Journal of Digital Imaging (2020) 33:191–203
https://doi.org/10.1007/s10278-019-00216-0

A Region-Based Deep Level Set Formulation for Vertebral Bone Segmentation of Osteoporotic Fractures

Faisal Rehman[1] · Syed Irtiza Ali Shah[2] · M. Naveed Riaz[3] · S. Omer Gilani[1] · Faiza R.[4]

Published online: 22 April 2019

Abstract
Accurate segmentation of the vertebrae from medical images plays an important role in computer-aided diagnoses (CADs). It provides an initial and early diagnosis of various vertebral abnormalities to doctors and radiologists. Vertebrae segmentation is very important but difficult task in medical imaging due to low-contrast imaging and noise. It becomes more challenging when dealing with fractured (osteoporotic) cases. This work is dedicated to address the challenging problem of vertebra segmentation. In the past, various segmentation techniques of vertebrae have been proposed. Recently, deep learning techniques have been introduced in biomedical image processing for segmentation and characterization of several abnormalities. These techniques are becoming popular for segmentation purposes due to their robustness and accuracy. In this paper, we present a novel combination of traditional region-based level set with deep learning framework in order to predict shape of vertebral bones accurately; thus, it would be able to handle the fractured cases efficiently. We termed this novel Framework as "FU-Net" which is a powerful and practical framework to handle fractured vertebrae segmentation efficiently. The proposed method was successfully evaluated on two different challenging datasets: (1) 20 CT scans, 15 healthy cases, and 5 fractured cases provided at spine segmentation challenge CSI 2014; (2) 25 CT image data (both healthy and fractured cases) provided at spine segmentation challenge CSI 2016 or xVertSeg.v1 challenge. We have achieved promising results on our proposed technique especially on fractured cases. Dice score was found to be 96.4 ± 0.8% without fractured cases and 92.8 ± 1.9% with fractured cases in CSI 2014 dataset (lumber and thoracic). Similarly, dice score was 95.2 ± 1.9% on 15 CT dataset (with given ground truths) and 95.4 ± 2.1% on total 25 CT dataset for CSI 2016 datasets (with 10 annotated CT datasets). The proposed technique outperformed other state-of-the-art techniques and handled the fractured cases for the first time efficiently.

Keywords Vertebral osteoporotic fracture · Vertebrae segmentation · Computer-aided diagnosis · Medical image analysis · Deep learning

Introduction

The vertebral column has an amazing complex structure which performs numerous functionalities for the body. It provides stability, protects the spinal cord, and helps in overall body movement [1]. It extends from the skull to the pelvis and comprised of 33 individual bones. Normally human vertebral column has five curves. Cervical is comprised of two curves, one is upper cervical curve and other is lower cervical curve. Thoracic has one concave curve which extends from T2 to T12. Lumber has a single convex curve, and the last one is sacral curve. These curves provide balance, absorption, flexibility, and distribution to overall body [2]. The cervical spine (neck) comprises seven bones (C1–C7). Thoracic spine (chest) has 12 bones (T1–T12). Lumbar spine (low back) consists of 5 bones (L1–L5). Sacrum and coccyx comprises the fused part. Figure 1 presents the labeled diagram of vertebral column.

Various abnormalities are related to human vertebral column like vertebral scoliosis [3], lumbar canal stenosis, vertebra degeneration [4], and osteoporotic vertebral compression

✉ Faisal Rehman
faisalrehman0003@gmail.com

[1] Department of Robotics & Intelligent Machine Engineering, School of Mechanical & Manufacturing Engineering (SMME), National University of Sciences and Technology (NUST), Islamabad, Pakistan

[2] Department of Aerospace Engineering, National University of Sciences and Technology (NUST), Islamabad, Pakistan

[3] Department of Mechatranics Engineering, College of Electrical Engineering (CEME), National University of Sciences and Technology (NUST), Islamabad, Pakistan

[4] POF's Hospital, Wah Cantt, Pakistan

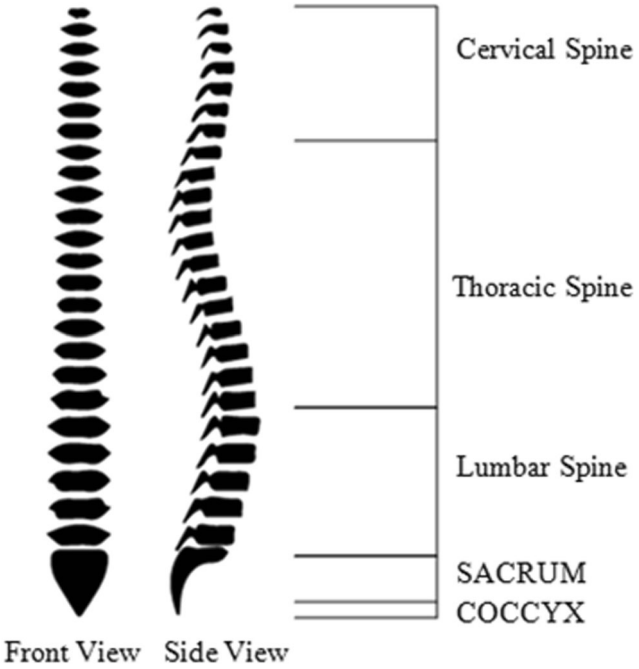

Fig. 1 Human vertebral column

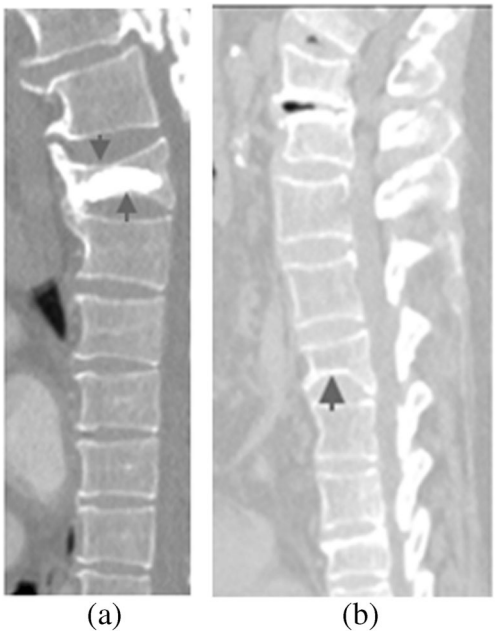

Fig. 2 Osteoporotic vertebral fractures are shown in **a** and **b** with arrows

fractures [5]. Out of these mentioned abnormalities, rapid increase in the ratio of osteoporotic vertebral fracture is an alarming situation worldwide. According to [6], annually 8.9 million vertebral fractures are caused by osteoporosis, which can be interpreted as, after every 3 s, a vertebral fracture is caused by osteoporosis in the world. Other statistics show that one out of three women and one out of five men above 50 years of age experience vertebral fracture in the world [7].

Osteoporosis is a condition in which bone mass is lowered and microlevel internal changes occurred in the bone tissues which make the bone fragile and may lead to fracture [5]. Vertebral osteoporosis is a condition of weakening and narrowing of vertebral bone due to osteoporosis, which is a cause of vertebral fracture. Figure 2 demonstrates the fracture of vertebral bone with arrows.

Vertebral fractures are the most common fracture which can be seen at early stage. Untreated vertebral fracture can lead to another fracture and may extend up to hip if remained untreated [8]. Figure 3 demonstrates percentage comparison of death cases with various osteoporotic fractures in men and women in 2010 on the data of around 30 different countries of the world [9]. All patients were 50 plus in age. It can be seen that hip fractures were caused more than vertebral fracture, but it should be noted that hip fracture is also an extension of uncured vertebral fracture [8].

With the help of the World Health Organization diagnostic criteria, future osteoporotic fractured cases were estimated by [9]. This estimation is drawn in Fig. 4, stratified by 5-year gaps till 2025. This is an increase rate of osteoporotic fracture between 60 and 64 years of age with the WHO statistics. This current situation is no doubt alarming and requires immediate measures to prevail this disease as it may lead to back pain, height loss, and various other complications [10].

In order to assist radiologists and clinicians, computer-aided diagnosis (CAD) provides an initial and important tool for various medical disorders like vertebral deformities. In clinical practice, CAD serves as an additional support for the doctors and radiologists [11]. It increases the efficiency of overall diagnostic system and saves the decision from medical error. The performance of CAD system is very important as it would directly affect the process of clinical decision making and the treatment.

CAD for vertebral disorders like osteoporotic fracture is indeed a challenging task due to complexity in its shape, low-contrast imaging, noise, and variation in field of views in radiological scans. Accurate diagnosis of osteoporotic vertebral fractures requires a strong and powerful vertebral segmentation algorithm to make the detection process efficient.

Various radiological scans are acquired like CT, MRI, and X-ray for the clinical diagnosis [11, 12]. These scan are performed in routine as a key element for the management of the patients subjected to various abnormalities. However, the patients subjected to vertebral disorders like osteoporotic vertebral fractures also need these radiological scans for their diagnosis. The selection of specific radiological scan is dependent on complexity and severity of patient's disease. After acquiring the appropriate radiological scan, it would be helpful for the radiologist to have better visualization of that scan or with some automatic abnormality detection. In this way, CAD is a helpful tool in radiological or clinical practice.

Previously, the problem of vertebrae segmentation was addressed with various traditional imaging techniques. Some of them are also reviewed in our literature review section.

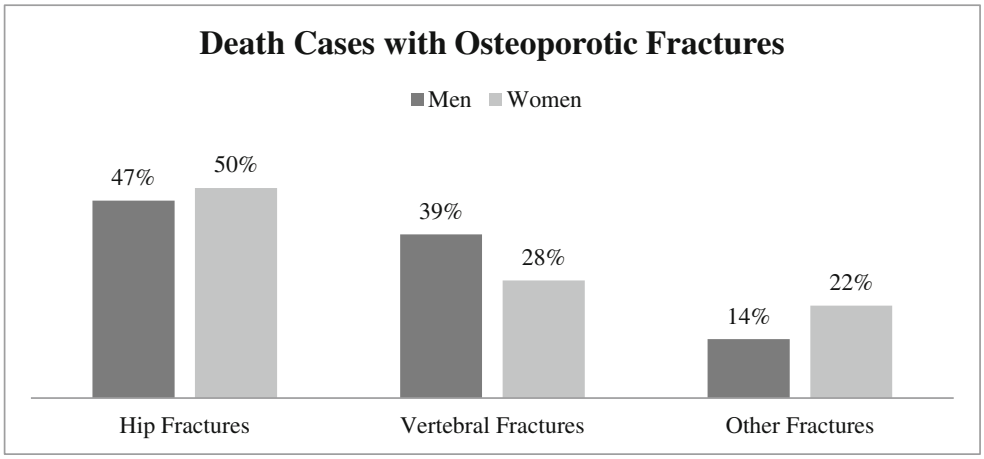

Fig. 3 Statistics of death cases with osteoporotic fractures in year 2010 on around 30 European countries data

Recently, deep learning–based techniques like convolutional neural networks and recurrent neural networks have proved state-of-the-art performance in many computer vision and medical image analysis tasks [11, 12, 26].

Convolutional neural network architectures like U-Net is a famous encoder-decoder network used for semantic segmentation [13]. It is an efficient framework for the subject segmentation but when dealing with large shape complexity, it has some limitations. U-Net performance may get worse in dealing with data of high topological shape variations. So, in such cases, we require an additional shape predictor to use with U-Net architecture in order to achieve appropriate segmentation results.

In this paper, the subject problem is addressed with the addition of region-based variational level set [14] in deep learning network. Level set–based methods [14] were previously used for various medical imaging and other segmentation tasks. These methods work on active contour models [15] and require an initialization to start the algorithm. Once the level set is initialized, the curve evolution will start and it will expend or shrink itself iteratively to approach the desired curve. Level sets can be further categorized in to two divisions. One is edge-based level sets, which use edge-based information to follow the aforementioned procedure and perform segmentation tasks. Other is region-based level sets (also called variational level sets), which use region-based approach instead of finding geometric boundaries to capture the geometric shapes accurately.

We have proposed a novel framework by unification of traditional region-based (variational) level set with a deep network. We termed the framework as "FU-Net." The combination of region-based level set with deep network provides us two-way advantage in one system. In this system, initially, the network is trained on the training dataset and then the pretrained network is given the input image to predict a probability output which is fed to the level set. The level set initializes with this probability output and iteratively updates the shape till desired. The improved output is then fed to the loss function which re-updates the weights and trains the network in a more efficient way. The final trained network is then

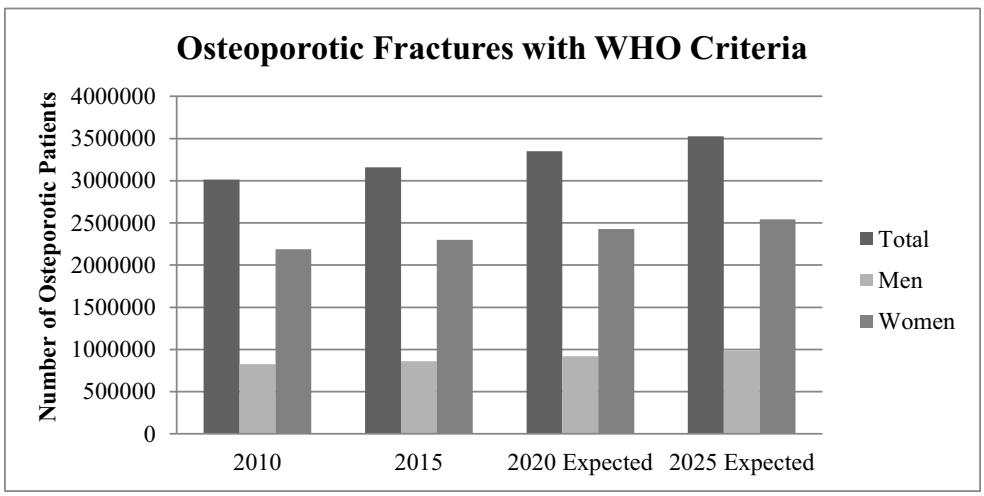

Fig. 4 WHO estimated statistics of osteoporotic fractured cases

available for testing. The main highlights of our system are as follows:

- Our work is a bridge between traditional level set method and current deep learned approach getting two-way benefits.
- We have formulated curve evolution using region-based level set under deep learned framework in order to improve the shape-based quality of fractured bone cases in vertebrae segmentation; thus, we called our network as FU-Net.
- Region-based deep level set provides high-level shape accuracy required in fractured cases which is not provided by simple U-Net architecture.
- With best of our knowledge, we are the first one to target vertebral fractured (osteoporotic) cases of CSI 2014 challenge with deep learning technique and achieved good results.
- Our algorithm is robust over multiple challenging datasets.
- Our overall results outperformed with other state of the arts, and our framework is baseline for various other medical imaging tasks.

The rest of the paper is organized as: In the "Literature Review" section, literature review is presented for both conventional and deep learning approaches used for vertebrae segmentation. The "Methodology" section presents our proposed methodology. The experiments and implementation details are given in the "Experiments" section. In the "Results" section, results are presented along with comparison benchmarks. Lastly, the "Discussions and Conclusions" section gives the conclusions and discussions on our work followed by the future directions.

Literature Review

In the past decades, various traditional segmentation techniques were proposed to address vertebra segmentation problem. For instance, Mahmoudi and Benjelloun et al. [16] performed X-ray vertebrae segmentation by applying conventional machine vision technique to find vertebral bone contours. These contours were used finally to get bone region segmentation. Klinder et al. [17] developed an automatic two-scale model-based approach for vertebrae segmentation. Global model used local vertebrae coordinate system to store vertebral bones shape and triangulated surface mesh was used to provide the shape information for local vertebral models. Fractured vertebrae segmentation was addressed by Roberts et al. [18] using active appearance model (AAM). The dataset used was computed radiographs of lumbar vertebrae of fractured cases taken from local hospital. The model was implemented in parts for different vertebral corners. The results achieved for fractured cases were satisfactory but not optimal. Benjelloun et al. [19] segmented cervical X-ray vertebrae using active shape model (ASM). A dual-model and dual-mode application is presented. The segmentation process was semiautomatic in which two points were manually created and used to initialize ASM search process for the subject segmentation. Mysling et al. [20] proposed an automatic framework to avoid manual initialization process of ASM for vertebral segmentation process.

In addition to these techniques, level set methods were also used by researchers for the subject segmentation, which are based on active contour model (ACM) [15]. For instance, Liu et al. [21] used level set–based method for segmentation of CT vertebrae images in sagittal, coronal, and axial domain. Hille et al. [22] used hybrid level set to perform segmentation of thoracic and lumbar vertebrae in MRI images. Rastgarpour et al. [23] combined local region-based level set with fuzzy cluster variation to address the solution of in-homogeneity in medical images for segmentation.

In contrast with conventional segmentation techniques which were primarily used for vertebra segmentation, deep learning techniques have achieved better performance and replaced these conventional techniques. For instance, Sekuboyina et al. [24] performed multiclass segmentation of lumber region using fully convolutional network (FCN). Initially, lumbar vertebrae localization is performed using multilayered perceptron, which performs nonlinear regression with global context. After localization, segmentation and labeling of lumbar vertebrae is performed using FCN. Later, Sekuboyina et al. [25] implemented 2D FCN for low-resolution localization of vertebrae and 3D FCN for high-resolution binary segmentations. Final segmentation map is obtained with the fusion of both networks. Janssens et al. [26] developed a framework for localization and segmentation of lumbar vertebrae using two consecutive fully convolutional neural networks. Lumbar region is initially localized with the bounding box, and then, lumbar vertebrae segmentation is carried out with the labeling of voxels. Lessmann et al. [27] combined instance memory with fully FCN and analyzed image patches in an iterative way to get vertebrae segmentation. Lessmann et al. [28] further improved his framework and evaluated performance on diverse and multimodal datasets. Lessmann et al. [28] used a memory component and perform instance segmentation to store information in the memory. The patches were iteratively analyzed by the network using that memory. In addition to segmentation, various other tasks like vertebral visibility prediction and vertebrae labeling were also performed.

The aforementioned methods discussed are relied on deep convolutional neural networks for vertebrae segmentation. However, many of these methods do not have specialty to handle fractured cases as osteoporotic vertebral fracture is a

very common abnormality in aged people. The basic difference of osteoporotic bone to a normal one is the narrowing of bone which can be handled with an addition of a shape parameterization term in the learning network of U-Net. For this purpose, Al Arif et al. [29] worked on segmentation of cervical X-ray images with an addition of a shape aware term in 2D convolutional neural network. This shape aware term is added in pixel-wise loss function during the network training. The term will take the Euclidean distance of predicted shape with the reference shape in order to achieve accuracy in segmented shape of cervical bones. However, this strategy simply finds the mean shape between predicted and the reference shape and unable to get high level of shape accuracy which is desired to get fractured vertebrae segmentation.

As discussed above, shape parameterization is missing in learning network to extract with high-level topological changes in shapes. For this purpose, level set–based approaches may be beneficial. Previously, these approaches were used for high-level shape extraction for various machine vision and medical imaging tasks [14, 21–23]. These techniques are based on active contour models [15]. One the difficulties in these techniques is the initialization, i.e., zero level set. Once they are initialized, the process of curve evolution starts. In this way, they iteratively approach the desired shape. The idea of combining classical level sets in deep learning network was studied and implemented by [30]. More specific for medical imaging, Ngo et al. [31] successfully solved the problem of left ventricle segmentation by combining classical level sets with deep learning network.

Methodology

We present a vertebral segmentation framework that is robust towards the complexity of shape, especially for vertebral osteoporosis. The method is based on the modified version of U-Net with optimized shape prediction. Our framework FU-Net is robust over different datasets and variations in vertebral shapes like in fractured cases. The proposed framework is presented in Fig. 5. In comparison with existing methods of literature, our proposed method captures the shape of vertebral bones precisely and efficiently.

Network Architecture

U-Net architecture is a "U"-shaped network comprised of contraction on the left side and expansion on the right side. Our segmentation network architecture is inspired by the U-net architecture [13] with a little bit modifications like we have kept uniform input and output dimensions (128 × 128) with the addition of padding in convolutional layers. On left side, i.e., contraction, our network has total nine convolutional layers and each follows a batch normalization and rectified linear unit layer (ReLU). After two consecutive convolutional layers, pooling layers of size 2 × 2 are used in order to reduce the dimension of image data on contraction side. The first pooling layer reduces dimensions to 64 × 64, second reduces to 32 × 32, and the third reduces to 16 × 16. On the right side, i.e., expansion, deconvolutional layers of size 2 × 2 are applied after each two consecutive convolutional layers to increase the image dimensions in an opposite way as applied in the contracting path. In order to avoid information loss in max pooling operation, after each expansion, concatenation is done between data in contracting path and data in expending path correspondingly. The overall system will take a single-channel input of 128 × 128 size vertebrae patch and vertebrae mask of same dimension is predicted as two channel output in probabilistic way. Figure 6 demonstrates the segmentation network.

DLS Formulation

In this section, we will present the traditional level set and then formulate under the deep network. As described before, variational or region-based level sets capture the shape of objects accurately instead of finding their geometric boundaries. For this purpose, Chan and Vese method [14] is a well-known traditional method used for image segmentation especially in medical imaging. Chan and Vese introduced energy functional as

$$E(m_1, m_2, \rho) = \mu \int_\gamma P(\rho)dxdy + \nu \int_\gamma Q(\rho)|\nabla(\rho)|dxdy + \int_\gamma \left(\varphi_1 |I - m_1|^2 P(\rho) + \varphi_2 |I - m_2|^2 (1 - P(\rho))\right) dxdy \quad (1)$$

where "I" is the input image, m_1 *and* m_2 are contours and ρ is the zero level set for initialization, and γ is the image domain $I_{x,y}$. $P(\omega)$ and $Q(\omega)$ are the functions of length and area, respectively, and μ, ν, φ_1, and φ_2 are positive parameters.

The minimization problem is given as

$$\min_{m_1, m_2, \rho} E(m_1, m_2, \rho) \quad (2)$$

Heaviside function and regularization are given in [14]. The expression for inside and outside contours is given as

$$m_1 = \frac{\int_\gamma I_{x,y} P(\rho_t)dxdy}{\int_\gamma P(\rho_t)dxdy} \,\&\, m_2 = \frac{\int_\gamma I_{x,y}(1 - P(\rho_t))dxdy}{\int_\gamma (1 - P(\rho_t))dxdy} \quad (3)$$

The gradient decent of ρ can be found with fixed m_1 and m_2 as

$$\frac{\partial \rho_t}{\partial t} = Q_\sigma \left(\rho \left[\nu\tau \left(\rho_t - \mu - \varphi_1 (I - m_1)^2 + \varphi_2 (I - m_2)^2\right]\right) \quad (4)$$

Here, curvature τ can be found by $\tau(\rho_t) = -div\left(\frac{\nabla \rho_t}{|\nabla \rho_t|}\right)$.

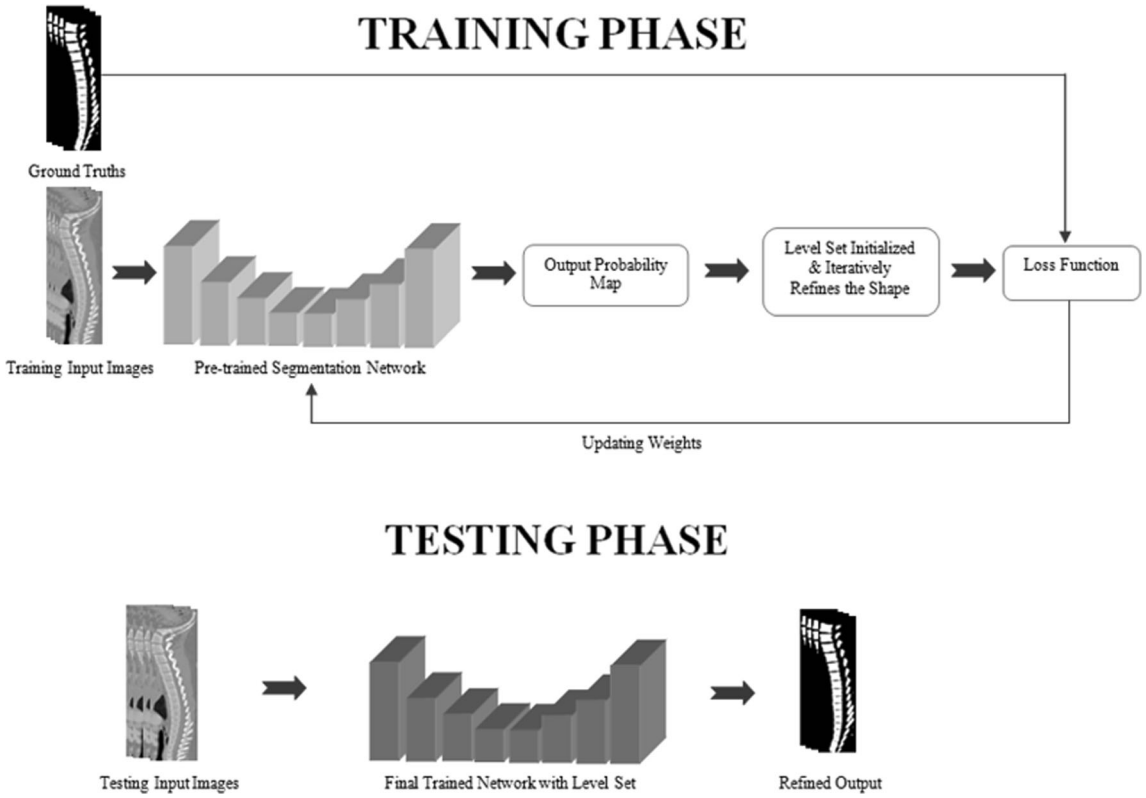

Fig. 5 System diagram of FU-Net

The curve will be updated will with the help of following expression:

$$\rho_{t+1} = \rho_t + \eta \frac{\rho_t}{\partial t} \tag{5}$$

In this way, the level set updates itself in time $t+1$ from the previous level set at time t and evolves the contour $\frac{\rho_t}{\partial t}$ with learning rate η.

Proposed Framework

Figure 5 shows the pipeline of our proposed framework for vertebrae segmentation. FU-net architecture presented above has been used as segmentation network. The network takes 128×128 size vertebrae patch as an input and predicts binary patch of same size as an output. Initially, the network is trained on the ground truths. Then, the pretrained network gives the input image to produce a probability output. The probability output of the learned network is then fed to level set, which further improves the shape and forward the improved output to the loss function for weight updating. In this way, segmentation is performed by deep network with use of an additional benefit of level set shape predictor.

Region-based level set gives big advantage when reforming in the deep framework. It helps in refining the shape of bones iteratively. In Eq. 1, "I" is the input image, ρ is the probabilistic predicted output of pretrained network given as zero level set. So, in the training phase, input images are fed to our pretrained network, the network predicts a probabilistic output. This output prediction is given again as input to level set for initialization as zero level set. Once level set is initialized, the curve evolution process will start in an iterative way following the Eqs. 4 and 5. The function ρ_t evolves the curve in time t using Eq. 4. Level set updates the shape in iterations using Eq. 5 till the desired level. At each time increment $t+1$, level set updates itself from previous level set at time t. Similarly, curve evolution $\frac{\rho_t}{\partial t}$ is done iteratively over time with a learning rate η. This process will continue till the desired output is achieved. The iterations were stopped on experimental basis and fixed for the particular dataset (average of 100). In this way, the network is trained again with level set. After this process, the output of level set is fed to loss function, where weights are updated in order to train the network properly.

The trained network is then available for testing. In order to test the trained network, vertebra patch is given at the input side. Segmented binary masks are available at the output side from the trained network.

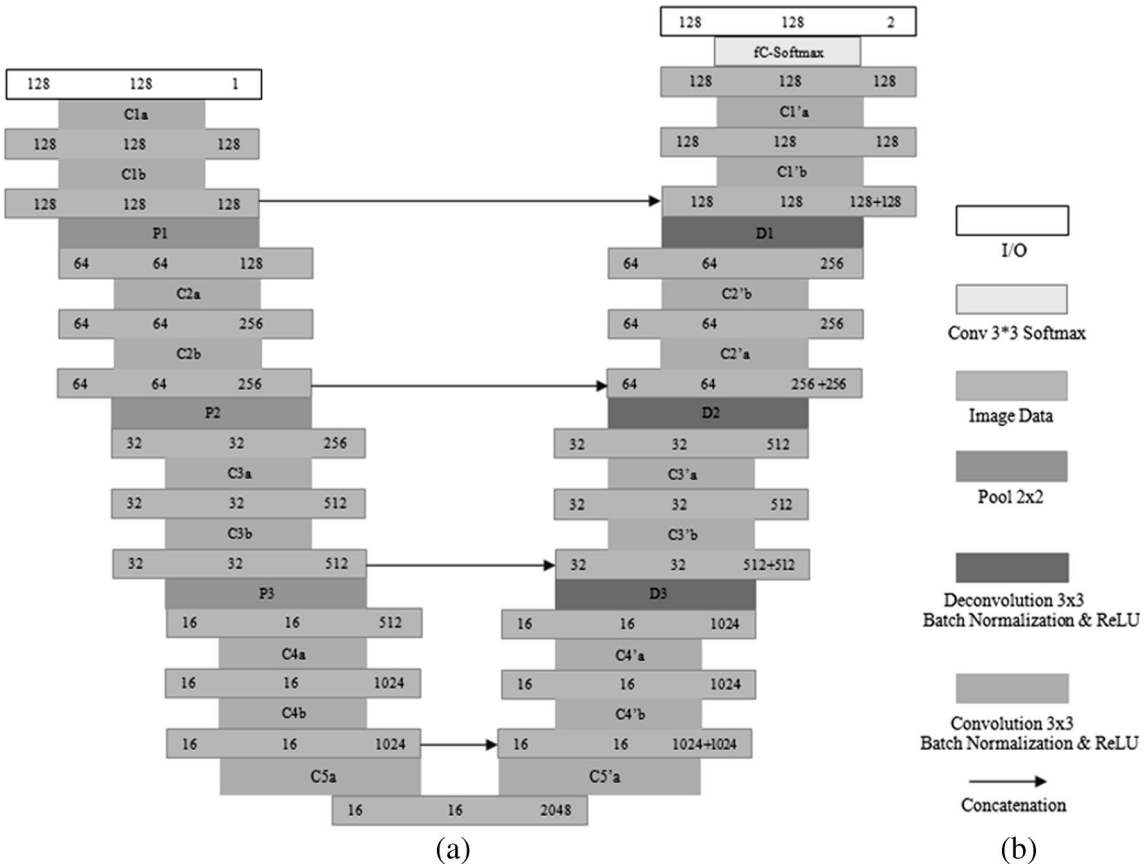

Fig. 6 U-Net architecture. **a** Segmentation network architecture. **b** Labels

Experiments

Datasets

We trained and tested our method on two publically available challenge datasets, CSI 2014 and xVertSeg.v1. The challenge organizers also provided the reference ground truths (except for ten cases of xVertSeg.v1), which helped us to compare our performance with researches on same datasets. Few images of these datasets are shown in Fig. 7.

The first dataset contains 20 CT scans, 15 healthy cases, and 5 fractured cases. This dataset has been provided at spine segmentation challenge (CSI 2014) Workshop on Computational Spine Imaging held in MICCAI, International Conference on Medical Image Computing and Computer Assisted Intervention 2014 (CSI 2014 challenge) [32]. The dataset was obtained from the medical center of University of California, Irvine (Orange, CA, USA). It consists of CT scans that visualize all thoracic and lumbar regions of vertebral column. The parameters of scans and reconstructions include 120 kVp, intravenous contrast and a slice thickness of 0.7 to 1.0 mm. The scans were taken with high spatial resolution as a single continuous CT data. The challenge organizers also provided reference segmentations, which were generated in a semi-automatic way with manual correction. The healthy data (15 CT scans) contain cases from young individuals (16–34 years). The fractured cases (5 CT osteoporotic cohort) were taken from aged persons (59–82 years). Each of them has at least one vertebral compression fracture in it.

The second dataset comprises 25 CT scans of lumbar vertebrae (both healthy and fractured cases) provided at spine segmentation challenge (CSI 2016) Workshop on Computational Spine Imaging held in (MICCAI) International Conference on Medical Image Computing and Computer Assisted Intervention 2016 (CSI 2016 or xVertSeg.v1 challenge) [33]. This dataset is also known as xVertSeg.v1 challenge. The slice thickness was between 1.0 and1.9 mm, and in-phase resolution of scan reconstruction was in between 0.29 and 0.80 mm. The data contain high variations in field of view, vertebral fractures, and deformities. For the first 15 cases, the challenge organizers provided referenced ground truths, but for last the 10 cases, reference ground truths were not available. Therefore, we annotated the contours with the help of interactive live wire tool [34]. The annotated contours were evaluated from two clinical experts (the details are mentioned in acknowledgements). Finally, binary masks were formed from the annotated contours, which were used for referenced segmentation.

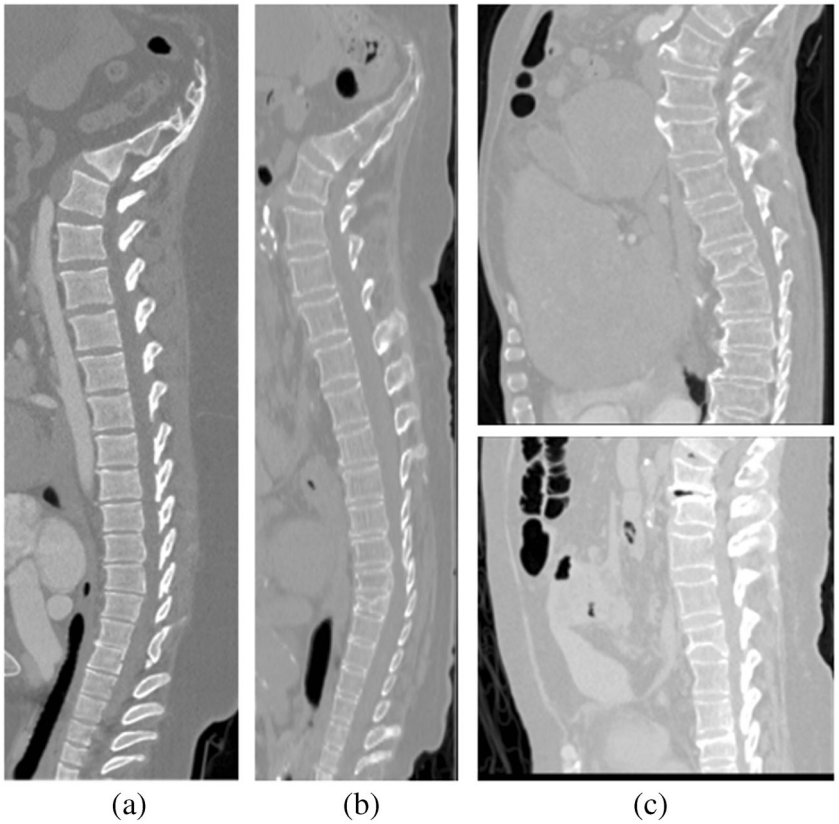

Fig. 7 Examples of datasets. **a** CSI 2014 Data Set. **b** CSI 2014 Dataset with fractured bone. **c** CSI 2016 Dataset with fractures

Evaluation Metrics

We have used dice score (DS) and mean absolute surface distance (ASD) in order to evaluate our performance and compare with state of the arts. These evaluation metrics are frequently used in recent works [24, 26, 27]. DS is used to measure the similarity to evaluate segmentation performance. It is calculated as [11]

$$\text{Dice score} = \frac{2 \mid O_p \cap G_t \mid}{\mid O_p \mid + \mid G_t \mid} \tag{6}$$

where $O_p$is the predicted output and G_t is the reference ground truth; it is calculated on individual vertebrae and then averaged over all scans. Similarly, mean ASD is average minimal distance calculated between two boundaries. It is calculated as [32]

$$ASD = \frac{1}{O_p} \sum_{i=1}^{O_p} \left|\left| \partial_i \left(O_p, O_g \right) \right|\right| \tag{7}$$

where O_p is the output segmentation surface, $O_g$is the reference ground truth surface, and ∂_i is the minimum distance between the two surfaces calculated on vertebrae in a similar fashion as described for DS.

Implementation Details

Figure 5 presents the system diagram, and Fig. 6 presents the deep network. Our proposed framework was implemented in Python using Tensorflow on windows desktop system Intel-(R) i-7 CPU and 1080 GTX graphics card with GPU memory 8 GB. The training time for CSI 2014 dataset was larger than xVertSeg.v1 datasets due to the coverage of large number of bones. The network was trained for 500 epochs, and early convergence was achieved successfully. Learning rate was used in decaying with momentum of 0.9. Dropout between two consecutive convolutional layers was used as 0.2. All experiments were carried out on 2D sagittal slices after augmentation, as data augmentation is necessary in order to enhance deep learning system network performance [13]. We augmented the datasets using various techniques like scaling, rotations explained in [35] to increase our training datasets above 5000 images. Finally, the arrangement of training and test sets were prepared with respect to state-of-the-art works in order to compare our segmentation performance with them. While dealing with the fractured case, we have prepared an arrangement for better learning of our system on fractured cases.

Our network, i.e., FU-Net, is a modified version of U-Net with level set combined with the network for better shape prediction for segmentation. This level set is initialized with the probability output of segmentation network, and it updates the shape of vertebral bones iteratively till desired

segmentation is achieved. The shape is updated iteratively using $\rho_{t+1} = \rho_t + \eta\frac{\rho_t}{\delta t}$ (Eq. 5). For testing the data from final trained network, the processing time for the subjects depend on total slices that varied from 30 to 150 s. The use of level set with deep learning has a great significance for segmentation of fractured cases, where the bones vary in shape. However, it also works well for normal healthy cases of vertebrae segmentation. In order to validate our proposed framework, we have compared our results with state-of-the-art techniques in the "Results" section.

Results

Segmentation Performance

MICCAI CSI 2014 CT dataset is comprised of thoracic and lumbar vertebrae. For the first 15 scans, it contains all healthy cases and last five contains fractured (osteoporotic) cases. Our algorithm smoothly handled normal cases. As far as fractured (osteoporotic) cases are concerned, our algorithm also performed better results. The last five fractures cases were not yet segmented with deep learning techniques with previous researchers. We are the first one to handle them with iterative level set approach combined with deep network. The DS for 15 CT cases was found to be 96.4 ± 0.8 and for all 20 cases (included fractured cases) was found to be 92.8 ± 1.9. The segmentation results are shown in Fig. 8.

The second dataset, i.e., CSI 2016 (xVerSeg.v1) dataset, is rich in field of views and variations in shape like fractures and other deformities. This data can be divided in to two categories. The first one contains 15 CT datasets with reference segmentations given, and the second one contains 10 CT datasets without reference segmentations. For the second part, we have done in-house procedure to get reference segmentations as described in dataset section. The DS for 15 CT datasets was found to be 95.2 ± 1.9 and dice score for total 25 CT datasets was 95.4 ± 2.1. The segmentation results show that DS increased with the increase in training data. The segmentation results of xVertSeg datasets are shown in Fig. 8.

In comparison with other vertebrae segmentation techniques, our proposed level set–based deep learning technique performed better in accuracy of segmentation results. Tables 1, 2, 3, 4, and 5 show the detailed comparisons of our results with the others.

Comparison of Other Methods on MICCAI CSI 2014 Challenge Datasets

As we have taken the use of level set to optimize the shape, our proposed technique achieved good DS. Table 1 shows the results of 15 CT datasets of CSI 2014 (healthy cases). The challenge winner [32] achieved DS $94.7 \pm 00\%$. Lessmann et al. [27, 28] used instance memory in deep network and adapted various schemes to get some improvements in DS. The results are presented in Table 1 which shows that combining level set in deep learning improves the accuracy in segmentation results.

Least work is done in fractured cases. To the best of our knowledge, we are the first to handle fractured cases with deep learning. In order to train the segmentation network properly for fractured cases, we have divided five fractured cases in to two categories, i.e., (2 cases + 3 cases). Similarly, healthy cases are also divided in two categories, i.e., (13 cases + 2 cases). Finally, the former healthy and fractured cases were combined and used for training, and later, healthy and fractured cases were combined and used for testing. Table 2 presents the results for all 20 cases (including fractured cases)

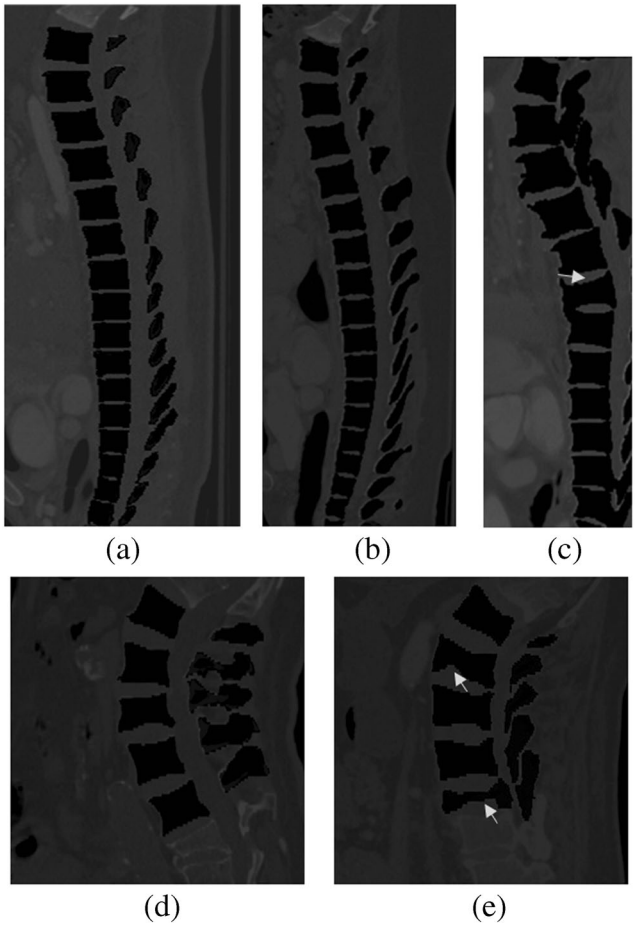

Fig. 8 Segmentation results of our proposed system (**a**, **b**). Results of CSI 2014 CT dataset. **c** Result from fractured case from CSI 2014 dataset (cropped for better visualization); arrow shows the fractured bone. **d** Result of xVertSeg.v1 dataset. **e** Result of fractured case of xVertSeg.v1 dataset with arrow, pointing the fractured bones

Table 1 Results with comparison of 10 training and 5 testing cases (healthy cases)

Dataset (MICCAI CSI 2014 Challenge) (10 healthy training cases and 5 healthy testing cases)	Dice score (%)	ASD (mm)
Challenge winner, J. Yao et al. (2016) [32]	94.7 ± 0.03	0.37 ± 00
Lessmann et al. (2018) [27]	94.8 ± 1.6	0.30 ± 0.1
Lessmann et al. (2018) [28]	96.3 ± 1.3	0.1 ± 0.1
Our proposed technique (FU-Net)	96.4 ± 0.8	0.1 ± 0.05

Comparison of Other Methods on MICCAI CSI 2016 Challenge Datasets

It is difficult to handle CSI 2016 dataset due to large variations in dataset, but the results were good. In order to compare our results with state of the arts, we have made three arrangements of dataset for training and testing; these arrangements are described and used in Tables 3, 4, and 5.

Table 3 shows the performance comparison of 15 healthy cases. In order to compare with Lessmann et al. [28], we have taken ten cases for training and five cases for testing purpose. The results show some improvement with use of variational level set in deep learning.

Janssens et al. [26] used the same dataset and performed leave three out cross-validation. In each time, 3 out of 15 cases were randomly chosen for testing and the remaining 12 were left for training, and this method was repeated for five times. In order to perform a rough comparison, we have randomly chosen three cases for testing in a similar fashion and the remaining 12 were used for the training purpose. The results are presented in Table 4.

As described in dataset section, reference ground truths of 10 out of 25 datasets were not given by the challenge organizers. These datasets are also considered for research with in-house ground truth formation scheme described before. The complete 25 scans were used as 15 for training and 10 for testing. The similar fashion was adapted by Sekuboyina et al. [24]. The results are presented in Table 5. A little bit increase in DS and decrease in surface density were recorded due to the increase of training dataset in our system.

Table 2 Results with comparison of 15 training and 5 testing (including fractured cases)

Dataset (MICCAI CSI 2014 Challenge) 15 healthy 5 unhealthy (fractured) cases	Dice score (%)	ASD (mm)
Challenge winner, J. Yao et al. (2016) [32]	89.7 ± 00	0.64 ± 00
Our proposed technique (FU-Net)	92.8 ± 1.9	0.41 ± 1.8

Table 3 Performance comparison on 15 cases (10 for training and 5 for testing)

Dataset (xVertSeg.v1 Challenge) 15 cases (10 for training and 5 for testing)	Dice score (%)	ASD (mm)
Lessmann et al. (2018) [28]	94.6 ± 2.2	0.3 ± 0.2
Our proposed technique (FU-Net)	95.1 ± 1.9	0.32 ± 0.18

Discussions and Conclusions

Discussions

This paper presents a novel combination of region-based level set technique with deep convolutional neural network. Region-based level set is based on active contour model which is used to extract region-based information (Vese and Chan et al. [14]). Region-based level set (Vese and Chan) is a very popular traditional method used for medical image segmentation. In similar way, convolutional neural networks are also used in recent years for medical image segmentation. We have taken the advantage of both techniques in a single framework.

Segmentation of vertebral fractures from images was neglected by recent researchers. Fractured cases require optimized shape prediction which is not provided by simple convolutional neural network. Level set–based convolutional neural network can handle these cases in a better way. Level set once initialized, performs iterations in a hidden way and improves the shape of bones. This process iteratively converge the shape of bones till the desired segmentation.

Figure 9 gives the box and whisker plot on the evaluation set of CSI 2014 dataset including fractured cases. The plot is drawn on DS per vertebrae at bone level. Table 2 gives the complete results. It is to be noted that the segmentation was very tough at upper thoracic and lower thoracic. This is because the data comprised of aged persons, and with aging, the bones of upper thoracic start merging and lose their original shape. Similarly, at lower thoracic, the plot shows the severe compression fractures at bones at T10 and T11. The results of lumbar vertebrae segmentation are more accurate than thoracic vertebrae.

Table 4 Performance comparison on 15 cases (12 for training and 3 for testing)

Dataset (xVertSeg.v1 Challenge) 15 cases (12 for training and 3 for testing)	Dice score (%)	ASD (mm)
Janssens et al. (2018) [26]	95.7 ± 0.8	0.37 ± 0.06
Our proposed technique (FU-Net)	95.2 ± 1.9	0.31 ± 0.18

Table 5 Performance comparison of 25 cases (15 for training and 10 for testing)

Dataset (xVertSeg.v1 Challenge) 25 cases (15 for training and 10 for testing)	Dice score (%)	ASD (mm)
Sekuboyina et al. (March 2017) [24]	94.3 ± 2.8	–
Our proposed technique (FU-Net)	95.4 ± 2.1	0.26 ± 0.23

As seen from Tables 3, 4, and 5 that the test results are improved with increase in training data. This shows that the learning of network is increased with increase in training data. This has been shown in Fig. 10.

One thing to be noted that Lessmann et al. [28] trained network on 10 CT data, Janssens et al. [26] trained network on 12 CT data, and Sekuboyina et al. [24] trained network on 15 CT data, but their dice scores were not increased smoothly because all of them were using difference techniques for vertebrae segmentation on same dataset.

We have trained and evaluated out framework on two different CT datasets. This helped us to check robustness of our proposed method over multiple datasets with variation in field of view, intensity, and resolutions. We have applied our method on both datasets without having any modification. The segmentation is performed smoothly without network overfitting problem

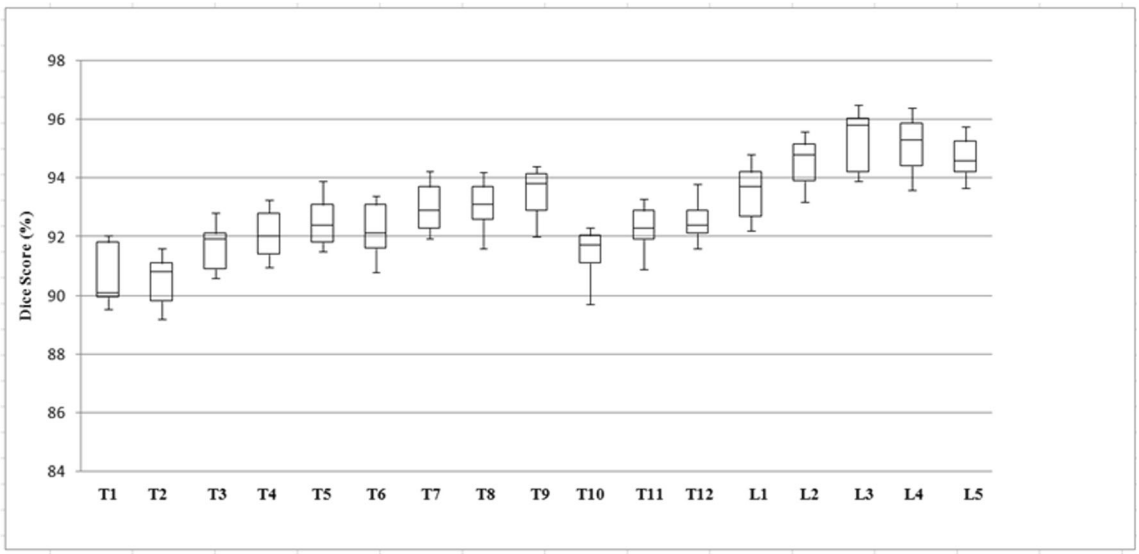

Fig. 9 Box and whisker plot on evaluation set of CSI 2014 dataset (including fractured cases)

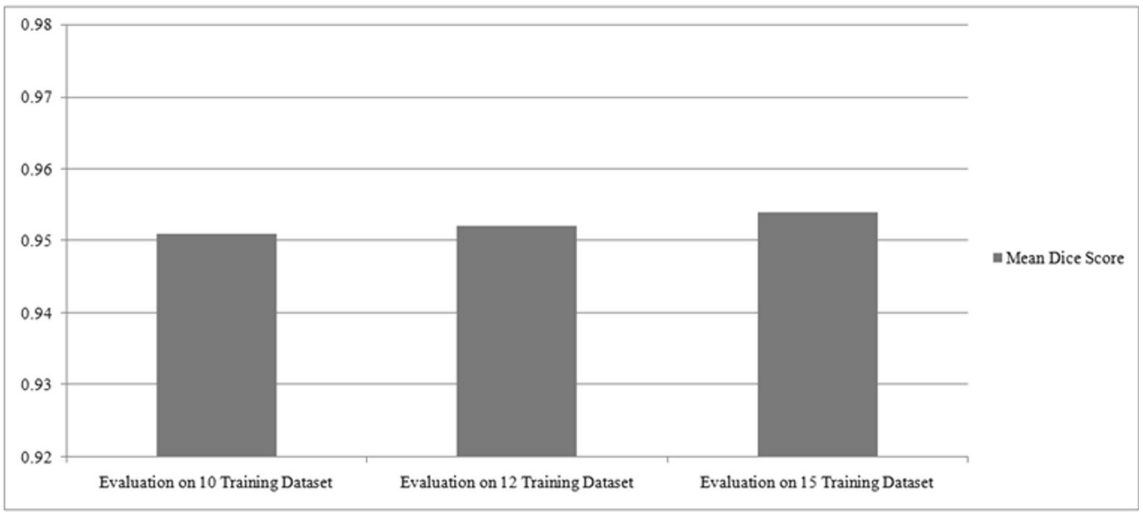

Fig. 10 Mean dice score evaluation on various combinations of training data

on the dataset. In some cases, our framework may not perform well but collectively, we have got competitive performance.

Conclusions and Future Directions

This work formulates region-based level set technique under deep learning framework. This work is also a bridge to link traditional image processing techniques with latest deep learning approaches giving two-way advantage. The framework successfully addresses the problem of vertebrae segmentation in CT image data. The region-based level set optimized predicted shape as desired. The shape convergence outperformed especially when dealing with fractured (osteoporotic) cases. The results in various ways are assessed and presented which show that our proposed framework, i.e., FU-Net is competitive over state-of-the-art methods.

In future, we have many directions to extend work in this domain like multimodal work can be carried out on X-ray or MRI image datasets. In this way, the performance will be evaluated on diverse datasets and it would be more helpful for diagnosis. In addition to that, the segmentation results on our dataset were improved with increase in training data; therefore, training data can be increased in future to improve the learning of the system. As we have achieved successful performance on our proposed framework, this framework can be employed in various other medical imaging tasks to solve the problems where iterative shape-based approach is required. Finally, this work may also be extended with unification of contextual level set in deep network to detect, segment out, and classify various medical imaging tasks.

Acknowledgments We thank the challenge organizers of MICCAI CSI 2014 and CSI 2016 (xVertSeg.v1) datasets for acquiring the datasets, preparing reference ground truth segmentations and making them publically available. We also acknowledge Dr. Faiza, focal person, POF's Hospital Wah Cantt, Pakistan, for providing us expert opinions and corrections from two clinical experts on annotated datasets of xVertSeg.v1 challenge.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Levangie PK, Norkin CC: Joint structure and function: a comprehensive analysis, 5th edition. Philadelphia: F.A. Davis Co, p. 140 Print 2011
2. Middleditch A, Olive J: Functional anatomy of the spine. In: 2nd, Vol. 1-3. Oxford: MCSP. Butterworth-Heinemann, 2005
3. Tang F-h et al.: Computer-generated index for evaluation of idiopathic scoliosis in digital chest images, a comparison with digital measurement. J Digit Imaging 21:113–120, Springer, 2007. https://doi.org/10.1007/s10278-007-9050-7.
4. Zhou Y, Liu Y, Chen Q, Gu G, Sui X: Automatic lumbar MRI detection and identification based on deep learning. J Digit Imaging, Springer, 2018. https://doi.org/10.1007/s10278-018-0130-7
5. Wang KC, Jeanmenne A, Weber GM, Thawait S, Carrino JA: An online evidence-based decision support system for distinguishing benign from malignant vertebral compression fractures by magnetic resonance imaging feature analysis. J Digit Imaging 24(3):507–515, Springer, 2010. https://doi.org/10.1007/s10278-010-9316-3
6. Johnell O, Kanis JA: An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. Osteoporos Int 17: 1726–1733, 2007. https://doi.org/10.1007/s00198-006-0172-4
7. Melton LJ, Atkinson EJ, O'Connor MK, O'Fallon WM, Riggs BL: Bone density and fracture risk in men. J Bone Miner Res 13:1915–1923, 1998
8. Das C, Baruah U, Panda A: Imaging of vertebral fractures. Indian J Endocr Metab 18(3):295–303, 2014. https://doi.org/10.4103/2230-8210.131140
9. Hernlund E et al.: Osteoporosis in the European Union: medical management, epidemiology and economic burden. Springer. Arch Osteoporos, 2013
10. Nevitt MC, Ettinger B, Black DM, Stone K, Jamal SA, Ensrud K, Segal M, Genant HK, Cummings SR: The association of radiographically detected vertebral fractures with back pain and function: a prospective study. Ann Intern Med 128:793–800, 1998
11. Anwar SM et al.: Medical image analysis using convolutional neural networks: A Review. Springer. J Med Syst 42:1–13, 2018
12. Litjens G, Kooi T, Ehteshami Bejnordi B, Setio A, Ciompi F, Ghafoorian M, van der Laak J, van Ginneken B, I. Sánchez C: A survey on deep learning in medical image analysis. Med Image Anal 42:60–88, 2017. https://doi.org/10.1016/j.media.2017.07.005.
13. Ronneberger O, Fischer P, Brox T: U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Berlin: Springer, 2015, pp. 234–241
14. Chan TF, Vese LA: Active contours without edges. TIP 10(2):266–277, 2001
15. Kass M, Witkin A, Terzopoulos D: Snakes: active contour models. Int J Comput Vis 1(4):321–331, 1988. https://doi.org/10.1007/BF00133570
16. Mahmoudi S, Benjelloun M: A new approach for cervical vertebrae segmentation. CIARP, 2007
17. Klinder T, et al: Spine segmentation using articulated shape models. Medical image computing and computer-assisted intervention: MICCAI, International Conference on Medical Image Computing and Computer-Assisted Intervention 11 Pt 1, 2008, pp 227–34
18. Roberts MG, et al: Segmentation of lumbar vertebrae using part-based graphs and active appearance models. Medical image computing and computer-assisted intervention: MICCAI, International Conference on Medical Image Computing and Computer-Assisted Intervention 12 Pt 2, 2009, pp 1017–24
19. Benjelloun M, Mahmoudi S, Lecron F: A framework of vertebra segmentation using the active shape model-based approach. Int J Biomed Imaging 2011:1–14, 2011
20. Mysling P, Petersen K, Nielsen M, Lillholm M: Automatic segmentation of vertebrae from radiographs: a sample-driven active shape model approach. In: Suzuki K, Wang F, Shen D, Yan P Eds. Machine learning in medical imaging. MLMI 2011. Lecture notes in computer science, Vol. 7009. Berlin: Springer, 2011

21. Liu X, Wu Y, Wang B: Spinal CT image segmentation based on level set method. 36th Chinese Control Conference (CCC), Dalian, 2017, pp 10956–10961. https://doi.org/10.23919/ChiCC.2017.8029105
22. Hille G, Saalfeld S, Tönnies K: Hybrid level-sets for vertebral body segmentation in clinical spine MRI. Procedia Computer Science. 90:22–27, 2016. https://doi.org/10.1016/j.procs.2016.07.005
23. Rastgarpour M, Shanbehzadeh J, Soltanian-Zadeh H: A hybrid method based on fuzzy clustering and local region-based level set for segmentation of inhomogeneous medical images. J Med Syst 38(8):68, 2014. https://doi.org/10.1007/s10916-014-0068-3
24. Sekuboyina A, Valentinitsch A, Kirschke JS, Menze BHA: Localisation-segmentation approach for multi-label annotation of lumbar vertebrae using deep nets. CoRR, abs/1703.04347, 2017
25. Sekuboyina A, Kukacka J, Kirschke JS, Menze BH, Valentinitsch A: Attention-driven deep learning for pathological spine segmentation. In: Computational methods and clinical applications in musculoskeletal imaging. Springer, volume 10734 of LNCS, 2018, pp 108–119. https://doi.org/10.1007/978-3-319-74113-0_10.
26. Janssens R, Zeng G, Zheng G: Fully automatic segmentation of lumbar vertebrae from CT images using cascaded 3D fully convolutional networks. In: IEEE 15th International Symposium on Biomedical Imaging (ISBI), 2018, pp 893–897. https://doi.org/10.1109/isbi.2018.8363715
27. Lessmann N, van Ginneken B, Išgum I: Iterative convolutional neural networks for automatic vertebra identification and segmentation in CT images. In: Medical imaging. Volume 10574 of Proceedings of SPIE, 2018, p 1057408
28. Lessmann N, van Ginneken B, de Jong PA, Išgum I: Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. In: Medical image analysis. Volume 53, 2019, pp 142–155. https://doi.org/10.1016/j.media.2019.02.005
29. Al Arif SMMR, Knapp K, Slabaugh G: Fully automatic cervical vertebrae segmentation framework for X-ray images. Comput Methods Prog Biomed 157:95–111, 2018. https://doi.org/10.1016/j.cmpb.2018.01.006
30. Kristiadi A, Pranowo P: Deep convolutional level set method for image segmentation. J ICT Res Appl 11:284, 2017. https://doi.org/10.5614/itbj.ict.res.appl.2017.11.3.5
31. Ngo TA, Lu Z, Carneiro G: Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. Med Image Anal 35:159–171, 2017
32. Yao J, Burns JE, Forsberg D, Seitel A, Rasoulian A, Abolmaesumi P, Hammernik K, Urschler M, Ibragimov B, Korez R, Vrtovec T, CastroMateos I, Pozo JM, Frangi AF, Summers RM, Li S: A multi-center milestone study of clinical vertebral CT segmentation. Comput Med Imaging Graph 49:16–28, 2016. https://doi.org/10.1016/j.compmedimag.2015.12.006
33. Ibragimov B, Korez R, Likar B, Pernus F, Xing L, Vrtovec T: Segmentation of pathological structures by landmark-assisted deformable models. IEEE Trans Med Imaging 36:1457–1469, 2017. https://doi.org/10.1109/stmi.2017.2667578
34. Barrett WA, Mortensen EN: Interactive live-wire boundary extraction. Med Image Anal 1(4):331–341, 1997. https://doi.org/10.1016/S1361-8415(97)85005-0
35. Krizhevsky A, Sutskever I, E Hinton G: ImageNet classification with deep convolutional neural networks. Neural Inf Process Syst 25, 2012, 1997. https://doi.org/10.1145/3065386