Augmenting Auto-context with Global Geometric Features for Spinal Cord Segmentation

Jeremy Kawahara[1], Chris McIntosh[1,2], Roger Tam[3], and Ghassan Hamarneh[1]

[1]Medical Image Analysis Lab., Simon Fraser University, Burnaby, Canada
`{jkawahar,cmcintos,hamarneh}@sfu.ca,`
[2]Princess Margaret Cancer Centre, University Health Network, Toronto, Canada
[3]MS/MRI Research Group, University of British Columbia, Vancouver, Canada
`roger.tam@ubc.ca`

Abstract. Anatomical shape variations are typically difficult to model and parametric or hand-crafted models can lead to ill-fitting segmentations. This difficulty can be addressed with a framework like auto-context, that learns to jointly detect and regularize a segmentation. However, mis-segmentation can still occur when a desired structure, such as the spinal cord, has few locally distinct features. High-level knowledge at a global scale (e.g. an MRI contains a single connected spinal cord) is needed to regularize these candidate segmentations. To encode high-level knowledge, we propose to augment the auto-context framework with global geometric features extracted from the detected candidate shapes. Our classifier then learns these high-level rules and rejects falsely detected shapes. To validate our method we segment the spinal cords from 20 MRI volumes composed of patients with and without multiple sclerosis and demonstrate improvements in accuracy, speed, and manual effort required when compared to state-of-the-art methods.

1 Introduction

Studies of multiple sclerosis (MS) show that spinal cord atrophy strongly contributes to the physical disability of a patient, motivating the need for quantifiable spinal cord measurements to evaluate the progression of pathology and the effectiveness of therapies [8]. Spinal cord segmentation is an important first step to extracting these measurements. However, segmenting the cord is challenging as it has an irregular cross-sectional shape whose signal and diameter change over the length of the cord. The image can be polluted by noise, suffer from motion artefacts, and its boundaries can be poorly defined where the cord contacts the spinal canal wall. When scanned at the most common resolution of 1 mm^3, the cord also contains a high number of partial volume voxels relative to the total size of the cord, which need to be accounted for in order to accurately measure the cord's true shape and volume [10]. Manual segmentation is a challenging and time consuming process that suffers from operator variability, making automated methods desirable. To address these challenges, several approaches have been proposed for automating spinal cord segmentation.

McIntosh et al. [6] segmented the cord using locally optimal 3D deformable organisms guided by a Hessian-based filter designed to adapt to the spinal cord's varying elliptical and tubular structure. They later extended this approach to use the cord's medial axis (found by a user-guided live-wire method) to guide the spinal crawler's cross-sectional shape fitting [7]. Horsfield et al. [3] fit an active surface model to the spinal cord and had the user specify the cord center-line by placing points on representative slices. Chen et al. [1] used a deformable atlas-based registration combined with a topology preserving classification to fully automate a crisp segmentation of the cord. In previous work [4], we represented the spinal cord's axial shape variations using probabilistic principal component analysis (PCA) and found the globally optimal path in 6D (three spatial and three principal components weights) between two user specified seed points.

Auto-context, proposed by Tu and Bai [11], is a general iterative learning framework used for segmentation that jointly learns the appearance and regularization distributions where the predicted class labels (the context) of the previous iteration are used as input to the current iteration. Auto-context was shown to improve segmentation results [11]; however, if the surrounding local context is incorrect it may re-enforce an incorrect segmentation. This was addressed in the recent work by Kontschieder et al. [5] who used a geodesic distance transform in an auto-context based segmentation approach to incorporate long range spatial context. Similarly, we extend auto-context to learn high-level problem specific information (global context), but rather than considering the geodesic distance, our probability mask represents distinct candidate shapes and encodes their relations to each other to encourage a single connected spinal cord.

Specifically, we propose to extract geometric features (e.g. volume) from the candidate shapes created in each iteration of auto-context. We define a shape as a distinct region of connected components composed of the same class label. Features from the candidate shapes are extracted and compared with each other to give global information about the other candidate shapes and are included in the auto-context framework. Thus the class label for a specific voxel is conditioned not only on its appearance and surrounding labels, but also on the geometric features of the shape it belongs to relative to other candidate shapes. We demonstrate that augmenting auto-context with global geometric context improves the original auto-context algorithm, and results in superior spinal cord segmentations when compared to a more traditional approach using hand-crafted gradient and intensity-based terms regularized by a PCA shape prior.

2   Methods

This section introduces auto-context, describes our global geometric features, explains our auto-context set-up, and outlines the training and testing processes.

Image Segmentation and Auto-context. We can think of image segmentation as a voxel labelling problem, where given an image $X = (x_1, \ldots, x_n)$ composed of n voxels, we want to find a set of labels such that each voxel i is

assigned a corresponding label, $Y = (y_1, \ldots, y_n)$. The label y_i can take on one of k possible values. Our objective is to find the optimal configuration Y^* out of the possible segmentations Y such that it maximizes the probability given the observed image, $Y^* = \text{argmax}_Y \, p(Y|X)$.

One approach to find Y^* is to apply Bayes' rule with a fixed $p(X)$ to give $p(Y|X) \propto p(X|Y)p(Y)$, where $p(X|Y)$ represents the likelihood of the image given a segmentation, and $p(Y)$ is the probability of that segmentation occurring. Hand-crafted data-terms and/or parametric models with restrictions on complexity and built-in assumptions (e.g. Gaussian) are often used [4, 6, 7].

Another approach is to directly model $p(Y|X)$. If we assume that the labels $y_i, i = 1, \ldots, n$ are independent and are conditioned only on a small patch of image voxels centred around the ith voxel, denoted as $N(i)$, then this can be modelled by $p(y_i|X_{N(i)})$, where X_i returns the intensities at voxel i. A discriminative model (e.g. decision forest [2]) can be used to learn this distribution; however, it does not consider the class labels of surrounding voxels. One way to model the interdependence of neighbouring labels is with conditional random fields where the class label y_i is dependent on a neighbouring class label y_j [9]. While this more closely approximates the true $p(Y|X)$, it still makes the assumption that each class label y_i is only dependent on a very small neighbourhood.

The auto-context model seeks to condition over a larger area of surrounding class labels to provide more "context" [11]. This is accomplished by introducing an iterative time-step t and training a series of classifiers using the discriminative probability (the classification confidence) map of the previous classification M^{t-1} where $M^t = (\boldsymbol{m}_1^t, \ldots, \boldsymbol{m}_n^t)$. Each vector $\boldsymbol{m}_i^t$ represents the probabilities of voxel i belonging to one of the k possible class labels, $\boldsymbol{m}_i = [p(y_i = 1), \ldots, p(y_i = k)]$. The initial class probability map M^0 is set to have uniform values. At time t, a classifier is trained to predict the true class label y_i given the image patch $X_{N(i)}$ and the context information $M^{t-1}(i)$, where M is centred at voxel i. Once the classifier is trained, the new probability map M^t is used in the next iteration $(t+1)$ and the algorithm repeats until M converges. The final output is a series of learned probability distributions,

$$m_i^t = p^t(y_i|X_{N(i)}, M^{t-1}(i)). \tag{1}$$

In testing, a novel image has the same features extracted and goes through the iterative classification process using the learned probability distributions p^t.

This formulation does not capture high-level information about the entire shape that the pixel belongs to nor does it consider the interaction among other candidate shapes. If we simply increase the size of $N(i)$, we increase the dimensionality of the feature space which can decrease the efficiency of our classifier. Thus this formulation is not well suited to capture high-level knowledge such as that the spinal cord is a single connected structure of a particular size.

Global Geometric Features. We propose to augment auto-context with global geometric features. At iteration t, instead of only using the previous probability map M^{t-1}, we also extract features from the candidate shapes found

within M^{t-1} and compare these features to each other to capture global information. More formally, we define a function $C(M, i)$ that takes the maximum a posteriori (MAP) class label of M to form a candidate shape S_q based on the regions of connected component with the same class label, $S_q = C(M, i)$. This divides M into Q distinct non-overlapping connected regions with the same class label, where each distinct region represents a candidate shape S_q (Fig. 1c,g). The shape S_q is composed of indices into the probability map M and the index for voxel i belongs to a single shape, $i \in S_q$.

   We define and utilize a single global geometric feature that encourages our method to detect only a single connected spinal cord. For voxel i we extract a feature $f(S_q, M)$ from the shape that i belongs to and compute the ratio between the current and the largest shape feature from the Q candidate shapes,

$$\phi(M, i) = \frac{f(C(M, i), M)}{\max(f(S_1, M), \ldots, f(S_Q, M))}. \tag{2}$$

   The feature we extract, $f(S_q, M)$, gives us an indication of the size and label confidence of the component being considered,

$$f(S_q, M) = \sum_{j \in S_q} M(j) \tag{3}$$

where the iterator j sums the probabilities belonging to shape S_q. If $\phi(M, i)$ returns a value of 1, then this indicates that voxel i belongs to the largest probable shape out of all the candidate shapes. We augment the previous auto-context model (1) with our global geometric feature ϕ,

$$m_i^t = p^t(y_i | X_{N(i)}, M_{N(i)}^{t-1}, \phi(M^{t-1}, i)). \tag{4}$$

This model considers *intensity information*, *local context*, and the proposed *global features* about the candidate shapes and the relations between them.

Auto-context Setup and Feature Design The original auto-context work used probabilistic boosting trees as the discriminative classifier, but other classifiers can be used [11]. We chose to use decision forests for our discriminative model due to their ability to generalize well to unseen data, handle both classification and regression, learn a distribution, and provide a probabilistic output [2]. This probabilistic score $\boldsymbol{m}_i$ is particularly important because it gives us an indication of the probability of being a member of the class by computing the percentage of observations of this class in a tree leaf averaged over all trees.

   The neighbourhood of the ith voxel, $N(i)$, is computed using a simple radial pattern where we sample those voxels that are the immediate neighbours (8 in 2D, 26 in 3D) of voxel i and those that are three voxels along the ray away from voxel i (Fig. 1e). This densely samples points close to our ith voxel and sparsely samples voxels further away which helps keep the size of the feature vector relatively small to allow for faster run-times. We use a 52-neighbour 3D version of the neighbourhood shown in Fig. 1e. To compute image appearance
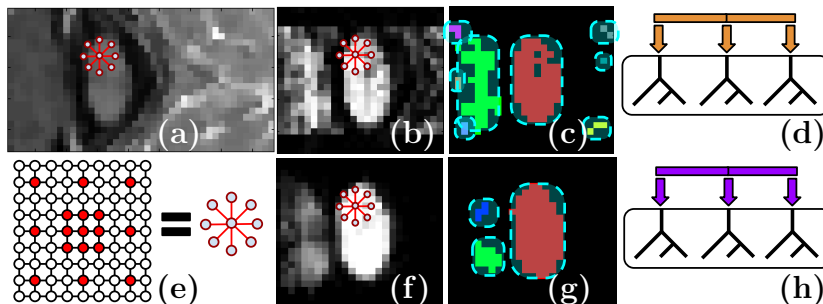
Fig. 1. An outline of our method: (a) The MRI data X. (b) The probability map M^t. (c) The MAP estimate of M produces distinct shapes (*dotted outline*) whose features $\phi(M, i)$ can be extracted. (d) Decision forest trained on a,b,c. (e) The neighbourhood $N(i)$ in 2D. (f) The probability map M^{t+1} produced by d. (g) Shapes computed from f. (h) Decision forest trained on a,f,g.

features we directly sample the neighbourhood intensities. In order to have an invariance to a shift in intensities and to capture the polarity (e.g. dark-to-bright transitions), we divide the intensity of neighbourhood voxels by the intensity of the voxel of interest, $X_{N(i)} = \{\frac{x_1}{x_i}, \ldots, \frac{x_{52}}{x_i}\}$.

Training and Testing Since the two patient groups should not be modelled as samples from a single distribution, we separate the multiple sclerosis (MS) from the non-MS patients to train two different classifiers specifically on the variability found in the two groups. We employ leave-one-out testing.

Our auto-context model is trained based on (4). Our training segmentations have a value of 1 inside the cord, 0 outside the cord, and a partial volume estimate for the cord's boundary. Each decision forest p^t is trained on a subset of this training data (2000 voxels from each volume) split between cord and background samples with the fuzzy border voxels omitted and samples of false positives explicitly included if they exist. To reduce over-fitting, for each training volume V, a separate decision forest that does not include the data from V is trained and used to predict the class membership of V. We repeat this process 5 times as we found that the results generally stabilized without further training (less than 0.01 difference in the mean Jaccard similarity index between consecutive iterations). The training produces a set of 5 trained decision forests, $\{p^1, \ldots, p^5\}$.

As our ground truth segmentations have a fuzzy border based on partial volume estimation not modelled by the decision forests, we train a separate regression forest p^r to capture a two-pixel thick partial volume border using the same intensity and probability features. It can take up to 6 hrs to fully train our auto-context model on a machine using 4 cores.

In the testing phase, a novel volume is classified by all the trained decision forests, $\{p^1, \ldots, p^5\}$, in series. We take the maximum a posterior estimate of our final classified volume M^5 and use the regression forest trained on border partial volumes p^r to produce our final segmentation (Fig. 2i,j).
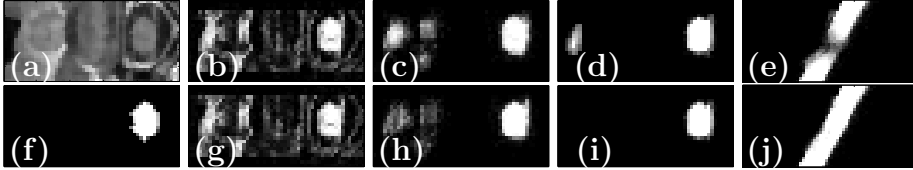
Fig. 2. Segmentation without global features *row 1*, and with global features *row 2*. (a) An axial slice of the cropped spinal cord. (f) The ground truth. (b,g) First iteration over-detects the cord. (c,h) Second iteration begins to regularize the segmentation. (d) Final fuzzy border segmentation with a false positive. (i) False positive removed by the global features. (e) Sagittal plane where weak appearance information splits the cord. (j) Corrected with global features.

3   Results

We validate our method using 20 MRI scans composed of 10 MS patients from a 1.5T scanner and 10 healthy patients from a 3.0T scanner (scans from different studies) with a voxel size of $0.976 \times 0.976 \times 1.000$ mm. Each scan was segmented by an expert who used an in-house method similar to that by Tench et al. [10]. We segment over vertebra C3 - C7 spanning a total of 80 slices.

To capture the partial volume effects (PVE) [10], we use a probabilistic version of the Jaccard index defined in [4] as, $J_{\mathrm{PVE}}(A, G) = \frac{\sum_x \sum_y \min(A(x,y), G(x,y))}{\sum_x \sum_y \max(A(x,y), G(x,y))}$ where A and G are the automated and ground truth segmentations. To highlight our methods improved ability to remove false positives, we compute the Hausdorff distance which measures the furthest distance between the closest points in the two segmentations, $H = \max_{a \in A} \left(\min_{g \in G} D(a, g) \right)$, where $D(a, g)$ computes the 3D Euclidean distance between points a, g. To measure the similarity between the automated and manual segmentation volumes, we calculate, $Vol_{\mathrm{S}} = 1 - \min(|(|A|_1/|G|_1) - 1|, 1)$ which returns a 0% accuracy if the automated volume overestimates the expert volume by more than 200% [7]. To indicate if our automated method is a useful consistent substitute to the manual method for computing spinal cord volume, we compute the Pearson's correlation coefficient between the volumes of the automated and manual segmentations.

We perform three experiments to validate our method. Our **first experiment**, compares our auto-context augmented with the global features method against a segmentation method we previously developed [4]. This previous work modelled the cord using hand-crafted gradient and intensity-based terms and a probabilistic PCA shape prior. Our new results compare favourably, with a mean J_{PVE} of 0.878 (previous 0.784) for MS and 0.928 (previous 0.832) for non-MS scans, validated over the same dataset. Method [4] had been favourably compared with [6, 7] on the same dataset.

Our **second experiment** is done to better understand our improvements and to highlight the problems with restricted parametric models and hand-crafted terms. To examine the effect that the restricted parametric PCA model has on the segmentation, we pass only the *ground truth* data (i.e. the cord is per-

fectly detected) to be regularized by the PCA model and A* optimizer of [4]. We find that even with the ground truth data, the restricted PCA model produces less accurate J_{PVE} results when compared to our auto-context model (Table 1 - row GT_{PCA}). To show the limitations of hand-crafted terms, we replace the gradient and intensity-based terms of [4] with the spinal cord probability map M^1 (trained only on intensity features) as input to be regularized by the PCA model. This improves the segmentation results (Table 1 - row PCA_{M}) over what was originally reported in [4] indicating that the trained classifier is better at detecting the cord.

Our **third experiment** compares auto-context with and without global features. To compare to the approach of Tu and Bai [11], we initially omit the global feature and rely only on the image X and probability maps M as input to the auto-context model to segment the spinal cord (Table 1 - row XM). We then reintroduce the global feature $\phi(M, i)$ and show improved segmentation results (Table 1 - row $XM\phi$). This demonstrates that it is useful to augment auto-context with the global shape feature (i.e. improvement over [11]).

Table 1. GT_{PCA} uses the *ground truth* as input to [4] to demonstrate the best the restricted PCA model can achieve. PCA_{M} uses the spinal cord probability map M^1 as input to the PCA-based method of [4]. XM is auto-context with the image and probability map features. $XM\phi$ is auto-context with the image, probability and global geometric feature. We compute the mean values of the probabilistic Jaccard index ($\overline{J_{\mathrm{PVE}}}$), Hausdorff distance ($\overline{H}$) and area similarity ($\overline{Vol_{\mathrm{S}}}$) in voxels between our auto-mated and ground segmentations over ten MS patients (*left columns*) and ten healthy (non-MS) patients (*right columns*). The Pearson's correlation coefficients (r) and p-values (p) between the automated and the ground truth segmentations are measured.

Case	$\overline{J_{\mathrm{PVE}}}$	$\overline{H}$	$\overline{Vol_{\mathrm{S}}}$	r	p	$\overline{J_{\mathrm{PVE}}}$	$\overline{H}$	$\overline{Vol_{\mathrm{S}}}$	r	p
GT_{PCA}	0.853	1.96	98.36	0.997	3×10^{-10}	0.881	1.57	98.97	0.993	9×10^{-9}
PCA_{M}	0.818	2.42	**96.38**	0.931	9×10^{-5}	0.847	2.04	93.65	0.896	4×10^{-4}
XM	0.856	15.35	94.50	0.932	9×10^{-5}	0.916	4.41	97.74	0.939	6×10^{-5}
$XM\phi$	**0.878**	**2.29**	96.32	**0.972**	$\mathbf{3\times10^{-6}}$	**0.928**	**1.82**	**98.87**	**0.991**	$\mathbf{3\times10^{-8}}$

We note that, once trained, our unoptimized implementation takes under 10 minutes to run for a novel cropped volume while [4] reported run-times between 1 and 5 hours for the same sized volumes. Also, while our method was tested on a cropped volume (Fig. 1a), it did not require any further user-input in contrast to [7, 3], and it was capable of capturing the PVE not modelled in [1].

4   Conclusion

We have proposed to augment auto-context with global geometric features that can capture high-level information and relationships between the candidate shapes found within a segmentation. The auto-context approach jointly detects and

regularizes a segmentation which allows for a flexible shape space capable of capturing subtle irregularities in the spinal cord. We have demonstrated improvements to accuracy and a lower running time when compared to a recently proposed method. Future work would validate our method over a larger spinal cord dataset with more diverse clinical parameters and investigate the development of other global geometric features.

References

1. Chen, M., Carass, A., Cuzzocreo, J., Bazin, P.L., Reich, D.S., Prince, J.L.: Topology preserving automatic segmentation of the spinal cord in magnetic resonance images. In: IEEE ISBI. pp. 1737–1740 (2011)
2. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Foundations and Trends® in Computer Graphics and Vision 7(2-3), 81–227 (2011)
3. Horsfield, M.A., Sala, S., Neema, M., Absinta, M., Bakshi, A., Sormani, M.P., Rocca, M.A., Bakshi, R., Filippi, M.: Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: Application in multiple sclerosis. Neuroimage 50(2), 446–455 (2010)
4. Kawahara, J., McIntosh, C., Tam, R., Hamarneh, G.: Globally optimal spinal cord segmentation using a minimal path in high dimensions. In: IEEE ISBI. pp. 836–839 (2013)
5. Kontschieder, P., Kohli, P., Shotton, J., Criminisi, A.: GeoF: Geodesic forests for learning coupled predictors. In: IEEE CVPR (2013)
6. McIntosh, C., Hamarneh, G.: Spinal crawlers: Deformable organisms for spinal cord segmentation and analysis. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006, LNCS, vol. 4190, pp. 808–815. Springer, Heidelberg (2006)
7. McIntosh, C., Hamarneh, G., Toom, M., Tam, R.: Spinal cord segmentation for volume estimation in healthy and multiple sclerosis subjects using crawlers and minimal paths. In: IEEE HISB. pp. 25–31 (2011)
8. Rocca, M., Horsfield, M., Sala, S., Copetti, M., Valsasina, P., Mesaros, S., Martinelli, V., Caputo, D., Stosic-Opincal, T., Drulovic, J., Comi, G., Filippi, M.: A multicenter assessment of cervical cord atrophy among MS clinical phenotypes. Neurology 76(24), 2096–2102 (2011)
9. Szummer, M., Kohli, P., Hoiem, D.: Learning CRFs using graph cuts. In: D. Forsyth, P.T., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 582–595. Springer, Heidelberg (2008)
10. Tench, C.R., Morgan, P.S., Constantinescu, C.S.: Measurement of cervical spinal cord cross-sectional area by MRI using edge detection and partial volume correction. J. Magn. Reson. Imaging 21(3), 197–203 (2005)
11. Tu, Z., Bai, X.: Auto-context and its application to high-level vision tasks and 3D brain image segmentation. IEEE TPAMI 32(10), 1744–1757 (2010)