MRI-SegFlow: a novel unsupervised deep learning pipeline enabling accurate vertebral segmentation of MRI images

Xihe Kuang, Jason PY Cheung, Honghan Wu, Socrates Dokos, *Member, IEEE*, Teng Zhang, *Member, IEEE*

Abstract— **Most deep learning based vertebral segmentation methods require laborious manual labelling tasks. We aim to establish an unsupervised deep learning pipeline for vertebral segmentation of MR images. We integrate the sub-optimal segmentation results produced by a rule-based method with a unique voting mechanism to provide supervision in the training process for the deep learning model. Preliminary validation shows a high segmentation accuracy achieved by our method without relying on any manual labelling.**

The clinical relevance of this study is that it provides an efficient vertebral segmentation method with high accuracy. Potential applications are in automated pathology detection and vertebral 3D reconstructions for biomechanical simulations and 3D printing, facilitating clinical decision making, surgical planning and tissue engineering.

I. Introduction

Magnetic Resonance Imaging (MRI) is widely used in the diagnosis of intervertebral disc degeneration. Vertebral segmentation is an essential part of many lumbar MRI automated analysis tasks, such as pathology detection and vertebral 3D reconstruction for further mechanical simulations or 3D printing [2].

Previous vertebral segmentation approaches can be divided into two categories: unsupervised and supervised. Most of the unsupervised segmentation methods are based on geometrical characteristics of the vertebrae [3-8], with specific templates, detectors or rules designed to identify and segment these. However, due to the complexity of the MR images, the accuracy and robustness of these unsupervised methods are low. Thus they cannot be used for clinical practice. Supervised methods typically train a segmentation model using manually masked vertebrae, and apply it on new images [9-11]. Using a convolutional neural network (CNN) [12], supervised methods can achieve high segmentation accuracy [10-11]. However, the laborious manual segmentation task is still required for the training process of supervised algorithms.

Original spine MR images contain more than required information for vertebral segmentation, which may increase the complexity of the learning task and reduce the accuracy of the output. Therefore, we proposed a regional CNN strategy. The training and predicting process of the network are restrained to a specific vertebral region to filter the irrelevant context. We hypothesize that a voting mechanism integrating a set of sub-optimal segmentation results can provide satisfactory supervision for the training of a deep learning network. Thus, manual labels will not be required.

In this paper, we aim to establish an unsupervised deep learning pipeline, MRI-SegFlow, for accurate vertebral segmentation. Objectives include, 1) selecting the region of interest (ROI) of each slice via a rule-based method; 2) integrating the ROIs via a voting mechanism to identify a volume of interest (VOI); 3) training a CNN model to segment vertebrae within the identified VOI.
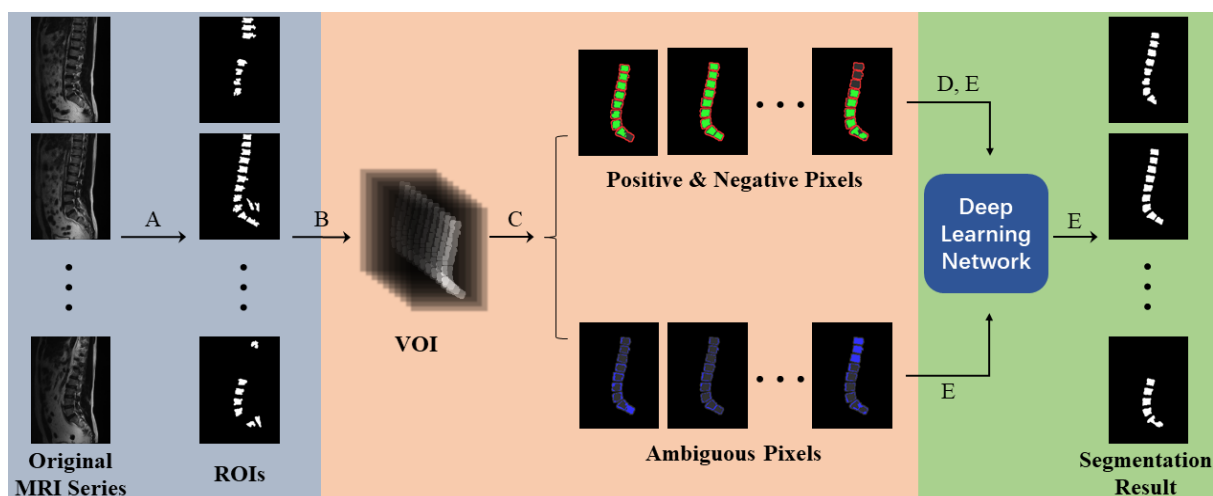

Figure 1. The MRI-SegFlow pipeline. A is the rule-based ROI detection component, B and C represent the VOI generation and triaging, followed by D and E for the network training and vertebral segmentation components.

Dr Teng Zhang, Mr Xihe Kuang, Dr Honghan Wu and A/Prof Jason Cheung are with the Department of Orthopedics and Traumatology, Faculty of Medicine, the University of Hong Kong (tgzhang@hku.hk, tony.kuangxh@hotmail.com, honghan.wu@gmail.com, cheungjp@hku.hk, contact Dr Zhang via phone: +852 39176989; fax: +852 28185210).

Socrates Dokos, is with the Graduate School of Biomedical Engineering, the University of New South Wales, Sydney, Australia, (email: s.dokos@unsw.edu.au)

II. Material and methodology

A. Dataset and Pipeline

243 volunteers were recruited by open advertisement if they are above 18 years old without any cancers and previous spinal surgeries. Their lumbar MR images were collected and used for the testing of our pipeline. The resolution of these images was 448×448. There were more than 15 slices in each MRI series. The average vertebral area in our MR images was 1650 pixels (range 800-2200), the average vertebral height was 40 pixels (range 30-55), and the average vertebral perimeter was 140 pixels (range 100-200), indicating the diversified nature of the dataset.

Briefly, our proposed MRI-SegFlow consisted of 3 major components, 1) the ROI detection (Figure 1: blue), 2) the voting mechanism (Figure 1: red), and 3) the CNN (Figure 1: green). The ROI detection is a rule-based vertebral detection process used to identify the ROI in each MRI slice. During the voting mechanism, ROIs of all slices were integrated to generate the VOI of the series, and then the pixels in the VOI were classified as 1) the vertebrae (positive pixels), 2) likely to be the vertebrae (ambiguous pixels) and 3) not the vertebrae (negative pixels). In the last component of our CNN-based vertebral segmentation, the network was trained by positive pixels and negative pixels to segment all the vertebrae in a VOI.

In the following sections, the rule-based ROI detection, the voting mechanism, and the CNN-based vertebral segmentation process will be described in more detail.

B. Rule-based ROI Detection

The ROI of each MRI slice was defined as the vertebral regions identified by the rule-based detecting process (Figure 1A). The local normalization [13] was firstly adopted to eliminate the pixel intensity variation in different tissues and to highlight vertebral edges (Figure 2B). The vertebral edges were then further enhanced with a Sobel filter, and after local thresholding, we obtained a binary mask in which the vertebrae were separated (Figure 2C). ROIs were all connected components in the mask having an area and perimeter within the range of vertebrae (Figure 2D).
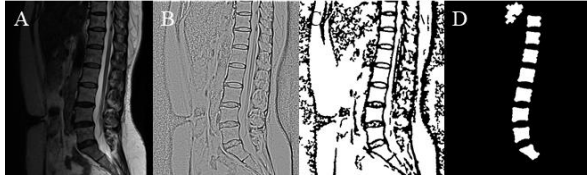

Figure 2. Rule-based ROI detection. A: original image, B: preprocessed image, C: edge enhanced and threshold image, D: ROI.

C. Voting Mechanism

The VOI of each MRI series consisted of all pixels that might belong to the vertebrae. Due to the consistent location of the vertebrae on each MRI slice, the VOI cross-section in different slices was the same. Due to spinal pathologies and the limited identification ability of the rule-based method, there were some errors in the ROIs, such as missing vertebrae and mistaken selection of non-vertebral regions (Figure 2D). Therefore, a novel voting mechanism was proposed to integrate the sub-optimal ROIs and identify the VOI of a MRI series.

In the following statements, the ROI of slice i and the VOI are denoted as $ROI_i(x,y)$ and $VOI(x,y)$. $ROI_i(x,y) = 1$, if the pixel at (x,y) is in the ROI of slice i; $ROI_i(x,y) = 0$, if it is not. A similar definition holds for $VOI(x,y)$.

The process of voting included:

Step 1: Assign weights to all pixels in ROIs and create a weight map for each MRI slice. Since the pixels at the edge of ROIs were more likely to be misjudged than those at the center, the central pixels were assigned more weight. The weight map of slice i, denoted as $W_i(x,y)$, is defined as:

$$W_i(x,y) = \begin{cases} w_c ROI_i(x,y), & d(x,y,B_1) \geq d_e \\ w_e ROI_i(x,y), & d(x,y,B_1) < d_e \end{cases} \quad (1)$$

where w_c and w_e represent the weight assigned to central and edge pixels, d_e is the cut-off for the separation of ROI center and edge, $B_1 = \{(x,y)|ROI_i(x,y) = 0\}$, and the $d(x,y,B_1)$ represents the distance from (x,y) to B_1, which is defined as:

$$d(x,y,B_1) = \min_{(a,b) \in B_1} \sqrt{(x-a)^2 + (y-b)^2} \quad (2)$$

Step 2: Calculate the voting result $V(x,y)$ as:

$$V(x,y) = \sum_i^N W_i(x,y) \quad (3)$$

where N represents the number of MRI slices.

Step 3: Define the vertebra central region R_{vc} in the VOI as:

$$R_{vc} = \{(x,y)|V(x,y) \geq T_h, d(x,y,B_2) \geq h_{ave}/8\} \quad (4)$$

where T_h is the higher voting cut-off, h_{ave} is the average vertebral height and $B_2 = \{(x,y)|V(x,y) < T_h\}$

Step 4: Define the potential vertebral region R_{pv} in the VOI as:

$$R_{pv} = \{(x,y)|V(x,y) \geq T_l\} \quad (5)$$

where T_l represents the lower voting cut-off.

Step 5: Check each connected component C_i in R_{pv}, and select it if $C_i \cap R_{vc} \neq \emptyset$. The determined vertebral region R_{dv} in the VOI consisted of all the selected connected components.

Step 6: In case some vertebral pixels were not identified in any ROI, the neighborhood of R_{dv} was also merged in the VOI. Therefore, the VOI is defined as (Figure 3A):

$$VOI(x,y) = \begin{cases} 1, & d(x,y,R_{dv}) \leq d_n \\ 0, & d(x,y,R_{dv}) > d_n \end{cases} \quad (6)$$

where d_n represents the maximum neighborhood distance.

Based on the voting result and the ROIs, the pixels in a VOI could be further classified into 3 categories: positive pixels, ambiguous pixels and negative pixels. These three categories for slice i (Figure 3B), denoted as Pos_i, Amb_i and Neg_i, are defined as follows:

$$Pos_i = \{(x,y)|ROI_i(x,y) = 1, V(x,y) \geq T_p\} \quad (7)$$

$$Neg_i = \{(x,y)|V(x,y) = 0\} \quad (8)$$

$$Amb_i = \{(x,y)|(x,y) \notin Pos_i \cup Neg_i\} \quad (9)$$

where the T_p represents the positive voting cut-off, and (x, y) satisfies $VOI(x, y) = 1$.

Figure 3. Voting process. A: VOI; B: Three pixel categories of VOI in one MRI slice: positive pixels (green), ambiguous pixels (blue), and negative pixels (red).

The positive pixels were identified in most ROIs and had a high probability of belonging to vertebrae, while the negative pixels were in the VOI but not included in any ROI.

D. Deep Learning Based Vertebrae Segmentation

A CNN was trained to further segment vertebrae from the VOI. Unlike the conventional network training process, in our pipeline no manual labels were involved. Instead, we relied on the positive and negative pixels in the VOIs to provide supervision (Figure 4).

The input of the CNN was a combination of the original MRI image, the local normalized image and the edge detection results. Since the local context was enough for segmentation, the network received a small patch of input images of size 24×24, one at a time, and could sufficiently determine whether the centre pixel of the patch belonged to the vertebrae. To improve the localization sensitivity of the model, coordinate channels [14] were introduced, which were two constant matrices of the same size as the input patches. The coordinate channels indicated the relative position of each pixel in the input patch, which could help the network to better understand the spatial distribution of features. Input to the network was processed by 3 successive convolution (Conv) blocks, each containing two convolutional layers with a kernel size of 3×3, and one maxpooling layer of kernel size 2×2. After each Conv block, the channels were doubled. Two fully connected layers followed the Conv blocks with 1024 hidden

units. All convolutional layers and fully-connected layers were activated via Rectified Linear Unit (ReLU), and the output layer was activated by softmax [12]. The output of the network was the probability that the center pixel of the input patch belonged to the vertebrae. After inferring all pixels in the VOI, the vertebral segmentation result was obtained.

The training process of the network could be divided into pre-training and finetuning. In the pre-training process, the same number of positive and negative pixels were selected randomly from all MRI series to train the network. Then the pre-trained model was fine-tuned on each MRI series to deal with individual differences. All positive and negative pixels from the MRI series were used in the fine-tuning. The optimized model was used to segment all vertebrae in the VOI of the MRI series.

III. IMPLEMENTATION AND PRELIMINARY VALIDATION

Sagittal lumbar MRIs obtained from the 234 patients were used for the validation of the proposed MRI-SegFlow pipeline, with the mid-9 slices selected for segmentation.

In the voting process, the weight of centre pixels w_c and edge pixels w_e were set to 2 and 1 respectively. The d_e and d_n were both set to 5. The T_h, T_l and T_p were set to 8, 1 and 4, respectively.

In the pre-training process, 100,000 positive pixels and 100,000 negative pixels were selected randomly from all MRI series. The pre-training took 5 epochs, and the fine-tuning generally took 1~3 epochs, depending on the different MRI images. Stochastic gradient descent (SGD) was used to optimize the model in both the pre-training and finetune stages.

Preliminary test results of the MRI-SegFlow pipeline (Figure 5B) were visually compared with the results of the rule-based method (Figure 5A). MRI-SegFlow achieved significantly higher segmentation accuracy by eliminating any missing vertebrae and image distortion.

IV. DISCUSSION

We have developed a novel unsupervised vertebral segmentation pipeline for sagittal lumbar MR images: MRI-SegFlow. It demonstrates high segmentation accuracy without
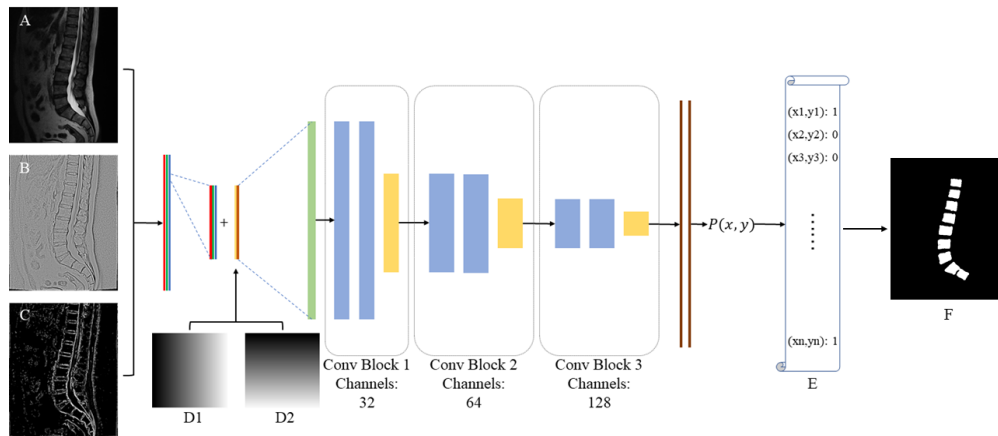

Figure 4. The pipeline of our deep learning-based vertebra segmentation. A, B and C represent the original MRI image, local normalized image and edge detection result, respectively, D is the coordinate channels, E presents the network output for all pixels in the VOI, which produces F, the vertebral segmentation result.

A: Rule-based vertebrae detection outcome with missing vertebrae and errrors

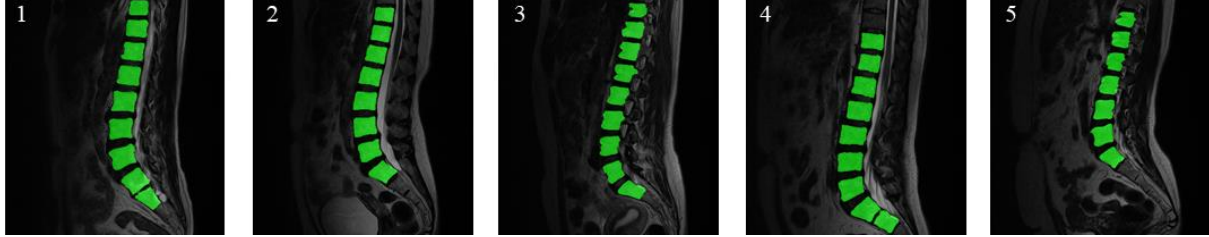B: Deep learning integrated outcome with accurate segmentation of vertebrae

Figure 5. Examples of vertebral segmentation results produced by the rule-based method (A), and the MRI-SegFlow pipeline (B)

relying on manual labels. We have adopted a regional CNN strategy in our pipeline. An ROI is determined first in each slice by a rule-based method and integrated through a unique voting mechanism to produce a VOI. The parameters in the rule-based ROI detection is empirically selected, they may need to be fine-tuned on another dataset. Since the ROI can be suboptimal, the fine-tuning process should be simple. The pixels within the VOI are further divided into positive pixels, ambiguous pixels and negative pixels. The model is trained by the positive and negative pixels to segment all vertebrae. The training and segmenting process are both within the pre-defined VOI, with any information not necessary for the segmentation being filtered.

Compared with previous reported supervised pipeline segmentation approaches [9-12] our method has the advantage of eliminating laborious manual labelling, achieving comparable segmentation accuracy. In comparison with unsupervised methods [3-8], our method solves the issues of missing vertebrae and image distortion. These outcomes are due to our integrated CNN that is able to learn and detect vertebrae having a wide range of feature variations caused not only by underlying disease, but by inconsistencies in image quality since the MRIs were acquired by different machines and human operators.

Extensive experiments need to be done to further validate the performance of our method. We can also transfer our MRI-SegFlow to other medical imaging modalities, or even to segment other anatomical structures in MRIs.

REFERENCES

[1]   C. W. Pfirrmann, A. Metzdorf, M. Zanetti, J. Hodler, and N. Boos, "Magnetic resonance classification of lumbar intervertebral disc degeneration," *Spine,* vol. 26, no. 17, pp. 1873-1878, 2001.

[2]   J. S. Ramos, M. T. Cazzolato, B. S. Faiçal, M. H. Nogueira-Barbosa, C. Traina, and A. J. Traina, "3DBGrowth: volumetric vertebrae segmentation and reconstruction in magnetic resonance imaging," in *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, 2019, pp. 435-440: IEEE.

[3]   S. Benameur, M. Mignotte, S. Parent, H. Labelle, W. Skalli, and J. de Guise, "3D/2D registration and segmentation of scoliotic vertebrae using statistical models," *Computerized Medical Imaging Graphics,* vol. 27, no. 5, pp. 321-337, 2003.

[4]   Z. Peng, J. Zhong, W. Wee, and J.-h. Lee, "Automated vertebra detection and segmentation from the whole spine MR images," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2006, pp. 2527-2530: IEEE.

[5]   M. S. Aslan, A. Ali, D. Chen, B. Arnold, A. A. Farag, and P. Xiang, "3D vertebrae segmentation using graph cuts with shape prior constraints," in *2010 IEEE International Conference on Image Processing*, 2010, pp. 2193-2196: IEEE.

[6]   P. H. Lim, U. Bagci, and L. Bai, "Introducing Willmore flow into level set segmentation of spinal vertebrae," *IEEE Transactions on Biomedical Engineering,* vol. 60, no. 1, pp. 115-122, 2012.

[7]   A. Rasoulian, R. Rohling, and P. Abolmaesumi, "Lumbar spine segmentation using a statistical multi-vertebrae anatomical shape+ pose model," *IEEE Transactions on Medical Imaging,* vol. 32, no. 10, pp. 1890-1900, 2013.

[8]   R. Korez, B. Ibragimov, B. Likar, F. Pernuš, and T. Vrtovec, "A framework for automated spine and vertebrae interpolation-based detection and model-based segmentation," *IEEE Transactions on Medical Imaging,* vol. 34, no. 8, pp. 1649-1662, 2015.

[9]   S. M. R. Al Arif, K. Knapp, and G. Slabaugh, "Fully automatic cervical vertebrae segmentation framework for X-ray images," *Computer methods programs in biomedicine,* vol. 157, pp. 95-111, 2018.

[10]   R. Janssens, G. Zeng, and G. Zheng, "Fully automatic segmentation of lumbar vertebrae from CT images using cascaded 3D fully convolutional networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 893-897: IEEE.

[11]   J.-T. Lu *et al.*, "Deepspine: Automated lumbar vertebral segmentation, disc-level designation, and spinal stenosis grading using deep learning," *arXiv preprint arXiv:.10215,* 2018.

[12]   A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[13]   M. Foracchia, E. Grisan, and A. Ruggeri, "Luminosity and contrast normalization in retinal images," *Medical image analysis,* vol. 9, no. 3, pp. 179-190, 2005.

[14]   R. Liu *et al.*, "An intriguing failing of convolutional neural networks and the coordconv solution," in *Advances in Neural Information Processing Systems*, 2018, pp. 9605-9616.