Supplementary Material
SPNet: Shape Prediction using a Fully Convolutional Neural Network

S M Masudur Rahman Al Arif[1], Karen Knapp[2] and Greg Slabaugh[1]

[1]City, University of London
[2]University of Exeter

This document contains mathematical details and additional supporting results for the submission titled 'SPNet: Shape Prediction using a Fully Convolutional Neural Network'.

1 Level-set Method

In the level set method [1, 2], the shapes are represented implicitly by an auxiliary function, $\Phi(.)$. The shape, $\boldsymbol{S}$, is denoted as the zero-level set of that function:

$$\boldsymbol{S} = \{\boldsymbol{p} | \Phi(\boldsymbol{p}) = 0\}, \tag{1}$$

where $\boldsymbol{p} \in \Omega_p$ and Ω_p is pixel-space over which the function is defined. The function, $\Phi(.)$, is a signed distance function (SDF) which is defined as:

$$\Phi(\boldsymbol{p}) = \begin{cases} d(\boldsymbol{p}, \boldsymbol{S}) & \text{if } \boldsymbol{p} \in \Omega_v^c \\ -d(\boldsymbol{p}, \boldsymbol{S}) & \text{if } \boldsymbol{p} \in \Omega_v \end{cases}, \tag{2}$$

where Ω_v is the set of pixels inside the object, which is a vertebra in our case, ^{c} represents the complement set, and d is defined as:

$$d(\boldsymbol{p}, \boldsymbol{S}) = \inf_{\boldsymbol{x} \in \boldsymbol{S}} D(\boldsymbol{p}, \boldsymbol{x}), \tag{3}$$

where inf denotes infimum and $D(\boldsymbol{a}, \boldsymbol{b})$ denotes the Euclidean distance between pixel position $\boldsymbol{a}$ and $\boldsymbol{b}$.

2 Conversion of Manual Annotations to SDFs

As mentioned in the paper, the manually annotated vertebral boundaries are available for all the 26,370 training vertebrae. Manual annotation for each of these vertebrae is converted into a signed distance function (SDF). To convert the vertebral shapes into an SDF, the pixels lying on the manually annotated vertebral boundary curve have been assigned zero values. Then all other pixels are assigned values according to Eqn. 2, where $\boldsymbol{S}$ represents the set of pixels with zero values. A few examples of training vertebrae with corresponding zero-level set pixels and SDFs are illustrated in Fig. 1.

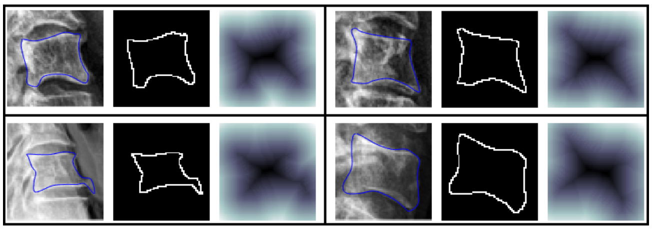

Fig. 1: Examples of training vertebrae: original image (left) with manually annotated vertebral boundaries (−), pixels at the zero-level set of the SDF (center) and the SDF (right). Darker tone represents negative values.

3 Conversion of SDFs to Shape Parameters

Once all the training vertebral shapes are converted to corresponding signed distance functions, we can apply principal component analysis on the SDFs. First, we compute the mean SDF, $\bar{\Phi}$, as:

$$\bar{\Phi} = \frac{1}{N}\sum_{n=1}^{N}\Phi_n, \tag{4}$$

where N is the number of training samples. We then extract the difference SDF (Φ_{d_n}) by subtracting the mean ($\bar{\Phi}$) from each SDF (Φ_n):

$$\Phi_{d_n} = \Phi_n - \bar{\Phi}. \tag{5}$$

The vectorized Φ_{d_n} are then arranged in a matrix, M:

$$\boldsymbol{\phi}_{d_n} = vec(\Phi_{d_n}), \tag{6}$$

$$M = [\boldsymbol{\phi}_{d_1}|\boldsymbol{\phi}_{d_2}|....|\boldsymbol{\phi}_{d_N}]. \tag{7}$$

The covariance matrix, C_M can then be computed as:

$$C_M = \frac{1}{N}MM^T. \tag{8}$$

The principal components of the variations of the training data can be extracted by singular value decomposition (SVD) of the matrix C_M:

$$[W, \Sigma, W_v^T] = svd(C_M), \tag{9}$$

where Σ is a diagonal matrix containing eigenvalues corresponding to the eigenvectors, which are arranged in a column-wise manner in W. The eigenvectors are sequentially arranged based on their corresponding eigenvalues. Now, each shape in the training data can be represented by the mean shape ($\bar{\boldsymbol{\phi}}$), matrix of eigenvectors (W) and a vector of shape parameters, $\boldsymbol{b}_n$:

$$\boldsymbol{\phi}_n = \bar{\boldsymbol{\phi}} + W\boldsymbol{b}_n. \tag{10}$$

For each training example we can compute $\boldsymbol{b}_n$ as:

$$\boldsymbol{b}_n = W^T(\boldsymbol{\phi}_n - \bar{\boldsymbol{\phi}}) = W^T\boldsymbol{\phi}_{d_n}. \tag{11}$$

Table 1: Dimensionality of different matrices and vectors.

Dimension	Matrix/Vector
64×64	$\varPhi_n$, $\bar{\varPhi}$, $\varPhi_{d_n}$
$4096 \times N$	M
4096×1	$\boldsymbol{\phi}_n$, $\bar{\boldsymbol{\phi}}$, $\boldsymbol{\phi}_{d_n}$, $\boldsymbol{b}_n$
4096×4096	C_M, W, V, U

These parameters are used as the ground truth ($\boldsymbol{b}_n^{GT}$) for training the proposed network. For simplicity, the mathematics in the original article is described for a single input image patch and the subscript $_n$ has been dropped. For SDFs defined over a pixel space of size 64×64 and a training dataset with N samples, the dimensionality of the matrices and vectors discussed in this section are summarized in Table 1.

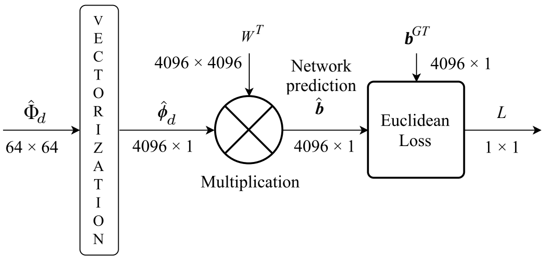

Fig. 2: Final layer.

4 Differentiability of the Proposed Loss Layer

The forward pass through the final layer (Fig. 2) can be summarized below. First, the output of the last convolutional layer of the SPNet ($\hat{\varPhi}_d$) is vectorized:

$$\hat{\boldsymbol{\phi}}_d = vec(\hat{\varPhi}_d). \tag{12}$$

Then the final prediction of network is computed as $\hat{\boldsymbol{b}}$:

$$\hat{\boldsymbol{b}} = W^T \hat{\boldsymbol{\phi}}_d, \tag{13}$$

or in element-wise form:

$$\hat{b}_i = \sum_{j=1}^{k} w_{ij} \hat{\phi}_{d_j}, \tag{14}$$

where w_{ij} is the value at the i-th row and j-th column of the transposed eigenvector matrix (W^T) and k is the number of eigenvectors. Finally, the loss is defined as:

$$L = \sum_{i=1}^{k} L_i, \tag{15}$$

where

$$L_i = \frac{1}{2}(\hat{b_i} - b_i^{GT})^2. \tag{16}$$

For back-propagation, the partial derivative of Eqn. 16 with respect to the input variable $\hat{b}_i$ can be expressed as:

$$\frac{\partial L_i}{\partial \hat{b}_i} = \hat{b_i} - b_i^{GT}. \tag{17}$$

Similarly, the partial derivative of Eqn. 14 with respect to the input, $\hat{\phi}_{d_j}$, can be expressed as:

$$\frac{\partial \hat{b}_i}{\partial \hat{\phi}_{d_j}} = w_{ij}. \tag{18}$$

5 Metrics

Two metrics are used to compare the predicted shapes with the ground truth shapes: the average point to ground truth curve error (E_{p2c}) and the Hausdorff distance (d_H) [3] between the prediction and ground truth shapes. These metrics are defined in Eqn. 19 and 20. The E_{p2c} represents on average how far the predicted shape points are from the ground truth, the second metric (d_H) denotes the maximum difference between the shapes. Both metrics are reported in pixels.

$$E_{p2c}(\hat{S}, S_{gt}) = mean\big\{min\{D(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{y} \in S_{gt}\} : \boldsymbol{x} \in \hat{S}\big\}, \tag{19}$$

$$d_H(\hat{S}, S_{gt}) = max\{\sup_{\boldsymbol{x}\in\hat{S}} \inf_{\boldsymbol{y}\in S_{gt}} D(\boldsymbol{x}, \boldsymbol{y}), \sup_{\boldsymbol{y}\in S_{gt}} \inf_{\boldsymbol{x}\in\hat{S}} D(\boldsymbol{x}, \boldsymbol{y})\}, \tag{20}$$

where $\hat{S}$ is the set points in the predicted shape, S_{gt} is the set of points in the manually annotated vertebral boundary curve, sup represents the supremum, inf represents the infimum and $D(\boldsymbol{x}, \boldsymbol{y})$ is the Euclidean distance between the point $\boldsymbol{x}$ and $\boldsymbol{y}$.

6 Parameters for the Chan-Vese Method

The Chan-Vese level set segmentation method (LS-CV) [2, 4] is a parametric method and finding a common set of parameters for the challenging X-ray dataset was difficult. A grid search method was followed to find a common set of parameters on a separate validation set of 40 images with 177 vertebrae. We also had to constrain the b-parameters to 1.2 times the standard deviation to produce acceptable results for this method.

7 Additional Results

7.1 Selection of the Number of Shape Parameters

Both of the initial shape predictor deep networks, SPNet and SP-FCNet, have been trained to regress all 4096 shape parameters. These parameters are related to the 4096 eigenvectors or modes of variations. The eigenvalues represent the variance in the training data along the corresponding eigenvectors. As the eigenvectors are ranked based on their eigenvalues, eigenvectors with small eigenvalues often result from noise and can be ignored. In Table 2, we report performance of our proposed SPNet on the validation set of 177 vertebrae when we consider a certain percentage of total variation at test time. The second row of the table indicates how many parameters are left when a certain percentage of variation is considered. Other parameters are replaced with zeros when converting back to the signed distance function. It can be seen that the lowest errors are found when 98% of the total variation is considered and only 18 b-parameters are kept. Based on this insight, both versions of our deep networks were modified and retrained to regress only 18 b-parameters.

7.2 Statistical Significance Test for SPNet-18

Table 3 reports the results of the statistical significance test between our proposed SPNet-18 and all other methods reported in Table 2 of the original article. It can be seen that the quantitative improvement of SPNet-18 over UNet-S in terms of the E_{p2c} metric is not statistically significant according to the paired t-test at a 5% significance level. However, the improvement in terms of the Hausdorff distance (d_H) passes the significance test.

7.3 Results with Fully Automatic Framework

Recently, the authors of the UNet-S paper [5] have proposed a fully automatic method for vertebral patch extraction [6]. This method has been able to extract 80% of our test vertebrae correctly. We then applied our proposed method on these extracted vertebra patches. The results are reported in Table 4 and shown in Fig. 3b. It should be noted that these errors are lower, because the fully automatic method fails to extract about 20% of the test vertebra in challenging situations.

These results also demonstrate our method's capability of adapting to minor variation in the scale and orientation of the extracted vertebral patches.

7.4 Additional Qualitative Results

Additional to qualitative results reported in the original article, here in Fig. 4, we show more examples from the test dataset.

Table 2: Effect of eigenvectors on errors for SPNet.

Variation (%)	90	95	98	99	99.5	99.8	100
No. of parameters	6	9	18	30	51	117	4096
Average E_{p2c}	1.34	1.23	**1.16**	1.19	1.21	1.23	1.25
Average d_H	4.83	4.62	**3.98**	4.15	4.31	4.49	4.68

Table 3: Statistical significance test (t-test).

SPNet-18 compared with:	Average E_{p2c}		Average d_H	
	h	p-value	h	p-value
LS-CV	1	$< 10^{-282}$	1	0
UNet	1	0.003	1	$< 10^{-05}$
UNet-S	0	0.827	1	0.035
SP-FCNet-18	1	$< 10^{-255}$	1	$< 10^{-151}$

Table 4: Results with fully automatic patch extraction process.

Metrics	Average E_{p2c}		Average d_H		$nVmR$	$FF\%$
Methods	Mean	Std	Mean	Std		
UNet-S	0.931	0.492	4.328	4.281	14	3.388
SPNet-18	**0.923**	**0.345**	**4.232**	**2.933**	**0**	**1.882**

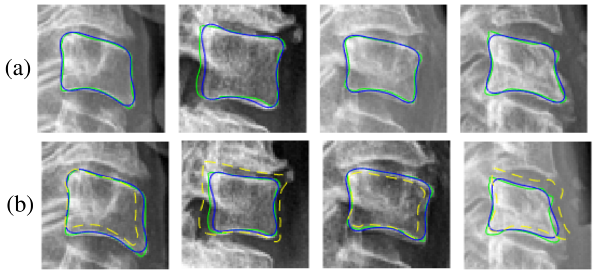

Fig. 3: (a) Results from the original test dataset (b) results for the same vertebra using fully automatic framework: Ground truth shape (—), predicted shape (—) and vertebra position in the original test dataset (- -).

References

1. M. E. Leventon, W. E. L. Grimson, and O. Faugeras, “Statistical shape influence in geodesic active contours,” in *International Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 316–323, IEEE, 2000. 1
2. A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky, “A shape-based approach to the segmentation of medical imagery using level sets,” *IEEE Transactions on Medical Imaging*, vol. 22, no. 2, pp. 137–154, 2003. 1, 4
3. D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, “Comparing images using the hausdorff distance,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, 1993. 4
4. T. F. Chan and L. A. Vese, “Active contours without edges,” *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001. 4
5. S. M. M. R. Al-Arif, K. Knapp, and G. Slabaugh, “Shape-aware deep convolutional neural network for vertebrae segmentation,” in *5th International Workshop and Challenge in Computational Methods and Clinical Applications for Musculoskeletal Imaging (MICCAI MSKI)*, pp. 12–24, Springer, 2018. 5

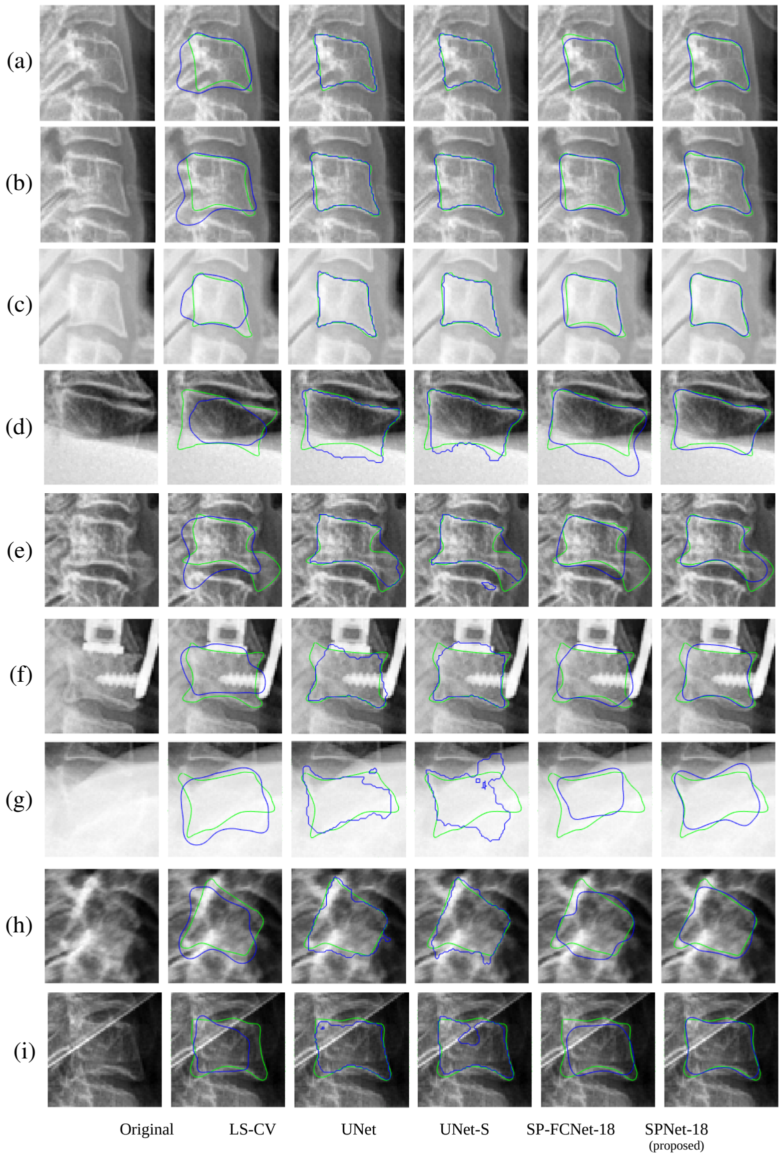

Fig. 4: Qualitative results. The predicted shape is plotted in blue (–) and the ground truth in green (–).

6. S. M. M. R. Al-Arif, K. Knapp, and G. Slabaugh, “Fully automatic cervical vertebrae segmentation framework for x-ray images,” *Elsevier Computer Methods and Programs in Biomedicine*, vol. 157, pp. 95–111, 2018. 5