithms in terms of DC, HD, RMSE and reliability metrics.

Declaration of competing interest

There is no conflict of interest of the authors in this work.

References

[1] Andrew J. Asman, Seth A. Smith, Daniel S. Reich, Bennett A. Landman, Robust GM/WM segmentation of the spinal cord with iterative non-local statistical fusion, in: Kensaku Mori, Ichiro Sakuma, Yoshinobu Sato, Christian Barillot, Nassir Navab (Eds.), International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2013, ISBN 978-3-642-40811-3, pp. 759–767, https://doi.org/10.1007/978-3-642-40811-395.

[2] Ismail Ben Ayed, Kumaradevan Punithakumar, Shuo Li, Ali Islam, Jaron Chong, Left ventricle segmentation via graph cut distribution matching, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2009, pp. 901–909.

[3] T. Behrens, K. Rohr, H.S. Stiehl, Robust segmentation of tubular structures in 3-D medical images by parametric object detection and tracking, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 33 (4) (August 2003) 554–561, https://doi.org/10.1109/TSMCB.2003.814305. ISSN 1083-4419, 1941-0492.

[4] Mark Bernhardt, Richard A. Hynes, Howard W. Blume, A.A. White 3rd, Cervical spondylotic myelopathy, J. Bone Joint Surg. 75 (1) (1993) 119–128.

[5] John Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. Intell. (6) (1986) 679–698.

[6] Min Chen, Aaron Carass, Jennifer Cuzzocreo, Pierre-Louis Bazin, Daniel S. Reich, Jerry L. Prince, Topology preserving automatic segmentation of the spinal cord in magnetic resonance images, in: 2011 IEEE International Symposium on Biomedical Imaging: from Nano to Macro, March 2011, pp. 1737–1740, https://doi.org/10.1109/ISBI.2011.5872741.

[7] Min Chen, Aaron Carass, Jiwon Oh, Govind Nair, Dzung L. Pham, Daniel S. Reich, et al., Automatic magnetic resonance spinal cord segmentation with topology constraints for variable fields of view, Neuroimage 83 (2013) 1051–1062.

[8] Esha Datta, Nico Papinutto, Regina Schlaeger, Alyssa Zhu, Julio Carballido-Gamio, Roland G. Henry, Gray matter segmentation of the spinal cord with active contours in mr images, Neuroimage 147 (2017) 788–799.

[9] Benjamin De Leener, Samuel Kadoury, Julien Cohen-Adad, Robust, accurate and fast automatic segmentation of the spinal cord, Neuroimage 98 (September 2014) 528–536, https://doi.org/10.1016/j.neuroimage.2014.04.051, 10538119.

[10] Benjamin De Leener, Julien Cohen-Adad, Samuel Kadoury, Automatic segmentation of the spinal cord and spinal canal coupled with vertebral labeling, IEEE Trans. Med. Imag. 34 (8) (August 2015) 1705–1718, https://doi.org/10.1109/TMI.2015.2437192. ISSN 0278-0062, 1558-254X.

[11] Benjamin De Leener, Manuel Taso, Julien Cohen-Adad, Virginie Callot, Segmentation of the human spinal cord, Magnetic Resonance Materials in Physics, Biology and Medicine 29 (2) (April 2016) 125–153, https://doi.org/10.1007/s10334-015-0507-2. ISSN 1352-8661.

[12] Dong Xue, Lei Yang, Tonghe Wang, Matthew Thomas, Leonardo Tang, Walter J. Curran, et al., Automatic multiorgan segmentation in thorax CT images using U-net-GAN, Med. Phys. 46 (5) (2019) 2157–2168, https://doi.org/10.1002/mp.13458. ISSN 2473-4209.

[13] Lawrence C. Evans, Partial differential equations. graduate studies in mathematics, American mathematical society 2 (1998), 1998.

[14] Mhamad Faour, Joshua T. Anderson, Arnold R. Haas, Rick Percy, Stephen T. Woods, Uri M Ahn, et al., Surgical and functional outcomes after multilevel cervical fusion for degenerative disc disease compared with fusion for radiculopathy: a study of workers compensation population, Spine 42 (9) (2017) 700–706.

[15] Michael G. Fehlings, K Daniel Riew Lindsay A Tetreault, James W. Middleton, Bizhan Aarabi, Paul M. Arnold, Darrel S. Brodke, Anthony S. Burns, Carette Simon, Robert Chen, et al., A clinical practice guideline for the management of patients with degenerative cervical myelopathy: recommendations for patients with mild, moderate, and severe disease and nonmyelopathic patients with evidence of cord compression, Global Spine J. 7 (3_suppl) (2017) 70S–83S.

[16] Raisa Z. Freidlin, Evren Ozarslan, Michal E. Komlosh, Lin-Ching Chang, Cheng Guan Koay, Derek K. Jones, et al., Parsimonious model selection for Tissue segmentation and classification applications: a study using simulated and experimental DTI data, IEEE Trans. Med. Imag. 26 (11) (November 2007) 1576–1584, https://doi.org/10.1109/TMI.2007.907294. ISSN 0278-0062, 1558-254X.

[17] Guanghua Fu, Huimin Lu, Joo Kooi Tan, Hyoungseop Kim, Xinglong Zhu, Jinhua Lu, Segmentation of spinal canal region in CT images using 3D region growing technique, in: 2018 International Conference on Information and Communication Technology Robotics, September 2018, pp. 1–4, https://doi.org/10.1109/ICT-ROBOT.2018.8549913.

[18] Charley Gros, Benjamin De Leener, Atef Badji, Josefina Maranzano, Dominique Eden, M. Sara, Dupont, et al., Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks, Neuroimage 184 (January 2019) 901–915, https://doi.org/10.1016/j.neuroimage.2018.09.081. ISSN 1053-8119.

[19] Mark A. Horsfield, Stefania Sala, Mohit Neema, Martina Absinta, Anshika Bakshi, Maria Pia Sormani, et al., Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: application in multiple sclerosis, Neuroimage 50 (2) (April 2010) 446–455, https://doi.org/10.1016/j.neuroimage.2009.12.121. ISSN 1053-8119.

[20] Jakubicek Roman, Jiri Chmelik, Petr Ourednicek, Jiri Jan, Deep-learning-based fully automatic spine centerline detection in CT data, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, July 2019, pp. 2407–2410, https://doi.org/10.1109/EMBC.2019.8856528.

[21] Fumihiko Kato, Yasutsugu Yukawa, Kota Suda, Masatsune Yamagata, Takayoshi Ueta, Normal morphology, age-related changes and abnormal findings of the cervical spine. part ii: magnetic resonance imaging of over 1,200 asymptomatic subjects, Eur. Spine J. 21 (8) (2012) 1499–1507.

[22] Jaehan Koh, Taehyong Kim, Vipin Chaudhary, Gurmeet Dhillon, Automatic segmentation of the spinal cord and the dural sac in lumbar mr images using gradient vector flow field, in: Annual International Conference of the IEEE Engineering in Medicine and Biology, IEEE, 2010, pp. 3117–3120, 2010.

[23] Jaehan Koh, Peter D. Scott, Vipin Chaudhary, Gurmeet Dhillon, An automatic segmentation method of the spinal canal from clinical MR images based on an attention model and an active contour model, in: 2011 IEEE International Symposium on Biomedical Imaging: from Nano to Macro, March 2011, pp. 1467–1471, https://doi.org/10.1109/ISBI.2011.5872677.

[24] C. Lenglet, M. Rousson, R. Deriche, DTI segmentation by statistical surface evolution, IEEE Trans. Med. Imag. 25 (6) (June 2006) 685–700, https://doi.org/10.1109/TMI.2006.873299. ISSN 0278-0062, 1558-254X.

[25] Chunming Li, Chenyang Xu, Changfeng Gui, Martin D. Fox, Distance regularized level set evolution and its application to image segmentation, IEEE Trans. Image Process. 19 (12) (2010) 3243–3254.

[26] Li Xiang, Jiao-Long Cui, Kin-Cheung Mak, Keith Dip-Kei Luk, Yong Hu, Potential use of diffusion tensor imaging in level diagnosis of multilevel cervical spondylotic myelopathy, Spine 39 (10) (2014) E615–E622.

[27] Allan R. Martin, Benjamin De Leener, Julien Cohen-Adad, David W. Cadotte, Aria Nouri, Jefferson R. Wilson, et al., Can microstructural mri detect subclinical tissue injury in subjects with asymptomatic cervical spinal cord compression? a prospective cohort study, BMJ Open 8 (4) (2018), e019809.

[28] David Mayerich, John Keyser, Hardware accelerated segmentation of complex volumetric filament networks, IEEE Trans. Visual. Comput. Graph. 15 (4) (July 2009) 670–681, https://doi.org/10.1109/TVCG.2008.196. ISSN 1077-2626, 1941-0506, 2160-9306.

[29] D.B. McCoy, S.M. Dupont, C. Gros, J. Cohen-Adad, R.J. Huie, A. Ferguson, et al., Convolutional neural network–based automated segmentation of the spinal cord and contusion injury: deep learning biomarker correlates of motor impairment in acute spinal cord injury, Am. J. Neuroradiol. 40 (4) (April 2019) 737–744, https://doi.org/10.3174/ajnr.A6020. ISSN 0195-6108, 1936-959X.

[30] Chris McIntosh, Ghassan Hamarneh, Spinal crawlers: deformable organisms for spinal cord segmentation and analysis, in: Rasmus Larsen, Mads Nielsen, Jon Sporring (Eds.), International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Berlin, Heidelberg, 2006, pp. 808–815, https://doi.org/10.1007/1186656599, 978-3-540-44708-5.

[31] Dipti Prasad Mukherjee, Irene Cheng, Nilanjan Ray, Vivian Mushahwar, Marc Lebel, Anup Basu, Automatic segmentation of spinal cord MRI using symmetric boundary tracing, IEEE Trans. Inf. Technol. Biomed. 14 (5) (September 2010) 1275–1278, https://doi.org/10.1109/TITB.2010.2052060. ISSN 1089-7771, 1558-0032.

[32] W New Peter, Raymond A. Cripps, B Bonne Lee, Global maps of non-traumatic spinal cord injury epidemiology: towards a living data repository, Spinal Cord 52 (2) (2014) 97.

[33] Aria Nouri, Lindsay Tetreault, Anoushka Singh, Spyridon K. Karadimas, Michael G. Fehlings, Degenerative cervical myelopathy: epidemiology, genetics, and pathogenesis, Spine 40 (12) (2015) E675–E693.

[34] David A. Parry, Timothy Booth, Peter S. Roland, Advantages of magnetic resonance imaging over computed tomography in preoperative evaluation of pediatric cochlear implant candidates, Otol. Neurotol. 26 (5) (2005) 976–982.

[35] Ferran Prados, M. Jorge Cardoso, Marios C. Yiannakas, Luke R. Hoy, Elisa Tebaldi, Hugh Kearney, et al., Fully automated grey and white matter spinal cord segmentation, Sci. Rep. 6 (October 2016) 36151, https://doi.org/10.1038/srep36151. ISSN 2045-2322.

[36] R Tyrrell Rockafellar, Roger J-B Wets, Variational Analysis, vol. 317, Springer Science & Business Media, 2009.

[37] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[38] Sascha Seifert, Irina Wachter, Gottfried Schmelzle, RÜdiger Dillmann, A knowledge-based approach to soft tissue reconstruction of the cervical spine, IEEE Trans. Med. Imag. 28 (4) (April 2009) 494–507, https://doi.org/10.1109/TMI.2008.2004659. ISSN 0278-0062, 1558-254X.

[39] Thorvald Julius Sørensen, A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and its Application to Analyses of the Vegetation on Danish Commons, I kommission hos E. Munksgaard, 1948.

[40] Y. Suetomi, T. Kanchiku, S. Nishijima, Y. Imajo, H. Suzuki, Y. Yoshida, et al., Application of diffusion tensor imaging for the diagnosis of segmental level of dysfunction in cervical spondylotic myelopathy, Spinal Cord 54 (5) (2016) 390.

[41] Lindsay Tetreault, Christina L. Goldstein, Paul Arnold, James Harrop, Alan Hilibrand, Aria Nouri, et al., Degenerative cervical myelopathy: a spectrum of related disorders affecting the aging spine, Neurosurgery 77 (suppl_1) (2015). S51–S67.

[42] A Tracy Jennifer, J.D. Bartleson, Cervical spondylotic myelopathy, Neurol. 16 (3) (2010) 176–187.

[43] Szabolcs Urbán, Attila Tanács, Atlas-based global and local RF segmentation of head and neck organs on multimodal MRI images, in: Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis, September 2017, pp. 99–103, https://doi.org/10.1109/ISPA.2017.8073577.

[44] Malinda Vania, Dawit Mureja, Deukhee Lee, Automatic spine segmentation from CT images using convolutional neural network via redundant generation of class labels, Journal of Computational Design and Engineering 6 (2) (April 2019) 224–232, https://doi.org/10.1016/j.jcde.2018.05.002. ISSN 2288-4300.

[45] Veeravagu Anand, Tyler Cole, Bowen Jiang, John K. Ratliff, Revision rates and complication incidence in single-and multilevel anterior cervical discectomy and fusion procedures: an administrative database study, Spine J. 14 (7) (2014) 1125–1131.

[46] Zhizhou Wang, B.C. Vemuri, DTI segmentation using an information theoretic tensor dissimilarity measure, IEEE Trans. Med. Imag. 24 (10) (October 2005) 1267–1277, https://doi.org/10.1109/TMI.2005.854516. ISSN 0278-0062, 1558-254X.

[47] Zian Wang, David Acuna, Huan Ling, Amlan Kar, Sanja Fidler, Object instance annotation with deep extreme level set evolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 7500–7508.

[48] Xiaoran Zhang, David Glynn Martin, Michelle Noga, Kumaradevan Punithakumar, Fully automated left atrial segmentation from mr image sequences using deep convolutional neural network and unscented kalman filter, in: IEEE International Conference on Bioinformatics and Biomedicine, IEEE, 2018, pp. 2316–2323, 2018.