

# Shape-aware Deep Convolutional Neural Network for Vertebrae Segmentation

S M Masudur Rahman Al Arif<sup>1</sup>, Karen Knapp<sup>2</sup> and Greg Slabaugh<sup>1</sup>

<sup>1</sup>City, University of London  
<sup>2</sup>University of Exeter

**Abstract.** Shape is an important characteristic of an object, and fundamental topic in computer vision. In image segmentation, shape has been widely used in segmentation methods, like the active shape model, to constrain a segmentation result to a class of learned shapes. However, to date, shape has been underutilised in deep segmentation networks. This paper addresses this gap by introducing a shape-aware term in the segmentation loss function. A deep convolutional network has been adapted in a novel cervical vertebrae segmentation framework and compared with traditional active shape model based methods. The proposed framework has been trained on an augmented dataset of 26370 vertebrae and tested on 792 vertebrae collected from a total of 296 real-life emergency room lateral cervical X-ray images. The proposed framework achieved an average error of 1.11 pixels, signifying a 36% improvement over the traditional methods. The introduction of the novel shape-aware term in the loss function significantly improved the performance by further 12%, achieving an average error of only 0.99 pixel.

## 1 Introduction

Deep learning has revolutionized the field of image classification [1–4], segmentation [5–7] and many other aspects of computer vision. Segmenting an anatomical body part in medical images is a challenging problem in the field. Although, training a deep network requires huge amount of data, which is usually not available for medical images, recent techniques using data augmentation have shown promising results for segmentation problem on medical images [8, 9]. Shape characteristics have long been used for image segmentation problems, especially in medical images [10–13]. Medical image modalities, e.g. X-ray, DXA, MRI, often produce noisy captures of anatomical body parts, where segmentation must rely on the shape information to produce reliable results. However, combining shape information in a deep segmentation network is not straightforward. In this paper, we try to solve the problem by introducing a novel shape-aware term in the segmentation loss function. To test its capability of shape preservation, we adapted the novel shape-aware deep segmentation network in a semi-automatic cervical vertebrae segmentation framework.

Segmenting the vertebrae correctly is a crucial part for further analysis in an injury detection system. Previous work in vertebrae segmentation has largely been dominated

by statistical shape model (SSM) based approaches [14–22]. These methods record statistical information about the shape and/or the appearance of the vertebrae based on a training set. Then the mean shape is initialized either manually or semi-automatically near the actual vertebra. The model then tries to converge to the actual vertebra boundary based on a search procedure. Recent work, [19–22] utilizes random forest based machine learning models in order to achieve shape convergence. In contrast to these methods, we propose a novel deep convolutional neural network (CNN) based method for vertebrae segmentation. Instead of predicting the shape of a vertebra, our framework predicts the segmentation mask of a vertebrae patch. In order to preserve the vertebra shape, a novel shape-aware loss term has been proposed. From a training set of 124 X-ray images containing 586 cervical vertebrae, 26370 vertebrae patch-segmentation mask pairs have been generated through data augmentation for training the deep network. The trained framework has been tested on dataset of 172 images containing 792 vertebrae. An average pixel-level accuracy of 97.01%, Dice similarity coefficient 0.9438 and shape error of 0.99 pixel have been achieved.

The key contributions of this work are two fold. First, the introduction of a novel shape-aware term in the loss function of a deep segmentation network which learns to preserve the shape of the target object and significantly improved the segmentation accuracy. Second, the application and adaptation of deep segmentation networks to achieve vertebrae segmentation in real life medical images which outperformed the traditional SSM based methods by 35%.

## 2 Data

A total of 296 lateral cervical spine X-ray images were collected from Royal Devon and Exeter Hospital in association with the University of Exeter. The age of the patients varied from 17 to 96. Different radiographic systems (Philips, Agfa, Kodak, GE) were used to produce the scans. Image resolution varied from 0.1 to 0.194 mm per pixel. The images include examples of vertebrae with fractures, degenerative changes and bone implants. The data is anonymized and standard research protocols have been followed. The size, shape, orientation of spine, image intensity, contrast, noise level all varied greatly in the dataset. For this work, 5 vertebrae C3-C7 are considered. C1 and C2 have an ambiguous appearance due to their overlap in lateral cervical radiographs, and our clinical experts were not able to provide ground truth segmentations for these vertebral

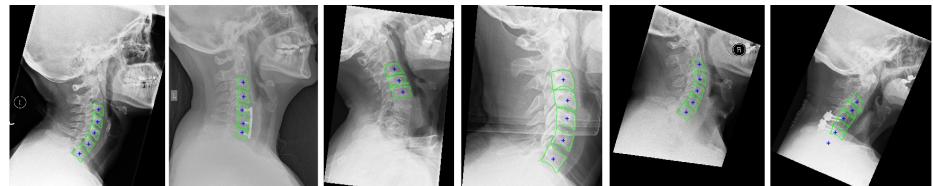


Fig. 1: X-Ray images and manual annotations. Center: blue plus (+) Vertebrae boundary curve (green).

bodies. For this reason they are excluded in this study, similar to other cervical spine image analysis research [15, 23]. Each vertebra from the images was manually annotated for the vertebral body boundaries and centers by an expert radiographer. A few examples with corresponding manual annotations are shown in Fig. 1.

The images were received in two sets. The first set of 124 images are used for training and the rest are kept for testing. The manually clicked center points and the vertebrae boundary curves are used extract the vertebrae image patch and corresponding segmentation masks. Different patch size and rotation angles are considered in order to augment the training data. After data augmentation, we ended up with 26370 vertebrae training patches. All the patches were then resized to  $64 \times 64$  pixel patches. The corresponding vertebrae curves were converted to binary segmentation masks of the same size. A few training vertebrae patches and corresponding overlaid segmentation masks are shown in Fig. 2. Similarly vertebrae patches were also collected from the test images. The orientation and scale for the test vertebrae were computed by the manually clicked center points only, shape information was not used. Our assumption is that the center points will be manually provided at test time, making the process semi-automatic. Some test vertebra are shown in Fig. 3. Note the differences in intensity, texture, and contrast, coupled with the possibility of surgical implants, making for a challenging problem on real-world data.

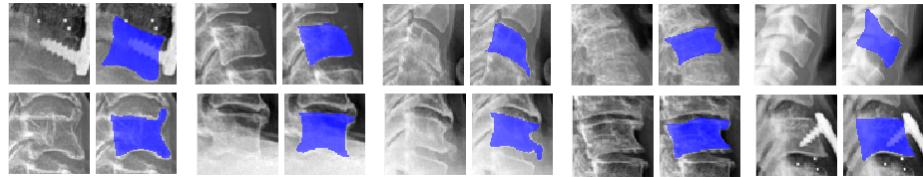


Fig. 2: Training vertebrae patches and corresponding segmentation masks (blue overlay).



Fig. 3: Examples of test vertebrae patches.

### 3 Methodology

Several deep segmentation networks have achieved outstanding performance in natural images [5–7]. However, medical images have their own set of challenges to overcome. The UNet architecture have shown excellent capability of segmenting different target objects in different medical image modalities [8, 9]. Following the literature, for our shape-aware vertebrae segmentation problem, we have chosen a modified version of the original UNet [8] deep segmentation network.

### 3.1 Network architecture

The UNet architecture consists of a contracting path and an expanding path in the network. The contracting path reduces the spatial dimension of an input image to a smaller version and the expanding path expands spatial dimension and results in a segmentation map at the output. In the original architecture [8], the spatial dimension output segmentation map is smaller than that of the input images due to the use of convolution layers without padding. In our version, we want to keep the spatial dimension of the input image and the output segmentation map same. Our architecture has nine convolutional layers in the contracting path. Each convolutional layer is followed by a batch normalization and a rectified linear unit (ReLU) layer. Three  $2 \times 2$  pooling layers, one each after two consecutive convolutional layers, reduce the input size of  $64 \times 64$  to a smaller dimension of  $8 \times 8$  at the end of the contracting path. This data is then forwarded through a mirrored expanding path of the network. The upsampling after every two convolution layer in the expanding path is achieved by a deconvolution layer with  $2 \times 2$  kernel size. The network shares intermediate information from the contracting path to the expanding path by concatenation of data. After each upsampling, the data in the expanding path is concatenated by the corresponding data from the contracting path. This helps the network to recover some of the information lost during max-pooling operation. Our network takes a single channel vertebra patch of spatial dimension  $64 \times 64$  and predicts a two channel probabilistic output for the prediction vertebrae mask of the same size. Fig. 4 details the network diagram. The number of filters in each convolutional/deconvolutional layer can be tracked from the intermediate data dimensions in Fig. 4. The total number of parameters in the network is 24,238,210.

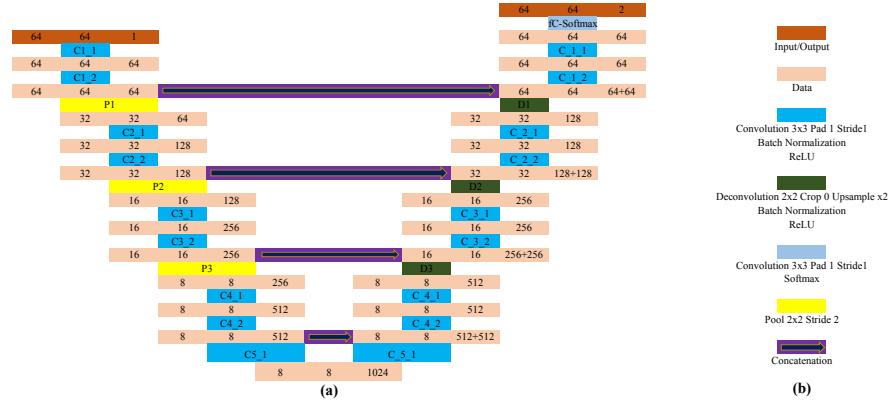


Fig. 4: UNet architecture: (a) Network diagram (b) Legends.

### 3.2 Loss function

Given a dataset of training image ( $x$ )-segmentation label ( $y$ ) pairs, training a deep segmentation network means finding a set of parameters  $\mathbf{W}$  that minimizes a loss function,  $L_t$ . The simplest form of the loss function for segmentation problem is the pixel-wise

log loss.

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W}} \sum_{n=1}^N L_t(\{x^{(n)}, y^{(n)}\}; \mathbf{W}) \quad (1)$$

where  $N$  is the number of training examples and  $\{x^{(n)}, y^{(n)}\}$  represents  $n$ -th example in the training set with corresponding manual segmentation. The pixel-wise segmentation loss per image can be defined as:

$$L_t(\{x, y\}; \mathbf{W}) = - \sum_{i \in \Omega_p} \sum_{j=1}^M y_i^j \log P(y_i^j = 1 | x_i; \mathbf{W}) \quad (2)$$

$$P(y_i^j = 1 | x_i; \mathbf{W}) = \frac{\exp(a_j(x_i))}{\sum_{k=1}^M \exp(a_k(x_i))} \quad (3)$$

where  $a_j(x_i)$  is the output of the penultimate activation layer of the network for the pixel  $x_i$ ,  $\Omega_p$  represents the pixel space,  $M$  is the total number of segmentation class labels and  $P$  are the corresponding class probabilities. However, this term doesn't constrain the predicted masks to conform to possible vertebra shapes. Since vertebrae shapes are known from the provided manual segmentation curves, we add a novel shape-aware term in the loss function to force the network to learn to penalize predicted areas outside the curve.

### 3.3 Shape-aware term

For training the deep segmentation network, we introduce a novel shape-based term,  $L_s$ . This term forces the network to produce a prediction masks similar to the training vertebra shapes. This term can be defined as:

$$L_s(\{x, y\}; \mathbf{W}) = - \sum_{i \in \hat{\Omega}_p} \sum_{j=1}^M y_i^j E_i \log P(y_i^j = 1 | x_i; \mathbf{W})$$

$$E_i = D(\hat{C}, C_{GT}); \quad (4)$$

where  $\hat{C}$  is the curve surrounding the predicted regions and  $C_{GT}$  is ground truth curve. The function,  $D(\cdot)$ , computes the average point to curve Euclidean distance between the predicted shape,  $\hat{C}$  and the ground truth shape,  $C_{GT}$ .  $\hat{C}$  is generated by locating the boundary pixels of the predicted mask. The redefined pixel space,  $\hat{\Omega}_p$ , contains the set of pixels where the prediction mask doesn't match the ground truth mask. These terms can also be explained using the toy example shown in Fig. 5. Given a ground truth mask (Fig. 5a) and a prediction mask (Fig. 5b),  $E_i$  is computed by measuring the average distance between the ground truth (green) curve and prediction (red) curve (Fig. 5c). Fig. 5d shows the redefined pixel space,  $\hat{\Omega}_p$ . This term adds additional penalty proportional to Euclidean distance between predicted and ground truth curve to the pixels that do not match the ground truth segmentation mask. In the case when the predicted mask is a cluster of small regions, especially during the first few epochs in training,  $E_i$  becomes very large because of the increase in the boundary perimeters from the disjoint predictions. Thus, this term also implicitly forces the network to learn to predict single connected prediction masks faster.

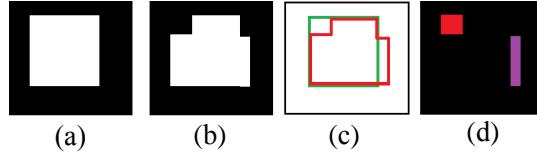


Fig. 5: Shape-aware loss: (a) Ground truth mask (b) Predicted mask (c) Ground truth shape,  $C_{GT}$  (green) and predicted shape,  $\hat{C}$  (red) (d) Refined pixel space,  $\hat{\Omega}_p$ : False positive (purple) and false negative (red).

### 3.4 Updated Loss Function

Finally, the loss function of Eqn. 1 can be extended as:

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W}} \sum_{n=1}^N L_t(\{x^{(n)}, y^{(n)}\}; \mathbf{W}) + L_s(\{x^{(n)}, y^{(n)}\}; \mathbf{W}) \quad (5)$$

The contribution of each term in the total loss can be controlled by introducing a weight parameter in Eqn. 5. However, in our case, best performance was achieved when both terms contributed equally.

## 4 Experiments

We have two versions of the deep segmentation network: UNet and UNet-S. ‘-S’ signifies the use of updated shape-aware loss function of Eqn. 5. The networks are trained for 30 epochs with batch size of 25 vertebrae patches. To update the network parameters, RMSprop version of mini-batch gradient descent algorithm is used [24]. Each network took around 30 hours to complete training in computer equipped with a NVIDIA Pascal Titan X GPU. In order to compare with the deep segmentation network based prediction results, three active shape model (ASM)-based shape prediction frameworks have been implemented. A simple maximum gradient-based image search based ASM (ASM-G) [14], a Mahalanobis distance based ASM (ASM-M) [15] and a random forest based ASM (ASM-RF) [21]. The later two have been used in cervical vertebrae segmentation in different datasets.

### 4.1 Inference and metrics

At test time, 792 vertebrae from 172 test images are extracted based on the manually clicked vertebral centers. These patches are forwarded through each of the networks to get the prediction masks. These prediction masks are compared with the ground truth segmentation mask to compute number pixels detected as true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Based on these measures two metrics are computed for each set of test vertebra patch and prediction masks: pixel-wise accuracy (pA) and Dice similarity coefficients (DSC). For the ASM based

shape predictors, the predicted shape is converted to a prediction map to measure these metrics.

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (6)$$

$$pA = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (7)$$

These metrics are well suited to capture the number of correctly segmented pixels, but they fail to capture the differences in shape. In order to compare the shape of the predicted mask appropriately with the ground truth vertebrae boundary, the predicted masks of the deep segmentation networks are converted into shapes by locating the boundary pixels. These shapes are then compared manually annotated vertebral boundary curves by measuring average point to curve Euclidean distance between them, similar to Eqn. 4. A final metric, called fit failure [20], is also computed which measures the percentage of vertebrae having an average point to ground truth curve error of greater than 2 pixels.

## 5 Results

Table 1 reports the average median, mean and standard deviation (std) metrics over the test dataset of 792 vertebrae for all the methods. The deep segmentation networks clearly outperform the ASM-based methods. Even the worst version of our framework, UNet achieves a 2.9% improvement in terms of pixel-wise accuracy and an increase of 0.055 for Dice similarity coefficient. Among the two version of deep networks, the use of novel loss function improves the performance by 0.31% in terms of pixel-wise accuracy. In terms of Dice similarity coefficient, the improvement is in the range of 0.006. Although, subtle, the improvements are statistically significant according to a paired t-test at a 5% significance level. Corresponding p-values between the two versions of the network are reported in Table 1. Bold fonts indicates the best performing metrics. Interestingly, among the ASM-based methods, the simplest version, ASM-G, performs better than the alternatives. Recent methods [15, 21], have failed to perform robustly on our challenging dataset of test vertebrae.

Table 1: Average quantitative metrics for mask prediction.

	Pixel-wise accuracy (%)				Dice similarity coefficient			
	Median	Mean	Std	p-value	Median	Mean	Std	p-value
ASM-RF	95.09	90.77	8.98		0.881	0.774	0.220	
ASM-M	95.09	93.48	4.92		0.900	0.877	0.073	
ASM-G	95.34	93.75	4.48		0.906	0.883	0.066	
UNet	97.71	96.69	3.04		0.952	0.938	0.048	
UNet-S	<b>97.92</b>	<b>97.01</b>	<b>2.79</b>	$7.17 \times 10^{-13}$	<b>0.957</b>	<b>0.944</b>	<b>0.044</b>	$7.76 \times 10^{-13}$

The average point to curve error for the methods are reported in Table 2. The deep segmentation framework, UNet, produced a 35% improvement over the ASM-based methods in terms of the mean values. The introduction of the novel loss term in the

Table 2: Average quantitative metric for shape prediction.

	Average point to curve error in pixels				Fit failure(%)
	Median	Mean	Std	p-value	
ASM-RF	1.82	2.59	1.85		43.43
ASM-M	1.54	1.88	1.05		32.70
ASM-G	1.38	1.73	0.99		26.89
UNet	0.77	1.11	1.29		8.59
UNet-S	<b>0.78</b>	<b>0.999</b>	<b>0.67</b>	0.0043	<b>6.06</b>

training further reduced the average error by 12% achieving the best error of 0.99 pixels. The most significant improvement can be seen in the fit failure which denotes the percentage of the test vertebrae having an average error of higher than 2 pixels. The novel shape-aware network, UNet-S, has achieved drop of around 37% from the ASM-RF method. The cumulative distribution of the point to curve error is also plotted in the performance curve of Fig. 6. It can be seen adaptation deep segmentation network provides a big improvement in area under the curve.

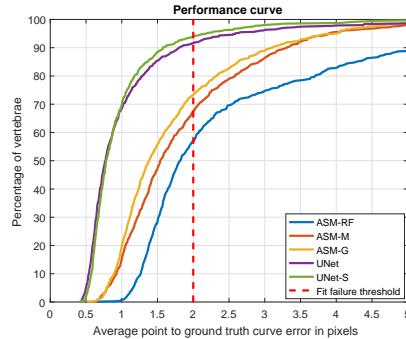


Fig. 6: Performance curve: Cumulative distribution of point to curve errors.

The box plot of the quantitative metrics are shown in Fig. 7. It can be seen that, even the worst outlier for shape-aware network, UNet-S, have a pixel-wise accuracy

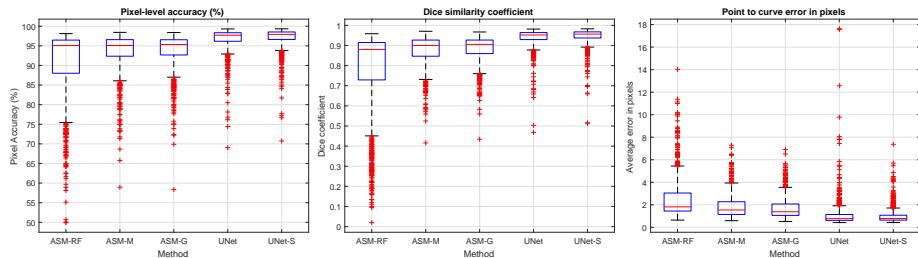


Fig. 7: Box plot of quantitative metrics: pixel-level accuracy (left), Dice similarity coefficients (middle) and point to manual segmentation curve error (right).

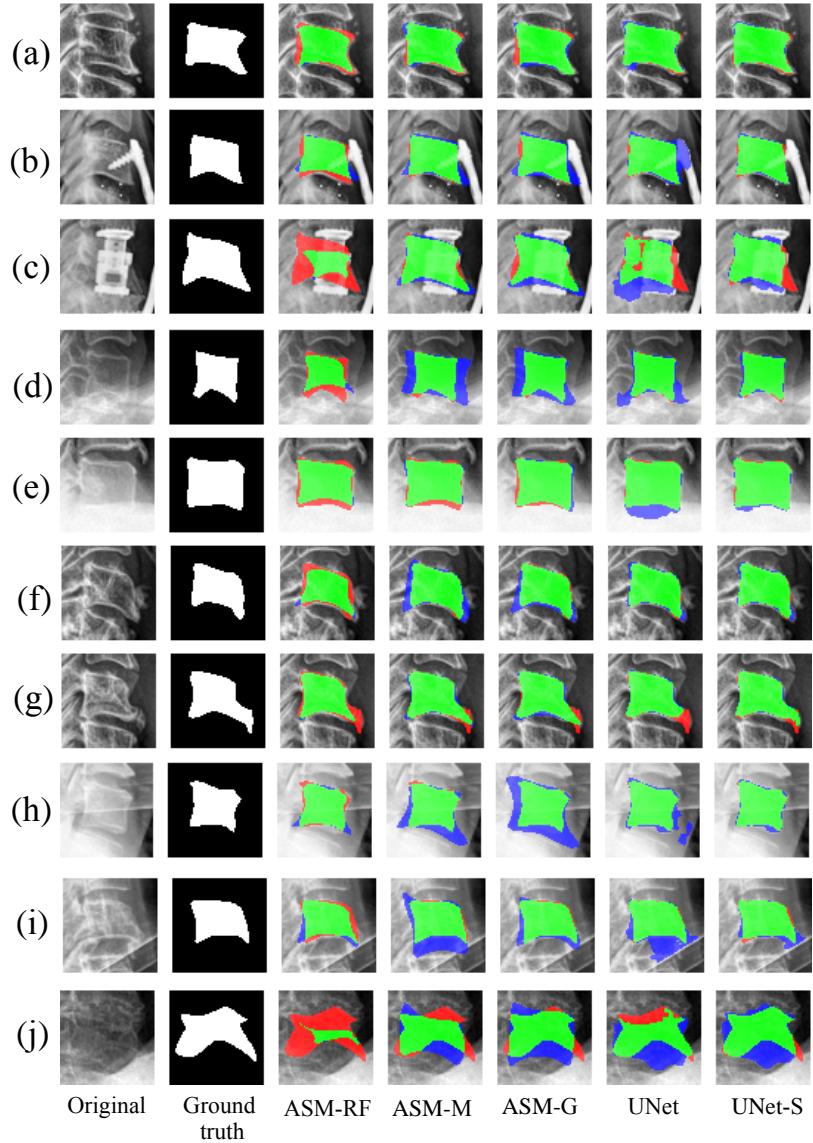


Fig. 8: Qualitative segmentation results: true positive (green), false positive (blue) and false negative (red).

higher than 70%, signifying the regularizing capability of the novel term. Most of the outliers are caused by bone implants, fractured vertebrae or abnormal artifacts in the images. A few examples for qualitative assessment are shown in Fig. 8. Fig. 8a shows an easy example where all the methods perform well. Examples with bone implants are shown in Fig. 8b and c. Fig. 8d and e shows vertebrae with abrupt contrast change. Vertebrae with fracture and osteoporosis are shown in Fig. 8f and g. Fig. 8g also shows

how UNet-S has been able capture the vertebrae fractures pattern. Fig. 8h and i show vertebrae with image artefacts. A complete failure case is shown in Fig. 8j. In all cases the shape-aware network, UNet-S, has produced better segmentation results than its counterpart.

### 5.1 Analysis on harder cases

Although statistically significant, the difference in performance between the UNet and UNet-S is subtle over the whole dataset of test vertebrae. This is because majority of the vertebrae are healthy and shape-awareness does not improve the results by a big margin. To show the shape-awareness capability of UNet-S a selection 52 vertebrae with severe clinical conditions are chosen. The average metrics for this subset of test vertebrae between UNet and UNet-S is reported in Table 3. An improvement of 1.2% and 0.02 have been achieved in terms of pixel-wise accuracy and Dice similarity coefficient, respectively. The difference over the whole dataset were only 0.31% and 0.006. The metric, point to curve error produces the most dramatic change. The novel shape-aware network, UNet-S, reduced the error by 22.9% for this subset of vertebrae with severe clinical conditions. Fig. 9 shows a few example of these subset of images.

Table 3: Comparison of UNet and UNet-S.

	Average quantitative metrics		
	Pixel-wise accuracy (%)	Dice coefficient	Point to curve error
UNet	94.01	0.91	1.61
UNet-S	95.21	0.93	1.24

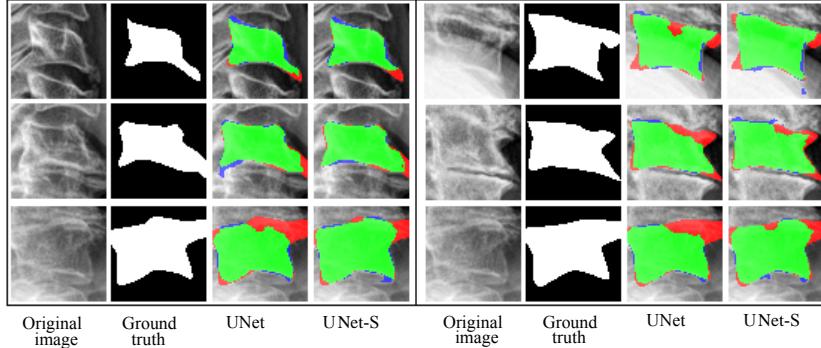


Fig. 9: Comparison of performance for vertebrae with severe clinical condition.

## 6 Conclusion

Deep segmentation networks have shown exciting application in different medical image modalities. The shape of an anatomical object is very important for automated computer aided diagnosis and injury detection. Our overarching goal is to build a computer aided system that can help the emergency department physicians to detect injuries with better accuracy. Towards this goal, in this paper, we proposed a robust semi automatic vertebrae segmentation method using deep convolutional neural networks that

incorporate the shape information in to achieve better segmentation accuracy. The proposed deep segmentation method has outperformed the traditional active shape model based approaches by a significant margin. In order to incorporate shape information with the mask prediction capability of the deep neural networks, a novel shape-aware loss function has been formulated. Inclusion of this novel term in the training provided significant quantitative and qualitative improvements. A maximum average pixel-level segmentation accuracy of 97.01%, Dice coefficient of 0.9438 and point to ground truth curve error of less than 1 pixel has been achieved over a diverse dataset of 792 test vertebrae collected from real life medical emergency rooms. Currently, we are working on a fully automatic localization framework to locate the vertebrae centers in arbitrary X-ray images. In the future, we will be using the segmented vertebrae column to automatically determine various clinical conditions like misalignment of the vertebral body, osteoporosis, bone density abnormality and type and severity of different vertebral fractures.

## Acknowledgements

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

## References

1. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.
2. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
3. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2015.
4. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
5. E. Shelhamer, J. Long, and T. Darrell, “Fully convolutional networks for semantic segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.
6. S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, “Conditional random fields as recurrent neural networks,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1529–1537, 2015.
7. H. Noh, S. Hong, and B. Han, “Learning deconvolution network for semantic segmentation,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1520–1528, 2015.
8. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, 2015.
9. A. BenTaieb and G. Hamarneh, “Topology aware fully convolutional networks for histology gland segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 460–468, Springer, 2016.

10. P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, “User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability,” *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.
11. C. Pluempitiwiriyawej, J. M. Moura, Y.-J. L. Wu, and C. Ho, “Stacs: New active contour scheme for cardiac mr image segmentation,” *IEEE Transactions on Medical Imaging*, vol. 24, no. 5, pp. 593–603, 2005.
12. J. Weese, I. Wächter-Stehle, L. Zagorchev, and J. Peters, “Shape-constrained deformable models and applications in medical imaging,” in *Shape Analysis in Medical Image Analysis*, pp. 151–184, Springer, 2014.
13. A. A. Farag, A. Shalaby, H. A. El Munim, and A. Farag, “Variational shape representation for modeling, elastic registration and segmentation,” in *Shape Analysis in Medical Image Analysis*, pp. 95–121, Springer, 2014.
14. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active shape models-their training and application,” *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
15. M. Benjelloun, S. Mahmoudi, and F. Lecron, “A framework of vertebra segmentation using the active shape model-based approach,” *Journal of Biomedical Imaging*, vol. 2011, p. 9, 2011.
16. M. A. Larhamam, S. Mahmoudi, and M. Benjelloun, “Semi-automatic detection of cervical vertebrae in X-ray images using generalized hough transform,” in *Image Processing Theory, Tools and Applications (IPTA), 2012 3rd International Conference on*, pp. 396–401, IEEE, 2012.
17. M. Roberts, T. F. Cootes, and J. E. Adams, “Vertebral morphometry: semiautomatic determination of detailed shape from dual-energy X-ray absorptiometry images using active appearance models,” *Investigative Radiology*, vol. 41, no. 12, pp. 849–859, 2006.
18. M. Roberts, E. Pacheco, R. Mohankumar, T. Cootes, and J. Adams, “Detection of vertebral fractures in DXA VFA images using statistical models of appearance and a semi-automatic segmentation,” *Osteoporosis International*, vol. 21, no. 12, pp. 2037–2046, 2010.
19. M. G. Roberts, T. F. Cootes, and J. E. Adams, “Automatic location of vertebrae on DXA images using random forest regression,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pp. 361–368, Springer, 2012.
20. P. Bromiley, J. Adams, and T. Cootes, “Localisation of vertebrae on DXA images using constrained local models with random forest regression voting,” in *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*, pp. 159–171, Springer, 2015.
21. S. M. M. R. Al-Arif, M. Gundry, K. Knapp, and G. Slabaugh, “Improving an active shape model with random classification forest for segmentation of cervical vertebrae,” in *Computational Methods and Clinical Applications for Spine Imaging: 4th International Workshop and Challenge, CSI 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers*, vol. 10182, p. 3, Springer, 2017.
22. T. F. Cootes, “Fully automatic localisation of vertebrae in ct images using random forest regression voting,” in *Computational Methods and Clinical Applications for Spine Imaging: 4th International Workshop and Challenge, CSI 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers*, vol. 10182, p. 51, Springer, 2017.
23. S. A. Mahmoudi, F. Lecron, P. Manneback, M. Benjelloun, and S. Mahmoudi, “GPU-based segmentation of cervical vertebra in X-ray images,” in *Cluster Computing Workshops and Posters (CLUSTER WORKSHOPS), 2010 IEEE International Conference on*, pp. 1–8, IEEE, 2010.
24. S. Ruder, “An overview of gradient descent optimization algorithms,” *arXiv preprint arXiv:1609.04747*, 2016.