

Article

# Automatic Deep Feature Learning via Patch-Based Deep Belief Network for Vertebrae Segmentation in CT Images

Syed Furqan Qadri <sup>1</sup>, Danni Ai <sup>2</sup>, Guoyu Hu <sup>2</sup>, Mubashir Ahmad <sup>2</sup>, Yong Huang <sup>2</sup>,  
Yongtian Wang <sup>1,2</sup> and Jian Yang <sup>2,\*</sup> <sup>1</sup> School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China<sup>2</sup> Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China

\* Correspondence: jyang@bit.edu.cn

Received: 23 October 2018; Accepted: 10 December 2018; Published: 25 December 2018



**Abstract:** Precise automatic vertebra segmentation in computed tomography (CT) images is important for the quantitative analysis of vertebrae-related diseases but remains a challenging task due to high variation in spinal anatomy among patients. In this paper, we propose a deep learning approach for automatic CT vertebra segmentation named patch-based deep belief networks (PaDBNs). Our proposed PaDBN model automatically selects the features from image patches and then measures the differences between classes and investigates performance. The region of interest (ROI) is obtained from CT images. Unsupervised feature reduction contrastive divergence algorithm is applied for weight initialization, and the weights are optimized by layers in a supervised fine-tuning procedure. The discriminative learning features obtained from the steps above are used as input of a classifier to obtain the likelihood of the vertebrae. Experimental results demonstrate that the proposed PaDBN model can considerably reduce computational cost and produce an excellent performance in vertebra segmentation in terms of accuracy compared with state-of-the-art methods.

**Keywords:** vertebrae segmentation; deep belief networks; supervised learning

## 1. Introduction

An accurate segmentation of individual vertebra in computed tomography (CT) images is an essential step for the clinical diagnosis and treatment of pathologic conditions. It is used in various applications of image-based vertebrae assessment, biomechanical modeling, and surgery planning [1]. It plays a vital role in computer-aided diagnoses (CAD) as it used for different clinical tasks to diagnose various diseases like scoliosis, kyphosis, spondylolisthesis, degenerative disc disease, etc. However, the exact vertebral boundaries cannot be determined due to the articulation of the vertebrae with each other that may induce vertebral overlaps in the segmentation of adjacent vertebrae. Despite an increasing attention in vertebra segmentation recently, robust and accurate spine segmentation methodologies are still deficient. Previous works on vertebra segmentation have used a multistep approach incorporating localization of spine and vertebra detection, followed by segmentation [2,3]. Multiple techniques have been applied to vertebra segmentation, such as active shape model and snake-based methods [4,5], level sets, or graph-based techniques using normalized cuts [6,7]. Most of these models depend on prior knowledge in the form of a statistical shape model or spine atlases. The performance of such methods is high on healthy spines with a good score of DICE coefficient. However, some of these methods fail in accurate segmentation with patients of osteoporotic fractures because such patients often suffer vertebral fractures in different stages and spinal deformities, such as

scoliosis. In such kind of fractures, model-dependent segmentation may fail due to that the variability of a unique shape is different from the mean shape model.

Machine learning methods can fulfill these requirements in the condition that sufficient datasets are available for their training. Without sufficient data [8], the machine learning methods can find many different hypotheses, that all give the same accuracy on training data and the problem appears when data available is too small. Many methods work by performing in local search form that gets stuck in local optima. For instance, neural network algorithms minimize an error function over training data by employ gradient descent, and decision tree algorithms employ a greedy splitting rule for growing of decision tree. In addition, in cases where there is enough training data, it may still computationally very difficult for the learning algorithm to find the best hypothesis. A previous work [9,10] proposed to solve the vertebra detection issue on arbitrary FOVs using random forests and multilayer perceptron. In the current scanning techniques, CT is an accurate modality to access the vertebral morphology. Several vertebra segmentation techniques from CT images have been proposed in the recent years. In an earlier study, vertebra segmentation is achieved by unsupervised approaches, such as region growing, adaptive threshold, boundary adjustment [11], graph cut [12], and watershed [13]. Lim et al. [14] presented the Willmore flow technique in a level set method to conduct a surface model evaluation. The edge- and region-based level set function combination approach [6] for vertebra segmentation is proposed in CT images. The hybrid morphological filter and Gaussian mixture for automatic initialization of level set is presented in [15] to handle topological variations. In region-based approach, vertebral body segmentation is conducted to detect the key features from statistical and heuristic methods in [16]. Another technique based on a watershed algorithm for automatic spinal column segmentation has been proposed previously [17].

The vertebra segmentation and labeling methodology presented from the watershed and mathematical morphology was proposed by Naegel et al. [18]. Klinder et al. [2] presented a method based on statistical shape model (SSM) by integrating detection, identification, and segmentation of the spine. A methodology-based edge detection and registration of a deformable surface model for the thoracic segmentation was proposed in a previous work [19]. Rasoulia et al. [20] incorporated both pose statistic and shape for lumbar vertebra segmentation. A deformed fence model for separation of lumbar vertebra from surrounding tissues is proposed in the study [21]. High order Markov random field (MRF) was used for individual vertebra segmentation from an articulated spine model [22]. A landmark-based shape model vertebra is achieved by using the transportation theory and aligned the model to a vertebra from CT images using game theory [23]. Spine segmentation [24] is conducted by partitioning the vertebra into several parts using an active shape model. The work [25] for the vertebral segmentation is presented from 3D superquadric model. It is presented as an atlas approach in the osteoporotic vertebra segmentation in a past work [26]. Vertebra detection is conducted from a statistical learning approach based on Adaboost proposed in [27].

### *Related Work*

The success of deep learning methods in medical imaging has facilitated the use of these techniques in various applications, such as image detection, classification, and segmentation. Given that deep learning is an automatic feature learning method, it has become a major topic in machine learning [28], computer vision [29], and various other areas, including medical diagnosis [30]. Professional knowledge is needed to extract features and apply generalization capability to different domains. Deep learning algorithms automatically learn effective features from the datasets and thus draw an increasing interest in the new era of research. Researchers express that excellent modeling accuracy can be achieved by the deep learning-based deep belief network (DBN) with fewer hidden nodes [31,32]. The deep belief network (DBN) is a deep learning model that consists of stacking Restricted Boltzmann Machines (RBMs) [31]; RBM is a primary unit of DBN. An RBM has a generative undirected graphical modeling type. It is defined that the lower layer is called a visible layer and a top layer called the hidden layer. The nodes of the visible and hidden layers are stochastic variables and

the weights are undirected between visible and hidden layer. The two-step DBN learning consists of unsupervised and supervised. Unsupervised feature learning is implemented by the CD algorithm which trains the stacked RBMs to obtain the initial weight, and supervised learning is implemented by the back-propagation algorithm.

Patch-based methods have obtained effective results for segmentation from medical images [33,34]. These methods label each voxel in the target images by image patch comparison, center on the voxel with patches from an atlas library, and allocate the most probable label in accordance with the closest and best match. Patch-based methods of label fusion have been presented and have given robust segmentation. These methods recently show improved results in different computer vision tasks, including super-resolution [35] and texture synthesis [36]. Image denoising [37] has given a promising performance and led to various patch-based segmentation tool development for the facilitation of medical imaging application methods [33,34,38,39]. In a previous study [40], the corner detection of cervical vertebrae using the patch-based methodology in X-ray images is proposed. This work is extended for cervical vertebra segmentation in X-ray images by using a deep fully convolutional neural network (CNN) and has achieved a DICE similarity coefficient of 84% with a shape error of 1.69 mm [41]. In this work, twelve patches around the vertebra are generated to provide the spatial context, and the vertebra corners are detected using a machine learning method. A liver segmentation from CT images is presented in [42] from deep stack autoencoder (DSAE). This deep learning-based method learns features from unlabeled data by using autoencoder and fine-tuned these features for segmentation the liver from other abdominal organs. An iterative convolutional neural network based automatic method for vertebrae identification and segmentation is proposed previously [43] to achieve a mean DICE coefficient of 0.948 and a mean surface distance of 0.29 mm. This method used the inherent order of vertebral column to make solve of detection problem and the interactive procedure is utilized for vertebrae identification and segmentation one by one in sequential order. In this method for enabling the contextual information, the vertebra in the low-resolution images is first roughly identified and localized, and afterward reanalyzed in the original high-resolution images to achieve a fine segmentation. A DBN based vertebrae segmentation is proposed in our preliminary work from CT images in [44].

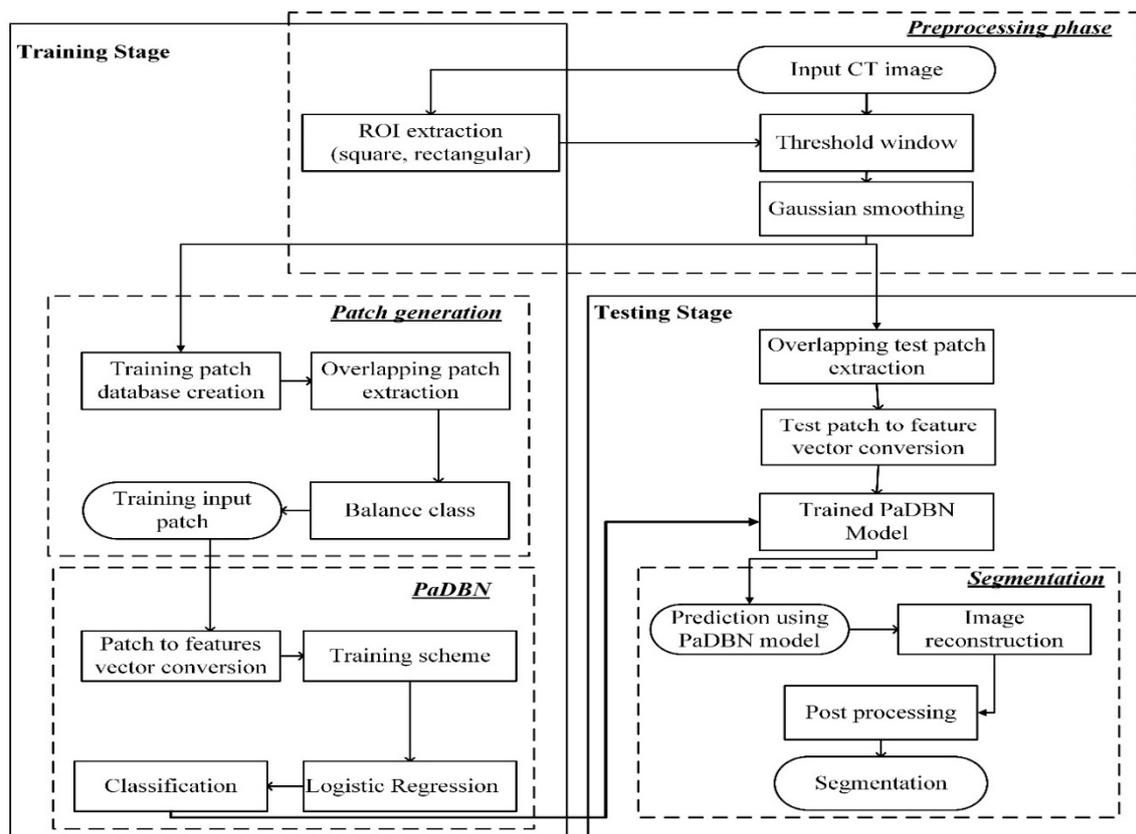
Unlike feature representation by the convolutional neural network (CNN) that comprises convolutional and subsampling tasks in learning a set of locally connected nodes through the locally receptive field for features extraction, our proposed PaDBN model is a full connection method for high-level feature learning. Unsupervised feature learning is implemented by the CD algorithm which trains the stacked RBMs to obtain the initial weight and supervised learning is implemented by the back-propagation algorithm to fine-tune the whole network and can optimize the deep-learned features for certain tasks, such as segmentation. While CNN is a partially connected approach that stresses the importance of locality while PaDBN is a fully connected approach which gets learning a single global weight matrix for features representation. The size of overlapping patches was set to  $32 \times 32$  pixels. Therefore, we choose PaDBN instead of CNN in our work.

The objective we pursue is to achieve the segmentation of vertebrae from CT images. In the present study, we propose based on our preliminary work [44], a new hierarchical deep learning feature representation PaDBN method for vertebrae segmentation. We adapted square and rectangular ROI extraction from input CT image for overlapping patch generation and used balance classes in training set to get high classification accuracy that was not done in our preliminary work [44]. Note, no matter what the proportion of samples in testing set, smaller and compact deep belief network architecture is adopted to minimize resources with better accuracy than our previous architecture. Our preliminary work describes a proof-of-concept strategy for improving vertebrae segmentation and result suggests that a fast correction method improves the segmentation and has potential to be incorporated as an additional implement into clinical usage. These preliminary results further motivated us to perform a full evaluation of PaDBN model for vertebrae segmentation accuracy. Moreover, this work also reported known limitation of the previous version and proposed a variation

of the method to circumvent them. Our proposed model is an unsupervised learning of input feature statistics from CT images. We transfer the obtained network weight matrix to back-propagation neural network with the same structure to start the supervised learning step. In supervised learning step, conjugate gradient is tested for back-propagation neural network learning. These learned features are further integrated into the patch matching framework to find patches for label propagation. Then, a deformable model is adopted for the vertebra segmentation by the vertebra likelihood mapping driven from patch matching. Finally, the postprocessing stage improves the segmentation by removing the overflow portion by different morphological operations. The output of the postprocessing step is the final vertebra segmentation. The proposed method is extensively evaluated on CT datasets from the 2014 MICCAI workshop on the spine (CSI 2014) [45]; these datasets are available on SpineWeb [46]. The experimental results show that the deep learning features of PaDBN achieve excellent segmentation accuracy. Compared with state-of-the-art vertebra segmentation methods, our algorithm obtains competitive segmentation accuracy.

## 2. Methodology

Similar to other deep learning-based frameworks, our PaDBN framework can be divided into two stages: training and testing. The PaDBN model for automatic vertebra segmentation further comprises the preprocessing phase, patch generation, PaDBN, and segmentation steps. An overview of our proposed method is shown in Figure 1.

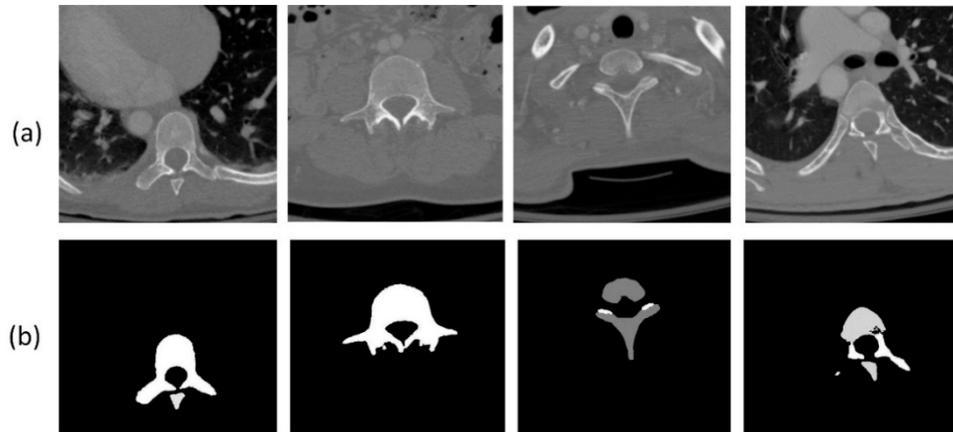


**Figure 1.** The proposed patch-based deep belief networks (PaDBN) model for vertebrae segmentation.

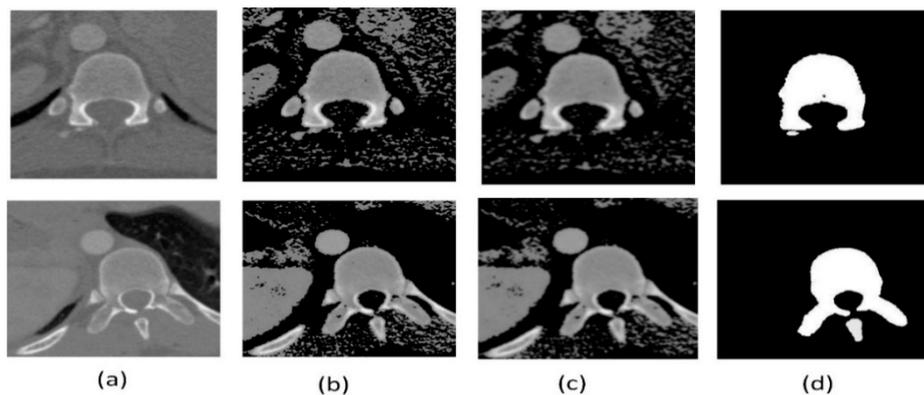
### 2.1. Preprocessing

The CT dataset can be represented by  $f: \Omega \times \{1 \dots s\} \subset \mathbb{R}^2 \times \mathbb{N} \rightarrow \mathbb{R}$ , where 's' is the number of slice images in a volume dataset,  $\Omega$  is domain, and  $f: \Omega \rightarrow \mathbb{R}$  is a bivariate function on a domain  $\Omega \subset \mathbb{R}^2$ . Prior to segmentation, the preprocessing phase is necessary for preparing the relevant information from

the CT images of the vertebra. This procedure includes different steps for preparing the meaningful features to train the model. We reduce the image size to the region of interest (ROI) by cropping the empty borders. The original  $512 \times 512$  pixel images shown in Figure 2 are cropped to bounding box around the vertebra in different pixel image sizes, and the datasets are shown in Figure 3.



**Figure 2.** Example images of CT vertebrae used in the experiment: (a) sample raw images  $512 \times 512$  pixels and (b) ground truths of selected samples, respectively.



**Figure 3.** Preprocessing phase: (a) ROIs by cropping the images using a bounding box; (b) applying threshold window; (c) Gaussian smoothing filter; and (d) respective ground truths.

The bounding box surrounding the vertebrae decreases the patch numbers for fast and efficient computation. We focus on segmenting the vertebrae and image patch containing the vertebra region. The bounding box size is sufficiently large to adjust the vertebra spatial variation across the images of the CT dataset. The overlapping patches are created from the ROI around the vertebrae. Two types of ROI geometry are considered square and rectangular ROIs. We applied the Gaussian filter to the CT images with a fixed kernel size to control the smoothness of images for achieving segmentation accuracy and attenuating the effect of noisy pixels. The large value of the Gaussian kernel is correlated with the large deformation, and the estimated transformation is used as the input of a small kernel of smoothness. As a preprocessing step, a rough threshold is applied on the entire CT volume to remove noise artifacts. The purpose of the threshold window is to distinguish the vertebrae from other soft tissues [47]. The reason for applying threshold is that vertebrae have higher intensities than other tissues in CT images but are also similar to other bony structure, such as ribs. The network learns the difference between vertebra positive samples from other bony structure as negative samples.

## 2.2. Patch Generation

We divide the cropped CT images into  $m \times m$  overlapping patches in accordance with the label on each pixel that is connected to a long-dimensional vector sequence and used as input for the network. The entire patches are created as binary class datasets, namely, vertebrae (positive sample) and nonvertebrae (negative samples), as shown in Figure 4. The overlapping patch of CT images is shuffled and then packed into the same size batches to the feed input of the PaDBN model.

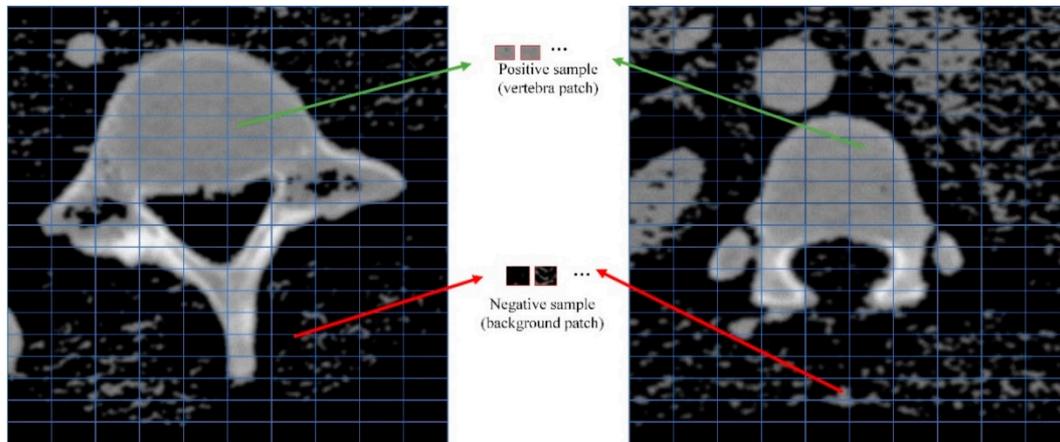


Figure 4. Positive and negative samples of the CT image.

To achieve optimal training, we use balanced training classes with equal number of samples (vertebra and background). This approach has a potential sampling bias in removing and selecting regions of sampling cases from the class with a large number of samples. The number of positive samples is equivalent to that of negative samples as balance training class that proves beneficial for model training. These steps reduce the computational time, generalization, and stability of the trained model. The purpose is to discriminate the vertebra from other objects in the CT image. Thus, the problem is two-class classification. Notably, no balanced class is done in the testing stage. For verification on balanced and unbalanced training classes, we conduct experiments to compare the accuracy on both datasets but achieve better results on balanced training class.

## 2.3. PaDBN Model

### 2.3.1. Patch to Feature Vector Conversion

The image informative representation is a feature that facilitates robust image segmentation. Deep learning, a subfield of machine learning, is used in the identification of meaningful features of the image. DBNs recently show impressive performance in various image analysis tasks [48–50]. In this study, DBN is used to convert the overlapped patches into feature vectors. To train and test our model, the image patches are converted into feature vectors. The proposed PaDBN model is used to extract the abstract features from CT images. The DBNs [31] have learning ability to extract a deep hierarchical representation of data images. Here, DBN is used as feature extractor because it can generate robust features that lead to the improvement in classification performance [48–51]. Preprocessing, i.e., cropping, normalization, smoothing, and balancing, is used to achieve the best representative features of raw images. This procedure is followed by utilizing the unsupervised modeling for extraction of deep structural features. The two hidden layers are used to extract the features from the vertebrae. The input visible layer and first hidden layer construct RBM1, and the first and second hidden layers form the RBM2 structure. The previous layer in RBM is used as a visible layer for the subsequent RBM. Considering CPU limitations and computational cost, the CT images after preprocessing are divided into  $32 \times 32$  overlapped patches to prevent any distortions. Thus,

each individual patch is presented by a row vector with 1024 dimensions. A total of 1024 dimension features from each individual patch are used in training the model. The model starts with 1024 units in the visible layer that shows high efficiency in feature selection and then finds the suitable number for these hyperparameters.

### 2.3.2. Training Scheme

A model is performed on the learned features  $x$  to generate the high-level features of patches  $p$  for characterizing the PaDBNs of vertebra CT images. The output of the former layer in DBN is used as an input for the subsequent layer after several RBMs are stacked in a hierarchical manner.

Particularly, DBN  $l$ th layer can be shown as

$$M_l = \{W^l, b^l, c^l\} \tag{1}$$

where  $W^l$  weight matrix shows the association between input and output nodes and  $b^l$  and  $c^l$  are the bias vector correspondence of  $l$ th RBMs input and output layer, respectively. The given input  $h^{l-1} = (h_1^{l-1}, h_2^{l-1}, \dots, h_{h_{l-1}}^{l-1})$ , the  $M_l$  output can denoted as

$$h^l = s(M_l, h^{l-1}) = (h_1^l, h_2^l, \dots, h_{h_l}^l) \tag{2}$$

where  $s(M_l, h^{l-1})$  shows the RBMs sampling procedure  $M_l$  with input nodes  $h^{l-1}$ . For training of a RBMs, the joint probability of  $l$ th layer can be represented as

$$P(h^{l-1}, h^l) = \frac{1}{Z} e^{-\mathcal{L}(h^{l-1}, h^l)} \tag{3}$$

The energy function can be modeled as

$$\mathcal{L}(h^{l-1}, h^l) = - \sum_{i=1}^{H_{l-1}} \sum_{j=1}^{H_l} w_{ij}^l h_i^{l-1} h_j^l - \sum_{i=1}^{H_{l-1}} b_i^l h_i^{l-1} - \sum_{j=1}^{H_l} c_j^l h_j^l \tag{4}$$

where  $w_{ij}^l$  belongs to real-valued weight linked between the nodes  $h_i^{l-1}$  and  $h_j^l$ . There are many ways to learn an RBM. Finally, the PaDBN is constructed by stacking of all learned RBMs using divergence procedure to learn the parameters of each RBMs [52]. The overlapping patches ( $m \times m$ ) are fed as the input to RBMs in a hierarchical manner. The two nodes output layer of PaDBN can be greedily trained as shown in Figure 5.

### 2.3.3. Regression Model and Classification

Two methods can be used to convert the DBN into the regression model. In the first method, the DBN is trained in an unsupervised manner and then converted into the neural network (NN) without adding any other layer. In this way, all the samples are fed forward through this NN, and the output is treated as samples for the regression model. In the second method, DBN is converted into the NN after its supervised training by adding an output layer (with two neurons) for regression on the top. Thereafter, NN is trained with the labeled dataset for fine tuning and treated as a regression model. In our work, we use the second method for the regression model. For classification, we used the sigmoid function for two-class logistic regression. Sigmoid function utilization is suitable for our binary classification problem because it maps the predicted values to probabilities. The two nodes in the output layer enable the network to select the most relevant features for representation of each image patch. Thus, the deep learning features have two dimensions. The function can be written as follows

$$\sigma(z) = \frac{1}{(1 + e^{-z})} \tag{5}$$

where  $z = \sum_{i=1}^n w_i x_i + \text{bias}$ .

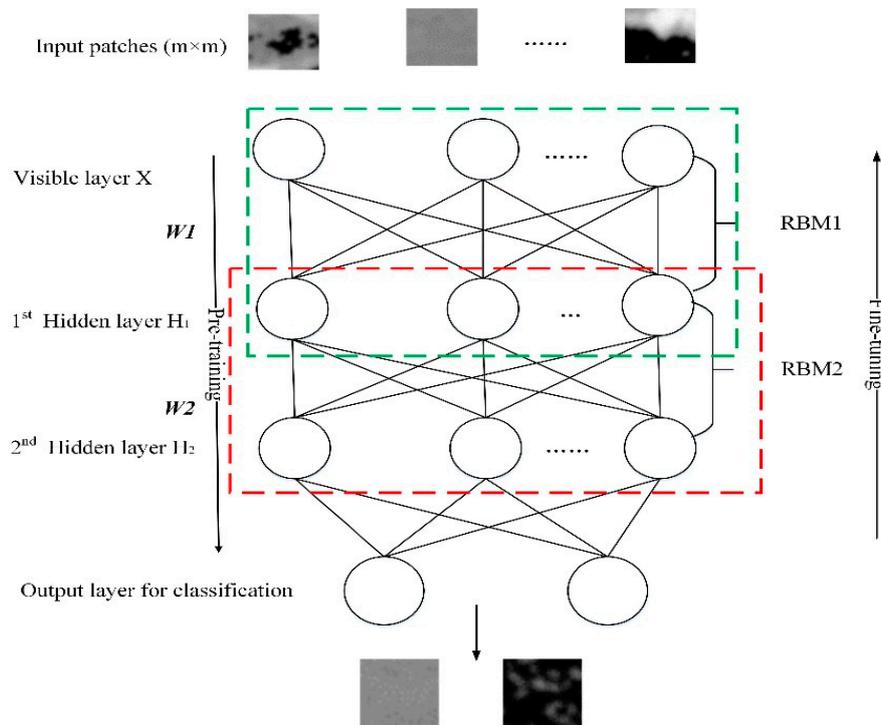


Figure 5. The structure of the proposed PaDBN Model.

#### 2.4. Segmentation

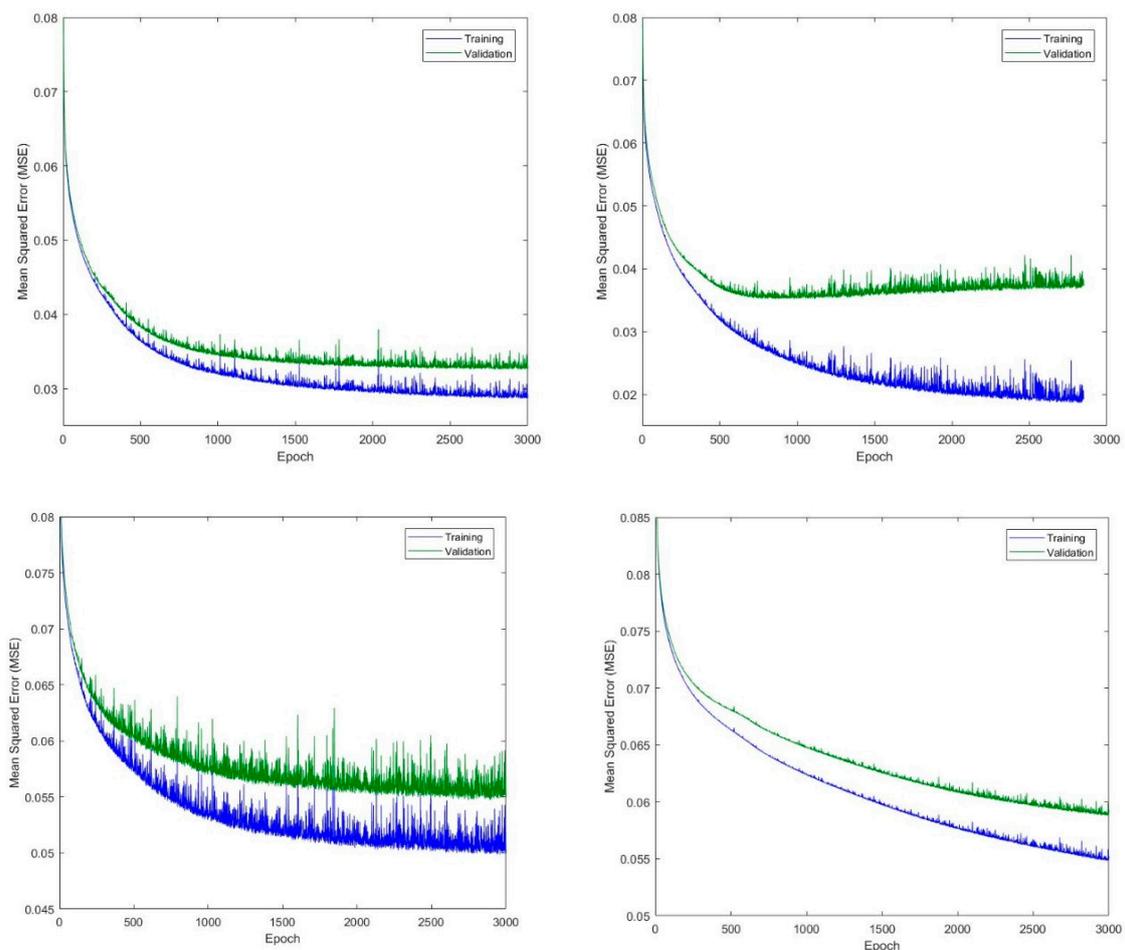
At the testing stage, image patches from unseen vertebrae are fed into the model. The prediction stage first predicts the test image class patch along with the trained model. After placing a distribution at each of the feature vectors, the final probability map is generated by adding all distributions. The proposed network focuses on vertebrae segmentation. The segmented image is obtained from the reconstruction of test patch probability map distribution. However, some flaws are observed in segmented images in boundary detection, and the results are poor. Some pixels of the background are segmented as vertebrae because of high contrast among vertebrae, ribs, and some other bony structure tissues. Moreover, some vertebra pixels are missing from foreground. To remove such flaws from the segmented images, we apply different morphological operations [11] as a postprocessing step to obtain the final segmentation.

### 3. Experimental Results

We evaluate our proposed model on CT datasets provided for CSI spine and vertebrae segmentation at the 2nd MICCAI workshop and corresponding ground truth obtained from trauma center during the daily medical examination. The CT scan comprises all thoracic (T1–T12) and lumbar (L1–L5) vertebrae with 1 mm of slice thickness, and the data resolution ranges from 0.31 mm to 0.45 mm for in-plane resolution. These datasets are obtained from the patient’s age group between 16 to 35 years old [45]. The proposed work is implemented in MATLAB 2017a as backend and trained on Dell desktop with a 3.6GHz Intel(R) i7 CPU, 32GB RAM, and Windows 7 ultimate operating system.

The training process evolves for binary segmentation of vertebrae where all the vertebrae have the same labels, and the model distinguishes between vertebrae and background in which features are learned in an unsupervised manner called pretraining. The activation conditional probability over hidden and visible units is given by the logistic function and is used as an activation function for the sigmoid function. Each RBM is trained by contrastive divergence (CD) algorithm [31]. The input image

patch of visible layer  $x$  constructs RBM1 with the first hidden layer  $h_1$ , and  $h_1$  is treated as a visible layer for constituting RBM2 with the second hidden layer  $h_2$ . In our training procedure, we use two RBMs to train our model. The parameters of RBM1 ( $W, b, c$ ) are obtained using CD algorithm. By fixing this layer parameter and training RBM2 from CD algorithm, the DBN global parameters are obtained as shown in Figure 6. A sparse penalty factor [31], which conducts the RBM sparsity by creating a connection of weight and hidden unit bias, is proposed on the basis of cross entropy. Cross entropy is used to distinguish between vertebra and nonvertebra using two small probability distributions to prevent the phenomenon of vertebra homogeneity during training. The 1024 raw dimension features are given as input to the network. The dataset is split into training 70% (training80%–validation20%) and testing 30%; seven CT images are used for training (training–validation) and three for testing.



**Figure 6.** The learning curve of the model.

The training data samples comprise the matrix with  $538,624 \times 1024$  dimensions and validation samples having a matrix with  $179,200 \times 1024$  dimensions. Five-hundred-and-thirty-eight-thousand-six-hundred-and-twenty-four samples were randomly chosen for training, and 179,200 samples were randomly chosen for validation. To achieve efficient training, the mini batch size was set to 128. The training samples were divided into 4028 mini batches, and validation samples were divided into 1400 mini batches. The architecture of PaDBN was selected as 1024-100-100-2, where the number of input nodes is 1024. The number of epochs in unsupervised pretraining is 50 to fully train every RBM of PaDBN. The 3000 epochs are used in fine-tuning. The model is trained to minimize the mean square error (MSE) between the predicted map and the ground truth. The output layer is used as a sigmoid layer to show the probability map of vertebra or nonvertebra objects. Our trained model can assess the labels of the arbitrarily sized image. Given a test image, we extract the overlapping patches of size 32

$\times 32$  and feed them to the trained model to generate prediction probability maps. For the overlapped image patch, the final probability maps will be the average probability map of overlapping patches that can be used to realize vertebra segmentation. Thus, the generalization and stability of the trained model can be guaranteed.

### Parameter Settings

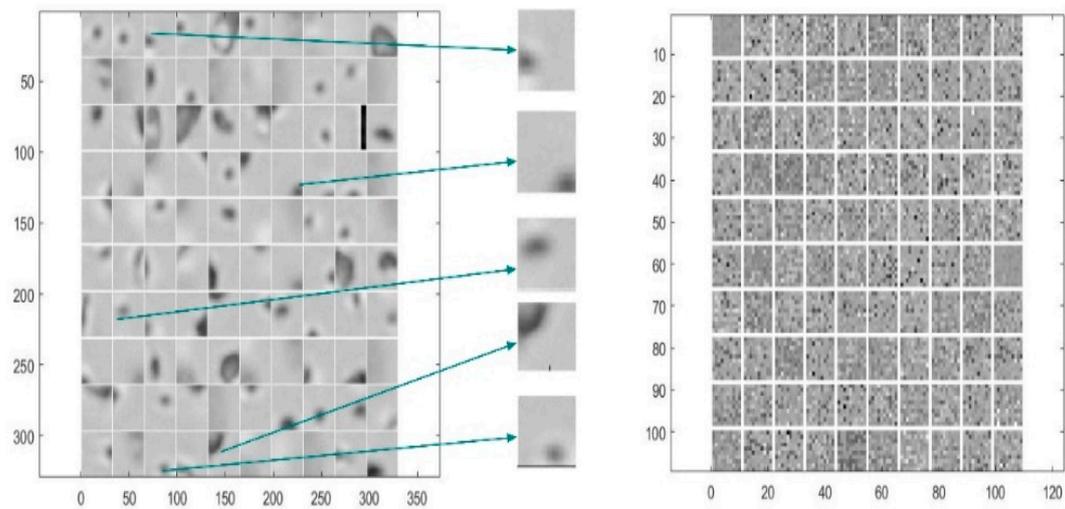
The parameters for the model are listed as the overlapped patch size  $32 \times 32$  pixel. The PaDBN model has two hidden layers. The numbers of nodes in each layer are 100 and 100, respectively. Two nodes in the output layer enable the model to select the two most relevant features to represent each patch. Thus, the deep learning features have two dimensions. To select a deep network structure, we used a random search [53] to select the structure for best performance. First, we tried to identify the range of the hyperparameters and after that, we selected the values randomly. Then, we trained our model with these selected values and we repeated this procedure until getting the best performance. The L2 regularization was set to 0.00001. The learning rate for pretraining was 0.05 and that for fine-tuning was 0.008. The momentum for pretraining was 0.5 and that for fine-tuning was 0.9. Same mini batch size 128 is used to achieve efficiency during pretraining and fine-tuning. We used 50 and 3000 epochs in pretraining and fine-tuning, respectively. The parameter combination scheme that achieves the top performance of our proposed method is shown in Table 1.

**Table 1.** Parameters combination scheme.

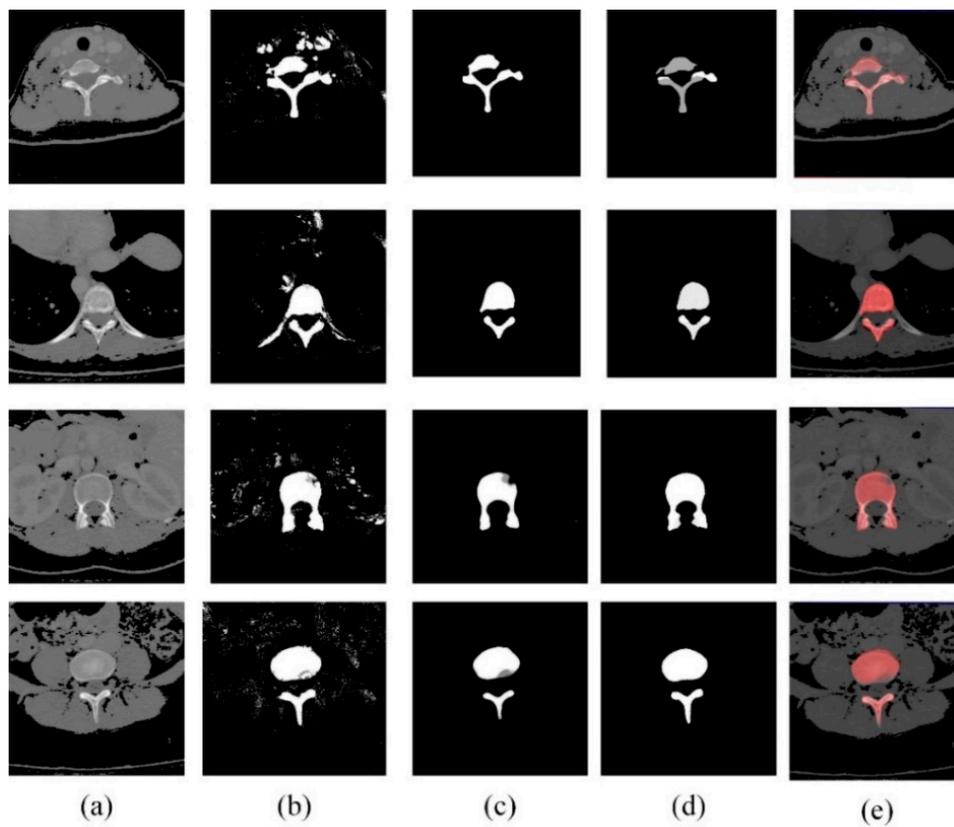
Pretraining		Fine-Tuning	
Hidden Layers	2	Scaling learning rate	1
No. of neurons in each layer	100-100	No. of epochs	3000
No. of epochs	50	Learning rate	0.008
Learning rate	0.05	Momentum	0.9
Momentum	0.5	Mini batch size	128
Mini batch size	128	weightPenaltyL2	0.00001

The learning curve of the PaDBN model is shown in Figure 6. A normalized training MSE of 0.026 and a normalized validation MSE of 0.032 were obtained after training. The learning curve shows high divergence before 500 epochs and becomes stable after 1700 epochs in the left side of Figure 6 (top left), but the (top right) graph shows the overfitting problem because the validation curve considerably deviates from the training curve after 500 epochs. The correct training curve, as shown in the (top left) graph, is obtained using the heuristic approach. Figure 6 (bottom) shows the poor training MSE graph with high learning rate (bottom left) and low learning rate (bottom right). Thus, weight initialization is important in deep learning. Figure 7 illustrates a typical feature representation of first and second hidden layers learned by two-RBM layer PaDBN model. Each square shows the weight between one hidden node and the pixel of the original image at the same position. The gray pixel indicates zero in the weight matrix, and white pixels denote positive values. The left side of Figure 7 shows the 100 node RBM1, amplified images show its weight visualization, and the right side of Figure 7 shows the weight visualization of 100 node RBM2.

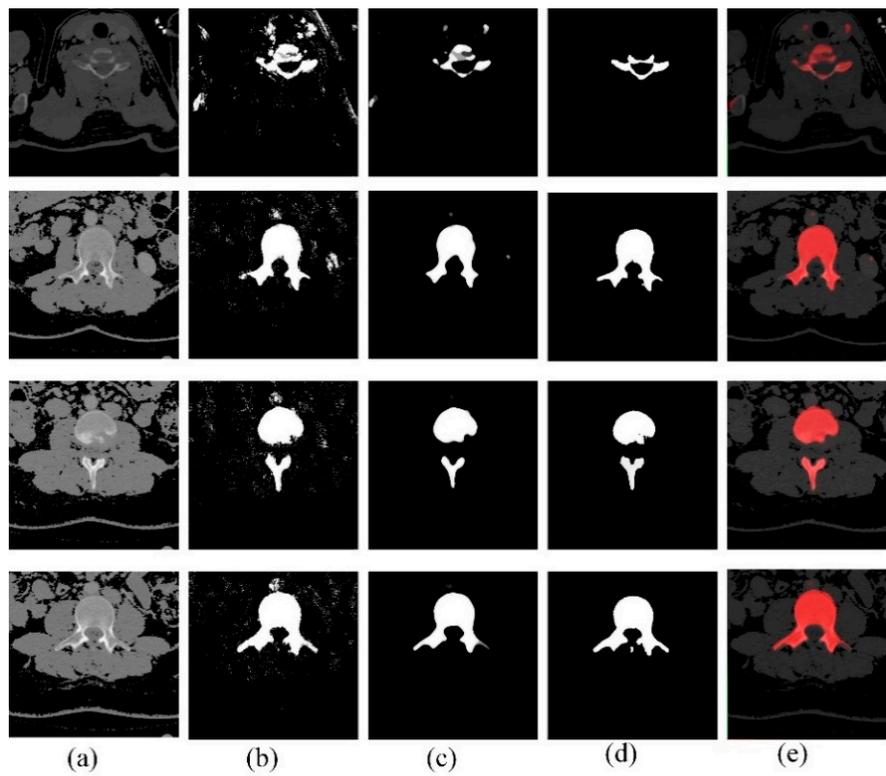
The segmentation results are shown in Figures 8–10 on different test axial images.



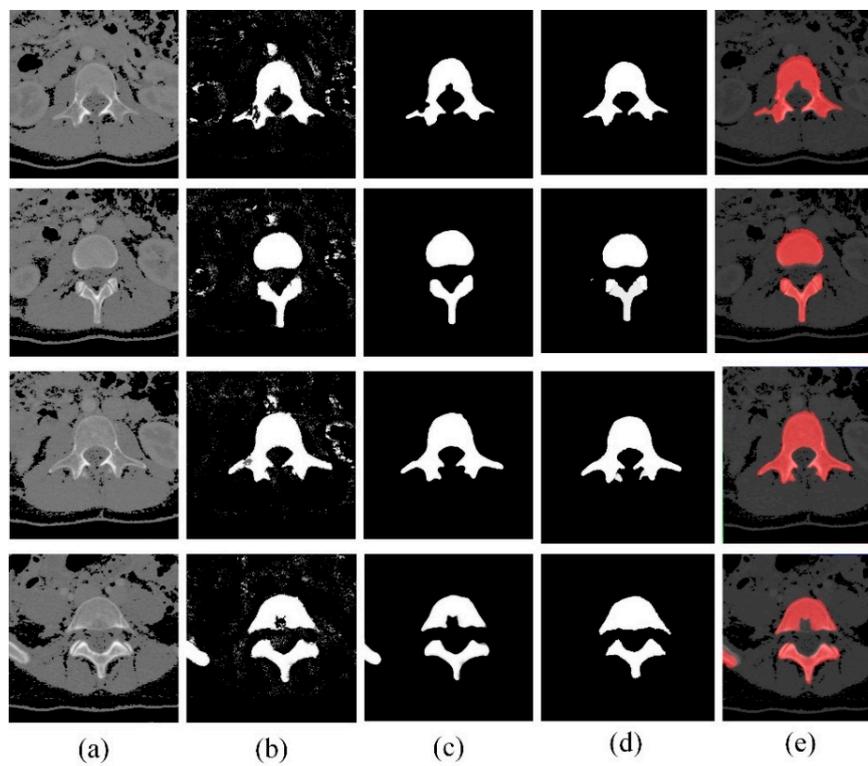
**Figure 7.** **Left side:** The visualization of 100 weights in the first layer of the deep belief network (DBN), restricted Boltzmann machines 1 (RBM1), and amplified images are some example of weights. **Right side:** The visualization of 100 weights in the second layer of DBN, RBM2.



**Figure 8.** Segmentation results from randomly selected axial slices of test CT image 1: (a) raw image; (b) initial segmentation; (c) final segmentation; (d) ground truth; and (e) segmentation superimposed on CT image.



**Figure 9.** Segmentation results from randomly selected axial slices of test CT image 2: (a) raw image; (b) initial segmentation; (c) final segmentation; (d) ground truth; and (e) segmentation superimposed on CT image.



**Figure 10.** Segmentation results from randomly selected axial slices of test CT image 3: (a) raw image; (b) initial segmentation; (c) final segmentation; (d) ground truth; and (e) segmentation superimposed on CT image.

The performance of vertebra segmentation is evaluated by specificity, sensitivity, precision, overall accuracy, F1 score, Matthews correlation coefficient (MCC) [54], DICE similarity index (DSI) [55], and Jaccard similarity index (JSI) [56] for quantitative assessment. We drew the confusion matrix to perform classic evaluation using four variables: true positive, which refers to the patches segmented correctly as vertebrae in the ground truth; false positive, which refers to the patches that are not classified as vertebrae but classified as vertebrae by method; true negative, which refers to the patches that are correctly classified as background; and false negative, which refers to the incorrectly classified patches as background. The proposed method achieves excellent results in classical measurements with an accuracy of 93.3%, a sensitivity of 91.1%, and a specificity of 93.4%. These results also confirm the improvement in the MCC results and indicate that our segmentation results show very satisfactory agreement with the ground truth (Table 2).

**Table 2.** Performance evaluation metrics by the confusion matrix.

	Specificity (%)	Sensitivity (%)	Precision (%)	Overall Accuracy (%)	F1 Score (%)	Matthews Correlation Coefficient (MCC) (%)
CT image 1	85.6	96.8	86.1	91.1	91.1	82.8
CT image 2	93.6	97.5	93.0	95.4	95.2	91.0
CT image 3	90.0	96.6	90.2	93.2	93.3	86.7
Overall	89.7	96.9	89.7	93.2	93.2	86.8

DSI measures the spatial overlapping between two binary images, and its value ranges from zero (nonoverlap) and one (perfect overlapped). The DSI is calculated for the individual vertebra and then averaged for overall vertebrae. Each vertebra in the segmentation reference is compared with the labels with which it has the largest overlapped. All automatic detected vertebrae, including the misclassified ones, are considered for this evaluation. Our proposed method improves the segmentation results of DSI with 86.1% similarity to the ground truth and JSI of 84.9% score (Table 3).

**Table 3.** Segmentation results.

	DICE Similarity Index (DSI)	Jaccard Similarity Index (JSI)
CT image 1	84.1	83.7
CT image 2	85.6	84.2
CT image 3	88.7	87.0
Overall	86.1	84.9

This research demonstrates that our proposed method can learn the complex segmentation task of vertebrae. Classification and segmentation errors are most often found in the first and last vertebrae, in which the model has less context availability. These vertebrae are often invisible, thereby making their segmentation difficult. DICE score is lower for the starting thoracic vertebrae because of ribs and intervertebral influence that are connected to the vertebrae. Some vertebrae that are not shown in the label annotations are still segmented by our method, thereby causing misclassification and low DICE score. These factors mislead segmentation. To show the generalization of our proposed method, we utilize the CSI 2014 dataset on the vertebra segmentation (Table 4).

**Table 4.** Our method compared with state-of-the-art methods on vertebrae segmentation.

Methods	DSI (%)
Y. Li et al. [15]	81.0
S. F. Qadri et al. [44]	85.0
A. Seitel et al. [57]	83.0
A. Sekuboyina et al. [58]	87.0
S. M. M. R. Al Arif et al. [41]	84.0
Our method	86.1

Compared with the findings of other related methods, our obtained results are satisfactory. For example, a mean DICE score of 85% is obtained in our preliminary work [44] for thoracic and lumbar vertebrae segmentation from CT images by using deep belief networks modeling but that has complex deep network architecture. DICE coefficient of 83% is reported in [57] by evaluation on thoracic and lumbar vertebrae, and the method utilizes the statistical multiobject shape and pose for vertebra segmentation. However, this model lacks automatic initialization and possible segmentation overlap at the boundaries for the closely spaced thoracic vertebra. Meanwhile, an average DICE score of 81% is obtained for lumbar vertebrae CT image segmentation using level set formulation in the presence of different levels of salt and pepper noises [15]. A mean DICE coefficient of 87% is achieved by deep learning method for spine segmentation, and the method is evaluated on thoracic and lumbar CT vertebrae images [58]. However, this model fails in the structural consistency of the vertebrae parts segmentation or sometimes entire vertebra from starting and ending slices. Another deep learning method [41] is proposed for cervical vertebrae segmentation from X-ray images and achieved DICE similarity coefficient of 84%; however, the patch-based center localization method has the limitation of not knowing which center belongs to which vertebra. Our method is implemented in MATLAB and thus consumes time to segment the vertebrae. A possible solution to improve the efficiency of our method is to implement it in Keras with TensorFlow backend in Python or in Caffe. In this way, the computational time will be reduced.

#### 4. Conclusions

In this paper, we propose a PaDBN learning model for vertebra segmentation in CT images on publicly available CSI 2014 dataset. The proposed model achieves promising results and is therefore applicable to the clinical setting. To address the challenges in ensuring robustness of feature representation and the large variations in the vertebrae, we propose to extract deep learning features by the PaDBN method. Then, the learned features are utilized under the patch matching model to estimate the vertebra likelihood map of the target image. Our proposed model based on deep learning shows superior performance and is more precise, robust, and generalized than traditional models that are dependent on predefined shapes and edges for fine-tuning. We believe that its performance can be further improved with a larger and more representative dataset than the current tested dataset. Compared with state-of-the-art methods, our framework demonstrates an excellent performance in vertebra segmentation. Our future research will extend the proposed method to different image modalities and spinal structures and will validate the method on large image databases with various spinal pathologies.

**Author Contributions:** Conceptualization, S.F.Q. and J.Y.; methodology, S.F.Q. and D.A.; software, S.F.Q. and M.A.; validation, S.F.Q., G.H. and M.A.; formal analysis, Y.H.; investigation, G.H.; writing—original draft preparation, S.F.Q.; writing—review and editing, J.Y. and D.A.; visualization, Y.H.; supervision, Y.W.; funding acquisition, J.Y.

**Funding:** This work was supported by the National Key Research and Development Program of China (2017YFC0110700).

**Conflicts of Interest:** The authors declare no conflicts of interest.

#### References

1. Pereanez, M.; Lekadir, K.; Castro-Mateos, I.; Pozo, J.; Lazary, A.; Frangi, A. Accurate Segmentation of Vertebral Bodies and Processes using Statistical Shape Decomposition and Conditional Models. *IEEE Trans. Med. Imaging* **2015**, *34*, 1627–1639. [[CrossRef](#)] [[PubMed](#)]
2. Klinder, T.; Ostermann, J.; Ehm, M.; Franz, A.; Kneser, R.; Lorenz, C. Automated model-based vertebra detection, identification, and segmentation in CT images. *Med. Image Anal.* **2009**, *13*, 471–482. [[CrossRef](#)] [[PubMed](#)]

3. Forsberg, D. Atlas-based registration for accurate segmentation of thoracic and lumbar vertebrae in CT data. In *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*; Springer: Cham, Switzerland, 2015; pp. 49–59.
4. Kadoury, S.; Labelle, H.; Paragios, N. Spine segmentation in medical images using manifold embeddings and higher-order MRFs. *IEEE Trans. Med. Imaging* **2013**, *32*, 1227–1238. [[CrossRef](#)] [[PubMed](#)]
5. Ibragimov, B.; Korez, R.; Likar, B.; Pernuš, F.; Vrtovec, T. Interpolation-based detection of lumbar vertebrae in CT spine images. In *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*; Springer: Cham, Switzerland, 2015; pp. 73–84.
6. Huang, J.; Jian, F.; Wu, H.; Li, H. An improved level set method for vertebra CT image segmentation. *Biomed. Eng. Online* **2013**, *2*, 48. [[CrossRef](#)] [[PubMed](#)]
7. Lootus, M.; Kadir, T.; Zisserman, A. Automated radiological grading of spinal MRI. In *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*; Springer: Cham, Switzerland, 2015; pp. 119–130.
8. Dietterich, T.G. Ensemble methods in machine learning. In *International Workshop on Multiple Classifier Systems*; Springer: Cagliari, Italy, 2000; pp. 1–15.
9. Glocker, B.; Feulner, J.; Criminisi, A.; Haynor, D.R.; Konukoglu, E. Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Nice, France, 2012; pp. 590–598.
10. Suzani, A.; Rasoulouian, A.; Seitel, A.; Fels, S.; Rohling, R.N.; Abolmaesumi, P. Deep learning for automatic localization, identification, and segmentation of vertebral bodies in volumetric MR images. In *Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling*; International Society for Optics and Photonics: Bellingham, WA, USA, 2015; Volume 9415, p. 941514.
11. Kang, Y.; Engelke, K.; Kalender, W.A. A new accurate and precise 3-D segmentation method for skeletal structures in volumetric CT data. *IEEE Trans. Med. Imaging* **2003**, *22*, 586–598. [[CrossRef](#)] [[PubMed](#)]
12. Aslan, M.S.; Ali, A.; Rara, H.; Farag, A.A. An automated vertebra identification and segmentation in CT images. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 233–236.
13. Li, H.; Wang, Z. A seepage flow model for vertebra CT image segmentation. In Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China, 17–18 January 2006; pp. 6364–6367.
14. Lim, P.H.; Bagci, U.; Bai, L. Introducing willmore flow into level set segmentation of spinal vertebrae. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 115–122. [[CrossRef](#)] [[PubMed](#)]
15. Li, Y.; Liang, W.; Tan, J.; Zhang, Y. A novel automatically initialized level set approach based on region correlation for lumbar vertebrae CT image segmentation. In Proceedings of the 2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings, Turin, Italy, 7–9 May 2015; pp. 291–296.
16. Blumfield, A.; Blumfield, E. Automated Vertebral Body Image Segmentation for Medical Screening. Google Patents US8891848B2, 18 November 2014.
17. Yao, J.; O'Connor, S.D.; Summers, R.M. Automated spinal column extraction and partitioning. In Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006, Arlington, VA, USA, 6–9 April 2006; pp. 390–393.
18. Naegel, B. Using mathematical morphology for the anatomical labeling of vertebrae from 3D CT-scan images. *Comput. Med. Imaging Graph.* **2007**, *31*, 141–156. [[CrossRef](#)]
19. Ma, J.; Lu, L.; Zhan, Y.; Zhou, X.; Salganicoff, M.; Krishnan, A. Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model. *Med. Image Comput. Comput. Assist. Interv.* **2010**, *13*, 19–27. [[PubMed](#)]
20. Rasoulouian, A.; Member, S.; Rohling, R.; Member, S. Lumbar Spine Segmentation Using a Statistical Multi-Vertebrae Anatomical Shape + Pose Model. *IEEE Trans. Med. Imaging* **2013**, *32*, 1890–1900. [[CrossRef](#)] [[PubMed](#)]
21. Kim, Y.; Kim, D. A fully automatic vertebra segmentation method using 3D deformable fences. *Comput. Med. Imaging Graph.* **2009**, *33*, 343–352. [[CrossRef](#)]
22. Kadoury, S.; Labelle, H.; Paragios, N. Automatic inference of articulated spine models in CT images using high-order Markov Random Fields. *Med. Image Anal.* **2011**, *15*, 426–437. [[CrossRef](#)] [[PubMed](#)]

23. Ibragimov, B.; Likar, B.; Pernuš, F.; Vrtovec, T. Shape representation for efficient landmark-based segmentation in 3-D. *IEEE Trans. Med. Imaging* **2014**, *33*, 861–874. [[CrossRef](#)] [[PubMed](#)]
24. Roberts, M.G.; Cootes, T.F.; Pacheco, E.; Oh, T.; Adams, J.E. Segmentation of lumbar vertebrae using part-based graphs and active appearance models. *Med. Image Comput. Comput. Assist. Interv.* **2009**, *12*, 1017–1024. [[PubMed](#)]
25. Štern, D.; Likar, B.; Pernuš, F.; Vrtovec, T. Parametric modelling and segmentation of vertebral bodies in 3D CT and MR spine images. *Phys. Med. Biol.* **2011**, *56*, 7505. [[CrossRef](#)] [[PubMed](#)]
26. Wang, Y.; Yao, J.; Roth, H.R.; Burns, J.E.; Summers, R.M. Multi-atlas segmentation with joint label fusion of osteoporotic vertebral compression fractures on CT. In *International Workshop on Computational Methods and Clinical Applications for Spine Imaging*; Springer: Germany, Munich, 2015; pp. 74–84.
27. Huang, S.-H.; Chu, Y.-H.; Lai, S.-H.; Novak, C.L. Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI. *IEEE Trans. Med. Imaging* **2009**, *28*, 1595–1605. [[CrossRef](#)] [[PubMed](#)]
28. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [[CrossRef](#)] [[PubMed](#)]
29. Farabet, C.; Couprie, C.; Najman, L.; LeCun, Y. Learning hierarchical features for scene labeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1915–1929. [[CrossRef](#)] [[PubMed](#)]
30. Carneiro, G.; Nascimento, J.C.; Freitas, A. The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. *IEEE Trans. Image Process.* **2012**, *21*, 968–982. [[CrossRef](#)]
31. Hinton, G.E.; Osindero, S.; Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)]
32. La Rosa, E.; Yu, W. Randomized algorithms for nonlinear system identification with deep learning modification. *Inf. Sci.* **2016**, *364*, 197–212. [[CrossRef](#)]
33. Coupé, P.; Manjón, J.V.; Fonov, V.; Pruessner, J.; Robles, M.; Collins, D.L. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *Neuroimage* **2011**, *54*, 940–954. [[CrossRef](#)] [[PubMed](#)]
34. Rousseau, F.; Habas, P.A.; Studholme, C. A supervised patch-based approach for human brain labeling. *IEEE Trans. Med. Imaging* **2011**, *30*, 1852–1862. [[CrossRef](#)] [[PubMed](#)]
35. Freeman, W.T.; Jones, T.R.; Pasztor, E.C. Example-based super-resolution. *IEEE Comput. Graph. Appl.* **2002**, *22*, 56–65. [[CrossRef](#)]
36. Efros, A.A.; Freeman, W.T. Image quilting for texture synthesis and transfer. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, Los Angeles, CA, USA, 12–17 August 2001; pp. 341–346.
37. Buades, A.; Coll, B.; Morel, J.-M. A non-local algorithm for image denoising. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, 20–25 June 2005; Volume 2, pp. 60–65.
38. Cordier, N.; Menze, B.; Delingette, H.; Ayache, N. Patch-based segmentation of brain tissues. In *MICCAI Challenge on Multimodal Brain Tumor Segmentation*; IEEE: Nagoya, Japan, 2013; pp. 6–17.
39. Wang, Z.; Donoghue, C.; Rueckert, D. Patch-based segmentation without registration: Application to knee MRI. In *International Workshop on Machine Learning in Medical Imaging*; Springer: Nagoya, Japan, 2013; pp. 98–105.
40. Al Arif, S.M.M.R.; Asad, M.; Gundry, M.; Knapp, K.; Slabaugh, G. Patch-based corner detection for cervical vertebrae in X-ray images. *Signal. Process. Image Commun.* **2017**, *59*, 27–36. [[CrossRef](#)]
41. Al Arif, S.M.M.R.; Knapp, K.; Slabaugh, G. Fully automatic cervical vertebrae segmentation framework for X-ray images. *Comput. Methods Programs Biomed.* **2018**, *157*, 95–111. [[CrossRef](#)] [[PubMed](#)]
42. Ahmad, M.; Yang, J.; Ai, D.; Qadri, S.F.; Wang, Y. Deep-stacked auto encoder for liver segmentation. In *Chinese Conference on Image and Graphics Technologies*; Springer: China, Beijing, 2017; pp. 243–251.
43. Lessmann, N.; van Ginneken, B.; Išgum, I. Iterative convolutional neural networks for automatic vertebra identification and segmentation in CT images. In *Medical Imaging 2018: Image Processing*; International Society for Optics and Photonics: Bellingham, WA, USA, 2018; Volume 10574, p. 1057408.
44. Qadri, S.F.; Ahmad, M.; Ai, D.; Yang, J.; Wang, Y. Deep Belief Network Based Vertebra Segmentation for CT Images. In *Chinese Conference on Image and Graphics Technologies*; Springer: China, Beijing, 2018; pp. 536–545.

45. Yao, J.; Burns, J.E.; Forsberg, D.; Seitel, A.; Rasoulian, A.; Abolmaesumi, P.; Hammernik, K.; Urschler, M.; Ibragimov, B.; Korez, R.; et al. A multi-center milestone study of clinical vertebral CT segmentation. *Comput. Med. Imaging Graph.* **2016**, *49*, 16–28. [[CrossRef](#)] [[PubMed](#)]
46. Available online: [Index@Spineweb.Digitalimaginggroup.Ca](mailto:Index@Spineweb.Digitalimaginggroup.Ca) (accessed on 11 October 2018).
47. Shen, W.; Yang, F.; Mu, W.; Yang, C.; Yang, X.; Tian, J. Automatic localization of vertebrae based on convolutional neural networks. In *SPIE Medical Imaging*; SPIE: Orlando, FL, USA, 2015; Volume 9413, p. 94132E.
48. Pinaya, W.H.L.; Gadelha, A.; Doyle, O.M.; Noto, C.; Zugman, A.; Cordeiro, Q.; Jackowski, A.P.; Bressan, R.A.; Sato, J.R. Using deep belief network modelling to characterize differences in brain morphometry in schizophrenia. *Sci. Rep.* **2016**, *6*, 38897. [[CrossRef](#)]
49. Khatami, A.; Khosravi, A.; Nguyen, T.; Lim, C.P.; Nahavandi, S. Medical image analysis using wavelet transform and deep belief networks. *Expert Syst. Appl.* **2017**, *86*, 190–198. [[CrossRef](#)]
50. Li, J.; Xi, B.; Li, Y.; Du, Q.; Wang, K. Hyperspectral classification based on texture feature enhancement and deep belief networks. *Remote Sens.* **2018**, *10*, 396. [[CrossRef](#)]
51. Kim, Y.; Lee, H.; Provost, E.M. Deep learning for robust feature generation in audiovisual emotion recognition. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 3687–3691.
52. Fischer, A.; Igel, C. Training restricted Boltzmann machines: An introduction. *Pattern Recognit.* **2014**, *47*, 25–39. [[CrossRef](#)]
53. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
54. Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C.A.F.; Nielsen, H. Assessing the accuracy of prediction algorithms for classification: An overview. *Bioinformatics* **2000**, *6*, 412–424. [[CrossRef](#)]
55. Dice, L.R. Measures of the amount of ecologic association between species. *Ecology* **1945**, *26*, 297–302. [[CrossRef](#)]
56. Jaccard, P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull. Soc. Vaudoise Sci. Nat.* **1901**, *37*, 547–579.
57. Seitel, A.; Rasoulian, A.; Rohling, R.; Abolmaesumi, P. Lumbar and Thoracic Spine Segmentation Using a Statistical Multi-object Shape + Pose Model. In *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*; Springer: Cham, Switzerland, 2015; pp. 221–225.
58. Sekuboyina, A.; Kukačka, J.; Kirschke, J.S.; Menze, B.H.; Valentinič, A. Attention-driven deep learning for pathological spine segmentation. In *International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging*; Springer: Quebec City, QC, Canada, 2017; pp. 108–119.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).